Threonyl tRNA synthetases as antibiotic targets and resistance mechanisms

Jonathon David Liston

John Innes Centre

Department of Molecular Microbiology

Norwich Research Park, Colney Lane, Norwich, NR4 7UH, UK

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

December 2022

Threonyl tRNA synthetases as antibiotic targets and resistance mechanisms

Jonathon David Liston, John Innes Centre, December 2022

Abstract

Threonyl tRNA synthetases (ThrRSs) catalyse the attachment of L-threonine to its cognate $tRNA^{Thr}$. These enzymes are essential for the translation of proteins. There are currently two known natural product ThrRS inhibitors, borrelidin and obafluorin, produced by *Streptomyces parvulus* Tü 4055 and *Pseudomonas fluorescens* ATCC 39502 respectively. Both of their biosynthetic gene clusters (BGCs) encode secondary ThrRSs which provide self-resistance mechanisms. In the borrelidin BGC, this is BorO and in the obafluorin BGC, this is ObaO. While the biosynthesis of both compounds is well understood, the mechanisms of self-resistance are not, and the mechanism of action of obafluorin remains elusive.

Here the structure of BorO was solved and it was found that both ThrRSs in the producer, *S. parvulus* (Sp), are resistant to borrelidin. Mutagenesis of the *Escherichia coli* target, EcThrRS, identified a L489T mutation which is sufficient to confer resistance by preventing borrelidin binding, explaining resistance by BorO. In SpThrRS, this residue is a glutamate and an EcThrRS, L489Q mutation is not sufficient to confer resistance, meaning that SpThrRS has a distinct borrelidin resistance mechanism. It was unexpectedly found that ObaO can confer resistance to borrelidin due to the presence of a methionine in this same position and the EcThrRS L489M mutant is resistant to borrelidin. Introduction of this mutation allowed the first structure of ObaO to be solved by X-ray crystallography. Cryogenic electron microscopy (Cryo-EM) structures of obafluorin bound to both EcThrRS and ObaO were solved, showing that EcThrRS covalently links to obafluorin through Y462, while the interaction is non-covalent for ObaO. Spontaneous resistant mutagenesis identified the serine (S463) immediately adjacent to Y462as an essential component of the ObaO resistance to obafluorin.

Finally, a survey of published genomes for additional copies of ThrRS identified BGCs encoding potentially novel natural product ThrRS inhibitors.

Acknowledgements

I would first like to thank Prof Barrie Wilkinson for having me in the group, for encouraging my curiosity, sage advice and trust throughout the years.

To Prof Dave Lawson and Dr Clare Stevenson I owe a special thanks for nurturing my excitement for all things protein, having endless patience in showing me how to do the techniques, the hours of data collection for awful crystals (or salt) and to Clare for the dozens and dozens of difficult, tiny and fragile crystals fished from crystallisation trays.

A very special thanks needs to go to Dr Sibyl Batey. One of the most dedicated, talented, patient and kind scientists (and people) I've met. Working with you was an absolute dream and you were sorely missed when you moved on to bigger and better things. Thank you for all of the encouragement and suggestions, the late-night chats in the lab and the seemingly endless well of knowledge and skills. Truly I could not have done this without you.

To Dr Joseph Sallmen, Dr Catriona Thompson and Dr Claudio Greco, thank you for all of your support both scientifically and personally. I feel privileged to have been able to meet you both as amazing scientists but also as people.

To Ben Scott and Hannah McDonald, I thank you for being with me from PhD day one. We really dragged each other through this, with the help of many a gin, wine, snack and horror film.

To Katie Noble and Dr Ainsley Beaton, thank you for letting me talk at you (if not necessarily to you) while trying to interpret my data, for the daily mochas and coparenting our office cactus, Jeffrey.

To Dr Damien Gayrard for being my housemate, keeping me sane during writing and bringing back gossip from the lab while I was in (self-imposed) lockdown to write.

A special thanks needs to go to Samar Tantush and Charlotte Hall, my best friends, for always being there to top up my glass and cheer me up. Your steady and stable friendship will always be the most valuable thing to me, even though you both abandoned me when you left Norwich.

I would also like to thank my parents, siblings and my dog, Walter. All of you have given me unconditional support and love throughout the PhD but also throughout life, with the surety that I can do anything I put my mind to.

Author’s declaration

The research described in this thesis was conducted entirely at the John Innes Centre between October 2018 and December 2022. All the data described are original and were obtained by the author, except where specific acknowledgement has been made. No part of this thesis has previously been submitted as for a degree at this or any other academic institution.

Abbreviations

2xYT: 2x Yeast Tryptone Medium

2,3-DHBA: (2,3)-dihydroxybenzoic acid

A: Adenylation

A3S: 3'-O-serinyl adenosine

A3T: 3'-O-threonyl adenosine

A3G: 3'-O-glycl adenosine

aaRS: Aminoacyl tRNA synthetase

ACB: Anticodon binding domain

ACP: Acyl carrier protein

ADP: Adenosine diphosphate

AHNB: (2S,3R)-2-amin-3-hydroxy-4-(4-nitrophenyl)butanoate

AIMP: aaRS interacting multifunctional protein

AlaRS: alanyl tRNA synthetase

AMP: Adenosine monophosphate

ArgRS: Arginyl tRNA synthetase

AsnRS: asparaginyl tRNA synthetase

AspRS: aspartyl tRNA synthetase

AspAMP: aspartyl adenylate

AT: Acyltransferase

ATP: Adenosine triphosphate

AU: Absorbance units

BGC: Biosynthetic Gene Cluster

BLAST: Basic Local Alignment Search Tool

BSA: Bovine Serum Albumin

C: Condensation

CD: Circular Dichroism

cDNA: Complimentary DNA

Clip-Seq: crosslinking immunoprecipitation sequencing

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

Cryo-EM: Cryogenic Electron Microscopy

CoA: Coenzyme A

CV: Column volume

CysRS: Cysteinyl tRNA synthetase

DBU: 1,8-diazabiocyclo[5.4.0]undec-7-ene

DH: Dehydratase

Dha: Dehydroalanine

Dhb: Dehydrobutyrine

DIC: Differential interference contrast

DLS: Diamond Light Source

DMAc: Dimethylacetamide

DMAPP: dimethylallylpyrophosphate

DMSO: Dimethyl sulfoxide

DNA: Difco Nutrient Agar

dNTP: Deoxyribonucleotide triphosphates

DTT: dithiothreitol

EB: Elution buffer

EcThrRS: *Escherichia coli* Threonyl tRNA synthetase

eIF4E: Eukaryotic translation initiation factor

ELSD: Evaporative Light Scattering Detector

EM: Electron microscopy

ER: Enoylreductase

ESKAPE: *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa* and *Enterobacter spp.*

EtBr: Ethidium bromide

FA: Formic Acid

FPLC: Fast Protein Liquid Chromatography

FPP: Farnesyl pyrophosphate

FSC: Fourier Shell Correlation

gDNA: Genomic DNA

GGPP: Geranylgeranyl pyrophosphate

GlnRS: Glutaminyl tRNA synthetase

GluRS: Glutamyl tRNA synthetase

GlyRS: Glycyl tRNA synthetase

GPP: Geranyl pyrophosphate

GYM: Glucose Yeast Extract Medium

HisRS: Histidyl tRNA Synthetase

HIV: Human Immunodeficiency Virus

HK: Housekeeping

HPLC: High Performance Liquid Chromatography

HRMS: High resolution mass spectrometry

hRSV: Human Respiratory Syncytial Virus

HRP: Horseradish Peroxidase

HsThrRS: *Homo sapiens* Threonyl tRNA Synthetase

HSV: Herpes Simplex Virus

IC_{50}: concentration at which 50% of protein activity has been abolished

IDT: Integrated DNA technologies

IleAMP: Isoleucyl adenylate

IleRS: Isoleucyl tRNA synthetase

IMAC: Immobilised Metal Affinity Chromatography

IPP: Isopentyl pyrophosphate

IPTG: Isopropyl β-D-1-thiogalactopyranoside

ITC: Isothermal Titration Calorimetry

iTOL: Interactive tree of life

JIC: John Innes Centre

JNK: c-Jun N-terminal Kinase

K_d: Dissociation constant

KR: Ketoreductase

KS: Ketosynthase

L: -NaCl Lennox Broth

L-TTA: L-threonine transaldolase

LB: Lysogeny Broth

LC-MS: Liquid Chromatography Mass Spectrometry

LeuAMP: Leucyl adenylate

LeuRS: Leucyl tRNA synthetase

LRMS: Low resolution mass spectrometry

LysRS: Lysinyl tRNA synthetase

MAG: Metagenome-assembled genome

MALDI-TOF: Matrix-assisted laser desorption/ionisation-time of flight

MBP: Maltose Binding Protein

MD: Molecular Dynamics

MeCN: Acetonitrile

MeOH: Methanol

MEP: Methyl-D-erythritol 4-phosphate

MetRS: Methionyl tRNA synthetase

MFS: Major Facilitator Superfamily

MIBiG: Minimum Information about a Biosynthetic Gene cluster

MIC: Minimum inhibitory concentration

ML: Maximum Likelihood

mLST8: mammalian lethal with Sec13 protein 8

mRNA: Messenger RNA

MRSA: Methicillin Resistant *Staphylococcus aureus*

MS: Mass Spectrometry

MSC: Multiple Synthetase Complex

MSSA: Methicillin Susceptible *Staphylococcus aureus*

mTOR: mechanistic target of rapamycin

mTORC: mTOR complex

MVA: mevalonate

MYM: Maltose Yeast Extract Medium

N1: Editing regulation domain

N2: Editing catalytic domain

NADPH: Nicotinamide adenine dinucleotide phosphate

NAG: *N*-acetyl glucosamine

NEB: New England BioLabs

NMR: Nuclear Magnetic Resonance

NP: Natural Product

NPS-TTD: Non-photosensitive trichothiodystrophy

NRPS: Non-ribosomal Peptide Synthase

OD: Optical density

OPM: Obafluorin Production Media

oriT: Origin of transfer

PBS: Phosphate buffered saline

PCR: Polymerase Chain Reaction

PCP: Peptide carrier protein

PDB: Protein Data Bank

PEG: Polyethylene glycol

PfThrRS: *Pseudomonas fluorescens* Threonyl tRNA Synthetase

PheRS: Phenylalanyl tRNA synthetase

Prep HPLC: Preparative High Performance Liquid Chromatography

PRP: pentapeptide repeat protein

ProAMP: Prolyl adenylate

ProRS: Prolyl tRNA synthetase

PKS: Polyketide Synthase

Psp: Phage shock protein

PYDG: Peptone Yeast Dextrin Glucose Medium

PylRS: Pyrrolysyl tRNA synthetase

QM: Quantum mechanics

Rags: Rag guanosine triphosphatase

RiPP: Ribosomally synthesised and post translationally modified peptide

RMSD: Root mean squared deviation

rRNA: Ribosomal RNA

SaThrRS: *Staphylococcus aureus* threonyl tRNA synthetase

SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SepRS: O-phosphoseryl tRNA synthetase

SerAMP: Serinyl adenylate

SerRS: Serinyl tRNA synthetase

SFM: Sofa Flour Mannitol Medium

SNA: Soft Nutrient Agar

SOC: Suppressor of Catabolism Medium

SPR: Surface Plasmon Resonance

SpThrRS: *Streptomyces parvulus* Threonyl tRNA synthetase

SUMO: Small Ubiquitin-like Modifier

SvThrRS: *Streptomyces venezuelae* Threonyl tRNA synthetase

TAR: Transformation-Associated Recombination

TARS: Human threonyl tRNA synthetase (labelled 1-3)

TB: Terrific Broth

TBST: Tris-Buffered Saline + Tween20

tCPDA: *trans*-cyclopentane 1,2-dicarboxylic acid

TE: Thioesterase

TEM: Transmission Electron Microscopy

TFA: Trifluoroacetic acid

TGS: ThrRS, GTPase and SpoT domain

THF: Tetrahydrofuran

ThrAMP: Threonyl adenylate

ThrRS: Threonyl tRNA synthetase

thrS: threonyl tRNA synthetase gene (in bacteria)

ThrSAA: 5'-O-(threonylsulfamoyl)adenosine

TIPS: Triisopropylsilane

TNF-α: Tumour necrosis factor α

tRNA: Transfer RNA

$tRNA^{Thr}$: Cognate tRNA for ThrRS

TrpRS: Tryptophanyl tRNA synthetase

TSB: Tryptone Soy Broth

TyrRS: Tyrosyl tRNA synthetase

UNE-T: N-terminal unique domain

UPP: undecaprenyl phosphate

UTR: Untranslated region (of mRNA)

UV: Ultraviolet

ValRS: Valinyl tRNA synthetase

VEGF: Vascular endothelial growth factor

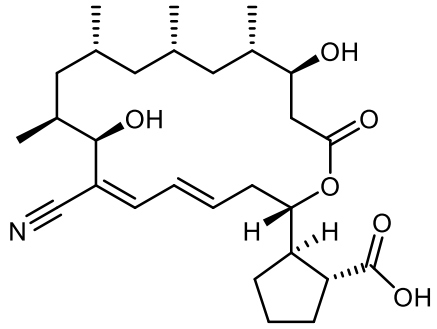
WGS: Whole genome sequencing

WT: Wild-type

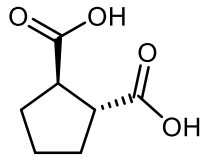
XRD: X-Ray diffraction

Compound Key

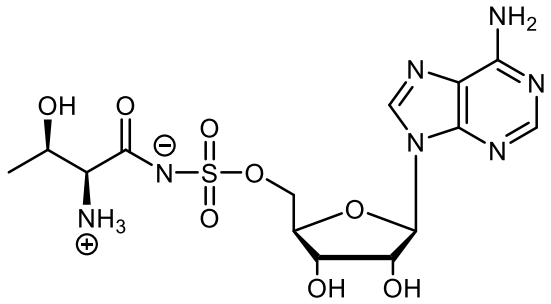
Borrelidin



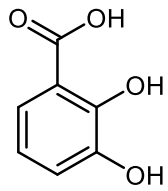
trans-cyclopentane-1,2-dicarboxylic acid (tCPDA)



5’-O-(threonylsulfamoyl)adenosine (ThrSAA)



2,3-dihydroxybenzoic acid (DHBA)



Obafluorin

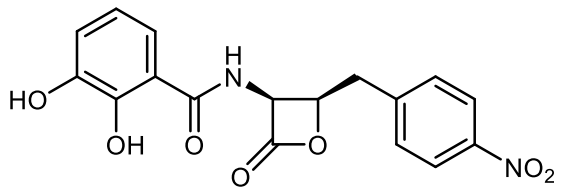


Table of Contents

Index of Tables

List of Figures

Chapter 1: Introduction

1.1 Bacterial Natural Products

With many millennia's head start on humanity, bacteria have evolved a diverse set of biological and chemical processes to provide an advantage that allows them to survive in their selected niches. We can, however, exploit this evolutionary advantage in order to tackle many of humanity's problems. For example, strains of *Geobacter* could be used to remove radioactive uranium from groundwater - their conductive pili can transfer electrons to minerals in the soil instead of using oxygen for respiration. This means they can chemically reduce the uranium, causing it to precipitate and preventing it from leaching into groundwater[1]. Additionally, a number of bacteria have demonstrated the ability to degrade plastics, reducing the issue of terrestrial and marine plastic accumulation[2].

Specialised metabolism in bacteria has also led to the evolution of complex pathways to produce diverse natural products (NPs) which are used by bacteria for a variety of different functions. Many of these examples provide the producer with a survival advantage during lifestyle or environmental changes. The chemical and biological activities of these compounds are highly diverse as well as being industrially and medically very valuable, and a wide range of microbial natural products are being used as pharmaceuticals in the clinic[3]. Although we often focus on the antibiotic properties of these molecules, they can also be involved in cell signalling, and to help the producing bacterium outcompete competitors for nutrients, fight off invading pathogens, protect against bacteriophages, or to attack the host they are trying to invade.

Natural products are part of specialised metabolism and are often interchangeably referred to as specialised metabolites, secondary metabolites or natural products[4]. Their production is not necessarily essential for the growth and survival of the bacterium under laboratory conditions but may be essential in an ecological context. This contrasts with primary metabolism which involves essential processes such as respiration, amino acid biosynthesis and nucleotide biosynthesis. Specialised metabolites are built using the products of primary metabolism, and specialised metabolic enzymes have generally evolved from primary metabolism copies which, in microorganisms, assemble into biosynthetic gene clusters (BGCs). The mechanism by which this clustering arises remains a mystery. BGCs encode most, if not all, of the genes which are required for the biosynthesis of a specific compound or group of related compounds (called congeners). Microorganisms can devote up to 15% of their genome to these BGCs to produce natural products[4].

BGCs encoding for the production of natural products in bacteria, fungi and plants can be identified using a variety of bioinformatic tools such as antiSMASH and PRISM[5-8]. However, for bacteria, which lack introns and exons, this process is significantly simpler than in eukaryotes, with the genes in the cluster often being more spatially linked, often on polycistronic mRNA with multiple genes

controlled by the same promoter. A major problem in natural product research is that a majority of these BGCs appear to be cryptic (non-producing) under laboratory conditions[9].

1.1.1 Actinomycetes and Natural Products

The *Actinomycetes* family is made up of a diverse array of Gram-positive bacteria, with a characteristically high G+C content in their genomes and a variety of diverse cell morphologies. Within this family there are many medically and environmentally important genera. These include pathogenic *Mycobacterium* species such as *Mycobacterium leprae* and *Mycobacterium tuberculosis* that are the causative agents of leprosy and tuberculosis respectively. Due to the danger represented by their infection in humans, there is as large amount of interest in the discovery of antimycobacterial drugs. Treatment of *Mycobacteria* infection is especially challenging, due to their thick, mycolic acid containing cell walls[10,11].

Frankia, another member of the actinomycete family, are involved in nitrogen fixation in the plant roots of *Actinorhizal* plants, which are often pioneer species due to their ability (thanks to *Frankia*) to grow on nitrogen-scarce soils. An example of this are alder trees (*Alnus glutinosa*) which are often found to be colonised by *Frankia alni*[12,13]. Other than their complex and often multicellular growth, a major feature of *Actinomycetes* is the presence of many BGCs in their genomes which produce a vast number of medically interesting compounds.

While *Streptomyces* are the most well studied genus of this group, many other members are understudied but still represent talented sources of natural products. For example, strains of *Micromonospora* produce gentamicin[14], a broad spectrum antibiotic frequently used clinically for the treatment of Gram-negative infections such as those caused by *Pseudomonas aeruginosa*. The clinically used antibiotic erythromycin[15] is produced industrially by *Saccharopolyspora erythraea*[16]. Compared to *Streptomyces*, these other genera of *Actinomycetes* are relatively underexplored, but no less talented in their capacity to produce natural products[17,18]. While not directly medically relevant, the *Actinomycete*, *Corynebacterium glutamicum* is extensively used industrially for the production of a variety of chemical entities including amino acids, such as glutamate which is neutralised to monosodium glutamate and used in the food industry, as well as amino-acid derived molecules such as pipecolic acid, and also fuels and polymers[19,20].

1.1.2 Streptomyces and Natural Products

Around 70-80% of all isolated antimicrobial natural products discovered in the so called “Golden Age” of antibiotics in the 1950s and 1960s[21] were derived from *Streptomyces*. Large portions (up to 15-20%) of *Streptomyces* genomes can be devoted to secondary metabolism[22] and most BGCs are cryptic (non-producing) in laboratory conditions[9], meaning that research is needed to fully identify and understand the biosynthesis of these complex natural products. The specialised metabolites of

Streptomyces have a wide range of biological activities including as immunosuppressants, herbicides, anti-fungals, anti-cancer drugs and antibiotics.

Streptomyces have a complex life cycle and grow as multicellular filamentous hyphae (see Figure 1.1 for an illustration). To begin, a spore of *Streptomyces* germinates on the surface of a media plate in the lab (or naturally on the surface of the soil) and begins to grow vegetative hyphae through extension and branching within the media before aerial hyphae sprout above the surface in complex filamentous branched structures. In certain environmental conditions, the hyphae can then undergo sporulation, in which the aerial hyphae septate to form spores, which can be released to start a new colony[23]. Another recently discovered[24] phase of *Streptomyces* growth is exploration, triggered by the presence of fungi and volatile signalling molecules. In this phase of growth, the *Streptomyces* can undergo spreading on the surface of a medium[24]. Depending on the current lifestyle stage of the *Streptomyces*, different natural products are produced.

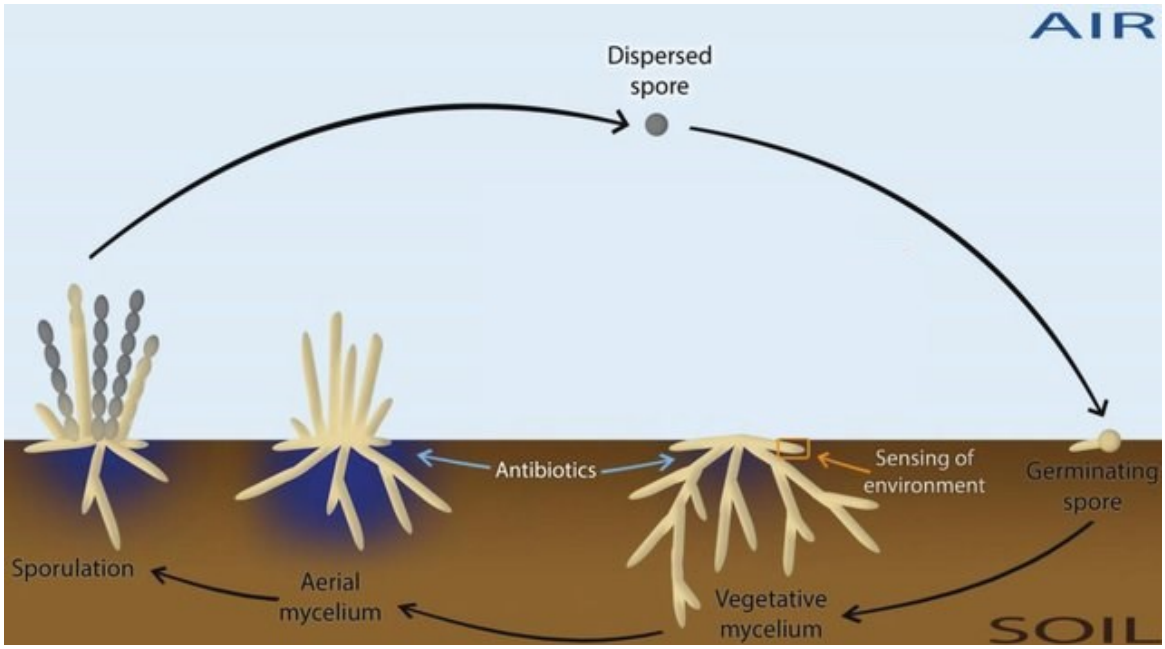


Figure 1.1. **Cartoon of the *Streptomyces* life cycle.** Each step is highlighted with blue to represent antibiotic (or other natural product) production. Sensing of environmental signals triggers transition from vegetative to aerial mycelium. Spore formation is associated with production of spore pigment, and therefore a change of colour of the colonies. Figure adapted from Urem *et al.* 2016[23]

Streptomyces scabies, among a small number of other species, has been identified as a plant pathogen causing potato blight[25] and while there are limited examples of *Streptomyces* acting as human pathogens, one of the very few is *Streptomyces somaliensis*. This causes actinomycetoma, which presents as deep tissue and bone infections, leading to the formation of tissue masses, resulting in tissue destruction and deformity[26]. Another, *Streptomyces* sp. TR1341 has been reported to be human lung associated, and was found to produce a variety of natural products which modulate the host immune system, as well as produce a variety of cytotoxic compounds.[27]

It has been suggested that the colonisation of humans by *Streptomyces* is actually beneficial, with the *Streptomyces* acting as coaches to train the immune system by using their natural products to modulate it.[28] An example of this is rapamycin, produced by *Streptomyces rapamycinicus,* which is used clinically following organ transplantation to prevent rejection[29]. These *Streptomyces* interactions could contribute to the hygiene hypothesis, which states that the increasing incidence of autoimmune diseases and allergies in modern, "western" countries is due to a lack of early interaction with the soil microbiome and an over sanitation of the environment during development[30].

1.1.3 Pseudomonads and Natural Products

Pseudomonads are Gram-negative bacteria which have a broad host range, with some species being soil or plant associated. *Pseudomonads* have a high genomic G+C content in their DNA, and are prolific producers of natural products. Many members of the *Pseudomonas* genera have been noted for their roles in disease but also in the production of natural products with potent biological activity [31-34]. Interestingly, there are a handful of known natural products where BGCs have been found both in *Pseudomonas* and *Streptomyces* genomes including Fosfomycin[35] and Bicyclomycin[36]. As is a common theme in natural products research, *Pseudomonas* strains are clearly a rich source of novel chemistry which might have the potential be used for medicines, but their biosynthetic potential are underexplored relative to *Streptomyces*[32].

Pseudomonads are of particular interest for human health, with *Pseudomonas aeruginosa* being a leading killer of patients with cystic fibrosis[37]. It is also classified as one of the ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) which have been identified as pathogens of high priority for the development of new treatments due to their virulence and multidrug resistance[37-39]. Additionally, *Pseudomonas* plant pathogens such as *Pseudomonas syringae* have been extensively studied for their prolific ability to be pathogens of fruit trees and other woody plants, leading to a major loss of yield and therefore costing billions of pounds to the agricultural industry per year[34].

Regardless of the genus of bacteria which is producing a specialised metabolite, they can broadly be categorised based upon the often-complex machinery involved in their biosynthesis. These can include polyketide synthases (PKSs), non-ribosomal peptide synthases (NRPSs), the ribosome, or terpene synthases.

1.1.4 Polyketide Synthases (PKSs)

Polyketide natural products are produced by plants, fungi and bacteria, and are produced by PKSs from simple acyl coenzyme A (CoA) precursor molecules to assemble complex structures[40]. There are three types of PKS: type I, type II and type III, although in recent years it has been thought that

such nomenclature is overly simplistic[40,41]. Type I PKSs are modular, multifunctional enzymes with each module being made up of catalytic domains having a distinct activity, and extending the polyketide chain[40,41] by one unit before passing on to the next module. Type II PKSs are multi-enzyme complexes which act iteratively to produce their products[40,41]. Type III PKSs are homodimeric and act by iteratively adding acyl CoA units to produce their polyketide chains[40]. Type I and II PKSs use acyl carrier proteins (ACPs), in a way similar to fatty acid synthases, to activate and hold acyl CoA substrates via a cysteine residue, whilst type III PKSs utilise acyl CoAs[40]. For all types of PKS, the carbon-carbon bond is formed by the catalytic activity of the ketosynthase (KS) domain and involves a decarboxylative Claisen-like condensation reaction[40]. Further modifications to the polyketide chain can occur through the activity of ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains[41]. The final step and domain will often be a thioesterase (TE) domain which will release the polyketide from the PKS either by cyclisation or hydrolysis. Following its release the natural product is often further modified by tailoring reactions to form the final natural product[41]. An example of a hypothetical PKS can be seen in Figure 1.2. The evolution of PKSs is unclear, although they appear to be related to fatty acid synthases[42].

One example of a clinically used polyketide is erythromycin (see Figure 1.3), which is used for respiratory tract infections, chlamydia, syphilis, skin infections and pelvic inflammatory disease. Produced by *Saccharopolyspora erythraea*, erythromycin is an inhibitor of the 23S ribosomal RNA (rRNA) in the 50S ribosome- erythromycin specifically targets bacteria over humans due to the differences between the prokaryotic and eukaryotic translational machinery[43]. Self-resistance to erythromycin in the producing organism comes from the *ermE* gene which encodes a methyltransferase which methylates at A2058 of the 23S rRNA, preventing the binding of erythromycin[44]. The promoter of the *ermE* gene, termed ermE* is regularly used in *Streptomyces* research as a strong promoter for gene expression and is an important tool for heterologous overexpression of genes in *Streptomyces spp.*[45].

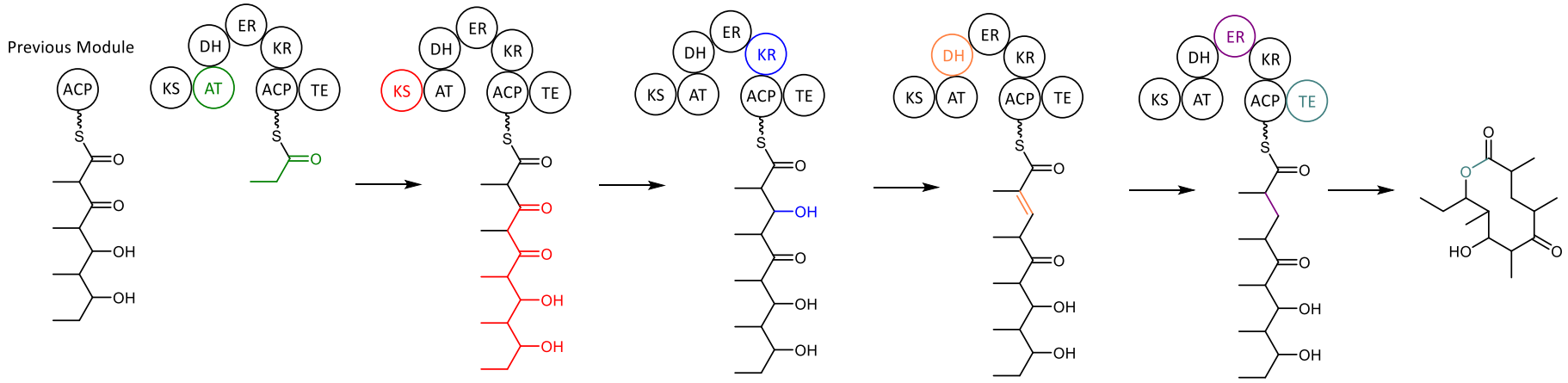


Figure 1.2. Example biosynthesis of an generic natural product by a PKS module. ACP = acyl carrier protein, AT = acyltransferase, selects and attaches the new acyl group to the ACP, shown in green. KS = ketosynthase, attaches the polyketide chain from the previous module to the new module, shown in red. KR = ketoreductase, reduces a ketone to an alcohol on the previously added acyl group, transformation shown in blue. DH = dehydratase, dehydrates an alcohol to an alkene, shown in orange. ER = enoyl reductase, reduces the alkene to an alkane, shown in purple. TE = thioesterase, macrocyclises (but not always) and releases the polyketide from the PKS, shown in aqua.

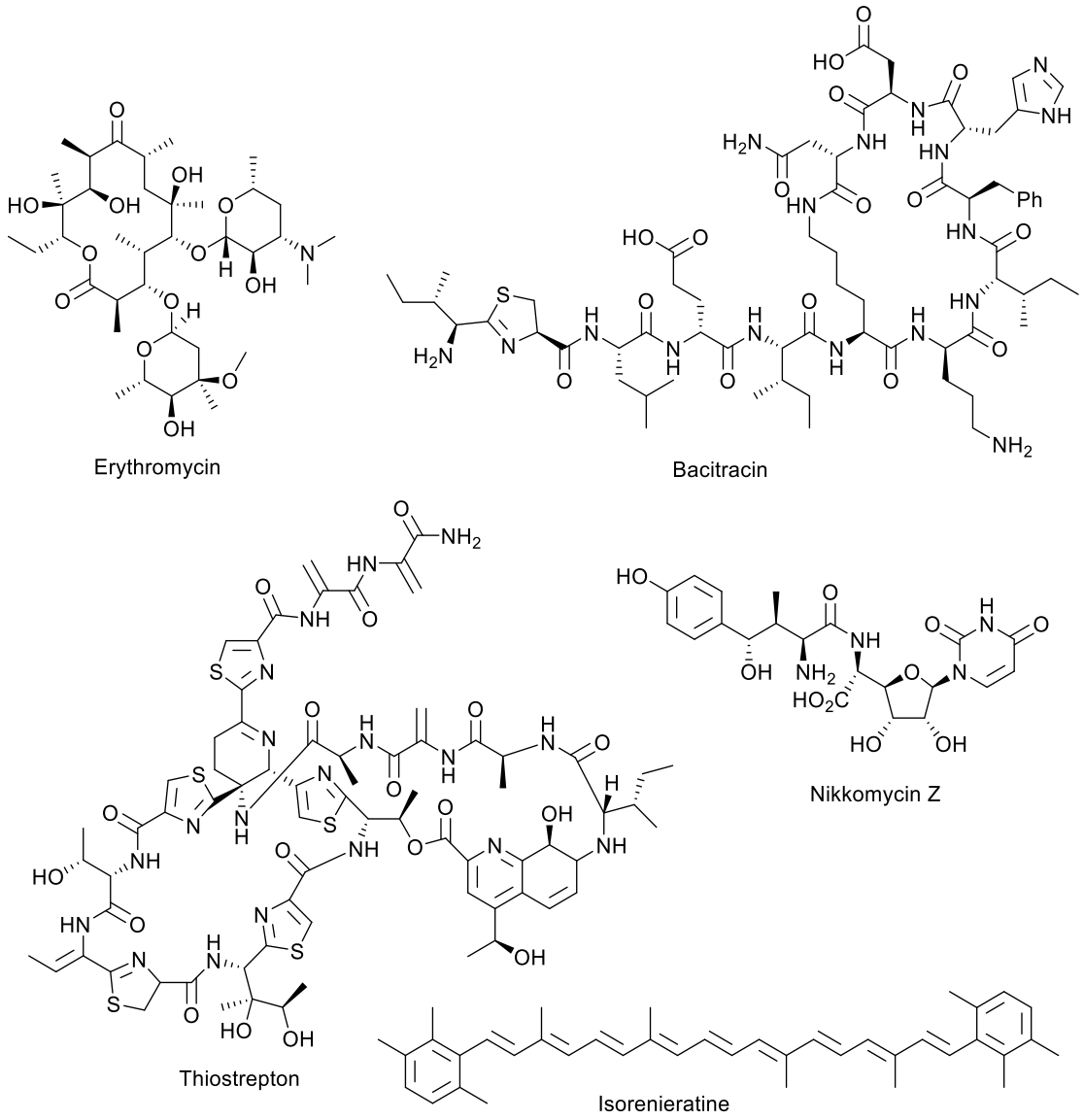


Figure 1.3. Examples of natural products produced by bacteria. Showing their diverse chemical structures.

1.1.5 Non-ribosomal Peptide Synthases (NRPSs)

NRPSs are enzymes which incorporate amino acids into secondary metabolites without ribosomally synthesising the peptides. Notably, unlike for ribosomes, the activity of NRPSs are not solely restricted to using the 20 proteinogenic amino acids but can also incorporate other amino acids such as ornithine or other monomers such as fatty acids[46]. BGCs for non-ribosomal peptides can also contain genes to make custom amino acids which are made bespoke for the NRPS product. NRPSs are modular, with each module being responsible for the incorporation of a single monomer[46]. There are three essential domains for peptide elongation: the adenylation (A) domain to select and activate the amino acids (or other monomers) for addition to the growing polypeptide by activation as an adenosine monophosphate (AMP) ester; the peptidyl carrier domain which tethers the growing peptide chain and the next monomer to be added to the growing chain; and the condensation (C) domain to catalyse the amide bond forming reaction. Finally, once the peptide

has been synthesised, a domain is often involved to release the peptide by hydrolysis or macrocyclisation, just like in PKSs[46]. An example of the synthesis of a hypothetical peptide by an NRPS can be seen in Figure 1.4.

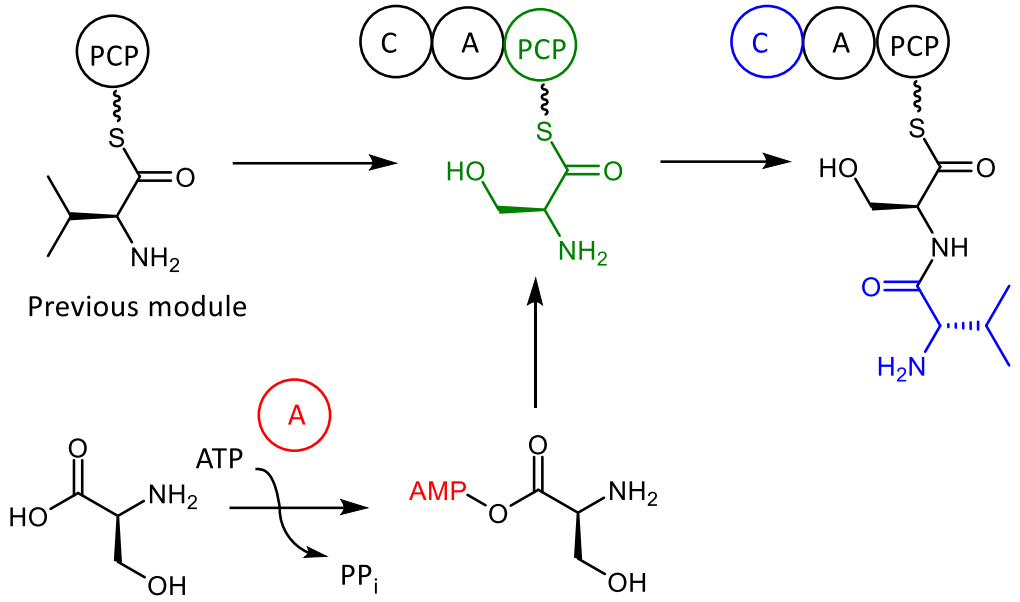


Figure 1.4. Example of biosynthesis for an imaginary dipeptide by an NRPS module. A = adenylation domain, activates the incoming amino acid by adenylating it, shown in red. PCP = Peptide carrier protein, attaches the adenylated amino acid by thioester formation, shown in green. C = condensation domain, catalyses amide bond formation, shown in blue.

An example of a clinically used non-ribosomal peptide is bacitracin (see Figure 1.3), produced by *Bacillus licheniformis* ATCC 10716 which is used to treat wound infections, pneumonia, empyema (pockets of pus in the body cavity) and eye infections[47]. Bacitracin works by inhibiting peptidoglycan synthesis via inhibition of the lipid II cycle; specifically bacitracin binds to the diphosphate lipid carrier undecaprenyl pyrophosphate (UPP) preventing dephosphorylation to undecaprenyl phosphate and therefore preventing recycling of the lipid carrier which is essential for peptidoglycan assembly in the cell wall, stalling peptidoglycan synthesis[48].

The *Bacillus* which produces bacitracin was first isolated from a knee scrape of a child, Margaret Tracy in 1943[49]. Self-resistance to bacitracin is primarily mediated by the BceAB ABC transporter which uses energy from ATP to sterically shield UPP and lipid II whilst also leading to a modification of the cell wall via a complex process that is not yet fully understood[50]. This is an unusual example of an ABC transporter involved in resistance because it does not appear to be simply exporting the natural product to produce accumulation. This transporter system is regulated by a two-component system made up of a sensor histidine kinase and an effector which de-repress transcription of the transport genes upon accumulation of bacitracin[51].

There is also, fascinatingly, a second layer of self-resistance. The UPP phosphatase, BcrC, is upregulated in the producer to compensate for bacitracin inhibition. More available UPP phosphatase leads to less free UPP for bacitracin to bind to by increasing flux through the reaction pathway, and so the lipid II cycle can proceed in the presence of bacitracin[48,51]. Additionally, a system similar to the phage-shock protein (Psp) system operates in the producer *Bacillus*. The phage shock protein system works by detection of membrane stress by the PspB-PspC complex and PspA. PspA is recruited to the membrane by PspB-PspC, releases its binding to PspF and accumulates beneath the damaged membrane or cell wall and appears to stabilise it. Release of PspF from PspA induces expression of the *pspA* and *pspG* operons- inducing increased expression of PspA, PspB, PspC and other proteins[52]. The membrane associated LiaI protein in the *Bacillus* producer of bacitracin recruits the PspA-like protein LiaH into patches which appear to stabilise the membrane beneath the damaged cell wall, protecting the producing *Bacillus* from stress. We therefore have three different mechanisms of resistance in the producer to its natural product: sequestration of the target by BceAB, overexpression of BcrC to compensate for bacitracin inhibition and LiaI to protect the damaged cell[48].

1.1.6 Ribosomally Synthesised Post-translationally Modified Peptides (RiPPs)

Ribosomally synthesised post-translationally modified peptides (RiPPs) are a large family of specialised metabolites made up of short peptides which are synthesised by the ribosome from the 20 proteinogenic amino acids which are then post-translationally modified by tailoring enzymes, to become the mature natural product. While much more recently identified and less well studied than polyketide compounds and non-ribosomal peptides, RiPPs appear to have a wide variety of biological activities and biosynthetic pathways, as well as some fascinating chemistry.

Generally, the peptide precursor of RiPPs has a core peptide, which is the part of the peptide which is modified to form the final peptide natural product, a “leader” peptide which is generally at the N-terminus and sometimes a “recognition sequence” at the C-terminus, both of which are involved in binding to modifying enzymes and peptide cyclisation and are proteolytically cleaved from the core peptide. In eukaryotes, an N-terminal signal peptide may also be present which targets the peptide to the correct subcellular locations/organelles[53].

An example of a RiPP which has progressed to clinical trials is thiostrepton. Thiostrepton targets the ribosome by binding to the 23S rRNA and inhibiting the action of elongation factors, preventing proper functioning of the ribosome. To combat this, self-resistance is encoded by a rRNA methyltransferase, *tsr*, outside of the BGC in the producer organism[54]- this is highly unusual as we expect self-resistance to be encoded in the BGC. This is an example of target modification as a mechanism of resistance or self-resistance. Thiostrepton is also widely used in *Streptomyces*

molecular biology as a selection marker and as an inducing agent for gene expression via induction of the *tipA* promoter[55].

1.1.7 Terpenes

Terpenes are compounds which are often volatile, being known to be the causative compounds of many scents, including limonene in citrus, pinene in pine trees and geosmin, the ubiquitous terpene in *Streptomyces* which causes petrichor, the smell of soil after rain. Geosmin also appears to have a role in insect attraction[56-58] which has led to it being suggested as a possible bait for mosquito traps; specifically, geosmin mediates egg laying in the yellow fever carrying mosquito, *Aedes aegtypti*[58].

Terpenes are formed from five carbon units, isoprenes, which are joined together to form a single linear polyene with branches. These five carbon units, dimethylallylpyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP), are both generated by either the mevalonate (MVA) or 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. Eukaryotes, archaea and some bacteria contain the MEP pathway, while other prokaryotes and plant plastids contain the MVA pathway, however both pathways end in DMAPP and IPP[59,60]. Geranyl pyrophosphate synthase dephosphorylates IPP, causing the formation of a carbocation. Addition of IPP to DMAPP, as seen in Figure 1.5, leads to the generation of the ten carbon geranyl pyrophosphate (GPP). A subsequent addition of IPP by farnesyl pyrophosphate synthase leads to the formation of the fifteen carbon farnesyl pyrophosphate (FPP) and a further addition of IPP leads to the formation of the twenty carbon geranylgeranyl pyrophosphate (GGPP). Classified based on the number of carbons forming the backbone, monoterpenes (C_{10}) such as the citrus scent molecule limonene are synthesised from GPP, sesquiterpenes (C_{15}) such as the *Streptomyces* scent compound geosmin are built from FPP and diterpenes (C_{20}) such as the antibiotic platensimycin are built from GGPP. Final terpenes are cyclised by a terpene cyclase, which holds the polyene in a defined, templated conformation which then allows for the loss of the pyrophosphate group and therefore formation of a carbocation. This then leads to cyclisations and rearrangements, before the final terpene skeleton is formed[60].

Triterpenes (C_{30}) such as the plant derived antibiotic, ursolic acid, are built from a head to head combination of two FPP molecules to form squalene which then goes on to form the final skeleton[61], as seen in Figure 1.5. Similarly, tetraterpenes (C_{40}) such as the cyanobacterial carotenoid, synechoxanthin are formed from two head-to-head GGPP molecules forming phytoene[62,63]. Tailoring enzymes such as cytochrome P450s, glycosyl transferases etc. make the final terpene structure. Terpene biosynthesis pathways are frequently promiscuous; a single terpene cyclase could produce dozens of different hydrocarbon skeletons. General terpene biosynthesis is less well

understood and predictable than NRPS and P KS based pathways and is therefore an area of great interest for pathway engineering[60].

Terpenes appear to be amazingly structurally and functionally diverse (see Figure 1.6). One such terpene produced by green photosynthetic bacteria and some *Actinomycetes*, such as *Streptomyces griseus*, is isorenieratine (see Figure 1.3). This carotenoid pigment molecule is a tetraterpene which plays a role in protection from photooxidative stress caused by sunlight. In green photosynthetic bacteria, such as cyanobacteria, isorenieratine and other carotenoids play a role in photosynthesis by quenching the reactive excited chlorophyll or singlet oxygen (a reactive oxygen species generated by excited chlorophyl acting on oxygen) as well as passing light energy to chlorophyl, being essential for the photosystems of cyanobacteria[64].

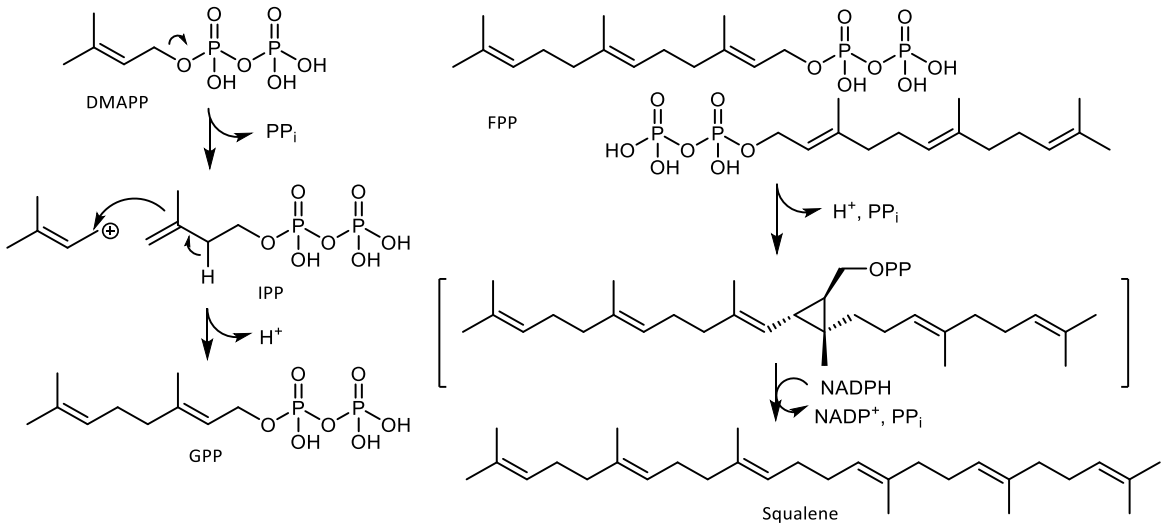


Figure 1.5. Generation of geranyl pyrophosphate and squalene. Intermediate of squalene formation showed in square brackets. DMAPP = dimethylallyl pyrophosphate, IPP = isopentenyl pyrophosphate, GPP = geranyl pyrophosphate, FPP = farnesyl pyrophosphate.

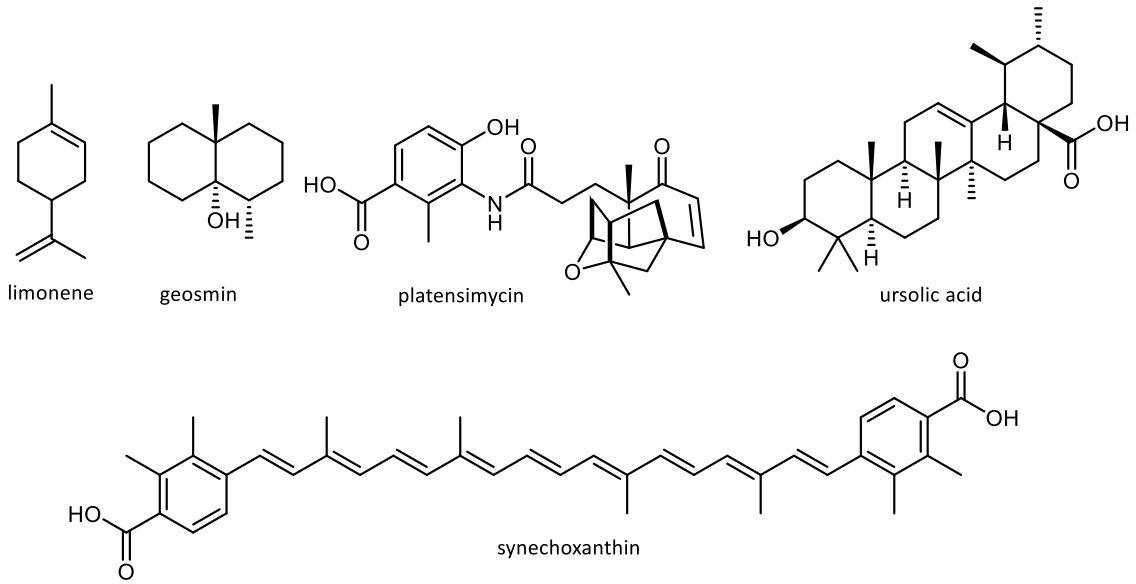


Figure 1.6. Structures of some examples of terpenes. Limonene is a monoterpene which gives citrus their aroma, geosmin is a sesquiterpene produced by Streptomyces which gives the distinctive smell of soil, platensimycin is a diterpene antibiotic compound produced by Streptomyces platensis, ursolic acid is a triterpene compound found in the skin of certain fruits such as apples, and synechoxanthin is a tetraterpene produced as the major carotenoid produced by the cyanobacterium *Synechococcus* sp. PCC 7002.

1.1.8 Other classes of natural products

Other than PKS, NRPS, RiPP and terpene pathway derived natural products, there are also a variety of other compounds produced by the gene products encoded within BGCs. As with all of the above examples, the biosynthetic genes for these compounds have evolved from primary metabolism and use building blocks from primary metabolism but continue to widen the chemical space which bacteria can access. This can also include hybrid clusters, for example where PKS and NRPS machinery work together to produce the final compound.

Nikkomycin Z is an antifungal nucleoside which targets cell wall biosynthesis by competitively inhibiting fungal chitin synthases. Nikkomycin is an analogue of uridine diphosphate-*N*-acetyl glucosamine (NAG), which is the substrate of chitin synthase, with chitin being a polymer of NAG. Nikkomycin Z is biosynthesised by *Streptomyces tendae*, from hydroxypyridylhomothreonine which is generated by the action of eleven genes, starting with the deamination of L-lysine and L-glutamate. Due to the lack of chitin synthases in the bacterial producer, there is no need for an self-resistance mechanism in this BGC.

1.2 Self-resistance in natural product producers

A key problem faced by organisms that produce natural products is the potential for self-killing, hence self-resistance mechanisms are required. These can take multiple forms and include: general efflux pumps to keep cytosolic concentrations low; detoxification methods such as acetyltransferases to render the natural products inactive, or second copies of housekeeping

proteins if these are the targets. It is for this reason that identification of essential housekeeping genes for which multiple copies exist is effective as a method of genome mining for new antimicrobials or for identifying the target of newly discovered antimicrobials. These self-resistance genes are often found as part of the BGC encoding production of the natural product, and so when the BGC is transferred to new strains, the new strain can gain resistance to its newly acquired compounds. See Figure 1.7 for a simplified summary of self-resistance mechanisms employed by bacteria.

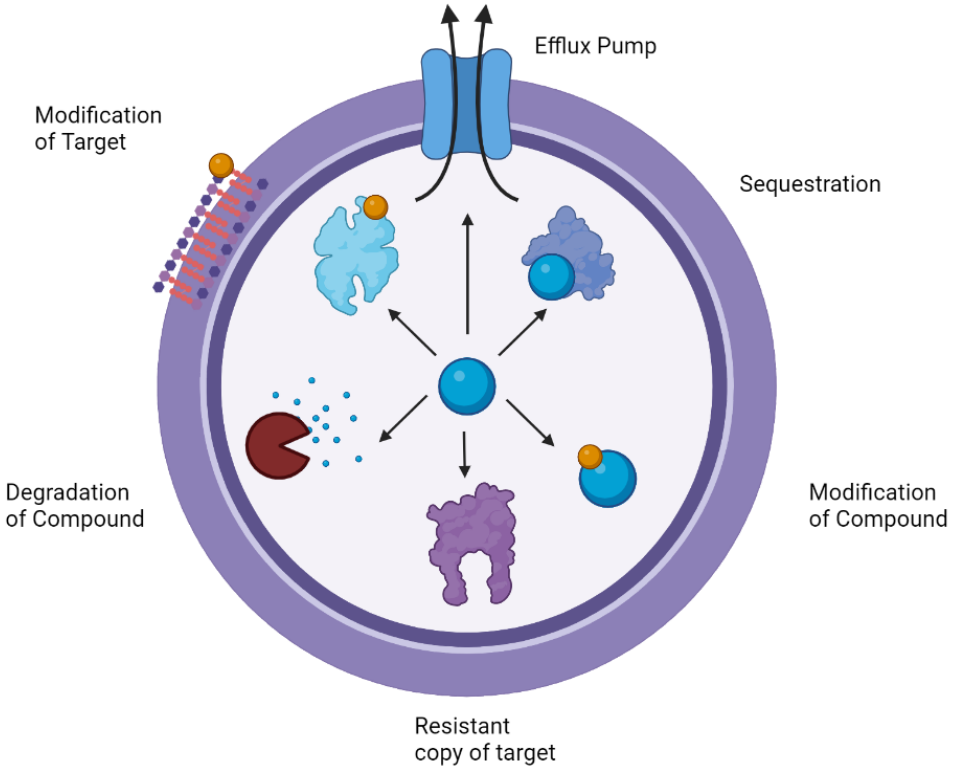


Figure 1.7. Cartoon representation of the main self-resistance mechnanisms deployed by natural product producers. Natural product molecule representaed by blue sphere, modifications representated by orange sphere. Figure generated in BioRender.

Because of the importance of self-resistance for survivability when producing natural products, the expression of self-resistance genes and biosynthetic genes must be tightly regulated in the cell; if the biosynthetic genes are expressed in the absence of the self-resistance gene, the natural product could be lethal to the producing cell[65]. If the self-resistance gene is expressed in the absence of biosynthetic genes this is less catastrophic but could lead to unnecessary metabolic drain.

1.2.1 Self-resistance by Efflux Pumps

Both general efflux pumps or specific efflux pumps can rapidly transfer the newly synthesised product out of the cytoplasm so that it cannot accumulate and harm the producing cell. For example, in the myxin producer, *Lysobacter antibioticus*, an efflux pump, LexABC confers self-resistance by preventing accumulation in the producer by active transport of myxin out of the cell[66].

1.2.2 Self-resistance by Sequestration

Another self-resistance mechanism involves the synthesis of proteins that can act to sequester antibiotic products in order to protect susceptible housekeeping proteins from being inhibited and thereby protecting the cell. In *Streptomyces* sp. CB03234, three "self-sacrifice proteins" confer self-resistance to the enediyne calicheamicin by binding it with nanomolar dissociation constant (K_d) values[67]. Enediynes target DNA via binding to the major groove and initiating double strand breaks following rearrangement to a transient diradical which is extremely reactive. This self-resistance mechanism has been identified in a number of enediyne producers.

In a variation on this theme, the molecular target of the natural product can be sequestered by the self-resistance protein, preventing binding and inhibition by the ligand. This was recently observed for the pyxidicyclines produced by the myxobacterium *Pyxidicoccus fallax*[68]. These compounds were discovered by screening for pentapeptide repeat protein (PRP) containing BGCs. PRPs are proteins with a fascinating structure which appears to mimic DNA and bind to gyrase and topoisomerase enzymes, preventing binding by gyrase and topoisomerase inhibitors[69,70].

1.2.3 Self-resistance by Compound Modification

Modification of the compound is also an effective strategy. These detoxification reactions reduce the affinity of the natural product for any housekeeping proteins; these "prodrugs" can then be safely exported from the cell, where the modification can be removed by a secreted enzyme yielding an active form of the natural product. An example of this can be found in streptomycin production in, *Streptomyces griseus*. The last enzyme of the biosynthetic pathway is streptomycin 6-phosphotransferase (AphD) which is co-regulated with the biosynthetic genes for streptomycin production. AphD phosphorylates and deactivates streptomycin, and the phosphorylated compound is then exported from the cell by transporters, StrVW before it is dephosphorylated by the phosphatase StrK outside of the cell to yield streptomycin[71].

1.2.4 Self-resistance by Resistant Copy of Target

To circumvent toxicity a resistant copy of the molecular target is often encoded in the BGC encoding a natural product. This can provide an idea as to both what the target is, and insight into the mode of action of the compound. For example, in *Streptomyces platensis*, self-resistance to the antibacterial molecule platensimycin is conferred by PtmP3, a resistant copy of the condensing enzyme FabF which is essential for fatty acid biosynthesis. Interestingly, PtmP3 can replace the activity of a second condensing enzyme FabH, which is also involved in fatty acid biosynthesis and is inhibited by platensimycin[72].

1.2.5 Self-resistance by Degradation of Compound

Most β-lactam antibiotic producing *Streptomyces* (and indeed most *Streptomyces*) encode β-lactamase genes which hydrolyse the β-lactam ring, destroying the natural product[73]. This is less

than ideal from the standpoint of producing a compound and so most *Streptomyces* producers of β-lactams also appear to have genes for penicillin binding proteins in their genomes, which appear to sequester the β-lactams and could be working together with β-lactamases. There is no specific link between β-lactam production and these proteins, so this may be a poor example of self-resistance mechanism[74].

1.2.6 Self-resistance by Target Modification

The target of the natural product can be chemically modified so that it cannot bind the natural product produced or has significantly lower binding affinity. Expression of the enzymes for these transformations are often linked to biosynthesis of the natural product, ensuring self-resistance. An example of this can be found in the 50S ribosomal subunit inhibitor tylosin, produced by *Streptomyces fradiae*. There are four resistance proteins encoded in the tylosin BGC; TlrA, B, C and D. Of these, TlrA, B and D are RNA methyltransferases which methylate the rRNA of the 50S ribosomal subunit and protect it from inhibition by tylosin, while TlrC is an efflux pump which exports the compound and prevents its accumulation at toxic concentrations[71]. This is an interesting example where multiple resistance mechanisms are employed by the producer organism to protect itself.

As noted, the identification of self-resistance genes is a potentially effective approach for genome mining to selectively discover new natural product inhibitors targeting a specific biological process; if we can understand self-resistance mechanisms in natural product producers, we can potentially identify the target or mechanism of action of the unknown compounds produced by cryptic BGCs before discovering the identity of the natural product[68,73,75].

1.3 Aminoacyl tRNA Synthetases

Essential components of translation, aminoacyl tRNA synthetases (aaRSs) acylate tRNA with their cognate amino acid at the 2’ or 3’ hydroxy group, allowing the tRNA to then bring the correct amino acid to the ribosome during translation for addition to the growing polypeptide chain. The tRNA is recognised by the cognate synthetase by binding at the anticodon binding domain, with the anticodon being the section of the tRNA which binds by base pairing to the codon on the mRNA that is being “read” by the ribosome[76,77]. In the aminoacylation reaction, an amino acid firstly reacts with ATP to form an amino acid-AMP intermediate. The activated amino acid is then joined to its cognate tRNA via the 2’ or 3’ hydroxyl group of the tRNA, depending on the class of aaRS. To maintain fidelity, aaRSs often include editing domains which can recognise misacylated tRNA or misadenylated amino acids and hydrolyse them[78-81].

Due to their essential role, aaRSs are attractive targets for drug discovery as inhibition of just one aminoacyl tRNA synthetase will usually lead to the death of the cell when translation is no longer possible. However, this is not always the case, with some aaRSs being capable of some tRNA

promiscuity, lacking proof-reading domains or having downstream enzymes which convert the amino acid attached to the tRNA once it has dissociated from the aaRS. Indeed, some bacterial strains do not appear to have a full complement of aaRSs and rely on this. On the other hand, many strains have multiple paralogues of their aaRSs, which can be involved in non-canonical functions or represent self-resistance mechanisms against aaRS targeting antibiotics[82].

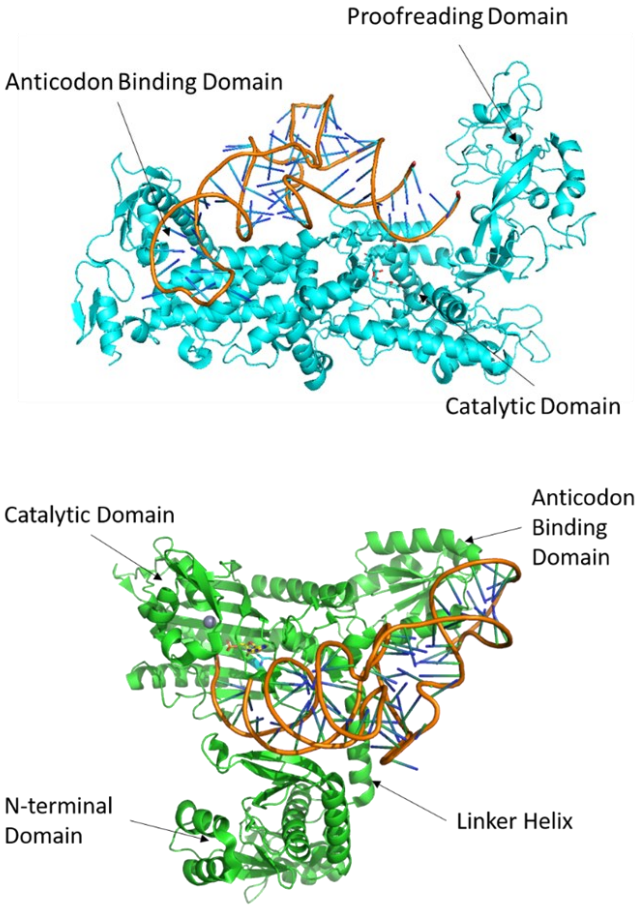
Thought to have been some of the first proteins to evolve, aaRS urzymes (primitive enzymes) appear to have evolved before the last universal common ancestor[83] with the genetic code then expanding based on the emergence of new aaRSs before the eventual "settling" of the genetic code[82,84-88], possibly replacing RNA-enzymes which have been suggested to precede them in the RNA world[83,86,88-90]. Horizontal gene transfer and gene duplication appears to have played an important role in the evolution of these enzymes, and the dissemination of secondary or tertiary copies throughout the kingdoms of life[82,91,92].

aaRSs are present in all forms of cellular life with little variation at the structural level between homologues of the same aminoacyl tRNA synthetase in different organisms. Broadly, there are two classes of aaRS; Class I which handle IleRS, LeuRS, ValRS, CysRS, MetRS, ArgRS, GlnRS, GluRS, LysRS, TyrRS and TrpRS and class II which handle: AlaRS, ProRS, HisRS, SerRS, ThrRS, GlyRS, AsnRS, AspRS, LysRS, SepRS, PylRS and PheRS[91]. These two classes are phylogenetically and structurally distinct from each other as can be seen in Figure 1.8, with subclasses based on the structures of their subunits (see Table 1.1). Although they have evolved from two highly distinct ancestors, these two classes of enzymes catalyse almost identical reactions, in an exquisite example of convergent evolution[93]. Amazingly, it appears that LysRSs evolved in two separate instances meaning that examples of Class I LysRS can only be found in some archaea and bacteria, while Class II LysRS are found in all other organisms. Class I LysRS can complement the knockout of class II LysRS in *E. coli*[94].

Table 1.1. Listing of subclasses of aaRSs, listing editing capability and quaternary structure; α = monomer, α_2 = homodimer, $\alpha_1\alpha_2$ = heterodimer, β denotes an inactive chain, α_4 = homotetramer ,adapted from Perona *et al*. 2012[91].

Subclass	Amino Acid	Editing Capability	Quaternary Structure
IA	Isoleucine	Yes	α
	Leucine	Yes	α
	Methionine	Yes	$\alpha_1\alpha_2$
	Valine	Yes	α
IB	Cysteine	No	$\alpha_1\alpha_2$
	Glutamine	No	α
	Glutamate	No	α

IC	Tryptophan	No	α_2
	Tyrosine	No	α_2
ID	Arginine	No	α
IE	Lysine	No	α
IIA	Glycine	No	α_2
	Histidine	No	α_2
	Proline	Yes	α_2
	Serine	Yes	α_2
	Threonine	Yes	α_2
IIB	Asparagine	No	α_2
	Aspartate	No	α_2
	Lysine	Yes	α_2
IIC	Alaine	Yes	α_2 (α in eukaryotes)
	Glycine	No	$(\alpha\beta)_2$
	Phenylalanine	Yes	$(\alpha\beta)_2$, α
	Pyrollysine	No	α_2
	O-phosphoserine	No	α_4



Class I

Amino Acids: Cys, Glu, Ile, Lys, Leu, Met, Gln, Arg, Val, Trp, Tyr

Rossman Fold for nucleotide binding

Usually aminoacylates 2' hydroxyl of tRNA

Often monomeric, can be dimeric

Class II

Amino Acids: Ala, Asp, Phe, Gly, His, Lys, Asn, Pro, Ser, Thr, Sep, Pyl

Anti-parallel β fold for nucleotide binding.

Usually aminoacylates 3' hydroxyl of tRNA

Usually dimeric or multimeric

Figure 1.8. Summary of class I and class II aaRSs. The example of Class I is IleRS, Protein Data Bank (PDB) file 1FFY[95] , the example of Class II is ThrRS, PDB file 1QF6[77]. In these examples, proofreading domains are present, partway through the split catalytic domain in the example of IleRS, and as the N-terminal domain for ThrRS. Figure generated in PyMol[96].

The architectures of the different aaRSs are diverse, with the core features of catalytic domain and anticodon binding domain being present in (almost) all proteins, but then with varying combinations of editing domains and other features. Class I aaRSs use Rossman Folds for their active sites, whereas class II uses an anti-parallel β fold for nucleotide binding[91]. Different aaRSs also use different mechanisms to differentiate between different amino acids at their active site, for example ThrRSs rely on an active site zinc for substrate recognition[97] while PheRS relies on a sandwich of π stacking interactions and hydrophobic interactions in the active site for substrate recognition[85]. In eukaryotes this structural diversity is further complicated by the addition of localisation signals and other domains involved in non-canonical functions such as interaction with members of signalling pathways, such as the tumour necrosis factor (TNF) pathway[98,99].

1.3.1 Threonyl tRNA synthetases (ThrRS)

ThrRS is a class IIA aaRS. In bacteria, the ThrRS comprises an N-terminal editing domain made of two halves: an editing regulation section (N1) which is ubiquitin like, and an editing catalytic section

(N2). This editing domain is connected to the catalytic domain which contains the signature anti-parallel β sheet for nucleotide binding and a C-terminal anticodon binding domain. The overall reaction carried out by ThrRSs can be seen in Figure 1.9, proceeding with adenylation to activate the threonine followed by aminoacylation of the cognate tRNA ($tRNA^{Thr}$) at the 3' hydroxyl group of its terminal ribose. When discussing specific amino acid residues, the numbering for the *E. coli* protein (EcThrRS) will be used throughout, unless otherwise stated. Where specific bases on the tRNA are discussed, the numbering is from the *E. coli* $tRNA^{Thr}$, encoded by the *thrW* gene. There are several different ThrRS structures available in the PDB, as seen in Table 1.2, which demonstrate the range of different conformations which the protein samples throughout its reaction cycle. An example of the structure of ThrRSs can be seen in Figure 1.10.

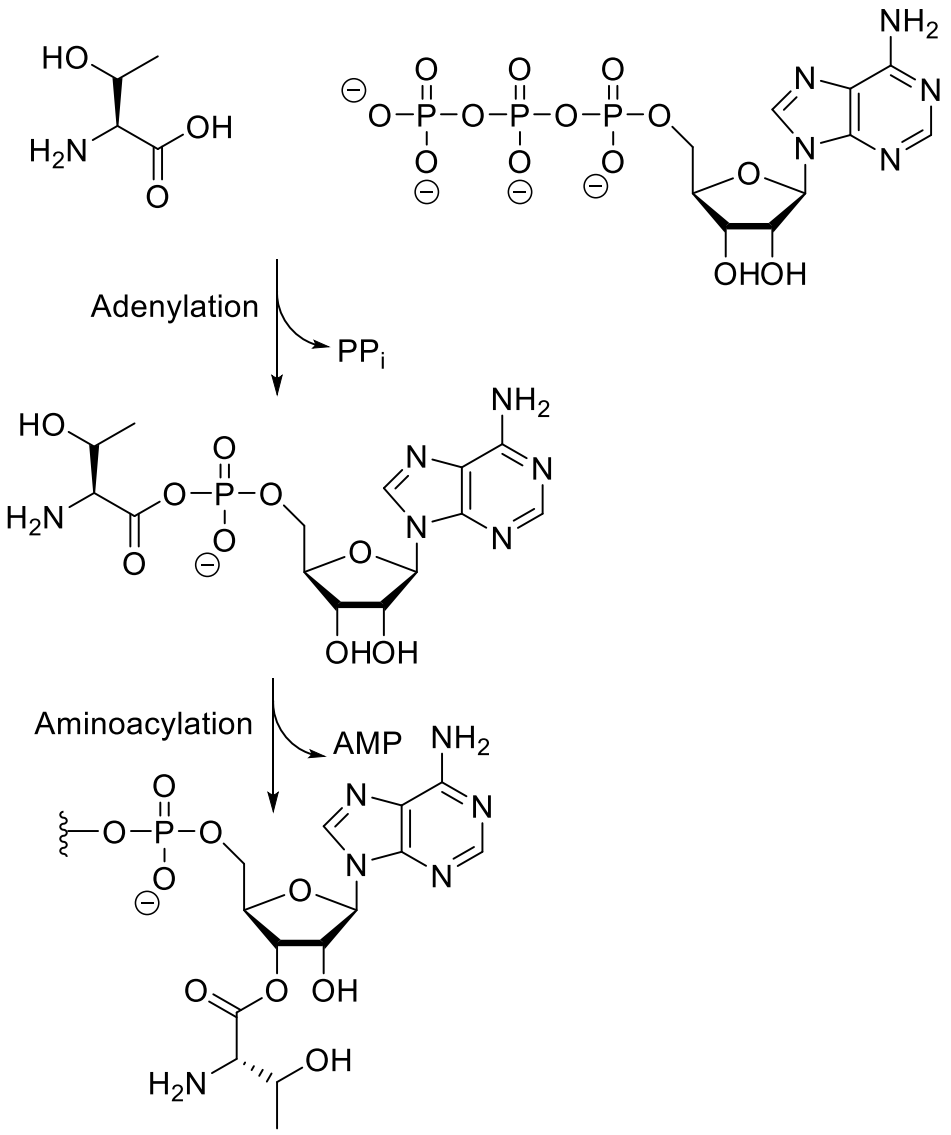


Figure 1.9. Reaction scheme for threonyl tRNA synthetases. Adenylation of threonine by ATP leads to formation of threonyl AMP (ThrAMP), which is followed by aminoacylation at the 3’ hydroxyl of the tRNA.

Table 1.2. A summary of some of the key structures deposited to the PDB. NMR = Nuclear Magnetic Resonance, XRD = X-Ray Diffraction. ACB = anticodon binding domain. ThrSAA = 5’-O-(threonylsulfamoyl)adenosine, SerSAA = 5’-O-(serinylsulfamoyl)adenosine, BC194 is a semisynthetic derivative of borrelidin, eIF4E is eukaryotic translation initiation factor 4E, A3S is 3’-O-serinyl adenosine, A3T is 3’-O-threonyl adenosine and A3G is 3’-O-glycyl adenosine. Blue = bacterial, green = eukaryotic, yellow = archaeal.

PDB File	Technique	Resolution (Å)	Organism	Domains Present	Ligands
1EVK[97]	XRD	2	*Escherichia coli*	Catalytic & ACB	Thr
1EVL[97]	XRD	1.55	*Escherichia coli*	Catalytic & ACB	ThrSAA
1FYF[100]	XRD	1.65	*Escherichia coli*	Catalytic & ACB	SerSAA
1KOG[101]	XRD	3.5	*Escherichia coli*	Full Protein	ThrSAA, 5'UTR of *thrS* mRNA
1NYQ[102]	XRD	3.2	*Staphylococcus aureus*	Full Protein	ThrSAA
1NYR[102]	XRD	2.8	*Staphylococcus aureus*	Full Protein	ATP, Thr
1QF6[103]	XRD	2.9	*Escherichia coli*	Full Protein	AMP, $tRNA^{Thr}$
1TJE[79]	XRD	1.5	*Escherichia coli*	Editing Domain	None
1TKE[79]	XRD	1.46	*Escherichia coli*	Editing Domain	Ser
1TKG[79]	XRD	1.5	*Escherichia coli*	Editing Domain	SerSAA
1TKY[79]	XRD	1.48	*Escherichia coli*	Editing Domain	A3S
1WWT	NMR	N/A	*Homo sapiens*	TGS Domain	None
1Y2Q[104]	XRD	1.95	*Pyrococcus abyssi*	Editing Domain	None
2HKZ[105]	XRD	2.1	*Pyrococcus abyssi*	Editing Domain	Ser
2HL0[105]	XRD	1.86	*Pyrococcus abyssi*	Editing Domain	A3S
2HL2[105]	XRD	2.6	*Pyrococcus abyssi*	Editing Domain	SerSAA
3A32[106]	XRD	2.3	*Aeropyrum pernix*	Catalytic & ACB	None
3PD3[107]	XRD	1.86	*Pyrococcus abyssi*	Editing Domain	A3T
3PD4[107]	XRD	2.4	*Pyrococcus abyssi*	Editing Domain	A3G
3UGQ[108]	XRD	2.1	Mitochondrial *Saccharomyces cerevisiae*	Catalytic & ACB	None
3UH0[108]	XRD	2	Mitochondrial, *Saccharomyces cerevisiae*	Catalytic & ACB	ThrSAA
4EO4[109]	XRD	2.87	Mitochondrial *Saccharomyces cerevisiae*	Catalytic & ACB	SerSAA
4P3N[110]	XRD	2.6	*Homo sapiens*	Catalytic & ACB	Borrelidin
4P3P[110]	XRD	2.1	*Escherichia coli*	Catalytic & ACB	Borrelidin

4TTV[111]	XRD	2.8	*Homo sapiens*	Catalytic & ACB	BC194
4YYE[112]	XRD	2.30	Mitochondrial *Saccharomyces cerevisiae*	Catalytic & ACB	ThrSAA, $tRNA^{Thr}$
5XLN[113]	XRD	1.9	*Homo sapiens*	UNE-T	eIF4E
5ZY9	XRD	2.5	*Phytophthora sojae*	Catalytic & ACB	Borrelidin
7L3O	XRD	2.1	*Cryptosporidium parvum*	Catalytic & ACB	None

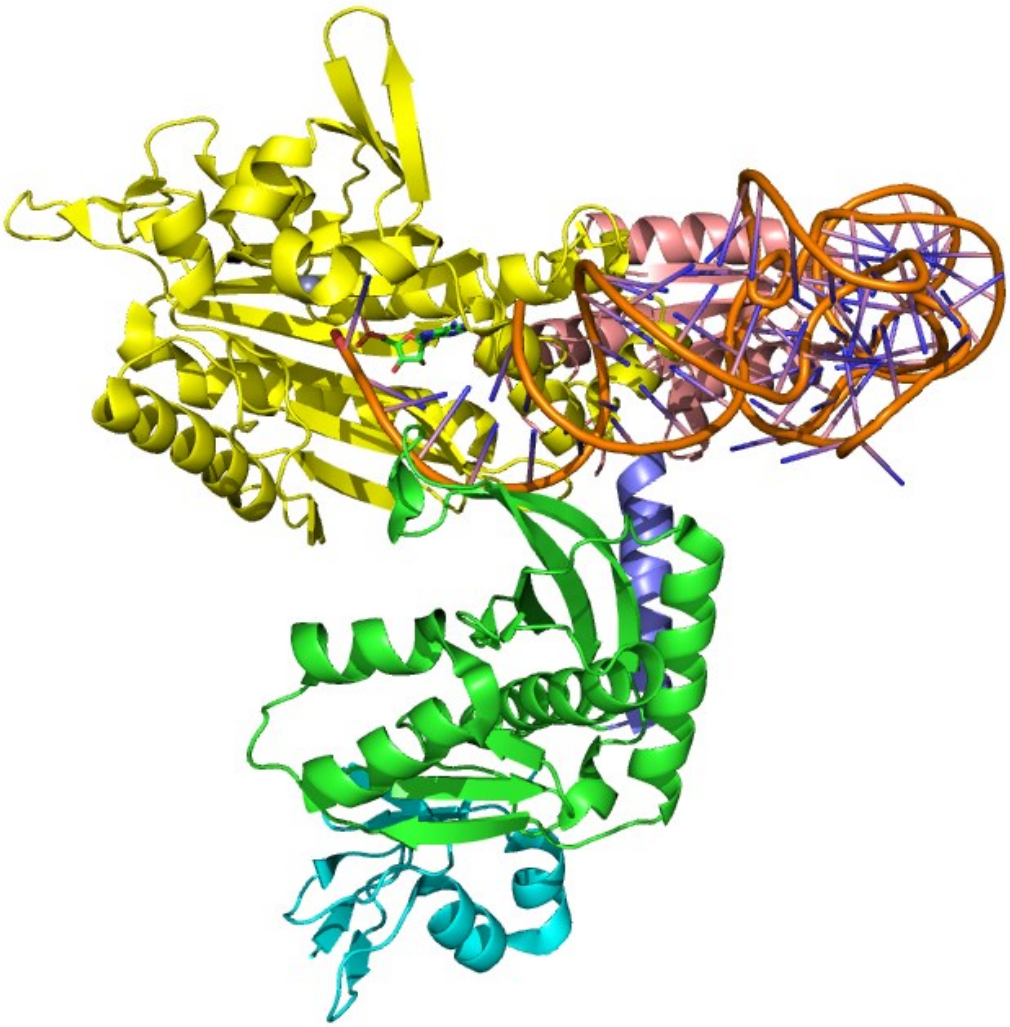


Figure 1.10. Structure of *E. coli* threonyl tRNA synthetase (EcThrRS) in complex with AMP and the cognate tRNA. N1 domain shown in cyan, N2 in green, linking helix in lavender, catalytic domain in yellow, anticodon binding domain in peach, tRNA in orange. Protein and tRNA shown in cartoon representation, AMP in stick representation and zinc as a grey sphere. PDB file 1QF6[77], figure generated in PyMol[96]. See Figure 1.11, Figure 1.12 and Figure 1.13 for zoomed in views of the aminoacylation active site, Figure 1.14 for a zoomed in view of the anticodon binding domain and Figure 1.16 for a zoomed in view of the editing active site.

The threonine substrate (see Figure 1.11) is primarily recognised by a Zn^{2+} ion bound to two histidines (and a cysteine (C334)- which binds to the threonine β-hydroxy and amine groups. Threonine is also orientated in the protein by a hydrogen bonding network between the β-hydroxyl of Thr and D383, and the Thr amine group and Y462 interact to further orientate the threonine and locate the carboxyl group ready for the adenylation reaction. A conformational change also occurs upon threonine binding which leads to interaction of R363 with the threonine carboxylate[102].

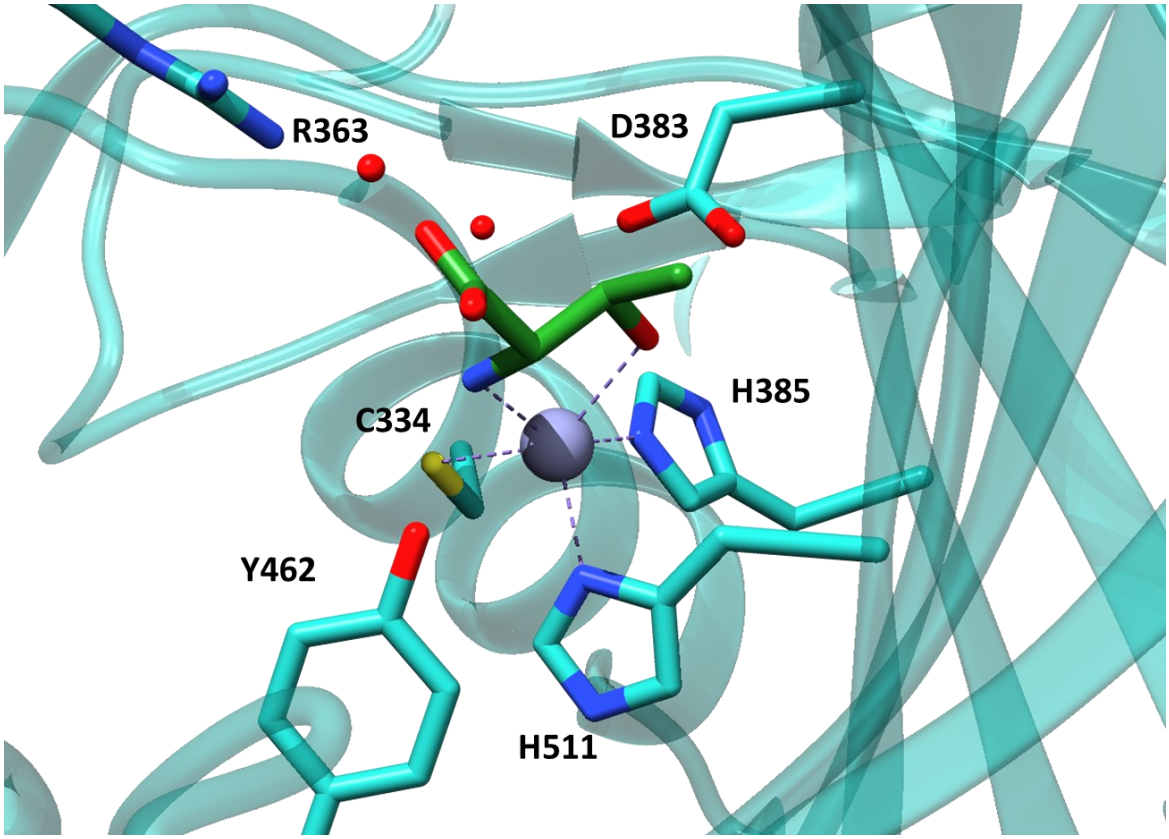


Figure 1.11. Crystal structures showing binding of threonine in ThrRS. *E. coli* ThrRS (EcThrRS), PDB file 1EVK[97], threonine substrate and important binding residues shown in stick representation with rest of protein shown in cartoon representation. Threonine is coloured green with the protein coloured cyan. Residues with explicit interactions with the zinc (represented as a grey sphere) or the threonine are labelled with amino acid identity and numbering, as per EcThrRS. Figure generated in UCSF Chimera[114].

Binding of threonine and ATP can occur in any order, with large scale conformational changes associated with each binding event. ATP binds to the active site in a bent conformation such that the β- and γ-phosphates bend back towards the adenosine (see Figure 1.12), with the phosphates coordinated to a Mg^{2+} ion which in turn also coordinated to the E365 residue. The γ-phosphate also binds to R375 which in turn forms a salt bridge to E365. Both of these residues are essential for the correct positioning of the pyrophosphate, and π stacking of the adenosine ring with R520 and F379 poise the ATP ready for action. F379 is a methionine in the *Staphylococcus aureus* ThrRS (SaThrRS)

crystal structure, the only ATP bound structure available on the PDB, with methionine playing the same role. K465 appears to be catalytically important by increasing the electropositivity of the α phosphate and stabilising the intermediate of the adenylation reaction, while R363 which is important in positioning of the threonine carboxylic acid group also binds the α phosphate. Binding of ATP appears to break a salt bridge between R476 and E527 which leads the protein to enter a conformation that allows tRNA binding to proceed in an effective way[102]. Productive tRNA binding will follow generation of the threonyl adenylate intermediate.

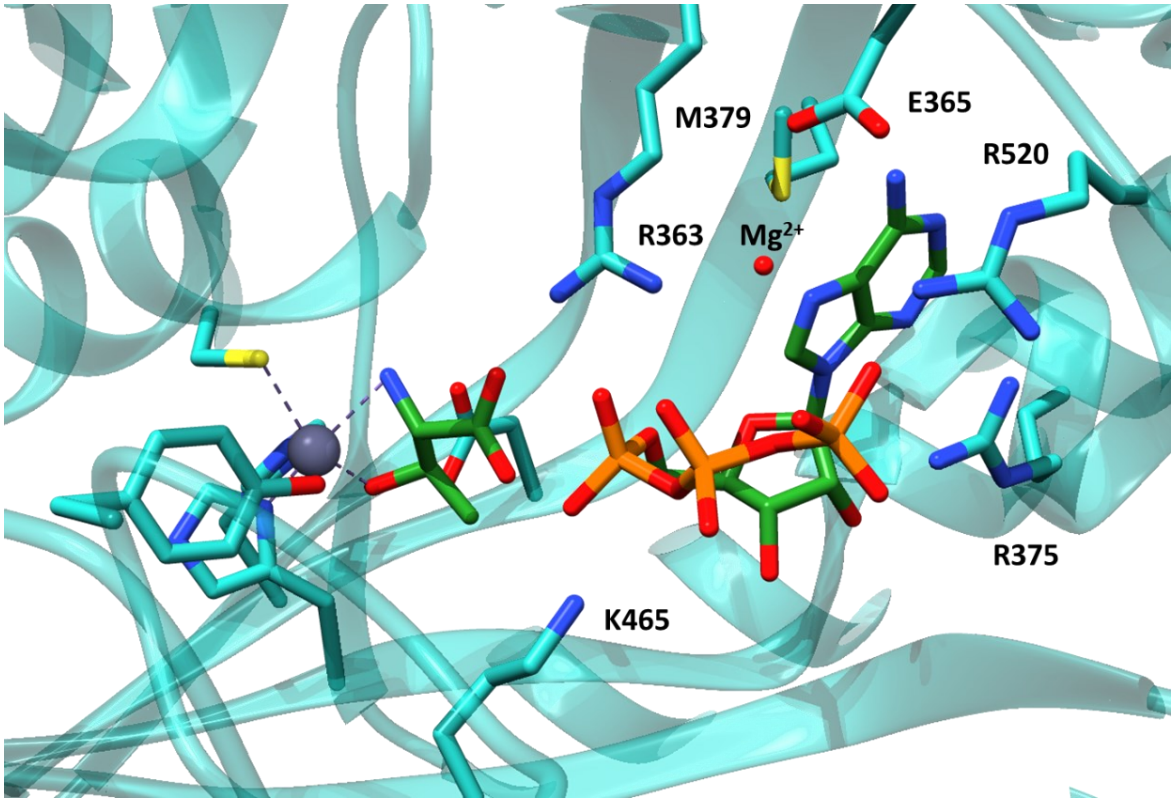


Figure 1.12. Crystal structures showing binding of ATP and threonine in ThrRS. *S. aureus* ThrRS (SaThrRS), PDB file 1NYR[102] threonine substrate, ATP and important binding residues shown in stick representation with rest of protein shown in cartoon representation. Threonine and ATP are coloured green with the protein coloured light blue. Residues with explicit interactions with ATP are labelled with amino acid identity and numbering, as per EcThrRS. Figure generated in UCSF Chimera[114].

Interaction with the tRNA occurs via major groove binding at the 3' end of the tRNA hairpin loop, with involvement of both the catalytic domain and the editing domain. The 3' end of the tRNA can insert into the active site (see Figure 1.13) of the catalytic domain for the aminoacylation reaction before swinging across to the editing domain without dissociation from the anticodon binding domain and the main major groove binding[77,79]. Binding of the terminal adenosine (A76) of the tRNA occurs via π-stacking of the adenosine with R363 and Y313 and hydrogen bonding of N6 with the main chain carbonyl of A316 via hydrogen bonding. A hydrogen bonding network of the 2' hydroxyl of the ribose of A76 with H309 and Y462 and the 3' hydroxyl of the ribose of A76 with Q484. The 4'

oxygen of the ribose of A76 hydrogen bonds to Y313. These interactions all place the 3' hydroxyl of the ribose of A76 in contact with the phosphate of the ThrAMP[77]. Of the 3' terminal end of the tRNA, C74 also makes hydrogen bonds with R375 but A73 and C75 make no direct contact with the protein[77].

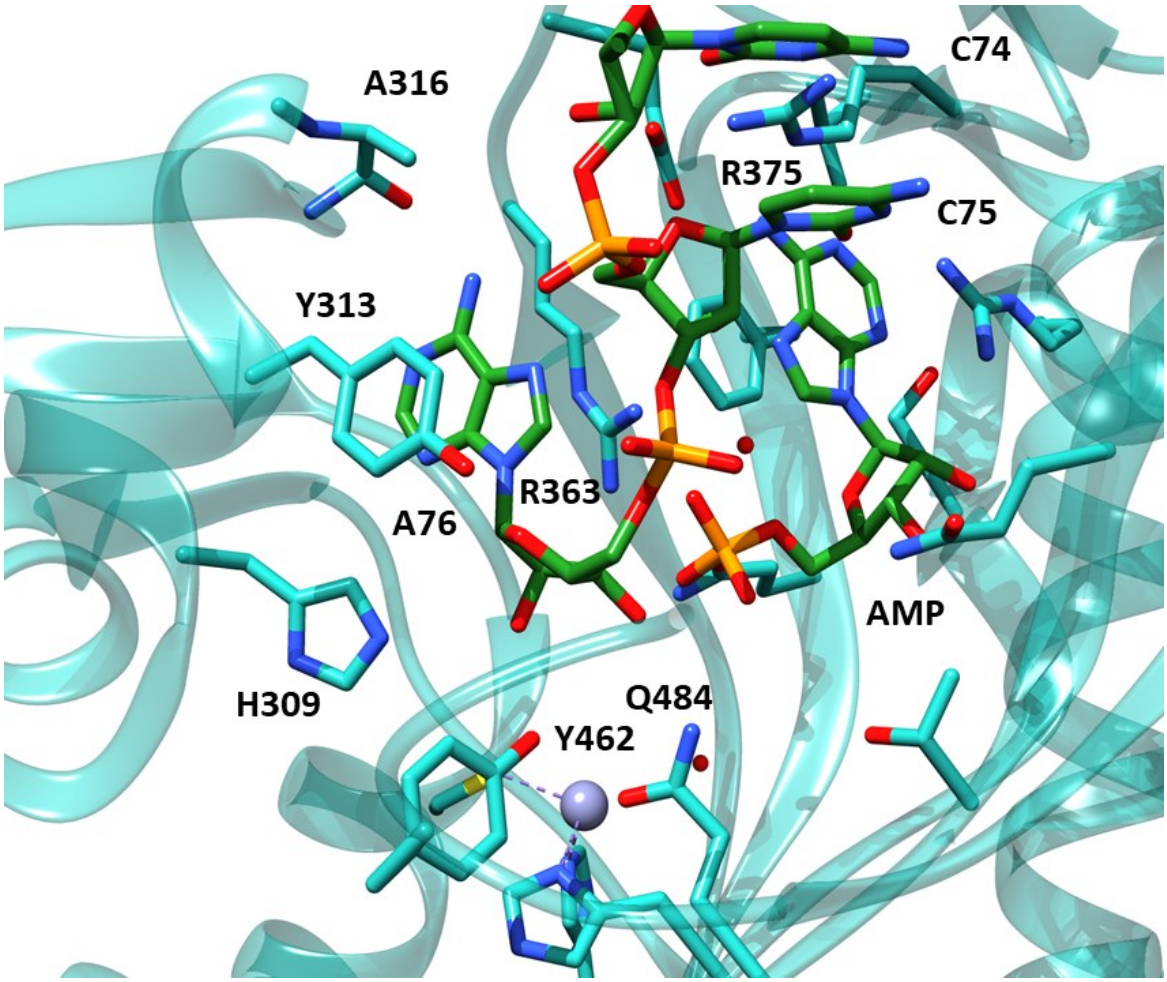


Figure 1.13. Crystal structures showing binding of the tRNA 3' end to the active site of ThrRS. EcThrRS, PDB file 1QF6[77]. tRNA, AMP and important binding residues shown in stick representation with rest of protein shown in cartoon representation. tRNA and AMP are coloured green with the protein coloured light blue. Residues with explicit interactions with tRNA are labelled with amino acid identity and numbering, as per EcThrRS, numbering of tRNA nucleotides is per the thrW encoded $tRNA^{Thr}$ from *E. coli*. Figure generated in UCSF Chimera[114].

The anticodon binding domain of ThrRS recognises the anticodon (C/G/U34, G35, U36) via binding to the major groove (see Figure 1.14)- G35 and U36 are the major recognition determinants, but all three anticodon bases interact with the protein. G35 and U36 are adjacently sat in a hydrophobic pocket composed of I547, I578, I582 and V595, with hydrogen bonding between the two bases themselves. G35 shows specific interactions with E600, and U36 with R609. C34 interacts with N575 but the interaction is unimportant for anticodon recognition[77]. Fascinatingly, it has been observed in *E. coli* that EcThrRS negatively regulates translation of itself via binding to a tRNA-like loop in the

5'UTR of its mRNA which prevents its translation and therefore prevents accumulation of EcThrRS in the cell[77].

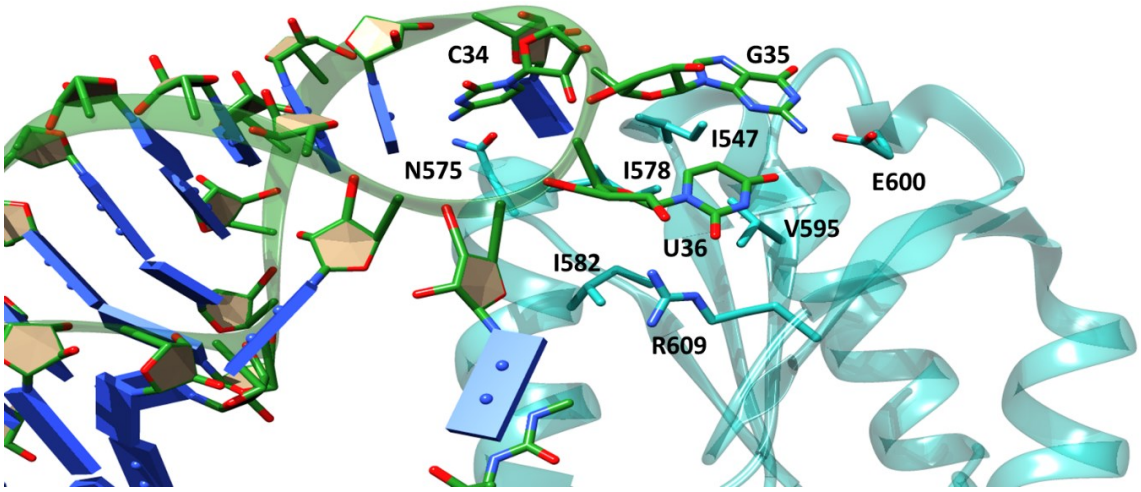


Figure 1.14. Crystal structures showing binding of the tRNA anticodon to the anticodon binding domain of ThrRS. EcThrRS, PDB file 1QF6[77]. tRNA anticodon and important binding residues shown in stick representation with rest of protein and tRNA shown in cartoon representation. tRNA is coloured green with the protein coloured light blue. Residues with explicit interactions with tRNA are labelled with amino acid identity and numbering, as per EcThrRS, numbering of tRNA nucleotides is per the thrW encoded $tRNA^{Thr}$ from E. coli. Figure generated in UCSF Chimera[114].

With ThrRSs being obligately dimeric, the cognate tRNA binds to both dimers. E323 in one monomer forms a salt bridge with R377 in the other monomer in the absence of $tRNA^{Thr}$ binding. Breaking of this salt bridge may affect binding in the second active site- possibly explaining why dimerisation is essential[102].

Following formation of the Thr-AMP intermediate, aminoacylation is catalysed by A76, the nucleotide at the 3' terminus of the tRNA. Electron transfer proceeds from the lone pair of a water molecule leading to deprotonation of His309, leading to proton transfer from the 2' hydroxyl of A76, which then deprotonates the 3' hydroxyl of A76 (stabilised by hydrogen bonding between 2' hydroxyl on the ribose of A76 and the phenolic hydroxyl of Y462), leading to nucleophilic attack of 3' hydroxyl and the carbonyl of ThrAMP and release of the AMP and aminoacylation[115], as illustrated in Figure 1.15.

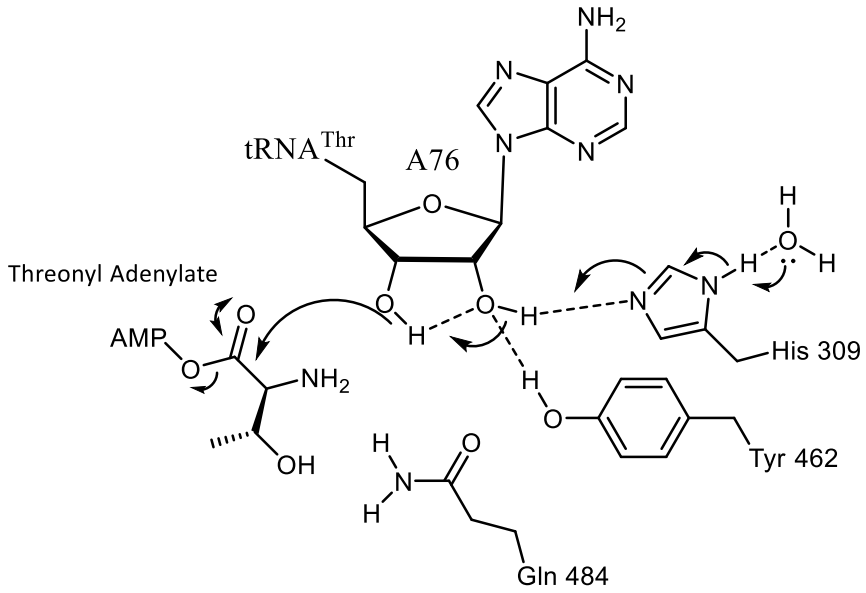


Figure 1.15. Proposed aminoacylation reaction mechanism. Adapted from Minajigi et al. 2008[115]

The N-terminal editing domain (see Figure 1.16) is essential for the proper function of EcThrRS; without it, serine is incorporated incorrectly into proteins, whereas with it serine is almost never incorporated[77]. In the editing domain, the β-hydroxyl group of the serine on mischarged $tRNA^{Thr}$ binds to the carbonyl and amino groups of G95, and the amine group is bound by the main chain carboxyl moiety of M181 and the side chain of H77[79]. Correctly charged $tRNA^{Thr}$ cannot bind to the editing site because of steric hinderance - the Thr methyl group would clash with H77, Y104 and D180 which are firmly constrained by a hydrogen bonding network[79]. The 2' hydroxyl on the ribose of A76 strongly binds to the sulfur of C182 which is pinned in place by hydrogen bonding with H186. Hydrolysis occurs via deprotonation of a water molecule by H73, leading to nucleophilic attack at the serine carboxyl group, leading to deprotonation of a second water, with stabilisation of intermediates by D180 and K156.

In summary, the catalytic mechanism of ThrRS is well understood, with many crystal structures of different stages of the reaction cycle being available. This means that it is an excellent target for drug discovery efforts- we can use these structures for *in silico* analysis as well as starting point to predict the effect of different interactions on the functionality of the protein.

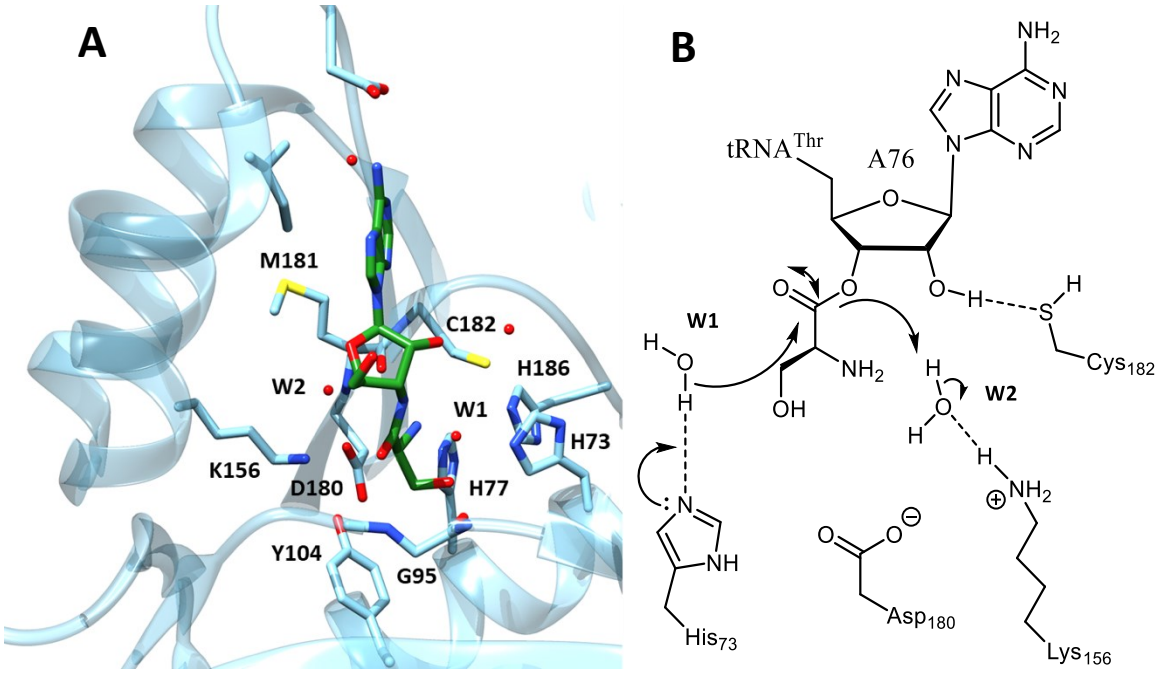


Figure 1.16. Mechanism of hydrolysis by the ThrRS editing domain. A) crystal structure showing binding of a 3' seryl adenosine analogue bound in the editing domain of ThrRS. EcThrRS, PDB file 1TKY[79]. 3' seryl adenosine analogue and important binding residues shown in stick representation with rest of protein shown in cartoon representation. 3' seryl adenosine analogue is coloured green with the protein coloured light blue. Residues with explicit interactions to the ligand are labelled with amino acid identity and numbering, as per EcThrRS. Figure generated in UCSF Chimera[114]. **B) proposed mechanism of hydrolysis by the editing domain**. Figure adapted from Dock-Bregeon *et al*. 2004[79].

1.3.2 Threonyl tRNA synthetases in eukaryotes

1.3.2.1 Human Threonyl tRNA synthetases

Humans possess three ThrRSs: TARS1, TARS2 and TARS3. TARS1 is cytoplasmic, encoded on chromosome 5, and contains, in addition to the normal editing, catalytic and anticodon domains, an N-terminal unique (UNE-T) domain and a TGS domain (ThrRS, GTPase and SpoT) which is structurally similar to ubiquitin (N1 in EcThrRS). TARS1 is secreted by vascular endothelial cells when exposed to pro-inflammatory cytokines such as tumour necrosis factor-α (TNF-α) and vascular endothelial growth factor (VEGF), with TARS1 itself having been demonstrated to induce angiogenesis and vascular endothelial cell migration[99].

Angiogenesis is the process by which new blood vessels are grown and is a key step in inflammation and carcinogenesis, in which it supports tumour growth[116]. The UNE-T domain has been demonstrated to be involved in VEGF translation via eukaryotic translation initiation factor 4E (eIF4E) binding[117]. Here the ThrRS anticodon binding domain targets the mRNA to be translated while eIF4E initiates translation. This eIF4E mediated translation initiation has been demonstrated to be capable of initiating translation of selected mRNAs via binding of the ThrRS anticodon binding domain to anticodon-like loop structures in 5' untranslated regions (5'UTRs)[113]. VEGF is involved in

the stimulation of angiogenesis. TARS1 also has a role in myogenic differentiation, negatively regulating myoblast differentiation via binding to Axin-1 which feeds into c-Jun N-terminal Kinase (JNK) signalling[118].

TARS2 is translated in the mitochondria. A non-canonical function of TARS2 is in sensing amino acid levels (specifically threonine levels) which it does in concert with the mechanistic target of rapamycin (mTOR). mTOR is a serine/threonine kinase, and forms the core of two distinct protein complexes, mTORC1 and mTORC2. mTORC1 acts as a central regulator of growth and proliferation, and is comprised of mTOR, Raptor (regulatory protein associated with mTOR) and mLST8 (mammalian lethal with Sec13 protein 8) as core components (with many, many other proteins involved) and reacts to multiple different signals[119].

One set of signals detected by mTORC1 are amino acids. For example, upon increase of threonine levels, mTORC1 is activated via association of threonine-bound TARS2 with Rag guanosine triphosphatases (Rags)[120], causing activation of mTORC1 on the lysosomal surface. mTORC1 activation leads to activation of protein, lipid and nucleotide synthesis and inhibition of autophagy[119], leading to cell growth and division. mTORC2 on the other hand is activated by growth factors to maintain survival and proliferation.

Autophagy is the process of controlled cell death in eukaryotes and is very tightly regulated. The presence of reactive oxygen species, lack of amino acids, and many other signals lead to autophagy as a way to maintain metabolism when nutrients are lacking and to remove damaged cells. The target cell is degraded in a controlled manner, leading to packaging and degradation of cell contents, leading to available nutrients for other cells in the body, and potentially to the removal of cells with damaged DNA[121]. Because of its role in the control of autophagy, mTORC1 has been found to be involved in cancer, aging, diabetes, neurodegenerative diseases, adipogenesis (formation of fat tissue) and muscle wasting diseases[119]. It has been suggested that inhibitors of mTORC1 could be used to treat insulin resistance[122], muscle atrophy[123], obesity[124], organ rejection following transplant[29], Alzheimer's disease[125], cancer[126] and modulation of aging[127]. This could all potentially be achieved by novel inhibitors of TARS2.

TARS3, is also known as ThrRS-like protein (ThrRS-L or TARSL2), is encoded on chromosome 15 and is found in both the cytoplasm and the nucleus (via a C-terminal nuclear localisation sequence). It has been suggested that TARS3 could be involved in aminoacylation of nuclear tRNAs and linked to their export from the nucleus[78,128,129]. Hundreds of mitochondrial tRNA-like genes are present in the nuclear genome including $tRNA^{Thr}$ copycats[130] and TARS3 could also have a role in aminoacylation of these molecules[78]. These tRNA copycats appear to be widely distributed in introns and appear to retain the canonical tRNA cloverleaf structure and have a potential role in splicing[131]. Additionally, vertebrate AlaRS is capable of mischarging $tRNA^{Thr}$, frequently without mistranslation occurring- it

has been suggested that the presence of two proof-reading capable ThrRSs in the cytoplasm are responsible for this protection from mistranslation[132].

The mammalian multiple tRNA synthetase complex (MSC) is a protein complex made up of a number of aaRSs, including GlnRS, ProRS, IleRS, LeuRS, MetRS, GluRS, LysRS, ArgRS and AspRS, along with other accessory proteins[133]. The function of the MSC appears to be to facilitate more efficient amino acid charging to tRNAs, and then delivery of charged tRNAs to the ribosome. The MSC is a source of intense study, with very little full complex structural information available, but it has been found that components of the MSC can dissociate in response to various stimuli, leading to non-canonical effects and a complex regulatory network[133]. An additional N-terminal domain of TARS3 has been found to be homologous to that of human ArgRS, containing two leucine zippers which are involved in recruitment to the MSC[134-136].

Along with the various *cis* acting factors involved in assembly of the complex (like the leucine zippers in TARS3 and human ArgRS)- there are *trans* factors also- including aaRS interacting multifunctional proteins (AIMPs). AIMP2 has been demonstrated to be involved in tumour suppression through its interaction with signalling pathways such as TGFβ, TNFα, Wnt and p53. But additionally members of the complex have been implicated in tumour suppression; LeuRS interacts with mTOR via RagD in the presence of Leucine (similar to the system observed with TARS2) and MetRS appears to have a role in DNA damage repair[133].

1.3.2.2 The role of threonyl tRNA synthetases in disease

A number of aminoacylation defects have been linked to disease in humans; the main ones which are linked to ThrRS are anti-synthetase syndrome, non-photosensitive trichothiodystrophy and cancer. Many of these are linked to non-canonical roles of ThrRS, including regulation of the immune system, which includes stimulation of endothelial cell migration, angiogenesis and the activation and maturation of dendritic cells[137].

Anti-synthetase syndrome is an autoimmune disease caused by the targeting of an aaRS by autoantibodies. This occurs most commonly for HisRS, ThrRS, AlaRS, GlyRS, IleRS, AsnRS, PheRS and TyrRS, while autoimmunity against LysRS, GlnRS, TrpRS and SerRS has been reported to be involved for other diseases[137]. Autoimmunity against ThrRS is caused by the PL-7 antibody, with clinical features including interstitial lung disease (scarring of lung tissue), fever, myositis (inflammation and degradation of muscle tissue), muscle weakness, arthritis, Gottron's sign (red rash over knuckles) and relapsing polychondritis (inflammation of cartilage)[138]. Of these, myositis is the most common presentation, but the disease is relatively rare and understudied, with small study sizes due to its rarity[137-140].

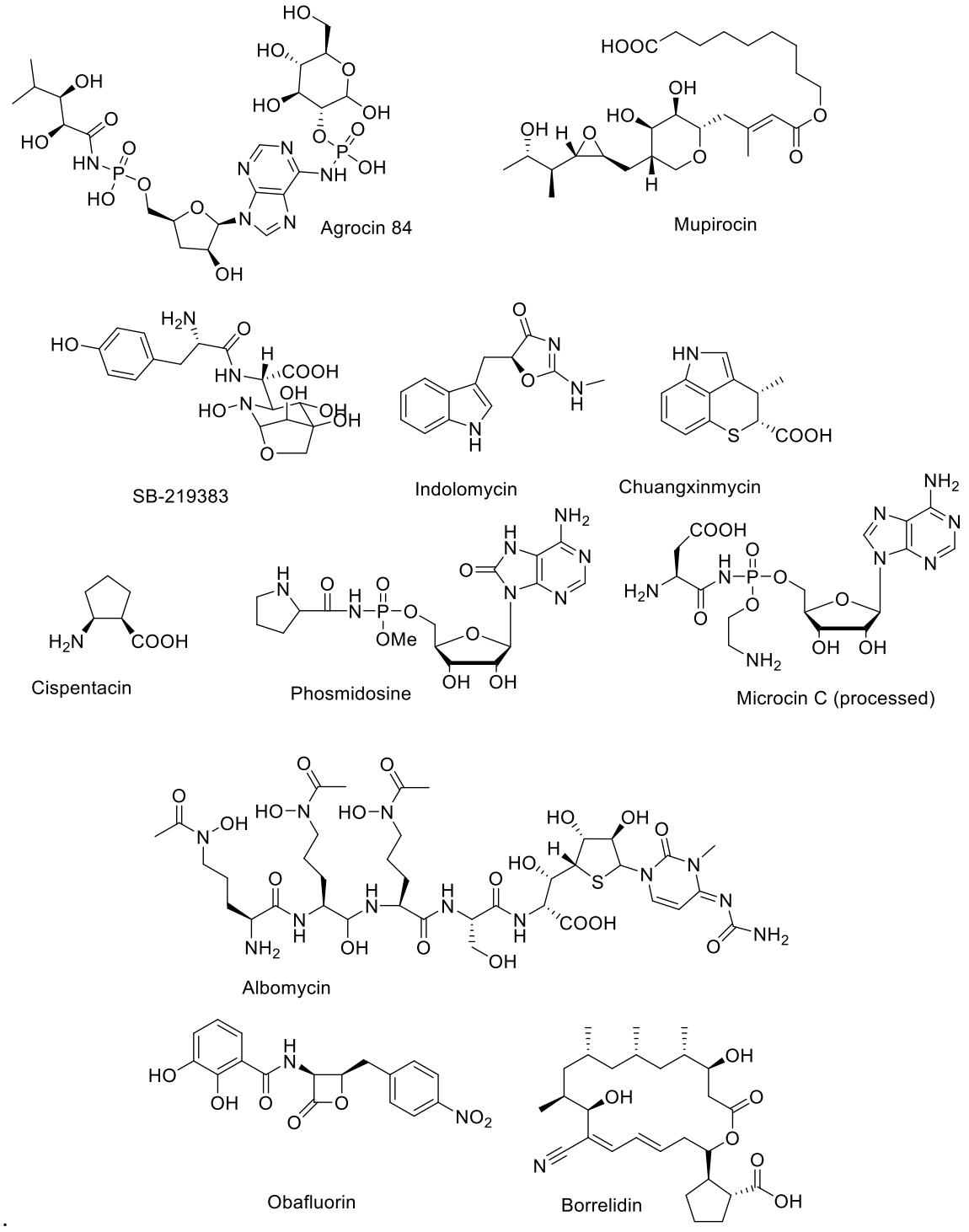
Mutations in human ThrRS have been linked with a form of non-photosensitive trichothiodystrophy (NPS-TTD), a rare connective tissue disorder with features including ichthyosis (persistent thick, dry skin), sparse and fragile hair, brittle nails, impaired intelligence, decreased fertility, short stature, anaemia and recurrent infections. The fragile hair will have a characteristic "tiger tail" pattern under polarised light due to a low cysteine content. NPS-TTD has been associated with mutations in a variety of different genes, with genetic variations in many individuals being uncharacterised to this point. In a group of individuals with non-characterised NPS-TTD, two individuals with ThrRS mutations were identified. One individual was heterozygous with TARS1 K276E (in the editing domain) in one allele and R638* (leading to almost complete truncation of the anticodon binding domain) in the other allele. The other individual was homozygous TARS1 L227P (in the editing domain)[141]. The mutations in both individuals led to decreased TARS protein levels and stability, which lead to haploinsufficiency and general reduced protein expression as well as the mischarging of tRNA molecules. Overall this caused the accumulation of misfolded proteins in the brain leading to neurological defects[141]. Non-photosensitive variants have been linked to AlaRS and MetRS also[142]. It is likely that other ThrRS mutations occur in humans, but too drastic a mutation would lead to lethality.

aaRSs activity has previously been shown to be highly upregulated in cancers and mutations in aaRSs have been identified in a number of cancers[143]. For example, a P42A mutation in GlyRS has been found in 40% of adenoid cystic carcinoma patients[144]. A correlation between ThrRS expression and advancement of ovarian cancer has also been observed[111]. Implication in cancer could be due to the increased supply of aminoacyl tRNAs for protein synthesis to support increased growth rates, but also due to the previously discussed non-canonical signalling roles of aaRSs. As previously discussed, TARS1 is secreted by endothelial cells and triggers angiogenesis in reaction to VEGF and TNFα exposure. TARS1 also has a role in VEGF upregulation; VEGF is thought to be the main angiogenesis promoter, as well as exhibiting immunosuppressive functions[145]. Inhibition of angiogenesis is a common strategy for the treatment of cancers; starved of nutrients and oxygen, cancer cells struggle to proliferate with angiogenesis blocked[145]. The importance of angiogenesis in cancer leads human ThrRS to be an attractive target for anti-cancer treatment, as long as the "normal" function of ThrRS is not too widely disrupted. This could be a good potential chemotherapy treatment, due to the inhibition of angiogenesis and immunosuppression caused by TARS1 regulation of VEGF as well as the inhibition of mTORC1 via TARS2 inhibition, potentially leading to the autophagy of cancer cells, if delivered in a targeted manner.

1.3.3 Aminoacyl tRNA synthetase inhibitors

A number of bacterially produced natural product inhibitors of aminoacyl tRNA synthetases have been identified, as detailed in Table 1.3 and Figure 1.17. Many of these molecules appear to mimic

the substrates of the reaction pathway or of the adenylated amino acid, and act as competitive inhibitors that bind in the active site of the protein. Three of these examples use a trojan horse mechanism in order to enter the cell. Of the inhibitors listed, one is commercially used as a biocontrol agent, one is clinically used, and one was previously used in the Soviet Union.



.

Figure 1.17. Structures of known bacterial natural product aminoacyl tRNA synthetase inhibitors.

Agrocin 84 (Figure 1.18) is produced by *Agrobacterium radiobacter* and acts as a non-hydrolysable LeuAMP analogue, with a 3'-deoxyarabinose in place of the ribose sugar and a D-glucopyranose-2-

phosphate attached to the N6 of the adenine ring. When pathogenic *Agrobacteria* (the causal agents of crown gall disease) infect plants, they introduce the genes required to produce agrocinopine A into the genome of the plant via their tumour inducing plasmids. Agrocinopine A is then produced by the plant and uptaken by the bacteria to use as a carbon source[146]. *A. radiobacter* strain K84 produces agrocin 84 to inhibit other *Agrobacteria* where it is uptaken as if it is agrocinopine A, due to the glucopyranose moiety which mimics that of Agrocinopine A (see Figure 1.18). Once it has entered the cell, the glucopyranose moiety is removed by a target phosphodiesterase before binding in a tRNA-dependent manner to the LeuRS, inhibiting it and killing the target[147]. Along with its self-resistant LeuRS, AgnB2, the BGC for Agrocin 84 contains a truncated ArgRS, AgnA which is thought to catalyse phosphoramidate bond formation. This protein is lacking an anti-codon binding domain but retains its catalytic domain[147]. The specificity of the trojan horse mechanism limits the scope of this compound to the treatment of crown gall disease and has led to this strain of *Agrobacterium* being used commercially for biocontrol[148].

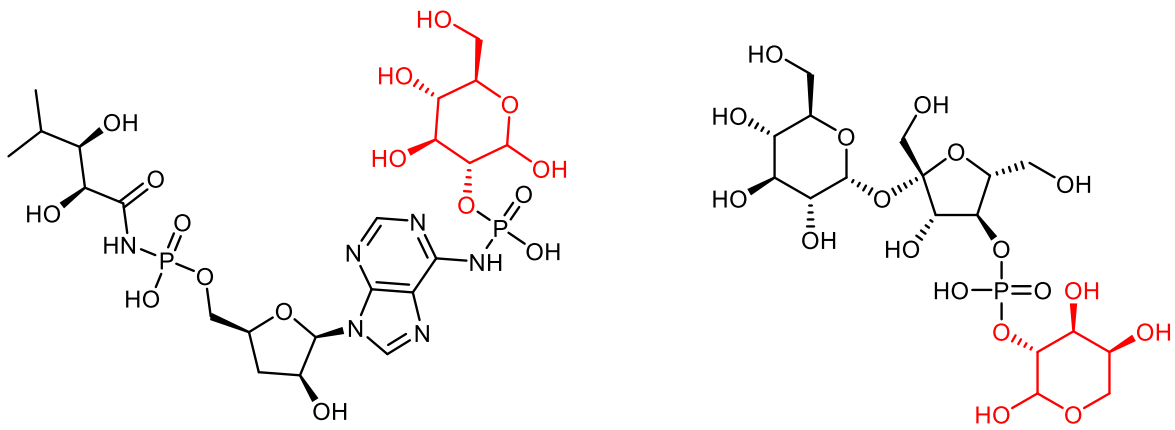


Figure 1.18. Structures of Agrocin 84 and Agrocinopine A. Glucopyranose moieties highlighted in red.

Albomycin (see Figure 1.17) is a seryl-adenylate analogue with a siderophore moiety attached. It can bind ferric iron and then enter the target cell via siderophore uptake proteins. Once inside the cell the Fe^{3+} will be reduced to Fe^{2+}, released and the siderophore moiety will be removed by an endogenous peptidase which liberates the "warhead" part of the molecule which is then free to inhibit its target[149]. There are unlikely to be side effects in humans because they lack the bacterial siderophore uptake machinery, and instead rely on metal binding proteins for uptake of iron. This means that albomycin does not penetrate human cells easily, which explains the high specificity of this trojan horse drug for bacteria. Indeed, no toxicity was observed in mouse models, and clinical trials in the Soviet Union were successful, with the compound being used for the treatment of bacterial infections[148]. It has been observed that albomycin effectively inhibits *E. coli, Staphylococcus aureus* and *Streptococcus pneumoniae*[149-151].

Possibly the most successful aaRS inhibitor to be used in the clinic is mupirocin (see Figure 1.17), which is on the World Health Organisation's list of essential medicines[152]. It is produced by a strain of *Pseudomonas fluorescens* and was discovered in 1971. Mupirocin is marketed as Bactroban® and is used topically and nasally against bacterial infection, especially methicillin resistant *Staphylococcus aureus* (MRSA). Mupirocin is an inhibitor of bacterial IleRS, and the producer *P. fluorescens* protects itself using a second resistant copy its housekeeping IleRS. Resistance to mupirocin in pathogenic bacteria has been shown to arise due to the acquisition of the *mupM* resistance gene on a plasmid[31]. Gram negative organisms, anaerobes and fungi are not susceptible to mupirocin, and healthy skin bacterial flora such as *Micrococcus, Corynebacteria* and *Propionibacteria* are not disrupted by its use. Systemic use is not possible due to rapid metabolism, limiting the scope of use to topical creams and nasal sprays.

1.3.3.1 Borrelidin

Borrelidin (see Figure 1.17) was discovered in 1949 and is produced by *Streptomyces rochei*. It was identified due to its anti-*Borrelia* activity, which is the origin for its name[153] (*Borrelia burgdorferi* being the spirochete bacteria responsible for Lyme disease). Soon afterwards, its anti-fungal[154,155], anti-viral[156] and general anti-bacterial activities were noted. Subsequently, borrelidin was show to possess potentially useful activity as an anti-malarial[157,158], a phytotoxin produced by a plant pathogenic strain of *Streptomyces*[159] and identified as an inhibitor of angiogenesis in mammals[160]. The discovery of anti-angiogenesis activity led to interest in studying borrelidin as a potential anti-cancer molecule; however, due to its generalised toxicity, it could not be developed for use in the clinic. Identification of its anti-malarial properties led to a semi-synthetic derivative (BC-194) being developed which could no longer bind to the human ThrRS (HsThrRS) but instead to the *Plasmodium falciparum* ThrRS alone- this has led to interest in its potential for development as an anti-malarial agent with a novel mechanism of action[157,158]. Due to its various biological activities, borrelidin has been the source of intense interest; it was the first known ThrRS inhibitor, and one of the most well characterised aaRS inhibitors in general. Borrelidin has found to be able to inhibit the growth all cell types, except some genera of *Archaea*[161] and so the natural product molecule represents a highly advantageous scaffold for structural modification to generate more specific congeners with specificity towards target organisms, over normal human cells.

Table 1.3. Details about known bacterial natural product inhibitors of aminoacyl tRNA synthetases. Type of aaRS targetted, producing strain, mode of action, self-resistance mechanism and status in approval for clinical trials is listed.

Natural Product	aaRS Target	Natural Product Type	Producer	Mode of Action	Self-Resistance Mechanism	Clinical Trial/Biocontrol Status
Agrocin 84	LeuRS	Nucleoside	*Agrobacterium radiobacter* strain K84	Trojan Horse, LeuAMP analogue, tRNA dependent	Additional resistant LeuRS	Producer is a commercially used biocontrol agent against Crown Gall Disease[147].
Mupirocin	IleRS	Polyketide	*Pseudomonas fluorescens* NCIMB 10586	IleAMP analogue	Additional, resistant IleRS	Approved for topical use for skin infections, non-bullous impetigo and nasal use for staphylococci, including MRSA[152].
SB-219383	TyrRS	Dipeptide	*Micromonospora* NCIMB 40684	Tyrosine binding site	Unknown BGC	Inhibition of *Pseudomonas aeruginosa* identified, no further work[162].
Indolmycin	TrpRS	Oxazole	*Streptomyces griseus*	Tryptophan binding site	Additional, resistant TrpRS	Identification as a potential antimycobacterial and inhibitor of MRSA, cleared *Helicobacter pylori* from Mongolian gerbils at 10mg/kg, no clinical trials[163].
Chuangxinmycin	TrpRS	Indole	*Actinoplanes tsinanensis* CPCC 200056	Tryptophan binding site	Additional, resistant TrpRS	Early clinical trials showed effectiveness in treating septicaemia, urinary and biliary infections, against *E. coli* and effectiveness in mice against *E. coli* and *Shigella dysenteriae*[164].

Cispentacin	ProRS	Amino Acid	*Bacillus cereus*	Proline binding site	Unknown	Demonstrated effective against *Candida* infection in mice[165]. No clinical trials.
Phosmidosine	ProRS	Nucleoside	*Streptomyces durhameusis*	ProAMP analogue	Unknown	Anti-cancer activity specific to tumour cells identified[166].
Microcin C	AspRS	RiPP	*Enterobacteriaceae*	Trojan Horse, AspAMP analogue	Acetyltransferase	In broiler chicks, improvement of growth performance, strengthened immune functions, enhanced intestinal barrier and regulated caecal microbiota- antibiotic alternative[167].
Albomycin	SerRS	NRPS	*Streptomyces griseus*	Trojan Horse, SerAMP analogue	Additional resistant SerRS	Effective against *Streptococcus pneumoniae* and *E. coli.* No toxicity in mice. Used in the Soviet Union for treatment of bacterial infections[150,151].
Obafluorin	ThrRS	NRPS	*Pseudomonas fluorescens* ATCC 39502	Unknown	Additional resistant ThrRS	Unknown mode of action, limited bioactivity assays.
Borrelidin	ThrRS	Polyketide	*Streptomyces parvulus* Tü4055	Allosteric site at bottom of catalytic pocket	Additional resistant ThrRS	Limited development due to toxicity, derivatives further explored[157].

To expand the potential usefulness of borrelidin even further, borrelidin was recently found to dissociate amyloid-β and tau fibril aggregation- the two proteins associated with the advancement of Alzheimer's disease[168]. Aggregation of these proteins is associated with neuronal cell death and so dissociation of these aggregates is an identified treatment for Alzheimer's disease. Two molecules, Nilotinib and Posiphen are in clinical trials as drug candidates to regulate amyloid-β and tau, but act indirectly, while borrelidin appears to directly prevent aggregation and dissociate aggregates by binding to hydrophobic pockets on amyloid-β[168]. As with all of the other potential uses of borrelidin, less toxic, more specific derivatives would have to be developed to facilitate development of a clinically useful drug candidate.

Recently, it was found that at sub growth-inhibitory concentrations, borrelidin can increase life span in *Saccharomyces cerevisiae* and *Caenorhabditis elegans* by upregulation of *GCN4* in yeast/ *atf-4* in *C. elegans*. In mammals, this is *ATF4*. Gcn4 translation is upregulated by the accumulation of uncharged tRNAs, and increased levels of ATF4 in mice lead to increased lifespan[169]. Borrelidin inhibits ThrRS in these organisms, leading to an increase in the pool of uncharged tRNAs and thus the observed life span extension. This once again shows that borrelidin is a drug with many potential uses, if a non-toxic derivative can be produced.

Following the identification of the biosynthetic gene cluster encoding its production in *S. rochei*, *borO* was swiftly identified as a ThrRS and an self-resistance gene via heterologous expression in *Streptomyces albus*. However, the exact mechanism for the resistance of BorO could only be guessed at on the basis of sequence alignments, with no crystal structures having been solved to date[110,170]. Borrelidin has been co-crystallised with TARS1 and EcThrRS with binding in the aminoacylation active site, simultaneously blocking binding of the amino acid, ATP and tRNA 2' hydroxy[76], as well as a fourth, non-catalytic subsite within the active site. This full-spectrum inhibition, as opposed to being just an analogue of one of the substrates of the reaction, is unusual for aaRS inhibitors. As observed in Table 1.3, other known inhibitors of aaRSs are either analogues of their aa-AMP intermediate or bind in the amino acid binding site and don't generally rely on this allosteric binding mode. An illustration of borrelidin binding, illustrating how it blocks threonine, ATP and tRNA binding, while binding in its own binding site, can be seen in Figure 1.19.

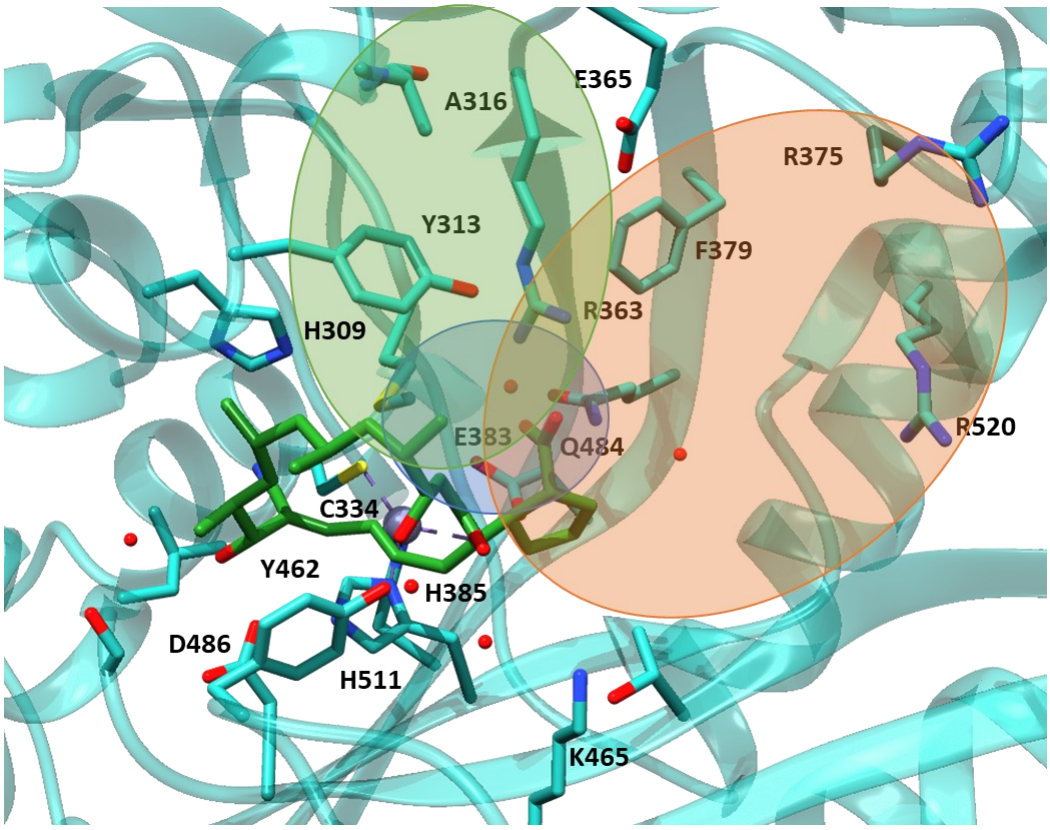


Figure 1.19. Crystal structures showing binding of the tRNA anticodon to the anticodon binding domain of ThrRS. EcThrRS, PDB file 4P3P. Borrelidin and important binding residues shown in stick representation with rest of protein shown in cartoon representation. Borrelidin is coloured green with the protein coloured light blue. Residues with explicit interactions to threonine, zinc, ATP or tRNA are labelled with amino acid identity and numbering, as per EcThrRS. Blue circle represents the approximate zone of threonine binding, orange circle shows the approximate zone of ATP binding and green circle shows the approximate zone of tRNA binding. Figure generated in UCSF Chimera[114].

1.3.3.2 Obafluorin

The second ThrRS targeting natural product to be discovered was obafluorin. Obafluorin was identified in 1984[171] and its produced by *Pseudomonas fluorescens* ATCC 39502, a plant-derived strain believed to be isolated from the Obal Garden Centre in Princeton, New Jersey in the USA. Obafluorin contains an unusual β-lactone group, along with catechol and a nitrobenzyl moieties, and the structure of obafluorin can be seen in Figure 1.17. Obafluorin is a potent inhibitor of threonine tRNA synthetases and exhibits broad spectrum antibacterial activity. Its biosynthesis (see Figure 1.20) involves the assembly of two chorismate-derived units: a catechol (2,3-dihydroxybenzoic acid, 2,3-DHBA), and the unusual (2S,3R)-2-amino-3-hydroxy-4-(4-nitrophenyl)butanoate (AHNB) which is produced through the action of a L-threonine transaldolase (L-TTA) from threonine and 4-nitrophenylacetaldehyde (see Figure 1.20). These two units are assembled by the NRPS ObaI, with assembly of the β-lactone moiety catalysed by an unusual TE domain in the offloading reaction[172,173]. The self-resistance gene *obaO* is found in the obafluorin

BGC in *P. fluorescens* and encodes a second threonyl tRNA synthetase in the producer genome that is resistant to obafluorin [170,174].

When aminoacylation assays were carried out, it was found that obafluorin inhibited EcThrRS with an IC_{50} (concentration at which 50% of the activity had been abolished) of 92 nM and while ObaO showed inhibition with increasing concentration of obafluorin, with an IC_{50} of 50nM the activity never dropped below 35% of the inhibited activity (see Figure 1.21)[174]. This unusual mechanism of resistance be partial inhibition warranted further study.

In the Wilkinson group, work is ongoing to understand the mode of action of obafluorin and the mechanism by which the self-resistance protein, ObaO confers resistance. Complementation experiments and *in vitro* activity assays have demonstrated that ObaO is essential for obafluorin self-resistance in *P. fluorescens* ATCC 39502[174]. *In vitro, o*bafluorin displays a highly unusual partial inhibition of ObaO, meaning that whilst obafluorin can fully inhibit EcThrRS, it does not fully inhibit aminoacylation by ObaO even at the highest concentrations tested. Studying obafluorin *in vitro* presents a challenge due to the chemical instability of the β-lactone moiety. This was first noted in the 1980s where it was noted that basic solution and MeOH were sufficient to ring open with a half-life of minutes[175].

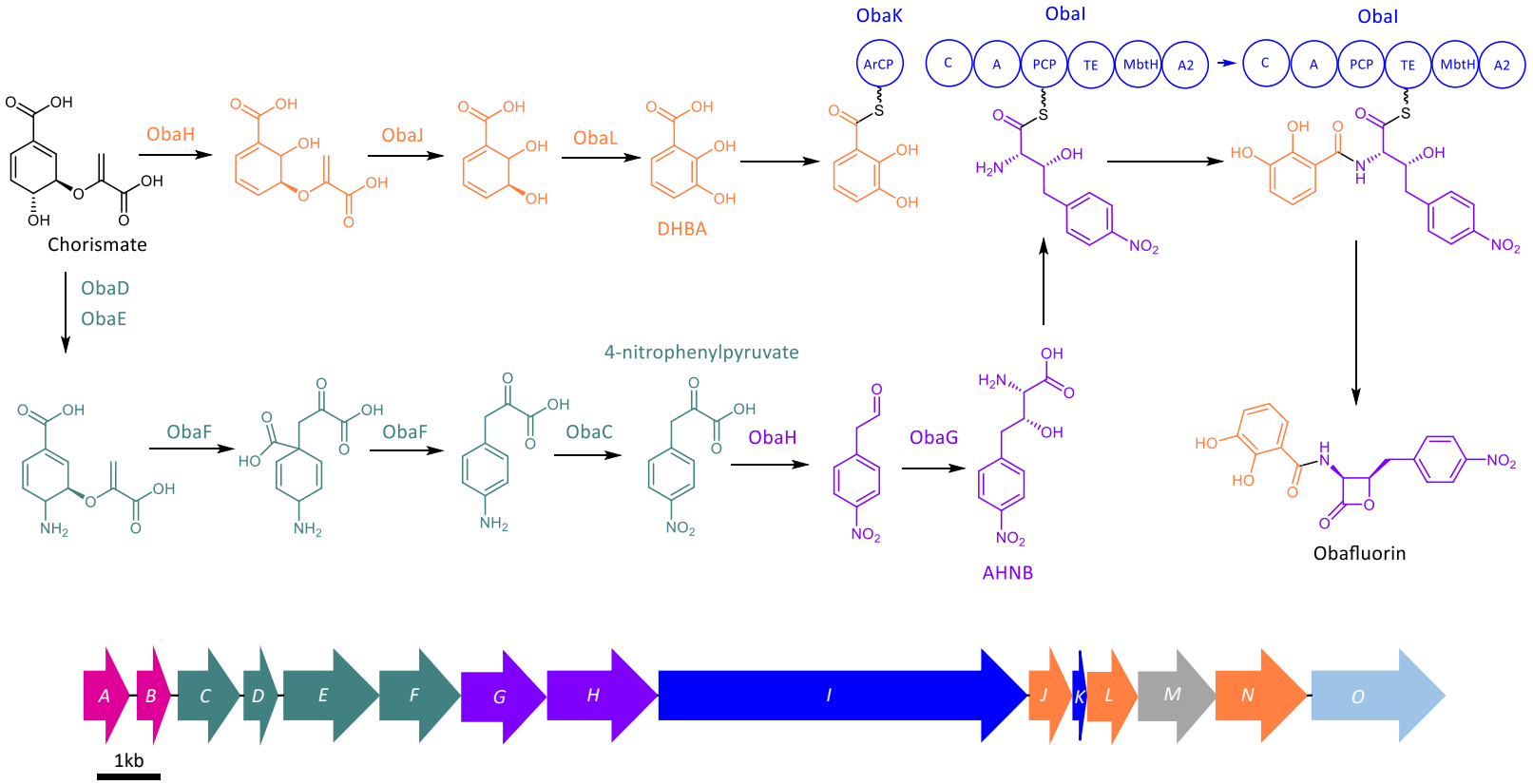


Figure 1.20. Proposed biosynthetic pathway and biosynthetic gene cluster for obafluorin. Starting from chorismate, which is generated by the primary metabolic shikimate pathway, the teal pathway, encoded by the teal genes, generates 4-nitrophenylpyruvate which is then converted to (2S,3R)-2-amino-3-hydroxy-4-(4-nitrophenyl)butanoate (AHNB) by the purple genes. The orange pathway shows the generation of 2,3-dihyrdroxybenzoic acid (2,3-DHBA), with the genes for this in orange. The dark blue NRPS proteins then work together to combine these building blocks into obafluorin. ArCP = aryl carrier protein, C = condensation domain, A = adenylation domain, PCP = peptide carrier protein, TE = thioesterase domain, MbtH = MbtH-like protein, essential for A domain activity. A2 = 2,3-DHBA adenylation domain. *obaA* and *B* in pink are regulators of the cluster, *obaM* in grey is involved in chorismate availability and obaO in light blue is the self-resistance gene. Figure adapted from Scott et al. 2017[33].

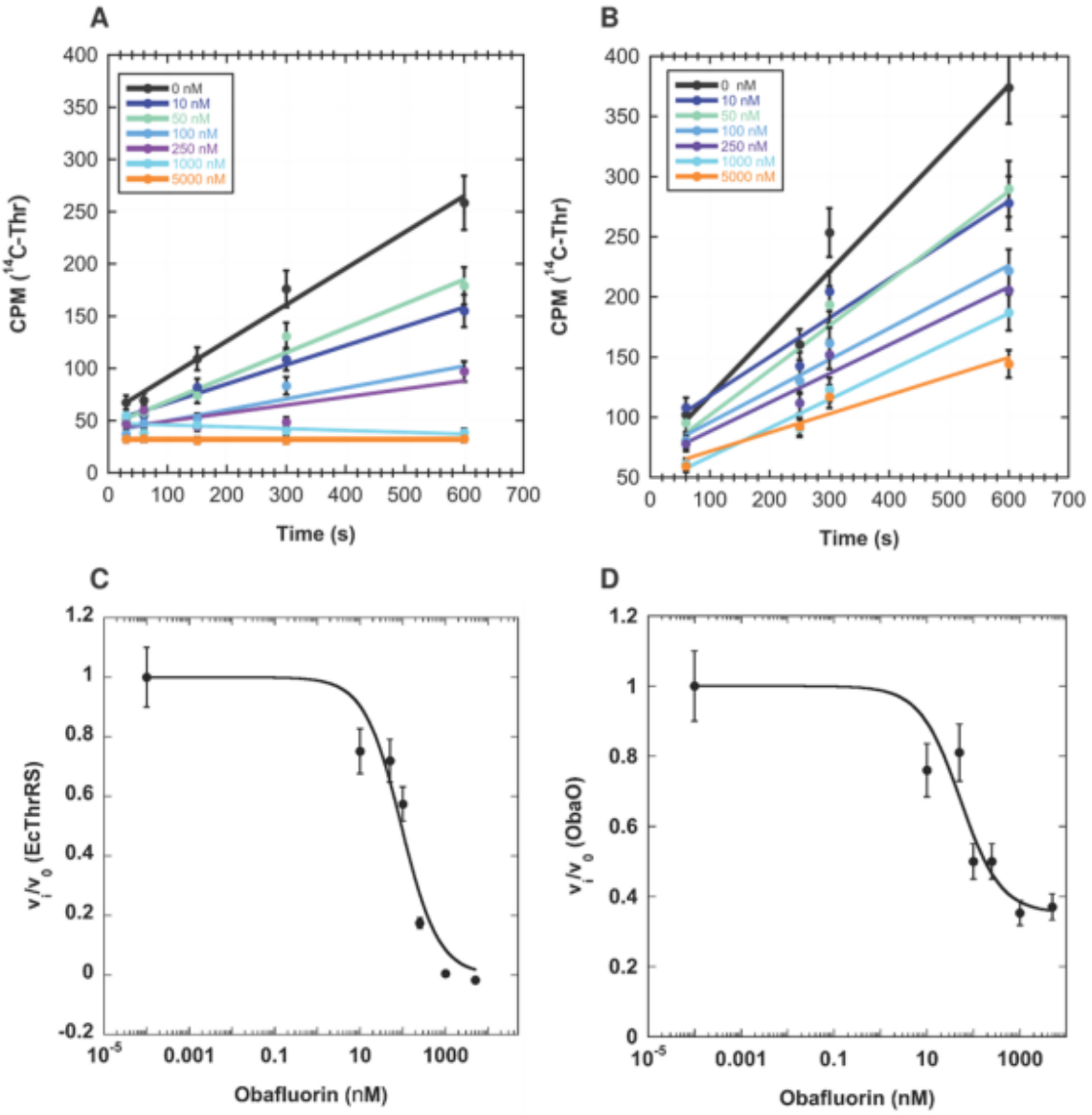


Figure 1.21. Complete and partial inhibition of EcThrRS and ObaO by obafluorin. Progress curves for **A)**, EcThrRS and **B)**, ObaO in the presence of 0–5000 nM obafluorin. Reactions (n = 3) included enzyme at 10 nM and saturating concentration of tRNA, threonine, and ATP. The progress curves were fit to a linear equation to derive initial rates. Error bars represent the standard error for each time point. Dose response curves for **C)**, EcThrRS and **D)**, ObaO were calculated from the initial rates by plotting the fractional velocity at each of seven different inhibitor concentrations against log [obafluorin]. CPM = counts per min. Taken from Scott et al. 2019[174].

1.3.3.3 Synthetic ThrRS inhibitors

The design of synthetic inhibitors based on the structures of the active site of protein targets has also been done for ThrRSs, with several synthetic inhibitor-bound ThrRS structures available on the PDB[176-179]; the structures of these compounds can be seen in Figure 1.22. These inhibitors are all designed to be homologues of ThrAMP or designed based on fragment hopping- where smaller

fragments are screened for binding, and then these fragments are covalently joined to form a single multi-site binding compound. While this study focusses on natural inhibitors, it is important to highlight this strategy. These synthetic inhibitors echo the inhibitor design developed by nature. All natural product aaRS inhibitors that have been discovered, except for borrelidin bind in the amino acid binding site, or are amino acyl-AMP analogues.

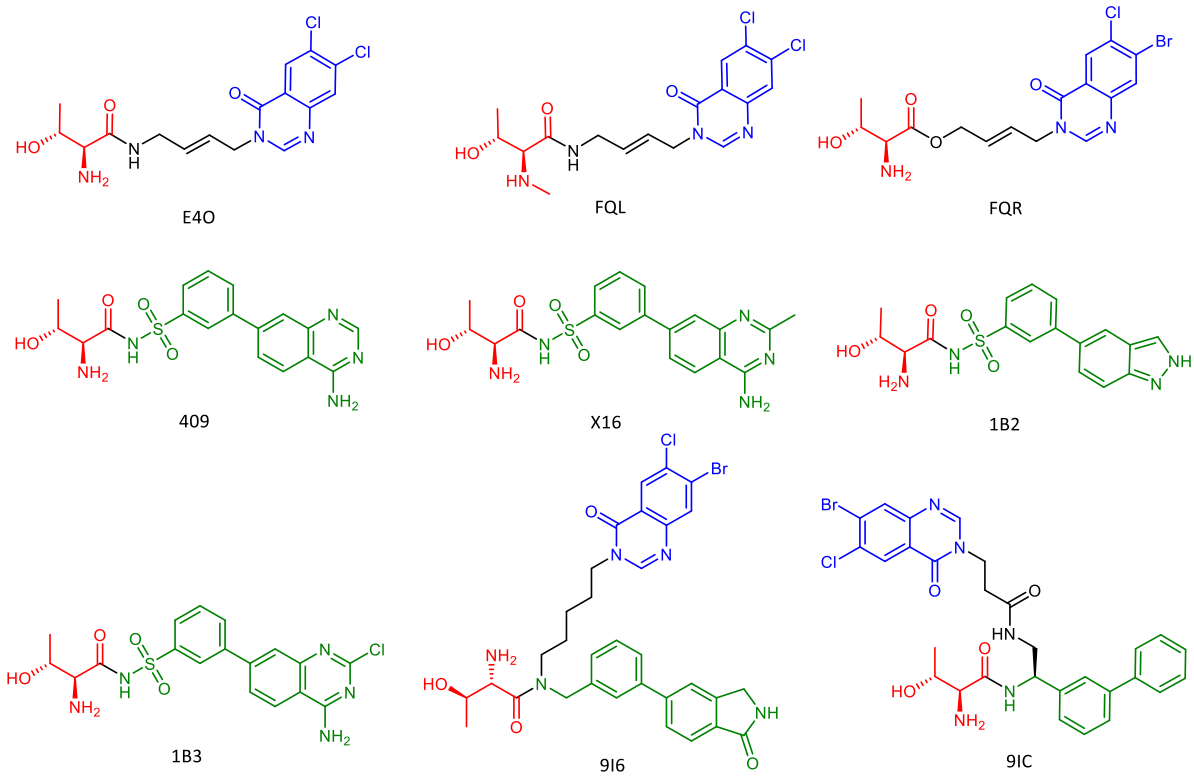


Figure 1.22. Structures of the synthetic ThrRS inhibitors with their structures bound to ThrRS uploaded to the PDB. The parts of the molecule binding in the threonine binding site is coloured in red, the parts of the molecule binding ATP binding site in green, and the parts of the molecule binding the tRNA binding site in blue. Figure generated in ChemDraw.

1.4 Project Aims

The aims of this project are three-fold:

1) To understand the mechanism of self-resistance in the borrelidin producer organism *Streptomyces parvulus*.
    a. To sequence the genome of *S. parvulus* in order to identify the tRNA and housekeeping ThrRS genes (*SpThrRS*).
    b. To purify BorO, SpThrRS and EcThrRS and solve their structures in the presence and absence of borrelidin
    c. To confirm that BorO does confer borrelidin resistance via knockout in the producer and heterologous expression and bioassay.
    d. To identify the mode of resistance to borrelidin, biophysically and structurally.
    e. To mutate EcThrRS to be resistant to borrelidin using structural and biophysical data as a guide.
2) To understand the mechanism of action of obafluorin against sensitive enzymes/organisms, and the self-resistance mechanism in the producer *Pseudomonas fluorescens*.
    a. To purify ObaO and EcThrRS and solve their structures in the presence and absence of obafluorin.
    b. To work out the residue(s) to which obafluorin binds, therefore elucidating a mechanism of action.
    c. To then elucidate the mechanism of resistance in ObaO using structural data and mutagenesis in ObaO, PfThrRS and EcThrRS.
3) To use the information we gained from the objectives 1 & 2 to expand the repertoire of natural product ThrRS inhibitors via genome mining.
    a. Search for homologues of BorO and examine their genomic context.
    b. Probe potential evolutionary history of BorO; can this inform us as to other ThrRS inhibitors?
    c. Search genomes for extra genomic copies of ThrRSs that are located in BGCs.

Chapter 2: Materials and Methods

2.1 General Materials

Chemicals and reagents were purchased from Merck Life Sciences, New England BioLabs (NEB), Fischer Scientific and Alfa-Aesar. Reagent or HPLC grade Solvents were purchased from Fisher Scientific. Borrelidin was purchased from BioAustralis (Smithfield, Australia) and Apollo Scientific (Bradbury, UK).

2.2 Buffers and Media

2.2.1 Culture Media

For all culture media, unless otherwise stated, deionised water was used to make up to the desired volumes. Unless otherwise stated, HCl/NaOH was used to adjust the pH as needed.

Lysogeny Broth (LB)

Difco™ Bacto Tryptone	10 g/L
Difco™ Yeast Extract	5 g/L
NaCl	5 g/L
Glucose	1 g/L
Agar	±20 g/L

Terrific Broth (TB)

Difco™ Yeast Extract	24 g/L
Difco™ Tryptone	12 g/L
Glycerol	4 mL/L

Suppressor of Catabolism (SOC) Medium

Tryptone	20 g/L
Yeast Extract	5 g/L
NaCl	0.58 g/L
KCl	0.19 g/L
$MgCl_2$	2 g/L
$MgSO_4$	2.5 g/L

Soft Nutrient Agar (SNA)

Difco™ Nutrient Broth	8 g/L

Agar	7 g/L

Difco Nutrient Agar (DNA)

Difco™ Nutrient Broth	4 g/L
Agar	10 g/L

-NaCl Lennox Broth (L)

Difco™ Yeast Extract	5 g/L
Difco™ Tryptone	10 g/L
Agar	±20 g/L

2x Yeast Tryptone Medium (2xYT)

Difco™ Yeast Extract	10 g/L
Difco™ Tryptone	16 g/L
NaCl	5 g/L

Maltose Yeast Extract Medium (MYM)

Maltose	4 g/L
Yeast Extract (Oxoid)	4 g/L
Malt Extract	10 g/L
Tap Water	500 mL/L
Trace Element Solution	1:500 dilution
pH	7.2

MAM Medium

Wheat starch	10 g/L
Spray dried corn steep liquor powder	2.5 g/L
Yeast Extract	3 g/L
$CaCO_3$	3 g/L
$FeSO_4 \times 7H_2O$	0.3 g/L
Agar	±20 g/L

Tap water	1 L
pH	7.3

Soya Flour Mannitol Medium (SFM)

Mannitol	20 g/L
Soya Flour	20 g/L
Agar	±20 g/L
Tap water	1 L

Glucose Yeast Extract Medium (GYM)

Glucose	4 g/L
Yeast Extract	4 g/L
Malt Extract	10 g/L
$CaCO_3$	2 g/L
Agar	12 g/L
pH	7.2

Bennett's Medium

Beef Extract	1 g/L
Glucose	10 g/L
N-Z amine A	2 g/L
Yeast Extract	1 g/L
Agar	±15 g/L
pH	7.3

Tryptone Soy Broth (TSB)

Difco™ Tryptone	17 g/L
Soy Peptone	3 g/L
NaCl	5 g/L
K_2HPO_4	2.5 g/L

Glucose	2.5 g/L

Obafluorin Production Medium (OPM)

Difco™ Yeast Extract	5 g/L
Glucose	5 g/L
$FeSO_4$	0.1 g/L
$MgSO_4$	0.1 g/L

NGY Seed Medium

Peptone	1 g/L
Yeast Extract	1 g/L
Glucose	4 g/L
pH	7.0

Peptone Yeast Dextrin Glucose Medium (PYDG)

Peptonised Milk Nutrient	15 g/L
Yeast Autolysate	1.5 g/L
Dextrin	45 g/L
Glycose	5 g/L
pH	7.0

2.2.2 Buffers

All buffers were made up with deionised water. Protein purification buffers were all filtered with Nalgene™ PES bottle top 0.22nm filters. All buffers were adjusted with HCl/NaOH as needed.

Full Length ThrRS Lysis Buffer

Tris-HCl	20 mM
NaCl	500 mM
$MgCl_2$	10 mM
pH	8.0

Full Length ThrRS Elution Buffer

Tris-HCl	20 mM

NaCl	500 mM
$MgCl_2$	10 mM
Imidazole	250 mM
pH	8.0

ΔN ThrRS Lysis Buffer

Tris-HCl	20 mM
NaCl	500 mM
$MgCl_2$	10 mM
Threonine	10 mM
pH	8.0

ΔN ThrRS Elution Buffer

Tris-HCl	20 mM
NaCl	500 mM
$MgCl_2$	10 mM
Threonine	10 mM
Imidazole	250 mM
pH	8.0

Transfer Buffer

Tris-HCl	25 mM
Glycine	192 mM
SDS	0.1% (v/v)
Methanol	20% (v/v)

Tris-Buffered Saline + Tween20 (TBST)

Tris-HCl	20 mM
NaCl	150 mM
Tween 20	0.1% (v/v)

pH	7.5

Developing Solution A

Tris-HCl	100 mM
Luminol	1.25 mM
Coumaric Acid	202.5 µM
pH	8

Developing Solution B

Tris-HCl	100 mM
H_2O_2	0.018% (v/v)
pH	8

Cryo-EM Buffer

Tris-HCl	20 mM
NaCl	100 mM
$MgCl_2$	10 mM
pH	8.0

ITC Buffer

Full Length ThrRS Lysis Buffer	95% (v/v)
DMSO	5% (v/v)
pH	8.0

2.3 Cultivation of Strains

E. coli was grown in liquid or solid LB, except for protein expression, where it was grown in Terrific Broth. *Staphylococcus aureus* strains were cultivated in TSB. *Streptomyces* strains were cultivated on solid MAM, SFM for *S. coelicolor*, Bennett's for *S. lividans* and MYM for *S. venezuelae*. Antibiotic selection was performed with: Kanamycin (50 µg/mL), Hygromycin (50 µg/mL), Apramycin (50 µg/mL), Tetracycline (25 µg/mL), Carbenicillin (100 µg/mL), Thiostrepton (50 µg/mL), Nitrofurantoin (100 µg/mL) and Nalidixic Acid (25 µg/mL), used as appropriate. When using hygromycin for *E. coli*, liquid L media and solid DNA media were used. *Streptomyces parvulus* ISOM-0991 and *S. parvulus* ISOM-1302 were obtained from Isomerase Therapeutics (Cambridge, UK). For all strains, spores or liquid cultures were stored at -80°C with 20% glycerol for cryopreservation.

Streptomyces spp. spores were harvested using standard methods[180]. Unless otherwise stated, overnight cultures were prepared by inoculation of 10mL LB with a single colony from a plate with the appropriate selection and incubated at 37°C 200rpm shaking overnight.

2.4 Generation of Constructs and Strains

Whenever plasmid DNA was required, it was prepared by use of the Qiagen miniprep kit as per the product manual from a 10 mL overnight culture, eluting with 30 µL of elution buffer (EB).

2.4.1 Polymerase Chain Reaction (PCR)

PCR primers were purchased from Integrated DNA technologies (IDT). Plasmid DNA was diluted to the appropriate concentration (1-10 ng/µL) using EB from the Qiagen miniprep kit. *Streptomyces* Genomic DNA (gDNA) templates were prepared using the FastDNA™ Spin Kit for Soil (MP Biomedicals) using an Omni Bead Ruptor 24 (Omni International) with Lysing Matrix E (MP Biomedicals) for 40 seconds at the 6 m/s speed setting for the lysis step, following the product manual for all other steps. *Pseudomonas* gDNA templates were prepared using the GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich) as per the product manual. PCR reactions were performed using an Applied Biosystems SimpliAmp Thermal Cycler (Thermo Fisher Scientific).

2.4.1.1 Q5 PCR

Where possible, for cloning purposes, Q5 polymerase (NEB) was used following the general method: reaction mix total volume 50 µL comprising: 10 µL Q5 Reaction Buffer, 10 µL High G+C Enhancer (when amplifying *Streptomyces* or *Pseudomonas* genes, replacing with nuclease-free water for amplification of *E. coli* codon optimised genes), 2.5 µL of each 10 µM primer, 2 µL of 1-10 ng/µL template DNA, 2.5 µL DMSO, 1 µL 10 mM dNTPs, 19 µL nuclease-free water, 0.5 µL Q5 polymerase. The following programme was generally used:

Step	**Temperature (°C)**	**Time (s)**	**Cycles**
Denaturation	98	60	1
Denaturation	98	30	32
Annealing	52	10	
Extension	72	30 per kb	
Extension	72	300	1

Where the general method gave unspecific amplification, a touchdown method was used, with the same reaction mix:

Step	**Temperature (°C)**	**Time (s)**	**Cycles**
Denaturation	98	60	1

Denaturation	98	30	10
Annealing	62-52 (-1 per cycle)	10	
Extension	72	30 per kb	
Denaturation	98	30	22
Annealing	52	10	
Extension	72	30 per kb	
Extension	72	300	1

2.4.1.2 Phusion PCR

For amplification of *borO* flanking regions for the *borO* knockout vector, Phusion polymerase (NEB) was used, with a reaction mix of: total volume 50 µL made up of 10 µL Phusion HF Buffer, 1 µL 10 mM dNTPs, 2.5 µL of each 10 µM primer, 7 µL of 8 ng/µL *S. parvulus* gDNA, 15 µL nuclease-free water, 10 µL 1 M betaine, 2.5 µL DMSO, 0.5 µL Phusion polymerase. The following programme was used:

Step	**Temperature (°C)**	**Time (s)**	**Cycles**
Denaturation	98	60	1
Denaturation	98	30	32
Annealing	52	10	
Extension	72	30 per kb	
Extension	72	300	1

2.4.1.3 Colony PCR

For colony PCR, the more error prone Taq polymerase was used. Green GoTaq master mix was used, with a reaction mix of total volume 25 µL made up of 12.5 µL Green GoTaq Master Mix, 1 µL each 10 µM primer, 1 µL template, 9.5 µL water. Sequencing primers used for colony PCR were designed to anneal 100 bp upstream of the insert site for forward primers and 100 bp downstream of the insert site for reverse primers. Templates were prepared depending on the strain. Single *E. coli* colonies were picked into 40 µL of sterile water and heated at 65 °C for 25 mins; an aliquot of this used directly as the template DNA in the PCR mix. *Streptomyces* colonies were picked into 50 µL of 50:50 DMSO:water and heated at 75°C for 25 mins; an aliquot of this used directly as the template DNA in the PCR mix. *Pseudomonas* colonies were picked into 50 µL of 50:50 DMSO:water and heated at 75 °C for 25 mins an aliquot of this used directly as the template DNA in the PCR mix. The following PCR programme was used:

Step	**Temperature (°C)**	**Time (s)**	**Cycles**

Denaturation	98	180	1
Denaturation	98	30	29
Annealing	51	40	
Extension	72	90	
Extension	72	300	1

2.4.2 Agarose Gel Electrophoresis

For the analysis of DNA samples, 1% agarose gels were prepared in TAE buffer (Formedium) with 5 µL per 100 mL 10 mg/mL ethidium bromide (EtBr, Invitrogen). Samples were prepared using Gel Loading Dye, Purple (6x), no SDS (NEB) and 10 µL sample per lane was loaded, with 5 µL 1 kb Plus DNA Ladder (NEB) as a marker. Gels were run at 120 V for 35 mins (small gels), 40 mins (medium gels), and 45 mins (large gels) using the Thermo Owl Easycast gel system. Gels were visualised using a G:Box F3 gel dock (Syngene) with the default setting for EtBr gels.

For gel extraction, samples were prepared with Gel Loading Dye, Purple (6x), no SDS (NEB) and 60 µL sample volume was loaded into extra wide lanes; these were otherwise run identically to analytical gels. Gels were cut under a UV lightbox, and imaging on a gel dock was skipped to minimise UV damage.

2.4.3 Construction of Plasmids

2.4.3.1 Traditional cloning

For traditional cloning, PCR amplification of inserts was followed by PCR cleanup using Qiagen PCR cleanup kit as per the product manual. NEB high fidelity restriction enzymes were used where possible, adding 1 µL of each enzyme and 5µL of 10x Cutsmart Buffer (NEB) per 50 µL reaction. The reaction was incubated for 3 hours at 37°C and deactivated at 80 °C for 20 minutes. Empty vectors were digested with the same reaction conditions, followed immediately by incubation with Calf Intestinal Phosphatase (NEB) for 1 h at 37 °C to dephosphorylate the terminal residues and prevent empty vector religation. Digested DNA was purified by gel extraction from a 1% agarose gel using a Qiagen gel extraction kit as per the product manual. Ligation reactions were achieved using T4 DNA ligase (NEB). Reactions of 20 µL total volume were made up by addition of 2 µL 10x T4 DNA ligase buffer and 1µL T4 DNA ligase to 17 µL total DNA with a twofold excess of insert over vector. These reactions were incubated at 4 °C overnight, before transformation of 2 µL ligation mix into 50µL DH5α *E. coli* cells via Heat-Shock as described in section 2.4.5.1.

2.4.3.2 Gibson cloning

For Gibson cloning, inserts with 30 bp overhangs were generated by PCR amplification. If the template for the PCR reaction had the same antibiotic resistance as the target vector, a PCR cleanup

was performed using a Qiagen PCR cleanup kit as per the product manual. This was followed by digestion using DpnI as described for other restriction digestions in section 2.4.3.1. If the template for the PCR reaction had a different antibiotic resistance to the target vector, the inserts were purified by extraction from a 1% agarose gel using a Qiagen gel extraction kit as per the product manual. Empty vectors were prepared as for traditional cloning. Gibson reactions were prepared by addition of 5 µL 2xGibson Master Mix (NEB) to 5µL of DNA with a two to fivefold excess of insert over vector. The Gibson reaction was then incubated at 50°C for 1 hour, followed by transformation of 2 µL Gibson reaction into 50µL NEBα *E. coli* cells via Heat-Shock as described in section 2.4.5.1.

2.4.3.3 Quikchange Mutagenesis

For mutagenesis via Quikchange, PCR amplification of the entire construct was performed using the Q5 DNA polymerase reaction described in section 2.4.1.1, generating nicked vector DNA containing the desired mutations. PCR product cleanup was performed with the Qiagen PCR cleanup kit as per the product manual. The template DNA from the PCR reaction was removed by digestion with DpnI for as described for other restriction digestions in section 2.4.3.1. The resulting nicked plasmid DNA was purified by gel extraction with the Qiagen gel extraction kit as per the product manual and 2µL of DNA transformed into 50µL DH5α *E. coli* cells via Heat-Shock as described in section 2.4.5.1.

E. coli codon optimised constructs were designed with flanking NdeI/HindIII restriction digest sites and synthesised in pET28 by Twist bioscience.

2.4.3.4 List of DNA Constructs used in this study

Table 2.1. Table of constructs used in this study. In the cloning method column, square brackets denote the template used for PCR and round brackets denote restriction enzymes used to linearise the backbone.

Name	**Backbone**	**Insert**	**Use**	**Cloning Method/Reference**
pJH10TS	pJH10TS	-	Constitutive overexpression of genes in *E. coli*/*P. fluorescens*.	Scott *et al.* 2017[33]
pJH10TS-BorO	pJH10TS	BorO	*E. coli*/*P. fluorescens* Bioassays	Gibson [pET28-BorO] (BmtI/KpnI)/This study
pJH10TS-FLAG-BorO	pJH10TS	N-FLAG BorO	*E. coli* Bioassays Western Blotting	Gibson [pET28-BorO] (BmtI/KpnI)/This study
pJH10TS-EcThrRS	pJH10TS	EcThrRS	*E. coli/P. fluorescens* Bioassays	Scott *et al.* 2019[174]

pJH10TS-FLAG-EcThrRS	pJH10TS	N-FLAG EcThrRS	*E. coli* Bioassays Western Blotting	Dr Sibyl Batey (JIC), This study
pJH10TS-SpThrRS	pJH10TS	SpThrRS	*E. coli* Bioassays	Gibson [pET28-SpThrRS] (BmtI/KpnI)/This study
pJH10TS-FLAG-SpThrRS	pJH10TS	N-FLAG SpThrRS	*E. coli* Bioassays Western Blotting	Gibson [pET28-SpThrRS] (BmtI/KpnI)/This study
pJH10TS-ObaO	pJH10TS	PfObaO	*E. coli/P. fluorescens* Bioassays	Scott *et al.* 2019[174]
pJH10TS-ΔN-ObaO	pJH10TS	ΔN PfObaO	*E. coli/P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-FLAG-ObaO	pJH10TS	N-FLAG PfObaO	*E. coli* Bioassays Western Blotting	Dr Sibyl Batey (JIC), This study
pUZ8002	pUZ8002	-	Conjugations	Kieser et al. 2000[180]
pMV306	pMV306	-	Cloning	Stover *et al.* 1991[181]
pIJ10700	pIJ10700	-	Cloning	Gust *et al.* 2004[182]
pGP9	pGP9	-	Conjugations	Zhang *et al.* 2009[183]
pIJ12057	pIH12057	-	Conjugations	Hong *et al.* 2005[184]
pBorOKO	pMV306	Hyg Cassette & BorO Flanking Regions	BorO Knockout in *S. parvulus*	Gibson [*S. parvulus* gDNA, pIJ10700, pMV306]/This study
pMB743	pMB743	-	Constitutive overexpression of genes in *Streptomyces*.	Dr Matt Bush (JIC), Bush *et al.* 2022[185]
pMB743-BorO	pMB743	BorO	*S. venezuelae* Bioassay/BorO Knockout Complementation	Gibson [*S. parvulus* gDNA] (NdeI/HindIII)/This study
pMB743-SpThrRS	pMB743	SpThrRS	*S. venezuelae* Bioassay/BorO Knockout Complementation	Gibson [*S. parvulus* gDNA] (NdeI/HindIII)/ This study

pET28a	pET28a	-	Inducible expression of N terminally-His tagged protein	Rosenberg *et al.* 1987[186]
pET29a	pET29a	-	Inducible expression of C terminally-His tagged protein	Rosenberg *et al.* 1987[186]
pET28-BorO	pET28a	BorO	Protein Production	Synthesised Codon Optimised/ This study
pET29-ΔN-BorO	pET29a	ΔN BorO	Protein Production	Traditional [pET28-BorO] (NdeI/XhoI) / This study
pET28-ΔN-BorO	pET28a	ΔN BorO	Protein Production	Traditional [pET28-BorO] (NdeI/XhoI) / This study
pET28-ΔN-BorO-ΔC	pET28a	ΔN BorO Δ661-675	Protein Production	Traditional [pET28-BorO] (NdeI/XhoI) / This study
pET29-ΔN-BorO-ΔC	pET29a	ΔN BorO Δ661-675	Protein Production	Traditional [pET28-BorO] (NdeI/XhoI) / This study
pET28-SpThrRS	pET28a	SpThrRS	Protein Production	Synthesised Codon Optimised/ This study
pET28-ΔN-SpThrRS	pET28a	ΔN SpThrRS	Protein Production	Traditional [pET28-SpThrRS] (NdeI/XhoI)
pET29-ΔN-SpThrRS	pET29a	ΔN SpThrRS	Protein Production	Traditional [pET28-SpThrRS] (NdeI/XhoI) / This study
pET28-EcThrRS	pET28a	EcThrRS	Protein Production	Scott *et al.* 2019[174]
pET29-ΔN-EcThrRS	pET29a	ΔN EcThrRS	Protein Production	T. Scott, PhD Thesis
pJH10TS-EcMut2	pJH10TS	EcThrRS N312Y	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study

pJH10TS-EcMut3	pJH10TS	EcThrRS F461H	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pJH10TS-EcMut4	pJH10TS	EcThrRS S488Y	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pJH10TS-EcMut5	pJH10TS	EcThrRS L489T	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pJH10TS-EcMut6	pJH10TS	EcThrRS N312Y, F461H	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pJH10TS-EcMut7	pJH10TS	EcThrRS N312Y, S488Y	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pJH10TS-EcMut8	pJH10TS	EcThrRS N312Y, L489T	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pJH10TS-EcMut9	pJH10TS	EcThrRS F461H, S488Y	*E. coli* Bioassays	Gibson [pJH10TS-EcMut3] (BmtI/KpnI) / This study
pJH10TS-EcMut10	pJH10TS	EcThrRS F461H, L489T	*E. coli* Bioassays	Gibson [pJH10TS-EcMut3] (BmtI/KpnI) / This study
pJH10TS-EcMut11	pJH10TS	EcThrRS S488Y, L489T	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pJH10TS-EcMut12	pJH10TS	EcThrRS N312Y, F461H, S488Y	*E. coli* Bioassays	Gibson [pJH10TS-EcMut6] (BmtI/KpnI) / This study
pJH10TS-EcMut13	pJH10TS	EcThrRS N312Y, F461H, L489T	*E. coli* Bioassays	Gibson [pJH10TS-EcMut6] (BmtI/KpnI) / This study

pJH10TS-EcMut14	pJH10TS	EcThrRS N312Y, S488Y, L489T	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pJH10TS-EcMut15	pJH10TS	EcThrRS F461H, S488Y, L489T	*E. coli* Bioassays	Gibson [pJH10TS-EcMut3] (BmtI/KpnI) / This study
pJH10TS-EcMut16	pJH10TS	EcThrRS N312Y, F461H, S488Y, L489T	*E. coli* Bioassays	Gibson [pJH10TS-EcMut6] (BmtI/KpnI) / This study
pET28-EcMut5	pET28a	EcThrRS L489T	Protein Production	Gibson [pJH10TS-EcMut5] (NdeI/HindIII) / This study
pET29-ΔN-EcMut5	pET29a	ΔN EcThrRS L489T	Protein Production	Gibson [pJH10TS-EcMut5] (NdeI/HindIII) / This study
pET29-ΔN-EcMut8	pET29a	ΔN EcThrRS N312Y, L489T	Protein Production	Gibson [pJH10TS-EcMut8] (NdeI/HindIII) / This study
pET29-ΔN-EcMut11	pET29a	ΔN EcThrRS S488Y, L489T	Protein Production	Gibson [pJH10TS-EcMut11] (NdeI/HindIII) / This study
pJH10TS-EcMutQ	pJH10TS	EcThrRS L489Q	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pET28-EcMutQ	pET28a	EcThrRS L489Q	Protein Production	Gibson [pJH10TS-EcMutQ] (NdeI/HindIII) / This study
pET29-ΔN-EcMutQ	pET29a	ΔN EcThrRS L489Q	Protein Production	Gibson [pJH10TS-EcMutQ] (NdeI/HindIII) / This study
pJH10TS-EcMutA	pJH10TS	EcThrRS D486A	*E. coli* Bioassay	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study

pET28-EcMutA	pET28a	EcThrRS D486A	Protein Production	Gibson [pJH10TS-EcMutA] (NdeI/HindIII) / This study
pET29-ΔN-EcMutA	pET29a	ΔN EcThrRS D486A	Protein Production	Gibson [pJH10TS-EcMutA] (NdeI/HindIII) / This study
pET28-BorOMutL	pET28a	BorO T510L	Protein Production	Quikchange [pET28-BorO] / This study
pET29-ΔN-BorOMutL	pET29a	ΔN BorO T510L	Protein Production	Gibson [pET29-ΔN-BorO] (NdeI/HindIII) / This study
pET28-SpMutL	pET28a	SpThrRS Q516L	Protein Production	Quikchange [pET28-SpThrRS] / This study
pET29-ΔN-SpMutL	pET29a	ΔN SpThrRS Q516L	Protein Production	Quikchange [pET29-ΔN-SpThrRS] / This study
pJH10TS-PfThrRS	pJH10TS	PfThrRS	*E. coli/P. fluorescens* Bioassays	Scott *et al.* 2019[174]
pJH10TS-BmObaO	pJH10TS	BmObaO	*E. coli/P. fluorescens* Bioassays	Gibson [pET28-BmObaO] (BmtI/KpnI) / This study
pJH10TS-ΔN-BmObaO	pJH10TS	ΔN BmObaO	*E. coli/P. fluorescens* Bioassays	Gibson [pET28-BmObaO] (BmtI/KpnI) / This study
pJH10TS-CsObaO	pJH10TS	CsObaO	*E. coli/P. fluorescens* Bioassays	Gibson [pET28-CsObaO] (BmtI/KpnI) / This study
pJH10TS-ΔN-CsObaO	pJH10TS	ΔN CsObaO	*E. coli/P. fluorescens* Bioassays	Gibson [pET28-CsObaO] (BmtI/KpnI) / This study
pJH10TS-PIObaO	pJH10TS	PIObaO	*E. coli/P. fluorescens* Bioassays	Gibson [pET28-PIObaO] (BmtI/KpnI) / This study

pJH10TS-ΔN-PIObaO	pJH10TS	ΔN PIObaO	*E. coli/P. fluorescens* Bioassays	Gibson [pET28-PIObaO] (BmtI/KpnI) / This study
pET28-ObaO	pET28a	PfObaO	Protein Production	Scott *et al.* 2019[174]
pET29-ΔN-ObaO	pET29a	ΔN PfObaO	Protein Production	T. Scott, PhD Thesis
pET28-BmObaO	pET28a	BmObaO	Protein Production	Synthesised Codon Optimised/ This study
pET29-ΔN-BmObaO	pET29a	ΔN BmObaO	Protein Production	Traditional [pET28-BmObaO] (NdeI/HindIII) / This study
pET28-CsObaO	pET28a	CsObaO	Protein Production	Synthesised Codon Optimised/ This study
pET29-ΔN-CsObaO	pET29a	ΔN CsObaO	Protein Production	Traditional [pET28-CsObaO] (NdeI/HindIII) / This study
pET28-PIObaO	pET28a	PIObaO	Protein Production	Synthesised Codon Optimised/ This study
pET29-ΔN-PIObaO	pET29a	ΔN PIObaO	Protein Production	Traditional [pET28-PIObaO] (NdeI/HindIII) / This study
pJH10TS-EcMutM	pJH10TS	EcThrRS L489M	*E. coli* Bioassays	Dr Sibyl Batey (JIC), This study
pET28-EcMutM	pET28a	EcThrRS L489M	Protein Production	Gibson [pJH10TS-EcMutM] (NdeI/HindIII) / This study
pET29-ΔN-EcMutM	pET29a	ΔN EcThrRS L489M	Protein Production	Gibson [pJH10TS-EcMutM] (NdeI/HindIII) / This study
pJH10TS-BmMutL	pJH10TS	BmObaO M490L	*E. coli* Bioassays	Gibson [pJH10TS-BmObaO] (BmtI/KpnI) / This study

pET28-BmMutL	pET28a	BmObaO M490L	Protein Production	Gibson [pJH10TS-BmMutL] (NdeI/HindIII) / This study
pET29-BmMutL	pET29a	ΔN BmObaO M490L	Protein Production	Gibson [pJH10TS-BmMutL] (NdeI/HindIII) / This study
pJH10TS-Chimera1	pJH10TS	EcThrRS Catalytic & Anticodon Binding Domain PfObaO Editing Domain	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-Chimera2	pJH10TS	EcThrRS Editing & Anticodon Binding Domains, PfObaO Catalytic Domain	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-Chimera3	pJH10TS	EcThrRS Editing & Catalytic Domain, PfObaO Anticodon Binding Domain	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-Chimera4	pJH10TS	EcThrRS Editing Domain PfObaO Catalytic & Anticodon Binding Domain	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-Chimera5	pJH10TS	EcThrRS Catalytic Domain, PfObaO Editing & Anticodon Binding Domain	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-Chimera6	pJH10TS	EcThrRS Anticodon	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study

		Binding Domain, PfObaO Editing & Catalytic Domains		
pJH10TS-EcC1	pJH10TS	EcThrRS, PfObaO catalytic subdomain 1	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-EcC2	pJH10TS	EcThrRS, PfObaO catalytic subdomain 2	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-EcC3	pJH10TS	EcThrRS, PfObaO catalytic subdomain 3	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-EcC4	pJH10TS	EcThrRS, PfObaO catalytic subdomain 4	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-EcC5	pJH10TS	EcThrRS, PfObaO catalytic subdomain 5	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-ObC1	pJH10TS	PfObaO, EcThrRS catalytic subdomain 1	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-ObC2	pJH10TS	PfObaO, EcThrRS catalytic subdomain 2	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-ObC3	pJH10TS	PfObaO, EcThrRS catalytic subdomain 3	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-ObC4	pJH10TS	PfObaO, EcThrRS catalytic subdomain 4	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-ObC5	pJH10TS	PfObaO, EcThrRS catalytic subdomain 5	*P. fluorescens* Bioassays	Dr Sibyl Batey (JIC), This study

pJH10TS-Ec305	pJH10TS	EcThrRS E305K	*E. coli* Bioassays	Gibson [pJH10TS-EcThrRS] (BmtI/KpnI) / This study
pJH10TS-Ob463	pJH10TS	PfObaO Y463F	*E. coli* Bioassays	Gibson [pJH10TS-ObaO] (BmtI/KpnI) / This study
pJH10TS-Ec463	pJH10TS	EcThrRS G463S	*E. coli* Bioassays	Dr Sibyl Batey (JIC), This study
pJH10TS-MKgeneO	pJH10TS	Compound 1 BGC *gene O*	*E. coli Bioassays*	Gibson [pET28-MpgeneO] (BmtI/KpnI) / This study
pET28-MKgeneO	pET28a	Compound 1 BGC *gene O*	Protein Production	Synthesised Codon Optimised / This study
pET29-ΔN-MKgeneO	pET29a	Compound 1 BGC *gene O*	Protein Production	Gibson [pET28-MpgeneO] (NdeI/HindIII) / This study

2.4.3.5 List of primers used in this study

Table 2.2. List of primers used in This study. Mutagenic regions shown in bold, Gibson overlaps underlined, restriction sites italics and non-annealing overhangs for restriction digest in lower case.

Name	**Sequence**	**Purpose**
pJH10TSSeqF	GCCGACATCATAACGGTTCTGGC	Colony PCR/Sanger sequencing of pJH10TS constructs
pJH10TSSeqR	GTTCCTCCTGCCAGTTGATGACC	Colony PCR/Sanger sequencing of pJH10TS constructs
BorOJHF	AGAATTCTAGTCAATTGGTCATTAATTAACTGC*GCTAGC*ATGTCCGTAATTCGCCCGACG	Construction of pJH10TS-BorO
BorOJHR	CTCCAGCGAGCTCTCTAGAATCGAT*GGTACC*TTACGCTTTGCTAAGAGGGCGAATATG	Construction of pJH10TS-BorO, pJH10TS-FLAG-BorO
BorOFLAGF	AGTCAATTGGTCATTAATTAACTGCGCT**A TGGATTATAAAGATGATGATGATAAA**ATGTCCGTAATTCGCCCGACGGCAGAAACT	Construction of pJH10TS-FLAG-BorO

EcThrRSJHF	AGTCAATTGGTCATTAATTAACTGC*GCTA GC*ATGCCTGTTATAACTCTTCCTGATGGC	Construction of pJH10TS-EcMut2, pJH10TS-EcMut3, pJH10TS-EcMut4, pJH10TS-EcMut5, pJH10TS-EcMut6, pJH10TS-EcMut7, pJH10TS-EcMut8, pJH10TS-EcMut9, pJH10TS-EcMut10, pJH10TS-EcMut11, pJH10TS-EcMut12, pJH10TS-EcMut13, pJH10TS-EcMut14, pJH10TS-EcMut15, pJH10TS-EcMut16, pJH10TS-EcMutQ, pJH10TS-EcMutA, pJH10TS-EcMutM, pJH10TS-Ec305
EcThrRSJHR	CCAGCGAGCTCTCTAGAATCGAT*GGTACC* TTATTCCTCCAATTGTTTAAGACTGCGGC	Construction of pJH10TS-EcMut2, pJH10TS-EcMut3, pJH10TS-EcMut4, pJH10TS-EcMut5, pJH10TS-EcMut6, pJH10TS-EcMut7, pJH10TS-EcMut8, pJH10TS-EcMut9, pJH10TS-EcMut10, pJH10TS-EcMut11, pJH10TS-EcMut12, pJH10TS-EcMut13, pJH10TS-EcMut14, pJH10TS-EcMut15, pJH10TS-EcMut16, pJH10TS-EcMutQ, pJH10TS-EcMutA, pJH10TS-EcMutM, pJH10TS-Ec305
SpThrRSJHF	AGTCAATTGGTCATTAATTAACTGC*GCTA GC*ATGAGCGATGTGCGTGTTATTATTCAG CG	Construction of pJH10TS-SpThrRS
SpThrRSJHR	CCAGCGAGCTCTGTAGAATCGAT*GGTACC* TTAAACTTGAGCACGTTCTTCAACTACTTT T	Construction of pJH10TS-SpThrRS, pJH10TS-FLAG-SpThrRS
SpThrRSFLAGF	AGTCAATTGGTCATTAATTAACTGCGCT**A TGGATTATAAAGATGATGATGATAAA**AT GAGCGATGTGCGTGTTATTATTCAGCG	Construction of pJH10TS-FLAG-SpThrRS

BorOKOUpF	GAATTCGAAGCTTATCGATGGCGGTGCTGCCAAGC	Construction of pBorOKO
BorOKOUpR	GGTCGACGGATCCCCGGAATCACTGGAATCTCTCCTCGG	Construction of pBorOKO
BorODownF	CGAAGCAGCTCCAGCCTACATGACCCACAGCCACGGG	Construction of pBorO*KO*
BorODownR	ACACAACGTCGCTTTGTTGGCGCGAGGCGCCGCAG	Construction of pBorO*KO*
HygF	CCCGAGGAGAGATTCCAGTGATTCCGGGGATCCGTCGAC	Construction of pBorO*KO*
HygR	GGCCCCGTGGCTGTGGGTCATGTAGGCTGGAGCTGCTTCG	Construction of pBorO*KO*
pMVBBF	CACACCTGCGGCGCCTCGCGCCAACAAAGCGACGTTGTGTC	Construction of pBorO*KO*
pMVBBR	AGCTGCTTGGCCAGCACCGCCATCGATAAGCTTCGAATTCTGC	Construction of pBorO*KO*
UpSeqF	GGCCGGCACTCCCTACGA	Sanger sequencing of pBorO*KO*, Colony PCR of *borO* knockout strains
DownSeqF	CGCAGGCGCAGTACGAA	Sanger sequencing of pBorO*KO*
DownSeqR	TTCGTACTGCGCCTGCG	Sanger sequencing and colony PCR of *borO* knockout strains
HygSeqF	CCCGGTGATCAAGCTGTT	Sanger sequencing of pBorO*KO*
pMVBBSeqF	GGTCCACCTACAACAAAGCT	Sanger sequencing of pBorO*KO*
pMBSeqF	TTGACGGCTGGCGAGAGGTG	Colony PCR/Sanger Sequencing of pMB743 constructs
pMBSeqR	GCGAGCTGAAGAAAGACAAT	Colony PCR/Sanger Sequencing of pMB743 constructs
BorOMBF	ATCGTCTAGAACAGGAGGCCC*CATATG*GTGTCTGTAATCCGTCCCACC	Construction of pMB743-BorO
BorOMBR	TGAGAACCCTAGGGGATCC*AAGCTT*TCAGGCCTTGGACAGCGGA	Construction of pMB743-BorO
BorOIntSeqF	TCGCGTTGGGGCCATTG	Sanger Sequencing of BorO WT constructs

BorOIntSeqR	ACCCCGATGTCCCTGG	Sanger Sequencing of BorO WT constructs
SpMBF	ATCGTCTAGAACAGGAGGCCC*CATATG*TCAGACGTCCGTGTGATC	Construction of pMB743-SpThrRS
SpMBR	TGAGAACCCTAGGGGATCC*AAGCTT*TCAGACCTGCGCGCGCTCCTC	Construction of pMB743-SpThrRS
SpIntSeqF	TCGAGCGCTGGCCGGCG	Sanger Sequencing of SpThrRS WT constructs
SpIntSeqR	CTCCTTCCGCTACCGCG	Sanger Sequencing of SpThrRS WT constructs
pET28SeqF	TGAGCGGATAACAATTCCCC	Colony PCR/Sanger Sequencing of pET28a constructs
pET28SeqR	GCTAGTTATTGCTCAGCGG	Colony PCR/Sanger Sequencing of pET28a constructs
pET29SeqF	GTAGAGGATCGAGATCGATC	Colony PCR/Sanger Sequencing of pET29a constructs
pET29SeqR	CAAGACCCGTTTAGAGGCCC	Colony PCR/Sanger Sequencing of pET29a constructs
DNBorOF	aaaaaa*CATATG*CGCGATCATCGTCGTATTGGG	Construction of pET29-ΔN-BorO, pET28-ΔN-BorO, pET28-ΔN-BorO-ΔC, pET29-ΔN-BorO-ΔC
DNBorOR	aaaaaa*CTCGAG*CGCTTTGCTAAGAGGGCGAATATG	Construction of pET29-ΔN-BorO
DNBorONHisR	aaaaaa*CTCGAG*TTACGCTTTGCTAAGAGGGCG	Construction of pET28-ΔN-BorO
DNBorODCNHisR	aaaaaa*CTCGAG*TTATGTGCGAATCAGGCCGGTTAC	Construction of pET28-ΔN-BorO-ΔC
DNBorODCCHisR	aaaaaa*CTCGAG*TGTGCGAATCAGGCCGGTTAC	Construction of pET29-ΔN-BorO-ΔC
DNSpThrRSF	aaaaaa*CATATG*CGTGATCATCGTAAATTGGG	Construction of pET29-ΔN-SpThrRS, pET28-ΔN-SpThrRS
DNSpThrRSNHisR	aaaaaa*CTCGAG*TTAAACTTGAGCACGTTCTTCAAC	Construction of pET28-ΔN-SpThrRS
DNSpThrRSR	aaaaaa*CTCGAG*TAACTTGAGCACGTTCTTCAACAACTAC	Construction of pET29-ΔN-SpThrRS

EcThrRSF	TTTAAGAAGGAGATATACATATGCCTGTTATAACTCTTCCTGATGGC	Construction of pET28-EcMut5, pET28-EcMutA, pET28-EcMutQ, pET28-EcMutM
EcThrRSR	TGGTGGTGGTGGTGCTCGAGTTATTCCTCCAATTGTTTAAGACTGCGGC	Construction of pET28-EcMut5, pET28-EcMutA, pET28-EcMutQ, pET28-EcMutM
DNEcThrRSF	GTTTAACTTTAAGAAGGAGATATACATATGCGCGACCACCGTAAAATC	Construction of pET29-ΔN-EcMut5, pET29-ΔN-EcMut8, pET29-ΔN-EcMut11, pET29-ΔN-EcMutA, pET29-ΔN-EcMutQ, pET29-ΔN-EcMutM
DNEcThrRSR	GTGGTGGTGCTCGAGTGCGGCCGC*AAGCTT*CTTTTCCTCCAATTGTTTAAGACTGCGGC	Construction of pET29-ΔN-EcMut5, pET29-ΔN-EcMut8, pET29-ΔN-EcMut11, pET29-ΔN-EcMutA, pET29-ΔN-EcMutQ, pET29-ΔN-EcMutM
EcThrRSN312YF	GTGGGAAAAAACCGGTCACTGGGAC**TAC**TACAAAGATGCAATGTTCACCACATCTTCTGA	Construction of pJH10TS-EcMut2, pJH10TS-EcMut6, pJH10TS-EcMut7, pJH10TS-EcMut8, pJH10TS-EcMut14
EcthrRSN312YR	CTCAGAAGATGTGGTGAACATTGCATCTTTGTA**GTA**GTCCCAGTGACCGGTTTTTTCCCA	Construction of pJH10TS-EcMut2, pJH10TS-EcMut6, pJH10TS-EcMut7, pJH10TS-EcMut8, pJH10TS-EcMut14
EcThrRSF461HF	CTGGGTGAAGGCGCT**CAC**TACGGTCCGAAAATTGAATTTACCC	Construction of pJH10TS-EcMut3, pJH10TS-EcMut6
EcThrRSF461HR	AATTTTCGGACCGTA**GTG**AGCGCCTTCACCCAGTTGATATTC	Construction of pJH10TS-EcMut3, pJH10TS-EcMut6
EcThrRSS488YF	GTACAGCTGGACTTC**TAT**TTGCCGTCTCGTCTGAGCGCTTC	Construction of pJH10TS-EcMut4, pJH10TS-EcMut7, pJH10TS-EcMut9, pJH10TS-EcMut12
EcThrRSS488YR	GCTCAGACGAGACGGCAA**ATA**GAAGTCCAGCTGTACTGTACCC	Construction of pJH10TS-EcMut4, pJH10TS-EcMut7, pJH10TS-EcMut9, pJH10TS-EcMut12

EcThrRSL489TF	GTACAGCTGGACTTCTCT**ACG**CCGTCTCGTCTGAGCGCTTC	Construction of pJH10TS-EcMut5, pJH10TS-EcMut8, pJH10TS-EcMut10, pJH10TS-EcMut13
EcThrRSL489TR	GCTCAGACGAGACGG**CGT**AGAGAAGTCCAGCTGTACTGTACCC	Construction of pJH10TS-EcMut5, pJH10TS-EcMut8, pJH10TS-EcMut10, pJH10TS-EcMut13
EcThrRSSL488YTF	GTACAGCTGGACTTC**TATACG**CCGTCTCGTCTGAGCGCTTC	Construction of pJH10TS-EcMut11, pJH10TS-EcMut14, pJH10TS-EcMut15, pJH10TS-EcMut16
EcThrRSSL488YTR	GCTCAGACGAGACGG**CGTATA**GAAGTCCAGCTGTACTGTACCC	Construction of pJH10TS-EcMut11, pJH10TS-EcMut14, pJH10TS-EcMut15, pJH10TS-EcMut16
EcThrRSL489QF	GTACAGCTGGACTTCTCT**CAG**CCGTCTCGTCTGAGCGCTTC	Construction of pJH10TS-EcMutQ
EcThrRSL489QR	GCTCAGACGAGACGG**CTG**AAGAGAAGTCCAGCTGTACTGTACCC	Construction of pJH10TS-EcMutQ
EcThrRSD486AF	GGTACAGTACAGCTG**GCG**TTCTCTTTGCCGTCTCGTCTGAG	Construction of pJH10TS-EcMutA
EcThrRSD486AR	AGACGGCAAAGAGAA**CGC**CAGCTGTACTGTACCCACCTGC	Construction of pJH10TS-EcMutA
BmObaOJHF	AGTCAATTGGTCATTAATTAACTGC*GCTAGC*ATGATCAGCATTGCGCTCCCAG	Construction of pJH10TS-BmObaO, pJH10TS-BmMutL
BmObaOJHR	CCAGCGAGCTCTCTAGAATCGAT*GGTACC*TCATGCCGGGCGCTCCGAAC	Construction of pJH10TS-BmObaO, pJH10TS-ΔN-BmObaO, pJH10TS-BmMutL
DNBmObaOJHF	ATTAATTAACTGC*GCTAGC*ATGCGCGACCACCGCAAAATTGGTAAG	Construction of pJH10TS-ΔN-BmObaO
CsObaOJHF	AGTCAATTGGTCATTAATTAACTGC*GCTAGC*ATGATTACTATTAGTTTGCCTGATGGATC	Construction of pJH10TS-CsObaO
CsObaOJHR	CTCCAGCGAGCTCTCTAGAATCGAT*GGTACC*TTATCCAATGGCCGTCTCCTCACG	Construction of pJH10TS-CsObaO, pJH10TS-ΔN-CsObaO
DNCsObaOJHF	ATTAATTAACTGC*GCTAGC*ATGCGCGACCACCGCAAGATTGCTAAG	Construction of pJH10TS-ΔN-CsObaO

PIObaOJHF	AATTCTAGTCAATTGGTCATTAATTAACTG C*GCTAGC*ATGGTTTCAATCGCTTTACCGG	Construction of pJH10TS-PIObaO
PIObaOJHR	CTCCAGCGAGCTCTCTAGAATCGAT*GGTA CC*TCAATCCACGACATTCTCACGCGTC	Construction of pJH10TS-PIObaO, pJH10TS-ΔN-PIObaO
DNPIObaOJHF	ATTAATTAACTGC*GCTAGC*ATGCGCGATC ATCGTAAGTTGGCCAAG	Construction of pJH10TS-ΔN-PIObaO
DNBmObaOF	aaaaaa*CATATG*CGCGACCACCGCAAAATT GGTAAG	Construction of pET29-ΔN-BmObaO
DNBmObaOR	aaaaaa*AAGCTT*TGCCGGGCGCTCCGAAC G	Construction of pET29-ΔN-BmObaO
DNCsObaOF	aaaaaa*CATATG*CGCGACCACCGCAAGATT GCTAAG	Construction of pET29-ΔN-CsObaO
DNCsObaOR	aaaaaa*AAGCTT*TCCAATGGCCGTCTCCTC ACGC	Construction of pET29-ΔN-CsObaO
DNPIObaOF	aaaaaa*CATATG*CGCGATCATCGTAAGTTG GCCAAG	Construction of pET29-ΔN-PIObaO
DNPIObaOR	aaaaaa*AAGCTT*ATCCACGACATTCTCACG CGTCAAG	Construction of pET29-ΔN-PIObaO
EcThrRSL489M F	GTACAGCTGGACTTCTCT**ATG**CCGTCTCG TCTGAGCGCTTC	Construction of pJH10TS-EcMutM
EcThrRSL489M R	GCTCAGACGAGACGG**CAT**AGAGAAGTCC AGCTGTACTGTACCG	Construction of pJH10TS-EcMutM
BmM490LF	AGGCAGATTACCTC**CTC**CCTGAGAAGCTC G	Construction of pJH10TS-BmMutL
BmM490LR	CGAGCTTCTCAGG**GAG**GAGGTAATCTGC CT	Construction of pJH10TS-BmMutL
EcE305KF	CTGTGG**AAA**AAAACCGGTCACTGGGACA ACTACAAAG	Construction of pJH10TS-Ec305
EcE305KR	TGTAGTTGTCCCAGTGACCGGTTTT**TTT**CC ACAGGACACGGTCCATCATG	Construction of pJH10TS-Ec305
ObaOJHF	AGTCAATTGGTCATTAATTAACTGC*GCTA GC*ATGGTCACTATCGCTCTACCGGACG	Construction of pJH10TS-Ob463
ObaOJHR	TCCAGCGAGCTCTCTAGAATCGAT*GGTAC C*TCAGCCGCCTGCTGATTGCTCGT	Construction of pJH10TS-Ob463

ObaOY463FF	AAGGCGCGTTC**TTC**AGCCCCAAGATCGAG TACCACCTG	Construction of pJH10TS-Ob463
ObaOY463FR	TACTCGATCTTGGGGCT**GAA**GAACGCGCC TTCG	Construction of pJH10TS-Ob463
PfThrRSSeqF	GACGAAGGCGTCACGGGCA	Amplification of PfThrRS and its 5'UTR from *P. fluorescens* ATCC 39502 gDNA
PfThrRSSeqR	GACACAATCCCAAGCTGCTCACCT	Amplification of PfThrRS and its 5'UTR from *P. fluorescens* ATCC 39502 gDNA
MpgeneOJHF	AGTCAATTGGTCATTAATTAACTGC*GCTA GC*ATGGTTCGTGATGCCGACGGT	Construction of pJH10TS-MpgeneO
MpgeneOJHR	CCAGCGAGCTCTCTAGAATCGAT*GGTACC* TTAACCAGCGGCGGGAATGGT	Construction of pJH10TS-MpgeneO
DNMpgeneOF	aaaaaa*CATATG*CGTGACCACCGCCGTAT	Construction of pET29-ΔN-MpgeneO
DNMpgeneOR	aaaaaa*AAGCTT*ACCAGCGGCGGGAATGG T	Construction of pET29-ΔN-MpgeneO

2.4.4 DNA sequencing

2.4.4.1 Whole Genome Sequencing (WGS)

2.4.4.1.1 S. parvulus WGS

Spores of *S. parvulus* were streaked for single colonies on MAM agar and incubated for 2 days at 30 °C before inoculating single colonies into 50 mL TSB and incubated for 2 days at 30°C and 200 rpm shaking. Cells were harvested by centrifugation using an Eppendorf centrifuge S810 R at 4000 rpm for 10 mins. Genomic DNA (gDNA) was prepared via the salting out method as previously described[180]. Whole genome sequencing was performed by Novogene using PCR-free Illumina sequencing. Illumina reads were trimmed with Trimmomatic, assembled using the SPAdes genome assembler and analysed using Quast, Icarus and antiSMASH by Dr Govind Chandra (JIC).

2.4.4.1.2 Micromonospora sp. KC207 WGS

For sequencing of *Micromonospora sp.* KC207, bacteria were streaked for single colonies on GYM agar and incubated for 2 weeks at 30°C. Single colonies were used to inoculate 50 mL liquid GYM which was incubated for 7 days at 30°C and 200 rpm shaking. Cells were harvested by centrifugation (as described above), washed with 1 mL sterile phosphate buffered saline (PBS) and then

resuspended in 0.5 mL DNA/RNA Shield buffer (Zymo Research, Irvine, USA) and shipped to MicrobesNG (Birmingham, UK). gDNA was extracted at MicrobesNG from the bacteria and both Illumina and nanopore sequencing was used. Illumina reads were assembled into contigs by MicrobesNG and data were analysed by Dr Govind Chandra. Nanopore sequencing gave a single contig which was used to scaffold for Illumina reads and they were compared to check for errors using bcftools. Differences between Illumina and nanopore were only in places where Illumina had poor coverage so nanopore was taken as the correct sequence. Nanopore contig was annotated using the Prokka annotation pipeline.

2.4.4.2 Sanger Sequencing

For the confirmation of plasmids constructed in This study, plasmid templates were prepared by extracting from *E. coli* using the Qiagen Miniprep Kit as per the product manual. DNA was eluted with 30 µL EB, and the concentration estimated using a DeNovix DS-11 Spectrophotometer (Nanodrop) and diluted to 100 ng/µL before sending to Source Bioscience (Cambridge, UK) with sequencing primers sent at 3.3 ng/µL. For the sequencing of the *P. fluorescens* PfThrRS gene, PCR products were produced using the general touchdown method as described in section 2.4.1.1. PCR products were cleaned up using the Qiagen PCR Clean-up kit as per the product manual, and eluted with 30 µL EB. The DNA concentration was estimated using Nanodrop and diluted to 10 ng/µL before sending to Source Bioscience (Cambridge, UK) with sequencing primers sent at 3.3 ng/µL.

2.4.5 Construction of Strains

2.4.5.1 Transformation of E. coli

Chemically competent *E. coli* cells were transformed with plasmids via heat shock at 42°C for 30 secs, and recovered at 37°C for 1 h before plating on LB agar containing the relevant antibiotics.

2.4.5.2 Conjugal transfer of plasmids into Streptomyces spp.

Constructs were introduced into *Streptomyces* spp. By conjugal transfer from *E. coli* using standard methods[180]. Both *E. coli* ET12567(pUZ8002) and *E. coli* S17 cells were used as donor strains for conjugal transfer, and in each case the same procedure was used.

2.4.5.3 Knockout of borO in Streptomyces parvulus ISOM-0991 and ISOM-1302 (⊿borO)

A suicide vector was constructed to replace the *borO* gene with a hygromycin resistance cassette in the *S. parvulus* genome. To achieve this, 1.5 kb of DNA from upstream and downstream of *borO* were amplified from *S. parvulus* ISOM-0991 gDNA. The hygromycin resistance cassette was amplified from pIJ10700 - this included an *oriT* for transfer of the plasmid into *Streptomyces*. The suicide vector backbone was pMV306 between the SalI and BmtI restriction sites- including the kanamycin resistance cassette and pBR1 derived high copy number origin. This plasmid while

originally used for *Mycobacterial* integration, and so will not replicate in *Streptomyces*, meaning that a crossover event with the genome is required for vector maintenance in *Streptomyces*.

Once constructed, the knockout vector was conjugated into both *S. parvulus* ISOM-0991 and ISOM-1302 as described in section 2.4.5.2. Overlays with nalidixic acid and hygromycin were done. The resulting plates were incubated at 30 °C for 7 days. The resulting exconjugants were streaked out on MAM with no antibiotics and incubated at 30 °C to allow for double crossover and ejection of the plasmid to occur. After 4 days, the different isolates were picked and patched onto MAM agar containing either hygromycin or hygromycin + kanamycin as described in section 2.3. Single colonies which grew on hygromycin, but not hygromycin + kanamycin, were again picked and patched onto MYM agar plus hygromycin. followed by colony PCR to detect a band the size of the hygromycin resistance cassette but lacking a band the size of the *borO* gene (3 kb vs 2.5 kb).

2.4.5.4 Transformation of Pseudomonas spp.

Conjugation into *Pseudomonas fluorescens* was performed using *E. coli* S17 cells as a donor, overlaying with tetracycline and streaking out with selection for single colonies and confirming introduction of plasmid using colony PCR.

2.4.5.5 List of strains used in this study

Table 2.3. List of strains used in this study.

Strain Name	**Description**	**Reference/Source of strain**
Methicillin susceptible *Staphylococcus aureus* (MSSA)	Bioassay strain, ATCC 6538P	American Type Culture Collection
Methicillin resistant *Staphylococcus aureus* (MRSA)	Bioassay strain, ATCC BAA-1717	American Type Culture Collection
E. coli 25922	Bioassay strain, ATCC 25922, WT	American Type Culture Collection
E. coli NR698	Membrane permeabilised bioassay strain, MC4100 (F^- *araD139* Δ(*argF-lac*)*U169, rpsL150, relA1, flbB5301, deoC1, ptsF25, rbsR), imp4213*	Ruiz *et al.* 2005[187]/Wilkinson lab strain
Bacillus subtilis EC1524	Bioassay strain, *trpC2*, subtilin BGC deleted	Widdick *et al.* 2003[188]/Wilkinson lab strain

E. coli DH5α	Cloning strain, F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG* ϕ80*dlac*Δ*(lacZ)M15* Δ(*lacIZYA-argF*)U169 *hsdR*17($r_K^-m_K^+$) λ^-	Thermo fisher scientific
E. coli NEB5α	Cloning strain, *fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
Streptomyces venezuelae NRRL B-65442	Bioassay strain	Gomez-Escribano *et al.* 2021[189] / Wilkinson lab strain
S. lividans 1326	Bioassay strain	Wilkinson lab strain
S. coelicolor M1146	Bioassay strain	Gomez-Escribano *et al.* 2011[190] / Wilkinson lab strain
S. albus J1074	Bioassay strain	Wilkinson lab strain
E. coli ET12567:pUZ8002	Donor strain for conjugations from *E. coli* into *Streptomyces*, *dam*$^-$ *dcm*$^-$ *hsdS*$^-$, pUZ8002	MacNeil *et al.* 1992[191] / Wilkinson lab strain
E. coli S17-1λ(pir)	Donor strain for conjugations from *E. coli* into *Pseudomonas* and *Streptomyces*, *recA thi pro hsd*(R^-M^+)RP4: 2-Tc::Mu-Km::Tn7 λpir SM^R Tp^R	Wilkinson lab strain
S. parvulus ISOM-0991	WT Borrelidin producing strain, Tü4055	Isomerase Therapeutics (Cambridge, UK)
S. parvulus Δ*borO*	*S. parvulus* ISOM-0991 with in-frame deletion of *borO*	This study
S. parvulus ΔborO:pMB743	S. parvulus ISOM-0991 with in-frame deletion of *borO*, contains pMB743	This study

S. parvulus ΔborO:pMB743-BorO	S. parvulus ISOM-0991 with in-frame deletion of *borO*, contains pMB743-BorO	This study
S. parvulus ΔborO:pMB743-SpThrRS	S. parvulus ISOM-0991 with in-frame deletion of *borO* contains pMB743-SpThrRS	This study
S. parvulus ISOM-1302	Borrelidin non-producing strain, WT carrying in-frame deletion of *borG*	Olano *et al.* 2004[192]/Isomerase Therapeutics
S. parvulus Δ*borG* Δ*borO*	*S. parvulus* carrying in frame deletions in *borO* and *borG*	This study
S. parvulus Δ*borG* Δ*borO*:pMB743	S. parvulus carrying in frame deletions in *borO* and *borG*, contains pMB743	This study
S. parvulus Δ*borG* Δ*borO*:pMB743-BorO	S. parvulus carrying in frame deletions in *borO* and *borG*, contains pMB743-BorO	This study
S. parvulus Δ*borG* Δ*borO*:pMB743-SpThrRS	S. parvulus carrying in frame deletions in *borO* and *borG*, contains pMB743-SpThrRS	This study
E. coli NiCo21(DE3)	Protein production strain, BL21(DE3) derivative *can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	New England Biolabs
P. fluorescens ATCC 39502	Obafluorin Producer, ATCC-39502, WT	American Type Culture Collection
P. fluorescens ATCC 39502 Δ*obaL* Δ*obaO*	Obafluorin Producer with in-frame *obaL* and *obaO* deletions	Scott *et al.* 2019[174] / Wilkinson lab strain

Micromonospora sp. KC207	Desert-derived strain containing potential novel ThrRS inhibitor BGC	Gifted by Prof. Nevzat Sahin, Ondokuz Mayis University, Turkey. Strain is unpublished, genome deposited on GenBank at accession ASM434861v1.
E. coli NR698:pJH10TS	*E. coli b*ioassay strain, pJH10TS empty vector negative control	Dr Sibyl Batey (JIC)
P. fluorescens Δ*obaL*Δ*obaO*:pJH10TS	*P. fluorescens* bioassay strain, Δ*obaL,* Δ*obaO*, pJH10TS empty vector negative control	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-BorO	*E. coli b*ioassay strain, pJH10TS-BorO	Dr Sibyl Batey (JIC)
P. fluorescens Δ*obaL*Δ*obaO*:pJH10TS-BorO	*P. fluorescens* bioassay strain, Δ*obaL,* Δ*obaO* pJH10TS-BorO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-FLAG-BorO	*E. coli b*ioassay strain, FLAG tagged BorO for western blotting. pJH10TS-FLAG-BorO	This study
P. fluorescens ΔobaLΔobaO:pJH10TS-EcThrRS	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO* pJH10TS-EcThrRS	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-EcThrRS	*E. coli b*ioassay strain, pJH10TS-EcThrRS	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-FLAG-EcThrRS	*E. coli b*ioassay strain, FLAG tagged EcThrRS for western blotting. pJH10TS-FLAG-EcThrRS	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-SpThrRS	*E. coli b*ioassay strain, pJH10TS-SpThrRS	This study
E. coli NR698:pJH10TS-FLAG-SpThrRS	*E. coli b*ioassay strain, FLAG tagged SpThrRS for western	This study

	blotting, pJH10TS-FLAG-SpThrRS	
P. fluorescens ΔobaLΔobaO:pJH10TS-ObaO	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO* pJH10TS-ObaO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-ObaO	*E. coli b*ioassay strain, pJH10TS-ObaO	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-ΔN-ObaO	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO* pJH10TS-ΔN-ObaO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-ΔN-ObaO	*E. coli b*ioassay strain, pJH10TS-ΔN-ObaO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-FLAG-ObaO	*E. coli b*ioassay strain, FLAG tagged ObaO for western blotting, pJH10TS-FLAG-ObaO	Dr Sibyl Batey (JIC)
S. venezuelae:pMB743	*S. venezuelae* bioassay strain, pMB743 empty vector negative control	This study
S. venezuelae:pMB743-BorO	*S. venezuelae* bioassay strain, pMB743-BorO	This study
S. *venezuelae*:pMB743-SpThrRS	*S. venezuelae* bioassay strain, pMB743-SpThrRS	This study
E. coli NiCo21:pET28-BorO	Protein production strain, pET28-BorO	This study
E. coli NiCo21:pET29-ΔN-BorO	Protein production strain, pET29-ΔN-BorO	This study
E. coli NiCo21: pET28-ΔN-BorO	Protein production strain, pET28-ΔN-BorO	This study
E. coli NiCo21:pET28-ΔN-BorO-ΔC	Protein production strain, pET28-ΔN-BorO-ΔC	This study
E. coli NiCo21:pET29-ΔN-BorO-ΔC	Protein production strain, pET29-ΔN-BorO-ΔC	This study
E. coli NiCo21:pET28-SpThrRS	Protein production strain, pET28-SpThrRS	This study

E. coli NiCo21:pET29-ΔN-SpThrRS	Protein production strain, pET29-ΔN-SpThrRS	This study
E. coli NiCo21:pET28-ΔN-SpThrRS	Protein production strain, pET28-ΔN-SpThrRS	This study
E. coli NiCo21:pET28-EcThrRS	Protein production strain, pET28-EcThrRS	This study
E. coli NiCo21:pET29-ΔN-EcThrRS	Protein production strain, pET29-ΔN-EcThrRS	This study
E. coli NR698:pJH10TS-EcMut2	*E. coli b*ioassay strain, pJH10TS-EcMut2	This study
E. coli NR698:pJH10TS-EcMut3	*E. coli b*ioassay strain, pJH10TS-EcMut3	This study
E. coli NR698:pJH10TS-EcMut4	*E. coli b*ioassay strain, pJH10TS-EcMut4	This study
E. coli NR698:pJH10TS-EcMut5	*E. coli b*ioassay strain, pJH10TS-EcMut5	This study
E. coli NR698:pJH10TS-EcMut6	*E. coli b*ioassay strain, pJH10TS-EcMut6	This study
E. coli NR698:pJH10TS-EcMut7	*E. coli b*ioassay strain, pJH10TS-EcMut7	This study
E. coli NR698:pJH10TS-EcMut8	*E. coli b*ioassay strain, pJH10TS-EcMut8	This study
E. coli NR698:pJH10TS-EcMut9	*E. coli b*ioassay strain, pJH10TS-EcMut9	This study
E. coli NR698:pJH10TS-EcMut10	*E. coli b*ioassay strain, pJH10TS-EcMut10	This study
E. coli NR698:pJH10TS-EcMut11	*E. coli b*ioassay strain, pJH10TS-EcMut11	This study
E. coli NR698:pJH10TS-EcMut12	*E. coli b*ioassay strain, pJH10TS-EcMut12	This study
E. coli NR698:pJH10TS-EcMut13	*E. coli b*ioassay strain, pJH10TS-EcMut13	This study
E. coli NR698:pJH10TS-EcMut14	*E. coli b*ioassay strain, pJH10TS-EcMut14	This study

E. coli NR698:pJH10TS-EcMut15	*E. coli b*ioassay strain, pJH10TS-EcMut15	This study
E. coli NR698:pJH10TS-EcMut16	*E. coli b*ioassay strain, pJH10TS-EcMut16	This study
E. coli NiCo21:pET28-EcMut5	Protein production strain, pET28-EcMut5	This study
E. coli NiCo21:pET29-ΔN-EcMut5	Protein production strain, pET29-ΔN-EcMut5	This study
E. coli NiCo21:pET29-ΔN-EcMut8	Protein production strain, pET29-ΔN-EcMut8	This study
E. coli NiCo21:pET29-ΔN-EcMut11	Protein production strain, pET29-ΔN-EcMut11	This study
E. coli NR698:pJH10TS-EcMutQ	*E. coli b*ioassay strain, pJH10TS-EcMutQ	This study
E. coli NiCo21:pET28-EcMutQ	Protein production strain, pET28-EcMutQ	This study
E. coli NiCo21:pET29-ΔN-EcMutQ	Protein production strain, pET29-ΔN-EcMutQ	This study
E. coli NR698:pJH10TS-EcMutA	*E. coli b*ioassay strain, pJH10TS-EcMutA	This study
E. coli NiCo21:pET28-EcMutA	Protein production strain, pET28-EcMutA	This study
E. coli NiCo21:pET29-ΔN-EcMutA	Protein production strain, pET29-ΔN-EcMutA	This study
P. fluorescens ΔobaLΔobaO:pJH10TS-PfThrRS	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-PfThrRS	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-PfThrRS	*E. coli b*ioassay strain, pJH10TS-PfThrRS	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-BmObaO	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-BmObaO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-BmObaO	*E. coli b*ioassay strain, pJH10TS-BmObaO	Dr Sibyl Batey (JIC)

P. fluorescens ΔobaLΔobaO:pJH10TS-ΔN-BmObaO	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-ΔN-BmObaO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-ΔN-BmObaO	*E. coli b*ioassay strain, pJH10TS-ΔN-BmObaO	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-CsObaO	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-CsObaO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-CsObaO	*E. coli b*ioassay strain, pJH10TS-CsObaO	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-ΔN-CsObaO	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-ΔN-CsObaO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-ΔN-CsObaO	*E. coli b*ioassay strain, pJH10TS-ΔN-CsObaO	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-PIObaO	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-PIObaO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-PIObaO	*E. coli b*ioassay strain, pJH10TS-PIObaO	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-ΔN-PIObaO	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-ΔN-PIObaO	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-ΔN-PIObaO	*E. coli b*ioassay strain, pJH10TS-ΔN-PIObaO	Dr Sibyl Batey (JIC)
E. coli NiCo21:pET28-ObaO	Protein production strain, pET28-ObaO	This study
E. coli NiCo21:pET29-ΔN-ObaO	Protein production strain, pET29-ΔN-ObaO	This study
E. coli NiCo21:pET28-BmObaO	Protein production strain, pET28-BmObaO	This study
E. coli NiCo21:pET29-ΔN-BmObaO	Protein production strain, pET29-ΔN-BmObaO	This study
E. coli NiCo21:pET28-CsObaO	Protein production strain, pET28-CsObaO	This study

E. coli NiCo21:pET29-ΔN-CsObaO	Protein production strain, pET29-ΔN-CsObaO	This study
E. coli NiCo21:pET28-PIObaO	Protein production strain, pET28-PIObaO	This study
E. coli NiCo21:pET29-ΔN-PIObaO	Protein production strain, pET29-ΔN-PIObaO	This study
E. coli NR698:pJH10TS-EcMutM	*E. coli b*ioassay strain, pJH10TS-EcMutM	Dr Sibyl Batey (JIC)
E. coli NiCo21:pET28-EcMutM	Protein production strain, pET28-EcMutM	This study
E. coli NiCo21:pET29-ΔN-EcMutM	Protein production strain, pET29-ΔN-EcMutM	This study
E. coli NR698:pJH10TS-BmMutL	*E. coli b*ioassay strain, pJH10TS-BmMutL	This study
E. coli NiCo21:pET28-BmMutL	Protein production strain, pET28-BmMutL	This study
E. coli NiCo21:pET29-BmMutL	Protein production strain, pET29-BmMutL	This study
P. fluorescens ΔobaLΔobaO:pJH10TS-Chimera1	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-Chimera1	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-Chimera2	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-Chimera2	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-Chimera3	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-Chimera3	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-Chimera4	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-Chimera4	Dr Sibyl Batey (JIC)
P. fluorescens ΔobaLΔobaO:pJH10TS-Chimera5	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-Chimera5	Dr Sibyl Batey (JIC)

P. fluorescens *ΔobaLΔobaO*:pJH10TS-Chimera6	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-Chimera6	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-EcC1	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-EcC1	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-EcC2	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-EcC2	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-EcC3	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-EcC3	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-EcC4	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-EcC4	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-EcC5	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-EcC5	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-ObC1	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-ObC1	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-ObC2	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-ObC2	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-ObC3	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-ObC3	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-ObC4	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-ObC4	Dr Sibyl Batey (JIC)
P. fluorescens *ΔobaLΔobaO*:pJH10TS-ObC5	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-ObC5	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-Ec305	*E. coli b*ioassay strain, pJH10TS-Ec305	This study

E. coli NR698:pJH10TS-Ob463	*E. coli b*ioassay strain, pJH10TS-Ob463	This study
P. fluorescens ΔobaLΔobaO:pJH10TS-Ec463	*P. fluorescens* bioassay strain, *ΔobaL, ΔobaO,* pJH10TS-Ec463	Dr Sibyl Batey (JIC)
E. coli NR698:pJH10TS-MpgeneO	*E. coli b*ioassay strain, pJH10TS-MpgeneO	This study
E. coli NiCo21:pET28-MpgeneO	Protein production strain, pET28-MpgeneO	This study
E. coli NiCo21: pET29-ΔN-MpgeneO	Protein production strain, pET29-ΔN-MpgeneO	This study

2.5 Bioactivity Assays

2.5.1 Liquid Bioactivity Assays

Aliquots of 10 mL fresh media (LB for *E. coli* and *B. subtilis*, TSB for *S. aureus*) were inoculated 1% with 100 µL of overnight cultures, generated as described in section 2.3, with the exception that *S. aureus* grown in TSB instead of LB. These cultures incubated at 37 °C and 250 rpm shaking until they reached an OD_{600} of 0.4-0.6. Cultures were then diluted to OD_{600} 0.07 using fresh media, and then further diluted 1:10. A 96-well plate was set up with 95 µL media per well, with 5 µL 20x concentrated compound stock or DMSO then added, as described in Table 2.4. To each well except for the media control wells, 100 µL of diluted culture were added, with 100 µL sterile media added for media controls. The plate was incubated for 20 hrs at 37 °C with 125 rpm shaking at 80% humidity. The OD_{600} was then measured using a BMG LabTECH CLARIOstar^{Plus} plate reader. For resazurin assays, 5 µL 6.75 mg/mL resazurin was added to each well, incubated for 4 hrs static at room temperature and then imaged.

Table 2.4. Layout for liquid bioactivity assays. Blue wells were filled with sterile water to aid in humidity control; orange wells contain bioindicator strain 1, green wells contain bioindicator strain 2. Numbers represent the final concentration of borrelidin in each well in µg/mL. PC is the positive control with 50 µg/mL kanamycin final concentration; NC is the negative control with DMSO added to match the concentration due to dissolution of borrelidin in DMSO; MC is the media control which is not inoculated with bacteria.

H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O
256	128	64	32	16	8	4	2	1	PC	NC	MC
256	128	64	32	16	8	4	2	1	PC	NC	MC
256	128	64	32	16	8	4	2	1	PC	NC	MC
H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O	H_2O

256	128	64	32	16	8	4	2	1	PC	NC	MC
256	128	64	32	16	8	4	2	1	PC	NC	MC
256	128	64	32	16	8	4	2	1	PC	NC	MC

2.5.2 Plate-based Bioactivity Assays

2.5.2.1 E. coli spot on lawn bioactivity assays

E. coli NR698 plate-based bioactivity assays were performed in soft nutrient agar (SNA). Universals containing 5 mL of LB with selection where appropriate was inoculated 1% from an overnight culture (generated as described in section 2.3). These were incubated at 37°C and 200 rpm shaking until they reached an OD_{600} of 0.3 - 0.4. This was then used to inoculate molten SNA with no selection at 60°C by adding 5 mL of culture per 50 mL SNA; 25 mL of the resulting agar was poured into each of 2 100 mm square plates (Thermo Scientific) and incubating at room temperature for 30 mins for the agar to set. For each concentration of borrelidin or obafluorin (1000, 256, 128, 64, 32, 16, 8, 4, 2, 1 µg/mL), 4 µL were spotted onto the plate and allowed to dry completely before incubation at room temperature for 16 hours. A negative control of DMSO or acetonitrile (MeCN) and a positive control of 50 µg/mL were used for each plate.

2.5.2.2 S. venezuelae spot on lawn bioactivity assays

Streptomyces venezuelae-based bioactivity assays were performed on MYM. For each 50 mL square MYM plate, 10 µL *S. venezuelae* spores were spread for confluence, allowed to dry for 15 minutes and then for each borrelidin or obafluorin concentration, 4 µL were spotted onto the plate and allowed to dry completely before incubation at 30°C for 3-4 days, taking daily pictures.

2.5.2.3 Micromonospora sp. KC207 overlay bioactivity assays

For overlay bioassays with *M.* sp. KC207, streaks of mycelia were plated on the centre of a 50 mL plate of the media to be tested and incubated at 30°C for either 1 or 2 weeks, until orange mycelia were well developed. Universals containing 5 mL of LB with no selection was inoculated to 1% from an overnight culture of the bioindicator strain to be tested (generated as described in section 2.3). These were incubated at 37°C with 200 rpm shaking until they reached an OD_{600} of 0.3 - 0.4. This was then used to inoculate molten SNA with no selection at 60°C by adding 5 mL of culture per 50 mL SNA; 25 mL of the resulting agar was poured over each plate of *Micromonospora*, allowed to set for 30 minutes and then incubated at 30°C overnight, before being imaged.

2.6 Feeding Experiments

2.6.1 Feeding experiments using Streptomyces parvulus strains

Ten microlitres of *S. parvulus* spores were inoculated into 30 mL of NGY media in 250 mL conical flasks and incubated at 30°C and 250 rpm for 2 days. These seed cultures were then used to

inoculate 1% into 30 mL of PYDG media in a 250 mL conical flask. *Trans*-cyclopentane 1,2-dicarboxylic acid (tCPDA) was added to give a final concentration of 1 mM (by addition of 30 µL of a 1 M solution in MeOH) on days 0 and 1. The resulting flasks were incubated at 30°C and 250 rpm shaking. After 5 days, 1 mL of culture was taken from each flask, acidified with 50 µL of 98-100% HPLC grade formic acid before extracting with 1 mL of ethyl acetate by shaking for 30 mins. The resulting mixture was clarified by centrifugation at 13000 rpm for 10 mins in a VWR Microstar 17 microcentrifuge. The ethyl acetate layer was removed and the solvent removed under reduced pressure using a Genevac EZ-2 on the low boiling point setting. The residue was resuspended in 200 µL methanol and analysed by HPLC-MS using an Agilent 1260 system fitted with a Phenomenex ® Kinetex® 5µm XB-C18 100Å (100 × 4.6 mm) column. Chromatography was achieved using gradient elution and the eluant monitored by UV absorbance at 210, 258, 277, 365 and 417 nm, in addition to Evaporative Light Scattering Detector (ELSD), and mass spectrometry in both positive and negative mode. The injection volume was 10 µL, and the flow rate 1 mL/min. Mobile phase A: water + 0.1% formic acid (FA); mobile phase B: MeCN + 0.1% FA. Elution gradient: T=0 min, 10%B; T=1 min, 10%B; T=11 min, 95%B; T=13 min, 95%B; T=13.1 min, 10%B; T=15 min, 10%B. Borrelidin production was identified by the presence of an ion with *m/z* of 488 in the negative mode and absorbances at 210, 258, 277 and 365 nm. Borrelidin production was quantified by integration of the 258 nm absorbance peak, and comparison to a calibration curve made with authentic borrelidin.

2.6.2 Feeding experiments using Pseudomonas fluorescens strains

Stocks of *P. fluorescens* were streaked onto LB agar plates supplemented with nitrofurantoin (and tetracycline if strains have pJH10TS construct, at concentrations listed in section 2.3) and incubated for 2 days at 30°C. Single colonies were used to inoculate 50 mL OPM in 250 mL conical flasks and incubated at 30°C and 200 rpm shaking. After 24 h this culture was used to inoculate (1%) 100 mL OPM in 500 mL conical flasks. Feeding was achieved by addition of 200 µL of a 100 mM 2,3-dihyrdoxy benzoic acid (2,3-DHBA) solution in DMSO (or 200 µL DMSO as control) and the resulting flasks incubated at 30°C and 200 rpm shaking. After 14 h images were taken.

2.6.2.1 Generation of P. fluorescens spontaneous obafluorin mutants

A feeding experiment with 2,3-DHBA added to growing cultures of the *P. fluorescens* Δ*obaO*Δ*obaL* mutant was set up as described in Section 2.6.2. However, the experiment was not terminated at 14 h and the cultures were allowed to grow to 72 h. All cultures which had begun to grow and turn purple were samples and streaked for single colonies on OPM agar containing 0.2 mM 2,3-DHBA and incubated for 4 days at 30°C. Single colonies which had turned purple were then streaked out on the same agar plus 2,3-DHBA, and in parallel used to inoculate 10 mL of liquid LB overnight at 30°C and 250 rpm shaking. These overnight cultures were then used as the seed culture for a repeat

feeding experiment as described in section 2.6.2. Any cultures which developed a purple colour at 14 h were then glycerol stocked and gDNA extracted for amplification and sequencing using the methods described in 2.4.3.

2.7 Light microscopy of S. parvulus sporulation in liquid culture

Cultures of WT, Δ*borG*, Δ*borO* and Δ*borG*Δ*borO S. parvulus* were grown at 30°C in production media from seed culture without feeding as described section 2.6.1. Flasks were inoculated in a staggered manner such that cultures which had incubated for each of 1, 2, 3, 4 and 5 days were ready at the same time. Aliquots of each culture (2 μL) were spotted onto a 1% agarose pad, allowed to dry for 10 min, and then covered with a glass coverslip. Differential interference contrast (DIC) images were taken of each sample using a Zeiss Axio Observer Z.1 inverted epifluorescence microscope fitted with a Zeiss Colibri 7 LED light source and a Zeiss Alpha Plan-Apo 100×/1.46 Oil DIC M27 objective. Images were analysed using Fiji[193].

2.8 Protein Purification

Fresh *E. coli* NiCo21 cells were transformed for protein expression as described in section 2.4.5.1. Ten millilitre LB overnight cultures were inoculated from a streak of multiple cells from the transformants to account for heterogeneity of protein production between different individuals. Liquid TB media (1 L) in a 2 L baffled conical flask was inoculated with 10 mL of the overnight culture (1%) and incubated at 37°C with 200 rpm shaking. When the culture had grown to an OD_{600} 0.6-0.8, protein expression was induced by the addition of IPTG at a final concentration of 0.1 mM and shaken at 200 rpm and 18°C for a further 16 h. The cells were harvested by centrifugation at 4000 rpm for 10 min and 4°C using a Sorvall Lynx 4000 centrifuge and the pellets stored at -80°C until needed.

For protein purification, cells were defrosted in lysis buffer, sonicated with a Sonics Vibracell sonicator for 10 min at 40% amplitude in a cycle of 10 sec on, 10 sec off, while on ice. The insoluble fraction and cell debris was removed by centrifugation at 16,000 rpm in a Sorvall Lynx 4000 centrifuge for 1 h at 4°C. The supernatant was collected and applied to ca. 3 g chitin resin (NEB) in a manual chromatography column, preequilibrated in lysis buffer. This was incubated with rotation at 4°C for 15 min before collection of the flow-through under gravity. The flowthrough from the chitin column was then applied to a 1 mL HisTrap excel (Cytiva) nickel column fitted to an Äkta Pure Protein Purification System (Cytiva) using an automated method for sample application. The column was washed with lysis buffer for 5 column volumes (CVs) before washes with 5 CVs each of 4%, 8%, 12% and 20% elution buffer. Target proteins were then eluted with 10 CVs of 100% elution buffer and captured into a 10mL loop by tracking the UV absorbance at 280 nm with a threshold of 110 mAU and collecting the 5mL with the greatest UV peak. This captured fraction was then applied to a HiLoad 16/600 Superdex 200 pg gel filtration column with isocratic elution with lysis buffer at

1 mL/min for 1.2 CVs. Peaks with a UV absorbance at 280 nm of greater than 10 mAU were collected in 2 mL fractions and analysed for protein purity using SDS-PAGE gels. Fractions containing proteins of the required size and purity were pooled and concentrated using Amicon spin columns at 5000xg in 10 minute increments using a Sigma 4-16KS centrifuge; 30 kDa cut-off columns were used for full-length protein, and 10 kDa cut-off columns for ΔN proteins. Protein samples were flash frozen in liquid nitrogen and stored at -80°C if not used immediately.

2.8.1 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Aliquots of 10 µL of protein sample were prepared by addition of 10 µL 2 x protein loading dye, incubation at 98°C 10 mins and loading of 10 µL per lane of gel. A 5 µL aliquot of colour protein standard (NEB) was used as a ladder. Samples were applied to Novex 12% Tris-Glycine Precast Gels (Thermo Fisher Scientific) in an Invitrogen Mini Gel Tank filled with Novex Tris-Glycine SDS Running Buffer and gels were ran at 150 V for 60 min. Gels were stained with InstantBlue Coomassie Protein Stain (Abcam) unless being used for Western blotting.

2.8.1.1 Protein Expression/Solubility Testing

The solubility and expression level of new proteins were tested in selected lysis buffers. To do this 50 mL of LB media was inoculated with 500 µL of an overnight culture (prepared as described in section 2.3) and incubated at 37°C and 200 rpm shaking until it reached an OD_{600} of 0.6-0.8. Protein expression was then induced by the addition of IPTG to a final concentration of 0.1 mM and the flasks were incubated at 18°C with 200 rpm shaking. After 16 h and aliquots of 1 mL was pipetted into a 2 mL Eppendorf tube and the cells harvested by centrifugation at 13000 rpm for 5 min at 4°C. The supernatant was discarded and the cells were resuspended in 1 mL of the lysis buffer to be tested. Cells were lysed via sonication with a MSE Soniprep 150 (Henderson Biomedical) with 3x30 sec pulses on ice, with 30 sec in between pulses. Soluble and insoluble fractions were then separated by centrifugation at 13000 rpm in a VWR Microstar 17 microcentrifuge for 10 min at 4°C. The supernatant was transferred to a fresh 1.5 mL Eppendorf tube and the pellet resuspended in 1mL deionised water. Aliquots of 10 µL were then analysed using SDS-PAGE as described in section 2.8.1.

2.8.2 Bradford Assay

Bradford assays were used to estimate protein concentrations using the Quick Start™ Bradford Protein Assay (Bio-Rad) as per the instruction manual, and in a 96 well plate format. A calibration curve of BSA at 1, 0.5, 0.25 and 0.125 mg/mL was used, and protein was measured neat in addition to 1:10 and 1:100 dilutions into the lysis buffer they were prepared in.

2.8.3 Western Blotting

Strains expressing single N-terminally FLAG-tagged (DYKDDDDK) protein were grown in 10 mL LB overnight cultures 37°C as described in section 2.3. From these cultures, 10 µL samples were taken

and run on SDS-PAGE as in section 2.8.1, but without staining. Proteins were transferred to a nitrocellulose membrane (Pall Corporation) using a Bio-Rad Trans-Blot Turbo transfer cell (10 V) in transfer buffer. After 1 h the membrane was rinsed with TBST then blocked with 5% (w/v) fat-free skimmed milk powder (Tesco) in TBST for 2 h at room temperature. Horseradish Peroxidase (HRP)-conjugated anti-FLAG antibody (Merck) diluted 1 in 20,000 in TBST was added to the membrane and incubated for 1 h, followed by 3x10 min washes with TBST before finally staining by adding 10 mL each of developing solutions A and B and allowing to develop at room temperature for precisely 1 minute. Imaging was achieved using an Amersham ImageQuant™ 500 CCD imaging system (Cytiva).

2.8.4 Protein Mass Spectrometry

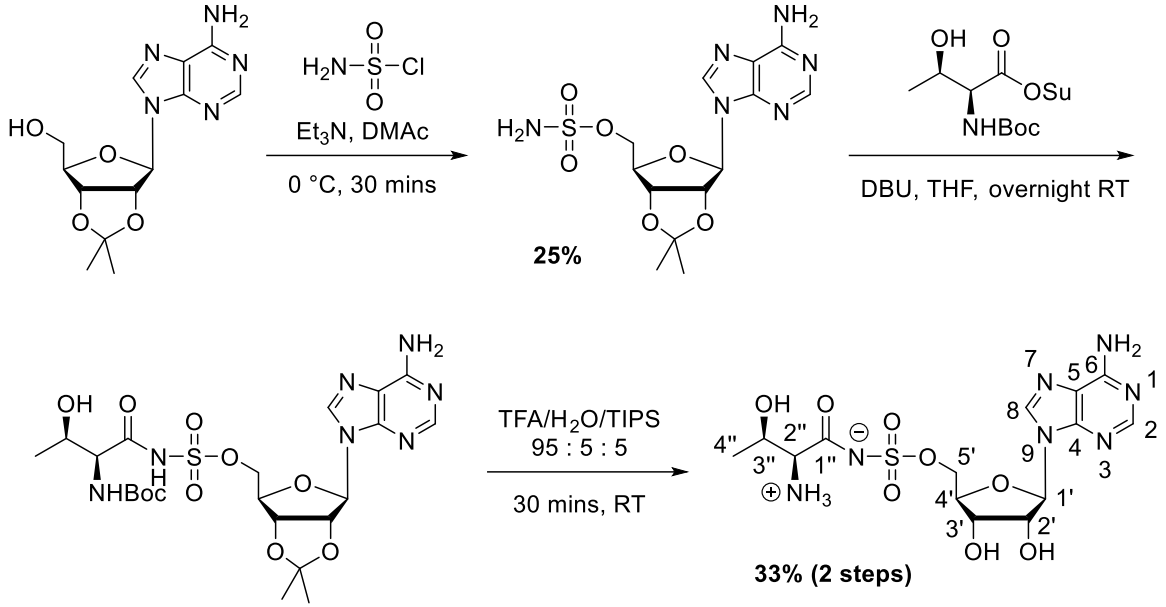
Where required, the identity of protein samples was verified via tryptic digest and MALDI-TOF mass spectrometry. To do this, the target protein was run on an SDS-PAGE gel as in section 2.8.1, with minimal Coomassie staining so that protein bands were just visible. Gel slices of the relevant protein bands were taken with a clean scalpel and prepared as described in Petre *et al.* 2017[194]. Subsequent tryptic digest and HRMS analysis was performed by Dr Carlo de-Oliveira Martins from the JIC proteomics facility. Proteins were digested using Trypsin (Promega) in a 1:20 trypsin:protein ratio for 8 h at room temperature in 50 mM ammonium bicarbonate, 5% (v/v) MeCN at pH 7.5. LC-MS/MS analysis was then run on an Orbitrap Eclipse tribrid mass spectrometer (Thermo Fisher Scientific) fitted with a nanoflow HPLC system (Dionex Ultimat3000, Thermo Fisher Scientific) as described in Petre *et al.* 2017[194].

2.9 Extraction and purification of Obafluorin

This study was performed by Dr Sibyl Batey, Dr Edward Hems and Melissa Davie (JIC). A 500 mL conical flask containing 100 mL of OPM was inoculated with a single colony of WT *P. fluorescens* ATCC 39502 and incubated at 25°C and 300 rpm shaking for 24 h. This seed culture was used to inoculate (1%) each of 16 x 500 mL OPM each in 2 L flasks and incubated at 25°C and 250 rpm shaking for 14 h. The resulting cultures were shaken vigorously with ethyl acetate (500 mL per culture), allowed then to separate for 2 h. The organic phase was separated and solvent removed under reduced pressure. The residue was dissolved in 1:1 MeCN:water (total volume 4 mL) and purified by flash chromatography on a Biotage Isolera One system, with a flow rate of 25 mL/min, and monitoring at wavelengths of 254 and 270 nm (Biotage SNAP 30 g C18 cartridge) with a mobile phase A of water and mobile phase B of MeCN. Gradient elution, with flow measured in column volumes (CVs); 0 CVs, 0%B; 4.5 CVs, 0%B; 5.5 CVs, 30%B; 13 CVs, 80%B; 13.5 CVs, 100%B; 15.5 CVs, 100%B. The fractions with UV absorbance corresponding to obafluorin were combined and the solvent removed under reduced pressure. Where necessary this material as further purified via preparative HPLC using a Dionex Ultimate 3000 HPLC system fitted with a Gemini-NX C18 column

(150 × 21.2 mm, 5.0 µm particle size). Flow rate 20 mL/min; mobile phase A, water; mobile phase B, MeCN; gradient elution: T=0 min, 5%B; T=2 min, 5%B; T=4 min, 45%B; T=12 min, 80%B; T=14 min, 80%B; T=15 min, 5%B. UV absorbance was monitored at 276 nm.

2.10 Synthesis of 5'-O-(threonylsulfamoyl)adenosine (ThrSAA)



Scheme 2.1. Chemical synthesis of ThrSAA, with atom numbering scheme for NMR assignments.

2',3'-*O*-isopropylideneadenosine and sulfamoyl chloride were purchased from Fluorochem (UK). All other solvents and reagents were purchased from Merck (UK). Wash and HPLC solvents were purchased from Fischer Scientific. Reactions were performed in oven dried glassware under an atmosphere of dry argon. Low resolution mass spectrometry (LRMS) was measured with an Advion Expression CMS spectrometer. Optical rotation values were measured using a Perkin Elmer®Model341 Polarimeter at 20 °C at a wavelength of 589 nm. NMR spectra were recorder on a Bruker Avance NEO 600 MHz spectrometer fitted with a cryoprobe (NEO600) operating at 297 K. NMR solvents were purchased from Merck (UK). Assignments were made with the aid of 2D COSY and HSQC experiments.

For high resolution mass spectrometry (HRMS) 0.1% (w/v) samples were prepared by diluting in methanol-0.1% formic acid (1:1) and infused into a Synapt G2-Si mass spectrometer (Waters, Manchester, UK) at 5–10 L/mi^{n} using a Harvard Apparatus syringe pump. The mass spectrometer was controlled by MassLynx 4.1 software (Waters). It was operated in high resolution and positive ion mode and calibrated using sodium formate. The sample was analysed for 2 min with 1 s MS scan time over the range of m/z 50–1200 with 2.5 kV capillary voltage, 40 V cone voltage, 350 °C cone temperature. Leu-enkephalin peptide (1 ng/µL, Waters) was infused at 10 L/min as a lock mass (m/z 556.2766) and measured every 10 s. The cone gas flow was set at 20 L/h, desolvation gas flow

at 700 L/h and nebuliser gas flow at 6 bar. Spectra were generated in MassLynx 4.1 by combining several scans, and peaks were centred using automatic peak detection with lock mass correction.

The first step of compound synthesis was performed by Dr Edward Hems (JIC). Triethylamine (Et_3N, 980 µL, 7.0 mmol) was added to a solution of 2',3'-*O*-isopropylideneadenosine (1.78 g, 5.8 mmol) in anhydrous dimethylacetamide (DMAc, 10 mL) and stirred under argon at 0 °C for 30 min. Sulfamoyl chloride (1.0 g, 8.7 mmol) was added slowly, and the resulting mixture stirred overnight at room temperature. Evaporation of DMAc was performed under reduced pressure and the crude residue was purified by flash C18 chromatography on a Biotage Isolera one system, with a flow rate of 25 mL/min, monitoring at wavelengths 210 and 254 nm (Biotage SNAP 30 g C18 cartridge, H_2O/MeCN, 1-50% over 6.5 column volumes). The fractions identified as containing the desired product by LRMS were combined and dried under reduced pressure to give 2',3'-*O*-(1-methylethylidene)-5'-*O*-sulfamoyladenosine (890 mg, 25%) as a white powder; ^{1}HNMR (DMSO-d_6, 600 MHz) 8.30 (s, 1H, H-8), 8.17 (s, 1H, H-2), 7.59 (bs, 2H, NH_2), 7.36 (bs, 2H, NH_2), 6.24 (d, $J_{1',2'}$ = 2.4 Hz, 1H, H-1'), 5.43 (dd, $J_{1',2'}$ = 2.4 Hz, $J_{2',3'}$ = 6.4 Hz, 1H, H-2'), 5.08 (dd, $J_{2',3'}$ = 6.4 Hz, $J_{3',4'}$ = 3.3 Hz, 1H, H-3'), 4.42-4.38 (m, 1H, H-4'), 4.24 (dd, $J_{4',5'a}$ = 5.4 Hz, $J_{5'a,5'b}$ = 10.7 Hz, 1H, H-5'a), 4.13 (dd, $J_{4',5'b}$ = 6.5 Hz, $J_{5'a,5'b}$ = 10.7 Hz, H-5'b), 1.55 (s, 3H, Me), 1.34 (s, 3H, Me); ^{13}CNMR (DMSO-d_6, 151 MHz), 156.6 (C6), 153.3 (C2), 149.2 (C4), 140.2 (C8), 119.6 (C5), 114.0 ($\underline{C}Me_2$), 89.6 (C1'), 84.1 (C4'), 83.8 (C2'), 81.6 (C3'), 68.5 (C5'), 27.4 (Me), 25.6 (Me); HRMS calc. for $C_{13}H_{19}N_6O_6S$ ($[M.H]^+$) 387.1087, found 387.1077 ($[M.H]^+$), Δ -2.6 ppm. The NMR spectra agreed with published data[2].

The next step of compound synthesis was done at two separate times, once by Dr Edward Hems, and once by myself. Boc-Thr-OSu (100 mg, 0.32 mmol) was added to a solution of 2',3'-*O*-(1-methylethylidene)-5'-*O*-sulfamoyladenosine (100 mg, 0.25 mmol) and 1,8-diazabiocyclo[5.4.0]undec-7-ene (DBU, 41 mg, 0.28 mmol) in anhydrous tetrahydrofuran (THF, 5 mL). The reaction mixture was stirred overnight under argon at room temperature. THF was removed under reduced pressure and the residue was dissolved into a mixture of trifluoroacetic acid (TFA)/triisopropylsilane (TIPS)/water (95:5:5, 10 mL) and stirred at room temperature for 30 min. The volatile components were removed under reduced pressure and the residue was purified by preparative-HPLC using a Dionex Ultimate 3000 HPLC system fitted with a Gemini-NX C18 column (150 × 21.2 mm, 5.0 µm particle size). Flow rate 20 mL/min; mobile phase A, water; mobile phase B, methanol; gradient elution: T=0 min, 2%B; T=3.5 min, 2%B; T=4 min, 6%B; T=14 min, 15%B; T=16 min, 80%B; T=17 min, 2%B; T = 17.5, 2%B; monitoring UV absorbance at wavelength 254nm. 5'-*O*-(threonylsulfamoyl)adenosine (47.0 mg, 33% over 2 steps) was isolated as a white powder; R_f 10.25 minutes (prep-HPLC gradient); $[\alpha]_D$ -2.7 (*c* 2.0, H_2O); ^{1}HNMR (DMSO-d_6, 600 MHz), 8.42 (s, 1H, H-8), 8.19 (s, 1H, H-2), 7.75 (bs, 3H, 2''-$N\underline{H_3}^+$), 7.49 (bs, 2H, Ad-$N\underline{H_2}$), 5.93 (d, $J_{1',2'}$ = 5.8 Hz, 1H, H-1'), 5.53 (bs, 1H, 2'-O$\underline{H}$), 5.41-5.14 (m, 2H, 3'-O$\underline{H}$ & 3''-O$\underline{H}$), 4.59 (dd, $J_{1',2'}$ = 5.8 Hz, $J_{2',3'}$ = 5.8 Hz, 1H, H-2'),

4.20-4.14 (m, 2H, H-3' & 5'a), 4.13-4.06 (m, 2H, H-4' & 5'b), 3.99-3.93 (m, 1H, H-3''), 3.22-3.17 (m, 1H, H-2''), 1.19 (d, $J_{3'',4''}$ = 6.4 Hz, 3H, H-4''); ^{13}CNMR (DMSO-d_6, 151 MHz), 171.2 (C1''), 155.8 (C6), 152.3 (C2), 149.9 (C4), 140.1 (C8), 119.3 (C5), 87.6 (C1'), 82.9 (C4'), 74.0 (C2'), 71.1 (C3'), 68.0 (C5'), 66.2 (C3''), 61.5 (C2''), 21.4 (C4''); HRMS ESI$^+$ calc. for $C_{14}H_{22}N_7O_8S$ ([M.H]$^+$) 448.1251, found 448.1240 ([M.H]$^+$), Δ -2.5 ppm. 5'-O-sulfamoyladenosine was then stored dry in ~1mg pre-weighed aliquots under nitrogen at -80°C to prevent compound degradation and resuspended in the required volume of protein buffer appropriate for its use.

2.11 Biophysical Analysis

2.11.1 Isothermal Titration Calorimetry (ITC)

For ITC studies, full length protein was used; this was purified in the absence of L-threonine. Protein samples were concentrated as much as possible without aggregation occurring before dialysis. To set up the experiments, 1.5 mL Eppendorf tubes were cut in half, 250 µL protein sample placed in the lid of the Eppendorf, a piece of SnakeSkin Dialysis Tubing (Thermo Fisher Scientific) placed over the sample and the lid closed, securing the dialysis membrane in place with the protein held in the lid with no air between the tubing and the sample. This apparatus was then dialysed in ITC buffer at 4°C overnight. Sample could be retrieved without opening the lid by extracting with a needle through the dialysis membrane. The protein concentration was then estimated using a Bradford Assay.

Protein samples were diluted to 20 µM with ITC buffer and loaded into the sample well of the MicroCal PEAQ-ITC apparatus (Malvern Panalytical). A sample of the analyte (200 µM in ITC buffer) was loaded into the apparatus syringe and a single 0.4 µL injection was followed by 18 x 2 µL injections. Binding was measured at 25°C with reference power 5 µcal/s, high feedback, stir speed 750 rpm, initial delay 60 s, injection spacing 150 s, and injection duration 4 s. Raw data were analysed using AFFINImeter v2.1802.5: the first injection was omitted, and the stoichiometric equilibrium method with the simple binding model was used. Figures were generated from AFFINImeter outputs using RStudio.

2.11.2 Mass Photometry

Protein samples were diluted in PBS buffer to a concentration of 10 nM – 100 nM, and movies recorded using default settings for 1 min on a Refeyn mass photometer, analysing the size of droplets using the DiscoverMP software with a calibration of BSA standard in PBS.

2.12 Structural Biology

2.12.1 X-Ray Crystallography

In general, crystal trials were set up with an initial protein concentration of 10-20 mg/mL . For co-crystallisations with borrelidin or ThrSAA, 0.3 µL of 200 mM borrelidin in DMSO or ThrSAA in protein

Lysis Buffer was added to 30 µL protein. Upon addition of borrelidin, precipitation of borrelidin was observed, which was resuspended gently with a pipette and centrifuged briefly to remove air bubbles. Crystallisation trials were set up using an Oryx8 (Douglas Instruments) liquid handling robot to dispense two drops per well, in an MRC 2-Drop Plate (Douglas Instruments) in a sitting drop vapour diffusion format, with 0.3 µL protein and 0.3 µL well solution per drop. Crystal trials were incubated at 20°C until crystals of suitable size for harvesting were observed. For initial screening for crystallisation, the PEGs suite (Qiagen) was generally used. The XP screen (Jena Biosciences) was also used to expand the chemical space explored- this screen contains the Anderson-Evans polyoxotungstate which supposedly promotes protein crystallisation, as well as providing an anomalous signal for phasing[195]. Subsequently, BorO crystallisation was optimised to 7 mg/mL protein, with seeding with crystals from well G10 of the PEGs screen diluted 1:10 and added at a 1:2:3 seed:protein:well solution. An optimisation of crystallisation conditions was designed, as summarised in Table 2.5. SpThrRS crystallisation was optimised with the same crystallisation conditions, using a protein concentration of 10 mg/mL. BmObaO concentration was optimised to 4 mg/mL, with seeding diluted 1:10 from screen crystals. An optimisation of crystallisation conditions was designed, as summarised in Table 2.6. For EcThrRS and its mutants, a PEGs screen was used, with protein concentrations at 10-20 mg/mL. Polyethylene glycol (PEG) has n number of ethylene units, with the PEG rating describing this number (for example, PEG 3350 is made of 3350 ethylene units). Smaller PEGs (less than 4000) are effective cryoprotectant. Cryoprotection solutions were set up as close as possible to the well solution, but with 40% PEG 3350 where well solution had PEG larger than 4,000, or 40% of whichever PEG was in the well otherwise. Where cocrystals were set up, compound was added to cryoprotection solutions at a final concentration of 2 mM.

Crystals were harvested with cryoprotection solution using Litholoops (Molecular Dimensions) before flash-cooling by plunging into liquid nitrogen. X-Ray diffraction data were collected at the Diamond Light Source on beamline i04. Initial data process and scaling was done with DIALS, with data reduction and merging performed using AIMLESS. MR-PHASER was used for molecular replacement, using either an AlphaFold model (for initial BmObaO models), partially built models from previous data collections (for final BorO and BmObaO models), fully solved models (in the case of *E. coli* point mutants) or PDB file 1EVL[97] (for the initial BorO model and WT EcThrRS models). Further model building and refinement was done in WinCoot[196] using CCP4i2[197]. EcThrRS:ThrSAA data was solved as part of the CCP4:Diamond Light Source Data Collection and Structure Solution Workshop 2021.

Table 2.5. Optimisation plate design for BorO, concentration PEG 3350 (% v/v) in black, concentration K_2SO_4 (M) in blue, concentration Tris-HCl pH 8.5 (M) in green.

	1	2	3	4	5	6	7	8	9	10	11	12
A	5 0.1 0.05	7.8 0.1 0.05	10.4 0.1 0.05	13.2 0.1 0.05	16 0.1 0.05	18.8 0.1 0.05	21.4 0.1 0.05	24.1 0.1 0.05	26.9 0.1 0.05	29.7 0.1 0.05	32.3 0.1 0.05	35.1 0.1 0.05
B	5 0.14 0.05	7.8 0.14 0.05	10.4 0.14 0.05	13.2 0.14 0.05	16 0.14 0.05	18.8 0.14 0.05	21.4 0.14 0.05	24.1 0.14 0.05	26.9 0.14 0.05	29.7 0.14 0.05	32.3 0.14 0.05	33.9 0.14 0.05
C	5 0.19 0.05	7.8 0.19 0.05	10.4 0.19 0.05	13.2 0.19 0.05	16 0.19 0.05	18.8 0.19 0.05	21.4 0.19 0.05	24.1 0.19 0.05	26.9 0.19 0.05	29.2 0.18 0.05	30.3 0.18 0.05	31.3 0.17 0.05
D	5 0.23 0.05	7.8 0.23 0.05	10.4 0.23 0.05	13.2 0.23 0.05	16 0.23 0.05	18.8 0.23 0.05	21.4 0.23 0.05	24.1 0.23 0.05	25.8 0.22 0.05	27 0.2 0.05	28 0.2 0.05	29 0.19 0.05
E	5 0.27 0.05	7.8 0.27 0.05	10.4 0.27 0.05	13.2 0.27 0.05	16 0.27 0.05	18.8 0.27 0.05	21 0.27 0.05	22.5 0.25 0.05	23.8 0.24 0.05	25 0.23 0.05	26.2 0.22 0.05	27.1 0.21 0.05
F	5 0.31 0.05	7.8 0.31 0.05	10.4 0.31 0.05	13.2 0.31 0.05	16 0.31 0.05	17.8 0.3 0.05	19.4 0.28 0.05	20.8 0.27 0.05	22.1 0.26 0.05	23.3 0.25 0.05	24.5 0.24 0.05	25.4 0.23 0.05
G	5 0.36 0.05	7.8 0.36 0.05	10.4 0.36 0.05	12.9 0.35 0.05	14.7 0.33 0.05	16.4 0.31 0.05	18 0.3 0.05	19.4 0.29 0.05	20.7 0.28 0.05	21.9 0.26 0.05	22.9 0.25 0.05	24 0.24 0.05
H	5 0.4 0.05	7.7 0.4 0.05	9.9 0.38 0.05	11.8 0.36 0.05	13.7 0.34 0.05	15.2 0.33 0.05	16.8 0.31 0.05	18.8 0.3 0.05	19.4 0.29 0.05	20.6 0.28 0.05	21.6 0.27 0.05	22.7 0.26 0.05

Table 2.6. Optimisation plate design for BmObaO, concentration PEG 200MME (% v/v) in black, PEG 3350 (% v/v) in blue, concentration Tris-HCl pH 7.5 (M) in green, concentration Tris-HCl pH 8.5 in orange.

	1	2	3	4	5	6	7	8	9	10	11	12
A	5 0.1	7.7 0.1	10.4 0.1	13.2 0.1	15.9 0.1	18.7 0.1	21.4 0.1	24.1 0.1	26.1 0.1	29.6 0.1	32.3 0.1	35 0.1
B	5	7.7	10.4	13.2	15.9	18.7	21.4	24.1	26.1	29.6	32.3	35

	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
C	5 0.1	7.7 0.1	10.4 0.1	13.2 0.1	15.9 0.1	18.7 0.1	21.4 0.1	24.1 0.1	26.1 0.1	29.6 0.1	32.3 0.1	35 0.1
D	5 0.1	7.7 0.1	10.4 0.1	13.2 0.1	15.9 0.1	18.7 0.1	21.4 0.1	24.1 0.1	26.1 0.1	29.6 0.1	32.3 0.1	35 0.1
E	5 5 0.1	7.7 5 0.1	10.4 5 0.1	13.2 5 0.1	15.9 5 0.1	18.7 5 0.1	21.4 5 0.1	24.1 5 0.1	26.1 5 0.1	29.6 5 0.1	32.3 5 0.1	35 5 0.1
F	5 5 0.1	5 7.7 0.1	5 10.4 0.1	5 13.2 0.1	5 15.9 0.1	5 18.7 0.1	5 21.4 0.1	5 24.1 0.1	5 26.1 0.1	5 29.6 0.1	5 32.3 0.1	5 35 0.1
G	5 5 0.1	6.4 6.4 0.1	7.7 7.7 0.1	9.1 9.1 0.1	10.4 10.4 0.1	11.8 11.8 0.1	13.2 13.2 0.1	14.6 14.6 0.1	15.9 15.9 0.1	17.3 17.3 0.1	18.6 18.6 0.1	20.1 20.1 0.1
H	5 5 0.1	6.4 6.4 0.1	7.7 7.7 0.1	9.1 9.1 0.1	10.4 10.4 0.1	11.8 11.8 0.1	13.2 13.2 0.1	14.6 14.6 0.1	15.9 15.9 0.1	17.3 17.3 0.1	18.6 18.6 0.1	20.1 20.1 0.1

2.12.1.1 X-Ray Fluorescence

Where required, of crystals used for structure determination also had their X-Ray fluorescence spectra measured at Diamond Light Source in order to confirm the presence of the Zn ion in the protein.

2.12.2 Cryogenic Electron Microscopy (Cryo-EM)

Samples were dialysed overnight 4°C into cryo-EM buffer using the procedure described in sample preparation for ITC in section 2.11.1. Sample concentration was then estimated using a Bradford Assay and diluted in cryo-EM buffer so that EcThrRS was at 11 mg/mL (with 10 µM obafluorin) and ObaO at 7 mg/mL (with 10 µM obafluorin). These concentrations correspond to a roughly 10:1 molar ratio of protein to obafluorin. These samples were then flash frozen in liquid nitrogen, and shipped on dry ice to the Cryo-EM Team at the University of Leeds, where EM grids were prepared and screened. For EcThrRS, the sample was diluted 1:5 with Cryo-EM buffer before preparing the grid. 3 µL of sample was added to the 1.2/1.3 copper Quantifoil grid (300 mesh), incubated for 30 seconds, blotted and plunge-frozen using Vitrobor Mark IV (FEI) (blot force 5, blot time 8). Data were collected on Titan Krios microscope operated at 300 kV at the nominal magnification 96 000 x using physical pixel size of 0.86 Å with a range of defoci set as - 1.2, -1.5, -1.8, -2.1, -2.4, -2.7, -3 µm. Movies were collected in counting mode on Falcon IV detector (Thermo) using EPU version 3.

Total dose on vacuum was 8.64 electron/Å/s and exposure time was set to generate a total dose of ~34.90 electrons/Å^2. Data was saved in EER format.

For ObaO, sample was diluted 1:10 and 5 mM DTT was added to reduce sample aggregation; grids were prepared in the same manner as for EcThrRS. Data were collected on Titan Krios microscope operated at 300 kV at the nominal magnification 120 000 x resulting in a calibrated physical pixel size of 0.68 Å. Total dose was set as 34.84 electrons/Å^2. Range of defoci -0.9, -1.2, -1.5, -1.8, -2.1, -2.4, -2.7, -3 µm was set and movies were collected using Falcon IVi detector (Thermo Scientific). Data was saved in EER format.

Processing was done in cryoSPARC v3.3.2 by Dr Dmitry Ghilarov. For EcThrRS 7005 movies were motion and CTF corrected in patch mode. To create an initial template for particle picking, an AlphaFold model of EcThrRS was used to generate a 3D volume using the ChimeraX molmap command; using this template, particles were picked from 1000 randomly selected movies and analysed to create an initial model. This volume was used to create the final templates to pick the whole dataset. 2 608 768 particles were picked, extracted with 2-fold binning and subjected to rounds of 2D and 3D (Ab initio) classification yielding 254 054 particles with high resolution features refining to Nyquist (3.55 Å). These were re-extracted unbinned with centring using 280 px box size and re-refined with local defocus and CTF correction (Non-uniform refinement in cryoSPARC). Refinement with C2 symmetry imposed resulted in 2.85 Å global resolution map as estimated by cryoSPARC using 0.143 FSC criterion.

For the ObaO, 7000 movies were collected and 6214 selected for further processing. 2 667 556 particles were picked using template picker and 2x binned images were subjected to rounds of 2D and 3D classification. Non-uniform refinement of the best final set of 224 636 particles (still 2x binned to improve the SNR) with per-particle defocus and CTF correction produced 2.95 Å map used for model real-space refinement. To assist model building, this map was locally filtered using DeepEMhancer[198].

Model building was then done by me. Initial models were built in WinCoot[196] and refined in real-space against a sharpened map using Ramachandran restraints in Phenix version 1.20.1-4487. For the ligands, SMILES strings describing the structures of obafluorin and obafluorin with a hydrolysed β-lactone ring were generated in ChemDraw Prime 16.0 and restraints were generated using the Grade webserver (Global Phasing Limited) at: http://grade.globalphasing.org/cgi-bin/grade/server.cgi. Restraints for the obafluorin-tyrosine and zinc—obafluorin linkages were produced manually using Phenix scripts. Model geometry was optimised in ISOLDE[199] (in UCSF ChimeraX[114]) followed by final refinement in Phenix.

2.13.2.1 Negative Staining Electron Microscopy

To pre-screen samples for Cryo-EM, samples of EcThrRS were prepared as for Cryo-EM without the addition of obafluorin. Grids were prepared at 1 in 99 dilutions by Jake Richardson from the Bioimaging Facility at JIC and stained with uranyl acetate. 435 TIFF images we collected using EPU at 45000x magnification at a FEI Talos F200C 200 KV transmission electron microscope equipped with the Gatan OneView 4k by 4k CMOS bottom-mounted camera and analysed in cryoSPARC v3.3.2 using the negative stain images pipeline.

2.13 Bioinformatic Analysis

2.13.1 Multiple Sequence Alignments

To identify homologs/paralogs, the protein Basic Local Alignment Search Tool (BLASTp) was used with default settings. Multiple sequence alignments were generated using MUSCLE[200,201] and manually trimmed as needed using MEGA X[202].

2.13.2 Construction of Phylogenetic Trees

Trimmed multiple sequence alignments were used to construct Maximum Likelihood phylogenetic tree in CIPRES[203] using RAxML[204] with 300 bootstraps. Trees were then visualised with the interactive tree of life (iTOL)[205].

2.13.3 Analysis of Biosynthetic Gene Clusters

Biosynthetic gene clusters were identified using antiSMASH bacterial version 6.0[5,206-210] with relaxed detection strictness and all extra features on.

2.13.4 Generation of Alphafold2 Models

The collabfold2 server was used to generate initial Alphafold2 models for cryo-EM, search models for molecular replacement for X-ray crystallography, and for structural analysis in the absence of structural data. Generally, models were generated as monomers using Amber to perform energy minimisation so that side chains are in chemically relevant orientations; otherwise, default settings were used. Generally, the top ranked structure was used for further analysis.

Chapter 3: Borrelidin resistance mechanisms in the producer, *Streptomyces parvulus*

3.1 Introduction

Although the biosynthesis of borrelidin has been being largely characterised (seen in Figure 3.1) it remains an actively investigated potential drug molecule. Importantly, Borrelidin has been shown to have a wide bioactivity profile, being potent against multiple different genera of bacteria, archaea and eukaryotes and multiple mammalian cell types.[110,154-161,168,211-220] During the biosynthesis of borrelidin, cytosolic protein BorG has been shown to be essential for the biosynthesis of the starter unit of the PKS, *trans*-cyclopentane 1,2 di-carboxylic acid (tCPDA). Cells lacking BorG (*ΔborG*) have restored borrelidin production when fed with exogenous tCPDA; interestingly a fed *ΔborG* strain is able to produce higher quantities than the wild-type strain[170], likely due to tCPDA production being a bottleneck during borrelidin biosynthesis.

Several different *Streptomyces* species are able to produce borrelidin, as has a species from the closely related genus *Nocardiopsis*. From these strains, while borrelidin A is the most prevalent congener, and other congeners have been isolated from a marine sediment derived *Streptomyces* sp. (borrelidin B)[214], a Saltern-derived *Nocardiopsis* sp. (borrelidin C-E)[217] and a mangrove-derived *Streptomyces rochei* (borrelidin F-I)[218] (see Figure 3.2). These allow for a rudimentary structure-activity relationship to be performed; most congeners were less potent than borrelidin A against both bacteria and cancer cell lines; however, borrelidin H appeared to have increased specificity for tumour cell lines vs healthy human cells and inhibited tumour cell migration[218]. Potential derivatives of borrelidin H could therefore be effective anti-cancer candidates. In this study, we will only be discussing borrelidin A (henceforth borrelidin).

The *borO* gene encodes a BGC-associated ThrRS which is resistant to borrelidin. Previous work has identified four possible amino acid positions as important for resistance.[110] These are found in a hydrophobic binding site adjacent to the aminoacylation active site, but their importance has not been validated and a structure of BorO has not been solved to date. The binding of borrelidin to sensitive ThrRS (e.g., EcThrRS) is facilitated by conformational change, which opens this hydrophobic borrelidin binding site. The protein is then held in a conformation which is incompatible with catalysis, and the threonine, $tRNA^{Thr}$ and ATP binding sites are blocked by borrelidin. In BorO, and some archaeal homologues, this site is thought to be sterically blocked or is missing entirely, hence their resistance to borrelidin (because borrelidin cannot bind). This has yet to be explored experimentally, only EcThrRS, HsThrRS and the ThrRS from the fungal plant pathogen *Phytophthora sojae* with borrelidin bound currently available on the Protein Data Bank (PDB)[110]. In previous work investigating the resistance to borrelidin, spontaneous resistant mutants to borrelidin in yeast were generated and showed increased expression of the native ThrRS. Unfortunately DNA sequencing did not exist when this study was conducted and therefore the genetic basis for the spontaneous resistance mechanisms remains elusive[221].

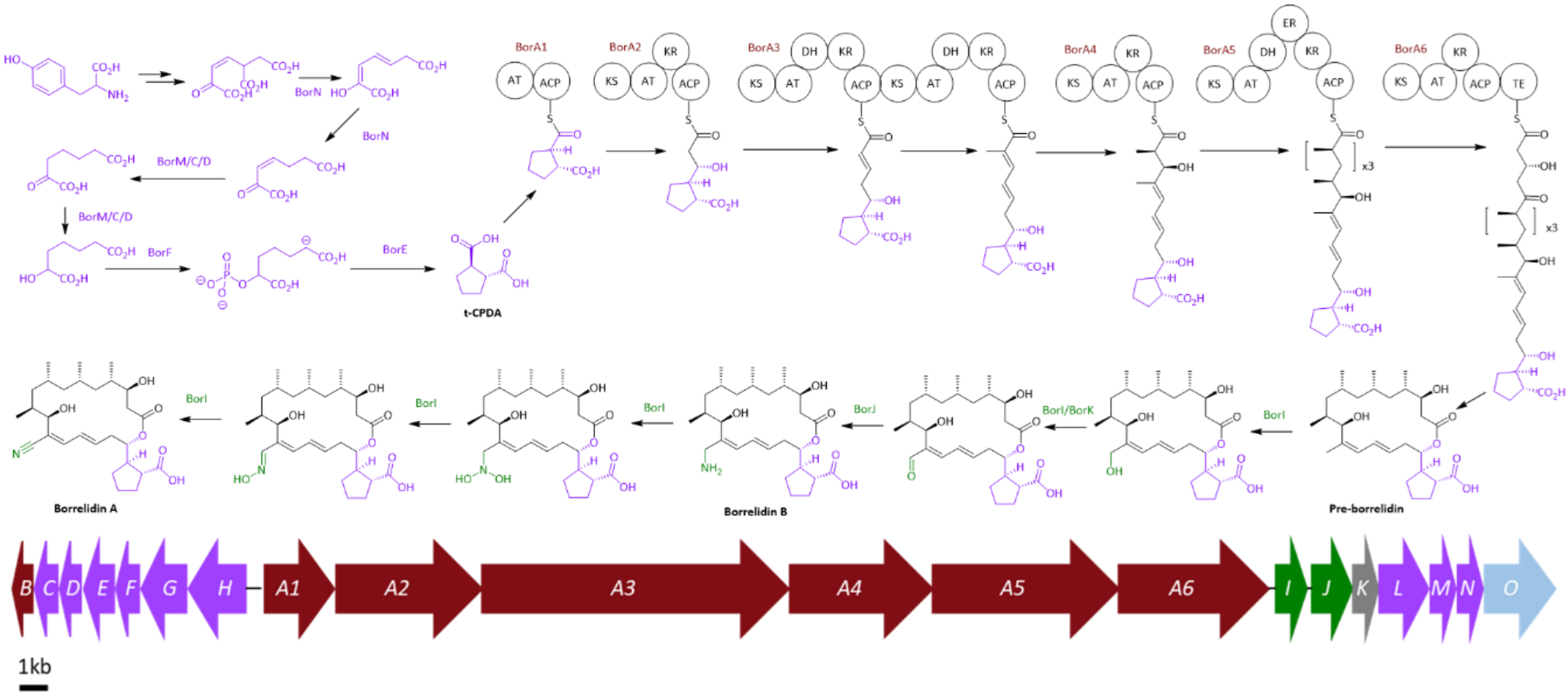


Figure 3.1. Proposed biosynthesis of borrelidin. Schematic shows the genes associated with: the production of starter unit (purple); genes with unknown function (grey); genes associated with nitrile formation (green); polyketide formation in dark (red); and the self-resistance gene is shown in light blue. The starter unit and its precursors are shown in purple throughout the biosynthesis scheme, while nitrile moiety formation is highlighted in green. Circles represent PKS domains, with AT = acetyltransferase, ACP = acyl carrier protein, KS = ketosynthase, KR = ketoreductase, DH = dehydrogenase, ER = enoyl reductase, TE = thioesterase. Key intermediates are labelled. Figure adapted from Olano *et al.* 2004a[170] and Olano *et al*. 2004b[213]

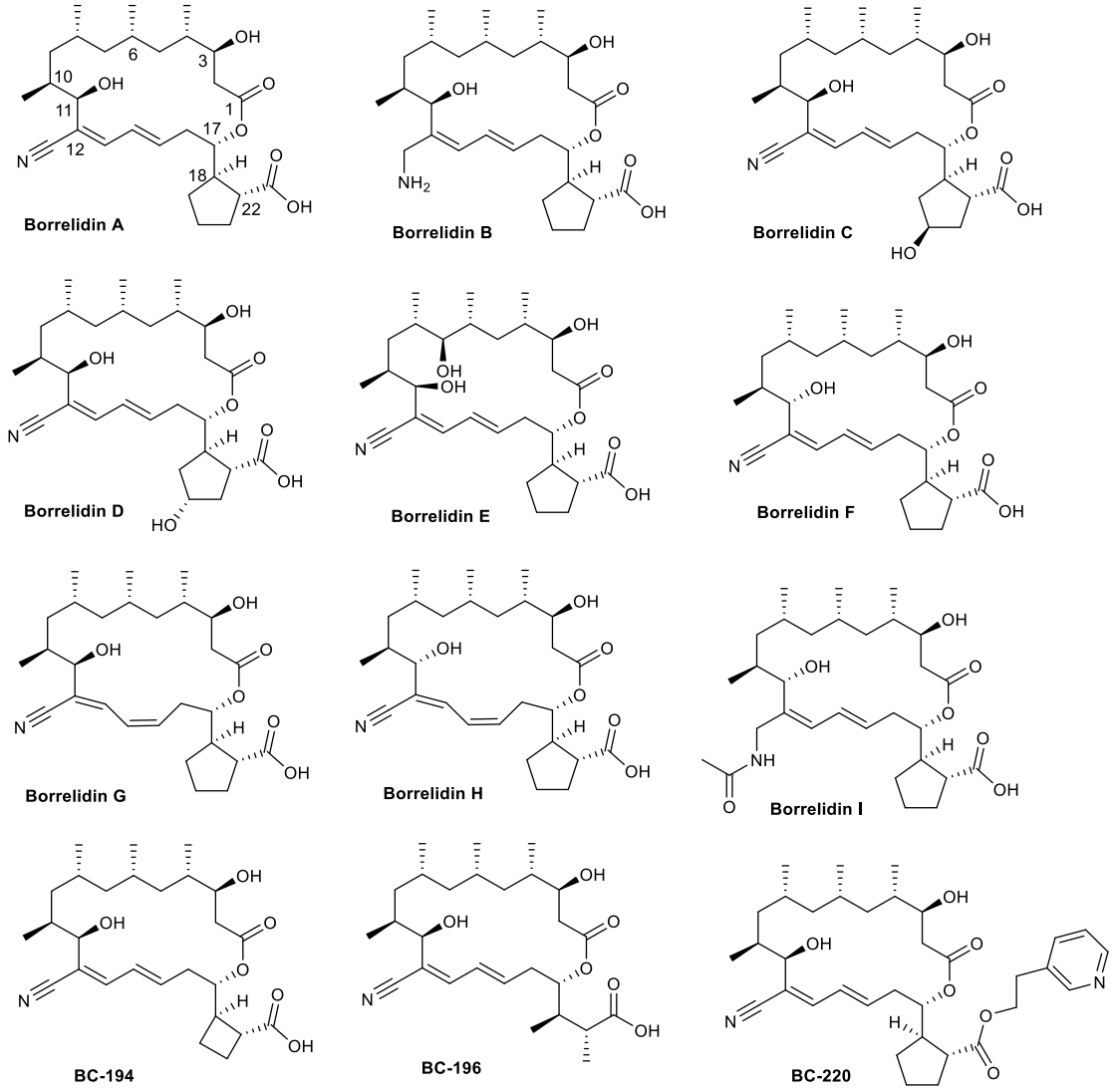


Figure 3.2. Structures of all of the identified borrelidin congeners, and semisynthetic derivatives.

The aim of this study was to try and unpick the specific borrelidin self-resistance determinants in BorO, by comparison to the ThrRS from *E. coli*, a known target of borrelidin and the housekeeping enzyme (SpThrRS) from the producer of borrelidin, *Streptomyces parvulus*.

In this study, I confirmed that borrelidin can inhibit the growth of the clinically relevant pathogens *B. subtilis*, *E. coli* and MRSA. I identified the gene for the housekeeping ThrRS, *SpThrRS* in the producing strain and showed that this housekeeping ThrRS can confer borrelidin resistance in a heterologous host. I found that borrelidin production is not compatible with the sporulation stage of the *Streptomyces* lifecycle, and, using a range of biochemical and biophysical techniques, I have shown that BorO and SpThrRS are unable to successfully bind borrelidin, whereas EcThrRS is able to. As well as this I have determined the key residues for resistance in these proteins.

3.2 Results and discussion

3.2.1 Borrelidin can inhibit a wide array of laboratory indicator strains.

Although previous work has demonstrated that borrelidin can inhibit the growth of a variety of strains, minimum inhibitory concentrations (MICs) have not always been reported. In order to determine a baseline of these in bioindicator strains, a series of bioassays were performed in liquid media. Bioassays were performed in a 96-well plate format, allowing a measurement of optical density (OD) to estimate growth following incubation at 37 °C for 20 hours with increasing concentrations of borrelidin. Cell viability and metabolism was also visualised with the addition of resazurin dye. Resazurin dye is converted from a blue pigment to a pink by living cells and thus can give a visual readout of minimum inhibitory concentration (MIC).[222]

It was observed that borrelidin can inhibit the growth of both Gram-negative and Gram-positive bacteria, including the clinically relevant strain, MRSA. The results of these assays showed that borrelidin is capable of inhibiting both methicillin susceptible *Staphylococcus aureus* (MSSA) and membrane permeabilised *Escherichia coli* (NR698) to an MIC of 16 µg/mL, whereas it is less effective against methicillin resistant *S. aureus* (MRSA) and wild-type *E. coli* (25922, with an MIC of 64 µg/mL. Interestingly, *B. subtilis* was unable to grow in the presence of any tested concentration of borrelidin. All results are summarised in Table 3.1. with graphs of OD_{600} and pictures of resazurin assays in Supplemental Figure 1 and Supplemental Figure 2.

Table 3.1. Minimum inhibitory concentration (MIC) for a variety of bioindicator strain. MICs given in both µg/mL and µM as measured in liquid bioassays by both OD_{600} and resazurin assays.

Bioassay Strain	**MIC (µg/mL)**	**MIC (µM)**
MSSA	16	32.7
MRSA	64	130.8
E. coli NR698	16	32.7
E. coli 25922	64	130.8
B. subtilis	< 1	<2.0

3.2.2 Sequencing of the S. parvulus genome leads to the identification of the housekeeping ThrRS

In order to identify the *Streptomyces parvulus* housekeeping ThrRS, the genome was sequenced for the first-time using Illumina sequencing (Novogene, Hong Kong). The G+C content of the genome was found to be 71.62%, the approximate length 8.4 Mbp and the L50 (number of contigs required to cover 50% of the genome) for the assembly was 4, and the L75 (number of contigs required to cover 75% of the genome) for the assembly was 8.

The borrelidin BGC itself could not be assembled. Specifically, the PKS genes sequenced poorly, a phenomenon regularly seen for the highly repetitive PKS DNA sequences, which frequently

"collapse" with Illumina short read sequencing. Fortunately, the BGC had previously been sequenced and deposited on GenBank and could be used to scaffold this section of the genome. Using the deposited BorO sequence as the query, a BLAST search against the *S. parvulus* genome revealed the housekeeping gene, hereafter referred to as SpThrRS. An antiSMASH analysis of the *S. parvulus* genome indicated that there are at least 35 possible biosynthetic gene clusters (Figure 3.3), of which 16 are very similar or identical to those previously studied and therefore likely to produce known compounds, or compounds chemically similar to known compounds. However, of those which have not been previously identified, it is possible that some may encode for novel antimicrobials.

In total, *S. parvulus* appears to have 8 terpene BGCs, 3 NRPS BGCs, 5 lanthipeptide BGCs, 2 other RiPP BGCs, 3 siderophore BGCs, 5 PKS BGCs (regions 8.1, 13.1 and 15.1 are all on short contigs, and so their full BGCs cannot be validated so have not been counted), 2 indole BGCs, 3 hybrid BGCs and an ectoine BGC.

As well as uncovering the housekeeping ThrRS, a gene encoding a truncated ThrRS with no N-terminal editing domain was identified. As well as examining the genome for ThrRs homologues, I also wanted to identify any potential $tRNA^{Thr}$ genes. The cognate $tRNA^{Thr}$ genes for ThrRS were identified; six were initially annotated, but two were found to be in the middle of annotated open reading frames and did not have an identifiable anticodon.

3.2.3 BorO is not functional in E. coli.

In a similar manner to work on the obafluorin resistance determinant, ObaO (discussed further in detail in Chapter 4 and published in Scott *et al.* 2019[174]) both *borO* and *SpThrRS* were cloned into the expression vector pJH10TS for heterologous expression in membrane permeabilised *E. coli* NR698 (hereafter NR698) as had already been done for EcThrRS and the ObaO homologues. To determine if BorO could confer resistance to borrelidin in *E. coli* NR598, cells that were inoculated into soft nutrient agar (SNA) were challenged with spots of 4 µL of a borrelidin containing solution, containing increasing concentrations, and grown for a further 16 hours at room temperature. Zones of inhibition indicated that BorO was not able to convey resistance to borrelidin in this strain. Interestingly, however, SpThrRS was able to confer resistance under these conditions (Figure 3.4). This was surprising as BorO has previously been shown to confer resistance when heterologously in *Streptomyces albus*[170] and biochemical work has shown that BorO and SpThrRS can be functionally expressed in *E. coli* and purified from NiCo21 cells (discussed in sections 3.2.7 & 8).

Region	Type	From	To	Most similar known cluster		Similarity
Region 1.1	NRPS-like	261,847	305,737	alanylclavam / 2-hydroxymethylclavam / 2-formyloxymethylclavam / clavam-2-carboxylate	Other:Non-NRP beta-lactam	12%
Region 1.2	furan	421,016	442,029	methylenomycin A	Other	9%
Region 1.3	lanthipeptide-class-v	471,658	513,990			
Region 1.4	lanthipeptide-class-iii	566,725	589,328	catenulipeptin	RiPP:Lanthipeptide	60%
Region 2.1	siderophore	59,095	69,197			
Region 2.2	lanthipeptide-class-i	269,288	294,477			
Region 2.3	RiPP-like	442,859	454,190			
Region 2.4	terpene	469,599	491,779	geosmin	Terpene	100%
Region 2.5	siderophore	641,453	654,626	paulomycin	Other	9%
Region 3.1	T3PKS	131,249	172,355	herboxidiene	Polyketide	8%
Region 3.2	NRPS-like	414,956	456,944	streptothricin	NRP	100%
Region 4.1	T1PKS , hglE-KS	1	67,934	azalomycin F3a	Polyketide	34%
Region 4.2	terpene	87,214	108,568	lysolipin I	Polyketide	4%
Region 4.3	indole	375,495	396,616	5-isoprenylindole-3-carboxylate β-D-glycosyl ester	Other	33%
Region 4.4	terpene	454,498	489,437	isorenieratene	Terpene	63%
Region 6.1	NRPS-like , T1PKS , NRPS	3	79,796	antimycin	NRP + Polyketide	93%
Region 6.2	terpene	124,980	150,563	isorenieratene	Terpene	100%
Region 6.3	indole	152,522	173,700	7-prenylisatin	Other	100%
Region 8.1	T1PKS	1	45,643	midecamycin	Polyketide	11%
Region 13.1	T1PKS	1	5,358			
Region 15.1	T1PKS	1	4,544			
Region 60.1	ectoine	522,181	532,579	ectoine	Other	100%
Region 60.2	melanin	1,415,073	1,425,681	istamycin	Saccharide	4%
Region 60.3	siderophore	1,511,535	1,523,307	desferrioxamin B / desferrioxamine E	Other	83%
Region 60.4	T2PKS	2,146,583	2,219,179	spore pigment	Polyketide	66%
Region 60.5	terpene	2,254,818	2,275,909	albaflavenone	Terpene	100%
Region 61.1	NRPS	2	79,071	lipopeptide 8D1-1 / lipopeptide 8D1-2	NRP	84%
Region 61.2	terpene	464,256	490,180	hopene	Terpene	100%
Region 61.3	T1PKS	595,830	689,004	vicenistatin	Polyketide	65%
Region 61.4	terpene	852,010	873,041			
Region 61.5	RiPP-like	884,412	894,627	informatipeptin	RiPP:Lanthipeptide	42%
Region 61.6	NRPS , T1PKS	1,153,813	1,266,541	borrelidin	Polyketide:Modular type I	88%
Region 61.7	lanthipeptide-class-i	1,373,097	1,397,682			
Region 61.8	lanthipeptide-class-iii	1,409,563	1,432,145	SapB	RiPP:Lanthipeptide	100%
Region 61.9	lassopeptide , T2PKS , butyrolactone	1,533,026	1,620,287	fluostatins M-Q	Polyketide	65%

Figure 3.3. antiSMASH report for the *S. parvulus* genome sequence. Each genome region is labelled based on BGC type and location in the genome. The most similar characterised BGC is then identified, along with a score for similarity to this BGC is then shown. The borrelidin BGC is region 61.6. The coelichelin (a siderophore) BGC is also in this region.

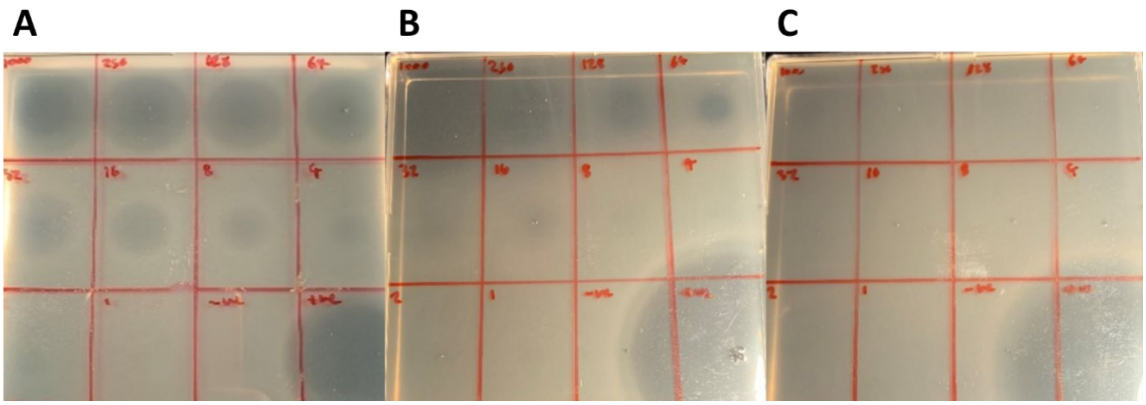


Figure 3.4. SpThrRS can confer resistance to borrelidin, but BorO does not, when expressed in *E. coli* NR698. A) EcThrRS. **B) BorO. C) SpThrRS.** *E. coli* NR698 was inoculated 1:10 into 25 mL soft nutrient agar which was poured into 100 mm square petri dishes. 4 µL of borrelidin (dissolved in DMSO) was spotted at concentrations of (top left to bottom left as indicated) 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL, negative control of DMSO only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates were then incubated at room temperature for 16 hours.

One possible reason for the lack of resistance conveyed by BorO in NR698 is the lack of expression of a fully functional protein. To check this, proteins were tagged with a FLAG sequence (DYKDDDDK) at their N-termini in pJH10TS and protein expression was determined using in a Western blot probed with an anti-FLAG antibody, as seen in Figure 3.5. It was found that EcThrRS, BorO and SpThrRS were all expressed from the pJH10TS plasmid in *E. coli* NR698. Unfortunately, in this assay the ObaO flag-tagged construct did not express at a detectable level in this assay. Previous Western blotting performed by Dr Sibyl Batey (JIC) did show expression of FLAG-tagged ObaO, and EcThrRS (see Supplemental Figure 3). Both the FLAG-tagged ObaO and SpThrRS conferred resistance to borrelidin in NR698, while FLAG-tagged EcThrRS and BorO could not, consistent with non-FLAG tagged results.

Convention normally dictates that the self-resistance gene encoded in a BCG conveys resistance to the natural product, whilst the housekeeping copy is sensitive to it. Following this convention, BorO should confer borrelidin resistance whilst SpThrRS should; however, I found the opposite to be true for heterologous expression in *E. coli*. To my knowledge, the presence of a self-resistance gene in a strain with a resistant housekeeping protein is rare and warrants further exploration.

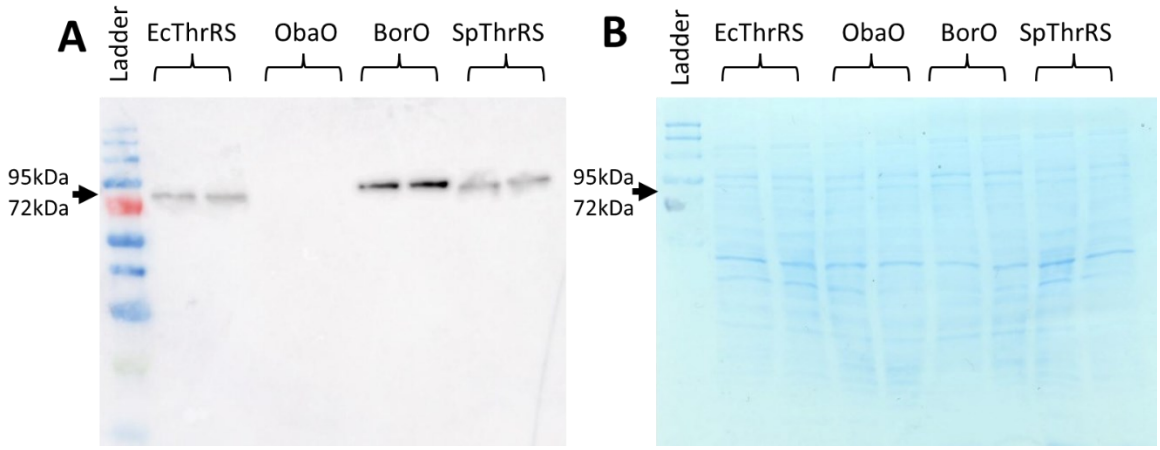


Figure 3.5. EcThrRS, BorO and SpThrRS were expressed in *E. coli* NR698. A) Western blot probed with a HRP conjugated anti-FLAG antibody. Biological duplicates for each sample are shown side by side. The arrow indicates the predicted size of ThrRSs, with the nearest sizes on the prestained broad range colour protein standard (NEB) protein ladder. **B) SDS-PAGE gel, Coomassie stained following transfer.** Same gel used for the Western blotting, stained after transfer as a protein loading control. As above ThrRS is indicated with an arrow. Same gel used for both parts of the figure, with the Coomassie staining acting to confirm equal sample loading.

3.2.4 BorO and SpThrRS can confer resistance to borrelidin in *Streptomyces venezuelae*

To find a *Streptomyces* strain which was appropriate for testing BorO resistance in *Streptomyces in vivo*, *Streptomyces coelicolor, Streptomyces venezuelae, Streptomyces albus* and *Streptomyces lividans* were screened for borrelidin sensitivity by a spot-on lawn bioassay. All of the strains showed some natural resistance when compared to MRSA and *E. coli.* This inherent resistance is likely to arise is two ways, with the *Streptomyces* outer membrane posing a barrier to borrelidin uptake, and the presence of multiple efflux pumps that are able to export borrelidin from the cell. In these assays *S. venezuelae* was the most susceptible to borrelidin with an MIC of 64 μg/mL, so all further experiments focused on this species.

To confirm that BorO confers resistance *in vivo*, the open reading frames of BorO and SpThrRS were cloned into pMB743 under an the constitutively active ermE* promoter and introduced into *S. venezuelae* by conjugal mating. pMB743 is a pIJ10770 based vector with a thiostrepton resistance cassette (and hygromycin for selection in *E. coli*), constructed by Dr Matt Bush (JIC)[185]. To determine if the expression of BorO or SpThrRS would increase *S. venezuelae* resistance to borrelidin, a spot-on lawn bioassay was then performed (Figure 3.6). The MIC for each strain was determined, this revealed that *S. venezuelae* containing empty vector, or with no vector, have an MIC for borrelidin of 64 μg/mL whereas the BorO and SpThrRS containing strains had full resistance to borrelidin, up to 1000 μg/mL. This confirms that both BorO and SpThrRS can convey resistance to borrelidin in non-native strains.

The ability of BorO to confer resistance to borrelidin in *S. venezuelae* but not in *E. coli*, while demonstrably being expressed in the latter, would suggest that there is a difference in the interaction of BorO and its substrates in *E. coli* and *S. venezuelae*. As we can tell from the work in section 3.2.7, BorO is expressed as a soluble and active protein in *E. coli*; as well as this, the abundance of threonine and ATP are the same regardless of the biological system. The only difference between these two strains are the $tRNA^{Thr}$ molecules in *E. coli* and *Streptomyces*. These $tRNA^{Thr}$ molecules have variations in their sequences, as seen in Supplemental Figure 4 and Supplemental Figure 5, and are likely to have differences in their post-transcriptional modifications, but it is hard to estimate what they are.

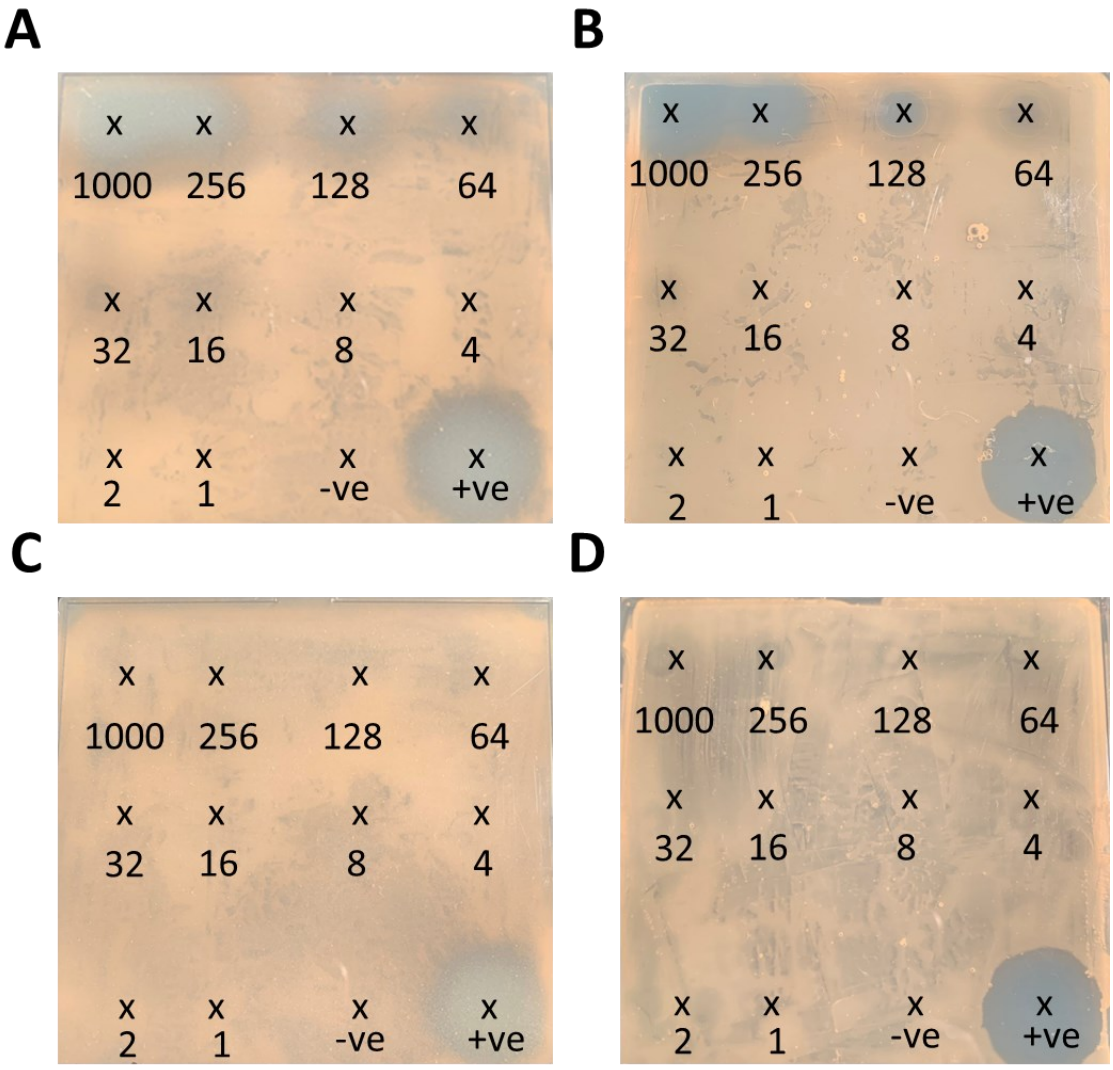


Figure 3.6. BorO and SpThrRS confer resistance to borrelidin in *S. venezuelae*. A) Wild-type *S. venezuelae*. B) empty vector control of *S. venezuelae* with pMB743. C) *S. venezuelae* expressing BorO. D) *S. venezuelae* expressing SpThrRS. Spores of each strain were spread for confluence on MYM media before spotting of 4µL borrelidin (at location marked X, dissolved in DMSO) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left as indicated, negative control of DMSO only bottom middle, and positive control of 50 µg/mL kanamycin bottom right. Plates were then incubated at 30°C for 16 hours.

3.2.5 Borrelidin production is altered depending on expression of borO

To assess the role of BorO in the borrelidin producing strain, I set out to create in-frame knockouts of *borO* in *S. parvulus*. In previous work on *S. parvulus*, all genetic manipulation had been done by protoplast transformation[170]. A significantly simpler method for the introduction of plasmids to *Streptomyces* is conjugation from *E. coli*, which relies on the introduction of an antibiotic resistance cassette, so *S. parvulus* was therefore tested for antibiotic susceptibility. It was found that *S. parvulus* is sensitive to a number of commonly used antibiotics including thiostrepton, kanamycin and hygromycin.

The transfer of the integrative plasmids pGP9[183] and pIJ12057[184] into *S. parvulus* by conjugation proved successful. This was the first time that plasmids have been successfully introduced into this strain by conjugation. The knockout vector was then designed as described in section 2.4.5.3. The regions upstream and downstream of *borO* proved to be difficult to amplify; however, with extensive optimisation of PCR conditions these were successfully amplified. Interestingly, amplification was only possible when using Phusion polymerase and with the addition of betaine. Betaine is an additive that can be added to PCR reactions in order to relax secondary structure elements in the DNA which may interfere with amplification[223]. This could suggest the presence of strongly defined secondary structure elements in these regions, which may be regulatory elements[224]. Once the knockout plasmid was assembled (pBorOKO), the knockouts strains were generated by conjugation of this plasmid into both *S. parvulus* WT and *S. parvulus* Δ*borG* strains and selection of exconjugants with hygromycin. These were then grown on antibiotic free media to allow allelic exchange, and resulting colonies checked for hygromycin resistance and kanamycin sensitivity (indicating allelic exchange has occurred and the knockout plasmid has been ejected). Finally, allelic exchange was confirmed by PCR. This generated two strains: *S. parvulus* Δ*borO* and *S. parvulus* Δ*borG*Δ*borO*.

Spot-on lawn bioassays of *S. parvulus* (hereafter WT) and *S. parvulus* Δ*borO* (hereafter Δ*borO*) were performed confirming that *S. parvulus* is resistant to borrelidin, even in the absence of *borO*, which once again confirms that SpThrRS conveys resistant to borrelidin, as seen in Figure 3.7.

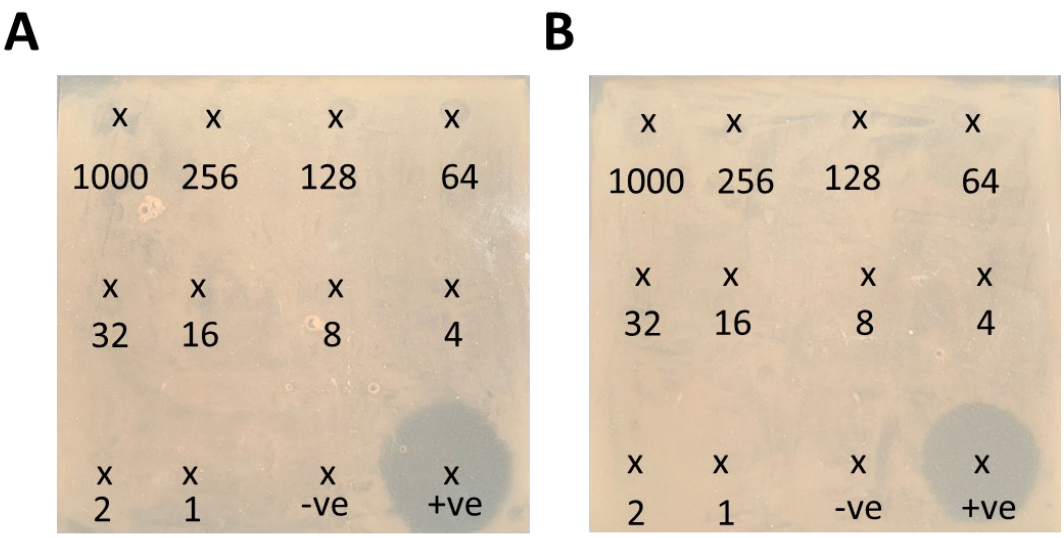


Figure 3.7. *S. parvulus* is resistant to borrelidin in the absence of *borO*. A) Wild-type *S. parvulus*. B) *S. parvulus* Δ*borO*. Spores of each strain were spread for confluence on each plate of MAM media before spotting on of 4µL borrelidin (dissolved in DMSO) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of DMSO only bottom middle, and positive control of 50 µg/mL kanamycin bottom right. Plates were then incubated at 30°C for 16 hours.

The ability of these strains to produce borrelidin was then monitored in order to see if they can both produce borrelidin and survive in the absence of their BGC located 'self-resistance gene'. This included examining, borrelidin production of the *S. parvulus* Δ*borG*Δ*borO* strain, and all strains were tested both without feeding tCPDA, and when fed with exogenous tCPDA, the starter unit required for biosynthesis as seen in Figure 3.8. This analysis was somewhat complicated by strains sporadically and unexpectedly not producing borrelidin, which correlated with a change in colour of the liquid culture, but this was overcome by performing sufficient repeat tests and replicates, and carefully monitoring the strain phenotypes as well as borrelidin titres. The colour change was reminiscent of the colour change that *S. parvulus* growing on solid culture undergoes when sporulating, as seen in Figure 3.9.

There was no significant change in borrelidin titre when *borO* was deleted, but complementation of the mutation with either *borO* or *spThrRS* led to a significant increase (ca. 2.5- to 3.5-fold increase in titre). Moreover, when all of these strains are fed with tCPDA the titre was found to increase significantly, with the complementation with *borO* leading to a doubling of titre compared to the fed WT strain (an approx. 5-fold increase over the unfed WT). Overall, the ectopic expression, possibly overexpression due to the strong ermE* promoter, of *borO* had a positive effect upon borrelidin production. Empty plasmid carriage in both Δ*borO* strains appeared to significantly suppress borrelidin production, an observation for which we have no current explanation (see Figure 3.8). No antibiotics were added for selection when plasmid containing strains were assayed.

The *S. parvulus* Δ*borG*Δ*borO* strain was originally produced due to a concern that it would be difficult to generate delete *borO* in a borrelidin producing strain. Thus, *borO* was also knocked out in the Δ*borG* strain, generating a double knockout. The *borG* product is required for the biosynthesis of borrelidin, via a role in producing the biosynthetic starter unit tCPDA; thus, the Δ*borG* strain produces only trace levels of borrelidin unless supplemented with exogenous tCPDA. As discussed further in Chapter 4, for obafluorin, *obaO* knockouts could only be generated in a single knockout strain which could not produce obafluorin, Δ*obaL*[174]. It was expected that Δ*borG* strains would produce no borrelidin or very little borrelidin when unfed, but when fed would have a similar production profile to the WT and Δ*borO* strains. The production profile of the Δ*borG* strains was, however, more complicated. Both Δ*borG* and Δ*borG*Δ*borO* strains produced none or very low levels of borrelidin unless exogenous tCPDA was added, after which they produced similar or slightly higher levels to the WT strain when supplemented with tCPDA. Empty plasmid carriage once again supressed borrelidin production (a phenomenon we cannot explain currently). However, when the Δ*borG*Δ*borO* strain was complemented with either BorO or SpThrRS, both appeared to have sporulated in liquid culture (characterised by a darkening of the culture), and therefore did not produce borrelidin as has been observed for the WT strain when grown on agar; it should be noted that *Streptomyces* strains that sporulate when grown in liquid culture are very rare. Photographs of the cultures from which borrelidin was extracted can be seen in Supplemental Figures 6-15.

We can see from this work that the absence of *borO* has an effect on borrelidin production, but borrelidin can still be produced in its absence, confirming that SpThrRS alone is able to convey self-resistance to borrelidin.

3.2.6 Streptomyces parvulus can sporulate in liquid culture

To confirm that the observed phenotype of darker colour change and lack of borrelidin production (even when fed with tCPDA) is truly a sporulation phenotype, cultures were observed by light microscopy at 1-5 days. It was found that cultures of *S. parvulus ΔborO* sporulated by day 3 and the development of spore pigment colouration in the liquid culture correlated with the sporulation of the culture. On the same day, lighter coloured, borrelidin producing cultures showed hyphal growth under microscopy (Figure 3.9C and Figure 3.10A), while the darker coloured, non-producing cultures had sporulated by day 3 (Figure 3.9D and Figure 3.10B). This suggests that we can correlate the development of a darkened culture colour with sporulation, in liquid culture.

It was found that *S. parvulus* frequently sporulated in liquid culture in borrelidin production media (PYDG). In different experiments, all the strains sporulated in liquid culture; however, the double knockout strains containing pMB743 expressing *borO* or *spThrRS* always sporulated, and Δ*borO* strains also frequently sporulated. Feeding tCPDA did not appear to affect sporulation, nor did the presence or absence of glucose in the media (data not shown), which has been shown to be

important for sporulation in *S. venezuelae*[225]. This observed sporulation in liquid media could be a stress response[226], could be a manifestation of a regulatory effect of BorO, or both. Sporulation in liquid culture under laboratory conditions in *Streptomyces* is rare, with the few reported examples including *S. venezuelae* and *S. griseus*.[225,227,228]

It is possible that the observed sporulation phenotype and variations in production levels of borrelidin is due to a regulatory function of BorO and/or SpThrRS. One of the ways in which these may be regulating borrelidin production could be through binding to tRNA-like secondary structure elements in mRNAs. It is not unprecedented for an aaRS paralogue to have a role in regulation. A paralogue of HisRS (HisZ) has previously been reported able to perform a regulatory role in histidine biosynthesis[229]. Notably this HisRS shows nonspecific RNA binding properties. Additionally, as mentioned in section 1.3.1 and 1.3.2.1, both EcThrRS and HsThrRS (specifically TARS1) have been demonstrated to have a regulatory role (auto-regulatory in the case of EcThrRS) by binding to secondary structure elements in mRNA.

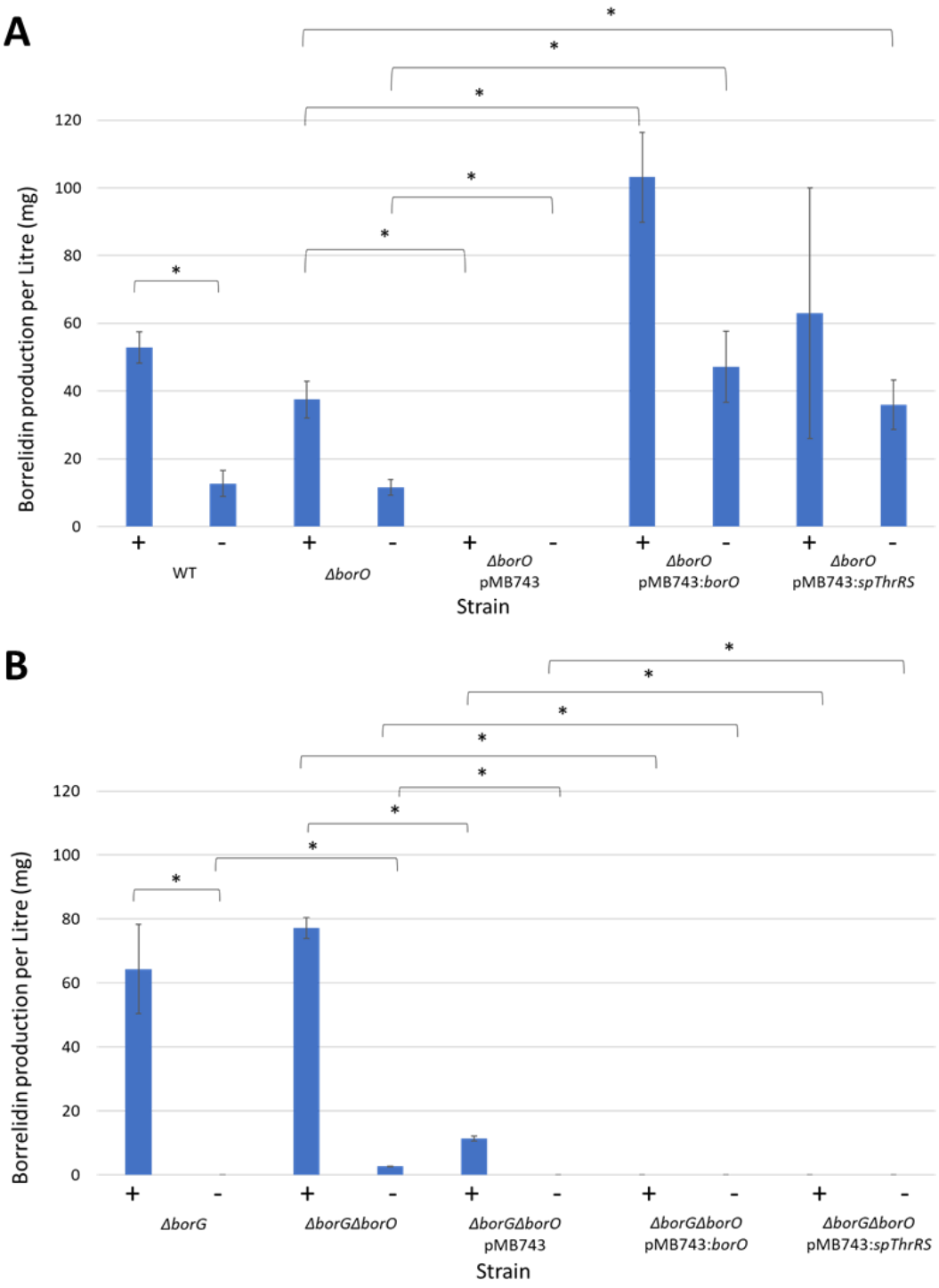


Figure 3.8. Production of borrelidin is varied in different strains of *S. parvulus*. A) *S. parvulus* WT, *ΔborO* and complemented strains B) *ΔborG* strains. Borrelidin production quantified by integration of the area under the curve measuring absorbance at 254nm, compared to a standard curve of a standard of borrelidin at known concentrations (see Supplemental Figure 16). Error bars represent the standard deviation. Significance is indicated by asterixis indicate significantly different samples, as judged by a Student's t-test ($p < 0.05$ *).

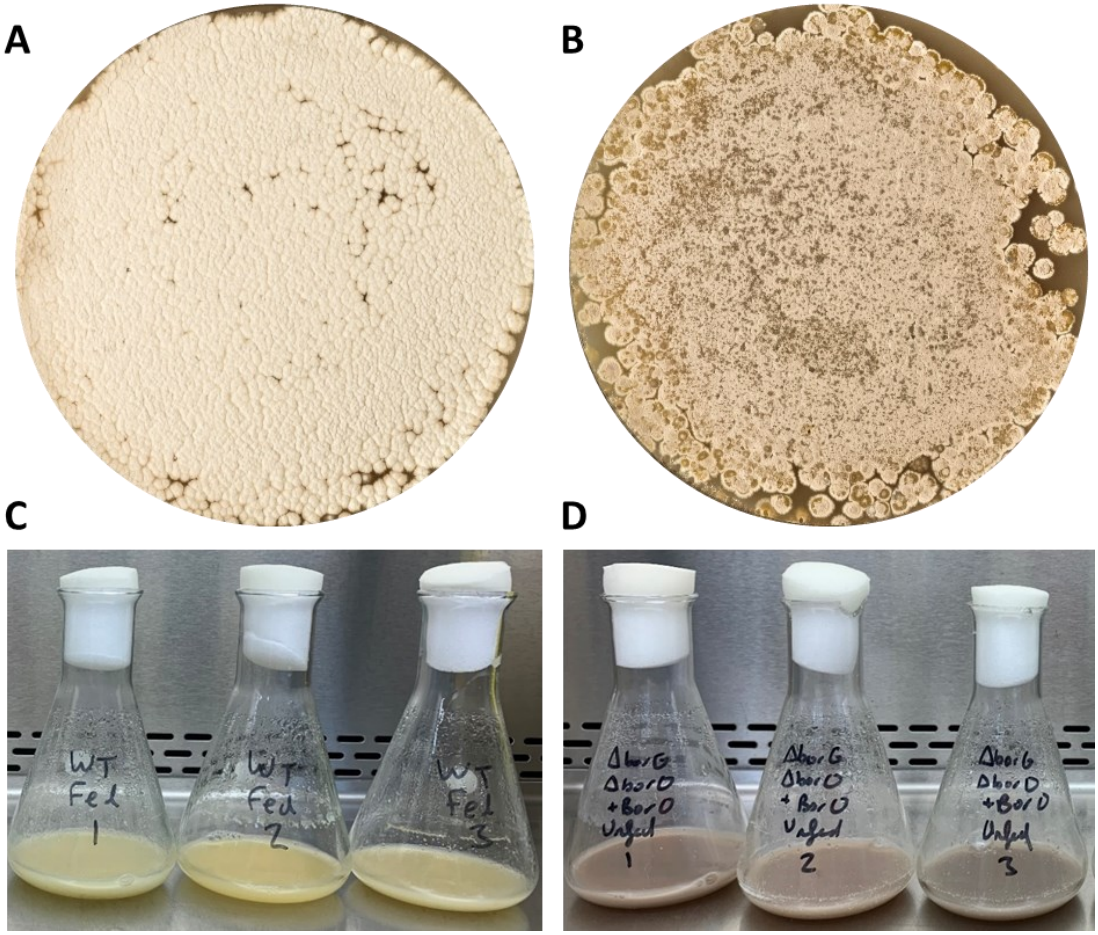


Figure 3.9. *S. parvulus* sporulation is linked to a darkening of the culture. A) *S parvulus* grown on solid SFM, not sporulating as judged by the appearance and lack of spore pigment. **B)** *S. parvulus* grown on solid MAM, sporulating. **C)** non-sporulating cultures in liquid culture showing, as indicated by the lighter colour **D)** sporulating cultures in liquid culture, as indicated by the darker colour.

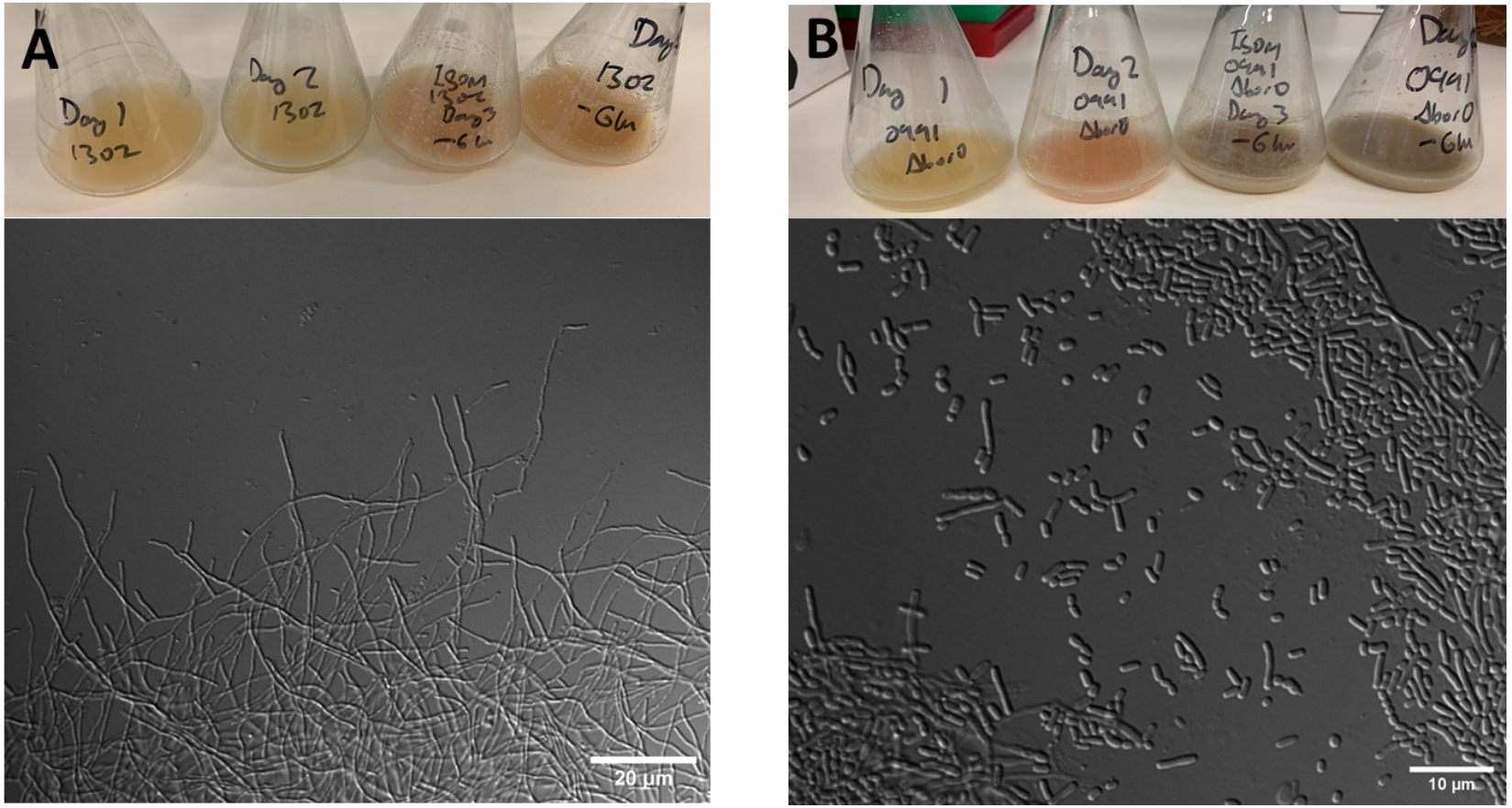


Figure 3.10. *S. parvulus* sporulates in liquid culture. A) Top panel: cultures of the non-sporulating strain *S. parvulus* Δ*borG* after 1, 2, 3 and 4 days of growth. Bottom panel: differential interference contrast (DIC) images of *S. parvulus* Δ*borG* from the day 3 flask showing normal hyphal growth. Cultures had flocculated which is normal for *Streptomyces* in liquid culture. **B)** Top panel: cultures of the sporulating strain *S. parvulus* Δ*borO* after 1, 2, 3 and 4 days of growth. Bottom panel: DIC images of *S. parvulus* Δ*borO* from the day 3 flask showing the formation of spores. Spores were suspended in the culture and only a few hyphae could be observed by microscopy.

3.2.7 Borrelidin does not bind to BorO or SpThrRS in vitro.

During the biochemical studies performed with EcThrRS and ObaO by previous lab members, it was found that both proteins were soluble in the same buffer, and that full length versions expressed best as soluble protein when produced with a N-terminal poly-His tag, while the N-truncated versions (lacking the editing domains) were best expressed as soluble protein when C-terminally poly-His tagged.

To check this was also the case for BorO, a disorder prediction was run on its sequence, suggesting that, potentially, the C-terminus could also be disordered. Both BorO and SpThrRS were synthesised in codon-optimised form for expression in *E. coli* by Twist Bioscience (San Francisco, USA) and a combination of ΔN and ΔC protein production constructs with either C- or N-terminally poly-His tags were generated, and protein expression and solubility were tested in *E. coli* NiCo21. It was found that for BorO, the same buffer, truncations and tag positions were the most suitable as for EcThrRS and ObaO. Because this common buffer could be used to produce soluble protein, to aid in the execution of biophysical experiments, and automated batch purification, this buffer would be used throughout.

In order to have a positive control for binding assays, 5'-O-(threonylsulfamoyl)adenosine (ThrSAA) was synthesised by Dr Edward Hems (JIC). ThrSAA is a non-hydrolysable analogue of threonyl-AMP (ThrAMP) as can be observed in Figure 3.11. ThrSAA has been used previously for studying the binding of ThrAMP to ThrRSs without being turned over[97,230]. While syntheses for this compound had previously been reported, purification always yielded the compound as a triethylammonium salt[231,232]. Dr Hems developed a procedure to yield the compound as a zwitterion and fully characterised the compound by 1D and 2D NMR, as seen in Supplemental Figures 17-22. Structures of ThrRS bound to this compound have already been solved as detailed in Table 1.2, section 1.3.1. The three-letter code for ThrSAA is TSB. If proteins can bind to ThrSAA, we know that they represent correctly folded ThrRS proteins.

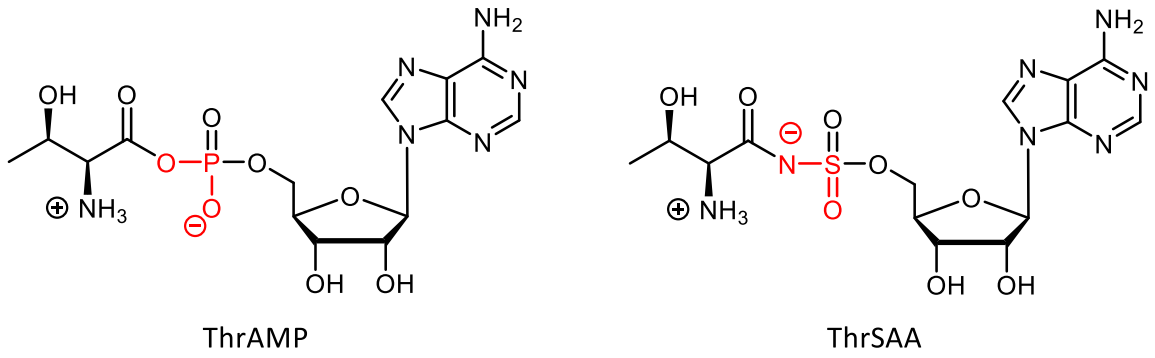


Figure 3.11. Structures of ThrAMP and the synthetic analogue, ThrSAA. Differences between the structures are shown in red. The sulfonamide linkage is much more resistant to hydrolysis that the phosphate.

The quality of the protein samples was checked by mass photometry to ensure that dimeric protein was present, and that the formation of higher order aggregates did not occur (see Figure 3.12). From previous work in the literature, we know that functional ThrRS proteins are dimeric. For all of the proteins tested in this study, both monomeric and dimeric forms could be observed. No higher order multimers are present, except a small quantity of tetramer for BorO. Previous mass photometry with N-terminally truncated ThrRSs showed that with increased dilution of protein sample, a higher proportion of the protein is found as a monomer rather than dimer, suggesting that oligomerisation may occur in a concentration dependent manner (Supplemental Figure 23).

Prior to determining the ability of EcThrRS, BorO and SpThrRS to bind borrelidin, ITC analysis revealed that EcThrRS, BorO and SpThrRS all bound ThrSAA with K_d values of 10.3 ± 1.06 nM (n = 3) for EcThrRS, 87.1 ± 12.0 nM (n = 3) for BorO, and 3.40 ± 0.68 nM (n = 3) for SpThrRS. This shows that all three proteins are folded correctly as they are able to bind the non-hydrolysable intermediate analogue with the expected affinity. When then testing binding of the three enzymes with borrelidin, only EcThrRS bound borrelidin with a K_d of 42.9 ± 4.22 nM (n =3), while BorO and SpThrRS showed no binding to borrelidin (Figure 3.13). Consistent with our findings, the K_i (the inhibition constant) for borrelidin against EcThrRS has previously been reported as 4nM[161], this is of the same magnitude as the measured K_d for binding.

We can therefore postulate that BorO and SpThrRS can confer resistance to borrelidin due to their inability to bind to borrelidin, at the tested concentrations. It is therefore likely that there are some structural differences between the structures of EcThrRS and BorO which prevent their binding to borrelidin.

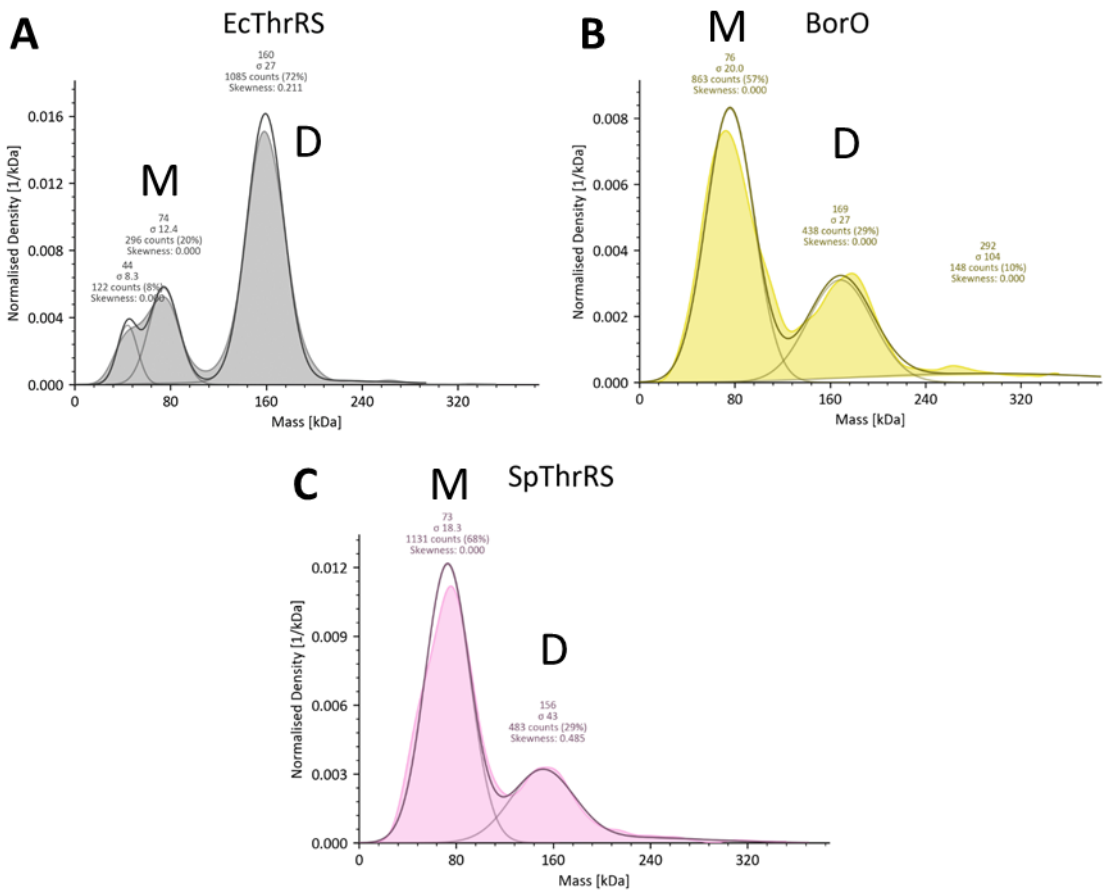


Figure 3.12. Mass photometry analysis of isolated proteins. Histograms of A) EcThrRS, B) BorO and C) SpThrRS. The average predicted mass in kDa is shown above each peak; with a protein concentration of 100 nM used for all three proteins. Proteins were measured in PBS. Peaks align to the sizes of the monomer (labelled with M) and the dimer (labelled with D). Presence of the dimer indicates the presence of functional protein.

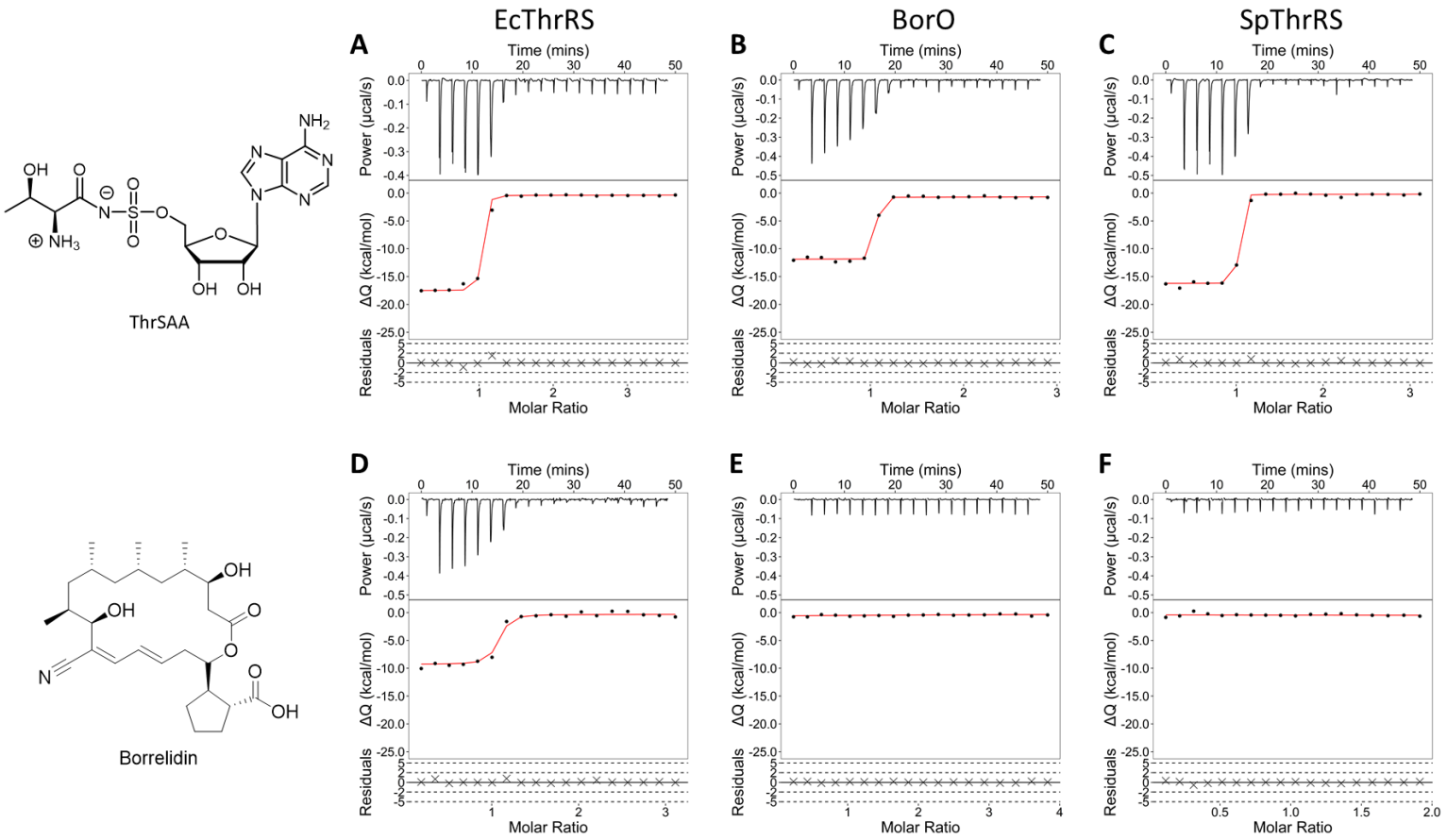


Figure 3.13. BorO and SpThrRS do not bind to borrelidin in ITC binding assays. Each lettered dataset shows the isotherms of the reaction in the top panel, the resulting binding curve in the middle panel, and the residuals (which measures the goodness of fit of the curve to the data) in the bottom panel. **A)** EcThrRS with ThrSAA; **B)** BorO with ThrSAA; **C)** SpThrRS with ThrSAA; **D)** EcThrRS with Borrelidin; **E)** BorO with borrelidin; **F**) SpThrRS with borrelidin. All three proteins show binding to the intermediate analog, ThrSAA, suggesting that they are folded correctly. Only EcThrRS shows binding to borrelidin. Figure generated in RStudio.

3.2.8 The Structure of BorO shows limited conformational change upon interaction with borrelidin

As already mentioned, previous work had suggested possible borrelidin resistance determinants based on multiple sequence alignments alone, with no structural information about BorO[110]. The next goal was therefore to solve the structures of both BorO and SpThrRS in order to directly compare them. Each protein was made up in Lysis buffer containing 10 mM threonine, so that threonine was available for binding to both crystallisation conditions (+/- borrelidin). Crystallisation of BorO at 7mg/mL using the PEGs screen (Qiagen) was attempted in the presence and absence of 2 mM borrelidin. Interestingly, successful crystallisation only occurred in the presence of borrelidin. Crystals were harvested, shipped to Diamond Light Source (DLS) for X-ray diffraction, and an initial 2.7 Å model of BorO could be solved. An X-Ray fluorescence scan confirmed the presence of zinc in the crystal, which we know is found in the active site of ThrRSs, suggesting that these are authentic crystals of BorO (see Figure 3.14).

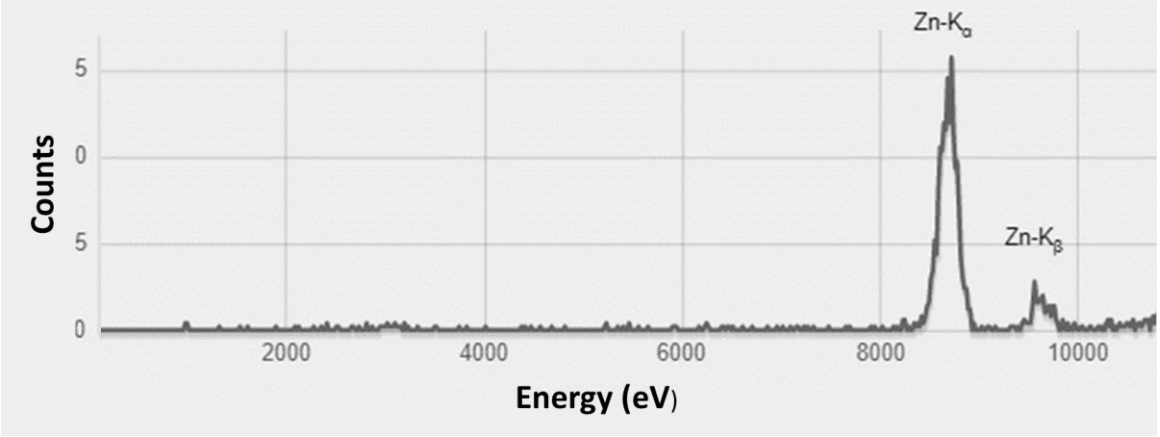


Figure 3.14. X-Ray Fluorescence shows that Zinc was present in BorO crystals. Annotated are the elements corresponding to the peaks. Graph taken directly from Diamond Light Source.

Optimisations of these crystallisation conditions were then designed as outlined in section 2.12.1. Using optimised crystallisation conditions and seeding with crystals from the first crystallisation trial to provide nucleation sites, further crystals could only be reliably grown in the presence of borrelidin. From these optimised crystals, data was collected, and a structure of BorO could be solved at a resolution of 2.5 Å, using the initial, low-resolution model for molecular replacement (Figure 3.15). This is the first structure of any *Streptomyces* ThrRS to be solved, and the first structure of a borrelidin self-resistance protein to be solved.

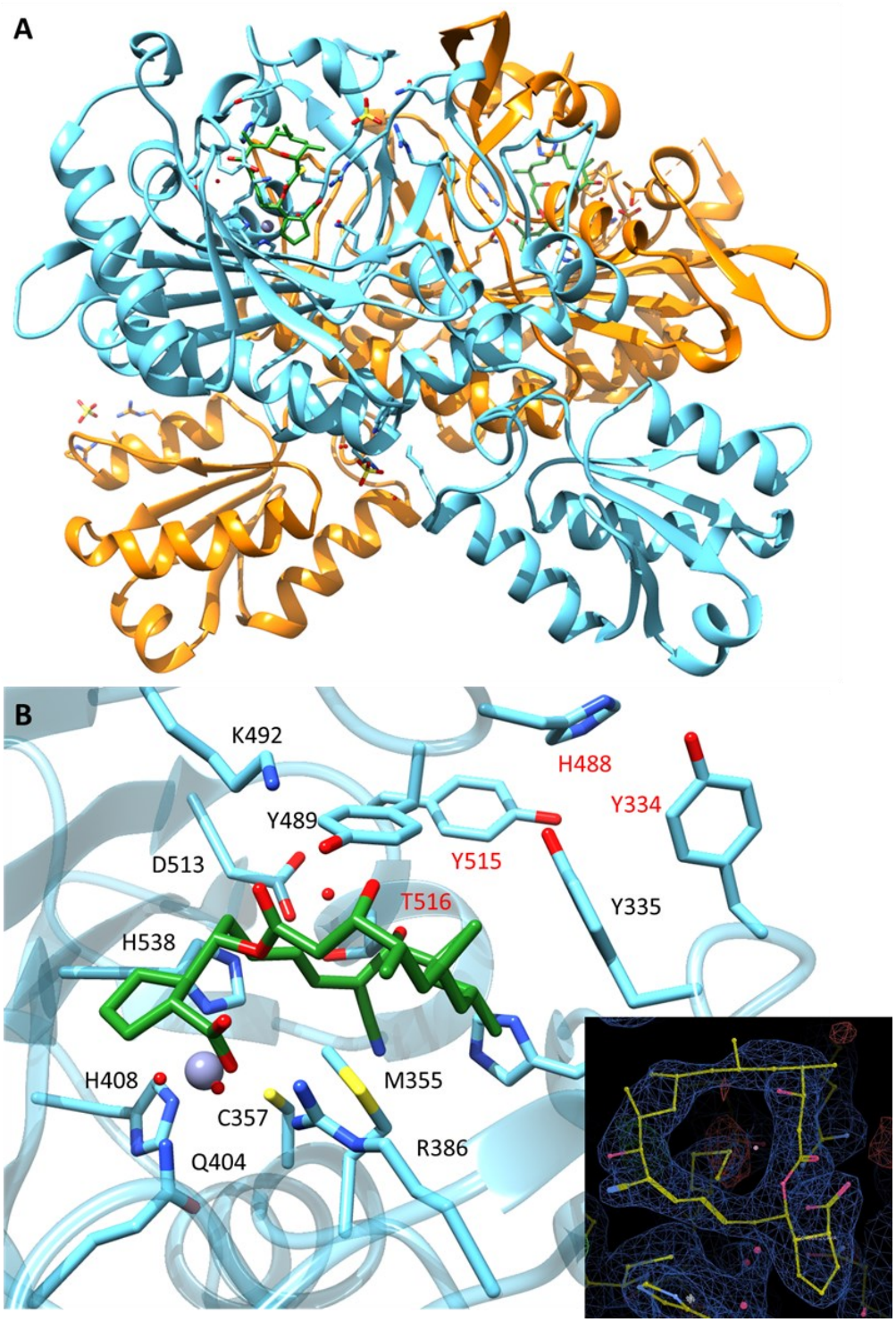


Figure 3.15. A structure of BorO was solved with borrelidin bound in the active site. A) A cartoon representation of the dimeric structure of BorO with monomers shown in blue and orange, with borrelidin in green. **B)** inset, bottom right: electron density map of borrelidin in the active site of BorO contoured to 1 rmsd. Main: borrelidin in the active site of BorO. Residues making contact with borrelidin or identified as a possible resistance determinant are labelled and shown in stick representation. Previously suggested resistance determinant residues (Fang *et al.* 2015)[110] are labelled in red. Borrelidin shown in stick representation in green, protein in blue in cartoon representation. Zinc shown as a grey sphere. Figure generated in Chimera and WinCoot.

Electron density was observed in the active site of the BorO structure which could be assigned as borrelidin. It was unexpected that borrelidin was found bound to the protein. Because in the ITC binding assays described in section 3.2.7, we observed that borrelidin did not bind to BorO. *In crystallo*, however, the protein is considerably more constrained than it is in solution, and borrelidin was added at a high molar excess compared to the protein. This would mean that while borrelidin does not bind to BorO under the conditions used in ITC, at this significantly higher concentration of borrelidin, binding can occur.

To ensure that I had crystallisation conditions for EcThrRS, a PEGs screen (Qiagen) was prepared with 20 mg/mL EcThrRS and either 2 mM borrelidin or 2 mM ThrSAA. Crystals grew in the presence of both ligands, and structures could be solved of EcThrRS bound to borrelidin at 1.9 Å and bound to ThrSAA at 1.75 Å (data collected as part of the Diamond CCP4 Data Collection and Structure Solution Workshop 2021). These structures were similar to those previously deposited (4P3P and 1EVL) but with moderately improved resolution for the borrelidin-bound structure, and moderately improved refinement statistics for the ThrSAA bound structure[97,110]. The structure solution statistics for crystal structures solved in this chapter can be found in Supplemental Tables 1-3.

By comparing the borrelidin-bound and ThrSAA-bound EcThrRS structures, it appears that a conformational change occurs upon binding with borrelidin, opening up the hydrophobic binding pocket allowing access to borrelidin. In the previously published ThrRS structures, this pocket is closed in the absence of borrelidin binding. By comparison, when borrelidin is bound to BorO, BorO undergoes a more moderate conformational change than that undergone by EcThrRS, as seen in Figure 3.16. The conformation of borrelidin in the BorO structure also appears to be different to that seen in both the EcThrRS and HsThrRS borrelidin-bound structures. This conformation adopted by BorO is unprecedented in other ThrRS structures deposited to the PDB. This would also go some way to suggest that this borrelidin binding is an artefact.

We can hypothesise that the conformation required for borrelidin binding is rarely adopted by BorO. Evidence for this includes the fact that BorO only ever crystalised in the presence of borrelidin and that we don't see evidence of borrelidin binding by ITC. Additionally, the conformation of both BorO and borrelidin in the obtained crystal structure was different to those observed in the crystal structures of sensitive enzymes. In sensitive enzymes, borrelidin can bind because the ThrRS can sample multiple different conformations, one (or many) of which will have the borrelidin hydrophobic pocket open and accessible. There will therefore be a high probability that when borrelidin encounters the protein, this binding site is open and borrelidin can bind. In BorO, however there are likely interactions which prevent this conformation from being sampled or prevent it from being sampled frequently; these are the resistance determinants for which we are hunting.

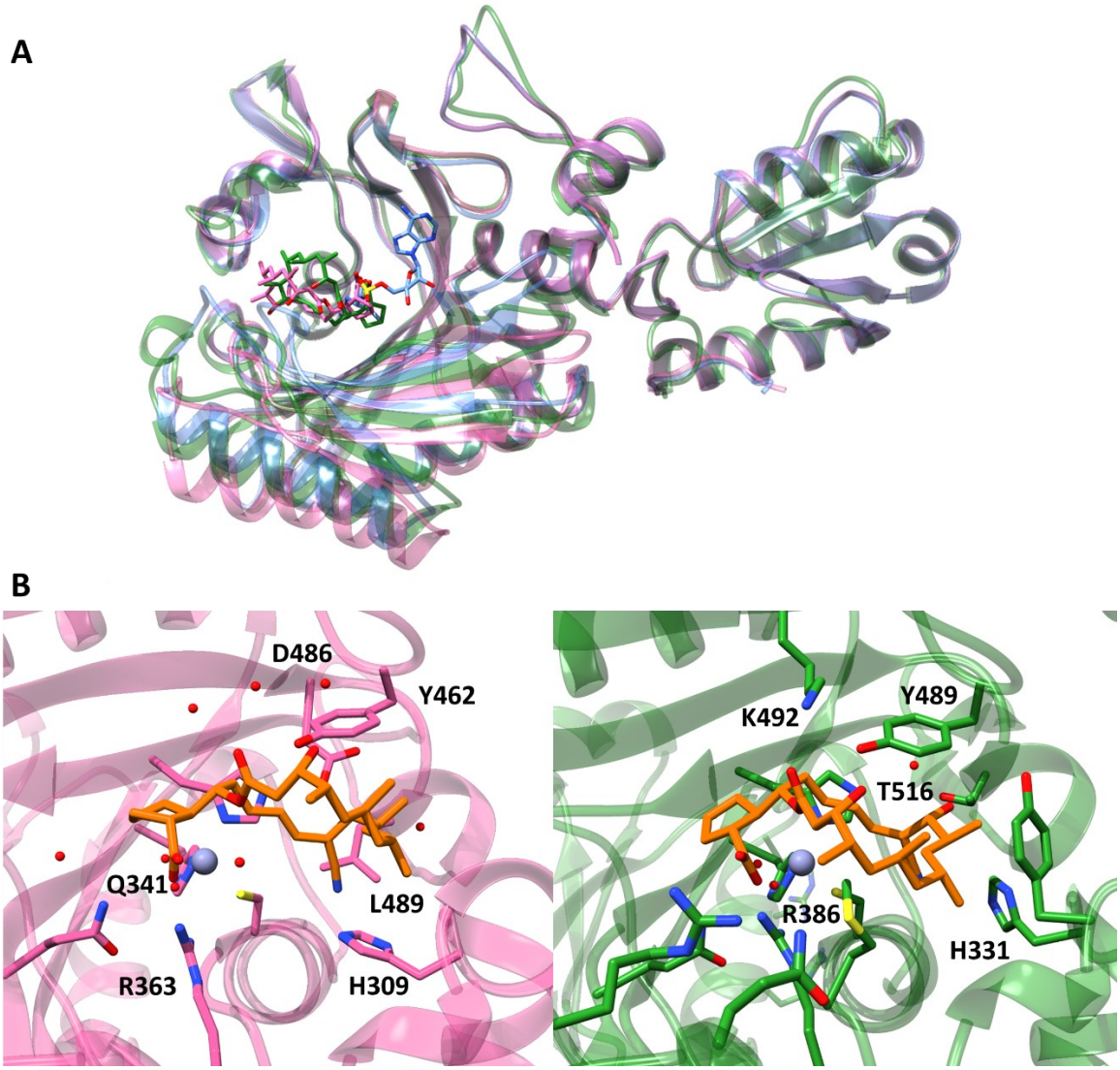


Figure 3.16. Comparison of BorO and EcThrRS borrelidin-bound structures reveals differences in conformation. A) Comparison of protein conformations for BorO and EcThrRS. BorO bound to borrelidin in green, EcThrRS bound to borrelidin in pink, EcThrRS bound to ThrSAA in blue. The conformation of BorO appears to be intermediate between that of the two EcThrRS structures. **B)** zoom-in of borrelidin bound in EcThrRS (left) and BorO (right), with BorO in green and EcThrRS in pink. Borrelidin shown in orange. Borrelidin appears to be bound in a slightly different orientation in BorO, as compared to EcThrRS. Zinc shown as a grey sphere. Figure generated in Chimera.

With the borrelidin binding pocket usually closed, borrelidin will be unable to bind to BorO most of the time, explaining why I was unable to detect binding using ITC. On the other hand, during crystallisation there was a high excess of borrelidin present, there are far more borrelidin molecules and therefore when BorO samples a conformation with the borrelidin binding site open, which is rare, it is more likely that borrelidin will be able to bind. Once bound, this conformation is clearly competent for crystal formation and once a crystal has started to form, the protein will be held in

this conformation due to the crystal contacts with other protein molecules (an artificial state which the protein is unlikely to encounter in solution).

While the solving of a borrelidin-bound structure is hugely useful (and will be discussed in more detail in later sections), it would also have been useful to also obtain an apo-structure or a threonine or ThrSAA bound structure for comparison. With threonine alone having not facilitated growth of crystals of BorO, crystallisation trials were prepared with ThrSAA at 2mM as had been done for borrelidin, with seeding and both in a PEGs screen (Qiagen) and in the same optimised screen as for borrelidin-bound BorO crystals. Unfortunately, this did not lead to crystals growing, suggesting that the conformation in which borrelidin can bind is required for crystallisation of BorO, possibly due to an increased variation of protein conformation with borrelidin bound vs. with just ThrSAA or threonine bound.

Having tried extensively to crystallise SpThrRS following a similar strategy as was used for BorO and being unsuccessful, an Alphafold2 model[233] was generated (statistics in Supplemental Figure 24). AMBER restraints were used to energy minimise such that the orientation of side chains is energetically favoured and chemically relevant. This can then be compared to crystal structures for the design of mutants- when a SpThrRS structure is mentioned, it is this AlphaFold model to which I am referring.

3.2.9 EcThrRS L489T is resistant to borrelidin by interaction with D486

With a structure of BorO having been obtained, the next goal was to mutate EcThrRS to gain resistance to borrelidin. Previous work, detailed in Fang *et al.* 2015[110] suggested that Y334, H488, Y515 and T516 (N312, F461, S488 and L489 in EcThrRS) could be responsible for the ability of BorO to confer resistance to borrelidin. The equivalent amino acids in SpThrRS and the *S. venezuelae* (SvThrRS) are listed in Table 3.2.

Table 3.2. Important residues for borrelidin resistance in EcThrRS, BorO, SpThrRS and SvThrRS.

Amino Acid in EcThrRS	Amino Acid in BorO	Amino Acid in SpThrRS	Amino Acid in SvThrRS
N312	Y334	H330	W329
F461	H488	Y482	F481
S488	Y515	N509	N508
L489	T516	Q510	L509

In order to test if these amino acid switches in EcThrRS can confer borrelidin resistance in *E. coli* NR698, EcThrRS was mutated at these positions. All 15 possible combinations of single, double, triple and quadruple mutations in EcThrRS were made. Of these, it was found that the L489T mutation alone was sufficient to confer resistance to borrelidin, as seen in Figure 3.17. All EcThrRS

mutants containing an L489T mutation were resistant to borrelidin, while mutants missing this mutation were sensitive to borrelidin (Figure 3.18).

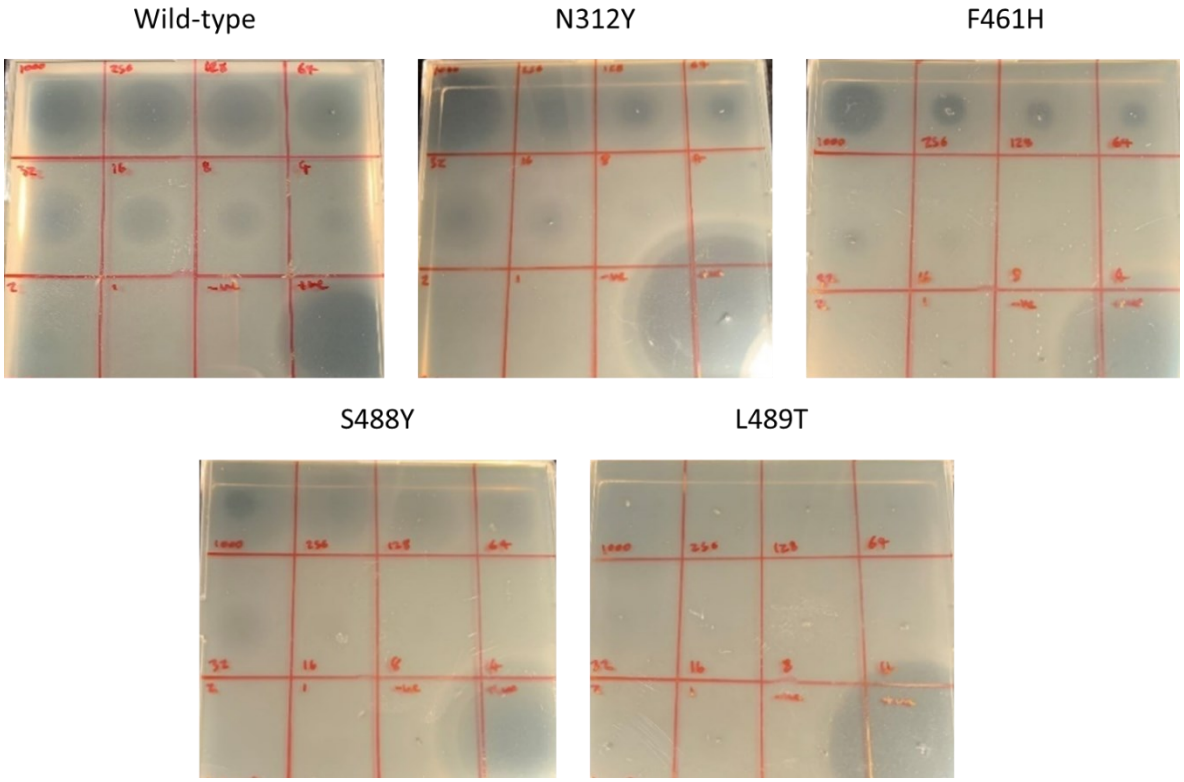


Figure 3.17. An L489T mutation is sufficient to make EcThrRS resistant to borrelidin. *E. coli* NR698 expressing EcThrRS with N312Y, F461H, S488Y or L489T point mutations were inoculated into soft nutrient agar to pour the plates before spotting on of 4 µL borrelidin (dissolved in DMSO) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of DMSO only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates incubated at room temperatrue for 16 hours. Of these single point mutations, L489T alone was sufficient to make EcThrRS fully resistant to borrelidin, up to 1 mg/mL.

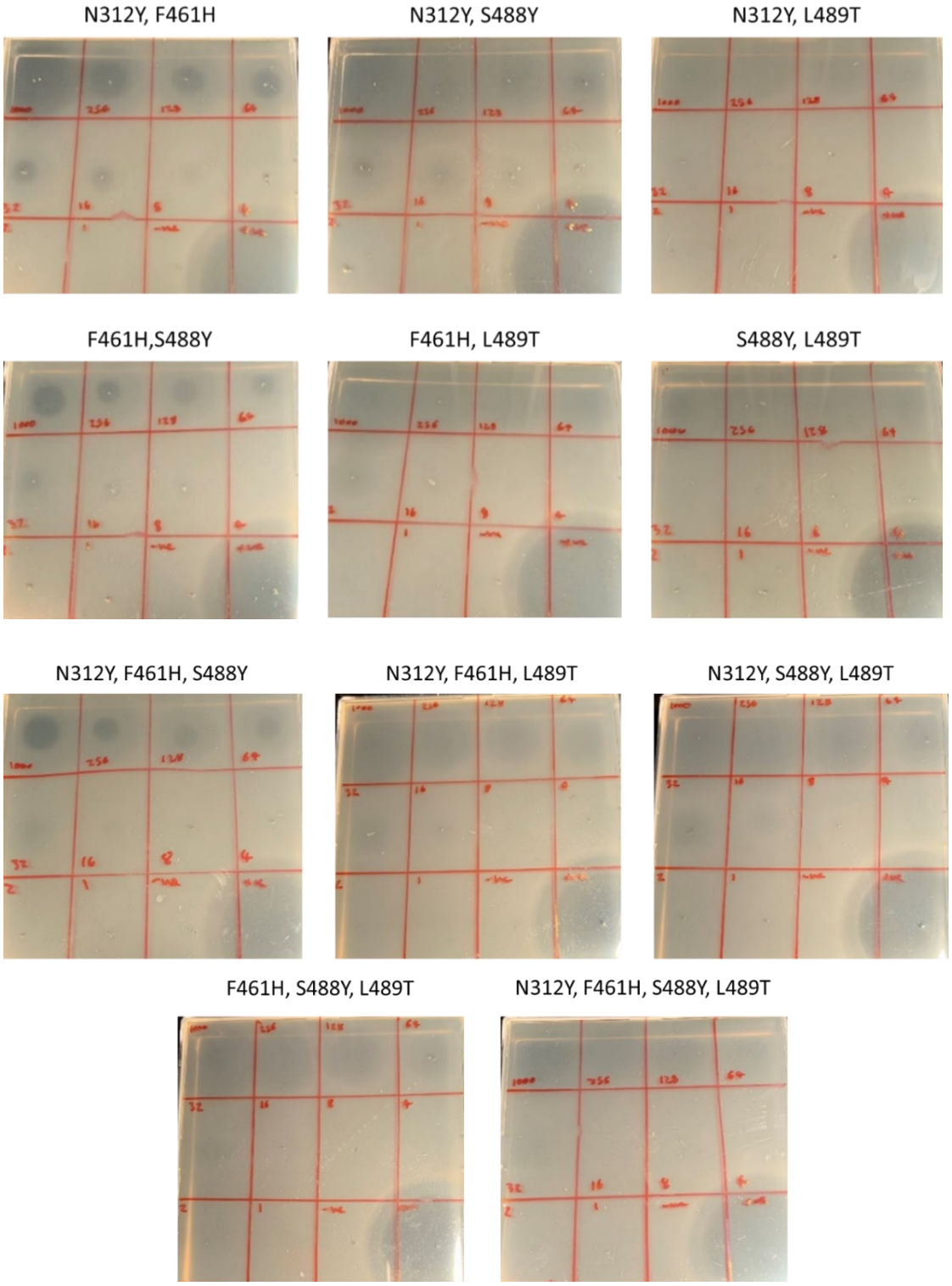


Figure 3.18. All *E. coli* NR698 strains expressing EcThrRS with a L489T mutation are resistant to borrelidin. *E. coli* NR698 expressing EcThrRS with combinations of N312Y, F461H, S488Y and L489T point mutations were inoculated into soft nutrient agar to pour the plates before spotting on of 4 µL borrelidin (dissolved in DMSO) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of DMSO only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates incubated at room temperature for 16 hours. All of the strains with L489T were resistant to borrelidin, up to 1 mg/mL.

In the BorO structure, T516 (the equivalent to L489 in EcThrRS) appears to be hydrogen bonded to the nearby D486 (Figure 3.19). The D486 sidechain appears to be involved in borrelidin binding by hydrogen bonding to the hydroxyl group near the nitrile moiety. D486 is completely conserved in all ThrRSs, including the *Archaeal* ThrRSs which are the most dissimilar to bacterial ThrRSs, suggesting that it is important for structure or function in these proteins. In order to verify the role of this residue in borrelidin binding, a D486A mutation was introduced to *EcThrRS* in pJH10TS and transformed into *E. coli* NR698. When challenged with borrelidin, the strain expressing EcThrRS D486A was fully resistant (Figure 3.20). This would suggest to us that this interaction is vital for binding of borrelidin in EcThrRS.

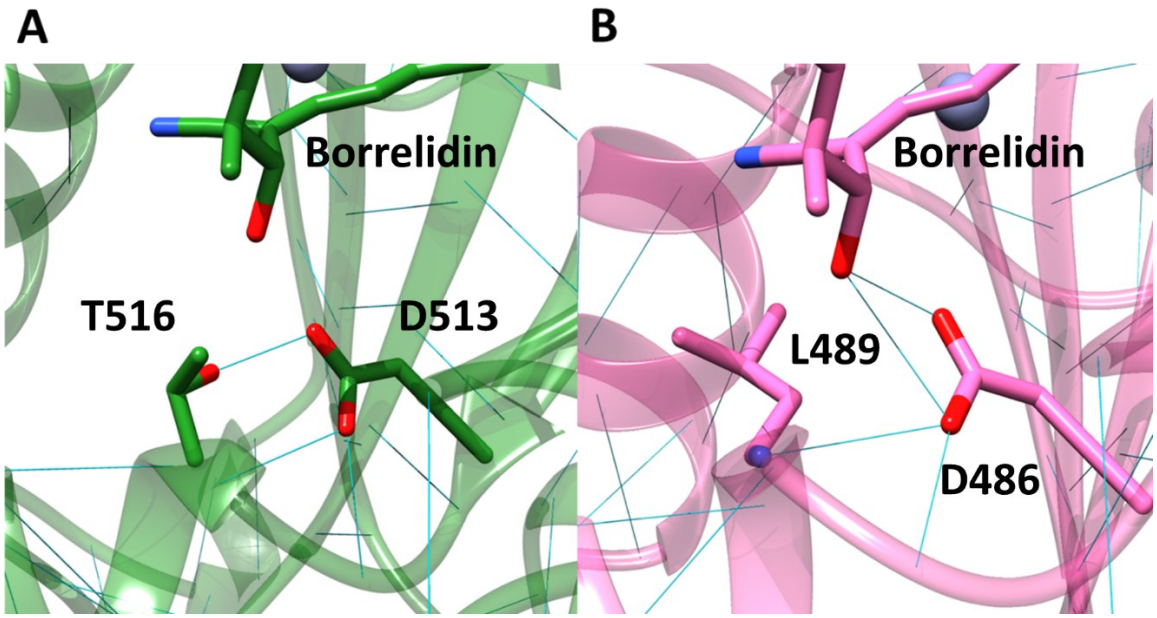


Figure 3.19. D486 hydrogen bonds to borrelidin in EcThrRS, but not in BorO. Hydrogen bonds shown as blue lines, identified using the FindHBond tool in Chimera. **A)** crystal structure of BorO with borrelidin bound, T516, D513 and borrelidin shown as sticks. D513 hydrogen bonds to the backbone of the protein and the β-hydroxyl of T516. **B)** crystal structure of EcThrRS with borrelidin bound, L489, D486 and borrelidin shown as sticks. D486 hydrogen bonds to the backbone of the protein and the hydroxyl group of borrelidin.

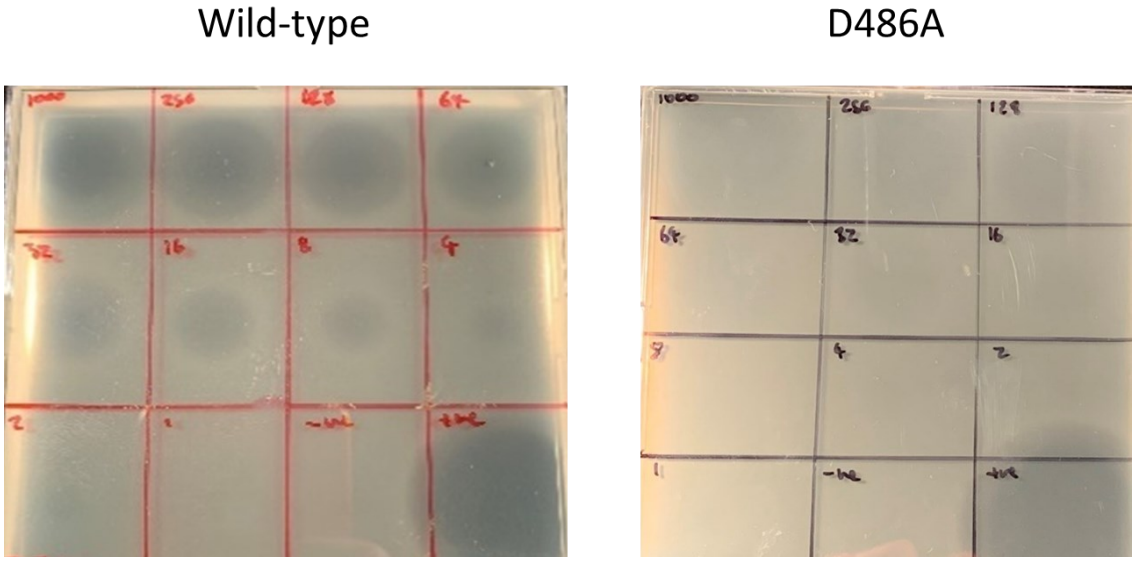


Figure 3.20. A D486A mutation is sufficient to make EcThrRS resistant to borrelidin. *E. coli* NR698 expressing EcThrRS WT or with D86A point mutation were inoculated into soft nutrient agar to pour the plates before spotting on of 4 µL borrelidin (dissolved in DMSO) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of DMSO only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates incubated for 16 hours at room temperature. EcThrRS D486A appears to confer resistance to borrelidin.

3.2.10 EcThrRS L489T does not bind to borrelidin

In order to further explore these resistance profiles, *EcThrRS* (L489T), *EcThrRS* (D486A) and *borO* (T516L) were cloned into pET28 for purification of full-length proteins, and the N-terminal editing domain was truncated as before and cloned into pET29 for purification of ΔN protein for crystallography. While both full-length and ΔN EcThrRS L489T expressed and could be purified in titres similar to the WT protein, both full-length and ΔN EcThrRS D486A could only be purified in small amounts, suggesting a problem with protein folding. BorO T516L also expressed less well than wild-type BorO, although BorO and SpThrRS both expressed significantly less well than EcThrRS.

The quality of the protein samples was checked by mass photometry to ensure that dimeric protein was present, and that the formation of higher order aggregates was not an issue (see Figure 3.21). For EcThrRS L489T, both dimeric and monomeric forms can be observed, but no higher order multimers are present. However, L486A was primarily monomeric, with many higher-order oligomers being detected, suggesting that there was a problem with folding of this protein. BorO T516L was found as both a monomer and dimer; however, the peaks do not look as smooth as for the WT protein, suggesting that there may be additional oligomerisation or some problems with this protein. Some higher order oligomers of mass roughly 320 are also present in the histogram, suggesting some protein aggregation.

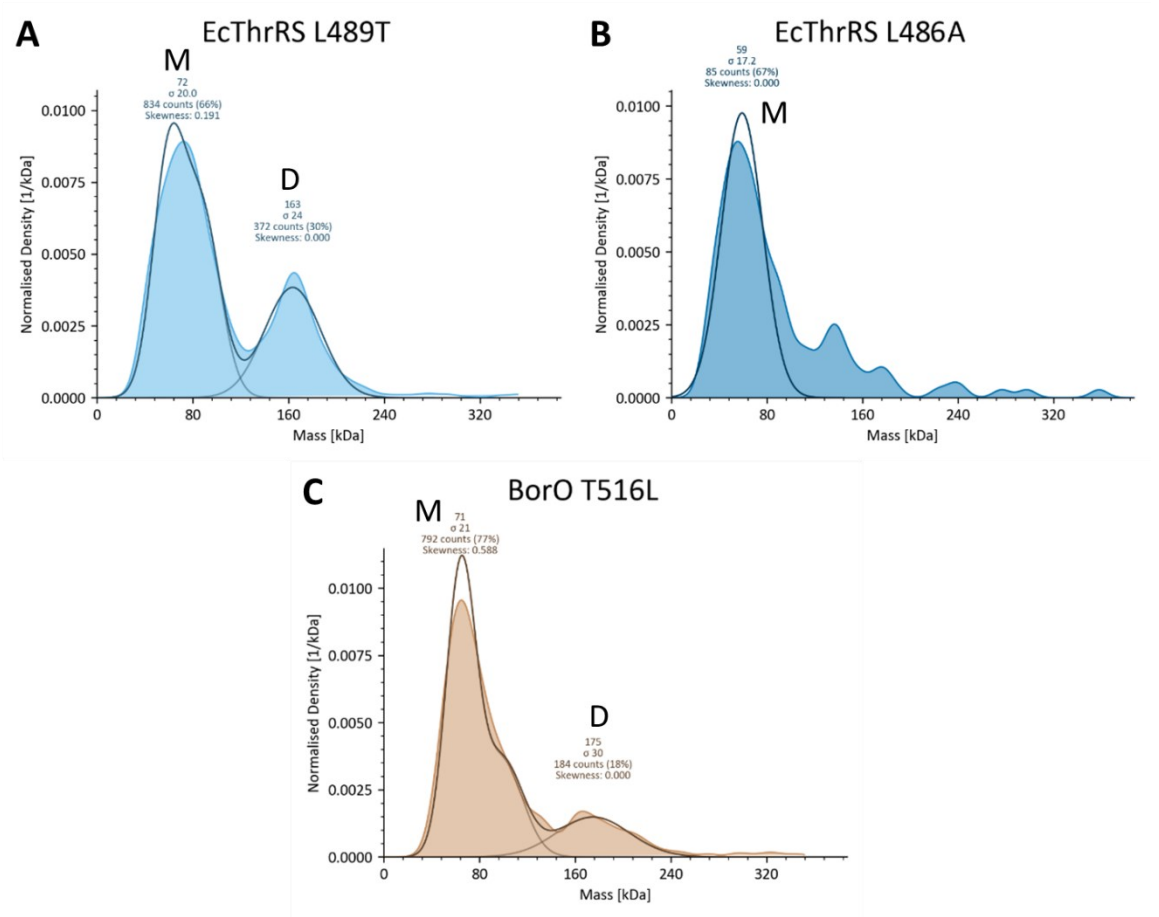


Figure 3.21. Mass photometry analysis of isolated proteins. Histograms of A) EcThrRS L489T, B) EcThrRS L486A and C) BorO T516L. The average predicted mass in kDa is shown above each peak; with a protein concentration of 100 nM used for all three proteins. Proteins were measured in PBS. Peaks align to the sizes of the monomer (M) and the dimer (D) for EcThrRS L489T and roughly for BorO T516L. Presence of the dimer indicates the presence of functional protein. EcThrRS L486A appears to be mainly made up of monomer and various higher order oligomers, suggesting that there is a problem with folding or that it is folded in a way that prevents dimerisation.

Binding of ThrSAA and borrelidin by EcThrRS L489T, EcThrRS D486A and BorO T516L was then assayed via ITC (Figure 3.22). Of these, EcThrRS L489T showed binding to ThrSAA with a K_d of 27.4 ± 9.10 nM (n = 3) and no binding to borrelidin (n = 3). We can therefore conclude that a L489T mutation in EcThrRS is sufficient to abolish binding to borrelidin. EcThrRS D486A showed no binding to the positive control, suggesting that the protein was not properly folded. BorO T510L did not give good enough ITC data for meaningful analysis, and with poor expression levels, it was impractical to purify enough protein for further ITC runs. The lack of binding to the positive control compound would suggest that EcThrRS D486A is not properly folded, and so further *in vitro* assays are impossible with this mutant protein.

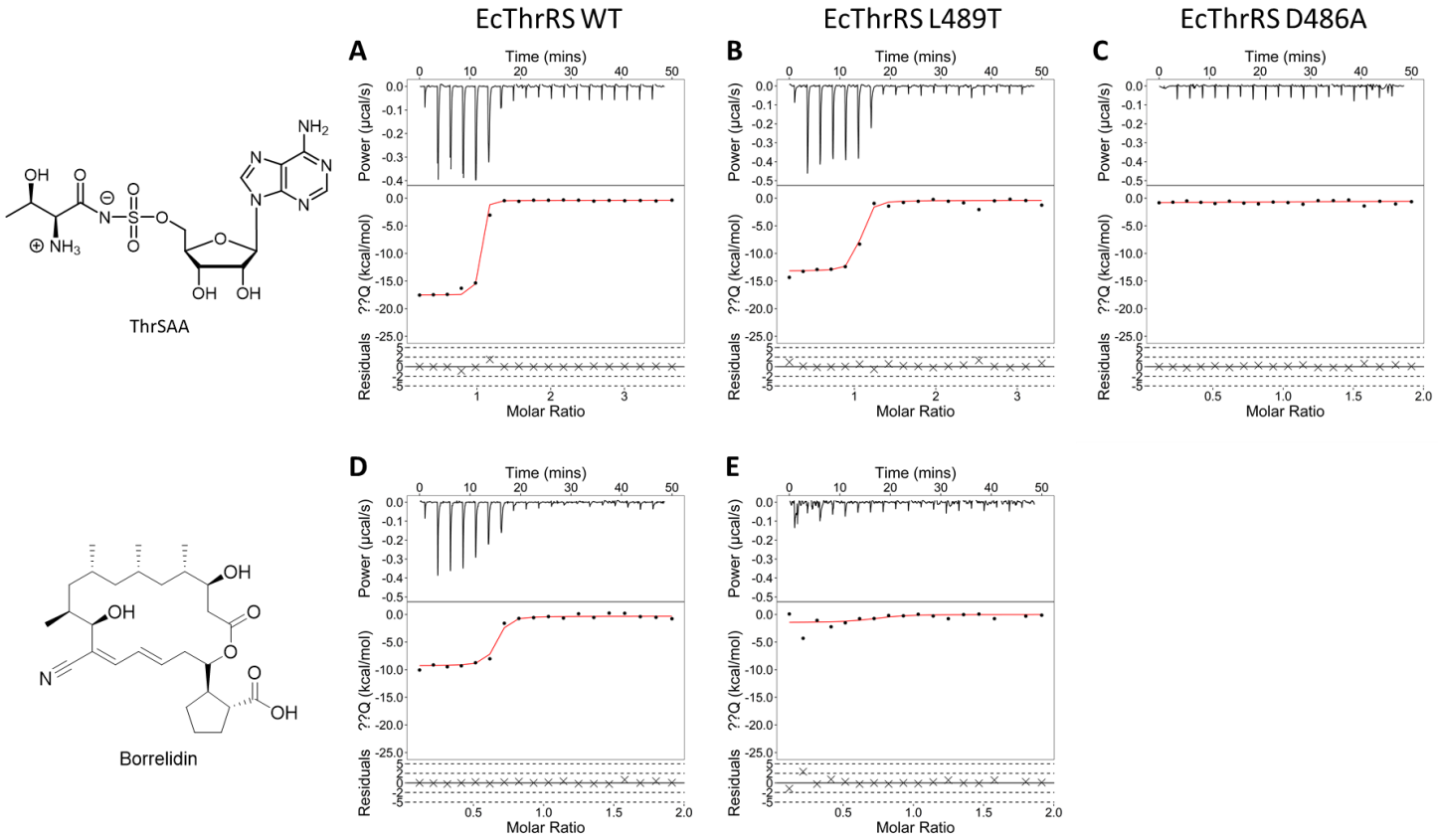


Figure 3.22. EcThrRS L489T does not bind to borrelidin in ITC binding assays, while EcThrRS D486A does not bind to the positive control compound. Each lettered dataset shows the isotherms of the reaction in the top panel, the resulting binding curve in the middle panel, and the residuals (which measures the goodness of fit of the curve to the data) in the bottom panel. **A)** Wild-type EcThrRS with ThrSAA, **B)** EcThrRS L489T with ThrSAA, **C)** EcThrRS D486A with ThrSAA, **D)** Wild-type EcThrRS with Borrelidin, **E)** EcThrRS L489T with borrelidin. Only WT EcThrRS and EcThrRS L489T show binding to the intermediate analog, ThrSAA, suggesting that they are folded correctly. Only WT EcThrRS shows binding to borrelidin. Figure generated in RStudio.

3.2.12 EcThrRS L489T borrelidin-bound structures.

To obtain structural data, crystal trials were prepared for EcThrRS D486A, EcThrRS L489T, EcThrRS N312Y:L489T, EcThrRS S488Y:L489T and BorO T510L. The PEGs screen (Qiagen) was used for all mutant proteins, with EcThrRS mutants at 20 mg/mL. The BorO mutant protein was trialled at 7 mg/mL with seeding with wild-type BorO seed crystals. The buffer used was Lysis buffer containing 10 mM threonine (details in section 2.2.2) and protein was prepared both with and without 2 mM borrelidin.

Crystals could be grown for all the EcThrRS mutant proteins with both threonine and borrelidin bound, except for EcThrRS D486A. As for the full-length EcThrRS D486A protein, it is likely that this mutant was not properly folded. All of these crystals were then sent to DLS for X-ray diffraction. Crystals of BorO T510L could be grown in the presence of borrelidin, but not in the presence of threonine alone. The BorO T510L and EcThrRS N312Y:L489T crystals did not diffract well enough for a structure to be solved.

Structures of EcThrRS L489T with threonine bound could be solved at a resolution of 2.7 Å, and with borrelidin bound at 2.2 Å. A structure from the EcThrRS S488Y:L489T crystal was solved to a resolution 2.85 Å with a pair of dimers in the asymmetric unit, with each dimer consisting of one borrelidin bound and one threonine bound monomer.

As argued for the BorO structure, the observed borrelidin binding could be an artefact of the crystallography conditions. At the concentrations used for ITC binding assays, in solution, borrelidin likely cannot bind, primarily due to the interaction of T489 with D486 which constrains the structure, and which must be broken for the conformational change required for borrelidin binding. Indeed, in all four L489T containing monomers (the dimer from the L489T structure, and one monomer from each dimer in the S488Y:L489T structure), the key hydroxyl group (on C11 of borrelidin, see Figure 3.2) of borrelidin is hydrogen bonded to both the aspartic acid and the threonine (Figure 3.23).

It is crucial that in all of these EcThrRS mutant crystal structures, borrelidin is bound in the same binding mode that it is in the wild-type EcThrRS crystal structure. This means that while this single point mutation can confer resistance to borrelidin on its own, there are likely other key mutations which are contributing to borrelidin resistance in BorO, which are not sufficient alone or have not been tested in this analysis. In order to assess the effect of the T516L mutation on BorO sensitivity to borrelidin, BorO T516L has been cloned into pMB743 for expression in *S. venezuelae*. However, time did not allow for generation and testing of this strain. It is likely that BorO has developed resistance which requires multiple mutations to disassemble.

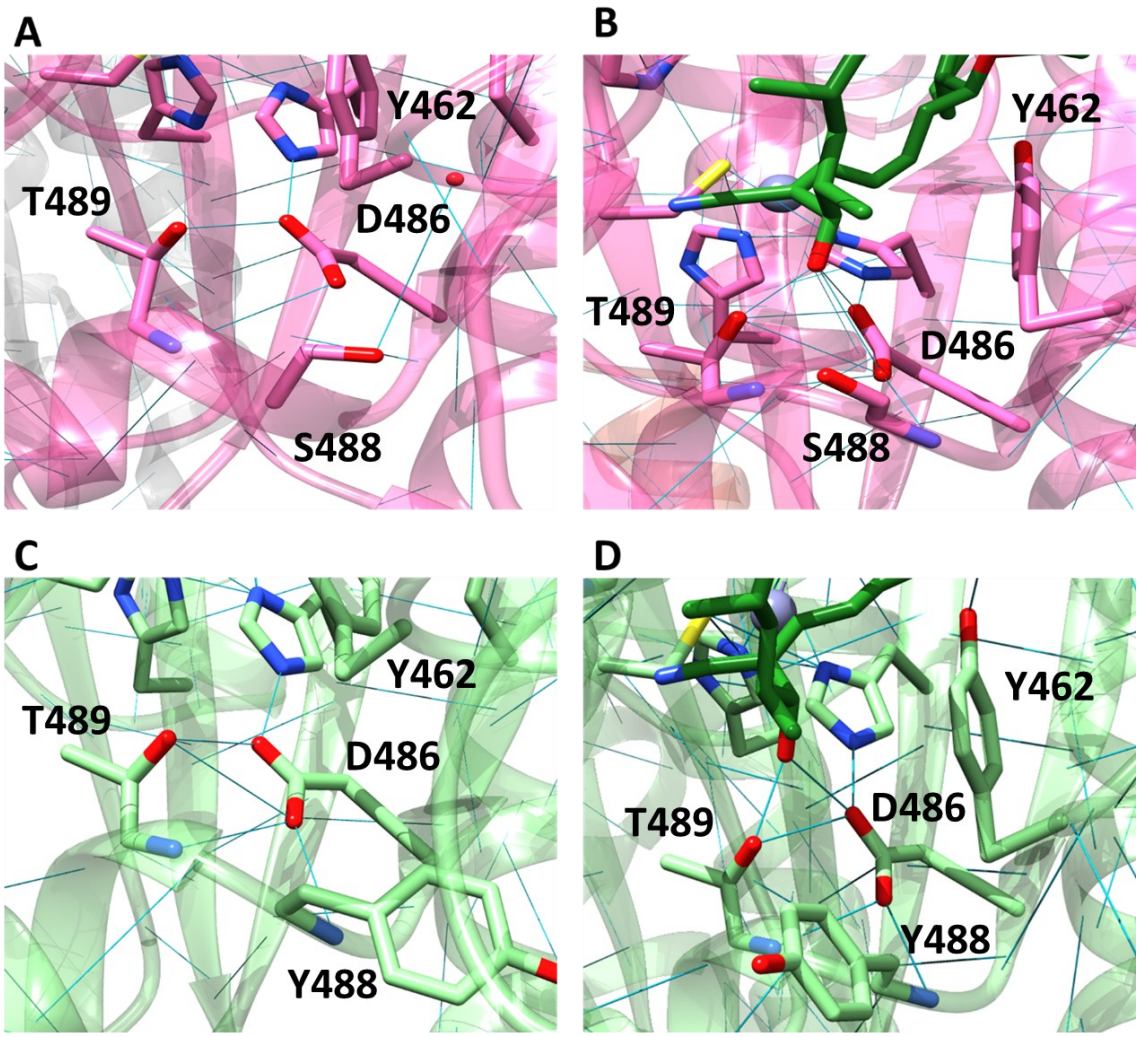


Figure 3.23. T489 in EcThrRS L489T hydrogens bonds to D486. A) Threonine bound structure of EcThrRS L489T**, B)** Borrelidin bound structure of EcThrRS L489T. **C)** Threonine bound structure of EcThrRS S488Y:L489T, **D)** Borrelidin bound structure of EcThrRS S488Y:L489T. Borrelidin shown in green, hydrogen bonds shown as blue lines, identified using the FindHBond tool in Chimera. In both structures, D486 is hydrogen bonding to the backbone of the protein, the β-hydroxyl of T489 and H511.

Previous work to mutate EcThrRS for borrelidin resistance showed that L489M and L489W were sufficient for borrelidin resistance *in vivo*, caused by a steric hinderance, resulting in the blocking of the hydrophobic borrelidin binding site[161]. The authors suggested that position 489 is the amino acid position that is most important for borrelidin resistance. Our data generally supports this hypothesis. However, our data also suggests that the way in which the self-resistance protein BorO conveys resistance is more subtle, by alteration of H-bonding in the borrelidin binding site as opposed to sterically filling it. BorO is resistant to borrelidin due to H-bonding between residue D513 (D486 in EcThrRS) and the threonine residue T516; this interaction is mimicked by the hydroxyl group at C11 of borrelidin. This part of the protein also appears to act as a hinge, allowing

for the conformational change required for borrelidin resistance. This hydrogen bonding between threonine and aspartic acid residues may aid in pinning the borrelidin binding site closed, preventing conformational change and in turn preventing borrelidin binding.

3.2.11 EcThrRS L489Q is not resistant to borrelidin

It has become clear that T516 (L489 in EcThrRS) is a key residue for borrelidin resistance in BorO. The amino acid in the equivalent position in SpThrRS is Q510. In the AlphaFold models for SpThrRS, the three top ranked models appear to have the nitrogen of the side chain amide of Q510 at the appropriate distance and orientation for hydrogen bonding to D507, the equivalent to D513 in BorO and D486 in EcThrRS (Figure 3.24). This interaction could mimic the effect observed for the interaction between T516 and D513 that we have just discussed for BorO; we know the L489T mutation is sufficient to confer borrelidin resistance upon EcThrRS, so we hypothesised that this residue alone could be conferring the borrelidin resistance observed for SpThrRS. This was assayed via the expression of an EcThrRS L489Q mutant in *E. coli* NR698 (using the pJH10TS vector). However, this mutation was not sufficient to confer full resistance to borrelidin as seen in Figure 3.25.

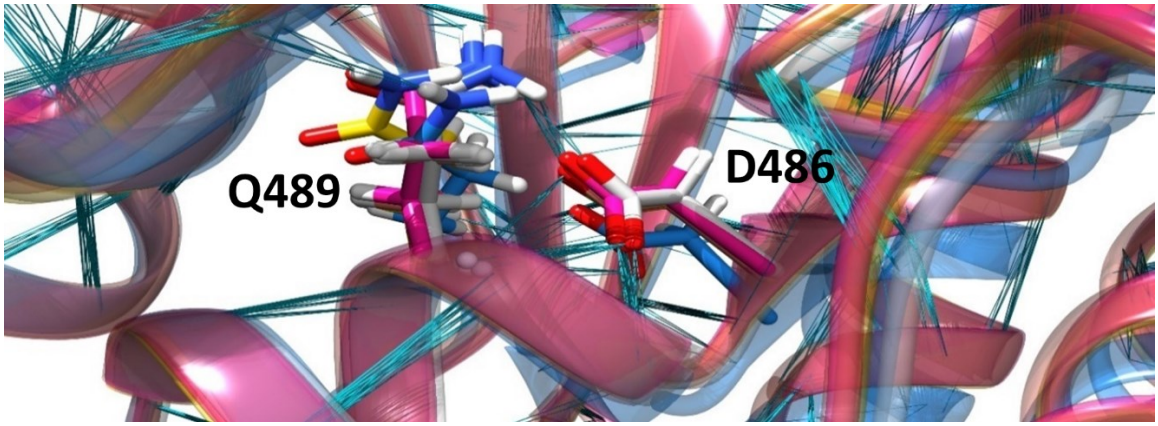


Figure 3.24. In some of the SpThrRS AlphaFold models, Q510 is hydrogen bonded to D507. The five models generated by AlphaFold are structurally aligned. Hydrogen bonds shown as blue lines, identified using the FindHBond tool in Chimera. Q510 and D507 shown as sticks. D486 hydrogen bonds to the backbone of the protein in all five models and to the amide of Q510 in 3 out of 5 models.

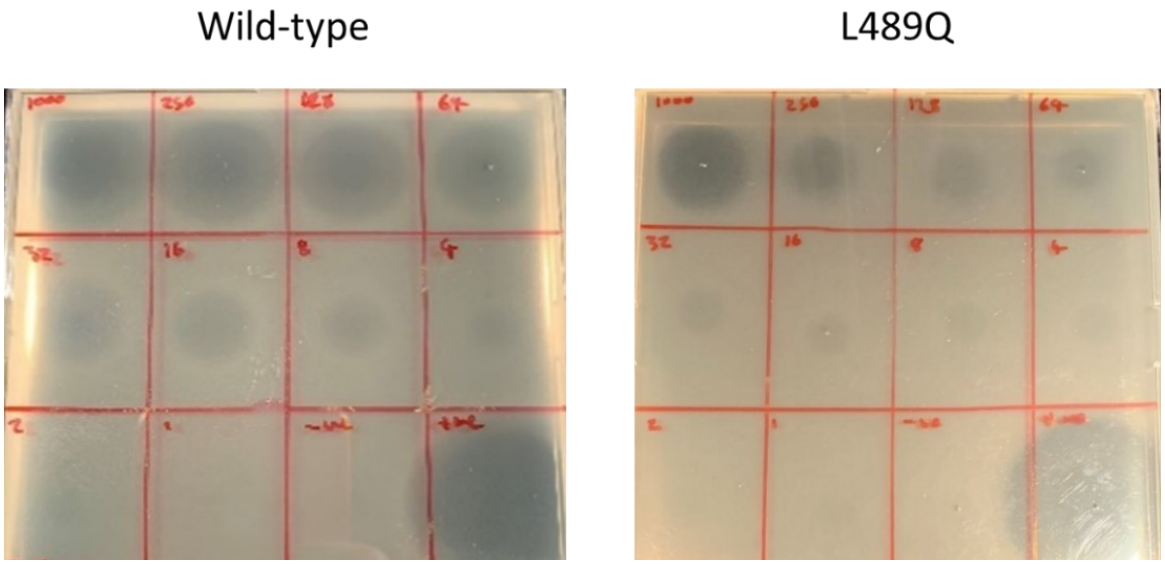


Figure 3.25. An L489Q mutation in EcThrRS is insufficient to confer resistance to borrelidin. *E. coli* NR698 expressing EcThrRS WT or the L489Q point mutant were inoculated into soft nutrient agar to pour the plates before spotting on of 4 µL borrelidin (dissolved in DMSO) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of DMSO only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates incubated for 16 hours at room temperature. EcThrRS L489Q does not appear to confer resistance to borrelidin.

3.2.12 EcThrRS L489Q and SpThrRS Q510L both bind to borrelidin

In order to assay the effect of the L489Q mutation on the binding of borrelidin by EcThrRS, the L489Q mutation was introduced to *EcThrRS* in pET28 allowing for the production of full-length N-terminal poly-His tagged protein. As well as this, the L489Q mutation was introduced into a ΔN poly-His tagged version of the protein which lacks the N-terminal editing domain. This truncated version was cloned into pET29 for production, and used for crystallography. A Q510L mutation was also introduced into *SpThrRS* and the SpThrRS Q510L protein purified to test if Q510 is a factor which prevents borrelidin binding in SpThrRS. The quality of the protein samples was checked by mass photometry to ensure that dimeric protein was present, and that the formation of higher order aggregates was not an issue (see Figure 3.26).

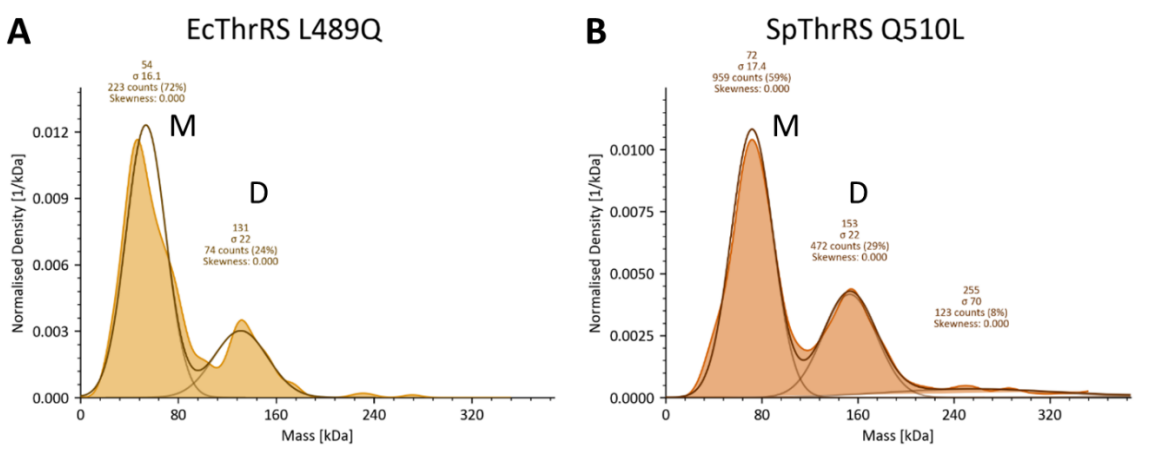


Figure 3.26. Mass photometry analysis of isolated proteins. Histograms of A) EcThrRS L489Q, B) SpThrRS Q510L. The average predicted mass in kDa is shown above each peak; with a protein concentration of 100 nM used for all three proteins. Proteins were measured in PBS. Peaks align to roughly the sizes of the monomer (M) and the dimer (D) for both proteins, with some small amounts of higher order oligomers also present, suggesting some aggregation. Presence of the dimer indicates the presence of functional protein.

When assayed by ITC (Figure 3.27), EcThrRS L489Q performed almost identically to wild-type EcThrRS with a K_d for ThrSAA binding of 18.7 ± 3.02 nM (n =3) and K_d for borrelidin binding of 14.4 ± 2.45 nM (n = 3). SpThrRS Q510L was also able to bind borrelidin; SpThrRS Q516L binds ThrSAA with a K_d of 19.9 ± 2.06 nM and binds to borrelidin with a K_d of 261 ± 16.8 nM. SpThrRS Q510L, while binding borrelidin, does so significantly more weakly than EcThrRS or EcThrRS L489Q. This would suggest that alone, Q510 cannot confer resistance, but it is involved somehow in the prevention of binding of borrelidin to SpThrRS, possibly alongside another mechanism. In SpThrRS, the amide from Q510 may be forming a hydrogen bond with the carbonyl of D507, but the amide-carbonyl hydrogen bond will be weaker than the hydroxyl-carboxyl hydrogen bond between T516 and D513 in BorO (the equivalent positions), necessitating a different resistance mechanism in SpThrRS.

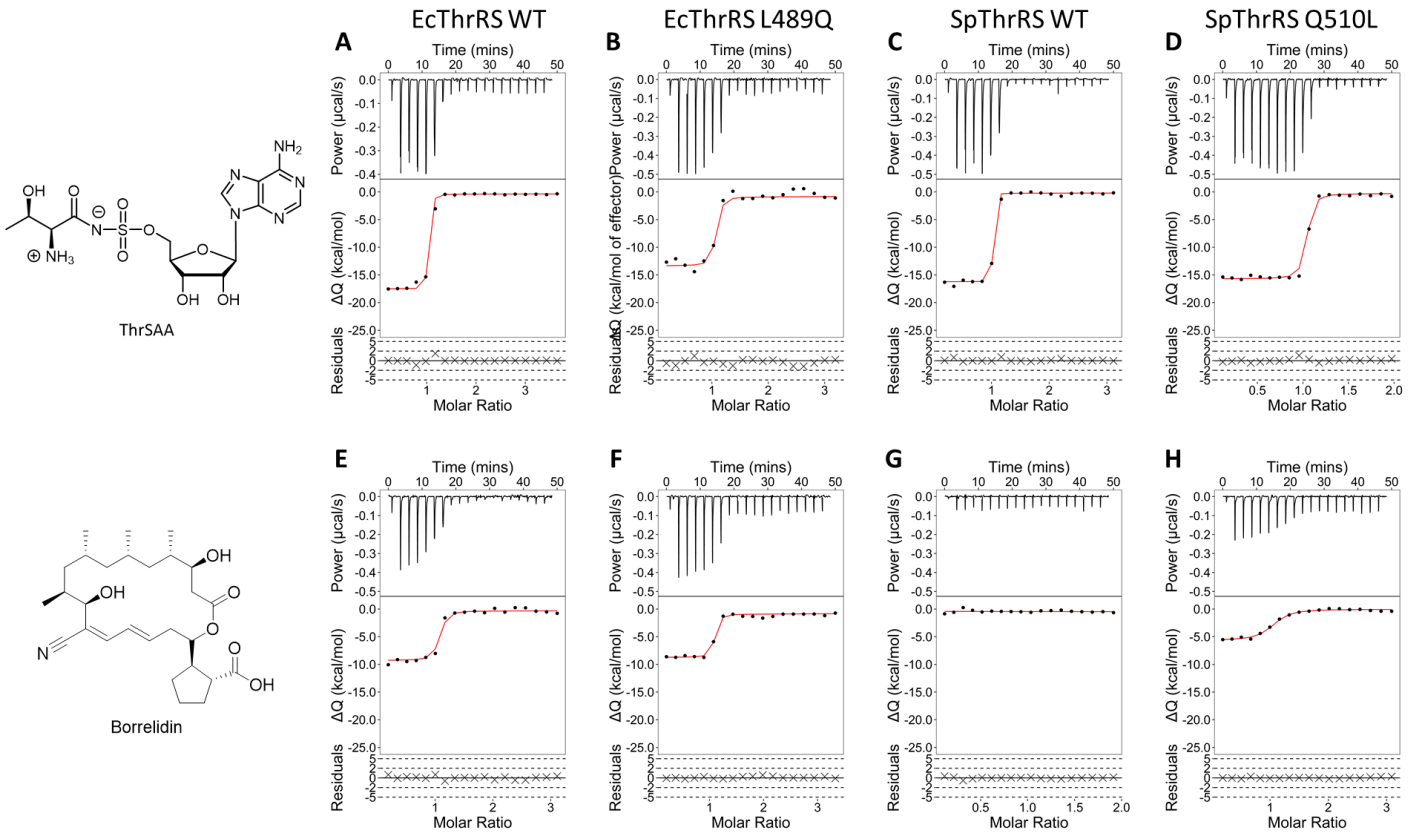


Figure 3.27. EcThrRS L489Q and SpThrRS Q510L both bind to borrelidin in ITC binding assays. Each lettered dataset shows the isotherms of the reaction in the top panel, the resulting binding curve in the middle panel, and the residuals (which measures the goodness of fit of the curve to the data) in the bottom panel. **A)** WT EcThrRS with ThrSAA, **B)** EcThrRS L489Q with ThrSAA, **C)** WT SpThrRS with ThrSAA, **D)** SpThrRS Q510L with ThrSAA, **E)** WT EcThrRS with Borrelidin, **F)** EcThrRS L489Q with borrelidin, **G**) WT SpThrRS with borrelidin, **H)** SpThrRS Q510L with borreldiin. All four proteins show binding to the intermediate analog, ThrSAA, suggesting that they are folded correctly. Both WT and L489Q EcThrRS show binding to borrelidin, as does SpThrRS Q510L. Figure generated in RStudio.

As a proof of concept and to confirm that EcThrRS L489Q binding to borrelidin prevents binding of ThrSAA, ITC competition assays with EcThrRS L489Q were performed (Figure 3.28). A protein-borrelidin complex was made by pre-incubating the protein with borrelidin at a 1:1 molar ratio, at a concentration of 20 µM. An ITC run was then performed with the protein-borrelidin complex in the cell, and ThrSAA injected in as a proxy for the substrates of the reaction was then performed. If ThrSAA was able to displace borrelidin as the more competitive EcThrRS-L489Q substrate, we would expect to see an ITC curve that reflects this displacement and binding. However, no binding events were observed, suggesting that in this protein, borrelidin can outcompete ThrSAA, and therefore for EcThrRS L489Q, borrelidin is a more favourable ligand than ThrSAA. As an additional control it would be useful in future to repeat this competition assay with wild-type EcThrRS and at least some proteins which do not bind to borrelidin, such as BorO or EcThrRS L489T. An example, with EcThrRS L489M, which can confer borrelidin resistance in *E. coli* NR698 can be seen in Figure 4.10 in section 4.2.5, in which ThrSAA is a more favourable substrate than borrelidin.

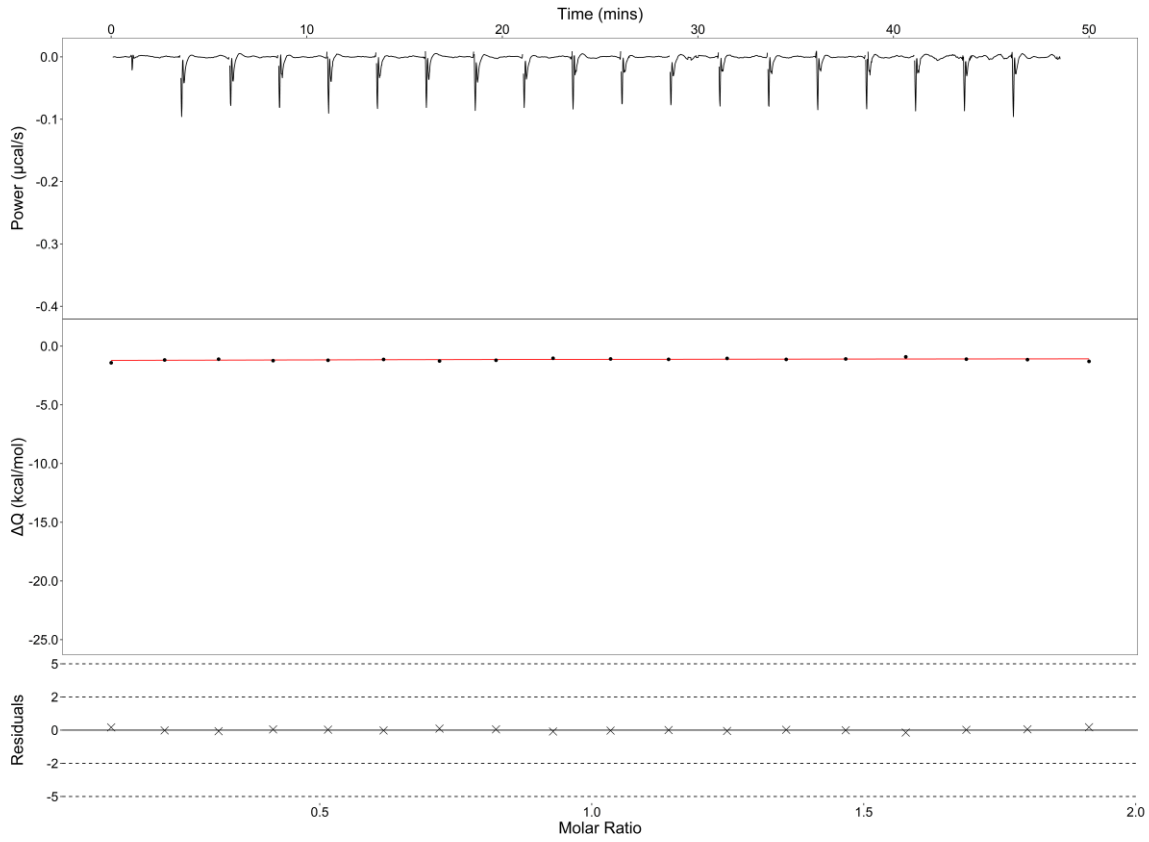


Figure 3.28. ITC plots for competition binding experiment for EcThrRS L489Q pre incubated with borrelidin against ThrSAA. The isotherm of the reaction is in the top panel, the resulting binding curve in the middle panel, and the residuals which measures the goodness of fit of the curve to the data in the bottom panel. EcThrRS L489Q was preincubated with borrelidin as if an ITC experiment had been undertaken. An ITC run was then performed by addition of ThrSAA using the syringe as normal. The resulting curve shows that ThrSAA is unable to bind to EcThrRS L489Q following incubation with borrelidin. Figure generated in RStudio.

3.2.13 Crystallography of EcThrRS L489Q and SpThrRS Q510L

A structure of EcThrRS L489Q bound to borrelidin could be solved to 1.5 Å, representing the highest resolution structure of a ThrRS mutant collected to date. For all of the EcThrRS mutants crystallised in this study, crystals grew in both the presence and absence of borrelidin, but diffraction was of demonstrably better quality in the presence of borrelidin.

In the structure of EcThrRS L489Q, we can see that Q489 bends away from D486, reminiscent of L489 in the WT protein, when borrelidin is bound (see Figure 3.29). The carbonyl of the side chain of L489Q also appears to be hydrogen bonded to the hydroxyl group on C11 of borrelidin. This suggests to us that the hypothesised interaction between D486 and Q489 is not happening- in BorO this interaction (between D513 and T516) is key for blocking borrelidin binding.

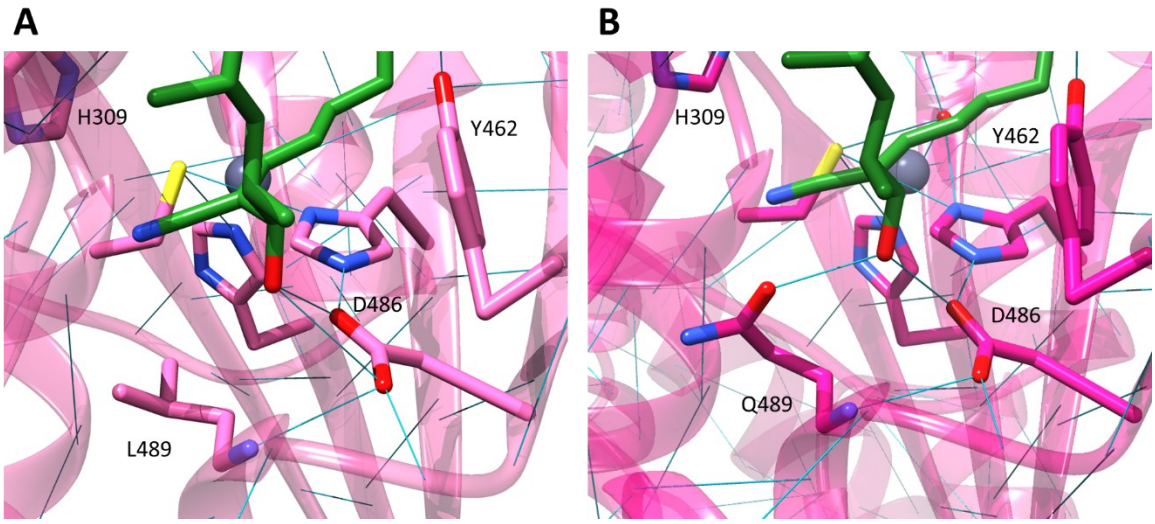


Figure 3.29. Q489 in EcThrRS L489Q does not hydrogen bond with D486 when borrelidin is bound. A) EcThrRS L489Q; B) EcThrRS WT. Hydrogen bonds are shown as blue lines, identified using the FindHBonds tool in Chimera. Q489 appears to be interacting with borrelidin but not D486. L489 does not interact with D486.

As with WT SpThrRS, no crystals could be grown for SpThrRS Q516L. After the work done in section 4.2.7, it had been hoped that improving the ability of SpThrRS to bind to borrelidin could improve the crystallisation of SpThrRS. In the case of ObaO, the creation of a mutant protein that is able to bind to borrelidin led to the successful resolution of a structure being solved.

3.2.14 The possible auxiliary resistance residue in SpThrRS can be identified from sequence and structural alignments

With no crystal structure of SpThrRS being available, we still wanted to identify additional possible self-resistance residues in SpThrRS. The AlphaFold model of SpThrRS was structurally aligned with that of EcThrRS bound to borrelidin and all residues within 5Å of borrelidin were selected. The resulting list of residues was then reduced by removal of those that were identical in both proteins.

The remaining residues were then manually screened for potential interactions which could be stabilising SpThrRS in a borrelidin binding pocket closed conformation. This list consisted of just five residues: H330, N357, Y482, N509 and Q510 in SpThrRS. This list corresponds to the exact positions of the four residues which were originally suggested in Fang *et al.* 2015[110] as the BorO resistance residues, plus N357 (see Table 3.3).

Table 3.3. The amino acid positions addressed in this analysis. Each row represents the same amino acid position in each protein.

SpThrRS	EcThrRS	SvThrRS	BorO	ObaO	SaThrRS
H327	H309	H326	H331	H308	H309
H330	N312	W329	Y334	N311	H312
N357	V338	N356	H361	I337	M340
Y482	F461	F481	H488	F462	F467
N509	S488	N508	Y515	H489	L494
Q510	L489	L509	T516	M490	L495

A sequence alignment of SaThrRS, SpThrRS, EcThrRS, SvThrRS, BorO and ObaO was then examined at these five positions; we know that SaThrRS, SvThrRS and EcThrRS are able to bind to borrelidin, and we know that ObaO only does not bind to borrelidin because of the methionine in position 489, relative to EcThrRS (as discussed in more detail in section 4.2.4-7), and we know that BorO and SpThrRS do not bind to borrelidin. We already know that position Q510 has a role in the blocking of borrelidin binding from the work done in section 3.2.12.

From this, we can hypothesise that Y482 is able to hinder the ability of SpThrRS to bind to borrelidin, possibly alongside Q510, which together could be conferring borrelidin resistance. The key interaction here could be H-bonding of the phenolic hydroxyl group of Y482 with H327 (Y461 and H309 in EcThrRS)- the AlphaFold structure can be seen in Figure 3.30. This interaction would be a pinning interaction holding the two halves of the protein together, keeping the active site in a more closed conformation and preventing the opening of the borrelidin binding site. In addition, the neighbouring residue, Y481 (Y462 in EcThrRS) has been shown to be important for threonine binding, $tRNA^{Thr}$ binding and the aminoacylation reaction itself, as discussed in section 1.3.1. Additionally, in section 4.2.11, it was demonstrated that Y462 (relative to EcThrRS) is essential for ThrRS activity. Additionally, H309 is completely conserved in ThrRSs because it is the key residue which catalyses the aminoacylation reaction[115].

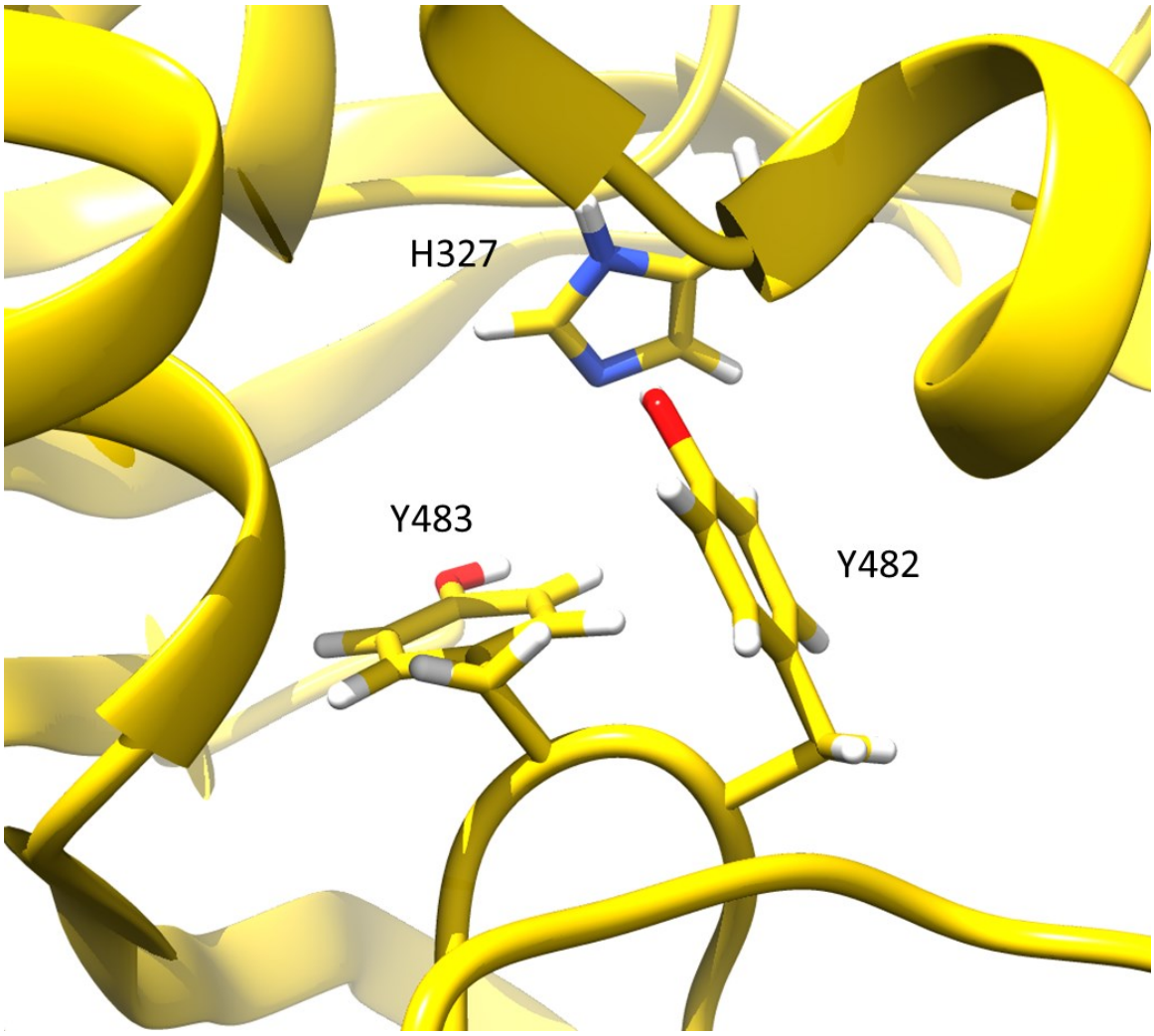


Figure 3.30. Y482 in SpThrRS may interact with H327 to keep the borrelidin binding site closed. The AlphaFold model for SpThrRS was used for this, Y483 and H327 (Y462 and H309 in EcThrRS) are essential for catalysing the aminoacylation reaction. Figure generated in Chimera.

Examination of a global alignment of ThrRS protein sequences (discussed in detail in chapter 5), shows that tyrosine in the 461 position (relative to EcThrRS) is only found in a handful of bacterial ThrRSs, and the Archaeal ThrRSs. Phenylalanine is highly conserved in this position in all other sequences except the BorOs, which all have a histidine in this position. Looking at the borrelidin resistance position (L489 in EcThrRS) in those proteins containing Y461 (relative to EcThrRS), in the Archaeal ThrRSs, we can find either an asparagine or a serine, while in the bacterial ThrRSs, there is always a glutamine in position 489 (relative to EcThrRS). This could give us a potential insight into the evolution of the SpThrRS "flavour" of borrelidin resistance.

3.3 Conclusions and Future Work

The work presented in this chapter explores the self-resistance mechanisms of the borrelidin producing strain *S. parvulus*. It would appear that, in *S. parvulus*, both BorO and SpThrRS are able to convey resistance to borrelidin, which is not only unexpected but also, to my knowledge, rarely observed. The only other report of a similar phenomenon is in the platensimycin producer, *S.*

platensis. In this strain, the BGC associated self-resistance gene, *ptmP3* can replace both the *fabF* and *fabH* genes, functionally. The housekeeping FabF in this strain is, however also resistant to platensimycin, while the FabH is sensitive[72]. From our current understanding of self-resistance mechanisms, it seems unusual that a BGC would contain a self-resistance determinant if the producing organism is already resistant. This raises the question as to why *S. parvulus* has a BGC-associated ThrRS, and if it has a role beyond self-resistance. Work in chapter 5 will begin to explore this; however, it would be interesting to undertake a full phylogenetic reconciliation of BorO. This technique allows an inference of the time that certain traits emerged to be given and was recently used by the Wright lab to explore the evolutionary history of glycopeptide antibiotic biosynthesis. In performing this analysis, two new glycopeptide antibiotics with unique bioactivity mechanisms were discovered[234,235].

3.3.1 The self-resistance mechanism of BorO

From this work, we can identify at least part of the self-resistance mechanism employed by BorO to prevent binding of borrelidin: we propose that T516 (L489 in EcThrRS) binds to D513 (D486 in EcThrRS) through a hydrogen bonding interaction. This interaction occurs in a hinge region of the protein, about which the borrelidin binding site can open and facilitate borrelidin binding. This interaction means that BorO samples a conformation with this site open less frequently. This in turn will mean that borrelidin is less likely to bind to BorO. In addition, this interaction appears to "sequester" the aspartic acid residue (D486), which is an important borrelidin binding residue in sensitive enzymes.

It is striking that in the borrelidin-bound structure of BorO, the borrelidin molecule is bound in a crooked conformation as compared to with all of the EcThrRS mutant crystal structures solved in this study. This may suggest that BorO is not solely relying on the T516 residue, which is found at the bottom of the borrelidin binding site, to confer resistance and that other self-resistance mechanisms are in place. However, T516 is sufficient to confer self-resistance to borrelidin, when this mutation is introduced into EcThrRS (L489T). Next steps to further elucidate this possibility would be to check the resistance profile of BorO T516L, as well as to generate further mutations in BorO if T516L retains resistance.

3.3.2 The self-resistance mechanism of SpThrRS

The housekeeping SpThrRS protein for *S. parvulus* is also resistant to borrelidin. It was demonstrated in this study that while it does not confer borrelidin resistance alone, the Q510 residue (L489 in EcThrRS) is involved in the prevention of borrelidin binding to SpThrRS. This would suggest that it is partially responsible for the observed borrelidin resistance phenotype; however, there may be another factor involved. Examination of the sequences and structures of other ThrRSs identified Y482 (F461 in EcThrRS) as a potential second residue involved in borrelidin resistance.

A next step in unpicking the mechanism of resistance of SpThrRS would be to introduce a F461Y mutation into EcThrRS and assay if this mutation alone can confer borrelidin resistance. This would be interesting because, if so, it would suggest that SpThrRS has evolved an independent mechanism to prevent borrelidin binding. If this single point mutation made in EcThrRS is not sufficient to confer borrelidin resistance, a double mutant of Y461:Q489 may be sufficient to confer borrelidin resistance. This would tell us that both of these residues are resistance determinants in SpThrRS. Regardless, it is interesting that in both SpThrRS and BorO, a distinct mutation when compared to sensitive ThrRSs, has developed specifically in the equivalent to position 489 in EcThrRS.

3.3.3 Future work

It was intriguing that BorO does not function in *E. coli* but does in *Streptomyces*. This indicates that there may be some specificity in the $tRNA^{Thr}$ binding properties of BorO, as compared to the other ThrRSs which have been studied. From previous work it is known that ObaO from *P. fluorescens* can confer resistance to borrelidin and obafluorin when expressed heterologously in *E. coli* (see Chapter 4), and also that SpThrRS can aminoacylate the *E. coli* $tRNA^{Thr}$, inferred from its ability to confer resistance to borrelidin in *E. coli* as demonstrated in section 3.2.3.

Building on this it would also be interesting to crystallise BorO with one of the *S. parvulus* $tRNA^{Thr}$ molecules. This $tRNA^{Thr}$ could be produced in *S. parvulus* to ensure that the native post-transcriptional modifications have been made, and then crystallisation with BorO could give insight as to the structural determinants mediating BorO's possible $tRNA^{Thr}$ specificity.

A possible regulatory role for BorO could be explored. As already mentioned, EcThrRS and HsThrRS have both been demonstrated to have a role in post-transcriptional regulation of genes. If a similar phenomenon is occurring with BorO, it would explain why it is maintained by *S. parvulus* and other strains which contain BorO homologs. The mRNA being bound to BorO could be explored by use of a crosslinking immunoprecipitation sequencing (CLIP-seq) experiment. In a CLIP-seq, covalent protein-RNA crosslinks are formed using UV radiation. Cells are then lysed and RNA digested by a mild RNAse- this leaves only those RNA molecules bound to proteins whole. The protein of choice is then either affinity purified or immunoprecipitated by use of a specific antibody. The protein is the digested by proteinase K, leaving just RNA molecules which have been bound to the protein of choice. From this RNA, a cDNA library can be generated by a reverse transcriptase and the cDNA library can then be sequenced, identifying RNA sequences which bind to the protein of interest.

Chapter 4: The obafluorin resistance protein, ObaO: resistance mechanisms and obafluorin mode of action

4.1 Introduction

Aminoacyl tRNA synthetases (aaRSs) represent a valuable target in bacterial natural product warfare; all cellular organisms rely on them for effective translation, and they are highly conserved in many parts of the tree of life. In addition to borrelidin already discussed, obafluorin represents a second threonyl tRNA synthetase (ThrRS)-targeting natural product. Obafluorin is a β-lactone antibiotic produced by *Pseudomonas fluorescens* ATCC 39502.

4.1.1 Beta-lactone natural products

Several bacterial β-lactone natural products have previously been described, including ebelactone, lipstatin and salinosporamide A which are generated from PKS, fatty acid and hybrid PKS/NRPS derived biosynthetic pathways respectively (see Figure 4.1A). These compounds all exhibit potent biological activity, and β-lactone NPs and molecules generally, act by covalent attachment to nucleophilic residues in the active sites of their protein targets- see Figure 4.1B.

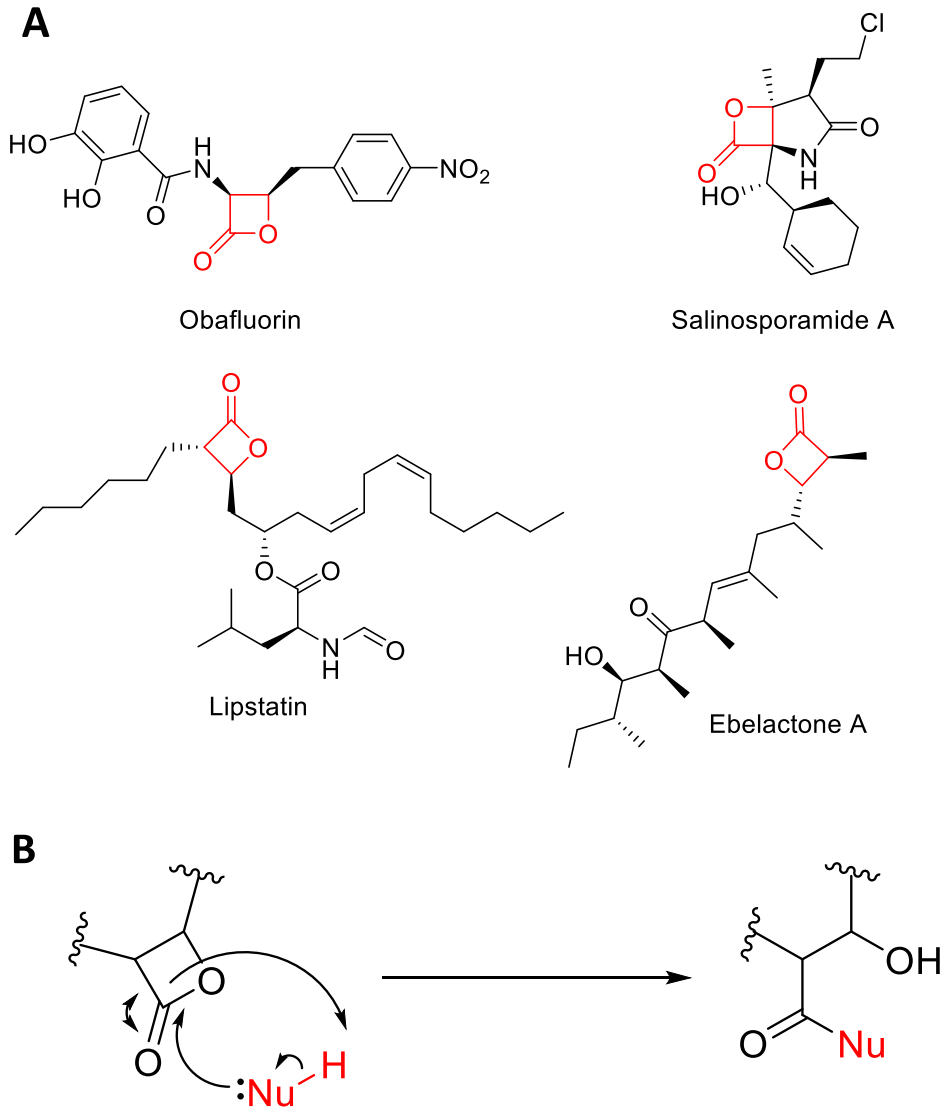


Figure 4.1. A) Examples of β-lactone containing natural products. Beta-lactone moiety shown in red. **B) Scheme of *β*-lactone ring opening**. Nucleophile generically annotated as Nu, in red.

4.1.2 Identification of ThrRS as the obafluorin target

The target of obafluorin was found to be ThrRS due to the presence of *obaO*, encoding a second ThrRS homologue that is located in the obafluorin BGC. Attempts to knockout *obaO* on its own were unsuccessful, until a double knockout with *obaL* was made; ObaL catalyses the final step in 2,3-DHBA biosynthesis. 2,3-DHBA is one of the two precursors used by the NRPS ObaI to assemble obafluorin, as discussed in section 1.3.3.1 and illustrated in Figure 1.20. When the Δ*obaL* strain is fed with exogenous 2,3-DHBA, the strain can grow and turn the growth medium purple- indicative of production of obafluorin. When the double knockout strain was fed with exogenous 2,3-DHBA, thereby initiating obafluorin biosynthesis, the strain did not grow. When complemented with *obaO*, the double knockout strain could grow and the growth medium turned purple, indicating the production of obafluorin (see Figure 4.2). The sensitive target of obafluorin, EcThrRS was unable to confer resistance when expressed in *P. fluorescens* Δ*obaO*Δ*obaL*.

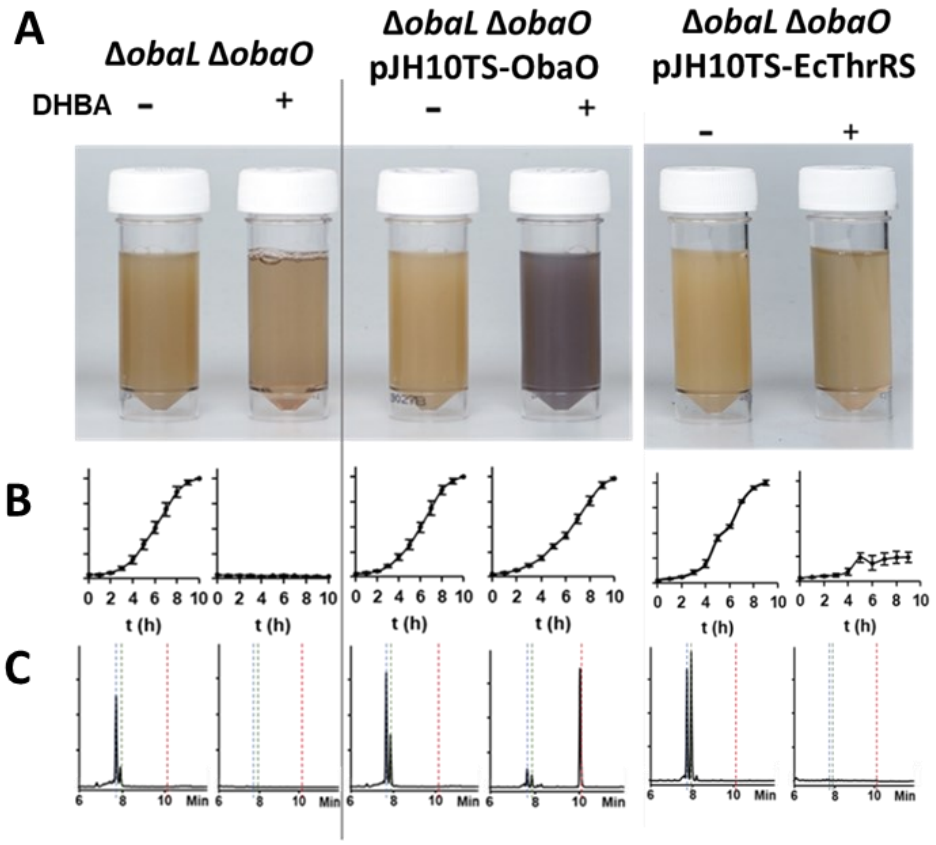


Figure 4.2. ObaO is the self-resistance determinant for obafluorin in *P. fluorescens* 39502. *P. fluorescens* Δ*obaL*Δ*obaO* WT or expressing EcThrRS or ObaO were grown ±2,3-dihydroxybenzoic acid (2,3-DHBA, 0.2 mM). The addition of 2,3-DHBA to growing WT and EcThrRS producing strains abolishes growth, while ObaO expressing strain maintains growth and obafluorin production. **A**, Cultures of each strain after 14 h growth; the purple coloration is indicative of obafluorin production. **B**, Log phase growth curves, showing complete absence of growth for Δ*obaL*Δ*obaO* when 2,3-DHBA is added. Each data point is the average of three biological repeats, and bars show the standard error. **C**, Representative HPLC chromatograms at 270 nm for each condition at 14 h; obafluorin elutes at 10.1 min (red dashed line). Adapted from Scott et al. 2019[174].

4.1.3 Unpicking obafluorin self-resistance/mode of action

Knowing that β-lactones usually acylate their biological targets, forming covalently attached intermediates, mass spectrometry (MS) experiments have been performed by other groups members (Dr Tom Scott, Dr Sibyl Batey and Melissa Davie). Purified EcThrRS was incubated with obafluorin and subsequent MS analysis indicated a major species for one acylation event with obafluorin, but also that additional acylation's could occur. Subsequent tryptic digest of these proteins samples followed by tandem MS analysis (carried out by the JIC Proteomics platform) indicated these that acylation could occur at many residues, but no major site of acylation could be identified. Thus, under the conditions tested, it appeared that obafluorin acts as a general acylating agent. However the existence of a BGC-situated self-resistance gene, which is a copy of the target in sensitive bacteria, would suggest that *in vivo* the interaction with obafluorin is likely to be more specific, or rather that if acylation is general, that only a specific one of these identified acylation events leads to inhibition[174]. Moreover, general acylation *in vivo* would suggest that obafluorin should be toxic, but the published data from an *in vivo* mouse infection model did not suggest this[175].

In the studies described above, a large excess of obafluorin was added to a highly concentrated sample of EcThrRS to prepare these samples, and in neutral-basic conditions. In these experiments, the positions which were acylated were highly variable between samples, with very little consensus. Therefore, this general acylation is likely to be an artifact of the experiment. General acylation also hindered co-crystallisation of obafluorin with the target EcThrRS and with the self-resistance protein PfObaO, and soaking already grown crystals with obafluorin also led to no good diffraction data.

Various point mutations in EcThrRS have already been tested by Dr Sibyl Batey and Melissa Davie, based on conserved polymorphisms and mass spectrometry data. These sets were A356G, G463S and L489M which are all strictly conserved in sensitive/resistant enzymes and K200A, K227A, K314A and K419A which were designed based on the results of the mass spectrometry data. When expressed in *P. fluorescens* Δ*obaO*Δ*obaL* fed with 2,3-DHBA, none of these single point mutations in EcThrRS conferred any resistance alone (unpublished).

The aim of this study was to try and elucidate the mechanism of action of obafluorin and unpick the specific obafluorin self-resistance determinants in ObaO, by comparing it to the ThrRS from *E. coli*, a known target of obafluorin.

In this study, I used a combination of structural, protein mutagenesis and biological activity assays approaches because from the analysis of sequence data alone it is not obvious where obafluorin binds or what the resistance determinants are. I confirmed that homologues of ObaO from other strains are resistant to obafluorin, and that these homologues require their editing function to be

intact to confer resistance. A structural study revealed the point of acylation by obafluorin in the target EcThrRS and structural work on ObaO combined with the generation of spontaneous resistant mutants against obafluorin allowed us to develop a model for the obafluorin resistance of ObaO.

4.2 Results and Discussion

4.2.1 ObaO Homologues confer resistance to obafluorin in P. fluorescens and E. coli

Given that efforts to structurally characterise *P. fluorescens* ATCC 35902 ObaO (PfObaO) had been unsuccessful, the first approach was to select ObaO homologues from other obafluorin BGCs that might be more amenable to crystallisation. The homologues chosen meant that one homologue from each obafluorin BGC containing genus (*Pseudomonas, Burkholderia and Chitiniphilus*) was represented. These were from the following strains: *Burkholderia multivorans* LMG 29306 (BmObaO), *Chitiniphilus shinanonensis* DSM 23277 (CsObaO) and *Pseudomonas sp.* Irchel s3a18 (PIObaO). To verify that these ObaO homologues confer obafluorin resistance, they were cloned into pJH10TS and transformed into both *E. coli* NR698 and *P. fluorescens ΔobaOΔobaL. E. coli* NR698 expressing these three genes were challenged with obafluorin in spot-on lawn bioassays and showed complete resistance to obafluorin (Figure 4.3). Similarly, when the double knockout *P. fluorescens ΔobaOΔobaL* strains expressing these genes were grown in obafluorin production media (OPM) containing 2,3-DHBA they were able to grow and produce obafluorin (see Figure 4.4). It should be noted, however that the BmObaO showed a very slight delay of growth in the feeding experiment. On the basis of these results, it can be concluded that BmObaO, CsObaO and PIObaO are self-resistance determinants and function in the same manner as PfObaO.

4.2.2 N-terminally truncated ObaO homologues do not confer obafluorin resistance

In order to also verify the importance of the N-terminal editing domain of ThrRS homologues *in vivo*, PfObaO, BmObaO, CsObaO, and PIObaO were cloned into pJH10TS with their N-terminal editing domain truncated in the same position that they would be for the crystallography experiments described earlier with EcThrRS. Although some ThrRSs which have been shown to function without their editing domains (those found in archaea)[236,237], none of the ΔN ObaO homologues could confer full resistance to obafluorin, likely due to misaminoacylation of $tRNA^{Thr}$ by serine (Figure 4.3 and Figure 4.4). For the spot-on lawn bioassays using *E. coli* NR698 expressing truncated ObaO homologues (Figure 4.3), it should be noted that the MIC for CsObaO was higher than for the other homologues. In archaea, which have no editing domain on their ThrRS, a Tyr-tRNA deacylase has been recruited for editing of misacylated tRNA[236].

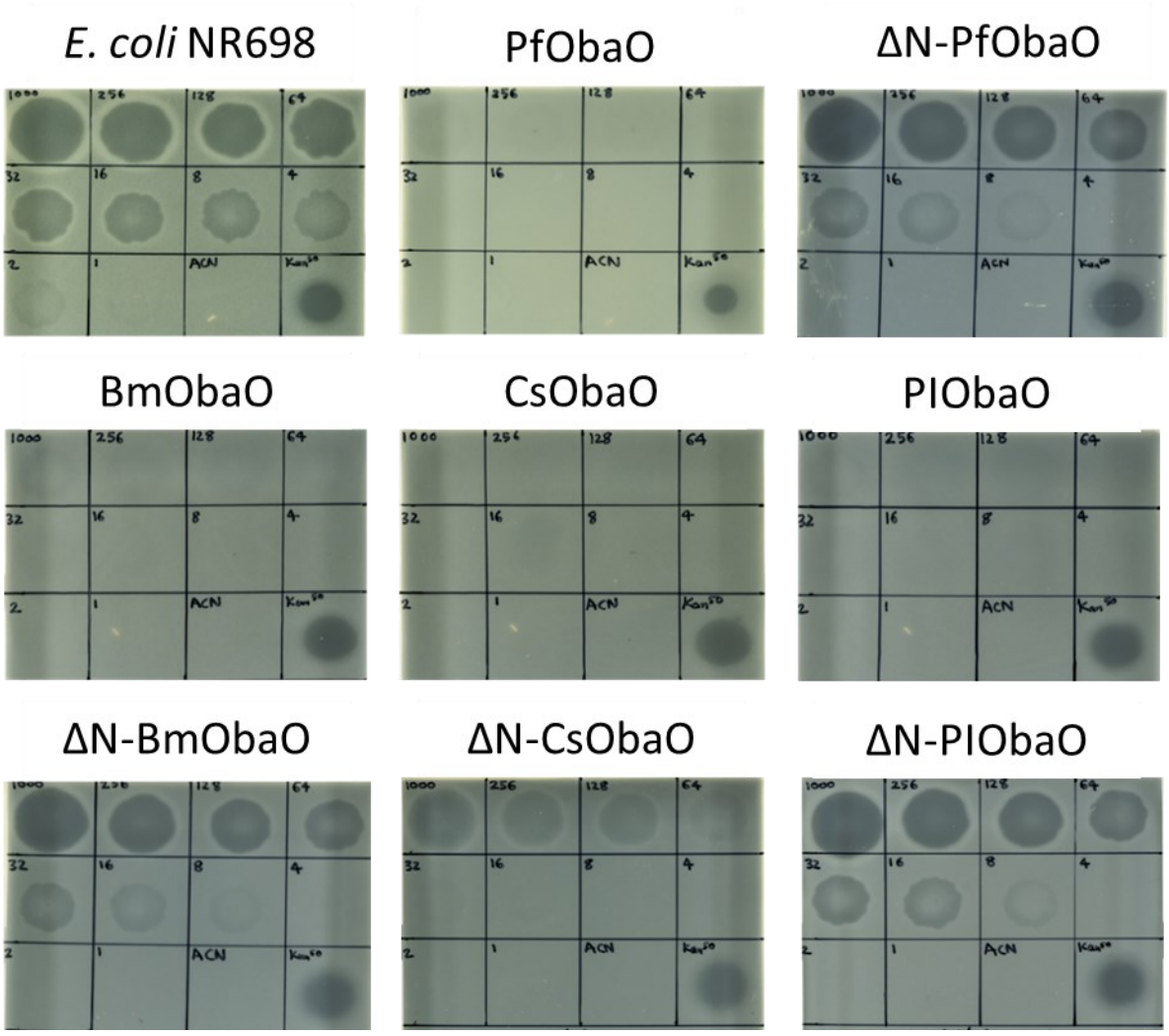


Figure 4.3. Full length ObaO homologues show full resistance, while ΔN ObaO homologues do not maintain obafluorin resistance. *E. coli* NR698 expressing ObaO homologs were inoculated into soft nutrient agar to pour the plates before spotting on of 4 µL obafluorin (dissolved in MecN) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of MeCN only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates incubated at room temperatrue for 16 hours. ΔN denotes the truncation of the N-terminal editing domain.

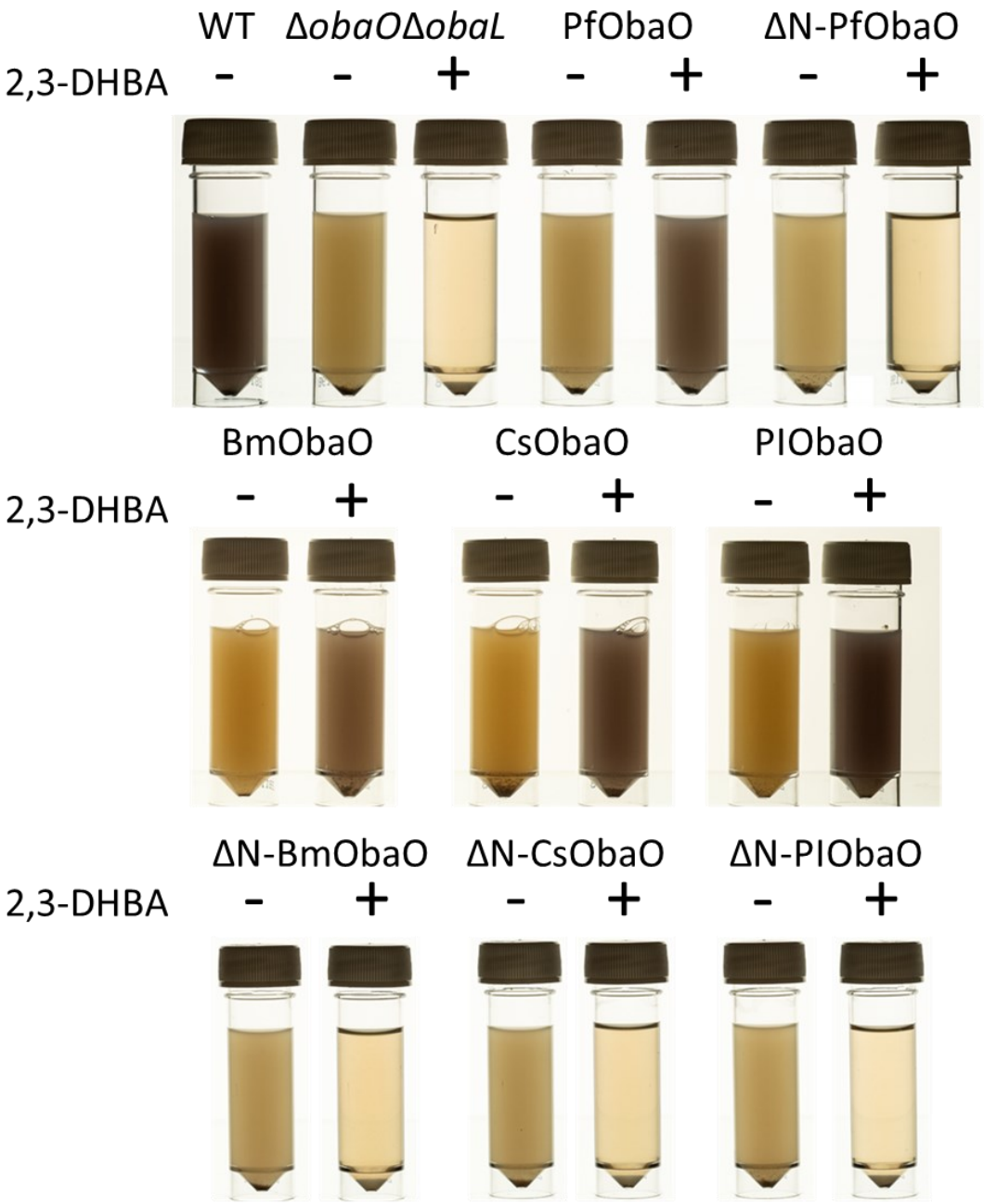


Figure 4.4. ObaO homologues could confer obafluorin resistance, but only with the editing domain intact, when expressed in *P. fluorescens* ΔobaOΔobaL. *P. fluorescens* Δ*obaL*Δ*obaO* WT or expressing ObaO homologues and WT *P. fluorescens* control were grown ±2,3-dihydroxybenzoic acid (2,3-DHBA, 0.2 mM). The addition of 2,3-DHBA to growing WT and ΔN-ObaO homologue producing strains abolishes growth, while ObaO expressing strains maintains growth and obafluorin production. Cultures of each strain photographed after 14 h growth; the purple coloration is indicative of obafluorin production. Unfed cultures are denoted by a minus sign, fed cultures are denoted by a plus sign. I created the plasmids, Dr Sibyl Batey performed the bioassays.

4.2.3 Crystallisation of BmObaO

All three ObaO homologues were truncated to remove the editing domain and cloned into pET29a. The resulting C-terminally hexa-histidine tagged proteins were purified using immobilised metal

affinity chromatography (IMAC) and gel filtration and dispensed into crystallisation trials with the PEGs screen (Qiagen), which has previously been successful for the other ThrRSs. The first crystallisation screens were set up with protein at 10 mg/mL and with or without 2 mM borrelidin (which was helpful for obtaining crystals of BorO, as discussed in Chapter 3). Obafluorin co-crystallisations had not previously been successful even for the readily crystallising EcThrRS, for previous lab members and so was not used for these crystallisation trials. As for the crystallisation of EcThrRS and BorO, the protein was in made up in Lysis buffer containing 10 mM threonine. Initial screening showed that, in most conditions, the protein had precipitated, indicating that a lower protein concentration should be used. Protein concentration for all three proteins was thus reduced to 4 mg/mL and the screen was repeated; under these conditions, crystals could be grown for BmObaO only (and only without borrelidin). These crystals were harvested, shipped to Diamond Light Source (DLS) and some X-ray diffraction could be observed, although it was of insufficient quality for further processing. An X-Ray fluorescence scan confirmed the presence of zinc in the crystal, which we know is found in the active site of the protein, suggesting these are authentic crystals of BmObaO (see Figure 4.5).

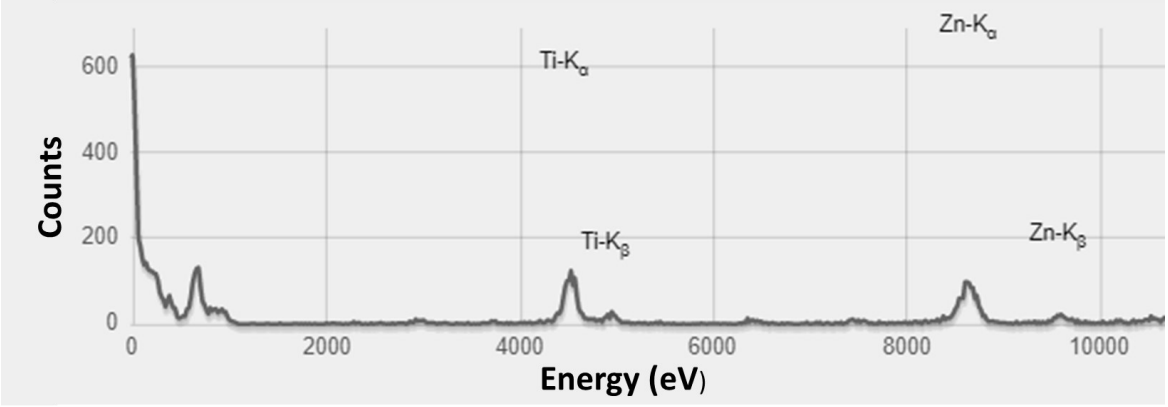


Figure 4.5. X-Ray Fluorescence graph from Diamond Light Source. Annotated are the elements corresponding to the peaks. Titanium is likely an artifact from the beamline hardware. An X-Ray Fluorescence graph for BorO, which did diffract can be seen in Figure 3.14.

Optimisations of these conditions were designed as outlined in section 2.12.1. Using optimised crystallisation conditions, set up with either the non-hydrolysable ThrAMP analogue, ThrSAA (discussed in section 3.2.7) or threonine at 2 mM, protein at 4 mg/mL, with and without seeding with crystals from the first crystallisation trial to provide nucleation sites. Crystals could reliably be grown in these conditions (example seen in Figure 4.6).

From these trials, 49 crystals were obtained and sent for diffraction but only a single crystal yielded data which was of sufficient quality for further processing. However, this showed strong anisotropy and PHASER was unable to phase correctly. This single diffracting crystal came from conditions with 0.1 M Tris pH 7.5, 5% PEG 2000 MME and 5% PEG 3350, with threonine added. While at the

Diamond CCP4 Data Collection and Structure Solution Workshop 2021, progress was made to solve the structure. PHASER had been unsuccessful because the data had pseudosymmetry and severe anisotropy, both of which could be mitigated somewhat by manually changing the symmetry of the space group, and using autoPROC for data processing and BUSTER for refinement. Overall, there were problems with the quality of the electron density, meaning that while an initial model could be built, the refinement statistics were never sufficient to consider this a real solution due to poor electron density in the anticodon binding domain due to the anisotropy. At this point, it became clear that it may be difficult to solve the structure of the enzyme bound to threonine or ThrSAA because of the poor quality of diffraction of crystals from this protein.

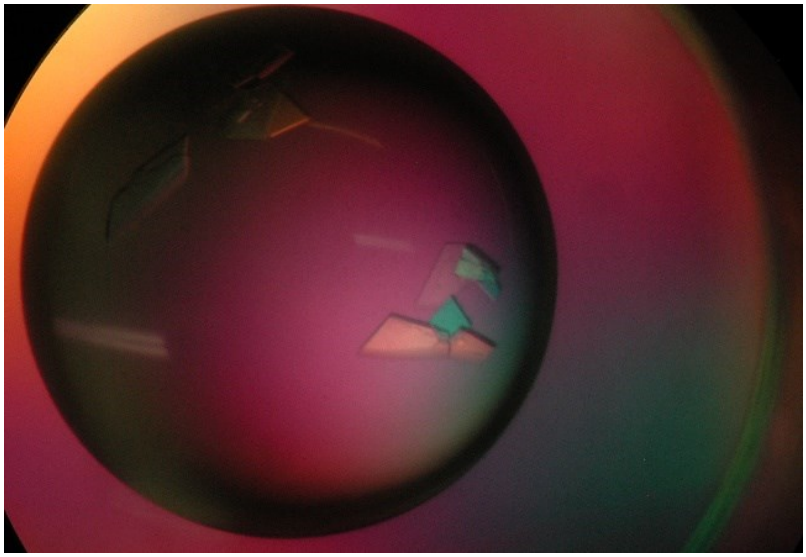


Figure 4.6. Example of a crystallisation drop containing crystals of BmObaO, Visualised at 10x magnification with a polarising light filter.

4.2.4 ObaO confers borrelidin resistance

To test whether ObaO could confer resistance to the non-cognate ThrRS inhibitor borrelidin, *E. coli* NR698 expressing PfObaO was assayed for resistance to borrelidin as shown in Figure 4.7. This showed that PfObaO confers resistance to borrelidin up to 1 mg/mL in spot-on lawn bioassays. Given this surprising observation we hoped that an understanding of how PfObaO is resistant to borrelidin could give insight into the mode of action of obafluorin. The same resistance could be conferred when each of BmObaO, CsObaO and PIObaO were expressed in *E. coli* NR698, and ΔN constructs of all these proteins failed to confer resistance from the results described in 4.2.3. The borrelidin resistance determinant is therefore conserved in all of the tested ObaO homologues.

Using ITC analysis as described in the previous chapter for BorO, we were able to demonstrate that borrelidin does not bind to PfObaO, and that in the accompanying control experiment the intermediate analogue ThrSAA displayed a K_d of 2.19 ± 0.14 nM (n=3) (indicating that the protein was functional and correctly folded). Binding curves can be seen in Figure 4.8A & E. Taken together

with the observation that no crystals could be grown for PfObaO in the presence of borrelidin, it appears that the resistance to borrelidin is due to an inability of borrelidin bind to bind the enzyme. This result mirrors that reported for BorO in the previous chapter.

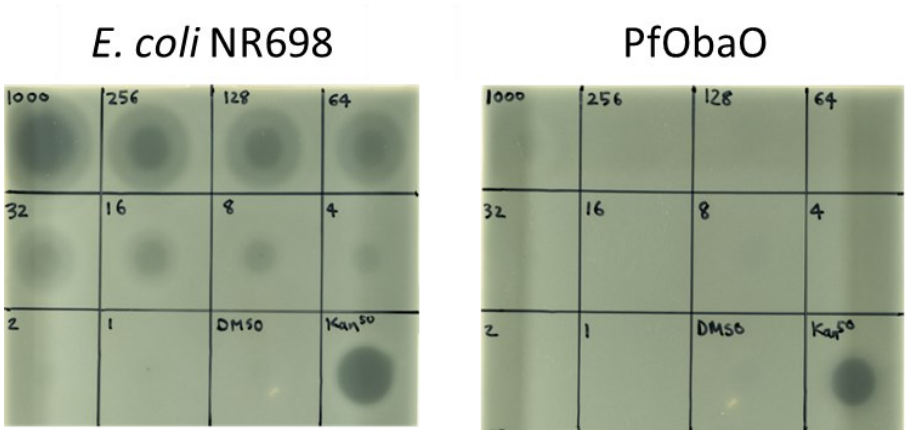


Figure 4.7, ObaO confers borrelidin resistance when expressed in *E. coli* NR698. *E. coli* NR698 expressing ObaO or WT *E. coli* NR698 was inoculated into soft nutrient agar to pour the plates before spotting on of 4µL borrelidin (dissolved in DMSO) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of DMSO only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates incubated at room temperatrue for 16 hours.

4.2.5 EcThrRS L489M is resistant to borrelidin but not obafluorin

While studying the borrelidin resistance of BorO and SpThrRS, detailed in Chapter 3, T516 in BorO was identified as being a key residue for borrelidin resistance. At this position in the ObaO homologues is a completely conserved methionine residue. This L489M mutation has previously been described[161] and was capable of conferring resistance to borrelidin. The authors suggested that this was due to increased steric bulk in the fourth, hydrophobic binding site which will prevent borrelidin binding. The L489M mutation had already been introduced into EcThrRS in pJH10TS by Dr Sibyl Batey as a possible obafluorin resistance residue due to its full conservation in ObaO analogues, however was unable to confer resistance to obafluorin when expressed in *E. coli* NR698. To investigate whether EcThrRS carrying the L489M mutation could confer resistance to borrelidin, the same strain was then assayed against borrelidin in a spot-on lawn bioassay alongside the control strain expressing WT EcThrRS; as shown in Figure 4.9) the L489M substitution clearly confers resistance to borrelidin.

To explore this further full length *EcThrRS*(L489M) was cloned into pET28a and N-terminally truncated *EcThrRS*(L489M) was cloned into pET29a for protein production. These proteins were purified and the full-length protein was subjected to ITC assays against borrelidin and the intermediate analogue ThrSAA, while the truncated protein was crystallised in the presence of borrelidin and the structure solved with borrelidin bound.

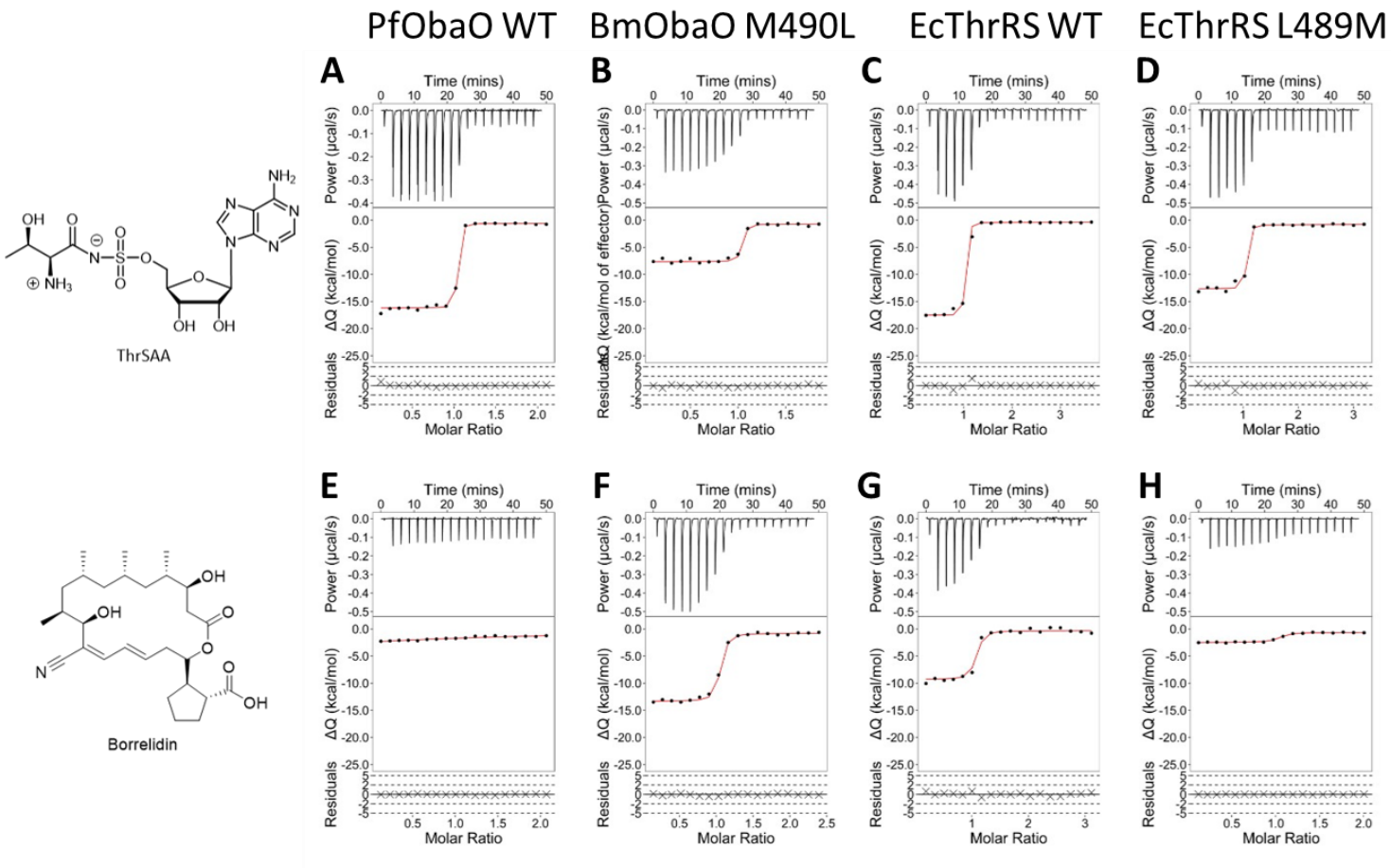


Figure 4.8. WT PfObaO and EcThrRS L489M cannot bind borrelidin, but BmObaO M490L and WT EcThrRS. Each lettered dataset shows the isotherms of the reaction in the top panel, the resulting binding curve in the middle panel, and the residuals (which measures the goodness of fit of the curve to the data) in the bottom panel. **A)** WT PfObaO with ThrSAA, **B)** BmObaO M490L with ThrSAA, **C)** WT EcThrRS with ThrSAA, **D)** EcThrRS L489M with ThrSAA, **E)** WT PfObaO with Borrelidin, **F)** BmObaO M490L with borrelidin, **G**) WT EcThrRS with borrelidin, **H)** EcThrRS L489M with borreldiin. All four proteins show binding to the intermediate analog, ThrSAA, suggesting that they are folded correctly. Figure generated in RStudio.

The ITC analysis revealed that EcThrRS(L489M) binds the intermediate analogue with a K_d of 4.12 ± 0.45 nM (n=3), and there significantly weaker binding to borrelidin with a K_d of 0.91 ± 0.29 µM (n=3), but the heat change of the binding for borrelidin is significantly smaller than expected (see Figure 4.8D and H). This significantly weaker binding would suggest that borrelidin is unable to outcompete the natural substrates of the enzyme when a methionine is in the borrelidin resistance position.

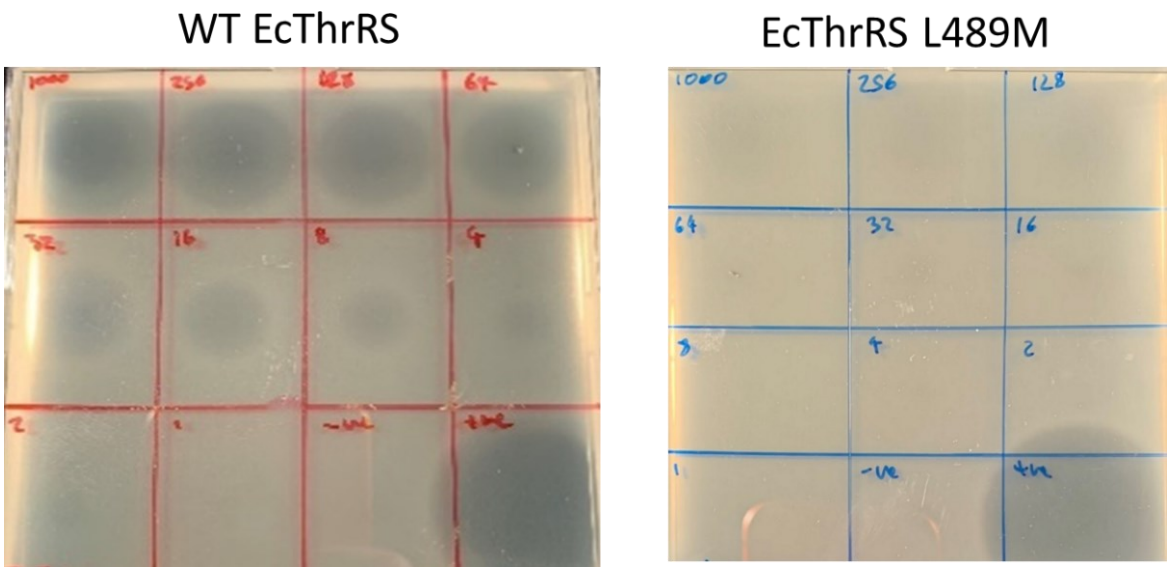


Figure 4.9. EcThrRS L489M confers resistance to borrelidin when expressed in *E. coli* NR698. *E. coli* NR698 expressing WT EcThrRS or EcThrRS L489M were inoculated into soft nutrient agar to pour the plates before spotting on of 4 µL borrelidin (dissolved in DMSO) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of DMSO only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates incubated at room temperatrue for 16 hours.

To confirm that EcThrRS L489M binding to borrelidin does not prevent binding of ThrSAA, ITC competition assays with EcThrRS L489M were performed (Figure 4.10). A protein-borrelidin complex was made by pre-incubating the protein with borrelidin at a 1:1 molar ratio, at a concentration of 20 µM. An ITC run was then performed with the protein-borrelidin complex in the cell, and ThrSAA injected in as a proxy for the substrates of the reaction was then performed. If ThrSAA was unable to displace borrelidin as the more competitive EcThrRS-L489M ligand, we would expect to see an ITC curve that reflects no binding. However, we were able to see an ITC curve that reflects the displacement of borrelidin and binding of ThrSAA. This suggests that in this protein, ThrSAA can outcompete borrelidin, and therefore for EcThrRS L489M, borrelidin is a less favourable ligand than ThrSAA. As an additional control it would be useful in future to repeat this competition assay with WT EcThrRS. An example, with EcThrRS L489Q, which cannot confer borrelidin resistance in *E. coli* NR698 can be seen in Figure 3.28 in section 3.2.12, in which borrelidin is a more favourable

ligand than borrelidin. It is interesting that this residue in the same position previously shown to be essential for borrelidin resistance is the position that confers borrelidin resistance in ObaO.

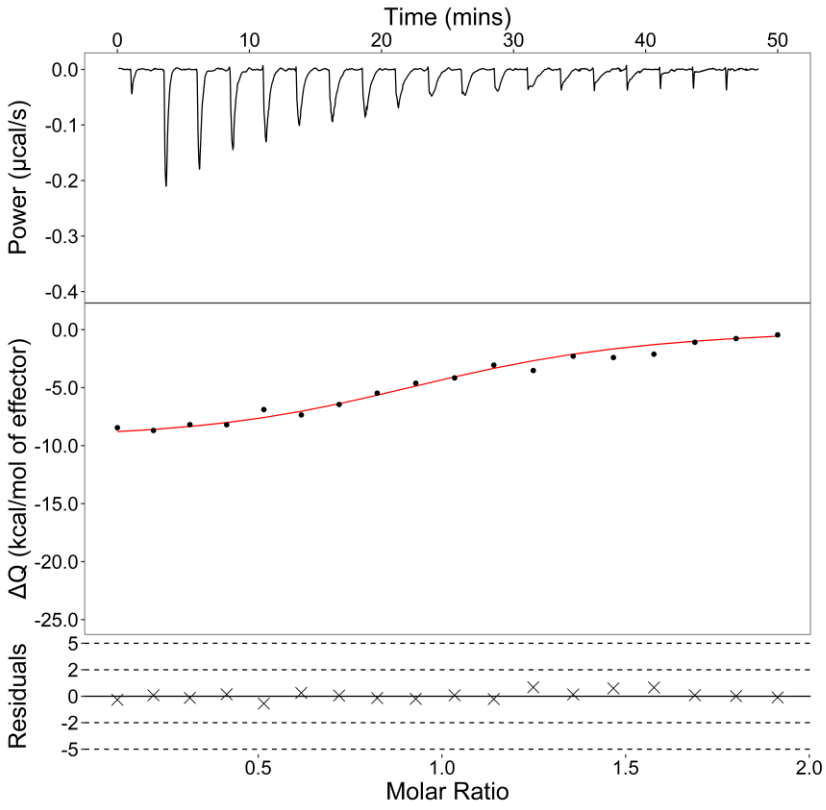


Figure 4.10. ITC plots for competition binding experiment for EcThrRS L489M pre incubated with borrelidin against ThrSAA. The isotherm of the reaction is in the top panel, the resulting binding curve in the middle panel, and the residuals which measures the goodness of fit of the curve to the data in the bottom panel. EcThrRS L489M was preincubated with borrelidin as if an ITC experiment had been undertaken. An ITC run was then performed by addition of ThrSAA using the syringe as normal. The resulting curve shows that ThrSAA is able to bind to EcThrRS L489Q following incubation with borrelidin. Figure generated in RStudio.

As well as using the positive binding control of ThrSAA in ITC experiments, the quality of the protein sample was checked by mass photometry to ensure that dimeric protein was present, and that of the formation of higher order aggregates was not an issue. The mass photometry histograms for EcThrRS L489M and BmObaO M490L can be seen in Figure 4.11. For both proteins, both dimer and monomer can be observed, but no higher order multimers are present. It was previously observed that with increased dilution, there is an increased proportion of monomer observed for these proteins (see Supplemental Figure 23).

A crystal structure of EcThrRS L489M in the presence of borrelidin was solved to 1.6 Å resolution. As was observed for BorO and its single point mutants discussed in Chapter 3, crystallisation with

borrelidin bound was likely to be possible due to the excess of borrelidin in the crystallisation conditions, but is also unsurprising, as the ITC data for this mutant indicated some ability to bind borrelidin, albeit significantly weaker than for the parent enzyme EcThrRS. This suggests that while the methionine in the hydrophobic binding pocket is able to interact with Y462 to keep the pocket closed, the protein can also sample additional conformations which break this interaction, allowing borrelidin to bind. The presence of large, hydrophobic residues (including methionine) have previously been shown to confer resistance to borrelidin and it was suggested that this was by physically blocking binding into the pocket[161]. Borrelidin binding in this protein can be seen in Figure 4.12.

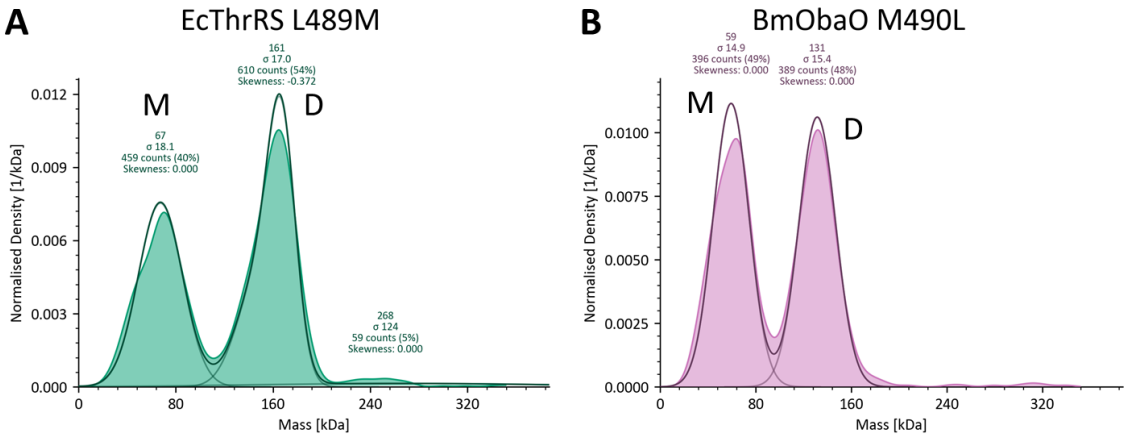


Figure 4.11. Mass photometry analysis of isolated proteins. Histograms of A) EcThrRS L489M and B) BmObaO M490L. The average predicted mass in kDa is shown above each peak; with a protein concentration of 100nM for both. See Figure 3.12A in Chapter 3 for the histogram for wild-type EcThrRS. Proteins were measured in PBS. Both proteins show peaks which align to the sizes of the monomer and the dimer. Presence of the dimer indicates the presence of functional protein.

While not strictly solved due to issues with anisotropy, the model built for apo BmObaO could provide us with some information; in the "borrelidin resistance" position, in the place of the leucine (L489) found in EcThrRS, there is a methionine in ObaO, which is conserved in all ObaO homologues. This methionine appears to possibly be interacting with Y463 via a methionine-aromatic bond. Y463 (Y462 in EcThrRS) is vital for threonine binding, tRNA binding and aminoacylation[115]. Additionally, this interaction may hold the "borrelidin binding" pocket closed, with the β carbons of the methionine and tyrosine residues being at approximately 5 Å in the ObaO model and the β carbons of the leucine and tyrosine being held at a 7 Å distance in the EcThrRS:borrelidin structure and roughly 4.5 Å in the EcThrRS:ThrSAA structure. The optimal methionine-aromatic interaction distance is 5 Å.[238] so in the model, the bond distances would be somewhat long. The model is also poor, and so these distances cannot necessarily be relied upon (see Figure 4.13). The observed lack

of binding could alternatively simply be due to steric hindrance of borrelidin binding in this site, as previously suggested in Ruan *et al.* 2005[161].

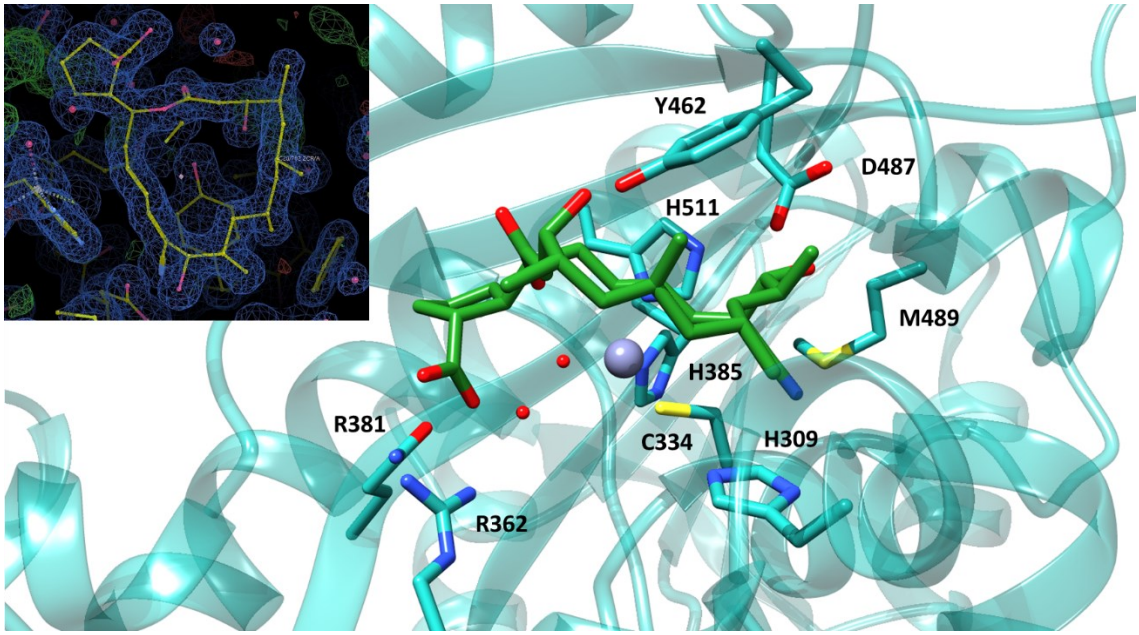


Figure 4.12. Crystal structure of borreldin bound to EcThrRS L489M. Inset, top left: electron density map, of borrelidin in the active site of EcThrRS L489M contoured to 1 rmsd. Main: borrelidin in the active site of EcThrRS L489M;residues making contact with borrelidin are labelled and shown in stick representation. Borrelidin shown in stick representation in green, protein in cyan in cartoon representation. Zinc shown as a grey sphere. Figure generated in Chimera.

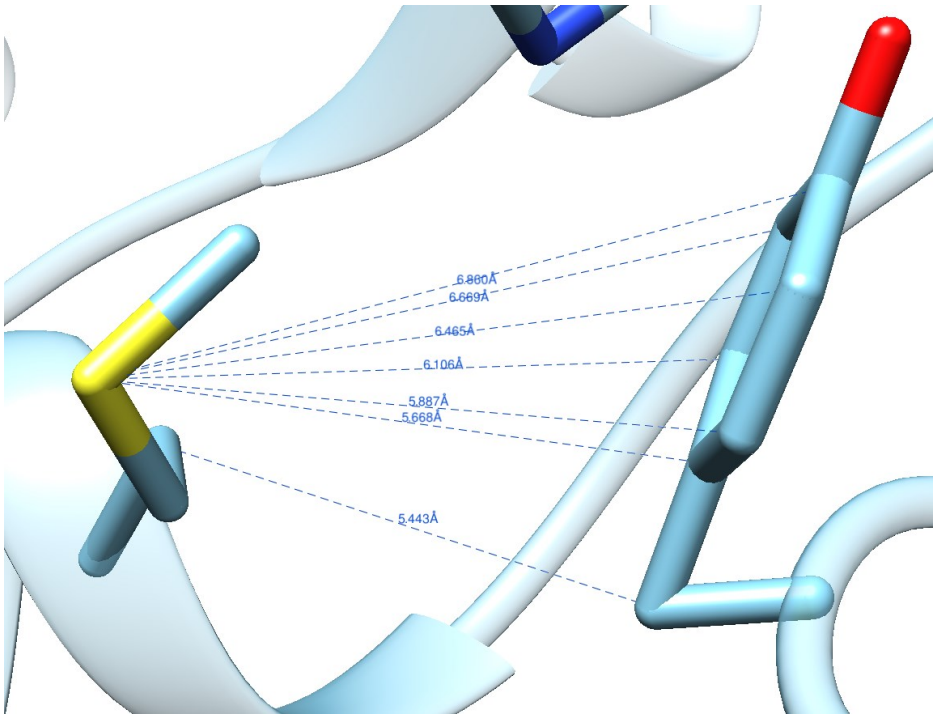


Figure 4.13. There is possibly a methionine-aromatic interaction between M490 and Y463 in a model of BmObaO. Distances shown in blue text, marked by dotted lines. Figure generated in Chimera.

4.2.6 BmObaO M490L can bind to borrelidin

To confirm that the BmObaO M490 residue (L489 in EcThrRS) is essential for borrelidin resistance , the M490L mutation was introduced into the pET28a construct containing *BmObaO* and the protein was purified. ITC experiments showed that BmObaO M490L binds borrelidin with a K_d of 2.38 ± 0.19 nM (n=2) and to ThrSAA with K_d of 4.17 ± 0.83 nM (n = 3) (Figure 4.8B and F); in contrast to the WT enzyme can only bind to ThrSAA (Figure 4.8A and 9E). This confirmed that M490, which is conserved in ObaO homologues, is the residue responsible for ObaO:borrelidin cross resistance. To further solidify this result, the gene encoding BmObaO M490L was cloned into pJH10TS and transformed into *E. coli* NR698. This strain was challenged with both borrelidin and obafluorin and was shown to have no resistance to either (Figure 4.14) in contrast to *E. coli* NR698 expressing ObaO which is resistant to both antibiotics. This could suggest that the protein is unfolded/non-functional in this strain but if the protein is properly folded, this suggests that M490 is important for both obafluorin and borrelidin resistance.

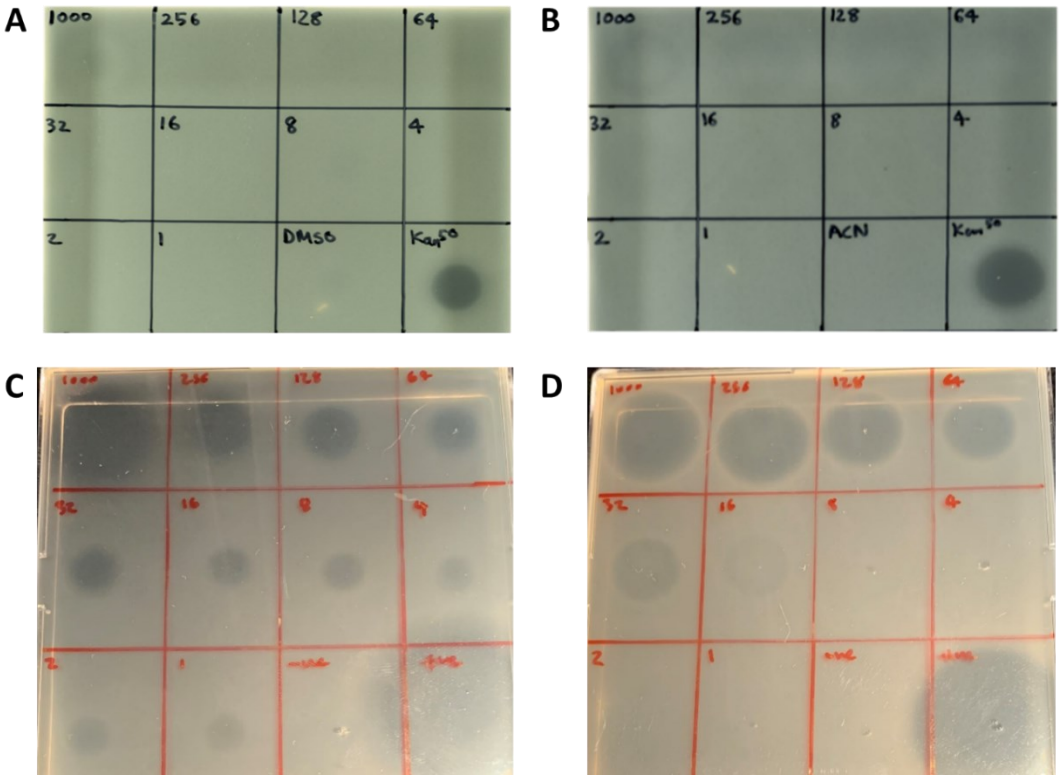


Figure 4.14. The BmObaO M490L mutation offers sensitity to both borreldin and obafluorin when expressed in *E. coli* NR698. A) *E. coli* NR698 expressing WT PfObaO challenged with borrelidin. B) *E. coli* NR698 expressing WT BmObaO challenged with obafluorin. C) *E. coli* NR698 expressing BmObaO M490L challenged with borrelidin. D) *E. coli* NR698 expressing BmObaO M490L challenged with obafluorin. Strain inoculated into soft nutrient agar to pour the plates before spotting on of 4 µL borrelidin or obafluorin (dissolved in DMSO or MeCN, respectively) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of DMSO or MeCN only bottom middle right, and positive control of kanamycin at 50 µg/mL bottom right. Plates incubated for 16 hours at room temperature.

4.2.7 A crystal structure of BmObaO M490L bound to borrelidin

Crystallisation of BmObaO M490L was at 4 mg/mL was attempted using the PEGs screen in the presence of 2 mM borrelidin. This mutant protein crystallised readily in the presence of borrelidin. This result supports the supposition that borrelidin binds to BmObaO M490L because the wild-type protein could never be crystallised in the presence of borrelidin. Crystals grown in 0.1 M HEPES pH 7.5, 15% PEG 20000 with 2 mM borrelidin (a different condition for crystallisation than was successful for the WT protein) successfully diffracted and the resulting data were used, along with the initial model from the previously incomplete BmObaO WT model for molecular replacement, in order to solve the structure of BmObaO M490L to 2.5 Å (Figure 4.15). This is the first structure of an obafluorin resistance (ObaO) protein to be solved and can be used to gain insight into the obafluorin resistance of these enzymes and, therefore, might give insights into the obafluorin mechanism of action. Borrelidin bound is in this structure in the same conformation as in EcThrRS and protein has adopted the same conformation observed by EcThrRS. The structure solution information for the crystal structures discussed in this chapter can be found in Supplemental Table 4.

With this structure in hand, some attempts at introducing rational mutations could be made. It was not immediately clear from the structure alone where obafluorin might bind, and the large conformational change caused by the presence of borrelidin meant that any guesses were not necessarily easily validated.

4.2.8 EcThrRS:ObaO Chimeras suggest a key subdomain interaction for obafluorin resistance

A series of EcThrRS:PfObaO chimeras were designed, assembled and assayed by Dr Sibyl Batey. These were produced in order to narrow down the residues/subdomains which are important for obafluorin resistance, and to provide any information possible regarding the mechanism of action. The first experiments found that only those chimeras which had the ObaO catalytic domain were able to confer resistance to obafluorin, when expressed in *E. coli* NR698 and challenged with obafluorin (Figure 4.16).

Knowing that the obafluorin resistance determinants reside in the catalytic domain, Dr Batey generated a series of subdomain chimeras in order to further isolate the location of these determinants The design is shown in Figure 4.17, with the catalytic domain split into a series of subdomains, C1-5. EcThrRS was used as the background, with the ObaO subdomains switching in. This resulted in the observation that some resistance could be seen with EcC2, but more at a higher level with EcC4. This would suggest that the major obafluorin self-resistance determinant(s) can be found in subdomain 4, while subdomain 2 carries a feature that partially confers resistance. The same experiments were repeated with an ObaO background, substituting in EcThrRS subdomains. Here resistance was only seen chimeras containing ObaC2 and ObaC4, the opposite of the expected

result (Figure 4.17). This surprising result could be due to the protein being unfolded for the other chimeras due to disrupted folding interactions. However this result does suggest that subdomain 4 is essential/critical, and that subdomain 2 is auxiliary, for conferring obafluorin resistance; ObC4 had a weaker resistance phenotype than ObC2.

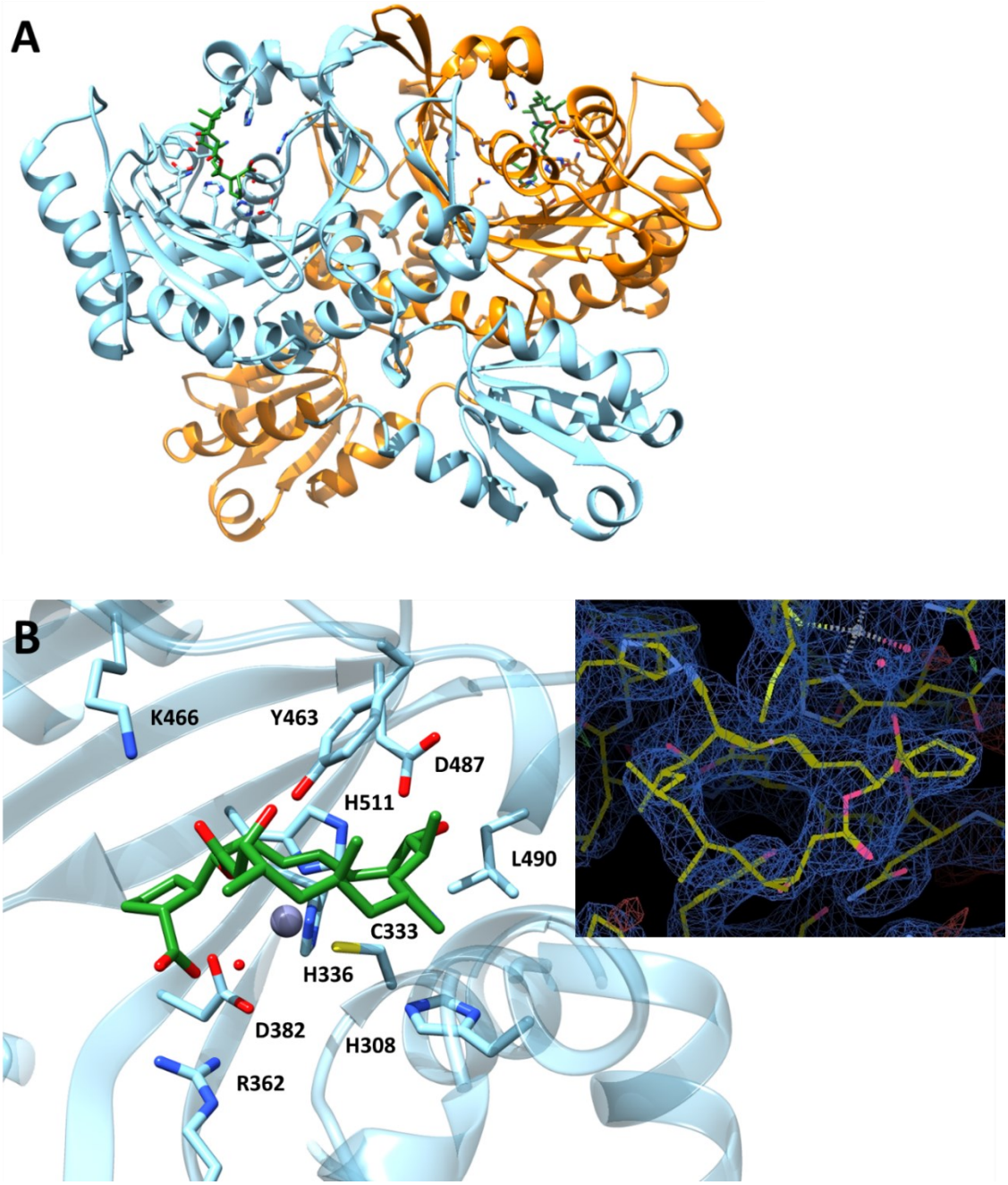


Figure 4.15. Crystal structure of BmObaO M490L. A) full protein showing dimer. One monomer in blue, the other in orange, borrelidin in green. **B)** inset, top right: electron density map, of borrelidin in the active site of EcThrRS L489M contoured to 1 rmsd. Main: borrelidin in the active site of EcThrRS L489M, with residues making contact with borrelidin labelled and shown in stick representation. Borrelidin shown in stick representation in green, protein in blue in cartoon preresntation. Zinc shown as a grey sphere. Figure generated in Chimera.

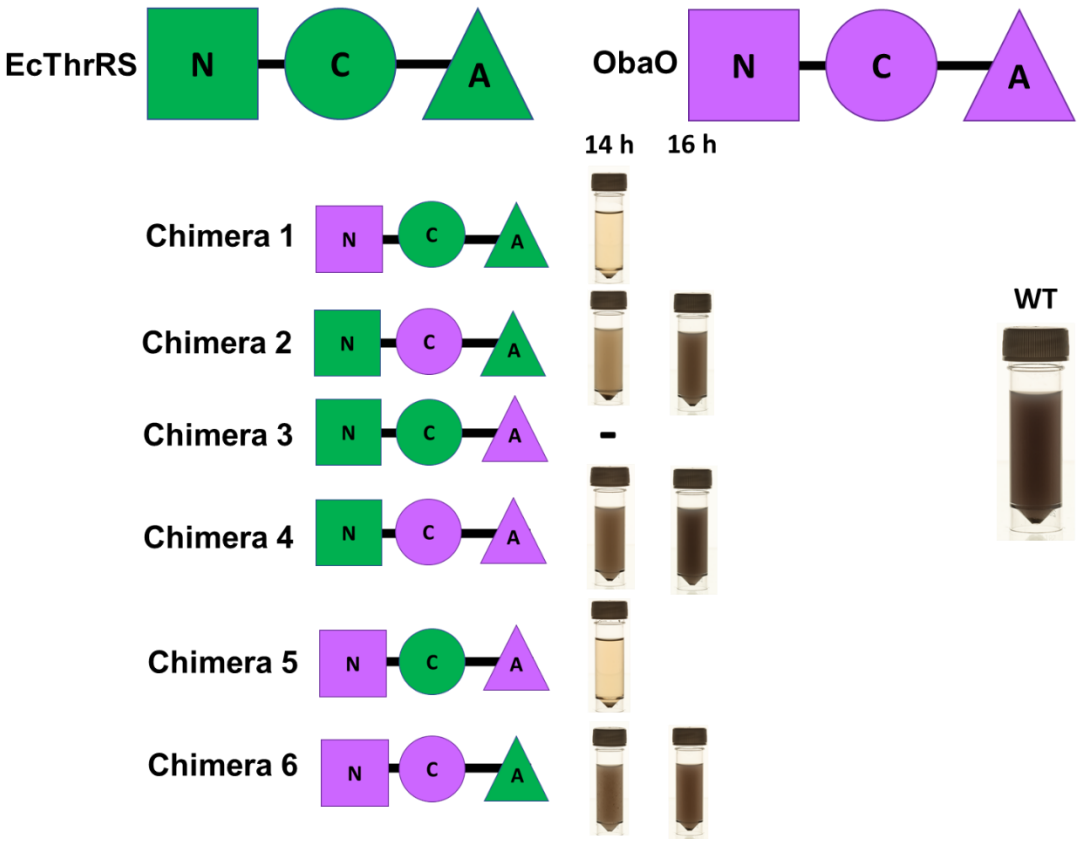


Figure 4.16. The ObaO catalytic domain is required to confer obafluorin resistance, in EcThrRS:ObaO chimeric protiens. EcThrRS domains shown in green, ObaO domains shown in purple. WT = wild-type *P. fluorescens*. N = N-terminal editing domain, C = aminoacylation catalytic domain, A = anticodon binding domain. For chimeras which conferred obafluorin resistance (observed by growth of *P. fluorescens* Δ*obaO*Δ*obaL* when grown in the presence of with 2,3-DHBA); aliquots of cultures pictured after 14 and 16 hours growth. Resistance is visualised by the development of a darker colour corresponding to obafluroin production. All chimeras containing an ObaO catalytic domain conferred resistance to obafluorin. Dr Sibyl Batey, unpublished.

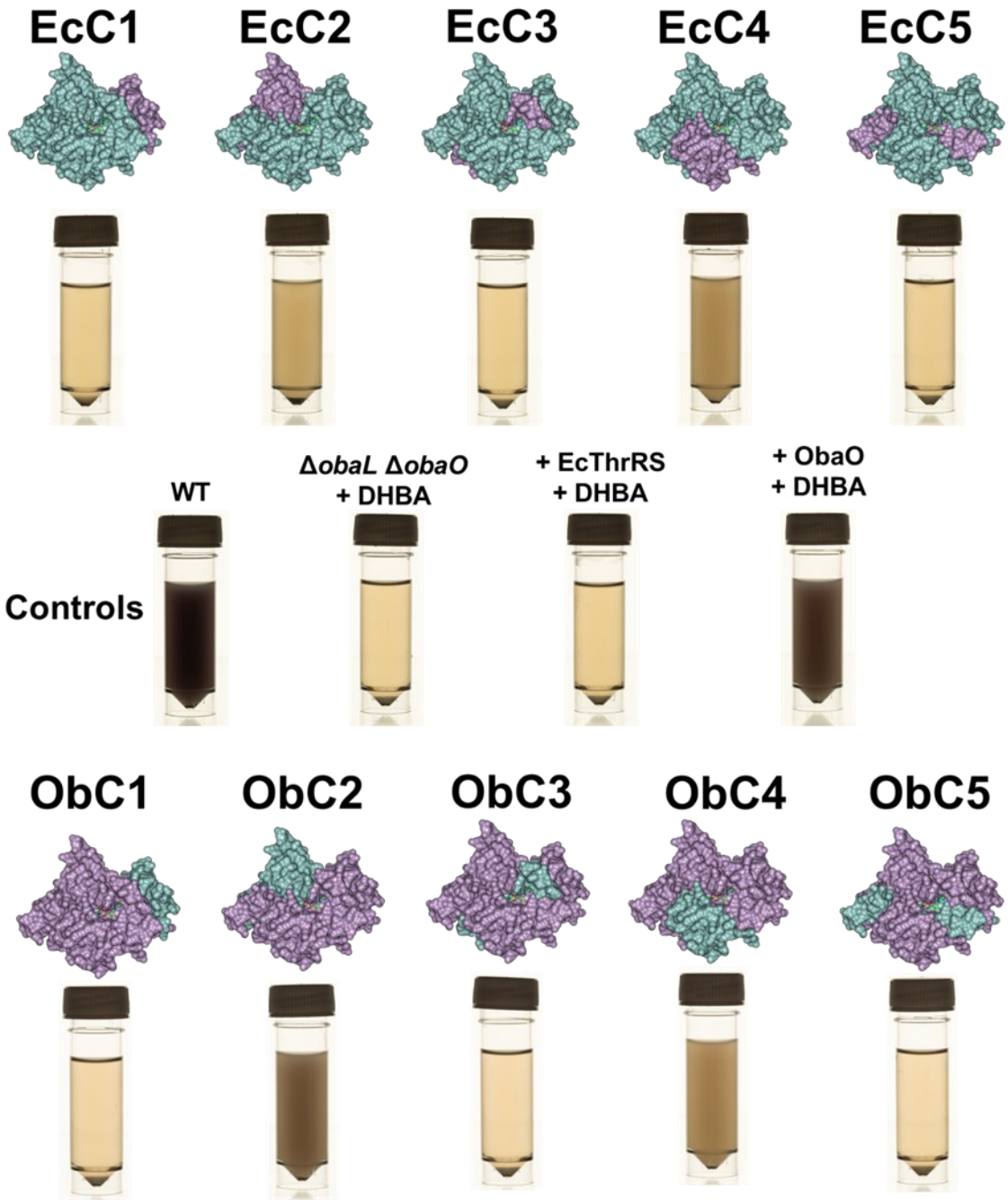


Figure 4.17. Catalytic subdomains 2 and 4 from ObaO are vital for conferring obafluorin resistance. EcThrRS domains shown in teal, ObaO domains shown in purple. Catalytic domain of PfObaO Alphafold model shown, in surface representation. WT = wild-type P. fluorescens, no 2,3-DHBA, all other samples grown in the presence of 2,3-DHBA. Catalytic domain split into the pictured subddomains C1-5 with EcC1 being EcThrRS C2-5 and ObaO C1, which ObC1 is ObaO C2-5 and EcThrRS C1. Obafluorin resistance observed by growth of *P. fluorescens* Δ*obaO*Δ*obaL* grown in the preence of 2,3-DHBA; aliquots of cultures pictured after 17 hours growth. Resistance visualised by the growth and development of a darker colour, corresponding to obafluorin production. Dr Sibyl Batey, unpublished. At least some resistance observed in EcC2, EcC4, ObC2 and ObC4.

4.2.9 The EcThrRS E305K mutant protein shows some resistance to obafluorin

By examining the multiple sequence alignment of the *Pseudomonas* housekeeping ThrRS, various ObaO homologue's and EcThrRS protein sequences, residue E305K was the only obvious difference in the subdomain chimera EcC2 between EcThrRS and ObaO that might constitute a potential extra site for obafluorin acylation. This residue site is located near the protein active site but points out towards the solvent, rather than into the active site.

This mutation was introduced into the EcThrRS-encoding sequence and cloned into pJH10TS, and the resulting plasmid transformed into *E. coli* NR698. The resulting strain was tested for obafluorin in spot-on lawn bioassays which indicated by a shift of MIC from 64 µg/mL obafluorin in EcThrRS WT to 1000 µg/mL in EcThrRS E305K- see Figure 4.18. This therefore suggests that while this mutation may be having an effect on self-resistance, it is insufficient to be the main resistance determinant, also explaining the difference observed between resistance of EcC2 and EcC4 previously. We can therefore determine that the main resistance determinant is most likely found in catalytic subdomain C4. The observed resistance is likely from titrating away enough obafluorin by acylating this additional surface lysine which will not affect the activity of the protein, reducing the effective concentration of obafluorin.

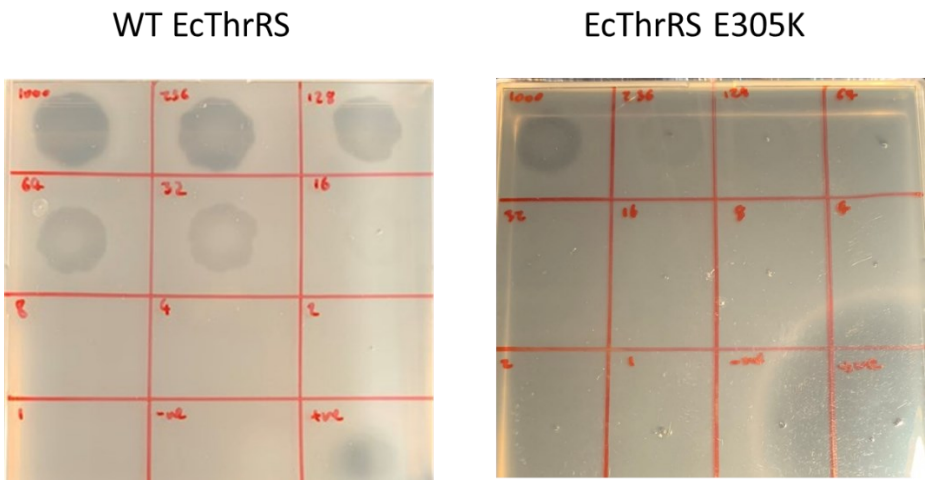


Figure 4.18. EcThrRS E305K confers a higher MIC to obafluorin than WT EcThrRS. *E. coli* NR698 expressing WT EcThrRS or EcThrRS E305K were inoculated into soft nutrient agar to pour the plates before spotting on of 4 µL obafluorin (dissolved in MecN) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of MeCN only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates incubated at room temperatrue for 16 hours.

4.2.10 Cryo-EM structures identify the mechanism of action of obafluorin by identifying Y462 as the point of attachment.

As noted already, we were unsuccessful in obtaining crystals of EcThrRS in the presence of obafluorin. This is possibly due to the generalised alkylation of EcThrRS by the excess of obafluorin,

or some other issue linked to the reactivity of obafluorin with components of the crystallisation buffer. Additionally, we were also unable to generate diffracting crystals of apo-ObaO (section 4.2.1 and 4.2.7). As such, a Cryo-EM approach taken as this technology deals with a protein in solution and can address significant heterogeneity. Additionally, we expected that random obafluorin alkylation events would average out, meaning that only obafluorin molecules binding in a specific place would be detected.

Pre-screening of EcThrRS by negative staining electron microscopy without the addition of obafluorin was first undertaken by Jake Richardson of the JIC Bioimaging Platform in order to verify that the protein would be suitable for cryo-EM data collection. This data was then analysed by Dr Dmitry Ghilarov as seen in Figure 4.19. With EcThrRS appearing to not precipitate when grids are prepared, and with a low-resolution model of EcThrRS built from this data, samples of EcThrRS and ObaO with obafluorin were sent to Leeds University for grid preparation and data collection.

A structure of EcThrRS was solved to ~2.9 Å resolution: this is the first ThrRS structure to have been solved by Cryo-EM. The map was produced by Dr Dmitry Ghilarov (JIC); using this map, I built the EcThrRS atomic model and refined it, using quantum mechanics (QM) methods to generate the restraints for both obafluorin in its ring opened state, and its covalent bond to the tyrosine residue Y462. Obafluorin can be seen in the ring opened form, and is covalently attached to the hydroxyl group of Y462 (Figure 4.20). Based on the positioning of the refined side chains, it would be reasonable to assume that the nitrobenzyl moiety of obafluorin makes π-stacking interactions with R363 and Y313. These are the residues known to form a π-stacking sandwich with the terminal adenosine of the tRNA (A76)[115]. The catechol moiety appears to coordinate the zinc ion through both phenoxy groups, mimicking the binding mode of threonine.

A structure of ObaO with obafluorin bound was also solved to ~3 Å. The overall map quality was lower than that for EcThrRS (seen in Figure 4.21D), mainly because of the problems with preferential orientation of particles which led to the decreased number of available high-resolution images. Future data collections with the addition of detergents or using support films could be done in order to mitigate this. The ObaO map displays some Coulomb potential density in the active site and had a conformation similar to that of the EcThrRS:obafluorin bound structure (see Figure 4.22). This indicates that obafluorin could be bound in the active site, but potentially at low occupancy. The obafluorin molecule was built into the model, but the map quality is poor (see Figure 4.21). All of the observed binding interactions seen for EcThrRS appear to be present, except that the Y463 (Y462 in EcThrRS) appears to be in a different orientation (see Figure 4.21E). Of the nucleophilic amino acids, tyrosine is not the strongest nucleophile, so for obafluorin to be ring opened by it, it is likely that the obafluorin and tyrosine will have to be held in a specific orientation with an appropriate residency time. This would suggest that there is something specific about the

orientation of the tyrosine in EcThrRS and susceptible housekeeping proteins, when compared to ObaO. Full structure solution information for both of these Cryo-EM structures can be found in Supplemental Table 5.

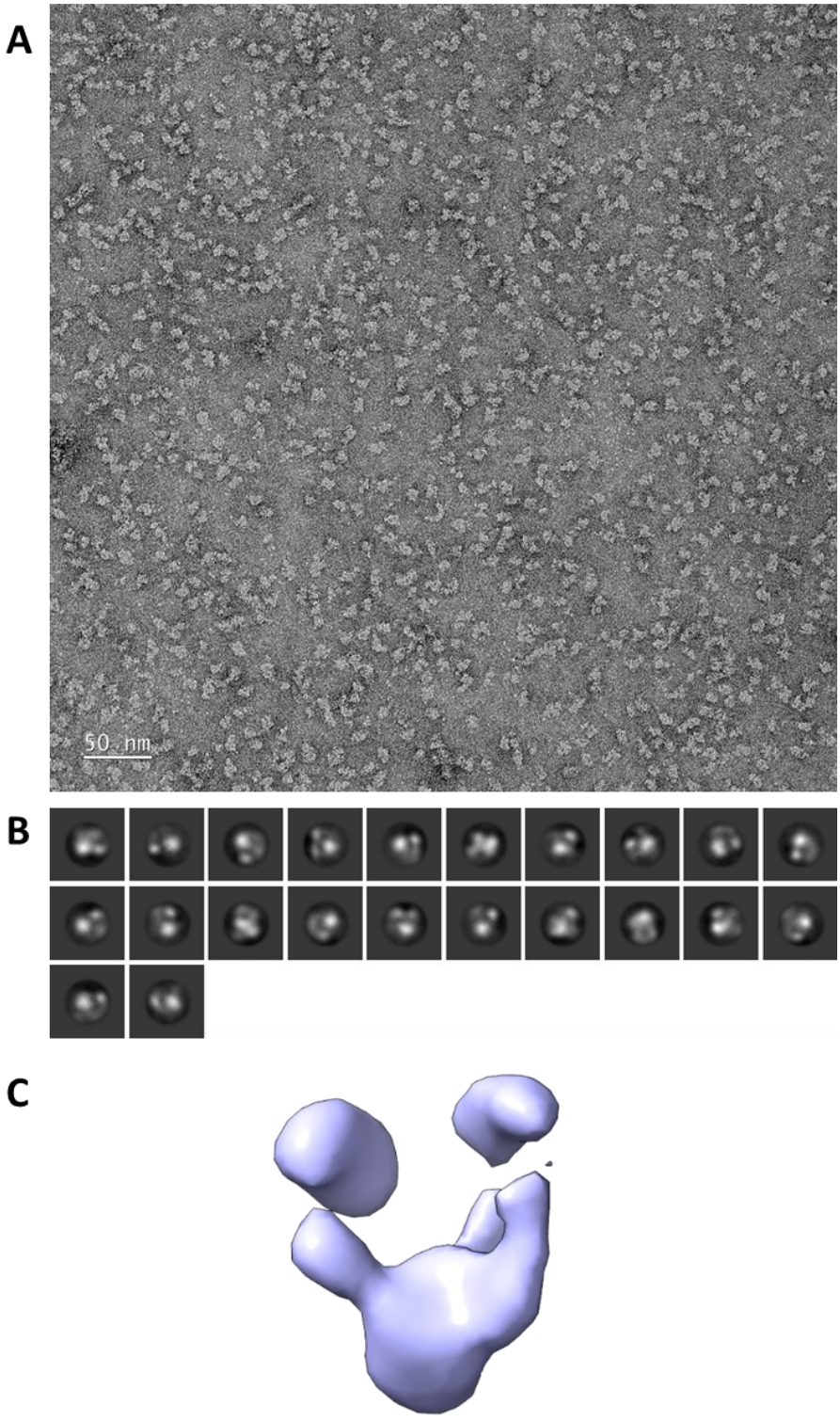


Figure 4.19. EM negative staining data for EcThrRS indicating that EcThrRS does not aggregate in grids, and that an low-resolution map can be built. A) Negatively stained transmission scanning electron micrograph (TEM) of EcThrRS imaged at 45000x magnification. **B)** Individual 2D classes collected from this data. **C)** Coarse grain EM model of EcThrRS derived from this data.

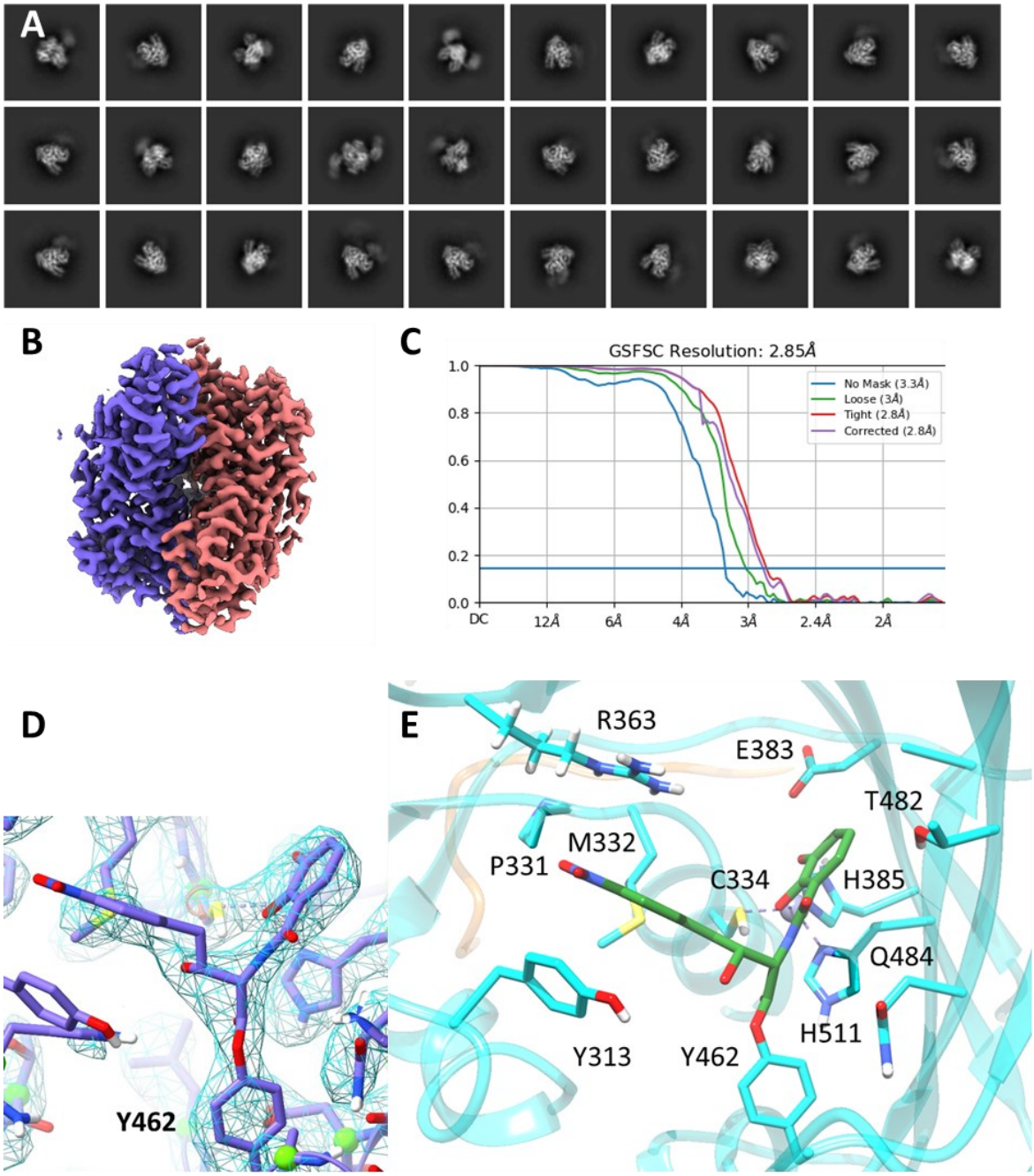


Figure 4.20. Cryo-EM data for EcThrRS showing that obafluorin forms a covalent atatchment to Y462. A) A selection of representative 2D classes for this dataset. **B)** An overall view map derived from this data, with one monomer in purple and the other in pink. (density for the editing domains is poor and not present at the selected contour level). **C)** Fourier shell correlation (FSC) curves for this data, indicating a resolution of 2.85 Å. **D**) Density map for the obafluorin molecule, showing clear bonding with the phenolic hydroxyl of Y462. Figure generated in ChimeraX. **E)** Refined structure of EcThrRS bound to obafluorin. Obafluorin or zinc-interacting residues are shown as sticks, obafluorin is shown as green sticks and the protein shown as cyan. Zinc is shown as a grey sphere. Figure generated in Chimera.

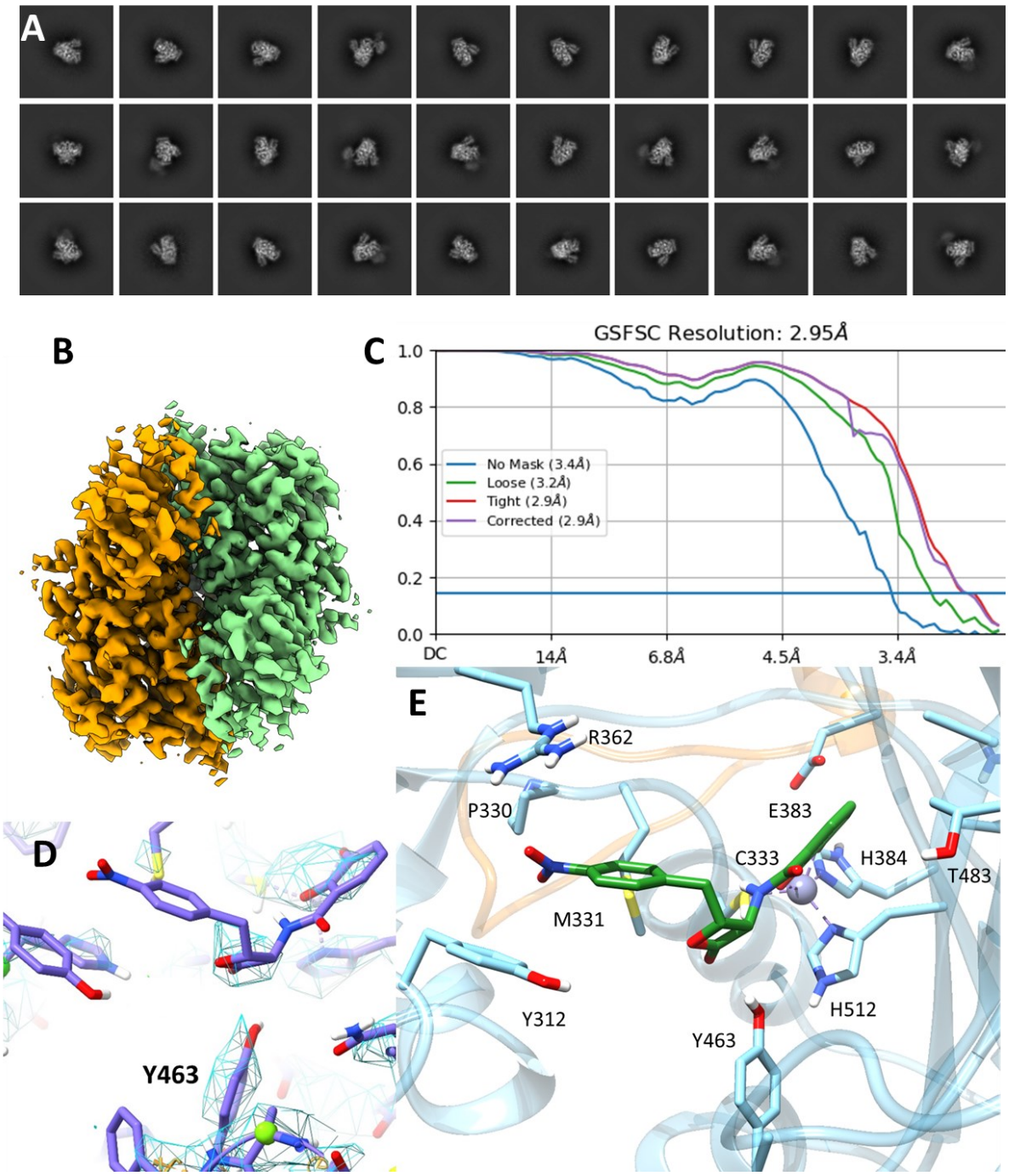


Figure 4.21. Cryo-EM data for ObaO showing that obafluorin may be bound at low occupancy, with no evidence of covalent attachment to Y463. A) A selection of representative 2D classes for this dataset. **B)** An overall view of the final map derived from this data, with one monomer in green and the other in orange. Editing domains are contoured out. **C)** Fourier shell correlation (FSC) curves for this data, indicating a resolution of 2.95 Å. **D**) Density map for the obafluorin molecule, showing clear bonding with the phenolic hydroxyl of Y462. Figure generated in ChimeraX. **E)** Refined structure of EcThrRS bound to obafluorin. Obafluorin or zinc- interacting residues are shown as sticks, obafluorin is shown as green sticks and the protein shown as light blue. Zinc is shown as a grey sphere. Figure generated in ChimeraX.

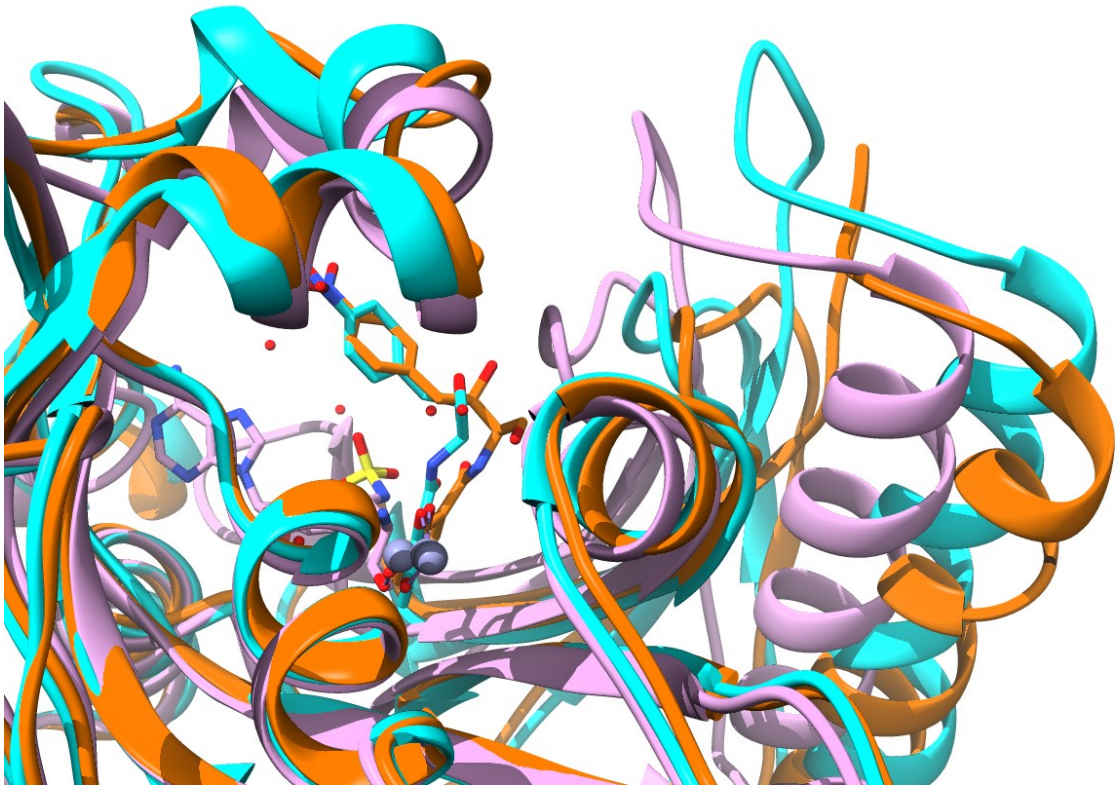


Figure 4.22. Crystal structures of EcThrRS bound to ThrSAA and Obafluorin, and ObaO bound to obafluorin. Ligands are shown in stick representation. EcThrRS bound to ThrSAA is shown in pink, ObaO bound to obafluorin in cyan and EcThrRS bound to obafluorin in orange. A large conformational shift is shown upon obafluorin binding vs ThrSAA binding, ObaO has a conformation at a mid-point between the two conformations sampled by EcThrRS. Figure generated in Chimera.

4.2.11 ObaO Y463F is an essential ThrRS residue

From crystal structures of ThrRSs with the substrates, Y462 has been identified as possibly vital for threonine binding, tRNA binding and aminoacylation[97,115], as discussed in Section 1.3.1 and Figures 1.11 and 1.15. To confirm the importance of the Y462 residue, the attachment site for obafluorin, a Y463F (equivalent to Y462 in EcThrRS) mutation was made in PfObaO in pJH10TS and transformed into *E. coli* NR698, and tested for obafluorin resistance as seen in Figure 4.23. PfObaO was chosen because if the phenolic hydroxyl was not vital for activity, we could observe obafluorin resistance when PfObaO was expressed in *E. coli* NR698. PfObaO Y463F was unable to confer resistance to obafluorin, suggesting that Y463 is essential for activity in PfObaO.

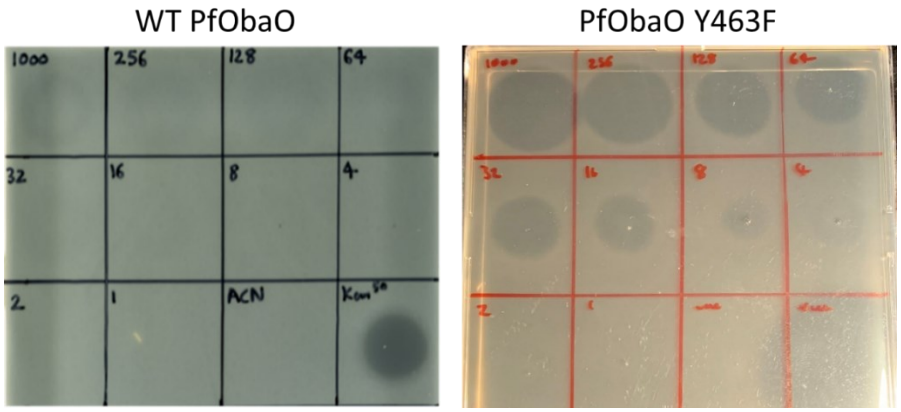


Figure 4.23. *E. coli* NR698 expressing PfObaO Y463F does not confer obafluorin resistance. *E. coli* NR698 expressing WT PfObaO or PfOBaO Y463F were inoculated into soft nutrient agar to pour the plates before spotting on of 4µL obafluorin (dissolved in MecN) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left, negative control of MeCN only bottom middle right, and positive control of 50 µg/mL kanamycin bottom right. Plates incubated at room temperatrue for 16 hours.

4.2.12 Spontaneous resistant mutants of the P. fluorescens ΔobaOΔobaL strain consistently found in displayed mutations corresponding to G462 in the PfThrRS gene.

One popular method for the identification of both the target of natural products, and for determining resistance mechanisms against them, is the generation of spontaneous resistant mutants. This is typically achieved by growth of the bacteria in the presence of sublethal concentrations of the antibiotic of interest. However, despite several attempts to achieve this with *E. coli* NR698 no mutants were generated (attempts were made by several previous group members). The reason for this was traced to the hydrolysis (and therefore inactivation) of exogenously added obafluorin on the necessary timescales. This is also related to the use of spot-on lawn bioassays rather than more typical disk diffusion assays: we and others have found that obafluorin reacts with the paper disk materials.

To circumvent this issue, we deduced that the *P. fluorescens* Δ*obaO*Δ*obaL* mutant could provide a solution to the problem. The logic was that if this strain was incubated in the presence of 2,3-DHBA, the only way the strain could produce obafluorin, and carry on growing, was to undergo mutation of the PfThrRS housekeeping gene. Moreover, obafluorin could be produced continuously and *in situ,* thus replacing material that degraded. This would be in contrast to the exogenous addition of obafluorin to *E. coli* strains which would rapidly degrade and not be replaced. Thus, Dr Sibyl Batey first attempted this approach by plating growing cultures of *P. fluorescens* Δ*obaO*Δ*obaL* onto agar plates containing 2,3-DHBA. Unfortunately, despite several variations, no resistant mutants were obtained.

Subsequently, this strain was used for a series of other experiments including its use as a ‘negative’ control, i.e., that when grown in the presence of 2,3-DHBA for the usual 14 hours, no growth was

observed, and the cultures did not turn purple (as no obafluorin was produced). However, on one occasion this strain was returned to the incubator and left for an extended period. It was noted that after 48 h the strain had begun to grow. Samples were then taken at 48, 72 & 96 hours, and at the latter time point and the cultures had turned purple indicating the production of obafluorin. This observation was consistent with our original hypothesis and indicated that an obafluorin resistant version of PfThrRS must have mutated. Additionally, Dr Batey had previously shown that WT colonies of *P. fluorescens* or colonies of *P. fluorescens* Δ*obaO*Δ*obaL* expressing ObaO (both of which are resistant to obafluorin) grow and turn a purple colour on solid OPM fed with 2 mM 2,3-DHBA. *P. fluorescens* Δ*obaO*Δ*obaL* on the other hand does not grow on solid OPM fed with 2 mM 2,3-DHBA. A purple phenotype and growth on solid OPM supplemented with 2mM 2,3-DHBA is therefore indicative of resistance to obafluorin. Selection of purple single colonies on OPM supplemented with 2 mM 2,3-DHBA is therefore a potential method for selection of spontaneous resistant mutants.

To then generate spontaneous resistant mutants purposefully, a set of 8 *P. fluorescens* Δ*obaO*Δ*obaL* cultures were grown in OPM, fed with 2 mM 2,3-DHBA and incubated at 30°C for 72 hours. Photographs were taken at 48 and 72 hours of growth, as seen Figure 4.24A and B. At 48 hours of growth, cultures 1, 4 and 5 had grown and darkened, indicating the production of obafluorin, and resistance to it. By 72 hours, all eight cultures had grown and darkened.

To select for single colonies of resistant *P. fluorescens* Δ*obaO*Δ*obaL*, samples of each of the eight cultures were then streaked out for single colonies on solid OPM plates supplemented with 2 mM 2,3-DHBA and incubated for a further 4 days. Colonies grew on all eight plates, but a small handful of colonies from culture number 8 had turned purple.

These single colonies were grown overnight in LB before then inoculating into fresh liquid OPM supplemented with 2 mM 2,3-DHBA to confirm that true resistance had been generated. Of the four colonies picked, two, 8.1 and 8.3, showed a purple colour, matching the wild-type at 14 hours. The housekeeping PfThrRS gene was amplified from these colonies and sent for Sanger sequencing to identify mutations that had arisen.

Dr Batey had also generated three mutants (SB2, 4 and 5) but without the selection for single colonies on solid, meaning that mixed populations were present in the sequencing. A full list of mutations can be found in Supplemental Table 6. Most of the mutations are silent (causing no change in amino acid), or in domains of the protein not predicted to be involved in resistance as judged from the chimera work described in section 4.2.8. In all samples, a mutation was found resulting in an amino acid switch in position 462 (G463 in EcThrRS, S464 in ObaO)- the sequencing chromatographs can be seen in Figure 4.25.

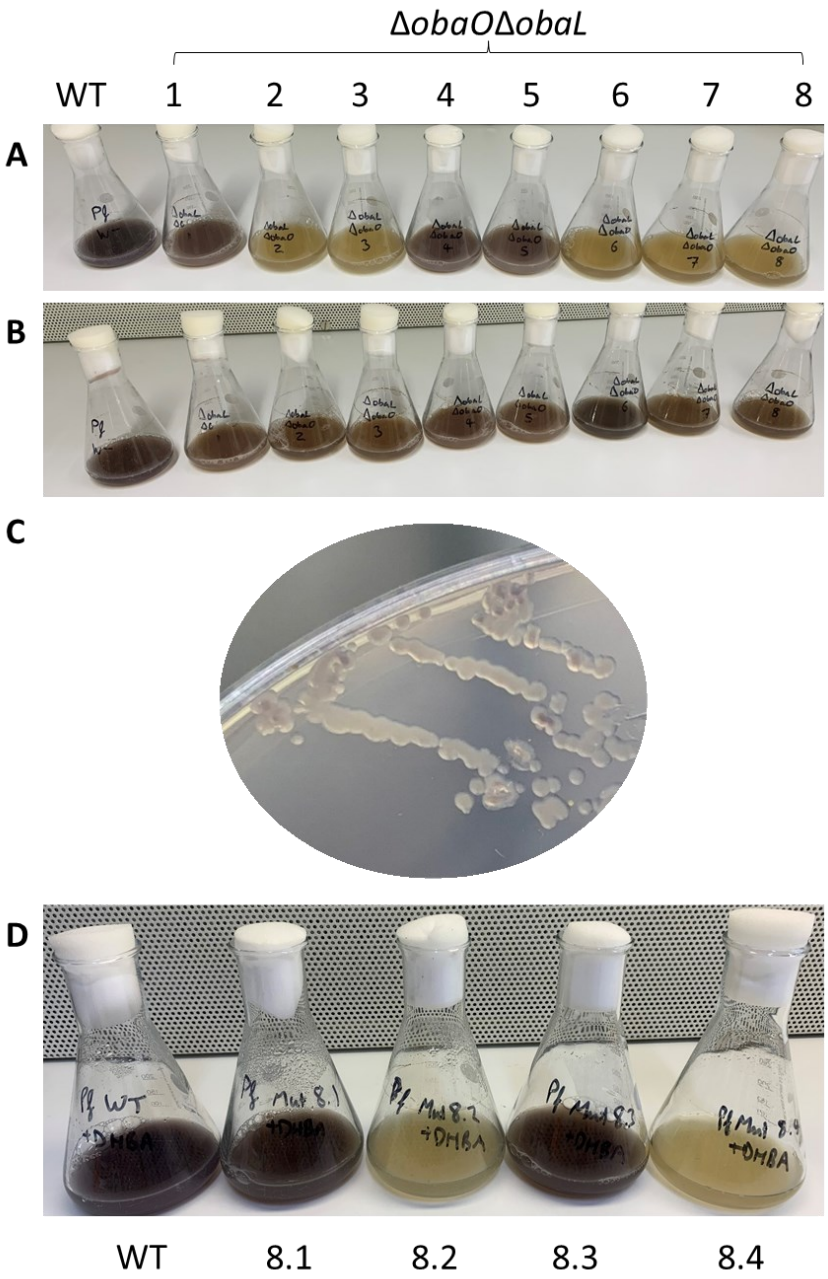


Figure 4.24. Spontaneous resistant mutants can be raised against obafluorin in *P. fluorescens* Δ*obaO*Δ*obaL*. Cultures of *P. fluorescens* Δ*obaL*Δ*obaO* and WT *P. fluorescens* as a control were grown at 30°C and fed with 0.2 mM 2,3-DHBA in liquid obafluorin production medium (OPM). The darkening of the culture is indicative of obafluorin production. **A) Cultures of each strain photographed after 48 h growth**; WT and cultures 1, 4 and 5 had grown and darkened at 48 hours. **B) Cultures of each strain photographed at 72 hours of growth.** All cultures had grown and changed colour by 72 hours. **C, Bacteria from culture 8 streaked out on solid OPM.** Media supplemented with 0.2 mM 2,3-DHBA. Some single colonies had turned purple after 4 days. **D, Spontaneous resistant mutants in *P. fluorescens* Δ*obaO*Δ*obaL* could be raised against obafluorin.** WT *P. fluorescens* and single purple colonies of resistant *P. fluorescens* Δ*obaL*Δ*obaO* were grown at 30°C and fed with 0.2 mM 2,3-DHBA in liquid OPM and incubated at 30°C for 14 hours. The darkening of the culture is indicative of obafluorin production. WT and cultures 8.1 and 8.3 had grown and darkened, indicating resistance to obafluorin.

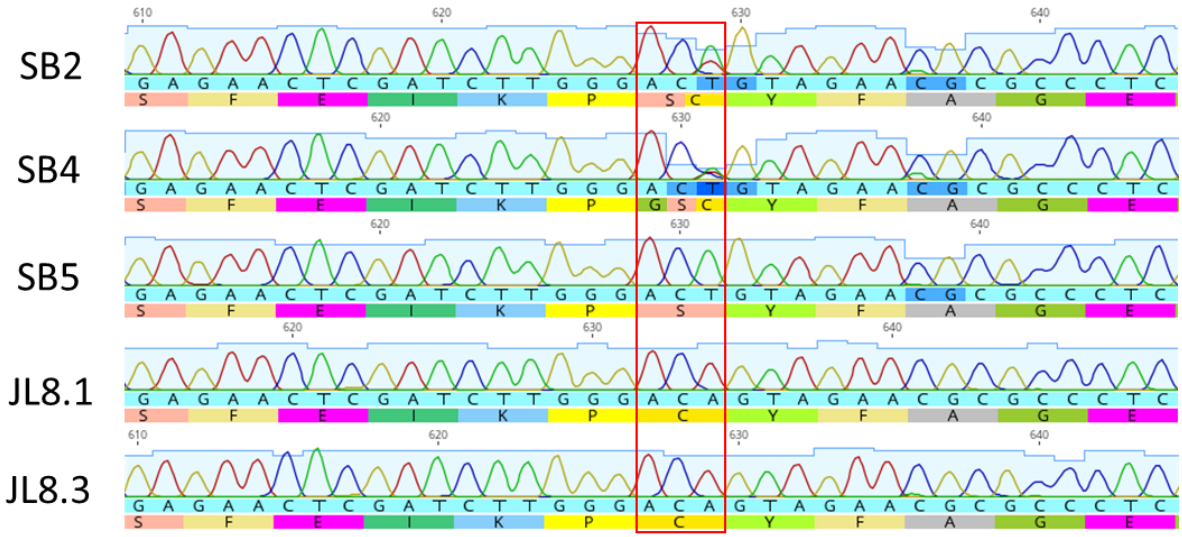


Figure 4.25. A consistent G463S/C mutation is observed in the PfThrRS gene in spontaneous obafluorin resistant mutants. SB2, 4 and 5 are from the resistant mutant cultures produced by Dr Sibyl Batey, JL 8.1 and 8.3 were produced by me, as seen in Figure *4.24*. Red box highlights the codon of interest, with all of the encoded amino acids in this position, for the mixed samples. Light blue shading of nucelotides indicates good quality sequencing, while darker blue indicates more poor quality sequencing. The amino acids encoded by each codon are shown below the nucleotide sequences. Figure generated in Geneious.

For all of the samples, a single nucleotide change from G to either A or T could be observed, leading to a mutation from a glycine to either a serine or a cysteine in the protein. This glycine, G462 (G463 in EcThrRS, S464 in ObaO) is the immediate neighbour of the tyrosine to which obafluorin covalently binds. In all of the ObaO homologues, a serine is observed in this position, and in all of the susceptible proteins, a glycine is present, so it is likely that this mutation is important for resistance. This position in the protein is in the C4 catalytic subdomain, the region identified as most likely to contain the resistance determinant in section 4.2.8. Glycine is unusual amongst the amino acids in that there is no β-carbon. This means there is an increased amount of conformational freedom for glycine, relative to all the other amino acids. Examining this part of the loop in both obafluorin-bound cryo-EM structures, it can be seen that there is a slight "bending" of the Y462 containing loop in EcThrRS at G463 when compared to ObaO- see Figure 4.26. This "bending" could perhaps be essential for proper positioning of the tyrosine with the obafluorin to facilitate acylation.

The ϕ and ψ angles describe the torsion angles of each amino acid in a peptide chain, as illustrated in Figure 4.27. Plotting the ϕ angle against the ψ angle for each amino acid in a structure gives us the Ramachandran plot. There are regions of the plot which are "favoured", "allowed", or "disallowed". Glycine, having more conformational freedom than most amino acids, can sample more areas of this chart without it being unfavourable. The Ramachandran plots for the obafluorin-bound proteins, specifically showing only G463 for EcThrRS or S464 for ObaO can be seen in Figure 4.28. The ϕ angle for the S464 in the obafluorin bound PfObaO structure was -173.7/-167.9° in

chain A and B respectively, with the ψ angle 154.7/158.4° in chain A and B respectively. This places the serine in the "allowed" region of the Ramachandran plot, in the top left quadrant of the plot. G463 in the obafluorin bound EcThrRS structure on the other hand, the ϕ angle is 163.2/160.3° in chain A and B respectively with the ψ angle 175.2/174.1° in chain A and B respectively. This places the glycine in the "favoured" region for glycines, in the top right quadrant of the plot. This conformation is disallowed for the serine and so it is unlikely that in ObaO, this conformation will be sampled, and so if the peptide needs to be in this conformation of obafluorin ring opening, mutation to serine or cysteine would prevent it.

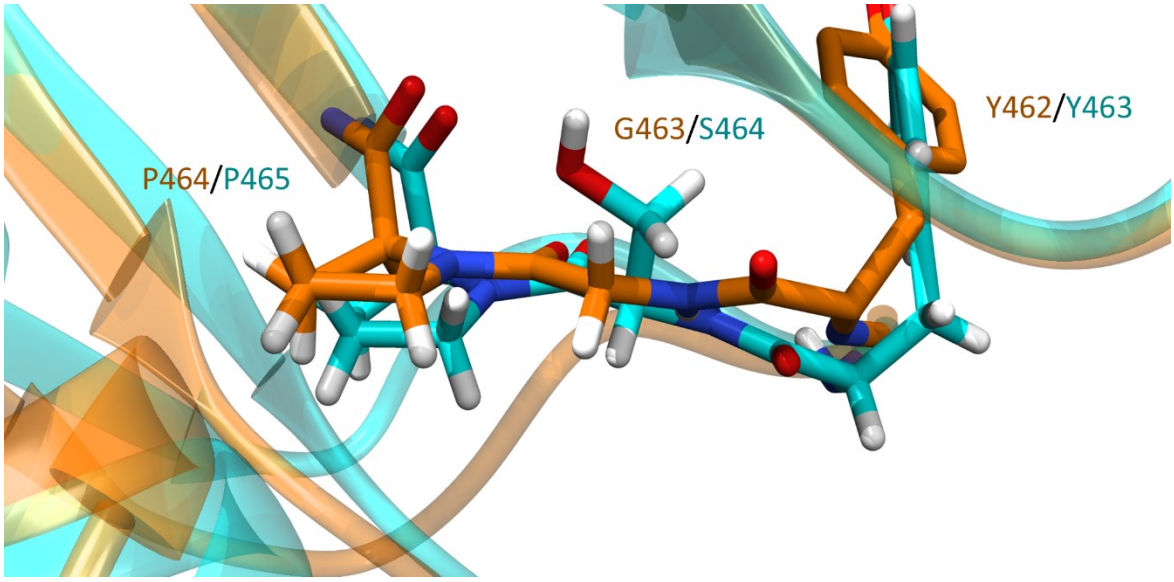


Figure 4.26. Comparison of the positioning of the P464/465, G463/S464 and Y462/463 in obafluorin bound EcThrRS and ObaO reveals a "bending" of the loop about G463/S464. EcThrRS shown in orange, ObaO in cyan. The amino acids on either side of G463 (S464 in ObaO) are shown in stick representation with the rest of the protein as cartoons. Figure generated in Chimera.

The G463S single point mutation had already been made by Dr Sibyl Batey in EcThrRS, and had no effect on resistance. This is intriguing because in the spontaneous resistant mutants, a G464S mutation in PfThrRS appears to be of conferring obafluorin resistance alone. In order to clarify what other mutations in PfThrRS relative to EcThrRS may be "poising" PfThrRS to become resistant to obafluorin, a structural alignment of the PfThrRS AlphaFold model and the EcThrRS and ObaO obafluorin-bound cryo-EM structures was analysed. When selecting all residues within 5 Å of the EcThrRS-bound obafluorin molecule in all of the structures, it was found that all residues that could be interacting with obafluorin were identical in EcThrRS and ObaO other than A316 in EcThrRS/N315 in PfThrRS (see Figure 4.29). In this position, in all of the *Pseudomonas* and *Burkholderia* proteins, whether housekeeping or resistant, an asparagine is found. This is positioned in the ObaO:obafluorin cryo-EM structure such that the asparagine could be interacting with the nitro group of obafluorin. This could lead to a level of constraint in the possible binding modes of

obafluorin in the active site which would prevent positioning of the obafluorin for ring opening when combined with the increased rigidity caused by the G463S mutation. The carbonyl of A316 has a hydrogen bonding interaction with the N6 of the adenosine ring of A76 from the tRNA in the EcThrRS:AMP:tRNA structure in PDB file 1QF6[77].

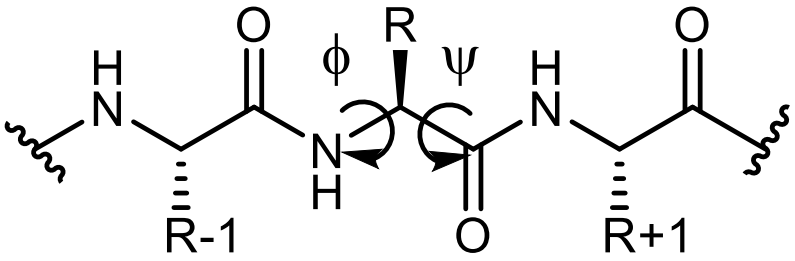


Figure 4.27. Schematic illustrating the phi (ϕ) and psi (ψ) angles of a given amino acid. With side chain R. The preceding amino acid side chain shown as R-1 and the following amino acid side chain shown as R+1.

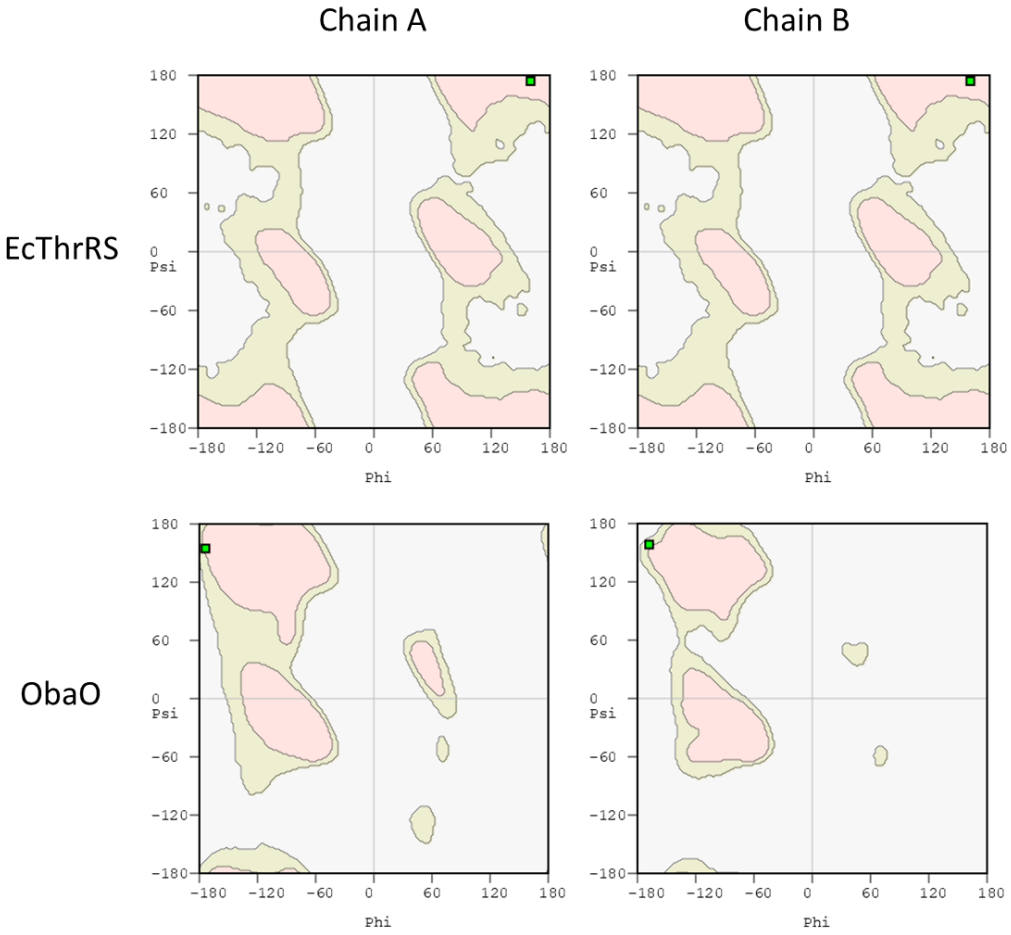


Figure 4.28. Ramachandran plots for the EcThrRS and ObaO obafluorin bound structures. Pink zones are preferred regions, yellow are allowed regions. The green square denotes G463 for EcThrRS and S464 for ObaO. Generated in Coot.

Examination of a global alignment of ThrRS sequences (discussed in detail in Chapter 5) reveals that this potential resistance residue is a glycine in all ThrRSs, except the ObaO homologues where it's

a serine, four other bacterial ThrRSs from *Phenylobacterium zucineum*, *Erythrobacter litoralis*, *Sphingopyxis alaskensis* and *Novosphingobium aromaticivorans* in which it is an alanine, and all of the *Archaeal* ThrRSs, where this position is variously valine, isoleucine, serine, asparagine, glutamate and leucine. If this glycine is indeed key for obafluorin sensitivity, it would appear that almost all bacteria and all eukaryote ThrRSs could be sensitive.

In summary, these resistant mutants, generated using a non-standard methodology, give us a strong indication as to the resistance mechanism of ObaO, which should now be further followed up.

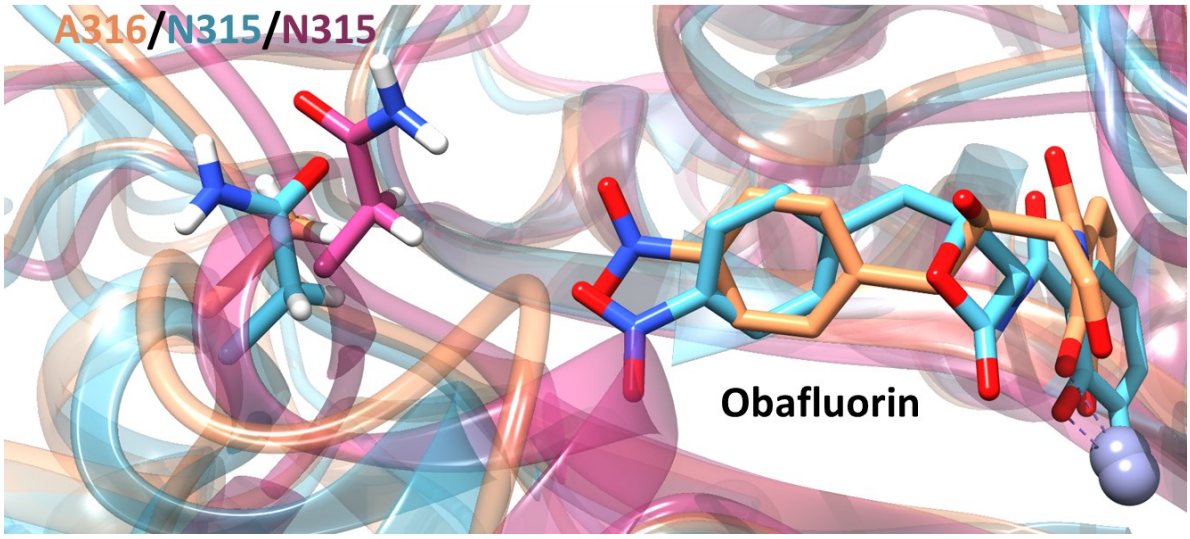


Figure 4.29. Structural alignment of obafluorin bound EcThrRS and ObaO and the Alphafold2 model of PfThrRS. Protein shown in cartoon representation, A316/N315 and obafluorin shown as sticks. Zinc ions shown as grey spheres. EcThrRS shown in orange, ObaO shown in blue, PfThrRS shown in pink. Generated in Chimera.

4.2.13 BorO is likely sensitive to obafluorin, but we cannot test it

As discussed in Chapter 3, BorO is not functional in *E. coli*; this means bioassays cannot be used to test for obafluorin susceptibility of *E. coli* heterologously expressing BorO. Given the similarity in the tRNA genes between *E. coli* and *Pseudomonas*, it is also likely that BorO will not be functional in *P. fluorescens*. When expressed heterologously in *P. fluorescens* Δ*obaO*Δ*obaL* and grown in the presence of 2,3-DHBA, BorO did not confer resistance, as seen in Figure 4.30. Given that we think that G463 is a key residue required for obafluorin sensitivity, examination of the BorO protein sequence revealed a glycine in this position (G490), suggesting that BorO is likely sensitive. In order to test this, the *S. venezuelae* strain discussed in Chapter 3 could be used for heterologous expression. However, spot-on lawn bioassays with Wild-type *S. venezuelae* revealed that it is naturally resistant to obafluorin (Figure 4.31). Consistent with this the *S. venezuelae* housekeeping enzyme SpThrRS also has glycine in this position (G504).

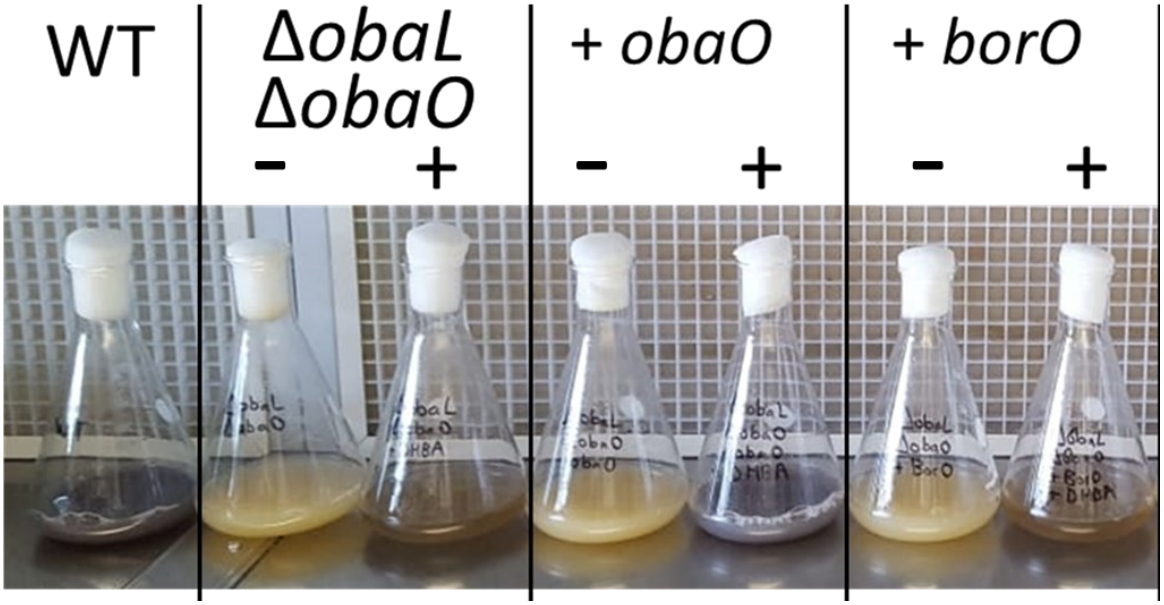


Figure 4.30. Expression of BorO in *P. fluorescens* does not confer resistance to obafluorin. Assayed by expression in *P. fluorescens* Δ*obaL*Δ*obaO* and growing in the presence of 2,3-DHBA. Cultures grown with (+) and without (-) 2,3-DHBA pictured at 17 hours. Purple colouration and growth when fed with 2,3-DHBA denotes obafluorin resistance. + indicates the addition of 2,3-DHBA; - indicates no addition. BorO expression construct built by me, bioassays performed by Dr Sibyl Batey.

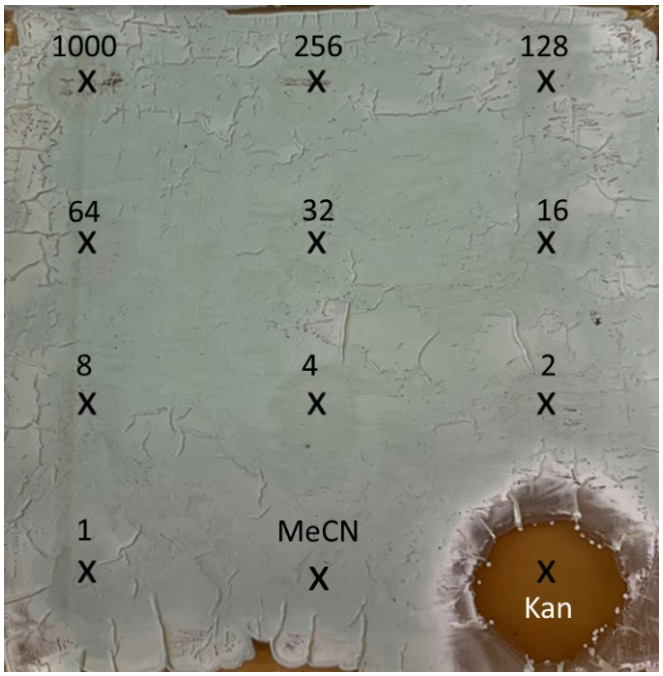


Figure 4.31. *S. venezuelae* is resistant to obafluorin. Spores of *S. venezuleae* were spread for confluence on MYM media before spotting of 4 µL obafluorin (at location marked X, dissolved in MeCN) at concentrations of 1000, 256, 128, 64, 32, 16, 8, 4, 2 and 1 µg/mL spotted from top left to bottom left as indicated, negative control of MeCN only bottom middle, and positive control of 50 µg/mL kanamycin bottom right. Plates were then incubated at 30°C for 2 days.

4.3 Conclusions and Future Work

4.3.1 A model for the mechanism of action of obafluorin

Using a combined structural, spontaneous mutation and bioactivity approach, we have ascertained that the mechanism of action of obafluorin involves covalent attachment to the phenolic hydroxyl group of a key active site tyrosine residue (Y462 in *E. coli* ThrRS) via ring opening of the β-lactone moiety. The nitrophenol moiety for interaction with R363 and Y313 via π stacking and binding via the catechol moiety to the metal ion bound in the active site, likely blocks binding of threonine in the active site, as well as poising obafluorin in the correct orientation for tyrosine to attack the β-lactone ring, preventing Y462 from participating in aminoacylation. The nitrophenol moiety is therefore utilising the same π-stacking sandwich that the adenosine ring of A76 of the $tRNA^{Thr}$ uses to bind in the catalytic domain active site. Additionally, the catechol moiety binds to the zinc in the same square pyramidal configuration that threonine does. It could therefore be viewed that the attached obafluorin acts as something of a threonyl tRNA analogue.

The structure of obafluorin bound EcThrRS shows that there is a large conformational change when obafluorin is bound. The loop which bears Y462 is displaced roughly 4 Å. This then leads the next loop along to be displaced by roughly 6 Å. This means that the active site is held in an open configuration, relative to the protein when ThrSAA is bound. A similar phenomenon was observed in borrelidin binding and inhibition, as discussed in Chapter 3.

4.3.2 Model for the ObaO obafluorin partial resistance mechanism

Our cryo-EM structure data suggests that the mechanism of resistance by ObaO appears to involve obafluorin binding non-covalently (and with low occupancy), with attack by the key tyrosine (Y463) limited due to obafluorin adopting a sub-optimal configuration that does not allow for ring opening. These properties appear to be conferred, at least in part, by the neighbouring residue S464, which is a glycine in susceptible proteins. Our analysis suggests that this change of residue limits the possible conformations which the protein adopt; this may also limit the residency time of obafluorin in the active site.

The importance of the ObaO S464 residue (G463 in PfThrRS/EcThrRS) was further highlighted by the presence of this mutation (and cysteine changes) in spontaneous resistant mutants, and the strict conservation of glycine at this position in sensitive enzymes (this residue is always serine in ObaO homologues). Mutation to a more conformationally restrained amino acid (as opposed to glycine) could be the key factor preventing optimal alignment of obafluorin and Y463. However, it appears that there must be an additional factor in the PfThrRS structure which predisposes it to gain of resistance with this single amino acid change; introducing the equivalent G463S mutation into EcThrRS does not yield obafluorin resistance. This additional factor could be Q315 which is an alanine in EcThrRS. From the structural alignments analysed in section 4.2.12, Q315 was identified

in both PfThrRS and ObaO which is not present in EcThrRS and could be poising PfThrRS to gain obafluorin resistance by mutation of G463 to serine or cysteine. Q315 may interact with the nitro group of obafluorin, affecting its positioning in the active site of PfThrRS and ObaO. By comparison, the smaller bulk and lack of hydrogen bonding capability of the side chain of alanine could be allowing more conformational freedom in EcThrRS. Non-covalent binding of obafluorin is likely to be reversible and the open conformation that is likely to be required for obafluorin binding is likely to be only a subpopulation of ObaO in the cell. These factors could explain the observed partial inhibition of ObaO; while obafluorin-bound ObaO will be incapable of aminoacylating tRNA, obafluorin is likely to dissociate, thus allowing a sub-population to always be active. Additionally, the nitrobenzyl group appears to be binding in the tRNA binding pocket, while the catechol group and β-lactone group bind in the threonine binding pocket. In general, proteins have less affinity for their products than their intermediates and substrates, so it likely that binding to ObaO is reasonably transient.

Obtaining the information that allowed us to propose a model for the obafluorin mechanism of action was challenging, in large part due to the chemical reactivity of obafluorin. Thus, in order to dissect the kinetics of the non-covalent binding of obafluorin to proteins, in particular resistant enzymes, it may be necessary to generate a more stable analogue. One possibility is to generate a β-lactam version, which would be roughly iso-steric with obafluorin. Beta-lactams are much less reactive than β-lactones, and this analogue would be significantly less likely to ring opon following nucleophilic attack by the relatively weak tyrosine hydroxyl nucleophile, but it should be able to form all of the other non-covalent bonding interactions required for obafluorin action. This would give us the opportunity to measure kinetics of binding and unbinding (giving an idea of whether residency time in the active site is an important factor), as well as potentially permitting solution of a crystal structure.

Other, secondary mechanisms of self-resistance within ObaO could include those conferred by, for example, ObaO K305 which appears to confer partial obafluorin resistance (as shown in section 4.2.9), from which surface decoration by obafluorin could occur, reducing the number of unhydrolyzed obafluorin molecules available to bind in the active site and thus titrating away the compound. Residue M490 in ObaO (L489 in EcThrRS) could be “sequestering” Y463 (the target residue of obafluorin) via a sulfur-aromatic interaction, pinning Y463 in place, and preventing the optimal orientation of the obafluorin molecule (and thus covalent attachment). This residue may also prevent the conformational change and active site opening which appear to be essential for obafluorin binding, in the same way that it appears to be for borrelidin binding. This convergent contribution to the resistance to both of these molecules is interesting, as it is unknown if borrelidin

has been “seen” by the obafluorin producer or its ancestors as ObaO was evolved. In the partially solved model for apo-BmObaO, it appears that M490 and Y463 might be interacting.

4.3.3 Future Work

With a model for obafluorin mechanism of action and resistance in hand, some further work is required to fully elucidate it. The G463/S463 mutations should be purposefully introduced to PfThrRS and expressed in *E. coli* NR698 and *P. fluorescens* Δ*obaO*Δ*obaL* to assay for resistance and to confirm that this single point mutation is responsible for conferring resistance in the resistant mutants generated in section 4.2.12. In addition, additional spontaneous resistant mutants should be independently generated in order to confirm that selection of this point mutation is consistent. Secondly, an EcThrRS double mutant with the A316Q mutation in combination with the G463S/C mutations should be constructed to confirm if this mutation in the *Pseudomonas* genes is what “pre-disposes” them for gain of resistance by this second mutation. The L489M and G463S/C mutations could also be made in EcThrRS to probe as to whether these mutations can work combinatorially, along with L489M and A316Q to confirm the importance of specifically the flexibility of G463 in obafluorin sensitivity, or if general flexibility in the binding pocket is the factor permitting obafluorin to acylate Y462. Additionally, a S464G mutation could be introduced to PfObaO. If this mutation can prevent PfObaO from conferring obafluorin resistance *in vivo*, this could tell us that it is required for resistance. Mass spectrometry of the EcThrRS:obafluorin complex and ObaO:obafluorin complex to confirm acylation at Y462 should be done with the lower concentration of obafluorin.

To explore the role of the G463 on the flexibility of EcThrRS, molecular dynamics (MD) simulations could be used. The EcThrRS G463S and PfObaO S464G mutations could be introduced *in silico* to the Cryo-EM structures of the two proteins (reported in section 4.2.10) and MD simulations run on the WT and mutant structures and the trajectories of the proteins analysed. Analyses of these trajectories would allow us to estimate the effect of a G463S mutation in EcThrRS or a S464G mutation on the flexibility of the protein around the obafluorin target, Y462 in EcThrRS/Y463 in PfObaO. Additionally, the conformations sampled by the protein could be analysed for those required for obafluorin binding. It would also be useful to solve a crystal structure of the PfThrRS protein, as well as PfThrRS G462S mutant, as ideally their cryo-EM structures in the presence of obafluorin.

A synthesis of a β-lactam version of obafluorin would be of great interest- this could be used as an obafluorin analogue in crystal trials, being less susceptible to nucleophilic attack. This compound could also be used to probe obafluorin resistance by ObaO; if EcThrRS and other sensitive enzymes were to show only partial inhibition to a β-lactam obafluorin analogue, and not be acylated by it, this would support the hypothesis that the partial inhibition phenotype is due to transient, non-

covalent obafluorin binding. If susceptible proteins are still acylated but the compound is less prone to ring-opening in neutral/basic conditions, then this could be an improved analogue for pre-clinical investigations.

Additionally, a probing of the evolutionary origins of ObaO would be interesting. This could give insight as to both the emergence of obafluorin biosynthesis and resistance in nature but also potentially lead to the discovery of novel ThrRS inhibitors.

Looking forward, testing of obafluorin for toxicity in eukaryotic cells would be of interest. This could be done by assaying for inhibition of the human protein and by testing for toxicity in human cell lines or *Galleria mellonella* larvae. The *Galleria* model has been used in our lab in multiple other projects (unpublished).

Chapter 5: Utilising knowledge of threonyl tRNA synthetases as resistance mechanisms for genome mining for novel antibiotics

5.1 Introduction

While the generic cycle of randomly culturing microbes from the soil, isolating those strains with bioactivity against a target bioindicator strain, and then identification of the active compounds has been used with great success for many years, there is a major problem of rediscovery[239]. It is therefore desirable to dereplicate strains for novelty, i.e., to focus effort on strains with the potential to produce novel natural products. Extraction of gDNA, sequencing of the genome, and subsequent analysis by BGC identifying software such as antiSMASH[5,206-210] provides a snapshot of the biosynthetic potential of any strain. Novelty can then be defined by selecting for genomes containing BGCs which are dissimilar to those that have already been discovered. This process is called genome mining.

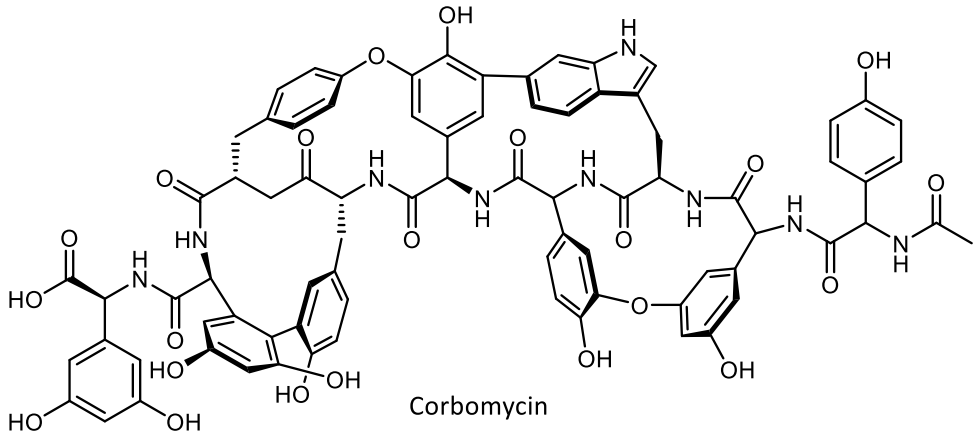
Since genome sequencing became cheaper and easier, the genomes of bacteria have been sequenced and deposited into databases such as the NCBI GenBank[240] at an astounding rate. This means that we can mine the genomes of bacteria for biosynthetic potential without having to culture the bacteria.

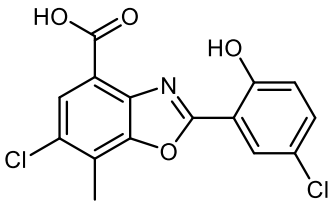
One fascinating example of a natural product discovered by genome mining was corbomycin (Figure 5.1), identified by using a phylogenetic approach[234]. Corbomycin is a glycopeptide. Other examples of glycopeptides include vancomycin, a clinically used antibiotic which inhibits cell wall synthesis, and which is produced by *Amycolatopsis orientalis*[241]. Self-resistance to glycopeptides is conferred by an *N*-methyltransferase which deactivates the glycopeptide. The evolutionary history of glycopeptide antibiotics were probed by phylogenetic reconciliation[235] and revealed that ancestral glycopeptide BGCs first assembled roughly 400 million years ago, and then diverged to give the different types of glycopeptide BGCs that exist today. Further analysis of this reconciliation, using the presence of known self-resistance genes, then led to the identification of the BGC for corbomycin, which has a divergent evolutionary history to vancomycin, but with no dedicated self-resistance determinant. Corbomycin inhibits autolysins, which are proteins essential for peptidoglycan remodelling of the cell wall, a process which is required for bacterial cell growth. Inhibition of autolysins is a novel activity for glycopeptides. Corbomycin has been shown to be effective against MRSA in a mouse infection model, with a high barrier to the formation of spontaneous resistance and so are highly promising for development as a candidate antibiotic[234].

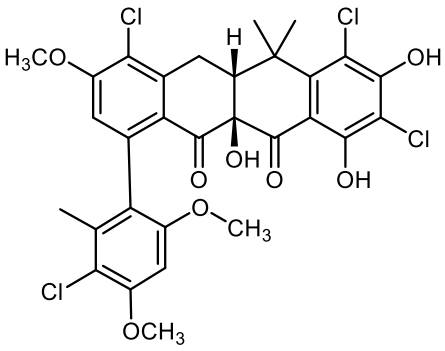
The use of phylogenetics for BGC discovery is a highly intriguing methodology. We have huge amounts of genome sequence data for the many different families of biosynthetic proteins and resistance genes, and so exploring uncharacterised branches of phylogenetic trees is an excellent place to look for novelty.

Another attractive strategy for the discovery of novel natural products is to target a specific self-resistance protein family of interest; in our case, ThrRSs. This is useful because you have knowledge

of the target and resistance mechanism for the natural product produced by the BGC of interest. For example, thiolactomycin (Figure 5.1) was discovered by systemically searching for duplicated fatty acid synthase genes that were co-located in a BGC. One such gene was identified with similarity to PtmP3, the platensimycin self-resistance gene, discussed in section 1.2.4. The BGC encoding this putative self-resistance gene was heterologously expressed and the product characterised as thiolactomycin. This compound had already been identified, so this same resistance gene was used as the bait to search for more clusters via a homology search- this then led to the identity of a series of thiolactomycin analogues[242]. This is the method which we would like to use to identify novel ThrRS inhibitors, by searching for novel potential ThrRS self-resistance genes.







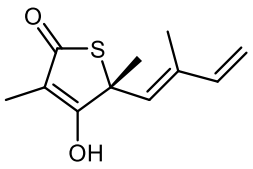


Figure 5.1, *S*elected examples of natural products identified by different genome mining strategies.

Another strategy for genome mining is to search environments which have historically been poorly sampled, extract environmental DNA, culture strains, or directly extract the metabolome of the environment. To date, most antibiotic discovery has looked at temperate soil derived strains for discovery. Looking in understudied environments such as polar or desert environments[243] can reveal both chemistry and biology which are rarely found in typical soil microbes. For example, this

year, the first novel natural product to be obtained from an atmospheric *Streptomyces* sp. was identified- caboxamycin B (Figure 5.1). The producer organism was isolated from storm clouds in a study which highlights clouds as a potential target for future sampling efforts[244]. Other underexplored ecological sources include environments where microbes and eukaryotes cohabitate- this can include insect-associated strains such as those in fungus farming ant colonies, those derived from marine sponges, plant rhizospheres, human microbiota or bacteria from other anthropogenic niches such as wastewater plants, city air or contaminated soil.

An example of a strain isolated from an understudied and competitive ecological niche which yielded useful natural products is *Streptomyces formicae* KY5. This strain was isolated from the plant ant, *Tetraponera penzigi*[245,246], which builds nests in specialised structures, called domatia, of the thorny acacia (*Acacia drepanolobium*), and protect the plant from herbivores[247,248]. The ants farm a *Chaetomium* fungus which grows in the domatia and is their food source; they harvest leaves to feed the fungus and in return, the fungus protects and feeds them. Several *Actinomycetes* can be found in the domatia and living on the ants, including *S. formicae* KY5. The role of the bacteria and their natural products are still not fully understood. Three classes of compounds are produced by a single BGC in *S. formicae* KY5, the formicapyridines, formicamycins and fasamycins. Furthermore, multiple congeners of each molecule are produced[249-251], with differing halogenation and *O*-methylation patterns and also differing bioactivity profiles. In this strain, the major product of the BGC is formicamycin J (see Figure 5.1), with the fasamycins being intermediates produced in small amounts and the formicapyridines being shunt products from the premature cleavage of the acyl carrier protein (ACP) produced in trace amounts in the wild type strain[249].

This chapter aims use knowledge from the study of the mechanisms of resistance to borrelidin and obafluorin, by BorO and ObaO, in order to identify BGCs which might encode the production of new ThrRS inhibitors via a self-resistance led genome mining approach. In so doing, the possible evolution of these proteins was also briefly explored. First, the possible evolutionary history of BorO was explored, with homologue identification being the first step. This was then used as the starting point for genome mining efforts, with a handful of BGCs containing BorO homologues having been identified. One particular BGC of interest was identified in the genome of a desert derived *Micromonospora* strain was identified by the existence of extra, uncharacterised ThrRS paralogues. A similar exercise revealed far less diversity in the genomic context of ObaO homologs. Finally, a global phylogenetic tree of ThrRSs informs us that borrelidin resistance is likely widespread in bacteria, most frequently due to the presence of amino acids equivalent to the L489M mutation in EcThrRS which was discussed at length in Chapter 4. Conversely, if our model for obafluorin self-resistance is correct, obafluorin resistance appears to be found solely amongst those proteins closely related to ObaO.

5.2 Results and Discussion

5.2.1 BorO homologues are found in Actinomycetes in the absence of a borrelidin BGC

With work having been done on BorO's mechanism of resistance, as outlined in Chapter 3, we were interested to explore the potential evolutionary history of BorO and in doing so, attempt to identify potential novel BGCs which produce ThrRS inhibitors. A BLAST search was used to identify putative BorO homologs; these were identified by a percentage identity of greater than 50% and the presence of a threonine residue in the equivalent position to EcThrRS L489, the essential resistance determinant as determined in Chapter 3. Surprisingly, more than half of the BorO homologues which we identified came from outside of borrelidin BGCs (Table 5.1). Strains containing BorO homologs have been sampled from all over the world, as have borrelidin-producing strains which have not had their genomes sequences, as listed in Table 5.2.

The genome sequences (where available) for all the strains containing a BorO homologue were downloaded, the number of ThrRSs was checked manually, and the locations of potential BGCs identified using antiSMASH. It was then noted whether or not any BorO was found in a BGC, the number of ThrRSs in the strain, as well as the identity of the "resistance residues" in the housekeeping enzyme; this was discussed at length in Chapter 3 and 4, but the presence of a methionine in the equivalent to position 489 in EcThrRS can confer borrelidin resistance. Additionally, from the work on the borrelidin resistant housekeeping protein, SpThrRS, we suspect that a combination of glutamine in position 489, relative to EcThrRS and a tyrosine in position 461, relative to EcThrRS, can confer borrelidin resistance. The borrelidin sensitivity of the housekeeping enzymes was therefore estimated, as shown in Table 5.1.

As was shown for *S. parvulus*, a majority of the sequenced borrelidin BGC containing strains appear to have putative borrelidin resistant housekeeping ThrRS encoding genes. This could tell us that the borrelidin BGC has been acquired by these strains relatively recently, that the housekeeping protein has evolved resistance since acquiring the borrelidin BGC or that BorO has an additional/alternative role to self-resistance, in these strains. For most of the strains which have a copy of *borO* in their genome, but no borrelidin BGC, it appears that their housekeeping ThrRS is predicted to be borrelidin sensitive. This could tell us that *borO* has been acquired by these strains as a resistance gene against borrelidin. Strikingly, we have two examples of metagenome-assembled genomes (MAGs) assembled from the same sample for which multiple distinct genomes contain genes encoding BorO homologues. One of these comes from the rhizosphere of a *Barbacenia macrantha* plant, and the other from a dinosaur bone fossil. The genomic context of BorO homologs is different in different MAGs in both of these samples, suggesting that BorO is being disseminated within these communities.

Table 5.1. List of strains containing a *borO* homologue. % identity relative to the *S. parvulus* BorO is presented, context of the *borO* is judged based on the proximity to a BGC based on antiSMASH analysis, or a manual inspection for homologues in short contigs/on the edges of contigs- for these if tentative biosynthetic genes are present, they are marked with a question mark. The number of ThrRSs in the assembly are listed where possible (N/A where a genome sequence was not available), housekeeping (HK) primary and secondary resistance residues are listed, in which 489 refers to the position relative to EcThrRS which has been found to be key for borrelidin resistance. A threonine or methionine in this position have been proven to confer borrelidin resistance. If a glutamine is in this position, it is theorised that if there is also a tyrosine in position 461 relative to EcThrRS, the protein will be resistant to borrelidin, as detailed in section 3.2.14. Details for any extra ThrRS encoding genes in the genome are also listed in brackets- some of these appear truncated, missing the N-terminal editing domain. Isolation source is listed where available. Rows are shaded green where housekeeping genes are predicted to be resistant to borrelidin, orange where housekeeping genes are predicted to be sensitive, yellow where resistance cannot be predicted and unshaded where housekeeping proteins haven't been identified.

Strain	**% Identity**	**In BGC?**	**Borrelidin BGC?**	**# ThrRSs**	**HK resistance residue (489)**	**HK 2° resistance residue (461)**	**Isolation Source**
Streptomyces parvulus Tü4055	100	Y	Y	3	Q (Truncated HK, Q)	Y (Truncated HK, F)	Tübingen Strain Collection
Streptomyces rochei str. Sal35	100	Y	Y	N/A	N/A	N/A	Forest soil, Shennongjia, Eastern Hubei Province, China
Streptomyces sp. SGAir0924	99.85	Y	Y	3	Q (Truncated HK, Q)	Y (Truncated HK, F)	Outdoor air, Singapore
Streptomyces sp. CRB46	99.84	Y	Y	2	Q	Y	Shotgun, *Cyperus rotundus* rhizosphere, Cemoro Sewu highland, East Java, Indonesia
Streptomyces sp. WAC02707	99.7	Y	Y	2	Q	Y	Soil, Enugu, Nigeria

Streptomyces sp. AVP053U2	97.04	Y	Y	2	L	F	*Styela clava*, tunicate in Long Island Sound, USA
Streptomyces griseocarneus	88.58	Y	Y	2	Q	Y	*Cyperus rotundus* rhizosphere, Cemoro sewu Dieng plateau, Indonesia
Streptomyces netropsis JCM 4063	88.28	Y	Y	3	L (Truncated HK, Q)	F (Truncated HK, F)	Soil, Hudson, New York, USA
Pseudonocardiales bacterium CP_BM_ER_28	81.62	N	N	1	N/A	N/A	Rock nearby *Barbacenia macrantha*-rhizosphere, Minas Gerais, Brazil
Pseudonocardiales bacterium CP_BE_RX_50	81.23	Y	N	1	N/A	N/A	Rhizosphere of *Vellozia epidendrioides*, Minas Gerais, Brazil
Pseudonocardiales bacterium CP_BM_ER_R9_29	81.16	Y	N	1	N/A	N/A	Rock nearby *Barbacenia macrantha*-rhizosphere, Minas Gerais, Brazil
Frankia sp. AiPs1	77.16	?	N	2	L	F	*Alnus incana* root nodule, Karttula, Finland
Frankia sp. CiP3	76.37	N	N	1	N/A	N/A	*Coriaria intermedia* nodules, Poblacion, Atok, Philippines
Frankia sp. Cj5	76.23	Y	N	1	N/A	N/A	*Coriaria japonica* nodules, Japan
Frankia sp. ACN1ag	75.93	N	N	1	N/A	N/A	Root nodules of *Alnus viridis crispa*, Atikokan, Ontario, Canada

Frankia torreyi str. Cpl1	75.78	N	N	3	L (L)	F (F)	Nodules from *Comptonia peregrina* Petersham, Massachusetts, USA
Frankia sp. CIT1	75.78	N	N	1	N/A	N/A	*Coriaria intermedia* nodules, Taiping Mountain, Taiwan
Pseudonocardiales bacterium CP_BM_ER_R8_32	75.78	N	N	2	L	F	Rock nearby *Barbacenia macrantha*-rhizosphere, Minas Gerais, Brazil
Actinobacteria bacterium OV320	75.71	Y	N	3	L (Truncated HK, Q)	F (Truncated HK, F)	Endosphere of plant *Populus trichocarpa*, Corvallis, Oregon
Frankia alni str. ACN14A	75.63	N	N	2	L	F	Root nodules of *Alnus viridis* ssp. *Crispa*, Tadoussac, Quebec, Canada
Frankia sp. QA3	75.63	N	N	2	L	F	Root nodule of *Alnus nitida,* Bahrin, District Swat, Pakistan
Nocardia donostiensis str. X1655	75.6	N	N	2	E	F	Human Bronchial sputum, San Sebastian, Spain
Frankia sp. CpI1-P	75.33	N	N	1	N/A	N/A	Nodules from *Comptonia peregrina* Petersham, Massachusetts, USA
Pseudonocardiaceae bacterium isolate Dino_bin31	74.29	N	N	2	M	F	Metagenome, centrosaurus fossil bone, Late Cretaceous North America, Alberta, Canada
Streptomyces sp. CT34	73.58	N	N	2	Q	Y	Ghanaian Soil Sample

Nitriliruptorales bacterium isolate Dino_bin24	73.11	?	N	2	L	F	Metagenome, centrosaurus fossil bone, Late Cretaceous North America, Alberta, Canada
Microthrixaceae bacterium isolate SSF14 UP4B101214	68.91	?	N	1	N/A	N/A	Metagenome assembly, wastewater treatment plant, Singapore
Tetrasphaera sp. isolate Aved_18-Q3-R54-62_MAXAC.378	58.96	Y	N	1	N/A	N/A	Metagenome, activated sludge, Avedore, Denmark
Tetrasphaera jenkinsii Ben 74 (DSM 17519)	57.7	N	N	1	N/A	N/A	Activated Sludge, Glenelg, S.A., Australia

Table 5.2. List of borrelidin producers with no deposited genome sequence, strain name, isolation location and reference listed.

Strain	Isolation Location	Reference
Streptomyces sp. RL09-241-NTF-B	Marine sediment, Cayucos, California, USA	Schulze *et al.* 2014[214]
Streptomyces sp. neau-D50	*Glycine max (*Soybean Root), Harbin, Heilongjiang province, China	Liu *et al*. 2012[155]
Streptomyces albovinaceous	Field soil sample, Lewiston Idaho	Singh *et al.* 1985[252]
Streptomyces str. MS-6-6	Western Saudi Arabia	Yassien *et al.* 2015[253]
Streptomyces heilongjiangensis	*Glycine max (*Soybean Root), Hulin, Heilongjiang province, China	Liu *et al.* 2013[254]
Streptomyces californicus	Soil, Gundlasingaram village, Warangal, India	Srinivasan *et al.* 2008[220]
Streptomyces C2989	Soil, Caravan site, Mablethorpe, Lincolnshire, UK	Lumb *et al.* 1965[156]
Streptomyces sp. OM-0060	Soil, Japan	Ishiyama *et al.* 2011[158]
Streptomyces coelicoflavus	Soil, Egypt	Hassan *et al.* 2016[255]
Streptomyces rochei MB037	Marine sponge *Dysidea arenaria*, South China Sea	Li *et al.* 2018[219]
Streptomyces sp. GK18	Potato lesions, Iran	Cao *et al.* 2012[159]
Nocardiopsis sp. HYJ128	Saltern topsoil, Jeung-do Island, Shinan-gun, Jeollanamdo, Korea	Kim *et al.* 2017[217]
Streptomyces rochei SCSIO Zj89	Mangrove-derived sediment sample, Yalongwan, China	Sun *et al.* 2018[218]

A phylogenetic tree of BorO homologues and the relevant housekeeping ThrRSs found in the genomes of BorO-containing strains was inferred using maximum likelihood, as seen in Figure 5.2. Because some of the housekeeping proteins were truncated to remove the editing domain, the alignment was trimmed to remove the editing domain for all proteins, along with other regions with poor alignment. This tree showed that BorO homologues clade distinctly to the housekeeping proteins, and that BorO homologues found in borrelidin BGCs clade distinctly to the other BorOs, but those BorOs from strains with a borrelidin-resistant housekeeping protein do not strictly clade together. This would suggest that there has been recent horizontal gene transfer of the whole borrelidin BGC between these strains; this could be from a strain which is naturally sensitive to

borrelidin to one naturally resistant, allowing the accepting strain to survive in the same environment as a borrelidin producer, and then being able to uptake and integrate the BGC. We cannot, however *ab initio,* guess the direction of horizontal gene transfer.

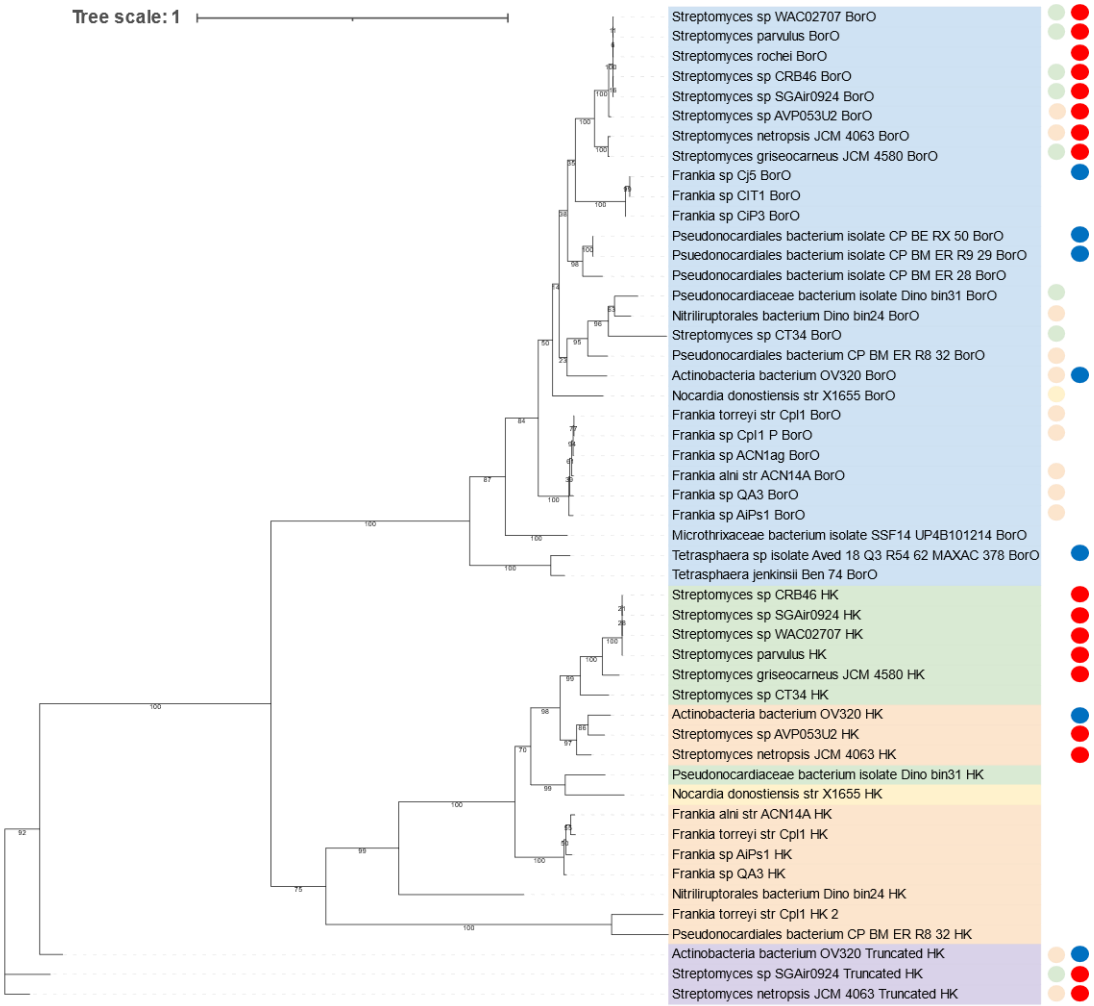


Figure 5.2. Maximum likelihood phylogenetic tree shows that BorO homologues are phylogenetically distinct to their housekeeping proteins. Branch lengths to scale, with bootstrap values on each branch. BorO homologues are coloured in blue, predicted resistant housekeeping ThrRSs coloured in green, predicted sensitive housekeeping ThrRSs coloured in orange, ThrRS for which we cannot predict resistance is shown in yellow and truncated ThrRS proteins should in purple. Circles in the first column to the right indicate the status of the relevant housekeeping protein, where it can be identified. In the second column, a red dot denotes the BorO being in a borrelidin BGC, blue dot denotes the BorO being in or near a non-borrelidin BGC. Figure generated in the interactive tree of life (iTOL)[205].

Four of the strains had what appears to be a truncated housekeeping protein with no N-terminal editing domain as the third *ThrRS* in the genome (the first being the housekeeping *ThrRS* and the second being the *borO*). The truncated housekeeping proteins appear to be derived from a separate group than these *Actinomycete* sequences, based on their positioning in the phylogenetic tree,

Figure 5.2. A BLAST search reveals the presence of these truncated ThrRSs in many *Streptomyces* genomes, but their function is not immediately obvious. antiSMASH analysis of the regions containing the top 20 BLAST hits of these homologues did not identify any being located within BGCs. A further exploration of these proteins would therefore be interesting but is beyond the scope of this thesis. From work done in Chapter 4, we know that ThrRSs generally need their editing domains to maintain fidelity so it would be especially interesting if these proteins can maintain fidelity without an editing domain.

Other than the known borrelidin BGCs, a few BorO homologues have been found in or near to other BGCs; these were identified in *Actinobacterium sp.* OV320, *Tetrasphaera sp.* Aved18, two of the *Pseudonocardia* strains from the *Barbacenia* rhizosphere sample, and *Frankia sp.* Cj5. The BorO homologues for these strains did not clade closely together, suggesting that the acquisition of these BorOs was at different points in evolution. These BGCs are also unrelated to each other. A housekeeping gene could only be identified in the *Actinobacterium sp.* OV320 strain, possibly due to the poor quality of the genome sequencing of the other strains and/or them being MAG derived.

5.2.1.1 A potentially novel BGC containing a BorO homologue in Actinobacterium sp. OV320

One of the strains in which a BorO homolog appears to be located in a non-borrelidin BGC is *Actinobacterium sp.* OV320. The housekeeping ThrRS in this strain is predicted to be sensitive to borrelidin. This would suggest to us that in this strain, the putative BorO homologue has been acquired as a resistance protein against borrelidin or for self-resistance to a compound produced by the strain. The possible biosynthetic genes for this BGC can be seen listed in Supplemental Table 7, and the BGC in Figure 5.3. While there were some other genomes with similar genes to genes 5-8, 14, 15 and 30-39 clustered together as identified in a multigene BLAST, the NRPSs are absent, suggesting that this BGC has inserted into this region in this genome. There are no close homologs to this BGC, suggesting that the product of the BGC may be novel.

The main biosynthetic machinery of this BGC appears to be a three module NRPS (encoded by genes 21 and 22), an extra putative adenylation domain (encoded by gene 18) and an extra putative PCP (encoded by gene 19).

There are multiple possible self-resistance genes in the BGC; genes 20 and 29 both encode putative major facilitator superfamily (MFS) transporters which are frequently involved in natural product export in natural product producers[256], preventing accumulation of the compound in the cell and therefore preventing toxicity, as discussed in Chapter 1. Additionally, the gene encoding the BorO homologue is located immediately following the main biosynthetic genes, and if it has not been inserted into the region by chance, could be indicative of a ThrRS targeting natural product. While we can't predict from the DNA sequence alone what the BGC is capable of producing, it would be interesting to obtain this strain in order to discover a new ThrRS inhibitor.

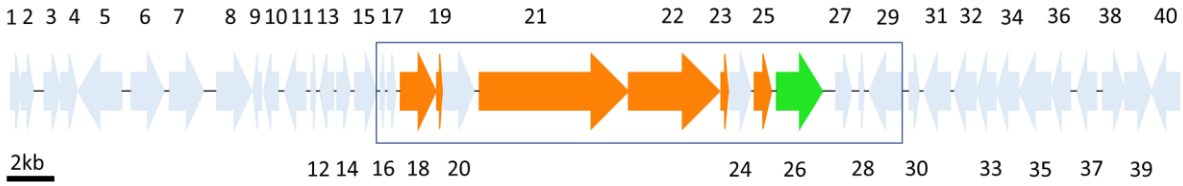


Figure 5.3. A novel BGC containing a *borO* homologue in *Actinobacterium* sp. OV320. Genes are numbered above/below the arrows. The BorO homolog is coloured in lime green, the NRPS-associated genes are coloured in orange and all other genes coloured in light blue. Blue box shows best guess at the actual BGC in this genome.

5.2.1.2 A BGC identified by metagenome assembly from a Tetrasphaera strain potentially containing a BorO homologue

A MAG which appears to contain a hybrid PKS/NRPS encoding BGC with a *borO* homologue was assembled from a sample derived from activated sludge- the material used for biological wastewater treatment in sewage works[257]. As far as I know, there has been no exploration of the antibiotic production potential of *Tetrasphaera* strains; work to date has mainly focussed on their phosphate and ammonia compound metabolism. This strain is not particularly talented, with only four BGCs identified in its genome by antiSMASH analysis (Figure 5.4); it is important, however, to point out that this is a low-quality MAG because antiSMASH does not perform well with short contigs. All four BGCs are novel, with only two papers addressing this ecosystem and assessing MAGs like this one[258,259]. Of these BGCs, our one of interest is region 1.1. The possible biosynthetic genes for this BGC can be seen listed in Supplemental Table 8, and the BGC in Figure 5.5.

Region	Type	From	To	Most similar known cluster		Similarity
Region 1.1	NRPS-like	192,214	246,771	isoindolinomycin	Polyketide	7%
Region 11.1	RRE-containing	106,837	127,130	glycopeptidolipid	NRP	5%
Region 24.1	hglE-KS	202,343	244,301	macrotetrolide	Polyketide	33%
Region 29.1	RiPP-like	138,560	149,414	7-deoxypactamycin	Polyketide:Iterative type I + Saccharide:Hybrid/tailoring	3%

Figure 5.4. antiSMASH report for *Tetrasphaera* sp. Aved_18-Q3-R54-62_MAXAC.378.

The main biosynthetic machinery of this BGC appear to be NRPS-related proteins (encoded by genes 22, 32-36) and putative PKS-related proteins (encoded by genes 9, 24-31). Genes 13-16 encode the proteins required to form an acetyl-CoA carboxylase, the biotin-dependent enzyme required to convert acetyl-CoA to malonyl-CoA and so are likely involved in precursor supply for the PKS.

To attempt to ascertain the borders of the BGC, the antiSMASH multigene blast was analysed; in other strains which were identified, gene 1 and 42-52 are all clustered together. This would suggest that at some point in the past, genes 2-41 were inserted into this genome. The ABC transporter system, the putative NRPS encoded by gene 22 and variations of the PKS genes appear to be co-located in a *Bacillus* strain and a *Paenibacillus* strain, putative malonyl-CoA synthesising enzymes from this BGC have homologs in the BGCs for isoindolinomycin, xantholipin and others.

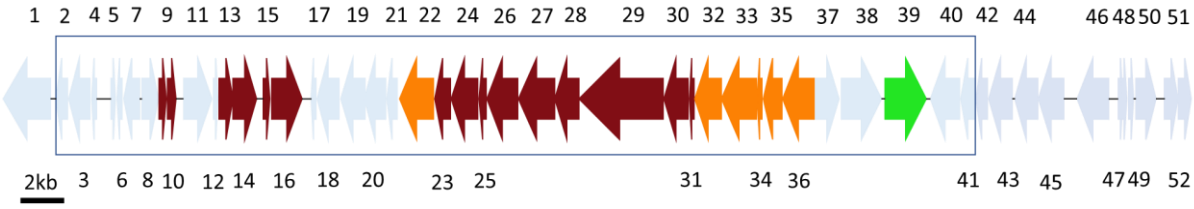


Figure 5.5. A novel BGC containing a borO homologue in *Tetrasphaera* sp. Aved_18-Q3-R54-62_MAXAC.378. Genes are numbered above/below the arrows. The BorO homolog is coloured in lime green, the NRPS-associated genes are coloured in orange, PKS-associated genes in dark red and all other genes coloured in light blue. Blue box shows best guess at the actual BGC in this genome.

There are two possible self-resistance mechanisms encoded by this BGC. Genes 18-20 encode an ABC transporter which likely exports the hypothetical product of this BGC and the BorO homologue is encoded by gene 39. The BorO could have nothing to do with this cluster or could indicate that the product of this BGC targets ThrRS. Because this is a MAG, if it were to be explored further, then the BGC would need to be synthesised and heterologously expressed. A problem with this is that Illumina sequencing was used to sequence this cluster. Short-read Illumina sequencing often struggles with sequencing PKS and NRPS genes due to their repetitive nature, and so can artificially truncate them in sequencing- this can be seen for many of the borrelidin BGCs available on GenBank. It would therefore be unadvisable to synthesise the BGC as it is, without being able to verify the size of the PKS/NRPS units via PCR, directly from gDNA. Therefore, while this cluster appears very intriguing and may include a ThrRS, it is a poor candidate for future work. If a similar BGC were to be identified in a cultured bacteria with a ThrRS still in it in future, however, this would be fascinating to explore further.

5.2.1.3 A novel lanthipeptide BGC from a plant rhizosphere MAG with a nearby BorO homologue

AntiSMASH revealed a single *borO* containing BGC in two metagenomes from the same sampling project, obtained from a rock near to the plant *Barbacenia macrantha*. This RiPP BGC was predicted to produce a lanthipeptide. The possible biosynthetic genes for this BGC can be seen listed in Supplemental Table 9 and the BGC in Figure 5.6. Lanthipeptides are a class of RiPPs which contain lanthionine bonds. In order to form these bonds, a dehydration reaction of serine or threonine residues must first occur, producing dehydroalanine (Dha) or dehydrobutyrine (Dhb). A cysteine residue elsewhere in the peptide can then attack the dehydroalanine, which can then either protonate to form a lanthionine bond, or go on for a second conjugate addition to another Dha/Dhb in order to then form a labionin bond between two labionin residues[260].

This cluster is predicted to be similar to that for catenulipeptin[261], SapB and labyrinthopeptin A1/2/3, being made up of just four genes; two transporters, a precursor peptide and a biosynthetic gene for lanthionine bond formation and cyclisation. It is therefore unlikely that the *borO*

homologue in this BGC is a self-resistance mechanism for the product of this BGC. SapB is involved in aerial hyphae formation in *Streptomyces* spp. and catenulipeptin appears to act as a surfactant. Labyrinthopeptin A1 has been demonstrated to have antiviral activity, with human immunodeficiency virus (HIV), human respiratory syncytial virus (hRSV) and herpes simplex virus (HSV) inhibition being reported[262,263], with Labyrinthopeptin inducing virolysis via binding to the membrane lipid phosphatidylethanolamine[264]. Their low cytotoxicity and unusual targeting mechanism means that they have garnered interest as development for future antiviral drug development.

Because the BGC was found in a MAG, the specific strain has not been cultured and so further work would require synthesis of the BGC and then expression heterologously. Additionally, it is unlikely that the BorO homologue is involved in self-resistance so this BGC should not be explored further for this work.

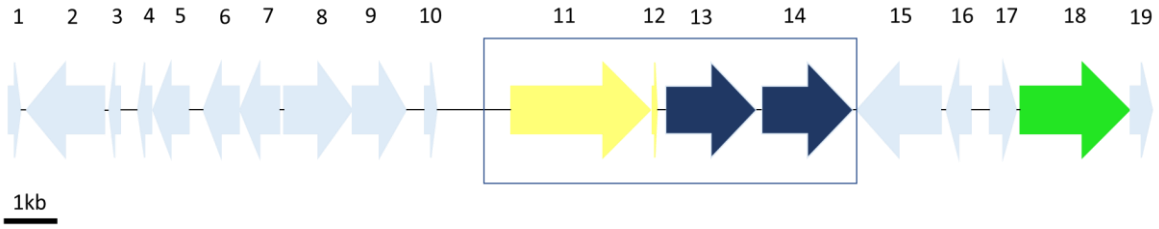


Figure 5.6. A novel BGC containing a borO homologue in *Pseudonocardiaceae*. Genes are numbered above the arrows. The BorO homolog is coloured in lime green, the lanthipeptide synthesis genes are coloured yellow and the transport genes are coloured dark blue. All other genes coloured light blue. Blue box shows best guess at the actual BGC in this genome.

5.2.1.4 A novel lanthipeptide BGC from a Frankia with a nearby BorO homologue

This BGC was identified in the genome of a *Frankia* strain which was first isolated from the root nodules of a *Coriaria japonica* plant in Japan, and encodes a lanthipeptide type RiPP BGC. The possible biosynthetic genes in this BGC can be seen listed in Supplemental Table 10, and the BGC in Figure 5.7. Unlike the earlier discussed *Pseudonocardia* BGC, there were no known BGCs identified as similar to this one. It appeared that genes 1-5 could be found in some of the other BorO containing *Frankia* strains in this analysis. The contigs containing the BorO homologues in these strains were aligned. For these strains, the lanthipeptide part of the BGC was in an assembly gap, suggesting that it may also be present in these strains, but could not be assembled. Based on homology to uncharacterised BGCs in *Streptomyces* strains, genes 6-14 are likely to be the legitimate members of the BGC.

There are two possible self-resistance genes in this BGC. Gene 8 encodes a putative GNAT family *N*-acetyltransferase. In Microcin C7 biosynthesis, the resistance gene, *mccE* is a homologue of this gene, and acts by acetylating the N-terminus of processed microcin C7, preventing AspRS inhibition.

The resulting protein could therefore either be a tailoring enzyme responsible for modifying the mature peptide NP, or it may be a self-resistance determinant which acts by acetylating and deactivating the product of the BGC. Gene 2 encodes the BorO homologue and could be a self-resistance gene if the lanthipeptide targets ThrRS, but may be in this genomic context by chance, if acting as an acquired borrelidin resistance gene.

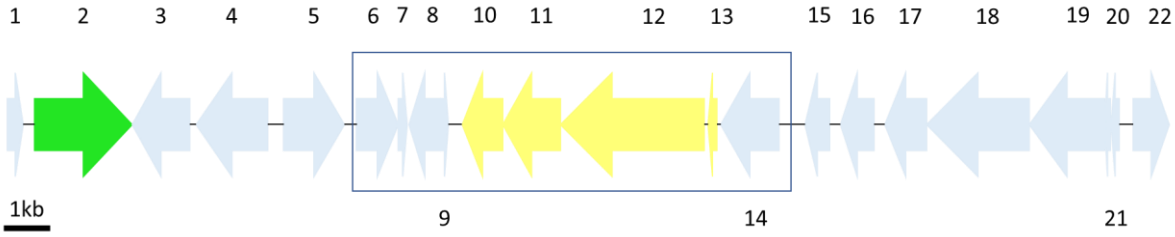


Figure 5.7. A novel BGC containing a borO homologue in *Frankia* sp. Cj5. Genes are numbered above/below the arrows. The BorO homolog is coloured in lime green and the lanthipeptide synthesis genes are coloured yellow. All other genes coloured light blue. Blue box shows best guess at the actual BGC in this genome.

In summary, we have identified four BGCs which contain genes encoding BorO homologues. However it is not clear for any of them that the BorO homologue is a legitimate self-resistance gene, instead it is likely to have been inserted in these genomic regions by chance. Of these, the BGC from *Actinobacterium* sp. OV320 seems to be the most promising lead for future work. Having followed this branch of enquiry, the BorO-containing homologue containing genomes were searched for other additional ThrRSs.

5.2.2 A potentially talented strain of Micromonospora strain from an underexplored ecological niche

When examining the genome of *Frankia torreyi*, a third ThrRS encoding gene was identified. The product of this gene clades more closely to housekeeping proteins than truncated housekeeping proteins or BorOs (see Figure 5.2) and appears similar to the putative housekeeping protein from one of the *Pseudonocardiales* strains. Interestingly, this third ThrRS homologue only gave one close BLAST hit at the time of searching (March 2020); this was from *Micromonospora* sp. KC207.

Micromonospora sp. KC207was isolated from desert soil at Darvaza in the Karakum Desert, Turkmenistan as part of the PhD thesis of Dr Hayrettin Saygın during a bacterial ecology project by the lab of Prof Nevzat Şahīn (Ondokuz Mayis University, Turkey). The genome sequence was deposited in GenBank, but no papers about this strain have been published to date. AntiSMASH analysis revealed the proximity of an NRPS-containing BGC and flanked by transposases; however, the annotated BGC appeared to be on the edge of the contig. We were therefore gifted the strain by Prof Şahīn and sequenced the genome ourselves.

5.2.2.1 Genome sequencing revealed a circular genome containing many novel BGCs

The strain was sent to MicrobesNG (Birmingham, UK) where the gDNA was extracted and sequenced using both Illumina and Nanopore sequencing. Dr Govind Chandra (JIC) then assembled the whole genome into a single contig. The genome was found to be circular of length 7,373,107 bp, (Figure 5.8). The G+C content for the genome was found to be 69.2%. The genome contains a CRISPR system and has regions of the genome with low G+C content, indicative of incorporation of genomic regions from other organisms such as other bacteria and bacteriophages.

AntiSMASH analysis indicates at least 28 BGCs (Figure 5.9), with many of the annotated BGCs appearing to be 'islands' comprising multiple different BGCs in close proximity. Our BGC of interest was located in region 9. Of the regions identified by antiSMASH, two appear to contain BGCs which are almost identical to already identified BGCs; region 10 appears to contain the rosamicin BGC, which produces rosamicin, salinipyrone A and pacificanone A[265], while region 14 appears to contain the BGCs for ketomemicin B3 and chloramphenicol but their functionality has not been demonstrated. These exact compounds may not be produced but the resultant natural products are likely to be chemically similar. Most of the remaining BGCs found in this genome have only low similarity to known BGCs, and so future exploration of these BGCs in this talented strain of *Micromonospora* could be an excellent way to discover new natural products. It has previously been noted that *Streptomyces* strains from desert environments contain a lot of novel chemistry, having diverged from non-desert strains[243]. It is likely that the same is true for desert-derived *Micromonospora* strains like this.

5.2.2.2 A BGC producing a potential novel ThrRS targeting compound can be identified

The BGC containing the homologue to the third *Frankia torreyi* ThrRS is found in Region 9 (Figure 5.10). For clarity, I will refer to the compound produced by this cluster as "Compound 1" and the gene*s A-O*. This BGC is flanked by a series of transposases. This would suggest that the BGC is in a transposon, a form of mobile genetic element. Transposons are widely found in both eukaryotic and prokaryotic genomes, and are thought to be one of the major driving forces in evolution. The region of the genome containing the BGC has a marked decrease in G+C content when compared to the rest of the genome and has many transposases. This suggests that the BGC originally entered the genome on a transposon from a different organism. In the fungus *Botrytis cinerea*, the botcinic acid BGC has alternating G+C balanced regions and A-T rich regions produced by transposable elements, which is evidence of transposon activity[266].

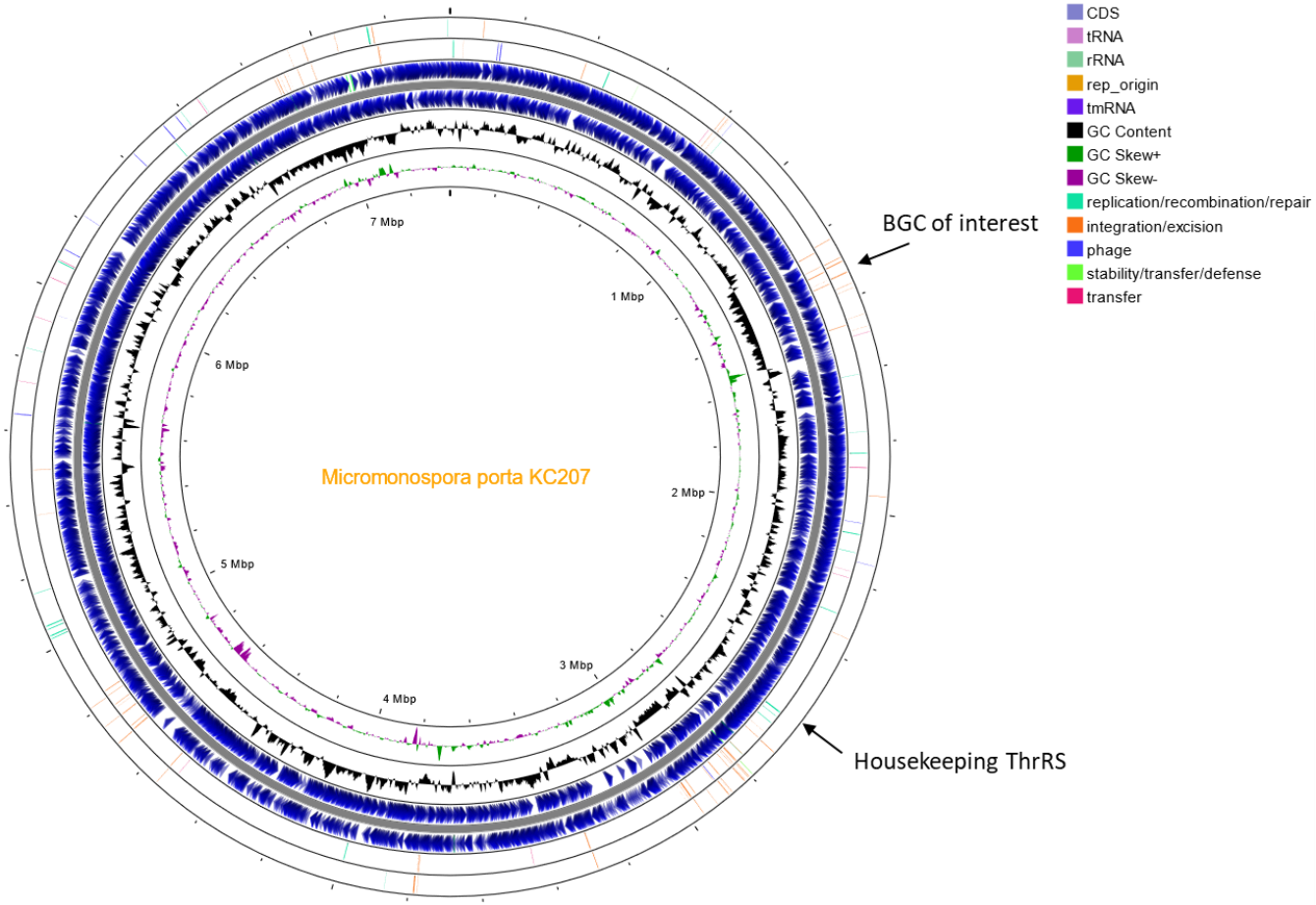


Figure 5.8. Visualisation of the *Micromonospora*. sp. KC207 genome. Outer two circles show features derived from mobile genetic images as coloured per the key, Annotated genes are then shown in dark blue, on the middle two circles. The next inner circle shows the G+C content at each position, then the GC skew is shown with a positive GC skew shown in green and negative GC skew in purple. Positions of the BGC of interest and the housekeeping ThrRS are marked. Figure generated in Proksee.

Region	Type	From	To	Most similar known cluster		Similarity
Region 1	terpene	206,337	226,912	carotenoid	Terpene	27%
Region 2	linaridin	244,222	264,764			
Region 3	thioamide-NRP	265,699	305,983	netropsin	NRP	22%
Region 4	terpene	342,182	362,391	phosphonoglycans	Saccharide	3%
Region 5	T1PKS	709,813	760,953	butyrolactol A	Polyketide	40%
Region 6	NRPS	873,373	918,104			
Region 7	T1PKS,PKS-like,oligosaccharide,other	1,019,561	1,087,023	avilamycin A / avilamycin C	Saccharide:Oligosaccharide	50%
Region 8	NAGGN	1,145,935	1,160,687			
Region 9	NRPS-like,NRPS	1,291,971	1,336,149	s56-p1	NRP	5%
Region 10	T1PKS	1,497,572	1,574,961	rosamicin / salinipyrone A / pacificanone A	Polyketide	86%
Region 11	NRPS	1,656,145	1,714,504	atratumycin	NRP	10%
Region 12	thiopeptide	2,825,166	2,868,512	nosiheptide	RiPP:Thiopeptide	23%
Region 13	T1PKS,phosphonate,transAT-PKS	2,965,268	3,230,921	macrotermycins	Polyketide	76%
Region 14	guanidinotides,NRPS-like	3,283,464	3,336,973	chloramphenicol	NRP	94%
Region 15	NRPS-like,NRPS,T1PKS	3,372,741	3,439,252	bleomycin	NRP:Glycopeptide + Polyketide:Modular type I + Saccharide:Hybrid/tailoring	9%
Region 16	terpene	3,510,263	3,530,399	2-methylisoborneol	Terpene	75%
Region 17	terpene,T1PKS	3,855,225	3,915,269	sporolide A / sporolide B	NRP + Polyketide:Enediyne type I	21%
Region 18	NRPS-like,NRPS,T1PKS	3,928,457	3,994,824	thiocoraline	NRP:Cyclic depsipeptide	5%
Region 19	T2PKS	4,053,307	4,125,876	formicamycins A-M	Polyketide	18%
Region 20	siderophore,lanthipeptide-class-iii	4,168,102	4,198,928	catenulipeptin	RiPP:Lanthipeptide	60%
Region 21	T1PKS	4,240,493	4,289,629	s56-p1	NRP	3%
Region 22	RiPP-like	4,396,173	4,405,818			
Region 23	NRPS,T1PKS,NRPS-like,other,T3PKS	4,581,369	4,711,574	kendomycin	Polyketide:Modular type I	55%
Region 24	other	4,716,338	4,757,156			
Region 25	RiPP-like	4,969,799	4,979,382	lymphostin / neolymphostinol B / lymphostinol / neolymphostin b	Polyketide + NRP	30%
Region 26	T3PKS	6,505,446	6,546,495	alkyl-O-dihydrogeranyl-methoxyhydroquinones	Terpene + Polyketide	57%
Region 27	RiPP-like	6,818,790	6,830,748			
Region 28	NRPS	6,884,459	6,927,671	A54145	NRP	8%

Figure 5.9. antiSMASH report for *Micromonospora* sp. KC207 genome. Each genome region is labelled based on BGC type and location in the genome. The most similar characterised BGC is then identified, along with a score for similarity to this BGC is then shown. The BGC of interest is region 9.

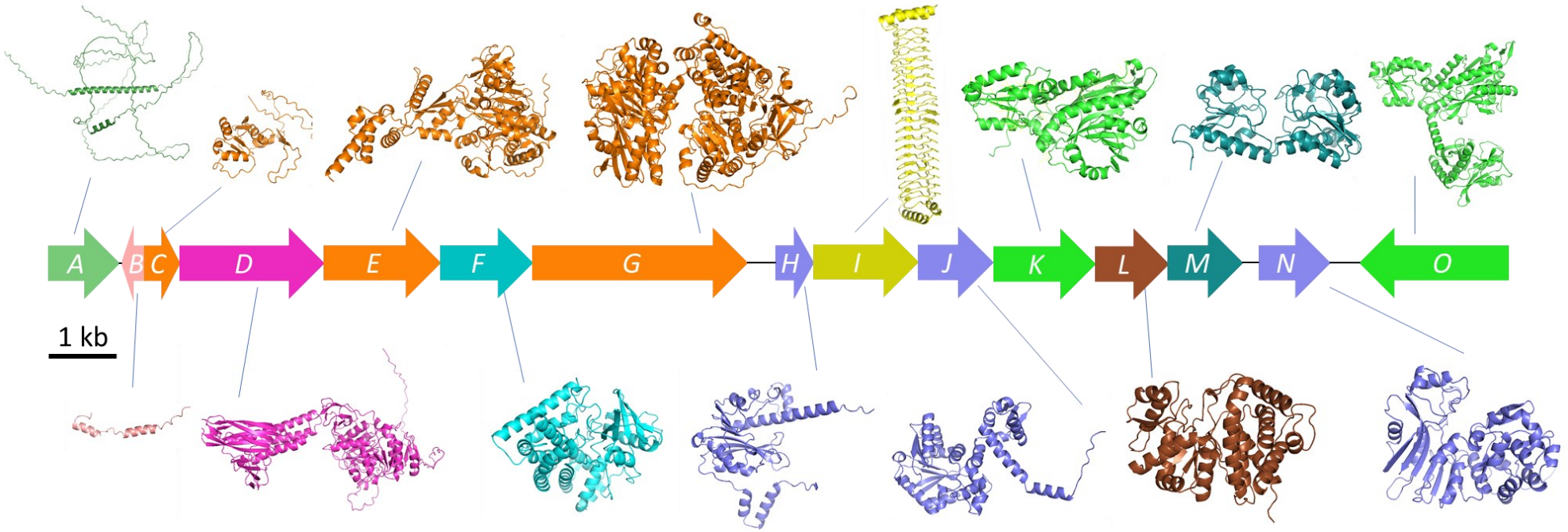


Figure 5.10. Putative compound 1 BGC. Genes are Labelled on the arrows. Each protein sequence was used to generate AlphaFold2 models. Protein structures are visualised in PyMol and gene arrows are coloured to match putative protein functions. Genes A and B are hypothetical proteins, coloured pale green and pale pink respectively. Regulatory genes are coloured hot pink, the oxidoreductase is coloured teal, the cytochrome P450 is coloured cyan, the ThrRS paralogues are coloured lime green, the methyltransferases are coloured lavender, the NRPS-associated genes are coloured orange, the glycosyltransferase is coloured brown and the pentapeptide repeat protein (PRP) is coloured dark yellow.

Overall, the BGC has no close homologues in other strains or much similarity to any known BGC (Figure 5.11) and BLAST of individual genes in the BGC (most of which are initially annotated as hypothetical) reveals only low identity hits which are assigned a putative function in Table 5.3. For this reason, the assignment of the functions for these genes can not necessarily be relied on. This also means, however, that this BGC, regardless of its product possibly contains some novel chemistry and biology. The lack of homologous BGCs found in other strains may be because the strain was isolated from desert soil which is generally under sampled when compared to temperate soils. With more samples becoming available from more diverse ecosystems, over time we may be able to get more information about homologous BGCs.

The main biosynthetic machinery encoded in this BGC appear to be NRPS proteins. The product of gene C looks similar to an NRPS in the AlphaFold structure, with a PCP domain and a partial adenylation domain, and due to this it may be non-functional. If the protein does have a function, it would be interesting to discover it. The adenylation domain of the NRPS encoded by gene *E* is predicted to be selective for leucine and the adenylation domain of the NRPS encoded by gene *G* is predicted to be selective for valine and the condensation domain in the NRPS encoded by gene *G* is identified as being LCL type- catalysing the condensation of two L-amino acids.

Gene *F* encodes a putative cytochrome P450. These proteins generally oxidise their substrates using a haem as a cofactor. In biosynthesis, they are extremely powerful, being able to functionalise the usually chemically inert C-H bond. They can catalyse many different reactions, including hydroxylation, ether formation, desaturation, epoxidation, dealkylation, oxidation, oxidative rearrangement, biaryl ring coupling, thiolation and nitration[267]. It is therefore difficult to predict the function of this protein.

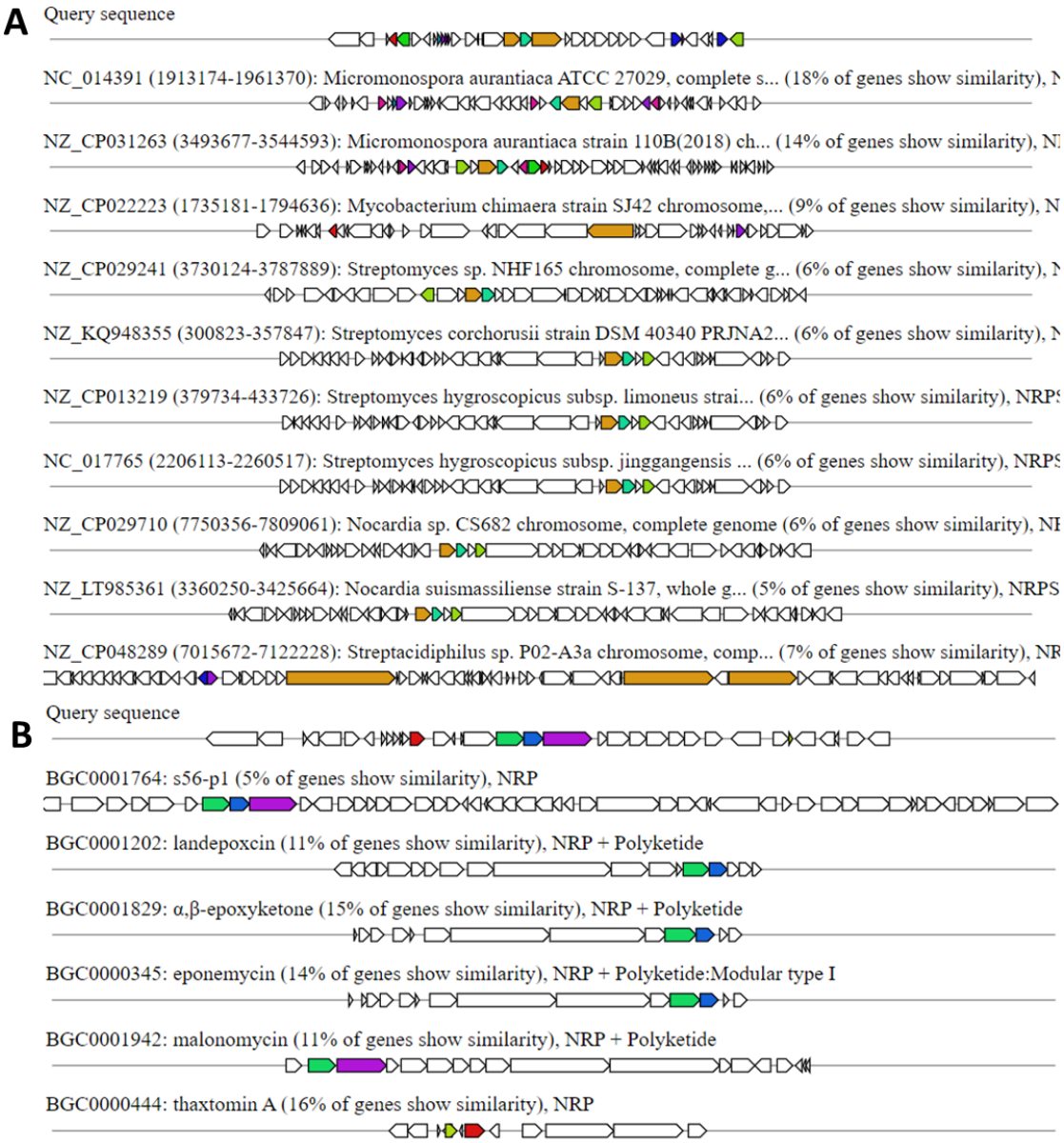


Figure 5.11. antiSMASH output comparing the compound 1 BGC to other genomes. A, multigene blast comparing to similar regions in other genomes. **B,** known cluster BLAST comparing the **1** BGC to other known BGCs.

Table 5.3, Table of BLAST hits and putative functions for the BGC from *Micromonospora* sp. KC207

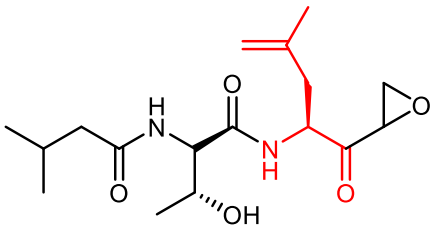
Gene	Closest non-hypothetical BLAST hit	Query Cover (%)	Percentage Identity (%)	Putative Function
Gene *A*	Helix-turn-helix domain containing protein, *Micromonospora humidisoli*	8	75.76	Unknown
Gene *B*	None	-	-	Unknown
Gene *C*	NRPS, *Streptomyces fulvoviolaceus*	54	55.05	Peptide synthesis, Unknown
Gene *D*	TFIIB-type zinc ribbon-containing protein, *Kitasatospora*	42	35.90	Regulation, Unknown
Gene *E*	NRPS, *Micromonospora* sp. STR1s_6	96	90.52	Peptide synthesis
Gene *F*	Cytochrome P450, *Micromonospora* sp. STR1s_6	100	94.32	Peptide Modification
Gene *G*	NRPS, *Micromonospora chokoriensis*	96	57.60	Peptide synthesis
Gene *H*	Class I SAM-dependent methyltransferase, *Candidatus Scalindua* sp.	83	42.77	Methylation
Gene *I*	Pentapeptide repeat-containing protein, *Saccharothrix ecbatanensis*	100	55.39	Self-resistance
Gene *J*	Methyltransferase domain-containing protein, *Bradyrhizobium ottawaense*	99	50.29	Methylation
Gene *K*	Class IIA aaRS C-terminal domain containing protein, *Streptomyces* sp. RB6PN25	98	43.04	Addition of threonine/serine to compound

Gene *L*	Glycosyltransferase, *Actinomadura chibensis*	95	43.19	Glycosylation
Gene *M*	2-hydroxyacid dehydrogenase, *Proteobacterium* bacterium	95	39.94	Reduction/Dehydrogenation
Gene *N*	SAM-dependent methyltransferase, *Streptosporangium saharense*	98	35.54	Methylation
Gene *O*	ThrRS, *Frankia torreyi*	100	68.05	Self-resistance

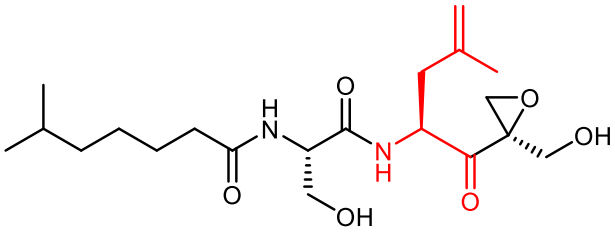
Strikingly, homologous proteins to those encoded by genes *E, F* and/or *G* are also found in other BGCs; s56-p1, landepoxcin, eponemycin and malonomycin. Homologous proteins to those encoded by genes *E, F and G* can be found in s56-p1 biosynthesis however the functions of these genes in s56-p1 biosynthesis remains unclear[268]. Homologous proteins to gene *E* and gene *G* found in malonomycin biosynthesis encode a dipeptide portion of the molecule which is then removed in the final malonomycin molecule, and so have not yet been characterised[269]. Homologous proteins to those encoded by genes *E* and *F* can be found in landepoxcin and eponemycin biosynthesis. In these BGCs, a homologous proteins to that encoded by gene *E* catalyses the introduction of leucine into the peptide and a homolog of the product of gene *F* catalyses the dehydration of the leucine side chain,[270,271] seen in Figure 5.12. This suggests to us that compound 1 will be made up of a L-dehydro-leucine and L-valine, which is then built on by the other members of the BGC.

One of the hypothetical genes, gene *K*, appears to encode something that is similar to a threonyl tRNA synthetase. The catalytic domain looks highly modified as compared to those with structures available on the PDB, and the protein has no editing domain (Figure 5.13). This would suggest that this protein cannot discriminate serine and threonine, and so is not a functional candidate as a resistance gene, but could be a biosynthetic gene.

In the AlphaFold model of the product of gene *K*, the substrate binding pocket appears to be more open and the long beta sheet in the catalytic domain which is usually five strands is elongated to nine. This extension of the beta sheet ends with a small pocket made of the four additional beta strands and a pair of alpha helices, with what appears to be a hydrophobic pocket. This could be an extended binding pocket for the substrate(s) of the enzyme or have an alternative function.



Landepoxin A



Eponemycin

Figure 5.12. Structures of landepoxcin A and eponemycin. Portions which are synthesised by Gene E and Gene F homologues are shown in red.

A comparison of the product of gene *K* to EcThrRS with respect to key residues for function can be found in Table 5.4. Generally, the threonine and tRNA binding regions of the product of gene *K* appear similar to EcThrRS and the zinc binding pocket of the product of gene *K* appears to be intact, although with aspartic acid in place of cysteine. This can be observed in some zinc finger proteins with minimal deviation in activity[272].

It appears that the product of gene *K* has all of the residues required for adenylation, although the adenosine ring binding part of the adenosine binding pocket is slightly modified- this could mean that a different substrate is accepted. While a biosynthetic ThrRS paralogue has yet to be reported, other aminoacyl tRNA synthetases have been shown to have non-canonical biosynthetic roles[92]. For example, in Agrocin 84 biosynthesis, AgnA is a truncated AsnRS which is lacking its anticodon-binding domain and so is thought to mediate addition of the methyl pentanamide group to the 5' phosphate of the nucleotide[147], as seen in Figure 5.14.

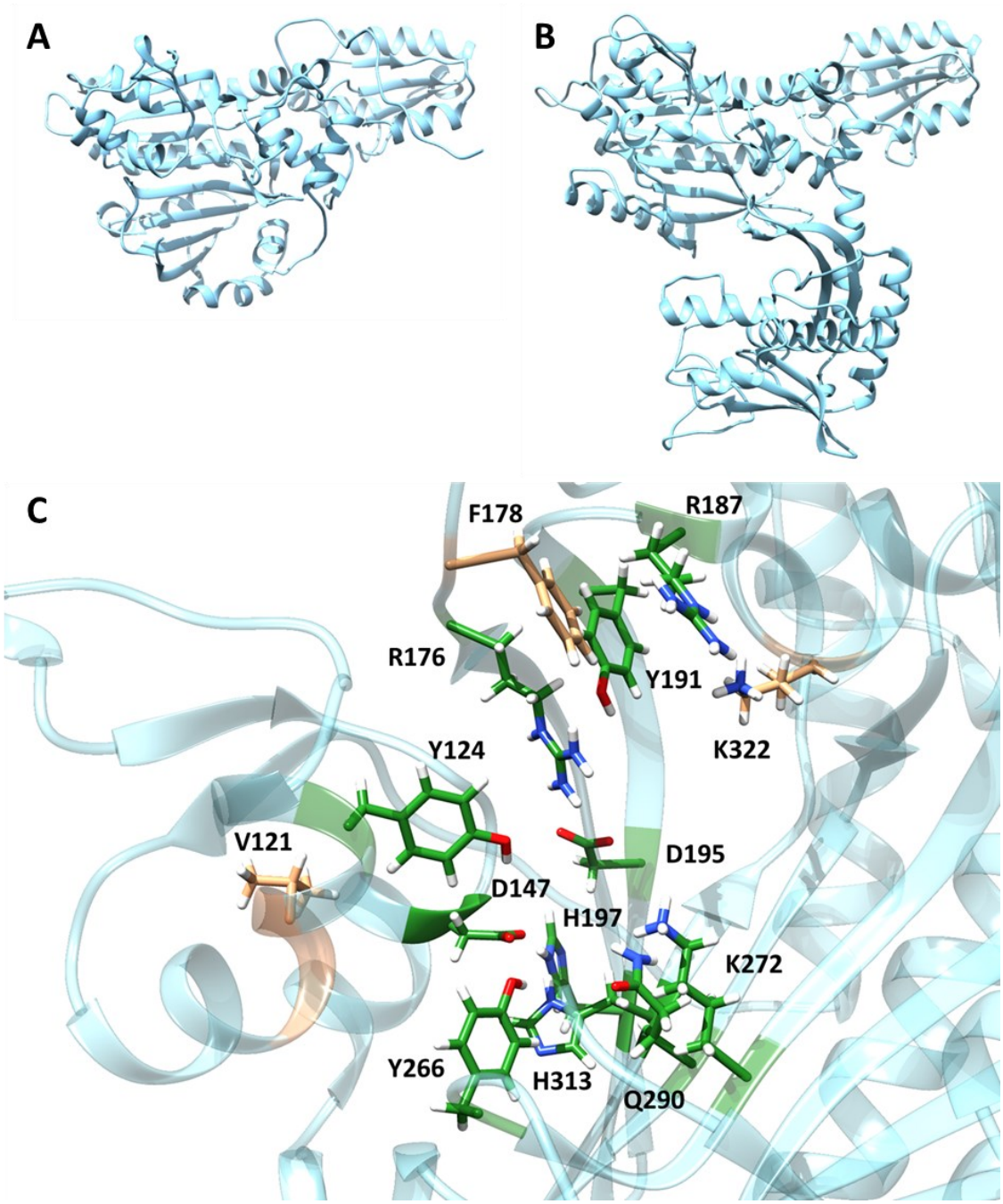


Figure 5.13. AlphaFold models of ThrRS paralogues in this cluster reveal a potential new function for the product of gene K. A) Gene K-encoded putative biosynthetic ThrRS with a novel structure as compared to other bacterial ThrRSs. **B**) Product of gene O- the putative resistance mechanism with a homologue in *Frankia torreyi*. **C)** The active site of the product of gene K, with the important active site residues for substrate binding and aminoacylation labelled as per their numbering in the product of gene K and shown as sticks. Those which are the same as in EcThrRS are shown in green, those which deviate shown in orange. Figure generated in Chimera. AlphaFold model statistics can be found in Supplemental Figure 25 and Supplemental Figure 26.

Table 5.4, Comparison to EcThrRS suggests that the product of gene K cannot aminoacylate but can adenylate. Key residues for activity in EcThrRS listed, with the equivalent reisdue in the product of gene K. Those which are likely to still be functional shaded in green, those which are not are shaded in orange. The function of the residue shown in the third column.

Key Residue in EcThrRS	**Residue in the product of gene *K***	**Purpose**
H309	V121	Catalysing Aminoacylation
Y313	Y124	tRNA terminal Adenosine binding
C334	D147	Zinc Coordination
R363	R176	Threonine/ATP/tRNA binding
E365	F178	ATP β and γ phosphate positioning
R375	R187	ATP γ-phosphate/tRNA C74 binding
F379	Y191	ATP adenosine binding
D383	D195	Threonine Binding
H385	H197	Zinc Coordination
Y462	Y266	Threonine/tRNA ribose binding
K465	K272	Catalysing Adenylation
Q484	Q290	tRNA ribose binding
H511	H313	Zinc Coordination
R520	K322	ATP adenosine binding
I547	V354	Anticodon recognition (X**GU**)
N575	N383	Anticodon Binding (**X**GU)
I578	L386	Anticodon recognition (X**GU**)
I582	L390	Anticodon recognition (X**GU**)
V595	V493	Anticodon recognition (X**GU**)
E600	E408	Anticodon recognition (X**G**U)
R609	D417	Anticodon recognition (XG**U**)

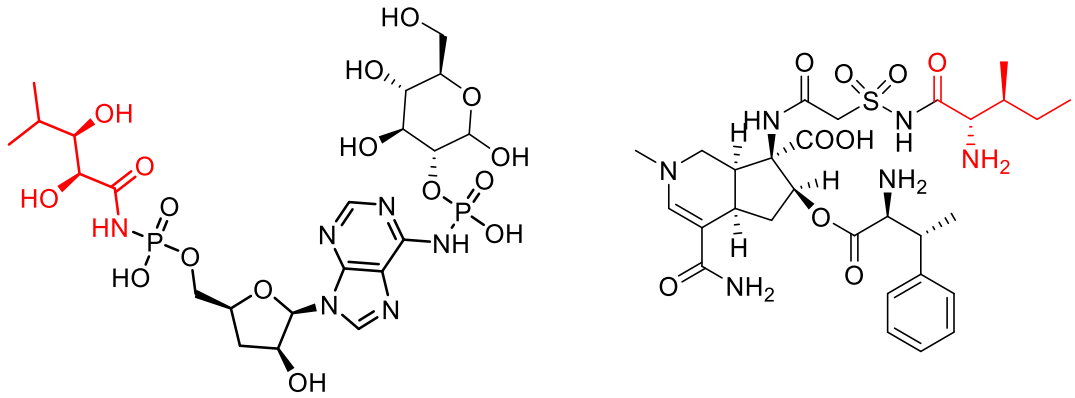


Figure 5.14. Structures of known natural products with aaRS paralogues inovlved in their biosynthesis. Portions of the compound added by aaRS paralogues coloured in red.

Crucially, the key residue for aminoacylation is missing in the product of gene *K* (H309 in EcThrRS), so while it appears to be capable of binding tRNA, threonine and ATP and adenylation, from a structural point of view, it is unlikely to aminoacylate tRNA. tRNA binding in this case could be structural, in order to ensure that the protein is in the correct conformation for action. In the biosynthesis of the sulfonamide SB-203208 (Figure 5.14), a paralogue of IleRS, SbzA catalyses the transfer of isoleucine onto the natural product during its biosynthesis, in a tRNA dependent manner[273].

There are two possible resistance mechanisms present in the compound 1 BGC; the ThrRS, encoded by gene *O*, which was the reason for this cluster being identified, and a pentapeptide repeat protein (PRP), encoded by gene I. PRPs, as discussed briefly in section 1.2.2, can be gyrase/topoisomerase inhibitor resistance proteins[68-70]. The product of gene *O* can be predicted to be sensitive to both borrelidin and obafluorin based in our mechanism of resistance models, discussed in Chapters 3 and 4, suggesting that this putative resistance gene is not conferring resistance to either of the two known ThrRS inhibitors. These factors, combined with the existence of a housekeeping ThrRS already in the strain would suggest that the product of gene *O* is a genuine resistance gene to a novel ThrRS inhibitor with a different mechanism of resistance than ObaO or BorO. This compound 1 could therefore be a dual topoisomerase/ThrRS inhibitor. This is not unprecedented- taxifolin is a plant polyphenol which was found to target both gyrase and IleRS by binding to the ATP binding site of gyrase and in the editing site of IleRS[274] (see Figure 5.15).

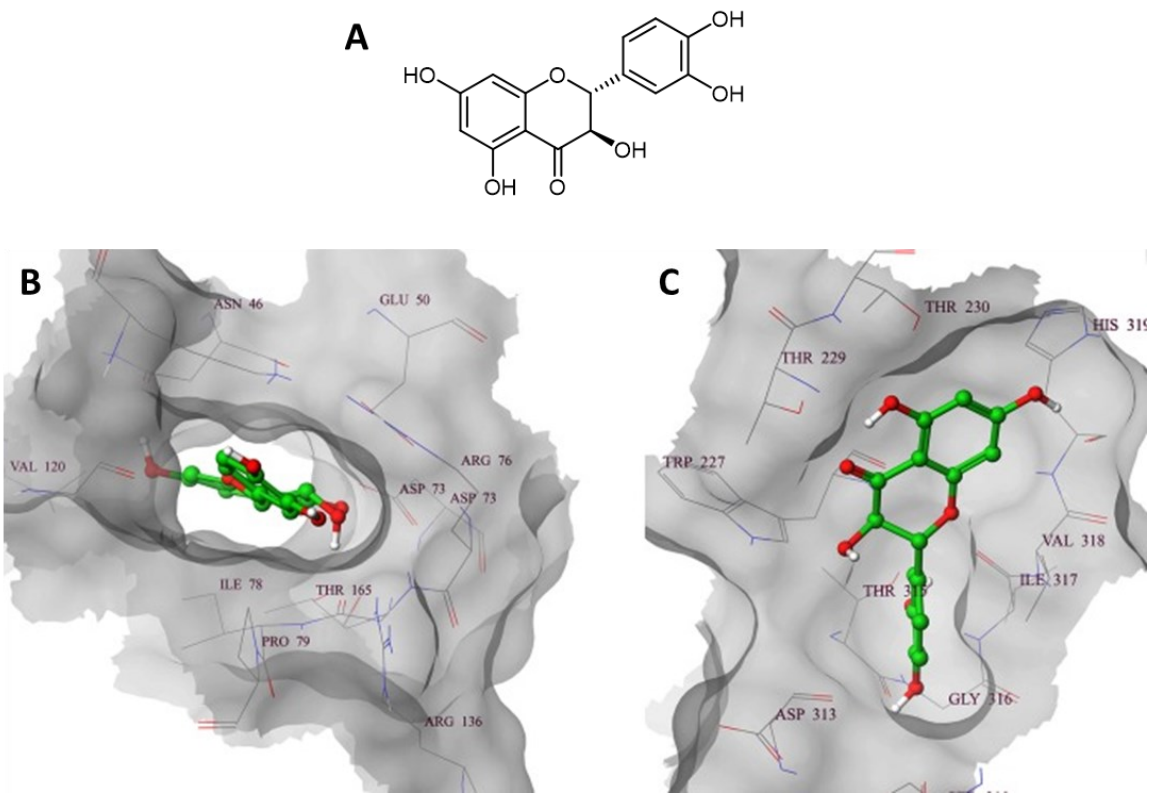


Figure 5.15. Structure of taxifolin and of it binding to its protein targets. A) Chemical structure of taxifolin **B)** Taxifolin shown in green bound in the DNA gyrase ATP-binding site. **C**) Taxifolin shown in green bound in the IleRS editing site. Figure adapted from Kozhikkadan Davis et al. 2018[274].

5.2.2.3 M. KC207 variably inhibited E. coli and B. subtilis on a variety of different media.

As an initial screen to check that *Micromonospora.* sp. KC207 is capable of producing bioactive compounds, overlay bioassays were performed with *E. coli* NR698, WT *E. coli* and *B. subtilis*, on a range of different *Streptoymces* media. *Micromonospora.* sp. KC207 was grown on agar plates as a streak, inoculated from a mycellium stock grown at 30°C for 1 week, until good growth could be observed. The bioindicator strain in Soft Nutrient Agar (SNA) was then overlayed on top of the *Micromonospora* and the plate was incubated at 30°C for a further day, before the plates were photographed.

The presense or absense of a zone of inhibition was noted, as seen in Table 5.5. Crucially, different inhibiton patterns were seen for different media, suggesting that different combinations of natural products are expressed on the plate in different conditions. The morphology of the *Micromonospora.* sp. KC207 colonies were also different. It was observed that *Micromonospora.* sp. KC207 starts life as a white colour, then develops an orange pigmentation, which then darkens, as it sporulates, becoming black. Example photographs can be seen in Supplemental Figures 17-22.

In these assays, *Micromonospora*. sp. KC207 grown on SFM produced compounds that inhibited both *E. coli* strains but not *B. subtilis*, suggesting a Gram-negative specific target and also the ability to easily penetrate the cell. *Micromonospora.* sp. KC207 grown on MYM produced compounds which could only kill *E. coli* NR698 and *B. subtilis*, suggesting a more general target, but poor membrane permeability. The antibacterial natural product(s) produced by *Micromonospora.* sp. KC207 on DNA, Bennett's and MAM only kills *B. subtilis*, suggesting Gram-positive specific target(s), whilst the antibacterial metabolite(s) produced by *Micromonospora.* sp. KC207 on GYM kills only NR698, suggesting a Gram-negative specific target and poor membrane mermeability. Because at this stage, we don't know what is being produced and in what quantities, it is difficult to draw any conclusions. Some of the zones of inhibition, such as those on GYM and MYM are quite small, suggesting a compound which is less potent, less diffusible and/or produced in low titres. On the other hand, on MAM, almost the entire plate of *B. subtilis* had been killed, likewise for Bennett's agar and DNA.

Table 5.5, bioactivity profile of *M*icromonospora. sp. KC207 against *E. coli* NR698, WT *E. coli* and *B. subtilis.* *Micromonospora*. sp. KC207 grown on different media for the same period of time before overlay with bioindicator strains. A tick represents the presence of a zone of inhibitor, cross represents the absense.

	SFM	MYM	MAM	GYM	Bennett's	DNA
NR698	✓	✓	×	✓	×	×
E. coli	✓	×	×	×	×	×
B. subtilis	×	✓	✓	×	✓	✓

These media were all reasonably rich but with different carbon sources, with DNA being the least rich. A more minimal media should be tested additionally in future in order to get a comparison- it may elicit different nautral products to be produced. Time allowed no further exploration of compound 1 production, but it is clear that *Micromonospora*. sp. KC207 is a talented strain, capable of producing a number of different antibiotics under laboratory conditions. The ThrRS paralogs in the BGC of compound 1 are of particular interest for future work.

5.2.2.4 A homolog of gene O from the compound 1 BGC can be found in a BGC in a strain of Cellulomonas

A final BLAST search revealed another deposited sequence encoding a ThrRS, homologous to gene *O* from the compound 1 BGC. This was identified in the genome of *Cellulomonas carbonis* str. CGMCC 1.10786, a motile *Actinomycete* isolated from coal mine soil in China[275]. This strain contained only four BGCs as seen in Figure 5.16; the BGC of interest is in region 10.1 and has no identifiable homologous BGCs in other strains. The possible biosynthetic genes for this BGC can be seen listed in Supplemental Table 11, and the BGC in Figure 5.17. Research on *Cellulomonas* has to date focussed on the ability of members of the genus to degrade cellulose[276]. The biosynthetic potential of this strain does seem limited, with no *Cellulomonas*-derived natural products having been deposited to the Natural Products Atlas[277,278], and no *Cellulomonas*-derived BGCs having been deposited on MIBiG[279], and with only 151 *Cellulomonas* genomes deposited on GenBank. When this information is taken together, it makes sense that this BGC has little homologue information. Thus, regardless of the target of this natural product, it would be interesting to characterise it. For comparison, there are 258 *Micromonospora*, 391 *Pseudomonas* and 619 *Streptomyces* genomes deposited on GenBank. Even *Frankia*, which is relative underexplored with only 23 deposited genomes, has published literature addressing their biosynthetic potential.[280]

Region	Type	From	To	Most similar known cluster		Similarity
Region 1.1	T3PKS	302,095	343,159	alkylresorcinol	Polyketide	100%
Region 2.1	lassopeptide	871,068	892,695			
Region 3.1	terpene	48,179	69,054	carotenoid	Terpene	66%
Region 10.1	NRPS	11,666	54,885			

Figure 5.16. antiSMASH report for *Cellulomonas carbonis* str. CGMCC 1.10786.

The main biosynthetic machinery encoded in this BGC are NRPS proteins (encoded by genes *13, 18, 21* and *24*). There are two possible resistance determinants in this BGC; gene *11* encodes a MFS transporter- this is likely involved in export of the product of the BGC, and therefore possibly also self-resistance. Gene *12* encodes a putative ThrRS which may be acting as a self-resistance gene in this BGC.

It is interesting that we have found homologs of the ThrRS originally identified in *Frankia torreyi* in four strains. Two of these strains also contained BorO homologs but no borrelidin BGC. The other two strains, which did not contain borrelidin BGCs, encoded homologues of this ThrRS located in a novel BGC for which no compound has been identified- *Cellulomonas carbonis* and *Micromonospora* sp. KC207. This presents two possible novel ThrRS inhibitors with BGCs in strains which have been isolated which can be characterised in future.

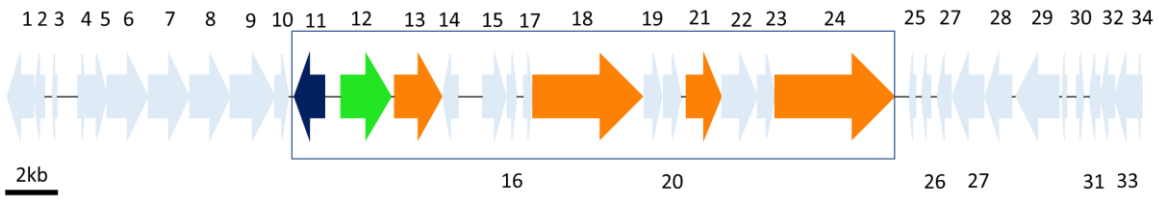


Figure 5.17. A novel BGC containing a ThrRS in *Tetrasphaera* sp. Aved_18-Q3-R54-62_MAXAC.378. Genes are numbered above/below the arrows. The ThrRS is coloured in lime green, the NRPS-associated genes are coloured in orange, MFS transporter in dark blue and all other genes coloured in light blue. Blue box shows best guess at the actual BGC in this genome.

5.2.3 ObaO homologues clade separately to their housekeeping ThrRSs

In order to search for novel BGCs using ObaO as a query, a BLAST search was used to identify 100 close homologues of both PfObaO and BmObaO, yielding two distinct lists of protein sequences. These 200 close homologues were aligned with MUSCLE and only those which we can predict to be resistant to obafluorin, based on not having glycine in the equivalent of position 463 (relative to EcThrRS) along with asparagine in position 316 (relative to EcThrRS), were investigated further. This reduced the list to 87 sequences. Of these, 23 had an alanine in the resistance position; a mutation which has yet to be characterised in the lab. These strains were mainly from *Rhodanobacter* strains (19/23) with one *Bacillus* example, two *Dyella* and one *Frateuria*. All other sequences had serine in position 463, relative to EcThrRS, with 15 *Pseudomonas* sequences, 2 *Chitiniphilus* sequences, 1 *Trinickia* sequence and 46 *Burkholderia* sequences (of which 37 were from different strains of *Burkholderia multivorans* and 6 were from different strains of *Burkholderia diffusa*). A representative from each species was chosen with the rest discarded. This left us with 45 sequences, details of which can be seen in Table 5.6.

Table 5.6. List of ObaO analogues identified by a BLAST search. % identity compared to the Pseudomonas fluorescens ATCC 39502 ObaO is displayed. The genomic context was examined to see if the ObaO homologue was in a BGC or specifically in an Obafluorin BGC. The number of ThrRSs in each genome is listed, the identity at each of the residues thought to be associated with obafluorin resistance is listed. Legitimate ObaO proteins are coloured in green, possibly obafluorin resistant ObaO-like proteins are shown in orange.

Strain	% Identity	In BGC?	Obafluorin BGC?	# ThrRSs
Pseudomonas fluorescens ATCC 39502	100	Y	Y	2
Pseudomonas viridiflava	99.69	Y	Y	2

Pseudomonas sp. 34 E 7	99.53	Y	Y	2
Pseudomonas orientalis	99.22	Y	Y	2
Pseudomonas salmasensis	97.49	Y	Y	2
Pseudomonas mandelii	92.62	Y	Y	2
Pseudomonadales bacterium RISCSPLOWO2_02_FULL_63_210	92.62	N	N	1
Pseudomonas sp. Irchel s3a18	81.29	Y	Y	2
Pseudomonas sp. Kh13	80.66	Y	Y	2
Pseudomonas sp. SK	80.50	Y	Y	2
Pseudomonas sp. EMN2	80.35	Y	Y	2
Pseudomonas anuradhapurensis	80.35	Y	Y	2
Pseudomonas sp. LAM2023	80.35	Y	Y	2
Pseudomonas parasichuanensis	80.19	Y	Y	2
Chitiniphilus eburneus	78.62	Y	Y	2
Chitiniphilus shinanonensis	77.67	Y	Y	2
Frateuria sp. Soil773	74.41	N	N	1
Dyella sp. RRB7	74.09	N	N	1
Rhodanobacter sp. 7MK24	73.62	N	N	1
Rhodanobacter glycinis	73.46	N	N	1
Rhodanobacter sp. B05	73.30	N	N	1
Rhodanobacter sp. T12-5	73.30	N	N	1
Dyella sp. S184	73.30	N	N	1
Rhodanobacter sp. LX_99	73.14	N	N	1
Rhodanobacter sp. MP1X3	73.14	N	N	1
Rhodanobacter sp. C05	73.14	N	N	1
Rhodanobacter sp. Root179	72.99	N	N	1
Rhodanobacter sp. Soil772	72.99	N	N	1
Rhodanobacter sp. C03	72.99	N	N	1
Rhodanobacter sp. C06	72.83	N	N	1
Bacillus sp. SRB_336	72.83	N	N	2
Rhodanobacter sp. Root561	72.83	N	N	1
Rhodanobacter sp. FDAARGOS 1247	72.83	N	N	1
Rhodanobacter spathiphylli	72.83	N	N	1

Rhodanobacter sp. A1T4	72.83	N	N	1
Rhodanobacter sp. MP7CTX1	72.83	N	N	1
Rhodanobacter sp. OK091	72.67	N	N	1
Rhodanobacter sp. C01	72.67	N	N	1
Burkholderia diffusa	71.09	Y	Y	2
Burkholderia stagnalis	70.77	Y	Y	2
Burkholderia sp. Bp8998	70.62	Y	Y	2
Burkholderia territorii	70.46	Y	Y	2
Burkholderia multivorans	70.46	Y	Y	2
Trinickia dinghuensis	69.98	N	N	2

The genomic context of these genes and the number of ThrRS genes in each genome were examined. It appears that for those sequences with an alanine in the equivalent to position 461 in EcThrRS, there are no secondary ThrRS genes in the genome, except for the *Bacillus* sequence. This would suggest that these proteins are housekeeping proteins; as would the fact that none of them are found in or near a BGC. If these proteins can confer resistance to obafluorin, it may give insight into the evolutionary history of ObaO and indeed obafluorin resistance, although we have no evidence that they would confer resistance.

With the exceptions of the sequences from the *Trinickia* and unassigned *Pseudomonadales* genomes, all legitimate ObaO homologues were only found in genomes encoding a clearly *bone fide* obafluorin BGC. The *Trinickia* ObaO is highly similar (93% identical) to BmObaO. This could suggest that it has been acquired by *Trinickia dinghuensis*, as a resistance gene against obafluorin, from an obafluorin producer.

A phylogenetic tree of the ObaO homologues and their housekeeping ThrRSs (Figure 5.18) reveal that ObaOs clade separately to the housekeeping proteins, with the *Pseudomonas* sequences all clading together, the *Chitiniphilus* sequences clading separate to them but closer than to the *Burkholderia* sequences. The *Pseudomonadales bacterium* sequence, which is not in an obafluorin BGC, clades among the *Pseudomonas* sequences, while the *Trinickia* sequence clades with the *Burkholderia* sequences. All the sequences with an alanine in the "resistance position" clade together, closer to the ObaO sequences than the housekeeping ThrRS sequences. The *Burkholderia*, *Trinickia* and *Chitiniphilus* housekeeping sequences clade together, while the *Pseudomonas* housekeeping sequences clade together and separate to the others. The second "normal" *Bacillus*

housekeeping ThrRS is significantly different to all the other sequences in this analysis and so was used as an outgroup.

All of this comes together to suggest that all ObaOs have a common origin, but at some point diverged into two groups, *Pseudomonas*-like and *Burkholderia*-like. It appears that the *Pseudomonales bacterium* ObaO is most similar to the ObaO from *Pseudomonas mandelii* and may have been acquired as a resistance gene in the community. The *Chitiniphilus* ObaOs are the most similar to the *Pseudomonas* ObaOs, despite their housekeeping ThrRSs being most similar to those from *Burkholderia*, suggesting that a horizontal gene transfer event may have taken place at some point. Likewise, the *Trinickia* ObaO is most similar to the *Burkholderia* ones and may have been acquired as a resistance gene in the absence of an obafluorin BGC. Of the ObaO-like proteins with an alanine in the "resistance" position, the *Bacillus* appears to be acquired from this group, with the *Bacillus* housekeeping ThrRS being only distantly related to these proteins. This would suggest that this protein could have a role in this specific *Bacillus* strain- possibly resistance.

A search for ObaO homologues showed that the situation for obafluorin resistance seems to be a lot simpler than for borrelidin resistance; generally, each ObaO homologue is in an obafluorin BGC, and obafluorin producer strains also encode obafluorin sensitive ThrRSs. While no novel BGCs could be identified from this analysis, this is a first step in exploring the evolutionary history of ObaO; this will not only allow us to understand how these proteins evolved, but will potentially allow us to discover other natural products in the future. Additionally, the discovery of ObaO homologues in the absence of an obafluorin BGC could be telling us that it can function as a transmissible resistance gene.

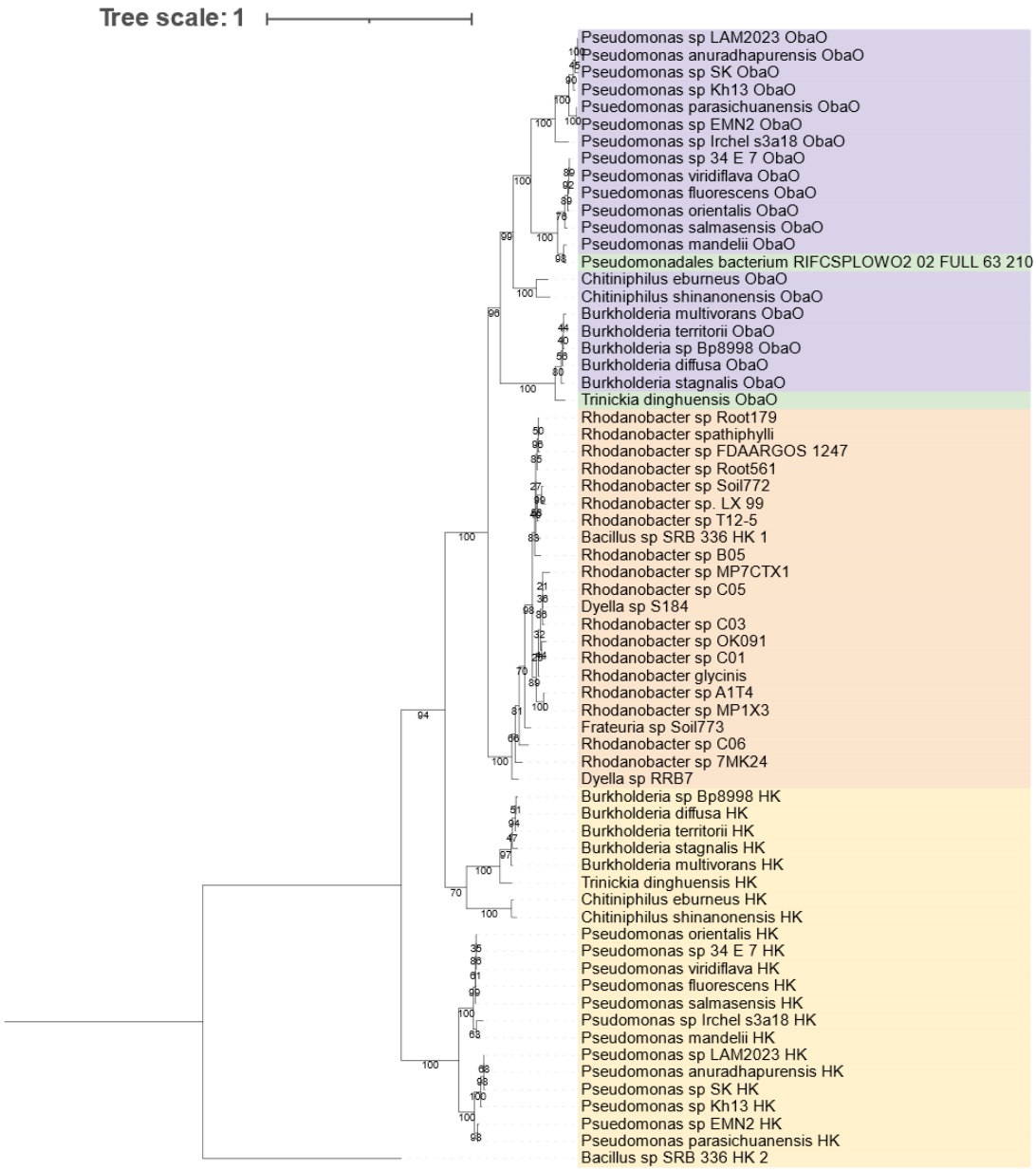


Figure 5.18. *M*aximum likelihood phylogenetic tree of ObaO homologues shows that they are phylogenetically distinct to their housekeeping proteins. Branch lengths to scale, with bootstrap values on each branch. *Bacillus* sp. SRB 336 second ThrRS claded far enough from all other sequences to use as outgroup, tree therefore was re-rooted at this branch. ObaO homologues are coloured in purple when found in obafluorin BGCs, green when not found in obafluorin BGCs, predicted possibly resistant housekeeping ThrRSs coloured in orange and predicted sensitive housekeeping ThrRSs coloured in yellow. Figure generated in the interactive tree of life (iTOL)[205].

5.2.4 A global phylogenetic tree indicates horizontal gene transfer as a major mechanism in the evolution of self-resistance genes

A global analysis of ThrRSs can give us clues about the evolution of our self-resistance genes of interest, the prevalence of borrelidin and obafluorin resistance in nature, and possibly the identification of novel BGCs which produce ThrRS targeting compounds.

A collection of ThrRS protein sequences had already been made previously in our lab, and a phylogenetic tree inferred as published in Scott *et al.* 2019[174]. The BorO homologue sequences, along with the sequences for the additional ThrRSs in their genomes, the ObaO homologues sequences and their additional housekeeping sequences, and the ThrRSs from *Micromonospora* sp. KC207 and *Cellulomonas carbonis* were added to this list. These sequences were then all aligned, the alignment manually trimmed, and a global phylogenetic tree of ThrRSs inferred (Figure 5.19). Sequences which are identical to each other were removed from the analysis. As for the BorO tree in section 5.2.1, the N-terminal editing domain is absent in this analysis due to its absence in truncated homologues, and other parts of the alignment which poorly aligned were trimmed. BorO homologues and housekeeping ThrRSs from *Actinomycetes* clade separately to each other. Likewise, ObaO homologues and their housekeeping proteins also clade separately.

The BorO sequences clade most closely with the ThrRSs from *Cyanobacteria*, *Deinococcota*, Green sulfur bacteria (*Chlorobiota*) and *Bacteroidota* – all of these are Gram-negative bacteria (Supplemental Figure 35). The next closest clade are a clade of bacteria-like archaeal sequences. The ThrRS from *Sulfolobus solfataricus,* which has previously been demonstrated to be sensitive to borrelidin[161], is in this group. The Archaeal proteins which had previously been shown to be resistant to borrelidin are found in the main clade of archaeal ThrRSs, which are distinct to those from bacteria. The existence of two separate archaeal ThrRS clades explains why only some archaeal ThrRSs are resistant to borrelidin[110,161]. It is interesting that BorO clades with sequences derived from Gram-negative bacteria. This may suggest an ancient horizontal gene transfer event which brought the ancestral BorO into *Actinomycetes.* The housekeeping genes for the BorO-containing strains generally clade with the other *Actinomycete* housekeeping genes.

The truncated housekeeping proteins cluster together, being most similar to those from Gram-positive *Bacillota* (Supplemental Figure 36). This could suggest that these ThrRS proteins were obtained from the *Bacillota* in a horizontal gene transfer event, and the editing domain truncated at some point. Their function in the bacteria containing them remains unclear.

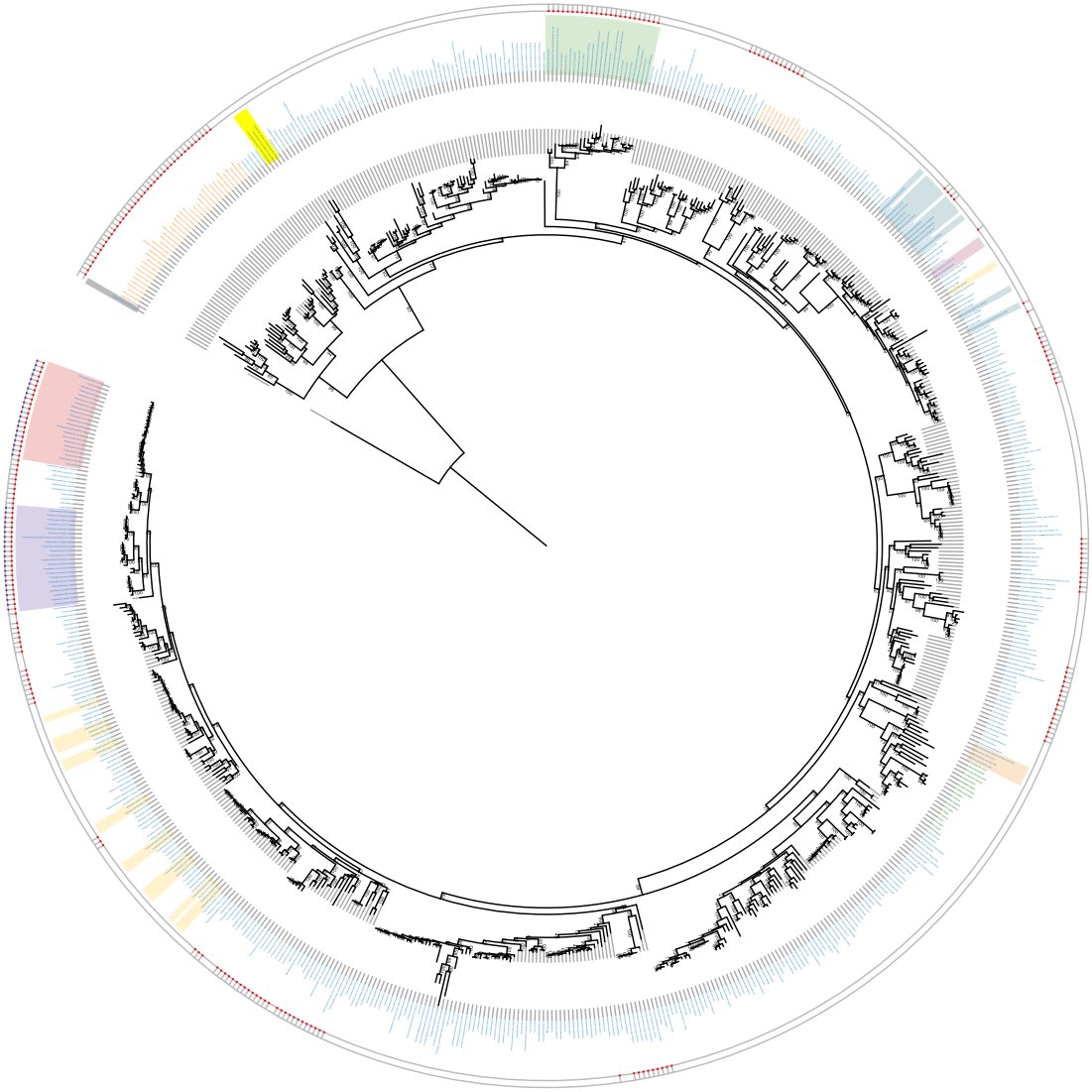


Figure 5.19. Maximum likelihood phylogenetic tree of ThrRS sequences from across the tree of life. Branch lengths to scale, with bootstrap values on each branch. Bacterial strain names in blue, archaeal strain names in orange, eukaryote strain names in green. BorO homologues are highlighted in green, BorO-containing strain housekeeping ThrRSs highlighted in blue and truncated housekeeping proteins highlighted in yellow, Gene O homologues in orange, Gene O-containing homologue ThrRSs highlighted in pink, Gene K is highlighted in grey, ObaO homologues highlighted in purple and ObaO-like proteins highlighted in red, ObaO containing strain housekeeping ThrRSs highlighted in pale yellow. Red exterior circles indicate predicted resistance to borrelidin, while exterior purple circle indicate predicted resistance to obafluorin. Figure generated in the interactive tree of life (iTOL).[205]

The homologues of the potential resistance gene, gene *O* found in the *M.* sp. KC207 compound 1 BGC and *C. carbonis* BGC discussed in sections 5.2.2.2 and 5.2.2.4 respectively, clade with the third ThrRS from *F. torreyi*. These sequences are placed most closely to the eukaryotic sequences (Supplemental Figure 37). This could suggest that these proteins were acquired from eukaryotes. Horizontal gene transfer from eukaryotes to bacteria has been observed before. For example, the mupirocin resistance protein, MupM has been shown to cluster closely with eukaryotic IleRSs, and so it has been suggested that a horizontal gene transfer event from eukaryotes was the source of this protein[281,282].

The protein encoded by gene *K*, the potential biosynthetic ThrRS from the compound 1 BGC, was the most distinct of all the sequences in the phylogenetic tree and so was used as the outgroup, with the tree being rooted at this node (Supplemental Figure 38). This therefore placed it closest to the basal archaeal clade, where all the proteins lack the borrelidin binding pocket[110,161]. It has been suggested that the archaea diverged rapidly from the bacterial and eukaryotic ThrRSs[236], so this protein may have evolved from these archaeal proteins, being transferred to the cell and then neofunctionalised for the potential biosynthetic role. It would be fascinating if other homologues could be found, to both characterise the reaction catalysed by this enzyme, and also to explore its evolution.

The ObaO homologues are placed amongst the other *Pseudomadota* sequences (Supplemental Figure 39). The placement of ObaO suggests that it may have evolved from gene duplication and development of self-resistance, rather than horizontal gene transfer.

This global phylogenetic tree (Figure 5.19) shows us that there are many strains which may have evolved borrelidin resistance- whether this is from exposure to borrelidin or chance is unknown. Most strains appear to have developed resistance by acquisition of methionine in position 489, relative to EcThrRS. This is the same mutation which confers borrelidin resistance by ObaO. This same mutation is found in multiple different parts of the phylogenetic tree, suggesting an example of convergent evolution.

The confinement of obafluorin resistance in the tree (according to our current model) to only ObaO homologues could be very telling regarding its reactivity. In nature, obafluorin may not be necessarily persistent enough to generate spontaneous resistance. We could generate spontaneous resistance in the producer, as discussed in section 4.2.12, but this was constantly "seeing" obafluorin over the period of the experiment due to its design, whereas in nature, this

may not be the case. This therefore presents a prime opportunity for less hydrolysis-prone analogues of obafluorin to be developed for more generally resistant strain.

5.3 Conclusions and Future Work

5.3.1 The compound 1 BGC contains some potentially interesting enzymes, and could encode an interesting natural product

Time allowed no further exploration of compound 1 production. The next step would be to grow the strain on solid and in liquid for all of the tested media, for 1 or 2 weeks, take extracts of each of these in solid and liquid in triplicate, and test their bioactivity, in an *E. coli* NR698 strain expressing gene *O* and/or gene *I*, the two possible self-resistance genes in the BGC. Differential activity between resistance gene expressing and empty-vector containing *E. coli* NR698 would allow for the identification of ThrRS and/or gyrase targeting compounds. Comparative metabolomics between extracts which are resisted by strains containing these self-resistance genes and those which are not could then for the identification of compound 1, while also confirming the target of compound 1.

Additionally, RNA extractions of each of these conditions could be taken and semi-quantitative PCR could be used to identify if genes from the compound 1 BGC are being expressed in those conditions. Combined this would then give at least growth conditions for compound 1 production, and once again comparative metabolomics could identify compound 1.

In parallel to support this, a TAR cloning strategy has been designed, involving PCR of sections of the compound 1 BGC, which can then be assembled in yeast to produce a vector for expression of the BGC. This vector can then be tested for growth in various different *Actinomycete* "superhosts" such as *Streptomyces coelicolor*[190,283]. This plasmid can then be modified via gene knockouts or refactoring of promoters in order to further explore compound 1 biosynthesis. Because molecular biology techniques for *Micromonospora* strains are limited, this would represent the best opportunity for the use of genetics to elucidate the biosynthesis of compound 1. This could be done by sequential deletion of the genes in the cluster, monitoring the titres of compound 1 (once it has been identified), as well as the accumulation of any intermediates, which could tell us which enzymes are responsible for which transformations, as well as potentially inform us as to the sequence of reactions. The potential biosynthetic protein which is personally of the most interest would be the product of gene *K*; crystallisation of the protein would be a good first step, to validate the AlphaFold model.

To start work on elucidating a structure of the product of gene *O* and to prepare for potential biophysical analysis of the resistance determinants, the *E. coli* codon optimised gene was synthesised in pET28 and cloned into pET29 with the N-terminal editing domain truncated, as had been done for all ThrRSs to date. No overexpression of Gene O was observed for the full-length protein, or the truncated protein. It would be useful to adjust the constructs in order to get overexpression of this protein. This could involve moving or removing the His-tag, the use of a SUMO tag or an MBP tag to improve the solubility and expression of the protein, changing the codon optimisation or screening different expression hosts. Work is ongoing in our lab to establish a generalised *Streptomyces* based protein expression platform- and this would be a good candidate for such a system.

It would also be interesting to characterise the potential activities of compound 1 and gyrase; the *E. coli* gyrase could be used for crystallography and/or cryo-EM. Supercoiling assays could be done, and unwinding assays with other topoisomerases could be done, in the presence or absence of compound 1. These, combined with binding assays such as ITC, and mutagenesis of any obvious binding/resistance residue could then allow us to characterise both the mechanism(s) of action of compound 1 and the potential modes of action of on the products of gene *I* and gene *O*.

5.3.2 Phylogeny of ThrRSs identifies some areas for future work

The homology search revealed a couple of interesting housekeeping proteins; the *Nocardia* strain with a BorO homologue present was isolated from the lung of a human, and has a glutamate in the borrelidin resistance position (L489 in EcThrRS)- it would be interesting to make this mutation in EcThrRS and test borrelidin resistance. Additionally, the presence of truncated housekeeping proteins in a number of strains would be interesting to explore further. It would be interesting to express and purify one of these and test for function and fidelity- if they are active and able to maintain fidelity, it would be interesting to unpick their role in the cell. Because in both humans and *E. coli*, ThrRSs have been demonstrated to bind to secondary structure elements in mRNA to either promote or prevent translation (discussed in sections 1.3.1 and 1.3.2.1), these proteins could be involved in regulation in some way.

Our understanding of borrelidin resistance, combined with our phylogeny found in section 5.2.4 can also inform us as to clinically relevant strains which could be targeted for inhibition by borrelidin or its potential future derivatives. For example, all of the ESKAPE pathogens are predicted to be sensitive to borrelidin.

During homology searches to identify ObaO homologues, a clade of ObaO-like housekeeping proteins from *Rhodanobacter* and related strains were identified, including one in a *Bacillus* strain which is also a second ThrRS gene in the genome. The global phylogenetic tree (Figure 5.19) indicated that these ObaO-like housekeeping proteins and the ObaO homologues are phylogenetically related to the housekeeping ThrRSs from other *Pseudomadota*, suggesting that ObaO may have evolved from a gene duplication and neofunctionalization event, rather than a cross-phyla horizontal transfer event. From our data we propose that obafluorin biosynthesis is not widely distributed in nature, suggesting that if an obafluorin homologue with more stability could be produced, it could be useful for the treatment of a variety of diseases. It would be interesting to see the effect of G462A (relative to EcThrRS) mutation in EcThrRS/PfThrRS as well as obtaining one of these *Rhodanobacter* strains to explore its resistance profile to obafluorin, which could validate our model.

A broader phylogenetic analysis of ThrRSs and then filtering for secondary copies of ThrRSs would be a parsimonious way to identify other novel ThrRS paralogues. In our tree alone, several strains have multiple ThrRSs in different portions of the tree, which could inform us as to ThrRSs with novel functions or in novel BGCs.

It would also be interesting to explore further the evolution of BorO and ObaO using phylogenetic reconciliation, as was done for glycopeptides and their self-resistance genes[234,235]. This might tell us how long-ago BorO and ObaO evolved, their evolutionary origins, and potentially identify more BGCs which contain ThrRS self-resistance genes.

5.3.3 Other potential novel ThrRS inhibitor BGCs have been identified

Of the novel BGCs identified in this study, those from *Cellulomonas carbonis* and *Actinobacterium sp.* OV320 are from cultured strains, which appear to have ThrRSs located in a BGC and which could be acting as a self-resistance gene. It could therefore be interesting to obtain these strains.

In this chapter, we have set out the beginnings of an exploration of the evolutionary history of borrelidin and obafluorin biosynthesis and resistance, as well as begun work on the characterisation of a potential novel ThrRS inhibitor produced by *M.* sp. KC207. We also identified BGCs in *Actinobacterium sp.* OV320 and *Cellulomonas carbonis* which may produce ThrRS targeting compounds and could increase the repertoire of known ThrRS inhibitors beyond borrelidin and obafluorin.

Chapter 6: Conclusions and Future Work

This thesis set out to understand the mechanism of action of the structurally unique natural product ThrRS inhibitor obafluorin, and the mechanisms of self-resistance of obafluorin and a second natural product ThrRS inhibitor, borrelidin, by their ThrRS self-resistance proteins ObaO and BorO respectively. Hypotheses for the mechanisms of resistance of both proteins were built and tested, and the mechanism of action of obafluorin was elucidated. A search of genomes for potential ThrRS self-resistance genes then led to the identification of BGCs encoding the production of potentially novel ThrRS inhibitors.

6.1 Borrelidin

Chapter 3 of this work represents a comprehensive analysis of self-resistance to borrelidin by the producer organism *Streptomyces parvulus*. The genome of *S. parvulus* was sequenced for the first time, and this thesis reports the first conjugations of plasmids into this strain and determines the first structure of BorO, which is the first ThrRS from a *Streptomyces* to be solved. The mutations previously suggested to be important for BorO resistance to borrelidin were examined through mutagenesis and it was found that just one residue, T516 (L489 in EcThrRS), is sufficient to confer borrelidin resistance when compared to the target in sensitive bacteria such as *E. coli*.

Unexpectedly, the *S. parvulus* housekeeping protein, SpThrRS is also resistant to borrelidin, as was the obafluorin resistance protein, ObaO as identified in Chapter 4. ObaO's resistance to borrelidin can be traced to the presence of a methionine residue in the same key position for conferring resistance in BorO (T516 in BorO, L489 in EcThrRS). This understanding of the mechanism of borrelidin resistance in ObaO led to the first structure of an ObaO homologue to be solved by mutation of BmObaO to allow binding of borrelidin, and co-crystallisation of this mutant with borrelidin. The amino acid in this key position in SpThrRS is glutamine (Q510) but introducing this mutation alone into the equivalent position into the sensitive *E. coli* enzyme (L489Q) was not sufficient to confer borrelidin resistance to EcThrRS; however, mutation of this Q residue to leucine does cause SpThrRS to be able to bind borrelidin weakly. This suggests that a more complex borrelidin resistance mechanism may be at play. Based on a structural alignment of a range of ThrRSs with known borrelidin sensitivity or resistance, the residue Y482 (F461 in EcThrRS) was identified as a possible second residue involved in borrelidin resistance in SpThrRS.

A structure of SpThrRS would be useful for elucidating the full details of the resistance to borrelidin observed for SpThrRS. The use of Cryo-EM to solve the structure of SpThrRS seems to be an obvious next step for the study of the borrelidin self-resistance observed by SpThrRS. The elucidation of the borrelidin resistance of SpThrRS would be interesting, with this being a rare example of an antibiotic

resistant housekeeping enzyme in a strain which also encodes for a self-resistance gene within the antibiotic BGC.

In Chapter 5, homologues of BorO were identified in the genomes of several different *Actinomycetes* isolated from different ecological niches, all over the world. This would suggest, along with the frequent reports of borrelidin production by different strains that borrelidin producers will be present in many ecological niches and that the acquisition of the ability to produce borrelidin is beneficial. Several strains were identified with putative copies of *borO* in the absence of a borrelidin BGC. Different metagenomes assembled from single samples of the same community showed that BorO homologues were encoded in different genomic contexts and in different strains from within the same community, suggesting a role for BorO and its homologues in acquired resistance. Additionally, most sequenced genomes containing borrelidin BGCs also contain predicted resistant housekeeping ThrRSs, possibly suggesting either recent acquisition of the BGC by these strains, the development of resistance in their housekeeping genes following the acquisition of the borrelidin BGC or an unknown regulatory role of *borO* in these strains.

To explore this postulated horizontal gene transfer of *borO* between different strains, a phylogenetic reconciliation could be performed. This would date the acquisition of *borO* and the borrelidin BGC by these strains, as well as indicating possible directions of horizontal gene transfer. This would tell us about the evolutionary trajectory that *borO* has taken through time and possibly allow us to identify new BGCs which produce ThrRS targeting natural products by the identification of distant BorO homologues.

The potential role of BorO as a regulatory protein is interesting and could be further explored; a Δ*borO* mutant of *S. parvulus* produced slightly less borrelidin, while overexpression of *borO* led to an increase in borrelidin producing, suggesting that BorO may have a role in the regulation of borrelidin production. With this in mind, it would be useful to obtain one of the borrelidin BGC-containing strains with a putatively susceptible housekeeping ThrRS, knock out *borO,* and then assay borrelidin production. This could then discriminate a self-inhibition phenotype vs a regulation phenotype. Further to this, possible BorO-RNA interactions in *S. parvulus* could be probed by the use of an RNA-protein co-immunoprecipitation assay, a CLIP-seq, revealing the specific RNA which BorO can bind to. If only the cognate tRNAs are observed, then it is unlikely that BorO has a regulatory role, but if mRNA transcripts are identified, this could inform us on the possible role of BorO in regulation.

Additionally, the presence of an additional, editing-deficient copy of ThrRS in some BorO-containing strains could be explored. These proteins do not appear to be associated with BGCs and so may have some other role in the bacteria. To start, the function and fidelity of these proteins could be explored biochemically using both radiolabelled serine and threonine and comparing the rate of aminoacylation of tRNA by both. An CLIP-Seq experiment could identify possible transcripts which the truncated proteins may be interacting with and identifying a possible regulatory role for the protein.

6.2 Obafluorin

Chapter 4 of this work represents a systematic approach to elucidate the mechanism of action of obafluorin and the mechanism of self-resistance by the BGC associated resistance gene product ObaO. This work was challenging due to the reactivity of obafluorin, and therefore novel strategies were designed to overcome this drawback. A cryo-EM structure of the target in sensitive bacteria, EcThrRS, with covalently bound obafluorin was solved and identified the phenolic hydroxyl of tyrosine 462 as the nucleophile which attacks and ring opens the electrophilic β-lactone ring of obafluorin. This residue is present in all ThrRSs and is vital for threonine and tRNA binding. This mechanism requires specific geometry and residency time of obafluorin in the active site to allow covalent modification of Tyr462. Obafluorin is positioned by interaction of the catechol moiety with the active site Zn^{2+} ion, and through π-stacking interactions of the nitrophenol group with R363 and Y313; this is the same pair of amino acids that position the adenosine ring of the terminal nucleotide of the tRNA, A76, in the active site during its aminoacylation. It can be postulated, therefore, that binding of obafluorin acts as an aminoacylated tRNA mimic. This contrasts with borrelidin which binds into a separate binding site within the active site that is usually cryptic, and in doing so blocks the binding of threonine, tRNA and ATP.

Cryo-EM of ObaO Using the same experimental setup we obtained a cryo-EM structure of ObaO with obafluorin bound in the active site, but with low occupancy and no evidence of the covalent bond having been made as was observed for EcThrRS. This suggests that obafluorin binds to ObaO non-covalently, which in part could explain the partial inhibition previously observed in aminoacylation assays[174]. Spontaneous resistant mutants to obafluorin consistently exhibited a single point mutation G463S in the *Pseudomonas fluorescens* housekeeping gene that is directly adjacent to the essential Y462 residue which makes a covalent linkage with obafluorin. This S463 residue is completely conserved in ObaO homologues and could be important for rigidifying the Y462 containing loop, leading to Y462 being unable to be placed in the active site with correct geometry for obafluorin ring opening. This could be further explored using molecular dynamics

simulations which explore the conformations sampled by this portion of the protein in ObaO, and compared to EcThrRS.

The synthesis of a β-lactam (in place of the β-lactone moiety) version of obafluorin would provide more stable analogues that might enable a more detailed analysis of the non-covalent binding interaction of obafluorin to EcThrRS and ObaO, for example by using ITC methods. β-lactams are less chemically reactive than β-lactones and the tyrosine nucleophile is relatively weak, and so it may be unlikely to be able ring-open a β-lactam analogue of obafluorin unless its geometry is perfect within the active site, and the residency time is increased. Such a molecule could also be used in aminoacylation assays; if a partial inhibition is observed for EcThrRS, this could confirm that non-covalent binding of obafluorin to ObaO is in part the cause of the partial inhibition observed for ObaO in aminoacylation assays.

Additional significant residues for obafluorin self-resistance have been identified through this work. BmObaO M510L was unable to confer obafluorin resistance and was first generated during the work to elucidate the borrelidin resistance of ObaO. Additionally, identified based on sequence alignments of ObaO homologues and work with chimeric EcThrRS:ObaO proteins, EcThrRS E305K was able to confer partial resistance to obafluorin. It would be useful to characterise the reasons for these observations, structurally and biophysically. It therefore remains clear that further work is required to fully unpick the full ObaO mechanism of self-resistance to obafluorin.

6.3 Mining for novel ThrRS Inhibitors

In Chapter 5, using the sequences of our known ThrRS inhibitor self-resistance proteins as bait, genomes were examined to identify those with copies of ThrRS additional to the housekeeping copy; in particular we were searching for second copies that were collocated with biosynthetic genes (i.e., as part of a BGC). BorO homologues located within non-borrelidin BGCs were identified in four different strains, and the one with the most promise for further study was found in *Actinobacterium* sp. OV320. A third copy of a BorO-like ThrRS was identified in the genome of *Frankia torreyi*, and this was then used as bait to identify homologues within BGCs encoded by *Micromonospora* sp. KC207 and *Cellulomonas carbonis*. *M.* sp. KC207 has been obtained from collaborators in Turkey and its genome sequenced by us using Illumina paired end technology; the resulting assembly allowed us to identify this strain as a novel species. The BGC from *M.* sp. KC207 looks promising as one encoding a potential new ThrRS inhibitor. This BGC contains two possible self-resistance genes, first a ThrRS, and secondly a pentapeptide repeat protein typically involved in resistance to type II topoisomerase inhibitors, and this suggests that the product of this BGC could be a dual targeting natural product. Additionally, the BGC contains a second, modified ThrRS

paralogue which could have a biosynthetic role; if so, this would be the first biosynthetic ThrRS to be identified. The structure and function of this possible biosynthetic ThrRS paralogue would be a good avenue of future research. If the compound produced by this BGC can be isolated and characterised, the possible biosynthetic step catalysed by this gene could be identified.

Moreover, the potential self-resistance proteins could be utilised for the identification of the product of this BGC. The self-resistance gene(s) could be expressed in a bioindicator strain such as *E. coli* NR698 and the bioactivity of different extracts of the producer when grown on different media could be examined; any differential readouts used to identify compounds to which these self-resistance genes confer resistance. Comparative metabolomics could then be used to identify the specific compound, when combined with a bioactivity-guided fractionation approach.

Obtaining the *Actinobacterium* sp. OV320 and *Cellulomonas carbonis* strains would allow us to begin work to identify the compounds produced by the target BGCs. If all three of these BGCs produce ThrRS targeting natural products, the repertoire of ThrRS inhibitors would therefore expand from two to five.

A global phylogenetic tree of ThrRS proteins was inferred which showed that borrelidin resistance is widely distributed throughout nature, but that obafluorin resistance appears to be found only in close homologues to ObaO (based on our current state of knowledge). A broader analysis involving all annotated ThrRS sequences, and then filtering for multiple ThrRS sequences in the same genome, would be a parsimonious way to identify further ThrRS natural product inhibitors. From the limited analysis performed in this study, 3 potential BGCs were identified which warrant further study.

In summary, this thesis has explored the mechanism of borrelidin self-resistance of the BGC situated BorO, identified the obafluorin mechanism of action and proposed a model for the mechanism of self-resistance to obafluorin by the BGC situated ObaO. Finally, three novel BGCs have been identified which contain putative ThrRS self-resistance genes and which therefore encode three potential novel ThrRS inhibitors.

This work could be further expanded. As demonstrated by the use of borrelidin cross-resistance to solve a structure of ObaO, characterisation of ThrRS self-resistance proteins can aid with the understanding of new ThrRS self-resistance mechanisms, and mechanisms of action.

Chapter 7: References

1 Cologgi, D. L. *et al.* Extracellular reduction of uranium via *Geobacter* conductive pili as a protective cellular mechanism. *Proc Natl Acad Sci U S A* **108**, 15248-15252, doi:10.1073/pnas.1108616108 (2011).

2 Atanasova, N. *et al.* Plastic Degradation by Extremophilic Bacteria. *Int J Mol Sci* **22**, doi:10.3390/ijms22115610 (2021).

3 Pham, J. V. *et al.* A Review of the Microbial Production of Bioactive Natural Products and Biologics. *Frontiers in Microbiology* **10**, doi:10.3389/fmicb.2019.01404 (2019).

4 Davies, J. Specialized microbial metabolites: functions and origins. *The Journal of Antibiotics* **66**, 361-364, doi:10.1038/ja.2013.61 (2013).

5 Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research* **39**, W339-W346, doi:10.1093/nar/gkr466 (2011).

6 Li, M. H. *et al.* Automated genome mining for natural products. *BMC Bioinformatics* **10**, 185, doi:10.1186/1471-2105-10-185 (2009).

7 Skinnider, M. A . *et al.* PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res* **45**, W49-w54, doi:10.1093/nar/gkx320 (2017).

8 Lakhani, J. *et al.* Chapter 3 - *In silico* detection tools to identify fungal secondary metabolites and their biosynthetic gene clusters in *New and Future Developments in Microbial Biotechnology and Bioengineering* (eds Joginder Singh & Praveen Gehlot) 23-35 (Elsevier, 2020).

9 Baltz, R. H. Gifted microbes for genome mining and natural product discovery. *J Ind Microbiol Biotechnol* **44**, 573-588, doi:10.1007/s10295-016-1815-x (2017).

10 Cazzaniga, G. *et al.* Natural products against key *Mycobacterium tuberculosis* enzymatic targets: Emerging opportunities for drug discovery. *European Journal of Medicinal Chemistry* **224**, 113732, doi:https://doi.org/10.1016/j.ejmech.2021.113732 (2021).

11 Batt, S. M. *et al.* Antibiotics and resistance: the two-sided coin of the mycobacterial cell wall. *Cell Surf* **6**, 100044, doi:10.1016/j.tcsw.2020.100044 (2020).

12 Diagne, N. *et al.* Use of *Frankia* and actinorhizal plants for degraded lands reclamation. *Biomed Res Int* **2013**, 948258, doi:10.1155/2013/948258 (2013).

13 Pujic, P. *et al.* Omics of the early molecular dialogue between *Frankia alni* and *Alnus glutinosa* and the cellulase synton. *Environmental Microbiology* **21**, 3328-3345, doi:https://doi.org/10.1111/1462-2920.14606 (2019).

14 Hodiamont, C. J. *et al.* Clinical Pharmacokinetics of Gentamicin in Various Patient Populations and Consequences for Optimal Dosing for Gram-Negative Infections: An Updated Review. *Clinical Pharmacokinetics* **61**, 1075-1094, doi:10.1007/s40262-022-01143-0 (2022).

15 Oriel, J. D. & Ridgway, G. L. Comparison of erythromycin and oxytetracycline in the treatment of cervical infection by *Chlamydia trachomatis*. *Journal of Infection* **2**, 259-262, doi:https://doi.org/10.1016/S0163-4453(80)90722-7 (1980).

16 Oliynyk, M. *et al.* Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nature Biotechnology* **25**, 447-453, doi:10.1038/nbt1297 (2007).

17 Sayed, A. M. *et al. Saccharopolyspora*: an underexplored source for bioactive natural products. *Journal of Applied Microbiology* **128**, 314-329, doi:https://doi.org/10.1111/jam.14360 (2020).

18 Hifnawy, M. S. *et al.* The genus *Micromonospora* as a model microorganism for bioactive natural product discovery. *RSC Advances* **10**, 20939-20959, doi:10.1039/D0RA04025H (2020).

19 Lee, J. Y. *et al.* The Actinobacterium *Corynebacterium glutamicum*, an Industrial Workhorse. *J Microbiol Biotechnol* **26**, 807-822, doi:10.4014/jmb.1601.01053 (2016).

20 Tsuge, Y. & Matsuzawa, H. Recent progress in production of amino acid-derived chemicals using *Corynebacterium glutamicum*. *World J Microbiol Biotechnol* **37**, 49, doi:10.1007/s11274-021-03007-4 (2021).

21 Berdy, J. Bioactive microbial metabolites. *J Antibiot (Tokyo)* **58**, 1-26, doi:10.1038/ja.2005.1 (2005).

22 Jacob, C. & Weissman, K. J. Unpackaging the Roles of *Streptomyces* Natural Products. *Cell Chem Biol* **24**, 1194-1195, doi:10.1016/j.chembiol.2017.09.013 (2017).

23 Urem, M. *et al.* OsdR of *Streptomyces coelicolo*r and the Dormancy Regulator DevR of *Mycobacterium tuberculos*is Control Overlapping Regulons. *mSystems* **1**, doi:10.1128/mSystems.00014-16 (2016).

24 Jones, S. E. *et al. Streptomyces* Volatile Compounds Influence Exploration and Microbial Community Dynamics by Altering Iron Availability. *mBio* **10**, e00171-00119, doi:10.1128/mBio.00171-19 (2019).

25 Ismail, S. *et al.* Investigation of *Streptomyces scabies* Causing Potato Scab by Various Detection Techniques, Its Pathogenicity and Determination of Host-Disease Resistance in Potato Germplasm. *Pathogens* **9**, doi:10.3390/pathogens9090760 (2020).

26 Kirby, R. *et al.* Draft genome sequence of the human pathogen *Streptomyces somaliensis*, a significant cause of actinomycetoma. *J Bacteriol* **194**, 3544-3545, doi:10.1128/jb.00534-12 (2012).

27 Herbrík, A. *et al.* A Human Lung-Associated *Streptomyces* sp. TR1341 Produces Various Secondary Metabolites Responsible for Virulence, Cytotoxicity and Modulation of Immune Response. *Frontiers in Microbiology* **10**, doi:10.3389/fmicb.2019.03028 (2020).

28 Bolourian, A. & Mojtahedi, Z. Immunosuppressants produced by *Streptomyces*: evolution, hygiene hypothesis, tumour rapalog resistance and probiotics. *Environmental Microbiology Reports* **10**, 123-126, doi:https://doi.org/10.1111/1758-2229.12617 (2018).

29 Baroja-Mazo, A. *et al.* Immunosuppressive potency of mechanistic target of rapamycin inhibitors in solid-organ transplantation. *World J Transplant* **6**, 183-192, doi:10.5500/wjt.v6.i1.183 (2016).

30 Okada, H. *et al.* The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. *Clin Exp Immunol* **160**, 1-9, doi:10.1111/j.1365-2249.2010.04139.x (2010).

31 Gurney, R. & Thomas, C. M. Mupirocin: biosynthesis, special features and applications of an antibiotic from a Gram-negative bacterium. *Applied Microbiology and Biotechnology* **90**, 11-21, doi:10.1007/s00253-011-3128-3 (2011).

32 Saati-Santamaría, Z. *et al.* Unveiling the genomic potential of *Pseudomonas* type strains for discovering new natural products. *Microbial Genomics* **8**, doi:https://doi.org/10.1099/mgen.0.000758 (2022).

33 Scott, T. A. *et al.* An L-threonine transaldolase is required for L-threo-β-hydroxy-α-amino acid assembly during obafluorin biosynthesis. *Nat Commun* **8**, 15935, doi:10.1038/ncomms15935 (2017).

34 Xin, X. F. *et al. Pseudomonas syringae*: what it takes to be a pathogen. *Nat Rev Microbiol* **16**, 316-328, doi:10.1038/nrmicro.2018.17 (2018).

35 Kim, S. Y. *et al.* Different biosynthetic pathways to fosfomycin in *Pseudomonas syringae* and *Streptomyces* species. *Antimicrob Agents Chemother* **56**, 4175-4183, doi:10.1128/aac.06478-11 (2012).

36 Vior, N. M. *et al.* Discovery and Biosynthesis of the Antibiotic Bicyclomycin in Distantly Related Bacterial Classes. *Applied and Environmental Microbiology* **84**, e02828-02817, doi:doi:10.1128/AEM.02828-17 (2018).

37 Davies, J. C. *Pseudomonas aeruginosa* in cystic fibrosis: pathogenesis and persistence. *Paediatr Respir Rev* **3**, 128-134, doi:10.1016/s1526-0550(02)00003-3 (2002).

38 Mulcahy, L. R. *et al. Pseudomonas aeruginosa* biofilms in disease. *Microb Ecol* **68**, 1-12, doi:10.1007/s00248-013-0297-x (2014).

39 Mulani, M. S. *et al.* Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review. *Frontiers in Microbiology* **10**, doi:10.3389/fmicb.2019.00539 (2019).

40 Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology* **7**, 285-295, doi:10.1016/s1367-5931(03)00020-6 (2003).

41 Khosla, C. Structures and Mechanisms of Polyketide Synthases. *The Journal of Organic Chemistry* **74**, 6416-6420, doi:10.1021/jo9012089 (2009).

42 Nivina, A. *et al.*. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chemical Reviews* **119**, 12524-12547, doi:10.1021/acs.chemrev.9b00525 (2019).

43 Staunton, J. & Wilkinson, B. Biosynthesis of Erythromycin and Rapamycin. *Chemical Reviews* **97**, 2611-2630, doi:10.1021/cr9600316 (1997).

44 Villsen, I. D. *et al.* ErmE methyltransferase recognizes features of the primary and secondary structure in a motif within domain V of 23 S rRNA. *J Mol Biol* **286**, 365-374, doi:10.1006/jmbi.1998.2504 (1999).

45 Bibb, M. J. *et al.* Cloning and analysis of the promoter region of the erythromycin resistance gene (ermE) of *Streptomyces erythraeus. Gene* **38**, 215-226, doi:10.1016/0378-1119(85)90220-3 (1985).

46 Strieker, M. *et al.*. Nonribosomal peptide synthetases: structures and dynamics. *Current opinion in structural biology* **20**, 234-240, doi:10.1016/j.sbi.2010.01.009 (2010).

47 Konz, D. *et al.* The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol* **4**, 927-937, doi:10.1016/s1074-5521(97)90301-x (1997).

48 Radeck, J. *et al.* Anatomy of the bacitracin resistance network in *Bacillus subtilis*. *Molecular Microbiology* **100**, 607-620, doi:https://doi.org/10.1111/mmi.13336 (2016).

49 Johnson, B. A. *et al.* Bacitracin: A New Antibiotic Produced by a Member of the *B. subtilis* Group. *Science* **102**, 376-377, doi:doi:10.1126/science.102.2650.376 (1945).

50 Gottstein, J. *et al.* New insights into the resistance mechanism for the BceAB-type transporter SaNsrFP. *Scientific Reports* **12**, 4232, doi:10.1038/s41598-022-08095-2 (2022).

51 Neumüller, A. M. *et al.*The two-component regulatory system BacRS is associated with bacitracin 'self-resistance' of *Bacillus licheniformis* ATCC 10716. *Eur J Biochem* **268**, 3180-3189, doi:10.1046/j.1432-1327.2001.02203.x (2001).

52 Flores-Kim, J. & Darwin, A. J. The Phage Shock Protein Response. *Annu Rev Microbiol* **70**, 83-101, doi:10.1146/annurev-micro-102215-095359 (2016).

53 Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30**, 108-160, doi:10.1039/c2np20085f (2013).

54 Kelly, W. L. *et al.* Thiostrepton Biosynthesis: Prototype for a New Family of Bacteriocins. *Journal of the American Chemical Society* **131**, 4327-4334, doi:10.1021/ja807890a (2009).

55 Murakami, T. *et al.* Thiostrepton-induced gene expression in *Streptomyces lividans. J Bacteriol* **171**, 1459-1466, doi:10.1128/jb.171.3.1459-1466.1989 (1989).

56 Neff, E. P. Stop and smell the geosmin. *Lab Animal* **47**, 270-270, doi:10.1038/s41684-018-0161-1 (2018).

57 Becher, P. G. *et al.* Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a soil arthropod to *Streptomyces* bacteria promoting spore dispersal. *Nature Microbiology* **5**, 821-829, doi:10.1038/s41564-020-0697-x (2020).

58 Melo, N. *et al.* Geosmin Attracts *Aedes aegypti* Mosquitoes to Oviposition Sites. *Current Biology* **30**, 127-134.e125, doi:https://doi.org/10.1016/j.cub.2019.11.002 (2020).

59 Wilding, E. I. *et al.* Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *J Bacteriol* **182**, 4319-4327, doi:10.1128/jb.182.15.4319-4327.2000 (2000).

60 Helfrich, E. J. N., *et al.* Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J Org Chem* **15**, 2889-2906, doi:10.3762/bjoc.15.283 (2019).

61 Wrońska, N. *et al.* The Synergistic Effect of Triterpenoids and Flavonoids-New Approaches for Treating Bacterial Infections? *Molecules* **27**, doi:10.3390/molecules27030847 (2022).

62 Feng, Y. *et al.* Crystal Structure of Geranylgeranyl Pyrophosphate Synthase (CrtE) Involved in Cyanobacterial Terpenoid Biosynthesis. *Frontiers in Plant Science* **11**, doi:10.3389/fpls.2020.00589 (2020).

63 Graham, J. E. *et al.* Synechoxanthin, an Aromatic C40 Xanthophyll that Is a Major Carotenoid in the Cyanobacterium *Synechococcus* sp. PCC 7002. *Journal of Natural Products* **71**, 1647-1650, doi:10.1021/np800310b (2008).

64 Maresca, J. A. *et al.*. Isorenieratene biosynthesis in green sulfur bacteria requires the cooperative actions of two carotenoid cyclases. *J Bacteriol* **190**, 6384-6391, doi:10.1128/jb.00758-08 (2008).

65 Sugiyama, M. Structural biological study of self-resistance determinants in antibiotic-producing actinomycetes. *The Journal of Antibiotics* **68**, 543-550, doi:10.1038/ja.2015.32 (2015).

66 Zhao, Y. *et al.* Resistance-Nodulation-Division Efflux Pump, LexABC, Contributes to Self-Resistance of the Phenazine Di-N-Oxide Natural Product Myxin in *Lysobacter antibioticus. Frontiers in Microbiology* **12**, doi:10.3389/fmicb.2021.618513 (2021).

67 Chang, C.-Y. *et al.* Resistance to Enediyne Antitumor Antibiotics by Sequestration. *Cell Chemical Biology* **25**, 1075-1085.e1074, doi:https://doi.org/10.1016/j.chembiol.2018.05.012 (2018).

68 Panter, F. *et al.* Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chemical Science* **9**, 4898-4908, doi:10.1039/C8SC01325J (2018).

69 Feng, L. *et al.* The pentapeptide-repeat protein, MfpA, interacts with mycobacterial DNA gyrase as a DNA T-segment mimic. *Proceedings of the National Academy of Sciences* **118**, e2016705118, doi:doi:10.1073/pnas.2016705118 (2021).

70 Mazurek, Ł. *et al.* Pentapeptide repeat protein QnrB1 requires ATP hydrolysis to rejuvenate poisoned gyrase complexes. *Nucleic Acids Research* **49**, 1581-1596, doi:10.1093/nar/gkaa1266 (2021).

71 Mak, S. *et al.* The expression of antibiotic resistance genes in antibiotic-producing bacteria. *Molecular Microbiology* **93**, 391-402, doi:https://doi.org/10.1111/mmi.12689 (2014).

72 Peterson, R. M. *et al.* Mechanisms of self-resistance in the platensimycin- and platencin-producing *Streptomyces platensis* MA7327 and MA7339 strains. *Chemistry & biology* **21**, 389-397, doi:10.1016/j.chembiol.2014.01.005 (2014).

73 Peterson, E. & Kaur, P. Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens. *Frontiers in Microbiology* **9**, doi:10.3389/fmicb.2018.02928 (2018).

74 Ogawara, H. Self-resistance in *Streptomyces*, with Special Reference to β-Lactam Antibiotics. *Molecules* **21**, doi:10.3390/molecules21050605 (2016).

75 Alanjary, M. *et al.* The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Research* **45**, W42-W48, doi:10.1093/nar/gkx360 (2017).

76 Fang, P. *et al.* Structural basis for full-spectrum inhibition of translational functions on a tRNA synthetase. *Nature Communications* **6**, 6402, doi:10.1038/ncomms7402(2015).

77 Sankaranarayanan, R. *et al.* The structure of threonyl-tRNA synthetase-tRNA^{Thr} complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **97**, 371-381 (1999).

78 Chen, Y. *et al.* A threonyl-tRNA synthetase-like protein has tRNA aminoacylation and editing activities. *Nucleic Acids Research* **46**, 3643-3656, doi:10.1093/nar/gky211 (2018).

79 Dock-Bregeon, A.-C. *et al.* Achieving Error-Free Translation: The Mechanism of Proofreading of Threonyl-tRNA Synthetase at Atomic Resolution. *Molecular Cell* **16**, 375-386, doi:https://doi.org/10.1016/j.molcel.2004.10.002 (2004).

80 Lue, S. W. & Kelley, S. O. An aminoacyl-tRNA synthetase with a defunct editing site. *Biochemistry* **44**, 3010-3016, doi:10.1021/bi047901v (2005).

81 Wei, W. *et al.* Pretransfer Editing in Threonyl-tRNA Synthetase: Roles of Differential Solvent Accessibility and Intermediate Stabilization. *ACS Catalysis* **7**, 3102-3112, doi:10.1021/acscatal.6b03051 (2017).

82 Chaliotis, A. *et al.* The complex evolutionary history of aminoacyl-tRNA synthetases. *Nucleic Acids Res* **45**, 1059-1068, doi:10.1093/nar/gkw1182 (2017).

83 Li, L. *et al.*. Aminoacylating Urzymes Challenge the RNA World Hypothesis. *Journal of Biological Chemistry* **288**, 26856-26863, doi:https://doi.org/10.1074/jbc.M113.496125 (2013).

84 de Farias, S. T. *et al.* Evolution of transfer RNA and the origin of the translation system. *Frontiers in Genetics* **5**, doi:10.3389/fgene.2014.00303 (2014).

85 Kaiser, F. *et al.* The structural basis of the genetic code: amino acid recognition by aminoacyl-tRNA synthetases. *Scientific Reports* **10**, 12647, doi:10.1038/s41598-020-69100-0 (2020).

86 O'Donoghue, P. & Luthey-Schulten, Z. On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol Mol Biol Rev* **67**, 550-573, doi:10.1128/mmbr.67.4.550-573.2003 (2003).

87 Ribas de Pouplana, L. Chapter Two - The evolution of aminoacyl-tRNA synthetases: From dawn to LUCA in *The Enzymes* Vol. 48 (eds Lluís Ribas de Pouplana & Laurie S. Kaguni) 11-37 (Academic Press, 2020).

88 Carter, C. W. *et al.* High-Resolution, Multidimensional Phylogenetic Metrics Identify Class I Aminoacyl-tRNA Synthetase Evolutionary Mosaicity and Inter-modular 'Coupling. *bioRxiv*, 2020.2004.2009.033712, doi:10.1101/2020.04.09.033712 (2020).

89 Rodin, S. N. & Rodin, A. S. Partitioning of Aminoacyl-tRNA Synthetases in Two Classes Could Have Been Encoded in a Strand-Symmetric RNA World. *DNA and Cell Biology* **25**, 617-626, doi:10.1089/dna.2006.25.617 (2006).

90 Carter, C. W. What RNA World? Why a Peptide/RNA Partnership Merits Renewed Experimental Attention. *Life* **5**, 294-320 (2015).

91 Perona, J. J. & Hadd, A. Structural Diversity and Protein Engineering of the Aminoacyl-tRNA Synthetases. *Biochemistry* **51**, 8705-8729, doi:10.1021/bi301180x (2012).

92 Krahn, N. *et al*. Diversification of aminoacyl-tRNA synthetase activities via genomic duplication. *Frontiers in Physiology* **13**, doi:10.3389/fphys.2022.983245 (2022).

93 Carter, C. W., Jr. Coding of Class I and II Aminoacyl-tRNA Synthetases. *Adv Exp Med Biol* **966**, 103-148, doi:10.1007/5584_2017_93 (2017).

94 Ibba, M. *et al.* Substrate recognition by class I lysyl-tRNA synthetases: a molecular basis for gene displacement. *Proc Natl Acad Sci U S A* **96**, 418-423, doi:10.1073/pnas.96.2.418 (1999).

95 Silvian, L. F. *et al*. Insights into editing from an ile-tRNA synthetase structure with $tRNA^{Ile}$ and mupirocin. *Science* **285**, 1074-1077 (1999).

96 Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* (2015).

97 Sankaranarayanan, R. *et al.* Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. *Nat Struct Biol* **7**, 461-465, doi:10.1038/75856 (2000).

98 Guo, M. *et al*. New functions of aminoacyl-tRNA synthetases beyond translation. *Nat Rev Mol Cell Biol* **11**, 668-674, doi:10.1038/nrm2956 (2010).

99 Williams, T. F. *et al*. Secreted Threonyl-tRNA synthetase stimulates endothelial cell migration and angiogenesis. *Scientific Reports* **3**, 1317, doi:10.1038/srep01317 (2013).

100 Dock-Bregeon, A. *et al.* Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem. *Cell* **103**, 877-884, doi:10.1016/s0092-8674(00)00191-4 (2000).

101 Torres-Larios, A. *et al.* Structural basis of translational control by Escherichia coli threonyl tRNA synthetase. *Nat Struct Biol* **9**, 343-347, doi:10.1038/nsb789 (2002).

102 Torres-Larios, A. *et al*. Conformational movements and cooperativity upon amino acid, ATP and tRNA binding in threonyl-tRNA synthetase. *J Mol Biol* **331**, 201-211, doi:10.1016/s0022-2836(03)00719-8 (2003).

103 Sankaranarayanan, R. *et al.* The Structure of Threonyl-tRNA Synthetase-tRNA^{Thr}Complex Enlightens Its Repressor Activity and Reveals an Essential Zinc Ion in the Active Site. *Cell* **97**, 371-381, doi:10.1016/S0092-8674(00)80746-1 (1999).

104 Dwivedi, S. *et al*. A D-amino acid editing module coupled to the translational apparatus in archaea. *Nat Struct Mol Biol* **12**, 556-557, doi:10.1038/nsmb943 (2005).

105 Hussain, T. *et al.* Post-transfer editing mechanism of a D-aminoacyl-tRNA deacylase-like domain in threonyl-tRNA synthetase from archaea. *Embo j* **25**, 4152-4162, doi:10.1038/sj.emboj.7601278 (2006).

106 Shimizu, S. *et al.* Two complementary enzymes for threonylation of tRNA in crenarchaeota: crystal structure of *Aeropyrum pernix* threonyl-tRNA synthetase lacking a cis-editing domain. *J Mol Biol* **394**, 286-296, doi:10.1016/j.jmb.2009.09.018 (2009).

107 Hussain, T. *et al*. Mechanistic insights into cognate substrate discrimination during proofreading in translation. *Proc Natl Acad Sci U S A* **107**, 22117-22121, doi:10.1073/pnas.1014299107 (2010).

108 Ling, J. *et al.* Yeast mitochondrial threonyl-tRNA synthetase recognizes tRNA isoacceptors by distinct mechanisms and promotes CUN codon reassignment. *Proc Natl Acad Sci U S A* **109**, 3281-3286, doi:10.1073/pnas.1200109109 (2012).

109 Ling, J. *et al*. The mechanism of pre-transfer editing in yeast mitochondrial threonyl-tRNA synthetase. *J Biol Chem* **287**, 28518-28525, doi:10.1074/jbc.M112.372920 (2012).

110 Fang, P. *et al.* Structural basis for full-spectrum inhibition of translational functions on a tRNA synthetase. *Nature Communications* **6**, 6402, doi:10.1038/ncomms7402 (2015).

111 Mirando, A. C. *et al.* Aminoacyl-tRNA synthetase dependent angiogenesis revealed by a bioengineered macrolide inhibitor. *Sci Rep* **5**, 13160, doi:10.1038/srep13160 (2015).

112 Holman, K. M. *et al*. The crystal structure of yeast mitochondrial ThrRS in complex with the canonical threonine tRNA. *Nucleic Acids Res* **44**, 1428-1439, doi:10.1093/nar/gkv1501 (2016).

113 Jeong, S. J. *et al.* A threonyl-tRNA synthetase-mediated translation initiation machinery. *Nature Communications* **10**, 1357, doi:10.1038/s41467-019-09086-0 (2019).

114 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).

115 Minajigi, A. & Francklyn, C. S. RNA-assisted catalysis in a protein enzyme: The 2'-hydroxyl of tRNA^{Thr} A76 promotes aminoacylation by threonyl-tRNA synthetase. *Proceedings of the National Academy of Sciences* **105**, 17748-17753, doi:doi:10.1073/pnas.0804247105 (2008).

116 Angelo, L. S. & Kurzrock, R. Vascular Endothelial Growth Factor and Its Relationship to Inflammatory Mediators. *Clinical Cancer Research* **13**, 2825-2830, doi:10.1158/1078-0432.CCR-06-2416 (2007).

117 Wolf, Y. I. *et al*. Evolution of aminoacyl-tRNA synthetases--analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res* **9**, 689-710 (1999).

118 Dai, C. *et al*.. A non-translational role of threonyl-tRNA synthetase in regulating JNK signaling during myogenic differentiation. *The FASEB Journal* **35**, e21948, doi:https://doi.org/10.1096/fj.202101094R (2021).

119 Saxton, R. A. & Sabatini, D. M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **168**, 960-976, doi:10.1016/j.cell.2017.02.004 (2017).

120 Kim, S. H. *et al.* Mitochondrial Threonyl-tRNA Synthetase TARS2 Is Required for Threonine-Sensitive mTORC1 Activation. *Mol Cell* **81**, 398-407.e394, doi:10.1016/j.molcel.2020.11.036 (2021).

121 Glick, D. *et al*. Autophagy: cellular and molecular mechanisms. *J Pathol* **221**, 3-12, doi:10.1002/path.2697 (2010).

122 Um, S. H. *et al.* Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity. *Nature* **431**, 200-205, doi:10.1038/nature02866 (2004).

123 Castets, P. *et al.* Sustained activation of mTORC1 in skeletal muscle inhibits constitutive and starvation-induced autophagy and causes a severe, late-onset myopathy. *Cell Metab* **17**, 731-744, doi:10.1016/j.cmet.2013.03.015 (2013).

124 Polak, P. *et al.* Adipose-specific knockout of raptor results in lean mice with enhanced mitochondrial respiration. *Cell Metab* **8**, 399-410, doi:10.1016/j.cmet.2008.09.003 (2008).

125 Spilman, P. *et al.* Inhibition of mTOR by rapamycin abolishes cognitive deficits and reduces amyloid-beta levels in a mouse model of Alzheimer's disease. *PLoS One* **5**, e9979, doi:10.1371/journal.pone.0009979 (2010).

126 Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55-61, doi:10.1038/nature10912 (2012).

127 Chen, C. *et al*. mTOR regulation and therapeutic rejuvenation of aging hematopoietic stem cells. *Sci Signal* **2**, ra75, doi:10.1126/scisignal.2000559 (2009).

128 Azad, A. K. *et al*. Role of nuclear pools of aminoacyl-tRNA synthetases in tRNA nuclear export. *Mol Biol Cell* **12**, 1381-1392, doi:10.1091/mbc.12.5.1381 (2001).

129 Steiner-Mosonyi, M. & Mangroo, D. The nuclear tRNA aminoacylation-dependent pathway may be the principal route used to export tRNA from the nucleus in *Saccharomyces cerevisiae*. *Biochemical Journal* **378**, 809-816 (2004).

130 Telonis, A. G. *et al*. Nuclear and mitochondrial tRNA-lookalikes in the human genome. *Front Genet* **5**, 344, doi:10.3389/fgene.2014.00344 (2014).

131 Hoser, S. M. *et al.* Intronic tRNAs of mitochondrial origin regulate constitutive and alternative splicing. *Genome Biology* **21**, 299, doi:10.1186/s13059-020-02199-6 (2020).

132 Chen, M. *et al.* Cross-editing by a tRNA synthetase allows vertebrates to abundantly express mischargeable tRNA without causing mistranslation. *Nucleic Acids Res* **48**, 6445-6457, doi:10.1093/nar/gkaa469 (2020).

133 Hyeon, D. Y. *et al.* Evolution of the multi-tRNA synthetase complex and its role in cancer. *Journal of Biological Chemistry* **294**, 5340-5351, doi:https://doi.org/10.1074/jbc.REV118.002958 (2019).

134 Kim, K. *et al.* Reinvestigation of aminoacyl-tRNA synthetase core complex by affinity purification-mass spectrometry reveals TARSL2 as a potential member of the complex. *PLoS One* **8**, e81734, doi:10.1371/journal.pone.0081734 (2013).

135 Park, S. J., Ahn, H. S., Kim, J. S. & Lee, C. Evaluation of Multi-tRNA Synthetase Complex by Multiple Reaction Monitoring Mass Spectrometry Coupled with Size Exclusion Chromatography. *PLoS One* **10**, e0142253, doi:10.1371/journal.pone.0142253 (2015).

136 Zhou, X. L. *et al.* Newly acquired N-terminal extension targets threonyl-tRNA synthetase-like protein into the multiple tRNA synthetase complex. *Nucleic Acids Res* **47**, 8662-8674, doi:10.1093/nar/gkz588 (2019).

137 Galindo-Feria, A. S. *et al*. Aminoacyl-tRNA Synthetases: On Anti-Synthetase Syndrome and Beyond. *Frontiers in Immunology* **13**, doi:10.3389/fimmu.2022.866087 (2022).

138 Labirua-Iturburu, A. *et al.* Anti-PL-7 (Anti-Threonyl-tRNA Synthetase) Antisynthetase Syndrome: Clinical Manifestations in a Series of Patients From a European Multicenter Study (EUMYONET) and Review of the Literature. *Medicine* **91** (2012).

139 Hervier, B. *et al.* Antisynthetase syndrome positive for anti-threonyl-tRNA synthetase (anti-PL7) antibodies. *European Respiratory Journal* **37**, 714-717, doi:10.1183/09031936.00104310 (2011).

140 Mathews, M. B. *et al*. Anti-threonyl-tRNA synthetase, a second myositis-related autoantibody. *Journal of Experimental Medicine* **160**, 420-434, doi:10.1084/jem.160.2.420 (1984).

141 Theil, A. F. *et al.* Bi-allelic TARS Mutations Are Associated with Brittle Hair Phenotype. *The American Journal of Human Genetics* **105**, 434-440, doi:https://doi.org/10.1016/j.ajhg.2019.06.017 (2019).

142 Botta, E. *et al.* Protein instability associated with AARS1 and MARS1 mutations causes trichothiodystrophy. *Hum Mol Genet* **30**, 1711-1720, doi:10.1093/hmg/ddab123 (2021).

143 Hu, J. *et al.* Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat Biotechnol* **31**, 522-529, doi:10.1038/nbt.2530 (2013).

144 Stephens, P. J. *et al.* Whole exome sequencing of adenoid cystic carcinoma. *J Clin Invest* **123**, 2965-2968, doi:10.1172/jci67201 (2013).

145 Lopes-Coelho, F. *et al*. Anti-Angiogenic Therapy: Current Challenges and Future Perspectives. *Int J Mol Sci* **22**, doi:10.3390/ijms22073765 (2021).

146 Ryder, M. H. *et al*.. Agrocinopine A, a tumor-inducing plasmid-coded enzyme product, is a phosphodiester of sucrose and L-arabinose. *J Biol Chem* **259**, 9704-9710 (1984).

147 Kim, J.-G. *et al.* Bases of biocontrol: Sequence predicts synthesis and mode of action of agrocin 84, the Trojan Horse antibiotic that controls crown gall. *Proceedings of the National Academy of Sciences* **103**, 8846-8851, doi:doi:10.1073/pnas.0602965103 (2006).

148 Travin, D. Y. *et al*. Natural Trojan horse inhibitors of aminoacyl-tRNA synthetases. *RSC Chem Biol* **2**, 468-485, doi:10.1039/d0cb00208a (2021).

149 Lin, Z. *et al.* Total synthesis and antimicrobial evaluation of natural albomycins against clinical pathogens. *Nature Communications* **9**, 3445, doi:10.1038/s41467-018-05821-1 (2018).

150 Pramanik, A. & Braun, V. Albomycin uptake via a ferric hydroxamate transport system of *Streptococcus pneumoniae* R6. *J Bacteriol* **188**, 3878-3886, doi:10.1128/jb.00205-06 (2006).

151 Gause, G. F. Recent studies on albomycin, a new antibiotic. *Br Med J* **2**, 1177-1179, doi:10.1136/bmj.2.4949.1177 (1955).

152 Francklyn, C. S. & Mullen, P. Progress and challenges in aminoacyl-tRNA synthetase-based therapeutics. *The Journal of biological chemistry* **294**, 5365-5385, doi:10.1074/jbc.REV118.002956 (2019).

153 Berger, J. *et al*. Borrelidin, a new antibiotic with antiborrelia activity and penicillin enhancement properties. *Arch Biochem* **22**, 476-478 (1949).

154 Gao, Y.-M. *et al.* Borrelidin, a Potent Antifungal Agent: Insight into the Antifungal Mechanism against *Phytophthora sojae*. *Journal of Agricultural and Food Chemistry* **60**, 9874-9881, doi:10.1021/jf302857x (2012).

155 Liu, C. X. *et al.* Antifungal activity of borrelidin produced by a *Streptomyces* strain isolated from soybean. *J Agric Food Chem* **60**, 1251-1257, doi:10.1021/jf2044982 (2012).

156 Lumb, M. *et al*. Isolation of Vivomycin and Borrelidin, Two Antibiotics with Anti-Viral Activity, from a Species of *Streptomyces* (C2989). *Nature* **206**, 263-265, doi:10.1038/206263a0 (1965).

157 Novoa, E. M. *et al.* Analogs of natural aminoacyl-tRNA synthetase inhibitors clear malaria *in vivo*. *Proceedings of the National Academy of Sciences* **111**, E5508-E5517, doi:doi:10.1073/pnas.1405994111 (2014).

158 Ishiyama, A. *et al.* Borrelidin, a potent antimalarial: stage-specific inhibition profile of synchronized cultures of *Plasmodium falciparum*. *The Journal of Antibiotics* **64**, 381-384, doi:10.1038/ja.2011.6 (2011).

159 Cao, Z. *et al*. Isolation of Borrelidin as a Phytotoxic Compound from a Potato Pathogenic *Streptomyces* Strain. *Bioscience, Biotechnology, and Biochemistry* **76**, 353-357, doi:10.1271/bbb.110799 (2012).

160 Moss, S. J. *et al.* Biosynthesis of the angiogenesis inhibitor borrelidin: directed biosynthesis of novel analogues. *Chemical Communications*, 2341-2343, doi:10.1039/B602931K (2006).

161 Ruan, B. *et al.* A unique hydrophobic cluster near the active site contributes to differences in borrelidin inhibition among threonyl-tRNA synthetases. *J Biol Chem* **280**, 571-577, doi:10.1074/jbc.M411039200 (2005).

162 Stefanska, A. L. *et al.* SB-219383, a novel tyrosyl tRNA synthetase inhibitor from a *Micromonospora* sp. I. Fermentation, isolation and properties. *J Antibiot (Tokyo)* **53**, 345-350, doi:10.7164/antibiotics.53.345 (2000).

163 Kanamaru, T. *et al. In vitro* and *in vivo* antibacterial activities of TAK-083, an agent for treatment of Helicobacter pylori infection. *Antimicrob Agents Chemother* **45**, 2455-2459, doi:10.1128/aac.45.9.2455-2459.2001 (2001).

164 (Authors not listed) Studies on a new antibiotic--Chuangxinmycin. *Sci Sin* **20**, 106-112 (1977).

165 Oki, T. *et al*. Cispentacin, a new antifungal antibiotic. II. *In vitro* and *in vivo* antifungal activities. *J Antibiot (Tokyo)* **42**, 1756-1762, doi:10.7164/antibiotics.42.1756 (1989).

166 Moriguchi, T. *et al.* First synthesis and anticancer activity of phosmidosine and its related compounds. *J Org Chem* **67**, 3290-3300, doi:10.1021/jo016176g (2002).

167 Dai, Z. *et al.* Effects of Antimicrobial Peptide Microcin C7 on Growth Performance, Immune and Intestinal Barrier Functions, and Cecal Microbiota of Broilers. *Frontiers in Veterinary Science* **8**, doi:10.3389/fvets.2021.813629 (2022).

168 Shin, J. *et al.* Borrelidin from Saltern-Derived Halophilic *Nocardiopsis* sp. Dissociates Amyloid-β and Tau Fibrils. *J Alzheimers Dis Rep* **5**, 7-13, doi:10.3233/adr-200247 (2021).

169 Robbins, C. E. *et al.* Cytosolic and mitochondrial tRNA synthetase inhibitors increase lifespan in a GCN4/atf-4-dependent manner. *iScience* **25**, 105410, doi:10.1016/j.isci.2022.105410 (2022).

170 Olano, C. *et al.* Biosynthesis of the Angiogenesis Inhibitor Borrelidin by *Streptomyces parvulus* Tü4055: Cluster Analysis and Assignment of Functions. *Chemistry & Biology* **11**, 87-97, doi:https://doi.org/10.1016/j.chembiol.2003.12.018 (2004).

171 WELLS, J. S. *et al*. Obafluorin, a Novel β-Lactone Produced by *Pseudomonas fluorescens*. *The Journal of Antibiotics* **37**, 802-803, doi:https://doi.org/10.7164/antibiotics.37.802 (1984).

172 Scott, T. A. *et al*. An L-threonine transaldolase is required for L-threo-β-hydroxy-α-amino acid assembly during obafluorin biosynthesis. *Nature Communications* **8**, 15935, doi:10.1038/ncomms15935 (2017).

173 Schaffer, J. E. *et al*. β-Lactone formation during product release from a nonribosomal peptide synthetase. *Nat Chem Biol* **13**, 737-744, doi:10.1038/nchembio.2374 (2017).

174 Scott, T. A. *et al.* Immunity-Guided Identification of Threonyl-tRNA Synthetase as the Molecular Target of Obafluorin, a β-Lactone Antibiotic. *ACS Chemical Biology* **14**, 2663-2671, doi:10.1021/acschembio.9b00590 (2019).

175 Tymiak, A. A. *et al*. Structure of obafluorin: an antibacterial β-lactone from *Pseudomonas fluorescens. The Journal of Organic Chemistry* **50**, 5491-5495, doi:10.1021/jo00350a010 (1985).

176 Guo, J. *et al.* Discovery of novel tRNA-amino acid dual-site inhibitors against threonyl-tRNA synthetase by fragment-based target hopping. *Eur J Med Chem* **187**, 111941, doi:10.1016/j.ejmech.2019.111941 (2020).

177 Teng, M. *et al.* Identification of bacteria-selective threonyl-tRNA synthetase substrate inhibitors by structure-based design. *J Med Chem* **56**, 1748-1760, doi:10.1021/jm301756m (2013).

178 Cai, Z. *et al.* Design, Synthesis, and Proof-of-Concept of Triple-Site Inhibitors against Aminoacyl-tRNA Synthetases. *J Med Chem* **65**, 5800-5820, doi:10.1021/acs.jmedchem.2c00134 (2022).

179 Guo, J. *et al.* Structure-guided optimization and mechanistic study of a class of quinazolinone-threonine hybrids as antibacterial ThrRS inhibitors. *Eur J Med Chem* **207**, 112848, doi:10.1016/j.ejmech.2020.112848 (2020).

180 Kieser, T. *et al*. *Practical Streptomyces Genetics*. (John Innes Foundation, 2000).

181 Stover, C. K. *et al.* New use of BCG for recombinant vaccines. *Nature* **351**, 456-460, doi:10.1038/351456a0 (1991).

182 Gust, B. *et al.* Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces. Adv Appl Microbiol* **54**, 107-128, doi:10.1016/s0065-2164(04)54004-2 (2004).

183 Zhang, C. *et al.* Discovery of okilactomycin and congeners from *Streptomyces scabrisporus* by antisense differential sensitivity assay targeting ribosomal protein S4. *The Journal of Antibiotics* **62**, 55-61, doi:10.1038/ja.2008.8 (2009).

184 Hong, H. J. *et al*. The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. *J Biol Chem* **280**, 13055-13061, doi:10.1074/jbc.M413801200 (2005).

185 Bush, M. J. *et al*. Hyphal compartmentalization and sporulation in *Streptomyces* require the conserved cell division protein SepX. *Nat Commun* **13**, 71, doi:10.1038/s41467-021-27638-1 (2022).

186 Rosenberg, A. H. *et al.* Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**, 125-135, doi:https://doi.org/10.1016/0378-1119(87)90165-X (1987).

187 Ruiz, N. *et al*. Chemical conditionality: a genetic strategy to probe organelle assembly. *Cell* **121**, 307-317, doi:10.1016/j.cell.2005.02.014 (2005).

188 Widdick, D. A. *et al.* Cloning and engineering of the cinnamycin biosynthetic gene cluster from *Streptomyces cinnamoneus cinnamoneus* DSM 40005. *Proc Natl Acad Sci U S A* **100**, 4316-4321, doi:10.1073/pnas.0230516100 (2003).

189 Gomez-Escribano, J. P. *et al. Streptomyces venezuelae* NRRL B-65442: genome sequence of a model strain used to study morphological differentiation in filamentous actinobacteria. *Journal of Industrial Microbiology and Biotechnology* **48**, doi:10.1093/jimb/kuab035 (2021).

190 Gomez-Escribano, J. P. & Bibb, M. J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol* **4**, 207-215, doi:10.1111/j.1751-7915.2010.00219.x (2011).

191 MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61-68, doi:10.1016/0378-1119(92)90603-m (1992).

192 Olano, C. *et al.* Biosynthesis of the Angiogenesis Inhibitor Borrelidin by *Streptomyces parvulus* Tü4055: Cluster Analysis and Assignment of Functions. *Chemistry & Biology* **11**, 87-97, doi:10.1016/j.chembiol.2003.12.018 (2004).

193 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).

194 Petre, B. *et al*. Protein-Protein Interaction Assays with Effector-GFP Fusions in *Nicotiana benthamiana. Methods Mol Biol* **1659**, 85-98, doi:10.1007/978-1-4939-7249-4_8 (2017).

195 Bijelic, A. & Rompel, A. Polyoxometalates: more than a phasing tool in protein crystallography. *ChemTexts* **4**, 10, doi:10.1007/s40828-018-0064-1 (2018).

196 Emsley, P. *et al*. Features and development of Coot. *Acta Crystallographica Section D* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

197 Potterton, L. *et al.* CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D* **74**, 68-84, doi:doi:10.1107/S2059798317016035 (2018).

198 Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Communications Biology* **4**, 874, doi:10.1038/s42003-021-02399-1 (2021).

199 Croll, T. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D* **74**, 519-530, doi:doi:10.1107/S2059798318002425 (2018).

200 Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113, doi:10.1186/1471-2105-5-113 (2004).

201 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

202 Kumar, S. *et al*. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol iol Evol* **35**, 1547-1549, doi:10.1093/molbev/msy096 (2018).

203 Miller, M. A. *et al*. in *2010 Gateway Computing Environments Workshop (GCE).* 1-8.

204 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

205 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**, W293-W296, doi:10.1093/nar/gkab301 (2021).

206 Blin, K. *et al.* antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research* **41**, W204-W212, doi:10.1093/nar/gkt449 (2013).

207 Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* **43**, W237-W243, doi:10.1093/nar/gkv437 (2015).

208 Blin, K. *et al.* antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research* **45**, W36-W41, doi:10.1093/nar/gkx319 (2017).

209 Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* **47**, W81-W87, doi:10.1093/nar/gkz310 (2019).

210 Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* **49**, W29-W35, doi:10.1093/nar/gkab335 (2021).

211 Kim, Y. *et al.* Aminoacyl-tRNA synthetase inhibition activates a pathway that branches from the canonical amino acid response in mammalian cells. *Proceedings of the National Academy of Sciences* **117**, 8900-8911, doi:doi:10.1073/pnas.1913788117 (2020).

212 Bhikshapathi, R. *et al*. Anti-HIV, anti-tubercular and mutagenic activities of borrelidin. *Indian Journal of Biotechnology* **9**, 265-270 (2010).

213 Olano, C. *et al.* Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: insights into nitrile formation. *Mol Microbiol* **52**, 1745-1756, doi:10.1111/j.1365-2958.2004.04090.x (2004).

214 Schulze, C. J. *et al.* Borrelidin B: Isolation, Biological Activity, and Implications for Nitrile Biosynthesis. *Journal of Natural Products* **77**, 2570-2574, doi:10.1021/np500727g (2014).

215 Sidhu, A. *et al.* Borrelidin Induces the Unfolded Protein Response in Oral Cancer Cells and Chop-Dependent Apoptosis. *ACS Med Chem Lett* **6**, 1122-1127, doi:10.1021/acsmedchemlett.5b00133 (2015).

216 Wakabayashi, T. *et al.* Borrelidin is an angiogenesis inhibitor; disruption of angiogenic capillary vessels in a rat aorta matrix culture model. *J Antibiot (Tokyo)* **50**, 671-676, doi:10.7164/antibiotics.50.671 (1997).

217 Kim, J. *et al.* Borrelidins C-E: New Antibacterial Macrolides from a Saltern-Derived Halophilic *Nocardiopsis* sp. *Mar Drugs* **15**, 166, doi:10.3390/md15060166 (2017).

218 Sun, J. *et al.* Borrelidins F–I, cytotoxic and cell migration inhibiting agents from mangrove-derived *Streptomyces rochei* SCSIO ZJ89. *Bioorganic & Medicinal Chemistry* **26**, 1488-1494, doi:https://doi.org/10.1016/j.bmc.2018.01.010 (2018).

219 Li, Y. *et al*. Comprehensive optimization of precursor-directed production of BC194 by *Streptomyces rochei* MB037 derived from the marine sponge *Dysidea arenaria. Applied Microbiology and Biotechnology* **102**, 7865-7875, doi:10.1007/s00253-018-9237-5 (2018).

220 Saisivam, S. *et al*. Isolation of borrelidin from *Streptomyces californicus*- An Indian soil isolate. *Indian Journal of Biotechnology*. **7**(3), (2008).

221 Nass, G. & Poralla, K. Genetics of borrelidin resistant mutants of *Saccharomyces cerivisiae* and properties of their threonyl-tRNA-synthetase. *Mol Gen Genet* **147**, 39-43, doi:10.1007/bf00337933 (1976).

222 Gabrielson, J. *et al.* Evaluation of redox indicators and the use of digital scanners and spectrophotometer for quantification of microbial growth in microplates. *Journal of Microbiological Methods* **50**, 63-73, doi:https://doi.org/10.1016/S0167-7012(02)00011-8 (2002).

223 Henke, W. *et al*. Betaine Improves the PCR Amplification of GC-Rich DNA Sequences. *Nucleic Acids Research* **25**, 3957-3958, doi:10.1093/nar/25.19.3957 (1997).

224 Georgakopoulos-Soares, I. *et al*. High-throughput techniques enable advances in the roles of DNA and RNA secondary structures in transcriptional and post-transcriptional gene regulation. *Genome Biology* **23**, 159, doi:10.1186/s13059-022-02727-6 (2022).

225 Glazebrook, M. A. *et al*. Sporulation of *Streptomyces venezuelae* in submerged cultures. *J Gen Microbiol* **136**, 581-588, doi:10.1099/00221287-136-3-581 (1990).

226 Yagüe, P. *et al*. Pre-sporulation stages of *Streptomyces* differentiation: state-of-the-art and future perspectives. *FEMS Microbiol Lett* **342**, 79-88, doi:10.1111/1574-6968.12128 (2013).

227 Daza, A. *et al*. Sporulation of several species of *Streptomyces* in submerged cultures after nutritional downshift. *J Gen Microbiol* **135**, 2483-2491, doi:10.1099/00221287-135-9-2483 (1989).

228 Kendrick, K. E. & Ensign, J. C. Sporulation of *Streptomyces griseus* in submerged culture. *J Bacteriol* **155**, 357-366, doi:10.1128/jb.155.1.357-366.1983 (1983).

229 Sissler, M. *et al.* An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis. *Proc Natl Acad Sci U S A* **96**, 8985-8990, doi:10.1073/pnas.96.16.8985 (1999).

230 Bovee, M. L. *et al*.Induced fit and kinetic mechanism of adenylation catalyzed by *Escherichia coli* threonyl-tRNA synthetase. *Biochemistry* **42**, 15102-15113, doi:10.1021/bi0355701 (2003).

231 Kato, D.-i. *et al.* Enantiodifferentiation of ketoprofen by Japanese firefly luciferase from *Luciola lateralis. Journal of Molecular Catalysis B: Enzymatic* **69**, 140-146, doi:https://doi.org/10.1016/j.molcatb.2011.01.008 (2011).

232 Moreau, C. *et al.* Structure–Activity Relationship of Adenosine 5′-diphosphoribose at the Transient Receptor Potential Melastatin 2 (TRPM2) Channel: Rational Design of Antagonists. *Journal of Medicinal Chemistry* **56**, 10079-10102, doi:10.1021/jm401497a (2013).

233 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

234 Culp, E. J. *et al.* Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature* **578**, 582-587, doi:10.1038/s41586-020-1990-9 (2020).

235 Waglechner, N. *et al*. Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nature Microbiology* **4**, 1862-1871, doi:10.1038/s41564-019-0531-5 (2019).

236 Rigden, D. J. Archaea recruited D-Tyr-tRNATyr deacylase for editing in Thr-tRNA synthetase. *RNA* **10**, 1845-1851, doi:10.1261/rna.7115404 (2004).

237 Korencic, D. *et al.* A freestanding proofreading domain is required for protein synthesis quality control in Archaea. *Proceedings of the National Academy of Sciences* **101**, 10260-10265, doi:doi:10.1073/pnas.0403926101 (2004).

238 Valley, C. C. *et al.* The methionine-aromatic motif plays a unique role in stabilizing protein structure. *J Biol Chem* **287**, 34979-34991, doi:10.1074/jbc.M112.374504 (2012).

239 Atanasov, A. G. *et al.* Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery* **20**, 200-216, doi:10.1038/s41573-020-00114-z (2021).

240 Clark, K. *et al*. GenBank. *Nucleic Acids Res* **44**, D67-72, doi:10.1093/nar/gkv1276 (2016).

241 Yim, G. *et al*. Glycopeptide antibiotic biosynthesis. *The Journal of Antibiotics* **67**, 31-41, doi:10.1038/ja.2013.117 (2014).

242 Bauman, K. D. *et al*. Genome mining methods to discover bioactive natural products. *Natural Product Reports* **38**, 2100-2129, doi:https://doi.org/10.1039/d1np00032b (2021).

243 Sivakala, K. K. *et al.* Desert Environments Facilitate Unique Evolution of Biosynthetic Potential in *Streptomyces*. *Molecules* **26**, doi:10.3390/molecules26030588 (2021).

244 Sarmiento-Vizcaíno, A. *et al.* Natural products, including a new caboxamycin, from *Streptomyces* and other *Actinobacteria* isolated in Spain from storm clouds transported by Northern winds of Arctic origin. *Frontiers in Chemistry* **10**, doi:10.3389/fchem.2022.948795 (2022).

245 Qin, Z. *et al.* Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African *Tetraponera* plant-ants. *Chemical Science* **8**, 3218-3227, doi:10.1039/C6SC04265A (2017).

246 Holmes, N. A. *et al.* Complete genome sequence of *Streptomyces formicae* KY5, the formicamycin producer. *Journal of biotechnology* **265**, 116-118, doi:10.1016/j.jbiotec.2017.11.011 (2018).

247 Young, T. P. *et al*. Ants on swollen-thorn acacias: species coexistence in a simple system. *Oecologia* **109**, 98-107, doi:10.1007/s004420050063 (1996).

248 Palmer, T. M. *et al.* Breakdown of an ant-plant mutualism follows the loss of large herbivores from an African savanna. *Science (New York, N.Y.)* **319**, 192-195, doi:10.1126/science.1151579 (2008).

249 Qin, Z. *et al*. A role for antibiotic biosynthesis monooxygenase domain proteins in fidelity control during aromatic polyketide biosynthesis. *Nature Communications* **10**, 3611, doi:10.1038/s41467-019-11538-6 (2019).

250 Qin, Z. *et al.* Formicamycin biosynthesis involves a unique reductive ring contraction. *Chemical Science* **11**, 8125-8131, doi:10.1039/D0SC01712D (2020).

251 Devine, R. *et al.* Re-wiring the regulation of the formicamycin biosynthetic gene cluster to enable the development of promising antibacterial compounds. *Cell Chemical Biology* **28**, 515-523.e515, doi:https://doi.org/10.1016/j.chembiol.2020.12.011 (2021).

252 Singh, S. K. *et al*.Treponemycin, a nitrile antibiotic active against *Treponema hyodysenteriae. Antimicrob Agents Chemother* **27**, 239-245, doi:10.1128/aac.27.2.239 (1985).

253 Yassien, M. A. *et al*.. Anti-Tuberculous Activity of Treponemycin Produced by a *Streptomyces* Strain MS-6-6 Isolated from Saudi Arabia. *Molecules* **20**, 2576-2590 (2015).

254 Liu, C. *et al. Streptomyces heilongjiangensis* sp. nov., a novel actinomycete that produces borrelidin isolated from the root surface of soybean [*Glycine max* (L.) Merr]. *Int J Syst Evol Microbiol* **63**, 1030-1036, doi:10.1099/ijs.0.041483-0 (2013).

255 Hassan, R. *et al*. Quorum Sensing Inhibiting Activity of *Streptomyces coelicoflavus* Isolated from Soil. *Frontiers in Microbiology* **7**, doi:10.3389/fmicb.2016.00659 (2016).

256 Vishwakarma, P. *et al*. Phylogenetic and conservation analyses of MFS transporters. *3 Biotech* **8**, 462, doi:10.1007/s13205-018-1476-8 (2018).

257 Liu, R. *et al*. Research advances of *Tetrasphaera* in enhanced biological phosphorus removal: A review. *Water Res* **166**, 115003, doi:10.1016/j.watres.2019.115003 (2019).

258 Liu, L. *et al.* Charting the complexity of the activated sludge microbiome through a hybrid sequencing strategy. *Microbiome* **9**, 205, doi:10.1186/s40168-021-01155-1 (2021).

259 Sánchez-Navarro, R. *et al.* Long-Read Metagenome-Assembled Genomes Improve Identification of Novel Complete Biosynthetic Gene Clusters in a Complex Microbial Activated Sludge Ecosystem. *mSystems*, e0063222, doi:10.1128/msystems.00632-22 (2022).

260 Repka, L. M. *et al*. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chemical Reviews* **117**, 5457-5520, doi:10.1021/acs.chemrev.6b00591 (2017).

261 Wang, H. & van der Donk, W. A. Biosynthesis of the class III lantipeptide catenulipeptin. *ACS Chem Biol* **7**, 1529-1535, doi:10.1021/cb3002446 (2012).

262 Haid, S. *et al.* Labyrinthopeptin A1 and A2 efficiently inhibit cell entry of hRSV isolates. *European Respiratory Journal* **50**, PA4124, doi:10.1183/1393003.congress-2017.PA4124 (2017).

263 Férir, G. *et al.* The lantibiotic peptide labyrinthopeptin A1 demonstrates broad anti-HIV and anti-HSV activity with potential for microbicidal applications. *PLoS One* **8**, e64010, doi:10.1371/journal.pone.0064010 (2013).

264 Prochnow, H. *et al.* Labyrinthopeptins Exert Broad-Spectrum Antiviral Activity through Lipid-Binding-Mediated Virolysis. *J Virol* **94**, doi:10.1128/jvi.01471-19 (2020).

265 Awakawa, T. *et al.* Salinipyrone and Pacificanone Are Biosynthetic By-products of the Rosamicin Polyketide Synthase. *Chembiochem* **16**, 1443-1447, doi:10.1002/cbic.201500177 (2015).

266 Porquier, A. *et al.* Botcinic acid biosynthesis in *Botrytis cinerea* relies on a subtelomeric gene cluster surrounded by relics of transposons and is regulated by the Zn2Cys6 transcription factor BcBoa13. *Current Genetics* **65**, 965-980, doi:10.1007/s00294-019-00952-4 (2019).

267 Rudolf, J. D. *et al*. Cytochromes P450 for natural product biosynthesis in *Streptomyces*: sequence, structure, and function. *Nat Prod Rep* **34**, 1141-1172, doi:10.1039/c7np00034k (2017).

268 Matsuda, K. *et al.* Genome Mining of Amino Group Carrier Protein-Mediated Machinery: Discovery and Biosynthetic Characterization of a Natural Product with Unique Hydrazone Unit. *ACS Chem Biol* **12**, 124-131, doi:10.1021/acschembio.6b00818 (2017).

269 Law, B. J. C. *et al.* A vitamin K-dependent carboxylase orthologue is involved in antibiotic biosynthesis. *Nature Catalysis* **1**, 977-984, doi:10.1038/s41929-018-0178-2 (2018).

270 Zabala, D. *et al.* A Flavin-Dependent Decarboxylase–Dehydrogenase–Monooxygenase Assembles the Warhead of α,β-Epoxyketone Proteasome Inhibitors. *Journal of the American Chemical Society* **138**, 4342-4345, doi:10.1021/jacs.6b01619 (2016).

271 Schorn, M. *et al.* Genetic Basis for the Biosynthesis of the Pharmaceutically Important Class of Epoxyketone Proteasome Inhibitors. *ACS Chemical Biology* **9**, 301-309, doi:10.1021/cb400699p (2014).

272 Kluska, K., Adamczyk, J. & Krężel, A. Metal binding properties, stability and reactivity of zinc fingers. *Coordination Chemistry Reviews* **367**, 18-64, doi:https://doi.org/10.1016/j.ccr.2018.04.009 (2018).

273 Hu, Z., Awakawa, T., Ma, Z. & Abe, I. Aminoacyl sulfonamide assembly in SB-203208 biosynthesis. *Nat Commun* **10**, 184, doi:10.1038/s41467-018-08093-x (2019).

274 Kozhikkadan Davis, C. *et al*. Taxifolin as dual inhibitor of Mtb DNA gyrase and isoleucyl-tRNA synthetase: *in silico* molecular docking, dynamics simulation and *in vitro* assays. *In Silico Pharmacol* **6**, 8-8, doi:10.1007/s40203-018-0045-5 (2018).

275 Shi, Z., Luo, G. & Wang, G. *Cellulomonas carbonis* sp. nov., isolated from coal mine soil. *Int J Syst Evol Microbiol* **62**, 2004-2010, doi:10.1099/ijs.0.034934-0 (2012).

276 Christopherson, M. R. *et al.* The genome sequences of *Cellulomonas fimi* and "*Cellvibrio gilvus*" reveal the cellulolytic strategies of two facultative anaerobes, transfer of "*Cellvibrio gilvus*" to the genus *Cellulomonas*, and proposal of *Cellulomonas gilvus* sp. nov. *PLoS One* **8**, e53954, doi:10.1371/journal.pone.0053954 (2013).

277 van Santen, J. A. *et al.* The Natural Products Atlas 2.0: a database of microbially-derived natural products. *Nucleic Acids Research* **50**, D1317-D1323, doi:10.1093/nar/gkab941 (2021).

278 van Santen, J. A. *et al.* The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. *ACS Central Science* **5**, 1824-1833, doi:10.1021/acscentsci.9b00806 (2019).

279 Kautsar, S. A. *et al.* MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research* **48**, D454-D458, doi:10.1093/nar/gkz882 (2019).

280 Udwary, D. W. *et al.* Significant natural product biosynthetic potential of actinorhizal symbionts of the genus frankia, as revealed by comparative genomic and proteomic analyses. *Appl Environ Microbiol* **77**, 3617-3625, doi:10.1128/aem.00038-11 (2011).

281 Brown, J. R., Zhang, J. & Hodgson, J. E. A bacterial antibiotic resistance gene with eukaryotic origins. *Curr Biol* **8**, R365-367, doi:10.1016/s0960-9822(98)70238-6 (1998).

282 Khoshnood, S. *et al.* A review on mechanism of action, resistance, synergism, and clinical implications of mupirocin against *Staphylococcus aureus. Biomedicine & Pharmacotherapy* **109**, 1809-1818, doi:https://doi.org/10.1016/j.biopha.2018.10.131 (2019).

283 Myronovskyi, M. *et al.* Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab Eng* **49**, 316-324, doi:10.1016/j.ymben.2018.09.004 (2018).

284 McKenzie, J. L. *et al.* A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria. *J Bacteriol* **194**, 2189-2204, doi:10.1128/jb.06790-11 (2012).

Appendix 1: Gene and Protein Sequences

A1.1 Gene and protein sequences used in this work

A1.1.1 EcThrRS

A1.1.1.1 Gene Sequence

ATGCCTGTTATAACTCTTCCTGATGGCAGCCAACGCCATTACGATCACGCTGTAAGCCCCATGGATGTTGCG
CTGGACATTGGTCCAGGTCTGGCGAAAGCCTGTATCGCAGGGCGCGTTAATGGCGAACTGGTTGATGCTT
GCGATCTGATTGAAAACGACGCACAACTGTCGATCATTACCGCCAAAGACGAAGAAGGTCTGGAGATCAT
TCGTCACTCCTGTGCGCACCTGTTAGGGCACGCGATTAAACAACTTTGGCCGCATACCAAAATGGCAATCG
GCCCGGTTATTGACAACGGTTTTTATTACGACGTTGATCTTGACCGCACGTTAACCCAGGAAGATGTCGAA
GCACTCGAGAAGCGGATGCATGAGCTTGCTGAGAAAAACTACGACGTCATTAAGAAGAAAGTCAGCTGGC
ACGAAGCGCGTGAAACTTTCGCCAACCGTGGGGAGAGCTACAAAGTCTCCATTCTTGACGAAAACATCGCC
CATGATGACAAGCCAGGTCTGTACTTCCATGAAGAATATGTCGATATGTGCCGCGGTCCGCACGTACCGAA
CATGCGTTTCTGCCATCATTTCAAACTAATGAAAACGGCAGGGGCTTACTGGCGTGGCGACAGCAACAACA
AAATGTTGCAACGTATTTACGGTACGGCGTGGGCAGACAAAAAAGCACTTAACGCTTACCTGCAGCGCCTG
GAAGAAGCCGCGAAACGCGACCACCGTAAAATCGGTAAACAGCTCGACCTGTACCATATGCAGGAAGAAG
CGCCGGGTATGGTATTCTGGCACAACGACGGCTGGACCATCTTCCGTGAACTGGAAGTGTTTGTTCGTTCT
AAACTGAAAGAGTACCAGTATCAGGAAGTTAAAGGTCCGTTCATGATGGACCGTGTCCTGTGGGAAAAAA
CCGGTCACTGGGACAACTACAAAGATGCAATGTTCACCACATCTTCTGAGAACCGTGAATACTGCATTAAG
CCGATGAACTGCCCGGGTCACGTACAAATTTTCAACCAGGGGCTGAAGTCTTATCGCGATCTGCCGCTGCG
TATGGCCGAGTTTGGTAGCTGCCACCGTAACGAGCCGTCAGGTTCGCTGCATGGCCTGATGCGCGTGCGT
GGATTTACCCAGGATGACGCGCATATCTTCTGTACTGAAGAACAAATTCGCGATGAAGTTAACGGATGTAT
CCGTTTAGTCTATGATATGTACAGCACTTTTGGCTTCGAGAAGATCGTCGTCAAACTCTCCACTCGTCCTGA
AAAACGTATTGGCAGCGACGAAATGTGGGATCGTGCTGAGGCGGACCTGGCGGTTGCGCTGGAAGAAAA
CAACATCCCGTTTGAATATCAACTGGGTGAAGGCGCTTTCTACGGTCCGAAAATTGAATTTACCCTGTATGA
CTGCCTCGATCGTGCATGGCAGTGCGGTACAGTACAGCTGGACTTCTCTTTGCCGTCTCGTCTGAGCGCTTC
TTATGTAGGCGAAGACAATGAACGTAAAGTACCGGTAATGATTCACCGCGCAATTCTGGGGTCGATGGAA
CGTTTCATCGGTATCCTGACCGAAGAGTTCGCTGGTTTCTTCCCGACCTGGCTTGCGCCGGTTCAGGTTGTT
ATCATGAATATTACCGATTCACAGTCTGAATACGTTAACGAATTGACGCAAAAACTATCAAATGCGGGCATT
CGTGTTAAAGCAGACTTGAGAAATGAGAAGATTGGCTTTAAAATCCGCGAGCACACTTTGCGTCGCGTCCC
ATATATGCTGGTCTGTGGTGATAAAGAGGTGGAATCAGGCAAAGTTGCCGTTCGCACCCGCCGTGGTAAA
GACCTGGGAAGCATGGACGTAAATGAAGTGATCGAGAAGCTGCAACAAGAGATTCGCAGCCGCAGTCTTA
AACAATTGGAGGAATAA

A1.1.1.2 Protein Sequence

MPVITLPDGSQRHYDHAVSPMDVALDIGPGLAKACIAGRVNGELVDACDLIENDAQLSIITAKDEEGLEIIRHSC
AHLLGHAIKQLWPHTKMAIGPVIDNGFYYDVDLDRTLTQEDVEALEKRMHELAEKNYDVIKKKVSWHEARETF
ANRGESYKVSILDENIAHDDKPGLYFHEEYVDMCRGPHVPNMRFCHHFKLMKTAGAYWRGDSNNKMLQRIY
GTAWADKKALNAYLQRLEEAAKRDHRKIGKQLDLYHMQEEAPGMVFWHNDGWTIFRELEVFVRSKLKEYQY
QEVKGPFMMDRVLWEKTGHWDNYKDAMFTTSSENREYCIKPMNCPGHVQIFNQGLKSYRDLPLRMAEFGS
CHRNEPSGSLHGLMRVRGFTQDDAHIFCTEEQIRDEVNGCIRLVYDMYSTFGFEKIVVKLSTRPEKRIGSDEMW
DRAEADLAVALEENNIPFEYQLGEGAFYGPKIEFTLYDCLDRAWQCGTVQLDFSLPSRLSASYVGEDNERKVPV
MIHRAILGSMERFIGILTEEFAGFFPTWLAPVQVVIMNITDSQSEYVNELTQKLSNAGIRVKADLRNEKIGFKIRE
HTLRRVPYMLVCGDKEVESGKVAVRTRRGKDLGSMDVNEVIEKLQQEIRSRSLKQLEE

A1.1.1.3 ΔN Protein Sequence

RDHRKIGKQLDLYHMQEEAPGMVFWHNDGWTIFRELEVFVRSKLKEYQYQEVKGPFMMDRVLWEKTGHW
DNYKDAMFTTSSENREYCIKPMNCPGHVQIFNQGLKSYRDLPLRMAEFGSCHRNEPSGSLHGLMRVRGFTQD
DAHIFCTEEQIRDEVNGCIRLVYDMYSTFGFEKIVVKLSTRPEKRIGSDEMWDRAEADLAVALEENNIPFEYQLG
EGAFYGPKIEFTLYDCLDRAWQVGTVQLDFSLPSRLSASYVGEDNERKVPVMIHRAILGSMERFIGILTEEFAGFF
PTWLAPVQVVIMNITDSQSEYVNELTQKLSNAGIRVKADLRNEKIGFKIREHTLRRVPYMLVCGDKEVESGKVA
VRTRRGKDLGSMDVNEVIEKLQQEIRSRSLKQLEE

A1.1.2 BorO

A1.1.2.1 Gene Sequence (Native)

GTGTCTGTAATCCGTCCCACCGCCGAAACCGAACGCGCAGTCGTGGTGGTCCCGGCTGGGACGACGTGCG
CCGACGCGGTCACCGCGGCAAAGCTGCCGCGCAATGGCCCCAACGCGATCGTCGTGGTGCGAGACCCGTC
CGGCGCCCTGCGTGACCTCGACTGGACCCCCGATTCCGACGTCGAGGTCGAGGCCGTCGCGTTGTCCAGC
GAGGACGGCCTCACGGTGCTGCGCCACTCCACGGCACACGTACTGGCCCAGGCGGTCCAGCAACTCTGGC
CGGAGGCCAGGCTCGGTATCGGCCCGCCGATCGAGAACGGCTTCTACTACGACTTCGACGTGGAGCGCCC
CTTCCAGCCAGAGGACCTCGAGCGCGTCGAGCAGCGGATGAAGGAGATCATCAAGTCCGGCCAGCGCTTC
TGCCGCCGCGAGTTCCCCGATCGGGAAGCGGCCCGTGCCGAGCTTGCCAAGGAGCCGTACAAGCTCGAGC
TCGTTGACCTCAAGGGCGACGTGGACGCCGCCGAGGCAATGGAGGTCGGCGGGAGCGACCTGACGATCT
ACGACAACCTCGACGCGAGAACTGGAGATGTGTGCTGGTCCGACCTCTGCCGCGGCCCCCACTTGCCGTCG
ACCCGCCTGATCCCGGCGTTCAAGCTGCTGCGCAACGCGGCAGCCTACTGGCGCGGCAGCGAGAAGAACC
CCCAACTGCAGCGCATCTACGGCACGGCCTGGCCGACCCGCGACGAGCTCAAGTCCCATCTCGCCGCCTTG
GAGGAGGCCGCCAAGCGTGACCACCGCCGCATCGGCGAGGAACTCGACCTCTTCGCGTTCAACAAGGAGA
TCGGCCGCGGCCTGCCGCTGTGGCTGCCCAACGGCGCGATCATCCGCGACGAACTCGAGGACTGGGCCCG
CAAGACCGAACGCAAGCTCGGCTACAAGCGCGTCGTCACCCCGCACATCACCCAGGAGGACCTTTACTACC

TCTCAGGCCATCTGCCTTACTACGCGGAGGACCTGTACGCGCCGATCGACATCGACGGCGAGAAGTACTAT
CTCAAGCCGATGAACTGCCCGCACCACCACATGGTGTACAAGGCGCGCCCGCACAGCTATCGCGACCTGCC
CTACAAGGTCGCCGAATACGGCACGGTGTACCGATTCGAGCGCAGCGGTCAGCTGCACGGCATGATGCGT
ACGCGCGGTTTCAGCCAGAATGACGCGCACATCTACTGCACGGCGGACCAGGCCAAGGACCAGTTCCTGG
AAGTCATGCGCATGCACGCGGACTACTACCGCACTCTGGGGATCAGCGACTTCTACATGGTGCTCGCGCTG
CGTGACTCGGCGAACAAGGACAAGTACCACGACGACGAGCAGATGTGGGAGGACGCTGAGCGGATCACC
CGGGAGGCCATGGAAGAGTCCGACATCCCCTTCCAGATCGACCTGGGCGGTGCCGCGCACTACGGCCCGA
AGGTCGACTTCATGATCCGAGCCGTCACCGGCAAGGAGTTCGCCGCCTCCACCAACCAGGTCGACCTGTAC
ACCCCGCAGCGTTTCGGGCTGACCTACCACGACTCCGACGGCACCGAGAAGCCCGTCGTGGTGATCCATCG
CGCTCCGCTCGGCTCGCACGAGCGCTTCACCGCCTATCTCACCGAGCACTTCGCAGGTGCCTTCCCGGTGTG
GTTGGCGCCGGAGCAGGTCCGGATCATTCCGATCGTGGAGGAACTCACGGACTACGCCGAGGAAGTCCGC
GACATGCTGCTGGACGCGGACGTGCGTGCCGACGTCGATGCCGGCGACGGCCGGCTGAATGCCAAGGTA
CGCGCGGCCGTCACCCGGAAGATCCCGCTCGTCGTGGTGGTCGGCAGGCGAGAGGCTGAGCAGCGCACC
GTAACCGTGCGCGACCGCTCCGGCGAGGAGACCCCGATGTCCCTGGAGAAGTTCGTGGCCCATGTCACTG
GACTCATCAGGACCAAGAGCCTGGACGGCGCCGGCCACATCCGTCCGCTGTCCAAGGCCTGA

A1.1.2.2 Gene Sequence (E. coli Codon Optimised)

ATGTCCGTAATTCGCCCGACGGCAGAAACTGAGCGTGCTGTAGTTGTAGTGCCTGCAGGCACAACCTGTGC
AGATGCTGTAACAGCAGCCAAACTCCCACGTAACGGTCCAAATGCCATTGTGGTTGTTCGCGATCCATCAG
GTGCGTTACGCGATTTAGATTGGACGCCAGACTCTGATGTGGAAGTGGAAGCGGTGGCACTGAGCAGTGA
AGATGGACTGACCGTACTCCGTCATAGCACTGCCCATGTCTTGGCGCAAGCCGTGCAACAGCTGTGGCCCG
AAGCACGCCTGGGCATTGGTCCCCCTATTGAAAACGGTTTCTATTATGATTTTGATGTAGAACGTCCGTTTC
AACCGGAAGACTTAGAACGTGTGGAACAACGTATGAAGGAAATTATTAAAAGCGGTCAACGTTTCTGTCG
TCGTGAATTTCCGGACCGCGAGGCCGCACGCGCGGAATTAGCAAAAGAACCATATAAACTGGAATTAGTG
GATCTGAAAGGTGATGTAGATGCAGCGGAAGCTATGGAAGTGGGCGGATCCGATCTCACTATTTATGATA
ATCTTGATGCCCGTACCGGTGACGTATGTTGGTCTGATTTGTGTCGTGGTCCGCATCTTCCAAGTACTCGTT
TGATTCCAGCATTTAAATTACTCCGTAATGCAGCGGCGTATTGGCGTGGTTCCGAAAAGAATCCACAGCTC
CAACGTATTTATGGAACCGCGTGGCCCACTCGTGATGAATTAAAATCACACCTGGCAGCGCTTGAAGAAGC
GGCTAAACGCGATCATCGTCGTATTGGGGAAGAGCTGGATTTATTTGCTTTTAATAAAGAAATTGGGCGTG
GTTTGCCACTCTGGCTCCCGAATGGTGCAATTATTCGTGATGAGCTTGAAGATTGGGCGCGTAAAACAGAG
CGTAAATTAGGATATAAACGTGTAGTAACGCCTCATATTACACAAGAAGATTTATATTATTTGTCCGGTCAC
CTGCCCTATTATGCTGAAGATCTTTATGCCCCAATTGATATCGATGGTGAGAAATATTACCTGAAACCAATG
AATTGTCCTCATCATCACATGGTCTATAAAGCCCGTCCTCATTCGTACCGTGATCTCCCTTATAAAGTTGCGG
AGTATGGAACCGTCTATCGCTTTGAACGTTCAGGCCAACTCCATGGAATGATGCGCACTCGTGGATTTTCG

CAAAACGATGCACATATTTACTGTACCGCCGATCAAGCTAAAGATCAATTTCTCGAAGTGATGCGTATGCAT
GCTGATTATTATCGTACCTTAGGAATTTCTGATTTCTATATGGTTCTGGCCCTTCGCGATTCCGCTAATAAAG
ATAAATACCATGATGATGAACAAATGTGGGAAGATGCGGAACGTATTACACGTGAAGCGATGGAAGAAA
GCGATATTCCGTTTCAAATTGATTTAGGTGGCGCGGCCCATTATGGGCCAAAAGTAGATTTTATGATTCGC
GCAGTTACAGGGAAAGAATTTGCAGCGTCTACGAATCAAGTAGATTTATATACGCCACAACGCTTTGGCCT
TACATATCATGATTCTGATGGAACAGAAAAGCCAGTAGTCGTTATTCACCGTGCCCCTTTAGGATCACATGA
ACGTTTTACGGCGTACCTTACGGAACATTTTGCGGGAGCTTTTCCCGTCTGGCTTGCCCCTGAACAAGTGCG
CATTATCCCAATTGTAGAAGAGCTGACCGATTATGCTGAAGAGGTGCGTGATATGCTTCTTGATGCCGATG
TCCGCGCGGATGTGGACGCAGGGGATGGGCGTTTGAACGCGAAAGTGCGTGCTGCGGTGACACGCAAAA
TTCCACTGGTGGTCGTCGTTGGACGCCGTGAAGCGGAACAACGTACAGTCACTGTCCGTGATCGTTCTGGA
GAAGAAACTCCCATGTCTTTAGAGAAATTTGTCGCACACGTAACCGGCCTGATTCGCACAAAATCATTAGAT
GGTGCGGGTCATATTCGCCCTCTTAGCAAAGCGTAA

A1.1.2.1.3 Protein Sequence

VSVIRPTAETERAVVVVPAGTTCADAVTAAKLPRNGPNAIVVVRDPSGALRDLDWTPDSDVEVEAVALSSEDGL
TVLRHSTAHVLAQAVQQLWPEARLGIGPPIENGFYYDFDVERPFQPEDLERVEQRMKEIIKSGQRFCRREFPDR
EAARAELAKEPYKLELVDLKGDVDAAEAMEVGGSDLTIYDNLDARTGDVCWSDLCRGPHLPSTRLIPAFKLLRN
AAAYWRGSEKNPQLQRIYGTAWPTRDELKSHLAALEEAAKRDHRRIGEELDLFAFNKEIGRGLPLWLPNGAIIR
DELEDWARKTERKLGYKRVVTPHITQEDLYYLSGHLPYYAEDLYAPIDIDGEKYYLKPMNCPHHHMVYKARPHS
YRDLPYKVAEYGTVYRFERSGQLHGMMRTRGFSQNDAHIYCTADQAKDQFLEVMRMHADYYRTLGISDFYM
VLALRDSANKDKYHDDEQMWEDAERITREAMEESDIPFQIDLGGAAHYGPKVDFMIRAVTGKEFAASTNQVD
LYTPQRFGLTYHDSDGTEKPVVVIHRAPLGSHERFTAYLTEHFAGAFPVWLAPEQVRIIPIVEELTDYAEEVRDML
LDADVRADVDAGDGRLNAKVRAAVTRKIPLVVVVGRREAEQRTVTVRDRSGEETPMSLEKFVAHVTGLIRTKSL
DGAGHIRPLSKA

A1.2.1.4 ΔN Protein Sequence

RDHRRIGEELDLFAFNKEIGRGLPLWLPNGAIIRDELEDWARKTERKLGYKRVVTPHITQEDLYYLSGHLPYYAED
LYAPIDIDGEKYYLKPMNCPHHHMVYKARPHSYRDLPYKVAEYGTVYRFERSGQLHGMMRTRGFSQNDAHIY
CTADQAKDQFLEVMRMHADYYRTLGISDFYMVLALRDSANKDKYHDDEQMWEDAERITREAMEESDIPFQID
LGGAAHYGPKVDFMIRAVTGKEFAASTNQVDLYTPQRFGLTYHDSDGTEKPVVVIHRAPLGSHERFTAYLTEHF
AGAFPVWLAPEQVRIIPIVEELTDYAEEVRDMLLDADVRADVDAGDGRLNAKVRAAVTRKIPLVVVVGRREAE
QRTVTVRDRSGEETPMSLEKFVAHVTGLIRT

A1.1.3 SpThrRS

A1.1.3.1 Gene Sequence (Native)

GTGTCAGACGTCCGTGTGATCATCCAACGCGATTCGGAGCAGGAAGAACGCGTGGTGACGACGGGCACTA
CGGCCGCCGAGCTCTTCGCCGGCCAGCGCTCGATCATCGCGGCGCGGGTGGCCGGTGAGCTCAAGGACCT
CGCCTACGAGGTCAAGGACGGCGAGACCGTCGAGGGCGTCGAGATCTCCTCCGAGGACGGCCTGGCCAT
CCTGCGCCACTCCACCGCGCACGTCATGGCGCAGGCCGTGCAGGAGCTCTTCCCCGAGGCCAAGCTGGGC
ATCGGCCCGCCCATCAAGGACGGCTTCTACTACGACTTCGACGTCGAGAAGCCCTTCCACCCCGATGACCTC
AAGGCCATCGAGAAGAAGATGCAGGAGATCCAGAAGCGCGGCCAGCGCTTCTCCCGCCGCGTCGTCACCG
ACGAGGCCGCCCGCGACGAGCTGGCCGGTGAGCCGTACAAGCTGGAGCTGATCGGTCTCAAGGGTGCCG
CCGGGCAGGCCGCCGACGGTGCCGACGCCGAGGTCGGCGCCGGTGAGCTGACCATCTACGACAACCTCG
ACGCCAAGACCGGCGAGCTGTGCTGGAAGGACCTCTGCCGAGGCCCCCACCTGCCGACCACCCGCGTCAT
CCCCGCCTTCAAGCTGATGCGGTCGGCCGCCGCCTACTGGCGCGGCAGCGAGAAGAACCCGCAGCTCCAG
CGCATCTACGGCACCGCCTGGCCGTCCAAGGACGAGCTGAAGGCGCACCTGGACTTCCTCGCCGAGGCCG
AGAAGCGCGACCACCGCAAGCTCGGCGCCGAGCTCGACCTGTTCTCCTTCCCCGACGAGCTGGGCCCCGG
CCTCGCGGTCTTCCACCCCAAGGGCGGCGTGATCCGCAAGGTCATGGAGGACTACTCGCGCCGCCGGCAC
GAGGTCTCCGGCTACGAGTTCGTGAACACCCCGCACATCTCGAAGGAGCACCTCTTCGAGATCTCCGGGCA
CCTGCCGCACTACTCGGAGGGCATGTTCCCGCCCATCCAGTTCGACGAGCAGAACTACCGCCTCAAGGCGA
TGAACTGCCCGATGCACAACCTGATCTTCAAGTCGCGGGGCCGCTCCTACCGTGAGCTGCCGCTGCGCCTG
TTCGAGTTCGGGACGGTGTACCGGTACGAGAAGTCGGGCGTCGTGCACGGGCTGACCCGCTCGCGCGGCT
TCACCCAGGACGACTCGCACATCTACTGCACCAAGGAGCAGATGCCCGAGGAGCTGGACACGCTCCTGAC
CTTCGTGCTCGACCTGCTGCGCGACTACGGCCTGACCGAGTTCGAGCTGGAGCTGTCCACCCGCGACGACT
CCGACAAGTTCATCGGCTCGGACGAGGACTGGGAGGAGGCCACGGAGGCGCTCCGCCTGGCCGCCGAGA
AGCAGAACCTGCCCCTGGTCCCGGACCCGGGCGGCGCCGCCTACTACGGCCCGAAGATCTCCGTGCAGGC
CAAGGACGCCATCGGCCGGTCCTGGCAGATGTCGACCATCCAGGTCGACTTCAACCAGCCCAAGCGGTTC
GGCCTGGAGTACACCGCGGCGGACGGCTCGCACCAGCAGCCGGTCATGATCCACCGGGCGCTGTTCGGCT
CCATCGAGCGCTTCTTCGGTGTGCTGCTGGAGCACTACGCGGGTGCCTTCCCGGCGTGGCTCGCGCCCGTC
CAGGCGCTCGGCATCCCGATCGGCGACGCGCACGTGCCGTACCTGCGGGAGTTCGCCGAGAAGGCCAGG
GCGGCCGGTCTGCGCGTCGAGGTGGACTCCTCCTCCGACCGCATGCAGAAGAAGATCCGCAACGCCCAGA
AGCAGAAGGTGCCCTTCATGGTCATCGCGGGCGACGAGGACATGGCGAACGGCGCCGTCTCCTTCCGCTA
CCGCGACGGCTCCCAGGAGAACGGCATCCCCGTCGACGAGGCCATCGCCAAGATCGCGAAGGTCGTCGAG
GAGCGCGCGCAGGTCTGA

A1.1.3.2 Gene Sequence (E. coli Codon Optimised)

ATGAGCGATGTGCGTGTTATTATTCAGCGTGACTCGGAACAAGAGGAGCGTGTAGTCACAACCGGGACCA
CTGCGGCAGAATTGTTTGCTGGTCAACGTAGCATTATTGCCGCCCGCGTCGCGGGCGAACTGAAAGATCTG
GCATATGAAGTTAAAGATGGTGAAACGGTGGAAGGAGTGGAAATTTCAAGCGAAGATGGTCTTGCGATTC
TTCGTCATAGCACTGCACATGTGATGGCCCAAGCTGTTCAAGAATTATTTCCTGAAGCAAAACTCGGAATTG
GTCCCCCGATTAAAGATGGGTTTTATTATGATTTTGATGTGGAGAAGCCGTTTCATCCTGACGATTTGAAGG
CGATTGAAAAGAAAATGCAAGAAATTCAGAAACGTGGTCAACGTTTCTCGCGTCGTGTAGTTACGGATGA
AGCGGCGCGTGATGAATTAGCGGGCGAACCATATAAACTCGAATTGATTGGCCTGAAAGGAGCAGCGGG
CCAAGCGGCAGATGGCGCAGATGCGGAAGTAGGAGCAGGGGAACTTACAATTTATGATAATCTGGATGC
GAAAACTGGAGAACTCTGTTGGAAAGATCTTTGTCGTGGTCCACATCTCCCAACGACACGTGTGATTCCGG
CGTTTAAACTTATGCGCAGCGCGGCTGCTTATTGGCGTGGGTCGGAAAAGAATCCTCAATTGCAACGTATT
TATGGAACTGCGTGGCCCAGCAAAGATGAACTTAAAGCCCATTTGGATTTCTTGGCGGAAGCAGAAAAGC
GTGATCATCGTAAATTGGGTGCAGAACTTGATCTCTTTAGTTTTCCTGATGAACTTGGTCCAGGTCTGGCAG
TGTTTCATCCAAAAGGTGGTGTCATTCGTAAAGTGATGGAAGATTATAGCCGTCGTCGTCATGAAGTAAGT
GGCTATGAATTTGTAAATACGCCACACATTAGCAAAGAACATCTGTTTGAAATTTCGGGCCATTTACCTCAT
TATTCTGAAGGAATGTTTCCGCCGATTCAATTCGATGAACAAAATTATCGCCTGAAAGCTATGAACTGTCCC
ATGCATAATCTTATTTTCAAAAGTCGTGGTCGTTCTTATCGCGAATTGCCCCTGCGTCTCTTTGAATTTGGCA
CTGTCTATCGTTATGAAAAGAGTGGTGTGGTTCATGGCTTAACGCGTAGTCGTGGGTTTACGCAAGATGAT
TCCCATATTTATTGTACGAAGGAACAGATGCCAGAAGAATTAGATACCCTTCTTACGTTTGTTCTGGATCTT
CTTCGTGATTATGGACTTACGGAATTTGAATTAGAACTCTCGACGCGTGATGATTCGGATAAATTTATTGGG
TCTGATGAAGATTGGGAAGAAGCTACCGAAGCCTTACGTTTAGCGGCGGAAAAGCAAAATCTTCCTTTAGT
TCCTGATCCTGGTGGTGCGGCGTATTATGGTCCGAAAATTAGCGTTCAAGCGAAAGATGCGATTGGGCGT
AGCTGGCAAATGAGCACTATTCAAGTGGATTTTAATCAACCGAAACGCTTTGGGTTGGAATATACAGCCGC
TGATGGGTCCCATCAACAACCTGTTATGATTCATCGCGCTTTGTTTGGTAGCATTGAACGTTTCTTTGGCGT
ACTTCTTGAACATTATGCCGGCGCGTTTCCTGCATGGCTTGCACCGGTACAAGCACTGGGAATTCCCATTGG
TGATGCACATGTTCCATATTTACGCGAATTTGCAGAAAAGGCGCGTGCTGCGGGCCTTCGTGTTGAAGTCG
ATAGTTCGTCTGATCGTATGCAAAAGAAAATTCGTAATGCGCAAAAGCAAAAGGTTCCGTTTATGGTAATT
GCCGGTGATGAAGATATGGCAAATGGAGCTGTTTCGTTTCGTTATCGTGATGGTTCACAAGAAAATGGAAT
TCCTGTGGATGAAGCAATTGCGAAAATTGCAAAAGTAGTTGAAGAACGTGCTCAAGTTTAA

A1.1.3.3 Protein Sequence

VSDVRVIIQRDSEQEERVVTTGTTAAELFAGQRSIIAARVAGELKDLAYEVKDGETVEGVEISSEDGLAILRHSTAH
VMAQAVQELFPEAKLGIGPPIKDGFYYDFDVEKPFHPDDLKAIEKKMQEIQKRGQRFSRRVVTDEAARDELAGE
PYKLELIGLKGAAGQAADGADAEVGAGELTIYDNLDAKTGELCWKDLCRGPHLPTTRVIPAFKLMRSAAAYWR

GSEKNPQLQRIYGTAWPSKDELKAHLDFLAEAEKRDHRKLGAELDLFSFPDELGPGLAVFHPKGGVIRKVMEDY
SRRRHEVSGYEFVNTPHISKEHLFEISGHLPHYSEGMFPPIQFDEQNYRLKAMNCPMHNLIFKSRGRSYRELPLRL
FEFGTVYRYEKSGVVHGLTRSRGFTQDDSHIYCTKEQMPEELDTLLTFVLDLLRDYGLTEFELELSTRDDSDKFIGS
DEDWEEATEALRLAAEKQNLPLVPDPGGAAYYGPKISVQAKDAIGRSWQMSTIQVDFNQPKRFGLEYTAADG
SHQQPVMIHRALFGSIERFFGVLLEHYAGAFPAWLAPVQALGIPIGDAHVPYLREFAEKARAAGLRVEVDSSSD
RMQKKIRNAQKQKVPFMVIAGDEDMANGAVSFRYRDGSQENGIPVDEAIAKIAKVVEERAQV

A1.1.3.4 ΔN Protein Sequence

RDHRKLGAELDLFSFPDELGPGLAVFHPKGGVIRKVMEDYSRRRHEVSGYEFVNTPHISKEHLFEISGHLPHYSEG
MFPPIQFDEQNYRLKAMNCPMHNLIFKSRGRSYRELPLRLFEFGTVYRYEKSGVVHGLTRSRGFTQDDSHIYCTK
EQMPEELDTLLTFVLDLLRDYGLTEFELELSTRDDSDKFIGSDEDWEEATEALRLAAEKQNLPLVPDPGGAAYYG
PKISVQAKDAIGRSWQMSTIQVDFNQPKRFGLEYTAADGSHQQPVMIHRALFGSIERFFGVLLEHYAGAFPAW
LAPVQALGIPIGDAHVPYLREFAEKARAAGLRVEVDSSSDRMQKKIRNAQKQKVPFMVIAGDEDMANGAVSF
RYRDGSQENGIPVDEAIAKIAKVVEERAQV

A1.1.4 SvThrRS

A1.1.4.1 Gene Sequence

GTGTCTGATGTCCGTGTGACCGTCCAGTCCGCCTCGGAAGCAGAGGAGAGGGCGGTGAGCGCGGGCACC
ACCGCCGGCGCCCTGTTCGCCGACGACCGCACCGTCATCGCCGCCCGCGTCGGTGGCGAGCTGAAGGACC
TGTCGTACGAGCTCGCCGACGGCGATGTCGTCGAGGGCGTCGAGATCTCCTCCCCGGACGGTCTCGACATC
CTGCGCCACTCGACCGCGCACGTCATGGCCCAGGCCGTGCAGGAGCTCTTCCCCGAGGCCAAGCTGGGCA
TCGGCCCGCCGGTCCGGGACGGCTTCTACTACGACTTCGACGTCGAGAAGCCGTTCACTCCTGAGGACCTC
AAGGTCATCGAGAAGAAGATGCAGGAGATCCAGAAGCGCGGCCAGCGCTTCGCCCGCCGGGTGGTGACC
GACGAGGACGCCCGCGCGGAGCTGGCGGACGAGCCGTACAAGCTGGAGCTCATCGGCATCAAGGGCTCG
GCCTCGACCGACGACGGCGCGAACGTCGAGGTGGGCGGCGGCGAGCTGACCATCTACGACAACCTCGAC
GCCAAGACCGGCGAGCTGTGCTGGAAGGACCTCTGTCGTGGTCCGCACCTGCCCACCACCCGGAACATCC
CGGCGTTCAAGCTGATGCGCAACGCCGCCGCCTACTGGCGCGGCAGCGAGAAGAACCCGATGCTCCAGCG
CATCTACGGCACCGCCTGGCCGTCGAAGGACGAGCTGAAGGCCCACCTCGACTTCCTCGCCGAGGCCGAG
AAGCGCGACCACCGCAAGCTGGGCAACGAGCTCGACCTCTTCTCCATCCCGGACGAGATCGGCTCCGGCCT
GGCGGTCTTCCACCCCAAGGGCGGCATCATCCGCCGGGTCATGGAGGACTACTCGCGCAAGCGCCACGAG
GAGGAGGGCTACGAGTTCGTCTACTCGCCGCACGCCACCAAGGGCAAGCTGTTCGAGAAGTCCGGTCACC
TCGACTGGTACGCCGAGGGCATGTACCCCCCCATGCAGCTCGACGAGGGCGTGGACTACTACCTCAAGCC
CATGAACTGCCCGATGCACAACCTGATCTTCGACGCGCGCGGTCGTTCCTACCGTGAACTGCCGCTGCGCC
TGTTCGAGTTCGGCACCGTGTACCGGTACGAGAAGTCGGGCGTCGTGCACGGCCTGACCCGCGCCCGCGG
CTTCACCCAGGACGACGCGCACATCTACTGCACCAAGGAGCAGATGGCGGAGGAGCTCGACAAGACCCTC

ACCTTCGTCCTGAACCTGCTGCGCGACTACGGTCTGACCGACTTCTACCTGGAGCTGTCCACCAAGGACCC
GGAGAAGTTCGTCGGCTCGGACGAGATCTGGGAGGAGGCCACCGCGGTCCTCCAGCAGGTCGCCGAGAA
GCAGGGCCTCCCGCTCACCCCCGACCCGGGCGGCGCCGCGTTCTACGGCCCGAAGATCTCGGTGCAGGCG
CGGGACGCCATCGGCCGCACCTGGCAGATGTCGACCGTGCAGCTCGACTTCAACCTGCCGGAGCGCTTCG
ACCTGGAGTACACCGCCCCGGACGGCACCAAGCAGCGCCCGGTCATGATCCACCGCGCGCTGTTCGGTTC
GATCGAGCGCTTCTTCGCCGTGCTCCTCGAGCACTACGCGGGCGCGATGCCGCCGTGGCTGGCGCCCGTCC
AGGCCACCGGCATCCCGATCGGCGACGCGCACGTCGACTACCTGCACGAGTTCGCCGCCAAGGCGAAGAA
GCAGGGCCTGCGGGTGGACGTGGACTCGTCCTCGGACCGGATGCAGAAGAAGATCCGCAACGCGCAGAA
GCAGAAGGTCCCGTTCATGATCATCGCGGGTGACGAGGACATGGCCGCCGGCGCCGTCTCCTTCCGCTACC
GCGACGGTTCGCAGGAGAACGGCATCCCCGTCGACGAGGCCATCGCCAAGATCGCCAAGATCGTCGAGG
ACCGCGTCCAGGTCTGA

A1.1.4.2 Protein Sequence

MSDVRVTVQSASEAEERAVSAGTTAGALFADDRTVIAARVGGELKDLSYELADGDVVEGVEISSPDGLDILRHST
AHVMAQAVQELFPEAKLGIGPPVRDGFYYDFDVEKPFTPEDLKVIEKKMQEIQKRGQRFARRVVTDEDARAEL
ADEPYKLELIGIKGSASTDDGANVEVGGGELTIYDNLDAKTGELCWKDLCRGPHLPTTRNIPAFKLMRNAAAYW
RGSEKNPMLQRIYGTAWPSKDELKAHLDFLAEAEKRDHRKLGNELDLFSIPDEIGSGLAVFHPKGGIIRRVMEDY
SRKRHEEEGYEFVYSPHATKGKLFEKSGHLDWYAEGMYPPMQLDEGVDYYLKPMNCPMHNLIFDARGRSYRE
LPLRLFEFGTVYRYEKSGVVHGLTRARGFTQDDAHIYCTKEQMAEELDKTLTFVLNLLRDYGLTDFYLELSTKDPE
KFVGSDEIWEEATAVLQQVAEKQGLPLTPDPGGAAFYGPKISVQARDAIGRTWQMSTVQLDFNLPERFDLEYT
APDGTKQRPVMIHRALFGSIERFFAVLLEHYAGAMPPWLAPVQATGIPIGDAHVDYLHEFAAKAKKQGLRVDV
DSSSDRMQKKIRNAQKQKVPFMIIAGDEDMAAGAVSFRYRDGSQENGIPVDEAIAKIAKIVEDRVQV

A1.1.5 PfObaO

A1.1.5.1 Gene Sequence

ATGGTCACTATCGCTCTACCGGACGGCAGTCGCAGAGATTTTCCAGAAGCTTTGACTGTGCAACAGCTGGC
TCAATCCATTGGCGCAGGCCTCGCTGCCGCGACGATTGGCGGCAAGGTCGACGGCACGCTGGTCGATGCC
AGCTACCTCCTGGAAACAGACGCTACCGTAGAGATTGTCACGACCAAAAGCCCGCAAGCGCTGGAGCTGA
TACGCCATTCGACGGCCCACTTGATGGCACAGGCGGTTCAGCGCCTGTACCCCGGCACGCAAGTAACGATA
GGCCCCGTTATTGATAATGGCTTCTATTACGACTTCGTGGCACCTCGTCCTTTCACAATGGACGATCTTCCGC
TCATCGAAGCCGAAATGACCAGGATCGTCAAAGAACAACTACCGGTTACCCGTACGCAATTACCACGGGAT
GAAGCACTGGCGTTTTTTGAACAACTGGGCGAAAGCTACAAGACGCAAATCATCGACGCCATACCCGCCG
GGGAAACGCTCTCACTTTATACCCAGGGCGAGTTCACCGACCTGTGCCGCGGGCCGCATGTGCCCAACACA
GCGAAGCTTGGCGCCTTCAAACTGATGAAAGTGGCTGGCGCCTACTGGCGCGGCGACTCCAACAACATCA
TGCTCAGCCGTATCTACGGCACTGCCTGGGGCAACGAAAAAGAGCTCAAGGCCTACTTGAATCAACTTCAA

GAAGCTGAAAAGCGTGATCACCGCAAACTGGCAAAACAGTTCGACTTGTTTCACCAACAAGAAGAAGCGC
CAGGCATGGTCTTTTGGCACCCCAAGGGCTGGTCGTTGTGGCAGACGGTTGAACAGTACATGCGCCGGGT
TTACCGCGATGGCGGCTACCGTGAAGTCAAGTCGCCGCAAGTACTCGACAGCACCCTGTGGAAGAAATCA
GGGCACTGGGACAACTACAAAGAAAACATGTTTGTCACTGAGTCCGAAAACCGCCAGTACGCACTCAAGC
CGATGAACTGCCCCGGCCACATCCAAATCTTCAAGCATGGCCTGCGCAGCCACCGTGAACTGCCGATCCGT
TACGGCGAGTTCGGCGGCTGCCACCGCAACGAGCCTTCGGGCGCTTTGCACGGCATTATGCGGGTGCGCG
CATTTACCCAGGATGACGGGCACATTTTCTGCACCGAAGAGCAAATCGCGGCCGAGATCAAGGCGTTCCAT
TACCAGGCAGTCAAGGTTTATGCCGACTTCGGCTTCACTGACATTGCGGTAAAAATCGCACTGCGTCCTGA
GCCAGGCAAACGCCTGGGCAGCGACGAAGTGTGGGACAAGGCCGAAAACCTGTTGCGTGAAGCCCTGTC
CGAGTGTGATGTGGAGTGGGAAGAACTGCCGGGCGAAGGCGCGTTCTACAGCCCCAAGATCGAGTACCA
CCTGCGTGACGCCATCGGGCGTGAATGGCAGGTAGGCACGATGCAGGTCGACTATCACATGCCGGATCGC
CTGGGCGCCGAATACGTCGATGAGCATTCGCAACGACGCAAGCCGGTCATGCTGCACCGCGCAATCGTGG
GCTCCCTGGAGCGGTTCCTTGGCATTCTCATCGAACACCATGCCGGTCAGTTCCCGCTCTGGCTTGCGCCGG
TGCAGGCCATCGTCGTGACGGTCACTGATGCACAGAACGATTATGCCGACCAGACACGCAATGACCTGGTC
CAACTTGGGTTCAGGGTGGAAGCTGATCTGCGCAATGAGAAAATCGGCTACAAGATTCGTGAAAGCACTC
TTCAACGCGTGCCTTACTTGCTCGTGGTCGGTGAGCGCGAAAAAGAAAACGGTACTGTCACCGTGCGCTCG
CGCGCAGGCGAAGACTTGGGAAGCATGACGATGGAAGCGCTGCACGCCTTCCTGTTGAACGAGCAATCAG
CAGGCGGCTGA

A1.1.5.2 Protein Sequence

MVTIALPDGSRRDFPEALTVQQLAQSIGAGLAAATIGGKVDGTLVDASYLLETDATVEIVTTKSPQALELIRHSTA
HLMAQAVQRLYPGTQVTIGPVIDNGFYYDFVAPRPFTMDDLPLIEAEMTRIVKEQLPVTRTQLPRDEALAFFEQ
LGESYKTQIIDAIPAGETLSLYTQGEFTDLCRGPHVPNTAKLGAFKLMKVAGAYWRGDSNNIMLSRIYGTAWGN
EKELKAYLNQLQEAEKRDHRKLAKQFDLFHQQEEAPGMVFWHPKGWSLWQTVEQYMRRVYRDGGYREVKS
PQVLDSTLWKKSGHWDNYKENMFVTESENRQYALKPMNCPGHIQIFKHGLRSHRELPIRYGEFGGCHRNEPS
GALHGIMRVRAFTQDDGHIFCTEEQIAAEIKAFHYQAVKVYADFGFTDIAVKIALRPEPGKRLGSDEVWDKAEN
LLREALSECDVEWEELPGEGAFYSPKIEYHLRDAIGREWQVGTMQVDYHMPDRLGAEYVDEHSQRRKPVMLH
RAIVGSLERFLGILIEHHAGQFPLWLAPVQAIVVTVTDAQNDYADQTRNDLVQLGFRVEADLRNEKIGYKIREST
LQRVPYLLVVGEREKENGTVTVRSRAGEDLGSMTMEALHAFLLNEQSAGG

A1.1.5.3 ΔN Protein Sequence

A1.1.6 BmObaO

A1.1.6.1 Gene Sequence

ATGATCAGCATTGCGCTCCCAGATGGTAGTCGTCGCCGTTATGACCGTCCGGTAACTGTTGCAGCCCTGGC
CGCAGACATCGGTCCAGGGCTGGCAAAGGCGGCTTTGGCCGGTAAGATTGACGGTAAACTCGTGGACTTG
GATTATTTAATCGATATTGACGCGACCGCCGAGATTGTTACCGAGAAACATCCGGACGCCTTGAGTATCAT
TCGTCATAGCTGTGCTCATCTCCTCGCGCAAGCGGTGCAGCGCCTGTATCCTGCCGCGCAGTTCTCGATCGG
ACCCGTAATTGAGAATGGGTTTTATTACGACATCTCAATTTCTCCGCCTTTGTCGGAAGACGATTTGCCCCG
CATTGAGACCGAGATGCGTGCAATTGTAGCCGAGGCTGTACCGGTATCACGTGCGGTGCTTAGCCGTGAC
GACGCTATTCGTTTCTTTTCTGAGCGCGGCCAAACCTACAAGGCCGAGATTGTGGCAAGTATTCCGGAGCA
CGAGCAGCTCACGATCTACACTCAGGGTGAGTTCTCTGACTTGTGTCGCGGCCCGCATGTTCCGAGCACGC
GCGCACTTCGTGCTTTTAAACTCATGAAAACTGCCGGAGCGTATTGGCGCGGTGACAGCAACAATGAGATG
CTTTGCCGTGTTTATGGTACGGCATGGTTAAATGATGCAGATTTGCAGGCGTATCTGCACCAATTGGCAGA
GGCCGAACGCCGCGACCACCGCAAAATTGGTAAGCAATTAGATCTGTTTCACATCCAAGAAGAGGCGCCG
GGCATGGTGTTCTGGCATCCCAAGGGCTGGAGCATGTGGCAGGTGGTTGAACAATATATTCGCCGCGTTT
ACGTGGAGTGCGGTTACCAAGAGGTGCGTGCCCCTCAGGTAGTCGATGTGTCGTTGTGGAAGCGCAGTGG
ACACTGGGATAATTATAAAGAGAACATGTTCTTCACCGAGTCAGAGAAGCGTGAATACGCCTTGAAGCCTA
TGAATTGTCCAGGTCATATTCAAATCTTTAAGCACGGCTTACGCTCATATCGTGACCTGCCTTTGCGTTACG
GCGAATTCGGTGGATGCCACCGTAATGAGGCTAGCGGCGCGCTGCACGGTATCATGCGTGTGCGTGCGTT
CACCCAAGACGATGGACACATCTTTTGTACAGAGGCGCAAATCGAGGACGAGGTGGCGGCCTTCCATCGT
CAAGCAATGAAGGTGTACGCGGACTTTGGTTTTGGCGATGATTCAATCGCCGTAAAGATCGCCCTGCGTCC
GGAGCTGCGTCTGGGCAGCGATGAGGTATGGGACCGCGCGGAAAACACCCTTCGCGATGCTCTGCGCAA
GTGTGGAGTCGAATGGGAAGAGCTGCCAGGCGAGGGCGCGTTCTATTCCCCCAAGGTCGAGTATCACCTT
AAAGACGCGATTGGCCGCGAGTGGCAAGTGGGTACTATCCAGGCAGATTACCTCATGCCTGAGAAGCTCG
GAGCCGAGTACATCGATGAGCGCAGCGAACGCCGTACCCCCGTTATGCTCCACCGTGCAATTGTGGGGAG
TATGGAGCGCTTCATCGGGATCCTTATTGAGCACCATGCGGGCCATCTCCCTGTTTGGCTGGCACCCATTCA
AGCGATGGTACTTAACGTAACGGACGCGCAACGTGAATATGTCCACGACGTTCGTCGCGCGTTGATCGGA
CACGGCGTCCGCGTAGACGTGGACGTCCGTAATGAGAAAATCGGCTACAAGGTTCGCGAGCATGTCCTGC
AGAAGGTGCCATACTTACTTGTGGCGGGCGAACGCGAGCGCGAAGCTGGTGTTGTGAGCGTCCGCGCCCA
CTCAGGTGAGGACTTAGGTACTATGACACTCGAAGCGTTCGCGGCACGCGTCCGTTCGGAGCGCCCGGCA
TGA

A1.1.6.2 Protein Sequence

MISIALPDGSRRRYDRPVTVAALAADIGPGLAKAALAGKIDGKLVDLDYLIDIDATAEIVTEKHPDALSIIRHSCAHL
LAQAVQRLYPAAQFSIGPVIENGFYYDISISPPLSEDDLPRIETEMRAIVAEAVPVSRAVLSRDDAIRFFSERGQTY

KAEIVASIPEHEQLTIYTQGEFSDLCRGPHVPSTRALRAFKLMKTAGAYWRGDSNNEMLCRVYGTAWLNDADL
QAYLHQLAEAERRDHRKIGKQLDLFHIQEEAPGMVFWHPKGWSMWQVVEQYIRRVYVECGYQEVRAPQVV
DVSLWKRSGHWDNYKENMFFTESEKREYALKPMNCPGHIQIFKHGLRSYRDLPLRYGEFGGCHRNEASGALH
GIMRVRAFTQDDGHIFCTEAQIEDEVAAFHRQAMKVYADFGFGDDSIAVKIALRPELRLGSDEVWDRAENTLR
DALRKCGVEWEELPGEGAFYSPKVEYHLKDAIGREWQVGTIQADYLMPEKLGAEYIDERSERRTPVMLHRAIVG
SMERFIGILIEHHAGHLPVWLAPIQAMVLNVTDAQREYVHDVRRALIGHGVRVDVDVRNEKIGYKVREHVLQK
VPYLLVAGERERE AGVVSVRAHSGEDLGTMTLEAFAARVRSERPA

A1.1.6.3 ΔN Protein Sequence

RDHRKIGKQLDLFHIQEEAPGMVFWHPKGWSMWQVVEQYIRRVYVECGYQEVRAPQVVDVSLWKRSGHW
DNYKENMFFTESEKREYALKPMNCPGHIQIFKHGLRSYRDLPLRYGEFGGCHRNEASGALHGIMRVRAFTQDD
GHIFCTEAQIEDEVAAFHRQAMKVYADFGFGDDSIAVKIALRPELRLGSDEVWDRAENTLRDALRKCGVEWEEL
PGEGAFYSPKVEYHLKDAIGREWQVGTIQADYLMPEKLGAEYIDERSERRTPVMLHRAIVGSMERFIGILIEHHA
GHLPVWLAPIQAMVLNVTDAQREYVHDVRRALIGHGVRVDVDVRNEKIGYKVREHVLQKVPYLLVAGERERE
AGVVSVRAHSGEDLGTMTLEAFAARVRSERPA

A1.1.7 CsObaO

A1.1.7.1 Gene Sequence

ATGATTACTATTAGTTTGCCTGATGGATCTAAGCGTGAATTTGCCGAACCCATCAGCGTACACGAGTTGGC
ATGCGCTATCGGACCGGGCCTTGGCGCCGCCGCACTTGCAGGCAAGGTTGACGGCAAATTAGTGGACACG
GCTCACCTTTTGCGTCATGATGCTACCGTAGAGATTGTTACTGACCGTCATCCAGATGCCCTGGAAGTCGTT
CGCCACTCTACAGCGCACCTGTTAGCGCAAGCGGTGCAGCGCCTCTACCCTGGCACACAAGTCACGATCGG
CCCAGTTATCGATAATGGATTTTACTATGATTTTGCCGGCGAGCGTCCTTTTACGGTCGAAGATCTTCCCGC
GATTGAGGCAGAGATGGCGCGTATTGCTAAAGAGGCACTCCCTGTAACCCGCAGTGAGAAGACCCGTGAG
CAAGCCGCTCAATTCTTCGAGGGTTTGGGTGAGCACTACAAGGTGGAAATTCTCCGTGATATCGCTGACGA
CCAACCTTTAAGCCTGTACACCCAAGGGGAGTTCACCGACCTGTGTCGTGGACCGCACGTGCCGAATACTG
GGAAGTTGCGCGCATTTAAGCTTATGAAGGTAGCGGGTGCATACTGGCGTGGCAACAGTGATAACGCCAT
GCTGAGCCGCATTTATGGCACAGCGTGGCTGAATGACAAAGACCTTAAGGCATACTTGCTGCAACTTGAGG
AAGCGGAGAAGCGCGACCACCGCAAGATTGCTAAGCTGTTAGACCTCTTCCATCAACAAGAGGAAGCTCC
CGGTATGGTTTTCTGGCACTACAAGGGCTGGGCGCTGTGGCAAGCGGTTGAACAATACATGCGTCGCGTG
TACCGTGATTCAGGTTATCGTGAAGTGAAGGCTCCTCAAGTCTTAGACGTCTCGCTGTGGCAACGCAGTGG
CCATTGGCAGAATTATCAAGAAAACATGTTTCTGACCGAGAGCGAGAAACGCCAATACGCTTTAAAGCCAA
TGAATTGTCCTGGGCATATCCAAATCTTCAAGCAGGGGCTTCGCTCGTATCGTGAATTGCCAATTCGCTACG
GTGAGTTTGGTGGATGTCATCGTAACGAGCCCAGCGGTGCACTCCACGGCATCATGCGTGTACGTGCCTTT
ACGCAAGATGACGGGCACGTTTTCTGTACTGAGGAACAAATTGCTGACGAGGTGCAGGCCTTTCATCGCCA

AGCCCTTAAAGTATACGCCGACTTTGGTTTTGACAATATTGCGGTTAAAATTGCTCTGCGTCCAGAAGCCGG
TAAGCGCCTGGGTAGTGACGAGGTGTGGGACAAGGCCGAGGCCCTTCTCCGTTCAGCACTGTCGGCGTGC
GGCGTAACATGGGAAGAGCTTCCGGGTGAAGGTGCGTTCTATTCGCCTAAGATTGAGTATCACTTAAAGG
ACGCTATTGGACGTGAGTGGCAAGTTGGCACCATGCAGGTTGACTACCTCATGCCGGAGCGCCTTGGAGC
TGAGTATATTGACGAGCACTCTCAACGCCGCTCGCCAGTCATGCTGCACCGTGCCATCGTTGGCTCTTTAGA
ACGCTTTATTGGAATCCTCATTGAGCACCACGCGGGCTACTTCCCAACGTGGCTGGCCCCGGTTCAAGCAA
TGTTGATGAACGTGACGGACGCACAAGCAGATTATGTCGAGGCCGTCCGTACGGCGCTTACCCGCGAGGG
GTTCCGTGTGGAGAGTGATTTACGCAATGAAAAGATTGGTTACAAGATTCGTGAAAATACCCTGCAACGCA
TTCCTTACCTCTTGGTCATTGGAGATCGTGAGAAGGAACACGGTACTGTAACCGTGCGTAGTCGCGCTGGC
GACGATTTGGGTACTATGACACCGGCCGAGTTCGCCGCACGCTTGCGTGAGGAGACGGCCATTGGATAA

A1.1.7.2 Protein Sequence

MITISLPDGSKREFAEPISVHELACAIGPGLGAAALAGKVDGKLVDTAHLLRHDATVEIVTDRHPDALEVVRHSTA
HLLAQAVQRLYPGTQVTIGPVIDNGFYYDFAGERPFTVEDLPAIEAEMARIAKEALPVTRSEKTREQAAQFFEGL
GEHYKVEILRDIADDQPLSLYTQGEFTDLCRGPHVPNTGKLRAFKLMKVAGAYWRGNSDNAMLSRIYGTAWL
NDKDLKAYLLQLEEAEKRDHRKIAKLLDLFHQQEEAPGMVFWHYKGWALWQAVEQYMRRVYRDSGYREVKA
PQVLDVSLWQRSGHWQNYQENMFLTESEKRQYALKPMNCPGHIQIFKQGLRSYRELPIRYGEFGGCHRNEPS
GALHGIMRVRAFTQDDGHVFCTEEQIADEVQAFHRQALKVYADFGFDNIAVKIALRPEAGKRLGSDEVWDKAE
ALLRSALSACGVTWEELPGEGAFYSPKIEYHLKDAIGREWQVGTMQVDYLMPERLGAEYIDEHSQRRSPVMLH
RAIVGSLERFIGILIEHHAGYFPTWLAPVQAMLMNVTDAQADYVEAVRTALTREGFRVESDLRNEKIGYKIRENT
LQRIPYLLVIGDREKEHGTVTVRSRAGDDLGTMTPAEFAARLREETAIG

A1.1.7.3 ΔN Protein Sequence

RDHRKIAKLLDLFHQQEEAPGMVFWHYKGWALWQAVEQYMRRVYRDSGYREVKAPQVLDVSLWQRSGHW
QNYQENMFLTESEKRQYALKPMNCPGHIQIFKQGLRSYRELPIRYGEFGGCHRNEPSGALHGIMRVRAFTQDD
GHVFCTEEQIADEVQAFHRQALKVYADFGFDNIAVKIALRPEAGKRLGSDEVWDKAEALLRSALSACGVTWEEL
PGEGAFYSPKIEYHLKDAIGREWQVGTMQVDYLMPERLGAEYIDEHSQRRSPVMLHRAIVGSLERFIGILIEHHA
GYFPTWLAPVQAMLMNVTDAQADYVEAVRTALTREGFRVESDLRNEKIGYKIRENTLQRIPYLLVIGDREKEHG
TVTVRSRAGDDLGTMTPAEFAARLREETAIG

A1.1.8 PIObaO

A1.1.8.1 Gene Sequence

ATGGTTTCAATCGCTTTACCGGACGGTAGCCGTCGTGAATTCCCGGAACCTCTTACTGTTCACCAATTAGCC
CAGAGCATTGGAATCGGCTTGGCCGCGGCAACTCTTGCCGGAAAAGTTGACGGTCGCCTTGTTGACGCAA
GCTTCGAACTTACTACGGATGCAACTGTCGAGGTTGTAACGTCAAAGTCGCAAGAGGCATTAGAGATTGTC
CGCCACAGCACCGCGCACTTACTGGCGCAGGCTGTACAACGCCTCTATCCTGGCACACAGGTAACTATCGG

GCCTGTGATTGACAATGGATTCTACTACGACTTCGTCACAGAACGCCCTTTTACCTTAGACGACTTACCCTT
GATTGAGGCAGAGATGAAGCGTATTGTGGACTCGAATTTACCCATTGAGCGTCGCACGCTGTCGCGCGAC
TTGGCCATCGACTTCTTTCACGCCTTGGGCGAGAGCTATAAGGTACGTATTATCGAGGATATTCCCGCGAG
CGAGACTTTGTCTTTGTACAGCCAAGGCGAGTTTACCGACTTGTGTCGCGGCCCTCATGTTCCCGACACTGG
GCGTTTAGGCGCATTCAAGCTCATGAAGGTTGCGGGTGCCTATTGGCGCGGGGACTCAAGCAACGCCATG
TTGAGTCGCATTTATGGCACCGCGTGGTTGAACGAGAAGCAGCTGAAAACCTACCTTAGCCAACTGGAAG
AGGCCGAGAAGCGCGATCATCGTAAGTTGGCCAAGCAGCTGAACCTGTTCCATCAACAGGAAGAGGCCCC
GGGCATGGTGTTCTGGCATCCCAAGGGTTGGTCGATCTGGCAAACGGTCGAGCAATATATGCGCAAGGTG
TATCGCGACAACGGATATCAAGAAGTCCGCTCCCCTCAAGTTGTCGACTCATCTTTGTGGCGCCGTTCAGGC
CACTGGGACAACTACAAGGAGAATATGTTTGTTACCGAGTCGGAAAACCGTCAGTACGCGTTAAAGCCAAT
GAATTGCCCTGGGCACATCCAGATCTTCAAGTTCGGCCTGCGTAGCTATCGTGAATTACCGATTCGTTATGG
CGAGTTCGGCGGATGTCATCGCAACGAACCAAGCGGAGCGCTTCATGGTATTATGCGTGTACGTGCTTTTA
CGCAGGACGATGGGCACATTTTCTGTACGGAGCAACAGATTGAGGCAGAAATCCAAACTTTCCATCACCAA
GCGCAAAAGGTGTATAAGGCGTTCGGGTTTGAAGACGTGGTGGTTAAAATTGCGCTCCGCCCGGAGCCTG
GTAAGCGCCTTGGCTCGGACGACGTTTGGGACAAGGCCGAGAATCAACTTCGCTCAGCCCTGTCTGCGTGT
GGGGTAAGTTGGGAAGAACTGCCGGGAGAAGGAGCGTTCTATAGCCCTAAGATCGAATATCATTTGCGCG
ACGCCATTGGACGTGAGTGGCAAGTCGGCACCATCCAGGTTGACTATCACATGCCGGACCGCCTGGGCGC
GGAATATGTCGACGAGCATAGCCAACGTCAACGTCCTGTCATGCTTCATCGCGCCATCGTTGGATCGATGG
AACGCTTTATCGGTATTTTAATCGAGCATCACGCGGGTGTTTTCCCCACGTGGTTGGCGCCCGTGCAAGCA
ATGGTGTTAACTGTAACGGATGCGCAGAATGCGTATGCTGAACAAGTTCGCGAGCAACTTCGCGGGATGG
GACTGCGTGTAGAGGCGGATCTTCGCAACGAAAAGATTGGCTACAAGATCCGCGAGAGTACGCTCCAACG
CGTGCCGTACCTGTTGGTCATCGGAGACCGCGAAAAGGAGAACGCTACCGTTACCGTCCGCGCTCGCTCG
GGTGAGGACCTTGGCACCCAATCAATTGCAGCGTTCGCCGAACGCTTGACGCGTGAGAATGTCGTGGATT
GA

A1.1.8.2 Protein Sequence

MVSIALPDGSRREFPEPLTVHQLAQSIGIGLAAATLAGKVDGRLVDASFELTTDATVEVVTSKSQEALEIVRHSTA
HLLAQAVQRLYPGTQVTIGPVIDNGFYYDFVTERPFTLDDLPLIEAEMKRIVDSNLPIERRTLSRDLAIDFFHALGE
SYKVRIIEDIPASETLSLYSQGEFTDLCRGPHVPDTGRLGAFKLMKVAGAYWRGDSSNAMLSRIYGTAWLNEKQ
LKTYLSQLEEAEKRDHRKLAKQLNLFHQQEEAPGMVFWHPKGWSIWQTVEQYMRKVYRDNGYQEVRSPQV
VDSSLWRRSGHWDNYKENMFVTESENRQYALKPMNCPGHIQIFKFGLRSYRELPIRYGEFGGCHRNEPSGALH
GIMRVRAFTQDDGHIFCTEQQIEAEIQTFHHQAQKVYKAFGFEDVVVKIALRPEPGKRLGSDDVWDKAENQLR
SALSACGVSWEELPGEGAFYSPKIEYHLRDAIGREWQVGTIQVDYHMPDRLGAEYVDEHSQRQRPVMLHRAIV

GSMERFIGILIEHHAGVFPTWLAPVQAMVLTVTDAQNAYAEQVREQLRGMGLRVEADLRNEKIGYKIRESTLQ
RVPYLLVIGDREKENATVTVRARSGEDLGTQSIAAFAERLTRENVVD

A1.1.8.3 ΔN Protein Sequence

RDHRKLAKQLNLFHQQEEAPGMVFWHPKGWSIWQTVEQYMRKVYRDNGYQEVRSPQVVDSSLWRRSGH
WDNYKENMFVTESENRQYALKPMNCPGHIQIFKFGLRSYRELPIRYGEFGGCHRNEPSGALHGIMRVRAFTQD
DGHIFCTEQQIEAEIQTFHHQAQKVYKAFGFEDVVVKIALRPEPGKRLGSDDVWDKAENQLRSALSACGVSWE
ELPGEGAFYSPKIEYHLRDAIGREWQVGTIQVDYHMPDRLGAEYVDEHSQRQRPVMLHRAIVGSMERFIGILIE
HHAGVFPTWLAPVQAMVLTVTDAQNAYAEQVREQLRGMGLRVEADLRNEKIGYKIRESTLQRVPYLLVIGDRE
KENATVTVRARSGEDLGTQSIAAFAERLTRENVVD

A1.1.9 PfThrRS

A1.1.9.1 Gene Sequence

ATGCCAACTATTACTCTACCCGACGGCAGTCAACGTTCATTCGATCATCCGGTTTCCGTAGCCGAGGTCGCC
GCATCCATTGGTGCCGGCCTGGCCAAGGCCACCGTGGCCGGCAAGGTCGATGGCCAGCTGGTCGATGCCA
GTGACCTGATCACCTCCGATGCCAGCCTGCAGATCATCACGCCCAAGGATCAAGAGGGGCTCGAGATTATT
CGCCACTCTTGCGCGCACCTGATTGGCCATGCGGTCAAGCAGCTGTACCCAACCGCCAAGATGGTGATCGG
CCCGGTAATCGAAGAAGGCTTCTATTACGACATCGCCTATGAGCGTCCTTTCACTCCGGACGACCTGGCGG
CCATCGAGCAGCGCATGCACGCGCTGATCGAAAAAGATTACGACGTGATCAAGAAGGTCACCCCGCGCGC
CGAAGTGATCGACGTGTTTACCGCCCGTGGCGAAGACTACAAGCTGCGCCTGGTGGAAGACATGCCGGAC
GAGCAGGCCATGGGTCTGTACTACCACGAAGAATACGTGGACATGTGCCGTGGCCCGCACGTGCCGAACA
CGCGCTTTCTCAAGTCGTTCAAGCTGACCAAGTTGTCCGGTGCCTACTGGCGCGGCGACGCAAAGAACGA
GCAACTGCAACGTATCTACGGCACCGCCTGGGCCGACAAGAAGCAGCTGGCCGCCTATATCCAGCGCATC
GAAGAAGCCGAGAAACGCGACCACCGCAAGATCGGCAAGCGCCTGAACCTGTTCCATCTGCAGGAAGAA
GCGCCGGGCATGGTGTTCTGGCATCCGAACGGCTGGACCCTGTACCAGGTGCTCGAGCAGTACATGCGCA
AGGTTCAGCGCGAAAACGGCTACCTCGAGATCAAGACCCCGCAGGTCGTCGATCGCAGCCTGTGGGAGAA
GTCCGGGCACTGGGCCAACTACGCCGACAATATGTTCACCACCCAGTCGGAAAACCGCGACTACGCTATCA
AGCCGATGAACTGCCCTTGCCATGTGCAGGTGTTCAATCAAGGCCTGAAGAGCTACCGCGAGTTGCCGAT
GCGTCTGGCCGAGTTCGGTGCCTGCCACCGCAACGAACCATCGGGTGCGCTGCACGGCATCATGCGCGTG
CGCGGCTTTACCCAGGACGACGCCCATATTTTCTGCACCGAAGAGCAGATGCAGGCTGAATCGGCCGCGTT
CATCAAGCTGACCATGGACGTTTACCGCGATTTCGGCTTCACCGAAGTCGAGATGAAACTGTCCACTCGTC
CGGAAAAACGCGTCGGCTCCGACGAGCTCTGGGATCGCGCCGAAGCAGCACTGGCCGCAGCACTCGACA
GCGCGGGCCTTGCGTACGACTTGCAGCCGGGCGAGGGCGCGTTCTACGGTCCCAAGATCGAGTTCTCGCT
GAAAGATTGCCTTGGCCGTGTCTGGCAGTGTGGTACCCTGCAGCTCGATTTTAACCTGCCGATCCGTCTGG
GAGCCGAATACGTCTGCGAAGACAACAGTCGTAAACACCCGGTTATGCTGCACCGGGCGATCCTCGGCTC

GTTCGAACGGTTCGTCGGCATCCTGATCGAGCACTACGAGGGCGCGTTCCCGGCGTGGCTGGCTCCGACC
CAGGCAGTGATCATGAATATCACTGATAAACAGGCAGATTTTGCCGCTGAAGTTGAAAAAACACTCAACGA
AAGCGGATTTCGTGCCAAGTCCGACTTGAGAAATGAAAAGATCGGCTTTAAAATCCGTGAGCATACTTTGC
TCAAGGTTCCCTATCTTTTGGTTATCGGAGATCGGGAAGTCGAGATGCAGACTGTCGCTGTGCGTACTCGT
GAAGGTGCTGACCTGGGCTCGATGCCCGTCGCCCAGTTCGCTGAGTTCCTCGCGCAAGCGGTTTCCCGGCG
TGGTCGCCCAGATTCGGAGTAA

A1.1.9.2 Protein Sequence

MPTITLPDGSQRSFDHPVSVAEVAASIGAGLAKATVAGKVDGQLVDASDLITSDASLQIITPKDQEGLEIIRHSCA
HLIGHAVKQLYPTAKMVIGPVIEEGFYYDIAYERPFTPDDLAAIEQRMHALIEKDYDVIKKVTPRAEVIDVFTARG
EDYKLRLVEDMPDEQAMGLYYHEEYVDMCRGPHVPNTRFLKSFKLTKLSGAYWRGDAKNEQLQRIYGTAWA
DKKQLAAYIQRIEEAEKRDHRKIGKRLNLFHLQEEAPGMVFWHPNGWTLYQVLEQYMRKVQRENGYLEIKTPQ
VVDRSLWEKSGHWANYADNMFTTQSENRDYAIKPMNCPCHVQVFNQGLKSYRELPMRLAEFGACHRNEPS
GALHGIMRVRGFTQDDAHIFCTEEQMQAESAAFIKLTMDVYRDFGFTEVEMKLSTRPEKRVGSDELWDRAEA
ALAAALDSAGLAYDLQPGEGAFYGPKIEFSLKDCLGRVWQCGTLQLDFNLPIRLGAEYVCEDNSRKHPVMLHRA
ILGSFERFVGILIEHYEGAFPAWLAPTQAVIMNITDKQADFAAEVEKTLNESGFRAKSDLRNEKIGFKIREHTLLKV
PYLLVIGDREVEMQTVAVRTREGADLGSMPVAQFAEFLAQAVSRRGRPDSE

A1.1.9.3 ΔN Protein Sequence

RDHRKLAKQFDLFHQQEEAPGMVFWHPKGWSLWQTVEQYMRRVYRDGGYREVKSPQVLDSTLWKKSGHW
DNYKENMFVTESENRQYALKPMNCPGHIQIFKHGLRSHRELPIRYGEFGGCHRNEPSGALHGIMRVRAFTQDD
GHIFCTEEQIAAEIKAFHYQAVKVYADFGFTDIAVKIALRPEPGKRLGSDEVWDKAENLLREALSECDVEWEELP
GEGAFYSPKIEYHLRDAIGREWQVGTMQVDYHMPDRLGAEYVDEHSQRRKPVMLHRAIVGSLERFLGILIEHH
AGQFPLWLAPVQAIVVTVTDAQNDYADQTRNDLVQLGFRVEADLRNEKIGYKIRESTLQRVPYLLVVGEREKEN
GTVTVRSRAGEDLGSMTMEALHAFLLNEQSAGG

A1.1.10 Micromonospora sp. KC207 gene O

A1.1.10.1 Gene Sequence (Codon Optimised)

ATGGTTCGTGATGCCGACGGTGCACTGCGTGATCTTGATTGGGTTCCTGAGGCTGATACTGAGGTGGAACT
GATCGAACTTAGTTCCCCGGATGGGCTGCGCGTCTTACGCCATTCAACGGCACACGTACTGGCTCAAGCTG
TTCAGGATCTTTTCCCGGAAGCTAAGCTCGGCATTGGTCCACCCACGGAGAATGGGTTCTATTACGACTTCG
GCGTTAGCAAGCCCTTTACCCCCGAGGATCTGGAACGTATCGAGGTGCGCATGAAAGAAATCGTCCGTTCC
GGACAGCTCTTCCGCCGTCGCGTGTTTCCCTCTGTGGGCGCAGCCCGCACAGAGCTCGCTGATGAACCATA
TAAGTTGGAGTTGATTGATATCAAAGGTAACGTGGACTTTTCAGAAGAGGTTATGGAAGTAGGGGAAGGT
CAACTGTCAATCTACGACAATGTAGATACCAAGAATGATAAGGTTGTATGGTCCGATTTGTGCCGTGGACC
CCATTTACCGCTTACTCGCCTGATCGGTGCGTTTAAACTTCTGCGCGTTGCAGGTGCATACTGGCGTGGTCA

AGAAGATCGCCCGATGTTACAGCGCATCTATGGTACGGCATGGGCGCGTCAAAAGGATTTAGACGATTAC
ATGTGGCGCCTGGAGCAGGCAGAACTGCGTGACCACCGCCGTATCGGGCGTGAATTGGAGCTCTTTCATTT
CGATAGCACGGCCCCTGGAATGCCCTACTGGTTACCGAAAGGTATGCGCGTTCTTAACAACTTACTCCAATT
CTGGCGTGACGAGCATGAGGCACGTGGATACGAGGAGATTGCGACGCCTCTGATTAATAATAAGAAGCTC
TGGGAGACGTCGGGCCATTGGGATCATTTTCGCGACGACATGTTCGTGATTGCGGGCGATGAATCTAGCA
CTATGGCGCTCAAGCCCATGAATTGTCCGAATGCCATGGTAGTTTTCAACCTGAAGACCCGCTCTTACCGCG
ATTTCCCTTTGCGCTTCTCAGATTCAGACCCTTTGCATCGCAACGAACGTTCGGGTACTCTCCATGGTCTCCT
TCGTGTTCAAAAGTTTCAACAAGACGATGCGCATGTCTTCGTAACCCCTGCGCAAATCCGTGATGAGTATG
AACGCATCTTCGACATTTGTGAGCGCTTCTACAAGATTTTCGGACTCACTTATCGCCTTCGCGTGGGCACGC
GTCCAGGCAAATTCATCGGTGATCGCGCTACTTGGGACTCTGCGGAAGAAACACTCTTGAGTATCGTACAC
GAGCGCACCGGCGGAGACTTTGACACGGAAGAAGGTGGCGGTGCGTTTTATGGCCCGAAGATCGACATCT
TAATGCAAGATGTACTCGGCCGTCAGTGGCAGACGGGTACCATTCAACTTGACTTCCAACTCCCGCGTCGT
TTTGGCTGTGTATATATTGATGATCAGGGTCAGCGCCAAACGCCAGTGGTAATCCACCGCGTAGTATATGG
AAGTTTAGAACGTTTCCTCGGTATCTACATTGAGCACACCGCAGGGAATTTCCCTTTGTGGTTAGCTCCTGT
CCATGTGGCGGTTCTTCCTATCTCCGACCAATTCACGGACTACGCGGAACATGTCGGCGCGCGTATCCGTG
AGGCTGGCGCGCGTGTGGAAGTCGATTCTCGTAATGAGACTCTCGGCAATCGCGTTCGTCTCACGCAACAG
AATAAGGTACCCCTGGTCGTCGTCGTGGGTGCGCGCGAGCAGGAAGATTCAACCGTTAGTCTTCGCCTTCG
CGGCGGGAAGAAGGTAGGAGCCCTGCCGTTAGCGGCCTTTGCGGATGGCCTCTCAGCGGACCTGCGTAA
GCGTTCCCCCGAATTTACCATTCCCGCCGCTGGTTAA

A1.1.10.2 Protein Sequence

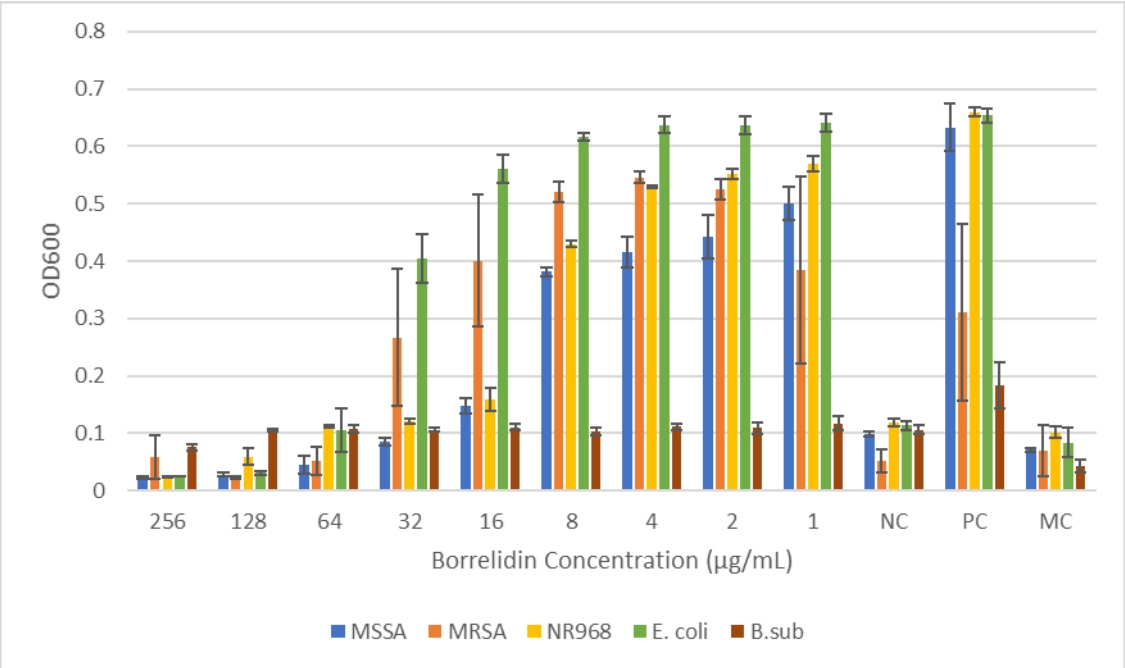
MVRDADGALRDLDWVPEADTEVELIELSSPDGLRVLRHSTAHVLAQAVQDLFPEAKLGIGPPTENGFYYDFGVS
KPFTPEDLERIEVRMKEIVRSGQLFRRRVFPSVGAARTELADEPYKLELIDIKGNVDFSEEVMEVGEGQLSIYDNV
DTKNDKVVWSDLCRGPHLPLTRLIGAFKLLRVAGAYWRGQEDRPMLQRIYGTAWARQKDLDDYMWRLEQA
ELRDHRRIGRELELFHFDSTAPGMPYWLPKGMRVLNNLLQFWRDEHEARGYEEIATPLINNKKLWETSGHWD
HFRDDMFVIAGDESSTMALKPMNCPNAMVVFNLKTRSYRDFPLRFSDSDPLHRNERSGTLHGLLRVQKFQQD
DAHVFVTPAQIRDEYERIFDICERFYKIFGLTYRLRVGTRPGKFIGDRATWDSAEETLLSIVHERTGGDFDTEEGG
GAFYGPKIDILMQDVLGRQWQTGTIQLDFQLPRRFGCVYIDDQGQRQTPVVIHRVVYGSLERFLGIYIEHTAGN
FPLWLAPVHVAVLPISDQFTDYAEHVGARIREAGARVEVDSRNETLGNRVRLTQQNKVPLVVVVGAREQEDST
VSLRLRGGKKVGALPLAAFADGLSADLRKRSPEFTIPAAG

A1.1.10.3 ΔN Protein Sequence

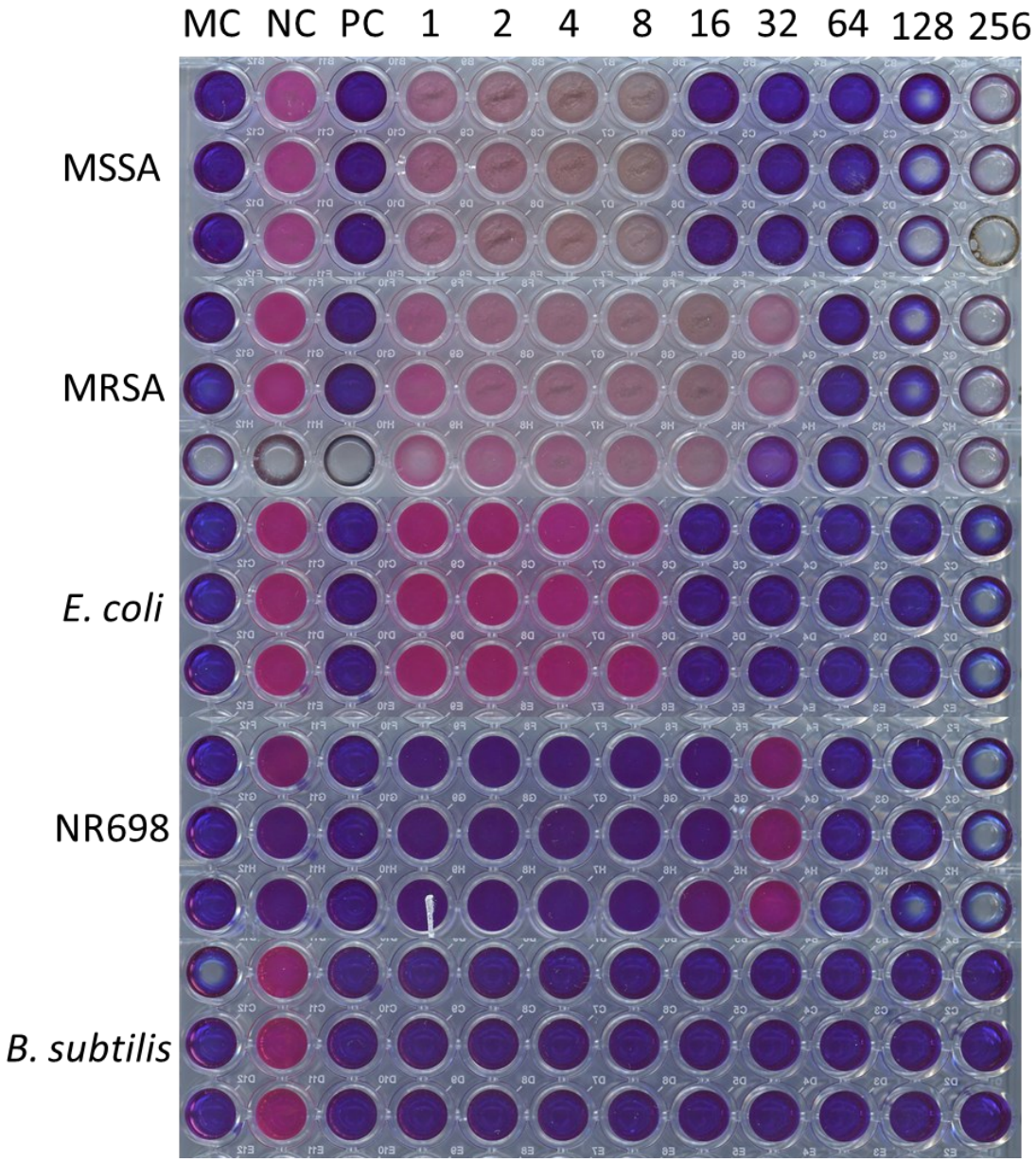
RDHRRIGRELELFHFDSTAPGMPYWLPKGMRVLNNLLQFWRDEHEARGYEEIATPLINNKKLWETSGHWDHF
RDDMFVIAGDESSTMALKPMNCPNAMVVFNLKTRSYRDFPLRFSDSDPLHRNERSGTLHGLLRVQKFQQDDA
HVFVTPAQIRDEYERIFDICERFYKIFGLTYRLRVGTRPGKFIGDRATWDSAEETLLSIVHERTGGDFDTEEGGGAF

YGPKIDILMQDVLGRQWQTGTIQLDFQLPRRFGCVYIDDQGQRQTPVVIHRVVYGSLERFLGIYIEHTAGNFPL
WLAPVHVAVLPISDQFTDYAEHVGARIREAGARVEVDSRNETLGNRVRLTQQNKVPLVVVVGAREQEDSTVSL
RLRGGKKVGALPLAAFADGLSADLRKRSPEFTIPAAG

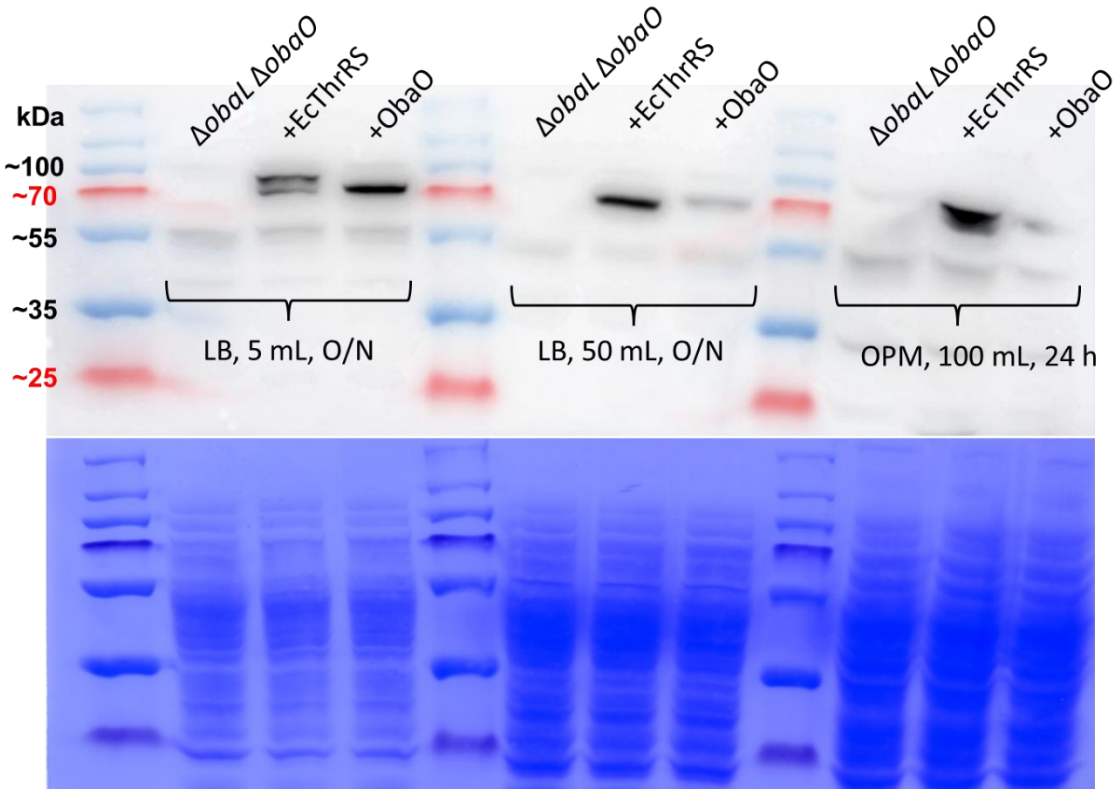
Appendix 2: Supplemental Information



Supplemental Figure 1. OD_{600} of each bioindicator strain after 20 hours at 37°C with increasing concentrations of borrelidin.



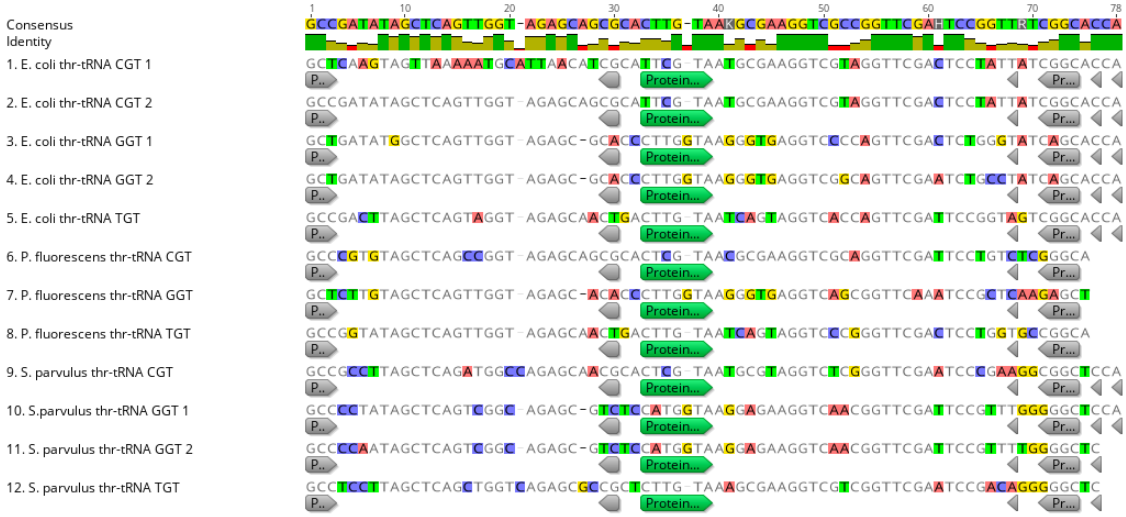
Supplemental Figure 2. Photographs of resazurin assays for MSSA, MRSA, *E. coli*, NR698 and *B. subtilis.* Numbers represent borrelidin concentration in µg/mL. MC = media control- no bacteria inoculated, NC = negative control, no added antibiotic and PC = positive control, 50 µg/mL kanamycin.



Supplemental Figure 3. EcThrRS and ObaO were expressed in *P. fluorescens* ΔobaOΔobaL. Top: Western blot probed with a HRP conjugated anti-FLAG antibody. Sizes on the prestained broad range colour protein marked. Bottom: SDS-PAGE gel to show loading of samples. Coomassie stained with InstantBlue. Experiment performed by Dr Sibyl Batey.

	S. parvulus thr-tRNA CGT	*S. parvulus* thr-tRNA GGT 1	*S. parvulus* thr-tRNA GGT 2	*S. parvulus* thr-tRNA TGT
E. coli thr-tRNA CGT 1	63.636	53.846	53.947	58.667
E. coli thr-tRNA CGT 2	74.026	67.532	65.333	72
E. coli thr-tRNA GGT 1	61.538	67.105	64.865	60.526
E. coli thr-tRNA GGT 2	61.538	68.421	66.216	65.789
E. coli thr-tRNA TGT	76.623	68.831	66.667	69.333
P. fluorescens thr-tRNA CGT	71.622	68.919	68.919	71.622
P. fluorescens thr-tRNA GGT	61.333	71.233	69.863	66.667
P. fluorescens thr-tRNA TGT	75.676	66.216	63.514	64.865

Supplemental Figure 4. Sequence identity matrix for tRNA sequences from *E. coli* and *P. fluorescens* with the *S. parvulus* tRNAs, coloured by sequence identity.



Supplemental Figure 5. Sequence alignment of *E. coli, P. fluorescens* and *S. parvulus* tRNA genes. Anticodon annotated by green arrows and protein binding sections annotated by grey arrows.

Supplemental Figure 6. Photograph of WT *S. parvulus* cultures on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.

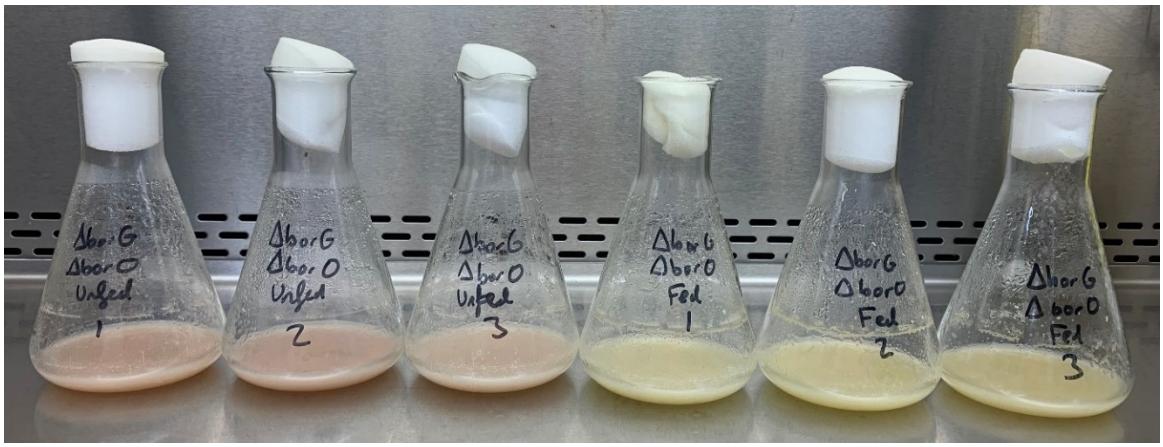
Supplemental Figure 7. Photograph of *S. parvulus ΔborO* cultures on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.

Supplemental Figure 8. Photograph of *S. parvulus ΔborO* cultures complemented with empty vector control on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.

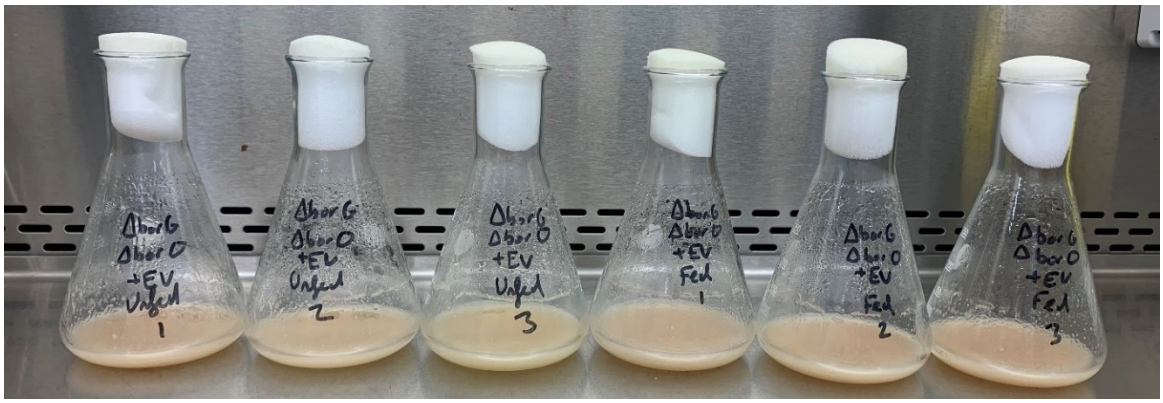
Supplemental Figure 9. Photograph of *S. parvulus ΔborO* cultures complemented with *borO* on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.

Supplemental Figure 10. Photograph of *S. parvulus ΔborO* cultures complemented with *SpThrRS* on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.

Supplemental Figure 11. Photograph of *S. parvulus ΔborG* on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.

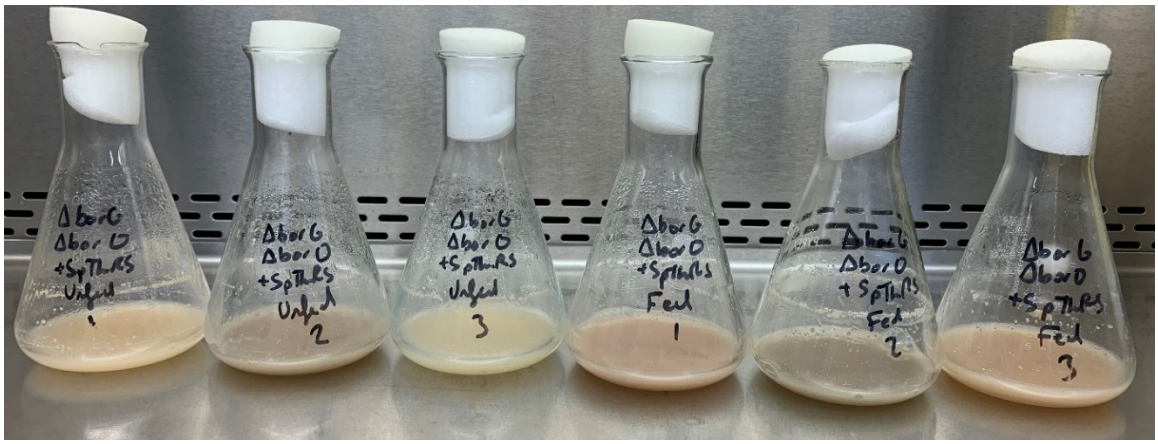


Supplemental Figure 12. Photograph of *S. parvulus ΔborGΔborO* cultures on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.

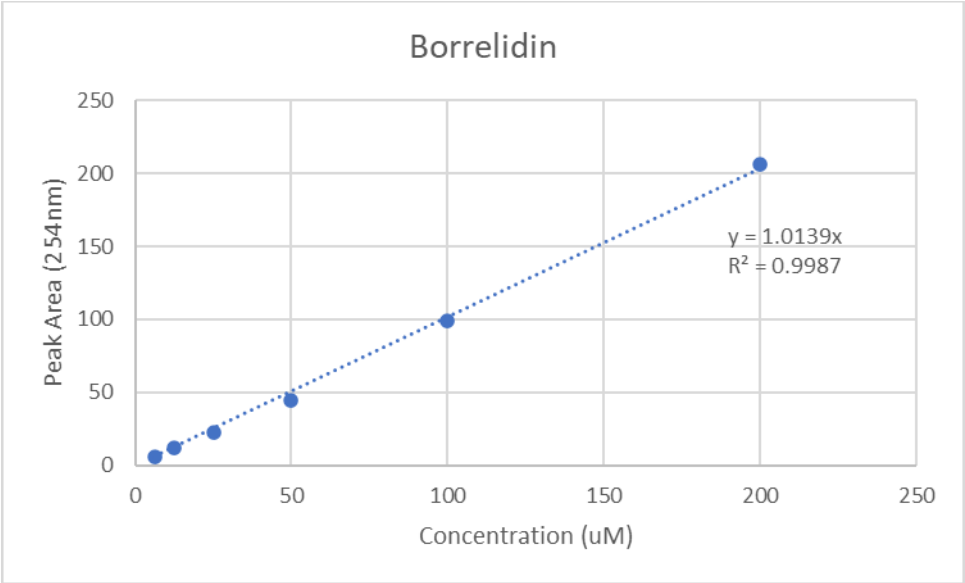


Supplemental Figure 13. Photograph of *S. parvulus ΔborGΔborO* cultures complemented with empty vector control on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.

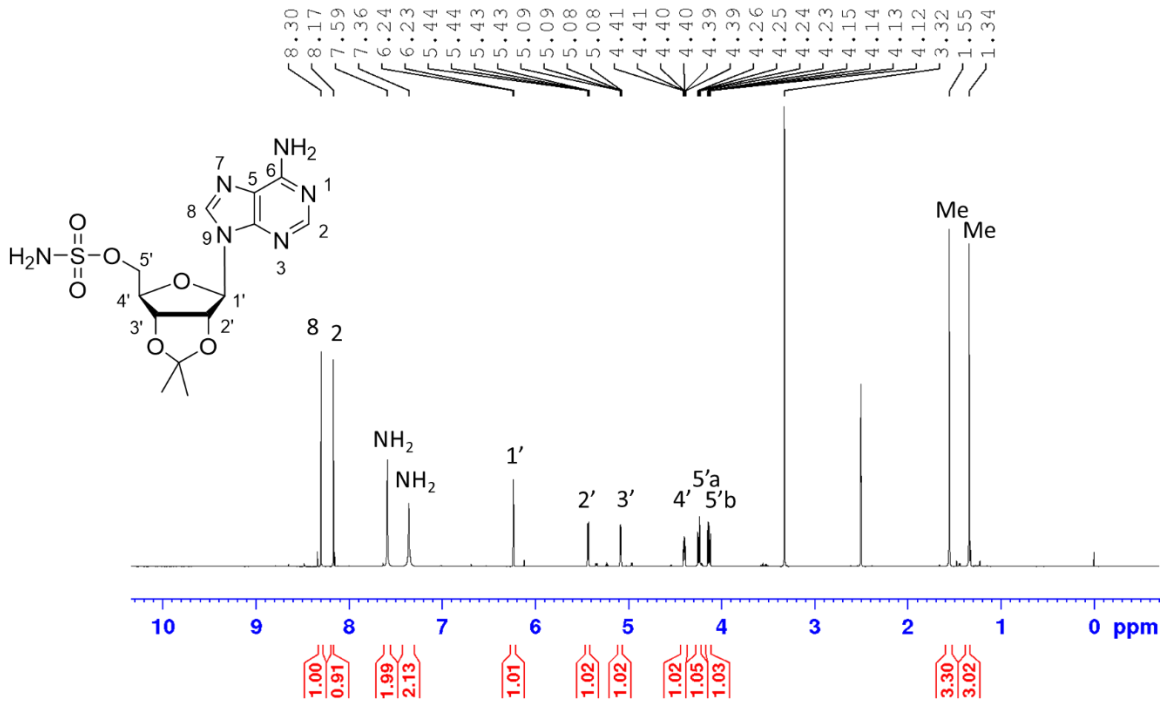
Supplemental Figure 14. Photograph of *S. parvulus ΔborGΔborO* cultures complemented *borO* on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.



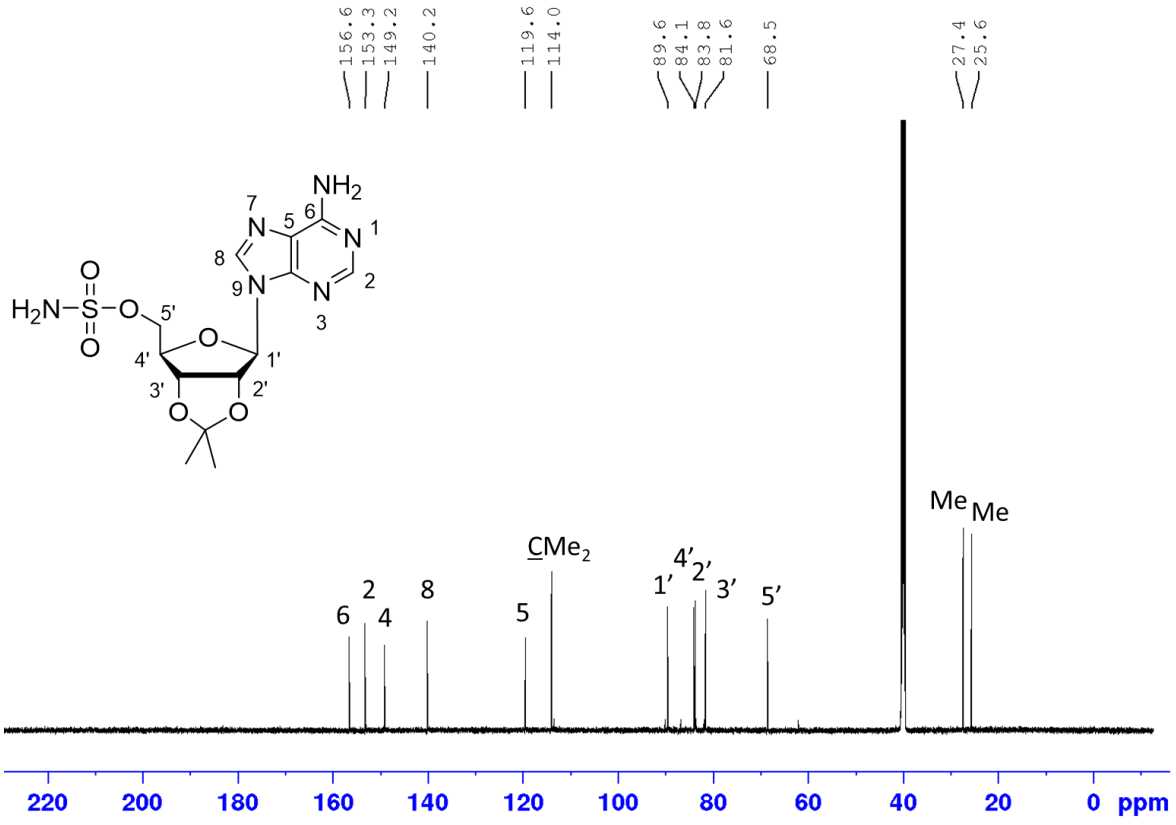
Supplemental Figure 15. Photograph of *S. parvulus ΔborGΔborO* cultures complemented with *SpThrRS* on day 5 of growth in borrelidin production media (PYDG). Left: biological triplicate of tCPDA unfed samples, right: biological triplicate of samples fed with tCPDA.



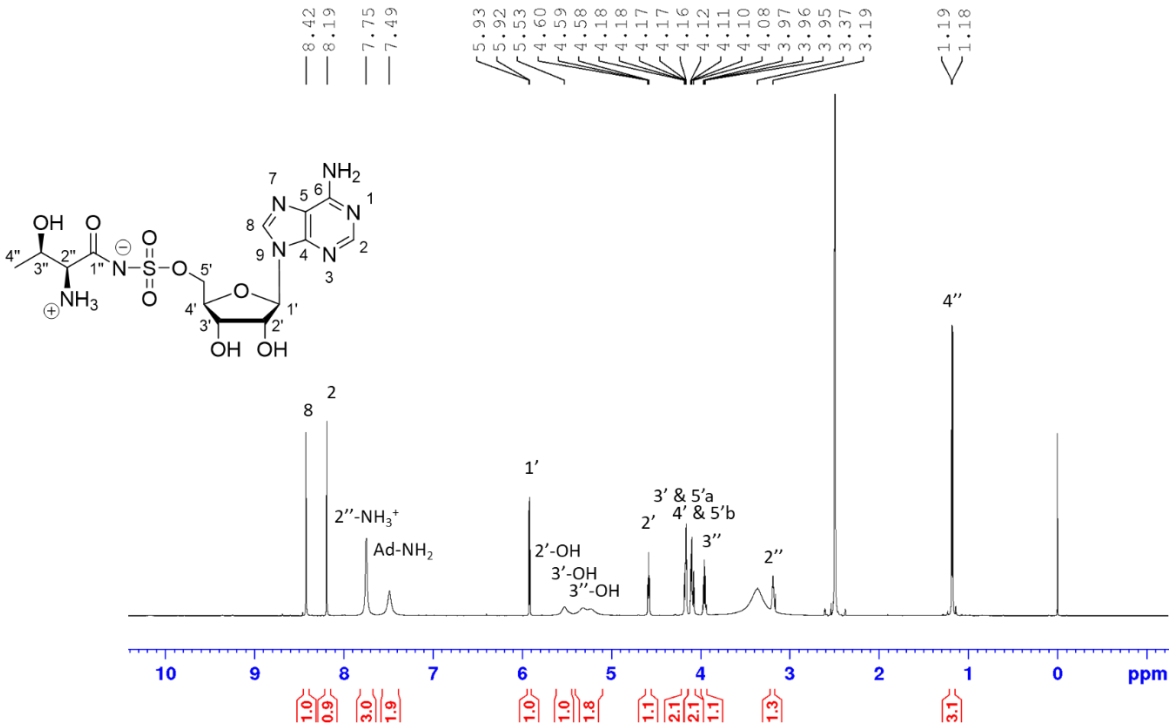
Supplemental Figure 16. Borrelidin calibration curve. Plotted known concentrations of borrelidin against their peak area at 254nm, assessed by HPLC as outlined in section 2.6.1.



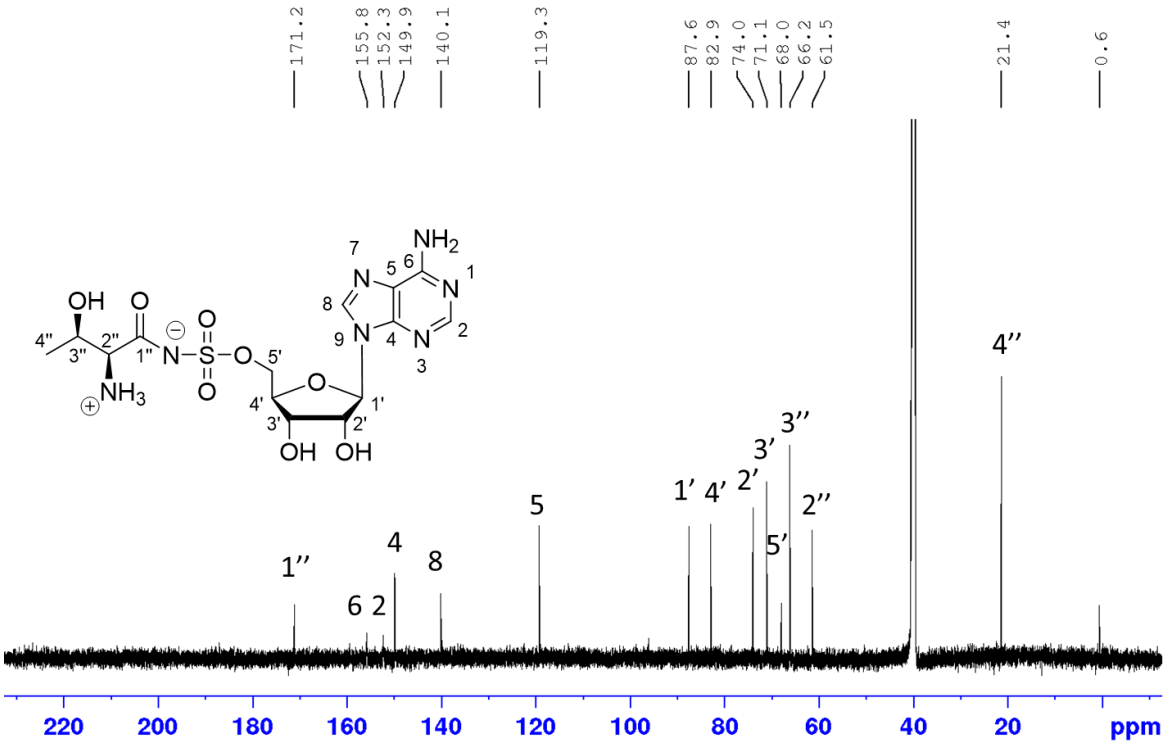
Supplemental Figure 17. ^{1}HNMR spectrum of 2',3'-*O*-(1-methylethylidene)-5'-*O*-sulfamoyladenosine. 600 MHz in DMSO-d_6.



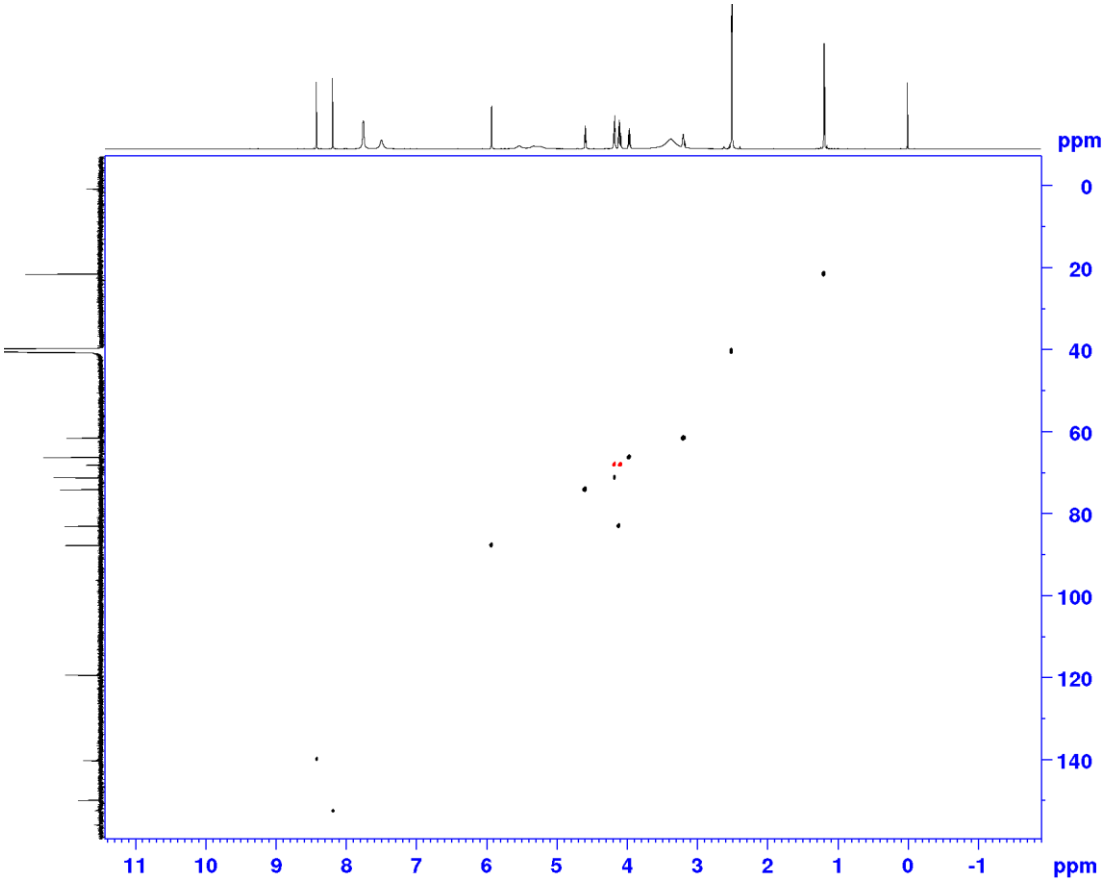
Supplemental Figure 18. ^{13}CNMR spectrum of 2',3'-*O*-(1-methylethylidene)-5'-*O*-sulfamoyladenosine. 600 MHz in DMSO-d_6.



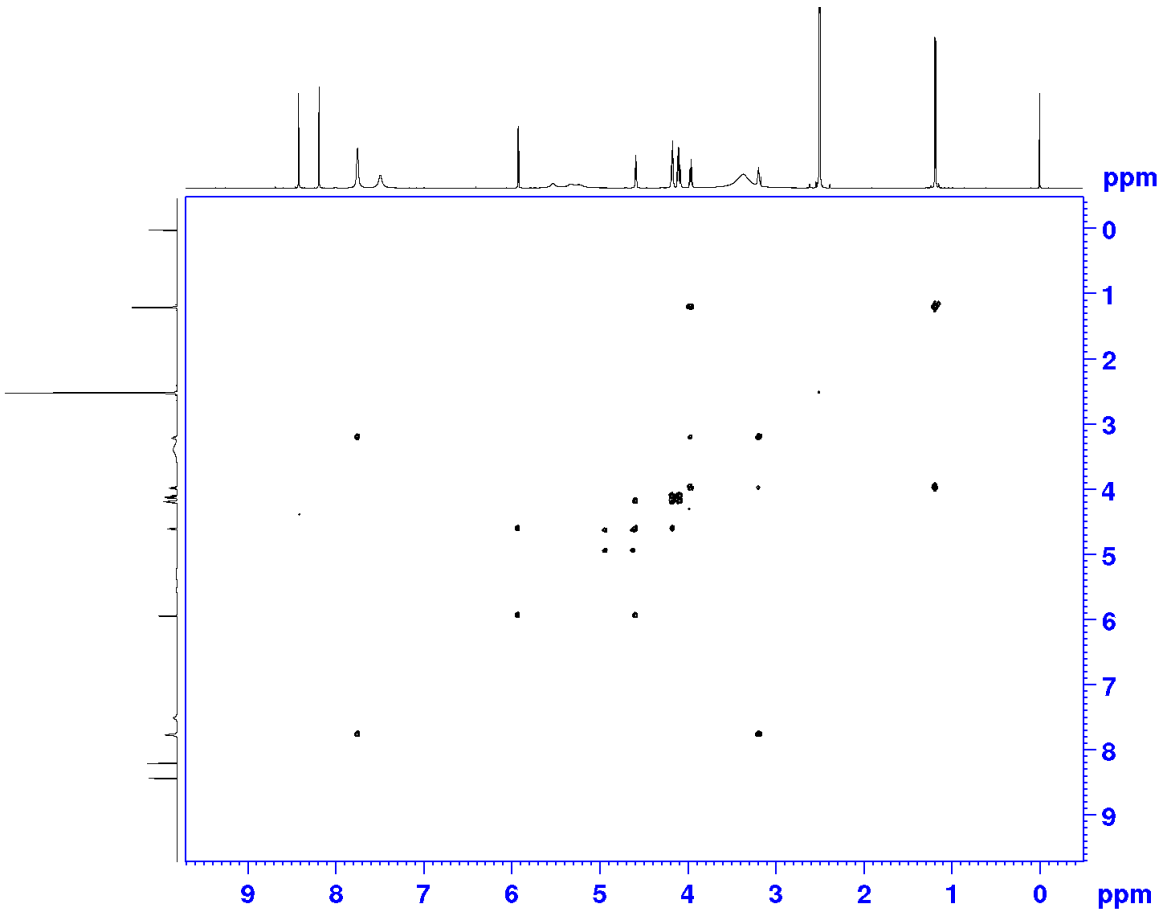
Supplemental Figure 19. ^{1}HNMR spectrum of ThrSAA. 600 MHz in DMSO-d_6.



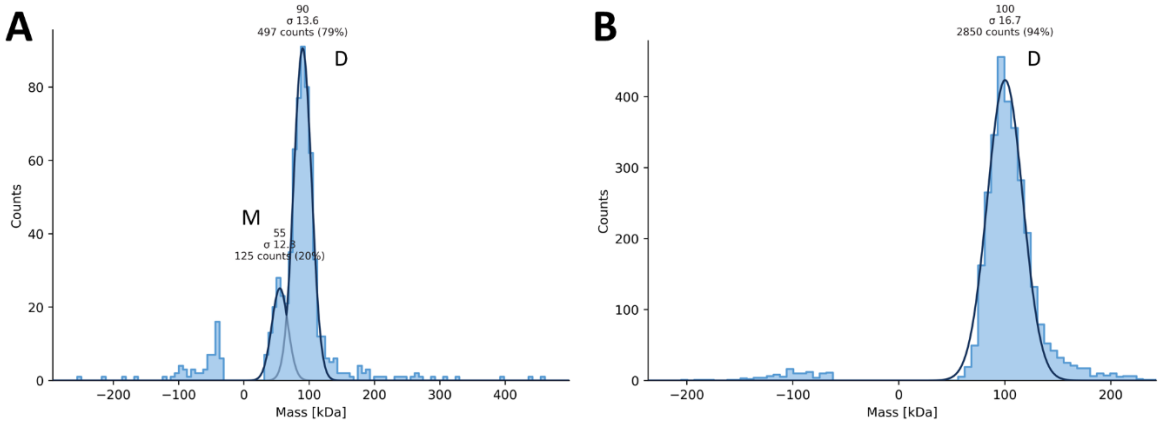
Supplemental Figure 20. ^{13}CNMR spectrum of ThrSAA. 600 MHz in DMSO-d_6.



Supplemental Figure 21. HSQC spectrum of ThrSAA. 600 MHz in DMSO-d_6.



Supplemental Figure 22. COSY spectrum of ThrSAA. 600 MHz in DMSO-d_6.



Supplemental Figure 23. Increased dilution leads to increase cproportion of the monomer. Histograms of A) ΔN BorO at 50nM, B) ΔN-BorO at 100nM. The average predicted mass in kDa is shown above each peak. Proteins were measured in PBS. Peaks align to the sizes of the monomer (labelled with M) and the dimer (labelled with D).

Supplemental Table 1. Structure solution statistics for BorO:Borrelidin, EcThrRS:ThrSAA and EcThrRS:Borrelidin.

Data set	**BorO:Borrelidin**	**EcThrRS:ThrSAA**	**EcThrRS:Borrelidin**
Data Collection			
Beamline	DLS-i04	DLS-i04	DLS-i04
Space group	P 4 1	P 21 21 21	P 21 21 21
a / b / c (Å)	77.26 / 77.26 / 158.98	86.75 / 108.68 / 112.67	94.57 / 107.23 / 109.16
α / β / γ (°)	90 / 90 / 90	90 / 90 / 90	90 / 90 / 90
Wavelength (Å)	0.999900	1.265153	0.979515
Resolution range (Å)	69.59 - 2.20	68.83 - 1.80	54.58 - 1.90
Inner (Å)	2.27 – 2.20	1.83 – 1.80	1.93 – 1.90
Total reflections	1161029	3369088	1194693
Unique reflections	47217	96823	87668
Completeness (%)	100	97.6	99.6
Wilson B-factor	38.8	26.8	28.0
R_{merge}	0.17	0.11	0.12
CC(1/2)	0.640	0.838	0.711
I/σI	0.47	1.55	1.41
Refinement			
Reflections used in refinement	47148	96737	87583
Reflections used for R_{free}	2385	5082	4445
R_{work} (%)	0.188	0.174	0.165
R_{free} (%)	0.225	0.199	0.195
No. non-hydrogen atoms	6698	7105	7238
Macromolecules	6409	6607	6573
Ligands	89	175	143
Solvent	200	323	522
Protein Residues	800	800	795
RMS (bonds)	0.0080	0.0108	0.0101

RMS (angles)	1.471	1.597	1.602
Average B factor (Å^2)	53.0	26.0	26.0
Macromolecules	53.7	24.77	25.6
Ligands	61.83	46.57	48.12
Solvent	43.13	35.92	39.49
Ramachandran plot Favoured (%)	98.00	98.62	99.25
Allowed (%)	2.00	1.38	0.75
Outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	1.53	1.96	1.83
Clashscore	2.91	5.15	4.43

Supplemental Table 2. Structure solution statistics for EcMut5:Thr, EcMut5:Borrelidin and EcMut11:Borrelidin/Thr.

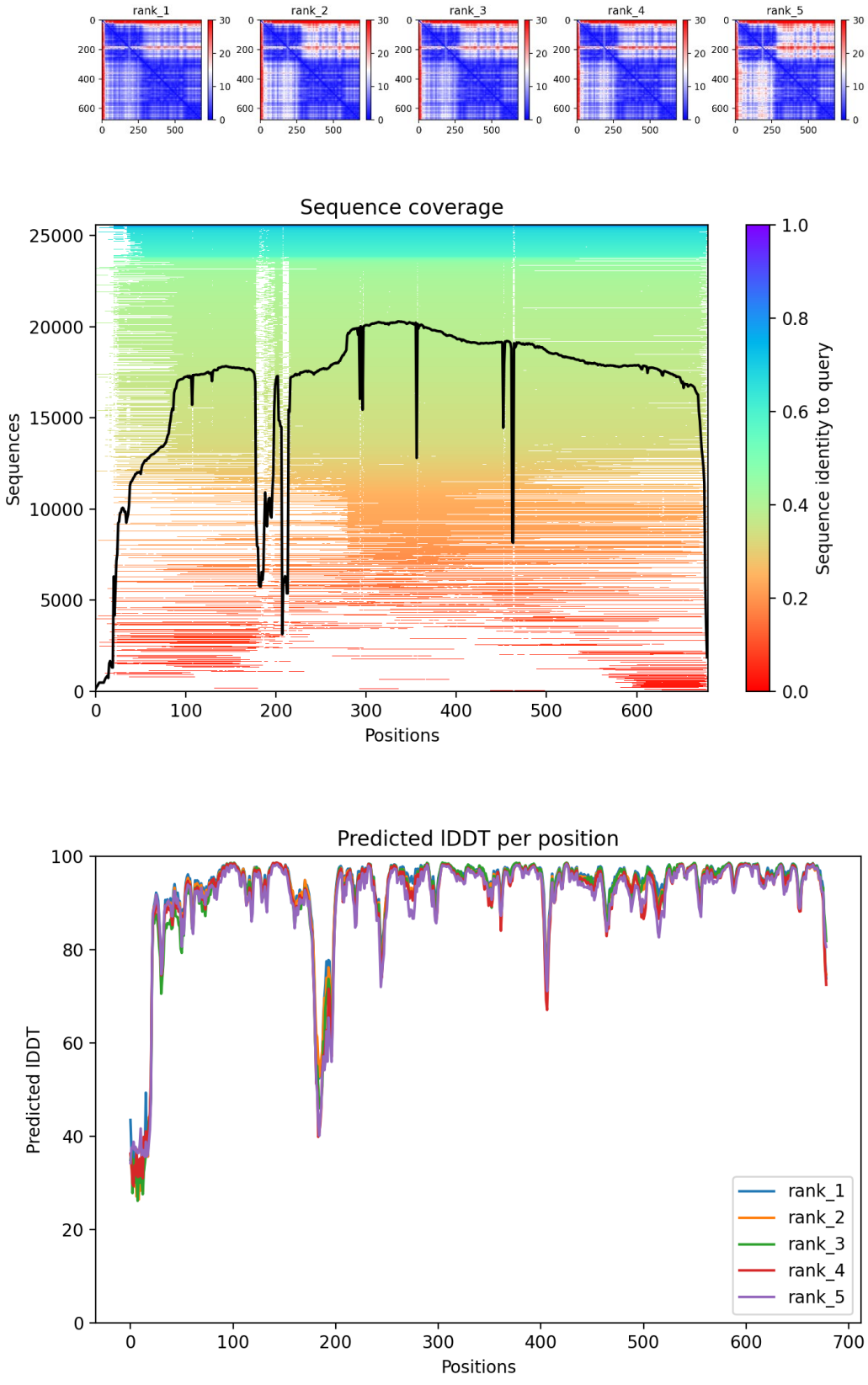
Data set	**EcMut5:Thr**	**EcMut5:Borrelidin**	**EcMut11**
Data Collection			
Beamline	DLS-i04	DLS-i04	DLS-i04
Space group	P 21 21 21	P 21 21 21	P 21 21 21
a / b / c (Å)	89.45 / 108.67 / 112.44	94.41 / 106.68 / 108.51	101.69 / 118.10 / 171.18
α / β / γ (°)	90 / 90 / 90	90 / 90 / 90	90 / 90 / 90
Wavelength (Å)	0.979499	0.979499	0.979499
Resolution range (Å)	70.10 - 2.70	59.31 – 2.20	70.37- 2.85
Inner (Å)	2.83 - 2.70	2.26 – 2.20	2.94 – 2.85
Total reflections	414513	758849	657569
Unique reflections	30822	56349	48926
Completeness (%)	100	100	99.8
Wilson B-factor	78.9	43.8	56.0
R_{merge}	0.14	0.21	0.22
CC(1/2)	0.572	0.428	0.792
I/σI	1.22	1.12	1.76
Refinement			

Reflections used in refinement	30747	56275	48825
Reflections used for R_{free}	1508	2764	2411
R_{work} (%)	0.210	0.197	0.210
R_{free} (%)	0.261	0.238	0.264
No. non-hydrogen atoms	6486	6835	13152
Macromolecules	6467	6595	13039
Ligands	12	108	74
Solvent	7	132	39
Protein Residues	797	797	1593
RMS (bonds)	0.0068	0.0097	0.0129
RMS (angles)	1.099	1.60	2.00
Average B factor ($Å^2$)	60.00	39.0	43.0
Macromolecules	61.17	38.97	43.86
Ligands	99.59	49.09	59.15
Solvent	72.63	44.46	40.86
Ramachandran plot Favoured (%)	96.49	97.99	97.68
Allowed (%)	3.51	2.01	2.32
Outliers (%)	0.38	0.00	0.00
Rotamer outliers (%)	3.47	3.51	6.84
Clashscore	4.40	4.83	4.76

Supplemental Table 3. Structure solution statistics for EcMutQ:Borrelidin.

Data set	**EcMutQ:Borrelidin**
Data Collection	
Beamline	DLS-i04
Space group	P 21 21 21
a / b / c (Å)	94.27 / 107.69 / 109.84
α / β / γ (°)	90 / 90 / 90

Wavelength (Å)	0.979499
Resolution range (Å)	59.66 - 1.50
Inner (Å)	1.53 – 1.50
Total reflections	4840096
Unique reflections	178623
Completeness (%)	99.9
Wilson B-factor	19.9
R_{merge}	0.11
CC(1/2)	0.720
I/σI	1.40
Refinement	
Reflections used in refinement	178510
Reflections used for R_{free}	8846
R_{work} (%)	0.165
R_{free} (%)	0.183
No. non-hydrogen atoms	7673
Macromolecules	6848
Ligands	148
Solvent	677
Protein Residues	799
RMS (bonds)	0.0123
RMS (angles)	1.773
Average B factor ($Å^2$)	19.0
Macromolecules	17.91
Ligands	33.93
Solvent	33.29
Ramachandran plot Favoured (%)	99.25
Allowed (%)	0.75
Outliers (%)	0.00
Rotamer outliers (%)	2.02
Clashscore	4.69



Supplemental Figure 24. AlphaFold prediction confidence and coverage information for SpThrRS

Supplemental Table 4. BmObaO:Thr, EcMutM:Borrelidin and BmMutL:Borrelidin structure solution information

Protein	**EcMutM:Borrelidin**	**BmMutL:Borrelidin**
Data Collection		
Beamline	DLS-i04	DLS-i04
Space group	P 21 21 21	C 1 2 1
a / b / c (Å)	94.37 / 107.14 / 108.54	145.38 / 72.70 / 108.04
α / β / γ (°)	90 / 90 / 90	90 / 126.27 / 90
Wavelength (Å)	0.979499	0.979499
Resolution range (Å)	53.57 - 1.60	61.97 - 2.55
Outer (Å)	1.63 – 1.60	2.66 – 2.55
Total reflections	1986081	204386
Unique reflections	143144	29998
Completeness (%)	98.4	99.0
Wilson B-factor	22.0	50.7
R_{merge}	0.09	0.12
CC(1/2)	0.631	0.646
I/σI	1.39	1.32
Refinement		
Reflections used in refinement	143043	31677
Reflections used for R_{free}	7122	1560
R_{work} (%)	0.168	0.212
R_{free} (%)	0.190	0.255
No. non-hydrogen atoms	7351	6343
Macromolecules	6719	6262
Ligands	130	72
Solvent	502	9
Protein Residues	794	786
RMS (bonds)	0.0120	0.0084
RMS (angles)	1.736	1.505

Average B factor (Å^2)	21.0	43.0
Macromolecules	20.01	44.54
Ligands	37.39	45.12
Solvent	32.80	49.98
Ramachandran plot Favoured (%)	99.12	96.69
Allowed (%)	0.88	2.66
Outliers (%)	0.00	0.38
Rotamer outliers (%)	1.64	4.66
Clashscore	5.59	7.63

Supplemental Table 5. EcThrRS:obafluorin and ObaO:obafluorin structure solution information.

Data set	**EcThrRS:Obafluorin**	**PfObaO:Obafluorin**
Data collection and processing		
Magnification	96 000	120 000
Voltage (kV)	300	300
Electron exposure (e^-/Å^2)	34.90	34.84
Defocus range (μm)	3.0-1.2	3.0-0.9
Pixel size (Å)	0.86	0.68
Symmetry imposed	C2	C2
Initial particle images	2 608 768	2 667 556
Final particle images	254 054	224 636
Map resolution (Å)	2.85	2.95
FSC threshold	0.143	0.143
Map resolution range (Å)	2.5-3	N/D
Refinement		
Initial model used	EcThrRS AlphaFold	PfObaO AlphaFold
Model map FSC (masked)	2.97	N/D
Model comparison		
Non-hydrogen atoms	6662	6604
Protein residues	810	804
Ligands	4	4

B-factors (Å^2)		
Protein	34.04	67.20
Ligand	35.41	77.96
RMS deviations		
Bond lengths (Å)	0.004	0.005
Bond angles (°)	0.616	0.968
Validation		
MolProbity score	0.93	1.11
Clashscore	0.53	1.78
Poor rotamers (%)	0.7	0.44
Ramachandran plot		
Favoured (%)	97.77	97.12
Allowed (%)	2.23	2.88
Disallowed (%)	0.00	0.00

Supplemental Table 6, Single point mutations observed in spontaneous resistant mutants.

Sample	**Mutation in DNA**	**Mutation in Protein**	**Residue in EcThrRS**	**Residue in ObaOs**	**Subdomain**	**Significance**
8.3	T334A	F112I	L	F/L	Editing	Minimal
8.3	G354A	Silent, A118	E	P	Editing	None
8.3	C356T	A119V	A	L/A/R	Editing	Minimal
8.1	C369A	Silent, R123	R	E	Editing	None
8.3	C373A	H125N	H	T/K/A/R	Editing	Minimal
8.3	G376A	A126T	E	R/A	Editing	Minimal
8.3	A628G	K210E	N	N/S/D	Editing	Minimal
8.1	C855G	Silent, R285	K	R/V	C1	None
8.1	C870T	Silent, L290	Q	R/Q	C1	None
8.1	A880G	T294A	G	S/A	C2	Minimal
8.1	C938T	A313V	K	K/Q	C2	Minimal
8.1	G948T	M316I	M	M	C2	Minimal
8.1, 8.3	C1041T	Silent, Y347	Y	H/Y	C2	None

SB2, SB4, SB5	G1384A	G462S	G	S	C4	Conserved in ObaOs
SB2, SB4, JL8.1, JL8.3	G1384T	G462C	G	S	C4	Similar to in ObaOs

Supplemental Table 7. Table of BLAST hits and putative functions for the BGC from *Actinobacterium* sp. OV320

Gene	Closest non-hypothetical BLAST hit (where possible)	Query Cover (%)	Percentage Identity (%)	Putative Function
Gene 1	Phage holin family protein, *Streptomyces* sp. 303MFCol5.2	90	97.76	Phage associated porin
Gene 2	DUF3618 domain containing protein, *Streptomyces asoensis*	100	94.39	Unknown
Gene 3	Winged helix-turn-helix domain containing protein, *Streptomyces* sp. 303MFCol5.2	100	86.42	DNA binding, regulation
Gene 4	SRPBCC domain containing protein, *Streptomyces* sp. 303MFCol5.2	100	98.38	Regulation
Gene 5	Endonuclease/exonuclease/phosphatase, *Streptomyces* sp. CMB-FB	100	98.03	Regulation
Gene 6	Hypothetical Protein, *Streptomyces* sp. SLBN-115	100	99.37	Unknown
Gene 7	DUF839 domain-containing protein, *Streptomyces asoensis*	100	99.18	Unknown
Gene 8	TROVE domain-containing protein, *Streptomyces* sp. 303MFCol5.2	100	98.49	RNA binding and stabilisation
Gene 9	SgcJ/EcaC family oxidoreductase, unclassified *Streptomyces*	100	99.19	Oxidation/Reduction
Gene 10	SDR family oxidoreductase, *unclassified Streptomyces*	100	97.89	Oxidation/Reduction

Gene 11	LysR family transcriptional regulator, *Streptomyces* sp. CMB-FB	100	95.62	Regulation
Gene 12	Muconolactone Delta-isomerase family protein, unclassified *Streptomyces*	100	98.04	Isomerisation
Gene 13	TetR/AcrR family transcriptional regulator, *Streptomyces asoensis*	98	94.63	Regulation
Gene 14	GntR family transcriptional regulator, *Streptomyces* sp. Root1310	100	98.62	Regulation
Gene 15	1-aminocyclopropane 1-carboxylate deaminase, unclassified *Streptomyces*	100	99.41	Deamination
Gene 16	None	-	-	Unknown
Gene 17	Hypothetical protein	94	40.00	Unknown
Gene 18	Amino acid adenylation domain-containing protein, *Allokutzneria albata*	98	56.32	Peptide synthesis
Gene 19	Phosphopantetheine-binding protein, *Streptomyces* sp. SM11	90	37.66	Peptide synthesis
Gene 20	MFS Transporter, *Kutzneria buriramensis*	90	69.10	Export/Self-resistance
Gene 21	NRPS, *Kutzneria buriramensis*	99	61.22	Peptide synthesis
Gene 22	NRPS, *Kutzneria buriramensis*	99	67.12	Peptide synthesis
Gene 23	MbtH family NRPS accessory protein, unclassified *Micromonospora*	97	56.52	Peptide synthesis
Gene 24	LLM class flavin-dependent oxidoreductase, *Catenulispora pinistramenti*	98	72.87	Monooxygenase
Gene 25	Alpha/beta fold hydrolase, *Streptomyces lydicamycinicus*	97	57.96	Possible Thioesterase

Gene 26	ThrRS, *Pseudonocardiales bacterium*	98	76.15	Self-resistance/Borrelidin resistance
Gene 27	Helix-turn-helix transcriptional regulator, *Nonomuraea dietziae*	100	61.78	Regulation
Gene 28	Hypothetical protein, *Streptomyces* sp. 303MFCol5.2	100	88.14	Unknown
Gene 29	MFS transporter, *Streptomyces* sp. 303MFCol5.2	100	97.19	Export/Self-resistance
Gene 30	TetR/AcrR family transcriptional regulator, unclassified *Streptomyces*	100	99.47	Regulation
Gene 31	TerD family protein, *Streptomyces* sp. 303MFCol5.2	100	99.24	Metal sensing/Nucleoside formation
Gene 32	Zinc-dependent alcohol dehydrogenase family protein, unclassified *Streptomyces*	100	98.48	Oxidoreductase
Gene 33	Carbohydrate ABC transporter permease, unclassified *Streptomyces*	100	100	Sugar Export
Gene 34	Sugar ABC transporter permease, *Streptomyces* sp. 303MFCol5.2	100	99.36	Sugar Export
Gene 35	Sugar ABC transporter substrate binding protein, unclassified *Streptomyces*	100	98.90	Sugar Export
Gene 36	DeoR/GlpR family DNA-binding transcriptional regulator, *Streptomyces* sp. CMB-FB	100	99.61	Regulation
Gene 37	NAD(P)-dependent oxidoreductase, *Streptomyces* sp. 303MFCol5.2	100	97.81	Oxidation/Reduction
Gene 38	5-dehydro-4-deoxyglucarate dehydratase, unclassified *Streptomyces*	100	99.68	Dehydratase
Gene 39	Hypothetical protein, unclassified *Streptomyces*	100	97.19	Unknown
Gene 40	MFS transporter, *Streptomyces* sp. SLBN-115	100	91.26	Transport/Self-resistance

Supplemental Table 8. Table of BLAST hits and putative functions for the BGC from *Tetrasphaera* sp. Aved_18-Q3-R54-62_MAXAC.378

Gene	Closest non-hypothetical BLAST hit	Query Cover (%)	Percentage Identity (%)	Putative Function
Gene 1	DNA topoisomerase subunit B, *Phycicoccus endophyticus*	97	84.47	DNA replication/topology control
Gene 2	Peptidase inhibitor family I36 protein, *Micromonospora* sp. MP36	82	33.56	Peptidase regulation
Gene 3	LuxR C-terminal-related transcriptional regulator, *Phycicoccus flavus*	99	61.88	Regulation
Gene 4	None	-	-	Unknown
Gene 5	None	-	-	Unknown
Gene 6	None	-	-	Unknown
Gene 7	None	-	-	Unknown
Gene 8	Transglycosylase domain-containing protein, *Actinomadura flavalba*	28	76.25	Peptidoglycan biosynthesis
Gene 9	4'-phosphopantetheinyl transferase superfamily protein, *Phycicoccus flavus*	82	63.53	ACP phosphopantetheine loading
Gene 10	4'-phsphopantetheinyl transferase superfamily protein, *Phycicoccus flavus*	100	61.07	ACP phosphopantetheine loading
Gene 11	Sigma-60 family RNA polymerase sigma factor, *Phycicoccus flavus*	47	58.60	DNA transcription initiation, regulation
Gene 12	Hypothetical protein, *Flexivirga oryzae*	98	50.51	Unknown

Gene 13	Acetyl-CoA carboxylase carboxyltransferase subunit beta, *Deferribacteres bacterium*	62	47.55	Acetyl-CoA formation
Gene 14	Acetyl-CoA carboxylase carboxyltransferase subunit alpha, *Actinobacteria bacterium*	96	56.28	Acetyl-CoA formation
Gene 15	Acetyl-CoA carboxylase biotin carboxyl carrier protein, *Candidatus Aminicenantes bacterium*	64	60.24	Acetyl-CoA formation
Gene 16	Acetyl-CoA carboxylase biotin carboxylase subunit, *Kribbella albertanoniae*	96	66.97	Acetyl-CoA formation
Gene 17	Hypothetical protein, *Flexivirga oryzae*	65	66.00	Unknown
Gene 18	ABC transporter permease, *Phycicoccus flavus*	100	63.66	Export/Self-resistance
Gene 19	ABC transporter permease, *Phycicoccus permease*	100	60.80	Export/Self-resistance
Gene 20	ABC transporter ATP-binding protein, *Phycicoccus flavus*	100	74.10	Export/Self-resistance
Gene 21	DUF2505 domain-containing protein, *Flexivirga oryzae*	92	51.22	Unknown
Gene 22	NRPS, *Phycicoccus flavus*	100	69.36	Peptide synthesis
Gene 23	Alpha/beta fold hydrolase, *Phycicoccus flavus*	99	50.61	Thioesterase
Gene 24	Beta-ketoacyl-synthase N-terminal like domain containing protein, *Paenibacillus* sp. 19GGS1-52	96	30.45	Fatty acid synthesis

Gene 25	Beta-ketoacyl synthase N-terminal-like domain containing protein, *Phycicoccus flavus*	74	78.49	Fatty acid synthesis
Gene 26	Beta-ketoacyl synthase N-terminal-like domain containing protein, *Phycicoccus flavus*	94	54.51	Fatty acid synthesis
Gene 27	Beta-ketoacyl synthase N-terminal like domain containing protein, *Flexivirga oryzae*	97	61.37	Fatty acid synthesis
Gene 28	Beta-ketoacyl synthase N-terminal like domain containing protein, *Flexivirga oryzae*	90	84.76	Fatty acid synthesis
Gene 29	SDR family NAD(P)-dependent oxidoreductase, *Flexivirga oryzae*	100	51.78	Fatty acid synthesis
Gene 30	ACP S-malonyltransferase, *Phycicoccus flavus*	97	65.54	Fatty acid synthesis
Gene 31	Acyl carrier protein, *Phycicoccus flavus*	97	62.65	Fatty acid synthesis
Gene 32	Condensation domain-containing protein, *Flexivirga oryzae*	88	35.11	Peptide synthesis
Gene 33	NRPS, *Phycicoccus flavus*	99	65.91	Peptide synthesis
Gene 34	Condensation domain-containing protein, *Paenibacillus spiritus*	76	39.44	Peptide synthesis
Gene 35	Condensation domain-containing protein, *Listeria monocytogenes*	84	30.12	Peptide synthesis

Gene 36	Phosphopantetheine-binding protein, *Phycicoccus flavus*	89	49.36	Peptide synthesis
Gene 37	None	-	-	Unknown
Gene 38	Anthranilate synthase family protein, *Tetrasphaera jenkinsii*	99	77.27	Precursor biosynthesis
Gene 39	ThrRS, *Tetrasphaera jenkinsii*	99	90.48	Borrelidin resistance/self-resistance
Gene 40	M4 family metallopeptidase, *Actinomycetia bacterium*	96	60.15	Proteolysis
Gene 41	DUF 1697 domain containing protein, *Terracocccus luteus*	91	63.48	Unknown
Gene 42	RNA-binding protein, *Branchiibus* sp. NY16-3462-2	84	53.01	RNA-binding
Gene 43	DNA replication/repair protein RecF, *Tetrasphaera* sp. F2B08	98	71.50	DNA replication/repair
Gene 44	Decarboxylating 6-phosphogluconate dehydrogenase, *Actinomycetia bacterium*	90	87.58	Pentose phosphate pathway
Gene 45	DNA polymerase III subunit beta, *Tetrasphaera jenkinsii*	100	93.39	DNA replication
Gene 46	Chromosomal replication initiator protein DnaA, *Tetrasphaera jenkinsii*	95	87.29	DNA replication
Gene 47	50S ribosomal protein L34, *Tetrasphaera jenkinsii*	100	97.78	Translation
Gene 48	Ribonuclease P protein component, *Tetrasphaera jenkinsii*	94	69.03	RNAse

Gene 49	Membrane protein insertion efficiency factor YidD, *Tetrasphaera jenkinsii*	100	87.36	Insertion of membrane proteins into the membrane
Gene 50	Membrane protein insertase YidC, *Tetrasphaera jenkinsii*	98	86.47	Insertion of membrane proteins into the membrane
Gene 51	RNA-binding protein, *Tetrasphaera jenkinsii*	84	84.05	RNA binding
Gene 52	Ribosomal RNA small subunit methyltransferase G, *Tetrasphaera jenkinsii* Ben 74	94	63.13	Ribosomal RNA methylation

Supplemental Table 9. Table of BLAST hits and putative functions for the BGC from *Pseudonocardia* sp.

Gene	**Closest non-hypothetical BLAST hit**	**Query Cover (%)**	**Percentage Identity (%)**	**Putative Function**
Gene 1	WhiB family transcriptional regulator, *Pseudonocardia* sp.	100	97.62	Regulation
Gene 2	Histidine kinase N-terminal domain-containing protein, *Pseudonocardiales bacterium*	100	91.92	Regulation
Gene 3	Biotin/lipoyl-binding carrier protein, *Pseudonocardiales bacterium*	100	98.59	Metabolism
Gene 4	Myocothiol system anti-sigma-R factor, *Pseudonocardiales bacterium*	94	82.22	Regulation

Gene 5	Sigma-70 family RNA polymerase sigma factor, *Pseudonocardia* sp.	87	93.60	Regulation
Gene 6	Alpha/beta hydrolase, *Pseudonocardiales bacterium*	96	79.72	Hydrolysis
Gene 7	SOS response- associated peptidase, *Pseudonocardiales bacterium*	92	100	DNA damage response
Gene 8	3-phosphoshikimate 1-carboxyvinyltransferase, *Pseudonocardiales bacterium*	100	87.44	Shikimate pathway
Gene 9	Ribosome small subunit-dependent GTPase A, *Pseudonocardiales bacterium*	100	92.04	Translation
Gene 10	Hypothetical protein, *Pseudonocardiales bacterium*	68	100	Unknown
Gene 11	Class III lanthionine synthetase LanKC, *Pseudonocardiales bacterium*	97	88.21	Lanthionine bond formation
Gene 12	SapB/AmfS family lanthipeptide, *Pseudonocardiales bacterium*	100	76.92	Lanthipeptide precursor peptide
Gene 13	ABC transporter ATP-binding protein, *Pseudonocardiales bacterium*	100	100	Export/Self-resistance

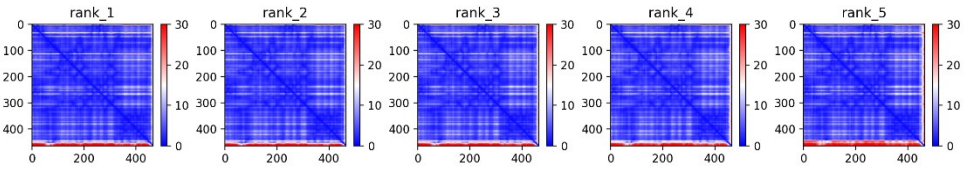
Gene 14	ABC transporter ATP-binding protein, *Pseudonocardiales bacterium*	99	91.71	Export/Self-resistance
Gene 15	L-glutamate gamma-semialdehyde dehydrogenase, *Pseudonocardiales bacterium*	99	91.21	Urea cycle
Gene 16	Hypothetical protein, *Pseudonocardiales bacterium*	100	88.02	Unknown
Gene 17	Rv3235 family protein, *Pseudonocardia* sp.	84	71.33	Unknown
Gene 18	ThrRS, *Pseudonocardiales bacterium*	96	99.85	Borrelidin resistance/self-resistance
Gene 19	Cold shock domain-containing protein, *Cellulomonas* sp.	91	35.20	Regulation

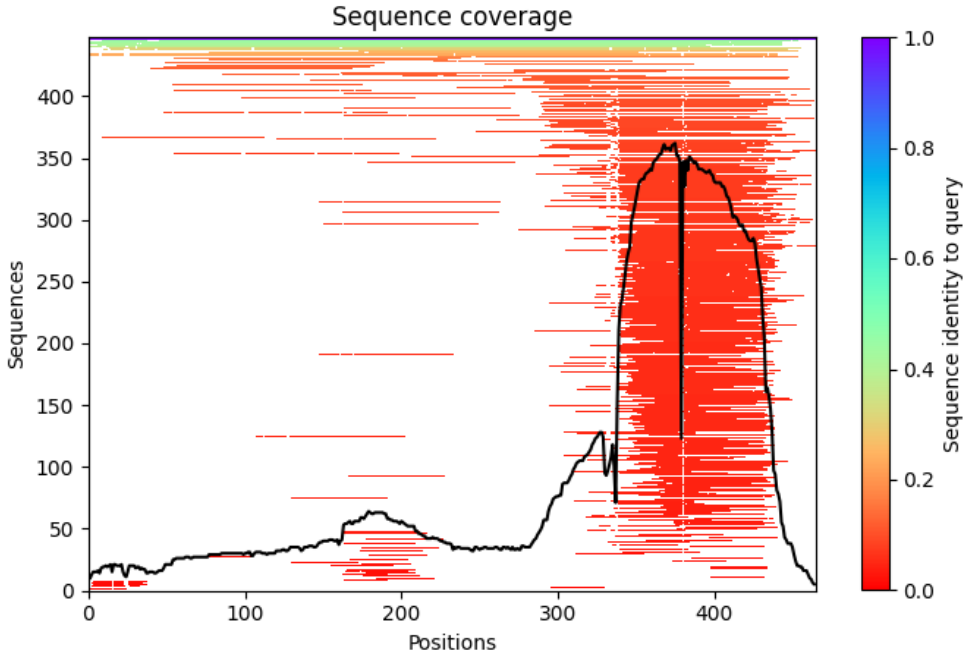
Supplemental Table 10. Table of BLAST hits and putative functions for the BGC from *Frankia* sp. Cj5

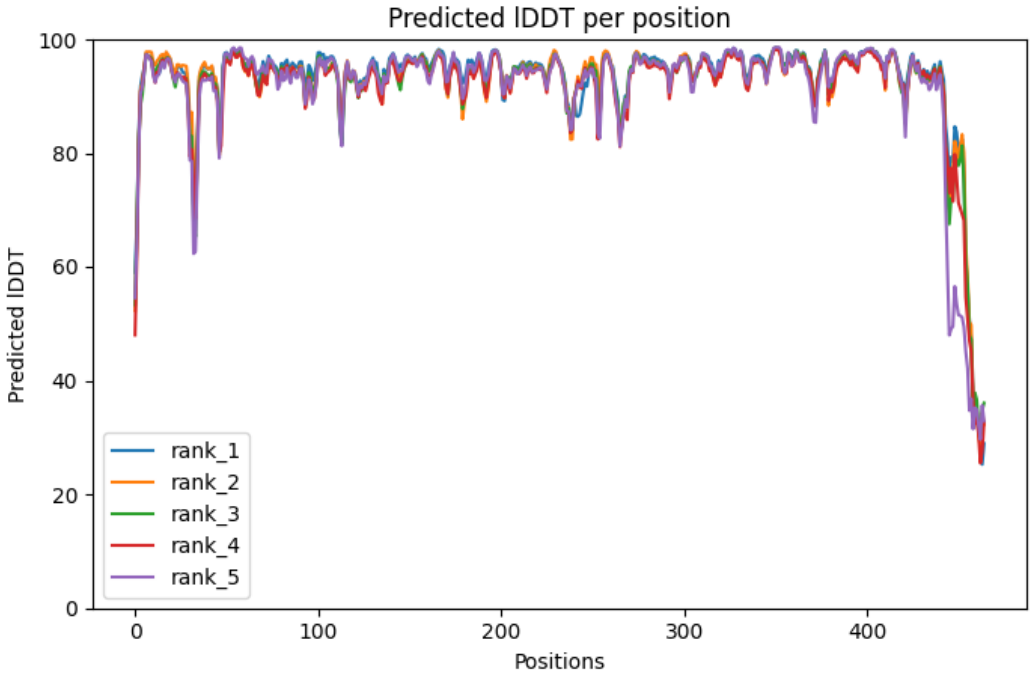
Gene	**Closest non-hypothetical BLAST hit**	**Query Cover (%)**	**Percentage Identity (%)**	**Putative Function**
Gene 1	Type II toxin-antitoxin system VapC family toxin, *Pseudonocardiales bacterium*	99	69.77	Toxin-Antitoxin system-metabolic flux regulation[284]
Gene 2	ThrRS, *Frankia* sp. CIT1	100	99.11	Borrelidin resistance/self-resistance
Gene 3	AAA+ family ATPase	94	69.70	Protein Transport/Protease activity

Gene 4	FAD-dependent oxidoreductase, *Frankia* sp. CIT1	99	98.46	Oxidation/Reduction
Gene 5	Cytochrome P450, *Frankia* sp. CIT1	100	99.52	Monooxygenation
Gene 6	Scr1 family system antitoxin-like transcriptional regulator, unclassified *Frankia*	100	100	Regulation
Gene 7	DUF397 domain-containing protein, *Frankia* sp. KB5	100	79.74	Regulation
Gene 8	GNAT family N-acetyltransferase, *Streptomyces* sp. PT12	96	54.59	Tailoring/Self-resistance
Gene 9	None	-	-	Unknown
Gene 10	Thiopeptide-type bacteriocin biosynthesis protein, *Frankia* sp. Ccl49	96	66.55	Lanthipeptide synthesis
Gene 11	Lanthionine synthetase C family protein, *Frankia* sp. R82	100	75.24	Lanthipeptide synthesis
Gene 12	Thiopeptide-type bacteriocin biosynthesis protein, *Frankia* sp. R82	99	70.29	Lanthipeptide synthesis
Gene 13	FxLD family lanthipeptide, *Frankia*	100	78.26	Lanthipeptide synthesis
Gene 14	Methyltransferase, FxLD system, *Frankia* sp. Ccl49	100	73.53	Methyltransferase
Gene 15	PGPGW domain-containing protein, *Frankia* sp. CIT1	100	99.43	Two component system, regulation
Gene 16	Response regulator, *Frankia* sp. CIT1	100	99.16	Two component system, regulation

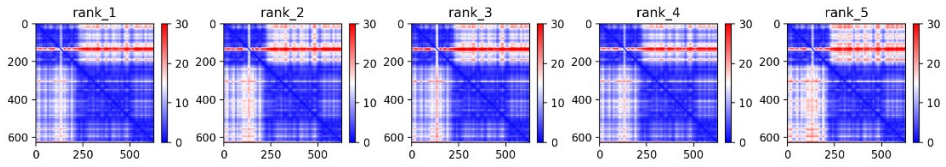
Gene 17	Potassium-transporting ATPase subunit C, *Frankia* sp. CiP3	99	96.25	Potassium transport
Gene 18	Potassium-transporting ATPase subunit KdpB, *Frankia* sp. CIT1	100	99.73	Potassium transport
Gene 19	Potassium-transporting ATPase subunit KdpA, *Frankia* sp. CiP3	100	99.10	Potassium transport
Gene 20	K(+) transporting ATPase subunit F, *unclassified Frankia*	100	96.55	Potassium transport
Gene 21	Hypothetical protein, *Frankia sp. CIT1*	100	98.68	Unknown
Gene 22	Enoyl-ACP reductase FabI, unclassified *Frankia*	98	99.61	Fatty acid synthesis

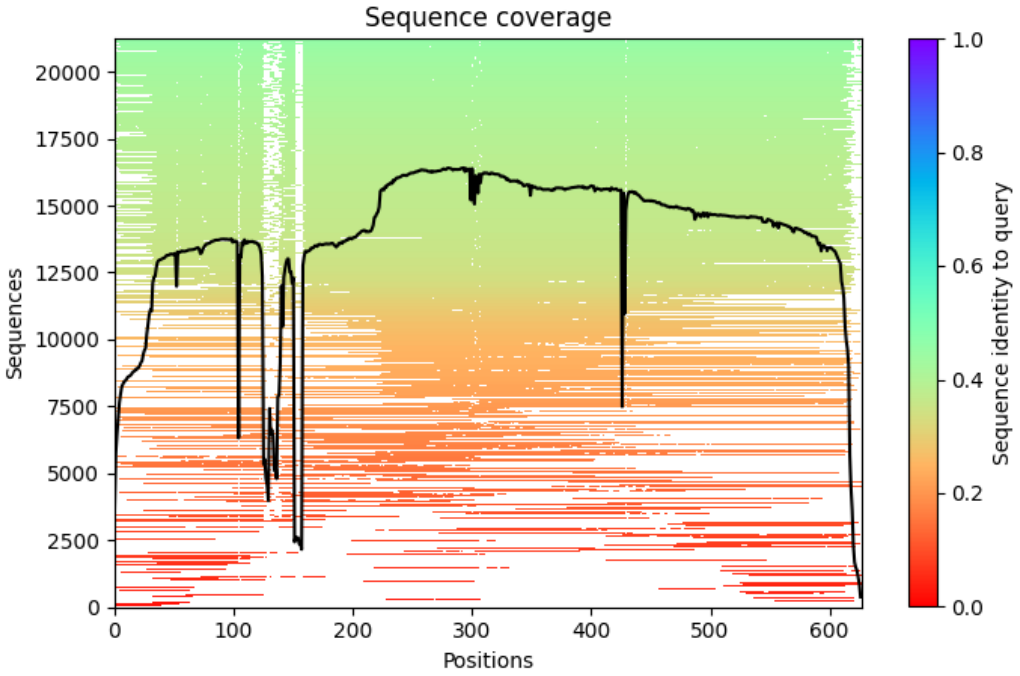


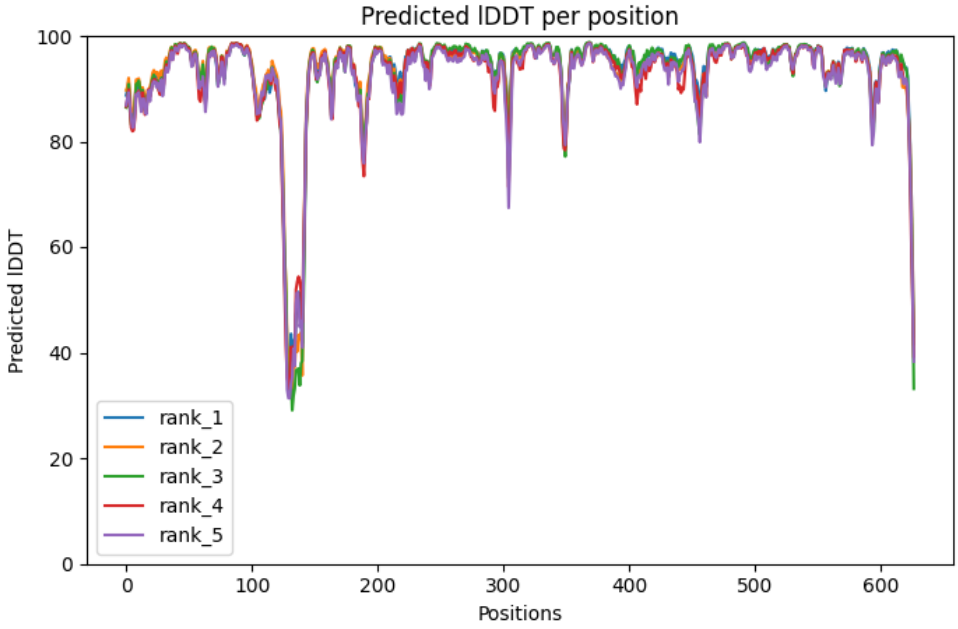




Supplemental Figure 25, AlphaFold prediction confidence and coverage information for the product of gene K from the compound 1 BGC.







Supplemental Figure 26, AlphaFold prediction confidence and coverage information for the product of gene O from the compound 1 BGC.

Supplemental Figure 17. Example streak plate overlay bioassays of M. sp. KC207 **on MYM. Left,** M. sp. KC207 grown on MYM overlayed with E. coli NR698. **Middle,** M. sp. KC207 grown on MYM overlayed with WT E. coli. **Right,** M. sp. KC207 grown on MYM overlayed with B. subtilis.

Supplemental Figure 18. Example streak plate overlay bioassays of M. sp. KC207 **on MAM. Left,** M. sp. KC207 grown on MAM overlayed with E. coli NR698. **Middle,** M. sp. KC207 grown on MAM overlayed with WT E. coli. **Right,** M. sp. KC207 grown on MAM overlayed with B. subtilis.

Supplemental Figure 19, example streak plate overlay bioassays of M. sp. KC207 **on GYM. Left,** M. sp. KC207 grown on GYM overlayed with E. coli NR698. **Middle,** M. sp. KC207 grown on GYM overlayed with WT E. coli. **Right,** M. sp. KC207 grown on GYM overlayed with B. subtilis.

Supplemental Figure 20. Example streak plate overlay bioassays of M. sp. KC207 **on DNA. Left,** M. sp. KC207 grown on DNA overlayed with E. coli NR698. **Middle,** M. sp. KC207 grown on DNA overlayed with WT E. coli. **Right,** M. sp. KC207 grown on DNA overlayed with B. subtilis.

Supplemental Figure 21. Example streak plate overlay bioassays of M. sp. KC207 **on Bennett's media. Left,** M. sp. KC207 grown on Bennett's media overlayed with E. coli NR698. **Middle,** M. sp. KC207 grown on Bennett's media overlayed with WT E. coli. **Right,** M. sp. KC207 grown on Bennett's media overlayed with B. subtilis.

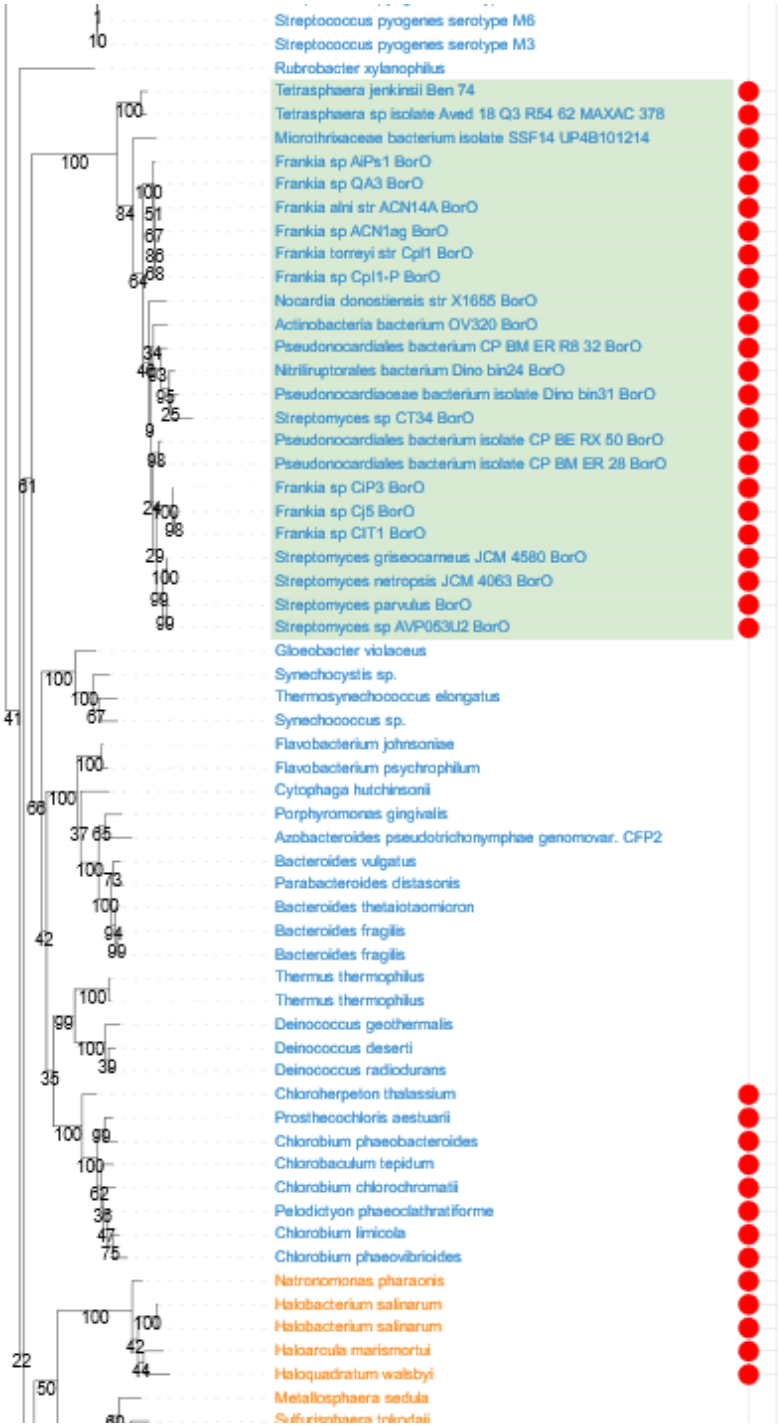
Supplemental Figure 22. Example streak plate overlay bioassays of M. sp. KC207 **on SFM.** Plates overlayed with resazurin to aid visualisation of zones of inhibition- blue zone around the M. sp. KC207 streak is the zone of clearance. **Left,** M. sp. KC207 grown on SFM overlayed with E. coli NR698. **Middle,** M. sp. KC207 grown on SFM overlayed with WT E. coli. **Right,** M. sp. KC207 grown on SFM overlayed with B. subtilis.

Supplemental Table 11. Table of BLAST hits and putative functions for the BGC from *Cellulomonas carbonis*

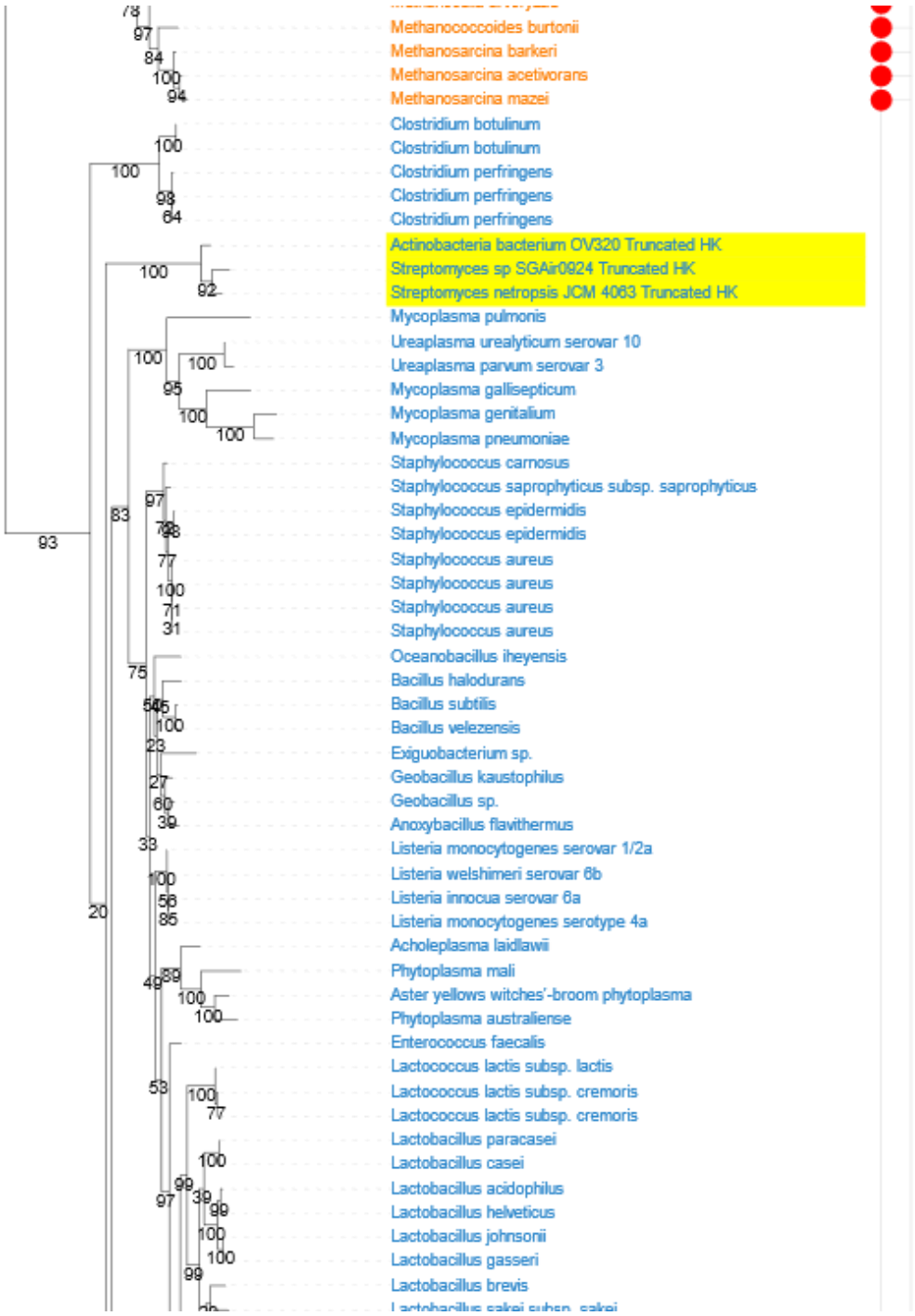
Gene	Closest non-hypothetical BLAST hit	Query Cover (%)	Percentage Identity (%)	Putative Function
Gene 1	M23 family metallopeptidase, *Actinotalea ferrariae*	92	82.30	Proteolysis
Gene 2	Hypothetical protein, *Actinotalea ferrariae*	97	68.70	Unknown
Gene 3	Helix-turn-helix domain-containing protein, *Georgenia soli*	72	65.08	Regulation
Gene 4	Hypothetical protein, *Actinotalea ferrariae*	96	84.27	Unknown
Gene 5	Hypothetical protein, *Actinotalea ferrariae*	100	61.81	Unknown
Gene 6	Hypothetical protein, *Actinotalea ferrariae*	79	81.95	Unknown
Gene 7	Hypothetical protein, *Actinotalea ferrariae*	100	91.45	Unknown
Gene 8	ATP/GTP-binding protein, *Cellulomonas cellasea*	94	86.95	Unknown
Gene 9	TraM recognition domain-containing protein, *Propionibacterium fredenreichii*	99	79.04	tRNA modification
Gene 10	DU4913 domain-containing protein, *Actinomycetia*	90	66.67	Unknown
Gene 11	MFS transporter, *Geodermatophilaceae bacterium*	98	54.31	Transport

Gene 12	ThrRS, *Frankia torreyi*	96	70.48	Self-resistance/acquired resistance
Gene 13	NRPS, *Nocardia arthritidis*	79	35.95	Peptide synthesis
Gene 14	4'-phosphopantetheinyl transferase superfamily protein, *Micromonospora* sp. MA102	97	47.03	Peptide synthesis
Gene 15	Radical SAM protein, *Nocardia donostiensis*	99	50.81	Tailoring Reaction
Gene 16	None	-	-	Unknown
Gene 17	PqqD family protein, *Nocardia donostiensis*	66	31.08	Chaperone for radical SAM
Gene 18	Amino acid adenylation domain-containing protein, *Vibrio* sp. Of7-15	95	25.83	Peptide synthesis
Gene 19	Methyltransferase domain-containing protein, *Anaerolineales bacterium*	96	35.29	Methylation
Gene 20	None	-	-	Unknown
Gene 21	Acyl carrier protein, *Streptomyces* sp. DH-12	17	40	Peptide synthesis
Gene 22	SidA/IucD/PvdA family monooxygenase, *Streptomyces* sp. ERV7	97	37.79	Monooxygenation
Gene 23	FMN-binding negative transcriptional regulator, *Streptomyces* sp. CFMR 7	94	41.10	Regulation
Gene 24	AMP-binding protein, *Streptomyces alkaliphilus*	90	36.30	Peptide synthesis
Gene 25	None	-	-	Unknown
Gene 26	Hypothetical protein, *Microlunatus elymi*	62	33.33	Unknown

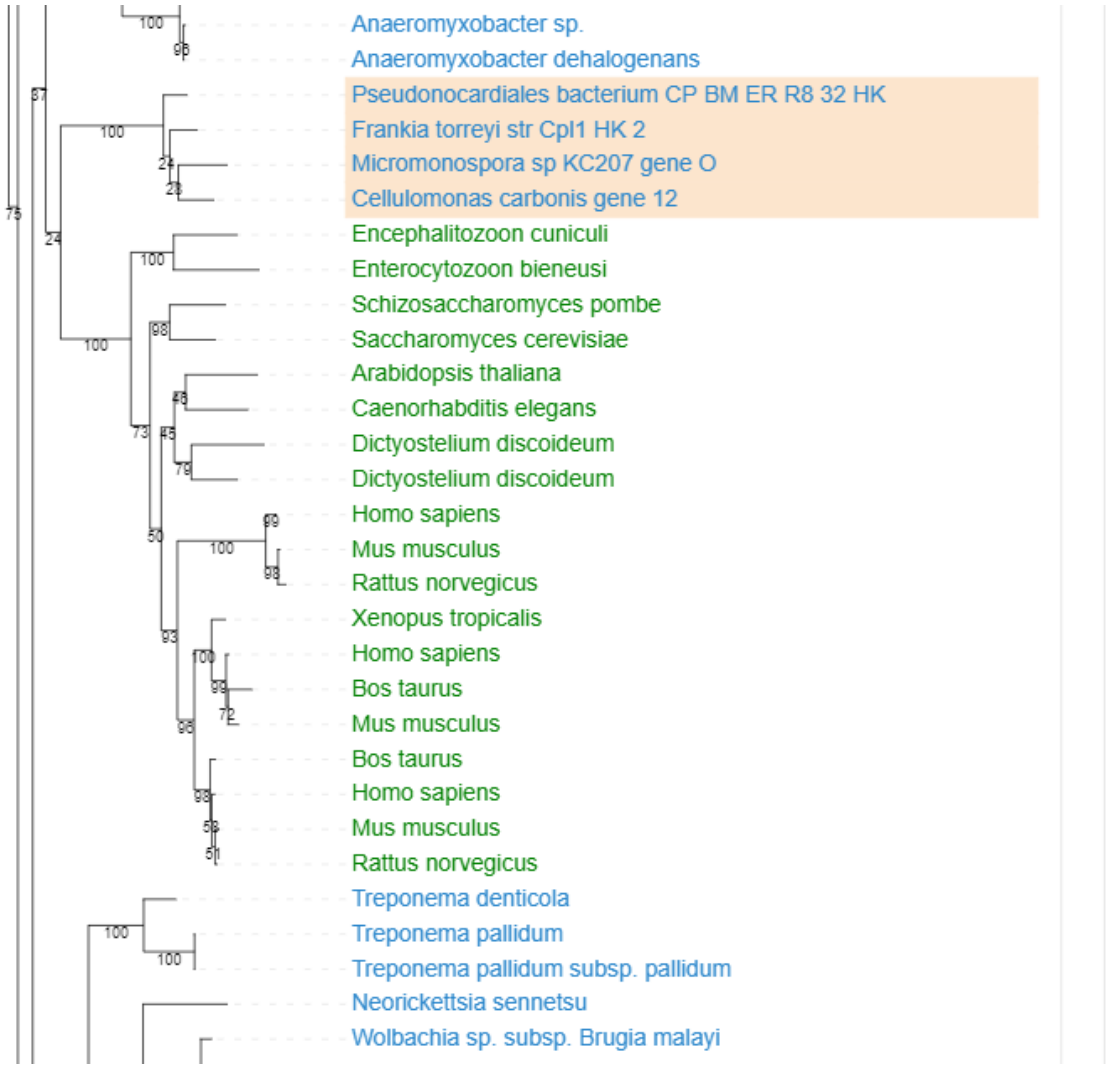
Gene 27	Hypothetical protein, *Cellulomonas hominis*	76	52.23	Unknown
Gene 28	ATP-binding protein, *Actinotalea ferrariae*	100	92.39	Unknown
Gene 29	Hypothetical protein, *Pseudoclavibacter helvolus*	98	48.21	Unknown
Gene 30	Relaxase, *Yonghaparkia*	99	74.26	DNA replication
Gene 31	MobC family plasmid mobilization relaxosome protein, *Intrasporangium chromatireducens*	100	79.55	Plasmid mobilisation
Gene 32	Hypothetical protein, *Actinotalea* sp. C106	89	70.30	Unknown
Gene 33	Hypothetical protein, *Actinotalea solisilvae*	95	29.07	Unknown
Gene 34	Site-specific recombinase, DNA invertase Pin, *Mycobacteroides abscessus*	99	74.21	DNA recombination
Gene 35	Fic family protein, *Microbacterium lacticum*	98	65.50	Protein phosphorylation
Gene 36	Antitoxin VbhA family protein, *Cellulomonadaceae bacterium*	98	76.56	Regulation



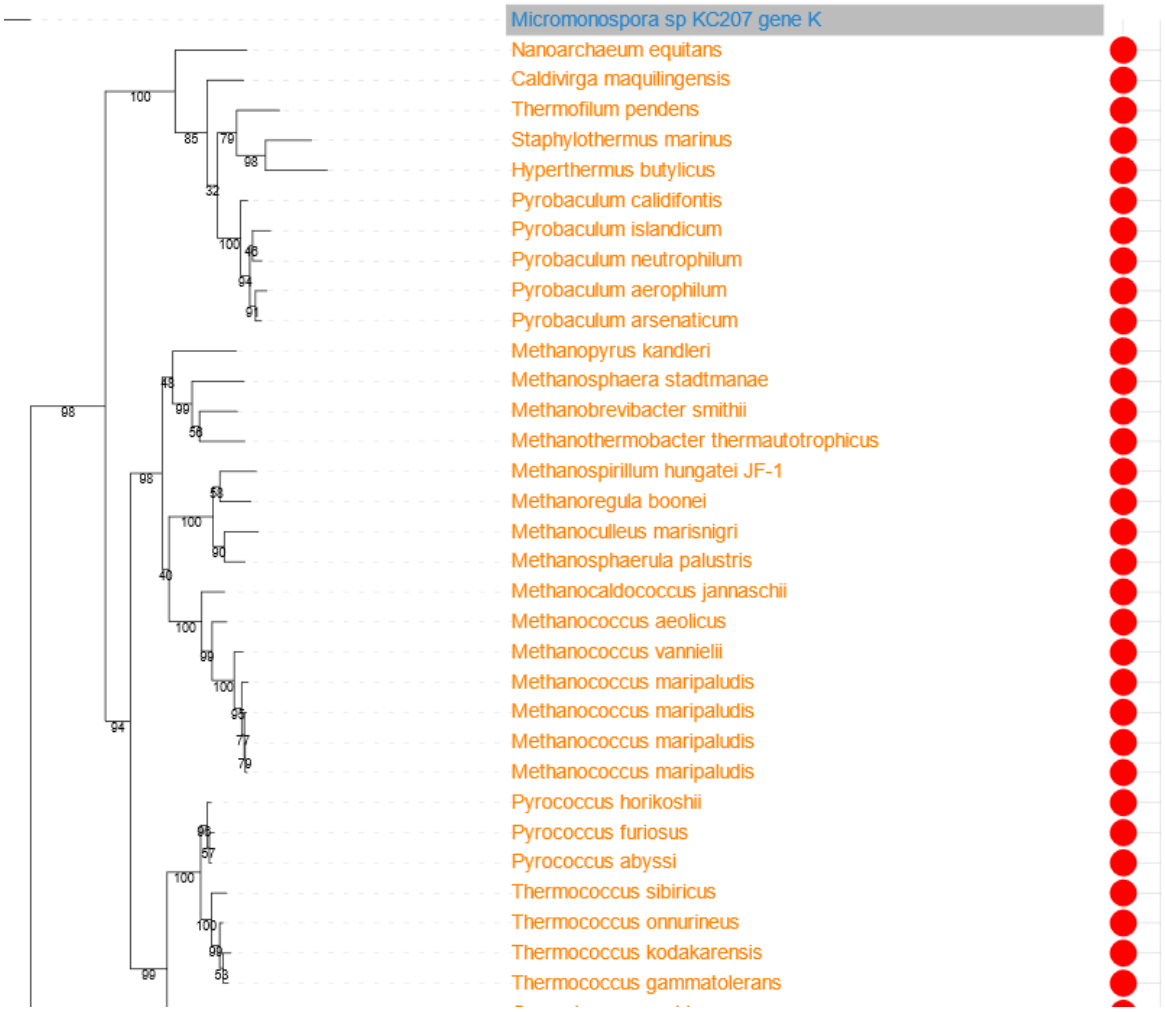
Supplemental Figure 35. Zoom in on BorO clade of maximum likelihood phylogenetic tree of ThrRS sequences from across the tree of life. Branch lengths to scale, with bootstrap values on each branch. Bacterial strain names in blue, archaeal strain names in orange. BorO homologues are highlighted in green, Red exterior circles indicate predicted resistance to borrelidin. Figure generated in the interactive tree of life (iTOL).



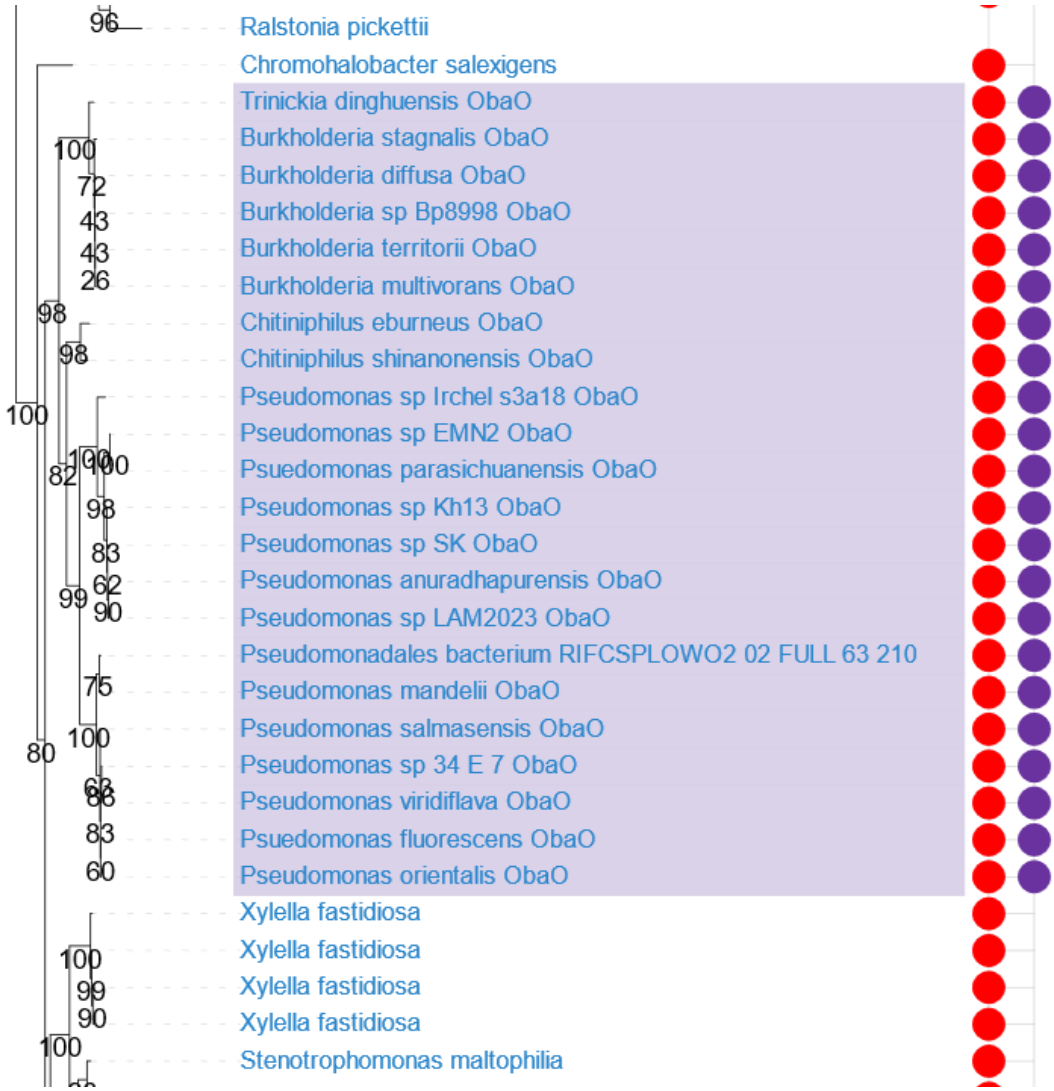
Supplemental Figure 36. Zoom-in on the truncated housekeeping ThrRS clade of the maximum likelihood phylogenetic tree of ThrRS sequences from across the tree of life. Branch lengths to scale, with bootstrap values on each branch. Bacterial strain names in blue, archaeal strain names in orange. *T*runcated housekeeping proteins highlighted in yellow. Red exterior circles indicate predicted resistance to borrelidin. Figure generated in the interactive tree of life (iTOL).



Supplemental Figure 37. *Z*oom-in on the *gene* O clade of the maximum likelihood phylogenetic tree of ThrRS sequences from across the tree of life. Branch lengths to scale, with bootstrap values on each branch. Bacterial strain names in blue, eukaryote strain names in green. Gene O homologues highlighted in orange. Red exterior circles indicate predicted resistance to borrelidin. Figure generated in the interactive tree of life (iTOL).



Supplemental Figure 38. Zoom-in on the *gene* K containing portion of the maximum likelihood phylogenetic tree of ThrRS sequences from across the tree of life. Branch lengths to scale, with bootstrap values on each branch. Bacterial strain names in blue, archaeal strain names in orange, eukaryote strain names in green. *Gene* K is highlighted *i*n grey. Red exterior circles indicate predicted resistance to borrelidin, while exterior purple circle indicate predicted resistance to obafluorin. Figure generated in the interactive tree of life (iTOL).



Supplemental Figure 39. zoom-in on the ObaO clade of the maximum likelihood phylogenetic tree of ThrRS sequences from across the tree of life. Branch lengths to scale, with bootstrap values on each branch. Bacterial strain names in blue, archaeal strain names in orange, eukaryote strain names in green. ObaO homologs highlighted in purple. Red exterior circles indicate predicted resistance to borrelidin, while exterior purple circle indicate predicted resistance to obafluorin. Figure generated in the interactive tree of life (iTOL).