Characterising the Two-Component System-Regulated Biosynthesis of Formicamycin in *Streptomyces formicae*

Katherine Noble

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

Doctor of Philosophy at the University of East Anglia

John Innes Centre

Department of Molecular Microbiology

University of East Anglia

School of Biological Sciences

February 2024

Abstract

A crucial part of overcoming antimicrobial resistance is in the development and overproduction of novel antibiotics. Isolation and genetic analysis of *Streptomyces formicae* identified the formicamycins; novel antibiotics with potent activity against clinically relevant, drug-resistant pathogens and a high barrier to the development of resistance. There are three known cluster-situated regulators that work cohesively to control activation, repression and export of these secondary metabolites. A two-component system, ForGF, was shown to be the main activator of this pathway and this project aimed to further elucidate its role and increase production levels of the compounds. Using a variety of molecular, biochemical and biophysical analyses such as surface plasmon resonance, gene-reporter fusion assays and qRT-PCR, it has been possible to identify the binding site of the response regulator and show the impact of binding on promoter activity and transcription levels. A combination of CRISPR/Cas9 mutagenesis and downstream analysis of *in vivo* and *in vitro* proteins have also been utilised to characterise the interaction of the two components with one another and their surroundings. It has been shown that manipulating ForGF and other regulators within the biosynthetic gene cluster leads to overexpression of the formicamycins, overcoming the problem of low production under standard laboratory conditions.

By exploiting this mechanism of control, it has also been possible to apply similar initial processes to other such cluster-situated two-component systems within the same strain. This genetic manipulation has resulted in changes to the bioactivity of a small library of strains when challenged with a number of pathogenic organisms. This has significant potential for further application for the targeted investigation and overproduction of other novel antimicrobials for clinical development.

This thesis is 261 pages and 67,711 words in length

Publications Arising from this Thesis

Devine R, McDonald H, Qin Z, Arnold C, **Noble K**, Chandra G, Wilkinson B, Hutchings M (2021) Re-wiring the regulation of the formicamycin biosynthetic gene cluster to enable the development of promising antibacterial compounds. *Cell Chemical Biology*. 28(4):515-523

Feeney MA, Newitt JT, …, **Noble K**, …, Duncan KR, Fernández-Martínez LT, Hutchings MI (2022) Actinobase: Tools and Protocols for Researchers Working on *Streptomyces* and other Filamentous Actinobacteria. *Microbial Genomics*. 8(7)

Devine R*, **Noble K***, McDonald H, Stevenson C, Grant C, de Oliviera Martins C, Wilkinson B, Hutchings M (2024) Formicamycin Biosynthesis is Controlled by Feedback and Redox Sensing Mechanisms. *In review*.

This work was supported by the UKRI Biotechnology and Biological Sciences Research Council Norwich Research Park Biosciences Doctoral Training Partnership [Grant number BB/M011216/1]. Candidate number 1001176

Table of Contents

Index of Tables

List of Figures

Acknowledgements

I would like to thank my supervisors, Professor Matt Hutchings and Professor Barrie Wilkinson, for their guidance, support and advice over the last four years. I never could have imagined that joining the Hutchings lab for 8 weeks as part of a Summer Studentship in my undergraduate degree would lead to this project, let alone a PhD or any of the things I've managed to achieve during my time in the lab.

A huge thank you to everyone in the Hutchings and Wilkinson labs both past and present. To Dr Nicolle Som and Dr Tom McLean who taught me to love (and hate) two-component systems, and both helped to guide and shape experiments that were crucial to the project.

A very special thank you to Dr Rebecca Devine, a truly brilliant scientist and excellent friend, absolutely none of this project would be possible without her unwavering support, incredible knowledge, and unimaginable skills. Thank you for taking so much time to teach me just a fraction of what you know, answer an abundance of questions, and help me make sense of the science when the science seemed to be completely upside down!

To my office pals Dr Jonathon Liston and Dr Ainsley Beaton, thank you for the synchronised snack time, endless mochas and letting me scream and shout about anything and everything throughout the day. Your combined friendship and scientific capabilities have really helped with every aspect of this project, and I don't think I'd be anywhere near where I am without the two of you.

To Dr Hannah McDonald for the floor time, weekend catch ups in the lab, and of course blasting Taylor Swift as loud as possible while somehow screaming the lyrics even louder.

I couldn't have asked for a better group of friends to make the difficult science days easier and I'm truly grateful to every single one of you for being such amazing people.

Thank you to everyone at the John Innes Centre, the expertise, encouragement and frequency of cake opportunities in the Chatt Atrium have been exceptional and made the Molecular Microbiology Department a wonderful place to complete this project.

I wouldn’t be anywhere close to having finished my PhD without the support of my friends and family outside of the lab. The final year of this project has given me some of the toughest days imaginable and I don’t think I’d have made it through without these people.

To my Mum and my Grandparents for always supporting me and being so intrigued and invested in my progress with science. All three of them have encouraged me to be the best scientist I can be, even since I was little, and have been proud of any achievement big or small along the way. I wrote the final chapter of this thesis alongside my Grandparents in hospital and finished it in the final weeks of my Grandma’s life. Having her throughout my life was an absolute privilege and she’s made me the person I am today; I hope finishing this thesis and getting my PhD will make her proud.

To Anna and Tom, the best friends I could ever hope to have and my chosen family. Thank you from the bottom of my heart for everything you’ve done and the support you’ve given me. This thesis wouldn’t have been written if it weren’t for all the love and encouragement you’ve given me. Thank you for believing in me, especially when I didn’t believe in myself.

Finally, I’d like to thank my amazing partner Sam. He’s been by my side since day one of my PhD and been there through every single high and low of the last four and a half years. The patience to tolerate “a quick ten minutes in the lab” on a Sunday night that almost always turned into an hour of unreturned calls while I convinced myself “this one extra thing won’t take long” has been outstanding. I can’t imagine how many times I must have rehearsed lab meetings and presentations for him to be able to quote my own slides back at me (yet he somehow still can’t spell formicamycin right on the first try). Thank you for being my best friend and greatest supporter and always encouraging me to keep going no matter what.

For Sophie, my first lab partner and best motivator to get this done for both of us.

In memory of Doreen Mary Noble

17th May 1937 – 18th February 2024

1. Introduction

1.1 Antibiotics

In order to survive, all organisms must overcome what Charles Darwin described as "the struggle for existence". Many have evolved unique ways of doing this, with microbes demonstrating a remarkable ability to interact and thrive in complex communities often vying for the same resources (Merritt & Kuehn, 2016). Some bacteria for example, produce antimicrobial secondary metabolites capable of eliminating such competition. These substances act to kill (bactericidal) or simply inhibit the growth (bacteriostatic) of other bacteria and in a secondary display of resource competition, humans have long exploited this antibiosis as means of tackling infectious diseases.

The term antibiotic was coined by Selman Waksman in 1943 to describe "a compound made by a microbe to destroy other microbes", and today that definition remains largely unchanged (Waksman & Woodruff, 1940). These compounds are categorised by the physiological components or biochemical pathways they disrupt in the target organism and/ or the mechanism through which they achieve this. Such effects can be brought about by; disruption of membranes (e.g. polymyxins), targeting nucleic acid production and/ or repair mechanisms (e.g. rifampicin and fluoroquinolones), interference with metabolic pathways (e.g. sulphonamides), disruption of cell wall biosynthesis (e.g. β-lactams) or inhibition of protein synthesis generally through binding to ribosomal subunits (e.g. aminoglycosides, chloramphenicol and tetracycline)(Sengupta et al., 2013; Sultan et al., 2018).

Antibiotics are now widely considered one of the greatest medical breakthroughs of the 20th century due to both their standalone capabilities and the medical advances that have been made possible through their use (Hutchings et al., 2019). One of the most significant benefits is their ability to differentiate between prokaryotic and eukaryotic targets, therefore focussing the impact on the pathogen with minimal effects on the host. Unlike some pharmaceuticals, antibiotics are curative measures, used to completely eradicate infection as opposed to simply being palliative. When the exact cause of an infection is unknown, an empiric approach allows a broad-spectrum antibiotic to be administered, progressing to definitive therapy using an alternative narrow spectrum antibiotic once the causative pathogen has been identified (Leekha et al., 2011). This extensive range of applications has meant that previously fatal infections have not only become treatable but preventable, extending the average life span by 23 years (Shrethsa L, 2005).

The informal use of antibiotics has been recorded throughout history, with Eber's papyrus dating back to 1500 BC detailing the use of mouldy bread, honey and soil as part of wound care and in the management of infections (Haas, 1999). However, in 1928 a Petri dish with a culture of *Staphylococcus aureus* bacteria was noted by Alexander Fleming to be contaminated with *Penicillium notatum*, which was inhibiting the bacterial growth (Fleming, 1929). With the assistance of Howard Florey, Ernst Chain and Norman Heatley, this observation led to the discovery, purification, and later commercial production of penicillin. This marked the first formal and regular use of an antibiotic in a clinical setting and the pertinent development promoted others to investigate the antibacterial properties of other such microbes. Later, in 1945, Dorothy Hodgkin would solve the now infamous β-lactam structure of penicillin (Hodgkin, 1949). The discovery of penicillin and a growing number of reports of the production of antimicrobial compounds by microorganisms led to the development of a systematic screening platform by Selman Waksman in the late 1930s (Lyddiard et al., 2016; Waksman & Woodruff, 1940). This quickly identified the soil as a brilliant source of these new strains and products, and worked to methodically screen samples against pathogens, marking the start of the "Golden Age of Antibiotic Discovery" that spanned the 1940s and 50s (Dubos, 1939).

The reason behind the wealth of potential was *Streptomyces*, a genus of soil-dwelling, chemoheterotrophic actinomycetes with extensive and complex metabolic capacity. Over 12,000 antibiotics have been characterised, falling into 38 classes, and more than two thirds of those currently prescribed are derivatives of *Streptomyces* natural products (Demain, 2009; Watve et al., 2001). However, the repeated discovery of known strains and their compounds meant that this pipeline dried up and discovery peaked in the mid-1950s. Pharmaceutical companies instead turned their efforts towards the development of synthetic, target-based molecules, but this proved largely unsuccessful. Several factors subsequently contributed to industry virtually abandoning research efforts and reducing funding. Initially this was driven by the belief that there was no longer a need for new drugs as experts claimed infectious diseases were defeated, making continued research less of a priority (Projan & Shlaes, 2004). Later came the realisation of poor investment return; antibiotics are generally used in acute conditions and are typically only required for a matter of days or weeks. However, pharmaceuticals used to treat chronic conditions (e.g. statins) or relieve symptoms (e.g. analgesics) typically demand long-term prescription or have a higher, more widespread requirement and therefore are more profitable (Overbye &

Barrett, 2005; Projan, 2003). The ensuing increased costs of development, pricing controls and pre-placed government restrictions on any new products have massively decreased the commercial potential of any new antibiotics (Norrby et al., 2005). The number of newly approved antibiotics has fallen significantly (68% reduction from 1980 to 2014) and currently only three large pharmaceutical companies are actively investigating new antibiotics (Bartlett et al., 2013; Ventola, 2015). Instead, the effort is being led by small and medium sized enterprises that are struggling to front the significant costs associated with this research and development pipeline (WHO, 2019). This lack of new discovery has in part contributed to the development of the antimicrobial resistance (AMR) crisis that is being faced on a global scale.

1.2 Antimicrobial Resistance

Even before the end of the Golden Age of Discovery, resistance was emerging and now, less than 100 years after the surge in discovery, death due to bacterial infection is once again posing a serious threat to humanity. A report commissioned by the UK government predicted that by the year 2050, up to 10 million people a year could die worldwide as a result of infectious diseases if nothing is done to combat AMR (O'Neill, 2016). The revolutionary nature of antibiotics meant that they were used relentlessly from the point of discovery. The apparent wealth of potential from the soil was widely hailed as a never-ending source of new treatments, leading many to dismiss even Alexander Fleming in his Nobel Prize Lecture in 1945 where he warned of resistance to penicillin (Fleming A, 1945). Before the structure of this ground-breaking antibiotic had even been solved, resistance was emerging, with four clinical isolates of *S. aureus* having been found to be res(Lobanovska & Pilla, 2017)ment (Lobanovska & Pilla, 2017). Despite Fleming's warnings, antibiotics were used in a vast number of settings, from clinical practice to agriculture, ultimately leading to an increased prevalence in the environment (Larsson, 2014). This sudden, large-scale exposure to selection pressure triggered an evolutionary response in pathogens, with many becoming resistant (Sykes, 2010). As a result, almost all antibiotics have progressively seen the development of resistance in organisms they were once capable of eliminating (**Figure 1.1**).

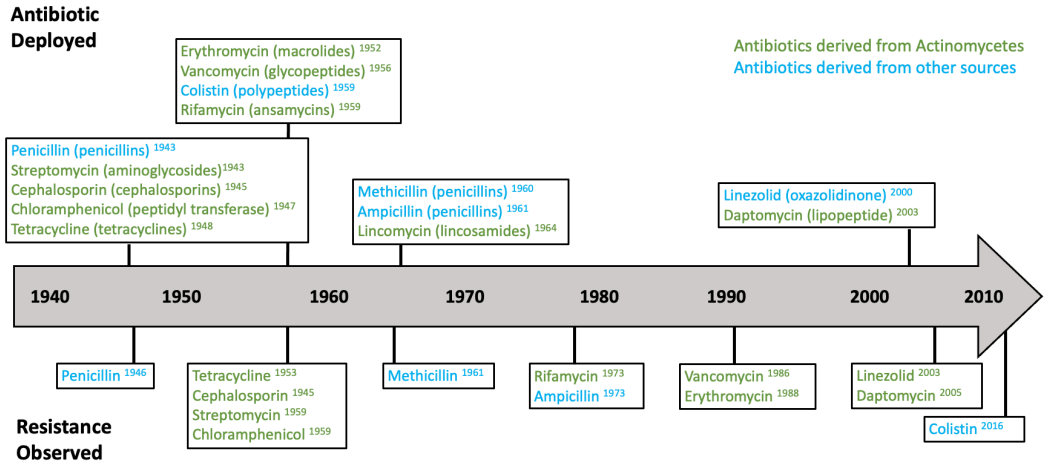


Figure 1.1 Timeline of antibiotic deployment (above) compared to initial observation of resistance (below). The first antibiotics to be produced from several key classes are shown in addition to their source. Note: some classes have multiple antibiotics detailed due to clinical significance and observed resistance does not denote a subsequent lack of clinical use. Figure adapted from (Hutchings, Truman and Wilkinson, 2019)

1.2.1 Mechanisms of Resistance

There are four main mechanisms through which bacteria resist antibiotics: limiting the uptake, active efflux, inactivation and modification of the target (**Figure 1.2**). The extent to which these are used depends on the type of bacteria, with variation seen between Gram-positive and Gram-negative bacteria due to their structural differences, however all four types have been observed across the domain (W. C. Reygaert, 2018).

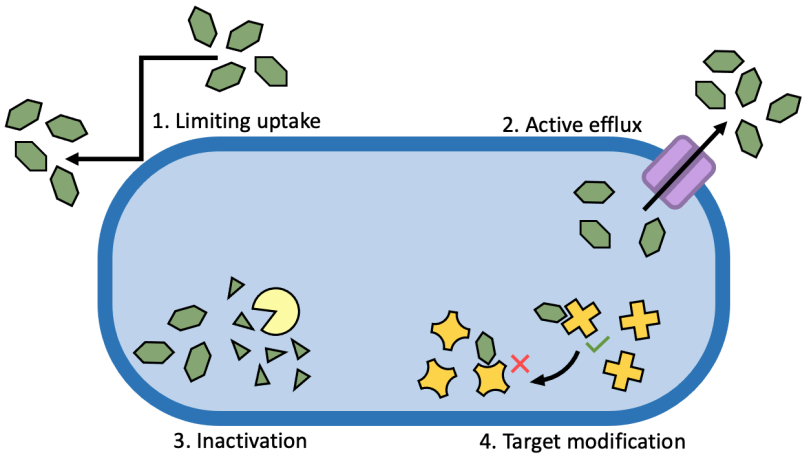


Figure 1.2 Diagrammatic representation of the four main mechanisms of antibiotic resistance employed by bacteria including limiting drug uptake, active efflux from the cell, inactivation of the drug and modification of the drug target. Adapted from (Reygaert, 2018)

By limiting the uptake of antibiotics into the cell, bacteria can physically block compounds from accessing their intracellular or periplasmic targets. In Gram-negative bacteria, the thick lipopolysaccharide layer acts as an effective barrier against antimicrobial agents (Blair et al., 2014). For example, the Gram-negative *Pseudomonas aeruginosa* has an innate low susceptibility to β-lactams as a result of its thick LPS outer membrane (OM) layer and a reduced number of porins being expressed in the OM (Hancock & Brinkman, 2002). The incredibly powerful vancomycin, which exerts its effects by inhibiting cell wall synthesis, is ineffective against Gram-negative bacteria due to its inability to cross the OM (Selim, 2022). Whilst using a physical barrier to reduce uptake is a mechanism largely employed by Gram-negative bacteria, there are well documented examples of Gram-positive bacteria also utilising this mechanism as a form of resistance. For example, *Mycobacterium tuberculosis* has an unusually thick peptidoglycan layer and a mycobacterial OM which means it is largely impermeable to hydrophilic antibiotics (Gygli et al., 2017). Similar thickening of the peptidoglycan layer has been seen in vancomycin intermediate *S. aureus* (VISA) strains but this only provides low level resistance (Miller et al., 2014).

Active efflux of an antibiotic out of the cell was first documented in 1980 and since then five distinct families have been documented in bacteria. These can range from substrate specific pumps, which generally confer resistance to a single antibiotic, to broad range pumps which transport a number of substrates out of the cell and are associated with multi-drug resistance. The five families are: the major facilitator superfamily (MFS), the small drug resistance family (SMR), the resistance-nodulation-cell-division family (RND), the ATP-binding family (ABC), and the multi-drug and toxic compound extrusion family (MATE) (Kumar & Schweizer, 2005; Poole, 2007). Each family uses a different energy source, transports a different range of substrates and holds different conformations, meaning that collectively efflux pumps are one of the most widespread mechanisms through which bacteria are known to be resistant to antibiotics. In Gram-positive bacteria the efflux pumps are typically encoded within the chromosome, conferring intrinsic resistance through mainly MFS and MATE families, however other pumps have been identified and some are known to be encoded on plasmids. In Gram-negative bacteria, all five families are well documented as mechanisms of resistance (Blair et al., 2014; Poole, 2007).

If bacteria are unable to block the antibiotic from entering the cell or remove it if it does manage to cross the external membrane, inactivation of the antibiotic is an alternative mechanism of preventing it from acting upon its cellular target. Whilst it was not

understood at the time, one of the now most studied and document examples of this is one of the mechanisms that caused Alexander Fleming to warn against resistance to penicillin. The β-lactamase enzymes act to hydrolyse the β-lactam ring, breaking it open and rendering the antibiotic unable to bind to their target penicillin binding proteins (PBPs). Various families of these enzymes, which are classified based on their molecular structure, are found between both Gram-positive and Gram-negative bacteria, conferring resistance to any antibiotics containing this β-lactam ring structure (Zapun et al., 2008). Enzymatic modification of the antibiotic can also involve the introduction of chemical groups to the compound such as acetylation, adenylation or phosphorylation (Blair et al., 2015). Often these changes mean that steric hindrance decreases the avidity of the antibiotic for its target, preventing it from exerting its effects on the cell.

A final common mechanism of resistance is for the bacteria to modify the target of the antibiotic, this can include increasing the production of the target to counteract inhibitory effects (e.g. VISA strains) or altering the target site to prevent antibiotic binding. Both of these examples can be seen in different mechanisms of penicillin and other β-lactam resistance. The β-lactam antibiotic normally exerts its effects by binding to PBPs with the β-lactam ring inhibiting the transpeptidases that facilitate the final crosslinking step in peptidoglycan formation, ultimately disrupting cell wall biosynthesis. However, some strains of the Gram-positive *S. aureus* have been shown to produce a modified copy of PBP (PBP-2a), which the β-lactam ring is unable to bind to, meaning it can no longer disrupt cell wall formation. This low-affinity copy of PBP, encoded by the *mecA* gene, is over-produced within the strain to minimise the impact of β-lactams on native copies of the target (Fishovitz et al., 2014).

1.2.2 The Spread of Resistance

When discussing the development of AMR, it is important to consider the ability of pathogenic bacteria to acquire and disseminate resistance, the human factors involved in further spread and of course the ways of minimising the impact of resistance.

There are three distinct ways in which bacteria develop resistance to antibiotics; having intrinsic resistance due to certain structural or functional components, acquiring genetic material that confers resistance or developing mutations that inhibit antibiotic activity (Cox

& Wright, 2013). For some bacteria, natural resistance is always actively expressed while for others it is only seen once cells have been exposed to the antibiotic. In Gram-negative bacteria and in mycobacteria, the presence of an outer membrane means many antibiotics are unable to permeate into the cell in the first place, while Gram-positive bacteria often overexpress genes encoding efflux pumps (e.g. NorA for quinolone resistance and TetA for tetracycline resistance) and antibiotic degrading enzymes (e.g. β-lactamases for penicillin resistance) in response to exposure (Ng et al., 1994; Wang et al., 2014). Most antibiotic-producing bacteria have co-evolved self-resistance mechanisms to minimise the toxicity of the secondary metabolites on the producing cell, supporting both competition and defence. This intrinsic resistance provides a suitable threat to our ability to out-compete bacteria and eradicate infections, however, the efficiency with which bacteria can share DNA, from one another coupled with short generation times means that this resistance can be disseminated incredibly rapidly, making the threat even more significant (Hawkey, 1998). The positioning of resistance genes on plasmids means they can be passed to other bacteria through both vertical and horizontal gene transmission (Munita & Arias, 2016). The development of an "environmental resistome" means that previously susceptible bacteria can obtain such genes, survive antibiotic attacks and in turn spread the resistance even further (Blair et al., 2015). Not only can resistance be obtained by acquiring genetic information from other bacteria, but mutation can occur within the bacterium's own DNA through erroneous replication or exposure to stressors such as UV radiation, starvation or antibiotics and other chemicals (W. C. Reygaert, 2018). Often mutations within the bacterial DNA are deleterious and will only be beneficial in the context of developing resistance if they occur in genes that are involved in the mechanisms previously outlined (Davies & Davies, 2010). Whilst these mutations provide a selection advantage against antibiotics, they generally do so at a cost to the organism e.g. methicillin resistant *S. aureus* (MRSA) has a significantly slower growth rate than its antibiotic susceptible counterpart (W. Reygaert, 2009).

The arguably impressive ability of bacteria to obtain and disperse resistance genes has contributed significantly to the development of AMR. However, the progression of this now global crisis has also been influenced drastically by human behaviour. Global advances in transport mean that countries are more connected than ever, with opposite sides of the world (England to Australia, ~9500 miles) being accessible in as little as 17 hours (Marks, 2009). Concerns about the potential of aviation playing a role in the spread of infectious

diseases were confirmed during the SARS-CoV-2 outbreak in which only one country of the 193 recognised by the United Nations was unaffected (as of January 2023), and transmission between the other 192 being largely attributed to international travel (List of Countries Without Coronavirus; Kucharski et al., 2020). Whilst this is an extreme example of a global pandemic, and viral transmission is typically more rapid than bacterial, the same principles can be applied to the epidemiological spread. Resistant strains and resistance genes can travel via human hosts to other countries and within those countries, introducing new issues and increasing the need for alternative antibiotics to be used. In the UK, different prescribing guidelines are in place in each NHS trust to factor in the current prevalent strains in the local area and ensure that appropriate antibiotics are prescribed based on the most up to date information surrounding resistance and susceptibility.

Additionally, a distinct lack of education surrounding the purpose, use and importantly limitations of antibiotics has resulted in a divide between the scientific/ medical community's understanding and that of the general public (Castro-Sánchez et al., 2016). Misconceptions about the appropriate context for antibiotic use and a lack of point of care testing has resulted in increased pressure on clinicians to prescribe them even when not necessarily indicated (Rather et al., 2017). In some countries, the availability of antibiotics remains largely unregulated to enable those without regular healthcare access to utilise their benefits. However, this also exacerbates the problem as a lack of clinical guidance mediated via over-the-counter availability and unregulated supply chains leads to misuse and in turn the development and spread of resistant strains through communities (Ayukekbong et al., 2017; Byrne et al., 2019). The (mis)use of antibiotics by humans is not just restricted to the clinical setting though; the agricultural industry is accountable for the majority of their consumption though prophylactic use and by association the development of resistance. This is an issue that went largely unnoticed, unmonitored and unregulated for a significant amount of time and is still a cause for concern today. The routine use of antibiotics to prevent and treat infection, improve feed conversion and as a growth promoter in farm animals has resulted in wide-spread generation of AMR which has passed into the environment and food chain, directly impacting on human health (Koike et al., 2017; Manyi-Loh et al., 2018).

Whilst the combination of these factors has led to a global health crisis considered to be one of the biggest threats to the modern world, the solutions to the problem also rely on having a good understanding of how each contributed. Incentives for pharmaceutical

companies to identify new antibiotics needs to be coupled with improved regulations and education surrounding their use and availability. Development of rapid point of care testing will allow for easier tracking of resistant infections while a greater understanding of inherent and acquired resistance mechanisms already present in the environment prior to their spread may also provide strong indication on how to prevent the same problems from occurring again so quickly. In order for any of these strategies to be utilised though, new compounds and products need to be found, to allow the solutions to be put in to practice and tackle the already significant number of resistant infections.

1.3 Natural Products

The term natural product is used to describe a compound produced by a living organism as part of its secondary metabolism and as such are interchangeably referred to as specialised or secondary metabolites. These products are not required for the growth, development, or reproduction of the producing organism under laboratory conditions but do typically provide it with some form of selection, growth or competition advantage. A significant number of these have been utilised by humans for a variety of purposes including food additives, dyes, pesticides, and spices. However, their anticancer, antifungal, antibiotic and immunosuppressant properties have proved hugely valuable in the medical fields. Depending on the therapeutic class, an estimated 30-66% of active drug molecules are natural products and up to 75% of antibiotics currently used today are either secondary metabolites or their derivatives (Pye et al., 2017; Ventola, 2015). They are thought to be better candidates as antimicrobials than synthetic alternatives due to their ease of transfer across cell membranes, chemical stability and higher, naturally optimised affinity for protein targets as a result of their cellular origin (Welsch et al., 2010).

Secondary metabolites are built using the products or intermediates of primary metabolism. While the number of “building blocks” available to create secondary metabolites may seem limited, the diversity and size of chemical structures seen across natural products is vast. This is in part a result of the enzymes that facilitate their synthesis, which generally evolved through gene duplication and divergence of those that serve functions in primary metabolism (Dias et al., 2012). These often form large, multi-enzyme complexes and the genes that encode them are clustered together in the genome. These biosynthetic gene clusters (BGCs) typically encode the core biosynthetic machinery,

accessory enzymes and regulatory, transport and self-resistance mechanisms involved in the production of secondary metabolites. BGCs can range from a few thousand base pairs to 100 kilobases in size and in some bacteria up to 15% of their genome is made up of these BGCs (Bibb, 2005; Davies, 2013). The broad range of enzymes that make up BGCs, means that natural products are an extremely diverse group of compounds. They are characterised based on their structural scaffold and/ or enzyme complexes involved in their production. Classes include polyketides, non-ribosomal peptides (NRPS), ribosomally synthesised and post-translationally modified peptides (RiPPs), terpenes and alkaloids (Atanasov et al., 2021).

The various classes of natural products are generally very well conserved among the organisms that produce them, in part due to the significant selective advantage they provide but also their compartmentalised encoding within the genome. The majority of BGCs contained within an organism's genome are not expressed under laboratory conditions meaning that we are not able to exploit the full potential of natural products available. However, genome sequencing techniques coupled with knowledge of BGCs means some of the potential compounds produced by an organism can be predicted.

1.3.1 Polyketides

Many clinically relevant natural products such as antibiotics (e.g. erythromycin), anti-parasitics (e.g. ivermectin) and chemotherapeutics (e.g. epothilone) are produced by polyketide synthases (PKSs) (Nivina et al., 2019). The PKS BGCs give rise to large multi-enzyme complexes or multi-domain enzymes that work systematically to generate highly decorated and extremely diverse molecules. The polyketide compounds are formed by sequential decarboxylative condensation of multiple extender units, typically using malonyl CoA or methylmalonyl CoA, to create an elongated carbon chain skeleton (Fischbach & Walsh, 2006). This chain then continues to be extensively modified in both oxidation and stoichiometric states by the PKS enzymes until the final product is made (Ray & Moore, 2016). These modifications, alongside final chain length, starter units and extender units all contribute towards the diversity of this group of natural products. PKS enzymes are separated into three groups, type I (T1PKS), type II (T2PKS) and type III (T3PKS), although this has since been considered an over-simplified classification system (Shen, 2003).

Type I Polyketide Synthases

Type I PKS are multifunctional enzymes that are organised into modules, with each having a distinct catalytic domain that is responsible for the addition of one unit before passing the extended chain on to the next module. Each module contains a ketosynthase (KS) domain, an acyltransferase (AT) domain and an acyl-carrier protein (ACP) domain, as well as any other domains required for the modification of the β-keto intermediate (Robbins et al., 2016). The AT loads the ACP with a single extender unit, on to which the KS can facilitate formation of a carbon-carbon bond between this extender unit and the growing product. The choice of extender unit used by the module is determined by the specificity of the AT domain (Oliynyk et al., 1996; Ruan et al., 1997). The other domains present may include: an enoylreductase (ER) that forms a single carbon-carbon bond, a dehydratase (DH) that forms a double carbon-carbon bond or a ketoreductase (KR) that forms a hydroxyl group (Dutta et al., 2014). Each module and the order that they exist in effectively acts as a "code" for the structure of each 2-carbon unit extension that will be added on to the chain at each stage of the process. Once the chain has passed through each module and been extended to its predetermined length, the ACP in the final module transfers the polyketide to the thioesterase (TE) unit where it is released from the multi-enzyme complex via a hydrolysis reaction (Hwang et al., 2020). Post-PKS modifications are carried out by accessory enzymes that create the final polyketide product.

One example of this modular T1PKS system is the biosynthesis of erythromycin A, a macrolide produced by *Saccharopolyspora erythraea*. It is formed by the combined action of a loading module, six extender modules and an end module, known as deoxyerythronolide B synthase (DEBS), and a series of post-PKS tailoring enzymes (Cane, 2010). A propionyl-CoA starter unit is loaded onto the ACP before each of the extender modules adds a methylmalonyl-CoA extender unit via the co-ordinated action of their AT and KS domains. Except for module three, each extender also includes a KR domain that modifies the β-keto group added by that module to a hydroxyl group. Module four also includes ER and DH domains which facilitate the addition of a single and double carbon-carbon bond, respectively. After the chain has passed through each module the poly-β-keto intermediate, is released from the ACP by the TE domain. It is then cyclised to form 6-deoxyerythronolide B before the post-PKS tailoring enzymes hydroxylase, O-methytransferase and glycosyl transferase enable the formation of erythromycin A (**Figure 1.3**) (Cummings et al., 2014; Weissman & Leadlay, 2005).

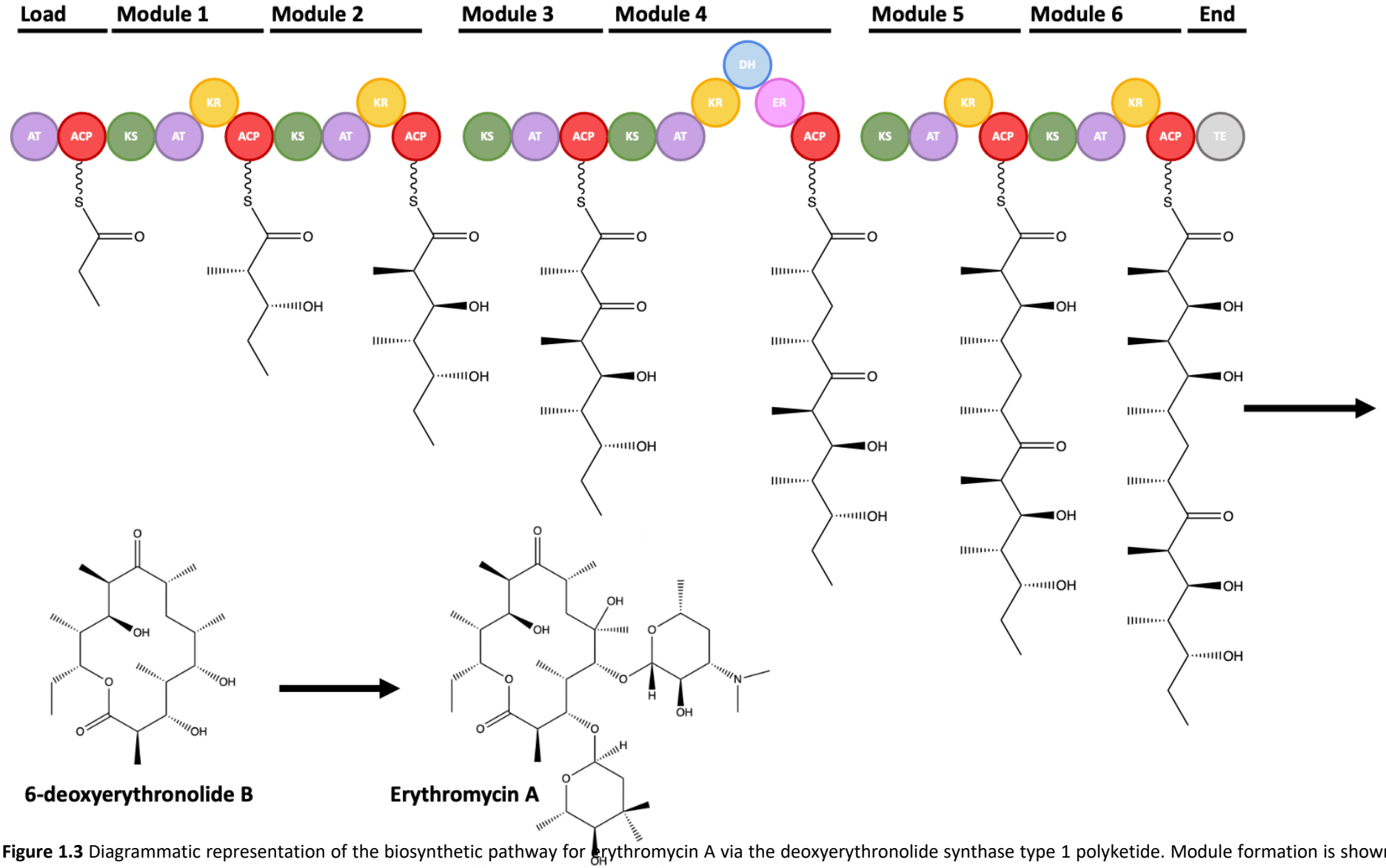


Figure 1.3 Diagrammatic representation of the biosynthetic pathway for erythromycin A via the deoxyerythronolide synthase type 1 polyketide. Module formation is shown through the lines above and individual domains are colour coded with abbreviations: **AT** acyltransferase, **ACP** acyl carrier protein, **KS** ketosynthase, **KR** ketoreductases, **DH** dehydratase, **ER** enoylreductase, **TE** thioesterase. Adapted from Cane *et al*., 2010.

Type II Polyketide Synthases

Type II PKS typically produce polycyclic aromatic compounds via a minimal set of individual enzymes that act iteratively to produce the final polyketide via discrete, monofunctional domains. A stand-alone ACP exists alongside two KS subunits (KS_α and KS_β) which form a heterodimer, the three of which are collectively referred to as a minimal PKS. The KS_α subunit facilitates the condensation carboxylation reaction between extender units while the KS_β subunit determines the final chain length and is also known as the chain length factor. The ACP acts as an anchor for the growing polyketide chain and shuttles it between each enzymatic domain alongside the malonyl-CoA extender units (McBride et al., 2023). Once the chain has been through each of the domains and has reached its predetermined length (usually between 10 and 30 carbons), the poly-β-keto intermediate is converted to the aromatic polyketide core by ketoreductases (KR), cyclases (CYC), aromatases (AR) and oxygenases (OX) (Tang et al., 2017). At this point a variety of post-PKS tailoring enzymes such as halogenases, oxygenases, methyltransferases and glycosyltransferases complete the final modifications, generating the range of structural diversities seen within this class (Hertweck et al., 2007). Predicting the final structure of T2PKS products is notoriously difficult due to the iterative addition of groups by multiple discreet enzymes as opposed to the action of one multi-enzyme complex. This diversity is best displayed in *Streptomyces* bacteria that employ a number of T2PKSs to produce an array of polyphonic compounds with a broad range of bioactivity. For example, actinorhodin, tetracyclines, doxorubicin,

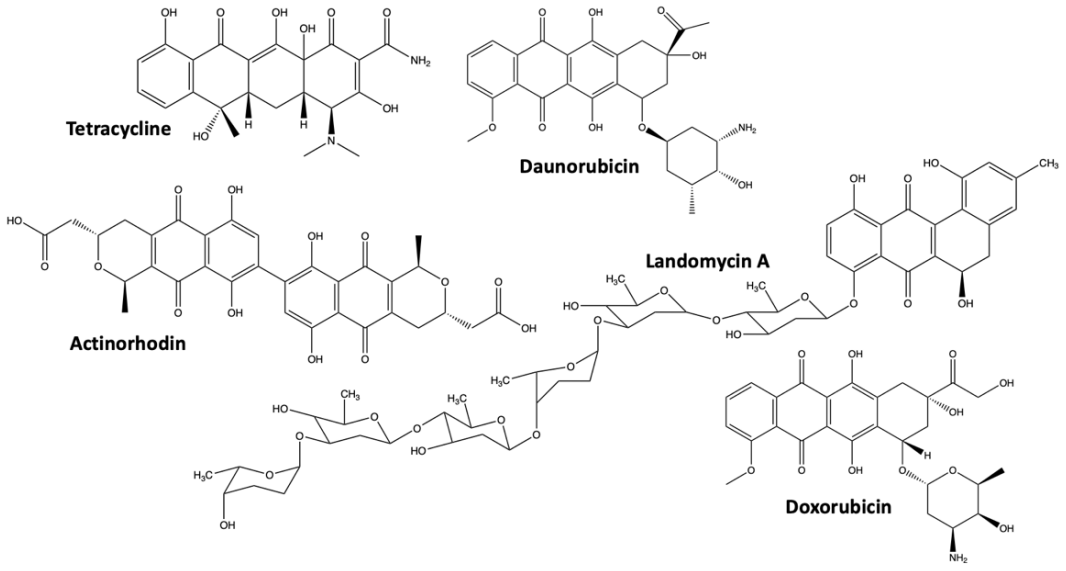


Figure 1.4 Structure of several natural products produced by type II polyketide synthases

merdermycin, landomycin A and daunorubicin all originate as natural products produced by various *Streptomyces* strains, some of which are currently used in clinical settings for their antimicrobial activity (**Fig 1.4**) (Risdian et al., 2019).

Actinorhodin is the best model compound for understanding T2PKSs including the ancillary enzymes involved in its biosynthesis and the post-PKS modifications that take place. The connection between the final product and the BGC responsible for its production was first made in 1984 (Malpartida & Hopwood, 1984). The actinorhodin minimal PKS is known to catalyse at least 18 separate reactions that are divided into loading, initiation, extension, cyclisation and release phases (Beltran-Alvarez et al., 2007). In the loading stage, malonate is used as the first building block and attached to the ACP where it is converted into acetyl-ACP to create the starter unit. This is then extended to a 16-carbon chain through seven extension reactions through iterative reactions by the KS subunits. The chain is then cyclised and released from the minimal PKS before a series of tailoring enzymes convert it to dihydrokalafungin (DHK). Finally, two DHK molecules are ligated to generate the six-ring actinorhodin final product (Okamoto et al., 2009; Y. Xu et al., 2012).

Type III Polyketide Synthases

Unlike types I and II, type III PKSs do not utilise ACP to anchor the growing polyketide chain, instead acyl-CoA is used directly as a substrate. Homodimeric enzymes act to iteratively carry out the priming, extension and cyclisation of the polyketide in a simplified version of the other previously described PKSs. Very few T3PKSs have been identified in bacteria with the significant majority being utilised by plants to produce natural products (Risdian et al., 2019). Of those identified, many are involved in the biosynthesis of precursor molecules that are required in later stages of a glycopeptide biosynthetic pathways, such as corbomycin. In addition, some T3PKSs have been exploited to produce novel compounds through precursor-directed biosynthesis (D. Yu et al., 2012).

1.3.2 Non-Ribosomal Peptides

Non-ribosomal peptide synthases (NRPS) are large, multi-functional enzymes that facilitate the production of peptide natural products without ribosomally synthesising the peptides. The mega-proteins form a highly organised modular complex where each active site is

responsible for catalysing a discreet reaction enabling amino acid monomers to be added to a growing peptide chain. Unlike ribosomes, NRPSs can use a wider range of monomers than the 20 proteinogenic amino acids and instead can utilise other amino acids such as ornithine and methylglutamate or other monomers such as fatty acids. The BGCs that encode the NRPS sometimes also include genes that produce amino acids specifically for the purpose of incorporation into the final natural product, allowing for a significant amount of diversity among this group (Walsh et al., 2013).

NRPS are minimal modules which consist of an adenylation (A) domain, a thiolation or peptidyl carrier protein (PCP), a condensation (C) domain and a TE domain. The A domain selects and activates the amino acid, the PCP tethers this activated amino acid alongside the growing polypeptide intermediate and the C domain catalyses the formation of a peptide bond between the two. Additional domains may also be present to facilitate chemical alterations such as epimerisation (E), cyclisation (CYC), methyltransferase (MT) and oxidation (O). Finally, the TE domain then releases the polypeptide chain from the NRPS via a hydrolysis reaction (Süssmuth & Mainz, 2017). The A domain is responsible for the diversity that is seen in this group of natural products as they can recognise a wide range of substrates. The substrate can be predicted based on the specificity-conferring residues within the binding pocket. In some cases, they can recognise several, structurally similar amino acids resulting in analogous final products (Wenski et al., 2022).

The modular systems that form NRPSs are like PKSs in that they are divided into three groups dependant on how the natural product progresses through the system. Type I is linear, where the number of modules is directly linked to the number of monomers that are put into the final product and the order in which they are incorporated into the nascent oligopeptide chain. This correlation between NRPS structure and the final product is known as the colinearity rule. In type II NRPSs the modules or domains are used iteratively, and the final biosynthetic product is made up of multiple repeated peptide sequences. Finally, type III NRPSs are non-linear where modules are rearranged and the final peptide product sequence is independent of the NRPS structure (Mootz, Schwarzer and Marahiel, 2002).

A prime example of a type I NRPS is the system responsible for the production of vancomycin, a tricyclic glycopeptide antibiotic made exclusively of non-proteinogenic components (Nolan & Walsh, 2009). This glycopeptide is produced by seven NRPS modules, each responsible for the addition of one amino acid subunit and are distributed into 3 enzyme complexes. The heptapeptide intermediate is generated as it progresses linearly

through the NRPS before being cyclised by cytochrome P450 oxygenases. These are recruited to the NRPS via a non-catalytic C domain-like X domain that is embedded in the mega-enzyme complex (Wenski et al., 2022) (**Fig 1.5**). The final vancomycin product is widely used in healthcare as a final line of treatment against Gram-positive bacterial infections such as MRSA.

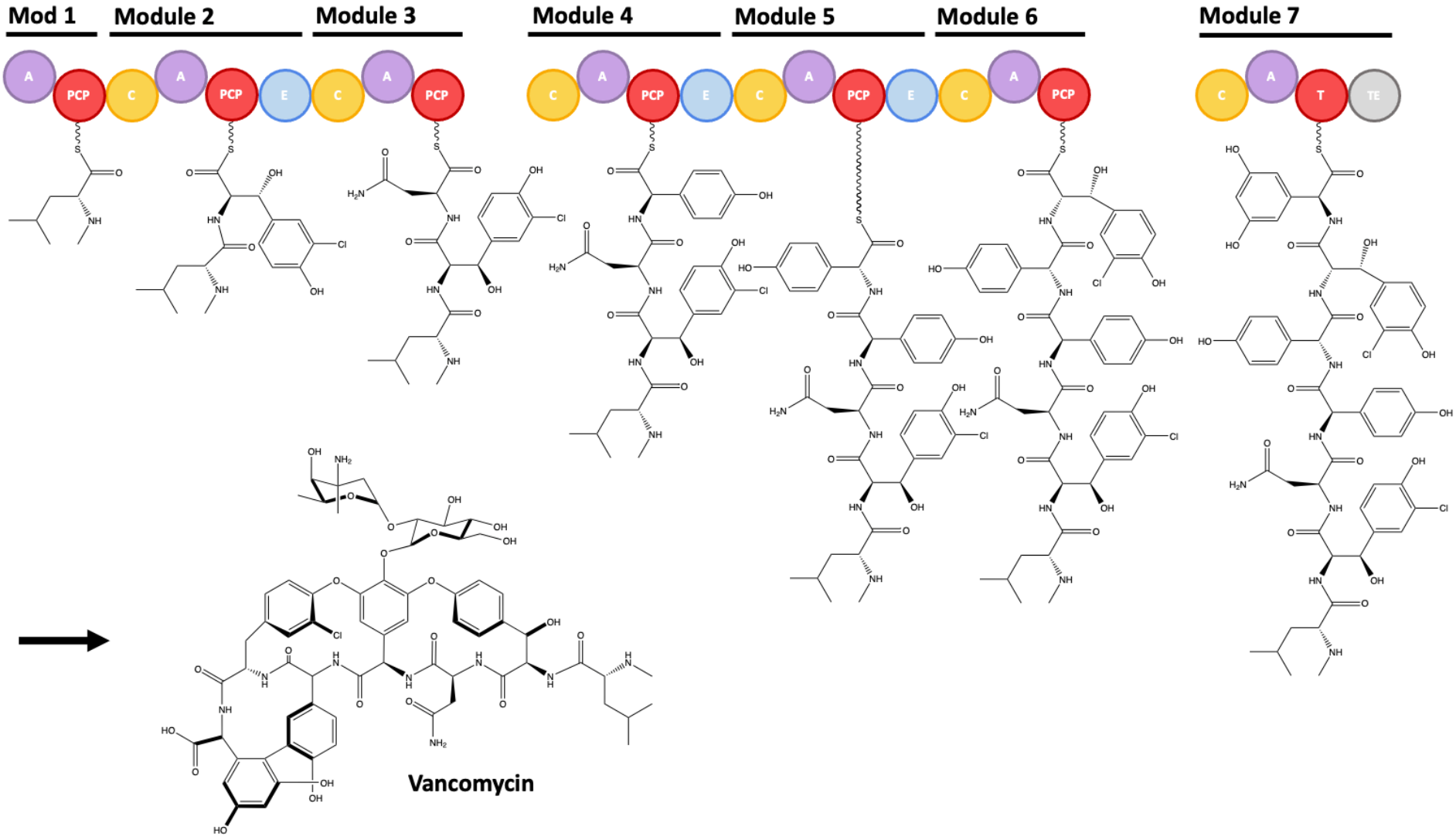


Figure 1.5 Diagrammatic representation of the biosynthetic pathway for vancomycin via the non-ribosomal peptide synthase. Module formation is shown through the lines above and individual domains are colour coded with abbreviations: **A** adenylation, **PCP** peptidyl carrier protein, **C** condensation, **E** epimerisation, **TE** thioesterase. Adapted from Wenski, Thiengmag and Helfrich, 2022

1.3.3 Ribosomally Synthesised and Post-Translationally Modified Peptides

Another large class of natural products are the ribosomally synthesised and post-translationally modified peptides, known as RiPPs. As suggested by the name, this family are made up of short peptides that are synthesised by the ribosome from the 20 proteinogenic amino acids and are then post-translationally modified by tailoring enzymes to create the final natural product. The variety of post-translational modifications that can be carried out results in a significant level of structural diversity within this class alongside a broad range of activity and targets. Typically, the precursor peptide for a RiPP consists of a core peptide, a leader peptide and occasionally a recognition sequence. The core peptide exists at the C-terminus and is the part of the precursor that is modified to create the final product. The leader peptide is present at the N-terminus and is involved in the recognition and binding of modifying enzymes and peptide cyclisation. If a recognition sequence is present, it sits at the C-terminus and carries out a similar function to the leader peptide. In eukaryotes, there may also be an N-terminal signal peptide that targets the peptide to the correct subcellular location (Arnison et al., 2013). The peptide undergoes a series of post-translational modifications before the leader peptide is cleaved by proteases to leave the mature peptide to be exported or used within the cell (**Fig 1.6**).

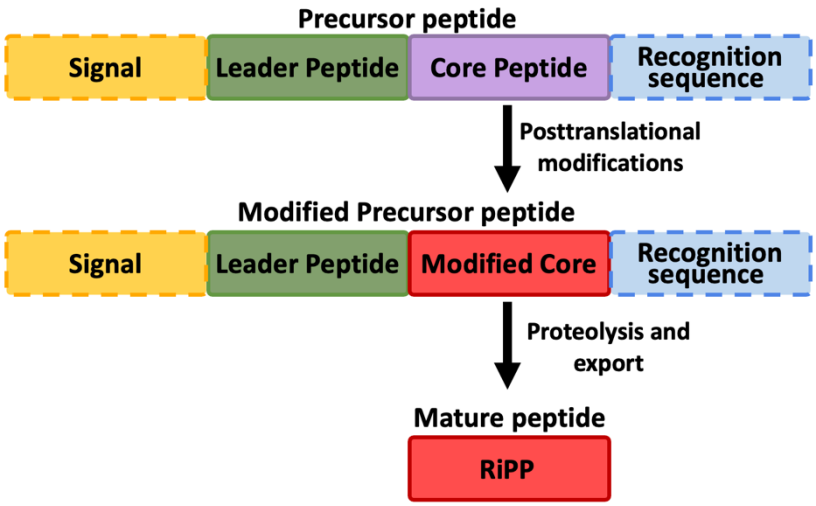


Figure 1.6 Diagrammatic representation of the general biosynthetic pathway of ribosomally synthesised and post-translationally modified natural products. The leader peptide acts to guide the post-translational modifications to the core peptide. Some contain an additional C-terminal recognition sequence and products of eukaryotic origin often contain an N-terminal signal sequence. Adapted from Arnison *et al.,* 2013

Whilst RiPPs all abide by the same biosynthetic principles, the diversity seen in this class of natural products is vast (**Fig 1.7**). This is largely due to the wide number of post-PKS tailoring enzymes that can act on the precursor peptides. While there is a high degree of homology between modifying enzymes within each family of RiPPs, there are no core domains or genes that are conserved among all RiPPs. The leader peptides act to guide the order and type of modifications that take place to generate the final mature peptide. This biosynthetic insight means it is possible to anticipate the precursor sequences and type of post-translational modifications that will take place for products within the same family, however the number and regioselectivity are often very hard to predict (Montalbán-López et al., 2021).

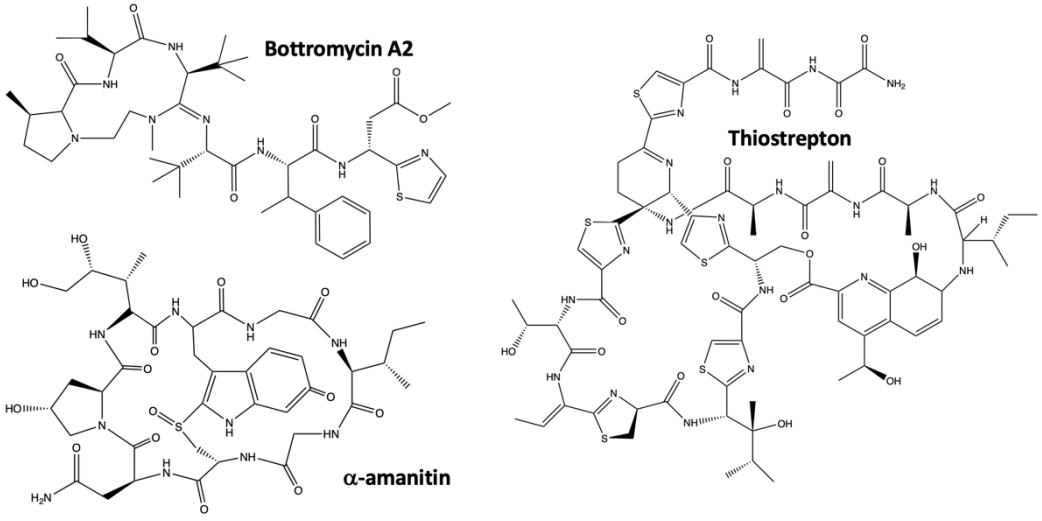


Figure 1.7 Structure of several ribosomally synthesised and post-translationally modified peptide natural products

1.3.4 Other Classes of Natural Products

In addition to the products derived from the PKS, NRPS and RiPP pathways, there are a variety of other compounds produced as part of secondary metabolism. The biosynthetic elements are more often than not encoded within BGCs and have evolved from those that are used in primary metabolism. The building blocks used in these systems to create the natural products are generally produced by the primary metabolism but can also be obtained from outside the cell. Terpenes, indole alkaloids and saccharides are just some examples of these classes of products that exist, but hybrid clusters have also been discovered where different elements of these natural product pathways work together to create the final product.

1.3.5 Discovery of New Natural Products

While an array of natural products and their analogues are already being used in a variety of settings both clinical and non-clinical, the need for bioactive substances continues to grow with increasing threats of AMR. Recent advances in genome sequencing technology means that the current knowledge and experience around identification and overproduction of these secondary metabolites can be applied to the discovery and application of new compounds. The biggest aspect of this is seen in the ongoing development of prediction tools which has revealed a wealth of unexploited potential in a large number of the antimicrobial producers that are already known to research efforts. The initial emergence of this idea of untapped natural products was seen when the full genome sequence of *Streptomyces coelicolor* was published in 2002 (Bentley et al., 2002). Despite having been a model organism since the 1960s, only five antibiotic gene clusters had been identified or described in any detail prior to this point. However, the genome sequence revealed BGCs for up to 30 secondary metabolites, the majority of which had not been characterised as they are not produced under normal laboratory conditions. Since then, the strain has been subject to significant investigation and provided an excellent platform for the development and application of synthetic biology tools. Predictive bioinformatic tools such as antiSMASH are used as part of the so called "genome mining" process (Blin et al., 2019). This allows for the identification of known BGCs to prevent the possibility rediscovery while also identifying these silent BGCs that may be the key to finding novel compounds with unique mechanisms of action against infectious diseases. Since the initial discovery of *S. coelicolor's* potential, *Streptomyces* bacteria have remained at the forefront of efforts to genome mine for natural products.

1.4 *Streptomyces* Bacteria

Belonging to the Actinomycetaceae family, *Streptomyces* are filamentous, Gram-positive bacteria characterised by their somewhat unusual life cycle and complex secondary metabolism. These facultative anaerobes are found in a multitude of environments but due to their dominance and the vast number of clinically relevant strains that have been isolated from the heterogenous environment, are commonly referred to as soil-dwelling. However, *Streptomyces* spp. have been isolated from marine, desert, volcanic and other such extreme environments (Sivalingam et al., 2019). The ability to survive such harsh conditions can, in

part, be attributed to their life cycle which encompasses three distinct development stages that are more comparable to that of fungi than other bacteria (**Figure 1.8**). Unigenomic spores that are encased in a thick hydrophobic outer layer contain small protective molecules such as heat shock proteins and are able to survive and remain dormant t(Flärdh & Buttner, 2009)itions (Flärdh & Buttner, 2009). Once favourable conditions are encountered, spores germinate to form filamentous vegetative hyphae that branch into dense and complex networks via apical tip extension. The essential proteins DivIVA and FtsZ have been shown to mediate the polar growth and compartmentalisation, respectively (Schlimpert et al., 2016). Under stress conditions (nutrient depletion, oxygen limitation), reproductive aerial mycelium is formed, projecting upwards and away from the vegetative hyphae. Following chromosome replication and segregation, division of the mycelium occurs through synchronised and largely simultaneous separation and is mediated by FtsZ to form prespores that undergo maturation (Bobek et al., 2017; Elliot et al., 2014). Development of hydrophobic aerial hyphae has been shown to be governed by and require the activity of the *bld* genes, while sporulation and maturation relied on the *whi* genes, both named due to the appearance of their respective mutants (bald or white). Once *Streptomyces* bacteria reach this stage, the spores can disperse to new, potentially more favourable environments to start the cycle again.

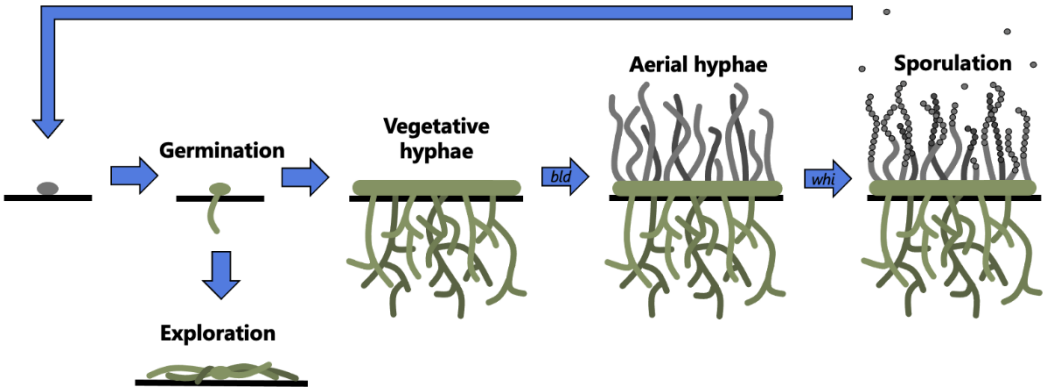


Figure 1.8 A basic overview of the lifecycle of *Streptomyces* to include both the classical and novel exploratory growth. In the classical pathway a single spore germinates and germ tubes grow by hyphal tip extension to create a branched network of vegetative hyphae within the soil. In response to a variety of signals including nutrient deprivation *bld* genes facilitate the upwards growth of aerial hyphae which then undergo chromosome segregation and septation using the *whi* genes to form spores. These spores are then dispersed into the environment to find nutrient rich conditions and begin the cycle again. In the divergent pathway of exploration signals such low glucose or the presence of fungi, cells rapidly develop as non-branching vegetative hyphae in a process independent of the *bld* and *whi* genes. Original figure based on data from Jones *et al*., 2017 and Jones *et al.,* 2018.

However, an additional developmental stage has recently been identified as an alternative reaction to nutrient depletion. The explorer phenotype, identified in the model organism *Streptomyces venezuelae* and replicated in some other *Streptomyces* strains is triggered by a reduction in glucose and/or iron availability to elicit a pH induced, relentless and rapid growth phase. Signalling between distant colonies to enter this growth phase is also made possible by release of the volatile organic(Jones & Elliot, 2018)ne (TMA) (Jones & Elliot, 2018). During exploration, non-branching vegetative hyphae develop significantly faster than is observed in the "classical" growth phase, alongside hydrophilic aerial hyphae. Confirmation of this new morphologically distinct development stage was shown when the explorer phenotype could still be seen in *bld* and *whi* mutants which were still capable of traversing biotic and abiotic surfaces (Jones et al., 2017). It is not fully understood whether these explorer cells sporulate or have the ability to return to the classical life cycle, so is instead considered an alternative branch of the *Streptomyces* development pathway (Jones & Elliot, 2018).

Streptomyces genomes contain DNA with an unusually high GC content, typically 67-78%, arranged in large linear chromosomes roughly 8-15 Mb in size (Law et al., 2018). The distinct phenotypic variation in secondary metabolite production between strains is largely attributed to the arrangement of the chromosome. A central core region contains essential genes involved in basic growth and development such as cell division, DNA replication, transcription and translation and central metabolism. This portion of the genome is highly conserved between *Streptomyces* species and other actinomycetes as shown in the high degree of synteny between *Mycobacterium tuberculosis* and *S. coelicolor* (Bentley et al., 2002). The chromosomal arms (roughly 1-2 Mb either end) contain largely non-essential and contingency genes involved in secondary metabolism and act as the source of most variation (Hopwood, 2006). The high level of chromosomal instability and the ability of genes to migrate between the core and contingent regions depending on essentiality results in incredibly varied recombination events between strains. Through the linear exchange of genetic material and/or gene duplication and divergence, strains can acquire or develop novel secondary metabolites and signalling molecules. The ability to effectively transfer such information is made possible by the existence of BGCs which include all the genes required to make one or several structurally similar natural products. BGCs typically contain the core biosynthetic machinery, accessory enzymes, expression control and self-resistance mechanisms involved in the production of secondary metabolites.

Advances in genome sequencing have provided insight into the potential of well-known *Streptomyces* strains and other actinomycetes. Such technology originated from the discovery that genes required for methlyenomycin were situated in clusters on just two transcriptional units (Chater & Bruton, 1985). These initial observations have been used to direct the development of high throughput computational analysis, which can screen genomes for cryptic BGCs. It was previously thought that most strains produced less than five secondary metabolites as this was all that was observed under standard laboratory conditions. However, the 651 *Streptomyces* genomes published on the antiSMASH database each contain 20 to 60 predicted BGCs (Blin et al., 2019; Ward & Allenby, 2018). The presence of silent gene clusters means many strains can continue to be used as a source of antibiotics with novel actions. However, traditional isolation methods have proved unsuccessful in exploiting their potential, instead efforts have turned to genome sequencing and editing tools alongside an in-depth understanding of regulation processes (Watve et al., 2001). Not only can this technology be used to unlock silent BGCs in well-known strains, but it can also identify significant potential in newly isolated ones alongside the genes that may be used to switch them on under lab conditions.

1.5 Regulation of Natural Products in *Streptomyces* spp.

The apparent arsenal of antimicrobial compounds available to most *Streptomyces* species through the BGCs they encode reflects the level of competition they must overcome to be able to thrive in various environments. Whilst this myriad of natural products is undoubtedly useful, their synthesis comes at a huge energetic cost to the producing organism. Such a vast capacity requires strict regulation to balance primary and secondary metabolism, preventing overproduction of natural products to the detriment of growth while maintaining the capability to outcompete other microbes. It is thought that the large number of regulatory elements encoded within the genome is related to this extensive secondary metabolism potential, allowing antimicrobials to be “selected” by the bacteria depending on factors such as the target or precursor availability (Devine et al., 2017; Van Wezel & McDowall, 2011).

A wide range of intra- and extra-cellular signals can trigger a response through different regulatory systems and cascades that function at different levels. High level regulation is mediated through global regulators that exist outside of BGCs and affect both the

biosynthesis of natural products and central metabolic processes by either directly interacting with relevant genes or indirectly controlling low level regulators. Cluster situated regulators (CSRs), as implied, are encoded within BGCs and control the production of their cognate secondary metabolite. However, some CSRs have been shown to have pleiotropic effects, mediating multiple biosynthetic pathways or acting on both primary and secondary metabolism. In addition, some clusters encode multiple regulators, some of which act to repress biosynthesis while others activate (Aigle & Corre, 2012). Regulatory cascades often involve cross-talk between these high and low level regulators which can themselves be separately triggered by a diverse array of signals (Chater, 2016; G. Liu et al., 2013). This makes the regulatory pathways of most natural products produced by *Streptomyces* species incredibly complex, and as such they are often far from fully characterised with activating signals or biochemical mechanisms remaining unknown. This might also explain why so many BGCs are not expressed, or rather their final products are not observed, under standard laboratory growth conditions. These clusters are thought to hold the greatest potential as they could encode the biosynthesis of new antimicrobials with yet unknown structures of compounds or molecules that have unique mechanisms of action (Rutledge & Challis, 2015).

Through a more in-depth knowledge of this metabolic regulation and by utilising modern genome editing technology, it may be possible to switch on silent biosynthetic pathways and engineer bacteria to overproduce such secondary metabolites for industrial production. It is first important to gain an in depth understanding of the different types of regulators present and combine knowledge on how they have already been shown to function with this genetic information of where they are present to work out how they could be used to identify new products.

1.5.1 Regulation of Growth, Development and Secondary Metabolism

The life cycle of *Streptomyces* bacteria is tightly regulated, and sporulation is associated with the onset of secondary metabolite biosynthesis, with the *bld* family of genes being able to co-ordinate the two. For example, BldD is a pleiotropic DNA binding protein that regulates the developmental progression from vegetative growth to the formation of aerial hyphae. This transition state regulator has a vast regulon comprising at least 167 genes, 42 of which encode further regulators (Tschowri et al., 2014). Among these targets are other

bld regulators and *whi* genes which are associated with the sporulation stage of their morphological differentiation (Bush et al., 2013; Den Hengst et al., 2010). For example, interaction with BldC, another DNA-binding protein, controls the expression of genes within the moenomycin A, actinorhodin and undecylprodigiosin BGCs in various *Streptomyces* species (Hunt et al., 2005; Makitrynskyy et al., 2020). BldD itself is also able to directly control the activity of promoters within the BGCs encoding erythromycin and daptomycin, further demonstrating its global capabilities. The activity of BldD is largely controlled by the ubiquitous second messenger, cyclic-di-GMP, which assembles into a tetramer that then bind two molecules of BldD in a dimer. Therefore, when cyclic-di-GMP levels are high, this complex is able to repress genes involved in the sporulation cascade via sequestration of BldD and interact with others involved in antibiotic biosynthesis. However, when cyclic-di-GMP levels are low, the complex does not form, transcripts are produced, and the cells progress to the next stage of development (Schumacher et al., 2017; Tschowri, 2016).

Another aspect of managing development and growth alongside secondary metabolism is the "stringent response" where resources are diverted away from growth and development and towards amino acid and/ or fatty acid biosynthesis by altering gene transcription when the cell is put under environmental stress. This observation was initially made when it was noted that antibiotic production in *Streptomyces* spp. was usually limited to cultures in the stationary phase of growth or growing at slower than normal rates (Strauch et al., 1991). The response is now known to be mediated by the guanosine tetraphosphate (ppGpp) which modulates gene expression via its interaction with RNA polymerase (Hesketh et al., 2007). The nucleotide is a universal stress response effector among bacteria and synthesised from GTP/ GDP and ATP when amino acids are used up (Hobbs & Boraston, 2019). As nutrients in the cell are depleted, concentrations of ppGpp increase and a shift in gene transcription is seen. Amino acid biosynthetic genes, including those that act as precursors for antimicrobials, are upregulated while those involved in growth and development are downregulated (Ochi, 1987; Strauch et al., 1991). Whilst a reduction in growth would initially seem counter-intuitive, this metabolic shift allows the *Streptomyces* spp. to out-compete other bacteria in the environment for the limited resources that are available by challenging them with antibiotics.

1.5.2 *Streptomyces* Antibiotic Regulatory Proteins (SARPs) Family

The *Streptomyces* antibiotic regulatory proteins, or SARPs, are a well-established family of activators of antibiotic biosynthesis. They exist in BGCs for a wide variety of natural product gene clusters and have only been found in actinobacteria, but mainly *Streptomyces* and are the largest family of regulators in the species (G. Liu et al., 2013). They contain a winged helix-turn-helix motif in the N-terminus that facilitates DNA binding, and their binding motifs are typically heptameric repeat sequences in the major groove of promoter regions of the genes that they regulate. The C-termini contain a bacteria activation domain that recruits RNA polymerase to the promoter site, facilitating transcription of the genes (Romero-Rodríguez et al., 2015).

When evaluating the prevalence of SARPs in *Streptomyces* BGCs, 57 out of 236 clusters reviewed were found to have at least one SARP encoded within them. While they are seen throughout clusters encoding a variety of antibiotic classes, 43 of these 57 were for polyketides (G. Liu et al., 2013). One of the most well-known SARPs is ActII-4 from *S. coelicolor* which initiates the production of actinorhodin by regulating the transcription of all five transcripts required for its biosynthesis. Other SARPs include DnrI (daunorubicin biosynthesis in *Streptomyces peucetius*), FdmR1 (fredericamycin in *Streptomyces griseus*) and CcaR (cephamycin-clavulanic acid in *Streptomyces clavuligerus*) (Chen et al., 2008; Sheldon et al., 2002).

1.5.3 Multiple Antibiotic Resistance Regulators Family

Named from the *E. coli* multiple antibiotic resistance regulator, the MarR family of transcription factors are seen throughout the bacterial kingdom. They exist as obligate homodimers with two, winged helix-turn-helix DNA-binding motifs. The recognition helix in each motif binds to a palindromic DNA sequence within consecutive major grooves while the corresponding wing sits in the minor groove (Gupta & Grove, 2014). MarR proteins often repress the transcription of target genes they regulate as these binding sequences typically fall within the promoter regions of transcripts, so the binding of RNA polymerase is physically blocked. However, there are several examples where they function as activators or even bifunctional to repress some genes while activating others. In association with the

palindromic binding sequence, this means they are also capable of regulating divergently oriented genes that would be expressed on different transcripts (Grove, 2017).

The activity of these regulators can be modulated through cysteine oxidation, but most are controlled by ligand or phenolic acid binding. When cysteine oxidation is the regulatory control mechanism, the formation of disulphide bonds in response to oxidative stress results in this change in DNA binding. Ligand and DNA binding takes place within the same domain, with a "hot spot" existing at the interface between the dimerization and DNA-binding interfaces. When a ligand binds, a conformational change occurs in the transcription factor, altering its interaction with the DNA binding sequence and exerting its regulatory effects on its target genes (Perera & Grove, 2010). If the MarR protein is responsible for controlling the expression of biosynthetic genes, the ligand it binds is often the substrate of the enzyme, thus providing a regulatory feedback mechanism. MarR proteins are often involved in the regulation of genes that control antibiotic export such as efflux pumps and in these cases the ligand is typically the compound that is exported by that efflux pump (Romero-Rodríguez et al., 2015).

The MarR family are known to regulate a multitude of processes within *Streptomyces* bacteria including central metabolism, stress responses, virulence and antibiotic biosynthesis. They are the fourth most abundant family of transcriptional regulator within *Streptomyces* species with an average of 50 encoded in each genome, but only a few of these have been fully characterised (Romero-Rodríguez et al., 2015). Examples include SAV4189 which activates avermectin biosynthesis in *Streptomyces avermitilis* and DptR3 which activates daptomycin biosynthesis in *Streptomyces roseosporus* (Guo et al., 2018; Q. Zhang et al., 2015).

1.5.4 LuxR Family

Another family of DNA-binding regulator proteins are those belonging to LuxR, named after the first protein within the family. In *Streptomyces* spp. there are an abundance of LuxR regulators, some of which are cluster situated and involved in the regulation of their cognate antibiotics, while others have a global effect on antibiotic biosynthesis. For example, SlgR2 and SlnM both regulate the production of streptolydigin and natamycin respectively in *Streptomyces lydicus* while AbsA2 is involved in the regulation of both

actinorhodin and undecylprodigiosin biosynthesis in *S. coelicolor* (Anderson et al., 2001; Gómez et al., 2012; Wu et al., 2014).

In general, LuxR proteins are less than 250 residues in length with an N-terminal autoinducer signal binding domain and a C-terminal helix-turn-helix DNA binding domain. They typically function as activators but may be repressors or even have a dual role in their regulatory actions (Patankar & González, 2009). A sub-family of LuxR proteins, known as large ATP-binding regulators of the LuxR family (LAL) are the predominant type that are found in *Streptomyces* bacteria and have a nucleotide triphosphate binding motif in their N-terminus. One well characterised example of this is PikD which functions as the primary biosynthetic activator of the multi-drug encoding macrolide PKS in *Streptomyces venezuelae*. Its activity regulates the production of pikromycin, methymycin, neomethymycin and narbomycin (Wilson et al., 2001).

1.5.5 TetR Family

Another significant group of regulators are the TetR family, named after the tetracycline repressor that was the first of the family to be characterised. These regulators are particularly abundant in bacteria that are exposed to a range of environmental changes, such as *Streptomyces* spp. They contain a helix-turn-helix DNA binding domain on the same polypeptide as a ligand binding domain with the binding of the latter inducing a conformational change that alters the DNA binding interaction. TetR family proteins are often encoded within their cognate BGC and regulate the expression of the self-resistance mechanism associated with the final product(s). For example, JadR functions within a complex regulatory system to activate jadomycin biosynthesis in *S. venezualae* (G. Xu et al., 2010).

1.5.6 Other Regulators

A large number of other regulator protein families exist across *Streptomyces* species, allowing for tight control over biosynthesis levels of the wide range of natural products that they encode. Examples include LysR, OmpR, LacI, WhiB and MerR all of which play a role in regulating various pathways and act within regulatory cascades to alter gene expression and antibiotic biosynthesis (Romero-Rodríguez et al., 2015). Another notable group of

regulators are two component systems, which are the focus of this work and discussed in more detail in later sections. The combinatorial effects of regulators in their activation, repression or switching capabilities means that the metabolic output of *Streptomyces* bacteria is constantly changing.

1.6 Two-Component Systems

A key part of surviving in a competitive environment with an ever-changing dynamic is the ability to sense these chemical and/ or physical changes as they happen and launch an appropriate response to them. For *Streptomyces* spp. this response often comes in the form of antibiotic biosynthesis, which, as already discussed, is a very tightly regulated process. This is partly achieved through the abundance of two-component systems (TCS) found throughout their genomes (McLean et al., 2019). These signal transducers act as a relay between a stimulus and the response that is elicited by the cell and are capable of responding to an abundance of triggers and transmitting signals that induce a variety of changes. TCSs fall into three distinct classes: classical, hybrid and phosphorelay. Classical systems are composed of a dimeric, typically transmembrane histidine kinase (HK) and a cognate response regulator (RR). In this system, an activating signal is detected in either the extracytosolic (or occasionally cytoplasmic) sensor domain of the HK. This signal recognition typically takes place through ligand binding which induces a conformational change in the intracellular kinase domain which autophosphorylates at a conserved histidine residue using ATP. This phosphoryl group is then transferred to a conserved aspartic acid residue in the receiver domain of the RR which also leads to a conformational change, often involving dimerization which activates the effector domain (**Fig 1.9**) (Zschiedrich et al., 2016). RRs are grouped into families based on the mechanism they use to modulate the expression of target genes; however, the majority fall in to the OmpR/PhoB subfamily with conserved N-terminal α/β receiver domains and C-terminal DNA binding domains (McLean, Wilkinson, et al., 2019). Hybrid systems have the HK and RR fused together in a complex, while phosphorelays use phosphotransferases to convey the signal between each component (Groisman, 2016). An additional two types of TCSs exist that can be seen in any of the three classes, convergent and divergent. Convergent systems are seen where multiple HKs are able to activate a single RR, meaning one regulator's activity can be triggered by multiple input signals (Lazar & Tabor, 2021). Divergent TCSs exist when a single HK is able to activate

multiple RRs through phosphorylation, meaning a myriad of processes can be seen in response to a single input signal (Fu et al., 2019). In most cases the genes encoding HKs and RRs are very close or next to one another and cannot function without the other. However, unpaired HKs and orphan RRs can be found in *Streptomyces* genomes, often with little or completely unknown functions. There has been some indication of cross regulation between unpaired HKs and orphan RRs, but this in itself raises questions as to their true status as "orphan" TCS elements (Raghavan & Groisman, 2010).

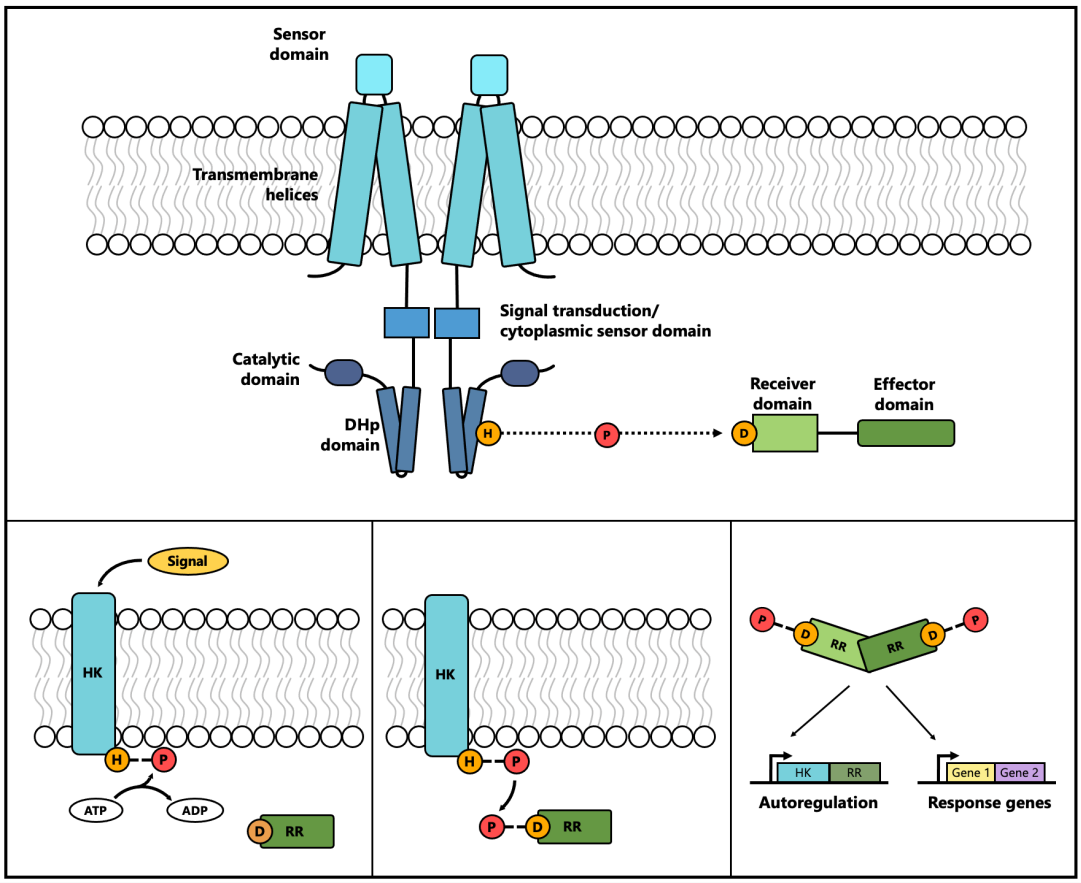


Figure 1.9 An overview of a classical two component system **A** The structural domains found within the histidine kinase (blue shades) and the response regulator (green shades), including the locations of the conserved (orange) histidine and aspartic acid residues, respectively, that relay the phosphate group **B** (**i**) upon receiving an activating signal in the extracytosolic sensor domain the histidine kinase autophosphorylates at the conserved histidine residue in the DHp domain (**ii**) this phosphoryl group is then relayed to the conserved aspartic acid residue in the receiver domain of the response regulator (**iii**) the response regulator dimerises and often acts as a transcription factor that binds to promoter sequences to activate or repress response genes.

Almost 70% of RRs bring about a change in gene expression through a DNA binding domain that facilitates binding to target genes to alter their activity e.g. the OmpR and NarL subfamilies (Martínez-Hackert & Stock, 1997). Some regulators have enzymatic activity that is involved in secondary messenger regulation as part of a cascade, while others bind to

alternative cellular components such as RNA or the transcriptional machinery to modulate gene expression (Djordjevic et al., 1998). The genes regulated by TCSs are vast in number and have been implicated in a broad range of functions. In *Streptomyces* and their close relatives, the ability of TCSs to mediate secondary metabolite biosynthesis in response to environmental changes provides great potential for upregulating the expression of BGCs. TCSs also act to demonstrate the incredible level of cross regulation that occurs during antibiotic biosynthesis. For example, deletion of *macRS* in *S. coelicolor* represses the production of actinorhodin while deletion of *phoRP* in the same organism results in actinorhodin upregulation (M. Liu et al., 2019; Martín, 2004).

TCSs can also be cluster-situated and control the expression of genes within the same BGC but more commonly they act globally to control the expression of genes throughout the genome. Those that are cluster-situated tend to be unique to the cluster and only seen in multiple species if they contain the same BGC. Global TCSs are not generally associated with any one gene cluster but exist outside of BGCs and have a more pleiotropic effect. Those with a global effect are more likely to be conserved across the genus as are the genes they control.

1.6.1 Highly Conserved Two Component Systems

Analysis of 93 *Streptomyces* genomes showed that there are 15 TCSs that are highly conserved throughout the genus (McLean et al., 2019). Of these, 12 have been characterised to varying extents while three remain largely unresearched and at present their roles and functions are unknown. Whilst those involved in processes other than regulation of antibiotic biosynthesis fall outside of the scope of this thesis, it is important to demonstrate the variety of systems that these proteins are able to influence and the fascinating number of roles they are able to carry out.

One such example of a highly conserved global regulator is MtrAB; which is not only seen throughout the *Streptomyces* species but is present in all actinomycetes. The RR MtrA has multiple binding sites throughout the genome with targets involved in DNA replication, cell division and antibiotic biosynthesis (Som, Heine, Holmes, Munnoch, et al., 2017). The extensive range of processes influenced by MtrA is demonstrated by the myriad of phenotypic and metabolic differences that are seen between wildtype and Δ*mtrA* mutant strains. In *S. venezualae, S. coelicolor* and *S. lividans* deletion of *mtrA* results in a bald

phenotype which is likely a result of MtrA interacting with several genes involved in aerial hyphae formation and others that regulate development and sporulation (Zhang et al., 2017). In *S. coelicolor* MtrA directly represses the production of the antibiotics actinorhodin and undecylprodigiosin via interactions with the promoters of *actII*-4 and *redZ* respectively, while in *S. venezualae* it represses the production of chloramphenicol and jadomycin with binding sites identified in gene promoters within each of their respective BGCs (Som et al., 2017; Zhu et al., 2020). Additional MtrA binding sites exist throughout the genome of most *Streptomyces* strains including within known BGCs and the implications of this binding are still being investigated.

Another example of a highly conserved TCS among *Streptomyces* spp., and arguably one of the most extensively studied systems is PhoPR. It acts as a major component of phosphate control in *Streptomyces* species, with the response regulator PhoP being classified as a master regulator. The system is able to control multiple processes in response to changes in the inorganic phosphate levels that are available from the external environment (McLean et al., 2019). PhoP is known to recognise and bind to promoters that contain PHO boxes which are highly conserved 11 nucleotide direct repeat units (Sola-Landa et al., 2005). Genes controlled by PhoP are involved in phosphate scavenging, storage, transport, mobilisation and antibiotic biosynthesis. In phosphate limiting conditions, PhoPR activates phosphate scavenging pathways to obtain the ester from both extra- and intra-cellular sources. It also delays morphological progression to the stationary phase until phosphate levels have returned to normal while upregulating the production of undecylprodigiosin and actinorhodin (Allenby et al., 2012).

The remaining conserved TCSs regulate a variety of systems and processes including developmental stages, primary and secondary metabolism. They are often associated with complex cross regulation and are able to respond to multiple signals to elicit multiple responses within the cell. The ability to sense and respond to these signals is essential for the survival of the cell and the co-ordinated regulation among the conserved systems remains at the forefront of scientific research interests.

1.6.2 Cluster Situated Two Component Systems

Alongside TCSs that are highly conserved, there are some that are unique to certain strains or BGCs and regulate the genes encoded in their cognate cluster. These are rare examples and often more specialised in terms of activating signal and have a discreet set of target genes that they regulate, although they may only control a single gene.

One such cluster-situated TCS is CinKR that is involved in the regulatory pathway of cinnamycin biosynthesis in *Streptomyces cinnamoneus*. The full mechanism of regulation is still under investigation, but is thought to be triggered by nutrient limitation that results in low level cinnamycin production. The lantibiotic then binds to the HK CinK which in turn phosphorylates CinR enabling the activation of a resistance associated transporter to prevent any deleterious effects of cinnamycin build up within the cell. Whilst CinRK does not directly regulate any biosynthetic genes, the presence of export enzymes is absolutely essential to cinnamycin production (O'Rourke et al., 2017).

Recently another cluster situated TCS has been identified and shown to be directly involved in the regulation of antibiotic production via interaction with biosynthetic genes. The novel TCS ForGF is the main activator of the biosynthetic pathway that produces the formicamycins, fasamycins and formicapyridines in *Streptomyces formicae*. The way in which this TCS regulates production of these compounds is the focus of this thesis and current knowledge surrounding the strain, its natural products and regulation will be discussed in more detail.

1.7 S*treptomyces formicae*

The discovery and utilisation of antibiotics in clinical and non-clinical settings has been through a significant timeline of events. From the point of formal discovery and classification to the development of resistance and the inevitable determination of humanity to overcome the problem. Rediscovery of known products remains a problem and attention has instead turned to exploring niche environments in the hope of finding novel bacterial stains and associated compounds. In association with current sequencing and genome refactoring tools it is hoped that any such strains could be utilised to their full antimicrobial producing potential.

1.7.1 Symbiosis, Discovery and Potential

Recently, the search for new strains has turned to under explored areas where mutualistic relationships exist between microbes and insects such as beetles, wasps and ants (Seipke et al., 2012). It is thought that there is a high possibility of symbiotic relationships that utilise natural products as a level of competition and defence in complex environments. One such relationship is seen in Attine leafcutter ants which cultivate the *Leucoagaricus gongylophorus* fungus garden. The ants forage for and harvest vegetation from their surroundings and bring it back to the fungus which digests it, producing nutrient-rich fruiting bodies that the ants are then able to use as their primary food source. In addition to plant material, the ants provide the fungus with protection from both mammalian and microbial threats (North et al., 1997). The latter of these is facilitated by a secondary symbiosis between the ants and *Pseudonocardia* bacteria that colonise crypts on the ant cuticle (Currie et al., 1999). The single strain of *Pseudonocardia* is passed between the colony via vertical transmission from the queen ant to all new worker ants. It produces anti-fungal agents such as dentigerumycin or nystatin P1 that prevent infection by pathogens like *Escovopsis weberi*, a co-evolved fungus with colony-collapsing potential (Barke et al., 2010; Oh et al., 2009). In association with a single strain of *Pseudonocardia*, *Streptomyces* species are acquired through horizontal transmission from the environment, generating a multitrophic community to provide more complex antimicrobial protection (Seipke et al., 2012). In this highly competitive environment, constant development is essential in what has been described as a ”chemical arms race” (Heine et al., 2018).

A protective mutualism may also occur in *Tetraponera penzigi* plant-ants where antibiotic producing bacteria are present within the ant’s domatia (specialised hollow nesting structures in the host plant) and may act to protect their resources(Baker et al., 2017). The ants colonise a myrmecophyte, *Vachellia drepanolobium* (commonly known as the Acacia plant), which they aggressively protect from herbivores through behavioural displays (Martins, 2010). In return the ants inhabit the domatia which form at the base of the hollowed-out thorns of swollen Acacia plants and where the ants cultivate their own fungus, *Chaetomium*. Within the bacterial community isolated from this environmental niche were Proteobacteria, Firmicutes and Bacteriodetes. Despite a distinct lack of actinobacteria within the community, it’s suspected that this relationship mirrors what is seen in Attine ants, potentially utilising a different class of antibiotic-producing bacteria as their predominant species (Qin et al., 2017). Actinobacteria were cultured from these domatia,

including *Streptomyces formicae* (KY5), which was confirmed as a novel strain through sequencing of the genome and comparison of key phylogenetic markers (16S RNA, *atpD*, *rpoB*, *gyrA*, *recA* and *trpB*) (Seipke et al., 2013). The 9.6 Mbp chromosome was shown to have a 71.38% GC content with 8162 protein coding sequences (Holmes et al., 2018). AntiSMASH analysis alongside manual inspection of the genome identified KY5 as a 'talented' stain containing at least 45 BGCs, more than the average 35 for other *Streptomyces* species. Bioassays of KY5 revealed it has potent activity against a variety of pathogens including *Bacillus subtilis*, *Candida albicans*, *Escherichia coli* and MRSA (**Fig. 1.10**) (Holmes et al., 2018; Qin et al., 2017). There is great potential for novel secondary metabolites to be produced from the myriad of gene clusters with low sequence identity to those with known products. Within the genome there are also well-conserved clusters such as those that produce geosmin, a conserved *Streptomyces* terpene associated with the smell of the rain. In addition, there are siderophore and osmolyte BGCs that encode desferroxamine and ectoine that likely aid with environmental stress response and survival. These "essential" clusters are all encoded towards the middle of the linear genome with the yet unknown clusters existing towards the chromosomal arms (**Table 1.1**).

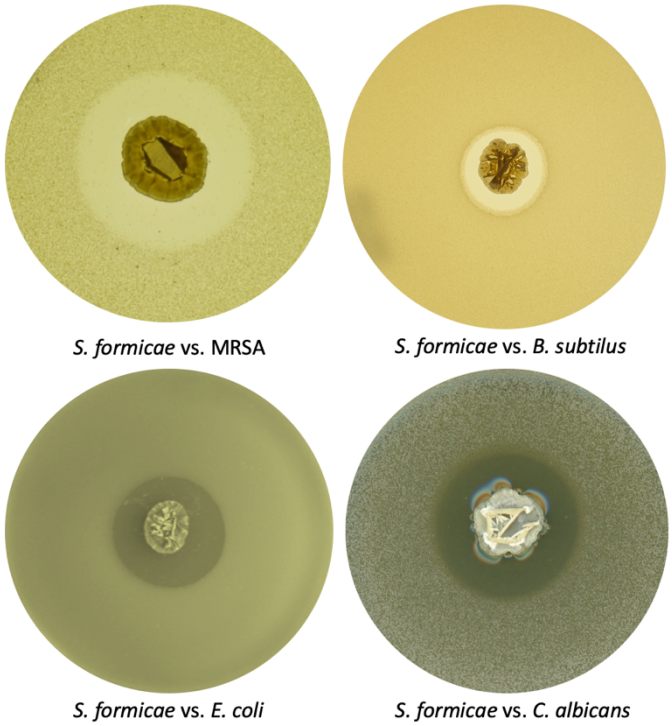


Figure 1.10 Bioassays of *Streptomyces formicae* against pathogens Methicillin-Resistant *Staphylococcus aureus*, *Bacillus subtilis*, *Escherichia coli* and *Candida albicans*.

Table 1.1 AntiSMASH (Version 6.0) analysis of the *Streptomyces formicae* genome. Where BGCs are further divided into a/b/c annotations, manual inspection has determined that the predicted gene clusters are likely to contain two or more BGCs that produce different natural products. It is therefore predicted that the strain may contain up to 49 BGCs.

BGC Number	Predicted BGC Type	BGC Number	Predicted BGC Type
1	RiPP Recognition Element Containing	24	NRPS
2	Other	25	Terpene
3	RiPP (Lanthipeptide)	26	Terpene
4	NRPS	27	Siderophore
5a	RiPP (Lanthipeptide)	28	Butyrolactone
5b	NRPS	29	RiPP
5c	T3PKS	30	Hybrid NRPS-T3PKS
6	NRPS	31	Terpene
7	T3PKS	32	Linear Azoline-containing Peptide (LAP)
8	NRPS	33	T2PKS
9	Terpene	34	Hybrid NRPS-T1PKS
10	Aminoglycoside	35	Redox Cofactor
11	NRPS	36	Terpene
12	NRPS	37a	RiPP (Lanthipeptide)
13	Hybrid NRPS-T1PKS	37b	T3PKS
14	Hybrid T1PKS-RiPP (Lanthipeptide)	37c	RiPP (Lassopeptide)
15	Terpene	38	T1PKS
16	Ectoine	39a	T3PKS
17	Hybrid T1PKS-NRPS-RiPP	39b	NRPS
18	RiPP (Lanthipeptide)	40	NRPS
19	Melanin	41	T1PKS
20	Siderophore	42a	NRPS
21	RiPP (Thiopeptide)	42b	β-Lactam
22	NRPS	42c	T1PKS
23	Linear Azoline-containing Peptide (LAP)		

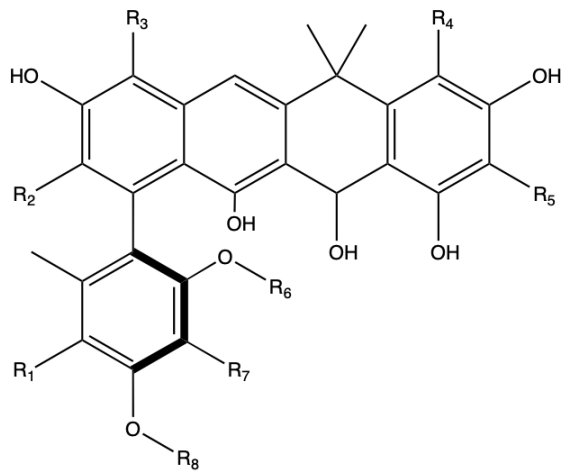
1.7.2 Novel Antibiotics from *Streptomyces formicae*

Initial investigations into the antimicrobial activity of KY5 used bioassay guided fractionation to isolate antibacterial and antifungal fractions. The latter proved difficult for structural elucidation while the former resulted in the purification and structural elucidation of 16 metabolites which were classified into two groups. The first three compounds were named fasamycins C-E due to their structural similarity to the previously described fasamycins A and B (Feng et al., 2012). The remaining compounds were named the formicamycins which

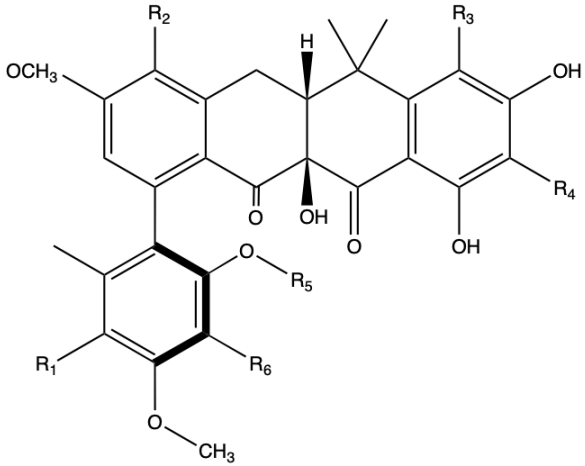
have pentacyclic ring structures and can be halogenated at up to four positions across the carbon back bone. Further study of the strain went on to identify an additional fasamycin congener and an entirely different class of secondary metabolites that were named the formicapyridines (Qin et al., 2017, 2019a). The structures of these compounds (**Fig 1.11**) allowed for the hypothesis that the biosynthetic pathway could be only carried out by a T2PKS with subsequent post-PKS modifications. Only one such T2PKS gene cluster was identified in the *S. formicae* genome, BGC 33 that was termed *for*. Using CRISPR/ Cas9, deletion of the *for* BGC attenuated fasamycin and formicamycin biosynthesis, while genetic complementation using an ePac carrying the entire BGC restored the production of 13 of the original 16 secondary metabolites to levels seen in the wildtype. Formicamycins K-M were only produced in the presence of sodium bromide; hence their production was not relevant in the context of genetic manipulation of the BGC. All 13 of these molecules exhibited potent activity against several strains on the World Health Organisation's watch list including MRSA and VRE (**Table 1.2**). Potency was shown to increase as the number of chlorine atoms incorporated into the compound increases and this again increases further when the chlorine groups are replaced by bromine. In general, the formicamycins were also shown to be more potent than the fasamycins (Qin et al., 2017). This is potentially due to their four halogenation positions as opposed to the two that are available in the fasamycins.

Table 1.2 MIC data for a selection of the fasamycins and formicamycins from *Streptomyces formicae* against *Bacillus subtilis*, methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococci*. Adapted from data in Qin et al., 2017.

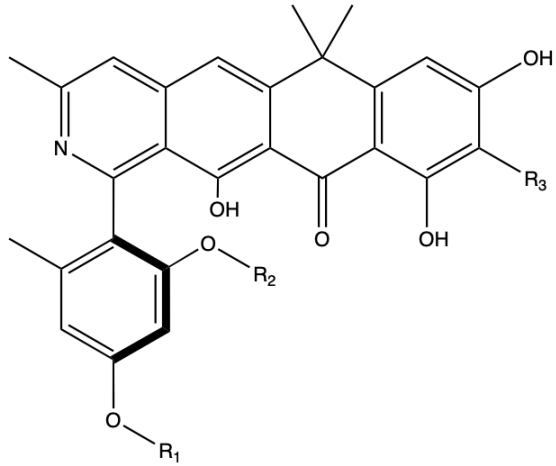
Compound	**Minimum Inhibitory Concentration (μM)**		
	B. subtilis	MRSA	VRE
Fasamycin C	<20	40	40
Fasamycin D	10	10	10
Fasamycin E	5	80	80
Formicamycin A	5	>80	>80
Formicamycin B	10	10	10
Formicamycin C	5	1.25	80
Formicamycin D	10	20	10
Formicamycin E	10	20	10
Formicamycin F	5	20	2.5
Formicamycin G	5	Not tested	Not tested
Formicamycin H	10	Not tested	Not tested
Formicamycin I	<2.5	<2.5	1.25
Formicamycin J	<20	0.625	1.25
Formicamycin K	<2.5	2.5	5
Formicamycin L	<2.5	1.25	2.5



Fasamycin C R_1 = H, R_2 = H, R_3 = H, R_4 = H, R_5 = H, R_6 = H, R_7 = H, R_8 = CH_3
Fasamycin D R_1 = H, R_2 = H, R_3 = Cl, R_4 = H, R_5 = H, R_6 = H, R_7 = H, R_8 = CH_3
Fasamycin E R_1 = Cl, R_2 = H, R_3 = H, R_4 = H, R_5 = H, R_6 = H, R_7 = H, R_8 = CH_3
Fasamycin F R_1 = H, R_2 = COOH, R_3 = H, R_4 = H, R_5 = H, R_6 = H, R_7 = H, R_8 = CH_3
Fasamycin L R_1 = Cl, R_2 = H, R_3 = Cl, R_4 = H, R_5 = Cl, R_6 = H, R_7 = H, R_8 = CH_3
Fasamycin M R_1 = Cl, R_2 = H, R_3 = Cl, R_4 = H, R_5 = H, R_6 = CH_3, R_7 = Cl, R_8 = H
Fasamycin N R_1 = Cl, R_2 = H, R_3 = Cl, R_4 = H, R_5 = H, R_6 = H, R_7 = H, R_8 = CH_3
Fasamycin O R_1 = Cl, R_2 = H, R_3 = Cl, R_4 = H, R_5 = Cl, R_6 = CH_3, R_7 = H, R_8 = CH_3
Fasamycin P R_1 = Cl, R_2 = H, R_3 = Cl, R_4 = Cl, R_5 = H, R_6 = CH_3, R_7 = Cl, R_8 = OH
Fasamycin Q R_1 = Cl, R_2 = H, R_3 = Cl, R_4 = Cl, R_5 = Cl, R_6 = H, R_7 = H, R_8 = CH_3



Formicamycin A R_1 = H, R_2 = Cl, R_3 = H, R_4 = H, R_5 = CH_3, R_6 = H
Formicamycin B R_1 = Cl, R_2 = Cl, R_3 = H, R_4 = H, R_5 = H, R_6 = H
Formicamycin C R_1 = H, R_2 = Cl, R_3 = Cl, R_4 = H, R_5 = CH_3, R_6 = H
Formicamycin D R_1 = Cl, R_2 = Cl, R_3 = Cl, R_4 = H, R_5 = H, R_6 = H
Formicamycin E R_1 = Cl, R_2 = Cl, R_3 = Cl, R_4 = H, R_5 = CH_3, R_6 = H
Formicamycin F R_1 = Cl, R_2 = Cl, R_3 = H, R_4 = Cl, R_5 = CH_3, R_6 = H
Formicamycin G R_1 = H, R_2 = Cl, R_3 = Cl, R_4 = Cl, R_5 = CH_3, R_6 = H
Formicamycin H R_1 = Cl, R_2 = H, R_3 = Cl, R_4 = Cl, R_5 = CH_3, R_6 = H
Formicamycin I R_1 = Cl, R_2 = Cl, R_3 = Cl, R_4 = Cl, R_5 = H, R_6 = H
Formicamycin J R_1 = Cl, R_2 = Cl, R_3 = Cl, R_4 = Cl, R_5 = CH_3, R_6 = H
Formicamycin K R_1 = H, R_2 = Cl, R_3 = Br, R_4 = Cl, R_5 = CH_3, R_6 = H
Formicamycin L R_1 = Cl, R_2 = Cl, R_3 = Br, R_4 = Cl, R_5 = CH_3, R_6 = H
Formicamycin M R_1 = H, R_2 = Br, R_3 = H, R_4 = H, R_5 = CH_3, R_6 = H
Formicamycin R R_1 = Cl, R_2 = Cl, R_3 = Cl, R_4 = Cl, R_5 = H, R_6 = Cl
Formicamycin S R_1 = Cl, R_2 = Cl, R_3 = Cl, R_4 = Cl, R_5 = CH_3 R_6 = Cl
Formicamycin T R_1 = H, R_2 = Cl, R_3 = H, R_4 = Cl, R_5 = CH_3, R_6 = H
Formicamycin U R_1 = Cl, R_2 = H, R_3 = Cl, R_4 = H, R_5 = CH_3, R_6 = H
Formicamycin V R_1 = Cl, R_2 = H, R_3 = Cl, R_4 = H, R_5 = CH_3, R_6 = H
Formicamycin W R_1 = H, R_2 = Cl, R_3 = Cl, R_4 = Cl, R_5 = CH_3, R_6 = H
Formicamycin X R_1 = H, R_2 = Cl, R_3 = Cl, R_4 = Cl, R_5 = OH, R_6 = H
Formicamycin Y R_1 = H, R_2 = H, R_3 = Cl, R_4 = Cl, R_5 = OH, R_6 = H



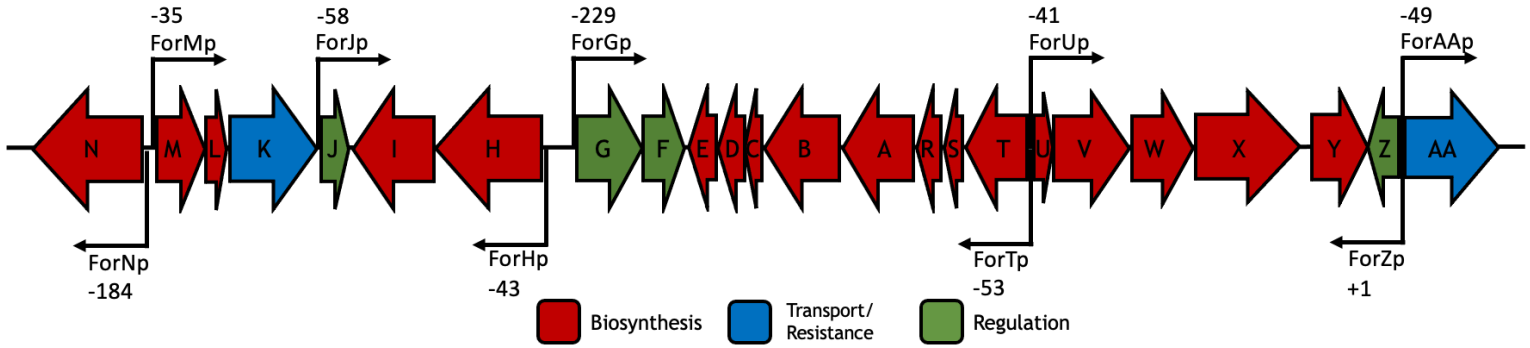
Formicapyridine A R_1 = H, R_2 = H, R_3 = H
Formicapyridine B R_1 = CH_3, R_2 = H, R_3 = H
Formicapyridine C R_1 = CH_3, R_2 = CH_3, R_3 = H
Formicapyridine D R_1 = H, R_2 = H, R_3 = Cl
Formicapyridine E R_1 = CH_3, R_2 = H, R_3 = Cl
Formicapyridine F R_1 = CH_3, R_2 = CH_3, R_3 = Cl
Formicapyridine G R_1 = H, R_2 = H, R_3 = Br
Formicapyridine H R_1 = CH_3, R_2 = H, R_3 = Br
Formicapyridine I R_1 = CH_3, R_2 = CH_3, R_3 = Br

Figure 1.11 Chemical structures of metabolites isolated from *Streptomyces formicae* and produced by the *for* biosynthetic gene cluster. Adapted from Qin *et al*., 2019

MRSA was maintained for 20 generations in subinhibitory concentrations of fasamycin or formicamycin without any observed resistance, showing significant potential for use of the compounds in a clinical setting, but this poses a challenge in identification of their mechanism of action (Qin et al., 2017). Fasamycins A and B (previously identified congeners in soil-derived DNA samples) have been shown to inhibit FabF in type II fatty acid biosynthesis (FASII), providing a potential line of investigation into their mechanism of action (Feng et al., 2012).

1.7.3 Biosynthesis and Regulation of Formicamycin

Based on a previously proposed biosynthetic pathway for the fasamycins, congeners identified from *S. formicae* were predicted to be precursors to the entirely new family of antibiotics called formicamycins. Of the 43 genes within the original ePAC clone of the *for* BGC, 24 genes expressed across nine transcripts were established to be part of this pathway through sequential deletion of genes from the edge of the cluster and determining the effect on formicamycin levels. Promoters were identified using cappable-RNA sequencing and showed 10 transcription start sites (**Fig. 1.12**) (Devine et al., 2021; Qin et al., 2017). Three regulatory elements were also identified within this core region of the cluster including two MarR family regulators (ForJ and ForZ), and the TCS (ForGF). Again, using CRISPR/Cas9, each of these regulatory elements was deleted to establish their roles. As previously described, deletion of the entire *for* cluster resulted in a total loss of formicamycin production, and interestingly deletion of the *forGF* TCS had the same effect, deeming it the main activator of the pathway. This TCS is comprised of a signal transducing HK (ForG) and a DNA-binding LuxR family RR (ForF). Basic modelling does not predict any transmembrane domains in ForG suggesting it may be cytosolic and its activating signal remains unknown. This TCS is the predominant focus of this thesis and will be discussed in significantly more detail throughout. Deleting the *forZ* gene reduced formicamycin production to around 65% of wildtype levels, with fasamycin precursors doubling. Meanwhile when the *forJ* gene was deleted, production of formicamycin increased 5-fold, meaning ForJ is a repressor of the pathway. While both ForJ and ForZ are MarR-regulators, they have relatively low sequence identity to one another suggesting they serve different functions or mechanisms in their regulatory action (Devine et al., 2021).



Gene and Number	Amino Acids	Putative Function	Gene and Number	Amino Acids	Putative Function
1. orf4	306	NAD-dependant epimerase/ dehydratase	23. forS	106	Monooxygenase/ cyclase
2. orf3	336	MarR family transcriptional regulator	24. forT	342	O-methyltransferase
3. orf2	199	Hypothetical protein	25. forU	119	Monooxygenase/ cyclase
4. orf1	170	Transposase	26. forV	430	Halogenase
5. forQ	422	Decarboxylase	27. forW	341	O-methyltransferase
6. forP	217	β-Lactamase (metallohydrolase)	28. forX	571	Monooxygenase
7. forO	259	Exodeoxyribonuclease III	29. forY	315	Oxidoreductase
8. forN	590	Acyl hydrolase	30. forZ	172	MarR family transcriptional regulator
9. forM	261	Methyltransferase	31. forAA	523	MFS family transporter
10. forL	113	PKS cyclase	32. forBB	220	LuxR family response regulator
11. forK	478	Na^+/ H^+ exchanger	33. forCC	417	Sensor histidine kinase
12. forJ	149	MarR family transcriptional regulator	34. orf6	321	ABC transporter
13. forI	455	ACC biotin carboxylase	35. orf7	284	ABC transporter permease
14. forH	607	ACC carboxyl transferase	36. orf8	529	Glutamate synthase
15. forG	363	Sensor histidine kinase	37. orf9	166	Hypothetical protein
16. forF	219	LuxR family response regulator	38. orf10	203	Hypothetical protein
17. forE	171	ACC biotin carboxy carrier protein	39. orf11	164	Hypothetical protein
18. forD	153	PKS cyclase/ dehydratase	40. orf12	247	Glutamate ABC transporter
19. forC	96	PKS ACP	41. orf13	310	ABC transporter substrate binding protein
20. forB	415	KS_β	42. orf14	214	ABC transporter permease
21. forA	420	KS_α	43. orf15	289	ABC transporter permease
22. forR	131	Cupin (cyclase/ monooxygenase)			

Figure 1.12 The genes within biosynthetic gene cluster 33 of *Streptomyces formicae* including the gene number, name, amino acid count and putative products. Included is as schematic of the genes responsible for formicamycin production, including annotations of their function and confirmed promoter sites. Adapted from Qin *et al*., 2017

Chromatin immunoprecipitation and sequencing data (ChIP-Seq) have provided greater insight into how these CSRs carry out their regulatory roles. ForF has been shown to bind between the divergent promoters for the *forGF* and *forHI* operons to potentially autoregulate its own expression and that or *forHI*. ForJ binds to multiple sites both within the *for* BGC and throughout the genome. Within the cluster these include the intergenic regions between divergent promoters controlling transcription of *forN*, *forMLK*, *forHI*, *forGF*, *forTRSABCDE* and *forUVWXY*, within the coding region of *forE* and its own promoter region pForJ (**Fig. 1.13**). This provides a potential mechanism for simply repressing transcription of all necessary machinery at these divergent promoter regions, a roadblock mechanism at *forE* extending to *forGF* and a similar mechanism for autoregulation at its own promoter (Devine et al., 2021; Qin et al., 2017). The ways in which these regulators bind and control gene expression to alter formicamycin biosynthesis is discussed in more detail throughout this thesis.

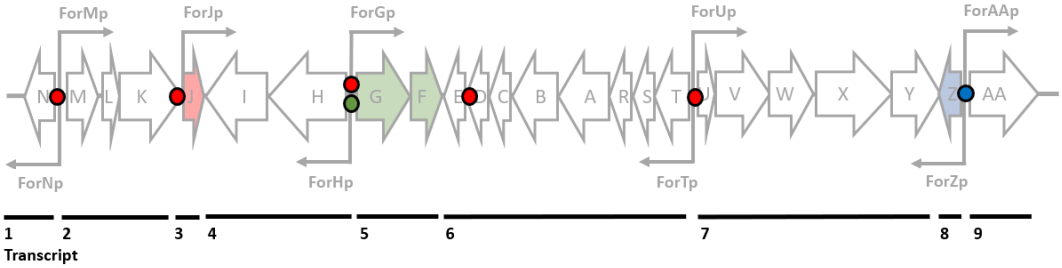


Figure 1.13 Cluster situated binding sites of the three formicamycin regulators. ForGF (green) binds the intergenic region between the divergent promoters ForGp / ForHp. ForJ (red) binds the intergenic regions between ForNp / ForMp, ForHp / ForGp and ForTp / ForUp as well as in the coding regions for ForJ and ForE. For Z (blue) binds the intergenic region between the divergent promoters for ForZp / ForAAp. Adapted from Devine *et al*., 2019

Isotope feeding experiments were used alongside genetic manipulation and analysis of the *for* BGC and comparative bioinformatics to establish a formicamycins biosynthetic pathway for the production of the formicamycins (**Fig 1.14**). ForABC act in unison to create the minimal polyketide synthase that is comprised of the two ketosynthase subunits KSα and KSβ (ForA and ForB, respectively) alongside the acyl carrier protein (ForC). These three enzymes act upon the poly-β-keto intermediate until the tridecaketide intermediate **1** is formed. A set of putative tailoring enzymes including dehydratases and cyclases (ForD, ForL and ForR) generate the next intermediate **2** before a hydrolase (ForN) and decarboxylase (ForQ) further convert this to the putative intermediate **3**. The first post-PKS modification is proposed to be a *gem*-demethylation at C18 as every formicamycin and fasamycin congener

contains two methyl groups at this position. The methyltransferase responsible for this addition is proposed to be ForT as it has the highest sequence identify with BenF, a SAM-dependent methyltransferase that carries out a similar reaction in the biosynthesis of benastatin. Alongside this, another methyltransferase (ForM or ForW) is proposed to carry out *O*-methylation at C3 to generate the non-halogenated **Fasamycin C**. This exact mechanism remains putative as it was not possible to isolate the putative intermediate 3 or any enzyme-free intermediates or congeners lacking the *gem*-dimethyl moiety so it may be that these methyltransferases instead act upon the ACP-bound poly-β-keto intermediate. Chlorination (ForV) of fasamycin C is the next suggested step as deletion of this gene results in accumulation of the previous intermediate and a complete lack of any other congeners. The ability to catalyse the addition of one or more chlorine molecules allows the biosynthetic pathway to progress and produce more congeners. In the example here, the mono-chlorinated **Fasamycin D** is used to demonstrate this. A C-ring modification is carried out by a Bayer-Villager monooxygenase (ForX) to generate the lactone intermediate **4** before a tertiary hydroxyl group is introduced at C10 via a flavin-dependent reductase (ForY), producing intermediate **5**. A second *O*-methylation takes places at C23 via either of the previously mentioned methyltransferases (ForM or ForW) to create intermediate **6**. The final step is proposed to take place through the combined action of the methyltransferases (ForM and ForW) and a promiscuous halogenase (ForV) to decorate the formicamycin structures with chlorine atoms, producing the different congeners as previously shown (Devine et al., 2021; Qin et al., 2017, 2019a).

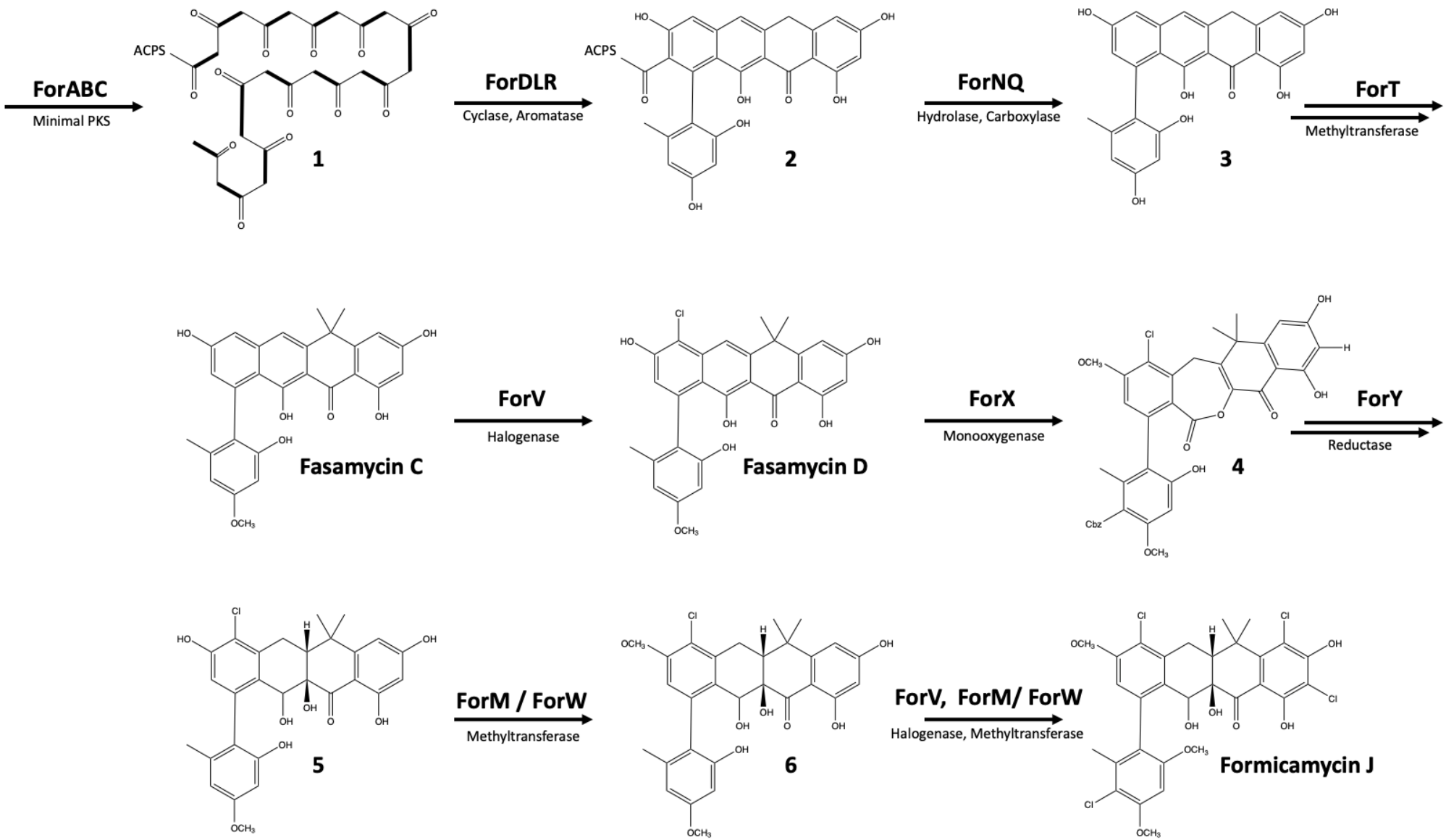


Figure 1.14 Proposed biosynthetic pathway of the formicamycins. Adapted from Qin *et al*., 2019

1.8 Aims of this Project

The aim of this thesis was to investigate the role of cluster-situated two-component systems in their control of antibiotic production in *Streptomyces formicae*. This was predominantly focussed on the TCS known to be involved in the regulation of formicamycin biosynthesis, ForGF, and built on previous work which fully characterised the biosynthetic pathway of formicamycin. A range of microbiological, biochemical and biophysical techniques were employed to elucidate the role of the TCS in the regulatory aspect this process. As part of this, a number of strains of *S. formicae* have been engineered to eradicate, overproduce or alter the biosynthesis of the formicamycins. The target DNA sequence of the response regulator was identified to provide insight into the mechanism of regulation and identify other targets outside of the BGC. Additionally, purification of these proteins in an attempt to solve their structure has revealed more information about their solubility and provided insight into the activating signal of the histidine kinase.

The generation of a small library of strains that overexpress the cluster-situated two component systems in *S. formicae* has resulted in further confirmation of the huge potential of natural products encoded within this strain. The changes in bioactivity and initial analyses of compounds will allow for further work to be conducted on this *Streptomyces* strain in the future to continue to unlock previously cryptic BGCs and the products they encode.

2. Materials and Methods

2.1 Chemicals and Reagents

Unless otherwise stated, the chemicals and reagents used were of laboratory grade or above and purchased from Sigma Aldrich (UK), Thermo Fisher Scientific (UK) or Merck (UK). All media and solutions were made using deionised water (dH_2O) unless otherwise stated.

2.2 Bacterial Strains and Growth Conditions

The bacterial strains used or generated in this study are listed in the appendix of this thesis (**Table 7.1**). Bacterial cultures were routinely grown in media prepared with analytical grade water (**Table 2.1**) and sterilised at 121 °C for 15 minutes. Any media additions (**Table 2.2**) were filter sterilised using 0.1 μM syringe filters. *Escherichia coli* strains were grown at 37 °C, statically when on agar plates and shaking at 220 rpm when in liquid media. *Streptomyces* spp. were grown at 30 °C, statically when on agar plates and shaking at 220 rpm when in liquid media. Any exceptions will be made clear as necessary. Growth of any culture was measured spectrophotometrically by recording optical density at 600 nm in a 1 cm cuvette.

Table 2.1 Composition of growth media used in this work

Media	Composition of 1L	Water	pH
LB	10 g tryptone 5 g yeast extract 10 g NaCl (omitted for selection with hygromycin) +/- 20 g agar	Deionised	N/A
SFM	20 g soy flour 20 g mannitol 20 g agar	Tap	N/A
MYM	4 g maltose 4 g yeast extract 10 g malt extract 20 g agar	50:50 tap: deionised	7.3

SNA	4 g Difco Nutrient Broth Powder 5 g agar	Deionised	N/A
TSB	17 g tryptone 3 g soya peptone 5 g NaCl 2.5 g dipotassium phosphate 2.5 g glucose	Deionised	7.3
2xYT	16 g tryptone 10 g yeast extract 5 g NaCl	Deionised	7.0

Table 2.2 Antibiotics used in this work and their final selective concentration

Media Addition	Selective Concentration (μg/mL)
Apramycin	50
Chloramphenicol	25
Hygromycin	50
Kanamycin	50
Nalidixic Acid	25

2.2.1 Preparation of *Streptomyces* Spore Stocks

A single *Streptomyces* colony was streaked onto SFM to give a confluent lawn and grown for 7-10 days or until spores were visible. To harvest the spores, 2 mL 40% glycerol was used alongside a single sterile cotton bud to lift spores from the plate. Harvested spore stocks were then stored at -80 °C.

2.2.2 Preparation of Glycerol Stocks

Stocks of *E. coli, B. subtilis, S. aureus* and *C. albicans* were produced by resuspending 1.5 mL of an overnight culture in a stocking media (LB-Miller, 20% glycerol). The harvested stocks were then stored at -80 °C.

2.3 General Microbiology

2.3.1 DNA Extraction

Genomic DNA was extracted from *Streptomyces* by centrifuging 1 mL overnight culture at 13,000 rpm, for 5 minutes and resuspending in 100 μL Solution I (50 mM Tris/ HCl pH 8.0, 10 mM EDTA). The sample was lysed by adding 200 μL Solution II (200 mM NaOH, 1% SDS) and inverting 10 times before adding 150 μL Solution III (3M potassium acetate pH 5.5) and inverting a further 5 times. Samples were centrifuged at 13,000 rpm for 5 minutes and the supernatant was extracted in 400 μL phenol:cholorform:isoamyl by vortexing for 2 minutes and centrifuging at 13,000 rpm for 5 minutes. The upper phase was transferred to a new tube before 600 μL isopropanol was added, and the samples incubated on ice for 10 minutes to precipitate. Samples were once again centrifuged at 13,000 rpm for 5 minutes before the DNA pellet was washed in 70% ethanol. The pellet was then air dried for 5 minutes before being resuspended in 30 μL dH_2O.

2.3.2 RNA Extraction

For all RNA work equipment such as microcentrifuge tubes and pipette tips were double autoclaved before use. RNase-free water was prepared by treating with diethyl pyrocarbonate (DEPC) (1% v/v) at 37°C overnight and autoclaving twice.

Samples were grown on top of cellophane discs on plates and harvested using a sterile spatula to transfer the biomass into a separate tube. This was then flash frozen in liquid nitrogen and stored at -80°C until extraction. Pellets were resuspended in 1 mL RLT Buffer (QIAGEN) which was supplnemented with β-mercaptoethanol (1% v/v). Samples were vortexed for 1 minute and the applied to a QIA-Shredder column (QIAGEN) and centrifuged for 2 minutes. Flow through was collected and mixed with 700 μL acidified phenol-chloroform for 1 minute. Samples were then incubated at room temperature for 3 minutes before centrifuging at 13,000 rpm for 20 minutes. The upper phase was collected and mixed with 0.5 volumes of 96% ethanol (made with DEPC treated water). Samples were then applied to a RNeasy Mini spin column (QIAGEN) and purified following manufacturer's instructions.

2.3.3 DNA and RNA Quantification

Extracted DNA or RNA was analysed using a Nanodrop 2000 UV-Vis Spectrophotometer. Depending on the concentration of the sample, further quantification was carried out using a Qubit Fluorimeter 2.0 using either the high-sensitivity or the broad-range kit as appropriate.

2.3.4 Primers

All primers were designed manually in A Plasmid Editor (ApE) and ordered from Integrated DNA Technologies (IDT) at 25 nmole with standard desalting and resuspended in dH_2O to a final concentration of 10 μM unless otherwise stated. A table of all primers used are listed in the appendix of this thesis (**Table 7.2**).

2.3.5 Polymerase Chain Reaction (PCR)

Depending on downstream application, two different DNA polymerases were used in this work. PCRBIO Taq DNA Polymerase (PCR Biosystems) was used for diagnostic and colony PCR and used single colonies as template DNA. For *E. coli* samples, colonies were picked and added directly into PCR mix (**Table 2.3**) whereas for *Streptomyces* spp. samples, colonies were first resuspended in 50% DMSO and incubated at 55°C with shaking at 180 rpm before 3 μL of this sample was added to the PCR mix (**Table 2.4**). Q5 High Fidelity Polymerase (NEB) was used for DNA fragment amplification for use in cloning and used either plasmid or genomic DNA as the template. PCRs were conducted using a C1000 Touch Thermal Cycler (BIORAD) with conditions (**Table 2.5**) being calculated based on the size of the target fragment and annealing temperature of primer pairs (NEB online Calculator) and polymerase manufacturer's instructions.

Table 2.3 Composition of individual PCR Mix used for PCRBIO Taq reactions

Reagent	Volume	Final Concentration
2xPCRBIO Taq	10 μL	1x
DMSO	1 μL	5%
Forward primer	0.5 μL	125 nM
Reverse primer	0.5 μL	125 nM
Template DNA	0.5 μL	variable
dH_2O	7.5 μL	

Table 2.4 Composition of individual PCR Mix used for Q5 Polymerase reactions

Reagent	Volume	Final Concentration
Q5 high fidelity polymerase	0.1 μL	
5x Q5 reaction buffer	2 μL	1x
5x Q5 high GC enhancer	2 μL	0.5x
dNTPs	0.25 μL	200 μM
Forward primer	0.5 μL	100 μM
Reverse primer	0.5 μL	100 μM
Template DNA	0.1 μL	variable
dH_2O	5.5 μL	

Table 2.5 Conditions of PCRs conducted during this work

Process	Temperature	Time	Number of Cycles
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	30 seconds	30-40
Annealing	Variable	30 seconds	
Extension	72 °C	30 seconds/ kb	
Final extension	72 °C	10 minutes	1
Hold	4 °C	Final hold	1

2.3.6 Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

For RT-PCR, extracted RNA was converted to cDNA using the LunaScript RT SuperMix (NEB) following manufacturer's instructions. Resultant cDNA was used as a template for PCR following protocols listed above.

2.3.7 Agarose Gel Electrophoresis and Recovery

Following PCR, amplified DNA was separated by electrophoresis on 1% (w/v) agarose gels made with TAE Buffer (45 mM Tris-acetate, 1 mM EDTA, pH 8.0) and ethidium bromide at a final concentration of 2 μg/mL. DNA samples were mixed with 0.16 volumes of Gel Loading Dye Purple 6x (NEB) and loaded onto the gel alongside a 1kB Plus ladder (NEB). Gels were submerged and run in 1 x TAE Buffer at 120 V for 1 hour and visualised using the ethidium bromide setting on a SYNGENE Imager.

Where PCR products were required for downstream application, bands were extracted from the agarose gels using the QIAQuick Gel Extraction Kit (QIAGEN) following the manufacturer's instructions.

2.3.8 Restriction Digest

Restriction enzymes (NEB) were used to digest PCR fragments and plasmid DNA using CutSmart Buffer (NEB). A total of 1 μg DNA was digested using 1 unit of the appropriate restriction enzyme(s) and placed at 37 °C for 30 minutes, after which the enzymes were heat inactivated at 65 °C for 10 minutes. Where the same enzyme was being used to target two separate cut sites within a plasmid, 2 μL shrimp alkaline phosphatase was added to dephosphorylate the digested DNA and prevent re-ligation. Digests were subsequently analysed by gel electrophoreses and desired bands excised and extracted as detailed above.

2.3.9 Gibson Assembly

Digested DNA fragments were assembled into a corresponding digested vector backbone using overlaps ranging between 18 and 24 nucleotides. Gel extracted DNA fragments were

incubated in a molar ratio of 1:3 (plasmid:insert) to a volume of 5 μL, alongside 5 μL Gibson Assembly master mix at 50 °C for an hour. Plasmid and fragment volumes were calculated using NEBioCalculator. The resulting mix was used to transform the newly generated plasmids into an appropriate *E. coli* strain.

2.3.10 Golden Gate Assembly

Small inserts such as synthetic protospacers were assembled into a relevant vector using golden gate assembly with the reaction mix (**Table 2.6**) being run in a C1000 Touch Thermocycler (**Table 2.7**). The resulting mix was transformed into an appropriate *E. coli* strain and screened for successful insertion using blue/ white X-Gal plates and confirmed by sequencing.

Table 2.6 Composition of individual Golden Gate Assembly reactions

Reagent	Volume/ Concentration
T4 Ligase Buffer	2 μL
T4 Ligase	1 μL
BbsI	1 μL
Vector backbone	100 ng
Insert	0.3 μL
dH_2O	Up to 20 μL

2.3.11 Ligation

Digested DNA fragments were assembled into a corresponding digested vector using overlaps ranging between 18 and 24 nucleotides. Gel extracted DNA fragments were incubated in a molar ration of 1:3 (plasmid:insert) alongside 2 μL T4 DNA Ligase to a volume of 20 μL. Plasmid and fragment volumes were calculated using NEBioCalculator. The mixture was incubated at room temperature for 10 minutes before the enzyme was heat inactivated at 65 °C for 10 minutes. The resulting mix was used to transform the newly generated plasmids into an appropriate *E. coli* strain.

2.3.12 Plasmid Preparation

Plasmid DNA was prepared using QIAprep Spin Miniprep Kit (QIAGEN). All steps performed according to manufacturer's instructions using columns and buffers supplied with the kit, except for DNA elution using 30 µL dH_2O rather than the recommended 50 µL, to increase concentration. A table of all plasmids used or generated as part of this work are listed in the appendix of this thesis (**Table 7.3**).

2.3.13 Sequencing

Plasmid constructs were confirmed using Mix2Seq offered by Eurofins Genomics. DNA concentration was measured and diluted appropriately according to manufacturer's instructions before test primers were added alongside DMSO 5% (v/v).

2.3.14 Gene Synthesis

Gene synthesis was performed by GenScript (US).

2.3.15 Preparation and Transformation of Chemically Competent Cells

A single colony of *E. coli* Top 10 or NiCo21 was used to inoculate 10 mL selective LB overnight and subcultured in a further 50 mL selective LB until OD_{600} was ~ 0.4. The subsequent culture was centrifuged at 4,000 rpm for 5 minutes at 4 °C and resuspended in 20 mL ice cold 100 mM $CaCl_2$, this process was repeated to wash the cells. The final pellet was resuspended in 2 mL ice cold 100 mM $CaCl_2$, an either aliquoted and flash frozen in liquid nitrogen to be stored at -80 °C or used immediately for transformation.

Heat shock transformations were performed by adding 2 µg plasmids to a 50 µL aliquot of the host strain and mixing gently before incubating on ice for 30 minutes. Cells were heat shocked at 42 °C for 30 seconds and then returned to ice for a further 2 minutes. To allow the cells to recover, 500 µL LB was added to the samples before they were incubated for 45 minutes at 37 °C, 220 rpm. After this time there were plated onto selective media and incubated overnight.

2.3.16 Preparation and Transformation of Electrocompetent Cells

A single colony of *E. coli* ET12567/pUZ8002, Top10 or NiCo21 was used to inoculate 10 mL selective LB overnight and subcultured in a further 50 mL selective LB until OD_{600} was ~ 0.4. The subsequent culture was centrifuged at 4,000 rpm for 5 minutes at 4 °C and resuspended in 20 mL ice cold 10% glycerol, this process was repeated to wash the cells. The final pellet was resuspended in 1 mL ice cold 10% glycerol, an either aliquoted and flash frozen in liquid nitrogen to be stored at -80 °C or used immediately for transformation.

Electrocompetent transformations were performed by adding 2 μg plasmid to a 50 μL aliquot of the host strain and electroporated in a MicroPulser Electroporator (BIORAD) at 200 Ω, 25 μF and 2.5 kV. An additional 500 μL LB-NaCl was added to samples which were recovered at 37 °C, 220 rpm for 45 minutes before being plated into LB + antibiotics and incubated at 37 °C overnight.

2.3.17 *In vivo* Plasmid Conjugations

Transformed *E. coli* ET12567/pUZ8002 were used as donors to introduce plasmids to *S. formicae* strains via conjugation. A single transformant colony was used to inoculate 10 mL selective LB overnight and subcultured in a further 10 mL selective LB until OD_{600} ~ 0.6. Samples were centrifuged at 4,000 rpm for 5 minutes and the pellet was washed with 10 mL LB. The final pellet was resuspended in 500 μL LB and added to 500 μL heat shocked *S. formicae* spores (20 μL spore stock in 500 μL 2xYT, incubated at 50 °C for 10 minutes). Samples were centrifuged at 13,000 rpm for 1 minute and the pellet was resuspended in residual liquid. A dilution series (100-10^{-2}) was made and 100 μL of each plated onto SFM + 10 mM $MgCl_2$. Plates were incubated at 30 °C for 16-20 hours, after which they were overlaid with 1 mL dH_2O + 0.5 mg Nal and relevant antibiotics. Samples were incubated for a further 5 days or until single colonies appeared (if only a lawn formed, plates were restreaked for single colonies).

2.4 Gene Editing in *Streptomyces formicae*

2.4.1 Construction of Gene Knockouts using CRISPR/Ca9

Clean deletions were made using the pCRISPomyces-2 vector (Cobb et al., 2015) which includes the Cas9 enzyme, a sgRNA scaffold with a BbsI site into which the protospacer can be assembled and an XbaI site for assembly of the repair template (**Fig 2.1).** Protospacers were designed to be 20 base pairs in length where the last 15 nucleotides plus the NGG sequence was unique in the genome to reduce off target effects. Forward and reverse sequences were ordered as single stranded DNA oligos from IDT and annealed by adding equal volumes of each oligo and heating to 95 °C for 5 minutes followed by gradual reduction of 0.1 °C per second until at 4 °C. Annealed protospacers were assembled into the vector via Golden Gate assembly.

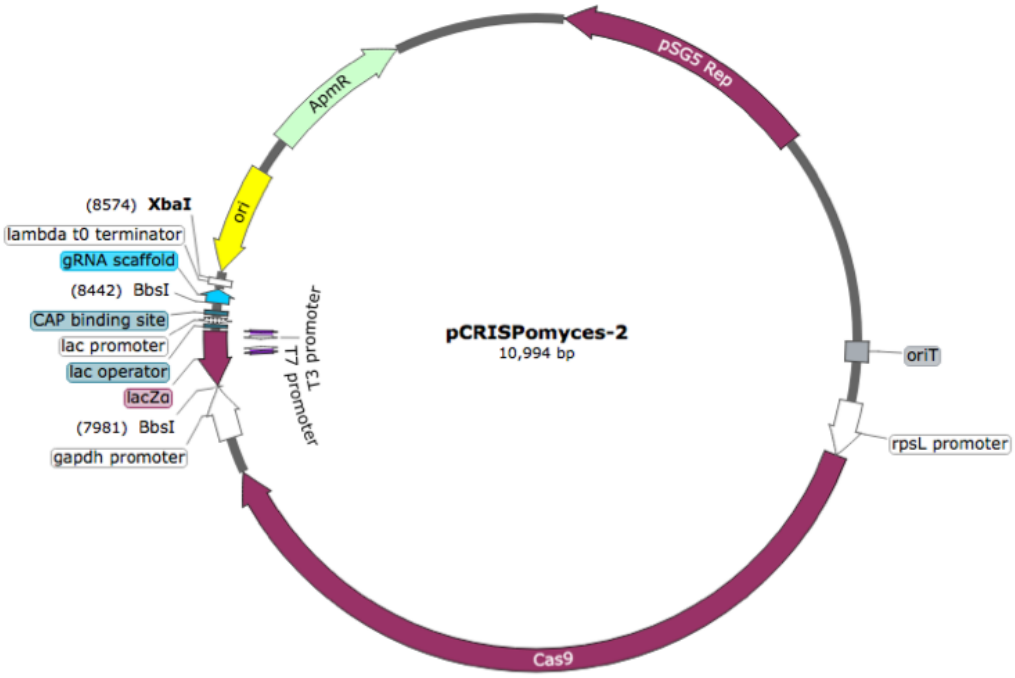


Figure 2.1 The pCRISPomyces-2 vector visualised and annotated in SnapGene. Relevant annotations include the Cas9 enzyme which causes a double stranded break in the target DNA as specified by the protospacer which can be integrated at the BsbI site. The gRNA scaffold is flanked by this BsbI site and an XbaI cut site, the latter of which is used to assemble the repair template. The vector also contains an Apramycin resistance cassette for easy selection of colonies containing the desired vector.

Primers were designed to amplify a 1 kb region either side of the target including a linker that would allow them to anneal to one another during assembly as well as the relevant sticky ends to allow assembly into the XbaI digestion site of pCRISPomyces-2. The repair

templates were PCR amplified while the pCRISPomyces-2 vector was digested with XbaI and all components were confirmed via gel electrophoresis and then recovered.

All fragments were assembled via Gibson Assembly and the resultant plasmid was transformed into *E. coli*, confirmed via PCR and sequencing, and then conjugated into the relevant *Streptomyces* strain. Once the required deletion had taken place, loss of the pCRISPomyces-2 plasmid was encouraged by restreaking colonies onto SFM and incubating at 37 °C (as the plasmid is highly sensitive to temperature) for multiple generations.

To confirm the deletion event has been successful, varying combinations of the original repair templates were used for PCR (**Fig 2.2**) products were compared between suspected successful mutants and the wildtype strain (wildtype PCR products will be longer than successful deletion mutants).

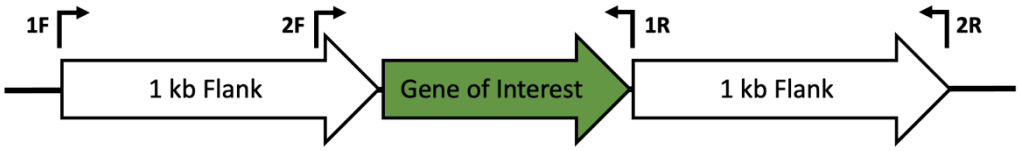


Figure 2.2 A visual representation of the primers used to assess whether the gene of interest had been successfully knocked out using the CRISPR/Cas9 system. Comparing the resultant bands from PCR amplifying the wildtype and the knockout strains should show bands with a size difference equivalent to that of the gene that has been knocked out.

2.4.2 Construction of Single Nucleotide Changes using CRISPR/Ca9

Point mutations were made using the pCRISPomyces-2 vector described above. Protospacers were designed to be 20 base pairs in length where the last 15 nucleotides plus the NGG sequence was unique in the genome to reduce off target effects. Forward and reverse sequences were ordered as single stranded DNA oligos from IDT and annealed by adding equal volumes of each oligo and heating to 95 °C for 5 minutes followed by gradual reduction of 0.1 °C per second until at 4 °C. Annealed protospacers were assembled into the vector via Golden Gate assembly.

Site directed mutagenesis via Gibson Assembly was used during the design of the repair template to introduce single point mutations at multiple sites within the gene of interest. The first base change was the desired point mutation to induce an amino acid change in the gene of interest. The second was a silent mutation to remove the NGG cut site from the

gene which utilised the redundancy in the genetic code to prevent a change in the resultant amino acid but remove the cut site of the Cas9 enzyme. Primers were designed to amplify 3 regions of variable length around the gene of interest including complementary overhangs that would allow them to anneal to one another in a specific order during assembly. The outermost regions included relevant sticky ends to allow the repair template to assemble into the XbaI digestion site of pCRISPomyces-2.

All fragments were assembled via Gibson Assembly and the resultant plasmid was transformed into *E. coli*, confirmed via PCR and sequencing, and then conjugated into the relevant *Streptomyces* strain. Once the required deletion had taken place, loss of the pCRISPomyces-2 plasmid was encouraged by restreaking colonies onto SFM and incubating at 37 °C (as the plasmid is highly sensitive to temperature) for multiple generations.

To confirm the desired point mutation had taken place, intergenic primers were used to PCR amplify the area of interest directly from the *Streptomyces*. Bands of the expected size were excised, extracted and sent to be confirmed via sequencing.

2.4.3 Genetic Complementation

To complement individual gene knockouts, the relevant gene was digested with NdeI and HindIII and ligated to include the gene's own native promoter in pMS82 (Gregory et al., 2003). The resultant plasmid was transformed into *E. coli*, confirmed via PCR and sequencing, and then conjugated into the relevant mutant *Streptomyces* strain which was further confirmed via PCR and stocked for long term storage.

2.4.4 Construction of Overexpression Mutants

To overexpress individual or sets of genes, the relevant fragments were digested and assembled downstream of the *ermE** promoter in pIJ10257 (Bibb et al., 1985). If several genes were being constitutively overexpressed at one, primers were designed to include appropriate overlaps to allow fragments to anneal to one another in a specific order. The resultant plasmid was transformed into *E. coli*, confirmed via PCR and sequencing, and then conjugated into the relevant mutant *Streptomyces* strain which was further confirmed via PCR and stocked for long term storage.

2.5 Analysis of Secondary Metabolites

2.5.1 Antimicrobial Activity Assay on Solid Media

To assess the antibiotic production by a strain in solid culture, agar plates were inoculated with 2 μL the relevant *Streptomyces* spore stock and incubated for 5-15 days. Indicator strains of *E. coli, MRSA, Bacillus subtilis* and *Candida albicans* were grown overnight in 10 mL LB at 30 °C, 220 rpm and were subcultured until OD 600 ~ 0.4. The subcultures were diluted into SNA (4:100 for MRSA and *C. albicans*, 1:10 for *B. subtilis* and *E.coli*) which was used to overlay the plate. Plates were grown at 30 °C overnight and imaged the following day to identify any zones of inhibition.

2.5.2 Chemical Extraction of Secondary Metabolites from Solid Media

All strains were initially grown in 10 mL TSB 2 days, after which 100 μL was plated onto the relevant media and grown statically for 5-15 days. Equal size agar plugs (1 cm^3) were taken and shaken in 1 mL ethyl acetate for 1 hour. After this, 300 μL of the ethyl acetate was transferred to a clean tube and the solvent removed under reduced pressure. Samples were either stored at -20 °C until analysis or used immediately.

2.5.3 Chemical Extraction of Secondary Metabolites from Liquid Media

All strains were initially grown in 10 mL TSB for 2 days, after which 100 μL was used to inoculate a further 10 mL of TSB in a sterile 50 mL falcon tube with sterile bungs and incubated at 30 °C and 250 rpm for 5-15 days. After this, 1 mL aliquots were removed and shaken with 1 mL ethyl acetate for 1 hour and centrifuged at 13,000 rom for 5 minutes. A 300 μL sample of the ethyl acetate fraction was transferred to a clean tube and the solvent removed under reduced pressure. Samples were either stored at -20 °C until analysis or used immediately.

2.5.4 HPLC and Mass Spectrometry Analysis of Secondary Metabolites

Dried extract was resuspended in 200 μL methanol before being analysed by HPLC Agilent 1290. Chromatography was undertaken using the following method: Phenomenex Gemini NX C18 column (150 3 4.6 mm); mobile phase A: water + 0.1% formic acid; mobile phase B: methanol. Elution gradient: 0–2 min, 50% B; 2–16 min, 50–100% B; 16–18 min, 100% B; 18–18.1 min, 100–50% B; 18.1–20 min, 50% B; flow rate 1 mL min1; injection volume 10 mL.

Titres of fasamycin and formicamycins were calculated by Dr Hannah McDonald by comparing the peak areas from this HPLC analysis to standard calibration curves, with account being made for changes in concentration between extractions. Calibration curves used standard solutions of fasamycin E and formicamycin I with UV absorption at 418 nm and 285 nm, respectively.

2.6 Protein Purification and Analysis

Proteins of interest were tagged with a 6xHis epitope and purified from whole cell lysate using NiCo21 as host strain and using Ni-NTA columns (Cytiva) in association with ÄKTA Pure FPLC.

2.6.1 Construction of Overexpression Plasmids

Overexpression plasmids were made using the pET28a (+) or pET29a (+) vectors (Invitrogen) as backbones, both of which contain a wide variety of digestion sites, a multiple cloning site with a 6xHis epitope either up or downstream for C-terminal or N-terminal tagging and an inducible T7 promoter under the control of a lac operon (**Fig 2.4**). Plasmids were generated via Gibson Assembly, using NdeI and HindIII as the digestion sites for inserting the gene of interest, and confirmed via sequencing. Where needed, codon optimised genes were ordered from GenScript and inserted into plasmids as above.

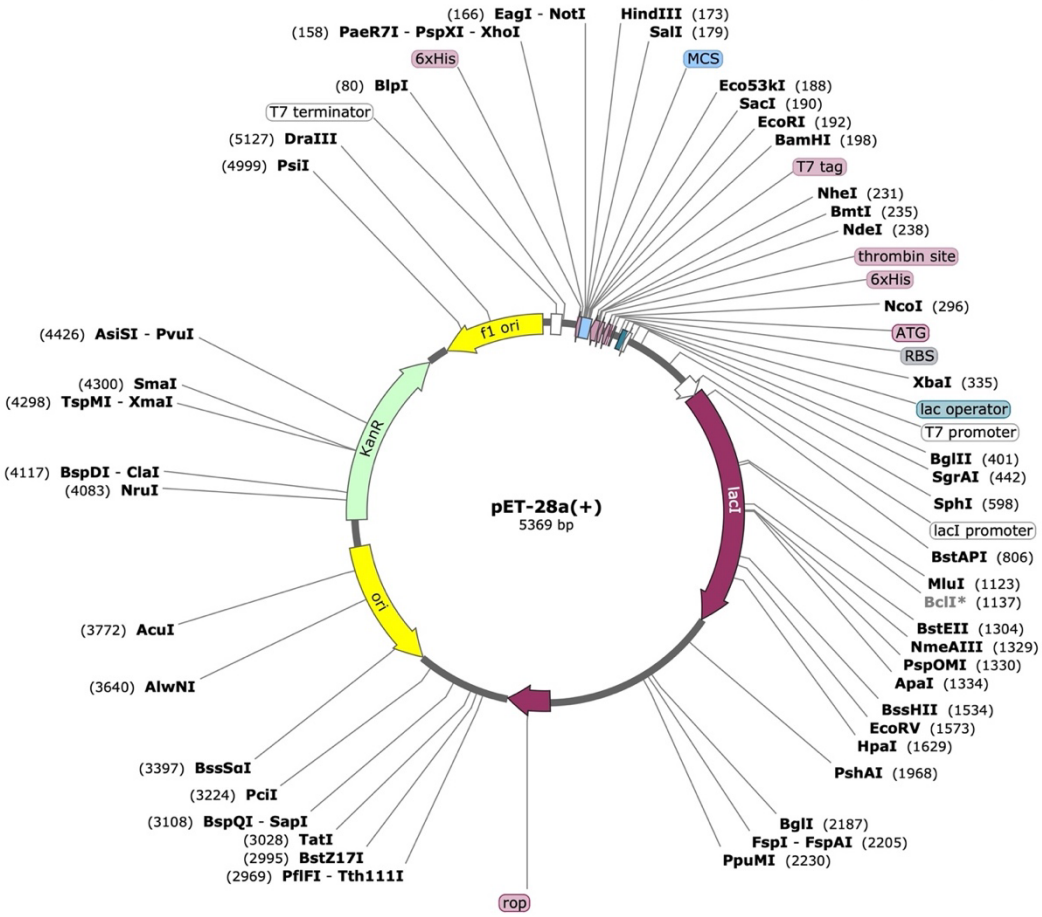


Figure 2.3 The pET28a(+) vector visualised and annotated in SnapGene. Relevant annotations include a multitude of restriction sites, for digest and insertion of the gene of interest, on either side of the Multiple Cloning Site (MCS) which is also flanked by a 6xHis epitope both up and downstream. There is also a lac operator under the control of a T7 promoter allowing the inducible expression of the gene of interest with IPTG. The vector also contains a kanamycin resistance cassette for easy selection of colonies containing the desired vector.

2.6.2 Test Overexpression

To ascertain high yielding conditions for each protein, small scale test expression assays for batch purification were conducted. Plasmids were transformed into NiCo21 and successful transformants were used to inoculate 10 mL selective LB overnight and subcultured in a further 100 mL selective LB until OD 600 ~0.6. Uninduced samples were collected by centrifuging 1 mL culture at 13,000 rpm for 1 min before the pellet was flash frozen in liquid nitrogen to be stored at -80 °C. Overexpression was induced in the remaining culture with the addition of IPTG. Expression conditions investigated routinely included concentration of IPTG (50 µM, 200 µM, 1 mM), temperature (18 °C, 30 °C, 37 °C) and time point (3 hr, 4 hr,

overnight). At each time point, 1 mL culture was collected and centrifuged at 13,000 rpm for 1 min before the pellet was flash frozen in liquid nitrogen to be stored at -80 °C.

2.6.3 Large Scale Harvest

Once high yielding overexpression conditions had been identified, the optimum conditions were routinely used to batch purify the protein of interest in 1 L selective LB in a 2 L conical flask. In the case of ForF (and its variants) and ForG, following fresh transformations of the overexpression plasmid into NiCo21 were used to inoculate 100 mL selective LB overnight and subcultured into a further 1 L selective LB until OG 600 ~0.6. Uninduced samples were collected and stored as detailed above. Overexpression wass induced in the remaining culture with the addition of IPTG (200 μM for ForG, 750 μM for ForF) and cultured at 18 °C, 190 rpm for a further 16 hours. Cells were collected and centrifuged at 4,000 rpm for 15 minutes at 4 °C and pellets were flash frozen in liquid nitrogen to be stored at -80 °C.

2.6.4 Protein Purification with ÄKTA Pure FPLC and His Trap Column

Harvested pellets were thawed on ice and resuspended in Lysis Buffer (20 μM Tris/HCl pH 8.0, 75 μM NaCl, 0.01 % Triton X-100, 1 mg/mL lysozyme) and incubated at room temperature of 30 minutes. The sample was then sonicated (Sonics Vibracell) on ice in 8 rounds of 30 sec on / 1 min off at 20 microns. Lysate was centrifuged at 15,000 rpm for 1 hour at 4 °C and supernatant was transferred to a new tube before being filter sterilised with a 0.22 μm filter and syringe. Where appropriate depending on the plasmid used, the supernatant was applied to 3 g chitin resin (NEB) in a pre-equilibrated manual chromatography column and incubated with rotation for 30 minutes at 4 °C before being collected under gravity. The resultant sample was loaded onto the ÄKTA Pure FPLC using HisTrap High Performance purification columns (Cytiva) with Ni Sepharose of either 1 mL or 5 mL volume and Unicorn 7.0 Software. Samples were passed over the Ni Sepharose column to allow binding of the 6xHis epitope with the attached protein. The column was then washed with Nickel Buffer A (50 mM Tris/HCl pH 8.0, 200 mM NaCl, 5 % glycerol) at increasing concentrations (4%, 6%, 8%) of Nickel Buffer B (50 mM Tris/HCl pH 8.0, 200 mM NaCl, 5 % glycerol, 500 mM Imidazole) to remove most Ni binding contaminants, each wash was collected in 1.9 mL fractions. The final sample was eluted in 100 % Nickel Buffer B and

the final elution was collected in 1 mL fractions. The full protocols used for both 1 mL and 5 mL Ni Sepharose columns in this work are available in the appendix.

Uninduced sample, a small amount of "pre-load" sample, flow through, and fractions from washes and elution were all assessed via SDS-PAGE to ensure the protein of interest had been successfully purified before moving on to the size exclusion step. Fraction plates were routinely stored at 4 °C to minimise degradation of protein.

2.6.5 Size Exclusion Chromatography with ÄKTA Pure FPLC

Successfully eluted fractions identified from the purification step were combined and concentrated if the total volume exceeded 5 mL. The resultant sample was loaded onto the ÄKTA Pure FPLC for size exclusion chromatography with the Hi Load 16/600 Superdex 300 prep grade column (Cytiva) to fractionate the appropriate range with the highest resolution. Flow through was collected in 1 mL fractions which were assessed via SDS-PAGE to identify the peak containing the protein of interest. Fractions from the successful peak were combined and quantified before being flash frozen in liquid nitrogen and stored at -80 °C in appropriate volume aliquots depending on downstream application.

2.6.7 Protein Quantification

Purified protein was analysed using a Quibit assay using the Qubit Fluorimeter 2.0 alongside the relevant Protein BSA test according to manufacturer's instructions. Where needed protein fractions were concentrated using Amicon Ultra Centrifugal filters at an appropriate volume and kDa cut off for the protein of interest.

2.6.8 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis

To confirm the presence of the protein of interest in crude samples, pellets were defrosted on ice and resuspended in 10 μL SDS Buffer (19:1 Laemmli buffer:β-mercaptoethanol) before boiling at 100 °C for 3 minutes. Lysate was centrifuged at 13,000 rpm for 10 minutes, supernatant transferred to a fresh Eppendorf (soluble fraction) and pellet (insoluble fraction) was resuspended in a further 20 μL SDS-Buffer.

For samples that had already been purified following large-scale harvest 10 μL of the fraction of interest was mixed with 10 μL SDS Buffer before boiling at 100 °C for 2 minutes. Lysate was centrifuged at 13,000 rpm for 10 minutes to and supernatant was used as the final fraction of interest.

Precast 12% TEO Tricine gels (AbCam) were used to resolve all denatured protein samples. Fractions were loaded onto gels submerged in RunBlue Buffer (AbCam) and run at 100 V for 90 minutes using PageRule 1kB Plus Protein Ladder as a marker. Protein bands were visualised by staining with IntantBlue Coomassie stain for 1 hour or overnight with gentle agitation and subsequently washed in dH_2O for 3 rounds of 10 minutes.

2.6.9 Western Blot

In order to identify bands containing the desired 6xHis epitope, SDS-PAGE was completed without the final staining steps. Instead, samples were transferred to nitrocellulose Biodyne A membrane in a Trans-Blot (BIORAD). Two layers of blotting paper equal to the size of the gel were soaked in Transfer Buffer (25 mM Tris/HCl pH 8.0, 192 mM Glycine, 0.1 % SDS, 20 % methanol) and placed onto the anode plate of the transfer cell. Nitrocellulose membrane equal to the size of the gel was soaked in 100 % methanol for 1 min before being washed in Transfer Buffer 5 min. This was then placed on top of the blotting paper followed by the SDS-PAG and another two layers of Transfer Buffer-soaked blotting paper. This was then transferred for 30 minutes at 2 A before the nitrocellulose membrane was incubated in Blocking Solution (5 % fat free skimmed milk powder, 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 1 % Tween) for 1 hour with gentle agitation. Anti-His antibody (QIAGEN) conjugated to horseradish peroxidase was diluted 1:20,000 in TBST (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 1 % Tween), the blocked membrane was incubated in 20 mL of this solution for 1 hour at room temperate with gentle agitation. The membrane was washed in 1xTBST for 3 rounds of 10 minutes before being developed for 1 minute in a 50:50 mixture of Solution A (10 mL 100 mM Tris/HCl pH 8.5, 100 μL luminol, 45 μL coumaric acid) and Solution B (10 mL 100 mM Tris/HCl pH8.5, 6 μL 30 % hydrogen peroxide). Fluorescence was detected with the ECL setting on a SYNGENE Imager.

2.6.10 Protein Mass Spectrometry

When required, protein samples were confirmed via tryptic digest and MALDI-TOF mass spectrometry. The sample of interest was run on an SDS-PAGE gel using minimal Coomassie staining to visualise bands. Gel slices of the relevant protein were taken and prepared by de-staining in 30% ethanol for 30 minutes at 65 °C until clear. Slices were then washed with 50 mM TEAB in 50 % acetonitrile before being incubated with 10 mM DTT for 30 minutes at 55 °C. The DTT solution was removed before adding 10 mM iodoacetamide in 10 mM TEAB and incubating in the dark on a vortexer for 30 minutes. Samples were then washed once with 50 mM TEAB in 50 % acetonitrile and again in just 50 mM TEAM while being vortexed. The buffer was removed and the gel slices measures before being sliced into 1x1 mm pieced and placed in a low bind tube. Sliced gels were then washed with 50 mM TEAB in 50 % acetonitrile for 20 minutes with vortexing and again with 100 % acetonitrile. Any solvent was removed, and tubes were thoroughly dried with a speedvac for 30 minutes. Subsequent tryptic digest and HRMS analysis was performed by Carlo de-Oliveira Martins from the JIC proteomics facility. Proteins were digested using trypsin in a 1:20 trypsin: protein ratio for 8 hours at room temperature in 50 mM ammonium bicarbonate MeCN (5% v/v) at pH 7.5. LC-MS/MS was then performed using an Orbitrap Eclipse tribrid mass spectrometer (Thermo Fisher Scientific) fitted with a nanoflow HPLC system (Dionex Ultimat3000) .

2.7 X-Ray Crystallography

Crystal trials were set up using an initial protein concentration of 10-20 mg/ mL with an Orxy8 (Douglas Instruments) liquid handling robot to dispense two drops per well, in a 96-well MRC 2-drop plate 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 and checked every few days to identify any potentially suitable crystals for harvesting. A combination of commercially available and bespoke screens were used.

2.8 β-Glucuronidase Assays

Strains were grown on SFM with a cellophane disc for 4 days before mycelium were harvested with a sterile metal spatula and resuspended in 1 mL Dilution Buffer (50 mM

Phosphate Buffer, 0.1 % (v/v) Triton X-100, 5 mM DTT, 10 mg/mL EDTA-free protease inhibitor). The concentration of protein in the lysate was quantified using a Nanodrop before 100 μL each sample was added to a 96 well plate to be freeze-thawed at -80 °C and 37 °C. The assay was initiated with the addition of 0.3 mg 4-nitrophenyl β-D-glucopyranoside (PNPG). Samples were loaded into a plate reader at 37 °C for optimum enzyme activity and quantified at 415 nm and 550 nm every 5 minutes for 40 minutes, inclusive. Softmax® Pro7 used to extract raw data which was used calculate Miller units.

2.9 Co-Immunoprecipitation

Strains were grown on SFM with a cellophane disc for 4 days before discs were removed and submerged in 10 mL 1% (v/v) formaldehyde solution made up in PBS for 20 minutes at room temperature to cross-link proteins to DNA. The mycelium was harvested by scraping with a metal spatula and centrifuged at 4,000 rpm for 10 minutes before the supernatant was removed. The pellet was washed in 10 mL 0.5 M glycine for 5 minutes and subsequently 25 mL ice cold PBS, with centrifuging and discarding of the supernatant between each step. After the final spin, pellets were flash frozen in liquid nitrogen and stored at -80 °C.

Pellets were resuspended in 750 μL Lysis Buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mg/mL lysozyme) and incubated at room temperature for 30 minutes. A further 750 μL IP Buffer (100 mM Tris-HCl pH 8.0, 500 mM NaCl, 1% Triton X-100) was added and samples mixed by pipetting. Samples were sonicated at 50 Hz in 8 rounds of 20 seconds on, 60 seconds off. This crude lysate was centrifuged at 13,000 rpm for 15 min at 4 °C before lysate was taken and moved to a fresh tube. Alongside this 50 μL per sample of Anti-FLAGM2 beads were washed twice in 3 mL 0.5 IP Buffer. Finally, 40 μL of beads were added to the lysate before being incubated overnight at 4 °C on a vertical rotor.

The following day lysate was removed, and the beads were washed 4 times with 500 μL 0.5 IP Buffer with centrifuging at 2,000 rpm maximum for 10 minutes at 4 °C. The final wash buffer was removed, and the sample was resuspended in 30 μL SDS Loading Buffer. Samples were boiled at 100 °C for 5 minutes and pelleted at 14,000 g for 1 minute before being loaded onto an SDS-PAGE gel (10% throughout). Samples were run a few millimetres into the gel before being gently but thoroughly washed with tap water. Gel slices were excised and measured before being stored at -20 °C.

Gel slices were destained with 1 mL 30 % ethanol for 30 minutes at 65 °C, this was repeated until clear. Slices were washed with 1 mL 50 mM TEAB, 50 % acetonitrile with vortexing at room temperature for 20 minutes before they were incubated in 1 mL 10 mM DTT n 50 mM TEAB for 30 minutes at 55 °C. The DTT solution was removed and 1 mL iodoacetamide added and samples were vortexed for 30 minutes at room temperature in the dark. This step was then repeated with 1 mL 50 mM TEAC. Buffer was removed and the gel slice cut in to 1x1 mm pieces using a sterile scalpel to be moved into a fresh LoBind microcentrifuge tube. Samples were washed once with 50 mM TEAB, 50 % acetonitrile and twice with 100 % acetonitrile, each wash for 20 minutes with vortexing at room temperature. A hole was pierced in the lid and samples were dried in a Genevac miVac centrifugal concentrator before being passed for processing, mass spectrometry and data processing by Dr Carlo de Oliveira Martins at the John Innes Centre.

2.10 Chromatin Immunoprecipitation (ChIP) Sequencing

Strains were grown on SFM with a cellophane disc for 4 days before discs were removed and submerged in 10 mL 1% (v/v) formaldehyde solution made up in PBS for 20 minutes at room temperature to cross-link proteins to DNA. The mycelium was harvested by scraping with a metal spatula and centrifuged at 4,000 rpm for 10 minutes before the supernatant was removed. The pellet was washed in 10 mL 0.5 M glycine for 5 minutes and subsequently 25 mL ice cold PBS, with centrifuging and discarding of the supernatant between each step. After the final spin, pellets were flash frozen in liquid nitrogen and stored at -80 °C.

Pellets were resuspended in 2 mL Lysis Buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mg/mL lysozyme) and incubated at 37 °C for 30 minutes. To fragment DNA, 1 mL IP Buffer (100 mM Tris-HCl pH 8.0, 500 mM NaCl, 1% v/v Triton X-100) was added and samples mixed by pipetting. Samples were sonicated at 50 Hz in 8 rounds of 20 seconds on, 60 seconds off. Crude lysate was removed and 25 μL mixed with 77 μL TE Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and extracted with 200 μL chloroform. A further 25 μL of this extract was removed and 2 μL RNase A (1 mg/mL) was added before the sample was incubated at 37 °C for 30 minutes and run on a 1 % agarose gel.

Once DNA fragments were confirmed to be at the desired range for sequencing, the remaining crude lysate was centrifuged at 4,000 rpm for 15 minutes at 4 °C. AntiFLAG M2

beads were prepared by washing 500 μL beads in 2.5 mL 0.5 IP Buffer. For each sample, 40 μL of prepared beads were incubated with the lysate on a vertical rotor overnight at 4 °C.

The following day lysate was removed, and the beads were washed 4 times with 500 μL 0.5 IP Buffer with centrifuging at 2,000 rpm maximum for 10 minutes at 4 °C. The bound DNA was eluted with 100 μL Elution Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) and samples were incubated at 65 °C overnight. A further 50 μL Elution Buffer was added to the samples which were incubated for 5 minutes at 65 °C before the 150 μL eluate was removed. Samples were purified by adding 2 μL proteinase K and incubating at 55 °C for 90 minutes. DNA was extracted in 150 μL phenol-chloroform and purified on a QIAquick column (QIAGEN). DNA was eluted in 50 μL EB Buffer (10 mM Tris-HCl pH 8.5) and quantified before being sent to GeneWiz for Illumina HiSeq platform sequencing.

Sequencing data were received as FASTQ files and processed by Dr Govind Chandra at the JIC. Raw reads were aligned to the reference genome and extracted co-ordinates were listed in .bed files which were visualised in Integrated Genome Browser to show enrichment peaks. A cut off of 2000 reads was sufficient to remove background noise and only visualise significant peaks indicating protein binding sites.

2.11 Surface Plasmon Resonance

2.11.1 ReDCaT Screening of Protein Binding Sites

Once the general DNA binding region of a protein of interest was identified by capable RNA-Seq and ChIP-Seq (performed by Dr Rebecca Devine), the sequencing of these binding regions were divided into a series of overlapping regions using the Promoter Oligo Overhang Programme (POOP). Each fragment was 40 nucleotides in length with 15 nucleotides of overlap between each. Reverse compliments of each DNA fragment were ordered, with a biotinylated linker at the 3' end which would allow the DNA to anneal to the chip. Forward and reverse sequences were annealed by adding 45 μL forward and 55 μL reverse and heating to 95 °C for 10 minutes followed by gradual reduction of 0.1 °C per second until reaching 4 °C. The annealed fragment was diluted to a final 1 μM stock with SPR Running Buffer (150 mM NaCl, 3 mM EDTA, 0.05% surfactant P20, 10 mM HEPES pH 7.4).

All experiments were run using a single Sensor Chip SA (GE Healthcare) on a Biacore 8K SPR system (Cytiva) with flow cells 1 and 2 being used as the reference and test flow cells, respectively. A biotinylated single stranded ReDCaT linker (100 nM) was bound to the chip by passing over at 5 μL/ min. Test DNA fragments were then injected over the flow cell at 10 μL/ min for 60 seconds before SPR Running Buffer was passed over the flow cell at the same rate for 120 seconds. The test protein was injected at the required concentration (routinely tested at 10 nM, 50 nM and 100 nM) at 50 μL/ min for 60 seconds before SPR Running Buffer was passed over the flow cell at the same rate for 360 seconds. The chip was regenerated with 1 M NaCl, 50 mM NaOH passed over the flow cell at 10 μL/ min which left only the ReDCaT linker bound to the chip.

Once a fragment had been identified as a candidate for protein binging, the exact binding site was further educated by foot printing. DNA sequences were ordered as described above but with successive 2 nucleotide truncation from both the 3' and 5' end to identify the point at which binding ceased once the actual binding site had been lost.

All sensograms were analysed using Biacore T200 BiaEvaluation software (GE Healthcare).

2.11.2 Amine Coupling

Amine coupling experiments were performed using a single Sensor Chip CM5 (Cytiva) on a Biacore 8K SPR system (Cytiva) using between 1 and 8 of the separate channels with 2 flow cells. This allowed flow cell 1 to be used as the reference while flow cell 2 was used as the test. A type 2 amine coupling kit (Cytiva) was used following the manufacturer's instructions. Protein samples were diluted to 6 μM in 10 mM acetate pH 4.0. The surface of the chip was activated with a 420 second injection of 50% N-hydoxysuccinimide (NHS) 50% 1-ethyl-3(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) at a flow rate of 10 μL/ min. Protein was then injected over all channels with a 420 second injection at 10 μL/min. Once covalent immobilisation was secured, the remaining chip surface was blocked with a 420 second injection of 1 M ethanolamine-HCl pH 8.5 at 10 μL/min. The binding of compounds could then be assessed.

3. Identifying and Characterising Cluster-Situated Two-Component Systems in *Streptomyces formicae*

The role played by any TCS in an organism can be hugely varied, and characterisation of the system to understand its regulatory mechanism and impact on the wider cell can be complex. The regulatory networks they can be involved in are vast and diverse, often relying on multiple other components for a response to be elicited. A combination of large, dimeric, membrane-bound sensor kinase proteins and smaller, dimeric, phosphorylation-dependant response regulator proteins also mean that both *in vitro* and *in vivo* work can be potentially difficult. Despite this, the current understanding of these regulators has proved their role to be significant in biological functions such as growth and development and natural product biosynthesis. Notably, TCSs have been shown to be responsible for the activation of natural product biosynthetic pathways, either directly or as part of a regulatory cascade (McLean et al., 2019). One such example is the focus of this thesis, ForGF, which is known to be the main activator of formicamycin biosynthesis. Interestingly, this TCS is cluster-situated and known to exert its regulatory function within the cluster in which it is found. This led to investigations into other cluster-situated regulators within *Streptomyces* species, and the potential of using molecular microbiology techniques to unlock cryptic BGCs through the TCSs. This chapter describes the bioinformatic analysis of cluster situated TCSs in *Streptomyces*, preliminary microbiological investigations into the exploitation of these regulators within *S. formicae* and discusses the potential and practicality of applying this knowledge to other *Streptomyces* strains and the wider bacterial community.

3.1 Identifying Two-Component Systems in *Streptomyces formicae*

A number of TCSs have been linked to the biosynthetic pathways of natural products, either through the direct activation or repression of the pathway, or by forming some part of the regulatory cascade. Manipulating the expression levels of these TCSs in turn has been shown to have a direct effect on the production of the secondary metabolites that they regulate. On this basis, it was hypothesised that cluster situated TCSs may form a good target for inducing the overproduction of compounds that are not normally expressed under laboratory conditions. Using a combination of antiSMASH 6.0 and Predicted Prokaryotic Regulatory Proteins (P2RP) softwares, the TCSs within the *S. formicae* genome were analysed (**Table 3.1**). This identified 68 TCSs across the genome including all 15 of the known highly conserved systems such as MtrAB, PhoRP and CutRS, defined using reciprocal BLASTn/p. Upon investigation of the location of these TCSs, it was identified that seven were situated within predicted BGCs. In addition, following manual inspection of the BGCs and their predicted proteins, a further four orphan RRs were found to be cluster situated: KY5_2304 (NarL family) within cluster 17, KY5_2385 within cluster 18, KY5_3254 within cluster 21 and KY5_8031 within cluster 42. All of these response regulators are part of the NarL family and contain typical LuxR helix-turn-helix domains. A final unpaired HK was also identified in the form of KY5_0087 within cluster 2 (**Fig 3.1**).

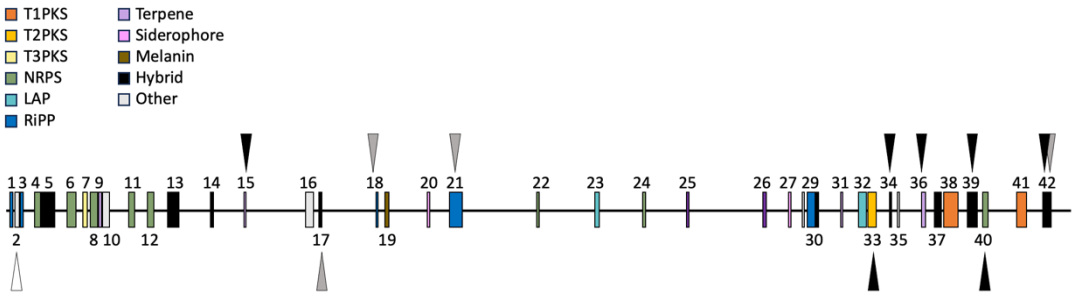


Figure 3.1 Map of the predicted biosynthetic gene clusters in *Streptomyces formicae*. Clusters are colour coded based on predicted products. Black arrows indicate the location of a cluster situated two component system, grey arrows indicate the location of a cluster situated orphan response regulator and white arrows indicate the location of a cluster situated unpaired sensor kinase. Figure based on data from antiSMASH 6.0 and P2RP analysis of the *S. formicae* genome with additional manual inspection.

For the purposes of this investigation, only the TCSs that were paired, and cluster situated were explored. The "pairing" of these TCSs was determined by the proximity of HK and RR genes to one another. The candidate TCSs were provisionally labelled based on the cluster number in which they were situated but will also be referred to by their gene names when necessary to distinguish the two components. Each of the seven cluster-situated TCSs were assessed to determine any homology with other known TCSs to determine if there were any known activating signals, mechanisms or regulatory modes of action that were likely to be relevant for consideration. The BGCs were also analysed for the likelihood of predicted products and any other known regulators that may be acting upon the cluster either at a cluster or global level. Following on from this analysis, each TCS was overexpressed using the high level constitutive *ermE** promoter in a both a wildtype and a Δ*forGF* (unable to produce formicamycin and has no activity against MRSA) background. From here each strain was assessed for bioactivity against a range of indicator strains on different media types and crude extracts of active samples were assessed by HPLC.

Table 3.1 Two-component systems in *Streptomyces formicae* as predicted with P2RP software. Annotations are a combination of HMMTOP and Pfam database information. Rows highlighted blue are the 15 highly conserved systems among *Streptomyces* bacteria and those highlighted green are cluster situated.

Gene Number (Name)	Conserved	Component - Type	Annotation	Details or Function
KY5 0188 KY5 0189	No	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 0412 KY5 0413	No	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1HATPase c 1 RR, 1 HTH LuxR	
KY5 0642 KY5 0643	No	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 4 TM, 1 HisKA 3, 1HATPase c	Immediately outside boundary of BGC 8
KY5 0876 KY5 0877	No	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 1 TM, 1 HAMP, 1 HisKA 3, 1 HATPase c	
KY5 1516 KY5 1517	No	HK – Classic RR – OmpR	4 TM, 1 HisKA, 1 HATPase 1 RR, 1 Trans reg c	Cluster situated within BGC 15 Predicted osmosensitive K+ histidine kinase
KY5 1705 KY5 1706	Yes	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 0 TM, 1 GAF,1 HisKA 3, 1 HATPase c	Upregulated in liquid culture in *S. coelicolor* (Yagüe et al., 2014)
KY5 2103 (MacS) KT5 2014 (MacR)	Yes	HK – Classic RR – NarL	6 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	Activates actinorhodin production and repressed aerial hyphae formation in *S. coelicolor* (M. Liu et al., 2021)
KY5 2127 KY5 2128	No	HK – Classic RR – OmpR	3 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	

KY5 2149 KY5 2150	Yes	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 4 TM, 1 HisKA 3, 1 HATPase c	Not responsible for antibiotic production in *S. coelicolor* (Yepes et al., 2011)
KY5 2210 KY5 2211	No	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 2369 KY5 2340	No	HK – Classic RR – NarL	4 TM, 1 PTS EIIC, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 2639 KY5 2640	No	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 4 TM, 1 HisKA 3, 1 HATPase c	
KY5 2675 KY5 2676	No	HK – Hybrid RR – CheY	4 TM, 7 HAMP, 1 GAF, 1 HisKA, 1 HATPase c, 1 RR 1 CheY	
KY5 2730 KY5 2732 KY5 2733	No	HK – Classic HK – Classic RR – OmpR	4 TM, 1 HisKA 3, 1 HATPase c 0 TM, 1 PAS 4,1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	
KY5 2829 KY5 2830	No	HK – Classic RR – OmpR	3 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	
KY5 2949 KY5 2950	No	HK – Classic RR – OmpR	3 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	
KY5 2978 KY5 2979	No	HK – Classic RR – NarL	3 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	

KY5 2981 KY5 2982	No	RR – LuxR HK – Classic	1 RR, 1 HTH LuxR 4 TM, 1 HisKA 3, 1 HATPase c	
KY5 3156 KY5 3157	No	RR – NarL HK – Classic	1 RR, 1 RR GuxR 2 TM, 1 HisKA 3, 1 HATPase c	
KY5 3176 (MtrB) KY5 3177 (MtrA)	Yes	HK – Classic RR – OmpR	2 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	Variety of roles in life cycle and antibiotic production (Som et al., 2017)
KY5 3224 (DraK) KY5 3225 (DraR)	Yes	HK – Classic RR – OmpR	2 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	Regulates a range of secondary metabolites in under pH stress in *S. coelicolor* (Z. Yu et al., 2012)
KY5 3568 KY5 3569	No	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 2 TM, 1 HAMP, 1 HisKA, 1 HATPase c	
KY5 3736 (PhoR) KY5 3737 (PhoP)	Yes	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 1 TM, 1 HisKA, 1 HATPase c	Phosphate limitation (Sola-Landa et al., 2003)
KY5 3797 (CssR) KY5 3798 (CssS)	Yes	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 2 TM, 1 HAMP, 1 HisKA, 1 HATPase c	Misfolded protein stress (Gullón et al., 2012)
KY5 3846 KY5 3847	No	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 4261 KY5 4262	No	HK – Classic RR – NarL	6 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	

KY5 4397 KY5 4398	No	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 4408 KY5 4409	No	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 4 TM, 1 HisKA 3, 1 HATPase c	
KY5 4411 KY5 4412	No	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 1 TM, 1 HAMP, 1 HisKA, 1 HATPase c	
KY5 4420 KY5 4421	No	HK – Classic RR – OmpR	1 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	
KY5 4470 (TunS) KY5 4471 (TunR)	Yes	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	Tunicamycin resistance (Wyszynski et al., 2012)
KY5 4500 (CseC) KY5 4501 (CseB)	No	HK – Classic RR – OmpR	2 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	Controls expression of *sigE* and the cell envelope stress response
KY5 4505 KY5 4506	No	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 4745 (AbrC3) KY5 4747 (AbrC2) KY5 4748 (AbrC1)	Yes	RR – NarL HK – Classic HK – Classic	1 RR, 1 HTH LuxR 4 TM, 1 HisKA 3, 1 HATPase c 4 TM, 1 HisKA 3, 1 HATPase c	Antibiotic regulation (Rodríguez et al., 2015)
KY5 4881 (EsrR) KT5 4882 (EsrS)	Yes	RR – NarL HK – Classic	1 RR, 1 HTH LurR 1 HATPase c	Envelope stress response (Kleine et al., 2017)

KY5 4909 KY5 4910	No	HK – Classic RR – NarL	5 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 5035 (AfsQ2) KY5 5036 (AfsQ1)	Yes	HK – Classic RR – OmpR	2 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	Nitrogen limitation (Shu et al., 2009)
KY5 5393 KY5 5394	No	HK – Classic RR – NarL	11 TM, 1 HisKA 3, 1 HATPase c, 1 MASE1 1 RR, 1 HTH LuxR	
KY5 5414 KY5 5415	Yes	HK – Classic RR – OmpR	1 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	
KY5 5495 KY5 5496	No	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 6 TM, 1 HisKA 3, 1 HATPase c	
KY5 5520 KY5 5521	No	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 1 TM, 1 HAMP, 1 HisKA, 1 HATPase c	
KY5 5540 KY5 5541	No	HK – Classic RR – OmpR	2 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	
KY5 5559 KY5 5560	No	RR – IclR HK – Classic	1 RR, 1 family HTH IclR 2 TM, 1 PAS, 1 HATPase, Probable incomplete	
KY5 5592 KY5 5593	No	HK – Classic RR – NarL	5 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	

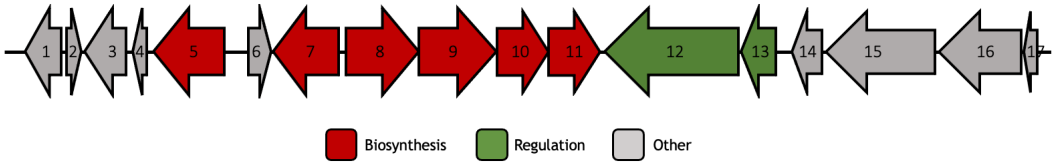
KY5 5876 (OsaA/C) KY5 5877 (OsaB)	Yes	HK – Hybrid RR – unclassified	1 TM, 12 HAMP, 1 GAF, 1 HisKA, 1 HATPase c, 1 RR 1 RR	Osmotic stress (Bishop et al., 2004)
KY5 5911 (GluR) KY5 5912 (GluK)	Yes	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 1 TM, 1 HAMP, 1 HisKA, 1 HATPase c	Glutamate uptake (L. Li et al., 2017)
KY5 5958 KY5 5959	No	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 5962 KY5 5963	No	RR – unclassified HK – Classic	1 RR, 1 HTH 11 2 TM, 1 HATPase c, Probable incomplete	
KY5 5991 (CutR) KY5 5992 (CutS)	Yes	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 1 TM, 1 HAMP, 1 HisKA, 1 HATPase c	Misfolded protein stress (McLean et al., 2023)
KY5 6003 (KdpD) KY5 6004 (KdpE)	No	HK – Classic RR – OmpR	4 TM, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	Potassium transport (Heermann et al., 2009)
KY5 6045 KY5 6046	No	HK – Classic RR – OmpR	3 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	
KY5 6320 KY5 6321	No	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 6 TM, 1 HisKA 3, 1 HATPase c	
KY5 6432 KY5 6433	No	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 0 TM, 1 HAMP, 1 HisKA, 1 HATPase c	

KY5 6501 KY5 6502	No	HK – Classic RR – NarL	5 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 6553 KY5 6554	No	RR – NarL HK - Classic	1 RR, 1 HTH LuxR 5 TM, 1 HisKA 3, 1 HATPase c	
KY5 6663 (ForG) KY5 6664 (ForF)	No	HK – Classic RR – NarL	0 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	Cluster situated within BGC 33, activates formicamycin biosynthesis
KY5 6680 KY5 6681	No	RR – LuxR HK – Classic	1 RR, 1 HTH LuxR 6 TM, 1 HiaKA 3, 1 HATPase c	
KY5 6759 KY5 6760	No	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	Cluster situated within BGC 34
KY5 7229 KY5 7230	No	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 0 TM, 2 GAF, 1 HisKA 3, 1 HATPase c	
KY5 7231 KY5 7232	No	HK – Classic RR – OmpR	0 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	
KY5 7292 KY5 7294	No	RR – NarL HK – Classic	0 TM, 1 RR, 1 HTH LuxR 6 TM, 1 HisKA 3, 1 HATPase c	Cluster situated within BGC 37
KY5 7446 KY5 7447	No	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 2 TM, 1 HAMP, 1 HisKA, 1 HATPase c	Cluster situated within BGC 39

KY5 7481 KY5 7482	No	RR – OmpR HK – Classic	1 RR, 1 Trans reg c 3 TM, 1 HAMP, 1 HisKA, 1 HATPase c	
KY5 7459 KY5 7550	No	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 0 TM, 2 GAF, 1 HisKA 3, 1 HATPase c	Cluster situated within BGC 40
KY5 7698 KY5 7699	No	RR – NarL HK – Classic	1 RR, 1 HTH LuxR 4 TM, 1 HisKA 3, 1 HATPase c	
KY5 7932 KY5 7933	No	HK – Classic RR – NarL	4 TM, 1 HisKA 3, 1 HATPase c 1 RR, 1 HTH LuxR	
KY5 8002 KY5 8003	No	HK – Classic RR – OmpR	2 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	Cluster situated within BGC 42
KY5 8111 KY5 8112	No	HK – Classic RR – OmpR	2 TM, 1 HAMP, 1 HisKA, 1 HATPase c 1 RR, 1 Trans reg c	

3.2 The Cluster 15 Situated TCS

The KY5 1516/ 1517 genes that form TCS 15 are situated within BGC 15, predicted to encode the biosynthetic pathway for a terpene product. However, the exact identity of this product has not been elucidated and remains difficult to predict due to the 0% of the genes within this cluster having any similarity with any genes in other known clusters. The sensor kinase (1516) and response regulator (1517) are encoded as two separate genes of 2547 and 684 nucleotides in length, respectively, and are likely contained on the same transcript (**Fig. 3.2**).



Gene and Number	Amino Acids	Putative Function	Gene and Number	Amino Acids	Putative Function
1. 1505	252	Hypothetical protein	10. 1514	338	Terpene cyclase
2. 1506	62	Hypothetical protein	11. 1515	343	Ocatprenyl diphosphate synthase
3. 1507	271	Putative secreted protein	12. 1516	848	Osmosensitive K+ channel histidine kinase
4. 1508	77	Putative small membrane protein	13. 1517	227	Osmosensitive K+ channel response regulator
5. 1509	441	Acetylornithine deacetylase	14. 1518	222	K+ transporting ATPase C chain
6. 1510	163	Hypothetical protein	15. 1519	713	K+ transporting ATPase B chain
7. 1511	419	Glucosyltransferase	16. 1520	554	K+ transporting ATPase A chain
8. 1512	465	Cytochrome P450	17. 1521	88	Hypothetical protein
9. 1513	475	Glutamate-1-semialdehyde aminotransferase			

Figure 3.2 The genes within biosynthetic gene cluster 15 of *Streptomyces formicae* including the gene number, amino acid count and putative products. Included is as schematic of the genes including annotations of their function as described in antiSMASH 7.0 and StrepDB

BLASTn analysis of TCS 15 revealed a high degree of synteny (94% DNA identity, 100% query cover) with the osmotic K^+ sensing TCS, KdpDE from *Staphylococcus aureus*. Further investigation shows that it is one of the most widely distributed TCSs among bacteria and archaea and is found in strains such as *E. coli* and *Salmonella typhimurium* (Xie et al., 2020). It has been established to play a role in the regulation of K^+ transport, through mediating expression of the high affinity K^+ uptake system Kdp-ATPase (Xue et al., 2011). It has also been demonstrated to have direct involvement in the regulation of a series of virulence factors in response to external K^+ levels (Freeman, Dorus and Waterfield, 2013). It is not uncommon for TCSs to have pleiotropic effects enabling the regulation of multiple processes, something that KdpDE is clearly capable of. Whilst the majority of *Streptomyces* strains, including *S. formicae*, are not pathogenic and therefore don't produce virulence factors, it is possible that this TCS with a relatively high level of synteny to KdpDE plays a

similar pleiotropic role to control the transport of K^+ and the biosynthesis of a natural product. It is highly likely that the RR of TCS 15 is able to regulate the transcription levels of genes 1518/ 1519/ 1520 as these encode the three protein subunits that comprise a K^+ transporting ATPase.

Generating the plasmid required for overexpression of these genes was relatively simple, as good quality genomic DNA could be isolated from *S. formicae* grown in liquid medium. This genomic DNA was used for PCR amplifying and cloning the operon containing TCS 15 into the pIJ10257 backbone for constitutive overexpression. A 3257-nucleotide region covering these genes was assembled as 3 fragments into the NdeI cut site using Gibson Assembly. Once generated and confirmed via PCR and sequencing, this plasmid (pIJ10257 1516/ 1517) was conjugated into the wildtype and Δ*forGF* backgrounds of *S. formicae.* Colonies were initially screened via antibiotic selection to determine the presence of the plasmid before being further confirmed via PCR. The strains were grown on SFM, MYM, LB

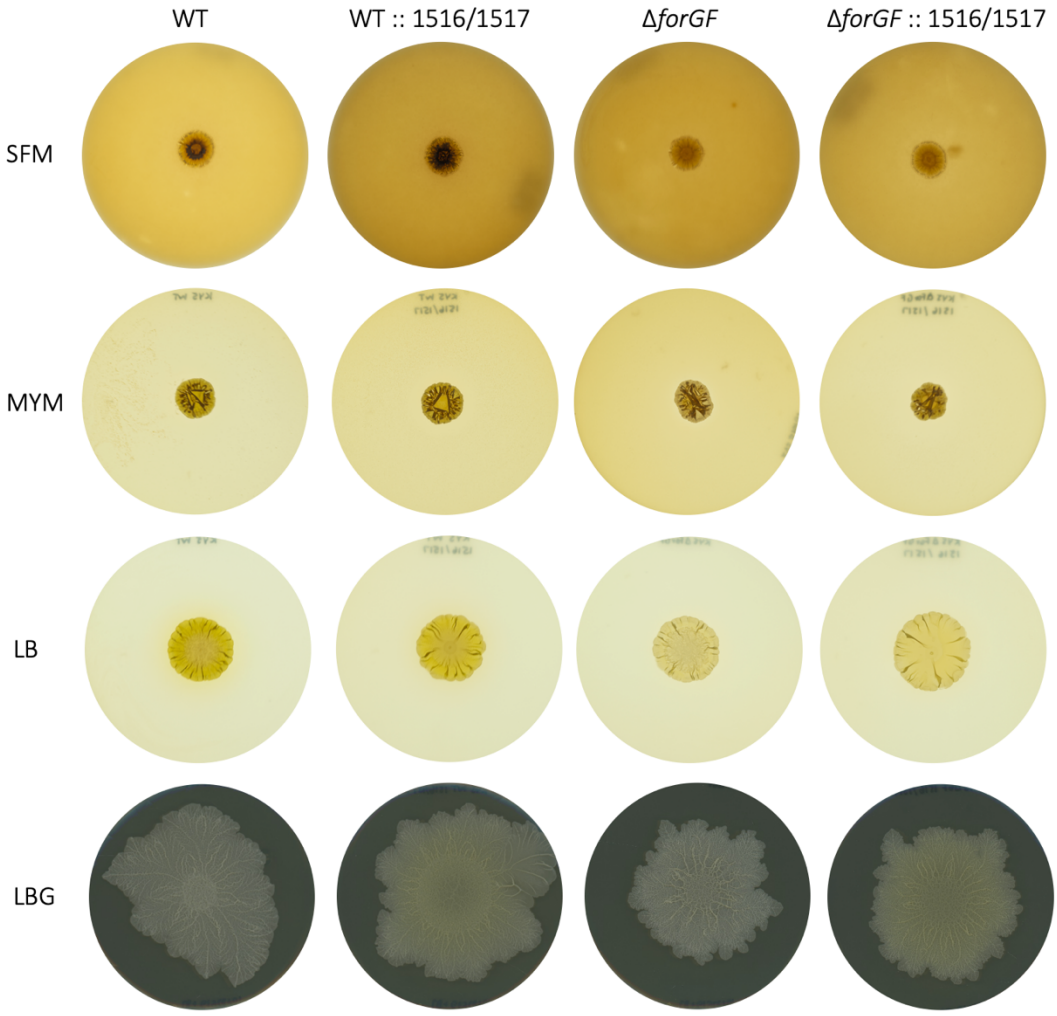


Figure 3.3 Phenotype comparison of *Streptomyces formicae* wildtype and Δ*forGF* strains following the introduction of pIJ10257 1516/1517 (TCS 15). Strains were spot plated onto SFM, MYM, LB and LB + Glycerol.

and LB + Glycerol to establish the phenotype across different media and assess any changes to phenotype following introduction of the plasmid (**Fig 3.3**). There were no apparent changes to any of the phenotypes across all media and all strains and replicates grew consistently with the parent wildtype and Δ*forGF* strains.

To determine any changes in antimicrobial activity that may have resulted from the overexpression of these genes, bioassays against *B. subtilis*, *E. coli*, *C. albicans* and MRSA were set up on SFM, MYM, LB and LB + Glycerol in biological and technical triplicate. When grown on LB + Glycerol and challenged with MRSA, the strains containing the overexpression plasmid displayed a significant increase in the observable zone of inhibition (**Fig 3.4**). A small zone of inhibition can be seen in the wildtype, and *S. formicae* has previously been established to have resistance to MRSA through the production of formicamycin, but not on this media. No such inhibition is visible in the *S. formicae* Δ*forGF* culture, likely due to the loss of formicamycin production. However, new inhibition and a dark-coloured compound are visible around the strains containing the TCS 15 overexpression plasmid. This suggests that overexpression of these genes resulted in the production of a compound not seen in the wildtype or Δ*forGF* strains that is active against MRSA.

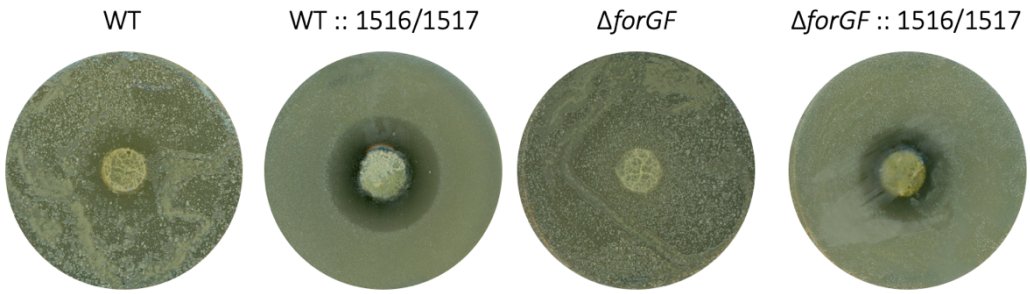


Figure 3.4 Bioassays of *Streptomyces formicae* wildtype and Δ*forGF* strains with and without the pIJ10257 1516/1516 (TCS 15) overexpression plasmids against MRSA. Grown on LB + Glycerol.

To further understand this bioactivity and establish whether it might be possible to identify the compound(s) responsible, crude extracts of these strains were analysed via HPLC (**Fig 3.5**). A series of new peaks were evident at UV 250 nm between the 14- and 16-minute retention times in both the wildtype and Δ*forGF* strains containing the overexpression plasmid. These peaks are likely to correlate with the new bioactivity observed in these strains and are worth pursuing further. As previously mentioned, there is a high likelihood that the response regulator of TCS 15 interacts with the genes encoding the K^+ ATPase transporter situated at the end of the same BGC based on the homology with the KdpDE

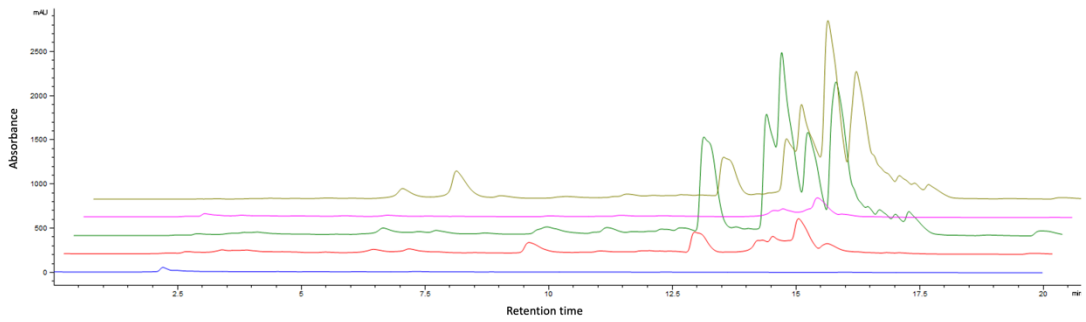


Figure 3.5 HPLC traces (UV 250 nm) showing crude extracts of LB + Glycerol media blank (blue), *S. formicae* wildtype (red), *S. formicae* :: 1516/1517 (dark green), *S. formicae* Δ*forGF* (pink) and *S. formicae* Δ*forGF* :: 1516/1517 (brown). HPLC (UV) LCMS analysis conducted in conjunction with Dr Hannah McDonald.

TCS. It is therefore possible that these new peaks are as a result of a metabolic shift in response to an influx of K^+ within the cell. However, in association with the new bioactivity it is more likely that these peaks are associated with an increase in the biosynthesis of a secondary metabolite as a result of the overexpression of TCS 15. The homologous KdpDE is also known to activate the biosynthesis of virulence factors and it is possible that TCS 15 is acting in a similar pleiotropic manner to act within multiple regulatory cascades including the ATPase and a natural product.

In addition to this increased bioactivity against MRSA when grown on LB + Glycerol, similar changes were seen when the strains were grown on LB and challenged with *B. subtilis* and *E. coli*. Once again, the strains containing the TCS 15 overexpression plasmid showed increases in the zones of inhibition against these pathogens compared to their parent wildtype and Δ*forGF* counterparts (**Fig 3.6**).

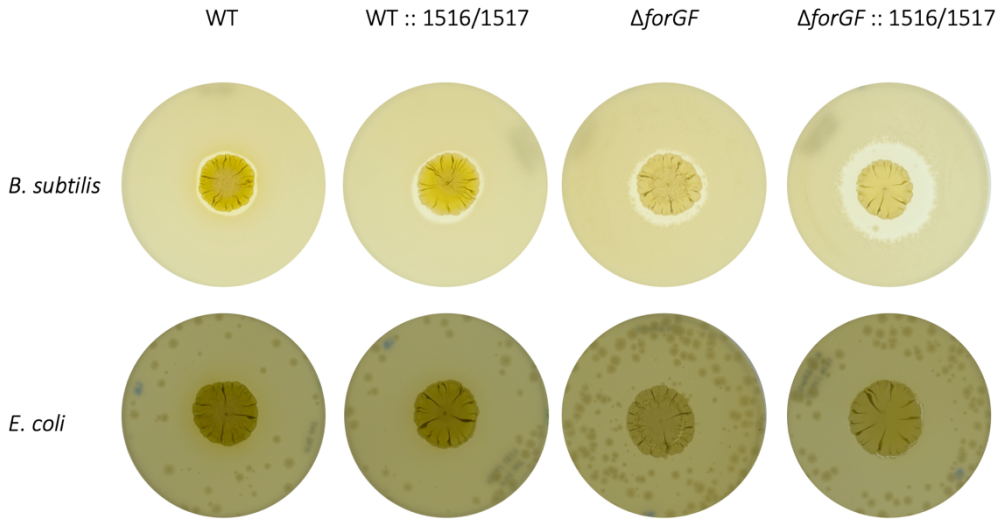


Figure 3.6 Bioassays of *Streptomyces formicae* wildtype and Δ*forGF* strains with and without the pIJ10257 1516/1516 (TCS 15) overexpression plasmids against *B. subtilis* and *E. coli*. Grown on LB.

For the *B. subtilis* challenge, a very small zone of inhibition can be seen in the wildtype and a slightly bigger zone around the Δ*forGF* strain. Previously, removal of the formicamycin BGC via CRISPR/Cas9 has been shown to maintain bioactivity against *B. subtilis* despite the loss of this potent antibiotic, hypothesised to be due to the activation of another BGC (Devine, unpublished). This small increase in bioactivity between the wildtype and Δ*forGF* strains is therefore expected, however, addition of the TCS 15 overexpression vector results in a further increase in bioactivity, demonstrated by the bigger zones of inhibition.

For the *E. coli* challenge, a reasonable zone of inhibition can be seen in the wildtype strain but reducing in the Δ*forGF* strain. Upon addition of the TCS 15 overexpression vector, both the wildtype and the Δ*forGF* strains show a marked increase in bioactivity.

Crude extracts of all four strains were analysed via HPLC alongside the LB media blank to again determine whether it might be possible to identify the compound(s) responsible for this bioactivity in the future (**Fig 3.7**). A series of new peaks were seen at a UV of 250 nm between the 13- and 16-minute retention times. Much smaller, but similar peaks can be seen at this retention time in both the wildtype and Δ*forGF* strains. This correlates with the hypothesis that BGC 15 is active at low background levels but adding a second copy of TCS 15 under the control of a constitutively active promoter results in a significant increase in the biosynthesis.

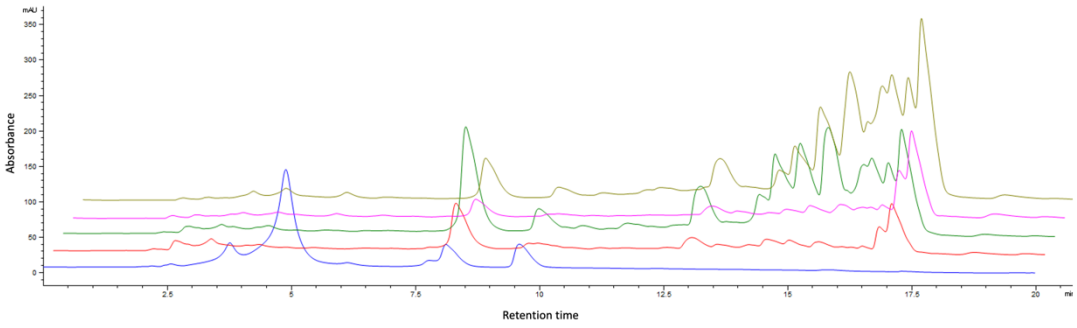


Figure 3.7 HPLC traces (UV 250 nm) showing crude extracts of LB media blank (blue), *S. formicae* wildtype (red), *S. formicae* :: 1516/1517 (dark green), *S. formicae* Δ*forGF* (pink) and *S. formicae* Δ*forGF* :: 1516/1517 (brown). HPLC (UV) LCMS analysis conducted in conjunction with Dr Hannah McDonald.

The range of increased bioactivity following the addition of the TCS 15 overexpression plasmid suggests that a secondary metabolite is overproduced as a result. The compound(s) appear to display broad bioactivity against both Gram-positive and Gram-negative bacteria. No changes in bioactivity were seen when these strains were challenged with *C. albicans,* suggesting that the compound(s) only possesses antibacterial properties (data not shown).

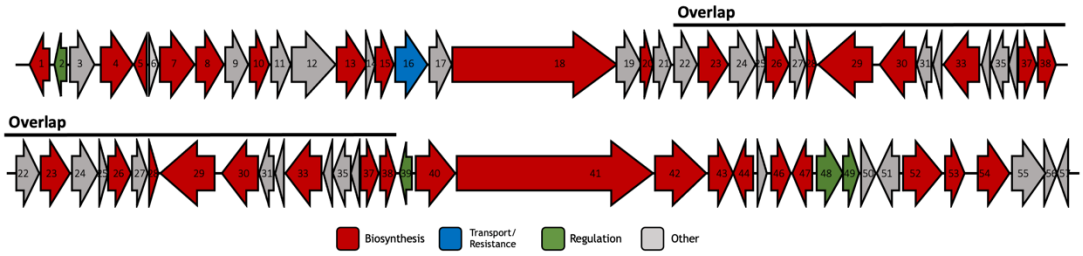
The presence of new peaks when analysed via HPLC correlate with the hypothesis that this cluster situated regulator is able to activate the biosynthesis of a natural product, however there is currently no data to confirm what this compound is or the location of the genes responsible for its biosynthetic pathway.

3.3 The Cluster 33 Situated TCS

The KY5 6663/ 6664 genes that form ForGF are situated within BGC 33, known to encode biosynthetic pathway for the formicamycins (Devine et al., 2021).The sensor kinase (6663) and response regulator (6664) are expressed as two separate genes of 1092 and 660 nucleotides in length, respectively, and are encoded on the same transcript under the control of a single promoter. The TCS, the cluster that it is situated in, and the regulation of the natural product encoded by the BGC are the focus of the remainder of this thesis. The bioinformatic analysis of these genes and impact of changes in expression levels on the production of secondary metabolites will be covered in more detail later on.

3.4 The Cluster 34 Situated TCS

The KY5 6759/ 6760 genes that form TCS 34 are situated within BGC 34, predicted to encode an NRP/ polyketide hybrid cluster. However, the exact identity of this product has not been elucidated and remains difficult to predict with any accuracy due to only 14% of the genes within this cluster having any similarity with any genes in other known clusters. The closest predicted products based on these genes are foxicin A/B/C, a set of nitrogen-containing quinone derivative compounds that act as siderophores by interacting with ferric ions and were first isolated from *Streptomyces diastatochromogenes* (Greule et al., 2017). The sensor kinase (6759) and response regulator (6760) are encoded as two separate genes of 1230 and 669 nucleotides in length, respectively, and are likely encoded by the same transcript (**Figure 3.8**). While the foxicin BGC is known to be widely distributed among *Streptomyces* bacteria and contains a two-component system (FoxRII/III), there is very low sequence identity between these TCS genes and those that make up TCS 34. The foxicin BGC is a hybrid cluster comprised of NRPS and T1PKS elements that work together to produce the final compounds (Greule et al., 2017). Closer inspection of the similarity of the genes between clusters reveals that the 14% of genes with similar sequences comes from the core



Gene and Number	Amino Acids	Putative Function	Gene and Number	Amino Acids	Putative Function
1. 6711	316	Esterase/ lipase/ thioesterase	30. 6740	571	Acyl-CoA dehydrogenase, short-chain specific
2. 6712	214	TetR transcriptional regulator	31. 6741	275	acyl-CoA thioesterase
3. 6713	359	L-lactate dehydrogenase	32. 6742	88	thiamine biosynthesis protein ThiS
4. 6714	468	Long chain fatty acid CoA ligase	33. 6743	589	Non-ribosomal peptide synthase
5. 6715	234	4'-phosphopantetheinyl transferase	34. 6744	116	4Fe-4S ferredoxin, iron-sulfur binding
6. 6716	88	Phosphopantetheine-binding-protein	35. 6745	316	Hypothetical protein
7. 6717	548	Long-chain-fatty-acid--CoA ligase	36. 6746	68	Hypothetical protein
8. 6718	422	3-oxoacyl-[acyl-carrier-protein] synthase, KASII	37. 6747	285	Alpha/beta hydrolase
9. 6719	354	4-hydroxyphenylpyruvate dioxygenase	38. 6748	259	Enoyl-[acyl-carrier-protein] reductase [NADH]
10. 6720	305	Phytanoyl-CoA dioxygenase	39. 6749	220	TetR/AcrR family transcriptional regulator
11. 6721	302	Arogenate dehydrogenase	40. 6750	577	Malonyl CoA-acyl carrier protein transacylase
12. 6722	663	Nitrogen regulatory protein P-II	41. 6751	2967	Malonyl CoA-acyl carrier protein transacylase
13. 6723	476	3-carboxy-cis,cis-muconate cycloisomerase	42. 6752	801	Polyketide synthase
14. 6724	95	Hypothetical Protein	43. 6753	374	Putative esterase/lipase
15. 6725	281	Alpha/beta hydrolase	44. 6754	300	Hydrolase, putative
16. 6726	481	Putative conserved transmembrane transport protein	45. 6755	153	Hypothetical Protein
17. 6727	352	Putative acyl-ACP desaturase, Stearoyl-ACP desaturase	46. 6756	332	N5,N10-methylenetetrahydromethanopterin reductase-related protein
18. 6728	2446	Malonyl CoA-acyl carrier protein transacylase	47. 6758	302	Class I SAM-dependent methyltransferase
19. 6729	390	Salicylate hydroxylase	48. 6759	409	Putative two-component system sensor kinase
20. 6730	140	Limonene-1,2-epoxide hydrolase	49. 6760	222	DNA-binding response regulator, LuxR family
21. 6731	277	2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase	50. 6761	252	DUF4232 domain-containing protein
22. 6732	372	3,7-dideoxy-D-threo-hepto-2, 6-diulosonate synthase	51. 6762	354	Lipoate-protein ligase A
23. 6733	473	Long-chain-fatty-acid--CoA ligase	52. 6763	613	Gamma-glutamyltranspeptidase
24. 6734	427	Short-chain alcohol dehydrogenase family	54. 6764	275	Putative inositol monophosphatase
25. 6735	82	Hypothetical Protein	54. 6765	472	NAD(P)/FAD-dependent oxidoreductase
26. 6736	402	3-oxoacyl-[acyl-carrier-protein] synthase, KASII	55. 6766	499	Methylated-DNA--protein-cysteine methyltransferase
27. 6737	235	Ketosynthase	56. 6767	166	Methylated-DNA--protein-cysteine methyltransferase
28. 6738	143	ComA operon protein 2	57. 6768	166	NTP pyrophosphohydrolase including oxidative damage repair enzyme
29. 6739	805	Butyryl-CoA dehydrogenase			

Figure 3.8 The genes within biosynthetic gene cluster 34 of *Streptomyces formicae* including the gene number, amino acid count and putative products. This hybrid cluster is comprised of NRPS and T1PKS biosynthetic genes so the potential overlap if these were to be separated into definitive BGCs is also shown. Included is as schematic of the genes including annotations of their function as described in antiSMASH 7.0. and StrepDB.

PKS genes and malonyl CoA carrier proteins, both of which are generic in the context of PKS BGCs and not specific to any particular cluster or product. The product of this cluster therefore remains completely unknown and difficult to predict with accuracy beyond the type of NP but holds the potential for a novel compound with unique structural components.

BLASTn analysis of TCS 34 revealed a reasonable degree of synteny (95% DNA sequency identity, 96% query cover) with two genes encoded within the *Streptomyces kanamyceticus* genome, 06875 and 06870. Interestingly, in *S. formicae* BGC 34 is next to the cluster that encodes the formicamycins, and antiSMASH analysis shows that these genes are in a similar respective location to BGC 12 in *S. kanamyceticus* which is predicted to encode the fasamycins (this BGC lacks homologues of *forXYZAA* and is therefore unable to make the formicamycins). Further inspection of antiSMASH analysis shows these two strains have a similar combination and order of predicted BGCs encoded, suggesting that one may be an evolutionary ancestor for the other or that they are evolutionarily related.

Generating the plasmid required for overexpression of these genes was relatively simple, as good quality genomic DNA could be isolated from *S. formicae* grown in liquid medium. This genomic DNA was used for PCR amplifying and cloning the operon containing TCS 34 into the pIJ10257 backbone for constitutive overexpression. A 1835-nucleotide region covering these genes was assembled as 3 fragments into the NdeI restriction site using Gibson Assembly. Once generated and confirmed via PCR and sequencing, this plasmid (piJ10257 1516/ 1517) was conjugated into the wildtype and Δ*forGF* backgrounds of *S. formicae.* Colonies were initially screened via antibiotic selection to determine the presence of the plasmid before being further confirmed via PCR. The strains were grown on SFM, MYM, LB and LB + Glycerol to establish the phenotype across different media (**Fig 3.9**). There were no apparent changes to any of the phenotypes across all media and all strains and replicates grew consistently with the wildtype and Δ*forGF* strains.

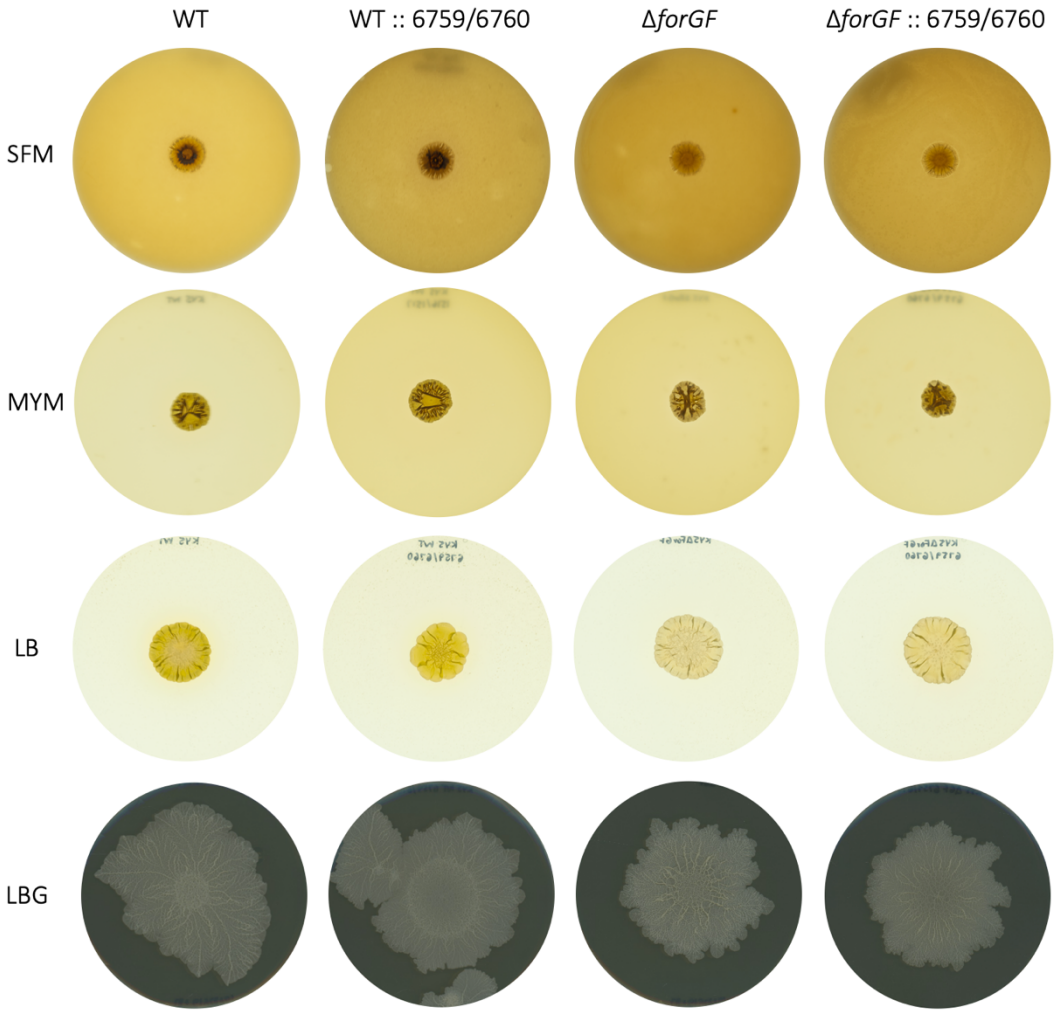


Figure 3.9 Phenotype comparison of *Streptomyces formicae* wildtype and Δ*forGF* strains following the introduction of pIJ10257 6759/6760 (TCS 34). Strains were spot plated onto SFM, MYM, LB and LB + Glycerol.

To determine any changes in antimicrobial activity that may have resulted from the overexpression of these genes, bioassays against *B. subtilis*, *E. coli*, *C. albicans* and MRSA were set up. When grown on MYM and challenged with MRSA, one of the strains containing the TCS 34 overexpression plasmid showed a change in bioactivity compared to their empty parent counterparts (**Fig 3.10**). While the wildtype background with the overexpression plasmid showed no significant change in the zone of inhibition around the strain, this can be expected due to the potency of formicamycins which will still be produced under these conditions. However, the Δ*forGF* strain shows a loss of some of this bioactivity, with the addition of the TCS 34 overexpression plasmid resulting in an increased and prominent zone of inhibition.

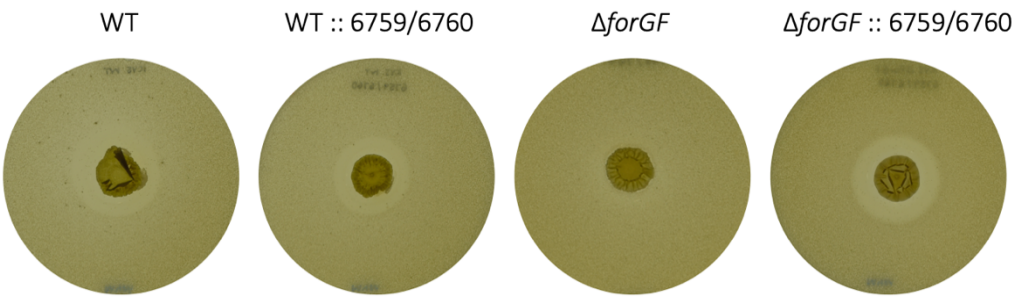


Figure 3.10 Bioassays of *Streptomyces formicae* wildtype and Δ*forGF* strains with and without the pIJ10257 6759/6760 (TCS 34) overexpression plasmids against MRSA. Grown on MYM.

To establish whether it would be possible to identify the compound(s) responsible for this new bioactivity in the Δ*forGF* strain, all four strains were analysed via HPLC alongside a media blank (**Fig 3.11**). A series of new peaks were seen in the wildtype and Δ*forGF* strains containing the TCS 34 overexpression plasmid between the 16- and 18- minute retention times. There is strong potential that these new peaks correlate with the new bioactivity as there are no apparent similar peaks either the wildtype or Δ*forGF* parent strains. This suggests that the compound(s) which may be responsible for the bioactivity are not normally produced, but the addition of a second copy of TCS 34 under the control of a constitutively active promoter has activated the biosynthesis in some way.

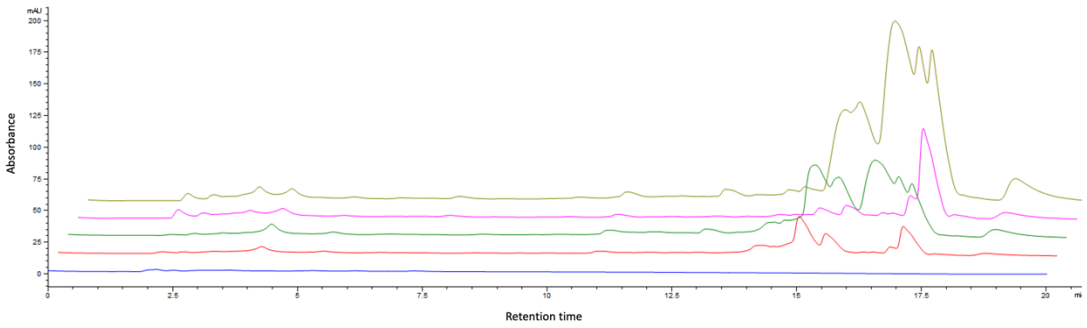


Figure 3.11 HPLC traces (UV 285 nm) showing crude extracts of MYM media blank (blue), *S. formicae* wildtype (red), *S. formicae* :: 6759/6760 (dark green), *S. formicae* Δ*forGF* (pink) and *S. formicae* Δ*forGF* :: 6790/6760 (brown). HPLC (UV) LCMS analysis conducted in conjunction with Dr Hannah McDonald.

In addition to the changes in bioactivity seen when grown on MYM and challenged with MRSA, there is also an increase in bioactivity when grown on LB and challenged with *B. subtilis* (**Fig 3.12**). There is a small zone of inhibition around the wildtype strain with an established increase in bioactivity in comparison in the Δ*forGF* strain. However, the addition of the TCS 34 overexpression plasmid results in a marked increase in bioactivity in both the wildtype and Δ*forGF* backgrounds.

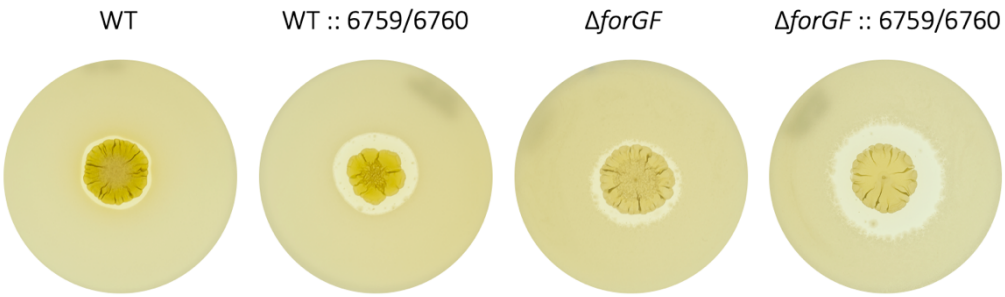


Figure 3.12 Bioassays of *Streptomyces formicae* wildtype and Δ*forGF* strains with and without the pIJ10257 6759/6760 (TCS 34) overexpression plasmids against *B. subtilis*. Grown on LB.

To establish whether there are any changes within the strains, all four were assessed by HPLC alongside a media blank (**Fig 3.13**). Once again, a series of new peaks were visible in both strains containing the TCS 34 overexpression plasmid with no apparent similar peaks present in the wildtype or Δ*forGF* counterparts. These peaks are present at 250 nm between the 13.5- and 15.5- minute retention times. This further supports the hypothesis that the compound responsible for this bioactivity is not normally expressed but its biosynthesis is somehow activated by the addition of a second copy of TCS 34 under the control of a constitutively active promoter.

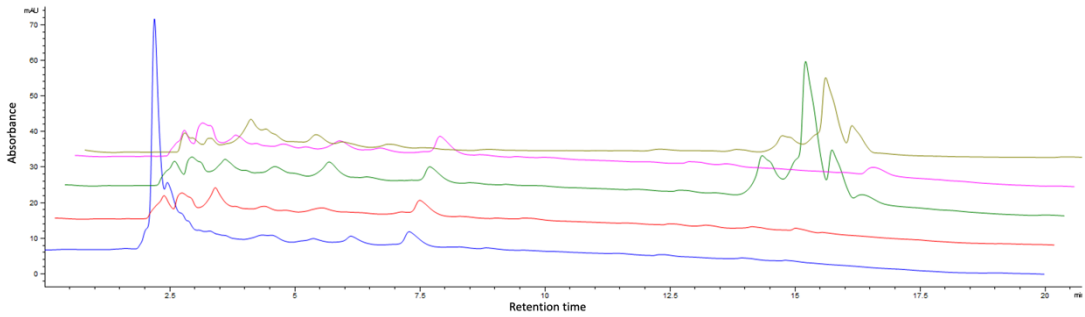


Figure 3.13 HPLC traces (UV 250 nm) showing crude extracts of LB media blank (blue), *S. formicae* wildtype (red), *S. formicae* :: 6759/6760 (dark green), *S. formicae* Δ*forGF* (pink) and *S. formicae* Δ*forGF* :: 6790/6760 (brown). HPLC (UV) LCMS analysis conducted in conjunction with Dr Hannah McDonald.

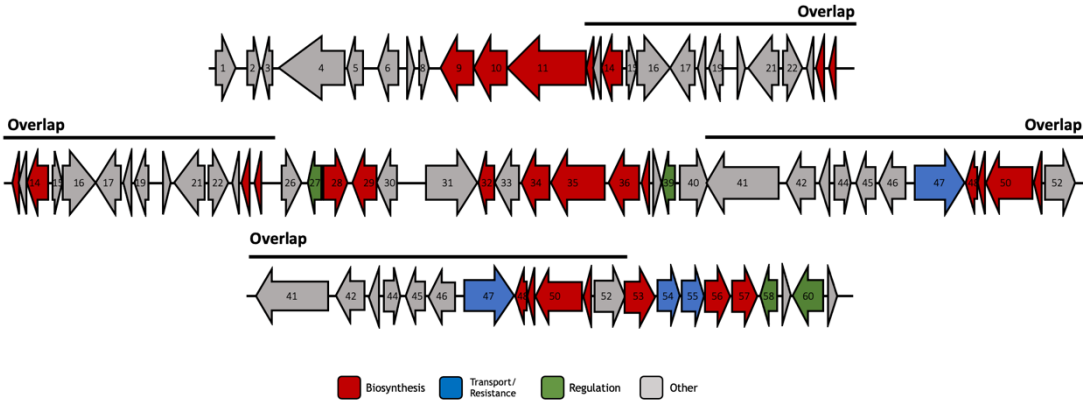
The increased bioactivity following the addition of the TCS 34 overexpression plasmid suggests that a secondary metabolite is overproduced as a direct result of this TCS. The compound appears to have specific activity against Gram-positive bacteria as no changes were seen when these strains were challenged with Gram-negative bacterial or fungal pathogens (data not shown). The presence of new peaks when analysed via HPLC suggest

that this is a compound not normally produced by the parent strains and that addition of a constitutively produced TCS has activated its biosynthesis. There are no current data to prove that the compound is produced by BGC 34 that TCS 34 is situated within, however this would be a reasonable first line of investigation to follow in order to identify the compound(s).

3.5 The Cluster 37 Situated TCS

The KY5 7292/ 7294 genes that form TCS 37 are situated within BGC 37, predicted to encode another NRP/ polyketide product. However, the exact identity of this product has not been elucidated and is difficult to predict due to only 22% of the genes in this cluster showing similarity to genes within any other known clusters. The closest predicted products based on these genes, are lagunapryone A/B/C, a set of cytotoxic acetogenins first isolated from an unknown marine actinomycete strain (Lindel et al., 1996). The sensor kinase (7294) and response regulator (7292) are encoded as two separate genes, both on the 5' stand of DNA, with another gene sitting between them on the 3' strand (**Fig 3.14**). Further analysis reveals that the 22% of genes with similarity to others include an alpha-pyrone synthase and a methyltransferase, both of which are largely unspecified genes often involved in PKS clusters meaning the product of this BGC remains relatively unpredictable.

BLASTn analysis of TCS 37 revealed a high degree of synteny (94% DNA sequence identity, 100% query cover) with two genes encoded within the *S. kanamyceticus* genome, 03855 and 03865. Within this strain there is also a third gene on the opposite DNA strain between these two, 03860 which also shows a similar level of similarity in its nucleotides. P2RP analysis predicts 03855 and 03865 to encode a sensor kinase and response regulator, respectively, although antiSMASH does not indicate that these genes are cluster situated within *S. kanamyceticus*, and there are currently no available data indicating their role.



Gene and Number	Amino Acids	Putative Function	Gene and Number	Amino Acids	Putative Function
1. 7235	258	Thiamin biosynthesis lipoprotein ApbE	32. 7266	217	Hypothetical protein
2. 7236	173	Transporter	33. 7267	319	putative acyl-ACP desaturase, Stearoyl-ACP desaturase
3. 7237	138	Putative transposase for insertion sequence element IS112	34. 7268	372	Chalcone synthase
4. 7238	825	Putative peptidoglycan bound protein (LPXTG motif)	35. 7269	710	Integral membrane protein
5. 7239	198	Hypothetical Protein	36. 7270	407	3-oxoacyl-[acyl-carrier-protein] synthase, KASII
6. 7240	257	IS5/IS1182 family transposase	37. 7271	89	Acyl carrier protein
7. 7241	104	Hypothetical protein	38. 7272	75	Hypothetical Protein
8. 7242	135	Acyl-CoA thioesterase	39. 7273	191	Transcriptional regulator, TetR family
9. 7243	409	Protein-L-isoaspartate O-methyltransferase	40. 7274	372	Saccharopine dehydrogenase
10. 7244	394	Lanthionine biosynthesis cyclase LanC	41. 4275	894	Putative signal transduction protein containing Nacht domain
11. 7245	1020	Lanthionine biosynthesis protein LanB	42. 7276	364	Methyl-accepting chemotaxis protein
12. 7246	48	FxLD family lantipeptide	43. 7277	115	Putative integral membrane protein
13. 7247	85	Hypothetical protein	44. 7278	215	Octanoate-[acyl-carrier-protein]-protein-N-octan oyltransferase
14. 7248	265	Hypothetical protein	45. 7279	262	Cobalt-zinc-cadmium resistance protein CzcD
15. 7249	140	NUDIX hydrolase	46. 7280	363	Small-conductance mechanosensitive channel
16. 7250	387	Hypothetical protein	47. 7281	627	ABC transporter ATP-binding/transmembrane protein
17. 7251	307	Hypothetical protein	48. 7282	147	Hypothetical protein
18. 7252	64	Hypothetical protein	49. 7283	87	Hypothetical protein
19. 7253	160	Hypothetical protein	50. 7284	635	Asparagine synthetase [glutamine-hydrolyzing]
20. 7254	79	Hypothetical protein	51. 7285	45	Hypothetical protein
21. 7255	413	XRE family transcriptional regulator	52. 7286	376	ROK family transcriptional regulator
22. 7256	230	HAD-superfamily hydrolase subfamily IA, variant 3	53. 7287	418	N-Acetyl-D-glucosamine ABC transport system, sugar-binding protein
23. 7257	31	Alcohol dehydrogenase	54. 7288	314	N-Acetyl-D-glucosamine ABC transport system, permease protein 1
24. 7258	125	Alcohol dehydrogenase	55. 7289	289	N-Acetyl-D-glucosamine ABC transport system, permease protein 2
25. 7259	65	Alcohol dehydrogenase	56. 7290	344	Myo-inositol 2-dehydrogenase
26. 7260	160	Transcriptional regulator	57. 7291	337	NADH-dependent dihydrogenase
27. 7261	190	Transcriptional regulator, TetR family	58. 7292	245	DNA-binding response regulator, LuxR family
28. 7262	316	NADP-dependent oxidoreductase	59. 7293	104	Hypothetical protein
29. 7263	304	ABC transporter substrate-binding protein	60. 7294	400	Two-component system sensor kinase
30. 7264	256	Hypothetical Protein	61. 7295	152	Lipoprotein, putative
31. 7265	650	Hypothetical protein			

Figure 3.14 The genes within biosynthetic gene cluster 37 of *Streptomyces formicae* including the gene number, amino acid count and putative products. This hybrid cluster is comprised of class 1 lanthipeptide, T3PKS and lassopeptide biosynthetic genes so the potential overlap if these were to be separated into definitive BGCs is also shown. Included is as schematic of the genes including annotations of their function as described in antiSMASH 7.0. and StrepDB.

Genomic DNA was again used for amplification and cloning of the operon encoding TCS 34 into the pIJ10257 vector for constitutive overexpression. A 1872-nucleotide region covering these genes was assembled as 3 fragments with overlaps generated to ensure that KY5 7293 would not be included. This was assembled into the NdeI restriction site using Gibson Assembly. Once generated and confirmed via PCR and sequencing, this plasmid (pIJ10257 7292/ 7294) was conjugated into the wildtype and Δ*forGF* backgrounds of *S. formicae.* Colonies were initially screened via antibiotic selection to determine the presence of the plasmid before being further confirmed via PCR. The strains were grown on SFM, MYM, LB and LB + Glycerol to establish the phenotype across different media (**Fig 3.15**). There were no apparent changes to any of the phenotypes across all media and all strains and replicates grew consistently with the wildtype and Δ*forGF* strains.

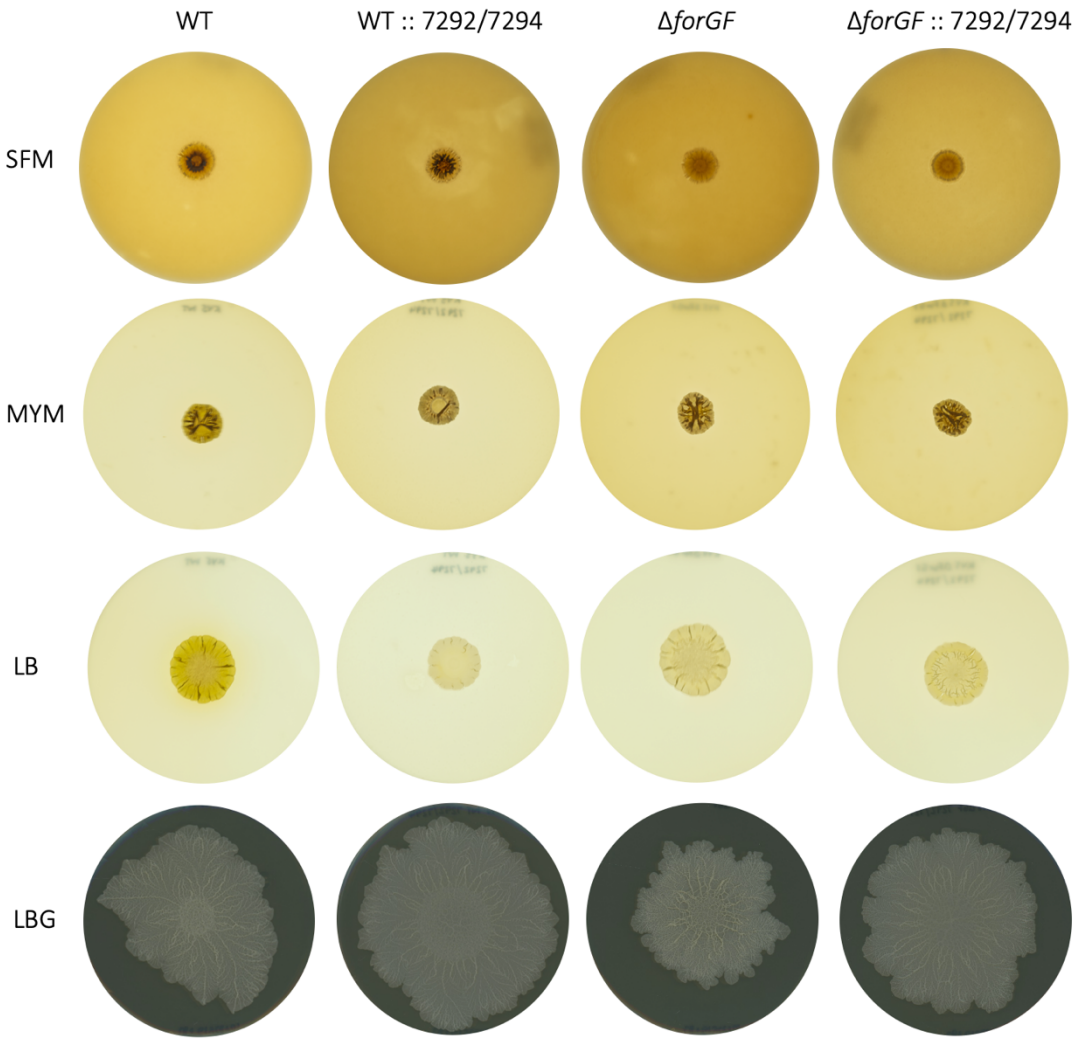


Figure 3.15 Phenotype comparison of *Streptomyces formicae* wildtype and Δ*forGF* strains following the introduction of pIJ10257 7292/7294 (TCS 37). Strains were spot plated onto SFM, MYM, LB and LB + Glycerol.

To determine any changes in antimicrobial activity that have resulted from the overexpression of these genes, bioassays against *B. subtilis*, *E. coli*, *C. albicans* and MRSA were set up. When grown on MYM and challenged with *C. albicans*, a significant increase in bioactivity was seen in the strains containing the TCS 37 overexpression plasmid (**Fig 3.16**). The wildtype strain showed a reasonable zone of inhibition around the strain and the Δ*forGF* strain showed a zone of inhibition roughly double this size. This is an established change in bioactivity against fungal pathogens by *S. formicae* and is thought to be through the resultant activation of BGC 6 (Devine, unpublished). However, addition of the TCS 37 overexpression plasmid further increased this zone of inhibition in both the wildtype and Δ*forGF* backgrounds to almost clear the plate of pathogen growth. This is a very significant increase in bioactivity following the addition of the plasmid.

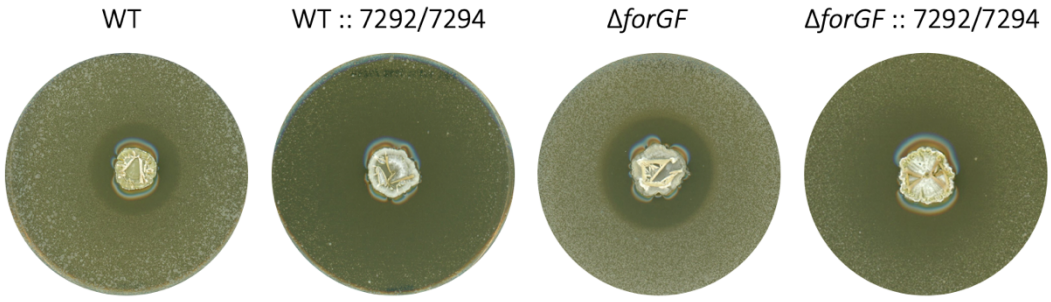


Figure 3.16 Bioassays of *Streptomyces formicae* wildtype and Δ*forGF* strains with and without the pIJ10257 7292/7294 (TCS 37) overexpression plasmids against *C. albicans*. Grown on MYM.

To establish whether it is possible to identify the compound(s) responsible for this new increased bioactivity, these strains were analysed via HPLC alongside a media blank (**Fig 3.17**). This identified a series of new peaks that are only present at significantly lower levels (if at all) in the wildtype and Δ*forGF* counterparts at the 14.5- to 16.5- minute retention time. It is unlikely that this compound would be produced at such low background levels due to the metabolic cost of secondary metabolism to the producing organism. It is

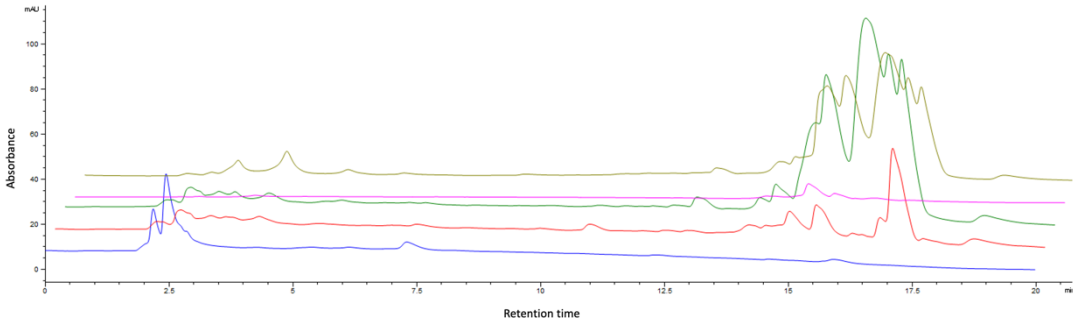


Figure 3.17 HPLC traces (UV 315 nm) showing crude extracts of LB media blank (blue), *S. formicae* wildtype (red), *S. formicae* :: 7292/7294 (dark green), *S. formicae* Δ*forGF* (pink) and *S. formicae* Δ*forGF* :: 7292/7294 (brown). HPLC (UV) LCMS analysis conducted in conjunction with Dr Hannah McDonald.

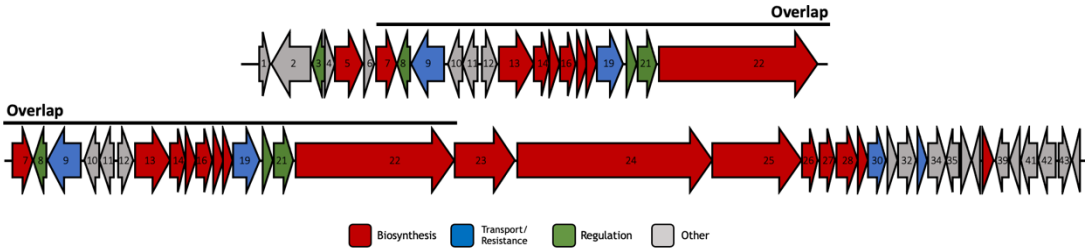
therefore more likely that the additional copy of TCS 37 under the control of a constitutively active promoter has resulted in the activation of the compound's biosynthetic pathway leading to overproduction and the associated bioactivity.

The new antifungal activity that is observed following introduction of the TCS 37 overexpression plasmid provides a promising route of investigation for the compound responsible. There were no notable changes to bioactivity when these strains were challenged with any bacterial pathogens, suggesting that this compound only possesses antifungal properties (data not shown). Whilst there is no direct evidence that TCS 37 acts within BGC 37, this would be a logical first line of investigation as part of identifying the source of this bioactivity and the BGC that encodes it.

3.6 The Cluster 39 Situated TCS

The KY5 7446/ 7447 genes that form TCS 39 are situated within BGC 39, predicted to encode another NRP/ polyketide product. Within this cluster, 92% of genes show similarity with the genes that make up the cluster that produces corbomycin in WAC 01529. This is a glycopeptide that is known to alter cell wall remodelling during growth by binding to peptidoglycan and inhibiting the action of autolysins (Culp et al., 2020). The sensor kinase (7447) and response regulator (7446) are encoded as two separate genes of 1281 and 213 nucleotides, respectively, and are likely encoded by the same transcript (**Fig 3.18**). Further investigation reveals that the sensor kinase is included in the 92% of genes that are similar to the corbomycin BGC, but the response regulator is not. This provides an interesting line of investigation as to whether the response regulator is present within the genome and may have been omitted as a sequencing/ assembly artefact, is situated elsewhere in the genome or whether the TCS exists as a hybrid in the corbomycin producing organism.

BLASTn analysis of TCS 39 revealed a high degree of synteny (96 % DNA sequence identity, 100 % query cover) with two genes encoded within the *S. kanamyceticus* genome, 02955 and 02960. P2RP analysis also predicts 02955 and 02960 to produce a sensor kinase and response regulator, respectively. AntiSMASH analysis of the *S. kanamyceticus* genome shows these genes to sit within BGC 4, which has high sequence identity to the cluster known to produce corbomycin. This suggests that BGC 4 of *S. kanamyceticus* and BGC 39 of *S. formicae* are likely of the same origin and further supports the idea that these strains are evolutionary ancestors.



Gene and Number	Amino Acids	Putative Function	Gene and Number	Amino Acids	Putative Function
1. 7427	252	Glyoxalase PtlL in pentalenolactone biosynthesis	23. 7449	1538	Siderophore biosynthesis non-ribosomal peptide synthetase modules
2. 7428	943	Hypothetical protein	24. 7450	4624	Siderophore biosynthesis non-ribosomal peptide synthetase modules
3. 7429	309	Putative lysR-family transcriptional regulator	25. 7451	2132	Siderophore biosynthesis non-ribosomal peptide synthetase modules, Bacillibactin synthetase component F
4. 7430	223	Succinate dehydrogenase cytochrome b subunit	26. 7452	392	Putative cytochrome P450 hydroxylase
5. 7431	652	Succinate dehydrogenase flavoprotein subunit	27. 7453	398	Putative cytochrome P450 hydroxylase
6. 7432	248	Succinate dehydrogenase iron-sulfur protein	28. 7454	537	Tryptophan halogenase
7. 7433	504	Long-chain-fatty-acid--CoA ligase	29. 7455	177	Putative oxidoreductase
8. 7434	292	AraC family transcriptional regulator	30. 7456	459	Sodium/Hydrogen exchanger
9. 7435	785	Excinuclease ABC subunit A	31. 7457	187	Hypothetical protein
10. 7436	365	(S)-2-hydroxy-acid oxidase	32. 7458	496	Putative peptidoglycan binding domain 1
11. 7437	368	4-hydroxyphenylpyruvate dioxygenase	33. 7459	249	ABC transporter, ATP-binding protein
12. 7438	373	Arogenate dehydrogenase	34. 7460	390	ABC transporter permease
13. 7439	799	Alpha-aminoadipate aminotransferase	35. 7461	347	LLM class flavin-dependent oxidoreductase
14. 7440	357	Chalcone synthase	36. 7462	187	Hypothetical protein
15. 7441	222	Enoyl-CoA hydratase	37. 7463	71	Dodecin domain-containing protein
16. 7442	452	Enoyl-CoA hydratase	38. 7464	428	Sugar transporter
17. 7443	281	Beta-ketoadipate enol-lactone hydrolase	39. 7465	320	Dimethylhistidine N-methyltransferase
18. 7444	74	MbtH protein	40. 7466	251	Glutamine amidotransferases class-II
19. 7445	705	Lipid A export ATP-binding/permease protein MsbA	42. 7467	447	Serine/threonine kinase
20. 7446	213	Two-component response regulator	42. 7468	458	Glutamate--cysteine ligase EgtA
21. 7447	426	Two-component sensor kinase	43. 7469	282	TIGR02452 family protein
22. 7448	3719	Siderophore biosynthesis non-ribosomal peptide synthetase modules, Bacillibactin synthetase component F	44. 7470	143	RNA 3'-terminal phosphate cyclase

Figure 3.18 The genes within biosynthetic gene cluster 39 of *Streptomyces formicae* including the gene number, amino acid count and putative products. This hybrid cluster is comprised of T3PKS and NRPS biosynthetic genes so the potential overlap if these were to be separated into definitive BGCs is also shown. Included is as schematic of the genes including annotations of their function as described in antiSMASH 7.0. and StrepDB.

Genomic DNA was used for amplification and cloning of the operon containing TCS 39 into the pIJ10257 vector for constitutive overexpression. A 1971-nucleotide region covering these genes was assembled as 3 fragments into the NdeI restriction site using Gibson Assembly. Once generated and confirmed via PCR and sequencing, this plasmid (pIJ10257 7446/ 7447) was conjugated into the wildtype and Δ*forGF* backgrounds of *S. formicae.*

Colonies were initially screened via antibiotic selection to determine the presence of the plasmid before being further confirmed via PCR. The strains were grown on SFM, MYM, LB and LB + glycerol to establish the phenotype across different media (**Fig 3.19**). There were no apparent changes to any of the phenotypes across all media and all strains and replicates grew consistently with the wildtype and Δ*forGF* strains.

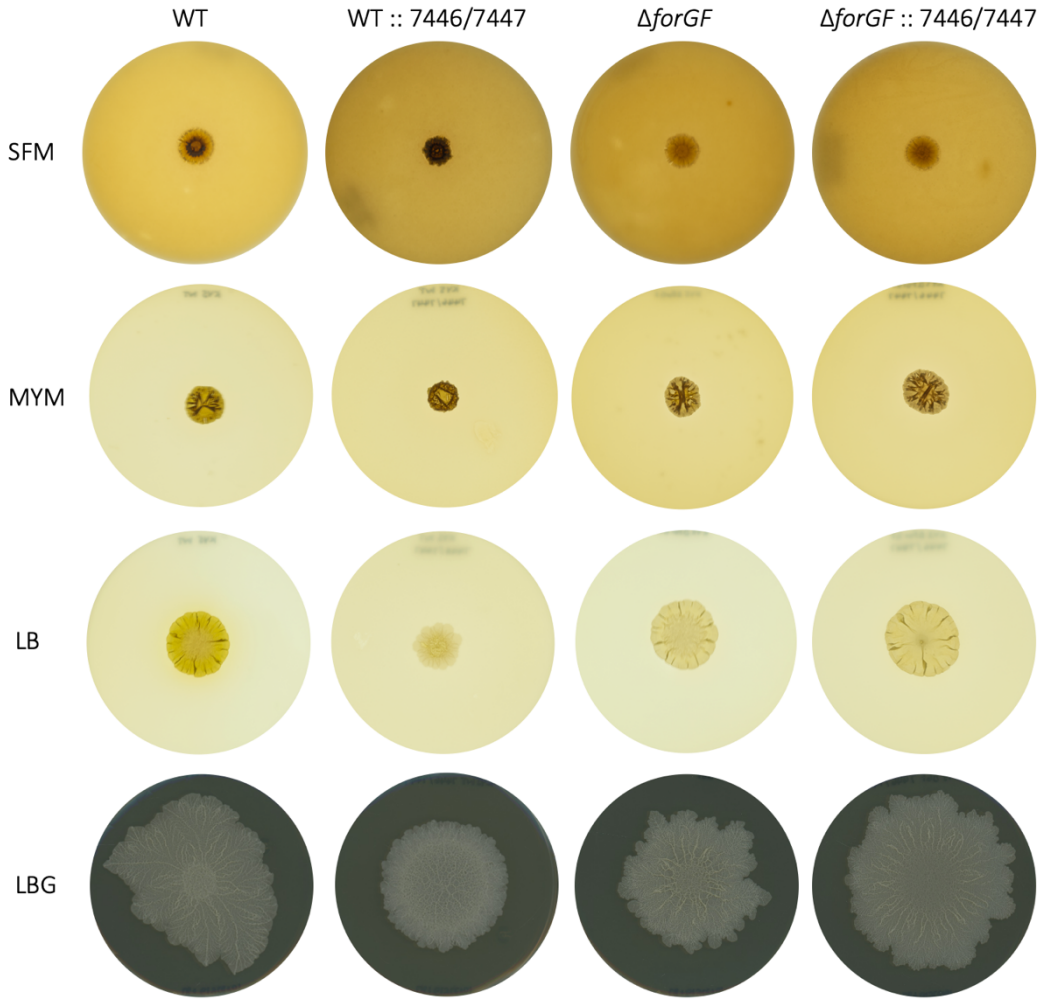


Figure 3.19 Phenotype comparison of *Streptomyces formicae* wildtype and Δ*forGF* strains following the introduction of pIJ10257 7446/7447 (TCS 39). Strains were spot plated onto SFM, MYM, LB and LB + Glycerol.

To determine any changes in antimicrobial activity that may have resulted from the overexpression of these genes, bioassays against *B. subtilis*, *E. coli*, *C. albicans* and MRSA were set up. When grown on MYM and challenged with *B. subtilis*, there was a small increase in bioactivity in the strains containing the overexpression plasmid in comparison to the wildtype and Δ*forGF* counterparts (**Fig 3.20**). A small zone of inhibition can be seen around the wildtype, with no significant change observed between this strain and the Δ*forGF* strain. However, following introduction of the TCS 39 overexpression plasmid, both

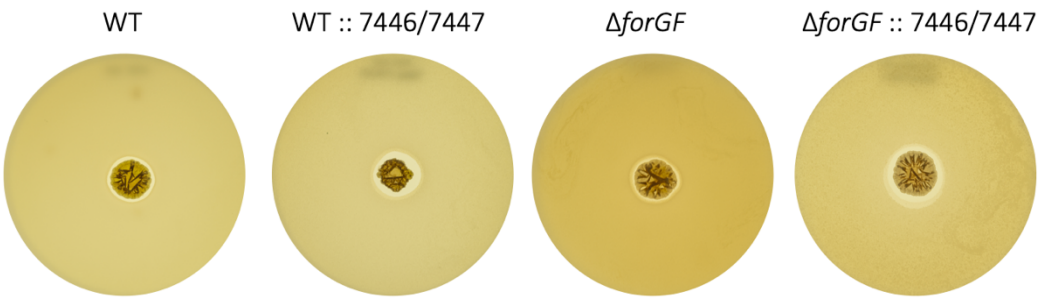


Figure 3.20 Bioassays of *Streptomyces formicae* wildtype and Δ*forGF* strains with and without the pIJ10257 7446/7447 (TCS 39) overexpression plasmids against *B. subtilis*. Grown on MYM.

strains showed a marked increase in the bioactivity against this *B. subtilis*. It is possible that this increase is as a result of the second copy of TCS 39 under the control of a constitutively active promoter.

To establish whether it may be possible to identify the compound(s) responsible for this bioactivity, these strains were assessed via HPLC alongside a media blank (**Fig 3.21**). A single distinctive peak can be seen at the 9-minute retention time in all four of these strains. However, it is significantly larger in the two strains containing the TCS 39 overexpression plasmid. It is possible that this peak is the compound associated with the increase bioactivity and correlated with the zones of inhibition seen across the four strains.

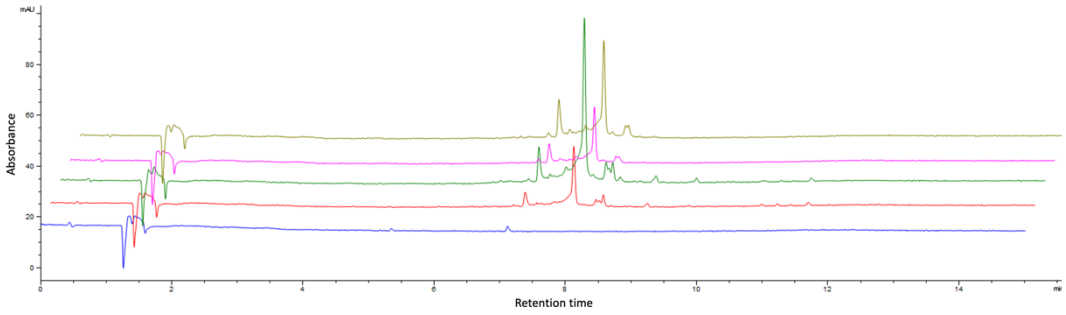


Figure 3.21 HPLC traces (UV 365 nm) showing crude extracts of LB media blank (blue), *S. formicae* wildtype (red), *S. formicae* :: 7292/7294 (dark green), *S. formicae* Δ*forGF* (pink) and *S. formicae* Δ*forGF* :: 7292/7294 (brown). HPLC (UV) LCMS analysis conducted in conjunction with Dr Hannah McDonald.

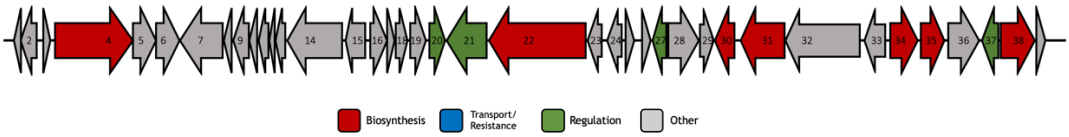
The zones of inhibition seen were relatively small across the bioassays conducted against *B. subtilis* and there were no changes to bioactivity seen in any of the other challenges, however, this is still a promising indication of a potential method of activating the biosynthesis of a bioactive natural product. Whilst there is no direct evidence that TCS 39 acts within BGC 39, this would be a logical first line of investigation as part of identifying the source of this bioactivity and the BGC that encodes it.

3.7 The Cluster 40 Situated TCS

KY5 7549/ 7550 genes that form TCS 40 are situated within BGC 40, predicted to encode the biosynthesis of a non-ribosomal peptide product. However, the exact identity of this product has not been elucidated and remains difficult to predict as there is very low sequence identity between this cluster and any other known cluster. The closest prediction is a 2% sequence identity with gausemycin A/B, a set of lipoglycopeptides isolated from *S. kanamyceticus* (Tyurin et al., 2021). The sensor kinase (7550) and response regulator (7549) exist as two separate genes of 1644 and 687 nucleotides respectively however they are convergent and encoded by separate transcripts (**Fig 3.22**).

BLASTn analysis of TCS 40 revealed a reasonable degree of synteny with two genes (84% DNA sequency identity, 97% query cover) in *S. kanamyceticus* but interestingly not with any genes associated with the gausemycin cluster.

Whilst initial attempts at cloning these genes into the overexpression plasmid pIJ10257were made, assembling the genes in the correct order with appropriate overlaps proved difficult. For the purpose of this preliminary investigation and with six other candidates of cluster situated TCSs, it was decided not to continue with this overexpression or to order synthetic DNA for cloning until experiments with other overexpression strains had proved successful.

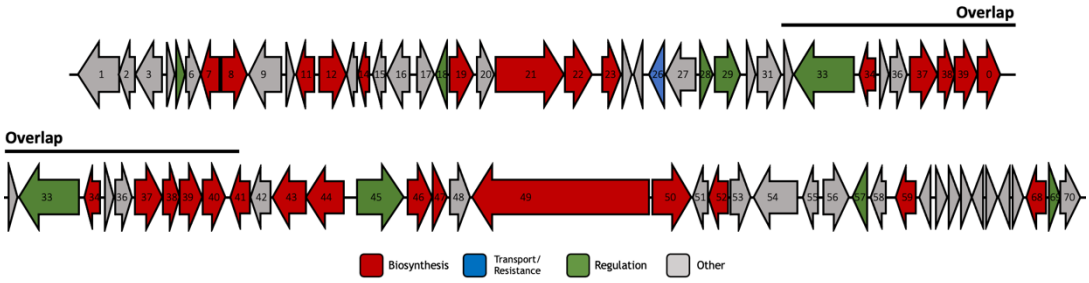


Gene and Number	Amino Acids	Putative Function	Gene and Number	Amino Acids	Putative Function
1. 7530	83	CDGSH iron-sulfur domain-containing protein	21. 7550	547	Putative two-component system histidine kinase
2. 7531	230	Putative methyltransferase	22. 7551	1324	Siderophore biosynthesis non-ribosomal peptide synthetase modules
3. 7532	121	putative membrane protein	23. 7552	221	Putative transmembrane protein
4. 7533	1087	Formate dehydrogenase O alpha subunit, selenocysteine-containing	24. 7553	210	Hsp18 transcriptional regulator
5. 7534	316	Formate dehydrogenase O beta subunit	25. 7554	143	Hsp20/alpha crystallin family protein
6. 7535	323	Formate dehydrogenase O putative subunit	26. 7555	148	DUF2267 domain-containing protein
7. 7536	581	NADH-quinone oxidoreductase subunit D	27. 7556	192	Transcriptional regulator, TetR family
8. 7537	115	Dehydrogenase	28. 7557	427	Oxidoreductase
9. 7538	251	Sodium:proton antiporter	29. 7558	187	Dihydrofolate reductase
10. 7539	90	DUF4040 domain-containing protein	30. 7559	265	Mandelate racemase
11. 7540	153	Hypothetical protein	31. 7560	607	Pyruvate Oxidase
12. 7541	58	Hypothetical protein	32. 7561	1031	Glycolate dehydrogenase, subunit GlcD
13. 7542	123	Glyoxalase/bleomycin resistance protein/dioxygenase	33. 7562	286	DUF72 domain-containing protein
14. 7543	784	Xylulose-5-phosphate phosphoketolase, Fructose-6-phosphate phosphoketolase	34. 7563	394	Threonine dehydrogenase-related Zn-dependent dehydrogenase
15. 7544	287	Universal stress protein family	35. 7564	329	3-oxoacyl-[acyl-carrier protein] reductase
16. 7545	243	Inosine-5'-monophosphate dehydrogenase	36. 7565	425	Hypothetical protein
17. 7546	87	Hypothetical protein	37. 7566	208	Tetracycline repressor protein
18. 7547	154	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinase	38. 7567	513	Salicylate hydroxylase
19. 7548	224	Inosine-5'-monophosphate dehydrogenase	39. 7568	136	Glyoxalase
20. 7549	228	Putative two-component system response regulator			

Figure 3.22 The genes within biosynthetic gene cluster 40 of *Streptomyces formicae* including the gene number, amino acid count and putative products. This hybrid cluster is comprised of NRPS, β-lactam and T1PKS biosynthetic genes so the potential overlap if these were to be separated into definitive BGCs is also shown. Included is as schematic of the genes including annotations of their function as described in antiSMASH 7.0. and StrepDB.

3.8 The Cluster 42 Situated TCS

The KY5 8002/ 8003 genes that form TCS 42 are situated within BGC 40, predicted to encode a NRP/ β-lactam/ polyketide product. Within this cluster, 52% of genes show similarity with the genes that make up the cluster that produces valclavam in *Streptomyces antibioticus*. This β-lactam has both antibiotic and antifungal activity through inhibiting methionine biosynthesis and inhibiting RNA synthesis, respectively (Nobary & Jensen, 2012; Röhl et al., 1987). The sensor kinase (8003) and response regulator (8002) are encoded as two separate genes of 1212 and 663 nucleotides, respectively, and are likely expressed on the same transcript (**Fig 3.23**).



Gene and Number	Amino Acids	Putative Function	Gene and Number	Amino Acids	Putative Function
1. 7975	595	Alpha-L-fucosidase	36. 8010	214	Putative 2-phosphosulfolactate phosphatase
2. 7976	266	Tat pathway signal sequence domain protein	37. 8011	457	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
3. 7977	404	Rhamnogalacturonides degradation protein RhiN	38. 8012	214	Nitroreductase
4. 7978	146	Putative lyase	39. 8013	335	Aldo-keto reductase
5. 7979	140	Transcriptional regulator, AraC family	40. 8014	379	Butyryl-CoA dehydrogenase
6. 7980	213	Glyoxalase	41. 8015	324	Clavaminate synthase 2 (Clavaminic acid synthetase 2) (CAS2) (CS2)
7. 7981	257	3-oxoacyl-[acyl-carrier protein] reductase	42. 8016	316	Agmatinase
8. 7982	387	EstA family serine hydrolase	43. 8017	509	(Carboxyethyl)arginine beta-lactam-synthase
9. 7983	508	Choline dehydrogenase	44. 8018	571	Acetolactate synthase large subunit
10. 7984	133	Mannose-6-phosphate isomerase	45. 8019	705	Transcriptional regulator, SARP family
11. 7985	280	Dimethylglycine N-methyltransferase	46. 8020	387	Glutamate N-acetyltransferase / N-acetylglutamate synthase
12. 7986	405	Putative Glycosyltransferase	47. 8021	192	NADH-FMN oxidoreductase
13. 7987	178	Mini-circle protein	48. 8022	264	Putative DeoR-family transcriptional regulator
14. 7988	208	Thioredoxin reductase	49. 8023	2689	Malonyl CoA-acyl carrier protein transacylase
15. 7989	177	Exonuclease SbcC	50. 8024	579	O-methyltransferase
16. 7990	378	Sugar dehydrogenase	51. 8025	283	Hypothetical protein
17. 7991	286	Acyl carrier protein	52. 8026	278	D-beta-hydroxybutyrate dehydrogenase
18. 7992	157	Transcriptional regulator, MarR family	53. 8027	291	putative DNA-binding protein
19. 7993	353	Enterochelin esterase	54. 8028	692	Peptidase M6
20. 7994	245	Hypothetical protein	55. 8029	234	Hypothetical protein
21. 7995	1041	Non-ribosomal peptide synthetase	56. 8030	441	Hydrolase
22. 7996	421	Amine oxidase, flavin-containing	57. 8031	236	DNA-binding response regulator
23. 7997	266	Aspartate aminotransferase	58. 8032	248	VCBS repeat-containing protein
24. 7998	123	PRC domain containing protein	59. 8033	316	KR domain-containing protein
25. 7999	47	Hypothetical Protein	60. 8034	181	Hypothetical protein
26. 8000	236	ABC transporter, ATP-binding protein	61. 8035	167	Hypothetical protein
27. 8001	490	ABC transporter, permease protein	62. 8036	195	DUF2975 domain-containing protein
28. 8002	220	Two-component response regulator	63. 8037	77	Transcriptional regulator, Cro/CI famil
29. 8003	403	Two-component system, sensor protein	64. 8038	182	DUF1062 domain-containing protein
30. 8004	166	Phage tail length tape-measure protein	65. 8039	127	Signal peptidase I
31. 8005	355	Cell envelope-associated transcriptional attenuator LytR-CpsA-Psr	66. 8040	171	Chitin binding protein
32. 8006	115	Hypothetical protein	67. 8041	153	Glyoxalase family protein
33. 8007	941	Serine/threonine-protein kinase PknB	68. 8042	326	L-fuco-beta-pyranose dehydrogenase
34. 8008	278	putative Arylesterase-related	69. 8043	140	Leucine-responsive regulatory protein, regulator for leucine (or lrp) regulon and high-affinity branched-chain amino acid transport system
35. 8009	131	Nuclear transport factor 2 family protein	70. 8044	300	Virginiamycin B lyase

Figure 3.23 The genes within biosynthetic gene cluster 40 of *Streptomyces formicae* including the gene number, amino acid count and putative products. This hybrid cluster is comprised of NRPS, β-lactam and T1PKS biosynthetic genes so the potential overlap if these were to be separated into definitive BGCs is also shown. Included is as schematic of the genes including annotations of their function as described in antiSMASH 7.0. and StrepDB.

BLASTn analysis of TCS 42 revealed a high degree of synteny (92% DNA sequence identity, 100% query cover) with two genes encoded within the *S. kanamyceticus* genome, 43015 and 43020. P2RP also predicts 43020 and 43015 to produce a sensor kinase and response regulator, respectively, although antiSMASH does not indicate that these genes would be cluster situated within *S. kanamyceticus*, and there is currently no available data indicating their role.

Genomic DNA was used for amplification and cloning of the operon encoding TCS 42 into the pIJ10257 vector for constitutive overexpression. A 1871-nucleotide region covering these genes was assembled as 3 fragments into the NdeI restriction site using Gibson Assembly. Once generated and confirmed via PCR and sequencing, this plasmid (pIJ10257 8002/ 8003) was conjugated into the wildtype and Δ*forGF* backgrounds of *S. formicae.* Colonies were initially screened via antibiotic selection to determine the presence of the plasmid before being further confirmed via PCR. The strains were grown on SFM, MYM, LB and LB + glycerol to establish the phenotype across different media (**Fig 3.24**). There were no apparent changes to any of the phenotypes across all media and all strains and replicates grew consistently with the wildtype and Δ*forGF* strains.

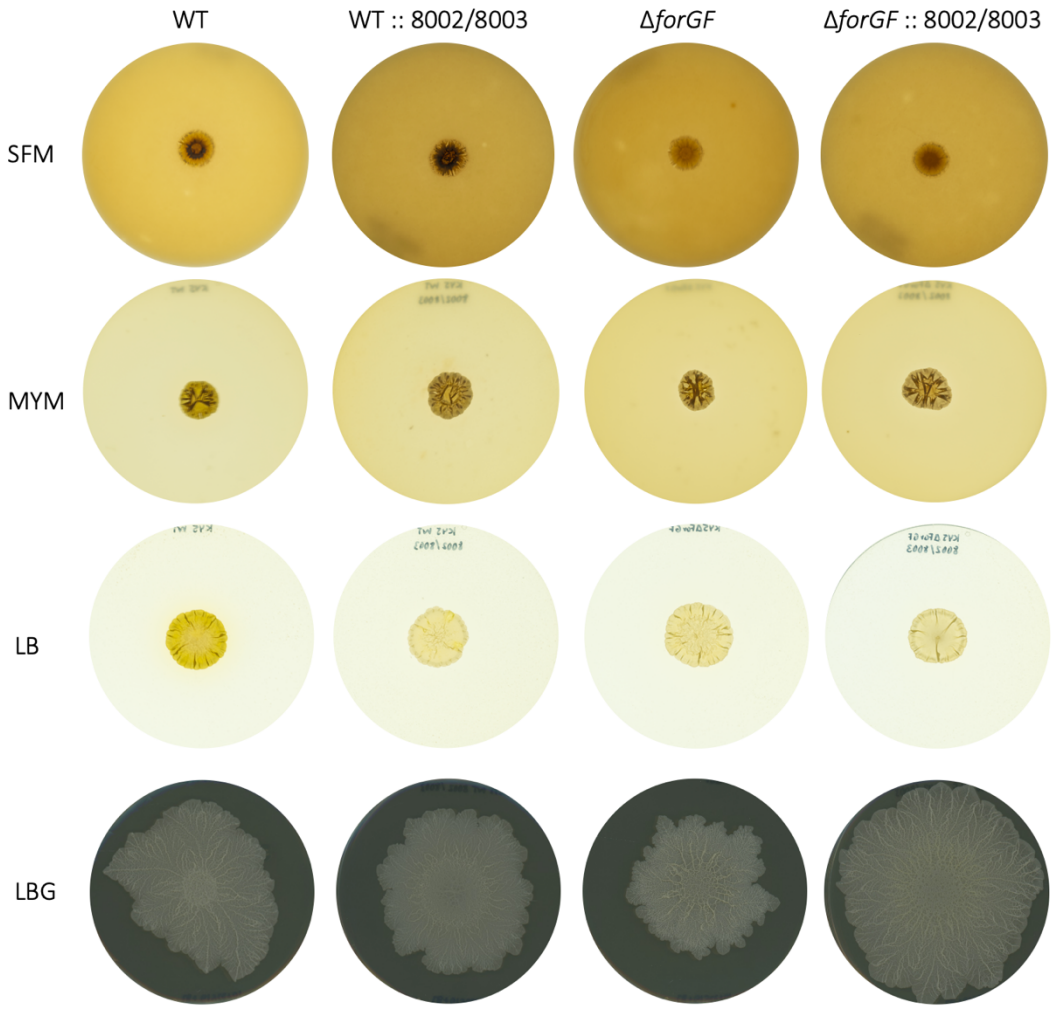


Figure 3.24 Phenotype comparison of *Streptomyces formicae* wildtype and Δ*forGF* strains following the introduction of pIJ10257 8002/8003 (TCS 42). Strains were spot plated onto SFM, MYM, LB and LB + Glycerol.

To determine any changes in antimicrobial activity that may have resulted from the overexpression of these genes, bioassays against *B. subtilis*, *E. coli*, *C. albicans* and MRSA were set up. When grown on LB and challenged with *B. subtilis*, the strains containing the overexpression plasmid appeared to lose any bioactivity that was previously present (**Fig 3.25**). The wildtype strain has a small zone of inhibition and the Δ*forGF* strain has a slightly larger zone of inhibition around the strain. However, both strains containing TCS 42 under the control of a constitutively active promoter have lost any bioactivity and have no zone of inhibition.

To establish whether this loss of bioactivity could be attributed to a compound no longer being present, all four strains were analysed via HPLC alongside a media blank (**Fig 3.26**). A distinctive peak between the 12.5- and 13.5- minute retention time is evident in both the

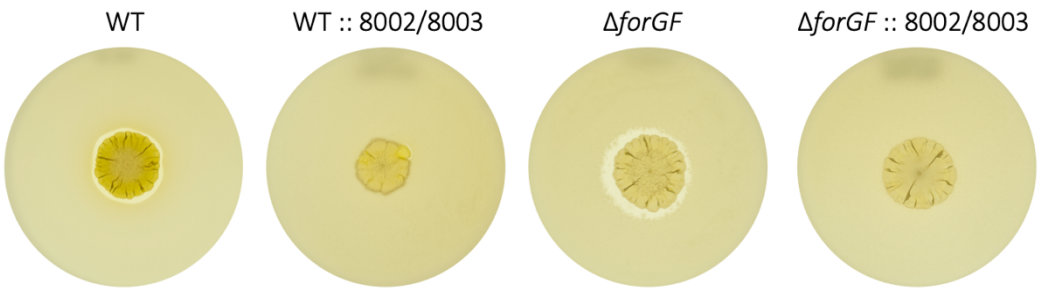


Figure 3.25 Bioassays of *Streptomyces formicae* wildtype and Δ*forGF* strains with and without the pIJ10257 8002/8003 (TCS 42) overexpression plasmids against *B. subtilis*. Grown on LB.

wildtype and Δ*forGF* strains but is not in either of the strains containing the TCS 42 overexpression plasmid. This suggests that introduction of the plasmid has resulted in the repression of the production of the compound that was previously providing bioactivity against *B. subtilis*. As previously discussed, TCSs are known to be able to both activate and repress various aspects of regulatory cascades, so it is possible that TCS 42 is responsible for the repression of this compound in the wildtype strains and overexpression has resulted in a complete shutdown of the biosynthetic pathway.

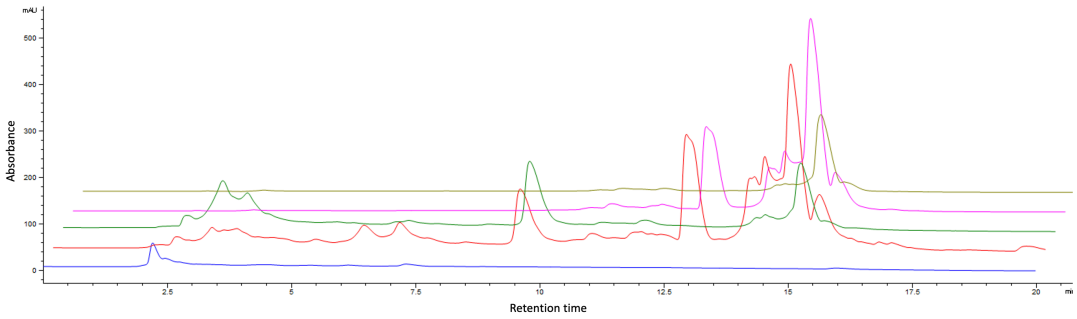


Figure 3.26 HPLC traces (UV 418 nm) showing crude extracts of LB media blank (blue), *S. formicae* wildtype (red), *S. formicae* :: 8002/8003 (dark green), *S. formicae* Δ*forGF* (pink) and *S. formicae* Δ*forGF* :: 8002/8003 (brown). HPLC (UV) LCMS analysis conducted in conjunction with Dr Hannah McDonald.

No further changes to bioactivity were observed in any of the other bioassays, however, the loss of bioactivity following introduction of this TCS provides a promising line of investigation. The ability to switch off a natural product through overexpression of TCS 42 suggests that deletion of the TCS may result in overproduction of the compound. Whilst there is no direct evidence that TCS 42 acts within BGC 42, this would be a logical first line of investigation as part of identifying the source of this loss of bioactivity and the BGC that encodes it.

3.9 Applications in Other *Streptomyces* Strains

The *Streptomyces* genus is widely renowned for its antibiotic producing abilities, but the current AMR crisis has raised understandable concerns about the development of new antibiotics for future use. While several approaches for tackling this issue are currently being discussed, the majority are reliant on new antimicrobials being discovered, with additional hope for molecules that function though novel mechanisms of action. However, there are no definitive and consistently successful methods for identifying such compounds, with a significant reduction in the number of new products being successful in reaching clinical trials since the 1950s. It has been shown here that through generally simple identification, targeting and overexpression of cluster-situated TCSs, new bioactivity against pathogens can be seen with all of those investigated.

With this in mind, it is important to assess the relevance and potential of applying this approach to other strains before determining if it could be a successful and worthwhile route of investigation for the future. Thus, the model organisms *S. coelicolor* and *S. venezuelae* were assessed for the number of cluster-situated TCSs contained within their genomes and the number of these clusters that encoded currently unknown products. Due to the previously observed similarity between TCSs at a nucleotide level and the structure of BGCs, *S. kanamyceticus* was also included in this assessment (**Table 3.2**).

Table 3.2 Biosynthetic gene clusters and two-component systems in *Streptomyces formicae, Streptomyces coelicolor, Streptomyces venezuelae* and *Streptomyces kanamyceticus* as predicted with P2RP and antiSMASH 7.0 software

Strain Name Accession no.	Number of BGCs	Number of TCSs	Number of Cluster-Situated TCSs (% of TCSs)	Number of Cluster-Situated Orphans
Streptomyces formicae CP022685	43	67	7 (10.4%)	5
Streptomyces coelicolor AL645882	26	40	5 (12.5%)	10
Streptomyces venezuelae ASM23038v1	32	58	6 (10.3%)	9
Streptomyces kanamyceticus GCA008704495	38	59	7 (11.9%)	3

An average of 11.2% of the BGCs within any of these strains contained a cluster-situated TCS, and on further inspection 11 of the 18 have less than 50% of genes with similarity to other known clusters (**Table 3.3**). An additional 3 which showed 100% gene similarity were actually part of predicted hybrid clusters with potential for these being two or more clusters with some degree of overlap. In all three of these cases, the TCS was encoded on the side where none of the genes showed similarity to those that produced a known product but still had core natural product biosynthetic genes. The remaining four TCSs were encoded within BGCs that had 53, 55, 87 and 92% similarity to other known clusters, demonstrating the potential of these producing further congeners of known products or different classes of natural products altogether.

Table 3.3 Genes that encode the cluster situated two-component systems alongside the predicted biosynthetic gene cluster and number of genes with similarity to other known clusters in *Streptomyces coelicolor, Streptomyces venezuelae* and *Streptomyces kanamyceticus* as predicted with P2RP and antiSMASH 7.0 software.

Genes	BGC	Predicted Cluster Type	Gene Similarity with Known BGCS
Sco_1312 Sco_1313	7	T3PKS/Lathipeptide	100% flaviolin
Sco_3339 Sco_3340	11	NRPS	87% CDA1/2/3/4
Sco_7099 Sco_7100	24	Other/ T3PKS	6% carbapenem
Sco_7516 Sco_7518	28	NRPS/T3PKS/Terpene	100% coelibactin
Sco_7578 Sco_7579	28	NRPS/T3PKS/Terpene	100% coelibactin
Sven_5289 Sven_5299	17	NI-Siderophore	6%
Sven_5333 Sven_5334	18	NI-Siderophore	22%
Sven_5368 Sven_5369	18	NI-Siderophore	22%
Sven_6076 Sven_6078	22	NRPS/RiPP	53% colibrimycin
Sven_6905 Sven_6906	28	NRPS	55% peucechelin
Sven_6986 Sven_6987	29	Terpene	5%
Skan_0347 Skan_0348	2	NRPS	12%
Skan_0533 Skan_0534	4	T2PKS/NRPS/T3PKS	92% corbomycin
Skan_0590 Skan_0591	4	T2PKS/NRPS/T3PKS	92% corbomycin
Skan_0769 Skan_0771	6	RiPP/Lassopeptide	50% ulleungdin
Skan_1459 Skan_1450	12	T2PKS	100% fasamycin
Skan_1961 Skan_1962	14	NRPS/T3PKS	42% CDA1/2/3/4
Skan_5161 Skan_1562	23	NRPS/PKS-like	12%

The concept of identifying and genetically manipulating the expression of cluster-situated TCSs within *Streptomyces* strains has significant potential for unlocking BGCs that are not currently expressed by wildtype strains under standard laboratory growth conditions. From initial assessment of these model organisms an additional 10% of the BGCs that they encode have the potential of being activated or repressed in this way, resulting in a change in the strain's bioactivity against pathogens. The ability to use strains that are already widely available, have established protocols within the laboratory environment and are known to be genetically tractable makes this an appealing route of further investigation. The number of BGCs encoded within other *Streptomyces* strains is well known but the inability to activate these remains an ongoing problem for the scientific community and the AMR crisis. However, this simple and easily investigated approach may be the key to identifying more antimicrobial agents for use in a wide range of environments.

3.10 Conclusions

This work has shown that the exploitation of cluster-situated two-component systems to unlock new bioactivity in *S. formicae* is very promising. The overexpression of TCS 15 resulted in new bioactivity against MRSA, *B subtilis* and *E. coli* covering both Gram-positive and Gram-negative bacteria in its new antibacterial activity. This was coupled with the potential of compound identification in the future after new peaks were seen in the HPLC analysis. In addition, the lack of any gene similarity between the cluster that TCS 15 is situated within with known clusters, beyond core PKS genes, holds great promise for the potential of this bioactivity being attributed to a novel compound. The overexpression of TCS 34 resulted in new bioactivity against MRSA and *B. subtilis* suggesting the activation of biosynthesis of a compound able to target Gram-positive bacteria. This was again coupled with new peaks from HPLC analysis, suggesting a potential of identifying the compound responsible for this increase in bioactivity with further investigations. The complete lack of any similarity of the genes contained in BGC 34, where TCS 34 is located, with any other known clusters again provides significant hope for the potential of a compound that may have a structure and/ or mechanism of action that has not previously been seen. Overexpression of TCS 37 saw an increase in antifungal activity rather than any changes to antibacterial bioactivity. An associated series of new peaks when analysed via HPLC provides a strong basis for further investigation to potentially identify the compound that

has provided this new bioactivity. Once again, the lack of any similar genes within BGC 37 to known gene clusters, beyond those commonly associated with PKS BGCs, suggests the potential for previously unseen compounds and the opportunity to investigate this increase in bioactivity much further. The overexpression of TCS 39 resulted in an increase in bioactivity against *B. subtilis* suggesting the presence of an antibiotic with activity against Gram-positive bacteria but no ability to target MRSA. A number of genes within BGC 39, where TCS 39 is situated, show similarity with those known to form part of the corbomycin BGC. However, corbomycin is a glycopeptide known to have activity against MRSA, and there is no evidence of TCS involvement in the regulation of this biosynthetic pathway. It is possible that BGC 39 encodes biosynthesis of a corbomycin-like molecule and overexpression of TCS 39 increased this production sufficiently to increase bioactivity against *B. subtilis* but not high enough to inhibit MRSA. However, the previously mentioned absence of the response regulator from known corbomycin BGCs suggests that TCS 39 may be involved in a regulating a different biosynthetic pathway. The new peaks seen in the HPLC analysis provide a strong platform for further investigations that would confirm or exclude the possibility of this compound being corbomycin or other such analogous compounds however this has not yet been investigated. Unfortunately attempts at overexpression of TCS 40 were unsuccessful at the genetic assembly stage, but clear evidence of successful activation of new bioactivity from the other cluster-situated TCSs appear to suggest that this would be a worthwhile additional line of investigation in the future. Overexpression of TCS 42 interestingly saw a reduction in bioactivity against *B. subtilis*, suggesting that an antibiotic is normally repressed by this regulator. The loss of peaks in these overexpression strains further points to this regulatory mechanism and provides a further basis for investigation through genetic manipulation of this TCS to either knock it out or render it inactive though point mutations. Around half of the genes encoded within BGC 42 where TCS 42 is situated show similarity to those known to be responsible for valclavam production. This would again be an interesting line of investigation to identify whether valclavam is produced by the wildtype strain and/ or whether this production is lost following overexpression of TCS 42.

With all of the above observations, there remain a lot of questions about the compounds responsible for the changes seen to the bioactivity against a range of pathogens and the BGCs that encode them. There is a distinct possibility that some or all of the RRs that make up these TCS do not target genes that form part of the BGC that they are situated in.

However, simply targeting and overexpressing cluster situated TCSs has consistently been shown to result in changes to bioactivity against a range of Gram-positive, Gram-negative and fungal pathogens. Identifying the targets of these RRs through further investigations alongside further pursuit of compound identification would be reasonable next lines of investigations (e.g ChIP-seq, RNA-seq and proteomics) to see if any of these changes in bioactivity hold promise for the development of new antimicrobial agents.

There is also significant potential for applying this approach to other *Streptomyces* strains as evidenced by the number of cluster-situated TCSs in just three other strains. A myriad of other *Streptomyces* strains are known to hold a wealth of potential from the BGCs they encode and altering the expression levels of TCSs in these strains may be the key to unlocking some of their cryptic natural products.

4. Characterising the Cluster-Situated Regulators of the Formicamycin Biosynthetic Gene Cluster

The synthesis of antimicrobial natural products often comes at great metabolic cost to the producing organism, and biosynthetic pathways are stringently regulated to ensure that these secondary metabolites are only ever produced when necessary. Controlling the expression of the coding genes means that all of this metabolic expense is saved before the pathway has the chance to begin. It also means that the litany of natural products that these strains are capable of producing as part of their competitive advantage are only used when required to provide the maximum impact while conserving the cell. Transcriptional regulation can be achieved by activators and repressors which mediate the recognition and/ or binding of RNA polymerase to the promoter regions of these genes. This binding can be inhibited by a repressor which physically blocks RNA polymerase binding or elongation in some way, meaning the gene(s) cannot be transcribed to messenger RNA which would subsequently be translated into protein by the ribosome. Activators enhance RNA polymerase binding, allowing the DNA strands around the promoter region and transcriptional start site to unwind and for the template strand of DNA to enter the active site of RNA polymerase for transcription. A higher level of transcription to mRNA and further translation into functional protein means there is more active protein present in the cell to carry out a variety of roles.

As previously discussed, transcriptional regulation in *Streptomyces* spp. is notoriously complex with a myriad of activators and repressors working both with and against one another to ensure that these secondary metabolites are only produced when absolutely necessary. The regulators that control the biosynthesis of secondary metabolites such as formicamycins can be cluster-situated and control their cognate natural products, or they can be global regulators that have pleiotropic effects throughout the genome to elicit a widespread regulatory effect. In addition, cross-cluster regulation is thought to be much more widespread than is currently reported. To add to this complexity, BGCs typically encode the genes required for the biosynthesis of a secondary metabolite on several operons that produce polycistronic mRNA rather than each gene being under the control of its own promoter and producing multiple strands of mRNA that must be individually translated. These genes generally have overlapping coding regions, thought to have been an evolutionary response known as translational coupling to make the production of multi-

enzyme complexes more efficient. If the functionality of one enzyme is linked to or dependant on another, it makes metabolic sense to have these genes encoded and expressed together as much as possible. In this system, the downstream gene is dependent on the expression of the upstream gene and the promoters of these transcripts are usually located directly upstream of the first gene's start codon in the intergenic region. These sites are often the targets of regulators, with binding allowing them to control transcriptional activity at the promoter.

The three cluster-situated regulators of formicamycin biosynthesis (ForJ, ForGF and ForZ) have been identified to have distinct roles in the regulatory pathway, acting upon various sites throughout the formicamycin BGC (Devine et al., 2021). This chapter will explore the roles of these regulators including their positions within the BGC, their binding sites and the impact of their binding on transcription and promoter activity.

4.1 An Overview of the Formicamycin BGC

The production of the enzymes responsible for actioning various aspects of a biosynthetic pathway begins with transcription to mRNA followed by translation to protein. Under normal circumstances, the first stage is carried out by RNA polymerase which must first recognise a promoter region for the genes it will be transcribing. Once bound, the promoter DNA around this transcriptional start site began to unwind, allowing the active catalytic site of the RNA polymerase to recognise the template for the enzymes to be produced. However, a multitude of factors may influence this initiation stage of the transcription process, including the binding of a transcriptional activator or the release of a transcriptional repressor. Activators enhance the binding and activity of RNA polymerase and repressors physically block it from accessing and/ or transcribing the DNA. To be able understand the mechanisms of regulation that act upon a biosynthetic pathway at the transcriptional level, it is important to understand the transcriptional organisation of the cluster.

Previous work in the Hutchings laboratory identified the transcription start sites, length of overlaps/ intergenic regions and the number of operons present in the formicamycin BGC through cappable RNA-sequencing (**Fig 4.1**). This used RNA that was extracted from 2-, 3- and 4-day old cultures of *S. formicae* and sent for capping and sequencing of the enzymatically modified 5' triphosphorylated end of the RNA at Vertis Biotech (no longer in business). The monophosphorylated mRNA was then compared with this tagged mRNA to identify the transcript start sites and establish the length and organisation of the transcripts through the presence of polycistronic mRNA.

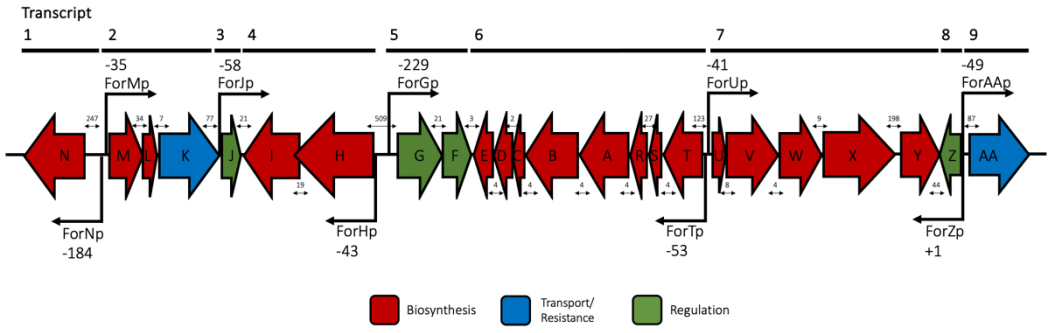


Figure 4.1 Transcriptional organisation of the formicamycin biosynthetic gene cluster. Promoter regions are detailed in relation to the start codon of the first gene in the transcript they control. The number of base pairs making up intergenic regions (above) and overlap (below) of genes are detailed with side-to-side arrows. Adapted from work by Dr Rebecca Devine with permission.

This established that throughout the cluster there are nine promoters, as indicated by the presence of these transcriptional start sites. Each of these are situated at varying positions in relation to the start codon of the first gene of the transcript under their control. In addition, intergenic regions range from a few base pairs to several hundred base pairs in length, with longer intergenic regions indicating a promoter that is responsible for driving the transcription of downstream genes. Several of these promoters are also divergent, driving the expression of genes encoded on opposite strands of DNA.

The acyl hydrolase encoded by *forN* is expressed on transcript 1 on its own, with the transcriptional start site positioned 184 base pairs upstream of the start codon. The methytransferase, PKS cyclase and Na^+/ H^+ exchanger encoded by *forM*, *forL* and *forK*, respectively, are all co-expressed on transcript 2 under the control of a promoter situated 35 base pairs upstream of the start codon for *forM*. The MarR regulator encoded by *forJ* is also expressed on its own as transcript 3 and is under the control of its own promoter that sits 58 base pairs upstream of the start codon. It has been proposed that the promoter region extends back into the coding region of *forK*, as the 77 base pair long intergenic region is not sufficient to accommodate a binding site for RNA polymerase as well as transcriptional regulators. The ACC biotin carboxylase and ACC carboxyl transferase encoded by *forH* and *forI* are responsible for the production of the malonyl CoA starter unit used by the For PKS. These genes are co-transcribed on transcript 4, under the control of a promoter situated 43 base pairs upstream of the start codon of *forH*. Similarly, the sensor kinase and response regulator encoded by *forG* and *forF*, respectively, are co-transcribed on transcript 5 and the transcriptional start site is situated some 229 base pairs upstream of the start codon of *forG*, suggesting there may be other binding sites for transcriptional regulators in this large intergenic region. The core PKS genes are encoded by *forABC* and are transcribed alongside a variety of PKS-tailoring enzymes as part of transcript 6. This is the longest transcript in the cluster, producing a final mRNA transcript that encompasses *forTSRABCDE*. All eight of these genes are under the control of a promoter situated 53 base pairs upstream of the start codon of *forT*. The remainder of the PKS tailoring enzymes that are critical to the production of formicamycin from the fasamycin precursors are produced as transcript 7 and includes *forUVWXY*. The promoter region of this transcript sits 41 base pairs upstream of the start codon of *forU*. Despite the large intergenic region of 198 base pairs between *forX* and *forY*, there is no transcriptional start site located in this region, indicating that *forY* must be included as part of the longer transcript controlled by the promoter p*forU*. There is also an

internal transcriptional start site situated in the coding region of *forV*, such internal promoters are known to exist in *Streptomyces* to ensure transcription of longer operons are maintained. They are generally present in transcripts where the terminal genes are situated some distance from the promoter region (Dangel et al., 2009). The co-transcription evident between transcripts 6 and 7 suggests the hypothetical formation of a multi-enzyme complex with dependency between functional enzymes and efficiency of their production being maintained at a transcriptional level. The final two transcripts are those that produce the MarR family regulator encoded by *forZ,* and the MFS family transporter encoded by *forAA*. The former of these, transcript 8, is produced by leaderless transcription with the mRNA transcript lacking a 5'UTR and Shine-Dalgarno sequence, meaning the start codon of *forZ* acts as a signal to initiate translation instead. Although not common among all bacteria, some 20% of actinobacterial genes with mapped transcriptional start sites are known to be leaderless. Transcript 9 is under the control of a promoter situated 49 base pairs upstream of the start codon of *forAA*.

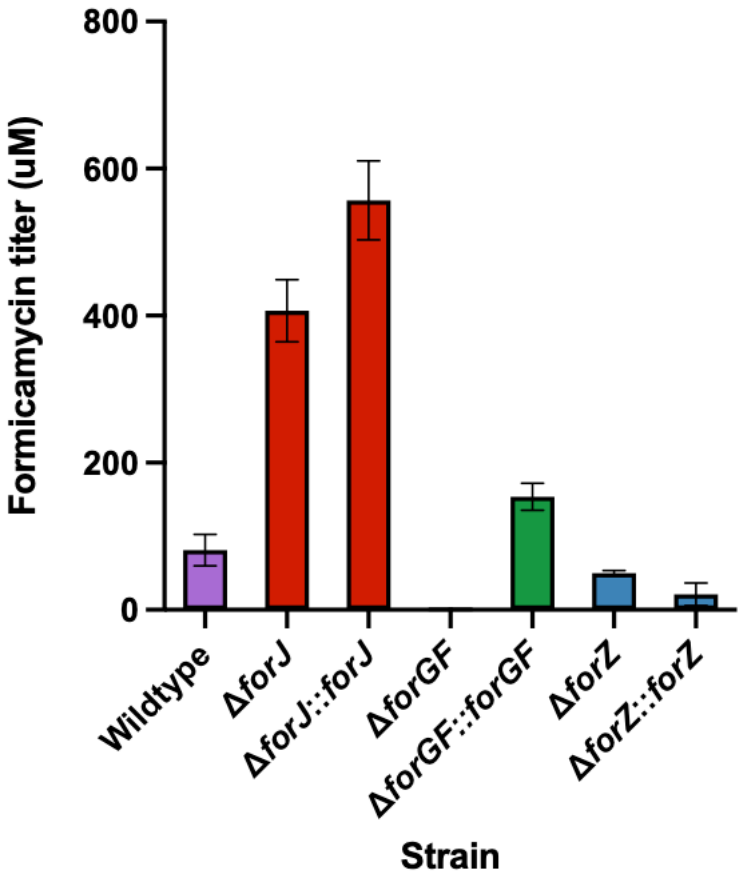


Figure 4.2 Titres of formicamycin production in μM between wildtype and regulator gene deletion strains alongside the complementation of these mutants. Values are the mean with error bars to represent standard deviation (wildtype n=16, mutants n=3). Adapted from Devine et al., 2021

Through this understanding of the transcriptional organisation, it was possible to next begin to evaluate the roles of the three cluster-situated regulators within the formicamycin BGC i.e. ForJ, ForGF, ForZ. This began with the generation of gene deletion mutants using CRISPR/Cas9 plasmid pCRISPomyces-2 followed by complementation of each gene deletion by reintroducing the gene *in trans* under the control of the native promoter. Each of these mutant strains were assessed in comparison to the wildtype strain to determine the impact of these regulators on formicamycin production (**Fig 4.2**). Throughout this thesis, where formicamycin production is measured in various mutants, data will be presented with a colour code based on the mutation in that strain: wildtype will be presented in purple, relating to the ForJ MarR regulator in red, relating to the ForGF two-component system in green and relating to the ForZ MarR regulator in blue. Following the loss of the MarR regulator ForJ, formicamycin production increased 5-fold in the Δ*forJ* strain compared to the wildtype, demonstrating its role as a repressor. In contrast, loss of the two-component system ForGF resulted in a complete loss of formicamycin production in the Δ*forGF* strain, deeming it an activator of the pathway. Finally, loss of the second MarR regulator ForZ, showed a reduction in formicamycin production to around 65% of wildtype levels in the Δ*forZ* strain, implying it is indirectly involved in the activation of formicamycin biosynthesis. Further to these data about transcriptional organisation and the overarching roles of the three cluster-situated regulators, ChIP-Seq was also performed to identify their binding sites. This revealed a multitude of binding peaks throughout the formicamycin cluster and several additional peaks throughout the rest of the genome (**Fig 4.3**). The binding sites of each regulator will be discussed in more detail later in this chapter. Whilst these data provide an invaluable insight into how these regulators may enact their control of the formicamycin biosynthetic pathway, they do not provide detail on their exact binding sites or impact on transcriptional activity. Transcriptional reporter assays utilise a reporter gene to provide a measurable indication of how active a promoter is under certain conditions and provide further detail on the mechanism of regulation. In this case, the β-glucuronidase gene (*gusA*) was fused to each of the promoters situated within the formicamycin BGC and

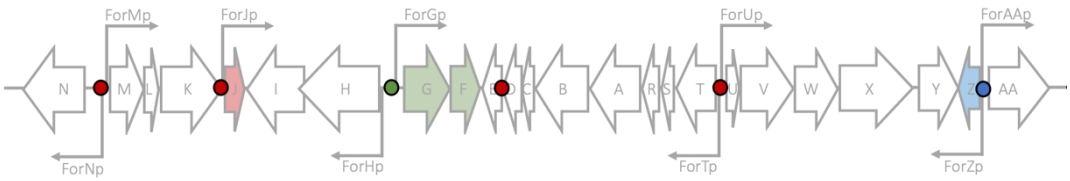


Figure 4.3 Cluster situated binding sites of the three formicamycin regulators. Binding sites of the main repressor ForJ are shown in red, the response regulator of the main activator ForGF in green and the additional MarR regulator ForZ in blue. Adapted from Devine *et al*., 2019

each construct was reintroduced into the native wildtype strain and the regulator deletion mutants. Through the hydrolysis of PNPG to galactose and chromophoric PNP, the activity of the β-glucuronidase gene product can be used as a measure of promoter activity once this substrate is added. The peak absorbance of this chromophore is observed at 415 nm and the intensity of the peak is dependent on the quantity of enzyme present, directly correlating with the transcriptional activity at the promoter driving *gusA* expression. By comparing the enzymatic activity at each promoter in the wildtype and the deletion mutants, it was possible to determine the effect of each regulator's binding on the transcriptional activity at each promoter.

4.2 The MarR Regulator, ForJ

After being established as the main repressor of the formicamycin biosynthetic pathway through the generation of *forJ* deletion mutants, it was decided to further investigate the binding and activity of ForJ. Deletion of *forJ* allowed for GUS activity driven by each promoter to be compared between the wildtype and Δ*forJ* strains and thus identify the effect of removing ForJ (**Fig 4.4**). A statistical increase in activity was seen in the absence of ForJ at six of the nine promoters: p*forJ*, p*forG*, p*forT*, p*forU*, p*forZ* and p*forAA*. This suggests

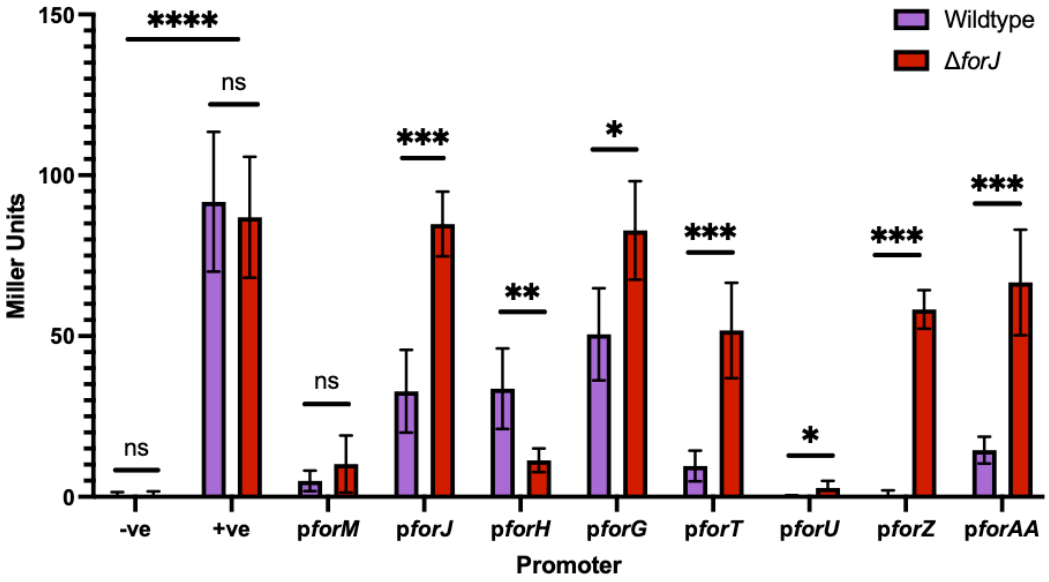


Figure 4.4 Colourimetric (β-glucuronidase) enzyme assay using PNPG to compare transcriptional activity of plasmid-borne *gusA*:promoter fusions between *Streptomyces formicae* wildtype and Δ*forJ* strains at day 4 of growth, n=27. Negative (pMF96) and positive (pMF23) control vectors lacking a promoter and under the control of *ermE** promoter, respectively, were used to confirm functional enzyme assay. Independent sample t-tests were performed within each promoter group to identify significant differences (**** $p < 0.001$, *** $p < 0.01$, ** $p < 0.025$, * $p < 0.05$) or no significant difference (ns).

that in wildtype *S. formicae*, ForJ exerts is repressor activity by binding across or near to these promoters and physically blocking RNA polymerase from transcribing the downstream genes. This is largely supported by the previously generated ChIP-Seq data, which identified ForJ binding sites in all of these promoters except the divergent p*forZ*/ p*forAA*.

To further understand the impact of ForJ binding, qRT-PCR was performed (**Fig 4.5**). These data aligned with the β-glucuronidase assay results and showed an increase in production of mRNA for transcripts under the control of all promoters with upregulated GusA activity in the absence of ForJ. As noted above, ChIP-Seq data identified ForJ binding sites in all of these promoters, with the exception of the divergent p*forZ*/p*forAA*. This indicates that there may be an alternative interaction involving ForJ that indirectly controls the regulation of *forZ* and *forAA*.

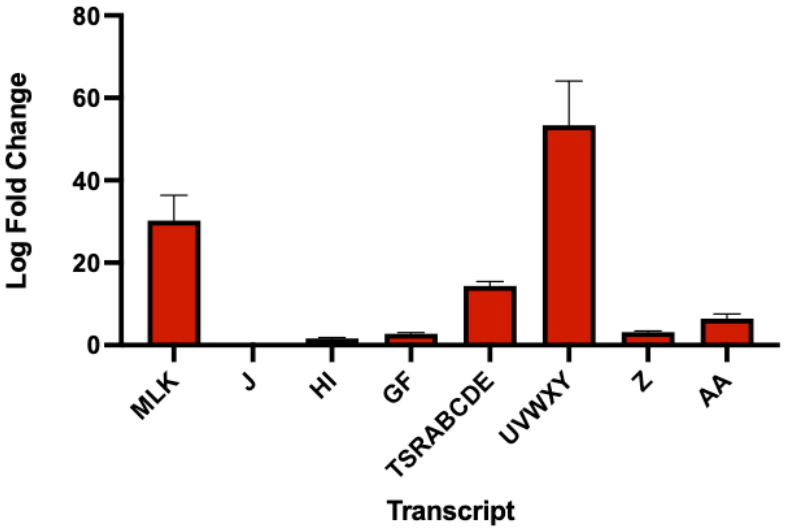


Figure 4.5 qRT-PCR analysis of mRNA production across the formicamycin biosynthetic gene cluster following the loss of the ForJ MarR regulator (main repressor of the pathway). All data shown are the log fold change in transcript levels relative to the wildtype.

When looking at the function of genes affected by the repression of ForJ, it is possible to gain an understanding of the impact ForJ has upon production of the formicamycin biosynthetic pathway. ForJ broadly represses the transcription of the formicamycin BGC, as indicated by the GUS and qRT-PCR data. This includes the core PKS genes (*forABC*) within the *forTRSABCDE* transcript, which are essential for the initiation of formicamycin biosynthesis and extending the polyketide chain, such that repression by ForJ prevents the pathway from the beginning. It has also been hypothesised that the expression of the biosynthetic genes used for post-PKS tailoring are also repressed until the polyketide chain has been synthesised. The genes encoding the three CSRs are all under the control of

promoters downregulated in the presence of ForJ, despite a lack of ForJ binding site in the promoter of ForZ. As described earlier, ForGF and ForZ are activators of formicamycin biosynthesis and transport, respectively, with their removal reducing the titres of formicamycin present within the cell. It is therefore logical that ForJ can repress their production in some way, and this is further supported by the observed increase in formicamycin production in the Δ*forJ* strain. The autorepression of promoter activity at p*forJ* (by ForJ) would prevent total pathway shutdown and allows fluctuation of regulator concentration within a narrow range, which is common among MarR family regulators (Grove, 2013).

A statistically significant reduction in promoter activity at p*forH* in the Δ*forJ* strain does not correspond with the current understanding of ForJ as a repressor, the function of the genes under its control, or the ChIP-seq and qRT-PCR data. The acetyl CoA carboxylase subunits encoded by *forHI* on a single transcript are key components in the initial steps of the formicamycin pathway. ChIP-seq has identified a ForJ binding site between the divergent *forHI* and *forGF* promoters and qRT-PCR has shown the mRNA levels of the *forHI* transcript to increase in the absence of ForJ which contradicts the GusA reporter assay. Similarly, the lack of significant change in the promoter activity at p*forM* does not align with existing knowledge and understanding of the pathway because ChIP-seq identified a binding site for ForJ at the divergent promoters p*forN* and p*forM* and qRT-PCR data shows an increase in mRNA levels of the *forMLK* transcript in the Δ*forJ* strain. These data suggest that this may be an anomaly and the investigations relating to p*forM* and p*forH* should be repeated to confirm whether the β-glucuronidase activity measured here is correct.

ChIP-seq data also identified a ForJ binding site in the coding region of *forE* which is part of the longest transcript in the cluster, *forTRSABCDE*, under the control of p*forT*. It is possible that ForJ bound to this site uses a roadblock mechanism to interfere with the progression of RNA polymerase along the DNA to inhibit its transcriptional activity. Elsewhere, ForJ has been shown to bind upstream of two other genes, KY5_3182, which encodes a putative MoxR-type ATPase, and KY5_5812, which encodes a hypothetical protein. The significance of these binding sites and functions of these gene products are unknown and have not been further investigated here.

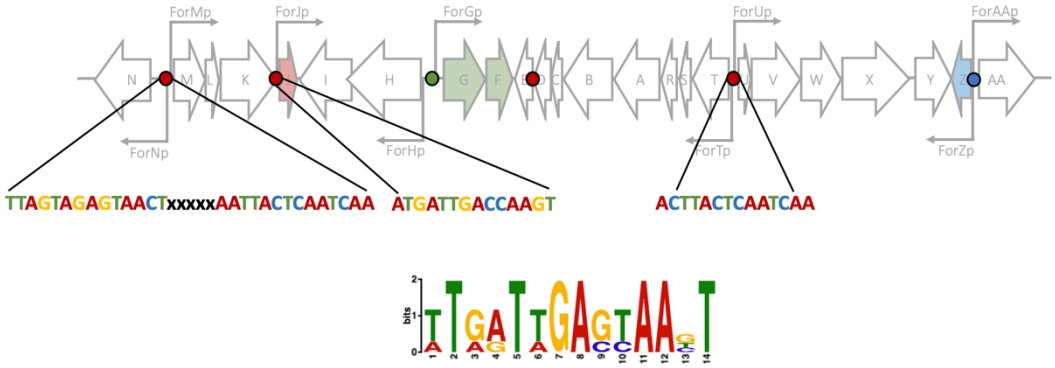


Figure 4.6 SPR identification of the ForJ binding sequence. A back-to-back binding sequence is seen in the *forM* promoter region, once in the *forJ* promoter region and again in the *forT* promoter region A 14-base pair consensus sequence was determined via MEME analysis. Figure adapted from Devine and Noble, 2024.

The final line of investigation into the ForJ repressor, carried out by Dr Rebecca Devine, was to identify the specific binding motif of this regulator (**Fig 4.6**). To achieve this, ReDCaT SPR was performed on the regions of DNA that showed enrichment peaks in the ChIP-seq data. Sequence specific binding of ForJ was only seen in the presence of a reducing agent, suggesting that this regulator is redox sensitive (Devine and Noble, in review). From the range of SPR probes designed, binding was observed at p*forM*, p*forJ* and p*forT* with subsequent alignment identifying a consensus sequence of a 14 base pair A/T rich area of DNA. This covers the transcriptional start sites of the *forMLK*, *forJ* and *forTSRABCDE* genes. In addition, the *forM* promoter contains two back-to-back ForJ binding sites which might allow dimeric binding for stronger repression. ForJ was shown to form multimers in response to oxidising conditions, through a single cysteine residue at position 68, perhaps allowing for stronger repression and may explain the lack of formicamycin production in liquid culture where the cells are more oxygenated. Oxidation is therefore believed to result in tighter binding to DNA which may also explain the presence of additional enrichment peaks seen in the ChIP-seq data as ForJ could be winding up the DNA.

In summary, the ForJ regulator is effective at repressing the formicamycin biosynthetic pathway through controlling the transcription of its biosynthetic enzymes contained within the BGC. It is capable of binding to the transcriptional start sites of a multitude of genes that encode essential components of the pathway and their repression results in complete shutdown.

4.3 The Two-Component System, ForGF

After being established as the main activator of the formicamycin biosynthetic pathway through the generation of Δ*forGF* deletion mutant strains, it was decided to further investigate the binding and activity of ForF, the response regulator of the ForGF two-component system. Deletion of *forGF* allowed for GUS activity driven by each promoter to be compared between the wildtype and Δ*forGF* strains to identify the effects of ForF binding (**Fig 4.7**). A statistical decrease in activity was seen in the absence of ForGF at three of the nine promoter sites: p*forG*, p*forH* and p*forAA*. This suggests that in wildtype *S. formicae*, the ForF response regulator exerts its activating activity by binding across or near to these promoters and enhancing transcription to mRNA by RNA polymerase. This is largely supported by the previously generated ChIP-seq data, which identified the intergenic region between the divergent *forHI* and *forGF* genes as a binding site for ForF but did not identify a ForF binding site between the divergent *forZ* and *forAA* genes.

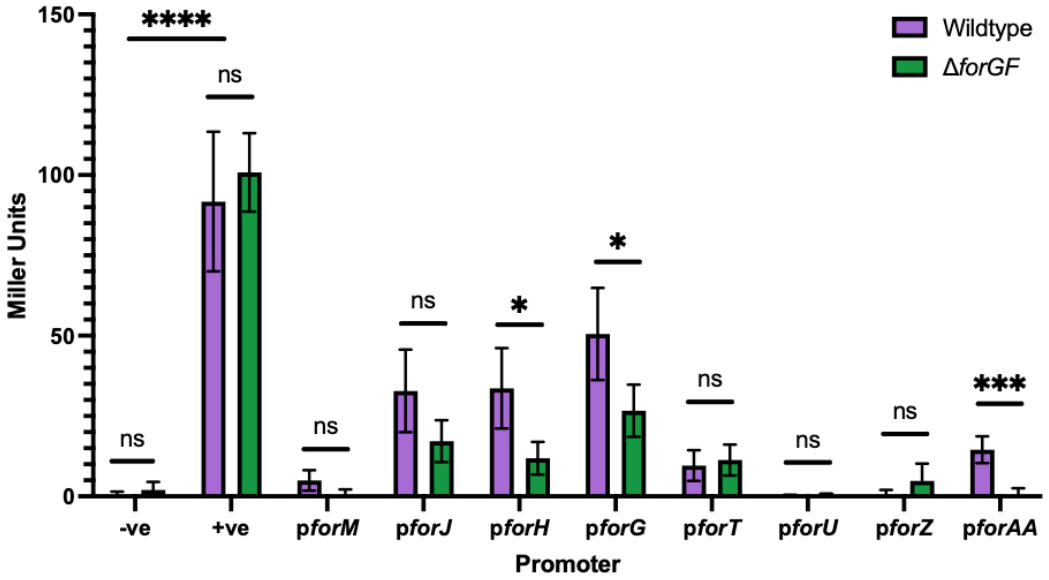


Figure 4.7 Colourimetric (β-glucuronidase) enzyme assay using PNPG to compare transcriptional activity of plasmid-borne *gusA*:promoter fusions between *Streptomyces formicae* wildtype and Δ*forGF* strains at day 4 of growth, n=27. Negative (pMF96) and positive (pMF23) control vectors lacking a promoter and under the control of *ermE** promoter, respectively, were used to confirm functional enzyme assay. Independent sample t-tests were performed within each promoter group to identify significant differences (**** $p < 0.001$, *** $p < 0.01$, ** $p<0.025$, * $p<0.05$) or no significant difference (ns).

To further understand the role of ForF, qRT-PCR was performed (**Fig 4.8**). These data aligned with the β-glucuronidase assay results and showed a decrease in production of mRNA for both the *forHI* and *forGF* transcripts in addition to the *forAA* transcript in the absence of ForGF. As noted above, the promoter of *forAA* does not have a ForF binding site, but there was a reduction in *forAA* promoter activity seen in the β-glucuronidase assay and a

corresponding reduction in *forAA* mRNA production in the qRT-PCR data in the absence for ForGF. This supports the idea of an indirect control mechanism regulating the expression of the transport gene *forAA*. No statistically significant changes in promoter activity or mRNA production were seen at the remaining promoter sites, but this was anticipated due to the lack of ForF enrichment peaks at these promoters in the ChIP-seq data.

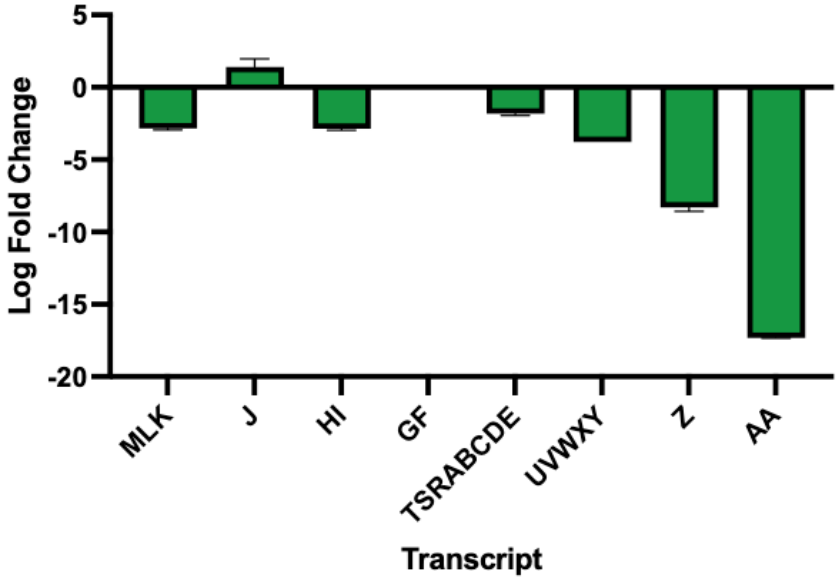


Figure 4.8 qRT-PCR analysis of mRNA production across the formicamycin biosynthetic gene cluster following the loss of the ForGF two-component system (main activator of the pathway). All data shown are the log fold change in transcript levels relative to the wildtype.

The ChIP-seq data also show that ForF binds upstream of KY5_0375 which encodes a putative NLPc/P60 family protein but is not within any currently predicted BGCs. The significance of this binding site is unknown but was confirmed using ReDCaT SPR as below.

An additional line of investigation into the ForF activator, was to identify the specific DNA binding sequences recognised by this response regulator. To achieve this, ReDCaT SPR was performed on the areas of DNA that showed enrichment peaks identified from the ChIP-seq data. The screened sites included the divergent promoters of *forHI* and *forGF* and the binding site upstream of KY5_0375. This identified one hit within the formicamycin BGC and two hits in the KY5_0375 intergenic region (**Fig 4.9**). All three of these were carried forward to the foot-printing stage where the DNA sequences were progressively truncated by two base pairs until binding was lost to identify the exact binding motif recognised by ForF (**Fig 4.10**). Across all SPR experiments, initial screening is carried out to identify appropriate concentrations of the protein being investigated with the intention of identifying a working concentration that results in an RMax % value as close to 100% for positive hits. Unfortunately, ForF proved to be unstable when diluted to concentrations lower than 500 nM, therefore RMax % values across these data are often much higher than 100% as

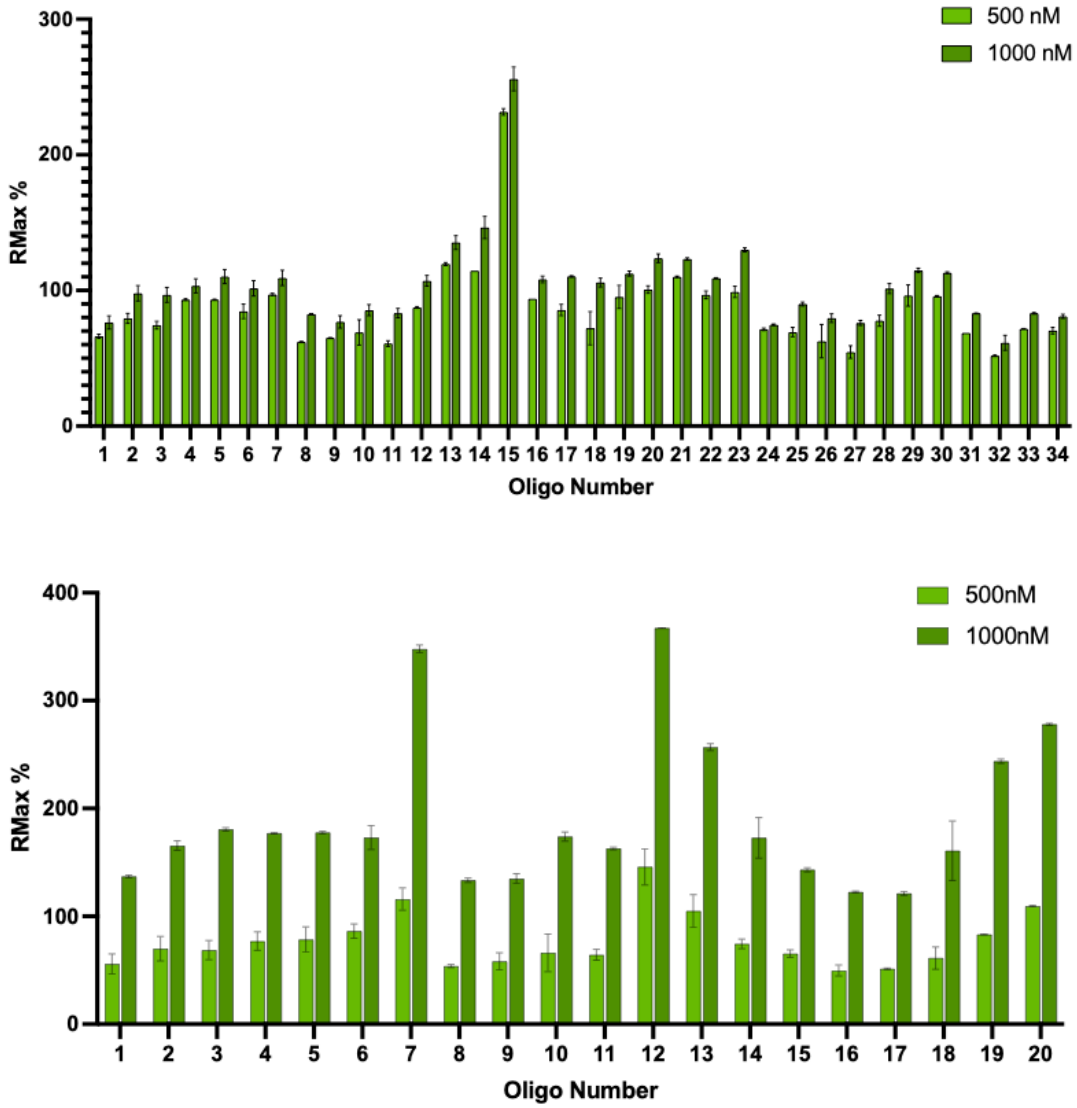


Figure 4.9 Initial SPR screening of the ChIP-seq binding enrichment peaks of ForF. The peaks across the divergent promoter p*forHI* and p*forGF* (top) and upstream of KY5_0357 (bottom) were split into 40 base pair long overlapping oligos, numbered sequentially. Hits were seen in oligo 15 of the divergent promoter and oligos 7 and 12 for the intergenic region.

the DNA is fully enriched with an inability to reduce the protein concentration being tested. Subsequent alignment identified a consensus sequence of a 12- base pair long DNA sequence which is consistent with the ChIP-seq data (**Fig 4.11**). Manual inspection of ChIP-seq data identified three other potential binding sites of ForF throughout the *S. formicae* genome that were only just below the previously used significance cut off, however none of these showed any indication of ForF binding from initial SPR analysis (data not shown) and did not contain the consensus sequence identified from sites where ForF binding was seen. Further searches of the entire *S. formicae* genome using the established binding motif did not identify any additional binding sites of relevance throughout the rest of the genome.

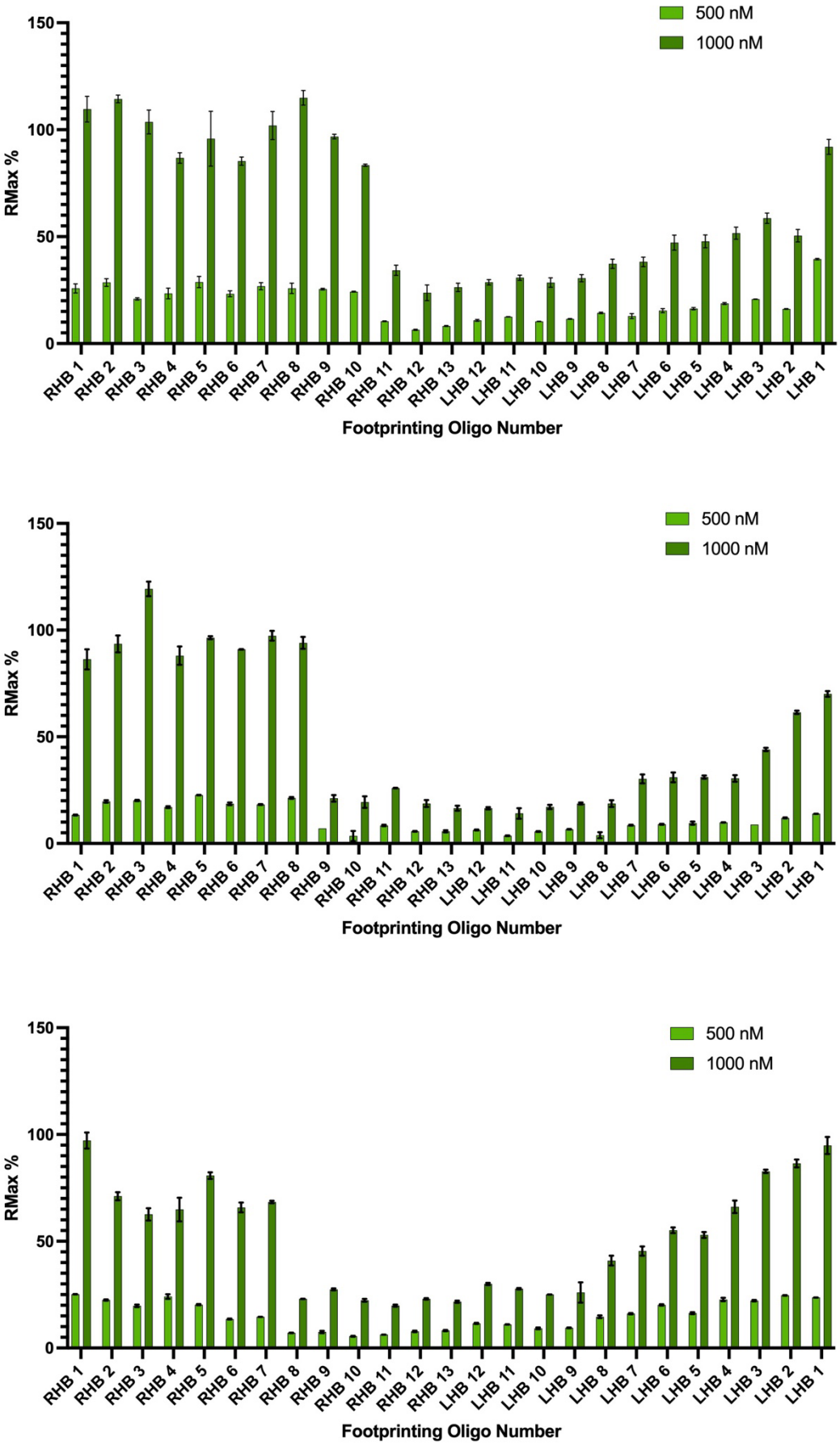


Figure 4.10 SPR foot-printing of the initial hits of ForF binding. The one hit from the divergent *forHI* and *forGF* promoters in the formicamycin promoter (top) showed drop off at RHB 10 with binding returning at LHB 1. The first hit from the intergenic region (middle) showed drop off at RHB 8 with binding returning at LHB 3 while the second hit (bottom) showed drop off at RHB 7 within binding returning at LHB 4.

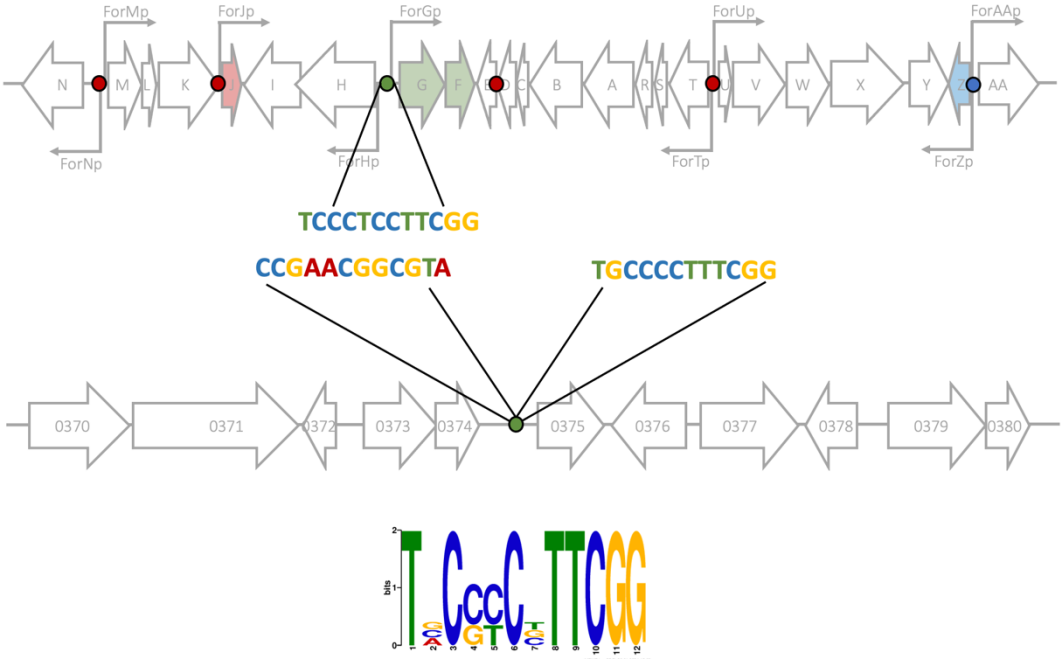


Figure 4.11 SPR identification of the ForF binding sequence within the formicamycin biosynthetic gene cluster identified as a 12 base pair motif across the divergent promoters p*forHI* and p*forGF*.

The results described above show that the ForGF TCS is effective in activating the formicamycin biosynthetic pathway, presumably by controlling the transcription of the *forHI* genes, which are required to boost production the malonyl CoA starter units of this pathway. ForF is capable of binding directly between the transcriptional start sites of these divergent promoters, resulting in autoregulation and the upregulation of *forHI*, potentially through the recruitment of enhancers.

4.4 The MarR Regulator, ForZ

After being established as a partial activator of the formicamycin biosynthetic pathway through the generation of Δ*forZ* mutant strains, it was decided to further investigate the binding and activity of ForZ. Deletion of *forZ* allowed for GUS activity driven by each promoter to be compared between the wildtype and Δ*forZ* strains to determine the effect of removing ForZ (**Fig. 4.12**). A statistical decrease in activity was seen in the absence of ForZ at two promoter sites: p*forJ* and p*forH*. However, a statistical increase in promoter activity was also seen in the absence of ForZ at another two promoter sites: p*forU* and p*forZ*. The increase at p*forU* is considered to be an anomaly here as the increased promoter activity is comparable to that of the negative control.

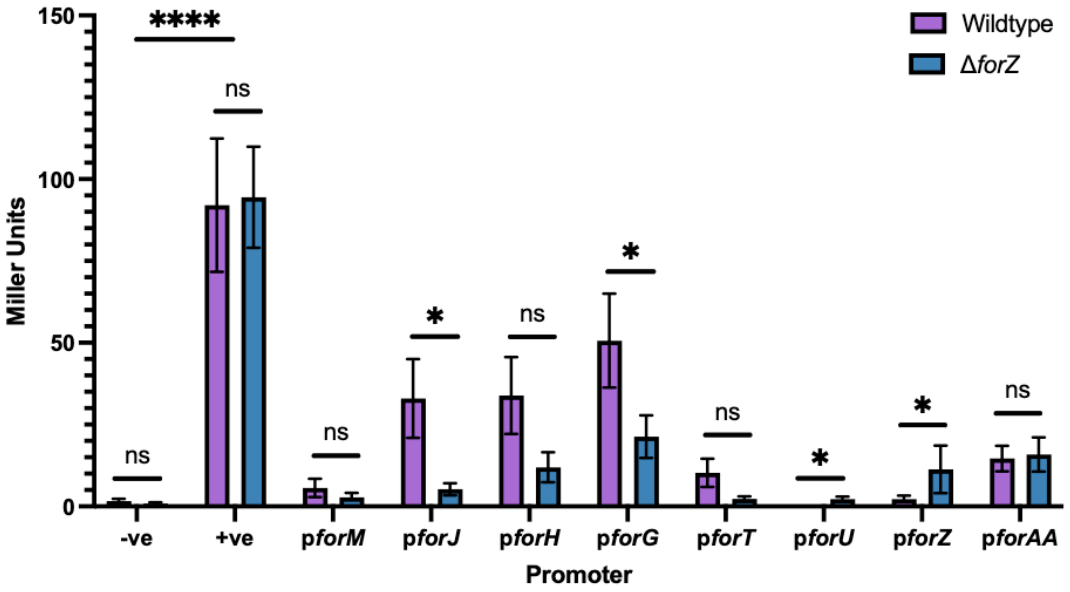


Figure 4.12 Colourimetric (β-glucuronidase) enzyme assay using PNPG to compare transcriptional activity of plasmid-borne *gusA*:promoter fusions between *Streptomyces formicae* wildtype and Δ*forGF* strains at day 4 of growth, n=27. Negative (pMF96) and positive (pMF23) control vectors lacking a promoter and under the control of *ermE** promoter, respectively, were used to confirm functional enzyme assay. Independent sample t-tests were performed within each promoter group to identify significant differences (**** $p < 0.001$, *** $p< 0.01$, ** $p<0.025$, * $p<0.05$) or no significant difference (ns).

To further understand the impact of ForZ binding on the transcription of genes throughout the BGC, qRT-PCR was performed (**Fig 4.13**). These data added to the complexity of understanding this regulator, with confirmation of a reduction in *forGF* promoter activity and *forGF* transcripts in the absence of ForZ, despite a lack of any ForZ binding peak upstream of *forGF* in the ChIP-seq data. However, where the β-glucuronidase assay identified no significant change in promoter activity at p*forM* or p*forAA*, qRT-PCR data

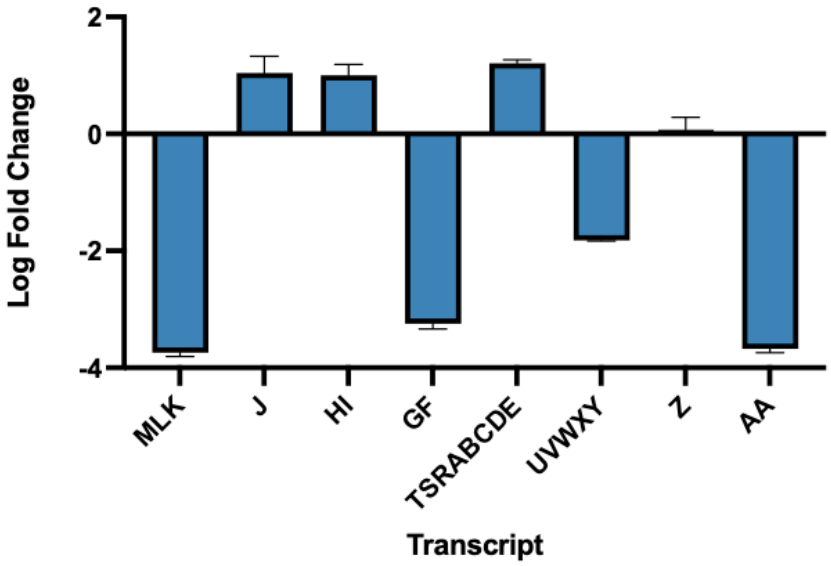


Figure 4.13 qRT-PCR analysis of mRNA production across the formicamycin biosynthetic gene cluster following the loss of the ForZ MarR regulator. All data shown are the log fold change in transcript levels relative to the wildtype.

showed a significant reduction in transcription of both the *forMLK* and *forAA* transcripts in the absence of ForZ. Similarly, while the β-glucuronidase assay showed a reduction in *forJ* and *forZ* promoter activity, the qRT-PCR data showed no significant changes in the *forJ* or *forZ* mRNA levels in the absence of ForZ. Despite this, these data suggest ForZ activates two promoters while repressing a third, despite not having binding sites at all of these locations. This activation and counter-repression activity has been observed before for MarR family regulators in other *Streptomyces* strains, some of which are also involved in the regulation of antibiotic biosynthetic pathways (Hünnefeld et al., 2019; Q. Zhang et al., 2015). It is possible there is cross-talk occurring between the three regulators of the *for* BGC, or alternatively a run-over impact of the previously mentions DNA-winding by ForJ. A similar process has been seen in *Streptomyces avermilitus* where a MarR regulator is able to indirectly alter the expression of a cluster situated regulator while directly upregulating its own expression (Guo et al., 2018). There may also be negative or positive feedback loops, with the production, or lack of, of fasamycins and formicamycins affecting the activity of the regulators and thus the expression of genes in the *for* BGC.

To investigate this further, Dr Rebecca Devine used ReDCaT SPR to identify the specific binding motif recognised by ForZ (**Fig 4.14**). To achieve this ReDCaT SPR was performed on the area of DNA that showed enrichment peaks identified from ChIP-seq data. Sequence specific binding of ForZ was observed against a single DNA probe in the form of an 8 base pair repeat sequence that covers the transcriptional start site of the *forAA* gene. This is consistent with the ChIP-seq data and suggests that ForZ exerts its regulatory effects by binding to this site and repressing the expression of the *forAA* MFS transporter gene by occluding the transcriptional start site, and thus blocking RNA polymerase. As there are several well documented examples of MarR regulators being regulated by the end products of the pathways they control, this line of investigation was pursued for ForZ. It was shown that the binding of ForZ to the *forAA* promoter is abolished in the presence of

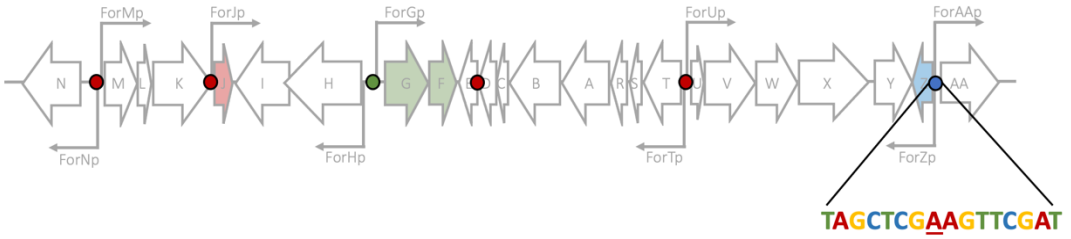


Figure 4.14 SPR identification of the ForZ binding sequence identified as an 8 base pair palindromic repeat across the *forZ* and *forAA* promoter regions. The A underlined in the binding motif is the transcription start site of *forAA*. Figure adapted from Devine and Noble, 2024.

formicamycins, and to a lesser extent fasamycins (Devine and Noble, in review). This provides a feedback mechanism such that when there are high levels of formicamycins in the cell, they bind to ForZ and inactivate its DNA binding activity to induce expression of the *forAA* promoter, and thus reducing the level of formicamycins by exporting them from the cell. This is likely also a self-resistance mechanism to ensure that the cell does not succumb to the toxic effects of the natural product it is producing.

4.5 Conclusions

Throughout this chapter, alongside the collaborative work in the Hutchings laboratory, the roles of the three cluster situated regulators in controlling formicamycin biosynthesis have been established (**Fig 4.15**). The two-component system ForGF is the primary activator of the pathway, with the response regulator binding to the divergent promoter responsible for controlling the transcription of *forGF* as well as *forHI*. While the activating signal for ForG is still under investigation, the autoregulation of ForGF enables a positive feedback mechanism to drive formicamycin biosynthesis. The upregulation of *forHI* enables the production of the ACC biotin carboxylase and ACC carboxyl transferase encoded by these genes that presumably boost the levels of the malonyl CoA starter unit of the formicamycin biosynthetic pathway. The MarR-family regulator ForJ is the primary repressor of the pathway, with binding sites across the BGC allowing it to physically block the binding of RNA polymerase, preventing transcription of the biosynthetic genes and the production of the core biosynthetic machinery needed to facilitate the production of formicamycin. The

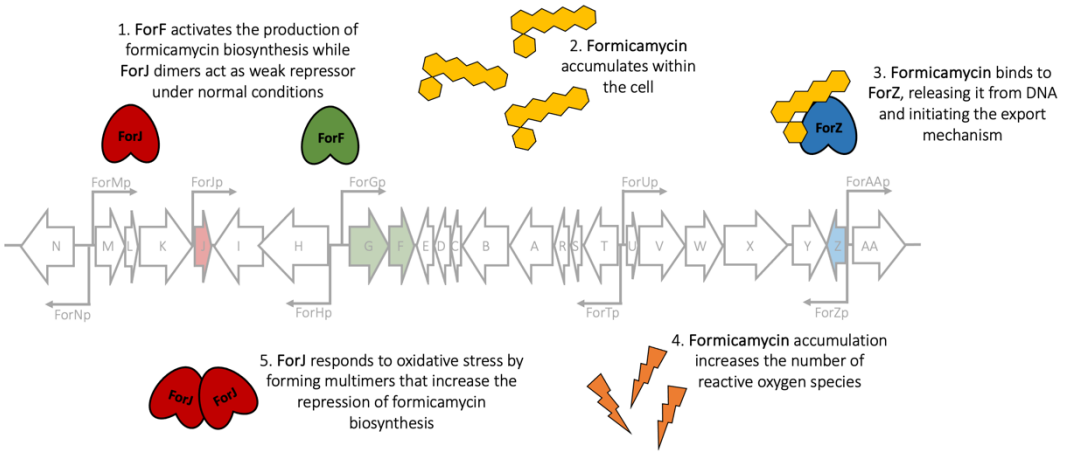


Figure 4.15 An overview of the regulatory network acting upon the formicamycin biosynthetic pathway mediated by the cluster situated regulators ForJ, ForGF and ForZ. Figure adapted from Devine and Noble, 2024.

formation of ForJ multimers in response to oxidising conditions results in tighter DNA binding and stronger transcriptional repression and ultimate downregulation of formicamycin biosynthesis (Devine and Noble, in review). This is further supported by the fact that the biosynthetic pathway responsible for producing the formicamycins has been shown to increase oxidative stress within the cell, meaning this pathway is shut down in response to such oxidative conditions to prevent harm being caused to the cell and the accumulation of these polyketides (Devine and Noble, in review). Alongside this, the role of ForZ has been determined and can explain the changes to formicamycin titres in response to loss of this MarR regulator. Under normal conditions, ForZ binds to the divergent promoter of *forZ* and *forAA*, repressing their production by physically blocking the binding of RNA polymerase and preventing their transcription to mRNA. However, once formicamycins accumulate they bind to ForZ, reducing its DNA binding capabilities and releasing it from this divergent promoter. This allows for increased transcription of *forAA* meaning formicamycins can be exported from the cell and once levels reduce again, ForZ DNA binding activity returns, and it once again represses *forAA*.

The formicamycins have already been established as a novel structural class of antibiotics with potential for clinical application. Their ability to inhibit the growth of clinically relevant pathogens with no evolution of resistance makes them promising for further development. The work outlined in this chapter demonstrates the complexity of the regulatory pathways controlling antibiotic biosynthesis, in this case involving multiple cluster-situated regulators that exemplify the depth of cross-regulation involved in the biosynthesis of secondary metabolites in *Streptomyces* species. Deciphering the action of key regulators and identifying the role of specific genes within the biosynthetic pathway is crucial in developing a strain that is capable of overproduction and export. In addition, an understanding of novel regulatory mechanisms used to regulate the biosynthesis of novel secondary metabolites may provide further insights into unlocking silent BGCs.

5. The Structure, Function, and Interaction of ForGF

Two-component systems are noted as being one of the most important regulatory elements involved in signal transduction in *Streptomyces* species. As briefly discussed earlier, most species contain dozens of these systems that are used to elicit an adaptive response to environmental stimuli, most notably the biosynthesis of natural products. When put into context with their role in the development and life cycle of *Streptomyces* species, TCSs are arguably one of their most important regulatory elements. The role and functions of typical TCSs across the bacterial community have been elucidated, but there are of course exceptions to how these systems are employed in different organisms.

A typical TCS includes a membrane bound HK and a cytoplasmic RR with each having a fixed role in terms of biochemical activity. The kinase is responsible for sensing the signal which triggers a conformational change that activates its kinase domain to enable autophosphorylation at a conserved histidine residue using ATP (Cruz-Bautista et al., 2023). The cognate RR catalyses the transfer of that phosphate to its own receiver domain. This typically induces a conformational change, often resulting in dimerisation, to release the DNA binding helix, such that the RR can bind to the target DNA to elicit a response to the original signal. In nature, a great degree of diversity can be found at almost every stage of this process, with a diverse range of chemical and physical input signals requiring equal accommodation in the domains within the HK responsible for their recognition. Equally, the RR is able to elicit responses through protein-protein, protein-RNA or protein-DNA interactions which themselves require different molecular determinants. However, signal transduction through autophosphorylation and phosphorelay to the RR remains the central mechanism of this process and very much relies on highly conserved domains (Goulian, 2010; Zschiedrich et al., 2016).

Recent developments in structural and biochemical studies have supported the investigation of the divergences from the typical two-component system, with particular focus on the molecular mechanisms of these outliers. A key aspect of this is elucidating the structure of the systems though purification to characterise the domains and hypothesise on their mechanism of action based on the predicted biochemical activity of each of these units. It is, however, still essential to gain further insight into the molecular mechanisms of the TCSs through *in vivo* studies.

An understanding of the binding, and general regulation of the formicamycin BGC by ForGF has been established through the previous chapters. Each stage of the TCS action can be investigated individually to provide an understanding of the stepwise mechanism through which activation is achieved. However, the signalling events that result in the activation of ForG remain unknown and its elucidation is one of the more significant challenges in this project. For most TCSs, an incredibly wide range of molecules have been known to trigger autophosphorylation in HKs. This chapter will explore the structures of the ForGF TCS and investigate how this impacts its interaction with other molecules in the cell and ultimately how this affects formicamycin production.

5.1 Structural Modelling and Purification

In the absence of X-ray crystallography data to determine the structure of a protein, modelling can be used to predict the 3D structure based on other reported proteins with similar sequence identities. Significant advances have been made in the performance and accessibility of artificial intelligence-based platforms, with the removal of requirements for these to be hosted locally, greatly increasing the ability of the end user to understand their protein without crystal structures. Whilst this is not always as accurate, especially when the protein of interest has very few orthologues with similar sequence identity or coverage that have had their structures elucidated, it can be useful in the initial stages of investigation or where protein crystal structure cannot be obtained.

Here, AlphaFold-2 was used to predict the 3D structures of the HK ForG and the cognate RR ForF (Jumper et al., 2021). Neither component has more than 31% sequence identity with any other previously reported structures, however, is it still possible to interpret a significant amount of functionality based on these molecular predictions and that of other software such as HMMTOP.

5.1.1 The Sensor Kinase, ForG

ForG is predicted to be unusual for a signal detecting HK, with a distinct lack of any transmembrane domains as predicted from primary sequence analysis (**Fig 5.1**). There is a reasonable degree of confidence in this model's predicted structure, considering the low sequence identity, and all of the amino acids present in the polypeptide have been included in this prediction (**Appendix Figure 1**). This would suggest that the sensor domain of this kinase sits in the cytoplasm of the cell rather than outside the cell. This provides additional intrigue as to the activating signal of this TCS, implying that it must either arise from within the cell or be a compound easily transported across the membrane. The remaining domains include an ATPase which facilitates the catalysis of ATP to ADP, freeing up the phosphate that will be used for autophosphorylation. Near to this is the DHp domain that contains the conserved histidine residue at position 175, which is the receiver of this phosphate. The domains predicted as part of this AlphaFold-2 structure align with those predicted to be present by HMMTOP analysis, further supporting the accuracy of the model. ForG is also

predicted to exist as a homodimer, which are seemingly facilitated through the coiled coils of the N-terminus.

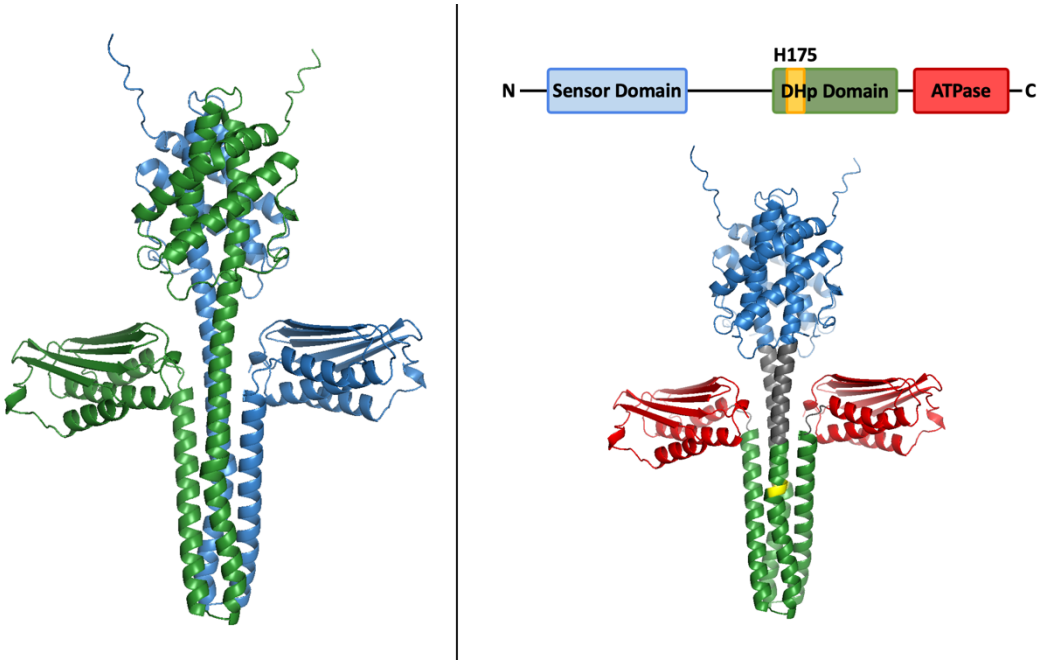


Figure 5.1 AlphaFold-2 modelling prediction of the structure of the histidine kinase ForG. The left-hand image shows its homodimeric form with monomer A in green and monomer B in blue. The right-hand image shows the predicted functional domains contained within the kinase. This includes the sensor domain (blue), DHp domain (green) and ATPase domain (red), also shown in yellow is the site of the conserved histidine residue that autophosphorylates on receipt of a signal.

In order to investigate the role and molecular function of ForG, purification attempts were made using a variety of *forG* overexpression conditions and vectors. Optimum expression conditions were identified with the use of a codon optimised sequence for expression in an *E. coli* host with the use of the pET29a(+) vector, induced with 200 μM IPTG and grown overnight at 18 °C. Initial trials indicated that the kinase was soluble, confirming predictions of this protein lacking any transmembrane domains. Purification was confirmed using SDS-PAGE and immunoblotting to identify specific protein bands that were tagged with the 6xHis epitope at the N- or C- terminus, when using the pET28a(+) and pET29a(+) vectors respectively, used during purification (**Fig 5.2**). Were this protein to contain any transmembrane domains, the purification attempts would either have shown a band at 37 kDa in the insoluble fraction or a smaller band size in the soluble and insoluble fractions as a result of the protein separating into two or more fragments. Transmembrane domains would also exclude the possibility of a positive signal being seen from immunoblotting of the soluble fraction when using pET28aa(+). This is due to the 6xHis epitope being attached to the N-terminus, so even if the protein had fragmented and predominant bands were seen

in the soluble fraction, only the N-terminal would contain the epitope that the antibodies bind to. Therefore, the combination of predominant bands at the expected molecular weight of 40 kDa in the soluble fraction alongside a positive immunoblot of this band confirmed that ForG does not contain transmembrane domains and likely exists cytosolically. Further analysis using mass spectroscopy with the help of Dr Carlo Martins (JIC proteomics), confirmed that the ForG is the band that could be predominantly seen in the soluble induced fraction, despite background contamination in the crude analysis (data not

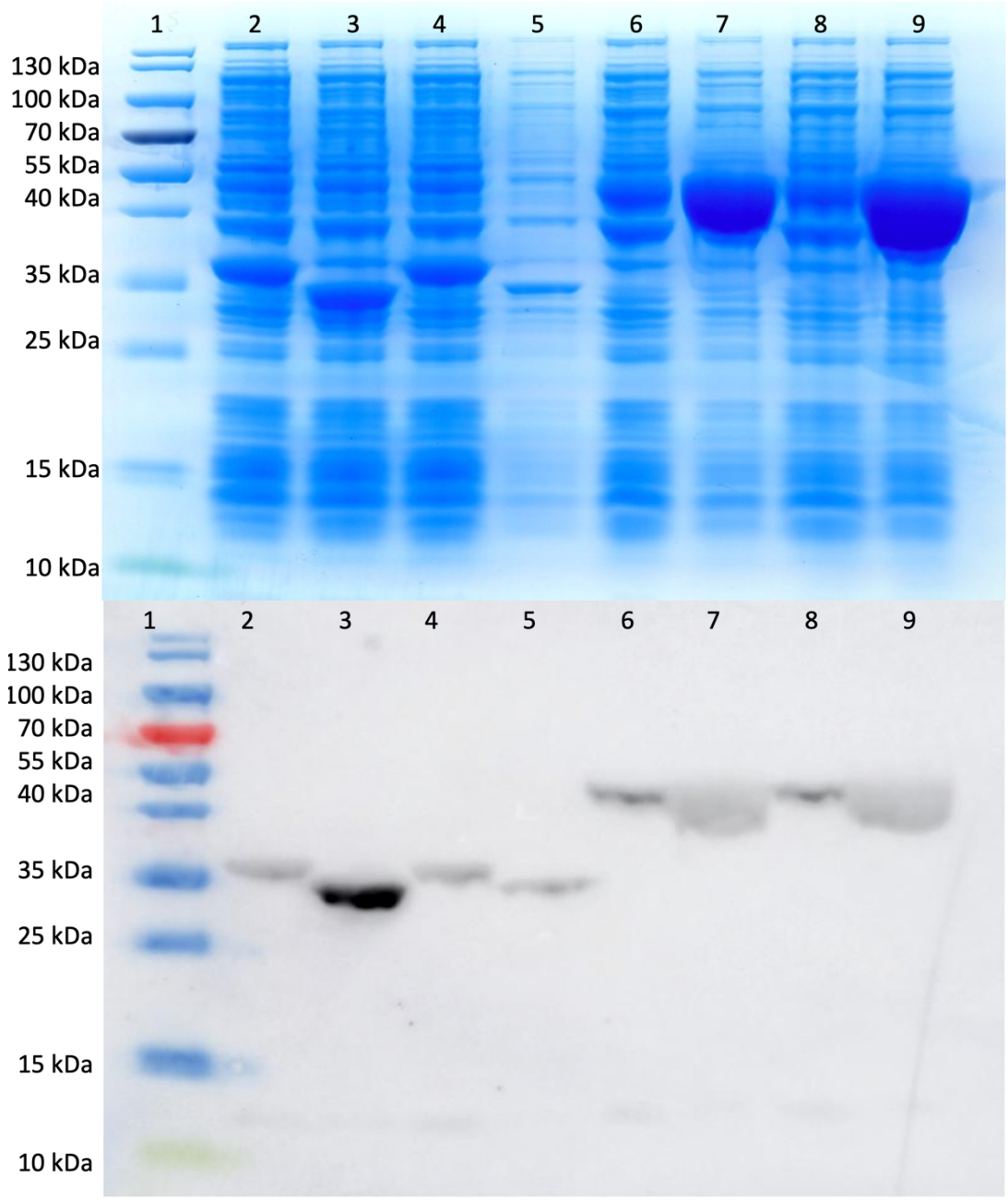


Figure 5.2 SDS-PAGE (top) and Western blot (bottom) analysis of crude samples of ForF and ForG to test overexpression conditions. Lanes from left to right: (1) Page Ruler Plus Ladder, (2) NiCo21 pET28a ForF soluble 200 mM IPTG, (3) NiCo21 pET29a ForF soluble 200 mM IPTG, (4) NiCo21 pET28a ForF soluble 500 mM IPTG, (5) NiCo21 pET29a ForF soluble 500 mM IPTG, (6) NiCo21 pET28a ForG soluble 200 mM IPTG, (7) NiCo21 pET29a ForG soluble 200 mM IPTG, (8) NiCo21 pET28a ForG soluble 500 mM IPTG, (9) NiCo21 pET29a ForG soluble 500 mM IPTG.

shown). It is important to note that during this purification process, significant background contamination was identified when the BL21 overexpression strain was used, so the final host used for all purification was NiCo21. This strain contains significantly fewer nickel-binding contaminants as the majority are either deleted or tagged with chitin-binding protein tags, meaning that less background contamination was seen in the purification process when using chitin column followed by a nickel column. Following the success of this overexpression in the NiCo21 host, the ForG samples were processed using fast pressure liquid chromatography (FPLC) in the form of an AKTA Pure (**Appendix Figure 3**). Once again, this showed great success with both purification and size exclusion resulting in the production of relatively pure protein at an average concentration of 3 mg/ mL allowing further *in vitro* investigations to be carried out.

In attempts to determine the actual structure of ForG, several sitting drop crystallisation trials were set up. Obtaining protein of a suitable concentration to continue these trials was relatively difficult, with the protein regularly precipitating when concentrated above 5 mg/ mL. Through a range of buffer trials, a suitable sample of ForG was concentrated to 9.5 mg/ mL and although screens with these samples have been unsuccessful in generating protein crystals, efforts are ongoing.

5.1.2 The Response Regulator, ForF

ForF is predicted to be a typical RR with a LuxR-type binding domain. Modelling predicts it is composed of a helix-turn-helix DNA binding domain at the C-terminus with a highly conversed receiver domain at the N-terminus that consists of five β-pleated sheets interspaced with another five α-helical structures (**Fig 5.3**). Within the receiver domain is a conserved aspartate residue at position 53, which likely accepts the phosphate group from the histidine residue in the cognate sensor kinase ForG once it has been autophosphorylated. Between the receiver and DNA binding domains is a flexible arm of 15 amino acids, it is likely that these change the conformation of the dimer between the phosphorylated and unphosphorylated states of ForF to change its DNA binding capabilities. For the majority of RRs, the phosphorylated form is the active form that enables DNA binding and either the repression or activation of the target genes that it controls. However, in some cases the phosphorylated state is actually the inactive form. This will be investigated at a later stage in this chapter.

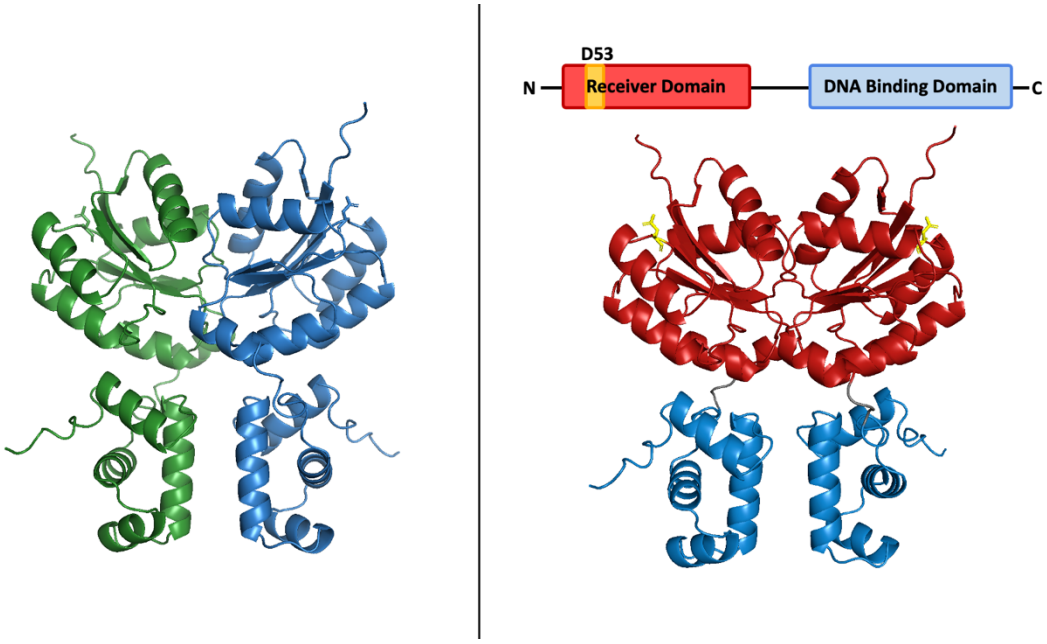


Figure 5.3 AlphaFold-2 modelling prediction of the structure of the response regulator ForF. The left-hand image shows its homodimeric form with monomer A in green and monomer B in blue. The right-hand image shows the predicted functional domains contained within the regulator. This includes the receiver domain (red) and DNA binding domain (blue) also shown in yellow is the site of the conserved aspartate residue that receives the phosphate group from the activated cognate histidine kinase, ForG.

In order to complete *in vitro* investigations into the role and molecular function of ForF, purification attempts were made using a variety of overexpression conditions and vectors. Optimum expression conditions were identified with the use of codon optimised sequences for expression in an *E. coli* host with the use of the pET29a(+) vector, induced at with 500 μM IPTG and grown overnight at 18 °C. Initial trials indicated that the RR was soluble, as expected of a protein such as this, which would normally act in the cytosol and lack any transmembrane domains. Purification was confirmed using SDS-PAGE and immunoblotting to identify specific protein bands that were tagged with the 6xHis epitope at the N-terminus used for purification (**Fig 5.2**).

Further analysis using mass spectroscopy with the help of Dr Carlo Martins (JIC Proteomics), confirmed that the ForF was the band that could be predominantly seen in the soluble induced fraction, despite background contamination in the crude analysis (data not shown). It is important to note that during this purification process, significant background contamination was identified when the BL21 overexpression strain was used, so the final host used for all purifications was NiCo21. Following the success of this overexpression in the NiCo21 host, the ForF samples were processed using fast pressure liquid chromatography (FPLC) in the form of an AKTA Pure (**Appendix Figure 4**). Once again, this

showed great success with both purification and size exclusion resulting in the production of relatively pure protein at an average concentration of 7 mg/ mL allowing further *in vitro* investigations to be carried out.

In attempts to determine the actual structure of ForF, several sitting drop crystallisation trials were set up. Through a range of buffer trials and concentration efforts a suitable sample of ForF was concentrated to 15 mg/ mL and although screens with these samples have been unsuccessful in generating protein crystals, efforts are ongoing.

5.2 The Sensing Domain of ForG

As previously discussed, identifying the activating signal(s) of HKs is notoriously difficult, with known examples including changes to temperature, light, small ligands such as amino acids or ions, and nutrient concentration (Ishii & Eguchi, 2021). The variety of input signals that these kinases are capable of responding is testament to the diversity of their sensor domains which can contain broad combinations of alpha helical and β-pleated sheet structures to accommodate their specific signal. Comparison of the sensor domain of ForG to other known kinases did not provide any further indication as to its activating signal. BLAST analysis of the sensor domain showed a relatively low homology with any other kinases and in addition, none of these hits appear to have any existing data about their

Table 5.1 Summary of BLAST analysis of the sensor domain of the sensor kinase ForG.

Description	Strain Name	Query Cover	% Identity	Accession Number
Sensor histidine kinase	*Streptomyces formicae*	100	100	WP_098245755.1
Sensor histidine kinase	*Streptomyces kanamyceticus*	100	97.33	WP_169801191.1
ATP binding protein	*Streptomyces morookaense*	92	49.29	WP_171078762.1
ATP binding protein	*Streptomyces* sp. ET3-23	92	48.57	WP_227725535.1
Sensor histidine kinase	*Streptomyces* sp.	92	46.10	QKO28691.1
ATP binding protein	*Streptomyces niveiscabiei*	100	39.33	WP_319735451.1

structure, function or interaction with other molecules within their respective producing organism (**Table 5.1**). In addition to this, analysis of the surface structure of the AlphaFold-2 model predictions shows multiple potential binding pockets across the sensor domain with varying depths and widths that would be able to facilitate a huge number of ligands, again providing no further indication as to the activating signal (**Fig 5.4**).

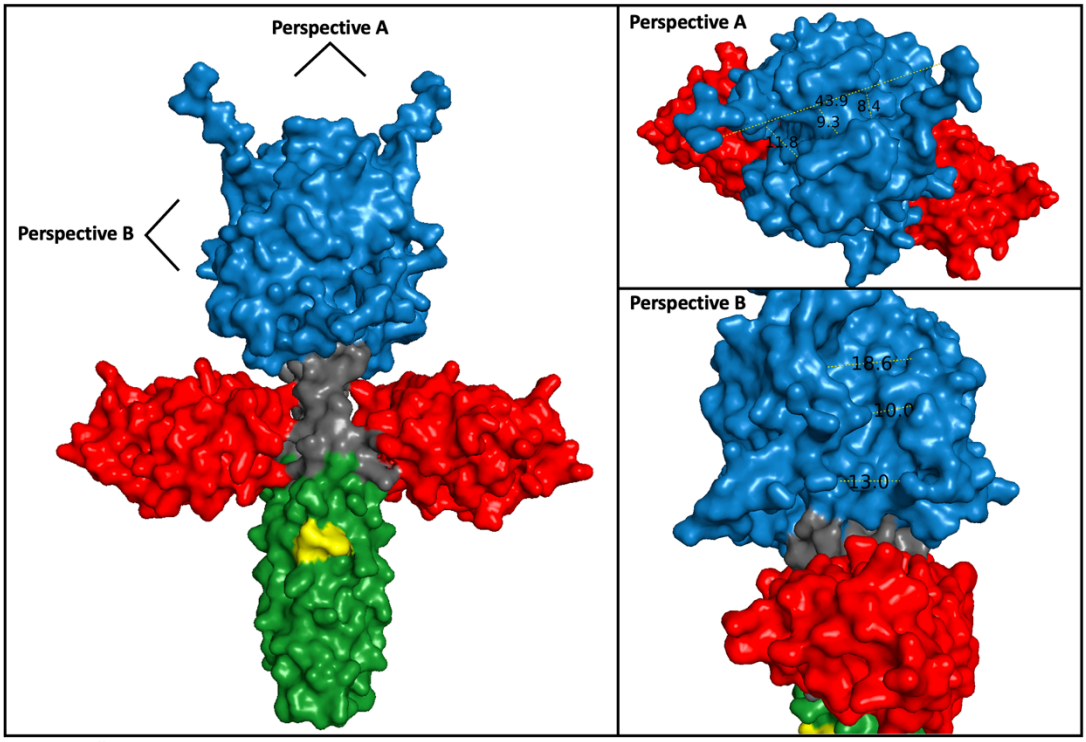


Figure 5.4 Visualisation of the surface of the ForG dimer model predicted by AlphaFold-2. The key domains are distinguished by colour include the sensor domain (blue), the DHp domain (green), the conserved histidine residue at position 175 (yellow) and the ATPase domain (red). The distances between amino acids across the sensor domain are shown in Angstroms (Å) from the above and side perspectives and have been calculated using the PyMOL measurement tool.

To begin *in vivo* investigations into the source of the signal that triggers the formicamycin biosynthesis, the previously generated Δ*forG* strain of *S. formicae* was complemented with a copy of the same gene containing a 3xFLAG tag to facilitate co-immunoprecipitation. The aim was to identify protein-protein interactions that occur *in vivo* with the hope of using these data to better understand how ForG functions within the cell. It is possible that another protein acts as a ligand for this sensor kinase and triggers the autophosphorylation of the conserved aspartate residue situated in the DHp domain. A total of 72 proteins were found to be in abundance below the Bayesian false discovery rate (BFDR) of 0.2 in comparison to the wildtype control. These were further sorted by enrichment, with further investigation being carried out on any proteins that has a greater than 10-fold increase, which accounted for the top 40 hits (**Table 5.2**).

Table 5.2 Proteins significantly enriched from co-immunoprecipitation of 3xFLAG ForG in *Streptomyces formicae*. Only those below the BFDR threshold of 0.2 and with a fold change greater than 10 have been listed here.

Gene Hit	Description	Fold Change
KY5_6977	Zinc metalloproteinase/ aureolysin	322.5
KY5_1974	2'3'-cyclic nucleotide 2'-phosphodiesterase	140
KY5_0089	TPR repeated containing protein	137.5
KY5_0764	Anchor A Family Protein	132.5
KY5_5660	Phospholipase C	97.5
KY5_7990	GSHD domain-containing protein	72.5
KY5_1430	Putative sensor-like histidine kinase	66.75
KY5_3853	NPCBM domain-containing protein	50
KY5_leuB	3-isopropylmalate dehydrogenase	50
KY5_2482	Peptidase M14 domain-containing protein	47.5
KY5_7239	Uncharacterised protein	42.5
KY5_6663	Sensor kinase (ForG)	40.05
KY5_5288	Putative reductase	40
KY5_7866c	Cytochrome c oxidase subunit	37.5
KY5_2046	Putative secreted alkaline phosphatase	37.5
KY5_1167c	Leucine-responsive regulatory protein	30
KY5_2221c	Neopullulanase / Maltodextrin glucosidase	30
KY5_5929	Siderophore synthetase component, ligase	30
KY5_1427c	Putative ATP/GTP-binding protein	30
KY5_2714c	PLD phosphodiesterase domain-containing protein	30
KY5_5546	Cholesterol esterase	27.5
KY5_3381	Cyclic pyranopterin monophosphate synthase	27.5
KY5_1429c	Multi-component regulatory system-10	27.5
KY5_6635	Glyco_hydro_18 domain-containing protein	27.5
KY5_0496c	Penicillin amidase family protein	27.5
KY5_5204c	Putative esterase	25
KY5_6287	Uncharacterized protein	22.5
KY5_5900	Protein RecA	22.5
KY5_1957c	Secreted protein	20

KY5_1547c	Bifunctional protein: zinc-containing alcohol dehydrogenase, quinone oxidoreductase	20
KY5_3555	Formyltetrahydrofolate deformylase	18
proS	Proline-tRNA ligase	17.5
KY5_5164c	Catalase	15
KY5_2754	Uncharacterized N-acetyltransferase	15
ppc	Phosphoenolpyruvate carboxylase	14.25
KY5_7901	D-alanyl-D-alanine carboxypeptidase	12.75
KY5_2011c	Putative membrane protein	12
polA	DNA polymerase I	12
KY5_7991	Uncharacterized protein	10.5
KY5_2507c	Putative secreted peptidase	10.25

The most abundant protein that was pulled down by ForG, showing a hugely significant 322-fold increase in comparison to the wildtype control, was a zinc metalloproteinase/ aureolysin. Metalloproteinases are ubiquitous among the bacterial community but the most extensively studied are those that have strong associations with pathogenic bacteria or where there is the potential of industrial application. Metalloproteinases act to cleave proteins with their catalytic activity being dependent on the presence of zinc, the proteolysis can take place at the terminal amino acid (exopeptidase) or in the middle of a polypeptide chain (endopeptidase). For example, aureolysin is a secreted extracellular protein that acts as a virulence factor in *S. aureus* as part of the staphylococcal proteolytic cascade, cleaving complement C3 to assist in evasion of the host immune system (Laarman et al., 2011; Shaw et al., 2004). However, *S. formicae* is not a pathogen, so it is unlikely that this metalloproteinase has a role in virulence. BLAST analysis of the primary sequence of this gene only showed similarity to genes within other *Streptomyces* strains with no hits for any genes for *S. aureus* or any other pathogenic bacteria. It is instead hypothesised that the zinc metalloproteinase may be acting to cleave a longer polypeptide into smaller protein fragments that may act as an activating ligand that binds to the sensor domain of ForG, initiating autophosphorylation. The 322-fold increase in this protein's abundance in the presence of FLAG-tagged ForG signals a likely significance in either the structure or functional role of the sensor kinase. There is limited evidence for such proteases being able to regulate the activity of bacterial histidine kinases in this way in *Caulobacter crescentus* (C. Zhang et al., 2022). As it has been possible to purify ForG independently of any other

proteins, it is unlikely that their interaction has a structural dependence therefore further investigations into their functional interaction are required, with attempts to co-purify the proteins and generation deletion mutants of KY5_6977 in both a wildtype and Δ*forG* background.

The second most abundant protein was the enzyme 2',3'-cyclic nucleotide 2' phosphodiesterase, encoded by KY5_1974, which is known to form a part of the metabolic pathways of nucleotides and 2',3'-cyclyic nucleotide 2' phosphodiesters are formed as intermediates during RNA hydrolysis by ribonuclease I. The relevance of this enzyme in its interaction with ForG is unknown and has not yet been investigated further.

The third most abundant protein has no specific designation as to is role or function beyond containing a tetratricopeptide repeat (TPR) and is encoded by KY5_0089. These structural motifs consist of 34 tandem repeat degenerate amino acids with the number of repeats known to range from three to 16. They are often attributed with facilitating protein-protein interactions and mediating the formation of multiprotein complexes (Blatch & Lä Ssle, 1999). Given the similar fold-changes seen between the proteins, it is possible that KY5_0089 enables a protein-protein interaction between ForG and KY5_1974 or between ForG and KY5_0764. The former of these showed a 140-fold increase and the latter a 132-fold increase, while KY5_0089 showed a 137-fold increase.

Interestingly, the fourth most abundant protein and one of these potential candidates for protein-protein interaction mediated by KY5_0089, is a choice-of-anchor A protein, encoded by KY5_0764. These are cell wall anchoring proteins that typically attach surface proteins to the cell wall of Gram-positive pathogens but have been documented to attach cytoplasmic and periplasmic proteins to the cell wall depending on the structure of the anchoring protein (Marraffini et al., 2006). It is entirely possible that a protein-protein interaction between ForG and KY5_0764 is mediated by KY5_0089 to anchor ForG to the cell wall to account for the lack of transmembrane domains in the sensor kinase. However, this would be an incredibly inefficient and atypical method given the highly conserved nature of TCSs across the bacterial community with the efficient transfer of signal to response through just two proteins. Equally, the entire formicamycin BGC has been heterologously expressed with successful production of formicamycin, independently of any of these proteins, suggesting that functionality of ForG is not reliant on any of them. Nevertheless, further investigations are underway to investigate this hypothesis with the

generation of deletion mutants of each of these proteins that showed a significant abundance after being pulled down from coimmunoprecipitation with ForG.

The final notable result from these data that will be briefly discussed is that the twelfth most abundant protein was ForG. This was highly expected given the predictions and current understanding of sensor kinases existing in dimeric forms.

5.3 Passing on the Signal

One of the key defining features of TCSs is their ability to relay a perceived signal and elicit a response in the activity of the cell. This signal is transmitted between the HK and the acting RR through the transfer of a phosphoryl group between conserved residues in each protein. Within ForG, this conserved residue is a histidine at position 175, which is situated within the DHp domain and predicted to autophosphorylate when the sensor domain receives an activating signal. Based on co-crystallised structures of the catalytic domains of other sensor kinases with ATP, it has been possible to predict this interaction as it would occur in ForG (**Fig 5.5**). While there is no resolved structure of ForG, there was a high degree of confidence

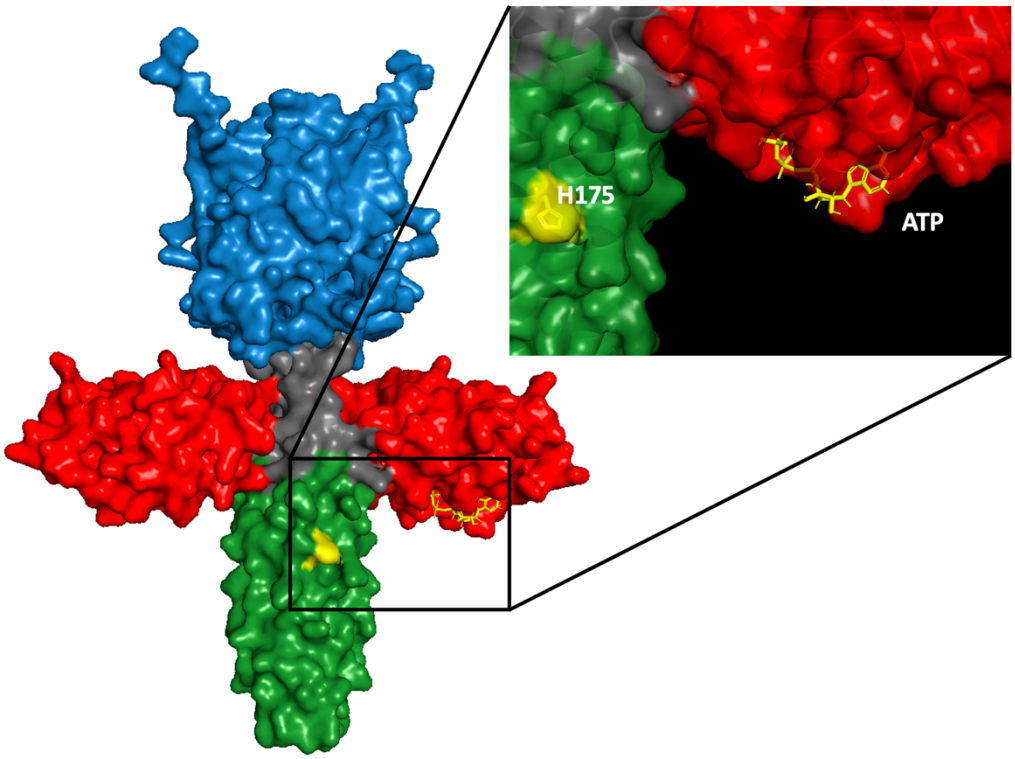


Figure 5.5 A detailed view of the predicted ATP binding pocked within the catalytic ATPase domain of the sensor kinase ForG. Also shown is the conserved histidine residue at position 175 within the DHp domain. Modelling of ForG is predicted by AlphaFold-2 and the ATP binding pocket has been predicted based on similar structures that have been co-crystallised with ATP.

in the prediction modelling of this domain as a highly conserved aspect of HKs. ATP is known to bind in similar such pockets which is often situated a short distance from the conserved histidine residue allowing the freed phosphoryl group to be transferred with minimal complications. In the case of ForG, the ATPase domain shown in red would catalyse the cleavage of ATP to ADP upon receipt of an activating signal, freeing a phosphoryl group to attach to the conserved H175 in the DHp domain.

To confirm the role and interaction of this histidine within ForG, mutations were made to change the residue, with the intention of rendering it unable to accept a phosphoryl group. A dual approach was taken to achieve this mutation, one using CRISPR/Cas9 to generate a single point mutation and the other using CRISPR/Cas9 to generate a total gene deletion that could be complemented with a mutated copy of the gene. In both cases the mutation aimed to change the encoded amino acid from a histidine to an alanine, which cannot be phosphorylated. The deletion of *forG* also allowed for the assessment of any redundancy in the genome in the activity of this kinase, as to whether any other HKs had the ability to activate its cognate RR ForF. The complementation of Δ*forG* with the *forG* H175A was done so under the control of a native promoter to ensure that it would be expressed at normal wildtype levels. As the point mutation being generated with CRISPR/Cas9 could not be confirmed by identifying differences in band sizes or any apparent loss or gain of base pairs, the resultant strains were confirmed via PCR amplification and sequencing of the *forG* gene to establish successful mutation. The production of formicamycin was subsequently measured to be able to understand the impact of the point mutation and gene deletion on the biosynthetic pathway (**Fig 5.6**). Gene deletion of *forG* abolished formicamycin production, as did deletion of *forGF* and production was restored following complementation of the mutant *in trans* using the *forG* gene under the control of its native promoter. However, complementation of the Δ*forG* mutant with a gene encoding ForG with the H175A point mutation did not restore formicamycin production. Similarly, the generation of the same mutation as a single point mutation using CRISPR/Cas9 also saw no formicamycin production from this strain suggesting ForG H175A is inactive.

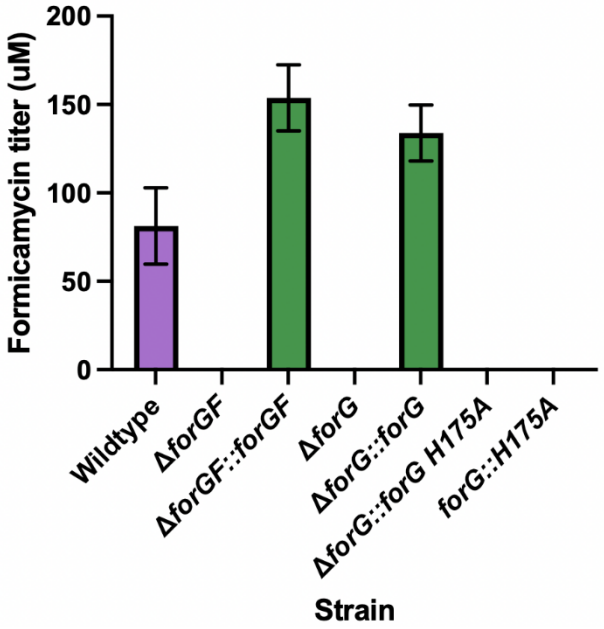


Figure 5.6 Titres of formicamycins in μM between wildtype and the *forG* gene deletion strains alongside the CRISPR/Cas9 generated point mutation and the complementation of the gene deletion with a *forG* allele containing the same mutation. Values are the mean with error bars to represent standard deviation (n=3).

These data demonstrate that ForG is able to act as a typical HK, despite its predicted structural difference from other such regulatory elements. The ability to turn an activating signal into a response in the form of formicamycin biosynthesis is dependent on the conserved histidine residue. Here it has been shown that mutation of that histidine to a phosphodeficient alanine completely shuts down production. In addition these results demonstrate that there is no cross talk between other HKS or small molecules and ForF that result in ForF becoming phosphorylated.

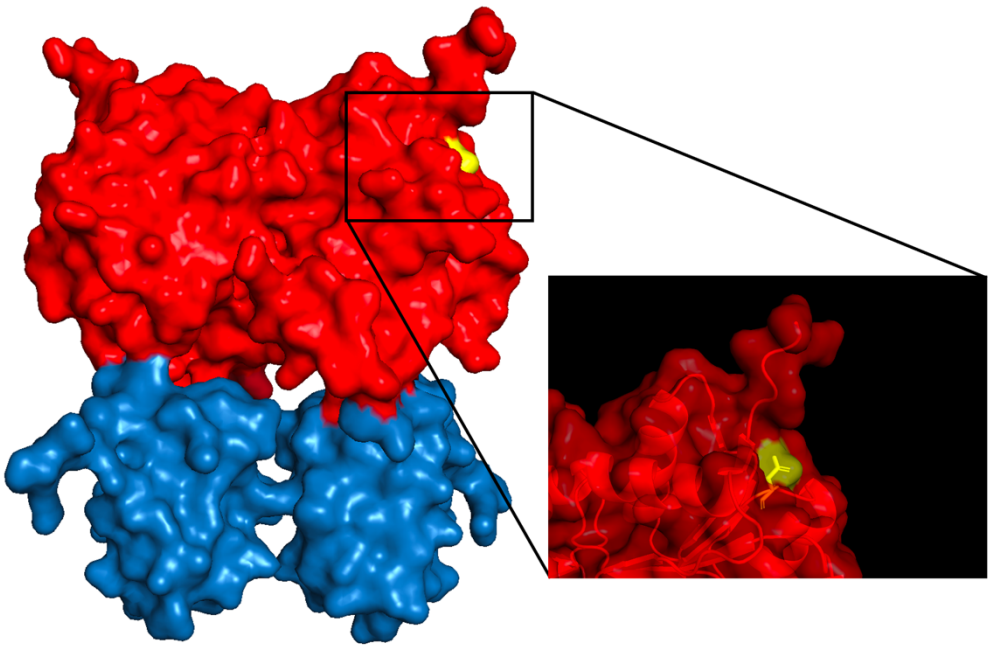


Figure 5.7 A detailed view of the site containing the conserved aspartate residue at position 53 within the receiver domain of the response regulator ForF. This residue accepts the phosphoryl group transferred from the conserved histidine residue within the cognate histidine kinase ForG. Modelling of ForF is predicted by AlphaFold-2.

Continuing this line of investigation, the phosphate group that attaches to the histidine residue in ForG would typically be transferred to a conserved aspartate residue within ForF. For most RRs, this residue is found between positions 50 and 55, with the majority existing at position 53. In ForF this is no different, with D53 sitting within the receiver domain of its predicted structure (**Fig 5.7**). Phosphotransfer would typically result in a conformational change in the RR, mediating a change in its DNA binding capabilities, enabling it to recognise and bind to a specific DNA sequence in or around its target genes. For some RRs, the phosphorylated form is its active DNA binding form while for others this phosphorylated state renders it inactive. To determine whether ForF is active in its phosphorylated or unphosphorylated state, the conserved aspartate residue was changed to a phosphomimetic glutamate and a phosphodeficient alanine. In some, but not all, RRs changing the aspartate to glutamate mimics the phosphorylated state and can alter the activity of RRs in the same way that phosphotransfer would in other RRs. Modelling of ForF containing each of these changes shows the potential conformational change that would occur (**Fig 5.8**). The confidence in each of these models is consistent with one another, and reasonable high across both the receiver and the DNA binding domains (**Appendix Figures 7.5 and 7.6**) Here it can be seen that the native protein has a largely open conformation, with a gap between the DNA binding domains, whereas the phosphomimetic variant containing a D53E change has a closed conformation, with the two DNA binding domains

crossing over one another. The phosphodeficient variant containing a D53A change is not completely identical to the native protein but does show the same open confirmation between the DNA binding domains.

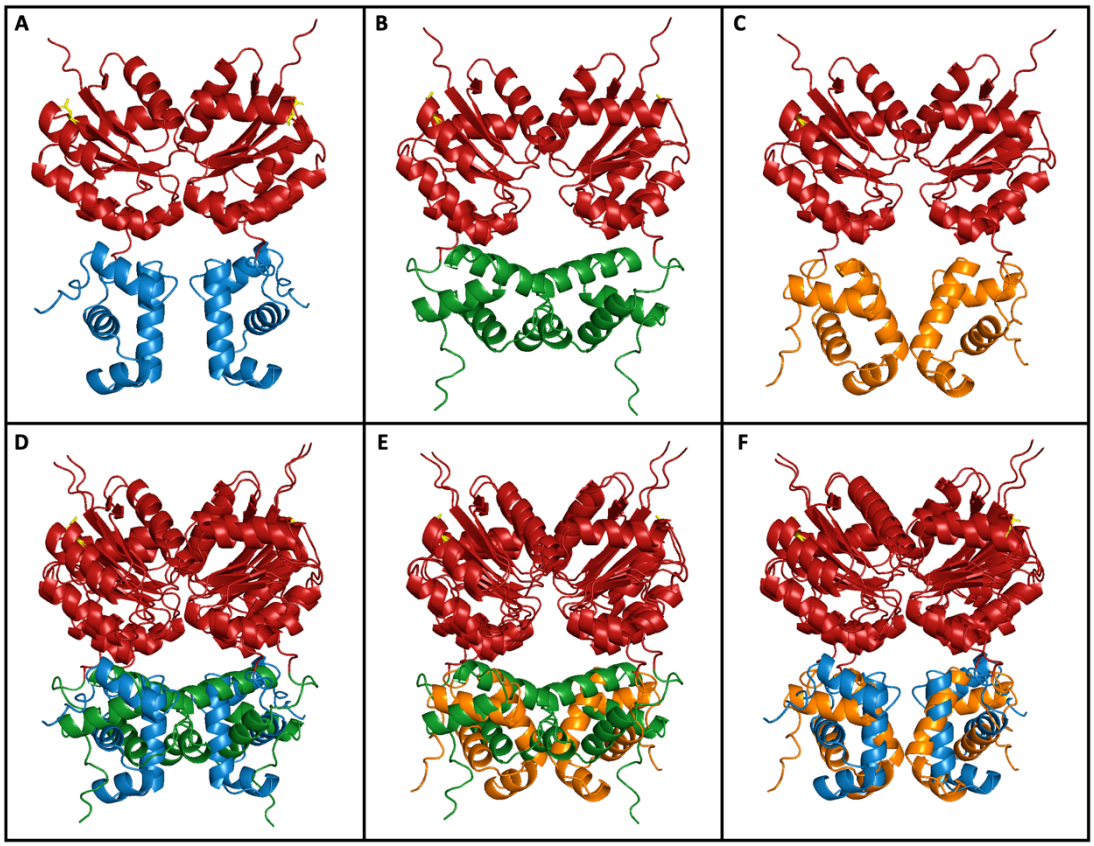


Figure 5.8 Visualisation of the impact of mutating D53 within the response regulator. **A** The native protein with the receiver domain (red) and DNA binding domain (blue) alongside the conserved aspartate residue (yellow) with an open confirmation of the DNA binding domain. **B** Phosphomimetic D53E mutation showing a closed conformation of the DNA binding domain (green). **C** Phosphodeficient D53A mutation showing an open confirmation of the DNA binding domain (orange). **D** Overlay of the native protein (red and blue) with the phosphomimetic D53E variant (red and green) to show the change in conformation of the DNA binding domain. **E** Overlay of the phosphomimetic D53E variant (red and green) with the phosphodeficient D53A variant (red and orange) to show the change in conformation of the DNA binding domain. **F** Overlay of the native protein (red and blue) with the phosphodeficient D53A variant (red and orange) to show the change in conformation of the DNA binding domain. All models generated using AlphaFold-2.

As with the mutations in *forG*, a dual approach was taken using CRISPR/Cas9 to generate a single point mutation alongside the generation of a total Δ*forF* gene deletion with complementation using a mutated copy of the gene. Once again, the production of formicamycin was measured to determine the effect of the point mutation and gene deletion on the formicamycin biosynthetic pathway (**Fig 5.9**). These data showed ForF, like ForG, is essential in the activation of formicamycin biosynthesis and there is no redundancy within the genome in the form of other response regulators that may be phosphorylated by active ForG to regulate gene expression and initiate the formicamycin biosynthetic pathway. The production of formicamycin was restored to wildtype levels upon complementation of the *forF* deletion mutant with the native gene under the control of its native promoter. Furthermore, complementing the deletion mutant with a copy of the *forF* gene containing a phosphomimetic mutation restored formicamycin biosynthesis, while complementing it with a *forF* allele containing a phosphodeficient mutation did not. The same pattern was seen when this mutation was made using CRISPR/Cas9 to bring about a single point mutation. These data suggest that in order to activate target gene expression and formicamycin biosynthesis, ForF must be in a phosphorylated state as mediated by the aspartate residue at position 53.

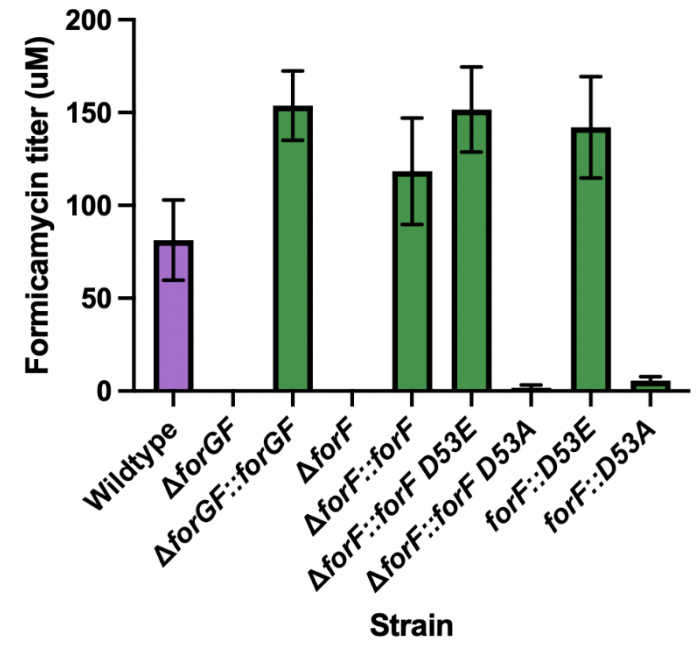


Figure 5.9 Titres of formicamycins in μM between wildtype and the *forF* regulator gene deletion strains alongside CRISPR/Cas9 generated point mutations to generate D53E (phosphomimetic) and D53A (phosphodeficient) ForF proteins. Data showing the complementation of the *forF* gene deletion mutant with this same variants are also shown. Values are the mean with error bars to represent standard deviation (n=3).

An additional proposed line of investigation into the transfer of the signal from ForG to ForF was to visualise this phosphorelay using radiolabelled ATP. This began with purification of the native proteins alongside the conserved residue variants mentioned here. The native ForG and ForF have been purified as described previously in this thesis, and the ForG H175A, ForF D53E and ForF D53A variants were purified using these same methods. Purification of all five of these proteins was successful, generating stocks of each at a consistent 5 mg/ mL using an AKTA Pure with a His trap column followed by size exclusion (**Appendix Figures 7-9**). Unfortunately, it was not possible to obtain the appropriate licencing to complete radioactive work within the timescale of this project. However, a protocol for purifying the proteins required for this investigation has been established and work remains in progress. The investigation would attempt to visualise the phosphotransfer from ForG to ForF by tracking γ-^{32}-labelled ATP and analysing the proteins using SDS-PAGE to with the hope of seeing this radioactive isotope move from one protein to the other.

To continue attempts to determine the structure of ForF in either its active or inactive state, a number of sitting drop crystallisation trials were set up with these phosphomimetic and phosphodeficient variants. Through a range of buffer trials for each protein, samples were successfully concentrated to a range of 6-15 mg/mL, but screens with these samples have been unsuccessful in generating protein crystals.

The final investigation carried out into the interaction of ForG and ForF was to perform amine coupling. As both native proteins had already been purified, as described earlier, it was possible to use SPR to understand the protein-protein interaction between the two. Amine coupling immobilises one protein of interest to a CM5 chip (containing a matrix of carboxymethylated dextran covalently attached to a gold chip) via the free amine at its N-terminus and flows the other over the chip surface while measuring any interaction that takes place between the two proteins. It is possible determine whether they bind to one another or interact in any way during this investigation. Amine coupling was performed in varying combinations of bound proteins and analytes at a range of concentrations to investigate the relationship between ForG and ForF. These data appeared to show no direct interaction between ForF and ForG in any of the combinations tested, nor do they show any interactions between ForG with itself or ForF with itself (data not shown). The latter of these was anticipated as both molecules already exist as dimers and have been purified as such, with no evidence or indication that either would form larger multimers. However, the former of these results was somewhat unexpected, as similar investigations into TCSs have

shown an increase in covalent binding between the two components at increasing concentrations (Hörnschemeyer et al., 2016; Kundu et al., 2021). However, amine coupling as a technique has been shown to reduce the binding affinity of two interacting proteins, as the immobilisation to the CM5 chip via amine groups can potentially block the binding site (Kortt et al., 1997). Instead, similar techniques can be used to immobilise the protein in varying orientations to limit the chance of this happening (Huang et al., 2009). There has also been some contradictory evidence to suggest that the interaction between a sensor kinase and its cognate response regulator is purely transient, mediated by the transfer of a phosphoryl group between conserved residues (Zapf et al., 2000).

Data from co-immunoprecipitation of ForF (shown below) and ForG did not identify one another as a protein that they bind to *in vivo* with any significance. This appears to align with the amine coupling data and suggests that the two do not normally bind to one another to facilitate the phosphotransfer. However, there must be some degree of recognition between the cognate components as point mutations have demonstrated no redundancy within the genome. This means that ForG cannot activate another RR, nor can ForF be activated by another HK to activate formicamycin biosynthesis. Therefore, these proteins must be able to recognise one another specifically to relay this signal.

To establish whether the two proteins co-localise within the cell, attempts were made at fluorescently tagging each with mCherry and mTurquoise. However, it was not possible to establish growth conditions that were simultaneously compatible with fluorescence microscopy and formicamycin production with ForG and ForF being expressed and active in the cell. It is hypothesised that this is due to the previously discussed increased repression by ForJ under oxidising conditions such as aerobic growth of liquid cultures, repressing the production of both ForG and ForF, however attempts to overcome this are ongoing,

5.4 The Response of ForF

Throughout this project, a significant amount of progress has been made in terms of understanding the role of ForF in eliciting its regulatory effects to control the production of formicamycin at a transcriptional level. The binding motif has been elucidated and the impact of this binding on both promoter activity and transcription of mRNA have been identified. This has provided significant insight into the mechanism of ForF and the way in which it is able to regulate formicamycin biosynthesis. Thus far ForF has been shown to act

like a typical RR in terms of its activation, predicted conformational changes on receipt of an activating phosphoryl group and action upon binding DNA.

As ForF is known to bind the intergenic region that covers the divergent promoters of *forGF* and *forHI* and activate the expression for *forHI*, a deletion mutant lacking *forHI* was generated using CRISPR/Cas9 to determine whether removing this operon disrupts formicamycin biosynthesis. It was hypothesised that its removal would eradicate formicamycin biosynthesis much like the Δ*forGF* strain, as this is the only known ForF target within the formicamycin BGC and therefore the most logical mechanism of activation of formicamycin biosynthesis. However, deletion of the *forHI* genes showed no significant change in the levels of formicamycin produced, with complementation of this mutant also maintaining production at wildtype levels (**Fig 5.10**). These data show that although the only seemingly relevant DNA binding activity of ForF is to bind to the divergent promoters *forGF-forHI* and activate transcription of its own genes *forGF* and the production of ACC biotin carboxylase and ACC carboxyl transferase encoded by *forHI*, activation of the formicamycin biosynthetic pathway is not achieved in this way. Thus, it further suggests that ForF must be carrying out another role to be able to enable pathway activation.

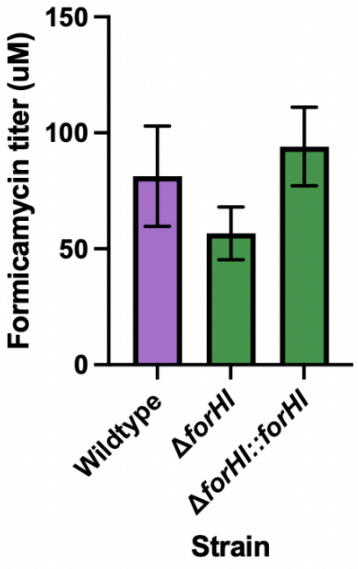


Figure 5.10 Titres of formicamycins in μM in wildtype and the *forHI* malonyl CoA starter unit deletion mutant. Values are the mean with error bars to represent standard deviation (n=3). Titres from the complemented strain are also shown.

It was therefore decided to investigate any other interactions that may be taking place between ForF and other proteins in the cell. To begin *in vivo* investigations, the previously generated Δ*forF* strain of *S. formicae* was complemented with a copy of the same gene encoding a 3xFLAG tag ForF protein to facilitate co-immunoprecipitation. A total of 94 proteins were found to be in abundance below the Bayesian false discovery rate (BFDR) of 0.2 in comparison to the wildtype control. These were further sorted by enrichment, with further investigation being carried out on any proteins that has a greater than 10-fold increase, which accounted for the top 52 hits (**Table 5.3**).

Table 5.3 Proteins significantly enriched from co-immunoprecipitation of 3xFLAG ForF in *Streptomyces formicae*. Only those below the BFDR threshold of 0.2 and with a fold change greater than 10 have been listed here.

Gene Hit	Description	Fold Change
KY5_2179c	Putative membrane protein	135
KY5_3288c	NAD-dependent oxidoreductase	100
KY5_6664	Two-component system response regulator (ForF)	90.5
glnD	Uridylyltransferase	77.5
KY5_4856	DNA-binding response regulator, LuxR family	60
KY5_1637c	Glutaryl-CoA dehydrogenase	58.5
KY5_2714c	PLD phosphodiesterase domain-containing protein	55
KY5_6811c	ABC transporter (Iron.B12.siderophore.hemin), ATP-binding component	55
KY5_7030	Cys-tRNA(Pro) deacylase YbaK	52.5
KY5_3337c	Putative two-component system response regulator	47.5
trpD	Anthranilate phosphoribosyltransferase	47.5
rplF	50S ribosomal protein L6	45
KY5_0790	Thioredoxin reductase	45
KY5_4163	Uncharacterized protein	42.5
KY5_2268c	Uncharacterized protein	42.5
KY5_5929	Siderophore synthetase component, ligase	42.5
glyQS	Glycine-tRNA ligase	42.5
KY5_1296	ATPase involved in DNA repair	37.5
KY5_5288	Putative reductase	37.5
KY5_5290	Sulphur carrier protein adenylyltransferase ThiF	37.5
KY5_3740	Uncharacterized protein	35

purL	Phosphoribosylformylglycinamidine synthase subunit PurL	35
ettA	Energy-dependent translational throttle protein EttA	30
KY5_2482	Peptidase_M14 domain-containing protein	30
KY5_3287c	Cyclopropane-fatty-acyl-phospholipid synthase	27.5
pheT	Phenylalanine--RNA ligase beta subunit	27.5
KY5_4161c	Putative DNA-binding protein	27.5
KY5_4282c	Putative membrane protein	25
KY5_0560	Aldehyde dehydrogenase	24.75
KY5_1167c	Leucine-responsive regulatory protein, regulator for leucine (Or lrp) regulon and high-affinity branched-chain amino acid transport system	22.5
psd	Phosphatidylserine decarboxylase proenzyme	22.5
KY5_3793	Transcriptional regulatory protein GlnR	22.5
KY5_5972c	DNA topoisomerase (ATP-hydrolyzing)	20
KY5_4328c	Putative secreted protein	20
KY5_1808	Putative DeoR-family transcriptional regulator	18
KY5_7760c	Uncharacterized protein	17.5
proS	Proline-tRNA ligase	17.5
KY5_0717	Streptogrisin-C (Serine protease C) (SGPC)	16.75
metG	Methionine--tRNA ligase	16.5
KY5_1430c	Putative sensor-like histidine kinase	15.75
KY5_2754	Uncharacterized N-acetyltransferase	15
ppc	Phosphoenolpyruvate carboxylase	14.25
rplT	50S ribosomal protein L20	13.5
KY5_0401	Adenylosuccinate lyase	12.38
dapD	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	12
pfp	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase	12
polA	DNA polymerase I	12
KY5_5674	D-3-phosphoglycerate dehydrogenase	11.88
KY5_2011c	Putative membrane protein	11.62
KY5_4039	tRNA-dependent lipid II--amino acid ligase	10.5
KY5_2653c	DNA-binding protein in cluster with Type I restriction-modification system	10.5
KY5_2179c	Putative membrane protein	135

The most abundant protein that was pulled down by ForF, was a putative membrane protein encoded by KY5_2179. No further information is available on the role or function of this protein, and BLAST analysis showed no homologues with any available data on this either. However, the 135-fold change in abundance in comparison to the wildtype suggests that there is clearly some significance in the relationship and interaction of these two proteins.

The second most abundant protein was an NAD-dependent oxidoreductase, encoded by KY5_3288. These enzymes catalyse the oxidation or reduction of their substrate while facilitating the opposite reaction in a nicotinamide adenine dinucleotide (NAD) cofactor (Sellés Vidal et al., 2018). It is possible that this is related to the ATP requirement of the TCS phosphorelay pathway, but this seems unlikely as the actual demand for ATP is at the HK, not the RR. There is also a possibility that this interaction is linked to the oxidative stress that is subsequently generated by the production of formicamycin in the cell. As discussed in **Chapter 3**, the biosynthetic pathway of formicamycin generates oxidative stress, so it may be that ForF binds to this oxidoreductase to promote the generation of NADH and subsequently minimise the impending impact of pathway activation.

The third most abundant protein-protein interaction observed in coimmunoprecipitation was ForF. This was highly expected as the regulator can only function in a dimeric state, otherwise it lacks the ability to bind DNA and exert its primary function through regulation of gene expression. This dimerisation appears to be mediated at a valine residue in the flexible loop between β5 and α4 and a lysine residue in the flexible loop between β6 and α5. It is likely that other residues and secondary structure are also involved in. this dimerisation, but this is not inherently apparent based on structural modelling alone (**Fig 5.11**).

The fourth most abundant protein to be pulled down by ForF was GlnD, a P_{II} uridylyltransferase encoded by KY5_5769. This enzyme catalyses the post-translational addition of a uridylyl molecule to a protein P_{II} using UTP as the second substrate. The P_{II} family of proteins are involved in signal transduction as part of nitrogen metabolism, with their de-uridylylation occurring during nitrogen starvation in response to the subsequent increased ammonia levels in the cell. Whilst the protein encoded by this gene is predicted to produce a uridylyltransferase, research has shown that the homologous GlnD within *S. coelicolor* acts as an adenylyltransferase (Hesketh et al., 2002). Here it acts to modify the P_{II}

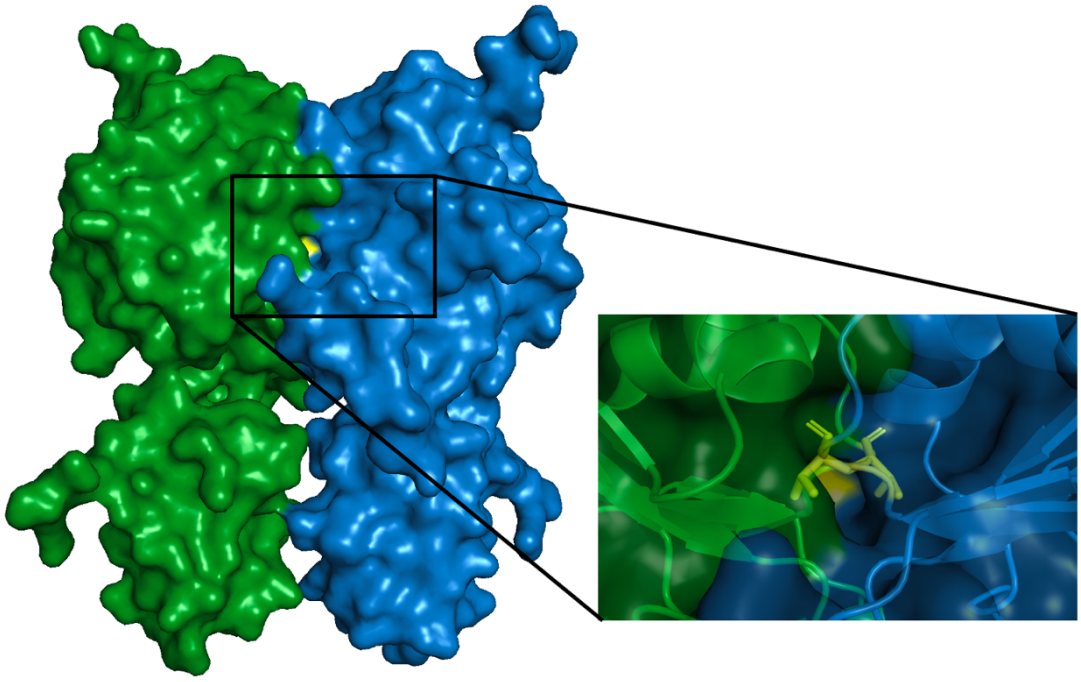


Figure 5.11 A detailed view of the dimerisation site of the ForF monomers to create the final homodimer. Each monomer is shown in green and blue with the lysine residues at position 108 highlighted in yellow. Modelling of ForF is predicted by AlphaFold-2.

protein GlnK via adenylation of a conserved tyrosine residue at position 51, again in response to low nitrogen concentration (Krysenko, 2023). There is no apparent structural or functional reason for ForF to bind to a uridylyltransferase or an adenylyltransferase, as this response regulator does not fall into the P_{II} family, nor is there any suggested link between nitrogen levels and the production of formicamycins.

The fifth and final most abundant protein that was pulled down by ForF that will be discussed here, was another LuxR family DNA binding RR, encoded by KY5_4856. There has been previous resports that orphan RRs such as BldM and WhiI can form heterodimers to regulate gene expression and control various developmental stages within *Streptomyces* species, with the ability to bind DNA targets that neither protein would normally regulate when acting as homodimers (Al-Bassam et al., 2014). Similarly, ScbR and ScbR2 also form a heterodimer and bind a promoter region that is not normally targeted by either homodimer (X. Li et al., 2017). This phenomenon has also been seen in *E. coli* where BglJ and RcsB act as a heterodimer to relieve repressive regulators (Venkatesh et al., 2010). Recent evidence has also shown that paired response regulators are also capable of this, such as the formation of the heterodimeric MtrA:WblE (unpublished data), their target(s) and impact hasn't been elucidated but it is likely that they also regulate different pathways to their homodimeric equivalents. Co-immunoprecipitation has supported the idea of a similar

interaction between ForF and KY5_4856, but the basis and significance of this interaction is currently unknown. However, it would be a worthwhile line of investigation to follow up in the future. It is currently hypothesised that ForF forms a heterodimer with this response regulator, and that this heterodimer is able to bind and regulate other DNA targets in the *S. formicae* genome. This may also explain the previously discussed lack of impact of *forHI* deletion on formicamycin production. As it has been shown that deletion of these started unit doesn't eradicate formicamycin production i.e. it is possible that in combination with KY5_4856, ForF forms a heterodimer that controls the formicamycin biosynthetic pathway through a different regulatory pathway. Further investigations are required to understand this interaction including the generation of a KY5_4856 deletion mutant and co-purification of the proteins.

5.5 Impact of ForGF on the Proteome

Typically, investigations into the function and activity of protein(s) would combine data on the binding and regulatory impact as has been outlined thus far, with RNA-seq to determine the comparative changes to gene expression in gene deletion strains. This provides further information on their putative regulons and is able to confirm the impact of a regulator on global mRNA levels which may extend beyond the previously understood binding sites. However, the relationship between mRNA levels and active protein within the cell does not necessarily correlate, and it is only the latter of these that typically results in a phenotypic change within a cell (McManus et al., 2015). A number of processes take place between the transcription of mRNA and the existence of a functional protein, including translation and post-translational modification. In addition, delays to protein synthesis, folding errors and translational regulators are all able to delay or otherwise alter the generation of active proteins (Hackl & Bechthold, 2015; Hesketh et al., 2007; Higo et al., 2011). This would traditionally have led to carrying out two-dimensional PAGE and immunoblotting or tandem mass spectrometry to measure protein abundance and activity. However, these techniques are notoriously challenging, have poor reproducibility, struggle to manage large data sets and are unreliable at providing accurate data for proteins that are at either end of the scales for size, pH and hydrophobicity (Gygi et al., 2000). To overcome this, the development of tandem mass tagging proteomics has allowed for a large-scale comparison of proteomes between strains with a reliable and reproducible data output (Thompson et al., 2003). This

technique uses isobaric chemical tags with an MS/MS reporter group and an amine reactive group that binds to the N terminus of proteins that have been extracted and digested. The MS/MS reporter group is varied between samples allowing them to be mixed and concurrently analysed via LC-MS/MS with up to 16 tags currently being available for use. The isotopic balancer means that differentially tagged proteins of the same mass will be identified as the same product during MS1 but can then be differentiated to their origin sample group during MS2.

On this basis, TMT proteomics was carried out on the wildtype and Δ*forGF*, Δ*forJ* and Δ*forZ* regulator gene deletion strains of *S. formicae* by Dr Rebecca Devine. Only the comparison of proteomes between the wildtype and Δ*forGF* strains will be discussed here as they are the most relevant to this project and the data is presented with her permission (**Fig 5.12**). A number of proteins were shown to have significant changes to their abundance in the absence of ForGF. As anticipated, the most significantly downregulated proteins included all of the biosynthetic enzymes that make up the formicamycin BGC which are essential for

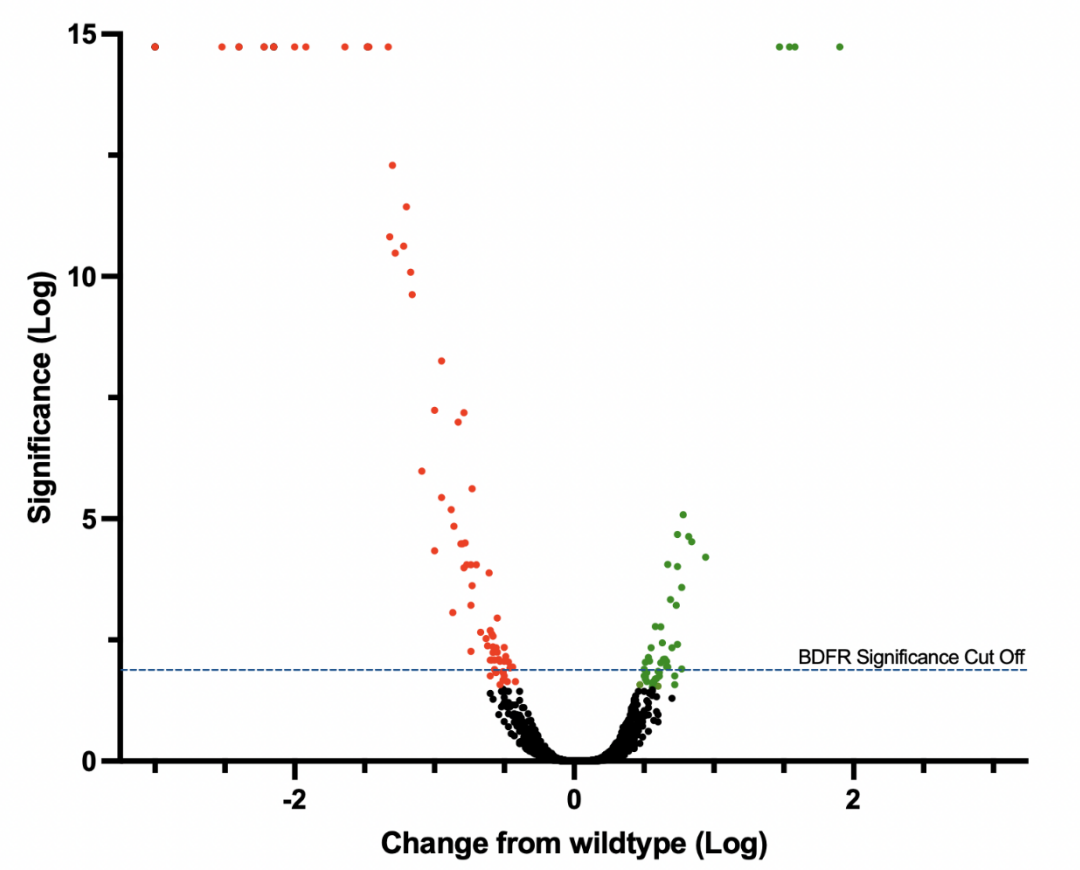


Figure 5.12 Volcano plot of the -log fold change of protein abundance against the -$\log_{10}$ p value between the *S. formicae* wildtype and *S. formicae* Δ*forGF* strains. The logBDFR significant cut off of 1.5 is shown with a blue dotted line and any proteins that showed a decreased change in abundance are represented with red dots while those that showed and increased change in abundance are represented with green dots.

the formation of various congeners as part of the biosynthetic pathway. As has been shown previously, gene deletion strains lacking ForGF do not produce any formicamycins and it is therefore logical that none of the biosynthetic enzymes are produced in this strain either. Also included in the most significantly downregulated proteins were KY5_3366 and KY5_3367, an MMPL family transporter and a TetR family transcriptional regulator, respectively. These proteins are encoded on a two-gene operon that are likely under the control of the same promoter. While there are no available data on transcriptional start sites around these genes, visual inspection of the intergenic region upstream of the start codon of KY5_3367 reveals a potential ForF binding site at -65 base pairs that may have been missed when searching the genome using the previously established binding motif. It is possible that ForF would bind here and be able to activate the transcription of these two genes. However, this remains a hypothesis and there are several investigations that would need to be carried out to establish whether this is actually the case, including identification of transcriptional start sites and confirmation of ForF binding. It is also unclear whether this potential interaction is in any way related to the production of formicamycin or whether ForF is involved in a different regulatory cascade.

Significantly fewer proteins were shown to be upregulated in the absence of ForGF, with a notable gap in significance between four proteins and any others that were upregulated. These four are KY5_1255 – KY5_1258, which are encoded by a series of four genes that are likely encoded within the same operon and under the control of the same promoter. As above though, there are no data currently available on transcriptional start sites around these genes, however, visual inspection of the intergenic region downstream of the start codon of KY5_1255 also identified a potential ForF binding site that may also have been missed when searching the genome using the previously established binding motif due to slight variations in the sequence. This potential binding site is some 621 base pairs downstream of the start codon of KY5_1255 for ForF could block expression by a roadblock mechanism. The proteins encoded by KY5_1255 – KY5_1258 include an XRE family transcriptional regulator, a TIR domain containing protein and two hypothetical proteins, respectively. BLAST analysis of the final two proteins did not provide any further information about their role or function in the cell. Once again, this remains a hypothetical interaction and a number of investigations would need to be carried out to understand the nature and relevance of this change in protein abundance.

A comparison between the previously established binding motif of ForF and the potential binding sites identified here was generated to be able to further understand how these sites may have been missed during previous attempts to identify any other binding sites throughout the genome (**Fig 5.13**). Here it can be seen that a number of nucleotides vary between these potential binding sites and the sequence that has been confirmed via SPR. As these potential sites were identified via visual inspection it is entirely possible that this is just confirmation bias and there is another mechanism through which these proteins show a change in abundance in a Δ*forGF* strain. All of these hypothetical mechanisms and interactions require significant further investigation to be able to understand the potential ways in which ForF is acting in the cell. Due to the time limitations of this project, it has not been possible to establish whether or not any binding takes place or any further impact if binding does take place at these genes. However, the knowledge of these interactions from TMT proteomics provides significant basis for further work to be carried out.

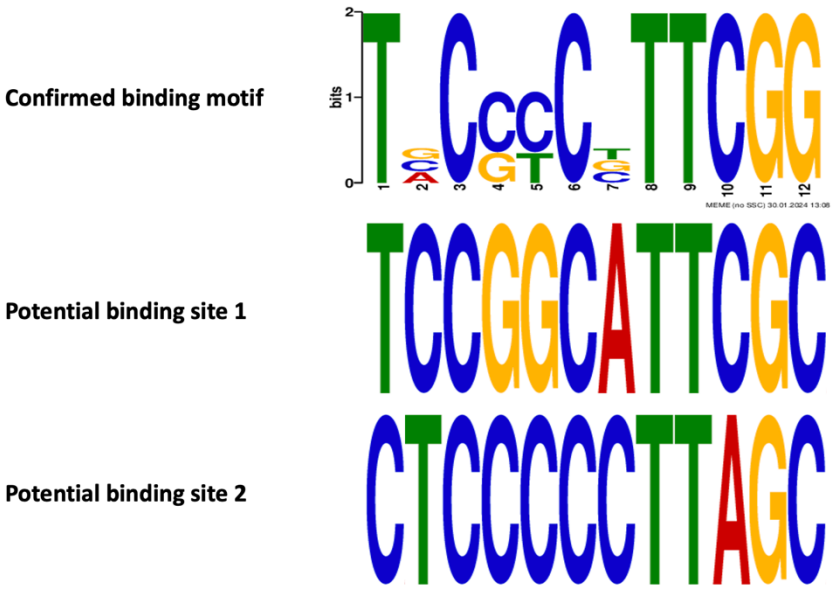


Figure 5.13 Comparison of the confirmed ForF binding site with the potential binding sites identified from visual inspection of the intergenic regions up/downstream of genes that encode proteins shown to have significant abundance changes in a Δ*forGF* strain compared to a wildtype strain of *S. formicae*.

5.6 Conclusions

In this final chapter, a number of investigations have been carried out to understand the ways in which ForG and ForF act as a TCS. Modelling of their respective structures has provided significant insight into how they carry out their regulatory functions to control formicamycin biosynthesis. The predicted lack of any transmembrane domains in the HK ForG has been confirmed by the ability to purify the complete protein from a soluble fraction of cell culture, despite the protein itself being somewhat unstable. This has established that the activating signal for the system must either originate from within the cell or be easily transported across the membrane to be able to initiate signal transduction. Assessment of the sensor domain has not provided much further insight into what this signal may be, with multiple potential binding pockets and no sequence homology with any other known sensor domains. However, analysis of protein-protein interactions via coIP has resulted in several potential avenues of investigation into this, with known interactions including a zinc metalloproteinase, a membrane anchoring protein, and a TPR repeat containing protein known to mediate the formation of multi-protein complexes. Further investigations are underway to understand these interactions and their relevance to the biosynthesis of formicamycin or any other process within the cell.

While it has not been possible to directly prove the phosphorelay mechanism between the HK and the RR, generation of phosphomimetic and phosphodeficient strain variants suggest with a high degree of certainty that the two act as a typical TCS. By preventing autophosphorylation of the conserved histidine residue within the ForG HK, via a phosphodeficient H175A variant, production of formicamycins was inhibited. Similarly, by preventing this signal from being passed on via phosphorylation to the conserved aspartate residue in the receiver domain of the RR, through a phosphodeficient D53A variant, no formicamycins were produced. The hypothesis that ForF needs to the phosphorylated to be in its active state was further supported through the generation of a phosphomimetic D53E variant which had a significantly increased production of formicamycins compared to the wildtype strain. The mechanism through which this active state allows DNA binding is demonstrated with the predicted models of these phosphomimetic and phosphodeficient variants. The DNA binding domain is in an open conformation in both the wildtype and phosphodeficient variant, but in a closed conformation in the phosphomimetic variant. While these variants are not a 100% accurate representation of the conformation of the dimer when phosphorylated and unphosphorylated, the formicamycin titre analysis of the

strains has shown that these are sufficient to yield a change to biosynthesis. Further to this, it has been demonstrated that there is no redundancy in the genome for either of the components with single *forF* and *forG* gene deletions producing the same phenotype as the full *forGF* operon deletion.

In the final stages of the regulatory action of the ForGF TCS, the RR ForF has been shown to bind at two intergenic regions in the genome. One of these is upstream of the *forHI* genes that encode the ACC biotin carboxylase and the ACC carboxyl transferase that result in the boosted production of the malonyl CoA starter unit of the formicamycin biosynthetic pathway. While a deletion of *forGF* has been shown to eradicate formicamycin production, deletion of the *forHI* genes that are normally activated by ForF does not affect formicamycin production. This has long been a source of confusion as to the mechanism through which ForGF is able to activate biosynthesis without activating the expression of any genes required for biosynthesis. However, further research here has provided some potential leads that are worth following up. Analysis of protein-protein interactions has shown that ForF strongly interacts with another LuxR family RR, with literature suggesting that heterodimeric RR complexes might be capable of binding targets that neither homodimer would normally regulate. In addition, a series of proteins have been shown to have significant changes in abundance in the absence of ForGF, suggesting that the TCS regulates their production somehow. The exact mechanism of this requires significant further investigation.

6. Discussion and Further Work

6.1 The Potential of Exploiting Cluster-Situated Two Component Systems

The aims of this work were to understand the role and impact of the CSRs, and specifically ForGF, on formicamycin biosynthesis. Through preliminary investigations into the impact of simple gene deletions on the entire biosynthetic pathway, including total pathway shut down or significant overexpression, it was hypothesised that other such regulators could be exploited in a similar way. Literature reviews concluded that this approach wasn't something that had previously been explored, with the roles of CSRs being more an incidental finding when investigating a specific compound rather than a starting point for identifying new compounds. TCSs have long been known to be involved in the regulation of secondary metabolism with their ability to relay an external signal into the cell and elicit a transcriptional response being an incredibly valuable resource for *Streptomyces* spp. These systems have been demonstrated to be involved in an extremely wide variety of processes such as primary and secondary metabolism, growth and development, and ion regulation (Cruz-Bautista et al., 2023; Jin et al., 2023). Some TCSs are able to respond to multiple extracellular signals to elicit different responses and are often involved in multiple regulatory cascades. Bioinformatic analysis has previously shown that there are 15 highly conserved TCSs across *Streptomyces* species, with only a handful of these being well characterised (McLean et al., 2019). On average each *Streptomyces* strain contains 90 HKs and 80 RRs, however this number can vary significantly and even fewer of these have been characterised (Romero-Rodríguez et al., 2015). Given the diverse number of processes that these systems have been linked to and the notable changes in the production of antimicrobials such as formicamycins, actinorhodin and cinnamycin as a direct result of TCS deletion, it was decided to investigate the potential of cluster-situated TCSs for activating the production of secondary metabolites.

Preliminary screening focussed on the *S. formicae* strain as previous work in the Hutchings lab had demonstrated total formicamycin pathway shut down following deletion of the *forGF* TCS operon. Analysis revealed seven cluster-situated TCSs, including ForGF in *S. formicae*, that could be targeted and the generation of overexpression mutants of each TCS operon showed huge potential for future investigations. Gene organisation of one of these TCSs, TCS 40, made the molecular cloning difficult and this was not pursued. However, as demonstrated in **Chapter 3**, the other five overexpression strains all showed changes in

their bioactivity with correlating changes to HPLC analysis as a direct result of introducing a second copy of the TCSs. Four of these showed increases in bioactivity to a wide range of indicator strains including Gram-positive and Gram-negative bacteria as well as fungal pathogens, with the fifth showing a reduction in activity against a Gram-positive organism. The BGCs that these TCS originate from are just as diverse in terms of the classes of natural product they are predicted to encode and include an unknown terpene and various combinations of NRP/ polyketide/ RiPP hybrids(Cruz-Bautista et al., 2023; Rodríguez et al., 2013) . The ability of these strains to inhibit the growth of clinically relevant pathogens with minimal genetic manipulation is remarkable and worthy of further investigation.

Using simple molecular cloning techniques coupled with crude HPLC analysis, the potential of exploiting cluster-situated TCSs has been demonstrated here. A significant degree of follow up is needed for each of these TCSs, with a number of these techniques having already been applied to the formicamycin BGC. Bioassay-guided fractionation and purification of each of these compounds from the *forGF* mutant background will prevent the potential of rediscovery and allow novel compounds to be the primary focus moving forward. At this point it would be crucial to identify the BGCs that encode these compounds by generating gene deletion strains and heterologous expression. While it seems logical to assume that the cluster in which the TCS is encoded is the producing cluster, the RRs of TCSs have been shown to bind multiple sites throughout the genome (McLean et al., 2019; Som et al., 2017; Zhu et al., 2020). It is therefore entirely possible that the overexpressed cluster-situated TCS is cross-regulating genes from another BGC, however, the class of compound will provide further indication as to the likely candidates to investigate.

Further to this it would be appropriate to continue in a similar approach to how the role of ForGF in the formicamycin BGC and biosynthetic pathway has been understood, with the characterisation of biosynthetic enzymes involved and how these data may be used to understand the biosynthetic pathway. Crucially, it is highly likely that there are other regulators at play and involved in controlling the biosynthesis of the bioactive substances that have been switched on here. Characterising them is also incredibly important to be able to better understand how to refactor a BGC to overproduce bioactive compounds with potential applications in an industrial or medical setting (Horbal et al., 2018; L. Li et al., 2021).

Not only has this work unlocked a significant number of potential lines of investigations into these compounds that have been shown to alter the bioactivity of a strain with significant

potential in terms of the number of BGCs that it encodes, but it has also provided a platform for application to other strains. *Streptomyces* species are renowned for their potential for natural product biosynthesis, which is why they have been at the forefront of antimicrobial research for eight decades (Alam et al., 2022; Donald et al., 2022; Zhao et al., 2019). However, as mentioned, rediscovery of known compounds and difficulty in understanding how to "turn on" BGCs that are not expressed under standard laboratory growth conditions has resulted in a lack of any new compounds being identified. Analysis of just three other *Streptomyces* strains was able to identify a number of other TCSs that this technique could be applied to. Through simple molecular techniques, bioassays and crude analysis it is possible to easily narrow down which would be worthwhile for further investigation. Given the number of TCSs and BGCs encoded within the average *Streptomyces* genome, it is likely that this approach will be applicable to all *Streptomyces* strains to some degree. Other one component regulators such as those in the MarR family are also known to be strongly involved in the control of natural product biosynthesis, and are prevalent throughout *Streptomyces* strains, again providing a further potential line of investigation into the control of secondary metabolites (Grove, 2013, 2017).

6.2 Characterisation of the Cluster-Situated MarR Regulators ForJ and ForZ

By understanding the role of cluster-situated regulators in a biosynthetic pathway, it is possible to develop an in depth understanding of the regulatory network that acts to control the production of a natural product. This can be exploited to rewire the BGC, allowing for overproduction of a compound through both increased biosynthesis and increased export to prevent toxicity to the producing organism. Previous work in the Hutchings lab demonstrated the roles of the two cluster situated MarR regulators ForJ and ForZ (Devine et al., 2021; Qin et al., 2017, 2019b). In this work, alongside collaborative research within the group, it has been possible to characterise these two regulators and gain a better understanding of the conditions under which they function and the mechanisms through which they act. Data from β-glucuronidase assays were able to support work completed by Dr Rebecca Devine to understand the impact of the binding of these two regulators on target promoter activity, complementing qRT-PCR data to provide an overall picture of the transcriptional impact of these regulators.

The main repressor of formicamycin biosynthesis, ForJ, binds at multiple sites throughout the BGC to repress transcription, presumably by physically blocking the binding of RNA polymerase, and preventing the core biosynthetic machinery from being produced. Work by Dr Rebecca Devine using SPR techniques allowed for elucidation of the ForJ binding motif recognised by the DNA-binding protein and demonstrated its ability to form multimers when put under oxidative stress (Devine and Noble, in review). Under normal conditions, the heterodimer would bind to these various sites, weakly repressing the production of the biosynthetic enzymes, but the biosynthesis of the formicamycins results in the accumulation of reactive oxygen species in the cell. This triggers the formation of ForJ multimers which have an increased repressive capability, slowing down the production of the formicamycins. As deletion of *forJ* was shown to increase the titres of formicamycin produced within the cell, this further supports the previously suggested analysis and genetic manipulation of cluster-situated MarR regulators within *Streptomyces* strains as a route to activating the production of other natural products.

The other cluster-situated MarR regulator, ForZ, was established to be involved in the regulation of the primary mechanism of self-resistance to the formicamycins. A low sequence homology with ForJ suggested that the two carried out quite different roles in the context of pathway regulation and once again collaborative investigations acted to demonstrate the exact role it carries out. Data from the β-glucuronidase assays were able to verify the sole binding site within the formicamycin BGC and aligned with qRT-PCR data to show the changes to promoter and transcriptional activity as a result of its binding. Further work by Dr Rebecca Devine demonstrated that this regulator is itself regulated by the levels of formicamycin within the cell. Under normal conditions the heterodimer would bind to repress the production of the main export pump, but accumulation of formicamycins results in their binding to ForZ, switching off its DNA binding activity and allowing the export pump to be produced (Devine and Noble, in review).

Work such as this acts to demonstrate the complexity of regulatory pathways, in this case involving two cluster-situated MarR family regulators that responds to different internal signals to balance the production of a secondary metabolite with its export. The intricate balance means that the cell should never be exposed to toxic levels of the compound as cycles of the pathway and end-product accumulation result in the repression of the pathway and the activation of export mechanisms. An understanding of these regulators has previously facilitated additional work in to be carried out in the lab to rewire the BGC to

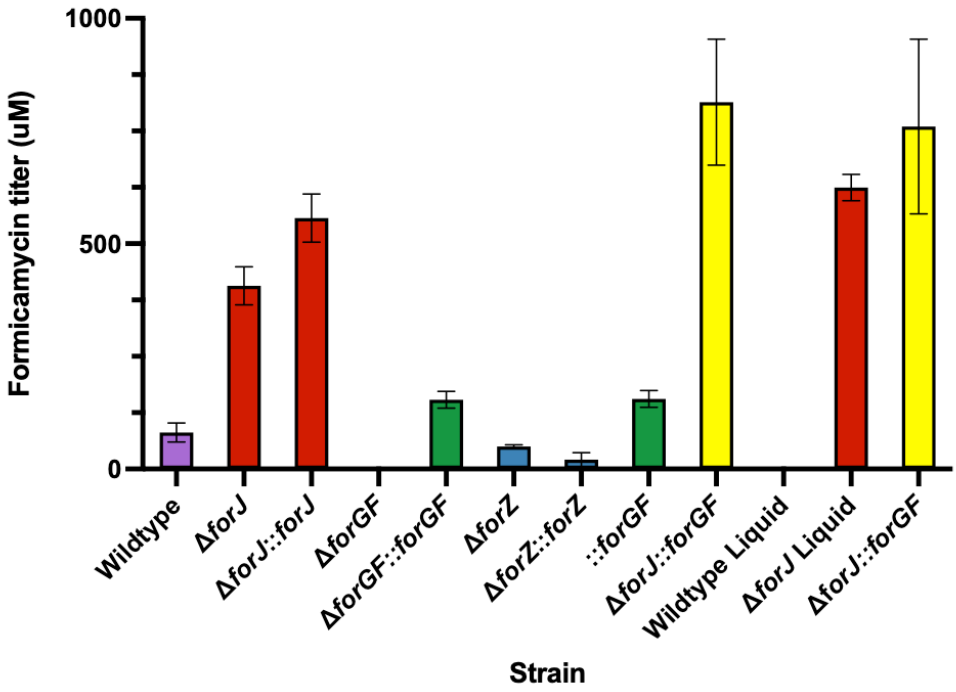


Figure 6.1 Titres of formicamycin production in μM between wildtype and regulator gene deletion strains alongside the complementation of these mutants and rewired strains with multiple mutations. Strains grown in liquid culture are indicated as such. Values are the mean with error bars to represent standard deviation (wildtype n=16, mutants n=3). Adapted from Devine et al., 2021

optimise the production of the formicamycins which hold great potential as bioactive molecules (**Fig 6.1**). For the purpose of this figure, the same previously established colour scheme has been followed, with rewired strains being shown in yellow. Here it was shown that this knowledge of the regulators can be exploited to overproduce the formicamycins and even result in their production in liquid culture, something that is not possible in the wildtype strain. This further allows for strains such as these to progress towards clinical trials and investigations into their possible applications in a wider setting. Without a thorough understanding of the way in which a compound is produced and the regulatory mechanisms that act upon the pathway it would not be possible to achieve this. Data such as these demonstrate the huge potential increases that can be achieved when the functional roles of cluster-situated regulators are so well understood. This also provides a strong basis to pursue investigations into other clusters, such as in **Chapter 3**, where these functions can potentially be exploited with the intention of overproducing the antimicrobial compounds that might otherwise not be seen in wildtype strains.

6.3 Characterisation of the Cluster-Situated Two-Component System ForGF

The primary focus of this work was to establish the role of the TCS ForGF in the regulation of formicamycin biosynthesis and understand how it functions and interacts with other molecules in the cell. As the main activator of the biosynthetic pathway, it is crucial to understand the molecular mechanisms that are involved in signal recognition and transduction to elicit a response in the cell. A range of *in vivo* and *in vitro* investigations have been carried out on various aspects of this process to uncover the ways in which the TCS carries out its regulatory role. Significant steps have been made in developing this understanding including using structural modelling to predict interactions with other molecules in the cell. Preliminary attempts at crystallisation have also been carried out to be able to establish a confirmed structure of both the HK and the RR and this has provided a strong basis for continued trials to obtain crystal structures.

While the activating signal of ForG remains unknown, structural modelling and successful purification from the soluble fraction has confirmed that the HK is atypical and does not contain any transmembrane domains. This has generated the hypothesis that the activating signal must either originate from within the cell or be able to easily cross the membrane to access ForG in the cytosol. Analysis of the other proteins that ForG interacts with in the cell has also provided insight into the source of this activating signal. Strong interaction with a zinc metalloproteinase suggests the potential of an activating signal that must be cleaved in order to act as a ligand for the sensor domain. A functional regulation by proteinases has been documented in other bacterial HKs, and a structural reliance is unlikely based on the separation of the genes in the genome and the ability to purify functional ForG independently of any other protein (C. Zhang et al., 2022). Equally a functional dependence is not likely as the production of formicamycins has been possible through heterologous expression of the formicamycin BGC and has not included this gene. Also observed were interactions with an anchoring protein known to anchor cytosolic proteins to the membrane, and a TPR repeat containing proteins that is capable of mediating the formation of multiprotein complexes (Blatch & Lä Ssle, 1999; Marraffini et al., 2006). The exact significance of each of these proteins is still unknown and the relevance of the interactions is still under investigation. The generation of gene deletion strains using molecular cloning techniques that have been well established in this strain will allow for further investigations to be carried out.

The signal transduction between ForG and ForF has been established through the generation of single point mutations in the conserved residues of each coding sequence. This has established that the HK would normally autophosphorylate as a conserved histidine residue at position 175 in the DHp domain, and when this is altered to a phosphodeficient alanine residue, the HK could not be phosphorylated and formicamycin production was eradicated. Similarly, it was shown that the RR would accept this phosphoryl group at a conserved aspartate residue at position 53 in the receiver domain, but when this is altered to a phosphodeficient alanine residue there is no formicamycin production either. To further support this, when the aspartate residue was altered to a phosphomimetic glutamate residue (D53E) ForF was shown to be constitutively active, with formicamycins once again being produced. These data further demonstrate the lack of cross talk and the specificity of the cognate HK and RR with neither being able to act without the other being present and functional. Further work should be able to demonstrate this phosphorelay using ^{32}P-ATP to radioactively track the phosphoryl group being transferred between the cognate components. It might also be possible to visualise the proteins *in vivo* using fluorescent labelling to determine whether ForG and ForF co-localise within the cell. Based on the relatively new understanding of the links between BGC regulation and oxidative stress alongside the generation of strains that are capable of producing the formicamycins in liquid cultures it is hoped that further attempts at this visualisation will be successful.

This work has also shown that the activity of ForF is likely to be more complex than originally thought. It was shown that the RR would normally bind the intergenic region between divergent promoters that control the transcription of the genes that encode the TCS and the ACC biotin carboxylase and the ACC carboxyl transferase that produce the malonyl CoA starter unit. It was also shown that this binding would normally increase the promoter activity and result in an increase mRNA production of both transcripts. Based on these data and the elucidation of the ForF binding motif using SPR, it was believed that it activated the production of *forHI* to be able to initiate the biosynthetic pathway by producing the malonyl CoA starter units. However, a *forHI* gene deletion strain showed no notable change in the levels of formicamycin produced, suggesting that ForF is in fact activating the pathway through another mechanism.

Analysing the *in vivo* protein-protein interactions of ForF, alongside abundance changes in the proteome has provided a hypothesis of this potential alternative mechanism. As detailed in **Chapter 5**, ForF has been shown to significantly interact with another LuxR family

RR, with evidence of other RRs being able to form heterodimers that can regulate different targets to either of their homodimeric forms (Al-Bassam et al., 2014) Hutchings lab, unpublished data). It is possible that heterodimer formation is occurring with ForF and KY5_4856, the other RR that it has been shown to interact with, with the possibility of the heterodimer regulating different targets to activate formicamycin biosynthesis Similarly, a four gene operon was shown to be significantly upregulated in the absence of ForGF, suggesting that ForF may also be acting to repress the production of these proteins somehow. The exact significance of these proteins to the formicamycin biosynthetic pathway or how an absence of the ForGF TCS results in a change in their abundance is currently unknown, however follow up investigations have been planned to better understand this relationship.

6.4 Final Conclusions

Overall, this project has successfully continued the characterisation of the regulatory mechanisms of formicamycin biosynthesis. Isolation of this novel antibiotic class highlights the wealth of unidentified natural products used by some of the most unlikely insects in some very unusual settings. The ability of the formicamycins to inhibit the growth of clinically relevant pathogens with minimal resistance makes them incredibly promising for future use. Deciphering the action of key regulators and identifying the roles of specific genes within the biosynthetic pathway is crucial in developing a strain that is capable of overproduction. In addition, an understanding of novel regulatory mechanisms used to control the biosynthesis of secondary metabolites has provided further insights into unlocking silent BGCs.

The depth of investigation into the ForGF TCS and other cluster-situated regulators involved in controlling this biosynthetic pathway demonstrate the significant potential for further investigation provided by the preliminary work into the other cluster-situated TCS within *S. formicae*. Overexpression of a cluster-situated TCS directly resulted in a change to the bioactivity in 100% of the five strains generated with associated changed to HPLC analysis. Three strains showed an increase in bacterial activity, the fourth lost all bacterial activity and the fifth showed new anti-fungal activity. Each of these new strains represents an extremely exciting potential for further investigations with the chance of unlocking novel compounds and taking one step closer to overcoming the antimicrobial resistance crisis and Darwin's struggle for existence.

7. Appendix

Appendix Table 1 Strains used and generated during this thesis

Strain	Description	Plasmid (Resistance)	Source
Escherichia coli			
Top10	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*Z ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G		Invitrogen
ET12567	*dam- dcm- hsdS-*	pUZ8002 (Tet)	
S17	*rec A thi pro hsd(R⁻M⁺)* RP4: 2-Tc::Mu-Km::Tn7 *λpir* SM Tp		
NEBα	*fhuA2 Δ(argF-lacZ)*U169 IphoA glnV44 Φ80*lac*Z Δ*M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*		NEB
DH5α	F- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG* Φ80*lac*Z Δ*M15* Δ(lacIZYA-argF)U169 *hsdR17 ($r_K m_K^+$) λ-*		Invitrogen
BL21	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdSλ DE3 = λsBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*		
NiCo21	BL21 derivative, *can:: CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::DB sly::CBD glmS6Ala Δ ΔhsdSλ DE3 = λsBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*		Lab stock
ECO 001-003	BL21 6xHis *forG* C-terminal	pKN007 (Kan)	This work
ECO 004-006	BL21 6xHis *forF* C-terminal	pKN008 (Kan)	This work
ECO 007-009	BL21 6xHis *forG* N-terminal	pKN009 (Kan)	This work
ECO 010-012	BL21 6xHis *forF* N-terminal	pKN010 (Kan)	This work
ECO 013-015	NiCo21 6xHis *forG* C-terminal	pKN007 (Kan)	This work

ECO 016-018	NiCo21 6xHis *forF* C-terminal	pKN008 (Kan)	This work
ECO 019-021	NiCo21 6xHis *forG* N-terminal	pKN009 (Kan)	This work
ECO 022-024	NiCo21 6xHis *forF* N-terminal	pKN010 (Kan)	This work
ECO 025-027	NiCo21 6xHis *forF D53E*	pKN032 (Kan)	This work
ECO 028-030	NiCo21 6xHis *forF D53A*	pKN033 (Kan)	This work
ECO 031-033	NiCo21 6xHis *forF D53N*	pKN034 (Kan)	This work
ECO 034-036	NiCo21 6xHis *forF D53E*	pKN032 (Kan)	This work
ECO 037-039	NiCo21 6xHis *forF D53A*	pKN032 (Kan)	This work
ECO 040-042	NiCo21 6xHis *forF D53N*	pKN033 (Kan)	This work
ECO 043-045	NiCo21 6xHis *forG* codon optimised C-terminal	pKN028 (Kan)	This work, Genewiz
ECO 047-049	NiCo21 6xHis *forF* codon optimised C-terminal	pKN029 (Kan)	This work, Genewiz
ECO 050-052	NiCo21 6xHis *forG* codon optimised N-terminal	pKN028 (Kan)	This work, Genewiz
ECO 053-055	NiCo21 6xHis *forF* codon optimised N-terminal	pKN029 (Kan)	This work, Genewiz
ECO 057-059	NiCo21 6xHis *forF* D53E codon optimised C-terminal	pKN038 (Kan)	This work, Genewiz
ECO 060-062	NiCo21 6xHis *forF* D53A codon optimised C-terminal	pKN039 (Kan)	This work, Genewiz
ECO 063-065	NiCo21 6xHis *forF* D53N codon optimised C-terminal	pKN040 (Kan)	This work, Genewiz
ECO 066-068	NiCo21 6xHis *forF* D53E codon optimised N-terminal	pKN041 (Kan)	This work, Genewiz
ECO 069-071	NiCo21 6xHis *forF* D53A codon optimised N-terminal	pKN042 (Kan)	This work, Genewiz

ECO 072-074	NiCo21 6xHis *forF* D53N codon optimised N-terminal	pKN043 (Kan)	This work, Genewiz
Bioassay Strains			
Bacillus subtilis	Wildtype strain 168, *trpC2*		Gift from Nicola Stanley Wall, University of Dundee
E. coli	Wildtype		Lab stock
Candida albicans	Clinical isolate		Norfolk and Norwich University Hospital (UK)
MRSA	Clinical isolate		Norfolk and Norwich University Hospital (UK)
***Streptomyces formicae* (KY5)**			
KY5 001-005	Wildtype		Lab stock
KY5 006-008	Δ*for*	pRD050	Dr Rebecca Devine (Hutchings lab)
KY5 009-011	Δ*forJ*	pRD026	Dr Rebecca Devine (Hutchings lab)
KY5 012-014	Δ*forGF*	pRD027	Dr Rebecca Devine (Hutchings lab)
KY5 015-017	Δ*forZ*	pRD028	Dr Rebecca Devine (Hutchings lab)

KY5 018-020	Δ*forJ: ΦBT1 forJ* p*forM*	pRD030 (Hyg)	Dr Rebecca Devine (Hutchings lab)
KY5 021-023	Δ*forGF: ΦBT1 forGF* p*forG*	pRD031 (Hyg)	Dr Rebecca Devine (Hutchings lab)
KY5 024-026	Δ*forZ: ΦBT1 forZ* p*forZ*	pRD032 (Hyg)	Dr Rebecca Devine (Hutchings lab)
KY5 027-029	Δ*forJ: ΦBT1 forJ 3xFlag*	pRD034 (Hyg)	Dr Rebecca Devine (Hutchings lab)
KY5 030-032	Δ*forGF: ΦBT1 forGF 3xFlag*	pRD035 (Hyg)	Dr Rebecca Devine (Hutchings lab)
KY5 033-035	Δ*forZ: ΦBT1 forZ 3xFlag*	pRD036 (Hyg)	Dr Rebecca Devine (Hutchings lab)
KY5 036-038	KY5 *ΦBT1 ermE* gus*	pMF96 (Hyg)	This work
KY5 039-041	KY5 *ΦC31 gus*	pMF23 (Apr)	This work
KY5 042-044	KY5 *ΦBT1* p*forM gus*	pKN011 (Hyg)	This work
KY5 045-047	KY5 *ΦBT1* p*forH gus*	pKN012 (Hyg)	This work
KY5 048-050	KY5 *ΦBT1* p*forG gus*	pKN013 (Hyg)	This work
KY5 051-053	KY5 *ΦBT1* p*forT gus*	pKN014 (Hyg)	This work
KY5 054-056	KY5 *ΦBT1* p*forU gus*	pKN015 (Hyg)	This work

KY5 057-059	KY5 *ΦBT1* p*forZ gus*	pKN016 (Hyg)	This work
KY5 060-062	KY5 *ΦBT1* p*forAA gus*	pKN017 (Hyg)	This work
KY5 063-065	Δ*forJ ΦBT1 ermE* gus*	pMF96 (Hyg)	This work
KY5 066-068	Δ*forJ ΦC31 gus*	pMF23 (Apr)	This work
KY5 069-071	Δ*forJ ΦBT1* p*forM gus*	pKN011 (Hyg)	This work
KY5 072-074	Δ*forJ ΦBT1* p*forH gus*	pKN012 (Hyg)	This work
KY5 075-077	Δ*forJ ΦBT1* p*forG gus*	pKN013 (Hyg)	This work
KY5 078-080	Δ*forJ ΦBT1* p*forT gus*	pKN014 (Hyg)	This work
KY5 081-083	Δ*forJ ΦBT1* p*forU gus*	pKN015 (Hyg)	This work
KY5 084-086	Δ*forJ ΦBT1* p*forZ gus*	pKN016 (Hyg)	This work
KY5 087-089	Δ*forJ ΦBT1* p*forAA gus*	pKN017 (Hyg)	This work
KY5 090-092	Δ*forGF ΦBT1 ermE* gus*	pMF96 (Hyg)	This work
KY5 093-095	Δ*forGF ΦC31 gus*	pMF23 (Apr)	This work
KY5 096-097	Δ*forGF ΦBT1* p*forM gus*	pKN011 (Hyg)	This work
KY5 098-100	Δ*forGF ΦBT1* p*forH gus*	pKN012 (Hyg)	This work
KY5 101-103	Δ*forGF ΦBT1* p*forG gus*	pKN013 (Hyg)	This work
KY5 104-106	Δ*forGF ΦBT1* p*forT gus*	pKN014 (Hyg)	This work
KY5 107-109	Δ*forGF ΦBT1* p*forU gus*	pKN015 (Hyg)	This work

KY5 110-112	Δ*forGF ΦBT1* p*forZ gus*	pKN016 (Hyg)	This work
KY5 113-115	Δ*forGF ΦBT1* p*forAA gus*	pKN017 (Hyg)	This work
KY5 116-118	Δ*forZ ΦBT1 ermE* gus*	pMF96 (Hyg)	This work
KY5 119-121	Δ*forZ ΦC31 gus*	pMF23 (Apr)	This work
KY5 122-124	Δ*forZ ΦBT1* p*forM gus*	pKN011 (Hyg)	This work
KY5 125-127	Δ*forZ ΦBT1* p*forH gus*	pKN012 (Hyg)	This work
KY5 128-130	Δ*forZ ΦBT1* p*forG gus*	pKN013 (Hyg)	This work
KY5 131-133	Δ*forZ ΦBT1* p*forT gus*	pKN014 (Hyg)	This work
KY5 134-136	Δ*forZ ΦBT1* p*forU gus*	pKN015 (Hyg)	This work
KY5 137-139	Δ*forZ ΦBT1* p*forZ gus*	pKN016 (Hyg)	This work
KY5 140-142	Δ*forZ ΦBT1* p*forAA gus*	pKN017 (Hyg)	This work
KY5 143-145	Δ*forG*	pKN001	This work
KY5 146-148	Δ*forF*	pKN002	This work
KY5 149-151	Δ*forG: ΦBT1 forG 3xFlag*	pKN018 (Hyg)	This work
KY5 152-154	Δ*forF: ΦBT1 forF 3xFlag*	pKN019 (Hyg)	This work
KY5 155-157	Δ*forF: ΦBT1 forF D53E*	pKN044 (Hyg)	This work
KY5 158-160	Δ*forF: ΦBT1 forF D53A*	pKN045 (Hyg)	This work
KY5 161-163	Δ*forF: ΦBT1 forF D53N*	pKN046 (Hyg)	This work

KY5 165-167	Δ*forG: ΦBT1 forG H175A*	pKN047 (Hyg)	This work
KY5 168-170	Δ*forG: ΦBT1 forG D176A*	pKN048 (Hyg)	This work
KY5 171-173	*forF D53E*	pKN020	This work
KY5 174-176	*forF D53A*	pKN021	This work
KY5 177-179	*forF D53N*	pKN022	This work
KY5 180-182	*forG H175A*	pKN023	This work
KY5 183-185	*forG D176A*	pKN024	This work
KY5 186-188	1516/1517	pKN049 (Hyg)	This work
KY5 189-191	6759/6760	pKN050 (Hyg)	This work
KY5 192-194	7292/7292	pKN051 (Hyg)	This work
KY5 195-197	7446/7447	pKN052 (Hyg)	This work
KY5 198-200	8002/8003	pKN053 (Hyg)	This work
KY5 201-203	Δ*forGF* 1516/1517	pKN049 (Hyg)	This work
KY5 204-207	Δ*forGF* 6759/6760	pKN050 (Hyg)	This work
KY5 208-210	Δ*forGF* 7292/7292	pKN051 (Hyg)	This work
KY5 211-213	Δ*forGF* 7446/7447	pKN052 (Hyg)	This work
KY5 215-217	Δ*forGF* 8002/8003	pKN053 (Hyg)	This work

KY5 218-220	Δ*forF: ΦBT1 forF mCherry*	pKN054 (Hyg)	This work, Susan Schlimpert
KY5 221-223	Δ*forF: ΦBT1 forF mTurquoise*	pKN055 (Hyg)	This work, Susan Schlimpert
KY5 224-227	Δ*forG: ΦBT1 forG mCherry*	pKN056 (Hyg)	This work, Susan Schlimpert
KY5 228-230	Δ*forG: ΦBT1 forG mTurquoise*	pKN057 (Hyg)	This work, Susan Schlimpert

Appendix Table 2 Primers used and generated during this thesis

Name	Description	Sequence
KN001	ForG KO sgRNA FOR	ACGCaggcgccagtcggcctggta
KN002	ForG KO sgRNA REV	AAACtaccaggccgactggcgcct
KN003	ForF KO sgRNA FOR	ACGCtggcgaagatgttgcgcaga
KN004	ForF KO sgRNA REV	AAACtctgcgcaacatcttcgcca
KN005	ForHI KO sgRNA FOR	ACGCccggatgctccacctggata
KN006	ForHI KO sgRNA REV	AAACtatccaggtggagcatccgg
KN007	ForF KO Repair 1 FOR	gctcggttgccgccgggcgtttttaTCTAGAaccactacgaacagcgcctg
KN008	ForF KO Repair 1 REV	GCTGCTGCGACCAGGCGAGCTCGCcatccgcgcctcatccgctc
KN009	ForF KO Repair 2 FOR	GCGAGCTCGCCTGGTCGCAGCAGCtgaggctcaggcgggttcga
KN010	ForF KO Repair 2 REV	gcaacgcggccttttacggttcctggccTCTAGAatcgtcgccgagctggagaa
KN011	ForG KO Repair 1 FOR	gctcggttgccgccgggcgtttttaTCTAGAgacgcggtgacgacgaggaa
KN012	ForG KO Repair 1 REV	GCTGCTGCGACCAGGCGAGCTCGCggcagcctcgttcacagcag
KN013	ForG KO Repair 2 FOR	GCGAGCTCGCCTGGTCGCAGCAGCtgaggcgcggatgcagaccg
KN014	ForG KO Repair 2 REV	gcaacgcggccttttacggttcctggccTCTAGAcgagcgcctggtcagactcg
KN015	ForHI KO Repair 1 FOR	gctcggttgccgccgggcgtttttaTCTAGAgttctgcttggccacgctca
KN016	ForHI KO Repair 1 REV	GCTGCTGCGACCAGGCGAGCTCGCgctgctcacggtcatcgtgt

KN017	ForHI KO Repair 2 FOR	GCGAGCTCGCCTGGTCGCAGCAGCgaggcggaccgtgcctaggc
KN018	ForHI KO Repair 2 REV	gcaacgcggccttttacggttcctggccTCTAGAtacgtcggcctcaagaccga
KN019	ForHI KO Repair 3 FOR	gctcggttgccgccgggcgtttttaTCTAGAgcaccagttcgacgttctgcttgg
KN020	ForHI KO Repair 3 REV	GCTGCTGCGACCAGGCGAGCTCGCgtgtggtcgggcgctgctcacggtcat
KN021	ForHI KO Repair 4 FOR	GCGAGCTCGCCTGGTCGCAGCAGCgaggcggaccgtgcctaggc
KN022	ForHI KO Repair 5 REV	gcaacgcggccttttacggttcctggccTCTAGAcgtgctggtcggcgtgctgc
KN023	ForG pET28a TEST FOR	CTGGTGCCGCGCGGCAGCCATATGatgaacgaggctgccactgatc
KN024	ForG pET28a TEST REV	GTGCTCGAGTGCGGCCGCAAGCTTtcatccgctcttcgtgggc
KN025	ForF pET28a TEST FOR	CTGGTGCCGCGCGGCAGCCATATGatgcagaccgtggtgacccg
KN026	ForF pET28a TEST REV	GTGCTCGAGTGCGGCCGCAAGCTTtcagccccggtcgccctg
KN027	pMF96 TEST FOR	gctcaatcaatcaccggatcc
KN028	pMF96 TEST REV	catgtccgtacctccgttg
KN029	p*forM* FOR	AAAAAcatatgtcttcggcgcacgacagacc
KN030	p*forM* REV	AAAAActcgagaccggctcccatcggttgc
KN031	p*forH* FOR	AAAAAcatatggcgctgctcacggtcatcg
KN032	p*forH* REV	AAAAActcgaggcagcctcgttcacagcag
KN033	p*forG* FOR	AAAAAcatatggcagcctcgttcacagcagc
KN034	p*forG* REV	AAAAActcgaggcgctgctcacggtcatcg

KN035	p*forT* FOR	AAAAAcatatgggcgaagccgaggctcatgc
KN036	p*forT* REV	AAAAActcgagcagatcgaccagcttctgctggtcc
KN037	p*forU* FOR	AAAAAcatatgcagatcgaccagcttctgctggtcc
KN038	p*forU* REV	AAAAActcgagggcgaagccgaggctcatgc
KN039	p*forZ* FOR	AAAAAcatatggaatccctgacgcgccgcg
KN040	p*forZ* REV	AAAAActcgaggacgatggtggtgtcgagcac
KN041	p*forAA* FOR	AAAAAcatatggacgatggtggtgtcgagcacc
KN042	p*forAA* REV	AAAAActcgaggaatccctgacgcgccgcg
KN043	BGC 30 1 FOR	GCTCGTGGGCGCTCAGCCGGTGGTGGTGGCCGCAGTCGGG
KN044	BGC 30 1 REV	CCCGACTGCGGCCACCACCACCGGCTGAGCGCCCACGAGCcctaccctacgtcctcctgc
KN045	BGC 30 2 FOR	GTGGCCGCAGTCGGGGCAGACCTGGAGGGCGCGCCGCAGC
KN046	BGC 30 2 REV	GCTGCGGCGCGCCCTCCAGGTCTGCCCCGACTGCGGCCACcctaccctacgtcctcctgc
KN047	BGC 30 3 FOR	AGGGCGCGCCGCAGCCGCGTGCGGTAGACGACCCCGTGGC
KN048	BGC 30 3 REV	GCCACGGGGTCGTCTACCGCACGCGGCTGCGGCGCGCCCTcctaccctacgtcctcctgc
KN049	BGC 30 4 FOR	AGACGACCCCGTGGCAGGCGTCGCAGACGGCCCAGTCGGA
KN050	BGC 30 4 REV	TCCGACTGGGCCGTCTGCGACGCCTGCCACGGGGTCGTCTcctaccctacgtcctcctgc
KN051	BGC 30 5 FOR	GACGGCCCAGTCGGAGTCGGAGTCGGAGTCGGTTCCGGTG
KN052	BGC 30 5 REV	CACCGGAACCGACTCCGACTCCGACTCCGACTGGGCCGTCcctaccctacgtcctcctgc

KN053	BGC 30 6 FOR	GAGTCGGTTCCGGTGTTCTGCGCCGGGGACCCGCGCCCCG
KN054	BGC 30 6 REV	CGGGGCGCGGGTCCCCGGCGCAGAACACCGGAACCGACTCcctaccctacgtcctcctgc
KN055	BGC 30 7 FOR	GGGACCCGCGCCCCGCCCCGGCCCGGTCGGCCGGGTCGGC
KN056	BGC 30 7 REV	GCCGACCCGGCCGACCGGGCCGGGGCGGGGCGCGGGTCCCcctaccctacgtcctcctgc
KN057	BGC 30 8 FOR	GTCGGCCGGGTCGGCATGGTGTGGTCGGGCGCTGCTCACG
KN058	BGC 30 8 REV	CGTGAGCAGCGCCCGACCACACCATGCCGACCCGGCCGACcctaccctacgtcctcctgc
KN059	BGC 30 9 FOR	CGGGCGCTGCTCACGGTCATCGTGTACCCCCTGTGCACGA
KN060	BCG 30 9 REV	TCGTGCACAGGGGGTACACGATGACCGTGAGCAGCGCCCGcctaccctacgtcctcctgc
KN061	BCG 30 10 FOR	ACCCCCTGTGCACGAAGCCTGCGTGATTCATCGGCTGCGC
KN062	BGC 30 10 REV	GCGCAGCCGATGAATCACGCAGGCTTCGTGCACAGGGGGTcctaccctacgtcctcctgc
KN063	BGC 30 11 FOR	ATTCATCGGCTGCGCCCGACCCTAATCGCGGCCGGTGGGG
KN064	BCG 30 11 REV	CCCCACCGGCCGCGATTAGGGTCGGGCGCAGCCGATGAATcctaccctacgtcctcctgc
KN065	BCG 30 12 FOR	TCGCGGCCGGTGGGGGCGCTCAACGCATCCGGCCCAAGGA
KN066	BCG 30 12 REV	TCCTTGGGCCGGATGCGTTGAGCGCCCCCACCGGCCGCGAcctaccctacgtcctcctgc
KN067	BCG 30 13 FOR	CATCCGGCCCAAGGACAGATCCGGACAGTAGGGGGTGGCA
KN068	BGC 30 13 REV	TGCCACCCCCTACTGTCCGGATCTGTCCTTGGGCCGGATGcctaccctacgtcctcctgc
KN069	BGC 30 14 FOR	CAGTAGGGGGTGGCATGGCGGTCTATGGCACACGGGGGAG
KN070	BCG 30 14 REV	CTCCCCCGTGTGCCATAGACCGCCATGCCACCCCCTACTGcctaccctacgtcctcctgc

KN071	BCG 30 15 FOR	TGGCACACGGGGGAGGGGACATCCCTCCTTCGGCAGAGCC
KN072	BCG 30 15 REV	GGCTCTGCCGAAGGAGGGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN073	BGC 30 16 FOR	TCCTTCGGCAGAGCCGCCGCGGGTCGGACGGGGGACGGAA
KN074	BCG 30 16 REV	TTCCGTCCCCCGTCCGACCCGCGGCGGCTCTGCCGAAGGAcctaccctacgtcctcctgc
KN075	BGC 30 17 FOR	GGACGGGGGACGGAACCTGTCCCAGTGGGCGGGGCCGCTG
KN076	BCG 30 17 REV	CAGCGGCCCCGCCCACTGGGACAGGTTCCGTCCCCCGTCCcctaccctacgtcctcctgc
KN077	BGC 30 18 FOR	TGGGCGGGGCCGCTGTCCGAATCGGCATATTGCGACATTG
KN078	BCG 30 18 REV	CAATGTCGCAATATGCCGATTCGGACAGCGGCCCCGCCCAcctaccctacgtcctcctgc
KN079	BCG 30 19 FOR	CATATTGCGACATTGACGGCGTGCGGGATGACCGGAGTAA
KN080	BCG 30 19 REV	TTACTCCGGTCATCCCGCACGCCGTCAATGTCGCAATATGcctaccctacgtcctcctgc
KN081	BGC 30 20 FOR	GGATGACCGGAGTAACGTCACTCGTAGCGTGGCGGGACCC
KN082	BGC 30 20 REV	GGGTCCCGCCACGCTACGAGTGACGTTACTCCGGTCATCCcctaccctacgtcctcctgc
KN083	BGC 30 21 FOR	AGCGTGGCGGGACCCTCAGGTGACACTTACCTGCCTGCCG
KN084	BGC 30 21 REV	CGGCAGGCAGGTAAGTGTCACCTGAGGGTCCCGCCACGCTcctaccctacgtcctcctgc
KN085	BGC 30 22 FOR	CTTACCTGCCTGCCGCGGGCCGAAGGCGGAGGGGCAGTGC
KN086	BGC 30 22 REV	GCACTGCCCCTCCGCCTTCGGCCCGCGGCAGGCAGGTAAGcctaccctacgtcctcctgc
KN087	BGC 30 23 FOR	GCGGAGGGGCAGTGCTGAACGACGATGATCGCGCACTCGA
KN088	BGC 30 23 REV	TCGAGTGCGCGATCATCGTCGTTCAGCACTGCCCCTCCGCcctaccctacgtcctcctgc

KN089	BGC 30 24 FOR	TGATCGCGCACTCGACCAGTACGGGGGACCGTGACCGGCA
KN090	BGC 30 24 REV	TGCCGGTCACGGTCCCCCGTACTGGTCGAGTGCGCGATCAcctaccctacgtcctcctgc
KN091	BGC 30 25 FOR	GGACCGTGACCGGCACACAGTGGTGGCCTGCCCGTGGGCC
KN092	BCG 30 25 REV	GGCCCACGGGCAGGCCACCACTGTGTGCCGGTCACGGTCCcctaccctacgtcctcctgc
KN093	BGC 30 26 FOR	GCCTGCCCGTGGGCCCGCACCCGCCCGGCCGGCGGGAGTC
KN094	BCG 30 26 REV	GACTCCCGCCGGCCGGGCGGGTGCGGGCCCACGGGCAGGCcctaccctacgtcctcctgc
KN095	BGC 30 27 FOR	CGGCCGGCGGGAGTCCGCGGCGCGCACCCGCACGCCCCGC
KN096	BGC 30 27 REV	GCGGGGCGTGCGGGTGCGCGCCGCGGACTCCCGCCGGCCGcctaccctacgtcctcctgc
KN097	BGC 30 28 FOR	ACCCGCACGCCCCGCCCTCCCGCGAGACCGCGCGACCGAA
KN098	BGC 30 28 REV	TTCGGTCGCGCGGTCTCGCGGGAGGGCGGGGCGTGCGGGTcctaccctacgtcctcctgc
KN099	BGC 30 29 FOR	GACCGCGCGACCGAACCGGGGAGGCTGCTGTGAACGAGGC
KN100	BGC 30 29 REV	GCCTCGTTCACAGCAGCCTCCCCGGTTCGGTCGCGCGGTCcctaccctacgtcctcctgc
KN101	BGC 30 30 FOR	TGCTGTGAACGAGGCTGCCACTGATCACTTAAGCCGCGGA
KN102	BGC 30 30 REV	TCCGCGGCTTAAGTGATCAGTGGCAGCCTCGTTCACAGCAcctaccctacgtcctcctgc
KN103	BGC 30 31 FOR	CACTTAAGCCGCGGACCCGAGATCGCCGAGGCCCTCGGGG
KN104	BGC 30 31 REV	CCCCGAGGGCCTCGGCGATCTCGGGTCCGCGGCTTAAGTGcctaccctacgtcctcctgc
KN105	BGC 30 32 FOR	CCGAGGCCCTCGGGGCCCGTCTCGAGTCCGTGCTCCACCA
KN106	BGC 30 32 REV	TGGTGGAGCACGGACTCGAGACGGGCCCCGAGGGCCTCGGcctaccctacgtcctcctgc

KN107	BGC 30 33 FOR	GTCCGTGCTCCACCACTACGAACAGCGCCTGCTCGCCTCG
KN108	BGC 30 33 REV	CGAGGCGAGCAGGCGCTGTTCGTAGTGGTGGAGCACGGACcctaccctacgtcctcctgc
KN109	BGC 30 34 FOR	GTGCTCCACCACTACGAACAGCGCCTGCTCGCCTCGGGCG
KN110	BGC 30 34 REV	CGCCCGAGGCGAGCAGGCGCTGTTCGTAGTGGTGGAGCACcctaccctacgtcctcctgc
KN111	BGC 4 1 FOR	GCGAAGGACCCGCCGCGTCACCGTGCCACCCCCGTCACAG
KN112	BGC 4 1 REV	CTGTGACGGGGGTGGCACGGTGACGCGGCGGGTCCTTCGCcctaccctacgtcctcctgc
KN113	BGC 4 2 FOR	CCACCCCCGTCACAGAGTTGATCACGCCCCTCGGCCAGAC
KN114	BGC 4 2 REV	GTCTGGCCGAGGGGCGTGATCAACTCTGTGACGGGGGTGGcctaccctacgtcctcctgc
KN115	BGC 4 3 FOR	GCCCCTCGGCCAGACCCCTGTGGCTATGCGCCCGTTGACC
KN116	BGC 4 3 REV	GGTCAACGGGCGCATAGCCACAGGGGTCTGGCCGAGGGGCcctaccctacgtcctcctgc
KN117	BGC 4 4 FOR	ATGCGCCCGTTGACCGCCACCGAAGGCCCGAAACCCGTCG
KN118	BGC 4 4 REV	CGACGGGTTTCGGGCCTTCGGTGGCGGTCAACGGGCGCATcctaccctacgtcctcctgc
KN119	BGC 4 5 FOR	GCCCGAAACCCGTCGCGCCGCACTCTCCGGAACGTCCGGA
KN120	BGC 4 5 REV	TCCGGACGTTCCGGAGAGTGCGGCGCGACGGGTTTCGGGCcctaccctacgtcctcctgc
KN121	BGC 4 6 FOR	TCCGGAACGTCCGGAGCCGTGTTGAAACGGACCAGACGAG
KN122	BGC 4 6 REV	CTCGTCTGGTCCGTTTCAACACGGCTCCGGACGTTCCGGAcctaccctacgtcctcctgc
KN123	BGC 4 7 FOR	AACGGACCAGACGAGCTACCCCGAACGGCGTAGCGAAGGG
KN124	BGC 4 7 REV	CCCTTCGCTACGCCGTTCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc

KN125	BGC 4 8 FOR	CGGCGTAGCGAAGGGTCCGCCAAGGCGTGGTCGAGAGTTG
KN126	BGC 4 8 REV	CAACTCTCGACCACGCCTTGGCGGACCCTTCGCTACGCCGcctaccctacgtcctcctgc
KN127	BGC 4 9 FOR	CGTGGTCGAGAGTTGCGTCAGAGGTGTACGAGGAGTGCGG
KN128	BCG 4 9 REV	CCGCACTCCTCGTACACCTCTGACGCAACTCTCGACCACGcctaccctacgtcctcctgc
KN129	BCG 4 10 FOR	GTACGAGGAGTGCGGAACGGACCATTCCGGTGCTGGTCGA
KN130	BGC 4 10 REV	TCGACCAGCACCGGAATGGTCCGTTCCGCACTCCTCGTACcctaccctacgtcctcctgc
KN131	BGC 4 11 FOR	TCCGGTGCTGGTCGAACCGTCGTCGGCCACCCGGACGGCT
KN132	BCG 4 11 REV	AGCCGTCCGGGTGGCCGACGACGGTTCGACCAGCACCGGAcctaccctacgtcctcctgc
KN133	BCG 4 12 FOR	GCCACCCGGACGGCTGCCCCTTTCGGGCGGTGTCCACCTG
KN134	BCG 4 12 REV	CAGGTGGACACCGCCCGAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN135	BCG 4 13 FOR	GGCGGTGTCCACCTGAACGGGTCAGTCGTGAGGAGCCTCA
KN136	BGC 4 13 REV	TGAGGCTCCTCACGACTGACCCGTTCAGGTGGACACCGCCcctaccctacgtcctcctgc
KN137	BGC 4 14 FOR	TCGTGAGGAGCCTCACAAAACGCCGTTTCGGCTCGGTTTT
KN138	BCG 4 14 REV	AAAACCGAGCCGAAACGGCGTTTTGTGAGGCTCCTCACGAcctaccctacgtcctcctgc
KN139	BCG 4 15 FOR	TTTCGGCTCGGTTTTTTCGCCCTTGTCGGGTCAAGATCCT
KN140	BCG 4 15 REV	AGGATCTTGACCCGACAAGGGCGAAAAAACCGAGCCGAAAcctaccctacgtcctcctgc
KN141	BGC 4 16 FOR	TCGGGTCAAGATCCTCTCCGACGACAAGCCCCCGCCACAG
KN142	BCG 4 16 REV	CTGTGGCGGGGGCTTGTCGTCGGAGAGGATCTTGACCCGAcctaccctacgtcctcctgc

KN143	BGC 4 17 FOR	AAGCCCCCGCCACAGCGGCGGGGCGGTCCGGGCGGACGCC
KN144	BCG 4 17 REV	GGCGTCCGCCCGGACCGCCCCGCCGCTGTGGCGGGGGCTTcctaccctacgtcctcctgc
KN145	BGC 4 18 FOR	GTCCGGGCGGACGCCGAGTCCTGCCGCCACCCGGATGACC
KN146	BCG 4 18 REV	GGTCATCCGGGTGGCGGCAGGACTCGGCGTCCGCCCGGACcctaccctacgtcctcctgc
KN147	BCG 4 19 FOR	GCCACCCGGATGACCGGTCGACATCGGTGCAACGGCAGGA
KN148	BCG 4 19 REV	TCCTGCCGTTGCACCGATGTCGACCGGTCATCCGGGTGGCcctaccctacgtcctcctgc
KN149	BGC 4 20 FOR	GCCACCCGGATGACCGGTCGACATCGGTGCAACGGCAGGA
KN150	BGC 4 20 REV	TCCTGCCGTTGCACCGATGTCGACCGGTCATCCGGGTGGCcctaccctacgtcctcctgc
KN151	6768600 1 FOR	GCCGCTCCACGGTCCCCCCACCCCCCACCCAAGGTGTTTC
KN152	6768600 1 REV	GAAACACCTTGGGTGGGGGGTGGGGGGACCGTGGAGCGGCcctaccctacgtcctcctgc
KN153	6768600 2 FOR	CACCCAAGGTGTTTCCATCGGCCCGCGCCGATTCACCCCT
KN154	6768600 2 REV	AGGGGTGAATCGGCGCGGGCCGATGGAAACACCTTGGGTGcctaccctacgtcctcctgc
KN155	6768600 3 FOR	CGCCGATTCACCCCTTCGGGCAGGCGCTCGGAGTGAACCA
KN156	6768600 3 REV	TGGTTCACTCCGAGCGCCTGCCCGAAGGGGTGAATCGGCGcctaccctacgtcctcctgc
KN157	6768600 4 FOR	GCTCGGAGTGAACCAGGGGAGCGGAAGGGGGCGCGTTCGC
KN158	6768600 4 REV	GCGAACGCGCCCCCTTCCGCTCCCCTGGTTCACTCCGAGCcctaccctacgtcctcctgc
KN159	6768600 5 FOR	AGGGGGCGCGTTCGCGCCGCGTTCTGTCCCGCTGATCACC
KN160	6768600 5 REV	GGTGATCAGCGGGACAGAACGCGGCGCGAACGCGCCCCCTcctaccctacgtcctcctgc

KN161	6768600 6 FOR	GTCCCGCTGATCACCGCCGACCTACGGTCGGAAGGACGGG
KN162	6768600 6 REV	CCCGTCCTTCCGACCGTAGGTCGGCGGTGATCAGCGGGACcctaccctacgtcctcctgc
KN163	6768600 7 FOR	CCGACCTACGGTCGGAAGGACGGGGCGCGGCGGGGAGGTT
KN164	6768600 7 REV	AACCTCCCCGCCGCGCCCCGTCCTTCCGACCGTAGGTCGGcctaccctacgtcctcctgc
KN165	6814000 1 FOR	GGATGGCGTGCTGGCCGTGATCCGCGGGGAGGCTCAGCCG
KN166	6814000 1 REV	CGGCTGAGCCTCCCCGCGGATCACGGCCAGCACGCCATCCcctaccctacgtcctcctgc
KN167	6814000 2 FOR	GGGGAGGCTCAGCCGTAAGGGTGGGCGGCTCACCCGTATG
KN168	6814000 2 REV	CATACGGGTGAGCCGCCCACCCTTACGGCTGAGCCTCCCCcctaccctacgtcctcctgc
KN169	6814000 3 FOR	CGGCTCACCCGTATGGGATCGCCGGGCAACCGTTTCCCTT
KN170	6814000 3 REV	AAGGGAAACGGTTGCCCGGCGATCCCATACGGGTGAGCCGcctaccctacgtcctcctgc
KN171	6814000 4 FOR	GCAACCGTTTCCCTTCGGCGTACGTCAAGTTGAGGGGAAG
KN172	6814000 4 REV	CTTCCCCTCAACTTGACGTACGCCGAAGGGAAACGGTTGCcctaccctacgtcctcctgc
KN173	6814000 5 FOR	CAAGTTGAGGGGAAGAGCAGCACGATGTCCCACCATCAGG
KN174	6814000 5 REV	CCTGATGGTGGGACATCGTGCTGCTCTTCCCCTCAACTTGcctaccctacgtcctcctgc
KN175	6814000 6 FOR	TGTCCCACCATCAGGATGTCCGGCCATTCTGATGGCGTAC
KN176	6814000 6 REV	GTACGCCATCAGAATGGCCGGACATCCTGATGGTGGGACAcctaccctacgtcctcctgc
KN177	6814000 7 FOR	ATTCTGATGGCGTACAAACTCCGTCCGCCCGGTTGCCCCA
KN178	6814000 7 REV	TGGGGCAACCGGGCGGACGGAGTTTGTACGCCATCAGAATcctaccctacgtcctcctgc

KN179	6814000 8 FOR	CGCCCGGTTGCCCCACCCTTTCAAGACCCCCATAAAATGA
KN180	6814000 8 REV	TCATTTTATGGGGGTCTTGAAAGGGTGGGGCAACCGGGCGcctaccctacgtcctcctgc
KN181	6814000 9 FOR	ACCCCCATAAAATGAACCTCCGGCAGGTGGCTACACGCTC
KN182	6814000 9 REV	GAGCGTGTAGCCACCTGCCGGAGGTTCATTTTATGGGGGTcctaccctacgtcctcctgc
KN183	6814000 10 FOR	GGTGGCTACACGCTCGAAAGGCGCCCCCGTGTCCATCGGC
KN184	6814000 10 REV	GCCGATGGACACGGGGGCGCCTTTCGAGCGTGTAGCCACCcctaccctacgtcctcctgc
KN185	6814000 11 FOR	CCCGTGTCCATCGGCAACTCCCCTGAAGACGACCGCCCTT
KN186	6814000 11 REV	AAGGGCGGTCGTCTTCAGGGGAGTTGCCGATGGACACGGGcctaccctacgtcctcctgc
KN187	6814000 12 FOR	AAGACGACCGCCCTTCAGACGACCGCCTCGCGGACGAGAT
KN188	6814000 12 REV	ATCTCGTCCGCGAGGCGGTCGTCTGAAGGGCGGTCGTCTTcctaccctacgtcctcctgc
KN189	6814000 13 FOR	CCTCGCGGACGAGATCCCTTCGGGTGACCGTCCGGCCGAC
KN190	6814000 13 REV	GTCGGCCGGACGGTCACCCGAAGGGATCTCGTCCGCGAGGcctaccctacgtcctcctgc
KN191	6814000 14 FOR	GACCGTCCGGCCGACGACCGGCCCTCGATCGGCCAGGTCC
KN192	6814000 14 REV	GGACCTGGCCGATCGAGGGCCGGTCGTCGGCCGGACGGTCcctaccctacgtcctcctgc
KN193	6814000 15 FOR	CGATCGGCCAGGTCCTGCAGCAGGCGCGCGTCGCCGCCGG
KN194	6814000 15 REV	CCGGCGGCGACGCGCGCCTGCTGCAGGACCTGGCCGATCGcctaccctacgtcctcctgc
KN195	6814000 16 FOR	GCGCGTCGCCGCCGGCCTGACCGTCGACGAGGTCAGCTCG
KN196	6814000 16 REV	CGAGCTGACCTCGTCGACGGTCAGGCCGGCGGCGACGCGCcctaccctacgtcctcctgc

KN197	6814000 17 FOR	GACGAGGTCAGCTCGTCCACCCGCGTGCGCATTCCGATCG
KN198	6814000 17 REV	CGATCGGAATGCGCACGCGGGTGGACGAGCTGACCTCGTCcctaccctacgtcctcctgc
KN199	6814000 18 FOR	GAGGTCAGCTCGTCCACCCGCGTGCGCATTCCGATCGTGC
KN200	6814000 18 REV	GCACGATCGGAATGCGCACGCGGGTGGACGAGCTGACCTCcctaccctacgtcctcctgc
KN201	9400100 1 REV	CGCTCTTCGTGGGTGCGTTGATAGGTCAGGTACTGGCCGT
KN202	9400100 2 FOR	ACGGCCAGTACCTGACCTATCAACGCACCCACGAAGAGCGcctaccctacgtcctcctgc
KN203	9400100 2 REV	TCAGGTACTGGCCGTAGAGGTAGAGCCCGACCGGGTACGA
KN204	9400100 3 FOR	TCGTACCCGGTCGGGCTCTACCTCTACGGCCAGTACCTGAcctaccctacgtcctcctgc
KN205	9400100 3 REV	CCCGACCGGGTACGACCAGCCGCCGATGGTGGTGGGTGTG
KN206	9400100 4 FOR	CACACCCACCACCATCGGCGGCTGGTCGTACCCGGTCGGGcctaccctacgtcctcctgc
KN207	9400100 4 REV	ATGGTGGTGGGTGTGGAGCGGGCGAGGGTCGAATCGACCA
KN208	9400100 5 REV	TGGTCGATTCGACCCTCGCCCGCTCCACACCCACCACCATcctaccctacgtcctcctgc
KN209	9400100 6 FOR	GGGTCGAATCGACCAGGGAGGTGGACCAGTCGGTGGTGTT
KN210	9400100 6 REV	AACACCACCGACTGGTCCACCTCCCTGGTCGATTCGACCCcctaccctacgtcctcctgc
KN211	9400100 7 FOR	CCAGTCGGTGGTGTTCCTGTTCAGGGATGCGGACATGGTC
KN212	9400100 7 REV	GACCATGTCCGCATCCCTGAACAGGAACACCACCGACTGGcctaccctacgtcctcctgc
KN213	9400100 1 FOR	GTTCAGGGATGCGGACATGGTCCCGGCCCTGGCCTCGGGC
KN214	9400100 1 REV	GCCCGAGGCCAGGGCCGGGACCATGTCCGCATCCCTGAACcctaccctacgtcctcctgc

KN215	ForF FLAG Tag 1 FOR	cgtctagaacaggaggccccatatgATGGACTACAAGGACCACGACGG
KN216	ForF FLAG Tag 1 REV	tgatcagtggcagcctcgttCCACCTCCGCCTGAACCGCC
KN217	ForF FLAG Tag 2 FOR	cgaaccggggaggctgctatgcagaccgtggtgacc
KN218	ForF FLAG Tag 2 REV	gcctgaaccgcctccaccgccccggtcgccctgcg
KN219	ForG FLAG Tag 1 FOR	cgtctagaacaggaggccccatatgATGGACTACAAGGACCACGACGG
KN220	ForG FLAG Tag 1 REV	tgatcagtggcagcctcgttCCACCTCCGCCTGAACCGCC
KN221	ForG FLAG Tag 2 FOR	GGCGGTTCAGGCGGAGGTGGaacgaggctgccactgatca
KN222	ForG FLAG Tag 2 REV	gagaacctaggatccaagcttgctgctcgccatcgaacccg
KN223	pCRISP-2 TEST FOR	AGGCTAGTCCGTTATCAACTTGAAA
KN224	pCRISP-2 TEST REV	TCGCCACCTCTGACTTGAGCGTCGA
KN225	pCRISP-2 spacer TEST	atacggctgccagataaggc
KN226	pIJ10257 TEST FOR	gatcttgacggctggcgagag
KN227	pIJ10257 TEST REV	gcgtcagcatatcatcagcgagc
KN228	pET28a / pET29a TEST FOR	TAATACGACTCACTATAGGG
KN229	pET28a / pET29a TEST REV	CTAGTTATTGCTCAGCGGT
KN230	ForF D53E Repair 1 FOR	gctcggttgccgccgggcgtttttttaTCTAGAagcatcatgtggcggttcca
KN231	ForF D53E Repair 1 REV	CCGACGTCGAGCAGCACCACTTCGGGCGACTTGGCCGCGATCA
KN232	ForF D53E Repair 2 FOR	TGATCGCGGCCAAGTCGCCCGAAGTGGTGCTGCTCGACGTCGG

KN233	ForF D53E Repair 2 REV	AACTtggcgaagatgttgcgcagaTGATGCTTCACGGTGCCCT
KN234	ForF D53E Repair 3 FOR	AGGGCACCGTGAAGCATCAtctgcgcaacatcttcgccaAGTT
KN235	ForF D53E Repair 3 REV	gcaacgcggccttttacggttcctggccTCTAGAactggacgtacgaggaggtg
KN236	ForF D53A Repair 1 FOR	gctcggttgccgccgggcgtttttaTCTAGAagcatcatgtggcggttcca
KN237	ForF D53A Repair 1 REV	CCGACGTCGAGCAGCACCACACGGGGCGACTTGGCCGCGATCA
KN238	ForF D53A Repair 2 FOR	TGATCGCGGCCAAGTCGCCCCGTGTGGTGCTGCTCGACGTCGG
KN239	ForF D53A Repair 2 REV	AACTtggcgaagatgttgcgcagaTGATGCTTCACGGTGCCCT
KN240	ForF D53A Repair 3 FOR	AGGGCACCGTGAAGCATCAtctgcgcaacatcttcgccaAGTT
KN241	ForF D53A Repair 3 REV	gcaacgcggccttttacggttcctggccTCTAGAactggacgtacgaggaggtg
KN242	ForF D53N Repair 1 FOR	gctcggttgccgccgggcgtttttaTCTAGAagcatcatgtggcggttcca
KN243	ForF D53N Repair 1 REV	CCGACGTCGAGCAGCACCACGTTGGGCGACTTGGCCGCGATCA
KN244	ForF D53N Repair 2 FOR	TGATCGCGGCCAAGTCGCCCAACGTGGTGCTGCTCGACGTCGG
KN245	ForF D53N Repair 2 REV	AACTtggcgaagatgttgcgcagaTGATGCTTCACGGTGCCCT
KN246	ForF D53N Repair 3 FOR	AGGGCACCGTGAAGCATCAtctgcgcaacatcttcgccaAGTT
KN247	ForF D53N Repair 3 REV	gcaacgcggccttttacggttcctggccTCTAGAactggacgtacgaggaggtg
KN248	dForF Comp Promoter FOR	gccgagaaccTAGGATCCAAGCTTcgtgtacccccctgtgcacg
KN249	dForF Comp Promoter REV	CTGGTACCATGCATAGATCTAAGCTTccgcgcctcatccgctcttc
KN250	dForF Comp ForF FOR	gccgagaaccTAGGATCCAAGCTTcgtgtacccccctgtgcacg

KN251	dForF Comp ForF REV	caccacggtctgcatccgcgcgcctcgttcacagcagcctc
KN252	dForG Comp FOR	gaggctgctgtgaacgaggcgcgcggatgcagaccgtggtg
KN253	dForG Comp REV	CTGGTACCATGCATAGATCTAAGCTTccgctgctcgccatcgaac
KN254	ForF TEST 1 FOR	ggccggcgggagtccgcggc
KN255	ForF TEST 1 REV	gccatcgaacccgcctgagc
KN256	ForF TEST 2 FOR	gtcgagttcctgctgcccacg
KN257	ForF TEST 2 REV	cctcggcatcgtcgccgagc
KN258	ForG TEST 1 FOR	cttggccatctgcatgagcg
KN259	ForG TEST 1 REV	gtcgtcgacgatcacgatgc
KN260	ForG TEST 2 FOR	gagaccgcgcgaccgaaccg
KN261	ForG TEST 2 REV	tggacgagctgtgccgcagc
KN262	pET28a ForF CO FOR	CTGGTGCCGCGCGGCAGCCATATGATGCAGACCGTGGTTACCCG
KN263	pET28a ForF CO REV	GTGCTCGAGTGCGGCCGCAAGCTTTAACCACGATCACCTTGC
KN264	pET28a ForG CO FOR	CTGGTGCCGCGCGGCAGCCATATGATGAATGAGGCGGCGACCG
KN265	pET28a ForG CO REV	GTGCTCGAGTGCGGCCGCAAGCTTTTAACCGCTTTTGGTCGGCAGC
KN266	pET29a ForF CO FOR	CTTTAAGAAGGAGATATACATATGATGCAGACCGTGGTTACCCG
KN267	pET29a ForF CO REV	GTGCTCGAGTGCGGCCGCAAGCTTTAACCACGATCACCTTGC
KN268	pET29a ForG CO FOR	CTTTAAGAAGGAGATATACATATGATGAATGAGGCGGCGACCG

KN269	pET29a ForG CO REV	GTGCTCGAGTGCGGCCGCAAGCTTTTAACCGCTTTTGGTCGGCAGC
KN270	pET29a ForF CO NO STOP REV	GTGCTCGAGTGCGGCCGCAAGCTTACCACGATCACCTTGCGGC
KN271	pET29a ForG CO NO STOP REV	GTGCTCGAGTGCGGCCGCAAGCTTACCGCTTTTGGTCGGCAGC
KN272	pET28a TEST FOR	TGAGCGGATAACAATTCCCC
KN273	pET28a TEST REV	GCTAGTTATTGCTCAGCGG
KN274	pET29a TEST FOR	GTAGAGGATCGAGATCGATC
KN275	pET29a TEST REV	CAAGACCCGTTTAGAGGCCC
KN276	pET28a ForF D53 FOR	CTGGTGCCGCGCGGCAGCCATATGATGCAGACCGTGGTGACCCG
KN277	pET28a ForF D53 REV	GTGCTCGAGTGCGGCCGCAAGCTTCAGCCCCGGTCGCCCTG
KN278	pET29a ForF D53 FOR	CTTTAAGAAGGAGATATACATATGATGCAGACCGTGGTGACCCG
KN279	pET29a ForF D53 REV	CTCGAGTGCGGCCGCAAGCTTGCCCCGGTCGCCCTG
KN280	BGC 30 RHB 1 FOR	TGGCACACGGGGGAGGGGACATCCCTCCTTCGGCAGAGCC
KN281	BGC 30 RHB 1 REV	GGCTCTGCCGAAGGAGGGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN282	BGC 30 RHB 2 FOR	GCACACGGGGGAGGGGACATCCCTCCTTCGGCAGAGCC
KN283	BGC 30 RHB 2 REV	GGCTCTGCCGAAGGAGGGATGTCCCCTCCCCCGTGTGCcctaccctacgtcctcctgc
KN284	BGC 30 RHB 3 FOR	ACACGGGGGAGGGGACATCCCTCCTTCGGCAGAGCC
KN285	BGC 30 RHB 3 REV	GGCTCTGCCGAAGGAGGGATGTCCCCTCCCCCGTGTcctaccctacgtcctcctgc
KN286	BGC 30 RHB 4 FOR	ACGGGGGAGGGGACATCCCTCCTTCGGCAGAGCC

KN287	BGC 30 RHB 4 REV	GGCTCTGCCGAAGGAGGGATGTCCCCTCCCCCGTcctaccctacgtcctcctgc
KN288	BGC 30 RHB 5 FOR	GGGGGAGGGGACATCCCTCCTTCGGCAGAGCC
KN289	BGC 30 RHB 5 REV	GGCTCTGCCGAAGGAGGGATGTCCCCTCCCCCcctaccctacgtcctcctgc
KN290	BGC 30 RHB 6 FOR	GGGAGGGGACATCCCTCCTTCGGCAGAGCC
KN291	BGC 30 RHB 6 REV	GGCTCTGCCGAAGGAGGGATGTCCCCTCCCcctaccctacgtcctcctgc
KN292	BGC 30 RHB 7 FOR	GAGGGGACATCCCTCCTTCGGCAGAGCC
KN293	BGC 30 RHB 7 REV	GGCTCTGCCGAAGGAGGGATGTCCCCTCcctaccctacgtcctcctgc
KN294	BGC 30 RHB 8 FOR	GGGGACATCCCTCCTTCGGCAGAGCC
KN295	BGC 30 RHB 8 REV	GGCTCTGCCGAAGGAGGGATGTCCCCcctaccctacgtcctcctgc
KN296	BGC 30 RHB 9 FOR	GGACATCCCTCCTTCGGCAGAGCC
KN297	BGC 30 RHB 9 REV	GGCTCTGCCGAAGGAGGGATGTCCcctaccctacgtcctcctgc
KN298	BGC 30 RHB 10 FOR	ACATCCCTCCTTCGGCAGAGCC
KN299	BGC 30 RHB 10 REV	GGCTCTGCCGAAGGAGGGATGTcctaccctacgtcctcctgc
KN300	BGC 30 RHB 11 FOR	ATCCCTCCTTCGGCAGAGCC
KN301	BGC 30 RHB 11 REV	GGCTCTGCCGAAGGAGGGATcctaccctacgtcctcctgc
KN302	BGC 30 RHB 12 FOR	CCCTCCTTCGGCAGAGCC
KN303	BGC 30 RHB 12 REV	GGCTCTGCCGAAGGAGGGcctaccctacgtcctcctgc
KN304	BGC 30 RBH 13 FOR	CTCCTTCGGCAGAGCC

KN305	BGC 30 RHB 13 REV	GGCTCTGCCGAAGGAGcctaccctacgtcctcctgc
KN306	BGC 30 LHB 12 FOR	TGGCACACGGGGGAGG
KN307	BGC 30 LHB 12 REV	CCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN308	BGC 30 LHB 11 FOR	TGGCACACGGGGGAGGGG
KN309	BGC 30 LHB 11 REV	CCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN310	BGC 30 LHB 10 FOR	TGGCACACGGGGGAGGGGAC
KN311	BGC 30 LHB 10 REV	GTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN312	BGC 30 LHB 9 FOR	TGGCACACGGGGGAGGGGACAT
KN313	BGC 30 LHB 9 REV	ATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN314	BGC 30 LHB 8 FOR	TGGCACACGGGGGAGGGGACATCC
KN315	BGC 30 LHB 8 REV	GGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN316	BGC 30 LHB 7 FOR	TGGCACACGGGGGAGGGGACATCCCT
KN317	BGC 30 LHB 7 REV	AGGGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN318	BGC 30 LHB 6 FOR	TGGCACACGGGGGAGGGGACATCCCTCC
KN319	BGC 30 LHB 6 REV	GGAGGGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN320	BGC 30 LHB 5 FOR	TGGCACACGGGGGAGGGGACATCCCTCCTT
KN321	BGC 30 LHB 5 REV	AAGGAGGGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN322	BGC 30 LHB 4 FOR	TGGCACACGGGGGAGGGGACATCCCTCCTTCG

KN323	BGC 30 LHB 4 REV	CGAAGGAGGGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN324	BGC 30 LHB 3 FOR	TGGCACACGGGGGAGGGGACATCCCTCCTTCGGC
KN325	BGC 30 LHB 3 REV	GCCGAAGGAGGGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN326	BGC 30 LHB 2 FOR	TGGCACACGGGGGAGGGGACATCCCTCCTTCGGCAG
KN327	BGC 30 LHB 2 REV	CTGCCGAAGGAGGGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN328	BGC 30 LHB 1 FOR	TGGCACACGGGGGAGGGGACATCCCTCCTTCGGCAGAG
KN329	BGC 30 LHB 1 REV	CTCTGCCGAAGGAGGGATGTCCCCTCCCCCGTGTGCCAcctaccctacgtcctcctgc
KN330	BGC 4 1 RHB 1 FOR	AACGGACCAGACGAGCTACCCCGAACGGCGTAGCGAAGGG
KN331	BGC 4 1 RHB 1 REV	CCCTTCGCTACGCCGTTCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN332	BGC 4 1 RHB 2 FOR	CGGACCAGACGAGCTACCCCGAACGGCGTAGCGAAGGG
KN333	BGC 4 1 RHB 2 REV	CCCTTCGCTACGCCGTTCGGGGTAGCTCGTCTGGTCCGcctaccctacgtcctcctgc
KN334	BGC 4 1 RHB 3 FOR	GACCAGACGAGCTACCCCGAACGGCGTAGCGAAGGG
KN335	BGC 4 1 RHB 3 REV	CCCTTCGCTACGCCGTTCGGGGTAGCTCGTCTGGTCcctaccctacgtcctcctgc
KN336	BGC 4 1 RHB 4 FOR	CCAGACGAGCTACCCCGAACGGCGTAGCGAAGGG
KN337	BGC 4 1 RHB 4 REV	CCCTTCGCTACGCCGTTCGGGGTAGCTCGTCTGGcctaccctacgtcctcctgc
KN338	BGC 4 1 RHB 5 FOR	AGACGAGCTACCCCGAACGGCGTAGCGAAGGG
KN339	BGC 4 1 RHB 5 REV	CCCTTCGCTACGCCGTTCGGGGTAGCTCGTCTcctaccctacgtcctcctgc
KN340	BGC 4 1 RHB 6 FOR	ACGAGCTACCCCGAACGGCGTAGCGAAGGG

KN341	BGC 4 1 RHB 6 REV	CCCTTCGCTACGCCGTTCGGGGTAGCTCGTcctaccctacgtcctcctgc
KN342	BGC 4 1 RHB 7 FOR	GAGCTACCCCGAACGGCGTAGCGAAGGG
KN343	BGC 4 1 RHB 7 REV	CCCTTCGCTACGCCGTTCGGGGTAGCTCcctaccctacgtcctcctgc
KN344	BGC 4 1 RHB 8 FOR	GCTACCCCGAACGGCGTAGCGAAGGG
KN345	BGC 4 1 RHB 8 REV	CCCTTCGCTACGCCGTTCGGGGTAGCcctaccctacgtcctcctgc
KN346	BGC 4 1 RHB 9 FOR	TACCCCGAACGGCGTAGCGAAGGG
KN347	BGC 4 1 RHB 9 REV	CCCTTCGCTACGCCGTTCGGGGTAcctaccctacgtcctcctgc
KN348	BGC 4 1 RHB 10 FOR	CCCCGAACGGCGTAGCGAAGGG
KN349	BGC 4 1 RHB 10 REV	CCCTTCGCTACGCCGTTCGGGGcctaccctacgtcctcctgc
KN350	BGC 4 1 RHB 11 FOR	CCGAACGGCGTAGCGAAGGG
KN351	BGC 4 1 RHB 11 REV	CCCTTCGCTACGCCGTTCGGcctaccctacgtcctcctgc
KN352	BGC 4 1 RHB 12 FOR	GAACGGCGTAGCGAAGGG
KN353	BGC 4 1 RHB 12 REV	CCCTTCGCTACGCCGTTCcctaccctacgtcctcctgc
KN354	BGC 4 1 RBH 13 FOR	ACGGCGTAGCGAAGGG
KN355	BGC 4 1 RHB 13 REV	CCCTTCGCTACGCCGTcctaccctacgtcctcctgc
KN356	BGC 4 1 LHB 12 FOR	AACGGACCAGACGAGC
KN357	BGC 4 1 LHB 12 REV	GCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN358	BGC 4 1 LHB 11 FOR	AACGGACCAGACGAGCTA

KN359	BGC 4 1 LHB 11 REV	TAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN360	BGC 4 1 LHB 10 FOR	AACGGACCAGACGAGCTACC
KN361	BGC 4 1 LHB 10 REV	GGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN362	BGC 4 1 LHB 9 FOR	AACGGACCAGACGAGCTACCCC
KN363	BGC 4 1 LHB 9 REV	GGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN364	BGC 4 1 LHB 8 FOR	AACGGACCAGACGAGCTACCCCGA
KN365	BGC 4 1 LHB 8 REV	TCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN366	BGC 4 1 LHB 7 FOR	AACGGACCAGACGAGCTACCCCGAAC
KN367	BGC 4 1 LHB 7 REV	GTTCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN368	BGC 4 1 LHB 6 FOR	AACGGACCAGACGAGCTACCCCGAACGG
KN369	BGC 4 1 LHB 6 REV	CCGTTCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN370	BGC 4 1 LHB 5 FOR	AACGGACCAGACGAGCTACCCCGAACGGCG
KN371	BGC 4 1 LHB 5 REV	CGCCGTTCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN372	BGC 4 1 LHB 4 FOR	AACGGACCAGACGAGCTACCCCGAACGGCGTA
KN373	BGC 4 1 LHB 4 REV	TACGCCGTTCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN374	BGC 4 1 LHB 3 FOR	AACGGACCAGACGAGCTACCCCGAACGGCGTAGC
KN375	BGC 4 1 LHB 3 REV	GCTACGCCGTTCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN376	BGC 4 1 LHB 2 FOR	AACGGACCAGACGAGCTACCCCGAACGGCGTAGCGA

KN377	BGC 4 1 LHB 2 REV	TCGCTACGCCGTTCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN378	BGC 4 1 LHB 1 FOR	AACGGACCAGACGAGCTACCCCGAACGGCGTAGCGAAG
KN379	BGC 4 1 LHB 1 REV	CTTCGCTACGCCGTTCGGGGTAGCTCGTCTGGTCCGTTcctaccctacgtcctcctgc
KN380	BGC 4 2 RHB 1 FOR	GCCACCCGGACGGCTGCCCCTTTCGGGCGGTGTCCACCTG
KN381	BGC 4 2 RHB 1 REV	CAGGTGGACACCGCCCGAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN382	BGC 4 2 RHB 2 FOR	CACCCGGACGGCTGCCCCTTTCGGGCGGTGTCCACCTG
KN383	BGC 4 2 RHB 2 REV	CAGGTGGACACCGCCCGAAAGGGGCAGCCGTCCGGGTGcctaccctacgtcctcctgc
KN384	BGC 4 2 RHB 3 FOR	CCCGGACGGCTGCCCCTTTCGGGCGGTGTCCACCTG
KN385	BGC 4 2 RHB 3 REV	CAGGTGGACACCGCCCGAAAGGGGCAGCCGTCCGGGcctaccctacgtcctcctgc
KN386	BGC 4 2 RHB 4 FOR	CGGACGGCTGCCCCTTTCGGGCGGTGTCCACCTG
KN387	BGC 4 2 RHB 4 REV	CAGGTGGACACCGCCCGAAAGGGGCAGCCGTCCGcctaccctacgtcctcctgc
KN388	BGC 4 2 RHB 5 FOR	GACGGCTGCCCCTTTCGGGCGGTGTCCACCTG
KN389	BGC 4 2 RHB 5 REV	CAGGTGGACACCGCCCGAAAGGGGCAGCCGTCcctaccctacgtcctcctgc
KN390	BGC 4 2 RHB 6 FOR	CGGCTGCCCCTTTCGGGCGGTGTCCACCTG
KN391	BGC 4 2 RHB 6 REV	CAGGTGGACACCGCCCGAAAGGGGCAGCCGcctaccctacgtcctcctgc
KN392	BGC 4 2 RHB 7 FOR	GCTGCCCCTTTCGGGCGGTGTCCACCTG
KN393	BGC 4 2 RHB 7 REV	CAGGTGGACACCGCCCGAAAGGGGCAGCcctaccctacgtcctcctgc
KN394	BGC 4 2 RHB 8 FOR	TGCCCCTTTCGGGCGGTGTCCACCTG

KN395	BGC 4 2 RHB 8 REV	CAGGTGGACACCGCCCGAAAGGGGCAcctaccctacgtcctcctgc
KN396	BGC 4 2 RHB 9 FOR	CCCCTTTCGGGCGGTGTCCACCTG
KN397	BGC 4 2 RHB 9 REV	CAGGTGGACACCGCCCGAAAGGGGcctaccctacgtcctcctgc
KN398	BGC 4 2 RHB 10 FOR	CCTTTCGGGCGGTGTCCACCTG
KN399	BGC 4 2 RHB 10 REV	CAGGTGGACACCGCCCGAAAGGcctaccctacgtcctcctgc
KN400	BGC 4 2 RHB 11 FOR	TTTCGGGCGGTGTCCACCTG
KN401	BGC 4 2 RHB 11 REV	CAGGTGGACACCGCCCGAAAcctaccctacgtcctcctgc
KN402	BGC 4 2 RHB 12 FOR	TCGGGCGGTGTCCACCTG
KN403	BGC 4 2 RHB 12 REV	CAGGTGGACACCGCCCGAcctaccctacgtcctcctgc
KN404	BGC 4 2 RBH 13 FOR	GGGCGGTGTCCACCTG
KN405	BGC 4 2 RHB 13 REV	CAGGTGGACACCGCCCcctaccctacgtcctcctgc
KN406	BGC 4 2 LHB 12 FOR	GCCACCCGGACGGCTG
KN407	BGC 4 2 LHB 12 REV	CAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN408	BGC 4 2 LHB 11 FOR	GCCACCCGGACGGCTGCC
KN409	BGC 4 2 LHB 11 REV	GGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN410	BGC 4 2 LHB 10 FOR	GCCACCCGGACGGCTGCCCC
KN411	BGC 4 2 LHB 10 REV	GGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN412	BGC 4 2 LHB 9 FOR	GCCACCCGGACGGCTGCCCCTT

KN413	BGC 4 2 LHB 9 REV	AAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN414	BGC 4 2 LHB 8 FOR	GCCACCCGGACGGCTGCCCCTTTC
KN415	BGC 4 2 LHB 8 REV	GAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN416	BGC 4 2 LHB 7 FOR	GCCACCCGGACGGCTGCCCCTTTCGG
KN417	BGC 4 2 LHB 7 REV	CCGAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN418	BGC 4 2 LHB 6 FOR	GCCACCCGGACGGCTGCCCCTTTCGGGC
KN419	BGC 4 2 LHB 6 REV	GCCCGAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN420	BGC 4 2 LHB 5 FOR	GCCACCCGGACGGCTGCCCCTTTCGGGCGG
KN421	BGC 4 2 LHB 5 REV	CCGCCCGAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN422	BGC 4 2 LHB 4 FOR	GCCACCCGGACGGCTGCCCCTTTCGGGCGGTG
KN423	BGC 4 2 LHB 4 REV	CACCGCCCGAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN424	BGC 4 2 LHB 3 FOR	GCCACCCGGACGGCTGCCCCTTTCGGGCGGTGTC
KN425	BGC 4 2 LHB 3 REV	GACACCGCCCGAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN426	BGC 4 2 LHB 2 FOR	GCCACCCGGACGGCTGCCCCTTTCGGGCGGTGTCCA
KN427	BGC 4 2 LHB 2 REV	TGGACACCGCCCGAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN428	BGC 4 2 LHB 1 FOR	GCCACCCGGACGGCTGCCCCTTTCGGGCGGTGTCCACC
KN429	BGC 4 2 LHB 1 REV	GGTGGACACCGCCCGAAAGGGGCAGCCGTCCGGGTGGCcctaccctacgtcctcctgc
KN430	ForF D53 Comp 1 FOR	gccgagaaccTAGGATCCAAGCTTcgtgtacccccctgtgcacg

KN431	ForF D53 Comp 1 REV	caccacggtctgcatccgcgcgcctcgttcacagcagcctc
KN432	ForF D53 Comp 2 FOR	gaggctgctgtgaacgaggcgcgcggatgcagaccgtggtg
KN433	ForF D53E Comp 2 REV	gacgtcgagcagcaccacttcgggcgacttggccgc
KN434	ForF D53E Comp 3 FOR	gcggccaagtcgcccgaagtggtgctgctcgacgtc
KN435	ForF D53A Comp 2 REV	gacgtcgagcagcaccacacgcgggcgacttggccgc
KN436	ForF D53A Comp 3 FOR	gcggccaagtcgcccgcgtgtggtgctgctcgacgtc
KN437	ForF D53N Comp 2 REV	gacgtcgagcagcaccacgttcgggcgacttggccgc
KN438	ForF D53N Comp 3 FOR	gcggccaagtcgcccgaacgtggtgctgctcgacgtc
KN439	ForF D53 Comp 3 REV	CTGGTACCATGCATAGATCTAAGCTTccgctgctcgccatcgaac
KN440	ForG H175A Comp 1 FOR	gaggctgctgtgaacgaggcgcgcggatgcagaccgtggtg
KN441	ForG H175A Comp 1 REV	CGTAGCCGACCCGGTCGGCCAGGTCGCGGGAGATG
KN442	ForG H175A Comp 2 FOR	CATCTCCCGCGACCTGGCCGACCGGGTCGGCTACG
KN443	ForG H175A Comp 2 REV	CTGGTACCATGCATAGATCTAAGCTTccgctgctcgccatcgaac
KN444	ForG D176A Comp 1 FOR	gaggctgctgtgaacgaggcgcgcggatgcagaccgtggtg
KN445	ForG D176A Comp 1 REV	CGTAGCCGACCCGGTCGGCCAGGTCGCGGGAGATG
KN446	ForG D176A Comp 2 FOR	CATCTCCCGCGACCTGCACGCACGGGTCGGCTACG
KN447	ForG D176A Comp 2 REV	CTGGTACCATGCATAGATCTAAGCTTccgctgctcgccatcgaac
KN448	pMS82 TEST FOR	gcaacagtgccgttgatcgtgctatg

KN449	pMS82 TEST REV	GCCAGTGGTATTTATGTCAACACCGCC
KN450	pIJ10257 1516 FOR	cgtctagaacaggaggccccatatgATGACCCGGGTCCTTGTG
KN451	pIJ10257 1516 REV	CCGCGTGCCATGGGTTCTCGAGTCCGTCCGCCGGTCGTCAGCTCTCGAAGC
KN452	pIJ10257 1517 1 FOR	GCTTCGAGAGCTGACGACCGGCGGACGGACTCGAGAACCCATGGCACGCGG
KN453	pIJ10257 1517 1 REV	gagaacctaggatccaagcttTCAGGTCGTGGCCGCCGCC
KN454	pIJ10257 1517 2 FOR	cgtctagaacaggaggccccatatgATGCCCGCGTGGCAGAAG
KN455	pIJ10257 1517 2 REV	gagaacctaggatccaagcttCTAGAGATCGACGCTCC
KN456	pIJ10257 6759/6760 FOR	cgtctagaacaggaggccccatatgATGGATGATCACCG
KN457	pIJ10257 6759/6760 REV	CGGAGCGGTGGTTCTGCGGCATTCACCTGGCGGGCAGCGTGACGC
KN458	pIJ10257 7292 FOR	GCGTCACGCTGCCCGCCAGGTGAATGCCGCAGAACCACCGCTCCG
KN459	pIJ10257 7292 REV	gagaacctaggatccaagcttTCAGCGCCGCCCGACCTGG
KN460	pIJ10257 7294 FOR	cgtctagaacaggaggccccatatgATGAGGGTGCTGGTGGTC
KN461	pIJ10257 7294 REV	GTATCGGCATCTCACATCACATCTTCAGATCCGGTTCCGGTCATCTCAGATCCG
KN462	pIJ10257 7446 FOR	CGGATCTGAGATGACCGGAACCGGATCTGAAGATGTGATGTGAGATGCCGATAC
KN463	pIJ10257 7446 REV	gagaacctaggatccaagcttCGGCGAGCTCGCCGCTCGC
KN464	pIJ10257 7447 FOR	cgtctagaacaggaggccccatatgATGACCGTTCGTACGTTCTC
KN465	pIJ10257 7447 REV	accgccaccgccagagccacctccgcctgaaccgcctccaccTCAGCGTGGCTGGTCCG
KN466	pIJ10257 7549 FOR	aagcttggatcctaggttctcTTACCTGCCTCGCAGCGGAAC

KN467	pIJ10257 7549 REV	ggtggaggcggttcaggcggaggtggctctggcggtggcggtATGGGTGGGGACGCGCAC
KN468	pIJ10257 7550 FOR	cgtctagaacaggaggccccatatgATGCGGGTGCTGGTGGTC
KN469	pIJ10257 7550 REV	CGGATGCCGACGTAGATCACGGCGACCATCAGCGCCGAGCAGAC
KN470	pIJ10257 8002 FOR	GTCTGCTCGGCGCTGATGGTCGCCGTGATCTACGTCGGCATCCG
KN471	pIJ10257 8002 REV	gagaacctaggatccaagcttTCACGCGCTCGGCCGAGG
KN472	pIJ10257 8003 FOR	GACCCGAAGAACATCGTCGCCCTCGCCATCTTC
KN473	pIJ10257 8003 REV	GAAGATGGCGAGGGCGACGATGTTCTTCGGGTC
KN474	pMS82 ForHI Comp FOR	gccgagaaccTAGGATCCAAGCTTgtcaatgtcgcaatatgccgattcg
KN475	pMS82 ForHI Comp REV	CTGGTACCATGCATAGATCTAAGCTTctaggcacggtccgcctcc
KN476	pET29a D53 CO 1 FOR	CTTTAAGAAGGAGATATACATATGATGCAGACCGTGGTTACCCG
KN477	pET29a D53E CO 1 REV	CATCCAGCAGAACCACTTCCGGGCTCTTCGCCG
KN478	pET29a D53E CO 2 FOR	CGGCGAAGAGCCCGGAAGTGGTTCTGCTGGATG
KN479	pET29a D53A CO 1 REV	CATCCAGCAGAACCACCGCCGGGCTCTTCGCCG
KN480	pET29a D53A CO 2 FOR	CGGCGAAGAGCCCGGCGGTGGTTCTGCTGGATG
KN481	pET29a D53N CO 1 REV	CATCCAGCAGAACCACGTTCGGGCTCTTCGCCG
KN482	pET29a D53N CO 2 FOR	CGGCGAAGAGCCCGAACGTGGTTCTGCTGGATG
KN483	pET29a D53 CO 2 REV	GTGCTCGAGTGCGGCCGCAAGCTTACCACGATCACCTTGCGGC
KN484	pMS82 ForF Promoter 1 FOR	gccgagaaccTAGGATCCAAGCTTcgtgtacccccctgtgcacg

KN485	pMS82 ForF Promoter 2 REV	caccacggtctgcatccgcgcgcctcgttcacagcagcctc
KN486	pMS82 ForF D53 Comp 2 FOR	gaggctgctgtgaacgaggcgcgcggatgcagaccgtggtg
KN487	pMS82 ForF D53E Comp 2 REV	gacgtcgagcagcaccacttcgggcgacttggccgc
KN488	pMS82 ForF D53E Comp 3 FOR	gcggccaagtcgcccgaagtggtgctgctcgacgtc
KN489	pMS82 ForF D53A Comp 2 REV	gacgtcgagcagcaccaccgcgggcgacttggccgc
KN490	pMS82 ForF D53A Comp 3 FOR	gcggccaagtcgcccgcggtggtgctgctcgacgtc
KN491	pMS82 ForF D53N Comp 2 REV	gtcgagcagcaccacgttgggcgacttggccgc
KN492	pMS82 ForF D53N Comp 3 FOR	gcggccaagtcgcccaacgtggtgctgctcgacgtc
KN493	pMS82 ForF D53 Comp 3 REV	CTGGTACCATGCATAGATCTAAGCTTccgctgctcgccatcgaacccgcctgagcctcag
KN494	pIJ10257 ForF FLAG 1 FOR	cgtctagaacaggaggccccatatgATGCAGACCGTGGTGACCCG
KN495	pIJ10257 ForF FLAG 2 REV	CCGTCGTGGTCCTTGTAGTCactaccgccaccgccagag
KN496	pIJ10257 ForF FLAG 3 FOR	GTCGTGGTCCTTGTAGTCTCAGCCCCGGTCGCCCTG
KN497	pIJ10257 ForF FLAG 3 REV	gagaacctaggatccaagctttcaCTTGTCGTCATCGTCCTTGTAGTC
KN498	pIJ10257 ForF Fluoro 1 FOR	cgtctagaacaggaggccccatatgATGCAGACCGTGGTGACCCGCATC
KN499	pIJ10257 ForF Fluoro 1 REV	cccttggagaccatTCAGCCCCGGTCGCCCTGC
KN500	pIJ10257 ForF Fluoro 2 FOR	GCAGGGCGACCGGGGCTGAatggtctccaaggg
KN501	pIJ10257 ForF Fluoro 2 REV	gagaacctaggatccaagctttcatcacttgtacagctc
KN502	pIJ10257 ForG Fluoro 1 FOR	cgtctagaacaggaggccccatatgatgaacgaggctgccactgatc

KN503	pIJ10257 ForG Flouro 1 REV	ctcctcgcccttggagaccatcgcctcatccgctcttcgtggg
KN504	pIJ10257 ForG Flouro 2 FOR	cccacgaagagcggatgaggcgatggtctccaagggcgaggag
KN505	pIJ10257 ForG Flouro 2 REV	gagaacctaggatccaagctttcatcacttgtacagctc
KN506	pIJ10257 ForF Fluoro 1 FOR	cgtctagaacaggaggccccatatgATGCAGACCGTGGTGACCCG
KN507	pIJ10257 ForF Fluoro 1 REV	CCGTCGTGGTCCTTGTAGTCactaccgccaccgccagag

Appendix Table 3 Plasmids used and generated during this thesis

Plasmid	Description	Resistance	Source
pUZ8002	RK2 derivative with a mutation in *oriT*	Kan	(Kieser et al., 2000)
pMS82	*ori*, pUC18, Hyg^R, *oriT*, RK2, *ΦBT1 int*	Hyg	(Gregory et al., 2003)
pIJ10257	*oriT, ΦBT1 attB-int*, Hyg^R, p*ermE**, pMS81 backbone	Hyg	(Hong et al., 2005)
pCRISPomyces-2	Apr^R, *oriT*, $rep^{pSG5(ts)}$, ori^{ColE1}, *sSpcas9*, synthetic guide RNA cassette	Apr	(Cobb et al., 2015)
pMF96	*ΦBT1 attB-int*, Hyg^R, *uidA CDS*	Hyg	(Feeney et al., 2017)
pMF23	*ΦC31 attB-int*, Apr^R	Apr	(Feeney et al., 2017)
pET28a	pBR322 origin and fI origin, Kan^R, expression vector	Kan	Invitrogen
pET29a	pBR322 origin and fI origin, Kan^R, expression vector	Kan	Invitrogen
pSS063	pIJ10257 origin, *mTurquoise2*	Hyg	A gift from Susan Schlimpert (John Innes Centre)
pSS172	pIJ10257 origin, *mCherry*	Hyg	A gift from Susan Schlimpert (John Innes Centre)
pKN001	pCRISPomyces-2 *for*G flanking DNA and gRNA	Apr	This work
pKN002	pCRISPomyces-2 *forF* flanking DNA and gRNA	Apr	This work
pKN003	pCRISPomyces-2 *forHI* flanking DNA and gRNA	Apr	This work
pKN004	pIJ10257 p*forG forF*	Hyg	This work

pKN005	pIJ10257 p*forG forG*	Hyg	This work
pKN006	pIJ10257 p*forH forHI*	Hyg	This work
pKN007	pET28a *forG*	Kan	This work
pKN008	pET28a *forF*	Kan	This work
pKN009	pET29a *forG*	Kan	This work
pKN010	pET29a *forF*	Kan	This work
pKN011	pMF96 p*forM* GUS	Hyg	This work
pKN012	pMF96 p*forH* GUS	Hyg	This work
pKN013	pMF96 p*forG* GUS	Hyg	This work
pKN014	pMF96 p*forT* GUS	Hyg	This work
pKN015	pMF96 p*forU* GUS	Hyg	This work
pKN016	pMF96 p*forZ* GUS	Hyg	This work
pKN017	pMF96 p*forAA* GUS	Hyg	This work
pKN018	pIJ10257 *forG* 3xFLAG	Hyg	This work
pKN019	pIJ10257 *forF* 3xFLAG	Hyg	This work
pKN020	pCRISPomyces-2 *forF* D53E	Apr	This work
pKN021	pCRISPomyces-2 *forF* D53A	Apr	This work
pKN022	pCRISPomyces-2 *forF* D53N	Apr	This work

pKN023	pCRISPomyces-2 *forG* H175A	Apr	This work
pKN024	pCRISPomyces-2 *forG* D176A	Apr	This work
pKN025	pMS82 p*forG forF*	Hyg	This work
pKN026	pMS82 p*forG forG*	Hyg	This work
pKN027	pMS82 p*forH forHI*	Hyg	This work
pKN028	pET28a *forG* codon optimised	Kan	This work, Genewiz
pKN029	pET28a *forF* codon optimised	Kan	This work, Genewiz
pKN030	pET29a *forG* codon optimised	Kan	This work, Genewiz
pKN031	pET29a *forF* codon optimised	Kan	This work, Genewiz
pKN032	pET28a *forF* D53E	Kan	This work
pKN033	pET28a *forF* D53A	Kan	This work
pKN034	pET28a *forF* D53N	Kan	This work
pKN035	pET29a *forF* D53E	Kan	This work
pKN036	pET29a *forF* D53A	Kan	This work
pKN037	pET29a *forF* D53N	Kan	This work
pKN038	pET28a *forF* D53E codon optimised	Kan	This work
pKN039	pET28a *forF* D53A codon optimised	Kan	This work
pKN040	pET28a *forF* D53N codon optimised	Kan	This work

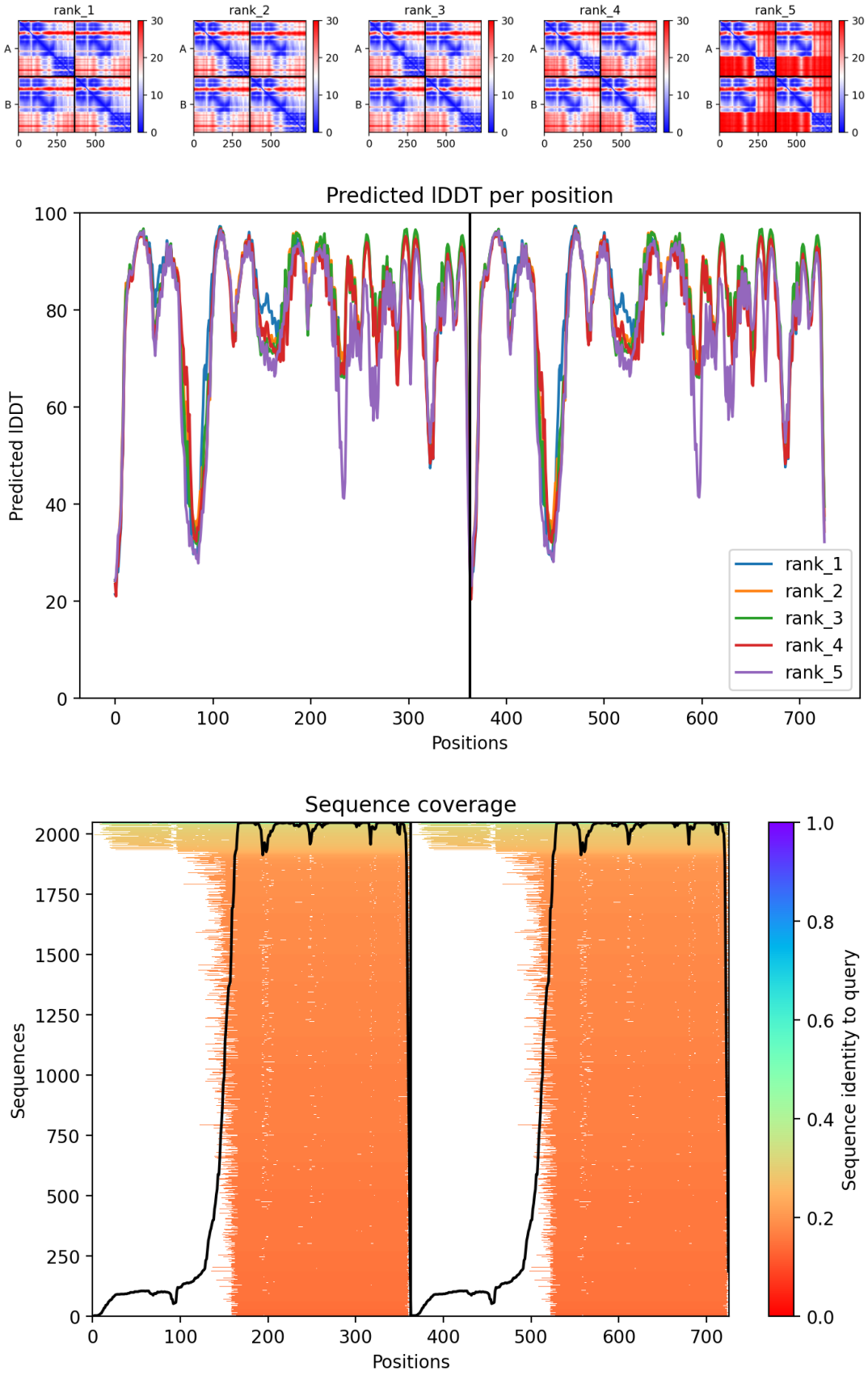
pKN041	pET29a *forF* D53E codon optimised	Kan	This work
pKN042	pET29a *forF* D53A codon optimised	Kan	This work
pKN043	pET29a *forF* D53N codon optimised	Kan	This work
pKN044	pMS82 p*forG forF* D53E	Kan	This work
pKN045	pMS82 p*forG forF* D53A	Hyg	This work
pKN046	pMS82 p*forG forF* D53N	Hyg	This work
pKN047	pMS82 p*forG forG* H175A	Hyg	This work
pKN048	pMS82 p*forG forG* D176A	Hyg	This work
pKN049	pIJ10257 1516/1517	Hyg	This work
pKN050	pIJ10257 6759/6760	Hyg	This work
pKN051	pIJ10257 7292/7294	Hyg	This work
pKN052	pIJ10257 7446/7447	Hyg	This work
pKN053	pIJ10257 8002/8003	Hyg	This work
pKN054	pIJ10257 p*forG forG mTurquoise*	Hyg	This work
pKN055	pIJ10257 p*forG forG mCherry*	Hyg	This work
pKN056	pIJ10257 p*forG forF mTurquoise*	Hyg	This work
pKN057	pIJ10257 p*forG forF mCherry*	Hyg	This work

Appendix Table 4 Summary of bioactivity in two-component system over-expression strains of *Streptomyces formicae*

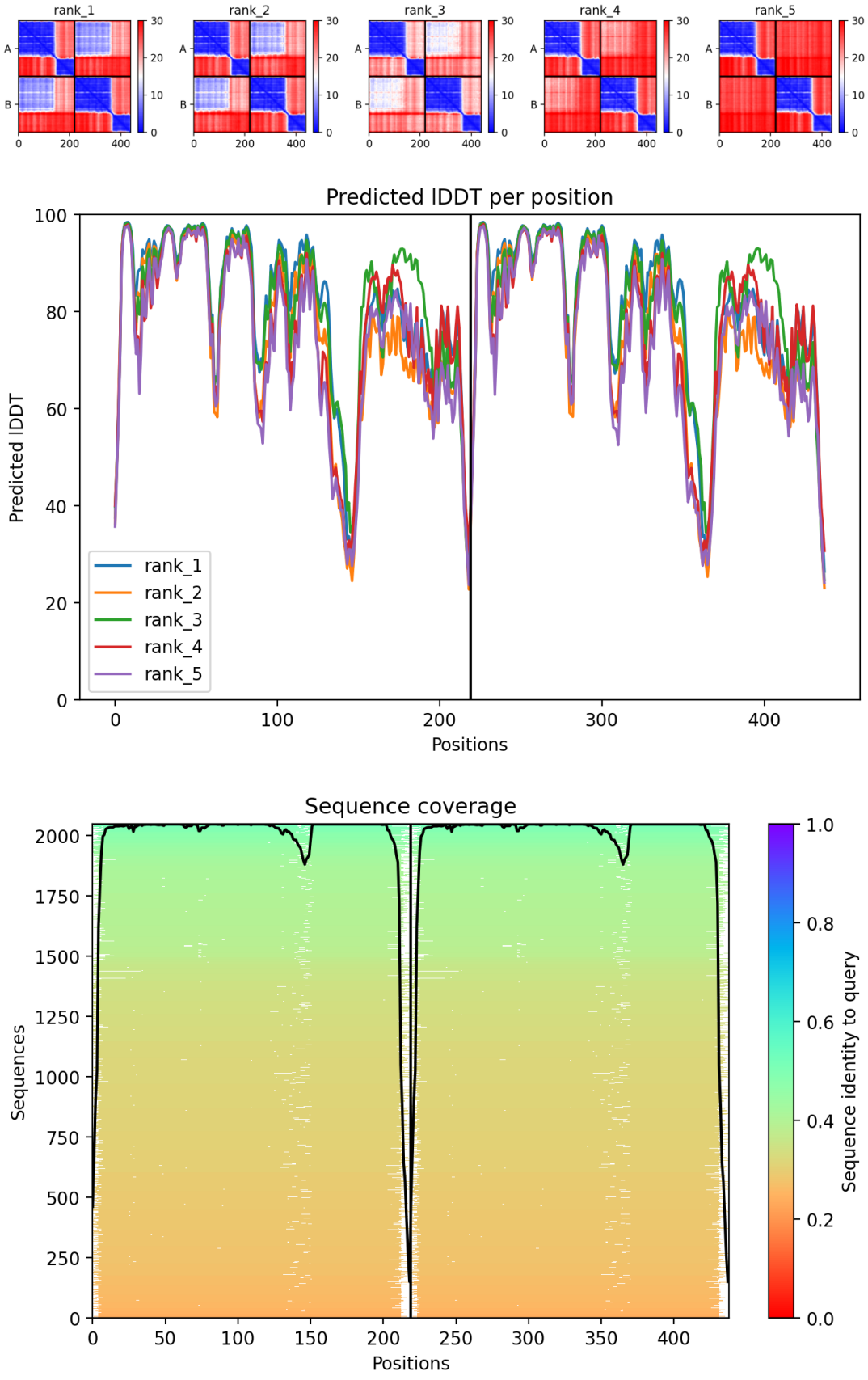
***Streptomyces* Strain**	**Overlay Strain**			
	MRSA	***Escherichia coli***	***Bacillus subtilis***	***Candida albicans***
KY5	SFM – No visible activity MYM – 3 mm zone LB – 2 mm zone LB + Glycerol – 5 mm zone	SFM – No visible activity MYM – 6 mm zone LB – 6 mm zone LB + Glycerol – 5 mm zone	SFM – No visible activity MYM – 1 mm zone LB – 1 mm zone LB + Glycerol – 2 mm zone	SFM – No visible activity MYM – 3 mm zone LB – 2mm zone LB + Glycerol – 2 mm zone
KY5 1516/1517	SFM – No change MYM – No change LB – No change LB + Glycerol – Increased	SFM – No change MYM – No change LB – Increased LB + Glycerol – No change	SFM – No change MYM – No change LB – Increased LB + Glycerol – No change	SFM – No change MYM – No change LB – No change No change LB + Glycerol –
KY5 6759/6760	SFM – No change MYM – Increased LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – Increased LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change
KY5 7292/7294	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – Increased LB – No change LB + Glycerol – No change

KY5 7446/7447	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – Increased LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change
KY5 8002/8003	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – Decreased LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change
Δ*forGF*	SFM – No visible activity MYM – 2 mm zone LB – 2 mm zone LB + Glycerol – No visible activity	SFM – No visible activity MYM – 4 mm zone LB – 4 mm zone LB + Glycerol – 5 mm zone	SFM – No visible activity MYM – 1 mm zone LB – 2 mm zone LB + Glycerol – 2 mm zone	SFM – No visible activity MYM – 6 mm zone LB – 2 mm zone LB + Glycerol – 3 mm zone
Δ*forGF* 1516/1517	SFM – No change MYM – No change LB – No change LB + Glycerol – Increased	SFM – No change MYM – No change LB – Increased LB + Glycerol – No change	SFM – No change MYM – No change LB – Increased LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change
Δ*forGF* 6759/6760	SFM – No change MYM – Increased LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – Increased LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change

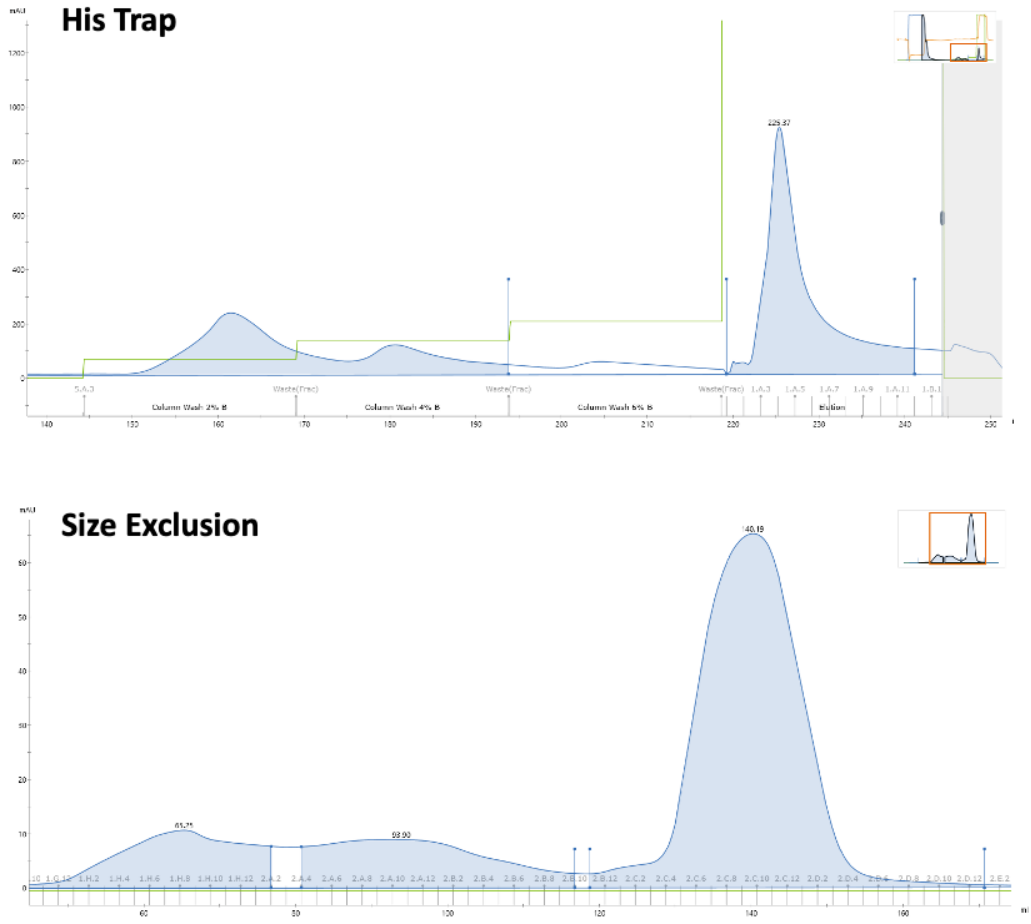
Δ*forGF* 7292/7294	SFM – No change MYM – No change LB – No change LB + Glycerol –	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – Increased LB – No change LB + Glycerol – No change
Δ*forGF* 7446/7447	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – Increased LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change
Δ*forGF* 8002/8003	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change	SFM – No change MYM – Decreased LB – No change LB + Glycerol – No change	SFM – No change MYM – No change LB – No change LB + Glycerol – No change



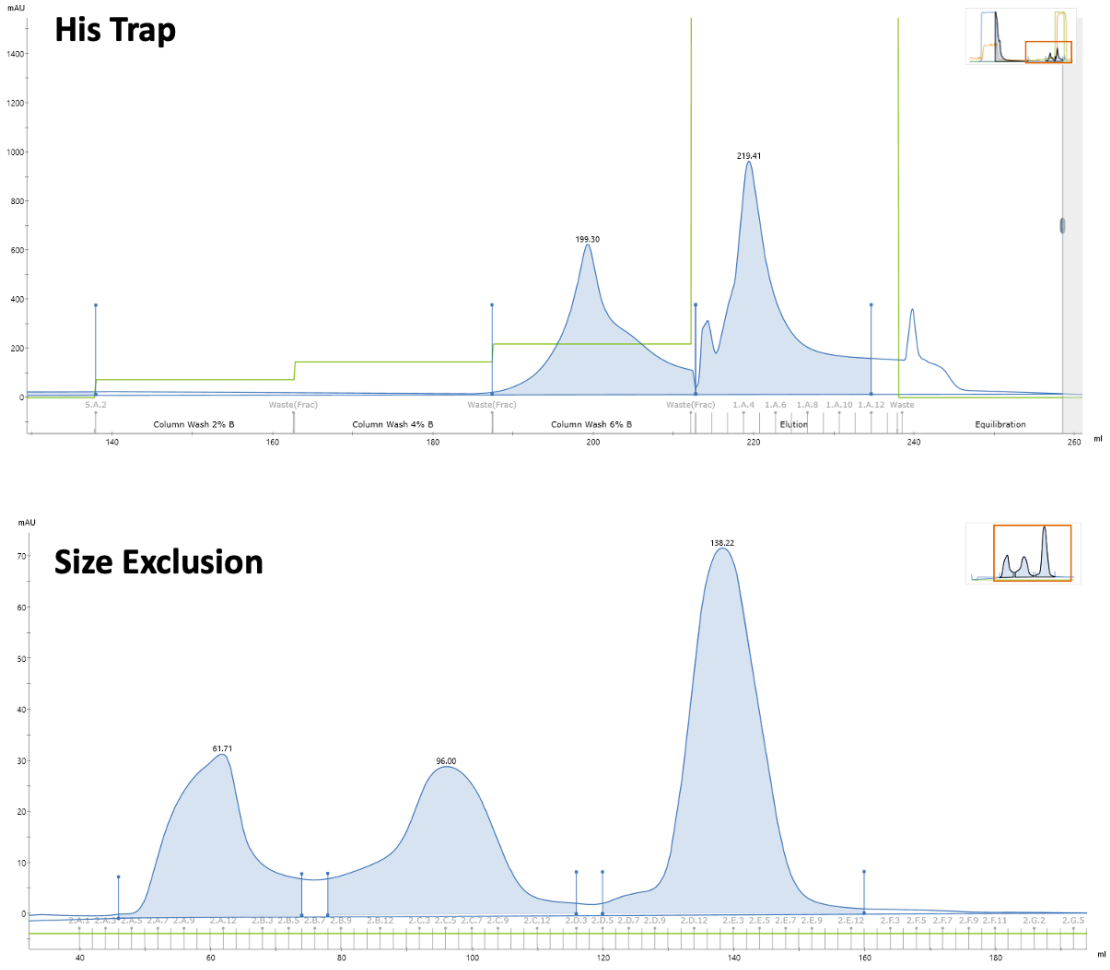
Appendix Figure 1 AlphaFold-2 prediction confidence and coverage information for the product of the ForG dimer (KY5_6663).



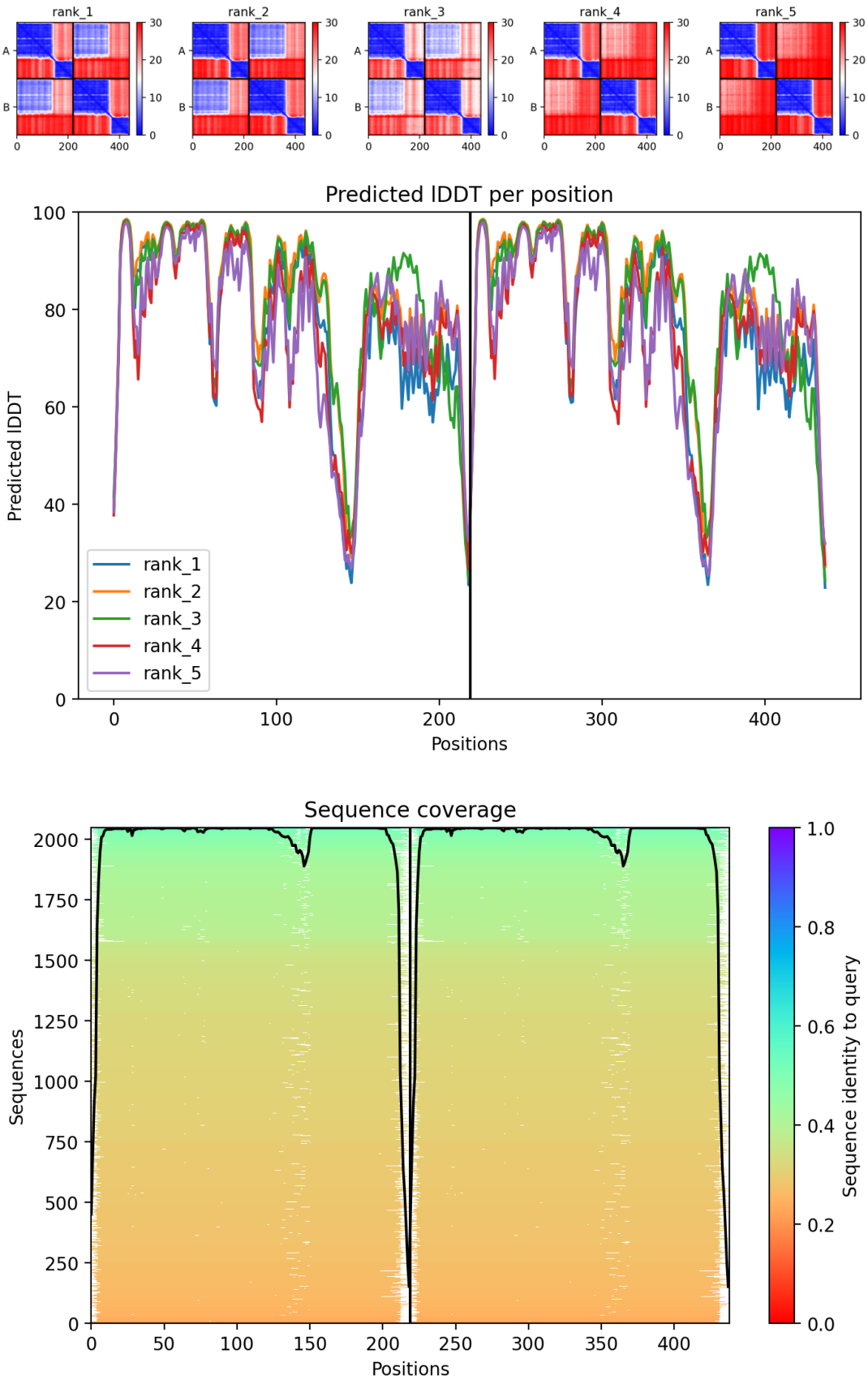
Appendix Figure 2 AlphaFold-2 prediction confidence and coverage information for the product of the ForF dimer (KY5_6664).



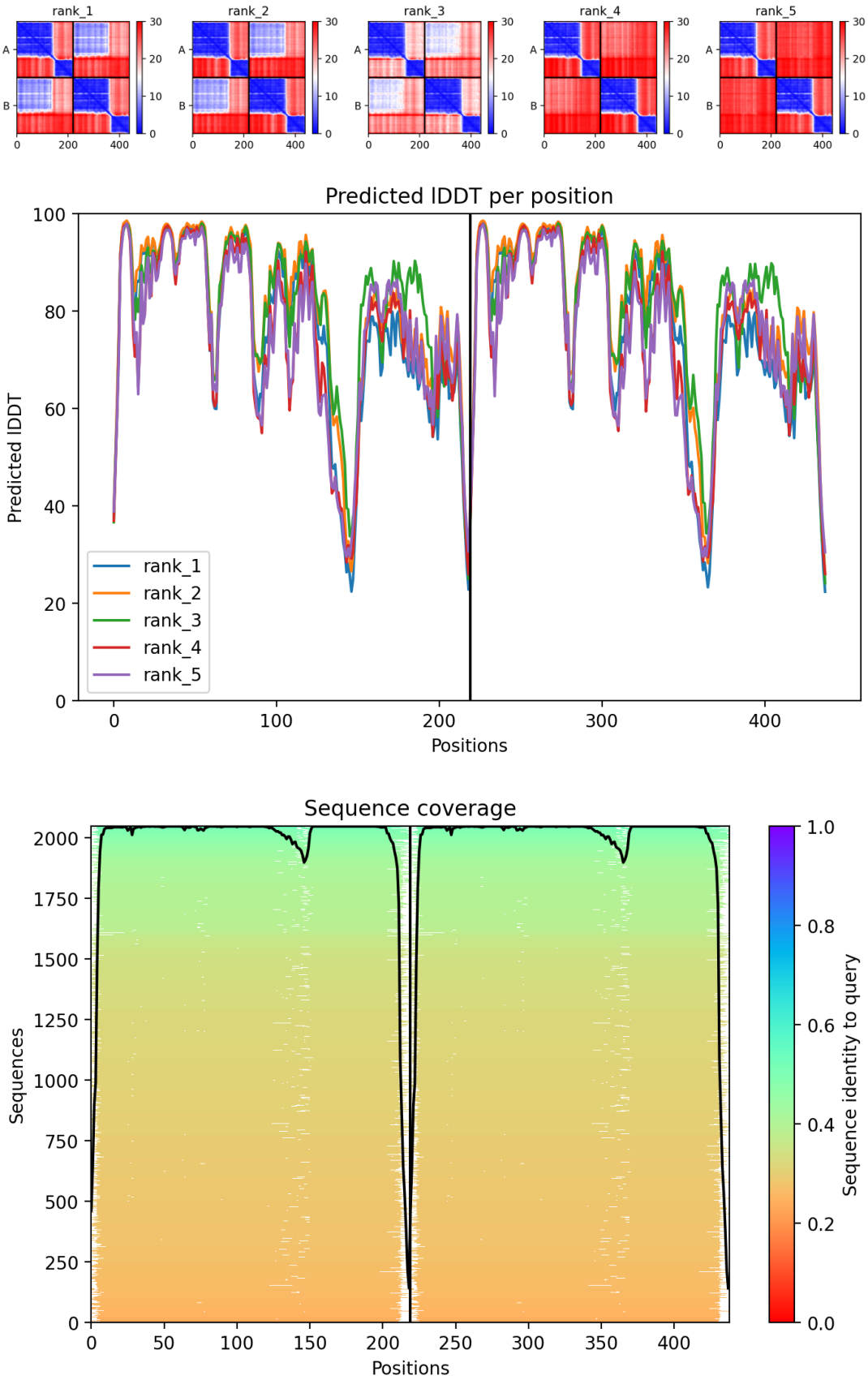
Appendix Figure 3 ATKA Pure purification of the native ForG using pET29a overexpression vector, a codon optimised gene insert and IPTG to induce production.



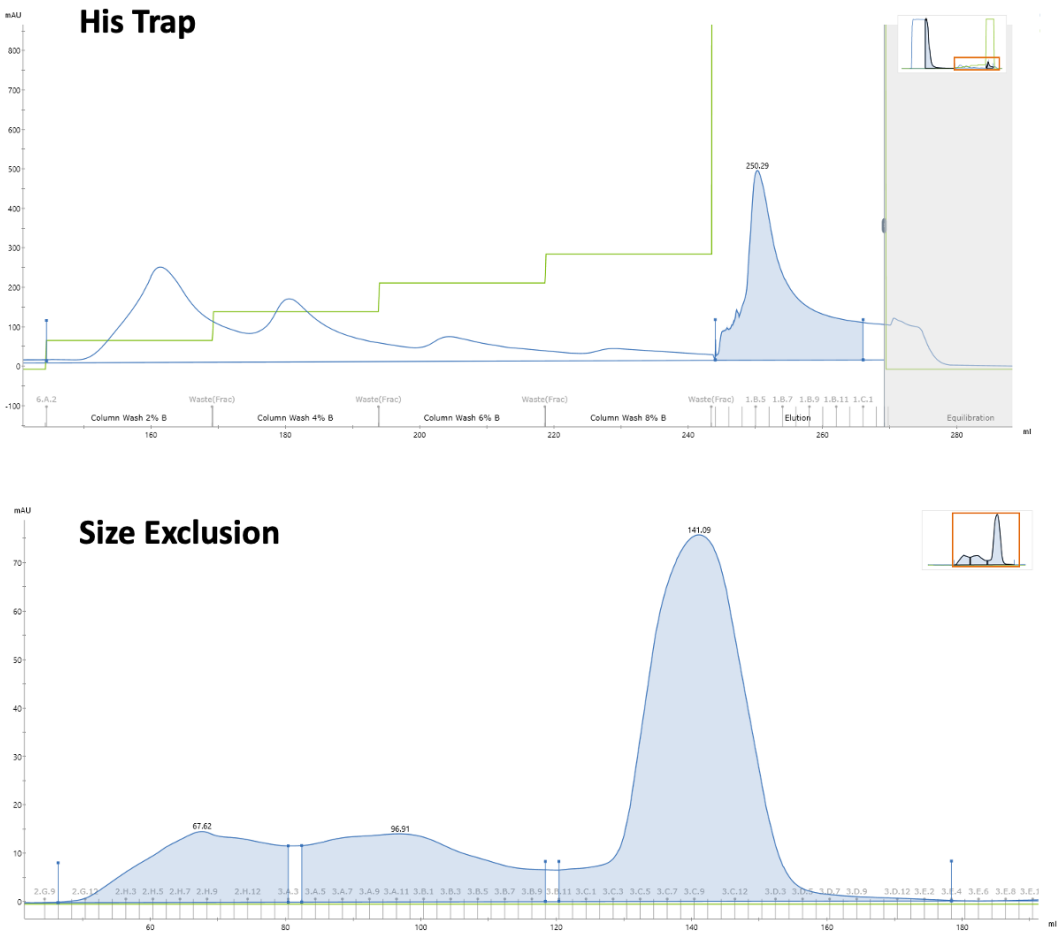
Appendix Figure 4 ATKA Pure purification of the native ForF using pET29a overexpression vector, a codon optimised gene insert and IPTG to induce production.



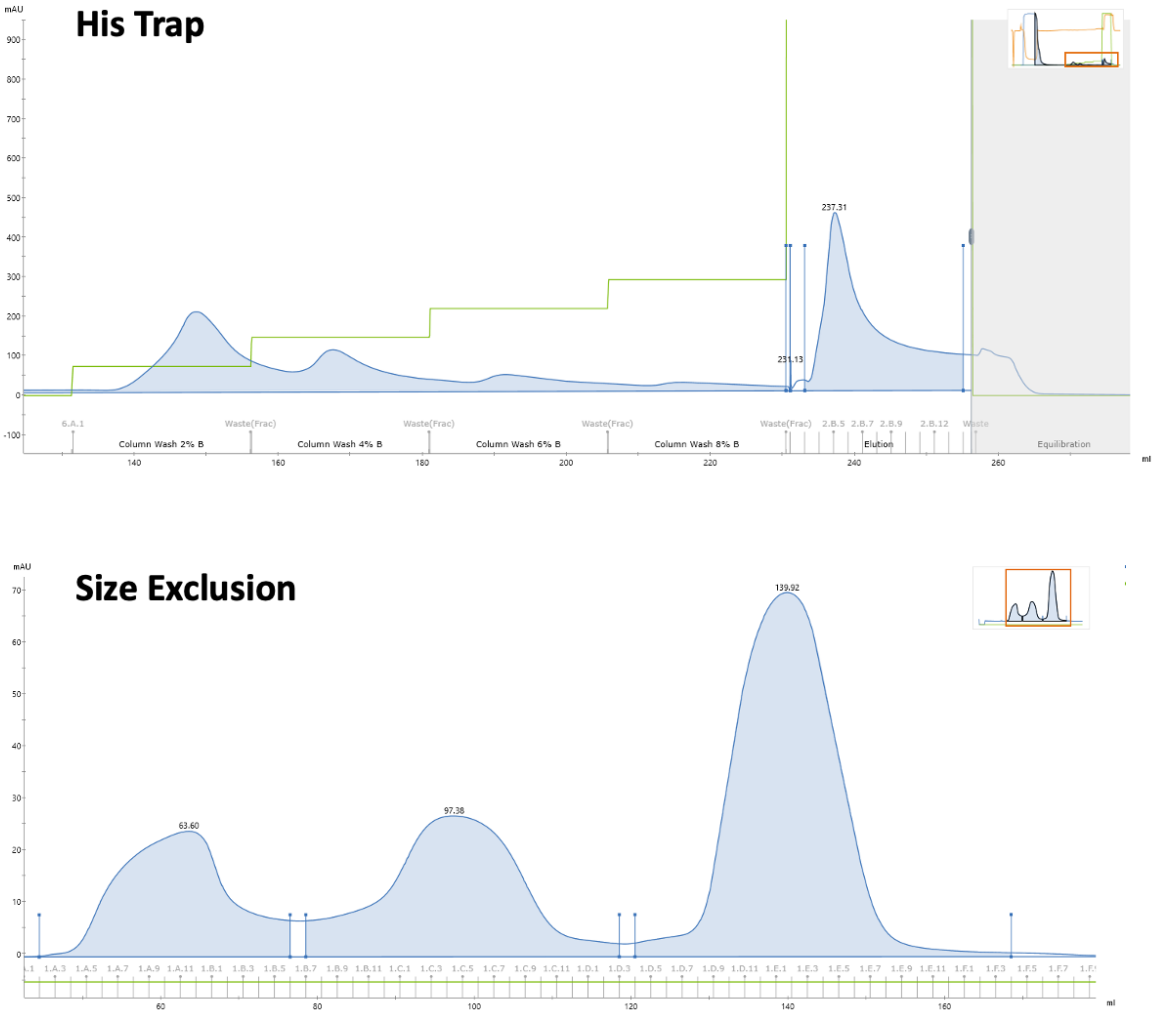
Appendix Figure 5 AlphaFold-2 prediction confidence and coverage information for the product of the ForF dimer containing a D53E mutation(KY5_6664).



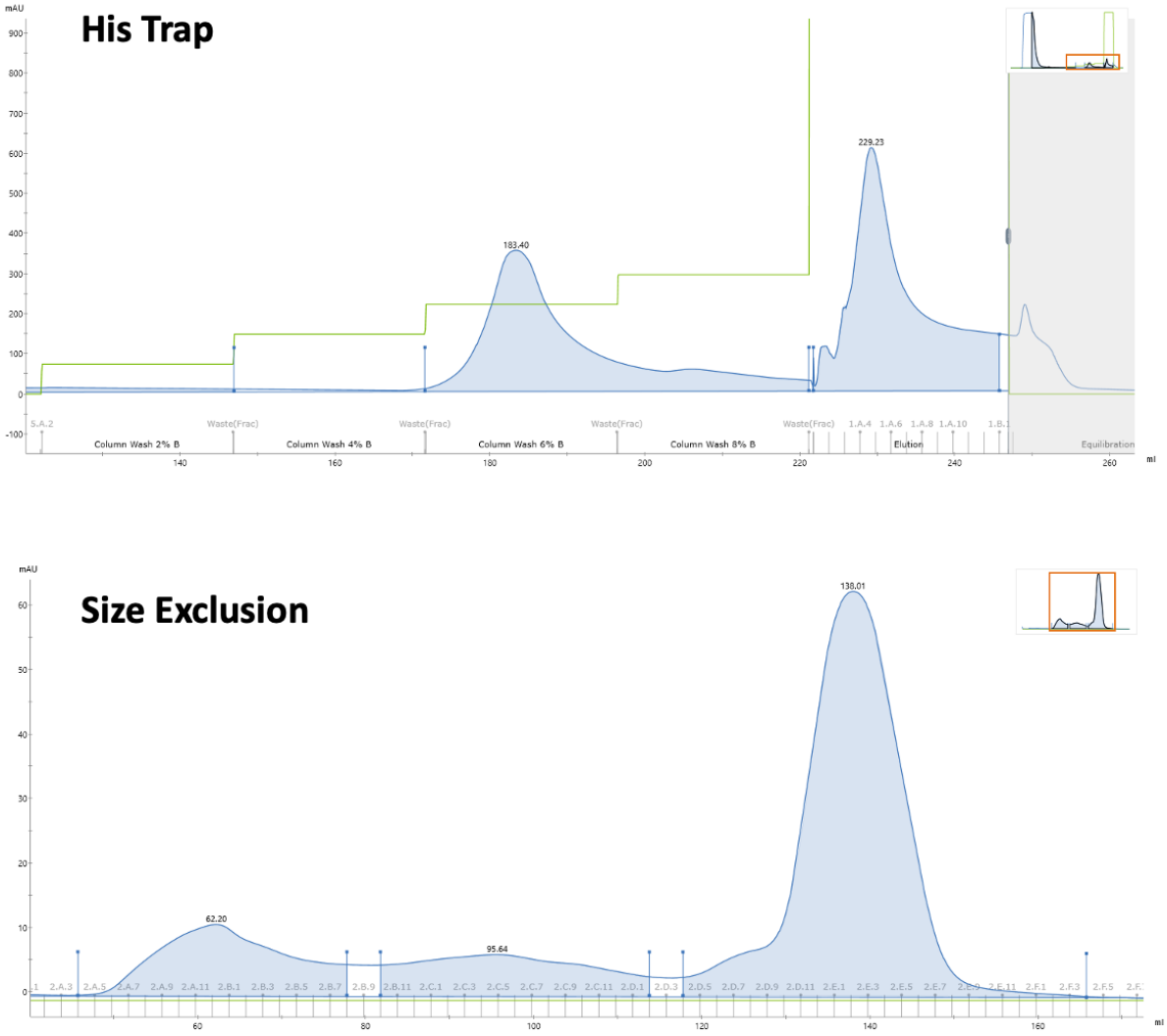
Appendix Figure 6 AlphaFold-2 prediction confidence and coverage information for the product of the ForF dimer containing a D5AE mutation(KY5_6664)



Appendix Figure 7 ATKA Pure purification of the ForG H175A variant using pET29a overexpression vector, a codon optimised gene insert and IPTG to induce production.



Appendix Figure 8 ATKA Pure purification of the ForF D53E variant using pET29a overexpression vector, a codon optimised gene insert and IPTG to induce production.



Appendix Figure 9 ATKA Pure purification of the ForF D53A variant using pET29a overexpression vector, a codon optimised gene insert and IPTG to induce production.

References

List of Countries Without Coronavirus | JANUARY 2022 - Koryo Tours. Retrieved September 14, 2023, from https://koryogroup.com/blog/are-there-countries-without-coronavirus

Aigle, B., & Corre, C. (2012). Waking up Streptomyces secondary metabolism by constitutive expression of activators or genetic disruption of repressors. *Methods in Enzymology*, *517*, 343–366. https://doi.org/10.1016/B978-0-12-404634-4.00017-6

Alam, K., Mazumder, A., Sikdar, S., Zhao, Y. M., Hao, J., Song, C., Wang, Y., Sarkar, R., Islam, S., Zhang, Y., & Li, A. (2022). Streptomyces: The biofactory of secondary metabolites. *Frontiers in Microbiology*, *13*, 968053. https://doi.org/10.3389/FMICB.2022.968053/BIBTEX

Al-Bassam, M. M., Bibb, M. J., Bush, M. J., Chandra, G., & Buttner, M. J. (2014). Response Regulator Heterodimer Formation Controls a Key Stage in Streptomyces Development. *PLOS Genetics*, *10*(8), e1004554. https://doi.org/10.1371/JOURNAL.PGEN.1004554

Allenby, N. E. E., Laing, E., Bucca, G., Kierzek, A. M., & Smith, C. P. (2012). Diverse control of metabolism and other cellular processes in Streptomyces coelicolor by the PhoP transcription factor: genome-wide identification of in vivo targets. *Nucleic Acids Research*, *40*(19), 9543–9556. https://doi.org/10.1093/NAR/GKS766

Anderson, T. B., Brian, P., & Champness, W. C. (2001). Genetic and transcriptional analysis of absA, an antibiotic gene cluster-linked two-component system that regulates multiple antibiotics in Streptomyces coelicolor. *Molecular Microbiology*, *39*(3), 553–566. https://doi.org/10.1046/J.1365-2958.2001.02240.X

Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. (2014). Retrieved March 7, 2019, from https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf

Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J.

S., … Van Der Donk, W. A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports*, *30*(1), 108. https://doi.org/10.1039/C2NP20085F

Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., Orhan, I. E., Banach, M., Rollinger, J. M., Barreca, D., Weckwerth, W., Bauer, R., Bayer, E. A., Majeed, M., Bishayee, A., Bochkov, V., Bonn, G. K., Braidy, N., Bucar, F., Cifuentes, A., D'Onofrio, G., Bodkin, M., … Supuran, C. T. (2021). Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery 2021 20:3*, *20*(3), 200–216. https://doi.org/10.1038/s41573-020-00114-z

Ayukekbong, J. A., Ntemgwa, M., & Atabe, A. N. (2017). The threat of antimicrobial resistance in developing countries: Causes and control strategies. In *Antimicrobial Resistance and Infection Control* (Vol. 6, Issue 1, p. 47). BioMed Central Ltd. https://doi.org/10.1186/s13756-017-0208-x

Baker, C. C. M., Martins, D. J., Pelaez, J. N., Billen, J. P. J., Pringle, A., Frederickson, M. E., & Pierce, N. E. (2017). Distinctive fungal communities in an obligate African ant-plant mutualism. *Proceedings. Biological Sciences*, *284*(1850), 20162501. https://doi.org/10.1098/rspb.2016.2501

Barke, J., Seipke, R. F., Grüschow, S., Heavens, D., Drou, N., Bibb, M. J., Goss, R. J., Yu, D. W., & Hutchings, M. I. (2010). A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant Acromyrmex octospinosus. *BMC Biology*, *8*(1), 109. https://doi.org/10.1186/1741-7007-8-109

Bartlett, J. G., Gilbert, D. N., & Spellberg, B. (2013). Seven ways to preserve the miracle of antibiotics. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *56*(10), 1445–1450. https://doi.org/10.1093/cid/cit070

Beltran-Alvarez, P., Cox, R. J., Crosby, J., & Simpson, T. J. (2007). Dissecting the component reactions catalyzed by the actinorhodin minimal polyketide synthase. *Biochemistry*, *46*(50), 14672–14681. https://doi.org/10.1021/BI701784C/ASSET/IMAGES/LARGE/BI701784CF00009.JPEG

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra,

G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., … Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature*, *417*(6885), 141–147. https://doi.org/10.1038/417141a

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology*, *8*(2), 208–215. https://doi.org/10.1016/J.MIB.2005.02.016

Bibb, M. J., Janssen, G. R., & Ward, J. M. (1985). Cloning and analysis of the promoter region of the erythromycin resistance gene (ermE) of Streptomyces erythraeus. *Gene*, *38*(1–3), 215–226. https://doi.org/10.1016/0378-1119(85)90220-3

Bishop, A., Fielding, S., Dyson, P., & Herron, P. (2004). Systematic insertional mutagenesis of a streptomycete genome: a link between osmoadaptation and antibiotic production. *Genome Research*, *14*(5), 893–900. https://doi.org/10.1101/GR.1710304

Blair, J. M. A., Richmond, G. E., & Piddock, L. J. V. (2014). Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. *Future Microbiology*, *9*(10), 1165–1177. https://doi.org/10.2217/FMB.14.66

Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. V. (2015). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, *13*(1), 42–51. https://doi.org/10.1038/nrmicro3380

Blatch, G. L., & Lä Ssle, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays*, *21*, 932–939. https://doi.org/10.1002/(SICI)1521-1878(199911)21:11

Blin, K., Pascal Andreu, V., de los Santos, E. L. C., Del Carratore, F., Lee, S. Y., Medema, M. H., & Weber, T. (2019). The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. *Nucleic Acids Research*, *47*(D1), D625–D630. https://doi.org/10.1093/nar/gky1060

Bobek, J., Šmídová, K., & Čihák, M. (2017). A waking review: Old and novel insights into the spore germination in Streptomyces. In *Frontiers in Microbiology* (Vol. 8, Issue NOV). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2017.02205

Bush, M. J., Bibb, M. J., Chandra, G., Findlay, K. C., & Buttner, M. J. (2013). Genes required for aerial growth, cell division, and chromosome segregation are targets of whia

before sporulation in Streptomyces venezuelae. *MBio*, *4*(5). https://doi.org/10.1128/mBio.00684-13

Byrne, M. K., Miellet, S., McGlinn, A., Fish, J., Meedya, S., Reynolds, N., & van Oijen, A. M. (2019). The drivers of antibiotic use and misuse: the development and investigation of a theory driven community measure. *BMC Public Health*, *19*(1), 1425. https://doi.org/10.1186/s12889-019-7796-8

Cane, D. E. (2010). Programming of Erythromycin Biosynthesis by a Modular Polyketide Synthase. *The Journal of Biological Chemistry*, *285*(36), 27517. https://doi.org/10.1074/JBC.R110.144618

Castro-Sánchez, E., Moore, L. S. P., Husson, F., & Holmes, A. H. (2016). What are the factors driving antimicrobial resistance? Perspectives from a public event in London, England. *BMC Infectious Diseases*, *16*(1). https://doi.org/10.1186/s12879-016-1810-x

Chater, K. F. (2016). Recent advances in understanding Streptomyces [version 1; referees: 4 approved]. In *F1000Research* (Vol. 5). Faculty of 1000 Ltd. https://doi.org/10.12688/f1000research.9534.1

Chater, K. F., & Bruton, C. J. (1985). Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. *The EMBO Journal*, *4*(7), 1893–1897. http://www.ncbi.nlm.nih.gov/pubmed/2992952

Chen, Y., Wendt-Pienkowski, E., & Shen, B. (2008). Identification and utility of FdmR1 as a Streptomyces antibiotic regulatory protein activator for fredericamycin production in Streptomyces griseus ATCC 49344 and heterologous hosts. *Journal of Bacteriology*, *190*(16), 5587–5596. https://doi.org/10.1128/JB.00592-08/SUPPL_FILE/JB00592_08_SI__REVISED___060508_.ZIP

Cobb, R. E., Wang, Y., & Zhao, H. (2015). High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synthetic Biology*, *4*(6), 723–728. https://doi.org/10.1021/SB500351F/SUPPL_FILE/SB500351F_SI_001.PDF

Cox, G., & Wright, G. D. (2013). Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. *International Journal of Medical Microbiology*, *303*(6–7), 287–292. https://doi.org/10.1016/J.IJMM.2013.02.009

Cruz-Bautista, R., Ruíz-Villafán, B., Romero-Rodríguez, A., Rodríguez-Sanoja, R., & Sánchez, S. (2023). Trends in the two-component system's role in the synthesis of antibiotics by Streptomyces. *Applied Microbiology and Biotechnology 2023 107:15*, *107*(15), 4727–4743. https://doi.org/10.1007/S00253-023-12623-Z

Culp, E. J., Waglechner, N., Wang, W., Fiebig-Comyn, A. A., Hsu, Y. P., Koteva, K., Sychantha, D., Coombes, B. K., Van Nieuwenhze, M. S., Brun, Y. V., & Wright, G. D. (2020). Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature*, *578*(7796), 582–587. https://doi.org/10.1038/S41586-020-1990-9

Cummings, M., Breitling, R., & Takano, E. (2014). Steps towards the synthetic biology of polyketide biosynthesis. *FEMS Microbiology Letters*, *351*(2), 116–125. https://doi.org/10.1111/1574-6968.12365

Currie, C. R., Scott, J. A., Summerbell, R. C., & Malloch, D. (1999). Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature*, *398*(6729), 701–704. https://doi.org/10.1038/19519

Dangel, V., Härle, J., Goerke, C., Wolz, C., Gust, B., Pernodet, J. L., & Heide, L. (2009). Transcriptional regulation of the novobiocin biosynthetic gene cluster. *Microbiology (Reading, England)*, *155*(Pt 12), 4025–4035. https://doi.org/10.1099/MIC.0.032649-0

Davies, J. (2013). Specialized microbial metabolites: functions and origins. *The Journal of Antibiotics 2013 66:7*, *66*(7), 361–364. https://doi.org/10.1038/ja.2013.61

Davies, J., & Davies, D. (2010). Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews : MMBR*, *74*(3), 417. https://doi.org/10.1128/MMBR.00016-10

Demain, A. L. (2009). Antibiotics: natural products essential to human health. *Medicinal Research Reviews*, *29*(6), 821–842. https://doi.org/10.1002/MED.20154

Den Hengst, C. D., Tran, N. T., Bibb, M. J., Chandra, G., Leskiw, B. K., & Buttner, M. J. (2010). Genes essential for morphological development and antibiotic production in Streptomyces coelicolor are targets of BldD during vegetative growth. *Molecular Microbiology*, *78*(2), 361–379. https://doi.org/10.1111/j.1365-2958.2010.07338.x

Devine, R., Hutchings, M. I., & Holmes, N. A. (2017). Future directions for the discovery of antibiotics from actinomycete bacteria. *Emerging Topics in Life Sciences*, *1*(1), 1–12. https://doi.org/10.1042/ETLS20160014

Devine, R., McDonald, H. P., Qin, Z., Arnold, C. J., Noble, K., Chandra, G., Wilkinson, B., & Hutchings, M. I. (2021). Re-wiring the regulation of the formicamycin biosynthetic gene cluster to enable the development of promising antibacterial compounds. *Cell Chemical Biology*, *28*(4), 515-523.e5. https://doi.org/10.1016/J.CHEMBIOL.2020.12.011

Dias, D. A., Urban, S., & Roessner, U. (2012). A Historical Overview of Natural Products in Drug Discovery. *Metabolites*, *2*(2), 303. https://doi.org/10.3390/METABO2020303

Djordjevic, S., Goudreau, P. N., Xu, Q., Stock, A. M., & West, A. H. (1998). Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(4), 1381–1386. https://doi.org/10.1073/pnas.95.4.1381

Donald, L., Pipite, A., Subramani, R., Owen, J., Keyzers, R. A., & Taufa, T. (2022). Streptomyces: Still the Biggest Producer of New Natural Secondary Metabolites, a Current Perspective. *Microbiology Research*, *13*(3), 418–465. https://doi.org/10.3390/MICROBIOLRES13030031/S1

Dubos, R. J. (1939). Studies on a bactericidal agent extracted from a soil bacillus: I. Preparation of the agent. Its activity in vitro. *Journal of Experimental Medicine*, *70*(1), 1–10. https://doi.org/10.1084/jem.70.1.1

Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L., & Skiniotis, G. (2014). Structure of a modular polyketide synthase. *Nature*, *510*(7506), 512–517. https://doi.org/10.1038/NATURE13423

Elliot, M. A., Buttner, M. J., & Nodwell, J. R. (2014). Multicellular Development in Streptomyces. In *Myxobacteria* (pp. 419–438). American Society of Microbiology. https://doi.org/10.1128/9781555815677.ch24

Feeney, M. A., Chandra, G., Findlay, K. C., Paget, M. S. B., & Buttnera, M. J. (2017). Translational control of the Sigr directed oxidative stress response in Streptomyces via

IF3-mediated repression of a noncanonical GTC start codon. *MBio*, *8*(3). https://doi.org/10.1128/MBIO.00815-17/SUPPL_FILE/MBO003173348ST2.DOCX

Feng, Z., Chakraborty, D., Dewell, S. B., Reddy, B. V. B., & Brady, S. F. (2012). Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis. *Journal of the American Chemical Society*, *134*(6), 2981–2987. https://doi.org/10.1021/ja207662w

Fischbach, M. A., & Walsh, C. T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. *Chemical Reviews*, *106*(8), 3468–3496. https://doi.org/10.1021/CR0503097/ASSET/CR0503097.FP.PNG_V03

Fishovitz, J., Hermoso, J. A., Chang, M., & Mobashery, S. (2014). Penicillin-binding protein 2a of methicillin-resistant Staphylococcus aureus. *IUBMB Life*, *66*(8), 572–577. https://doi.org/10.1002/IUB.1289

Flärdh, K., & Buttner, M. J. (2009). Streptomyces morphogenetics: Dissecting differentiation in a filamentous bacterium. In *Nature Reviews Microbiology* (Vol. 7, Issue 1, pp. 36–49). Nature Publishing Group. https://doi.org/10.1038/nrmicro1968

Fleming, A. (1929). On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *British Journal of Experimental Pathology*, *10*(3), 226. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2048009/

Fleming A. (1945). *Sir Alexander Fleming – Nobel Lecture - NobelPrize.org*. https://www.nobelprize.org/prizes/medicine/1945/fleming/lecture/

Freeman, Z. N., Dorus, S., & Waterfield, N. R. (2013). The KdpD/KdpE Two-Component System: Integrating K+ Homeostasis and Virulence. *PLoS Pathogens*, *9*(3). https://doi.org/10.1371/JOURNAL.PPAT.1003201

Fu, J., Qin, R., Zong, G., Liu, C., Kang, N., Zhong, C., & Cao, G. (2019). The CagRS two-component system regulates clavulanic acid metabolism via multiple pathways in Streptomyces clavuligerus F613-1. *Frontiers in Microbiology*, *10*(FEB), 426463. https://doi.org/10.3389/FMICB.2019.00244/BIBTEX

Gómez, C., Olano, C., Méndez, C., & Salas, J. A. (2012). Three pathway-specific regulators control streptolydigin biosynthesis in Streptomyces lydicus. *Microbiology (United*

Kingdom), *158*(10), 2504–2514. https://doi.org/10.1099/MIC.0.061325-0/CITE/REFWORKS

Goulian, M. (2010). Two-component signaling circuit structure and properties. In *Current Opinion in Microbiology* (Vol. 13, Issue 2, pp. 184–189). NIH Public Access. https://doi.org/10.1016/j.mib.2010.01.009

Gregory, M. A., Till, R., & Smith, M. C. M. (2003). Integration site for Streptomyces phage ϕBT1 and development of site-specific integrating vectors. *Journal of Bacteriology*, *185*(17), 5320–5323. https://doi.org/10.1128/JB.185.17.5320-5323.2003/ASSET/65BA751E-E8BE-4E84-AFF1-E9A247466F3D/ASSETS/GRAPHIC/JB1730267002.JPEG

Greule, A., Marolt, M., Deubel, D., Peintner, I., Zhang, S., Jessen-Trefzer, C., De Ford, C., Burschel, S., Li, S. M., Friedrich, T., Merfort, I., Lüdeke, S., Bisel, P., Müller, M., Paululat, T., & Bechthold, A. (2017). Wide distribution of foxicin biosynthetic gene clusters in Streptomyces strains - An unusual secondary metabolite with various properties. *Frontiers in Microbiology*, *8*(FEB), 245162. https://doi.org/10.3389/FMICB.2017.00221/BIBTEX

Groisman, E. A. (2016). Feedback Control of Two-Component Regulatory Systems. *Annual Review of Microbiology*, *70*(1), 103–124. https://doi.org/10.1146/annurev-micro-102215-095331

Grove, A. (2013). MarR family transcription factors. In *Current Biology* (Vol. 23, Issue 4, pp. R142–R143). Cell Press. https://doi.org/10.1016/j.cub.2013.01.013

Grove, A. (2017). Regulation of Metabolic Pathways by MarR Family Transcription Factors. *Computational and Structural Biotechnology Journal*, *15*, 366. https://doi.org/10.1016/J.CSBJ.2017.06.001

Gullón, S., Vicente, R. L., & Mellado, R. P. (2012). A Novel Two-Component System Involved in Secretion Stress Response in Streptomyces lividans. *PLOS ONE*, *7*(11), e48987. https://doi.org/10.1371/JOURNAL.PONE.0048987

Guo, J., Zhang, X., Lu, X., Liu, W., Chen, Z., Li, J., Deng, L., & Wen, Y. (2018). SAV4189, a MarR-family regulator in Streptomyces avermitilis, activates avermectin biosynthesis. *Frontiers in Microbiology*, *9*(JUN). https://doi.org/10.3389/FMICB.2018.01358/FULL

Gupta, A., & Grove, A. (2014). Ligand-binding pocket bridges DNA-binding and dimerization domains of the urate-responsive MarR homologue MftR from Burkholderia thailandensis. *Biochemistry*, *53*(27), 4368–4380. https://doi.org/10.1021/BI500219T

Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., & Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(17), 9390–9395. https://doi.org/10.1073/PNAS.160270797/ASSET/0B9D7317-E2EB-4720-835D-4394FC3086C3/ASSETS/GRAPHIC/PQ1602707003.JPEG

Gygli, S. M., Borrell, S., Trauner, A., & Gagneux, S. (2017). Antimicrobial resistance in Mycobacterium tuberculosis: mechanistic and evolutionary perspectives. *FEMS Microbiology Reviews*, *41*(3), 354–373. https://doi.org/10.1093/FEMSRE/FUX011

Haas, L. F. (1999). Papyrus of Ebers and Smith. *Journal of Neurology, Neurosurgery, and Psychiatry*, *67*(5), 578. https://doi.org/10.1136/JNNP.67.5.578

Hackl, S., & Bechthold, A. (2015). The Gene bldA, a Regulator of Morphological Differentiation and Antibiotic Production in Streptomyces. *Archiv Der Pharmazie*, *348*(7), 455–462. https://doi.org/10.1002/ARDP.201500073

Hancock, R. E. W., & Brinkman, F. S. L. (2002). Function of pseudomonas porins in uptake and efflux. *Annual Review of Microbiology*, *56*, 17–38. https://doi.org/10.1146/ANNUREV.MICRO.56.012302.160310

Hawkey, P. M. (1998). The origins and molecular basis of antibiotic resistance. *BMJ (Clinical Research Ed.)*, *317*(7159), 657–660. https://doi.org/10.1136/BMJ.317.7159.657

Heermann, R., Lippert, M. L., & Jung, K. (2009). Domain swapping reveals that the N-terminal domain of the sensor kinase KdpD in Escherichia coli is important for signaling. *BMC Microbiology*, *9*, 133. https://doi.org/10.1186/1471-2180-9-133

Heine, D., Holmes, N. A., Worsley, S. F., Santos, A. C. A., Innocent, T. M., Scherlach, K., Patrick, E. H., Yu, D. W., Murrell, J. C., Vieria, P. C., Boomsma, J. J., Hertweck, C., Hutchings, M. I., & Wilkinson, B. (2018). Chemical warfare between leafcutter ant symbionts and a co-evolved pathogen. *Nature Communications*, *9*(1), 2208. https://doi.org/10.1038/s41467-018-04520-1

Hertweck, C., Luzhetskyy, A., Rebets, Y., & Bechthold, A. (2007). Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural Product Reports*, *24*(1), 162–190. https://doi.org/10.1039/B507395M

Hesketh, A., Chen, W. J., Ryding, J., Chang, S., & Bibb, M. (2007). The global role of ppGpp synthesis in morphological differentiation and antibiotic production in Streptomyces coelicolor A3(2). *Genome Biology*, *8*(8), R161. https://doi.org/10.1186/gb-2007-8-8-r161

Hesketh, A., Fink, D., Gust, B., Rexer, H. U., Scheel, B., Chater, K., Wohlleben, W., & Engels, A. (2002). The GlnD and GlnK homologues of Streptomyces coelicolor A3(2) are functionally dissimilar to their nitrogen regulatory system counterparts from enteric bacteria. *Molecular Microbiology*, *46*(2), 319–330. https://doi.org/10.1046/J.1365-2958.2002.03149.X

Higo, A., Horinouchi, S., & Ohnishi, Y. (2011). Strict regulation of morphological differentiation and secondary metabolism by a positive feedback loop between two global regulators AdpA and BldA in Streptomyces griseus. *Molecular Microbiology*, *81*(6), 1607–1622. https://doi.org/10.1111/J.1365-2958.2011.07795.X

Hobbs, J. K., & Boraston, A. B. (2019). (p)ppGpp and the Stringent Response: An Emerging Threat to Antibiotic Therapy. *ACS Infectious Diseases*, *5*(9), 1505–1517. https://doi.org/10.1021/acsinfecdis.9b00204

Hodgkin, D. C. (1949). The X-ray analysis of the structure of penicillin. *Advancement of Science*, *6*(22), 85–89. https://europepmc.org/article/MED/18134678

Holmes, N. A., Devine, R., Qin, Z., Seipke, R. F., Wilkinson, B., & Hutchings, M. I. (2018). Complete genome sequence of Streptomyces formicae KY5, the formicamycin producer. *Journal of Biotechnology*, *265*, 116–118. https://doi.org/10.1016/J.JBIOTEC.2017.11.011

Hong, H. J., Hutchings, M. I., Hill, L. M., & Buttner, M. J. (2005). The role of the novel Fem protein VanK in vancomycin resistance in Streptomyces coelicolor. *The Journal of Biological Chemistry*, *280*(13), 13055–13061. https://doi.org/10.1074/JBC.M413801200

Hopwood, D. A. (2006). Soil To Genomics: The *Streptomyces* Chromosome. *Annual Review of Genetics*, *40*(1), 1–23. https://doi.org/10.1146/annurev.genet.40.110405.090639

Horbal, L., Marques, F., Nadmid, S., Mendes, M. V., & Luzhetskyy, A. (2018). Secondary metabolites overproduction through transcriptional gene cluster refactoring. *Metabolic Engineering*, *49*, 299–315. https://doi.org/10.1016/J.YMBEN.2018.09.010

Hörnschemeyer, P., Liss, V., Heermann, R., Jung, K., & Hunke, S. (2016). Interaction Analysis of a Two-Component System Using Nanodiscs. *PLOS ONE*, *11*(2), e0149187. https://doi.org/10.1371/JOURNAL.PONE.0149187

Huang, Z., Hwang, P., Watson, D. S., Cao, L., & Szoka, F. C. (2009). Tris-Nitrilotriacetic Acids of Sub-nanomolar Affinity Toward Hexahistidine Tagged Molecules. *Bioconjugate Chemistry*, *20*(8), 1667. https://doi.org/10.1021/BC900309N

Hünnefeld, M., Persicke, M., Kalinowski, J., & Frunzke, J. (2019). The MarR-Type Regulator MalR Is Involved in Stress-Responsive Cell Envelope Remodeling in Corynebacterium glutamicum. *Frontiers in Microbiology*, *10*(MAY), 1039. https://doi.org/10.3389/fmicb.2019.01039

Hunt, A. C., Servín-González, L., Kelemen, G. H., & Buttner, M. J. (2005). The bldC developmental locus of Streptomyces coelicolor encodes a member of a family of small DNA-binding proteins related to the DNA-binding domains of the MerR family. *Journal of Bacteriology*, *187*(2), 716–728. https://doi.org/10.1128/JB.187.2.716-728.2005

Hutchings, M., Truman, A., & Wilkinson, B. (2019). Antibiotics: past, present and future. In *Current Opinion in Microbiology* (Vol. 51, pp. 72–80). Elsevier Ltd. https://doi.org/10.1016/j.mib.2019.10.008

Hwang, S., Lee, N., Cho, S., Palsson, B., & Cho, B. K. (2020). Repurposing Modular Polyketide Synthases and Non-ribosomal Peptide Synthetases for Novel Chemical Biosynthesis. *Frontiers in Molecular Biosciences*, *7*, 540848. https://doi.org/10.3389/FMOLB.2020.00087/BIBTEX

Ishii, E., & Eguchi, Y. (2021). Diversity in Sensing and Signaling of Bacterial Sensor Histidine Kinases. *Biomolecules*, *11*(10). https://doi.org/10.3390/BIOM11101524

Jin, S., Hui, M., Lu, Y., & Zhao, Y. (2023). An overview on the two-component systems of Streptomyces coelicolor. *World Journal of Microbiology and Biotechnology 2023 39:3*, *39*(3), 1–17. https://doi.org/10.1007/S11274-023-03522-6

Jones, S. E., & Elliot, M. A. (2018). 'Exploring' the regulation of Streptomyces growth and development. In *Current Opinion in Microbiology* (Vol. 42, pp. 25–30). Elsevier Ltd. https://doi.org/10.1016/j.mib.2017.09.009

Jones, S. E., Ho, L., Rees, C. A., Hill, J. E., Nodwell, J. R., & Elliot, M. A. (2017). Streptomyces exploration is triggered by fungal interactions and volatile signals. *ELife*, *6*. https://doi.org/10.7554/eLife.21738

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, *596*(7873), 583–589. https://doi.org/10.1038/S41586-021-03819-2

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., & Hopwood, D. A. (2000). *Practical Streptomyces Genetics. A Laboratory Manual.* 613. https://search.worldcat.org/title/456092084

Kleine, B., Chattopadhyay, A., Polen, T., Pinto, D., Mascher, T., Bott, M., Brocker, M., & Freudl, R. (2017). The three-component system EsrISR regulates a cell envelope stress response in Corynebacterium glutamicum. *Molecular Microbiology*, *106*(5), 719–741. https://doi.org/10.1111/MMI.13839

Koike, S., Aminov, R., & Mackie, R. (2017). *Agricultural use of antibiotics and antibiotic resistance The interaction of genetic and environmental factors in the provoking of rheumatoid arthritis View project Ruminal bacteria inoculation View project AGRICULTURAL USE OF ANTIBIOTICS AND ANTIBIOTIC RESISTANCE*. https://www.researchgate.net/publication/315613886

Kortt, A. A., Oddie, G. W., Iliades, P., Gruen, L. C., & Hudson, P. J. (1997). Nonspecific amine immobilization of ligand can Be a potential source of error in BIAcore binding experiments and may reduce binding affinities. *Analytical Biochemistry*, *253*(1), 103–111. https://doi.org/10.1006/ABIO.1997.2333

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Sun, F., Jit, M., Munday, J. D., Davies, N., Gimma, A., van Zandvoort, K., Gibbs, H., Hellewell, J., Jarvis, C. I., Clifford, S., Quilty, B. J., Bosse, N. I., … Flasche, S. (2020). Early dynamics

of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(20)30144-4

Kumar, A., & Schweizer, H. P. (2005). Bacterial resistance to antibiotics: active efflux and reduced uptake. *Advanced Drug Delivery Reviews*, *57*(10), 1486–1513. https://doi.org/10.1016/J.ADDR.2005.04.004

Kundu, J., Verma, A., Verma, I., Bhadada, S. K., & Sharma, S. (2021). Molecular mechanism of interaction of Mycobacterium tuberculosis with host macrophages under high glucose conditions. *Biochemistry and Biophysics Reports*, *26*. https://doi.org/10.1016/J.BBREP.2021.100997

Krysenko, S. (2023). Impact of Nitrogen-Containing Compounds on Secondary Metabolism in Streptomyces spp.—A Source of Metabolic Engineering Strategies. *SynBio 2023, Vol. 1, Pages 204-225*, *1*(3), 204–225. https://doi.org/10.3390/SYNBIO1030015

Laarman, A. J., Ruyken, M., Malone, C. L., van Strijp, J. A. G., Horswill, A. R., & Rooijakkers, S. H. M. (2011). Staphylococcus aureus metalloprotease aureolysin cleaves complement C3 to mediate immune evasion. *Journal of Immunology (Baltimore, Md. : 1950)*, *186*(11), 6445–6453. https://doi.org/10.4049/JIMMUNOL.1002948

Larsson, D. G. J. (2014). Antibiotics in the environment. In *Upsala Journal of Medical Sciences* (Vol. 119, Issue 2, pp. 108–112). Informa Healthcare. https://doi.org/10.3109/03009734.2014.896438

Law, J. W.-F., Tan, K.-X., Wong, S. H., Ab Mutalib, N.-S., & Lee, L.-H. (2018). Taxonomic and Characterization Methods of Streptomyces: A Review. *Progress In Microbes & Molecular Biology*, *1*(1). https://doi.org/10.36877/pmmb.a0000009

Lazar, J. T., & Tabor, J. J. (2021). Bacterial two-component systems as sensors for synthetic biology applications. *Current Opinion in Systems Biology*, *28*, 100398. https://doi.org/10.1016/J.COISB.2021.100398

Leekha, S., Terrell, C. L., & Edson, R. S. (2011). General principles of antimicrobial therapy. *Mayo Clinic Proceedings*, *86*(2), 156–167. https://doi.org/10.4065/mcp.2010.0639

Li, L., Jiang, W., & Lu, Y. (2017). A Novel Two-Component System, GluR-GluK, Involved in Glutamate Sensing and Uptake in Streptomyces coelicolor. *Journal of Bacteriology*, *199*(18). https://doi.org/10.1128/JB.00097-17

Li, L., Maclntyre, L. W., & Brady, S. F. (2021). Refactoring biosynthetic gene clusters for heterologous production of microbial natural products. *Current Opinion in Biotechnology*, *69*, 145–152. https://doi.org/10.1016/J.COPBIO.2020.12.011

Li, X., Wang, J., Shi, M., Wang, W., Corre, C., & Yang, K. (2017). Evidence for the formation of ScbR/ScbR2 heterodimers and identification of one of the regulatory targets in Streptomyces coelicolor. *Applied Microbiology and Biotechnology*, *101*(13), 5333–5340. https://doi.org/10.1007/S00253-017-8275-8/FIGURES/3

Lindel, T., Jensen, P. R., & Fenical, W. (1996). Lagunapyrones A-C: Cytotoxic acetogenins of a new skeletal class from a marine sediment bacterium. *Tetrahedron Letters*, *37*(9), 1327–1330. https://doi.org/10.1016/0040-4039(96)00014-7

Liu, G., Chater, K. F., Chandra, G., Niu, G., & Tan, H. (2013). Molecular Regulation of Antibiotic Biosynthesis in Streptomyces. *Microbiology and Molecular Biology Reviews*, *77*(1), 112–143. https://doi.org/10.1128/mmbr.00054-12

Liu, M., Xu, W., Zhu, Y., Cui, X., & Pang, X. (2021). The Response Regulator MacR and its Potential in Improvement of Antibiotic Production in Streptomyces coelicolor. *Current Microbiology*, *78*(10), 3696–3707. https://doi.org/10.1007/S00284-021-02633-3/FIGURES/4

Liu, M., Zhang, P., Zhu, Y., Lu, T., Wang, Y., Cao, G., Shi, M., Chen, X. L., Tao, M., & Pang, X. (2019). Novel two-component system MacRS is a pleiotropic regulator that controls multiple morphogenic membrane protein genes in Streptomyces coelicolor. *Applied and Environmental Microbiology*, *85*(4). https://doi.org/10.1128/AEM.02178-18

Lobanovska, M., & Pilla, G. (2017). Focus: Drug Development: Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future? *The Yale Journal of Biology and Medicine*, *90*(1), 135. /pmc/articles/PMC5369031/

Lyddiard, D., Jones, G. L., & Greatrex, B. W. (2016). Keeping it simple: lessons from the golden era of antibiotic discovery. *FEMS Microbiology Letters*, *363*(8). https://doi.org/10.1093/femsle/fnw084

Makitrynskyy, R., Tsypik, O., Desir`, D., Nuzzo, D., Paululat, T., Zechel, D. L., & Bechthold, A. (2020). Secondary nucleotide messenger c-di-GMP exerts a global control on natural product biosynthesis in streptomycetes. *Nucleic Acids Research*, *48*(3), 1583–1598. https://doi.org/10.1093/nar/gkz1220

Malpartida, F., & Hopwood, D. A. (1984). Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature 1984 309:5967*, *309*(5967), 462–464. https://doi.org/10.1038/309462A0

Manyi-Loh, C., Mamphweli, S., Meyer, E., & Okoh, A. (2018). Antibiotic use in agriculture and its consequential resistance in environmental sources: Potential public health implications. In *Molecules* (Vol. 23, Issue 4). MDPI AG. https://doi.org/10.3390/molecules23040795

Marks, P. (2009). Aviation plans for a future pandemic. *New Scientist*, *202*(2709), 18–19. https://doi.org/10.1016/S0262-4079(09)61370-1

Marraffini, L. A., DeDent, A. C., & Schneewind, O. (2006). Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. *Microbiology and Molecular Biology Reviews*, *70*(1), 192. https://doi.org/10.1128/MMBR.70.1.192-221.2006

Martín, J. F. (2004). Phosphate Control of the Biosynthesis of Antibiotics and Other Secondary Metabolites Is Mediated by the PhoR-PhoP System: an Unfinished Story Downloaded from. *JOURNAL OF BACTERIOLOGY*, *186*(16), 5197–5201. https://doi.org/10.1128/JB.186.16.5197-5201.2004

Martínez-Hackert, E., & Stock, A. M. (1997). Structural relationships in the OmpR family of winged-helix transcription factors. In *Journal of Molecular Biology* (Vol. 269, Issue 3, pp. 301–312). Academic Press. https://doi.org/10.1006/jmbi.1997.1065

Martins, D. J. (2010). Not all ants are equal: obligate acacia ants provide different levels of protection against mega-herbivores. *African Journal of Ecology*, *48*(4), 1115–1122. https://doi.org/10.1111/j.1365-2028.2010.01226.x

McBride, C. M., Miller, E. L., & Charkoudian, L. K. (2023). An updated catalogue of diverse type II polyketide synthase biosynthetic gene clusters captured from large-scale nucleotide databases. *Microbial Genomics*, *9*(3). https://doi.org/10.1099/MGEN.0.000965

McLean, T. C., Beaton, A. D. M., Martins, C., Saalbach, G., Chandra, G., Wilkinson, B., & Hutchings, M. I. (2023). Evidence of a role for CutRS and actinorhodin in the secretion stress response in Streptomyces coelicolor M145. *Microbiology*, *169*(7), 1358. https://doi.org/10.1099/MIC.0.001358

McLean, T. C., Lo, R., Tschowri, N., Hoskisson, P. A., Al Bassam, M. M., Hutchings, M. I., & Som, N. F. (2019). Sensing and responding to diverse extracellular signals: An updated analysis of the sensor kinases and response regulators of streptomyces species. *Microbiology (United Kingdom)*, *165*(9), 929–952. https://doi.org/10.1099/mic.0.000817

McLean, T. C., Wilkinson, B., Hutchings, M. I., & Devine, R. (2019). Dissolution of the Disparate: Co-ordinate Regulation in Antibiotic Biosynthesis. *Antibiotics*, *8*(2). https://doi.org/10.3390/ANTIBIOTICS8020083

McManus, J., Cheng, Z., & Vogel, C. (2015). Next-generation analysis of gene expression regulation--comparing the roles of synthesis and degradation. *Molecular BioSystems*, *11*(10), 2680–2689. https://doi.org/10.1039/C5MB00310E

Merritt, J., & Kuehn, S. (2016). Resource competition: When communities collide. *ELife*, *5*(2016JULY). https://doi.org/10.7554/eLife.18753

Miller, W. R., Munita, J. M., & Arias, C. A. (2014). Mechanisms of antibiotic resistance in enterococci. *Expert Review of Anti-Infective Therapy*, *12*(10), 1221. https://doi.org/10.1586/14787210.2014.956092

Montalbán-López, M., Scott, T. A., Ramesh, S., Rahman, I. R., Van Heel, A. J., Viel, J. H., Bandarian, V., Dittmann, E., Genilloud, O., Goto, Y., Grande Burgos, M. J., Hill, C., Kim, S., Koehnke, J., Latham, J. A., Link, A. J., Martínez, B., Nair, S. K., Nicolet, Y., … Van Der Donk, W. A. (2021). New developments in RiPP discovery, enzymology and engineering. *Natural Product Reports*, *38*(1), 130–239. https://doi.org/10.1039/D0NP00027B

Mootz, H. D., Schwarzer, D., & Marahiel, M. A. (2002). *Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases***. https://doi.org/10.1002/1439-7633

Munita, J., & Arias, C. (2016). Mechanisms of Antibiotic Resistance. *American Society for Microbiology*, *4*(1). https://doi.org/10.1128/microbiolspec.VMBF-0016-2015

Ng, E. Y. W., Trucksis, M., & Hooper, D. C. (1994). Quinolone resistance mediated by norA: physiologic characterization and relationship to flqB, a quinolone resistance locus on the Staphylococcus aureus chromosome. *Antimicrobial Agents and Chemotherapy*, *38*(6), 1345–1355. https://doi.org/10.1128/AAC.38.6.1345

Nivina, A., Yuet, K. P., Hsu, J., & Khosla, C. (2019). Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chemical Reviews*, *119*(24), 12524–12547. https://doi.org/10.1021/ACS.CHEMREV.9B00525/ASSET/IMAGES/LARGE/CR9B00525_0010.JPEG

Nobary, S. G., & Jensen, S. E. (2012). A comparison of the clavam biosynthetic gene clusters in Streptomyces antibioticus Tü1718 and Streptomyces clavuligerus. *Canadian Journal of Microbiology*, *58*(4), 413–425. https://doi.org/10.1139/W2012-012

Nolan, E. M., & Walsh, C. T. (2009). How Nature Morphs Peptide Scaffolds into Antibiotics. *Chembiochem : A European Journal of Chemical Biology*, *10*(1), 34. https://doi.org/10.1002/CBIC.200800438

Norrby, S. R., Nord, C. E., & Finch, R. (2005). Lack of development of new antimicrobial drugs: a potential serious threat to public health. *The Lancet Infectious Diseases*, *5*(2), 115–119. https://doi.org/10.1016/s1473-3099(05)01283-1

North, R. D., Jackson, C. W., & Howse, P. E. (1997). Evolutionary aspects of ant-fungus interactions in leaf-cutting ants. *Trends in Ecology & Evolution*, *12*(10), 386–389. https://doi.org/10.1016/S0169-5347(97)87381-8

Ochi, K. (1987). A rel mutation abolishes the enzyme induction needed for actinomycin synthesis by streptomyces antibioticus. *Agricultural and Biological Chemistry*, *51*(3), 829–835. https://doi.org/10.1080/00021369.1987.10868121

Oh, D.-C., Poulsen, M., Currie, C. R., & Clardy, J. (2009). Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis. *Nature Chemical Biology*, *5*(6), 391–393. https://doi.org/10.1038/nchembio.159

Okamoto, S., Taguchi, T., Ochi, K., & Ichinose, K. (2009). Biosynthesis of Actinorhodin and Related Antibiotics: Discovery of Alternative Routes for Quinone Formation Encoded in the act Gene Cluster. *Chemistry & Biology*, *16*(2), 226–236. https://doi.org/10.1016/J.CHEMBIOL.2009.01.015

Oliynyk, M., Brown, M. J. B., Cortes, J., Staunton, J., & Leadlay, P. F. (1996). A hybrid modular polyketide synthase obtained by domain swapping. *Chemistry & Biology*, *3*(10), 833–839. https://doi.org/10.1016/S1074-5521(96)90069-1

O'Rourke, S., Widdick, D., & Bibb, M. (2017). A novel mechanism of immunity controls the onset of cinnamycin biosynthesis in Streptomyces cinnamoneus DSM 40646. *Journal of Industrial Microbiology and Biotechnology*, *44*(4–5), 563–572. https://doi.org/10.1007/S10295-016-1869-9/FIGURES/10

Overbye, K. M., & Barrett, J. F. (2005). Antibiotics: Where did we go wrong? In *Drug Discovery Today* (Vol. 10, Issue 1, pp. 45–52). Elsevier Current Trends. https://doi.org/10.1016/S1359-6446(04)03285-4

Patankar, A. V., & González, J. E. (2009). Orphan LuxR regulators of quorum sensing. *FEMS Microbiology Reviews*, *33*(4), 739–756. https://doi.org/10.1111/J.1574-6976.2009.00163.X

Perera, I. C., & Grove, A. (2010). Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *Journal of Molecular Cell Biology*, *2*(5), 243–254. https://doi.org/10.1093/JMCB/MJQ021

Poole, K. (2007). Efflux pumps as antimicrobial resistance mechanisms. *Annals of Medicine*, *39*(3), 162–176. https://doi.org/10.1080/07853890701195262

Projan, S. J. (2003). Why is big Pharma getting out of antibacterial drug discovery? In *Current Opinion in Microbiology* (Vol. 6, Issue 5, pp. 427–430). Elsevier Ltd. https://doi.org/10.1016/j.mib.2003.08.003

Projan, S. J., & Shlaes, D. M. (2004). Antibacterial drug discovery: is it all downhill from here? *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *10 Suppl 4*(4), 18–22. https://doi.org/10.1111/J.1465-0691.2004.1006.X

Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H., & Linington, R. G. (2017). Retrospective analysis of natural products provides insights for future discovery trends. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(22), 5601–5606. https://doi.org/10.1073/PNAS.1614680114/SUPPL_FILE/PNAS.1614680114.SD01.CSV

Qin, Z., Devine, R., Hutchings, M. I., & Wilkinson, B. (2019a). A role for antibiotic biosynthesis monooxygenase domain proteins in fidelity control during aromatic polyketide

biosynthesis. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-11538-6

Qin, Z., Devine, R., Hutchings, M. I., & Wilkinson, B. (2019b). Aromatic polyketide biosynthesis: fidelity, evolution and engineering. *BioRxiv*, 581074. https://doi.org/10.1101/581074

Qin, Z., Munnoch, J. T., Devine, R., Holmes, N. A., Seipke, R. F., Wilkinson, K. A., Wilkinson, B., & Hutchings, M. I. (2017). Formicamycins, antibacterial polyketides produced by Streptomyces formicae isolated from African Tetraponera plant-ants. *Chemical Science*, *8*(4), 3218–3227. https://doi.org/10.1039/C6SC04265A

Raghavan, V., & Groisman, E. A. (2010). Orphan and hybrid two-component system proteins in health and disease. In *Current Opinion in Microbiology* (Vol. 13, Issue 2, pp. 226–231). NIH Public Access. https://doi.org/10.1016/j.mib.2009.12.010

Rather, I. A., Kim, B. C., Bajpai, V. K., & Park, Y. H. (2017). Self-medication and antibiotic resistance: Crisis, current challenges, and prevention. In *Saudi Journal of Biological Sciences* (Vol. 24, Issue 4, pp. 808–812). Elsevier B.V. https://doi.org/10.1016/j.sjbs.2017.01.004

Ray, L., & Moore, B. S. (2016). Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Natural Product Reports*, *33*(2), 150–161. https://doi.org/10.1039/C5NP00112A

Reygaert, W. (2009). Methicillin-Resistant Staphylococcus aureus (MRSA): Molecular Aspects of Antimicrobial Resistance and Virulence. *American Society for Clinical Laboratory Science*, *22*(2), 115–119. https://doi.org/10.29074/ASCLS.22.2.115

Reygaert, W. C. (2018). An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiology*, *4*(3), 482. https://doi.org/10.3934/MICROBIOL.2018.3.482

Risdian, C., Mozef, T., & Wink, J. (2019). Biosynthesis of Polyketides in Streptomyces. *Microorganisms*, *7*(5). https://doi.org/10.3390/MICROORGANISMS7050124

Robbins, T., Liu, Y. C., Cane, D. E., & Khosla, C. (2016). Structure and mechanism of assembly line polyketide synthases. *Current Opinion in Structural Biology*, *41*, 10–18. https://doi.org/10.1016/J.SBI.2016.05.009

Rodríguez, H., Rico, S., Díaz, M., & Santamaría, R. I. (2013). Two-component systems in Streptomyces: key regulators of antibiotic complex pathways. *Microbial Cell Factories*, *12*(1), 127. https://doi.org/10.1186/1475-2859-12-127

Rodríguez, H., Rico, S., Yepes, A., Franco-Echevarría, E., Antoraz, S., Santamaría, R. I., & Díaz, M. (2015). The two kinases, AbrC1 and AbrC2, of the atypical two-component system AbrC are needed to regulate antibiotic production and differentiation in Streptomyces coelicolor. *Frontiers in Microbiology*, *6*(MAY), 140985. https://doi.org/10.3389/FMICB.2015.00450/ABSTRACT

Röhl, F., Rabenhorst, J., & Zähner, H. (1987). Biological properties and mode of action of clavams. *Archives of Microbiology*, *147*(4), 315–320. https://doi.org/10.1007/BF00406126/METRICS

Romero-Rodríguez, A., Robledo-Casados, I., & Sánchez, S. (2015). An overview on transcriptional regulators in Streptomyces. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1849*(8), 1017–1039. https://doi.org/10.1016/J.BBAGRM.2015.06.007

Ruan, X., Pereda, A., Stassi, D. L., Zeidner, D., Summers, R. G., Jackson, M., Shivakumar, A., Kakavas, S., Staver, M. J., Donadio, S., & Katz, L. (1997). Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *Journal of Bacteriology*, *179*(20), 6416–6425. https://doi.org/10.1128/JB.179.20.6416-6425.1997

Rutledge, P. J., & Challis, G. L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. In *Nature Reviews Microbiology* (Vol. 13, Issue 8, pp. 509–523). Nature Publishing Group. https://doi.org/10.1038/nrmicro3496

Schlimpert, S., Flärdh, K., & Buttner, M. (2016). Fluorescence time-lapse imaging of the complete S. venezuelae life cycle using a microfluidic device. *Journal of Visualized Experiments*, *2016*(108), e53863. https://doi.org/10.3791/53863

Schumacher, M. A., Zeng, W., Findlay, K. C., Buttner, M. J., Brennan, R. G., & Tschowri, N. (2017). The Streptomyces master regulator BldD binds c-di-GMP sequentially to create a functional BldD 2-(c-di-GMP) 4 complex. *Nucleic Acids Research*, *45*(11), 6923–6933. https://doi.org/10.1093/nar/gkx287

Seipke, R. F., Barke, J., Heavens, D., Yu, D. W., & Hutchings, M. I. (2013). Analysis of the bacterial communities associated with two ant-plant symbioses. *MicrobiologyOpen*, *2*(2), 276–283. https://doi.org/10.1002/mbo3.73

Seipke, R. F., Kaltenpoth, M., & Hutchings, M. I. (2012). *Streptomyces* as symbionts: an emerging and widespread theme? *FEMS Microbiology Reviews*, *36*(4), 862–876. https://doi.org/10.1111/j.1574-6976.2011.00313.x

Selim, S. (2022). Mechanisms of gram-positive vancomycin resistance (Review). *Biomedical Reports*, *16*(1). https://doi.org/10.3892/BR.2021.1490

Sellés Vidal, L., Kelly, C. L., Mordaka, P. M., & Heap, J. T. (2018). Review of NAD(P)H-dependent oxidoreductases: Properties, engineering and application. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1866*(2), 327–347. https://doi.org/10.1016/J.BBAPAP.2017.11.005

Sengupta, S., Chattopadhyay, M. K., & Grossart, H. P. (2013). The multifaceted roles of antibiotics and antibiotic resistance in nature. In *Frontiers in Microbiology* (Vol. 4, Issue MAR, p. 47). Frontiers Research Foundation. https://doi.org/10.3389/fmicb.2013.00047

Shaw, L., Golonka, E., Potempa, J., & Foster, S. J. (2004). The role and regulation of the extracellular proteases of Staphylococcus aureus. *Microbiology*, *150*(1), 217–228. https://doi.org/10.1099/MIC.0.26634-0/CITE/REFWORKS

Sheldon, P. J., Busarow, S. B., & Hutchinson, C. R. (2002). Mapping the DNA-binding domain and target sequences of the Streptomyces peucetius daunorubicin biosynthesis regulatory protein, DnrI. *Molecular Microbiology*, *44*(2), 449–460. https://doi.org/10.1046/J.1365-2958.2002.02886.X

Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*, *7*(2), 285–295. https://doi.org/10.1016/S1367-5931(03)00020-6

Shrethsa L. (2005). Life Expectancy in the United States. *CRS Report for Congress*. https://digital.library.unt.edu/ark:/67531/metadc810624/

Shu, D., Chen, L., Wang, W., Yu, Z., Ren, C., Zhang, W., Yang, S., Lu, Y., & Jiang, W. (2009). afsQ1-Q2-sigQ is a pleiotropic but conditionally required signal transduction system

for both secondary metabolism and morphological development in Streptomyces coelicolor. *Applied Microbiology and Biotechnology*, *81*(6), 1149–1160. https://doi.org/10.1007/S00253-008-1738-1

Sivalingam, P., Hong, K., Pote, J., & Prabakar, K. (2019). Extreme environment streptomyces: Potential sources for new antibacterial and anticancer drug leads? *International Journal of Microbiology*, *2019*. https://doi.org/10.1155/2019/5283948

Sola-Landa, A., Moura, R. S., & Martin, J. F. (2003). The two-component PhoR-PhoP system controls both primary metabolism and secondary metabolite biosynthesis in Streptomyces lividans. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(10), 6133–6138. https://doi.org/10.1073/PNAS.0931429100

Sola-Landa, A., Rodríguez-García, A., Franco-Domínguez, E., & Martín, J. F. (2005). Binding of PhoP to promoters of phosphate-regulated genes in Streptomyces coelicolor: identification of PHO boxes. *Molecular Microbiology*, *56*(5), 1373–1385. https://doi.org/10.1111/J.1365-2958.2005.04631.X

Som, N. F., Heine, D., Holmes, N. A., Munnoch, J. T., Chandra, G., Seipke, R. F., Hoskisson, P. A., Wilkinson, B., & Hutchings, M. I. (2017). The Conserved Actinobacterial Two-Component System MtrAB Coordinates Chloramphenicol Production with Sporulation in Streptomyces venezuelae NRRL B-65442. *Frontiers in Microbiology*, *8*(JUN), 1145. https://doi.org/10.3389/fmicb.2017.01145

Som, N. F., Heine, D., Holmes, N., Knowles, F., Chandra, G., Seipke, R. F., Hoskisson, P. A., Wilkinson, B., & Hutchings, M. I. (2017). The MtrAB two-component system controls antibiotic production in Streptomyces coelicolor A3(2). *Microbiology*, *163*(10), 1415. https://doi.org/10.1099/MIC.0.000524

Strauch, E., Takano, E., Baylts, H. A., & Bibb, M. J. (1991). The stringent response in Streptomyces coelicolor A3(2). *Molecular Microbiology*, *5*(2), 289–298. https://doi.org/10.1111/J.1365-2958.1991.TB02109.X

Sultan, I., Rahman, S., Jan, A. T., Siddiqui, M. T., Mondal, A. H., & Haq, Q. M. R. (2018). Antibiotics, resistome and resistance mechanisms: A bacterial perspective. In *Frontiers in Microbiology* (Vol. 9, Issue SEP, p. 2066). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2018.02066

Süssmuth, R. D., & Mainz, A. (2017). Nonribosomal Peptide Synthesis—Principles and Prospects. *Angewandte Chemie International Edition*, *56*(14), 3770–3821. https://doi.org/10.1002/ANIE.201609079

Sykes, R. (2010). The 2009 Garrod lecture: the evolution of antimicrobial resistance: a Darwinian perspective. *The Journal of Antimicrobial Chemotherapy*, *65*(9), 1842–1852. https://doi.org/10.1093/jac/dkq217

Tang, G. L., Zhang, Z., & Pan, H. X. (2017). New insights into bacterial type II polyketide biosynthesis. *F1000Research*, *6*, 172. https://doi.org/10.12688/F1000RESEARCH.10466.1

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., & Hamon, C. (2003). Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry*, *75*(8), 1895–1904. https://doi.org/10.1021/AC0262560/ASSET/IMAGES/LARGE/AC0262560F00006.JPEG

Tschowri, N. (2016). Cyclic dinucleotide-controlled regulatory pathways in Streptomyces species. In *Journal of Bacteriology* (Vol. 198, Issue 1, pp. 47–54). American Society for Microbiology. https://doi.org/10.1128/JB.00423-15

Tschowri, N., Schumacher, M. A., Schlimpert, S., Chinnam, N. B., Findlay, K. C., Brennan, R. G., & Buttner, M. J. (2014). Tetrameric c-di-GMP mediates effective transcription factor dimerization to control streptomyces development. *Cell*, *158*(5), 1136–1147. https://doi.org/10.1016/j.cell.2014.07.022

Tyurin, A. P., Alferova, V. A., Paramonov, A. S., Shuvalov, M. V., Kudryakova, G. K., Rogozhin, E. A., Zherebker, A. Y., Brylev, V. A., Chistov, A. A., Baranova, A. A., Biryukov, M. V., Ivanov, I. A., Prokhorenko, I. A., Grammatikova, N. E., Kravchenko, T. V., Isakova, E. B., Mirchink, E. P., Gladkikh, E. G., Svirshchevskaya, E. V., … Korshun, V. A. (2021). Gausemycins A,B: Cyclic Lipoglycopeptides from Streptomyces sp.*. *Angewandte Chemie (International Ed. in English)*, *60*(34), 18694–18703. https://doi.org/10.1002/ANIE.202104528

Van Wezel, G. P., & McDowall, K. J. (2011). The regulation of the secondary metabolism of Streptomyces: New links and experimental advances. In *Natural Product Reports* (Vol.

28, Issue 7, pp. 1311–1333). The Royal Society of Chemistry. https://doi.org/10.1039/c1np00003a

Venkatesh, G. R., Koungni, F. C. K., Paukner, A., Stratmann, T., Blissenbach, B., & Schnetz, K. (2010). BglJ-RcsB heterodimers relieve repression of the Escherichia coli bgl operon by H-NS. *Journal of Bacteriology*, *192*(24), 6456–6464. https://doi.org/10.1128/JB.00807-10/SUPPL_FILE/2010_BGLJ_RCSB_SUPPLEMENT_REV.PDF

Ventola, C. L. (2015). The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *Pharmacy and Therapeutics*, *40*(4), 277. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4378521/

Waksman, S. A., & Woodruff, H. B. (1940). Bacteriostatic and Bactericidal Substances Produced by a Soil Actinomyces. *Experimental Biology and Medicine*, *45*(2), 609–614. https://doi.org/10.3181/00379727-45-11768

Walsh, C. T., O'Brien, R. V., & Khosla, C. (2013). Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. *Angewandte Chemie (International Ed. in English)*, *52*(28), 7098. https://doi.org/10.1002/ANIE.201208344

Wang, W., Guo, Q., Xu, X., Sheng, Z. K., Ye, X., & Wang, M. (2014). High-level tetracycline resistance mediated by efflux pumps Tet(A) and Tet(A)-1 with two start codons. *Journal of Medical Microbiology*, *63*(Pt 11), 1454–1459. https://doi.org/10.1099/JMM.0.078063-0

Ward, A., & Allenby, N. (2018). Genome mining for the search and discovery of bioactive compounds: The Streptomyces paradigm. *FEMS Microbiology Letters*, *365*(24). https://doi.org/10.1093/femsle/fny240

Watve, M., Tickoo, R., Jog, M., & Bhole, B. (2001). How many antibiotics are produced by the genus Streptomyces ? *Archives of Microbiology*, *176*(5), 386–390. https://doi.org/10.1007/s002030100345

Weissman, K. J., & Leadlay, P. F. (2005). Combinatorial biosynthesis of reduced polyketides. *Nature Reviews. Microbiology*, *3*(12), 925–936. https://doi.org/10.1038/NRMICRO1287

Welsch, M. E., Snyder, S. A., & Stockwell, B. R. (2010). Privileged scaffolds for library design and drug discovery. *Current Opinion in Chemical Biology*, *14*(3), 347–361. https://doi.org/10.1016/J.CBPA.2010.02.018

Wenski, S. L., Thiengmag, S., & Helfrich, E. J. N. (2022). Complex peptide natural products: Biosynthetic principles, challenges and opportunities for pathway engineering. *Synthetic and Systems Biotechnology*, *7*(1), 631–647. https://doi.org/10.1016/J.SYNBIO.2022.01.007

WHO. (2019). 2019 antibacterial agents in clinical development: an analysis of the antibacterial clinical development pipeline. *World Health Organization*.

Wilson, D. J., Xue, Y., Reynolds, K. A., & Sherman, D. H. (2001). Characterization and analysis of the pikD regulatory factor in the pikromycin biosynthetic pathway of Streptomyces venezuelae. *Journal of Bacteriology*, *183*(11), 3468–3475. https://doi.org/10.1128/JB.183.11.3468-3475.2001/ASSET/6C804AA4-80B7-4DC2-8F2C-2B6007D335FA/ASSETS/GRAPHIC/JB1110029003.JPEG

Wu, H., Liu, W., Dong, D., Li, J., Zhang, D., & Lu, C. (2014). SlnM gene overexpression with different promoters on natamycin production in Streptomyces lydicus A02. *Journal of Industrial Microbiology and Biotechnology*, *41*(1), 163–172. https://doi.org/10.1007/S10295-013-1370-7

Wyszynski, F. J., Lee, S. S., Yabe, T., Wang, H., Gomez-Escribano, J. P., Bibb, M. J., Lee, S. J., Davies, G. J., & Davis, B. G. (2012). Biosynthesis of the tunicamycin antibiotics proceeds via unique exo-glycal intermediates. *Nature Chemistry 2012 4:7*, *4*(7), 539–546. https://doi.org/10.1038/nchem.1351

Xie, M., Wu, M., & Han, A. (2020). Structural insights into the signal transduction mechanism of the K+-sensing two-component system KdpDE. *Science Signaling*, *13*(643), 2970. https://doi.org/10.1126/SCISIGNAL.AAZ2970/SUPPL_FILE/AAZ2970_SM.PDF

Xu, G., Wang, J., Wang, L., Tian, X., Yang, H., Fan, K., Yang, K., & Tan, H. (2010). "Pseudo" γ-butyrolactone receptors respond to antibiotic signals to coordinate antibiotic biosynthesis. *Journal of Biological Chemistry*, *285*(35), 27440–27448. https://doi.org/10.1074/jbc.M110.143081

Xu, Y., Willems, A., Au-Yeung, C., Tahlan, K., & Nodwell, J. R. (2012). A two-step mechanism for the activation of actinorhodin export and resistance in Streptomyces coelicolor.

MBio, *3*(5). https://doi.org/10.1128/MBIO.00191-12/SUPPL_FILE/MBO005121346S1.DOCX

Xue, T., You, Y., Hong, D., Sun, H., & Sun, B. (2011). The Staphylococcus aureus KdpDE two-component system couples extracellular K+ sensing and Agr signaling to infection programming. *Infection and Immunity*, *79*(6), 2154–2167. https://doi.org/10.1128/IAI.01180-10

Yagüe, P., Rodríguez-García, A., López-García, M. T., Rioseras, B., Martín, J. F., Sánchez, J., & Manteca, A. (2014). Transcriptomic Analysis of Liquid Non-Sporulating Streptomyces coelicolor Cultures Demonstrates the Existence of a Complex Differentiation Comparable to That Occurring in Solid Sporulating Cultures. *PLOS ONE*, *9*(1), e86296. https://doi.org/10.1371/JOURNAL.PONE.0086296

Yepes, A., Rico, S., Rodríguez-García, A., Santamaría, R. I., & Díaz, M. (2011). Novel Two-Component Systems Implied in Antibiotic Production in Streptomyces coelicolor. *PLOS ONE*, *6*(5), e19980. https://doi.org/10.1371/JOURNAL.PONE.0019980

Yu, D., Xu, F., Zeng, J., & Zhan, J. (2012). Type III polyketide synthases in natural product biosynthesis. *IUBMB Life*, *64*(4), 285–295. https://doi.org/10.1002/IUB.1005

Yu, Z., Zhu, H., Dang, F., Zhang, W., Qin, Z., Yang, S., Tan, H., Lu, Y., & Jiang, W. (2012). Differential regulation of antibiotic biosynthesis by DraR-K, a novel two-component system in Streptomyces coelicolor. *Molecular Microbiology*, *85*(3), 535–556. https://doi.org/10.1111/J.1365-2958.2012.08126.X

Zapf, J., Sen, U., Madhusudan, Hoch, J. A., & Varughese, K. I. (2000). A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. *Structure (London, England : 1993)*, *8*(8), 851–862. https://doi.org/10.1016/S0969-2126(00)00174-X

Zapun, A., Contreras-Martel, C., & Vernet, T. (2008). Penicillin-binding proteins and β-lactam resistance. *FEMS Microbiology Reviews*, *32*(2), 361–385. https://doi.org/10.1111/J.1574-6976.2007.00095.X

Zhang, C., Zhao, W., Duvall, S. W., Kowallis, K. A., & Childers, W. S. (2022). Regulation of the activity of the bacterial histidine kinase PleC by the scaffolding protein PodJ. *The Journal of Biological Chemistry*, *298*(4). https://doi.org/10.1016/J.JBC.2022.101683

Zhang, Q., Chen, Q., Zhuang, S., Chen, Z., Wen, Y., & Li, J. (2015). A MarR Family Transcriptional Regulator, DptR3, Activates Daptomycin Biosynthesis and Morphological Differentiation in Streptomyces roseosporus. *Applied and Environmental Microbiology*, *81*(11), 3753–3765. https://doi.org/10.1128/AEM.00057-15

Zhang, Wu, L., Zhu, Y., Liu, M., Wang, Y., Cao, G., Chen, X. L., Tao, M., & Pang, X. (2017). Deletion of mtrA inhibits cellular development of Streptomyces coelicolor and alters expression of developmental regulatory genes. *Frontiers in Microbiology*, *8*(OCT). https://doi.org/10.3389/FMICB.2017.02013/FULL

Zhao, Q., Wang, L., & Luo, Y. (2019). Recent advances in natural products exploitation in Streptomyces via synthetic biology. *Engineering in Life Sciences*, *19*(6), 452. https://doi.org/10.1002/ELSC.201800137

Zhu, Y., Zhang, P., Zhang, J., Wang, J., Lu, Y., & Pang, X. (2020). Impact on Multiple Antibiotic Pathways Reveals MtrA as a Master Regulator of Antibiotic Production in Streptomyces spp. and Potentially in Other Actinobacteria. *Applied and Environmental Microbiology*, *86*(20). https://doi.org/10.1128/AEM.01201-20

Zschiedrich, C. P., Keidel, V., & Szurmant, H. (2016). Molecular Mechanisms of Two-Component Signal Transduction. In *Journal of Molecular Biology* (Vol. 428, Issue 19, pp. 3752–3775). Academic Press. https://doi.org/10.1016/j.jmb.2016.08.003