Strathclyde Institute of Pharmacy and Biomedical Sciences

Investigating *Streptomyces clavuligerus* Linear Replicons for Improved Clavulanic Acid Production

Elmira Mohit

Thesis presented in fulfilment of the requirement for the degree of Doctor of Philosophy

2023

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date: 30th September 2023

Acknowledgements

Firstly, I would like to thank Dr Paul Herron who has provided me with his continuous support and encouragement over the last four years. Paul, you have greatly enhanced my scientific capabilities and provided invaluable guidance in my development to becoming an independent researcher. I could not have asked for a better mentor, and I am grateful for the friendship we have forged.

I would like to acknowledge and thank my funders IBioIC and BBSRC. I would also like to thank GSK for making this PhD possible, in particular Dr Ben Huckle and Dr Steve Kendrew. Steve, your interest, and advice in my work and in my career has been extremely helpful and your perspective on this project always provided me with a lot of inspiration.

I would like to thank Prof Gilles van Wezel and Dr Le Zhang for providing us with the plasmids necessary to complete this study.

Thank you to current and former members of the Herron group: Jana, Tiago, Gillian, Lis and David for all your help over the last year. Lis, thank you for all your knowledge and guidance when I started and for providing me with the framework for my thesis. I would also like extend my thanks to the rest of SIPBS level 6: the loveliest, funniest, and kindest group of co-workers. I am grateful to have shared an office and a lab with you all. In particular, I would like to thank Eilidh, Adriana, Molly and Robyn for their friendship and for providing me with a glass of wine and a good laugh when I really needed it, I really appreciate you.

I would like to sincerely thank my family and friends, especially my mother Zohreh, who always supported and believed in me. Thank you to my chosen family, Jodie and Sarah, I do not know what I would have done without you both in the last years.

Lastly, I would like to thank my partner, Charlie. Charlie, I feel so fortunate to have met you, and your endless support and encouragement has made this thesis possible. Thanks for being the most wonderful person.

Table of Contents

Abstract

Increasing antimicrobial resistance against ß-lactam antibiotics through bacterially produced ß-lactamases has prompted research into effective enzyme inhibitors, such as clavulanic acid. *Streptomyces clavuligerus* is the primary producer of clavulanic acid, which is used in various drugs such as Augmentin®. *Streptomyces clavuligerus* contains a dynamic genome and is composed of four giant linear plasmids (GLPs), pSCL1, pSCL2, pSCL3 and pSCL4 and its chromosome. Various genes essential for the maintenance of linear replicons, such as *tap* and *tpg* which encode telomeric terminal proteins (TPs), are found on three out of four GLPs (pSCL2, pSCL3, pSCL4). Previous work demonstrated a circularised chromosome and loss of plasmid after Cas9 mediated cleavage of the largest GLP, pSCL4, potentially due to the absence of *tap-tpg*. To determine the role of *tap-tpg* in chromosomal and plasmid linearity, we optimised and tested their inactivation using CRISPR-dCas9 multiplexing, targeting *tap-tpg*$_4$ on pSCL4, *tap-tpg*$_3$ on pSCL3 and *tap-tpg*$_2$ on pSCL2. We used Illumina short-read sequencing to genotypically analyse our strains, which highlighted that the knockdown of multiple *tap-tpg* genes resulted in mutant strains with various combinations of lost plasmids and terminal ends. Noticeably, only the mutants with silenced *tap-tpg*$_4$ demonstrated the loss of multiple plasmids, and pSCL2 loss was suggested to be a direct consequence of the CRISPR-dCas9 system, as all mutants and vector controls lost the plasmid. Additionally, we demonstrated chromosome circularisation and the loss of the 13-nucleotide binding site of Tap for our *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ silenced mutant EM10, suggesting that overall *tap-tpg* expression levels in *S. clavuligerus* affect end-patching. The mutant strains were phenotypically characterised and mutants which had lost pSCL3, pSCL2 and/or pSCL1 had a significantly higher specific growth rate, therefore we confirmed that these replicons were essential for strain fitness. We also confirmed that *S. clavuligerus* does not require pSCL3, pSCL2 and pSCL1 for clavulanic acid production, as strains that had lost more replicons did not produce more clavulanic acid. Therefore, biotechnologically, the loss of replicons is of importance in terms of strain fitness rather than clavulanic acid production in *S. clavuligerus.* Moreover, to study the process of end-patching and determine the binding activity of TPs, overexpression of Tap4, Tap3 and Tap2 proteins was optimised and Tap4 was shown to potentially bind to the chromosomal ssDNA telomeres. With this study we have elucidated the importance of *tap-tpg* in the plasmid maintenance of *S. clavuligerus* and confirmed the essentiality of

replicons, highlighting that whilst most replicons were dispensable, they are necessary for maintaining strain fitness and optimum growth.

1.0 Introduction

Since their discovery in the 1940s, the use of antibiotics to combat infectious diseases has increased by 65% (WHO, 2018). Despite progress in the research, development and processing of novel antibiotics, no new antibiotics have been clinically approved in the last two decades (Ventola, 2015). Increasing usage of antibiotics has led to antimicrobial resistance in both developing and developed areas of the world, resulting in the rise of bacterial 'superbugs', like *Pseudomonas aeruginosa* and methicillin-resistant *Staphylococcus aureus* (MRSA) (de Lima Procópio *et al.*, 2012).

The antibiotic mechanism of action targets vital bacterial cellular functions in order to kill or prevent proliferation of that organism. Cellular functions vital to growth include DNA replication, RNA synthesis, protein synthesis, cell wall synthesis, membrane integrity and metabolic pathway inhibition (de Lima Procópio *et al.,* 2012). Reasons for bacterial resistance include genetic changes in the bacterial genome that encode functions to prevent the inhibition of these cellular functions. These changes may involve attaining resistance through horizontal gene transfer or mutations (Aminov, 2009) that hinder antibiotic action through changes in the site of drug interaction (Andersson and Hughes, 2010). Interestingly, many bacterial resistance genes originate in the biosynthetic gene cluster responsible for the production of that antibiotic and act as a natural defence mechanism. This is true for all *Streptomyces* species, which produce a large variety of bioactive secondary metabolites (Kinashi, 2011).

1.1 Resistance to ß-lactam Antibiotics

Antibiotic resistance by pathogenic microbes has been a recognised problem for the last 50 years, and even more so in the past decade due to the reduction in efficacy of clinically approved antibiotics since the early 2000s (**Figure 1.1.**). Generally, antibiotics can be categorised by their mechanisms of action. One of the most widely used group of narrow-spectrum antibiotics are penicillins, first found to be produced by *Penicillum chrysogenum* (Behrens and Corse, 1948). Penicillins, alongside carbapenems, monobactams and cephalosporins, form one of the largest antibiotic classes: the ß-lactams (Fisher, Meroueh and Mobashery, 2005).

ß-lactams target the peptidoglycan layer of bacterial cell walls, which consist of N-acetyl-glucosamine (NAG) and N-acetyl-muramic acid (NAM), and therefore destroy structural integrity by breaking down pentapeptide bridges between its component NAG and NAM chains (Long *et al.,* 2005). NAG and NAM cross linkages are catalysed by

transpeptidases or penicillin binding proteins (PBP) (Zapun, Vernet and Pinho, 2008). ß-lactam antibiotics mimic the specific transpeptidase substrates and thus inhibit further transpeptidation reactions, breaking the cross-linked bacterial cell wall of Gram-positive organisms (Hamed *et al.*, 2013). Ultimately, transpeptidase inhibition causes autolysis.

Penicillins are made up of a bicyclic structure forming 6-aminopenicillanic acid (6-APA). Their dipeptide structure is formed by L-cysteine and D-valine condensation, which produces the thiazolidinic and ß-lactam rings (Fernandes, Amador and Prudêncio, 2013). The narrow spectrum of activity of penicillin lead researchers towards semi-synthetic and synthetic forms of the antibiotic. In the 1970s Weissenburger and van der Hoeven developed a chemical splitting technology, allowing for the extraction of the 6-APA nucleus of penicillin, allowing for its modification into a range of semi-synthetic forms such as ampicillin, flucloxacillin, methicillin and amoxicillin (**Figure 1.1.**) (Weissenburger and Hoeven, 1970).

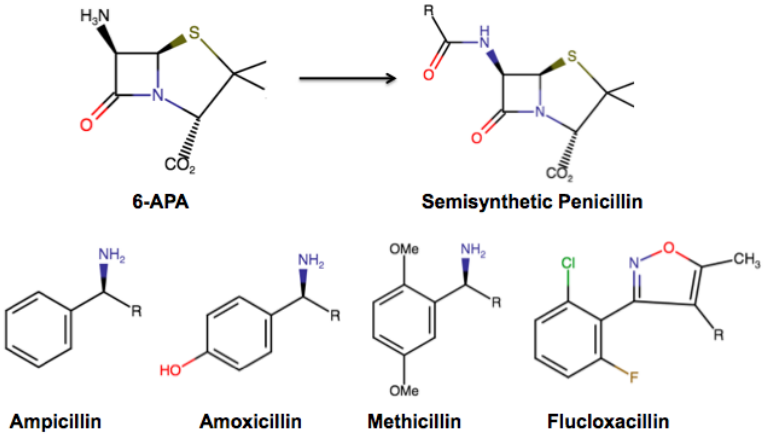


Figure 1.1. Semi-synthetic penicillin formed by the addition of R groups to the 6-APA nucleus. Adapted from Hamed *et al.,* (2013)

One of the problems with synthetic and semi-synthetic ß-lactam antibiotics, such as amoxicillin, lies in commercial production, due to the reactive bicyclic core ring structure. Chemical degradation is a result of problematic production and is one of many issues that ß-lactam antibiotics faced from their first clinical usage in the 1940s (Fernandes *et al.,* 2013).

Resistance against this class of antibiotics is prevalent and is mediated through bacterial defence mechanisms such as efflux pumps to eject ß-lactams (Pages *et al.*, 2009), PBP upregulation and PBP mutations which block the binding of the inhibitor, to the essential enzymes (Kliebe *et al.*, 1985). An example of PBP mutations can be seen in *Streptococcus pneumoniae*, where resistance to penicillin occurs by PBP enzyme alterations (Severin, Figueiredo and Tomasz, 1996). On the other hand, studies by Nakagawa *et al.*, (1984) and Suzuki *et al.*, (1980), highlight that PBPs are vital to the integral function of Gram-positive bacteria, therefore penicillin resistant pneumococci may synthesize peptidoglycan with altered biochemical structures, which could be another reason for resistance (Hakenbeck, 1999). The synthesis of an altered peptidoglycan layer was seen in *Escherichia coli* class A and 1B PBPs (Nakagawa *et al.,* 1984; Suzuki *et al.,* 1980).

Another mechanism for resistance against ß-lactam antibiotics is the production of bacterial ß-lactamases which hydrolyse the ß-lactam ring structure. This was demonstrated by Aggarwal, Chaudhary and Bala, (2008), who investigated the extended-spectrum ß-lactamases (ESBL) present in *P. aeruginosa* from a range of clinical samples. A total of 30 of 148 clinical samples infected with *P. aeruginosa* were ESBL positive and also multi-drug resistant (Aggarwal, Chaudhary and Bala, 2008). Characteristically, ß-lactamases can be divided into different classes based on their amino acid sequences and structural mechanisms, highlighted by the Ambler Molecular Class shown in **Table 1.1.** (Ambler, 1980). Group B ß-lactamases use Zn^{2+} cations to allow for inhibition and are sometimes called metallo-ß-lactamases, whereas ß-lactamases belonging to groups A, C and D need serine residues in their active sites to enable inhibition (Ambler, 1980; Schultsz and Geerlings, 2012).

Table 1.1. Ambler classification of ß-lactamases targeting various ß-lactam antibiotics.

AMBLER MOLECULAR CLASS	*TYPE*	*CHARACTERISTICS/FUNCTION*	*ENZYME EXAMPLES*
A	Narrow-spectrum ß-lactamases	Penicillin hydrolysis	TEM-1, TEM-2, SHV-1
A	Extended-spectrum ß-lactamases	Narrow and extended-spectrum ß-lactam hydrolysis	SHV-2, CTX-M-15, PER-1
A	Serine carbapenemases	Carbapenem hydrolysis	KPC-1, IMI-1, SME-1
B	Metallo-ß-lactamases	Carbapenem hydrolysis	VIM-1, IMP-1
C	Cephalosporinases	Cephamycin and oxymino ß-lactam hydrolysis; mediated chromosomally	AmpC, ACT-1, CMY-2, P99
D	OXA-type enzymes	Oxacilin, oxymino ß-lactam and carbapenem hydrolysis	OXA Enzymes

*Adapted from Toussaint and Gallagher, (2015).

Amoxicillin and ampicillin are particularly ß-lactamase sensitive, due excessive usage against a range of Gram-positive and negative organisms. Their sensitivity is down to the replacement of the phenylacetic acid structural component with phenylglycine (Fernandes *et al.,* 2013). ESBL producing bacteria are often also resistant to different classes of antibiotics such as fluoroquinolones or aminoglycosides, highlighting the necessity of finding novel treatments (Aggarwal, Chaudhary and Bala, 2008).

Genes encoding ß-lactamases in bacterial genomes lead to ß-lactam resistance though a range of different enzymatic activities (TEM-1, SHV-2) shown in **Table 1.1** Two types of chromosomal encoded ß-lactamases exist: constitutive and inducible. The latter requires ß-lactam agents to initiate their expression and exhibit resistance as ß-lactamase expression is not found in the absence of ß-lactam agents (Kong, Schneper and Mathee, 2010). The first cloned and sequence resistance gene, *ampC*, was found in *E. coli* K-12 in the 1980s (Jaurin and Grundström, 1981). It was observed that *ampC* expression in *E. coli* was directly proportional to nutrient availability and growth rate.

Nonetheless, examples of inducible *ampC* can be seen for *Enterobacter cloacae*, where cephalopsorinase expression is controlled via the activation of this gene (Hennessey, 1967; Gootz and Sanders, 1983). Unlike the *ampC* in *E. coli*, *E. cloacae* differentially expresses *ampC* in the absence and presence of ß-lactam agents, suggesting that there may be differences in regulation of these genes between different species (Gootz and Sanders, 1983).

Despite differences in regulation, the transfer of antibiotic genes between species is not uncommon. Many of the genes that confer resistance are found on bacterial plasmids. Plasmid-mediated resistance allows bacteria to become multi-drug resistant by being extremely mobile (Schultsz and Geerlings, 2012). The recombination of resistance genes within the same plasmid or horizontal gene transfer between different plasmids and the chromosome can thus confer resistance genes to species that were previously susceptible to ß-lactams. Susceptibility to ß-lactamases resulting in antimicrobial resistance and decreasing drug effectiveness against pathogenic microbes are the main reasons why ß-lactamase inhibitors, such as clavulanic acid, are therefore often combined with the above mentioned semi-synthetic penicillins.

1.2 Clavulanic Acid as a ß-lactamase inhibitor

Clavulanic acid was first discovered when produced in a *Streptomyces clavuligerus* fermentation broth soon after its initial isolation and classification by Higgens and Kástner, (1971) and belongs to the ß-lactam clavam group (Nagarajan *et al.*, 1971). Clavulanic acid is an inhibitor of ß-lactamases produced by a series of pathogenic bacteria such as *Staphylococcus*, and ß-lactamases produced through plasmid mediated processes in bacteria such as *E. coli* and *Klebsiella* (Boon and Beale, 1987; Beytur *et al.*, 2014). On its own, clavulanic acid possesses limited antibacterial activity and is therefore used in combination with other ß-lactam antibiotics such as amoxicillin. The inhibitory properties of clavulanic acid make it an unusual, non-classical ß-lactam compound, as it has previously been shown to inhibit a variety of ß-lactamases, such as narrow-spectrum ß-lactamases, ESBLs and even cephalosporinases (Shahid *et al.*, 2009).

The wide range of inhibition of clavulanic acid was also witnessed by Hugonnet and Blanchard, (2007), who identified a ß-lactamase critical resistance factor in *Mycobacterium tuberculosis.* The chromosomally encoded BlaC, an Ambler class A ß-lactamase, resulted in extremely high *M. tuberculosis* cephalosporinase and

penicillinase activity. Enzyme inhibition, which was tested with tazobactam, sulbactam and clavulanic acid, was most stable and quick with clavulanic acid treatment, compared to the other two (Hugonnet and Blanchard, 2007). The study underlines the main aspects necessary for effective ß-lactamase inhibition: stability and fast inhibition and emphasises the potential use of clavulanic acid against multiple drug resistant and extremely drug resistant organisms. Thus, clinical therapy strategies have involved utilising clavulanic acid in combination treatments and are particularly successful against ESBLs when regular antibiotic treatments are ineffective (Aggarwal, Chaudhary and Bala, 2008). An example of a clavulanic acid and amoxicillin combination drug is called Augmentin® (GlaxoSmithKline), a commercially produced antibiotic launched in the UK in 1981 (Geddes, Klugman and Rolinson, 2007).

In the clinic, clavulanic acid combination treatments are used to treat a range of infections such as pneumonia, urinary tract infections, cellulitis and often prescribed pre and post-surgery to prevent surgical site infections (Huttner *et al.*, 2020). The use of amoxicillin in combination with clavulanic acid depends on the severity of disease and identification of the organism of infection. In cases where ß-lactam resistant organisms, such as *Staphylococcus aureus*, have been identified co-amoxiclav combination treatments (amoxicillin and clavulanic acid) have shown higher effectiveness (Leflon-Guibout *et al.*, 2000; Huttner *et al.*, 2020). However, even in the past, inhibitor resistances have been reported against clavulanic acid, particularly TEM ß-lactamases but also in some ESBLs, as shown in research by Lemozy *et al.*, (1995). Despite combination treatments, resistances against clavulanic acid and other forms of ß-lactam antibiotics still persist, many of which are caused by mutations in the bacterial genome leading to amino acid residue replacement in ß-lactamases, resulting in a resistant phenotype (Leflon-Guibout *et al.*, 2000).

1.3 Clavulanic Acid Background and Biosynthesis

To fully understand reasons behind clavulanic acid inhibitor resistance and to comprehend what factors affect the production and function of the compound, the biosynthesis and genetic makeup of *S. clavuligerus*, needs to be investigated. Overall, the most extensive research regarding clavulanic acid biosynthesis and associated gene clusters has been performed for *S. clavuligerus* (Jensen and Paradkar, 1999). Clavulanic acid has a characteristic ß-lactam and oxazolidine ring and its inhibitory properties were

explained by investigating its particular nucleus 3R, 5R stereochemistry (Brown, 1986). The molecular nucleus of clavulanic acid is of a similar to classic penicillin, the only difference being oxygen in the oxazolidine ring rather than sulphur in penicillin's thiazolidine ring and modifications to its C-6 and C-2 groups. These modifications allow the compound to avoid cleavage by ß-lactamases and irreversibly inhibit the enzymes (Baggaley, Brown and Schofield, 1997). Other reported clavulanic acid producers include *Streptomyces katsurahamanus* and *Streptomyces jumonjinensis*, who also produce cephamycin C, although information regarding their genome and biosynthesis has only been released recently (AbuSara *et al.*, 2019).

In addition to clavulanic acid, *S. clavuligerus* produces more than 20 metabolites including other ß-lactam antibiotics such as cephamycin C and penicillin N (Nabais and da Fonseca, (1995); Baggaley *et al.,* 1997). Over eight structural genes are associated with clavulanic acid biosynthesis, all of which are part of its biosynthetic gene cluster (BGC), located adjacent to the *S. clavuligerus'* cephamycin BCG (Jensen and Paradkar, 1999).

Regulation of *S. clavuligerus* antibiotic BGCs is structured in a hierarchical way consisting of various levels ranging from pathway-specific transcriptional regulatory genes to global regulatory genes. The bottom of the hierarchy consists of transcriptional activators regulating biosynthetic genes in the 'late' stages **(Figure 1.2.)**. One of these is *claR*, leading to clavulanic acid production from clavaminic acid. The preceeding level consists of genes, such as *ccaR*, encoding antibiotic protein families and is located next to other antibiotic gene clusters for example the BGC for cephamycin C (Alexander and Jensen, 1998).

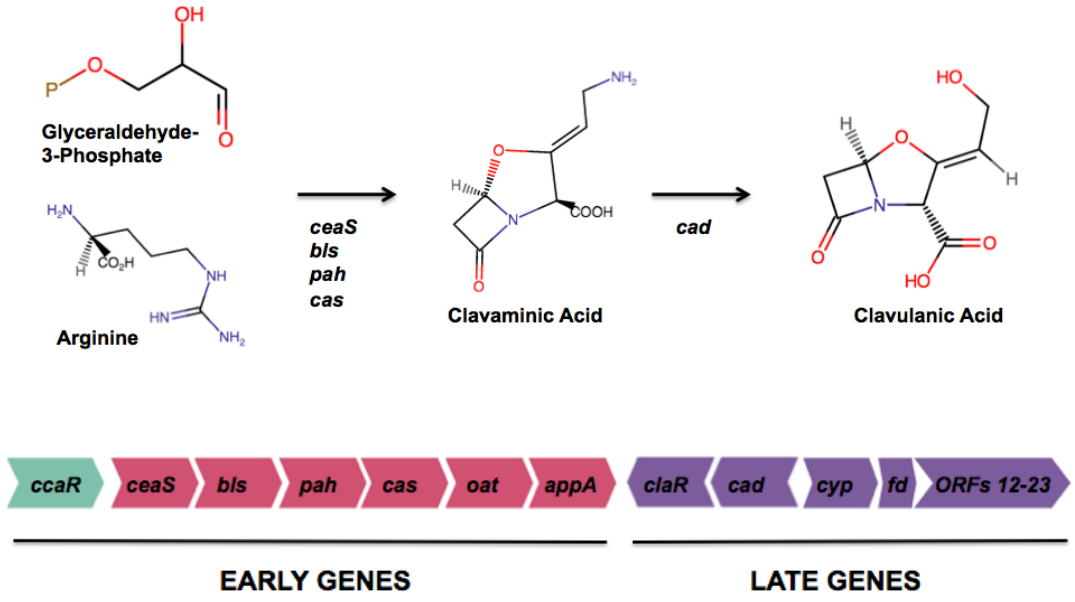


Figure 1. 2. Clavulanic acid chemical structure and biosynthesis. Summary of main intermediates and genes involved in the biosynthesis of clavulanic acid by *S. clavuligerus* from glyceraldehyde-3-phosphatate and arginine. Key genes involved in the pathway include carboxyethylarginine (*ceaS*), ß-lactam synthase (*bls*), proclavaminic acid aminohydrolase (*pah*), clavaminic acid synthase (*cas*) and clavulanic acid dehydrogenase (*cad*). Adapted from Algora-Gallardo, (unpublished data).

The clavulanic acid precursors include arginine and D-glyceraldehyde-3-phosphate, and initial reactions include condensation, hydroxylation, hydrolysis, and cyclization reactions to form clavaminic acid (**Figure 3.).** Two sets of genes, referred to as 'early genes', encode clavaminic acid formation in *S. clavuligerus*. These early genes are located in three parts of the chromosome (Paradkar, 2013). The clavaminic acid encoded by the early genes is also referred to as the intermediate, as the production of clavaminic acid can either result in 3S and 5S clavams, or clavulanic acid biosynthesis. The production of clavulanic acid is encoded for by the 'late genes'. The 'late genes' in the *S. clavuligerus* BGC responsible for clavulanic acid production include essential genes such as *cad* and *gcaS*. Mutations in these genes have diminished *S. clavuligerus'* ability to produce clavulanic acid, highlighting their essentiality to clavulanic acid biosynthesis (Jensen *et al.*, 2000). Industrially, clavulanic acid biosynthesis through *S. clavuligerus* bacterial fermentations depends on fermenter stability, strain sensitivity and productivity. Upstream processing becomes increasingly important when looking at the challenges in maximising downstream clavulanic acid production. Therefore, understanding the genetic components and growth mechanisms leading to *Streptomyces* metabolite production, particularly *S. clavuligerus*, allows for more targeted upstream genetic manipulation to improve productivity of the strains.

1.4 The Importance of *Streptomyces* Secondary Metabolism

1.4.1 *Streptomyces* Survival and Growth

The discovery of clavulanic acid production via *S. clavuligerus* highlights the versatility and usefulness of *Streptomyces* secondary metabolism. *S. clavuligerus* belongs to the genus *Streptomyces*, which are one of around 120 genera belonging to the *Actinobacteria*, a diversely distributed group consisting of Gram-positive bacteria. First discovered in the 1870's, Actinomycetes are known for their G+C rich DNA, normally over 70% for most *Streptomyces* species (Wright and Bibb, 1992).

Streptomyces are morphologically different to most Gram-positive bacteria, as their growth and germination involves the production of hyphae that form a mycelium. The mycelium differentiates into aerial hyphae that subsequently segregate into spore chains. The unique ability of *Streptomyces* to sporulate under favourable temperature and nutrient conditions is associated with a coordinated and specific metabolism which can be split into primary (essential) and secondary (non-essential) denominations, and makes them extremely versatile bacteria (Niu *et al.*, 2016).

Streptomyces are distributed in a range of natural habitats, such as lakes and soils, or in symbiotic relationships with other species e.g. plants or fungi (Seipke, Kaltenpoth and Hutchings, 2012; Bakker *et al.*, 2014). To survive these diverse environments, it is thought that the organisms produce secondary metabolites. Most importantly, their secondary metabolism allows for the production of bioactive metabolites, such as antiviral and antifungal agents as well as antibiotics (Hopwood, 2006; Khan *et al.*, 2011) active against competing and parasitic microorganisms. The first antibiotic to be discovered produced by *Streptomyces* was streptomycin, produced by *Streptomyces griseus* (*S. griseus*) (de Lima Procópio *et al.,* 2012). Today, around 80% of antibiotics are produced through secondary metabolism from the genus *Streptomyces*, as highlighted in **Figure 1.3.**

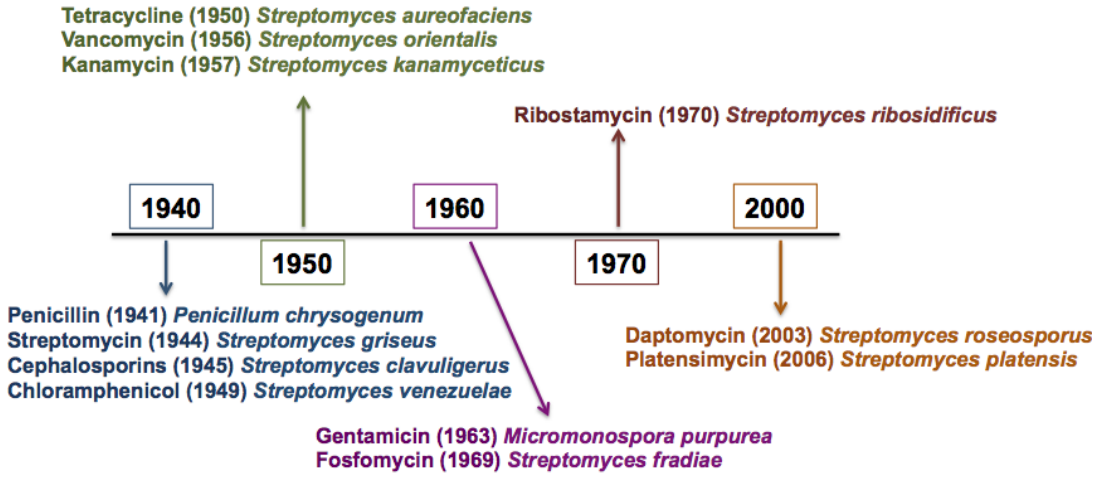


Figure 1.3. History of antibiotic discovery highlighting key findings related to *Streptomyces* species. Adapted from de Lima Procópio *et al.,* (2012)**.**

Competitive *Streptomyces* survival in marine and soil environments has also led to the production of secondary metabolites in the form of DNA intercalating molecules, functioning to inhibit *E. coli* phages (Kronheim *et al.*, 2018). The study by Kronheim *et al.,* (2018) investigated the inhibitory effects of two secondary metabolites, doxorubicin and daunorubicin, on the replication of twelve diverse *Streptomyces* phages using a range of *Streptomyces* species. Their results highlighted a clear inhibition of phage replication (10^5-fold) in the presence and phage propagation in the absence of doxorubicin and daunorubicin, showing an alternative role for *Streptomyces* secondary metabolites (Kronheim *et al.,* 2018) and potentially explains the abundance of anti-cancer agents produced by a soil bacterium.

1.4.2 Development and Regulation of *Streptomyces* Metabolism

Streptomyces development and co-regulation of physiological differentiation is controlled by a series of genes, such as the *bld* genes, which activate erection of aerial hyphae. It is thought that favourable nutrient conditions lead to the growth of vegetative or substrate mycelia and unfavourable conditions result is the growth of aerial mycelia (Piette *et al.*, 2005). The *bld* genes control the growth of the aerial mycelia, and proceeding sporulation of Streptomycetota is controlled by *whi* genes. The *whi* genes thus control segregation into spore chains and coiling of the mycelial branches. Both *bld* and *whi* genes are also involved in the switch between primary and secondary metabolism of the organisms in response to environmental conditions (Piette *et al.,* 2005).

Secondary metabolite production is regulated by global regulatory mechanisms, which are found at the next level up in the hierarchy. Global regulatory mechanisms allow for *Streptomyces* physiological differentiation and response to environmental stimuli, triggering the production of their secondary metabolites, which is essential for their survival in competitive environments. Regulation functions at transcription, translation, and post-translation levels, permitting gene cascades which allow for targeted gene expression.

S. clavuligerus global regulatory mechanisms involve three components: *bld* genes, autoregulator production and the stringent response (Paradkar, 2013). The *bld* genes, forming the BldD regulon, link antibiotic gene cluster expression to morphological differentiation. An example of this is *bldA*, a rare leucyl tRNA which translates TTA codon containing genes. The *bld* genes are key to Streptomycetota growth, as genes with TTA codons are only found in regulatory genes for secondary metabolism, aerial mycelium

formation and sporulation. Additionally, *bldG* genes play a role in sigma factor regulation, which in turn regulates genes responsible for the organism's morphological differentiation (Paradkar, 2013). In *S. clavuligerus*, the *ΔbldG* strains presented halted morphological differentiation, clavulanic acid and cephamycin C production, highlighting the importance of these gene cascades for growth and in the production of metabolites (Bignell *et al.*, 2005).

1.5 The Genome Architecture of *Streptomyces clavuligerus*

As mentioned in section 1.4, *Streptomyces* are known for their ability to produce a range of natural products and antibiotics. Underlying reasons for their ability to produce such a large variety of secondary metabolites is the genetic structure, starting with their genomic topology and architecture. Streptomycetota possess linear chromosomes consisting of conserved and variable regions, where the conserved regions carry most essential genes for primary metabolism and variable regions carry species specific genes for non-essential secondary metabolism (Hopwood, 2006). In addition to circular plasmids, many *Streptomyces* species, including *S. clavuligerus,* possess extra-chromosomal giant linear plasmids (GLPs) whose replication is independent of the chromosome (Medema *et al.*, 2010).

1.5.1 The Giant Linear Plasmids of *Streptomyces clavuligerus*

GLPs are essential for secondary metabolism of most actinobacteria and determine various other functions such as bacterial phytopathogenicity and compound biosynthesis (Hopwood, 2006). However, GLP§s are considered unstable, due to their dynamic nature. It has previously been shown that GLPs encode genes that play a large role in secondary metabolism, and that there is no evidence linking them to primary metabolism (Medema *et al.,* 2010). Investigating the essentiality of their role in *Streptomyces* metabolite production, e.g., for clavulanic acid, has therefore become increasingly important.

More recent genome publications underlining the genome architecture and genome similarities between most *Streptomyces* bacteria include *Streptomyces bottropensis* ATCC and *Streptomyces rimosus*, which have provided insight into their topology (Pethick *et al.*, 2013; Zhang *et al.*, 2013). However, often times these sequences are in a draft genome, which is not representative of the full scope of the genetic features of the species. Complete or 'closed' genome sequences exist for *Streptomyces avermitilis (S. avermitilis)*, *Streptomyces coelicolor* (*S. coelicolor*) A3(2) and *Stretptomyces rochei*

(S. rochei), allowing for comparison of more conserved regions of the bacterial genome (Bentley *et al.*, 2002; Ikeda *et al.*, 2003). The *S. clavuligerus* ATCC 2764 type strain has been sequenced by Medema *et al.,* (2010) and Song *et al.*, (2010), allowing for more insight into the GLPs and their genetic structure. Additionally, there have been genome publications of an industrial clavulanic acid producing *S. clavuligerus* strain F613-1, which has led to analytic comparisons between wild-type and genetically modified strains, however none of the mentioned *S. clavuligerus* genomes are complete (Cao *et al.*, 2016).

The *S. clavuligerus*, DSM 738 type strain contains a 6.76 Mb chromosome and four giant linear plasmids (GLPs) (Algora-Gallardo*.,* unpublished data; Medema *et al.,* 2010). The discovery of *S. clavuligerus'* additional replicons dates back to research by Kirby, (1978), who discovered that holomycin was produced by an alternate genetic element to the chromosome. The *S. clavuligerus* chromosome encodes 23 secondary metabolite clusters, the clavulanic acid cluster being one, and pulse-field gel electrophoresis (PFGE) has revealed that its extra-chromosomal GLPs consists of one megaplasmid (1.79 Mb) and three smaller, linear plasmids (10.5, 149.4, and 444.2 kb, respectively) (Wu and Roy, 1993). These are referred to as pSCL1-4 as seen in **Table 1.2**.

Table 1.2. Chromosome and GLP architecture of *Streptomyces clavuligerus*, *Streptomyces coelicolor* and *Streptomyces griseus* based on the literature.

	Streptomyces clavuligerus					***Streptomyces coelicolor***	***Streptomyces griseus***
	Chr	**pSCL4**	**pSCL3**	**pSCL2**	**pSCL1**	**Chr**	**Chr**
Sequence length	**6.8 Mb**	**1.8 Mb**	**0.45 Mb**	**0.15 Mb**	**0.01 Mb**	**8.7 Mb**	**9.0 Mb**
GC%	**72%**	**72%**	**72%**	**72%**	**72%**	**72%**	**71%**
rRNA operons	**6**	**0**		**0**	**0**	**6**	**6**
Coding sequences	**5,700**	**1,581**	**528**	**175**	**12**	**7,825**	**7,238**
Secondary metabolite clusters	**23**	**17**	**1**	**0**	**0**	**20**	**34**

*Adapted from Medema *et al.,* (2010) and Algora-Gallardo *et al.,* (unpublished data), Chr = chromosome.

Dynamic evolutionary events include the formation of pSCL4, which have furthered the understanding of the role GLPs have for *Streptomyces*. pSCL4 has evolved through recombination of pSCL1 with the chromosome, seen by an overlapping 7 kb DNA sequence between pSCL1 and 4 (Medema *et al.,* 2010). Homologous recombination and single crossovers between plasmids and the chromosome are a few of the common horizontal transfer mechanisms of *Streptomyces* bacteria (Kinashi, 2011). Particularly the transfer of antibiotic production clusters involved in secondary metabolism between GLPs, and the chromosome have been reported for *S. rimosus* and *S. coelicolor* (Pandza *et al.*, 1998; Yamasaki and Kinashi, 2004). Research by Pandza *et al.,* (1998) and Yamasaki and Kinashi, (2004) showed transfer of the *S. rimosus* oxytetracycline cluster from the chromosome to its pPGZ3 plasmid and transfer of the *S. coelicolor* methylenomycin production cluster from SCP1 to the chromosome, highlighting the interchangeability of these mobile genetic elements.

1.6 Regulation of Linear Growth and Replication

Genetic features that play a role in the linear replication of Streptomycetota can be witnessed in the chromosome of *S. clavuligerus*, such as an origin of replication site (*oriC*) flanked by *dnaA* and *dnaN* genes, allowing for autoregulation and control of replication (Medema *et al.,* 2010). The flanked *oriC* is also evident in other species such as *S. coelicolor* and *Streptomyces lividans* (*S. lividans*) (Zakrzewska-Czerwińska, Majka and Schrempf, 1995). Additionally, an *oriC* site can also be seen in pSCL4, the largest

linear replicon of *S. clavuligerus*, with high homology to similar plasmids such as pSLA2-L in *S. rochei* or SCP1 in *S. coelicolor*, highlighting similarities between chromosomal and plasmid replication (Kinashi, 2011; Medema *et al.,* 2010). Other common genetic components shared amongst most *Streptomyces* linear replicons include partitioning systems that involve proteins such as ParA, ParB, SMC (structural maintenance of chromosome) and FtsK.

Partitioning of replicons allows *Streptomyces* species to develop spores from aerial hyphae without resulting in a loss of plasmids throughout subcultures, important for the maintenance of genetic integrity (Hsu and Chen, 2010). Partitioning and transferring of GLPs is thought to depend on the highly conserved *parAB* operon, as seen for SLP2 in *S. lividans*, SCP1 and SCP2 in S. coelicolor. During partitioning, ParB recruits ParA to the *parS* site to form the essential ParB-*parS* complex which is vital symmetry during the process. The *parAB* operon can also be found in *Streptomyces* linear chromosomes, for example in *S. coelicolor* where the operon is clustered close to the *oriC* site (Hsu and Chen, 2010). In *S. clavuligerus*, both the chromosome and the GLPs (aside from pSCL1) contain genes encoding ParA and ParB. The *parAB* genes found within its 6.8 kB linear chromosome show high similarity to *S. griseus* and *S. avertimitilis* chromosomal genes, unlike the *parAB* genes on its GLP, pSCL4, which more closely resemble genes on the SCP1 plasmid of *S. coelicolor*, highlighting differences between the chromosome and GLPs when considering conserved genes (Medema *et al.,* 2010).

SMC proteins are a form of molecular motor, maintaining chromosomal structure, organisation, and condensation of the nucleoid. In *S. coelicolor* SMCs and associated proteins, ScpA and ScpB are responsible for growth and viability. This was underlined in previous research, testing various single and double knockout mutants in *S. coelicolor*. The Δ*scpA* and Δ*scpAB* double knockout presented altered, bilobed morphological shapes in *S. coelicolor* aerial hyphae when compared to the wild type (WT). When a plasmid encoding the *scpAB*$^{+}$ genes was reintroduced to the WT *S. coelicolor* strain, a reduction of the bilobal morphology was witnessed, emphasising the role of *Streptomyces* SMC proteins for linear growth (Dedrick, Wildschutte and McCormick, 2009).

Similar to the SMC proteins, FtsK is a DNA motor protein and contributes to chromosome segregation, vital during vegetative growth leading to sporulation, particularly abundant during aerial hyphae growth. More specifically, FtsK (belonging to the FtsK/SpoIIIE

family) separates replicated chromosomes (daughter cells) during sporulation by closing a septum (Thoma and Muth, 2016). The function of FtsK becomes important during bacterial conjugation, which involves DNA translocation and the TraB plasmid encoded protein in *Streptomyces* species. Streptomycetota are therefore thought to have adapted their FtsK segregation mechanisms for the transfer of plasmids, particularly during conjugation (Sepulveda, Vogelmann and Muth, 2011).

The above-mentioned genetic features of *Streptomyces* species play a part in ensuring chromosomal linearity is passed on as replication occurs. Homologues of proteins such as ParAB, FtsK and SMC are found in various bacterial species such as *E. coli*, where they play a role in the transfer of chromosomal DNA to daughter cells. This brings into question how organisms with linear growth, such as *Streptomyces*, ensure segregation of linear chromosomes by clearing DNA from sites of septal closure (Wang *et al.*, 2007). While it was speculated that FtsK was responsible for DNA transfer and chromosome partitioning, knockout mutants in *S. coelicolor* have shown that these processes do not solely depend on the protein, as mutants still exhibited in-tact, complete chromosomes (Wang *et al.,* 2007).

The reasons behind the importance of chromosomal linearity and the question whether ancestral forms of the bacterial chromosome were linear or circular thus remain unclear (Zhou *et al.*, 2011). It has however become evident, that chromosomal linearity presents evolutionary advantages that allow for a larger chromosome size in bacterial species, such as *Streptomyces*. Central regions of the chromosome often contain conserved genetic regions with essential genes for growth and sporulation, unlike the chromosomal ends which contain species specific genes. In addition, this form of linear, genetic organisation allows a separation between the telomeric ends and the core region of the chromosome, leading to less functional disruption (Zhou *et al.,* 2011). These archetypal telomeres also encode genes vital for secondary metabolism and antibiotic production. However even within the archetypal telomeres, various conserved, palindromic regions can be witnessed, highlighting their importance in maintaining linearity throughout replication of the chromosome and plasmids (Alexander Steinbüchel, 2007). Several key genes have been associated with the maintenance of chromosomal linearity in Streptomycetota including *tap, tpg* and *ttr*.

1.7 Telomeres and terminal proteins of *Streptomyces*

Streptomyces, along with other genera such as *Kitasatospora* and *Rhodococcus* posess linear chromosomes complemented by linear or circular plasmids. *Streptomyces* linear replicons can be identified by their invertron-structured terminal regions, also referred to as terminal inverted repeats (TIRs) (Qin and Cohen, 1998). A large diversity of TIRs can be witnessed in closely related strains, particularly highlighting the transfer of genetic material or allelic exchange between the chromosome and GLPs through horizontal gene transfer (Tidjani, Bontemps and Leblond, 2020). The genetic exchange between chromosomes and GLPs could be a potential reason behind the metabolic diversity witnessed in Streptomycetota, and underline part of the significance of linear replicons (Zhang *et al.,* 2020).

Telomere sequences of most *Streptomyces* chromosomes can be divided into archetypal and non-archetypal single strand overhangs of DNA. Archetypal telomeres can be seen in the various *Streptomyces* species, such as *S. coelicolor* A3(2), *S. rochei* pSLA2, *S. lividans* in addition to the *S. clavuligerus* DSM 783 chromosome and GLPs (**Figure 1.4**). Archetypal telomeres occur at the terminal 150-160 bp and are the most characterised type. They contain an extremely conserved 13 palindromic motif (palindrome I) alongside 3-6 more palindromes at the end of the chromosome, forming complementary stem structures referred to as either ‘rabbit ears’ or ‘clover leaves’ (Tidjani, Bontemps and Leblond, 2020; Algora-Gallardo *et al.*, 2021a). In the ‘rabbit ears’ model, palindrome I is complementary to a segment of palindrome IV, leaving palindromes II and III to form hairpin structures connected by A and G at the ends. In the ‘clover leaf’ model, palindromes I-IV each form a hairpin loop (Yang, Tseng and Chen, 2015). Non- archetypal telomeres can be witnessed in the SCP1 GLP of *S. coelicolor* in addition to *S. griseus* and two linear plasmids pLR1 and pLR2 (Goshi *et al.*, 2002; Zhang *et al.*, 2006a; Ohnishi *et al.*, 2008). Non-archetypal telomeres tend to be more variable, however still contain palindromes.

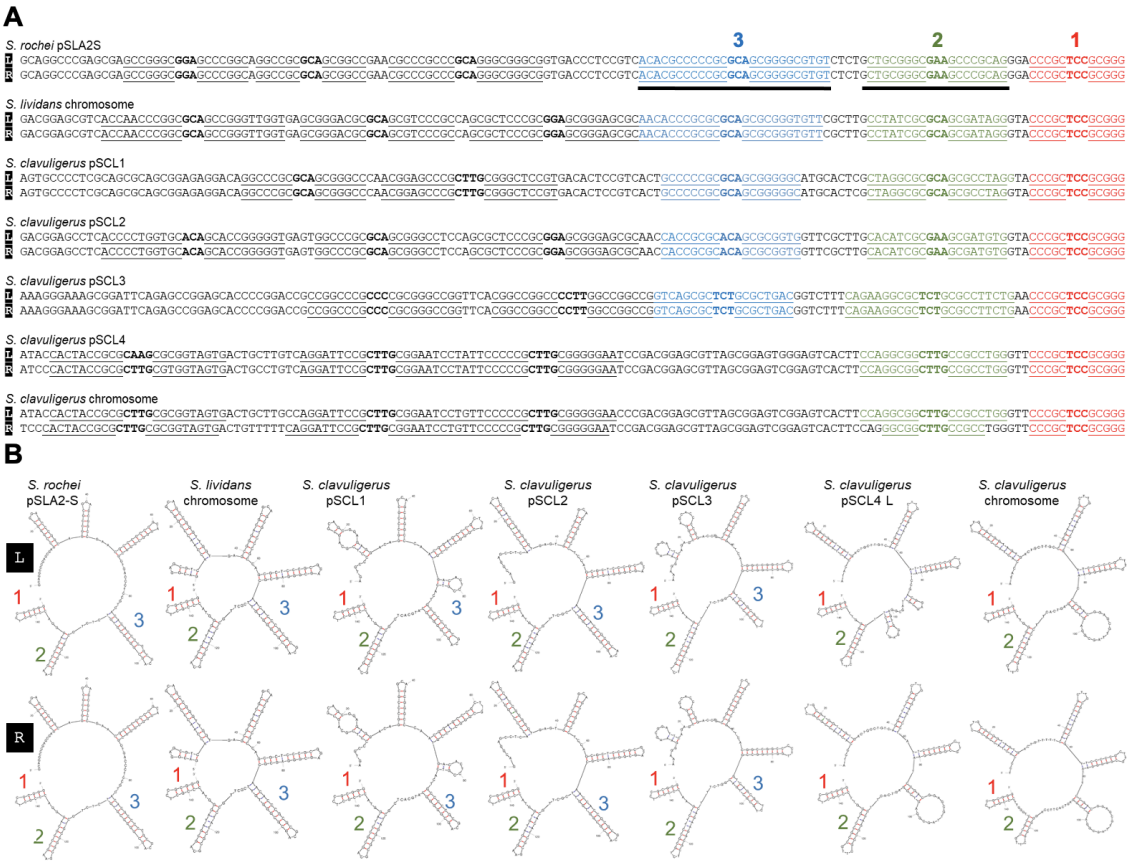


Figure 1. 4. **Clover-leaf based model of *Streptomyces* archetypal telomeres.** The *S. clavuligerus* chromosome and GLP, pSCL4 present unusual telomeres compared to the *S. lividans* and *S. rochei* chromosome and plasmid. Terminal sequences are highlighted in colours matching palindromes 1 (red), 2 (green) and 3 (blue), with the 13-bp conserved sequence highlighted as palindrome 1. Structures were predicted using Mfold (Zuker, 2003).

It is the interaction of the above-mentioned palindromic sequences with two proteins, Tap and Tpg, that allows for effective DNA replication and linear growth. New *Streptomyces* DNA is synthesized by protein priming, a concept that has been investigated in various other organisms such as eukaryotic adenoviruses and bacteriophages (Salas, 1991; Salas *et al.*, 2016). DNA replication is initiated at the origin of replication (oriC), continuing bi-directionally towards both ends of the chromosome allowing for DNA duplication, except at the 3' ends, leading to single stranded overhangs which cannot be replicated due to the failure to prime the Okazaki fragments of the lagging strand (Yang, Tseng and Chen, 2015; Yang *et al.,* 2017, 2017). These single-stranded gaps are filled up by a process called 'end patching', where DNA synthesis is primed by terminal proteins, preventing loss of genetic information and chromosomal shortening throughout rounds of replication (Tidjani, Bontemps and Leblond, 2020). The terminal protein Tap, which contains a DNA binding helix-turn-helix domain and nuclear localisation signals, binds to palindromes II and III allowing for interaction with Tpg (Yang *et al.,* 2017) (**Figure 1.5**).

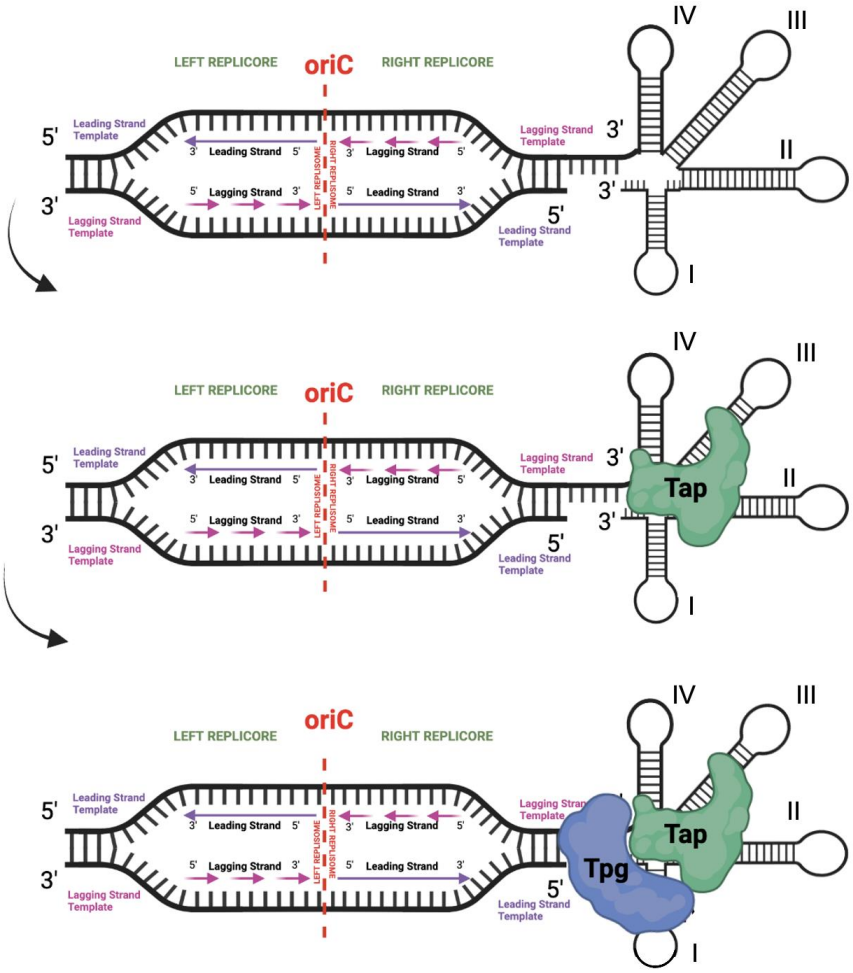


Figure 1. 5. Overview of the telomeric end-patching process lead by Tap and Tpg, allowing for maintencence of chromosome and plasmid ends. Replication is initiated at oriC and proceeds bi-directionally towards both ends. Okazaki fragment maturation of lagging strand leaves single strand overhangs at 3' ends of new DNA, which fold into loops: palindromes I-IV. Tap interacts with palindromes II and III, recruiting Tpg which primes the lagging strand, allowing extension. After end-patching, terminal proteins stay covalently attached to DNA ends. Own Illustration.

Additionally, replication in *Streptomyces* with both archetypal and non-archetypal telomeres are aided by DNA motifs. More typical DNA motifs can be seen for archetypal telomeres, represented by a 3-nucleotide 5′-GCA-3′ in the central part of the palindrome, which are thought to play a protective part during replication, resisting nucleases which target the single strand overhangs. 4-nucleotide or 3-nucleotide loops can be seen in non-archetypal telomeres (Tidjani, Bontemps and Leblond, 2020). Although their function is not fully determined, these motifs play a role in recruiting other terminal proteins (homologues of Tap and Tpg), such as Tac and Tpc in SCP1 for *S. coelicolor* (Huang *et al.,* 2007). It is thought that non-archetypal telomeres can be capped by both archetypal and non-archetypal terminal proteins, sharing the helix-turn-helix domains as well as nuclear localisation signals and both archetypal and non-archetypal telomeres have been observed in different replicons of the same organism (Tsai *et al.,* 2008).

The specific function of Tap and Tpg has been witnessed in *S. clavuligerus* (Algora-Gallardo unpublished data, 2019; Medema *et al.,* 2010). pSCL4 is the largest of four GLPs and has been shown to carry 25 gene clusters encoding secondary metabolism proteins, two more than on the organism's chromosome, establishing its biological role (Medema *et al.,* 2010). The chromosome of *S. clavuligerus*, amongst other organisms such as the *Rhodococcus* species (RHA-1), lacks the presence of the *tap* and *tpg* genes (McLeod *et al.,* 2006). However, one copy of *tap-tpg* can be found on pSCL2 (*tap-tpg*$_2$), pSCL3 (*tap-tpg*$_3$) and pSCL4 (*tap-tpg*$_4$) GLPs as shown in **Figure 1.6**. In contrast organisms like *Streptomyces hygroscopicus subsp*. *jinggangensis* carry on copy of *tpg* on the chromosome and plasmid, however two copies of *tap* on the chromosome (Wu *et al.*, 2012; Algora-Gallardo *et al.*, 2021a). The paper by Algora-Gallardo *et al.,* (2021) underlines *that*, like *S. clavuligerus*, various strains such as *S. rochei* 7434AN4 (Nindita *et al.,* 2015, 2019) contain *tap-tpg* on their GLPs but not on the chromosome, suggesting that plasmids may be required for chromosomal end-patching as they would encode the necessary proteins to do so.

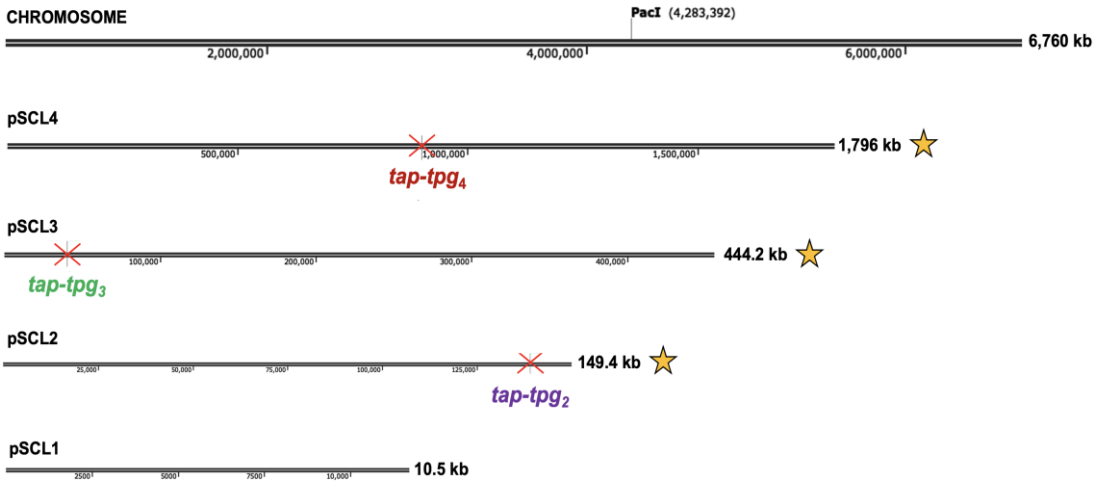


Figure 1. 6 The five linear replicons and three terminal proteins of *S. clavuligerus*. The presence of terminal proteins on GLPs is marked with a star, showing that pSCL4, pSCL3 and pSCL2 contain one copy of *tap-tpg* marked according to location and named according to plasmid. No copy of *tap-tpg* is found on the chromosome or pSCL4. Chromosome = 6,760 kb, pSCL4 = 1,796 kb, pSCL3 = 444.2 kb, pSCL2 149.4 kb, pSCL1 10.5 kb. Not to scale.

To determine how essential these terminal proteins are on pSCL4, and whether the genes on the chromosome alone allow effective growth, previous work at Strathclyde by Algora-Gallardo, (unpublished data) investigated the effect of pSCL4 removal on *S. clavuligerus'* chromosome. The removal of pSCL4 lead to chromosome circularisation, potentially due to the lack of *tap-tpg*$_4$, disabling linear growth as seen for *S. rochei* and *S. lividans* in the study by Bao and Cohen, (2001). A schematic representation can be seen in **Figure 1.7**.

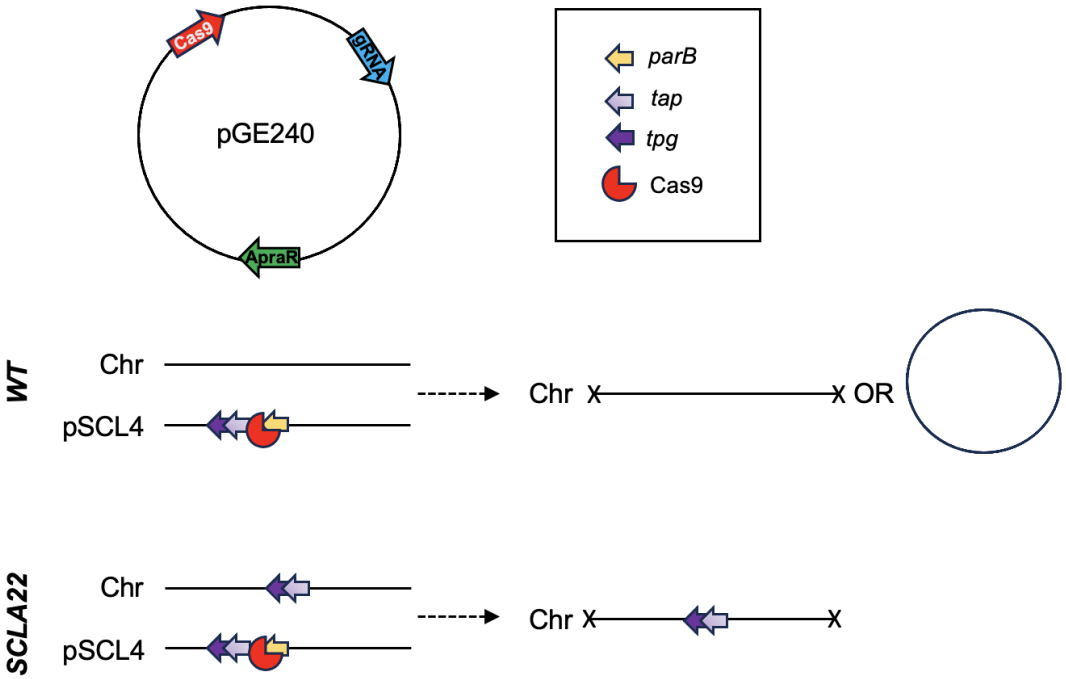


Figure 1. 7 Cas9 mediated cleavage of pSCL4 eliminates plasmid from *S. clavuligerus* DSM 738. The pGE240 plasmid contains a CRISPR-Cas9 system and carries a spacer-guide RNA complex (sgRNA) targeting *parB* on pSCL4, leading to entire plasmid loss and loss of chromosome ends (X) leading to some mutants presenting circularised chromosomes Algora-Gallardo, (unpublished data). A pGE240 based plasmid, SCLA22 was complemented with a copy of *tap-tpg*$_4$, and *S. clavuligerus* mutants were derived from the strain, showing pSCL4 loss and similar loss of chromosome ends to WT and highlighting the relationship between *tap-tpg*$_4$ and the chromosome. Adapted from Algora-Gallardo, (unpublished data).

To summarise, telomere replication depends on the interaction of palindromic sequences with Tap and Tpg and is also a mechanism allowing for increased production of secondary metabolites, such as clavulanic acid, as the terminal proteins work to maintain the linear chromosomal integrity and allow better growth. The significance of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ in the production of clavulanic acid by *S. clavuligerus* has not yet been determined due to challenges in finding appropriate, targeted genetic engineering tools. Targeted genetic engineering tools have not just presented difficulties for *S. clavuligerus*, but for actinomycetes altogether due to their diverse genomes (Tong *et al.*, 2015). An area that has only recently been explored for *Streptomyces* is utilising modern genetic tools such as clustered regularly interspaced short palindromic repeats (CRISPR) cas9, dcas9 or the novel CRISPR-BEST (Base Editing SysTem) which have all shown promising results in targeted gene editing (Cobb, Wang and Zhao, 2015; Huang *et al.*, 2015; Tong *et al.*, 2015, 2019).

1.8 Scope and Aims of the Project

Clavulanic acid combined with amoxicillin (Augmentin® – GSK) is one of the most commonly prescribed antibiotics to treat wide-spectrum beta-lactamase producing bacterial infections (Elander, 2003). Clavulanic acid production involves *S. clavuligerus* fermentation and, in the UK, it is produced industrially at a plant in Irvine, Scotland. The yield of production remains a constant condition in need of optimisation, and clavulanic acid yield tend to be lower in the UK compared to overseas competitors of Augmentin® production (GSK - personal communication). Recognising the complex genome architecture of *S. clavuligerus* is vital to genetically engineering strains that possess increased fermenter stability and clavulanic acid productivity, necessary for the continuation of clavulanic acid production by GSK. The goal of this project is to investigate the genome dynamics of *S. clavuligerus*, focussing on the GLP-chromosome relationship, with the goal of understanding the essentiality of *tap-tpg*$_4$, *tap-tpg*$_3$, and *tap-tpg*$_2$ in the maintenance of plasmid and chromosome linearity and thus effective growth.

This project will build on from earlier work at Strathclyde, which also looked at pSCL4 significance and demonstrated that Cas9 mediated cleavage of pSCL4 lead to loss of plasmid and chromosomal circularisation (see **Figure 1.7**). Complementation of the chromosome with *tap-tpg*$_4$ did not affect chromosome linearity, highlighting the relationship between the GLP and chromosome. Therefore, it remains unclear whether

it was the loss of *tap-tpg*$_4$ which resulted in the loss of pSCL4 and chromosomal circularisation.

Therefore, the aim of this project is to investigate the significance and relationship of *tap-tpg*$_4$ with chromosomal linearity. To look at the significance of *tap-tpg*$_4$, additional terminal protein-encoding genes on the GLPs of *S. clavuligerus* DSM 738, pSCL3 (*tap-tpg*$_3$) and pSCL2 (*tap-tpg*$_2$), will be characterised. The role of *tap-tpg*$_4$, *tap-tpg*$_3$, and *tap-tpg*$_2$ in the genome dynamics of *S. clavuligerus* will be demonstrated by constructing CRISPR-dCas9 and CRISPR-BEST knockdowns targeted to silence or block these genes and/or encoded proteins.

Once the significance of Tap and Tpg have been confirmed for plasmid and chromosomal linearity and growth, the necessity of the GLPs will be considered, with the objective of curing the GLPs to decrease the total metabolic burden elicited by their non-essential genes.

The Tap and Tpg proteins of Streptomycetota have been previously characterised, for example in *S. lividans* (Bao and Cohen, 2001), and their role in 3' end-patching during linear replication has been identified. The *S. clavuligerus* DSM 738 terminal proteins remain uncharacterised and will therefore be overexpressed and purified, to determine their involvement with the 3' terminal ends and further elucidate the relationship between Tap on GLPs and the chromosome.

Understanding the interaction of terminal proteins with replicon ends is not just important in finding out replicon essentiality in *S. clavuligerus,* but for our general understanding of the biology of *Streptomyces*. If we can understand the importance of the various *tap-tpgs*, it will eventually allow for manipulation of the genes that play a role in linear growth and increased CA production.

2.0 Materials and Methods

2.1 Media and Antibiotics

Media used to culture microbial strains was autoclaved at 121 °C for sterilisation. Antibiotics and solutions were filter sterilised using syringes with 0.22 µm filters.

Table 2. 1. List of ingredients used to make up media for the culture and growth of microbial strains used in this report.

MEDIUM	INGREDIENTS	AMOUNT
GYM AGAR	Agar	- 12 g
	$CaCO_3$	- 2 g
	Glucose	- 4 g
	Malt extract	- 10 g
	Yeast extract	- 4 g
	Distilled water (dH_2O)	- 1000 mL
LURIA-BERTANI (LB) AGAR	Agar	- 40 g
	Sodium Chloride (NaCl)	- 20 g
	Tryptone	- 40 g
	Yeast extract	- 20 g
	dH_2O	- 1000 mL
LB BROTH	NaCl	- 20 g
	Tryptone	- 40 g
	Yeast extract	- 20 g
	dH_2O	- 1000 mL
L3M9 AGAR	Dextrin	- 4 g
	Trehalose dihydrate	- 10 g
	Dipotassium phosphate	- 0.5 g
	NaCl	- 1 g
	Magnesium sulphate	- 1 g
	Calcium chloride	- 0.5 g
	Casamino acids	- 2 g
	MOPS buffer	- 10.5 g
	Agar trace salts	- 1 mL
	Agar	- 30 g
	dH_2O	- 1000mL
TSB	Casein peptone	- 17 g
	Soya peptone	- 3 g
	NaCl	- 5 g
	K_2HPO_4	- 2.5 g
	dH_2O	- 1000 mL
SOY FLOUR MANNITOL (SFM)	Agar	- 16 g
	Mannitol	- 16 g
	Soy Flour	- 16 g
	Tap water	- 1000 mL
S.O.C MEDIA	Glucose (1 M)	- 20 mL
	Magnesium chloride ($MgCl_2$) (1 M)	- 10 mL
	Magnesium sulphate ($MgSO_4$) (1 M)	- 10 mL
	NaCl (5 M)	- 2 mL
	Potassium chloride (KCl) (1 M)	- 2.5 mL

	Tryptone Yeast Extract dH_2O	- 20 g - 5 g - 1000 mL
2 X YT	NaCl Tryptone Yeast extract dH_2O	- 5g - 16 g - 10 g - 1000 mL
ZY MEDIA (STUDIER, 2005)	Tryptone Yeast Extract dH_2O	- 10 g - 5 g - 1000 mL
50 X 5052	Glycerol D-Glucose Lactose	- 0.5 % (w/v) - 0.05% (w/v) - 0.2 % (w/v)
20 X NPS	$(NH_4)_2\ SO_4$ KH_2PO_4 Na_2HPO_4	- 0.5 M - 1 M - 1 M
ZYP-5052	$MgSO_4$ 50 x 5052 20 x NPS	- 1 M - 20 mL - 50 mL

Table 2. 2. List of antibiotics used to select for the appropriate microbial strains shown as stock and final concentrations in media.

ANTIBIOTIC	STOCK CONCENTRATION (μg/mL)	FINAL CONCENTRATION (μg/mL)
AMPICILLIN	100 in dH_2O	100
APRAMYCIN	50 in dH_2O	50
CHLORAMPHENICOL	25 in 100% ethanol	25
KANAMYCIN	25 in dH_2O	25
NALIDIXIC ACID	25 in dH_2O	25

2.2 Microbial Strains and Plasmids

All the microbial strains used in the experiments for this report and their genetic adaptations are shown in **Table 2.3.** *S. clavuligerus* strains DSM 738 and DSM 41826 were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). All the plasmids used in this report are highlighted in **Table 2.4.** (- symbol means lost plasmid)

Table 2. 3. List of microbial strains, highlighting their respective characteristics and sources.

Strain	Genotypic Characteristics	Reference
E. coli **Subcloning Efficiency™ DH5α Competent Cells**	*F- Φ80lacZΔM15 Δ(lacZYA-argF), U169, recA1, endA1, hsdR17(rk-, mk+), phoA, supE44, thi-1, gyrA96, relA1 λ-*	Invitrogen
E. coli **ET12567/pUZ8002**	*dam-13::Tn9 dcm-6 hsdM hsdR recF143, zij201::Tn10, galK2, galT22, ara14, lacYI, xylS, leuB6, thi-1, tonA31, rpsL136, hisG4, tsx78, mtli, glnV44, F-*	(MacNeil et al., 1992)
E. coli **BL21 (DE3)**	*F– ompT hsdSB (rB–, mB–) gal dcm (DE3)*	(Wood, 1966)
S. clavuligerus **DSM 738**	Wildtype	(Higgens and Kástner, 1971)
S. coelicolor **A3(2) M145**	Wildtype	(Bentley et al., 2002)
SCO-pSET	*S. coelicolor* M145 conjugated with pSET152	This study
SCO-cBEST	*S. coelicolor* M145 conjugated with pCRISPR-cBEST	This study
SCO-pEM4	*S. coelicolor* M145 conjugated with pEM4	This study
SCO-1370	*S. coelicolor* M145 conjugated with pGWS1370	This study
SCLAV-pSET	*S. clavuligerus* DSM 738 conjugated with pSET152	This study
SCLAV-pLIS22	*S. clavuligerus* DSM 738 conjugated with pLIS22	This study
SCLAV-pEM3	*S. clavuligerus* DSM 738 conjugated with pEM3	This study
SCLAV-cBEST	*S. clavuligerus* DSM 738 conjugated with pCRISPR-cBEST	This study
SCLAV-pEM4	*S. clavuligerus* DSM 738 conjugated with pEM4	This study
SCLAV-1370 also (1370)	pSCL2- *S. clavuligerus* DSM 738	This study

EM2	pSCL2- *S. clavuligerus* DSM 738	This study
EM3	pSCL1-, pSCL2-, pSCL3- *S. clavuligerus* DSM 738	This study
EM4	pSCL2- *S. clavuligerus* DSM 738	This study
EM5	pSCL2- *S. clavuligerus* DSM 738	This study
EM6	pSCL2-, pSCL3- *S. clavuligerus* DSM 738	This study
EM7	pSCL1-, pSCL2- *S. clavuligerus* DSM 738	This study
EM8	pSCL1-, pSCL2-, pSCL3- *S. clavuligerus* DSM 738	This study
EM9	pSCL2- *S. clavuligerus* DSM 738	This study
EM10	pSCL2-, pSCL3- *S. clavuligerus* DSM 738	This study
M5	pSCL4-, pSCL3- *S. clavuligerus* DSM 738	Algora-Gallardo

Table 2. 4. List of plasmids used and/or constructed in this report.

PLASMID	DESCRIPTION	ANTIBIOTIC RESISTANCE	RESTRICTION SITES USED	REFERENCE
pGEM®-T EASY	Cloning vector	Ampicillin	N/A (TA cloning)	Promega
pGWS1370	pSET152 based vector with removed NcoI site used for CRISPR-dCas9 cloning	Apramycin	BamHI, NcoRI	Larson and Hershberger, (1986), van Wezel, (2019)
pSET152	Shuttle vector *E. coli/Streptomyces*	Apramycin	EcoRI, XbaI	(Bierman et al., 1992)
pLIS22	pMS82 plus *tap-tpg* operon	Hygromycin	SacI	Algora-Gallardo *et al.,* (unpublished data)
pEM3	pGWS1370 backbone + CRISPR-dCas9 *tap-tpg* spacer insert (pSCL4)	Apramycin	BamHI, NcoRI	This report
pEM3.1	pGWS1370 backbone + CRISPR-dCas9, different *tap-tpg* spacer insert (pSCL4)	Apramycin	BamHI, NcoRI	This report
pEM3.2	pGWS1370 backbone + CRISPR-dCas9 *tap-tpg* spacer insert (pSCL3)	Apramycin	BamHI, NcoRI	This report
pEM3.3	pGWS1370 backbone + CRISPR-dCas9 *tap-tpg* spacer insert (pSCL2)	Apramycin	BamHI, NcoRI	This report
pEM3.1.2	pGWS1370 backbone + CRISPR-dCas9 *tap-tpg* spacer insert (pSCL2)	Apramycin	EcoRI, BamHI/BglII	This report
pEM3.1.3	pGWS1370 backbone + CRISPR-dCas9 *tap-tpg* spacer insert (pSCL2)	Apramycin	EcoRI, BamHI/BglII	This report
pEM3.2.3	pGWS1370 backbone + CRISPR-dCas9 *tap-tpg* spacer insert (pSCL2)	Apramycin	EcoRI, BamHI/BglII	This report
pEM3.1.2.3	pGWS1370 backbone + CRISPR-dCas9 *tap-tpg* spacer insert (pSCL2)	Apramycin	EcoRI, BamHI/BglII	This report
pEM4	pSET152 backbone + CRISPR-dCas9 *tap-tpg* spacer insert	Apramycin	EcoRI, XbaI	This report
pCRISPR-cBEST	CRISPR BEST plasmid	Apramycin	NcoI	Tong *et al.,* (2020)
pUC57_34640	Genscript Tap4 codon optimised vector	Ampicillin	EcoRI, BamHI	This report
pUC57_28815	Genscript Tap3 codon optimised vector	Ampicillin	EcoRI, BamHI	This report
pUC57_28545	Genscript Tap2 codon optimised vector	Ampicillin	EcoRI, BamHI	This report

pBGS19	Kanamycin resistant analogue of pUC8 and pUC9	Kanamycin	EcoRI, BamHI	(Spratt *et al.*, 1986)
pBGS19_34640	Tap_4 codon optimised sequence cloned into pBGS19	Kanamycin	BamHi, NdeI	This report
pBGS19_28815	Tap_3 codon optimised sequence cloned into pBGS19	Kanamycin	BamHI, NdeI	This report
pBGS19_28545	Tap_3 codon optimised sequence cloned into pBGS19	Kanamycin	BamHI, NdeI	This report
pET15b+	pET vector with N terminus His Tag	Ampicillin	BamHI, NdeI	Addgene
pET15+_34640	Tap_4 codon optimised sequence cloned into pET15+	Ampicillin	BamHI, NdeI	This report
pET15+_28815	Tap_3 codon optimised sequence cloned into pET15+	Ampicillin	BamHI, NdeI	This report
pET15+_28545	Tap_2 codon optimised sequence cloned into pET15+	Ampicillin	BamHI, NdeI	This report

2.3 Primers

Primers used for polymerase chain reaction (PCR) amplifications, CRISPR-dcas9 sgRNA construction, CRISPR-cBEST sgRNA construction and for sequencing can be found in **Table 7.** Primer design was completed with SnapGene software and ordered through Integrated DNA Technologies (IDT).

Table 2. 5. List of primers designed and used in this report with highlighted spacer sequences and underlined protospacer adjacent motif.

NAME	GENE TARGET	DIRECTION	5' - 3' SEQUENCE	TM (°C)
sgRNA_F_PRIMEROLD	*tap-tpg* on pSCL4	F	catgCCATGG**acggtatccgaagccgagac**GTTTTAGAGCTAGAAATAG	67.6
sgRNA_F_PSCL4-TAP	*tap-tpg* on pSCL4	F	catgCCATGG**cgggcgtacgcctgtcgctt**GTTTTAGAGCTAGAAATAG	73.0
sgRNA_F_PSCL3-TAP	*tap-tpg* on pSCL3	F	catgCCATGGg**cgagatccactcgggttgc**GTTTTAGAGCTAGAAATAG	71.0
sgRNA_F_PSCL2-TAP	*tap-tpg* on pSCL2	F	catgCCATGG**cgaagcccgctgcatcgtcg**GTTTTAGAGCTAGAAATAG	72.0
sgRNA_R	*dcas9*	R	ctagGGATCCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTACGCCTACGTAAAAAAAGCACCGACTCGGTGCC	85.7
SPACER_SPECIFIC_SEQUENCE	*tap-tpg*	F	ACG GTA TCC GAA GCC GAG AC	59.0
dCAS9_R3SEQ_R	*dcas9*	R	GATCGAGTGCCGGTCGGTGT	64.0
LIS_RT_TAP_RV	*Tap on pSCL4*	F	TGTCTTGACCATCTTGCGGTGCATGTT	65.0
ELMIRA_TAP_RV	*Tap on pSCL4*	R	AAGCTGACACAGCGCGGCTT	66.0
LIS-SCL4-L-END_F	*pSCL4 L telomere*	F	TGCGGCACTTGCCCAGTACAAGACCC	66.4
LIS-SCL4-L-END_R	*pSCL4 L telomere*	R	GGTGGTCTGTCCCACGGCACTTCACAC	66.7
LIS-SCL4-R-END_F	*pSCL4 R telomere*	F	GGGGTCCTGGAGGAAGGCGTAGTTCTTGT	66.1
LIS-SCL4-R-END_R	*pSCL4 R telomere*	R	TACACGAAGAAGCCCTTCGACGGGGGC	67.3
LIS-CHROMTELPROBE_F	*Chromo-some L telomere*	F	TTGATCCGAAACGCTTTAAGTTCAGGCTCG	62.1
LIS-CHROMTELPROBE_R	*Chromo-some L telomere*	R	TAGACGATGGGACTGCGTTGTCCGAGTCG	65.8

LIS-CAOCHR-R-END_F	*Chromo-some R telomere*	F	ACGGACCGCATCTGCTGGACATGCTC	66.4
LIS-CAOCHR-R-END_F	*Chromo-some R telomere*	R	GTTGTCTGTCGACCGCAGCCTGAAGCC	66.6
cBEST_SP 1_ssDNA	*tap-tpg on pSCL4*	F	CGGTTGGTAGGATCGACGGC**tcgtcgcccagccttccagc**GTTTTAGAGCTAGAAATAGC	75.0
cBEST_SP 2_ssDNA	*tap-tpg on pSCL4*	F	CGGTTGGTAGGATCGACGGC**agcgccccacgccaccagctc**GTTTTAGAGCTAGAAATAGC	61.0
cBEST_SP 3_ssDNA	*tap-tpg on pSCL4*	F	CGGTTGGTAGGATCGACGGC**cgcccagccttccagcaggc**GTTTTAGAGCTAGAAATAGC	69.0
ADDGENE_M13/PUC_R	*lacZ*	R	AGCGGATAACAATTTCACACAGG	58.0
M13_R_29	*lacZ*	R	CAGGAAACAGCTATGAC	49.0
PCRISPR-BEST_SP 1_F	*tap-tpg spacer1 on pEM4*	F	TCGTCGCCCAGCCTTCCAGC	64.0
TAP4CO_F	*Tap$_4$ codon optimised sequence (34640)*	F	GGTGACGACCGGTTGTCTCT	59.2
TAP4CO_R	*Tap$_4$ codon optimised sequence*	R	TTTCCTGGACGAAGAGGCCT	58.4
TAP3CO_F	*Tap$_3$ codon optimised sequence (28815)*	F	CGGTCGCATACACAACACAA	55.8
TAP3CO_R	*Tap$_3$ codon optimised sequence*	R	TCGTATCGAGTCCGATCCGC	58.7
TAP2CO_F	*Tap$_2$ codon optimised sequence (28545)*	F	ACCTCTTCACCCCACAGCAG	59.5
TAP2CO_R	*Tap$_2$ codon optimised sequence*	R	TGCCTTGGACGGTGGTTTAC	57.8
EM_RPOB_F	*rPOB on Chr Clav*	F	TGCCCCGCTGCGCATCTCCTTT	67.0
EM_RPOB_R	*rPOB on Chr Clav*	R	GTCTCGCCCGCGTGCAGCTTGAC	67.0
TAP2_CHECK_F	*tap$_2$ on pSCL2*	F	CGGAGAACGAGCTGTTCAGT	57.1
TAP2_CHECK_R	*tap$_2$ on pSCL2*	R	CAGCTCGAAGTCGAGGTGGA	58.9

2.4 Bacterial Cultivation and Growth

S. coelicolor strains were grown on SFM agar at 30 °C for 5 days, single colonies were selected and re-streaked onto new SFM agar plates and grown for another 5-7 days until aerial mycelia were formed. *S. clavuligerus* strains were grown on L3M9 agar at 26 °C for 7 days, single colonies were selected and re-streaked onto new L3M9 agar plates and grown for another 7-14 days. Spores were disturbed with sterile cotton buds and harvested using 5 mL 20% (v/v) glycerol. Spore suspensions were filtered through a 10 mL syringe filled with cotton and stored in cryovials at -80 °C and -20 °C. Spore stocks were used for inoculation of SFM and L3M9 plates during conjugation.

For liquid media growth, one *Streptomyces clavuligerus* colony was cut out and inoculated in TSB in a 50 mL baffled Erlenmeyer flask (VWR) for 3-5 days without antibiotics, or with 50 µg/mL apramycin.

E. coli strains were grown in LB broth at 37°C and 250 revolutions per minute (RPM) overnight (Infors HT Multitron). Appropriate antibiotics were added to plasmid-containing strains with resistances (see **Table 2.4.**). 500 µL of the overnight cultures was added to 500 µL of 50% (v/v) glycerol and stored at -80 °C for use in multi-step cloning experiments.

2.5 Cell Dry Weight Determination

Cells were cultured according to section 2.4 for a period of 108 hours to establish a growth curve. One mL samples were taken every 12 hours and frozen at -20 °C with cultures set up in biological triplicates for each *Streptomyces clavuligerus* strain. Dry weight was determined by applying defrosted cells to a 1 µm pore size, micro glass fibre filter (LifeTechnologies Limited) dried. Biomass was filtered through a Buchner Funnel (KIF Laboport and NALGENE 180 PVC metric tube) and the micro glass filter with the biomass was dried to a constant mass. Cell dry weight was determined by subtracting mass of the filter from the filter with biomass (mg/mL). The specific growth rate (µ) was calculated by plotting the $\ln(OD_{600})$ over time of the log phase of growth, which yields a linear correlation where the slopes equals µ. Data was plotted using GraphPad Prism (GraphPad Software, La Jolla California USA), and statistics were analysed by performing One-way ANOVA followed by Fisher's multiple comparison test.

2.6 Quantification of Clavulanic Acid in Supernatant

A 1.4M imidazole solution was prepared by dissolving 10g of imidazole (VWR chemicals) in 7.2 mL of 10M hydrochloric acid and topping up measuring cylinder to 100 mL with dH_2O. Clavulanic acid standards were prepared by dissolving lithium clavulanate in dH_2O and filter sterilising. Standards were diluted to 400, 350, 300, 250, 200 and 100 µg/mL and the linear equation was used to calculate clavulanic acid production. 1mL was taken from streptomyces clavuligerus cultures and spun down at 4000 x g for 5 minutes. Eight µL of the supernatant was mixed with 200 µL of imidazole solution and incubated for 30 minutes at room temperature. Absorbance was measured at 324 nm on the SpectraMax plate reader, and all samples and standards were pipetted in technical triplicates. Data was plotted using GraphPad Prism and One-way ANOVA was performed followed by Tukey's multiple comparison test.

2.7 CRISPRi/CRISPR-BEST Protocol – construction of plasmids using multi-step cloning.

2.7.1 Spacer selection

SnapGene and CRISPy-Web were used to generate the small guide RNA (sgRNA) spacer for gene silencing of *tap-tpg* in *S. clavuligerus* and introduction of stop codon using the following genome sequence: GCF_028752555.1. The spacers were 20 base pair DNA sequences (highlighted in bold in **Table 2.5.**) chosen to target non-template (coding) strands (Qi et al., 2013).

2.7.2 Plasmid DNA purification

10 mL overnight cultures of *E. coli* DH5-α cells previously transformed with pGWS1370, pCRISPR-cBEST and pSET152 with appropriate antibiotics (**Table 2.5.**) were purified using the spin column-based Wizard® Plus SV Minipreps DNA Purification System Kit (Promega). The purified DNA was eluted in 30 µL of nuclease free dH_2O and the final concentration was measured on the NanoDrop 2000c Spectrophotometer (Thermo Fischer Scientific).

2.7.3 Introduction of spacers into sgRNA scaffold and sgRNA confirmation

CRISPR-dCas9

Polymerase chain reaction (PCR) was set up using GoTaq® DNA Polymerase (Promega) and run in the Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems) thermal cycler.

Table 2. 6. PCR reaction set-up to introduce spacer into sgRNA scaffold

COMPONENT	VOLUME (µL)
5 X GREEN GOTAQ® BUFFER	10
DNTPS (10 MM)	1
FORWARD PRIMER (20 MM)	1.25
REVERSE PRIMER (20 MM)	1.25
DNA POLYMERASE (5U/ML)	0.25
TEMPLATE PLASMID (PGWS1370)	1
DMSO	2.5
DH_2O	Up to 50

Table 2. 7. Thermal cycler conditions for PCR reaction to introduce spacer into sgRNA.

STAGE	TEMPERATURE (°C)	TIME (MINS)	# OF CYCLES
INITIAL DENATURATION	95	3:00	1
DENATURATION	95	0:30	
ANNEALING	60	0:30	35
EXTENSION	72	6:00	
FINAL EXTENSION	72	5:00	1
HOLD	4	∞	-

Another PCR was performed to check for the sgRNA cascade in various transconjugants which used the Q5 (New England Biolabs) Hi fidelity DNA polymerase kit. Reaction components and conditions were as follows.

Table 2. 8. Ideal PCR set-up for probe amplifications and *S. clavuligerus* transconjugant confirmation.

COMPONENT	VOLUME (µL)
5 X REACTION BUFFER	5
DNTPS (10 MM)	0.5
FORWARD PRIMER (10 MM)	1.25
REVERSE PRIMER (10 MM)	1.25
DNA POLYMERASE	0.25
TEMPLATE GDNA/PDNA (50 NG/4.5NG)	1
DMSO	2.5
GC ENHANCER	5
DH_2O	Up to 20

Table 2. 9. Ideal thermal cycler conditions for general PCR reaction used in this work.

STAGE	TEMPERATURE (°C)	TIME (MINS)	# OF CYCLES
INITIAL DENATURATION	98	0:30	1
DENATURATION	98	0:10	
ANNEALING	72 Δ	0:30	26
EXTENSION	72	2:00s	
FINAL EXTENSION	72	5:00	1
HOLD	10	∞	-

CRISPR-cBEST

ssDNA oligonucleotides were designed and ordered (IDT) according to the templates in **Table 2.6**. and diluted in 1x NEBuffer2 to a concentration of 0.2 mM.

2.7.4 CRISPR-dCas9 PCR purification and TA cloning into pGEM®-T Easy

PCR amplicons from **Section 2.7.3** were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). To TA clone, reactions were set-up with pGEM®-T Easy vector using the 2X Rapid Ligation Buffer (Promega). Cloning reactions were set up in PCR tubes and reaction components included 2 µL 2X Rapid Ligation Buffer - T4 DNA Ligase, 1 µL of 50 ng/µL pGEM®-T Easy vector DNA, 1 µL of 25-30 ng/ µL PCR amplicon insert DNA, 1 µL T4 DNA Ligase (3 Weiss units/µl) and 5 µL of dH_2O to make up a 10 µL reaction volume. Negative control was set up using only pGEM®-T Easy without insert and background control was set up using vector plus dH_2O. Cloning reaction was incubated for one hour at room temperature and subsequently heat inactivated at 65°C for 10 minutes.

2.7.5 Linearisation of CRISPR plasmids by Restriction Digestion

CRISPR-dCas9

pGEM®-T Easy, pGWS1370, pEM3 and pEM3.1-3.3 were digested according to Promega. 1 µg/µL of plasmid DNA, 2 µL Restriction Enzyme 10x Buffer, 0.2 µL acetylated BSA and dH_2O were mixed in an Eppendorf tube by pipetting. For a single digestion, 1 µL of the according restriction enzyme was added to make up a total volume of 20 µL. The reaction was incubated at 37°C for 3 hours. For a double digest, 0.5 µL of

one enzyme was added to the reaction mix initially and after a 1-hour incubation at 37°C, 0.5 µL of the second enzyme was added and incubated at the same temperature for an additional 2 hours. Negative controls included all reaction components except the restriction enzyme/s.

CRISPR-cBEST

CRISPR-cBEST was digested using FastDigest NcoI in a 20 uL digestion reaction containing the following components:

Table 2. 10. CRISPR-cBEST digestion reaction.

COMPONENT	AMOUNT (µL)	FINAL CONCENTRATION
PLASMID DNA	10	40 ng/µL
FASTDIGEST NCOI	1	
10X FASTDIGEST BUFFER	2	1x
DDH_2O	7	
TOTAL VOLUME	20	

1 µg plasmid DNA was digested and incubated at 37°C for 30 minutes. 1 µL of FastAP thermosensitive alkaline phosphatase was added to the reaction and incubated for additional 10 mins at 37°C. Negative controls included all reaction components, except restriction enzymes.

2.7.6 Ligations to insert sgRNA cassete into digested CRISPR plasmids

CRISPR-dCas9

Generally, ligations were set up following a vector to insert DNA ratio of 1:3 and 1:6. Purified plasmid DNA constructs which were restriction digested were ligated to shuttle or cloning vectors overnight at 4°C using T4 DNA ligase (Promega) and Ligase 10x Buffer (Promega) to make up a 10 µL reaction volume.

CRISPR-cBEST

The ssDNA oligonucleotides containing the 20 bp protospacer sequences were cloned into linearised pCRISPR-cBEST using the HiFi DNA Assembly Kit (New England Biolabs) and incubated for 1h at 50°C.

Table 2. 11. **HiFi DNA Assembly to make pEM4.**

COMPONENT	AMOUNT (µL)	FINAL CONCENTRATION
LINEARISED PLASMID DNA	1 (30 ng)	1.5 ng/µL
5 UM SSDNA OLIGONUCLEOTIDE	5	1.25 µM
2X NEBUILDER HIFI DNA ASSEMBLY MASTER MIX	10	1x
DDH_2O	4	
TOTAL VOLUME	20	

2.7.7 Transformations into *E. coli* DH5α and BL21(DE3) competent cells

3 µL of ligation products were added to 50 µL Eppendorf tubes of *E.coli* Subcloning Efficiency™ DH5α Competent Cells (Invitrogen). 2.5 µL of pUC19 was added to 50 µL of competent cells as a positive control and dH_2O was used as a no DNA negative control. Tubes were placed on ice for 30 minutes, heat shocked in a 42°C water bath for 1 minute and placed on ice for another 2 minutes. 950 µL of pre-warmed LB or S.O.C media was added to each transformation and tubes were incubated at 37°C at 250 RPM for 1 hour. Transformations were spun down at 6000 RPM in a centrifuge (Eppendorf) for 1.5 minutes and the supernatant was pipetted off, leaving 100 µL. The cells were resuspended in the remaining media and transformations was plated onto an LB plate with the appropriate antibiotics and left overnight at 37°C. For *E. coli* BL21 (DE3) competent cell (Invitrogen) transformations, the same protocol was followed as for *E.coli* DH5α Competent Cells, however only 250 µL of pre-warmed S.O.C media was added to heat shocked cells.

For pGEM®-T Easy and pBGS19 cloning, transformed *E. coli* cells were plated on LB agar containing 100 µg/mL ampicillin or 50 µg/ML kanamycin, 50 µg/mL X-gal and 100 µM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Blue-white colony screening was used to select successful transformants. For plasmids without a lacZ operon such as pSET152 based plasmids and pET15b+ solely the according antibiotic (was added to LB plates and a colony PCR or Sanger Sequencing was conducted to select successful transformants.

2.7.8 Transformations into *E. coli* ET12567/pUZ8002 competent cells

2 µL of pEM2 was added to 150 µL Eppendorf tubes of *E. coli* ET12567/pUZ8002 chemically competent cells (pre-made). 2 µL of pSET152 without the dcas9 and the

spacer was used as a vector control and added to 150 µL of the competent cells. Additionally, 2 µL of pGWS1050 was added to 150 µL of competent cells as a CRISPR-dcas9 control, and 2 µL of dH_2O was added 150 µL of competent cells as a negative control. Tubes were placed on ice for 30 minutes, heat shocked in a 42°C water bath for 1 minute and placed on ice for another 2 minutes. 950 µL of pre-warmed S.O.C media was added to each transformation and tubes were incubated at 37°C at 250 RPM for 3.5 hours. Transformations were spun down at 6000 RPM in a centrifuge (Eppendorf) for 1.5 minutes and the supernatant was pipetted off, leaving 100 µL. The cells were resuspended in the remaining media and transformations were plated onto LB agar containing 50 µg/mL apramycin, 25 µg/mL kanamycin and 25 µg/mL chloramphenicol and left overnight at 37°C.

2.7.8 Sanger Sequencing

Forward and reverse primers **SPACER_SPECIFIC_SEQUENCE** and **dCAS9_R3SEQ_R** were chosen to determine correct plasmid construction for dCas9 cloning. **ADDGENE_M13/PUC_R** primer was used to screen CRISPR-cBEST constructs. Eppendorf tubes containing 15 µL of 50 ng/µL plasmid DNA plus 2 µL of 10 µM forward or reverse primers were prepared to make up a total volume of 17 µL. Pre-existing sequencing primers were used for pBGS19 and pET vector cloning checks (**M13_R_29**). Eppendorf tubes were labelled and shipped off via the Eurofins TubeSeq Sanger Sequencing service. Amplified sequence results were analysed using the SnapGene® software.

2.8 Intergenic Conjugation

E. coli ET12567/pUZ8002 containing pEM3, pSET152, pGWS1370, pCRISPR-cBEST and pEM4 was conjugated with *S. clavuligerus* DSM 738 and *S. coelicolor* M145 according to Kieser et al., (2000). *E. coli* ET12567/pUZ8002 transformed with the according plasmids was grown in LB media supplied with 50 µg/mL apramycin, 25 µg/mL kanamycin and 25 µg/mL chloramphenicol at 37°C and 250 RPM. Overnight cultures we back diluted 1:10 in 15 mL of fresh LB media and grown until an OD_{600} of 0.4-0-6 was reached. Cells were spun down at maximum RPM using a Thermo Scientific™ Megafuge and washed twice with an equal volume of LB before being resuspended in 0.1 volume of LB. 100 µL of *S. clavuligerus* spores and 5 µL of *S. coelicolor* spores and were added to separate Eppendorf tubes and gently mixed with 500 µL of 2xYT media. Tubes were

heat shocked at 50°C for 10 minutes and 500 µL of *E. coli* ET12567/pUZ8002 strains with according plasmids was added. The *S. clavuligerus* mixture was plated onto L3M9 + 10 mM $MgCl_2$ agar and incubated at 26°C for 16 hours. The *S. coelicolor* mixture was plated onto SFM + 10 mM $MgCl_2$ agar and incubated at 30°C for 16 hours. After 16 hours, all the plates were overlaid with 1 mL dH_2O mixed with 0.5 mg nalidixic acid and 1 mg of apramycin for antibiotic selection. Transconjugants were streaked onto L3M9 and SFM agar and spores were harvested.

2.9 Genomic DNA Isolation (Salting Out Method)

One colony was cut out from transconjugant plates and grown at 26°C in baffled 50 mL flasks of TSB media supplied with apramycin. The flasks were left to grow for 2-4 days at 200 RPM, split into 2x25 mL falcons and centrifuged at 4000 x g for 30 minutes. The pelleted mycelium was resuspended in 3 mL STE buffer (75 mM NaCl, 25mM EDTA pH 8, 20mM Tris-HCl pH 7.5) and 150 µL of 20mg/mL lysozyme was added and incubated at 37°C, 250RPM until clear. 84 µL of 20 mg/mL proteinase K solutions was added alongside 360 µL 10% SDS and sample was incubated at 55°C for 2-2.5h, inverting occasionally. 1200 µL of 5M NaCl was added, and sample was left to cool to 37°C or lower. 3 mL chloroform was added, and tubes were inverted and incubated at room temperature for 30 minutes. Sample was centrifuged ay 4000xg for 15 minutes and top layer was removed and transferred to a new 15 mL falcon tube. The solution was incubated on ice for 2 minutes and an equal volume of ice-cold isopropanol was added and mixed by inversion then left at -20°C overnight. The next day the DNA was spooled onto a sealed Pasteur pipette and stored in sterile Eppendorf before being rinsed in 3mL 70% ethanol and centrifuged for 10 minutes. The leftover ethanol was pipetted off and samples were left to air dry before a final resuspension in 100 µL dH_2O. The DNA concentration was measured with a NanoDrop 2000c Spectrophotometer (ThermoFisher) at 260 nm and an absorbance ratio 260/280 of >1.8 was used as a standard for pure DNA.

2.10 Agarose Gel Electrophoresis

Purple Gel Loading Dye (6x) (New England Biolabs) was added to the PCR, colony PCR and restriction digestion samples. Samples were loaded on a 3.0% (to visualise difference in fragments of 20 bp) 1.5% or 1.0% agarose gel. Either a 100 bp DNA Ladder or 1 kb Plus DNA Ladder (New England Biolabs) was used for experiments. Gel electrophoresis was performed at 65 V - 90 V for 60 to 90 minutes. UltraPure™ Ethidium Bromide, 10 mg/mL (ThermoFisher) was added to the agarose gels and visualised using the Syngene Gel documentation system InGenius.

2.11 Southern Blotting

The protocol is based on vacuum-based transference of DNA where a probe was labelled with digoxigenin (DIG) using the DIG DNA Labelling and Detection Kit (Roche Life Science) which is based on the random priming of hexanucleotides dATP, dCTP, dGTP and DIG-labelled dUTP by Klenow Enzyme. A chromogenic substrate was used for detection.

Table 2. 12. Table of solutions used for Southern Blotting.

APPLICATION	SOLUTION	COMPONENTS
DNA TRANSFER	Depurination solution	19 mL 13M HCl in 1L dH_2O
	Denaturation solution	87mg NaCL, 20mg NaOH in 1L dH_2O
	Neutralisation solution	77.8 g ammonium acetate in 1L dH_2O
	20x SSC	175.3 g NaCl, 88.2 g sodium citrate in 1L dH_2O
	2x SSC	20 x SSC diluted 10x in dH_2O
PRE-HYBRIDISATION & HYBRIDISATION	Maleic acid buffer	22.2 g Maleic acid, 17.6 g NaCl in 1L dH_2O
	10% Blocking solution	5 g blocking reagent in 50 mL maleic acid buffer
	Standard hybridization buffer (SHB)	5x SSC, 0.1% N-laurylsarcosine (w/v), 0.02% SDS (w/v), 1% blocking solution (v/v)
STRINGENCY WASHES	Stringency buffer A	2x SSC, 0.1% SDS
	Stringency buffer B	0.2x SSC, 0.1% SDS
DETECTION	Detection buffer	12.1 g Tris-HCl, 5.85 g NaCl in 1L dH_2O
	Washing buffer	0.3% (w/v) Tween-20 in maleic acid buffer
	Antibody solution	Anti-DIG-AP diluted 1:5000 in 1% blocking solution
	Colour-substrate solution	200 µL of NBT/BCIP, 10 mL detection buffer

2.11.1 Probe Labelling

The probe and ladder were labelled using random primed DNA labelling, using 15 µL containing 1-2 µg the PCR amplified and purified target sequence/ladder. DNA was

boiled at 95°C for 10 minutes, followed by incubation on ice for 10 minutes. The reaction was set up according to **Table 12**, mixed and incubated overnight. The reaction was stopped by adding 2 µL 0.2 M EDTA (pH 8).

Table 2. 13. Random primed DNA labelling reaction set-up.

COMPONENT	AMOUNT (µL)
DENATURED DNA	15
HEXANUCLEOTIDE MIX 10X	2
DIG DNA LABELLING MIX	2
KLENOW ENZYME	1

2.11.2 Gel Electrophoresis

2.5 µg of gDNA (*S. clavuligerus*) was digested with 2.5 µL of SacI (New England Biolabs) according to manufacturer's instructions and left overnight at 37°C. DNA was loaded onto a 0.7% agarose gel and left to run for 3-4h at 60 V. Lambda digested with *Hind*III and diluted 10x was used as a size marker.

2.11.2 DNA transfer to nylon membrane

Three sheets of Whatman paper (largest layer) were cut and soaked in 2x SSC buffer and placed as the first layer onto vaccum apparatus. One sheet of Hybond N+ nylon membrane (GE Healthcare) was cut slightly smaller than Whatman paper and placed on top followed by the plastic window, cut smaller than the membranes and gel. The transfer lid was clipped onto vacuum apparatus and the gel was placed on top of the plastic window in the centre of the tray. 50 mL depurination buffer was pipetted onto the gel using a stripette gun (ThermoFisher) and a pressure of 50 mbar was applied and left for 20 minutes. The same process was repeated with 50 mL denaturation buffer and 50 mL neutralisation buffer. 50 mL of 20x SSC was added onto gel and the vacuum was applied for 1-2 hours. Finally, the membrane was exposed to UV for 2 minutes in a UV crosslinker (UVP) and washed in 2xSSC for 2 minutes.

2.11.3 Hybridisation

The membrane was wrapped in mesh and inserted into Hybaid tube. 20 mL preheated (65°C) SHB was added, and tube was left to rotate in hybridisation oven at 65°C for 1 hour. Probes were denatured at 100°C for 15 minutes, then left on ice for 10 minutes. The old prehybridization SHB was decanted and replaced with 20 mL SHB plus denatured probes and left to hybridise overnight at 65°C.

2.11.3 Stringency Washes

The membrane was incubated twice with 150 mL Stringency Buffer A in Hybaid tube at 65°C for 15 minutes. The process was repeated with 150 mL Stringency Buffer B.

2.11.3 Detection

The membrane was rinsed briefly with washing buffer followed by a 30-minute wash with 100 mL freshly prepared blocking solution at room temperature. The blocking solution was replaced with 30 mL antibody solution and incubated for another 30 minutes in the opened hybridisation oven. The antibody solution was decanted and replaced with 100 mL washing buffer and washed twice for 15 minutes to remove unbound antibody. The membrane was removed from mesh and washed with 20 mL detection buffer in a semi-sealed plastic bag for 2 minutes. The detection buffer was decanted, and 10 mL of freshly prepared colour-substrate solution was added to membrane, sealed, and left in the dark until colour developed. Once colour was detected, the membrane was washed with tap water.

2.12 Conjugation Efficiency Optimisation

2.12.1 Conjugation Efficiency and Frequency using frozen spores

S. clavuligerus DSM 738 and *S. coelicolor* M145 spores were counted by performing serial dilutions of harvested spores of which the 10^{-7} and 10^{-8} dilutions were plated and counted, respectively. The experiments were performed in triplicate and the number of spores/mL was calculated using the following formula:

Spores/mL = average number of colonies x dilution x 10^2

The conjugation efficiency was calculated based on the number of recipient cells. To calculate the conjugation based on this, the number of recipient spores was altered using a range of 10^5, 10^6, 10^7, 10^8, 10^9 and 10^{10} spores and the number of donor cells was kept constant. 3.5 x 10^5 donor *E. coli* ET12567/pUZ8002 cells, transformed with pSET152 or pGWS1370, were used for each conjugation. Conjugations were performed as mentioned in **Section 2.8**, however were measured in triplicate, and diluted to 10^{-2} before plating for *S. coelicolor* experiments. *S. clavuligerus* conjugations were not diluted before plating. The number of colonies were counted, and the average colony number was divided by the number of recipient spores added, giving an efficiency. The conjugation frequency was calculated as a percentage of the efficiency.

2.12.2 Conjugation Efficiency and Frequency using fresh spores

The conjugation efficiency experiments were conducted using fresh spores by harvesting one half, one or two plates of spores, heat shocking them at 50 °C and adding them to 500 µL of *E. coli* ET12567/pUZ8002 cells grown to an OD of 0.5 (roughly 4 x 10^8 cells). The number of spores was quantified from the corresponding number of plates by harvesting one plate of spores using 5mL of 20% glycerol (v/v), spinning down the spore suspension and resuspending in 1 mL of glycerol to allow for subsequent dilution and storage. Other than that, conjugations were performed as mentioned in Section 2.5, however were measured in triplicate for each plate number of spores added. The number of colonies were analysed, and conjugation efficiencies and frequencies were calculated based on the number of spores added.

2.13 Tap Protein overexpression

Table 2. 14. Solutions to prepare for protein overexpression.

SOLUTION	COMPONENTS
5 X SDS LOADING DYE	50 mM Tris HCl pH 6.8 2% SDS 10% Glycerol 1% ß-mercaptoethanol 12.5 mM EDTA 0.02% Bromophenol Blue
PMSF SOLUTION	100 mM PMSF in isopropanol
SDS 10 X RUNNING BUFFER	14.4 % Glycine 3% Tris-Base, pH 8 1% SDS Up to 1 L with deionised water, pH 8.3
COOMASSIE BLUE STAINING SOLUTION	0.1% Coomassie R250 10% Acetic Acid 40% Methanol
DE-STAINING SOLUTION	30% Methanol (v/v) 10% Acetic Acid (v/v)

2.13.1 His-Tag Protein Expression using Autoinduction Media

Colonies from fresh *E.coli* BL21 (DE3) transformations containing Tap pET15b+ overexpression plasmids were picked and inoculated in 10 mL overnights containing LB with 100 µg/mL of ampicillin. *E. coli* BL21 (DE3) colonies containing only pET15b+ vector were used as the negative control. 1 mL aliquot of overnight was spun down and used as an uninduced control. 50 mL ZY-5052 autoinduction media was inoculated in Erlenmeyer 250 mL flasks with a 500 µL aliquot of the BL21 overnight cultures and grown at 37°C, 30°C and 26°C for 20, 22, 26 and 24 hours at 250 RPM. Cultures were pelleted by centrifugation at 4000 x g for 20 minutes. Pellets were weighed and resuspended in BugBuster® Master Mix (MM) (Merck Millipore Novagen™) at 5mL BugBuster® MM per mg pellet concentration. The serine protease inhibitor PMSF was added to a concentration of 1mM and the BugBuster® MM was incubated at room temperature until the lysate was clear. Lysate was centrifuged at max RPM for 20 minutes at 4°C to remove cell debris and supernatant was retained for SDS-PAGE analysis.

2.13.2 His-Tag Protein Expression using IPTG

A) Colonies from fresh *E.coli* BL21 (DE3) transformations containing Tap pET15b+ overexpression plasmids were picked and inoculated in 10 mL overnights containing LB

with 100 µg/mL of ampicillin. *E. coli* BL21 (DE3) colonies containing only pET15b+ vector were used as the positive control and uninduced (no IPTG) cultures were used as the negative control. Overnight cultures were diluted back to an OD_{600} of 0.05 and grown until they reached OD_{600} 0.4-0.6 in 50mL Erlenmeyer flasks with LB and 100 µg/mL of ampicillin. Once the correct OD_{600} was reached, 0.1, 0.2, 0.3 and 0.5 mM of IPTG was added to the flasks. Flasks grown at 37 °C were grown for 4 hours following induction with IPTG. Flasks grown at 18 °C and 26 °C were grown overnight at 250 RPM following induction with IPTG.

B) Entire transformation plates from fresh *E.coli* BL21 (DE3) transformations containing Tap pET15b+ overexpression plasmids were harvested using LB with 100 µg/mL ampicillin. 5 mL of LB was pipetted onto each transformation plate and colonies were gently mixed into the LB using a plastic spreader. Another 5 mL of LB plus ampicillin was pipetted onto the plates and mixed using the spreader. The LB was pipetted off the plate and into a 50 mL Erlenmeyer flask with fresh LB and 100 µg/mL ampicillin. The culture was grown to an OD_{600} of 0.7 and IPTG was added at 0.1, 0.2, 0.3 mM concentrations and grown overnight at 18 °C at 250 RPM.

Cultures (A) and (B) were pelleted by centrifugation at 4000 x g for 20 minutes. Pellets were weighed and resuspended in BugBuster® MM at 5 mL BugBuster® MM per mg pellet concentration. The serine protease inhibitor PMSF was added to a concentration of 1mM and the BugBuster® MM mixture was incubated at room temperature until the lysate was clear. Lysate was centrifuged at 4200 x g for 20 minutes at 4°C to remove cell debris and supernatant was retained for quantification using a Bradford Protein Assay and SDS-PAGE analysis.

2.13.3 Inclusion body solubilisation

Pellets were weighed and resuspended in BugBuster® MM. To determine presence of absence of inclusion bodies in the IPTG-induced overexpression cultures grown at 18 °C, the cultures were spun down at 4200 x g for 20 minutes at 4°C and the pellet was retained and washed three more times with 1:10 diluted BugBuster® MM. Finally, the pellet was resuspended in 1:10 diluted BugBuster® MM and quantified using a Bradford Protein Assay. The remaining resuspended pellet was boiled at 95 °C in 5x SDS sample loading dye and stored at -20°C for SDS-PAGE analysis.

2.13.4 SDS-PAGE & Coomassie Blue Staining

Protein lysate samples were mixed with 5x SDS dye to make a 1x working solution and 20 µg per well was loaded onto NovexTM 4-20% Tris-Glycine Mini-Gels (ThermoFisher Scientific™). Gels were run in 1x running buffer for 60 minutes at 140 V. SDS-PAGE gels were stained in Coomassie Blue for 1 hour, de-stained for 30 minutes at room temperature and left in water overnight on a shaking-platform to be imaged with the Azure™ biosystem C200.

2.14 Western Blot using Chemiluminescence

Table 2. 15. Solutions to prepare for Western Blotting.

SOLUTION	COMPONENTS
10 X TBS STOCK SOLUTION	200 mM Tris Base 1500 mM NaCl pH 7.6
1 X TBS-T (WASH BUFFER)	100 mL 10 x TBS solution 900 mL dH_2O 1 mL Tween-20
TRANSFER BUFFER	1.44% Glycine 0.3% Tris-HCl 0.1% SDS 20 mL methanol Up to 1 L with dH_2O
BLOCKING SOLUTION	5 g Non-fat dry milk 100 mL 1x TBS-T

2.14.1 Protein Transfer, Wash and Antibody Treatment

The nitrocellulose transfer membrane was prepared according to BioRad BlotModule II instructions using a wet-sandwich transfer system and run at 25V, 100 mA for 1.45 hours. Following transfer, membrane was washed 4 times in deionised water for 5 minutes each time to remove residual transfer buffer. The membrane was incubated for one hour in blocking solution at room temperature. All membrane incubations and washed were set up with agitation. Primary 6x Anti-His-Mouse antibody (ThermoFisher) was diluted 1:1,000 in the blocking solution. Following the blocking, the membrane was incubated with the primary antibody in blocking solution overnight at 4 °C. The next day, the nitrocellulose membrane was washed 3x for 10 minutes in wash buffer. The membrane was incubated in a 1:50,000 dilution of the secondary Anti-Mouse-Goat-HRP antibody

(ThermoFisher) in blocking solution for one hour at room temperature and washed 6 times with washing buffer for 5 minutes each to remove unbound secondary antibody.

2.14.2 Chemiluminescent detection

A 1:1 working solution of the SuperSignal West Pico Plus (ThermoFisher) chemiluminescent substrate was pipetted into a square plate and the membrane was incubated in the substrate for 4 minutes. The blot was removed, drained, and kept in a plastic seal for imaging with the ChemiDocMP Imaging System (BioRad).

2.15 Electrophoretic Mobility Shift Assay

200 bp ssDNA sequences (5' to 3') of the left-hand-side (LHS) and right-hand-side (RHS) *S. clavuligerus* telomeres were synthesised and ordered (Azenta). 300 nM ssDNA was incubated on ice in binding buffer (1 mM EDTA, 20 mM Tris HCl pH 8, 5 mM $MgCl_2$, 0.5 mM DTT, 5% (v/v) glycerol, 0.5 µg/ml polydI-dC) without and with in- creasing amounts of crude Tap4 protein lysate in a total volume of 16 µL. After mixing, samples were incubated for 20 min at 30°C, followed by a 10 min incubation on ice, and then loaded onto 2.5% agarose gel in Novex™ 0.5x TBE Running Buffer. Agarose gel electrophoresis was carried out in 0.5x TBE Running Buffer at 50 V at 4°C for a period of 3 hours. Finally, the gel was stained with a 0.5 µg/ml working stock of ethidium bromide for 30 minutes, de-stained in 0.5x TBE Running Buffer for 1 hour and imaged with the Azure™ biosystem C200 imaging doc.

2.16 Whole genome sequencing of *S. clavuligerus* mutant strains

2.16.1 MicrobesNg HiSeq Illumina short read sequencing of *S. clavuligerus* mutant strains

S. clavuligerus dCas9 mutants, controls and DSM 738 strains were sent to MicrobesNG in Birmingham for short-read Illumina sequencing. Genomic DNA was quantified using the NanoDrop 2000c Spectrophotometer (Thermo Fischer Scientific) and diluted with nuclease-free dH_2O to a concentration of 80 ng/µL to a total volume of 100 µL DNA integrity was check by running 1:10 diluted samples on a 0.8% agarose gel and performing electrophoresis.

DNA samples were sequenced using 2x250 bp paired end reads with 30x depth coverage on lllumina NovaSeq 6000 (Illumina, San Diego, USA). Genomic DNA libraries were made using the Nextera XT Library Prep Kit (Illumina, San Diego, USA) and following the manufacturer's protocol. DNA quantification and library preparation were carried out with a Hamilton Microlab STAR automated liquid handling system (Hamilton Bonaduz AG, Switzerland).

Reads were adapter trimmed using Trimmomatic version 0.30 with a quality cut-off of Q15 (Bolger, Lohse and Usadel, 2014). De novo assembly was performed on samples using SPAdes version 3.7 (Bankevich *et al.*, 2012), and contigs are annotated using Prokka (Seemann, 2014).

2.16.2 Bioinformatic analysis of the *S. clavuligerus* mutant strains

S. clavuligerus dCas9 mutant analysis was performed on Galaxy Australia by comparing the trimmed, paired-end reads to the closed genome reference sequence obtained from NCBI: GCA_028752555.1 (Algora-Gallardo). The reference sequence was made up of five replicons representing the *S. clavuligerus* chromosome CP086229, pSCL4 CP086230, pSCL3 CP086231, pSCL2 CP086232 and pSCL1 CP086233. Bowtie2 (Version 2.5.0) (Langmead and Salzberg, 2012) and BWA-MEM2 (Burrows-Wheeler Aligner) Version 2.2.1 (Vasimuddin *et al.*, 2019) were used to map the reads against the *S. clavuligerus* reference genome with output files being in binary alignments (BAM) format with a binary alignments index. Additionally, BAM files were converted into BIGWIG files using bamCoverage (Galaxy EU). Genome assembly quality was assessed with QUAST (Gurevich *et al.*, 2013) and mapped reads were visualised using Integrative Genomics Viewer version 2.12.3 (Robinson *et al.*, 2011) and JBrowse2 (Diesh *et al.*,

2022). Qualimap V.2.2.1 (Okonechnikov, Conesa and García-Alcalde, 2016) was used to assess mutant genome quality and mean coverage. Qualimap was used from the terminal using the command:

cd qualipmap_v2.2.1\ 3/
./ qualimap

A nucleotide BLAST database was created using the command:

makeblastdb -in .fna -out -dbtype nucl

Illumina sequences of interest of mutant strains in this study were then aligned to the database and screened for sequence similarity using the command:

blastn -db -query .fna -out .txt

2.17.2 Bioinformatic analysis of the *S. clavuligerus* terminal proteins

Phylogenetic trees were constructed on ClustalW (Thompson, Higgins and Gibson, 1994) and MEGA11 (Tamura, Stecher and Kumar, 2021) with 100 bootstraps and edited on iTOL (Letunic and Bork, 2021). Tap and Tpg amino acid alignments were made in Uniprot using Clustal Omega (Sievers et al., 2011). *S. clavuligerus* DSM 738 (NCBI: GCA_028752555.1) Tap and Tpg protein physical and chemical parameters were calculated using the ExPASy tool on the Swiss Institute of Bioinformatics server (Gasteiger *et al.*, 2003) and 3D models of the terminal proteins were computed using Alphafold2 (Jumper *et al.*, 2021). The .pdb files of the predicted protein structures produced by Alphafold were then run though PDBe Fold v2.59, which was used to obtain a more detailed overview of the similarities and differences between the proteins of interest (Krissinel and Henrick, 2004). Lastly, the PSIPRED v4.0 workbench was used to predict secondary structures of *S. clavuligerus* DSM 738 Tap and Tpg proteins (McGuffin, Bryson and Jones, 2000).

3.0 Genetic Manipulation of *Streptomyces clavuligerus* using CRISPR tools.

Genetic manipulation of Streptomycetota has become an important method of improving strains to increase secondary metabolite yield. There is a lack of molecular tools for DNA manipulation of streptomycetes. Additionally, members of the genus *Streptomyces*, have a high G-C content making expression of exogenous DNA difficult. Despite these limitations, popular genetic tools, such as CRISPR-based gene editing and RNA interference (RNAi), have shown successes in *Streptomyces* (Baltz, 2016). In this chapter, we demonstrated our first aim and utilised these toolkits to target the genome of *S. clavuligerus*, with the goal of creating strains with increased fermenter stability and clavulanic acid production. In particular, we aimed to determine replicon essentiality by optimising CRISPRi methods, to elucidate the significance and relationship of *tap-tpg*$_4$ in the genome dynamics of *S. clavuligerus.*

3.1 Comparison and selection of CRISPRi-based toolkits.

CRISPR-Cas9 toolkits have changed methods of streptomycete genome engineering, avoiding time-consuming traditional genome editing protocols, such as allelic replacement through homologous recombination (Alberti and Corre, 2019). Since their discovery, the main goal in *Streptomyces* genome engineering has been to enhance the biosynthesis of natural products aiming to improve yields. Successful CRISPR-Cas9 genome editing methods have been developed in several labs worldwide mainly for use in *S. coelicolor*, however further species such as *S. lividans*, *Streptomyces albus* and *Streptomyces viridochromogenes* have also been studied (Cobb, Wang and Zhao, 2015; Huang *et al.*, 2015; Tong *et al.*, 2015; Zeng *et al.*, 2015). Plasmids such as pCRISPomyces-1 and 2 (Cobb *et al.,* 2015), pKCCas9dO (Huang *et al.,* 2015), pCRISPR-Cas9 / pCRISPR-dCas9 (Tong *et al.,* 2015), and pWHU2653 (Zeng *et al.,* 2015) are now widely available for use, functioning with seperate trans-activating CRISPR RNA (tracrRNA) and CRISPR RNA (crRNA) or a tracrRNA/crRNA fusion molecule called sgRNA. An overview of recently applied CRISPR-based toolkits is described in **Table 3.1**, highlighting the individual genetic control mechanisms of the required plasmids (promoters, selectable markers, plasmid clearance) and target organisms.

Table 3. 1. Overview of recent CRISPR-based toolkits for in vivo genetic engineering of streptomycete genomes.

	pCRISPR-dCas9	***pCRISPR-BEST***	***pCRISPomyces-1***	***pCRISPomyces-2***	***pKCCas9dO***	***pCRISPR-Cas9***	***pWHU2653***
Selectable marker	Apramycin	Apramycin	Apramycin	Apramycin	Apramycin	Apramycin	Apramycin
Promoter for* Cas9 *expression	Constitutive: *gapdh*	Inducible: *tipAp*	Constitutive: *rpsLp*	Constitutive: *rpsLp*	Inducible: *tipAp*	Inducible: *tipAp*	Constitutive: *aac(3)IVp*
Promoter for sgRNA expression	Constitutive: *ermE*p*	Constitutive: *ermE*p*	Constitutive: *gapdhp* for crRNA and rpsLp for tracrRNA	Constitutive: *gapdhp* for	Constitutive: *j23199p*	Constitutive: *ermE*p*	Constitutive: *ermE*p*
Plasmid clearance	Integrative	Temperature-sensitive rep origin from pSG5	Temperature-sensitive rep origin from pSG5	Temperature-sensitive rep origin from pSG5	Temperature-sensitive rep origin from pSG5	Temperature-sensitive rep origin from pSG5	CodA counter-selectable marker
Test species	*S. coelicolor A3(2)*	*S. coelicolor*	*S. lividans*	*S. lividans, S. vidriochromogens, S. albus, S. formicae*	*S. coelicolor* M145,	*S. coelicolor A3(2)*	*S. coelicolor* M145
Reference	(Qi *et al.,* 2013), This study	(Tong *et al.,* 2019)	(Cobb, Wang and Zhao, 2015)	(Cobb, Wang and Zhao, 2015)	(Huang *et al.,* 2015)	(Tong *et al.,* 2015)	(Zeng *et al.,* 2015)

Adapted from Alberti and Corre, (2019).

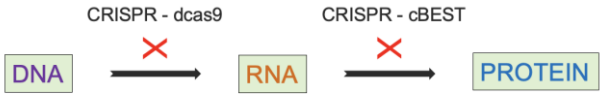
The two systems investgated in this chapter were CRISPR – dCas9 and CRISPR-BEST which, after conjugation with the recipient organism, either interfere with transcription, blocking the RNA polymerase or translation, which leads to a truncated protein. CRISPR-dCas9 is a type II CRIPSR system from *Streptococcus pyogenes* with a mutated catalytic domain preventing it from cleaving the target sequence. Instead, it binds and blocks transcription and requires a spacer and guide RNA (gRNA) to function, also called sgRNA. The sgRNA is of particular interest when looking at organisms with diverse and difficult genomes, as it is bound to a target sequence which is complementary to a 20-base pair (bp) spacer sequence (found as part of or adjacent to the gRNA), stopping transcription and consequently acting as a gene silencer, due to the lack of endonuclease activity (Qi *et al.*, 2013) (**Figure 3.1 A**).

CRISPR-dCas9 avoids the main drawback of regular Cas9 systems, as it does not cause double stranded breaks (DSBs) (Zeng *et al.*, 2015; Tong *et al.*, 2019). pCRISPR-dCas9 is based on pSET152, which uses a phiC31 integrase for plasmid integration into the host genome after conjugation with the organism of interest. More specifically, the delivery plasmid (pEM3 in this chapter) contains an *attP* site, which recombines with the *attB* site in the host genome to form *attL* and *attR* sites (Combes *et al.,* 2002), allowing for integration and transcriptional blockage as shown by **Figure 3.1 B**.

The most recent *Streptomyces* base editing tool also functions without DSBs and is referred to as CRISPR-BEST (Tong *et al.,* 2019). CRISPR-BEST can be applied to convert a cytidine (C):guanosine (G) bp to thymidine (T):adenosine (A) or A:T bp to G:C using a deaminase based base editor, allowing for gene inactivation. More specifically, C:T conversion turns a CAG codon to a TAG codon. In the case of such a nonsense mutation, the mRNA contains a new UAG stop codon. The study by Tong *et al.,* (2019) evaluated the efficiency and off-target effects for CRISPR-cBEST systems by targeting the ActIORF1 gene for actinorhodin biosynthesis in *S. coelicolor*, observing phenotypic and genotypic changes. CRISPR-cBEST has a larger editing window than CRISPR-aBEST, due to the frequency of G or C bases in *Streptomyces* (Tong *et al.,* 2019). Therefore, the CRISPR-cBEST system was also tested in a non-model organism, *Streptomyces griseofuscus*, successfully introducing a stop codon and avoiding any polar effect.

All of the mentioned CRISPR-Cas9 based *Streptomyces* gene editors (**Table 3.1**) have shown success in genome editing in different *Streptomyces spp*.. More often, the systems showed higher efficiency and more data exists for model organisms like *S. coelicolor* (Alberti and Corre, 2019). *S. clavuligerus* DSM 738 contains a highly dynamic genome with a four GLPs and is an industrially relevant strain. CRISPR-Cas9 based systems have shown successes in this organism however, these DSBs have resulted in the loss of entire plasmids, and targeted gene knockouts present difficulties during transconjugant screening (Gomez-Escribano *et al.,* 2021). Thus, CRISPR editing systems that did not result in DSBs, like CRISPR-dCas9 and CRISPR-BEST were selected and applied to our instable organism of interest: *S. clavuligerus* DSM 738.

A



B

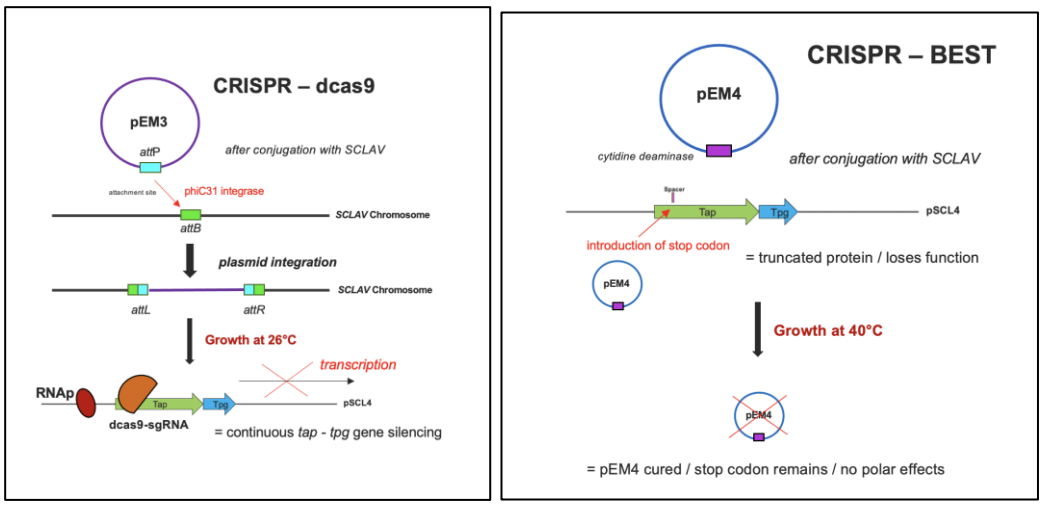


Figure 3. 1. CRISPR-dCas9 and CRISPR-BEST block transcription and translation respectively. A) Graphical representation of transcriptional and translational interference. **B)** Chromosomal integration and transcriptional blockage after conjugation with *S. clavuligerus* and CRISPR-cBEST plasmid used in this study targeting *tap-tpg*$_4$. For CRISPR-dCas9, the spacer binding site is highlighted by the orange dCas9 protein. For CRISPR-BEST, the spacer binding site is labelled on the *tap*$_4$ gene. pEM3 = CRISPR-dCas9 plasmid used in this study targeting *tap-tpg*$_4$ and pEM4 = CRISPR-cBEST plasmid targeting *tap-tpg*$_4$.

3.2 Construction of a CRISPR-dCas9 containing vector targeting *tap-tpg*$_{4.}$

Tap and Tpg play a vital role in the maintenance of *Streptomyces* linear replicons through end patching. The importance of *tap-tpg*$_{4,}$ *tap-tpg*$_{3}$ and *tap-tpg*$_{2}$ in the maintenance of plasmid and chromosomal linearity in *S. clavuligerus* has not yet been established, however previous work highlighted that cutting the largest GLP, pSCL4, resulted in plasmid loss and chromosome circularisation, showing that genes on pSCL4 are necessary for chromosome linearity (**Figure 1.7**) (Algora-Gallardo, unpublished data). To determine whether chromosome circularisation was due to the loss of *tap-tpg* on pSCL4, the significance of *tap-tpg*$_{4}$ in telomeric replication was investigated through a CRISPR-dCas9 knockdown vector, targeting the genes.

3.2.1 Cloning overview to create the integrative, *tap-tpg*$_{4}$ targeting plasmid, pEM3.

In order to construct a plasmid to silence *tap-tpg*$_{4}$ expression, the CRISPR-dCas9 vector included three components: an encoded dCas9, a gRNA and a 20 bp spacer targeting *tap*$_{4}$ in *S. clavuligerus* allowing for transcriptional silencing. A CRISPR-dCas9 knockdown vector including all three components was created through a multi-step restriction enzyme cloning process, resulting in pEM3 (with a pSET152 backbone). An overview is highlighted in **Figure 3.2 A.**

PCR was used to introduce *tap*$_{4}$ spacer into the gRNA scaffold. This created a spacer and gRNA complex, which targeted *tap*$_{4}$ in *S. clavuligerus* through an RNA guided dCas9 (sgRNA) and a 20 bp spacer sequence which was introduced into the 76 bp gRNA scaffold, already present on the pGWS1370 vector (van Wezel, 2019). The spacer sequence was complementary to *tap*$_{4}$ on pSCL4 of *S. clavuligerus* DSM 738, and was also expected to target *tpg*$_{4,}$ as *tap* and *tpg* are predicted to be part of the same operon. Additionally, the spacer was flanked by a protospacer adjacent motif (PAM) sequence: NGG, which served as a binding signal for dCas9 and consisted of the codons TGG. **sgRNA_F-primerOLD** and **sgRNA_R** primers were used to amplify the 183 bp sequence using pGWS1370 DNA, confirmed in **Figure 3.2 B.**

We used restriction enzyme cloning to construct a CRIPSR-dCas9 vector with all three essential components: dCas9, gRNA and spacer, and the PCR amplified sgRNA sequence shown in **Figure 3.2. B** was cloned into pGWS1370, making the integrative

vector pEM3 (**Figure 3.2 C**). To achieve this, the sgRNA was initially cloned into pGEM®-T Easy and both the pGEM®-T Easy vector plus the sgRNA and pGWS1370 were digested with BamHI and NcoRI and subsequently ligated. Therefore, we created a plasmid with all the components necessary for *tap*$_4$ and *tpg*$_4$ silencing.

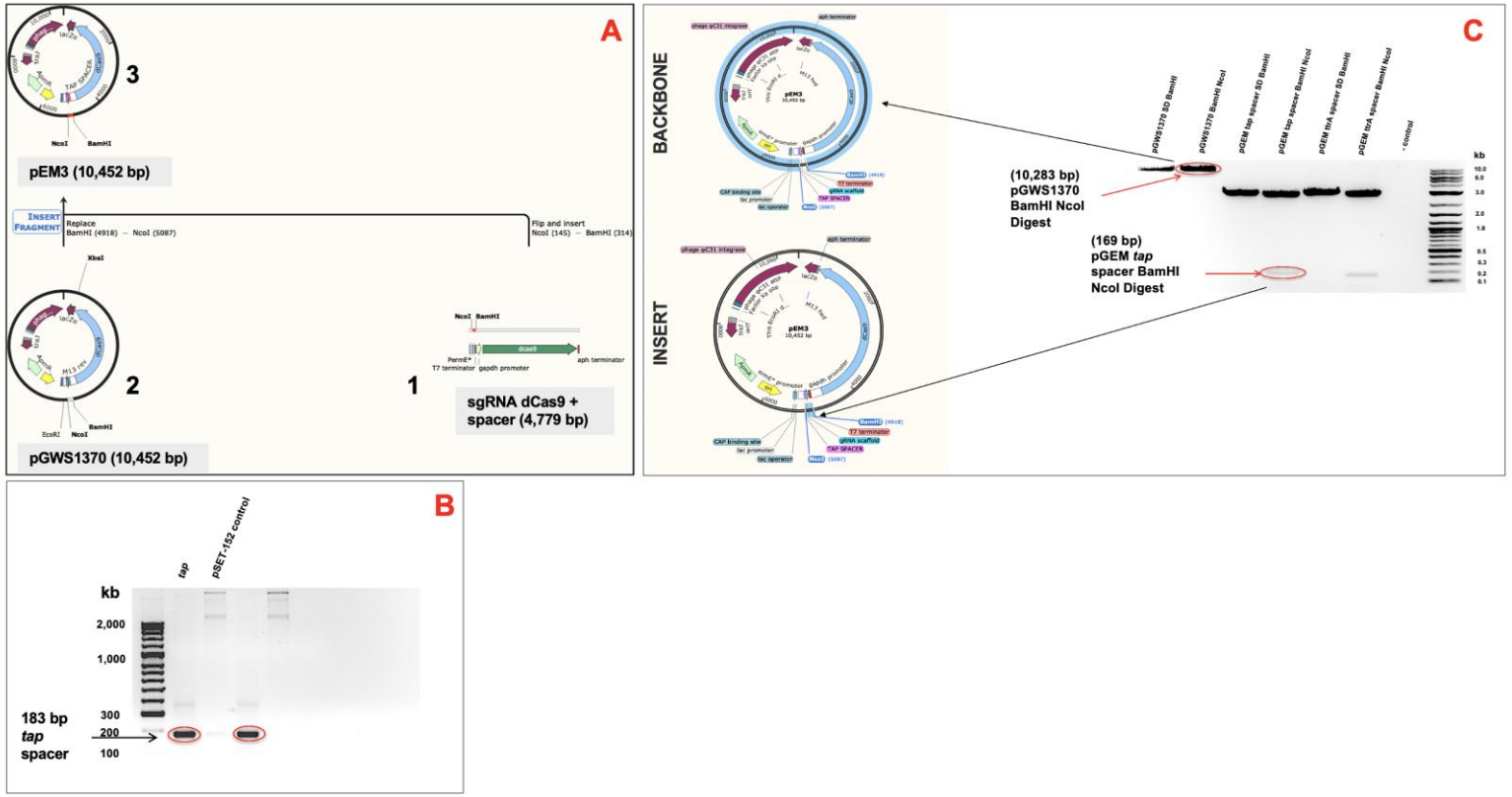


Figure 3.2. Overview of the multi-step cloning process involved in the construction of pEM3 for the silencing of *tap-tpg*$_4$. A) 1) sgRNA was amplified using PCR and cloned into pGEM®-T Easy allowing for storage. 2) Both pGWS1370 and pGEM®-T Easy plus the sgRNA were digested with NcoI and BamHI, linearising the vector backbone and providing compatible ends for insert. 3) sgRNA fragment was ligated with pGWS1370 to create pEM3. **B)** Agarose gel showing the amplification of the 183 bp sgRNA with *tap*$_4$ specific spacer sequence. Forward and reverse primers amplified a 183 bp sequence including the *tap*$_4$ specific spacer sequence, circled in red. pSET152 DNA was used as the negative control in the PCR and showed a faint band, as it did not contain the sgRNA, preventing the forward primer from annealing but still presenting sequence similarities with the reverse primer. **C)** Restriction enzyme digestion of the pGWS1370 backbone and pGEM®-T Easy / sgRNA insert produced 10,283bp and 169bp fragments used for ligation of pEM3.

3.3 Sanger sequencing confirms *tap-tpg*$_4$ specific spacer.

Confirmation of the genetic makeup of pEM3 was accomplished by verifying the presence or absence of the spacer sequence through Sanger sequencing. **M13_R** primer was used to sequence roughly 1kb of pEM3. Sequencing results were analysed on SnapGene and assembled via the CAP3 Sequence Assembly Program. Results showed that the 20 bp spacer sequence was present in pEM3 as highlighted by **Figure 3.3.** Once the genetic makeup of pEM3 was verified, we aimed to conjugate the plasmid with *S. clavuligerus* to determine how *tap-tpg*$_4$ silencing will affect replicon linearity and maintenance. Conjugations had not been optimised for *S. clavuligerus* using the vectors in this chapter. Therefore, there was a need to optimise conjugations and determine the best method to achieve the highest conjugation efficiency.

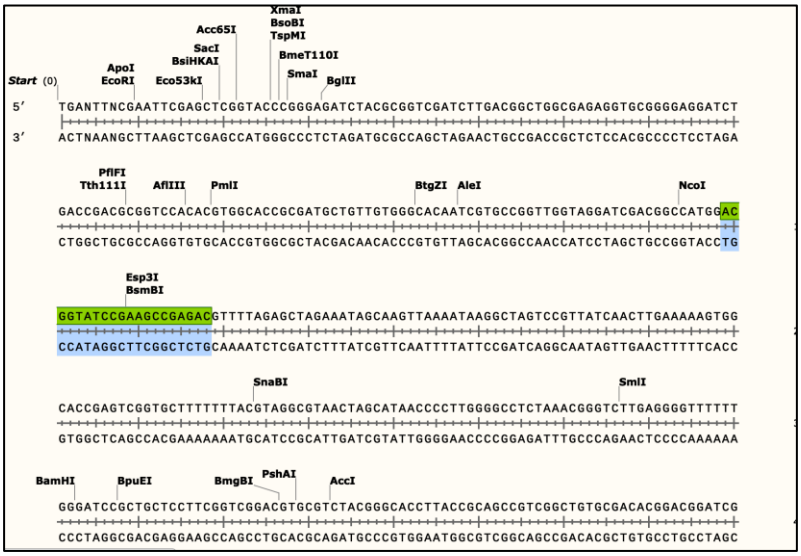


Figure 3. 3. Sanger sequencing confirms cloning of 20 bp spacer sequence. M13_R primer was used to amplify around 1 kb of pEM3, confirming presence of spacer sequence.

3.4 Optimising conjugation efficiencies for integrative vector controls: pSET152 and pGWS1370.

Following on from section 3.3 in this chapter, we aimed to introduce pEM3 into *S. clavuligerus* through conjugation to allow *tap₄* gene silencing. Conjugation with *S. clavuligerus* DSM 738, had previously been more difficult than with model organisms, like *S. coelicolor,* and required optimisation. pSET-152 is commonly used as an integrating vector for heterologous cloning and expression of genes in Actinobacteria, such as in *Micromonospora spp.* and *Streptomyces spp.* (Li, Zhou and Deng, 2003; Sun *et al.,* 2014). pSET-152 was a favourable vector in this chapter, as it presented higher conjugation efficiency, and was considered more stable than temperature-sensitive replicons depicted earlier (**Table 3.1**).

Higher efficiencies were needed in an organism like *S. clavuligerus*, whose dynamic genome can lead to an increased frequency of random homologous mutations and genetic instability (Algora-Gallardo, unpublished data). In this chapter we also utilised the dCas9-based vector pGWS1370. To increase the number of transconjugants, and consequently knockdown-mutants using pEM3, both pSET-152 and pGWS1370 vectors were chosen with the aim to find optimal conjugation conditions for both vectors with *S. clavuligerus*. We chose to focus on the number of recipient cells as our independent variable, and, by adding 10-fold increasing values of recipient spores, measured the number of transconjugant colonies and conjugation efficiency. Altering the number of recipient spores has been a common method to calculate conjugation efficiencies (Wang and Jin, 2014; Zhang *et al.,* 2019).

Initial experiments used previously frozen *S. coelicolor* M145 recipient spores as an experimental control and point of comparison to *S. clavuligerus*. 10^5 to 10^{10} *S. coelicolor* and *S. clavuligerus* recipient spores were added to an unchanged number of *E. coli* ET12567/pUZ8002 donor cells grown to OD_{600} 0.5 ($3.5x10^6$ cfu). For conjugation optimisation experiments using fresh spores, ½, 1 and 2 plates of recipient spores were used.

3.4.1 Increasing number of recipient vs donor cells in *S. coelicolor* conjugations leads to an increase in number of transconjugants and decrease in conjugation efficiency.

S. coelicolor M145 spores were used as an experimental control, as well as to investigate how the number of transconjugants and conjugation efficiency differed from our organism of interest, *S. clavuligerus*. **Figure 3.4.** highlights the difference between *S. coelicolor* number of transconjugants and conjugation efficiency, with the addition on increasing recipient spores. For both **Figure 3.4 A** and **3.4 B** an overall increase in the number transconjugants was seen when more recipient spores were added to the conjugations, regardless of whether pSET-152 or pGWS1370 was used as the integrating vector. Conversely, an overall decrease in conjugation efficiency was observed when more recipient spores were added to the conjugations. For SCO-pSET transconjugants the conjugation efficiency after the addition of 10^5 recipient spores was $2.6x10^{-3}$ and decreased to $9.1x10^{-7}$ with the addition 10^{10} recipient spores. For SCO-1370 transconjugants the conjugation efficiency after the addition of 10^5 recipient spores was $4.9x10^{-4}$ and decreased to $2.2x10^{-7}$ with the addition 10^{10} recipient spores (**Table 3.2**).

Our aim in this chapter was to obtain the maximum number of transconjugants after conjugation with pEM3. Thus, the number of transconjugants obtained after the addition of variable recipient spores presented a much more relevant result, regardless of the conjugation efficiency. **Figure 3.4 A** highlights that the number of transconjugants for SCO-pSET152 conjugations increased from 200 to 9000 when 10^5 vs 10^{10} recipient spores were added. A noticeable outlier was the slight decrease in number of colonies when adding 10^8 vs 10^9 spores, with 10^8 spores being the standard number added to *S. coelicolor* conjugations (Kieser, T *et al.,* 2000), suggesting that this value needs to be tailored to the organism of interest, as we did in this Chapter. Finally, the number of transconjugants for SCO-1370 conjugations increased from 49 to 2179 when 10^5 vs 10^{10} recipient spores were added (see **Table 3.2.**).

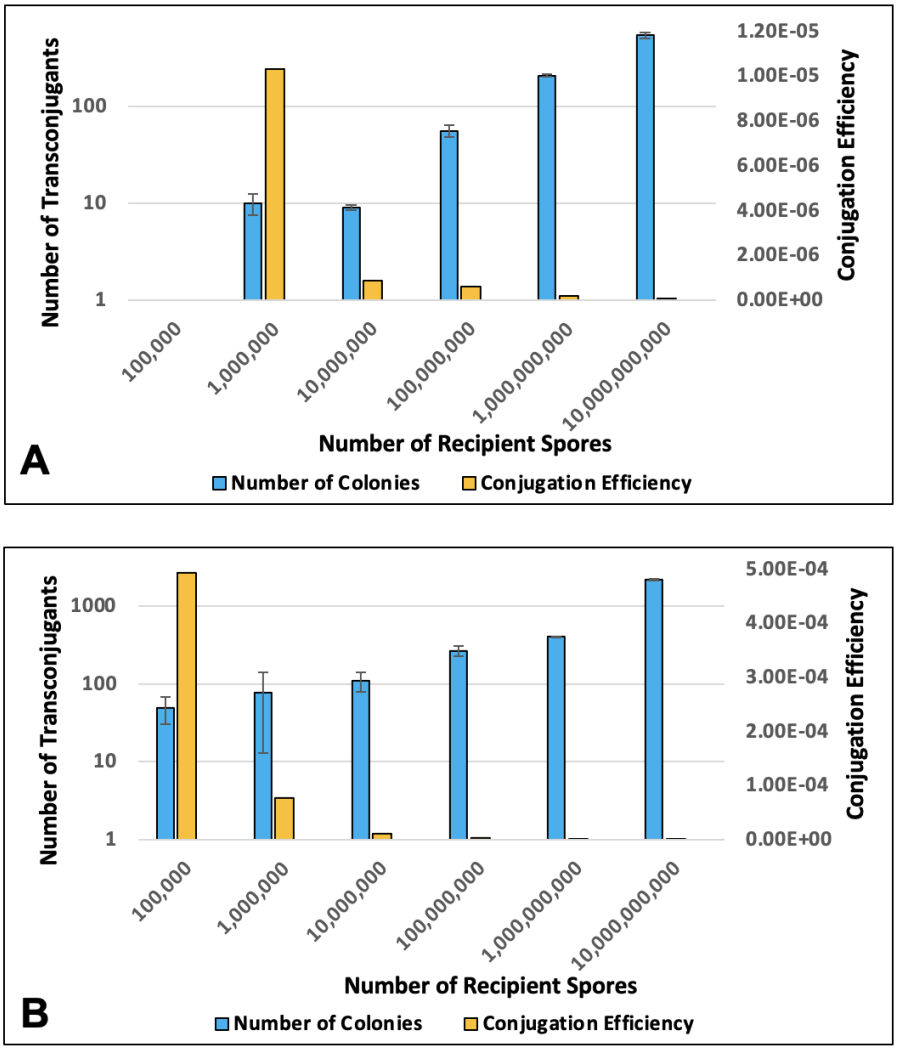


Figure 3. 4 Higher number of recipient spores added to *S. coelicolor* M145 conjugations leads to an increase in number of transconjugant colonies but a decrease conjugation efficiency. Transconjugants were derived from conjugations of **A)** *S. coelicolor* with pSET-152 carrying *E. coli* ET12567/pUZ8002. **B)** *S. coelicolor* with pGWS1370 carrying *E. coli* ET12567/pUZ8002.

3.4.2 *S. clavuligerus* DSM 738 conjugation optimisation highlights similar trend to *S. coelicolor* control: increasing number of recipient vs donor cells leads to an increase in number of transconjugants and decrease in conjugation efficiency.

The experiments for *S. clavuligerus* conjugation optimisation were based on *S. coelicolor* transconjugant colony number and conjugation efficiencies depicted in Section 3.4.1. Frozen spore stocks were used as recipient spores and mixed with a constant number of donor *E. coli* cells. In this section, fresh spores were also harvested and used as recipient cells in conjugation optimisation experiments, as the use of frozen spores in conjugations presented a significantly lower transconjugant number and conjugation efficiency value compared to *S. coelicolor* experiments.

Figure 3.5 A showed that the conjugation efficiencies for SCLAV-pSET conjugations with frozen recipient spores, could only be measured after the addition of 10^6 recipient spores. This presented the highest efficiency of $1.0x10^{-5}$ (10^6 recipient spores added), which fell to $5.2x10^{-8}$ (10^{10} recipient spores added). The number of transconjugants for SCLAV-pSET conjugations increased with the addition of more recipient spores. The addition of 10^6 spores resulted in 56 transconjugants and the addition of 10^{10} recipient spores resulted in 544 colonies. The total number of colonies were lower for SCLAV-pSET transconjugants and lower for efficiencies than for *S. coelicolor*.

Conjugations with pGWS1370 carrying donor *E. coli* cells were only successful with *S. coelicolor* as the recipient, as no transconjugants were recovered for *S. clavuligerus* experiments. As pGWS1370 conjugations were only successful using *S. coelicolor* spores, modifications to the optimisation experiments were designed for *S. clavuligerus*: the main difference being – using fresh spores instead of frozen spores (**Figure 3.5 A, B, C**.). This idea came after conversation with colleagues at GSK. The fresh spores were harvested directly from L3M9+$MgCl_2$ plates and either ½, 1 or 2 plates of recipient cells were added to the donor cells. The total number of recipient spores added to each conjugation was computed where 1/2, 1 and 2 plates of spores corresponded to an average of $6.9x10^8$, $1.4x10^9$ and $2.8x10^9$ total spores respectively.

Overall, results highlighted similar trends between using fresh spores and frozen spores in both number of colonies and conjugation efficiency for SCLAV-pSET and SCLAV-

1370. SCLAV-pSET conjugations presented the highest number of transconjugants for the *S. clavuligerus* data (**Figure 3.5 B** and **Figure 3.5 C**). SCLAV-pSET conjugations also showed that the average number of transconjugants increased from 832 to 1608, when ½ vs 2 plates of fresh recipient spores were used. In contrast, the conjugation efficiency decreased from $1.2x10^{6}$ to $5.8x10^{7}$ when ½ vs 2 plates of fresh recipient spores were added.

SCLAV-1370 conjugations presented similar trends for the number of transconjugants after conjugation with fresh spores, however the numbers were significantly lower. The average number of transconjugants increased from 3 to 5, when ½ vs 2 plates of fresh recipient spores were used, highlighting the low conjugation efficiency of pGWS1370 compared to pSET-152, perhaps due to the encoded dCas9. Notably, utilisation of fresh recipient spores resulted in transconjugant colonies compared to frozen recipient spores in *S. clavuligerus*, which did not result in any transconjugant growth. The conjugation efficiency for SCLAV-1370 presented a decrease when more recipient spores were added, with the highest efficiency being at $4.29x10^{-9}$ for ½ plate and the lowest at $1.4x10^{-9}$ for 1 plate.

Total number of colonies, standard deviation and conjugation efficiency values are summarised in **Tables 3.2.** and **3.3**. Consequently, the number of transconjugants and conjugation efficiency show that the best ratio of donor to recipient spores depends on the desired experimental outcomes. For us, the outcomes were to obtain the highest number of transconjugants, which meant using a lower donor to recipient ratio, i.e., increasing the number of recipient spores. The highest efficiency presented the opposite: a higher donor to recipient ratio.

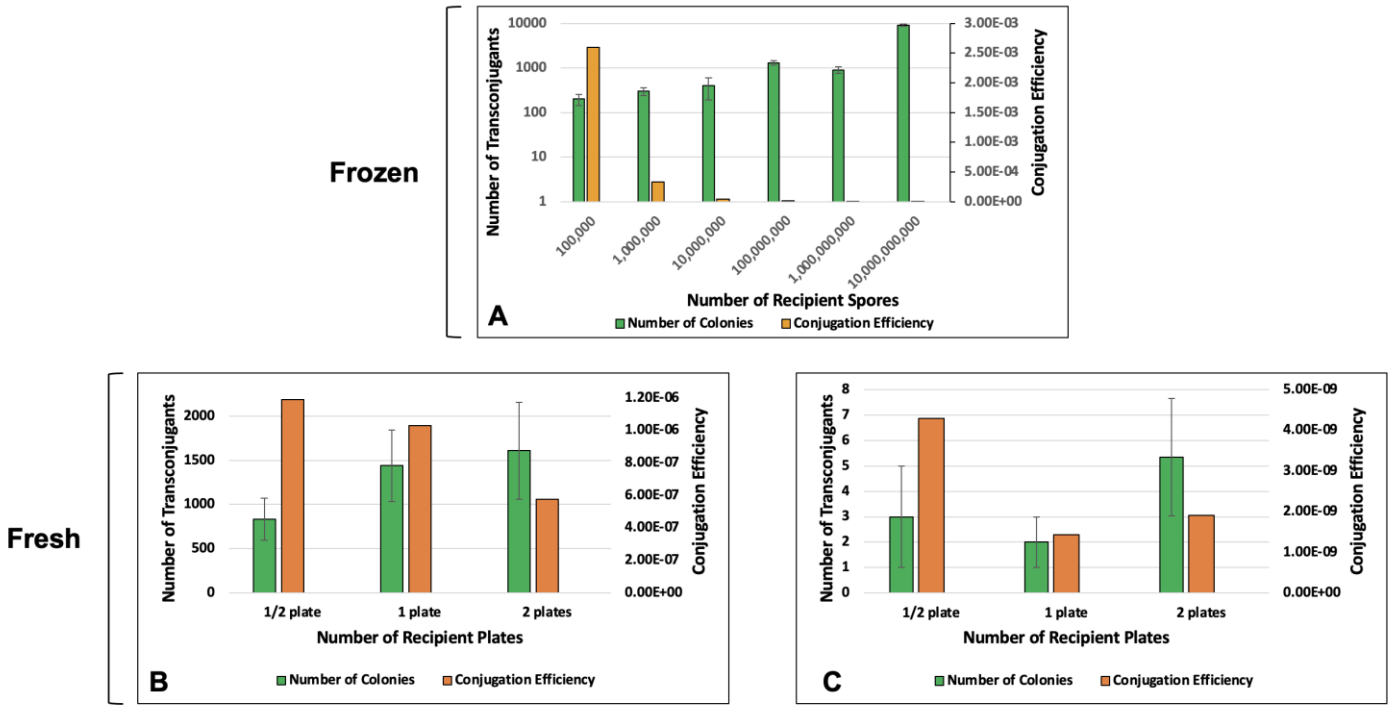


Figure 3. 5. Higher number of spores added to conjugations leads to an increase in number of *S. clavuligerus* DSM 738 transconjugant colonies but a decrease conjugation efficiency. A) SCLAV-pSET number of transconjugant colonies increased and conjugation efficiency decreased using frozen recipient spores. **B)** SCLAV-pSET number of transconjugant colonies increased and conjugation efficiency decreased using fresh plates of recipient spores. **C)** SCLAV-1370 conjugations presented the highest number of transconjugant colonies after using 2 fresh recipient spore plates, but highest conjugation efficiency after using ½ plate of recipient *S. clavuligerus* spores.

Table 3.2. Summary of number of colonies and conjugation efficiency using frozen spores.

	SCLAV-pSET			*SCO-pSET*			*SCO-1370*		
Number of Spores	**Number of colonies**	**SD**	**Conjugation Efficiency**	**Number of colonies**	**SD**	**Conjugation Efficiency**	**Number of colonies**	**SD**	**Conjugation Efficiency**
10^5	0	0	0	200	± 57.74	$2.6x10^{-3}$	49	± 19.04	$4.9x10^{-4}$
10^6	10	± 2.52	$1.0 x 10^{-5}$	300	± 57.74	$3.3x10^{-4}$	77	± 63.79	$7.7x10^{-5}$
10^7	9	± 0.58	$8.6x10^{-7}$	400	± 208.17	$4.3x10^{-5}$	110	± 30.35	$1.1x10^{-5}$
10^8	56	± 7.37	$5.9x10^{-7}$	1300	± 152.75	$1.3x10^{-5}$	265	± 39.50	$2.7x10^{-6}$
10^9	208	± 5.69	$2.0x10^{-7}$	900	± 152.75	$1.0x10^{-6}$	400	± 7.55	$4.0x10^{-7}$
10^{10}	544	± 36.30	$5.17x10^{-8}$	9000	± 556.77	$9.1x10^{-7}$	2173	± 38.55	$2.2x10^{-7}$

Table 3.3. Summary of number of colonies and conjugation efficiency using fresh spores.

	SCLAV-pSET			*SCLAV-1370*		
Number of Spores	**Number of colonies**	**SD**	**Conjugation Efficiency**	**Number of colonies**	**SD**	**Conjugation Efficiency**
1/2 plate	832	± 237.62	$1.2x10^{-6}$	3	± 2.00	$4.3x10^{-9}$
1 plate	1440	± 404.61	$1.0x10^{-6}$	2	± 1.00	$1.4x10^{-9}$
2 plates	1608	± 550.61	$5.8x10^{-7}$	5	± 2.31	$1.91x10^{-9}$

3.5 Conjugation of pEM3 into *Streptomyces clavuligerus* resulted in transconjugants with phenotypic differences.

The conclusions from conjugation optimisation trials in section 3.4 were used as a basis for conjugation of pEM3 into *S. clavuligerus* and allowed for targeted *tap-tpg*$_4$ gene silencing. These included using fresh spores rather than frozen spores and harvesting up to two plates for a higher number of colonies, which was the main aim to screen a range of transconjugants. As mentioned earlier, the targeted *tap-tpg*$_4$ spacer sequence was specific to the *tap* gene found in pSCL4 of *S. clavuligerus*. However, as Tap and Tpg proteins are commonly expressed amongst *Streptomyces* species, particularly those known to have linear plasmids, pEM3 transformants were conjugated into both *S. clavuligerus* DSM 738 and *S. coelicolor* M145, making SCLAV-pEM3 and SCO-pEM3 transconjugant strains. pEM3 contained *aac(3)IV*, the apramycin resistance gene, therefore transconjugants that grew in the presence of the antibiotic were selected for further analysis.

S. coelicolor and *S. clavuligerus* transconjugants were confirmed by patching colonies onto new agar plates containing apramycin and nalidixic acid. Additionally, vector controls such as pSET-152 and pGWS1370 were transformed into *E. coli* ET12567/pUZ8002 competent cells, with the aim of observing phenotypic changes between *tap-tpg*$_4$ knockdown transconjugants. pGWS1370 was selected as a control to observe how the CRISPR-dCas9 system affected the growth of *S. coelicolor* and *S. clavuligerus* transconjugants, as the plasmid contained the dCas9 and the sgRNA but not the *tap-tpg*$_4$ specific spacer sequence.

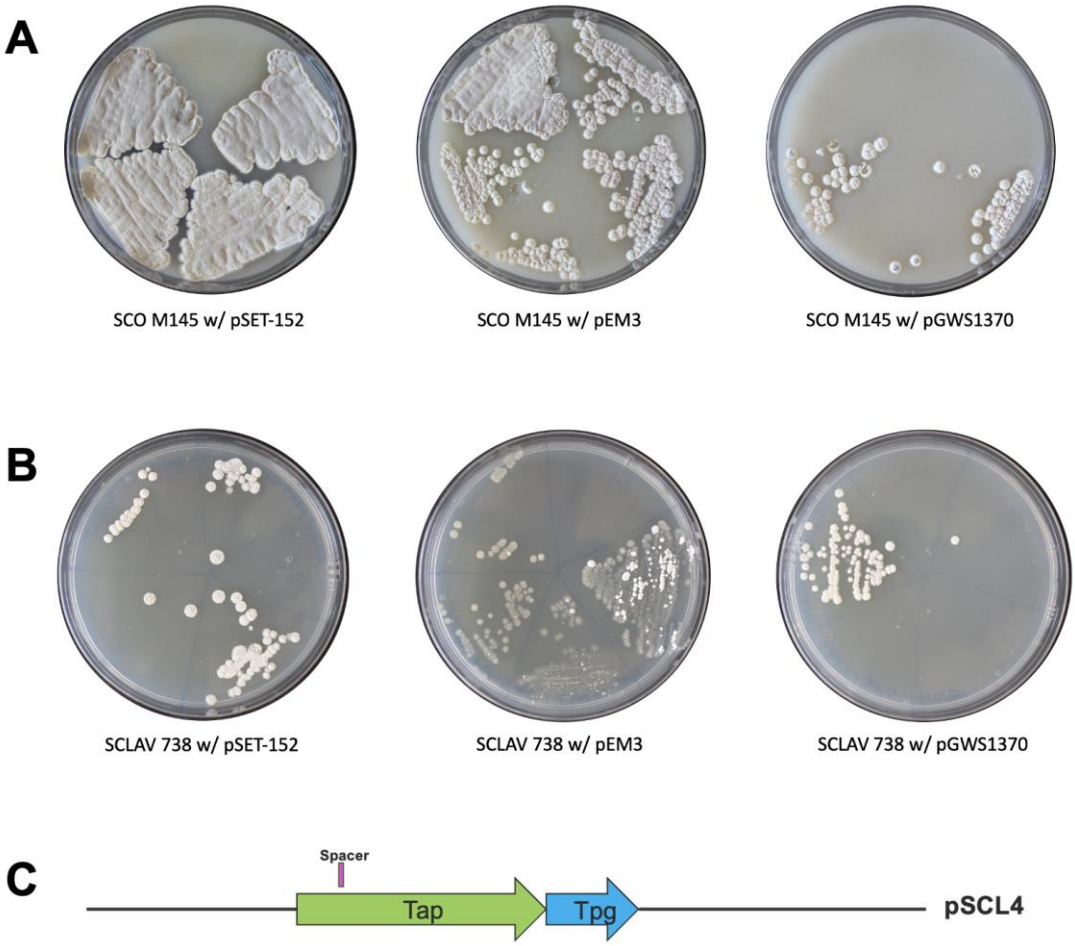


Figure 3. 6. Transconjugant plates highlighting phenotypic differences in *S. coelicolor* and *S. clavuligerus* growth and sporulation. A) All *S. coelicolor* M145 plates show healthy growth and sporulation, with less dense growth for SCO-1370 and SCO-pEM3 re-streaked transconjugants. **B)** Transconjugant plates highlighting phenotypic differences in *S. clavuligerus* growth and sporulation with targeted tap_4 spacer. SCLAV-pSET and SCLAV-1370 transconjugants show healthy sporulation after re-streaking on selective antibiotic plates, whereas SCLAV-pEM3 transconjugants struggle to sporulate after the same amount of time, exhibiting a bald phenotype. **C)** Diagram representing location of pEM3 *tap-*tpg_4 targeted spacer, 221 bp from start codon.

S. clavuligerus conjugations were set up using the same controls as *S. coelicolor* conjugations, which were a no plasmid control, a pSET-152 control and a pGWS1370 vector control (**Figure 3.6 B**). No growth was observed in the no plasmid control plate, highlighting the reliability of the method and results. Steady sporulation and growth were observed for all the *S. coelicolor* transconjugants presenting signs of aerial hyphae and grey coloured spores. Selected *S. clavuligerus* transconjugants presented similar trends to *S. coelicolor* transconjugants (**Figure 3.6 B**). Aerial hyphal growth was observed for SCLAV-pSET, and SCLAV-1370, with signs of sporulation, however colonies lacked the typical dark, green-coloured *S. clavuligerus* DSM 738 spores. As such, SCLAV-pEM3 transconjugants showed two phenotypes: bald and beginning to sporulate.

A bald *S. clavuligerus* phenotype has previously been associated with genotypic changes, highlighting the importance of observing morphological changes in genetically unstable *Streptomyces spp*.. These changes involved circularisation of the chromosome which were identified by screening for the loss of *S. clavuligerus* telomere ends (Algora-Gallardo, unpublished data). Therefore, the two phenotypes were genetically assessed to determine whether the CRISPR-dCas9 targeting tap_4 (**Figure 3.6 C**) resulted in shortening of pSCL4 telomeres, with the hypothesis the telomeres of bald colonies would be affected. However, later results (see section 3.7) highlighted that there was no difference in pSCL4 ends between bald vs. sporulated colonies. These phenotypic changes were important in characterising mutants, as they led to a range of bald and sporulating transconjugant mutants being picked in Chapter 3.0.

In order to determine whether the pEM3 plasmid had integrated in the *S. clavuligerus* chromosome, we picked three different colonies initially and isolated genomic DNA. One colony was chosen at random and used in a PCR, which allowed for confirmation of the *S. clavuligerus* transconjugant colonies. We confirmed the transconjugant colonies by checking for the presence or absence of the tap_4 specific sgRNA in SCLAV-pEM3 alongside the vector, WT and negative controls using **Spacer_specific_sequence** and **dCas9_R3Seq_R** primers (**Figure 3.7**).

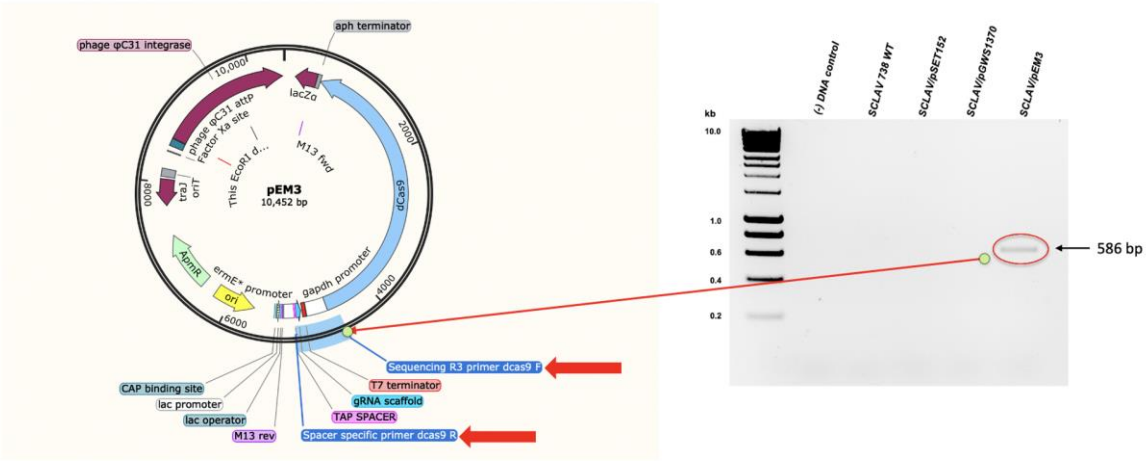


Figure 3.5. Plasmid integration in CRISPR-dCas9 transconjugants was confirmed via PCR. Primers amplified a 586 bp sequence and were specific to the designed sgRNA, shown by red arrows. A band was expected for the SCLAV-pEM3 gDNA sample, as this would have been the only sample with the *tap*$_4$ specific spacer, which served as the forward primer. A no-DNA negative control was used and did not include the spacer sequence.

3.6 Southern Blot confirms presence of pSCL4 after *tap-tpg*$_4$ gene silencing.

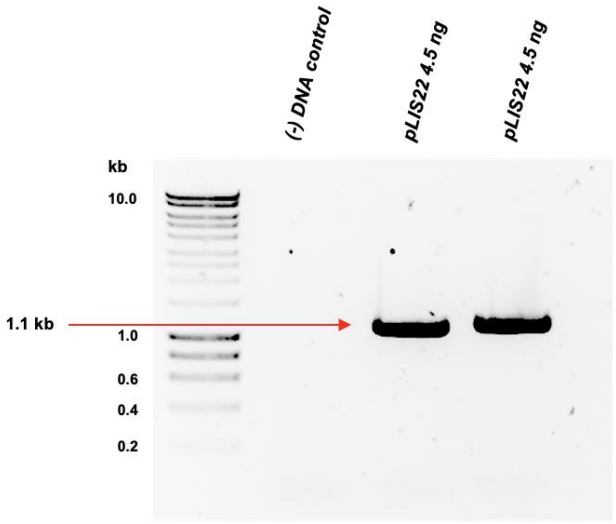
Cutting pSCL4 has previously led to plasmid loss which caused *S. clavuligerus* DSM 738 chromosome circularisation (Algora-Gallardo, unpublished data). The plasmid was cut at *parB*, which was the reason for plasmid loss, however in our hypothesis (Chapter 1.0) we speculated that the reason behind chromosome circularisation was the absence of *tap-tpg*$_4$. The absence of *tap-tpg*$_4$ could have affected the chromosome as the Tap and Tpg proteins allow for linear end-patching and the chromosomal telomere arms were shortened as well (Algora-Gallardo, unpublished data). It remains unclear whether *tap-tpg*$_4$ bind to the hairpins on the chromosome and what significance they hold in chromosome end-patching. To clarify whether the loss of pSCL4 and the resulting chromosome circularisation shown by Algora-Gallardo, (unpublished data) was due to the absence of *tap-tpg*$_4$ on pSCL4, we performed a Southern Blot.

We determined the presence or absence of *tap-tpg*$_4$ in the genomic DNA of the *S. clavuligerus* transconjugants and used it to confirm the presence or absence of pSCL4. This was done by designing a 1.1 kb labelled probe which bound to the *tap*$_4$ gene during probe hybridisation using **Lis_RT_Tap_Rv** and **Elmira_Tap_Rv** primers. pLIS22 DNA (see Chapter 2.0) was used to amplify the 1.1 kb sequence (**Figure 3.8 A**), which was gel purified and labelled with DIG-dUTP. pLIS22 is a pMS82 based integrating plasmid, which was complemented with *tap-tpg*$_4$ on the chromosome. Plasmid DNA was used for probe amplification instead of gDNA, as this resulted in less unspecific binding during probe hybridisation. The gDNA was digested with SacI meaning that when the unbound probe washed away, a 4.4 kb fragment was visible if a copy of *tap-tpg*$_4$ existed on pSCL4 and if the plasmid was still present, as depicted by **Figure 3.8 B**, and by the blot shown in **Figure 3.9 B**.

Through the Southern blot, we aimed to understand whether there was a loss of pSCL4 if the expression of tap-tpg$_4$ was silenced to determine the role of the *tap* and *tpg* genes in plasmid retention and chromosome circularity. For this we used the following strains: *S. clavuligerus* WT, SCLAV-pLIS22, SCLAV-pEM3, SCLAV-1370 (referred to as 1370) and pLIS22 plasmid DNA. Additionally, we accounted for various potential experimental outcomes, highlighted in **Figure 3.9 A**. As such, the absence of a band on the membrane at 4.4 kb after probe hybridisation and washing could have indicated the loss of pSCL4 or could suggest the dCas9 gene silencing had detrimental effects. The presence of a band at 4.4 kb as seen by **Figure 3.9 B.** suggested that the plasmid could have remained linear. Potential reasons behind the maintenance of plasmid linearity were that the dCas9

did not have any effects on *tap-tpg*$_4$ expression, pSCL4 circularisation, or that *tap-tpg*$_4$ silencing may not cause the loss of the pSCL4 and its linearity. To summarise, the Southern blot in **Figure 3.9 B.** highlighted a band at 4.4 kb, confirming the presence of a copy of *tap-tpg* on pSCL4 for SCLAV-pEM3, but most importantly, *tap-tpg*$_4$ gene silencing did not lead to plasmid loss.

A



B

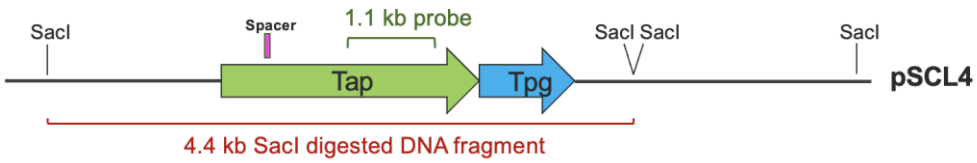


Figure 3. 6. Probe PCR amplification and design targeting *tap-tpg*$_4$ for use in Southern Blot. A) pLIS22 DNA was used to amplify a 1.1 kb probe which was analysed on a 1.0% TAE agarose gel for 90 minutes at 80V and subsequently purified. **B)** Diagram representing Southern Blot probe binding location on pSCL4. The 4.4 kb SacI digested fragment is shown in red.

A

B

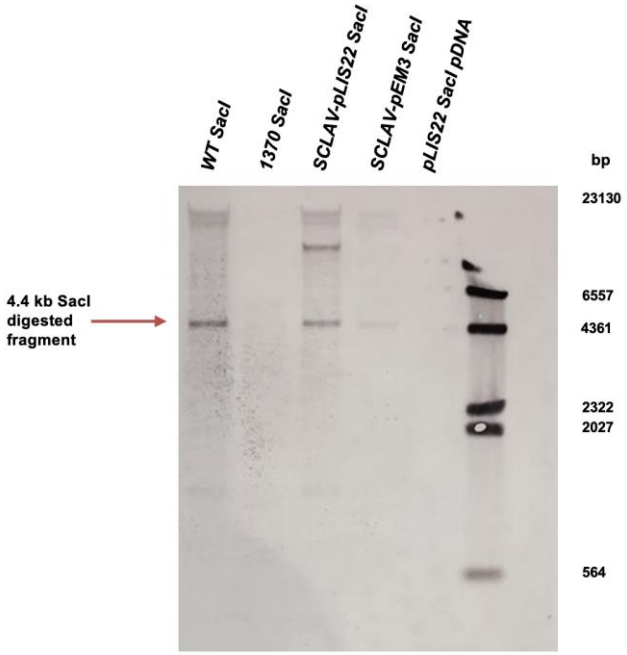
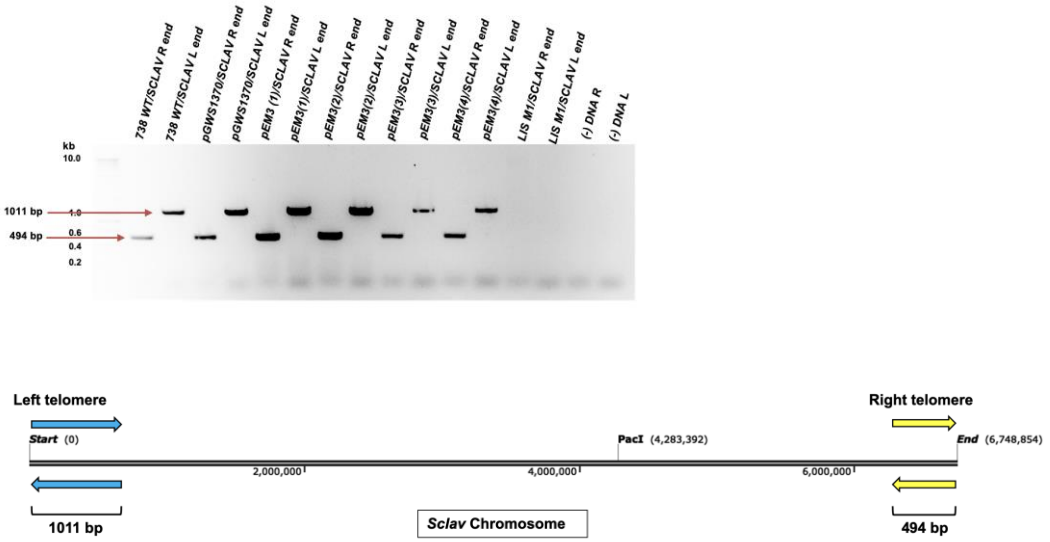


Figure 3. 7. Diagram representing Southern Blot experimental results and their meanings in terms of plasmid loss and linearity. A) Potential outcomes of the presence and absence of a band at 4.4 kb **B)** Southern Blot of SCLAV-pEM3 gDNA highlighting presence of 4.4kb tap-tpg$_4$ band and pSCL4 plasmid. Controls used included *S. clavuligerus* WT gDNA, SCLAV 1370 gDNA (1370), SCLAV-pLIS22 gDNA and pLIS22 plasmid DNA (definite positive control).

3.7 Linear *S. clavuligerus* pSCL4 and chromosome telomere ends were confirmed by PCR.

The Southern Blot in Section 3.6 proved that there was a copy of *tap-tpg*$_4$ present on pSCL4, confirming the plasmid retention, but it was not known whether pSCL4 or *S. clavuligerus* chromosome ends circularised through dCas9 gene silencing. Therefore, we constructed PCR primers to test whether the SCLAV-pEM3 chromosome was circularised, potentially due to *tap-tpg*$_4$ silencing. The primers annealed to the terminal repeat sequences of the *S. clavuligerus* DSM 738 chromosomal telomeres, which had previously been recovered and sequenced by Algora-Gallardo, (unpublished work). The same idea was applied for pSCL4 telomere end primers. **Lis-ChromTelProbe_F**, **Lis-ChromTelProbe_R**, **Lis-CaoChr-R-end_F** and **Lis-CaoChr-R-end_F** were used to amplify *S. clavuligerus* chromosome left and right ends (**Figure 3.10 A**). **Lis-SCL4-L-end_F**, **Lis-SCL4-L-end_R**, **Lis-SCL4-R-end_F** and **Lis-SCL4-R-end_R** were used to amplify pSCL4 left and right ends (**Figure 3.10 B**). Results showed bands for all the positive controls, but also for the SCLAV-pEM3 mutant. The presence of a band for our SCLAV-pEM3 strain suggested the presence of chromosomal and plasmid ends, and thus replicon linearity despite dCas9 knockdown of *tap-tpg*$_4$.

A



B

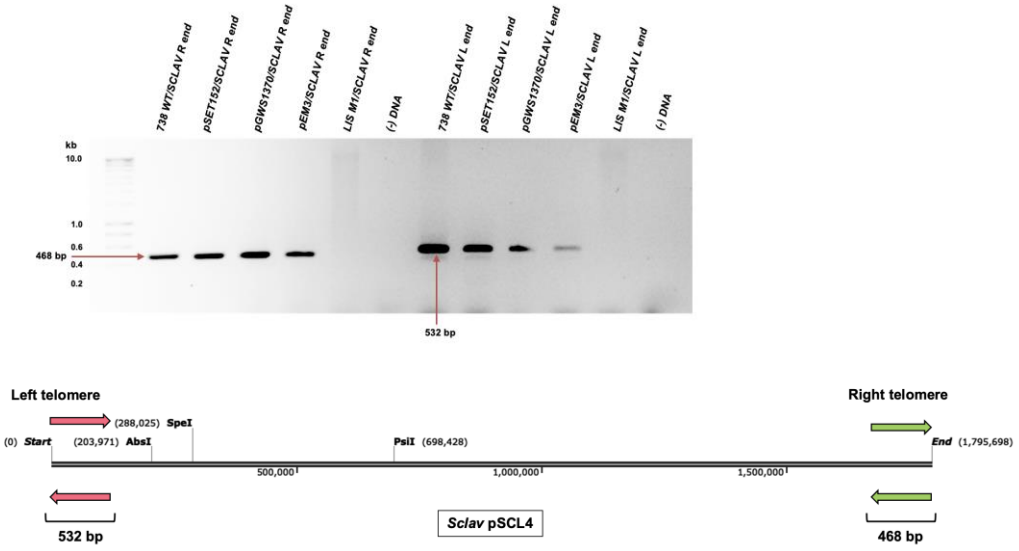


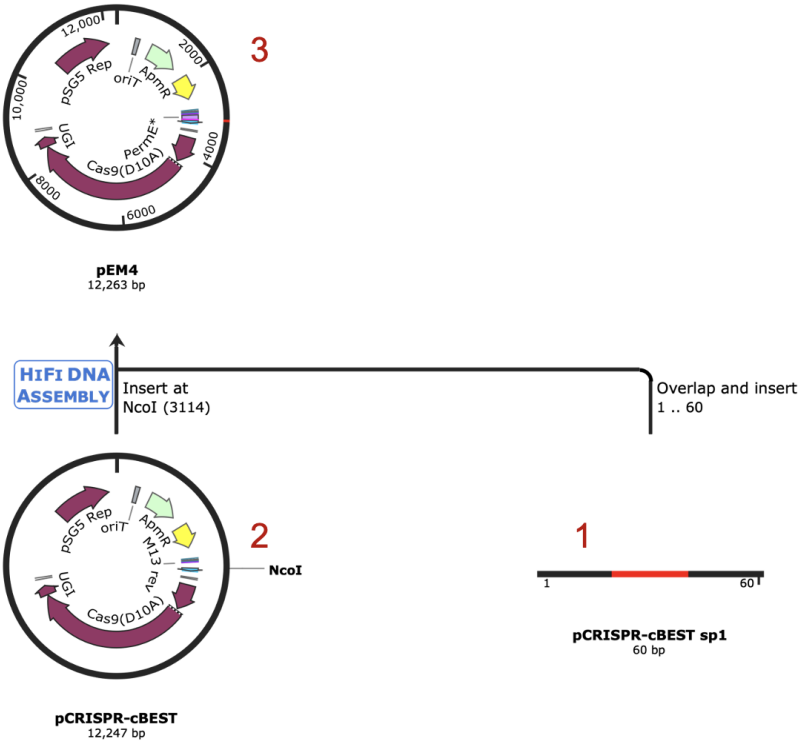
Figure 3.8. PCR highlighting linear *S. clavuligerus* chromosome and pSCL4 ends with representative diagram of primer locations. A) WT and mutant *S. clavuligerus* gDNA was used to amplify left and right chromosome ends resulting in 1011 bp and 494 bp amplicons, respectively. Not to scale **B)** WT and mutant *S. clavuligerus* gDNA was used to amplify left and right pSCL4 ends resulting in 532 bp and 468 bp amplicons, respectively. Not to scale.

3.8 CRISPR-BEST as an alternate CRISPRi method for translational blockage of *tap-tpg*$_4$.

Earlier in the chapter, we described various CRISPR-based methods (**Table 3.1**) and concluded that CRISPR vectors without resulting DSBs would be favourable when conjugated with a genetically unstable organism like *S. clavuligerus*. Amongst these vectors, we looked at transcriptional blockage using CRISPR-dCas9 and, in this section, aim to elucidate the effects of CRISPR-BEST. We used pCRISPR-BEST to investigate whether the effects of replicon linearity were specific to the CRISPR-dCas9 knockdown system. Another reason why pCRISPR-BEST was chosen to test the significance of *tap-tpg*$_4$ was that unlike the CRISPR-dCas9 gene silencing mechanism, which blocked *tap-tpg*$_4$ mRNA expression, the CRISPR-BEST system introduced a stop codon to the target sequence through a nonsense mutation, blocking translation. The CRISPR-BEST system uses deaminase based base editing for stop codon introduction, which results in a truncated target protein and a truncated target protein is one of the advantages of blocking translation over blocking transcription as it does not result in any downstream polar effects, i.e., translation will terminate at the stop codon (**see Figure 3.1 B**). Of the two Addgene plasmids (aBEST and cBEST), pCRISPR-cBEST was chosen and a spacer designed to target *tap-tpg*$_4$ was designed using CRISPy-Web. We chose pCRISPR-cBEST as *S. clavuligerus* is a GC rich organism, therefore the Cas9-D10A Nickase Protein (D10A) encoded by the plasmid had a wider target for C:G to T:A deamination.

To target *tap*$_4$ on *S. clavuligerus*, leading to a truncated Tap$_4$, we introduced a spacer sequence into the gRNA cassette. This was achieved by ordering a 60 bp ssDNA oligonucleotide and cloning the sequence into the CRISPR-cBEST, a pSG5 replicon-based vector, through HiFi DNA Assembly (**Figure 3.11 A**). The ssDNA oligonucleotide included a 20 bp spacer sequence complementary to 20 bp on pSCL4 in *S. clavuligerus*, allowing it to target *tap*$_4$. HiFi DNA assembly overcame the problem of any 5' - and 3' - end mismatches which would normally be seen for T4 ligase ligations. pCRISPR-cBEST was linearised using NcoI for compatible ends with the ssDNA oligonucleotide (**Figure 3.11 B**). This construct was named pEM4, and consisted of a *tap*$_4$ spacer, gRNA and Cas9(D10A). Sanger sequencing confirmed the construction using **Addgene_M13/pUC_R** and **pCRISPR-BEST_sp1_F** primers.

A



B

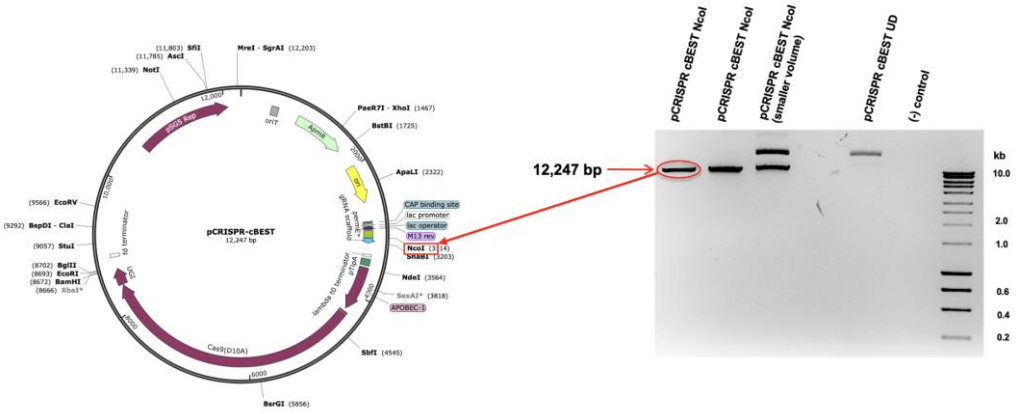


Figure 3. 9. pEM4 cloning to target *tap₄* A) Overview of the CRISPR-cBEST cloning process using HiFi DNA Assembly to create pEM4 with spacer targeting *tap*$_4$. **B)** Gel electrophoresis to verify pCRISPR-cBEST linearisation via NcoI restriction enzyme digestion**.**

To determine the effects of a *tap-tpg*$_4$ translational blockage via the CRISPR-cBEST system, the plasmid was conjugated with *S. clavuligerus*. After confirmation of correct spacer insertion and position on pEM4, it was transformed, and successful transformants were selected for conjugation using *S. clavuligerus* DSM 738 recipient spores and *S. coelicolor* M145 recipient spores. Once again, *S. coelicolor* was used as an experimental control and as a point of comparison to check whether the CRISPR-cBEST system functioned in a model organism.

S. coelicolor conjugations were set up using two controls: the pCRISPR-cBEST plasmid without the spacer and a no plasmid vector control. These allowed us to observe the effect of the spacer on organism growth/sporulation. SCO-pEM4 transconjugants presented less sporulated colonies and confluent growth compared to the vector control. The SCO-cBEST vector control showed first signs of sporulation and actinorhodin production, suggesting that the spacer in pEM4 interfered with the onset of secondary metabolite production. Nonetheless, both the theoretical mutant and vector control transconjugants presented little to no sporulation after 5 days, highlighting the detrimental effects of such a CRISPR system for *Streptomyces*, although it was uncertain whether the mutation did indeed take place (**Figure 3.12 A**).

Successful transconjugant colonies were also patched onto new plates, supplied with apramycin, and grown at the same temperature as initial conjugation, to determine plasmid retention. To see whether the plasmid was eliminated from its host, the successful transconjugants grown initially at 30 °C were also re-streaked and grown at 40 °C on SFM supplied solely with nalidixic acid (**Figure 3.12 B**). The difference in temperature allowed for growth, avoiding potential long-term toxicity and off target effects. The primary phenotypic difference compared to transconjugant growth at 30 °C was the lack of sporulation in addition to larger, more yellow-coloured colonies around the edges of the re-streaked lawn quarters.

A

B

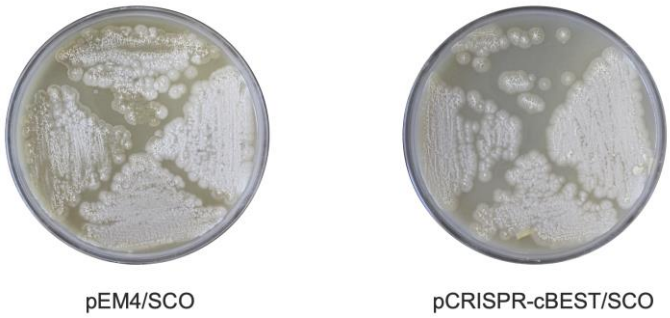


Figure 3.10. SCO-CRISPR-cBEST transconjugant plates highlighting phenotypic differences in *S. coelicolor* growth and sporulation before and after vector loss. A) Steady growth can be seen for the SCO-pEM4 and SCO-cBEST transconjugants after 5 days of incubation at 30 °C. SCO-pEM4 transconjugants show little to no sporulation. SCO-cBEST transconjugants show first signs of grey-coloured spores and actinorhodin production. Negative control shows no growth. **B)** Transconjugant plates after growth at 40 °C to remove temperature sensitive replicon. Transconjugants for both the control and SCO-pEM4 show lack of sporulation, however colonies are larger with first signs of aerial hyphae growth. pEM4/SCO = pCRISPR-cBEST plasmid with *tap*$_4$ specific spacer conjugated with *S. coelicolor*, pCRISPR-cBEST/SCO = pCRISPR-cBEST conjugation with *S. coelicolor*.

Transconjugants from the SCLAV-cBEST conjugations did not grow, despite several attempts at the experiment and the use of optimal conjugation conditions described in section 3.4. Growth was seen for *S. coelicolor* transconjugants, highlighting that the issue might lay with the temperature sensitive replicon for *S. clavuligerus.* Although a lot remains to be explored about repair and recombination in *Streptomyces*, localised genetic instability can be a result of pSG5 replicon activation. The natural way to cure the replicon is to increase the temperature, which may have been a reason why *S. clavuligerus* did not grow. A way we tried to overcome the high temperatures required for replicon loss was to lower the temperature. Lowering the temperature however also causes rearrangements and genome instability due to pSG5 replication fork generation, which interferes with the natural replication fork of the host organism, posing toxic effects (Wlodek *et al.,* 2017). Plasmids with a pSG5 based replicon have previously caused issues during conjugation with *S. clavuligerus* (Steve Kendrew - personal communication). Due to a shortage of time we decided to focus the rest of the work in this thesis on dCas9-based gene editing to determine the significance of *tap-tpg* silencing of all five replicons.

3.9 Summary

In this chapter, we investigated CRIPSR-dCas9 and CRISPR-BEST transcriptional interference of *tap* and *tpg* genes via multi-step cloning in *S. clavuligerus* and *S. coelicolor.* Initial CRISPR-dCas9 experiments focussed on designing and amplifying a gRNA and 20 bp spacer sequence, to create a CRISPR system specific to tap_4, targeting the co-transcribed tpg_4 as well. **Figure 3.3.** confirmed a 183 bp amplified gRNA + spacer sequence, which meant that the pEM3 construct included all the necessary components for conjugation with the recipient organisms. Integrated plasmid cloning has been a common method allowing for conjugal transfer of DNA from *E. coli* to *Streptomyces* and methods have been developed to introduce vectors into more difficult strains such as *Streptomyces fradiae* (Bierman *et al.*, 1992). Similarly, there have been advances made in DSB genome editing, since 2015, the first time CRISPR/Cas9 systems were applied to *Streptomyces*. These editing strategies have been created in the form of CRISPR/Cas9 system plasmids such as the pCRISPomyces system (Cobb, Wang and Zhao, 2015) who tested their Golden-Gate Assembly cloned plasmid in *S. lividans*. Cobb *et al.,* (2015) also highlighted the importance of the spacer sequence and PAM sequence to efficiently knockout a target sequence, however unlike the results in this chapter, Cobb *et al.,* (2015) used multiple sgRNAs and relied on DSBs and homologous recombination to allow for gene knockouts.

Figures 3.6 A. and **3.6 B.** show the differences in *S. coelicolor* and *S. clavuligerus* transconjugant growth after pEM3, pGWS1370 and pSET152 conjugations. *S. coelicolor* presented sporulated, healthier-looking colonies compared to *S. clavuligerus*, which showed patchy bald and some white, sporulated colonies for SCLAV-pEM3 transconjugants after re-streaking. These phenotypic differences highlight *S. clavuligerus'* sensitivity to integrating vectors, as normally healthy, sporulated colonies present a dark green colour in contrast to the grey/white. For more sensitive organisms such as *S. clavuligerus,* DSBs could impose toxic off-target effects (Qi *et al.*, 2013).

We supported the phenotype of SCLAV-pEM3 mutants and confirmed them genotypically though PCR in **Figure 3.7**., which showed that SCLAV-pEM3 contained a tap_4 specific spacer, unlike the controls. A Southern blot provided more genetic information and was performed with the main purpose of determining plasmid-chromosome interactions and linearity, by confirming pSCL4 linearity. The Southern blot

results in **Figure 3.9.** suggested various outcomes: 1) that the plasmid remained linear, i.e. the dCas9 did not have any significant effects on *tap-tpg* expression, 2) that the plasmid circularised or 3) that *tap-tpg* silencing may indeed not affect the loss of the plasmid or plasmid linearity.

To determine the effects of dCas9 gene silencing on linearity, the plasmid and chromosome ends of SCLAV-pEM3, SCLAV-1370, SCLAV-pSET and SCLAV-LIS M1 were confirmed via PCR (**Figure 3.10**). *Streptomyces* are known to have extremely conserved, palindromic sequences at the ends of their chromosome and plasmids, as seen for *S. rocheii*, *S. lividans* and *S. avermitilis* amongst other species (Nindita *et al.*, 2015). Deletions at the ends of *Streptomyces* GLPs, have led to chromosomal circularisation. Disruption of *tap-tpg* genes in *S. lividans* and *S. rocheii* have highlighted their importance and effect on chromosome linearity and underlined the relationship between the GLPs and the chromosome.

The interactions between GLPs and the chromosomes in *Streptomyces* have further been investigated in the study by (Bao and Cohen, 2001) who discovered that *tpgL* found on the chromosome affected the ability of *S. lividans* to maintain chromosomal linearity and retain its pSLA2-S GLP. On the other hand, Nindita *et al.,* (2015) highlighted the *tapRM-tpgRM* genes on the pSLA2-M GLP in *S. rocheii* worked to maintain linearity in the chromosome. The study used plasmid curing and complementation experiments to verify results and emphasises high sequence similarity of *tap-tpg* amongst *S. lividans* and *S. rocheii* (Nindita *et al.,* 2015). Studies by Bao and Cohen, (2001) and Nindita *et al.,* (2015) show, *tap-tpg* homologs on GLPs substitute their function to the chromosome or vice versa to maintain linear growth for Streptomyces. Another example of this was proven by (Yang *et al.*, 2013), who established the dependency of the single *tpg* variant on *S. lividans'* SLP2 plasmid. Most significantly, they highlighted that *tpg* was still functional, but required the presence of chromosomal tap for the purpose of end patching, thus linear growth (Yang *et al.,* 2013).

Therefore, we speculated that an alternate reason explaining the intact, linear ends of pSCL4 and the chromosome, shown by **Figure 3.10 A.** and **Figure 3.10 B.**, could be the substitution of a functionally inactive pSCL4 *tap-tpg* by pSCL3 or pSCL2 *tap-tpg*. To test this hypothesis, CRISPR-dCas9 sgRNA multiplexing, i.e., multiple spacers targeting

pSCL4, pSCL3 and pSCL2 *tap-tpg*, could be implemented allowing for identification of any interaction between the terminal proteins on the different plasmids.

This this chapter we applied the CRISPR-cBEST vector and tested translational blockage of Tap4 in *S. coelicolor* and *S. clavuligerus*. The cytidine deaminase base editing process, like CRISPR-dCas9, was chosen due to its suitability for sensitive *Streptomyces* strains, such as *S. clavuligerus*, as the introduction of a stop codon via deaminase-based editing led to a truncated protein compared to a DSB. Results in **Figure 3.12 A** and **Figure 3.12 B.** showed the problematic nature of conjugal transfer in *S. clavuligerus*, as only SCO-cBEST and SCO-pEM4 transconjugants grew. Furthermore, **Figure 3.12 A** and **Figure 3.12 B** presented the phenotypic differences (like lack of sporulation) when the *S. coelicolor* transconjugants were grown at 40 °C, to expel the pSG5-based temperature sensitive replicon. This too could result in problems in future *S. clavuligerus* experiments as the expulsion of the replicon is performed under temperatures above the optimal *S. clavuligerus* growth temperature of 26°C. Compared to the pSG5 replicative pCRISPR-cBEST vector described by Tong *et al.,* (2019), integrated vectors such as pSET152-based systems have shown far more stable repression effects (Baltz, 2012), which are mimicked in the results of this chapter.

Finally, we investigated the fundamental process allowing for gene silencing in *Streptomyces*: conjugations. More specifically, conjugation efficiencies were investigated as shown by **Figure 3.4**. and **Figure 3.5**., highlighting the necessity to use multiple plates of fresh recipient spores to obtain a higher number of transconjugants. The conjugation efficiency experiments in this chapter considered the number of recipient cells as the independent variable for calculations. Particularly for genetically complex and metabolically vital Streptomycetes, conjugation optimisation experiments are key to constructing a successful genetic engineering tool, and have been frequently observed in the literature (Du *et al.,* 2012; Wang and Jin, 2014; Zhang *et al.,* 2019). One example is the study by Zhang *et al.*, (2019), who looked at optimising conjugal transfer for *Streptomyces kanamyceticus* for kanamycin B production. Unlike the results described in this chapter, the study by Zhang *et al.,* (2019) looked at several key factors for conjugations using integrative pSET152 based vectors, which included the time required to heat shock spores, donor-to-recipient ratios and antibiotic overlaying times. Zhang *et al.,* (2019) found that both the number of colonies and the frequency increased

when the number of spores added to the *E. coli* ET12567 cells increased to it's maximum ratio of 100:1, providing a conjugation frequency of $6.7x10^{-6}$ compared to our highest efficiency in this chapter of $2.6x10^{-1}$ for SCO-pSET and $1.0x10^{-3}$ for SCLAV-pSET using frozen spores. Notably, the frequency presents a different ratio of number of transconjugants over total viable cell number and for the puporses of this Chapter, this value did not aid our decision of which conjugation method to adapt. A vital conclusion from the study by Zhang *et al.,* (2019) was that the number of recipient spores alone made no difference to conjugation efficiencies and frequencies, but indeed the donor to recipient ratio. So, despite adding only a maximum of 10^9 spores, their number of *E. coli* ET12567 cells was at 10^7, 100 times more than in this study. In a study by Du *et al.*, (2012) the *Streptomyces lincolnensis* conjugation frequency reached its maximum of $1.1x10^{-4}$ when using 0.5 mL of mycelial culture and 10^8 donor *E. coli* ET12567 cells, slightly higher than Zhang *et al.,* (2019), although not directly comparable due to different methods. However, both studies by Du *et al.,* (2012) and Zhang *et al.,* (2019) used mycelia to test their conjugal transfer rates. In the study by Zhang *et al.,* (2019), mycelia yielded a higher conjugation frequency of $5.3x10^{-5}$, 7.5 times higher than the frequency when using spores.

Further studies by Du *et al.,* (2012) and Wang and Jin, (2014) looked at the addition of $CaCl_2$ and $MgCl_2$ to the conjugation media for a variety of strains. Wang and Jin, (2014) investigated conjugations using a pSET-152 integrative vector in *Streptomyces netropsis* SD-07, and determined an increase of 10-10,000 times in conjugation frequency when an optimal amount of 60mM was added of each compound individually. Tests were also conducted on *S. coelicolor*, *Streptomyces lavendulae*, *S. venezuelae* with the conclusion that $CaCl_2$ concentrations were a crucial factor for increasing conjugation efficiency and frequency. The study by Du *et al.,* (2012) used both multiple copy and integrative vectors to optimise *Streptomyces lincolnensis* conjugations and in contrast to Wang and Jin, (2014), found that the optimal concentration of $MgCl_2$ is 20 mM, similar to the concentration used in this chapter.

The idea of using fresh recipient spores compared to frozen recipient spores was also explored in this chapter. The conjugation efficiency did not increase for *S. clavuligerus* when fresh spores were used compared to when frozen spores were used, however the number of colonies went from 554 to 1608 colonies for the most efficient conjugation: SCLAV-pSET. Additionally, the use of frozen recipient spores resulted in growth of transconjugant colonies for the previously unsuccessful SCLAV-1370 conjugation

showing an increasing trend when more spore plates were harvested for experiments. Due to the fact that the CRISPR vector backbones have resulted in difficult conjugations, particularly for the SCLAV-1370 vector control and SCLAV-cBEST vector control, results in these conjugation efficiency experiments showed a lot of promise. Not in regards to conjugation efficiencies as these all generally decreased when increased number of recipient spores (frozen or fresh) were added, but in regards to number of colonies, as these are necessary to proceed with analysis of CRISPR mutants. Therefore, the conjugation efficiency experiments conclude two factors that aided increased number of spores: the necessity of using increased number of spores when using frozen recipient spores or to use a minimum of 2 plates of fresh recipient spores for difficult conjugations. These two factors were directly applicable to the CRISPR-cBEST system from this chapter, as there were no transconjugants for SCLAV-pEM4 and SCLAV-cBEST conjugations and were applied, having no effects.

The outcome of the CRISPR-cBEST conjugations emphasised the need to further explore CRISPR-dCas9 to develop an effective method in the gene silencing of *tap-tpg* for *S. clavuligerus*. The results in this chapter highlight the need to develop the CRISPR-dCas9 gene silencing method to allow us to determine the significance of the three *tap-tpg* pairs on all of the GLPs of *S. clavuligerus*, not just pSCL4, as results when targeting solely tap_4 were inconclusive.

4.0 Characterisation of the Replicon Structure of CRISPR-dCas9 Multiplexed Mutants Through Illumina Sequencing.

In Chapter 3.0 we showed that the knockdown of *tap-tpg*$_4$ resulted in unchanged chromosome and pSCL4 ends, therefore it is hypothesised that the single sgRNA method of gene silencing either did not function, that *tap-tpg*$_4$ do not play a great role in maintaining the chromosomal and pSCL4 ends or that the missing *tap-tpg*$_4$ on pSCL4 were complemented by *tap-tpg* in another plasmid, compensating for the absence of *tap-tpg*$_4$ expression. It is therefore speculated that an alternate reason explaining the intact, linear ends of both pSCL4 and the chromosome, could be the compensation of a functionally inactive pSCL4 *tap-tpg* by pSCL3 or pSCL2 *tap-tpg*.

This phenomenon has previously been witnessed in *Streptomyces lavendulae subs. lavendulae* CCM 3239 (Novakova *et al.*, 2021) or by Nindita *et al.*, (2015) who showed complementation of the *S. rochei* chromosome with *tap-tpg* resulted in linear chromosome maintenance of pSLA2-cured strains, suggesting that there is a crucial relationship between *tap* and *tpg* on plasmids and the chromosome. It has also been hypothesised in *S. clavuligerus* between pSCL4 and the chromosome, i.e., the absence of *tap-tpg* from the chromosome was compensated by Tap-Tpg4 encoded in trans, this was suggested by Algora-Gallardo *et al.,* (unpublished data), who demonstrated that the loss of pSCL4 resulted in chromosomal circularisation and loss of terminal ends. Thus, the potential of *tap-tpg*$_3$ and *tap-tpg*$_2$ interacting with pSCL4 and/or the chromosome after knockdown of *tap-tpg*$_4$ from the GLP could be an explanation for presence of chromosome and pSCL4 ends. In this chapter we aimed to elucidate which one of these hypotheses is true, by carrying out experiments to knockdown *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ on all these *S. clavuligerus* GLPs carrying a copy of the co-transcribed genes.

4.1 Construction of a CRISPR-dCas9 multiplexed vector targeting *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ in *S. clavuligerus*.

For the multiplexed sgRNA targeting of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$, new primers (**sgRNA_F_pSCL4-tap**, **sgRNA_F_pSCL3-tap**, **sgRNA_F_pSCL2-tap**) were designed to amplify the sgRNA cassette using pGWS1370 as template DNA. Multiplexing of sgRNAs is a method of targeting various genes or various parts of the same gene to maximise gene knockout or knockdown effects (McCarty *et al.*, 2020). The resulting PCR

reaction (**Figure 4.1 A**) with the brightest band was DNA purified for all three sgRNAs and cloned into pGEM®-T Easy, following transformation into DH5- α cells. pGEM®-T Easy plus the cloned sgRNAs for *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ and pGWS1370 were restriction enzyme digested with BamHI and NcoI and gel purified (**Figure 4.1 B**). The 183 bp insert (sgRNA cassette) was ligated with the 10,277 bp vector backbone (pGWS1370) to create pEM3.1, pEM3.2 and pEM3.3, which had spacers targeting *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$, respectively. The plasmids were used to transformed DH5-α chemically competent cells and experimentally confirmed through a colony PCR (**Figure 4.1C**) and Sanger sequencing, which highlighted the insertion of the sgRNA cassette targeting the according *tap-tpgs* on the GLPs.

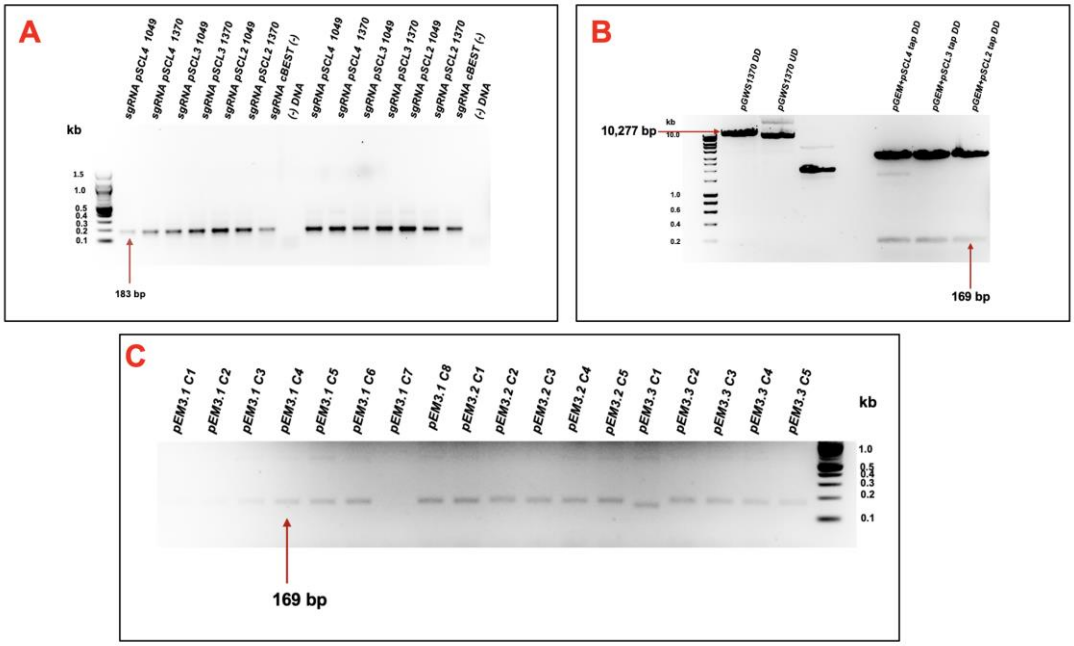


Figure 4.1. Cloning and confirmation of CRISPR-dCas9 *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ targeting plasmids. A) Gel electrophoresis of PCR confirming 183 bp sgRNA amplicons for CRISPR-dCas9 multiplexing targeting *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$. **B)** Gel electrophoresis of pGEM®-T Easy/sgRNA and pGWS1370 restriction enzyme digestion confirming 169 bp and 10,277 bp bands. Where: DD = double digest with BamHI and NcoI, UD = no enzyme added. **C)** Colony PCR verifying presence of pEM3.1, pEM3.2 and pEM3.3 transformants to see a 20bp difference in transformants if the spacer sequence was present. Expected 169 bp fragment can be seen for all transconjugants except pEM3.1 C7 and pEM3.3 C1, which show no band or a smaller band, highlighting the absence of the spacer sequence and incorrect transformants. C1-C8 = colony 1 – colony 8.

BioBrick cloning was used to multiplex the sgRNA cassettes, which included the *ermE** promoter, the spacer, the guide RNA and a T7 terminator. Creating a plasmid with multiple sgRNA cassettes involved digesting the recipient plasmid with BglII and EcoRI and the donor plasmid with BamHI and EcoRI, flanking the cassettes as BioBrick parts (**Figure 4.3**). These were analysed on an agarose gel, cut out and DNA purified (**Figure 4.4**). BglII and BamHI have compatible restriction sites, therefore ligation of the donor and recipient insert, and backbone sequences recreated the restriction sites. The process of cloning individual parts was repeated to create pEM3.1, pEM3.2, pEM3.3, pEM3.1.2 (targets *tap-tpg*$_4$ and *tap-tpg*$_3$), pEM3.1.3 (targets *tap-tpg*$_4$ and *tap-tpg*$_2$), pEM3.2.3 (targets *tap-tpg*$_3$ and *tap-tpg*$_2$) and pEM3.1.2.3 (targets *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$).

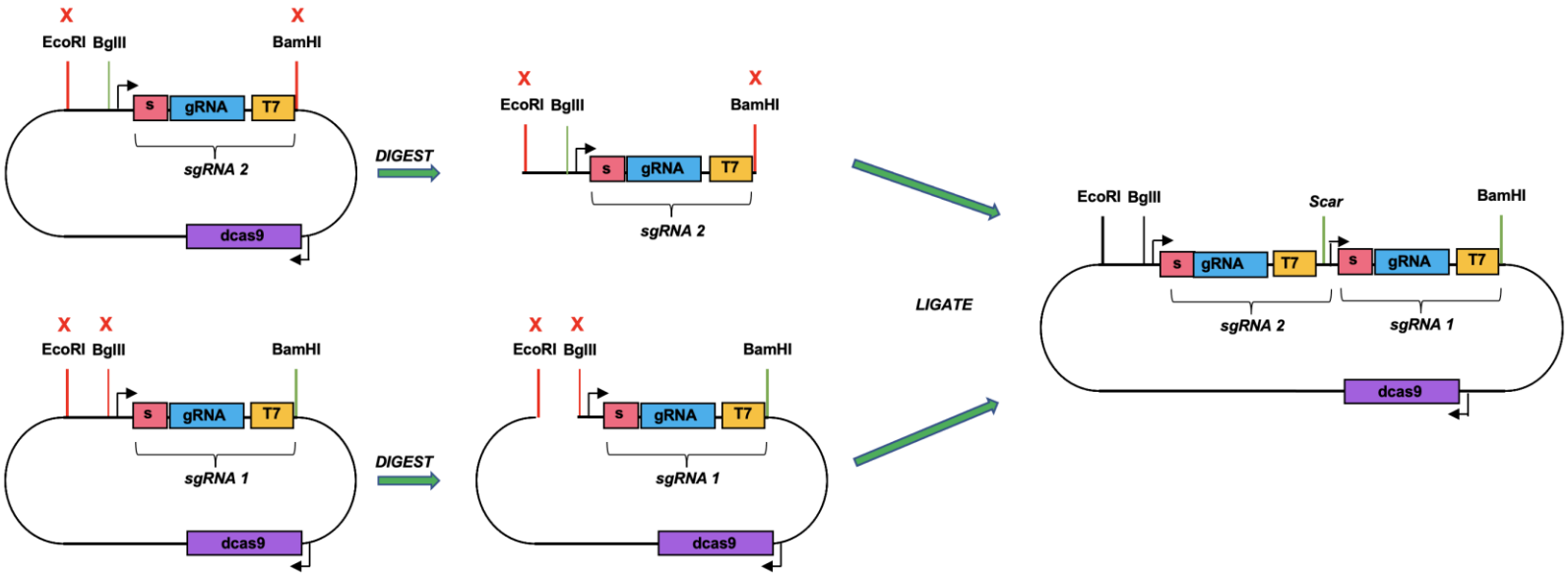


Figure 4. 2. Overview of the BioBrick assembly method for combination of multiple sgRNA cassettes. The recipient plasmid (pEM3.1) was digested with BglII and EcoRI and the donor plasmid was digested with BamHI and EcoRI allowing for the cloning of the *ermE**, sgRNA and T7 terminator. Linearised plasmids were ligated together to form the multiplexed plasmid, leaving a BamHI/BglII scar sequence. This process was repeated until the correct number and combination of sgRNA cassettes were cloned into the recipient plasmid Adapted from Larson et al., (2013).

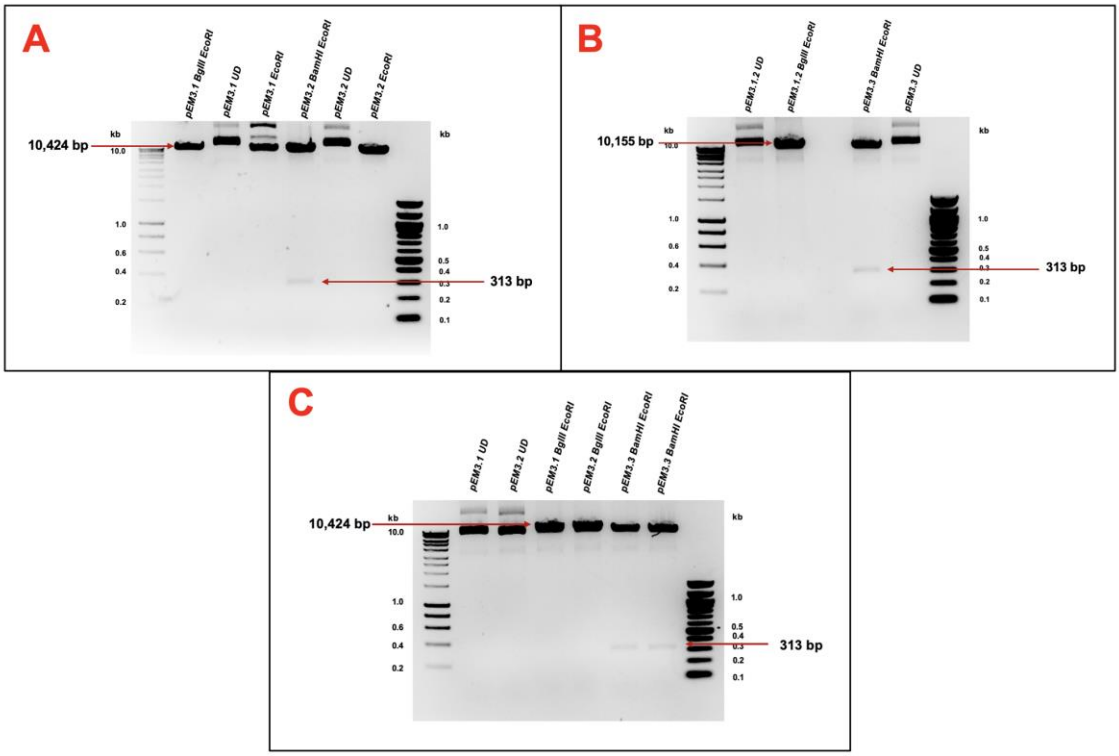


Figure 4. 3. Gel electrophoresis of digested BioBrick parts for the creation of multiplexed sgRNA plasmids targeting all *tap-tpgs* on GLPs in *S. clavuligerus*. A) Agarose gel showing 10,424 bp pEM3.1 recipient digest with BglII/EcoRI and 313 bp pEM3.2 donor digest with BamHI/EcoRI used to create pEM3.1.2. **B)** Agarose gel showing 10,155 bp pEM3.1.2 recipient digest with BglII/EcoRI and 313 bp pEM3.3 donor digest with BamHI/EcoRI used to create pEM3.1.2.3. C) Agarose gel showing 10,424 bp pEM3.1 and pEM3.2 recipient digest with BglII/EcoRI and 313 bp pEM3.3 donor digests with BamHI/EcoRI used to create pEM3.1.3 and pEM3.2.3.

4.2 Conjugation with CRISPR-dCas9 multiplexed plasmids highlights loss of *S. clavuligerus* replicon ends.

We aimed to identify the effects of multiple knockdowns of *tap-tpg* in *S. clavuligerus* and to determine the relationship of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ with the five replicons, therefore CRISPR-dCas9 plasmids pEM3.1, pEM3.2, pEM3.3, pEM3.1.2, pEM3.1.3 and pEM3.1.2.3 were transformed into *E. coli* ET12567/pUZ8002 and subsequently conjugated with our recipient organism, *S. clavuligerus* DSM 738. Conjugation followed an optimised method which used fresh recipient spores (shown in chapter 3.0) and allowed the donor *E. coli* cells to grow to an OD_{600} of 0.5. This resulted in a number of exconjugant colonies with genotypes highlighted in **Table 4.1**. The map of a representative plasmid is shown in **Figure 4.5**, which shows the pGWS1370-based plasmid pEM3.1, including the pGWS1370 backbone, *ermE** promoter, sgRNA scaffold and T_7 terminator.

Table 4. 1. Multiplexed CRISPR-dCas9 plasmids and their *S. clavuligerus tap-tpg* target.

Plate	***Plasmid***	***Spacer Target***
1	pEM3.1	***tap-tpg*$_4$**
2	pEM3.2	***tap-tpg*$_3$**
3	pEM3.3	***tap-tpg*$_2$**
4	pEM3.1.2	***tap-tpg*$_4$, *tap-tpg*$_3$**
5	pEM3.1.3	***tap-tpg*$_4$, *tap-tpg*$_2$**
6	pEM3.2.3	***tap-tpg*$_3$, *tap-tpg*$_2$**
7	pEM3.1.2.3	***tap-tpg*$_4$, *tap-tpg*$_3$, *tap-tpg*$_2$**

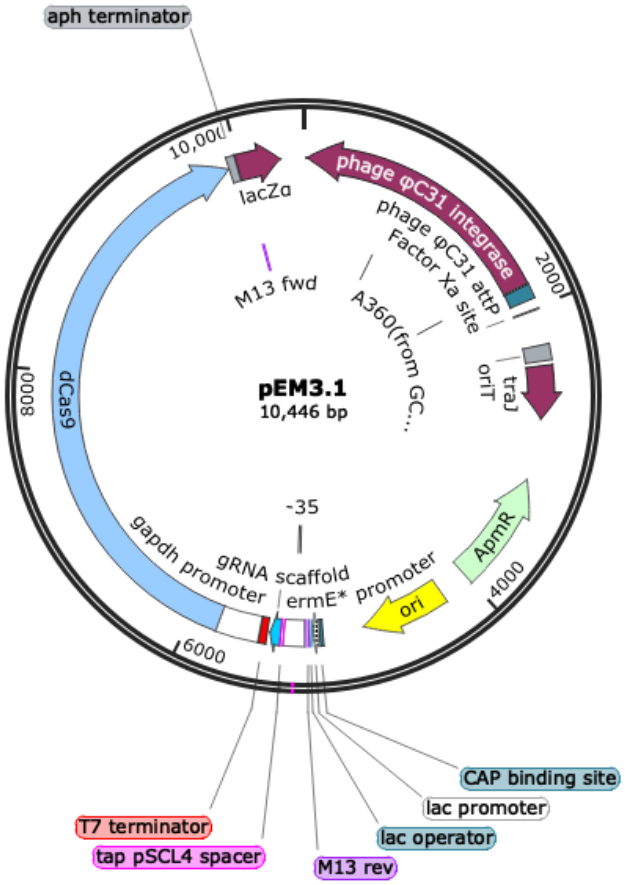


Figure 4. 4. Diagram of pEM3.1 plasmid with a *tap-tpg₄* specific spacer sequence upstream of the gRNA scaffold and T7 terminator. For pEM3.1 to pEM3.1.2.3 plasmids the only difference lies in the number of spacer sequences, *ermE** promoters and T7 terminators (more spacer sequences = more promoter and terminator sequences as each cassette needs its own promoter and terminator to allow for gene silencing).

10 single colonies were picked for each of the 7 conjugation plates highlighted in **Table 4.1** and spore stocks were prepared of 70 colonies in total. Initially, pEM3.1 (*tap-tpg*$_4$ targeting*)* transconjugants were screened for the loss of the chromosome and pSCL4 ends as this would indicate replicon linearity. Similar to the mutant screening in chapter 3.0, pEM3.1 transconjugant mutants were screened via colony PCR (**Figure 4.5**). This generated mixed results due to the filamentous nature and high %GC content of *Streptomyces*, meaning that DNA isolation was not straightforward, but the colonies needed to be extensively boiled, vortexed, centrifuged and resuspended in water.

The colony PCR checked pEM3.1 transconjugant replicon linearity, and primers amplified approximately 300-600 base pairs of chromosomal and pSCL4 telomere ends. There were no amplicons for the majority of picked mutants, but there was an amplicon for the WT *S. clavuligerus* gDNA and for pEM3.1 colony 6 for the LHS/RHS of the chromosome ends as well as the LHS pSCL4 end (**Figure 4.5 A, B, C**). The presence of an amplicon suggested linearity and full-length telomeres. The absence of an amplicon for all other pEM3.1 transconjugants indicated that parts of the telomeres were missing and have shortened. This could also mean that the chromosome and pSCL4 could have shortened because they circularised, however the circularisation point at which LHS/RHS replicon ends met was unclear and was further clarified through Illumina sequencing in section 4.3.

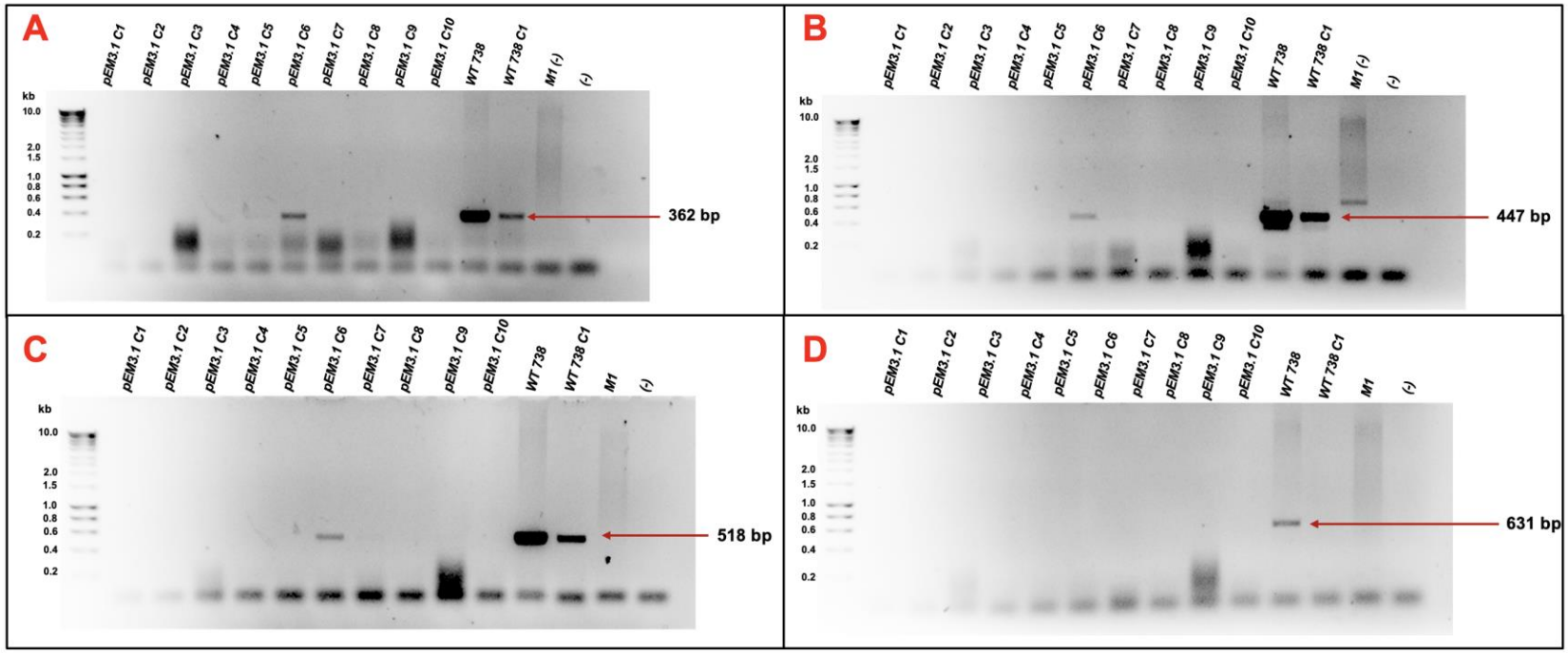


Figure 4. 5 PCR highlighting LHS and RHS chromosome and pSCL4 telomere end amplification for CRISPR-dCas9 multiplexed mutants. A) Amplification of LHS chromosome ends for *S. clavuligerus* DSM 738 with integrated vector pEM3.1. **B)** Amplification of RHS chromosome ends for *S. clavuligerus* DSM 738 with integrated vector pEM3.1. C) Amplification of LHS pSCL4 ends for *S. clavuligerus* DSM 738 with integrated vector pEM3.1. **D)** Amplification of RHS pSCL4 ends for *S. clavuligerus* DSM 738 with integrated vector pEM3.1. C1-C10 = pEM3.1 transconjugant colony 1- pEM3.1 transconjugant colony 10.

The colony PCRs in **Figure 4.5** provided a means of identifying potential mutants by demonstrating the presence or absence of telomeres for pEM3.1, however the time of screening was cut down by selecting a range of pEM3.1, pEM3.2, pEM3.3, pEM3.1.2, pEM3.1.3, pEM3.2.3 and pEM3.1.2.3 mutants to be taken forward for Illumina sequencing. Therefore, we picked a total of 8 transconjugant colonies representative of all multiplexing genotypes (**Table 4.2**), which were named EM2-EM10 in addition to the pGWS1370 transconjugant vector controls. The genomes of these 8 transconjugants contained the integrative, multiplexed CRISPR-dCas9 plasmids, and represented the full spectrum of potential knockdown effects.

So far, phenotypic signs in *S. clavuligerus* such as lack of sporulation or patchy, white colonies indicated unhealthy strains especially when considering CRISPR knockdown mutants (Chapter 3.0). We suspected that the variations in phenotype were due to adverse, off-target, toxic effects of both systems or unspecific binding of the gRNA (Rostain *et al.*, 2023). Therefore, additional factors for choosing these transconjugant colonies were differences in colour, odour, growth pattern and baldness. The final decision to choose the transconjugants mentioned in **Table 4.2** was based primarily on their phenotypic characteristics, and whether the colonies grew from a previously aliquoted spore stock. It was clear that some mutants exhibited phenotypes very similar to the WT while others did not, irrespective of their genotypes (**Figure 4.6**). Most importantly, we observed that these mutants were all morphologically different when compared to each other, despite some of them (EM2/EM3:*tap-tpg*$_4$ and EM6/EM7:*tap-tpg*$_4$&*tap-tpg*$_3$) containing CRISPR-dCas9 plasmids with the same *tap-tpg* targets.

Table 4.2. Mutant strain names for *S. clavuligerus* DSM 738 conjugated with multiplexed CRISPR-dCas9 plasmids.

Mutant	*Plasmid/Colony*	*CRISPR-dCas9 tap/tpg target*
EM2	pEM3.1 c6	pSCL4
EM3	pEM3.1 c10	pSCL4
EM4	pEM3.2 c9	pSCL3
EM5	pEM3.3 c2	pSCL2
EM7	pEM3.1.2 c9	pSCL4 + pSCL3
EM8	pEM3.1.3 c8	pSCL4 + pSCL2
EM9	pEM3.2.3 c6	pSCL3 + pSCL2
EM10	pEM3.1.2.3 c7	pSCL4 + pSCL3 + pSCL2
1370	pGWS1370	dCas9 control

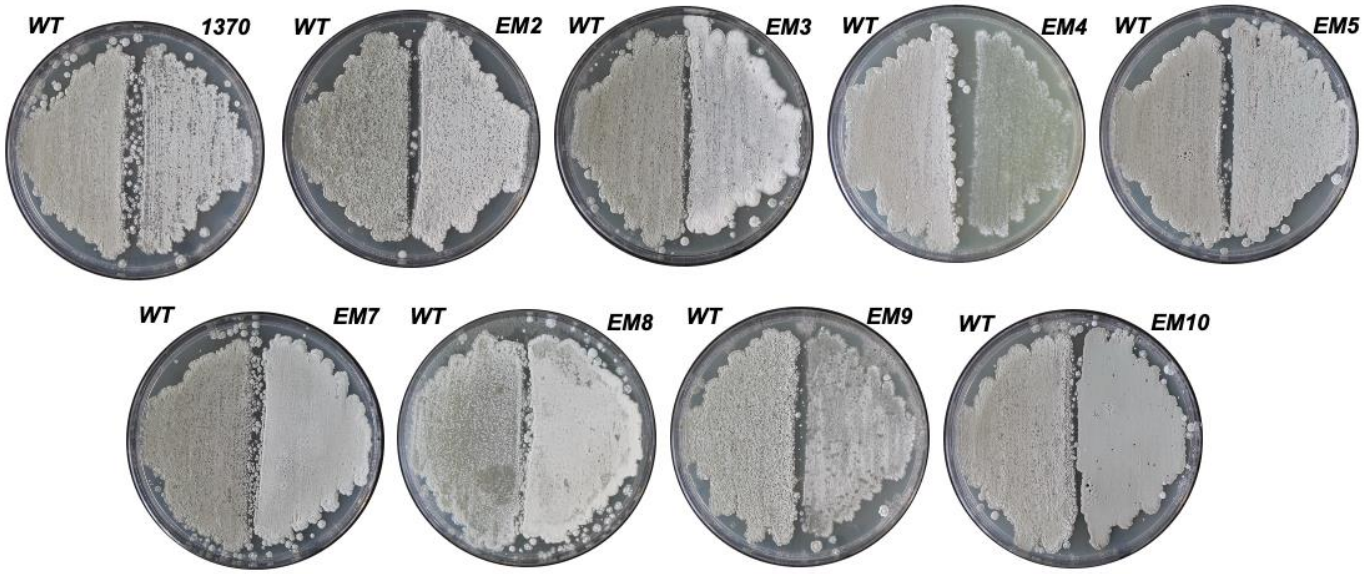


Figure 4. 6. Morphological differences between various *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ knockdown mutant strains. EM2 - EM10 in addition to 1370 (pGWS1370 vector control) were streaked alongside the *S. clavuligerus* DSM 738 WT strain on L3M9 agar and grown at 26°C for 2 weeks. The WT strain presents a green colour indicating sporulation and some patchiness, which is mirrored by EM5, EM7 and EM10. A bald lawn can be seen for EM4, which struggled to grow in liquid medium. White-coloured mycelia can be seen for 1370, EM2, EM3, EM8 and EM9 highlighting sub-optimal growth conditions and potential genetic instability.

4.3 Illumina Sequencing proves loss of entire plasmids and replicon ends from CRISPR-dCas9 mutant strains.

The results of the mutant colony PCRs screening for loss of chromosome and pSCL4 ends, highlighted a loss of telomere ends for most colonies (**Figure 4.5**). Screening replicon ends using a method like colony PCR provided initial indications as to which EM mutants to exclude, however was not reliable for screening replicons for loss of ends and potential circularisation, as insufficient DNA isolation could also be misinterpreted as a negative (loss of replicon ends) result. Out of the 8 chosen strains, genomic DNA from 8 were prepared and sent to MicrobesNG for Illumina sequencing to evaluate whether there was any replicon shortening and/or circularisation (EM3, EM4, EM5, EM7, EM8, EM9, EM10, 1370). The strains were chosen solely based on their genotype, i.e., each one of the mutants had a different, multiplexed plasmid integrated within its chromosome as depicted by **Table 4.3**. This allowed for a wider analysis of *tap-tpg* knockdown effects on replicon dynamics and general characteristics. Overall, Illumina sequencing results showed that even though some mutants had the same CRISPR-dCas9 *tap-tpg* knockdown plasmids, they exhibited a difference in lost plasmids, also illustrated by differences in strain morphologies (previous section).

4.3.1 *tap-tpg*$_4$ transcriptional silencing leads to loss of multiple GLPs.

In order to further investigate how the lost plasmids compared to the WT strain, we mapped the EM mutant reads onto the *S. clavuligerus* genome Algora-Gallardo *et al.,* (2023) using BWA-MEM2 on Galaxy and visualised the read mapping for each mutant on the entire replicon using IGV and JBrowse2. Table 4.3 summarises the results of read mapping with particular focus on the missing plasmids from EM mutants and controls. There was no trend between which *tap-tpg* operon was targeted and which plasmids were missing, although other trends were clear: only the *tap-tpg*$_4$ targeting CRISPR-dCas9 sgRNA allowed for combined pSCL3, pSCL2 and pSCL1 replicon loss, underlining the importance of the *tap-tpg* operon of pSCL4. The loss of pSCL3, similarly to pSCL1, can only be witnessed for mutants with sgRNA targeting *tap-tpg*$_4$, highlighting that there could be similarity between pSCL3 and pSCL4 *tap-tpg* genes. Additionally, only the *tap-tpg*$_4$ targeting CRISPR-dCas9 sgRNA lead to a loss of pSCL1, potentially due to the evolutionary history of pSCL4 i.e., the recombination of pSCL1 and the chromosome (Medema *et al.*, 2010).

Table 4.3. Illumina sequencing results of CRISPR-dCas9 multiplexed mutants EM3-EM10 showing the presence and absence of replicons.

Mutant	*CRISPR-dCas9 tap/tpg target*	*Missing Plasmids*
EM3	pSCL4	pSCL3, 2, 1
EM4	pSCL3	pSCL2
EM5	PSCL2	pSCL2
EM7	pSCL4 + pSCL3	pSCL2, 1
EM8	pSCL4 + pSCL2	pSCL3, 2, 1
EM9	pSCL3 + pSCL2	pSCL2
EM10	pSCL4 + pSCL3 + pSCL2	pSCL3, 2
1370	dCas9 control	pSCL2

4.3.2 CRISPR-dCas9 system results in loss of pSCL2 in *S. clavuligerus* DSM 738

We observed another trend from the Illumina sequencing results: the loss of pSCL2 for all of the CRISPR-dCas9 based mutants. This included both the EM strains and the 1370 control, which did not have an integrated *tap-tpg* specific spacer and was used as a dCas9 vector control. As pSCL2 was lost from 1370, we speculated that the loss of pSCL2 was due to CRISPR-dCas9 off-target effects, as this strain did not have a targeted *tap-tpg* spacer. The potential for the gRNA sequence itself to share enough homology with the target DNA have contributed to higher chances of Cas9 off-target effects (Zhao *et al.*, 2017). sgRNA targeting specificity (during pre-binding, binding, and cleavage) can be affected by many factors, such upstream seed sequence complexity, especially when looking at Cas9 cleavage systems. Little has been researched about off-target effects of the sgRNA complex together in dCas9 systems, however the high GC content of the *S. clavuligerus* replicons and the complexity of terminal repeats could have been contributing factors for off-target dCas9 effects, particularly when examining pSCL2 as the LHS and RHS terminal sequences share the most homology compared to pSCL1, pSCL3 and pSCL4 (Algora-Gallardo, unpublished data).

4.4 Mapping of EM mutant reads to *S. clavuligerus* DSM 738 genome confirms missing replicon ends.

To examine the EM mutants more closely and decipher how *tap-tpg* silencing could have resulted in the loss of entire plasmids, the loss of telomere ends was considered. This has previously been observed in *S. clavuligerus* mutants where another replicon has been lost (Algora-Gallardo, unpublished data). For this, we mapped our Illumina reads for EM mutants against the reference *S. clavuligerus* genome and observed a) the loss of replicons and b) that replicons which had not been lost showed a low mapping quality (MAPQ) number towards their telomere ends. **Table 4.4.** highlights the number of base pairs missing from the LHS and RHS of the chromosome, pSCL4, pSCL3 and pSCL1 for all EM mutants. The number of bp was counted by disregarding reads at the telomere ends that had a MAPQ value of 0. MAPQ numbers showed the probability of misplaced reads due to poor base quality, bad alignment, or ambiguity. A larger MAPQ indicates a higher chance of the alignment being unambiguous and this was accounted for by increasing BWA-MEM2 (Galaxy Europe) seed-lengths, as a longer seed length setting in BWA produces a more unambiguous and stringent alignment (Deng *et al.*, 2022).

High read coverage and MAPQ quality were an indicator of correct mutant alignment to the reference genome. The reads showed full coverage (MAPQ=60) for all retained replicons, particularly towards the centre – a phenomenon also observed by Algora-Gallardo, (unpublished data). We observed low MAPQ values at replicon ends for EM mutants which can be explained by looking at the underlying biology behind linear replication: the replication fork starts at the centre of the replicon and moves bi-directionally towards both ends. This highlights another potential reason why parts of the replicon end had no reads mapping against the reference genome or why the reads were of such low mapping quality: inconsistent segregation of replicons.

After the mutant reads were mapped onto the *S. clavuligerus* genome (Algora-Gallardo, unpublished data), the BAM alignment files were condensed into BigWig files. **Table 4.5.** provides an overview of the dCas9 mutants compared to the WT, including the M5 and 1370 controls. Similar trends can be witnessed for missing telomere ends between mutants with the same lost plasmids, highlighted by matching colours on **Table 4.4** and **Table 4.5**. An example of this was seen for mutants that had only lost pSCL2 (EM4, EM5 and EM9), resembling the 1370 vector control, which showed similar numbers of bp lost from the ends of the chromosome, pSCL1 and pSCL3, highlighted in blue in **Table 4.5.** Interestingly, the WT reads mapped up until the replicon ends with MAPQ values above

0, which was why as little as 6 bp missing from replicon ends was considered noteworthy for the EM mutant and 1370 control strains.

Two strains were used as controls when mapping EM mutant reads to the S*. clavuligerus* reference which were the 1370 vector control strain and a previously unsequenced strain derived from work by Algora-Gallardo, (unpublished data) called M5, missing pSCL3 and pSCL4. 1370 and M5 were used to determine the effects of CRISPR-dCas9 and a targeted *tap-tpg* spacer on replicon loss (1370), but also as a point of comparison and a mutant with varying lost replicons (M5). M5 was created with CRISPR-Cas9, using a high copy number *Streptomyces* expression plasmid (pGE240 derived from pIJ86), which caused a double stranded break at *parB*, encoding the essential plasmid partitioning protein ParB. Additionally, Cas9 systems are prone to more off-target, toxic effects when compared to dCas9 systems, as they do not just cause gene silencing, but they cause DSBs at target sites, leading to homologous recombination (potentially the cause of pSCL4 loss). **Table 4.4** showed that M5 lost substantial chunks of the chromosome ends, unlike the EM mutant strains.

Our decision to target *tap-tpg* in this chapter was based on work by Algora-Gallardo, (unpublished data), which showed that the loss of *tap-tpg*$_4$ due to pSCL4 loss resulted in chromosome circularisation. Algora-Gallardo, (unpublished data) also showed that the loss of a large number of base pairs from *S. clavuligerus* chromosome ends was an indicator of circularisation, which will be looked at later in this chapter for EM mutants. M5 showed a loss of 210,556 bp from the RHS of its chromosome ends. Comparatively, the highest number of bp lost from EM mutant replicon ends was 16 bp from the RHS telomere of EM9. **Table 4.4** also shows that the highest number of bp lost from pSCL4 telomeres was seen for EM7 (167 bp from RHS), the highest number of bp lost from pSCL3 replicon ends was seen for EM5 (889 bp from RHS) and the highest number of bp lost from pSCL1 telomeres was seen for EM10 (78 bp from LHS).

Table 4.4. Missing LHS and RHS telomere base pairs of CRISPR-dCas9 mutants and dCas9 vector (1370) and pSCL4&pSCL3 minus control (M5).

	EM3	*EM4*	*EM5*	*EM7*	*EM8*	*EM9*	*EM10*	*1370*	*M5*
Chr LHS	19	10	9	8	9	9	13	9	77,850
Chr RHS	15	9	8	6	8	16	0	9	210,556
pSCL4 RHS	81	7	9	9	8	9	9	16	-
pSCL4 LHS	47	7	62	167	7	6	9	14	-
pSCL3 RHS	-	759	799	755	-	822	-	832	-
pSCL3 LHS	-	853	889	888	-	887	-	616	-
pSCL1 LHS	-	7	59	-	-	6	78	14	10
pSCL1 RHS	-	8	7	-	-	12	32	11	11

Table 4.5. Overview of CRISPR-dCas9 mutants and controls highlighting similar trends in missing telomeres between strains with the same lost plasmids.

Mutant	*pSCL1*	*pSCL2*	*pSCL3*	*pSCL4*	*Chr*
EM3	O	O	O	X	X
EM4	X	O	X	X	X
EM5	X	O	X	X	X
EM7	O	O	X	X	X
EM8	O	O	O	X	X
EM9	X	O	X	X	X
EM10	X	O	O	X	X
1370	X	O	X	X	X
M5	I	I	O	O	X
WT Control	I	I	I	I	I

X =	missing telomere ends
O =	missing plasmid
I =	intact plasmid

4.5 Missing chromosome ends on M5 lead to loss of structural proteins and proteins involved in secondary metabolism.

Work by Algora-Gallardo, (unpublished data), showed that the loss of pSCL4 lead to chromosomal circularisation, highlighted by the loss of replicon ends. We therefore hypothesised that the shortening of M5 chromosome ends was a direct effect of loss of pSCL4. The reasons behind chromosomal shortening remained unclear and we wondered whether chromosomal shortening of M5 was a result of pSCL4 missing, or whether shortening was a result of certain genes, such as *tap-tpg₄,* missing from pSCL4. Our mutants lost fewer bp from their chromosome ends (lost > 10 bp) when compared to M5 (lost >1000 bp), underlining the hypothesis that the loss of pSCL4 leads to chromosomal shortening. The mutants created in this chapter contained targeted *tap-tpg* spacers and used dCas9, meaning that once the spacer and gRNA bound to the target gene, gene expression was silenced and there was no need for homologous or non-homologous end joining (NHEJ). M5 was created by conjugating *S. clavuligerus* with a Cas9-based plasmid, thus the genetic shifts and loss of pSCL4 were a direct impact of the genome plasticity resulting from the Cas9 induced DSB (Hoff *et al.*, 2018). To allow us to understand how the loss of chromosomal ends affected M5, we took a closer look at the Illumina sequences from regions of the chromosome where no reads mapped to mutant Genbank files using IGV, and determined what protein-coding genes were lost from the terminal regions of the chromosome.

Figure 4.7 zooms in on the missing chromosome ends for EM mutants and M5. Unlike the WT, 1370, and all the derived EM mutants, the M5 Illumina reads show the loss of large regions of chromosome ends. M5 lost a total number of 16 genes encoding a variety of structural proteins and proteins involved in secondary metabolism from the LHS of the chromosome. These included several type I polyketide synthase proteins involved in fatty acid synthesis such as acyl-carrier proteins e.g., S-malonyl-transferases and cytochrome P450 enzymes, which play a role in membrane phospholipid and glycolipid formation for *Streptomyces spp*. Type I polyketide synthases are important for secondary metabolite production, however, they are not considered essential for strain survival. M5 lost a total number of 159 protein coding genes from the RHS chromosome, some of which were categorised into three types of biosynthetic gene clusters: terpenes, non-ribosomal peptide synthetase and type I polyketide synthases. Several proteins were involved with secondary metabolite production such as aminohydrolases, antibiotic biosynthesis monooxygenases involved in polyketide biosynthesis, peptidoglycan

binding proteins and homoserine O-acetyltransferases were identified, highlighting that a higher number of biosynthetic gene clusters were lost from the M5 RHS chromosome end compared to LHS chromosome end. The loss of genes playing a role in secondary metabolite production vs primary metabolism from chromosome ends suggests the telomere loss observed in M5 may have had effects on the strain's phenotype, but not on the vegetative mycelial growth and enzyme production during the first stages of the *S. clavuligerus* life cycle, such as enzymes involved in glycolysis and the tricarboxylic acid cycle.

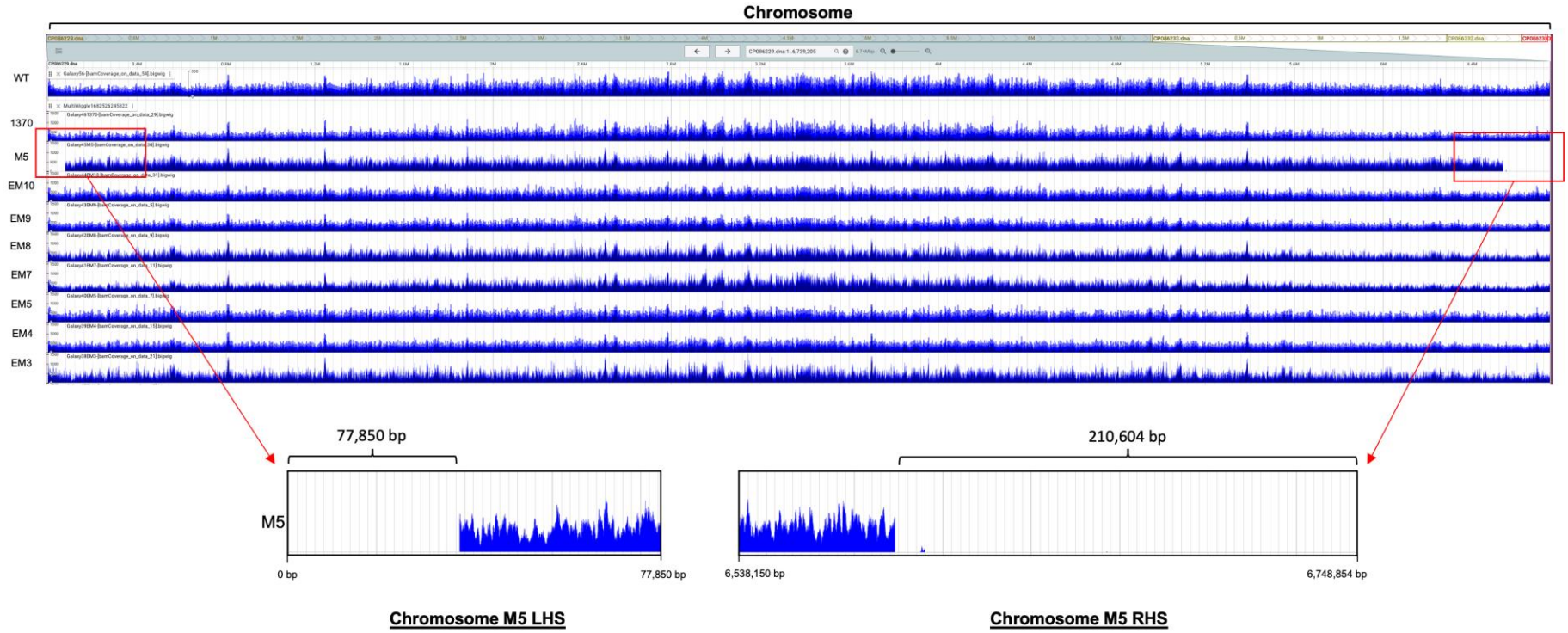


Figure 4. 7. Sequence reads from CRISPR-dCas9 mutants mapped onto the assembled *S. clavuligerus* DSM 738 chromosome. Missing chromosome ends can be seen for the M5 mutant. View from JBrowse2. Chromosome length = 6,748,854 bp.

4.6 Sequence Read islands on pSCL4 and pSCL3 are a result of transposable elements.

In the previous section Illumina sequencing reads were investigated and areas with unmapped reads (EM mutants) or missing reads (M5) were mapped against *S. clavuligerus* DSM 738 focussing on the chromosome. In this section we focus on pSCL4 and pSCL3 lllumina reads shown by **Figure 4.8** which depicts pSCL4 and pSCL3 read coverage. When we mapped the Illumina reads to pSCL4, we observed mapped reads for the WT and all EM mutants while no reads mapped for M5. An interesting observation we made was that the M5 mutant showed a loss of pSCL4, however small peaks were visible at around 1,405,500 bp – 1,406,850 which were also present for the WT (**Number 1, Figure 4.8**). We thought the peaks could correspond to ribosomal RNA (rRNA) encoding operons, i.e., the peaks could be a result of shared sequence similarity of M5 and the WT.

To determine what proteins are encoded by the M5 and chromosome Illumina sequence peaks **Number 1, Figure 4.8**, the peak areas were analysed using GenBank files. The Genbank sequence showed that this region primarily included an IS110-like element IS116 family transposase with terminal inverted repeats flanking the gene and is not uncharacteristic for an organism with such random genetic variability as shown by Leskiw *et al.*, (1990), who witnessed this insertion for the first time. After performing a nucleotide BLAST on the sequence, the transposase had a 100% sequence identity and E-value of 0.0 to various regions on *S. clavuligerus* plasmids, such as ATCC 27064 pCLA1, the F1D-5 pSCL1 plasmid as well as a few other F1DN strain pF1DNA plasmids, highlighting that the peaks witnessed are not unusual and correspond to a common transposable element in *S. clavuligerus*.

To investigate whether these transposable elements and islands of reads were seen on other replicons within the *S. clavuligerus* DSM 738 genome, we created a command line BLAST database. The BLAST database contained the *S. clavuligerus* DSM 738 genome as the reference file, and fasta files of the query sequences, such as the IS110-like element IS116 family transposase segment, were used to screen for significant alignments. In the case of the above-mentioned transposase, which was seen for the M5 mutant on the missing plasmid pSCL4, the peak sequence (**number 1**, **Figure 4.8**) was identical to a region on pSCL2 with a similar peak, showing that the pSCL4 transposon inserted on pSCL2 and why reads (peaks) mapped to pSCL4, despite the GLP being cured.

Figure 4.8 showed that there was almost double the pSCL4 read coverage for EM7 in first 743,730 bp of the replicon compared to the rest of the replicon (**number 2**, **Figure 4.8**). More specifically, an area of higher coverage (743,730 bp in total) was seen towards the start of pSCL4 for EM7 and an area of low coverage (691,524 bp in total) was seen towards the end of the replicon with the centre of the replicon presenting average coverage compared to the other mutants and WT. Reasons behind the difference in coverage remain unclear, but could perhaps be due to more repetitive sequences on the LHS of pSCL4 compared to other regions. Additionally, Medema *et al.*, (2010) suggested that pSCL4 did indeed stem from recombination of pSCL1 with the chromosome, so it is not surprising to see higher coverage for regions of the plasmid, perhaps due to inter-replicon similarities. We also speculated that a high number of genomic islands in that region could be a contributor to genome plasticity, an area which will be explored in Section 4.8.

Figure 4.8 also shows the missing pSCL3 replicon from M5, EM10, EM8 and EM3. Unlike M5, EM10, EM8 and EM3 are CRISPR-dCas9 mutants, which have a spacer silencing *tap-tpg*$_4$ only*, tap-tpg*$_4$ & *tap-tpg*$_2$*, and tap-tpg*$_4$*, tap-tpg*$_3$ & *tap-tpg*$_2$, respectively. Mutants that retained pSCL3, like EM4, EM5 and EM7, showed a loss of pSCL3 ends (**Table 4.4**: around 700-900 bp from LHS and RHS). Additionally, the loss of pSCL3 ends occurred for all mutants which also lost pSCL2, suggesting that the loss of pSCL2 leads to missing pSCL3 ends. Mutants which lost pSCL3 also presented read islands, similar to read islands when pSCL4 was lost (M5). The small read islands are shown on **Figure 4.8** by **number 3** and **4** red boxes, and the read island size was consistent in the EM mutants, mapping over 6,850 bp (**number 3**, **Figure 4.8**) and 2,000 bp (**number 4, Figure 4.8**) for M5, EM3, EM8 and EM10.

The 6,850 bp long on pSCL3 (**number 3, Figure 4.8**) was investigated using the WT GenBank file, and the hypothetical proteins in the region were identified using BLASTp. Proteins included in the region spanning from 251,750bp – 258,600 bp consisted of peptide biosynthesis proteins (lasso peptides / cysteine proteases that play a role in antimicrobial activity and enzyme stability through cleaving precursor peptides), vacuolar proteins, bacterial co-enzymes such as PQQ synthesis protein D, transcriptional regulators, and many hypothetical proteins with >98% sequence similarities to other Actinobacteria. The largest protein coding region (**number 3, Figure 4.8**) was identified to have high sequence similarity to a IS*1182 Streptomyces* transposase, like results found for peaks seen on pSCL4 (**number 1**, **Figure 4.8**). The genes encoding IS*1182*

Streptomyces transposases contain conserved sequences and are often homologous to other parts of the *S. clavuligerus* genome, highlighting their genetic mobility as the encoded transposases allow for excision and insertion of mobile transposons.

Once again, the sequences of read islands present in **number 3 and 4** on **Figure 4.8** were screened using a command-line BLAST database. The 6,850 bp sized island on **number 3** on **Figure 4.8** was BLASTed against the *S. clavuligerus* DSM 738 genome and showed 12 sequences providing significant alignments, 9 of which had an E-value of 0.0, indicating that the sequences were identical. The **number 3** peak seen presented significant similarity to sequences on the chromosome, as well as pSCL3. There was an equivalent match on the chromosome for each hypothetical protein in the 6,850 bp read island highlighting genome plasticity and genome rearrangements due to CRISPR-dCas9 off-target effects, but also due to DNA transposases, which were present in the largest region of the **number 3** peak in **Figure 4.8**.

We also look at the encoded proteins in the 2,000 bp region of reads for strains that had lost pSCL3 (**number 4, Figure 4.8**), and similarly found that the read island was located over transposable elements. An identified IS*Azo13*-like element IS*Scl2* family transposase gene spanned over 1,695 bp of the 2,000 bp region highlighted, encoding the IS*Scl2* family transposase. The 2,000 bp read island (**number 4** on **Figure 4.8**) was BLASTed against the *S. clavuligerus* DSM 738 genome and showed four matching sequences producing significant alignments, two of which produced a higher score than the rest (100% match), although E-values were 0.0, or close to zero for all four significant alignments, highlighting their reliability. The four significant alignments from the BLAST database output showed similar outputs as for previous read islands (e.g., **number 3, Figure 4.8**) for mutants that had lost pSCL3, as the identified transposase contained an identical sequence on the chromosome. So, for each peak seen on pSCL3, there was also a sequence region of identical bases on the chromosome, highlighting transposition.

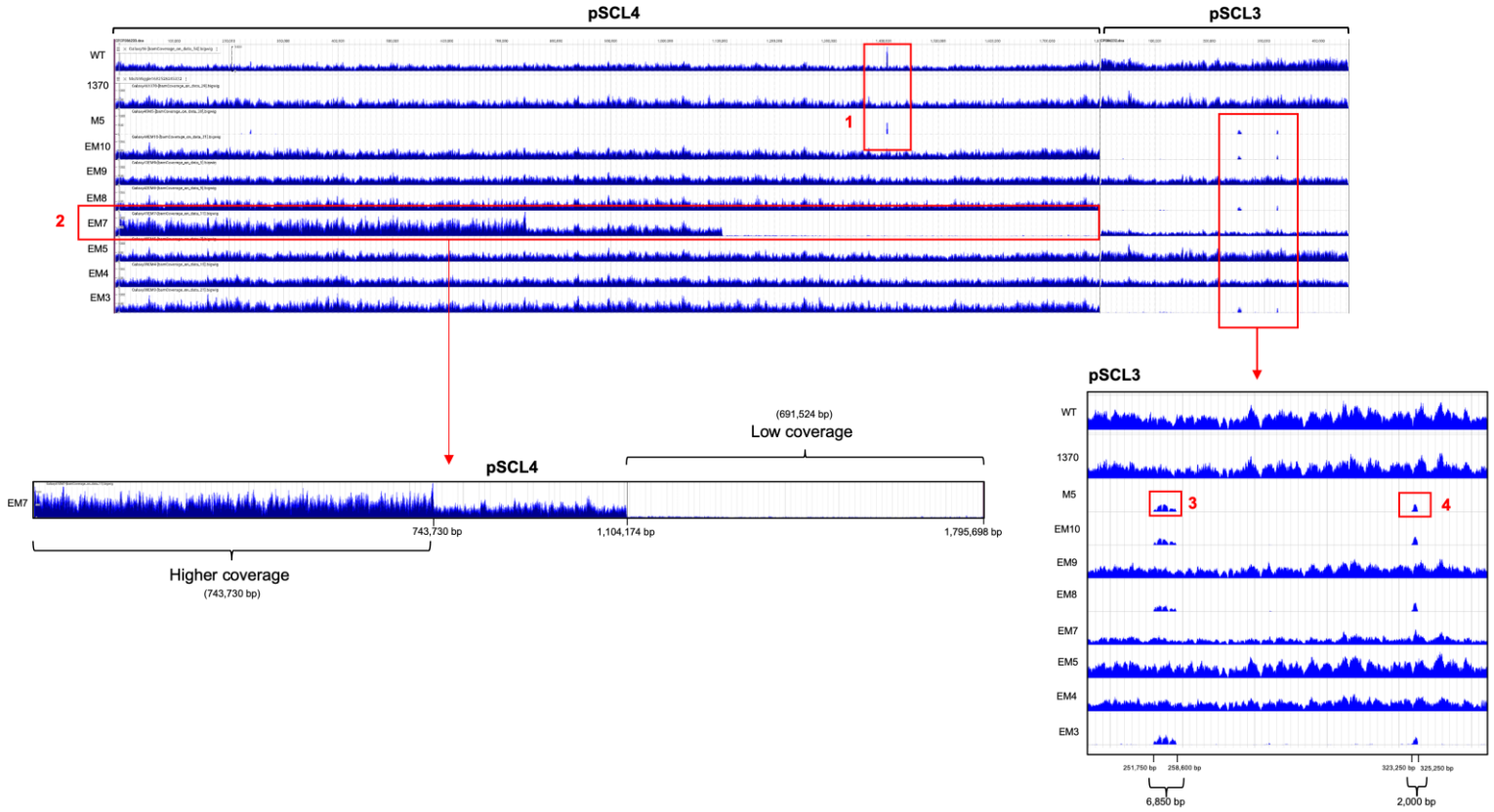


Figure 4. 8. CRISPR-dCas9 mutant strains sequence reads mapped onto *S. clavuligerus* DSM 738 pSCL4 and pSCL3. Illumina reads of the *S. clavuligerus* DSM 738 WT, 1370, M5 and EM mutant strains were mapped on the WT assembly highlighting pSCL4 and pSCL3 mapped reads and read islands. View from JBrowse2. 1= 1,405,500 bp – 1,406,850 bp.

4.7 Read islands on pSCL2 and pSCL1 are a result of transposable elements.

We determined that the read islands present in **Figure 4.8** were a result of transposon insertions and potential recombination events, which meant that segments of lost replicons had recombined and inserted themselves into other replicons, seen for our EM mutants and controls. In this section we map our reads onto pSCL2 and pSCL1 sequences, as shown by **Figure 4.9**, which depicts pSCL2 and pSCL1 read coverage after reads were mapped to the *S. clavuligerus* DSM 738 GLPs. We witnessed small islands (**numbers 1 and 2**, **Figure 4.9**), of mapped reads in mutants that lost pSCL2, which meant all mutants including 1370. These islands were 1,000 bp (**number 1, Figure 4.9**) and 1,200 bp (**number 2, Figure 4.9**) long and occurred in the first third and last third regions of the replicon. The 1,000 bp region of reads (**number 1, Figure 4.9**) included two protein coding genes, one of which was a WhiB family transcriptional regulator, an important transcription factor controlling sporulation and septation. The Wbl family was initially characterised in 1972 (Chater, 1972), and is found in Actinobacteria, sharing paralogues with other organisms within this phylum (Bush, 2018). The 1,000 bp read island seen for strains that had lost pSCL2 (**number 1, Figure 4.9**) was explored using the command-line BLAST database with *S. clavuligerus* DSM 738 as the reference genome, similarly to section 4.6, and results highlighted five sequences producing significant alignment. The WhiB transcriptional regulator on pSCL2 was homologous to the same regulator on pSCL4, suggesting that the pSCL2 genes encoding the regulator must have recombined with pSCL4. All mutants that had lost pSCL2, still maintained pSCL4 in their genomes, supporting the recombination events.

The 1,200 bp read island seen on pSCL2 (**number 2, Figure 4.9**) contained genes encoding a IS110-like element IS116 family transposase, like the transposase witnessed for pSCL4 read islands (**Figure 4.8**). Previously, IS116 like transposases have inserted themselves into the genome of *S. clavuligerus* through multi-copy plasmids such as pIJ702 (Leskiw *et al.*, 1990). The IS110-like element IS116 family transposase contained conserved regions resembling regions in other transposase proteins found in actinomycetes: IS900 of *Mycobacterium paratuberculosis* and IS110 from *S. coelicolor* A3(2) (Leskiw *et al.*, 1990), which highlights recombination events.

The 1,200 bp read island for strains that had lost pSCL2 (**number 2, Figure 4.9**) was further explored using the command-line BLAST database and results highlighted three sequences producing significant alignments with the pSCL2 read island, two of which had high alignment scores and were therefore considered. Both of these sequences

encoded the IS110-like element IS116 family transposase, which was homologous to the sequence on pSCL4 (E-value:0.0). Like the other transposase encoding genes highlighted in **Figures 4.8** and **4.9**, it confirms one of two options: there was no transposition, and the transposon exists in both replicons, as shown by the sequence identity between transposons on two replicons. However, the most likely reason behind the read islands on pSCL2 was mis-mapping of the reads and that the transposable elements previously existing on the lost replicon, i.e., pSCL2 had inserted into another replicon, i.e., pSCL4 in this case producing sequence homology.

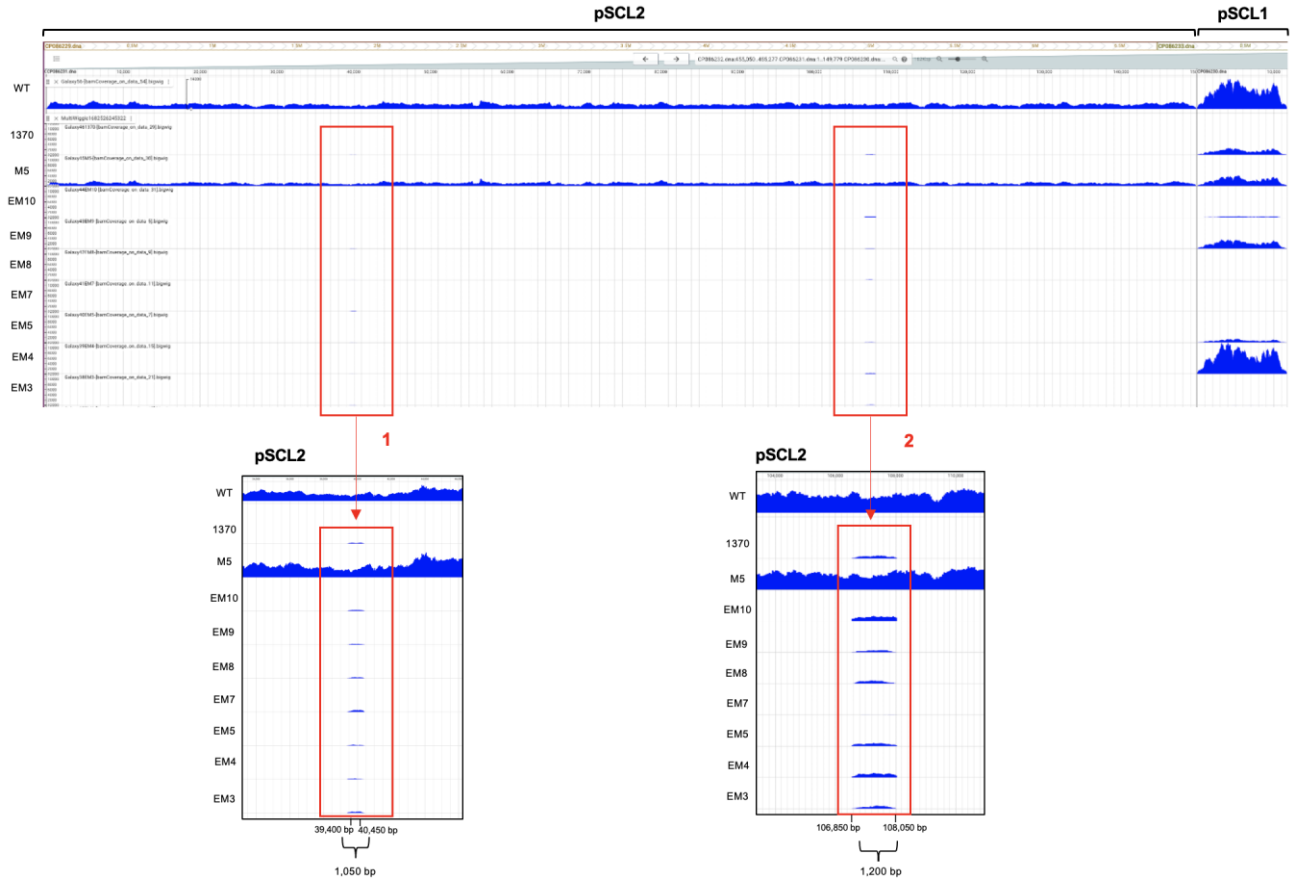


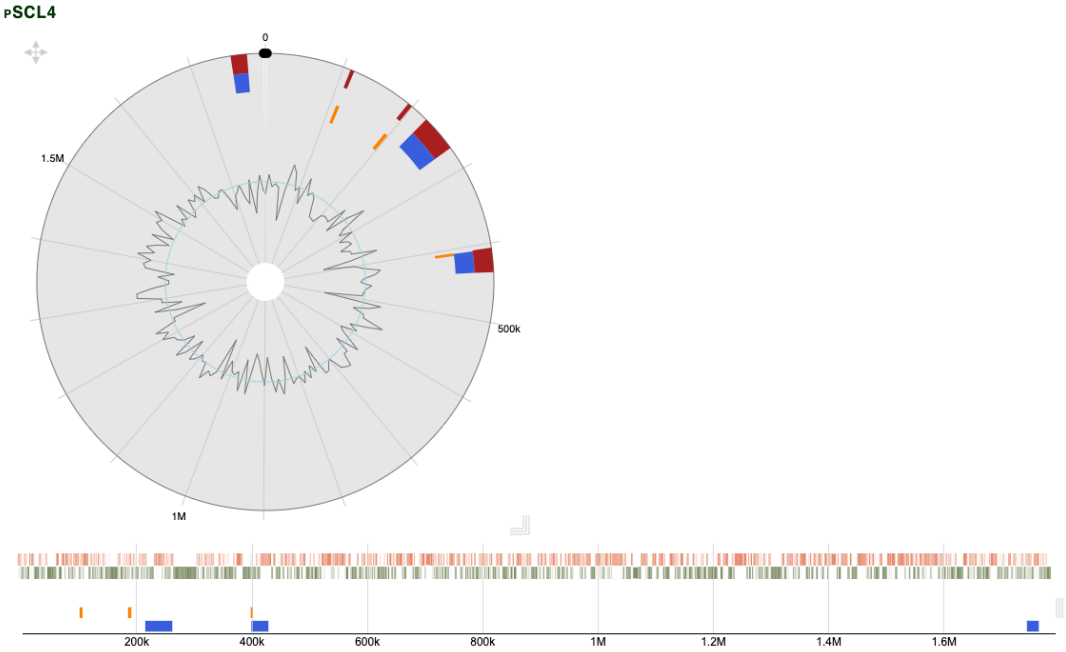
Figure 4. 9. CRISPR-dCas9 mutant strains mapped onto *S. clavuligerus* DSM 738 pSCL2 and pSCL1. Illumina reads of the *S. clavuligerus* DSM 738 WT, 1370, M5 and EM mutant strains were mapped on the WT assembly highlighting pSCL2 and pSCL1 mapped reads and read islands. pSCL2 length = 149,779 / pSCL1 length = 11,696.

4.8 Missing EM7 regions are genomic islands.

The Illumina read coverage for our EM mutants and controls remained consistent for replicons which had not been lost. The replicon ends presented lower read coverage, which we believed was due to TIR variability, and mapped Illumina sequences at the centre of the replicons generally showed higher read coverage, particularly visible by pSCL1 mapped reads in **Figure 4.9**. We discovered that one mutant, EM7, did not follow this trend and an area of higher coverage (743,730 bp in total) was seen towards the start of pSCL4 for EM7 and an area of low coverage (691,524 bp in total) was seen towards the end of the replicon (**Figure 4.8**). In section 4.6 we speculated that one reason behind a difference in read coverage was due to genomic islands i.e., larger sequence areas with evidence of horizontal gene transfer. To determine whether the small read islands (seen for mutants which had lost pSCL4, pSCL3, pSCL2) and the variable read coverage (seen on pSCL4 for EM7) was due to genomic islands, GenBank files were uploaded to Islandviewer 4 (Bertelli *et al.*, 2017). pSCL4, pSCL3 and pSCL2 GenBank files (CP086229.1 – 33.1, see Chapter 2) were uploaded individually with particular focus on the area of low coverage for EM7 on pSCL4 (**Number 2, Figure 4.8**).

Islandviewer 4 results predicted a total of 6 genomic islands for the first third of pSCL4, suggesting that the higher read coverage of the first third of EM7 could be due to recombination (**Figure 4.10 A** and **B**). The predicted genomic islands showed sequence islands encoded hypothetical proteins, various types of transcriptional regulators, beta-lactamase class C-like and penicillin binding proteins (PBPs), and, unsurprisingly, putative transposases. These putative transposases were positioned on the LHS of both blue coloured islands (**Figure 4.10 A**). Genomic islands generally surround a core genome (centre of replicon normally for *Streptomyces*), highlighting signs of present or past mobility (Juhas *et al.*, 2009). This mobility could stem from the any form of transfer of genetic material like conjugation, transduction or transformation largely depending on the transport system (i.e., type IV secretion system in *Streptomyces*). Collectively, these observations provided us with a better understanding about the increased read coverage seen in the first third of pSCL4 for EM7.

A



B

Start	End	Size	GI Prediction Method	View	External Annotations	Download (all data)
109,304	114,236	4,932	Predicted by at least one method			Sequence, Genes, Proteins
193,165	198,912	5,747	Predicted by at least one method			Sequence, Genes, Proteins
223,055	269,991	46,936	Predicted by at least one method			Sequence, Genes, Proteins
405,894	409,957	4,063	Predicted by at least one method			Sequence, Genes, Proteins
407,076	436,601	29,525	Predicted by at least one method			Sequence, Genes, Proteins
1,751,918	1,772,578	20,660	Predicted by at least one method			Sequence, Genes, Proteins
109,304	114,236	4,932	SIGI-HMM			Sequence, Genes, Proteins
193,165	198,912	5,747	SIGI-HMM			Sequence, Genes, Proteins
405,894	409,957	4,063	SIGI-HMM			Sequence, Genes, Proteins
223,055	269,991	46,936	IslandPath-DIMOB			Sequence, Genes, Proteins
407,076	436,601	29,525	IslandPath-DIMOB			Sequence, Genes, Proteins
1,751,918	1,772,578	20,660	IslandPath-DIMOB			Sequence, Genes, Proteins

Figure 4. 10. Islandviewer 4 shows that pSCL4 contains six predicted genomic islands. All six islands were predicted by more than one method: SIGI-HMM and IslandPath-DIMOB.

4.9 Copy number increases for pSCL1 in *tap-tpg*$_3$ knockdown strains, but decreases compared to the WT.

In sections 4.6, 4.7 and 4.8 we mapped our Illumina sequencing reads to the DSM 738 genome which gave us a general idea of the replicon read coverage in various CRISPR-dCas9 knockdown strains. Upon first inspection, Bigwig files showed that pSCL1 had a higher read coverage across the whole plasmid for EM4, EM9, M5 and the WT (**Figure 4.10**). To investigate the reasons behind increased read coverage for pSCL1, the mean coverage was calculated for all strains and controls using Qualimap v.2.2.1 (Okonechnikov, Conesa and García-Alcalde, 2016). The mean coverage was used as an estimation of plasmid copy number and was used to quantify *S. clavuligerus* plasmid copy numbers relative to the chromosome. Mean coverage values for the chromosome were normalised to 1 to allow for comparison and estimation of copy number and copy number values were thus calculated for the rest of the GLPs as a replicon:chromosome ratio as highlighted in **Figure 4.11** and visualised via gel electrophoresis in **Figure 4.12**.

The idea to investigate pSCL1 copy number was based on previous work, which showed an increase in pSCL1 copy number for *S. clavuligerus* strains that had lost their replicon ends and/or entire plasmids, as seen for our *S. clavuligerus* CRISPR-dCas9 mutants. In work conducted by Algora-Gallardo, (unpublished data) the calculated mean coverage ratios for pSCL1 increased considerably in strains that had circular chromosomes and missing plasmids (pSCL3 and pSCL4), which was a concept that was explored for the mutants in this chapter.

We made various observations from the mean coverage ratios depicted in **Figure 4.11**, one of which was that the copy number of pSCL1 was higher compared to other replicons in CRISPR-dCas9 *tap-tpg*$_3$ targeting knockdown strains (EM4, EM9, 1370, M5). We could not measure copy number in strains that had lost pSCL1 (EM3, EM7 or EM8), and plasmid loss was highlighted by mean coverage ratios (MCR) (EM3:0.02, EM7:0.0, EM8:0.0). The strains that had maintained pSCL1 were EM4 (MCR:20.2) and EM9 (MCR:5.99) which had *tap-tpg*$_3$ and *tap-tpg*$_3$ & *tap-tpg*$_4$ CRISPR-dCas9 spacer targets, respectively. A higher pSCL1 copy number was also witnessed for EM5 (MCR:2.06), 1370 (MCR:4.14), M5 (MCR:5.42) but interestingly also for the WT *S. clavuligerus* strain (MCR:27.2). When we compared the EM mutants to the WT, the mutants showed a lower pSCL1 copy number relative to the WT.

A lower pSCL1 copy number was only seen for EM10 (MCR:0.55) compared to the WT, but also compared to its other replicons. Potential reasons behind lower pSCL1 copy number compared to the WT for all mutants could be due to dCas9 off-target effects. We speculated that the lack of *tap-tpg* expression on one replicon affecting replicon growth and maintenance of another replicon was another reason behind lower pSCL1 copy number. An example of this could be seen for EM10 which had a spacer targeting *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$, and showed a lower pSCL1 copy number compared to EM4 and EM9. Perhaps the higher number of CRISPR-dCas9 spacer targets in EM10, compared to EM4 (spacer targeted *tap-tpg*$_3$) or EM9 (spacer targeted *tap-tpg*$_3$ and *tap-tpg*$_2$) resulted in lower *tap-tpg* expression for EM10 and thus affected plasmid replication.

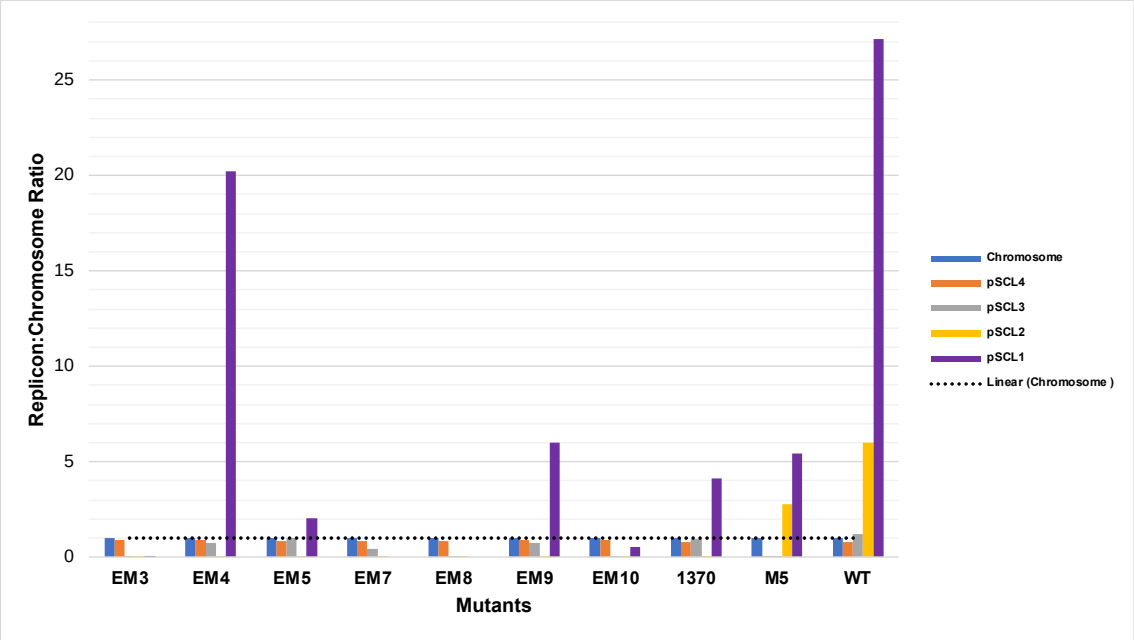


Figure 4. 11. Mean coverage for each replicon of the CRISPR-dCas9 mutants and controls using Illumina reads. EM3, EM4, EM5, EM7, EM9, EM10, M5 and WT (*S. clavuligerus* DSM 738) shown as replicon:chromosome ratio, with increased copy number observed for pSCL1. Normalised chromosome mean coverage is shown by dotted line.

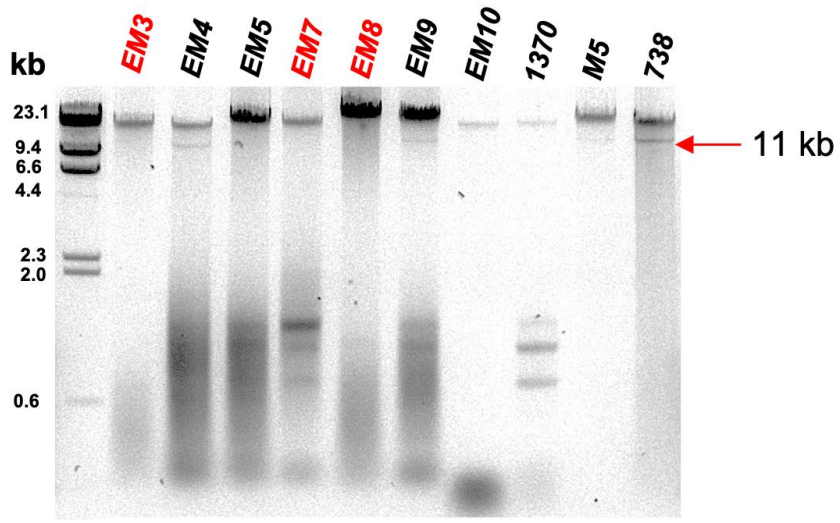


Figure 4. 12. Gel electrophoresis of CRISPR-dCas9 mutant gDNA showing the presence and absence of pSCL1. Controls used were the WT *S. clavuligerus* (738) and M5 mutant, which was missing pSCL4 and pSCL3 but had previously been confirmed to contain pSCL1 and pSCL2 via pulse field gel electrophoresis (Algora-Gallardo, unpublished data). 500 ng of isolated gDNA was loaded into each well and the darkness of the pSCL1 bands correlate with the plasmid copy number for each mutant. An 11kb band was seen for EM4, EM5, EM9, EM10, M5 and the WT, shown in black. EM3, EM7 and EM8 were shown to have lost pSCL1 when the Illumina reads were mapped onto the DSM 738 *S. clavuligerus* reference genome, highlighted in red. Ladder used: λ Phage DNA (Promega).

4.10 Strain EM10 carries a circular chromosome.

We based our aims and mutant construction in this chapter on previous work by Algora-Gallardo, (unpublished data), which showed that the loss of chromosome ends in *S. clavuligerus* Illumina sequencing reads was an indication of replicon circularisation. Replicon circularisation was therefore explored for the CRISPR-dCas9 mutants in this chapter in addition to the controls. This provided us with an insight into the significance of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ and plasmid loss in maintaining replicon linearity. The loss of pSCL4 has previously resulted in *S. clavuligerus* chromosome circularisation (Algora-Gallardo, unpublished data). The only strain which lost pSCL4 and thus presented a similar loss of chromosome ends to previous work by Algora-Gallardo, (unpublished data), was strain M5, confirming the necessity of pSCL4 for maintenance of chromosome ends. The strains constructed in this chapter presented negligible bp losses to the chromosome ends compared to M5.

In section 4.4 we demonstrated that there were similar trends in missing telomeres between mutants with the same lost plasmids. The five replicons of the EM mutants and controls were thus analysed for replicon linearity with the hypothesis that a large loss in replicon ends (like the chromosome in M5) suggested replicon circularisation. To do so, the first 300,000 bp of the LHS chromosome and pSCL4 ends were copied and pasted to the last 300,000 bp of the ends, creating a file which allowed us to identify any reads spanning across the region between the LHS and RHS telomeres (**Figure 4.13**). The same was done for the first 200,000 bp LHS and last 200,000 bp RHS ends in pSCL3 and for the entirety of pSCL1 (pSCL2 was lost from dCas9 mutants and not of interest for M5 control, as our focus was chromosome circularisation). After this was done, Illumina reads were aligned to the newly constructed 'end-to-end' genomes using BWA-MEM2, and the sequences were screened for split reads across the centre, which would indicate circularisation as shown by **Figure 4.13**.

Figure 4.13 showed the aligned Illumina reads of EM10, which was the only mutant strain to present any split reads. We discovered a read spanning across the circularisation point (purple), with a read mate pair spanning across the ends of the EM10 chromosome. This read was investigated further and shown to have a 13 bp deletion, in line with the observed chromosome end deletions (LHS=13 bp) from **Table 4.4**. The forward pair or left alignment consisted of 249 bp and the reverse pair or right alignment

consisted of 192 bp, making the read consist of a total of 451 bp (From 299,979 bp – 300,475 bp). Additionally, the reads had a MAPQ value of 60, indicating that the read coverage quality was considered abundant. To double check whether the read was found anywhere else on the *S. clavuligerus* DSM 738 genome, i.e., to rule out any misalignments, the read was BLASTed against the whole genome. Results highlighted a total of four matches, two of which were from the RHS end of the *S. clavuligerus* chromosome and two which were from the LHS end, supporting the IGV alignment in **Figure 4.13**.

Interestingly, EM10 presented a 451 bp split read for all other replicons, which was concerning, therefore we BLASTed these reads against the *S. clavuligerus* DSM 738 genome for further investigation. Results showed that the 13 bp mate pair that spanned across the ends of all replicons (aside from pSCL3 and pSCL2, as these were lost) showed 100% matches with the chromosome ends, like BLAST results of the 451 bp read spanning across the EM10 chromosome ends. These results showed that pSCL4 and pSCL1 remained linear, which was also supported by MAPQ values of less than 3.

To summarise, **Figure 4.13** showed that only one read spanned across the 13 bp point of circularisation for EM10. Considering that we used Illumina short-read sequencing, one bridged read equated to one cell in the mutant cell culture used for gDNA extraction. Therefore, one read meant that one EM10 cell was carrying a circular chromosome. The significance behind one bridged read for EM10 vs multiple, i.e., multiple cells containing a circular chromosome, is vital for understanding the underlying biology of linear growth in *Streptomyces*. We presumed that the reason behind linear chromosomes for most CRISPR-dCas9 mutants in this chapter, as well as M5, could be because *S. clavuligerus* cells with a circular chromosome were not fit enough to survive and grow in the cultures. Thus, the cells which carried linear chromosomes dominated a population in which the genomes were already impacted by the stress of dCas9 off-target effects, potentially explaining why the strain morphologies were also variable.

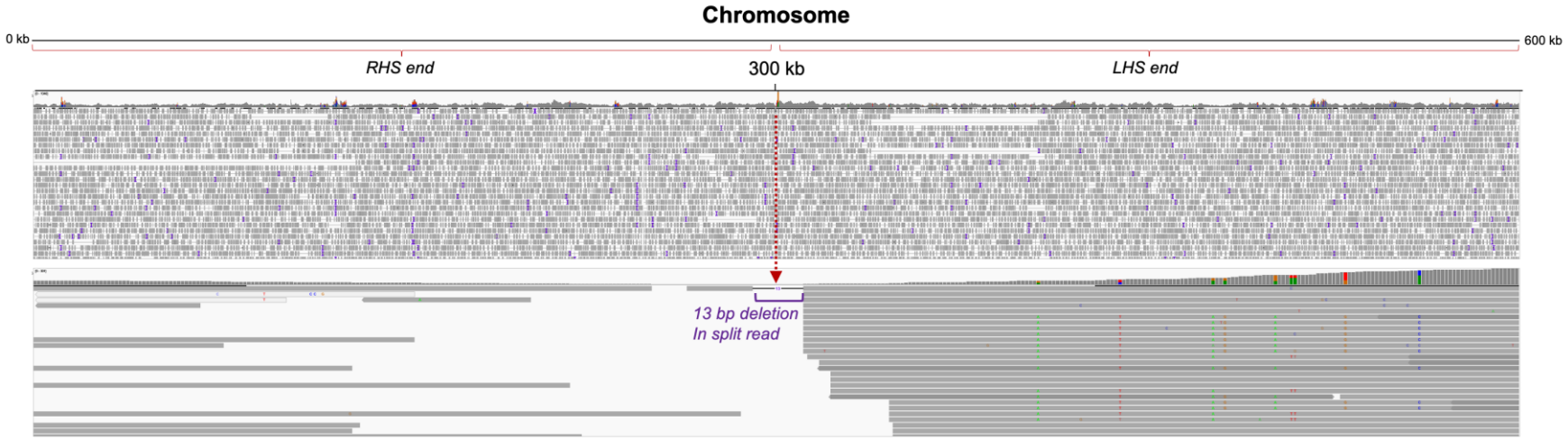


Figure 4.13. Aligned Illumina reads show that EM10 carries a circular chromosome. 13 bp deletion seen for split read was seen spanning across the 300kb centre of RHS and LHS of EM10 alignment with *S. clavuligerus* DSM 738. View from IGV.

To obtain physical evidence of chromosome circularisation a PCR analysis was performed using primers that amplified across the EM10 region of circularisation. The primers were located 263 bp from the RHS of the EM10 chromosome and 256 bp from the LHS, making the predicted amplicon, as result of a chromosome circularisation, 519 bp (**Figure 4.14 A**). This was confirmed by PCR in **Figure 4.14 B**, which shows EM mutant chromosome circularisation checks, with our main focus being EM10. The predicted amplicon sizes for the other EM mutants were > 500bp, and the amplicons from the PCR in **Figure 4.14 B** highlight unspecific binding and no band at the correct amplicon size, unlike EM10. Incorrect amplicon size supported the sequencing results indicating that these mutants did not have circularised chromosomes. The mutants presented one or multiple extremely bright bands, which is another factor that negates the sequencing results, as none of the other mutants contained any bridged reads. The circularisation check PCR for EM10 presented a 519 bp band, which was extremely faint, mirroring the fact that only one read mapped across the bridged LHS and RHS terminal sequence.

To establish the exact point at which circularisation occurred, we used the FASTA file of the WT *S. clavuligerus* DSM 738 600 kb chromosome ends and mapped the 451 split EM10 read across the centre region. **Figure 4.15** showed that a region of 13 bp was missing from the read, and that the point of circularisation comprised of the sequence CTCC*-13bp-*GCGG, highlighted by the asterisk on **Figure 4.15 A**. Additionally, the 13 bp deletion from the LHS end of EM10 was confirmed, highlighted by the blue dotted line on **Figure 4.15 A** which shows the meeting point (300-301 kb) between RHS and LHS ends, when the 300 kb LHS and RHS ends were aligned end-to-end, confirming circularisation.

The significance of the loss of 13 bp specifically was also investigated by considering the telomere topology of *S. clavuligerus'* archetypal telomeres. Telomere topology plays a great role in archetypal telomeres, as the shape of the hairpins and the number of hairpins per LHS and RHS telomere end determines the binding point of Tap and Tpg, as Tap generally binds to palindromes II and III, as established by Yang *et al.*, (2017). The binding of Tap allows for linear growth as it binds to a specific sequence on the 3' single stranded overhang during linear replication and recruits Tpg. Not only does Tpg protect the linear ends from exonucleases, but it also primes synthesis of the 5' terminus

of the lagging strand, highlighting the importance of maintaining telomere topology. Additionally, the binding of Tap to specific telomere sequences (which make up the hairpins) is essential, and the presence of palindromic terminal repeat sequences make up the binding sites for Tap (Bao and Cohen, 2003). This crucial information about telomere replication relates back to the missing 13 bp from the EM10 circular chromosome, as seen in **Figure 4.15 B**. Palindrome 1 consists of 13 bp and forms the last hairpin loop of *S. clavuligerus* chromosomal telomeres. Therefore, we wondered whether the missing 13 bp was an indication of the loss of an entire hairpin. This was of particular significance, as EM10 contained an integrated CRISPR-dCas9 vector, which silenced *tap-tpg*$_4$*, tap-tpg*$_2$ *and tap-tpg*$_3$.

The relationship between hairpin maintenance and *tap-tpg* silencing remains unclear. A potential explanation behind the 13 bp loss could be that *tap-tpg* silencing resulted in an absence of Tpg recruitment. As previously mentioned, Tpg protects the linear strand from exonucleases and thus, the absence of Tpg resulted in a lack of protection of chromosomal telomere ends. Overall, results showed that there was a relationship between silencing *tap-tpg* and chromosome circularisation for EM10. The extent of this relationship remains to be explored.

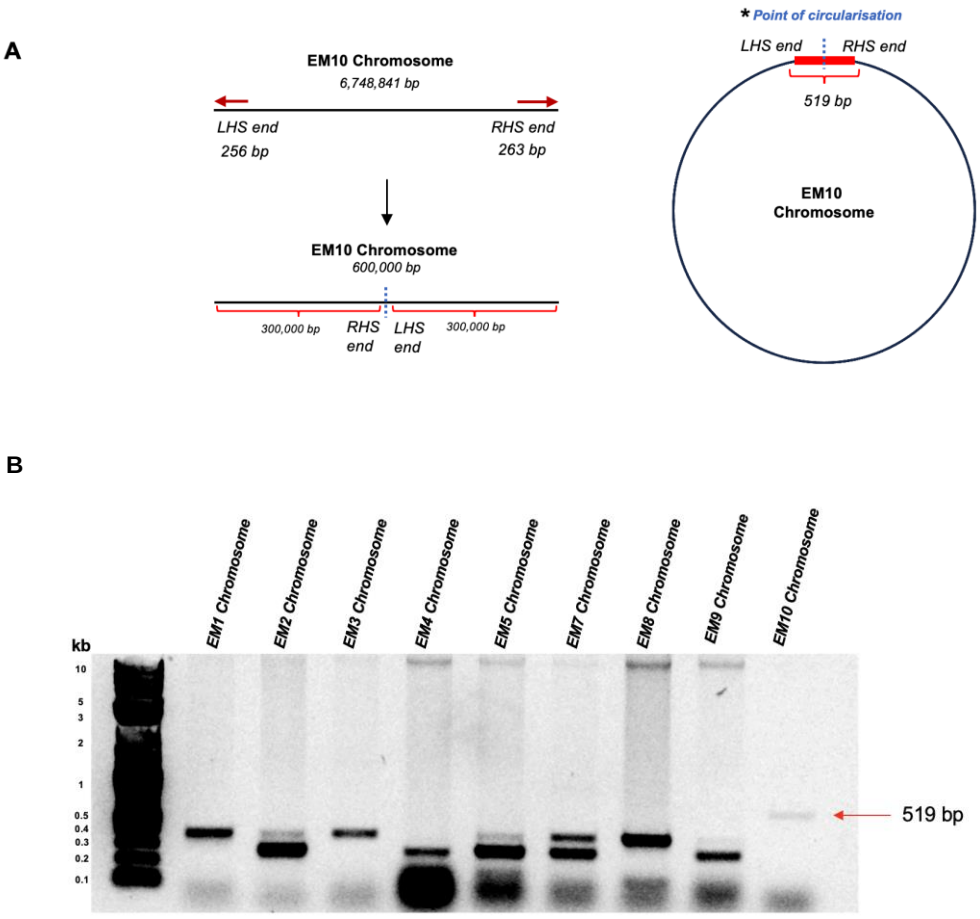


Figure 4.14. EM10 carries a circular chromosome. A) Schematic representation of chromosomal circularisation**. B)** PCR confirmation of EM10 chromosome circularisation. Agarose gel electrophoresis highlights a 519 bp amplicon because of forward and reverse primers amplifying 256 bp of the LHS and 263 bp of the RHS chromosome ends, to confirm circularisation. Other mutants were tested for chromosome circularisation and showed unspecific band amplification.

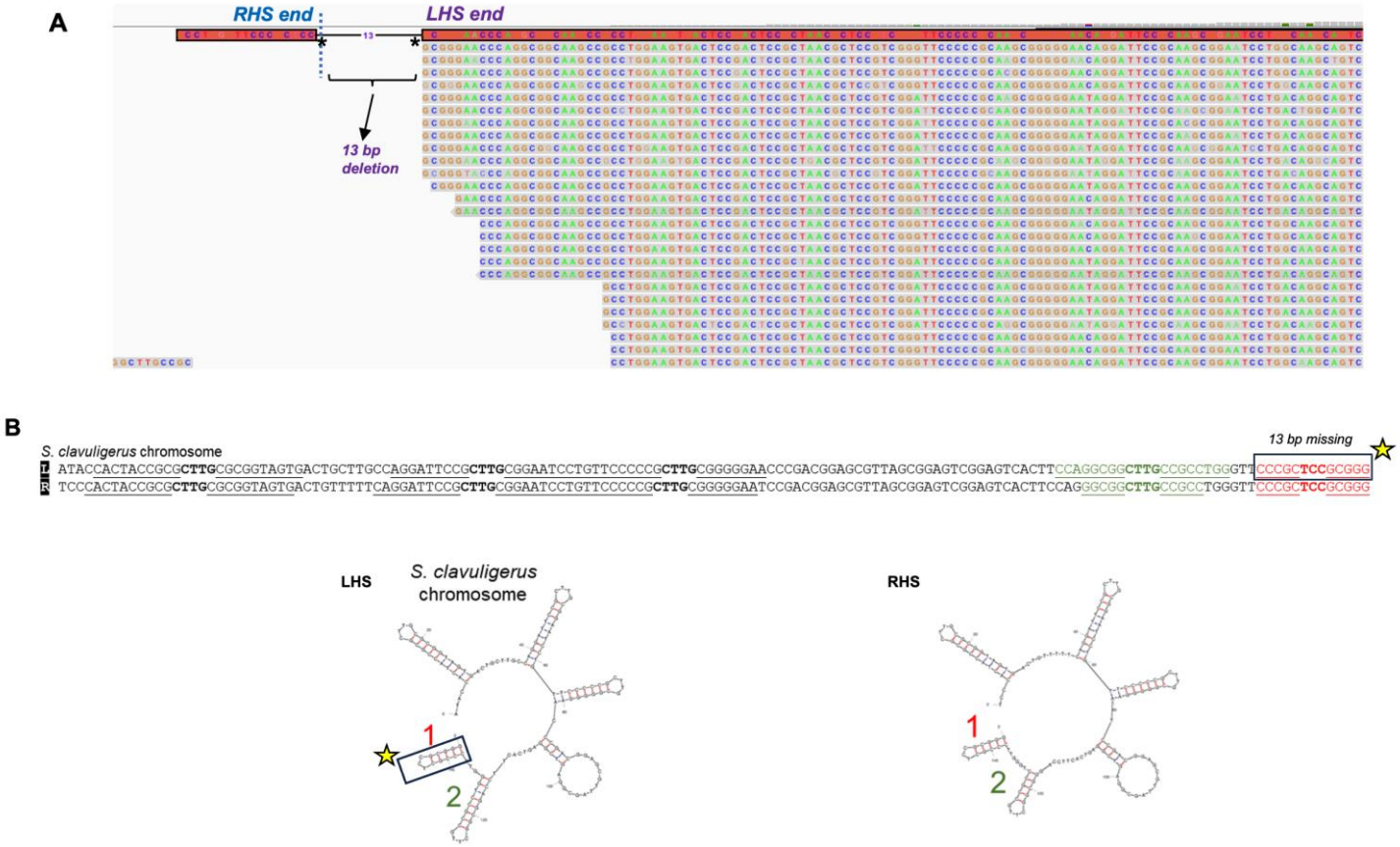


Figure 4.15. Loss of palindrome I from EM10 circularised chromosome. A) Illumina read spanning across EM10 point of circularisation, highlighting 13 bp deletion form LHS telomere **B)** 13 bp deletion corresponds to missing terminal hairpin of LHS *S. clavuligerus* end. Missing bp location relative to the archetypal telomeric hairpins is indicated by a star. ssDNA archetypal hairpins were generated using Mfold (Zuker, 2003).

4.11 Identification of the missing annotation of the genes on the lost plasmids

As highlighted in the sections before, the EM mutants and controls lost various combinations of GLPs, highlighting that many coding sequences must have also been lost from the mutants. One common trend witnessed for all CRISPR-dCas9 mutants including the dCas9 vector control was the loss of pSCL2. To evaluate the significance of the loss of pSCL2, i.e., if there are any essential missing genes that were lost from the mutants due to loss of the replicon, estimated genome sequences were constructed using QUAST Circos plots in Galaxy (Gurevich *et al.*, 2013) (**Figure 4.16**). The estimated genome sequences were constructed for four mutants that represented all combinations of lost plasmids. These were EM3 (lost pSCL3, pSCL2 and pSCL1), EM5 (lost pSCL2), EM7 (lost pSCL2 and pSCL1), and EM10 (lost pSCL3 and pSCL2). The Circos plots in **Figure 4.16** allowed for visualisation of the sequencing data, using both the assembled *S. clavuligerus* DSM 738 genome and the paired-end Illumina reads obtained for the EM mutants and M5. These reads were analysed using QUAST, a quality assessment tool for genome assemblies, and allowed for coverage data visualisation and genome statistics. The size of WT assembled genome which was used for the analysis was 9,161,391 bp in total, the loss of plasmids was thus also highlighted in the total length of the QUAST statistics.

Table 4. 6. Characteristics of CRISPR-dCas9 chosen mutants including total genome size taken from QUAST results.

Mutant	***CRISPR-dCas9 target***	***Missing Plasmids***	***Genome Size (bp)***
WT	None	None	9,161,391
1370	dCas9 control	pSCL2	8,981,493
M5	Cas9 ParB	pSCL4, 3	6,598,867
EM3	pSCL4	pSCL3, 2, 1	8,544,552
EM5	pSCL2	pSCL2	8,977,309
EM7	pSCL4 + pSCL3	pSCL2, 1	8,466,537
EM10	pSCL4 + pSCL3 + pSCL2	pSCL3, 2	8,556,248

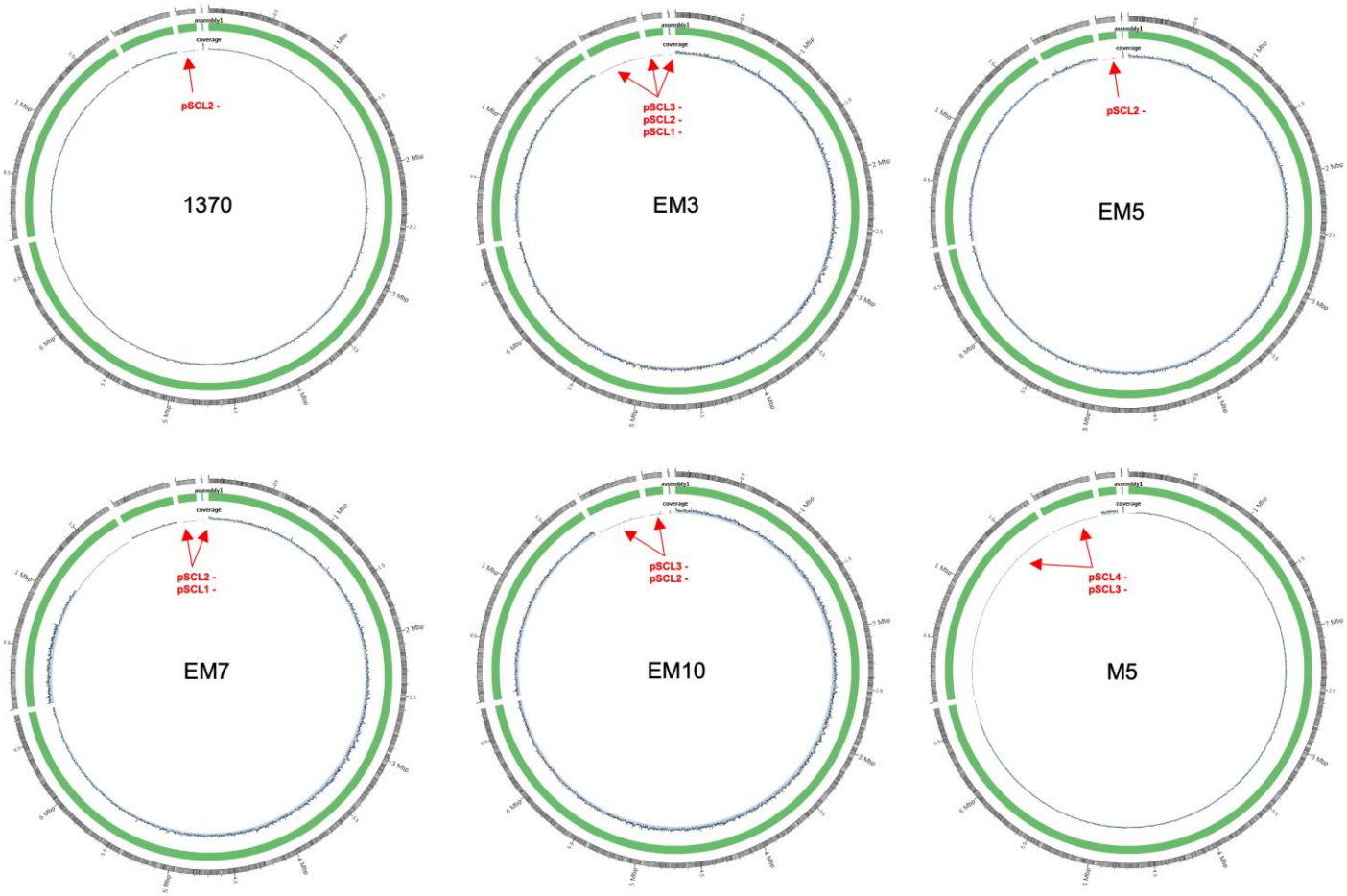


Figure 4.16. CRISPR-dCas9 system results in *S. clavuligerus* mutants with varying genotypes. Circos plots show 1370 control (pSCL2) EM3 (pSCL4-, pSCL3-, pSCL2-), EM5 (pSCL2-), EM7 (lost pSCL2-, pSCL1-), EM10 (pSCL3-, pSCL2-) and M5 (pSCL4, pSCL3).

pSCL2 was lost for all EM dCas9 mutants, meaning all the coding sequences found on this plasmid were lost as well, which we thought may have had significant consequences for the physiology of *S. clavuligerus* mutants. To compare the genomes of the EM mutants with varying lost plasmids, the five replicons of *S. clavuligerus* were annotated using RAST and compared to a whole genome annotation, allowing for the identification of any significant missing subsystems compared to the WT. Additionally, the M5 strain was used as a point of comparison and included in the analysis, as this strain also lost pSCL4, the largest of the GLPs. **Table 4.7** shows that most subsystems were present on the chromosome followed by pSCL4. None of our CRISPR-dCas9 mutants lost pSCL4, and the only other plasmid that contained any subsystems was pSCL3. A summary of the subsystems present for the strains are highlighted in **Table 4.8**, and the numbers in bold show the missing subsystems due to loss of pSCL3, which were a total of 9 subsystems for EM3 and EM10 and due to loss of pSCL4 and pSCL3, which were a total of 78 subsystems for M5.

The loss of pSCL3 resulted in the loss of two protein metabolism subsystems, four DNA metabolism subsystems, and three amino acid and derivatives subsystems, clarifying that these were also the only subsystems present on the GLP. A lost protein metabolism subsystem was identified as a protein-L-isoaspartate O-methyltransferase, important for maintaining protein integrity and damage repair. The lost DNA metabolism and amino acid subsystems on pSCL3 were DNA replication, ligation, and structural proteins. DNA structural proteins included essential partitioning proteins like Smc, but also DNA-binding proteins such as IHF, responsible for chromosome condensation and morphological development (Nanji *et al.*, 2019). Other DNA-binding proteins were nucleoid-associated proteins i.e., H-NS-equivalent type protein, Lsr2, important in regulating antibiotic production, particularly to repress specialised secondary metabolism as shown by Gehrke *et al.*, (2019) in *Streptomyces venezuelae.* Additionally, homologs of HU nucleoid associated proteins were missing from pSCL3, HupA and HupS. HU-like proteins are differentially regulated by *hupA* and *hupS* in *Streptomyces* species and were noteworthy, as the encoded proteins were involved in spore maturation. Both the pSCL3-free strains, EM3 and EM10, however showed signs of sporulation (EM3 less so), emphasising that the loss of these nucleoid associated proteins and other regulatory proteins from pSCL3 did not affect sporulation. There were no subsystems present on pSCL2 and pSCL1, underlining that their loss may have had less of an impact on growth and sporulation of strains than the loss of pSCL3 and pSCL4.

The fact that 78 subsystems were missing from M5 underlines that the chromosome and pSCL4 contained the most subsystems, and that comparably most subsystems are on the chromosome of *S. clavuligerus*. Most of the lost subsystems on pSCL4 were genes encoding cell wall and capsule subsystems like capsular and extracellular polysaccharides in addition to amino acids and aromatic compounds. These amino acids and aromatic compounds included precursors for antibiotic synthesis such as the chorismite pathway leading to tryptophan synthesis and the PABA proteins. Additionally, various Rrf2 family transcriptional regulators were missing which played a role in the iron-sulphur transport regulation of *S. clavuligerus* and the organisms' response to nitric oxide (Crack *et al.*, 2015).

It made sense that some of the subsystems lost on pSCL3 and pSCL4 were involved in antibiotic synthesis as there are 17 putative biosynthetic gene clusters on pSCL4 and one on pSCL3 (Algora-Gallardo, unpublished data). Furthermore, biosynthetic genes often cluster together and can be located on more than one plasmid, yet still function to produce precursors, such as the clavam biosynthetic gene cluster involved in the production of our compound of interest in this study, clavulanic acid. Various genes such as *ceaS1*, *bls1*, *pah1* and *oat1* are clustered on pSCL4, and others such as *cas1* are located on the chromosome, perhaps why the cluster does not show up as a lost subsystem and why the loss of pSCL4 did not have major effects on clavulanic acid production, which is shown in section 4.13 of this chapter. The differences in missing subsystems were an estimation of the differences in lost coding sequences between the mutants and the WT and only hold value when compared to the physiological effects of the mutants. To characterise these effects, we decided to investigate mutant growth and clavulanic acid production (section 4.12 & section 4.13).

Table 4. 7. RAST summary of *S. clavuligerus* DSM 738 characteristics. Adapted from (Algora-Gallardo, unpublished data).

	Chromosome	***pSCL4***	***pSCL3***	***pSCL2***	***pSCL1***
Size (bp)	6,748,854	1,795,698	455,364	149,779	11,696
terminal repeat length (bp)	293	194	1,597	413	917
Coding sequences	5,825	1,728	528	175	12
rRNA operons	18	0	0	0	0
tRNA genes	65	0	1	0	0
BGC's	25	17	1	0	0

Table 4.8. List of RAST annotation with total subsystem comparison to WT (*S. clavuligerus* DSM 738). Bold = missing subsystems due to plasmid loss.

Subsytems	*WT*	*EM3*	*EM5*	*EM7*	*EM10*	*M5*
Cofactors, Vitamins, Prosthetic Groups, Pigments	187	187	187	187	187	187
Cell Wall and Capsule	54	54	54	54	54	**40**
Virulence, Disease and Defense	37	37	37	37	37	**34**
Potassium metabolism	13	13	13	13	13	13
Photosynthesis	0	0	0	0	0	0
Miscellaneous	29	29	29	29	29	**26**
Phages, Prophages, Transposable elements, Plasmids	1	1	1	1	1	1
Membrane Transport	34	34	34	34	34	**27**
Iron acquisition and metabolism	37	37	37	37	37	37
RNA Metabolism	45	45	45	45	45	**43**
Nucleosides and Nucleotides	102	102	102	102	102	**99**
Protein Metabolism	200	**198**	200	200	**198**	**196**
Cell Division and Cell Cycle	0	0	0	0	0	0
Motility and Chemotaxis	0	0	0	0	0	0
Regulation and Cell signaling	26	26	26	26	26	**24**
Secondary Metabolism	17	17	17	17	17	17
DNA Metabolism	96	**92**	96	96	**92**	**92**
Fatty Acids, Lipids, and Isoprenoids	164	164	164	164	164	**155**
Nitrogen Metabolism	10	10	10	10	10	10
Dormancy and Sporulation	10	10	10	10	10	10
Respiration	108	108	108	108	108	**107**
Stress Response	43	43	43	43	43	43
Metabolism of Aromatic Compounds	23	23	23	23	23	**19**
Amino Acids and Derivatives	402	**399**	402	402	**399**	**388**
Sulfur Metabolism	8	8	8	8	8	**4**
Phosphorus Metabolism	27	27	27	27	27	27
Carbohydrates	313	313	313	313	313	**309**

4.12 Characterisation of WT and mutant growth of *S. clavuligerus*.

To investigate the phenotypic effects of lost plasmids in *S. clavuligerus* DSM 738 CRISPR-dCas9 mutants, growth curves were obtained for EM3, EM5, EM7 and EM10 alongside the WT and 1370 strains in liquid cultures. For this, one colony was inoculated in 50mL of TSB media and grown for a period of 108 hours. 1mL aliquots were taken every 12 hours and cell dry weight was measured to obtain an accurate representation of growth as a means of biomass. Culturing *S. clavuligerus* DSM 738 in a liquid culture was previously optimised, however these conditions did not apply to EM mutants. Optimisation of growth involved increased aeration, using one solid colony, shaking at 250 RPM, and using large springs to achieve even greater aeration. Previous growth trials of mutants showed pelleting of strains, which did not sporulate as well initially, such as 1370 or EM3. Therefore, measuring biomass was determined as the most suitable method of determining growth, as supposed to measuring the optical density, as total biomass would not be affected by physiological changes in the strains i.e., pelleting.

The graphs in **Figure 4.17** showed that *S. clavuligerus* WT growth followed the most representative curve, meaning an exponential (log) phase was followed by a stationary phase. Despite the WT resembling the most sigmoidal shaped growth curve, irregular growth was seen for the strain as initial growth in the log phase plateaued briefly before proceeding to grow exponentially again. Reasons behind irregular *S. clavuligerus* growth were thought to be the removal of flasks during biomass sample collection, however, could also be due to the varying phenotypic characteristics witnessed in *S. clavuligerus* growth in earlier sections of this chapter.

Nonetheless, EM mutant growth exhibited phases resembling log phases proceeding into stationary phase followed by a phase resembling decreased growth, suggesting the cells in the flasks had lysed. There was variation in standard deviation, as the biological triplicates grew differently. Overall, all strains began to show signs of cell lysis after 84 hours of growth, highlighting that the optimal amount of time for growing the 738 WT and EM mutant strains before lysis occurs (e.g., for gDNA isolation) is around 60-72 hours which is roughly 3 days. The mutant strain for which the least biomass was harvested, which also showed stunted growth (starting growth after 24 hours) was EM3, a pSCL3, pSCL2 and pSCL1 free strain, suggesting that the lack of coding sequences on these plasmids was indeed affecting growth. The other three mutant strains EM5, EM7, and EM10 started to grow between 12-24 hours, however their log phase began after 24 hours.

To investigate the effect of lost plasmids on exponential growth and determine which strain grew significantly faster than the WT, the specific growth rate (μ) was determined. μ was determined by measuring the linear time points during the log phase of growth and measuring the slope. The specific growth rate was used as a measure to the growth of the mutant strains compared to the WT control. The results showed μ values of 0.02, 0.03, 0.03, 0.01, 0.08 and 0.10 for WT, 1370, EM3, EM5, EM7 and EM10 respectively (**Figure 4.18**). Statistical analyses were performed using Fisher's LSD test and showed a significant increase in μ for EM7 compared to the WT ($p<0.03$), EM10 compared to the WT ($p<0.01$) and EM10 compared to the vector control 1370 ($p<0.02$) (**Figure 4.18**).

These results highlighted the phenotypic variation seen when growing a heterogenous population of cells within a liquid culture. They also concluded a significant increase in *S. clavuligerus* growth compared to the WTs for mutant strains which had lost pSCL3 and pSCL2 (EM10) and pSCL2 and pSCL1 (EM7), suggesting that the loss of these plasmids did not affect specific growth rates and that these plasmids are not essential.

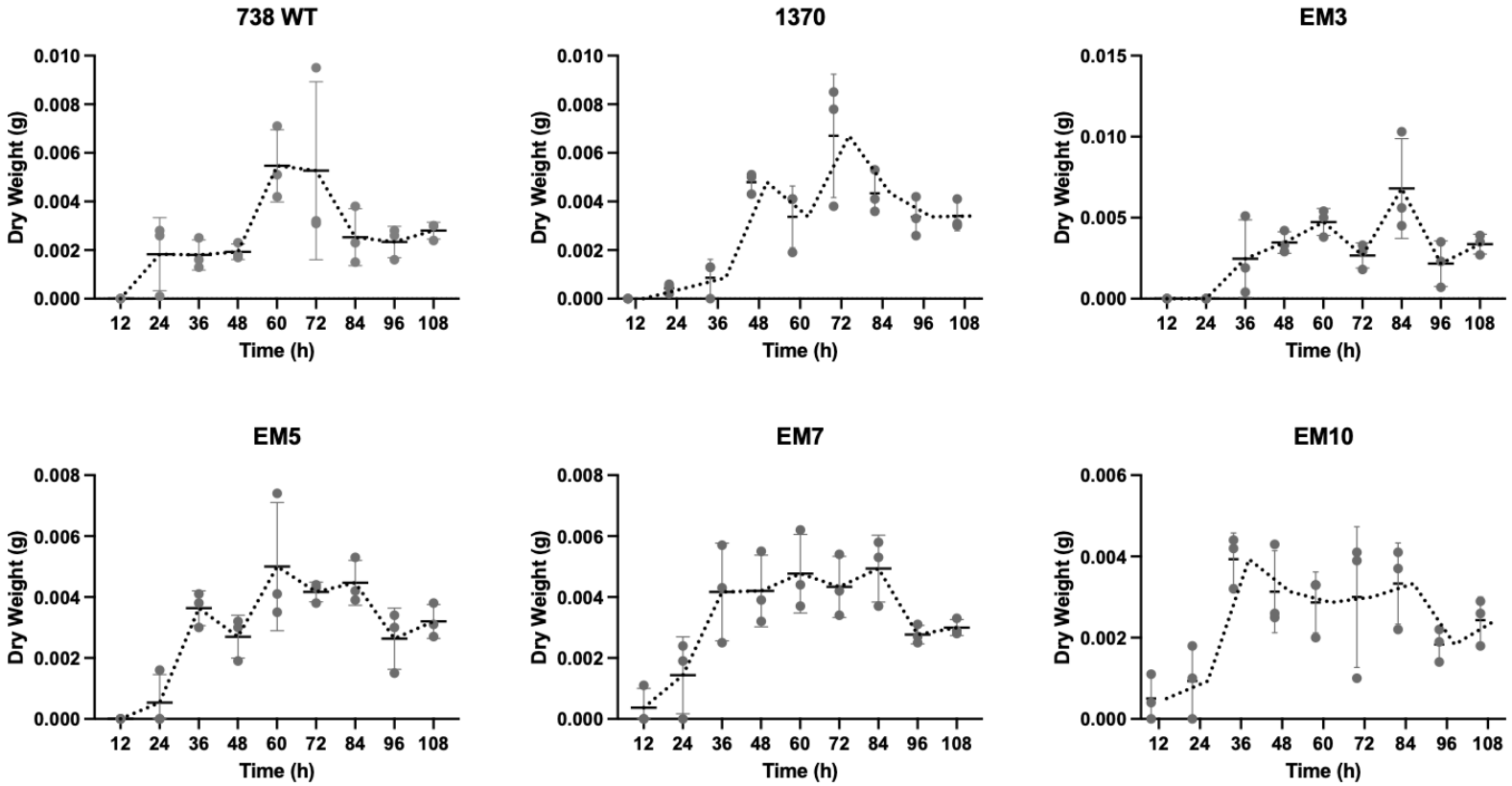


Figure 4.17. Growth of plasmid-free strains in liquid cultures over a period of 108 hours. Strains were measured in biological triplicates and cell dry weight (g) was used as an indicator of growth. Error bars represent the mean with standard deviation with a dotted trendline to highlight sigmoidal resemblance.

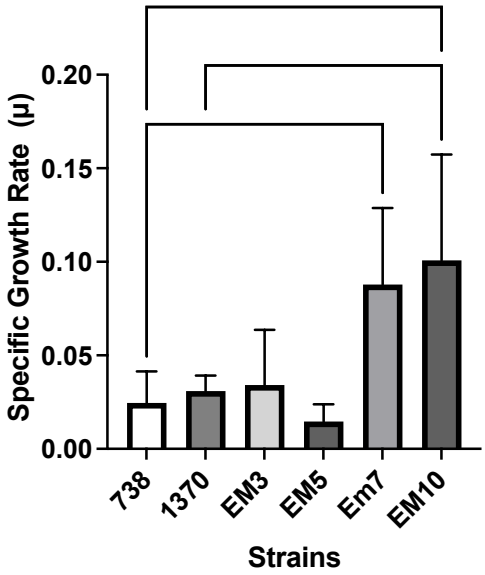


Figure 4.18. Specific growth rate (μ) of the plasmid-free strains in liquid cultures. Time points that followed a linear correlation in the log phase were used to calculate μ. Error bars show standard deviation. *: $p<0.05$.

4.13 Clavulanic acid production titres of *S. clavuligerus* strains with varying lost plasmids.

We hypothesised that the reduction of the number of plasmids in the *S. clavuligerus* genome would allow for improved growth due a decreased metabolic burden, which was seen in the previous chapter by an increased specific growth rate of mutants which were pSCL1 and pSCL2 negative (EM7), and mutants which were pSCL2 and pSCL3 negative (EM10). To establish whether increased growth correlates with an increase in clavulanic acid production, additional samples were taken of strains that were grown in liquid culture (**Figure 4.17**) for the purposes of growth curves. This allowed direct comparison of strain growth with clavulanic acid production, with measurements taken every 12 hours for a period of 108 hours.

Samples were taken of the WT, 1370, EM3, EM5, EM7 and EM10 flasks in biological triplicates and *S. clavuligerus* mutant cells were spun down so that 8 µL of the supernatant could be used to perform clavulanic acid assays. The amount of clavulanic acid was determined via the addition of imidazole, producing a derivative molecule with a wavelength that absorbs at 324 nm, measured by spectrophotometer. Potassium clavulanate standards were used to create a standard curve, and the standard curve equation was used to calculate the amount of clavulanic acid produced, solving for y, where $y = mx + c$ (x = clavulanic acid production µg/mL).

Figure 4.19 A highlights an increase in clavulanic acid production over time for all strains, however clavulanic acid titres were incomparable to industrial titres, with the highest amount of clavulanic acid produced being 0.0003 µg/mL and industrial titres being around 200 µg/mL (Steve Kendrew, GSK). To compare clavulanic acid production between the WT and mutants and to see whether plasmid loss affected production, the clavulanic acid production titres at 108 hours were plotted, highlighting significant reductions for strains with missing plasmids (EM3 and EM7) (**Figure 4.19 B**). This fell in line with previous data (Algora-Gallardo, unpublished data), who showed a reduction in clavulanic acid production in strains with lost pSCL4. In this work, EM3 and EM7 had not lost pSCL4, however as demonstrated in section 4.6, translocations of parts of the pSCL4 were common in the mutants and could have resulted in translocation of clavulanic acid encoding genes, as some of the early clavam biosynthetic genes are present on pSCL4 (Paradkar, 2013). Additionally, the 1370 control showed a significant reduction in clavulanic acid production compared to the WT, and when we considered

Figure 4.18 from the previous section, it became evident that this reduction was connected to a decreased specific growth rate for 1370. This cannot be said for EM3 or EM7.

Overall, minimal clavulanic acid was produced and reasons behind these low production values could be the media (TSB) used, as the WT itself does not present high production, which is important to note. Using the same media did however allow for a direct comparison of WT and mutant growth over time, showing a minimal clavulanic production increase.

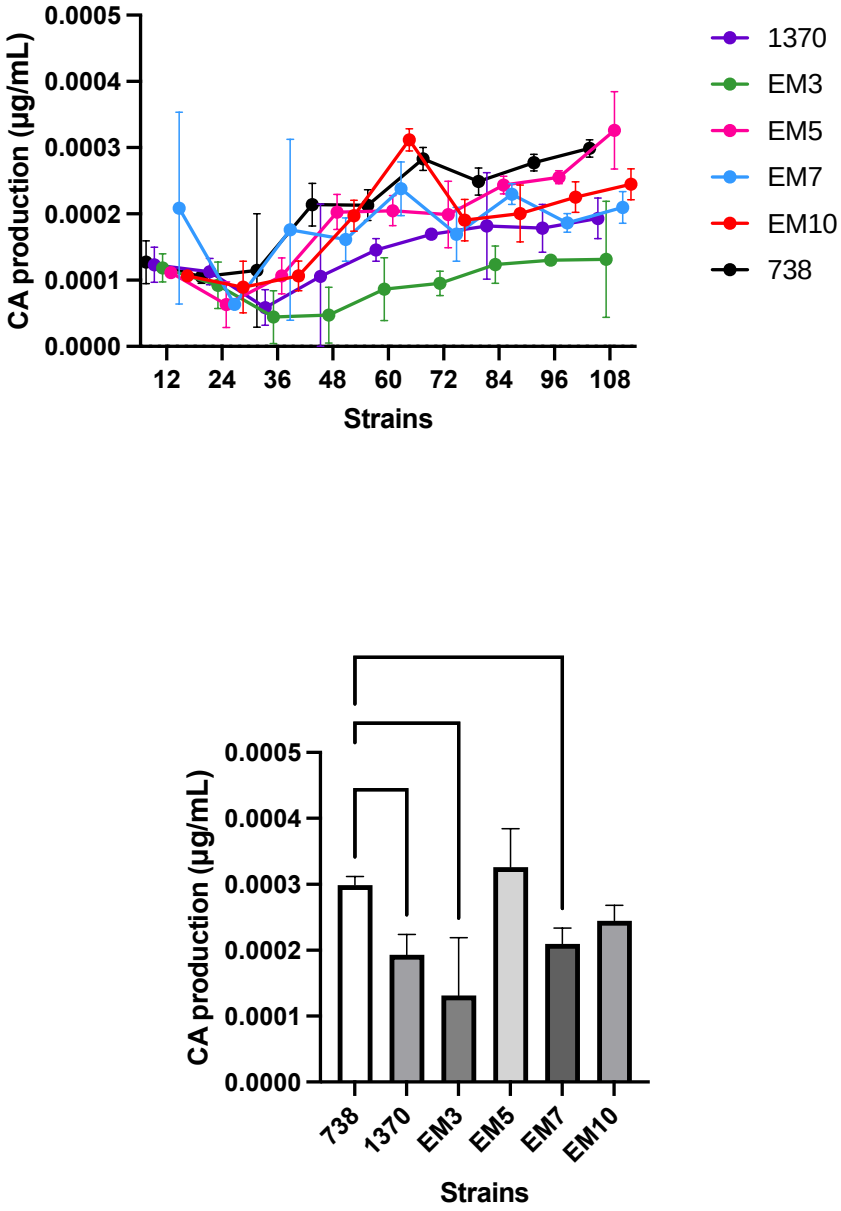


Figure 4.19. Clavulanic acid production (µg/mL) of the plasmid free strains in liquid cultures. A) Amount of clavulanic acid was measured every 12 hours over a period of 108 hours, highlighting increased production over time. **B)** Clavulanic acid production after 108 hours, showing significant decrease in production for strains with missing plasmids and *tap-tpg*$_4$ targeting spacers. Error bars show standard deviation. *: p<0.05, **:p<0.005, ***:p<0.0005.

4.14 Summary

In this chapter we report the loss of various GLPs from *S. clavuligerus* mutant strains. This was achieved by utilising a modern genetic toolkit, CRISPR-dCas9, which had a targeted spacer resulting in transcriptional silencing of the *tap* genes. More specifically, the CRISPR-dCas9 system was multiplexed with multiple spacer cassettes to create knockdowns targeting various combinations of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$. These knockdowns created EM3, EM4, EM5, EM7, EM8, EM9 and EM10 which were analysed genotypically highlighting the loss of GLPs. We took forward strains representing all of the combinations of lost plasmids for phenotypic analysis: these were EM3 (lost pSCL3, pSCL2 and pSCL1), EM5 (lost pSCL2), EM7 (lost pSCL2 and pSCL1), and EM10 (lost pSCL3 and pSCL2).

The concept of using multiple spacers has been used in studies by Qi *et al.,* (2013), Cobb *et al.,* (2015), Jiang *et al.,* (2015) and Tong *et al.,* (2015) who used multiple sgRNAs to target genes upstream and downstream of their target genes. This would be difficult to apply to the current study since *tap* and *tpg* are under the same operon, and gene silencing of multiple genes would not determine the significance of the terminal proteins for *S. clavuligerus* growth and thus clavulanic acid production. Multiplexing spacers was completed using BioBrick cloning, a concept shown by Larson *et al.,* (2013). The study by Larson *et al.,* (2013) described an in depth protocol for prokaryotic and eukaryotic CRISPR-dCas9 application, however focussed on model organisms such as *E. coli* for their application.

The value of multiplexing CRISPR sgRNAs in organisms like *E. coli* and *S. cerevisiae* has also been shown by Zhang *et al.*, (2017) and McCarty *et al.*, (2019) who showed increased transcriptional repression when multiple genes were targeted in parallel, using CRISPR-dCas12, a system with increased multiplexing ability as it possessed its own gRNAs. Multiplexing sgRNAs for our system, CRISPR-dCas9 is an inexpensive and straightforward method that can lead to larger scale sgRNA library preparation and gene regulation (Larson *et al.*, 2013). Limitations for both CRISPR-dCas9 and CRISPR-dCas12 were highlighted by all three aforementioned studies, the main limitation being the need for a PAM sequence, which limited the availability of target sites in the genome of interest (Larson *et al.*, 2013; Zhang *et al.*, 2017; McCarty *et al.*, 2019). One way to overcome PAM sequence limitations, suggested by Larson *et al.,* (2013), was to use Cas9/Cas12 homologs, which had different PAM sequences and targeted a wider range of template sequences, thereby reducing nonspecific binding.

CRISPR-dCas9 showed far less spacer specificity in *S. clavuligerus* DSM 738, used in this chapter. The EM mutants in this chapter showed genetic variability, which we determined through Illumina sequencing of the strains which showed the varying genotypes in mutants, despite many of them having the same CRISPR-dCas9 spacer targets. The WT, 1370, EM3, EM4, EM5, EM7, EM8, EM9 and EM10 strains were initially sent off for sequencing and proved the loss of entire plasmids and replicon ends when mapped against the WT assembly. Another phenomenon observed in this chapter was that all the CRISPR-dCas9 mutants presented the loss of pSCL2, which was a direct result of using this CRISPR system, as the pSET152 control *S. clavuligerus* strain (which did not have the dCas9 and/or sgRNA cassette) still retained the plasmid, sharing sequence homology with the WT, however the 1370 CRISPR-dCas9 vector control did not.

Reasons behind the loss of pSCL2 remained unclear, however we speculated that the sgRNA cassette shared homology with *tap-tpg*$_2$ and therefore led to a knockdown and off-target effects on pSCL2. Off-target effects on other parts of the target organism's genome are not uncommon for CRISPR-dCas9 systems, as described by Wu, Kriz and Sharp, (2014) and Modrzejewski *et al.*, (2020). Additionally, a recent study by Rostain *et al.*, (2023), emphasised the usefulness of dCas9 in targeted engineering but equally compared the benefits of the system to the adverse, off-target effects, which were caused by nucleotide identity of target sequences with the PAM-proximal sequence. The conclusions made by Rostain *et al.,* (2023) link to the results presented in this chapter, as one reason behind loss of pSCL2 in this chapter could have been sequence similarity of the PAM sequence with the target sequence causing non-specific, off-target dCas9 silencing, which is also underlined in both studies. More specifically, a term called 'bad-seed' sequence was used by both studies to highlight that the PAM adjacent sequence was responsible for increased off-target effects, particularly when it shared sequence homology with the target genome (Cui *et al.*, 2018; Rostain *et al.*, 2023). The genome wide screening and RNA-seq experiments conducted in both studies concluded that bad-seed sequences could have as little identity as 4 nucleotides with the target genome which could affect increased toxicity due to random gene repression. A method to overcome this was suggested to be regulation of the dCas9 promoter, which is an idea that can be applied to this study, as the constitutive *ermE**p promoter is known to be extremely strong (Wang *et al.*, 2013).

Apart from losing pSCL2, the mapped Illumina sequences of our EM mutants in section 4.4 showed genotypic similarities and differences, which we believed were a result of the *tap-tpg* sgRNA target. These were highlighted as, EM mutants contained combinations of spacers targeting *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$, and had the same lost plasmids despite different spacer targets. Interestingly, strains with *tap-tpg*$_4$ targeting spacers were the only strains that lost pSCL3, suggesting non-specific silencing. Strains with similar lost plasmids despite having different *tap* genes silenced showed that the encoded Tap proteins in *S. clavuligerus* may not solely bind to the telomeres of the plasmid they are encoded on. This concept was supported by previous work (Algora-Gallardo, unpublished data), who showed that the loss of *tap-tpg*$_4$ caused chromosomal circularisation.

Chromosomal circularisation was also seen for one of our mutants, EM10, which had lost pSCL2 and pSCL3. Previous work by Algora-Gallardo, (unpublished data) demonstrated the loss of pSCL4 and pSCL3 due CRISPR-Cas9 cleavage at the *parB* site on pSCL4 (seen for strain M5). Additionally, Algora-Gallardo, (unpublished data) showed that the chromosome circularised for their strains, and it was hypothesised that this was due to the loss of *tap-tpg*$_4$. In this chapter the chromosome of EM10 circularised, and it was the only mutant with a circularised replicon. We suggested the loss of pSCL3 and pSCL2 from EM10 as potential reasons behind EM10 chromosome circularisation. This showed the potential of pSCL3 and pSCL2 encoded Tap patching the chromosome ends, as no other mutant had lost this variation of plasmids and the loss of these plasmids resulted in circularisation. We built on the idea that Tap and Tpg do not just act on one replicon and speculated that another reason why EM10 had a circularised chromosome could have been Tap and Tpg bioavailability. Tap and Tpg are encoded by 3 out of 5 replicons in *S. clavuligerus* for the mutant suggesting higher bioavailability when more plasmids are retained. The loss of plasmids might have affected their availability for other replicons and thus affect replicon linearity. Similarly, the transcriptional blockage of *tap* and *tpg* could affect bioavailability. The dCas9 sgRNA in EM10 targeted all the *tap* genes, including *tap-tpg*$_4$, supporting previous work showing that the loss of Tap4 leads to chromosomal circularisation, but also could have meant that less bioavailability of the Tap protein could have led to circularisation.

In this chapter we aligned our EM mutant Illumina sequences to the *S. clavuligerus* reference genome, which allowed us to see one split read, which mapped across the 13 bp region of circularisation of EM10. The fact that only one read mapped across this

region showed that cells with a circular chromosome were not fit enough to survive and grow, linked to a low number of reads in that region after sequencing. Interestingly the 13 bp deletion seen for the split read coincided with 13 bp lost from the LHS end of EM10, showing the loss of the 1st archetypal telomeric palindrome from the strain. Genotypic effects, aside from chromosome circularisation, of the loss of the hairpin remain unclear, however strains were also analysed phenotypically. EM10 presented a dense spore lawn compared to the WT and to other mutant strains, such as EM3. Additionally, the specific growth rate of EM10 was significantly higher than then WT 738 and the dCas9 WT 1370 strains, being the only strain with statistically significant higher growth when compared to both controls. This poses the question whether the loss of a terminal hairpin and chromosome circularisation lead to better growth or whether growth is completely unrelated to both. It is known that during end patching, Tap recruits Tpg, which physically prevents telomere end damage due to exonucleases (Bao and Cohen, 2003). This protective characteristic is unnecessary when a replicon circularises, making Tap more bioavailable to be used by the rest of the replicons if our 'bioavailability' concept is indeed true.

The loss of entire replicons of the dCas9 mutants in this chapter highlighted that the loss of plasmids resulted in loss of replicon ends, as well as the presence of read islands. We showed that these read islands were mostly a result of transposable elements, and that they occurred when reads of the plasmid-free strains were mapped onto the *S. clavuligerus* DSM 738 WT assembly. Genes encoding transposases were found to border the read island regions, highlighting that the sequences had re-inserted themselves into another part of the genome, i.e., the chromosome or pSCL4 for dCas9 mutants in this study. Genome rearrangements and duplications are common amongst genetically edited *Streptomyces* species and remain one of the challenges of editing such dynamic genomes. Duplications of genes could therefore also be an underlying reason behind read islands, but also stabilise the loss of essential genes and missing subsystems from strains with lost plasmids. The importance of genomic rearrangements and duplications was underlined when referring this information back to the method of creating mutants in this chapter. Not only was *tap-tpg*$_4$ silenced, but multiplexed CRISPR-dCas9 systems were used to block transcription of all the *tap-tpg* genes in *S. clavuligerus*. The deletion of multiple copies of essential genes, i.e., synchronous deletions, has previously resulted in synthetic lethality in *Streptomyces* (Bu *et al.*, 2020). Therefore, the synthetic lethality of silencing of all the *tap-tpg* genes could have impacted

cell growth and lead to cell death, explaining why none of the mutants in this study presented more than three replicons lost as the EM mutants in this chapter were the strains which survived culturing, i.e., had essential core genes allowing them to grow and be sub-cultured on solid and liquid media.

In this chapter we calculated the replicon copy number in reference to the chromosome read coverage by measuring mean coverage depth. The mean coverage depth showed that pSCL1 copy number was the highest for mutants and controls that had retained the plasmid, falling in line with the fact that smaller plasmids generally have higher copy numbers due to their smaller metabolic burden (Fong *et al.*, 2007). However, the EM mutants and controls showed lower mean coverage and thus copy number of pSCL1 compared to the WT. A lower pSCL1 copy number for plasmid free mutants compared to the *S. clavuligerus* WT was contrary to previous work by Algora-Gallardo, (unpublished data), who identified an increase in pSCL1 copy number for strains that had lost replicon ends or entire plasmids. Reasons behind a lower copy number of pSCL1 as well as other replicons such as pSCL4 and pSCL3 in EM mutant strains compared to the WT could be that the knockdown of *tap-tpg* resulted major instability and genomic rearrangements which can cause variations in copy number (Periwal and Scaria, 2015). Generally, mean coverage depth was <1 per chromosome for pSCL4 and pSCL3 for EM mutants, and 2-20 copies per chromosome for pSCL1 in strains that hat retained the replicon. The WT presented a pSCL1 copy number of 27 per chromosome copy, which falls at around an average when comparing to the literature, highlighting copy numbers as low as 7 per chromosome for 356 kb plasmid SCP1 from *S. coelicolor* A3(2) (Yamasaki *et al.*, 2003) and as high as 60 per chromosome for 18 kb plasmid pSLA2-S in *S. rochei* (Hirochika and Sakaguchi, 1982). pSLA2-S has a comparable size to pSCL1 (11 kb), indicating that despite pSCL1 copy number being relatively high for mutants and WT in this study, an underlying reason could be plasmid size.

Additionally, we annotated the EM mutant and WT genomes with RAST, which allowed us to identify missing subsystems due to the loss of plasmids. Most subsystems were found on pSCL4 and the chromosome, highlighted by a fewer number of total subsystems for M5, as this strain had lost a large region of the RHS chromosome end as well as both pSCL3 and pSCL4. Previous work showed a total of 8,268 coding sequences present on *S. clavuligerus* DSM 738 (Algora-Gallardo, unpublished data), underlining that the core genome of the strain represented around 40% of the total genome, comparable to other *Streptomyces* species (Zhou *et al.*, 2012) as most genes

are found on the chromosome. Since all the dCas9 mutants in this study possessed an in-tact chromosome and pSCL4, with minimal bp lost toward the LHS and RHS telomeres, the core genes encoded were preserved and thus genes involved in primary and secondary metabolism, e.g., biosynthetic gene clusters, were preserved. Nevertheless, the loss of pSCL4 and pSCL3 from M5 and pSCL3 from strains like EM3 and EM10 were the main contributors to increased numbers of missing subsystems such as protein metabolism subsystems i.e., methyltransferases and DNA structural and binding proteins i.e., Smc.

Finally, we investigated physiological characteristics of the dCas9 plasmid-free strains, such as specific growth rate and clavulanic acid biosynthesis. Clavulanic acid biosynthesis was observed over a period of 108h and showed minimal titres of the antibiotic (max 0.0003 µg/mL), especially compared to industrial tires of around 200 µg/mL, however production increased over time. The plasmid-free EM mutant strains did not produce more clavulanic acid than the WT, contrary to the theory, posed in the *Aims* section of chapter 1.0, which predicted increased clavulanic acid production due to a decrease in metabolic burden via plasmid loss. Previously, Álvarez-Álvarez *et al.*, (2014), reported a 20%-30% decrease in clavulanic acid production when pSCL4 was lost. Part of the early genes of clavulanic acid biosynthesis were located on pSCL4, explaining why a loss in pSCL4 resulted in less production (Paradkar, 2013). None of the dCas9 plasmid-free strains had lost pSCL4, however a significant decrease in clavulanic acid production was still seen for 1370, EM3 and EM7 compared to the WT.

We measure the specific growth rate (µ) for EM mutants with varying lost plasmids, and EM7 (pSCL2-, pSCL1-) and EM10 (pSCL3-, pSCL2-) showed a significantly higher growth rate compared to the WT, one of which (EM10) contained a circular chromosome, highlighting that chromosome circularisation resulted in better growth. The specific growth rates also disproved the hypothesis presented in the *Aims* section of this chapter 1.0, which predicted that strains with increased number of lost plasmids would decrease the metabolic burden, potentially increasing growth. This chapter highlighted that the number of plasmids lost was not necessarily the cause of significantly higher or lower growth compared to the WT, as strains with a greater number of lost plasmids presented higher genetic instability, linked to poor sporulation and lower specific growth rates (like EM3). Comparison of the chromosomes of *Streptomyces* species has highlighted bilateral synteny and origin islands, which are part of the conserved, central region of the chromosome, and the telomere arms are therefore subject to rearrangements and

recombination (Algora-Gallardo *et al.*, 2021). Chromosome circularisation of EM10 thus may not have had as great of an impact on growth, as the central regions of the strain remained in-tact, and selective pressures could have resulted in cells with linear replicons to grow, particularly since *tap-tpg*$_4$ was still expressed on pSCL4.

Furthermore, genetic instability has been previously witnessed in studies targeting terminally vs centrally located genes via Cas9 systems, which differ to the system in this study, however, highlight that the location of the gene being edited on the replicon is essential for stability and replication (Hoff *et al.*, 2018). In this chapter, *tap-tpg* genes were silenced rather than targeted for DSBs due to the previously established instability caused by DSBs, with the aim of witnessing their interactions with replicon ends in cis and in trans, to determine their significance in *S. clavuligerus*. The location of *tap-tpg* genes is not always terminal. On pSCL4, *tap-tpg*$_4$ is located centrally at around 900,000 bp from the replicon ends. *tap-tpg*$_3$ on pSCL3 is located closer to the left arm, at around 40,000 bp and *tap-tpg*$_2$ in pSCL2 is located around 13,000 bp from the RHS end, highlighting that for pSCL3 and pSCL2 the Tap-Tpg proteins are encoded by genes located far closer towards the ends of the replicons that are more prone to recombination. To summarise, the reasons behind the loss of replicons from EM mutants and 1370 in this chapter remain unclear. Based on the literature and the results in this chapter, we presumed that the susceptibility of pSCL3 and pSCL2 *tap-tpg* to rearrangements and the combination of CRISPR-dCas9, as well as rounds of sub-culturing (Gomez-Escribano *et al.*, 2021), could have resulted in genome rearrangements, lost replicon ends and ultimately, lost replicons in *S. clavuligerus* DSM 738.

5.0 Role of the Terminal Proteins of *S. clavuligerus* DSM 738 in End-patching.

In section 4.0 we showed that inhibiting the transcription of *tap-tpg*$_4$, *tap-tpg*$_3$, and *tap-tpg*$_2$ with CRISPR-dCas9 produced mutants with a variation of lost plasmids and made it clear that there are functional similarities between the Tap and Tpg proteins of *S. clavuligerus* DSM 738. The question of whether Tap and Tpg can act in cis or in trans remains open in *S. clavuligerus*. Evidence suggests that Tap4 is essential for maintenance of chromosome linearity, a concept previously proven by Gomez-Escribano *et al.*, (2021), and multiplexing *tap-tpg* knockdown plasmids showed the loss of various plasmids, even when the *tap-tpg* on the plasmid was not silenced. Therefore, it remains unclear whether Tap4 (pSCL4), Tap3 (pSCL3) and Tap2 (pSCL2) lead end-patching of the 3' terminal repeat ends of any other replicon. To explore the process of end-patching further and better understand how Tap binds to the 3' single-stranded DNA overhang, allowing for priming of the 5' end lagging strand during replication in *S. clavuligerus*, we set out to characterise the Tap4, Tap3 and Tap2 proteins with the aim of overexpressing, purifying, and determining their relationship with the terminal ssDNA ends.

5.1 Genetic and phylogenetic characterisation of the terminal proteins of *S. clavuligerus*

Terminal proteins (TPs) are highly conserved in their size as well as their structure and are encoded as an operon in most *Streptomyces* species (Yang *et al.*, 2002). These operons are often located towards the end of the replicon, as seen for *tap-tpg*$_3$ and *tap-tpg*$_2$, as well as other *Streptomyces* species (Yang *et al.*, 2017; Algora-Gallardo *et al.*, 2021a). The pSCL4 megaplasmid of *S. clavuligerus* is a notable exception (**Figure 5.1.**), with the operon positioned at the centre of the replicon, which may be significant in terms of operon conservation as operons positioned in the centre of linear replicons are more conserved compared to the variable replicon arms (Hoff *et al.*, 2018). The reasons for the differing *tap-tpg*$_4$ operon position remains unclear, however its central position on the replicon may be indicative of increased importance relative to other end-proximal *tap-tpg* operons, and indicative of a need to maintain it by positioning it closer to the origin. Previous studies by Yang *et al.*, (2017) and Algora-Gallardo *et al.*, (2021) have highlighted that certain *Streptomyces* species tend to carry the TPs on their chromosome like *S. hygroscopicus subsp. Jinggangensis* (Wu *et al.*, 2012), whilst others do not .

Instead, the strains which do not carry the TPs on their chromosome tend to carry them on the GLPs, like *S. clavuligerus*. In some cases (Wu *et al.*, 2012), there are multiple *tap* and/or *tpg* copies which are present on the same replicon or on two or more replicons (i.e., both the chromosome and plasmids), by *S. clavuligerus* DSM 738 in this work (**Figure 5.1.**).

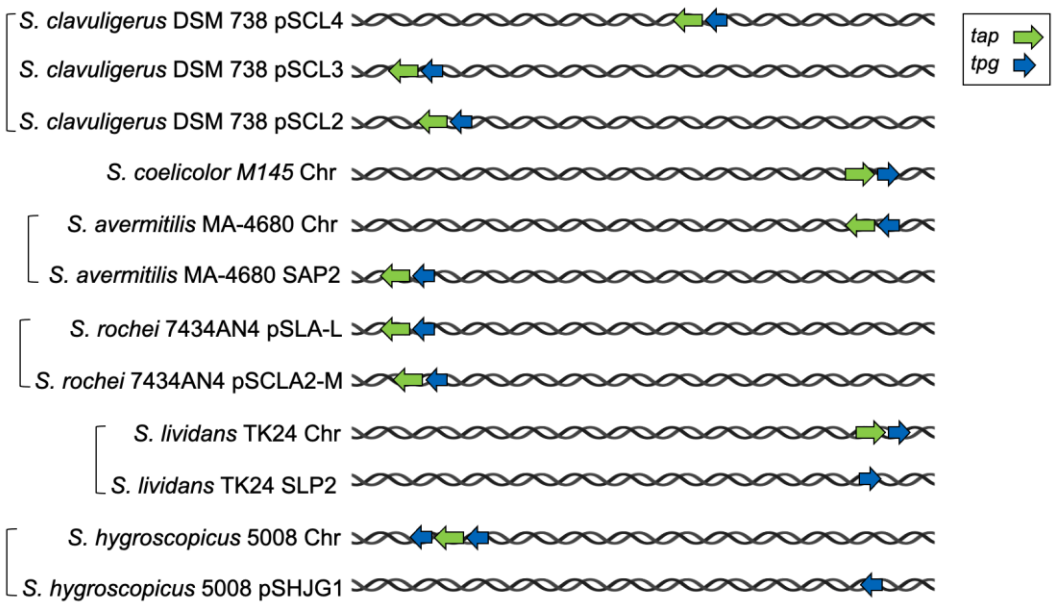


Figure 5.1. Graphical representation of the location of *tap-tpg* operons on archetypal *Streptomyces* replicons. For most strains, *tap* and *tpg* are located towards replicon ends, aside from pSCL4 in *S. clavuligerus* DSM 738, where they are centrally located. Additional copies of *tpg* are common for Streptomyces species with archetypal telomeres, such as *S. hygroscopicus* 5008. Other species, like *S. lividans* TK24 possess replicons only encoding *tpg*, common when both *tap* and *tpg* and already encoded on the chromosome. Not to scale.

Whilst linear replication begins at *oriC* for *Streptomyces* alongside other bacteria, e.g., *Borrelia burgdorferi*, end-patching and the presence of Tap and Tpg proteins remains unique to the *Streptomyces* (Yang *et al.*, 2017). TP end-patching is of importance in more extensively studied systems like phage ø29 (Salas, 1991) as well as adenoviruses (Liu, Naismith and Hay, 2003), however these examples of TP end-patching differ mechanistically from *Streptomyces*. In the phage ø29 and adenoviruses, primed synthesis means that the entire replicon duplicates from one end to another. In *S. clavuligerus* bi-directional linear replication and maturation of the Okazaki fragments leaves a single-stranded gap at the telomere ends (roughly 150-300bp), which is patched by Tap and Tpg (Yang *et al.*, 2017).

The highly conserved TP sequences are characteristic of archetypal telomeres, which are the most frequently found chromosomal and plasmid telomeres in *Streptomyces spp.*, however non-archetypal telomeres have recently been identified, contributing to a total of six different types of telomere structures including the SCP1-type telomere from *S. coelicolor* (Huang *et al.*, 2007), the Sg13350-type telomere from *S. griseus* (Ohnishi *et al.*, 2008), the Sg2247-type telomere (Algora-Gallardo *et al.*, 2021), the pLR1 telomere from *Streptomyces sp.* 44030 and the pLR2 telomere from *Streptomyces* sp. 440414 (Zhang *et al.*, 2006b). To establish the sequence similarity between the Tap and Tpg proteins on *S. clavuligerus* DSM 738 and the genera *Streptomyces* and *Kitasatospora*, a maximum likelihood phylogenetic tree was created using MEGA11 (Tamura, Stecher and Kumar, 2021). This tree included the Tap and Tpg amino acid sequences of 20 *Streptomyces* strains, representing a total of 25 different Tap or Tpgs on replicons of these strains. This tree was based on information provided in the study by Algora-Gallardo *et al.*, (2021), as well as unpublished work (Algora-Gallardo, unpublished data), who identified complete *Streptomyces* telomere sequences. *Streptomyces* and *Kitasatospora* sequences that were not flanked by the last 150 bp of known telomeres (mentioned before) were excluded. Additionally, the replicons of the chosen strains possessed both LHS and RHS telomeres, and only archetypal telomere-containing strains were chosen for phylogenetic analysis of the Tap and Tpg proteins.

An unrooted phylogenetic tree for was chosen to visualise the sequence similarity between Tap proteins and was centred at the mid-point, as the ancestral Tap in the data remained unknown (**Figure 5.2**) (Algora-Gallardo *et al.*, 2021). Notably, the *S. clavuligerus* DSM 738 Tap proteins were all found in different nodes of the phylogenetic tree, illustrating their differing ancestral relatedness, and potentially role in end-patching.

It is noticeable that the Tap2 proteins of *S. clavuligerus* are the most similar to those of other *Streptomyces spp*., particularly *Streptomyces globisporus* C1027 SGLP1, which was positioned in the same clade. Phylogenetically, Tap3 clustered together with the TP of *Kitasatospora setae* KM-6054, as well as the pSA3239 plasmid from *S. lavendulae*. Tap4 clustered together with the *S. anulatus* ATCC1123 chromosome Tap and plasmid Tap, in addition to the chromosomal terminal protein of *S. antibioticus* DSM 41481. Overall, Tap4 and Tap3 were found to be most similar to each other, whereas Tap2 was found to cluster with TP sequences of organisms that were phylogenetically distant to the other Tap proteins of *S. clavuligerus*, such as *S. coelicolor* A3(2), *S. avermitilis* MA4680 or *S. lividans* TK24.

The Tpg maximum likelihood phylogenetic tree shown in **Figure 5.3** presented similar trends to the Tap tree. The Tpg2 protein was most similar to terminal proteins on chromosomes and plasmids of organisms like *S. avermitilis* MA4680, *S. hygroscopicus* subs.jinggangensis 5008, *S. ambofaciens* ATCC23877, *S. lividans* TK24, *S. glaucescens* GLA0, *S. coelicolor* A3(2), forming clades. Tpg2 also shares nodes with Tpg3, although all the Tpg proteins of *S. clavuligerus* DSM 738, like the Tap proteins, were positioned several nodes apart on separate branches. Tpg4 shared a clade with *S. anulatus* ATCC1123, like Tap4, as well as *S. antibioticus* 41841. The maximum likelihood phylogenetic trees highlight, that Tap-Tpg4, Tap-Tpg3 and Tap-Tpg2, despite being found in the same organism, are phylogenetically distinct, and presents further evidence that they may have originated from gene duplication or recombination events which meant that over time, their sequences diverged.

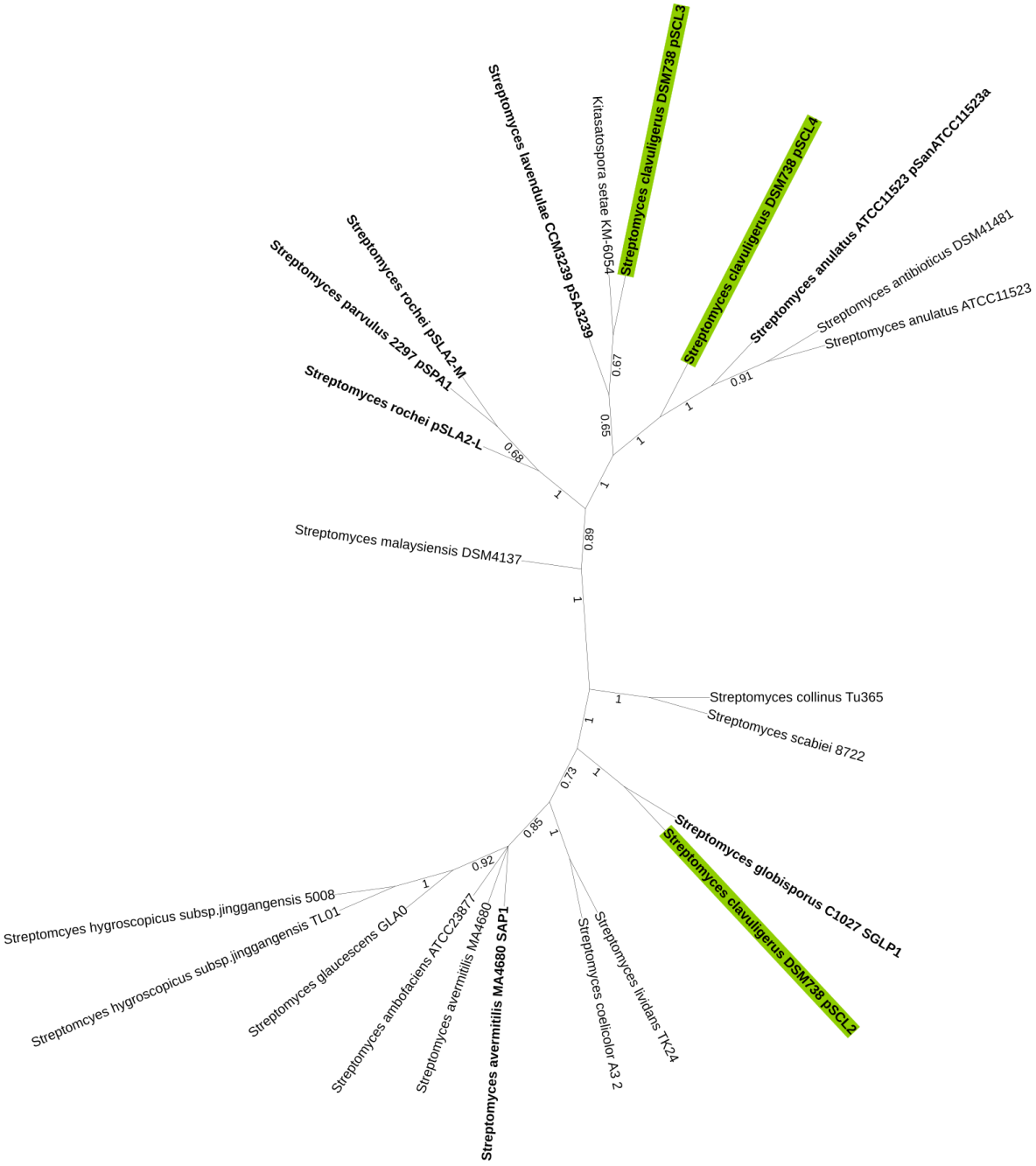


Figure 5. 2. Maximum likelihood phylogenetic trees of Tap proteins found on Streptomycetota with archetypal telomeres. *S. clavuligerus* terminal proteins are highlighted in green. Tree constructed on ClustalW and MEGA11 with 100 bootstraps and edited on iTOL (Letunic and Bork, 2021). The percentage of replicate trees that showed associated taxa clustered together in the 100 replicate bootstrap test are shown beside branches (Felsenstein, 1985). Regular = found on chromosome / Bold = found on plasmid.

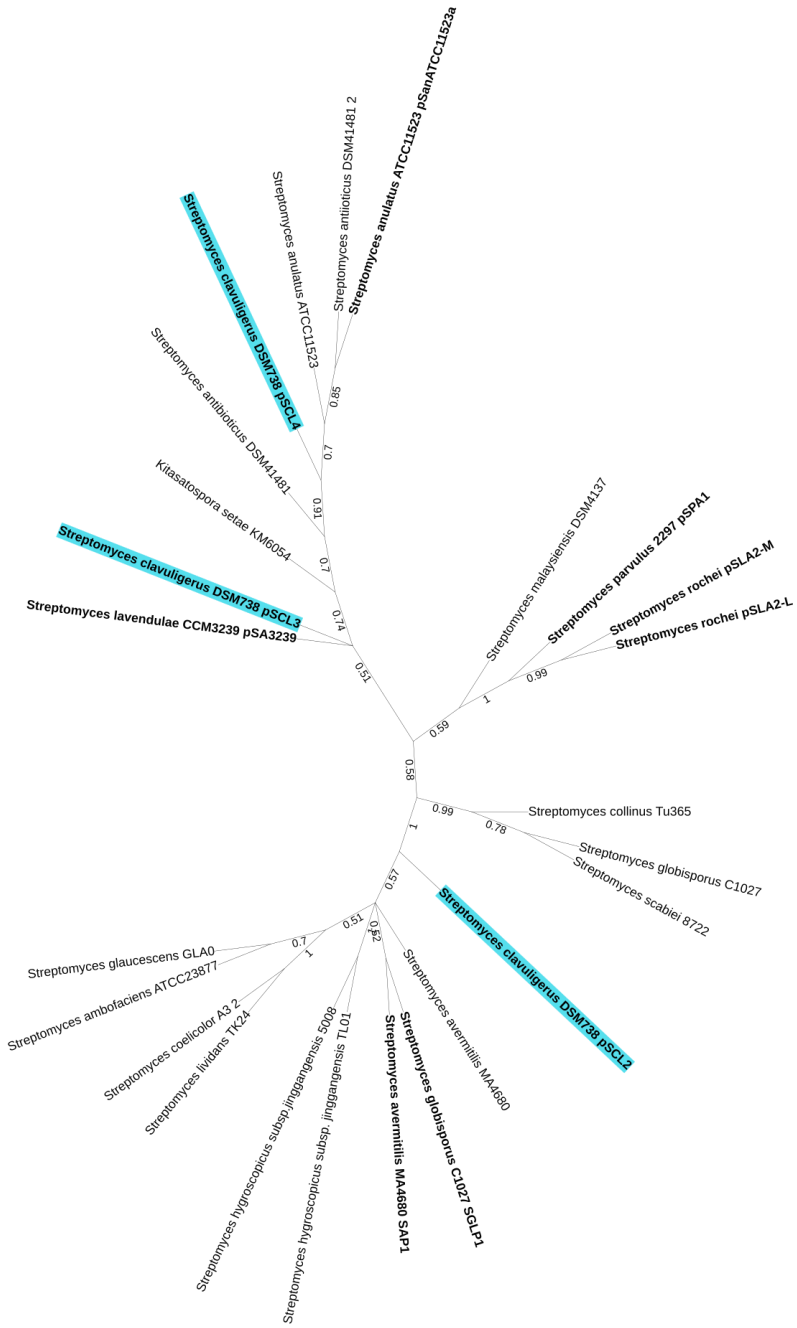


Figure 5. 3. Maximum likelihood phylogenetic trees of Tpg proteins found on Streptomycetota with archetypal telomeres. *S. clavuligerus* terminal proteins are highlighted in blue. Tree constructed on ClustalW and MEGA with 100 bootstraps and edited on iTOL (Letunic and Bork, 2021). The percentage of replicate trees that showed associated taxa clustered together in the 100 replicate bootstrap test are shown beside branches (Felsenstein, 1985). Regular = found on chromosome / Bold = found on plasmid.

5.2 Predicted secondary protein parameters and alignment of Tap-Tpg4, Tap-Tpg3 and Tap-Tpg2.

Streptomyces clavuligerus DSM 738 encodes 3 different Tap proteins and 3 different Tpg proteins, of varying amino acid length shown in **Table 5.1.** The protein parameters were calculated using the ExPASy tool on the Swiss Institute of Bioinformatics server (Gasteiger *et al.*, 2003). Tap proteins are generally larger in size than Tpg proteins and comprise a length of around 735-859 amino acids and atomic mass of around 79-88 kDa. Their large physical size and amino acid residue composition contributes to a variety of chemical factors such as the aliphatic index, which has been associated with a protein's thermostability and is often associated with the instability index too (Guruprasad, Reddy and Pandit, 1990). A higher aliphatic index highlights increased thermostability, and is thus linked to a lower instability index (Idicula-Thomas and Balaji, 2005). *S. clavuligerus* DSM 738 Tpg proteins are around 184-192 amino acids long with an atomic mass of around 20-21 kDa and thus exhibit a lower instability index and a higher aliphatic index, along with a different composition of amino acids contributing to their instability, as some amino acids like lys(K) and Gly(G) have been statistically shown to be associated with higher protein instability (Guruprasad, Reddy and Pandit, 1990).

Table 5.1 Physical and chemical parameters of the Tap and Tpg proteins of *S. clavuligerus* DSM 738.

NAME	AMINO ACIDS	MOLECULAR WEIGHT (KDA)	THEORETICAL PI	INSTABILITY INDEX	ALIPHATIC INDEX	CHARGED RESIDUES (+)	CHARGED RESIDUES (-)
TAP4	859	88.2	6.57	50.60	78.85	82	85
TAP3	832	88.7	5.89	43.20	78.23	84	96
TAP2	735	79.0	8.43	35.08	83.74	90	86
TPG4	184	21.3	10.25	46.83	81.20	37	29
TPG3	192	21.0	9.72	33.56	80.47	33	25
TPG2	186	20.6	10.03	32.4	84.57	34	23

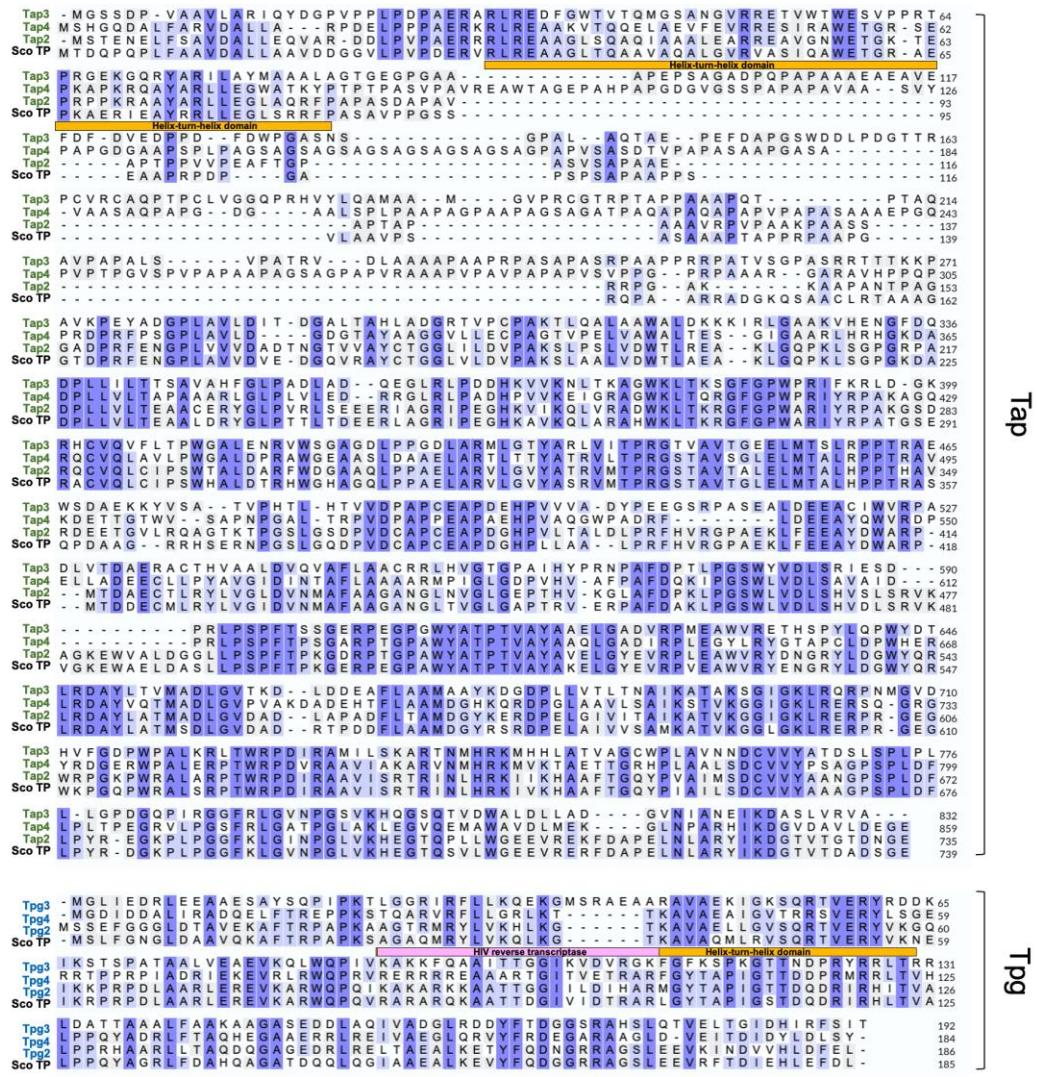
Generally, *Streptomyces* terminal proteins are characterised by several functional domains, such as the helix turn helix (HTH) domain and nuclear localisation signals as well as a ß-strand that potentially involved in protein-protein interactions (Yang *et al.*, 2013). HTH domains are associated with DNA-binding proteins, and we hypothesised that some of the Tap and Tpg proteins may therefore include HTH domains, allowing

them to bind the telomere ends during terminal replication. In order to investigate this, we analysed the predicted domain structures using PROSITE (Sigrist *et al.*, 2013) for each of the six Tap and Tpg proteins of *S. clavuligerus* DSM 728, alongside an additional archetypal TP from *S. coelicolor* for comparison (**Figure 5.4**). PROSITE scanned the four *S. clavuligerus* and *S. coelicolor* amino acid sequences in search of Tpg domains, however none of the sequences contained a helix turn helix motif using the tool, suggesting that the Tpg proteins cannot bind DNA. Studies by Bao and Cohen, (2001 and 2003) and Yang *et al.*, (2002) however identified helix-turn-helix motifs on Tpg, suggesting the capability of the protein to partake in protein-primed DNA synthesis, which was also experimentally proven in the studies. Therefore, the sequences of the four Tap and four Tpg proteins were aligned using ClustalO (Sievers *et al.,* 2011), and literature was used to identify a helix-turn helix motif in addition to a HIV reverse transcriptase motif (**Figures 5.4A**), which was first identified by Bao and Cohen, (2001) for *S. lividans* and *S. rochei*. The presence of a reverse transcriptase motif and helix turn helix domain on Tpg illustrated the role of Tpg in priming the lagging strand during linear replication, a concept that was disputed until this domain was identified (Bao and Cohen, 2001).

Figure 5.4 B depicts the percentage identity matrix of the Tap and Tpg proteins of pSCL4, pSCL3 and pSCL2 when aligned with CLUSTALO and the SCO chromosomal TPs, showing variability. This approach once again highlights that Tap2 and Tpg2 were more similar to the Tap and Tpg proteins of SCO, than the Tap and Tpg proteins of pSCL3 and pSCL4 of *S. clavuligerus* DSM 738. This may indicate that the genes encoding these proteins were gained through recombination or horizontal gene transfer i.e., through telomere switching, or that their functions differ from those of Tap-Tpg3 and Tap-Tpg4 (Tidjani, Bontemps and Leblond, 2020) (also shown by phylogenetic tree in **Figure 5.2** and **Figure 5.3**). CLUSTALO indicated an average 46.37% identity of Tap3 with Tap4, Tap2 and SCO TP and an average 43.31% sequence identity of Tpg3 with Tpg4, Tpg2 and SCO TP. The results of CLUSTALO also showed a slightly higher average percentage identity of 52.34% between Tap4 and the other *S. clavuligerus* and *S. coelicolor* Tap proteins (the lowest % identity was between Tap4 and Tap3) and also showed an average 46.92% identity between Tpg4 with the other *S. clavuligerus* and *S. coelicolor* Tpg proteins. Overall, Tap3 and Tpg3 show the lowest percentage identities

when compared to other terminal proteins of the same organism, perhaps due to a difference in their structure or evolutionary recombination events.

A



B

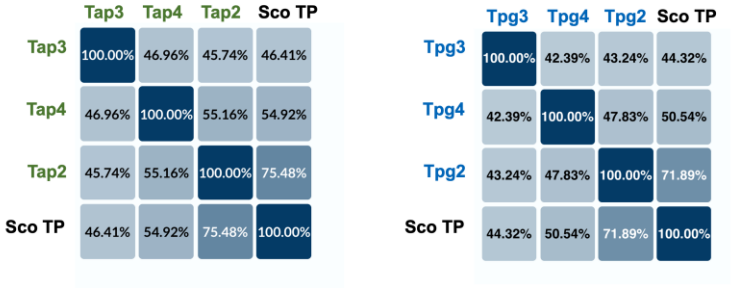


Figure 5. 4. Alignment and percentage identity matrix of the Tap and Tpg amino acid sequences of *S. clavuligerus* pSCL4, pSCL3 and pSCL2 and the *S. coelicolor* chromosome. Alignments made in Uniprot using Clustal Omega (Sievers et al., 2011).

5.3 Predicted secondary structure of Tap-Tpg4, Tap-Tpg3 and Tap-Tpg2

To find out the predicted secondary structure of Tap and Tpg in *S. clavuligerus*, and determine the implications between structural differences of the proteins, 3D models of the TPs were computed using Alphafold2 (Jumper *et al.*, 2021). Alphafold2 has demonstrated to predict highly accurate protein structures which are similar to those derived through conventional structural approaches such as X-ray crystallography (Jumper *et al.*, 2021). We therefore decided to submit the amino acid sequences of our proteins of interest (Tap4, Tap3, Tap2, Tpg2, Tpg3, Tpg4) to Alphafold2 for structural prediction and to confirm the previous results produced with PROSITE and Uniprot. Alphafold2 allowed for the prediction of an accurate 3D structure based on the per-residue confidence scores (pLDDT) which were between 0 and 100, where lower scores represent unstructured (in isolation) parts of the local protein structure. **Figure 5.5** shows the *S. clavuligerus* DSM 738 Tap protein structure predictions at rank 0. The pLDDT scores indicate that for the most part, the Alphafold2 AI algorithm is confident in the prediction of the central part of the amino acid residues, highlighted by the colour dark blue. Lower pLDDT scored can be witnessed by the colour red, shown by a larger, disordered beta sheet fragment of the Tap proteins. For all Tap proteins, the NH_2-terminus (N-terminus) and COOH - terminus (C-terminus) are marked. The predicted Tap4 structure showed an exposed N-terminus and a hidden C-terminus, the predicted structure of Tap3 showed exposed N and C termini, while the predicted structure of Tap2 also indicated an exposed N-terminus and hidden C-terminus. The position of the termini become important when considering any sort of protein overexpression work using terminus tagging i.e., His-Tagging (demonstrated in section 5.4 of this chapter), therefore identification of the tag's exposure was essential (Bao and Cohen, 2003). This could be concluded from the Alphafold2 ranked structures of the TPs by determining which N and C termini were exposed and played a role later in this chapter, as exposed termini allowed for exposure of the His-Tags in downstream steps like protein purification, permitting for nickel column binding of the His-tag and elution.

Figure 5.6 shows the Tpg Alphafold2 predicted structures, showing high pLDDT scores for Tpg4, Tpg3 and Tpg2 structures. The alpha helix and beta sheet folding shows a less complex structure for Tpg proteins, also underlined by the visibility of the N and C termini in all *S. clavuligerus* Tpg proteins. All regions seem to show high confidence scores, except the N terminus of Tpg2, which shows a <50 score. The secondary structures of

Tap proteins are further underlined by the pSPIRED structural analysis (McGuffin, Bryson and Jones, 2000) in **Figures 5.7**, **5.8** and **5.9**. PSIPRED analysed PSI-BLAST (position specific iterated) outputs and performed its accurate secondary structure prediction based on two feed-forward neural networks. The Tpg areas of lower confidence as well as secondary structure predictions aligned with the amino acid residues can be seen by **Figure 5.10**.

pLDDT alone was not sufficient in determining whether AlphaFold2 is confident in the relative positions of the domains, further analysis i.e., the predicted aligned error (PAE), was investigated to determine confidence in the domain positions in a multi-domain protein such as Tap. The predicted aligned error is highlighted in Figure 5.5. as a 2D plot, with residue number running along both axes. The colour at coordinates x,y indicates the predicted error at residue x where lighter green indicates higher error and lower confidence. Confidence is higher within a domain than between domains, often domain structure is visible in the PAE plot, as shown in **Figure 5.5**, which show one main domain and smaller domains with relatively high confidence scores around the central diagonal axes, however not around it. Low PAE scores could also be a result of the predicted structure demonstrating a protein in isolation, therefore any other interactions i.e., through a linker or a surrounding complex are disregarded and as such at position x, the equivalent y PAE value might be higher than expected. Tap comes from a family of lambda-repressor-like DNA-binding domains, therefore it may form homodimers with another Tap monomer, allowing it to bind to the target DNA as seen for the bacteriophage lambda (CATH: Protein Structure Classification Database, Sillitoe *et al.*, 2019). The smaller domains correspond to roughly the first 0-100 residues and the larger domain corresponds to the last 300 to 850 residues for Tap4 and Tap3 and the last 150 to 750 residues of Tap2. This underlines the similarities in PAE between Tap structures, as well as parallels between number of domains, supporting the terminal protein similarities shown in previous sections of this chapter. PAE graphs for Tpg proteins highlight large areas with high confidence and low error, and do not present any specific areas of higher confidence, suggesting that there is only one main domain for *S. clavuligerus* Tpg proteins and a low PAE overall in isolation.

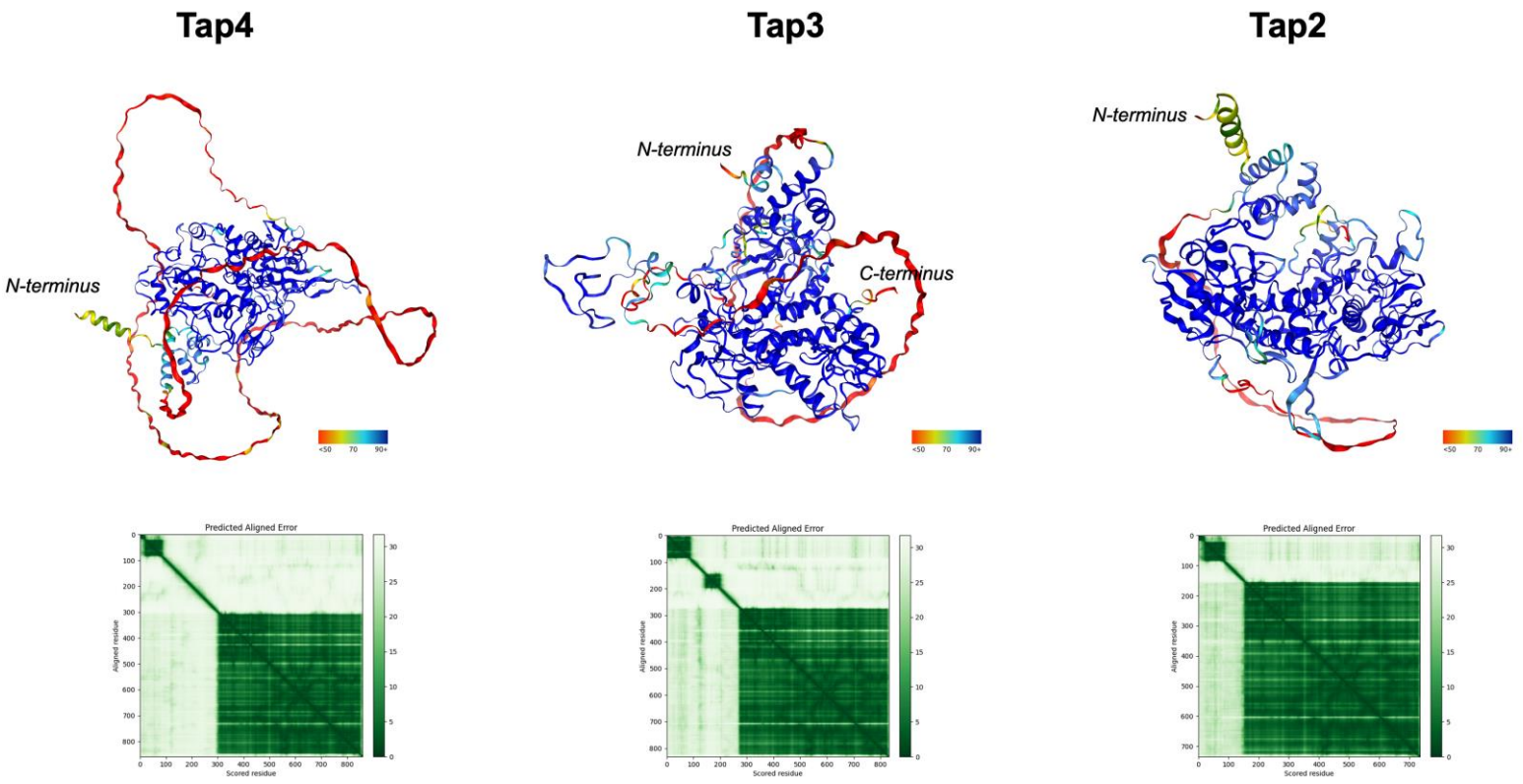


Figure 5. 5. Alphafold2 3D model predictions and predicted pLDDT and PAE graphs of Tap4, Tap3 and Tap2 structures. N and C termini are highlighted, and 3D model predictions of the proteins follow a pLDDT scoring system between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation, shown by colour-scheme key below each protein structure. Corresponding PAE plots are depicted below Tap 3D structures. (Jumper *et al.,* 2021).

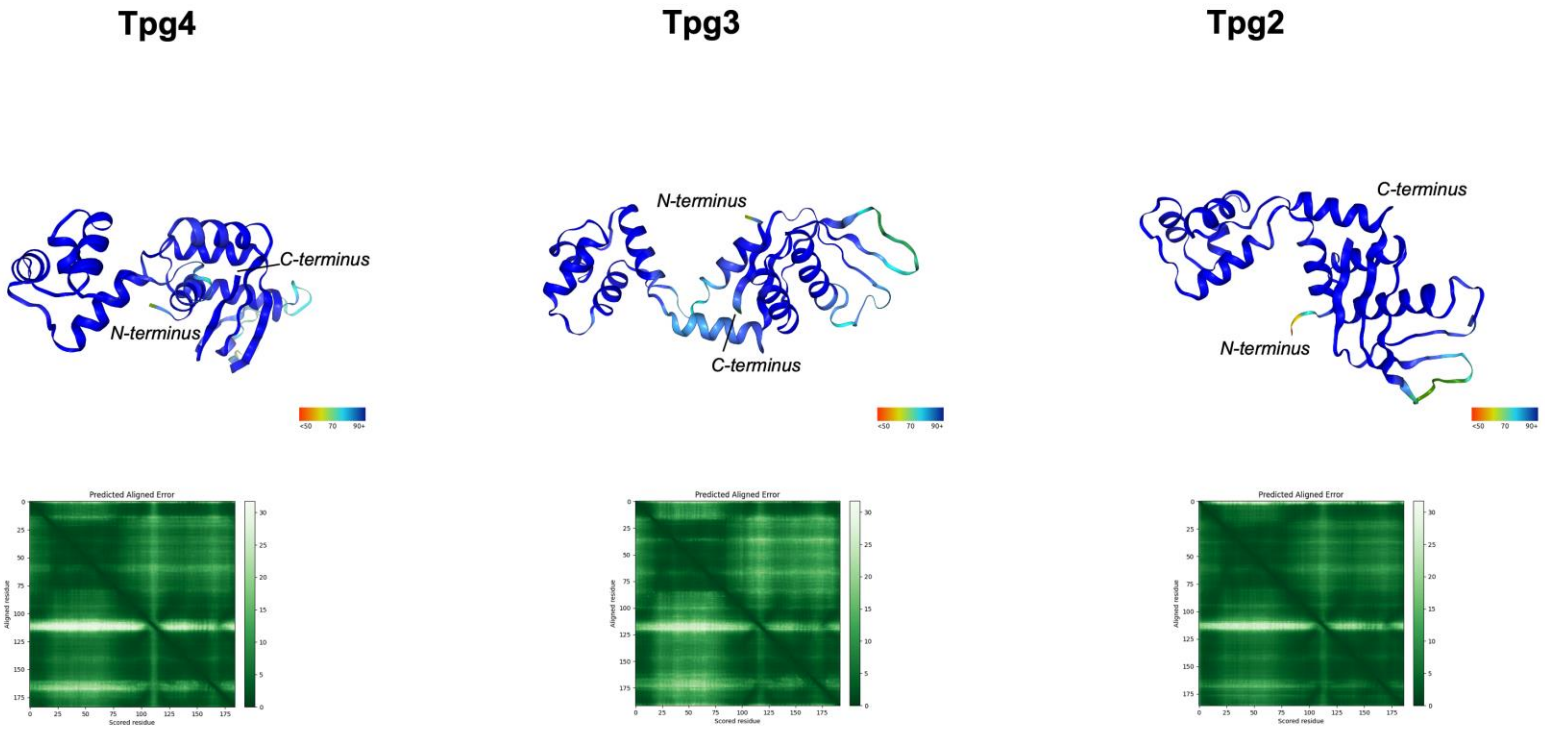


Figure 5. 6. Alphafold2 3D model predictions and predicted pLDDT and PAE graphs of Tpg4, Tpg3 and Tpg2 structures. N and C termini are highlighted, and 3D model predictions of the proteins follow a pLDDT scoring system between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation, shown by colour-scheme key below each protein structure. Corresponding PAE plots are depicted below Tpg 3D structures. (Jumper *et al.,* 2021).

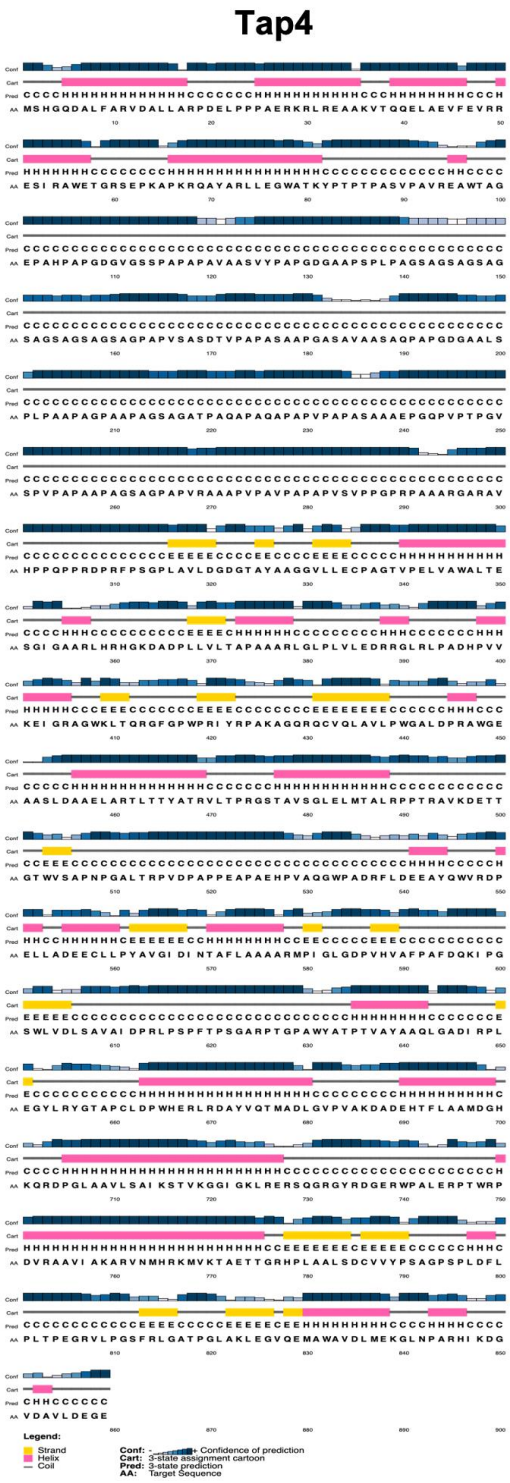


Figure 5. 7. Predicted secondary structure of *S. clavuligerus* Tap amino acid sequences. The PSIPRED v4.0 workbench was used to predict secondary structures highlighting strands (yellow), helices (pink) and coils (grey line). Confidence predictions are shown as blue coloured bars and target amino acid sequence are displayed below the secondary structure characteristics. (McGuffin, Bryson and Jones, 2000)

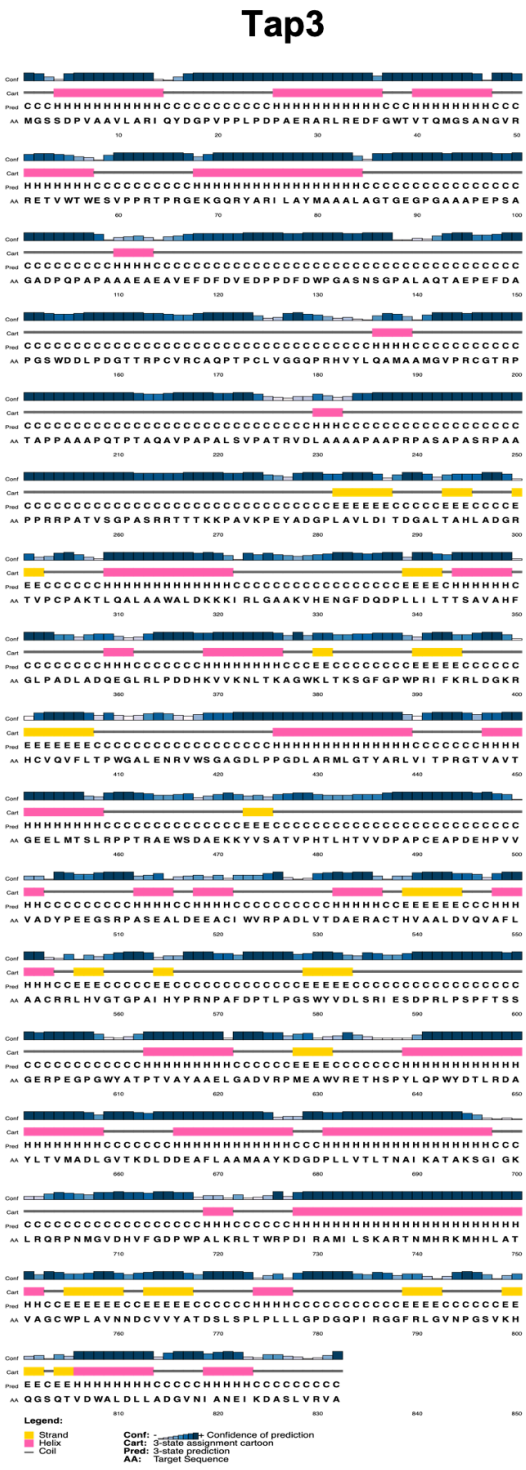


Figure 5. 8. Predicted secondary structure of *S. clavuligerus* Tap amino acid sequences. The PSIPRED v4.0 workbench was used to predict secondary structures highlighting strands (yellow), helices (pink) and coils (grey line). Confidence predictions are shown as blue coloured bars and target amino acid sequence are displayed below the secondary structure characteristics.(McGuffin, Bryson and Jones, 2000)

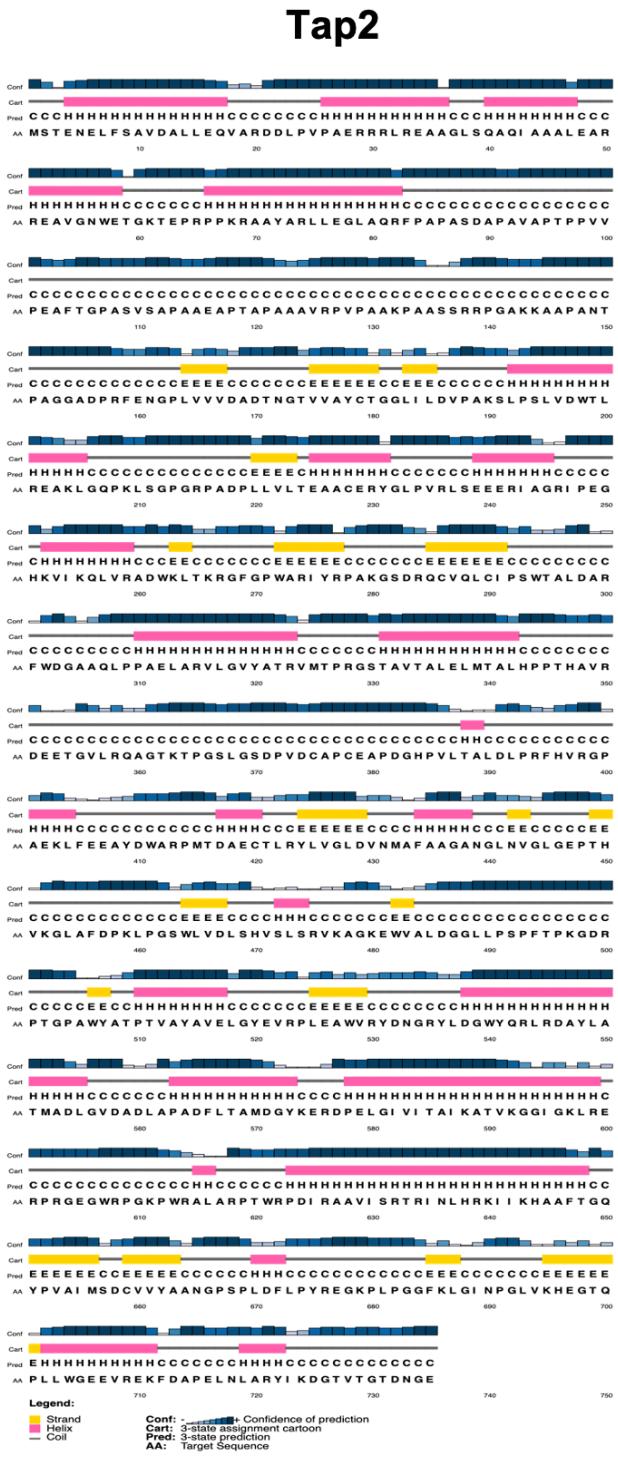


Figure 5. 9. Predicted secondary structure of *S. clavuligerus* Tap amino acid sequences. The PSIPRED v4.0 workbench was used to predict secondary structures highlighting strands (yellow), helices (pink) and coils (grey line). Confidence predictions are shown as blue coloured bars and target amino acid sequence are displayed below the secondary structure characteristics. (McGuffin, Bryson and Jones, 2000)

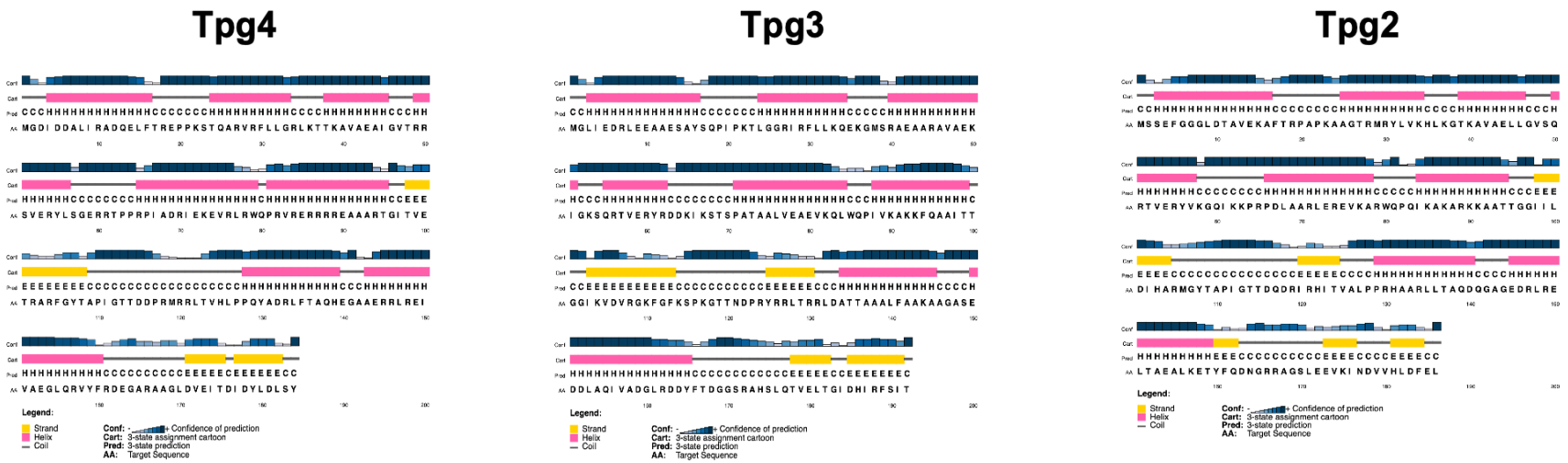


Figure 5. 10. Predicted secondary structure of *S. clavuligerus* Tpg amino acid sequences. The PSIPRED v4.0 workbench was used to predict secondary structures highlighting strands (yellow), helices (pink) and coils (grey line). Confidence predictions are shown as blue coloured bars and target amino acid sequence are displayed below the secondary structure characteristics. (McGuffin, Bryson and Jones, 2000)

To obtain a more detailed overview of the similarities and differences between our proteins of interest, the. pdb files of the predicted protein structures produced by Alphafold were run though PDBe Fold v2.59. (Krissinel and Henrick, 2004). The output presented various scores of a pairwise comparison of 3D structure alignment and secondary structure matching (**Table 5.2**) visualising the previously predicted shared domains of the proteins. All *S. clavuligerus* DSM 738 Tap proteins were compared to each other and the same was done for the Tpg proteins, demonstrating similar trends to the percentage identity of amino acid residues (**Figure 5.4**).

The quality of alignment (Q) score considers alignment length and root mean square deviation (RMSD) and presents a more meaningful score, when the RMSD is lower, and the number of aligned residues (NALGN) is higher and uses measurements between 0 and 1. This can often be affected by alignment length, i.e., if the protein is smaller, the alignment length is also smaller and therefore the Q score could overall be higher as the RMSD lowers. The highest Q score was seen between Tpg4 and Tpg2, which presented the highest percentage identity as well. For the Tap proteins, Tap4 and Tap2 presented the highest percentage identity, however their Q score is not representative of that in a dataset also looking at Tpg, as the Tpg proteins contain a much shorter chain of amino acids, as mentioned earlier. P and Z scores measure the statistical significance, and the P score is used to calculate the Z score in terms of Gaussian statistics. A higher Z-score thus indicated a higher statistical significance, and the highest Z-score can be seen between Tap4 and Tpg3 structure alignments. The Z score is lower for Tpg proteins, highlighting less significance.

RMSD is a commonly used method to identify similarity between protein structures. It shows the root mean square between carbon atoms of protein chains and is measured in angstroms. Therefore, a higher RMSD value indicates how close or separated a pair of matched carbon atoms is at the best superimposed 3D position, calculated by the PDBe Fold algorithm. A low RMSD value, as seen for Tap4 and Tap2 alignments which present the lowest value (1.13 angstroms), is therefore an indicated that the two structures are matched and thus a smaller value represents a more similar structure. Generally, the RMSD is a lot lower between Tap proteins than between Tpg proteins, with the highest score being 3.19 angstroms for Tpg3 and Tpg2 alignments.

Further metrics such as the number of aligned residues (NALGN), the number of gaps between residues (NGAPS) and the secondary structure matching as a percentage (SEQ%) were investigated (**Table 5.2**). Overall, alignments between Tap proteins presented higher NALIGN scores, which also makes sense as their sequence length is longer. Within the Tap protein comparisons, Tap3 and Tap2 show the highest NALGN values, however Tap4 and Tap2 alignments show the highest secondary structure match (SEQ%) and lowest number of gaps. For the Tpg proteins, the highest NALGN score was the highest between Tpg4 and Tpg2 and the lowest for Tpg4 and Tpg3. Similar to the percentage identity scores, the Tap and Tpg proteins seem to have the least in common with Tap and Tpg on pSCL3. Tpg4 and Tpg2 3D structure alignments also presented the lowest NGAPS values and highest SEQ% match, indicating that they are structurally the most similar of the Tpg proteins.

Table 5.2. Pairwise comparison of 3D structure alignment of Tap and Tpg proteins in S. clavuligerus DSM 738.

	Tap4-Tap3	***Tap4-Tap2***	***Tap3-Tap2***	***Tpg4-Tpg3***	***Tpg4-Tpg2***	***Tpg3-Tpg2***
Q SCORE	0.34	0.39	0.40	0.28	0.69	0.31
P score	42.4	38.1	29.7	4.06	11.1	1.02
Z score	22.3	21.7	19.8	8.15	9.70	3.69
RMSD	1.25	1.13	1.25	2.03	1.81	3.19
Nalgn	533	532	537	121	180	153
Ngaps	12	12	16	6	2	8
Seq%	0.53	0.58	0.50	0.40	0.45	0.33

5.4 pET vector construction and target cloning of Tap4, Tap3 and Tap2.

Various studies have explored the relationship of *Streptomyces* TPs with telomere ends to characterise and better understand linear replication (Bao and Cohen, 2003; Yang *et al.*, 2013; Yang, Tseng and Chen, 2015). The studies focussed on the terminal proteins of S*. lividans* and *S. coelicolor,* using various expression hosts like *E. coli* and *S. lividans* to overexpress their TPs. The S. clavuligerus DSM 738 TPs remain uncharacterised as no studies have attempted to overexpress and/or purify Tap and Tpg from pSCL4,

pSCL3 or pSCL2. Optimising overexpression and consequently purifying DNA-binding proteins is important for understanding what role they play in interacting with DNA sequences and whether Tap and Tpg act in cis or in trans on the various replicons of *S. clavuligerus* to determine their essentiality.

To overexpress Tap4, Tap3, and Tap2, the amino acid sequences of the proteins were codon-optimised for the expression host of choice, which was *E. coli* BL21 (DE3). Target genes were eventually cloned into a pET vector, which was under control of T7 bacteriophage transcription, where induction of expression occurred by providing the host cells with T7 RNA polymerase. DNA sequences were codon optimised using the GenScript codon optimisation tool and subsequently ordered from the company, directly cloned into the vector pUC57. The codon optimised sequences were constructed based on the predicted protein structures from Section 5.3, as a His-Tag was added onto the N-terminus of each Tap protein as the predicted structures produced by Alphafold2 indicated that these termini were exposed and suitable for tagging. The N-terminal end was predicted to be the most exposed by Alphafold2 and was thus picked for His-6 tagging, allowing for nickel column purification of the recombinant proteins and identification in later steps such as with a Western Blot. Additionally, the codon optimised sequences were designed to include EcoRI and NdeI restriction sites on the N-terminal end and a BamHI site on the C terminal end, which facilitated the restriction cloning of the sequences into the multiple cloning site (MCS) of two vectors: pBGS19 and pET15b+. pBGS19 includes a kanamycin resistance gene and pET15b+ includes an ampicillin resistance gene, which allowed for transformant selection during the transformation steps of the restriction cloning process highlighted in **Figure 5.11**.

Initially, both the pBGS19 and Tap-pUC57 (pUC57_34640, pUC57_28815, pUC57_28345) constructs were double digested with EcoRI and BamHI, the DNA was purified and the whole reaction was ligated overnight. White-coloured transformants (Tap sequences inserted into MCS = interruption of the lacZ gene in which the MCS is positioned) that were kanamycin resistant, but ampicillin sensitive were picked and sent off for Eurofins Sanger sequencing. The Tap4/pBGS19, Tap3/pBGS19 and Tap2pBGS19 colonies with the correct insert were grown overnight and DNA was isolated. This allowed for restriction digestion with BamHI and NdeI of the Tap/pBGS19 constructs as well at the pET15b+ vector (**Figure 5.11**). Agarose gel electrophoresis

produced insert fragments of 2584 bp, 2506 bp and 2212 bp for Tap4, Tap3 and Tap2 and a backbone fragment of 5696 bp for the pET15b+ vector. These bands were excised, and gel purified, which allowed subsequent ligation to create the final construct, as the DNA fragments had compatible sticky ends. Additionally, transformants after step 2 shown in **Figure 5.11**, were screened in this way.

A PCR was performed to verify the final pET15b+ vector Tap construct, as well as whole plasmid sequencing. **TAP4CO_F/R, TAP3CO_F/R,** and **TAP2CO_F/R** primers were used to amplify amplicons of 725 bp, 551 bp, and 675 bp for Tap4/pET15b+, Tap3/pET15b+ and Tap2/pET15b+ respectively. **Figure 5.12** highlights the gel electrophoresis and amplified DNA fragments of the previously mentioned sizes, which were confirmed with Sanger sequencing and represented the final vectors used for transformation into *E. coli* BL21 (DE3) for recombinant protein overexpression.

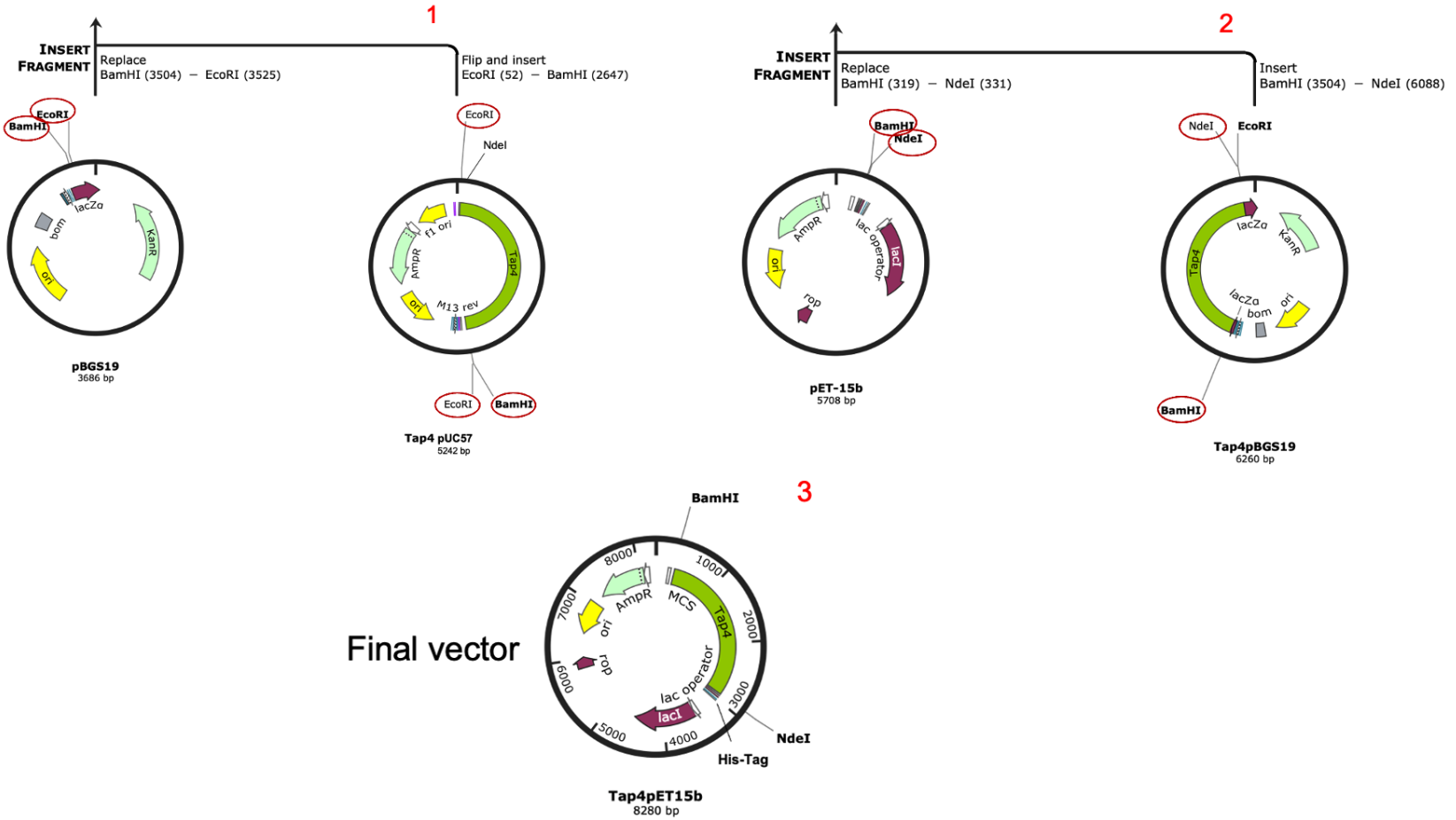
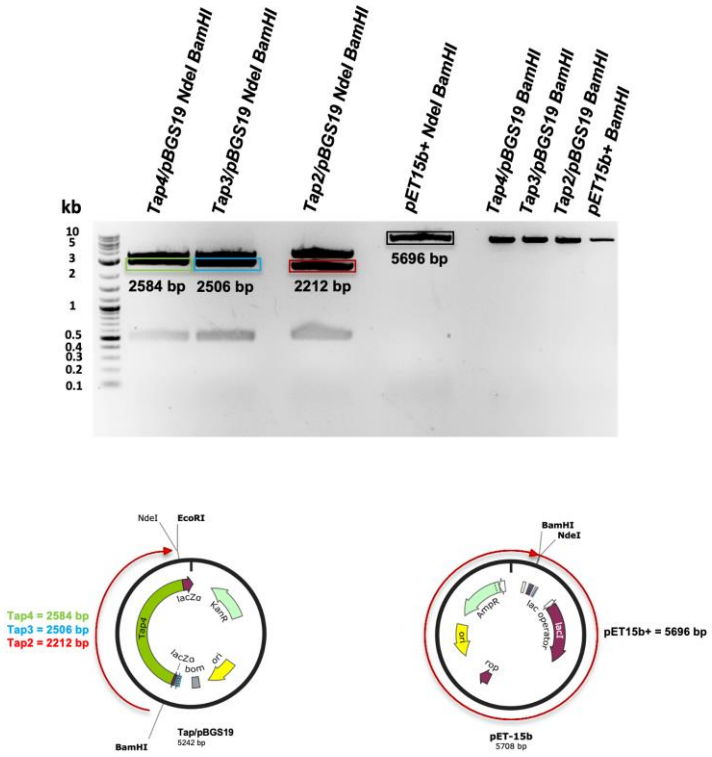


Figure 5.11. Overview of the Tap4 pET-vector restriction cloning process. 1) EcoRI/BamHI double restriction digestion of both Tap4/pUC57 and the pBGS19 vector to allow for ligation into Tap4/pBGS19. **2)** Tap4/pBGS19 and pET15b+ restriction digestion with BamHI and NdeI to make the final 8.2 kb vector (**3**): Tap4/pET15b+. Enzymes used in the restriction digestion process are highlighted in red circles

A



B

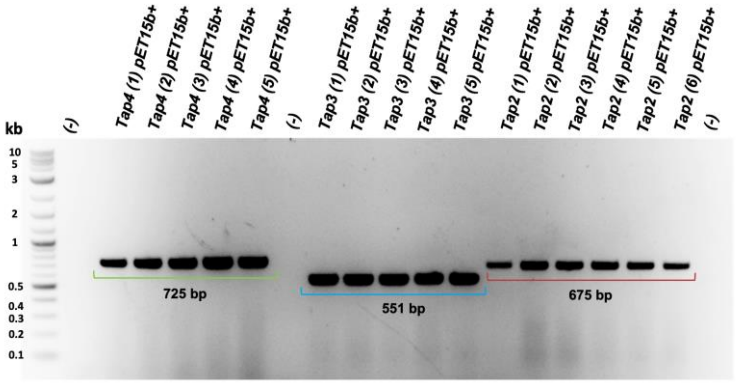


Figure 5.12. Restriction digest cloning overview to create final pET-vector. A) The three Tap/pBGS19 constructs along with the pET15b+ vector were restriction digested with NdeI and BamHI and cut out of the agarose gel, purified, and ligated. Fragments of 2584 bp, 2506 bp and 2212 bp for Tap4, Tap3 and Tap2 inserts and 5696 bp for the pET15b+ backbone. **B)** PCR and agarose gel to confirm final Tap overexpression constructs of 725 bp (Tap4/pET15b=), 551 bp (Tap3/pET15b+) and 675 bp (Tap/pET15b+).

5.5 Optimisation of Tap4, Tap3 and Tap2 protein overexpression.

To determine the interaction of the Tap and Tpg proteins with the telomere ends in *S. clavuligerus* DSM 738 and inform us about which of the 3 Tap and Tpg protein combinations bind to the telomeres of the organism, we optimised recombinant protein overexpression. For this, we focussed on the Tap proteins, as these bind to the 5' telomere ends and recruit Tpg. Optimisation of *S. clavuligerus* Tap recombinant protein expression was essential, given that the protein had previously never been overexpressed. TPs have previously been overexpressed from other organisms like *S. coelicolor* (Bao and Cohen, 2003) and *S. lividans* (Yang *et al.*, 2013; Yang, Tseng and Chen, 2015) to better understand the biological and biochemical mechanisms of telomere binding and TP interactions during linear replication, such as deoxynucleotidylation of TPs. The latter studies overexpressed the TPs in *E. coli* overexpression strains, similar to methods used in this chapter. The cloned pET15b+/Tap4, Tap3 and Tap2 vectors were transformed into *E. coli* BL21 (DE3) and various methods were tested to determine the optimal conditions for overexpression which included: growth media, incubation/overexpression duration, induction method, and temperature.

The pET vector expression system consists of various genetic components which make it suitable for protein overexpression in a regulated manner, once in an appropriate expression host, *E. coli* BL21 (DE3). The expression of recombinant proteins encoded by pET vectors is tightly regulated by T7 polymerase encoded on the expression host cell chromosome, as the vector contains a T7 promoter and lac repressor for regulation of expression with lactose or isopropyl β- d-1-thiogalactopyranoside (IPTG), and a lac operator downstream of the promoter, allowing for inducible expression of our gene of interest (*tap*$_4$, *tap*$_3$ and *tap*$_2$). Inducible expression is vital during recombinant protein overexpression, as it can be controlled and provides higher recombinant protein yields (Studier and Moffatt, 1986). The chromosomally encoded T7 polymerase is regulated under *lac*UV5, therefore expression is normally induced by the addition of IPTG.

5.5.1 Autoinduction media does not result in protein overexpression of Tap4, Tap3 and Tap2.

The first set of Tap overexpression optimisation experiments focussed on utilising the *lac* operon encoded on the vector by using autoinduction media supplied with lactose, glucose and glycerol (Studier, 2005). Autoinduction in *E. coli* is generally well understood: glucose is initially used as a source of carbon for growth, however becomes depleted, leading the cells to relieve catabolite repression and depend on other carbon sources, in this case lactose and glycerol. Lactose utilisation by the cells allows for the production of allolactose through beta-galactosidase, which functions as the *lac* operon inducer, as it binds and releases the repressor, and increases protein expression (Blommel *et al.*, 2007). **Figure 5.13 A** illustrates an SDS-PAGE gel of crude Tap4, Tap3 and Tap2 protein extracts after autoinduction media optimisation trials at various temperatures. *E. coli* BL21(DE3) cells were inoculated into autoinduction media and grown at 37 °C and 30°C for a period of 16 and 20 hours to avoid beta-lactam antibiotic degradation by a secreted beta lactamase enzyme and compensate for antibiotic degradation by a natural drop in pH during fermentation. There were no clear, dark bands at the expected molecular mass for Tap4, Tap3 and Tap2 proteins (88.2 kDa, 88.7 kDa and 79.0 kDa respectively) compared to the uninduced (UI) control, highlighting that the method of autoinduction and/or the temperature was not suited for overexpression of these proteins by the host cells. The cells were thus grown using the same autoinduction method, at a lower temperature of 26 °C. **Figure 5.13 B** shows SDS-PAGE gels of Tap4, Tap3 and Tap2 crude extracts after growth in autoinduction media at 26°C using the empty pET-15b+/*E.coli* BL21 (DE3) transformant cells as an empty vector control. Additionally, 1 mL samples were taken of the cultures at various time points to determine whether utilisation of the carbon sources in the autoinduction media visibly affected protein overexpression. Similarly, no clear, darker bands can be seen at the appropriate molecular masses for any of the Tap proteins when compared to the vector control, indicating the need for further optimisation.

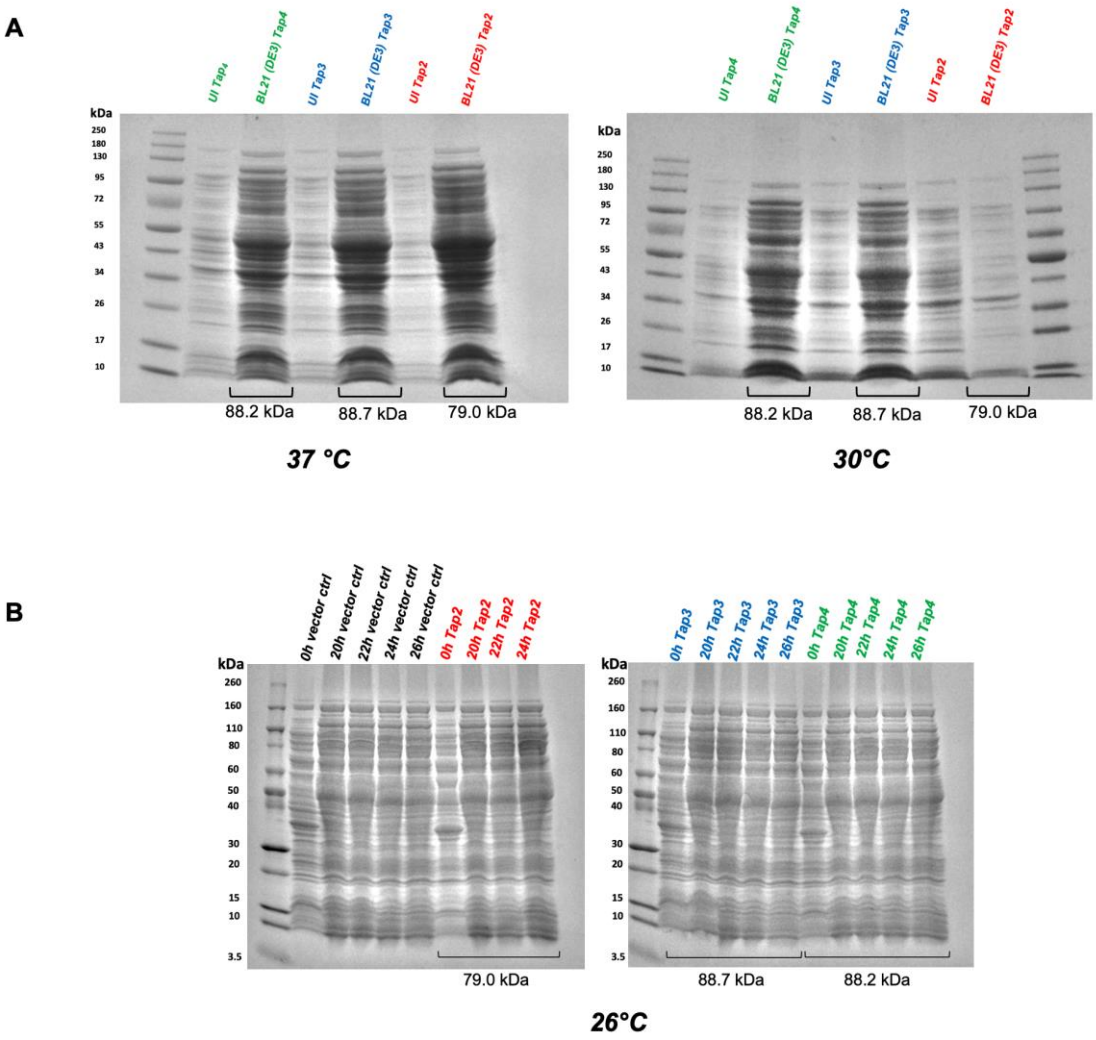


Figure 5.13. Tap protein shows no overexpression using autoinduction media at various temperatures. A) Tap4, Tap3 and Tap2 overexpression trial at 37 °C and 30 °C using cultures grown in LB as a uninduced (UI) control and run on a Tris-Glycine 4-20% SDS-PAGE gel. Expected band sizes for Tap4, Tap3 and Tap2 proteins are highlighted at 88.2 kDa, 88.7 kDa and 79.0 kDa respectively with no clear, dark band at those molecular masses at either temperature compared to the control. **B)** Tap4, Tap3 and Tap2 overexpression trial at 26°C using the empty pET-15b+/*E. coli* BL21 (DE3) transformant cells as empty vector control. Here, samples were taken over a period of 20-26 hours with 0 hours acting as an UI control. No clear, darker bands can be seen at the appropriate molecular masses for any of the Tap proteins, especially compared to the vector control. Protein soluble lysate was loaded at 20 µg total protein per well.

5.5.2 Tap4, Tap3 and Tap2 protein expression through IPTG induction and temperature reduction.

The use of autoinduction media as a mode of induction for Tap proteins produced unsatisfactory results, indicating the need to further optimise the overexpression process. The mode of induction was therefore explored, with the *lac* analogue IPTG used to substitute allolactose. IPTG induction (like lactose) induced protein expression by binding to the lac repressor encoded by *lacI*, derepressing the T7 RNA polymerase promoter upstream of the recombinant protein coding sequences of pET vectors, allowing for T7 polymerase to bind to the promoter and for protein expression to occur. To determine the optimal concentration of IPTG, 0.2 mM, 0.5 mM, and 1.0 mM of IPTG were initially used to induce *E. coli* BL21 (DE3) cells containing the constructed Tap4, Tap3 and Tap2 overexpression vectors at 37 °C. This presented no difference in the protein lysate bands between the three concentrations of IPTG when they were run on an SDS-PAGE gel and visualised (**Figure 5.14 A**).

Figure 5.14 A shows that there was however a band at approximately 110 kDa present for all Tap proteins and absent in the uninduced control protein lysate sample. Differences in molecular mass of proteins compared to their actual mass are common and could be due to protein degradation. Additionally, larger band appearances could be due to disulphide bond formation, between two cysteine residues. Disulphide bond formation tends to be common in proteins with a high number of cysteine residues, and the dense packaging of proteins with disulphide bridges allows faster migration through SDS gels than in their reduced form (de Marco, 2009). To check the possibility that disulphide bond formation was responsible for unexpected migration of our Tap proteins we added a reducing agent, i.e., β-mercaptoethanol, when denaturing the protein sample before SDS-PAGE loading. To check whether protein size could be affected by the method of lysate preparation and/or loading onto an SDS-PAGE gel, two types of sample loading buffer were tested for the IPTG induced samples at 37°C (**Figure 5.14 A**). Protein lysate was loaded at 20 µg total protein per well into a SDS-PAGE gel, once without a reducing agent (using Novex Tris-Glycine SDS sample loading buffer) and once with a reducing agent (using sample loading buffer supplemented with β-mercaptoethanol). No difference was seen between cell lysates with and without reducing agent, and no clear, dark bands can be seen at the expected Tap molecular masses. Further analysis showed

that Tap is not a cysteine-rich protein (ExPASy), supporting the similar banding observed between the SDS-PAGE gel with and without reducing agent.

It was still unclear if the Tap proteins were expressed under any of the conditions tested. We hypothesised that the lack of visible Tap expression may be due to the high predicted instability and solubility indexes of the protein, shown in section 5.2. In order to test this, we tested incubation of our cultures at lower temperatures, which has been shown in the literature to improve expression levels for such proteins (Idicula-Thomas and Balaji, 2005; Francis and Page, 2010). Lower incubation temperatures were used in order to facilitate the expression of correctly folded protein, especially if the protein has a high instability and solubility index (Idicula-Thomas and Balaji, 2005; Francis and Page, 2010). IPTG was thus added to *E. coli* BL21(DE3) cultures that had reached an OD_{600} of around 0.9 to induce production of Tap4, Tap3 and Tap2, and cultures were returned to the incubator. **Figure 5.14 B** depicts the protein expression after addition of 0.1 mM, 0.2 mM and 0.3 mM of IPTG at 18 °C overnight. Protein lysate was loaded at 20 µg total protein per well and pleasingly, bands at the expected molecular masses (underlined at band) of the Tap proteins were visible on the SDS-PAGE gel, when compared to the vector control or to previous experiments with higher post-induction temperatures. Despite there being a stronger band in the Tap lysate samples, the vector control also presented a range of fainter bands, and the SDS-PAGE gel alone was not enough to determine whether the protein was overexpressed. No difference was seen between the UI controls and induced samples, suggesting that 0.1mM IPTG may be sufficient to induce Tap expression, or that the expression conditions still had to be optimised.

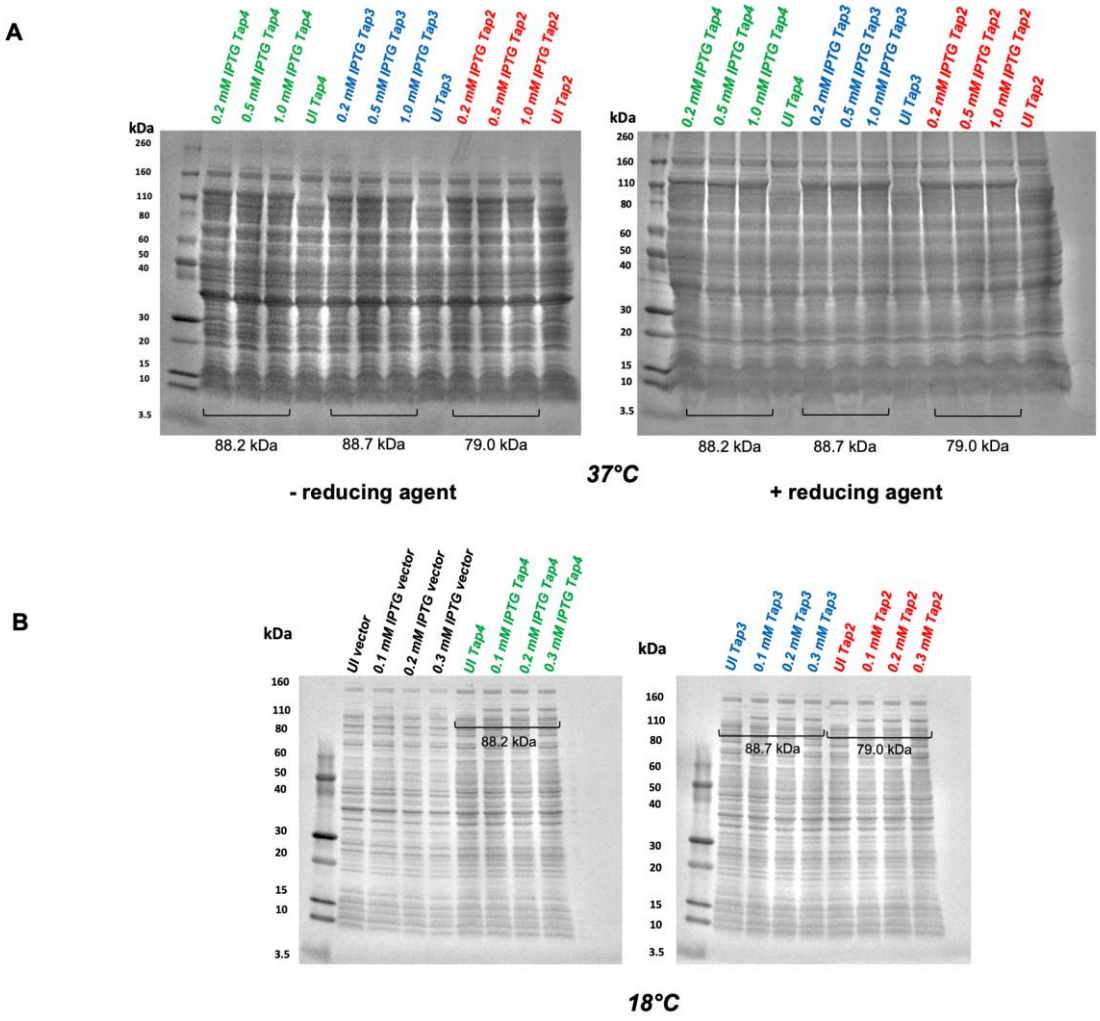


Figure 5.14. Tap protein overexpression using IPTG as the mode of induction. A) 0.2 mM, 0.5 mM and 1.0 mM of IPTG were used to induce *E. coli* BL21 (DE3) cells containing the constructed Tap4, Tap3 and Tap2 overexpression vectors at 37 °C. Protein soluble lysate was loaded at 20 µg total protein per well into a Tris-glycine 4-20% SDS-PAGE gel, once without a reducing agent (using Novex Tris-Glycine SDS sample loading buffer) and once with a reducing agent (using sample loading buffer supplied with β-mercaptoethanol). No difference was observed between cell lysates with and without reducing agent, and no clear, dark bands were observed at the expected (underlined) Tap molecular masses. **B)** 0.1 mM, 0.2 mM, and 0.3 mM of IPTG was used to induce *E. coli* BL21 (DE3) cells containing the constructed Tap4, Tap3 and Tap2 overexpression vectors at 18 °C. Protein soluble lysate was loaded at 20 µg total protein per well with visible bands at the expected molecular masses (underlined at band) of the Tap proteins compared to the vector control. No difference was seen between the UI controls and induced samples.

The unclear results in **Figure 5.14**, prompted a trial immobilised metal affinity chromatography (IMAC) purification of Tap4, to determine whether the protein was being overexpressed in the soluble part of the cell lysate. **Figure 5.15** shows the SDS-PAGE gel of the uncleared lysate, washed lysate, and purified fractions using the 18 °C, 0.2mM IPTG harvested soluble protein lysate (**Figure 5.14 B**), as IPTG induction did not present any significant overexpression effects. The soluble protein lysate was loaded onto a HisTrap FF Crude histidine-tagged protein purification nickel column and eluted by hand at a flow rate of 1 min/mL. A band can be seen at 88.2kDa, the expected size of the Tap4 protein, as indicated by **Figure 5.15**, however the band was most clear in the uncleared lysate and washed lysate suggesting inefficient His-Tag binding to the nickel column. Low protein recovery after elution of Tap4 underlined that the His-Tag was not binding to the metal ions of the column, suggesting that Tap4 was being washed away before elution with a high concentration of imidazole, although varying concentrations of imidazole were not tested and thus the problem with Tap4 elution may have also been a result of sub-optimal imidazole concentrations.

Despite Tap4 being tagged with a His-tag at the N-terminus, many bands were visible on the SDS-PAGE gel depicted in **Figure 5.15**. Potential reasons behind this could be that nickel binds non-specifically to endogenously expressed proteins of *E.coli* BL21(DE3), for example chaperones or proteases (Guisbert *et al.*, 2004). One method to overcome this would be to add more imidazole to the wash buffer, to control the amount of off-target binding, although given the faint bands at around 88.2 kDa this would not have resulted in a high yield of purified Tap4. Overall, the results from the nickel column purification did not present clear purification of Tap4 and therefore demonstrated the need to optimise expression conditions. The nickel column purification also provided further proof, along with the results from **Figure 5.14 B**, that Tap4 was not being overexpressed in the soluble cell lysate, presenting the need to solubilise the insoluble fraction of the cell lysate i.e., the cell pellet, and isolate inclusion bodies.

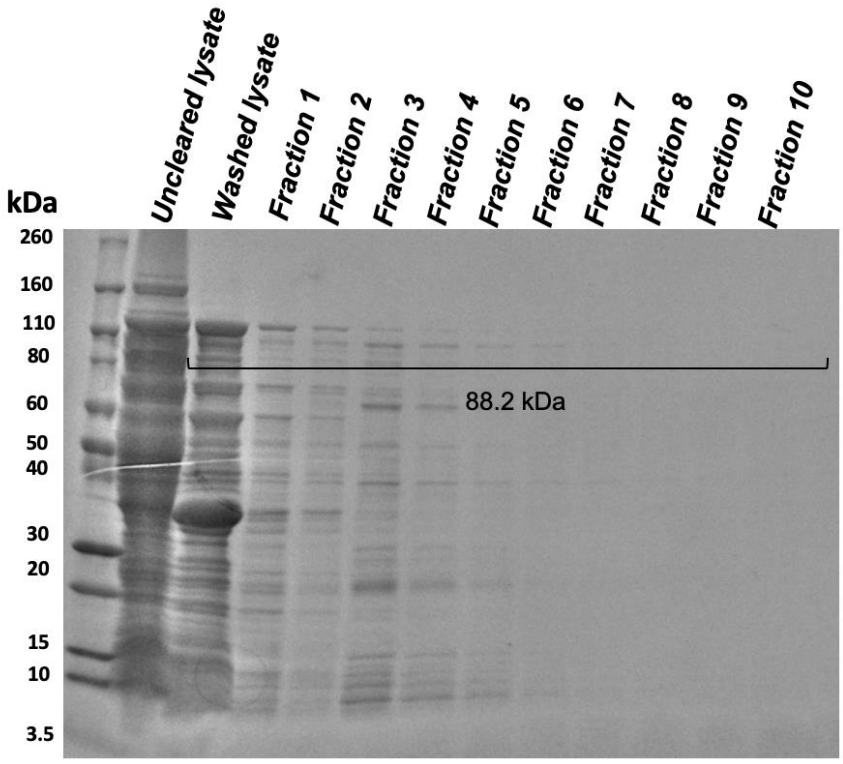


Figure 5. 15. IMAC nickel column purification of Tap4 soluble cell lysate. Tris-Glycine SDS-PAGE gel shows the uncleared lysate (soluble cell lysate), washed lysate (wash buffer) and purified fractions using the 18 °C, 0.2mM IPTG harvested soluble protein lysate. Protein was bound with 5mM imidazole, then eluted using 500mM imidazole in the elution buffer and collected in 10 fractions, labelled on the gel. A faint band can be seen at around 88.2 kDa for Tap4. Unspecific binding of endogenous proteins to the nickel column resulted in numerous, thicker bands present in all elution fractions.

5.5.3 Tap4, Tap3 and Tap2 protein overexpression results in Tap inclusion body formation.

Since the use of autoinduction media (Section 5.5.1) and IPTG (Section 5.5.2) induction at various temperatures and conditions did not clarify whether Tap4 was being expressed in the soluble cell lysate (supernatant) after chemical lysis, the insoluble fraction (pellet) was chemically lysed to isolate and solubilise any potential inclusion bodies. Inclusion body formation can present challenges in protein recovery, especially in recombinant *E. coli* protein expression, where high levels of expression can lead to protein inclusion body aggregation (Singh *et al.*, 2015). Inclusion body formation in bacteria can be due to a multitude of reasons, one of which is the lack of protein sorting compartments in bacterial cells compared to eukaryotic cells. Strong promoters and induction conditions tend to overload the protein control mechanisms, such as the heat shock response, which is regulated by chaperones and proteases assisting in protein folding and removal of misfolded proteins (Guisbert *et al.*, 2004). Overloading these control mechanisms in *E. coli*, can lead to inclusion body formation, which is a protective mechanism. There are two types of inclusion bodies: classical and non-classical, the latter of which exists in a biologically active state and therefore does not require protein refolding (Singh *et al.*, 2015).

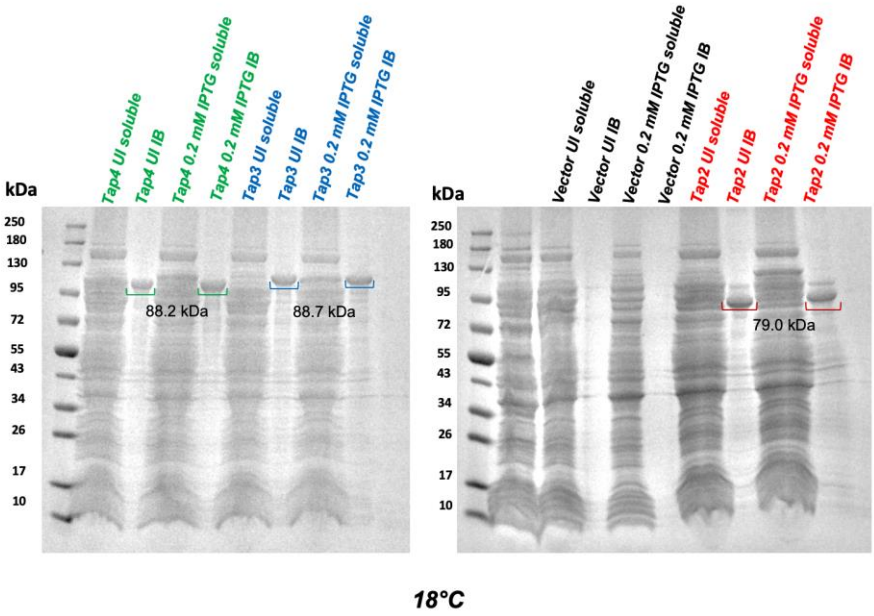
We predicted that inclusion body aggregation was the reason Tap4, Tap3 and Tap2 were not visible when soluble crude extracts were run on SDS-PAGE gels in **Figure 5.13** and **Figure 5.14**. To solubilise inclusion bodies, *E. coli* BL21 (DE3) Tap overexpression cultures were induced with IPTG and grown overnight at 18°C. After centrifugation, the supernatant was discarded and the cell pellet was retained for chemical lysis, as studies have shown that cell disruption by mechanical methods resulted in the accumulation of protein molecules from the soluble fraction (Peternel and Komel, 2010). **Figure 5.16 A** shows the results of Tap4, Tap3 and Tap2 insoluble protein aggregate solubilisation using a non-denaturing solubilisation agent, BugBuster MM (see Chapter 2.0). Bands were present at slightly higher than the predicted 88.2 kDa, 88.7 kDa and 79.0 kDa molecular masses of the Tap proteins and confirmed that these proteins had aggregated in the form of inclusion bodies. This could potentially be due to post translational modification or incomplete denaturation before electrophoresis. The solubilisation of Tap4, Tap3 and Tap2 inclusion bodies resulted in relatively pure proteins, demonstrated by the absence of other endogenously expressed proteins in the SDS-PAGE gel lanes

containing inclusion body cell lysate, and the presence of many endogenously expressed proteins in the soluble vector lanes. The absence of other endogenously expressed proteins on the SDS-PAGE gel after inclusion body washes with BugBuster MM were an indication that Tap proteins were non-classical inclusion bodies, as mild solubilisation using reagents with Tris-HCl and lysozyme has previously lead to the solubilisation of non-classical inclusion bodies, and lead to active recombinant protein extraction, without refolding (Jevševar *et al.*, 2005).

There was only a slight difference between uninduced and induced cultures, which was characteristic of a leaky *lac*UV5 promoter (Du *et al.*, 2021). The fact that there was visible Tap protein expression (after inclusion body solubilisation) when low concentrations and no concentrations of IPTG were added, further underlines the strength of the promoter and the subsequent potential of protein aggregation. Tap protein overexpression was confirmed by the absence of bands at the according molecular masses in the vector control lanes (labelled in black) in **Figure 5.16 A**.

Figure 5.16 B further confirmed that Tap4, Tap3 and Tap2 formed inclusion bodies, as they were detected via chemiluminescent western blot detection using a primary antibody (Mouse-Anti-His) which bound to the His-tag and a secondary antibody (Goat-Anti-Mouse-HRP) which bound to the primary antibody and was conjugated with horseradish peroxidase (HRP). Addition of the luminol substrate allowed for oxidation of the luminol by peroxide, producing 3-aminophthalate, emitting light at 425 nm, equivalent to the bands seen on **Figure 5.16 B** for Tap4, Tap3 and Tap2. The empty pET15b+ vector control showed no band on blot, as it lacked a His-tag and antibodies did not bind.

A



B

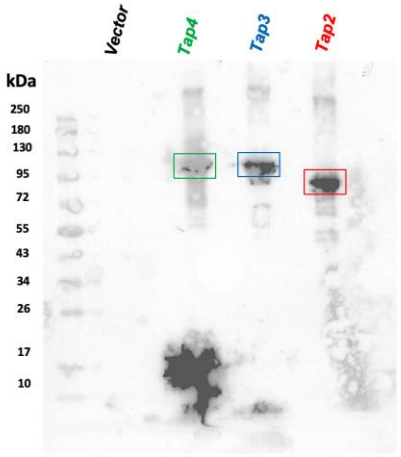


Figure 5.16. Tap 4, Tap3 and Tap2 inclusion body solubilisation and immunoblot detection. A) Protein soluble cell lysate and inclusion body lysate was loaded at 20 µg total protein per well into a Tris-glycine 4-20% SDS-PAGE gel. Bands for Tap4 (88.2 kDa), Tap3 (88.7 kDa) and Tap2 (79.0 kDa) were present for solubilised inclusion bodies and absent for vector control. No difference was observed between uninduced IPTG control and induced samples. **B)** Tap4, Tap3 and Tap2 western blot detection using chemiluminescence, confirming all three proteins were present as inclusion bodies.

5.6 Electrophoretic Mobility Shift Assay (EMSA) reveals Tap4 and chromosomal telomere interaction.

In the previous section, we confirmed the overexpression of Tap proteins through Western Blotting and determined a method of solubilising the formed inclusion bodies during recombinant protein overexpression in *E. coli* BL21 (DE3). In Chapter 3.0, we elucidated relationships between silencing the *tap-tpg* genes of *S. clavuligerus*, showing that this led to the loss of entire replicons, particularly when *tap-tpg*$_4$ was silenced. One of our main aims for the work done in all of the chapters of this study was to build on the work by Algora-Gallardo, (unpublished data), who demonstrated a relationship between the loss of *tap-tpg*$_4$ and chromosomal circularisation, and we hypothesised that this was due to the lack end-patching TPs. In this section, we aimed to test this hypothesis and therefore designed an electrophoretic mobility shift assay (EMSA) to help us understand whether the *S. clavuligerus* TPs do indeed interact with this ssDNA overhangs and cause a 'shift' in the gel. As previous work (Algora-Gallardo, unpublished data) had shown the potential of a Tap4 and chromosome terminal end interaction, we used our solubilised crude lysates of Tap4 (from **Figure 5.16A** cultures) and tested the protein's ability to bind to the 3' ends of *S. clavuligerus* terminal hairpin ssDNA.

To demonstrate the relationship of Tap4 with *S. clavuligerus* chromosome ends, we ordered 200 bp ssDNA oligonucleotides, which retained the hairpin shapes of the organism's archetypal LHS and RHS chromosome telomeres. It was vital for us to test the relationship of Tap4 with ssDNA rather than dsDNA, as an interaction with ssDNA would directly confirm the underlying mechanisms of terminal 'end-patching'. Additionally, we ordered the ssDNA oligonucleotides rather than dsDNA, as dsDNA required denaturing at higher temperatures to separate strands potentially resulting in the loss of ssDNA hairpin structure.

The resulting EMSA in **Figure 5.17** tested the binding activity of the 200 bp chromosome telomere ssDNA after incubation with no or increasing concentrations of Tap4. The ssDNA/protein reactions and controls were incubated for 20 minutes at 30 °C and cooled on ice before running on an agarose gel to determine whether any bound complexes would show slower migration through the gel compared to the free DNA and unbound controls. We included controls which tested all reaction components by themselves, these were polydI-dC on its own, ssDNA with polydI-dC, telomere ssDNA on its own,

and lastly, Tap4 with polydI-dC and also on its own. Results showed bright bands for ssDNA controls (highlighted in red), suggesting that the DNA had migrated through the agarose gel. The wells containing Tap4 showed that the protein failed to enter the gel. This was also seen in the wells containing the Tap4/telL and Tap4/telR reactions in with increasing Tap4 concentrations. Interestingly there was also far less migrated, unbound DNA (highlighted by purple brackets, **Figure 5.17**) at the bottom of the Tap4/telL and Tap4/telR wells compared to ssDNA control wells, indicating that there was some form of retention of the ssDNA, perhaps through Tap4-telomere interaction.

Another observation, critical to our understanding of **Figure 5.17**, was that there was less unbound ssDNA visible on the gel as the Tap4 protein concentration increased, suggesting that increasing the protein concentration allowed for binding of more ssDNA within the EMSA reaction. This interpretation of results was based on what was seen on the agarose gel, and the fact only crude rather than purified protein was used may have also affected the protein migration due to increased protein-protein interactions of total protein lysates, although unlikely. The phenomenon of terminal DNA/TP protein complexes failing to enter the gel, as seen by **Figure 5.17**, has previously been reported in research by (Lin *et al.*, 1993) who showed that the chromosome of *S. lividans* was linear. Using pulsed-field gel electrophoresis, their research showed that linear fragments of *S. lividans* chromosome telomeres failed to enter the gels due to the binding of TP to the S. *lividans* telomere sequences. The addition of proteinase K to the 'free' telomere DNA and protein reactions resulted in better DNA migration through the pulsed field gel, highlighting that the retention of the DNA in the plugs was a result of TPs being bound to telomere DNA, as seen for Tap4 and *S. clavuligerus* chromosome telomeres in **Figure 5.17**.

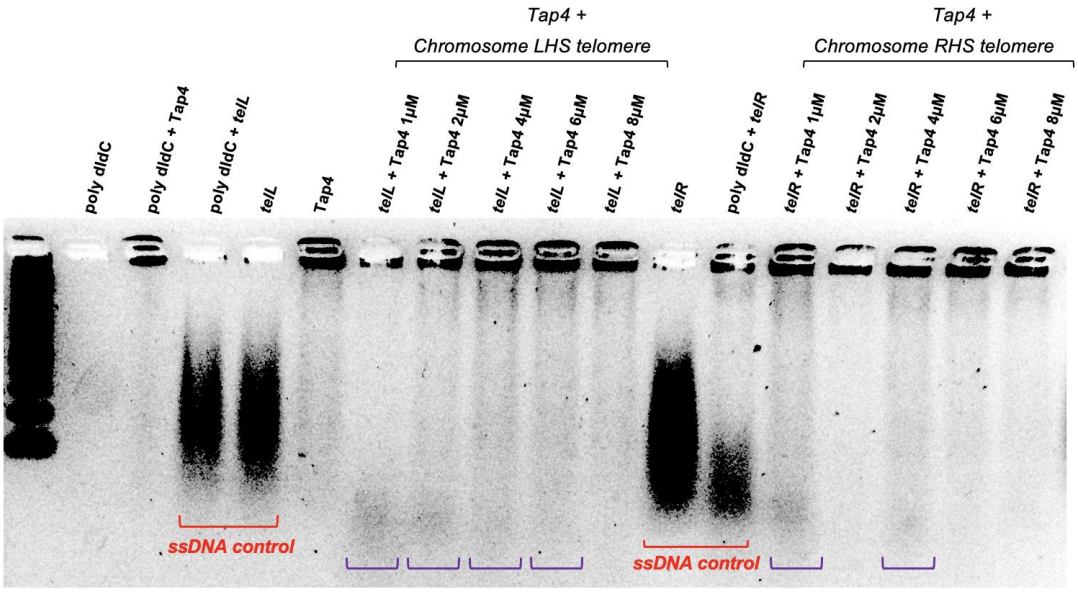


Figure 5. 17. EMSA gel retardation agarose gel demonstrates *S. clavuligerus* DSM 738 Tap4 and chromosome telomere binding. The 200 bp chromosome telomere ssDNA was pre incubated in the absence (telL / telR wells) or presence (telL+Tap4 / telR +Tap4 wells) of Tap4 isolated crude protein at increasing concentrations of 1 µM, 2 µM, 4 µM, 6 µM and 8 µM. polydI-dC was added to prevent unspecific protein binding and was tested with Tap4, telomere ssDNA, and on its own, highlighted in red. Unbound ssDNA is highlighted by the purple brackets. Potential Tap4/chromosome telomere complexes were retained at the top of the wells. After running, 0.8% TAE gel was stained with ethidium bromide and photographed. Where, telR = RHS ssDNA chromosome telomere, telL = LHS ssDNA chromosome telomere, Tap4 = solubilised crude protein lysate.

5.7 Summary

In this chapter we report the role of the terminal proteins of *S. clavuligerus* DSM 738 by looking at their relative locations on the GLPs and in reference to other *Streptomyces* species. Furthermore, we characterised Tap-Tpg4, Tap-Tpg3 and Tap-Tpg2 in *S. clavuligerus* phylogenetically and by looking at their physical, chemical, and structural characteristics. This allowed us to construct pET vector-based plasmids for the overexpression of recombinant Tap4, Tap3 and Tap2 in *E. coli*. Overexpression conditions for the Tap proteins were optimised, leading to the conclusion that Tap proteins formed inclusion bodies. This was overcome by inclusion body solubilisation using chemical lysis, a gentle method which produced Tap proteins with increased purity. Tap protein expression was confirmed through Western Blotting and Tap4 protein interactions with chromosomal telomere ssDNA was tested in an EMSA.

Various *Streptomyces* strains were used to compare the relative location of the *tap-tpg* operon, and in some replicons, just *tpg* (**Figure 5.1**). A noticeable factor when comparing TP operon location was that most operons were located towards the ends of either plasmids (*S. clavuligerus, S. avermitilis, S. rochei, S. lividans, S. hygroscopicus*) or chromosomes (*S. coelicolor, S. avermitilis, S.lividans, S.hygroscopicus*). The terminal inverted repeat ends of organisms with linear replicons i.e., the 'arms' have been shown to be variable (Hoff *et al.*, 2018), contributing to genetic instability within the genus. Genomic instability has also been associated with major rearrangement events, such as deletions in the 'arms' of the replicons, leading to events discussed in Chapter 4.0 such as replicon circularisation (Algora-Gallardo, unpublished data). From our data (**Figure 5.1** and **Figure 5.2/5.3**), which compares the TP operon and terminal proteins of *Streptomyces* strains with archetypal telomeres, pSCL4 stood out as the only replicon with a centrally located *tap-tpg* operon (**Figure 5.1**). We speculated whether the central location of the operon meant it was a more important operon. We can also relate this back to Chapter 4.0, which highlighted that only *tap-tpg*$_4$ silencing resulted in multiple plasmid loss and chromosomal circularisation. Put together, these results show that there is more to be understood about Tap4 interaction with the terminal ends the *S. clavuligerus* replicons.

Phylogenetically, the TPs of pSCL4 and pSCL3 seem to be the most similar to each other, but least similar to other *Streptomyces* archetypal TPs (**Figure 5.2** and **Figure**

5.3). This was highlighted in an unrooted phylogenetic tree comparing the terminal proteins of *Streptomyces* and *Kitasatospora* strains with closed, whole genome sequences. Horizontal gene transfer imposes variability in the telomeres of Streptomycetota, a concept shown by Hoff *et al.*, (2018). Horizontal transfer in the telomere arms have also been shown to be caused by recombination events along the centre of the chromosome, which lead to higher recombination frequency towards the terminal ends (Hoff *et al.*, 2018). The phylogenetic trees also showed that were similarities between the chromosome and plasmid Tap and Tpg proteins and given that many Streptomycetota plasmids have been acquired horizontally and many replicons are subject to recombination by transposable elements, these similarities are justified.

We investigated the similarities and differences of *S. clavuligerus* DSM 738 TPs with *S. coelicolor* A3(2) TPs, comparing the three GLPs with the equivalent TPs, Tac and Tcp, present on the *S. coelicolor* chromosome. The trends in percentage similarity of aligned amino acid sequences shown in **Figure 5.4** corresponded with the phylogenetic similarities of the TPs. To summarise, Tap2 and Tpg2 presented the highest percentage similarity with the *S. coelicolor* TPs and the lowest similarity with its own *S. clavuligerus* Tap4 and Tap3. Most often, the explanation behind similarities of one GLP to a replicon in another *Streptomyces spp*. are evolutionary events, which have been made easier to track with advances in sequencing and computational analysis. Research by Wu *et al.*, (2006), indicated the homology between encoded proteins on pSCL2 and pSLA2-L, particularly the RepL1 and RepL2 proteins responsible for replication initiation and copy number control (translationally identical proteins were found on pSCL2). These results support the similarities of Tap2 and Tpg2 with other *Streptomyces spp*. in terms of genetic homology and evolutionary events leading to certain plasmids, like pSCL2, conserving genes encoding essential proteins which are similar to other species, like pSLA2-L from *S. lividans*.

The structural characteristics of Tap and Tpg proteins were elucidated by 3D model prediction using Alphafold2 (**Figure 5.6** and **Figure 5.7**), and pairwise comparison of the 3D structure of the *S. clavuligerus* TPs allowed for identification of structural metrics. The AlphaFold2 predicted 3D structure pLDDTs provided per-residue confidence scores, presenting overall high confidence in the predicted structures of Tpg, however mixed confidence in Tap structural prediction was indicated by a score <50. Similar PAE graph

trends were seen for Tap4, Tap3 and Tap2, supporting the pLDDT scores of the 3D structures, as high confidence and low error was demonstrated for the first ~100 residues and last 500 residues. Approximately 200 residues thus showed low confidence and high error scores, which corresponded to values of <50 in the pLDDT structure highlighted in **Figure 5.6**. PAE plots showed high confidence and low error for the predicted 3D structures of Tpg. These confidence scores were of great importance in this chapter and contribute to the overall characterisation of *S. clavuligerus* Tap and Tpg proteins, as crystal structures do not currently exist in the literature for TPs.

TP binding mechanisms have been the focus of recent studies, and have helped to identify deoxynucleotidylation sites, important for determining the location of telomere end-patching (Yang, Tseng and Chen, 2015; Yang *et al.*, 2017). As such, these studies, as well as work form Algora-Gallardo (unpublished data), highlight various domains shown in **Figures 5.4**, **5.7**, **5.8**, **5.9**, and **5.10**. The Tap and Tpg amino acid alignment in Figure 5.4, shows the helix-turn-helix domain, which is vital for Tap and Tpg functionality. Additionally, a HIV reverse transcriptase motif was identified on Tpg (Bao and Cohen, 2001). The *S. clavuligerus* Tap and Tpg beta-strands, alpha helices, and coils were highlighted by **Figures 5.7-5.10**, and supported by Yang *et al.*, (2013) who look at Tpg homolog amino acid sequences encoded by *Streptomyces* chromosomes and linear plasmids.

The 3D structural alignment analyses in Section 5.3 led to the identification of the 3D positions of the N and C termini of Tap and Tpg proteins in *S. clavuligerus* DSM 738. Thus, we constructed Tap overexpression vectors with the goal of eventually testing the interaction of Tap with the 3' terminal ssDNA overhangs in an EMSA. The pET-vector based cloning was supported by research conducted by Bao and Cohen, (2003), Yang *et al.*, (2013), and Yang, Tseng and Chen, (2015), all of whom looked at overexpressing the terminal proteins of *S. lividans.* Whilst the studies by Yang *et al.*, (2013), and Yang, Tseng and Chen, (2015) expressed the Tap and Tpg proteins in *E. coli* BL21(DE3) like in this chapter, the study by Bao and Cohen, (2003) overexpressed the protein in *S. lividans*, avoiding the need to codon optimise the amino acid sequences for the expression host. The methods of induction differed on the basis on the overexpression host: thiostrepton induction was used by Bao and Cohen, (2003) and IPTG was used to induce expression in the other two studies. In this chapter, we tested two different

methods of induction: autoinduction media, and IPTG. Overall, our results highlighted that the method of induction itself made little difference, whilst other factors did, such as lowering the temperature. Additionally, as the majority of studies that overexpressed Tap and Tpg proteins used IPTG, this method was taken forward and optimised at a lower temperature, which is also where the first faint bands at the according molecular masses for Tap4, Tap3 and Tap2 were seen (**Figure 5.14**). Nickel column purification (**Figure 5.15**) was tested, as the proteins were His-tagged, yielding very faint bands at the roughly 80-90 kDa molecular masses. Purification of the protein was not optimised i.e., by changing imidazole binding and elution buffer concentrations, however the absence of a strong band in the SDS-PAGE gel of the protein lysate before (**Figure 5.14**) and after (**Figure 5.15**) purification, presented the need to check for inclusion bodies.

In this chapter we report the formation of inclusion bodies during the overexpression of Tap4, Tap3 and Tap2. Many reasons can lead up to the formation of inclusion bodies when overexpressing recombinant proteins in *E. coli*, and their dynamic nature makes it difficult to identify in what state these proteins reside in the cell. As such, an equilibrium between aggregated inclusion bodies and folded, overexpressed proteins exists, presenting the need for more gentle solubilisation buffers to conserve protein structure. Tap protein inclusion bodies were identified after solubilisation using the lysozyme containing BugBuster MM lysis buffer. Various methods exist for inclusion body solubilisation, such as adding a high concentration of chaotropes and denaturants like guanidine hydrochloride and urea to the lysis buffer, and subsequently re-folding the protein in a column (often urease mediated) (Fischer *et al.*, 1992; Rudolph and Lilie, 1996). Reducing agents have also been added to proteins with a high number of cysteine bonds to prevent incorrect disulphide bond formation. These include beta-mercaptoethanol or dithiothreitol, but interfere with column purification methods, such as nickel column purification used in this chapter, as the reducing agents react with the free nickel bound to the column. Alternatively, mechanical lysis using a French Press or sonication solubilises inclusion bodies, but all these methods tend to be harsh and affect the bioactivity of the recombinant protein (Singh *et al.*, 2015). Generally, both the addition of chaotropes (denaturing solubilisation agents) and mechanical lysis necessitates refolding of the protein but can also lead to aggregation during the process (Singh *et al.*, 2012), as shown in **Figure 5.18**. Due to time constraints, optimisation of protein refolding

and buffer concentrations was not possible, therefore a non-denaturing solubilisation agent was used, containing benzonase nuclease and rLysozyme™.

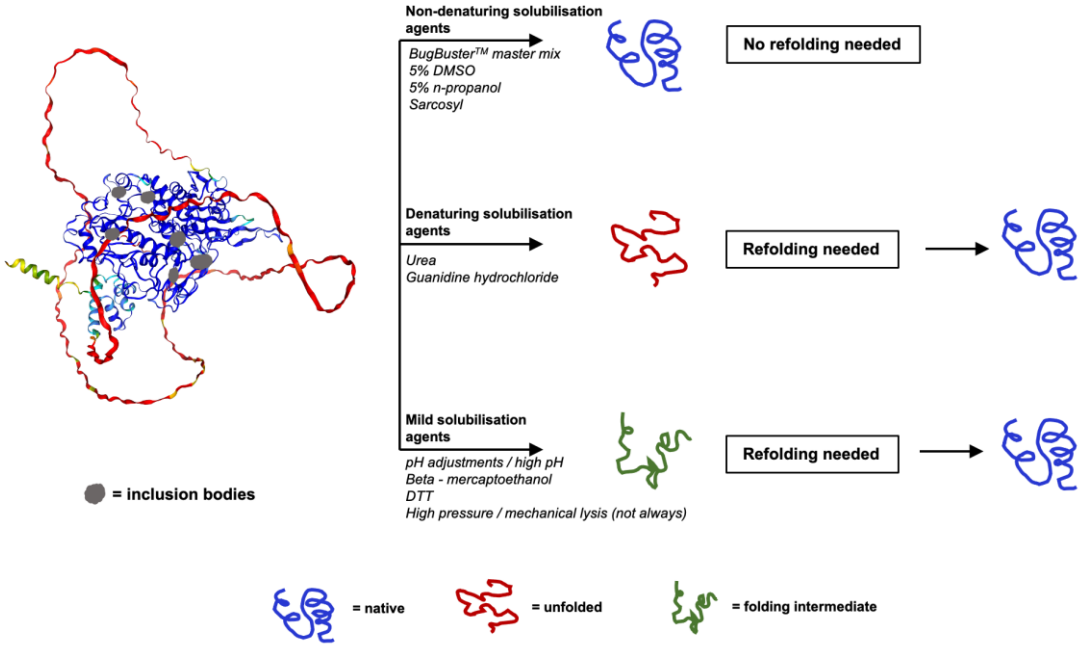


Figure 5. 18. Comparison of different solubilisation methods for recombinant protein recovery from inclusion bodies. Adapted from Singh et al., (2015)

Inclusion body formation during overexpression was unexpected, as previous studies investigating Tap and Tpg overexpression did not encounter them. These studies also used different conditions during overexpression and cell lysis, mentioned earlier (Yang, Tseng and Chen, 2015), which included mechanical lysis (Bao and Cohen, 2003). Additionally, the studies investigated TPs of *S. lividans*, whose TPs would have had different physical and chemical parameters leading to optimal overexpression in different conditions to S*. clavuligerus* TPs. Much remains unknown about inclusion bodies, however recent reviews have outlined a linkage between the physical and chemical parameters of eukaryotic and prokaryotic proteins and their likelihood to form insoluble aggregates, many of which are dominated by beta-strand structure and form inclusion bodies of amyloid nature (de Groot, Sabate and Ventura, 2009). Thus, the inclusion body formation of Tap4, Tap3 and Tap2 can be linked with the parameters such as instability and aliphatic indices shown in **Table 5.1**. The *S. clavuligerus* Tap proteins presented higher instability and a lower aliphatic index compared to the Tpg proteins, suggesting less thermostability and a shorter in vivo half-life compared to Tpg4, Tpg3 and Tpg2 (Idicula-Thomas and Balaji, 2005). We therefore wondered whether there were any mechanisms to compensate the presumably shorter in-vivo half-life of Tap proteins, referred back to the *tap-tpg* operon. When considering the *tap-tpg* operon in *S. clavuligerus* DSM 738, we established that the position of *tap* is more proximal to the promoter region of the operon and *tpg* lies downstream of *tap*, highlighting higher mRNA transcription levels for *tap*, which could be one of the methods in which the organism compensates for the shorter in vivo half-life of the encoded Tap protein.

The solubility of the proteins also plays a role in inclusion body formation, as there is increased solubility with increasing positive or negative charges of the residues (Idicula-Thomas and Balaji, 2005), with greater differences in average positively and negatively charged residues seen for Tpg (**Table 5.1**). The physical and chemical parameters such as limited recombinant protein solubility which led to aggregation, and other factors like high levels of transcription intrinsic to T7 promoters contribute to inclusion body formation. So, combined with the physical and chemical Tap protein parameters, overexpression conditions can be altered to prevent aggregation. These include temperature of culture incubation and IPTG concentration which both improved when reduced for the Tap proteins of *S. clavuligerus* in this chapter. Specifically lowering the temperature for induction and cold-shocking the cultures before incubating them at lower

temperatures, slowed protein synthesis and has previously resulted in a higher percentage of correctly folded proteins within the formed inclusion bodies (Peternel and Komel, 2010). Conditions such as cell lysis buffers, and methods of chemical vs mechanical lysis to test which reasons may be responsible for Tap4, Tap3 and Tap2 inclusion body formation were not optimised. To summarise, in this chapter we presented a non-denaturing method of inclusion body solubilisation of the Tap proteins, producing pure crude cell lysates (**Figure 5.16**) and finally confirmed their presence though immunoblotting.

Lastly, Tap protein expression was confirmed through Western Blotting and Tap4 protein interactions with chromosomal telomere ssDNA were tested in an EMSA. Results showed potential ssDNA/Tap4 interactions as unbound ssDNA migrated through the agarose gel, however migration was retarded when Tap4 was added to the ssDNA and ran on the agarose gel (**Figure 5.17**). More specifically, the Tap4 protein failed to enter the gel, and ssDNA/Tap4 binding was further highlighted as the addition of increasing concentrations of Tap4 resulted in less unbound telomere DNA in the agarose gel. The EMSA results in this chapter were supported by Lin *et al.*, (1993), who demonstrated a similar phenomenon in their study focussing on *S. lividans* telomere and TP binding. Binding of TPs to 'free' *S. lividans* chromosome telomere ends were confirmed by retention of the DNA in the plugs that were not treated with proteinase K and migration in the plugs that were. The degradation of TPs by proteinase K allowed for better migration of the DNA through the gel, also observed in this chapter in **Figure 5.17**. Our EMSA agarose gel results however also demonstrated, that when Tap4 was run on the gel on its own, it was retained in the well and did not migrate through the gel. Potential reasons behind the lack of migration of Tap4 may have been the usage of crude Tap4 lysate instead on purified protein, but also that our crude lysate contained solubilised inclusion bodies which may have existed as a soluble/insoluble mixture in the lysate, affecting migration. Nonetheless, the hypothesis by Lin *et al.*, (1993) was also tested with chromosomal DNA from *S. lividans* (ZX7), *Streptomyces moderatus* ATCC23443, *Streptomyces parvuius* ATCC 12434, *S. rochei* 7434-AN4, *S. antibioticus* IMRU 3720, *S. coelicolor* A3(2) M130, *Streptomyces lipmanii* ATCC 27357, and E. coli DH1. The addition of proteinase K resulted in retention of the DNA in the plugs and indicated the presence of covalently bound TPs, supporting the results in this chapter. To conclude, based on the results of the EMSA in this chapter, we suggest that Tap4 binds to the

ssDNA chromosome telomere overhangs in *S. clavuligerus* DSM 738. However, the inclusion of more controls, such as a specific ssDNA or dsDNA competitor control, would increase our confidence in the conclusions drawn.

6.0 Discussion

In this work, we built upon the ideas of Algora-Gallardo, (unpublished data), who demonstrated that removal of pSCL4 from *S. clavuligerus* DSM 738 resulted in chromosomal circularisation, potentially because of the absence of *tap-tpg*$_4$ and the encoded TPs. *Streptomyces clavuligerus* possesses four GLPs, and TPs are encoded on three out of the four (pSCL4, pSCL3 and pSCL2). Therefore, this work aimed to decipher the relationship between all of the TPs with the replicon ends by optimising a CRISPR-dCas9 multiplexing system to allow for silencing of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$.

In chapter 3.0, we optimised *tap-tpg*$_4$ silencing by looking at the fundamental experimental processes allowing us to implement the CRISPR systems into our target organism. For this, we used CRISPR-dCas9 and CRISPR-cBEST with *tap*$_4$ specific spacers to allow for transcriptional silencing (dCas9) or translational blockage (cBEST) through the introduction of a stop codon. Conjugation optimisation experiments were essential as conjugations with *S. clavuligerus* DSM 738 proved difficult and had not been optimised for the main vectors used in this study. pSET-152 and pGWS1370 vectors were therefore transformed into *E. coli* and conjugated with *S. clavuligerus*. By increasing the number of donor *Streptomyces* spores compared to recipient *E. coli* cells, we concluded that the number of transconjugant colonies increased, but the conjugation efficiency decreased. We compared adding fresh vs frozen recipient spores to the donor cells and determined that the use of fresh spores increased the number of resulting transconjugant colonies and made conjugations between *S. clavuligerus* DSM 738 and previously unsuccessful integrating vectors work. These results were used for CRISPR-BEST conjugations in chapter 3.0, as our goal was to increase the number of transconjugant colonies, allowing us to isolate mutant strains.

Both CRISPR-dCas9 and CRISPR-cBEST were used to determine the significance of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ in *S. clavuligerus* DSM 738, one of which was an integrative vector (CRISPR-dCas9) and one of which was based on the temperature sensitive pSG5 replicon, commonly used in other CRISPR systems (chapter 3.0). Whilst conjugation optimisation experiments aided in the conjugation of CRISPR-dCas9 based plasmids with *S. clavuligerus* DSM 738, the optimisation experiments did not aid in

CRISPR-cBEST based conjugations with our target organism. We believed the CRISPR-cBEST system did not function in *S. clavuligerus,* as pSG5 replicons had previously shown instability and the high number of off-target effects led to unsuccessful conjugations (GSK-personal communication). Our main aim in applying both CRISPR systems to our target organism were to allow for *tap-tpg* silencing, rather than using CRISPR-Cas9, which caused DSBs, associated with high levels of rearrangements (Hoff *et al.*, 2018).

Therefore, the significance of *tap-tpg*$_4$ was explored by CRISPR-dCas9 gene silencing, presenting mixed results. Previous work highlighted that the absence of *tap-tpg*$_4$ resulted in chromosome circularisation (Algora-Gallardo, unpublished data), a concept which we believed applied to mutants in this study, such as SCLAV-pEM3. Chapter 3.0 results highlighted that this was not the case for SCLAV-pEM3, as PCR was used to amplify the linear ends of the knockdown strain, underlining chromosome linearity, confirmed with a Southern Blot. Given that the absence of *tap-tpg* from the chromosome was previously compensated for by Tap-Tpg4 encoded in trans (Algora-Gallardo, unpublished data), we continued to explore whether chromosome linearity was a result of Tap-Tpg3 and/or Tap-Tpg2 acting in trans on the chromosome in addition to what we knew about Tap-Tpg4.

In chapter 4.0 we multiplexed CRISPR-dCas9 to target all combinations of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ in order to determine the significance of the TPs and to determine whether silencing them will have any effects on replicon linearity in cis or in trans. These dCas9 knockdowns created EM3, EM4, EM5, EM7, EM8, EM9 and EM10 which were sent off for Illumina sequencing. Illumina sequencing highlighted that multiplexing and silencing multiple *tap* genes resulted in plasmid loss (**Table 4.3** and **Table 4.5**). The loss of multiple plasmids resulted in strains with different morphologies, however plasmid loss showed no relationship with strain morphology i.e., strains which had lost multiple plasmids were not necessarily of balder, or of a non-sporulating phenotype. A few trends were however observed: strains with integrated *tap-tpg*$_4$ targeting CRISPR-dCas9 vectors (EM3, EM7, EM8, EM10) showed loss of multiple GLPs. This was not the case when *tap-tpg*$_3$ and/or *tap-tpg*$_2$ were targeted. The importance of the encoded Tap-Tpg4 proteins have previously been highlighted by Algora-Gallardo, (unpublished work), who showed that when *tap-tpg*$_4$ was lost and the chromosome was complemented with a copy of *tap-tpg*$_4$, the resulting strains still presented a circularised chromosome and loss

of chromosome telomeres i.e., they were unchanged from the pSCL4-free WT. *tap-tpg*$_4$ complementation on the *S. clavuligerus* DSM 738 chromosome therefore did not have any effects on chromosome linearity, highlighting the significance of Tap-Tpg4 encoded by pSCL4 in end patching of the chromosome. Additionally, mutants in this study had all lost pSCL2, which was a direct consequence of using CRISPR-dCas9 and reasons behind this remain unknown, although we speculated that a bad PAM seed sequence may have been a reason behind this.

The initial screening of mutants was reduced, and final strains were analysed genotypically. These were: EM3 (lost pSCL3, pSCL2 and pSCL1), EM5 (lost pSCL2), EM7 (lost pSCL2 and pSCL1), and EM10 (lost pSCL3 and pSCL2). The Illumina sequences of these strains were mapped to the *S. clavuligerus* DSM 738 WT genome, showing that the loss of plasmids and the integrated CRISPR-dCas9 had strong off-target effects, but also that there were a lot of genome rearrangements led by transposable elements. Research by Gomez-Escribano *et al.*, (2021) showed that pSCL3 was lost in one of their *S. clavuligerus* strains through basic culturing methods. This was unintentional and highlights two things about our study which are, that huge genome rearrangements can happen randomly and that perhaps stressing the strain through harsh culture conditions or a gene silencing via CRISPR systems can increase the chances of these rearrangements occurring, as shown by genotypic analysis our EM mutants in chapter 4.0.

The EM mutants in this study showed missing numbers of base pairs from their replicon ends as highlighted by **Table 4.4**. Chromosome circularisation was therefore also investigated in chapter 4.0 based on the fact that work by Algora-Gallardo, (unpublished data) witnessed loss of terminal ends as a sign of chromosome circularisation (**Figure 4.14** and **Figure 4.15**). Analysis and mapping of Illumina reads on IGV illustrated a split read across the terminal repeat ends of EM10, highlighting that this strain had a circular chromosome. The Illumina sequences also showed that no other strains presented signs of circularisation for any of the replicons, once again highlighting a previously mentioned relationship between Tap-Tpg4 and the chromosome, i.e., that our *tap-tpg*$_4$ targeting CRISPR-dCas9 based strain was the only strain in the study to show chromosome circularisation. We believed this was a direct impact of *tap-tpg*$_4$ silencing, however results

in chapter 4.0 did not clarify this, as other *tap-tpg* genes were silenced in EM10 through the multiplexing system.

We made a critical observation in chapter 4.0 in terms of telomere topology of EM10, which carried a circular chromosome. We demonstrated that the predicted telomere shapes of the *S. clavuligerus* chromosome contained hairpin loops also referred to as palindromes. When we linked the number of base pairs lost from telomere ends and investigated the hairpin structure of the *S. clavuligerus* chromosome telomeres, we found that the 13 bp lost from the LHS telomere, as a result of EM10 chromosome circularisation, matched the 13bp loop of palindrome I. This made us wonder whether it was the 13bp lost from palindrome I that resulted in chromosome circularisation, as no other EM mutants showed the same number of bp lost from their LHS or RHS telomeres. Our theory was prompted by Yang, Tseng and Chen, (2015), who demonstrated that Tap was required for the deoxynucleotidylation of Tpg and that telomere end patching is initiated by TP priming of the first 13 nucleotides. Therefore, we speculated that the loss of the 13 bp from EM10 telomeres meant that end-patching could not occur. Since EM10 was the only strain with a silenced *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ it suggests that the TPs encoded by pSCL3 and pSCL2 prime the terminal ends in trans. We suggest that the silenced gene expression of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ therefore lead to a decrease of terminal protein bioavailability or no bioavailability at all, and terminal end-patching could therefore not be initiated.

In chapter 4.0, the EM strains were also physiologically tested, and growth curves were constructed in addition to clavulanic acid production measurements. The physiological analyses of the strains showed that EM mutants with varying lost plasmids, EM7 (pSCL2, pSCL1-) and EM10 (pSCL3-, pSCL2), showed a significantly higher specific growth rate compared to the WT, highlighting that the loss of plasmids resulted in better growth, perhaps due to a decreased metabolic burden. Unfortunately, these strains showed that the growth was not an indication of better clavulanic acid production, as minimal titres of clavulanic acid were produced (max 0.0003 µg/mL), especially compared to industrial titres of around 200 µg/mL.

In chapter 5.0 we investigated the role of the *S. clavuligerus* DSM 738 terminal proteins in end patching through recombinant protein overexpression of Tap with the aim of testing their binding specificity to the 3' terminal ssDNA overhangs of *S. clavuligerus* telomeres. Initially, we noticed differences in the *tap-tpg* operon location for *S. clavuligerus*, where most operons were encoded towards the ends of the replicons (pSCL2 and pSCL3 *tap-tpg*), and only the *tap-tpg* pair on pSCL4 was encoded towards the centre of the replicon. We speculated whether the central location of the operon meant it was a more important operon, as it was located in a more conserved region of the replicon, as previously (chapter 4.0) only *tap-tpg*$_4$ silenced strains showed multiple plasmid loss and circularisation.

Both Tap and Tpg 3D structures were predicted through Alphafold2 and pLDDT scores, as well as physical and chemical parameters, were investigated for the TPs of *S. clavuligerus*, as this had not been done before (**Figure 5.5** and **Figure 5.6**). Higher confidence scores were seen for Tpg structural predictions vs Tap, Tpg was also shown to be more thermostable, also showing a lower instability index, highlighting the overall stability of the protein. 3D structures were used to identify the N and C termini for subsequent pET vector cloning to allow recombinant protein overexpression. In order to eventually test protein-DNA interaction in an EMSA, we chose to overexpress Tap, as Tap is the protein which binds to the ssDNA 3' overhangs of the *S. clavuligerus* telomeres, acting to recruit Tpg. 3D structures presented an exposed N-terminus for most Taps, therefore we decided to add a histidine tag to the N-terminus of the Tap4, Tap3 and Tap2 proteins, which allowed us to detect whether protein had been overexpressed through immunoblotting. Before this could be done, overexpression of the codon-optimised Tap proteins in *E. coli* BL21 (DE3) was optimised. Optimisation of Tap recombinant protein overexpression concluded that different induction methods (IPTG was chosen over autoinduction media) as well as lowering the recombinant *E. coli* culture temperatures from 37°C to 18°C proved to be the most effective (**Figure 5.16**).

Despite optimising culture conditions for the overexpression of Tap4, Tap3 and Tap2, SDS-PAGE gels of crude protein lysate only showed faint bands at 88.2 kDa, 88.7 kDa and 79.0 kDa, highlighting that the proteins were not overexpressed after all. As our lysis method and crude extract was a fraction of the soluble protein cell lysate, we tested the insoluble part of the lysate (cell pellet) for inclusion bodies (**Figure 5.16**). To do so, the

inclusion bodies were solubilised by gently washing the spun-down cell pellet with BugBuster MM. Inclusion body formation during overexpression was unexpected, as previous studies investigating Tap and Tpg overexpression did not encounter them, however these studies also used differing lysis methods, such as mechanical lysis (Bao and Cohen, 2003; Yang, Tseng and Chen, 2015), suggesting that BugBuster MM was not the most effective lysis method for breaking the cells containing Tap recombinant protein. We solubilised the inclusion bodies (BugBuster MM lysis) using a non-denaturing method as opposed to a denaturing method (chemical and mechanical lysis), as a non-denaturing method did not require additional refolding of the protein. Protein inclusion bodies are formed through the cell's protective mechanism and are thought to be influenced by various factors, touched upon in chapter 5.0 (section 5.7), such as the protein's instability and aliphatic indices and solubility. Additionally high expression levels of strong promoters, like the inducible T7 promoter used in this study, have shown to overwhelm the *E. coli* heat shock response, i.e., endogenous chaperones and proteases, which were all factors we believe contributed to the formation of Tap4, Tap3 and Tap2 inclusion bodies.

Lastly, we tested the binding capacity of Tap4 with the telomere ends of the *S. clavuligerus* DSM 738 chromosome. The previously demonstrated relationship between the loss of pSCL4 and lack of chromosomal end-patching shown by Algora-Gallardo, (unpublished data) made us wonder whether it was the loss of pSCL4 as a whole that resulted in chromosome circularisation or whether it was *tap-tpg4*. We knew from chapter 4.0, that silencing *tap-tpg*$_4$ resulted in multiple plasmid loss, but it was difficult to determine the extent of this due to the off-target effects demonstrated by the CRISPR-dCas9 system and general *S. clavuligerus* genome plasticity. Nonetheless, we chose to investigate the relationship further, building on work in this study as well as by Algora-Gallardo, (unpublished data), and demonstrated that there was binding of Tap4 with the 3' ssDNA overhangs of the chromosome in an EMSA (**Figure 5.17**). The results of the EMSA agarose gel did not show a shift in the DNA-protein complexes formed, in fact they showed the opposite: the Tap4-ssDNA complexes failed to enter the gel and were stuck in the wells (**Figure 5.17**). This was contrasted by the unbound ssDNA controls, which migrated through the agarose gel, highlighting that it was the potential interaction of the TPs with the ssDNA which led to ssDNA-protein complex retention in the wells, a concept which was shown by Lin *et al.*, (1993). The experimental procedure and result

of the EMSA presented various limitations, one of which was that solubilised Tap4 inclusion bodies were used in their crude form and purified protein might have demonstrated a different relationship to the ssDNA overhangs. Additionally, an ssDNA competitor sequence was not tested, therefore the specificity of Tap4 with the *S. clavuligerus* ssDNA telomere overhangs was not certain. Optimisation of ssDNA and testing a wider range of protein concentrations was necessary and may have affected the clarity of the gel leading to more conclusive results. Limitations existed, however the evidence supporting our EMSA results was demonstrated in the literature (Lin *et al.*, 1993).

7.0 Conclusions and future work

7.1 Conclusions

To conclude, with this project, we aimed to investigate the significance of *tap-tpg*$_4$ and its relationship with replicon linearity in *S. clavuligerus* DSM 738. This was difficult to determine, as the genes encoding TPs are found on three out of four *S. clavuligerus* GLPs and initial results highlighted the potential of TPs acting in trans on other replicons, as silencing *tap-tpg*$_4$ did not cause chromosome circularisation or loss of telomeres. We proceeded to test this concept by multiplexing our CRISPR-dCas9 silencing method to determine how *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ knockdowns would affect the terminal ends of the five *S. clavuligerus* replicons. Our conjugation optimisation results from chapter 3.0 laid the framework for CRISPR-dCas9 multiplexing in chapter 4.0, where we Illumina sequenced mutants with knocked down *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$, creating strains with various combinations of lost plasmids and terminal ends. These strains demonstrated genotypic differences, some of which were linked to their phenotypes. We demonstrated the loss of the 13-nucleotide binding site of Tap for our *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$ silenced mutant EM10, and showed that the EM10 chromosome circularised, suggesting that a potential overall reduction in *tap-tpg* gene expression in *S. clavuligerus* affects end-patching. Put together with the pSCL4-chromosome relationship demonstrated through CRISPR-Cas9 curing of the plasmid by Algora-Gallardo, (unpublished data), we support the hypothesis that Tap-Tpg4 partakes in end-patching. This concept was highlighted in chapter 5.0 of this study, showing chromosome terminal hairpin and Tap4 binding activity, although more research needs to be done to confirm the specificity of binding.

One of our final aims was to understand the interaction of the terminal proteins with replicon ends and to determine their essentiality and biotechnological implications. Due to the fact that some of our EM mutants lost pSCL3, pSCL2 and pSCL1, which significantly increased their specific growth rate and affected strain morphologies, we confirm that these replicons are essential for strain fitness, but that *S. clavuligerus* does not require them for clavulanic acid production, as strains that had lost more replicons did not produce more clavulanic acid. Thus, curing pSCL3, pSCL2 and pSCL1 agreed with the hypothesis highlighted in our aims, which suggested a decrease in *S. clavuligerus* metabolic burden as a result of a decrease in replicons. It was important to

note that none of our strains lost pSCL4, however the essentiality of the plasmid had previously been highlighted by Algora-Gallardo, (unpublished work), who showed that despite being dispensable, *S. clavuligerus* pSCL4-free strains exhibited primary and secondary deterioration making pSCL4 essential. Deterioration was perhaps due to loss of essential elements on the plasmid or the effect of plasmid deletion, mirrored for the EM mutants in this study. Therefore, in the biotechnological context the loss of replicons is of importance in terms of strain fitness rather than clavulanic acid production in *S. clavuligerus* DSM 738.

7.2 Future Work

Future experiments would focus on confirming the knockdown effects of CRISPR-dCas9. This would be done through RT-qPCR, RNA-seq and followed up with transcriptomic analysis in order to see how silencing affects gene expression of *tap-tpg*$_4$, *tap-tpg*$_3$ and *tap-tpg*$_2$, and to see whether gene knockdown affects the *S. clavuligerus* transcriptome. This would allow us to show how lack of *tap-tpg* expression could have led to chromosomal circularisation in our EM mutant. Additionally, the expression of genes involved in the production of clavulanic acid could have been analysed for the mutants, which would be vital for GSK and would allow them to improve their industrial *S. clavuligerus* strains.

Another experiment which would have been interesting to follow through with would have been to focus on adjusting CRISPR-dCas9 and CRISPR-BEST Cas9 expression levels by replacing the corresponding constitutive promoters with inducible promoters and/or introducing riboswitches (Rudolph, Vockenhuber and Suess, 2013). CRISPR-cBEST contained a pTipA promoter to control Cas9 expression, therefore the toxicity of the replicon could have been reduced by altering thiostrepton concentrations. The off-target effects of dCas9 have been shown over the last year (Rostain *et al.*, 2023), highlighting that the reduction of expression of dCas9 and/or the sgRNA, would allow for reduced off-target effects, as both of these components had strong, constitutive promoters.

In this study, we characterised strains based on morphology however, these morphological observations would have held a lot more value if they were confirmed through microscopy. One example of this would be to Schwedock stain the mycelia (Schwedock *et al.*, 1997) and use fluorescence microscopy in to look at *S. clavuligerus*

strain fitness by determining mycelial branch lengths, hyphae widths and inter-branch distances, as demonstrated by Algora-Gallardo, (unpublished data).

Lastly, we confirmed the loss of pSCL3, pSCL2 and pSCL1 from *S. clavuligerus* in this study and developed strains with various combinations of lost plasmids. In this study one of our mutants, EM7, was a pSCL2- and pSCL3- strain. We also sequenced a pSCL4- and pSCL3- strain, M5, which was created by Algora-Gallardo, (unpublished data). In order to determine plasmid essentiality one idea for future work could be to conjugate our multiplexed CRISPR-dCas9 plasmids, which we know cause the loss of pSCL3, pSCL2 and/or pSCL1, with the M5 strain to create a fully plasmid-free strain. This would allow us to determine the biotechnological implications by assessing clavulanic acid production and overall strain fitness of plasmid-free strains.

8.0 References

AbuSara, N.F. *et al.* (2019) 'Comparative Genomics and Metabolomics Analyses of Clavulanic Acid-Producing Streptomyces Species Provides Insight Into Specialized Metabolism', *Frontiers in Microbiology*, 10. Available at: https://doi.org/10.3389/fmicb.2019.02550.

Aggarwal, R., Chaudhary, U. and Bala, K. (2008) 'Detection of extended-spectrum β-lactamase in Pseudomonas aeruginosa', *Indian Journal of Pathology and Microbiology*, 51(2), p. 222. Available at: https://doi.org/10.4103/0377-4929.41693.

Alberti, F. and Corre, C. (2019) 'Editing streptomycete genomes in the CRISPR/Cas9 age', *Natural Product Reports*, 36(9), pp. 1237–1248. Available at: https://doi.org/10.1039/C8NP00081F.

Alexander, D.C. and Jensen, S.E. (1998) 'Investigation of the Streptomyces clavuligerus Cephamycin C Gene Cluster and Its Regulation by the CcaR Protein', *Journal of Bacteriology*, 180(16), pp. 4068–4079.

Alexander Steinbüchel (no date) *Microbial Linear Plasmids*. Springer Verlag.

Algora-Gallardo, L. *et al.* (2021a) *Bilateral symmetry of linear streptomycete chromosomes*, p. 2021.03.09.434596. Available at: https://doi.org/10.1101/2021.03.09.434596.

Algora-Gallardo, L. *et al.* (2021b) 'Bilateral symmetry of linear streptomycete chromosomes', *Microbial Genomics*, 7(11), p. 000692. Available at: https://doi.org/10.1099/mgen.0.000692.

Álvarez-Álvarez, R. *et al.* (2014) 'A 1.8-Mb-reduced Streptomyces clavuligerus genome: relevance for secondary metabolism and differentiation', *Applied Microbiology and Biotechnology*, 98(5), pp. 2183–2195. Available at: https://doi.org/10.1007/s00253-013-5382-z.

Ambler, R.P. (1980) 'The structure of beta-lactamases', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 289(1036), pp. 321–331. Available at: https://doi.org/10.1098/rstb.1980.0049.

Aminov, R.I. (2009) 'The role of antibiotics and antibiotic resistance in nature', *Environmental Microbiology*, 11(12), pp. 2970–2988. Available at: https://doi.org/10.1111/j.1462-2920.2009.01972.x.

Andersson, D.I. and Hughes, D. (2010) 'Antibiotic resistance and its cost: is it possible to reverse resistance?', *Nature Reviews. Microbiology*, 8(4), pp. 260–271. Available at: https://doi.org/10.1038/nrmicro2319.

Baggaley, K.H., Brown, A.G. and Schofield, C.J. (1997) 'Chemistry and biosynthesis of clavulanic acid and other clavams', *Natural Product Reports*, 14(4), pp. 309–333. Available at: https://doi.org/10.1039/NP9971400309.

Bakker, M.G. *et al.* (2014) 'Diffuse symbioses: roles of plant-plant, plant-microbe and microbe-microbe interactions in structuring the soil microbiome', *Molecular Ecology*, 23(6), pp. 1571–1583. Available at: https://doi.org/10.1111/mec.12571.

Baltz, R.H. (2012) 'Streptomyces temperate bacteriophage integration systems for stable genetic engineering of actinomycetes (and other organisms)', *Journal of Industrial Microbiology & Biotechnology*, 39(5), pp. 661–672. Available at: https://doi.org/10.1007/s10295-011-1069-6.

Baltz, R.H. (2016) 'Genetic manipulation of secondary metabolite biosynthesis for improved production in Streptomyces and other actinomycetes', *Journal of Industrial Microbiology & Biotechnology*, 43(2), pp. 343–370. Available at: https://doi.org/10.1007/s10295-015-1682-x.

Bankevich, A. *et al.* (2012) 'SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing', *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 19(5), pp. 455–477. Available at: https://doi.org/10.1089/cmb.2012.0021.

Bao, K. and Cohen, S.N. (2001) 'Terminal proteins essential for the replication of linear plasmids and chromosomes in Streptomyces', *Genes & Development*, 15(12), pp. 1518–1527. Available at: https://doi.org/10.1101/gad.896201.

Bao, K. and Cohen, S.N. (2003) 'Recruitment of terminal protein to the ends of Streptomyces linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication', *Genes & Development*, 17(6), pp. 774–785. Available at: https://doi.org/10.1101/gad.1060303.

Bao, K. and Cohen, S.N. (2004) 'Reverse transcriptase activity innate to DNA polymerase I and DNA topoisomerase I proteins of Streptomyces telomere complex', *Proceedings of the National Academy of Sciences of the United States of America*, 101(40), pp. 14361–14366. Available at: https://doi.org/10.1073/pnas.0404386101.

Behrens, O.K. and Corse, J. (1948) 'Biosynthesis of penicillins; preparation and evaluation of precursors for new penicillins', *The Journal of Biological Chemistry*, 175(2), pp. 771–792.

Bentley, S.D. *et al.* (2002) 'Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)', *Nature*, 417(6885), pp. 141–147. Available at: https://doi.org/10.1038/417141a.

Bertelli, C. *et al.* (2017) 'IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets', *Nucleic Acids Research*, 45(W1), pp. W30–W35. Available at: https://doi.org/10.1093/nar/gkx343.

Beytur, A. *et al.* (2014) 'Oral Amoxicillin-Clavulanic Acid Treatment in Urinary Tract Infections Caused by Extended-Spectrum Beta-Lactamase–Producing Organisms', *Jundishapur Journal of Microbiology*, 8(1). Available at: https://doi.org/10.5812/jjm.13792.

Bierman, M. *et al.* (1992) ‘Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp’, *Gene*, 116(1), pp. 43–49. Available at: https://doi.org/10.1016/0378-1119(92)90627-2.

Bignell, D.R.D. *et al.* (2005) ‘Expression of ccaR, Encoding the Positive Activator of Cephamycin C and Clavulanic Acid Production in Streptomyces clavuligerus, Is Dependent on bldG’, *Antimicrobial Agents and Chemotherapy*, 49(4), pp. 1529–1541. Available at: https://doi.org/10.1128/AAC.49.4.1529-1541.2005.

Blommel, P.G. *et al.* (2007) ‘Enhanced Bacterial Protein Expression During Auto-induction Obtained by Alteration of Lac Repressor Dosage and Medium Composition’, *Biotechnology progress*, 23(3), pp. 585–598. Available at: https://doi.org/10.1021/bp070011x.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) ‘Trimmomatic: a flexible trimmer for Illumina sequence data’, *Bioinformatics*, 30(15), pp. 2114–2120. Available at: https://doi.org/10.1093/bioinformatics/btu170.

Boon, R.J. and Beale, A.S. (1987) ‘Response of Streptococcus pyogenes to therapy with amoxicillin or amoxicillin-clavulanic acid in a mouse model of mixed infection caused by Staphylococcus aureus and Streptococcus pyogenes.’, *Antimicrobial Agents and Chemotherapy*, 31(8), pp. 1204–1209.

Brown, A.G. (1986) ‘Clavulanic acid, a novel beta-lactamase inhibitor--a case study in drug discovery and development’, *Drug Design and Delivery*, 1(1), pp. 1–21.

Bu, Q.-T. *et al.* (2020) ‘Comprehensive dissection of dispensable genomic regions in Streptomyces based on comparative analysis approach’, *Microbial Cell Factories*, 19(1), p. 99. Available at: https://doi.org/10.1186/s12934-020-01359-4.

Bush, M.J. (2018) ‘The actinobacterial WhiB-like (Wbl) family of transcription factors’, *Molecular Microbiology*, 110(5), pp. 663–676. Available at: https://doi.org/10.1111/mmi.14117.

Cao, G. *et al.* (2016) ‘Complete Genome Sequence of Streptomyces clavuligerus F613-1, an Industrial Producer of Clavulanic Acid’, *Genome Announcements*, 4(5). Available at: https://doi.org/10.1128/genomeA.01020-16.

Chater, K.F. (1972) ‘A Morphological and Genetic Mapping Study of White Colony Mutants of Streptomyces coelicolor’, *Microbiology*, 72(1), pp. 9–28. Available at: https://doi.org/10.1099/00221287-72-1-9.

Cobb, R.E., Wang, Y. and Zhao, H. (2015) ‘High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System’, *ACS Synthetic Biology*, 4(6), pp. 723–728. Available at: https://doi.org/10.1021/sb500351f.

Combes, P. *et al.* (2002) ‘The streptomyces genome contains multiple pseudo-attB sites for the (phi)C31-encoded site-specific recombination system’, *Journal of Bacteriology*, 184(20), pp. 5746–5752. Available at: https://doi.org/10.1128/JB.184.20.5746-5752.2002.

Crack, J.C. *et al.* (2015) 'NsrR from Streptomyces coelicolor Is a Nitric Oxide-sensing [4Fe-4S] Cluster Protein with a Specialized Regulatory Function*', *Journal of Biological Chemistry*, 290(20), pp. 12689–12704. Available at: https://doi.org/10.1074/jbc.M115.643072.

Cui, L. *et al.* (2018) 'A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9', *Nature Communications*, 9(1), p. 1912. Available at: https://doi.org/10.1038/s41467-018-04209-5.

Dedrick, R.M., Wildschutte, H. and McCormick, J.R. (2009) 'Genetic Interactions of smc, ftsK, and parB Genes in Streptomyces coelicolor and Their Developmental Genome Segregation Phenotypes', *Journal of Bacteriology*, 191(1), pp. 320–332. Available at: https://doi.org/10.1128/JB.00858-08.

Deng, Z.-L. *et al.* (2022) 'Rapid and accurate identification of ribosomal RNA sequences via deep learning', *Nucleic Acids Research*, 50(10), p. e60. Available at: https://doi.org/10.1093/nar/gkac112.

Diesh, C. *et al.* (2022) 'JBrowse 2: A modular genome browser with views of synteny and structural variation'. bioRxiv, p. 2022.07.28.501447. Available at: https://doi.org/10.1101/2022.07.28.501447.

Du, F. *et al.* (2021) 'Regulating the T7 RNA polymerase expression in E. coli BL21 (DE3) to provide more host options for recombinant protein production', *Microbial Cell Factories*, 20, p. 189. Available at: https://doi.org/10.1186/s12934-021-01680-6.

Du, L. *et al.* (2012) 'An Efficient Intergeneric Conjugation of DNA from Escherichia coli to Mycelia of the Lincomycin-Producer Streptomyces lincolnensis', *International Journal of Molecular Sciences*, 13(4), pp. 4797–4806. Available at: https://doi.org/10.3390/ijms13044797.

Elander, R.P. (2003) 'Industrial production of beta-lactam antibiotics', *Applied Microbiology and Biotechnology*, 61(5–6), pp. 385–392. Available at: https://doi.org/10.1007/s00253-003-1274-y.

Felsenstein, J. (1985) 'Confidence Limits on Phylogenies: An Approach Using the Bootstrap', *Evolution*, 39(4), pp. 783–791. Available at: https://doi.org/10.2307/2408678.

Fernandes, R., Amador, P. and Prudêncio, C. (2013) 'β-Lactams: chemical structure, mode of action and mechanisms of resistance', *Reviews in Medical Microbiology*, 24(1), p. 7. Available at: https://doi.org/10.1097/MRM.0b013e3283587727.

Fischer, B. *et al.* (1992) 'A novel sequential procedure to enhance the renaturation of recombinant protein from Escherichia coli inclusion bodies', *Protein Engineering*, 5(6), pp. 593–596. Available at: https://doi.org/10.1093/protein/5.6.593.

Fisher, J.F., Meroueh, S.O. and Mobashery, S. (2005) 'Bacterial Resistance to β-Lactam Antibiotics: Compelling Opportunism, Compelling Opportunity †', *Chemical Reviews*, 105(2), pp. 395–424. Available at: https://doi.org/10.1021/cr030102i.

Fong, R. *et al.* (2007) 'Characterization of a Large, Stable, High-Copy-Number Streptomyces Plasmid That Requires Stability and Transfer Functions for Heterologous

Polyketide Overproduction', *Applied and Environmental Microbiology*, 73(4), pp. 1296–1307. Available at: https://doi.org/10.1128/AEM.01888-06.

Francis, D.M. and Page, R. (2010) 'Strategies to Optimize Protein Expression in E. coli', *Current Protocols in Protein Science*, 61(1), pp. 5241–52429. Available at: https://doi.org/10.1002/0471140864.ps0524s61.

Gasteiger, E. *et al.* (2003) 'ExPASy: The proteomics server for in-depth protein knowledge and analysis', *Nucleic Acids Research*, 31(13), pp. 3784–3788. Available at: https://doi.org/10.1093/nar/gkg563.

Geddes, A.M., Klugman, K.P. and Rolinson, G.N. (2007) 'Introduction: historical perspective and development of amoxicillin/clavulanate', *International Journal of Antimicrobial Agents*, 30, pp. 109–112. Available at: https://doi.org/10.1016/j.ijantimicag.2007.07.015.

Gehrke, E.J. *et al.* (2019) 'Silencing cryptic specialized metabolism in Streptomyces by the nucleoid-associated protein Lsr2', *eLife*, 8, p. e47691. Available at: https://doi.org/10.7554/eLife.47691.

Gomez-Escribano, J.P. *et al.* (2021) 'Genome editing reveals that pSCL4 is required for chromosome linearity in Streptomyces clavuligerus', *Microbial Genomics*, 7(11), p. 000669. Available at: https://doi.org/10.1099/mgen.0.000669.

Gootz, T.D. and Sanders, C.C. (1983) 'Characterization of beta-lactamase induction in Enterobacter cloacae', *Antimicrobial Agents and Chemotherapy*, 23(1), pp. 91–97. Available at: https://doi.org/10.1128/aac.23.1.91.

Goshi, K. *et al.* (2002) 'Cloning and Analysis of the Telomere and Terminal Inverted Repeat of the Linear Chromosome of Streptomyces griseus', *Journal of Bacteriology*, 184(12), pp. 3411–3415. Available at: https://doi.org/10.1128/JB.184.12.3411-3415.2002.

de Groot, N.S., Sabate, R. and Ventura, S. (2009) 'Amyloids in bacterial inclusion bodies', *Trends in Biochemical Sciences*, 34(8), pp. 408–416. Available at: https://doi.org/10.1016/j.tibs.2009.03.009.

Guisbert, E. *et al.* (2004) 'A chaperone network controls the heat shock response in E. coli', *Genes & Development*, 18(22), pp. 2812–2821. Available at: https://doi.org/10.1101/gad.1219204.

Gurevich, A. *et al.* (2013) 'QUAST: quality assessment tool for genome assemblies', *Bioinformatics (Oxford, England)*, 29(8), pp. 1072–1075. Available at: https://doi.org/10.1093/bioinformatics/btt086.

Guruprasad, K., Reddy, B.V.B. and Pandit, M.W. (1990) 'Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence', *Protein Engineering, Design and Selection*, 4(2), pp. 155–161. Available at: https://doi.org/10.1093/protein/4.2.155.

Hakenbeck, R. (1999) ‘β-Lactam-Resistant Streptococcus pneumoniae: Epidemiology and Evolutionary Mechanism’, *Chemotherapy*, 45(2), pp. 83–94. Available at: https://doi.org/10.1159/000007170.

Hamed, R.B. *et al.* (2013) ‘The enzymes of β-lactam biosynthesis’, *Nat. Prod. Rep.*, 30(1), pp. 21–107. Available at: https://doi.org/10.1039/C2NP20065A.

Hennessey, T.D. (1967) ‘Inducible beta-lactamase in Enterobacter’, *Journal of General Microbiology*, 49(2), pp. 277–285. Available at: https://doi.org/10.1099/00221287-49-2-277.

Higgens, C.E. and Kástner, R. (1971) ‘Streptomyces clavuligerus sp. nov., a β-Lactam Antibiotic Producer’, in. Available at: https://doi.org/10.1099/00207713-21-4-326.

Hirochika, H. and Sakaguchi, K. (1982) ‘Analysis of linear plasmids isolated from Streptomyces: association of protein with the ends of the plasmid DNA’, *Plasmid*, 7(1), pp. 59–65. Available at: https://doi.org/10.1016/0147-619x(82)90027-0.

Hoff, G. *et al.* (2018) ‘Genome plasticity is governed by double strand break DNA repair in Streptomyces’, *Scientific Reports*, 8(1), p. 5272. Available at: https://doi.org/10.1038/s41598-018-23622-w.

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

Hsu, C.-C. and Chen, C.W. (2010) ‘Linear Plasmid SLP2 Is Maintained by Partitioning, Intrahyphal Spread, and Conjugal Transfer in Streptomyces’, *Journal of Bacteriology*, 192(1), pp. 307–315. Available at: https://doi.org/10.1128/JB.01192-09.

Huang, C.-H. *et al.* (2007) ‘The telomere system of the Streptomyces linear plasmid SCP1 represents a novel class’, *Molecular Microbiology*, 63(6), pp. 1710–1718. Available at: https://doi.org/10.1111/j.1365-2958.2007.05616.x.

Huang, H. *et al.* (2015) ‘One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces’, *Acta Biochimica Et Biophysica Sinica*, 47(4), pp. 231–243. Available at: https://doi.org/10.1093/abbs/gmv007.

Hugonnet, J.-E. and Blanchard, J.S. (2007) ‘Irreversible Inhibition of the Mycobacterium tuberculosis β-Lactamase by Clavulanate’, *Biochemistry*, 46(43), pp. 11998–12004. Available at: https://doi.org/10.1021/bi701506h.

Idicula-Thomas, S. and Balaji, P.V. (2005) ‘Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli’, *Protein Science : A Publication of the Protein Society*, 14(3), pp. 582–592. Available at: https://doi.org/10.1110/ps.041009005.

Ikeda, H. *et al.* (2003) ‘Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis’, *Nature Biotechnology*, 21(5), pp. 526–531. Available at: https://doi.org/10.1038/nbt820.

Jaurin, B. and Grundström, T. (1981) 'ampC cephalosporinase of Escherichia coli K-12 has a different evolutionary origin from that of beta-lactamases of the penicillinase type', *Proceedings of the National Academy of Sciences of the United States of America*, 78(8), pp. 4897–4901. Available at: https://doi.org/10.1073/pnas.78.8.4897.

Jensen, S.E. *et al.* (2000) 'Enzymes catalyzing the early steps of clavulanic acid biosynthesis are encoded by two sets of paralogous genes in Streptomyces clavuligerus', *Antimicrobial Agents and Chemotherapy*, 44(3), pp. 720–726. Available at: https://doi.org/10.1128/aac.44.3.720-726.2000.

Jensen, S.E. and Paradkar, A.S. (1999) 'Biosynthesis and Molecular Genetics of Clavulanic Acid', *Antonie van Leeuwenhoek*, 75(1), pp. 125–133. Available at: https://doi.org/10.1023/A:1001755724055.

Jevševar, S. *et al.* (2005) 'Production of Nonclassical Inclusion Bodies from Which Correctly Folded Protein Can Be Extracted', *Biotechnology Progress*, 21(2), pp. 632–639. Available at: https://doi.org/10.1021/bp0497839.

Juhas, M. *et al.* (2009) 'Genomic islands: tools of bacterial horizontal gene transfer and evolution', *Fems Microbiology Reviews*, 33(2), pp. 376–393. Available at: https://doi.org/10.1111/j.1574-6976.2008.00136.x.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

Khan, S.T. *et al.* (2011) 'Streptomyces associated with a marine sponge Haliclona sp.; biosynthetic genes for secondary metabolites and products', *Environmental Microbiology*, 13(2), pp. 391–403. Available at: https://doi.org/10.1111/j.1462-2920.2010.02337.x.

Kieser, T *et al.* (2000) *Practical Streptomyces Genetics*. Norwich, UK: John Innes Foundation.

Kinashi, H. (2011) 'Giant linear plasmids in Streptomyces : a treasure trove of antibiotic biosynthetic clusters', *The Journal of Antibiotics*, 64(1), pp. 19–25. Available at: https://doi.org/10.1038/ja.2010.146.

Kirby, R. (1978) 'An unstable genetic element affecting the production of the antibiotic holomycin by Streptomyces clavuligerus', *FEMS Microbiology Letters*, 3(5), pp. 283–286. Available at: https://doi.org/10.1111/j.1574-6968.1978.tb01948.x.

Kliebe, C. *et al.* (1985) 'Evolution of plasmid-coded resistance to broad-spectrum cephalosporins.', *Antimicrobial Agents and Chemotherapy*, 28(2), pp. 302–307.

Kong, K.-F., Schneper, L. and Mathee, K. (2010) 'Beta-lactam Antibiotics: From Antibiosis to Resistance and Bacteriology', *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*, 118(1), pp. 1–36. Available at: https://doi.org/10.1111/j.1600-0463.2009.02563.x.

Krissinel, E. and Henrick, K. (2004) 'Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions', *Acta Crystallographica Section D:*

Biological Crystallography, 60(12), pp. 2256–2268. Available at: https://doi.org/10.1107/S0907444904026460.

Kronheim, S. *et al.* (2018) 'A chemical defence against phage infection', *Nature*, 564(7735), pp. 283–286. Available at: https://doi.org/10.1038/s41586-018-0767-x.

Langmead, B. and Salzberg, S.L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, 9(4), pp. 357–359. Available at: https://doi.org/10.1038/nmeth.1923.

Larson, M.H. *et al.* (2013) 'CRISPR interference (CRISPRi) for sequence-specific control of gene expression', *Nature Protocols*, 8(11), pp. 2180–2196. Available at: https://doi.org/10.1038/nprot.2013.132.

Leflon-Guibout, V. *et al.* (2000) 'Epidemiological Survey of Amoxicillin-Clavulanate Resistance and Corresponding Molecular Mechanisms in Escherichia coli Isolates in France: New Genetic Features of blaTEM Genes', *Antimicrobial Agents and Chemotherapy*, 44(10), pp. 2709–2714.

Lemozy, J. *et al.* (1995) 'First characterization of inhibitor-resistant TEM (IRT) beta-lactamases in Klebsiella pneumoniae strains', *Antimicrobial Agents and Chemotherapy*, 39(11), pp. 2580–2582. Available at: https://doi.org/10.1128/aac.39.11.2580.

Leskiw, B.K. *et al.* (1990) 'Discovery of an insertion sequence, IS116, from Streptomyces clavuligerus and its relatedness to other transposable elements from actinomycetes', *Microbiology*, 136(7), pp. 1251–1258. Available at: https://doi.org/10.1099/00221287-136-7-1251.

Letunic, I. and Bork, P. (2021) 'Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation', *Nucleic Acids Research*, 49(W1), pp. W293–W296. Available at: https://doi.org/10.1093/nar/gkab301.

de Lima Procópio, R.E. *et al.* (2012) 'Antibiotics produced by Streptomyces', *The Brazilian Journal of Infectious Diseases*, 16(5), pp. 466–471. Available at: https://doi.org/10.1016/j.bjid.2012.08.014.

Lin, Y.-S. *et al.* (1993) 'The chromosomal DNA of Streptomyces lividans 66 is linear', *Molecular Microbiology*, 10(5), pp. 923–933. Available at: https://doi.org/10.1111/j.1365-2958.1993.tb00964.x.

Liu, H., Naismith, J.H. and Hay, R.T. (2003) 'Adenovirus DNA replication', *Current Topics in Microbiology and Immunology*, 272, pp. 131–164. Available at: https://doi.org/10.1007/978-3-662-05597-7_5.

de Marco, A. (2009) 'Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli', *Microbial Cell Factories*, 8(1), p. 26. Available at: https://doi.org/10.1186/1475-2859-8-26.

McCarty, N.S. *et al.* (2019) 'Rapid Assembly of gRNA Arrays via Modular Cloning in Yeast', *ACS Synthetic Biology*, 8(4), pp. 906–910. Available at: https://doi.org/10.1021/acssynbio.9b00041.

McCarty, N.S. *et al.* (2020) ‘Multiplexed CRISPR technologies for gene editing and transcriptional regulation’, *Nature Communications*, 11(1), p. 1281. Available at: https://doi.org/10.1038/s41467-020-15053-x.

McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) ‘The PSIPRED protein structure prediction server’, *Bioinformatics*, 16(4), pp. 404–405. Available at: https://doi.org/10.1093/bioinformatics/16.4.404.

McLeod, M.P. *et al.* (2006) ‘The complete genome of Rhodococcus sp. RHA1 provides insights into a catabolic powerhouse’, *Proceedings of the National Academy of Sciences*, 103(42), pp. 15582–15587. Available at: https://doi.org/10.1073/pnas.0607048103.

Medema, M.H. *et al.* (2010) ‘The Sequence of a 1.8-Mb Bacterial Linear Plasmid Reveals a Rich Evolutionary Reservoir of Secondary Metabolic Pathways’, *Genome Biology and Evolution*, 2, pp. 212–224. Available at: https://doi.org/10.1093/gbe/evq013.

Modrzejewski, D. *et al.* (2020) ‘Which Factors Affect the Occurrence of Off-Target Effects Caused by the Use of CRISPR/Cas: A Systematic Review in Plants’, *Frontiers in Plant Science*, 11. Available at: https://www.frontiersin.org/articles/10.3389/fpls.2020.574959 (Accessed: 8 July 2023).

Nabais, A.M.A. and da Fonseca, M.M.R. (1995) ‘The effect of solid medium composition on growth and sporulation ofStreptomyces clavuligerus; spore viability during storage at +4°C’, *Biotechnology Techniques*, 9(5), pp. 361–364. Available at: https://doi.org/10.1007/BF00638871.

Nagarajan, R. *et al.* (1971) ‘Beta-lactam antibiotics from Streptomyces’, *Journal of the American Chemical Society*, 93(9), pp. 2308–2310. Available at: https://doi.org/10.1021/ja00738a035.

Nakagawa, J. *et al.* (1984) ‘Functional biosynthesis of cell wall peptidoglycan by polymorphic bifunctional polypeptides. Penicillin-binding protein 1Bs of Escherichia coli with activities of transglycosylase and transpeptidase’, *The Journal of Biological Chemistry*, 259(22), pp. 13937–13946.

Nanji, T. *et al.* (2019) ‘Streptomyces IHF uses multiple interfaces to bind DNA’, *Biochimica Et Biophysica Acta. General Subjects*, 1863(11), p. 129405. Available at: https://doi.org/10.1016/j.bbagen.2019.07.014.

Nindita, Y. *et al.* (2015) ‘The tap-tpg gene pair on the linear plasmid functions to maintain a linear topology of the chromosome in Streptomyces rochei’, *Molecular Microbiology*, 95(5), pp. 846–858. Available at: https://doi.org/10.1111/mmi.12904.

Nindita, Y. *et al.* (2019) ‘The genome sequence of Streptomyces rochei 7434AN4, which carries a linear chromosome and three characteristic linear plasmids’, *Scientific Reports*, 9(1), p. 10973. Available at: https://doi.org/10.1038/s41598-019-47406-y.

Niu, G. *et al.* (2016) ‘Specialised metabolites regulating antibiotic biosynthesis in Streptomyces spp.’, *FEMS Microbiology Reviews*, 40(4), pp. 554–573. Available at: https://doi.org/10.1093/femsre/fuw012.

Novakova, R. *et al.* (2021) ‘The linear plasmid pSA3239 is essential for the replication of the Streptomyces lavendulae subsp. lavendulae CCM 3239 chromosome’, *Research in Microbiology*, p. 103870. Available at: https://doi.org/10.1016/j.resmic.2021.103870.

Ohnishi, Y. *et al.* (2008) ‘Genome sequence of the streptomycin-producing microorganism Streptomyces griseus IFO 13350’, *Journal of Bacteriology*, 190(11), pp. 4050–4060. Available at: https://doi.org/10.1128/JB.00204-08.

Okonechnikov, K., Conesa, A. and García-Alcalde, F. (2016) ‘Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data’, *Bioinformatics*, 32(2), pp. 292–294. Available at: https://doi.org/10.1093/bioinformatics/btv566.

Pages, J.-M. *et al.* (2009) ‘Efflux Pump, the Masked Side of ß-Lactam Resistance in Klebsiella pneumoniae Clinical Isolates’, *PLOS ONE*, 4(3), p. e4817. Available at: https://doi.org/10.1371/journal.pone.0004817.

Pandza, S. *et al.* (1998) ‘Recombination between the linear plasmid pPZG101 and the linear chromosome of Streptomyces rimosus can lead to exchange of ends’, *Molecular Microbiology*, 28(6), pp. 1165–1176. Available at: https://doi.org/10.1046/j.1365-2958.1998.00877.x.

Paradkar, A. (2013) ‘Clavulanic acid production by Streptomyces clavuligerus : biogenesis, regulation and strain improvement’, *The Journal of Antibiotics*, 66(7), pp. 411–420. Available at: https://doi.org/10.1038/ja.2013.26.

Periwal, V. and Scaria, V. (2015) ‘Insights into structural variations and genome rearrangements in prokaryotic genomes’, *Bioinformatics*, 31(1), pp. 1–9. Available at: https://doi.org/10.1093/bioinformatics/btu600.

Peternel, Š. and Komel, R. (2010) ‘Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells’, *Microbial Cell Factories*, 9(1), p. 66. Available at: https://doi.org/10.1186/1475-2859-9-66.

Pethick, F.E. *et al.* (2013) ‘Draft Genome Sequence of the Oxytetracycline-Producing Bacterium Streptomyces rimosus ATCC 10970’, *Genome Announcements*, 1(2), p. e0006313. Available at: https://doi.org/10.1128/genomeA.00063-13.

Piette, A. *et al.* (2005) ‘From dormant to germinating spores of Streptomyces coelicolor A3(2): new perspectives from the crp null mutant’, *Journal of Proteome Research*, 4(5), pp. 1699–1708. Available at: https://doi.org/10.1021/pr050155b.

Qi, L.S. *et al.* (2013) ‘Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression’, *Cell*, 152(5), pp. 1173–1183. Available at: https://doi.org/10.1016/j.cell.2013.02.022.

Qin, Z. and Cohen, S.N. (1998) ‘Replication at the telomeres of the Streptomyces linear plasmid pSLA2’, *Molecular Microbiology*, 28(5), pp. 893–903. Available at: https://doi.org/10.1046/j.1365-2958.1998.00838.x.

Robinson, J.T. *et al.* (2011) ‘Integrative Genomics Viewer’, *Nature biotechnology*, 29(1), pp. 24–26. Available at: https://doi.org/10.1038/nbt.1754.

Rostain, W. *et al.* (2023) 'Cas9 off-target binding to the promoter of bacterial genes leads to silencing and toxicity', *Nucleic Acids Research*, p. gkad170. Available at: https://doi.org/10.1093/nar/gkad170.

Rudolph, M.M., Vockenhuber, M.-P. and Suess, B. (2013) 'Synthetic riboswitches for the conditional control of gene expression in Streptomyces coelicolor', *Microbiology (Reading, England)*, 159(Pt 7), pp. 1416–1422. Available at: https://doi.org/10.1099/mic.0.067322-0.

Rudolph, R. and Lilie, H. (1996) 'In vitro folding of inclusion body proteins', *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 10(1), pp. 49–56.

Salas, M. (1991) 'Protein-priming of DNA replication', *Annual Review of Biochemistry*, 60, pp. 39–71. Available at: https://doi.org/10.1146/annurev.bi.60.070191.000351.

Salas, M. *et al.* (2016) 'DNA-Binding Proteins Essential for Protein-Primed Bacteriophage Φ29 DNA Replication', *Frontiers in Molecular Biosciences*, 3. Available at: https://doi.org/10.3389/fmolb.2016.00037.

Schultsz, C. and Geerlings, S. (2012) 'Plasmid-Mediated Resistance in Enterobacteriaceae', *Drugs*, 72(1), pp. 1–16. Available at: https://doi.org/10.2165/11597960-000000000-00000.

Schwedock, J. *et al.* (1997) 'Assembly of the cell division protein FtsZ into ladder-like structures in the aerial hyphae of Streptomyces coelicolor', *Molecular Microbiology*, 25(5), pp. 847–858. Available at: https://doi.org/10.1111/j.1365-2958.1997.mmi507.x.

Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics*, 30(14), pp. 2068–2069. Available at: https://doi.org/10.1093/bioinformatics/btu153.

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

Sepulveda, E., Vogelmann, J. and Muth, G. (2011) 'A septal chromosome segregator protein evolved into a conjugative DNA-translocator protein', *Mobile Genetic Elements*, 1(3), pp. 225–229. Available at: https://doi.org/10.4161/mge.1.3.18066.

Severin, A., Figueiredo, A.M. and Tomasz, A. (1996) 'Separation of abnormal cell wall composition from penicillin resistance through genetic transformation of Streptococcus pneumoniae.', *Journal of Bacteriology*, 178(7), pp. 1788–1792.

Shahid, M. *et al.* (2009) 'Beta-lactams and beta-lactamase-inhibitors in current- or potential-clinical practice: a comprehensive update', *Critical Reviews in Microbiology*, 35(2), pp. 81–108. Available at: https://doi.org/10.1080/10408410902733979.

Sievers, F. *et al.* (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*, 7(1), p. 539. Available at: https://doi.org/10.1038/msb.2011.75.

Sigrist, C.J.A. *et al.* (2013) 'New and continuing developments at PROSITE', *Nucleic Acids Research*, 41(D1), pp. D344–D347. Available at: https://doi.org/10.1093/nar/gks1067.

Sillitoe, I. *et al.* (2019) 'CATH: expanding the horizons of structure-based functional annotations for genome sequences', *Nucleic Acids Research*, 47(Database issue), pp. D280–D284. Available at: https://doi.org/10.1093/nar/gky1097.

Singh, A. *et al.* (2015) 'Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process', *Microbial Cell Factories*, 14(1), p. 41. Available at: https://doi.org/10.1186/s12934-015-0222-8.

Singh, S.M. *et al.* (2012) 'Solubilization of inclusion body proteins using n-propanol and its refolding into bioactive form', *Protein Expression and Purification*, 81(1), pp. 75–82. Available at: https://doi.org/10.1016/j.pep.2011.09.004.

Song, J.Y. *et al.* (2010) 'Draft Genome Sequence of Streptomyces clavuligerus NRRL 3585, a Producer of Diverse Secondary Metabolites', *Journal of Bacteriology*, 192(23), pp. 6317–6318. Available at: https://doi.org/10.1128/JB.00859-10.

Spratt, B.G. *et al.* (1986) 'Kanamycin-resistant vectors that are analogues of plasmids pUC8, pUC9, pEMBL8 and pEMBL9', *Gene*, 41(2–3), pp. 337–342. Available at: https://doi.org/10.1016/0378-1119(86)90117-4.

Studier, F.W. (2005) 'Protein production by auto-induction in high-density shaking cultures', *Protein Expression and Purification*, 41(1), pp. 207–234. Available at: https://doi.org/10.1016/j.pep.2005.01.016.

Studier, F.W. and Moffatt, B.A. (1986) 'Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes', *Journal of Molecular Biology*, 189(1), pp. 113–130. Available at: https://doi.org/10.1016/0022-2836(86)90385-2.

Sun, F.-H. *et al.* (2014) 'Development of an intergeneric conjugal transfer system for xinaomycins-producing Streptomyces noursei Xinao-4', *International Journal of Molecular Sciences*, 15(7), pp. 12217–12230. Available at: https://doi.org/10.3390/ijms150712217.

Suzuki, H. *et al.* (1980) 'In vitro peptidoglycan polymerization catalysed by penicillin binding protein 1b of Escherichia coli K-12', *FEBS Letters*, 110(2), pp. 245–249. Available at: https://doi.org/10.1016/0014-5793(80)80083-4.

Tamura, K., Stecher, G. and Kumar, S. (2021) 'MEGA11: Molecular Evolutionary Genetics Analysis Version 11', *Molecular Biology and Evolution*, 38(7), pp. 3022–3027. Available at: https://doi.org/10.1093/molbev/msab120.

Thoma, L. and Muth, G. (2016) 'Conjugative DNA-transfer in Streptomyces , a mycelial organism', *Plasmid*, 87–88, pp. 1–9. Available at: https://doi.org/10.1016/j.plasmid.2016.09.004.

Tidjani, A.-R., Bontemps, C. and Leblond, P. (2020) 'Telomeric and sub-telomeric regions undergo rapid turnover within a Streptomyces population', *Scientific Reports*, 10(1), p. 7720. Available at: https://doi.org/10.1038/s41598-020-63912-w.

Tong, Y. *et al.* (2015) 'CRISPR-Cas9 Based Engineering of Actinomycetal Genomes', *ACS synthetic biology*, 4(9), pp. 1020–1029. Available at: https://doi.org/10.1021/acssynbio.5b00038.

Tong, Y. *et al.* (2019) 'Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST', *Proceedings of the National Academy of Sciences*, p. 201913493. Available at: https://doi.org/10.1073/pnas.1913493116.

Toussaint, K.A. and Gallagher, J.C. (2015) 'β-Lactam/β-Lactamase Inhibitor Combinations: From Then to Now', *Annals of Pharmacotherapy*, 49(1), pp. 86–98. Available at: https://doi.org/10.1177/1060028014556652.

Tsai, H.-H. *et al.* (2008) 'Terminal proteins of Streptomyces chromosome can target DNA into eukaryotic nuclei', *Nucleic Acids Research*, 36(10), p. e62. Available at: https://doi.org/10.1093/nar/gkm1170.

Vasimuddin, Md. *et al.* (2019) 'Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems', in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 314–324. Available at: https://doi.org/10.1109/IPDPS.2019.00041.

Ventola, C.L. (2015) 'The Antibiotic Resistance Crisis', *Pharmacy and Therapeutics*, 40(4), pp. 277–283.

Wang, L. *et al.* (2007) 'Role of an FtsK-Like Protein in Genetic Stability in Streptomyces coelicolor A3(2)', *Journal of Bacteriology*, 189(6), pp. 2310–2318. Available at: https://doi.org/10.1128/JB.01660-06.

Wang, W. *et al.* (2013) 'An Engineered Strong Promoter for Streptomycetes', *Applied and Environmental Microbiology*, 79(14), pp. 4484–4492. Available at: https://doi.org/10.1128/AEM.00985-13.

Wang, X.-K. and Jin, J.-L. (2014) 'Crucial factor for increasing the conjugation frequency in Streptomyces netropsis SD-07 and other strains', *FEMS Microbiology Letters*, 357(1), pp. 99–103. Available at: https://doi.org/10.1111/1574-6968.12507.

Weissenburger, H.W.O. and Hoeven, M.G. van der (1970) 'An efficient nonenzymatic conversion of benzylpenicillin to 6-aminopenicillanic acid', *Recueil des Travaux Chimiques des Pays-Bas*, 89(10), pp. 1081–1084. Available at: https://doi.org/10.1002/recl.19700891011.

Wlodek, A. *et al.* (2017) 'Diversity oriented biosynthesis via accelerated evolution of modular gene clusters', *Nature Communications*, 8(1), p. 1206. Available at: https://doi.org/10.1038/s41467-017-01344-3.

Wood, W.B. (1966) 'Host specificity of DNA produced by Escherichia coli: bacterial mutations affecting the restriction and modification of DNA', *Journal of Molecular Biology*, 16(1), pp. 118–133. Available at: https://doi.org/10.1016/s0022-2836(66)80267-x.

Wright, F. and Bibb, M.J. (1992) 'Codon usage in the G+C-rich Streptomyces genome', *Gene*, 113(1), pp. 55–65. Available at: https://doi.org/10.1016/0378-1119(92)90669-g.

Wu, H. *et al.* (2012) 'Genomic and transcriptomic insights into the thermo-regulated biosynthesis of validamycin in Streptomyces hygroscopicus 5008', *BMC Genomics*, 13(1), p. 337. Available at: https://doi.org/10.1186/1471-2164-13-337.

Wu, W. *et al.* (2006) 'Prediction and functional analysis of the replication origin of the linear plasmid pSCL2 in Streptomyces clavuligerus', *Canadian Journal of Microbiology*, 52(4), pp. 293–300. Available at: https://doi.org/10.1139/w05-126.

Wu, X., Kriz, A.J. and Sharp, P.A. (2014) 'Target specificity of the CRISPR-Cas9 system', *Quantitative biology*, 2(2), pp. 59–70. Available at: https://doi.org/10.1007/s40484-014-0030-x.

Wu, X. and Roy, K.L. (1993) 'Complete nucleotide sequence of a linear plasmid from Streptomyces clavuligerus and characterization of its RNA transcripts.', *Journal of Bacteriology*, 175(1), pp. 37–52.

Yamasaki, M. *et al.* (2003) 'Limited regions of homology between linear and circular plasmids encoding methylenomycin biosynthesis in two independently isolated streptomycetes', *Microbiology (Reading, England)*, 149(Pt 5), pp. 1351–1356. Available at: https://doi.org/10.1099/mic.0.26102-0.

Yamasaki, M. and Kinashi, H. (2004) 'Two Chimeric Chromosomes of Streptomyces coelicolor A3(2) Generated by Single Crossover of the Wild-Type Chromosome and Linear Plasmid SCP1', *Journal of Bacteriology*, 186(19), pp. 6553–6559. Available at: https://doi.org/10.1128/JB.186.19.6553-6559.2004.

Yang, C.-C. *et al.* (2002) 'The terminal proteins of linear Streptomyces chromosomes and plasmids: a novel class of replication priming proteins', *Molecular Microbiology*, 43(2), pp. 297–305.

Yang, C.-C. *et al.* (2013) 'Mutational Analysis of the Terminal Protein Tpg of Streptomyces Chromosomes: Identification of the Deoxynucleotidylation Site', *PLoS ONE*, 8(2). Available at: https://doi.org/10.1371/journal.pone.0056322.

Yang, C.-C. *et al.* (2017) 'Telomere associated primase Tap repairs truncated telomeres of Streptomyces', *Nucleic Acids Research*, 45(10), pp. 5838–5849. Available at: https://doi.org/10.1093/nar/gkx189.

Yang, C.-C., Tseng, S.-M. and Chen, C.W. (2015) 'Telomere-associated proteins add deoxynucleotides to terminal proteins during replication of the telomeres of linear chromosomes and plasmids in Streptomyces', *Nucleic Acids Research*, 43(13), pp. 6373–6383. Available at: https://doi.org/10.1093/nar/gkv302.

Zakrzewska-Czerwińska, J., Majka, J. and Schrempf, H. (1995) 'Minimal requirements of the Streptomyces lividans 66 oriC region and its transcriptional and translational activities.', *Journal of Bacteriology*, 177(16), pp. 4765–4771. Available at: https://doi.org/10.1128/jb.177.16.4765-4771.1995.

Zapun, A., Vernet, T. and Pinho, M.G. (2008) 'The different shapes of cocci', *FEMS Microbiology Reviews*, 32(2), pp. 345–360. Available at: https://doi.org/10.1111/j.1574-6976.2007.00098.x.

Zeng, H. *et al.* (2015) 'Highly efficient editing of the actinorhodin polyketide chain length factor gene in Streptomyces coelicolor M145 using CRISPR/Cas9-CodA(sm) combined system', *Applied Microbiology and Biotechnology*, 99(24), pp. 10575–10585. Available at: https://doi.org/10.1007/s00253-015-6931-4.

Zhang, H. *et al.* (2013) 'Draft Genome Sequence of Streptomyces bottropensis ATCC 25435, a Bottromycin-Producing Actinomycete', *Genome Announcements*, 1(2). Available at: https://doi.org/10.1128/genomeA.00019-13.

Zhang, R. *et al.* (2006a) 'Diversity of Telomere Palindromic Sequences and Replication Genes among Streptomyces Linear Plasmids', *Applied and Environmental Microbiology*, 72(9), pp. 5728–5733. Available at: https://doi.org/10.1128/AEM.00707-06.

Zhang, R. *et al.* (2006b) 'Diversity of telomere palindromic sequences and replication genes among Streptomyces linear plasmids', *Applied and Environmental Microbiology*, 72(9), pp. 5728–5733. Available at: https://doi.org/10.1128/AEM.00707-06.

Zhang, S. *et al.* (2019) 'Establishment of a highly efficient conjugation protocol for Streptomyces kanamyceticus ATCC12853', *MicrobiologyOpen*, 8(6), p. e00747. Available at: https://doi.org/10.1002/mbo3.747.

Zhang, X. *et al.* (2017) 'Multiplex gene regulation by CRISPR-ddCpf1', *Cell Discovery*, 3(1), pp. 1–9. Available at: https://doi.org/10.1038/celldisc.2017.18.

Zhang, Z. *et al.* (2020) 'Antibiotic production in Streptomyces is organized by a division of labor through terminal genomic differentiation', *Science Advances*, 6(3), p. eaay5781. Available at: https://doi.org/10.1126/sciadv.aay5781.

Zhao, D. *et al.* (2017) 'CRISPR/Cas9-assisted gRNA-free one-step genome editing with no sequence limitations and improved targeting efficiency', *Scientific Reports*, 7(1), p. 16624. Available at: https://doi.org/10.1038/s41598-017-16998-8.

Zhou, Z. *et al.* (2011) 'The -omics Era- Toward a Systems-Level Understanding of Streptomyces', *Current Genomics*, 12(6), pp. 404–416. Available at: https://doi.org/10.2174/138920211797248556.

Zhou, Z. *et al.* (2012) 'Genome plasticity and systems evolution in Streptomyces', *BMC Bioinformatics*, 13(10), p. S8. Available at: https://doi.org/10.1186/1471-2105-13-S10-S8.

Zuker, M. (2003) 'Mfold web server for nucleic acid folding and hybridization prediction', *Nucleic Acids Research*, 31(13), pp. 3406–3415. Available at: https://doi.org/10.1093/nar/gkg595.