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

Permanent WRAP URL:

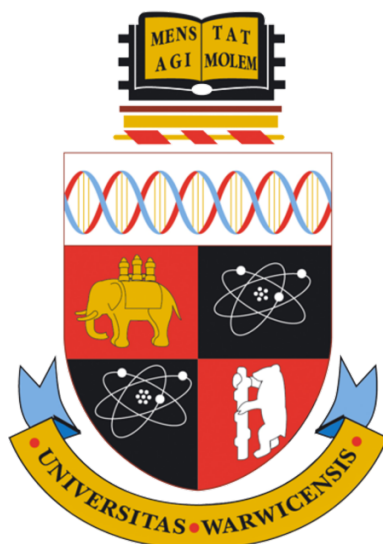http://wrap.warwick.ac.uk/175062

warwick.ac.uk/lib-publications

Exploring epoxyketone synthases and their biosynthetic potential

Marlene Luise Rothe

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

University of Warwick
Department of Chemistry
September 2022

Contents

List of figures

List of schemes

List of tables

Acknowledgements

First and foremost, I would like to thank Prof. Dr Greg Challis for the outstanding opportunity to be part of his research group and to work on this project. I am very grateful for all the experiences I had, for everything I have learned and for all the support I received. I would like to express my sincere thanks to my co-supervisor Prof. Dr Fabrizio Alberti for hosting me at the SLS laboratories in the beginning of my PhD and after lockdown and for his mentorship and constant support. I sincerely thank Lona for all her support throughout my PhD and her continued guidance, especially during the final weeks of writing my thesis.

I would like to thank my advisory panel, Dr David Fox and Dr Christophe Corre, and all supporting staff of the Department of Chemistry, including Dr Lijiang Song, Lynette Walsh, Dr Ivan Prokes, ChemTech, and chem stores for their constant help and hard work. I would like to thank the SynBio CDT for the generous funding, support, and opportunities. Thank you to all my CDT friends for being there throughout the whole process and very special thanks go to my Warwick cohort friends Muna, Jack, and Laurence for making this time so enjoyable and memorable.

I am also thanking Dr Matt Jenner, Dr Jinlian Zhao, and Munro Passmore for their help with mass spectrometry experiments and Dr Chuan Huang and Yu Zhang for their work with the tryptopeptin clusters. I also am especially grateful to Munro and Mia Foran for their outstanding help and support in the biology lab, thank you! I would like to thank all members, past and present, of the Corre and Alberti group who I met. I am very grateful for how kindly I was welcomed in SLS by Rohini, Jack H, Dan Ro, Dan T, Patrick, and Valentin and appreciate all the kindness, smiles, support, and help I received from Jack W, Jingfan, Richard, Duha, Paula, Pan, Micky, Sophie, Sopida, Dan Ri, and Matt over the past few years. I would like to thank all Challis group members, past and present, who I had the pleasure of meeting. Thank you to Callum who was there in the very beginning, thank you to Shanshan, Chuan, Xinyun, Jinlian, Chris F, Arun, Emzo, Doug, Chris, Christian, Maz, Jake, Ioanna, Matt B, Helen, Fang, Fayrouz, Barbara, Yu, and Mabilly for being there and for everything I have learnt from you. Special thanks to Munro, Mia, James, and Dan V for joining the group and making the time in the lab, office, and in Edinburgh so fun. Thank you for lovely chats, pub lunches, all the help, and for brightening up every day on campus!

I'm struggling to find words to express how grateful I am to Becky, for always being there and for being such a great companion throughout this journey. I could not dream of anyone else I would have wanted to go through the past few years together with and I am so grateful to have found such a wonderful friend, thank you for everything!

I would also like to thank Warwick Mountains for all the fun adventures and being such an amazing club and group of lovely people. I sincerely thank Kenilworth Art Society for such a kind and supportive welcome into the group and for all the creative and lovely times together.

Finally, I would like to thank all my wonderful friends and my family for their constant kindness and amazing friendships throughout all my adventures, including incredible support during my PhD. Special thanks go to my Kenilworth friends, thank you for all the late nights, cakes, swims, and for making my time here as amazing and fun as it was! I couldn't have done it this happy without any of you and I have no words to express my gratitude and love to all of you. I am especially grateful for my wonderful parents and thank them for always being there for me and for their constant love and support. Thank you for everything!

Declaration

I confirm that the presented work has been prepared in accordance with the guidelines of the University of Warwick on the presentation of a research thesis for the degree of Doctor of Philosophy. The submitted thesis has been composed by myself and all included work is original research carried out by me, unless stated otherwise and except the outlined cases below:

> In Chapter 4.2.3 and 4.3.2, the *ttp* and *tpt* clusters were cloned, transformed, and heterologously expressed by Dr Chuan Huang.

> In Chapter 4.3.2, the metabolite expression and MS/MS analysis was carried out by Yu Zhang.

No material has been submitted to any other application of any degree, or to any other institution. Results from other researchers and contributions from collaborators are referenced appropriately throughout the thesis.

Marlene Luise Rothe   _21ˢᵗ September 2022_

Marlene Luise Rothe   Date

Abbreviations

°C	Celsius
A	Adenylation (domain)
Aa	Amino acid
Ac	Acetyl
ACAD	Acyl-CoA dehydrogenase
ACP	Acyl carrier protein
Amp	Ampicillin
AMP	Adenosine monophosphate
APS	Ammonium persulfate
AT	Acyl transferase
ATP	Adenosine triphosphate
BGC	Biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
Bn	Benzyl
Boc	*tert*-Butyloxycarbonyl protecting group
bp	Base pairs
Bu	Butyl
BUSCO	Benchmarking Universal Single-Copy Orthologs
C	Condensation (domain)
Cbz	Carboxybenzyl
CDI	Carbonyl diimidazole
cMT	Carbon-methyltransferase
CoA	Coenzyme A
COMU	(1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)-dimethylamino-morpholino-carbenium hexafluorophosphate

COSY	Correlation spectroscopy
Cpa	Cyclopentenylalanine
DCC	*N,N'*-Dicyclohexylcarbodiimide
DCM	Dichloromethane
dH$_2$O	Deionized water
DIEA	*N,N*-Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dppf	1,1'-Bis(diphenylphosphino)ferrocene
dr	Diastereomeric ratio
E	Epimerization (domain)
EDC	1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted ion chromatogram
equiv.	Equivalent
ESI	Electrospray ionization
Et	Ethyl
EWG	Electron withdrawing group
FAD	Flavin adenine dinucleotide
FAS	Fatty acid synthase
FMN	Flavin mononucleotide
HMBC	Heteronuclear multiple bond correlation
HMM	Hidden Markov model
HOBt	Hydroxybenzotriazole

HPLC	High performance liquid chromatography
HR	High resolution
HSQC	Heteronuclear single quantum coherence
IMAC	Immobilized metal ion affinity chromatography
IMWG	International Myeloma Working Group
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infra-red
Kan	Kanamycin
KS	Ketosynthase
LB	Luria-Bertani (medium)
LC	Liquid chromatography
MALDI	Matrix assisted laser desorption/ionization
Me	Methyl
MLP	MbtH-like protein
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MT	Methylation (domain)
MYG	Malt extract, yeast extract, glucose
m/z	Mass to charge ratio
NADH	Reduced nicotinamide adenine dinucleotide
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NMI	*N*-methylimidazole
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthetase

OD$_{600}$	Optical density at 600 nm
*O*MT	*O*-methyltransferase
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PLE	Pig liver esterase
ppm	parts per million
PPT	Phosphopantetheine
Q	Quadrupole
RiPP	Ribosomally synthesised post-translationally modified peptide
RNA	Ribonucleic acid
RT	Room temperature
SAH	*S*-adenosyl homocysteine
SAM	*S*-adenosyl methionine
SAR	Structure activity relationship
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SFM	Soya flour mannitol (medium)
SMM	Supplemented minimal medium
SMMS	Supplemented minimal medium soli
TBE	Tris, boric acid, EDTA buffer
TCFH	*N,N,N',N'*-tetramethylchloroformamidinium hexafluorophosphate
TE	Thioesterase (domain)
TEMED	Tetramethylenediamine
TFA	Trifluoric acid
THF	Tetrahydrofuran

TLC	Thin layer chromatography
TMS	Trimethylsilane
TOF	T ime-of-flight
TSB	Tryptic soy broth
UHPLC	Ultra high performance liquid chromatography
UV	Ultra-violet

Abstract

Epoxyketone proteasome inhibitors derived from natural products are invaluable for the treatment of multiple myeloma and are currently in clinical trials against autoimmune disorders. Epoxyketone metabolites are produced by various bacterial phyla, including Actinobacteria, Cyanobacteria, and Proteobacteria. Their epoxyketone pharmacophore is assembled by acyl-CoA dehydrogenase-like enzymes called epoxyketone synthases. These enzymes use α-dimethyl-β-keto carboxylic acid substrates assembled by a combination of non-ribosomal peptide synthetases, polyketide synthases, and fatty acid synthases. To explore whether it is possible to replace chemically challenging and unsustainable methods for epoxyketone formation in synthetic proteasome inhibitor production with an enzyme catalysed procedure, the substrate tolerance of epoxyketone synthases was investigated.

Chemically stable methyl ester analogues of the substrate for epoxyketone synthase EpnF were synthesised with an alanine, leucine, cyclopentenylalanine, phenylalanine, and tryptophan residue in place of the native 4,5-dehydroleucine residue. To probe the substrate tolerance of EpnF, coupled enzyme assays were performed using pig liver esterase, to generate the carboxylic acid substrate *in situ*, and EpnF. While the alanine, leucine, cyclopentenylalanine, and phenylalanine derived substrates were tolerated by EpnF, no epoxyketone formation was observed for the tryptophan-derived substrate analogue. Although the phenylalanine derived analogue was accepted by EpnF, a precursor of the proteasome inhibitor oprozomib, which has a phenylalanine residue in the same position, was not converted to the corresponding epoxyketone.

A bioinformatics approach was developed to identify and analyse novel epoxyketone biosynthetic gene clusters *in silico*. A cryptic gene cluster hypothesised to direct the production of the tryptophan derived epoxyketone tryptopeptin A was identified in *Streptomyces sparsogenes* ATCC 25498. Comparison of the metabolite profiles with that of known tryptopeptin A producer *Streptomyces* sp. maeda85 indicated that *S. sparsogenes* ATCC 25498 does not produce tryptopeptin A. However, collaborative work to clone the gene cluster and express it in a heterologous host showed that it directs the production of several tryptopeptin related metabolites.

1. Introduction

1.1 Proteasomes

Proteasomes are proteolytic complexes within cells that regulate protein degradation. The constitutive proteasome or 26S proteasome regulates degradation processes of misfolded, damaged, or short-lived proteins and therefore is crucial for cells' survival. When cells are presented with inflammatory stress, however, constitutive proteasomes can be replaced with immunoproteasomes which can assist more effectively in the immune response of the cell.[1]

1.1.1   Constitutive proteasome

The 26S proteasome is a highly selective and efficient proteolytic complex in eucaryotic cells which regulates various essential cellular processes by degrading numerous types of proteins. It recognizes polyubiquitin markers which are covalently installed on proteins to signal and initiate their proteolytic degradation.[2–5] The constitutive proteasome weights approximately 1.5 MDa and structure elucidation by electron microscopy revealed that it consists of the barrel shaped 20S proteasome subunit, the catalytic core particle, and the 19S proteasomes, which are attached on either end of the 20S barrel (Figure 1.1).[2,6] While the 19S unit recognises the ubiquitylated markers and initiates protein unfolding, the 20S barrel contains the proteolytic active site.[2]

Figure 1.1 Schematic drawing of the 26S constitutive proteasome from the side (left) and down the catalytic barrel formed of two heptameric α- and two heptameric β-units.

The ATP dependant 19S regulatory unit recognizes ubiquitinated proteins, deubiquitylates and unfolds them, and transfers them through the α-rings into the proteolytic site.[3,7]

The 20S proteasome weights approximately 750 kDa and a cylindric shape with a barrel in the centre was observed from electron microscopy structure elucidation.[2] The ring is formed by axial stacking of two heptameric outer α-rings and two heptameric inner β-rings, resulting in an αββα motif which is conserved among most eucaryotic species.[2,6] The α-rings form the outer sides of the barrel and are comparatively narrow forming a barrier for regular proteins and cell organelles. Each of the heptameric β-rings contains three subunits with active *N*-terminal threonine residues which hydrolyse marked proteins by cleaving their peptide bonds and thus forming the proteolytically active site of the proteasome. Of those six catalytic threonine residues, two cleave preferentially after acidic amino acids, two after basic, and two after hydrophobic ones which leads to the proteins being broken down into peptides between 3 and 15 amino-acids long which subsequently are further hydrolysed into individual amino acids by peptidases.[2,8] The proteolytic mechanism involves a nucleophilic attack of the threonine hydroxy group on the peptide bond carbonyl leading to a tetrahedral intermediate which eliminates the downstream product to form an acyl-enzyme intermediate. Hydrolysis of the acyl-enzyme complex subsequently releases the upstream product and regenerates the catalytic threonine (Scheme 1.1).[8]
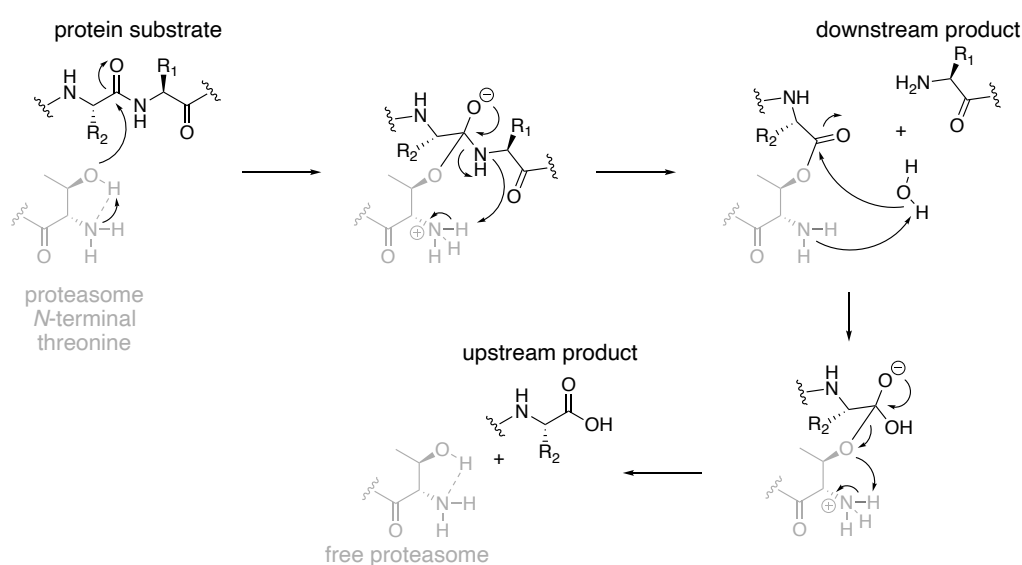
Scheme 1.1 Catalytic mechanism of proteolytic degradation of proteins facilitated by the constitutive proteasome.

The proteasome activity is crucial for several cell processes. It is involved in protein quality control by destroying marked damaged proteins, and transcriptional regulation by rapidly degrading transcriptional factors. Additionally, it is responsive to changing physiological conditions which makes it invaluable for cell homeostasis, cell composition, cell cycle regulation, and apoptosis.[4,6] Due to its vital functions, various cell types are highly dependent on the 26S proteasome.

1.1.2   Immunoproteasome

Immunoproteasomes replace the constitutive proteasomes when a cell is put under inflammatory conditions. The immunoproteasome digests infectious particles into fragments to present to relevant cells of the immune system for rapid development of an effective immune response.[1] It has also been linked to increased production of cytokines, small proteins to alert the immune system, and to the differentiation in T cells which are crucial for the immune response.[1,9,10] All of these activities make the immunoproteasome invaluable for several inflammatory fighting processes. Additionally, recent findings suggest that the immunoproteasome might also be involved in skeletal muscle differentiation and detailed studies of immunoproteasomes in the lung showed novel activities, suggesting that it might fulfil various roles in different cell types that remain unknown.[11,12]

The immunoproteasome is distinguished from the constitutive proteasome through different β-subunits. In the heptameric β-ring of the catalytic core, three β-subunits are replaced with specific immuno-subunits, homologous to the constitutive β-subunits but functionally different (Figure 1.2).[10] Additionally, the 19S regulatory protein of the constitutive proteasome is replaced with a smaller 11S regulator unit. The immuno-subunits encoding genes are in vicinity to genes encoding for antigen processing transporters which are involved in the immune response as well. Expression of the immuno-subunits is induced by the presence of cytokines and other inflammatory signalling markers and their upregulation leads to almost exclusive generation of the immunoproteasome instead of the constitutive proteasome.[1]
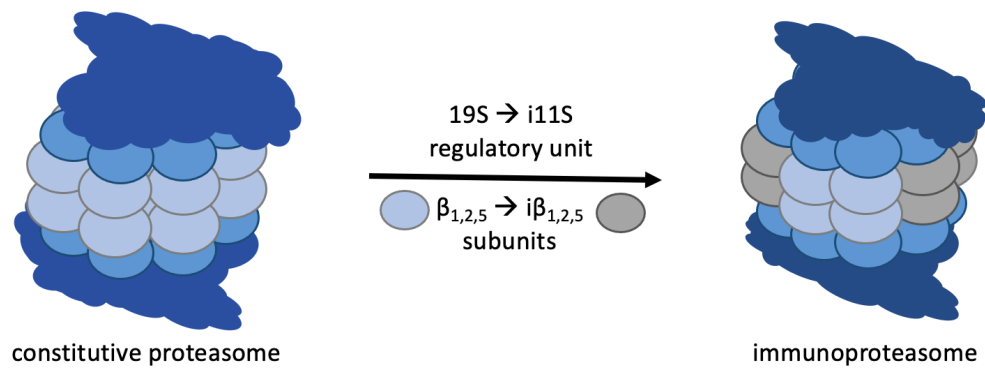
Figure 1.2 Schematic visualisation of the transformation from the constitutive proteasome to the immunoproteasome under the influence of inflammatory stress.

1.1.3 Proteasome inhibition as a treatment strategy

Human life expectancy is increasing continuously, which changes the landscape of diseases and ailments which require treatment. Nowadays, some of the most challenging diseases to cure or control are chronic illnesses – diseases that persistently last for long amounts of time. In particular, diseases related to our highly complex immune system, such as different types of cancer and autoimmune diseases remain to be fully understood.[13] Ongoing research is revealing more insight into symptoms, mechanisms, and related treatment options for these immunological challenges.[14,15] Over the last decades, some cancer types like the blood cell cancer multiple myeloma as well as several autoimmune disorders have been related to misfunctioning proteasomes which has led to proteasomes being targeted for treatment of multiple myeloma and autoimmune disorders.[16,17]

1.1.3.1 Multiple myeloma

Multiple myeloma is a type of blood cancer which is characterised by malignant plasma cells present in the bone marrow and in later stages in the peripheral blood, kidneys, and other organs.[14] Plasma cells usually produce antibodies as an immune response to fight off foreign substances, however, infected plasma cells produce increased amount of monoclonal proteins which are secreted into tissues and organs.[18] Cases of multiple myeloma have been on the rise globally over the last decades. While the growth rate of the number of cases varies among countries, the overall multiple myeloma cases worldwide increased by 126% since 1990.[19]

Until 2014, multiple myeloma used to be diagnosed only after end-organ damage was observed (e.g. hypercalcaemia, renal insufficiency, anaemia, bone disease with lytic lesions) which prevented early diagnosis and early treatment options. Advanced research linked several biomarkers with an increased risk of suffering from myeloma-related organ damage which led the International Myeloma Working Group (IMWG) to update the criteria for diagnosis to include these biomarkers in the official diagnosis of the disease.[20] Diagnosis now includes blood and urine sampling to test for monoclonal protein concentration as well as bone marrow biopsy to investigate plasma cells. Early diagnosis and start of treatment in combination with progressing treatment options continuously prolong the survival and improve the quality of life of multiple myeloma patients.[21,22]

Developed in the 1980s, autologous stem cell transplants was the first break through in treating this clonal plasma cell disease, followed by the development of "IMIDs" thalidomide and lenalidomide which are immunomodulatory drugs (Figure 1.3).[22,23] More recently developed treatment options are monoclonal antibodies, histone deacetylating agents, and proteasome inhibitors.[23] Proteasome inhibitors act on the constitutive proteasome to prevent degradation of ubiquitinated proteins to interrupt the cell cycle progression of infected cells.[2,5] Treatment of multiple myeloma depends on the health state of the patient, however, proteasome inhibitors are given in different doses throughout all combination of treatment options and are therefore crucial for tackling this blood cell cancer.[24–26]

thalidomide (**1**)   lenalidomide (**2**)
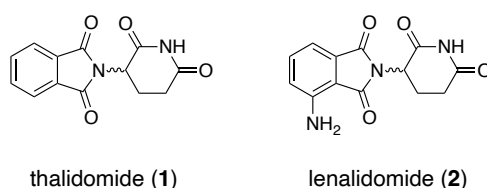
*Figure 1.3 Structures of immunomodulatory drugs thalidomide (**1**) and lenalidomide (**2**).*

Systematic analyses by Cowan *et al.* have shown how the burden of multiple myeloma is significantly increased in low income countries, especially in sub-Saharan Africa, which is another reason why it is crucial that the production of proteasome inhibitors becomes more accessible and affordable.[19]

1.1.3.2 Autoimmune diseases

Autoimmune disorders are diseases of the immune system in which the immune response is directed against itself.[15,17] There are almost 100 different types of autoimmune disorders which impact nearly 5% of the population and the rate is constantly increasing.[27] While the precise reasons for the outbreak of many autoimmune diseases remain mostly unknown, correlations with some genetic and environmental factors, e.g. exposure to chemicals or radiation, have been drawn.[28,29] Autoimmune diseases can be systemic or effect specific organs and tissues, they display a broad range of symptoms, and are diagnosed based on different biomarkers.[30]

Recent research revealed that the immunoproteasome is involved in many crucial processes during immune responses such as T cell expansion and differentiation, viral protein degradation, and cytokine production, which makes the immunoproteasome a broad target for various types of autoimmune disorders.[2,5,9,17] Its expression is induced by inflammatory related factors and increased amounts are found at sites of infection and inflammation in autoimmune disorders.[31,32] Diseases such as systemic lupus erythematosus, amyotrophic lateral sclerosis, Sjögren's syndrome, and Crohn's disease show increased expression of the immunoproteasome, and studies show promising results for targeting this immune-optimised proteasome.[33–37]

1.2 Proteasome inhibitors derived from natural products

The majority of currently used drugs are derived from secondary metabolites, which makes biologically active natural products invaluable for human health (Figure 1.4).[38–41] Pharmaceutically relevant compounds are found in animals, plants, fungi, and bacteria and exhibit a broad range of activities useful for treatments of various diseases and ailments. Some life changing examples of natural product derived drugs are the discovery of antibiotics by Sir Alexander Flemming, the development of aspirin from the bark of willow trees, and clinically widely used antitumour agents such as bleomycin.[42–44] More recently, a rising number of proteasome inhibitors has been discovered, predominantly in the actinomycete genera of *Streptomyces*, a type of Gram-positive filamentous bacteria.[45]
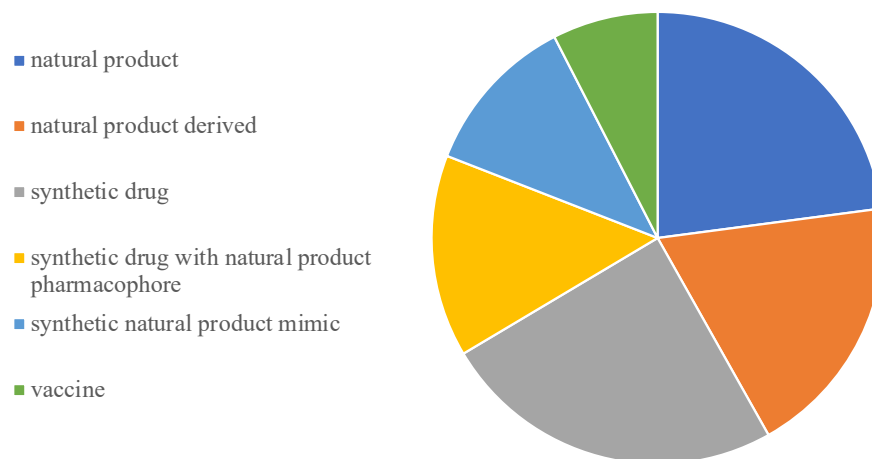
Figure 1.4 Newly approved drugs between 1881 and 2019 and their type based on analysis by Newman et al.[41]

The proteolytic mechanism of proteasomes was identified in experiments involving site-directed mutagenesis and experiments with known *N*-terminal nucleophilic hydrolase inhibitors, which react with the catalytic hydroxyl in the active site forming a hemiacetal resembling the enzymatic transition state.[46,47] These studies did not only give insight into the proteolytic mechanism, but also provided useful information for the search for proteasome inhibitors. Throughout the last decades, several known biologically active natural products were re-investigated for potential proteolytic activities and new proteasome inhibiting natural product classes have been identified.

1.2.1 Peptide aldehydes

Some of the earliest naturally occurring proteasome inhibitors were leupeptins, a class of peptide aldehydes isolated from various *Streptomyces* strains (Figure 1.5).[48,49] The protease inhibiting activity of leupeptin (**3**) was discovered in 1969 and started the development and studies of numerous synthetic peptide aldehydes.[50]
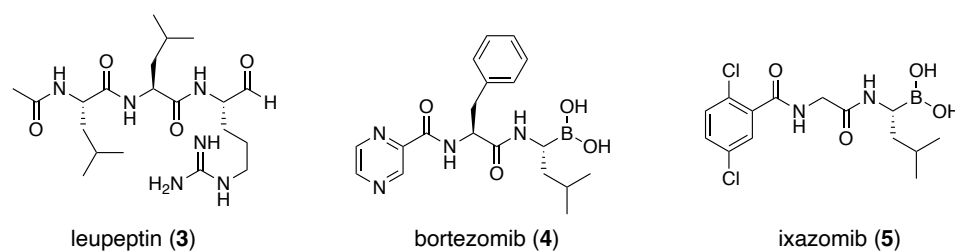
*Figure 1.5 Structures of peptide aldehyde proteasome inhibitor leupeptin (**3**) and thereof derived clinically used boronate proteasome inhibitors bortezomib (**4**) and ixazomib (**5**).*

The peptide fragment of the proteasome inhibitors was found to be important for association to the substrate binding pocket while the pharmacophore reacts with the active site. Because aldehydes form reversible hemiacetal adducts, researchers investigated alternative pharmacophores that could react with the *N*-terminal catalytic threonine residues of the proteasome. The discovery of peptide aldehyde natural products has led to the development of synthetic peptide vinyl sulfones and peptide boronates proteasome inhibitors such as bortezomib (**4**) and ixazomib (**5**) which are currently used to treat multiple myeloma.[51–54]

1.2.2 Lactacystin and β-lactones

Another naturally occurring proteasome inhibitor is lactacystin (**6**), a natural product found in *Streptomyces*.[55] It irreversibly modifies one of the β-subunits of eucaryotic proteasomes by reacting with the proteasome's catalytic threonine.[56] Investigation of the reaction mechanism of lactacystin revealed β-lactone intermediate omuralide (**7**) which led to the development of cell permeable β-lactone proteasome inhibitors.[57,58] Natural product β-lactone salinosporamide A (**8**) was also found to induce apoptosis in multiple myeloma cells and has therefore served as a lead compound for analogue development (Figure 1.6).[59–61]

lactacystin (**6**) omuralide (**7**) salinosporamide A (**8**)
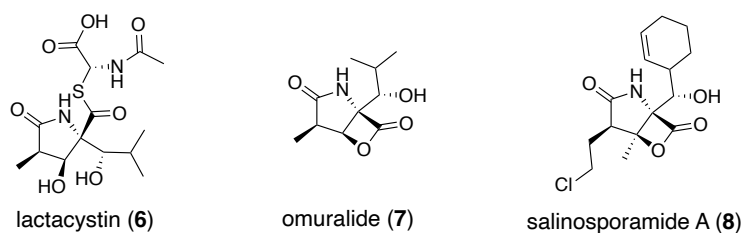
*Figure 1.6 Structures of proteasome inhibitors utilising β-lactones: Lactacystin (**6**), synthetic β-lactone omuralide (**7**), and natural product β-lactone salinosporamide A (**8**).*

1.2.3 Macrocyclic peptides

Macrocyclic peptides exhibit a broad range of biological activities and TMC-95 metabolites isolated from an *Apiospora montagnei* were investigated for their proteasome inhibiting abilities.[62] Especially TMC-95 A (**9**) showed specific proteasome inhibition by interacting non-covalently with the 20S catalytic core resulting in a reversible inhibition.[63] Other classes of macrocyclic peptides such as argyrins were identified as proteasome inhibitors after argyrin A (**10**) was investigated as growth inhibitor of mammalian cells.[64] Conducting structure activity relationship studies with phepropeptin A (**11**) revealed its ability of β-subunit inhibition and linked the activity of phepropeptins to important van der Waals and hydrophobic interactions between this class of macrocycles and the proteasome (Figure 1.7).[65]

TMC-95A (**9**) argyrin A (**10**) phepropeptin A (**11**)
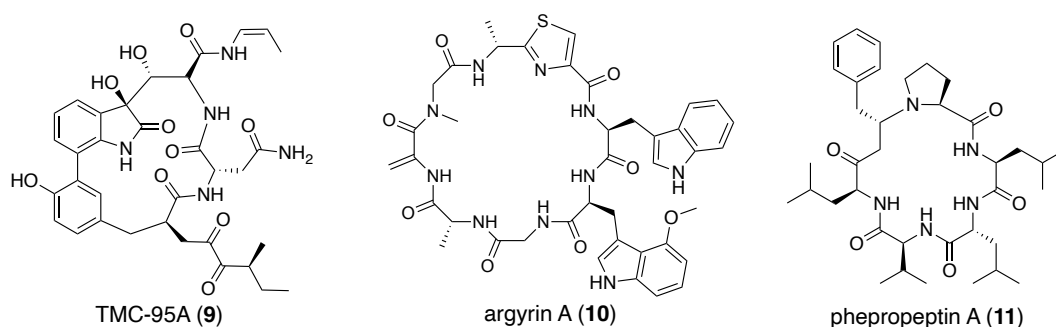
*Figure 1.7 Structures of natural product macrocyclic peptide proteasome inhibitors TMC-95A (**9**), argyrin A (**10**), and phepropeptin A (**11**).*

1.2.4 Epoxyketones

Lastly, epoxyketone natural products such as eponemycin (**12**) and epoxomicin (**13**) have been isolated and studied due to their anti-tumour activity, however, they were found to exhibit additional proteasome inhibiting properties.[66,67] Mechanistic studies

provided insight into the irreversible inhibition of the *N*-terminal threonine residue of the proteasome's active site (further discussed in section 1.3.3).[68] Epoxyketone natural products served as lead compounds in the development of multiple myeloma drugs such as oprozomib (**14**) and immunoproteasome inhibitors are currently in clinical trial (further discussed in section 1.3.2) (Figure 1.8).[69–71]

eponemycin (**12**)   epoxomycin (**13**)   oprozomib (**14**)
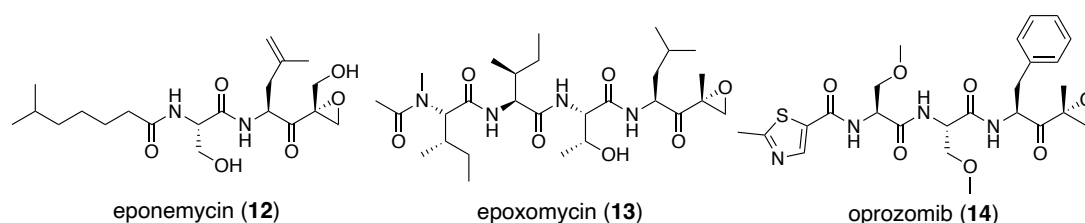
*Figure 1.8 Structures of epoxyketone proteasome inhibitor natural products eponemycin (**10**), epoxomicin (**11**), and thereof derived synthetic epoxyketone proteasome inhibitor oprozomib (**12**).*

1.2.5   Novel proteasome inhibiting natural product classes

As proteasome inhibitors become more successful drugs for multiple myeloma and autoimmune disorder treatments, the research for novel and improved proteasome inhibitors continues. The development of commercially available activity test kits for different proteasomes and specific proteasome subunits enabled more efficient activity testing of potential proteasome inhibitors and several classes of natural products showed proteasome inhibiting activities in such *in vitro* experiments or preliminary *in vivo* studies. For example, terpenoids of the agosterols family, stronglyphorines, and petrosaspongiolide exhibited potent preliminary proteasome inhibition. Additionally, several natural products belonging to compound classes of alkaloids, flavonoids, polyketides, and macrolides showed inhibiting activities against different types of proteasomes.[62]

1.3 Epoxyketone proteasome inhibitors

Epoxyketone proteasome inhibitors consist of a central peptide fragment of 2-4 amino acids with the epoxyketone warhead attached to the *C*-terminus and various possible functional groups at the *N*-terminus (Figure 1.9). Ever since the proteasome inhibiting activities of the epoxyketone pharmacophore have been discovered, clinically used epoxyketone drugs against multiple myeloma were developed and derived epoxyketones are currently in clinical trial against numerous autoimmune diseases.[37,62]
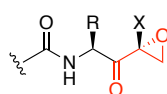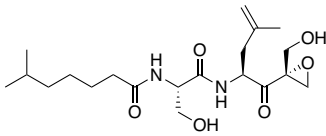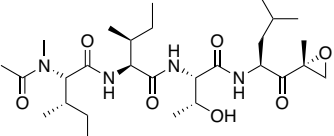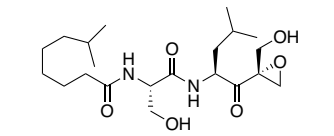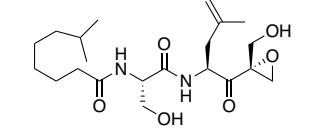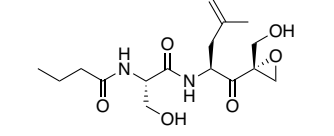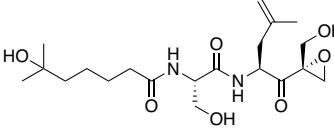
X = Me, CH$_2$OH or H

Figure 1.9 General structure of α,β-epoxyketone proteasome inhibitors. R represents amino acid residues, and the N-terminal moiety can vary widely.

1.3.1 Epoxyketone natural products

The first known epoxyketone proteasome inhibitor, eponemycin (**12**) was discovered in *Streptomyces hygroscopicus* extracts in the search for novel antitumor agents.[72] Several other epoxyketone secondary metabolites have since been extracted from a broad range of actinomycetes (Figure 1.10). While the majority of epoxyketones have been found experimentally in extracts of actinomycetes, some of the more recently discovered epoxyketones such as landepoxcin A and clarepoxcin A were identified through bioinformatic analysis of a soil metagenome library and heterologous expression of the biosynthetic gene clusters.[73]

eponemycin (**12**) *S. hygroscopicus* (1990)[72]	epoxomicin (**13**) *Actinomycetes* Q96-17 (1992)[69]	epopromycin A (**15**) *S.* sp. NK04000 (1997)[74]
epopromycin B (**16**) *S.* sp. NK04000 (1997)[74]	TMC-86A (**17**) *S.* sp. TC1084 (1999)[75]	TMC-86B (**18**) *S.* sp. TC1084 (1999)[75]
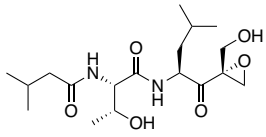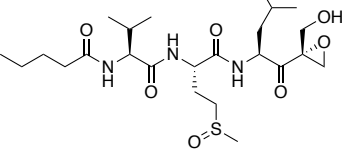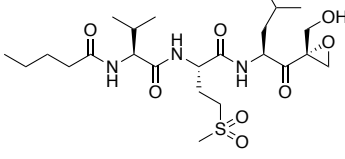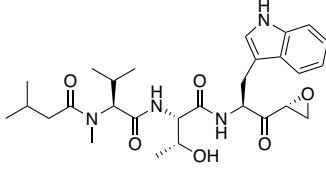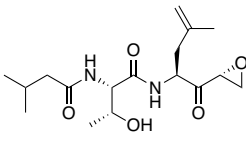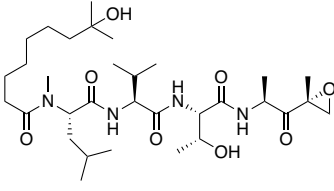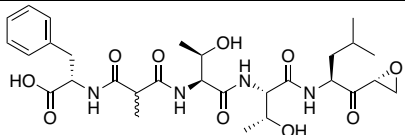
Figure 1.10 Structures of all known epoxyketone natural products.

1.3.2 Epoxyketone pharmaceuticals

Based on the proteasome inhibiting properties of epoxyketone natural products, several epoxyketone drugs have been developed. Epoxomicin (**13**) served as a lead compound in the development of the first FDA approved epoxyketone drug against multiple myeloma.[69] Elofsson *et al.* followed a classical medicinal chemistry approach of systematically changing elements of the compound.[67] Varying the *N*-terminal group of the epoxyketone and testing different amino acids resulted in the structure of YU-101 (**26**) which was picked up by Proteolix for further structural optimisation experiments.[79] The results of these studies ultimately provided carfilzomib (**27**) which is currently manufactured by Onyx Pharmaceuticals as Kyprolis™ and administered to multiple myeloma patients (Figure 1.11).
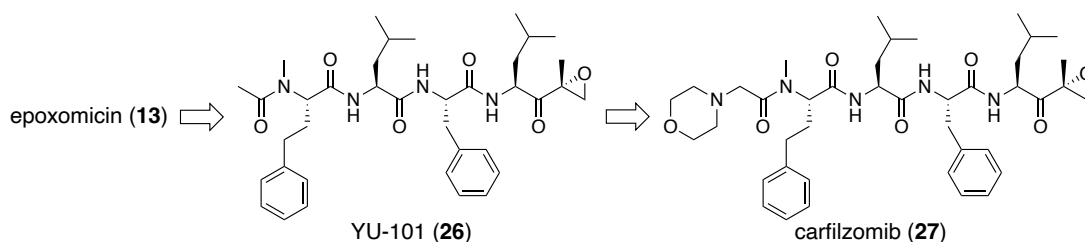
*Figure 1.11 Structures of relevant compounds for the development of clinically used epoxyketone proteasome inhibitors. YU-101 (**24**) was developed from natural product epoxomicin (**11**) and picked up by Proteolix for further improvements resulting carfilzomib (**25**).*

12

A major downside of the first generation epoxyketone proteasome inhibitor carfilzomib is that it has to be administered intravenously, which is accompanied by dosage difficulties and discomfort of the patients. Zhou *et al.* attempted to develop an orally available second-generation multiple myeloma epoxyketone drug and generated novel carfilzomib analogues.[70] Truncation of carfilzomib by one amino acid and screening for various *N*-terminal groups and different amino acids led to the discovery of oprozomib (**14**) (Figure 1.12). The novel epoxyketone proteasome inhibitor showed suitable oral bioavailability and the most promising proteasome inhibition of the tested compounds.[80] Oprozomib exhibited promising results when tested in different formulations in early clinical trials and is currently in phase Ib and II by Amgen.[71]
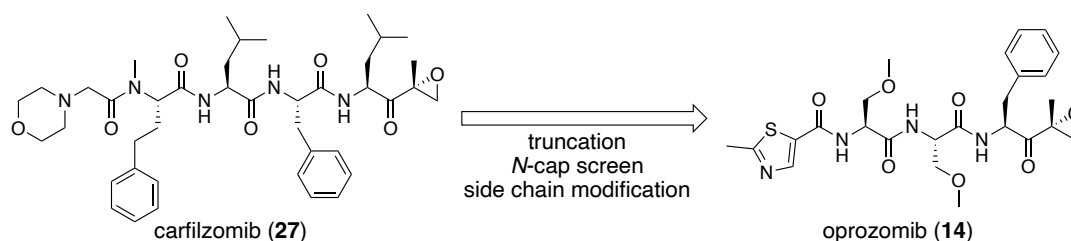
Figure 1.12 Structure of second generation epoxyketone proteasome inhibitor oprozomib and how it was developed from carfilzomib.

More recently, epoxyketones have also been studied for their activity against the immunoproteasome. To investigate the interactions of drug candidates and the immunoproteasome, Johnson *et al.* modelled various epoxyketone analogues into the catalytic site of the immunoproteasome (Figure 1.13).[37] This identified interactions between the peptide moiety of the epoxyketone and the proteasome such as beneficial van der Waals interactions which could enhance the binding affinity between drug candidate and proteasome. Differences in the interactions between the proteasome and the different moieties of epoxyketone drugs are believed to be crucial for the selectivity of compounds such as ONX 0914 (**28**) and KZR-616 (**29**) towards the immunoproteasome without loss in the proteolytic functions of the constitutive proteasome.[37,81]
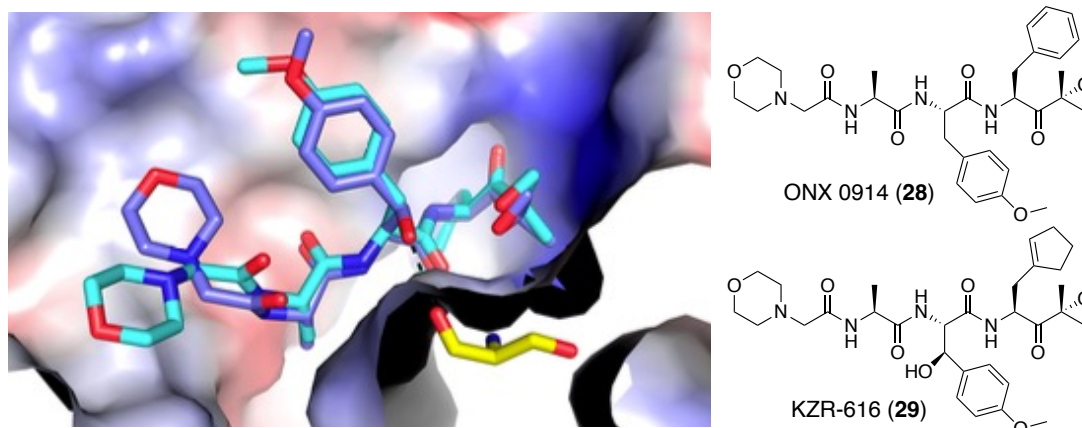
Figure 1.13 Structures of immunoproteasome inhibitor candidates ONX 0914 (26) and KZR-616 (27) modelled into the immunoproteasome binding site by Johnson et al. to explore interactions between epoxyketone peptide chain and the proteasome active site. KZR-616 demonstrates an increased binding affinity and solubility compared to ONX 0914.

Alongside the structural models, Johnson *et al.* probed potential epoxyketone immunoproteasome inhibitors searching for more potent autoimmune disorder drug candidates. A library of epoxyketones with varying amino acids and *N*-terminal groups was synthesised based on structural models and co-crystal structures of humanised yeast proteasome and known immunoproteasome inhibitor PR-924 (**30**) (Figure 1.14).[82] Compounds were tested with various immunoproteasome subunits to determine their inhibiting activities and the effective change of cytokine expression was assessed. Out of all the candidates tested, the most promising activities and beneficial physicochemical properties were observed in KZR-616 (**29**), which is currently in clinical trial against rheumatic diseases.
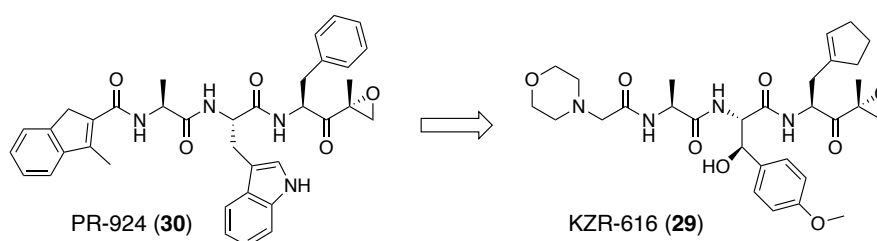
Figure 1.14 Structures of immunoproteasome inhibitor epoxyketones PR-924 (28) and KZR-616 (27).

1.3.3 Mechanism of proteasome inhibition

Crystallisation of epoxomicin (**13**) bound to the 20S catalytic core of the proteasome showed the morpholino adduct **31** being formed between the pharmacophore and the

N-terminal catalytic threonine.[68] The mechanism was therefore hypothesised to proceed via a nucleophilic attack of the threonine side chain oxygen on the electrophilic epoxyketone carbonyl group, followed by opening of the epoxide ring by the threonine amine to form the morpholino adduct, resulting in irreversible proteasome inhibition (Scheme 1.2).

Almost two decades after the first epoxyketone-proteasome crystal structure was published, higher resolution data of the proteasome bound to the second generation epoxyketone proteasome inhibitor oprozomib (**14**) provided insight into a possible alternative mechanism. Instead of the previously observed morpholino adduct between the pharmacophore and the *N*-terminal threonine, Schrader *et al.* identified the 1,4-oxazepano adduct **32**.[82] Based on the newly observed 7-membered ring, Carmony *et al.* published a revised proposed inhibition mechanism in which the epoxide ring is opened by attack of the catalytic threonine on the other epoxide carbon.[83] The oxazepano adduct is believed to be the kinetically favoured product and this mechanistic insight opens new possibilities for the design of novel proteasome inhibitors.
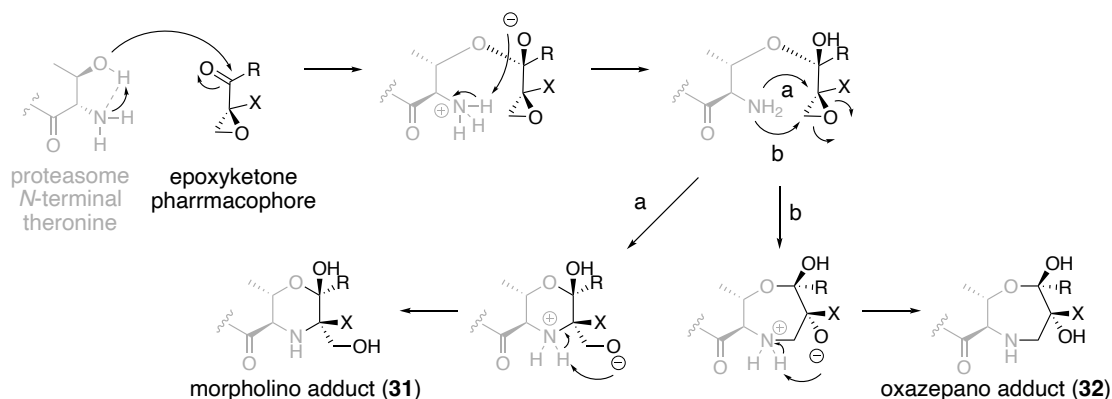
Scheme 1.2 Proposed possible mechanisms of epoxyketone proteasome inhibition. The proteasome N-terminal catalytic threonine nucleophilically attacks the epoxyketone carbonyl group before it irreversibly forms either the morpholino adduct 31 or oxazepano adduct 32.

1.3.4   Epoxyketone chemical synthesis

Over the past decades, several groups have attempted the synthesis of epoxyketone proteasome inhibitors and chemical synthesis is the current method of production for industrial epoxyketone drugs.[67,70,84–87] Assembling the peptide part of the epoxyketones of interest is achieved by standard and straight forward solid phase

peptide coupling reactions and desired *N*-terminal groups can usually be attached in a single additional coupling step (Figure 1.15).[88–92]
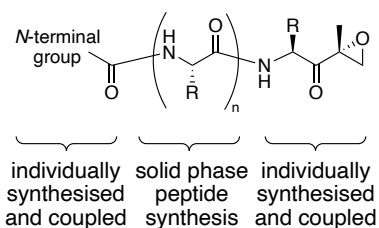
Figure 1.15 General approach for the synthesis of epoxyketone pharmaceuticals. The N-terminal group, peptide fragment, and epoxyketone warhead are individually assembled and subsequently coupled to form the epoxyketone drug.

The most challenging part of the synthesis is the formation of the epoxyketone pharmacophore itself. The epoxide was originally formed by oxidation of a β-carbonyl alkene (**34**) to the epoxyketone warhead (**35**) (Scheme 1.3). While different oxidising agents and slight differences in reaction conditions have been published, this step continues to exhibit no or very poor selectivity, with both enantiomers being formed.
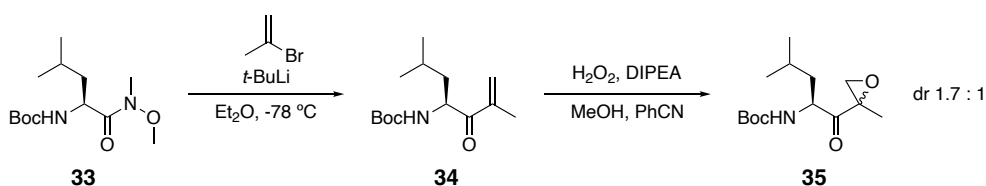
Scheme 1.3 Original reaction conditions of synthetic epoxyketone formation.

Since the original synthesis of carfilzomib which included industrially critical components such as cryogenic conditions and 76% yield with 1.7:1 selectivity of the epoxide formation, some optimisations have been achieved.[85] A minor improvement to a dr of 2:1 has been reported over a decade later which replaced the isopropenylbromide and BuLi with isopropenylmagnesium bromide and overall more suitable industrial conditions.[88] As Carfilzomib became an increasingly relevant asset for the Amgen portfolio, a strategic approach was applied to search for improved conditions to form the epoxyketone pharmacophore.[93] Amgen and other parties interested in epoxyketone synthesis investigated alternative epoxidation reagents and studies on various protecting groups revealed that the choice of amine protection has an influence on the diastereoselectivity of the epoxidation.[94] The ultimate significant

breakthrough was a chiral manganese complex (**36**) that catalysed the epoxide formation in high yields with high diastereoselectivity and low catalyst loading, although it favoured the undesired diastereomer. Amgen however took these findings and developed its currently used process, which involves the selective catalytic epoxidation of *N*-Boc-D-leucine morpholine amide followed by highly efficient epimerization to the desired epoxyketone fragment (**40**) which can then be coupled to the peptide moiety (Scheme 1.4).[95] Increased selectivity in the catalysed epoxide formation, the development of a continuous process, and purification by crystallisation made the current large scale epoxyketone production safer and more efficient compared to the original synthetic strategy.
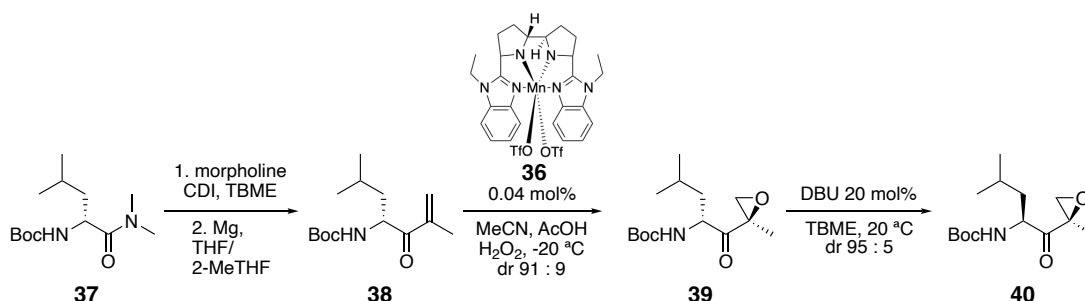
Scheme 1.4 Improved synthesis by Amgen for catalytic asymmetric epoxyketone formation followed by epimerization to the desired diastereomer.

However, despite all efforts, the epoxyketone formation is still far from optimal. Since the commercially relevant drugs are based on natural products, the biosynthesis of epoxyketones could provide valuable insight into epoxyketone formation.

1.3.5 Epoxyketone biosynthetic gene clusters

The biosynthetic gene clusters of epoxyketones that have previously been studied were identified as non-ribosomal peptide synthetase (NRPS)/ polyketide synthase (PKS) hybrid clusters with additional modifying and regulatory protein encoding genes.[96,97] While the specific order and predicted functions of genes in the known epoxyketone clusters vary, they all encode an NRPS, a PKS, an acyl-CoA dehydrogenase (ACAD) homologue, and a 20S proteasome subunit, presumably for self-resistance (Figure 1.16).
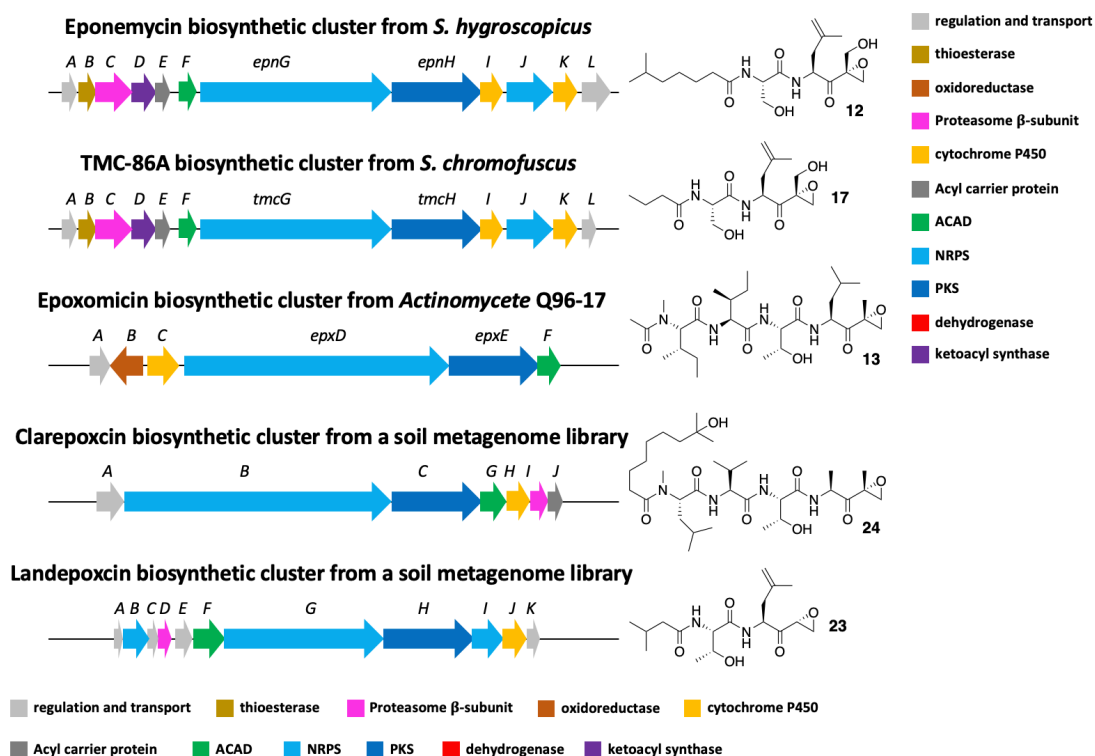
Figure 1.16 Known biosynthetic gene clusters of epoxyketone natural products eponemycin[72], TMC-86A[75], epoxomicin[69], clarepoxcin[73], and landepoxcin[73]. Predicted functions of the encoded genes are included.

The biosynthetic gene clusters of the secondary metabolites eponemycin (**12**) isolated from *S. hygroscopicus* ATCC53709 and epoxomicin (**13**) from *Streptomyces* strain ATCC 53904 were identified by Moore *et al.*[96] The link of the *epn* and *epx* cluster to eponemycin and epoxomicin production respectively was confirmed by heterologous expression of the identified clusters and metabolite analysis.

Structurally related epoxyketone TMC-86A (**17**) – which differs from eponemycin only in the shorter fatty acid chain – and its producing gene cluster in *S. chromofuscus* ATCC49982 have been studied in the Challis group. Gene knockout experiments within the *tmc* cluster confirmed the TMC-86A producing cluster and revealed mechanistic insights.[97]

The most recent identified gene clusters of landepoxcin (**23**) and clarepoxcin (**24**) were identified by Owen *et al.* through a multiplex genome mining approach of a soil metagenome database and confirmed by heterologous expression of the clusters followed by metabolite analysis.[73]

1.3.6 Biosynthesis of epoxyketone TMC-86A

The proposed biosynthesis of TMC86-A (**17**) involves the fatty acid synthase (FAS) catalysed condensation of an acetyl-CoA primer to a malonyl unit bound to an acyl carrier protein (ACP) which is subsequently reduced to butanoyl-ACP, forming the fatty acid end of the natural product.[98] The butanoyl moiety is elongated by two NRPS modules by first an L-serine followed by a 4,5-dehydro-L-leucine residue which is previously desaturated by a cytochrome P450 from L-leucine.[99] The peptide intermediate is extended by the PKS module by a malonate unit which subsequently is methylated in the α-position by a *C*-methyltransferase domain (cMT) and cofactor *S*-adenosyl methionine (SAM).[100] The intermediate is then hydrolysed to form an α-dimethyl-β-keto carboxylic acid substrate for the flavin dependent epoxyketone synthase which forms the epoxyketone warhead.[97] (Figure 1.17).[97]
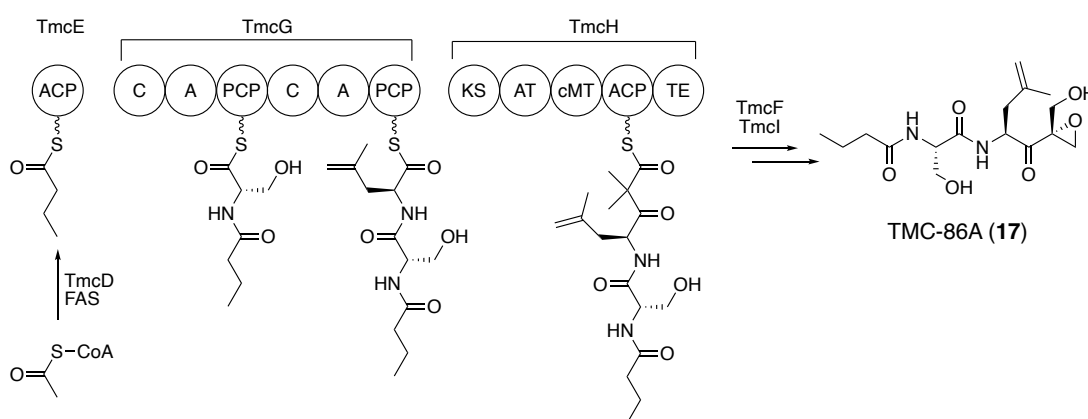
*Figure 1.17 Proposed pathway of the biosynthesis of TMC-86A (**17**).*

1.3.6.1 Expression regulation

The first gene in the *tmc* cluster that is predicted to be affiliated with TMC-86A production is the LuxR transcriptional regulator *tmcA*. LuxR transcription factors have a repressing helix-turn-helix DNA binding motif and a sensing region which can bind inducer molecules.[101] This transcriptional regulation mechanism is advantageous for an organism because it can induce the production of beneficial secondary metabolites. However, the exact transcription regulation of the *tmc* cluster has not been studied yet.

1.3.6.2 Self-resistance

It is crucial for an organism to develop a resistance mechanism against the secondary metabolites it is producing. The *tmc* cluster encodes for a proteasome β-subunit homologue TmcC. A similar self-resistance mechanism has been observed in *Salinispora tropica*, a marine actinomycete that produces the proteasome inhibitor Salinosporamide A (**8**).[102] The corresponding biosynthetic gene cluster encodes for a proteasome binding pocket containing mutations that ensure resistance against its own produced chemical.[103] It is expected that expression of TmcC results in a similar mechanism of self-resistance for *S. chromofuscus*.

1.3.6.3 Fatty acid biosynthesis

Bacterial fatty acid synthesis is catalysed by multiple enzymes and involves cycles of condensation, reduction, dehydration, and carbon-carbon bond reduction (Figure 1.18).[104,105] The fatty acid biosynthetic (Fab) proteins are widely studied and broadly conserved across the bacteria kingdom.
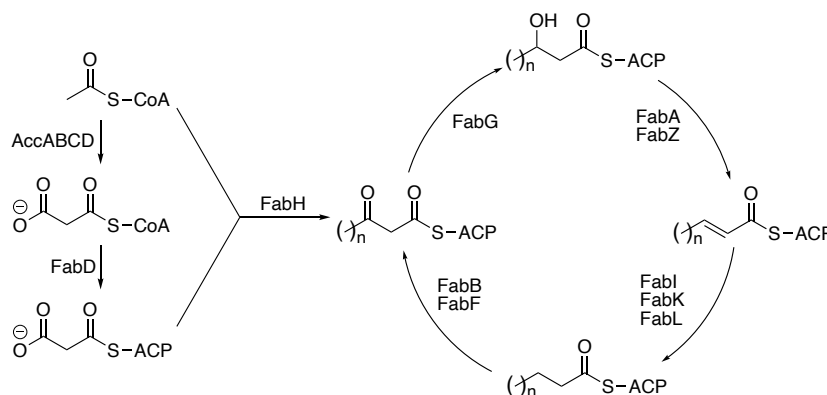
Figure 1.18 Bacterial fatty acid catalytic cycle.

The *tmc* cluster encodes for a FabH homologue TmcD, which is hypothesised to specifically load the acetyl-CoA moiety on the ACP-like TmcE.[106] Subsequently, this serves as a substrate for a standard fatty acid synthesis cycle with FabG, FabA, and FabI to result in butanoyl-ACP which will form the butanoic acid moiety of TMC-86A (Figure 1.19).[107–109]
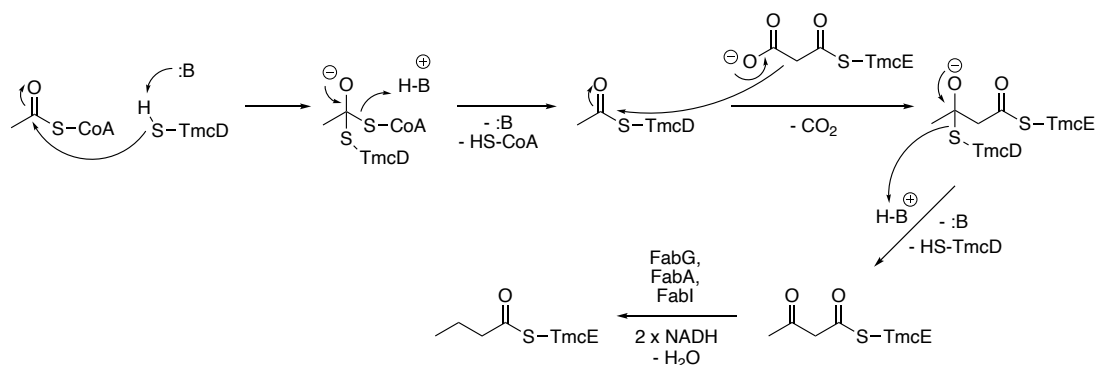
Figure 1.19 Mechanism of the formation of the butanoyl fragment in the TMC-86A biosynthetic process.

1.3.6.4 Dehydroleucine formation

Challis group member Dr Chuan Huang performed a series of gene knockout experiments to reveal that TmcJ and TmcK both play a critical role in the formation of the 4,5-dehydroleucine residue in TMC-86A. TmcJ is an NRPS homologue of one module containing an A domain and a PCP domain alongside an MbtH-like protein (MLP) while MLPs are highly conserved small proteins in bacteria which are known to promote the folding, stability, and activity of NRPSs.[110] TmcK being a putative cytochrome P450, Dr Huang proposed that TmcJ is loaded with L-leucine which is reduced subsequently by TmcK and hydrolytically released, probably by the thioesterase homologue TmcB (Figure 1.20). The L-dehydroleucine is then loaded onto the second module of the main NRPS TmcG.
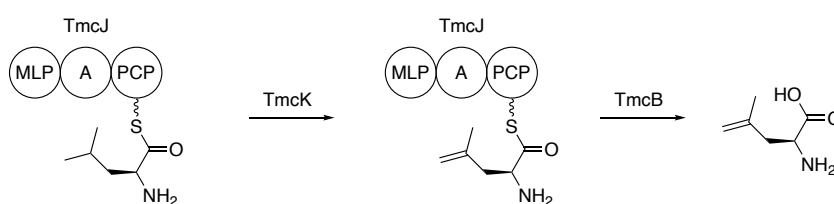
Figure 1.20 Proposed biosynthesis of 4,5-dehydroleucin in the biosynthesis of TMC-86A.

1.3.6.5 Non-ribosomal peptide biosynthesis

The TmcE butanoyl thioester is elongated by the NRPS TmcG to form the peptide backbone of the epoxyketone TMC-86A. NRPS enzymes are multifunctional and contain several catalytic domains which are organised into modules.[99] The key domains in the TMC-86A biosynthesis are Adenylation (A) domains, Peptidyl Carrier

Protein (PCP) domains, and Condensation (C) domains, and are described in detail below. EpnG and TmcG each contain two modules which are both composed of an A, C, and PCP domain to first attach an L-serine residue to the butanoic acid fragment, then incorporate an L-dehydroleucine residue (Figure 1.21).
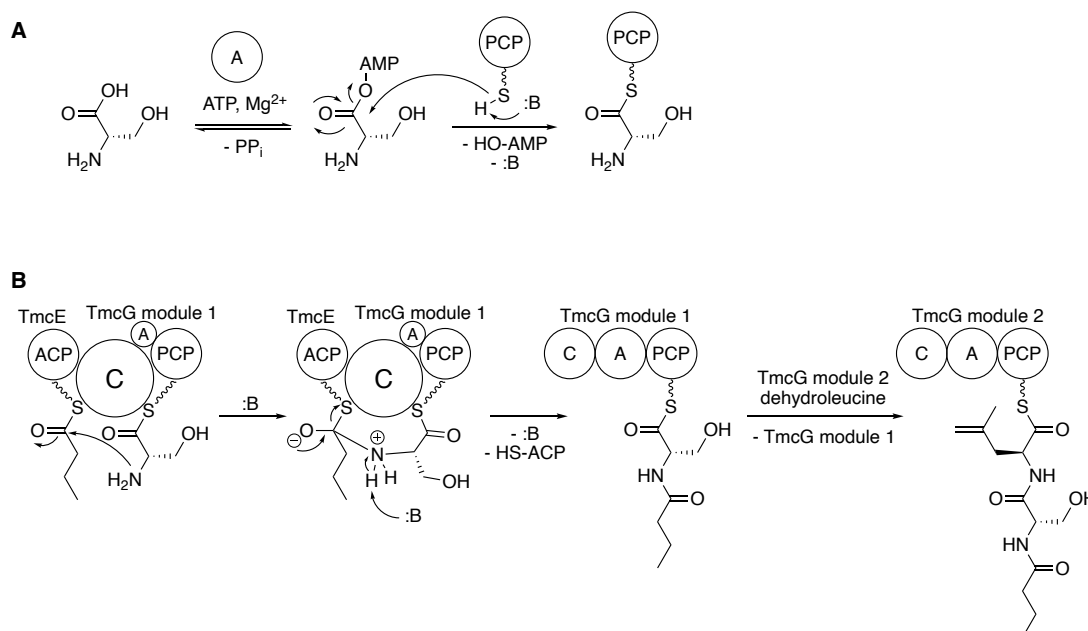
Figure 1.21 Mechanism of the non-ribosomal peptide assembly in TMC-86A biosynthesis. A. Amino acid activation using the A domain and ATP. B. Condensation domain mediated transpeptidation forming first a peptide bond between the TmcE bound butanoyl fragment and activated serine of the first NRPS module, followed by peptide bond formation between activated leucine of the upstream NRPS module and serine of the downstream NRPS module. The product of the NRPS assembly is the fatty acid peptide backbone of TMC-86A.

1.3.6.5.1 A domain

The purpose of an adenylation domain is to activate an amino acid to be loaded on a PCP domain. Using ATP, the amino acid is converted into an amino acyl adenylate which is subsequently attacked by the phosphopantetheine (PPT) thiol of the neighbouring PCP domain (Figure 1.21A). A domains are approximately 550 aa long and consist of two subunits, a small *C*-terminal fragment and a large *N*-terminal subunit, in between which the active site is located. This substrate binding pocket is flanked by a loop which covers the active site during catalysis to initiate the amino acid activation and to shield the involved ATP molecule from water.[111] The catalytic pocket also contains eight amino acid residues that coordinate the amino acid substrate

and are therefore responsible for the specificity of amino acid incorporation in the formed natural product.[112]

1.3.6.5.2 PCP domain

Only 80-110 aa in length, the PCP domain serves as a transporter of the activated amino acid on to the next fragment in the biosynthetic assembly. PCP domains are nonpolar proteins arranged in a bundle of four α-helices while the active site is located just before the second helix which determines the PCP interactions with neighbouring proteins.[99] Posttranslational modification of a serine residue in the active site with PPT converts this domain into its active holo form.

1.3.6.5.3 C domain

Condensation domains catalyse peptide bond formation between the fragments of two adjacent modules. The approximately 450 aa long domain contains one nucleophilic acceptor site for the downstream activated amino acid and one electrophilic donor site which interacts with the upstream PCP thioester substrate. The C domain forms an αβα-sandwich while solvent channels allow the domain to accept two substrates simultaneously. By facilitating the nucleophilic attack of the downstream amino acid on the upstream acyl group, C domains are enabling a directional assembly of the amino acids determined by the A domain specificity (Figure 1.21B).[99]

1.3.6.6 Polyketide synthase extension

The final extension of the backbone of TMC-86A is facilitated by a type I PKS enzyme. TmcH elongates the peptide backbone with a malonyl unit, followed by di-methylation in the α-position.[97] Similar to NRPSs, type I PKS enzymes are multifunctional and formed of modules which are composed of catalytic domains (Figure 1.22 and Figure 1.23).[113] The key domains involved in TMC-86A biosynthesis are an acyl carrier protein (ACP) domain, acyl transferase (AT) domain, ketosynthase (KS) domain, *C*-methyltransferase (cMT) domain, and a thioesterase (TE) domain.
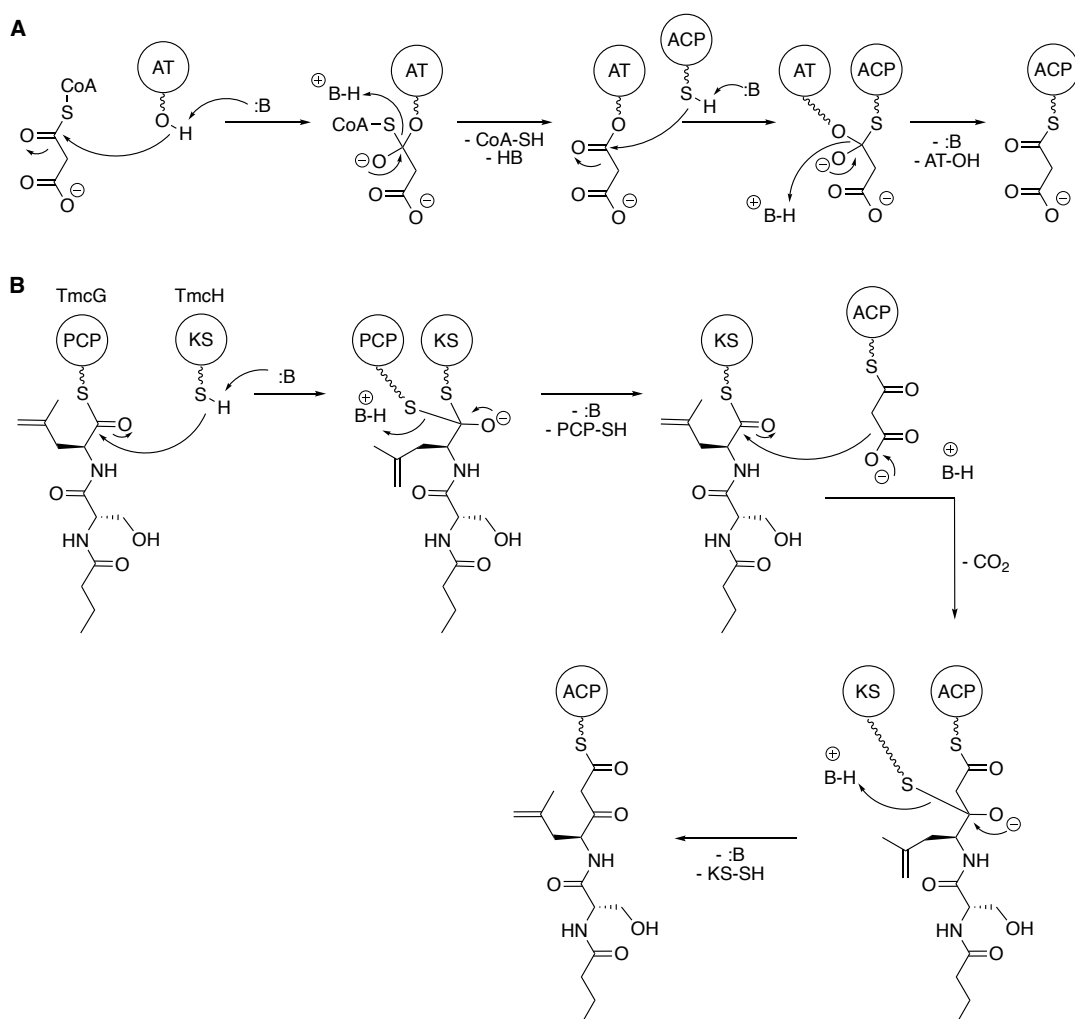
Figure 1.22 Detailed mechanism of the AT, ACP, and KS domains involved in the polyketide assembly in the TMC-86A biosynthesis. A. The AT domain transfers the malonyl unit onto the PKS ACP domain. B. The NRPS product is transferred to the KS domain and fused to the malonyl unit on the ACP domain by carbon-carbon bond formation catalysed by the KS domain.

1.3.6.6.1   ACP domain

Acyl carrier protein domains are approximately 80 aa long and organised in a bundle of four helices similar to PCP domains in NRPS modules. ACP domains function as a transporter for extender units and growing polyketide intermediates. The substrates are covalently bound to a PPT thiol which activates them to be fused to the assembled polyketide of the upstream module.[113]

1.3.6.6.2   AT domain

Acyl transferase domains contain two subunits of a combined length of approximately 300 aa and are responsible for loading ACP domains with specific extender units.

Common extender units are malonyl-CoA and methyl malonyl-CoA which can be transferred to the AT domain by nucleophilic attack by the AT catalytic serine residue which was previously activated by an adjacent histidine residue. Subsequently the malonyl-AT is transferred to the PPT thiol of the ACP domain (Figure 1.22A).[113]

1.3.6.6.3 KS domain

Ketosynthase domains contain approximately 430 aa residues and facilitate the chemically challenging carbon-carbon bond formation between an extender unit and a polyketide intermediate of the upstream module. KS domains form alternating layers of α-helices and β-sheets and exist in dimers. The two active sites connect both substrates at the dimer interface to catalyse the newly formed carbon-carbon bond with one catalytic serine and two catalytic histidine residues in each substrate binding pocket. KS domains have been studied broadly and found to be the most conserved domain in PKS modules. They also exhibit substrate selectivity and are the most crucial part in the assembly of polyketide backbones (Figure 1.22B).[113]

1.3.6.6.4 cMT domain

C-Methyltransferase domains are similar to MT domains in NRPS modules and use SAM as a cofactor to facilitate the methylation in the α-position of an ACP β-keto thioester. Methyltransferases in PKS modules are approximately 320 aa long and contain two subunits, one that binds the substrate and one that binds the cofactor. MT domains are often structurally inserted into other domains and catalyse the carbon deprotonation to facilitate the nucleophilic attack on the electrophilic methyl group of cofactor SAM (Figure 1.23A).[113]
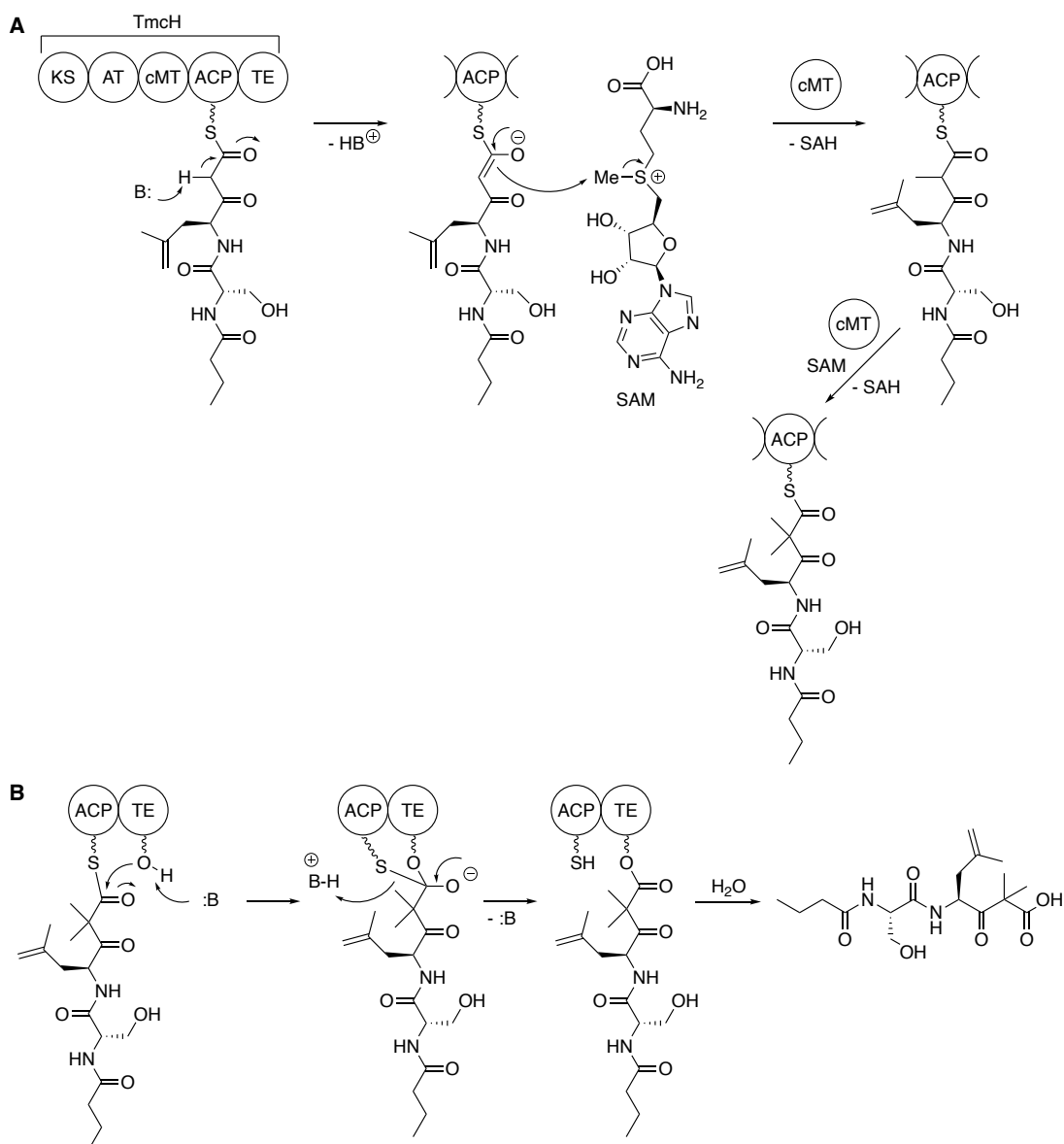
Figure 1.23 Detailed mechanism of the cMT and TE domain in the polyketide assembly in the TMC-86A biosynthesis. A. The malonyl α-carbon is subsequently di-methylated by two cycles of cMT mediated methylation using the cofactor SAM. B. The fully assembled backbone of TMC-86A is transferred to the TE domain and hydrolysed by water to release the α-dimethyl-β-keto carboxylic acid.

1.3.6.6.5   TE domain

Thioesterase domains are 240-290 aa long and facilitate the release of the assembled polyketide from the final PKS module. The substrate of the final module's ACP domain is loaded on the TE domain forming an acyl intermediate which is subsequently released. The polyketide backbone can either be cyclised by a nucleophilic group in the assembled molecule or hydrolysed by water to release a carboxylic acid (Figure 1.23B).[113]

1.3.6.7 Epoxyketone formation

All known epoxyketone producing biosynthetic gene clusters contain a conserved flavin-dependent acyl-CoA dehydrogenase (ACAD) homologue which is responsible for creating the epoxyketone pharmacophore. The assembled backbone is released from the final PKS module to form an α-dimethyl-β-keto carboxylic acid which subsequently is converted into the epoxyketone by the ACAD homologue TmcF before it is oxidised by cytochrome P450 TmcI to the final product TMC-86A (Figure 1.24).
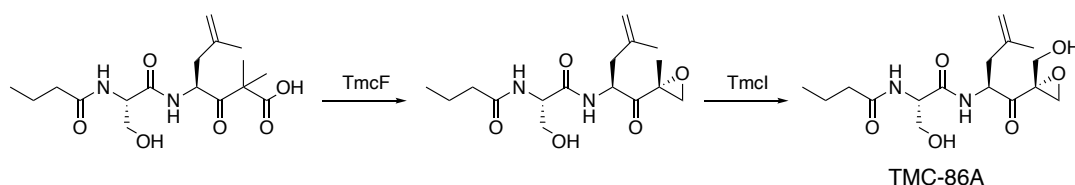
Figure 1.24 Final steps in the biosynthetic assembly of TMC-86A. TmcF forms the epoxyketone and TmcI is hydroxylating the methyl group.

1.3.6.8 Biosynthesis of other epoxyketones

Epoxyketone natural products contain a backbone assembled by an FAS, an NRPS, and a PKS, although the individual components vary between compounds. The *N*-terminal fatty acid chains consist of different length and branching while some are hydroxylated. Various amino acids appear to be incorporated by the different NRPSs although the majority of known epoxyketones contain a leucine or dehydroleucine adjacent to the epoxyketone. All earlier discovered epoxyketones appear to be assembled from α-dimethyl-β-keto and some were additionally hydroxylated. Some of the more recently discovered epoxyketones do not have the additional methyl group at the epoxyketone stereocenter which suggests they were formed from monomethylated substrates.

Because the structure of eponemycin and TMC-86A are very similar, Challis group member Dr Chuan Huang performed a series of experiments to compare the *epn* and *tmc* clusters which revealed that the *epn* cluster also produces TMC-86A. Heterologous expression experiments demonstrated that the *epn* fatty acid synthase in

fact produces a butanoyl residue and that TMC-86A is the main metabolite of the *epn* cluster when the *epn* fatty acid synthase is overproduced. It is currently assumed that the fatty acid synthase of the *epn* cluster in *S. hygroscopicus* is not well expressed and that the ordinary fatty acid 6-methylheptanoic acid is used instead leading to production of eponemycin a shunt metabolite.

1.4 Epoxyketone synthase EpnF

Epoxyketone synthases are Acyl-CoA dehydrogenase (ACAD) homologues and were thus suspected to catalyse the epoxyketone formation via a dehydrogenation mechanism similar to known ACADs. ACADs catalyse the α,β-dehydrogenation using flavin as a cofactor and the desaturation catalytic mechanism of medium chain length acyl-CoA has been studied thoroughly.[114] The α-hydrogen is removed by a glutamic acid residue in the ACAD active site as a proton and the β-hydrogen is transferred to the flavin cofactor as a hydride in a concerted manner while further ACAD hydrogen donor residue stabilise the substrate and lower the pK_a of the α-hydrogen (Scheme 1.5).
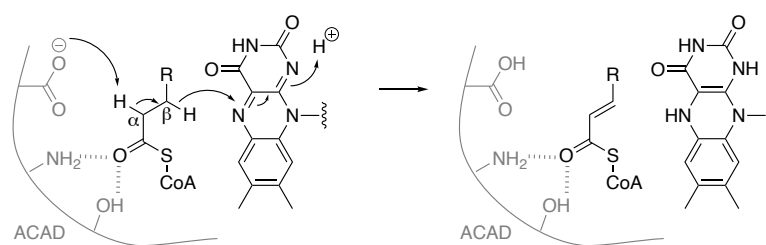
Scheme 1.5 Mechanism of the ACAD catalysed α,β-desaturation.

To elucidate the mechanism of epoxyketone synthase catalysed epoxide formation, previous Challis group member Dr Cartwright investigated the *tmc* and *epn* clusters and the corresponding epoxyketone synthases. Attempts to overproduce and purify TmcF were not successful, however, a protocol for EpnF overproduction in *E. coli* and purification has been established. *In vitro* enzyme assays with epoxyketone synthase EpnF were developed to study its activity and reaction mechanism.[97]

1.4.1   EpnF overproduction and purification

The epoxyketone synthase encoding gene *epnF* was amplified from *S. hygroscopicus* genomic DNA and cloned into a pET151 expression vector which was transformed into *E. coli* One Shot BL21 Star (DE3) cells for overexpression. The pET151 expression system allows for controlled induction of enhanced protein expression (Appendix 8.1). The desired protein is expressed with a hexa-histidine (His$_6$) tag at the *N*-termini which allows for purification by immobilized metal affinity chromatography (IMAC).

EpnF was overproduced, purified with Ni-affinity chromatography, and analysed by ESI-Q-TOF mass spectrometry and peptide mass finger printing to confirm the correct protein has been purified. A yellow colour with a UV absorbance maximum of approximately 450 nm suggested the presence of a flavin cofactor which was later confirmed to be flavin adenine dinucleotide (FAD).[115]

1.4.2 EpnF *in vitro* assays with α-dimethyl-β-keto carboxylic acid substrate

Purified EpnF was used in *in vitro* enzyme assays to produce epoxyketones from synthetic substrates. The initial hypothesis was that the α-dimethyl-β-keto carboxylic acid released from the PKS module would decarboxylate to the isopropyl ketone to act as the EpnF substrate. Incubation of the isopropyl ketone however did not show conversion to the desired epoxyketone, suggesting the α-dimethyl-β-keto carboxylic acid is the substrate which was confirmed by observation of epoxyketone formation when incubating the acid with EpnF (Scheme 1.6).
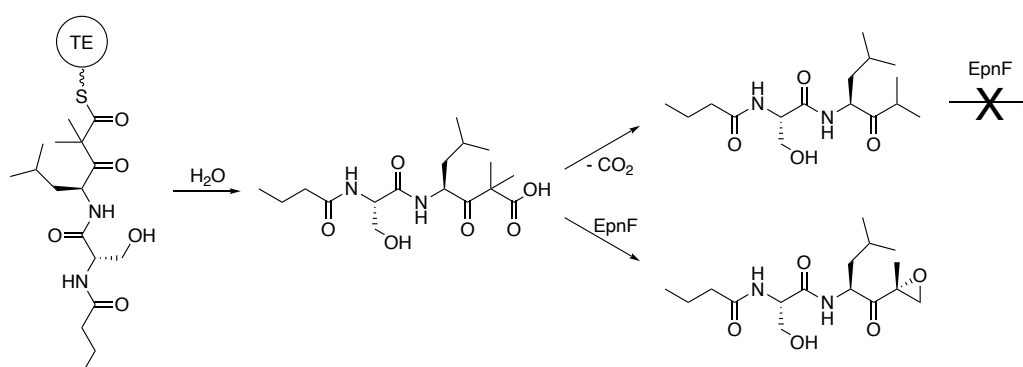
Scheme 1.6 Reaction scheme of TMC-86A precursor once it reached the TE domain. The TE-thioester is hydrolysed to the carboxylic acid which either decarboxylates to the corresponding isopropyl ketone, which is not converted by EpnF, or the carboxylic acid which functions as the suitable substrate for EpnF and is transformed into the corresponding epoxyketone.

The mechanism was proposed to proceed via an unprecedented decarboxylation-dehydrogenation-monooxygenation reaction (Scheme 1.7).[97,114] Decarboxylation of the α-dimethyl-β-keto carboxylic acid could generate an enolate intermediate which would eliminate a β-hydride to the flavin cofactor to form an α,β-unsaturated ketone while still being bound to the enzyme. However, because known ACAD enzymes catalyse a concerted desaturation mechanism, it is also possible that the decarboxylation and β-hydride elimination to the cofactor occur simultaneously to

generate the α,β-unsaturated ketone. The reduced flavin subsequently reacts with molecular oxygen to form flavin hydroperoxide which reacts with the unsaturated ketone in an addition-1,3-elimination to yield the final epoxyketone product.
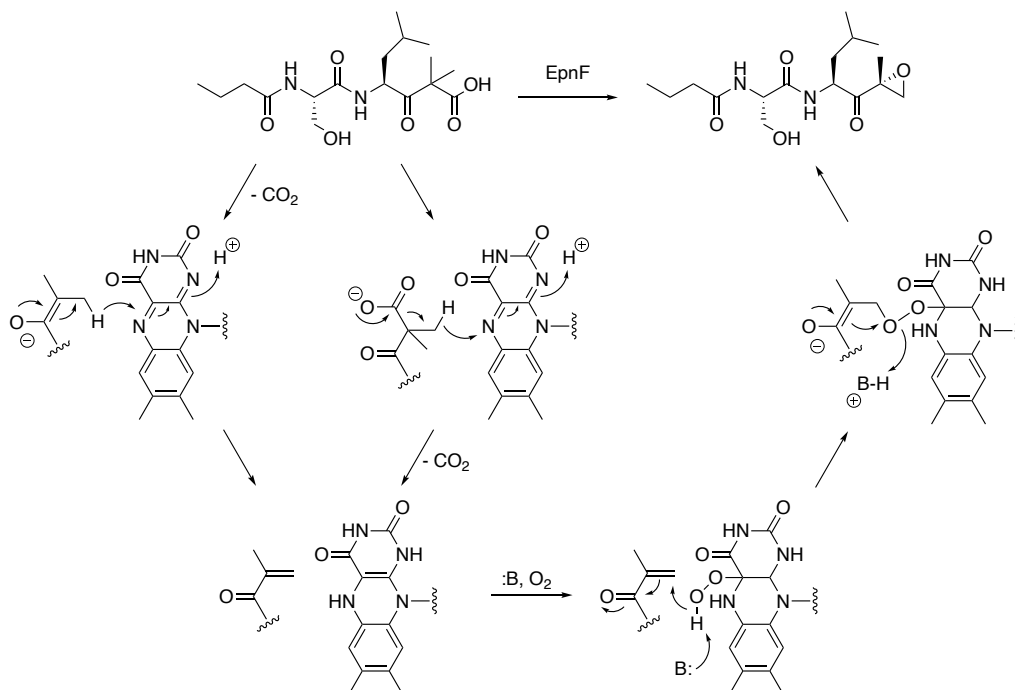
Scheme 1.7 *Proposed reaction mechanism of epoxyketone formation by EpnF with the α-dimethyl-β-keto carboxylic acid substrate.*

The main challenge of these *in vitro* assays is the instability of the acid substrate due to its propensity to spontaneously decarboxylate and tautomerize to form ketones. The decarboxylation of α-dimethyl-β-keto carboxylic acids is facilitated by the β-carbonyl which is an electron withdrawing group (EWG) and hydrogen bonding partner for the carboxylic acid hydroxyl group.[116] This leads to a cyclic transition state forming an enol after loss of CO_2 and tautomerization produces isopropyl ketones (Scheme 1.8).
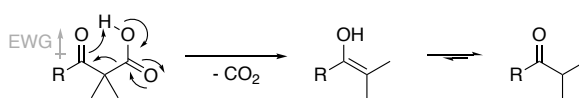
Scheme 1.8 *Mechanism of the spontaneous decarboxylative degradation reaction of α-dimethyl-β-keto carboxylic acids. The degradation is initiated by an intramolecular deprotonation in a cyclic transition state which releases carbon dioxide and forms an enol that subsequently tautomerizes to the corresponding isopropyl ketone.*

1.4.3 Coupled pig liver esterase (PLE) and EpnF enzyme assays

To overcome the obstacle of an unstable substrate for epoxyketone synthases, previous Challis group member Dr Cartwright explored using chemically stable methyl ester precursors. Methyl esters could be hydrolysed into the acids by rather harsh chemical conditions which would probably require an additional work up step before any enzyme could be added. Alternatively, esters can be cleaved under mild conditions by promiscuous ester hydrolases to generate carboxylic acids *in situ*. The commercially available pig liver esterase (PLE) was tested with α-dimethyl-β-keto methyl ester **31** and provided ester hydrolysis to the desired α-dimethyl-β-keto carboxylic acid epoxyketone synthase substrate **32**. PLE is a promiscuous serine hydrolase which catalyses ester cleavage of a variety of substrates with a catalytic triad of serine, histidine, and aspartate residue (Figure 1.25).[117] The activated catalytic serine nucleophilically attacks ester substrates to form an acyl-ester intermediate that is subsequently cleaved hydrolytically.[118] Substrate promiscuity of PLE is ensured by using mixtures of isoenzymes with different substrate specificities.[119]
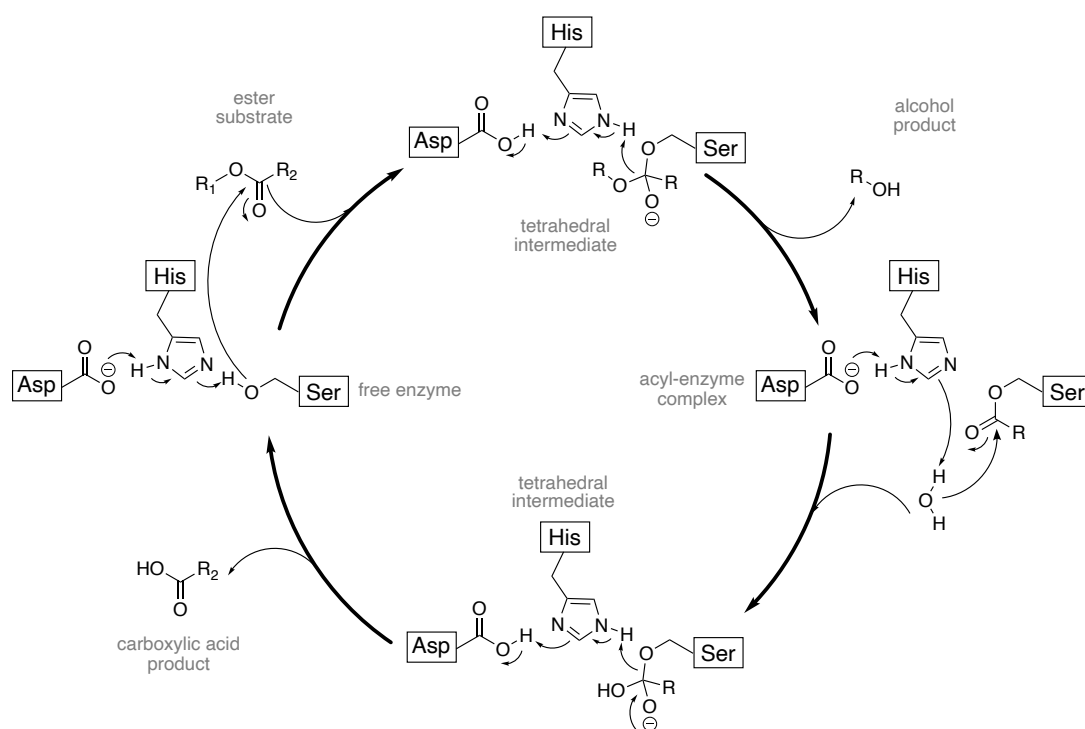
Figure 1.25 Mechanism of the aspartate, histidine, serine catalytic triad facilitated ester hydrolysis.

The substrate precursor **41** was then incubated with PLE and EpnF simultaneously in coupled enzyme assays to generate the epoxyketone synthase substrate **42** *in situ* to be

converted by EpnF instantaneously. Production of the desired epoxyketone **43** was observed providing proof of principle for the use of stable methyl ester precursors in coupled enzyme assays to investigate epoxyketone synthase activity (Scheme 1.9).
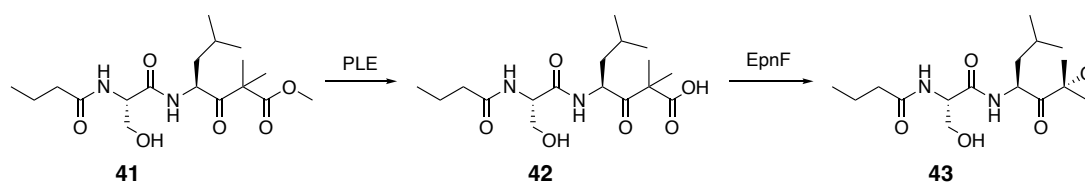
*Scheme 1.9 Expected reactions of the coupled enzyme assays with PLE and EpnF. PLE hydrolyses the chemically stable methyl ester substrate **41** to the corresponding carboxylic acid **42** which is subsequently transformed into the epoxyketone **43** by EpnF.*

1.4.4 Preliminary investigation into EpnF substrate tolerance

The majority of epoxyketone natural products contain an L-leucine or dehydro-L-leucine residue adjacent to the epoxyketone warhead, although there are exceptions like clarepoxcin A with an L-alanine and tryptopeptin A with an L-tryptophan residue. To start the investigation of the substrate tolerance of EpnF towards non-native substrates with varying amino acid residues, Dr Cartwright synthesised an alanine derived β-keto ester substrate analogue. This sterically less hindered substrate analogue was incubated in the coupled enzyme assays with PLE and EpnF to confirm that EpnF is able to catalyse the epoxyketone formation on this substrate analogue.[115]

1.5 Project aims

Epoxyketone proteasome inhibitors play a major role in the treatment of multiple myeloma and gained further interest for potential treatment of autoimmune disorders. FDA approved drugs such as carfilzomib and drug candidates in clinical trials like oprozomib and KZR-616 are assets to the portfolios of pharmaceutical companies like Amgen. During the large-scale production of epoxyketone drugs, one of the main drawbacks remains to be the challenging and inefficient epoxyketone formation. All synthetic epoxyketone proteasome inhibitors are based on the natural products which are known to be produced biosynthetically by highly selective and efficient catalytic enzymes. Recent studies on the reaction mechanism of epoxyketone synthase EpnF offer novel insights into biologically catalysed epoxyketone formation. Exploring epoxyketone synthases further could be the first step towards the development of a more sustainable enzyme catalysed epoxyketone formation for proteasome inhibitor production (Scheme 1.10). Based on previous research from the Challis group, this project was focusing on different aspects to further investigate epoxyketone synthases by working toward the following aims.
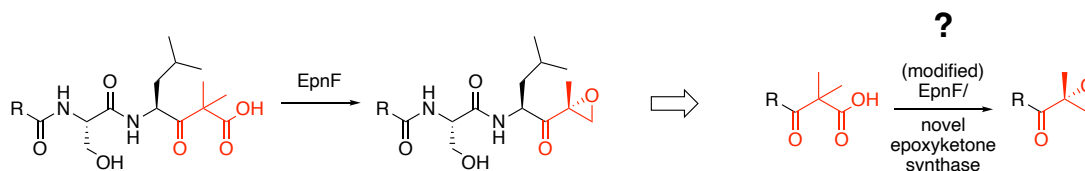
Scheme 1.10 Visual representation of the project's underlying research question: If EpnF facilitates the epoxyketone formation of the native α-dimethyl-β-keto carboxylic acid substrate, can EpnF or novel or modified epoxyketone synthases efficiently catalyse the general epoxyketone formation from these substrates for chemoenzymatic proteasome inhibitor production?

1.5.1   Exploring the substrate tolerance of epoxyketone synthase EpnF

The first aim of this project was to further investigate the substrate tolerance of epoxyketone synthase EpnF. A coupled enzyme assay with PLE and EpnF was previously developed that enables testing chemically stable α-dimethyl-β-keto methyl ester substrate analogues with epoxyketone synthases. Reproducing and optimising the chemical substrate synthesis will provide a suitable starting point to produce a set of substrate analogues. The developed coupled enzyme assay with PLE and EpnF is expected to be reproduced and conditions optimised to investigate if EpnF is able to

convert the synthesised substrate analogues into epoxyketones (Scheme 1.11). Clinical trial drugs oprozomib and KZR-616 contain a phenylalanine and cyclopentenylalanine residue respectively next to the epoxyketone which would make phenylalanine and cyclopentenylalanine derived α-dimethyl-β-keto methyl ester substrate analogues commercially relevant targets to test.
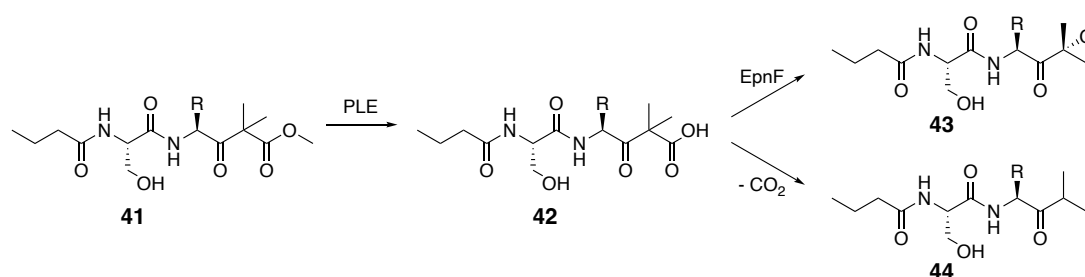
Scheme 1.11 Coupled enzyme assays with PLE and EpnF to test different substrate analogues with varying amino acids (R) next to the site of reaction.

1.5.2 Towards commercially relevant epoxyketone proteasome inhibitors

EpnF is obtainable as recombinant purified enzyme and exhibits *in vitro* activity which makes it the first epoxyketone synthase that can be studied for chemoenzymatic purposes. The epoxyketone proteasome inhibitor oprozomib is currently in clinical trials which makes it a potential target for chemoenzymatic production. To investigate whether EpnF's activity goes beyond simple substrate analogues and if it could be used to produce commercially relevant epoxyketone drugs, an α-dimethyl-β-keto methyl ester precursor for oprozomib will be synthesised and tested in the coupled PLE/EpnF enzyme assays (Scheme 1.12).
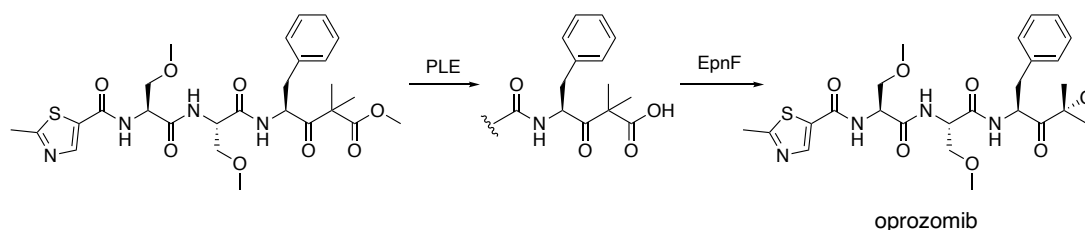
Scheme 1.12 Potential reactions in coupled enzyme assays with PLE, EpnF, and a stable oprozomib methyl ester precursor.

1.5.3 Identification and analysis of novel biosynthetic gene clusters

EpnF is currently the only documented epoxyketone synthase that has been purified and studied *in vitro*. If epoxyketone synthases could be used chemoenzymatically or potentially as part of novel designed biosynthetic pathways for proteasome inhibitor drugs, it is important to gain more insight into these enzymes and investigate other epoxyketone synthases. Especially for proteasome inhibitors like oprozomib with a phenylalanine residue adjacent to the epoxyketone, it would be valuable to study epoxyketone synthases that are able to convert aromatic substrates. This project aims for the *in silico* identification and analysis of novel potential epoxyketone synthases with specificity towards aromatic substrates. If novel epoxyketone synthases can be identified, it would be of interest to clone and express the encoding genes to overproduce and purify novel epoxyketone synthases for *in vitro* studies.

2. Exploring the substrate tolerance of epoxyketone synthase EpnF

2.1 Synthesis of EpnF substrate analogues

2.1.1 Choice of substrate analogues

The first aim of this project was to investigate the substrate tolerance of epoxyketone synthase EpnF by testing its activity with various substrate analogues. Previous studies by former PhD student Dr Josh W. Cartwright suggested that the native α-dimethyl-β-keto carboxylic acid substrates are unstable due to decarboxylation and thus not suitable substrates for *in vitro* enzyme assays. The synthesis of a α-dimethyl-β-keto methyl ester substrate provided a chemically stable alternative, of which the ester can be cleaved *in situ* by pig liver esterase (PLE) to minimize decarboxylation side reactions.[115] Based on these studies, a range of α-dimethyl-β-keto methyl ester substrates was synthesised with different amino acids next to the α-dimethyl-β-keto group.

Since the native substrate of EpnF contains an L-leucine residue in the position next to the site of epoxide formation, a leucine derived α-dimethyl-β-keto methyl ester substrate was of interest. Cleavage of the methyl ester *in situ* would form the native substrate for EpnF and would provide a good reference assay to compare the enzyme's activity with the native substrate to non-native substrates.

Previous preliminary studies by Dr Cartwright also included an alanine derived substrate because it is present in the natural product epoxyketone clarepoxcin A. Additionally, an alanine residue next to the α-dimethyl-β-keto methyl ester would provide a substrate with minimal sterical hinderance and minimal interactions in the active site. To reproduce Dr Cartwright's preliminary studies and provide a sterically minimally hindered substrate analogue, an alanine derived α-dimethyl-β-keto methyl ester was synthesised.

While the majority of epoxyketone secondary metabolites incorporate L-leucine adjacent to the epoxyketone moiety, the other amino acid in this position observed besides clarepoxcin with the alanine residue, is L-tryptophan in tryptopeptin A (**22**).

A tryptophan derived substrate enables to investigate EpnF's activity towards aromatic side chains and substrates that are sterically significantly more demanding than the native substrate.

To screen different functional groups and chemical properties, the goal was to synthesise substrates with a polar, a negatively charged, and a positively charged amino acid residue. The planned substrate analogues containing serine, glutamic acid, and lysine next to the α-dimethyl-β-keto group had to be made from side chain protected starting materials that added an additional deprotection step at the end of the synthetic route of the other substrate analogues.

To increase the application towards clinically relevant epoxyketones, a phenylalanine and a cyclopentenylalanine (Cpa) derived substrate were synthesised based on the epoxyketone drugs oprozomib and KZR-616 which are currently in clinical trials for treatment of multiple myeloma and autoimmune diseases respectively.[37,71] Finding an epoxyketone synthase that is able to convert aromatic or cyclopentenylalanine derived substrates would provide a promising starting point for further studies towards the chemoenzymatic synthesis of epoxyketone proteasome inhibitors like oprozomib (**14**) and KZR-616 (**29**).

The set of chosen substrate analogues covers a range of hydrophobic, polar, and charged side chains which provides a suitable starting point to investigate the substrate tolerance of epoxyketone synthase EpnF (Figure 2.1).
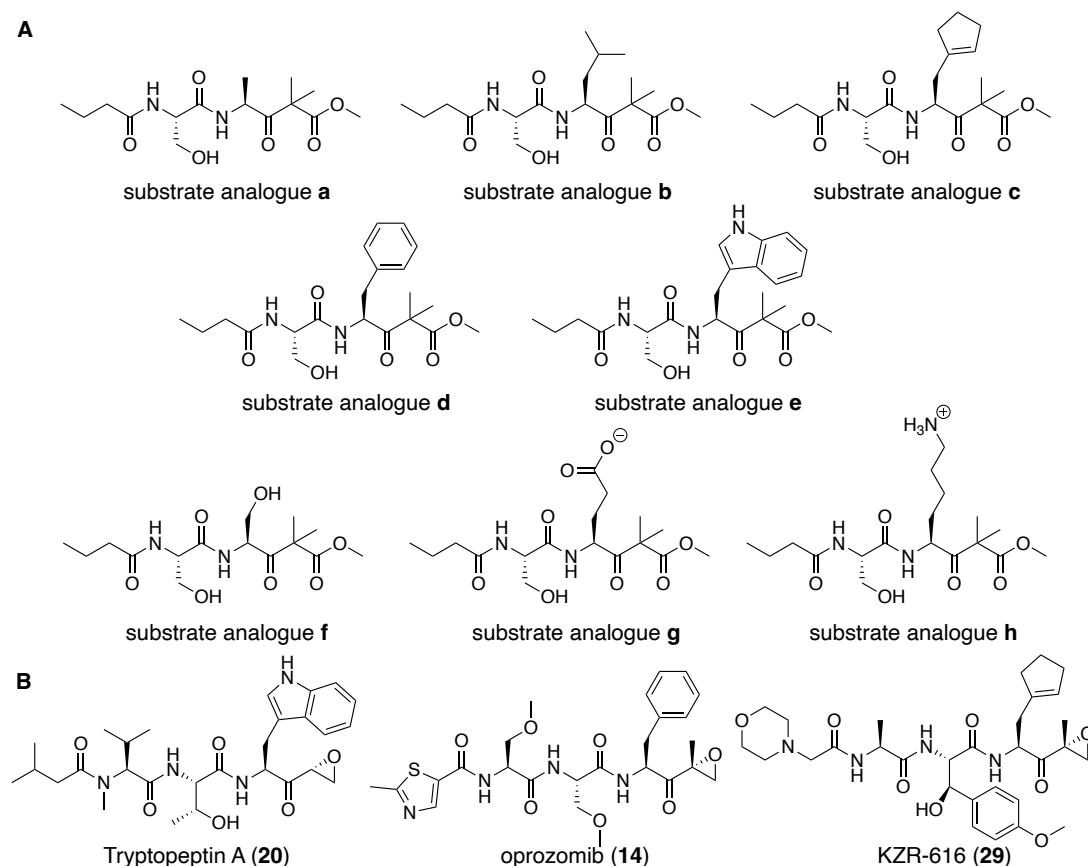
*Figure 2.1 Structures of substrate analogues and commercially relevant epoxyketones. A. Chemical structures of substrate analogues **a-h** to be synthesised for EpnF in vitro activity assays. B. Structures of epoxyketone natural product tryptopeptin A (**20**) and proteasome inhibitors oprozomib (**14**) and KZR-616 (**29**) which are currently in clinical trial for treatment of multiple myeloma and autoimmune disorder respectively.*

2.1.2 Proposed synthetic strategy

The synthetic strategy to obtain the chosen substrate analogues was based on Dr Cartwright's synthetic route for stable methyl ester substrate analogues.[115] A serine butanoic acid fragment is synthesised (fragment 1) which is then coupled to an amino acid α-dimethyl-β-keto-methyl ester fragment (fragment 2) as shown in Scheme 2.1.
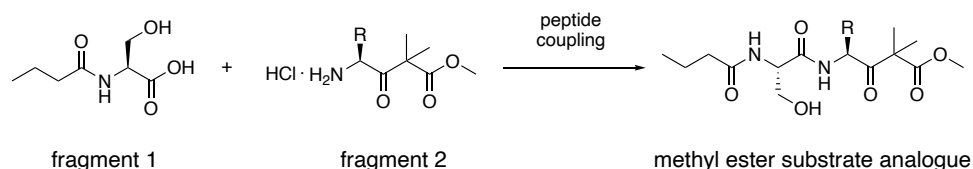
Scheme 2.1 Synthetic strategy to produce α-dimethyl-β-keto-methyl ester substrate analogues with different amino acid residues R. Fragment 2 will be synthesised separately for each chosen substrate analogue to subsequently be coupled to fragment 1.

For each desired substrate analogue, a fragment 2 with the appropriate amino acid residue will be synthesised from the corresponding *N*-Boc-protected L-amino acid (condition details described in Chapter 2.1.3). Alanine, leucine, phenylalanine, and tryptophan starting materials are commercially available as *N*-Boc-protected amino acids and were purchased. Serine, glutamic acid, and lysine derived fragment 2 compounds needed to be synthesised from side chain protected starting materials. Protecting groups were chosen based on commercial availability of the starting material and orthogonal removal conditions of the protecting groups throughout the synthetic route. Benzyl (bn) side chain protected serine, benzyl ester side chain protected glutamic acid, and carboxybenzyl (cbz) side chain protected lysine were all purchased with additional *N*-Boc-protection. For the cyclopentenylalanine derived substrate analogue, the *N*-Boc-protected artificial amino acid for the synthesis of cyclopentenylalanine fragment 2 had to be synthesised (details described in Chapter 2.1.3).

2.1.3   Synthesis of substrate analogue fragments

The synthesis of fragment 1 was achieved by following Dr Cartwright's two step protocol while replacing DCC with EDC to simplify the work up of the first step of the synthesis. L-Serine methyl ester hydrochloride (**45**) was coupled to butyric acid with EDC and triethylamine to form methyl ester **46** and peroxidative deprotection yielded fragment 1, butyryl-L-serine (**47**) (Scheme 2.2).
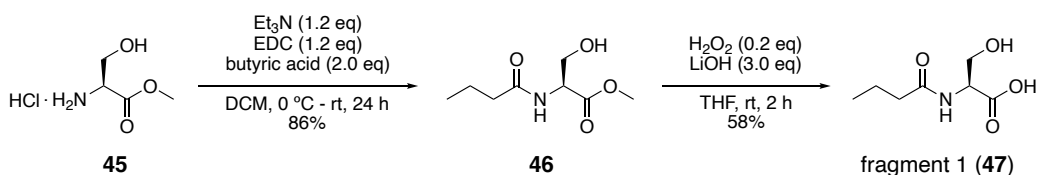
Scheme 2.2 Reaction conditions and yields for the synthesis of fragment 1.

The various fragment 2 molecules were generated from the corresponding *N*-Boc-L-amino acids which were all purchased besides the non-proteogenic amino acid cyclopentenylalanine which needed to be synthesised. *N*-Boc-L-cyclopentenyl alanine was assembled following the published procedure of Johnson *et al.* according to Scheme 2.3.[37]

Cyclopentanone (**48**) and triflic anhydride were used to prepare cyclopentenyl trifluoromethanesulfonate (**49**) which was coupled to *N*-Boc-β-idoalanine methyl ester immediately to yield *N*-Boc-cyclopentenylalanine methyl ester (**50**). Deprotection of the methyl ester under basic conditions provided *N*-Boc-cyclopentenylalanine (**51**) in an overall yield similar to the published procedure.
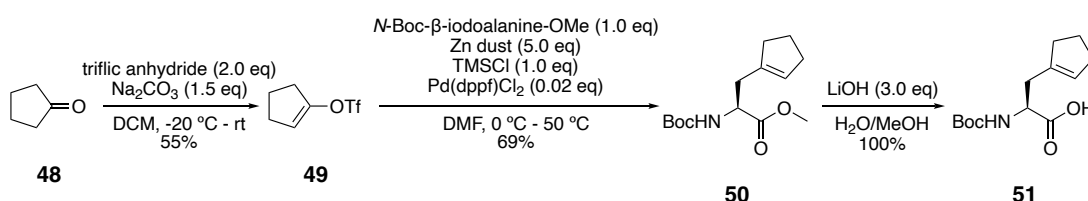
Scheme 2.3 Synthetic route to produce N-Boc protected cyclopentenylalanine.

With all starting materials in hand, the different *N*-Boc-L-amino acids (**52a-h**) were coupled with methyl potassium malonate to β-keto esters **53a-h** in 56-95% yield. The coupling with *N*-Boc-(cbz)-lysine (**52h**) provided the lowest yield of 56% while the other reactions fall into a range of 72-96% yield. It is possible that the sterically demanding properties of the cbz protected lysine side chain make the coupling reaction less efficient. Di-methylation of the extended amino acid malonate esters to **54a-h** and deprotection with HCl in dioxane provided all desired fragment 2 molecules **55a-h** (Scheme 2.4). Individual yields for each step are shown in Appendix **Error! Reference source not found.**.
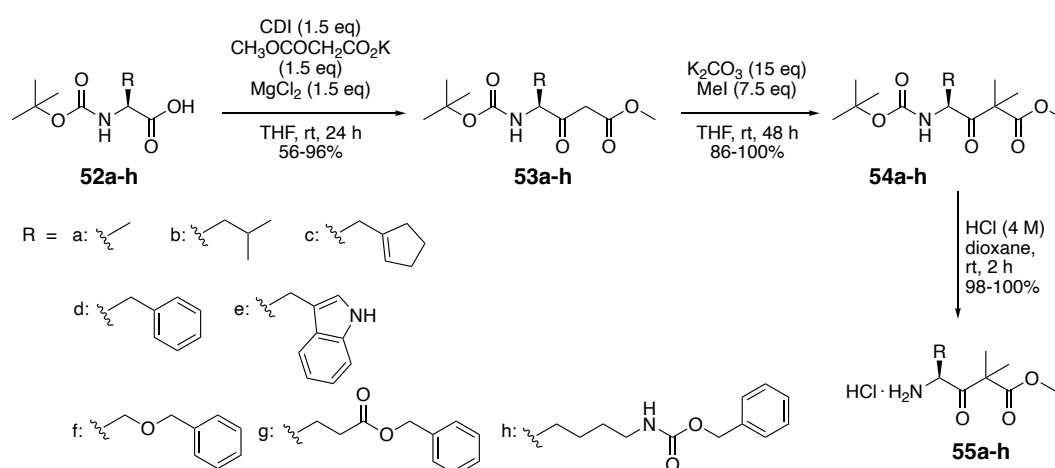
Scheme 2.4 Synthetic scheme for the fragment 2 molecules to assemble the derived α-dimethyl-β-keto methyl ester substrates.

2.1.4 Fragment coupling to assemble α-dimethyl-β-keto methyl ester substrate analogues

2.1.4.1 Initial coupling conditions to obtain alanine, leucine, and phenylalanine derived substrates

The first fragment 2 compounds that were synthesised were the alanine, leucine, and phenylalanine derived fragments **55a**, **55b**, and **55d**. Initially, the final coupling step of fragment 1 and fragment 2 was attempted with the published coupling conditions of EDC, HOBt, and triethylamine. The alanine, leucine, and phenylalanine derived methyl ester substrate analogues **56a,b,d** were obtained in 10-23% yield. Characterisation of the final products by NMR revealed that the compounds were produced as diastereomeric mixtures, probably due to racemisation of the alanine, leucine, and phenylalanine α-stereocenters (Scheme 2.5). The diastereomers appeared to have identical retention times in LCMS experiments since a single peak was observed for all three final products. Because the diastereomers were inseparable by silica gel column chromatography it was decided to use the diastereomeric α-dimethyl-β-keto methyl ester substrate analogues in upcoming experiments.
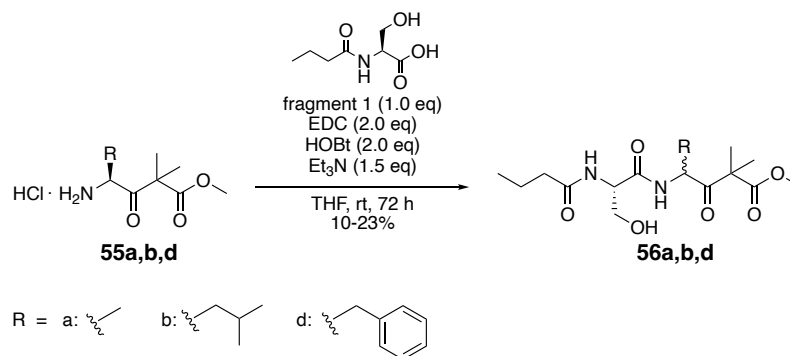
*Scheme 2.5 Final coupling of fragment 1 to alanine, leucine, and phenylalanine derived fragment 2 with EDC, HOBt, and triethylamine to yield the final substrate analogues **56a**, **56b**, and **56d** respectively.*

Enough material of the substrate analogues was generated (26-52 mg), however, due to low yields and safety issues with HOBt additional coupling conditions for the final step were tested.

2.1.4.2 Coupling condition optimisation

The phenylalanine substrate was not only obtained in very low yield of 16%, but it also was the only coupled fragment so far with commercial relevance due its identical amino acid in the equivalent position of the proteasome inhibitor oprozomib. Additionally, the majority of the remaining fragment 2 compounds to be coupled contained aromatic functional groups all of which is why the different coupling conditions were tested in the phenylalanine derived fragment coupling reaction.

Availability and safety issues of HOBt called for an alternative coupling additive and low yields might benefit from entirely different coupling conditions. Due to availability and recommendations, three additional coupling conditions were tested before the synthesis of further substrate analogues was attempted. The alternative additive oxyma was tested with coupling agent EDC, the combination reagent COMU was investigated, and coupling with *N*-methylimidazole (NMI) and *N,N,N',N'*-tetramethylchloroformamidinium hexafluorophosphate (TCFH) was attempted after it has been found to be suitable for challenging peptide coupling reactions (Figure 2.2).[120]
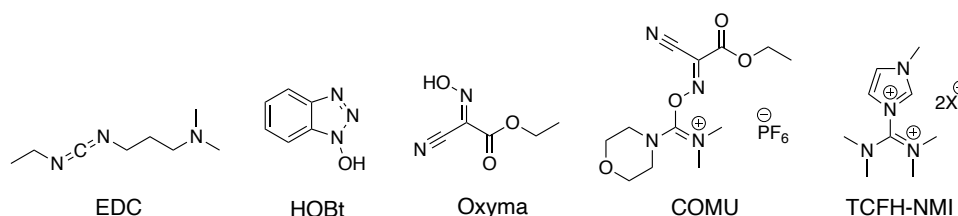
Figure 2.2 Structures and names of the tested coupling reagents and additives.

The alternative additive oxyma structurally differs from HOBt and facilitates peptide couplings through the same mechanism but without the risk of explosion. It was tested with EDC to couple fragment 1 to phenylalanine fragment 2 **55d** which provided a yield in the same range as the reactions with EDC and HOBt (11%). While the use of the combined reagent COMU improved the yield slightly (44%) the combination of NMI and TCFH provided the highest yield of 54% (Scheme 2.6).[120] Further substrate analogues were therefore coupled to fragment 1 with NMI-TCFH.
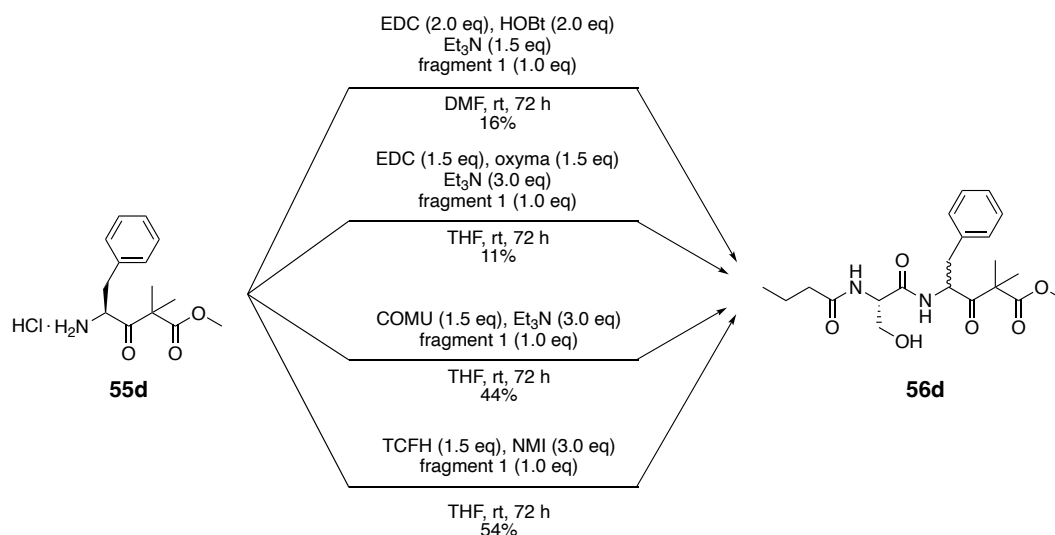
Scheme 2.6 Comparison of the four different coupling conditions tested for the coupling of fragment 1 (47) and phenylalanine derived fragment 2.

2.1.4.3 Final coupling of remaining substrate analogues

The cyclopentenylalanine, tryptophan, and side chain protected serine, glutamic acid, and lysine derived fragment 2 compounds **55c,e-h** were each coupled to fragment 1 using the optimised reaction conditions with NMI-TCFH (Scheme 2.7). Yields ranged from 17-38% and generated sufficient material of all desired α-dimethyl-β-keto methyl ester substrate analogues **56c,e-h** for full spectroscopic characterisation and planned experiments.
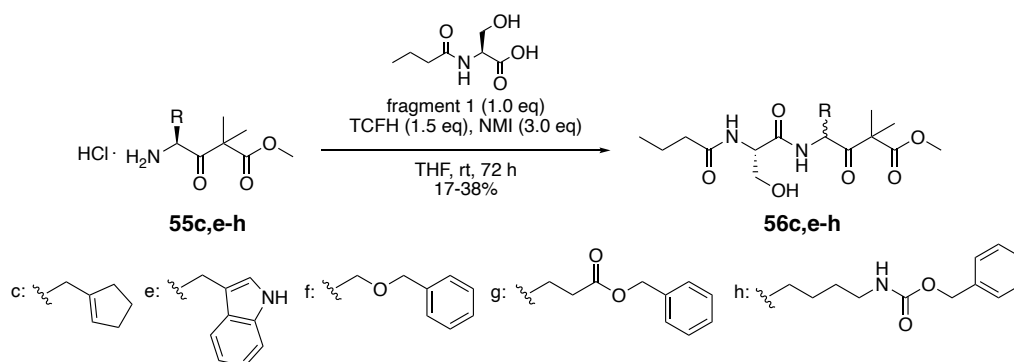
Scheme 2.7 Final coupling of fragment 1 to cyclopentenylalanine, tryptophan, serine, glutamic acid, and lysine derived fragment 2 to the final product α-dimethyl-β-keto methyl ester substrate analogues **56c**, **56e**, **56f**, **56g**, *and* **56h** *respectively.*

The coupling conditions could be further optimised by screening for additional coupling reagents and various reaction conditions such as different solvents or bases.

Additionally, protecting the serine hydroxy group might lead to an increased yield of the coupling reaction. However, because suitable amounts of all desired substrate analogues were obtained it was proceeded with the synthesised substrate derivatives and no further conditions were tested.

2.1.5 Attempted side chain deprotection of serine, glutamic acid, and lysine derived substrate analogues

The deprotection of all the side chains of **55f-h** was attempted via hydrogenation with Pd/C catalyst and hydrogen gas. Unfortunately, the deprotection of all three substrates yielded in cyclised products as the consequence of an intramolecular nucleophilic attack of the side chain functional group on the ester (Scheme 2.8). It was decided to not further pursuit the synthesis of these substrates because they appear to be fundamentally instable. Additionally, there are no known natural products with any polar or charged residues in this position and there is no clinical relevance of a serine, glutamic acid, or lysine side chain at this position.
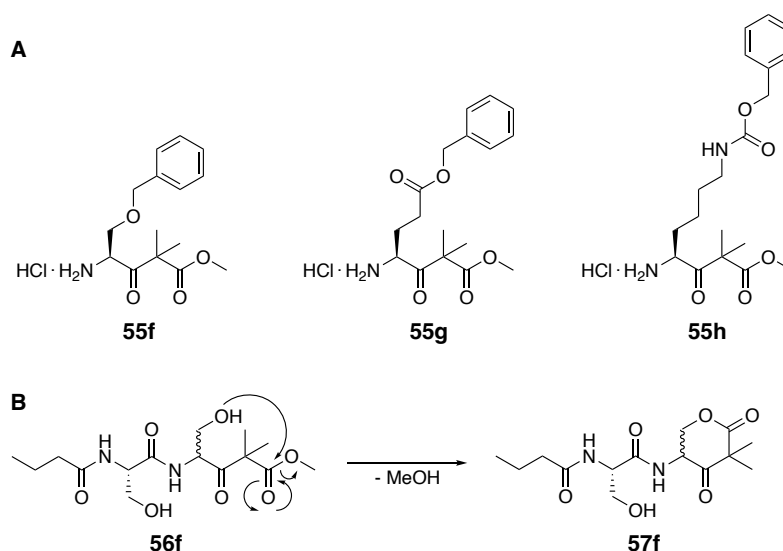
*Scheme 2.8 Side chain protected fragment 2 molecules and cyclisation side reaction. A. Structures of the serine, glutamic acid, and lysine side chain protected fragment 2 **55f-h**. B. Intramolecular cyclisation of the deprotected serine derived substrate (**56f**) to the cyclised product (**57f**).*

2.2 Overproduction and purification of epoxyketone synthase EpnF

The substrate tolerance of epoxyketone synthase EpnF was tested using *in vitro* enzyme assays with the previously synthesised substrate analogues. To obtain epoxyketone synthase EpnF, the published procedure for EpnF overproduction and purification was followed.[97] Overproduction of the desired protein in *E. coli* BL21 (DE3) containing the pET151 vector harbouring *epnF* was induced by the addition of IPTG. Additionally, at the point of induction, the flavin cofactor precursor riboflavin was added to the cultures to boost flavin production. EpnF was subsequently purified from the cell lysate by Nickel affinity chromatography to yield the desired protein (20.5 mg/L). An SDS-PAGE gel showed the presence of the purified protein and analysis by intact protein high-resolution mass spectrometry confirmed the molecular weight as 68.920 kDa (calculated for EpnF with the His6-tag, V5 tag, and TEV site: 69.921 kDa) (Figure 2.3).
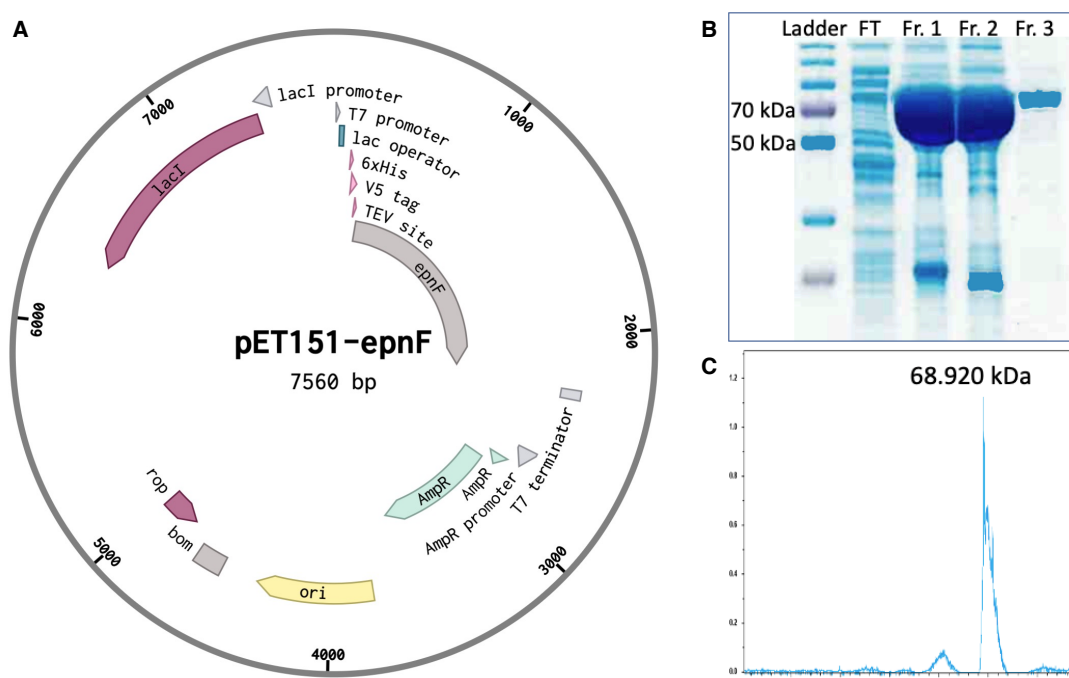
Figure 2.3 EpnF purification procedure. A. Plasmid map of pET151 harbouring epnF. B. SDS-PAGE gel of fractions from purification of EpnF (FT = flow through, Fr. 1 = 100 mM, Fr. 2 = 200 mM, Fr. 3 = 300 mM imidazole in elution buffer). C. Deconvoluted mass spectrum of EpnF.

2.3 Coupled PLE and EpnF *in vitro* enzyme assays with phenylalanine derived substrate analogue

2.3.1 Proof of concept

To explore the biosynthetic potential of EpnF, enzyme assays were conducted with the purified epoxyketone synthase and substrate analogues. Initial assays were set up with the phenylalanine derived α-dimethyl-β-keto methyl ester **56d** due to its commercial relevance. The phenylalanine derived substrate analogue **56d** was incubated with PLE and EpnF at 37 ºC for 3 hours. The assays were quenched with twice the volume of acetonitrile + 0.1% formic acid, centrifuged, and the supernatant was analysed by liquid chromatography mass spectrometry (LCMS). The relevant compounds during the coupled *in vitro* enzyme assays are expected to be the methyl ester substrate analogue **56d**, the hydrolysed carboxylic acid substrate **58d**, the corresponding epoxyketone product **59d**, and the spontaneous decarboxylated side product **60d** (Scheme 2.9). Obtained LCMS ion chromatograms were therefore extracted for these four species to observe enzyme activities.
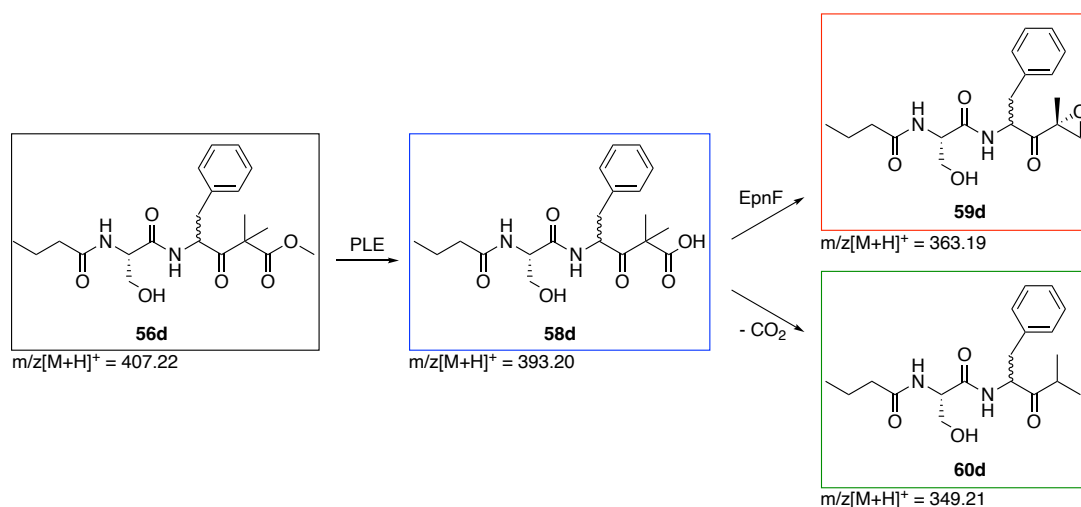
*Scheme 2.9 Expected reactions in enzyme assays with pig liver esterase (PLE), EpnF, and phenylalanine derived α-dimethyl-β-keto methyl ester **56d**. PLE is expected to cleave the methyl ester to the corresponding α-dimethyl-β-keto carboxylic acid **58d** which will be converted to the epoxyketone **59d** by EpnF or spontaneously decarboxylate to side product **60d**. m/z of the [M+H]+ ion of each species that will be monitored in LCMS chromatograms are shown.*

In the first experiment with 10 μM PLE and 100 μM EpnF, ester hydrolysis but no conversion to the epoxyketone product **59d** was observed. Setting up assays under the

same conditions with additional cofactor flavin adenine dinucleotide (FAD, 2 mM final concentration) showed conversion to the desired product (Figure 2.4).
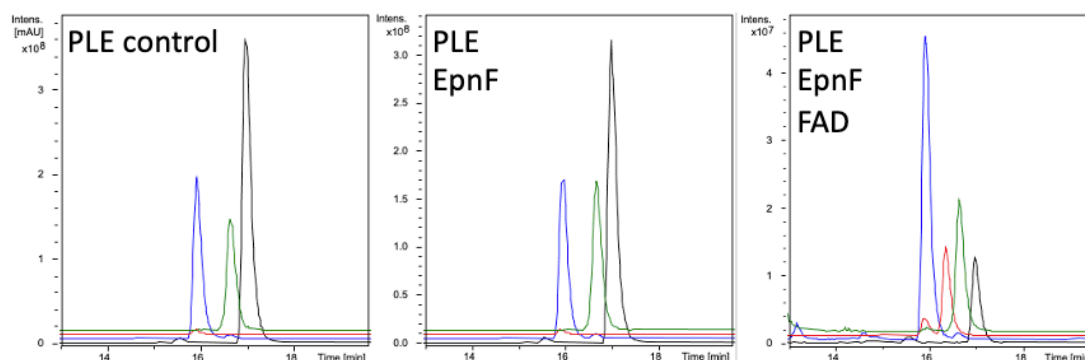
*Figure 2.4 Extracted ion chromatograms (EICs) from LCMS analysis of the reaction of phenylalanine derived α-dimethyl-β-keto methyl ester **56d** with EpnF and PLE. Black: EIC for m/z 407.22 corresponding to [M+H]$^+$ for the substrate **56d**. Blue: EIC for m/z 393.20 corresponding to [M+H]$^+$ for the product of ester cleavage **58d**. Red: EIC for m/z 363.19 corresponding to [M+H]$^+$ for the epoxyketone product **59d**. Green: EIC for m/z 349.21 corresponding to [M+H]$^+$ for the decarboxylated side product **60d**. Left: PLE only control reaction. Middle: Incubation of the substrate with PLE and EpnF. Right: Incubation of the substrate with PLE, EpnF, and additional FAD.*

Although these results show that EpnF is able to convert a phenylalanine derived substrate into the desired epoxyketone in *in vitro* coupled enzyme assays, the assay conditions appear to be suboptimal. The overall turnover is low, only small amounts of epoxyketone are produced and the decarboxylation reaction produces slightly more of the side product compared to the desired epoxyketone product. It is possible that EpnF is not very stable under the tested *in vitro* conditions and degrades quickly, however, it could also be that it is just not very active or efficient in general. It seems to be highly dependent on the additional cofactor boost despite the incorporation of cofactor during expression. Because only low levels of epoxyketone production were observed, a series of condition optimisation experiments has been conducted.

2.3.2   Condition optimisation

2.3.2.1 Length of the assays

To get a better understanding of the reactivity of EpnF and for how long it remains active, several time points of the assays were analysed. Quenching the assays after 1 hour, 3 hours, and 6 hours revealed that there was an increase in epoxyketone production in the 3-hour assays compared to the 1-hour time point. However, no EpnF activity was observed after 3 hours of incubation since the epoxyketone level

remained unchanged in the 6-hour assay compared with the 3-hour time point (Figure 2.5). This suggests that EpnF is degrading within 3 hours of the start of the coupled *in vitro* assays. All following assays were therefore incubated for 3 hours before quenching and LCMS analysis.
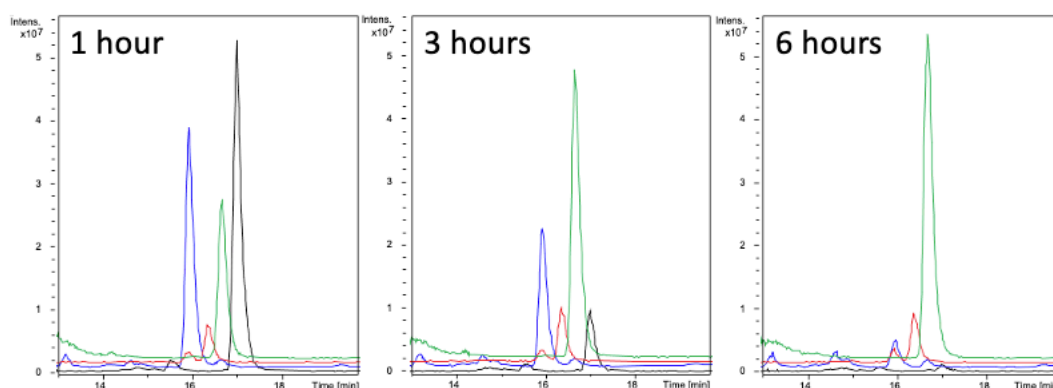
*Figure 2.5 Extracted ion chromatograms (EICs) from LCMS analysis of the reaction of phenylalanine derived α-dimethyl-β-keto methyl ester **56d** with EpnF and PLE. Black: EIC for m/z 407.22 corresponding to [M+H]+ for the substrate **56d**. Blue: EIC for m/z 393.20 corresponding to [M+H]+ for the product of ester cleavage **58d**. Red: EIC for m/z 363.19 corresponding to [M+H]+ for the epoxyketone product **59d**. Green: EIC for m/z 349.21 corresponding to [M+H]+ for the decarboxylated side product **60d**. The substrate was incubated for 1 hour (left), 3 hours (middle), and 6 hours.*

2.3.2.2 Temperature

It appears that PLE is hydrolysing the methyl ester substrate analogue faster than EpnF is catalysing the epoxyketone formation which leads to significant amounts of decarboxylated side product. To investigate if the activity of EpnF is significantly temperature dependant and if there is a temperature optimum, assays were set up at a range of different temperatures. Incubation temperatures of 4 °C, 15 °C, room temperature (approximately 20 °C), 25 °C, 30 °C, and 37 °C were tested. Incubation at 30 °C provided the highest production of epoxyketone product which suggests that EpnF works the most efficiently at this temperature. Additionally, the ester cleavage catalysed by PLE occurs slightly slower at 30 °C and is therefore optimised for the coupled enzyme assay at this temperature. At 37 °C the ester hydrolysis is too fast for EpnF to keep up and significant amounts of decarboxylated side product are produced, while at much lower temperatures PLE is too slow, resulting in a low concentration of hydrolysed EpnF substrate (Figure 2.6). All following assays were therefore incubated for 3 hours at 30 °C before quenching and LCMS analysis.
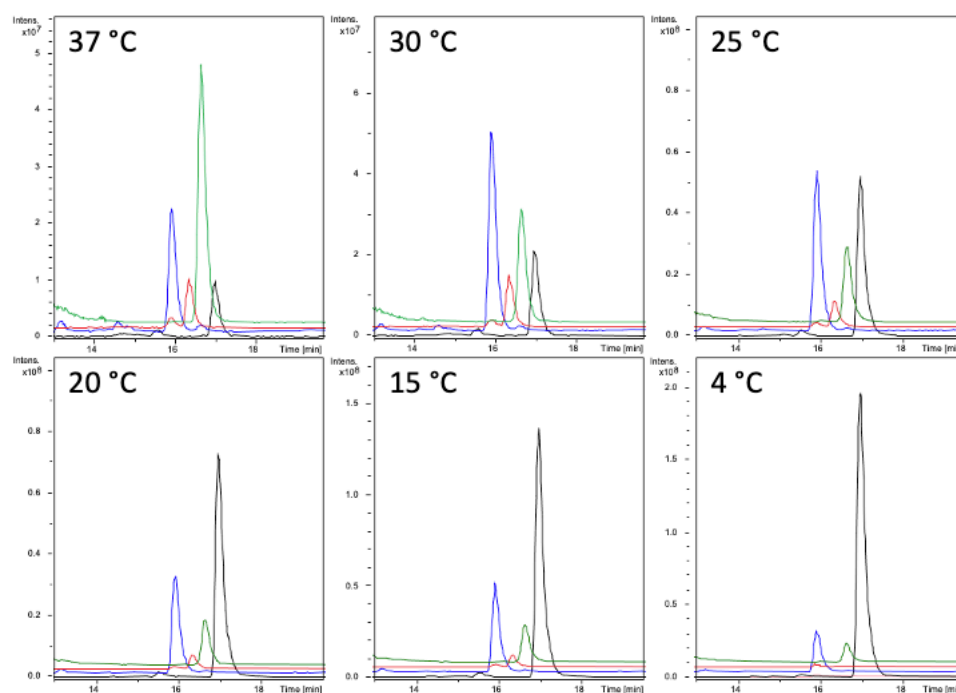
*Figure 2.6 Extracted ion chromatograms (EICs) from LCMS analysis of the reaction of phenylalanine derived α-dimethyl-β-keto methyl ester **56d** with EpnF and PLE. Black: EIC for m/z 407.22 corresponding to $[M+H]^+$ for the substrate **56d**. Blue: EIC for m/z 393.20 corresponding to $[M+H]^+$ for the product of ester cleavage **58d**. Red: EIC for m/z 363.19 corresponding to $[M+H]^+$ for the epoxyketone product **59d**. Green: EIC for m/z 349.21 corresponding to $[M+H]^+$ for the decarboxylated side product **60d**. The temperature at which the substrate was incubated are included with each set of assay traces.*

2.3.2.3 PLE and EpnF concentrations

To investigate if different concentrations of both enzymes could improve the epoxyketone production or optimise the PLE to EpnF ratio for more efficient enzyme usage, various concentrations of PLE and EpnF were tested. Subsequent experiments showed that reducing the EpnF concentration to 50 μM did not affect production of the epoxyketone product, however, using less than that reduced the amount of epoxyketone formed. A PLE concentration of 5 μM still provided sufficient ester hydrolysis, however, PLE concentrations below 5 μM resulted in less carboxylic acid substrate being formed which led to significantly less epoxyketone production (Figure 2.7). All following assays were therefore set up with a PLE to EpnF ratio of 1:10 using 50 μM EpnF and 5 μM PLE final assay concentrations.
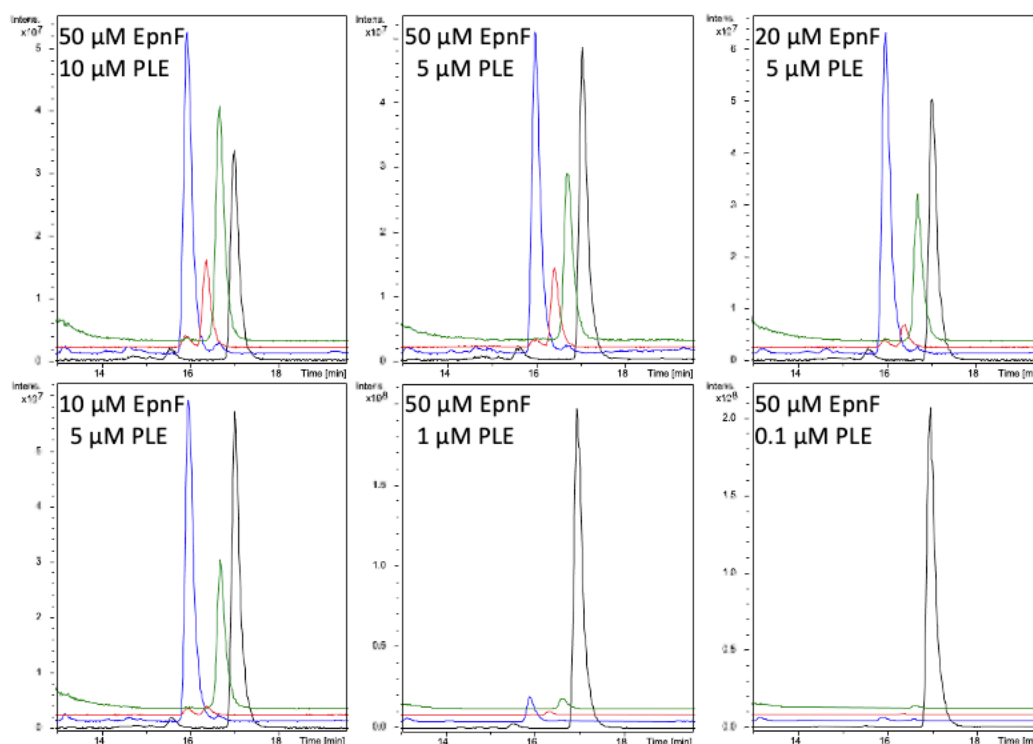
*Figure 2.7 Extracted ion chromatograms (EICs) from LCMS analysis of the reaction of phenylalanine derived α-dimethyl-β-keto methyl ester **56d** with EpnF and PLE. Black: EIC for m/z 407.22 corresponding to [M+H]$^+$ for the substrate **56d**. Blue: EIC for m/z 393.20 corresponding to [M+H]$^+$ for the product of ester cleavage **58d**. Red: EIC for m/z 363.19 corresponding to [M+H]$^+$ for the epoxyketone product **59d**. Green: EIC for m/z 349.21 corresponding to [M+H]$^+$ for the decarboxylated side product **60d**. The concentrations of EpnF and PLE are included with each set of assay traces.*

2.3.2.4 Control reactions

To ensure that any observed reactivity was a result of the expected assay component and not a consequence of other unknown factors, several control reactions were set up at the final optimised conditions and analysed by LCMS. Controls without EpnF in the presence of FAD showed no conversion to the desired epoxyketone confirming that FAD did not facilitate any epoxyketone formation without an epoxyketone synthase present. No reaction of the substrate was observed in a control with just FAD present to confirm that FAD was not reacting with the substrate in unpredicted ways. A control with all components present but no PLE showed no conversion of the substrate either to confirm that the PLE formed carboxylic acid was indeed the substrate for EpnF epoxyketone production (Figure 2.8).
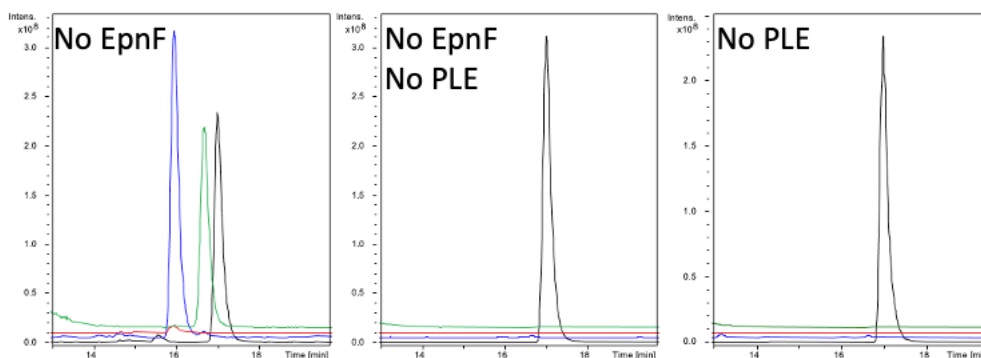
*Figure 2.8 Extracted ion chromatograms (EICs) from LCMS analysis of the reaction of phenylalanine derived α-dimethyl-β-keto methyl ester **56d** with EpnF and PLE. Black: EIC for m/z 407.22 corresponding to [M+H]⁺ for the substrate **56d**. Blue: EIC for m/z 393.20 corresponding to [M+H]⁺ for the product of ester cleavage **58d**. Red: EIC for m/z 363.19 corresponding to [M+H]⁺ for the epoxyketone product **59d**. Green: EIC for m/z 349.21 corresponding to [M+H]⁺ for the decarboxylated side product **60d**. Omitted components of the control reactions are included with each set of assay traces.*

2.3.3   MS/MS analysis

MS/MS analysis of extracts of the reaction using optimised assay conditions was used to analyse ion fragmentation of the expected substrate **56d**, carboxylic acid **58d**, epoxyketone **59d**, and side product **60d**. All four species showed the same fragmentation pattern, presumably resulting from loss of water and cleavage of the peptide bond between phenylalanine and serine (Figure 2.9). This further confirms the structures of the proposed products.
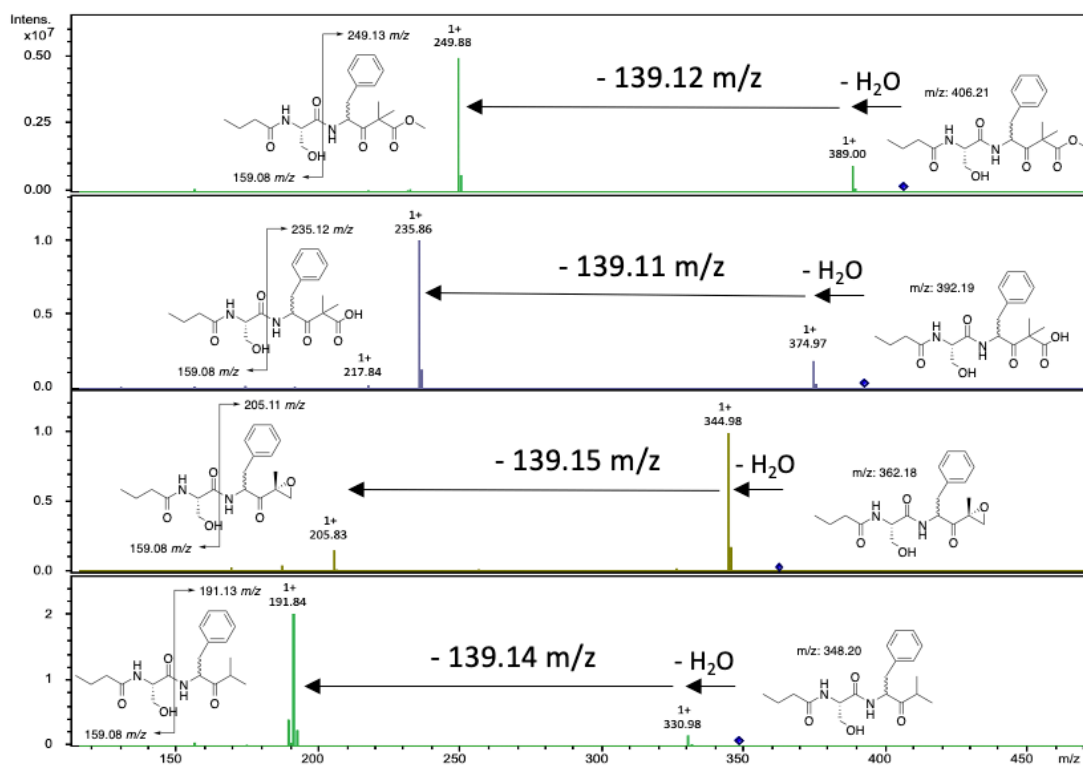
Figure 2.9 MS/MS analysis of expected compounds in coupled enzyme assays. For each species, the loss of one water molecule and a fragmentation between the serine and phenylalanine peptide bond was observed.

2.4 Coupled enzyme assays with substrate analogues

To further explore the substrate tolerance of EpnF, all the remaining synthetic substrate analogues were tested in coupled enzyme assays. The expected detectable compounds in the coupled PLE and EpnF assays are the α-dimethyl-β-keto methyl esters **56a-c,e**, the hydrolysed α-dimethyl-β-keto carboxylic acids **58a-c,e**, the potentially formed epoxyketones **59a-c,e**, and the decarboxylated side products **60a-c,e** (Scheme 2.10).
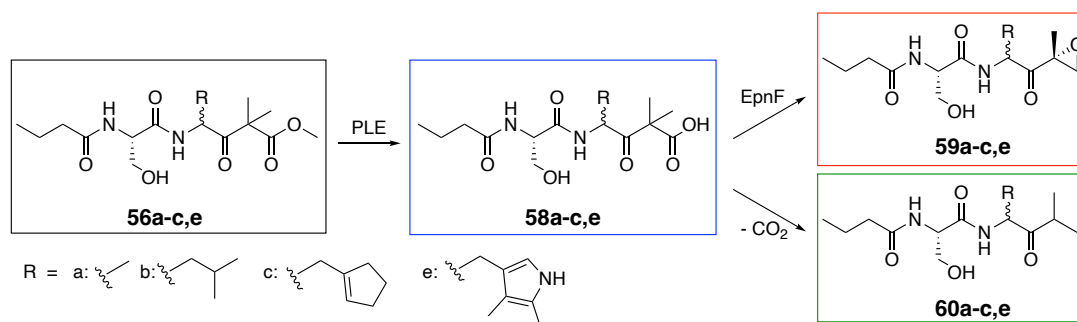
Scheme 2.10 Expected reactions in enzyme assays with pig liver esterase (PLE), EpnF, and the remaining α-dimethyl-β-keto methyl ester derived from alanine, leucine, cyclopentenylalanine, and tryptophan, 56a, 56b, 56c, and 56e respectively. PLE is expected to cleave the methyl ester to the corresponding α-dimethyl-β-keto carboxylic acids 58a-c,e which will be converted to epoxyketones 59a-c,e by EpnF or spontaneously decarboxylate. The coloured boxes refer to the corresponding [M+H]⁺ m/z extracted for in the EIC in the LCMS traces shown below.

Incubation of the alanine, leucine, cyclopentenylalanine, and tryptophan derived substrate analogue each with PLE confirmed that all the desired carboxylic acid substrates **58a-c,e** were generated *in situ* (Figure 2.10A). Subsequently each substrate was incubated with PLE and EpnF under the previously optimised assay conditions and analysed by LCMS. Epoxyketone production was observed for the alanine, leucine, and cyclopentenylalanine derived substrates while no epoxyketone was formed from the tryptophan derived α-dimethyl-β-keto carboxylic acid (Figure 2.10B).
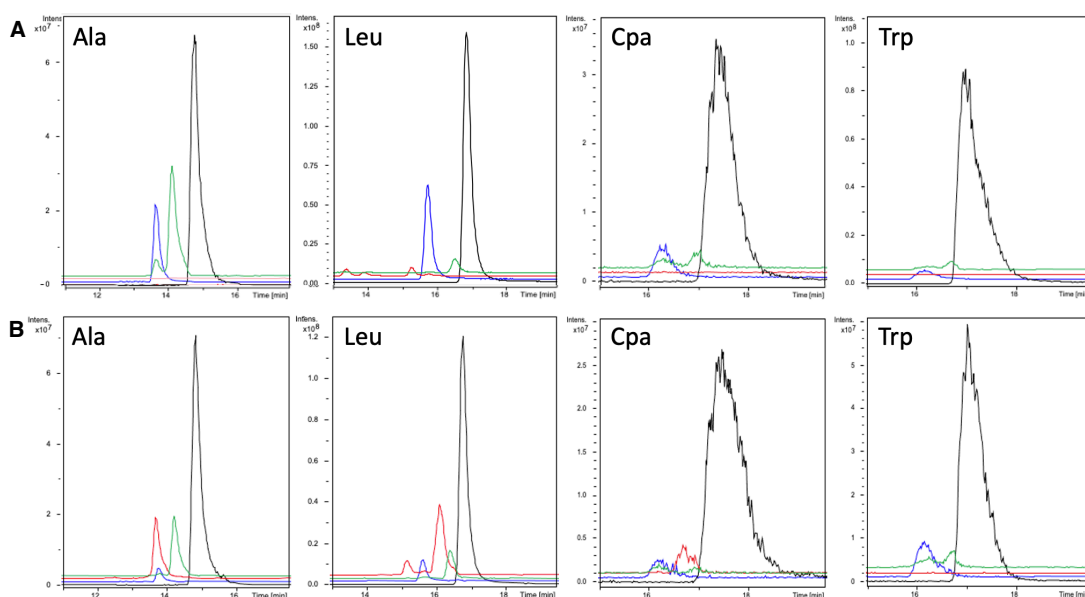
*Figure 2.10 Extracted ion chromatograms (EICs) from LCMS analysis of the in vitro enzyme assays with α-dimethyl-β-keto methyl ester analogues derived from alanine, leucine, cyclopentenylalanine, and tryptophan. A. Control reactions of each substrate incubated with PLE only. Black: EIC for m/z corresponding to $[M+H]^+$ for the substrates **56a-c,e**. Blue: EIC for m/z corresponding to $[M+H]^+$ for the product of ester cleavage **58a-c,e**. Red: EIC for m/z corresponding to $[M+H]^+$ for the epoxyketone product **59a-c,e**. Green: EIC for m/z corresponding to $[M+H]^+$ for the decarboxylated side product **60a-c,e**. The substrate of each assay is included with each set of LCMS traces B. Coupled enzyme assays of each substrate incubated with PLE and EpnF.*

Epoxyketone production for the alanine and leucine derived substrates was expected since EpnF previously tolerated these substrate analogues in preliminary studies by Dr Cartwright. Comparing the amounts of epoxyketone products produced it appears that EpnF transformed the native leucine substrate most efficiently under the tested conditions. The cyclopentenylalanine derived substrate was transformed into the desired epoxyketone by EpnF although the turnover appeared to be lower compared to the native substrate as well.

No epoxyketone production was observed when incubating tryptophan with PLE and EpnF (Figure 2.10B). It is possible that the tryptophan residue is sterically too demanding or that the large aromatic moiety prevents the substrate from being accepted by the active site otherwise. The observation that EpnF does not tolerate a tryptophan derived substrate analogue suggests that EpnF has a limited substrate tolerance and cannot catalyse epoxyketone formation for a broad range of substrates.

2.5 Conclusions

In summary, a synthetic strategy for preparing α-dimethyl-β-keto methyl esters was developed and provided alanine, leucine, cyclopentenylalanine, phenylalanine, tryptophan, side chain protected serine, side chain protected glutamic acid, and side chain protected lysine derived α-dimethyl-β-keto methyl ester analogues. While it was not possible to deprotect the later three substrates without cyclisation of the products, the alanine, leucine, cyclopentenylalanine, phenylalanine, and tryptophan derived analogues were fully characterised and available as substrates for enzyme assays.

Overproduction and purification procedures for the epoxyketone synthase EpnF were successfully reproduced and yielded the epoxyketone synthase in 20.5 mg/L. Protein purity and protein mass were confirmed by SDS-PAGE gel analysis and intact protein high-resolution mass spectrometry to ensure the correct protein was purified for the planned *in vitro* studies.

Coupled enzyme assays with pig liver esterase (PLE) and epoxyketone synthase EpnF were initially set up with the phenylalanine derived α-dimethyl-β-keto methyl ester due to its commercial relevance. EpnF activity was observed and the corresponding epoxyketone was formed. To optimise the coupled enzyme assay conditions, several variables have been changed independently. The duration of EpnF activity was found to be 3 hours and a temperature of 30 ℃ provided a suitable balance of slightly decreased PLE activity and highest observed turnover by EpnF. PLE and EpnF concentrations were optimised and various controls under the improved assay conditions confirmed that observed reactivity was a result of the expected assay component. Additionally, MS/MS analysis revealed fragmentation patterns of the different detected compounds in the coupled enzyme assays which were consistent with the expected structures of the proposed assay products.

Despite all efforts to improve assay conditions and increase epoxyketone production, the enzymatic turnover remained low. It is possible that EpnF is only tolerating the native *SS*-diastereomer which would decrease the amount of potentially formed epoxyketone to approximately half of the amount of substrate added. Additionally, it is unknown if PLE prefers one substrate diastereomer over the other which could lead to slower hydrolysis of the *SS*-diastereomer and decrease the amount of potentially

formed epoxyketone even further. This could be investigated in the future by separating the substrate diastereomers and probing them separately in PLE and coupled enzyme assays with PLE and EpnF. Alternatively, if diastereomeric separation on column chromatography with a suitable chiral stationary phase was possible the PLE and coupled enzyme assays with PLE and EpnF could be analysed by chiral LCMS to explore potential chiral resolution by either PLE, EpnF, or both.

The substrate tolerance of EpnF was tested under the optimised assay conditions with the alanine, leucine, cyclopentenylalanine, and tryptophan derived α-dimethyl-β-keto methyl esters analogues. Incubation of each of the substrates with PLE showed hydrolysis of the methyl ester to the corresponding carboxylic acid confirming that PLE is a suitable esterase to generate the desired EpnF α-dimethyl-β-keto carboxylic acid substrates *in situ*. Coupled enzyme assays with EpnF revealed epoxyketone formation of the alanine, leucine, and cyclopentenylalanine substrates, however no tryptophan derived epoxyketone was formed. While EpnF appears to be the most efficient for the native leucine derived substrate, it does show activity towards non-native substrates that contain residues of smaller or similar steric hindrance compared to the native substrate such as alanine, leucine, cyclopentenylalanine, and phenylalanine. Because EpnF is catalysing epoxyketone formation for the phenylalanine derived substrate, it appears that it can tolerate aromatic substrates to some extent. Therefore, it is hypothesised that the tryptophan derived substrate is not tolerated due to the increased steric hinderance of its larger aromatic side chain.

EpnF exhibits *in vitro* activity and substrate tolerance towards some non-native substrates which is a beneficial starting point for the development of a biocatalyst. The ability to catalyse epoxyketone formation with a phenylalanine and cyclopentenylalanine residue next to the site of reaction is useful insight for the potential development of a chemoenzymatic production of proteasome inhibitor drugs oprozomib and KZR-616 respectively. The developed coupled enzyme assays with the optimised conditions provide a method to test substrates with EpnF for enzymatic epoxyketone formation and further substrate analogues, including commercially relevant substrates, could be tested.

Although the conditions for the coupled enzyme assays were improved and various aspects optimised, the overall turnover to the desired epoxyketone products is still far

from ideal. EpnF appears to degrade within 3 hours which suggests that the epoxyketone synthase is very unstable and therefore prevented from being efficient under *in vitro* conditions. Additionally, the catalytic ability seems to be limited and the reaction rate very low since only a slow build-up of epoxyketone products is observed. EpnF also seems to have limited substrate tolerance which prevents it from producing a broad range of epoxyketones and would require severe modifications to enhance its activity to a suitable level for broader biocatalytic applications. All these observations reveal that EpnF is accompanied by several challenging aspects and makes it desirable to additionally investigate alternative epoxyketone synthases.

3. Towards the chemoenzymatic synthesis of commercially relevant epoxyketones

3.1 Synthesis of the oprozomib precursor

The first part of this project showed that EpnF can form epoxyketones from α-dimethyl-β-keto methyl esters in *in vitro* coupled enzyme assays. To investigate the potential of EpnF to be used in the chemoenzymatic production of commercially relevant epoxyketones, the α-dimethyl-β-keto methyl ester substrate for oprozomib **61** was synthesised. The synthetic strategy was based on the published procedure for oprozomib by Zhou *et al.*, which involves coupling the *N*-terminal thiazole **62** to fragment **63** previously assembled from a di-*O*-methylated serine fragment **64** and the epoxyketone warhead which will be replaced with the α-dimethyl-β-keto methyl ester fragment **55d** (Scheme 3.1).[70] However, since the published coupling reactions were set up with HOBt it was decided to use the previously tested NMI-TCFH coupling conditions instead (Chapter 2.1.4.2).
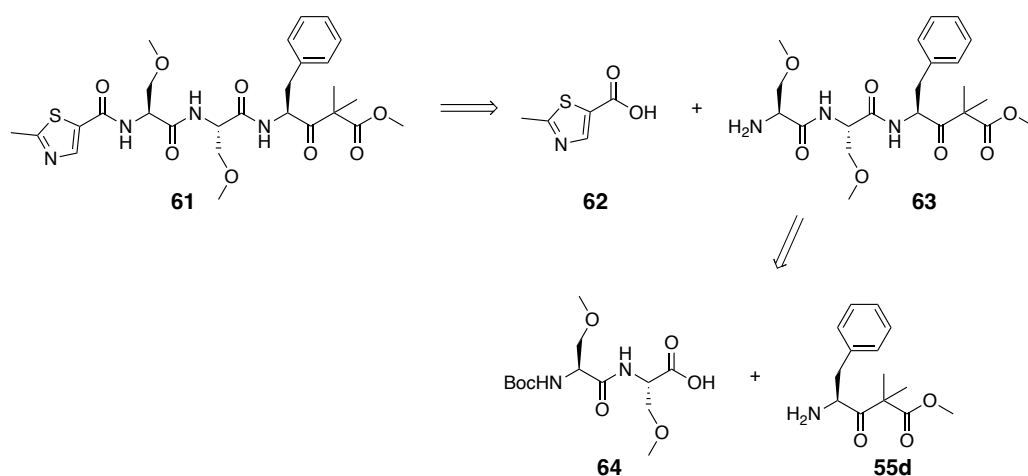
*Scheme 3.1 Disconnection of the oprozomib methyl ester precursor **61** analogue to the oprozomib synthetic strategy by Zhou et al.[70] The N-terminal thiazole **62** is attached in the final step to the fragment **63** that is previously assembled from di-O-methylated serine **64** and phenylalanine α-dimethyl-β-keto methyl ester **55d**.*

N-Boc-*O*-methylated serine (**65**) was purchased, and the acid functional group was protected with a benzyl group to yield **66** (Scheme 3.2). Deprotection of the *N*-Boc group under acidic conditions provided H$_2$N-O-Me-Ser-OBn **67**. Coupling both *O*-

methylated serine fragments under the previously tested NMI-TCFH coupling conditions (Chapter 2.1.4.2) and deprotection of the benzyl group by hydrogenation yielded the di-*O*-methylated serine fragment **64**. The same NMI-TCFH coupling conditions were used to couple **64** with the phenylalanine α-dimethyl-β-keto methyl ester fragment **55d** prepared as described in section 2.1.3. The *N*-Boc protecting group was subsequently removed in acidic conditions to provide **63** as a diastereomeric mixture. The oprozomib precursor was generated by coupling thiazole **62** to fragment **63** with NMI-TCFH. Although all the previous coupling reactions had consistent yields of 47% and 65%, this final coupling step unfortunately provided only 10% yield of the oprozomib precursor **61**. It is possible that the NMI-TCFH coupling conditions are not ideal for thiazoles compared to amino acids. However, since sufficient amount of the oprozomib precursor was produced for full spectroscopic characterisation and enzyme assays (12 mg), it was decided to not further optimise this synthetic route.
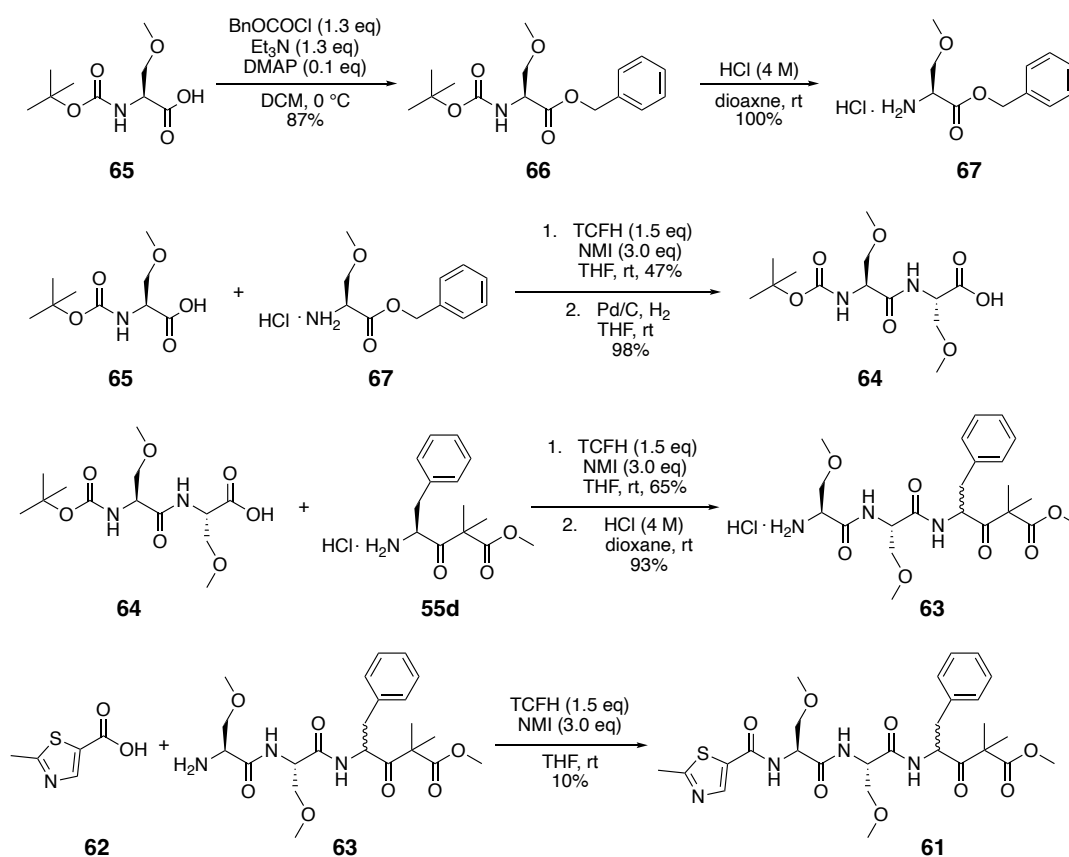
Scheme 3.2 Synthetic route for the α-dimethyl-β-keto methyl ester oprozomib precursor **61**. Reaction conditions and yields are shown for each step.

3.2 Coupled enzyme assays with PLE, EpnF, and oprozomib precursor

The oprozomib precursor was assayed with PLE to investigate if the esterase would cleave the methyl ester, and with PLE and epoxyketone synthase EpnF to reveal if EpnF was able to produce the commercially relevant epoxyketone oprozomib. Enzyme assays were set up under the previously optimised assay conditions (Chapter 2.3.2) and analysed by LCMS by extracting for the m/z of expected compounds (Figure 3.1A).

Incubating the oprozomib precursor **61** with PLE for 3 hours resulted in ester hydrolysis to the α-dimethyl-β-keto acid **68** as well as formation of the decarboxylated side product **69**. However, coupled enzyme assays in which the oprozomib precursor was incubated with PLE and EpnF for 3 hours did not result in epoxyketone formation suggesting oprozomib precursor **61** is not a substrate for EpnF (Figure 3.1B).
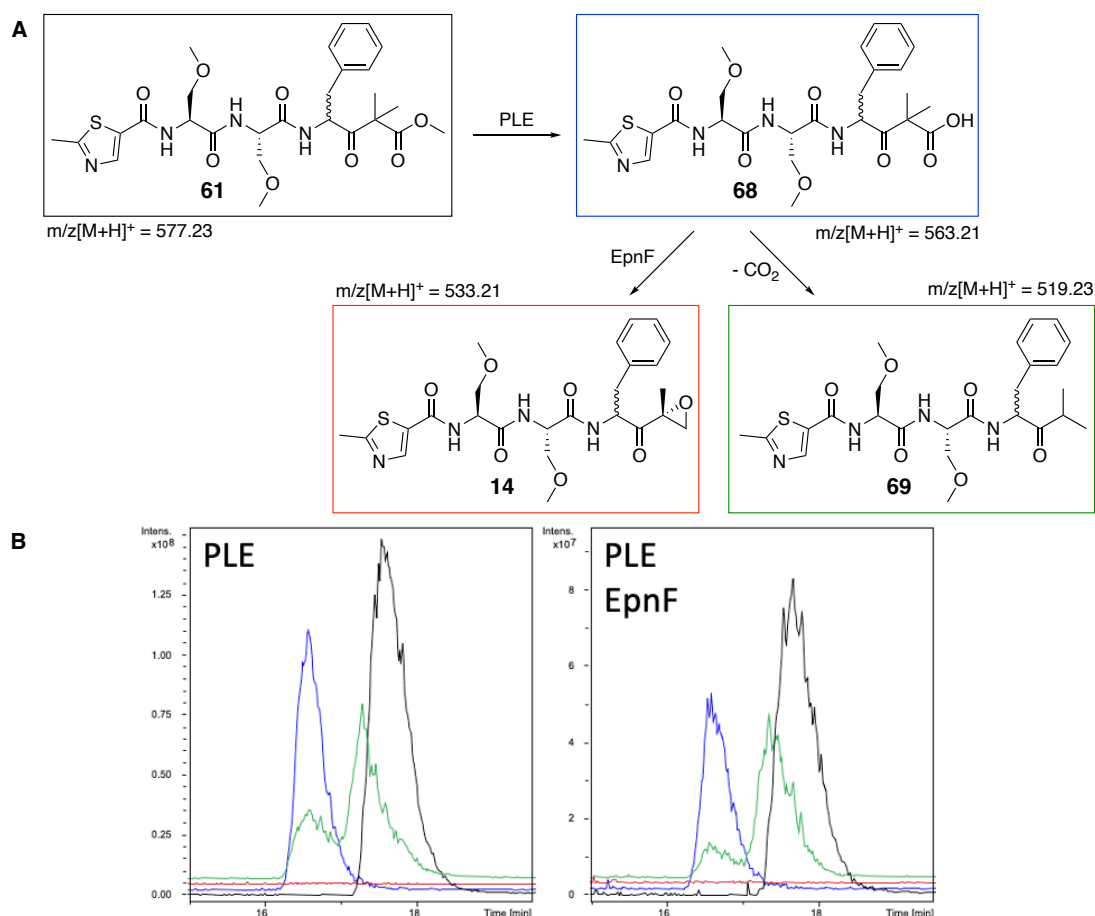
*Figure 3.1 Coupled enzyme assays with synthetic oprozomib precursor **61**, PLE, and EpnF. A. Expected reactions in coupled enzyme assays. PLE is expected to cleave the methyl ester **61** to the corresponding α-dimethyl-β-keto carboxylic acid **68** which could be converted to the epoxyketone oprozomib **14** by EpnF or spontaneously decarboxylate to side product **69**. m/z of the [M + H]⁺ ion of each species is*

Since EpnF is able to convert the phenylalanine derived substrate **58d**, either the absence of the butyric acid serine fragment or the presence of the thiazole di-*O*-methylated serine is responsible for EpnF not tolerating the oprozomib precursor (Figure 3.2). One possibility is that the serine group of **58d** engages in crucial hydrogen bonding interactions for the substrate to dock into the active site which would not be provided with *O*-methylated serine residues in the oprozomib substrate. It could also be that the larger oprozomib precursor is sterically too hindered to be accepted by EpnF or that the butyric acid moiety is otherwise involved in crucial interactions with the active site.

phenylalanine derived substrate **58d**   oprozomib substrate **68**
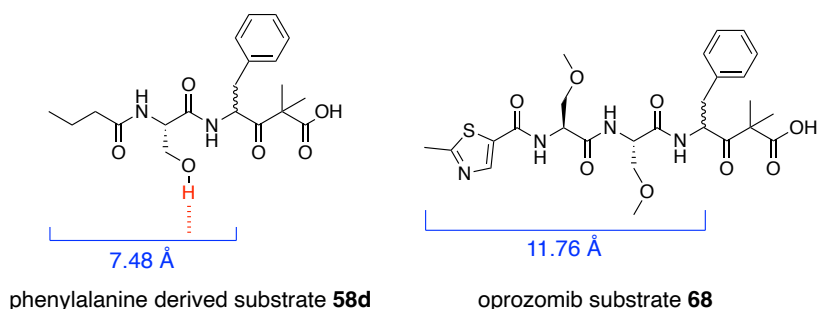
*Figure 3.2 Comparison of EpnF accepted phenylalanine derived substrate **58d** and oprozomib precursor **68** which is not converted by EpnF under the tested conditions. The additional hydrogen bond donor in **58d** is highlighted in red and the difference in steric hindrance is visualised by the distance from the terminal methyl group carbon to the phenylalanine α-carbon in minimised energy configuration of the molecules.*

3.3 *N*-Boc phenylalanine α-dimethyl-β-keto methyl ester assays

To investigate potential reasons for the inactivity of EpnF toward the oprozomib precursor, *N*-Boc protected phenylalanine α-dimethyl-β-keto methyl ester **55d** was tested in *in vitro* enzyme assays (Figure 3.3). Since **70** is sterically similarly hindered to the phenylalanine derived substrate **58d** it is expected to be converted to the corresponding epoxyketone if EpnF's inactivity towards the oprozomib precursor is due to steric restrictions of the active site. However, if the properties of the serine group next to phenylalanine in **58d** are crucial for interactions with EpnF, it is expected that an *N*-Boc protected phenylalanine α-dimethyl-β-keto acid will not be converted into an epoxyketone since it is lacking the crucial hydroxy group.

phenylalanine derived substrate **58d**   phenylalanine derived fragment **70**
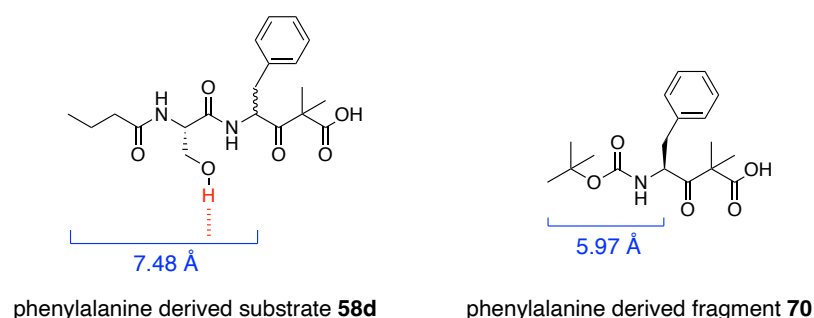
*Figure 3.3 Comparison of EpnF accepted phenylalanine derived substrate **58d** and the N-Boc protected phenylalanine α-dimethyl-β-keto carboxylic acid **70**. The additional hydrogen bond donor in **58d** is highlighted in red and the difference in steric hindrance is visualised by the distance from the terminal methyl group carbon to the phenylalanine α-carbon in minimised energy configuration of the molecules.*

N-Boc protected phenylalanine α-dimethyl-β-keto methyl ester **58d** was tested in *in vitro* enzyme assays with PLE to confirm the methyl ester cleavage and in coupled enzyme assays with PLE and EpnF to investigate EpnF's activity towards this substrate (Figure 3.4A). While PLE seemed to cleave the methyl ester **58d** and to generate the desired carboxylic acid substrate **70**, coupled assays with PLE and EpnF did not result in epoxyketone formation suggesting *N*-Boc protected phenylalanine fragment **70** is not a substrate for EpnF either (Figure 3.4B). This is an indicator that the hydroxy group of the serine engages in crucial interactions with the active site of EpnF or could mean that the active site is otherwise very specific towards the serine butyric acid residue next to the phenylalanine α-dimethyl-β-keto carboxylic acid. It is possible that the additional amide bond or linear butyl chain in **58d** are crucial for

substrate binding or that the specific steric demand of the *tert*-butyl group in **70** hinders this substrate from properly interacting with the active site.
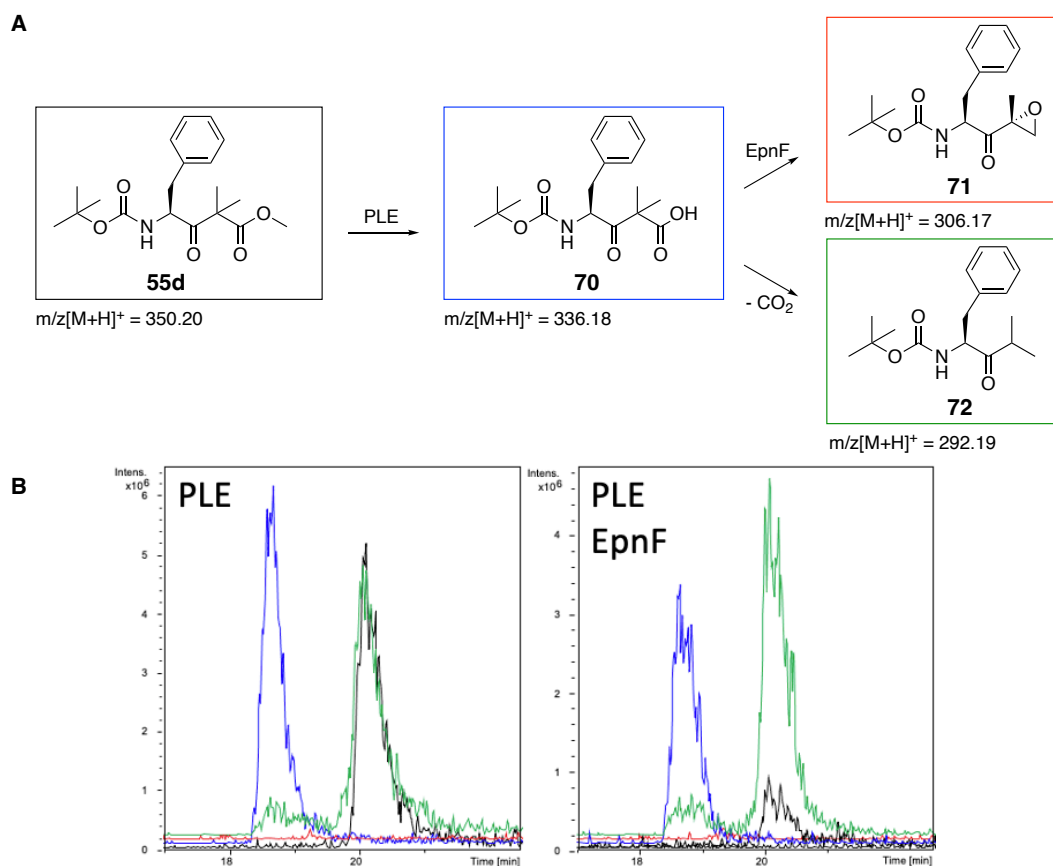
*Figure 3.4 Coupled enzyme assays with synthetic N-Boc protected phenylalanine α-dimethyl-β-keto methyl ester **55d**, PLE, and EpnF. A. Expected reactions in coupled enzyme assays. PLE is expected to cleave the methyl ester **55d** to the corresponding α-dimethyl-β-keto carboxylic acid **70** which could be converted to the shown epoxyketone **71** by EpnF or spontaneously decarboxylate to side product **72**. m/z of the [M + H]⁺ ion of each species is shown. B. Extracted ion chromatograms (EICs) from LCMS analysis of the reaction of N-Boc protected phenylalanine α-dimethyl-β-keto methyl ester **55d** with PLE only and PLE, EpnF, and FAD. Black: EIC for m/z 350.20 corresponding to [M + H]⁺ for the substrate **55d**. Blue: EIC for m/z 336.18 corresponding to [M + H]⁺ for the product of ester cleavage **70**. Red: EIC for m/z 306.17 corresponding to [M + H]⁺ for the epoxyketone product **71**. Green: EIC for m/z 292.19 corresponding to [M + H]⁺ for the decarboxylated side product **72**.*

3.4 Conclusions

In summary, a successful synthetic strategy for an α-dimethyl-β-keto methyl ester oprozomib derivative was developed to provide 12 mg of the chemically stable oprozomib precursor for *in vitro* enzyme assays.

Experiments with pig liver esterase (PLE) showed hydrolysis of the methyl ester to confirm that PLE is a suitable esterase to generate the desired oprozomib α-dimethyl-β-keto acid substrate *in situ*. Experiments with EpnF did not show any conversion to an epoxyketone under the tested conditions which suggests that the oprozomib precursor is not tolerated by EpnF.

Testing EpnF with *N*-Boc protected phenylalanine α-dimethyl-β-keto methyl ester showed that this substrate is also not converted into the desired epoxyketone in coupled enzyme assays. These findings suggest that the hydroxy group of the serine is crucial for interactions with the active site of EpnF, or that the butyric acid serine residue is otherwise essential for interactions with the active site.

If the serine residue is crucial for the interactions with the active site, it could be possible to synthesise the oprozomib precursor with serine instead of *O*-methylated serine and incubate this alternative substrate with PLE and EpnF. It is possible that there would still be no turnover to the epoxyketone due to the steric hinderance. However, if EpnF was able to catalyse the epoxyketone formation for a serine derived oprozomib precursor the serine group of the epoxyketone could subsequently be methylated to form the proteasome inhibitor oprozomib. Although there are currently no studies available for late-stage *O*-methylations in the oprozomib synthesis, investigating this would still be an option for a chemoenzymatic approach to oprozomib.

Overall, EpnF converts some non-native substrates to the predicted epoxyketones, however, it did not exhibit the desired *in vitro* catalytic activity. Substrates that are sterically more hindered and deviate significantly from the native substrate are not accepted by EpnF which indicates that EpnF as it is, is not suitable for chemoenzymatic production of commercially relevant epoxyketones.

Known epoxyketone synthases could function as lead enzymes for enzyme engineering experiments to create an epoxyketone synthase with optimised stability and activity towards specific commercially relevant substrates. To find a suitable epoxyketone synthase as a starting point for enzyme engineering experiments and for future studies towards chemoenzymatic epoxyketone production in general, it would be highly relevant to gain more insight into various epoxyketone synthases. Most of the known epoxyketone biosynthetic gene clusters like the *epn* cluster incorporate a leucine related residue next to the epoxyketone pharmacophore. However, novel epoxyketone proteasome inhibitors like oprozomib and KZR-616 contain aromatic or artificial amino acids in this position which is why it would be important to identify and analyse novel epoxyketone synthases with specificities towards different residues.

4. Identification and analysis of novel epoxyketone biosynthetic gene clusters

4.1 *In silico* identification of potential epoxyketone synthases

EpnF showed reasonable substrate tolerance towards substrates similar to the native substrate, however, it could not accept the oprozomib precursor. Furthermore, its lack of stability and low turnover is suboptimal. We therefore wanted to find an epoxyketone synthase that naturally acts on α-methyl-β-keto carboxylic acid substrates with aromatic amino acids next to the site of reaction. Such an epoxyketone synthase would be expected to be found in the biosynthesis of tryptopeptin A (**22**), first isolated from *Streptomyces* sp. KUSC-G1123, which contains a tryptophan residue next to the epoxyketone warhead (Figure 4.1).[77]
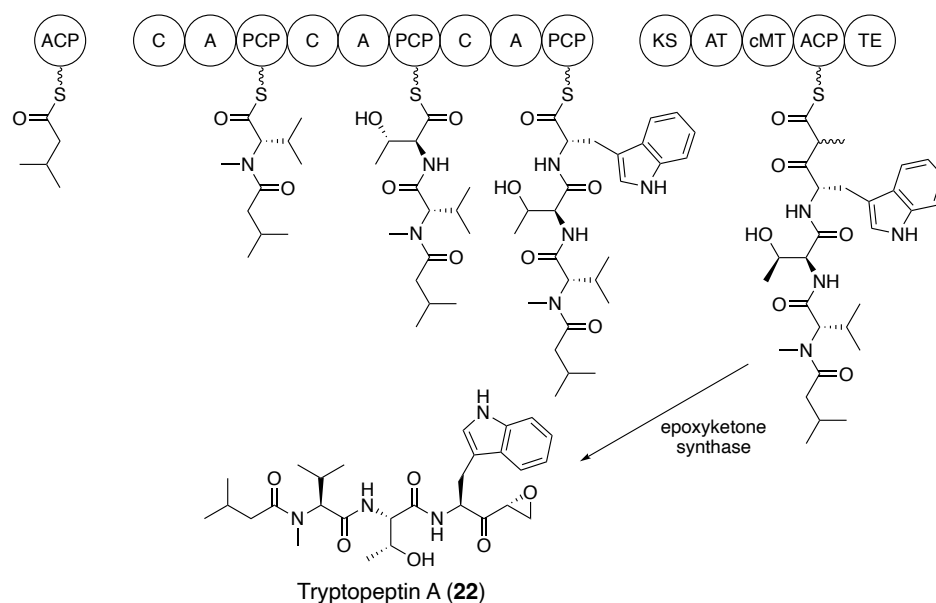
Figure 4.1 Structure and proposed pathway of the biosynthesis of the aromatic epoxyketone natural product tryptopeptin A (22).

The biosynthesis of tryptopeptin A is expected to be similar to the assembly of eponemycin, involving an acyl carrier protein, an NRPS assembling *N*-methylated valine, threonine, and tryptophan, a type I PKS, and an epoxyketone synthase to form the epoxyketone warhead (Figure 4.1). The potential biosynthetic gene cluster for

tryptopeptin A would therefore be expected to resemble the eponemycin biosynthetic gene cluster.

4.1.1 Identification of potential epoxyketone biosynthetic gene clusters with clusterTools

ClusterTools is a bioinformatic tool that allows for multiple criteria searches to identify putative biosynthetic gene clusters in genomic sequence databases.[121] Multiple requirements can be submitted in a search such as sequences for homologues that have to be in proximity to known cluster elements like specific NRPS or PKS modules. The database is then searched for all submitted requirements and provides an output of clusters that are predicted to match the requests. For the described experiment, a redundant bacterial database with approximately 11500 genomes created in 2016 by clusterTool's author Dr Emzo de los Santos from the "Prokaryotic RefSeq Representative Genomes" repository from the National Center for Biotechnology Information (NCBI) was used.

To identify potential tryptopeptin A biosynthetic gene clusters, a search for clusters with an EpnF homologue, a ketosynthase (KS) domain, an adenylation (A) domain, and a condensation (C) domain was conducted. This search identified several clusters containing one *epnF* homologue, an NRPS, and a PKS encoding gene, similar to known epoxyketone biosynthetic gene clusters (Figure 4.2). Because the first hit only contained a very short NRPS it was rejected and the second, third, and forth hits highlighted in Figure 4.2 were further analysed.
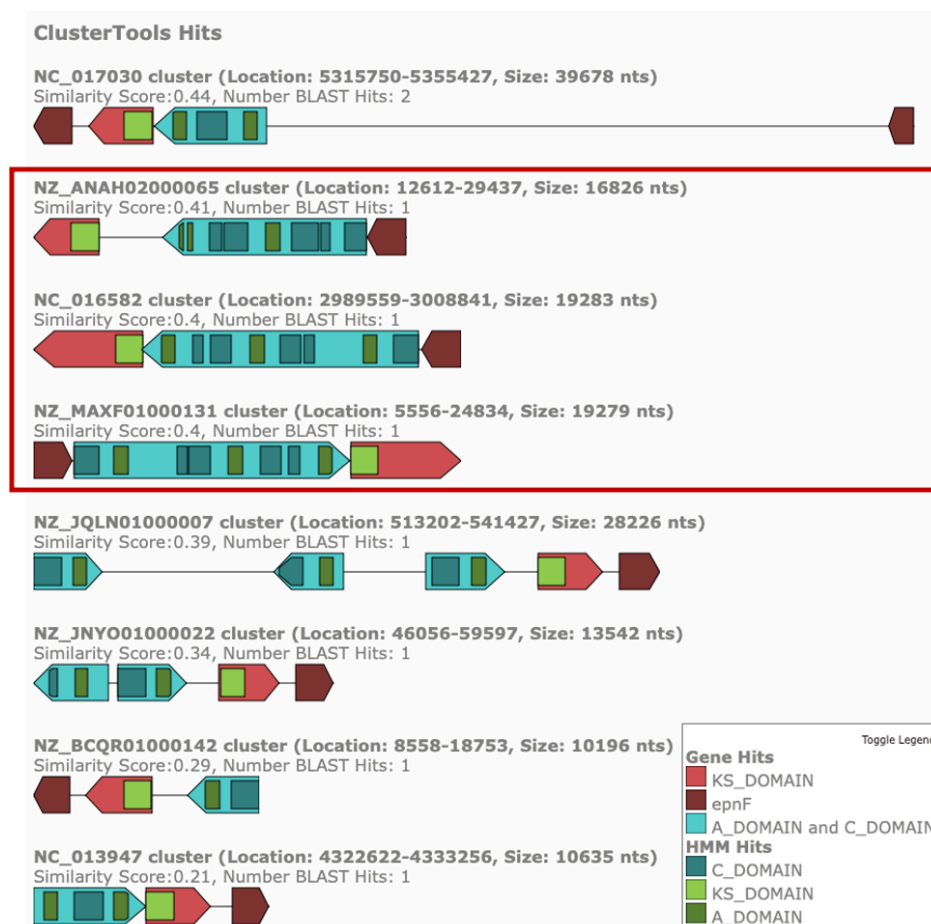
Figure 4.2 ClusterTools output for the search for epnF homologues, KS domains, A domains, and C domains. The three clusters of interest from Cystobacter fuscus (NZ_ANAH02000065), Streptomyces bingchenggensis (NC_016582), and Streptomyces sparsogenes (NZ_MAXF01000007) are highlighted.

4.1.2 Cluster comparison

The contig of *Streptomyces hygroscopicus* with the *epn* cluster and the corresponding contigs of *Streptomyces sparsogenes* ATCC 25498, *Streptomyces bingchenggensis* BCW-1, and *Cystobacter fuscus* DSM 2262 of the three clusters of interest were submitted to antiSMASH to allow for cluster comparison and to get more detailed insight into the organisation of the biosynthetic gene clusters.[122]

The antiSMASH KnownClusterBlast output revealed 16-25% gene similarity between all the *Streptomyces* clusters of interest with landepoxcin **23** and epoxomicin **13** clusters. The *C. fuscus* cluster of interest showed 10% gene similarity to the clarepoxcin (**24**) cluster (all mentioned clusters shown in Chapter 1.3.5 Figure 1.16).

The eponemycin biosynthetic gene cluster and the three clusters of interest were aligned with clinker to reveal similarities and differences between the clusters (Figure

4.3).[123] The clusters from *S. sparsogenes* and *S. bingchenggensis* were closely aligned, appeared to be highly similar in the core functional protein expressing genes, and exhibit 79% identity to each other when comparing with Basic Local Alignment Search Tool (BLAST). The *C. fuscus* cluster however does not show much similarity with any of the other clusters besides the *epnF* homologue and the NRPS encoding gene.
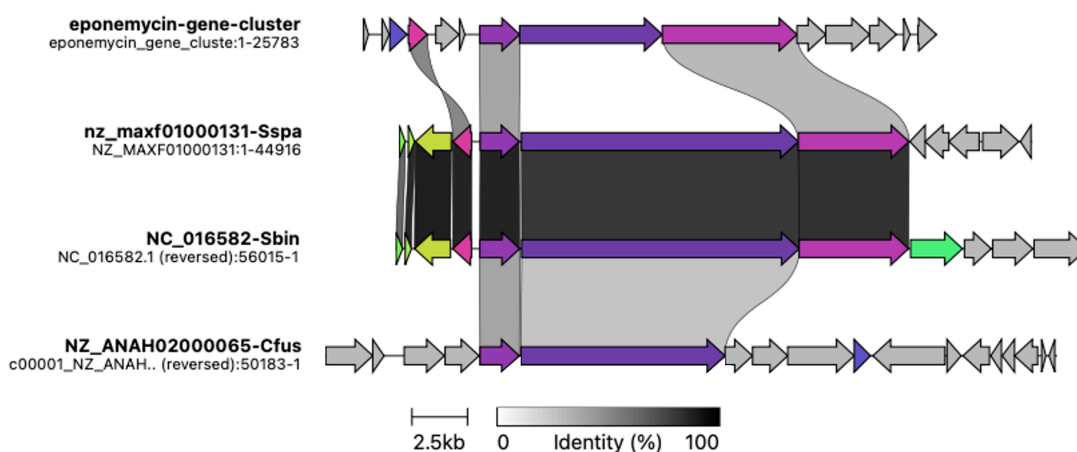
Figure 4.3 Cluster alignments of the eponemycin biosynthetic gene cluster and the potential epoxyketone cluster of interest. Clusters were aligned to epnF and its homologues with clinker.[123]

4.1.3 A domain specificity prediction

Predicting the specificities of A domains in an NRPS can give further insight into what metabolite might be directed by a biosynthetic gene cluster. A domains contain highly conserved amino acid residues in their active sites which determine their specificity.[112] Increasing amounts of available data for A domain sequences and confirmed specificities continuously improve the prediction of the relevant residues in the specificity pockets of A domains which leads to more accurate predictions about which amino acid is activated and incorporated into the NRPS assembled backbone. Analysis of these residues and corresponding specificities can provide the predicted amino acid sequence built by the NRPS.[124]

The open access webtool developed by Bachmann *et al.* was used to predict the A domain specificity of the NRPSs encoded by all three clusters of interest (Table 4.1).[124] The NRPS encoded by the *C. fuscus* cluster harbours four A domains. The specificity of the first A domain could not be predicted, the second, third, and fourth

70

A domains were predicted to encode a threonine, a leucine, and another threonine respectively. These amino acid recognition predictions suggest the *C. fuscus* cluster is very unlikely to direct the biosynthesis of tryptopeptin A.

Table 4.1 Outcome of A domain specificity predictions.[124] Listed are amino acids of assigned function in the substrate specificity pockets of the adenylation domains in the NRPS encoded by three clusters of interest and their predicted substrate specificity.

	A domain 1	A domain 2	A domain 3	A domain 4
C. fuscus	No hit	Thr	Leu	Thr
Residues active site	DSAXIAEV	DFWNIGMV	DAWFFGAV	DFWNIGTV
S. bingchenggensis	Val	Thr	No hit	
Residues active site	DAYWWGGT	DFWNIGMV	DASVALCV	
S. sparsogenes	Val	Thr	No hit	
Residues active site	DAYWWGGT	DFWNIGMV	DASVALCV	

The *S. sparsogenes* and *S. bingchenggensis* clusters encode NRPSs containing three adenylation domains with identical predicted amino acid residues in the substrate specificity pockets. Comparing the two NRPSs with BLAST revealed that they are 97% identical to each other suggesting that they will exhibit the same substrate preferences. The first A domain has a predicted specificity for valine, the second for threonine, which is what would be expected for tryptopeptin A. Although no clear prediction could be made for the third A domain, the amino acid residues in the substrate specificity pocket are quite small and the majority is hydrophobic, which is consistent with binding a large hydrophobic amino acid such as tryptophan (Figure 4.4).
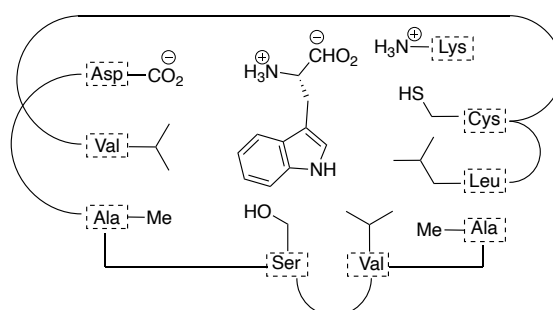
Figure 4.4 Tryptophan mapped into the substrate specificity pocket of the third A domain of the NRPS encoded by the Streptomyces clusters of interest.

These findings suggest that it is possible that the identified clusters in *S. sparsogenes* and *S. bingchenggensis* direct the biosynthesis of tryptopeptin A, however, only the *S. sparsogenes* ATCC 25498 strain was commercially available which is why this cluster was chosen for more detailed analysis and following *in vivo* experiments.

4.1.4 Cluster analysis with antiSMASH and BLAST

To completely characterise the *S. sparsogenes* cluster of interest Basic Local Alignment Search Tool (BLAST) was used to predict the function of each of the genes in the cluster (as defined by the antiSMASH result from the published Illumina data). The predicted function of the EpnF homologue was an acyl-CoA dehydrogenase (ACAD) which is consistent with the predictions for epoxyketone synthases in known epoxyketone biosynthetic gene clusters. As well as the EpnF homologue, NRPS, and PKS the biosynthetic gene cluster contained a proteasome β-subunit which probably is relevant for the species resistance towards the produced epoxyketone. Additionally, the cluster includes an oxidoreductase homologue, predicted dehydrogenases, and an aminotransferase downstream of the NRPS/PKS (Figure 4.5). The PKS prediction by antiSMASH included a KS domain, an AT domain, and a cMT domain which suggests the incorporation of a methylated malonyl after the NRPS assembled backbone which is in agreement with the structure of tryptopeptin A (**22**).
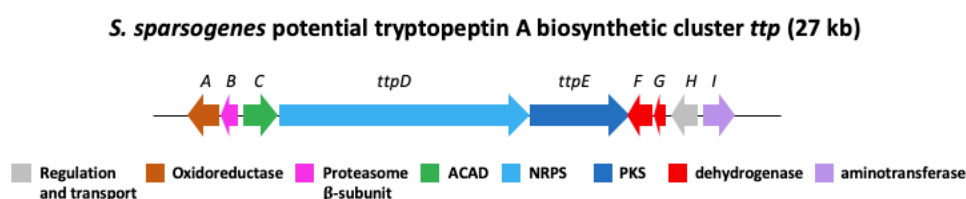
Figure 4.5 Potential tryptopeptin A producing cluster ttp from S. sparsogenes containing the epoxyketone synthase homologue encoding gene ttpC, a predicted oxidoreductase, proteasome β-subunit, NRPS, PKS, dehydrogenases, and an aminotransferase analogue alongside regulatory genes.

To further investigate the identified biosynthetic gene cluster and find out if it really produces tryptopeptin A the relevant *S. sparsogenes* strain was ordered from the German Collection of Microorganisms and Cell Cultures and cultured for metabolite expression, extraction, and analysis.

4.2 Analysis of metabolite profile of *Streptomyces sparsogenes* and *S. coelicolor* containing the *ttp* cluster

Streptomyces sparsogenes ATCC 25498 was ordered from the German Collection of Microorganisms and Cell Cultures and the freeze-dried cell pellet was successfully reactivated on different media (Figure 4.6). The strain grew in white colonies that formed grey spores after 10 days and spore stocks were prepared.
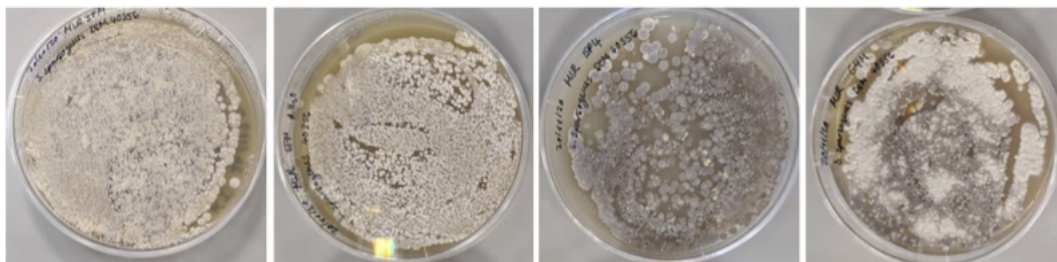
Figure 4.6 Streptomyces sparsogenes growing on SFM prepared with tap water, SFM prepared with distilled water, ISP4, and GYM (left to right).

A strong confirmation that the identified biosynthetic gene cluster is able to produce tryptopeptin A would be the observation of tryptopeptin A production in *S. sparsogenes*. The strain was grown in liquid cultures and on plates for metabolite expression and extraction to analyse the produced metabolites by HRMS.

4.2.1 Choice of growth medium

Correspondence about growth conditions for tryptopeptin A producers was established with the first author of the first tryptopeptin A isolation publication, Dr Yuta Tsunematsu (Assistant Professor University of Shizuoka, Japan).[77] Dr Tsunematsu was not able to provide information about growth conditions for the published strain due to industrial collaboration on the previous tryptopeptin A publication. However, he was able to provide information on an alternative strain, *Streptomyces* species maeda85, that he found to produce tryptopeptin A as well. He observed tryptopeptin A production by growing *S.* sp. maeda85 on medium containing malt extract, yeast extract, and glucose (MYG medium) which was therefore the medium of choice for metabolite expression and extraction experiments with the *S. sparsogenes* strain of interest.

4.2.2 *Streptomyces sparsogenes* ATCC 25498 metabolites

Cultures of *S. sparsogenes* were grown in liquid MYG medium and on MYG agar plates in triplicate and incubated each for 5 and 10 days. Metabolites were extracted with ethyl acetate and analysed by HRMS. There were clear differences between the media control and the culture extracts which suggests that *S. sparsogenes* is producing secondary metabolites (Figure 4.7).
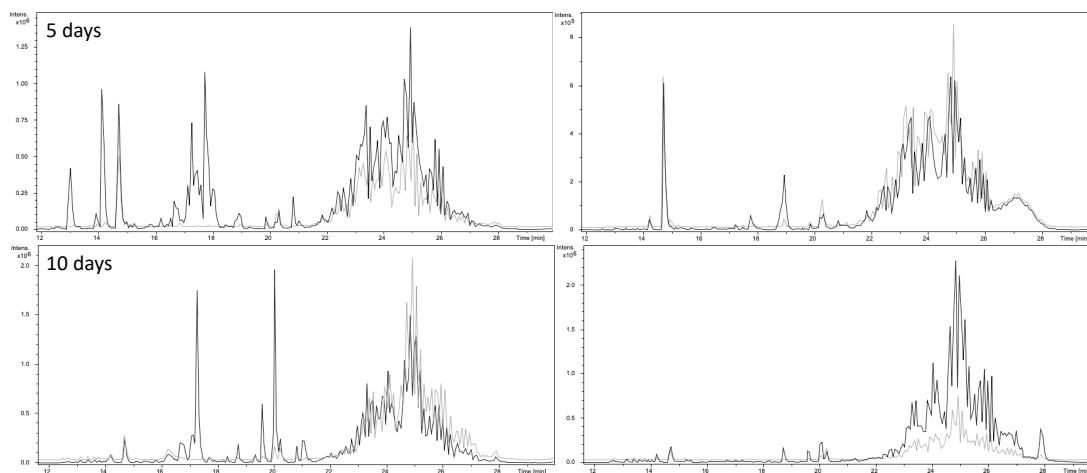
Figure 4.7 Base peak chromatograms of high-resolution mass spectrometry analysis of extracts of S. sparsogenes grown in MYG (black) and medium control (grey) after 5 days (top) and 10 days (bottom). Left: Extracts from MYG liquid media. Right: Extracts from MYG agar plates.

When extracting for expected ion masses of tryptopeptin A, no correlating mass was found in the traces. The major metabolite peaks from the *S. sparsogenes* extracts were analysed for their masses and chemical formulae were predicted. Searching for the predicted formulae in the dictionary of natural products found hits for *Streptomyces* and other actinomycete natural products such as ribostamycin and usabamycin A.[125–127]

There are several possibilities why no tryptopeptin A was produced by *S. sparsogenes*. First of all, it is possible that the tested conditions such as media, temperature, and growth time did not trigger expression of the *ttp* cluster. Secondly, it might be that the cluster is expressed, but that there are functional issues with one or more of the biosynthetic proteins. To further investigate the potential tryptopeptin A biosynthetic gene cluster from *S. sparsogenes* it was decided to clone the *ttp* cluster for heterologous expression. Dr Chuan Huang, senior postdoc in the Challis group,

therefore cloned the relevant cluster and transformed it into *Streptomyces coelicolor* M1154 for heterologous expression and cluster specific metabolite analysis.

4.2.3   Heterologous expression of the *ttp* cluster in *S. coelicolor* and potential metabolites of interest

If the identified cluster was fully expressed and functional, production of tryptopeptin A would be expected. However, if the epoxyketone synthase was inactive or different fatty acid residues were accepted, different metabolites could be produced as well. Dr Huang cloned the *ttp* cluster with approximately 3000 base pairs extension downstream and approximately 11000 base pairs extension upstream of the predicted cluster boundaries into *S. coelicolor* M1154 (Appendix 8.2). Heterologous expression and metabolite analysis revealed that the cloned cluster directed the production of several metabolites structurally related to tryptopeptin A (Figure 4.8).
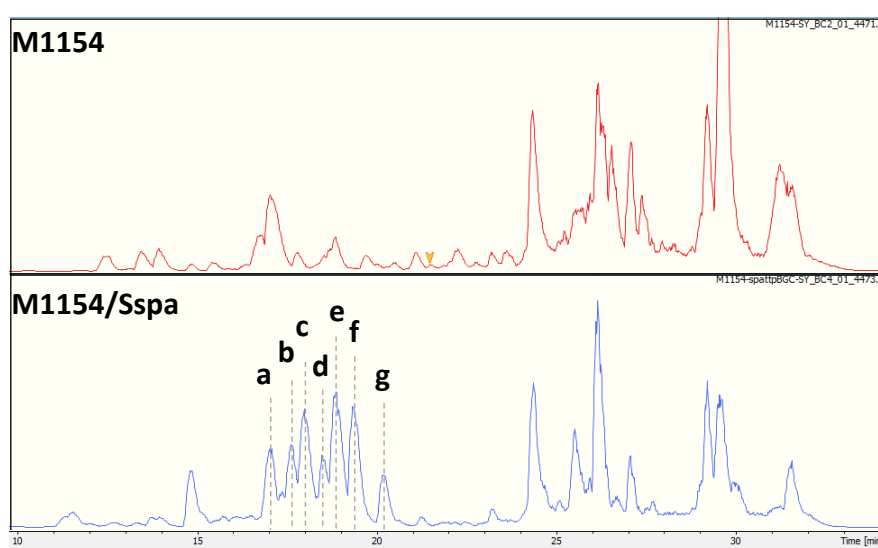
Figure 4.8 LCMS results from heterologous expression of the ttp cluster and metabolite extraction by Dr Huang. Top: S. coelicolor M1154 control metabolite expression and extraction. Bottom: S. coelicolor M1154 harbouring potential tryptopeptin cluster ttp from S. sparsogenes. Labelled peaks correspond to tryptopeptin A related metabolites discussed below.

Analysis of the metabolite peaks for their exact masses, predicted chemical structures, and fragmentation patterns revealed that peak a and d appear to correspond to metabolites with the same mass and predicted chemical formula as well as peak c and f appear to have the same mass and chemical formula prediction. In depth analysis of the observed metabolites by Dr Huang lead to five potential metabolite structures

related to tryptopeptin A (Figure 4.9). The mass and predicted structure of proposed metabolite **73** corresponds to peak a and d, proposed metabolite **74** to peak b, proposed metabolite **75** to peak c and f, proposed metabolite **76** to peak e, and proposed metabolite **77** to peak g. Potential metabolite **76** would include the epoxyketone pharmacophore although it would contain a different fatty acid residue compared to tryptopeptin A. If the observed mass and predicted structure indeed belong to the predicted structure, the epoxyketone synthase in the *ttp* cluster would be active and catalyse epoxyketone formation. Because none of the other predicted structures contained the epoxyketone pharmacophore, it is possible that there are issues with the encoded epoxyketone synthase such as inefficient expression, stability issues, or that it is otherwise low- or non-functional.
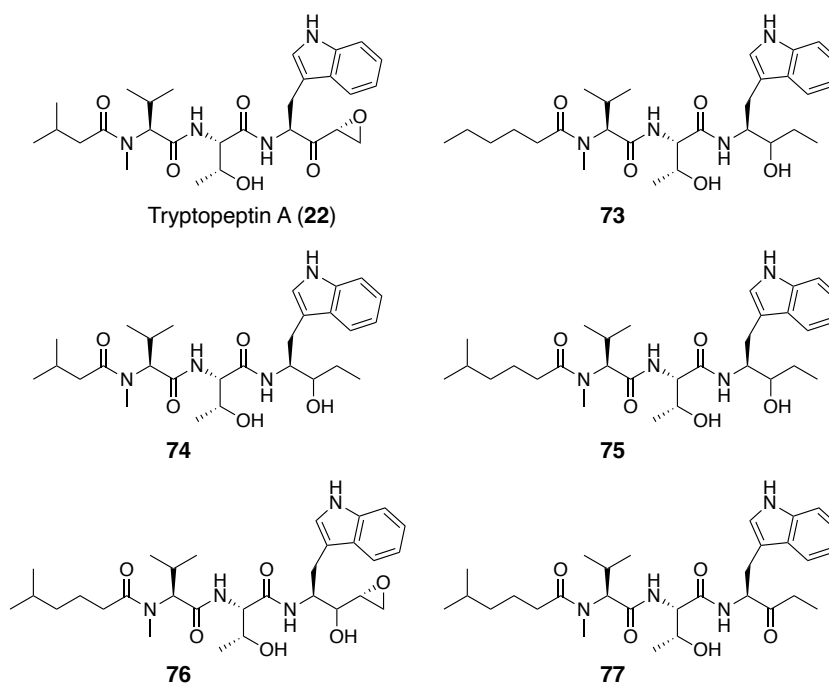
Figure 4.9 Structure of possible metabolites produced by Streptomyces strains containing the potential tryptopeptin A gene cluster.

The LCMS traces of the extracts from *S. sparsogenes* were subsequently extracted for the expected masses of the observed tryptopeptin A related metabolites from the heterologous expression experiments. However, none of the predicted structures were produced by *S. sparsogenes* under the used conditions. While it is still possible that expression of the *ttp* cluster was simply not triggered under the tested growth conditions, it could also be that proteins at the beginning of the assembly line for

tryptopeptin A are misfolded, not properly expressed, or otherwise non-functional or inhibited in *S. sparsogenes*. Performing proteomic experiments on *S. sparsogenes* could reveal more insight into if the predicted tryptopeptin A cluster is expressed under the tested conditions.

Because no tryptopeptin A production was observed in neither the metabolite expression of *S. sparsogenes*, nor in the heterologous expression of the *ttp* cluster, it was unclear if the potential epoxyketone synthase is properly expressed or functional at all. Comparing the *ttp* cluster to a known tryptopeptin A biosynthetic gene cluster would provide further insight and it was decided to investigate the known tryptopeptin A producer *Streptomyces* sp. maeda85 in more detail. A sample of *S.* sp. maeda85 and Illumina sequencing data of the strain was kindly provided by Dr Tsunematsu.

4.3 Analysis of metabolite profile of the known tryptopeptin A producer *Streptomyces* species maeda85

Because no tryptopeptin A or related metabolites were found to be produced under the tested conditions in *S. sparsogenes*, the known tryptopeptin A producer *Streptomyces* species maeda85 provided by Dr Tsunematsu was further investigated. *S.* sp. maeda85 was reactivated on MYG agar and grew in white colonies that formed spores after 10 days (Figure 4.10). The strain also produced yellow oily droplets that turned the colonies and surrounding medium yellow. Spore stocks were prepared after 10 days of growth and stored at – 80 °C.
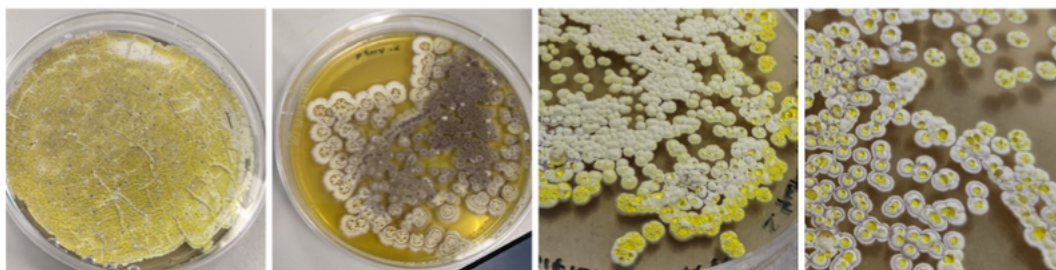
Figure 4.10 Streptomyces sp. maeda85 growing on MYG agar plates.

4.3.1 *Streptomyces* sp. maeda85 metabolite extraction and analysis

Cultures of *Streptomyces* sp. maeda85 were grown in MYG liquid medium and on MYG agar plates in triplicate and incubated each for 5 and 10 days. Metabolites were extracted with ethyl acetate and analysed by HRMS. Clear differences between the media control and the culture extracts suggest that *Streptomyces* sp. maeda85 is producing secondary metabolites (Figure 4.11).
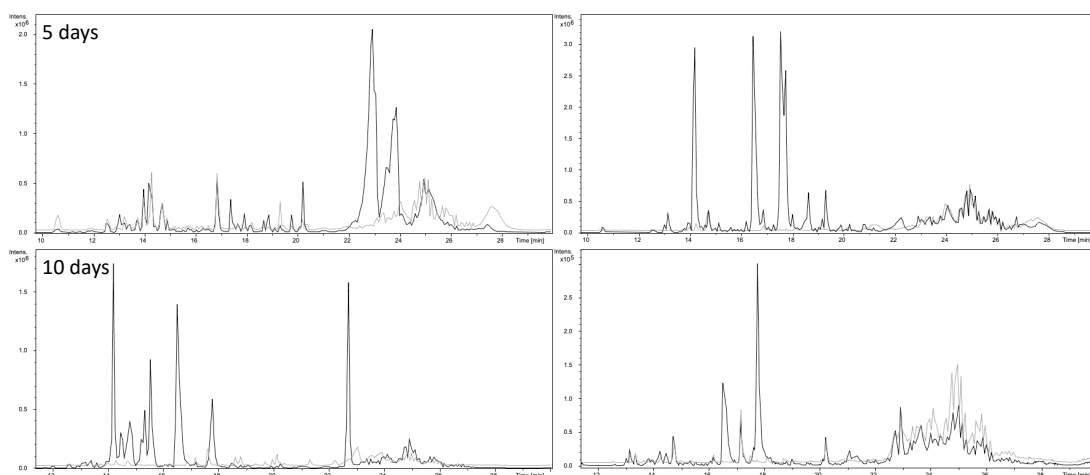
Figure 4.11 Base peak chromatograms of high-resolution mass spectrometry analysis of extracts of S. sp. maeda85 grown in MYG (black) and medium control (grey) after 5 days (top) and 10 days (bottom). Left: Extracts from MYG liquid media. Right: Extracts from MYG agar plates.

Searching the LCMS chromatogram of the extracts from *Streptomyces* sp. maeda85 grown on solid MYG medium for 5 days identified a small peak with the expected m/z of tryptopeptin A ($C_{28}H_{41}N_4O_6$ ([M+H]$^+$): 529.3021, $C_{28}H_{40}NaN_4O_6$ ([M+Na]$^+$): 551.2840). Chemical formula prediction revealed the correct chemical formula for tryptopeptin A ($C_{28}H_{41}N_4O_6^+$ for [M+H]$^+$, $C_{28}H_{40}N_4O_6Na^+$ for [M+Na]$^+$). Furthermore, MS2 data showed fragmentation between the serine and valine peptide bond with chemical formulae of the fragments being predicted as expected as well. This confirms that *Streptomyces* sp. maeda85 expresses the tryptopeptin A biosynthetic gene cluster and produces tryptopeptin A, albeit in low amounts (Figure 4.12).
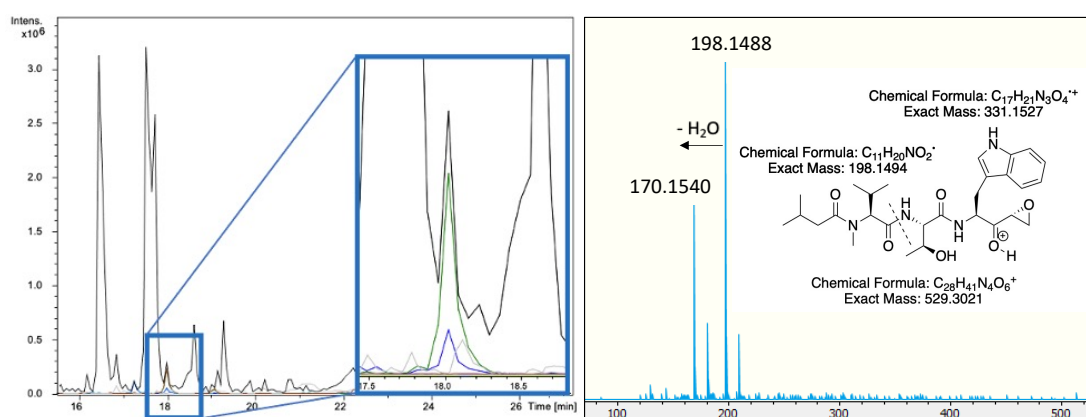
Figure 4.12 LCMS analysis of metabolite extracts from S. sp. maeda85 after 5 days of growth on plates. Left: Base peak chromatogram (black), and extracted chromatograms for m/z 529.3021 (blue) and m/z 551.2840 (green) corresponding to [M+H]$^+$ and [M+Na]$^+$ adduct of tryptopeptin A respectively. Right: MS2 fragments of m/z 529.3021 corresponding to [M+H]$^+$ with corresponding structure.

Furthermore, tryptopeptin A related compounds **73**, **74**, and **75** were also observed with the same molecular weight and predicted chemical formulae as the metabolites that were observed when the *ttp* cluster from *S. sparsogenes* was heterologously expressed in *S. coelicolor* (Figure 4.13).
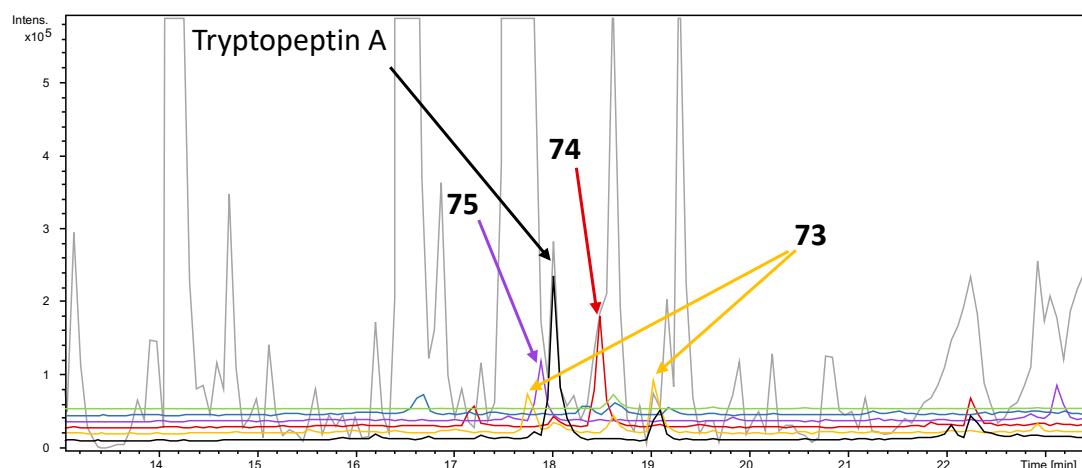
*Figure 4.13 LCMS results from metabolite extraction of S. sp. maeda85 solid cultures after 5 days of growth. The base peak chromatogram (grey) was extracted for m/z corresponding to [M+H]$^{+}$ of tryptopeptin A related metabolites **73** ($C_{29}H_{47}N_4O_5$, m/z 531.3531, orange), **74** ($C_{28}H_{45}N_4O_5$, m/z 517.3384, red), **75** ($C_{30}H_{49}N_4O_5$, m/z 545.3697, purple), **76** ($C_{30}H_{47}N_4O_6$, m/z 559.3490, blue), and **77** ($C_{30}H_{47}N_4O_5$, m/z 543.3541, green).*

Additionally, the major metabolite peaks from the *S.* sp. maeda85 extracts were analysed for mass and chemical formulae were predicted. Searching for the predicted formulae in the dictionary of natural products found hits for *Streptomyces* natural product geralcin A and terpene natural products such as lycocernuine.[128,129] Overall, the metabolic profile of *S.* sp. maeda85 appears significantly different from *S. sparsogenes* both for liquid and plate cultures.

To investigate the potential tryptopeptin A biosynthetic gene cluster from *S.* sp. maeda85 it was decided to conduct heterologous expression experiments. Dr Huang, therefore cloned the relevant cluster and transformed it into *Streptomyces coelicolor* M1154 for heterologous expression and cluster specific metabolite analysis.

4.3.2 Heterologous expression of the *tpt* cluster in *S. coelicolor*

The potential tryptopeptin A directing biosynthetic gene cluster from *S.* sp. maeda85 was cloned with approximately 13000 base pairs extension downstream and approximately 5000 base pairs extension upstream of the expected cluster boundaries by Dr Huang (Appendix 8.2). The cluster was transformed into *Streptomyces coelicolor* M1154, heterologously expressed, and examined for metabolite production by Dr Huang (Figure 4.14).
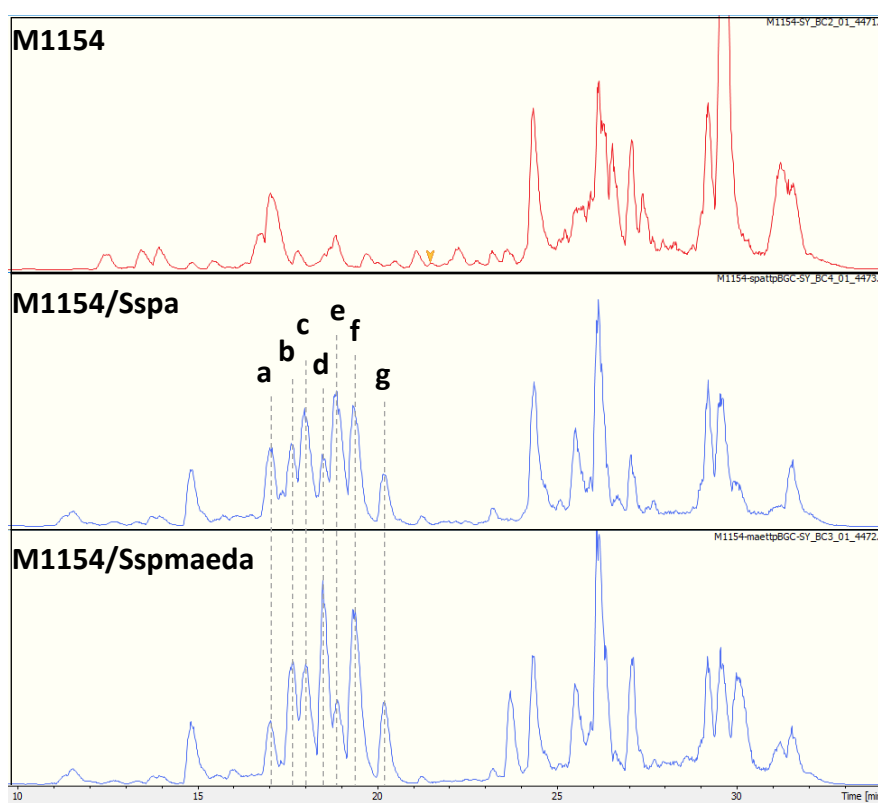
Figure 4.14 LCMS results from heterologous expression of the S. sp. maeda85 cluster of interest and metabolite extraction by Dr Huang. Top: S. coelicolor M1154 control metabolite expression and extraction. Middle: M1154 harbouring potential tryptopeptin cluster ttp from S. sparsogenes. Bottom: M1154 harbouring potential tryptopeptin cluster from S. sp. maeda85. Labelled peaks correspond to tryptopeptin A related metabolites.

Analysis of the metabolite peaks for their exact masses, predicted chemical structures, and fragmentation patterns revealed that the metabolites produced by the cloned cluster from *S.* sp. maeda85 appeared identical to the tryptopeptin A related metabolites observed in heterologous expression experiments with the *ttp* cluster from *S. sparsogenes*. The metabolic profiles of *S. coelicolor* harbouring each heterologously expressed cluster appear similar which could mean that the two

clusters are identical or at least that all involved biosynthetic proteins are highly homologous.

Additionally, Challis group member Yu Zhang performed metabolite expression and extraction experiments of *S. coelicolor* harbouring each the *ttp* and *tpt* cluster for tandem mass spectrometry analysis to further confirm the metabolite structures predicted by Dr Huang (Figure 4.15). All five metabolites **73**-**77** predicted by Dr Huang were observed in both metabolic profiles.
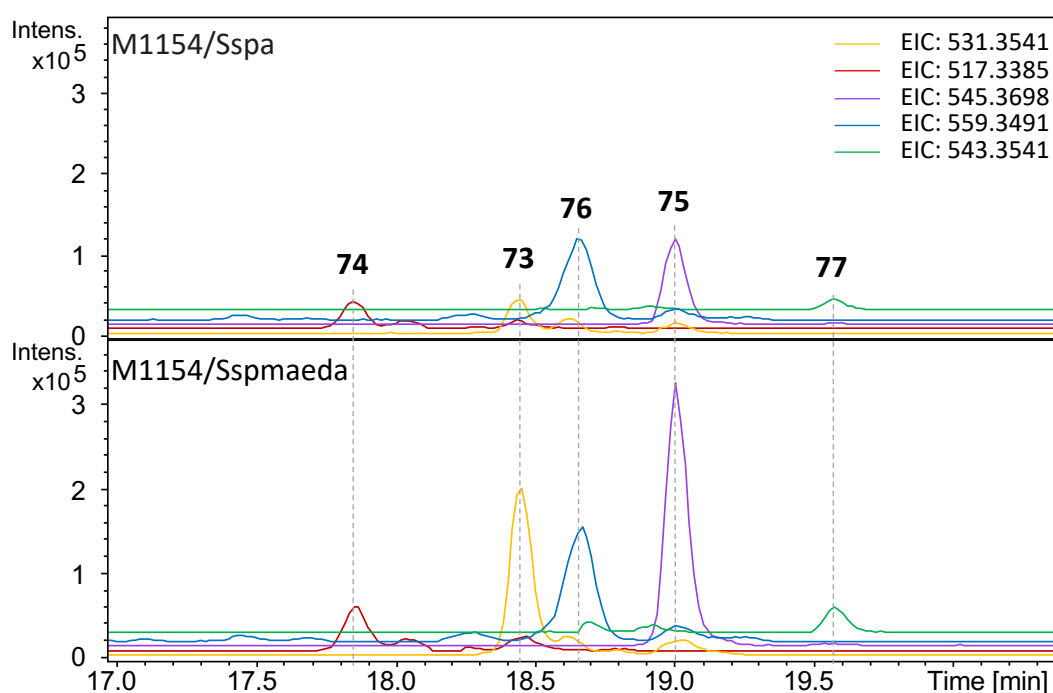
*Figure 4.15 LCMS results from metabolite extraction of S. coelicolor harbouring the ttp cluster from S. sparsogenes (top) and the tpt cluster from S. sp. maeda85 (bottom). The base peak chromatogram was extracted for m/z corresponding to [M+H]⁺ of tryptopeptin A related metabolites **73** ($C_{29}H_{47}N_4O_5$, m/z 531.3541, orange), **74** ($C_{28}H_{45}N_4O_5$, m/z 517.3385, red), **75** ($C_{30}H_{49}N_4O_5$, m/z 545.3698, purple), **76** ($C_{30}H_{47}N_4O_6$, m/z 559.3491, blue), and **77** ($C_{30}H_{47}N_4O_5$, m/z 543.3541, green).*

The observed peaks of the five tryptopeptin A related metabolites **73**-**77** were analysed by Yu Zhang to investigate ion fragmentation patterns and determine the masses and predicted chemical formulae of the observed fragments. MS/MS analysis of metabolites **73**-**77** revealed fragments that matched the predicted structures by fragmentation between peptide bonds, side chains, and loss of carbon monoxide or water (Figure 4.16A-D). However, the fragmentation of metabolite matching the mass of predicted structure **77** was observed in very low intensity and no concluding analysis could be made (Figure 4.16E). These experiments are strong indicators that

the structures predicted by Dr Huang are the metabolites produced both by the *ttp* and *tpt* cluster. Both clusters appear to incorporate a variety of fatty acid chains and seem to mostly produce tryptopeptin A related metabolites without the epoxyketone pharmacophore which indicates low turnover and difficulties with the involved epoxyketone synthases.
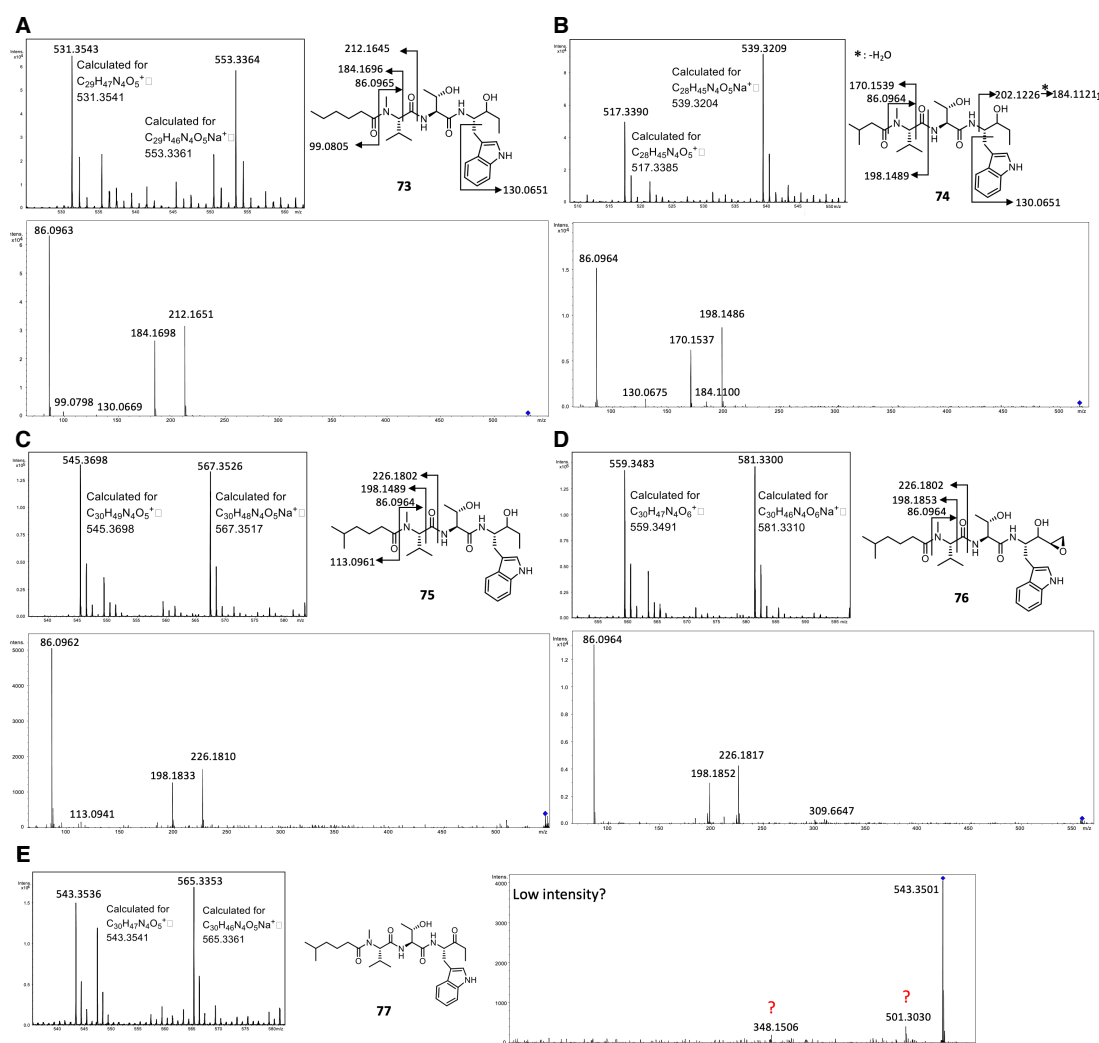
Figure 4.16 MS/MS analysis of predicted tryptopeptin A related metabolites 73-77. A. MS/MS analysis and predicted fragmentation of expected metabolite 73. B. MS/MS analysis and predicted fragmentation of expected metabolite 74. C. MS/MS analysis and predicted fragmentation of expected metabolite 75. D. MS/MS analysis and predicted fragmentation of expected metabolite 76. E. MS/MS analysis and predicted fragmentation of expected metabolite 77.

Detailed cluster comparison of the *ttp* cluster and the cluster of interest in *S.* sp. maeda85 would provide more insight into whether the clusters are identical or could

highlight their differences. Because the potential cluster of interest in *S.* sp. maeda85 was cut off on the edge of a contig in the provided Illumina sequencing data by Dr Tsunematsu the full genome of *S.* sp. maeda85 was re-sequenced with nanopore sequencing. The published genomic data of *S. sparsogenes* ATCC 25498 was Illumina sequencing data split between 310 contigs which is why it was decided to re-sequence with nanopore sequencing as well to compare the clusters of interest and full genomes as accurately as possible.

4.4 Genomic DNA extraction, nanopore sequencing, and genome assembly and annotation

4.4.1 *Streptomyces sparsogenes* ATCC. 25498

Isolation of genomic DNA with the QIAGEN MagAttract® HMW DNA Kit yielded high molecular weight DNA at a concentration of 26.2 ng/µL for sequencing experiments. The supplier's protocol for Gram-positive bacteria was followed and consistent high molecular weight was confirmed by agarose gel electrophoresis (Figure 4.17).
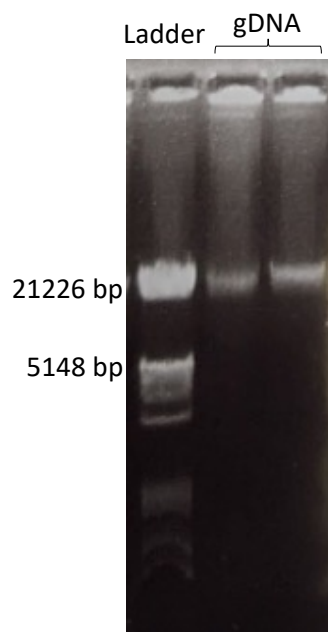
Figure 4.17 Agarose gel (0.6%) of genomic DNA extracted from S. sparsogenes with the QIAGEN MagAttract® HMW DNA Kit and λ DNA + EcoRI + HinDIII marker.

Illumina sequencing data for *S. sparsogenes* was previously reported as a draft genome on one linear chromosome of a length of 10049240 base pairs spread out over 310 contigs.[130] Because the data is split between numerous contigs, whole-genome sequencing with Oxford nanopore technologies was performed. The high molecular weight DNA was prepared according to the native barcoding for genomic DNA protocol and sequenced on the MinION Mk1B flow cell. From the generated nanopore data, the barcodes were trimmed, and the data was assembled, polished, and annotated. The total genome size was found to be 9943733 base pairs long with a long linear

chromosome contig of 9885695 base pairs and a short contig of 58038 base pairs possibly being a plasmid (Table 4.2).

Table 4.2 Statistics of the genome assembly from nanopore data of S. sparsogenes ATCC 25498.

Parameters	S. sparsogenes
Numbers of contigs	2
Total contigs length	9951953
Mean contig size	4975976.5
Contig size first quartile	9893942
Median contig size	9893942
Contig size third quartile	9893942
Longest contig	9893942
Shortest contig	58011
Contigs > 500 nt	2
Contigs > 1K nt	2
Contigs > 10K nt	2
Contigs > 100K nt	1
Contigs > 1M nt	1
N50	9893942
L50	1
N80	9893942
L80	1

The genome assembly was assessed with Benchmarking Universal Single-Copy Orthologs (BUSCO) for expected core conserved genes and a final BUSCO score of 95.2% was achieved (Table 4.3).[131]

Table 4.3 Details from the BUSCO genome assembly assessment showing the conserved genes identified within the scaffold of S. sparsogenes genome assembly.

BUSCO Scaffold Stat	S. sparsogenes
Percentage BUSCO	95.2%
Complete BUSCO's	1503
Complete and single copy BUSCO's	1493
Complete and duplicate BUSCO's	10
Fragmented BUSCO's	33
Missing BUSCO's	43
Total BUSCO groups searched	1579

The acquired data provided full genome coverage and a more complete genome sequence than the previously published Illumina data. While Zhang *et al.* reported 41 potential biosynthetic gene clusters found in their Illumina sequencing data, analysis

of the nanopore sequencing data with antiSMASH predicted a total of 44 biosynthetic gene clusters of various types including the potential tryptopeptin A gene cluster (Table 4.4). Of the 44 clusters, the majority of clusters was identified as commonly known biosynthetic gene cluster types: 6 NRPS clusters, 5 PKS clusters, 7 NRPS/PKS hybrid clusters, 6 terpene clusters, 4 RiPP-like clusters, 6 lanthipeptide clusters, and 10 other clusters. AntiSMASH predictions of the most similar known clusters revealed 100% similarity for 9 clusters in *S. sparsogenes* of which some are common *Streptomyces* secondary metabolites such as coelibactin and geosmin. Region 1.12 corresponds to the potential tryptopeptin A cluster of interest.

Table 4.4 AntiSMASH output overview of the S. sparsogenes genome assembled and annotated from nanopore sequencing data. Clusters are assigned a type and most similar known cluster.

Identified secondary metabolite regions using strictness 'relaxed'

contig_1 (Genus species)

Region	Type	Most similar known cluster		Similarity
Region 1.1	NRPS	coelibactin	NRP	100%
Region 1.2	hglE-KS, T1PKS	A-500359 A / A-500359 B	NRP	5%
Region 1.3	NRPS	lactonamycin	Polyketide	5%
Region 1.4	transAT-PKS, NRPS	phthoxazolin	NRP + Polyketide	18%
Region 1.5	NRPS, T1PKS, thiopeptide, thioamitides, LAP	hygrocin A / hygrocin B	Polyketide	19%
Region 1.6	T2PKS	anthrabenzoxocinone	Polyketide	100%
Region 1.7	nucleoside	tubercidin	Other	81%
Region 1.8	ranthipeptide	rabelomycin / dehydrorabelomycin / fluostatin F / fluostatin G / fluostatin H	Polyketide:Type II	4%
Region 1.9	terpene, LAP	brasilicardin A	Terpene + Saccharide	46%
Region 1.10	butyrolactone	A-factor	Other	100%
Region 1.11	terpene	geosmin	Terpene	100%
Region 1.12	T1PKS, NRPS	landepoxcin	NRP + Polyketide	22%
Region 1.13	lanthipeptide-class-i			
Region 1.14	terpene	hopene	Terpene	61%
Region 1.15	lassopeptide	anantin C	RiPP	75%
Region 1.16	T2PKS	spore pigment	Polyketide	83%
Region 1.17	RiPP-like			
Region 1.18	NRPS	enduracidin	NRP	6%
Region 1.19	siderophore			
Region 1.20	terpene	heronamide A / heronamide B / heronamide C / heronamide D / heronamide E / heronamide F	NRP + Polyketide	8%
Region 1.21	NRPS-like	echoside A / echoside B / echoside C / echoside D / echoside E	NRP	88%
Region 1.22	lipolanthine, lanthipeptide-class-iii	naphthomycin A	Polyketide	15%
Region 1.23	siderophore	desferrioxamine E	Other	100%
Region 1.24	lanthipeptide-class-v, thioamitides	neothioviridamide	RiPP	100%
Region 1.25	lassopeptide	citrulassin A	RiPP	75%
Region 1.26	ladderane, arylpolyene	atratumycin	NRP	26%
Region 1.27	NRPS, T2PKS, RRE-containing, oligosaccharide, PKS-like	antibiotic HKI 10311129 / antibiotic A2121-1 / cervimycin D / cervimycin C / cervimycin A	Polyketide + Saccharide:Hybrid/tailoring	54%
Region 1.28	RRE-containing			
Region 1.29	lanthipeptide-class-i			
Region 1.30	lanthipeptide-class-ii	SBI-06990 A1 / SBI-06989 A2	RiPP:Lanthipeptide	100%
Region 1.31	lanthipeptide-class-iii	SapB	RiPP:Lanthipeptide	75%
Region 1.32	ectoine	ectoine	Other	100%
Region 1.33	siderophore			
Region 1.34	T1PKS, transAT-PKS	nanchangmycin	Polyketide	87%
Region 1.35	amglyccycl	hydroxystreptomycin	Saccharide	72%
Region 1.36	transAT-PKS-like, NRPS, transAT-PKS	oxalomycin B	NRP + Polyketide	45%
Region 1.37	terpene	pentalenolactone	Terpene	52%
Region 1.38	RiPP-like			
Region 1.39	NRPS-like, terpene	isorenieratene	Terpene	85%
Region 1.40	NRPS	A54145	NRP	6%
Region 1.41	NRPS, PKS-like, lassopeptide	foxicins A-D	NRP + Polyketide	19%
Region 1.42	terpene	pristinol	Terpene	100%
Region 1.43	T1PKS, NRPS	griseoviridin / fijimycin A	NRP:Cyclic depsipeptide + Polyketide:Trans-AT type I	11%
Region 1.44	T1PKS	sanglifehrin A	NRP + Polyketide	15%

4.4.2 *Streptomyces* sp. maeda85

The genomic DNA was isolated with the QIAGEN MagAttract® HMW DNA Kit as described for *S. sparsogenes* (Chapter 4.4.1) and DNA was obtained at a concentration of 54.6 ng/µL. Consistent high molecular weight distribution was confirmed by gel

electrophoresis before proceeding with the native barcoding for genomic DNA protocol for DNA sequencing on the MinION Mk1B flow cell (Figure 4.18).
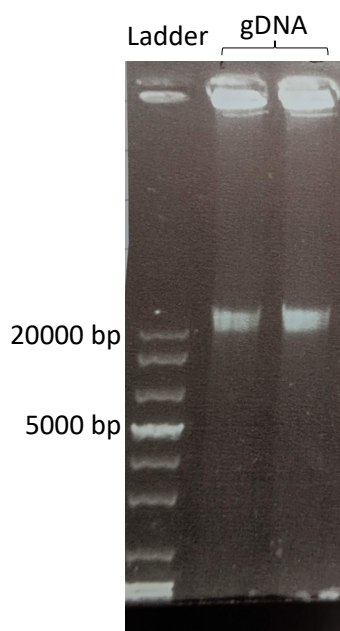
Figure 4.18 Agarose gel (0.6%) of genomic DNA extracted from S. sp. maeda85 with the QIAGEN MagAttract® HMW DNA Kit and GeneRuler 1 kb Plus DNA ladder.

The MinION Mk1B flow cell was used to sequence the high molecular weight DNA that was prepared according to the native barcoding for genomic DNA protocol. The barcode of the generated nanopore data were trimmed and the data was assembled, polished, and annotated to generate a full genome sequence of 12410231 base pairs. Additionally, the assembled genome was polished with Illumina data provided by Dr Tsunematsu to reveal a more accurate *S.* sp. maeda85 genome of the size of 12410045 base pairs (Table 4.5). The genome appears to be split between seven contigs with the chromosome probably being the largest contig of 11.6 Mb and six plasmids between 33.1 kb and 436.8 kb long.

Parameters	*S.* sp. maeda85	*S.* sp. maeda85 polished
Numbers of contigs	7	7
Total contigs length	12410231	12410045
Mean contig size	1772890.14	1772863.57
Contig size first quartile	79029	79028
Median contig size	127988	127985
Contig size third quartile	11613803	11613619
Longest contig	11613803	11613619
Shortest contig	33097	22098
Contigs > 500 nt	7	7
Contigs > 1K nt	7	7
Contigs > 10K nt	7	7
Contigs > 100K nt	3	3
Contigs > 1M nt	1	1
N50	11613803	11613619
L50	1	1
N80	11613803	11613619
L80	1	1

The genome assemblies were assessed with BUSCO for core conserved genes and a final BUSCO score of 98.8% and 99.2% were achieved for the nanopore sequenced genome and the genome polished with Illumina data respectively (Table 4.6).

Table 4.6 Details from the BUSCO genome assembly assessment showing the conserved genes identified within the scaffold of S. sp. maeda85 genome assemblies.

BUSCO Scaffold Stat	*S.* sp. maeda85	*S.* sp. maeda85 polished
Percentage BUSCO	98.8%	99.2%
Complete BUSCO's	1560	1566
Complete and single copy BUSCO's	1548	1554
Complete and duplicate BUSCO's	12	12
Fragmented BUSCO's	10	6
Missing BUSCO's	9	7
Total BUSCO groups searched	1579	1579

Analysis of the full annotated genome with antiSMASH predicted a total of 52 biosynthetic gene clusters of various types including the potential tryptopeptin A gene cluster (Table 4.7). Of the 52 clusters, the majority of clusters was identified as commonly known biosynthetic gene cluster types: 8 NRPS clusters, 9 PKS clusters, 7 NRPS/PKS hybrid clusters, 7 terpene clusters, 6 RiPP clusters, 5 lanthipeptide clusters, and 10 other clusters. AntiSMASH predictions of the most similar known

clusters revealed 100% similarity for 6 clusters in *S.* sp. maeda85 of which some are common *Streptomyces* secondary metabolites such as geosmin. Region 1.17 corresponds to the potential tryptopeptin A cluster of interest. The *S.* sp. maeda85 genome and cluster predictions are compared to the *S. sparsogenes* genome and predicted clusters from Chapter 4.4.1.

Table 4.7 AntiSMASH output overview of the S. sp. maeda85 genome assembled and annotated from nanopore sequencing data. Clusters are assigned a type and most similar known cluster.

Identified secondary metabolite regions using strictness 'relaxed'

c00001_contig_.. (original name was: contig_1_pilon_pilon_pilon_pilon)

Region	Type	Most similar known cluster		Similarity
Region 1.1	terpene	colabomycin E	Polyketide:Type II	6%
Region 1.2	nucleoside	tubercidin	Other	81%
Region 1.3	T1PKS	nystatin	Polyketide	52%
Region 1.4	T1PKS	meilingmycin	Polyketide	53%
Region 1.5	furan , NRPS , T1PKS , hglE-KS	atratumycin	NRP	7%
Region 1.6	T1PKS	ebelactone	Polyketide	5%
Region 1.7	terpene	pristinol	Terpene	100%
Region 1.8	NAPAA , NRPS , ladderane	RP-1776	Polyketide + NRP:Cyclic depsipeptide	16%
Region 1.9	PKS-like , NRPS	decaplanin	NRP:Glycopeptide	5%
Region 1.10	T1PKS	foxicins A-D	NRP + Polyketide	12%
Region 1.11	lanthipeptide-class-ii			
Region 1.12	NRPS-like , ranthipeptide	echoside A / echoside B / echoside C / echoside D / echoside E	NRP	11%
Region 1.13	terpene	2-methylisoborneol	Terpene	100%
Region 1.14	butyrolactone			
Region 1.15	terpene	geosmin	Terpene	100%
Region 1.16	redox-cofactor	reedsmycins	Polyketide	10%
Region 1.17	T1PKS , NRPS , LAP , thiopeptide , thioamitides	landepoxcin	NRP + Polyketide	16%
Region 1.18	NAPAA	foxicins A-D	NRP + Polyketide	7%
Region 1.19	terpene	hopene	Terpene	76%
Region 1.20	lassopeptide	chaxapeptin	RiPP	57%
Region 1.21	T2PKS	spore pigment	Polyketide	83%
Region 1.22	transAT-PKS , NRPS	phthoxazolin	NRP + Polyketide	39%
Region 1.23	RiPP-like			
Region 1.24	siderophore			
Region 1.25	terpene	meridamycin	NRP + Polyketide	5%
Region 1.26	NRPS-like	echoside A / echoside B / echoside C / echoside D / echoside E	NRP	88%
Region 1.27	siderophore	desferrioxamin B	Other	100%
Region 1.28	lanthipeptide-class-i			
Region 1.29	NRPS , NRPS-like	phosphonoglycans	Saccharide	3%
Region 1.30	lassopeptide	citrulassin A	RiPP	50%
Region 1.31	ladderane , arylpolyene	atratumycin	NRP	31%
Region 1.32	NRPS , T2PKS , RRE-containing , oligosaccharide , PKS-like	saquayamycin Z	Polyketide:Type II + Saccharide:Oligosaccharide	47%
Region 1.33	lanthipeptide-class-i			
Region 1.34	lanthipeptide-class-ii	SBI-06990 A1 / SBI-06989 A2	RiPP:Lanthipeptide	100%
Region 1.35	lassopeptide	SSV-2083	RiPP:Lanthipeptide	36%
Region 1.36	lassopeptide			
Region 1.37	lanthipeptide-class-iii	SapB	RiPP:Lanthipeptide	75%
Region 1.38	PKS-like	chlortetracycline	Polyketide	14%
Region 1.39	ectoine	ectoine	Other	100%
Region 1.40	siderophore			
Region 1.41	T1PKS	nanchangmycin	Polyketide	60%
Region 1.42	NRPS-like , T1PKS	himastatin	NRP	8%
Region 1.43	NRPS-like , T1PKS , other , NRPS , transAT-PKS	amipurimycin	Polyketide	66%
Region 1.44	redox-cofactor			
Region 1.45	NRPS	atratumycin	NRP	7%
Region 1.46	T1PKS , RiPP-like	amycomicin	Polyketide	50%
Region 1.47	hglE-KS	heterocyst glycolipids	Other	28%
Region 1.48	NRPS	streptolydigin	NRP + Polyketide:Modular type I + Saccharide:Hybrid/tailoring	7%
Region 1.49	T1PKS	meilingmycin	Polyketide	27%
Region 1.50	terpene	actinomycin D	NRP	21%
Region 1.51	RiPP-like			
Region 1.52	NRPS , NRPS-like	BE-43547A1 / BE-43547A2 / BE-43547B1 / BE-43547B2 / BE-43547B3 / BE-43547C1 / BE-43547C2	NRP:Cyclic depsipeptide + Polyketide:Modular type I	13%

4.4.2.1 Cluster analysis with antiSMASH

The tryptopeptin A cluster in *S.* sp. maeda85 was further analysed with AntiSMASH and BLAST. Predicted functions of the encoded proteins were assigned the corresponding genes (Figure 4.19). The analysed biosynthetic gene cluster contains a proteasome β-subunit which probably is relevant for the species resistance towards the produced epoxyketone, and an oxidoreductase homologue was found as well. While the *S. sparsogenes* cluster contains additional predicted dehydrogenases and an aminotransferase downstream of the NRPS/PKS unit, genes in the equivalent positions in the *S.* sp. maeda85 cluster were predicted to encode for regulatory and transport proteins.

Figure 4.19 Potential tryptopeptin A producing biosynthetic gene cluster tpt from S. sp. maeda85 with the epoxyketone synthase homologue encoding gene ttpC.

This detailed cluster analysis revealed that the *ttp* cluster in *S. sparsogenes* and this potential tryptopeptin A biosynthetic gene cluster in *S.* sp. maeda85, the *tpt* cluster, are different clusters after all. While their main biosynthetic genes, the ACAD homologue, the NRPS, and the PKS are very similar, they contain different genes encoding for other functional and regulatory proteins.

4.4.2.2 Cluster comparison of S. sp. maeda85 tryptopeptin cluster

Including the *S.* sp. maeda85 cluster of interest in the cluster alignments with the previously identified clusters of interest for potential tryptopeptin A production confirms high cluster similarity with the *S. bingchenggensis* cluster of interest and high gene similarity with the core genes of the *S. sparsogenes* cluster of interest (Figure 4.20).
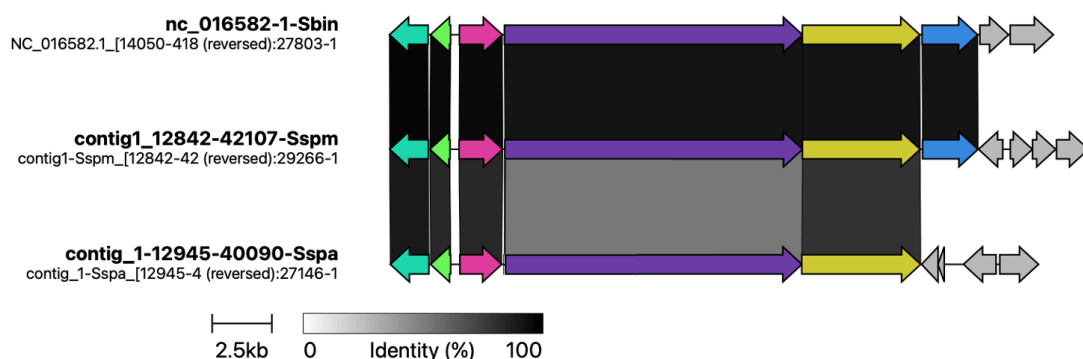
Figure 4.20 Cluster alignments of the potential tryptopeptin A biosynthetic gene clusters in S. bingchenggensis, S. sp. maeda85, and S. sparsogenes. Clusters were aligned to the epoxyketone synthase homologues with clinker.[123]

4.4.2.3 A domain specificity prediction

A domain specificity prediction for NRPS in the cluster of interest from *S.* sp. maeda85 revealed identical amino acid residues in the substrate specificity pockets to the ones predicted for the clusters of interest from *S. sparsogenes* and *S. bingchenggensis* (Table 4.8). The predicted amino acids incorporated in the potentially produced natural product are therefore as well a valine residue, a threonine residue, and an unpredicted substrate. Overall, the NRPS in the cluster of interest in the *S.* sp, maeda85 strain shows 97% similarity to the NRPS in the *ttp* cluster in *S. sparsogenes* and 99% similarity to the NRPS in the *S. bingchenggensis* strain.

Table 4.8 Outcome of A domain specificity predictions of clusters of interest in identified Streptomyces strains.[124] *Listed are amino acids of assigned function in the substrate specificity pockets of the adenylation domains in the NRPS encoded by the clusters of interest and their predicted substrate specificity.*

	A domain 1	A domain 2	A domain 3
S. bingchenggensis	Val	Thr	No hit
Residues active site	DAYWWGGT	DFWNIGMV	DASVALCV
S. sparsogenes	Val	Thr	No hit
Residues active site	DAYWWGGT	DFWNIGMV	DASVALCV
***S.* sp. maeda85**	Val	Thr	No hit
Residues active site	DAYWWGGT	DFWNIGMV	DASVALCV

4.4.2.4 Phylogenetic analysis of epoxyketone synthases and homologues

Sequences of known epoxyketone synthases and novel potential epoxyketone synthases identified in this work were aligned with ClustalW to generate a phylogenetic tree and visualise evolutionary relationships between them (Figure 4.21).[132] The epoxyketone synthases involved in the landepoxcin and clarepoxcin biosynthesis formed a clade with EpxF, the epoxyketone synthase involved in the epoxomicin biosynthesis. EpnF and TmcF formed their own clade with a very close evolutionary link. The three novel predicted epoxyketone synthases from *Streptomyces* strain formed their own clade while the ACAD homologue in *S. bingchenggensis* and *S.* sp. maeda85 appeared to be the most similar. Comparing both potential epoxyketone synthases with BLAST revealed that they are 96% identical to each other with only very few different residues. The potential epoxyketone synthases from the identified *C. fuscus* strain appears to be evolutionary the most disconnected from all the other epoxyketone synthases.
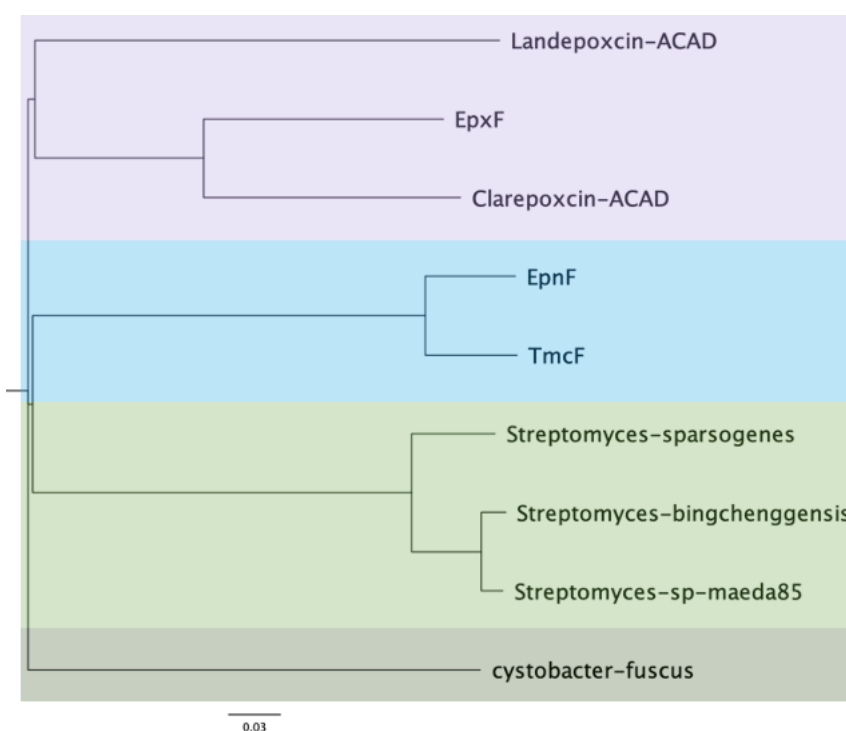
Figure 4.21 Phylogenetic tree of known epoxyketone synthases and novel predicted epoxyketone synthases.

4.4.3 Genome comparison

The genomes of *S. sparsogenes* and *S.* sp. maeda85 were re-sequenced with nanopore sequencing and resulted in high quality complete genome sequences for both strains while the genome of *S.* sp. maeda85 was additionally polished with Illumina sequencing data and is therefore of slightly higher quality. Both strains were predicted to harbour many putative biosynthetic gene clusters by AntiSMASH. In comparison, more identified clusters in *S. sparsogenes* show 100% similarity with known clusters, however, *S.* sp. maeda85 seems to harbour more biosynthetic gene clusters overall, especially more PKS clusters (Figure 4.22).
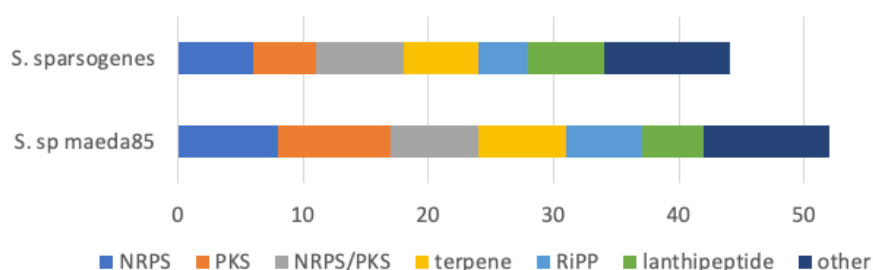
Figure 4.22 Comparison of the biosynthetic gene clusters and their predicted metabolite in S. sparsogenes and S. sp. maeda85.

Several known clusters are present in both strains such as biosynthetic gene clusters that direct the production of geosmin, pristinol, and ectoine (Table 4.9). Additionally, some seemingly novel clusters appear in both strains as well, like a cluster with 88% similarity to echoside clusters, a cluster with 83% similarity to a polyketide spore cluster, and a cluster 81% similar to tubercidin cluster. Similarity in this context refers to the percentage of genes which are homologous to genes in the known cluster while the individual enzymes can differ in specificity.

Table 4.9 Table of clusters predicted in both sequenced strains listed with their type and similarity to clusters in S. sparsogenes and S. sp. maeda85.

type	known cluster	similarity
terpene	geosmin	100
lanthipeptide	SBI-06990 A1/2	100
other	ectoine	100
terpene	pristinol	100
NRPS	echoside A-E	88
PKS	spore pigment	83
other	tubercidin	81
lanthipeptide	SapB	75

Overall, the shared clusters appear in different areas in the genomes of both strains and there are several biosynthetic gene clusters that are only present in one of the strains (Table 4.10). While *S. sparsogenes* and *S.* sp. maeda85 seem to share some biosynthetic gene clusters, they are obviously two different strains each harbouring the information to produce a variety of secondary metabolites.

Table 4.10 Table of known clusters predicted for S. sparsogenes (left) and S. sp. maeda85 (right) listed with the cluster type.

S. sparsogenes

type	known cluster	similarity
NRPS	coelibactin	100
PKS	anthrabenzoxocinone	100
other	A-factor	100
other	desferrioxamine E	100
lanthipeptide	neothloviridamide	100
terpene	pristinol	100

S. sp. maeda85

type	known cluster	similarity
terpene	2-methylisoborneol	100
other	desferrioxamine B	100

4.5 Overproduction and purification of novel potential epoxyketone synthases from *S. sparsogenes* (TtpC) and *S.* sp. maeda85 (TptC)

To test the *in vitro* activity of the potential epoxyketone synthases TtpC and TptC from *S. sparsogenes* and *S.* sp. maeda85 respectively, the corresponding genes were each amplified from genomic DNA, cloned into an expression vector, overproduced, and purified.

4.5.1 Cloning *ttpC* and *tptC* into expression vector pET151

The TtpC and TptC encoding genes were each amplified by PCR and cloned into pET151 protein expression vector with the Invitrogen Champion™ pET151 Directional TOPO® Expression Kit (Appendix 8.1, Figure 4.23).
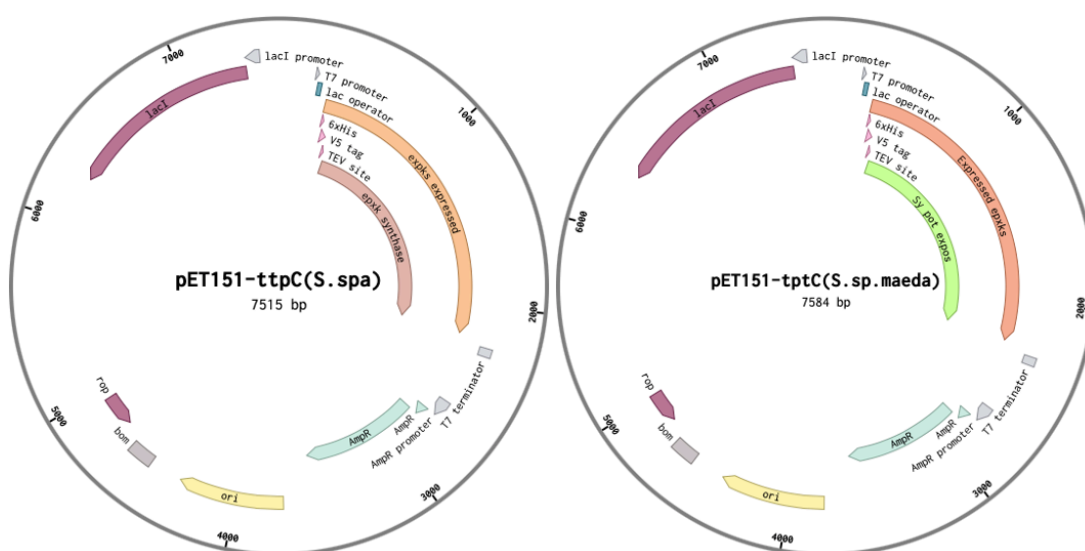
Figure 4.23 Plasmid maps of the desired plasmids pET151-ttpC (left) and pET151-tptC (right).

Restriction digests of both plasmids with EcoRV revealed that the correct genes were inserted into the vector (Figure 4.24). Sequencing with T7 forward and reverse primers of each plasmids ensured that the genes were inserted in the correct orientation and without any mutations (Appendix 8.3).
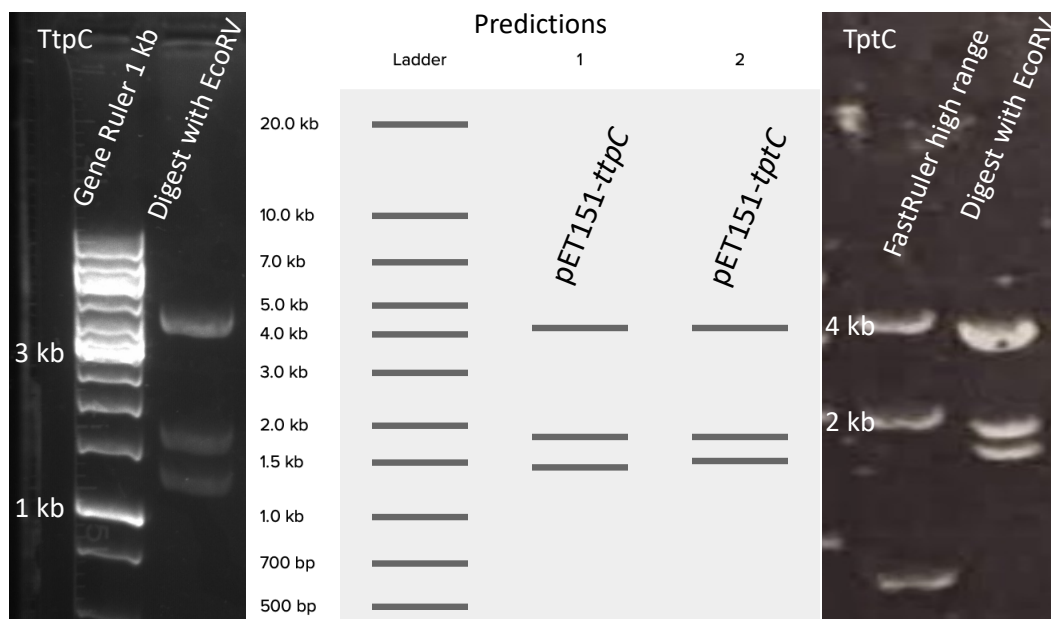
Figure 4.24 Agarose gel (1%) of plasmid pET151-ttpC (left) and pET151-tptC (right) digested with EcoRV and predictions thereof.

4.5.2 Overproduction and purification of potential epoxyketone synthases TtpC and TptC

The potential epoxyketone synthases TtpC and TptC were overproduced as described for EpnF (Chapter 2.2) and purified by Nickel affinity chromatography to yield the desired proteins (TtpC in 14.5 mg/L, TptC in 13.2 mg/L). SDS-PAGE gel analysis suggested overproduction of both desired proteins significantly pure (Figure 4.25). Analysis by intact protein high-resolution mass spectrometry was attempted for TttC and TptC, however, unfortunately the obtained resolution was too low for accurate analysis. Because the cloned genes were confirmed by sequencing and the SDS-PAGE gel suggests the correct proteins were purified, it is highly likely that TtpC and TptC were the proteins obtained.
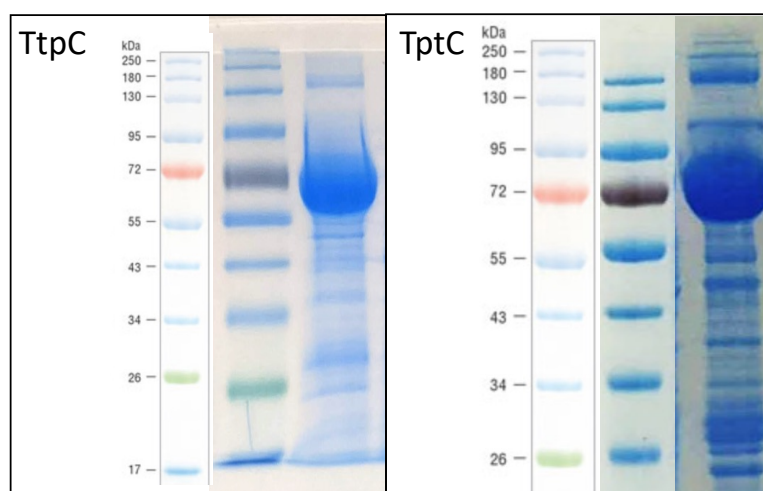
Figure 4.25 SDS-PAGE gels and HRMS of purified TtpC (left) and TptC (right).

4.6 Coupled enzyme assays with PLE and novel epoxyketone synthases

To explore if TtpC and TptC were catalytically active *in vitro*, enzyme assays were conducted with the tryptophan derived α-dimethyl-β-keto methyl ester **56e** as it was available from previous experiments and it was the most similar to the proposed L-tryptophan containing natural tryptopeptin A substrate.

Previously optimised assay conditions were used, containing substrate **56e** (1 mM), PLE (5 μM), epoxyketone synthase (50 μM), and FAD (2 mM) and were incubated at 30 °C for 3 hours (Chapter 2.3.2). The assays were quenched with twice the volume of acetonitrile + 0.1% formic acid, centrifuged, and the supernatant was analysed for expected compounds by LCMS (Figure 4.26A).
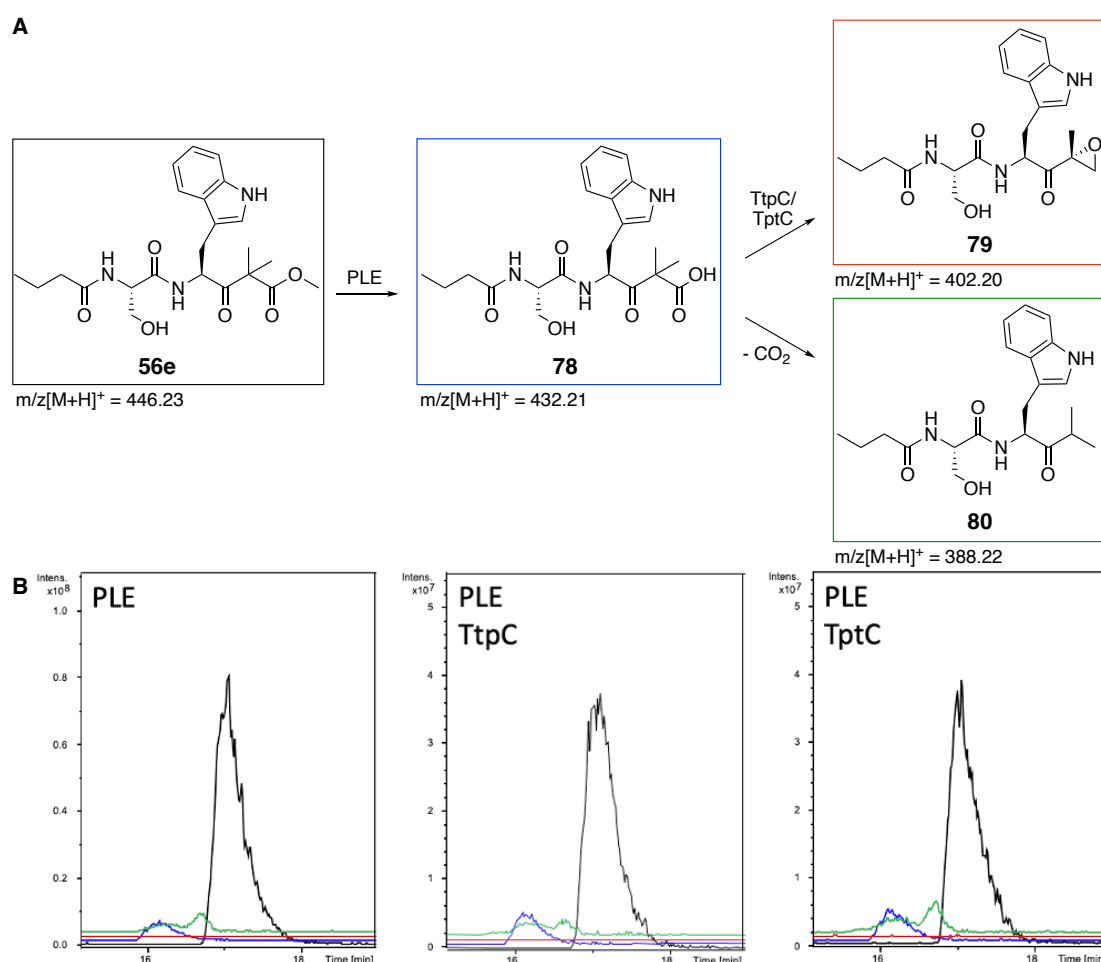
*Figure 4.26 Coupled enzyme assays with tryptophan derived α-dimethyl-β-keto methyl ester **56e**, PLE, EpnF homologues, and FAD. A. Expected reactions in enzyme assays with pig liver esterase (PLE), TtpC/TptC, and tryptophan derived substrate analogue **56e**. PLE is expected to cleave the methyl ester to the corresponding α-dimethyl-β-keto carboxylic acid **78** which will be converted to epoxyketone **79** by potential epoxyketone synthase TtpC/TptC or spontaneously decarboxylate to side product **80**. m/z of the [M+H]⁺ ion of each species are shown. B. Extracted ion chromatograms (EICs) from LCMS analysis of the reaction of tryptophan derived α-dimethyl-β-keto methyl ester **56e** with EpnF and PLE.*

101

A control reaction of **56e** and PLE showed that PLE converted the methyl ester of the substrate into the desired acid. However, no further products were observed upon incubation with TtpC or TptC suggesting the epoxyketone synthases were not active towards the tryptophan derived substrate under the tested conditions (Figure 4.26B). Since there was no tryptopeptin A production observed in *S. sparsogenes* it is possible that TtpC is just not functional. However, because *S.* sp. maeda85 did produce tryptopeptin A it is assumed that TptC is a functional enzyme *in vivo* and able to form epoxyketones. It could be that either that TptC is highly substrate specific and does not accept α-dimethyl substrates or substrates different from the native substrate in general. It is also possible that the enzyme is not stable enough under the tested conditions or that it is not functional *in vitro* in general.

4.7 Conclusions

In summary, bioinformatic analyses with clusterTools, antiSMASH, and A domain sequence alignments have revealed a biosynthetic gene cluster in *Streptomyces sparsogenes* ATCC 25498 and *S. bingchenggensis* BCW-1 that was predicted to direct the production of tryptopeptin A, an epoxyketone natural product with an aromatic amino acid next to the epoxyketone warhead.

Streptomyces sparsogenes ATCC 25498 was retrieved from the German Collection of Microorganisms and Cell Cultures, cultured, and analysed for tryptopeptin A metabolite production. *S. sparsogenes* did not produce tryptopeptin A in liquid or solid media after 5 or 10 days of growth and no tryptopeptin A related metabolites were found. It is possible that either the gene cluster expression was not triggered by the tested growth conditions, that there are inherent expression difficulties for the found cluster, or that some expressed enzymes are not functional or otherwise inhibited.

Additionally, the confirmed tryptopeptin A producer *Streptomyces* species maeda85 was obtained and cultured. Small amounts of tryptopeptin A were found in ethyl acetate extracts from *S.* sp. maeda85 grown for 5 days on solid media. HRMS analysis of the extracts confirmed the chemical formula and expected fragmentation patterns for tryptopeptin A. Three predicted metabolites related to tryptopeptin A were found in the extracts as well. These results suggest that the identified biosynthetic gene cluster in *S.* sp. maeda85 is expressed and that the relevant proteins for tryptopeptin A production were functional. However, since other tryptopeptin related metabolites without the epoxyketone pharmacophore were detected, the identified cluster does not seem to direct the tryptopeptin A production exclusively and there might be issues with the expression or functionality of the corresponding epoxyketone synthase.

To obtain more complete and higher accuracy genomic data for both *Streptomyces* strains of interest, *S. sparsogenes* and *S.* sp. maeda85 were re-sequenced with Oxford MinION nanopore sequencing technologies. High molecular weight genomic DNA from both *Streptomyces* strains was isolated and both genomes were successfully sequenced, assembled, and annotated. Full genome coverage was achieved for *S. sparsogenes* which provided an improvement to the previously published Illumina sequencing data. Analysis of the whole genome with antiSMASH identified the

potential tryptopeptin A gene cluster along with 43 other predicted biosynthetic gene clusters, which are more than the previously reported 41 gene clusters identified from the Illumina sequencing data. For *S.* sp. maeda85, full genome coverage was achieved as well and the obtained nanopore data was polished with Dr Tsunematsu's Illumina sequencing data. Genome analysis with antiSMASH identified 55 biosynthetic gene clusters of various types, the potential tryptopeptin A cluster being one of them.

The two novel epoxyketone synthases TtpC and TptC from *S. sparsogenes* and *S.* sp. maeda85 respectively were overproduced, purified, and tested in coupled enzyme assays with pig liver esterase (PLE) and the tryptophan derived α-dimethyl-β-keto methyl ester substrate. While PLE hydrolysed the methyl ester substrate derivative to the desired α-dimethyl-β-keto carboxylic acid, no epoxyketone production was observed under the tested conditions. Since *S.* sp. maeda85 produced small amounts of tryptopeptin A, TptC is expected to be active *in vivo*, however, the activity could not be reproduced *in vitro*. It is possible that the enzyme needs *in vivo* conditions to be stabilised in order to be active or that it is just not active under the tested conditions. Screening different enzyme assay conditions could reveal if there were any specific *in vitro* conditions that would allow TptC to produce epoxyketones. Additionally, structural studies and directed evolution experiments with the obtained epoxyketone synthases might optimise their stability and general *in vitro* activities.

5. Conclusion and future work

5.1 Synthesis of α-dimethyl-β-keto methyl ester substrate analogues

A small library of novel α-dimethyl-β-keto methyl ester substrate analogues was synthesised following published strategy by Zabala *et al.* and the synthetic route was optimised by using NMI-TCFH coupling conditions in the final peptide bond formation.[97] Alanine, leucine, cyclopentenylalanine, phenylalanine, tryptophan, protected serine, protected glutamic acid, and protected lysine derived α-dimethyl-β-keto methyl ester analogues were produced and fully characterised. While the side chain deprotection reactions of the later three analogues were unsuccessful due to unwanted intramolecular side reactions, the remaining analogues were available as substrates for enzyme assays in their diastereomeric form. Additionally, an α-dimethyl-β-keto methyl ester precursor of the proteasome inhibitor oprozomib was synthesised and the *N*-Boc phenylalanine α-dimethyl-β-keto methyl ester intermediate was prepared to serve as enzyme substrate as well providing overall seven α-dimethyl-β-keto methyl ester substrates analogues to be tested with epoxyketone synthases (Figure 5.1).
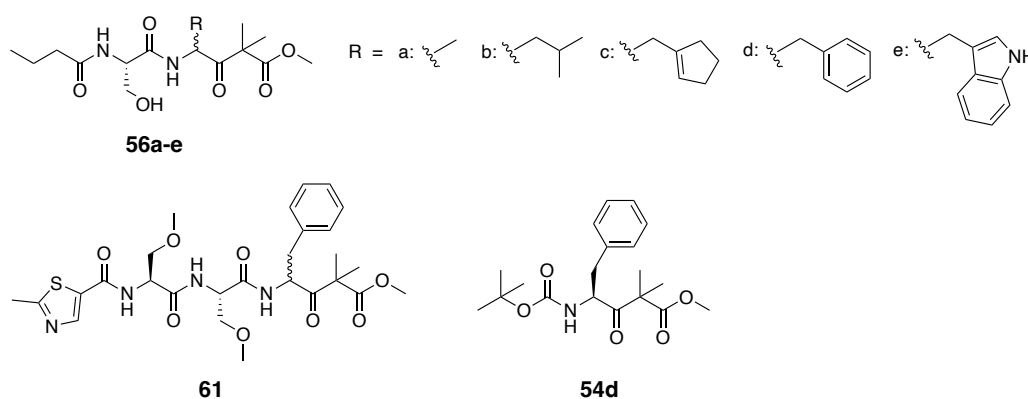
Figure 5.1 Synthetic substrate analogues for coupled enzyme assays.

5.2 *In vitro* coupled enzyme assays condition optimisation

The epoxyketone synthase EpnF was overproduced and purified to yield 20.5 mg/L of the recombinant protein for *in vitro* experiments. Pig liver esterase (PLE) was used to generate the α-dimethyl-β-keto carboxylic acid substrate in *situ* from the synthetic methyl ester analogues in coupled enzyme assays with EpnF to test the epoxyketone synthase's substrate tolerance. Initially, EpnF was probed in coupled enzyme assays with the commercially relevant phenylalanine derived substrate analogue which revealed low turnover to the epoxyketone and therefore was followed by a series condition optimisation experiments. A temperature of 30 ºC provided the highest observed epoxyketone formation while PLE was slowed down compared to 37 ºC which produced lower amounts of decarboxylated side product. EpnF appeared to be inactive after 3 hours which is why all future assays were incubated for this time and final concentrations of 5 μM PLE and 50 μM EpnF provided sufficient substrate generation by PLE and efficient enzyme usage.

While it was possible to slightly optimise the assay conditions, enzyme stability and low epoxyketone formation remained challenging. The enzyme stability and duration of activity could be optimised by testing further assay conditions such as different buffers or a range of pH values. Although a 2-fold excess of additional cofactor compared to substrate was used in the assays, it would be possible to investigate the option of using regenerative cascades within the coupled enzymes to improve catalytic activity of EpnF.

Using the purified enzyme makes the assays quite unsustainable due to the laborious and material consuming purification process. Now that the enzyme's activity is probed and characterised, it would be an option to test if desired activity is observed when incubating substrates with the lysate instead of the purified enzyme which would make the whole process more efficient. If lysate was a suitable replacement for the purified enzyme in the assays, investigations towards freeze drying the lysate and using suspensions thereof would further improve the economy of coupled enzyme assays.

5.3 EpnF forms novel epoxyketones from non-native substrates

Optimised enzyme conditions were used to probe synthetic substrate analogues with EpnF in coupled enzyme assays with PLE. PLE successfully converted all α-dimethyl-β-keto methyl ester substrate derivatives into the desired α-dimethyl-β-keto carboxylic acids *in situ*. EpnF catalysed epoxyketone formation for the alanine, leucine, cyclopentenylalanine, and phenylalanine derived substrates while the highest turnover was observed for the native leucine and sterically less hindered alanine derived substrate. The tryptophan derived substrate was not converted into the desired epoxyketone which leads us to believe that the active site of EpnF is sterically hindered for the larger aromatic residue (Table 5.1).

Table 5.1 Overview of substrate tolerance of EpnF tested in coupled enzyme assays with PLE.
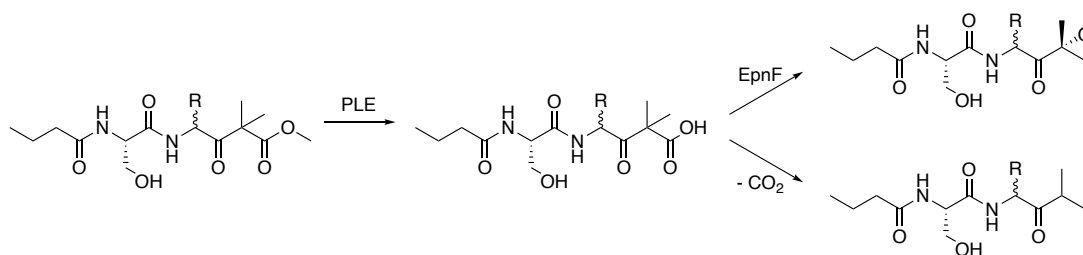
Enzyme	Substrate	Activity
EpnF	R =	✓
EpnF	R =	✓
EpnF	R =	✓
EpnF	R =	✓
EpnF	R =	✗
EpnF		✗
EpnF		✗

The α-dimethyl-β-keto methyl ester precursor for oprozomib was not tolerated by EpnF suggesting that EpnF in its native form is not suitable for chemoenzymatic production of this commercially relevant proteasome inhibitor. To gain further insight

into reasons for rejection of the sterically more hindered substrate, EpnF was also tested with the *N*-Boc phenylalanine α-dimethyl-β-keto methyl ester fragment. This shorter substrate was not tolerated by EpnF either which suggests that the hydroxyl group of the serine residue in the other α-dimethyl-β-keto methyl ester substrates is crucial for interactions with the active site since this group was missing from the short fragment and is methylated in the oprozomib precursor. It is also possible that the butyric acid serine residue is otherwise essential for interactions with the active site.

It was demonstrated that EpnF is able to convert some non-native substrates into novel epoxyketones, however, the substrate tolerance is limited. To improve the epoxyketone synthase activity to a broader substrate range or specific commercially relevant substrates, EpnF would have to be altered which could be achieved through enzyme engineering. Structural understanding of the active site would provide insight into which residues could be changed for improved substrate tolerance and would provide a starting point for mutations of active site residues. Since no crystal structures are available for any epoxyketone synthases, preliminary structural models of EpnF were generated with AlphaFold.[133] While the homology to ACAD enzymes became very clear, the area which most likely is the active site was predicted with low confidence meaning the structural arrangement of active site residues might differ from the model. While it is very useful to know which residues might be involved in the active site, the low confidence significantly limits the options for designed residue mutations in the active site.

Another option to improve the enzymatic activity would be engineering of EpnF through directed evolution.[134] A large library of mutants could be produced by error prone PCR and high throughput probing of the generated mutants with for example the oprozomib precursor could search for an optimised candidate for chemoenzymatic proteasome inhibitor production.

5.4 *In silico* identification and analysis of novel putative tryptopeptin A directing biosynthetic gene clusters

To investigate epoxyketone synthases with specificity towards aromatic amino acids next to the formed epoxyketone warhead, a bioinformatic approach was developed for *in silico* identification and analysis of putative tryptopeptin A biosynthetic gene clusters. Novel epoxyketone clusters were identified with clusterTools and prediction of A domain specificities identified clusters likely to incorporate the valine, threonine, tryptophan sequence expected for tryptopeptin A production. A cluster of interest in *Streptomyces sparsogenes* ATCC 25498 was further analysed with antiSMASH and BLAST to characterise the cluster *in silico* and predict the functionality of the encoded proteins. The putative tryptopeptin A cluster contained a proteasome β-subunit which is conserved across epoxyketone directing clusters alongside with the ACAD homologue potential epoxyketone synthase. The relevant *Streptomyces sparsogenes* strain was ordered and cultivated to further investigate the potential tryptopeptin A cluster.

5.5 Metabolic profiling and full genome sequencing of *Streptomyces sparsogenes* ATCC 25498 and *Streptomyces* sp. maeda85

5.5.1 Metabolite expression, extraction, and analysis

Streptomyces sparsogenes ATCC 25498 was grown in liquid medium and on plates for 5 and 10 days to investigate metabolite production, however, no tryptopeptin A was observed. Heterologous expression of the identified cluster in *Streptomyces coelicolor* achieved by Challis group member Dr Huang revealed production of tryptopeptin A derived metabolites which were absent in the metabolite extracts of the *Streptomyces* strain. This suggests that the putative tryptopeptin A cluster was not expressed under the tested conditions or that enzymes in the beginning of the biosynthetic assembly were misfolded, truncated, or otherwise unfunctional *in vivo*. For comparison with a known tryptopeptin A biosynthetic gene cluster and the option to investigate a functional novel epoxyketone synthase, the confirmed tryptopeptin A producer *Streptomyces* sp. maeda85 was kindly provided by Dr Yuta Tsunematsu, author or the first tryptopeptin A isolation publication. Small amounts of tryptopeptin A were observed in extracts of *S.* sp. maeda85 plate cultures and tryptopeptin A related metabolites identified by Dr Huang were produced as well. Heterologous expression of the expected tryptopeptin A cluster from *S.* sp. maeda85 in *S. coelicolor* performed by Dr Huang revealed that this known tryptopeptin A directing cluster and the putative tryptopeptin A cluster from *S. sparsogenes* exhibited very similar metabolic profiles including identical tryptopeptin A related compounds which suggests that both clusters direct the production of the same metabolites. Because no tryptopeptin A was observed in neither of experiments with the heterologously expressed clusters it is expected that the corresponding epoxyketone synthases are accompanied by difficulties in expression, stability, activity, or other factors that could lead to low- or non-functional enzymes.

To investigate if the putative epoxyketone synthases are properly expressed under the tested conditions, proteomic analysis could be attempted and might provide insight into potential issues with these enzymes. Additionally, different expression conditions or cloning the clusters of interest into alternative strains would be an option to further investigate if tryptopeptin A production could be triggered in different circumstances.

5.5.2 Full genome sequencing, assembly, annotation, and comparison

The available Illumina data of both relevant *Streptomyces* strains was split between hundreds of contigs which is why it was decided to re-sequenced both genomes with Oxford MinION nanopore sequencing technologies. High molecular weight genomic DNA was isolated and full genome sequencing was accomplished for *S. sparsogenes* and *S.* sp. maeda85. Genome assembly and annotation of the generated nanopore data resulted in BUSCO scores of 95.2% and 98.8% respectively. The *S.* sp. maeda85 nanopore data was additionally polished with the Illumina sequencing data provided by Dr Tsunematsu to improve the BUSCO score to 99.2%. The genomes of *S. sparsogenes* and *S.* sp. maeda85 were analysed with antiSMASH and predicted to harbour 44 and 52 biosynthetic gene clusters respectively. Alongside the putative tryptopeptin A cluster in both strains, several clusters were predicted to produce known metabolites such as geosmin and pristinol. Four biosynthetic gene clusters in both strains were highly homologous and predicted to direct the production of novel metabolites and numerous other different clusters were predicted in both strains while *S.* sp. maeda85 appeared to harbour significantly more predicted PKS clusters than *S. sparsogenes*.

5.6 Overproduction, purification, and *in vitro* testing of novel epoxyketone synthases

The novel epoxyketone synthases encoding genes *ttpC* and *tptC* were successfully amplified from genomic DNA from *S. sparsogenes* and *S.* sp. maeda85 respectively and cloned into pET151 expression vectors. Potential epoxyketone synthases TtpC and TptC were overproduced and purified, however neither of them catalysed epoxyketone formation in coupled assays with PLE and the tryptophan derived α-dimethyl-β-keto methyl ester substrate analogue **56e**. It is possible that the novel epoxyketone synthases are very substrate specific and would only tolerate a tryptopeptin A precursor **81** or at least only substrates more similar to the native substrates such as α-monomethyl-β-keto carboxylic acids (Figure 5.2). Additionally, it might be that these enzymes are just inherently unstable under the tested conditions or *in vitro* in general. Since tryptopeptin A production was observed in *S.* sp. maeda85, TptC appeared to be active *in vivo* in the natural host strain. However, because there was no epoxyketone synthase activity detected in heterologous expression experiments, difficulties with the functionality of the novel epoxyketone synthases were already expected.
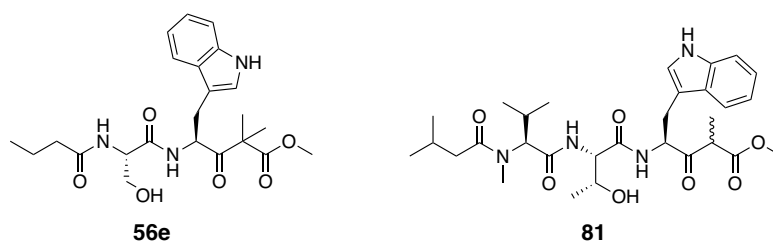
*Figure 5.2 Structures of tryptophan derived substrate analogues **56e** and tryptopeptin A substrate **81**.*

Synthesising the tryptopeptin A precursor **81** and testing it in *in vitro* assays with the novel epoxyketone synthases would reveal if these enzymes were highly substrate specific. Probing alternative overproduction and purification methods or different *in vitro* conditions would be an option to investigate if *in vitro* activity of the putative epoxyketone synthases can be achieved under different circumstances.

5.7 Concluding remarks

Overall, these studies provided insight into the substrate tolerance of known epoxyketone synthase EpnF and an *in silico* approach to identify novel epoxyketone synthase biosynthetic gene clusters was developed. While the chemoenzymatic production of commercially relevant epoxyketone proteasome inhibitors was unsuccessful, it was crucial to further investigate known and novel epoxyketone synthases for more insight into this class of enzymes.

In order to proceed towards more sustainable production options of epoxyketone proteasome inhibitors, EpnF or the identified novel epoxyketone synthases could serve as lead enzymes for directed evolution experiments towards functional biocatalysts. Additionally, with the increasing development of biosynthetic tools and methods, *de novo* construction of bacterial pathways including epoxyketone synthases for natural product derived proteasome inhibitors could be of interest in the future to aim for epoxyketone proteasome inhibitor production through fermentation entirely.

6. Materials and methods

6.1 Chemistry

6.1.1 Materials, instruments, and general conditions

All commercially available chemicals, reagents, and solvents were purchased from Acros Organic, Fischer Scientific, Sigma Aldrich or VWR and used without further purification. All aqueous solutions were prepared with deionised water. Room temperature (rt) refers to ambient temperatures of 20-22 °C, 0 °C refers to an ice water bath, and heating of experiments was achieved with thermostatically controlled oil baths. Thin-layer chromatography analysis (TLC) was conducted using aluminium-backed plates pre-coated with Merck silica gel 60 F254 and compounds were visualised using UV radiation and potassium permanganate stain (2 g $KMnO_4$, 10 g K_2CO_3, 1.25 mL 2.5 M NaOH in 200 mL water). Concentration of samples in *vacuo* was achieved by evaporation of solvents on a BUCHI Rotavapor R-200 or R-210 connected to a BUCHI Vacuum Pump V-700. Silica column chromatography was performed on 40-63 μM, 40-60 Å silica gel (Sigma Aldrich).

NMR spectra were recorded on Bruker Advance AV-300 and HD-400 MHz spectrometers. Chemical shifts are reported in parts per million (ppm) referenced from $CDCl_3$ (δH: 7.26 ppm and δC: 77.2 ppm) or MeOD (δH: 3.31 ppm and δC: 49.0 ppm) and assignments were made with the aid of COSY, HMBC, and HSQC spectra. Coupling constants (J) are rounded to the nearest 0.5 Hertz (Hz) and multiplicities are given as multiplet (m), singlet (s), doublet (d), triplet (t), quartet (q), quintet (quin.), sextet (sext.), or combinations thereof. Infrared analyses were performed using an Alpha Bruker Platunium ATR single reflection diamond ATR module. Optical rotations were obtained using an Optical Activity Ltd AA-1000 millidigree auto-ranging polarimeter (589 nm) and specific rotations are given in units of 10^{-1} deg cm^2 g^{-1}. Low resolution mass spectra (LRMS) were recorded using an Agilent 6130B single Quad (ESI). High resolution mass spectra (HRMS) were obtained were obtained by either Dr Lijiang Song or Lynette Walsh on a Bruker MaXis ESI spectrometer. Samples from enzyme assays were analysed on a C18 column using a

Bruker amaZon speed ETD instrument and a gradient from 5-100% acetonitrile in water each supplemented with 0.1% formic acid at a flow rate of 0.2 mL/min.

6.1.2 Chemical synthesis

6.1.2.1 Methyl butyryl-L-serine (46)
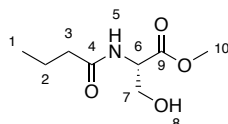
To a solution of L-serine methyl ester hydrochloride (**45**) (2.01 g, 12.9 mmol) in DCM (50 mL) was added Et_3N (2.15 mL, 15.4 mmol, 1.2 equiv.) at 0 °C under argon. After 10 minutes EDC (4.93 g, 25.7 mmol, 2.0 equiv.) and butyric acid (1.4 mL, 15 mmol, 2.0 equiv.) was added, the reaction mixture was warmed up to room temperature and left to stir overnight. The reaction was quenched by addition of 1 M aqueous HCl (10 mL), the layers were separated, the aqueous layer was extracted with EtOAc (3 × 10 mL), and the combined organic layers were dried with $MgSO_4$, filtered, and concentrated in *vacuo*. The resulting residue was purified by silica gel chromatography (EtOAc) to give the desired product **46** as a colourless oil (2.08 g, 11.0 mmol, 86%).

R_f = 0.41 (EtOAc, UV/$KMnO_4$).

^{1}H NMR (400 MHz, $CDCl_3$): δ 0.97 (3H, t, J = 7.5 Hz, H1), 1.70 (2H, sixt., J = 7.5 Hz, H2), 2.26 (2H, t, J = 7.5 Hz, H3), 3.80 (3H, s, H10), 3.95 (1H, dd, J = 9.5, 4.0 Hz, H7), 3.97 (1H, dd, J = 4.0, 9.5 Hz, H7), 4.70 (1H, quin., J = 3.5 Hz, H6), 6.36 (1H, br s, H5).

^{13}C NMR (101 MHz, $CDCl_3$): δ 14.0 (C1), 19.3 (C2), 38.7 (C3), 53.2 (C10), 55.0 (C6), 64.1 (C7), 171.2 (C9), 173.5 (C4).

IR: ν 3308, 3078, 2959, 2875, 2121, 1739, 1646, 1533, 1460, 1437, 1370, 1337, 1206, 1147, 1079, 983, 855, 806.

HRMS (ESI) calculated for $C_8H_{15}NNaO_4$ ([M+Na]$^+$): 212.0893; found: 212.0896.

$[\alpha]_D^{20}$ = + 26.0 (c 1.14, $CHCl_3$).

6.1.2.2 Butyryl-L-serine (47)
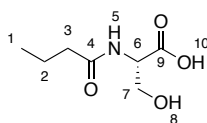
To a solution of **46** (1.99 g, 11.0 mmol) in THF (50 mL) was added 30% (w/w) H_2O_2 (0.25 mL, 2.4 mmol, 0.22 equiv.) and aqueous LiOH (43.5 mL, 0.76 M, 33.1 mmol, 3.0 equiv.) dropwise. After 2 hours, THF was removed in *vacuo* and the remaining residue was diluted with H_2O (15 mL). The aqueous layer was washed with diethyl ether (2 × 20 mL), acidified to pH 1 with 1 M aqueous HCl (10 mL), and sodium chloride was added (8 g). The aqueous layer was extracted with EtOAc (3 × 50 mL), and the combined organic layers dried with $MgSO_4$, filtered, and concentrated in *vacuo* to give the desired product **47** as a clear oil (0.5283 g, 3.033 mmol, 28%).

R_f = 0.46 (EtOAc, UV/KMnO$_4$).

^{1}H NMR (400 MHz, MeOD): δ 0.97 (3H, t, J = 7.5 Hz, H1), 1.66 (2H, sixt., J = 7.5 Hz, H2), 2.27 (2H, t, J = 7.5 Hz, H3), 3.84 (1H, dd, J = 9.5, 4.0 Hz, H7), 3.86 (1H, dd, J = 9.5, 4.0 Hz, H7), 4.49 (1H, t, J = 3.5 Hz, H6).

^{13}C NMR (101 MHz, MeOD): δ 14.0 (C1), 20.2 (C2), 38.8 (C3), 56.1 (C6), 63.0 (C7), 165.3 (C9), 175.9 (C4).

IR: ν 3314, 3091, 2964, 2936, 2876, 1721, 1627, 1536, 1462, 1416, 1375, 1205, 1143, 1075, 1046, 902.

HRMS (ESI) calculated for $C_7H_{13}NNaO_4$ ([M+Na]$^+$): 198.0737; found: 198.0741.

$[\alpha]_D^{20}$ = + 8.2 (*c* 0.516, MeOH).

6.1.2.3 Methyl (S)-4-((tert-butoxycarbonyl)amino)-3-oxo-5-phenylpentanoate (53d)
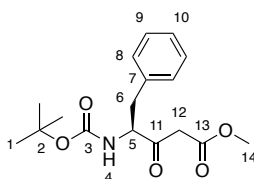
To a solution *N*-Boc-L-phenylalanine (**52d**) (2.00 g, 7.54 mmol) in THF (40 mL) was added CDI (1.83 g, 11.3 mmol, 1.5 equiv.) under argon. After stirring the reaction mixture at room temperature for 2 hours, methyl potassium malonate (1.70 g, 11.3 mmol, 1.5 equiv.) and $MgCl_2$ (1.08 g, 11.3 mmol, 1.5 equiv.) were added and the reaction was stirred overnight. The solvent was removed in *vacuo* and the crude residue was partitioned between EtOAc (40 mL) and 1 M aqueous HCl (40 mL). The organic layer was collected, and the aqueous phase was extracted with EtOAc (3 × 40 mL). Organic extractions were combined, washed with 5% $NaHCO_3$ (w/w in H_2O, 2 × 40 mL) and brine (40 mL), dried with $MgSO_4$ and filtered. The solvent was removed in *vacuo* to provide the desired product **53d** as white crystals (2.01 g, 6.26 mmol, 83%).

R_f = 0.62 (EtOAc/hexane, 1:1, UV/$KMnO_4$).

^{1}H NMR (400 MHz, $CDCl_3$): δ 1.40 (9H, s, H1), 2.97 (1H, dd, J = 13.5, 7.5 Hz, H6), 3.14 (1H, dd, J = 6.0, 14.0 Hz, H6), 3.48 (2H, m, H12), 3.71 (3H, s, H14), 4.56 (1H, dt, J = 7.0, 13.5 Hz, H5), 4.99 (1H, br s, H4), 7.17 (2H, d, J = 7.0 Hz, H8), 7.25 (1H, d, J = 7.0 Hz, H10), 7.30 (2H, dd, J = 7.0, 14.0 Hz, H9).

^{13}C NMR (101 MHz, $CDCl_3$): δ 28.4 (C1), 37.1 (C6), 46.8 (C12), 52.6 (C14), 60.6 (C5), 81.2 (C2), 127.2 (C10), 128.9 (C9), 129.4 (C8), 136.2 (C7), 162.5 (C3),167.7 (C13), 202.2 (C11).

IR: ν 3362, 3027, 3000, 2971, 2954, 2936, 2903, 1750, 1719, 1687, 1604, 1515, 1454, 1440, 1402, 1368, 1312, 1264, 1248, 1203, 1164, 1139, 1095, 1074, 1048, 1028, 994, 954, 853.

HRMS (ESI) calculated for $C_{17}H_{23}NNaO_5$ ([M+Na]$^+$): 344.1468; found: 344.1468.

$[\alpha]_D^{20}$ = - 4.6 (*c* 0.658, $CHCl_3$).
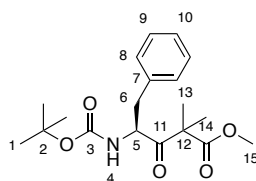
To a solution of **53d** (1.97 g, 6.13 mmol) in THF (30 mL), K_2CO_3 (8.47 g, 61.3 mmol, 10 equiv.) and methyl iodide (1.9 mL, 31 mmol, 5.0 equiv.) were added under argon. The reaction was stirred overnight and more K_2CO_3 (4.23 g, 30.6 mmol, 5.0 equiv.) and methyl iodide (0.95 mL, 15.32 mmol, 2.5 equiv.) were added. The reaction was stirred overnight, quenched with H_2O (20 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 25 mL) and the combined organic layers were washed with brine (3 × 25 mL), dried with $MgSO_4$, filtered, and concentrated in *vacuo* to give the desired product **54d** as a clear oil (1.95 g, 5.58 mmol 91%).

R_f = 0.71 (EtOAc/hexane, 1:1, UV/$KMnO_4$).

^1H NMR (400 MHz, $CDCl_3$): δ 1.32 (3H, s, H13), 1.34 (9H, s, H1), 1.35 (3H, s, H13), 2.78 (1H, dd, *J* = 8.0, 14.0 Hz, H6), 3.09 (1H, dd, *J* = 5.5, 14.0 Hz, H6), 3.66 (3H, s, H15), 4.73 (1H, d br, *J* = 10.0 Hz, H4), 4.85 (1H, dt, *J* = 8.0, 6.0 Hz, H5), 7.15 (2H, d, *J* = 7.0 Hz, H8), 7.22 (1H, d, *J* = 7.0 Hz, H10), 7.28 (2H, dd, *J* = 7.0, 7.0 Hz, H9).

^{13}C NMR (101 MHz, $CDCl_3$): δ 21.8 (C13), 22.3 (C13), 28.3 (C1), 38.7 (C6), 52.7 (C15), 55.0 (C2), 56.5 (C5), 80.1 (c12), 126.9 (C10), 128.6 (C9), 129.7 (C8), 136.6 (C7), 173.6 (C14), 207.6 (C3).

IR: ν 3407, 3377, 2979, 2951, 2933, 1739, 1706, 1605, 1500, 1454, 1438, 1390, 1366, 1310, 1246, 1149, 1093, 1068, 1048, 1025, 987, 957, 933, 875, 854, 842.

HRMS (ESI) calculated for $C_{19}H_{27}NNaO_5$ ([M+Na]$^+$): 372.1781; found: 372.1785.

$[\alpha]_D^{20}$ = - 30.3 (*c* 0.451, $CHCl_3$).

6.1.2.5 Methyl (S)-4-amino-2,2-dimethyl-3-oxo-5-phenylpentanoate hydrochloride (55d)
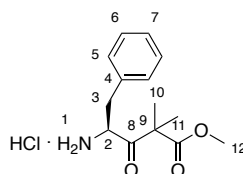
To **54d** (1.95 g, 5.81 mmol) was added 4 M HCl in dioxane (40 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 2 hours to provide the product **55d** as a colourless oil (1.63 g, 5.70 mmol, 98%).

R_f = 0.51 (EtOAc, UV/KMnO$_4$).

^{1}H NMR (400 MHz, MeOD): δ 1.39 (3H, s, H10), 1.49 (3H, s, H10), 2.83 (1H, dd, J = 9.5, 14.5 Hz, H3), 3.34 (1H, d, J = 4.0 Hz, H3), 3.79 (3H, s, H12), 4.75 (1H, dd, J = 7.0, 9.5 Hz, H2), 7.29 (2H, d, J = 7.0, H5), 7.35 (1H, d, J = 7.0 Hz, H6), 7.39 (2H, d, J = 7.0 Hz, H7).

^{13}C NMR (101 MHz, MeOD): δ 22.5 (C10), 22.6 (C10), 37.8 (C3), 53.5 (C12), 56.2 (C9), 57.8 (C2), 129.2 (C7), 130.4 (C6), 130.5 (C5), 134.4 (C4), 174.8 (C11), 205.5 (C8).

IR: ν 3229, 3009, 2981, 2845, 2768, 2622, 2549, 2430, 1951, 1713, 1489, 1454, 1435, 1241, 1193, 1183, 1153, 1114, 1028, 1001, 984, 945, 864.

HRMS (ESI) calculated for C$_{14}$H$_{20}$NO$_3$ ([M+H]$^+$): 250.1438; found: 250.1446.

6.1.2.6 (4S)-4-(2-butyramido-3-hydroxypropanamido)-2,2-dimethyl-3-oxo-5-phenylpentanoate (56d)
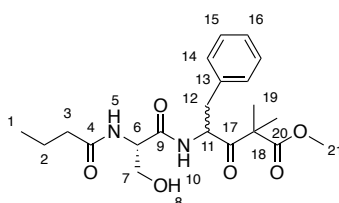
Conditions 1: To a solution of **47** (0.137 g, 0.803 mmol) in DMF (5 mL) was added HOBt (0.224 g, 1.61 mmol, 2.0 equiv.), EDC (0.308 g, 1.61 mmol, 2.0 equiv.) and Et$_3$N (0.09 mL, 1.3 mmol, 1.5 equiv.) and the solution was stirred for 1 hour. Amine

119

55d (0.205 g, 0.823 mmol) in DMF (5 mL) with Et$_3$N (0.09 mL, 1.3 mmol, 1.5 equiv.) was added and the reaction stirred for 3 days. The reaction mixture was quenched with 1 M aqueous HCl (10 mL) and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with 1 M aqueous LiCl (40 mL), 1 M aqueous HCl (40 mL), 1 M aqueous NaHCO$_3$ (40 mL) and brine (3 × 40 mL). The organic layer was dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56d** as a yellow oil (0.0519 g, 0.126 mmol, 16%).

Conditions 2: To a solution of **47** (0.0512 g, 0.294 mmol) in THF (3 mL) was added NMI (0.07 mL, 0.8 mmol, 3.0 equiv.) and TCFH (0.107 g, 0.382 mmol, 1.3 equiv.) and the solution was stirred for 15 minutes. Amine **55d** (0.120 g, 0.295 mmol, 1.0 equiv.) was added and the reaction stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (10 mL), THF concentrated in *vacuo* and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (3 × 40 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56d** as a yellow oil (0.0638 g, 0.157 mmol, 53%).

Conditions 3: To a solution of **47** (0.0439 g, 0.252 mmol) in THF (5 mL) was added EDC (0.0792 g, 0.413 mmol, 1.6 equiv.), oxyma (0.0589 g, 0.414 mmol, 1.6 equiv.) and Et$_3$N (0.1 mL, 0.72 mmol, 2.9 equiv.) and the solution was stirred for 15 minutes. Amine **55d** (0.108 g, 0.264 mmol, 1.0 equiv.) was added and the reaction stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (10 mL), THF concentrated in vacuo and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (3 × 40 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56d** as a yellow oil (0.0114 g, 0.0281 mmol, 11%).

Conditions 4: To a solution of **47** (0.0467 g, 0.268 mmol) in THF (5 mL) was added COMU (0.167 g, 0.390 mmol, 1.5 equiv.) and Et$_3$N (0.10 mL, 0.717 mmol, 2.9 equiv.) and the solution was stirred for 15 minutes. Amine **55d** (0.107 g, 0.163 mmol, 1.0 equiv.) was added and the reaction stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (10 mL), THF concentrated in vacuo

and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (3 × 40 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56d** as a yellow oil (0.0474 g, 0.117 mmol, 44%).

R$_f$ = 0.25 (MeOH/DCM, 5:95, UV/KMnO$_4$).

Diastereomer 1: ^{1}H NMR (400 MHz, MeOD): δ 0.93 (3H, t, J = 7.5 Hz, H1), 1.27 (3H, s, H19), 1.32 (3H, s, H19), 1.62 (2H, sixt., J = 7.5 Hz, H2), 2.20 (2H, t, J = 4.5 Hz, H3), 2.78 (1H, dd, J = 14.0, 8.5 Hz, H12), 3.09 (1H, dd, J = 14.0, 5.5 Hz, H12), 3.49 (1H, d, J = 6.0 Hz, H7), 3.63 (3H, s, H21), 3.68 (1H, d, J = 4.0 Hz, H7), 4.35 (1H, br t, J = 6.0 Hz, H6), 5.11 (1H, br t, J = 5.5 Hz, H11), 7.23 (5H, m, H14-16).

^{13}C NMR (101 MHz, MeOD): δ 14.0 (C1), 20.2 (C2), 22.1 (C19). 22.3 (C19), 38.7 (C3), 38.9 (C12), 53.2 (C21), 56.1 (C11), 56.4 (C6), 56.4 (C18), 62.8 (C7), 127.8 (C16), 129.4 (C15), 130.6 (C14), 138.2 (C13) 171.6 (C9), 174.7 (C20), 176.1 (C4), 207.3 (C17).

Diastereomer 2: ^{1}H NMR (400 MHz, MeOD): δ 0.93 (3H, t, J = 7.5 Hz, H1), 1.28 (3H, s, H19), 1.33 (3H, s, H19), 1.62 (2H, sixt., J = 7.5 Hz, H2), 2.21 (2H, t, J = 4.0 Hz, H3), 2.78 (1H, dd, J = 14.0, 8.5 Hz, H12), 3.09 (1H, dd, J = 14.0, 5.5 Hz, H12), 3.49 (1H, d, J = 6.0 Hz, H7), 3.64 (3H, s, H21), 3.68 (1H, d, J = 4.0 Hz H7), 4.36 (1H, br t, J = 6.0 Hz, H6), 5.13 (1H, br t, J = 5.5 Hz, H11), 7.23 (5H, m, H14-16).

^{13}C NMR (101 MHz, MeOD): δ 14.0 (C1), 20.2 (C2), 22.2 (C19), 22.4 (C19), 38.7 (C3), 38.9 (C12), 53.2 (C21), 56.2 (C11), 56.4 (C6), 56.5 (C18), 62.9 (C7), 127.8 (C16), 129.5 (C15), 130.6 (C14), 138.2 (C13), 171.7 (C9), 174.8 (C20), 176.2 (C4), 207.4 (C17).

IR: ν 3305, 3065, 2962, 2935, 2875, 1743, 1714, 1642, 1527, 1455, 1436, 1386, 1257, 1203, 1148, 1055.

HRMS (ESI) calculated for C$_{21}$H$_{30}$N$_2$NaO$_6$ ([M+Na]$^+$): 429.1996; found: 429.1999.

$[\alpha]_D^{20} = -33.5$ (*c* 0.377, MeOH).

6.1.2.7 Methyl (S)-4-((tert-butoxycarbonyl)amino)-6-methyl-3-oxoheptanoate (53b)
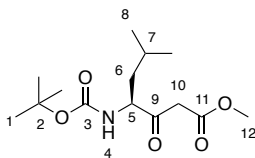
To a solution *N*-Boc-L-leucine (**52b**) (0.502 g, 2.17 mmol) in THF (12 mL) was added CDI (0.529 g, 3.26 mmol, 1.5 equiv.) under argon. After stirring the reaction mixture at room temperature for 2 hours, methyl potassium malonate (0.513 g, 3.28 mmol, 1.5 equiv.) and $MgCl_2$ (0.315 g, 3.30 mmol, 1.5 equiv.) were added and the reaction was stirred overnight. The reaction was quenched with 1 M aqueous HCl (10 mL), THF was removed in *vacuo* and the aqueous layer was extracted with EtOAc (3 × 20 mL). Organic extractions were combined, washed with 5% $NaHCO_3$ (w/w in H_2O, 2 × 40 mL) and brine (40 mL), dried with $MgSO_4$ and filtered. The solvent was removed in *vacuo* to provide the desired product **53b** as clear oil (0.449 g, 1.56 mmol, 72%).

$R_f = 0.5$ (EtOAc/hexane, 1:2, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 0.95 (6H, dd, *J* = 3.0, 6.5 Hz, H8), 1.38 (1H, m, H7), 1.44 (9H, s, H1), 1.65 (2H, m, H6), 3.57 (2H, m, H10), 3.74 (3H, s, H12), 4.34 (1H, t, *J* = 8.0 Hz, H5), 4.89 (1H, d, *J* = 7.0 Hz, H4).

^{13}C NMR (101 MHz, CDCl$_3$): δ 21.8 (C8), 23.4 (C7), 28.5 (C1), 40.0 (C6), 46.3 (C10), 52.6 (C12), 58.3 (C5), 80.5 (C2), 135.8 (C3), 167.6 (C11), 202.9 (C9).

HRMS (ESI) calculated for $C_{14}H_{25}NNaO_5$ ([M+Na]$^+$): 310.1625; found: 310.1620.

6.1.2.8 Methyl (S)-4-((tert-butoxycarbonyl)amino)-2,2,6-trimethyl-3-oxoheptanoate (54b)
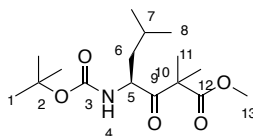
To a solution of **53b** (0.430 g, 1.50 mmol) in THF (10 mL), K_2CO_3 (2.07 g, 15.00 mmol, 10 equiv.) and methyl iodide (0.47 mL, 7.5 mmol, 5.0 equiv.) were added under an argon atmosphere. The reaction was stirred overnight and more K_2CO_3 (5.0 equiv.) and methyl iodide (2.5 equiv.) were added. The reaction was stirred overnight, quenched with H_2O (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 15 mL) and the combined organic layers were washed with brine (2 × 25 mL), dried with $MgSO_4$, filtered, and concentrated in *vacuo* to give the desired product **54b** as yellow oil (0.473 g, 1.50 mmol 100%).

R_f = 0.58 (EtOAc/hexane, 1:2, UV/$KMnO_4$).

^{1}H NMR (400 MHz, $CDCl_3$): δ 0.93 (6H, dd, J = 6.5 Hz, H8), 1.05 (1H, m, H7), 1.41 (3H, s, H11), 1.42 (12H, s, H1 and H11), 1.65 (2H, m, H6), 3.73 (3H, s, H13), 4.64 (1H, dt, J = 3.0, 10.0 Hz, H5), 4.77 (1H, d, J = 9.5 Hz, H4).

^{13}C NMR (101 MHz, $CDCl_3$): δ 21.6 (C8), 22.2 (C11), 22.5 (C11), 23.7 (C7), 28.5 (C1), 42.1 (C6), 52.8 (C13), 54.0 (C5), 54.8 (C10), 79.8 (C2), 135.9 (C3), 173.6 (C12), 208.8 (C9).

HRMS (ESI) calculated for $C_{16}H_{29}NNaO_5$ ([M+Na]$^+$): 338.1938; found: 338.1937.

6.1.2.9 Methyl (S)-4-amino-2,2,6-trimethyl-3-oxoheptanoate (55b)
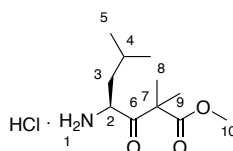
To **54b** (0.4726 g, 1.498 mmol) was added 4 M HCl in dioxane (20 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 2 hours to provide the product **55b** as white powder (0.3771 g, 1.498 mmol, 100%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 1.00 (6H, dd, J = 6.5, 12.0 Hz, H5), 1.41 (1H, m, H4), 1.46 (3H, s, H8), 1.51 (3H, s, H8), 1.82 (2H, m, H3), 3.76 (3H, s, H10), 4.47 (1H, br s, H2), 8.80 (2H, br s, H1).

^{13}C NMR (101 MHz, $CDCl_3$): δ 21.0 (C5), 23.0 (C8), 23.7 (C8), 24.0 (C4), 39.7 (C3), 53.1 (C10), 54.5 (C2), 54.7 (C7), 173.3 (C9), 202.8 (C6).

HRMS (ESI) calculated for $C_{11}H_{22}NNaO_3$ [(M-Na)$^+$]: 216.1594; found: 216.1592.

6.1.2.10   Methyl   (S)-4-((S)-2-butyramido-3-hydroxypropanamido)-2,2,6-trimethyl-3-oxoheptanoate (56b)
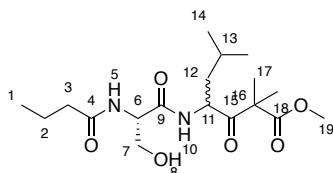
To a solution of **47** (0.122 g, 0.697 mmol) in THF (7 mL) was added HOBt (0.194 g, 1.39 mmol, 2.0 equiv.), EDC (0.297 g, 1.39 mmol, 2.0 equiv.) and Et$_3$N (0.19 mL, 1.4 mmol, 2.0 equiv.) and the solution was stirred for 1 hour. Amine **55b** (0.150 g, 0.697 mmol) in THF (2 mL) and Et$_3$N (0.10 mL, 0.70 mmol, 1.0 equiv.) was added and the reaction was stirred for 3 days. The reaction mixture was quenched with 1 M aqueous HCl (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 15 mL) and the combined organic layers were washed with brine (2 × 20 mL). The organic layer was dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (MeOH/DCM, 5:95 to 1:9) to yield the product **56b** as yellow oil (0.0261 g, 0.0702 mmol, 10%).

R$_f$ = 0.19 (MeOH/DCM, 1:99, UV/KMnO$_4$).

Diastereomer1: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.92 (6H, t, *J* = 6.0 Hz, H14), 0.96 (3H, t, *J* = 7.5 Hz, H1), 1.39 (3H, s, H17), 1.41 (3H, s, H17), 1.44 (1H, m, H12), 1.46 (1H, m, H12), 1.63 (1H, m, H13), 1.67 (2H, m, H2) 2.25 (2H, t, *J* = 7.5 Hz, H3), 3.60 (1H, dd, *J* = 4.5, 7.5 Hz, H7), 3.72 (3H, s, H19), 4.11 (1H, dd, *J* = 3.0, 11.5 Hz, H7), 4.46 (1H, m, H6), 4.92 (1H, m, H11), 6.59 (1H, br d, 7.4 Hz, H5), 7.07 (1H, br d, 9.0 Hz, H10).

^{13}C NMR (101 MHz, CDCl$_3$): δ 13.8 (C1), 19.1 (C2), 21.4 (C14), 22.5 (C17), 23.6 (C14), 25.0 (C13), 38.4 (C3), 41.4 (C12), 52.9 (C19), 53.0 (C11), 54.6 (C16), 55.0 (C6), 62.8 (C7), 173.5 (C18), 174.2 (C4), 207.2 (C15).

Diastereomer 2: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.90 (6H, t, *J* = 3.5 Hz, H14), 0.95 (3H, t, *J* = 7.5 Hz, H1), 1.40 (3H, s, H17), 1.43 (3H, s, H17), 1.44 (1H, m, H12), 1.46 (1H, m, H12), 1.63 (1H, m, H13), 1.67 (2H, m, H2), 2.21 (2H, t, *J* = 7.0 Hz, H3),

3.08 (1H, dd, 3.2, J = 8.0 Hz, H7), 3.75 (3H, s, H19), 4.06 (1H, dd, J = 3.0, 11.5 Hz, H7), .4.42 (1H, m, H6), 4.92 (1H, m, H11), 6.59 (1H, br d, 7.5 Hz, H5), 7.07 (1H, br d, 9.0 Hz, H10).

^{13}C NMR (101 MHz, CDCl$_3$): δ 13.8 (C1), 19.1 (C2), 21.3 (C14), 22.4 (C17), 23.5 (C14), 25.0 (C13), 38.4 (C3), 41.2 (C12), 52.8 (C19), 53.0 (C11), 53.6 (C16), 55.0 (C6), 62.8 (C7), 173.5 (C18), 174.2 (C4), 207.2 (C15).

HRMS (ESI) calculated for C$_{18}$H$_{32}$N$_2$NaO$_6$ ([M+Na]$^+$): 395.2153; found: 395.2147.

6.1.2.11 *Methyl (S)-4-((tert-butoxycarbonyl)amino)-3-oxopentanoate (53a)*
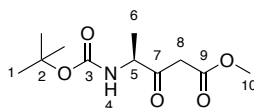
To a solution *N*-Boc-L-alanine (**52a**) (0.501 g, 2.65 mmol) in THF (12 mL) was added CDI (0.648 g, 3.99 mmol, 1.5 equiv.) under argon. After stirring the reaction mixture at room temperature for 2 hours, methyl potassium malonate (0.622 g, 3.98 mmol, 1.5 equiv.) and MgCl$_2$ (0.378 g, 3.97 mmol, 1.5 equiv.) were added and the reaction was stirred overnight. The reaction was quenched with 1 M aqueous HCl (10 mL), THF was removed in *vacuo* and the aqueous layer was extracted with EtOAc (3 × 20 mL). Organic extractions were combined, washed with 5% NaHCO$_3$ (w/w in H$_2$O, 2 × 40 mL) and brine (40 mL), dried with MgSO$_4$ and filtered. The solvent was removed in *vacuo* to provide the desired product **53a** as clear oil (0.593 g, 2.42 mmol, 92%).

R$_f$ = 0.42 (EtOAc/hexane, 1:2, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.36 (3H, d, J = 7.0 Hz, H6), 1.45 (9H, s, H1), 3.57 (2H, m, H8), 3.75 (3H, s, H10), 4.38 (1H, t, J = 7.0 Hz, H5), 5.10 (1H, br s, H4).

^{13}C NMR (101 MHz, CDCl$_3$): δ 17.2 (C6), 28.5 (C1), 45.7 (C8), 52.6 (C10), 55.6 (C5), 80.4 (C2), 167.4 (C9), 202.4 (C7).

HRMS (ESI) calculated for C$_{11}$H$_{19}$NNaO$_5$ ([M+Na]$^+$): 268.1155; found: 268.1157.

6.1.2.12　Methyl　　　　(S)-4-((tert-butoxycarbonyl)amino)-2,2-dimethyl-3-oxopentanoate (54a)
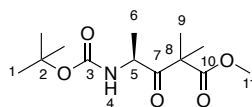
To a solution of **53a** (0.583 g, 2.38 mmol) in THF (10 mL), K_2CO_3 (3.29 g, 23.8 mmol, 10 equiv.) and methyl iodide (0.74 mL, 12 mmol, 5.0 equiv.) were added under an argon atmosphere. The reaction was stirred overnight and more K_2CO_3 (5.0 equiv.) and methyl iodide (2.5 equiv.) were added. The reaction was stirred overnight, quenched with H_2O (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 15 mL) and the combined organic layers were washed with brine (2 × 25 mL), dried with $MgSO_4$, filtered, and concentrated in *vacuo* to give the desired product **54a** as yellow solid (0.634 g, 2.32 mmol 98%).

R_f = 0.59 (EtOAc/hexane, 1:2, UV/$KMnO_4$).

^{1}H NMR (400 MHz, $CDCl_3$): δ 1.26 (3H, d, J = 7.0 Hz, H6), 1.41 (3H, s, H9), 1.43 (12H, s, H1 and H9), 3.73 (3H, s, H11), 4.66 (1H, dt, J = 7.0, 14.0 Hz, H5), 4.97 (1H, d, J = 7.5 Hz, H4).

^{13}C NMR (101 MHz, $CDCl_3$): δ 19.2 (C6), 22.3 (C9), 22.5 (C9), 28.5 (C1), 51.6 (C5), 54.7 (C8), 80.3 (C2), 173.7 (C10), 208.7 (C7).

HRMS (ESI) calculated for $C_{13}H_{23}NNaO_5$ ([M+Na]$^+$): 296.1468; found: 296.1465.

6.1.2.13　Methyl　(S)-4-amino-2,2-dimethyl-3-oxopentanoate　hydrochloride (55a)
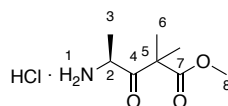
To **54a** (0.606 g, 2.22 mmol) was added 4 M HCl in dioxane (20 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 2 hours to provide the product **55a** as white powder (0.465 g, 2.22 mmol, 100%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 1.46 (3H, s, H6), 1.52 (3H, s, H6), 1.63 (3H, d, J = 6.5 Hz, H3). 3.76 (3H, H8), 4.63 (1H, br H2), 8.66 (2H, br s, H1).

^{13}C NMR (101 MHz, CDCl$_3$): δ 16.9 (C3), 22.5 (C6), 23.0 (C6), 52.7 (C2), 53.2 (C5), 54.3 (C8), 172.7 (C7), 204.6 (C4).

HRMS (ESI) calculated for C$_8$H$_{15}$NNaO$_3$ ([M+Na]$^+$): 196.0944; found: 197.0976.

6.1.2.14 Methyl (S)-4-((S)-2-butyramido-3-hydroxypropanamido)-2,2-dimethyl-3-oxopentanoate (56a)
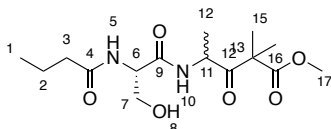
To a solution of **47** (0.0912 g, 0.527 mmol) in THF (7 mL) was added HOBt (0.161 g, 1.05 mmol, 2.0 equiv.), EDC (0.221 g, 1.05 mmol, 2.0 equiv.) and Et$_3$N (0.16 mL, 1.1 mmol, 2.0 equiv.) and the solution was stirred for 1 hour. Amine **55a** (0.101 g, 0.577 mmol, 1.1 equiv.) and Et$_3$N (0.10 mL, 0.70 mmol, 1.0 equiv.) was added and the reaction was stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 15 mL) and the combined organic layers were washed with brine (2 × 20 mL). The organic layer was dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56a** as yellow oil (0.0402 g, 0.118 mmol, 23%)

R$_f$ = 0.61 (MeOH/DCM, 1:9, UV/KMnO$_4$).

Diastereomer 1: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.94 (3H, t, *J* = 7.5 Hz, H1), 1.29 (3H, d, *J* = 5.5 Hz, H12), 1.40 (3H, s, H15), 1.42 (3H, s, H15), 1.67 (2H, sixt., *J* = 7.5 Hz, H2), 2.23 (2H, t, *J* = 5.5 Hz, H3), 3.59 (1H, dd, *J* = 1.5, 5.0 Hz, H7), 3.72 (3H, s, H17), 4.04 (1H, dd, *J* = 3.5, 11.5, H7), 4.45 (1H, m, H6), 4.91 (1H, m, H11), 6.57 (1H, br d, *J* = 6.5 Hz, H5), 7.22 (1H, br d, *J* = 8.0 Hz, H10).

^{13}C NMR (101 MHz, CDCl$_3$): δ 13.8 (C1), 18.7 (C12), 19.1 (C2), 22.4 (C15), 22.5 (C15), 38.4 (C3), 50.6 (C11), 52.8 (C17), 53.8 (C6), 54.7 (C14), 62.8 (C7), 170.3 (C10), 173.6 (C16), 174.2 (C4), 207.4 (C13).

Diastereomer 2: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.96 (3H, t, *J* = 7.5 Hz, H1), 1.29 (3H, d, *J* = 5.5 Hz, H12), 1.41 (3H, s, H15), 1.44 (3H, s, H15), 1.67 (2H, sixt.,

J = 7.5 Hz, H2), 2.26 (2H, t, *J* = 7.5 Hz, H3), 3.62 (1H, dd, *J* = 1.5, 5.0 Hz, H7), 3.74 (3H, s, H17), 4.10 (1H, dd, *J* = 3.0, 11.5 Hz, H7), 4.45 (1H, m, H6), 4.91 (1H, m, H11), 6.57 (1H, br d, *J* = 6.5 Hz, H5), 7.22 (1H, br d, *J* = 8.0 Hz, H10).

^{13}C NMR (101 MHz, CDCl$_3$): δ 13.8 (C1), 18.7 (C12), 19.2 (C2), 22.4 (C15), 22.6 (C15), 38.4 (C3), 50.7 (C11), 52.9 (C17), 53.9 (C6), 54.7 (C14), 62.9 (C7), 170.3 (C10), 173.6 (C16), 174.2 (C4), 207.4 (C13).

HRMS (ESI) calculated for C$_{15}$H$_{26}$N$_2$NaO$_6$ ([M+Na]$^+$): 353.1683; found: 353.1679.

6.1.2.15   Methyl   (S)-4-((tert-butoxycarbonyl)amino)-5-(1   H-indol-3-yl)-3-oxopentanoate (53e)
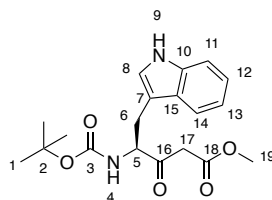
To a solution *N*-Boc-L-tryptophan (**52e**) (0.605 g, 1.99 mmol) in THF (12 mL) was added CDI (0.490 g, 3.02 mmol, 1.5 equiv.) under argon. After stirring the reaction mixture at room temperature for 2 hours, methyl potassium malonate (0.4717 g, 3.02 mmol, 1.5 equiv.) and MgCl$_2$ (0.290 g, 3.04 mmol, 1.5 equiv.) were added and the reaction was stirred overnight. The reaction was quenched with 1 M aqueous HCl (10 mL), THF was removed in *vacuo* and the aqueous layer was extracted with EtOAc (3 × 20 mL). Organic extractions were combined, washed with 5% NaHCO$_3$ (w/w in H$_2$O, 2 × 40 mL) and brine (40 mL), dried with MgSO$_4$ and filtered. The solvent was removed in *vacuo* to provide the desired product **53e** as orange oil (0.657 g, 1.83 mmol, 91%).

R$_f$ = 0.49 (EtOAc/hexane, 1:2, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.41 (9H, s, H1), 3.47 (2H, m, H17), 3.25 (2H, m, H6), 3.66 (3H, s, H19), 4.67 (1H, m, H5), 5.13 (1H, br d, *J* = 7.1 Hz, H4), 7.03 (1H, s, H8), 7.14 (3H, t, *J* = 9.5 Hz, H11), 7.12 (1H, t, *J* = 7.3 Hz, H14), 7.36 (1H, d, *J* = 7.9 Hz, H13), 7.62 (1H, d, *J* = 7.5 Hz, H12), 8.15 (1H, br s, H9).

^{13}C NMR (101 MHz, CDCl$_3$): δ 27.0 (C6), 28.4 (C1), 46.9 (C17), 52.5 (C19), 60.0 (C5), 79.9 (C2), 109.7 (C7), 111.4 (C14), 118.9 (C11), 120.0 (C12), 122.5 (C13), 123.1 (C8), 127.1 (C15), 136.2 (C10), 167.0 (C18), 201.1 (C16).

HRMS (ESI) calculated for C$_{19}$H$_{24}$N$_2$NaO$_5$ ([M+Na]$^+$): 383.1577; found: 383.1586.

6.1.2.16 Methyl (S)-4-((tert-butoxycarbonyl)amino)-5-(1H-indol-3-yl)-2,2-dimethyl-3-oxopentanoate (54e)
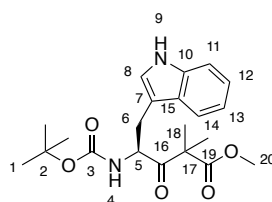
To a solution of **53e** (0.648 g, 1.80 mmol) in THF (10 mL), K$_2$CO$_3$ (2.59 g, 18.7 mmol, 10 equiv.) and methyl iodide (0.58 mL, 9.38 mmol, 5.0 equiv.) were added under an argon atmosphere. The reaction was stirred overnight and more K$_2$CO$_3$ (5.0 equiv.) and methyl iodide (2.5 equiv.) were added. The reaction was stirred overnight, quenched with H$_2$O (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 20 mL), and the combined organic layers were washed with brine (2 × 40 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo* to give the desired product **54e** as yellow solid (0.616 g, 1.59 mmol 88%).

R$_f$ = 0.53 (EtOAc/hexane, 1:2, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.30 (3H, s, H18), 1.36 (12H, s, H1, H18), 3.07 (1H, dd, *J* = 7.2, 15.0 Hz, H6), 3.21 (1H, dd, *J* = 5.9, 15.0 Hz, H6), 3.57 (3H, s, H20), 4.82 (1H, br d, H4), 4.96 (1H, m, H5) 7.01 (1H, s, H8), 7.16 (2H, m, H12, H13), 7.34 (1H, d, *J* = 7.9 Hz, H11), 7.60 (1H, d, *J* = 7.9 Hz, H14), 8.06 (1H, br s, H9).

^{13}C NMR (101 MHz, CDCl$_3$): δ 21.9 (C18), 22.3 (C18), 28.2 (C6), 28.4 (C1), 52.5 (C20), 54.8 (C17), 55.9 (C5), 80.2 (C2), 110.4 (C7), 111.2 (C14), 118.9 (C11), 119.8 (C12), 120.1 (C10), 122.3 (C13), 123.1 (C8), 135.7 (15), 173.1 (C19), 207.9 (C16).

HRMS (ESI) calculated for C$_{21}$H$_{28}$N$_2$NaO$_5$ ([M+Na]$^+$): 411.1890; found: 411.1891.
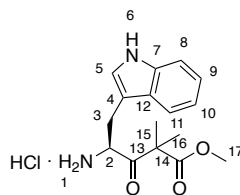
To **54e** (0.756 g, 1.95 mmol) was added 4 M HCl in dioxane (20 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 2 hours to provide the product **55e** (0.632 g, 1.95 mmol, 100%).

¹H NMR (400 MHz, CDCl₃): δ 1.44 (3H, s, H15), 1.53 (3H, s, H15), 2.99 (1H, m, H3). 3.27 (1H, m, H3), 3.71 (3H, s, H17), 4.68 (1H, br s, H2), 7.10 (3H, br m, H5, H9, H10), 7.29 (2H br m, H8 H11), 8.07 (3H, br s, H6, H1).

¹³C NMR (101 MHz, CDCl₃): δ 22.7 (C15), 22.9 (C15), 26.6 (C3), 53.3 (C17), 55.0 (C14), 55.9 (C2), 105.5 (C4), 112.2 (C11), 118.1 (C8), 120.5 (C9), 122.0 (C10), 126.8 (C5), 128.1 (C12), 136.5 (C7), 173.2 (C16), 204.4 (C13).

6.1.2.18   Methyl   (S)-4-((S)-2-butyramido-3-hydroxypropanamido)-5-(1H-indol-3-yl)-2,2-dimethyl-3-oxopentanoate (56e)
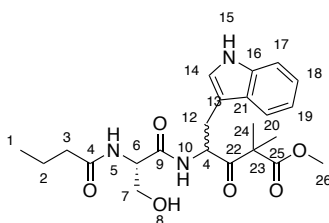
To a solution of **47** (0.0701 g, 0.398 mmol) in THF (7 mL) was added NMI (0.10 mL, 1.72 mmol, 3.0 equiv.) and TCFH (0.169 g, 0.603 mmol, 1.5 equiv.) and the solution was stirred for 15 minutes. Amine **55e** (0.0.131 g, 0.402 mmol, 1.0 equiv.) was added and the reaction was stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 15 mL), and the combined organic layers were washed with brine (2 × 20 mL). The organic layer was dried with MgSO₄, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56e** (0.0462 g, 0.104 mmol, 26%).

$R_f = 0.29$ (MeOH/DCM, 1:9, UV/KMnO$_4$).

Diastereomer 1: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.92 (3H, t, $J = 7.7$ Hz, H1), 1.18 (3H, s, H24), 1.25 (3H, s, H24), 1.63 (2H, m, H2), 2.18 (2H, t, $J = 6.9$ Hz, H3), 2.97 (1H, dd, $J = 5.8, 14.6$ Hz, H12), 3.25 (1H, dd, $J = 6.0, 13.8$ Hz, H12), 3.51 (1H, s, H26), 3.55 (1H, m, H7), 3.67 (1H, m, H7), 4.39 (1H, t, $J = 5.2$ Hz, H6), 5.24 (1H, t, $J = 16.9$ Hz, H11), 7.01 (1H, s, H12), 7.08 (2H, t, $J = 6.9$ Hz, H18, H19), 7.30 (1H, d, $J = 6.7$ Hz, H17), 7.58 (1H, d, $J = 7.7$ Hz, H20).

^{13}C NMR (101 MHz, CDCl$_3$): δ 14.0 (C1), 20.2 (C2), 22.0 (C24), 22.3 (C24), 29.1 (C12), 38.7 (C3), 53.1 (C26), 55.3 (C11), 56.6 (C6), 56.9 (C23), 62.9 (C7), 110.6 (C13), 112.3 (C20), 119.3 (C17), 119.9 (C18), 122.4 (C16), 124.7 (C19), 128.8 (C14), 137.8 (C21), 171.7 (C9), 174.9 (C25), 176.6 (C4), 208.1 (C22).

Diastereomer 2: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.96 (3H, t, $J = 7.7$ Hz, H1), 1.20 (3H, s, H24), 1.28 (3H, s, H24), 1.63 (2H, m, H2), 2.25 (2H, t, $J = 7.7$ Hz, H3), 2.97 (1H, dd, $J = 5.8, 14.6$ Hz, H12), 3.25 (1H, dd, $J = 6.0, 13.8$ Hz, H12), 3.51 (1H, s, H26), 3.55 (1H, m, H7), 3.67 (1H, m, H7), 4.39 (1H, t, $J = 5.2$ Hz, H6), 5.24 (1H, t, $J = 16.9$ Hz, H11), 7.01 (1H, s, H12), 7.03 (2H, t, $J = 6.9$ Hz, H18, H19), 7.29 (1H, d, $J = 6.7$ Hz, H17), 7.53 (1H, d, $J = 7.7$ Hz, H20).

^{13}C NMR (101 MHz, CDCl$_3$): δ 14.0 (C1), 20.2 (C2), 22.1 (C24), 22.3 (C24), 29.1 (C12), 38.7 (C3), 53.1 (C26), 55.3 (C11), 56.4 (C6), 56.9 (C23), 62.9 (C7), 110.6 (C13), 112.3 (C20), 119.3 (C17), 120.0 (C18), 122.5 (C16), 124.7 (C19), 128.8 (C14), 137.8 (C21), 171.7 (C9), 174.9 (C25), 176.6 (C4), 208.1 (C22).

HRMS (ESI) calculated for C$_{23}$H$_{31}$N$_3$NaO$_6$ ([M+Na]$^+$): 468.2105; found: 468.2096.

6.1.2.19   *Cyclopent-1-en-1-yl trifluoromethanesulfonate (49)*
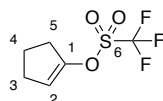
To a solution of cyclopentanone (**48**) (0.80 mL, 8.8 mmol) in anhydrous DCM (30 mL) was added Na$_2$CO$_3$ (1.41 g, 13.3 mmol, 1.5 equiv.) and triflic anhydride (5.0 g, 17.72 mmol, 2 equiv.) dropwise at -10 °C under an argon atmosphere. The cooling bath was removed, and the reaction stirred overnight before it was quenched

with water (30 mL) and extracted with DCM (3 × 30 mL). The combined organic layers were washed with brine (90 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo* to provide the crude product as dark blue liquid (1.05 g, 4.86 mmol, 55%). ^{1}H NMR and COSY spectra confirmed the formation of the desired product **49** which was used in the next step without further purification.

^{1}H NMR (400 MHz, CDCl$_3$): δ 2.02 (2H, quint., *J* = 7.4 Hz, H4), 2.42 (2H, m, H3), 2.57 (2H, m, H5), 5.63 (1H, s, H2).

6.1.2.20 Methyl (S)-2-((tert-butoxycarbonyl)amino)-3-(cyclopent-1-en-1-yl)propanoate (50)
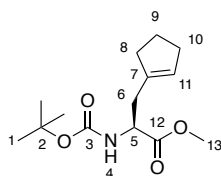
To a solution of zinc dust (1.60 g, 24.3 mmol, 5 equiv.) in DMF (6 mL) was added TMSCl (0.62 mL, 4.9 mmol, 1 equiv.) dropwise and the mixture was stirred at room temperature for 1 hour. The upper clear layer was removed, the bottom layer was washed with DMF (2 × 10 mL), cooled to 0 °C and *N*-Boc-β-iodoalanine-OMe (1.61 g, 4.90 mmol, 1 equiv.), Triflate **49** (1.05 g, 4.90 mmol), and Pd(dppf)Cl$_2$ (0.671 g, 0.970 mmol, 0.02 equiv.) was added. The reaction mixture was stirred at room temperature overnight, quenched with brine (10 mL) and extracted with EtOAc (3 × 10 mL). The combined organic layers were washed with brine (30 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (EtOAc/hexane, 1:99 to 1:9) to yield the product **50** as a yellow oil (0.898 g, 3.33 mmol, 69%).

R$_f$ = 0.28 (EtOAc/hexane, 1:9, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.44 (9H, s, H1), 1.86 (2H, quint., *J* = 6.9 Hz, H9), 2.22 (2H, m, H8), 2.29 (2H, m, H10), 2.54 (2H, m, H6), 3.72 (3H, m, H13), 4.41 (1H, m, H5), 4.94 (1H, br d, *J* = 6.0 Hz, H4), 5.46 (1H, s, H11).

^{13}C NMR (101 MHz, CDCl$_3$): δ 23.7 (C9), 28.5 (C1), 32.6 (C10), 34.4 (C6), 34.9 (C8), 52.2 (C5), 53.4 (C13), 80.8 (C2), 128.5 (C11), 139.7 (C7), 173.3 (C12).

HRMS (ESI) calculated for $C_{14}H_{23}NNaO_4$ ([M+Na]$^+$): 292.1519; found: 292.1515.

6.1.2.21 Methyl (S)-2-((tert-butoxycarbonyl)amino)-3-(cyclopent-1-en-1-yl)propanoic acid (51)
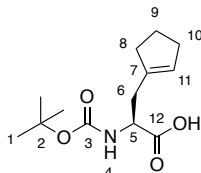
To a solution of **50** (0.338 g, 1.25 mmol) in water/methanol (2:1, 10 mL) was added LiOH (0.0901 g, 3.76 mmol, 3 equiv.) and the reaction was stirred at room temperature overnight. Methanol was removed *in vacuo*, the aqueous layer washed with DCM (10 mL), acidified with HCl (1 M, aqueous) to pH 3, and extracted with DCM (3 × 10 mL). The combined organic layers were dried with MgSO$_4$, filtered, and concentrated in *vacuo* to provide the product **51** as opaque oil (0.320 g, 1.25 mmol, 100%).

^{1}H NMR (400 MHz, CDCl$_3$): δ1.44 (9H, s, H1), 1.87 (2H, quint., J = 6.3 Hz, H9), 2.23 (2H, m, H8), 2.30 (2H, m, H10), 2.54 (2H, m, H6), 4.41 (1H, br m, H5), 4.94 (1H, br d, J = 6.0 Hz, H4), 5.50 (1H, s, H11).

^{13}C NMR (101 MHz, CDCl$_3$): δ23.6 (C9), 28.4 (C1), 32.6 (C10), 33.9 (C6), 34.8 (C8), 52.1 (C5), 80.4 (C2), 128.7 (C11), 138.8 (C7), 177.5 (C12).

HRMS (ESI) calculated for $C_{13}H_{20}NO_4$ ([M-H]$^-$): 254.1398; found: 254.1401.

6.1.2.22 Methyl (S)-2-((tert-butoxycarbonyl)amino)-5-(cyclopent-1-en-1-yl)-3-oxopentanoate (53c)
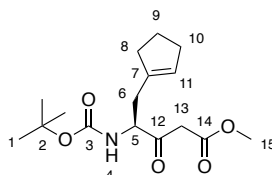
To a solution *N*-Boc-L-cyclopentenylalanine (**51**) (0.297 g, 1.17 mmol) in anhydrous THF (10 mL) was added CDI (0.283 g, 1.75 mmol, 1.5 equiv.) under argon. After stirring the reaction mixture at room temperature for 1 hour, methyl potassium malonate (0.273 g, 1.75 mmol, 1.5 equiv.) and MgCl$_2$ (0.166 g, 1.75 mmol,

1.5 equiv.) were added and the reaction was stirred overnight. The reaction was quenched with 1 M aqueous HCl (8 mL), THF was removed in *vacuo* and the aqueous layer was extracted with EtOAc (3 × 10 mL). Organic extractions were combined, washed with 5% NaHCO$_3$ (w/w in H$_2$O, 2 × 30 mL) and brine (30 mL), dried with MgSO$_4$ and filtered. The solvent was removed in *vacuo* to provide the desired product **53c** as yellow oil (0.321 g, 1.03 mmol, 88%).

R$_f$ = 0.38 (EtOAc/hexane, 1:9, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.43 (9H, s, H1), 1.86 (2H, quint., J = 6.8 Hz, H9), 2.22 (2H, m, H8), 2.30 (2H, m, H10), 2.52 (2H, m, H6), 3.55 (2H, t, J = 19.5 Hz, H13), 3.73 (3H, s, H15), 4.41 (1H, br m, H5), 4.96 (1H, br s, H4), 5.47 (1H, s, H11).

^{13}C NMR (101 MHz, CDCl$_3$): δ23.6 (C9), 28.4 (C1), 32.6 (C10), 32.8 (C6), 34.9 (C8), 46.3 (C13), 52.6 (C15), 58.3 (C5), 80.4 (C2), 128.7 (C11), 139.0 (C7), 167.6 (C14), 202.4 (C12).

HRMS (ESI) calculated for C$_{16}$H$_{25}$NNaO$_5$ ([M+Na]$^+$): 334.1625; found: 334.1620.

6.1.2.23   Methyl (S)-4-((tert-butoxycarbonyl)amino)-5-(cyclopent-1-en-1-yl)-2,2-dimethyl-3-oxopentanoate (53c)
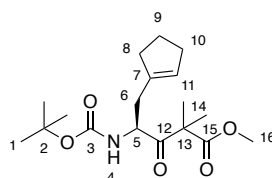
To a solution of **53c** (0.316 g, 1.02 mmol) in THF (10 mL), K$_2$CO$_3$ (1.40 g, 10.2 mmol, 10 equiv.) and methyl iodide (0.32 mL, 5.08 mmol, 5.0 equiv.) were added under an argon atmosphere. The reaction was stirred overnight and more K$_2$CO$_3$ (5.0 equiv.) and methyl iodide (2.5 equiv.) were added. The reaction was stirred overnight, quenched with H$_2$O (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 20 mL), and the combined organic layers were washed with brine (2 × 40 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo* to give the desired product **54c** as yellow oil (0.328 g, 0.966 mmol 95%).

R$_f$ = 0.34 (EtOAc/hexane, 1:9, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ1.40 (15H, s, H1,H14), 1.84 (2H, quint., J = 7.0 Hz, H9), 2.21 (2H, m, H8), 2.30 (2H, m, H10), 2.45 (2H, m, H6), 3.72 (3H, s, H16), 4.70 (1H, br s, H5), 5.44 (1H, s, H11).

^{13}C NMR (101 MHz, CDCl$_3$): δ22.1 (C14), 22.5 (C14), 23.7 (C9), 28.4 (C1), 32.7 (C10), 34.6 (C6), 34.9 (C8), 52.7 (C16), 54.3 (C5), 79.9 (C2), 128.2 (C11), 134.7 (C14), 139.0 (C7), 173.8 (C15), 208.1 (C12).

HRMS (ESI) calculated for C$_{18}$H$_{29}$NNaO$_5$ ([M+Na]$^+$): 362.1938; found: 362.1934.

6.1.2.24 Methyl (S)-4-((tert-butoxycarbonyl)amino)-5-(cyclopent-1-en-1-yl)-2,2-dimethyl-3-oxopentanoate (55c)
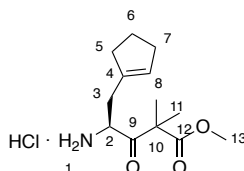

To **54c** (0.311 g, 0.916 mmol) was added 4 M HCl in dioxane (20 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 2 hours to provide the product **55c** as brown oil (0.252 g, 0.915 mmol, 100%).

R$_f$ = 0.34 (EtOAc/hexane, 1:9, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.46 (3H, s, H11), 1.54 (3H, s, H11), 1.90 (2H, br m, H6), 2.18 (2H, m, H5), 2.34 (2H, m, H7), 2.68 (2H, m, H3), 3.76 (3H, s, H13), 4.61 (1H, br s, H2), 5.80 (1H, s, H8).

^{13}C NMR (101 MHz, CDCl$_3$): δ 22.8 (C11), 23.0 (C11), 23.5 (C6), 32.3 (C7), 32.9 (C3), 34.8 (C5), 53.2 (C13), 54.4 (C2), 132.9 (C8), 134.7 (C11), 136.0 (C4), 172.8 (C12), 203.9 (C9).

HRMS (ESI) calculated for C$_{13}$H$_{22}$NO$_3$ ([M+Na]$^+$): 240.1594; found: 240.1588.

6.1.2.25 Methyl (S)-4-((S)-2-butyramido-3-hydroxypropanamido)-5-(cyclopent-1-en-1-yl)-2,2-dimethyl-3-oxopentanoate (56c)
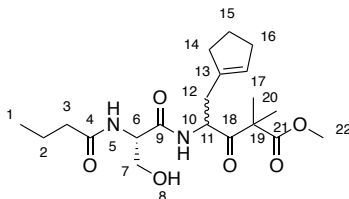
To a solution of **47** (0.0384 g, 0.221 mmol) in THF (7 mL) was added NMI (0.05 mL, 0.7 mmol, 3.0 equiv.) and TCFH (0.0923 g, 0.331 mmol, 1.5 equiv.) and the solution was stirred for 15 minutes. Amine **55c** (0.0607 g, 0.221 mmol, 1.0 equiv.) was added and the reaction was stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 15 mL), and the combined organic layers were washed with brine (2 × 20 mL). The organic layer was dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56c** (0.0152 g, 0.0383 mmol, 17%).

R$_f$ = 0.40 (MeOH/DCM, 5:95, UV/KMnO$_4$).

Diastereomer 1: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.93 (3H, t, *J* = 9.3 Hz, H1), 1.38 (3H, s, H20), 1.42 (3H, s, H20), 1.64 (2H, quint., *J* = 7.0 Hz, H2), 1.81 (2H, m, H15), 2.18 (2H, t, *J* = 7.0 Hz, H3), 2.22 (2H, m, H14), 2.30 (2H, m, H16), 2.53 (2H, m, H12), 3.57 (1H, dd, *J* = 4.8, 19.2 Hz, H7), 3.72 (3H, s, H22), 4.00 (1H, dd, *J* = 3.6, 11.9 Hz, H7), 4.40 (1H, br s, H6), 4.96 (1H, m, H11), 5.38 (1H, s, H17), 6.63 (1H, br d, *J* = 6.6 Hz, H5), 7.14 (1H, br s, *J* = 9.2 Hz, H10).

^{13}C NMR (101 MHz, CDCl$_3$): δ 13.8 (C1), 19.1 (C2), 22.3 (C20), 22.5 (C20), 23.6 (C15), 32.6 (C16), 33.8 (C12), 34.7 (C3), 38.3 (C14), 52.8 (C22), 53.3 (C11), 54.9 (C19), 62.7 (C7), 128.2 (C17), 139.1 (C13), 170.8 (C9), 174.0 (C21), 207.0 (C18).

Diastereomer 2: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.93 (3H, t, *J* = 9.3 Hz, H1), 1.20 (3H, s, H20), 1.24 (3H, s, H20), 1.64 (2H, quint., *J* = 7.0 Hz, H2), 1.81 (2H, m, H15), 2.18 (2H, t, *J* = 7.0 Hz, H3), 2.22 (2H, m, H14), 2.30 (2H, m, H16), 2.53 (2H, m, H12), 3.54 (1H, dd, *J* = 4.8, 19.2 Hz, H7), 3.70 (3H, s, H22), 3.97 (1H, dd, *J* = 3.6, 11.9 Hz, H7), 4.40 (1H, br s, H6), 4.96 (1H, m, H11), 5.45 (1H, s, H17), 6.86 (1H, br d, *J* = 6.6 Hz, H5), 7.14 (1H, br s, *J* = 9.2 Hz, H10).

¹³C NMR (101 MHz, CDCl₃): δ 13.8 (C1), 19.1 (C2), 20.3 (C20), 21.5 (C20), 23.5 (C15), 32.6 (C16), 33.8 (C12), 34.7 (C3), 38.4 (C14), 52.8 (C22), 53.1 (C11), 55.0 (C19), 63.0 (C7), 128.7 (C17), 138.9 (C13), 170.5 (C9), 173.8 (C21), 207.0 (C18).

HRMS (ESI) calculated for $C_{20}H_{32}N_2NaO_6$ ([M+Na]⁺): 419.2153; found: 419.2146.

6.1.2.26   Cyclopentenylalanine
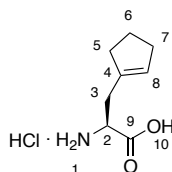
To **51** (0.0139 g, 0.0544 mmol) was added 4 M HCl in dioxane (20 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 3 hours to provide the product as white powder (0.0095 g, 0.050 mmol, 92%).

¹H NMR (400 MHz, MeOD): 1.94 (2H, quint., J = 7.4 Hz, H6), 2.35 (4H, m, H5,7), 2.70 (1H, dd, J = 8.9, 14.9 Hz, H3), 2.80 (1H, dd, J = 4.5, 14.9 Hz, H3), 4.10 (1H, t, J = 6.0 Hz, H2), 5.66 (1H, s, H8).

¹³C NMR (101 MHz, MeOD): δ 24.3 (C6), 33.2 (C7), 33.4 (C3), 35.2 (C5), 52.4 (C2), 131.4 (C8), 138.3 (C4), 171.7 (C9).

6.1.2.27   Methyl   (S)-5-(benzyloxy)-4-((tert-butoxycarbonyl)amino)-3-oxopentanoate (53f)
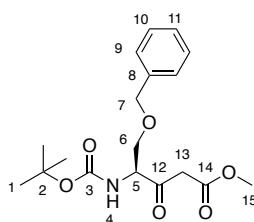
To a solution *N*-Boc-L-Ser(OBzl)-OH (**52f**) (0.596 g, 2.00 mmol) in THF (10 mL) was added CDI (0.4865 g, 3.000 mmol, 1.5 equiv.) under argon. After stirring the reaction mixture at room temperature for 2 hours, methyl potassium malonate (0.468 g, 3.00 mmol, 1.5 equiv.) and MgCl₂ (0.285 g, 3.00 mmol, 1.5 equiv.) were added and the reaction was stirred overnight. The reaction was quenched with 1 M aqueous HCl (10 mL), THF was removed in *vacuo* and the aqueous layer was extracted with EtOAc

(3 × 15 mL). Organic extractions were combined, washed with 5% NaHCO$_3$ (w/w in H$_2$O, 2 × 30 mL) and brine (30 mL), dried with MgSO$_4$ and filtered. The solvent was removed in *vacuo* to provide the desired product **53f** (0.526 g, 1.541 mmol, 77%).

R$_f$ = 0.61 (EtOAc/hexane, 1:2, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.44 (9H, s, H1), 3.59 (2H, d, J = 3.0 Hz, H13), 3.66 (1H, dd, J = 4.5, 10.0 Hz, H6), 3.70 (3H, s, H15), 3.92 (1H, dd, J = 3.5, 9.5 Hz, H6), 4.47 (1H, br s, H5), 4.50 (2H, d, J = 3.0 Hz, H7), 5.47 (1H, br d, J = 7.0 Hz, H4), 7.31 (5H, m, H9-11).

^{13}C NMR (101 MHz, CDCl$_3$): δ 28.4 (C1), 46.7 (C13), 52.5 (C15), 60.9 (C5). 69.4 (C6), 73.6 (C7), 80.4 (C2), 127.9 (C11), 128.1 (C10), 128.6 (C9), 137.4 (C8), 167.3 (C14), 201.0 (C12).

HRMS (ESI) calculated for C$_{18}$H$_{25}$NNaO$_6$ ([M+Na]$^+$): 374.1574; found: 374.1578.

6.1.2.28 Methyl (S)-5-(benzyloxy)-4-((tert-butoxycarbonyl)amino)-2,2-dimethyl-3-oxopentanoate (54f)
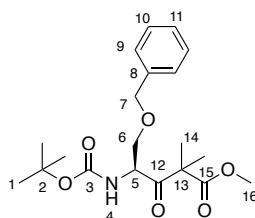
To a solution of **53f** (0.407 g, 1.16 mmol) in THF (10 mL), K$_2$CO$_3$ (1.60 g, 11.6 mmol, 10 equiv.) and methyl iodide (0.56 mL, 12 mmol, 5.0 equiv.) were added under an argon atmosphere. The reaction was stirred overnight and more K$_2$CO$_3$ (5.0 equiv.) and methyl iodide (2.5 equiv.) were added. The reaction was stirred overnight, quenched with 1 M aqueous HCl (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 15 mL) and the combined organic layers were washed with brine (2 × 25 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo* to give the desired product **54f** as yellow solid (0.440 g, 1.16 mmol 100%).

R$_f$ = 0.68 (EtOAc/hexane, 1:2, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.39 (3H, s, H14), 1.40 (3H, s, H14), 1.44 (9H, s, H1), 3.59 (1H, dd, J = 5.0, 9.5 Hz, H6), 3.65 (3H, s, H16), 3.79 (1H, dd, J = 5.0, 9.5 Hz, H6), 4.49 (2H, s, H7), 4.78 (1H, dt, J = 4.5, 9.0 Hz, H5), 5.15 (1H, br d, J = 9.0 Hz, H4), 7.23 (5H, m, H9-11).

^{13}C NMR (101 MHz, CDCl$_3$): δ 22.3 (C14), 22.6 (C14), 28.4 (C1), 52.6 (C16), 54.1 (C13), 56.6 (C5), 69.9 (C6), 73.4 (C7), 81.3 (C2), 127.7 (C11), 127.9 (C10), 128.5 (C9), 137.7 (C8), 173.6 (C15), 206.8 (C12).

HRMS (ESI) calculated for C$_{20}$H$_{29}$NNaO$_6$ ([M+Na]$^+$): 402.1886; found: 402.1887.

6.1.2.29 Methyl (S)-4-amino-5-(benzyloxy)-2,2-dimethyl-3-oxopentanoate hydrochloride (55f)
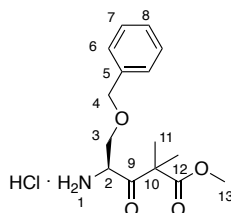
To **54f** (0.412 g, 1.09 mmol) was added 4 M HCl in dioxane (20 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 2 hours to provide the product **55f** (0.334 g, 1.06 mmol, 98%).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.39 (3H, s, H11), 1.48 (3H, s, H11), 3.59 (3H, s, H13), 3.55 (2H, m, H3), 4.54 (2H, s, H4), 4.88 (1H, br s, H2), 7.29 (5H, br s, H6-8), 8.71 (2H, br s, H1).

^{13}C NMR (101 MHz, CDCl$_3$): δ 22.8 (C11), 23.0 (C11), 53.1 (C13), 54.6 (C10), 56.8 (C2), 66.9 (C3), 73.7 (C4), 128.1 (C8), 128.1 (C7), 128.6 (C6), 136.9 (C5), 173.0 (C12), 201.8 (C9).

HRMS (ESI) calculated for C$_{15}$H$_{22}$NNaO$_4$ ([M+Na]$^+$): 280.1543; found: 280.1547.

6.1.2.30 Methyl (S)-5-(benzyloxy)-4-((S)-2-butyramido-3-hydroxypropanamido)-2,2-dimethyl-3-oxopentanoate (56f)
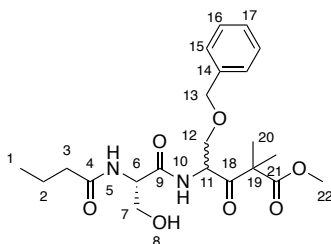
To a solution of **47** (0.201 g, 0.719 mmol) in THF (5 mL) was added NMI (0.17 mL, 2.15 mmol, 3.0 equiv.) and TCFH (0.304 g, 1.0836 mmol, 1.5 equiv.) and the solution was stirred for 15 minutes. Amine **55f** (0.108 g, 0.266 mmol) was added and the reaction stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (10 mL), THF concentrated in vacuo and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (3 × 40 mL), dried with MgSO₄, filtered, and concentrated in vacuo. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56f** as a yellow oil (0.102 g, 0.233 mmol, 33%).

R_f = 0.65 (MeOH/DCM, 5:95, UV/KMnO₄).

Diastereomer 1: ^{1}H NMR (400 MHz, CDCl₃): δ 0.93 (3H, t, J = 6.9 Hz, H1), 1.38 (3H, s, H20), 1.42 (3H, s, H20), 1.65 (2H, m, H2), 2.23 (2H, t, J = 6.9 Hz, H3), 3.60 (2H, m, H7, H12), 3.64 (3H, s, H22), 3.78 (1H, t, J = 6.3 Hz, H12), 4.06 (1H, t, J = 11.7 Hz, H7), 4.45 (2H, s, H13), 4.51 (1H, m, H6), 5.07 (1H, m, H11), 6.56 (1H, br d, J = 6.8 Hz, H5), 7.29 (5H, m, H15-17).

^{13}C NMR (101 MHz, CDCl₃): δ 13.7 (C1), 19.1 (C2), 22.3 (C20), 22.6 (C20), 38.4 (C3), 52.8 (C22), 53.9 (C6), 54.8 (C19), 55.6 (C11), 62.8 (C7), 69.3 (C12), 73.4 (C13), 127.8 (C16), 128.0 (C17), 128.6 (C15), 137.6 (C14), 171.8 (C4), 174.0 (C21), 205.0 (C18).

Diastereomer 2: ^{1}H NMR (400 MHz, CDCl₃): δ 0.91 (3H, t, J = 7.7 Hz, H1), 1.36 (3H, s, H20), 1.39 (3H, s, H20), 1.63 (2H, m, H2), 2.17 (2H, t, J = 7.8 Hz, H3), 3.60 (2H, m, H7, H12), 3.64 (3H, s, H22), 3.74 (1H, t, J = 7.3 Hz, H12), 4.07 (1H, t, J = 11.7 Hz, H7), 4.49 (2H, s, H13), 4.51 (1H, m, H6), 5.07 (1H, m, H11), 6.52 (1H, br d, J = 7.3 Hz, H5), 7.29 (5H, m, H15-17).

¹³C NMR (101 MHz, CDCl$_3$): δ 13.7 (C1), 19.1 (C2), 22.3 (C20), 22.5 (C20), 38.4 (C3), 52.8 (C22), 54.3 (C6), 54.8 (C19), 55.4 (C11), 62.8 (C7), 69.4 (C12), 73.5 (C13), 127.8 (C16), 128.0 (C17), 128.6 (C15), 137.6 (C14), 171.8 (C4), 174.0 (C21), 205.0 (C18).

HRMS (ESI) calculated for C$_{22}$H$_{32}$N$_2$NaO$_7$ ([M+Na]$^+$): 459.2102; found: 459.2098.

6.1.2.31 Methyl (S)-4-((S)-2-butyramido-3-hydroxypropanamido)-2,2-dimethyl-3-oxopentanoate (53g)
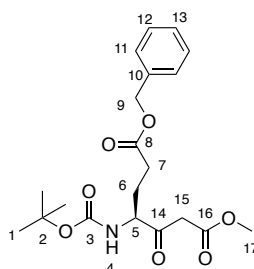
To a solution *N*-Boc-L-Glu(OBzl)-OH (**52g**) (0.677 g, 2.00 mmol) in THF (10 mL) was added CDI (0.486 g, 3.00 mmol, 1.5 equiv.) under argon. After stirring the reaction mixture at room temperature for 2 hours, methyl potassium malonate (0.468 g, 3.00 mmol, 1.5 equiv.) and MgCl2 (0.285 g, 3.00 mmol, 1.5 equiv.) were added and the reaction was stirred overnight. The reaction was quenched with 1 M aqueous HCl (10 mL), THF was removed in *vacuo* and the aqueous layer was extracted with EtOAc (3 × 15 mL). Organic extractions were combined, washed with 5% NaHCO$_3$ (w/w in H$_2$O, 2 × 30 mL) and brine (30 mL), dried with MgSO$_4$ and filtered. The solvent was removed in *vacuo* to provide the desired product **53g** (0.756 g, 1.92 mmol, 96%).

R$_f$ = 0.54 (EtOAc/hexane, 1:2, UV/KMnO$_4$).

¹H NMR (400 MHz, CDCl$_3$): δ 1.43 (9H, s, H1), 1.85 (1H, sextet, *J* = 7.5 Hz, H6), 2.26 (1H, m, H6), 2.47 (2H, m, H7), 3.59 (2H, s, H15), 3.73 (3H, s, H17), 4.41 (1H, m, H5), 5.12 (2H, s, H9), 5.20 (1H, br d, *J* = 7.5 Hz, H4), 7.35 (5H, br s, H11-13).

¹³C NMR (101 MHz, CDCl$_3$): δ 26.0 (C6), 28.1 (C1), 30.1 (C7), 46.0 (C15), 52.6 (C17), 59.0 (C5), 66.8 (C9), 80.5 (C2), 128.4 (C13), 128.5 (C12), 128.7 (C11), 135.8 (C10), 167.3 (C16), 172.8 (C8), 201.6 (C14).

HRMS (ESI) calculated for $C_{20}H_{27}NNaO_7$ ([M+Na]$^+$): 416.1680; found: 416.1683.

6.1.2.32   *7-benzyl 1-methyl (S)-4-((tert-butoxycarbonyl)amino)-2,2-dimethyl-3-oxoheptanedioate (54g)*
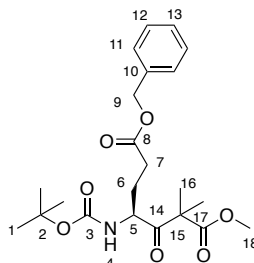
To a solution of **53g** (0.586 g, 1.49 mmol) in THF (10 mL), K_2CO_3 (2.06 g, 14.9 mmol, 10 equiv.) and methyl iodide (0.46 mL, 7.4 mmol, 5.0 equiv.) were added under an argon atmosphere. The reaction was stirred overnight and more K_2CO_3 (5.0 equiv.) and methyl iodide (2.5 equiv.) were added. The reaction was stirred overnight and more K_2CO_3 (5.0 equiv.) and methyl iodide (2.5 equiv.) were added. The reaction was stirred overnight, quenched with 1 M aqueous HCl (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 15 mL) and the combined organic layers were washed with brine (2 × 25 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo* to give the desired product **54g** as yellow solid (0.546 g, 1.295 mmol 86%).

R_f = 0.73 (EtOAc/hexane, 1:2, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.42 (12H, s, H1, H16), 1.43 (3H, s, H16), 1.73 (1H, m, H6), 2.10 (1H, m, H6), 2.43 (2H, m, H7), 3.71 (3H, s, H18), 4.66 (1H, dt, *J* = 3.5, 9.5 Hz, H5), 4.91 (1H, br d, *J* = 10.0 Hz, H4), 5.12 (2H, s, H9), 7.35 (5H, br s, H11-13).

^{13}C NMR (101 MHz, CDCl$_3$): δ 22.1 (C16), 22.3 (C16), 27.8 (C6), 28.4 (C1), 30.1 (C7), 52.8 (C18), 54.8 (C5), 54.9 (C15), 66.6 (C9), 80.2 (C2), 128.4 (C12,13), 128.7 (C11), 135.7 (C10), 172.2 (C17), 173.6 (C8), 207.7 (C14).

HRMS (ESI) calculated for $C_{22}H_{31}NNaO_7$ ([M+Na]$^+$): 444.1993; found: 444.1992.

6.1.2.33 7-benzyl 1-methyl (S)-4-((S)-2-butyramido-3-hydroxypropanamido)-2,2-dimethyl-3-oxoheptanedioate hydrochloride (55g)
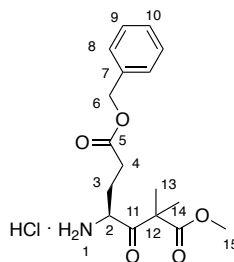

To **54g** (0.567 g, 1.35 mmol) was added 4 M HCl in dioxane (20 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 2 hours to provide the product **55g** (0.480 g, 1.34 mmol, 100%).

^{1}H NMR (400 MHz, CDCl[3]): δ 1.45 (3H, s, H13), 1.50 (3H, s, H13), 2.10 (1H, m, H3), 2.33 (1H, m, H3), 2.66 (1H, m, H4), 2.84 (1H, m, H4), 3.68 (3H, s, H15), 4.73 (1H, br s, H2), 5.08 (2H, m, H6), 7.30 (5H, m, H8-10), 8.74 (2H, br s, H1).

^{13}C NMR (101 MHz, CDCl[3]): δ 22.4 (C13), 22.8 (C13), 25.4 (C3), 29.4 (C4), 53.2 (C15), 54.7 (C2), 55.3 (C12), 66.9 (C6), 128.4 (C10), 128.5 (C9), 128.7 (C8), 135.7 (C7), 172.4 (C14), 172.9 (C5), 204.1 (C11).

HRMS (ESI) calculated for $C_{17}H_{23}NNaO_5$ ([M+Na]$^+$): 344.1468; found: 344.1468.

6.1.2.34 7-benzyl 1-methyl (S)-4-((S)-2-butyramido-3-hydroxypropanamido)-2,2-dimethyl-3-oxoheptanedioate (56g)
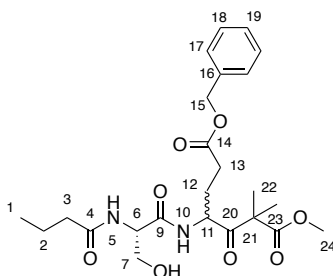
To a solution of **47** (0.203 g, 0.632 mmol) in THF (5 mL) was added NMI (0.15 mL, 1.9 mmol, 3.0 equiv.) and TCFH (0.263 g, 0.938 mmol, 1.5 equiv.) and the solution was stirred for 15 minutes. Amine **55g** (0.103 g, 0.590 mmol) was added and the reaction stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (10 mL), THF concentrated in vacuo and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (3 × 40 mL), dried with

MgSO$_4$, filtered, and concentrated in vacuo. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56g** as a yellow oil (0.112 g, 0.234 mmol, 38%).

R$_f$ = 0.57 (MeOH/DCM, 5:95, UV/KMnO$_4$).

Diastereomer 1: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.90 (3H, t, J = 7.5 Hz, H1), 1.39 (3H, s, H13), 1.40 (3H, s, H13), 1.62 (2H, m, H2), 1.80 (1H, m, H12), 2.14 (1H, m, H12), 2.18 (2H, t, J = 7.5 Hz), 2.43 (2H, m, H13), 3.61 (1H, m, H7), 3.71 (3H, s, H24), 4.03 (1H, m, H7), 4.41 (1H, m, H6), 4.98 (1H, m, H11), 5.09 (2H, s, H18), 6.64 (1H, br d, J = 7.8 Hz), 7.34 (5H, br s, H17-19).

^{13}C NMR (101 MHz, CDCl$_3$): δ 13.8 (C1), 19.1 (C2), 22.0 (C22), 22.3 (C22), 27.1 (C12), 30.0 (C13), 38.3 (C3), 52.0 (C24), 52.9 (C11), 53.6 (C6), 54.1 (C21), 62.7 (C7), 66.7 (C15), 128.4 (C18), 138.6 (C19), 128.7 (C17), 135.8 (C16), 169.6 (C9), 171.0 (C14), 172.6 (C4), 174.1 (C23), 206.5 (C20).

Diastereomer 2: ^{1}H NMR (400 MHz, CDCl$_3$): δ 0.93 (3H, t, J = 7.5 Hz, H1), 1.21 (3H, s, H13), 1.23 (3H, s, H13), 1.62 (2H, m, H2), 1.80 (1H, m, H12), 2.14 (1H, m, H12), 2.23 (2H, t, J = 5.4 Hz), 2.43 (2H, m, H13), 3.61 (1H, m, H7), 3.68 (3H, s, H24), 4.03 (1H, m, H7), 4.41 (1H, m, H6), 5.10 (1H, m, H11), 5.09 (2H, s, H18), 6.62 (1H, br d, J = 7.8 Hz), 7.34 (5H, br s, H17-19).

^{13}C NMR (101 MHz, CDCl$_3$): δ 13.8 (C1), 19.1 (C2), 19.1 (C22), 20.2 (C22), 27.4 (C12), 30.3 (C13), 38.4 (C3), 52.0 (C24), 52.9 (C11), 53.6 (C6), 54.0 (C21), 63.0 (C7), 66.9 (C15), 128.4 (C18), 138.5 (C19), 128.7 (C17), 135.8 (C16), 169.6 (C9), 171.0 (C14), 172.6 (C4), 177.9 (C23), 213.2 (C20).

HRMS (ESI) calculated for C$_{24}$H$_{34}$N$_2$NaO$_8$ ([M+Na]$^+$): 501.2207; found: 501.2202.

6.1.2.35 Methyl (S)-8-(((benzyloxy)carbonyl)amino)-4-((tert-butoxycarbonyl)amino)-3-oxooctanoate (53h)
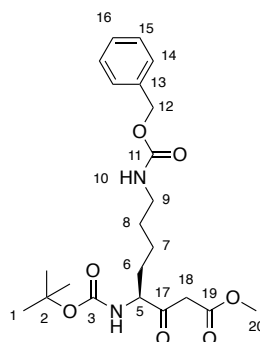
To a solution *N*-Boc-L-Lys(Z)-OH (**52h**) (0.761 g, 2.00 mmol) in THF (10 mL) was added CDI (0.486 g, 3.00 mmol, 1.5 equiv.) under argon. After stirring the reaction mixture at room temperature for 2 hours, methyl potassium malonate (0.465 g, 3.00 mmol, 1.5 equiv.) and $MgCl_2$ (0.284 g, 3.00 mmol, 1.5 equiv.) were added and the reaction was stirred overnight. The reaction was quenched with 1 M aqueous HCl (10 mL), THF was removed in *vacuo* and the aqueous layer was extracted with EtOAc (3 × 15 mL). Organic extractions were combined, washed with 5% $NaHCO_3$ (w/w in H_2O, 2 × 30 mL) and brine (30 mL), dried with $MgSO_4$ and filtered. The solvent was removed in *vacuo* to provide the desired product **53h** (0.471 g, 1.11 mmol, 56%).

R_f = 0.33 (EtOAc/hexane, 1:2, UV/$KMnO_4$).

^{1}H NMR (400 MHz, $CDCl_3$): δ 1.38 (2H, m, H7), 1.43 (9H, s, H1), 1.54 (2H, m, H8), 1.59 (1H, m, H6), 1.89 (1H, m, H6), 3.19 (2H, m, H9), 3.56 (2H, m, H18), 3.73 (3H, s, H20), 4.31 (1H, m, H5), 5.10 (2H, s, H12), 5.19 (1H, br d, *J* = 6.5 Hz, H4), 7.35 (5H, m, H14-16).

^{13}C NMR (101 MHz, $CDCl_3$): δ 22.3 (C7), 28.4 (C1), 29.6 (C8), 30.5 (C6), 40.5 (C9), 45.0 (C18), 52.6 (C20), 59.7 (C5), 66.8 (C12), 80.7 (C2), 128.3 (C15-16), 128.7 (C15), 136.7 (C13), 156.8 (C11), 167.4 (C17), 202.3 (C19).

HRMS (ESI) calculated for $C_{22}H_{32}N_2NaO_7$ ([M+Na]$^+$): 459.2102; found: 459.2105.

6.1.2.36 Methyl (S)-8-(((benzyloxy)carbonyl)amino)-4-((tert-butoxycarbonyl)amino)-2,2-dimethyl-3-oxooctanoate (54h)
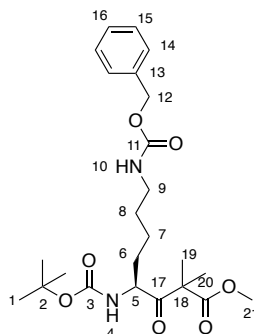
To a solution of **53h** (0.383 g, 0.877 mmol) in THF (10 mL), K_2CO_3 (1.21 g, 8.77 mmol, 10 equiv.), and methyl iodide (0.27 mL, 4.386 mmol, 5.0 equiv.) were added under an argon atmosphere. The reaction was stirred overnight and more K_2CO_3 (5.0 equiv.) and methyl iodide (2.5 equiv.) were added. The reaction was stirred overnight and more K_2CO_3 (5.0 equiv.) and methyl iodide (2.5 equiv.) were added. The reaction was stirred overnight, quenched with 1 M aqueous HCl (10 mL) and THF was removed in *vacuo*. The aqueous layer was extracted with EtOAc (3 × 15 mL) and the combined organic layers were washed with brine (2 × 25 mL), dried with $MgSO_4$, filtered, and concentrated in *vacuo* to give the desired product **54h** as yellow solid (0.397 g, 0.855 mmol, 97%).

R_f = 0.42 (EtOAc/hexane, 1:2, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.35 (2H, m, H7), 1.40 (3H, s, H18), 1.41 (9H, s, H1), 1.43 (3H, s, H18), 1.55 (2H, m, H8), 1.62 (2H, m, H6), 3.18 (2H, br d, J = 6.0 Hz, H9), 3.71 (3H, s, H, H21), 4.58 (1H, dt, J = 4.5, 10.5 Hz, H5), 4.90 (1H, br d, J = 9.0 Hz, H4), 5.09 (2H, s, H12), 5.33 (5H, m, H14-16).

^{13}C NMR (101 MHz, CDCl$_3$): δ 22.2 (C7), 22.6 (C19), 22.5 (C19), 28.4 (C1), 32.4 (C6), 40.8 (C9), 52.8 (C21), 54.8 (C18), 55.2 (C5), 66.8 (C12), 80.1 (C2), 128.2 (C15-16), 128.6 (C14), 135.4 (C3), 136.6 (C13), 157.1 (C11), 173.4 (C17), 207.8 (C20).

HRMS (ESI) calculated for $C_{24}H_{36}N_2NaO_7$ ([M+Na]$^+$): 487.2415; found: 487.2412.
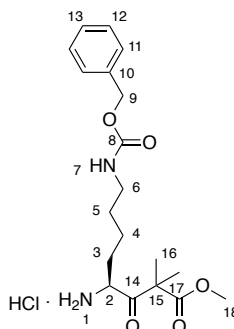
To **54h** (0.410 g, 0.882 mmol) was added 4 M HCl in dioxane (20 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 2 hours to provide the product **55h** as (0.353 g, 0.882 mmol, 100%).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.43 (3H, s, H16), 1.48 (3H, s, H16), 1.58 (4H, m, H4-5), 1.83 (1H, m, H3), 1.92 (1H, m, H3), 3.17 (2H, br s, H6), 3.70 (3H, s, H18), 4.56 (1H, br s, H2), 5.07 (2H, s, H9), 7.31 (5H, br s, H11-13), 8.67 (2H, br s, H1).

^{13}C NMR (101 MHz, CDCl$_3$): δ 21.5 (C4), 22.6 (C16), 22.9 (C16), 25.4 (C5), 28.7 (C3), 40.2 (C6), 53.1 (C18), 54.5 (C15), 56.1 (C2), 67.2 (C9), 128.0 (C13), 128.1 (C12), 128.6 (C11), 136.6 (C10), 156.9 (C17), 204.4 (C14).

HRMS (ESI) calculated for C$_{19}$H$_{29}$N$_2$O$_5$ ([M-H]$^-$): 365.2071; found: 365.2074.

6.1.2.38 Methyl (S)-8-(((benzyloxy)carbonyl)amino)-4-((S)-2-butyramido-3-hydroxypropanamido)-2,2-dimethyl-3-oxooctanoate (56h)
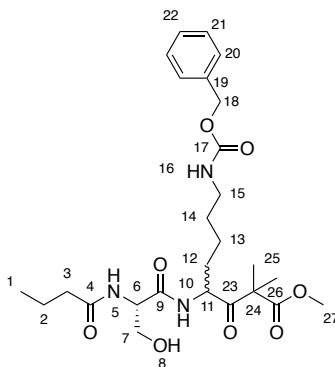
To a solution of **47** (0.188 g, 0.516 mmol) in THF (5 mL) was added NMI (0.13 mL, 1.6 mmol, 3.0 equiv.) and TCFH (0.23135 g, 0.823 mmol, 1.5- equiv.) and the solution was stirred for 15 minutes. Amine **55h** (0.103 g, 0.5892 mmol) was added

and the reaction stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (10 mL), THF concentrated in vacuo and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (3 × 40 mL), dried with $MgSO_4$, filtered, and concentrated in vacuo. The crude product was purified by silica gel chromatography (MeOH/DCM, 1:99 to 1:9) to yield the product **56h** as a yellow oil (0.0846 g, 0.162 mmol, 31%).

$R_f = 0.14$ (MeOH/DCM, 5:95, UV/$KMnO_4$).

Diastereomer 1: ^{1}H NMR (400 MHz, $CDCl_3$): δ 0.92 (3H, t, $J = 7.0$ Hz, H1), 1.28 (2H, m, H13), 1.39 (3H, s, H25), 1.41 (H25), 1.52 (3H, m, H12, H14), 1.65 (2H, m, H2), 1.86 (1H, m, H12), 2.21 (2H, m, H3), 3.16 (2H, br s, H15), 3.61 (1H, dd, J = 3.4, 10.2 Hz, H7), 3.73 (3H, s, H27), 4.06 (1H, m, H7), 4.44 (1H, br s, H6), 4.89 (1H, q, $J = 3.7$ Hz, H11), 5.10 (2H, s, H15), 6.65 (1H, br s, H5), 7.16 (1H, br s, H10), 7.25 (5H, br s, H20-22).

^{13}C NMR (101 MHz, $CDCl_3$): δ 13.8 (C1), 19.2 (C2), 22.2 (C25), 22.5 (C25), 29.3 (C14), 29.7 (C13), 31.2 (C12), 38.3 (C3), 40.6 (C15), 52.9 (C27), 53.7 (C6), 53.9 (C24), 54.1 (C11), 62.7 (C7), 66.9 (C18), 128.2 (C22), 128.3 (C21), 128.7 (C20), 137.9 (C19), 156.9 (C17), 171.9 (C26), 174.9 (C4), 206.3 (C23).

Diastereomer 2: ^{1}H NMR (400 MHz, $CDCl_3$): δ 0.94 (3H, t, $J = 7.2$ Hz, H1), 1.28 (2H, m, H13), 1.38 (3H, s, H25), 1.39 (H25), 1.52 (3H, m, H12, H14), 1.65 (2H, m, H2), 1.86 (1H, m, H12), 2.21 (2H, m, H3), 3.16 (2H, br s, H15), 3.61 (1H, dd, J = 3.4, 10.2 Hz, H7), 3.70 (3H, s, H27), 4.06 (1H, m, H7), 4.52 (1H, br s, H6), 4.89 (1H, q, $J = 3.7$ Hz, H11), 5.10 (2H, s, H15), 6.56 (1H, br s, H5), 7.08 (1H, br s, H10), 7.25 (5H, br s, H20-22)

^{13}C NMR (101 MHz, $CDCl_3$): δ 13.8 (C1), 19.2 (C2), 22.2 (C25), 22.5 (C25), 29.3 (C14), 29.7 (C13), 31.2 (C12), 38.3 (C3), 40.6 (C15), 52.9 (C27), 53.7 (C6), 53.9 (C24), 54.1 (C11), 62.7 (C7), 66.9 (C18), 128.2 (C22), 128.3 (C21), 128.7 (C20), 137.9 (C19), 156.9 (C17), 171.9 (C26), 174.9 (C4), 206.3 (C23).

HRMS (ESI) calculated for $C_{26}H_{39}N_3NaO_8$ ([M+Na]$^+$): 544.2629; found: 544.2618.

6.1.2.39 Benzyl N-(tert-butoxycarbonyl)-O-methyl-L-serinate (66)
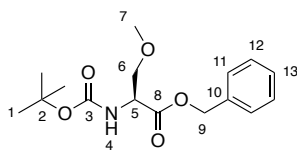
To a solution of *N*-Boc-*O*-methyl-L-serine (**65**) (0.052 g, 0.237 mmol) in anhydrous DCM (5 mL) was added Et₃N (0.04 mL, 0.3 mmol, 1.3 equiv.) and DMAP (0.002 g, 0.02 mmol, 0.1 equiv.) at 0 °C under argon. Benzyl chloroformate (0.04 mL, 0.3 mmol, 1.3 equiv.) was added slowly and the reaction mixture left to stir for 3 hours. The reaction was quenched by addition of brine (10 mL), the layers were separated, the aqueous layer was extracted with DCM (3 × 10 mL) and the combined organic layers were dried with MgSO₄, filtered, and concentrated in *vacuo*. The resulting residue was purified by silica gel chromatography (EtOAc/hexane, 1:9 to 1:1) to yield the desired product **66** (0.0617 g, 0.199 mmol, 87%).

R_f = 0.28 (EtOAc/hexane, 1:9, UV/KMnO₄).

¹H NMR (400 MHz, CDCl₃): δ 1.44 (9H, s, H1), 3.29 (3H, s, H7), 3.60 (1H, dd, *J* = 3.1, 9.4 Hz, H6), 3.81 (1H, dd, *J* = 2.8, 9.3 Hz, H6), 4.46 (1H, br d, *J* = 8.5 Hz, H5), 5.14 (1H, d, *J* = 12.4 Hz, H9), 5.27 (1H, d, *J* = 12.4 Hz, H9), 5.39 (1H, br d, *J* = 8.0 Hz, H4), 7.34 (5H, m, H11-13).

¹³C NMR (101 MHz, CDCl₃): δ 28.3 (C1), 54.1 (C5), 59.2 (C7), 67.1 (C9), 72.5 (C6), 80.0 (C2), 128.1 (C13), 128.3 (C12), 128.5 (C11), 135.5 (C10), 170.6 (C8).

HRMS (ESI) calculated for C₁₆H₂₃NNaO₅ ([M+Na]⁺): 332.1468; found: 332.1468.

6.1.2.40 Benzyl O-methyl-L-serinate (67)
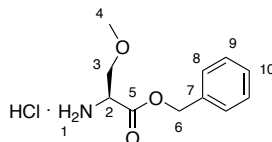
To *N*-Boc-methyl-L-serine-OBn (**66**) (0.0559 g, 0.181 mmol) was added 4 M HCl in dioxane (2 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 1 hour to provide the product **67** as colourless oil (0.0443 g, 0.180 mmol, 100%).

R_f = 0.32 (EtOAc/hexane, 1:1, KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 3.29 (3H, s, H4), 3.83 (2H, br s, H3), 4.27 (1H, s, H2), 5.15 (1H, d, J = 12.0 Hz, H6), 5.26 (1H, d, J = 12.0 Hz, H6), 7.33 (5H, m, H8-10), 7.91 (2H, br s, H1).

^{13}C NMR (101 MHz, CDCl$_3$): δ 53.7 (C2), 59.2 (C4), 68.7 (C6), 69.0 (C3), 128.5 (C10), 128.8 (C9), 129.0 (C8), 134.4 (C7), 167.6 (C5).

HRMS (ESI) calculated for C$_{11}$H$_{16}$NO$_3$ ([M+Na]$^+$): 210.1125; found: 210.1129.

6.1.2.41 Benzyl N-(N-(tert-butoxycarbonyl)-O-methyl-L-seryl)-O-methyl-L-serinate
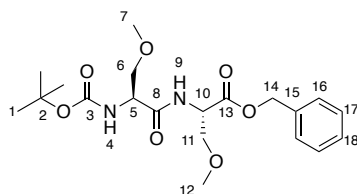
To a solution of *N*-Boc-*O*-methylserine (**65**) (0.0896 g, 0.409 mmol) in THF (3 mL) was added NMI (0.10 mL, 1.2 mmol, 3.0 equiv.) and TCFH (0.172 g, 0.612 mmol, 1.5 equiv.) and the solution was stirred for 15 minutes. Amine **67** (0.100 g, 0.408 mmol, 1.0 equiv.) was added and the reaction stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (5 mL), THF concentrated in *vacuo* and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (3 × 40 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (EtOH/hexane, 1:1) to yield the product as a colourless oil (0.0688 g, 0.168 mmol, 41%).

R_f = 0.39 (EtOAc/hexane, 1:1, UV/KMnO$_4$).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.45 (9H, s, H1), 3.29 (3H, s, H7), 3.34 (3H, s, H12), 3.47 (1H, m, H6), 3.60 (1H, m, H11), 3.77 (1H, m, H6), 3.84 (1H, m, H11), 4.28 (1H, br s, H5), 4.76 (1H, m, H10), 5.15 (1H, dd, J = 3.0, 12.3 Hz, H14), 5.25 (1H, dd, J = 2.2, 12.4 Hz, H14), 5.40 (1H, br s, H4), 7.21 (1H, br d, J = 7.3 Hz, H9), 7.34 (5H, m, H16-18).

^{13}C NMR (101 MHz, CDCl$_3$): δ 28.4 (C1), 52.9/53.1 (C10), 54.0 (C5), 59.2 (C12), 59.4 (C7), 67.4 (C14), 72.2 (C11), 72.3 (C6), 80.4 (C2), 128.2 (C18), 128.5 (C17), 128.7 (C16), 135.9 (C15), 170.4 (C13), 171.0 (C8).

HRMS (ESI) calculated for C$_{20}$H$_{30}$N$_2$NaO$_7$ ([M+Na]$^+$): 433.1945; found: 433.1941.

6.1.2.42   N-(N-(tert-butoxycarbonyl)-O-methyl-L-seryl)-O-methyl-L-serine (64)
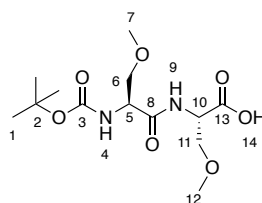
To a solution of di-protected di-*O*-methyl serine (0.221 g, 0.0538 mmol) in THF (3 mL) was added Pd/C (approximately 5 mg) and hydrogen gas was bubbled through the solution for 15 minutes. The reaction was stirred under hydrogen gas for 3 hours, filtered through celite, and the solvent was removed in *vacuo* to yield the product **64** (0.166 g, 0.0519 mmol, 96%).

^{1}H NMR (400 MHz, CDCl$_3$): δ 1.45 (9H, s, H1), 3.37 (3H, s, H7), 3.38 (3H, s, H12), 3.51 (1H, m, H6), 3.62 (1H, dd, *J* = 3.2, 9.5 Hz, H11), 3.80 (1H, dd, *J* = 4.2, 9.5 Hz, H6), 3.90 (1H, dd, *J* = 3.2, 5.1 Hz, H11), 4.33 (1H, m, H5), 4.71 (1H, m, H10), 5.49 (1H, br s, H4), 7.34 (1H, br d, *J* = 7.4 Hz, H9).

^{13}C NMR (101 MHz, CDCl$_3$): δ 28.4 (C1), 52.6 (C10), 52.8 (C5), 59.2 (C12), 59.5 (C7), 71.8 (C11), 72.1 (C6), 80.5 (C2), 170.7 (C13), 173.0 (C8).

HRMS (ESI) calculated for C$_{13}$H$_{24}$N$_2$NaO$_7$ ([M+Na]$^+$): 343.1476; found: 343.1476.

6.1.2.43   Methyl   (6S,9S,12S)-12-benzyl-6,9-bis(methoxymethyl)-2,2,14,14-tetramethyl-4,7,10,13-tetraoxo-3-oxa-5,8,11-triazapentadecan-15-oate
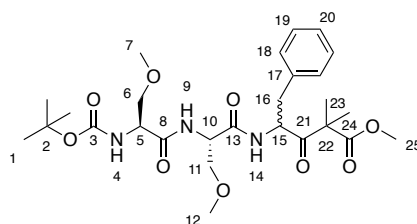
To a solution of **64** (0.151 g, 0.468 mmol) in THF (3 mL) was added NMI (0.11 mL, 1.4 mmol, 3.0 equiv.) and TCFH (0.197 g, 0.702 mmol, 1.5 equiv.) and the solution was stirred for 15 minutes. Amine **55d** (0.133 g, 0.468 mmol, 1.0 equiv.) was added and the reaction stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (5 mL), THF concentrated in *vacuo* and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (3 × 40 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (EtOH/hexane, 1:1) to yield the product (0.167 g, 0.303 mmol, 65%).

R$_f$ = 0.21 (EtOAc/hexane, 1:1, UV/KMnO$_4$).

Diastereomer 1: ^{1}H NMR (400 MHz, MeOD): δ 0.43 (3H, s, H23), 1.09 (3H, s, H23), 1.45 (9H, s, H1), 3.06 (2H, d, *J* = 3.8 Hz, H16), 3.33 (3H, s, H7), 3.35 (3H, s, H12), 3.61 (3H, m, H6, H11), 3.73 (3H, s, H25), 3.79 (1H, m, H6), 4.29 (1H, br s, H5), 4.42 (1H, t, *J* = 4.6 Hz, H15), 4.63 (1H, br s, H10), 7.21 (5H, m, H18-20).

^{13}C NMR (101 MHz, MeOD): δ 19.2 (C23), 21.5 (C23), 28.6 (C1), 37.9 (C16), 52.9 (C25), 54.1 (C10), 54.2 (C5), 59.3 (C12), 59.4 (C7), 63.9 (C15), 72.9 (C11), 73.2 (C6), 79.4 (C2), 124.4 (C22), 128.1 (C20), 129.4 (C19), 131.5 (C18), 135.2 (C17), 171.3 (C13), 173.1 (8), 178.7 (C24), 213.4 (C21).

Diastereomer 2: ^{1}H NMR (400 MHz, MeOD): δ 0.43 (3H, s, H23), 1.09 (3H, s, H23), 1.45 (9H, s, H1), 3.06 (2H, d, *J* = 3.8 Hz, H16), 3.33 (3H, s, H7), 3.35 (3H, s, H12), 3.61 (3H, m, H6, H11), 3.73 (3H, s, H25), 3.79 (1H, m, H6), 4.29 (1H, br s, H5), 4.42 (1H, t, *J* = 4.6 Hz, H15), 4.63 (1H, br s, H10), 7.21 (5H, m, H18-20).

^{13}C NMR (101 MHz, MeOD): δ 19.2 (C23), 21.5 (C23), 28.6 (C1), 37.9 (C16), 52.9 (C25), 54.1 (C10), 54.2 (C5), 59.3 (C12), 59.4 (C7), 63.9 (C15), 72.9 (C11), 73.2 (C6), 79.4 (C2), 124.4 (C22), 128.1 (C20), 129.4 (C19), 131.5 (C18), 135.2 (C17), 171.3 (C13), 173.1 (8), 178.7 (C24), 213.4 (C21).

HRMS (ESI) calculated for C$_{22}$H$_{33}$N$_3$NaO$_7$ ([M+Na]$^+$): 474.2211; found: 474.2208 (Boc group fell off during MS analysis).
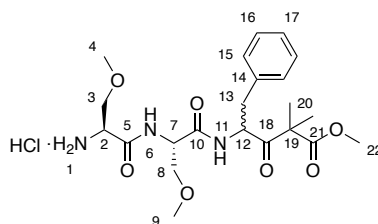
To di-protected di-*O*-methyl serine phenylalanine (0.127 g, 0.230 mmol) was added 4 M HCl in dioxane (3 mL) under argon and stirred at room temperature. The solvent was removed in *vacuo* after 1 hour to provide the product **63** as colourless oil (0.105 g, 0.214 mmol, 93%).

Diastereomer 1: ^{1}H NMR (400 MHz, MeOD): δ 0.44 (3H, s, H20), 1.09 (3H, s, H20), 3.06 (2H, d, *J* = 3.8 Hz, H13), 3.34 (3H, s, H4), 3.34 (3H, s, H9), 3.35 (3H, s, H22), 3.60 (1H, t, *J* = 5.3 Hz, H3), 3.68 (2H, m, H3, H8), 3.74 (1H, t, *J* = 6.1 Hz, H8), 4.15 (1H, m, H2), 4.42 (1H, t, *J* = 4.6 Hz, H12), 4.69 (1H, m, H7), 7.20 (5H, m, H15-17).

^{13}C NMR (101 MHz, MeOD): δ 18.5 (C20), 20.5 (C20), 37.2 (C13), 46.5 (C19), 52.5 (C22), 54.1 (C7), 54.2 (C2), 58.3 (C9), 58.4 (C4), 63.4 (C12), 72.9 (C8), 73.2 (C3), 127.4 (C17), 129.8 (C16), 130.5 (C15), 135.2 (C14), 171.3 (C10), 173.1 (5),178.7 (C21), 213.8 (C18).

Diastereomer 2: ^{1}H NMR (400 MHz, MeOD): δ 0.44 (3H, s, H20), 1.09 (3H, s, H20), 3.06 (2H, d, *J* = 3.8 Hz, H13), 3.34 (3H, s, H4), 3.34 (3H, s, H9), 3.35 (3H, s, H22), 3.60 (1H, t, *J* = 5.3 Hz, H3), 3.68 (2H, m, H3, H8), 3.74 (1H, t, *J* = 6.1 Hz, H8), 4.15 (1H, m, H2), 4.42 (1H, t, *J* = 4.6 Hz, H12), 4.69 (1H, m, H7), 7.20 (5H, m, H15-17).

^{13}C NMR (101 MHz, MeOD): δ 18.5 (C20), 20.5 (C20), 37.2 (C13), 46.5 (C19), 52.5 (C22), 54.1 (C7), 54.2 (C2), 58.3 (C9), 58.4 (C4), 63.4 (C12), 72.9 (C8), 73.2 (C3), 127.4 (C17), 129.8 (C16), 130.5 (C15), 135.2 (C14), 171.3 (C10), 173.1 (5),178.7 (C21), 213.8 (C18).

HRMS (ESI) calculated for $C_{22}H_{33}N_3NaO_7$ ([M+Na]$^+$): 474.2211; found: 474.2213.

6.1.2.45 Methyl (S)-4-((S)-3-methoxy-2-((S)-3-methoxy-2-(2-methylthiazole-5-carboxamido)propanamido)propanamido)-2,2-dimethyl-3-oxo-5-phenylpentanoate (61)
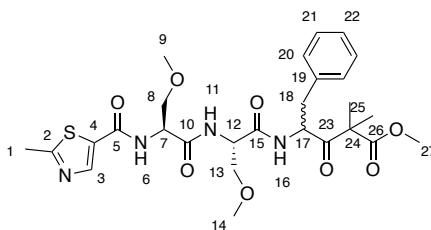
To a solution of thiazole **62** (0.0294 g, 0.205 mmol) in THF (3 mL) was added NMI (0.05 mL, 0.6 mmol, 3.0 equiv.) and TCFH (0.0863 g, 0.308 mmol, 1.5 equiv.) and the solution was stirred for 15 minutes. Amine **63** (0.102 g, 0.205 mmol, 1.0 equiv.) was added and the reaction stirred overnight. The reaction mixture was quenched with 1 M aqueous HCl (5 mL), THF concentrated in *vacuo* and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (3 × 40 mL), dried with MgSO$_4$, filtered, and concentrated in *vacuo*. The crude product was purified by silica gel chromatography (MeOH/DCM, 2:98 to 95:5) to yield the product **61** as white solid (0.0120 g, 0.0207 mmol, 10%).

R$_f$ = 0.25 (MeOH/DCM, 5:95, UV/KMnO$_4$).

Diastereomer 1: ^{1}H NMR (400 MHz, MeOD): δ 1.31 (3H, s, H25), 1.34 (3H, s, H25), 2.71 (3H, s, H1), 2.79 (1H, m, H18), 3.07 (1H, m, H18), 3.27 (3H, s, H14), 3.31 (3H, s, H9), 3.41 (3H, s, H27), 3.50 (1H, m, H13), 3.56 (1H, m, H13), 3.69 (1H, m, H8), 3.81 (1H, m, H8), 4.44 (1H, m, H12), 4.82 (1H, m, H7), 5.11 (1H, m, H17), 7.19 (5H, m, H20-22), 8.25 (1H, s, H3).

^{13}C NMR (101 MHz, MeOD): δ 19.1 (C1), 22.2 (C25), 22.4 (C25), 38.9 (C18), 54.7 (C12), 54.8 (C7), 55.5 (C17), 56.1 (C27), 56.5 (C24), 59.2 (C14), 59.4 (C9), 72.8 (C13), 72.9 (C8), 127.8 (C22), 129.4 (C21), 130.5 (C20), 135.8 (C4), 138.2 (C19), 144.6 (C3), 171.9 (C15), 173.0 (C10), 173.2 (C2), 174.4 (C26), 207.2 (C23).

Diastereomer 2: ^{1}H NMR (400 MHz, MeOD): δ 1.30 (3H, s, H25), 1.33 (3H, s, H25), 2.71 (3H, s, H1), 2.79 (1H, m, H18), 3.07 (1H, m, H18), 3.23 (3H, s, H14), 3.33 (3H, s, H9), 3.39 (3H, s, H27), 3.50 (1H, m, H13), 3.56 (1H, m, H13), 3.69 (1H, m,

H8), 3.81 (1H, m, H8), 4.44 (1H, m, H12), 4.82 (1H, m, H7), 5.11 (1H, m, H17), 7.19 (5H, m, H20-22), 8.27 (1H, s, H3).

^{13}C NMR (101 MHz, MeOD): δ 19.1 (C1), 22.2 (C25), 22.4 (C25), 38.9 (C18), 54.7 (C12), 54.9 (C7), 55.7 (C17), 56.3 (C27), 56.5 (C24), 59.3 (C14), 59.5 (C9), 72.8 (C13), 72.9 (C8), 127.7 (C22), 129.5 (C21), 130.5 (C20), 135.8 (C4), 138.2 (C19), 144.5 (C3), 171.9 (C15), 173.0 (C10), 173.2 (C2), 174.4 (C26), 207.2 (C23).

HRMS (ESI) calculated for $C_{27}H_{36}N_4NaO_8S$ ([M+Na]$^+$): 599.2146; found: 599.2135.

6.2 Biology

6.2.1 Materials and instruments

All broth and medium ingredients were purchased from Sigma Aldrich, Thermo Scientific, Difco, and Bacton Dickinson Microbiology unless otherwise stated. Restriction enzymes and buffers, polymerases, DNA loading dye and markers, Gel Red, DNA ligase, and buffers were purchased from New England Biolabs and Cambridge Bioscience unless otherwise stated. GeneJET Gel Extraction Kit, GeneJET PCT Purification Kit, and GeneJET Plasmid Miniprep Kit were purchased from Thermo Scientific. The Champion™ pET151 Directional TOPO® Expression Kit was purchased from Invitrogen. MagAttract® High Molecular Weight DNA extraction kit was purchased from QIAGEN. The FastDNA™ SPIN Kit for Soil was purchased from Fisher Scientific.

Media was autoclaved at 121 °C and 1 bar for 20 minutes. Ministart syringe filters (Sartosius) were used for sterile filtering. *E. coli* cultures were incubated in New Brunswick Scientific Innova shakers and enzyme assays in Thermo Scientific Heraeus static incubators. Ampicillin stock solution was used at 100 mg/mL in water and prepared and sterilised by filtration through 0.2 μm syringe filter or obtained from the tech team preparation room in the School of Life Sciences. PCR reactions were carried out on an Eppendorf Mastercycler Personal. Agarose gel electrophoresis were performed in a BioRad tank PowerPac 300 and visualised with a UVP imaging system UV transilluminator. Optical densities were measured using a Thermo Scientific BioMate 3 spectrophotometer. Cultures and samples were centrifuged using Eppendorf models 5415D, 5424, 5804R, and 5810R or a Thermo Scientific Sorvall RC 6 Plus equipped with an SS34 or SLA3000 rotor. Cell lysis was conducted using a Constant Systems Ltd TS-Series Cabinet cell disruptor (One-Shot mode, 20 KPSI). Proteins were visualised using acrylamide SDS-PAGE gel electrophoresis performed in a BioRad tank and a PowerPacTM machine. Both protein and DNA concentrations were determined using a Thermo Scientific NanoDrop Lite spectrophotometer. UHPLC-ESI-Q-TOF-MS analysis of purified proteins was performed by Dr Chuan Huang and Dr Matt Jenner on a Dionex UltiMate 3000 RS UHPLC connected to an ACE 3 C4-300 reverse phase column (Advanced Chromatography Technologies, Aberdeen, UK; 100 × 2.1 mm, 5 μM, 30 °C) coupled to a Bruker maXis II mass

spectrometer. Samples from enzyme assays were analysed using a Bruker amaZon speed ETD instrument and a C18 column (100 × 2.1 mm, 1.8 μm) with a gradient from 5-100% acetonitrile in water each supplemented with 0.1% formic acid at a flow rate of 0.2 mL/min.

6.2.2   Media compositions and preparation

6.2.2.1 Luria-Bertani medium (LB)

Either LB broth powder (25 g), or separate components of yeast extract (5 g), tryptone (10 g), and NaCl (10 g), were dissolved in 1 L dH$_2$O, Bacto agar (15 g, only for agar preparation) was added and everything sterilised by autoclaving. For experiments carried out in the School of Life Sciences, LB broth and agar plates were prepared by the technical team and obtained from their preparation room.

6.2.2.2 Glucose yeast malt medium (GYM)

Glucose (4.0 g), yeast extract (4.0 g), malt extract (10.0 g), CaCO$_3$ (2.0 g), and agar (12.0 g, only for agar preparation) were dissolved in 1 L dH$_2$O and pH was adjusted to 7.2 before adding agar and sterilizing by autoclaving.

6.2.2.3 Malt yeast glucose (MYG)

Same as GYM but without CaCO$_3$ and without pH adjustment.

6.2.2.4 ISP4 medium

Soluble starch (10 g), MgSO$_4$*7H$_2$O (1 g), NaCl (1 g), (NH$_4$)$_2$SO$_4$ (2 g), CaCO$_3$ (2 g), and agar (20 g, only for agar preparation) were dissolved in 1 L dH$_2$O and sterilised by autoclaving. Trace element solution (0.1 g FeSO$_4$*7H$_2$O, 0.1 g MnCl$_2$*4H$_2$O, 0.1 g ZnSO$_4$*7H$_2$O in 100 mL dH$_2$O) was filter sterilised and added (1 mL).

6.2.2.5 Soya flour mannitol (SFM)

Mannitol (20 g), soya flour (20 g), and agar (20 g) were dissolved in 1 L tap water and sterilised by autoclaving.

6.2.2.6 Tryptic soy broth (TSB)

TSB powder was purchased from Sigma Aldrich and was prepared according to supplier's instruction.

6.2.2.7 Supplemented minimal medium (SMM):

The following solutions were added to a total volume of 1 L dH$_2$O:

- 81.9 mL PEG 6000 (6.1% w/v in dH$_2$O)
- 2.5 mL Mg$_2$SO$_4$*7H$_2$O, 5 mM final
- 10 mL TES buffer (0.25 M, pH 7.2), 25 mM final
- 1 mL NaH$_2$PO$_4$ + K$_2$HPO$_4$ (50 mM each), 1 mM final
- 2 mL glucose (50% w/v)
- 25 mL glycine
- Sterilise by autoclaving and add 1 mL of sterile trace element solution (0.1 g FeSO$_4$*7H$_2$O, 0.1 g MnCl$_2$*4H$_2$O, 0.1 g ZnSO$_4$*7H$_2$O, 0.1 g CaCl$_2$*6H$_2$O, NaCl in 100 mL dH$_2$O).

6.2.3   Buffer compositions and preparation

6.2.3.1 SDS-PAGE sample buffer (5×)

SDS (10% w/v), 2-mercaptoethanol (10 mM), glycerol (20% w/v), Tris-base (0.2 M), and bromophenol blue (0.05% w/v) was dissolved in deionised H$_2$O and the pH adjusted to 6.8. The solution was diluted to 1 × with protein sample when used.

6.2.3.2 SDS-PAGE running buffer (10×)

Tris-base (250 mM), glycine (2 M), and SDS (1% w/v) were dissolved in deionised H$_2$O and the pH adjusted to 8.0. The solution was diluted to 1 × with protein sample when used.

6.2.3.3 Tris/Borate/EDTA (TBE) buffer (10×)

Tris-HCl (890 mM), boric acid (890 mM), and EDTA (20 mM) were dissolved in deionised H_2O. The solution was diluted to $1 \times$ before use.

6.2.4 Bacterial strains

The following strains were used:

- *E. coli* BL21 Star™ (DE3) One Shot® chemically competent cells (Invitrogen, high-level recombinant protein expression)
- *E. coli* BL21 (DE3) One Shot® chemically competent cells (lab stock)
- *E. coli* One Shot® TOP10 chemically competent cells (Invitrogen, transformation, replication, and DNA isolation)
- *Streptomyces sparsogenes* ATCC 25498 (German Collection of Microorganisms and Cell Cultures, DSMZ 40356)
- *Streptomyces* species maeda85 (provided by Dr Yuta Tsunematsu)

6.2.5 Plasmids

The following plasmids were used:

- pET151 (Invitrogen, overexpression of proteins with an *N*-terminal His$_6$-tag)
- pET151-*epnF* (provided by Dr Chuan Huang)
- pET151-*ttpC* (assembled with Invitrogen Champion™ pET151 Directional TOPO® Expression Kit)
- pET151-*tptC* (assembled with Invitrogen Champion™ pET151 Directional TOPO® Expression Kit)

6.2.6 Primers

PCR primers were designed based on sequence information from GeneBank deposits and obtained MinION sequencing data or designed according to provider's kit instructions. All primers were obtained from Integrated DNA Technologies.

- *ttpC* forward primer: CACCATGTCTCGTCCGCTCAACGCC
- *ttpC* reverse primer: CGACCCGTCGGTGACGATGTC
- *tptC* forward primer: CACCTCTCGTCCGCTCAACGCCATC

- *tptC* reverse primer: CGACACGTCGGGGCCGGACGC

6.2.7 Experimental procedures

6.2.7.1 Chemical competent cells

An overnight culture of *E. coli* BL21 (DE3) in LB media was set up and incubated at 37 °C. The next day, 100 mL autoclaved LB media was inoculated with 1 mL of the overnight culture and incubated at 37 °C until optical density 0.65. The cultures were span down (4 °C, 4000 rpm, 10 min) and the supernatant discarded. The cell pellet was resuspended in 30 mL washing solution (30 mM CH_3COOK, 80 mM $CaCl_2$, pH 5.8), centrifuged (4 °C, 4000 rpm, 10 min), supernatant was discarded, and the washing procedure repeated one more time. The cell pellet was resuspended in 1 mL storage solution (80 mL washing solution + 14 mL glycerol) on ice and stored in aliquots of 50 µL at – 80 °C.

6.2.7.2 Heat-shock transformation

One aliquot of *E. coli* BL21 (DE3) chemical competent cells was thawed on ice. Plasmid pET151 harbouring the gene of interest was added to the cells (1 µL), mixed and kept on ice for 30 minutes. The cells were heat shocked at 42 °C for 45 seconds and put back on ice for 2 minutes. LB media (450 µL) was added, and the cells were incubated at 37 °C for 1 hour. The cells (20 µL and 200 µL) were plated on LB agar with ampicillin and incubated at 37 °C overnight. The next day, 4 colonies were picked and grown in LB media with ampicillin (100 µg/mL) overnight and purification with a plasmid miniprep kit (GeneJET, Thermo Fisher) yielded the pure plasmids. The plasmid was confirmed to be the correct assembly by restriction digest and sequencing of the incorporation region of the gene of interest.

6.2.7.3 Restriction digest for plasmid characterisation

Digestion reactions were set up as follows and incubated for 1 hour at 37 °C: milli Q water (7.5 µL), appropriate restriction enzyme (0.5 µL), buffer according to the enzyme (1 µL), purified plasmid (1 µL). Gel loading dye (2 µL) was added to the digestion reactions and the mixtures were loaded and run on an agarose gel (1%). The

digest pattern was compared to the predicted pattern and fragment sizes of the target plasmid.

6.2.7.4 EpnF overproduction and purification

Overnight cultures of *E. coli* BL21 (DE3) containing pET151 harbouring EpnF were grown in LB with ampicillin (100 µg/mL) and 1 mL was inoculated in each 2 × 1 L LB media with ampicillin (100 µg/mL). The cultures were incubated at 37 °C until an optical density of 0.75 was reached. The cultures were cooled to 15 °C, IPTG (100 µM) and riboflavin (500 µM) was added, and the cultures were incubated at 15 °C overnight. The cells were harvested by centrifugation (4000 rpm, 20 min, 4 °C), resuspended in 15 mL wash buffer (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, 20 mM imidazole, pH 8) and either stored at − 80 °C or lysed with a cell disruptor. The lysate was spun down (17000 rpm, 60 min, 4 °C) and the supernatant was loaded onto a Ni-column while the flow through was collected. The column was washed with wash buffer (10 mL) and the desired protein was eluted with elution buffer of increasing imidazole concentration (100 mM, 200 mM and 300 mM in 20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, 3 mL each). The fractions of the elution were concentrated to a volume of < 2.5 mL with a 30 kDa spin filter and desalted with a PD-10 desalting column and concentration buffer (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 8). Further concentration yielded 20.6 mg/L EpnF stored in 10% glycerol aliquots of 0.34 mM.

6.2.7.5 Enzyme assays

To screen for conditions, incubation times of 1 hour, 3 hours, and 6 hours and incubation temperatures of 4 °C, 15 °C, 20 °C, 30 °C, and 37 °C were tested with the phenylalanine derived substrate analogue as well as PLE/EpnF concentrations of 10µM/100µM, 10µM/50µM and 5µM/50µM and the necessity of FAD (2 mM). Final assays were set up to a total volume of 200 µL in concentration buffer and incubated for 3 hours at 30 °C. The PLE assay contained 1 mM substrate and 5 µM PLE and a negative control was set up without the PLE. The coupled PLE/EpnF assay contained 1 mM substrate, 5 µM PLE, 5 µM EpnF, and 2 mM FAD and a negative control was set up the same way but without EpnF. The assays were quenched by adding 200 µL

acetonitrile + 0.1% formic acid to 100 μL assay. The quenched assay was spun down, and the supernatant filtered and analysed in positive mode LCMS.

6.2.7.6 Streptomyces sparsogenes ATCC 25498 (DSM 40356) reactivation

Streptomyces sparsogenes was ordered as freeze-dried cell pellet from the German Collection of Microorganisms and Cell Cultures (DSMZ, order number DSM 40356). The ampoule was opened according to the DSM protocol "Opening of ampoules and rehydration of dried cultures". Half of the cell pellet was stored in a sterile cryogenic tube at − 80 °C and 0.5 mL sterile water was added to the other half of the pellet to rehydrate it for 30 minutes at room temperature. Of the rehydrated culture, 50 μL each was spread out on 2 plates each with GYM, ISP4, and SFM (one made with tap water, one made with distilled water) agar and incubated at 30 °C. White and grey cultures were visible after 3 days of incubation and one plate of each medium was stored in the fridge, one plate of each medium was incubated for another 7 days before preparing spore stocks.

6.2.7.7 Spore stocks

Spore stocks were taken from either GYM or MYG agar plates of cultures incubated for 10 days at 30 °C. Sterile water (10 mL) was poured onto the surface of bacterial lawn and spores were scraped off using a sterile wedge-spreader. The spore suspension was transferred with a 10 mL sterile serological pipette into a sterile 5 mL syringe barrel fitted with sterile cotton wool. The suspension was pushed through the cotton wool with a sterile plunger into a sterile falcon tube and centrifuged (4000 rpm, 5 minutes). The supernatant was discarded, and the pelleted cells resuspended in 1 mL sterile water with glycerol (25% w/v). Spore stocks were stored at − 80 °C.

6.2.7.8 Genomic DNA extraction for cloning ttpC

The Fisher Scientific FastDNA™ SPIN Kit for Soil was used to extract genomic DNA from *S. sparsogenes* for cloning experiments. The supplier's protocol was followed, and DNA eluted in 100 μL DES buffer (provided by supplier) in 73.1 ng/μL.

6.2.7.9 Genomic DNA extraction for sequencing and cloning tptC

High molecular weight DNA was extracted using the QIAGEN MagAttract® HMW DNA Kit. Spore stocks were inoculated in TSB medium (10 µL in 10 mL medium) in a sterile falcon tube and incubated at 30 °C and 180 rpm for 3 days. Cultures (5 mL) were centrifuged (4000 rpm, 10 minutes), the supernatant was discarded, and the pellet was used according to the supplier's protocol. The protocol "Manual Purification of High-Molecular-Weight Genomic DNA from Gram-Positive Bacteria" by QIAGEN was followed, and DNA eluted in 400 µL nuclease free water at 26.2 ng/µL for *S. sparsogenes* and in 100 µL nuclease free water at 149 ng/µL for *Streptomyces* sp. maeda85.

6.2.7.10 Nanopore sequencing and genome assembly and annotation

The Oxford nanopore technologies protocol "Native barcoding genomic DNA kit" (EXP-NBD104 and EXP-NBD114) and "Ligation Sequencing Kit" (SQK-LSK109) were used alongside the long fragment buffer to prepare the sequencing library from the isolated genomic DNA. Sequencing was performed on a MinION Mk1B flow cell and guppy 4.4.1+1c81d6233 was used for basecalling. The sequencing data was assembled with Flye 2.8.2-b1691 and polished four times with Minimap2 2.11-r797 and Racon v1.4.20. Final polishing was achieved with Medaka 1.2.1 and BUSCO 5.0.0 was used to check the assembly. Prokka was used in a conda environment for genome annotation and checking the annotation with BUSCO provided a result of 95.2% for *S. sparsogenes* and 98.8% for *Streptomyces* sp. maeda85. For the genome of *S.* sp. maeda85, assembled and polished nanopore data was additionally polished with short-read Illumina sequencing data provided by Dr Yuta Tsunematsu. Bowtie2 version 2.4.2 and Pilon version 1.23 were used to yield an improved full genome that was annotated with Prokka to provide a BUSCO score of 99.2%.

6.2.7.11 Metabolite expression and extraction

Liquid media: 5 µL spore stocks was used to inoculate 50 mL MYG medium, then incubated at 30 °C and 180 rpm for either 5 or 10 days. The growth was conducted in triplicate and a control of 50 mL MYG medium was used.

Solid media: 5 µL spore stock was spread on MYG agar plates and incubated at 30 °C for 5 or 10 days. The growth was conducted in triplicate and a control MYG agar plate was used.

After the incubation period, the lawn of the plate cultures and the surface of the control plate was scraped into 50 mL distilled water and treated like the liquid cultures from then on. Cultures and controls were acidified to pH 3 by dropwise addition of 1 M HCl and extracted with ethyl acetate (3 × 50 mL) and the combined organic layers were dried over MgSO₄, filtered, and concentrated in *vacuo*. The extract was redissolved in 1.5 mL HPLC grade acetonitrile/water (1:1) and filtered prior to HRMS analysis.

6.2.7.12 *PCR amplification PCR for ttpC and tptC cloning*

To amplify the TtpC encoding gene, a PCR reaction was set up with DNA extracted by the soil kit and *ttpC* primers (12.5 µL Q5 HF polymerase, 1.25 µL forward primer, 1.25 µL reverse primer, 2 µL genomic DNA, 8 µL nuclease free water) and incubated in a thermocycler (98 °C for 60 seconds, 35 cycles of [98 °C for 10 seconds, 73 °C for 20 seconds, 72 °C for 25 seconds], 72 °C for 180 seconds, hold at 4 °C). The PCR reaction was purified by DNA gel electrophoresis and extraction with the Thermo Scientific GeneJET Gel Extraction Kit according to the supplier's protocol providing 15 µL of the desired gene amplification product in 25.1 ng/µL.

6.2.7.13 *ttpC and tptC cloning reaction*

The cloning of *ttpC* and *tptC* each into pET151 was achieved with the Invitrogen Champion™ pET151 Directional TOPO® Expression Kit. The supplier's materials and protocols were followed to yield *E. coli* BL21 Star™ (DE3) One Shot® cells with expression vector pET151-*ttpC* and pET151-*tptC* respectively. Plasmid purification, restriction digest, and sequencing with T7 primers confirmed the successful cloning of pET151-ttpC and pET151-*tptC* respectively.

6.2.7.14 *TtpC and TptC overproduction and purification*

Overnight cultures of *E. coli* BL21 Star™ (DE3) One Shot® containing pET151 harbouring each *ttpC* and *tptC* were grown in LB with ampicillin (100 µg/mL) and

10 mL were used to inoculate inoculated in each 2 × 1 L LB media with ampicillin (100 µg/mL). The cultures were incubated at 37 °C until $OD_{600} = 0.75$ was reached. The cultures were cooled to 15 °C, IPTG (500 µM) and riboflavin (500 µM) was added, and the cultures were incubated at 15 °C overnight. The cells were harvested by centrifugation (4000 rpm, 20 min, 4 °C), resuspended in 15 mL wash buffer (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, 20 mM imidazole, pH 8), and lysed by sonication on ice (10 minutes total sonication time, 30 seconds on, 60 seconds off cycle) for TptD and by cell disrupter for TtpC. The lysate was centrifuged (17000 rpm, 60 min, 4 °C) and the supernatant was loaded onto a Ni-column while the flow through was collected. The column washed with wash buffer (10 mL) and the desired protein was eluted with elution buffer (200 mM imidazole in 20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, 3 mL each). The fractions of the elution were concentrated to a volume of < 2.5 mL with a 30 kDa spin filter (expected protein size is 68.7 kDa) and desalted with a PD-10 desalting column and concentration buffer (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 8). The desalted protein was stored in 10% glycerol aliquots at − 80 °C.

6.2.7.15 TtpC and TptC enzyme assyays

Assays were set up to a total volume of 200 µL in concentration buffer and incubated for 3 hours at 30 °C. The PLE assay contained 1 mM substrate and 5 µM PLE and a negative control was set up without the PLE. The coupled PLE/EpnF assay contained 1 mM substrate, 5 µM PLE, 50 µM TtpC or TptC, and 2 mM FAD and a negative control was set up the same way but without TtpC/TptC. The assays were quenched by adding 100 µL acetonitrile + 0.1% formic acid to 50 µL assay. The quenched assay was centrifuged (14000 rpm, 10 minutes), and the supernatant was filtered and analysed by HRMS.

7. References

1   M. Groettrup, C. J. Kirk and M. Basler, *Nat. Rev. Immunol.*, 2010, **10**, 73–78.

2   K. Tanaka, *Proc. Japan Acad. Ser. B Phys. Biol. Sci.*, 2009, **85**, 12–36.

3   G. C. Lander, E. Estrin, M. E. Matyskiela, C. Bashore, E. Nogales and A. Martin, *Nature*, 2012, **482**, 186–191.

4   S. H. Lecker, A. L. Goldberg and W. E. Mitch, *J. Am. Soc. Nephrol.*, 2006, **17**, 1807–1819.

5   A. Ichihara and K. Tanaka, *Mol. Biol. Rep.*, 1995, **21**, 49–52.

6   P. C. A. Da Fonseca and E. P. Morris, *J. Biol. Chem.*, 2008, **283**, 23305–23314.

7   G. A. Collins and A. L. Goldberg, *Cell*, 2017, **169**, 792–806.

8   A. F. Kisselev and A. L. Goldberg, *Chem. Biol.*, 2001, **8**, 739–758.

9   W. Chen, C. C. Norbury, Y. Cho, J. W. Yewdell and J. R. Bennink, *J. Exp. Med.*, 2001, **193**, 1319–1326.

10   H. Kimura, P. Caturegli, M. Takahashi and K. Suzuki, *J. Immunol. Res.*, 2015, **2015**, 541984.

11   Z. Cui, S. M. Hwang and A. V. Gomes, *Mol. Cell. Biol.*, 2014, **34**, 96–109.

12   I. E. Keller, O. Vosyka, S. Takenaka, A. Kloß, B. Dahlmann, L. I. Willems, M. Verdoes, H. S. Overkleeft, E. Marcos, S. Adnot, S. M. Hauck, C. Ruppert, A. Günther, S. Herold, S. Ohno, H. Adler, O. Eickelberg and S. Meiners, *Sci. Rep.*, 2015, **5**, 1–12.

13   Z. Elkoshi, *Front. Immunol.*, 2022, **13**, 1–13.

14   S. K. Kumar, V. Rajkumar, R. A. Kyle, M. Van Duin, P. Sonneveld, M. V. Mateos, F. Gay and K. C. Anderson, *Nat. Rev. Dis. Prim.*, 2017, **3**, 1–20.

15   L. Wang, F. S. Wang and M. E. Gershwin, *J. Intern. Med.*, 2015, **278**, 369–395.

16   E. E. Manasanch and R. Z. Orlowski, *Nat. Rev. Clin. Oncol.*, 2017, **14**, 417–433.

17   M. Basler, S. Mundt, A. Bitzer, C. Schmidt and M. Groettrup, *Clin. Exp. Rheumatol.*, 2015, **33**, 74–79.

18   A. Gulla and K. C. Anderson, *Haematologica*, 2020, **105**, 2358–2367.

19   A. J. Cowan, C. Allen, A. Barac, H. Basaleem, I. Bensenor, M. P. Curado, K. Foreman, R. Gupta, J. Harvey, H. Dean Hosgood, M. Jakovljevic, Y. Khader, S. Linn, D. Lad, L. Mantovani, V. M. Nong, A. Mokdad, M. Naghavi, M.

Postma, G. Roshandel, K. Shackelford, M. Sisay, C. T. Nguyen, T. T. Tran, B. T. Xuan, K. N. Ukwaja, S. E. Vollset, E. Weiderpass, E. N. Libby and C. Fitzmaurice, *JAMA Oncol.*, 2018, **4**, 1221–1227.

20   S. V. Rajkumar, M. A. Dimopoulos, A. Palumbo, J. Blade, G. Merlini, M. V. Mateos, S. Kumar, J. Hillengass, E. Kastritis, P. Richardson, O. Landgren, B. Paiva, A. Dispenzieri, B. Weiss, X. LeLeu, S. Zweegman, S. Lonial, L. Rosinol, E. Zamagni, S. Jagannath, O. Sezer, S. Y. Kristinsson, J. Caers, S. Z. Usmani, J. J. Lahuerta, H. E. Johnsen, M. Beksac, M. Cavo, H. Goldschmidt, E. Terpos, R. A. Kyle, K. C. Anderson, B. G. M. Durie and J. F. S. Miguel, *Lancet Oncol.*, 2014, **15**, e538–e548.

21   K. C. Anderson, *Clin. Cancer Res.*, 2016, **22**, 5419–5427.

22   D. Kazandjian and O. Landgren, *Semin. Oncol.*, 2016, **43**, 682–689.

23   G. K. Raju, K. Gurumurthi, R. Domike, D. Kazandjian, O. Landgren, G. M. Blumenthal, A. Farrell, R. Pazdur and J. Woodcock, *Clin. Pharmacol. Ther.*, 2018, **103**, 67–76.

24   M. Attal, V. Lauwers-Cances, C. Hulin, X. Leleu, D. Caillot, M. Escoffre, B. Arnulf, M. Macro, K. Belhadj, L. Garderet, M. Roussel, C. Payen, C. Mathiot, J. P. Fermand, N. Meuleman, S. Rollet, M. E. Maglio, A. A. Zeytoonjian, E. A. Weller, N. Munshi, K. C. Anderson, P. G. Richardson, T. Facon, H. Avet-Loiseau, J.-L. Harousseau and P. Moreau, *N. Engl. J. Med.*, 2017, **376**, 1311–1320.

25   P. L. McCarthy, S. A. Holstein, M. T. Petrucci, P. G. Richardson, C. Hulin, P. Tosi, S. Bringhen, P. Musto, K. C. Anderson, D. Caillot, F. Gay, P. Moreau, G. Marit, S. H. Jung, Z. Yu, B. Winograd, R. D. Knight, A. Palumbo and M. Attal, *J. Clin. Oncol.*, 2017, **35**, 3279–3289.

26   S. V. Rajkumar, *Am. J. Hematol.*, 2016, **91**, 719–734.

27   G. E. Dinse, C. G. Parks, C. R. Weinberg, C. A. Co, J. Wilkerson, D. C. Zeldin, E. K. L. Chan and F. W. Miller, *Arthritis Rheumatol.*, 2020, **72**, 1026–1035.

28   K. P. Liao, F. Kurreeman, G. Li, G. Duclos, S. Murphy, R. Guzman, T. Cai, N. Gupta, V. Gainer, P. Schur, J. Cui, J. C. Denny, P. Szolovits, S. Churchill, I. Kohane, E. W. Karlson and R. M. Plenge, *Arthritis Rheum.*, 2013, **65**, 571–581.

29   F. W. Miller, L. Alfredsson, K. H. Costenbader, D. L. Kamen, L. M. Nelson, J. M. Norris and A. J. De Roos, *J. Autoimmun.*, 2012, **39**, 259–271.

30   A. Davidson and B. Diamond, *N. Engl. J. Med.*, 2001, **345**, 340–350.

31   P. M. Kloetzel, *Nat. Rev. Mol. Cell Biol.*, 2001, **2**, 179–188.

32   I. Shachar and N. Karin, *J. Leukoc. Biol.*, 2013, **93**, 51–61.

33   Y. Wang, W. Zhou and Z. Zhang, *Rheumatol. (United Kingdom)*, 2015, **54**, 194–195.

34   K. Puttaparthi and J. L. Elliott, *Exp. Neurol.*, 2005, **196**, 441–451.

35   T. Egerer, L. Martinez-Gamboa, A. Dankof, B. Stuhlmüller, T. Dörner, V. Krenn, K. Egerer, P. E. Rudolph, G. R. Burmester and E. Feist, *Arthritis Rheum.*, 2006, **54**, 1501–1508.

36   A. Visekruna, N. Slavova, S. Dullat, J. Gröne, A. J. Kroesen, J. P. Ritz, H. J. Buhr and U. Steinhoff, *Int. J. Colorectal Dis.*, 2009, **24**, 1133–1139.

37   H. W. B. Johnson, E. Lowe, J. L. Anderl, A. Fan, T. Muchamuel, S. Bowers, D. C. Moebius, C. Kirk and D. L. McMinn, *J. Med. Chem.*, 2018, **61**, 11127–11143.

38   D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461–477.

39   D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2012, **75**, 311–335.

40   H. Yuan, Q. Ma, L. Ye and G. Piao, *Molecules*, 2016, **21**, 559.

41   D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2020, **83**, 770–803.

42   A. Flemming, *J. Exp. Pathol.*, 1929, **10**, 226–236.

43   M. R. Montinari, S. Minelli and R. De Caterina, *Vascul. Pharmacol.*, 2019, **113**, 1–8.

44   H. Umezawa, M. Ishizuka, K. Maeda and T. Takeuchi, *Cancer*, 1967, **20**, 891–895.

45   M. Nett, H. Ikeda and B. S. Moore, *Nat. Prod. Rep.*, 2009, **26**, 1362–1384.

46   T. P. Dick, A. K. Nussbaum, M. Deeg, W. Heinemeyer, M. Groll, M. Schirle, W. Keilholz, S. Stevanović, D. H. Wolf, R. Huber, H. G. Rammensee and H. Schild, *J. Biol. Chem.*, 1998, **273**, 25637–25646.

47   E. Seemüller, A. Lupas, D. Stock, J. Löwe, R. Huber and W. Baumeister, *Science (80-.).*, 1995, **268**, 579–582.

48   T. Aoyagi, T. Takeuchi, A. Matsuzaki, K. Kawamura, S. Kondo, M. Hamada, K. Maeda and H. Umeazawa, *J. Antibiot. (Tokyo).*, 1969, **12**, 283–286.

49   M. Mullowney, R. A. McClure, M. T. Robey, N. L. Kelleher and R. J. Thomson, *Nat. Prod. Rep.*, 2018, **35**, 847–878.

50   K. L. Rock, C. Gramm, L. Rothstein, K. Clark, R. Stein, L. Dick, D. Hwang and A. L. Goldberg, *Cell*, 1994, **78**, 761–771.

51   M. Bogyo, J. S. McMaster, M. Gaczynska, D. Tortorella, A. L. Goldberg and H. Ploegh, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 6629–6634.

52   J. Adams, M. Behnke, S. Chen, A. A. Cruickshank, L. R. Dick, L. Grenier, J. M. Klunder, Y.-T. Ma, L. Plamondon and R. L. Stein, *Bioorganic Med. Chem. Lett.*, 1998, **8**, 333–338.

53   J. Adams and M. Kauffman, *Cancer Invest.*, 2004, **22**, 304–311.

54   N. Gupta, M. J. Hanley, P. M. Diderichsen, H. Yang, A. Ke, Z. Teng, R. Labotka, D. Berg, C. Patel, G. Liu, H. van de Velde and K. Venkatakrishnan, *Clin. Pharmacol. Ther.*, 2019, **105**, 376–387.

55   S. Omura, T. Fujimoto, K. Otoguro, K. Matsuzaki, R. Moriguchi, H. Tanaka and Y. Sasaki, *J. Antibiot. (Tokyo).*, 1991, **44**, 113–116.

56   G. Fenteany, R. F. Standaert, W. S. Lane, S. Choi, E. J. Corey and S. L. Schreiber, *Science*, 1995, **268**, 726–731.

57   L. R. Dick, A. A. Cruikshank, A. T. Destree, L. Grenier, T. A. McCormack, F. D. Melandri, S. L. Nunes, V. J. Palombella, L. A. Parent, L. Plamondon and R. L. Stein, *J. Biol. Chem.*, 1997, **272**, 182–188.

58   B. C. Potts, M. X. Albitar, K. C. Anderson, S. Baritaki, C. Berkers, B. Bonavida, J. Chandra, D. Chauhan, J. C. Cusack, W. Fenical, I. M. Ghobrial, M. Groll, P. R. Jensen, K. S. Lam, G. K. Lloyd, W. McBride, D. J. McConkey, C. P. Miller, S. T. C. Neuteboom, Y. Oki, H. Ovaa, F. Pajonk, P. G. Richardson, A. M. Roccaro, C. M. Sloss, M. A. Spear, E. Valashi, A. Younes and M. A. Palladina, *Mol. Cell. Biochem.*, 2011, **11**, 254–284.

59   R. H. Feling, G. O. Buchanan, T. J. Mincer, C. A. Kauffman, P. R. Jensen and W. Fenical, *Angew. Chemie Int. Ed.*, 2003, **42**, 355–357.

60   D. Chauhan, L. Catley, G. Li, K. Podar, T. Hideshima, M. Velankar, C. Mitsiades, N. Mitsiades, H. Yasui, A. Letai, H. Ovaa, C. Berkers, B. Nicholson, T. H. Chao, S. T. C. Neuteboom, P. Richardson, M. A. Palladino and K. C. Anderson, *Cancer Cell*, 2005, **8**, 407–419.

61   R. R. Manam, K. A. McArthur, T. H. Chao, J. Weiss, J. A. Ali, V. J. Palombella, M. Groll, G. K. Lloyd, M. A. Palladino, S. T. C. Neuteboom, V. R. Macherla and B. C. M. Potts, *J. Med. Chem.*, 2008, **51**, 6711–6724.

62   G. E. Hubbell and J. J. Tepe, *RSC Chem. Biol.*, 2020, **1**, 305–332.

63   M. Groll, Y. Koguchi, R. Huber and J. Kohno, *J. Mol. Biol.*, 2001, **311**, 543–548.

64 I. Nickeleit, S. Zender, F. Sasse, R. Geffers, G. Brandes, I. Sörensen, H. Steinmetz, S. Kubicka, T. Carlomagno, D. Menche, I. Gütgemann, J. Buer, A. Gossler, M. P. Manns, M. Kalesse, R. Frank and N. P. Malek, *Cancer Cell*, 2008, **14**, 23–35.

65 R. Sekizawa, I. Momose, N. Kinoshita, H. Naganawa, M. Hamada, Y. Muraoka, H. Iinuma and T. Takeuchi, *J. Antibiot. (Tokyo).*, 2001, **54**, 874–881.

66 L. Meng, B. H. B. Kwok, N. Sin and C. M. Crews, *Cancer Res.*, 1999, **59**, 2798–2801.

67 L. Meng, R. Mohan, B. Kwok, M. Elofsson, N. Sin and C. Crews, *Proc. Natl. Acad. Sci.*, 1999, **96**, 10403–10408.

68 M. Groll, K. B. Kim, N. Kairies, R. Huber and C. M. Crews, *J. Am. Chem. Soc.*, 2000, **122**, 1237–1238.

69 M. Hanada, K. Sugawara, K. Kaneta, S. Toda, Y. Nishiyama, K. Tomita, H. Yamamoto, M. Konishi and T. Oki, *J. Antibiot. (Tokyo).*, 1992, **45**, 1746–1752.

70 H.-J. Zhou, M. Aujay, M. Bennett, M. Dajee, S. Demo, Y. Fang, M. Ho, J. Jiang, C. Kirk, G. Laidig and E. Lewis, *J. Med. Chem.*, 2009, **52**, 3028–3038.

71 P. Hari, J. V. Matous, P. M. Voorhees, K. H. Shain, M. Obreja, J. Frye, H. Fujii, A. J. Jakubowiak, D. Rossi and P. Sonneveld, *Blood Cancer J.*, 2019, **9**, 66.

72 K. Sugawara, M. Hatori, Y. Nishiyama, K. Tomita, H. Kamei, M. Konishi and T. Oki, *J. Antibiot. (Tokyo).*, 1990, **43**, 8–18.

73 J. G. Owen, Z. Charlop-Powers, A. G. Smith, M. A. Ternei, P. Y. Calle, B. V. B. Reddy, D. Montiel and S. F. Brady, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 4221–4226.

74 K. Tsuchiya, S. Kobayashi, T. Nishikiori, T. Nakagawa and K. Tatsuta, *J. Antibiot. (Tokyo).*, 1997, **50**, 261–263.

75 Y. Koguchi, J. Kohno, S.-I. Suzuki, M. Nishio, K. Takahashi, T. Ohnuki and S. Komatsubara, *J. Antibiot. (Tokyo).*, 1999, **52**, 1069–1076.

76 A. R. Pereira, A. J. Kale, A. T. Fenley, T. Byrum, H. M. Debonsi, M. K. Gilson, F. A. Valeriote, B. S. Moore and W. H. Gerwick, *ChemBioChem*, 2012, **13**, 810–817.

77 Y. Tsunematsu, S. Nishimura, A. Hattori, S. Oishi, N. Fujii and H. Kakeya, *Org. Lett.*, 2015, **17**, 258–261.

78 L. Keller, A. Plaza, C. Dubiella, M. Groll, M. Kaiser and R. Müller, *J. Am. Chem. Soc.*, 2015, **137**, 8121–8130.

79 K. B. Kim and C. M. Crews, *Nat. Prod. Rep.*, 2013, **30**, 600–604.

80 P. Lawasut, D. Chauhan, J. Laubach, C. Hayes, C. Fabre, M. Maglio, C. Mitsiades, T. Hideshima, K. C. Anderson and P. G. Richardson, *Curr. Hematol. Malig. Rep.*, 2012, **7**, 258–266.

81 T. Muchamuel, M. Basler, M. A. Aujay, E. Suzuki, K. W. Kalim, C. Lauer, C. Sylvain, E. R. Ring, J. Shields, J. Jiang, P. Shwonek, F. Parlati, S. D. Demo, M. K. Bennett, C. J. Kirk and M. Groettrup, *Nat. Med.*, 2009, **15**, 781–787.

82 J. Schrader, F. Henneberrg, R. Mata, K. Tittmann, T. R. Schneider, H. Stark, G. Bourenkov and A. Chari, *Science*, 2016, **353**, 594–598.

83 K. Carmony, W. Lee and K. B. Kim, *ChemBioChem*, 2016, **17**, 2115–2117.

84 M. J. Lee, D. Bhattarai, J. Yoo, Z. Miller, J. E. Park, S. Lee, W. Lee, J. J. Driscoll and K. B. Kim, *J. Med. Chem.*, 2019, **62**, 4444–4455.

85 N. Sin, B. K. Kyung, M. Elofsson, L. Meng, H. Auth, B. H. B. Kwok and C. M. Crews, *Bioorganic Med. Chem. Lett.*, 1999, **9**, 2283–2288.

86 F. Moschona, M. Savvopoulou, M. Tritopoulou, D. Tatarraki and G. Rassias, *Catalysts*, 2020, **10**, 1–65.

87 A. Spaltenstein, J. J. Leban, J. J. Huang, K. R. Reinhardt, O. H. Viveros, J. Sigafoos and R. Crouch, *Tetrahedron Lett.*, 1996, **37**, 1343–1346.

88 P. Phiasivongsa, L. C. Sehl, W. D. Fuller and G. J. Laidig, *US-8367617-B2*.

89 M. S. Smyth and G. J. Laidig, *US8207125B2*.

90 M. S. Smyth and G. J. Laidig, *US20120094930A1*.

91 P. Phiasivongsa, L. C. Sehl, D. F. William and G. J. Laidig, *US8921583B2*.

92 G. J. Laidig, P. Radel and M. S. Smyth, *US20050256324A1*.

93 P. K. Dornan, T. Anthoine, M. G. Beaver, G. C. Cheng, D. E. Cohen, S. Cui, W. E. Lake, N. F. Langille, S. P. Lucas, J. Patel, W. Powazinik, S. W. Roberts, C. Scardino, J. L. Tucker, S. Spada, A. Zeng and S. D. Walker, *Org. Process Res. Dev.*, 2020, **24**, 481–489.

94 Shanghai Haoyuan Chemical Technology Co Ltd, *CN104230857A*.

95 K. Hoeferl-Prantz and T. Wilhelm, *US20160194354A1*.

96 M. Schorn, J. Zettler, J. P. Noel, P. C. Dorrestein, B. S. Moore and L. Kaysser, *ACS Chem. Biol.*, 2014, **9**, 301–309.

97 D. Zabala, J. W. Cartwright, D. M. Roberts, B. J. C. Law, L. Song, M. Samborskyy, P. F. Leadlay, J. Micklefield and G. L. Challis, *J. Am. Chem. Soc.*, 2016, **138**, 4342–4345.

98   Y. Li, G. Florova and K. A. Reynolds, *J. Bacteriol.*, 2005, **187**, 3795–3799.

99   R. Finking and M. A. Marahiel, *Annu. Rev. Microbiol.*, 2004, **58**, 453–488.

100   J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380–416.

101   C. L. Santos, M. Correia-Neves, P. Moradas-Ferreira and M. V. Mendes, *PLoS One*, 2012, **7**, 46758.

102   T. A. M. Gulder and B. S. Moore, *Angew. Chemie - Int. Ed.*, 2010, **49**, 9346–9367.

103   A. J. Kale, R. P. McGlinchey, A. Lechner and B. S. Moore, *ACS Chem. Biol.*, 2011, **6**, 1257–1264.

104   Y. M. Zhang, H. Marrakchi, S. W. White and C. O. Rock, *J. Lipid Res.*, 2003, **44**, 1–10.

105   G. Gago, L. Diacovich, A. Arabolaza, S. C. Tsai and H. Gramajo, *FEMS Microbiol. Rev.*, 2011, **35**, 475–497.

106   C. Davies, R. J. Heath, S. W. White and C. O. Rock, *Structure*, 2000, **8**, 185–195.

107   C. Y. Lai and J. E. Cronan, *J. Bacteriol.*, 2004, **186**, 1869–1878.

108   K. Finzel, C. Nguyen, D. R. Jackson, A. Gupta, S. C. Tsai and M. D. Burkart, *Chem. Biol.*, 2015, **22**, 1453–1460.

109   R. J. Heath, N. Su, C. K. Murphy and C. O. Rock, *J. Biol. Chem.*, 2000, **275**, 40128–40133.

110   E. A. Felnagle, J. J. Barkei, H. Park, A. M. Podevels, M. D. McMahon, D. W. Drott and M. G. Thomas, *Biochemistry*, 2010, **49**, 8815–8817.

111   M. Saraste, P. R. Sibbald and A. Wittinghofer, *Trends Biochem. Sci.*, 1990, **15**, 430–434.

112   G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.*, 2000, **7**, 211–224.

113   A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050–1073.

114   S. Ghisla and C. Thorpe, *Eur. J. Biochem.*, 2004, **271**, 494–508.

115   J. W. Cartwright, *Univ. Warwick*.

116   R. D. Bach and C. Canepa, *J. Org. Chem.*, 1996, **61**, 6346–6353.

117   A. Rauwerdink and R. J. Kazlauskas, *ACS Catal.*, 2015, **5**, 6153–6176.

118   S. Lange, A. Musidlowska, C. Schmidt-Dannert, J. Schmitt and U. T. Bornscheuer, *ChemBioChem*, 2001, **2**, 576–582.

119   P. Domínguez de María, C. A. García-Burgos, G. Bargeman and R. W. van Gemert, *Synth.*, 2007, **10**, 1439.

120   G. L. Beutner, I. S. Young, M. L. Davies, M. R. Hickey, H. Park, J. M. Stevens and Q. Ye, *Org. Lett.*, 2018, **20**, 4218–4222.

121   E. L. C. de los Santos and G. L. Challis, *bioRxiv*, 2017, 4–7.

122   K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. Van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2021, **49**, W29–W35.

123   C. L. M. Gilchrist and Y. H. Chooi, *Bioinformatics*, 2021, **37**, 2473–2475.

124   B. O. Bachmann and J. Ravel, *Methods Enzymol.*, 2009, **458**, 181–217.

125   T. Shomura, N. Ezaki, T. Tsuruoka, T. Niwa, E. Akita and T. Niida, *J. Antibiot. (Tokyo).*, 1970, **23**, 155–161.

126   H. Baud, A. Betencourt, M. Peyre and L. Penasse, *J. Antibiot. (Tokyo).*, 1977, 720–723.

127   S. Sato, F. Iwata, S. Yamada, H. Kawahara and M. Katayama, *Bioorganic Med. Chem. Lett.*, 2011, **21**, 7099–7101.

128   G. Le Goff, M.-T. Martin, C. Servy, S. Cortial, P. Lopes, A. Bialecki, J. Smadja and J. Ouazzani, *J. Nat. Prod.*, 2012, **75**, 915–919.

129   W. A. Ayer, J. K. Jenkins, K. Piers and S. Valverde-Lopez, *Can. J. Chem.*, 1967, **45**, 445–450.

130   H. Zhang, Q. Zhou, T. Lou, S. Wang and H. Ruan, *J. Glob. Antimicrob. Resist.*, 2017, **11**, 159–160.

131   F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov, *Bioinformatics*, 2015, **31**, 3210–3212.

132   F. Madeira, M. Pearce, A. R. N. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov and R. Lopez, *Nucleic Acids Res.*, 2022, **50**, 276–279.

133   M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov and M. Steinegger, *Nat. Methods*, 2022, **19**, 679–682.

134   F. H. Arnold, *Angew. Chemie - Int. Ed.*, 2018, **57**, 4143–4148.

8. Appendix
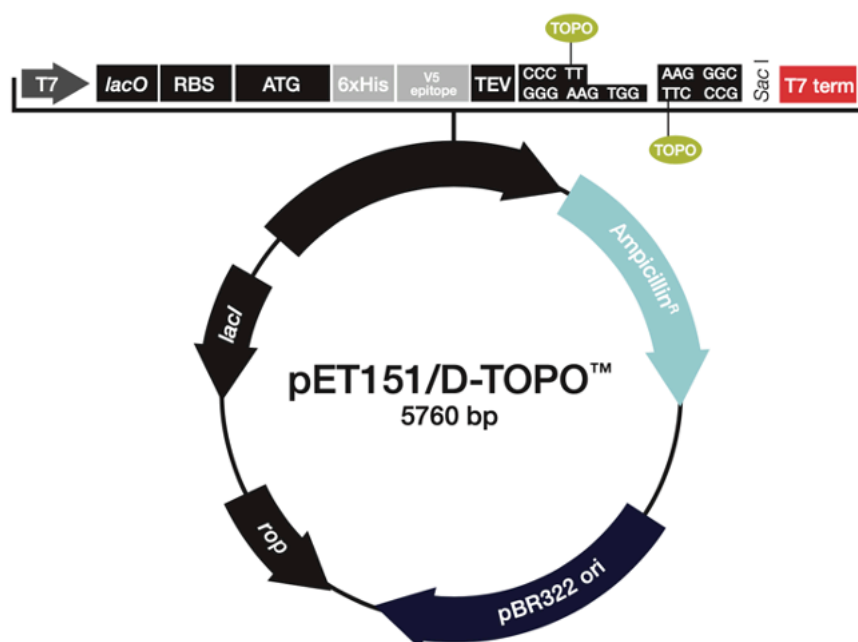
8.1 Plasmid map pET151

Figure 8.1 Plasmid map of the pET151 expression vector included in the Champion^{TM} pET151 directional TOPO^{TM} expression kit ordered from ThermoFisher Scientific (https://www.thermofisher.com/order/catalog/product/K15101).

8.2 Cluster boundaries for heterologous expression experiments by Dr Huang

Figure 8.2 Cluster boundaries used for cloning the S. sparsogenes putative tryptopeptin A cluster for heterologous expression by Dr Huang. The grey "L" and "R" markers represent the boundaries on each site. Ctg34_4 is the putative epoxyketone synthase and ctg34_14 the putative aminotransferase.
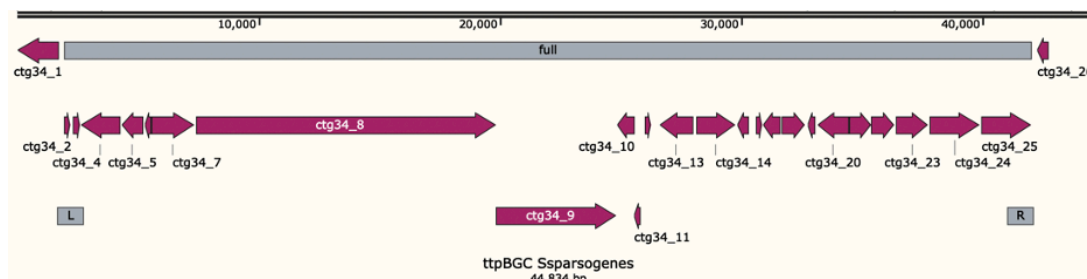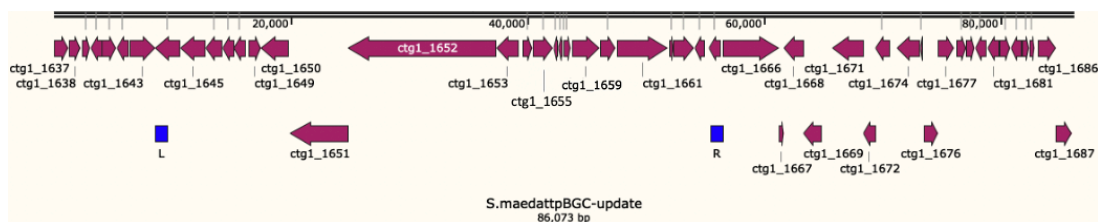

Figure 8.3 Cluster boundaries used for cloning the S. sp. maeda85 putative tryptopeptin A cluster for heterologous expression by Dr Huang. The blue "L" and "R" markers represent the boundaries on each site. Ctg1_1655 is the putative epoxyketone synthase and ctg1_1645 the proposed final regulatory gene in the predicted cluster.

8.3 Sequencing of pET151 vectors harbouring putative epoxyketone synthases

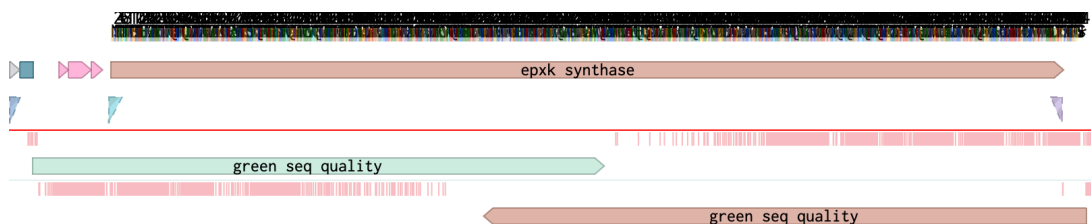*Figure 8.4 Sequencing with T7 and T7 reverse primers of pET151-ttpC plasmid. TtpC was amplified from S. sparsogenes genomic DNA and the plasmid assembled with the directional TOPO cloning kit.*

Figure 8.5 Sequencing with T7 and T7 reverse primers of pET151-tptC plasmid. TptC was amplified from S. sp. maeda85genomic DNA and the plasmid assembled with the directional TOPO cloning kit.

8.4 Putative functions of genes in the identified tryptopeptin A biosynthetic gene clusters

8.4.1 Potential tryptopeptin A cluster in *Streptomyces sparsogenes* ATCC 25498

Gene	aa length	predicted function	nearest homologue	identity
Sspa_01578	538	oxidoreductase	FAD-binding oxidoreductase protein	91%
Sspa_01577	281	Proteasome β-subunit	proteasome subunit beta	88%
Sspa_01576	584	ACAD	acyl-CoA/acyl-ACP dehydrogenase	90%
Sspa_01575	4130	NRPS	non-ribosomal peptide synthetase	81%
Sspa_01574	1659	PKS	type I polyketide synthase	83%
Sspa_01573	223	dehydrogenase	4-hydroxythreonine-4-phosphate dehydrogenase	59%
Sspa_01572	87	dehydrogenase	4-hydroxythreonine-4-phosphate dehydrogenase	91%
Sspa_01571	454	transport	uncharacterized protein YgbK (DUF1537 family)	56%
Sspa_01570	540	aminotransferase	PLP-dependent aminotransferase family protein	66%

8.4.2 Potential tryptopeptin A cluster in *Streptomyces* sp. maeda85

Gene	aa length	predicted function	nearest homologue	identity
ctgl_1655	538	oxidoreductase	FAD binding oxidoreductase	61%
ctgl_1654	281	Proteasome β-subunit	proteasome subunit beta	98%
ctgl_1653	609	ACAD	acyl-CoA dehydrogenase	95%
ctgl_1652	4155	NRPS	NRPS	92%
ctgl_1651	1643	PKS	Type I PKS	92%
ctgl_1650	774	Transport/regulation	serine/threonine protein kinase	94%
ctgl_1649	337	regulator	LacI family transcriptional regulator	98%
ctgl_1648	307	transporter	sugar ABC transporter permease	98%
ctgl_1647	324	transporter	carbohydrate ABC transporter permease	99%
ctgl_1646	446	transporter	sugar ABC transporter substrate binding protein	99%

8.5 Individual yields of substrate analogue syntheses

Table 2 Individual yields during the synthesis of substrate analogues referring to Scheme 2.4-7.

Substrate	Malonate extension	Dimethylation	Deprotection	EDC/HOBt Coupling	TCFH–NMI coupling
a	92%	98%	100%	23%	-
b	72%	100%	100%	10%	-
c	88	95%	100%	-	17%
d	83	91%	98%	16%	53%
e	91	88%	100%	-	26%
f	77	100%	985	-	33%
g	96	86%	100%	-	38%
h	56	97%	100%	-	31%