Engineering *Streptomyces clavuligerus* for Efficient Growth on Sustainable Carbon Sources

Anna Sofia Birke

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

Strathclyde Institute for Pharmacy and Biomedical Sciences

University of Strathclyde, Glasgow

September 2020

ABSTRACT

Streptomyces clavuligerus is the industrial producer of the β-lactamase inhibitor, clavulanic acid (CA). This compound is one of the components of the co-formulation generically known as co-amoxiclav. This pharmaceutical combines CA and amoxycillin and is industrially produced by GlaxoSmithKline (GSK), which markets the drug under the trade name Augmentin. The industrial-scale fermentation with *S. clavuligerus*, the process by which CA is produced by GSK, requires human food-grade and costly carbon sources as feedstocks. To make the industrial production of CA more sustainable and economical, GSK would like to source industrial by-products as feedstocks. Yet, *S. clavuligerus* has a limited carbon catabolic profile, its preferred carbon source being glycerol, which does not allow the use of alternative sources of carbon, such as glucose. *S. clavuligerus* was considered naturally glucose auxotrophic due to the lack of a glucose uptake system. In streptomycetes, glucose is transported into the cell via a glucose permease (GlcP) and subsequently phosphorylated by a glucokinase (Glk). Although genes encoding GlcP and Glk, *glcP* and *glk*, respectively, are present on the *S. clavuligerus* genome, expression of the former is too low to enable sufficient uptake of the hexose to support growth. Industrial production strain improvement traditionally involved random mutagenesis and subsequent rounds of selection of strains exhibiting desirable traits, such as improved antibiotic titres. Yet, with this approach, altered phenotypes cannot be linked to changes in genotypes. Therefore, the focus of this thesis was to gain an understanding of carbon utilisation of two industrial strains provided by GSK: SC2 and SC6. Subsequently, a targeted approach to improving glucose utilisation was taken by expressing streptomycete *glcP* and *glk* genes. This showed that, contrary to previous reports on CCR in a wildtype background, glucose triggers repression of CA biosynthesis in the earlier production strain. Furthermore, the strains were found to be de-regulated in their production of extracellular proteases, which might affect nutrient uptake and utilisation during fermentation. Protein-protein interaction between GlcP and Glk from *S. clavuligerus*, further, revealed a lack of interaction. GlcP and Glk are known to interact in *S. coelicolor*, a glucose utilising streptomycete. Overall, this thesis provides findings of aspects of the central carbon metabolism of an industrially relevant organism. Further, advantages of rational strain improvement are highlighted by heterologous expression of genes of interest and being able to directly link observed phenotypes.

Phylogenetic analyses of GlcP and Glk represented in the ActDES database contributed to work presented in the following publication:

Schniete, J. K., Selem-Mojica, N., Birke, A. S., Cruz-Morales, P., Hunter, I. S., Barona-Gomez, F. and Hoskisson, P. A. (2021) 'ActDES–a curated actinobacterial database for evolutionary studies', *Microbial genomics*, p. 498.

CONTENTS

I would like to thank Mama, Papa, Ellen and Alex (and Ella) for their love and support. Without you, I would not be the person I am today. Thank you for enabling me to follow my own (academic) path and having my back all along. Thank you, also, to my grandparents who have always been so supportive of me.

Last but not least, I would like to thank David for being the kind, caring and amazing partner that he is. Thank you for always being there for me, giving me space to grow and believing in me, especially when I did not believe in myself. I am so grateful I was able to experience the Glasgow chapter with you and I am excited to see what the Berlin chapter will bring.

"All right," said Deep Thought. "The Answer to the Great Question..." [...]

"Of Life, the Universe and Everything..." said Deep Thought. [...]

"Is..." said Deep Thought, and paused. [...]

"Forty-two," said Deep Thought, with infinite majesty and calm."

From The Hitchhiker's guide to the Galaxy by Douglas Adams

1 INTRODUCTION

Bacteria from the genus *Streptomyces* are filamentous soil bacteria known for their ability to produce bioactive specialised metabolites, such as antibiotics (Hopwood, 2007). One of the first clinically relevant antibiotics, however, was penicillin. This β-lactam antibiotic, discovered in 1928, is produced not by a streptomycete but by the fungus *penicillium ssp.*(Fleming, 1929). The first antibiotics with clinical relevance to be discovered from *Streptomyces* were actinomycin and streptomycin. In the years and decades that followed, a period commonly referred to as the golden era of antibiotic drug discovery, 70-80% of bioactive specialised compounds to be identified came from streptomycetes (Waksman *et al.*, 1941; Schatz *et al.*, 1944; Lyddiard *et al.*, 2016).

The treatment of bacterial infections with antibiotics is one of the greatest advances in the history of human and veterinary medicine. Yet, with the increase in antibiotic use, dissemination of pre-existing antibiotic resistance-conferring genes, as well as the emergence and spread of novel antibiotic resistances have increased in the clinic. For instance, penicillinase, a β-lactamase that cleaves and, thus, irreversibly inactivates penicillin, was discovered prior to the drug's introduction to the clinic in 1943 (Abraham *et al.*, 1940; Ventola, 2015). With β-lactam antibiotics being commonly administered, there has been a sharp rise in the cumulative number of β-lactamases identified since the 1990s (Davies *et al.*, 2010), although this effect might also partly be due to more and better molecular tools being available for identifying resistance-conferring genes compared to previous decades. With the number of (multi-)resistant bacterial strains being on the rise, among them human pathogens, such as *Escherichia coli*, *Klebsiella pneumoniae* and *Pseudomonas aeruginosa* and *Acinetobacter* species, bacterial infections have once again become a global health threat (World Health Organisation, WHO). It has been estimated that, if the global community fails to effectively tackle the issue of antimicrobial resistance (AMR, includes resistance to antibiotics) by 2050, 10 million lives might be at risk due to bacterial infections (O'Neill, 2016). However, interventions for closer and more thorough monitoring of resistance levels, awareness campaigns and research efforts lead by health organisations such as the WHO and the European Centre for Disease Prevention and Control (ECDC) have brought into effect national and international programmes to combat the crisis that appear to have slowed down

the previously explosive spread of resistance. The latest ECDC report, released in November 2019, presented data showing that changes in resistance percentages among Gram-negative bacteria were small between 2015 and 2018, though the overall level of resistance was (still) high (European Centre for Disease Prevention and Control, 2019).

One strategy implemented in the clinic to effectively treat infections caused by bacteria resistant to β-lactam antibiotics is co-formulating the β-lactam antibiotic amoxicillin with the β-lactamase inhibitor clavulanic acid (CA). CA was discovered in 1971 as a natural product of *S. clavuligerus* (Higgens *et al.*, 1971; Howarth *et al.*, 1976; Reading *et al.*, 1977) and is the focus of this project. The co-formulated product is generically known as co-amoxiclav. The pharmaceutical company GlaxoSmithKline (GSK) markets it under the trade name of Augmentin. Industrial CA production by GSK is a fermentation-based process that uses *S. clavuligerus* strains that have undergone the iterative in-house strain improvement programme that started back in the late 1970s. Methods of random mutagenesis applied as part of the programme are UV-radiation and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG) treatment. Subsequently, strains exhibiting desirable traits, such as increased CA production, were selected (personal communication with GSK). Given the clinical relevance of Augmentin and the ever-increasing pressures of resources, it is essential to make the CA production process as sustainable and economical as possible to be able to provide an affordable pharmaceutical product in the future. Therefore, GSK are moving towards using sustainable feedstocks, such as industrial by-products, for CA fermentations, as opposed to expensive and unsustainable, human food-grade feedstocks.

With the focus on carbon source utilisation, the aim of this thesis was to investigate the carbon utilisation profile of early CA production lineages, SC2 and SC6, and generate strains that are able to metabolise sustainably sourced glucose for the production of CA in an industrial setting. Unlike in the random GSK production strain improvement programme, a targeted approach of genetic modification was used.

1.1 RATIONAL STRAIN IMPROVEMENT OF *STREPTOMYCES* FOR SUSTAINABLE ANTIBIOTIC PRODUCTION

Traditionally, the improvement of *Streptomyces* strains for industrial antibiotic production has relied on random mutagenesis, for instance by exposing cells to UV radiation or treatment with NTG. Subsequently, strains were selected that exhibited a desired trait, such as improved antibiotic tires. Given that this strategy is inherently random, the introduction of mutations is untargeted, leading to the accumulation of mutations until relevant genes are hit. This can cause strains to become genetically unstable. Additionally, mutations might have pleiotropic effects, some of which might impact the overall viability of the cell (Lal *et al.*, 1996). Many pharmaceutical companies have engaged in large random mutagenesis and screening programmes to improve production and industrial performance of their strains. However, the development of strains with improved traits, have not been linked to genetic changes. The inability to link genotypes to phenotypes has hampered with researchers' abilities to rationally design antibiotic producing strains. Nowadays, the increase in genomic data and genetic tools available for *Streptomyces* render rational strain design possible, thus, allowing for industrial strains to be engineered that will allow the pharmaceutical industry to become sustainable. In 2015, the emission intensity of the pharmaceutical industry was three times that of the automobile industry, showing that the production of pharmaceuticals is far from being sustainable and, thus, not future-proof (Jackson *et al.*, 2018; Belkhir *et al.*, 2019).

In the upcoming years, the demand for sustainably produced pharmaceuticals and the pressure to keep costs low are going to increase as the changes to life on earth through human-induced climate change become apparent. Human-induced climate change is driven by greenhouse gas emissions, such as CO_2 emissions, which have increased 6-fold since the 1950s (Romm, 2018). The increased emissions of CO_2 compared to pre-industrial times is mainly caused by the usage of non-renewable resources, such as the fossil fuels, across industries. While the global communities' immense expenditure and demand for resources is currently not compatible with the earth's biocapacity, which is the amount of ecological resources that can be generated in a given time (compare earth overshoot day on 29th of December in 1970 to the 22nd of August in 2020, https://www.overshootday.org/), unequal access to available resources is a major issue that is perpetuated by the impact of climate change on land usage and crop yields (Iizumi *et al.*, 2018; Mbow *et al.*, 2019). Paradoxically, while 821 million people on earth are under- and/or malnourished (Mbow *et al.*, 2019), an

estimate of 88 million tonnes of food are wasted annually in Europe alone. That is the equivalent of 170 million tonnes of CO_2 that is wasted every year. Food waste, in general, is estimated to contribute approximately 8% of global greenhouse gas emissions (https://ec.europa.eu/food/safety/food_waste_en).

Climate change is also predicted to cause a rise in the number of local and global outbreaks of infectious diseases caused by viral and bacterial pathogens (Altizer *et al.*, 2013; Mbow *et al.*, 2019). Consequently, the demand for pharmaceuticals will increase, leaving the biotechnology industry facing the challenge of having to meet this growing demand by providing affordable products whilst producing them in a sustainable, ideally carbon neutral, manner. GSK, the industrial producer of CA, have released data showing that 48% of their carbon emissions originate from purchased goods and services (https://www.gsk.com/en-gb/responsibility/environment/carbon/#). By transitioning from unsustainable, expensive, and potentially human food-grade feedstocks to sustainably sourced feedstocks, such as industrial by-products and food waste, carbon emissions and the economic impact of CA production could be reduced. Switching from one feedstock to another is not always possible without genetically adapting the production strain. Rational strain improvement offers an approach to engineering production strains in a targeted, time and resource efficient manner that might play a crucial role in facilitating the switch to sustainable antibiotic production in the pharmaceutical industry. For a rational strain engineering approach to be taken, a thorough understanding of the producing organism's metabolism, in this case of *S. clavuligerus*, is required.

1.2 OVERVIEW OF THE GENUS *STREPTOMYCES*

Streptomyces are Gram positive, aerobic, spore-forming and mycelial actinomycetes that were first defined based on these morphological traits by Waksman and Henrici in 1943 (Waksman *et al.*, 1943). Aside from antibiotics, they produce a range of specialised metabolites with anticancer, antifungal, antiparasitic and immunosuppressive properties (Hopwood, 2007). They are ubiquitous in nature and colonise terrestrial and marine soil habitats where they pursue a saprophytic lifestyle and, thus, play an important role in decomposing organic matter (Embley *et al.*, 1994; Chater *et al.*, 2010).

Streptomyces is the type genus of the family *Streptomycetaceae* that is found within the order Actinomycetales, within the phylum Actinobacteria (Stackebrandt *et al.*, 1997). The genus was originally defined as a group of aerobic, spore-forming actinomycetes that pursue a saprophytic lifestyle in the soil (Waksman *et al.*, 1943). The *Streptomycetaceae* family contains 615 individual taxa, of which the majority are *Streptomyces*. Other genera in this family are *Kitasatospora* and *Streptacidiphilus*, which are genotypically and phenotypically similar to *Streptomyces* (Omura *et al.*, 1982; Zhang *et al.*, 1997; Kim *et al.*, 2003). In fact, the *Kitasatospora* genus was originally proposed to be synonymous with *Streptomyces*. This was later reversed based on the higher amount of meso-diaminopimelic acid in *Kitasatospora* whole-cell hydrolysates compared to that of *Streptomyces* (Wellington *et al.*, 1992; Zhang *et al.*, 1997). *Streptacidiphilus* are acidophilic *Streptomyceneae* that grow at pHs as low as 3.5 – 6.5, a unique characteristic of *Streptacidiphilus* compared to *Streptomyces* and *Kitasatospora* (Kim *et al.*, 2003). In the past, phylogenetic analyses of *Streptomyceneae*, such as those carried out as part of the international *Streptomyces* project (ISP), relied on identifying and classifying morphological characteristics, such as pigment colour, under standardised experimental conditions (Shirling *et al.*, 1966). Additionally, chemotaxonomic analyses of the cell wall composition were used to distinguish streptomycetes from other actinomycetes (Lechevalier *et al.*, 1970). Newer phylogenetic studies compare ribosomal gene sequences and ribosomal protein sequences, such as 16S rRNA and RpoB, respectively, to produce robust clades within highly similar species (Han *et al.*, 2012).

Prior to advances in the fields of chemotaxonomy and molecular phylogeny, one of the unique characteristics that sets streptomycetes apart from other bacteria, which is also the reason why they were initially classified as fungi, is their complex life cycle (presented in more detail in section 1.3, Figure 1-1). However, chemotaxonomic studies of the cell wall composition and ultrastructure later confirmed their position among the true bacteria (Cummins *et al.*, 1958). Their cell wall is typical for Gram-positive bacteria and contains LL-diaminopimelic acid (LL-DAP) and glycine, which is characteristic for cell wall type I (Cummins *et al.*, 1958; Lechevalier *et al.*, 1970).

Another unique feature of streptomycetes is their genomes. These consist of a large, when compared to the 4.6 Mb *E. coli* genome, linear chromosome of around 6.2 – 11.9 Mb and can

also encompass several plasmids. Streptomycete genomes also have high G+C % contents that range from 61% to 80% (Van Keulen, 2014). The genome of the model organism *Streptomyces. Coelicolor*, for instance, has a 8.6 MB genome with a 72% G+C content (Bentley *et al.*, 2002). Additionally, there are two extrachromosomal plasmids present in this species, SCP1 and SC2, the former of which is linear with a size of 365 kb and the latter of which is 31 kb with a circular configuration (Haug *et al.*, 2003; Bentley *et al.*, 2004). The *S. clavuligerus* genome is comprised of a 6.8 Mb chromosome and four plasmids, pSCL1 – pSCL4, with an overall G+C % content of 72% (Wu *et al.*, 1993; Netolitzky *et al.*, 1995; Medema *et al.*, 2010). The giant linear plasmids, pSCL4, of the ATCC 27064 strain was analysed in more detail and was reported to contain numerous specialised metabolite gene clusters for the synthesis of polyketides, non-ribosomal peptides, terpenes and β-lactams (Medema *et al.*, 2010). Overall, *S. clavuligerus* produces a variety of antibiotics, including CA, cephamycin C, cephalosporin and holomycin (Higgens *et al.*, 1971; Kenig *et al.*, 1979).

1.3 *STREPTOMYCES* ECOLOGY AND LIFE CYCLE

The soil is a challenging habitat characterised by a diverse community comprised of microbes, animals, and plants. The levels of nutrients, such as carbon nitrogen, and phosphorus, fluctuate. In order to thrive in the soil, streptomycetes produce and secrete a range of compounds, among which are polymer-degrading exoenzymes, proteases and bioactive specialised metabolites (natural products) (Chater *et al.*, 2010; van der Meij *et al.*, 2017). To adapt to changes in nutrient availability, streptomycetes undergo a complex life cycle comprised of several stages that are morphologically and physiologically distinct (Figure 1-1). Development starts with a dormant spore, from which germ tubes emerge once environmental conditions are suitable, a process known as germination. Subsequently, vegetative hyphae form the vegetative mycelia, which increase the range in which *Streptomyces* can access nutrients.
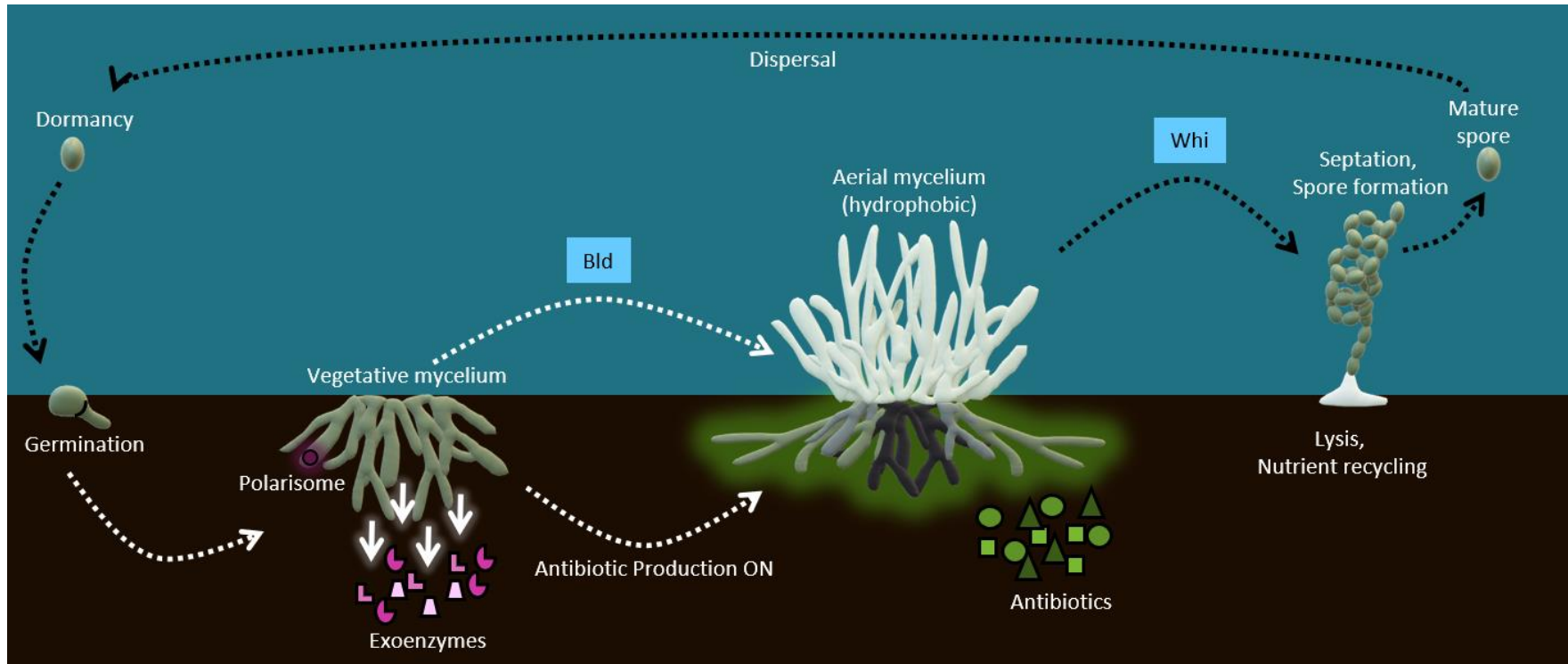
Figure 1-1: Schematic of the streptomycete life cycle and key regulators involved. During germination, germ tubes emerge from the dormant spore. Hyphae that grow into the substrate by tip extension branch laterally, thus, creating the substrate/vegetative mycelium (green). The polarisome (purple) plays a central role in this process. Substrate hyphae secrete exoenzymes (purple symbols) to degrade insoluble polymers. Once nutrients become scarce, a morphological switch occurs that is tightly connected with physiological changes (Bld cascade), such as the onset of antibiotic production. Proteases secreted into the substrate breakdown the vegetative mycelium into building blocks, that are recycled for the formation of aerial hyphae. Aerial hyphae are surrounded by a hydrophobic sheath that enables growth out of the aqueous substrate. Subsequently, they begin to form spore chains, a process controlled by the *whi* genes. Individual spores are dissemination through the air.

7

Building blocks from vegetative mycelium are degraded and recycled to form aerial hyphae. At this stage, production of specialised metabolites, such as antibiotics, is initiated. Individual aerial hyphae will eventually form spore chains, from which individual spores are dispersed through the air (Flärdh *et al.*, 2009).

1.3.1 THE DORMANT SPORE

At the start of the life cycle is the dormant spore, which is generally thought to be metabolically inactive. Yet, glucose metabolism via the pentose phosphate pathway and glycolysis occurs through the catabolism of trehalose stored inside the spore. Trehalose is composed of two α-α-1-1 glycosidic linked glucose molecules and is broken down into glucose by the glycosidic hydrolase trehalase that colocalises with trehalose in the spore cytoplasm (Salas *et al.*, 1984; McBride *et al.*, 1987, 1990). Spores also contain active ribosomes and a pool of messenger RNA (mRNA), which are preserved in the trehalose to facilitate the biosynthesis of proteins required for germination (Martín *et al.*, 1986; Mikulík *et al.*, 2002). The wall of the spore consists of an inner and an outer wall, these are approximately 12 μm and 30 μm thick, respectively (Glauert *et al.*, 1961). The structure of the spore wall is defined by the process of cross-wall formation of aerial hyphae and septation that occur during sporulation, the process that proceeds germination (Wildermuth *et al.*, 1970). Maintenance of the spore wall is an essential part of dormancy, as was demonstrated by deletion of the structural cell wall component, NepA. In the absence of this highly insoluble protein, spores were shown to germinate under poor conditions (Dalton *et al.*, 2007; De Jong *et al.*, 2009). Interestingly, a *ΔnepA* strain of *S. coelicolor*, so lacking the structural cell wall component of the dormant spore, produced a bald phenotype, referring to the absence of aerial hyphae, on minimal medium supplemented with glucose as carbon source (Dalton *et al.*, 2007).

1.3.2 GERMINATION: DARKENING, SWELLING, GERM TUBE EMERGENCE

Germination of a spore marks the beginning of the vegetative phase of the life cycle. The onset of germination depends on the intracellular Ca^{2+} concentration, which is maintained by the EF-hand calcium-binding protein, CabC. Its absence results in premature germination of spores at the spore chain. The phenotype produced by the *cabC* mutant is induced in

wildtype (WT) *S. coelicolor* by cultivation in the presence of 10 mM CaCl$_2$. This shows that premature germination is caused by an increase in intracellular Ca^{2+} (Wang *et al.*, 2008).

There are three distinct phases involved in germination: darkening, swelling of the spore, at which point protein biosynthesis increases in rate, and subsequent emergence of a germ tube (Hardisson *et al.*, 1978). No exogenous nutrients are required for the phase of darkening, which refers to the transition from bright to dark under a phase contrast microscope. This is presumably due to consumption of the internally stored trehalose (McBride *et al.*, 1987). The darkening phase is associated with a loss in spore hydrophobicity. Given that germination requires an aqueous environment, the loss of hydrophobicity ultimately leads to water influx, causing the spore to swell. The swelling phase requires an exogenous source of carbon and both carbon and nitrogen sources are necessary for the emergence of the germ tube to occur as carbon flux trough the pentose phosphate and glycolytic pathway increases (Hardisson *et al.*, 1978; Salas *et al.*, 1984). The process is, further, dependent on the synthesis of cyclic adenosine 3`, 5`-monophosphate (cAMP), of which intracellular levels peak during germination (Hardisson *et al.*, 1978; Süsstrunk *et al.*, 1998). The second messenger cAMP is synthesised by adenylate cyclase, Cya, the absence of which results in germination initiation comparable to that of the wildtype but with a severe delay of germ tube emergence. This phenotype was reversed by exogenous cAMP (Süsstrunk *et al.*, 1998). The cAMP-receptor, CRP, also plays an important role in germination. A *crp* mutant developed germ tubes sporadically and less often than the wildtype, a phenotype that could not be suppressed by exogenous cAMP (Süsstrunk *et al.*, 1998; Derouaux *et al.*, 2004; Piette *et al.*, 2005).

Either one or two germ tubes grow out of the spore and elongate by apical tip extension, giving rise to vegetative hyphae that branch laterally (Glauert *et al.*, 1961; Locci *et al.*, 1980; Gray *et al.*, 1990; Daniel *et al.*, 2003). SsgA, a member of the SsgA-like protein (SALP) family that encompasses proteins that coordinate cell division and sporulation, is thought to be involved in determining the site of germ tube emergence, as a SsgA-GFP fusion protein colocalises with the germ tube (van Wezel *et al.*, 2000; Noens *et al.*, 2005, 2007). While DNA synthesis does not take place during dormancy, germination coincides with the onset of DNA synthesis. This is carried out by a single replisome, the DNA replication machinery, located in the spore. Once a critical hyphae length is reached, replisomes form proximal to the tips of

the hyphae. As the hyphae continue to extend, the replisomes follow the tip (Wolánski *et al.*, 2011). The number of replisomes double/triple in germinating spores compared to dormant spores and further double inside the germ tubes (Ruban-Ośmiałowska *et al.*, 2006).

1.3.3 DIGGING DEEP: VEGETATIVE MYCELIA

Streptomycetes develop a network of hyphae that forage into the substrate to scavenge for nutrients (see 1.3.3.1). The hyphae that emerge grow by tip extension and branch laterally, resulting in two new hyphal tips, which also grow by tip extension (Flärdh *et al.*, 2009; McCormick *et al.*, 2012). This is a form of polarised growth whereby new peptidoglycan cell walls are established near the hyphal tip. New cell poles must also be established after branching, which requires the involvement of the coiled-coil protein, DivIVA. This protein was discovered as a homologue of DivIVA from *Bacillus subtillis* (Cha *et al.*, 1997; Edwards *et al.*, 1997; Daniel *et al.*, 2003; Flärdh, 2003). In *S. coelicolor*, partial deletion of *divIVA* or reduction of DivIVA results in irregularly shaped hyphae as a result of the absence of polar growth. Overexpression of DivIVA results in hyperbranching (Flärdh, 2003). Like the DivIVA homologue from *B. subtillis*, the protein co-localises with the cell poles and is also found on the inside of the lateral hyphal wall. Such DivIVA foci mark where new hyphae emerge in the branching process. DivIVA foci also co-localise with the cell wall synthesis machinery, with which it communicates bidirectionally, a process which involves the Ser/Thr kinase, AfsK (Hempel *et al.*, 2008, 2012). At the newly determined cell poles, DivIVA and another coiled-coil protein, Scy (***S****treptomyces* **cy**toskeletal element), which is thought to stabilise DivIVA, form a multiprotein complex, known as the polarisome or tip organising centre (TIPOC) (Walshaw *et al.*, 2010; Hempel *et al.*, 2012; Holmes *et al.*, 2013). Scy directly interacts with another protein found in the hyphal extension area, the filament-forming protein, FilP (Fuchino *et al.*, 2013; Holmes *et al.*, 2013). The intermediate filaments provided by FilP during polarised growth have been proposed to act as a stress-bearing structure while the tip undergoes extension (Fuchino *et al.*, 2013). Recently, evidence has been presented that FilP also affects the size and position of DivIVA and, thus, of the polarisome. This, is turn, might affect the shape of the hypha under certain growth conditions (Fröjd *et al.*, 2019).

Enzymes that are secreted into the environment include cellulases, lignocellulases, chitin-hydrolysing enzymes and xylanases (Schrempf, 2001; Chater *et al.*, 2010). Cellulose and chitin are the most and second most abundant polysaccharides on earth and, when broken down, they represent an important source of carbon for streptomycetes. Cellulose and lignin promote growth of high-density colonies (Schlatter *et al.*, 2009). The type of enzymes secreted can increase the adaptability of *Streptomyces* species to their ecological niche. The model organism *S. coelicolor* A3(2), for instance, is known to produce an agar-degrading enzyme, agarase, which allows colonies to grow on substrates containing agar as the sole carbon source (Stanier, 1942; Hodgson *et al.*, 1981; Buttner *et al.*, 1987).

Streptomyces also secrete several types of proteases, some of which are involved in providing proteinaceous nitrogen sources for growth. *S. clavuligerus*, the producer of CA, also produces proteases in response to nutrient depletion and in an ammonium-dependent manner (Aharonowitz, 1979; Bascarán *et al.*, 1990; Porto *et al.*, 1996). Other proteases, such as the trypsin-like protease, are required for the degradation of old mycelium that has stopped growing and is broken down to serve as a nutrient source for aerial hyphae (Manteca *et al.*, 2005). Younger and still growing substrate mycelium are protected from proteolytic activity by the presence of protease inhibitors, for instance, leupeptin produced by *S. exfoliatus* SMF13 (Kim *et al.*, 1995; Chater *et al.*, 2010). *Streptomyces* are used industrially to produce proteases. For instance, *S. griseus* is the industrial producer of pronase, a protease cocktail with versatile applications (Wi *et al.*, 2006; Razzaq *et al.*, 2019).

1.3.4   The Switch from vegetative to aerial mycelium: Antibiotic Production

The switch from vegetative to reproductive aerial mycelium is a complex process, triggered by changes in environmental conditions. This transition period coincides with the onset of antibiotic production (Bibb, 2005; Hopwood, 2007). Although antibiotic production is non-essential for the survival of the cell, antibiotics serve multiple different purposes in nature, which can be contrasting to the use of antibiotics in modern medicine (Hopwood, 2007). Given the bioactivity of antibiotics, it was initially thought that their role in the soil was to kill off or inhibit the growth of competing microorganisms (Waksman, 1961). However, it has since become apparent that antibiotics, for the most part, occur at sub-inhibitory

concentrations in the soil (Davies, 2006; Yim *et al.*, 2007; Aminov, 2009). Studies have shown that sub-lethal concentrations of antibiotics alter gene expression profiles of non-producers, such as of *Salmonella typhimurium* and *Streptococcus pneumoniae*, *in vitro* (Goh *et al.*, 2002; Ng *et al.*, 2003; Yim *et al.*, 2006). The different effect of or response to a molecule at a low vs high concentration, whereby the low dose is beneficial and the high dose is detrimental, is known as hormesis (Calabrese *et al.*, 2005; Davies, 2006; Ray *et al.*, 2014). Low doses of protein biosynthesis inhibiting antibiotics, including streptomycin and rifampicin, can shift nutrient use and overlap in co-cultures, thereby (re-)defining nutritional boundaries within a niche occupied by several bacterial strains (Jauri *et al.*, 2013). In contrast, when streptomycetes enter symbiosis with animals or plants, where they can colonise both the rhizosphere and endosphere, the production of bioactive specialised metabolites aid in protecting their symbiotic partners from invasion of insects, microbes, or fungi (Chater *et al.*, 2010; Bulgarelli *et al.*, 2013). A strain of *S. albus*, for instance, that was isolated from drunken horse grass, *Achnatherum inebrians*, produces a natural product with insecticidal activity (Shi *et al.*, 2013). More recently, *S. formicae* was isolated from the African plant-ant *Tetraponera*. This species produces a range of natural products known as formicamycins, which are active against Gram-positive bacteria and fungi (Qin *et al.*, 2017).

1.3.5 GLOBAL REGULATION OF ANTIBIOTIC BIOSYNTHESIS DURING DEVELOPMENT

1.3.5.1 THE A-FACTOR CASCADE, ADPA AND THE AFSKR SYSTEM

Control of morphological development and specialised metabolism is exerted by a range of global regulatory pathways, one of which is the butanolide system. γ-butyrolactones are "bacterial hormones", which are diffusible and effective even at very low concentrations (Dyson 2011). One of the best studied butanolide systems is the A-factor cascade in the streptomycin-producing *S. griseus* (Horinouchi *et al.*, 1992). A-factor binds a homodimer of the repressor-type receptor, ArpA, which controls expression of genes involved in later stages of development and specialised metabolism (Hara *et al.*, 1982; Onaka *et al.*, 1995; Onaka and Horinouchi, 1997). Binding of A-factor to ArpA causes a conformational change in the N-terminal DNA binding domain of the repressor protein, thus, resulting in dissociation of ArpA from the previously bound DNA. (Onaka and Horinouchi, 1997; Onaka, Sugiyama, *et al.*, 1997).

In *S. coelicolor*, CprA and CprB, two ArpA homologues that share 35% of their amino acid sequence with ArpA, have been identified. The former exerts a positive influence on specialised metabolite production and sporulation as a *cprA* disruption mutant produces less actinorhodin and undecylprodigiosin and sporulation was delayed. The role of CprB, in contrast, appears to be that of a negative regulator of specialised metabolism. A *cprB* disruption mutant overproduced actinorhodin and sporulated earlier (Onaka *et al.*, 1998). ArpA influences antibiotic biosynthesis through other factors, one of which is AdpA, the A-factor responsive transcriptional activator. ArpA recognises palindromic recognition sequences, knows as ARE box, one of which is present in the promoter of *adpA* (Onaka and Horinouchi, 1997; Ohnishi *et al.*, 1999, 2005). AdpA, in turn, activates transcription of antibiotic biosynthetic regulators, such as *strR* in *S. griseus* and the positive regulators of CA biosynthesis in *S. clavuligerus*, *ccaR* and *claR* (see 1.4.2) (Ohnishi *et al.*, 1999; López-García *et al.*, 2010; Pinilla *et al.*, 2019).

The expression of genes involved in A-factor biosynthesis are under the control of AfsR, which is part of the AfsK-AfsR system, a serine/threonine phosphorylation system involved in regulation development and specialised metabolism (Umeyama *et al.*, 2002). AfsR is a member of the **S**treptomyces **a**ntibiotic **r**egulatory **p**rotein (SARP) family (Wietzorrek *et al.*, 1997; Bibb, 2005). AfsK is the kinase that phosphorylates DivIVA mentioned previously (see 1.3.3) and it also phosphorylates AfsR (Hong *et al.*, 1991). Heterologous expression of an additional copy of *afsR* has been shown to increase antibiotic production in several *Streptomyces*, including *S. lividans*, *S. peucetius* and *S. clavuligerus*. CA was the antibiotic that was enhanced in the latter species (Parajuli *et al.*, 2005). Given the positive control of expression of *ccaR* and *claR* (see 1.4.2) by AdpA, the increase in CA in the *afsR* expression strain might be the result if elevated *adpA* expression (López-García *et al.*, 2010; Pinilla *et al.*, 2019).

To date, one autoregulatory γ-butyrolactones receptor has been identified in *S. clavuligerus* (Kim *et al.*, 2004; Santamarta *et al.*, 2005). Deletion of the gene encoding the γ-butyrolactones receptor, which has the two names *scaR* and *brp*, leads to an overproduction of both cephamycin C and CA supporting the idea that Brp functions as a repressor in the absence of its ligand. The Brp recognition sequence is located upstream of the *ccaR* gene

(Santamarta *et al.*, 2005). Expression of *brp*, in turn, is under the control of AreB, an IcIR-type transcription regulator that also binds upstream of *ccaR*, although this requires an unidentified small molecule, and influences expression of leucine biosynthetic genes (Santamarta *et al.*, 2005, 2007).

1.3.5.2   STRINGENT RESPONSE

The onset of antibiotic production is also triggered by nutrient depletion, such as nitrogen and phosphorus limitation (Challis *et al.*, 2003). A decrease in growth or arrest thereof is an important step in initiation of antibiotic biosynthesis, which is likely to involve guanosine 5`-diphosphate 3`-diphosphate, ppGpp. Synthesis of ppGpp is carried out by ppGpp synthetase RelA and the (p)ppGpp synthetase/hydrolase SpoT (Haseltine *et al.*, 1973; Gentry *et al.*, 1996). During nitrogen limitation or amino acid starvation, (p)ppGpp inhibits rRNA synthesis as a way of conserving energy, by binding to the β-subunit of the RNA polymerase. This process is known as stringent response (Chatterji *et al.*, 1998). In *S. coelicolor* A3(2), RelA is required for undecylprodigiosin and actinorhodin production under conditions of nitrogen limitation (Chakraburtty *et al.*, 1997). Moreover, when ppGpp synthesis is induced independently from the stringent response, for instance by heterologous expression of a truncated version of *relA*, expression of pathway-specific regulators of antibiotic biosynthesis gene clusters, such as the actinorhodin gene cluster *actII-ORFIV*, is increased in *S. coelicolor* (Hesketh *et al.*, 2001, 2007).

A potential relationship between the stringent response and specialised metabolite production in *S. clavuligerus* has been investigated and conflicting data have been presented over the years. While Bascarán *et al.* (1991) were the first to investigate the effect of ppGpp levels on cephalosporin biosynthesis and reported no obligatory relationship between the processes, cephamycin C and CA production were reported to be abolished in the RelA homologue knockout mutants of *S. clavuligerus* ATCC 27064, $\Delta relA$ and Δrsh (Bascarán *et al.*, 1991; Jin *et al.*, 2004). In contrast to these findings, it has also been reported that the $\Delta relA$ mutant overproduces these compounds, suggesting, for the first time, that (p)ppGpp might act as a negative regulator of antibiotic production. The discrepancies between these findings might be due to differences in parental strains and growth media (Gomez-Escribano *et al.*,

2008). Thus, the role of RelA in controlling antibiotic production in *S. clavuligerus* this remains to be clarified.

1.3.5.3   PHOSPHATE LIMITATION

Phosphate limitation is another external factor that triggers the onset of antibiotic production via the PhoR-PhoP two-component system, first discovered to negatively regulate actinorhodin and undecylprodigiosin production in *S. lividans* (Sola-Landa *et al.*, 2003). PhoR is activated under phosphate deprivation conditions, the histidine kinase PhoR autophosphorylates and subsequently transfers the phosphoryl group onto the response regulator PhoP. Subsequently, PhoP recognises and binds a consensus sequence, known as a *pho* box, thereby regulating transcription (Martín, 2004). In *S. coelicolor*, *pho* boxes have been detected in over 100 different gene clusters with the consensus sequence being comprised of a varying number of 11 nucleotide direct repeats (Sola-Landa *et al.*, 2008). *Pho* boxes have also been detected in 31 *S. clavuligerus* DSM 738 genes, but not in any of the CA biosynthetic genes (Salehghamari *et al.*, 2012). Although cephalosporin, cephamycin C and CA biosynthetic genes are known to be the subject of repression by inorganic phosphate (Aharonowitz *et al.*, 1977; Lübbe *et al.*, 1985; Lebrihi *et al.*, 1987; Jhang *et al.*, 1989), the underlying mechanism remain to be elucidated. An overlap of PhoP and AdpA binding sites of antibiotic biosynthetic gene promoters has been reported, showing multiple signals are integrated in regulation of antibiotic production (Zheng *et al.*, 2019).

1.3.5.4   CARBON CONTROL

Carbon control over antibiotic production in the form of carbon catabolite repression, CCR, is well documented (Sanchez *et al.*, 2010). In non-antibiotic producing bacteria, CCR ensures sequential utilisation of a preferred carbon source before an alternative carbon source by repressing the expression of genes required for uptake and catabolism of the alternative carbon source (Stülke *et al.*, 1999; Brückner *et al.*, 2002). This is also the case in *Streptomyces*, in which often glucose represses a host of uptake systems required for uptake and catabolism of alternative carbon sources, such as glycerol, agarose and maltose (Hodgson, 1982). In Gram-negative organisms and low GC Gram-positive bacteria, CCR is mediated by components of the glucose-transporting phosphotransferase system (PTS), with HPr phosphorylation being the key step (Saier, 1989; Reizer *et al.*, 1993). However, glucose is not

transported via a PTS in *Streptomyces* and a HPr-deficient strain of *S. coelicolor* was still able to repress expression of *glpK*, the glycerol kinase gene, in the presence of glucose (Parche *et al.*, 1999; Nothaft, Dresel, *et al.*, 2003). Instead, the glucokinase Glk plays a central role in mediating glucose dependent CCR, which also triggers repression of antibiotic production. Glk is one of the two components that comprise the GlcP-Glk system, the major streptomycete glucose uptake system (van Wezel *et al.*, 2005). The *S. clavuligerus* GlcP-Glk system is the focus of this thesis, therefore glucose uptake and the involvement of Glk in CCR are described in greater detail in section 1.3.11.

Glucose is not the only carbon source to trigger a CCR response in streptomycetes. Glycerol, mannose, fructose and lactose repress antibiotic production in various *Streptomyces* (Demain, 1989). In *S. clavuligerus*, for instance, glycerol, which is the preferred carbon source of this species, and maltose have been identified as triggers of CCR-mediated repression of cephamycin C production. Cephalosporin biosynthesis is repressed in the presence of glycerol and starch (Aharonowitz *et al.*, 1978; Hu *et al.*, 1984; Lebrihi *et al.*, 1988).

1.3.6 BLD REGULATORY CASCADE

Growth of the aerial hyphae is initiated once nutrients in the environment become limited. Important regulators of this process are encoded by *bld* genes, the absence of which results in a lack of aerial hyphae, thus, resulting in a bald (bld) phenotype (Merrick, 1976). To date, 12 *bld* genes, *bldA* to *bldN*, and their gene products have been identified and characterised (Claessen *et al.*, 2006). To gain an understanding of the hierarchy that defines the Bld cascade, complementation experiments were carried out, whereby two strains, each one deficient in a different *bld* gene, were grown in proximity to one another. Certain pairs of strains were able to complement each other and, thus, form aerial hyphae. Such complementation experiments and characterisation of the *bld* gene products has allowed establishment of the following hierarchy: [*bldJ*[1]]<[*bldK/bldL*]<[*bldA/bldH*]<[*bldG*]<[*bldC*]<[*bldD*] (Willey *et al.*, 1993; Nodwell *et al.*, 1996, 1998, 1999).

---

[1] Formerly known as *bld261*

At the top of the cascade is BldD, a global regulator and transcription factor that autoregulates negatively by binding to its own promoter as a homodimer (Elliot *et al.*, 1999; Elliot, Locke, *et al.*, 2003). BldD, for which the crystal structure of the DNA binding domain has been solved, represses expression of developmental genes (Elliot *et al.*, 2001; Kim *et al.*, 2006). This requires the second messenger, 3`, 5`-cyclic diguanylic acid (c-di-GMP), which facilitates dimerisation of BldD (Tschowri *et al.*, 2014). The BldD regulon encompasses approximately 167 genes, among which are *bldA*, *adpA* (*bldH*), the gene product of which is involved in regulating specialised metabolism (see 1.3.5), *bldC*, *bldM*, *bldN* and the aforementioned *ssgA* (see 1.3.2).

bldC, the second *bld* gene in the hierarchy, encodes a small DNA-binding protein related to the DNA-binding domain of the MerR family of transcription regulators. It has been shown to be required for antibiotic biosynthesis in *S. coelicolor* (Willey *et al.*, 1993; Hunt *et al.*, 2005). One level below *bldC* is *bldG*, which encodes an anti-anti-sigma factor (Champness, 1988; Bignell *et al.*, 2000). More recently, a direct link between *bldG*, and CcaR, one of the cluster-situated positive regulators of cephamycin C and CA biosynthesis in *S. clavuligerus*, was demonstrated (Pérez-Llarena *et al.*, 1997; Bignell *et al.*, 2005). The presence of *bldG* transcripts were necessary for *ccaR* expression. Like in *S. coelicolor*, there are both monocistronic and polycistronic transcripts of *bldG* and *orf3* in *S. clavuligerus* (Bignell *et al.*, 2000, 2005). The *bld* gene one step down from *bldG*, is *bldA*, which encodes the only tRNA in streptomycetes that recognises the rare UUA codon (leucine) (Merrick, 1976; Lawlor *et al.*, 1987). A gene that contains the TTA codon is *adpA* and expression of this gene, thus, requires *bldA*. In a positive feedback loop, *bldA* expression is dependent on the transcriptional activation of AdpA (Higo *et al.*, 2011). AdpA was discussed previously as a regulator of antibiotic production (see 1.3.5), its's regulon, however, encompasses not only regulators of antibiotic biosynthetic pathways but also a genes involved in development and protease production (see 1.3.7). BldKL is an oligopeptide importer and belongs to the family of **A**TP-**b**inding **c**assette (ABC) transporters and BldJ is thought to be involved in producing a signal, potentially an oligopeptide, that is then imported by BldKL (Nodwell *et al.*, 1996).

The bald phenotype exhibited by all *bld* mutants, except for *bldB* mutants, is dependent on the carbon source available in the growth medium (Pope *et al.*, 1996). This was used as a

criterium in early studies to distinguish between different classes of *bld* mutants (Merrick, 1976). Further investigation of the mutant carbon metabolisms revealed a lack of carbon catabolite repression (CCR), as was demonstrated by analysing the expression levels of the *gal1P* gene, which are normally glucose-sensitive and galactose-dependent. In five of the six tested mutant strains, *bldA*, *bldC*, *bldD*, *bldG* and *bldH*, *gal1P* expression levels were high in the presence of glucose when no expression occurred in the wildtype (Pope *et al.*, 1996).

1.3.7 Aerial Mycelium

The Bld cascade regulates the formation of the hydrophobic sheath that surrounds aerial hyphae. This sheath, also known as the rodlet layer, is necessary for hyphae to grow out of the aqueous substrate (McCormick *et al.*, 2012). It is composed of three different types of surface proteins: the lantibiotic-like spore-associated peptide SapB (AmfS in *S. griseus*), the chaplins and the rodlins (Guijarro *et al.*, 1988; Claessen *et al.*, 2002, 2003; Elliot *et al.*, 2003), with the interaction of the chaplins and rodlins being essential for the formation of fibrils that compose the insoluble layer found on spores (Claessen *et al.*, 2004). Although *bld* mutants are blocked in aerial mycelium production and, thus, do not sporulate, mutant phenotypes can be rescued by external supplementation with the SapB biosurfactant. This results in substrate mycelium growing in a fashion that resembles aerial hyphal growth but sporulation does not occur (Willey *et al.*, 1991, 1993, 2006). AdpA plays a central role at this stage of development and is also required for the formation of the hydrophobic sheath by controlling transcription of *ramR* (*amfR* in *S. griseus*). RamR is the transcriptional activator of the *ramCSAB* operon (*amfTSBA* in *S. griseus*) (Ueda *et al.*, 2002, 2005; Wolański *et al.*, 2011). Deletion of *adpA* in *S. clavuligerus* produces a conditional phenotype characterised by sparse aerial mycelium and a complete lack of sporulation (López-García *et al.*, 2010).

The secretion of proteases to locally degrade old vegetative mycelium, as well as the production and secretion of protease inhibitors to protect new mycelium are important steps during the transition period from vegetative to aerial mycelium. Expression of genes encoding proteases and protease inhibitors are also regulated by AdpA (Akanuma *et al.*, 2009; Wolański *et al.*, 2011).

1.3.8 SPORULATION

Once aerial hyphae have grown out of the substrate, the apical part of an aerial hypha will undergo multiple rounds of cell division producing unigenomic pre-spore compartments (Kieser *et al.*, 2000). Prior to the formation of pre-spore compartments, however, the formation of a cross wall, known as basal septum, occurs that divides the part of the hypha that will undergo septation and the subapical part of the hypha stem (Wildermuth *et al.*, 1970; McCormick *et al.*, 2012). In *S. coelicolor*, sporulation is governed by a set of regulatory genes called *whi* genes. The genes were named after their corresponding mutant phenotypes that produce a white coloured pigment instead of the grey pigment characteristic for *S. coelicolor* spores. *WhiG* encodes the RNA polymerase sigma factor σ^{whiG}, the expression of which is controlled by BldD binding to its promoter. σ^{whiG}, in turn, activates the expression of other *whi* genes required for the onset of sporulation, namely *whiA*, *whiB*, *whiG*, *whiH* and *whiJ*. (Chater *et al.*, 1989; Elliot *et al.*, 2001). Consequently, *sigF* and *ftsz* encoding σ^{F} and FtsZ, are expressed. Large amounts of FtsZ protein assemble to form a ring-like structure, called the Z-ring, inside a cell that is about to divide. This is a requirement for cell division and, thus, for sporulation septation to occur in the first place (Chater, 2001). Once spores have been formed, they might be dispersed by movement of animals living in the soil, such as worms and springtails, or they might be distributed aerially attached to the surface of soil particles (Lloyd, 1969).

1.3.9 CENTRAL CARBON AND SPECIALISED METABOLISM

Carbon metabolism is comprised of catabolic and anabolic reactions that provide cellular energy in the form of reducing agents, as well as building blocks required for growth and biomass production. *Streptomyces* metabolic pathways are often divided into primary or central carbon and secondary metabolism. Recently, updated definitions for primary, secondary, and specialised metabolism have been suggested, allowing the distinction of specialised metabolism as a sub-category of secondary metabolism. Primary metabolic pathways have been defined as evolutionary conservative metabolic pathways that are maintained within a lineage as their contribution is essential for autonomous cellular growth and survival (Chevrette *et al.*, 2020). Primary metabolism is, henceforth, referred to as central carbon metabolism. Metabolic robustness of central carbon metabolic pathways can be

19

increased by gene family expansion events of enzymes (Schniete *et al.*, 2018; Chevrette *et al.*, 2020). Intermediates of central carbon metabolism serve as precursors for further metabolic reactions, such as the biosynthesis of antibiotics during secondary metabolism (Chevrette *et al.*, 2020). Specialised metabolism is a form of secondary metabolism that results in the formation of specialised metabolites or natural products, the niche-specific ecological functions of which are known. In contrast to central carbon metabolism, secondary metabolism evolves and adapts through enzyme promiscuity and acquisition. Although secondary metabolism can increase a species' fitness, it is non-essential (Aharonowitz *et al.*, 1978; Chevrette *et al.*, 2020).

Increasing metabolic intermediates from central carbon metabolism for improved antibiotic production by streptomycetes is a rational strain improvement strategy (for example Li and Townsend, 2006; Ryu *et al.*, 2006). Therefore, glycolysis, the pentose phosphate pathway, the tricarboxylic acid (TCA) cycle, anaplerotic reactions, as well as individual enzymes involved in catalysing central carbon metabolic reactions have been studied extensively over the past years (Hodgson, 2000; Schniete *et al.*, 2018; Llamas-Ramírez *et al.*, 2020). Central carbon metabolism in *S. clavuligerus* has received considerable attention due to the strain being naturally unable to utilise glucose as a carbon source (Aharonowitz *et al.*, 1978; García Domínguez *et al.*, 1988; Garcia-Dominguez *et al.*, 1989; Pérez-Redondo *et al.*, 2010). Beyond studies of carbon uptake, however, only a small number of studies have focused on identifying targets within central carbon metabolism for rational strain improvement of CA producing strains. The following sections focus on the current understanding of central carbon metabolic pathways in *S. clavuligerus* (Figure 1-2) with regards to CA production and present approaches taken to increase precursor availability by modifying targets within these pathways.

1.3.10   CENTRAL CARBON METABOLIC PATHWAYS IN *S. CLAVULIGERUS* AND TARGETS OF RATIONAL STRAIN IMPROVEMENT

In *S. clavuligerus*, glyceraldehyde-3-P (GAP) and arginine are the precursors for CA biosynthesis (Figure 1-2) (Townsend *et al.*, 1985; Valentine *et al.*, 1993; Khaleeli *et al.*, 1999). GAP is an intermediate of glycolysis, a pathway central to carbon metabolism.
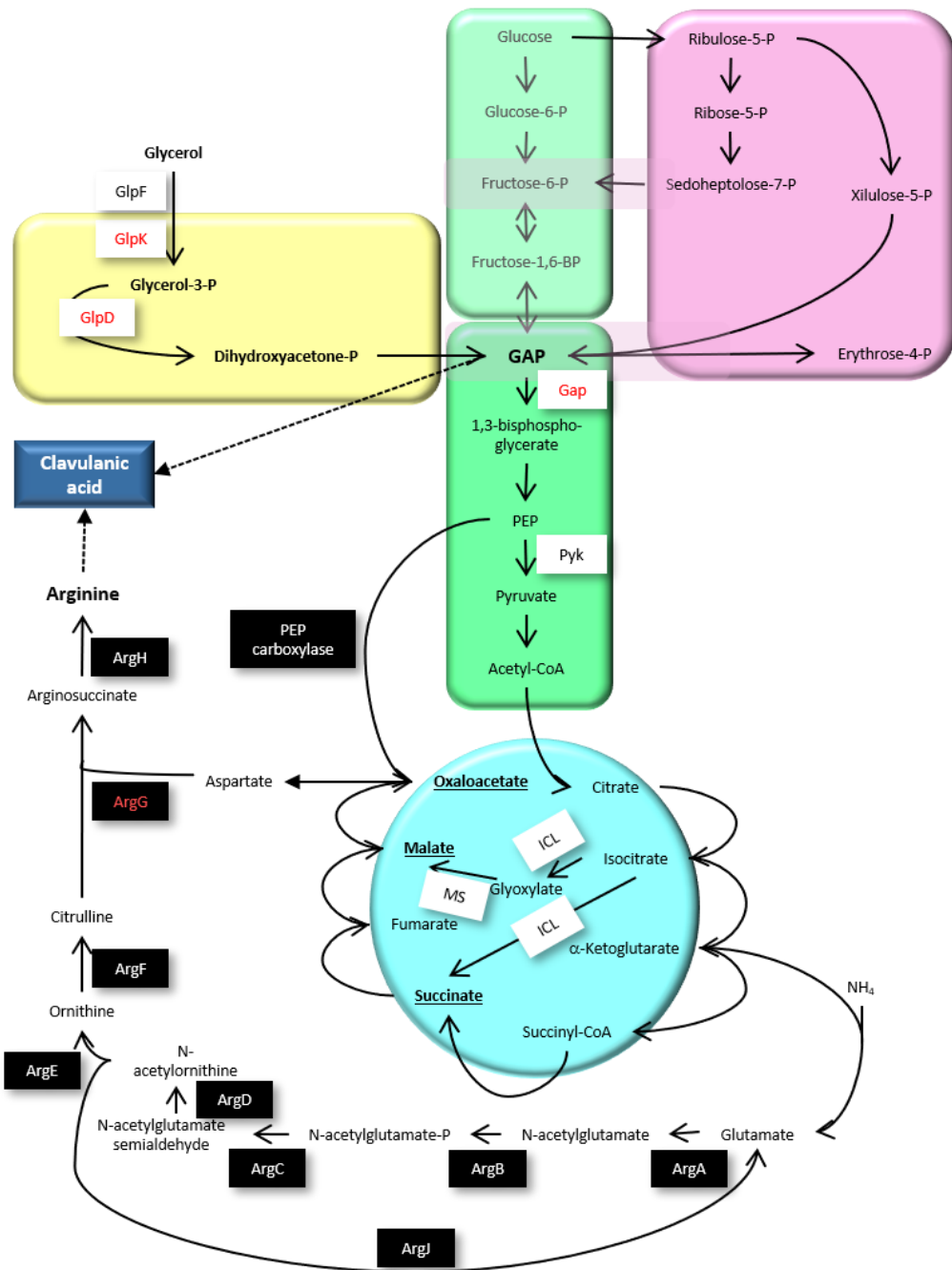
Figure 1-2: Central carbon metabolic pathways and CA biosynthesis precursors. Enzymes are shown in white and black boxes, enzymes that have been targeted during rational strain improvement are shown in red. Yellow box: glycerol uptake and catabolism catalysed by GlpF (=glycerol permease), GlpK (glycerol kinase), GlpD (=glycerol-3-P dehydrogenase). Green box: glycolytic reactions. GAP=glyceraldehyde-3-P; Gap=glyceraldehyde-3-P dehydrogenase; PEP=phosphoenol pyruvate; Pyk=pyruvate kinase. Pink box: Pentose phosphate pathways. Blue circle: TCA cycle and glyoxylate shunt. ICL=Isocitrate lyase; MS=malate synthase. Arginine biosynthetic pathway catalysed by ArgABCDEFGHJ.

Due to its inability to utilise glucose, GAP is not formed from glucose via glucose-6-P, fructose-6-P and fructose-1,6-bisphosphate in *S. clavuligerus*, but is instead formed from glycerol via glycerol-3-P and dihydroxyacetone-P (DHAP). The conversion from glycerol to glycerol-3-P is carried out by glycerol kinase (GlpK) and the conversion from glycerol-3-P to DHAP is catalysed by glycerol-3-P dehydrogenase (GlpD) (Minambres *et al.*, 1992; Baños *et al.*, 2009). The corresponding genes, *glpK* and *glpD*, of which there are two each, are located within an operon containing a gene encoding a glycerol permease (*glpF*). The availability of glycerol is a crucial factor for CA biosynthesis, which is only produced under excess glycerol conditions (Kirk *et al.*, 2000; Chen *et al.*, 2002; Ramirez-Malule *et al.*, 2018). The uptake and subsequent catabolism of glycerol carried out by Glp enzymes can be increased by expressing multiple copies of the *glp* genes. This positively affected CA production (Baños *et al.*, 2009).

Once GAP has been formed, approximately 80-90% of GAP feeds into downstream glycolytic reactions if glycerol is in excess. Under glycerol-limited conditions, 100% of GAP proceeds through central carbon metabolism, leaving only limited amounts of GAP to serve as a precursor of CA biosynthesis (Kirk *et al.*, 2000). During glycolysis, GAP is converted to 1,3-bisphospho-glycerate by glyceraldehyde-3-P dehydrogenases, of which *S. clavuligerus* also has two, Gap1 and Gap2 (Li *et al.*, 2006). Disruption of the corresponding genes, *gap1* and *gap2*, leads to elevated CA production during fermentations due to the increased pool of GAP available for CA production (Chen *et al.*, 2003; Li *et al.*, 2006). In fact, the pool of available GAP in the *gap* mutants increases to such an extent that availability of arginine, the other precursor, becomes the rate limiting factor. Conversely, arginine feeding has no effect on CA yields produced by the WT (Chen *et al.*, 2003; Li *et al.*, 2006). External feeding of the *gap1* mutant with arginine increases CA yields to >200% of WT yields (Li *et al.*, 2006).

After the conversion from GAP to 1,3-bisphospho-glycerate, the latter is further converted to phosphoenolpyruvate (PEP), to subsequently form pyruvate, a reaction catalysed by pyruvate kinase (Pyk). Two Pyk-encoding genes, *sclav_1203* and *sclav_4329*, are encoded by the *S. clavuligerus* genome (Medema *et al.*, 2010; Cao *et al.*, 2016). Gene family expansion in central carbon metabolism has previously been studied by characterising Pyk1 and Pyk2 from *S. coelicolor*, both of which are constitutively expressed.

Yet, the two enzymes differ in their enzymatic activities. It was, thus, concluded that duplicated enzymes of central carbon metabolism might increase an organism's fitness by fulfilling distinct physiological roles and, thereby, elevating the overall metabolic robustness of the cell (Schniete *et al.*, 2018). Pyk and PEP carboxylase, which catalyses the formation of oxaloacetate from PEP, compete for PEP as a substrate. Interestingly, carbon flux through the latter enzyme was found to be increased during CA biosynthesis with carbon flux through Pyk being reduced (Ramirez-Malule *et al.*, 2018). This highlights the importance of studying and understanding carbon flux through the phosphoenolpyruvate-pyruvate-oxaloacetate (PEP-PYR-OXA) node for improved secondary metabolite production. The two Pyk enzyme from *S. clavuligerus*, however, have not yet been characterised in that way.

Pyruvate is converted to acetyl-CoA, which subsequently enters the TCA cycle. A link has been found between the accumulation of TCA intermediates and CA biosynthesis in *S. clavuligerus*: Succinate, malate and oxaloacetate (Ramirez-Malule *et al.*, 2018). Succinate and malate are also formed during anaplerotic reactions as part of the glyoxylate pathway, which is active in *S. clavuligerus* under phosphate and/or carbon limited conditions (Soh *et al.*, 2001; Ramirez-Malule *et al.*, 2018). Isocitrate lyase (ICL) and malate synthase (MS), the two enzymes required for the formation of succinate and malate from isocitrate have been identified in *S. clavuligerus* (Chan *et al.*, 1998; Soh *et al.*, 2001). ICL is responsible for generating succinate and glyoxylate from isocitrate, and malate synthase catalyses the formation of malate from glyoxylate and acetyl-CoA.

Arginine, the second precursor molecule, is synthesised from glutamate, which in turn is formed from α-ketoglutarate, a TCA cycle intermediate (Rodríguez-García *et al.*, 1995, 2000; Rodríguez-García *et al.*, 1997). The arginine biosynthetic genes have been identified in *S. clavuligerus*, most of which are organised in the *argCJBDRGH* cluster. Amplification of *argG*, the product of which is an argininosuccinate synthetase (ASS), which catalyses the conversion from citrulline to arginosuccinate, increases CA production 2.3-fold compared to the WT (Rodríguez-García *et al.*, 1995). Expression of the *arg* genes in the presence of L-arginine is repressed by ArgR, also encoded in the *arg* cluster. The repressor recognises ARG boxes, imperfect 18-nt long palindromic tandem repeats, one of which is located upstream of *argC*, the first gene in the *arg* cluster (Rodríguez-García *et al.*, 1997; Rodríguez-García *et al.*, 2000).

Interestingly, *argR* deletion has a pleiotropic effect on expression of CA biosynthetic genes with some of the genes being upregulated while others are downregulated (Yin *et al.*, 2012).

1.3.11   CARBOHYDRATE UPTAKE SYSTEMS AND REGULATORS INVOLVED IN CCR

Transport of solutes across a biological membrane that encloses the cytoplasm is essential for survival. Solutes naturally move from an area with a higher concentration to that with a lower concentration, a process known as diffusion (Fick's law). Diffusion does not require a transport system (Stillwell, 2016). The cell envelope, comprised of a layer of peptidoglycan (PG) and the lipid bilayer, separates *Streptomyces* from their soil environment (Rahlwes *et al.*, 2019). Transporters embedded in the cytoplasmic membrane allow transport of solutes against their concentration gradient and enable solutes, for which the membrane is impermeable, to pass. Two types of transport are distinguished: channel and carrier mediated passive transport (facilitated diffusion) and active transport. Facilitated diffusion relies on the energy provided by the concentration gradient of a solute for transport, while active translocation of a solute requires additional energy to transport the solute against its electrochemical gradient. The energy for primary active transport is usually ATP hydrolysis, whereas co-transport of an ion is required to generate an electrochemical gradient as the energy source for secondary active transport (Stillwell, 2016).

In silico analysis of the *S. coelicolor* genome predicted the presence of 53 genes encoding carbohydrate uptake systems, with the largest family of transporters being ATP-binding cassette (ABC) transporters, followed by phosphoenolpyruvate (PEP)-dependent phosphotransferase (PTS) transporters, followed by transporters of the Major Facilitator Superfamily (MFS), all of which conduct active forms of transport (Bertram *et al.*, 2004). A single aqua-glyceroporin (GlpF, see 1.3.10, Figure 1-2) that allows facilitated diffusion of glycerol across the membrane, has been identified in *Streptomyces* (Smith *et al.*, 1988a; Minambres *et al.*, 1992; Baños *et al.*, 2009). The following carbohydrate transport systems and the involvement of associated proteins in regulating CCR are described in the subsequent sections: Glucose uptake via glucose permease (H$^+$-symporter, secondary active transport), ABC transport systems (primary active transport) for maltose, xylose and (GlcNAc)$_2$, GlcNAc uptake via a PEP-dependent PTS (PEP:PTS, primary active transport), Glycerol uptake via the aquaglyceroporin (facilitated diffusion).

The major glucose transport system in streptomycetes is the glucose permease-glucokinase (GlcP-Glk) uptake system (van Wezel *et al.*, 2005). The system has been termed "major" due to evidence supporting the existence of a constitutive, low-affinity uptake system (Sabater *et al.*, 1973; Hodgson, 1982). GlcP is a member of the MFS membrane transporters, which is one of the largest classes of secondary active transporters. It carries out symport of glucose with H^+ (Marger *et al.*, 1993; van Wezel *et al.*, 2005). In keeping with the characteristics of MFS transporters, the predicted topology for GlcP are 12 transmembrane (TM) domains that form two larger domains comprised of six plus six TM helix bundles. These bundles are connected by a cytoplasmic loop (van Wezel *et al.*, 2005; Reddy *et al.*, 2012).

Import of glucose into the bacterial cell is accompanied by phosphorylation of the substrate. Bacterial glucose transport is usually carried out by a PEP:PTS and substrate phosphorylation is, thus, conducted by a component of the PTS (Siebold *et al.*, 2001). PEP:PTSs also exist in streptomycetes, however not for glucose (see 1.3.11.4). Phosphorylation of glucose is, instead, carried out by the GlcP-associated Glk, which is independent of a PTS (Angell *et al.*, 1992; Mahr *et al.*, 2000). Glk is a member of the ROK (**r**epressor, **o**pen reading frame, **k**inase) family that encompasses sugar kinases and transcriptional repressors (Titgemeyer *et al.*, 1994; Mahr *et al.*, 2000). It is encoded by *glk* and is expressed constitutively, yet its catalytic activity varies in a glucose-dependent manner (Ikeda *et al.*, 1984; Angell *et al.*, 1992; Mahr *et al.*, 2000; van Wezel *et al.*, 2007). It is thought that Glk is modified post-translationally by carbon flux through the GlcP-Glk system, for which a physical interaction between the two components is required (van Wezel *et al.*, 2007; Sun *et al.*, 2020). The nature of physical interaction of GlcP and Glk has not been elucidated to date.

There are two copies of *glcP* genes in *S. coelicolor*, *glcP1* and *glcP2.* These are identical except for a single base pair. Interestingly, only *glcP1* is located in the core region of the linear chromosome, which is where essential housekeeping genes are generally found, *glcP2* is located near the arm of the chromosome (van Wezel *et al.*, 2005). Glk in *S. coelicolor* is encoded by *glkA* (Angell *et al.*, 1992). Despite being unable to utilise glucose efficiently (Garcia-Dominguez *et al.*, 1989), both *glcP* and *glk* are present on the *S. clavuligerus* chromosome (Medema *et al.*, 2010; Song, Jeong, *et al.*, 2010; Cao *et al.*, 2016). The genes in

S. clavuligerus are, henceforth, referred to as *glcP* and *glk* and are to be distinguished from the genes in *S. coelicolor*, which are *glcP1*, the duplicated gene *glcP2*, and *glkA*. Although *S. clavuligerus* expresses its native *glcP* at a level that is too low to enable sufficient utilisation of glucose, which is presumably due to a weak promoter, GlcP and Glk are functional proteins that enable growth of *S. clavuligerus* on glucose when *glcP* is expressed from the *glcP1* promoter (Pérez-Redondo *et al.*, 2010). This also shows that glucose catabolism downstream of glucose import and phosphorylation is active in *S. clavuligerus*, which might explain how *S. clavuligerus* is able to grow on maltose. (Aharonowitz *et al.*, 1978; Garcia-Dominguez *et al.*, 1989; Pérez-Redondo *et al.*, 2010). Interestingly, Glk appears to be constitutively active in *S. clavuligerus* irrespective of the presence of glucose (Garcia-Dominguez *et al.*, 1989; Pérez-Redondo *et al.*, 2010).

1.3.11.2   ROLE OF GLK IN CCR

In the presence of glucose, Glk plays a central role in the repression of uptake and/or utilisation of other sugars, such as arabinose, glycerol, galactose and fructose, as well extracellular polymers, for instance agar, at the transcriptional level (Hodgson, 1982; Angell *et al.*, 1992, 1994; Kwakman *et al.*, 1994; Romero-Rodríguez *et al.*, 2017). Studies investigating CCR in *Streptomyces* were conducted in *glk* mutants that are defective in glucose utilisation and, therefore, resistant to 2-deoxyglucose (Dog[R]). 2-DOG is a glucose molecule of which the 2-hydroxyl group has been replaced by a hydrogen, which makes it non-metabolisable. These *glk* mutants have also lost the ability to repress glycerol kinase and agarase activity. Heterologous expression of *glk* from *Zymomonas mobilis* in these mutants restores glucose utilisation, yet repression of agarase-encoding *dagA* is not re-established (Angell *et al.*, 1994; Kwakman *et al.*, 1994). Early studies aiming to investigate the function of Glk in CCR found that by introducing and expressing *glk* in Dog[R] *S. coelicolor* strains, glucose dependent CCR was only partially re-established (Hodgson, 1982; Angell *et al.*, 1994). This indicates that Glk alone is not sufficient for glucose triggered CCR in glucose-utilising *Streptomyces*. Indeed, expression of the gene adjacent to *glk*, *sco2127*, was required for full glucose triggered repression (Ikeda *et al.*, 1984; Angell *et al.*, 1994; Forero *et al.*, 2012). SCO2127 has been suggested to stimulate *glk* and *glcP* expression in *S. peucetius* var. *caesius* with the expression profile of *sco2127* matching that of *glcP* in glucose containing media (Guzman *et al.*, 2005; Chávez *et al.*, 2009). Dog[R] mutants of *S. peucetius* var. *caesius* are less

sensitive to CCR, exhibit reduced glucose uptake and low Glk activity (Segura, 1996; Escalante *et al.*, 1999). Yet, by expressing *sco2127*, a reversal of the Dog[R] phenotype was observed (Guzman *et al.*, 2005). SCO2127 mediated CCR has been suggested to act independently from Glk, however, this remains to be clarified (Forero *et al.*, 2012). Protein-protein interaction analyses have identified BldKB, the ABC transporter in the Bld cascade (see 1.3.6), as an interaction partner of SCO2127, eluding to the possibility of SCO2127 acting on the interface of CCR and developmental control (Chávez *et al.*, 2011).

Glk lacks a DNA-binding motif, making it likely that Glk exercises its global control of CCR through interactions with transcriptional regulators, such as MalR, XylR, DasR and GylR, though this has not been demonstrated conclusively (Mahr *et al.*, 2000; van Wezel *et al.*, 2007). Moreover, GlkA-independent CCR mechanisms have been identified (Flores *et al.*, 1993; Gubbens *et al.*, 2012) and analysing CCR in *Streptomyces* other than *S. coelicolor* has expanded the widely accepted model for Glk in CCR in this genus (Ordóñez-Robles *et al.*, 2017).

More recently, evidence for the presence of more than one version of the Glk has led to the idea that Glk might form a complex with GlcP at the membrane, while another, cytosolic version mediates CCR (van Wezel *et al.*, 2007). In 2020, a novel form of post-translational modification (PTM) of Glk was reported in *S. roseosporus*, which is crotonylation. This form of PTM was shown to negatively regulate glucokinase activity while positively regulating Glk-dependent CCR. This means that crotonylated Glk has low kinase activity while CCR-mediated repression of gene expression is high and vice versa. Glk and other enzymes involved in glucose catabolism were found to be crotonylated in response to glucose (Sun *et al.*, 2020). *S. coelicolor* gene homologs of *cobB* and *hdac*, the two de-crotonylases, and *kct1*, the crotonyl-transferase, involved in this form of PTM in *S. roseosporus* are *sco0452*, *sco3330* and *sco3175*. Interestingly, although homologues for *hdac* (*sclav_2356*) and *kct1* (*sclav_2309*) are present in *S. clavuligerus*, there is no gene homologous to *cobB*. In *S. roseosporus*, CobB is primarily responsible to erasing crotonylation and maintaining low crotonylation of Glk during secondary metabolism (Sun *et al.*, 2020).

As mentioned previously (see 1.3.5.4), glucose does not repress antibiotic production in *S. clavuligerus*. Random mutagenesis of *S. clavuligerus* NRRL 3585 resulted in a glucose

utilisation mutant, *gut-1*, which was able to transport glucose and galactose (García Domínguez *et al.*, 1988; Garcia-Dominguez *et al.*, 1989). Strikingly, the *gut-1* mutant produced CA and cephamycin C at levels comparable to the WT strain (García Domínguez *et al.*, 1988; Garcia-Dominguez *et al.*, 1989), showing that glucose import and phosphorylation in the *gut-1* mutant occurs independently of Glk or confirming that Glk does not control antibiotic biosynthesis in *S. clavuligerus* in general.

1.3.11.3   ABC TRANSPORTERS AND REGULATORS INVOLVED IN CCR

ABC transporters conduct active transport of solutes with the energy being provided by an ATPase. Unlike transport mediated via GlcP-Glk and PTS, sugars and carbohydrates that are translocated via an ABC transporter are not phosphorylated. Bacterial ABC transport systems consist of a transmembrane channel with the canonical six plus six TM structure (see MFS transporters, 1.3.11.1) and has an ATPase domain. The substrate is bound by a periplasmic binding protein (Jones *et al.*, 2004).

Multiple substrates for ABC transporters have been identified in *Streptomyces*, such as maltose, xylose and *N`-N`*-diacetylchitobiose, (GlcNAc)$_2$ (van Wezel *et al.*, 1997; Seo *et al.*, 2002; Świątek *et al.*, 2013; Lu *et al.*, 2020). The maltose transporting ABC system, MalEFG, is encoded by *malEFG*. Expression of the operon is controlled by MalR, which acts as a repressor of *malEFG* expression in the presence of glucose. The response of MalR to glucose has been proposed to be mediated either through interaction of MalR with Glk or through modification of MalR by Glk (van Wezel *et al.*, 1997).

Xylose is also transported via an ABC transporter, which is encoded by *xylFGH*. Expression of the operon is regulated by Rok7b7 (Figure 1-3). In *S. coelicolor* and *S. avermitilis*, Rok7b7 acts as a global regulator of xylose and glucose uptake, as well as CCR. Its regulon encompasses over 160 genes, among which is *sco2494*, which encodes a pyruvate phosphate dikinase (Świątek *et al.*, 2013; Lu *et al.*, 2020). *S. coelicolor* has two pyruvate phosphate dikinases, PpdK1 (SCO0208) and PpdK2 (SCO2494), which catalyse the generation of PEP from pyruvate and phosphate under ATP consumption (Andrews *et al.*, 1969; Hiltner, 2015). PpdK2 was shown to play a role in Glk-independent CCR and to be upregulated in a Glk-deficient strain (Gubbens *et al.*, 2012).
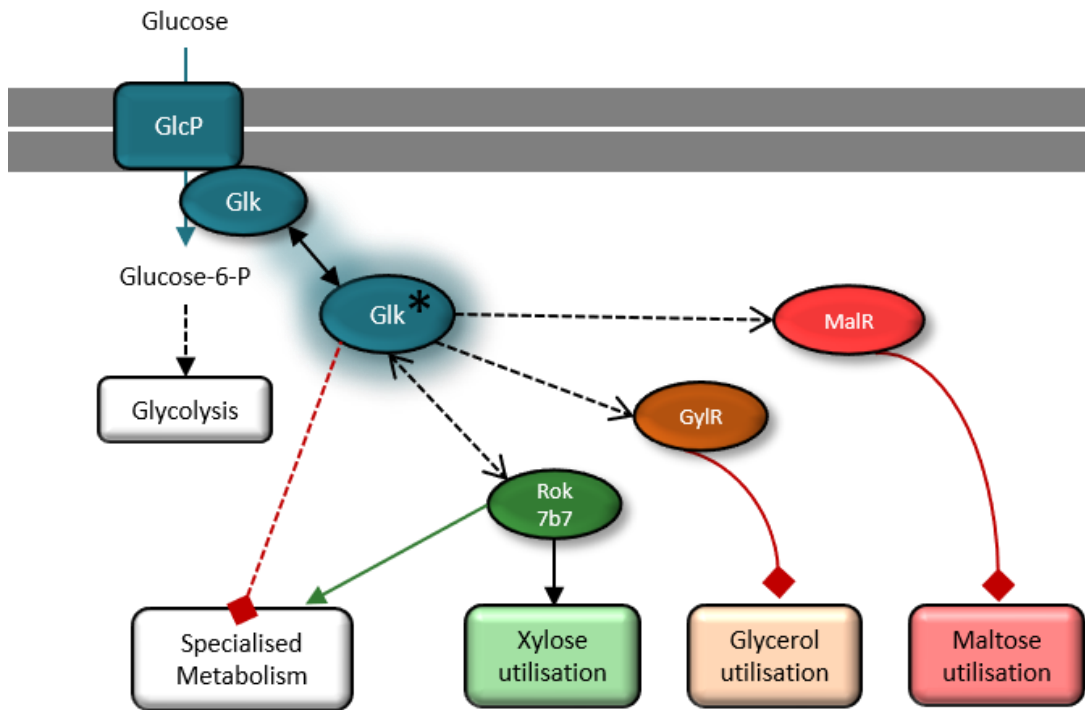
Figure 1-3: Schematic of GlcP-Glk system and the proposed role for Glk in CCR (adapted from van Wezel *et al.*, 2007). Glucose is transported into the cell via GlcP (blue rectangle) and phosphorylated by Glk (blue oval). Glucose-6-P is further catabolised during glycolysis. GlcP and Glk physically interact at the membrane. Glk is post-translationally modified (double-ended arrow) to Glk*. This cytosolic Glk* might interact with regulators involved in carbon utilisation, such as the regulators of maltose and glycerol utilisation, MalR (coral) and GylR (orange). MalR and GylR repress expression of maltose and glycerol utilisation genes, respectively, in the presence of glucose (red, diamond-shaped arrows). Glk and Rok7b7 might influence each other (dotted, double-ended arrow). Glk represses specialised metabolism, for instance actinorhodin production in *S. coelicolor* (red, diamond-shaped arrow), whereas Rok7b7 positively regulates it (green arrow).

Multiple streptomycete transporters exist for GlcNAc, which is a monomer of the biopolymer chitin, and is, thus, ubiquitous in the soil, as it is found in fungal cell walls and invertebrate exoskeletons In *S. olivaceoviridis*, an ABC uptake system, NgcEFG, for GlcNAc was discovered. NgcE, the substrate binding protein of the system, effectively binds GlcNAc and its dimer, *N`-N`*-triacetylchitobiose (GlcNAc)$_2$, suggesting that NgcEFG might be able to transport both of these substrates in nature (Xiao *et al.*, 2002). Uptake of the substrates is regulated by NgcR, a ROK-family regulator, that shares 92% of its amino acid sequence with Rok7b7. Discovery of NgcEFG lead to the identification of the DasABC transporter in *S. coelicolor* (Figure 1-4). Yet, unlike NgcE, DasA is only able to bind (GlcNAc)$_2$ (Saito *et al.*, 2007). Once (GlcNAc)$_2$ has been imported by DasABC, it is converted to GlcNAc and subsequently phosphorylated by NagK, producing GlcNAc-6-P. GlcNAc-6-P, in turn, is deacetylated by NagA (Świątek *et al.*, 2012).

Upstream of the *dasABC* operon is *dasR*, which encodes a GntR-type master regulator of carbon metabolism, development, and secondary metabolite production. It was first described in *S. griseus* (Seo *et al.*, 2002). DasR controls GlcNAc-dependent growth and development of *Streptomyces*, such as by regulating the DasABC transporter, expression of *nagK* and *nagA*, as well as the GlcNAc-specific PTS system (see 1.3.11.4) (Seo *et al.*, 2002; Nothaft *et al.*, 2003b; Nothaft *et al.*, 2010; Świątek *et al.*, 2012). Furthermore, it also controls expression of secondary metabolite biosynthetic genes by recognising and binding to so called **D**asR-**r**esponsive **e**lements (*dre*), as well as non-canonical sites of cluster-situated regulator genes (Seo *et al.*, 2002; Rigali *et al.*, 2006, 2008; Świątek-Połatyńska *et al.*, 2015). GlcNAc-6-P and GlcN-6-P are both allosteric effectors of DasR, preventing its binding to DNA. Conversely, high concentrations of organic or inorganic phosphate enhance DasR binding to its targets (Tenconi *et al.*, 2015). Given that GlcNAc induces antibiotic production in *S. clavuligerus* (Rigali *et al.*, 2008), DasR in *S. clavuligerus* might play a role similar to the one illustrated in *S. coelicolor*. However, the role of DasR in *S. clavuligerus* has not yet been investigated in more detail.
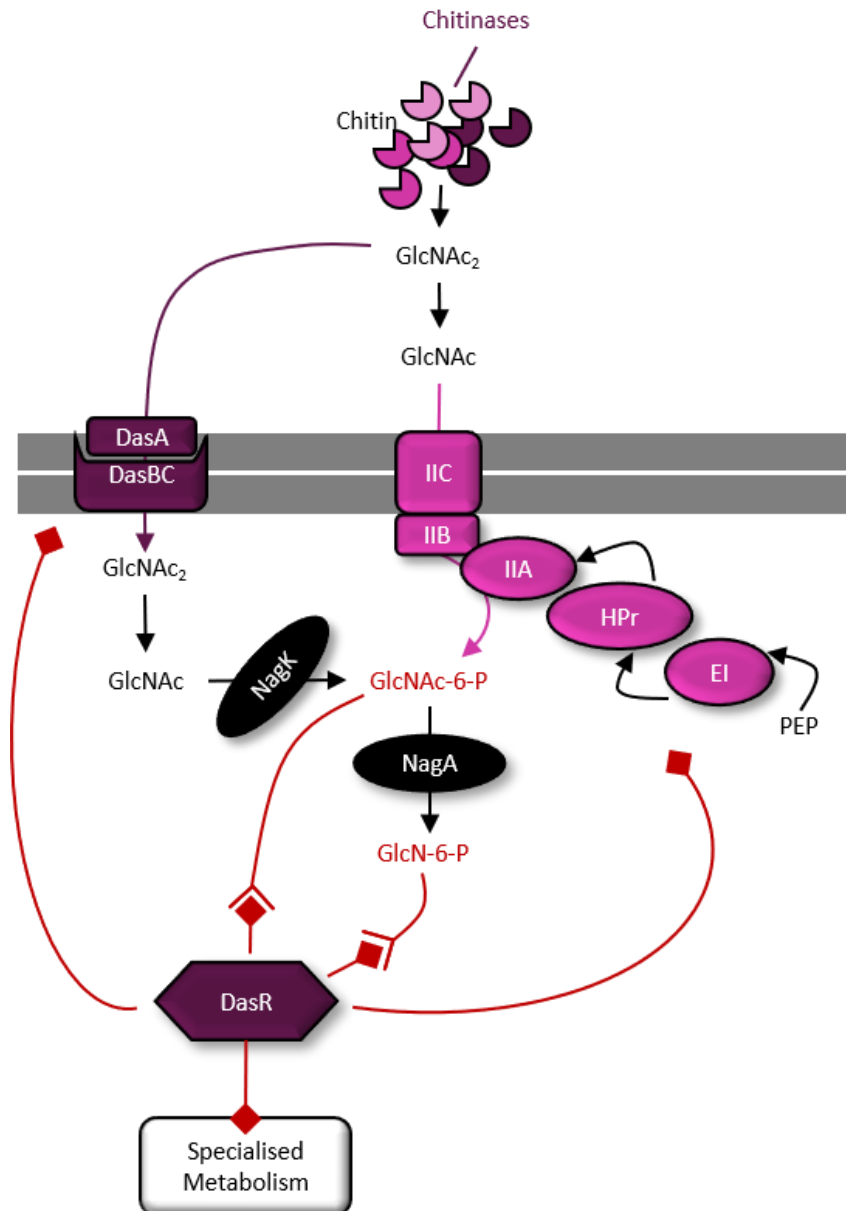
Figure 1-4: GlcNAc and (GlcNAc)$_2$ import via DasABC and the PEP:PTS (adapted from Rigali *et al.*, 2006). Chitin in the soil is degraded by chitinases (pink and purple sectors). (GlcNAc)$_2$ is imported by DasABC (purple), converted to GlcNAc and phosphorylated by NagK. GlcNAc is transported via the PEP-PTS (pink), is directly converted to GlcNAc-6-P, and further converted to GlcN-6-P by NagA. GlcNAc-6-P and GlcN-6-P are both allosteric inhibitors of DasR. DasR inhibits expression DasABC and PEP-PTS genes, as well as *nagA* and *nagK*. DasR is also involved in repression of specialised metabolism.

Uptake and phosphorylation of carbohydrates via PEP:PTS is a form of active transport that relies on group translocation of a phosphoryl group with phosphoenolpyruvate serving as the phosphoryl donor (Roseman, 1969). The phosphoryl translocation chain involves multiple components, most of which are soluble and one of which is an integral membrane permease. Soluble components are the histidine-containing phosphocarrier protein HPr and enzyme I (EI), as well as the EII complex, consistent of EIIA and EIIB. The third component of the EII complex is EIIC, the integral membrane permease responsible for importing the sugar. Sugar import via the EIIC permease is associated with phosphorylation of the sugar through EIIB (Saier Jr *et al.*, 1992; Deutscher *et al.*, 2014).

Group translocation begins with the phosphorylation of HPr, which then transfers the phosphoryl group onto EI. Subsequently, the phosphoryl group is transferred onto EIIA, EIIB and EIIC. EI and HPr are universal energy coupling enzymes and are not specific to the substrate transported by the PEP:PTS. For instance, HPr from *S. coelicolor* is able to phosphorylate EII^{Glucose} from *Bacillus subtilis*, EII^{Lactose} from *Staphylococcus aureus* and EIIA^{Mannitol} from *E. coli* although the streptomycete PEP:PTS does not transport any of these substrates (Parche *et al.*, 1999). Indeed, the *Streptomyces* PEP:PTS transports fructose and GlcNAc, yet low levels of HPr activity have been detected in *S. coelicolor* in the presence of glucose (Titgemeyer *et al.*, 1995; Parche *et al.*, 1999; Wang *et al.*, 2002; Nothaft *et al.*, 2003a; Nothaft *et al.*, 2003b;). The PTS has been studied in detail in *S. coelicolor* and *S. olivaceoviridis* (Wang *et al.*, 2002; Nothaft *et al.*, 2003a; Nothaft *et al.*, 2003b). PTS-mediated uptake of fructose and GlcNAc involves the transfer of a phosphoryl group from PEP to EI, encoded by *ptsI*, which is subsequently transferred to HPr, encoded by *ptsH* (Parche *et al.*, 1999, 2000). HPr activity was shown to be induced in the presence of fructose and deletion of *pstH* results in a fructose negative phenotype (Butler *et al.*, 1999; Nothaft *et al.*, 2003b). The phosphoryl group is transferred from HPr onto IIA, either EIIA^{crr} if GlcNAc is the substrate or EII^{Fructose} if the transported sugar is fructose. The gene encoding EIIA^{crr}, *crr*, is located upstream of *ptsI*. Promoter sequences were detected upstream of both genes, but reverse transcriptase PCR showed that *crr* and *ptsI* are transcribed together. Transcription of *crr* and *ptsI* is stimulated in the presence of GlcNAc (Nothaft *et al.*, 2003b). Furthermore, a common *cis*-element was located upstream of *ptsI*, but also upstream of chitinase-encoding genes (Nothaft *et al.*,

2003b), highlighting the biological relevance of a GlcNAc-transporting system in the context of chitin utilisation in the soil. While the absence of *crr* only results in the absence of growth on GlcNAc, a *pstI* null mutant is unable to grow on GlcNAc and fructose (Nothaft *et al.*, 2003a; Nothaft *et al.*, 2003b). The next group of enzymes in the phosphoryl group translocation chain of the GlcNAc-transporting PTS consists of enzyme IIB, NagF and PtsC1 in *S. coelicolor* and *S. olivaceoviridis*, respectively, and enzyme IIC (NagE2, PtsC2) with the former carrying out phosphorylation of the incoming substrate to form *N'*-acetylglucosamine-6-phosphate (GlnNAc-6-P) (Wang *et al.*, 2002; Rigali *et al.*, 2006). Both NagE2 and PtsC2 are the high-affinity permeases in the PEP:GlcNAc PTS in *S. coelicolor* and *S. olivaceoviridis*, respectively , and their absence causes a GlcNAc negative phenotype in the respective strains (Wang *et al.*, 2002; Nothaft *et al.*, 2010).

Interestingly, *nagE2* is located downstream of *nagF* and *nagE1*. Yet, deletion of *nagE1* does not cause the GlcNAc negative phenotype observed in the absence of *nagE2*. Further, expression of the *nag* regulon, the set of genes dependent on the presence and sensing of extracellular GlcNAc, is dependent on NagE2 and expression of HPr was entirely lost in a *nagE2* null mutant. Expression of the *pst* and *nagE* genes is regulated by the global GntR-type regulator DasR (Figure 1-4), which acts as a repressor of gene expression in the absence of GlnNAc (Rigali *et al.*, 2004, 2006).

1.3.11.5 Facilitated Diffusion of Glycerol

In streptomycetes, glycerol crosses the lipid bilayer via facilitated diffusion mediated by an aquaglyceroporin, the first one of which was discovered in 1992 (Preston *et al.*, 1991; Borgnia *et al.*, 1999). Aquaporins are transmembrane water pores. However, aqua-glyceroporins also allow passage of small, uncharged molecules, such as glycerol, ammonia and urea, across the membrane (Borgnia *et al.*, 1999; Stillwell, 2016).

The facilitator of glycerol transport in *Streptomyces* is the glycerol permease, GlpF. Phosphorylation of the incoming glycerol to glycerol-3-P is carried out by GlpK and subsequent conversion to DHAP is catalysed by glycerol-3-P dehydrogenase, GlpD (Seno *et al.*, 1983, 1984; Biro *et al.*, 1987). In *S. coelicolor*, the genes encoding GlpF, GlpK and GlpD are organised in one transcriptional unit to form the glycerol utilisation operon, *gylCABX*, the

promoter of which is glycerol-inducible and glucose-repressible in *S. coelicolor*. Upstream of *gylCABX*, there is the gene encoding GylR, which is transcribed from its separate glycerol-induced promoter (Smith *et al.*, 1988b, 1988a; Bolotin *et al.*, 1990; Hindle *et al.*, 1994). GylR is an autoregulator that is also directly involved in regulating glycerol utilisation in the presence of glucose as it acts as a transcriptional repressor by binding upstream of the operon (Hindle *et al.*, 1994). Additionally, NdgR, an IclR-type regulator, has been identified as a transcriptional regulator of the operon that binds upstream of *gylC* in *S. coelicolor*. NdgR was formally known as a nitrogen-dependent growth regulator that is also involved in controlling antibiotic production (Yang *et al.*, 2009; Lee *et al.*, 2015, 2017).

While the *S. coelicolor* genome encodes a single glycerol utilisation cluster, *S. clavuligerus* has two operons encoding genes involved in glycerol catabolism that differ in their number of genes (Minambres *et al.*, 1992; Baños *et al.*, 2009). The larger operon contains the genes encoding GylR (*gylR*), GlpF1 (*glpF1*), GlpK1 (*glpK1*) and GlpD (*glpD1*). The smaller operon is comprised of *glpF2*, *glpK2* and *glpD2*, the gene products of which share 50%, 70% and 32% of their amino acid sequence with GlpF1, GlpK1 and GlpD1, respectively (Figure 1-5) (Baños *et al.*, 2009).

While the expression of *gylR-gylF1K1D1* is induced by glycerol, *gylF2-gylK2* is constitutively expressed, albeit at a level too low to support growth with glycerol as a sole carbon source (Minambres *et al.*, 1992; Baños *et al.*, 2009). Interestingly, the inducible glycerol uptake system was found to be repressed by serine and cysteine (Minambres *et al.*, 1992), which might imply that there is cross-talk between amino acid catabolism and glycerol utilisation in *S. clavuligerus*.
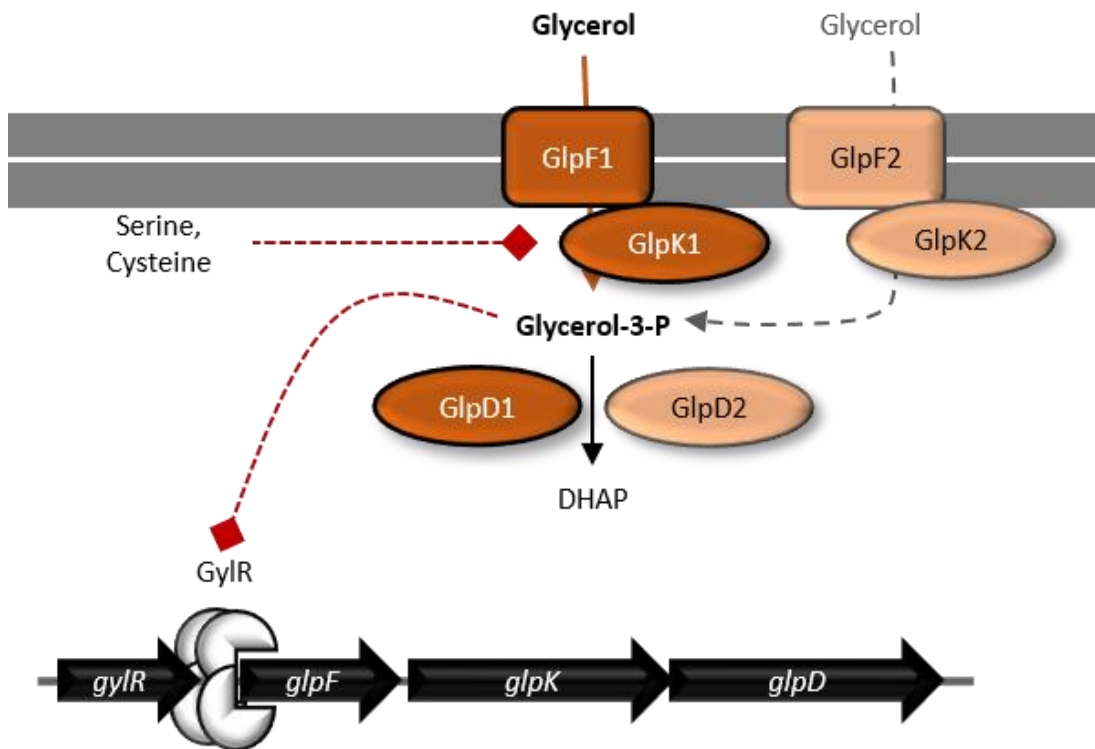
Figure 1-5: Glycerol uptake and phosphorylation in *S. clavuligerus*. Small amounts of glycerol are imported and phosphorylated via the constitutive glycerol uptake system (light orange), GlpF2 and GlpK2. The catabolic product, glycerol-3-P, triggers dissociation of GylR (white sectors, red, diamond-shaped arrow) from upstream of the glycerol utilisation cluster, *glpFKD*. The inducible glycerol uptake system, GlpF1K1D1 (orange), is repressed by L-serine and L-cysteine (red, diamond-shaped arrow).

1.4   CLAVULANIC ACID BIOSYNTHESIS BY *S. CLAVULIGERUS*

A few years after *S. clavuligerus* was discovered, it was found to produce the β-lactamase inhibitor clavulanic acid (CA). CA is classified as a β-lactam antibiotic and is comprised of the characteristic β-lactam ring fused to an oxazolidine ring, which is a thiozolidine ring in other penicillins and cephalosporins. Furthermore, CA lacks the acylamino-side chain present in other β-lactam antibiotics (Howarth *et al.*, 1976; Reading *et al.*, 1977). This bicyclic structure is a characteristic of clavam antibiotics although CA has a unique 3R, 5R stereochemistry. It is also the only clavam that exhibits β-lactamase inhibitory properties (Saudagar *et al.*, 2008). CA does not exhibit potent antibiotic activity and alone it is not an effective treatment for bacterial infections. In combination with amoxycillin, however, a 5,000-fold decrease in the minimum inhibitory concentration (MIC) of amoxycillin is observed against methicillin resistant *Staphylococcus aureus* (Saudagar *et al.*, 2008). According to the Bush-Jacoby-Medeiros functional classification system for β-lactamases, CA inhibits Group 2 β-lactamases of the molecular classes A and D, which include penicillin-hydrolysing enzymes, carbenicillin-hydrolysing enzymes, cephalosporinases, as well as broad and extended spectrum β-lactamases, such as those from *Klebsiella* p*seudomonas* and *Salmonella* (Bush *et al.*, 1995).

1.4.1   BIOSYNTHETIC PATHWAY

The biosynthetic pathway of CA has not yet been entirely elucidated. However, feeding studies and ^{13}C-NMR studies using ^{13}C-labelled molecules have allowed the identification of arginine and GAP as precursor key molecules. Intermediates have also been identified with these methods (Valentine *et al.*, 1993; Khaleeli *et al.*, 1999; Song, Jensen, *et al.*, 2010). The current understanding of the steps required for the biosynthesis of CA are described in the following paragraph, genes encoding enzymes that are involved in a reaction are mentioned in parentheses (Figure 1-6). The CA gene cluster is approximately 30 kb in size and is located adjacent to the cephamycin gene cluster in *S. clavuligerus* (Ward *et al.*, 1993). Additionally, there is a paralogous CA gene cluster that contains some of the biosynthetic genes (Jensen *et al.*, 2000).
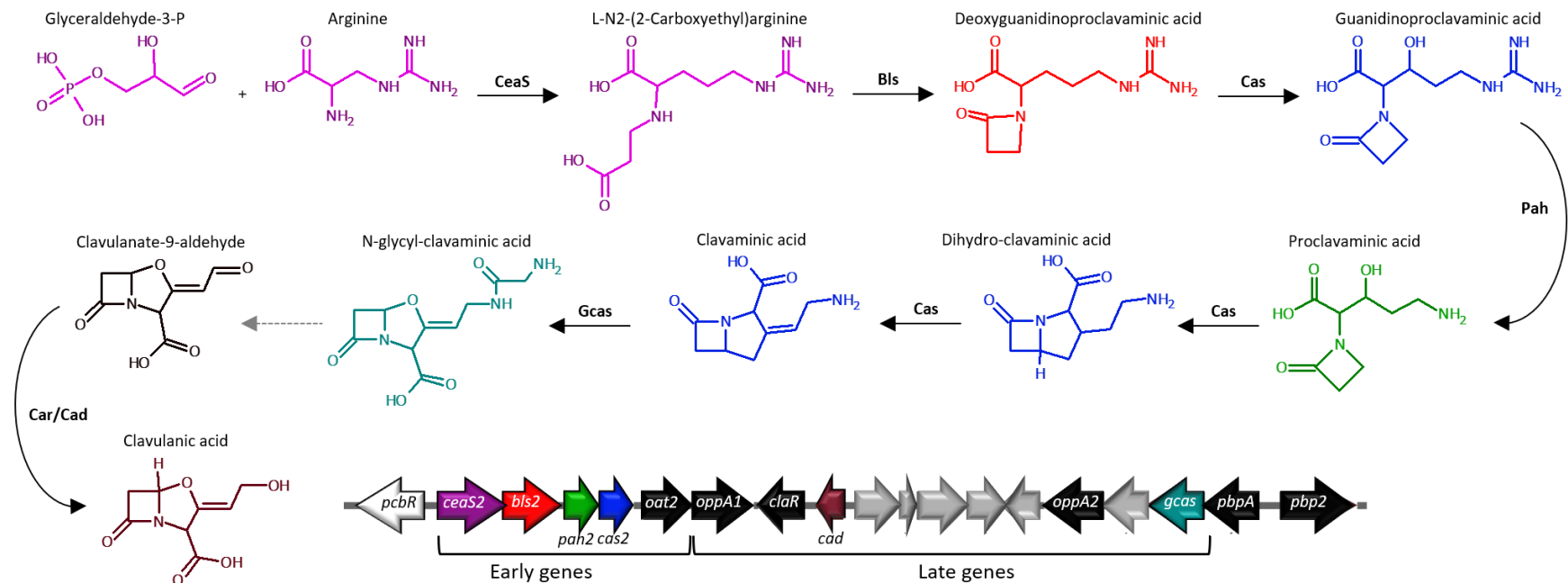
Figure 1-6: CA biosynthetic pathway and gene cluster (adapted from Pinilla *et al.*, 2019). Precursor and intermediate molecules of CA biosynthetic pathway. Enzymes catalysing a reaction are shown above the reaction arrows. Early and late genes encoding the enzymes that catalyse formation of intermediate molecules are colour-coded with the intermediates. Genes involved in cephamycin C biosynthesis are shown in black.

The reaction is catalysed by the thiamine pyrophosphate (TPP)-dependent L-N2-(2-carboxyethyl) arginine (Cea) synthase CeaS, encoded by *ceaS2* and *ceaS1* on the paralogous gene cluster (Khaleeli *et al.*, 1999; Jensen *et al.*, 2000). Cea is subsequently converted into deoxyguanidinoproclavaminic acid (DGPC) by a β-lactam synthetase, Bls (*bls2* and *bls1*). This step results in the formation of the characteristic β-lactam ring. The mechanism of ATP/Mg^{2+}-dependent intramolecular closure to form the β-lactam ring is different from the oxidative cyclisation that occurs during biosynthesis of penicillin (Bachmann *et al.*, 1998, 2000; Jensen *et al.*, 2000).

The following enzymatic reactions are catalysed by the clavaminic acid synthase, Cas (*cas2* and *cas1*) and the proclavaminate amidinohydrolase, Pah (*pah2* and *pah1*). The hydroxylation of DGPC, resulting in the formation of guanidinoproclavaminic acid, is catalysed by Cas. Subsequently, the guanidine group of guanidinoproclavaminic acid is removed by Pah, generating the intermediate proclavaminic acid (Baldwin *et al.*, 1993; Elson *et al.*, 1993; Jensen *et al.*, 2000). The latter, in turn, serves as substrate for an oxidative ring closure catalysed by Cas. The formation of this ring results in the formation of the first bicyclic molecule of the pathway, namely dihydro-clavaminic acid (Elson *et al.*, 1987; Salowe *et al.*, 1991). Dihydro-clavaminic acid is converted to clavaminic acid by Cas. The gene encoding the enzyme that is required for the next catalytic step, the N-glycyl-clavaminic acid synthase (Gcas), is located downstream of the cluster-situated transcriptional activator encoding *claR* gene in *orf17* (see 1.4.2), unlike the previously mentioned genes which are found upstream of *claR*, which marks the border of "early" and "late" biosynthetic genes. Gcas catalyses the conversion from clavaminic acid to N-glycyl-clavaminic acid in an ATP-dependent manner as one of the final reactions leading up to CA formation (Mellado *et al.*, 2002; Jensen *et al.*, 2004; Arulanantham *et al.*, 2006).

The last intermediate is known to be clavulanate-9-aldehyde. However, it has not yet been elucidated how N-glycyl-clavaminic acid is converted into clavulanate-9-aldehyde. This is a highly unstable molecule with β-lactamase activity, which is thought to be due to its 3R, 5R stereochemistry (Nicholson *et al.*, 1994; Fulston *et al.*, 2001). From clavulanate-9-aldehyde, a final reduction reaction is necessary to form CA. This NADPH-dependent reaction is carried

out by the CA dehydrogenase (Cad), encoded by the *cad* gene (Pérez-Redondo *et al.*, 1998; MacKenzie *et al.*, 2007).

1.4.2 CLUSTER-SITUATED REGULATORS

CcaR is a cluster-situated autoregulatory positive regulator that governs clavulanic acid, cephamycin C and clavam production in *S. clavuligerus* (Aidoo *et al.*, 1994; Pérez-Llarena *et al.*, 1997; Alexander *et al.*, 1998). It is a member of the SARP family (Wietzorrek *et al.*, 1997; Bibb, 2005). CcaR regulates its own expression by binding to the *ccaR* promotor. It also binds to bidirectional *cefD-cmcI* promotor within the cephamycin C biosynthetic gene cluster, thereby controlling the expression of CA and cephamycin C biosynthetic genes. CcaR interacts with the other positive regulator for CA biosynthesis, ClaR (Pérez-Llarena *et al.*, 1997; Santamarta *et al.*, 2002). The *ccaR* gene contains the rare TTA codon (position 38) and its expression was, thus, expected to depend on *bldA* to provide the leucyl tRNA for translation as is the case in *S. coelicolor* (Lawlor *et al.*, 1987; Leskiw *et al.*, 1991). Surprisingly, in *S. clavuligerus*, *ccaR* mRNA and CcaR levels in a *bldA* disruption mutant are comparable to that of the WT, indicating that *ccaR* is efficiently mistranslated (Trepanier *et al.*, 2002). Aside from the aforementioned AdpA (see 1.3.5.1), BldG, an anti-anti-sigma factor and component of the Bld cascade (see 1.3.6) influences *ccaR* expression, presumably via AdpA (Bignell *et al.*, 2005; Ferguson *et al.*, 2016). A *bldG* null mutant is not only defective in development but also in cephamycin C and CA biosynthesis (Bignell *et al.*, 2005).

Interestingly, gene expression of 186 genes were altered in a *ccaR* deletion mutant. Among the genes that were significantly downregulated were the genes encoding the glycerol-inducible uptake system, *glpF2*, *glpK2*, the regulator gene, *gylR*, as well as *glpK1* and *glpD*. Another gene that was downregulated was *gap2*, the product of which is responsible for converting GAP to 1,3-bisphosphoglycerate (see Figure 1-2). Genes involved in arginine synthesis were also significantly downregulated. In contrast, *glpF1*, which encodes the constitutive glycerol permease was slightly upregulated. During exponential growth of this strain, *glcP* the was also significantly upregulated (Álvarez-Álvarez *et al.*, 2014). Furthermore, CcaR is known to repress holomycin production, which is consistent with the stark upregulation of holomycin biosynthetic genes in the *ccaR* mutant (De la Fuente *et al.*, 2002; Álvarez-Álvarez *et al.*, 2014)

ClaR is a LysR-type regulator of CA biosynthesis that is dependent on CcaR (Pérez-Redondo *et al.*, 1998; Álvarez-Álvarez *et al.*, 2014). ClaR also positively regulates expression of CA biosynthetic genes, which is why it has been used as a reporter gene, fused to a kanamycin resistance cassette, to screen for strains with enhanced CA production (Paradkar, Aidoo and Jensen, 1998; Qin *et al.*, 2017). ClaR mainly regulates expression of late CA biosynthetic genes (Paradkar *et al.*, 1998). However, neither expression of the early nor late CA biosynthetic genes was entirely abolished in a *claR* disruption mutant. Moreover, and in contrast to the *ccaR* deletion mutant, no genes involved in central carbon or nitrogen metabolism were discovered among the genes with altered expression profiles. However, a gene involved in aerial mycelium was affected by the absence of ClaR, resulting in a bald phenotype (Martínez-Burgo *et al.*, 2015).

1.5 SCOPE OF THE PHD PROJECT

GSK have developed CA production strain lineages through random mutagenesis (UV radiation and NTG treatment) and subsequent selection. Two early strains from their industrial lineages, SC2 and SC6, are known to differ in their CA titres. They also have different carbon source requirements. GSK would like to use sustainably sourced glucose as a source of carbon during fermentations. However, WT *S. clavuligerus* has a natural glucose auxotrophic phenotype and a thorough understanding of SC2 and SC6 central carbon metabolism is required for rational strain improvement. Given that the reason for auxotrophy is caused by insufficient expression of *glcP* (Pérez-Redondo *et al.*, 2010), we hypothesised that by heterologously expressing *glcP* in SC2 and SC6, the strains would be able to utilise glucose as a carbon source for biomass production. Glucose is not known to trigger CCR of antibiotic production in *S. clavuligerus* (Aharonowitz *et al.*, 1978; García Domínguez *et al.*, 1988; Lebrihi *et al.*, 1988), thus, repression of clavulanic acid biosynthesis by glucose was not expected to occur.

1.5.1 AIMS AND OBJECTIVES

The aim of this project was to develop a greater understanding of the glucose utilisation paradox in WT and industrial *S. clavuligerus* through the construction of glucose utilising *S. clavuligerus* strains. This aim was broken down into the following specific objectives:

- Computational approach to understanding the diversity of Actinobacterial GlcP and Glk
- Characterisation of carbon utilisation and CA production by SC2 and SC6
- Construction of SC2 and SC6 strains for heterologous expression of *glcP* and *glk* and phenotypic characterisation of these strains
- Investigation of protein-protein interactions between GlcP and Glk proteins

2 MATERIAL AND METHODS

Chemicals and media components used in this project were supplied by Thermo-Fisher, Sigma-Aldrich, VWR and Melford Biolaboratories Ltd. Enzymes and kits required for molecular biology work were provided by Promega, New England Biolabs and Qiagen. Suppliers are mentioned accordingly in parentheses.

2.1 TSOFTWARE AND BIOINFORMATICS

The software used to draw chemical structure was ACD/Chemsketch (available from https://www.acdlabs.com/resources/freeware/chemsketch/index.php). Protein sequences used for constructing phylogenetic trees and the corresponding genes were either extracted from the Actinobacterial database, ActDES (Schniete *et al.*, 2021), the KEGG database (Kanehisa *et al.*, 2000) or StrepDB (http://strepdb.*Streptomyces*.org.uk). Homologous proteins and genes in a database were searched for using BLAST, which was either accessed through the NCBI server or the Gnome terminal, an open source project set up by Miguel de Icaza in 1996 (Altschul *et al.*, 1990; German, 2003). JalView was used to carry out multiple sequence alignments (Clamp *et al.*, 2004; Waterhouse *et al.*, 2009). Genome loci were visualised in SnapGene (from GSL Biotech LLC). Accession numbers for the amino acid and nucleotide sequences that were not taken from the Actinobacterial database are listed in Table 2-1. The nucleotide and amino acid sequences were extracted from the European Nucleotide Archive (ENA), the Universal Protein Resource KnowledgeBase (UniProtKB) and the National Centre for Biotechnology Information (Boutet *et al.*, 2007; Leinonen *et al.*, 2010; Sayers *et al.*, 2020).

2.1.1 CONSTRUCTION OF PHYLOGENETIC TREES

The phylogeny of sequences was determined by aligning the sequences using MUSCLE (Edgar, 2004). Trees were calculated in MEGA X (Kumar *et al.*, 2016, 2018). The statistical model and method used were the LG model and Maximum likelihood, respectively, with a bootstrap of 1000 (Felsenstein, 1985; Le *et al.*, 2008). The calculated trees were exported as Newick files and edited in FigTree v1.4.4 (available at http://tree.bio.ed.ac.uk/software/figtree/) (Olsen, 1990).

Table 2-1: Proteins and genes that were extracted from the KEGG, StrepDB and NCBI databases. Strains and their identifiers are listed as they are used throughout this thesis. Gene and protein identifiers for the European Nucleotide Archive (ENA) and the Universal Protein Resource KnowledgeBase (UniProtKB), respectively, are shown.

Species	Strain		*glcP* Gene	ENA ID	UniProt ID	*glk* Gene	ENA	Uni-ProtKB
S. albulus	NK660	DC74	DC74_5710	AIA06160.1	A0A059W8R5	DC74_2611	AIA03113.1	A0A059W1L6
S. albus	DSM 41398	SLNWT	SLNWT_1654	AJE82030.1	A0A0B5ERV7	SLNWT_5756	AJE86132.1	A0A0B5F6W5
S. avermitilis	MA-4680	SAVERM	SAVERM_2657	BAC70368.1	Q82JU7	SAVERM_6074	BAC73785.1	Q82AI3
S. clavuligerus	ATCC 27064	SCLAV	SCLAV_4529	EFG09600.1	E2Q8Y6	SCLAV_1340	EFG06419.1	E2Q0M6
S. coelicolor	M145	SCO	SCO5578	CAA22421.1	Q9ZBQ1	SCO2126	CAB51974.1	P0A4E1
			SCO7153	CAC01642.1				
S. globisporus	C-1027	WQO	WQO_26035	ALU96485.1	A0A0U3C3W9	WQO_08430	ALU93378.1	A0A0U3LRS2
S. griseus	IFO 13350	SGR	SGR_1900	BAG18729.1	B1VYX4	SGR_5377	BAG22206.1	B1VZT1
S. hygroscopicus subsp. jinggangensis	5008	SHJG	SHJG_6698	AEY91965.1	H2JXH7	SHJG_3609	AEY88881.1	H2JWX8
S. lavendulae subsp. lavendulae	CCM 3239	SLAV	SLAV_11790	ATZ24222.1	A0A2K8PBV4	SLAV_26570	ATZ27106.1	A0A2K8PK38
S. lividans	TK24	SLIV	SLIV_03515	AIJ11727.1	D6EC77	SLIV_27075	AIJ16326.1	D6EIX7
			SLIV_10580	AIJ13116.1	A0A076LYP3			
S. rapamycinicus	NRRL 5491	M271	M271_15845	AGP54743.1	A0A0A0NF82	M271_35710	AGP58547.1	A0A0A0NG57
S. scabiei	87.22	SCAB	SCAB_26321	CBG69736.1	C9Z3S0	SCAB_67551	CBG73754.1	C9YTW6
S. tsukubaensis	NRRL 18488	B7R87	B7R87_24980	AZK96755	A0A5H2U845	B7R87_07225	AZK93690	A0A5H2UH54
S. venezuelae	NRRL B-65442.	SVEN/vnz	SVEN_5273/vnz_26080	CCA58559.1	F2R5F5	SVEN_1786/vnz_08755	CCA55073.1	F2RI90

43

Multi-locus phylogenetic trees were calculated and constructed using the autoMLST web server (Alanjary *et al.*, 2019) available at https://automlst.ziemertlab.com/. The closest related strains and an outgroup organism were selected automatically by the algorithm. An ultrafast bootstrap of 1,000 was applied to the concatenated alignment. The tree was exported in the newick format (.nwk) and edited in FigTree v1.4.4 as before.

2.1.2   SECONDARY AND TERTIARY STRUCTURE PREDICTION

Structure predictions for membrane proteins were carried out using PROTTER (secondary) and I-TASSER (tertiary, 3D) (Roy *et al.*, 2010; Omasits *et al.*, 2013; Yang and Zhang, 2015; Yang and Yan, *et al.*, 2015). Obtained .pdb files were viewed in Jmol Viewer (open-source Java viewer for chemical structures in 3D, available at http://www.jmol.org/).

2.1.3   SYNONYMOUS AND NON-SYNONYMOUS CHANGES IN *glcP* AND *glk* GENES

The Synonymous Non-synonymous Analysis Program SNAP, v2.1.1 (available from http://www.datamonkey.org/), was used for estimating the number of synonymous and non-synonymous changes per nucleotide site and the number of potential synonymous and non-synonymous changes (Bromberg *et al.*, 2008). The ratio of dN to dS was calculated manually. A Z-based test of neutrality (dN = dS) was applied to the sequences to calculate the statistical significance of a sequence being under positive (dN < dS) or purifying (dN > dS) selection. The statistical tests were carried out in MEGA X (Kumar *et al.*, 2018). Significant differences between dS and dN values and the dN/dS ratios obtained for *glcP* and *glk* were tested using a two-tailed, homoscedastic *t*-test.

2.1.4   PROMOTER SEQUENCE PREDICTIONS

Promoter predictions were carried out using the BProm webtool (available at www.softberry.com), in addition to manual comparison between known -35, -10 and RBS sequences upstream of *glcP* and *glk* genes. Sequences were aligned in JalView (Clamp *et al.*, 2004; Waterhouse *et al.*, 2009).

2.1.5   *IN SILICO* SEARCH FOR CARBOHYDRATE TRANSPORTER-ENCODING GENES IN *S. CLAVULIGERUS*

The systematic search for carbohydrate transporters was carried out as previously described for *S. coelicolor* (Bertram *et al.*, 2004), using the transporter classification database (available

at http://www.tcdb.org/) (Saier *et al.*, 2006, 2009, 2014, 2016). Hits were categorised according to the types of transport systems.

2.2   CULTIVATION AND BACTERIAL PHYSIOLOGY

Cells were generally grown in Universal tubes (supplied by Elkay Laboratory Products Limited (UK)) or 250 ml Erlenmeyer flasks (either supplied by GSK or purchased from Duran, Wertheim (DE)) and grown in shaking incubators. Cultures were transferred into sterile 50 ml centrifuge tubes (supplied by Corning Incorporated, New York (USA)) for centrifugation in the Heraeus Megafuge 40R. Smaller volumes were centrifuged in a Microfuge 16 Centrifuge by Beckman Coulter (benchtop centrifuge). Most media components and chemicals were provided by the following suppliers: Sigma-Aldrich Co., 3050 Spruce St, St. Louis (USA); Fisher Scientific UK Ltd (part of Thermo Fisher Scientific), Bishop Meadow Road Loughborough LE11 5RG (UK); VWR International Ltd. (part of Avantor), Hunter Boulevard, Magna Park, Lutterworth LE17 4XN (UK); Merck Chemicals, The Old Brickyard Gillingham Dorset SP8 4XT (UK); Melford Biolaboratories Ltd., Bildeston Rd, Ipswich (UK). Petri dishes and square plates were from Thermo Scientific and Gosselin, respectively.

2.2.1   BACTERIAL STRAINS

All bacterial strains that were used and/or generated during this project are listed in Table 2-2. *S. clavuligerus* and *S. coelicolor* strains were grown on L3M9 and MS agar, respectively, for sporulation and conjugation experiments. Ex-conjugants of *S. clavuligerus* and *S. coelicolor* were selected for on GYM agar and nutrient agar, respectively, containing the appropriate antibiotics. *Streptomyces* spores were generally heat-activated at 50°C for 10 minutes. Volumes of 100 - 200 µl spore stocks were added to 800 – 900 µl of TSB or inoculation medium in the case of *S. clavuligerus* or 2xYT in the case of *S. coelicolor* (Aharonowitz *et al.*, 1978; Kieser *et al.*, 2000). Heat-activation was carried in a dry bath (MD-MINI by Major Science). Heat-activated spores were used set up seed cultures for pre-germination, which were left to grow for 48 hours at 26°C and 24 hours at 30°C (250 rpm) for *S. clavuligerus* and *S. coelicolor*, respectively. *E. coli* cells were generally grown on LB-agar or cultured in LB broth (Sambrook *et al.*, 1989).

Table 2-2: Bacterial strains used and generated in this project.

Species	Genotype	Reference
S. clavuligerus DSM 738	wildtype	DSMZ
S. clavuligerus SC2	GSK production strain	GSK culture collection
S. clavuligerus SC2 ASB0	integrated pIJ6902 vector (φC31)	This study
S. clavuligerus SC2 ASB1	integrated pIJ6902+*glcP*(SCLAV)	This study
S. clavuligerus SC2 ASB2	integrated pIJ6902+*glk*(SCLAV)	This study
S. clavuligerus SC2 ASB3	integrated pIJ6902+*glcP*(SCO)	This study
S. clavuligerus SC2 ASB4	integrated pIJ6902+*glk*(SCO)	This study
S. clavuligerus SC6	GSK production strain	This study
S. clavuligerus SC6 ASB0	integrated pIJ6902 vector (φC31)	GSK culture collection
S. clavuligerus SC6 ASB2	integrated pIJ6902+*glk*(SCLAV)	This study
S. clavuligerus SC6 ASB3	integrated pIJ6902+*glcP*(SCO)	This study
S. clavuligerus SC6 ASB4	integrated pIJ6902+*glk*(SCO)	This study
S. coelicolor M145	Wildtype, SCP1-, SCP2-	(Feitelson *et al.*, 1983)
S. coelicolor M145 ASB0	integrated pIJ6902 vector (φC31)	this study
S. coelicolor M145 ASB1	integrated pIJ6902+*glcP*(SCLAV)	this study
S. coelicolor M145 ASB2	integrated pIJ6902+*glk*(SCLAV)	this study
S. coelicolor M145 ASB3	integrated pIJ6902+*glcP*(SCO)	this study
S. coelicolor M145 ASB4	integrated pIJ6902+*glk*(SCO)	this study
S. coelicolor M145 ASB8	integrated pIJ6902+*glk*(SGR)	this study
S. coelicolor BAP20	M145 Δ*glcP1*::apr Δ*glcP*2::hyg	Van Wezel *et al.*, 2005
S. coelicolor BAP20 ASB0	integrated pIJ6902 vector (φC31)	this study
S. coelicolor BAP20 ASB1	integrated pIJ6902+*glcP*(SCLAV)	this study
S. coelicolor BAP20 ASB2	integrated pIJ6902+*glk*(SCLAV)	this study
S. coelicolor BAP20 ASB3	integrated pIJ6902+*glcP*(SCO)	this study
S. coelicolor BAP20 ASB4	integrated pIJ6902+*glk*(SCO)	this study
E. coli DHh5α	*F- ,φ80, lacZΔM15, Δ(lacZYA-argF), U169, recA1, endA1, hsdR17, (rk-, mk+), phoA ,supE44, thi-1, gyrA96, relA1*	Grant *et al.*, 1990 Invitrogen
E. coli ET12567/pUZ8002	*dam-13::Tn9, dcm-6, hsdM, hsdR recF143, zij201::Tn10, galK2, galT22, ara14, lacYl, xylS, leuB6, thi-1, tonA31, rpsL136, hisG4, tsx78, mtli, glnV44, F-*	MacNeil *et al.*, 1992
E. coli BTH101	*F-, cya-99, araD139, galE15, galK16, rpsL1 (Str r), hsdR2, mcrA1, mcrB1, relA1*	Euromedex (ref. EUK001), Battesti *et al.*, 2012

2.2.1 Preparation of Glycerol Stocks

Stocks of *Streptomyces* spores were prepared by growing the strains until spores had formed. This took five to seven days for *S. coelicolor* on MS agar and up to four weeks for *S. clavuligerus* on L3M9. The spores were harvested into a 20% glycerol solution using a sterile cotton bud. The solution was then passed through a sterilised 20 ml syringe containing cotton wool to remove mycelial fragments. Once filtered, the spores were transferred into a cryogenic vial (provided by StarLab) and stored at -20°C (Kieser *et al.*, 2000). Stocks of *E. coli* cells were prepared from overnight cultures grown in LB medium. Aliquots of 500 µl of cells were mixed with 500 µl of 50% glycerol. The cells were shock-frozen in liquid nitrogen and stored at -80°C (Sambrook *et al.*, 1989).

2.2.2 Media

Unless otherwise stated, media were prepared and sterilised on the same day, either by autoclaving at 121°C for 20 minutes or by filter-sterilising using a 0.22 µm pore filter (MF-Millipore™ by Merck) and syringes provided by BD PlastiPak if ingredients were heat-sensitive. Recipes for the used media are shown in Table 2-3. Distilled water was provided by the Purite water system (Neptune Life Science). Carbon sources other than glucose that were added to media were balanced so that the molar carbon ratio was equivalent to that of glucose, i.e., if the recipe required 5 g/L of glucose (6 carbon atoms/molecule), 10 ml glycerol (3 carbon atoms/molecule) were added.

Antibiotic stocks and their working concentrations are listed in Table 2-4. Carbenicillin disodium salt and kanamycin monosulphate were purchased from Melford Biolaboratories Ltd. Thiostrepton, nalidixic acid sodium salt and chloramphenicol were supplied by Sigma-Aldrich Co. Apramycin sulphate and Hygromycin B were purchased from Duchefa Biochemie B.V, 2003 RV Haarlem (NL) and Invivo Gen, San Diego (USA), respectively.

Blue-white screening of *E. coli* required the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, supplied by Formedium as IPTG Dioxan Free powder) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal, supplied by PROMEGA as 50 mg/ml stock solution) to LB-agar plates at final concentrations of 0.1 mM and 40 µg/ml, respectively.

Table 2-3: Media used for cultivation of *Streptomyces* and *E. coli*.

Media		Reference
1x YT	16 g Tryptone 10 g Yeast extract 5 g NaCl 1000 ml distilled H_2O	Kieser *et al.*, 2000
Glucose Yeast extract Malt extract (GYM)	4 g Glucose 4 g Yeast extract 10 g Malt extract 2 g $CaCo_3$[2] 12 g Agar-agar[2] pH adjusted to 7.2	Kieser *et al.*, 2000
Basal inoculum medium	0.6 g $MgSO_4$ 3.5 g K_2HPO_4 100 μl of 10x trace salts stock solution[3] 10 ml glycerol (autoclaved separately and added just before use) 2 g L-asparagine 1 g Yeast extract 1 g NH_4Cl Made up to 1000 ml with 0.1 M 3-(N-morpholino) propane sulfonic acid (MOPS) buffer	Aharonowitz *et al.*, 1978
L3M9	0.3 g Dextrin 10 g α-α-Trehalose 0.5 g Di-potassium hydrogen orthophosphate 1 g NaCl 1 g $MgSO_4$ 0.5 g CaCl 2 g Casamino acids 11 g MOPS powder 1 ml 10x trace salts stock solution[4] 30 g Agar-agar made up to 1000 ml with distilled H_2O pH adjusted to 6.8 ± 0.2	GSK
Lysogeny Broth (LB)	10 g Tryptone 5 g Yeast extract 10 g NaCl 1000 ml distilled H_2O Added 10 g Agar-agar for solid medium	Bertani, 1951
Mannitol Soya Flour (MS) agar	20 g Soya Flour (Holland and Barrett) 20 g Mannitol 20 g Agar-agar 1000 ml tap water	Hobbs *et al.*, 1989
Milk agar	50 g Dry milk powder (Morrison's) 5 g Tryptone 2.5 g Yeast extract	Modified from Hardy Diagnostics

---

[2] Only added when preparing solid version of this media.
[3] Contained 1 g each of $FeSO_4.7H_2O$, $MnCl_2.4H_2O$, $ZnSO_4.H_2O$ and $CaCl_2$ per 100 ml distilled H_2O.
[4] Contained 0.5 g each of $FeSO_4$, $MnSO_4$, $ZnSO_4$ per 50 ml of distilled H_2O.

	12.5 g Agar-agar 5 g Glucose[5] pH adjusted to 6.8 – 7.0	
Minimal liquid medium (NMMP)	0.4 g $(NH_4)_2SO_4$[6] 0.5 g Casamino acids 0.6 $MgSO_4.7H_2O$ 50 g PEG 6000 100 µl Minor elements solution[7] Dissolve in 800 ml distilled H_2O After autoclaving add 150 ml NaH_2PO_4 / K_2HPO_4 buffer (0.1M, pH 6.8)[8] 5 g Glucose in 50 ml distilled H_2O[9]	Modified from Hodgson, 1982
Minimal Medium	1 g $(NH_4)_2SO_4$ 0.5 g K_2HPO_4 0.2 g $MgSO_4.7H_2O$ 0.01 g $FeSO_4.7H_2O$ 10 g Agar-agar 5 g Glucose[4] 1000 ml distilled H_2O Adjusted pH to 7.0-7.2 prior to adding agar and carbon source	Modified from Hopwood, 1967
Nutrient agar	21 g nutrient broth No.1 powder (Sigma) 10 g Agar-agar 1000 ml distilled H_2O	
Tryptone Soya Broth (TSB)	30 g TSB powder 1000 ml distilled H_2O	

---

[5] Carbon sources were added from 50% (glucose) and 80% (glycerol) stock solutions, which were either filter-sterilised or autoclaved separately.

[6] Not added for ammonium-free version of this media.

[7] Contained 1 g each of $ZnSO_4.7H_2O$, $FeSO_4.7H_2O$, $MnCl_2.4H_2O$, $CaCl_2$, anhydrous in 100 ml distilled H_2O

[8] Prepared my combining equal volumes of 0.1 M NaH_2PO_4 and K_2HPO_4 solutions and adjusting the pH to 6.8, buffer was autoclaved separately.

[9] Prepared separately and filter-sterilised before combining with the rest of the media.

Table 2-4: Antibiotic stocks and working concentrations.

Antibiotic	Stock[10] [mg/ml]	Final concentration [μg/ml]
Carbenicillin	100	100
Apramycin	50	50
Kanamycin	50	50
		25
Chloramphenicol	25 in Ethanol	25
Thiostrepton	50 in DMSO	50
	0.5 in DMSO	0.5
Nalidixic acid	25	25
Hygromycin	50	50
		100

[10] Dissolved in distilled H_2O unless otherwise specified

Streptomyces growth curves based on absorbance readings (600 nm) were carried out in multi-well plates, either 24-well (Nunc, Sigma-Aldrich) or 96-well plates (Tpp, Trasadingen, (CH) and shaken continuously for up to 40 hours at a constant temperature (26°C for *S. clavuligerus* and 30°C for *S. coelicolor*) in the Synergy HT microplate reader by BioTek. The multi-well plates were covered with corresponding lids. Prior to inoculating the multi-well plates, spores were heat-activated and pre-germinated in a seed culture as described previously (see 2.2.1). If the final culture medium differed from the pre-germination medium, the cells were washed twice in 0.25M TES buffer. The buffer was prepared by dissolving 57.3 g/L of TES free acid (Sigma) in distilled water, adjusting the pH to 6.8. and autoclaving. The cells were only washed once in fresh medium if the pre-germination medium was the same as the final culture medium.

After washing, the cell pellets were resuspended in 1 ml of final culture medium and the optical density (OD) at 600 nm was determined spectrophotometrically (Biochrom WPA Biowave CO800 Cell-Density Meter). The cell solutions were then normalised to an OD_{600} of 0.4. The working volumes in 24-well and 96-well plates were 1.5 ml and 0.2 ml, respectively. The normalised cell solutions were added to the medium-containing wells at a 1:9 ratio. Previous work has shown that an OD_{600} of 0.04 corresponds to 1.5×10^5 pre-germinated spores (Schniete, 2015). Following cultivation, cells and supernatants were collected from the wells for further analyses, which consisted of manual absorbance readings, determination of cell dry weight (CDW), spotting onto agar plates, RNA isolation, determination of CA concentrations and protease activity.

Data were analysed in Excel; the standard error was determined for each set of values / time point. The natural logarithm (LN) of each value was determined before plotting the averaged LN values against the time. The specific growth rate μ h^{-1} of a culture was calculated manually using the following formula: $\ln N_t - \ln N_0 = \mu(t-t_0)$, with N_t and N_0 as the culture concentrations at timepoints t and t_0, respectively, that mark the time in hours at the end and start of exponential phase. Statistical significance across all strains was determined using a one-way ANOVA, significance between two sets of determined growth rates was conducted with

Tukey's HSD test, using the following tool: https://www.statskingdom.com/180Anova1way.html.

2.2.4 LARGE SCALE *STREPTOMYCES* GROWTH CURVES

Large scale growth curves of *S. clavuligerus* were carried out in 250 ml baffled Erlenmeyer flasks (provided by GlaxoSmithKline) containing 50 ml of culture medium. The cultures were inoculated with an OD_{600} of between 0.01 and 0.001 depending on the OD_{600} of the seed culture. Samples were taken as required for absorbance readings (600 nm) and determination of CDW. Downstream analysis of the taken samples was the same as described in the previous section. Data analysis was carried out as described for small scale growth curves (see 2.2.3).

2.2.5 CELL DRY WEIGHT DETERMINATION

The dry weight of cells contained in 1 ml of a culture sample was determined by firstly washing the cells twice before resuspending the pellet in 1 ml of distilled water. The cells were applied to a micro glass fibre filter (purchased from Fisherbrand, 1 μm pore size) that had dried to a constant mass. The biomass was filtered through a Buchner Funnel (KIF Laboport and NALGENE 180 PVC metric tube) and subsequently washed with five volumes of distilled water. The filter containing the biomass was dried to a constant mass. The biomass was determined by subtracting the mass of the filter from the filter containing the biomass (mg/ml).

2.2.6 SPOTTING ONTO AGAR PLATES

Pre-germinated spores and cell solutions that had been normalised to an OD_{600} were used for spotting onto milk agar or minimal medium agar plates. For this, 1 μl of the solutions was spotted onto 6-well plates (NUNC, Sigma-Aldrich), which contained 1.5×10^3 spores or cells. The plates were incubated for up to 4 weeks and pictures were taken were taken with a Canon EOS 1200D camera.

2.2.7 CA ASSAY

CA concentrations in culture supernatant were determined according to a protocol provided by GSK. Briefly, 8 μl of supernatant that had previously been cleared by centrifugation for 5

minutes at top speed in a benchtop centrifuge, were added to 200 μl of imidazole. The imidazole solution as prepared by dissolving 10 g imidazole (VWR Chemicals) in 7.2 ml 10M HCl and made up to a final volume of 100 ml with distilled water, pH 7.0±0.2. CA standard solutions of the following concentrations were assayed along with the samples: 400 μg/ml, 350 μg/ml, 300 μg/ml, 250 μg/ml, 200 μg/ml, 100 μg/ml, for which the appropriate amount of lithium clavulanate (supplied by GSK) was dissolved in distilled water and filter sterilised. The standard solutions were stored at -80°C. Samples (technical triplicates) were analysed by measuring absorbance at 324 nm following a 30-minute incubation at room temperature. A calibration curve was constructed from the CA standards, which was used to determine CA concentrations in the culture supernatants. The data were plotted in Excel with error bars indicating the standard deviation calculated for each set of values. Statistical significance was determined using either a one-way ANOVA (yields from more than two strains) or a Tukey's HSD test for comparing yields between two strains (webtool: https://www.statskingdom.com/180Anova1way.html).

2.2.8 AZOCASEIN PROTEASE ASSAY

Protease activity from culture supernatant cleared as described previously was determined according the protocol by Ginther adapted for *S. clavuligerus* (Ginther, 1979; Bascarán *et al.*, 1990). Supernatant aliquots of 10 μl was mixed with 90 μl 0.2M Tris-HCl (prepared by dissolving 24.2 g Tris Base in 1 L dH$_2$0 and adjusted the pH to 7.2 using 1M HCl) and 1 mM CaCl$_2$ (prepared by dissolving 0.147 g in 1 L dH$_2$0) containing 1% (w/v) azocasein protease substrate. Equal volumes of culture medium and proteinase K (1 μl in 19 μl water) served as negative and positive control, respectively. Biological and technical replicates, either duplicates or triplicates, were analysed alongside each other. The assay mixtures were incubated at 35°C for 1 hour. The reactions were stopped using 100 μl 10% trichloroacetic acid (diluted 100% trichloroacetic acid in water) and subsequently centrifuged at 3,000 xg for 15 minutes. From the supernatant, 80 μl was transferred into a microtiter plate containing 20 μl 1.8M NaOH (prepared by dissolving 7.2 g NaOH in 100 ml dH$_2$0) and mixed by pipetting. The absorbance was read at 440 nm. Protease activity was calculated as the change in absorbance of a sample compared with medium only as a function of the biomass. The data were plotted in Excel with error bars indicating the standard deviation calculated for each set of values. Statistical significance was tested using a one-way ANOVA for differences across

all strains. Statistically significant differences between two strains were tested using a Tukey's HSD Test (webtool: https://www.statskingdom.com/180Anova1way.html*)*.

2.3   MOLECULAR BIOLOGY

The enzymes and kits used for this study were all from Promega, New England Biolabs, Bioline and Qiagen. Suppliers and pieces of equipment are mentioned accordingly. Plasmids and PCR products were sequenced by Eurofins (Cologne, Germany).

2.3.1   PLASMIDS, COSMIDS, GBLOCKS AND OLIGOS

The plasmids used and generated in this study are shown in Table 2-5. The original plasmids all came from the Hoskisson group's plasmid collection. Oligos and gBlocks were obtained from Integrated DNA Technologies (IDT, WC1A 9NR London, UK).

2.3.1   PLASMID MINIPREP

Plasmids were isolated from bacterial cells by inoculating overnight cultures of 5 – 7 ml of LB broth (containing the appropriate antibiotic(s)) that had been inoculated with a single colony from the original transformation plate. Overnight cultures were grown in a shaking incubator at 37℃ and 300 rpm. The cultures were centrifuged for 7 minutes at 4200 rpm and the supernatant was discarded. Plasmid extractions were carried out according to Wizard® Plus SV Minipreps DNA Purification System from Promega. DNA concentrations and OD_{260}/OD_{280} ratios were determined as before.

2.3.2   GENOMIC DNA ISOLATION

Genomic DNA was isolated from two to three-day-old bacterial cultures using the Isolate II genomic DNA kit by Bioline and was carried out following the supplier's manual for hard-to-lyse bacteria. The quantity and quality of genomic DNA was determined using a NanoDrop 2000c spectrophotometer.

Table 2-5: Overview of plasmids used on this project.

Plasmid	Size [kb]	Features	References
pIJ6902	7.3	ΦC31, ori ColE1, a*pra^R*, t*hio^{R,}* p*tipA*	Huang *et al.,* 2005
pIJ6902-ASB1	8.7	ΦC31, ori ColE1, a*pra^R*, t*hio^R*, p*tipA*, *sclav_4529* (*glcP*)	This study
pIJ6902-ASB2	8.2	ΦC31, ori ColE1, a*pra^R*, t*hio^R*, p*tipA*, *sclav_1340* (*glk*)	This study
pIJ6902-ASB3	8.7	ΦC31, ori ColE1, a*pra^R*, t*hio^R*, p*tipA*, *sco5578* (*glcP*)	This study
pIJ6902-ASB4	8.2	ΦC31, ori ColE1, a*pra^R*, t*hio^R*, p*tipA*, *sco2126* (*glk*)	This study
pIJ6902-ASB8	8.2	ΦC31, ori ColE1, a*pra^R*, t*hio^R*, p*tipA*, *sgr_1900* (*glk*)	This study
pKT25	3.4	ori p15A, *kan^R*, C-terminal tag for BACTH	Karimova *et al.,* 1998, Euromedex
pKT25-Zip	3.5	ori p15A, *kan^R*, C-terminal leucine zipper region	Karimova *et al.,* 1998, Euromedex
pKT25-ASB1	4.8	ori p15A, *kan^R*, C-terminal tag, *sclav_4529* (*glcP*)	This study
pKT25-ASB2	3.4	ori p15A, *kan^R*, C-terminal tag, *sclav_1340* (*glk*)	This study
pKT25-ASB3	4.8	ori p15A, *kan^R*, C-terminal tag, *sco5578* (*glcP*)	This study
pKT25-ASB4	3.4	ori p15A, *kan^R*, C-terminal tag, *sco2126* (*glk*)	This study
pUT18	3.0	ori ColE1, *amp^R*, N-terminal tag for BACTH	Karimova *et al.,* 1998, Euromedex
pUT18-ASB1	4.4	ori ColE1, *amp^R*, N-terminal tag, *sclav_4529* (*glcP*)	This study
pUT18-ASB2	3.9	ori ColE1, *amp^R*, N-terminal tag, *sclav_1340* (*glk*)	This study
pUT18-ASB3	4.4	ori ColE1, *amp^R*, N-terminal tag, *sco5578* (*glcP*)	This study
pUT18-ASB4	3.9	ori ColE1, *amp^R*, N-terminal tag, *sco2126* (*glk*)	This study
pUT18C	3.0	ori ColE1, *amp^R*, C-terminal tag for BACTH	Karimova *et al.,* 1998, Euromedex
pUT18C-Zip	3.1	ori ColE1, *amp^R*, C-terminal leucine zipper region	Karimova *et al.*, 1998, Euromedex
pUT18C-ASB1	4.4	ori ColE1, *amp^R*, C-terminal tag, *sclav_4529* (*glcP*)	This study
pUT18C-ASB2	3.9	ori ColE1, *amp^R*, C-terminal tag, *sclav_1340* (*glk*)	This study
pUT18C-ASB3	4.4	ori ColE1, *amp^R*, C-terminal tag, *sco5578* (*glcP*)	This study
pUT18C-ASB4	3.9	ori ColE1, *amp^R*, C-terminal tag, *sco2126* (*glk*)	This study
pGEM-T Easy	3.0	ori ColE1, *amp^R*, MCS within β-galactosidase	Promega

2.3.1 RNA EXTRACTION

Culture samples (1 ml) were taken and centrifuged in a benchtop centrifuge at top speed for 5 minutes. The supernatant was removed, and the cell pellets were carefully resuspended in 500 μl RNA later (SigmaAldrich). The pellets were stored at -20°C until extraction. To extract the RNA from the samples, the tubes were thawed on ice before centrifuging as before. The RNA later solution was removed, and the cells were resuspended in 135 μl of TE buffer (pH 8.0, supplied by SigmaAldrich) containing 15 μl of 50 mg/ml lysozyme. The lysozyme stock was prepared by dissolving 50 mg of lysozyme powder supplied by Sigma to 1 ml of sterile water and sterilised by filtration. Cell lysis was carried out at 37°C for 1 hour and stopped by adding 525 μl of RTL buffer provided by the Quiagen RNAeasy mini kit, to which 10 μl β-mercaptoethanol (Sigma) had been added. The samples were mixed, centrifuged and the supernatant transferred into a new microcentrifuge tube to which 375 μl absolute ethanol was added. RNA extraction using the RNAeasy mini kit column was carried out following the supplier's protocol, skipping the on-column DNase I digest step. This step was carried out off-column with 40 μl RNA/DNA samples using the NewEngland Biolabs DNase I enzyme. The digest was carried out according to the supplier's manual. The reactions were stopped by adding 2 μl of 0.25 M EDTA (final concentration of 5 mM EDTA/sample). The resulting RNA was cleaned up using the RNAeasy mini kit column as before with the difference that the RNA was eluted in 40 μl RNase-free water twice. RNA concentrations and OD_{260}/OD_{280} ratios were determined using the NanoDrop 2000c spectrophotometer.

2.3.1 POLYMERASE CHAIN REACTIONS

Polymerase chain reactions (PCR) were either carried out with the Q5 high fidelity DNA polymerase provided by New England Biolabs, the taq-based polymerase OneTaq by Promega or the hot-start taq polymerase provided in the OneStep Reverse Transcriptase-PCR kit by Quiagen. The reactions were set up according to the manufacturer's protocols with around 100 ng of genomic DNA (gDNA), 10-30 ng of plasmid DNA or 50 ng of RNA. The PCR programmes varied depending on the polymerase and downstream application of the PCR products. In general, three distinctly different PCR programmes were used: (1) amplification of genes from gDNA or plasmids for cloning using the Q5 polymerase, (2) colony PCRs using the OneTaq and (3) Reverse Transcriptase PCRs from RNA.

Figure 2-1 shows schematics of the PCR programmes used for these applications. PCRs were carried out in the Applied Biosystems Veriti 96-well thermal cycler or the 2720 thermal cycler. Primers are listed Table 2-6. PCR products were visualised by agarose gel electrophoresis (see 2.3.3). Gene sequences for designing primers were taken either from KEGG or StrepDB. The primers were designed using the The SnapGene® software (from GSL Biotech; available at snapgene.com).

2.3.1.1 AMPLIFICATION OF GENES FOR CLONING PURPOSES

The amplification of DNA from genomic *Streptomyces* DNA required the addition of either 0.5% dimethyl sulfoxide (DMSO) or the appropriate amount of the GC enhancer solution (OneTaq)/Q solution (RT kit) provided by the enzyme kit. The schematics show the programmes for when a touchdown PCR was carried out. Amplification of bands using the *glcP* and *glk* primers from plasmids were conducted at 62°C and 58°C, respectively. The extension times of all programmes varied with the expected size of the amplified product (sizes are shown in Table 2-6).

2.3.1.2 COLONY PCRS FOR *STREPTOMYCES* AND *E. COLI* COLONIES

Colony PCRs of *Streptomyces* were set up by dissolving colonies in 50 µl of 50% DMSO and boiling for 30 minutes at 60°C prior to mixing the PCR reactions. Colony PCRs of *E. coli* were carried by picking a single colony from a plate, transferring it to a PCR tube containing 50 µl of distilled, nuclease-free water and mixing. Subsequently, 5 µl of cell solution was used for PCR, which was carried after a 5 minute boiling step at 94°C (Woodman, 2008). To ensure amplification, a touchdown approach was often taken with the annealing temperature starting from 60°C and decreasing to 45°C.

2.3.1.3 REVERSE TRANSCRIPTASE (RT-) PCR

RT-PCRs were carried out using the OneStep RT-PCR kit by Qiagen with RNA extracted from liquid cultures serving as templates. To ensure the samples were not contaminated with gDNA, additional reactions were set up that were not incubated at 50°C (reverse transcription step).
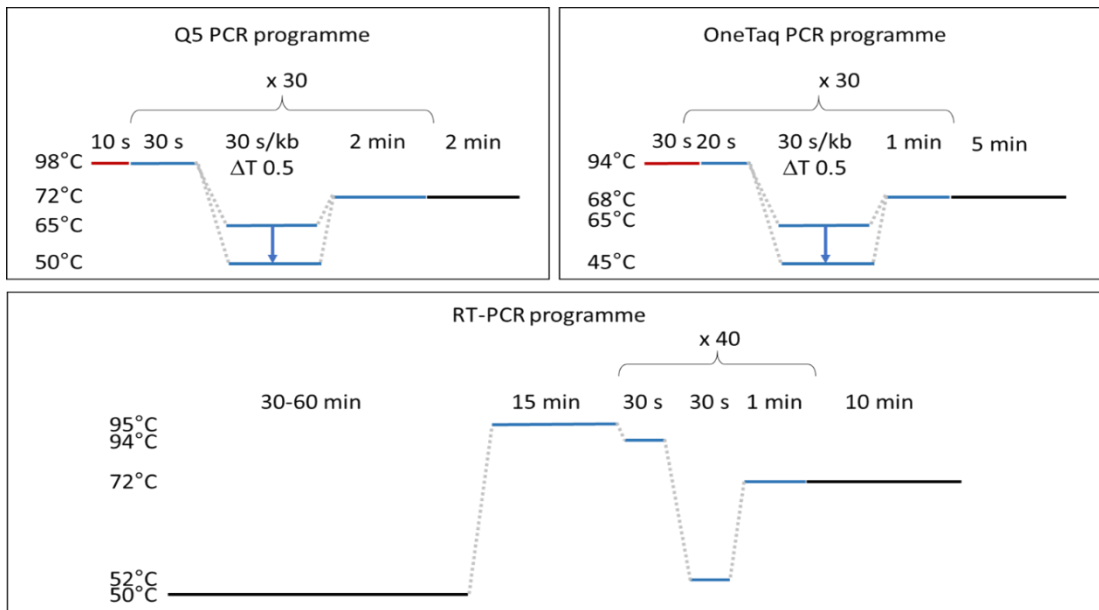
Figure 2-1: Schematic of PCR programmes. Three different types of PCR programmes were used for Q5, OneTaq or the hot start polymerase for RT-PCR. The 30- to 60-minute stage at 50°C during the RT-PCR programme was the reverse transcriptase step. The reverse transcriptase was heat-inactivated during the subsequent 15-minute 94°C stage.

Table 2-6: Primers used for cloning, confirmation of DNA and reverse transcriptase PCR.

Gene / Template	Orientation	Sequence	Restriction site	Size of Amplicon	Application
sclav_4529 (glcP)	Forward	GTACCATATGGTGAGCAGCAGGACTCCCC	NdeI	1.4 kb	cloning into pIJ6902
	Reverse	GTACAGATCTTCAGCCCATCTCCTCCAGCG	BglII		
sclav_1340 (glk)	Forward	GTACCATATGATGGGACTCACCATCGGCGT	NdeI	0.9 kb	cloning into pIJ6902
	Reverse	GTACAGATCTTCAGCCCTGGCGGGCGAG	BglII		
sco5578 (glcP)	Forward	GTACCATATGGTGGCCAGCACATCGCAGGC	NdeI	1.4 kb	cloning into pIJ6902
	Reverse	GTACAGATCTTCAGCCCATTTCCTCCAGGGC	BglII		
sco2126 (glk)	Forward	same as SCLAV_1340 Forward	NdeI	0.9 kb	cloning into pIJ6902
	Reverse	GTACAGATCTTCACATGATCGGGTCGGGTTCTC	BglII		
sven_5273 (glcP)	Forward	GTACCATATGGTGTCCAATACCGCGCAGGC	NdeI	1.4 kb	cloning into pIJ6902
	Reverse	GTACAGATCTTTAGCCCATCTCCTCCAACGCCT	BglII		
sven_1786 (glk)	Forward	same as SCLAV_1340 Forward	NdeI	0.9 kb	cloning into pIJ6902
	Reverse	same as SCLAV_1340 Reverse	BglII		
sgr_1900 (glcP)	Forward	GTACCATATGGTGACCAGCACTGCGAACGGAC	NdeI	1.4 kb	cloning into pIJ6902
	Reverse	GTACAGATCTCTACCCCATCTCCTCCAACGC	BglII		
sgr_5377 (glk)	Forward	same as SCLAV_1340 Forward	NdeI	0.9 kb	cloning into pIJ6902
	Reverse	GTACAGATCTTCAGCCCTGGCGGGCCAG	BglII		
sclav_4529 (glcP)	Forward	GCATTCTAGACGTGAGCAGCAGGACTCCC	XbaI	1.4 kb	BACTH: cloning into pKT25 and pUT18C
	Reverse	CTAGGGATCCTCAGCCCATCTCCTCCAGCGG	BamHI		
	Forward	GCATTCTAGAGTGAGCAGCAGGACTCCCCCGT	XbaI		BACTH: cloning into pUT18
	Reverse	CTAGGGATCCAGGCCCATCTCCTCCAGCGG	BamHI		
sclav_1340 (glk)	Forward	GCATTCTAGACATGGGACTCACCATCGGCG	XbaI	0.9 kb	BACTH: cloning into pKT25 and pUT18C
	Reverse	CTAGGGATCCTCAGCCCTGGCGGGCGAG	BamHI		
	Forward	GCATTCTAGAATGGGACTCACCATCGGCG	XbaI		BACTH: cloning into pUT18
	Reverse	CTAGGGATCCGAGCCCTGGCGGGCGAGGTC	BamHI		
sco5578 (glcP)	Forward	GCATTCTAGACGTGGCCAGCACATCGCAG	XbaI	1.4 kb	BACTH: cloning into pKT25 and pUT18C
	Reverse	GACTGAATTCTCAGCCCATTTCCTCCAGGGC	EcoRI		
	Forward	GCATTCTAGAGTGGCCAGCACATCGCAGG	XbaI		BACTH: cloning into pUT18
	Reverse	GACTGAATTCTCGCCCATTTCCTCCAGGGC	EcoRI		
sco2126 (glk)	Forward	same as SCLAV_1340 Reverse (BACTH, pKT25)	XbaI	0.9 kb	BACTH: cloning into pKT25 and pUT18C
	Reverse	CTAGGGATCCTCACATGATCGGGTCGGGTTCT	BamHI		
	Forward	same as SCLAV_1340 Reverse (BACTH, pUT18)	XbaI		BACTH: cloning into pUT18
	Reverse	CTAGGGATCCTCACATGATCGGGTCGGGTTCT	BamHI		

Gene expression of *glcP* and *glk* was analysed semi-quantitatively by measuring mean brightness values of each band in ImageJ (https://imagej.nih.gov/ij/; ImageJ Control Panel > Analyze > Measure). The brightness of *glcP* and *glk* bands were subsequently normalised to the respective *hrdB* band, i.e., the brightness of the *glcP* band from DSM 738 was divided by the brightness obtained for *hrdB* from DSM 738. The bar charts were constructed in Excel.

2.3.2   RESTRICTION DIGESTS

Restriction enzymes were obtained from Promega and New England Biolabs. Digests were carried out using 100 ng to 1 mg of DNA and were set up according to the suppliers' specifications. Digests were incubated in Eppendorf tube heat blocks and subsequently analysed by agarose gel electrophoresis.

2.3.3   AGAROSE GEL ELECTROPHORESIS

Agarose gels of 0.8 – 1.8% were prepared by dissolving agarose (UltraPure) in 1x TAE buffer, which was diluted from a 50x TAE stock solution. The stock was prepared by combining 242 g Tris Base, 57.1 ml glacial acetic acid (Merck) and 200 ml 0.25 M EDTA stock solution and making the final volume of 1000 ml up with distilled water. The agarose in 1x TAE buffer was melted in the microwave and allowed to cool to hand-warm prior to adding ethidium bromide (10μl/100ml 1% agarose). Loading dye (6x, Promega or NEB) was added to the samples before loading them onto the gel. Gels were run at 80 - 120V (PowerPac by Bio-Rad) for 30 – 90 minutes. DNA bands were visualised with the VWR Genosmart computer and H6Z0812 camera.

2.3.4   DNA GEL EXTRACTION AND PCR PRODUCT CLEAN UP

DNA extraction from agarose gels and purification of PCR products were carried out using the PROMEGA Wizard SV Gel and PCR Clean-Up System. In the first case, the band of interest was excised from the gel and transferred into a microcentrifuge tube and the agarose gel was melted before proceeding with the supplier's protocol. This step was skipped when cleaning up PCR products. Concentrations of DNA and OD_{260}/OD_{280} ratios were determined using the NanoDrop 2000c spectrophotometer as before.

2.3.5 Ligations of Vectors and Inserts

Ligation of PCR products with linearised vectors was conducted using the T7 ligase from Promega. The amount of DNA required was calculated with the following equation:

$$\frac{ng\ vector\ x\ kb\ size\ of\ insert}{kb\ size\ of\ vector} \ x\ insert : vector\ molar\ ratio = ng\ of\ insert$$

The insert to vector ratio was increased for hard-to-clone inserts. The ligation reaction mixtures were either incubated overnight at 4°C or at 24°C for 2–3 hours and subsequently used to transform competent *E. coli* cells. 50 – 100 ng of DNA was used to transform 50 µl of competent *E. coli*.

2.3.6 Preparation of chemically competent and electrocompetent *E. coli* cells

A culture of *E. coli* was grown overnight and used to inoculate 25 ml of LB the following day, which was left to grow at 37°C, 250 rpm, until an OD_{600} of 0.4 – 0.6 was reached. The cells were transferred into 50 ml centrifuge tubes and incubated on ice for 10 minutes. They were then washed twice in ice cold 0.1 M $CaCl_2$ (prepared by dissolving 11 g $CaCl_2$ in 1000 ml distilled water) and finally resuspended in 5 ml 0.1 M $CaCl_2$ containing 15% glycerol. The cells were aliquoted into cryovials and shock-frozen in liquid nitrogen. The vials were stored at -80°C (Sambrook *et al.*, 2006).

2.3.7 Heat-Shock Transformation of *E. coli*

Aliquots (50 µl) of competent *E. coli* cells were either prepared in the laboratory or bought from Invitrogen (Subcloning Efficiency Dh5α). The cells were thawed on ice prior to adding 50-100 ng of DNA with which the cells were incubated on ice for 25 minutes. The cells were then heat-shocked for 45 seconds at 42°C and incubated on ice for another 2 minutes. Subsequently, LB broth (500 µl) was added and the cells were left to recover for 45 – 60 minutes at 37°C, 250 rpm. The cells were briefly centrifuged at 6,000 rpm and most of the supernatant was removed before plating them on LB agar plates containing the appropriate antibiotics. Plates were incubated for 16-18 hours at 37°C.

Conjugation was used for the intergenic transfer of plasmids based on pIJ6902, which carries the ϕC31 integrase gene from the *E. coli* strain ET12567/pUZ8002 to *Streptomyces* strains and was carried out as described by Kieser (Bierman *et al.*, 1992; Kieser *et al.*, 2000). Firstly, the methylation-deficient *E. coli* strain ET12578/pUZ8002 was transformed with the plasmid to be conjugated via heat-shock. pIJ6902 contains the origin of transfer sequence, *oriT*, and the gene encoding TraJ, both of which are required for transfer of the plasmid from the donor to the recipient cell (Fürste *et al.*, 1989; Ziegelin *et al.*, 1989). Transfer functions were provided by the *E. coli* strain through the non-transmissible vector pUZ8002 (Kieser *et al.*, 2000). Transformed *E. coli* were selected for with kanamycin (25 μg/ml), chloramphenicol (25 μg/ml) and apramycin (50 μg/ml). Single colonies were subsequently checked by colony PCR. Overnight cultures of the transformed *E. coli* were set up in LB broth containing the same concentrations of kanamycin, chloramphenicol and apramycin. The following day, the overnight cultures were used to inoculate 25 ml of LB broth containing antibiotics. Cells for the overnight cultures were added to the medium in a ratio of 1:25, which is less diluted than described by (Kieser *et al.*, 2000). The cultures were grown at 37°Cm 250 rpm for 5 to 8 hours before they reached an OD$_{600}$ of between 0.4 and 0.6. The cells were then washed twice by centrifuging at 4,200 rpm for 8 minutes at 4°C and resuspending the pellets in fresh LB without antibiotics. Following the last wash step, cell pellets were resuspended in a volume of fresh LB broth (no antibiotics) that would allow adding 500 μl of *E. coli* cells to 500 μl of heat-activated *Streptomyces* spores. The bacteria were gently mixed by pipetting and subsequently centrifuged at 6,000 rpm for 2 minutes. Most of the supernatant was removed and the pellets resuspended in the remaining medium before plating onto either L3M9 agar and MS agar containing 10mM MgCl$_2$ (supplied by Sigma-Aldrich, added from 1M stock) for *S. clavuligerus* and *S. coelicolor*, respectively. The plates were incubated at 30°C for no longer at 18 hours and subsequently overlaid with 1 ml of sterile water containing 0.5 mg/ml nalidixic acid and the appropriate antibiotic for selection. The plates were incubated for a further 5 to 10 days at either 26°C (*S. clavuligerus*) or 30°C (*S. coelicolor*).

2.3.9 QUALITATIVE BACTERIAL TWO HYBRID ASSAY

Qualitative bacterial two hybrid (BACTH) assays were carried out by co-transforming around 25 ng of each plasmid into competent *E. coli* BTH101 via heat-shock. Transformed cells were plated onto LB containing 50 µg/ml kanamycin and ampicillin, 100 µg/ml streptomycin, as well as 40 µg/ml 5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside (X-Gal). Transformations were incubated for up to 48 hours at 30°C (Karimova *et al.*, 1998). Plasmids containing leucine zipper sequences, namely pKT25-zip and pUT18-zip, were used as positive controls while combinations of pKT25 and pUT18C, or pKNT25 and pUT18 served as negative controls. Multiple colonies were picked from each transformation plate and re-streaked onto square plates of the same medium, from which two colonies were subsequently chosen to use for quantification.

2.3.10 QUANTIFICATION OF BACTH INTERACTIONS

Colonies of *E. coli* BTH101 that had previously been co-transformed with pKT25/pUT18 plasmids, were transferred into wells of a 96-well microtiter plate containing 250 µl LB and 50 µg/ml kanamycin, ampicillin and streptomycin. The cells were grown in a plate reader at 37°C and continuous fast shaking for around five hours and subsequently used to inoculate polypropylene blocks containing 1 ml LB and the antibiotics, as well as 0.5 mM IPTG. The IPTG was added from a 0.5 M stock, which was prepared by dissolving 1.19 g in 10 ml distilled water and sterilising by filtration. The polypropylene blocks were placed in a shaking incubator set to 37°C and 220 rpm to grow overnight. The following morning, absorbance readings of the overnight cultures were taken by transferring 50 µl from the blocks to a fresh 96-well microtiter plate containing 150 µl LB. Fresh polypropylene blocks were filled with 1 ml lysis buffer and 40 µl chloroform. The lysis buffer was made up on the day of the assay and contained 100 ml Z-buffer, 270 µl β-mercaptoethanol and 50 µl 10% sodium dodecylsulfate (SDS). The Z-buffer was prepared by dissolving 16.1g Na_2HPO_4, 5.5g $NaH_2PO_4.H_2O$, 0.75g KCl, 0.246g $MgSO_4.7H_2O$ in 1000 ml distilled water and adjusting the pH to 7.0. The SDS solution (10%) was prepared by dissolving 1 g of 99% SDS in 10 ml of water. Volumes of 70 µl from the overnight cultures were added to the lysis buffer-containing blocks. The cells were mixed with the lysis buffer by pipetting up and down 15 times. Subsequently, 100 µl were transferred from each well into a fresh microtiter plate and 50 µl

substrate were added. The substrate solution contained 4 mg of O-nitrophenyl β–D-galactopyranoside (ONGP) per ml of Z-buffer. The timepoint of substrate addition was recorded along with the timepoint of stopping the reaction by adding 50 µl of a stop solution (1 M Na_2CO_3, prepared by dissolving 105.9 g in 1000 ml of distilled water). Absorbance readings at 420 nm and 550 nm were taken for each well. Enzymatic activity was calculated according to the formula shown below:

$$1000 * \frac{OD420 - (1.75 * OD550)}{t(min) * V(ml) * A600}$$

With t(min) being the time between adding the substrate and stopping the reaction and V(ml) being the volume of culture added to the lysis buffer. Co-transformations were analysed in biological and technical duplicates (Miller, 1972; Griffith *et al.*, 2002). The data were plotted in Excel with error bars indicating the standard deviation calculated for each set of values. Statistical significance was tested using a one-way ANOVA when comparing β-galactosidase activities from more than two strains. A *t*-test was used to compare activities of two strains.

3 Phylogeny and Computational Analysis of the Streptomycete Glucose Uptake System

Most *Streptomyces* internalise and catabolise glucose via glycolysis (Hodgson, 2000). *S. clavuligerus*, however, does not efficiently utilise glucose as a carbon source (Aharonowitz *et al.*, 1978; Garcia-Dominguez *et al.*, 1989). This is allegedly due to the lack of an effective glucose uptake system (GlcP-Glk), despite the genes *glcP* (*sclav_4529*) and *glk* (*sclav_1340*) being encoded on the genome (van Wezel *et al.*, 2005; Pérez-Redondo *et al.*, 2010). This apparent glucose auxotrophy is unique to *S. clavuligerus* among this genus.

This chapter presents results obtained from a bioinformatics approach to understanding the phylogeny of the GlcP-Glk system in Actinobacteria, making use of ActDES, which is a curated database containing 612 high quality genomes from 80 genera of Actinobacteria (Figure 3-1) (Schniete *et al.*, 2021).Moreover, a more detailed analysis of GlcP and Glk amino acid and the corresponding nucleotide (nt) sequences was carried out for a group of 14 *Streptomyces* . The genetic contexts of *glcP* and *glk* in these strains were compared. This was done to gain a deeper understanding of the level of conservation of these genome regions as adjacent genes can be functionally linked and, thus, be under a similar constraint (Pellegrini *et al.*, 1999). This is particularly interesting given that Glk is an orphan kinase and *glcP* and *glk* are found in distant genome locations in *S. coelicolor* (Angell *et al.*, 1992; van Wezel *et al.*, 2005). A correlation of low gene expression and an accumulation of synonymous and non-synonymous base changes has been established in the context of pseudogenes (Balakirev *et al.*, 2003). Therefore, to investigate the possibility of *glcP* becoming a pseudogene due to the absence of a promoter, the occurrence of base changes in *glcP* and *glk* sequences within the group of selected *Streptomyces* was examined (Bromberg *et al.*, 2008). Moreover, sequences starting 200 nt upstream from the start codon of *glcP* genes were analysed, potential promoter sequences and ribosome binding sites were predicted and compared across the strains.
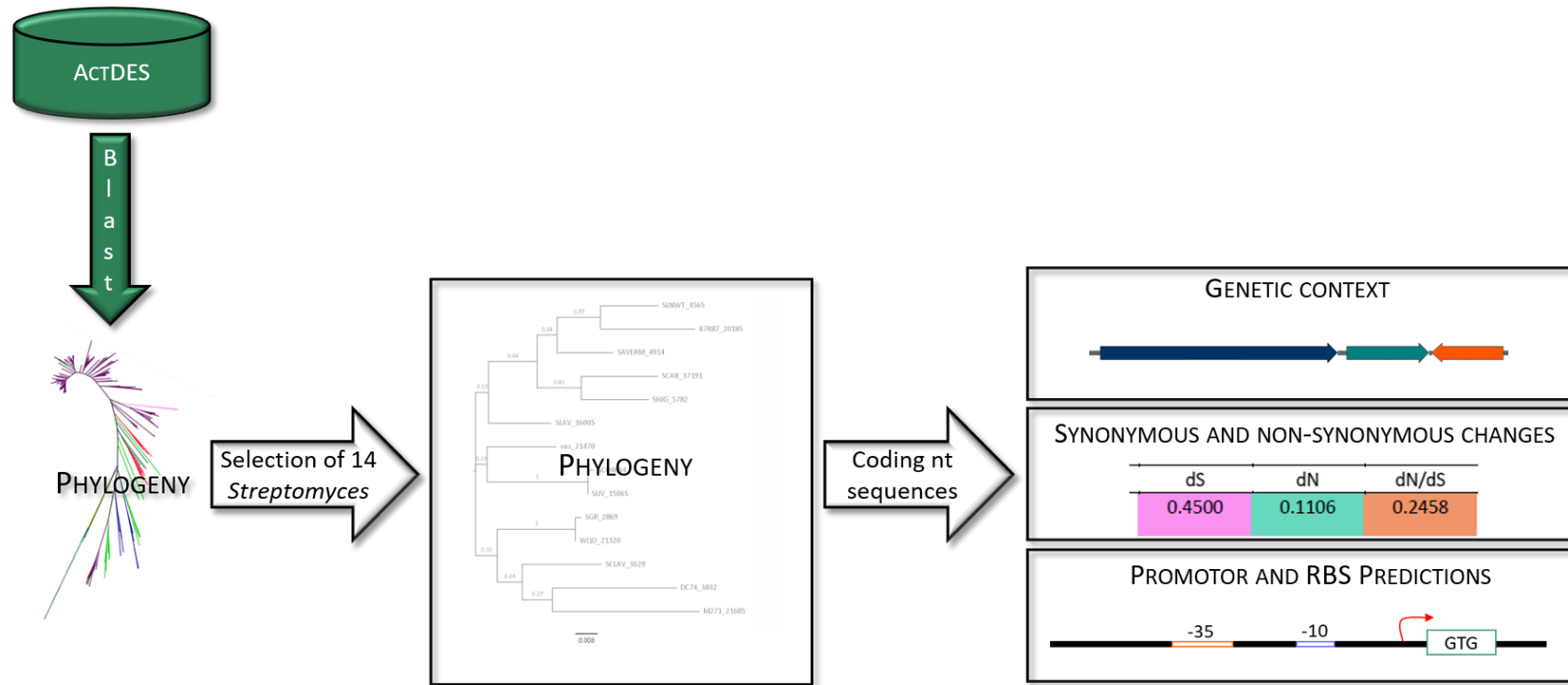
Figure 3-1: Schematic showing the workflow of the bioinformatics approach to examine the phylogeny of GlcP and Glk amongst Actinobacteria, focusing on *Streptomyces*. GlcP and Glk sequences from Actinobacteria were extracted from ActDES (Schniete *et al.*, 2021), aligned using MUSCLE (Edgar, 2004) and used to construct phylogenetic trees in MEGA X (Kumar *et al.*, 2018). GlcP and Glk sequences from a group of 14 *Streptomyces* strains were analysed phylogenetically (Clamp *et al.*, 2004; Waterhouse *et al.*, 2009). The genetic context of *glcP* and glk from these strains was compared, the occurrence of synonymous and non-synonymous changes was examined using SNAP (Bromberg *et al.*, 2008) and potential promoter sequences and ribosome binding sites (RBS) were predicted manually and using BProm (Salamov *et al.*, 2011).

66

Potential promoter and RBS sequences were initially predicted using BProm, a tool designed to predict bacterial promoters based on their sequence similarity with known *E. coli* promoters (Salamov *et al.*, 2011). As this proved unsuccessful in many cases, promoter sequence predictions were carried out manually with identified promoter sequences from *S. coelicolor* serving as templates.

The last part presents a collection of genes that were identified on the *S. clavuligerus* genome that might be involved in carbohydrate uptake systems. The genes presented were identified by adopting the *in silico* strategy developed by Bertram *et al.* that led to the identification of 53 genes involved in carbohydrate transport systems in *S. coelicolor* (Bertram *et al.*, 2004). The *in silico* search was carried out with the aim of providing native genes as targets for broadening the carbon utilisation profile of *S. clavuligerus* for rational strain improvement.

3.1 GlcP AND Glk HOMOLOGUES AMONGST ACTINOBACTERIA

GlcP and Glk amino acid sequences from ActDES were extracted using the basic local alignment search tool, blast (Altschul *et al.*, 1990; Schniete *et al.*, 2021). The initial blast search was conducted using SCO5578 (GlcP1) and SCO2126 (GlkA) as queries, which are the components of the GlcP-Glk system in *S. coelicolor*. A relaxed e-value of 1×10^{-1} yielded 3710 and 7609 hits, respectively (Table 3-1). Implementing a strict e-value-based cut-off of 1×10^{-4} only reduced the number of hits to 2284 for GlcP and 7085 for Glk, due to a lack of query sequence specificity. Therefore, GlcP and Glk signature sequences (see 3.1.1) were constructed from conserved motifs that were concatenated and submitted as queries. This reduced the number of obtained hits to 480 (GlcP) and 2055 (Glk) using the relaxed e-value, 405 (GlcP) and 781 (Glk) using the strict e-value.

3.1.1 CONSTRUCTION OF SIGNATURE SEQUENCES

GlcP and Glk signature sequences are based on amino acid motifs within MFS-family glucose permeases and ROK-family glucokinases that have been previously identified as highly conserved across species and even across kingdoms (Mueckler *et al.*, 1985; Maiden *et al.*, 1987; Kayano *et al.*, 1988; Miyazono *et al.*, 2012). The GlcP-Glk system has been studied extensively in *S. coelicolor*, which is why SCO5578 and SCO2126 were used as templates.

Table 3-1: Hits obtained from the ActDES database with different queries and e-values (Schniete *et al.*, 2021). Query sequences were the complete amino acid sequences of SCO5578 (GlcP1) and SCO2126 (GlkA) or the corresponding signature sequences. A relaxed (1×10^{-1}) or strict (1×10^{-4}) e-value was applied to the blast search. The 405 and 781 hits for GlcP and Glk (bold), respectively, that were obtained with signature sequences and the strict e-value were used to construct the phylogenetic trees.

Query	e-value	N_{Hits}
SCO5578 (GlcP1)	1×10^{-1}	3710
SCO5578 (GlcP1)	1×10^{-4}	2289
GlcP signature sequence	1×10^{-1}	480
GlcP signature sequence	**1×10^{-4}**	**405**
SCO2126 (GlkA)	1×10^{-1}	7609
SCO2126 (GlkA)	1×10^{-4}	7085
Glk signature sequence	1×10^{-1}	2055
Glk signature sequence	**1×10^{-4}**	**781**

GlcP1 is predicted to have 12 transmembrane (TM) regions and a cytoplasmic loop that connects TM regions 6 and 7 (Figure 3-2A). Four of the five motifs are in TM regions 1, 4, 7 and 10. The fifth motif is found in the extracellular region connecting TM regions 10 and 11. No streptomycete GlcP structure has been solved so far, therefore, 3D model predictions were carried out for GlcP1 (Figure 3-2B). The predicted model has a confidence score (c-score) of 0.32 (c-score ranges from -5 to 2, the higher the c-score, the more confident a model prediction) and is shown with an estimated resolution of 6.5±3.9Å (Zhang, 2008; Roy *et al.*, 2010; Yang and Zhang, 2015; Yang and Yan, 2015). The positions of the conserved motifs as part of the TM helices 1, 4, 7 and 10 are highlighted in the model. In total, five motifs were extracted from SCO5578 that were concatenated (residue range is mentioned in parentheses): (1) GFLFGYDSSVI (34-44), (2) AMWRIIGGFAIGMASVIGPA (117–136), (3) QQLVGINVVFY (287-297), (4) LIAAHVFVLFFALSWGVVVWVL (380-401), and (5) GEMFPNRIRAAAL (403-415) to build the following GlcP signature sequence: GFLFGYDSSVIAMWRIIGGFAIGMASVIGPAQQFVGINVAFYLIAAHVFVLFASWGVVVWVFGEMFPN RIRAAAL.

In a sequence alignment of GlcP amino acid sequences from different *Streptomyces*, it is obvious that, while the selected motifs are relatively conserved across species and genera, there is some variation (Figure 3-3). Given that amino acids sequences of SCO5578 and SCO7153, GlcP1 and GlcP2 from *S. coelicolor*, are identical, only SCO5578 is shown in the alignment. Sequences from *S. clavuligerus* (SCLAV), *S. avermitilis* (SAVERM), *S. griseus* (SGR), *S. lividans* (SLIV, SLI), *S. venezuelae* (SVEN), *S. scabiei* (SCAB), *S. albus* (SLNWT), *S. globisporus* (WQO), *S. hygroscopicus subsp. jinggangensis* (SHJG), *S. rapamycinicus* (M271) and *S. albulus* (DC74) are included in the alignment.

There are three different sequences for *S. lividans*, SLIV_03515 and SLIV_10580 from *S. lividans* TK24 and SLI_5859 from *S. lividans* 66. The former two sequences were extracted from the KEGG database, the latter sequence was taken from StrepDB. However, SLI_10580 is missing 31 amino acids from the N-terminus of the protein, caused by sequence information missing from the start of the corresponding gene.
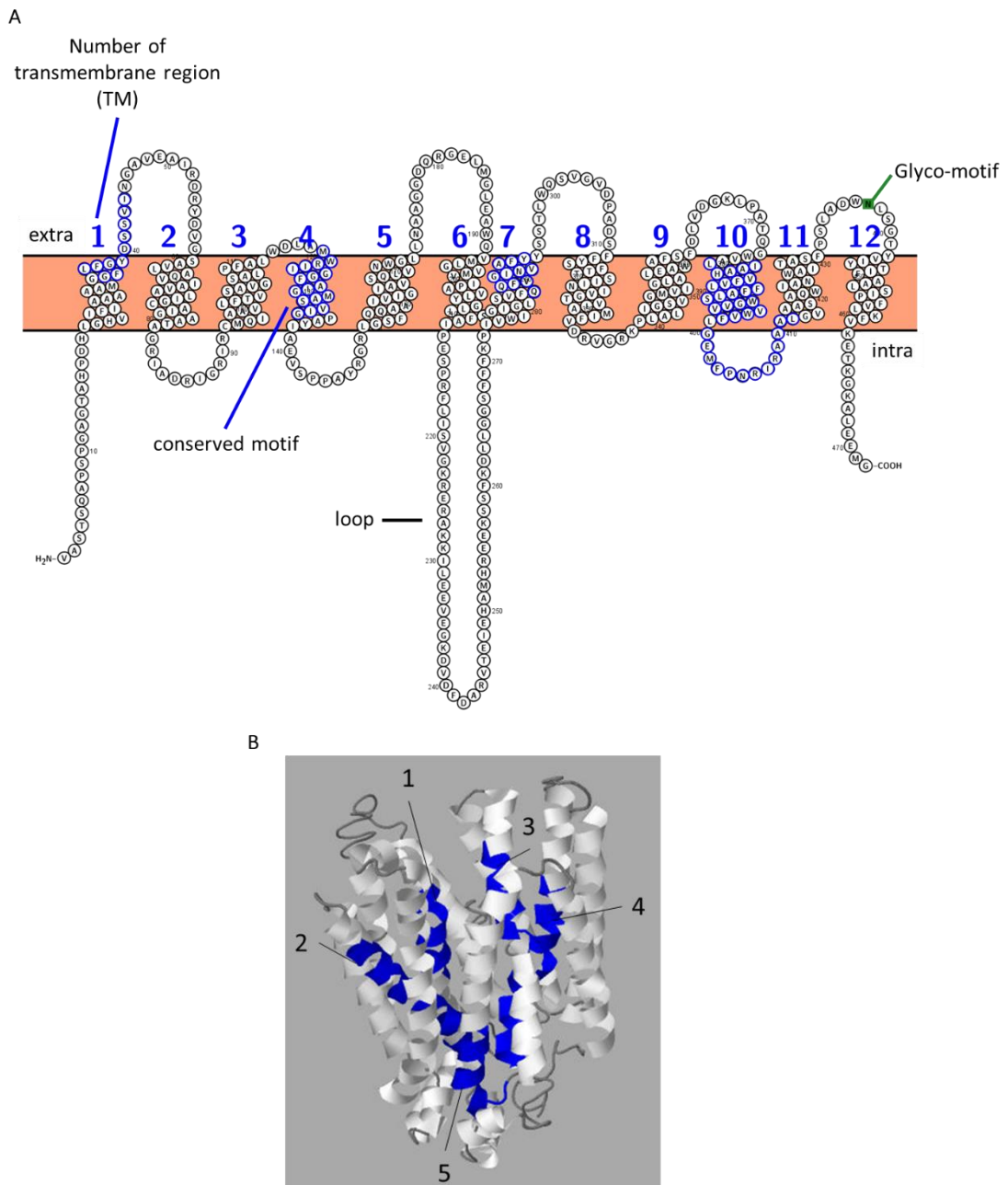
Figure 3-2: Secondary structure of SCO5578 and predicted 3D model. (A) Secondary and (B) tertiary structure predictions were carried out using PROTTER and I-TASSER, respectively, showing the positions of the conserved residues (blue) used in the concatenated GlcP signature sequence (Zhang, 2008; Omasits *et al.*, 2013). Conserved residues are located within TM regions 1, 4, 7 and 10. The 3D model was predicted with a c-sore of 0.32 and is shown with an estimated resolution of 6.5±3.9Å.
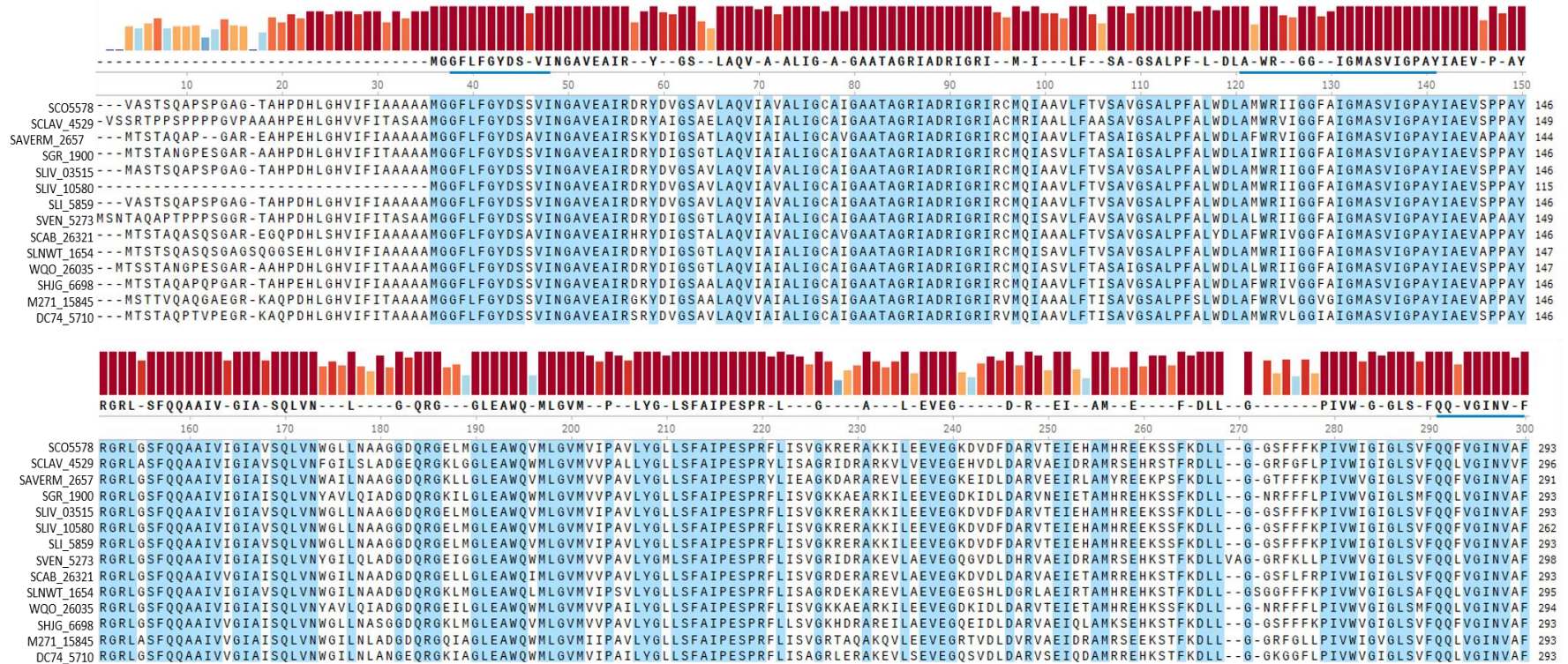
Figure 3-3: GlcP amino acid sequence alignment from 12 different *Streptomyces* strains. Sequences were taken from *S. coelicolor* (SCO5578), *S. clavuligerus* (SCLAV_4529), *S. avermitilis* (SAVERM_2657), *S. griseus* (SGR_1900), *S. lividans* TK24 (SLIV_03515, SLIV_10580) and *S. lividans* 66 (SLI_5859), *S. venezuelae* (SVEN_5273), *S. scabiei* (SCAB_26321), *S. albus* (SLNWT_1654), *S. globisporus* (WQO_25035), *S. hygroscopicus subsp. jinggangensis* (SHJG_6698), *S. rapamycinicus* (M271_15845), *S. albulus* (DC74_5710). Residues identical in 95% or more of the sequences are shown in blue. The coloured bars at the top indicate the level of conservation of a residue (the redder the bar, the more conserved the residue), the consensus sequence is shown below. Residues are underlined if they are part of a conserved motif used to construct the GlcP signature sequence.

71
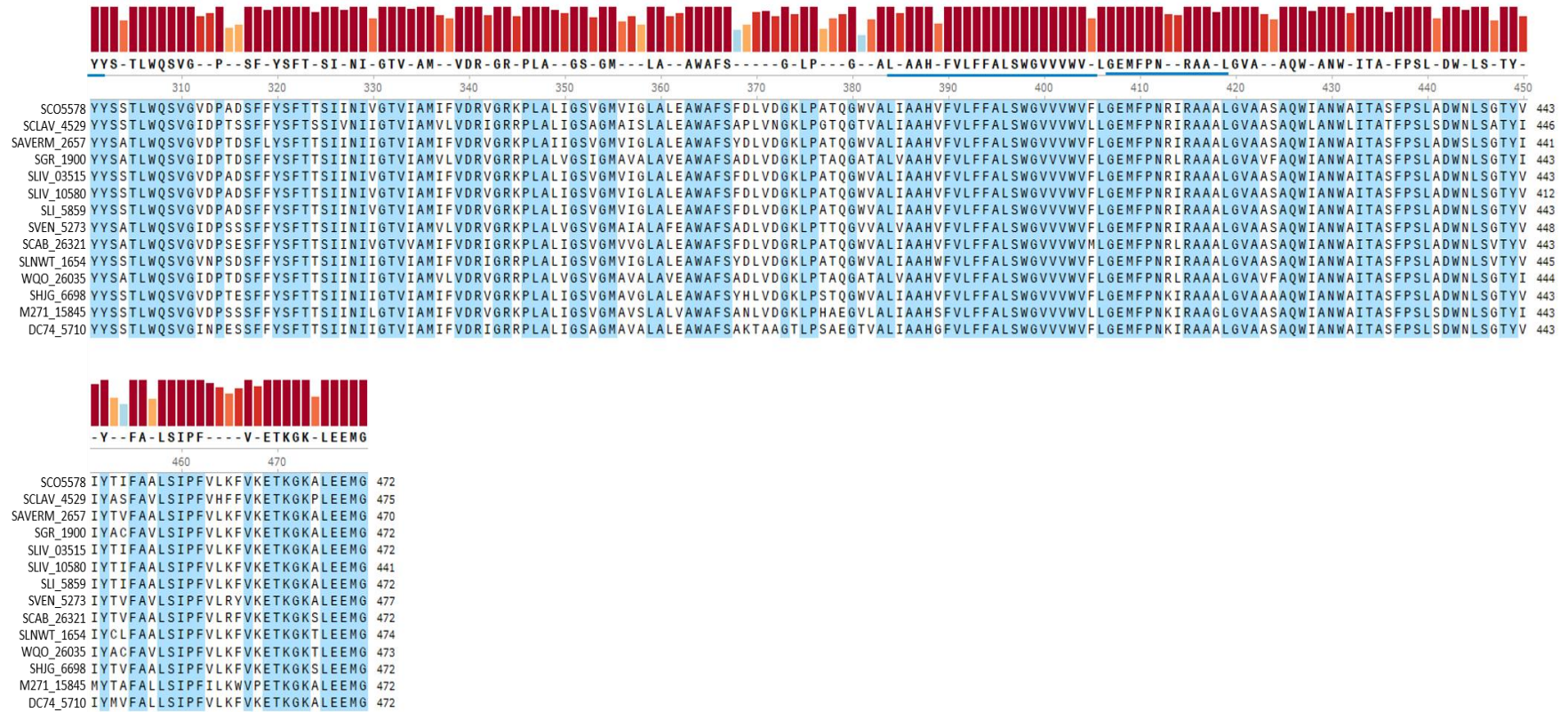
Figure 3-3 (continued): GlcP sequence alignment continued. Residues present in 95% of the sequences are highlighted in blue and residues contained in the signature sequence are underlined blue in the consensus sequence as before.

The *sliv_10580* gene sequence was, therefore, extended to include 103 bases upstream from the proposed start for later analyses and is, henceforth, referred to as *sliv_10580* ext (protein: SLIV_10580 ext). *S. lividans* 66 is synonymous with *S. lividans* 1326, the latter being the strain represented in the ActDES database (Schniete *et al.*, 2021). The *S. lividans* 66 genome has a single *glcP* gene, *sli_5859*, according to the StrepDB database. *S. lividans* is similar to *S. coelicolor* phylogenetically. It might, therefore, be expected for *S. lividans* to also possess two copies of the *glcP* gene. As this appears to be the case for the TK24 strain, a blast search was conducted and a second *glcP* gene was identified on the *S. lividans* 66 genome (*sli_7370*). However, sequence information is missing, and the truncated gene consists of 870 nt instead of 1418 nt, which is the length of both *S. coelicolor glcP* genes. Sequences taken from *S. lividans 66* were, therefore, excluded from further analyses.

3.1.1.2   Glk Signature Sequence

Unlike GlcP1, GlkA is not a transmembrane protein. It is composed of helices and sheets (Figure 3-4) (Zhang, 2008; Roy *et al.*, 2010; Yang and Zhang, 2015; Yang and Yan, *et al.*, 2015). In the presence of glucose, Glk is predicted to form a stable homotetramer (Imriskova *et al.*, 2005). Only two motifs were used to construct the Glk signature sequence (residue range is mentioned in parentheses): (1) LGTGLGGGIIIGNKLRRGHFGVAAEFGH (130-157) and (2) CGCGSQGCWEQYAS (167-180). The resulting sequence is LGTGLGGGIIIGNKLRRGHFGVAA**EFGH**CGCGSQGCW**E**QYAS, purple and red residues are the ExGH and Zn-binding regions (CXCGXXGCXE) characteristic for members of the ROK-family. Glucose-binding sites within the signature sequence are highlighted in bold. The signature sequence is in the middle of the amino acid sequence (centre of the 3D model). The 3D model was predicted with a c-score of 1.74, which is higher than the c-score for the GlcP1 model (0.32). The alignment of Glk sequences from the same *Streptomyces* strains shows a higher number of motifs found in 95% of sequences compared to the number of residues that reach this threshold in the GlcP alignment (Figure 3-5).
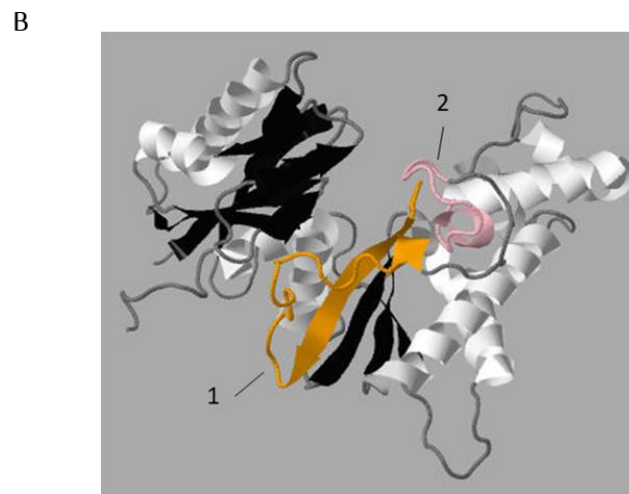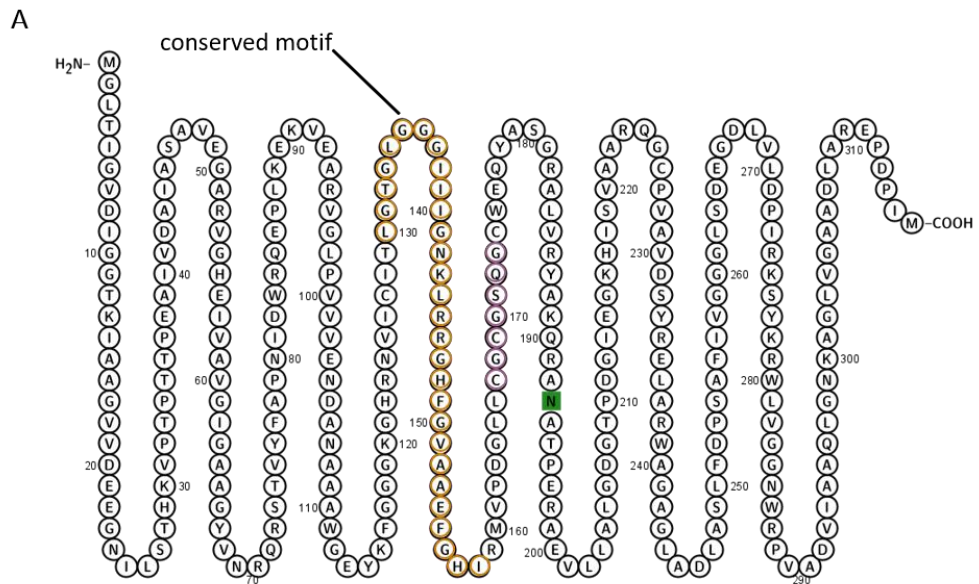
Figure 3-4: Secondary structure of SCO2126 and predicted 3D model. The models were constructed with (A) PROTTER and (B) I-TASSER showing the positions of the two adjacent conserved motifs that were used to construct the Glk signature sequence (Zhang, 2008; Omasits *et al.*, 2013). Conserved residues (orange in A, yellow and pink in B) contain the characteristic ExGH motif and Zn-binging region (CXCGXXGCXE). The 3D model was predicted with a c-score of 1.74 and is shown with an estimated resolution of 2.8±2.1Å.
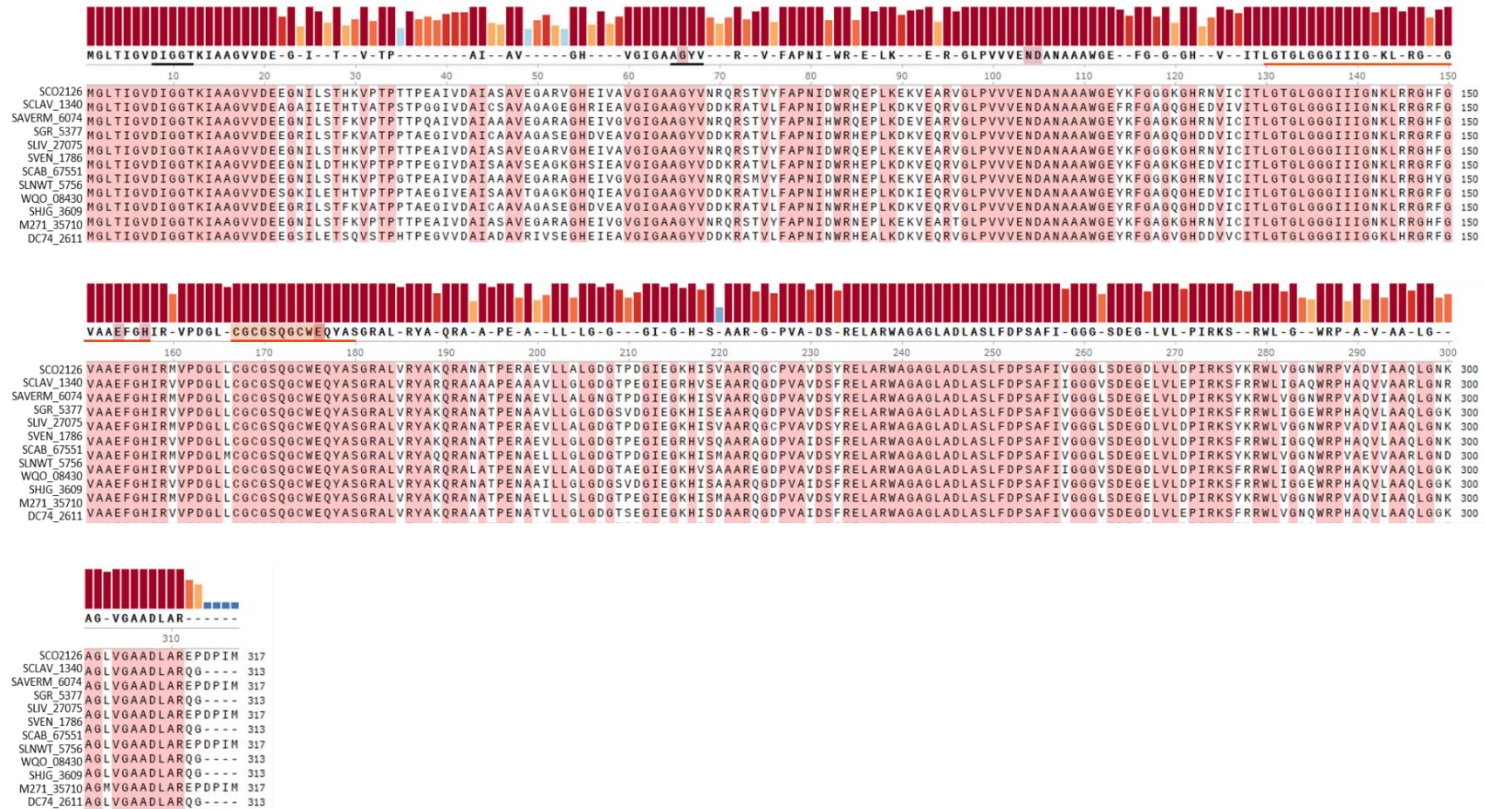
Figure 3-5: Sequence alignment of Glk sequences from *Streptomyces* strains. Sequences were taken from *S. coelicolor* (SCO2126), *S. clavuligerus* (SCLAV_1340), *S. avermitilis* (SAVERM_6074), *S. griseus* (SGR_5377), *S. lividans* TK24 (SLIV_27075) *S. venezuelae* (SVEN_1786), *S. scabiei* (SCAB_67551), *S. albus* (SLNWT_5756), *S. globisporus* (WQO_08430), *S. hygroscopicus subsp. jinggangensis* (SHJG_3609), *S. rapamycinicus* (M271_35710), *S. albulus* (DC74_2611). The red highlighted residues are shared by at least 95% among the sequences. Conserved motif residues are underlined orange, the ATP-binding site and catalytic site are underlined black. Glucose-binding sites and the Zn-binding region are highlighted in red and orange in the consensus sequence, respectively.

3.1.2 ActDES RpoB PHYLOGENETIC TREE

Hits obtained using the GlcP and Glk signature sequences as queries were aligned and phylogenetic trees were constructed. The topology of these trees was compared to a previously constructed RpoB tree (Figure 3-6) (Schniete *et al.*, 2018). RpoB is the RNA polymerase β-subunit and using this sequence for phylogenetic analyses delivers a higher phylogenetic resolution than 16S rDNA especially when differentiating between closely related strains (Case *et al.*, 2007; Vos *et al.*, 2012). The use of *rpoB*-based phylogeny for *Streptomyces* is a suitable method for distinguishing phylogenetically highly similar strains. Furthermore, and in contrast to 16S rDNA, the single-copy *rpoB* gene is protein-encoding, meaning phylogenetic analyses can also the carried out using the respective amino acid sequences (RpoB) (Kim *et al.*, 2004; Case *et al.*, 2007). The branches in all phylogenetic trees are coloured according to the order the species belongs to, from which the protein sequence was extracted.

The RpoB tree is comprised of three main branches, one of which is almost entirely made up of sequences taken from representatives of the order Streptomycetales (purple). It also contains a much smaller clade representing Catelunisporales (light pink). The branch to the right is predominantly made up of Micrococcales (navy blue), Propionibacteriales (yellow) and Frankiales (turquoise) RpoB. Sequences from species found in the orders of Pseudonocardiales (forest green), Corynebacteriales (green), Micromonospororales (pink), Glycomycetales (black), Actinomycetales (light blue) and Streptosporangiales (red) constitute the branch on the left. Five sequences were allocated positions within the phylogenetic tree that were not in accordance with the general position of the corresponding order (Hiltner, 2015). These were RpoB sequences from *Streptomyces carneus* (ID: 103200_819), *Streptomyces ssp.* (NRRL F3213, ID: 104209_3190), *Rhodococcus rhodni* (ID:104518_2496), *Actinospica acidiphila* (ID: 104678_6451) and *Microtetraspora glauca* (ID:104669_4572). Yet, the number of misplaced sequences is low and the RpoB tree offers a robust phylogenetic basis for assuming the phylogenetic distance of species.
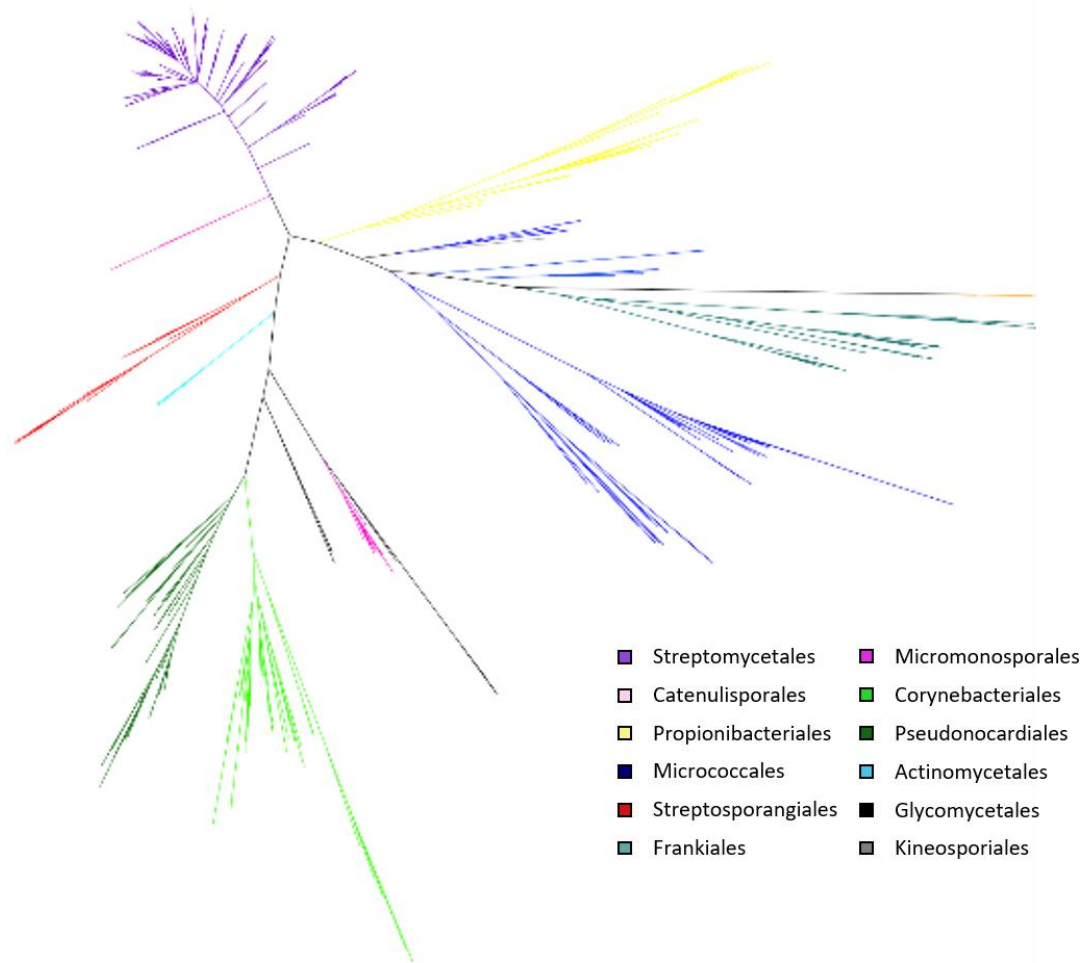
Figure 3-6: Evolutionary analysis of Actinobacterial RpoB sequences (Schniete *et al*., 2018). Orders represented in the tree are shown in the legend. Sequences were extracted from the ActDes database, aligned using MUSCLE prior to constructing this phylogenetic tree using the Maximum likelihood algorithm integrated into MEGA X (Edgar, 2004; Kumar *et al.*, 2018; Schniete *et al.*, 2021). The tree is comprised of three distinct branches, and sequences stemming from species of the same order were generally placed in the same clade. Exceptions to this were *Streptomyces carneus* (ID: 103200_819), *Streptomyces ssp.* (NRRL F3213, ID: 104209_3190), *Rhodococcus rhodni* (ID:104518_2496), *Actinospica acidiphila* (ID: 104678_6451) and *Microtetraspora glauca* (ID:104669_4572), which were misplaced on the tree.

3.1.3 Phylogenetic Diversity of GlcP among Actinobacteria

Evolutionary analysis of 405 Actinobacterial GlcP sequences (Figure 3-7) produced a phylogenetic tree with a topology that is dissimilar to the RpoB tree shown previously. Unlike the RpoB tree, the GlcP tree is not comprised of three distinct branches. The radial tree is composed of a single branch that is curved where the majority of Streptomycetales (purple) sequences are present (group I). The curvature infers greater evolutionary change for a sub-set of Streptomycetales, the majority of which are members of the *Streptomyces* genus, as the tip of the branch diverges from the base of the tree. Other Streptomycetales genera represented in group I, *Streptacidiphilus* and *Kitasatospora* are more basal with the tree. Sequences that were misplaced among the Streptomycetales clade are *Actinospica acidiphila* JNYX (ID:104678_2047, Catenulisporales), *Nocardia convaca* NBRC 100430 (ID: 102745_6156, Corynebacteriales) and *Nocardia niigatensis* NBRC 100313 (ID:102806_6534, Corynebacteriales).

The tip of the curved end of the tree is composed of two distinct clades of Streptomycetales, one of which contains the *S. clavuligerus* GlcP sequence (SCLAV_4529, white arrow), the other contains GlcP1 from *S. coelicolor* (SCO5578, black arrow). The former forms a sub-clade with GlcP from *S. tsukubaensis* NRRL 18488, the sequences are 87.4% identical. Interestingly, this species has been reported to utilise glucose, as well as glycerol as carbon sources and both sugars trigger a CCR response of antibiotic biosynthesis (Kim *et al.*, 2007; Martínez-Castro *et al.*, 2013; Ordóñez-Robles *et al.*, 2017). Both SCO5578 and the duplicated transporter, SCO7153 (GlcP2), form a stable clade with homologues from *S. lividans* 1326 and *S. coelicoflavus* ZG0656, with which they share 100.0% and 97.8% of their amino acid sequence, respectively.

In the phylogenetic tree, there is a second, much smaller and phylogenetically distant Streptomycetales clade (group II) located at the opposite end of the tree to group I. While sequences found in group II only share approximately 50% of their amino acid sequence with SCO5578, group I sequences are 65% - 100% identical to SCO5578. Within the group I clade, there is a branch of Cetanulisporales (light pink) that is comprised of two species, one of which is another *Actinospica* (*Actinospica acidiphila* was mentioned earlier because it was misplaced among the Streptomycetales clade).
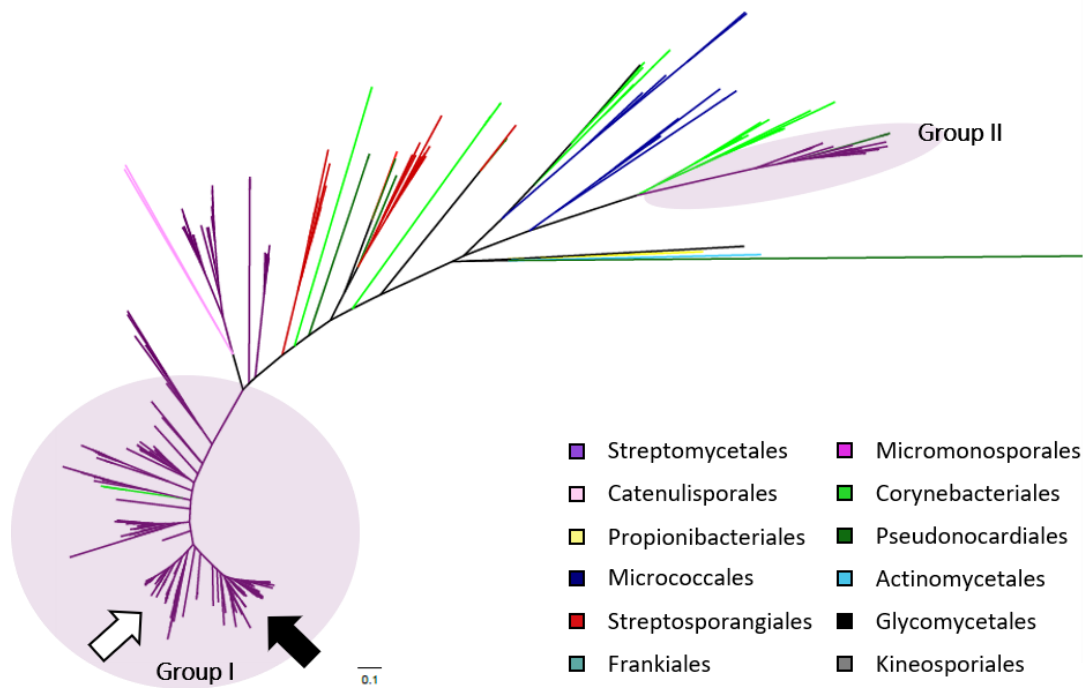
Figure 3-7: Evolutionary analysis of 405 GlcP sequences obtained from ActDES (Schniete *et al.*, 2021). Branches are coloured according to their order and the orders represented in the tree are shown in the legend. Extracted sequences were aligned using MUSCLE (Edgar, 2004). The phylogenetic tree was constructed with a bootstrap of 1,000 using the Maximum likelihood method in MEGA X (Kumar *et al.*, 2018). The tree is drawn to scale, branch lengths are measured in the number of substitutions per site as indicated by the scale bar. The positions of *S. clavuligerus* (white arrow) and *S. coelicolor* (black arrow) GlcP sequences are indicated. There are two phylogenetically distinct Streptomycetales clades, group I and group II, that contain sequences with more than 65% and approximately 50% amino acid sequence identity with SCO5578, respectively.

The branch from which the Catenulisporales clade emerges also holds a Streptomycetales clade largely comprised of *Streptacidiphilus* and *Kitasatorspora*. The Catenulisporales clade, therefore, almost serves as a phylogenetic marker separating the *Streptomyces* and non-*Streptomyces* Streptomycetales.

The size and number of Streptomycetales clades in this tree suggest a level of diversity of GlcP proteins across Actinobacteria. An example that highlights the diversity of GlcP within group I is *Kitasatospora aureofaciens* (former *Streptomyces aureofaciens*, Labeda *et al.*, 2017). In total, there are three different strains of *K. aureofaciens* represented in the tree: JNWR01, JOER01 and JODU01. The former two strains, JNWR1 and JOER01, each have two GlcP proteins (ID: 103181_2156, 104282_6108, 103197_3030 and 103197_6598) that share 68.1% and 67.5% of their amino acid sequence with each other, respectively. These are only around 65.0% identical to SCO5578 and are placed in the Streptomycetales clade on the far side of the Catenulisporales branch found in group I. In contrast, GlcP from the third strain, JODU01 (ID: 102744_5673), is placed near SCO5578 on the tree and has an almost identical amino acid sequence (97.5%).

Group II exclusively contains *Streptomyces* sequences and two *Sciscionella* (Pseudonocardiales) sequences, GlcP from *Sciscionella marina* DSM 45152 (ID:104667_6476) and GlcP from *Sciscionella* SE31 JALM01 (ID:104717_4741), highlighting that the phenomenon of having two phylogenetically distinct GlcP proteins might be unique characteristic of *Streptomyces*. In group II, the *Sciscionella* sequences separate the unique streptomycete GlcP sequences from (strains not represented in group I) and additional streptomycete GlcP sequences (strains represented in group I). For instance, the GlcP sequence from *Streptomyces ssp*. NRRL_F-5126 (ID: 104216_4897) located adjacent to the *Sciscionella marina* sequence in group II is the only GlcP-like sequence from this strain and it only shares 49.8% of its amino acid sequence with SCO5578. The GlcP sequence on the other side of the two *Sciscionella* sequences is from *Streptomyces* AcH_505 (ID: 103247_7115, 48.4% amino acid sequence identity with SCO5578). This strain has a GlcP1 homologue (ID: 103247_5245) located in group I that is 80.4% identical to SCO5578. The presence of group II and the fact that it contains unique GlcP sequences, as well as GlcP from species that are also represented in group II strongly suggests that group II sequences might have been

acquired horizontally. Unlike the clades in the RpoB tree, no neat division between the clades is visible for Streptosporaniales (red), Corynebacteriales (green), Pseudonocardiales (forest green) and Propionibacteriales (yellow) as representative sequences are scattered across clades. For instance, *Rhodococcus* sequences are found in every Corynebacteriales clade on the tree. This might be due to horizontal gene transfer (HGT), which may have occurred more often in *Rhodococci*. Among one of the Corynebacteriales clades, *Pilimelia anulata* JOFP01 (ID:105533_3553, Micromonosporales) forms a clade with *Nocardia asteroides* NBRC 15531 (ID:102748_1874). As the sequences are 100.0% identical in their amino acid sequence, it is possible the *Pilimelia* strain acquired its GlcP through horizontal gene transfer (HGT).

HGT and gene family expansion, which is another mechanism by which organisms can increase their biochemical diversity, can increase metabolic robustness and plasticity of the organism in a dynamic environment (Wagner, 2008; Hiltner *et al.*, 2015; Schniete *et al.*, 2018). The basis of gene family expansion is "genetic redundancy". Yet, many genes labelled as "redundant" encode proteins that have diverged functionally to yield greater biochemical diversity (Bentley *et al.*, 2002; Schniete *et al.*, 2018). The GlcP phylogenetic tree displays the diversity of glucose permeases among Actinobacteria found in the ActDES database. The two distinct branches of Streptomycetales in both trees suggests the presence of phylogenetically distant GlcP and Glk proteins among this order, which may have arisen through gene family expansion and/or HGT.

3.1.4   GENE FAMILY EXPANSION OF ROK-FAMILY GLUCOKINASES IN STREPTOMYCETALES

The streptomycete glucose uptake system is composed of GlcP and Glk. The proteins are, thus, functionally linked (van Wezel *et al.*, 2005). Given that functionally linked proteins can be exposed to the same functional constraints, gene family expansion might also have occurred for Actinobacterial Glk proteins. Therefore, an evolutionary analysis of 781 Glk sequences was carried out. The topology of the Glk phylogenetic tree (Figure 3-8) is comprised of four distinct branches that emerge from the basal branch and, thus, resembles the RpoB tree more than the GlcP tree. The top and bottom clades are comprised of Streptomycetales (purple) sequences, labelled group I and group II, and, unlike the GlcP tree, they are similar in size.

81
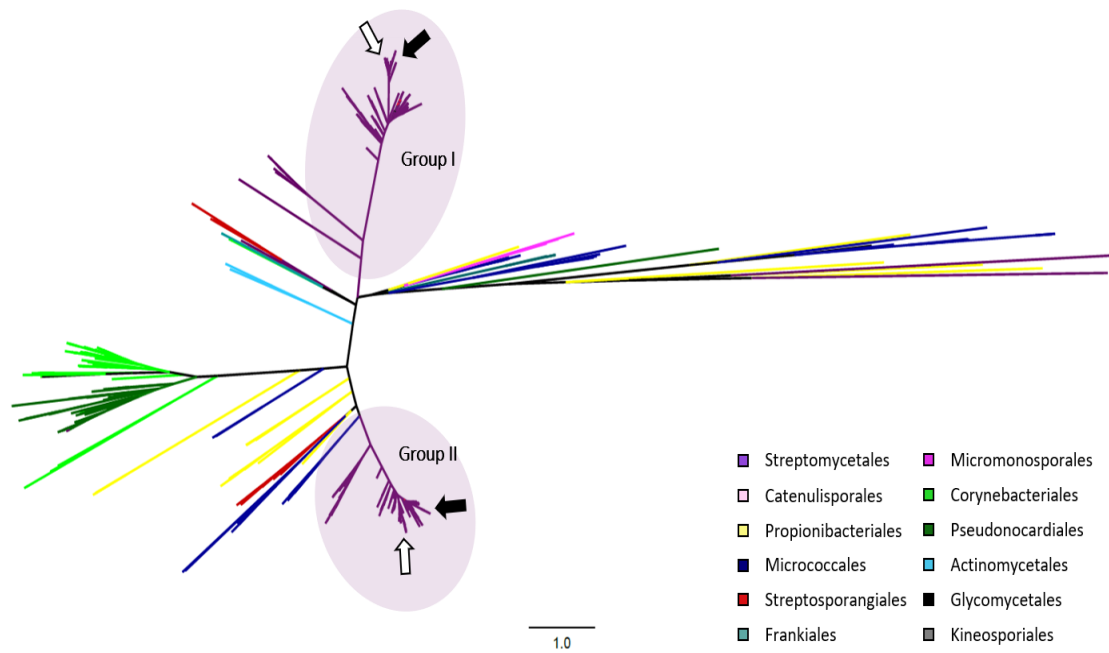
Figure 3-8: Evolutionary analysis of 781 Glk sequences obtained from ActDES (Schniete *et al*., 2021). The sequences were aligned with MUSCLE (Edgar, 2004). The Maximum likelihood method integrated into MEGA X was used to construct the phylogenetic tree with a bootstrap of 500 (Kumar *et al.*, 2018). The tree is drawn to scale with the scale bar showing the branch length equivalent to one substitution per site. The clades are coloured by order (see legend). Two phylogenetic Streptomycetales clades are present in the tree, group I and group II, which introduce a level of symmetry to the topology of the tree. Group I sequences share at least 80.0% of their amino acid sequence with SCO2126, group II sequences share the same level of sequence similarity with SCO6260, which is not associated with GlcP. The positions of *S. clavuligerus* (white arrow) and *S. coelicolor* (black arrow) Glk sequences in groups I and II are indicated.

Despite being phylogenetically distant, the sequences found in group I share at least 80.0% of their amino acids sequence with the GlcP-associated glucokinase, GlkA (SCO2126), from *S. coelicolor*. As before, *S. coelicolor* (SCO2126, black arrow) and *S. clavuligerus* (SCLAV_1340, white arrow) sequences are found in distinct subclades of group I. The former forms a clade with *S. lividans* 1326 (ID: 103283_5482) and *S. violaceoruber* (ID: 104361_2661), both of which are 100.0% identical to SCO2126. The *S. clavuligerus* GlkA-homologue forms a clade with *S. tsukubaensis* (ID: 104320_1425).

Group II encompasses sequences that are similar to SCO6260, a non GlcP-associated ROK-family glucokinase. Generally, if a strain is represented in group I and group II in the Glk tree, the respective sequences share approximately 50% of their amino acid sequence. For instance, SCO2126 and SCO6260 from *S. coelicolor* are 45.7% identical, SCLAV_1340 and the SCO6260-homologue, SCLAV_5022, are 51.1% identical. SCO6260 and SCLAV_5022 in group II are also located in phylogenetically distinct sub-clades.

Interestingly, *Microtetrara glauca* JOFO01 (IDs: 104669_1243 and 104669_2394), a Streptosporangiales species, is represented in both Streptomycetales groups, enhancing the symmetry observed in this tree. Simultaneously, given that non-Streptomycetales are largely excluded from groups I and II, this suggests a greater phylogenetic diversity of ROK-family glucokinases within the Streptomycetales than orders of Actinobacteria. An example that illustrates the variety of Glk sequences amongst Streptomycetales specifically are Glk sequences from *Streptomyces niveus* NCIMB_11891. The species is not only represented in group I (ID: 103303_1668) and group II (ID: 103303_6408), but also possesses a third ROK-family Glk protein distant from both Streptomycetales groups, which is placed in the distal branch that extends to the right (ID: 103303_7638). This distal branch contains sequences from Propionibacteriales (yellow), Actinomycetales (pink), Micrococcales (navy blue), Pseudonocariales (forest green) and members of the *Streptacidiphilus* genus. The clade also contains a *Modestobacter* KNN45-2b sequence (ID: 104679_2235). The genus *Modestobacter* was previously classified as Frankiales but has been reassigned to the order of Geodermatophilales (Sen *et al.*, 2014). This indicates that horizontal acquisition in addition to duplication might be occurring within the ROK-family of glucokinases of *Streptomyces niveus*.

Strains whose Glk are in the Corynebcateriales/Pseudonocardiales clade are represented once in the tree, suggesting that no gene expansion has occurred in members of these orders. Alternatively, the use of the signature query sequence or the nature of the ActDES database do not allow for other Glk orthologues to be picked up in these orders.

Flanking the Corynebcateriales/Pseudonocardiales clade are two smaller Streptosporangiales clades, one nearer to group I that shares a branch with *Actinomycetales* (turquoise) and *Corynebacteriales* and one closer to group II, placed in between Propionibacteriales (yellow) and Micrococcales (navy blue). The former Streptosporangiales clade consists of sequences from five different *Nocardiopsis* species, as well as one from *Streptomonospora alba*. The *Nocardiopsis* species are also represented in the latter Streptosporangiales clade (nearer to group II), suggesting that gene expansion may have also occurred amongst this genus.

This phylogenetic analysis suggests gene expansion of ROK-family glucokinases may have occurred in Streptomycetales and the *Nocardiopsis* genus by gene duplication and/or HGT. The increased frequency with which this appears to have occurred for Glk proteins compared to GlcP may be linked to the multi-functionality of Glk in Streptomycetales as sugar kinase and mediator of CCR (Angell *et al.*, 1992, 1994; van Wezel *et al.*, 2007).

3.2   GLCP AND GLK AMONG *STREPTOMYCES*

The Actinobacteria-wide phylogenetic analysis of GlcP and Glk has highlighted that gene family expansion might have occurred particularly among Streptomycetales. To gain a more in-depth understanding of the level of conservation of SCO5578 (GlcP1) and SCO2126 (GlkA) orthologues and their coding genes among the *Streptomyces* genus, evolutionary analyses of GlcP and Glk amino acid sequences from 14 *Streptomyces* species were carried out and the genetic context of the corresponding *glcP* and *glk* genes were compared. Sequences from the following strains were selected: *S. albulus* NK660 (DC74), *S. albus* DSM 41398 (SLNWT), *S. avermitilis* MA-4680 (SAVERM), *S. globisporus* C-1027 (WQO), *S. griseus subsp. griseus* NBRC 13350 (SGR), *S. hygroscopicus subsp. jinggangensis* 5008 (SHJG), *S. lavendulae subsp. lavendulae* CCM 3239 (SLAV), *S. lividans* TK24 (SLIV), *S. rapamycinicus* NRRL 5491 (M271), *S. scabiei* 87.22 (SCAB), *S. venezuelae* NRRL B-65442 (SVEN/vnz) and *S. tsukubensis* NRRL 18488

(B7R87). These strains were selected based on the availability of whole genome sequencing data on StrepDB and KEGG. *S. tsukubaensis* was included due to the close phylogenetic relationship with *S. clavuligerus* as shown in the RpoB tree (Figure 3-6).

3.2.1 MULTI-LOCUS SPECIES TREE OF *STREPTOMYCES* STRAINS

Analogously to the Actinobacteria-wide evolutionary analyses, RpoB protein sequences were initially used to establish phylogenetic relationships between the strains. However, the resulting tree had low bootstrap 1000 values, indicating instability of resulting clades. Therefore, a multi-locus approach was taken using the autoMLST web server (Figure 3-9) (Alanjary *et al.*, 2019).). Bootstrap values calculated for these sequences, which include user-submitted sequences of 14 *Streptomyces* strains and closely related strains chosen by autoMLST, are high pointing at the high confidence of these clades.

Among the *Streptomyces* sequences, the following species form stable clades: *S. globiosporus* (WQO) and *S. griseus* (SGR), *S. venezuelae* (SVEN) and *S. lavendulae subsp. lavendulae* (SLAV), *S. clavuligerus* (SCLAV) and *S. tsukubaensis* (B7R87), *S. albulus* (DC74) and *S. rapamycinicus* (M271), *S. avermitilis* (SAVERM) and *S. scabiei* (SCAB), as well as *S. coelicolor* (SCO) and *S. lividans* (SLIV). *S. albus* (SLNWT) and *S. hygroscopicus subsp. jinggangensis* (SHJG) have been placed on their own, so without another user-submitted sequence but with closely related strain.

3.2.2 GlcP AND Glk PHYLOGENY AMONG THE SELECTION OF *STREPTOMYCES*

Phylogenetic trees with GlcP (GlcP1 orthologues) and Glk (GlkA orthologues) sequences from the same *Streptomyces* were constructed and compared to the previous tree (Figure 3-10). Stable clades present in the GlcP and Glk trees that are also found in the autoMLST tree were the following: *S. globiosporus* (WQO) and *S. griseus* (SGR), *S. clavuligerus* (SCLAV) and *S. tsukubaensis* (B7R87), *S. albulus* (DC74) and *S. rapamycinicus* (M271), and *S. coelicolor* (SCO) and *S. lividans* (SLIV). Support values for these pairings varied between the GlcP and Glk tree, with bootstrap values in the latter tree tending to be lower compared to the former.
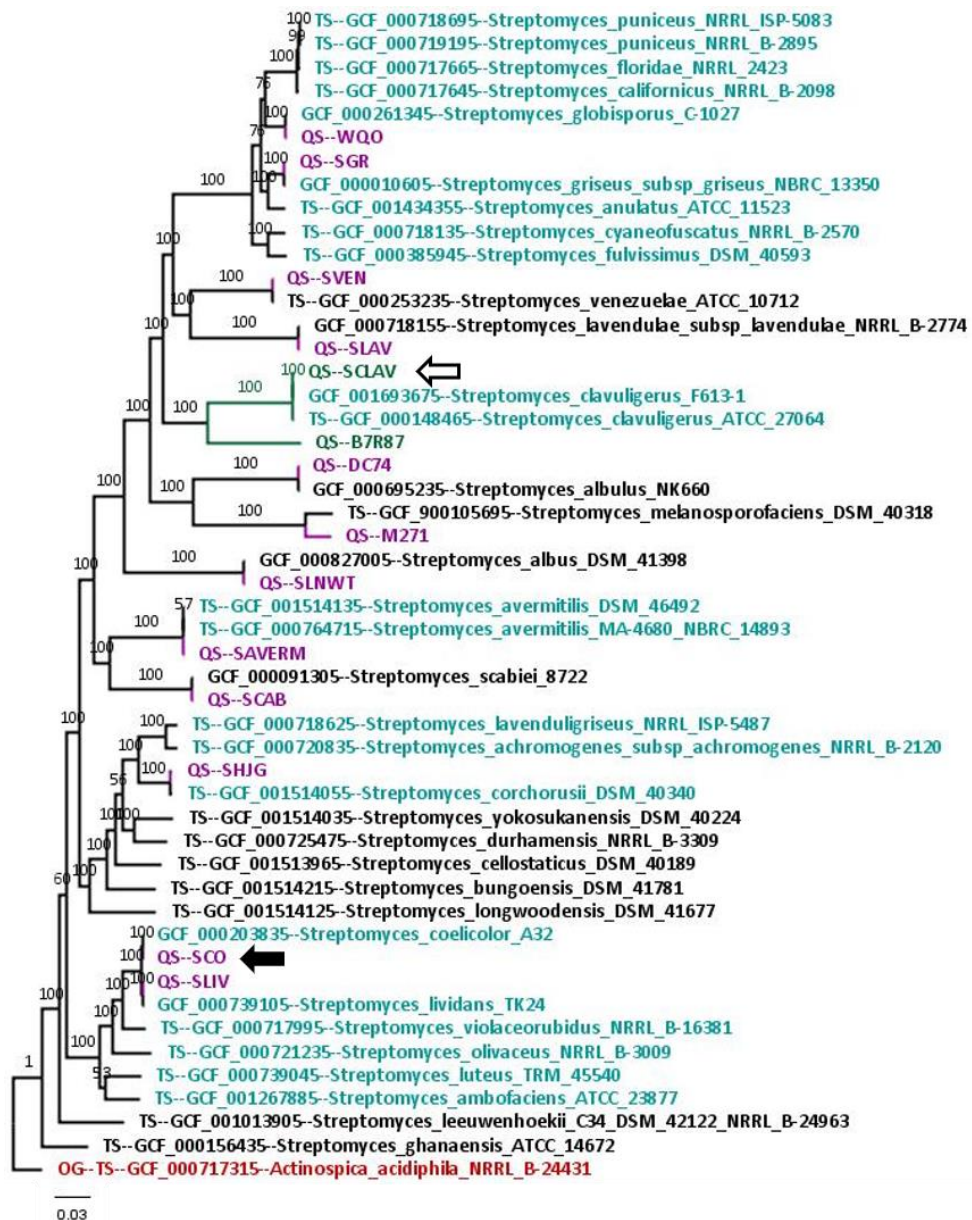
Figure 3-9: Multi-locus species tree of 14 Streptomyces species. The tree was calculated using the autoMLST web server interface (default setting, concatenated tree, bootstrap 1,000)(Alanjary *et al.*, 2019). The tree is drawn to scale and branch lengths are measured by number of substitutions per site as indicated by the scale bar. The branch labels show bootstrap values, with 100 indicating the highest stability of a clade. The tree contains sequences taken from type strains (TS) and the following user-submitted sequences (QS, shown in purple/green): *S. albulus* (DC74), *S. albus* (SLNWT), *S. avermitilis* (SAVERM), *S. clavuligerus* (SCLAV), *S. coelicolor* (SCO), *S. globisporus* (WQO), *S. griseus* (SGR), *S. hygroscopicus* (SHJG), *S. lavendulae* (SLAV), *S. lividans* (SLIV), *S. rapamycinicus* (M271), *S. scabiei* (SCAB), *S. tsukubaensis* (B7R87), *S. venezuelae* (SVEN). Gene family clusters (GCF) are indicated. Average nucleotide identities (ANI) of min. 90% (blue). The outgroup (OG) is shown in red.
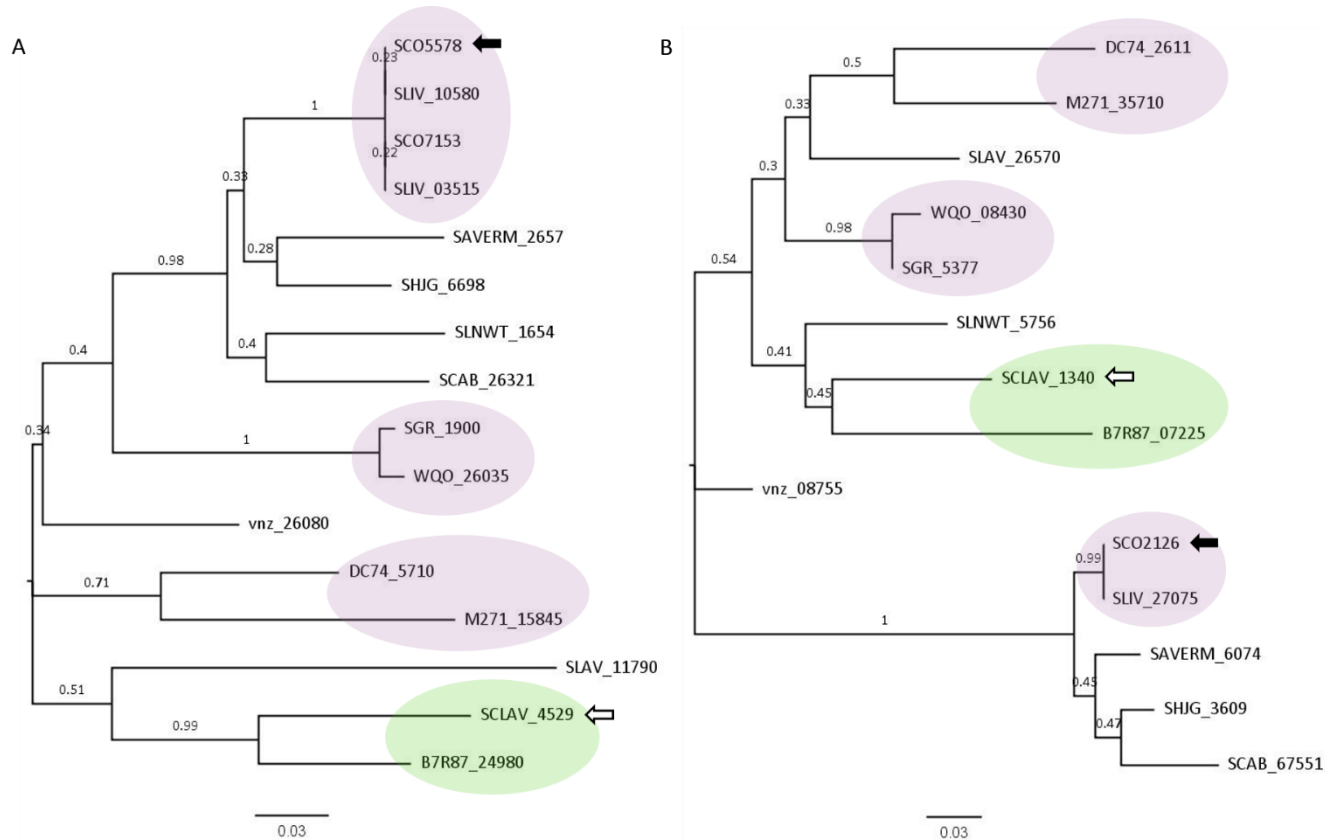
Figure 3-10: Evolutionary analysis of (A) GlcP1 and (B) GlkA orthologues from 14 different *Streptomyces* (abbreviations as before). Sequences were aligned using MUSCLE, phylogenetic trees were built using the Maximum likelihood method and a bootstrap of 1,000, which are displayed as decimal branch labels (Edgar, 2004; Kumar *et al.*, 2018). The trees are drawn to scale, and branch lengths are measured in number of substitutions per site as indicated by the scale bars. Clades conserved from the autoMLST tree are circled in purple, a green circle is drawn around the *S. clavuligerus/S. tsukubaensis* clade. The positions of *S. clavuligerus* (white arrow) and *S. coelicolor* (black arrow) sequences are indicated.

For instance, the DC74/M271 clades have bootstrap values of 0.7 and 0.5 in the GlcP and Glk tree, respectively. Similarly, the SCLAV/B7R87 are have bootstrap values of approximately 1.0 (GlcP) and 0.5 (Glk). *S. albus* (SLNWT) is loosely grouped together with *S. scabiei* in the GlcP tree (bootstrap=0.4) and the *S. clavuligerus/S. tsukubaensis* clade in the Glk tree (bootstrap=0.4). *S. hygroscopicus* (SHJG) has been clustered with *S. avermitilis* GlcP and *S. scabiei* Glk sequences, albeit with low support values of 0.3 and 0.5, respectively. Interestingly, *S. venezuelae* GlcP and Glk are not located in any of the larger clades. This might suggest divergence of these proteins compared to GlcP and Glk from closely related strains, such as *S. lavendulae*.

To summarise, these phylogenetic analyses indicate the presence of distinct groups of GlcP1 and GlkA orthologues among these strains, many of which correspond to stable clades formed in the corresponding autoMLST tree

3.3   GENETIC CONTEXT OF *GLCP* AND *GLK*

Comparing the genetic context of a given gene, that is the genes located upstream and downstream and their (predicted) gene products and evaluating the level of gene synteny present in these loci can aid identification of gene homologues and predict the function of a gene product in different species. Homologous genes often have similar genes located upstream and downstream. Adjacent genes can be organised in translational units, operons, or gene clusters, to coordinate their expression as the proteins they encode might be functionally related. Functionally linked proteins tend to evolve in a correlated manner as they are exposed to the same kind of constraint (Pellegrini *et al.*, 1999).

Although GlcP and Glk are both required for glucose important in *Streptomyces*, Glk is an orphan kinase and *glcP* and *glk* are located at distinct positions on the genome (Angell *et al.*, 1992; van Wezel *et al.*, 2005). While *glcP* expression is induced by glucose, *glk* is expressed constitutively (van Wezel *et al.*, 2005, 2007; Pérez-Redondo *et al.*, 2010). Evolutionary analyses of GlcP and Glk across Actinobacteria have indicated that, although gene family expansion is likely to have occurred or to be occurring in both cases, the evolutionary forces driving expansions might differ. To investigate this, gene synteny of genes located upstream and downstream of *glcP* and *glk* sequences from the group of *Streptomyces* strains were

compared. Regions of each genome of 5,000 nt upstream and downstream of *glcP* and *glk* are presented, genes encoding products with the same predicted functions have the same colour and hypothetical protein-encoding genes are shown in grey. The genome loci are presented adjacent to a phylogenetic tree constructed using the genetic context sequences.

3.3.1 GENE VARIATION DOWNSTREAM OF *glcP* IN THE GENETIC CONTEXT

The genetic context of *glcP* sequences from the group of *Streptomyces* strains appears to be more conserved upstream of *glcP* than downstream thereof, where there is greater variation in gene synteny (Figure 3-11, Table 3-2). Upstream of *glcP*, there are genes that encode a chromosome segregation protein, *smc*, a small leucine zipper protein-encoding gene, *slzA*, and an acylphosphatase-encoding gene. These are present in almost all shown loci with the exceptions of the duplicated *glcP* genes in *S. coelicolor* (*sco7153*) and *S. lividans* (*sliv_03515*), as well as the *S. rapamycinicus* (*m271*) genetic context locus, although this is presumably the result of missing sequence information and mis annotation. Further, the acylphosphatase gene is missing from the *S. clavuligerus* (*sclav*) and *S. tsukubaensis* (*b7r87*) loci, yet it is present further upstream of *slzA* on these genomes.

The level of conservation upstream of *glcP* is not present downstream of *glcP* where there is more variation in synteny. The most common genes in the downstream regions encode a luciferase-like monooxygenase, a putative nucleobase:cation symporter, the signal recognition particle receptor FtsY, a putative DNA polymerase/primase and a NsdA-type regulator. The *S. clavuligerus* and *S. tsukubaensis* loci contain a gene that codes for an acetyltransferase. The order of the genes downstream *b7r87_24980* suggests that the acetyltransferase might have been acquired horizontally by *S. clavuligerus* and *S. tsukubaensis.* While the former strain lacks the monooxygenase gene in this locus, both the monooxygenase and the acetyltransferase gene are present in the *S. tsukubaensis* locus.

The *S. coelicolor* and *S. lividans glcP1* loci form a stable sub-clade at the bottom of the largest group in the genetic context tree, which also contains the loci sequences from *S. hygroscopicus* (*shjg*), *S. avermitilis* (*saverm*), *S. scabiei* (*scab*) and *S. albus* (*slnwt*). Their respective GlcP sequences are also loosely grouped in the GlcP phylogenetic tree
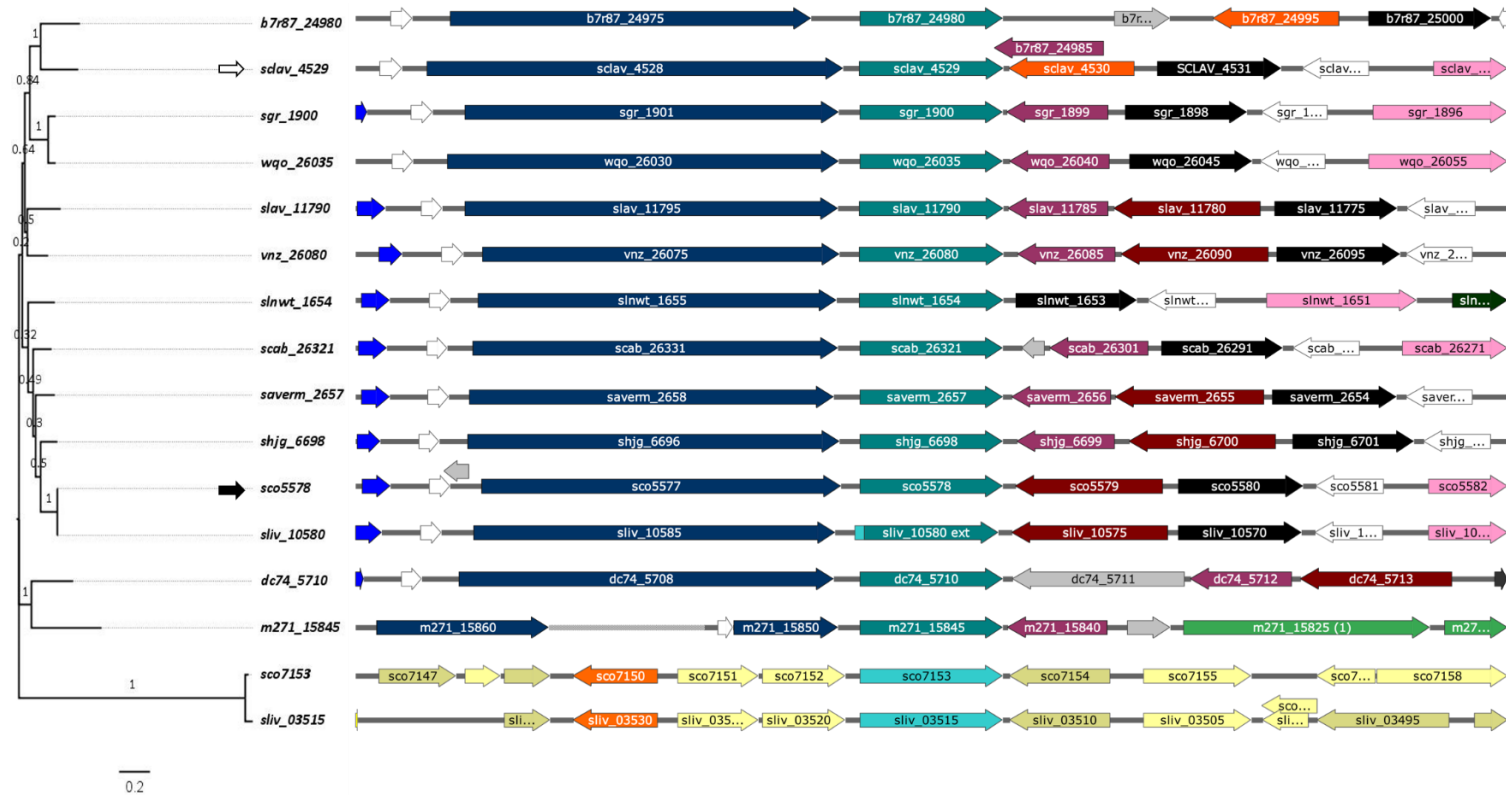
Figure 3-11: Genetic context of *glcP* from *Streptomyces* strains and corresponding phylogenetic tree. The sequences were aligned using MUSCLE and the Maximum likelihood method for tree construction and a bootstrap of 500 (Edgar, 2004; Kumar *et al.*, 2018). The branch lengths are drawn to scale in number of substitutions per site, the tips have been aligned (dashed grey lines) for visualisation purposes. Genome sequences comprised of the regions 5,000 nt upstream of downstream of *glcP* (shown in the centre, turquoise) were manually compared using available sequence information from StrepDB, KEGG, and NCBI databases. Genes encoding products with the same predicted function are shown in same colour (see Table 3-2).

Table 3-2: List of genes shown in the genetic context loci in Figure 3-11. Colours, names, and predicted gene product functions are listed for all genes except for the genome loci of the duplicated *glcP* genes in *S. coelicolor* and *S. lividans*.

Colour	Location relative to *glcP*	Name (length in nt), predicted gene product function
turquoise		*glcP* (1,500), glucose permease
teal	upstream	*smc* (3,500), chromosome partition protein
white	upstream	*slzA* (200), small leucine zipper protein
blue	upstream	gene encoding an acylphosphatase (200)
red	downstream	gene encoding a nucleobase:cation symporter-1 (1,400)
purple	downstream	gene encoding a luciferase-like monooxygenase (1,000)
black	downstream	*ftsY* (1,200), fused signal recognition protein
white	downstream	gene encoding a DNA polymerase/primase (700)
light pink	downstream	*nsdA* (1,500), regulator
green (1)	downstream	*narlL* (2,400), sensor histidine kinase
green (2)	downstream	gene encoding a LuxR-type transcriptional regulator (600)
orange	downstream	gene encoding an acetyltransferase (1,200)
dark green	downstream	gene encoding an ammonium transporter (1,300)

The *S. coelicolor* and *S. lividans glcP1* loci lack the monooxygenase-like gene present on the loci of *S. hygroscopicus*, *S. avermitilis*, *S. scabiei* and *S. albus*. Further, the duplicated genes, *sco7153* and *sliv_03515*, are located in an entirely different location on the genome and, as the duplication event was limited to the *glcP* coding sequences plus 19 bases upstream of the start codon, the gene context is distinct from the ones of *sco5578* and *sliv_10580* ext (van Wezel *et al.*, 2005). *S. albulus* (*dc74*) and *S. rapamycinicus* (*m271*) loci form a separate clade in the tree as their genetic context sequences differ in the genes located downstream of *dc74_5710* and *m271_15845*, respectively, compared to the other sequences. *dc74_5710* and *m271_15845* share 85% of their nt sequence. The genes downstream of *dc74_5710* encode a hypothetical protein, the monooxygenase and a permease. Those downstream of *m271_15845* encode the monooxygenase, a small hypothetical protein, as well as a sensor histidine kinase and a LuxR-type regulator.

By comparing the GlcP phylogenetic tree and the adjacent genome loci, there appears to be comprehensible patterns of genes that may have evolved from an original locus that may have contained *smc*, *slzA* and the acylphosphatase gene upstream of *glcP*, and the monooxygenase, the nucleotide:cation symporter, *ftsY*, the DNA polymerase/primase and the NsdA-type regulator gene downstream of *glcP*. Strikingly, most of the divergence observed for these genetic context loci is focused downstream of *glcP*, in the immediate vicinity of which no transposon like sequences or phage genes were detected.

3.3.2 GENE SYNTENY IN THE *glk* GENE CONTEXT

Analogously to the gene context comparison carried out for *glcP* genes, genome regions 5,000 nt upstream and downstream of *glk* were extracted from the *Streptomyces* genomes (Figure 3-12, Table 3-3). The sequences were used to construct a phylogenetic tree, as before. Unlike with the *glcP* gene context sequences, there is hardly any variance in the genes as the loci are almost identical. Directly upstream of *glk*, there is a small gene encoding a protein, for which no further function has been predicted. In *S. coelicolor*, however, SCO2127 has been shown to be essential to restore glucose dependent CCR, suggesting its involvement in this regulatory process. Moreover, expression of *sco2127* has been observed to have a stimulatory effect on *glkA* expression, and potentially also, on *glcP* expression (Guzman *et al.*, 2005).
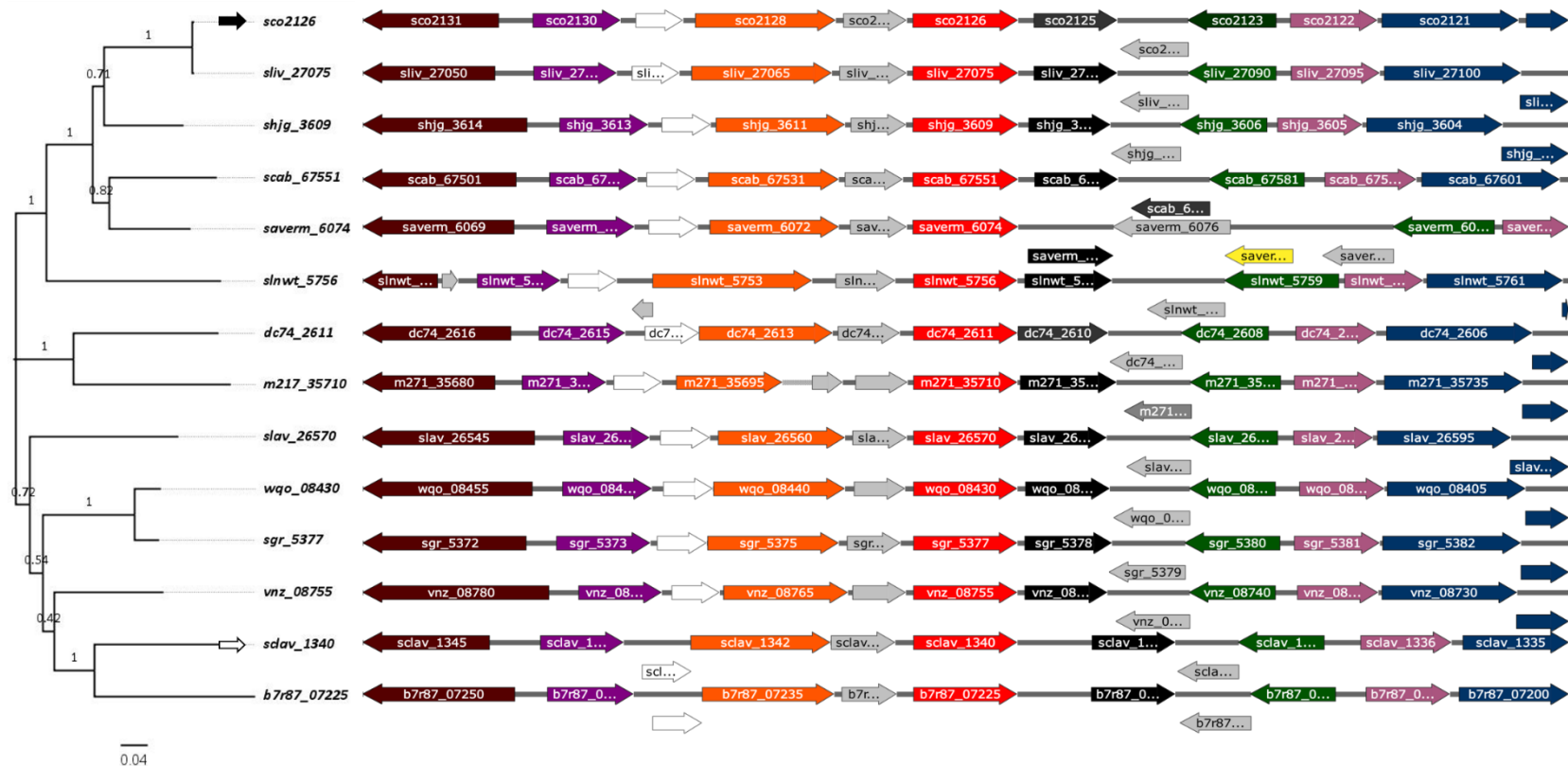
Figure 3-12: Gene context sequences containing 5,000 nt upstream and downstream of *glk* (red) in *Streptomyces* strains and corresponding phylogenetic tree. The tree was constructed using MUSCLE-aligned sequences and using the Maximum likelihood method with a bootstrap of 500 (Edgar, 2004; Kumar *et al.*, 2018). The tree is drawn to scale, the distance for 0.04 substitutions per site is indicated by the scale bar. Tips have been aligned for better visualisation. Genes for which products have the same predicted functions are shown in the same colours (Table 3-3). Positions of *sclav_1340* and *sco2126* loci are highlighted by the white and black arrow, respectively.

93

Table 3-3: Genes defined in the *glk* genetic context loci in Figure 3-12. Colours, position of genes relative to *glk* and, names, lengths in nucleotides (nt) and predicted gene product functions are shown.

Colour	Location relative to *glk*	Name (length in nt), predicted gene product function
red		*glk* (900), glucokinase
grey	upstream	gene encoding a hypothetical protein (600)
orange	upstream	*asrA*-ATPase (1,200), arsenite/tail-anchored protein-transporting ATPase
white	upstream	gene encoding a polyketide cyclase (400)
purple	upstream	gene encoding a metallophosphoesterase (800)
brown	upstream	gene encoding an long chain fatty acid CoA ligase (1,800)
grey	downstream	genes encoding hypothetical proteins (700 and 600)
black	downstream	gene encoding an endonuclease (700)
yellow	downstream	*saverm_6077* (600), putative RNA polymerase ECF-subfamily sigma factor
green	downstream	gene encoding an esterase/lipase (1,500)
pink	downstream	gene encoding a 1-acyl-sn-glycerol-3-phosphate acyltransferase (800)
blue (1)	downstream	gene encoding a sensor histidine kinase (1,200)
blue (2)	downstream	gene encoding a LuxR-type regulator (600)

Upstream of this highly conserved gene, there are genes encoding an ArsA-ATPase, a polyketide cyclase, a metallophosphoesterase and a long chain fatty acid CoA ligase. Additional hypothetical protein-encoding genes are found upstream of *slnwt_5756* in *S. albus* and *m271_35710* in *S. rapamycinicus*. Downstream of *glk*, there are genes encoding an endonuclease, a hypothetical protein, an esterase/lipase, a 1-acyl-sn-glycerol-3-phosphate acyltransferase and two genes shown in blue that encode the sensor histidine kinase and response regulator of a two-component system. *S. albulus* and *S. coelicolor* allegedly lack the esterase/lipase encoding gene (grey) when it is present in both of their respective clade partners' genetic context loci, *S. rapamycinicus* (*m271_35715*) and *S. lividans* (*sliv_27080*). In fact, *dc74_2610* and *m271_35715* share 75% of their nt sequence and *sco2125* and *sliv_27080* are identical. It can, therefore, be deduced that *dc74_2610* and *sco2125* also encode an esterase/lipase.

The composition of the clades in the phylogenetic tree constructed with the genetic context sequences for *glk* resembles the clade compositions in the corresponding Glk tree (Figure 3-10B). For instance, *S. clavuligerus* and *S. tsukubaensis* form a stable clade (bootstrap=1) in the genetic context phylogenetic tree, which they also do in the Glk tree. In fact, *S. albus* (*slnwt_5756*) and *S. venezuelae* (*vnz_08755*) are the only strains whose genetic context loci have not been grouped with strains they were grouped with in the Glk phylogenetic tree.

Interestingly, the intergenic region between *glk* and the esterase/lipase-encoding gene in *S. clavuligerus* and *S. tsukubaensis* is extended by around 500 nt, to lengths of 678 nt between *sclav_1339* and *sclav_1340* and 668 nt between *b7r87_07225* and *b7r87_07220*. The additional bases, however, do not extend the *glk* open reading frame. The gene encoding the hypothetical protein downstream of the esterase/lipase is annotated as a putative membrane protein in *S. rapamycinicus* (*m271_35720*) and *S. scabiei* (*scab_67571*). Based on sequence homology, it is likely that for *S. albulus* and *S. hygroscopicus*, which form clades with *S. rapamycinicus* and *S. scabiei* in the Glk phylogenetic tree, respectively, *dc74_2609* and *shjg_3607* also encode putative membrane proteins as they share 66% and 74% of their nt sequences with *m271_35720* and *scab_67571*, respectively.

Comparing the *glk* genetic context, the corresponding phylogenetic tree and the Glk phylogenetic tree suggests that the *glk* gene context is evolving differently and, potentially, more slowly than the *glcP* gene context in this group of strains. Given the observation of gene family expansion in the form of a second ROK-family glucokinase in Streptomycetales, the evolutionary pressure acting on *glk* might have been shifted onto the second glucokinase, which might explain the level of conservation observed in the genetic context.

3.4   SYNONYMOUS AND NON-SYNONYMOUS CHANGES IN *GLCP* AND *GLK* CODING SEQUENCES

Evolutionary analyses of GlcP and Glk protein sequences using phylogenetic tools have indicated that gene family expansion might have occurred among *Streptomyces* species for both proteins. Yet, the level of conservation among streptomycete Glk appears to be greater than that for streptomycete GlcP, possibly due to the emergence of a second group of ROK-family glucokinases. To confirm the higher level of conservation among Glk proteins, additional analyses using the coding sequences, *glcP* and *glk*, were conducted. Evolutionary divergence can be estimated by examining the number of synonymous, not amino acid-altering, or non-synonymous, amino acid-altering, base substitutions per synonymous and non-synonymous site, respectively. The ratio of non-synonymous changes (dN) to synonymous changes (dS), dN/dS, indicates whether a protein-encoding DNA sequence is under functional constraint and, thus, under purifying selective pressure ($dN/dS < 1$) or diversification in the form of positive selection ($dN/dS > 1$) (Nei *et al.*, 1986; Kimura, 1991; Korber-Irrgang, 2000; Hurst, 2002).

3.4.1   THE NATIVE *glcP* GENE IN *S. CLAVULIGERUS* IS THE SUBJECT OF PURIFYING SELECTIVE PRESSURE

Given that the native *glcP* gene in *S. clavuligerus* allegedly lacks a promoter sequence causing glucose auxotrophy (Aharonowitz *et al.*, 1978; Garcia-Dominguez *et al.*, 1989; Pérez-Redondo *et al.*, 2010), this may have resulted in the accumulation of synonymous and non-synonymous substitutions within *glcP*. The ratio of dN/dS might, thus, have been shifted towards 1, indicating neutrality, whereas *glcP* in other glucose-utilising strains would be expected to be under purifying selective pressure. This hypothesis was tested using codon-aligned *glcP* sequences following the Nei and Gojobori method integrated into the SNAP tool (Nei *et al.*, 1986; Ota *et al.*, 1994; Ganeshan *et al.*, 1997).

96

The average dS, dN values and dN/dS ratios for pairwise alignments of all *glcP* sequences with every *glcP* sequence are presented in Table 3-4. The overall average dS and dN values for all pairwise comparisons of *glcP* sequences were 0.4500 and 0.1106, respectively, yielding an average dN/dS ratio of 0.2458. The average dS and dN values from pairwise comparisons ranged from 0.3755 (*sghj_6698*) to 0.5584 (*b7r87_24980*) and 0.0931 (*sghj_6698*) to 0.1542 (*slav_11790*), respectively.

The lowest and highest average dN/dS ratios were 0.2131 for pairwise comparisons with *b7r87_24980* from *S. tsukubaensis* and 0.3340 for *slav_11790* from *S. lavendulae*. As all the dN/dS ratios are nearing zero, this infers purifying selective pressure as the main evolutionary force acting on the *glcP* gene across the species. The dN/dS ratio for *sclav_4529* of 0.2244 is comparable to the total average for all pairwise comparisons of 0.2458, which indicates that the native *glcP* gene in *S. clavuligerus* is either still is or at least has been until recently, the subject of purifying selective pressure. This was confirmed by carrying out a codon based Z-test to calculate that the probability of strict neutrality, dN = dS, can be ruled out and an alternative hypothesis, here dN < dS for purifying selection, can be accepted (Kumar *et al.*, 2018). The probabilities calculated for each of the pairwise comparisons in the test for strict neutrality and the test for purifying selection are 0.0000, showing that the null hypothesis dN = dS can be rejected and confirming the likelihood of purifying selection controlling the accumulation of non-synonymous changes across the *glcP* gene, including *sclav_4529*.

Exceptions to this consensus are *sco5578*, *sco10580* ext., *sco7153*, yet with these genes sharing between 99.7% and 99.9% of their nt sequence, this was to be expected (Table 3-5). In fact, per pairwise comparison, a single synonymous change was detected. Interestingly, while probability values for both tests for *sco5578*, *sco7153* and *sliv_10580* ext. yielded values well above 0.05, values for the pairwise comparison of *sliv_03515* with *sco7153* are below 0.05, 0.0275 in the strict neutrality test and 0.0138 in the purifying selection test. However, no non-synonymous changes were detected in *sliv_03515* when compared to *sco5578*, *sco7153* or *sliv_10580* ext., indicating that this might be an artefact of the test.

Table 3-4: Jukes-Cantor corrections for multiple hits of the proportion of synonymous and non-synonymous changes observed per total number of predicted synonymous and non-synonymous sites, dS and dN, respectively (Korber-Irrgang, 2000). The values and ratios presented are averages of each *glcP*/*glk* sequence compared to the sequence specified in the first column. dN/dS ratios are smaller than one, indicating a purifying selective pressure is acting on the genes. The sequences are in order of the respective phylogenetic trees (Figure 3-10). P-values (*t*-test) showing that there is a significant difference between *glcP* and *glk* dS values, but not dN values or dN/dS ratios are shown.

Average	dS	dN	dN/dS
glcP pairwise comparisons	0.4500	0.1106	0.2458
sco5578	0.4628	0.1012	0.2187
sliv_10580 ext.	0.4543	0.0993	0.2186
sco7153	0.4613	0.0993	0.2153
sliv_03515	0.4589	0.0994	0.2166
saverm_2657	0.4151	0.1,000	0.2409
shjg_6698	0.3755	0.0931	0.2479
slnwt_1654	0.4328	0.1039	0.2401
scab_26321	0.4297	0.0979	0.2278
sgr_1900	0.4112	0.1066	0.2592
wqo_26035	0.4037	0.1050	0.2601
vnz_26080	0.4029	0.0986	0.2447
dc74_5710	0.4688	0.1210	0.2581
m271_15844	0.4403	0.1309	0.2973
slav_11790	0.4617	0.1542	0.3340
sclav_4529	0.5567	0.1249	0.2244
b7r87_24980	0.5584	0.1190	0.2131
glk pairwise comparisons	0.3958	0.1035	0.2615
dc76_2611	0.3708	0.1066	0.2875
m271_35710	0.3824	0.1111	0.2905
slav_26570	0.3643	0.1058	0.2904
wqo_08430	0.3704	0.087	0.2349
sgr_5377	0.3808	0.0841	0.2209
slnwt_5756	0.3777	0.0977	0.2587
sclav_1340	0.4717	0.1066	0.2260
b7r87_07225	0.5044	0.1163	0.2306
vnz_08755	0.3375	0.0921	0.2729
sco2126	0.3642	0.112	0.3075
sliv_27075	0.3698	0.1121	0.3031
saverm_6074	0.4165	0.1021	0.2451
shjg_3609	0.3787	0.1062	0.2804
scab_67551	0.4478	0.1112	0.2483
p-value	0.0026	0.1201	0.0528

Table 3-5: Codon-based Z-test results for *glcP* genes from *S. coelicolor* and *S. lividans*. Probability values of rejecting the null hypothesis, dN = dS, in the codon-based Z-test of strict neutrality (turquoise) and in favour of accepting the alternative hypothesis dN < dS for purifying selection (pink) are shown. Significant values (p-value<0.05, *t*-test) are highlighted in turquoise and pink for neutrality and purifying selection, respectively. The test was carried out in MEGA X (Kumar *et al.*, 2018).

	p-value			
	sco5578	*sliv_10580 ext.*	*sco7153*	*sliv_03515*
sco5578		0.1602	0.1602	0.0863
sliv_10580 ext.	0.0801		0.1602	0.0863
sco7153	0.0801	0.0801		0.0275
sliv_03515	0.0431	0.0431	0.0138	

Investigating the prevalence of synonymous and non-synonymous changes in *glcP* coding sequences across these species, has revealed a strong purifying pressure as the driver for limiting amino acid changing substitutions across the coding sequence. As the uptake and subsequent catabolism of glucose is a central pathway for many *Streptomyces*, this is not surprising. Interestingly, *glcP* from *S. clavuligerus* was no exception to this consensus. However, the d*S* and d*N* values for both *S. clavuligerus* and *S. tsukubaensis* are higher than the calculated average d*S* and d*N* values for all pairwise comparisons between *glcP* sequences, indicating that more base substitutions have occurred in these sequences when the other sequences included in the analysis are used as reference points. This, in turn, indicates greater evolutionary divergence for *glcP* in *S. clavuligerus* and *S. tsukubaensis* compared to *glcP* sequences from the other *Streptomyces* strains.

3.4.2   d*S* IS SIGNIFICANTLY SMALLER FOR *glk* SEQUENCES THAN FOR *glcP* SEQUENCES

The bottom half of Table 3-4 shows d*S*, d*N* values and d*N*/d*S* ratios for pairwise comparisons of *glk* sequences with the sequence specified in the first column. Average d*S* and d*N* values for all pairwise comparisons of *glk* sequences are 0.3958 and 0.1035, respectively, which is lower than average d*S* (0.4500) and d*N* (0.1106) values obtained for *glcP*. The difference between d*S* values calculated for *glcP* and *glk* sequences is significant, meaning that the proportion of synonymous changes per potential synonymous sites is smaller for *glk* sequences than for *glcP* sequences (p-value=0.0026, *t*-test). This suggests that base changes occur more frequently in *glcP* in this group of strains, which agrees with what had been observed for evolutionary analyses of GlcP and Glk. The difference of d*N* values between *glcP* and *glk* sequences, however, is not statistically significant (p-value=0.1202). Within a *glk* sequence, the calculated d*S* value was generally higher, ranging from 0.3375 (*vnz_08755*) to 0.5044 (*b7r87_07225*), while d*N* values ranged from 0.0841 (*sgr_5377*) to 0.1163 (*b7r87_07225*).

The total average ratio of d*N*/d*S* for all *glk* sequences is 0.2615, which is comparable to the average d*N*/d*S* ratio for *glcP* (0.2458) and, thus, not significant (p-value=0.0528). With dN/dS ratios being <1, this infers that purifying selective pressure is also acting on *glk* sequences. The codon-based Z-based test of neutrality and using the alternative hypothesis of d*N*<d*S* as a test for purifying selection was used to confirm this observation, for which all probabilities

but one were below 0.05. This indicated the statistical significance for accepting the alternative hypothesis of purifying selection. The probability values for strict neutrality and purifying selection for *sco2126* and *sliv_27075* were 0.1604 and 0.0802, respectively, meaning the null hypothesis cannot be rejected, indicating strict neutrality (*sco2126* and *sliv_27075* are identical but for two synonymous base changes).

Uniquely, d*S* and d*N* values for *glk* sequences from *S. clavuligerus* and *S. tsukubaensis* were higher than the calculated averages for *glk* pairwise comparisons, which was also the case for *glcP* sequences. This suggests that not only *glcP*, but also *glk* in these species have diverged more from the group of *glk* sequences included in the analysis than any of those sequences have.

3.5 PROMOTER SEQUENCE AND RIBOSOME BINDING SITE PREDICTIONS FOR *GLCP* AND *GLK*

GlcP from *S. clavuligerus* (SCLAV_4529) is functional when *sclav_4529* is expressed in a glucose-negative *E. coli* strain (Pérez-Redondo *et al.*, 2010). Given this information and the observation that *sclav_4529* is under purifying pressure as other streptomcyete *glcP* sequences, this suggests that a promoter sequence may not be absent. Therefore, the regions 200 nt upstream from the start codon of *glcP* sequences were compared. Potential -35, -10 and ribosome binding sites (RBS) were predicted upstream of each of the *glcP* sequences based on predictions made for *sco5578* (*glcP1*), which is known to have a functional promoter (Pérez-Redondo *et al.*, 2010). The upstream regions of *glcP* appear to be conserved across the strains (Figure 3-13, top alignment). The predicted -35 site, spacer and -10 motif present in almost all the upstream regions are the ones predicted for *sco5578*, which are 5`-TTGACGT-3` and 5`-AGGCATAGT-3`, respectively. The consensus sequences for -35 sites, spacers and -10 sites for all sequences except for the duplicated *glcP* loci in *S. coelicolor* (*sco7153*) and *S. lividans* (*sliv_03515*), *S. clavuligerus* (*sclav_4529*) and *S. tsukubaensis* (*b7r87_24980*) regions, are 5`-TTGACK-3` (-35), 5`-TMRDDWMWTRA-3` (spacer) and 5`-ABARMATARB-3` (-10), respectively. The RBS sequences are even more conserved with the consensus for all loci being 5`-HGGAGH-3`. There is but a single exception to this, *scab_26321* (*S. scabiei*), which, based to the alignment, is lacking an RBS.

101
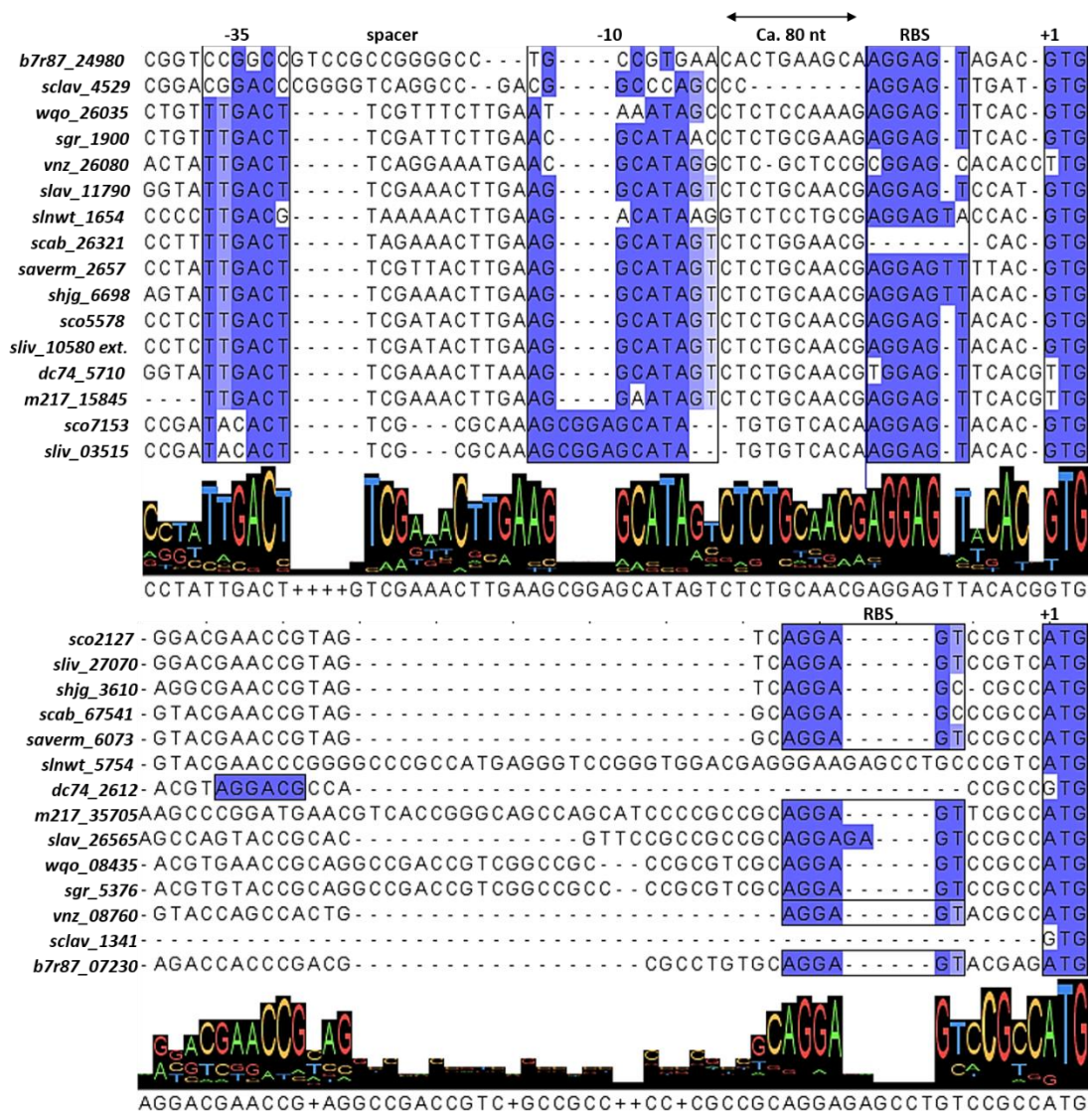
Figure 3-13: Multiple sequence alignment of the region upstream from the *glcP* (top) and *sco2127* orthologues (bottom). Start codons (blue boxes), promoter sequences and RBS were predicted either based on the *glcP1* promoter or by BProm (Clamp *et al.*, 2004; Waterhouse *et al.*, 2009; Salamov *et al.*, 2011). The sequences are in the same order as the corresponding genetic context sequences are arranged in Figure 3-11**Error! Reference source not found.** and Figure 3-12. No promoter sequences were identified upstream of *sco2127* orthologues.

The base changes and insertions upstream of *sco7153* and the identical *sliv_03515* compared to the respective *glcP1* upstream regions affect the -35, spacer and -10 regions but not the predicted RBS and might explain the low expression of the duplicated genes compared to *glcP1* in the presence of glucose (van Wezel *et al.*, 2005). The *S. clavuligerus* (*sclav_4529*) upstream region is distinctly different from all others except for that of *S. tsukubaensis* (*b7r87_24980*). The latter species is known to metabolise glucose but *glcP* expression is not induced by glucose (Ordóñez-Robles *et al.*, 2017).

There are major differences in the predicted -35 site, the spacer and the -10 site, with consensus sequences being 5`-CSGRCC-3`, 5`-SKSSGYTCRGGSC-3` followed by GA for *sclav_4529* and C for *b7r87_24980*, and 5`-YGSCSYRRS-3`, respectively. A recent study examining transcription start sites (TSS) at the -35 and -10 position in a high-quality *S. clavuligerus* genome sequence presented the following consensus sequences: 5`-NTGAC-3` for the -35 site and 5`-TANNNT-3` for the -10 site. The -35 and -10 sequences were found to be conserved in 60.4% and 91.0% of all TSS (Hwang *et al.*, 2019). Yet, the corresponding -35 and -10 sequences predicted upstream of *sclav_4529* from the alignment were 5`-CGGGCC-3` and 5`-CGGCCCAGC-3` and did, thus, not resemble the consensus for these sites. This suggests the absence of a promoter upstream of *sclav_4529* as it has been postulated previously (Pérez-Redondo *et al.*, 2010). In contrast, the similarities between *sclav_4529* and *b7r87_24980* upstream regions also suggest that there might be a promoter upstream of *sclav_4529* that might not resemble any consensus. Further, the alignment highlights not only the differences in the transcriptional regulatory sites of *sco7153* and *sliv_03515* compared to *sco5578* and *sliv_10580* ext., but also the distinction between glucose-dependent and glucose-independent promoter sequences, such as the ones for *sco5578* and *b7r87_24980*.

No promoter sequences nor potential RBS sites were predicted directly upstream of any of the *glk* genes. It has previously been shown in *S. coelicolor* that *glkA* expression is driven from a promoter of a gene encoding a hypothetical protein encoded by the gene upstream of *glkA*, *sco2127* (Guzman *et al.*, 2005). Therefore, aligned sequences of 200 nt upstream of *sco2127* and orthologous genes were compared (Figure 3-13, bottom alignment). The region directly upstream of the *sco2127* orthologues is highly conserved. This includes the start codon that

is 5`-ATG-3` for all genes but *dc74_2612* from *S. albulus* and *sclav_1341* from *S. clavuligerus*. These genes start with 5`-GTG-3`. Predicted RBS are also very similar with exceptions of those predicted upstream of *dc74_2612*, *slav_26565* (*S. lavendulae*) and *sclav_1341*. While the former has three additional bases separating the start codon from the RBS, there are two additional bases present within the predicted RBS upstream of *slav_26565*. The sequence directly upstream of *sclav_1341* is absent from the alignment. This is not an artefact of the alignment caused by low resemblance of the *S. clavuligerus* sequence with the consensus sequence, but rather it is due to this part of the sequence being entirely absent in *S. clavuligerus* (the region upstream of *sclav_1341* re-joins the alignment, Figure 3-14). With the *sclav_1341* upstream region potentially lacking an RBS, this gene might not be transcribed at all in *S. clavuligerus*. This is particularly interesting as the region upstream of *b7r87_07230* resembles the consensus sequence. Unlike *S. clavuligerus*, *S. tsukubaensis* is capable of internalising and catabolising glucose (Ordóñez-Robles *et al.*, 2017).

3.6 *IN SILICO* PREDICTION OF GENES INVOLVED IN CARBOHYDRATE TRANSPORT IN *S. CLAVULIGERUS*

Wildtype *S. clavuligerus* only efficiently utilises glycerol, maltose and starch as carbon sources (Aharonowitz *et al.*, 1978; Garcia-Dominguez *et al.*, 1989). The presence of the *glcP* and *glk* genes on its genome, however, demonstrates, that phenotypes do not necessarily reflect the metabolic potential of an organism (Garcia-Dominguez *et al.*, 1989; Pérez-Redondo *et al.*, 2010). To determine how many other carbohydrate uptake systems are encoded by the *S. clavuligerus* genome, a systematic *in silico* approach was taken following the methods of a study that identified carbohydrate importer-encoding genes in *S. coelicolor* (Bertram *et al.*, 2004). An overview of the genes that encode putative carbohydrate uptake systems that belong to the transporter families of MFS, ABC, PTS and aquaporin-like transporters as defined in the transporter classification database is given (Table 3-6) (Saier Jr *et al.*, 2014). In total, 75 genes were identified as either carbohydrate transporter-encoding genes, as genes coding for uptake system components or associated transcriptional regulators. This is very similar to 53 genes that were identified in *S. coelicolor*, that number does not include potential regulators (Bertram *et al.*, 2004). Twelve carbohydrate transporting ABC-system operons were identified, most of which are predicted to transport maltose. Unlike *S. coelicolor*, *S. clavuligerus* has two glycerol uptake systems encoded by *glpF1* and *glpF2* (Minambres *et al.*, 1992; Baños *et al.*, 2009).
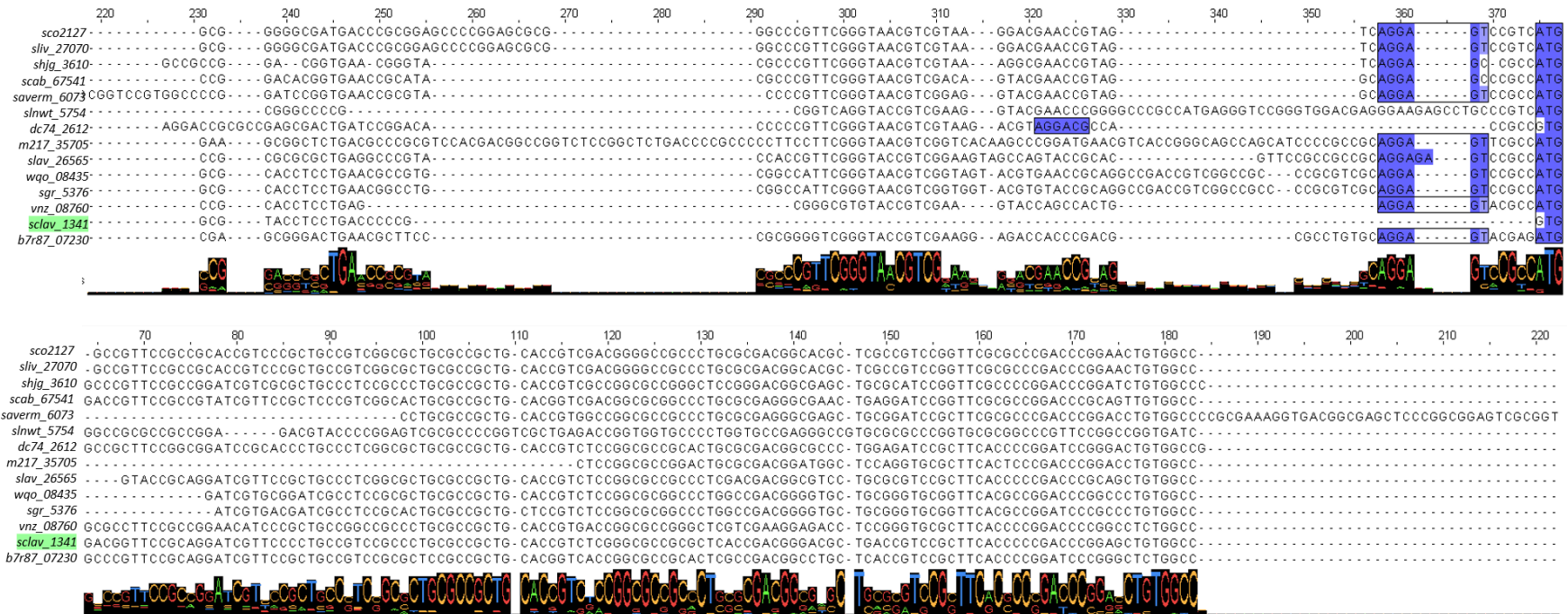
Figure 3-14: Extended multiple sequence alignment of the region upstream of *sco2127* orthologues (Clamp *et al.*, 2004; Waterhouse *et al.*, 2009). The alignment highlights that the absence of an RBS or promoter sequence upstream of *sclav_1341* is not an artefact of the alignment. The position of the *sclav_1341* upstream region is highlighted (green). The 377 nt long alignment displays an extension of the alignment in Figure 3-13 to show that the sequence upstream of *sclav_1341* lacks an identifiable RBS.

Table 3-6: List of genes encoding putative carbohydrate uptake systems and genes included in a given operon on the *S. clavuligerus* genome. The genes were identified following the strategy developed for *S. coelicolor* (Bertram *et al.*, 2004). Predicted protein functions, the number of transmembrane regions (TM domains) and protein orthologues in *S. coelicolor* (SCO) are shown. Genes encoding the glycerol and glucose permeases are highlighted in bold.

Gene	Product Function Prediction	TM domains	SCO Protein Orthologue
sclav_0197	Sugar permease	12	SCO2289
sclav_0284	ATP/GTP-binding protein	0	SCO1075
sclav_0285	Sugar permease	5	
sclav_0286	Sugar permease	5	
sclav_0287	Sugar-binding protein	0	SCO0065
sclav_0289	ROK-family transcriptional regulator	0	SCO1077
sclav_0317	Hypothetical protein	0	
sclav_0480	MFS transporter	12	SCO3331
sclav_0623, ptsA	Phosphoenolpyruvate-dependent sugar phosphotransferase	0	SCO1390
sclav_0624	Phosphoenolpyruvate-protein phosphotransferase	0	SCO1391
sclav_0631, **glpF1**	Glycerol uptake facilitator protein	5	
sclav_0632, glpK1	Glycerol kinase	0	
sclav_0684	MFS transporter	10	SCO1448
sclav_0876, gylR	Glycerol operon regulatory protein, IclR-family	0	SCO1658
sclav_0877, **glF2**	Glycerol uptake facilitator protein	6	SCO1659
sclav_0878, glkk2	Glycerol kinase	0	SCO1660
sclav_0879	Glycerol-3-phosphate dehydrogenase	0	SCO1661
sclav_0880	GntR-type regulator	0	
sclav_0881, araH	L-arabinose permease	10	
sclav_0882	ATP-binding protein	0	
sclav_0883	Sugar-binding protein/LacI transcriptional regulator	0	
sclav_1447, malE	Sugar-binding protein, CUT1 family	0	SCO2231
sclav_1448, malF	Sugar permease	6	SCO2230
sclav_1449, malG	Sugar permease	5	SCO2229
sclav_1450, aglA	α-glucosidase	0	SCO2228
sclav_1451	LacI-family transcriptional regulator	0	SCO2232
sclav_1826	ROK-family transcriptional regulator	0	SCO2657
sclav_1827	Sugar-binding protein	0	SCO2658
sclav_1828	Sugar permease	6	SCO2659
sclav_1829	Sugar permease	6	SCO2660
sclav_1830	Sugar hydrolase	0	SCO2661
sclav_1831	N-acetylglucosamine kinase	0	SCO2662
sclav_2016, ptsB, nagF	Phosphotransferase system component	IIB	SCO2905
sclav_2017	*nagE1*	5	SCO2906
sclav_2018, ptsC, nagE2	Phosphotransferase system component	IIC	SCO2907
sclav_2058	Sugar hydrolase	0	SCO2943
sclav_2059	Sugar permease	6	SCO2944
sclav_2060	Sugar permease	6	SCO2945
sclav_2061	Sugar-binding protein	0	SCO2946

sclav_2084	Sugar-binding protein	0	SCO2978
sclav_2085	Sugar permease	6	SCO2979
sclav_2086	Sugar permease	6	SCO2980
sclav_2087	Integral membrane protein	9	
sclav_4093, dasR	transcriptional repressor	0	SCO5231
sclav_4094, dasA	Sugar-binding protein	0	SCO5232
sclav_4095, dasB	Sugar permease	5	SCO5233
sclav_4096, dasC	Sugar permease	6	SCO5234
sclav_4097	Sugar hydrolase	0	SCO5235
sclav_4333	Sugar permease	6	SCO5428
sclav_4334	Sugar permease	6	SCO5429
sclav_4335	Solute-binding lipoprotein	0	SCO5430
sclav_4337	Helix-turn-helix protein	0	
sclav_4529, glcP	Glucose permease	12	**SCO5578**
sclav_4833	Solute-binding lipoprotein	0	SCO6005
sclav_4834	Sugar permease	5	SCO6006
sclav_4835	Sugar permease	6	SCO6007
sclav_4837	ROK-family transcriptional regulator (*rok7b7*)	0	SCO6008
sclav_4838	Solute-binding protein	0	SCO6009
sclav_4839, xylF	ATP-binding protein	0	SCO6010
sclav_4840, xylG	Xylose permease	12	SCO6011
sclav_4841, xylH	1-deoxy-D-xylulose-5-phosphate synthase 2	0	SCO6013
sclav_4884	Sugar permease	6	SCO6086
sclav_4885	Sugar permease	6	SCO6087
sclav_4886	Solute-binding protein	0	SCO6088
sclav_5018	GntR-family transcriptional regulator	0	SCO6256
sclav_5019	Sugar-binding lipoprotein	0	SCO6257
sclav_5020	Sugar permease	7	SCO6258
sclav_5021	ATP-binding protein	0	SCO6259
sclav_5022	Sugar kinase	0	**SCO6260**
sclav_5119	ROK-family transcriptional regulator	0	SCO6566
sclav_5120	ATP-binding protein	0	SCO6567
sclav_5121	Ribose/xylose/arabinose/galactoside permease	8	SCO6568
sclav_5122	Sugar-binding protein	0	SCO6569
sclav_5333	MFS transporter	11	
sclav_5695	MFS transporter	12	SCO3915

Interestingly, *sclav_4837* encodes Rok7b7, a ROK-family transcriptional regulator that has been characterised in *S. coelicolor* and, very recently, in *S. avermitilis* (see 1.3.11.3) (Świątek *et al.*, 2013; Lu *et al.*, 2020). In *S. coelicolor* and *S. avermitil*is, *rok7b7* is located directly upstream of the xylose ABC transporter genes, *xylFGH*. It can, thus, be inferred that that *sclav_4839*, *sclav_4840* and *sclav_4841* also encode a xylose ABC transport system. While the gene encoding *rok7b7* was upregulated in a *Glk*A-deficient strain in *S. coelicolor* (Gubbens *et al.*, 2012), in a *Δrok7b7* strain, *ppdK2* was also down-regulated. The latter enzyme is involved in *Glk*A-independent CCR in *S. coelicolor*, suggesting that expression of *rok7b7* and *ppdk2* might be coordinated (Gubbens *et al.*, 2012; Świątek *et al.*, 2013; Lu *et al.*, 2020). Given that the *glcP* genes in *S. coelicolor* and *S. avermitilis* have been identified as targets of Rok7b7 and the fact that Rok7b7 also affects antibiotic biosynthesis (Świątek *et al.*, 2013; Lu *et al.*, 2020), it would be interesting to investigate the role of this regulator in WT *S. clavuligerus* and compare it to expression profiles of industrial strains.

Carrying out the *in silico* search for genes that might be involved in carbohydrate uptake in *S. clavuligerus* highlights how little is understood about carbohydrate uptake systems and their control are in this species. The genes further might provide targets for improving an industrial strain's carbon catabolism and broadening the organism's carbon utilisation profile.

3.7   SUMMARY

The evolutionary analyses carried out with GlcP and Glk sequences from the ActDES database (Schniete *et al.*, 2021) revealed the presence of more than one Streptomycetales clade in the corresponding phylogenetic trees (section 3.1.3, Figure 3-7 and section 3.1.4, Figure 3-8). This suggests gene family expansion may have occurred for both components of the glucose uptake system. Yet, Glk sequences found in the same group were more similar in their amino acid sequence than GlcP sequences found in the same group. The higher level of conservation also appears to be present in the genetic context of *glk* in the group of selected *Streptomyces* strains, with little variation in gene synteny. This is in contrast to the genetic context of *glcP*, where there is some variation in gene synteny, especially downstream of *glcP* (section 3.3, Figure 3-11, Figure 3-12). Calculation of d*N*/d*S* ratios for *glcP* and *glk* indicates that these genes are under purifying selective pressure. The analysis showed that, although there are

no significant differences in dN/dS ratios across *glcP* and *glk* genes from this group of strains, dS is significantly smaller in *glk* sequences. Furthermore, dS and dN values calculated for pairwise comparisons of *glcP* and *glk* sequences with the respective sequences from *S. clavuligerus* and *S. tsukubaensis* were higher than the average dS and dN values. This suggests that *glcP* and *glk* in these strains may have diverged farther from the consensus sequences of the group (Table 3-4). Taken together, these observations suggest that evolution of GlcP and Glk might be uncoupled, which is not unlikely given that Glk is an orphan kinase and also acts as a mediator of CCR (Angell *et al.*, 1994; van Wezel *et al.*, 2007), yet both proteins are under functional constraint.

Examining the sequences directly upstream of *glcP* and *glk* highlighted the following: i) promoter sequences for *glcP* vary in a strain-dependent manner, ii) the -35 and -10 sequences predicted upstream of *sclav_4529* based on the alignment are dissimilar to identified consensus sequences in this species (Hwang *et al.*, 2019) and iii) there does not appear to be a promoter sequence or RBS upstream of *sclav_1340*. The latter observation might be interesting given that the homologous gene in *S. coelicolor*, *sco2127*, has shown to be important for *glk* expression, presumably due to the two genes being co-ordinately expressed from the promoter upstream of *sco2127* (Guzman *et al.*, 2005).

Following the approach that had been taken to identify transporter-encoding genes in *S. coelicolor* (Bertram *et al.*, 2004), a list of 75 genes potentially involved in carbohydrate uptake systems in *S. clavuligerus* was compiled and presented. These could serve as targets for future production strain improvement using a rational approach to understand and improve central carbon metabolism in *S. clavuligerus*.

4 Carbon Utilisation of and CA Production by SC2 and SC6

To gain insight into the previously reported unusual carbon source utilisation of *S. clavuligerus* (Aharonowitz *et al.*, 1978; Garcia-Dominguez *et al.*, 1989; Pérez-Redondo *et al.*, 2010), the WT strain DSM 738 and two commercial lineages, SC2 and SC6, were studied (Figure 4-1). DSM 738 is identical to the strain used to study glucose auxotrophy of WT *S. clavuligerus* (Pérez-Redondo *et al.*, 2010). SC2 and SC6 are early GSK CA production strains that are the results of random mutagenesis and selection as part of the GSK strain improvement programme. The aim of this chapter is to characterise DSM 738, SC2 and SC6 in terms of their growth, development, and CA production in the presence of glycerol and glucose and to, thus, verify catabolic differences first described in an internal GSK report (Dr Steve Kendrew, GSK, personal communication).

The type strain at the start of the GSK strain improvement process was DSM 738 (Figure 4-1A). Subsequent random mutagenesis, whole cell mutagenesis using UV radiation and treatment with NTG (Dr Andrew Collis, GSK, personal communication), and selection gave rise to SC2. Additional rounds of mutagenesis and selection eventually resulted in SC6 with three intermediate strains (SC3 – SC5). Internal GSK reports have described SC2 and SC6 as having distinctly different phenotypes with regards to their carbon source requirements. While SC2 was reported to grow normally on media containing glycerol as the main carbon source (designated car$^+$ phenotype), SC6 requires the addition of either maltose or dextrin (designated car$^-$ phenotype). This difference is not obvious when growing the strains on rich media, such as MS agar, GYM agar and L3M9 agar (Figure 4-1B). On these media, a developmental gradient was observed from MS to GYM to L3M9. All three strains exhibited poor growth with arrested development at the vegetative stage on MS agar, the sporulation medium of choice for many *Streptomyces* including *S. coelicolor*. Improved growth, white pigment production and a "fuzzy" morphology, indicative of aerial hyphae were observed on GYM agar. Only on L3M9 agar, did colonies produce the characteristic grey/green pigment indicative of sporulation of *S. clavuligerus* (Higgens *et al.*, 1971). L3M9 is a sporulation medium developed by GSK that has been adapted to the needs of their industrial *S. clavuligerus* lineages.
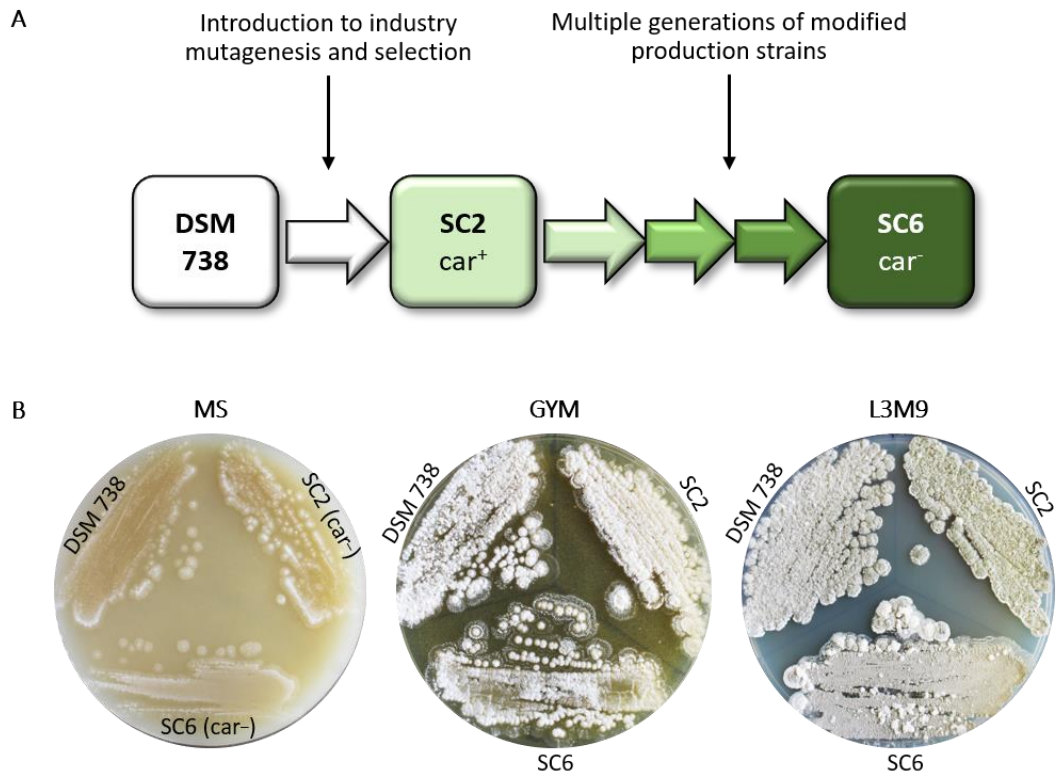
Figure 4-1: DSM 738, SC2 and SC6. (A) Schematic of the lineage of the strains used in this project: DSM 738 as the ancestor of SC2. Three industrial strains lie between SC2 and SC6. SC2 exhibits the designated car$^+$ phenotype, SC6 has altered carbon source requirements (designated car$^-$ phenotype). (B) Growth on MS agar (left), GYM agar (centre) and L3M9 agar (right, industrial sporulation media) after three weeks. Sporulation was detected on L3M9 medium (grey-greenish pigment production).

4.1   Growth of SC2 and SC6 in the Presence of Glucose

DSM 738, SC2 and SC6 were cultivated in defined media (solid and liquid) to confirm the previously reported car[+] and car[–] phenotypes and to gain an understanding of further differences in the strains' carbon utilisation profile. Growing the strains in defined media posed a major technical challenge. Therefore, cultivation conditions were first optimised in complex media (TSB) prior to growing them in defined liquid media (NMMP) containing either glucose (0.5%) or glycerol (1.0%). The latter carbon source was included in the analysis because it is the preferred carbon source of the wildtype strain (Aharonowitz *et al.*, 1978; Garcia-Dominguez *et al.*, 1989). Glycerol is also currently being used as a feedstock during industrial CA production.

The original NMMP recipe calls for 0.05% (w/v) casamino acids, which not only serve as a source of nitrogen for *Streptomyces* but also as a source of carbon (Kieser *et al.*, 2000). WT *S. clavuligerus* simultaneously utilises amino acids, such as L-alanine, L-asparagine, L-histidine and L-threonine as source of nitrogen and carbon (Aharonowitz *et al.*, 1978). The focus of the phenotypic characterisation was to identify differences in glycerol and glucose utilisation between the strains. Thus, casamino acids were substituted with ammonium sulphate, which provides nitrogen but not carbon. Yet, this resulted in total absence of growth or heavy pelleting of the cultures, which prevented the assessment of growth. Dispersed growth was observed using a combination of casamino acids (0.005% (w/v)) and ammonium sulphate. This modified NMMP medium is, henceforth, referred to as low amino acid NMMP. Growth was recorded for the strains under five different conditions: TSB, NMMP supplemented with glucose or glycerol, and low amino acid NMMP with glucose, either 0.5% (w/v) or 1.0% (w/v). Growth and development were also analysed on solid minimal medium in the presence or absence of glucose and glycerol. Growth rates for DSM 738, SC2 and SC6 strains in the tested media were significantly different (Figure 4-2, p-value=7.2×10^{-12} in a one-way ANOVA). An overview of all growth rates is shown in Table 4-1.

112
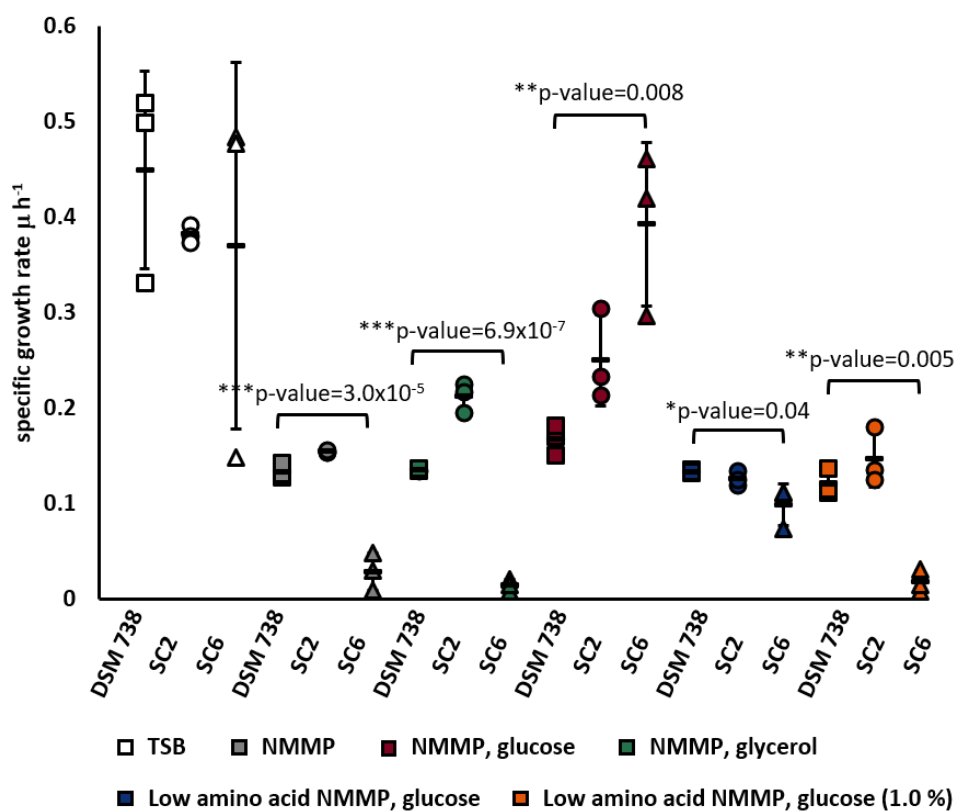
Figure 4-2: Specific growth rates of *S. clavuligerus* cultures. DSM 738, SC2 and SC6 are represented by squares, circles, and triangles, respectively. Culture media are shown in the legend, when stated glycerol (1.0%) or glucose (0.5% or 1.0%) were added. Differences between growth rates across all strains and conditions are significant (one-way ANOVA). Significance levels for each media are shown accordingly.

Table 4-1: Overview of mean specific growth rates (μ h^{-1}) of *S. clavuligerus* cultures presented in Figure 4-2. Data were collected from three independent cultures during exponential growth per strain per medium. Significance levels across all strains is indicated by one-way ANOVA p-values; p-values from Tukey's HSD test for significant differences between SC2 and SC6 cultures are shown in the last column. NS: not significant.

Medium / Strain	DSM 738	SC2	SC6	one-way ANOVA	Tukey's HSD test
TSB	0.45±0.10	0.38±0.01	0.37±0.19	NS	NS
NMMP	0.13±0.01	0.15±0.00	0.03±0.02	3×10^{-5}	0.001
NMMP, glycerol 1.0% (w/v)	0.14±0.00	0.21±0.02	0.01±0.01	7×10^{-7}	0.001
NMMP, glucose (0.5%)	0.17±0.02	0.25±0.05	0.39±0.09	0.008	NS
Low amino acid NMMP, glucose (0.5%)	0.13±0.00	0.13±0.01	0.10±0.02	0.04	NS
Low amino acid NMMP, glucose (1.0%)	0.12±0.01	0.15±0.03	0.02±0.01	0.0005	0.001

4.1.1 GROWTH OF DSM 738, SC2 AND SC6 IN TSB

Cultivating the strains in liquid media required a pre-germination step, which was carried out for 48 hours in a glycerol-based medium. The subsequent growth in TSB (glucose-based medium) was dispersed.

There were no significant differences between DSM 738, SC2 and SC6 specific growth rates in this media, which were 0.45 ± 0.10 h^{-1}, 0.38 ± 0.01 h^{-1} and 0.37 ± 0.19 h^{-1}, respectively (Figure 4-2, Table 4-1). However, data points collected for TSB-grown cultures are spread out, which might have contributed to the absence of a significant differences. The corresponding growth curves for DSM 738 and SC2 are similar while the SC6 growth curve is distinctly different (Figure 4-3). The DMS 738 and SC2 growth curves exhibit an exponential and stationary phase into which cultures transition after around ten hours. The SC6 growth curve, in contrast, peaks at approximately 15 hours and subsequently plummets prior to rising again after 24 hours. The second rise in data points is correlated with an increase in error bars, indicating a broader range of data points. After 40 hours, cell aggregates were visible, suggesting the drop and subsequent rise in absorbance might have been caused by fragmentation and a change in pellet morphology. Fragmentation is the process of cell pellet disassembly. A study investigating fragmentation of *S. lividans* in TSB containing sucrose (10% (w/v)) found that fragmentation occurs when cultures enter stationary phase. Death of mycelial fragments happens when no fresh nutrients are added to the culture and it coincides with hyphae emergence from aging pellets (Zacchetti *et al.*, 2018). Differences in pellet sizes, in addition to fragmentation and changes in pellet morphology, might account for the large data distribution as indicated by the larger error bars post 24 hours for SC6 cultures (van Veluw *et al.*, 2012; Nieminen *et al.*, 2013). However, this would have to be confirmed by microscopy.
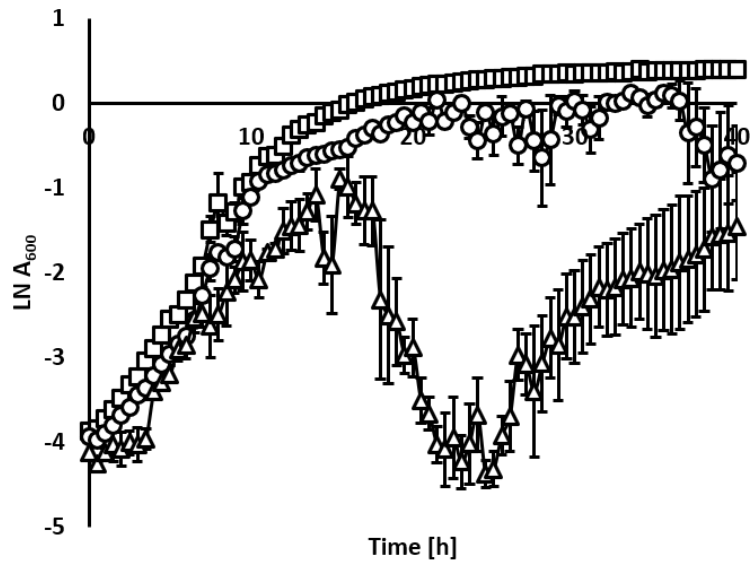
Figure 4-3: Growth curves of *S. clavuligerus* cultures in TSB. Growth of DSM 738 cultures is represented by squares, SC2 and SC6 growth are represented by circles and triangles, respectively. Strains were grown 24-well plates containing TSB. The LN of absorbance (600nm) is shown over the course of 40 hours. Error bars indicate the standard error for three data points per time point.

Growing the strains in TSB allowed optimisation of the pre-germination and cultivation protocol for subsequent cultivation in NMMP containing 0.05% (w/v) casamino acids (Hopwood *et al.*, 1985). In 500 ml *S. coelicolor* cultures grown in NMMP containing 0.2% (w/v) casamino acids and glucose, glucose utilisation did not commence prior to amino acid depletion (van Wezel, *et al.*, 1997). Given that WT *S. clavuligerus* utilises amino acids as a source of carbon (Aharonowitz *et al.*, 1978), the strains were grown in NMMP lacking any additional source of carbon other than the casamino acids to assess if the casamino acids would support growth of all three strains, prior to cultivating them in the presence of glycerol or glucose.

The growth rates for DSM 738, SC2 and SC6 growing in NMMP were 0.13 ± 0.01 h^{-1}, 0.15 ± 0.00 h^{-1} and 0.03 ± 0.02 h^{-1}, respectively, differences being statistically significant when comparing all three strains (p-value=3.0×10^{-5}, one-way ANOVA, Figure 4-2, Table 4-1). SC2 and SC6 growth rates were significantly different with a p-value of 0.001 (Tukey's HSD test). The differences in growth rates are reflected in the corresponding growth curves (Figure 4-4A). DSM 738 and SC2 cultures exhibited exponential growth that lasted up until 26 hours and 18 hours, respectively. Absorbance maxima for DSM 738 cultures were recorded at 27 hours, after which values decreased before slightly increasing after 34 hours. The biphasic growth phenotype resembles growth behaviour that occurs due to nutrient exhaustion, described for streptomycetes growing under nitrate-limitations. (Karandikar *et al.*, 1997). SC2 cultures transitioned from an exponential phase that lasted until approximately 14 hours into pre-stationary phase into lytic phase from approximately 31 hours. In contrast, the SC6 growth curve is flat due to poor growth of the cultures. Pellet formation caused the relatively large error bars on this growth curve. This suggests that while DSM 738 and SC2 were able to utilise the available casamino acids present in the media as nitrogen and carbon sources, SC6 was unable to do so to an extent that supported notable and dispersed growth.

Supplementing NMMP with glycerol (1.0% (w/v)) resulted in growth rates of 0.14 ± 0.00 h^{-1} for DSM 738, 0.21 ± 0.02 h^{-1} for SC2 and 0.01 ± 0.01 h^{-1} for SC6. These are significantly different with a p-value of 6.9×10^{-7} (one-way ANOVA).
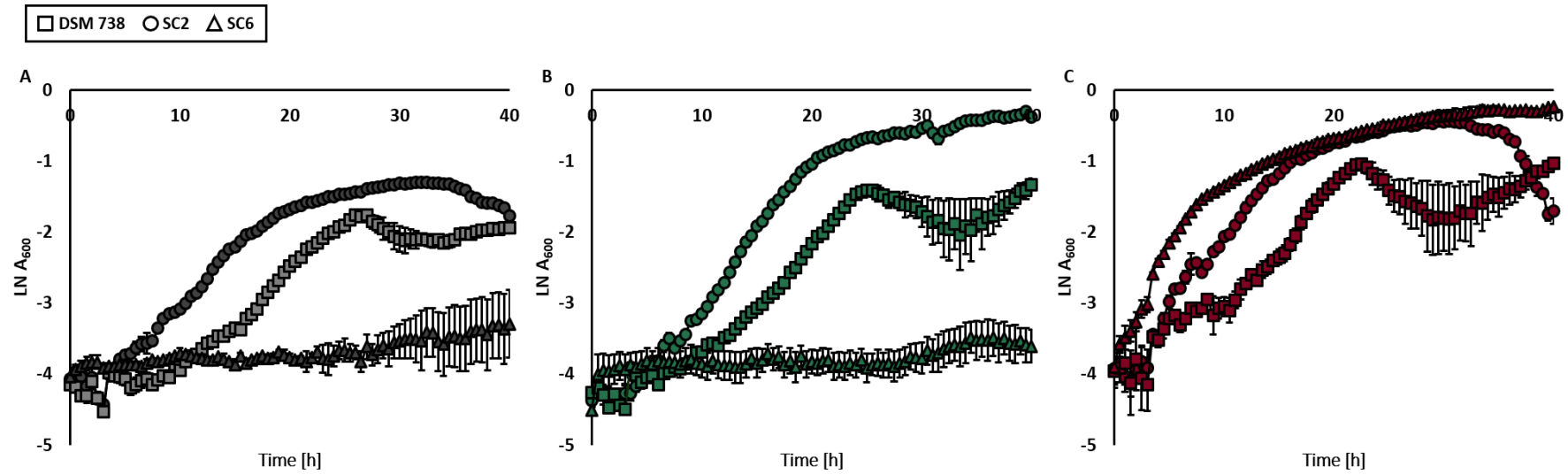
Figure 4-4: Growth curves of *S. clavuligerus* cultures in minimal media. Growth of DSM 738 cultures is represented by squares, SC2 and SC6 growth are represented by circles and triangles, respectively. Strains were grown 24-well plates containing (A, grey) NMMP, NMMP with 1.0% glycerol (B, green) or NMMP with 0.5% glucose (C, red). The LN of absorbance (600nm) is shown over the course of 40 hours. Error bars indicate the standard error for three data points per time point.

The difference between growth rates for SC2 and SC6 cultures was also significant (p-value=0.001, Tukey's HSD test). While the corresponding growth curves (Figure 4-4B) for DSM 738 and SC6 resemble the ones recorded in the absence of glycerol, the presence of glycerol resulted in an extended growth phase for SC2 cultures, from 14 hours to 20 hours. Furthermore, even after 40 hours of cultivation, no lytic phase was detected for SC2 cultures, which was previously the case in NMMP.

The addition of glucose to NMMP, did not alter growth rates for DSM 738 or SC2 greatly, 0.17 ± 0.02 h^{-1} for DSM 738, 0.25 ± 0.05 h^{-1} for SC2, yet had a major impact on growth of SC6 cultures (Figure 4-2, Table 4-1). The average specific growth rate for SC6 cultures growing in this medium was 0.39 ± 0.09 h^{-1}. The difference across growth rates for all three strains was significant, with a p-value of 0.008 (one-way ANOVA). Yet, the difference between SC2 and SC6 growth rates was not significant (Tukey's HSD test). DSM 738 and SC2 growth curves resemble those recorded in NMMP, the major difference to NMMP supplemented with glycerol-grown cultures being the pronounced lysis that occurred in SC2 cultures after 30 hours (Figure 4-4C). In stark contrast, SC6 growth curve is characterised by a steep but short exponential phase (up to 7 hours), a long pre-stationary phase (up to 34 hours) and a short stationary phase.

The slightly enhanced growth rates of DSM 738 and SC2 cultures in NMMP containing glucose might be the result of the strains utilising the amino acids in the media followed by residual glucose uptake after the point of amino acid exhaustion. Residual glucose uptake has been shown to occur in WT *S. clavuligerus* (Aharonowitz *et al.*, 1978; Garcia-Dominguez *et al.*, 1989). SC6, however, had not been able to grow in NMMP nor NMMP containing glycerol, indicating that, unlike DSM 738 and SC2, SC6 might be able to efficiently utilise glucose as a source of carbon in the presence of amino acids as nitrogen source.

4.1.3   EFFECT OF REDUCED AMINO ACID AVAILABILITY ON GROWTH OF DSM 738, SC2 AND SC6

Observed growth of DSM 738 and SC2 in NMMP supplemented with glucose might have been supported primarily due to the presence of casamino acids in the media. Therefore, the amount of casamino acids was reduced 10-fold in the presence of glucose with most of the nitrogen provided by ammonium sulphate. Two different glucose concentrations were tested, 0.5% (w/v) and 1.0% (w/v), and growth rates were determined as before (Figure 4-2,Table 4-1). In the presence of 0.5% (w/v) glucose, the reduction of amino acids resulted in decreased growth rates for all strain compared to NMMP supplemented with the same amount of glucose, from 0.17 ± 0.02 h^{-1} to 0.13 ± 0.00 h^{-1} for DSM 738, from 0.25 ± 0.05 h^{-1} to 0.13 ± 0.01 h^{-1} for SC2 and from 0.39 ± 0.09 to 0.10 ± 0.02 h^{-1} for SC6. The differences in growth rates across strains was significant (p-value=0.04, one-way ANOVA), yet this difference was not between SC2 and SC6 cultures. The overall reduction in growth rate might indicate that nitrogen availability was the growth rate-limiting factor in the low amino acid NMMP medium. This hypothesis is supported by the observation that, although the growth rate for SC6 in this medium is reduced compared to SC6 cultures growing in NMMP containing glucose, the growth phase of these cultures was extended from 8 hours to 20 hours (Figure 4-5A), suggesting the same amount of carbon supported growth for an extended period (12 hours). The growth phases of DSM 738 and SC2 cultures lasted up until 30 hours, which might have been due to these strains being able to use the reduced amount of casamino acids present in the medium.

The increase in glucose concentration (1.0% (w/v)) had a detrimental effect on SC6, for which the growth rate dropped to 0.02 h^{-1} $\pm$ 0.01, but not for DSM 738 (0.12 ± 0.01 h^{-1}) or SC2 (0.15 ± 0.03 h^{-1}) with an overall statistically significant difference between the cultures growth rates (p-value=0.0005, one-way ANOVA). The difference between SC2 and SC6 growth rates was also significant (p-value=0.001, Tukey's HSD test). This may indicate that a second, low-affinity, high-capacity transport system for glucose is present in *S. clavuligerus* that is inactive in SC6 but not in DSM 738 or SC2. This would not enable SC6 to grow in higher glucose concentrations, a phenomenon that has been observed previously in Actinobacteria (Briczinski *et al.*, 2008). The detrimental effect of the doubled glucose concentration is also visible in the corresponding growth curve (Figure 4-5B).
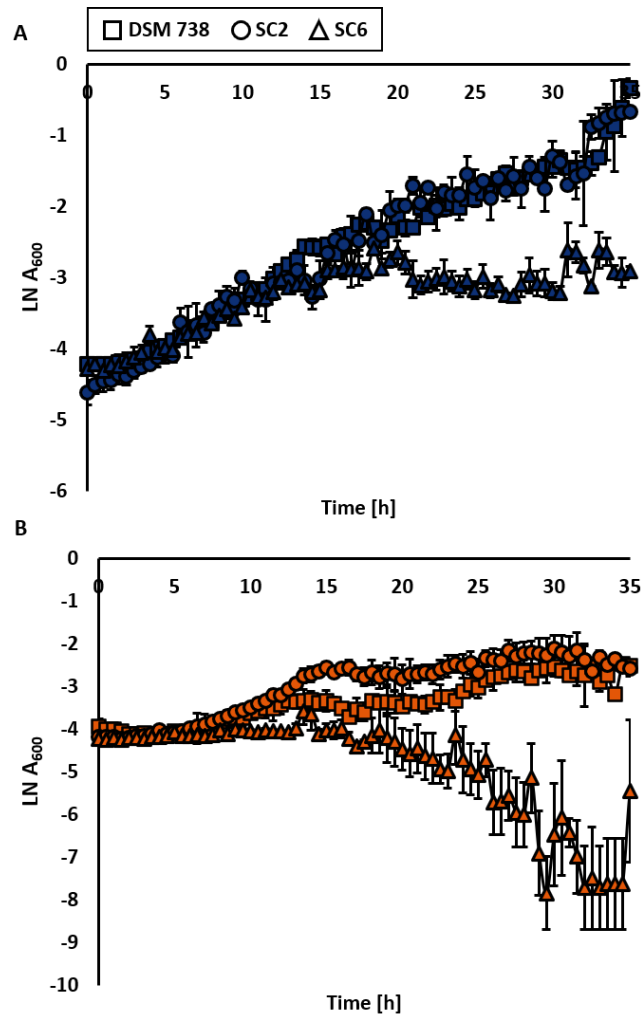
Figure 4-5: Growth of *S. clavuligerus* cultures. DSM 738 (squares), SC2 (circles) and SC6 (triangles) were grown in (A) NMMP with 0.5% glucose or (B) NMMP with 1.0% glucose. Growth is shown as LB of absorbance (600 nm) over the course of 35 hours. Error bars indicate the standard error across three data points per time point.

4.1.4 GROWTH AND DEVELOPMENT OF DSM 738, SC2 AND SC6 ON SOLID MINIMAL MEDIA

The onset of antibiotic production is correlated with the morphological switch from vegetative to aerial mycelium (McCormick *et al.*, 2012). A global regulator of development that also plays a role in the regulation of CA and cephamycin C production in *S. clavuligerus*, is BldG (Lawlor *et al.*, 1987). Deletion of *bldG* not only results in a bald phenotype, so lacking aerial hyphae, but also CA production is abolished due to a lack of *ccaR* expression (Bignell *et al.*, 2005). Therefore, the strains were grown on solid minimal media containing casamino acids only, casamino acids and glycerol or casamino acids and glucose to assess the strains developmental phenotypes in the presence of these carbon sources.

Surprisingly, despite the clear differences in growth rates and phenotypes in liquid minimal media, DSM 738, SC2 and SC6 colonies had the same colony morphologies when grown on solid minimal media (Figure 4-6). In the absence of glycerol, all colonies grew sparsely, and no aerial hyphae were visible, indicating development had arrested at the vegetative stage. The presence of glycerol, however, resulted in in denser growth and aerial mycelium formation. This suggests that, despite not being able to grow with glycerol as a carbon source in liquid media, SC6 must be able to sense, import and/or catabolise glycerol when grown on solid media. There are two types of glycerol uptake systems in *S. clavuligerus*, the inducible system, GlpF1K1D1, which is the main glycerol transport system in the presence of glycerol, and a constitutive one, comprised of GlpF2 and GlpK2. Activity of the latter system has been reported to be insufficient to support growth of *S. clavuligerus* with glycerol as the main carbon source (Minambres *et al.*, 1992; Baños *et al.*, 2009). Yet, flux through or sensing of glycerol by GlpF2K2 might be the reason for the observed colony morphology on agar.

122
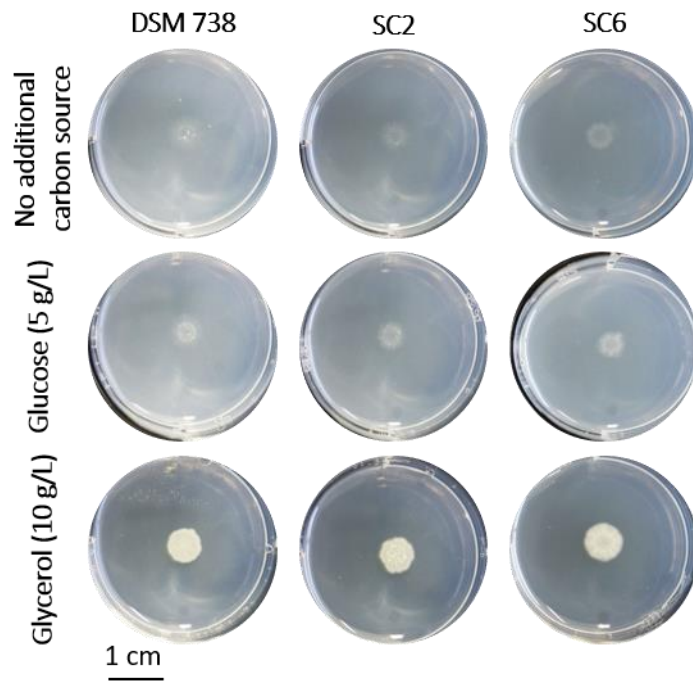
Figure 4-6: Growth of DSM 738, SC2 and SC6 on minimal media containing casamino acids and either no additional carbon source or supplemented with glucose, or glycerol. Images were taken after a four-week incubation period (scale bar = 1 cm). Spores were pre-germinated and normalised prior to spotting (1 μl with 1.5×10^3 pre-germinated spores). Only colonies growing in the presence of glycerol were dense and formed aerial mycelium.

4.2 EXPRESSION OF THE NATIVE GLUCOSE UPTAKE SYSTEM IN *S. CLAVULIGERUS* STRAINS

The apparent glucose auxotrophic phenotype observed for WT *S. clavuligerus* is thought to be caused by the low expression levels of *glcP* (Pérez-Redondo *et al.*, 2010). Therefore, expression of *glcP* and *glk* was investigated in DSM 738, SC2 and SC6.

RT-PCR was carried out using RNA extracted from TSB-grown DSM 738, SC2 and SC6 cultures after 24 hours since high growth rates were recorded across the strains. Expression of *glcP* and *glk* was compared across the strains in relation to *hrdB* expression (Figure 4-7). The *hrdB* gene encodes the major vegetative sigma factor σ^{hrdB} and is expressed constitutively throughout growth in liquid culture (Buttner *et al.*, 1990). It is, therefore, used as a control.

To establish if differences in *glcP* and/or *glk* expression exist between the strains, oligonucleotides for *hrdB*, *glcP*, and *glk* were designed to allow multiplexing of the reverse transcriptase (RT-)PCR and to semi-quantitatively compare the expression of these genes between strains. Prior to carrying out semi-quantitative RT-PCR, *hrdB*, *glcP* and *glk* were amplified from DSM 738, SC2 and SC6 genomic DNA (gDNA) to ensure amplicons had the expected sizes, 345 base pairs (bp), 553 bp, and 442 bp, respectively (Figure 4-7A). Two bands were amplified with the *glk*-specific primers; however, amplification of this additional band was avoided by decreasing the annealing temperature in the RT-PCR to 50°C. While individual amplicons were obtained for *hrdB*, *glcP* and *glk* from RNA when an RT-step was included in the amplification process, no amplicons were obtained when the RT-step was omitted (Figure 4-7B).

All three strains expressed *hrdB*, *glcP* and *glk* under the tested condition. Quantifying the abundance of *glcP* and *glk* transcripts relative to *hrdB* transcripts showed that *glcP* expression in SC6 was increased compared to *glcP* expression in DSM 738 and SC2 (Figure 4-7C). Relative *glcP* expression levels in DSM 738 and SC2 were comparable, 1.0 and 0.8, respectively, whereas, *glcP* expression was elevated in SC6, being 1.4-fold that of *hrdB* expression in this strain. DSM 738 expressed *glk* at a comparable level to *hrdB* (1.1), while expression of *glk* in SC2 (0.7) was reduced compared to this strain's *hrdB* expression. Unlike the relative expression of *glcP* in SC6, relative expression of *glk* was identical to *hrdB* expression in this strain (1.0). These data suggest that *glcP* expression is elevated in SC6 compared to DSM 738 and SC2, while *glk* expression is comparable across the strains.
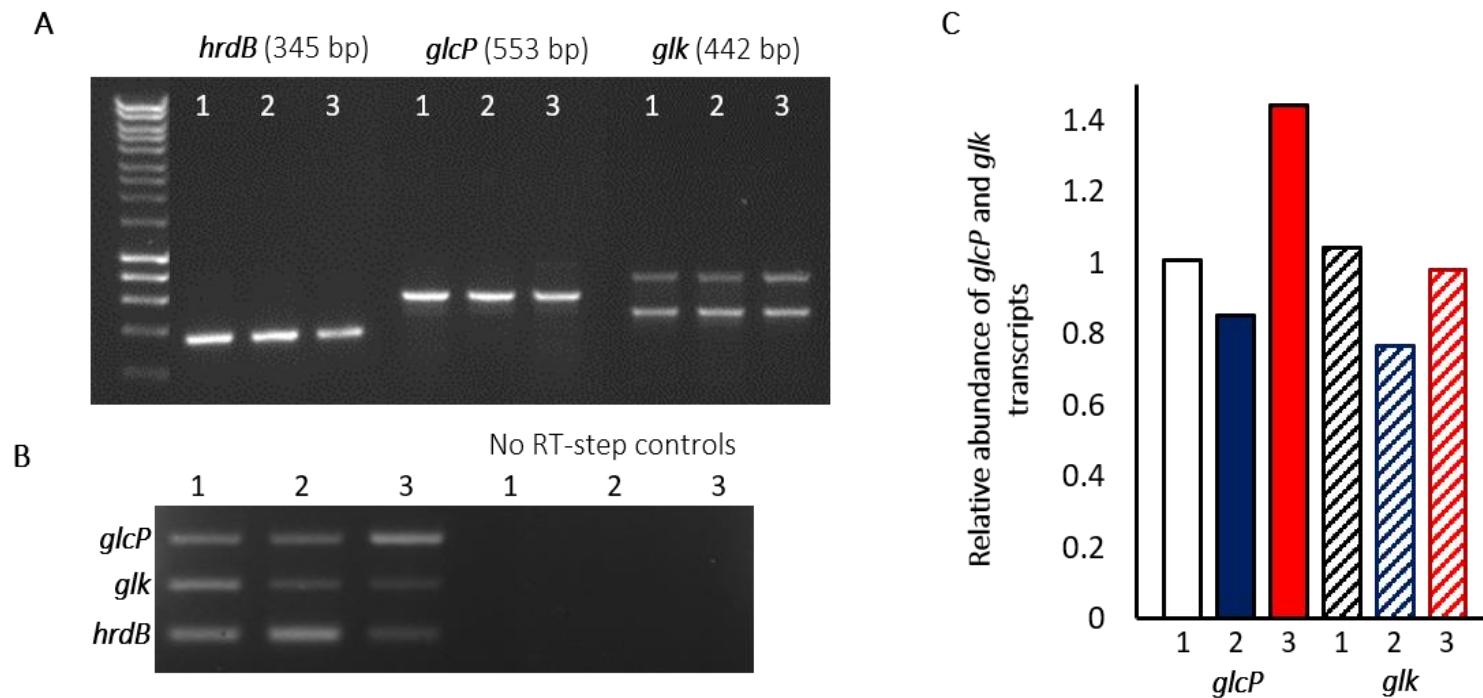
Figure 4-7: Transcriptional analysis of *glcP* and *glk* in DSM 738, SC2 and SC6 by semi-quantitative reverse transcriptase PCR. The agarose gels (1.0%) show (A) PCR bands for *hrdB*, *glcP* and *glk* obtained using the RT-PCR primers using gDNA (1: DSM 738; 2: SC2; 3: SC6) and (B) bands obtained from RNA after 30 cycles. The RNA was isolated from pre-germinated spores that had subsequently been cultivated in TSB for 24 hours. No amplicons are visible when the RT-step was omitted (No RT-step controls). (C) Bar charts show relative abundance of *glcP* and *glk* transcripts from DSM 738 (white), SC2 (blue) and SC6 (red) relative to *hrdB* expression. Semi-quantitative analysis was carried out by comparison of the brightness of corresponding bands on agarose gels.

SC2 and SC6 have been modified to produce the specialised metabolite CA in an industrial setting. Therefore, the next step in characterising SC2 and SC6 was the determination of CA production by these strains compared to the commercial reference strain, DSM 738, in the presence of glucose with this carbon source being of interest to GSK.

Determining CA yields from culture supernatant posed a major challenge due to a range of difficulties. Firstly, to compare CA production between strains, the yield of CA per cell dry weight (CDW) had to be ascertained, which proved problematic due to wide fluctuations in the data. Therefore, a correlation between the absorbance of a culture at 600 nm and CDW was established to eliminate the error caused by varying CDW values. For this purpose, DSM 738, SC2 and SC6 were grown in TSB (50 ml) over the course of 50 hours and CDW was determined at 24 and 50 hours, when the cultures were in distinct phases of their growth (Figure 4-8vA). While DSM 738 and SC2 cultures had reached their absorbance maxima at 36 hours, OD_{600} being 8.40±0.10 and 5.67±1.53, respectively, SC6 absorbance was highest at the 50-hour time point (OD_{600}=6.27±1.27). CDW was determined and normalised to an absorbance of 0.1 (Figure 4-8B). The values fluctuated widely for the 50-hour samples yielding an average CDW of 0.00. CDW 24-hour samples, in contrast. were comparable across replicates, when absorbance was at 0.47±0.05 for DSM 738, 1.65±0.12 for SC2 and 0.89±0.04 for SC6, respectively. Thus, a correlation coefficient between absorbance and biomass was established for cultures with dispersed growth from measuring CDW of 14 independent cultures at the 24-hour time point. A biomass of 0.59 mg/ml was determined to be equal to an OD_{600} of 0.1 with a standard deviation of 0.02.

Secondly, CA was only detected when DSM 738, SC2 and SC6 were grown in 50 ml cultures. The assay protocol was provided by GSK. The method, however, was reported back in 1982 and is based on the absorbance of the reaction product of CA with imidazole, which is 1-(4-aza-8-hydroxy-6-oxo)oct-2-en-1-oyl-imidazole. The absorbance maximum of this product is at 312 nm and is specific for intact clavulanic acid. Residual imidazole in the samples does not interfere with absorbance as the maximal absorbance of imidazole occurs at 280 nm. The lower sensitivity limit for this method was found to be 2 μg/ml (Bird *et al.*, 1982).

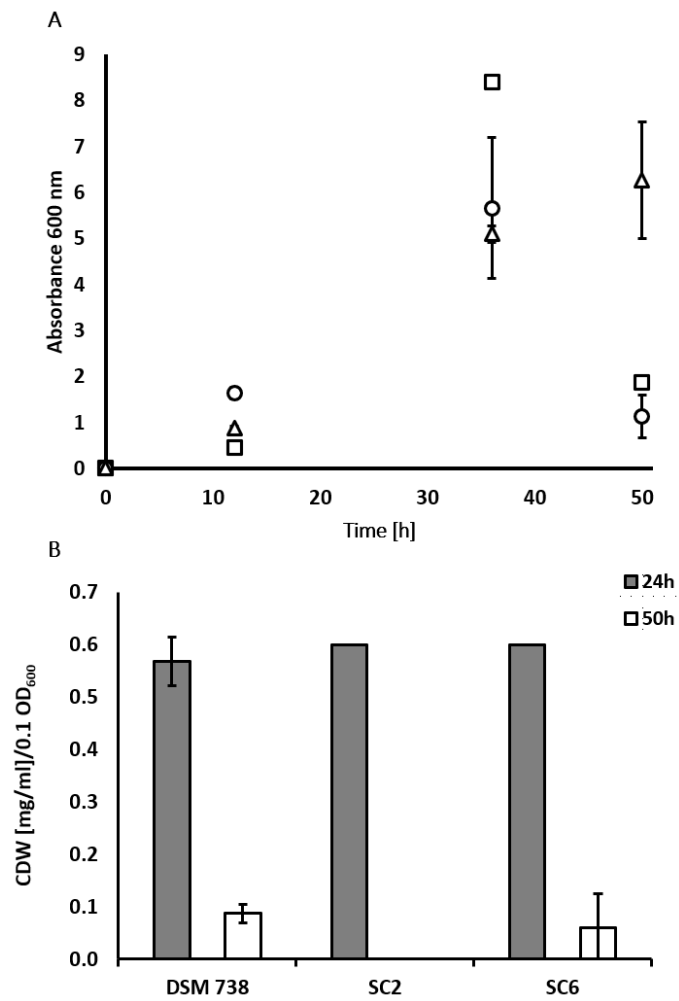Figure 4-8: Correlation coefficient between CDW and absorbance for *S. clavuligerus* cultures. (A) Culture absorbance (600 nm) at 24, 36 and 50 hours for 50 ml TSB cultures. DSM 738, SC2 and SC6 growth is represented by squares, circles, and triangles, respectively. (B) CDW of these cultures after 24 and 50 hours, error bars indicate the standard deviation for three data points per culture per time point.

The absence of CA in supernatant of cultures with a volume smaller than 50 ml (multi-well plates) might, therefore, have been caused by CA concentrations being below 2 μg/ml at the assayed time point (end-point measurement). The strains were, thus, cultivated in 50 ml TSB and NMMP, both containing 0.5% (w/v) glucose. No CA was detected in NMMP. However, CA was produced in TSB. Cultures were inoculated to an OD_{600} of 0.004, which is the equivalent of 1.5×10^4 pre-germinated spores (Hiltner, 2015). Samples were taken manually at four time points, after 24, 65, 90 and 140 hours, to minimise the loss in culture volume. Absorbance, pH, CA concentrations (Figure 4-9) were measured for DSM 738, SC2 and SC6 cultures and the yield of CA/CDW was determined (Figure 4-10).

Absorbance at the four time points was comparable across the strains, with maxima measured after 65 hours, which were 11.0, 12.0 and 7.0 for DSM 738, SC2 and SC6 cultures, respectively (Figure 4-9A). After 140 hours, the absorbance had plummeted to 0.6, 0.5 and 0.4 for DSM 738, SC2 and SC6 cultures, respectively. An increase in pH over time across all cultures was observed, from 6.8 at 0 hours to 7.0 after 24 hours to 8.2 for samples taken after 65 hours. The concentrations and the calculated yields of CA determined for DSM 738, SC2 and SC6 cultures fluctuated in a time-and strain-dependent manner (Figure 4-10). After 24 hours, CA concentrations in DSM 738 and SC6 cultures were comparable, 53±3 μg/ml and 57±2 μg/ml, respectively, while the concentrations measured in SC2 cultures were significantly higher, 216±4 μg/ml. The difference in CA concentrations at this time point across all strains was significant (p-value=1.2×10^{-13}, one-way ANOVA), as were the differences between DSM 738 and SC2 cultures (p-value=0.001, Tukey's HSD test) and between SC2 and SC6 cultures (p-value=0.001, Tukey's HSD test).

Previous studies have shown that the onset of CA production occurs when cell density is high (Paradkar *et al.*, 1995; Sánchez *et al.*, 1996; Ferguson *et al.*, 2016). This might explain why CA yields after 24 hours were high with mean values of 91±67 μg/mg, 45±7 μg/mg and 112±59 μg/mg for in DSM 738, SC2 and SC6 cultures, respectively. Although the overall concentration of CA in SC2 cultures was higher than the concentration measured in SC6 cultures, the latter had produced less biomass after 24 hours. Thus, the CA yield determined for SC6 cultures was greater than the yield from SC2 cultures, albeit it non-significant.
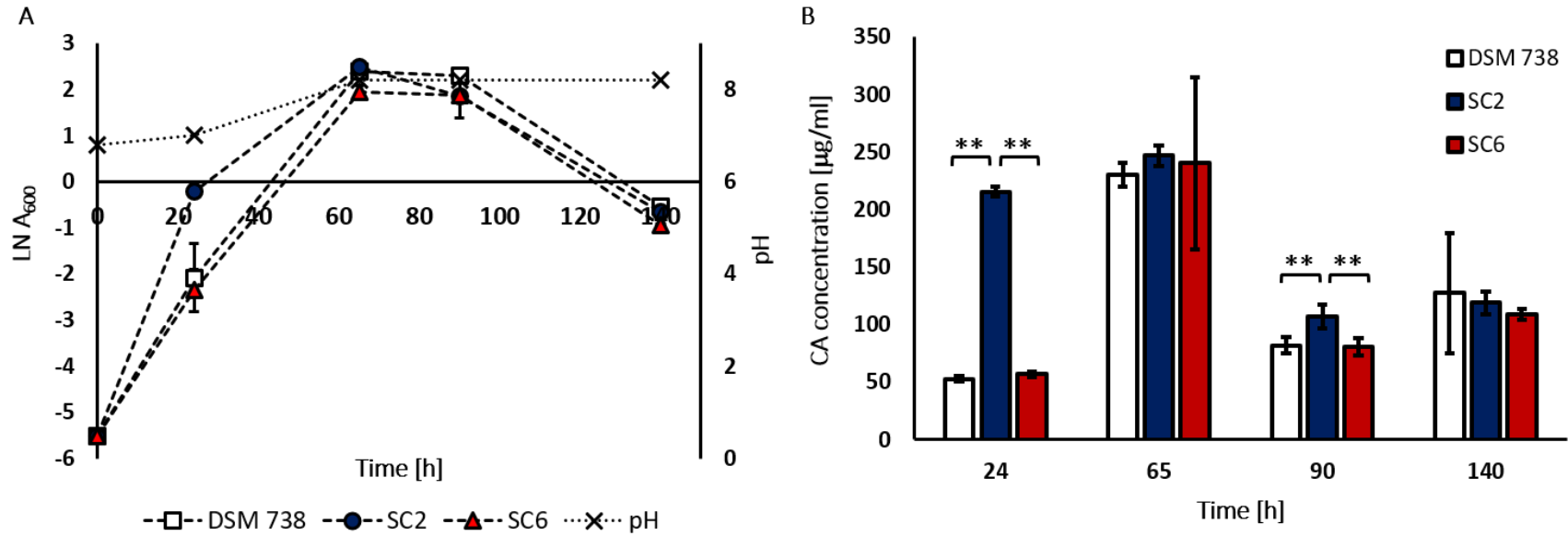
Figure 4-9: Growth and CA concentration in *S. clavuligerus* cultures. (A) DSM 738, SC2 and SC6 were grown in 50 ml TSB cultures over the course of 140 hours, OD_{600} and pH was assessed at 24, 65, 90 and 140 hours. Growth is shown as LN Absorbance (600 nm), DSM 738, SC2 and SC6 are represented by white squares, blue circles, and red triangles, respectively. Culture pH is depicted as x, which was the same across all cultures. (B) CA concentration is shown as white, blue, or red bars for DSM 738, SC2 and SC6, respectively, at given timepoints. Error bars indicate standard deviation. P-values <0.05 from one-way ANOVA were 1.2×10^{-13} at 24 h and 0.003 at 140 hours for concentrations across all strains. Statistically significant concentration differences between two strains are indicated by asterisks (*post hoc* Tukey's HSD test), with p-values of 0.001 between DSM 738 and SC2, and SC2 and DSM 738 at 24 h and p-values of 0.006 and 0.004 between DSM 738 and SC2, and SC2 and SC6 at 65 h, respectively.

A

B

hours	24		65		90		140	
	C_{CA} [μg/ml]	CA yield [μg/mg]	C_{CA} [μg/ml]	CA yield [μg/mg]	C_{CA} [μg/ml]	CA yield [μg/mg]	C_{CA} [μg/ml]	CA yield [μg/mg]
DSM 738	53±3	91±67	230±11	4±0.4	82±7	1±0.1	127±52	36±11
SC2	216±4	45±7	247±9	3±0.1	107±10	3±1	119±10	38±11
SC6	57±2	112±59	240±75	6±3	81±8	2±0.4	109±5	47±2
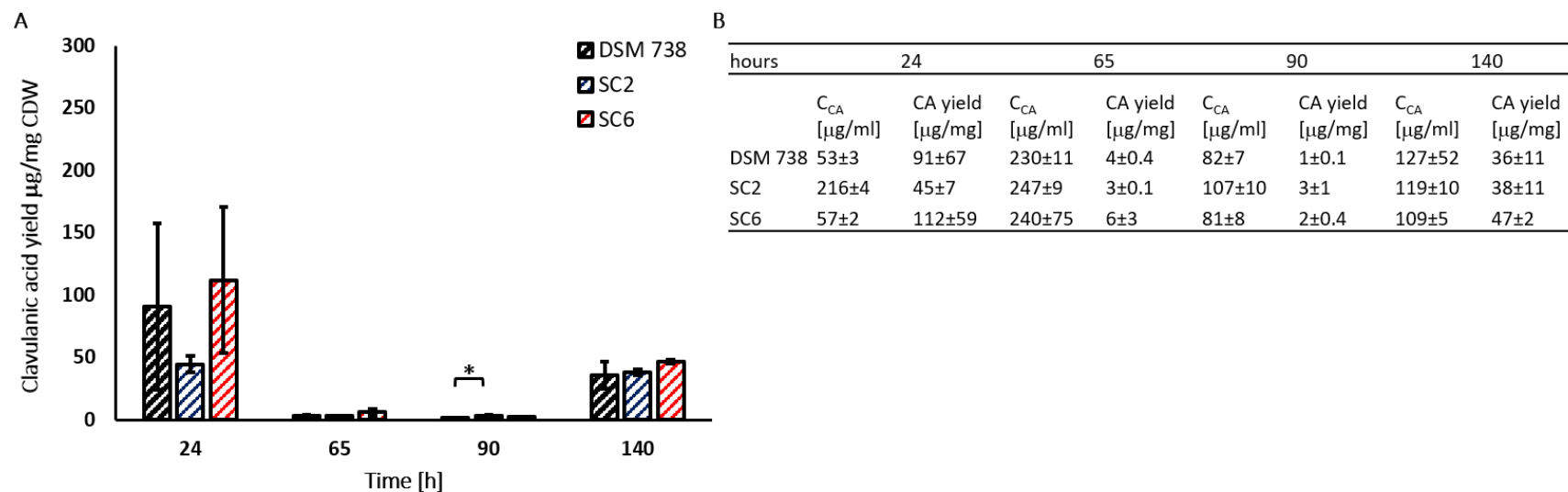
Figure 4-10: CA yields produced by *S. clavuligerus* cultures. (A) Samples were taken after 24, 65, 90, and 140 hours from TSB-grown DSM 738, SC2 and SC6. CA yields are depicted as CA concentrations in μg per mg CDW. CA yields in DSM 738, SC2 and SC6 cultures are shown as black and white, blue, and white, red, and white-striped bars. A significant difference between CA yields from all strains was detected after 90 hours (p-value=0.04, one-way ANOVA), caused by a significant difference between CA yields produced by DSM 738 and SC6 cultures (p-value=0.03, Tukey's HSD test, indicated by asterisk). (B) Overview of CA concentrations (C_{CA}) and yields in cultures at the timepoints.

At the 65-hour timepoint, the mean CA concentration measured in DSM 738 cultures was 230±11 μg/ml, the mean concentrations in SC2 and SC6 cultures being 247±9 μg/ml and 240±75 μg/ml, respectively. The associated CA yields in each set of cultures were low compared to the previous timepoint, 4±0.4 μg/ml, 3±0.1 μg/ml and 6±3 μg/ml for DSM 738, SC2 and SC6 cultures, respectively. The drop in yields compared to the earlier timepoint was caused by two factors: the increase in pH across all cultures would have promoted the well-established degradation of CA in neutral and basic pH (Bersanetti *et al.*, 2005), and the increase in biomass compared to the 24 hour time point would have decreased the yield if the production rate stayed the same.

After 90 hours, concentrations for DSM 738, SC2 and SC6 cultures were 82±7 μg/ml, 107±10 μg/ml and 81±8 μg/ml, respectively, which are significantly different (p-value=0.003, one-way ANOVA). The concentration determined for SC2 cultures was significantly higher than for DSM 738 cultures (p-value=0.006, Tukey's HSD test) and SC6 cultures (p-value=0.004, Tukey's HSD test). This indicates that the production rate has slowed down compared to the previous timepoint and more CA has degraded, which is reflected in the mean yields, 1±0.1 μg/ml, 3±1 μg/ml, and 2±0.4 μg/ml for DSM 738, SC2 and SC6 cultures, respectively, the difference across the strains being significant (p-value=0.04, one-way ANOVA). Yields produced in SC6 cultures were significantly higher than those produced by DSM 738 cultures (p-value=0.03, Tukey's HSD test). As the culture absorbance decreased between 90 and 140 hours, the concentrations slightly increased (127±52 μg/ml for DSM 738 cultures, 119±10 μg/ml for SC2 cultures, and 109±5 μg/ml for SC6 cultures) and the CA yields determined for DSM 738, SC2 and SC6 cultures were elevated compared to the 90-hour timepoint, 36±11 μg/ml, 38±3 μg/ml, and 47±2 μg/ml, respectively. The biomass present in SC6 cultures was lower than the biomass in SC2 cultures, resulting in the CA yield determined for the former to be higher than the latter. Differences in CA concentrations and yields were not significant at this timepoint.

Residual glucose uptake is known to occur in WT *S. clavuligerus* (Garcia-Dominguez *et al.*, 1989; Pérez-Redondo *et al.*, 2010) and SC6 is able to grow in the presence of glucose as the main carbon source. These data show that CA was present in all cultures at all sampling timepoints, indicating that the presence of glucose in the media does not cause full

repression of CA biosynthesis in any of the strains. This confirms what has previously been reported with regards to CCR in WT *S. clavuligerus*.

4.4 EXTRACELLULAR PROTEASE ACTIVITY

S. clavuligerus produces nutritious extracellular protease(s), the main proteolytic activity having been attributed to a zinc-dependent metalloprotease, encoded by *sclav_4359* (Bascarán *et al.*, 1990). Extracellular protease activity appears to be correlated with cephalosporin and cephamycin C biosynthesis in a nitrogen source-dependent manner (Aharonowitz, 1979; Braña *et al.*, 1985; Bascarán *et al.*, 1990, 1991). The correlative relationship between cephamycin C and protease production was also found in *S. lactamdurans* (Ginther, 1979). Yet, no such relationship has been reported for CA biosynthesis and the production of extracellular protease in *S. clavuligerus*, despite cephamycin C and CA production co-occurring during cultivation (Sánchez *et al.*, 1996). Given that casamino acids provide most of the nitrogen in seed fermentations, which can influence the yield of a natural product during fermentations, it is important to understand any cross-regulation that might exist between the production of extracellular protease(s) and CA biosynthesis in industrial strains. Therefore, extracellular protease activity was investigated under the same conditions under which CA production was previously examined (see 4.3). Furthermore, unlike CA production, proteolytic activity by extracellular protease(s) on agar is visible to the naked eye. If CA and protease(s) were produced simultaneously, observing proteolytic activity might serve as a proxy for studying carbon-dependent regulation of secondary metabolism in *S. clavuligerus*. Thus, the production of extracellular protease(s) on solid media in the presence of glycerol and glucose was investigated to test whether proteolytic activity is subject to carbon-dependent regulation.

4.4.1 PROTEASE PRODUCTION AND CA BIOSYNTHESIS IN LIQUID MEDIA

Protease activity was measured from supernatant of TSB-grown cultures (Figure 4-11) and proteolytic activity was detected in all samples at all timepoints. Generally, protease activity measured in DSM 738 cultures was lower than the activities detected in SC2 and SC6 cultures (Figure 4-11A). Significant differences in measured protease activity across all strains were detected after 24 hours (p-value=0.02, one-way ANOVA) and 65 hours (p-value=0.009, one-way ANOVA).
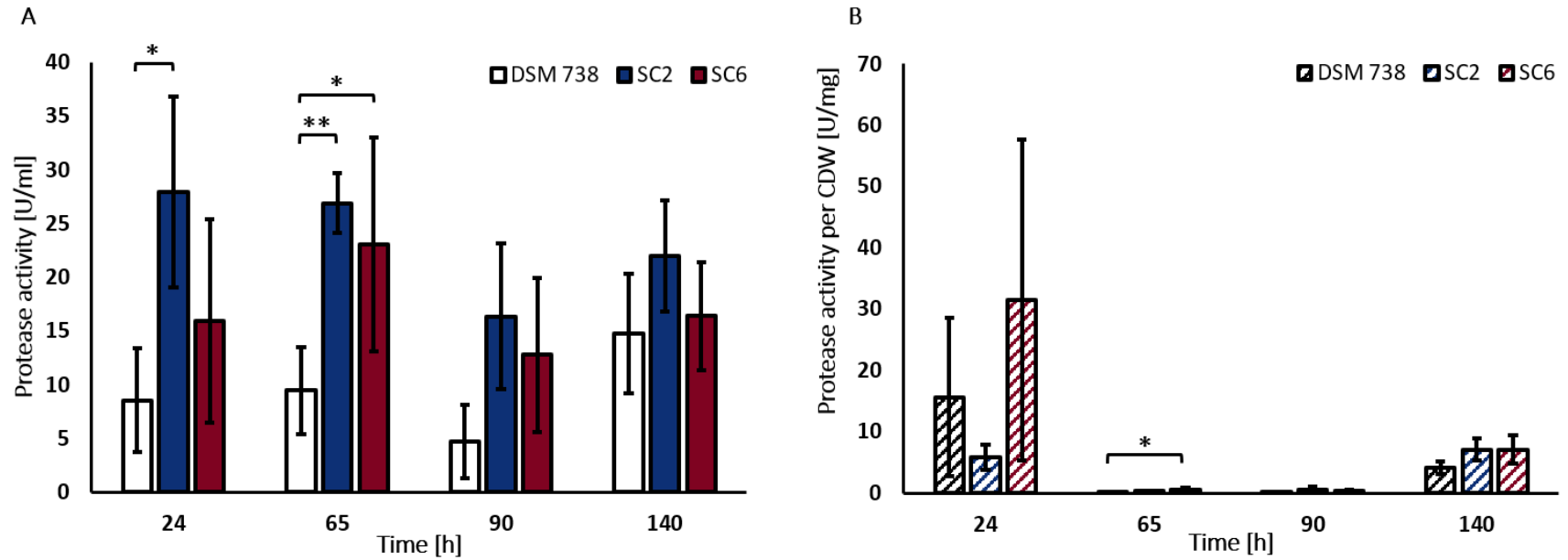
Figure 4-11: Protease activity in *S. clavuligerus* cultures. (A) Protease activity (U/ml) and (B) protease activity per CDW (U/mg, B) of TSB-grown DSM 738 (white), SC2 (blue) and SC6 (red) cultures are shown at the indicated timepoints with the error bars indicating the standard deviation. Significant differences between protease activity in cultures were identified by ANOVA and *post hoc* Tukey's HSD test. The following significant differences in protease activities were identified: between SC2 and DSM 738 at 24 h (*p-value=0.01), between SC2 and DSM 738 at 65 h (**p-value=0.009), between DSM 738 and SC6 at 65 h (*p-value=0.03). Protease activity per CDW of DSM 738 and SC6 were significantly different at 65 h with a p-value of 0.02.

Activities measured in DSM 738 samples at 24 and 65 hours were comparable, 8.6±4.8 U/ml and 9.5±4.1 U/ml, respectively, and dropped to 4.7±3.4 U/ml at 90 hours before increasing to 14.8±5.6 U/ml at 140 hours. These activities are significantly lower than the activities measured in SC2 samples after 24 hours with 28.0±0.1 U/ml (p-value=0.01, Tukey's HSD test), after 65 hours with 26.9±2.8 U/ml (p-value=0.009, Tukey's HSD test). Protease activities measured for SC2 cultures after 90 hours were 16.4±6.8 U/ml, and 22.0±5.2 U/ml after 140 hours. Protease activities measured in SC6 cultures at the four different timepoints were 16.0±9.5 U/ml (24 hours), 23.1±9.9 U/ml (65 hours, significant difference to DSM 738 cultures with p-value=0.03 in Tukey's HSD test), 12.8±7.2 U/ml (90 hours), and 16.4±5.0 U/ml (140 hours).

To allow a direct comparison between CA yields and protease activity, protease activities were calculated per mg CDW (Figure 4-11B). The graph highlights that protease activity/biomass was highest in all cultures after 24 hours and 140 hours, which is also when the highest CA yields were measured (Figure 4-10A). The only timepoint, at which differences in U/CDW between all strains were significant, however, was the 65-timepoint (p-value=0.03, one-way ANOVA). The differences in growth between the strains (Figure 4-9A) translated into SC6 producing the highest activity/biomass, 31.4±26.3 U/mg, followed by DSM 738, 15.6±12.9 U/mg, and SC2 with 5.9±2.0 U/mg after 24 hours. Given the increase in biomass over the time, the activities per biomass were low after 65 and 90 hours with DSM 738 having produced 0.1±0.1 U/mg at both time points, SC2 having produced 0.4±0.04 U/mg and 0.6±0.4 U/mg. Activity/biomass for SC6 cultures were 0.6±0.3 U/mg and 0.3±0.2 U/mg at 65 (significantly higher than DSM 738 U/mg, p-value=0.02, Tukey's HSD test) and 90 hours, respectively. DSM 738 also produced the lowest activities per biomass after 140 hours with 4.1±1.0 U/mg, while SC2 produced 7.1±1.8 U/mg and SC6 produced 7.0±2.3 U/mg.

Furthermore, RNA seq analysis data (courtesy of Dr John T Munnoch) from SC6 fermentations comparing the expression of genes at 52 and 65 hours, showed that expression of *ccaR*, the transcriptional regulator of early CA biosynthetic genes, and the gene encoding the metalloprotease gene, *sclav_4359*, was increased at the later timepoint (Table 4-2). This has previously been reported for WT *S. clavuligerus* (Álvarez-Álvarez *et al.*, 2014; Pinilla *et al.*,

2019) and is elaborated on in more detail in section 4.5, in conjunction with other changes in gene expression during CA biosynthesis.

These data show that CA biosynthesis and protease production co-occur in liquid, suggesting the processes might be co- or cross-regulated. This hypothesis is supported by the fact that both CA biosynthesis and extracellular protease activity is repressed by ammonium (Romero *et al.*, 1984; Bascarán *et al.*, 1990; Ives *et al.*, 1997). The next step was, thus, to investigate if protease activity was the subject of CCR.

4.4.2 EXTRACELLULAR PROTEASE ACTIVITY ON SOLID MEDIA

The production of extracellular protease(s) on solid media was investigated by spotting pre-germinated DSM 738, SC2 and SC6 spores on tryptone and milk powder-containing agar, henceforth referred to as milk agar, either lacking an additional carbon source, or containing glucose or glycerol. Each colony emerged from 1 µl spot inoculations containing around 1.5×10^3 pre-germinated spores, colony images were taken after four weeks of growth (Figure 4-12). To ensure the exposure to glycerol present in the pre-germination media would not skew the protease production on milk agar, spores were also cultivated in the presence of glucose (TSB) and subsequently spotted alongside spores that had only been pre-germinated for comparison. Pre-conditioning in TSB did not alter the observed phenotypes. Thus, only a representative set of images is shown for each strain on each type of milk agar.

After the four-week incubation period, DSM 738 and SC2 colonies covered the majority the surface area of the agar (9.6 cm^2) with colonies growing in the presence of glycerol being the largest. In the absence of glycerol, DSM 738 and SC2 colonies were white in the centre, which is indicative of aerial mycelia, but not around the edges. The agar had been cleared entirely of the opaque milk protein due to extracellular proteolytic activity. Yet, in the presence of glycerol, the agar was still opaque, indicating the absence of proteolytic activity caused by repression/depression or absence of induction of protease production. Furthermore, colonies had an altered morphology, appearing yellow in the centre instead of white and exhibiting increased crenulation. The yellow pigment is indicative of holomycin production (Kenig *et al.*, 1979). In contrast, development, and proteolytic activity of SC6 colonies appeared to be carbon source independent and, thus, glycerol insensitive.
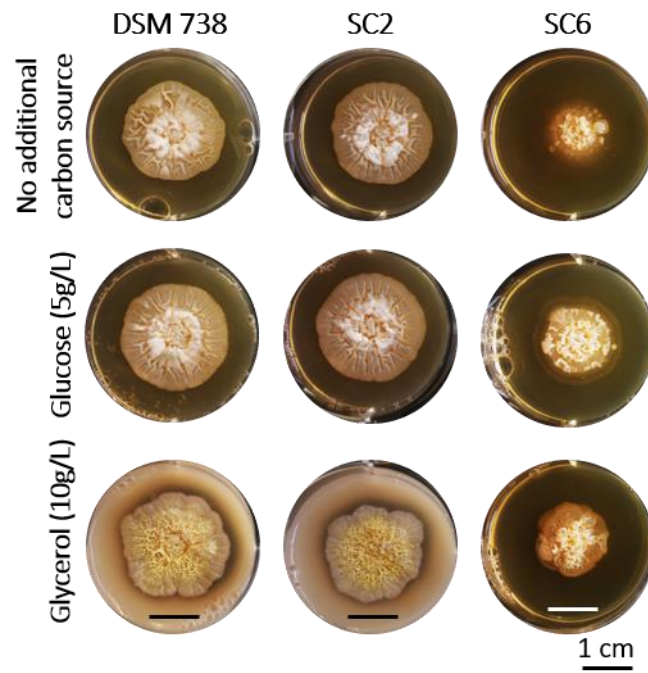
135

Figure 4-12: Agar clearing of *S. clavuligerus* colonies due to extracellular protease activity. Spotted colonies of DSM 738, SC2 and SC6 were grown on milk agar either lacking an additional carbon source or supplemented with glucose (0.5%) or glycerol (1.0%). Images were taken after a four-week incubation period (the scale bar = 1 cm).

All SC6 colonies were small and heterogeneous with sparse areas of aerial mycelium formation. The agar was cleared by the colonies on all types of milk agar, no yellow pigment was visible. These data suggest that proteolytic activity by DSM 738 and SC2 colonies is repressed in the presence of glycerol. The polyol is known to repress cephamycin C production in WT *S. clavuligerus* (Lebrihi *et al.*, 1988), but not CA as glycerol excess is essential for CA biosynthesis (Chen *et al.*, 2003) SC6, in contrast, seems insensitive to glycerol with regards to protease activity and exhibits an impaired developmental phenotype on milk agar. This might suggest that sensing and/or uptake and/or catabolism of glycerol might be impaired in SC6, which might also affect CA biosynthesis.

4.4.3   GLYCEROL DEPENDENT REPRESSION OF PROTEASE PRODUCTION IN WT *S. CLAVULIGERUS*

Given these findings and data from the literature, a model for glycerol-dependent repression of protease production by DSM 738 and SC2 is proposed (Figure 4-13). On milk agar, amino acids present in tryptone and yeast extract, which are found at concentrations of 5 g/L and 2.5 g/L, respectively, are taken up and utilised as a source of carbon and nitrogen prior to uptake and catabolism of glycerol (Figure 4-13A). L-cysteine and L-serine, for which a constitutive transporter exists in *S. clavuligerus* and which is found in tryptone, repress the major, inducible glycerol uptake system (Minambres *et al.*, 1992). Repression of glycerol uptake presumably takes place at the transcriptional level and could involve GylR or a signal produced thereof. GylR is the negative autoregulator and repressor of the *gylF1K1D1* operon (Hindle *et al.*, 1994). Once amino acids are depleted, signals involved in the cellular response to amino acid starvation might induce expression of *sclav_4359* in the absence of glycerol (Figure 4-13B). Although RelA, a key regulator involved in the stringent response triggered by amino acid starvation, is significantly downregulated in a *ccaR* disruption mutant, no clear relationship between the stringent response and protease production has been established for *S. clavuligerus* (Bascarán *et al.*, 1991; Álvarez-Álvarez *et al.*, 2014). Therefore, the production of extracellular protease is not suggested to be a result of the stringent response in the model presented here, yet this possibility cannot be entirely excluded. Expression of *sclav_4359* and the subsequent extracellular protease activity result in degradation of the milk protein present in the agar (50 g/L), thereby providing the cell with amino acids.
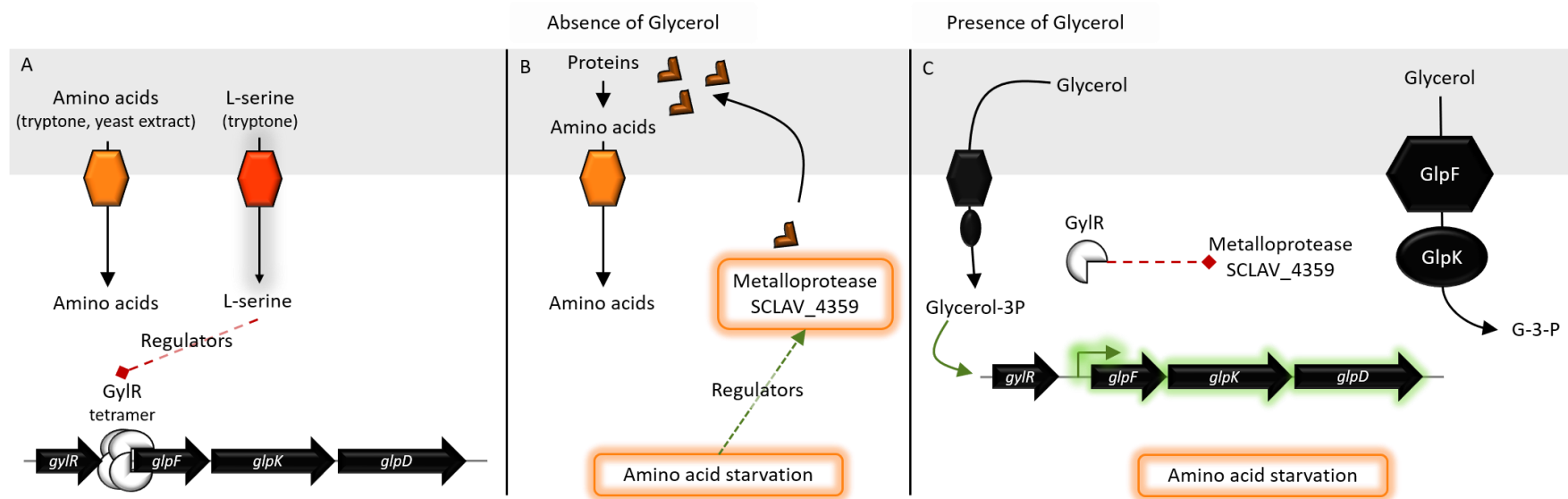
Figure 4-13: Schematic illustrating a proposed model for glycerol-dependent regulation of protease production in DSM 738 and SC2. Negative modulation (repression) is represented by red dotted diamond-shaped arrows, positive modulation is indicated by the green (dotted) arrows. (A) On milk agar, amino acids are preferentially utilised by *S. clavuligerus* and the import of L-serine (5 g/L in tryptone) represses expression of *glpF1K1D1* (Minambres *et al.*, 1992). This may involve GylR (Hindle *et al.*, 1994) or a signal generated by GylR. (B) Once amino acids become depleted, *sclav_4359* expression is induced in the absence of glycerol, which might be a direct result of the cell's response to amino acid starvation and involve other regulators. (C) In the presence of glycerol, small amounts of glycerol are imported via the minor, constitutive glycerol uptake system, which produces glycerol-3-phosphate, the inducer of GylR (Seno *et al.*, 1983; Smith *et al.*, 1988b; Hindle *et al.*, 1994). Dissociation of GylR upstream of *glpF* allows transcription of the operon. Unbound GylR may be involved in repression of *sclav_4359* expression.

138

In the presence of glycerol (Figure 4-13C), residual amounts of glycerol are internalised and phosphorylated by the minor, constitutive glycerol uptake system, GlpF2K2, to form glycerol-3-phosphate (G-3-P), the inducer of *gylF1K1D1* expression (Seno *et al.*, 1983; Smith *et al.*, 1988a; Hindle *et al.*, 1994). GylR dissociates from the DNA in the presence of G-3-P, leading to transcriptional activation. Unbound GylR might, thus, be involved in repression of *sclav_4359* expression when glycerol is available during amino acid starvation. This may require additional regulators. Given that CA biosynthesis and protease activity co-occur during cultivation in liquid (Figure 4-9, Figure 4-10, Figure 4-11), there is a possibility that signals and regulators that govern production of CA are also involved in regulating in metalloprotease production. The glycerol-insensitive protease phenotype observed for SC6 might be the result of the native GlcP-Glk system interfering with existing mechanisms that repress *sclav_4359* expression in the presence of glycerol. Given the conditional glycerol-negative phenotype SC6 exhibits (GSK report, see 4.1.2, Figure 4-4B), alterations in glycerol uptake and catabolism might impact the strains response to glycerol on milk agar.

4.5   CHANGES IN GENE EXPRESSION IN SC6 DURING CA BIOSYNTHESIS

SC6 is the most industrial strain among the strains used throughout this project. It was confirmed to exhibit a glycerol-negative phenotype, designated car$^-$, but simultaneously has improved growth in the presence of glucose, which might be the result of increased *glcP* expression. To gain an understanding whether glucose and glycerol uptake systems are expressed during CA biosynthesis in SC6, the expression of genes involved in glucose and glycerol uptake and catabolism were compared using RNAseq analysis data (courtesy of Dr John T Munnoch, Table 4-2). The data shows $\log_2$ fold changes in gene expression from 52 hours of cultivation compared to 65 hours. Compared to the earlier time point, CA biosynthesis was increased at the later time point, which is why biosynthetic regulatory genes, *ccaR* ($\log_2$ fold change=0.560) and *claR* ($\log_2$ fold change=0.573), were upregulated slightly, albeit not significantly (p-values=0.092 and 0.077, respectively). Generally, a two-fold change in gene expression, which is the equivalent of a $\log_2$ fold change of 1, is used as cut-off for identifying meaningful changes in gene expression (Quackenbush, 2002). However, the $\log_2$ fold changes in expression of genes shown in the table were also compared to the $\log_2$ fold changes of *ccaR* and *claR* as indicators for the increase in secondary metabolism that occurred between 52 and 65 hours.

Table 4-2: RNA-seq analysis of key genes involved in CA biosynthesis and regulation thereof, the metalloprotease-encoding *sclav_4359*, *adpA* and genes involved in central carbon metabolism. Significant $\log_2$ fold changes in gene expression at 65 hours (when CA production is occurring) compared to 52 hours in GSK production media in SC6 are shown. P-values showing the statistical significance of the $\log_2$ fold changes are listed in the last column (p=value<0.05 are **bold**). Genes whose expression was upregulated at the later timepoint when CA was being produced are highlighted in green, downregulated genes are shown in red.

Gene	$\log_2$ Fold change	p-value
ccaR	0.560	0.092
claR	0.573	0.077
pah2	1.118	**0.001**
cas2	0.837	**0.042**
oat2	0.743	**0.030**
oppA1	0.885	**0.008**
Metalloprotease, *sclav_4359*	0.702	0.060
gylR	0.728	0.421
glpF1	-0.009	0.987
glpK1	-0.204	0.714
glpD1	2.077	**0.004**
glpF2	-0.240	0.650
glpK2	-0.635	0.123
glcP	-0.513	0.225
glk	-0.233	0.670
adpA, sclav_1957	0.391	0.251

4.5.1 METALLOPROTEASE GENE *SCLAV_4359*

As mentioned previously, protease activity in *S. clavuligerus* is largely attributed to a single, potentially zinc-dependent, metalloprotease of 41.7 kDa (Bascarán *et al.*, 1990). This protease is encoded by *sclav_4359*. RNA sequencing data obtained from SC6 at two different time points during fermentation (52 hours and 65 hours), with CA being produced at the later timepoint, have revealed that *sclav_4359* expression is upregulated ($\log_2$ fold change=0.702, p-value=0.060) at 65 hours compared to 52 hours. This coincides with the upregulation of *ccaR* ($\log_2$ fold change=0.560, p-value=0.092), responsible for transcriptional activation of early CA biosynthetic genes, and *claR* ($\log_2$ fold change=0.573, p-value=0.077). ClaR, which regulates expression of late CA biosynthetic genes, is known to be regulated by CcaR (Pérez-Redondo et al., 1998; Santamarta et al., 2011). Further genes involved in CA biosynthesis that were upregulated at the later timepoint were *pah2* ($\log_2$ fold change=1.118, p-value=0.001), *cas2* ($\log_2$ fold change=0.837, p-value=0.042), *oat2* ($\log_2$ fold change=0.743, p-value=0.030), and *oppA1* ($\log_2$ fold change=0.885, p-value=0.008). These genes are considered early CA biosynthetic genes and are, thus, directly under the transcriptional control of CcaR.

4.5.2 GENES INVOLVED IN GLYCEROL AND GLUCOSE UPTAKE

Central carbon metabolic genes coding for components of the major glycerol uptake system, *glpF1*, *glpK1* and *glpD1*, as well as the repressor of the glycerol operon *gylR*, were included in the RNAseq analysis (Baños et al., 2009). CcaR negatively regulates *glpF1* and *glpK1* expression during CA biosynthesis in WT *S. clavuligerus* (Álvarez-Álvarez et al., 2014). This was also observed for SC6, in which *glpF1* and *glpK1* were weakly downregulated at the 65-hour timepoint compared to the 52-hour timepoint (*glpF1*: $\log_2$ fold change=-0.009, p-value=0.987; *glpK1*: $\log_2$ fold change=-0.204, p-value=0.714). Expression of *glpD1*, in contrast, was strongly upregulated in SC6 during CA biosynthesis with a $\log_2$ fold change=2.077 (p-value=0.004). This had previously not been observed in the WT (Álvarez-Álvarez et al., 2014). Genes encoding the minor uptake system, *glpF2K2* were downregulated at 65 hours, with $\log_2$ fold change=-0.240 (p-value=0.650) for *glpF2* and $\log_2$ fold change=-0.635 (p-value=0.123) for *glpK2*. Also downregulated were *glcP* and *glk*, with a $\log_2$ fold change=-0.513 (p-value=0.225) for *glcP* and a $\log_2$ fold change=-0.233 (p-value=0.670) for *glk*. This suggests that the expression of genes encoding components of the glycerol and glucose uptake systems in SC6 are downregulated during CA biosynthesis.

Interestingly, in a *ccaR* deletion mutant, the expression of *glpF1* and *glcP* varied in a growth-phase dependent manner, while all other genes mentioned here were consistently downregulated. During exponential phase, *glpF1* and *glcP* expression was upregulated compared to the WT strain, while expression of these genes was downregulated during stationary phase (Álvarez-Álvarez *et al.*, 2014). This indicates that while expression of *glcP* and *glk* might be either directly or indirectly influenced by CcaR, regulation of their expression during CA biosynthesis might involve other regulators.

4.5.3   THE GLOBAL REGULATOR ADPA

The coordinated upregulation of CA biosynthetic genes and *sclav_4359* during CA production conditions has previously reported for WT *S. clavuligerus* (Pinilla et al., 2019). This study also showed that expression of *adpA* (*sclav_1957*), which encodes the global transcriptional regulator AdpA, is upregulated when CA is produced. In SC6, *adpA* was also weakly upregulated at the later timepoint when *ccaR* was expressed (log$_2$ fold change=0.391, p-value=0.251). AdpA has been suggested to act as a positive modulator of CcaR, (López-García, Santamarta and Liras, 2010; Pinilla *et al.*, 2019). The relationship between AdpA and CcaR, however, appears to be reciprocal as *adpA* was significantly downregulated in a *ccaR* deletion mutant (Álvarez-Álvarez *et al.*, 2014). However, *ccaR* expression is not entirely dependent on the presence of AdpA, which requires *bldA* to provide the rare leucyl-tRNA for translation of UUA codons (Takano *et al.*, 2003). Although *ccaR* also starts with a TTA codon, its expression does not depend on *bldA* and the absence of AdpA resulted in downregulation but not abolishment of *ccaR* expression (Trepanier *et al.*, 2002; López-García *et al.*, 2010). The involvement of AdpA in regulating secondary metabolite production is well established in other streptomycetes as it impacts expression of over 1000 genes (Ohnishi *et al.*, 2005; Akanuma *et al.*, 2009; Higo *et al.*, 2012) and might, thus, also act as a regulatory link in co-regulating protease production an CA biosynthesis in *S. clavuligerus* (López-García *et al.*, 2010).

4.6   GENOME-WIDE COMPARISON OF SC2 AND SC6

SC2 and SC6 were mutagenised by UV radiation and treatment with NTG for production strain improvement. While exposure to UV light induces a range of damages that do not only directly affect DNA but also the intracellular pool of nucleotides, NTG treatment most often

induces base substitutions. Frameshift mutations are also elicited, albeit to a lesser extent. (Delić *et al.*, 1970). To gain an understanding of the genomic evolution of and any major genetic differences between SC2 and SC6, and compared to WT *S. clavuligerus*, chromosomal genome sequences were compared to one another (chromosome sequences were obtained from a related project in the laboratory and were kindly shared by Dr John T Munnoch). In total, 48 mutations were detected in 45 in coding sequences (*sclav*), of which 12 occurred in SC2 and 35 were identified in SC6 (Table 4-3 and Table 4-4). On the chromosome, no genes were detected in either of the strains that were entirely absent in another strain.

Most changes were single base substitutions (29) of which 8 were synonymous and 21 were non-synonymous and most of which were transitions while only 5 were transversions. The remaining mutations were deletions (12) and insertions (7), 14 of which caused frameshifts in the corresponding genes. With exceptions of *sclav_1236a/1256b*, which is one open reading frame, and *sclav_4749*, only one change occurred per gene.

Gene products affected by mutations have a wide variety of (predicted) functions, including membrane proteins, transcriptional regulators, and proteases, all of which might help to understand the observed differences in carbon utilisation phenotypes of SC2 and SC6. Nine genes were found to differ in SC2, but not in SC6, compared to the ATCC 27064 sequence, three of which encode putative transcription regulators (*sclav_2126*, *sclav_2132* and *sclav_3187*). The first two regulator-encoding genes, *sclav_2126*, *sclav_2132*, are in proximity to one another on the chromosome and *sclav_2127* to *sclav_2131* have also been predicted to encode transcriptional regulators. While the one-base insertion in *sclav_2126* results in the reading frame being extended by 95 bases, which translated to 31 amino acids, there is no frameshift in *sclav_2131* (three-base insertion). There is merely an additional glutamic acid at position 197 in the amino acid sequence.

The third regulator gene, *sclav_3187*, codes for a helix-turn-helix (HTH) motif-containing LacI-family type regulator. The four-base deletion in this sequence causes a complete frameshift and a premature stop codon after 330 amino acids when the original protein consists of 340 residues. As all three gene sequences in SC6 are identical to the wildtype, this indicates that their gene products might have physiologically essential roles.

Table 4-3: Specific mutations on the SC2 (pink), SC6 (green) or both (white) chromosome compared to the ATCC 27064 sequence. Column 1 specifies synonymous (S) or non-synonymous (N) changes, column 2 specifies the type of change: substitution (sub), insertion (ins) or deletion (del). Frameshifts are shown in column 3 (Y=yes) or not (N=no). Inserted or deleted bases are shown in column 4. The last two columns show original and altered codons (changed amino acids).

Gene	Strain	1	2	3	4	Original Codon/s (AA)	Altered Codon/s (AA)
sclav_0130	SC6	N	sub			**C**GG (R278)	**T**GG (Y278)
sclav_0190	SC6		ins	Y	**CGCC**	CTC (L477) - GCC (A478) - CAC (H479)	CTC (L477) - GCC (A478) - CGC (R497)
sclav_0198	SC6	S	sub			GT**G** (V189)	GT**A** (V189)
sclav_0208	SC6		ins	N	**CTC**	CTC (L47) - GTG (V48)	CTC (L47) - CTC (L48) - GTG (V49)
sclav_0371	SC2		del	Y	**GG**	CTG (L384)- **GG**A (G385) - CGG R(385)	CTG (L384) - ACG (T385) - GCT (A386)
sclav_0475	SC6		del	Y	**CG**	GCC (A772) - G**CG** (773) - CCG (P774)	GCC (A772) - GCC (A773) - GAC (D774)
sclav_0533	SC2		del	N	**TCG**	G**TC** (167) - **G**GG (G168) - AGC (S169)	GGG (G167) - AGC (S168)
sclav_0627	SC6	N	sub			CA**T** (H274)	CA**G** (Q274)
sclav_0735	SC6	N	sub			G**CC** (A99)	G**AA** (E99)
sclav_0990	SC6	N	sub			**A**CG (T15)	**G**CG (A15)
sclav_1039	SC6	S	sub			CC**C** (P191)	CC**T** (P191)
sclav_1177	SC6	N	sub			TA**C** (Y108)	T**G**C (C108)
sclav_1236a	SC2, SC6		ins	N		[11]Inserted bases	
sclav_1236b	SC2, SC6	S	sub			CC**T** (P5)	CC**C** (P5)
		N	sub			AT**G** (M8)	A**C**G (T8)
sclav_1310	SC6		del	Y	**C**	TCC (S5) - CCC (P6) - **C**AC (H7)	TCC (S5) - CCC (P6) - ACG (T7)
sclav_1366	SC6	S	sub			GT**G** (V247)	G**C**G (A247)
sclav_1428	SC6	S	sub			GT**A** (V116)	GT**G** (V116)
sclav_1458	SC6	S	sub			GC**G** (A54)	GC**C** (A54)
sclav_1550	SC6	N	sub			**A**TC (I184)	**G**TC (V184)
sclav_1642	SC 6	N	sub			TA**C** (Y231)	T**G**C (C231)
sclav_1974	SC6	N	sub			TA**C** (Y399)	T**G**C (C399)
sclav_1980	SC6	N	sub			C**G**C (R52)	C**A**C (H52)
sclav_2126	SC2		ins	Y	**G**	CTG (L309) - GAC (D310) - TCG (S311)	GCT (A309) - **G**GA (G310) - CTC (L311)
sclav_2132	SC2		ins	N	**GGA**	GAG (E196) - GAC (D197) - GAC (D198)	GAG (E196) - GA**G** (E197) - **GA**C (D198) - GAC (D199)
sclav_2152	SC6	N	sub			C**C**G (P48)	**C**AG (Q48)
sclav_2480	SC2		del	Y	**T**	TCT (R11) - **T**GG (T12) - TGG (T13)	TCT (R11) - GGT (P12) - GGT (P12)
sclav_3187	SC2		del	Y	**CCGA**	TCC (S295) - **TCC** (S296) - **GA**C (D297)	TCC (S295) - TCC (S296)
sclav_3236	SC6	N	sub			AT**G** (M1)	AT**A** (I1)
sclav_3343	SC2	N	sub			G**T**A (V34)	G**C**A (A34)
sclav_3618	SC6		ins	Y	**G**	GG**T** (G78) - CTC (L79) - GAT (D80)	GGG (G79) - TCT (S79) - CGA (R80)
sclav_3702	SC6	N	sub			**T**AT (Y256)	**C**AT (H256)
sclav_3737	SC6		del	Y	**G**	ACG (T46) - G**G**C (G47)	ACG (T46) - GCC (A47)
sclav_3951	SC6	N	sub			A**CC** (W39)	A**GT** (S39)
sclav_4138	SC6	N	sub			TA**C** (Y260)	T**G**C (C260)
sclav_4333	SC6	N	sub			**T**AC (Y249)	**C**AC (H249)
sclav_4515	SC6	S	sub			CG**G** (R595)	CG**T** (R595)
sclav_4536	SC6		del	Y	**C**	CCC (P28) - C**C**G (P29) - GGG (G30)	CCC(P28) - CGG (R29) - GGG (G30)
sclav_4714	SC6	N	sub			**T**AC (Y89)	**C**AC (H89)
sclav_4749	SC2		ins	Y	**C**	GTC (V2024) - GCC (A2025) - TAC (Y2026)	GTC (V2024) - G**CC** (A2025) - CTA (L2026)
	SC6	N	sub			**T**AC (Y2026)	**C**AC (H2026)
sclav_4782	SC6	S	sub			GT**A** (V10)	GT**G** (V10)
sclav_4924	SC2		del	Y	**G**	GCG (A517) - G**C**C (G518) - GAG (E519)	GCG (A217) - GCG (A218) - AGG (R219)
sclav_4978	SC6	N	sub			TA**C** (Y795)	T**C**C (S795)
sclav_5122	SC2		del	Y	**C**	CCC (P150) - **C**AG (Q151) - GCG (A152)	CCC (P150) - AGG (R151) - CGT (R152)
sclav_5196	SC6		sub			**T**AC (Y712)	**C**AC (H712)

[11] GCCCCACAGCCCGGACCGCACCGGCCCGCACCCACCGGGAACGGCTCCGGCCCCGCGAGCGACCCCGCGAACGGC

sclav_5218	SC6		del	N	**CAG**	**CAG** (D578)

Table 4-4: List of genes for which sequences differ in SC2 (pink), SC6 (green) or both (white) from the ATCC 27064 sequence. The predicted gene product function and the affected strain are specified. Mutations were detected by alignment of ATCC 27064, SC2 and SC6 sequences.

Gene	Predicted Gene Product Function	Strain
sclav_0130	XRE family transcriptional regulator	SC6
sclav_0190	Mn or Mg-dep protein phosphatase	SC6
sclav_0198	integral membrane protein	SC6
sclav_0208	isopenicillin N synthase family oxygenase, 2-oxoglutarate-dependent ethylene/succinate-forming enzyme	SC6
sclav_0371	carboxypeptidase G2 / metallopeptidase	SC2
sclav_0475	goadsporin biosynthetic protein	SC6
sclav_0533	dethiobiotin synthetase, BioD	SC2
sclav_0627	glycosyl hydrolase	SC6
sclav_0735	hypothetical protein	SC6
sclav_0990	glycosyl transferase	SC6
sclav_1039	secreted subtilisin-like protease	SC6
sclav_1177	zinc finger SWIM domain protein	SC6
sclav_1236a, sclav_1236b (one ORF)	ADP- ribosylglycohydrolase	SC2, SC6
sclav_1310	methyltransferase	SC6
sclav_1366	Cytochrome C heme-binding subunit, QcrC	SC6
sclav_1428	arsenate reductase, ArsC	SC6
sclav_1458	hypothetical protein	SC6
sclav_1550	hypothetical protein	SC6
sclav_1642	putative DNA-binding protein	SC6
sclav_1974	glycosyl hydrolase	SC6
sclav_1980	Putative polar amino acid ABC transporter (glutamine) permease	SC6
sclav_2126	transcriptional regulator	SC2
sclav_2132	transcriptional regulator	SC2
sclav_2152	hypothetical protein	SC6
sclav_2480	hypothetical protein	SC2
sclav_3187	LacI-type transcription regulator	SC2
sclav_3236	FAD dependent oxidoreductase	SC6
sclav_3343	hypothetical protein	SC2
sclav_3618	MaoC_dehydratas domain-containing protein	SC6
sclav_3702	Inositol-5-monophosphate dehydrogenase	SC6
sclav_3737	serine/threonine protein kinase	SC6
sclav_3951	fructose-1,6-bisphosphatase II, GlpX	SC6
sclav_4138	N, O-diacetyl muramidase, lysozyme	SC6
sclav_4333	integral membrane protein	SC6
sclav_4515	ATP-dependent DNA helicase, RecG	SC6
sclav_4536,	PII-adenylyltransferase, GlnD	SC6
sclav_4714	hypothetical protein	SC6
sclav_4749	Non-ribosimal peptide synthetase, CdaPS2	SC2, SC6
sclav_4782	Deoxyuridine 5'-triphosphate nucleotidohydrolase	SC6
sclav_4924	hypothetical protein	SC2
sclav_4978	subtilisin-like protease	SC6
sclav_5122	putative ABC transporter substrate-binding protein	SC2
sclav_5196	glycosyl hydrolase	SC6
sclav_5218	transcription accessory protein	SC6

The most interesting and potentially relevant mutation in SC6 is the single base deletion in *sclav_4536* that encodes the adenylyltransferase GlnD, and that is located downstream of *amtB* and *glnK*. In *S. coelicolor*, *amtB-glnK-glnD* form a transcriptional unit that is under the control of GlnR and PhoP, global regulators of nitrogen and phosphate metabolism, respectively (Fink *et al.*, 2002; Rodríguez-García *et al.*, 2009; Wang *et al.*, 2012). The mutated gene in SC6 encodes a truncated version of GlnD consisting of 405 amino acids. The original protein is comprised of 1015 amino acids. The role of GlnD in nitrogen metabolism has been studied in *S. coelicolor*, where it carries out reversible and irreversible post-translational modification of GlnK, a PII signalling protein, in response to changes nitrogen conditions (Hesketh *et al.*, 2002). The *amtB-glnK-glnD* operon helps to control the glutamine pool within cells and requires interplay with central carbon metabolism (to provide 2-oxogluterate as an animation substrate) and central nitrogen metabolism, through control of ammonium uptake. GlnR, for which the binding sequence gTnAc – n_6 – GaAAc – n_6 has been identified, binds upstream of *amtB* (Fink *et al.*, 2002; Tiffert *et al.*, 2008).

The three GlnR binding boxes identified in *S. coelicolor* are also present upstream of *amtB* in the *S. clavuligerus* strains: TTCAT (a3)– n_6 – GTAAC (b3) – n_{15} – GTCAC (a2) – n_6 – GAAAC (b2) – n_6 – CGGCA (a1) – n_6 – GAAAC (b1) – n_6, suggesting the operon might also under the regulatory control of GlnR in *S. clavuligerus* (Li *et al.*, 2018). Despite neither GlnD nor GlnK being necessary to ensure glutamine synthetase (GS) adenylation in response to changes in nitrogen conditions in *S. coelicolor*, these data suggest that the *amtB-glnK-glnD* operon is also expressed in a GlnR-regulated manner in response to nitrogen availability in all three strains. In SC6, however, GlnD might not be capable of (de-)adenylating GlnK due to it being truncated.

The RNAseq analysis shows that while *amtB* expression increased weakly in SC6 ($\log_2$ fold change=0.536, p-value=0.510), during CA production (Table 4-5), expression of *glnK* ($\log_2$ fold change=-1.322, p-value=0.201) and *glnD* ($\log_2$ fold change=-1.105, p-value=0.170), was downregulated. Although none of these changes were significant, the $\log_2$ fold changes for *glnK* and *glnD* were <-1, indicating that the downregulation of these genes was > than 2-fold at 65 hours compared to 52 hours.

Table 4-5: RNA-seq analysis of *amtB-glnK-glnD*. Significant $\log_2$ fold changes in gene expression at 65 hours (when CA production is occurring) compared to 52 hours in GSK production media in SC6 are shown. P-values (*t*-test) showing the statistical significance of the $\log_2$ fold changes are listed in the last column.

Gene	Log$_2$ fold change	p-value
amtB	0.536	0.510
glnK	-1.322	0.201
glnD	-1.105	0.170

Nonetheless, the lack of confidence of significance due to high p-values does not allow to draw conclusions about the biological relevance of the observed changes in gene expressions in this operon.

However, in a previous study, *amtB* and *glnK* were downregulated while *glnD* expression was unchanged in a *ccaR* deletion mutant (Álvarez-Álvarez *et al.*, 2014). This suggests that i) *amtB* and *glnK* (but not *glnD*) might form an operon in *S. clavuligerus* or ii) both genes are positively influenced by CcaR. Yet only *amtB* is weakly upregulated with a similar $\log_2$ fold change as *ccaR* at 65 hours compared to the later time point in SC6, while expression of *glnK* and *glnD* was downregulated. This suggests the frameshift that occurred in *glnD* may have a polar effect on *glnK* expression in SC6, by either directly affecting expression of *glnK* due to the absence of functional GlnD or by indirectly altering regulatory mechanisms of *glnK* expression. Alternatively, it may indicate the presence of mono- as well as polycistronic transcripts. Given the (non-essential) involvement of GlnD and GlnK in nitrogen metabolism, the altered expression of *glnK-glnD* might contribute to SC6 being unable to grow in NMMP with casamino acids as the sole carbon and nitrogen source (Figure 4-4A). It might, further, impact the regulation of extracellular protease production, which was altered in SC6 (Figure 4-12).

Another mutation that might be involved in the phenotypic differences between SC2 and SC6 observed in relation to proteolytic activity is the non- synonymous base substitution in *sclav_4978*. This mutation results in a radical replacement of the hydrophobic tyrosine at position 795 with the polar amino acid serine.

4.7   SUMMARY

DSM 738, SC2 and SC6 differ in terms of their carbon utilisation and secondary metabolite production. While DSM 738 and SC2 were able to grow in NMMP containing only casamino acids or with added glycerol or glucose, SC6 only grew well in the presence of glucose (see 4.1.1, 4.1.2). This indicates a shift in catabolic capacities had occurred in favour of glucose metabolism, which was potentially due to elevated *glcP* expression that was observed in SC6 (see 4.2). CA was detected in all TSB-grown cultures at all sampled timepoints, with SC6 consistently producing higher yields of CA (see 4.3). Protease activity was also detected under

the same experimental conditions, indicating production of extracellular protease and CA was co-occurring in all cultures (see 4.4.1). Data obtained from examining protease activity on milk agar suggest that extracellular protease activity is repressed by glycerol in DSM 738 and SC2, but not in SC6 (see 4.4.2). This may be linked to the radical replacement in the subtilisin-like protease encoded by *sclav_4978*. A model was presented for glycerol dependent CCR of protease production (see 4.4.3, Figure 4-13). Changes in gene expression occurring in SC6 when *ccaR* and CA biosynthetic genes are upregulated showed that *sclav_4359* and *adpA* are also upregulated, while all genes involved in glycerol and glucose uptake were downregulated, with exceptions of *gylR* and *glpD* (Table 4-2). The phenotypic differences between SC2 and SC6 were supported with data obtained from comparing their chromosomal sequences and identifying mutations in either of the strains compared to the WT sequence (Table 4-3, Table 4-4). Most changes occurred in SC6, of which one is particularly interesting: *glnD* that encodes an adenylyltransferase is truncated in this strain due to a base deletion causing a premature stop codon. The role of GlnD in nitrogen metabolism in *S. clavuligerus* has not been investigated, yet in *S. coelicolor* a complete deletion of this gene was found to not be detrimental (Hesketh *et al.*, 2002). Nonetheless, the major phenotypic and physiological differences of SC2 and SC6 might be, at least in part, explained by the absence of functional GlnD.

5 Heterologous Expression of *glcP* and *glk* from *S. clavuligerus* and *S. coelicolor* in *Streptomyces* Strains

In this chapter, the effects of heterologous expression of streptomycete *glcP* and *glk* genes in SC2 and SC6 were examined with regards to growth, development, and CA production. The link between carbon utilisation, protease production and CA biosynthesis that might exist in *S. clavuligerus* was further investigated in these strains. Given that most studies of the *Streptomyces* glucose uptake system focus on *S. coelicolor*, the *glcP* and *glk* genes from *S. clavuligerus* and *S. coelicolor*, *sclav_4529 and sclav_1340*, *sco5578* and *sco2126*, respectively, were chosen for heterologous expression in SC2 and SC6. As mentioned in the previous chapter, *sclav_4529* nucleotide sequences are identical across DSM 738, SC2 and SC6. This is also the case for *sclav_1340.* Therefore, genes amplified from DSM 738 gDNA were used in the cloning process as representative *S. clavuligerus* genes. SC2 and SC6 heterologously expressing *sclav_4529*, *sclav_1340*, *sco5578* or *sco2126* are henceforth referred to as constructed strains, designated ASB1, ASB2, ASB3 and ASB4, respectively.

In addition to characterising *S. clavuligerus* constructed strains, antibiotic production in GlcP-deficient *S. coelicolor* strains, BAP20 (*ΔglcP1::hyg, ΔglcP2:: aacC4*, van Wezel *et al.*, 2005), complimented with *sclav_4529* or *sco5578* or heterologously expressing *sclav_1340* or *sco2126* was examined in the presence of a range of carbon sources. Pigment production by *S. coelicolor* is indicative of the switch to secondary metabolite production having occurred and, unlike CA production by *S. clavuligerus*, pigment production on agar is visible to the naked eye. Pigment production by these strains was, thus, assessed to identify possible CCR mechanisms occurring when *sclav_4529* or *sco5578* or either of the *glk* genes were heterologously expressed in an *S. coelicolor* background. This would allow assessment of the effect that Glk from *S. clavuligerus* might have on antibiotic production in the model organism.

5.1 Construction of SC2, SC6 and *S. coelicolor* Strains For Expression of *glcP* and *glk*

Heterologous expression of DNA in *Streptomyces* is commonly achieved using integrating vectors that contain a phage integrase gene, *int,* and a phage attachment site, *attP* (Kieser *et al.*, 2000). The pIJ6902 vector (see Figure 5-1A for plasmid map) and into which the *glcP* and

glk genes were cloned, contains a φC31 integrase gene from the temperate *Streptomyces* φC31 phage. The φC31 integrase is a member of the family of large serine integrases and catalyses the irreversible integration of *attP* onto the bacterial genome. φC31 integrase, thus, catalyses the recombination between *attP* and *attB*, the bacterial attachment site, and it does so in the absence of any accessory factors (Thorpe *et al.*, 1998). The integration site for φC31 in streptomycetes is located within a pirin homologue, encoded by *sco3798* in *S. coelicolor* (Kuhstoss *et al.*, 1991; Smith, 2015). Based on sequence homology, this gene is *sclav_3010* in *S. clavuligerus*. Although derivatives of the φC31-derived pSET152 plasmid have been used for genetic manipulation of *S. clavuligerus*, no study demonstrating the use of pIJ6902 in *S. clavuligerus* exists to date to the author's knowledge (Trepanier *et al.*, 2002; Li *et al.*, 2006; Jnawali *et al.*, 2011; Kizildogan *et al.*, 2017).

The genes of interest, *glcP* and *glk*, were cloned under the control of the *tipA* promoter, *ptipA*. This is a thiostrepton-inducible promoter, from which gene expression can be increased up to 200-fold in the presence of thiostrepton (Murakami *et al.*, 1989). Thiostrepton is a peptide antibiotic capable of inhibiting protein biosynthesis by binding to 23S rRNA within the 50S ribosomal subunit and inhibiting the function of elongation factor G (EF-G) (Thompson *et al.*, 1979; Rodnina *et al.*, 1999). EF-G is a GTPase that functions as a molecular motor driving translocation during elongation of protein biosynthesis (Rodnina *et al.*, 1997). Inherently thiostrepton-resistant *Streptomyces*, such as *S. azureus*, produce a methylase that methylates a thiostrepton binding site on the 23S rRNA, thus, preventing its binding to the ribosomal subunit (Thompson *et al.*, 1982). Although gene expression from *ptipA* is increased by thiostrepton, the promoter is leaky due to the nature of its regulation and the function of its gene products (Holmes *et al.*, 1993; Chiu *et al.*, 1999). Genetic complementation in *S. coelicolor* has been demonstrated in the absence of thiostrepton (Hoskisson *et al.*, 2015). Further, elevated external osmolarity and the resulting increase in negative DNA supercoiling have been shown to affect *pTipA* activity in *S. lividans* (Ali *et al.*, 2002).
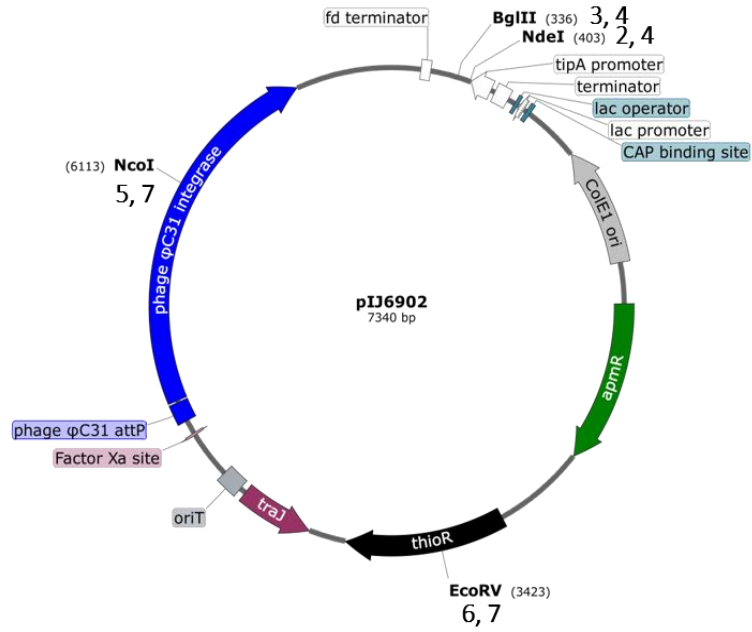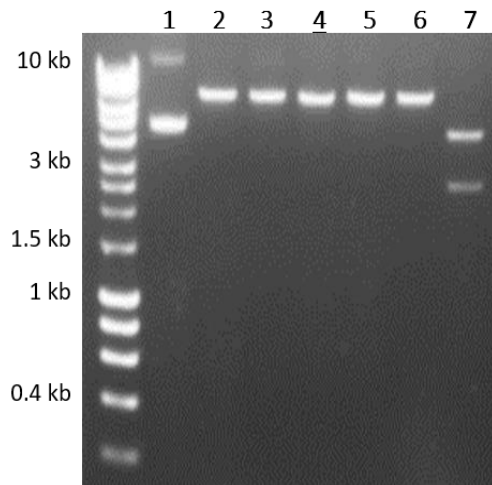
Figure 5-1: (A) Plasmid map of the integrating vector pIJ6902 and (B) agarose gel (1.0%) showing bands that resulted from restriction digest of the plasmid. Highlighted features on the map are φC31 integrase gene (blue), the apramycin (green) and thiostrepton (black) resistance conferring genes, as well as the *tipA* promoter (*ptipA*, white). The agarose gel shows bands of undigested plasmid (1), bands from plasmid digest with *Nde*I (2), *Bgl*II (3), the combination of *Nde*I and *Bgl*II (4), *Nco*I (5), *EcoR*V (6) and the combination of *Nco*I and *EcoR*V (7). The numbers adjacent to the restriction enzymes on the plasmid map indicate the lanes of the gel that show bands corresponding to DNA fragments that result from digest using the indicated enzyme(s).

The addition of sucrose to cultivation media lead to a 10-fold increase in *pTipA* activity *in vivo* (Ali *et al.*, 2002). The *tipA* gene encodes two gene products, TipAL, both of which covalently bind thiostrepton (Chiu *et al.*, 1996). The thiostrepton-TipAL complex acts as autogenous transcriptional activator of *tipA* (Holmes *et al.*, 1993). In some streptomycetes, among which are *S. coelicolor* and *S. clavuligerus*, thiostrepton induces the expression of other thiostrepton-inducible proteins (Tips), some of which have been characterised as resistance-conferring proteins (Vasant Kumar *et al.*, 1994). In *S. clavuligerus*, a 15-kDa Tip has been identified, though not further characterised. Moreover, two genes encoding other members of the TipAS protein family are present on the *S. clavuligerus* genome, *sclav_3133* is located on the chromosome and encodes a 239 amino acid-long protein and *sclav_p1179* is located on pSCL4 and encodes 272 amino acid-long protein. Calculated molecular weights for SCLAV_3133 and SCLAV_p1179 are 28.24 kDa and 29.27 kDa (calculated using the protein molecular weight calculator available at https://www.bioinformatics.org/sms/prot_mw.html), respectively, which is larger than the previously identified Tip of 15 kDa. Both *sclav_3133* and *sclav_p1179* are present on the SC2 and SC6 chromosomes, as well as both copies of pSCL4 (unpublished sequence data from Dr John T Munnoch). This suggests an uncharacterised, thiostrepton-dependent regulon may exist in *S. clavuligerus*.

Antibiotic resistance markers present on pIJ6902, *aac(3)IV* and *tsr*, which encode an aminoglycoside acetyltransferase and methylase, respectively, confer resistance to apramycin (*apmR*) and thiostrepton (*thioR*) (Huang *et al.* 2005). Apramycin and thiostrepton were, thus, used to select for successful integration events of pIJ6902 into the respective genomes. Restriction enzymes *Nde*I, *Bgl*II, *Nco*I and *EcoR*V for which recognition sites are shown on the pIJ6902 map, cut either within the multiple cloning site or downstream of *ptipA* as is the case for *Nde*I, *Bgl*II. These enzymes were used to facilitate unidirectional insertion of *glcP* or *glk*. *Nco*I and *EcoR*V cut within the integrase and thiostrepton resistance-conferring gene, respectively. Restriction digests using these enzymes was, therefore, used to confirm the plasmid map (Figure 5-1B). Uncut plasmid migrating through the agarose gel produced two distinct bands that correspond to its nicked and supercoiled configuration. Single digests using either of the enzymes resulted in 7.3 kb size fragments. The *Nde*I and *Bgl*II restriction sites are 67 bp apart. Therefore, the double digest of pIJ6902 using *Nde*I and *Bgl*II yielded

two fragments, of which the smaller one was too small to visualise on a 1.0% agarose gel. Carrying out a double digest with *Nco*I and *EcoR*V, however, resulted in two fragments sizes 2.7 kb and 4.6 kb, for which both bands of the expected sizes are visible on the gel.

5.1.1 CONSTRUCTION OF PIJ6902 DERIVATIVES CONTAINING *glcP* OR *glk*

Genes chosen for heterologous expression in *S. clavuligerus* and *S. coelicolor* strains were *glcP* and *glk* from *S. clavuligerus*, *S. coelicolor*, *S. venezuelae* and *S. griseus*. pIJ6902 derivatives with these genes cloned downstream of *ptipA*, were constructed. The cloning process consisted of four steps (Figure 5-2), the first of which was the amplification of *glcP* and *glk* from *S. clavuligerus* (DSM 738), *S. coelicolor* (M145), *S. venezuelae* (ATCC 10712) and *S. griseus* (NBRC 13350) gDNA. The oligonucleotide sequences for gene amplification were designed to include restriction enzyme recognition sites for *Nde*I and *Bgl*II in the forward and reverse primer, respectively, replacing the *glcP* start codon, 5`-GTG-3`, with 5`-ATG-3` intrinsic to the *Nde*I recognition sequence, 5`-CAT**ATG**-3` (Figure 5-2A). PCR products of the correct sizes, 1.5 kilobases (kb) for *glcP* and 1.0 kb for *glk*, were obtained from *S. clavuligerus*, *S. coelicolor* and *S. griseus* gDNA (Figure 5-3), although only a faint band is visible for *sgr_1900*. PCR products amplified from *S. venezuelae* gDNA were 0.8 kb and 0.7 kb and, thus, smaller than excepted. No products were obtained from non-template controls (NTC).

DNA was subsequently purified from agarose gel slices excised at the expected band sizes. The DNA polymerase had created blunt-ended DNA fragments, which were modified using *Nde*I and *Bgl*II to yield fragments compatible for ligation with pIJ6902 digested with the same enzymes (Figure 5-2B). Ligation of *sclav_4529* (*glcP*), *sclav_1340* (*glk*), *sco5578* (*glcP*), *sco2126* (*glk*), *sven_5273* (*glcP*), *sven_1786* (*glk*), *sgr_5377* (*glk*) with pIJ6902 resulted in construction of the following plasmids: pIJ6902-ASB1, pIJ6902-ASB2, pIJ6902-ASB3, pIJ6902-ASB4, pIJ6902-ASB5, pIJ6902-ASB6 and pIJ6902-ASB8, respectively (Figure 5-2C, D). No transformants were obtained using the ligation mixture containing *sgr_1900* and pIJ6902. Therefore, pIJ6902-ASB7 was later synthesised by IDT. The presence of the inserts on the constructed plasmids was confirmed by restriction digest using *Nde*I and *Bgl*II and further digests were carried out using *Nco*I and *EcoR*V (Figure 5-4 for pIJ6902-ASB1 and pIJ6902-ASB2, Figure 5-5 for pIJ6902-ASB3 and pIJ6902-ASB4).

154

A

*Nde*I

GTAC**CATATG**

Amplification of *glcP* and *glk*

Forward Primer

5`-...AGGAGTTGAT**GTG**AGCAGCAGGACTCCCCCGTCACCAC.................GGGCAAGCCGCTGGAGGAGATGGGCTGAGCCCCGGCCC...-3`
3`-...TCCTCAACTACACTCGTCGTCCTGAGGGGGCAGTGGTG.................CCCGTTCGGCGACCTCCTCTACCCG**ACT**CGGGGCCGGG...-5`

Reverse Primer

TCTAGACATG
*Bgl*II

B

Digest of PCR products with *Nde*I and *Bgl*II

5`-**TATG**AGCAGCAGGACTCCCCC.................CGCTGGAGGAGATGGGC**TGAA**-3`
3`-**AC**TCGTCGTCCTGAGGGGG.................GCGACCTCCTCTACCCG**ACTTCTAG**-5`

C

Target Vector Linearisation

fd terminator

<BgIII> (7273)

<NdeI> (0)

tipA promoter
terminator
lac operator
lac promoter
CAP binding site

lambda t0 terminator

phage φC31 integrase

ori

pIJ6902
7273 bp

apmR

phage φC31 attP
Factor Xa site

oriT

traJ

thioR

D

Ligation

M13 fwd

fd terminator

Reverse Primer (336 .. 361)

glcP(SCLAV)

phage φC31 integrase

Forward Primer (1751 .. 1775)
tipA promoter
M13 rev

pIJ6902-ASB

ori
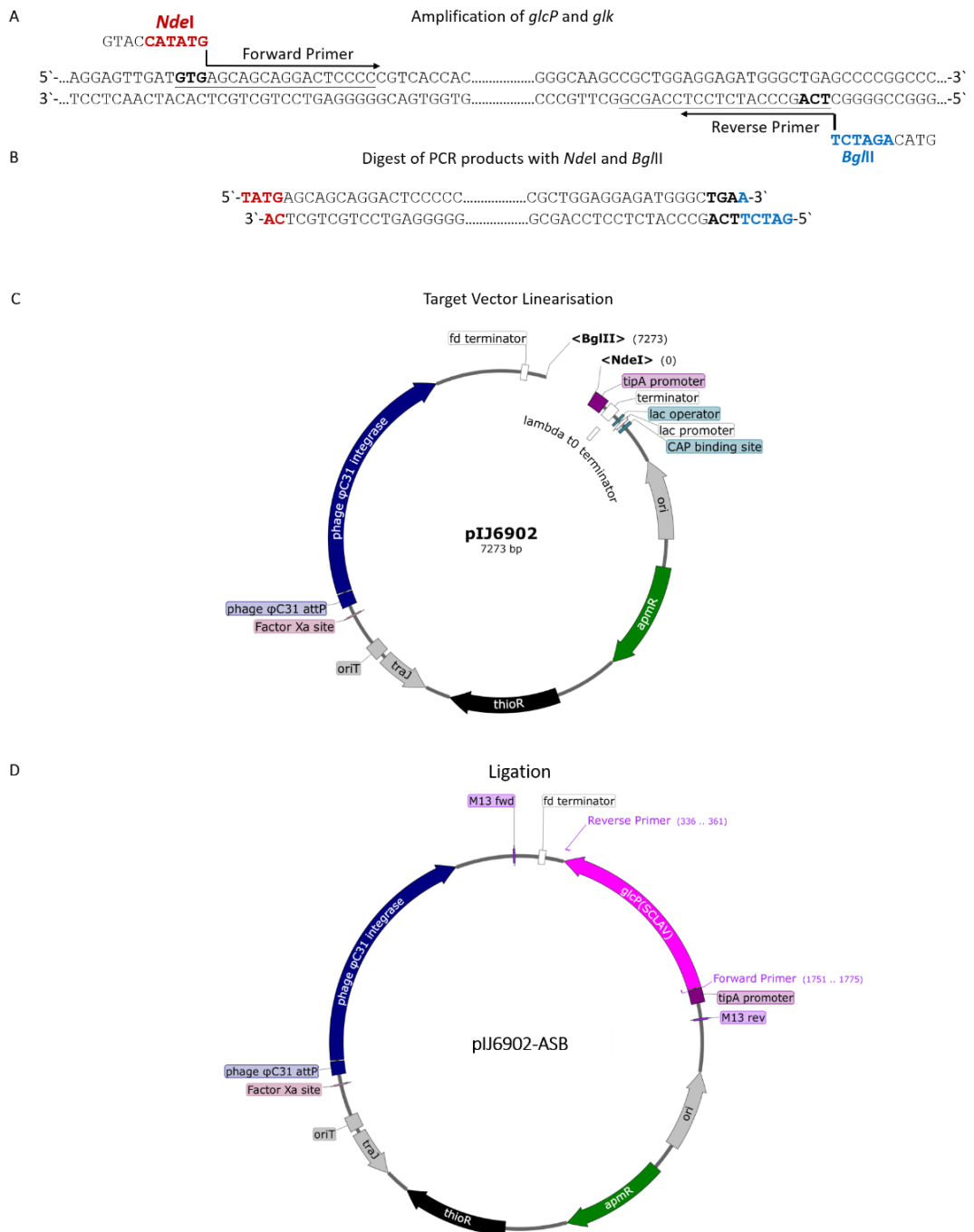
phage φC31 attP
Factor Xa site

oriT

traJ

apmR

thioR

Figure 5-2: Schematic of the cloning process to construct pIJ6902 plasmids containing *glcP* and *glk* genes. (A) The forward and reverse primers (black arrows) to amplify *glcP* and *glk* were designed to include *Nde*I (red) or *Bgl*II (blue) restriction sites, respectively, with the 5`-ATG-3` in the *Nde*I site replacing the start codon5`-GTG-3` of the original gene sequence of *glcP*. (B) The blunt-ended PCR products were modified with *Nde*I and *Bgl*II to yield sticky ends with 5` overhangs. (C) Restriction digest of pIJ6902 resulted in complementary sticky ends on the target vector suitable for ligation with the inserts from the previous step. (D) Inserts and vector were ligated, the ends produced by each enzyme directing the orientation of the insert.
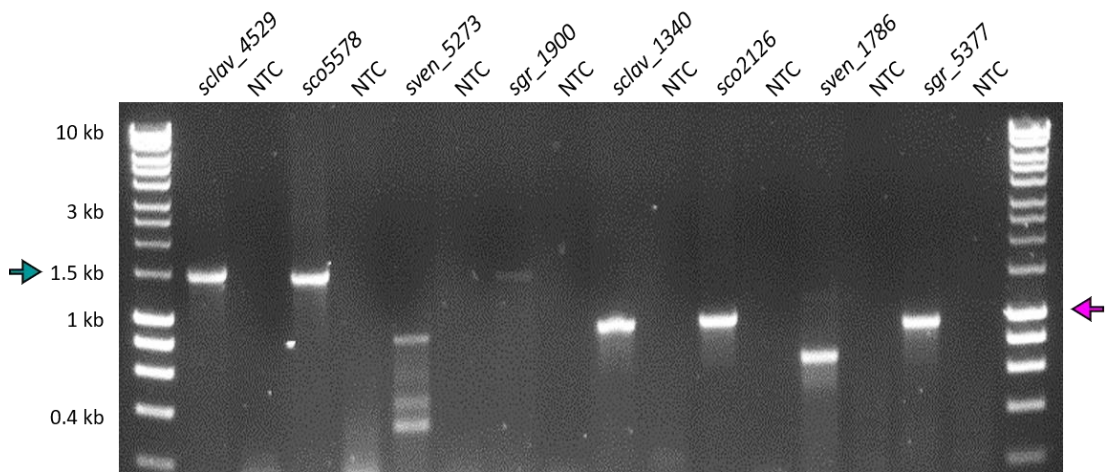
Figure 5-3: Agarose gel (1.0%) showing bands that resulted from PCR amplification of *glcP* and *glk* genes. The genes were amplified from *S. clavuligerus* DSM 738, *S. coelicolor* M145, *S. venezuelae* ATCC 10712 and *S. griseus* NBRC 13350 gDNA. Bands of 1.5 kb are visible for *sclav_4529*, *sco5578* and *sgr_1900* (turquoise arrow), 1.0 kb can be seen for *sclav_1340*, *sco2126*, *sgr_5377* (pink arrow). No bands of the correct sizes can be seen for *S. venezuelae* PCR products. No DNA was amplified in the absence of gDNA (non-template controls, NTC).
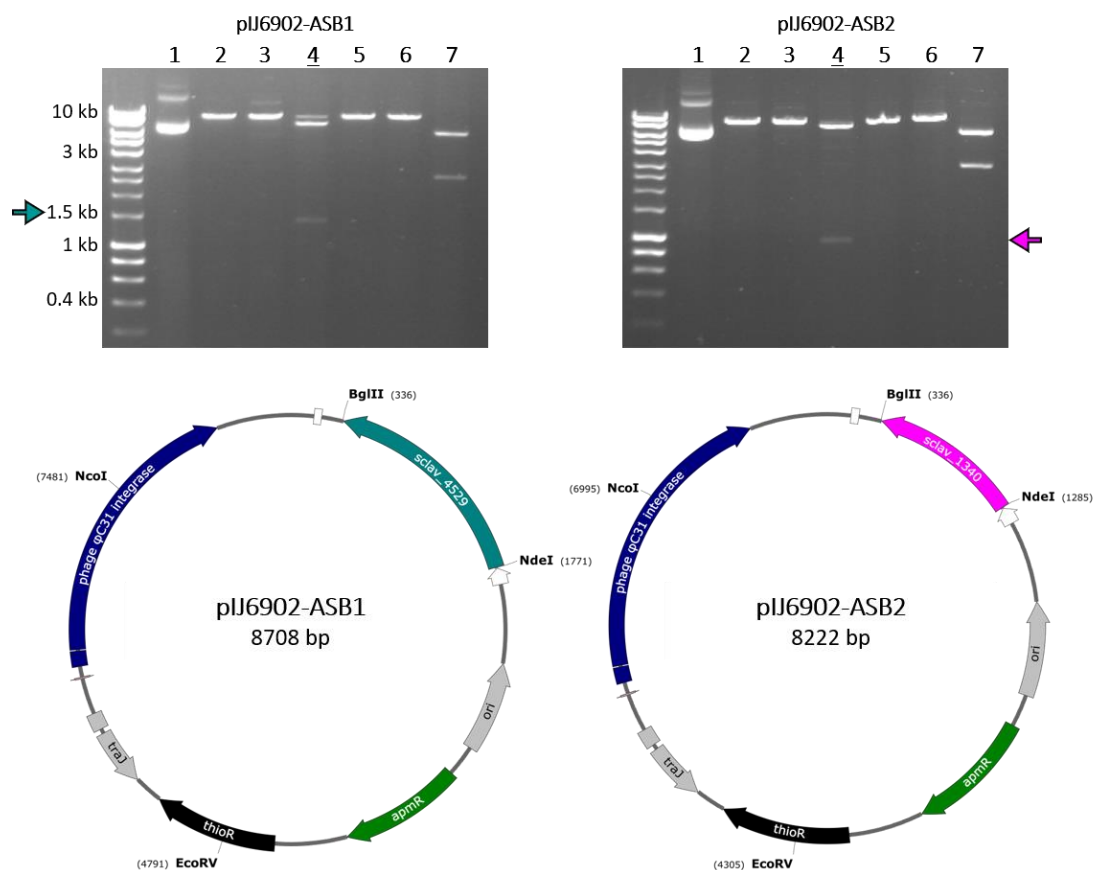
Figure 5-4: Agarose gels (1.0%) showing bands resulting from restriction digests of pIJ6902-ASB1 and pIJ6902-ASB2 and corresponding *in silico* cloned plasmid maps. Undigested plasmid (lane 1) resulted in three bands, the top, middle, and bottom bands corresponding to the nicked, linear, and supercoiled configurations of the plasmids. Plasmid DNA digested with *Nde*I, *Bgl*II, both enzymes, *Nco*I, *EcoR*V or both enzymes was run in lanes 2, 3, 4, 5, 6 and 7, respectively. *Nde*I and *Bgl*II sites flank the inserts, which are shown in turquoise (*glcP*) and pink (*glk*) on the maps. The positions of the expected band sizes for sclav_4529 and sclav_1340 are indicated by the turquoise and pink arrow, respectively.
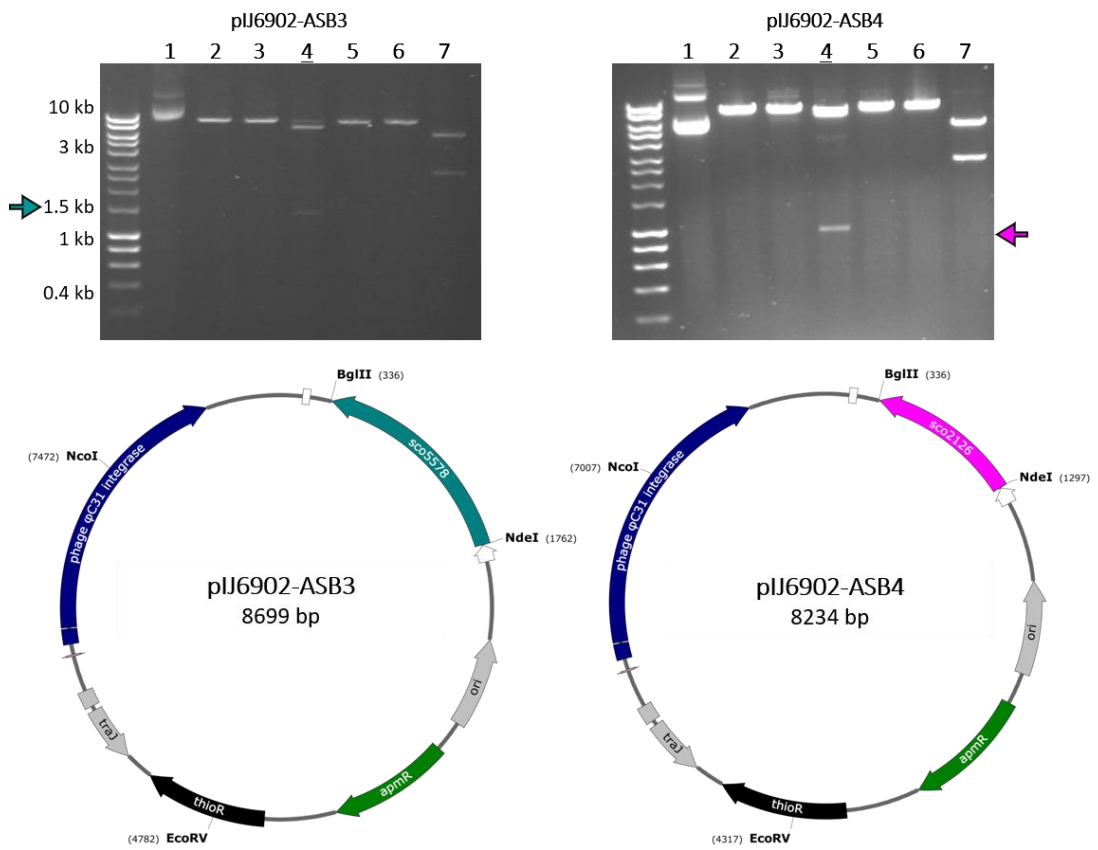
Figure 5-5: Agarose gels (1.0%) showing bands resulting from restriction digests of pIJ6902-ASB3 and pIJ6902-ASB4 and corresponding *in silico* cloned plasmid maps. Bands corresponding to the nicked, linear, and supercoiled configurations of untreated plasmid are shown in lanes 1. The numbers of the lanes correspond to restriction digests of pIJ6902-ASB3 and pIJ6902-ASB4 as before, with either *Nde*I, *Bgl*II or a combination of both enzymes, *Nco*I, *EcoR*V or both enzymes in lanes 2, 3, 4, 5, 6 and 7.

In silico cloned plasmid maps are shown below the respective agarose gels for pIJ6902-ASB1, pIJ6902-ASB2, pIJ6902-ASB3 and pIJ6902-ASB4. The first lanes of each gel show undigested plasmid, with two or three bands clearly visible. The top, middle and bottom bands correspond to the nicked, linear, and supercoiled configuration of the plasmids, respectively.

As expected, single digests of pIJ6902-ASB1, pIJ6902-ASB2, pIJ6902-ASB3 and pIJ6902-ASB4 carried out with *Nde*I, *Bgl*II, *Nco*I or *EcoR*V resulted in single bands of 8.7 kb for pIJ6902-ASB1 and pIJ6902-ASB3 and 8.2 kb for pIJ6902-ASB2 and pIJ6902-ASB4. Fainter bands are also visible in some of these lanes and are presumably due to partial digests or over-digests of the plasmids. Double digests of the plasmids using *Nde*I and *Bgl*II confirmed the presence of the inserts and their sizes with the smaller of the two bands in lane 4 of each gel corresponding to the 1.5 kb *glcP* and 1.0 kb *glk* inserts.

Digests carried out with *Nco*I and *EcoR*V resulted in fragments of 6.0 kb and 2.7 kb for pIJ6902-ASB1 and pIJ6902-ASB3 and fragments of 5.5 kb and 2.7 kb for pIJ6902-ASB2 and pIJ6902-ASB4. Restriction digests carried out with pIJ6902-ASB5, pIJ6902-ASB6 and pIJ6902-ASB8 showed that these plasmids had not been successfully constructed (Figure 5-6). While the bands for pIJ6902-ASB5 resemble those obtained for pIJ6902-ASB1 and pIJ6902-ASB3, *sven_5273* has an intrinsic *Bgl*II restriction site, which should have resulted in an additional, smaller band on the gel. Although the absence of this band may have been due to its size and the fact that DNA was run on a 1.0% agarose gel, sequencing of this insert later revealed that it was not *sven_5273*. The band for the insert in pIJ6902-ASB6 was too small to be *sven_1786* and attempts to amplify the full-length gene failed. As gDNA isolated from *S. venezuelae* served as template in both cases, it is not unlikely that contamination or degradation of the gDNA might have been the cause of unsuccessful amplification. Further attempts to re-isolate *S. venezuelae* gDNA were not carried out due to time constraints.

5.1.2 CONJUGATION OF PIJ6902-ASB DERIVATIVES INTO *S. CLAVULIGERUS* AND *S. COELICOLOR*

Successfully constructed pIJ6902-ASB plasmids (inserts confirmed by sequencing), pIJ6902-ASB1, pIJ6902-ASB2, pIJ6902-ASB3 and pIJ6902-ASB4, were conjugated into *S. clavuligerus* and *S. coelicolor* strains. Conjugation requires a donor methylation-deficient *E. coli* strain that has been transformed with the plasmids to be conjugated (Mazodier *et al.*, 1989).
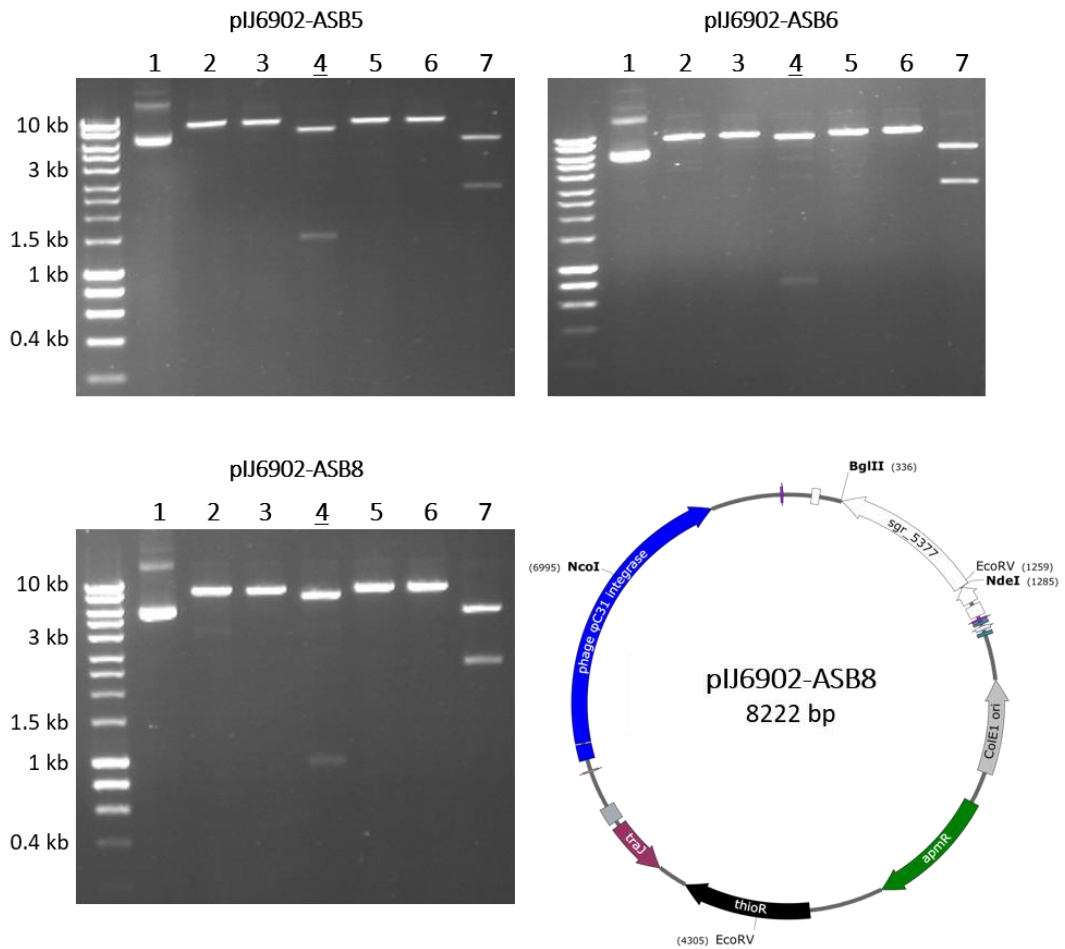
Figure 5-6: Agarose gels (1.0%) showing bands resulting from the same restriction digests carried out for the previous constructs and *in silico* cloned map of pIJ6902-ASB8. Although bands visible on the gel showing pIJ6902-ASB5 bands, the insert on this plasmid, *sven_5273*, has an intrinsic *Bgl*II site, which should have resulted in additional bands in lanes 3 and 4. The size of the insert on pIJ6902-ASB6, *sven_1786*, which is visible in lane 4 is smaller than the expected 1.0 kb. Although the size of the insert on pIJ6902-ASB8 has the expected size of 1.0 kb, *sgr_5377* has an intrinsic *EcoR*V site (see map). However, no additional bands are in lanes 6 and 7 on the gel.

Streptomyces cells that successfully integrated the vector into their genomes were selected for by growing them in the presence of apramycin and thiostrepton at working concentrations of 50 µg/ml each (Figure 5-7A). While the parental strains SC2 and SC6 (and DSM 738) grew on GYM agar, no growth was detected in the presence of the two antibiotics.

Strains containing either the empty pIJ6902 vector, henceforth referred to as SC2-ASB0 and SC6-ASB0, or strains containing pIJ6902-ASB1, pIJ6902-ASB2, pIJ6902-ASB3 or pIJ6902-ASB4, were able to grow when apramycin and thiostrepton were present. Unfortunately, multiple rounds of conjugating pIJ6902-ASB1 into SC6 were unsuccessful as all potential ex-conjugants were sensitive to the antibiotics. Given that several mutations have occurred on the SC6 chromosome compared to SC2, the lack of ex-conjugants might have been caused by alterations within the site of integration on the genome.

The pIJ6902 constructs were also conjugated into *S. coelicolor* strains, the WT strain M145 and a glucose permease deficient strain BAP20, $\Delta glcP1::hyg$, $\Delta glcP2::aacC4$ (van Wezel *et al.*, 2005). A representative set of M145 colonies are shown for these strain (Figure 5-7B). While the parental M145 strain grew well on GYM and even produced some actinorhodin, as indicated by the blue pigment, it was the only strain unable to grow in the presence of apramycin and thiostrepton. Successfully constructed strains for heterologous expression of *glcP* and *glk* genes from the *tipA* promoter are henceforth labelled analogously to the *S. clavuligerus* strains, e.g., M145-ASB1 for M145 containing pIJ6902-ASB1.

5.2 PHENOTYPIC CHARACTERISATION OF CONSTRUCTED STRAINS IN GLUCOSE-RICH MEDIA

Phenotypic characterisation of constructed strains followed a similar structure to the characterisation carried out for DSM 738, SC2 and SC6 in the previous chapter, the focus being on growth and behaviour in the presence of glucose. Constructed strains were compared to the respective empty vector control strains, SC2-ASB0 or SC6-ASB0. All cultures contained sub-MIC concentrations (0.5 µg/ml) of thiostrepton (Murakami *et al.*, 1989). Despite this being a very low concentration of the antibiotic, differences in growth and developmental phenotypes were observed between the parental and empty vector control strains, which are mentioned accordingly. Growth of the constructed strains was recorded in TSB and low amino acid NMMP supplemented with glucose, 0.5% and 1.0% (w/v). Specific growth rates were determined (Table 5-1).
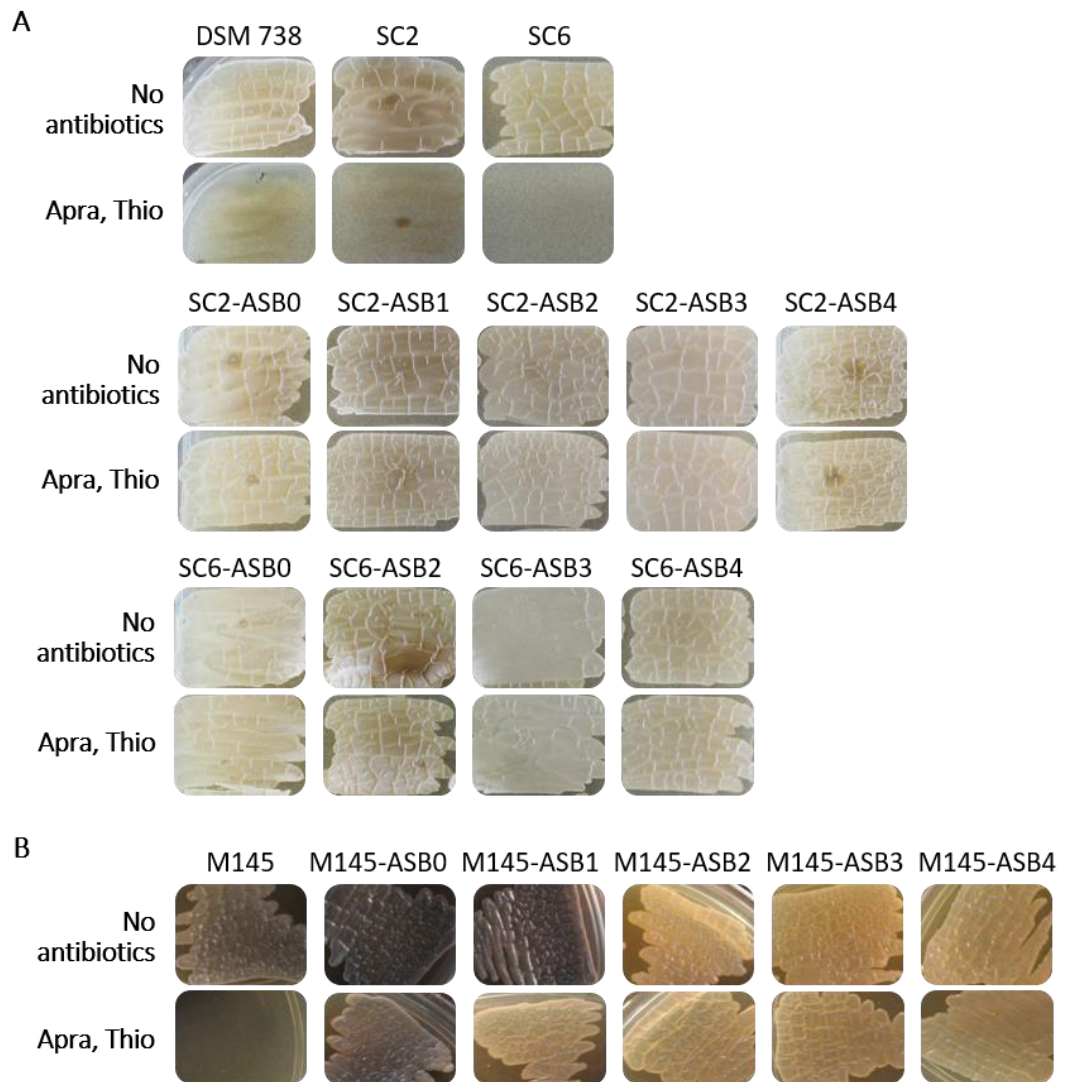
Figure 5-7: Parental strains and ex-conjugants on GYM agar. (A) *S. clavuligerus* and (B) *S. coelicolor* colonies were grown on GYM with and without apramycin (50 μg/ml) and thiostrepton (50 μg/ml). The parental strains DSM 738, SC2, SC6 and M145 only grew absence of the antibiotics. All other strains grew well in the presence of the antibiotics. Empty vector control strains are labelled ASB0, strains carrying pIJ6902-ASB1 (*sclav_4529*), pIJ6902-ASB2 (*sclav_1340*), pIJ6902-ASB3 (*sco5578*) and pIJ6902-ASB4 (*sco2126*) are labelled ASB1, ASB2, ASB3 and ASB4, respectively.

Table 5-1: Specific growth rates μ h^{-1} of SC2 and SC6 cultures: (A) growth rates and standard deviation determined from TSB- and NMMP-(supplemented with 5 g/L or 10 g/L of glucose)grown cultures are shown. (B) Significance levels (p-values, Tukey's HSD test) for growth rate differences between respective ASB0 strains and the strains shown in the top row. Specific growth rates for SC2 cultures in NMMP with glucose (1.0%) could not be determined. NS: not significant; N/A: not applicable (due to a lack of data points).

A

Growth rates μh^{-1}	SC2-ASB0	SC2-ASB1	SC2-ASB2	SC2-ASB3	SC2-ABS4	SC6-ASB0	SC6-ASB2	SC6-ASB3	SC6-ASB4
TSB	0.43±0.04	0.27±0.03	0.25±0.01	0.31±0.01	0.31±0.01	0.27±0.01	0.24±0.01	0.37±0.01	0.27±0.02
Low amino acid NMMP, glucose (5 g/L)	0.11±0.01	0.07±0.02	0.12±0.02	0.13±0.02	0.12±0.01	0.13±0.00	0.06±0.00	0.16±0.02	0.08±0.00
Low amino acid NMMP, glucose (10 g/L)	0.06±0.00	--	--	--	--	0.16±0.03	0.14±0.02	0.16±0.02	0.10±0.01

B

p-values	SC2-ASB1	SC2-ASB2	SC2-ASB3	SC2-ABS4	SC6-ASB2	SC6-ASB3	SC6-ASB4
TSB	0.001	0.001	0.001	0.001	NS	0.001	NS
Low amino acid NMMP, glucose (5 g/L)	0.03	NS	NS	NS	0.001	NS	0.001
Low amino acid NMMP, glucose (10 g/L)	N/A	N/A	N/A	N/A	NS	NS	0.02

163

5.2.2 GROWTH DIFFERENCES IN RICH MEDIA

Specific growth rates calculated for SC2-ASB0 and SC6-ASB0 cultures in TSB containing thiostrepton were 0.43 ± 0.04 h^{-1} and 0.27 ± 0.01 h^{-1}, respectively. These growth rates are comparable for those determined for the parental strains, SC2 (0.38 ± 0.01 h^{-1}) and SC6 (, 0.37 ± 0.19 h^{-1}), growing in TSB in the absence of thiostrepton.

Specific growth rates determined for the other constructed SC2 strains were 0.27 ± 0.03 h^{-1} for SC2-ASB1, 0.25 ± 0.01 h^{-1} for SC2-ASB2, 0.31 ± 0.01 h^{-1} for SC2-ASB3 and 0.31 ± 0.01 h^{-1} for SC2-ASB4 (Table 5-1, Figure 5-8). Across all SC2 cultures, differences in specific growth rates were significant (p-value= 4.1×10^{-5}, one-way ANOVA). In a comparison with the empty vector control, the differences between SC2-ASB0 and SC2-ASB1 (p-value= 0.001, Tukey's HSD), SC2-ASB0 and SC2-ASB2 (p-value= 0.001, Tukey's HSD), SC2-ASB0 and SC2-ASB3 (p-value= 0.001, Tukey's HSD), as well as SC2-ASB0 and SC2-ASB4 (p-value= 0.001, Tukey's HSD) were significant.

Specific growth rates determined for SC6 cultures were 0.24 ± 0.01 $h^{-}1$ for SC2-ASB2, 0.25 ± 0.01 h^{-1} for SC2-ASB2, 0.37 ± 0.01 $h^{-}1$ for SC6-ASB3 and 0.27 ± 0.02 h^{-1} for SC6-ASB4 (Table 5-1, Figure 5-8). Differences in specific growth rates across all SC6 cultures were significant (p-value= 1.3×10^{-5}, one-way ANOVA). Compared to the empty vector control, however, only the difference between SC6-ASB0 and SC6-ASB3 (p-value= 0.001, Tukey's HSD test) was significant. The respective growth curves recorded for the strains growing in TSB can be divided into ones that have a of distinct stationary phase, and ones lacking stationary (Figure 5-9). Growth curves of SC2 constructed strains appeared uniform across all strains, but SC2-ASB1, transitioning from exponential into pre-stationary phase after 12 hours. The SC2-ASB1 growth curve, in contrast, peaks at around the same time before steadily decreasing. Unlike absorbance minima described previously (see TSB-grown SC6, Figure 4-3), this is not correlated with an increase in error bars. SC6-ASB2 and SC6-ASB4 growth curves resemble that of SC2-ASB1 in that absorbance peaks at 14 hours before a steady decline occurs, again not accompanied by large error bars SC6-ASB3, the strain expressing *sco5578*, appears to have grown in a biphasic manner, whereby absorbance peaked at 14 hours, decreased until 24 hours, increased a second time, and plateaued from 27 hours onwards.
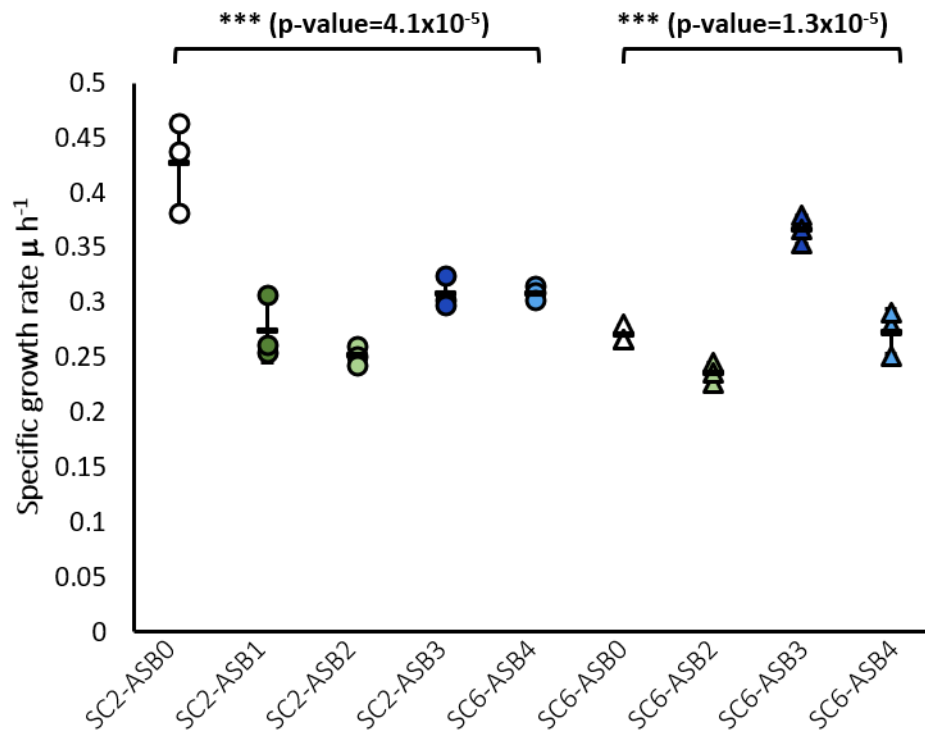
Figure 5-8: Specific growth rates (μ h^{-1}) for SC2 and SC6 constructed strains growing in TSB. Differences across SC2 cultures were significant (***p-value=4.1x10^{-5}, one-way ANOVA), as well as across SC6 cultures (***p-value=1.3x10^{-5}, one-way ANOVA).
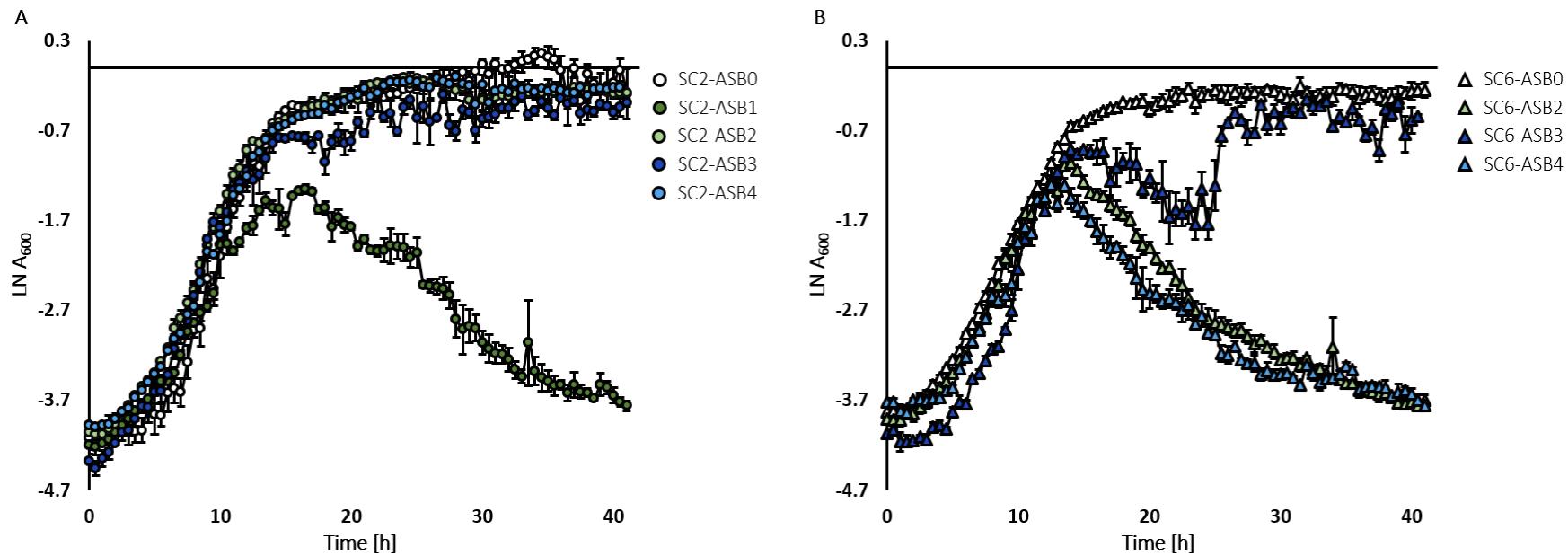
Figure 5-9: Growth curves of SC2 and SC6 cultures. (A) SC2 and (B) SC6 constructed strains were grown in 24-well plates in TSB over 40 hours. The medium contained thiostrepton (0.5 μg/ml) The strains are represented according to the legends. The LN of absorbance (600 nm) is shown. Error bars indicate the standard error for across replicates.

This growth pattern resembles diauxic growth, characterised by nutrient starvation and a resulting metabolic adaptation phase during which a transient growth arrest occurs, termed diauxic lag (Monod, 1942). During the growth arrest, major changes in gene expression occur, some of which have been described for *S. coelicolor*, that affect central and secondary metabolism, as well as differentiation. For instance, during diauxie, the onset of expression of genes involved in stress responses to energy starvation, such as the glutamine synthetase, GlnA, occurs (Puglia *et al.*, 1995; Vohradsky *et al.*, 1997, 2000; Novotna *et al.*, 2003). SC6-ASB3 is the only strain that was observed to follow a diauxic growth pattern. This suggests it might be the only strain capable of metabolically adapting in response to nutrient exhaustion under these conditions. This hypothesis is supported by the fact that neither SC6-ASB2 nor SC6-ASB4 growth was sustained in this media after 14 hours, which might have been the result of nutrient depletion and a lack of metabolic adaptation.

5.2.3 DIFFERENCES BETWEEN PARENTAL AND EMPTY VECTOR CONTROL STRAINS IN DEFINED MEDIA

Next, the strains were grown in low casamino acid (0.005% (w/v)) NMMP supplemented with glucose, either 0.5% (w/v) or 1.0% (w/v) and containing thiostrepton (0.5 μg/ml). Specific growth rates of SC2-ASB0 and SC6-ASB0 cultures grown in NMMP with 0.5% (w/v) glucose were 0.11±0.01 h^{-1} and 0.13±0.00 h^{-1}, respectively (Table 5-1, Figure 5-10). These were comparable to the parental strains SC2 (0.13±0.01 h^{-1}) and SC6 (0.10±0.02 h^{-1}) growing in the same media in the absence of thiostrepton (Table 4-1, Figure 4-2).

In NMMP with 1.0% (w/v) glucose, specific growth rates were 0.06±0.00 h^{-1} for SC2-ASB0 and 0.16±0.03 h^{-1} for SC6-ASB0. In the thiostrepton-free version of the media, the parental strains SC2 and SC6 had grown at rates of 0.15±0.03 h^{-1} and 0.02±0.01 h^{-1}, respectively. The addition of thiostrepton to SC2 and SC6 cultures in this media had no measurable effect on the growth of these strains, indicating that either the presence of pIJ6902 and/or the combination of the presence of the plasmid and the thiostrepton is affecting the physiology of SC2-ASB0 and SC6-ASB0.
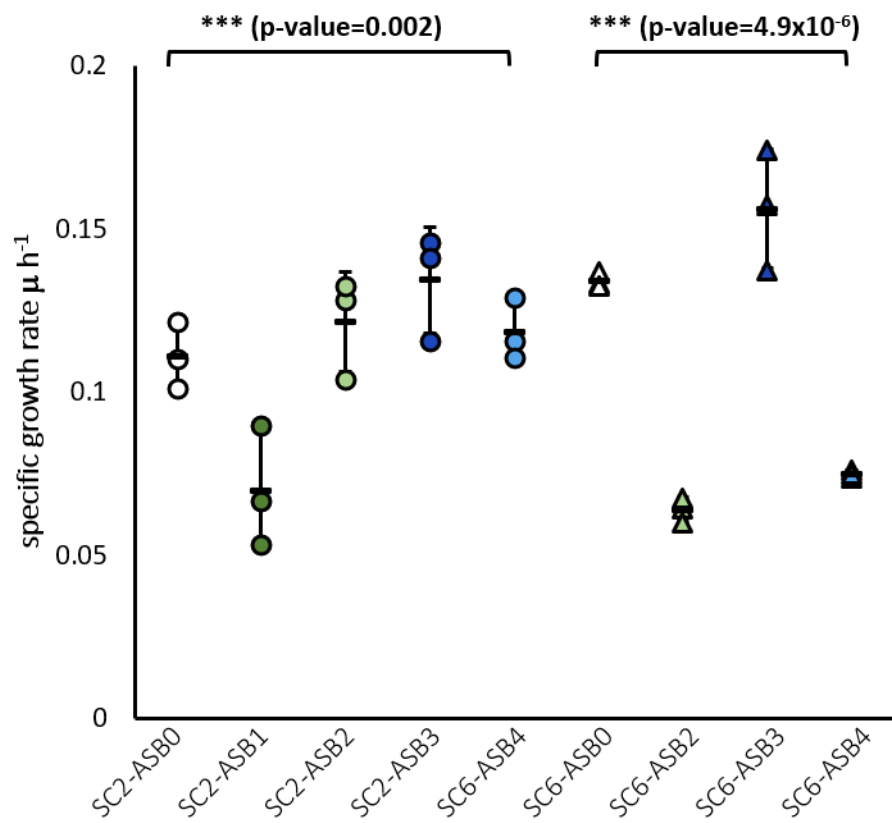
167

Figure 5-10: Specific growth rates (μ h^{-1}) for SC2 and SC6 constructed strains grown in reduced amino acids NMMP with glucose (0.5%). Differences across SC2 cultures were significant (***p-value=0.002, one-way ANOVA), as well as across SC6 cultures (***p-value=4.9x10^{-6}, one-way ANOVA).

In NMMP containing 0.5% (w/v) glucose, specific growth rates of SC2 constructed strains were comparable to SC2-ASB0 (0.11±0.01 h^{-1}), SC2-ASB1 (expressing *sclav_4529*) being the exception. Specific growth rates for SC2-ASB1, SC2-ASB2, SC2-ASB3 and SC2-ASB4 were 0.07±0.02 h^{-1}, 0.12±0.02 h^{-1}, 0.13±0.02 h^{-1} and 0.12±0.01 h^{-1}, respectively (Figure 5-10). The difference in specific growth rates across all strains was significant (p-value=0.002, one-way ANOVA). In a comparison with SC2-ASB0, however, only the difference in specific growth rates between SC2-ASB0 and SC2-ASB1 was significant (p-value=0.03, Tukey's HSD test). This infers that only heterologous expression of *glcP* from *S. clavuligerus* had a significant, growth rate-reducing effect in a SC2 background under the tested conditions.

This is also reflected in the corresponding growth curves (Figure 5-11A). While the curves representing growth of SC2-ASB0 and SC2-ASB2 (expressing *sclav_1340*) cultures are identical, and growth curves of SC2-ASB3 (expressing *sco5578*) and SC2-ASB4 (expressing *sco2126*) resemble the SC2-ASB0 curve, SC2-ASB1 cultures plateaued after only 17 hours. At this point, growth of the other strains was still sustained by the media. After 22 hours, absorbance of SC2-ASB1 cultures declined, resembling the curve recorded for SC6 in the same media. Given *glcP* expression in SC6 has been shown to be elevated compared to SC2 (see 4.2), this suggest that the shorter growth phase observed for SC6 and SC2-ASB1 cultures in NMMP containing 0.5% glucose might be the direct result of an increase of the GlcP-Glk transporter systems and, thus, glucose internalisation.

Specific growth rates determined for SC6 constructed strains growing in NMMP with 0.5% (w/v) glucose were 0.06±0.00 h^{-1} for SC6-ASB2, 0.16±0.02 h^{-1} for SC6-ASB3 and 0.08±0.00 h^{-1} for SC6-ASB4. The difference in growth rates was significant (p-value=4.9x10^{-6}, one-way ANOVA), as well as the differences between SC6-ASB0 and both *glk*-expressing strains, SC6-ASB2 (p-value= 0.001, Tukey's HSD test), and SC6-ASB4 (p-value= 0.001, Tukey's HSD test).

Interestingly, it is the SC6-ASB3 growth curve that stands out (Figure 5-11B). Growth curves of SC6-ASB0, SC6-ASB2 and SC6-ASB4 exhibit a long exponential phase (up to 40 hours in the case of SC6-ASB2).
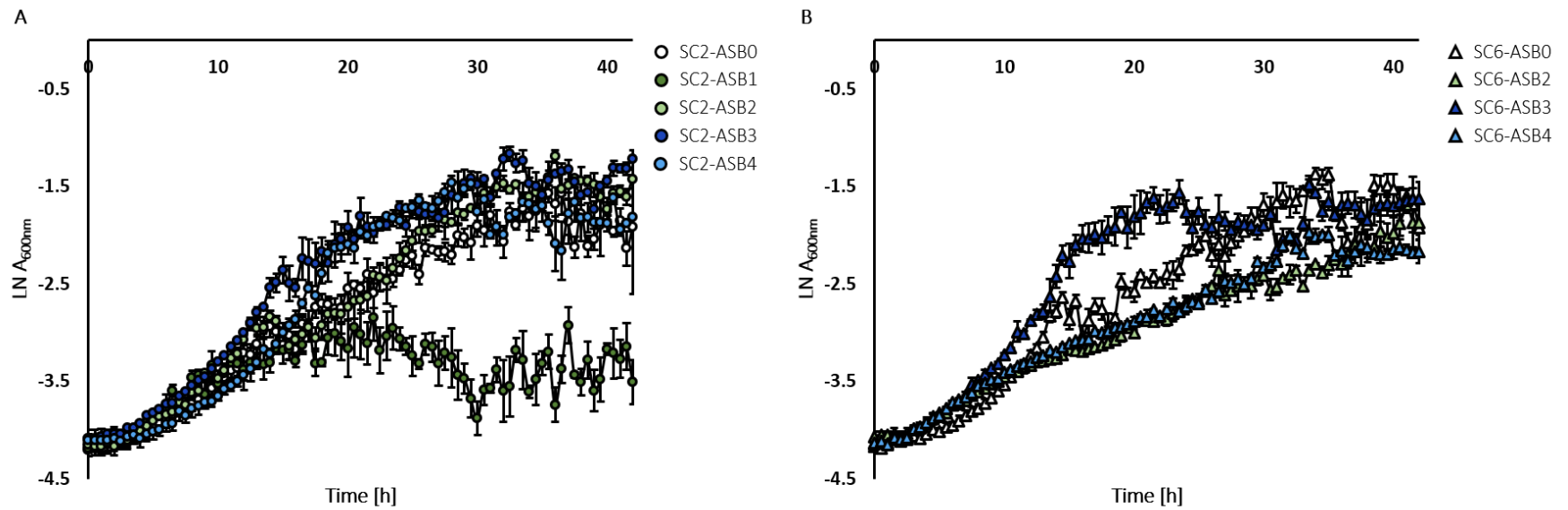
Figure 5-11: Growth curves of SC2 and SC6 cultures growing in low amino acid NMMP with glucose (0.5%). (A) SC2 and (B) SC6 constructed strains were grown in 24-well plates in TSB over 40 hours. The medium contained thiostrepton (0.5 µg/ml) The strains are represented according to the legends. The LN of absorbance (600 nm) is shown. Error bars indicate the standard error for across replicates.

A minor increase in absorbance can be seen after 31 hours, indicating that these cultures may have also undergone biphasic growth.

In NMMP containing 1.0% (w/v) glucose, specific growth rates of SC2-ASB1, SC2-ASB2, SC2-ASB3 and SC2-ASB4 cultures could not be determined due to very poor growth and pelleting, which is also obvious from the large error bars of the corresponding growth curves (Figure 5-12, Figure 5-13A). This behaviour resembled SC6 cultures growing under the same conditions, suggesting that the phenotype observed for SC6 in this media may, at least in part, be a result of the elevated *glcP* expression and the resulting increase of the GlcP-Glk uptake systems.

Specific growth rates for SC6 cultures were 0.14 ± 0.02 h^{-1} for SC6-ASB2, 0.16 ± 0.02 h^{-1} for SC6-ASB3 and 0.10 ± 0.01 μ h^{-1} for SC6-ASB4 (Figure 5-12), the difference across all strains being significant (p-value=0.01, one-way ANOVA). In a comparison with the empty vector control strains, the only significant difference between growth rates was between SC6-ASB0 and SC6-ASB4 (p-value=0.0.02, Tukey's HSD test). The SC6-ASB4 growth curve, further, highlights a clear difference in culture growth with absorbance peaking after 27 hours before steadily declining (Figure 5-13B).

These observations show that expression of additional copies of *glcP* and *glk* genes in SC2 and SC6 (including heterologous genes) has different impacts on growth rates and culture growth over time in a medium-, expressed genes-, and in strain background-dependent manner. SC2-ASB1 exhibited poorer growth than SC2-ASB0 across all tested media, indicating that expression of an additional copy of the native *glcP* gene in SC2 does not provide the strain with the tools to grow better in a glucose-rich environment compared to the parental strain. Expression of *sclav_1340* and *sco2126* in SC6 appeared to have a similar effect on growth across all tested media, indicating that Glk from *S. clavuligerus* and GlkA from *S. coelicolor* might elicit a similar response in SC6 under these experimental conditions. SC6-ASB3 exhibited growth patterns that enhance the diauxic growth of the strain, which might indicate that this strain is able to adjust metabolically in a way that the other strains cannot.
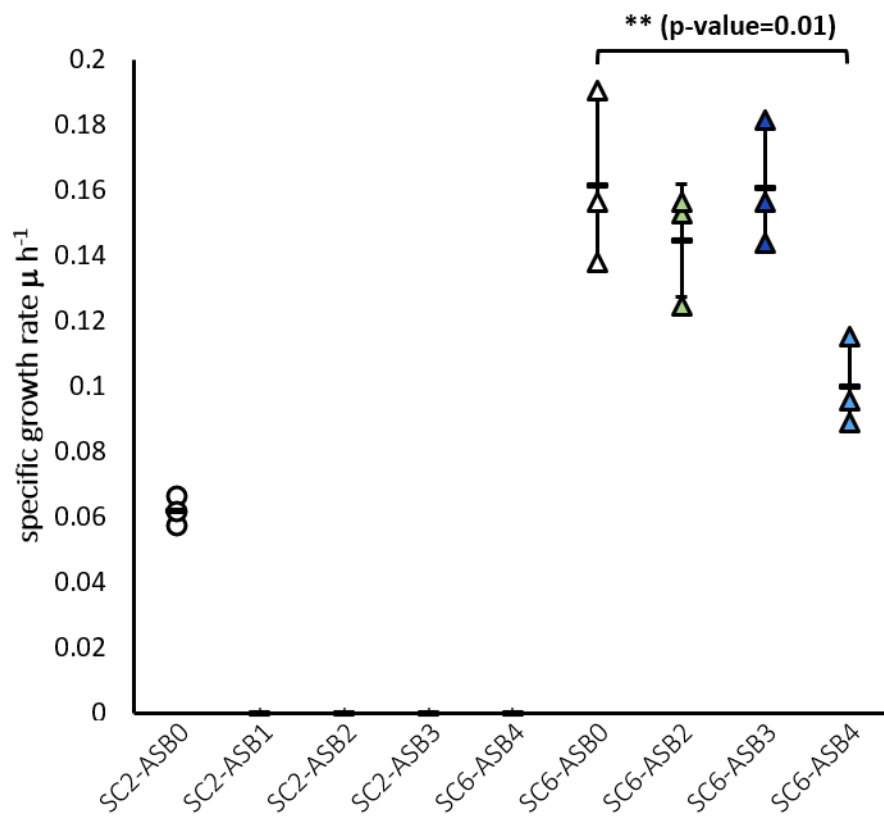
Figure 5-12: Specific growth rates (μ h^{-1}) for SC2 and SC6 constructed strains grown in reduced amino acids NMMP with glucose (0.1%). Differences across SC6 cultures were significant (**p-value=0.01, one-way ANOVA).
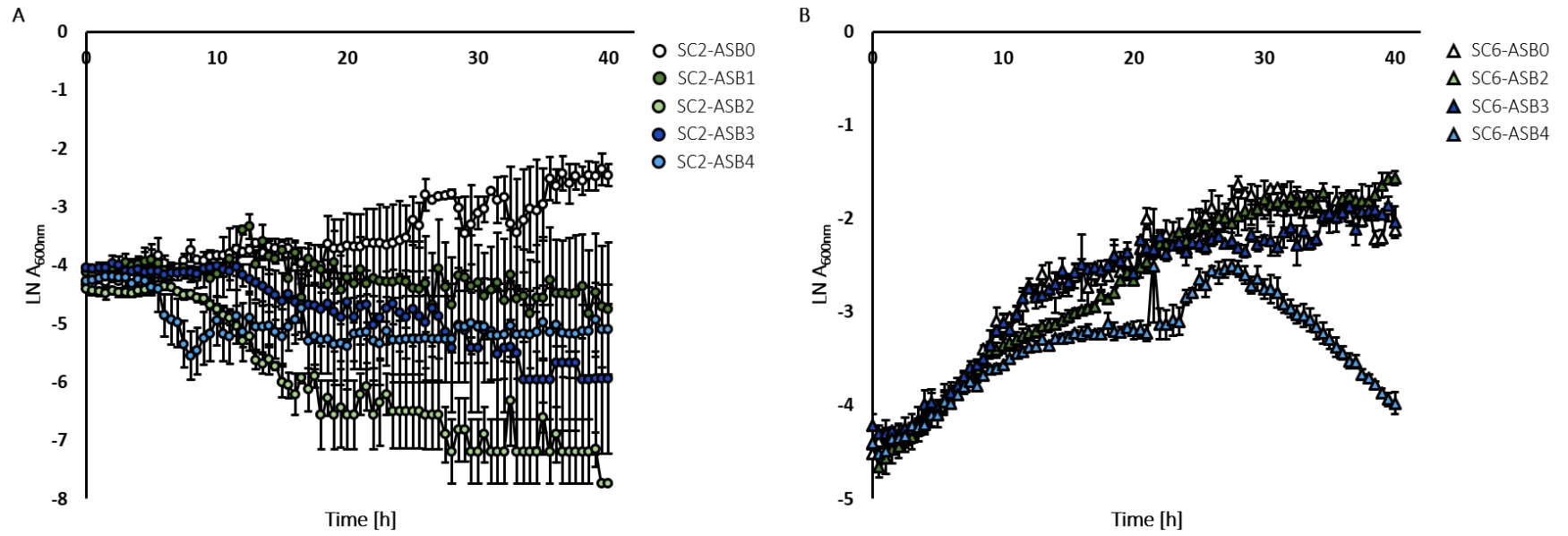
Figure 5-13: Growth curves of SC2 and SC6 cultures growing in low amino acid NMMP with glucose (1.0%). (A) SC2 and (B) SC6 constructed strains were grown in 24-well plates in TSB over 40 hours. The medium contained thiostrepton (0.5 µg/ml) The strains are represented according to the legends. The LN of absorbance (600 nm) is shown. Error bars indicate the standard error across replicates.

5.2.5 GROWTH AND DEVELOPMENT ON SOLID MINIMAL MEDIA

To assess whether heterologous expression of *glcP* or *glk* affects morphological differentiation, the constructed strains were grown on minimal media agar as before (see 4.1.4). The strains were spotted onto the agar either directly after a 48-hour pre-germination period or after having grown in TSB containing 0.5 μg/ml thiostrepton. Conditioning prior to spotting had previously not influenced growth and development of the parental strains (Figure 4-6), nor did it appear to cause different phenotypes in the constructed strains, except for SC6-ASB4 colonies. These colonies differed in their morphology depending on whether the cells had been pre-conditioned in glycerol-based pre-germination media or in glucose-based TSB. Thus, only a representative set of images is shown for each strain and two sets of images of SC6-ASB4 colonies are shown.

Growing SC2, SC2-ASB0 and *glcP*- and *glk*-expressing SC2 strains on minimal medium resulted in the same growth pattern for each strain, where the absence of glycerol resulted in sparse growth with most of the colonies being comprised of substrate mycelium (Figure 5-14A). The addition of glycerol to the medium, supported denser growth and aerial hyphae emerged in the centre of the colonies, which was most clearly visible for SC2 and SC2-ASB1 (Figure 5-14B). The expression of *glcP* or *glk* genes did not appear to majorly impact the development of SC2 constructed strains on this medium, despite the clear differences in growth of the corresponding liquid cultures. Given the absence of developmental phenotypes on solid medium, this suggests that regulators involved in morphological development are not affected by heterologous expression of *glcP* or *glk* in SC2.

Colonies of SC6 constructed strains resembled colonies of SC2 constructed strains with growth being seemingly poor in the absence of glycerol, the exception being SC6-ASB4 colonies (Figure 5-15A). While no SC6-ASB4 colonies grew in the absence of an additional carbon source when pre-germinated spores were spotted onto the agar, SC6-ASB4 colonies grown from TSB cultured-cells developed on all types of agar.
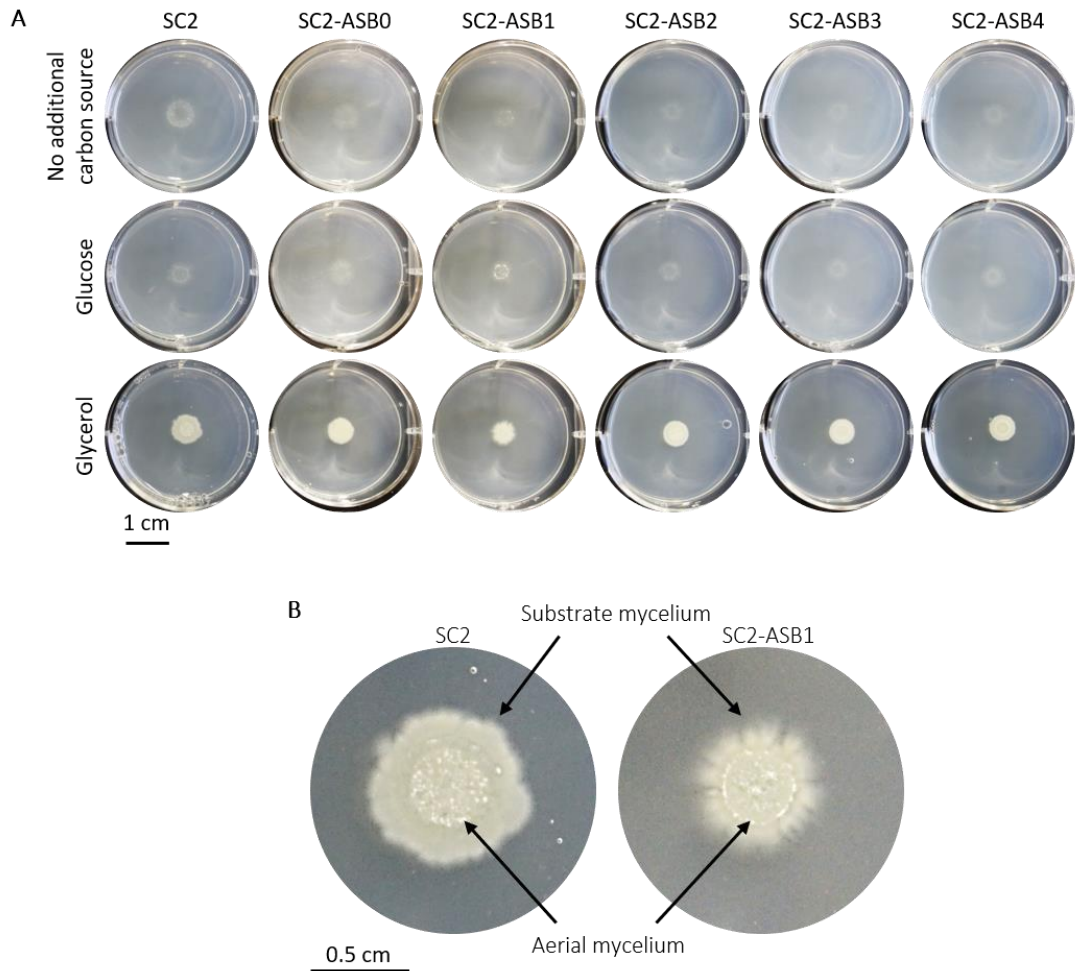
174

Figure 5-14: Development and growth of SC2 strains on minimal medium. (A) Colonies were imaged after a four-week incubation period (scale bar = 1 cm or 0.5 cm). The agar either lacked an additional carbon source (but contained casamino acids) or was supplemented with glucose or glycerol. (B) Aerial hyphae can be seen on enlarged images of SC2 and SC2-ASB1 colonies growing in the presence of glycerol.
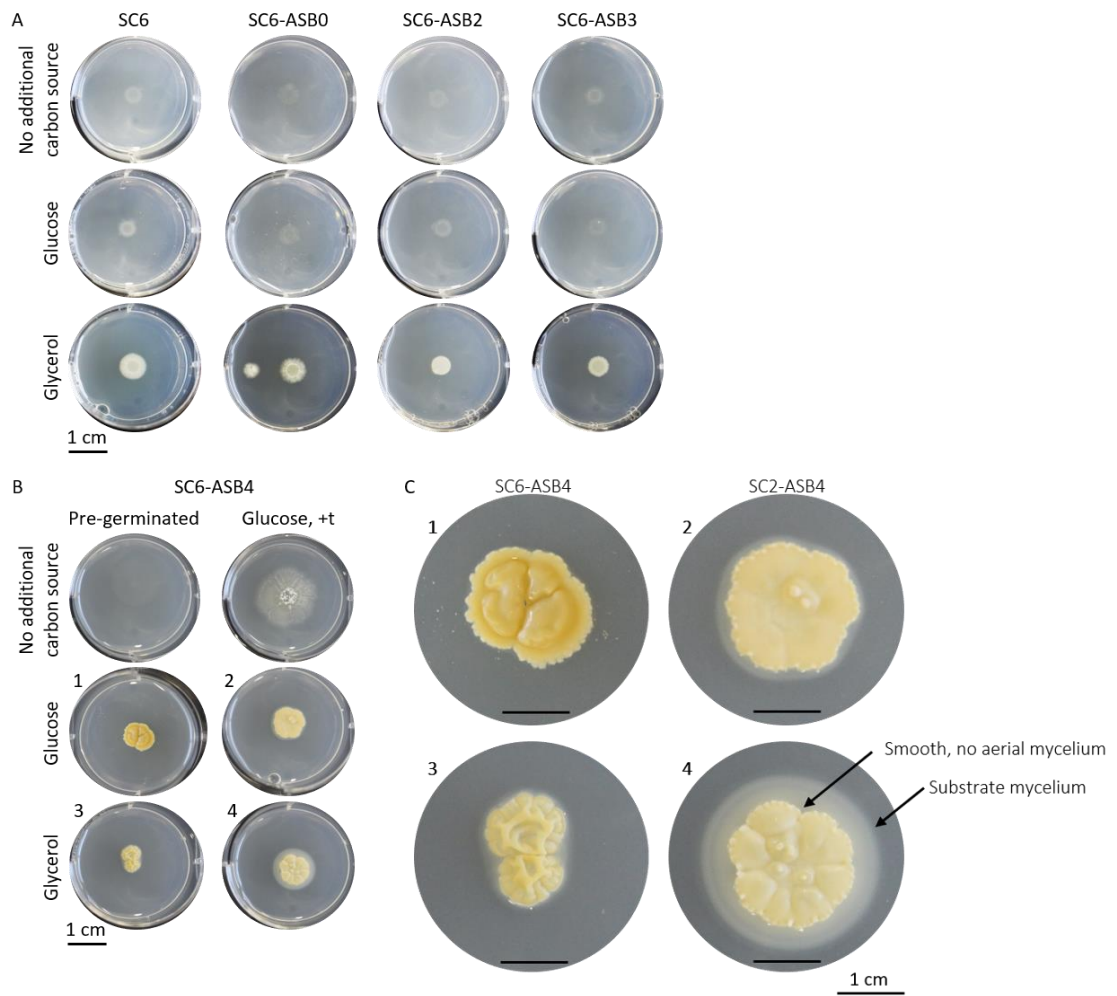
Figure 5-15: SC6, SC6-ASB0 and *glcP-* and *glk*-expressing SC6 colonies on minimal medium. Images were taken after four weeks of growth, scale bars indicate 1 cm. (A) Agar contained either no additional carbon source, glucose, or glycerol. (B) SC6-ASB6 colonies that emerged from spots of pre-germinated spores or spots taken from TSB-grown cultures differed in size. (C) Enlarged images (scale bar = 1 cm) of SC6-ASB4 colonies on glucose and glycerol-containing agar.

This was even the case in the absence of glucose and glycerol. On glucose and glycerol-containing agar, large, smooth, bald, and yellow colonies developed (Figure 5-15B and C). The yellow colour is indicative of the presence of substrate mycelia and might also indicate holomycin production (Higgens *et al.*, 1971; Kenig *et al.*, 1979). Moreover, microcolonies emerged within these colonies, which increased their heterogenous appearance. These are likely spontaneous mutants and highlight the genetic instability of SC6.

The absence of developmental phenotypes when *glcP* or *glk* are heterologously expressed, SC6-ASB4 being the exception, indicates that the GlcP-Glk system may not play a vital role in morphological differentiation in *S. clavuligerus*. In contrast, SC6-ASB4 exhibited a distinctly bald, potentially holomycin-(over)producing phenotype that was not observed for SC6 expressing its native *glk* gene (SC6-ASB2). This suggests GlkA from *S. coelicolor* may be involved in regulating development and morphological differentiation in a way that Glk from *S. clavuligerus* does not.

5.3   EFFECT OF HETEROLOGOUS EXPRESSION OF *glcP* AND *glk* ON CA PRODUCTION BY SC2 AND SC6

To determine the effects of heterologous expression of *glcP* and *glk* genes on SC2 and SC6 secondary metabolism, CA production in liquid medium was investigated. Contrary to observations made for DSM 738, SC2 and SC6, CA concentrations were detected in 1.5 ml culture supernatant. Therefore, endpoint measurements of CA after 40 hours of growth in TSB were carried out (Figure 5-16).

After 40 hours of growth in TSB with thiostrepton, the yield of CA produced by SC2-ASB0 was of 2.22±3.45 μg/mg CDW. While yields produced by SC2-ASB2, SC2-ASB3 and SC2-ASB4 were comparable, 2.33±3.61 μg/mg, 1.64±2.98 μg/mg and 3.07±1.79 μg/mg, respectively, CA production by SC2-ASB1 was undetectable (0.00±0.00 μg/mg). However, the difference between SC2-ASB0 and SC2-ASB1 yields was not significant. The lack of CA in SC2-ASB1 culture supernatants suggests that the presence of the native GlcP-Glk system in SC2 might cause depression and/or repression of CA biosynthesis (Figure 5-17). Given that TSB-grown SC2-ASB1 cultures had already reached lysis after 40 hours, the absence of CA from the supernatant might also be the result of CA degradation.
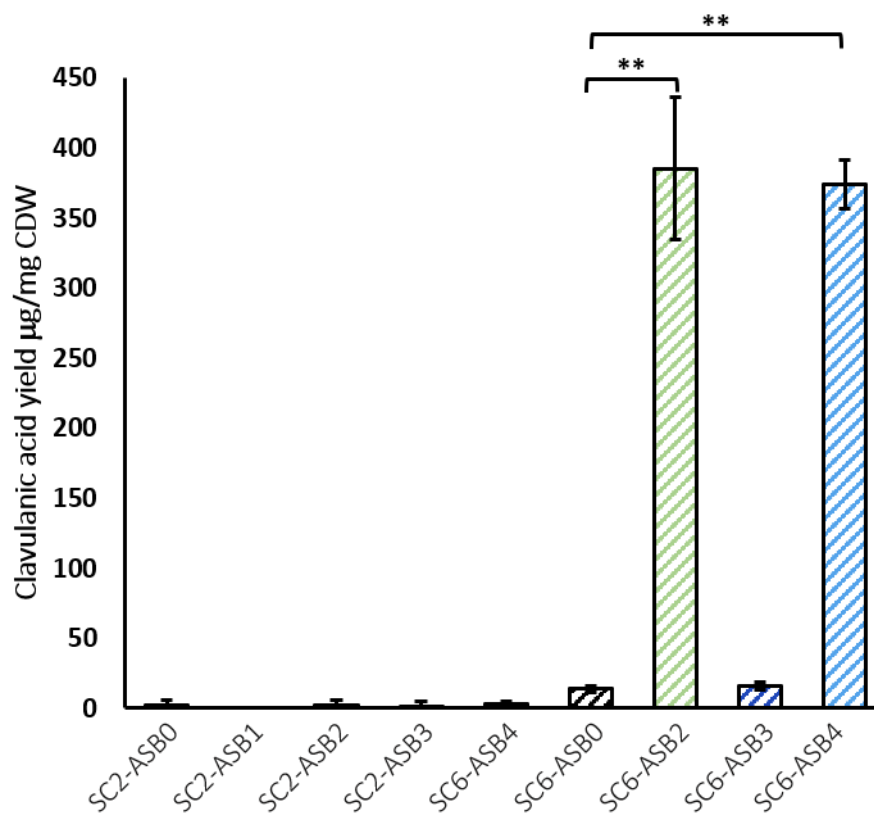
Figure 5-16: CA yields produced by *S. clavuligerus*. Yields were measured as μg/ml of produced CA per mg of CDW. Samples were taken after 40 hours of cultivation in TSB with thiostrepton (0.5 μg/ml). Clavulanic acid assays were carried out in biological duplicate and technical triplicate. Error bars indicate the standard deviation across replicates. Yields determined for SC6 cultures were significantly different (p-value=1.2×10^{-17}, one-way ANOVA, not indicated), with the difference between SC6-ASB0 and SC6-ASB2, as well as the difference between SC6-ASB0 and SC6-ASB4 being statistically different (**p-value=0.001, Tukey's HSD).
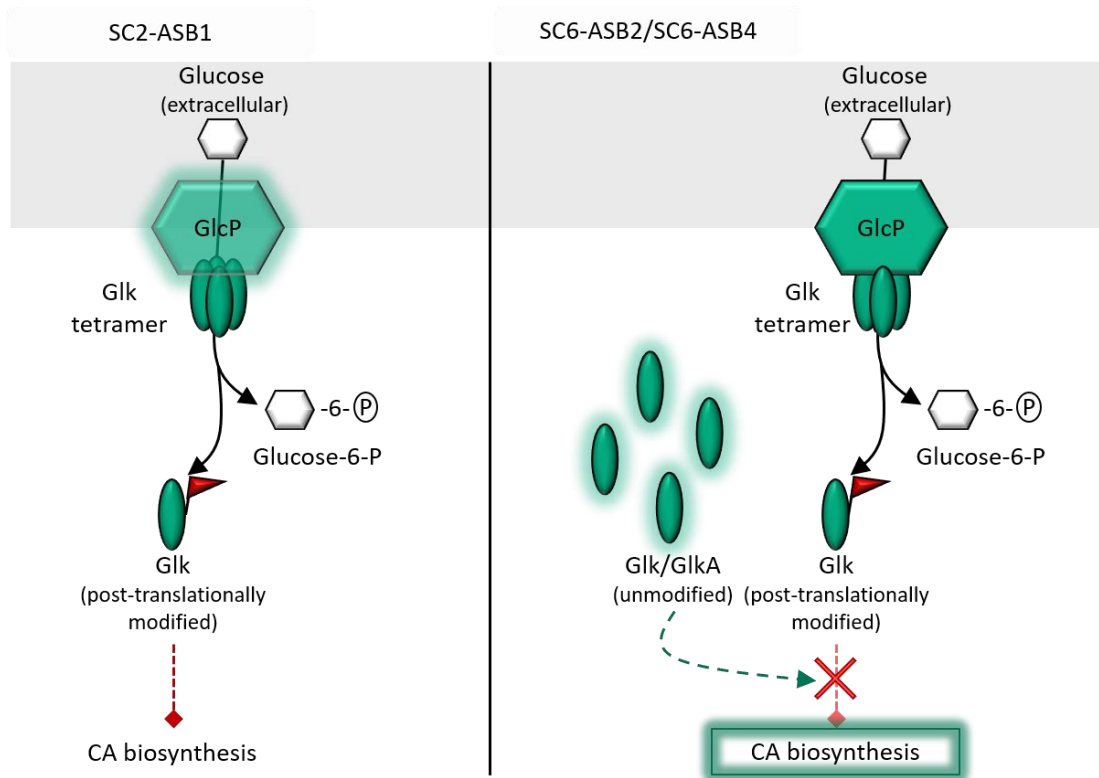
Figure 5-17: Schematic illustrating a possible mechanism of CCR that represses CA biosynthesis in SC2-ASB1 and enhances CA production in SC6-ASB2 and SC6-ASB4. Expression of *sclav_4529* in SC2 enables glucose uptake through the native GlcP-Glk system, resulting in post-translational modification of Glk, which depresses or represses CA biosynthesis (left). In SC6, where there is already a functioning GlcP-Glk system in place, heterologous expression of *sclav_1340* or *sco2126* results in an excess of unmodified Glk or GlkA, respectively. These unmodified kinases relieve the existing glucose-dependent depression of CA biosynthesis.

SC6 has been shown to produce significantly more CA than SC2 in TSB (Figure 4-10). This was also the case for CA yields produced by SC6-ASB0, which were 13.86±2.06 µg/mg, and, thus, significantly higher than the yields produced by SC2-ASB0 (p-value=0.0001, *t*-test). While SC6-ASB3 produced comparable yields of CA (16.05±3.09 µg/mg), the *glk*-expressing SC6 strains SC6-ASB2 and SC6-ASB4 produced 28-fold and 23-fold higher yields than the empty vector control strain, 385.51±50.88 µg/mg (p-value=0.001, Tukey's HSD) and 374.38±17.62 µg/mg (p-value=0.001, Tukey's HSD), respectively.

SC6 is likely to have a functioning GlcP-Glk system and has been shown to grow well in NMMP containing glucose (Figure 4-4). Given this glucose-positive phenotype and given that the expression of Glk results in such a profound rise in CA production levels, this might indicate that glucose does depress CA production in SC6 and that the excess of intracellular Glk in SC6-ASB2 and SC6-ASB4 that is not interacting with GlcP and, thus, lacking the posttranslational modification (PTM) status necessary for Glk to mediate CCR, might be responsible for the increase in CA yields in these strains (Figure 5-17) (van Wezel *et al.* 2007). As mentioned in the introduction (see 1.3.11.2), a novel form of PTM was recently identified in *S. roseosporus* that influences primary and secondary metabolism. Therefore, the PTM involved in modulating secondary metabolism in these strains might be crotonylation, which positively regulated CCR in *S. roseosporus* (Sun *et al.*, 2020).

5.4 PROTEOLYTIC ACTIVITY IN LIQUID CULTURES OF SC2 AND SC6 CONSTRUCTED STRAINS

Data presented in the literature and in the previous chapter suggest the existence of a physiological link between clavulanic acid biosynthesis and the production/activity of nutritious proteases in liquid (Aharonowitz 1979; Braña *et al.* 1985; Bascarán *et al.* 1990; Bascarán *et al.* 1991; Sánchez and Brana 1996, see 4.4.1). Given the complete lack of CA production by SC2-ASB1 and the increase in CA yields in SC6-ASB2 and SC6-ASB4 cultures compared to the vector control and parental strain, protease activity was assessed under the same experimental conditions to establish whether extracellular proteolytic activity produced by these strains was also altered.

Unfortunately, protease activity of TSB-grown cultures, measured in U/mg CDW, was generally low across all cultures (Figure 5-18). Activities measured in SC2 cultures were 0.03±0.01 U/mg for SC2-ASB0, 0.04±0.22 U/mg for SC2-ASB1 and 0.03±0.02 U/mg for SC2-ASB2. In SC2-ASB3 and SC2-ASB4 culture supernatant, activities of 0.07±0.07 U/mg and 0.03±0.02 U/mg were determined, respectively. Activities measured for SC6-ASB0 and SC6-ASB3 cultures were similarly low, 0.10±0.06 U/mg and 0.04±.01 U/mg, respectively. Although proteolytic activities measured in SC6-ASB2 and SC6-ASB4 cultures were high in comparison, namely 1.23±0.82 U/mg and 1.21±0.92 U/mg, respectively, these increases were not statistically significant compared to the empty vector control culture.

Wildtype *S. clavuligerus* is known to produce high amounts of extracellular protease in liquid medium after nutritional shift-down, showing that nutrient availability prevents extracellular protease activity (Bascarán *et al.*, 1990). Low protease activity in TSB-grown cultures might, thus, be indicative of nutrients being available at the time of sampling.

5.4.1 CARBON SOURCE DEPENDENCE OF EXTRACELLULAR PROTEASE ACTIVITY ON SOLID MEDIUM

In the previous chapter, the presence of glycerol in the media was found to prevent proteolytic degradation of milk protein by DSM 738 and SC2 but not SC6 (see 4.4.3, Figure 4-12), suggesting that protease production by WT(-like) *S. clavuligerus* might be the subject of glycerol dependent CCR. The glycerol insensitive phenotype observed for SC6 might be the result of expression of the native GlcP-Glk system, the presence of which might be interfering with CCR pathways to regulate protease production. Thus, the proteolytic degradation of milk protein by SC2 and SC6 constructed strains was tested in the presence of glucose and glycerol. Analogously to spotting onto minimal medium, the difference in protease activity and colony morphologies were compared in a carbon source-dependent manner, but also between colonies of the same strain that had been spotted directly after pre-germination and such grown from TSB-grown cultures with thiostrepton (0.5 μg/ml).
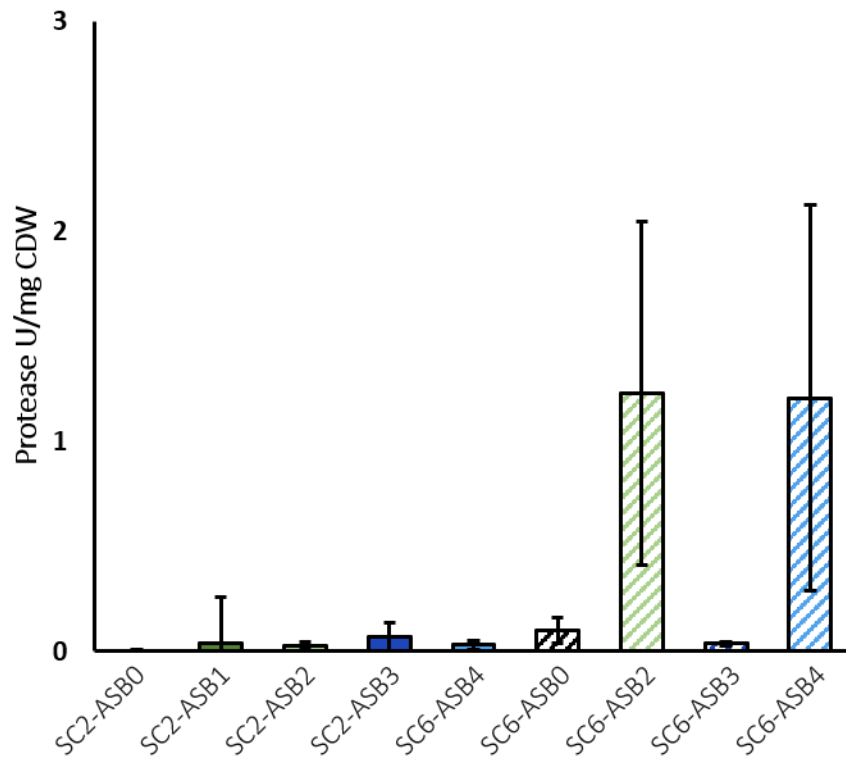
Figure 5-18: Proteolytic activities [U/mg CDW] of SC2 and SC6 cultures. Samples were taken after a 40-hour incubation period. The strains were grown in TSB with thiostrepton (0.5 µg/ml). Error bars indicate the standard deviation across replicates. Differences were not statistically significant.

Protease activity was clearly visible on milk agar with thiostrepton (0.5 μg/ml), on which colonies of SC2 constructed strains behaved like the parental strain (SC2) with glycerol causing the absence of extracellular protease activity (Figure 5-19). While this was the case for SC2-ASB0, SC2-ASB2, SC2-ASB3 and SC2-ASB4, regardless of whether the colonies had emerged from pre-germinated spores or TSB-grown cells, SC2-ASB1 colonies grown from pre-germinated spores behaved distinctly differently (red box). The media was cleared entirely by all SC2-ASB1 colonies irrespective of the presence of glucose or glycerol. The colony morphology strongly resembled that observed for SC6 on this media (Figure 4-12). However, unlike SC6, the glycerol-insensitive phenotype exhibited by SC2-ASB1 was conditional. In the previous chapter, it was proposed that the glycerol-insensitive phenotype observed for SC6 might be due to the native GlcP-Glk system interfering with regulatory pathways of amino acid catabolism, glycerol catabolism and, thus, protease production. The fact that this same phenotype is also exhibited by SC2-ASB1, albeit it conditional, supports this hypothesis.

Based on these data and the model presented for the regulation of protease production in WT *S. clavuligerus* in the presence of glycerol (Figure 4-13), a model is proposed for the control of this process in *S. clavuligerus* strains expressing their native *glcP-glk* system (SC6 and SC2-ASB1, Figure 5-20): While the uptake and conversion of glycerol to glycerol-3-phosphate under amino acid starvation conditions would normally prevent the secretion and, thus, activity of extracellular protease(s) in the WT, a process which may be GylR-dependent (left panel), the presence of the native GlcP-Glk system is sufficient to prevent proteolytic activity in the presence of glycerol (right panel). This disruption might depend on the interaction of GlcP with Glk, since *glk* expression occurs in WT *S. clavuligerus* (see 4.2) and does not result in a glycerol-insensitive protease phenotype (see 4.4.2, Figure 4-12). It is, further, proposed that Glk might interact with GylR in a way that prevents dissociation of the repressor from upstream of the glycerol utilisation operon. This, in turn, might prevent unbound GylR to interact with other targets, such as the metalloprotease gene, *sclav_4359*. Carbon flux through GlcP-Glk in SC2-ASB1 might result in the absence of an interaction between Glk and GylR due to post-translational modification of the former (van Wezel *et al.*, 2007), which might explain why the glycerol-insensitive phenotype is reversed when SC2-ASB1 spores are pre-conditioned in TSB.
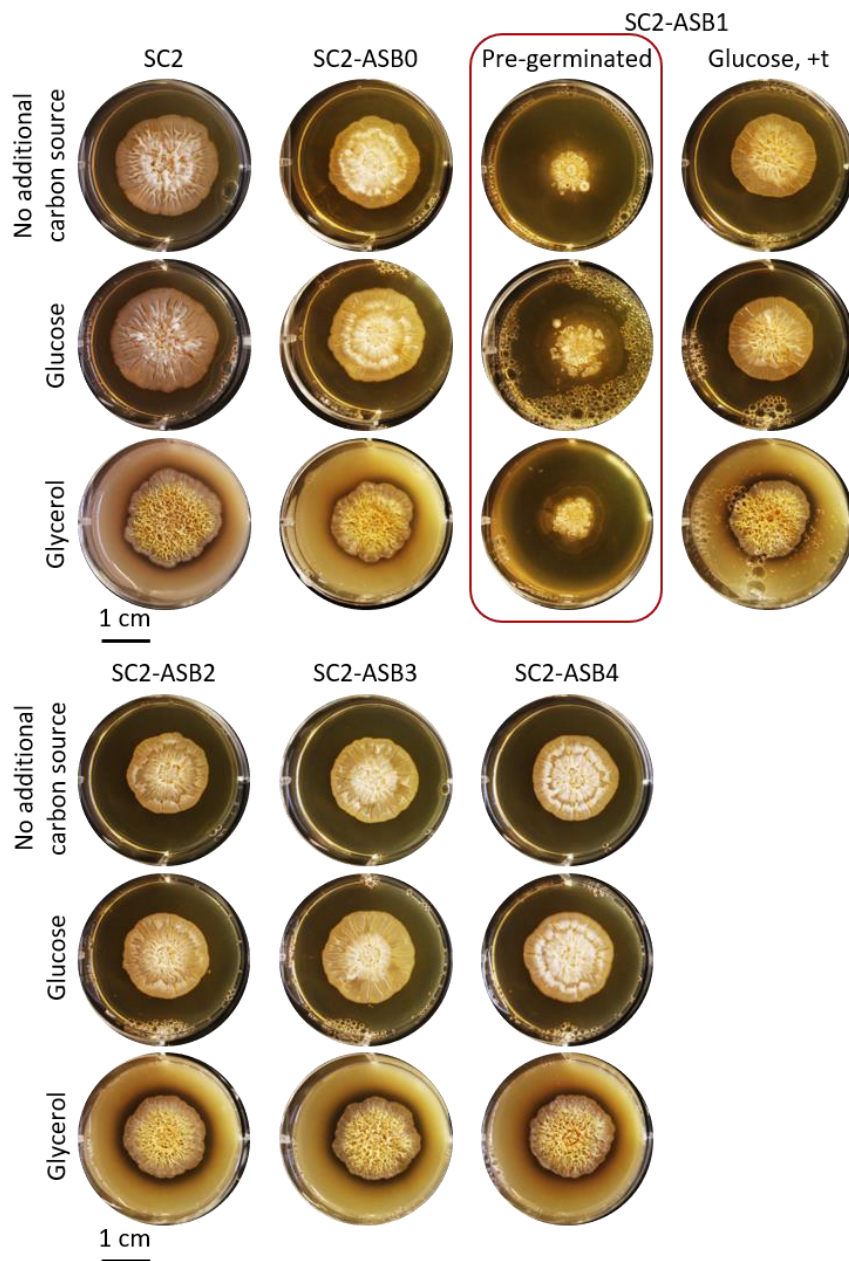
Figure 5-19: Clearing of agar from milk protein from agar by SC2 strains. The agar contained thiostrepton (0.5 µg/ml) unless the parental strain, SC2, was spotted. Pre-germinated spores were spotted onto milk agar in 6-well plates and images were taken after a four-week incubation period (scale bar = 1 cm). The colonies were grown either in the absence of an additional carbon source or in the presence of glucose or glycerol. For SC2-ASB1 colonies, one set shows pre-germinated spore-borne colonies (Pre-germinated), the other shows colonies spotted from cells that had been cultured in TSB containing 0.5% glucose and 0.5 µg/ml thiostrepton (Glucose, t).
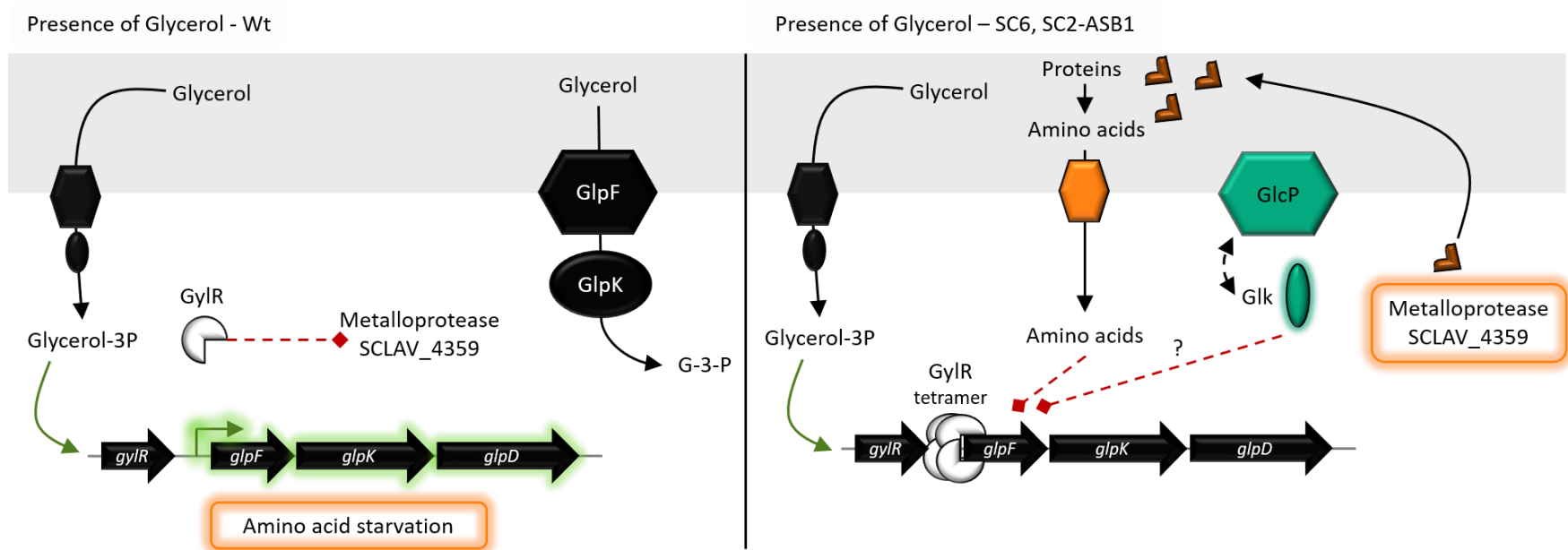
Figure 5-20: Schematic of proposed interplay of protease and carbon catabolic regulatory mechanisms in SC2-ASB1 and SC6. Inhibitory regulation is indicated by the red diamond arrows, induction is represented by the green arrows. In WT *S. clavuligerus*, L-serine is constitutively transported and represses the major glycerol uptake system (Minambres *et al.*, 1992). Glycerol, in turn, suppresses the production and activity of extracellular proteases (Figure 4-12). In SC2-ASB1 and SC6, protease activity clears the agar of milk protein in the presence of glycerol and glucose. This can be explained by the activity of Glk, which is post-translationally activated by carbon flux through the GlcP-Glk uptake system (van Wezel *et al.*, 2007). It is thought to interact with GylR, the autoregulatory repressor of the *glpFKD* operon. The presence of the native GlcP-Glk system disrupts existing regulatory mechanisms of amino acid and glycerol catabolism resulting of the glycerol-insensitive phenotype of SC2-ASB1 and SC6.

SC6, however, has also exhibited an amino acid-negative phenotype when cultivated in NMMP containing casamino acids as carbon and nitrogen source (see 4.1.2, Figure 4-4). This suggests that the irreversibility of glycerol insensitivity in terms of protease production by SC6 might have additional metabolic causes.

5.4.1.2 ABSENCE OF PROTEOLYTIC ACTIVITY OF SC6-ASB4 COLONIES

The effects of *glcP* and *glk* expression in SC6 constructed strains on milk agar were more diverse (Figure 5-21), starting with the major discrepancies between SC6 and SC6-ASB0 in terms of agar clearing. Protease production by the latter strain resembled that of SC2 where milk protein degradation was not observed in the presence of glycerol. Differences between parental and empty vector control strains had also been observed in liquid, highlighting the gap in knowledge about the responses triggered by thiostrepton in *S. clavuligerus*. SC6-ASB2, SC6-ASB3 and SC6-ASB4 colonies all had slightly different colony morphologies and cleared the agar to varying extents. For SC6-ASB2, the difference between pre-germinated spore-borne colonies and TSB culture-borne colonies was visible, the former resembled SC6 colonies with being small, heterogenous and having cleared the agar completely, while the latter appeared to have degraded the milk protein to a lesser extent.

While SC6-ASB3 colonies resembled SC6-ASB0 colonies (no agar clearing in the presence of glycerol), SC6-ASB4 colonies were larger and more crenulated than any of the other colonies. No clearing was observed regardless of the absence or presence of a carbon source. In fact, these colonies strongly resembled BAP20 (*ΔglcP1::hyg, ΔglcP2::*aacC4) colonies growing on milk agar (van Wezel *et al.* 2005, Figure 5-22). Compared to the WT strain, M145, growing on the same medium (no additional carbon source), BAP20 formed larger, more crenulated colonies that did not produce any blue pigment, indicating the role that the GlcP/Glk system has in regulating antibiotic production in *S. coelicolor*. The resemblance of SC6-ASB4 colonies with BAP20 colonies suggests that heterologous expression of *sco2126* in SC6 causes similar morphological and physiological changes than the absence of GlcP in *S. coelicolor*. The stark contrast of proteolytic activity produced by SC6-ASB4 in liquid versus the absence thereof on solid medium also highlights that secondary metabolite production may be controlled differently in liquid and on solid substrate.

186
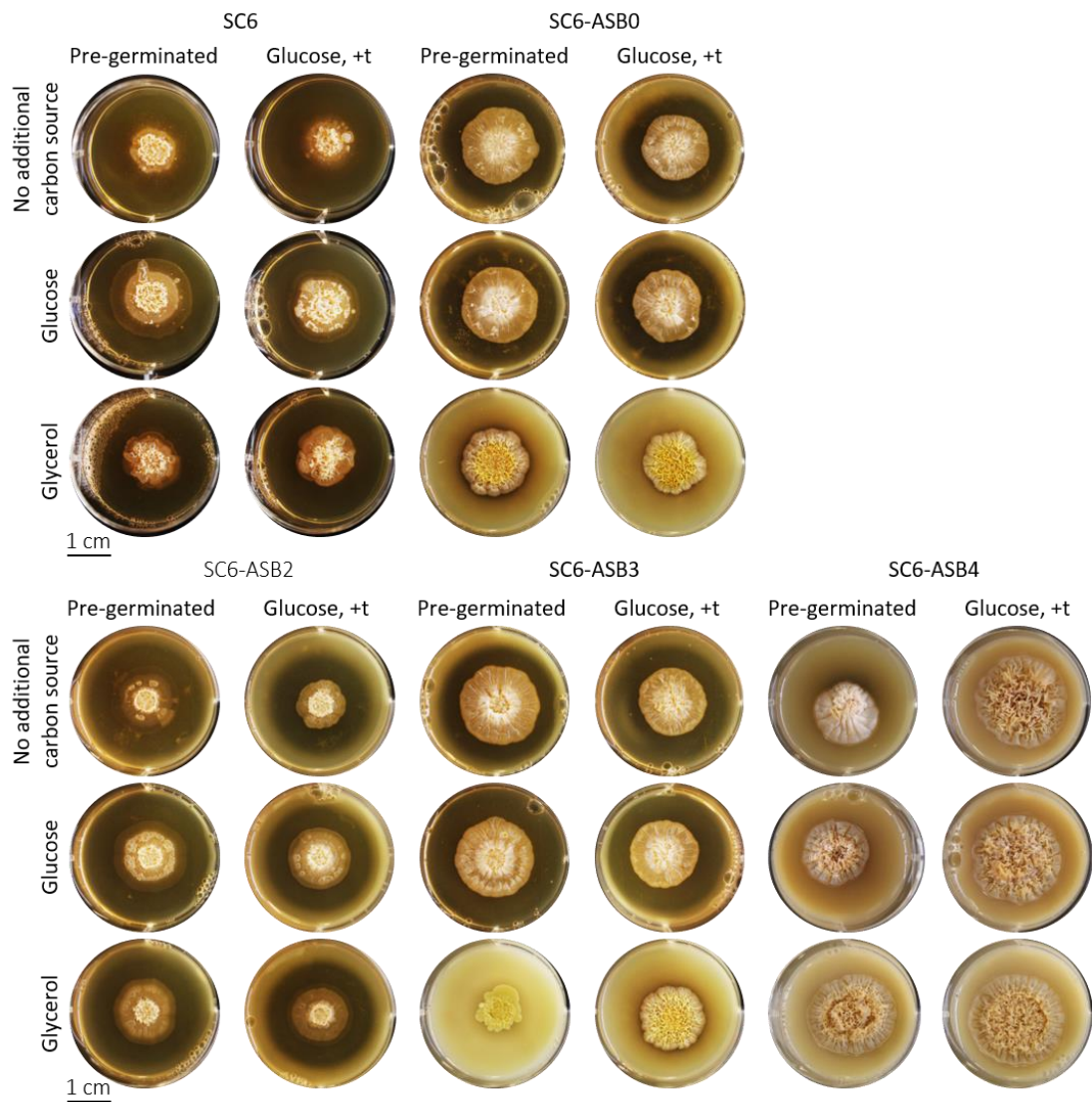
Figure 5-21: Clearing of agar from milk protein by SC6 strains. The agar contained thiostrepton (0.5 μg/ml) unless SC6 was spotted. Colonies were spotted onto milk agar (6-well plates, scale bar = 1 cm) either directly after a 48-hour long pre-germination period (Pre-germinated) or after having grown in TSB containing 0.5 μg/ml of thiostrepton (Glucose, t). Images were taken after a four-week incubation. The agar either lacked an additional carbon source or contained glucose or glycerol.
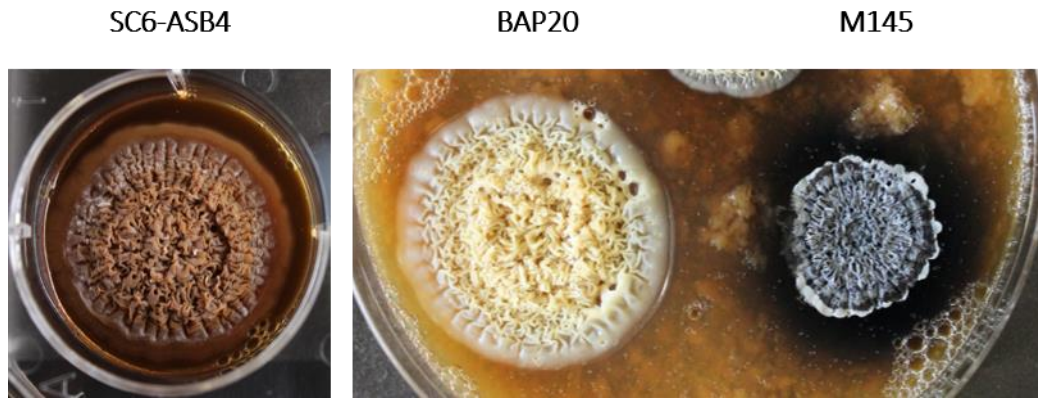
Figure 5-22: SC6-ASB4, BAP20 and M145 colonies on milk agar. SC6-ASB4 was spotted on agar (6-well plate) containing glycerol and thiostrepton (0.5 µg/ml), the *S. coelicolor* strains were grown on thiostrepton-free agar lacking an additional carbon source (standard size petri dish). Images were taken after a six-week incubation period.

The glucose uptake system in *Streptomyces* and the involvement of GlkA in glucose dependent CCR has been studied in detail in *S. coelicolor* (Angell *et al.*, 1994; van Wezel *et al.*, 2005, 2007). One of the advantages of studying CCR in *S. coelicolor* is that the pigmented antibiotics, actinorhodin and undecylprodigiosin, are visible to the naked eye and the biosynthesis of which is dependent on glucose catabolism (Hodgson, 1980). Actinorhodin is a pH indicator that is blue above a pH of 8.5 and red below that pH (Brockmann *et al.*, 1950, 1955). Expression of *sco2126* in SC6 produced a promising increase in CA yield in liquid (Figure 5-16), as well as a repressed proteolytic phenotype on milk agar (Figure 5-21). To assess whether the expression of *S. clavuligerus glcP* or *glk* genes in *S. coelicolor* also results in altered antibiotic production phenotypes, BAP20 strains heterologously expressing either *sclav_4519* (BAP20-ASB1) or *sco5578* (BAP20-ASB3) were grown in the presence of glucose and pigment production was examined. The strains were grown in NMMP containing glucose (0.5% (w/v)) for 40 hours, after which actinorhodin concentrations were determined spectrophotometrically in culture supernatant (Figure 5-23).

Determining actinorhodin concentrations from small scale liquid cultures (96-well plates) showed that the concentration produced by M145 was approximately four times lower than that produced by the GlcP-deficient strain BAP20, 4.57±0.70 µM/ml and 19.16±3.03 µM/ml, respectively. This might be attributed to the repressive effect glucose would have had on the antibiotic production in the WT strain prior to glucose exhaustion (Kwakman *et al.*, 1994). Actinorhodin concentrations measured in BAP20-ASB0 cultures were reduced compared to the parental cultures, 11.23±1.72 µM/ml, which might reflect the metabolic burden placed on the cell by maintaining the plasmid on the genome. Interestingly, heterologous expression of *sclav_4529* in BAP20 (BAP20-ASB1) resulted in actinorhodin concentrations comparable to those produced by BAP20 cultures (16.74±2.67 µM/ml), which is roughly a 1.5-fold increase compared to concentrations in BAP20-ASB0 cultures. In contrast, actinorhodin concentrations in BAP20-ASB3 (expresses *sco5578*) cultures, were 0.88±0.19 µM/ml, which is approximately 13-fold reduced compared to concentrations determined for BAP20-ASB0 cultures, respectively.
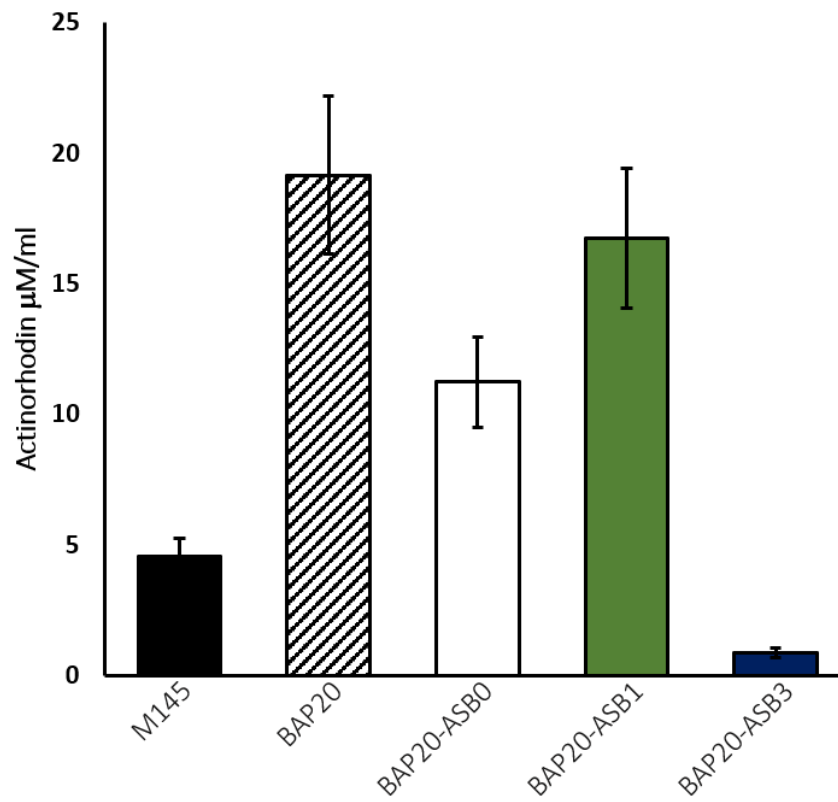
Figure 5-23: Actinorhodin concentration in *S. coelicolor* cultures. Strains were grown in NMMP with glucose (0.5%) for 40 hours. Endpoint actinorhodin concentrations were assessed spectrophotometrically from culture supernatant. Samples were measured in triplicate. Error bars indicate the standard deviation. Differences in actinorhodin concentration were not statistically significant.

This suggests that only heterologous expression of *sco5578* in BAP20 restores glucose-dependent repression of actinorhodin biosynthesis under the tested conditions. Pigment production by and morphological development of BAP20 strains were also assessed on GYM and milk agar plates containing a range of carbon sources. Pre-germinated spores of M145, BAP20, BAP20-ASB0, BAP20-ASB1 and BAP20-ASB3 were spotted onto GYM-agar (12-well plates) and incubated for 168 hrs, Figure 5-24). While M145 colonies had not formed aerial mycelia at this point, which is to be expected given that the glucose present in the media would have prevented developmental progression, BAP20 colonies were large, crenulated, and fuzzy in places. This indicates development progressed faster for BAP20 colonies than for M145 colonies in the same amount of time, presumably due to the inability of BAP20 to utilise glucose. The empty vector control colony, BAP20−ASB0, differed from the parental strain in that red and blue pigments were produced. Heterologous expression of *sclav_4529* in BAP20 (BAP20-ASB1) resulted in a bald phenotype and colonies were markedly smaller than colonies of any other BAP20 strain. The surface was smooth and shiny, yet when these colonies were re-streaked onto MS agar, they had a WT *S. coelicolor* morphology. In contrast, BAP20-ASB3 colonies were comparable to other BAP20 colonies in size and crenulation. Pigment production, however, was visibly enhanced. The dissimilarity of BAP20-ASB3 and M145 colonies might be the result to of complementation of *glcP* at an ectopic location in the chromosome (*attB* site), which results in incomplete complementation. Additionally, *glcP* is expressed from *ptipA* in BAP20-ASB3 and not the gene's native promoter. While expression from *ptipA* is leaky, expression from the *glcP* promoter is induced by glucose, a metabolic fine-tuning that is lacking in BAP20-ASB3.

On milk agar containing glucose, glycerol, arabinose, xylose, mannitol, maltose, fructose, or lactose, M145-ASB0 produced actinorhodin in the presence of all carbon sources but glycerol (Figure 5-25), as did BAP20-ASB0, albeit to a lesser extent. While M145-ASB1 colonies resembled M145-ASB0 colonies, BAP20-ASB1 colonies were small, bald, shiny, and white on all types of media but on milk agar supplemented with xylose. This phenotype was also observed for BAP20-ASB1 growing on GYM agar (Figure 5-24).
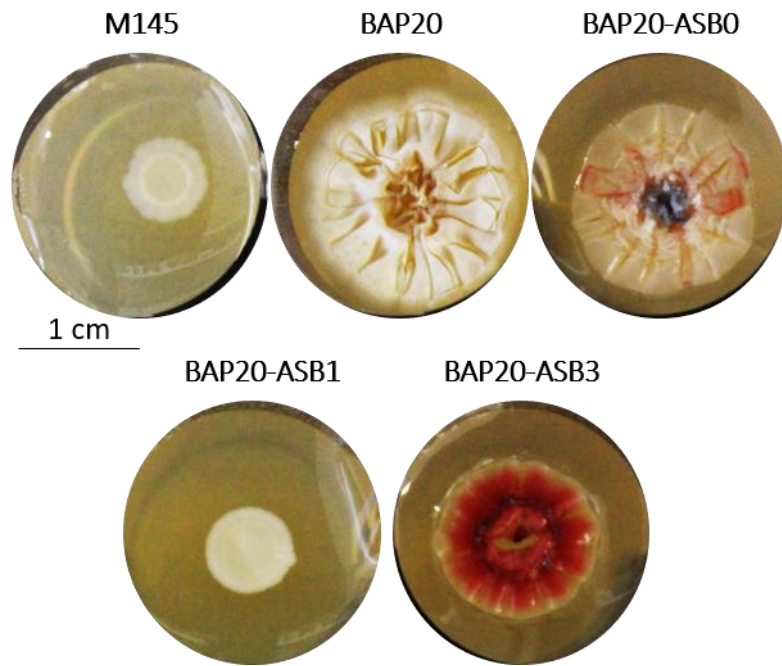
Figure 5-24: Growth, development, and pigment production by *S. coelicolor* strains on GYM agar (scale bar = 1 cm). Images were taken after a six-week incubation period. Red and blue pigments were produced.
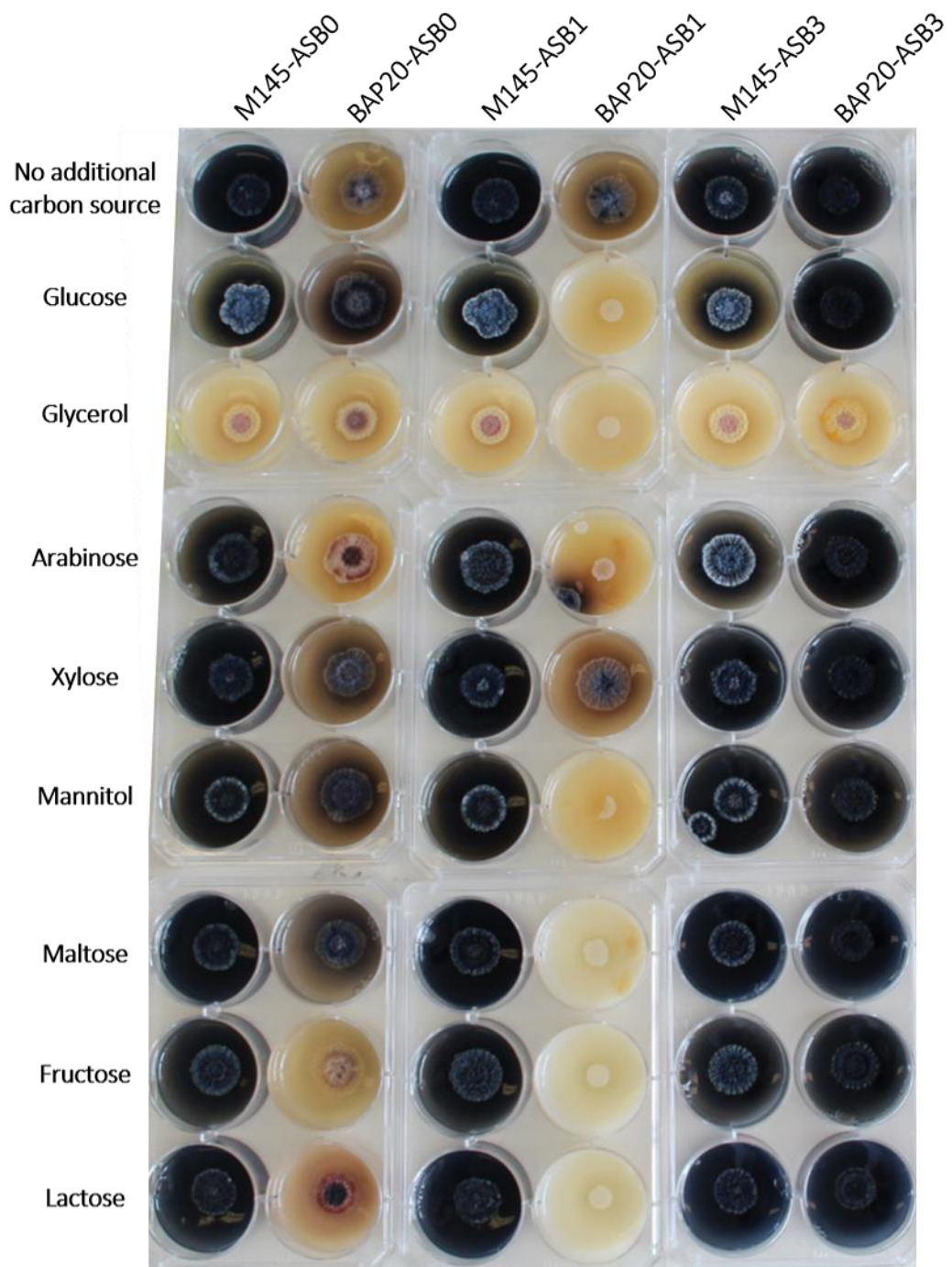
Figure 5-25: Pigment production by M145 and BAP20 strains on milk agar. The strains were pre-germinated in 2x YT medium prior to spotting on milk agar containing the indicated carbon sources. Colonies were grown on 6-well plates. Images were taken after a six-week incubation period.

In contrast, BAP20-ASB3 appeared to produce more pigment than M145-ASB3 or M145-ASB0, which had also been previously observed to be the case on GYM agar. Given the Glk-independent but Rok7b7-dependent regulation of xylose utilisation in *S. coelicolor*, the progression of development observed for BAP20-ASB1 strain in the presence of xylose might simply indicate that Rok7b7 is not affected by the absence of GlcP nor the presence of pIJ6902-ASB1 (Świątek *et al.*, 2013).

5.6 SUMMARY

Heterologous expression of *glcP* and *glk* genes in SC2 and SC6 produces a range of growth phenotypes and appears to affect secondary metabolism in different ways (overview of SC2 strains given in Table 5-2, overview of SC6 strains given in Table 5-3). While no significant differences between SC2-ASB0 and SC2-ASB2 to SC2-ASB4 were detected in terms of growth rates in TSB and NMMP containing glucose, CA yields produced in liquid cultures, as well as development and proteolytic activities observed on solid medium of SC2-ASB1 cultures were markedly reduced (Figure 5-16). These data suggest that the presence of the native GlcP-Glk system in SC2 does not facilitate faster growth rates or improved CA titres in the presence of glucose. This, in turn, might be due to glucose triggering repression or depression of CA in SC2-ASB1. This strain also exhibited a conditional glycerol-insensitive protease phenotype that resembled SC6 growing on milk agar (Figure 5-19). This infers that the glycerol-insensitivity of SC6 might be, at least partly, caused by the strains increased *glcP* expression (see 4.2). Strains that produced significantly higher CA titres than any of the other strains were SC6-ASB2 and SC6-ASB4, both of which expressed *glk* genes, *sclav_1340* or *sco2126*, respectively. Based on the assumption that glucose causes repression or depression of CA biosynthesis, the excess of Glk proteins in SC6-ASB2 and SC6-ASB4 due to heterologous expression of *sclav_1340* or *sco2126* might be relieving CCR, resulting in increases CA production. This, however, might require the presence of a functioning GlcP-Glk system, which is why no such increase in CA yield was observed for SC2-ASB2 and SC2-ASB4. SC6-ASB4 showed no proteolytic activity on milk agar and colonies had an altered morphology that resemble the BAP20 strain. Furthermore, heterologous expression of *sclav_4529* in BAP20 produced a bald, shiny colony morphology with an altered pigment production behaviour, indicating that GlcP and Glk proteins from *S. clavuligerus* and *S. coelicolor* might not be compatible with one another when expressed in either of the species.

Table 5-2: Overview of growth rates and CA yields produced by SC2 constructed strains. Heterologously expressed components of the glucose uptake system (GlcP=hexagon, Glk=oval tetramer) is indicated by the glow of the objects, *S. clavuligerus* proteins are shown in green, *S. coelicolor* proteins are shown in blue. The growth rates are shown as percentage of the rate calculated for the SC2-ASB0 cultures in the indicated medium (error indicated shows propagated error according to Gauss). CA yields are shown as μg/ml CDW and alterations in colony morphologies and clearance of agar (indicative of protease production) are listed.
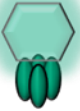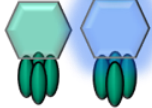
	Strain	μ h⁻¹ TSB 0.5% glucose	μ h⁻¹ NMMP 0.5% glucose	CA yield [μg/mg CDW]	Morphology and agar clearance on solid
	SC2-ASB0	100%	100%	2.22±3.45	Parental morphology, clearance of agar suppressed by glycerol
	SC2-ASB1	63±9%	64±19%	---	SC6-like morphology, no suppression of protease activity by glycerol
	SC2-ASB2	58±6%	109±21%	2.33±1.64	Like SC2-ASB0
	SC2-ASB3	72±7%	118±21%	3.61±1.64	Like SC2-ASB0
	SC2-ASB4	72±7%	109±13%	3.07±1.79	Like SC2-ASB0

Table 5-3 Overview of growth rates and CA yields produced by SC6 constructed strains. Heterologously expressed components of the glucose uptake system (GlcP=hexagon, Glk=oval tetramer) is indicated by the glow of the objects, *S. clavuligerus* proteins are shown in green, *S. coelicolor* proteins are shown in blue. The growth rates are shown as percentage of the rate calculated for the SC6-ASB0 cultures in the indicated medium (error indicated shows propagated error according to Gauss). CA yields are shown as μg/ml CDW and of colony morphologies and clearance of agar from milk protein (indicative of protease production).

	Strain	μ h^{-1} TSB 0.5% glucose	μ h^{-1} NMMP 0.5% glucose	CA yield [μg/mg CDW]	Morphology and agar clearance on solid
	SC6-ASB0	100%	100%	13.86±2.06	SC2-like morphology, clearance of agar suppressed by glycerol
	SC6-ASB2	89±5%	46±3%	385.51±50.88	SC6-like morphology, no suppression by glycerol
	SC6-ASB3	137±6%	120±14%	16.05±3.09	Like ASB6-ASB0
	SC6-ASB4	100±8%	58±1%	374.38±17.62	Bald morphology, yellow on minimal agar, crenulated on milk agar, no clearance (BAP20-like morphology on milk agar)

6 Determination of Protein-Protein Interaction Between Glucose Permeases and Glucokinases from *S. clavuligerus* and *S. coelicolor*

In the previous chapter, expression of *sco2126* in SC6 (SC6-ASB4) resulted in significantly increased CA titres produced by this strain, as well as developmental and protease production phenotypes that deviated from the parental strain. In addition, *sclav_4529* expression in BAP20 (BAP20-ASB1) resulted in a developmentally arrested strain. This indicates that non-cognate GlcP and Glk may interact differently than the cognate GlcP and Glk proteins, thus, resulting in morphological and physiological impairment (van Wezel *et al.*, 2007).

To understand the level of interaction between cognate and non-cognate GlcP and Glk of *Streptomyces*, bacterial two hybrid assays (BACTH) were carried out. At the basis of this assay is the catalytic activity of subdomains, T15 and T18, of the adenylate cyclase (CyaA) from *Bordetella pertussis* (Figure 6-1A-D). Adenylate cyclase catalyses the synthesis of cAMP from ATP (Confer *et al.*, 1982). When T25 and T18 are fused to interacting proteins, bringing the subdomains into physical proximity, reassembly of the catalytic domain permits cAMP synthesis (Ladant, 1988; Ladant *et al.*, 1989; Karimova *et al.*, 1998). Once synthesised, cAMP binds the catabolite-activator protein, CAP, a transcriptional regulator that forms a complex with cAMP, and binds cAMP/CAP-dependent promoters, such as the *lac* promoter. The gene located downstream of the *lac* promoter, *lac Z*, encodes a β-galactosidase enzyme, a widely used reporter gene. β-galactosidase catalyses the reaction from *O*-nitrophenol-β-D-galactoside to *O*-nitrophenol (ONP), which absorb at 420 nm, and galactose. The specific enzyme activity of β-galactosidase is measured in Miller units and is calculated as the change of the absorbance at 420 nm, which correlates with the amount of produced ONP, as a function of the reaction time and divided by the volume and absorbance (600 nm) of the bacterial culture (Miller, 1972). The intermembrane GlcP consists of 12 transmembrane helices that are connected by a cytoplasmic loop (Figure 6-1E, see Figure 3-2 for 3D model). The N- and C-terminal ends are predicted to be intracellular. The glucokinase forms an intracellular tetramer in the presence of glucose (Miyazono *et al.*, 2012). To elucidate which parts of the permeases and kinases are involved in potential interactions, N- and C-terminal fusion proteins of GlcP and Glk with T18 and T25 were generated (Karimova *et al.*, 1998).
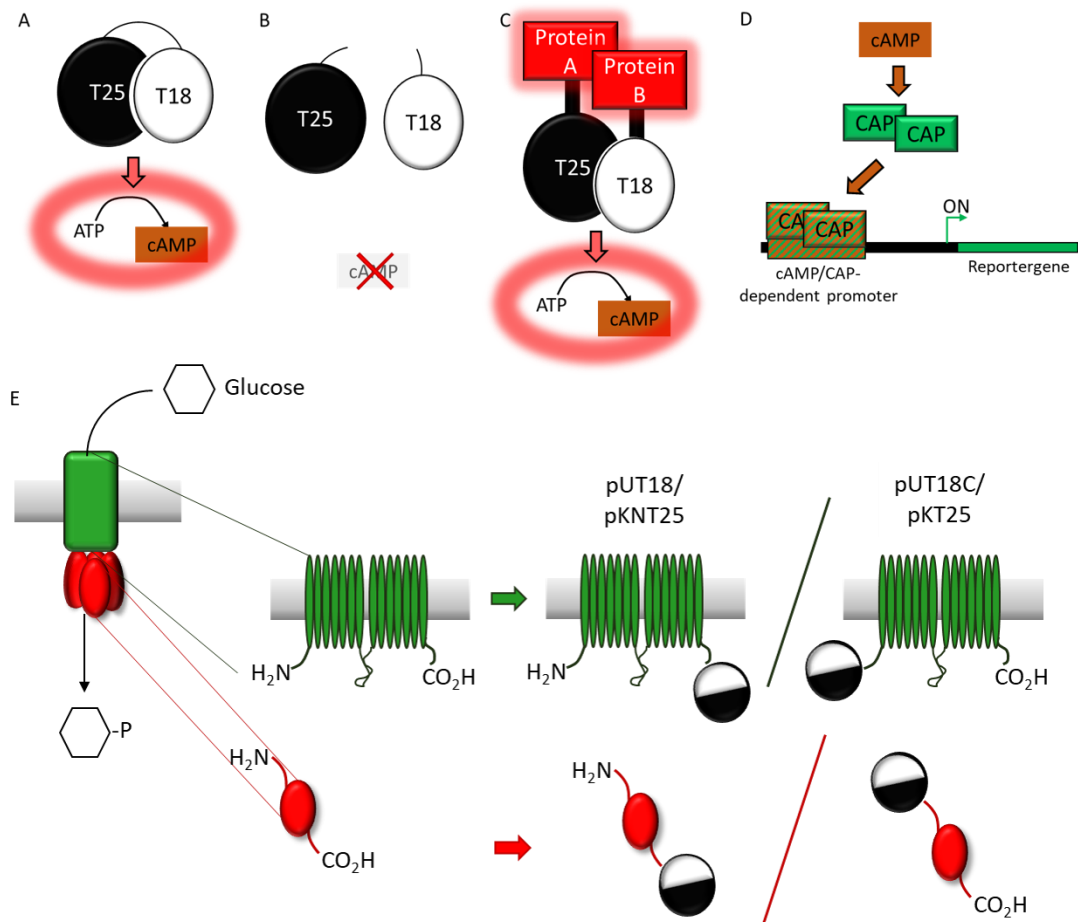
Figure 6-1: (A-D) Schematic of the principle of the bacterial two hybrid and (E) the GlcP- and Glk-fusion proteins with T25 and T18. (A) The T25 and T18 proteins form the catalytic domain of the adenylate cyclase, which generates cAMP from ATP. (B) T25 and T18 are unable to assemble and no cAMP is generated. (C) Fusion of the subdomains to interacting proteins (Proteins A and B) brings them into physical proximity. Reassembly occurs, and cAMP is formed. (D) CAP binds cAMP/CAP-dependent promoters in the presence of cAMP and activates reporter gene expression, the activity of which is measurable. (E) Here, GlcP (green) and Glk (red) were fused to T25 (black) or T18 (white) either C- or N-terminally. GlcP is composed of 12 transmembrane helices connected by a cytoplasmic loop. The N- and C-terminal ends of GlcP1 (SCO5578) are located intracellularly. The Glk monomers were fused to one of the subdomains at either end.

The cloning procedure included *glcP* and *glk* genes from *S. clavuligerus* and *S. coelicolor*. Thus, plasmids were cloned containing *sclav_4529*, *sclav_1340*, *sco5578* and *sco2126*. Vector backbones were pUT18, pUT18C (Figure 6-2), pKT25 and pKNT25 (Figure 6-3), the former two containing the T18 sequence, the latter two the T25 sequence. Multiple cloning sites (MCS) on the vectors are located upstream and downstream of the respective T genes. Restriction enzyme recognition sequences flanking the T genes are *Xba*I, *EcoR*I and *BamH*I. These enzymes were used to linearise the target vectors by double digests with either *Xba*I and *EcoR*I or *Xba*I and *BamH*I, which subsequently facilitated unidirectional insertion of *glcP* and *glk*. Since some of the genes of interest contain intrinsic *EcoR*I or *BamH*I sites, this directed the combination of restriction enzymes that were used to clone the genes onto the vector.

The inserts were amplified from gDNA using oligonucleotides that contained required restriction sites, which allowed modification of the ends of the PCR products to make them compatible with the ends of the vectors (Figure 6-4). To maintain a single open reading frame, two different sets of oligonucleotides were required per gene, one set for cloning the gene upstream and one set for cloning the gene downstream of the respective T gene (Figure 6-4A). A touchdown PCR approach was taken for gene amplification, which produced products of the correct sizes for *glcP* (1.5 kb) and *glk* (1.0 kb) from *S. clavuligerus* and *S. coelicolor* (Figure 6-4B).

The presence of the inserts was confirmed by restriction digest using either *Xba*I and *BamH*I or *Xba*I and *EcoR*I (Figure 6-5). Analogous to previous cloning of *glcP* and *glk* into pIJ6902, the plasmids containing *sclav_4529*, *sclav_1340*, *sco5578*, or *sco2126* were named ASB1, ASB2, ASB3 and ASB4. Unfortunately, no positive clones for pKNT25 carrying inserts were detected after the first round of transformations. Due to time constraints, transformations with these plasmids were not repeated. However, pUT18, pUT18C and pKT25 plasmids containing the *glcP* and *glk* genes from *S. clavuligerus* and *S. coelicolor* were successfully cloned and were used for subsequent BACTH assays.
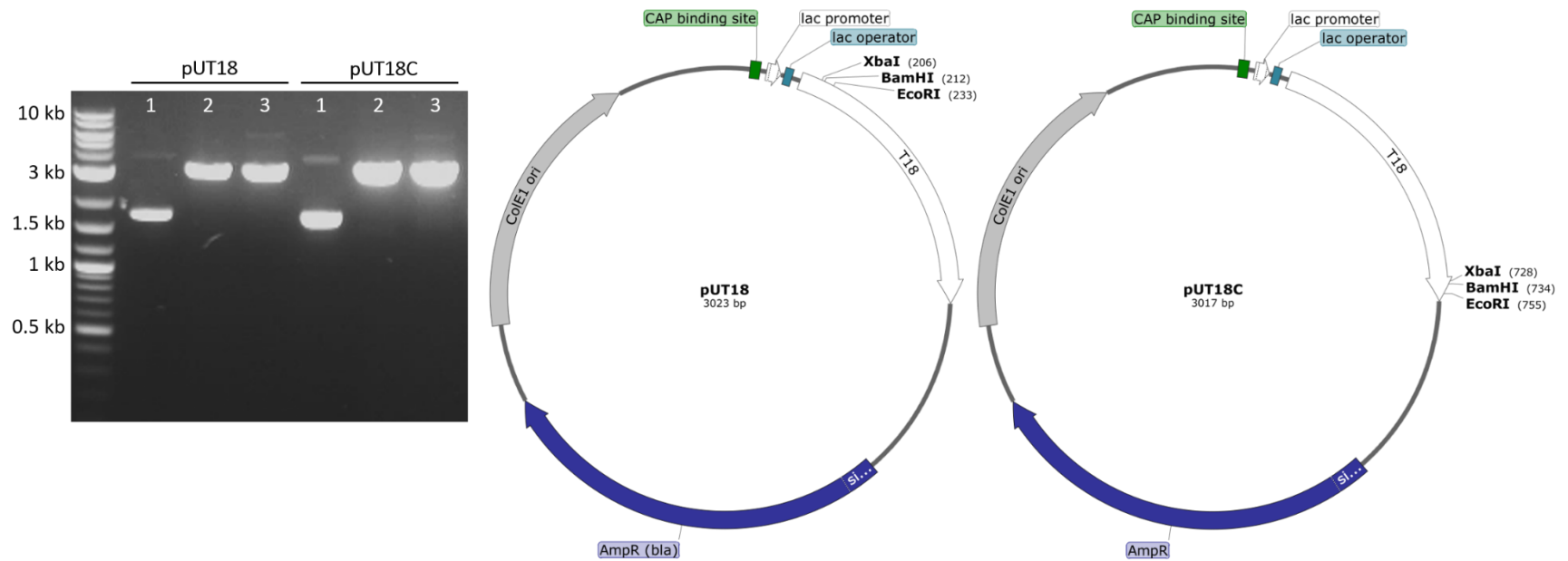
Figure 6-2: Agarose gel (1.0%) showing bands resulting from running undigested (1) or *Xba*I/*EcoR*I (2) or *Xba*I/*BamH*I (3) digested pUT18 and pUT18C and corresponding plasmids maps. Two bands are visible for undigested plasmid, the top and bright band corresponding to nicked and supercoiled plasmid configurations. Bands that resulted from double digests with either of the combination of restriction enzymes corresponded to a size of 3 kb. Features highlighted on the maps are the CAP binding site (green), the lac promoter (white) and operator sequences (light blue), the β-lactamase gene (dark blue) of which the gene product confers resistance to ampicillin, as well as the ColE1 high copy number origin of replication (grey).
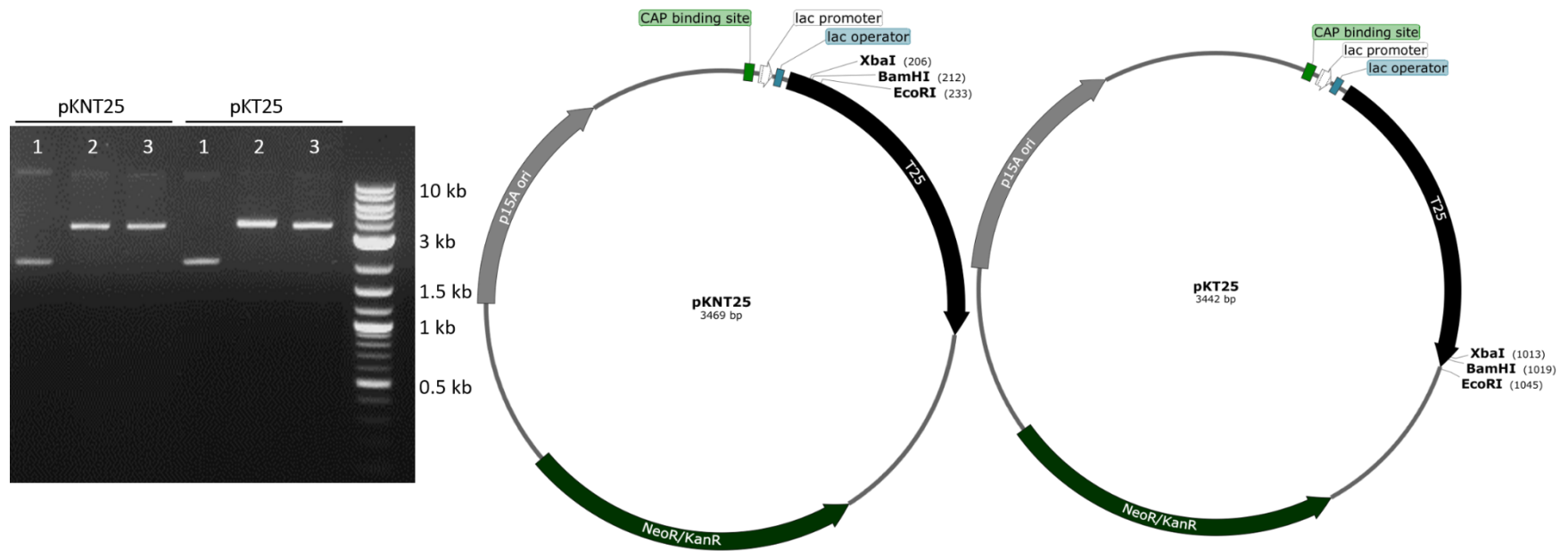
Figure 6-3: Agarose gel (1.0%) showing bands resulting from restriction digest of pKNT25 and pKT25 plasmids using *Xba*I and *EcoR*I (2) or *Xba*I and *BamH*I (3) and corresponding plasmid maps. The bright bands visible in lanes 2 and 3 correspond to the size of linearised pKNT25 and pKT25 plasmids, which is 3.4 kb. Features highlighted on the maps are the T25 sequence (black), the neomycin/kanamycin resistance conferring gene aminoglycoside transferase (dark green), the low-copy number origin of replication p15A (grey), as well as the CAP binding site (light green), the lac promoter (white) and operator (blue) sequence.
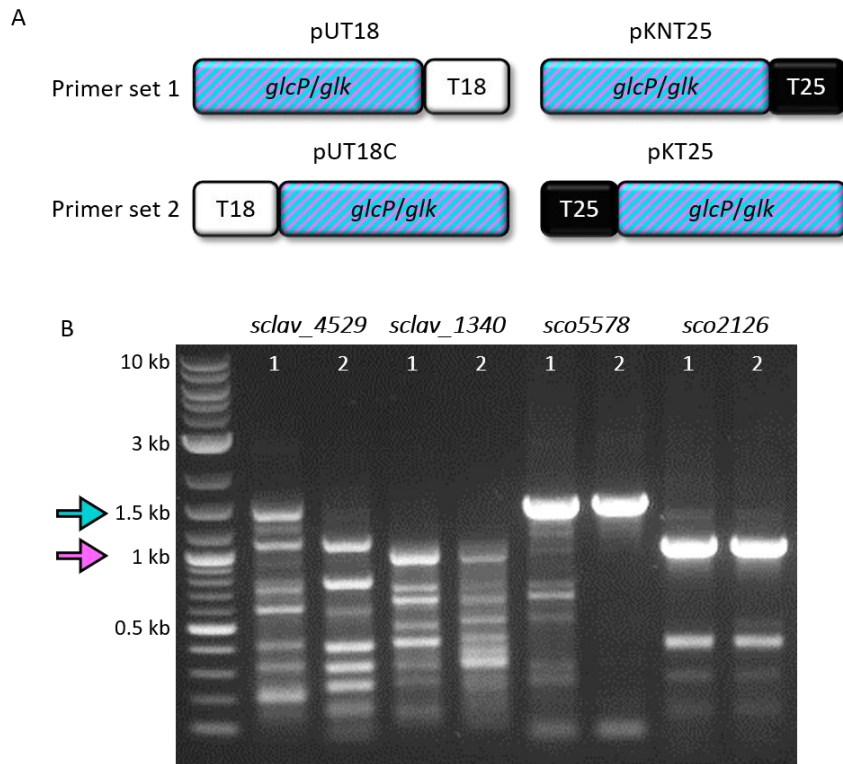
Figure 6-4: Schematic showing the positions of the T25 and T18 genes relative to cloned genes and amplified PCR products. (A) Two different sets of primers (1 and 2) required for in-frame cloning with sequence encoding T fragments. (B) Agarose gel (1.0%) showing PCR products of *glcP* and *glk* from *S. clavuligerus* (*sclav_4529* and *sclav_1340*), *S. coelicolor* (*sco5578* and *sco2126*).
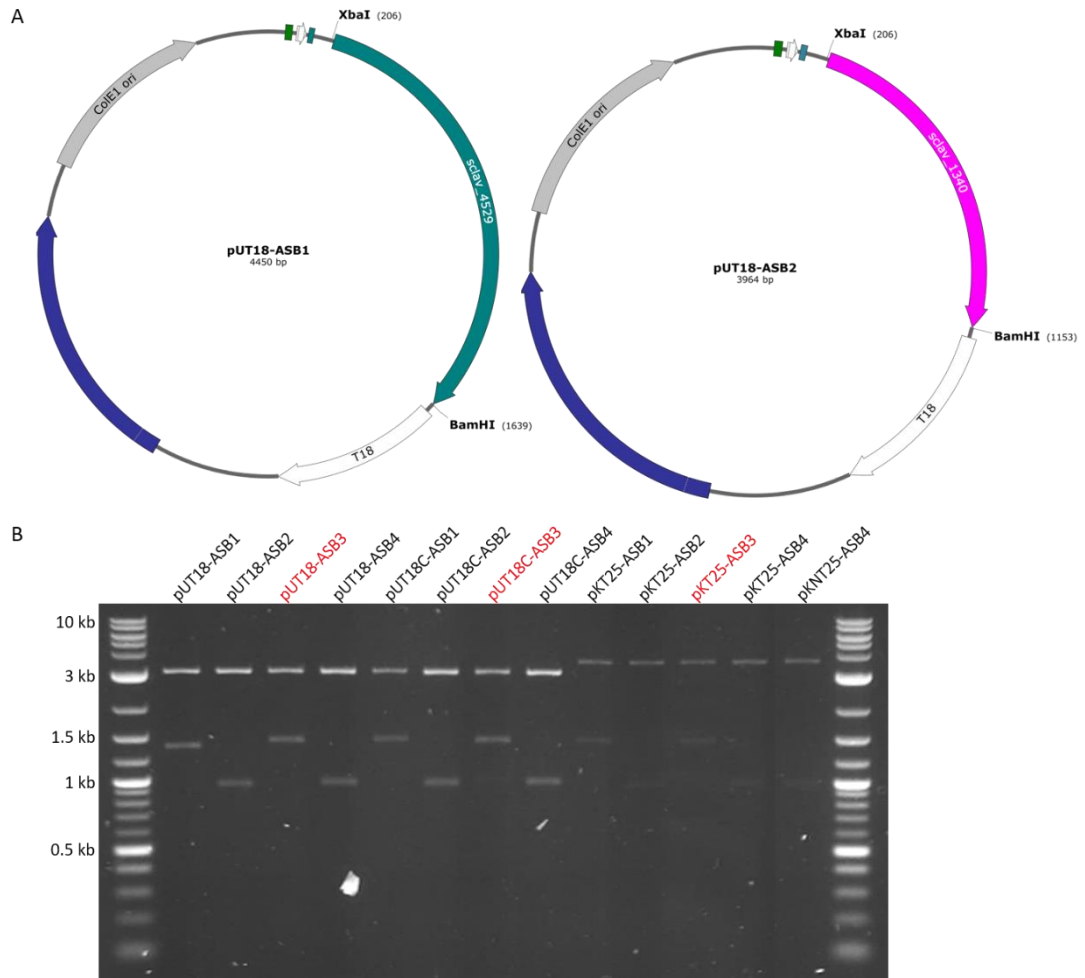
Figure 6-5: BACTH plasmids pUT18-ASB1 and pUT18-ASB2. (A) Plasmid maps of the *in silico* cloned plasmids. The *glcP* and *glk* inserts are highlighted in turquoise and pink, respectively. The maps are shown as representative maps for all constructed plasmids. (B) Agarose gel (1.0%) showing bands resulting from restriction digest of the constructed plasmids. The smaller bands correspond to the inserted *glcP* and *glk* genes. S*co2126* was cloned using *Xba*I and *EcoR*I (red).

6.2   Qualitative Interaction of GlcP and Glk

The phenotype produced as the result of protein-protein interaction in a BACTH assay is caused by the enzymatic activity of the β-galactosidase enzyme, which, in the presence of the substrate, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), turns co-transformed colonies blue. Due to the lack of pKNT25 constructs, co-transformations were carried out using pUT18, pUT18C and pKT25 constructs along with pUT18C-zip and pKT25-zip as positive controls (Figure 6-6). pUT18C-zip and pKT25-zip contain complementary leucine zipper sequences that facilitate T18 and T25 (Karimova *et al.*, 1998). Co-transforming with the empty vectors served as negative control. Competent *E. coli* (BTH101) cells were co-transformed with derivatives of pUT18C and pKT25 (Figure 6-7).

While most colonies were either unambiguously white or blue on most plates, there were mixed populations of blue and white colonies on several transformation plates. The blue colonies on these plates may have arisen from spontaneous Lac$^+$ revertants, which occur at a frequency of 10^{-8} in BTH101 (Karimova *et al.*, 2005). Six individual colonies were selected from each co-transformation plate and re-streaked onto larger plates, a representative colony is shown for every co-transformation (Figure 6-8). As expected, cells obtained from co-transforming with pUT18C-zip and pKT25-zip were blue and colonies containing the empty vectors were white. Most co-transformations resulted in the colonies being white. However, this was not the case for *E. coli* that had been co-transformed with complementary plasmids (T18 and T25) that contained either of the *glcP* genes, *sclav_4529* or *sco5578*, providing the respective T sequences were fused to the N-terminal end of the permeases (see T18-SCLAV_4529 and T25-SCLAV_4529; T18-SCO5578 and T25-SCLAV_4529; T18-SCLAV_4529 and T25-SCO5578; T18-SCO5578 and T25-SCO5578). This suggests GlcP from different species might be able to physically interact in a heterologous host. Given the necessity of the N-terminal location of the T proteins both for intra- and inter-species GlcP interactions, this indicates that the nature of the interaction between cognate GlcP and non-cognate GlcP might be similar.
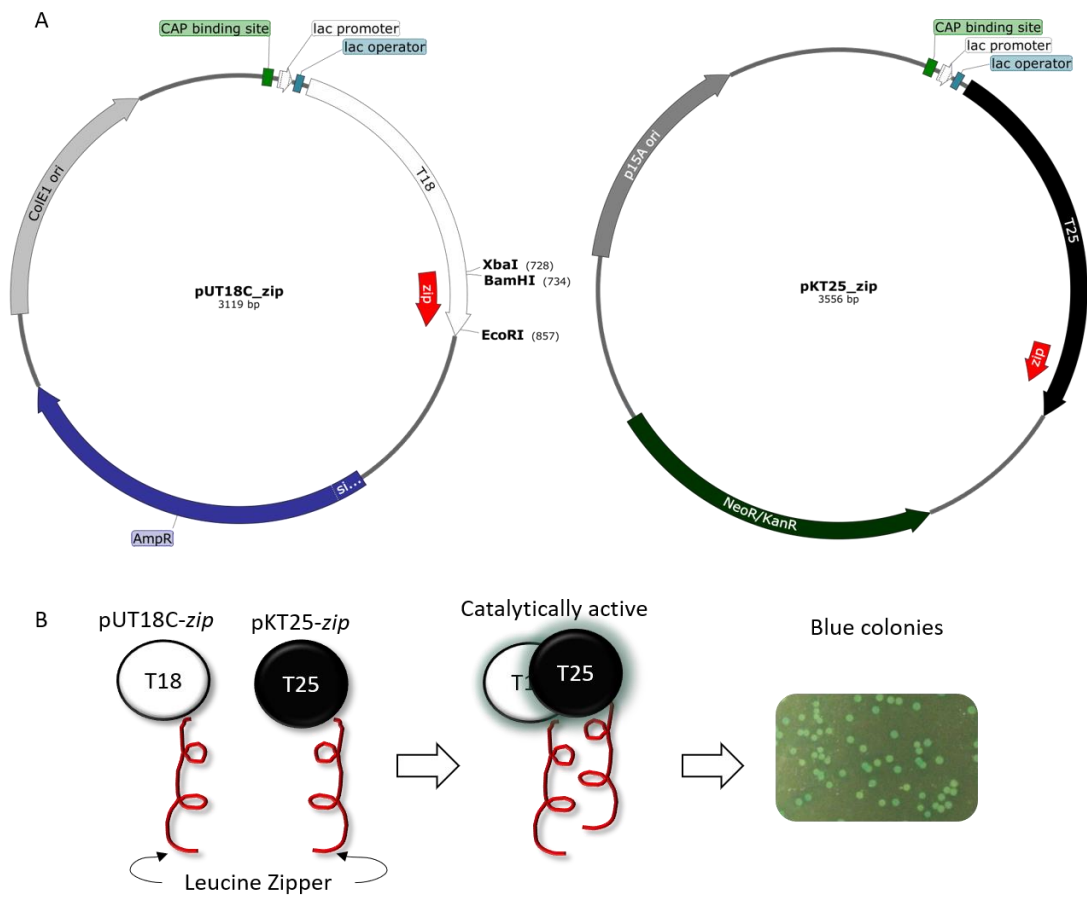
Figure 6-6: (A) Positive control plasmids, pUT18C-zip and pKT25-zip vector maps and schematic showing the leucine zipper facilitating physical proximity of T18 and T25. The leucine zipper sequences are highlighted in red in the plasmid maps. (B) The leucine zipper sequences direct and facilitate assembly of the T18-T25 fusion protein, which becomes catalytically active. In the presence of X-gal, co-transformed colonies of *E. coli* turn blue due to β-galactosidase activity.
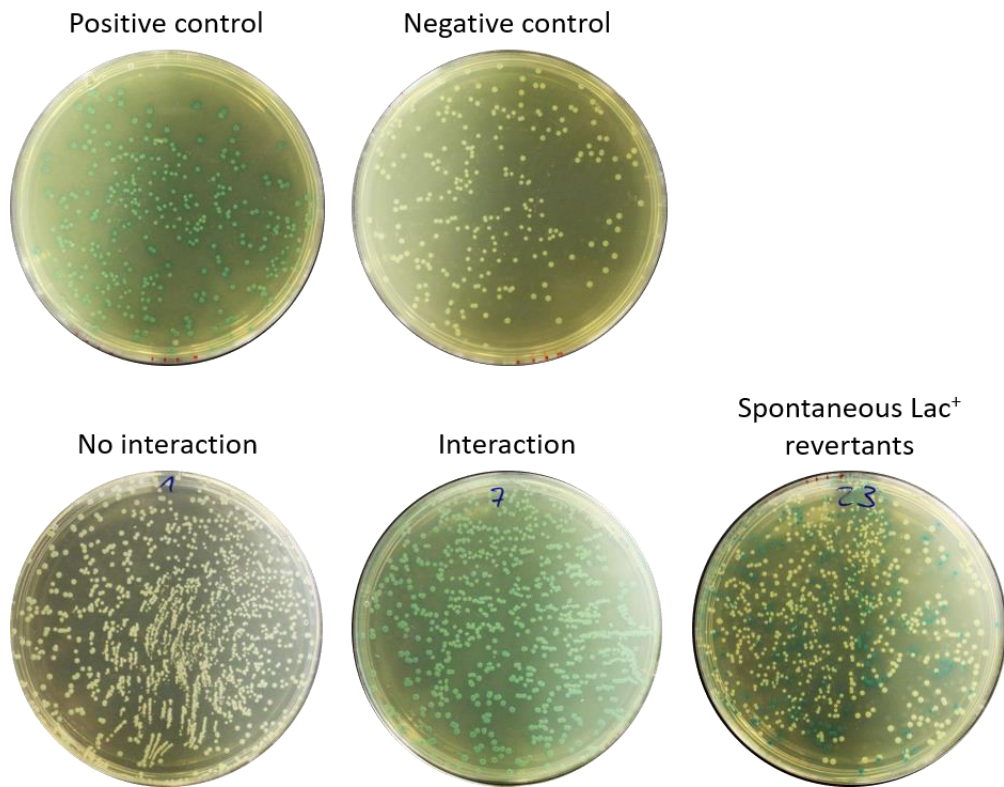
Figure 6-7: Co-transformation plates showing blue, white, and mixed populations. *E. coli* cells co-transformed with pUT18C-zip and pKT25-zip were blue (positive control). Co-transformants of pUT18 and pKT25 were white. Blue colonies in the mixed populations might be spontaneous Lac$^+$ revertants. This mutation occurs at a frequency of 10^{-8} in BTH101 *E. coli*.
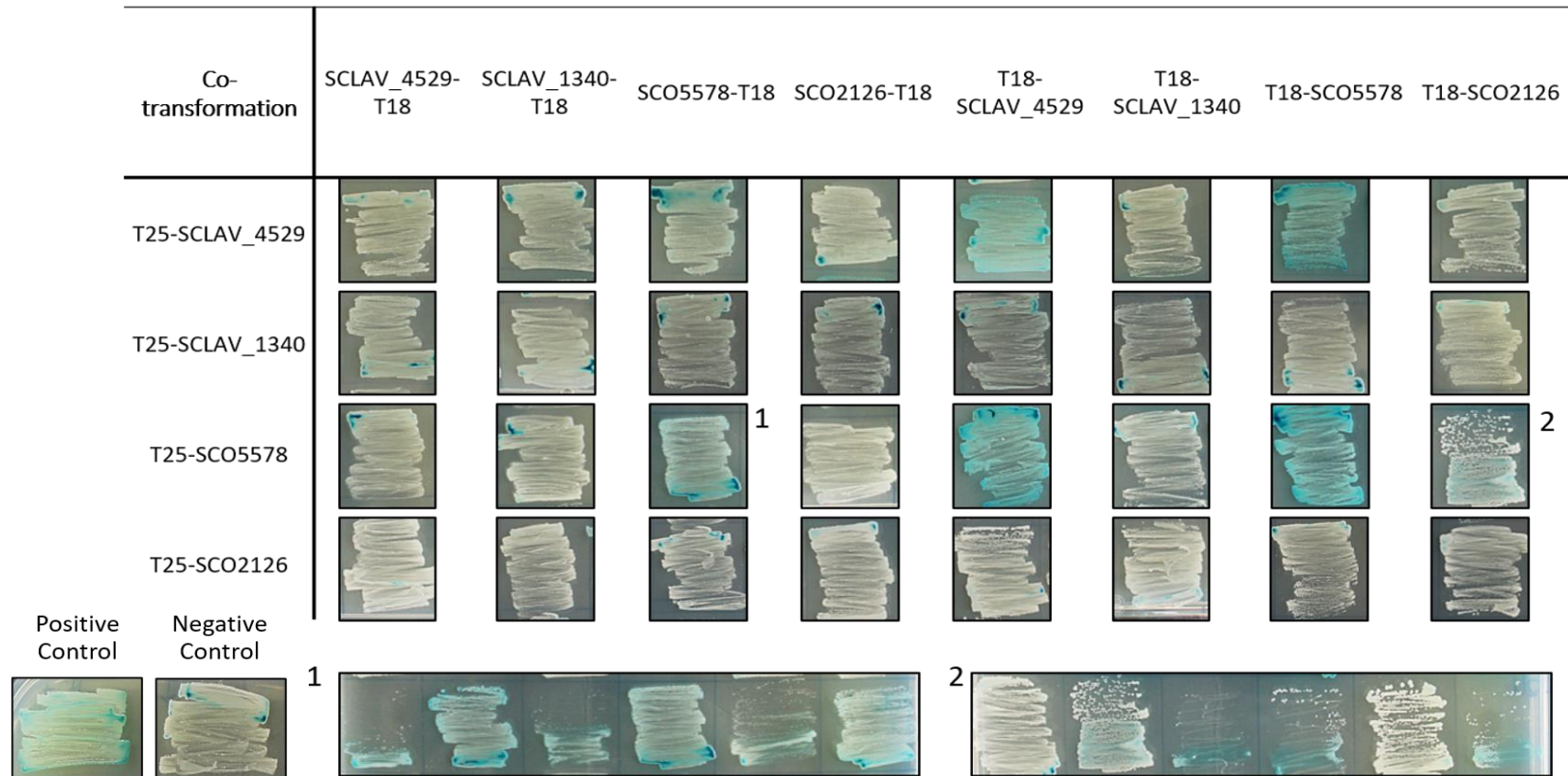
Figure 6-8: Blue and white colonies obtained from qualitative BACTH assays. Colonies for the positive and negative controls are shown in the bottom left. Potential interaction partners are listed in the top row and first column of the table. Blue colonies are visible from interaction between SCLAV_4529 and SCO5578, as well as SCO5578 and SCO2126. (1) Light blue colonies emerged from co-transformants of T25-SCO5578 and SCO5578-T18. (2) Four of six colonies expressing the T25-SCO5578 and T18-SCO2126 constructs were light blue.

Aside from these interactions, two other potential interactions were observed based on the colour of co-transformed colonies: SCO5578-T18 and T25-SCO5578, another GlcP-GlcP interaction, however, here T18, is fused to SCO5578 at the C-terminal, as well as T18-SCO2126 and T25-SCO5578. In the former case, colonies were distinctly less blue in comparison to the previously described interactions (Figure 6-8, 1) and, in the latter case, only four of the six re-streaked colonies from the original transformation plate appeared blue (Figure 6-8, 2). Interestingly, aside from the T18-SCO2126 and T25-SCO5578 interaction, no other GlcP-Glk co-transformations produced blue colonies nor did any Glk-Glk co-transformations.

Glucose uptake via the GlcP/Glk system and CCR mediated by GlkA is known to require direct interaction of the two proteins in *Streptomyces* and Glk has been shown to form homotetramers (Mahr *et al.*, 2000; Imriskova *et al.*, 2005; van Wezel *et al.*, 2007). The lack of observed GlcP-Glk, as well as Glk-Glk interactions might be due to several factors: i) the interaction might not be taking place in the heterologous host, *E. coli*, ii) these interactions may require C-terminally fused T sequences, iii) partial interactions may be less apparent on agar plates or transient in nature.

6.3   QUANTIFICATION OF GLCP AND GLK INTERACTIONS

Quantitative β-galactosidase activity assays were carried out to identify partial interactions, as well as to confirm interactions previously observed on X-Gal containing agar plates. The assays were conducted using co-transformed *E. coli* cultures in biological and technical duplicate (Griffith *et al.*, 2002). The quantification of β-galactosidase activity confirmed the interactions previously observed between SCLAV_4529 and SCO5578 when T18 or T25 are fused to the permeases N-terminally (Figure 6-9 and Table 6-1). The mean activities (shown as Miller units [MU]) measured in these co-transformed cultures were 1508.01±165.25 MU, 1461.75±388.11 MU, 444.36±679.42 MU and 2193.47±282.94 MU with the interaction between T18-SCO5578 and T25-SCO5578 producing the highest β-galactosidase activity resulting from a GlcP-GlcP interaction. This interaction was comparable to the activity measured in positive control cultures (2975.31±769.44 MU).
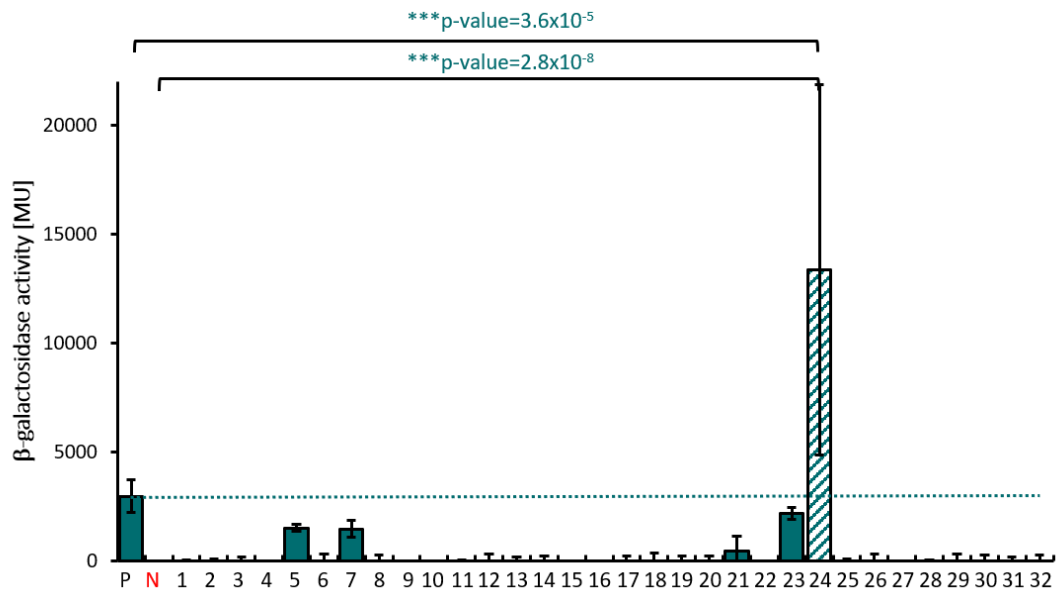
Figure 6-9: Mean β-galactosidase activity from liquid BACTH *E. coli* cultures (see Table 6-1). Activity was measured after 3 and 10 minutes, samples were assayed in biological and technical duplicate. Error bars indicate the standard error across samples. Differences in activities across all cultures were significant (***p-value=0.0001, one-way ANOVA). A significantly higher activity compared to the controls was measured from an interaction between T18-SCO2126 and T25-SCO5578 (Tukey's HSD test).

Table 6-1: Mean β-galactosidase activity measured from co-transformation cultures. Mean activities are shown in Miller units [MU], the standard error for biological and technical duplicates is shown. The highlighted row shows the interactions that produced a significantly higher activity than the controls (see Figure 6-9).

	Protein I	Protein II	Mean activity [MU]	Error
Positive control	T18-zip	T25-zip	2975	769
Negative control	NA	NA	0	48
1	SCLAV_4529-T18	T25-SCLAV_4529	0	35
2	SCLAV_1340-T18	T25-SCLAV_4529	0	95
3	SCO5578-T18	T25-SCLAV_4529	0	154
4	SCO2126-T18	T25-SCLAV_4529	0	1
5	T18-SCLAV_4529	T25-SCLAV_4529	1508	165
6	T18-SCLAV_1340	T25-SCLAV_4529	0	88
7	T18-SCO5578	T25-SCLAV_4529	1462	388
8	T18-SCO2126	T25-SCLAV_4529	0	89
9	SCLAV_4529-T18	T25-SCLAV_1340	0	95
10	SCLAV_1340-T18	T25-SCLAV_1340	0	56
11	SCO5578-T18	T25-SCLAV_1340	0	105
12	SCO2126-T18	T25-SCLAV_1340	0	85
13	T18-SCLAV_4529	T25-SCLAV_1340	0	66
14	T18-SCLAV_1340	T25-SCLAV_1340	0	66
15	T18-SCO5578	T25-SCLAV_1340	0	151
16	T18-SCO2126	T25-SCLAV_1340	0	3
17	SCLAV_4529-T18	T25-SCO5578	0	114
18	SCLAV_1340-T18	T25-SCO5578	0	95
19	SCO5578-T18	T25-SCO5578	0	74
20	SCO2126-T18	T25-SCO5578	0	117
21	T18-SCLAV_4529	T25-SCO5578	444	369
22	T18-SCLAV_1340	T25-SCO5578	0	42
23	T18-SCO5578	T25-SCO5578	2193	283
24	T18-SCO2126	T25-SCO5578	13378	8348
25	SCLAV_4529-T18	T25-SCO2126	0	453
26	SCLAV_1340-T18	T25-SCO2126	0	88
27	SCO5578-T18	T25-SCO2126	0	94
28	SCO2126-T18	T25-SCO2126	0	66
29	T18-SCLAV_4529	T25-SCO2126	0	87
30	T18-SCLAV_1340	T25-SCO2126	0	77
31	T18-SCO5578	T25-SCO2126	0	46
32	T18-SCO2126	T25-SCO2126	0	87

Although interactions were measured between T18-SCLAV_4529 with T25-SCLAV_4529, T18-SCLAV_4529 with T25-SCLAV_4529 and T18-SCO5578 with T25-SCO5578, as well as T18-SCLAV_4529 and T25-SCO5578, activities were not significantly higher compared to the negative control. However, the difference between the positive and negative control in the assay did also not produce a statistically significant difference in the statistical analysis (ANOVA and *post hoc* Tukey's HSD test). Sinilarly, the difference in activity between control cultures and co-transformed cultures of T18-SCLAV_4529 and T25-SCO5578, however, was not significant, inferring that the interaction between GlcPs from *S. clavuligerus* and *S. coelicolor* is not significantly different from the negative control. Yet, this is likely due to the large variation in data points, which, in turn, may be caused by the nature of the interaction being transient, especially in the absence of the substrate glucose. Nevertheless, these data confirm the inter- and intraspecies interactions of GlcPs detected on the agar plates and support the observation that GlcP proteins interact via their N-terminal ends. A schematic summary of these findings is shown in Figure 6-10.

While no activity for the interaction between SCO5578-T18 and T25-SCO5578 was measured despite colonies appearing light blue on agar plates and highlighting the dependence of GlcP-GlcP interaction on N-terminally fused T sequences, the interaction of T18-SCO2126 with T25-SCO5578 produced a β-galactosidase activity of 13377.85±8498.04 MU. This was significant with p-values for this interaction compared to the positive and negative control cultures of 3.6×10^{-5} and 2.8×10^{-8}, respectively. No further GlcP-Glk or Glk-Glk interactions were detected. In the latter case, it is not unlikely that this might be due to the interaction being measured in *E. coli*. However, given that an interaction was measured between GlcP1 and GlkA from *S. coelicolor* but not between the homologous proteins from *S. clavuligerus*, this may suggest that SCLAV_4529 and SCLAV_1340 do not physically interact under the tested conditions, or that the nature of their interaction might be transient and, therefore, not easily detectable by BACTH.
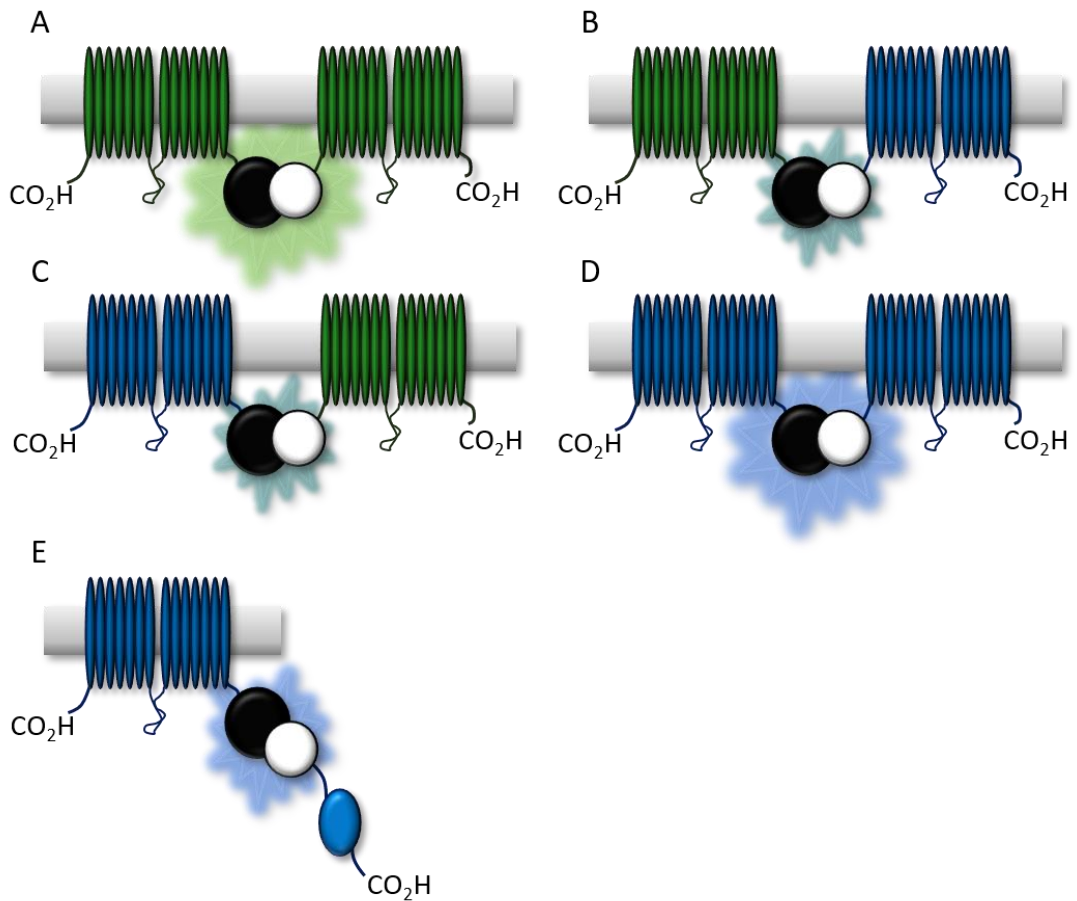
Figure 6-10: Schematic summarising the results obtained from quantitative and qualitative BACTH. (A) The glucose permease from *S. clavuligerus*, SCLAV_4529, (green) interacts with itself. (B) It also interacts with the GlcP from *S. coelicolor*, SCO5578 (blue), providing the permeases are fused to T sequences N-terminally. Higher β-galactosidase activities were measured from SCLAV_4529 interacting with itself than with SCO5578. (C) The reverse situation of constructs, with SCO5578 and SCLAV_4529 fused to T25 and T18, respectively, is depicted. (D) SCO5578 interacting with itself is shown. (E) A GlcP-Glk interaction was detected between SCO5578 and SCO2126.

Investigating the interactions of GlcP and Glk proteins from *S. clavuligerus* and *S. coelicolor* has revealed the capability of GlcP to interact with GlcP from another *Streptomyces* species. Both intra- and interspecies interactions were observed that appear to be dependent on the physical proximity of the N-terminal ends of the permeases. No GlcP-GlcP interaction was detected when the T sequence of one of the potential interaction partners was fused to the C-terminal end (Figure 6-10).

The only GlcP-Glk interaction detected was between SCO5578 and SCO2126, but not SCLAV_4529 and SCLAV_1340 (Figure 6-8, Figure 6-9). GlcP-Glk interaction is required for glucose uptake and phosphorylation by this system (van Wezel *et al.*, 2007). Thus far, glucose auxotrophy in *S. clauvligerus* has been thought to be caused by low expression of *glcP* (Pérez-Redondo *et al.*, 2010), yet these data suggest that lack of protein-protein interaction between GlcP and Glk might also play a role. This might, further, explain the fact that glucose does not trigger repression of antibiotic production in WT *S. clavuligerus* (Aharonowitz *et al.*, 1978; Lebrihi *et al.*, 1988). No interactions between Glk proteins was detected, although Glk has been demonstrated to form homotetramers (Imriskova *et al.*, 2005). This suggests either that the constructs used to test Glk-Glk interaction may require T sequences to be fused to the N-terminal end of the proteins, which could be easily tested by completing the cloning of pKNT25 constructs and including them in the analysis. Alternatively, it might indicate that *E. coli* is not a suitable host for testing Glk-Glk interaction as this might require additional factors. Although GlkA (*S. coelicolor*) homotetramers are stable, tetramers formed by Glk from *S. peucetius* var. *caesius* dissociates easily, a process that is prevented by glucose (Imriskova *et al.*, 2005). Therefore, simply the absence of glucose in the culture medium might have prevented detecting Glk-Glk interaction.

7 DISCUSSION

Understanding central carbon metabolic function in industrial organisms, including substrate uptake and precursor supply in industrially relevant organisms in essential for. improving industrial process efficiency. It is also important for generating sustainable processes for industrial metabolite production.

7.1 GENE DUPLICATION AND HGT OF CENTRAL CARBON METABOLISM GENE FAMILIES IN *STREPTOMYCES*

Gene family expansion through gene duplication and HGT is known to contribute to genome diversification in *Streptomyces* (Ventura *et al.*, 2007; Zhou *et al.*, 2012). Gene duplication can lead to neofunctionalization of the duplicated gene, thereby expanding the enzymatic repertoire of this gene family. However, non-functionalisation may also occur, due to functional redundancy (Lynch *et al.*, 2000; Zhang, 2003; Ventura *et al.*, 2007). A duplicated gene may then undergo pseudogenization, which is the process of accumulation of mutations caused by a lack of selection pressure. The pseudogene is either not expressed or functionless (Zhang, 2003).

The Actinobacteria-wide phylogenetic trees of GlcP (Figure 3-7) and Glk (Figure 3-8) suggest that gene duplication has contributed to gene family expansion, an example for which are the two *glcP* genes in *S. coelicolor.* While *glcP1* encodes the major glucose permease in this species, *glcP2* is not expressed, although GlcP1 and GlcP2 have identical amino acid sequences (van Wezel *et al.*, 2005). Aside from *S. coelicolor*, duplicated *glcP* and *glk* genes appear to be present in multiple Actinobacteria represented in the ActDES database (Schniete *et al.*, 2021). *S. ghanaensis* ATCC_14672, for instance, has three GlcPs, which share between 68% and 89% of their amino acid sequence. In contrast to GlcP2 from *S. coelicolor*, the additional GlcPs in *S. ghanaensis* might have gained novel enzymatic functions or kinetics which might be driving their diversification, as it has been observed for pyruvate kinases (Pyk) in *S. coelicolor*. The Pyk enzymes in this species, encoded by *pyk1* and *pyk2*, are 69% identical and differ in their enzyme activities and kinetics. Moreover, a *pyk1* transposon insertion mutant overproduced actinorhodin under certain conditions, while this was not observed for the *pyk2* transposon insertion mutant (Schniete *et al.*, 2018). This highlights the importance

of understanding gene family expansion of central carbon metabolic genes for specialised metabolite production.

While gene duplication can lead to a higher gene dosage within gene families, HGT is the primary route to the acquisition of new genes (Ochman *et al.*, 2000). Evidence has been presented that suggests HGT might be the main driver of gene family expansion in bacteria (Treangen *et al.*, 2011). HGT within Actinobacterial GlcPs is showcased by GlcP from *Pilimelia anulata* (Micromonosporales) and *Nocardia asteroides* (Corynebacteriales). These permeases have identical amino acid sequences and form a stable clade in the phylogenetic tree, despite the host organisms belonging to phylogenetically distant orders. The species, further, have distinct lifestyles. Although both occur in the soil, *Pilimelia* is known for its ability to degrade keratin and *Nocarida asteroides* causes nocardiosis in humans (Beaman *et al.*, 1994; Vobis *et al.*, 2015).

Interestingly, strains represented in group I and group II in the Glk phylogenetic tree, do not necessarily also have two GlcPs. For instance, *S. aureocirculatus* has two Glk enzymes, which share 50% of their amino acid sequence, yet it has only one GlcP protein. This indicates that, although GlcP and Glk form a glucose transporting unit, they may not be under the same functional constraint. They might, thus, evolve as separate proteins, contributing to lineage-specific expansion, which is associated with an increase in adaptability to an ecological niche (Zhou *et al.*, 2012).

The observation that multiple GlcP and Glk proteins are present primarily among members of the *Streptomyces* genus might be a result of there being 290 streptomycete genomes represented in a database consistent of a total of 612 genomes (Schniete *et al.*, 2021).

7.2 GENOME EXPANSION IN *S. CLAVULIGERUS*

The *S. clavuligerus* genome has been shaped by gene family expansion events of central carbon and secondary metabolism. For instance, while it has a single copy of *glcP*, multiple glycerol permease genes (*glpF1* and *glpF2*) are encoded on the chromosome, the gene products of which are 56% identical (Minambres *et al.*, 1992; Baños *et al.*, 2009; Medema *et al.*, 2010; Song *et al.*, 2010; Cao *et al.*, 2016). The two permeases are part of two distinct glycerol uptake systems (a minor constitutive one and a major inducible one). These allow *S.*

clavuligerus to "sense" glycerol in its environment and express the major uptake system in response (Figure 5-20)(Minambres *et al.*, 1992). A clear example of HGT of biosynthetic gene clusters (BGCs) is illustrated by two copies of biosynthetic genes for a putative antimicrobial peptide identified on the ATCC 27064 genome. These are thought to have originated from the Hymenoptera order of insects (Ayala-Ruano *et al.*, 2019). Also, a tacrolimus biosynthetic gene cluster-bearing strain *S. clavuligerus* strain (CKD 1119) was identified in 2016. Tacrolimus is an immunosuppressive polyketide, commonly administered in the clinic to prevent graft rejection (Kino *et al.*, 1987). It is naturally produced by *S. tsukubaensis* and not *S. clavuligerus* (Wu *et al.*, 1993; Kim *et al.*, 2007; Medema *et al.*, 2010; Song *et al.*, 2010; Cao *et al.*, 2016).

S. clavuligerus is not the only organism that produces CA and cephamycin C, as *S. jumonjinensis* and *S. katsurahamanus* also possess the corresponding biosynthetic gene clusters (BGCs), for which whole genome sequences were recently made available (Ward *et al.*, 1993; Jensen, 2012; AbuSara *et al.*, 2019). The presence of CA BGC-like genes in *S. pratensis* ATCC 33331 (formerly *S. flavogriseus*) and *Saccharomonospora viridis* DSM 43017 (Jensen, 2012; Álvarez-Álvarez *et al.*, 2013) and the presence of additional genes within the CA and cephamycin C BGCs that do not appear to be involved in biosynthesis of the metabolites, raises the question of origin and distribution of the CA and cephamycin C BGC among and beyond streptomycetes. This has recently been addressed by carrying out a phylogenomic analysis of CA and cephamycin BGCs in producers, such as *S. clavuligerus*, and non-producers, including *S. antibioticus* Tü1718 (Goomeshi Nobary *et al.*, 2012; Chevrette *et al.*, 2020). The study revealed that CA and cephamycin C biosynthetic genes are widespread among *Streptomyces*, the analysis included 25 strains. Moreover, it was concluded that chemical diversity within an organism can arise from crosstalk and interplay between BGCs and subclusters, which is the case for *S. clavuligerus* (Chevrette *et al.*, 2020).

These examples highlight that genome expansion has occurred in *S. clavuligerus*. The single *glcP* gene and its low expression might, thus, reflect minor role glucose plays in the metabolic network and natural ecology of *S. clavuligerus*.

Natural glucose auxotrophy of WT *S. clavuligerus* has been attributed to the low expression of the native *glcP* gene, *sclav_4529* (Pérez-Redondo et al., 2010). Indeed, the commonly used promoter sequence prediction tool BProm failed to predict such a sequence upstream of *sclav_4529* that resembled the consensus sequence determined for the group of analysed streptomycetes sequences (Figure 3-13). However, a recently published analysis of promoters and transcription start sites (TSS) in *S. clavuligerus* has demonstrated a general lack of knowledge about the structure and nature of such sequences in this species (Hwang et al., 2019). Promoter sequences and TSS in streptomycetes are often fundamentally different than in other bacteria, such as *E. coli* and are, therefore, not easily predicted using available bioinformatic tools. Although genome-scale analysis of promoter sequences and TSS for *S. coelicolor* and *S. clavuligerus* have been published (Jeong *et al.*, 2016; Hwang *et al.*, 2019), the focus of these studies was to provide tools to "unlock" cryptic BGCs of secondary metabolites and not to understand expression of genes involved in central carbon metabolism. However, RNA sequencing of *S. clavuligerus* growing during exponential, transition, late exponential, and stationary phase showed that *glcP* is expressed continuously throughout cultivation in liquid, albeit at a low level (Hwang *et al.*, 2019, supplementary data). This agrees with the *glcP* transcription observed in DSM 738, SC2 and SC6 TSB-grown cultures (Figure 4-7).

SC6 appears to express *glcP* at an elevated level compared to DSM 738 and SC2 (Figure 4-7), although the semi-quantitative RT-PCR should be repeated, and RNA sequencing analysis of SC6 suggested that *glcP* expression was downregulated during CA biosynthesis when *ccaR* expression was upregulated (Table 4-2). The RNA sequencing data presented here was obtained from two different timepoint, 52 and 65 hours, and the increase in *ccaR* expression was not significant. Neither was the decrease in *glcP* expression. The experiment should, therefore, be repeated choosing multiple timepoints for extraction of RNA when the culture is in distinct phases of growth. Nevertheless, these data indicate transcriptional repression of *glcP* may occur *ccaR* is expressed at higher levels. In fact, this is supported by transcriptomic data obtained from a *ccaR* deletion mutant, in which *glcP* expression was significantly upregulated during the exponential growth phase compared to the WT (Álvarez-Álvarez *et al.*, 2014). Furthermore, transcriptomic data obtained from WT *S. clavuligerus*

217

show that *ccaR* expression is lowest during the transition phase from exponential growth to the second growth phase, which coincides with a peak in *glcP* expression (Hwang *et al.*, 2019, supplementary data). These findings indicate that *glcP* expression in *S. clavuligerus* might be low either as the direct or indirect result of *ccaR* expression. CcaR is a cluster-situated transcription regulator. Thus, *glcP* repression might be the result of events occurring further downstream of CcaR but that are, nonetheless, CcaR-dependent. This has been proposed previously for genes involved in central carbon and nitrogen metabolism that lacked a CcaR recognition sequence but were among 186 genes that were differentially expressed in the *ccaR* deletion mutant (Álvarez-Álvarez *et al.*, 2014).

Despite not naturally utilising glucose, *glk* is expressed constitutively in WT *S. clavuligerus* (Pérez-Redondo et al., 2010). Expression of *glk* was confirmed in DSM 738, SC2 and SC6 (Figure 4-7). Yet, sequence alignments of the regions upstream of *glk* genes revealed the absence of any promoter-like sequences or transcriptional start sequences (Figure 3-13). Both *glk* (*sclav_1340*) and the *sco2127* homologue, *sclav_1341*, are expressed at high levels throughout all phases of growth in liquid medium in WT *S. clavuligerus*. Yet, *sclav_1340* expression steadily increases with time while *sclav_1341* expression declines (Hwang *et al.*, 2019, supplementary data). The fact that no promoter or RBS sequence was predicted upstream of *sclav_1341* using convectional tools highlights the necessity for investigating expression of central carbon metabolic genes in *S. clavuligerus*.

The role of *sco2127* and its hypothetical gene product *SCO2127* was investigated in *S. coelicolor* and *S. peucetius var. caesius*, where it was essential for glucose uptake and glucose dependent CCR. It positively impacts *glk*A expression, GlkA activity and has also been proposed to stimulate *glcP* expression (Angell *et al.*, 1994; Guzman *et al.*, 2005). SCO2127 further interacts with SCO5113, BldKB, and SCO2582, a putative metalloendopeptidase (Chávez *et al.*, 2011). In the *bld* hierarchy, the signal produced by BldK influences *bldA* and *adpA* (*bldH*) as the next components of the signalling cascade (Nodwell *et al.*, 1996). Thus, SCO2127 might not only effect glucose uptake, catabolism and CCR, but may also influence morphological development through interaction with BldK. Surprisingly, Chávez *et al*. (2011) also showed that SCO2127 is not pulled down together with neither GlcP nor GlkA in *S. coelicolor*, indicating that SCO2127 might not physically interact with either of these proteins. However, SCO2127 was not expressed until the end of exponential growth, while previous

work has shown that *glcP1* expression is induced by glucose. In contrast, *glk*A is expressed constitutively. The lack of physical interaction between SCO2127 and GlcP1 in *S. coelicolor* might have, therefore, been due to time-dependent differential expression profiles (van Wezel *et al.*, 2005, 2007; Chávez *et al.*, 2011). However, these findings raise the question as to how SCO2127 would logistically be able to stimulate *glk* and potentially also *glcP* expression if their expression profiles were not compatible. Thus, the role of SCO2127 in stimulating glucose uptake remains unclear. Interestingly, the deletion of *sco2127* also results in an intracellular accumulation of nitrite caused by downregulation of *sco0216* and *sco0217* that encode the α and β chain of nitrate reductase (Cuauhtemoc *et al.*, 2016), indicating that SCO2127 might further act as a link between central carbon and nitrogen metabolism. A stark difference in gene expression between *sclav_1341* and *sclav_4529* was observed in WT *S. clavuligerus*, which does not suggest that SCLAV_1341 stimulates expression of the latter gene in *S. clavuligerus* (Hwang *et al.*, 2019, supplementary data). Nevertheless, the role of SCLAV_1341 in central carbon metabolism and CCR would be worth investigating in *S. clavuligerus*.

Investigating protein-protein interaction between GlcP and Glk proteins from *S. clavuligerus* and *S. coelicolor* showed that cognate and non-cognate GlcPs interact. Although statistically non-significant, a signal was also detected from an interaction between GlcP1 and GlkA from *S. coelicolor*. Yet no interactions between Glk proteins, nor between GlcP and Glk from *S. clavuligerus* were observed (Figure 6-10). This raises the question whether the impaired glucose utilisation of *S. clavuligerus* is, at least partly, limited by a lack of physical interaction between the native GlcP and Glk. However, this may have also been an artefact of expression in the heterologous host, *E. coli*.

7.4 WORKING MODEL FOR CCR IN *S. CLAVULIGERUS*

Research of CCR in streptomycetes has been focused on the role of Glk (Angell *et al.*, 1992, 1994; van Wezel *et al.*, 2007). More recently, however, Glk-independent mechanisms of glucose-dependent catabolite repression have been investigated (Gubbens *et al.*, 2012; Romero-Rodríguez *et al.*, 2017). Phenotypic characterisation of DSM 738, SC2, SC6 and the constructed strains have allowed the development of a working models for glycerol dependent CCR, though not of CA biosynthesis, in WT *S. clavuligerus* (Figure 7-1).

219
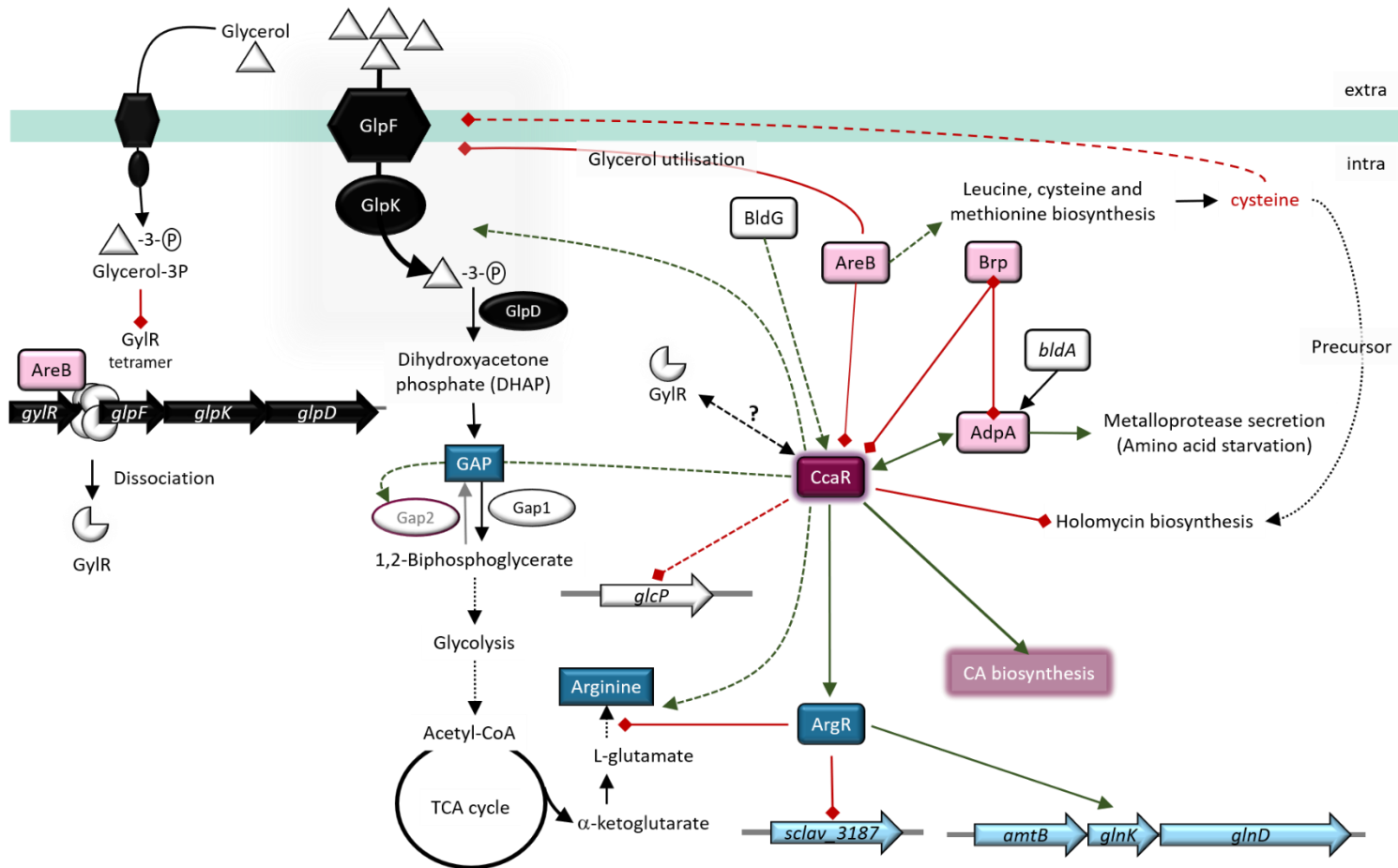
Figure 7-1: Schematic of regulators involved in central carbon, nitrogen, and specialised metabolism in WT *S. clavuligerus*. Reactions are represented by black arrows; positive and negative effects of regulators are shown as green and red, diamond-shaped dashed arrows, respectively. Direct activation and repression by regulators are represented by green arrows and red, diamond-shaped dashed arrows, respectively.

This model integrates information obtained from data presented in chapters 4 and 5, and transcriptomic data published previously (Álvarez-Álvarez *et al.*, 2014; Hwang *et al.*, 2019; Pinilla *et al.*, 2019), as well as data presented for CCR in *S. tsukubaensis* (Ordóñez-Robles *et al.*, 2017). *S. tsukubaensis*, with which *S. clavuligerus* forms stable clades in all GlcP and Glk phylogenetic trees (Figure 3-7, Figure 3-8), shares 98.4% of its 16S rRNA with *S. clavuligerus* (Martínez-Castro *et al.*, 2013). The two species are metabolically more similar than *S. clavuligerus* and *S. coelicolor*, for which the majority of CCR models have been established (Romero-Rodríguez *et al.*, 2017). Expression of *glcP* in *S. tsukubaensis* was also low and its expression is insensitive to glucose. Additionally, a lack of an RBS was reported.

In *S. tsukubaensis*, both glucose and glycerol inhibit antibiotic production (Ordóñez-Robles *et al.*, 2017). Regulators and components involved in the working model for glycerol-dependent CA biosynthesis and CCR, as well as their proposed roles are explained in detail in the following sections.

7.4.1   GLYCEROL UPTAKE SYSTEM AND ROLE OF GLYCEROL IN *S. CLAVULIGERUS*

Glycerol provides energy and precursors for CA biosynthesis in *S. clavuligerus*, while simultaneously triggering repression cephamycin C and cephalosporin production (Aharonowitz *et al.*, 1978; Hu *et al.*, 1984; Lebrihi *et al.*, 1988; Chen *et al.*, 2003). However, both CA biosynthesis, as well as repression of cephamycin C production occur in a glycerol concentration dependent manner (Lebrihi *et al.*, 1988; Chen *et al.*, 2003). Excess glycerol conditions provide a large enough intracellular GAP pool for energy generation and precursor supply, while glycerol concentrations above 54 mM will repress cephamycin C production. Paradoxically, biomass production under the same conditions will continue to increase up until a concentration of 100 mM glycerol (Lebrihi *et al.*, 1988). Concentration dependent repression of antibiotics is common. For instance, repression of tacrolimus and anthracycline production by *S. tsukubaensis* and *S. peucetius* var. *caesius*, respectively, are only triggered above a concentration of approximately 100 mM glucose (Guzman *et al.*, 2005; Chen *et al.*, 2012).

S. clavuligerus has two glycerol uptake systems, a minor constitutive system and a major, inducible system, with the phosphorylated catabolite glycerol-3-phosphate probably serving

as inducer for expression of the major system (Minambres *et al.*, 1992; Baños *et al.*, 2009). This is the case in *S. coelicolor* and *E. coli*, yet in contrast to GlpF in these species, glycerol utilisation by *S. clavuligerus* is not repressed by glucose (Sanno *et al.*, 1968; Sweet *et al.*, 1990; Minambres *et al.*, 1992; Hindle *et al.*, 1994*)*. Given the preference of *S. clavuligerus* for glycerol over glucose and the pleiotropic effects it has on secondary metabolite production, glycerol was placed at the top of the working model. Unfortunately, there is still a lack of understanding of the impact of glycerol on transcriptomics and proteomics in *S. clavuligerus*, as studies have not focused on glycerol-induced changes (see for example Ferguson *et al.*, 2016; Ünsaldı *et al.*, 2017; Pinilla *et al.*, 2019).

Furthermore, there is a knowledge gap of the role of GylR in *S. clavuligerus*, potential interaction partners and alternative regulators of the *gyl* operon. To date, only one other regulator is known to bind within the glycerol utilisation cluster in S. *coelicolor*, NdgR, the *S. clavuligerus* homologue of which, AreB, is a known repressor of *ccaR* and is discussed in more detail in section 7.4.2 (Santamarta *et al.*, 2005, 2007; Lee *et al.*, 2017). Another example for diverse control of glycerol utilisation was recently provided by *Mycobacterium smegmatis*. Expression of the *gyl* operon in this actinomycete is under tripartite control of GylR, the cAMP receptor protein Crp and the alternative sigma factor SigF, which allows the integration of glycerol availability, cellular energy state and cAMP levels (Bong *et al.*, 2019). These examples show that glycerol utilisation in *S. clavuligerus* might be the subject to multiple regulators that allow the integration of various metabolic signals.

7.4.2 AreB links CA biosynthesis, Glycerol and Amino Acid Metabolism

CcaR is regulated by BldG, AreB, the global regulator AdpA and γ-butyrolactone receptor Brp (Paradkar, 2013). AreB is a negative regulator of *ccaR* expression that is also involved in regulating amino acid biosynthesis (Santamarta *et al.*, 2005, 2007). AreB is a homologue of NdgR, a regulator involved in governing morphological differentiation, quorum sensing, amino acid metabolism, glycerol utilisation and antibiotic production in a nitrogen source dependent manner (Yang *et al.*, 2009; Kim *et al.*, 2012; Lee *et al.*, 2015, 2017). In *S. clavuligerus*, AreB binds the bidirectional promoter of *adpA* and the *leuCD* operon, thereby acting as an autoregulator and positive modulator of leucine biosynthetic genes (Santamarta *et al.*, 2007). NdgR from *S. coelicolor* has been implicated in cysteine and methionine

biosynthesis by functioning as a positive regulator of biosynthetic genes involved in these pathways (Kim *et al.*, 2012). Given the high level of conservation of NdgR homologues in *Streptomyces* (Yang *et al.*, 2009), the role of AreB in *S. clavuligerus* in the CCR model (Figure 7-1) is one that links control of CA biosynthesis through negatively regulating *ccaR* expression to amino acid metabolism through acting as a positive modulator of amino acid biosynthetic genes. Cysteine is of particular interest given that it is the precursor of holomycin biosynthesis, a pathway that is blocked by CcaR, and also due to the repressive effect cysteine has on glycerol utilisation (Kenig *et al.*, 1979; Minambres *et al.*, 1992; De la Fuente *et al.*, 2002; Álvarez-Álvarez *et al.*, 2014; Liras, 2014). This implies a level of antagonism of amino acid catabolism and glycerol catabolism in *S. clavuligerus*, which is supported by the more recent finding that NdgR negatively regulates expression of the glycerol utilisation operon by binding upstream of the glycerol facilitator gene in *S. coelicolor* (Lee *et al.*, 2017). The observation that SC6 was impaired in both utilisation of casamino acids and glycerol as sources of carbon in liquid minimal media might indicate that mechanisms involved in regulating NdgR-dependent pathways may have been affected by genetic alterations that have occurred throughout the strain improvement process.

7.4.3 ADPA REGULATES CA BIOSYNTHESIS AND PROTEASE PRODUCTION

Overlaps in the NdgR and AdpA regulons exist in *S. coelicolor*, *S. peucetius* and also in *S. clavuligerus* as both AreB and AdpA impact *ccaR* expression (Santamarta *et al.*, 2007; Yang *et al.*, 2009; López-García *et al.*, 2010). Yet, while the influence of AreB on *ccaR* expression is weak and negative, AdpA and *CcaR* appear to positively regulate each other (Santamarta *et al.*, 2005, 2007; Álvarez-Álvarez *et al.*, 2014). AdpA homologues are known to positively control cluster-situated regulators of antibiotic production in *Streptomyces*, such as *S. coelicolor*, *S. griseus*, *S. ansochromogenes*, morphological development and replication (Ohnishi *et al.*, 1999, 2005; Pan *et al.*, 2009; López-García *et al.*, 2010; Higo *et al.*, 2011; Wolański *et al.*, 2011, 2014). Usually, control of antibiotic biosynthesis is exerted by AdpA binding to promoters of cluster-situated regulator genes, however in *S. ghanaensis*, AdpA directly binds to promoters of moenomycin biosynthetic genes, thereby skipping the level of cluster-situated regulators (Makitrynskyy *et al.*, 2013). AdpA orthologues have also been identified in members of *Pseudonocardiaceae* and *Micromonosporaceae*, although no antibiotic production regulating AdpA orthologue has been identified to date (Ostash *et al.*,

223

2015). In streptomycetes, the AdpA regulon encompasses between 100 and 500 genes, among which are proteases and protease inhibitors (Kato *et al.*, 2002, 2005; Tomono *et al.*, 2005; Higo *et al.*, 2011, 2012; Guyet *et al.*, 2014). The metalloendopeptidase SgmA is of particular interest as this protease shares 70.5% of its amino acid sequence with the zinc-dependent metallopeptidase SCLAV_4539. In *S. griseus*, AdpA binds two sites upstream of *sgmA*, thus, acting as a transcriptional activator, as well as a time-dependent enhancer of transcription. Interestingly, *sgmA* disruption mutants produced a yellow pigment (Kato *et al.*, 2002), which was also observed for SC6-ASB4 colonies grown on milk agar. SC6-ASB4 was the only strain that did not degrade the milk protein in the agar. In contrast to *sgmA* disruption mutants, however, SC6-ASB4 also failed to produce aerial mycelium (Kato *et al.*, 2002). Transcriptomic data obtained from SC6 (Table 4-2) and WT *S. clavuligerus* (Pinilla *et al.*, 2019) have shown that upregulation of *adpA* and *sclav_4529* coincide. Unfortunately, no information on protease production in the *ΔadpA* mutant is available (López-García *et al.*, 2010). Based on the role of AdpA in *S. griseus* and *S. coelicolor*, as well as transcriptomic data obtained from SC6 and WT *S. clavuligerus* strain, AdpA is placed above CcaR in the CA biosynthesis regulatory cascade (Figure 7-1) but also between antibiotic production and protease production (López-García *et al.*, 2010; Álvarez-Álvarez *et al.*, 2014; Pinilla *et al.*, 2019). Expression of *adpA* depends on *bldA*, to provide the rare TTA codon, and its repressor Brp, which directly binds the *adpA* promoter (Nguyen *et al.*, 2003; Takano *et al.*, 2003; López-García *et al.*, 2010).

7.4.4 BRP REGULATES *ccaR* EXPRESSION DIRECTLY AND INDIRECTLY

Brp is a γ-butyrolactone receptor that binds to recognition sequences upstream of *adpA* and *ccaR*, thereby acting as a transcriptional repressor (Santamarta *et al.*, 2005; López-García *et al.*, 2010). The corresponding gene, *sclav_p0894*, is also regulated by Brp and located on the giant linear mega plasmid, pSCL4 (Santamarta *et al.*, 2005; Medema *et al.*, 2010). Paralogous CA biosynthetic genes have also been identified on the plasmid, along with biosynthetic gene clusters (BGGs) for 24 other secondary metabolites (Jensen *et al.*, 2000; Medema *et al.*, 2010). In total, 58 biosynthetic gene clusters (BGCs) for secondary metabolites have been identified in *S. clavuligerus* to date, most of which are not expressed under standard laboratory conditions (Hwang *et al.*, 2019). Crosstalk appears to exist between secondary metabolite BGCs and regulators situated on the chromosome and on pSCL4. *S. clavuligerus*

lacking pSCL4, for instance, produces reduced amounts of CA and cephamycin C, due to a downregulation of late biosynthetic genes, while holomycin production is increased (Álvarez-Álvarez *et al.*, 2017). Cross-talk between plasmids and chromosomes are not uncommon (Vial *et al.*, 2020), and Brp might be one of the regulators contributing to the plasmid-chromosome cross-talk that regulates secondary metabolite production in *S. clavuligerus*.

7.4.5 ArgR Links Arginine Biosynthesis and CA Biosynthesis

ArgR is the transcriptional repressor of arginine biosynthetic genes that are organised in the *argCJBDRGH* cluster (Rodríguez-García *et al.*, 1997, 2000). ArgR appears to be positively influenced by CcaR in *S. clavuligerus* (Figure 7-1) (Álvarez-Álvarez *et al.*, 2014). In *S. coelicolor*, it has been shown to affect the expression of 1544 genes throughout cultivation in liquid, yet only 44 genes were identified upstream of which there are ArgR binding sites (ARG boxes) (Pérez-Redondo *et al.*, 2012; Botas *et al.*, 2018). Two genes, for which ARG boxes were identified are *sco4158* and *sco5583* (*amtB*). The former gene encodes an a putative LacI-type regulator, the expression of which has been found to be upregulated in *S. coelicolor* in response to glucose (Romero-Rodríguez *et al.*, 2016). The gene encoding the *S. clavuligerus* homologue, *sclav_3187* (SCLAV_3187 and SCO4158 are 83% identical) is altered in SC2, where a frameshift was detected compared to the WT reference sequence and the SC6 sequence (Table 4-3, Table 4-4). SC2 exhibited longer growth phases in NMMP containing casamino acids compared to DSM 738 (Figure 4-4), an effect that was absent when most of the nitrogen was provided by ammonium sulphate (Figure 4-5). The role of SCLAV_3187 in *S. clavuligerus* is still unknown yet might be worth investigating.

The second gene upstream of which an ARG box was detected is *amtB* (Botas *et al.*, 2018). This gene forms an operon with *glnK* and *glnD* and all three genes were under expressed in a *ΔargR* mutant (Fink *et al.*, 2002; Hesketh *et al.*, 2002; Álvarez-Álvarez *et al.*, 2014; Botas *et al.*, 2018). GlnD from *S. coelicolor* and *S. clavuligerus* share 74% of their amino acid sequence and the corresponding GlnK proteins are 92% identical. Amino acid residues that are adenylated in *S. coelicolor* are also present in the *S. clavuligerus* homologue. Therefore, it is not improbable that GlnD also adenylates GlnK in DSM 738 and SC2 in response to ammonium availability, as is the case in *S. coelicolor* (Hesketh *et al.*, 2002). SC6, however, in which GlnD is truncated (Table 4-3, Table 4-4), exhibited poor growth in NMMP

supplemented with casamino acids and lacking glucose (Figure 4-4). This might be the phenotypic manifestation of malfunctioning of GlnD or the complete absence thereof. Moreover, GlnK has been identified as a pleiotropic regulator of development and secondary metabolism in *S. coelicolor* (Waldvogel *et al.*, 2011), highlighting the importance of understanding the roles of GlnD and GlnK in *S. clavuligerus*. SC6 produced significantly higher CA yields compared to DSM 738 and SC2, the genetic basis of which is still unknown and may be linked to GlnK.

7.5 GLUCOSE TRIGGERED CCR OF CA BIOSYNTHESIS IN *S. CLAVULIGERUS* STRAINS

Although neither DSM 738 nor SC2 produced large yields of CA, the presence of glucose in the medium did not cause repression of its biosynthesis in either of the strains. CA biosynthesis by SC2-ASB1, however, was entirely abolished at 27 mM glucose. Biomass production by this strain was almost entirely absent at 54 mM glucose. WT *S. clavuligerus* is known to import residual amount of glucose (Aharonowitz *et al.*, 1978; Garcia-Dominguez *et al.*, 1989), yet the internal concentration of glucose and associated catabolites may be too low to trigger CCR in these strains. Glucose dependent repression of CA biosynthesis in SC2-ASB1 might, thus, simply be the result of increased carbon flux through GlcP-Glk.

Alternatively, given that *S. clavuligerus* does not naturally utilise glucose, the absence of CA production by SC2-ASB1 might be caused by a shortage in GAP, the rate limiting precursor molecule (Ives *et al.*, 1997; Chen *et al.*, 2003), if glycolytic genes are not upregulated in response to glucose, for which no evidence has been presented thus far. Given that the absence of CA production was only observed for SC2-ASB1 and not for SC2-ASB3 (expressing *sco5578*), this suggests only cognate GlcP and Glk proteins elicit a CCR response. However, no protein-protein interaction between GlcP and Glk from *S. clavuligerus* was detected in the BACTH assays (Figure 6-8, Figure 6-9), which may have been due to expression in the heterologous host, *E. coli*.

Although SC6 appeared to utilise glucose efficiently, which is presumably also caused by an increase in expression of its native *glcP* gene, this strain produced significantly higher CA yields than SC2. Yet, heterologous expression of *glk* genes *sclav_1340* and *sco2126* further increased those yields. As mentioned in the introduction, regulation of secondary

226

metabolism through crotonylation might be impaired in WT-like strains given the absence of *cobB*, the de-crotonylase. Crotonylation of Glk in *S. roseosporus* positively regulated CCR (Sun *et al.*, 2020). An abundance of Glk lacking this PTM in *glk*-expressing strains, such as SC6-ASB2 and SC6-ASB4, might further relieve CCR resulting in greater CA yields. The role of crotonolyation as a regulatory mechanism of secondary metabolism remains to be investigated in *S. clavuligerus*.

8 CONCLUSION AND FUTURE WORK

The phylogenetic analyses of GlcP and Glk among Actinobacteria using the ActDES database and bioinformatic analyses carried out with the group of streptomycetes have provided insight into the occurring gene family expansion through gene duplication and HGT. Given that *S. clavuligerus* has only a single GlcP, yet two glycerol permeases (Minambres *et al.*, 1992; Baños *et al.*, 2009; Schniete *et al.*, 2021), an understanding of the evolutionary history of central carbon metabolism in this species might be gained by carrying out the analogous database searches for components of Actinobacterial glycerol uptake systems. This might be especially interesting given the multiple examples of HGT that appear to have occurred in *S. clavuligerus* for genes involved in secondary metabolism (Kim *et al.*, 2007; Ayala-Ruano *et al.*, 2019). HGT might have, therefore, also played a role in the acquisition or gene family expansion of genes involved in central carbon metabolism. A solid understanding of the evolutionary history of central carbon metabolism in *S. clavuligerus* might further aid in unlocking the species' full metabolic potential. Targeting any of the presented genes potentially involved in carbohydrate uptake and catabolism may also contribute to this.

Central carbon, nitrogen and secondary metabolism in WT and industrial *S. clavuligerus* are under the control of a network of regulators, the roles of some have not yet been investigated, such as SCLAV_3187, GlnK and GlnD. Random mutagenesis introduces a plethora of mutations into potentially important genes, the altered gene products of which might significantly alter growth, viability, and CA production in industrial strains, as it appears to be the case for SC6, without being able to comprehend the underlying genetic changes. Therefore, rational strain improvement strategies are the preferred approach to production strain improvement. Yet, heterologous expression of *glcP* in *S. clavuligerus*, which was a rational approach based on the knowledge available on central carbon metabolism in this

species, produced the adverse effect, namely total absence of CA biosynthesis. This highlights the importance of gaining a more thorough understanding of the metabolic network in WT and industrial *S. clavuligerus*. Thus, transcriptomic analyses of DSM 738, SC2, SC6, as well as the constructed strains in the presence of glycerol or glucose might aid in understanding the pleiotropic effects of these carbon sources on gene expression. This approach was taken when constructing the CCR model for *S. tsukubaensis*, in which glucose and glycerol elicit a repressive response (Ordóñez-Robles *et al.*, 2017).

Moreover, examining the impact and function of Glk in *S. clavuligerus* might be of interest by generating strains that lack Glk. The native *glk* gene is constitutively expressed in WT *S. clavuligerus* (Pérez-Redondo *et al.*, 2010), yet the physiological function and a potential involvement in CCR have not yet been investigated. The role of crotonylation might also aid in understanding why CCR-dependent regulatory mechanisms operate the way they do in this species. It might, thus, be helpful to investigate crotonylation patterns, especially since *S. clavuligerus* appears to be lacking CobB, analogously to the way it was examined in *S. roseosporus* (Sun *et al.*, 2020).

9 REFERENCES

Abraham, E. P. and Chain, E. (1940) 'An enzyme from bacteria able to destroy penicillin', *Nature*, 146(3713), p. 837.

AbuSara, N. F., Piercey, B. M., Moore, M. A., Shaikh, A. A., Nothias, L.-F., Srivastava, S. K., Cruz-Morales, P., Dorrestein, P. C., Barona-Gómez, F. and Tahlan, K. (2019) 'Comparative Genomics and Metabolomics Analyses of Clavulanic Acid-Producing *Streptomyces* Species Provides Insight Into Specialized Metabolism', *Frontiers in Microbiology*, p. 2550.

Aharonowitz, Y. (1979) 'Regulatory interrelationships of nitrogen metabolism and cephalosporin biosynthesis', *Genetics of industrial microorganisms. American Society for Microbiology*, pp. 210–217.

Aharonowitz, Y. and Demain, A. L. (1977) 'Influence of inorganic phosphate and organic buffers on cephalosporin production by *Streptomyces clavuligerus*', *Archives of Microbiology*, 115(2), pp. 169–173.

Aharonowitz, Y. and Demain, A. L. (1978) 'Carbon catabolite regulation of cephalosporin production in *Streptomyces clavuligerus*', *Antimicrobial Agents and Chemotherapy*, 14(2), pp. 159–164.

Aidoo, K. A., Wong, A., Alexander, D. C., Rittammer, R. A. R. and Jensen, S. E. (1994) 'Cloning, sequencing and disruption of a gene from *Streptomyces clavuligerus* involved in clavulanic acid biosynthesis', *Gene*, 147(1), pp. 41–46.

Akanuma, G., Hara, H., Ohnishi, Y. and Horinouchi, S. (2009) 'Dynamic changes in the extracellular proteome caused by absence of a pleiotropic regulator AdpA in *Streptomyces griseus*', *Molecular microbiology*, 73(5), pp. 898–912.

Alanjary, M., Steinke, K. and Ziemert, N. (2019) 'AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential', *Nucleic acids research*, 47(W1), pp. W276–W282.

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.

Ali, N., Herron, P. R., Evans, M. C. and Dyson, P. J. (2002) 'Osmotic regulation of the *Streptomyces lividans* thiostrepton-inducible promoter, *ptipA*', *Microbiology*, 148(2), pp. 381–390.

Altizer, S., Ostfeld, R. S., Johnson, P. T. J., Kutz, S. and Harvell, C. D. (2013) 'Climate change and infectious diseases: from evidence to a predictive framework', *Science*, 341(6145), pp. 514–519.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool.', *Journal of molecular biology*, 215(3), pp. 403–410.

Álvarez-Álvarez, R., Martínez-Burgo, Y., Pérez-Redondo, R., Braña, A. F., Martín, J. F. and Liras, P. (2013) 'Expression of the endogenous and heterologous clavulanic acid cluster in *Streptomyces flavogriseus*: why a silent cluster is sleeping', *Applied microbiology and biotechnology*, 97(21), pp. 9451–9463.

Álvarez-Álvarez, R., Martínez-Burgo, Y., Rodríguez-García, A. and Liras, P. (2017) 'Discovering the potential of *S. clavuligerus* for bioactive compound production: cross-talk between the chromosome and the pSCL4 megaplasmid', *BMC genomics*, 18(1), pp. 1–13.

Álvarez-Álvarez, R., Rodríguez-García, A., Santamarta, I., Pérez-Redondo, R., Prieto-Domínguez, A., Martínez-Burgo, Y., Liras, P., Álvarez-Álvarez, R., Rodríguez-García, A., Santamarta, I., Pérez-Redondo, R., Prieto-Domínguez, A., Martínez-Burgo, Y. and Liras, P. (2014) 'Transcriptomic analysis of *Streptomyces clavuligerus ΔccaR::tsr*: effects of the cephamycin C-clavulanic acid cluster regulator CcaR on global regulation', *Microbial biotechnology*, 7(3), pp. 221–231.

Aminov, R. I. (2009) 'The role of antibiotics and antibiotic resistance in nature', *Environmental microbiology*, 11(12), pp. 2970–2988.

Andrews, T. J. and Hatch, M. D. (1969) 'Properties and mechanism of action of pyruvate, phosphate dikinase from leaves', *Biochemical Journal*, 114(1), pp. 117–125.

Angell, S., Lewis, C. G., Buttner, M. J. and Bibb, M. J. (1994) 'Glucose repression in *Streptomyces coelicolor* A3 (2): a likely regulatory role for glucose kinase', *Molecular and General Genetics MGG*, 244(2), pp. 135–143.

Angell, S., Schwarz, E. and Bibb, M. J. (1992) 'The glucose kinase gene of *Streptomyces coelicolor* A3 (2): its nucleotide sequence, transcriptional analysis and role in glucose repression', *Molecular microbiology*, 6(19), pp. 2833–2844.

Arulanantham, H., Kershaw, N. J., Hewitson, K. S., Hughes, C. E., Thirkettle, J. E. and Schofield, C. J. (2006) 'ORF17 from the clavulanic acid biosynthesis gene cluster catalyzes the ATP-dependent formation of N-glycyl-clavaminic acid', *Journal of Biological Chemistry*, 281(1), pp. 279–287.

Ayala-Ruano, S., Santander-Gordón, D., Tejera, E., Perez-Castillo, Y. and Armijos-Jaramillo, V. (2019) 'A putative antimicrobial peptide from Hymenoptera in the megaplasmid pSCL4 of *Streptomyces clavuligerus* ATCC 27064 reveals a singular case of horizontal gene transfer with potential applications', *Ecology and evolution*, 9(5), pp. 2602–2614.

Bachmann, B. O., Li, R. and Townsend, C. A. (1998) 'β-Lactam synthetase: a new biosynthetic enzyme', *Proceedings of the National Academy of Sciences*, 95(16), pp. 9082–9086.

Bachmann, B. O. and Townsend, C. A. (2000) 'Kinetic mechanism of the β-lactam synthetase of *Streptomyces clavuligerus*', *Biochemistry*, 39(37), pp. 11187–11193.

Balakirev, E. S. and Ayala, F. J. (2003) 'Pseudogenes: are they "junk" or functional DNA?', *Annual review of genetics*, 37(1), pp. 123–151.

Baldwin, J. E., Lloyd, M. D., Wha-Son, B., Schofield, C. J., Elson, S. W., Baggaleyb, K. H. and Nicholson, N. H. (1993) 'A substrate analogue study on clavaminic acid synthase: possible clues to the biosynthetic origin of proclavamic acid', *Journal of the Chemical Society, Chemical Communications*, (6), pp. 500–502.

Baños, S., Pérez-Redondo, R., Koekman, B. and Liras, P. (2009) 'Glycerol utilization gene cluster in *Streptomyces clavuligerus*', *Applied and environmental microbiology*, 75(9), pp. 2991–2995.

Bascarán, V., Hardisson, C. and Braña, A. F. (1990) 'Regulation of extracellular protease production in *Streptomyces clavuligerus*', *Applied microbiology and Biotechnology*, 34(2), pp. 208–213.

Bascarán, V., Sánchez, L., Hardisson, C. and Braña, A. F. (1991) 'Stringent response and

initiation of secondary metabolism in *Streptomyces clavuligerus*', *Microbiology*, 137(7), pp. 1625–1634.

Battesti, A. and Bouveret, E. (2012) 'The bacterial two-hybrid system based on adenylate cyclase reconstitution in *Escherichia coli*', *Methods*, 58(4), pp. 325–334.

Beaman, B. L. and Beaman, L. (1994) '*Nocardia* species: host-parasite relationships.', *Clinical Microbiology Reviews*, 7(2), pp. 213 LP – 264.

Belkhir, L. and Elmeligi, A. (2019) 'Carbon footprint of the global pharmaceutical industry and relative impact of its major players', *Journal of Cleaner Production*, 214, pp. 185–194.

Bentley, S. D., Brown, S., Murphy, L. D., Harris, D. E., Quail, M. A., Parkhill, J., Barrell, B. G., McCormick, J. R., Santamaria, R. I. and Losick, R. (2004) 'SCP1, a 356 023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2)', *Molecular microbiology*, 51(6), pp. 1615–1628.

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H. and Harper, D. (2002) 'Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2)', *Nature*, 417(6885), pp. 141–147.

Bersanetti, P. A., Almeida, R. M. R. G., Barboza, M., Araújo, M. L. G. C. and Hokka, C. O. (2005) 'Kinetic studies on clavulanic acid degradation', *Biochemical engineering journal*, 23(1), pp. 31–36.

Bertani, G. (1951) 'Studies on lysogenesis I.: the mode of phage liberation by lysogenic *Escherichia coli*', *Journal of bacteriology*, 62(3), p. 293.

Bertram, R., Schlicht, M., Mahr, K., Nothaft, H., Saier, M. H. and Titgemeyer, F. (2004) '*In silico* and transcriptional analysis of carbohydrate uptake systems of *Streptomyces coelicolor* A3(2)', *Journal of bacteriology*, 186(5), pp. 1362–1373.

Bibb, M. J. (2005) 'Regulation of secondary metabolism in streptomycetes', *Current opinion in microbiology*, 8(2), pp. 208–215.

Bierman, M., Logan, R., O'brien, K., Seno, E. T., Rao, R. N. and Schoner, B. E. (1992) 'Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces spp*.', *Gene*, 116(1), pp. 43–49.

Bignell, D. R. D., Tahlan, K., Colvin, K. R., Jensen, S. E. and Leskiw, B. K. (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.

Bignell, D. R. D., Warawa, J. L., Strap, J. L., Chater, K. F. and Leskiw, B. K. (2000) 'Study of the *bldG* locus suggests that an anti-anti-sigma factor and an anti-sigma factor may be involved in Streptomyces coelicolor antibiotic production and sporulation.', *Microbiology*, 146(9), pp. 2161–2173.

Bird, A. E., Bellis, J. M. and Gasson, B. C. (1982) 'Spectrophotometric assay of clavulanic acid by reaction with imidazole', *Analyst*, 107(1279), pp. 1241–1245.

Biro, S. and Chater, K. F. (1987) 'Cloning of *Streptomyces griseus* and *Streptomyces lividans* genes for glycerol dissimilation', *Gene*, 56(1), pp. 79–86.

Bolotin, A. and Biro, S. (1990) 'Nucleotide sequence of the putative regulatory gene and major promoter region of the *Streptomyces griseus* glycerol operon', *Gene*, 87(1), pp. 151–152.

Bong, H.-J., Ko, E.-M., Song, S.-Y., Ko, I.-J. and Oh, J.-I. (2019) 'Tripartite Regulation of the *glpFKD* Operon Involved in Glycerol Catabolism by GylR, Crp, and SigF in *Mycobacterium smegmatis*', *Journal of Bacteriology*, 201(24), pp. e00511-19.

Borgnia, M., Nielsen, S., Engel, A. and Agre, P. (1999) 'Cellular and molecular biology of the aquaporin water channels', *Annual review of biochemistry*, 68(1), pp. 425–458.

Botas, A., Pérez-Redondo, R., Rodríguez-García, A., Álvarez-Álvarez, R., Yagüe, P., Manteca, A. and Liras, P. (2018) 'ArgR of *Streptomyces coelicolor* is a pleiotropic transcriptional regulator: effect on the transcriptome, antibiotic production, and differentiation in liquid cultures', *Frontiers in microbiology*, 9, p. 361.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. and Bairoch, A. (2007) 'Uniprotkb/swiss-prot', in *Plant bioinformatics*, pp. 89–112.

Braña, A. F., Wolfe, S. and Demain, A. L. (1985) 'Ammonium repression of cephalosporin production by *Streptomyces clavuligerus*', *Canadian journal of microbiology*, 31(8), pp. 736–

743.

Briczinski, E. P., Phillips, A. T. and Roberts, R. F. (2008) 'Transport of Glucose by *Bifidobacterium animalis* subsp. *lactis* Occurs via Facilitated Diffusion', *Applied and Environmental Microbiology*, 74(22), pp. 6941 LP – 6948.

Brockmann, H. and Hieronymus, E. (1955) 'Über Actinomycetenfarbstoffe, V. Mitteil. 1): Zur Konstitution des Actinorhodins, III. Mitteil1)', *Chemische Berichte*, 88(9), pp. 1379–1390.

Brockmann, H. and Pini, H. (1950) 'Über Actinomycetenfarbstoffe, I. Mitteil.: Actinorhodin, ein roter, antibiotisch wirksamer Farbstoff aus Actinomyceten', *Chemische Berichte*, 83(2), pp. 161–167.

Bromberg, Y., Yachdav, G. and Rost, B. (2008) 'SNAP predicts effect of mutations on protein function', *Bioinformatics*, 24(20), pp. 2397–2398.

Brückner, R. and Titgemeyer, F. (2002) 'Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization', *FEMS microbiology letters*, 209(2), pp. 141–148.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Van Themaat, E. V. L. and Schulze-Lefert, P. (2013) 'Structure and functions of the bacterial microbiota of plants', *Annual review of plant biology*, 64, pp. 807–838.

Bush, K., Jacoby, G. A. and Medeiros, A. A. (1995) 'A functional classification scheme for beta-lactamases and its correlation with molecular structure.', *Antimicrobial agents and chemotherapy*, 39(6), p. 1211.

Butler, M. J., Deutscher, J., Postma, P. W., Wilson, T. J. G., Galinier, A. and Bibb, M. J. (1999) 'Analysis of a *ptsH* homologue from *Streptomyces coelicolor* A3 (2)', *FEMS microbiology letters*, 177(2), pp. 279–288.

Buttner, M. J., Chater, K. F. and Bibb, M. J. (1990) 'Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2).', *Journal of Bacteriology*, 172(6), pp. 3367 LP – 3378.

Buttner, M. J., Fearnley, I. M. and Bibb, M. J. (1987) 'The agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2): nucleotide sequence and transcriptional analysis', *MGG Molecular &*

General Genetics, 209(1), pp. 101–109.

Calabrese, E. J. and Blain, R. (2005) 'The occurrence of hormetic dose responses in the toxicological literature, the hormesis database: an overview', *Toxicology and applied pharmacology*, 202(3), pp. 289–301.

Cao, G., Zhong, C., Zong, G., Fu, J., Liu, Z., Zhang, G. and Qin, R. (2016) 'Complete Genome Sequence of *Streptomyces clavuligerus* F613-1, an Industrial Producer of Clavulanic Acid', *Genome Announcements*, 4(5), pp. e01020-16.

Case, R. J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W. F. and Kjelleberg, S. (2007) 'Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies', *Appl. Environ. Microbiol.*, 73(1), pp. 278–288.

Cha, J.-H. and Stewart, G. C. (1997) 'The *divIVA* minicell locus of *Bacillus subtilis*.', *Journal of bacteriology*, 179(5), pp. 1671–1683.

Chakraburtty, R. and Bibb, M. (1997) 'The ppGpp synthetase gene *relA* of *Streptomyces coelicolor* A3(2) plays a conditional role in antibiotic production and morphological differentiation.', *Journal of Bacteriology*, 179(18), pp. 5854–5861.

Challis, G. L. and Hopwood, D. A. (2003) 'Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species', *Proceedings of the National Academy of Sciences*, 100(suppl 2), pp. 14555–14561.

Champness, W. C. (1988) 'New loci required for *Streptomyces coelicolor* morphological and physiological differentiation.', *Journal of bacteriology*, 170(3), pp. 1168–1174.

Chan, M. and Sim, T.-S. (1998) 'Malate synthase from *Streptomyces clavuligerus* NRRL3585: cloning, molecular characterization and its control by acetate', *Microbiology*, 144(11), pp. 3229–3237.

Chater, K. F. (2001) 'Regulation of sporulation in *Streptomyces coelicolor* A3 (2): a checkpoint multiplex?', *Current opinion in microbiology*, 4(6), pp. 667–673.

Chater, K. F., Biró, S., Lee, K. J., Palmer, T. and Schrempf, H. (2010) 'The complex extracellular biology of *Streptomyces*', *FEMS Microbiology Reviews*.

Chater, K. F., Bruton, C. J., Plaskitt, K. A., Buttner, M. J., Méndez, C. and Helmann, J. D. (1989) 'The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility σ factor of *B. subtilis*', *Cell*, 59(1), pp. 133–143.

Chatterji, D., Fujita, N. and Ishihama, A. (1998) 'The mediator for stringent control, ppGpp, binds to the β-subunit of *Escherichia coli* RNA polymerase', *Genes to Cells*, 3(5), pp. 279–287.

Chávez, A., Forero, A., Sánchez, M., Rodríguez-Sanoja, R., Mendoza-Hernández, G., Servín-Gonzalez, L., Sánchez, B., García-Huante, Y., Rocha, D. and Langley, E. (2011) 'Interaction of SCO2127 with BldKB and its possible connection to carbon catabolite regulation of morphological differentiation in *Streptomyces coelicolor*', *Applied microbiology and biotechnology*, 89(3), pp. 799–806.

Chávez, A., García-Huante, Y., Ruiz, B., Langley, E., Rodríguez-Sanoja, R. and Sanchez, S. (2009) 'Cloning and expression of the *sco2127* gene from *Streptomyces coelicolor* M145', *Journal of industrial microbiology & biotechnology*, 36(5), pp. 649–654.

Chen, D., Zhang, Qi, Zhang, Qinglin, Cen, P., Xu, Z. and Liu, W. (2012) 'Improvement of FK506 production in *Streptomyces tsukubaensis* by genetic enhancement of the supply of unusual polyketide extender units via utilization of two distinct site-specific recombination systems', *Applied and environmental microbiology*, 78(15), pp. 5093–5103.

Chen, K.-C., Lin, Y.-H., Tsai, C.-M., Hsieh, C.-H. and Houng, J.-Y. (2002) 'Optimization of glycerol feeding for clavulanic acid production by *Streptomyces clavuligerus* with glycerol feeding', *Biotechnology Letters*, 24(6), pp. 455–458.

Chen, K.-C., Lin, Y.-H., Wu, J.-Y. and Hwang, S.-C. J. (2003) 'Enhancement of clavulanic acid production in *Streptomyces clavuligerus* with ornithine feeding', *Enzyme and microbial technology*, 32(1), pp. 152–156.

Chevrette, M. G., Gutiérrez-García, K., Selem-Mojica, N., Aguilar-Martínez, C., Yañez-Olvera, A., Ramos-Aboites, H. E., Hoskisson, P. A. and Barona-Gómez, F. (2020) 'Evolutionary dynamics of natural product biosynthesis in bacteria', *Natural Product Reports*.

Chiu, M. L., Folcher, M., Griffin, P., Holt, T., Klatt, T. and Thompson, C. J. (1996) 'Characterization of the covalent binding of thiostrepton to a thiostrepton-induced protein from *Streptomyces lividans*', *Biochemistry*, 35(7), pp. 2332–2341.

Chiu, M. L., Folcher, M., Katoh, T., Puglia, A. M., Vohradsky, J., Yun, B.-S., Seto, H. and Thompson, C. J. (1999) 'Broad Spectrum Thiopeptide Recognition Specificity of the *Streptomyces lividans* TipAL Protein and Its Role in Regulating Gene Expression', *Journal of Biological Chemistry*, 274(29), pp. 20578–20586.

Claessen, D., De Jong, W., Dijkhuizen, L. and Wösten, H. A. B. (2006) 'Regulation of *Streptomyces* development: reach for the sky!', *Trends in microbiology*, 14(7), pp. 313–319.

Claessen, D., Rink, R., de Jong, W., Siebring, J., de Vreugd, P., Boersma, F. G. H., Dijkhuizen, L. and Wösten, H. A. B. (2003) 'A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in *Streptomyces coelicolor* by forming amyloid-like fibrils', *Genes & development*, 17(14), pp. 1714–1726.

Claessen, D., Stokroos, I., Deelstra, H. J., Penninga, N. A., Bormann, C., Salas, J. A., Dijkhuizen, L. and Wösten, H. A. B. (2004) 'The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlins and chaplins', *Molecular microbiology*, 53(2), pp. 433–443.

Claessen, D., Wösten, H. A. B., Keulen, G. van, Faber, O. G., Alves, A. M. C. R., Meijer, W. G. and Dijkhuizen, L. (2002) 'Two novel homologous proteins of *Streptomyces coelicolor* and *Streptomyces lividans* are involved in the formation of the rodlet layer and mediate attachment to a hydrophobic surface', *Molecular microbiology*, 44(6), pp. 1483–1492.

Clamp, M., Cuff, J., Searle, S. M. and Barton, G. J. (2004) 'The jalview java alignment editor', *Bioinformatics*, 20(3), pp. 426–427.

Confer, D. L. and Eaton, J. W. (1982) 'Phagocyte impotence caused by an invasive bacterial adenylate cyclase', *Science*, 217(4563), pp. 948–950.

Cuauhtemoc, L.-C., Nidia, M.-C., Alba, R.-R., Sara, C.-L., Esteban, M., Romina, R.-S., Ruiz-Villafán, B. and Sergio, S. (2016) 'Deletion of the hypothetical protein SCO2127 of *Streptomyces coelicolor* allowed identification of a new regulator of actinorhodin production', *Applied microbiology and biotechnology*, 100(21), pp. 9229–9237.

Cummins, C. S. and Harris, H. (1958) 'Studies on the cell-wall composition and taxonomy of Actinomycetales and related groups', *Microbiology*, 18(1), pp. 173–189.

Dalton, K. A., Thibessard, A., Hunter, J. I. B. and Kelemen, G. H. (2007) 'A novel compartment, the 'subapical stem'of the aerial hyphae, is the location of a sigN-dependent, developmentally distinct transcription in *Streptomyces coelicolor*', *Molecular microbiology*, 64(3), pp. 719–737.

Daniel, R. A. and Errington, J. (2003) 'Control of cell morphogenesis in bacteria: two distinct ways to make a rod-shaped cell', *Cell*, 113(6), pp. 767–776.

Davies, J. (2006) 'Are antibiotics naturally antibiotics?', *Journal of Industrial Microbiology and Biotechnology*, 33(7), pp. 496–499.

Davies, J. and Davies, D. (2010) 'Resistance origins and evolution of antibiotic', *Microbiology and Molecular Biology reviews. Microbiol Mol Biol Rev*, 74(3), pp. 417–433.

Delić, V., Hopwood, D. A. and Friend, E. J. (1970) 'Mutagenesis by *N*-methyl-*N*'-nitro-*N*-nitrosoguanidine (NTG) in *Streptomyces coelicolor*', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 9(2), pp. 167–182.

Demain, A. L. (1989) 'Carbon source regulation of idiolite biosynthesis', *Regulation of secondary metabolism in Actinomycetes*, pp. 127–134.

Derouaux, A., Halici, S., Nothaft, H., Neutelings, T., Moutzourelis, G., Dusart, J., Titgemeyer, F. and Rigali, S. (2004) 'Deletion of a cyclic AMP receptor protein homologue diminishes germination and affects morphological development of *Streptomyces coelicolor*', *Journal of bacteriology*, 186(6), pp. 1893–1897.

Deutscher, J., Aké, F. M. D., Derkaoui, M., Zébré, A. C., Cao, T. N., Bouraoui, H., Kentache, T., Mokhtari, A., Milohanic, E. and Joyet, P. (2014) 'The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions', *Microbiology and Molecular Biology Reviews*, 78(2), pp. 231 LP – 256.

Edgar, R. C. (2004) 'MUSCLE: multiple sequence alignment with high accuracy and high throughput', *Nucleic acids research*, 32(5), pp. 1792–1797.

Edwards, D. H. and Errington, J. (1997) 'The *Bacillus subtilis* DivIVA protein targets to the division septum and controls the site specificity of cell division', *Molecular microbiology*,

24(5), pp. 905–915.

Elliot, M. A., Bibb, M. J., Buttner, M. J. and Leskiw, B. K. (2001) 'BldD is a direct regulator of key developmental genes in *Streptomyces coelicolor* A3(2)', *Molecular microbiology*, 40(1), pp. 257–269.

Elliot, M. A., Karoonuthaisiri, N., Huang, J., Bibb, M. J., Cohen, S. N., Kao, C. M. and Buttner, M. J. (2003) 'The chaplins: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in *Streptomyces coelicolor*', *Genes & development*, 17(14), pp. 1727–1740.

Elliot, M. A. and Leskiw, B. K. (1999) 'The BldD protein from *Streptomyces coelicolor* is a DNA-binding protein', *Journal of bacteriology*, 181(21), pp. 6832–6835.

Elliot, M. A., Locke, T. R., Galibois, C. M. and Leskiw, B. K. (2003) 'BldD from *Streptomyces coelicolor* is a non-essential global regulator that binds its own promoter as a dimer', *FEMS microbiology letters*, 225(1), pp. 35–40.

Elson, S. W., Baggaley, K. H., Davison, M., Fulston, M., Nicholson, N. H., Risbridger, G. D. and Tyler, J. W. (1993) 'The identification of three new biosynthetic intermediates and one further biosynthetic enzyme in the clavulanic acid pathway', *Journal of the Chemical Society, Chemical Communications*, (15), pp. 1212–1214.

Elson, S. W., Baggaley, K. H., Gillett, J., Holland, S., Nicholson, N. H., Sime, J. T. and Woroniecki, S. R. (1987) 'Isolation of two novel intracellular β-lactams and a novel dioxygenase cyclising enzyme from *Streptomyces clavuligerus*', *Journal of the Chemical Society, Chemical Communications*, (22), pp. 1736–1738.

Embley, T. M. and Stackebrandt, E. (1994) 'The molecular phylogeny and systematics of the actinomycetes', *Annual Reviews in Microbiology*, 48(1), pp. 257–289.

Escalante, L., Ramos, I., Imriskova, I., Langley, E. and Sanchez, S. (1999) 'Glucose repression of anthracycline formation in *Streptomyces peucetius* var. *caesius*', *Applied microbiology and biotechnology*, 52(4), pp. 572–578.

European Centre for Disease Prevention and Control (2019) *Surveillance of antimicrobial resistance in Europe 2018.*

Feitelson, J. S. and Hopwood, D. A. (1983) 'Cloning of a *Streptomyces* gene for an *O*-methyltransferase involved in antibiotic biosynthesis', *MGG Molecular & General Genetics*, 190(3), pp. 394–398.

Felsenstein, J. (1985) 'Confidence limits on phylogenies: an approach using the bootstrap', *Evolution*, 39(4), pp. 783–791.

Ferguson, N. L., Peña-Castillo, L., Moore, M. A., Bignell, D. R. D. and Tahlan, K. (2016) 'Proteomics analysis of global regulatory cascades involved in clavulanic acid production and morphological development in *Streptomyces clavuligerus*', *Journal of industrial microbiology & biotechnology*, 43(4), pp. 537–555.

Fink, D., Weissschuh, N., Reuther, J., Wohlleben, W. and Engels, A. (2002) 'Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in *Streptomyces coelicolor* A3(2)', *Molecular microbiology*, 46(2), pp. 331–347.

Flärdh, K. (2003) 'Essential role of DivIVA in polar growth and morphogenesis in *Streptomyces coelicolor* A3(2)', *Molecular microbiology*, 49(6), pp. 1523–1536.

Flärdh, K. and Buttner, M. J. (2009) '*Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium', *Nature Reviews Microbiology*, 7(1), pp. 36–49.

Fleming, A. (1929) 'On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*', *British journal of experimental pathology*, 10(3), p. 226.

Flores, M. E., Ponce, E., Rubio, M. and Huitron, C. (1993) 'Glucose and glycerol repression of α-amylase in *Streptomyces kanamyceticus* and isolation of deregulated mutants', *Biotechnology letters*, 15(6), pp. 595–600.

Forero, A., Sánchez, M., Chávez, A., Ruiz, B., Rodríguez-Sanoja, R., Servín-González, L. and Sánchez, S. (2012) 'Possible involvement of the *sco2127* gene product in glucose repression of actinorhodin production in *Streptomyces coelicolor*', *Canadian journal of microbiology*, 58(10), pp. 1195–1201.

Fröjd, M. J. and Flärdh, K. (2019) 'Apical assemblies of intermediate filament-like protein FilP are highly dynamic and affect polar growth determinant DivIVA in *Streptomyces venezuelae*',

Molecular microbiology, 112(1), pp. 47–61.

Fuchino, K., Bagchi, S., Cantlay, S., Sandblad, L., Wu, D., Bergman, J., Kamali-Moghaddam, M., Flärdh, K. and Ausmees, N. (2013) 'Dynamic gradients of an intermediate filament-like cytoskeleton are recruited by a polarity landmark during apical growth', *Proceedings of the National Academy of Sciences*, 110(21), pp. E1889–E1897.

Fulston, M., Davison, M., Elson, S. W., Nicholson, N. H., Tyler, J. W. and Woroniecki, S. R. (2001) 'Clavulanic acid biosynthesis; the final steps', *Journal of the Chemical Society, Perkin Transactions 1*, (9), pp. 1122–1130.

Fürste, J. P., Pansegrau, W., Ziegelin, G., Kröger, M. and Lanka, E. (1989) 'Conjugative transfer of promiscuous IncP plasmids: interaction of plasmid-encoded products with the transfer origin', *Proceedings of the National Academy of Sciences*, 86(6), pp. 1771 LP – 1775.

Ganeshan, S., Dickover, R. E., Korber, B. T., Bryson, Y. J. and Wolinsky, S. M. (1997) 'Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease.', *Journal of virology*, 71(1), pp. 663–677.

Garcia-Dominguez, M., Martin, J. F. and Liras, P. (1989) 'Characterization of sugar uptake in wild-type *Streptomyces clavuligerus*, which is impaired in glucose uptake, and in a glucose-utilizing mutant.', *Journal of bacteriology*, 171(12), pp. 6808–6814.

García Domínguez, M., Liras Padín, P., Laiz Trobajo, L., García Calzada, J. and Martín Martín, J. F. (1988) 'Obtention and characterization of a *S. clavuligerus gut-1* mutant able to utilize glucose for growth and cephamycin and clavulanic acid production'.

Gentry, D. R. and Cashel, M. (1996) 'Mutational analysis of the Escherichia coli *spoT* gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation', *Molecular microbiology*, 19(6), pp. 1373–1384.

German, D. M. (2003) 'The GNOME project: a case study of open source, global software development', *Software Process: Improvement and Practice*, 8(4), pp. 201–215.

Ginther, C. L. (1979) 'Sporulation and the production of serine protease and cephamycin C by *Streptomyces lactamdurans*', *Antimicrobial Agents and Chemotherapy*, 15(4), pp. 522–526.

Glauert, A. M. and Hopwood, D. A. (1961) 'The fine structure of *Streptomyces violaceoruber*

(*S. coelicolor*) III. The walls of the mycelium and spores', *The Journal of Cell Biology*, 10(4), pp. 505–516.

Goh, E.-B., Yim, G., Tsui, W., McClure, J., Surette, M. G. and Davies, J. (2002) 'Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics', *Proceedings of the National Academy of Sciences*, 99(26), pp. 17025–17030.

Gomez-Escribano, J. P., Martin, J. F., Hesketh, A., Bibb, M. J. and Liras, P. (2008) '*Streptomyces clavuligerus relA*-null mutants overproduce clavulanic acid and cephamycin C: negative regulation of secondary metabolism by (p)ppGpp', *Microbiology*, 154(3), pp. 744–755.

Goomeshi Nobary, S. and Jensen, S. E. (2012) 'A comparison of the clavam biosynthetic gene clusters in *Streptomyces antibioticus* Tü1718 and *Streptomyces clavuligerus*.', *Canadian journal of microbiology*, 58(4), pp. 413–425.

Grant, S. G., Jessee, J., Bloom, F. R. and Hanahan, D. (1990) 'Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants.', *Proceedings of the National Academy of Sciences*, 87(12), pp. 4645 LP – 4649.

Gray, D. I., Gooday, G. W. and Prosser, J. I. (1990) 'Apical hyphal extension in *Streptomyces coelicolor* A3(2)', *Microbiology*, 136(6), pp. 1077–1084.

Griffith, K. L. and Wolf Jr, R. E. (2002) 'Measuring β-galactosidase activity in bacteria: cell growth, permeabilization, and enzyme assays in 96-well arrays', *Biochemical and biophysical research communications*, 290(1), pp. 397–402.

Gubbens, J., Janus, M., Florea, B. I., Overkleeft, H. S. and van Wezel, G. P. (2012) 'Identification of glucose kinase-dependent and-independent pathways for carbon control of primary metabolism, development and antibiotic production in *Streptomyces coelicolor* by quantitative proteomics', *Molecular microbiology*, 86(6), pp. 1490–1507.

Guijarro, J., Santamaria, R., Schauer, A. and Losick, R. (1988) 'Promoter determining the timing and spatial localization of transcription of a cloned *Streptomyces coelicolor* gene encoding a spore-associated polypeptide.', *Journal of bacteriology*, 170(4), pp. 1895–1901.

Guyet, A., Benaroudj, N., Proux, C., Gominet, M., Coppée, J.-Y. and Mazodier, P. (2014) 'Identified members of the *Streptomyces lividans* AdpA regulon involved in differentiation

and secondary metabolism', *BMC microbiology*, 14(1), p. 81.

Guzman, S., Carmona, A., Escalante, L., Imriskova, I., Lopez, R., Rodríguez-Sanoja, R., Ruiz, B., Servín-González, L., Sanchez, S. and Langley, E. (2005) 'Pleiotropic effect of the SCO2127 gene on the glucose uptake, glucose kinase activity and carbon catabolite repression in *Streptomyces peucetius* var. *caesius*', *Microbiology*, 151(5), pp. 1717–1723.

Han, J.-H., Cho, M.-H. and Kim, S. B. (2012) 'Ribosomal and protein coding gene based multigene phylogeny on the family *Streptomycetaceae*', *Systematic and applied microbiology*, 35(1), pp. 1–6.

Hara, O. and Beppu, T. (1982) 'Mutants blocked in streptomycin production in *Streptomyces griseus*-the role of A-factor', *The Journal of antibiotics*, 35(3), pp. 349–358.

Hardisson, C., Manzanal, M.-B., Salas, J.-A. and Suárez, J.-E. (1978) 'Fine structure, physiology and biochemistry of arthrospore germination in *Streptomyces antibioticus*', *Microbiology*, 105(2), pp. 203–214.

Haseltine, W. A. and Block, R. (1973) 'Synthesis of guanosine tetra-and pentaphosphate requires the presence of a codon-specific, uncharged transfer ribonucleic acid in the acceptor site of ribosomes', *Proceedings of the National Academy of Sciences*, 70(5), pp. 1564–1568.

Haug, I., Weissenborn, A., Brolle, D., Bentley, S., Kieser, T. and Altenbuchner, J. (2003) '*Streptomyces coelicolor* A3(2) plasmid SCP2*: deductions from the complete sequence', *Microbiology*, 149(2), pp. 505–513.

Hempel, A. M., Cantlay, S., Molle, V., Wang, S.-B., Naldrett, M. J., Parker, J. L., Richards, D. M., Jung, Y.-G., Buttner, M. J. and Flärdh, K. (2012) 'The Ser/Thr protein kinase AfsK regulates polar growth and hyphal branching in the filamentous bacteria *Streptomyces*', *Proceedings of the National Academy of Sciences*, 109(35), pp. E2371–E2379.

Hempel, A. M., Wang, S., Letek, M., Gil, J. A. and Flärdh, K. (2008) 'Assemblies of DivIVA mark sites for hyphal branching and can establish new zones of cell wall growth in *Streptomyces coelicolor*', *Journal of bacteriology*, 190(22), pp. 7579–7583.

Hesketh, A., Chen, W. J., Ryding, J., Chang, S. and Bibb, M. (2007) 'The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces*

coelicolor A3(2)', *Genome biology*, 8(8), p. R161.

Hesketh, A., Fink, D., Gust, B., Rexer, H., Scheel, B., Chater, K., Wohlleben, W. and Engels, A. (2002) 'The GlnD and GlnK homologues of *Streptomyces coelicolor* A3(2) are functionally dissimilar to their nitrogen regulatory system counterparts from enteric bacteria', *Molecular microbiology*, 46(2), pp. 319–330.

Hesketh, A., Sun, J. and Bibb, M. (2001) 'Induction of ppGpp synthesis in *Streptomyces coelicolor* A3(2) grown under conditions of nutritional sufficiency elicits *actII-ORF4* transcription and actinorhodin biosynthesis', *Molecular microbiology*, 39(1), pp. 136–144.

Higgens, C. E. and Kastner, R. E. (1971) '*Streptomyces clavuligerus* sp. nov., a β-lactam antibiotic producer', *International Journal of Systematic and Evolutionary Microbiology*, 21(4), pp. 326–331.

Higo, A., Hara, H., Horinouchi, S. and Ohnishi, Y. (2012) 'Genome-wide distribution of AdpA, a global regulator for secondary metabolism and morphological differentiation in *Streptomyces*, revealed the extent and complexity of the AdpA regulatory network', *DNA research*, 19(3), pp. 259–274.

Higo, A., Horinouchi, S. and Ohnishi, Y. (2011) 'Strict regulation of morphological differentiation and secondary metabolism by a positive feedback loop between two global regulators AdpA and BldA in *Streptomyces griseus*', *Molecular microbiology*, 81(6), pp. 1607–1622.

Hiltner, J. K. (2015) 'Systems biology approach for metabolomic engineering in *Streptomyces*: the Phosphoenolpyruvate-Pyruvate-Oxaloacetate node'.

Hiltner, J. K., Hunter, I. S. and Hoskisson, P. A. (2015) 'Tailoring specialized metabolite production in *Streptomyces*', in *Advances in applied microbiology*, pp. 237–255.

Hindle, Z. and Smith, C. P. (1994) 'Substrate induction and catabolite repression of the *Streptomyces coelicolor* glycerol operon are mediated through the GyIR protein', *Molecular microbiology*, 12(5), pp. 737–745.

Hobbs, G., Frazer, C. M., Gardner, D. C. J., Cullum, J. A. and Oliver, S. G. (1989) 'Dispersed growth of *Streptomyces* in liquid culture', *Applied microbiology and biotechnology*, 31(3), pp.

272–277.

Hodgson, D. A. (1980) 'Carbohydrate utilization in *Streptomyces coelicolor* A3(2)'.

Hodgson, D. A. (1982) 'Glucose repression of carbon source uptake and metabolism in *Streptomyces coelicolor* A3(2) and its perturbation in mutants resistant to 2-deoxyglucose', *Microbiology*, 128(10), pp. 2417–2430.

Hodgson, D. A. (2000) 'Primary metabolism and its control in streptomycetes: a most unusual group of bacteria', *Advances in microbial physiology*, 42, pp. 47–238.

Hodgson, D. A. and Chater, K. F. (1981) 'A Chromosomal Locus Controlling Extracellular Agarase Production by *Streptomyces coelicolor* A3(2), and its Inactivation by Chromosomal Integration of Plasmid SCP1', *Microbiology*, 124(2), pp. 339–348.

Holmes, D. J., Caso, J. L. and Thompson, C. J. (1993) 'Autogenous transcriptional activation of a thiostrepton-induced gene in *Streptomyces lividans*.', *The EMBO journal*, 12(8), pp. 3183–3191.

Holmes, N. A., Walshaw, J., Leggett, R. M., Thibessard, A., Dalton, K. A., Gillespie, M. D., Hemmings, A. M., Gust, B. and Kelemen, G. H. (2013) 'Coiled-coil protein Scy is a key component of a multiprotein assembly controlling polarized growth in *Streptomyces*', *Proceedings of the National Academy of Sciences*, 110(5), pp. E397–E406.

Hong, S.-W., Kito, M., Beppu, T. and Horinouchi, S. (1991) 'Phosphorylation of the AfsR product, a global regulatory protein for secondary-metabolite formation in Streptomyces coelicolor A3(2).', *Journal of bacteriology*, 173(7), pp. 2311–2318.

Hopwood, D. A. (1967) 'Genetic analysis and genome structure in *Streptomyces coelicolor*.', *Bacteriological reviews*, 31(4), p. 373.

Hopwood, D. A. (2007) *Streptomyces* in nature and medicine: the antibiotic makers.

Hopwood, D. A., Bibb, M. J., Chater, K. F., Kieser, T., Bruton, C. J., Kieser, H. M., Lydiate, D. J., Smith, C. P., Ward, J. M. and Schrempf, H. (1985) 'Genetic manipulation of *Streptomyces*–A laboratory manual'.

Horinouchi, S. and Beppu, T. (1992) 'Regulation of secondary metabolism and cell

differentiation in *Streptomyces*: A-factor as a microbial hormone and the AfsR protein as a component of a two-component regulatory system', *Gene*, 115(1–2), pp. 167–172.

Hoskisson, P. A., Sumby, P. and Smith, M. C. M. (2015) 'The phage growth limitation system in *Streptomyces coelicolor* A(3) 2 is a toxin/antitoxin system, comprising enzymes with DNA methyltransferase, protein kinase and ATPase activity', *Virology*, 477, pp. 100–109.

Howarth, T. T., Brown, A. G. and King, T. J. (1976) 'Clavulanic acid, a novel β-lactam isolated from *Streptomyces clavuligerus*; X-ray crystal structure analysis', *Journal of the Chemical Society, Chemical Communications*, (7), pp. 266b – 267.

Hu, W.-S., Braña, A. F. and Demain, A. L. (1984) 'Carbon source regulation of cephem antibiotic production by resting cells of *Streptomyces clavuligerus* and its reversal by protein synthesis inhibitors', *Enzyme and Microbial Technology*, 6(4), pp. 155–160.

Huang, J., Shi, J., Molle, V., Björn, S., David, W., Maureen J., B., Nitsara, K., Chih-Jian, L., Camilla M., K., Mark J., B. and Stanley N., C. (2005) 'Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*', *Molecular Microbiology*, 58(5), pp. 1276–1287.

Hunt, A. C., Servín-González, L., Kelemen, G. H. and Buttner, M. J. (2005) 'The *bldC* developmental locus of *Streptomyces coelicolor* encodes a member of a family of small DNA-binding proteins related to the DNA-binding domains of the MerR family', *Journal of bacteriology*, 187(2), pp. 716–728.

Hurst, L. D. (2002) 'The Ka/Ks ratio: diagnosing the form of sequence evolution.', *Trends in genetics: TIG*, 18(9), p. 486.

Hwang, S., Lee, N., Jeong, Y., Lee, Y., Kim, W., Cho, S., Palsson, B. O. and Cho, B.-K. (2019) 'Primary transcriptome and translatome analysis determines transcriptional and translational regulatory elements encoded in the *Streptomyces clavuligerus* genome', *Nucleic acids research*, 47(12), pp. 6114–6129.

Iizumi, T., Shiogama, H., Imada, Y., Hanasaki, N., Takikawa, H. and Nishimori, M. (2018) 'Crop production losses associated with anthropogenic climate change for 1981–2010 compared with preindustrial levels', *International Journal of Climatology*, 38(14), pp. 5405–5417.

Ikeda, H., Seno, E. T., Bruton, C. J. and Chater, K. F. (1984) 'Genetic mapping, cloning and physiological aspects of the glucose kinase gene of *Streptomyces coelicolor*', *Molecular and General Genetics MGG*, 196(3), pp. 501–507.

Imriskova, I., Arreguín-Espinosa, R., Guzmán, S., Rodriguez-Sanoja, R., Langley, E. and Sanchez, S. (2005) 'Biochemical characterization of the glucose kinase from *Streptomyces coelicolor* compared to *Streptomyces peucetius var. caesius*', *Research in microbiology*, 156(3), pp. 361–366.

Ives, P. R. and Bushell, M. E. (1997) 'Manipulation of the physiology of clavulanic acid production in *Streptomyces clavuligerus*', *Microbiology*, 143(11), pp. 3573–3579.

Jackson, J. and Belkhir, L. (2018) 'Assigning firm-level GHGE reductions based on national goals-Mathematical model & empirical evidence', *Journal of Cleaner Production*, 170, pp. 76–84.

Jauri, P. V., Bakker, M. G., Salomon, C. E. and Kinkel, L. L. (2013) 'Subinhibitory antibiotic concentrations mediate nutrient use and competition among soil *Streptomyces*', *PLoS One*, 8(12), p. e81064.

Jensen, S. E. (2012) 'Biosynthesis of clavam metabolites', *Journal of Industrial Microbiology & Biotechnology*, 39(10), pp. 1407–1419.

Jensen, S. E., Elder, K. J., Aidoo, K. A. and Paradkar, A. S. (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.

Jensen, S. E., Paradkar, A. S., Mosher, R. H., Anders, C., Beatty, P. H., Brumlik, M. J., Griffin, A. and Barton, B. (2004) 'Five additional genes are involved in clavulanic acid biosynthesis in *Streptomyces clavuligerus*', *Antimicrobial agents and chemotherapy*, 48(1), pp. 192–202.

Jeong, Y., Kim, J.-N., Kim, M. W., Bucca, G., Cho, S., Yoon, Y. J., Kim, B.-G., Roe, J.-H., Kim, S. C. and Smith, C. P. (2016) 'The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2)', *Nature communications*, 7(1), pp. 1–11.

Jhang, J., Wolfe, S. and Demain, A. L. (1989) 'Phosphate regulation of ACV synthetase and

cephalosporin biosynthesis in *Streptomyces clavuligerus*', *FEMS microbiology letters*, 57(2), pp. 145–150.

Jin, W., Ryu, Y. G., Kang, S. G., Kim, S. K., Saito, N., Ochi, K., Lee, S. H. and Lee, K. J. (2004) 'Two relA/spoT homologous genes are involved in the morphological and physiological differentiation of *Streptomyces clavuligerus*', *Microbiology*, 150(5), pp. 1485–1493.

Jnawali, H. N., Yoo, J. C. and Sohng, J. K. (2011) 'Improvement of clavulanic acid production in *Streptomyces clavuligerus* by genetic manipulation of structural biosynthesis genes', *Biotechnology letters*, 33(6), pp. 1221–1226.

Jones, P. M. and George, A. M. (2004) 'The ABC transporter structure and mechanism: perspectives on recent research', *Cellular and Molecular Life Sciences CMLS*, 61(6), pp. 682–699.

De Jong, W., Manteca, A., Sanchez, J., Bucca, G., Smith, C. P., Dijkhuizen, L., Claessen, D. and Wösten, H. A. B. (2009) 'NepA is a structural cell wall protein involved in maintenance of spore dormancy in *Streptomyces coelicolor*', *Molecular microbiology*, 71(6), pp. 1591–1603.

Kämpfer, P. and Glaeser, S. P. (2012) 'Prokaryotic taxonomy in the sequencing era–the polyphasic approach revisited', *Environmental microbiology*, 14(2), pp. 291–317.

Kanehisa, M. and Goto, S. (2000) 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research*, 28(1), pp. 27–30.

Karandikar, A., Sharples, G. P. and Hobbs, G. (1997) 'Differentiation of *Streptomyces coelicolor* A3(2) under nitrate-limited conditions', *Microbiology*, 143(11), pp. 3581–3590.

Karimova, G. and Ladant, D. (2005) 'A bacterial two-hybrid system based on a cyclic AMP signaling cascade', *Protein-protein interactions*, pp. 499–515.

Karimova, G., Pidoux, J., Ullmann, A. and Ladant, D. (1998) 'A bacterial two-hybrid system based on a reconstituted signal transduction pathway', *Proceedings of the National Academy of Sciences*, 95(10), pp. 5752–5756.

Kato, J., Chi, W.-J., Ohnishi, Y., Hong, S.-K. and Horinouchi, S. (2005) 'Transcriptional control by A-factor of two trypsin genes in *Streptomyces griseus*', *Journal of Bacteriology*, 187(1), pp. 286–295.

Kato, J., Suzuki, A., Yamazaki, H., Ohnishi, Y. and Horinouchi, S. (2002) 'Control by A-factor of a metalloendopeptidase gene involved in aerial mycelium formation in *Streptomyces griseus*', *Journal of bacteriology*, 184(21), pp. 6016–6025.

Kayano, T., Fukumoto, H., Eddy, R. L., Fan, Y. S., Byers, M. G., Shows, T. and Bell, G. I. (1988) 'Evidence for a family of human glucose transporter-like proteins. Sequence and gene localization of a protein expressed in fetal skeletal muscle and other tissues.', *Journal of Biological Chemistry*, 263(30), pp. 15245–15248.

Kenig, M. and Reading, C. (1979) 'Holomycin and an antibiotic (MM 19290) related to tunicamycln, metabolites of *Streptomyces clavuligerus*', *The Journal of antibiotics*, 32(6), pp. 549–554.

Van Keulen, G. (2014) 'The Family *Streptomycetaceae*', in.

Khaleeli, N., Li, R. and Townsend, C. A. (1999) 'Origin of the β-lactam carbons in clavulanic acid from an unusual thiamine pyrophosphate-mediated reaction', *Journal of the American Chemical Society*, 121(39), pp. 9223–9224.

Kieser, T., Bibb, M., Buttner, M. and Chater, K. (2000) Practical *Streptomyces* Genetics.

Kim, B.-J., Kim, C.-J., Chun, J., Koh, Y.-H., Lee, S.-H., Hyun, J.-W., Cha, C.-Y. and Kook, Y.-H. (2004) 'Phylogenetic analysis of the genera *Streptomyces* and *Kitasatospora* based on partial RNA polymerase β-subunit gene (*rpoB*) sequences', *International journal of systematic and evolutionary microbiology*, 54(2), pp. 593–598.

Kim, H.-S. and Park, Y.-I. (2007) 'Lipase activity and tacrolimus production in *Streptomyces clavuligerus* CKD 1119 mutant strains', *Journal of microbiology and biotechnology*, 17(10), p. 1638.

Kim, H. S., Lee, Y. J., Lee, C. K., Choi, S. U., Yeo, S.-H., Hwang, Y. Il, Yu, T. S., Kinoshita, H. and Nihira, T. (2004) 'Cloning and characterization of a gene encoding the γ-butyrolactone autoregulator receptor from *Streptomyces clavuligerus*', *Archives of microbiology*, 182(1), pp. 44–50.

Kim, I., Lee, C., Kim, M., Kim, Jeong-Mok, Kim, Ji-Hye, Yim, H., Cha, S. and Kang, S. (2006) 'Crystal structure of the DNA-binding domain of BldD, a central regulator of aerial mycelium

formation in *Streptomyces coelicolor* A3(2)', *Molecular microbiology*, 60(5), pp. 1179–1193.

Kim, I. S. and Lee, K. J. (1995) 'Physiological roles of leupeptin and extracellular proteases in mycelium development of *Streptomyces exfoliatus* SMF13', *Microbiology*, 141(4), pp. 1017–1025.

Kim, S. B., Lonsdale, J., Seong, C.-N. and Goodfellow, M. (2003) '*Streptacidiphilus* gen. nov., acidophilic actinomycetes with wall chemotype I and emendation of the family *Streptomycetaceae* (Waksman and Henrici (1943) AL) emend. Rainey et al. 1997', *Antonie van Leeuwenhoek*, 83(2), pp. 107–116.

Kim, S. H., Lee, B.-R., Kim, J.-N. and Kim, B.-G. (2012) 'NdgR, a common transcriptional activator for methionine and leucine biosynthesis in *Streptomyces coelicolor*', *Journal of bacteriology*, 194(24), pp. 6837–6846.

Kimura, M. (1991) 'Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics', *Proceedings of the National Academy of Sciences*, 88(14), pp. 5969–5973.

Kino, T., Hatanaka, H., Miyata, S., Inamura, N., NISHIYAMA, M., YAJIMA, T., GOTO, T., OKUHARA, M., KOHSAKA, M. and AOKI, H. (1987) 'FK-506, a novel immunosuppressant isolated from a *Streptomyces*', *The Journal of antibiotics*, 40(9), pp. 1256–1265.

Kirk, S., Avignone-Rossa, C. A. and Bushell, M. E. (2000) 'Growth limiting substrate affects antibiotic production and associated metabolic fluxes in *Streptomyces clavuligerus*', *Biotechnology Letters*, 22(22), pp. 1803–1809.

Kizildogan, A. K., Jaccard, G. V., Mutlu, A., Sertdemir, I. and Öycengiy, G. (2017) 'Genetic engineering of an industrial strain of *Streptomyces clavuligerus* for further enhancement of clavulanic acid production', *Turkish Journal of Biology*, 41(2), pp. 342–353.

Korber-Irrgang, B. (2000) 'HIV signature and sequence variation analysis'.

Kuhstoss, S. and Rao, R. N. (1991) 'Analysis of the integration function of the streptomycete bacteriophage φC31', *Journal of molecular biology*, 222(4), pp. 897–908.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) 'MEGA X: molecular evolutionary genetics analysis across computing platforms', *Molecular biology and evolution*,

35(6), pp. 1547–1549.

Kumar, S., Stecher, G. and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets', *Molecular Biology and Evolution*, 33(7), pp. 1870–1874.

Kwakman, J. H. and Postma, P. W. (1994) 'Glucose kinase has a regulatory role in carbon catabolite repression in *Streptomyces coelicolor*.', *Journal of bacteriology*, 176(9), pp. 2694–2698.

De la Fuente, A., Lorenzana, L. M., Martín, J. F. and Liras, P. (2002) 'Mutants of *Streptomyces clavuligerus* with disruptions in different genes for clavulanic acid biosynthesis produce large amounts of holomycin: possible cross-regulation of two unrelated secondary metabolic pathways', *Journal of bacteriology*, 184(23), pp. 6559–6565.

Labeda, D. P., Dunlap, C. A., Rong, X., Huang, Y., Doroghazi, J. R., Ju, K.-S. and Metcalf, W. W. (2017) 'Phylogenetic relationships in the family Streptomycetaceae using multi-locus sequence analysis.', *Antonie van Leeuwenhoek*, 110(4), pp. 563–583.

Ladant, D. (1988) 'Interaction of *Bordetella pertussis* adenylate cyclase with calmodulin. Identification of two separated calmodulin-binding domains.', *Journal of Biological Chemistry*, 263(6), pp. 2612–2618.

Ladant, D., Michelson, S., Sarfati, R., Gilles, A. M., Predeleanu, R. and Barzu, O. (1989) 'Characterization of the calmodulin-binding and of the catalytic domains of *Bordetella pertussis* adenylate cyclase.', *Journal of Biological Chemistry*, 264(7), pp. 4015–4020.

Lal, R., Khanna, R., Kaur, H., Khanna, M., Dhingra, N., Lal, S., Gartemann, K.-H., Eichenlaub, R. and Ghosh, P. K. (1996) 'Engineering antibiotic producers to overcome the limitations of classical strain improvement programs', *Critical reviews in microbiology*, 22(4), pp. 201–255.

Lawlor, E. J., Baylis, H. A. and Chater, K. F. (1987) 'Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2).', *Genes & Development*, 1(10), pp. 1305–1310.

Le, S. Q. and Gascuel, O. (2008) 'An improved general amino acid replacement matrix', *Molecular biology and evolution*, 25(7), pp. 1307–1320.

Lebrihi, A., Germain, P. and Lefebvre, G. (1987) 'Phosphate repression of cephamycin and clavulanic acid production by *Streptomyces clavuligerus*', *Applied microbiology and biotechnology*, 26(2), pp. 130–135.

Lebrihi, A., Lefebvre, G. and Germain, P. (1988) 'Carbon catabolite regulation of cephamycin C and expandase biosynthesis in *Streptomyces clavuligerus*', *Applied microbiology and biotechnology*, 28(1), pp. 44–51.

Lechevalier, M. P. and Lechevalier, H. (1970) 'Chemical composition as a criterion in the classification of aerobic actinomycetes', *International Journal of Systematic and Evolutionary Microbiology*, 20(4), pp. 435–443.

Lee, B.-R., Bhatia, S. K., Song, H.-S., Kim, J., Kim, W., Park, H., Yoon, J.-J., Park, S.-H., Hwang, D. and Kim, B.-G. (2017) 'The role of NdgR in glycerol metabolism in *Streptomyces coelicolor*', *Bioprocess and biosystems engineering*, 40(10), pp. 1573–1580.

Lee, B.-R., Yi, D.-H., Song, E., Bhatia, S. K., Lee, J. H., Kim, Y.-G., Park, S.-H., Lee, Y. K., Kim, B.-G. and Yang, Y.-H. (2015) 'Increased vulnerability to physical stress by inactivation of NdgR in *Streptomyces coelicolor*', *Applied biochemistry and biotechnology*, 175(8), pp. 3673–3682.

Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tárraga, A., Cheng, Y., Cleland, I., Faruque, N., Goodgame, N. and Gibson, R. (2010) 'The European nucleotide archive', *Nucleic acids research*, 39(suppl_1), pp. D28–D31.

Leskiw, B. K., Bibb, M. J. and Chater, K. F. (1991) 'The use of a rare codon specifically during development?', *Molecular microbiology*, 5(12), pp. 2861–2867.

Li, R. and Townsend, C. A. (2006) 'Rational strain improvement for enhanced clavulanic acid production by genetic engineering of the glycolytic pathway in *Streptomyces clavuligerus*', *Metabolic Engineering*, 8(3), pp. 240–252.

Li, Z., Liu, X., Wang, Jingzhi, Wang, Y., Zheng, G., Lu, Y., Zhao, G. and Wang, Jin (2018) 'Insight into the Molecular Mechanism of the Transcriptional Regulation of *amtB* Operon in *Streptomyces coelicolor*', *Frontiers in microbiology*, 9, p. 264.

Liras, P. (2014) 'Holomycin, a dithiolopyrrolone compound produced by *Streptomyces clavuligerus*', *Applied microbiology and biotechnology*, 98(3), pp. 1023–1030.

Llamas-Ramírez, R., Takahashi-Iñiguez, T. and Flores, M. E. (2020) 'The phosphoenolpyruvate-pyruvate-oxaloacetate node genes and enzymes in *Streptomyces coelicolor* M-145', *International Microbiology*, pp. 1–11.

Lloyd, A. B. (1969) 'Dispersal of streptomycetes in air', *Microbiology*, 57(1), pp. 35–40.

Locci, R. and Schaal, K. P. (1980) 'Apical growth in facultative anaerobic actinomycetes as determined by immunofluorescent labeling', *Zentralblatt für Bakteriologie. 1. Abt. Originale A, Medizinische Mikrobiologie, Infektionskrankheiten und Parasitologie*, 246(1), pp. 112–118.

López-García, M. T., Santamarta, I. and Liras, P. (2010) 'Morphological differentiation and clavulanic acid formation are affected in a *Streptomyces clavuligerus adpA*-deleted mutant', *Microbiology*, 156(8), pp. 2354–2365.

Lu, X., Liu, X., Chen, Z., Li, J., van Wezel, G. P., Chen, W. and Wen, Y. (2020) 'The ROK-family Regulator Rok7B7 Directly Controls Carbon Catabolite Repression, Antibiotic Biosynthesis, and Morphological Development in *Streptomyces avermitilis*', *Environmental Microbiology*.

Lübbe, C., Wolfe, S. and Demain, A. L. (1985) 'Repression and inhibition of cephalosporin synthetases in *Streptomyces clavuligerus* by inorganic phosphate', *Archives of microbiology*, 140(4), pp. 317–320.

Lyddiard, D., Jones, G. and Greatrex, B. (2016) 'Keeping it simple: lessons from the golden era of antibiotic discovery', *FEMS Microbiology Letters*, 363(8).

Lynch, M. and Conery, J. S. (2000) 'The evolutionary fate and consequences of duplicate genes', *science*, 290(5494), pp. 1151–1155.

MacKenzie, A. K., Kershaw, N. J., Hernandez, H., Robinson, C. V, Schofield, C. J. and Andersson, I. (2007) 'Clavulanic Acid Dehydrogenase: Structural and Biochemical Analysis of the Final Step in the Biosynthesis of the β-Lactamase Inhibitor Clavulanic Acid†', *Biochemistry*, 46(6), pp. 1523–1533.

MacNeil, D. J., Gewain, K. M., Ruby, C. L., Dezeny, G., Gibbons, P. H. and MacNeil, T. (1992) 'Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector', *Gene*, 111(1), pp. 61–68.

Mahr, K., van Wezel, G. P., Svensson, C., Krengel, U., Bibb, M. J. and Titgemeyer, F. (2000)

'Glucose kinase of *Streptomyces coelicolor* A3(2): large-scale purification and biochemical analysis', *Antonie Van Leeuwenhoek*, 78(3), pp. 253–261.

Maiden, M. C. J., Davis, E. O., Baldwin, S. A., Moore, D. C. M. and Henderson, P. J. F. (1987) 'Mammalian and bacterial sugar transport proteins are homologous', *Nature*, 325(6105), pp. 641–643.

Makitrynskyy, R., Ostash, B., Tsypik, O., Rebets, Y., Doud, E., Meredith, T., Luzhetskyy, A., Bechthold, A., Walker, S. and Fedorenko, V. (2013) 'Pleiotropic regulatory genes *bldA*, *adpA* and *absB* are implicated in production of phosphoglycolipid antibiotic moenomycin', *Open biology*, 3(10), p. 130121.

Manteca, A., Fernandez, M. and Sanchez, J. (2005) 'A death round affecting a young compartmentalized mycelium precedes aerial mycelium dismantling in confluent surface cultures of *Streptomyces antibioticus*', *Microbiology*, 151(11), pp. 3689–3697.

Marger, M. D. and Saier Jr, M. H. (1993) 'A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport', *Trends in biochemical sciences*, 18(1), pp. 13–20.

Martín, J. F. (2004) 'Phosphate control of the biosynthesis of antibiotics and other secondary metabolites is mediated by the PhoR-PhoP system: an unfinished story', *Journal of bacteriology*, 186(16), pp. 5197–5201.

Martín, M. C., Díaz, L. A., Manzanal, M. B. and Hardisson, C. (1986) 'Role of trehalose in the spores of *Streptomyces*', *FEMS microbiology letters*, 35(1), pp. 49–54.

Martínez-Burgo, Y., Álvarez-Álvarez, R., Rodríguez-García, A. and Liras, P. (2015) 'The pathway-specific regulator ClaR of *Streptomyces clavuligerus* has a global effect on the expression of genes for secondary metabolism and differentiation', *Applied and environmental microbiology*, 81(19), pp. 6637–6648.

Martínez-Castro, M., Salehi-Najafabadi, Z., Romero, F., Pérez-Sanchiz, R., Fernández-Chimeno, R. I., Martín, J. F. and Barreiro, C. (2013) 'Taxonomy and chemically semi-defined media for the analysis of the tacrolimus producer "*Streptomyces tsukubaensis*"', *Applied microbiology and biotechnology*, 97(5), pp. 2139–2152.

Mazodier, P., Petter, R. and Thompson, C. (1989) 'Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species.', *Journal of Bacteriology*, 171(6), pp. 3583–3585.

Mbow, C., Rosenzweig, C., Barioni, L. G., Benton, T. G., Herrero, M., Krishnapillai, M., Liwenga, E., Pradhan, P., Rivera-Ferre, M. G. and Sapkota, T. B. (2019) 'Food security', in *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security and greenhouse gas fluxes in terrestrial ecosystems*.

McBride, M. J. and Ensign, J. C. (1987) 'Metabolism of endogenous trehalose by *Streptomyces griseus* spores and by spores or cells of other actinomycetes.', *Journal of bacteriology*, 169(11), pp. 5002–5007.

McBride, M. J. and Ensign, J. C. (1990) 'Regulation of trehalose metabolism by *Streptomyces griseus* spores.', *Journal of Bacteriology*, 172(7), pp. 3637 LP – 3643.

McCormick, J. R. and Flärdh, K. (2012) 'Signals and regulators that govern *Streptomyces* development', *FEMS microbiology reviews*, 36(1), pp. 206–231.

Medema, Marnix H, Trefzer, A., Kovalchuk, A., van den Berg, M., Müller, U., Heijne, W., Wu, L., Alam, M. T., Ronning, C. M. and Nierman, W. C. (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.

Medema, Marnix H., Trefzer, A., Kovalchuk, A., van den Berg, M., Müller, U., Heijne, W., Wu, L., Alam, M. T., Ronning, C. M., Nierman, W. C., Bovenberg, R. A. L., Breitling, R. and Takano, E. (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(1), pp. 212–224.

van der Meij, A., Worsley, S. F., Hutchings, M. I. and van Wezel, G. P. (2017) 'Chemical ecology of antibiotic production by actinomycetes', *FEMS Microbiology Reviews*, 41(3), pp. 392–416.

Mellado, E., Lorenzana, L. M., Rodríguez-Sáiz, M., Díez, B., Liras, P. and Barredo, J. L. (2002) 'The clavulanic acid biosynthetic cluster of *Streptomyces clavuligerus*: genetic organization of the region upstream of the car gene', *Microbiology*, 148(5), pp. 1427–1438.

Merrick, M. J. (1976) 'A morphological and genetic mapping study of bald colony mutants of *Streptomyces coelicolor*', *Microbiology*, 96(2), pp. 299–315.

Mikulík, K., Bobek, J., Bezoušková, S., Benada, O. and Kofroňová, O. (2002) 'Expression of proteins and protein kinase activity during germination of aerial spores of *Streptomyces granaticolor*', *Biochemical and biophysical research communications*, 299(2), pp. 335–342.

Miller, J. H. (1972) 'Assay of B-galactosidase In: Experiments in molecular genetics'.

Minambres, B., Reglero, A. and Luengo, J. M. (1992) 'Characterization of an inducible transport system for glycerol in *Streptomyces clavuligerus*', *The journal of antibiotics*, 45(2), pp. 269–277.

Miyazono, K., Tabei, N., Morita, S., Ohnishi, Y., Horinouchi, S. and Tanokura, M. (2012) 'Substrate recognition mechanism and substrate-dependent conformational changes of an ROK family glucokinase from *Streptomyces griseus*.', *Journal of bacteriology*, 194(3), pp. 607–616.

Monod, J. (1942) 'Studies on the growth of bacterial cultures', *Hermann and Cie, Paris, France*.

Mueckler, M., Caruso, C., Baldwin, S. A., Panico, M., Blench, I., Morris, H. R., Allard, W. J., Lienhard, G. E. and Lodish, H. F. (1985) 'Sequence and structure of a human glucose transporter', *Science*, 229(4717), pp. 941–945.

Murakami, T., Holt, T. G. and Thompson, C. J. (1989) 'Thiostrepton-induced gene expression in *Streptomyces lividans*.', *Journal of bacteriology*, 171(3), pp. 1459–1466.

Nei, M. and Gojobori, T. (1986) 'Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions.', *Molecular biology and evolution*, 3(5), pp. 418–426.

Netolitzky, D. J., Wu, X., Jensen, S. E. and Roy, K. L. (1995) 'Giant linear plasmids of β-lactam antibiotic producing *Streptomyces*', *FEMS microbiology letters*, 131(1), pp. 27–34.

Ng, W.-L., Kazmierczak, K. M., Robertson, G. T., Gilmour, R. and Winkler, M. E. (2003) 'Transcriptional regulation and signature patterns revealed by microarray analyses of *Streptococcus pneumoniae* R6 challenged with sublethal concentrations of translation

inhibitors', *Journal of bacteriology*, 185(1), pp. 359–370.

Nguyen, K. T., Tenor, J., Stettler, H., Nguyen, L. T., Nguyen, L. D. and Thompson, C. J. (2003) 'Colonial differentiation in *Streptomyces coelicolor* depends on translation of a specific codon within the *adpA* gene', *Journal of bacteriology*, 185(24), pp. 7291–7296.

Nicholson, N. H., Baggaley, K. H., Cassels, R., Davison, M., Elson, S. W., Fulston, M., Tyler, J. W. and Woroniecki, S. R. (1994) 'Evidence that the immediate biosynthetic precursor of clavulanic acid is its N-aldehyde analogue', *Journal of the Chemical Society, Chemical Communications*, (11), pp. 1281–1282.

Nieminen, L., Webb, S., Smith, M. C. M. and Hoskisson, P. A. (2013) 'A flexible mathematical model platform for studying branching networks: experimentally validated using the model actinomycete, *Streptomyces coelicolor*', *PloS one*. 2013/02/18, 8(2), pp. e54316–e54316.

Nodwell, J. R. and Losick, R. (1998) 'Purification of an extracellular signaling molecule involved in production of aerial mycelium by *Streptomyces coelicolor*', *Journal of bacteriology*, 180(5), pp. 1334–1337.

Nodwell, J. R., McGovern, K. and Losick, R. (1996) 'An oligopeptide permease responsible for the import of an extracellular signal governing aerial mycelium formation in *Streptomyces coelicolor*', *Molecular microbiology*, 22(5), pp. 881–893.

Nodwell, J. R., Yang, M., Kuo, D. and Losick, R. (1999) 'Extracellular complementation and the identification of additional genes involved in aerial mycelium formation in *Streptomyces coelicolor*', *Genetics*, 151(2), pp. 569–584.

Noens, E. E. E., Mersinias, V., Traag, B. A., Smith, C. P., Koerten, H. K. and van Wezel, G. P. (2005) 'SsgA-like proteins determine the fate of peptidoglycan during sporulation of *Streptomyces coelicolor*', *Molecular microbiology*, 58(4), pp. 929–944.

Noens, E. E., Mersinias, V., Willemse, J., Traag, B. A., Laing, E., Chater, K. F., Smith, C. P., Koerten, H. K. and van Wezel, G. P. (2007) 'Loss of the controlled localization of growth stage-specific cell-wall synthesis pleiotropically affects developmental gene expression in an ssgA mutant of Streptomyces coelicolor', *Molecular microbiology*, 64(5), pp. 1244–1259.

Nothaft, H., Dresel, D., Willimek, A., Mahr, K., Niederweis, M. and Titgemeyer, F. (2003) 'The

phosphotransferase system of *Streptomyces coelicolor* is biased for N-acetylglucosamine metabolism', *Journal of bacteriology*, 185(23), pp. 7019–7023.

Nothaft, H., Parche, S., Kamionka, A. and Titgemeyer, F. (2003) '*In vivo* analysis of HPr reveals a fructose-specific phosphotransferase system that confers high-affinity uptake in *Streptomyces coelicolor*', *Journal of bacteriology*, 185(3), pp. 929–937.

Nothaft, H., Rigali, S., Boomsma, B., Swiatek, M., McDowall, K. J., Van Wezel, G. P. and Titgemeyer, F. (2010) 'The permease gene *nagE2* is the key to N-acetylglucosamine sensing and utilization in *Streptomyces coelicolor* and is subject to multi-level control', *Molecular microbiology*, 75(5), pp. 1133–1144.

Novotna, J., Vohradsky, J., Berndt, P., Gramajo, H., Langen, H., Li, X., Minas, W., Orsaria, L., Roeder, D. and Thompson, C. J. (2003) 'Proteomic studies of diauxic lag in the differentiating prokaryote *Streptomyces coelicolor* reveal a regulatory network of stress-induced proteins and central metabolic enzymes', *Molecular microbiology*, 48(5), pp. 1289–1303.

O'Neill, J. (2016) 'Tackling drug-resistant infections globally: Final report and recommendations.', *HM Government and Welcome Trust: UK*.

Ochman, H., Lawrence, J. G. and Groisman, E. A. (2000) 'Lateral gene transfer and the nature of bacterial innovation', *nature*, 405(6784), pp. 299–304.

Ohnishi, Y., Kameyama, S., Onaka, H. and Horinouchi, S. (1999) 'The A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus*: identification of a target gene of the A-factor receptor', *Molecular microbiology*, 34(1), pp. 102–111.

Ohnishi, Y., Yamazaki, H., Kato, J.-Y., Tomono, A. and Horinouchi, S. (2005) 'AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in *Streptomyces griseus*', *Bioscience, biotechnology, and biochemistry*, 69(3), pp. 431–439.

Olsen, G. (1990) 'Newick's 8: 45', *Tree format standard*.

Omasits, U., Ahrens, C. H., Müller, S. and Wollscheid, B. (2013) 'Protter: interactive protein feature visualization and integration with experimental proteomic data', *Bioinformatics*, 30(6), pp. 884–886.

Omura, S., Takahashi, Y., Iwai, Y. and Tanaka, H. (1982) '*Kitasatosporia*, a new genus of the order Actinomycetales', *The Journal of antibiotics*, 35(8), pp. 1013–1019.

Onaka, H., Ando, N., Nihira, T., Yamada, Y., Beppu, T. and Horinouchi, S. (1995) 'Cloning and characterization of the A-factor receptor gene from *Streptomyces griseus*.', *Journal of Bacteriology*, 177(21), pp. 6083–6092.

Onaka, H. and Horinouchi, S. (1997) 'DNA-binding activity of the A-factor receptor protein and its recognition DNA sequences', *Molecular microbiology*, 24(5), pp. 991–1000.

Onaka, H., Nakagawa, T. and Horinouchi, S. (1998) 'Involvement of two A-factor receptor homologues in *Streptomyces coelicolor* A3(2) in the regulation of secondary metabolism and morphogenesis', *Molecular microbiology*, 28(4), pp. 743–753.

Onaka, H., Sugiyama, M. and Horinouchi, S. (1997) 'A mutation at proline-115 in the A-factor receptor protein of *Streptomyces griseus* abolishes DNA-binding ability but not ligand-binding ability.', *Journal of bacteriology*, 179(8), pp. 2748–2752.

Ordóñez-Robles, M., Santos-Beneit, F., Albillos, S. M., Liras, P., Martín, J. F. and Rodríguez-García, A. (2017) '*Streptomyces tsukubaensis* as a new model for carbon repression: transcriptomic response to tacrolimus repressing carbon sources', *Applied microbiology and biotechnology*, 101(22), pp. 8181–8195.

Ostash, B., Yushchuk, O., Tistechok, S., Mutenko, H., Horbal, L., Muryn, A., Dacyuk, Y., Kalinowski, J., Luzhetskyy, A. and Fedorenko, V. (2015) 'The *adpA*-like regulatory gene from *Actinoplanes teichomyceticus*: *in silico* analysis and heterologous expression', *World Journal of Microbiology and Biotechnology*, 31(8), pp. 1297–1301.

Ota, T. and Nei, M. (1994) 'Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site.', *Molecular biology and evolution*, 11(4), pp. 613–619.

Pan, Y., Liu, G., Yang, H., Tian, Y. and Tan, H. (2009) 'The pleiotropic regulator AdpA-L directly controls the pathway-specific activator of nikkomycin biosynthesis in *Streptomyces ansochromogenes*', *Molecular microbiology*, 72(3), pp. 710–723.

Paradkar, A. (2013) 'Clavulanic acid production by *Streptomyces clavuligerus*: Biogenesis, regulation and strain improvement', *Journal of Antibiotics*, 66(7), pp. 411–420.

Paradkar, A. S., Aidoo, K. A. and Jensen, S. E. (1998) 'A pathway-specific transcriptional activator regulates late steps of clavulanic acid biosynthesis in *Streptomyces clavuligerus*', *Molecular microbiology*, 27(4), pp. 831–843.

Paradkar, A. S. and Jensen, S. E. (1995) 'Functional analysis of the gene encoding the clavaminate synthase 2 isoenzyme involved in clavulanic acid biosynthesis in *Streptomyces clavuligerus.*', *Journal of bacteriology*, 177(5), pp. 1307–1314.

Parajuli, N., Viet, H. T., Ishida, K., Tong, H. T., Lee, H. C., Liou, K. and Sohng, J. K. (2005) 'Identification and characterization of the *afsR* homologue regulatory gene from *Streptomyces peucetius* ATCC 27952', *Research in Microbiology*, 156(5–6), pp. 707–712.

Parche, S., Nothaft, H., Kamionka, A. and Titgemeyer, F. (2000) 'Sugar uptake and utilisation in *Streptomyces coelicolor*: a PTS view to the genome', *Antonie Van Leeuwenhoek*, 78(3–4), pp. 243–251.

Parche, S., Schmid, R. and Titgemeyer, F. (1999) 'The phosphotransferase system (PTS) of *Streptomyces coelicolor*: identification and biochemical analysis of a histidine phosphocarrier protein HPr encoded by the gene ptsH', *European journal of biochemistry*, 265(1), pp. 308–317.

Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. and Yeates, T. O. (1999) 'Assigning protein functions by comparative genome analysis: protein phylogenetic profiles', *Proceedings of the National Academy of Sciences*, 96(8), pp. 4285–4288.

Pérez-Llarena, F. J., Liras, P., Rodriguez-Garcia, A. and Martin, J. F. (1997) 'A regulatory gene (ccaR) required for cephamycin and clavulanic acid production in *Streptomyces clavuligerus*: amplification results in overproduction of both beta-lactam compounds.', *Journal of bacteriology*, 179(6), pp. 2053–2059.

Pérez-Redondo, R., Rodríguez-García, A., Botas, A., Santamarta, I., Martín, J. F. and Liras, P. (2012) 'ArgR of *Streptomyces coelicolor* is a versatile regulator', *PloS one*, 7(3), p. e32697.

Pérez-Redondo, R., Rodríguez-García, A., Martín, J. F. and Liras, P. (1998) 'The *claR* gene of *Streptomyces clavuligerus*, encoding a LysR-type regulatory protein controlling clavulanic acid biosynthesis, is linked to the clavulanate-9-aldehyde reductase (*car*) gene', *Gene*, 211(2), pp. 311–321.

Pérez-Redondo, R., Santamarta, I., Bovenberg, R., Martín, J. F. and Liras, P. (2010) 'The enigmatic lack of glucose utilization in *Streptomyces clavuligerus* is due to inefficient expression of the glucose permease gene', *Microbiology*, 156(5), pp. 1527–1537.

Piette, A., Derouaux, A., Gerkens, P., Noens, E. E. E., Mazzucchelli, G., Vion, S., Koerten, H. K., Titgemeyer, F., De Pauw, E. and Leprince, P. (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.

Pinilla, L., Toro, L. F., Laing, E., Alzate, J. F. and Ríos-Estepa, R. (2019) 'Comparative Transcriptome Analysis of *Streptomyces clavuligerus* in Response to Favorable and Restrictive Nutritional Conditions', *Antibiotics*, 8(3), p. 96.

Pope, M. K., Green, B. D. and Westpheling, J. (1996) 'The *bld* mutants of *Streptomyces coelicolor* are defective in the regulation of carbon utilization, morphogenesis and cell–cell signalling', *Molecular microbiology*, 19(4), pp. 747–756.

Porto, A. L. F., Campos-Takaki, G. M. and Lima Filho, J. L. (1996) 'Effects of culture conditions on protease production by *Streptomyces clavuligerus* growing on soy bean flour medium', *Applied Biochemistry and Biotechnology*, 60(2), pp. 115–122.

Preston, G. M. and Agre, P. (1991) 'Isolation of the cDNA for erythrocyte integral membrane protein of 28 kilodaltons: member of an ancient channel family', *Proceedings of the National Academy of Sciences*, 88(24), pp. 11110–11114.

Puglia, A. M., Vohradsky, J. and Thompson, C. J. (1995) 'Developmental control of the heat-shock stress regulon in *Streptomyces coelicolor*', *Molecular microbiology*, 17(4), pp. 737–746.

Qin, R., Zhong, C., Zong, G., Fu, J., Pang, X. and Cao, G. (2017) 'Improvement of clavulanic acid production in *Streptomyces clavuligerus* F613-1 by using a *claR-neo* reporter strategy', *Electronic Journal of Biotechnology*, 28, pp. 41–46.

Qin, Z., Munnoch, J. T., Devine, R., Holmes, N. A., Seipke, R. F., Wilkinson, K. A., Wilkinson, B. and Hutchings, M. I. (2017) 'Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African Tetraponera plant-ants', *Chemical science*, 8(4), pp. 3218–3227.

Quackenbush, J. (2002) 'Microarray data normalization and transformation', *Nature genetics*, 32(4), pp. 496–501.

Rahlwes, K. C., Sparks, I. L. and Morita, Y. S. (2019) 'Cell walls and membranes of Actinobacteria', in *Bacterial Cell Walls and Membranes*, pp. 417–469.

Ramirez-Malule, H., Junne, S., Cruz-Bournazou, M. N., Neubauer, P. and Ríos-Estepa, R. (2018) '*Streptomyces clavuligerus* shows a strong association between TCA cycle intermediate accumulation and clavulanic acid biosynthesis', *Applied microbiology and biotechnology*, 102(9), pp. 4009–4023.

Ray, S. D., Farris, F. F. and Hartmann, A. C. (2014) 'Hormesis', in Wexler, P. B. T.-E. of T. (Third E. (ed.), pp. 944–948.

Razzaq, A., Shamsi, S., Ali, A., Ali, Q., Sajjad, M., Malik, A. and Ashraf, M. (2019) 'Microbial proteases applications', *Frontiers in bioengineering and biotechnology*, 7, p. 110.

Reading, C. and Cole, M. (1977) 'Clavulanic acid: a beta-lactamase-inhibiting beta-lactam from *Streptomyces clavuligerus*', *Antimicrobial agents and chemotherapy*, 11(5), pp. 852–857.

Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I. and Saier Jr, M. H. (2012) 'The major facilitator superfamily (MFS) revisited', *The FEBS journal*, 279(11), pp. 2022–2035.

Reizer, J., Romano, A. H. and Deutscher, J. (1993) 'The role of phosphorylation of HPr, a phosphocarrier protein of the phosphotransferase system, in the regulation of carbon metabolism in gram-positive bacteria', *Journal of cellular biochemistry*, 51(1), pp. 19–24.

Rigali, S., Nothaft, H., Noens, E. E. E., Schlicht, M., Colson, S., Müller, M., Joris, B., Koerten, H. K., Hopwood, D. A. and Titgemeyer, F. (2006) 'The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development', *Molecular microbiology*, 61(5), pp. 1237–1251.

Rigali, S., Schlicht, M., Hoskisson, P., Nothaft, H., Merzbacher, M., Joris, B. and Titgemeyer, F. (2004) 'Extending the classification of bacterial transcription factors beyond the helix–turn–helix motif as an alternative approach to discover new cis/trans relationships', *Nucleic acids*

research, 32(11), pp. 3418–3426.

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A. W., Hopwood, D. A. and van Wezel, G. P. (2008) 'Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by Streptomyces', *EMBO reports*, 9(7), pp. 670–675.

Rodnina, M. V, Savelsbergh, A., Katunin, V. I. and Wintermeyer, W. (1997) 'Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome', *Nature*, 385(6611), pp. 37–41.

Rodnina, M. V, Savelsbergh, A., Matassova, N. B., Katunin, V. I., Semenkov, Y. P. and Wintermeyer, W. (1999) 'Thiostrepton inhibits the turnover but not the GTPase of elongation factor G on the ribosome', *Proceedings of the National Academy of Sciences*, 96(17), pp. 9586–9590.

Rodríguez-García, A., de La Fuente, A., Pérez-Redondo, R., Martín, J. F. and Liras, P. (2000) 'Characterization and expression of the arginine biosynthesis gene cluster of *Streptomyces clavuligerus*', *Journal of molecular microbiology and biotechnology*, 2(4), pp. 543–550.

Rodríguez-García, A., Martín, J. F. and Liras, P. (1995) 'The *argG* gene of *Streptomyces clavuligerus* has low homology to unstable arG from other actinomycetes: effect of amplificatio on clavulcanic acid biosynthesis', *Gene*, 167(1–2), pp. 9–15.

Rodríguez-García, A., Sola-Landa, A., Apel, K., Santos-Beneit, F. and Martín, J. F. (2009) 'Phosphate control over nitrogen metabolism in *Streptomyces coelicolor*: direct and indirect negative control of *glnR*, *glnA*, *glnII* and *amtB* expression by the response regulator PhoP', *Nucleic Acids Research*, 37(10), pp. 3230–3242.

Rodríguez-García, A., Ludovice, M., Martín, J. F. and Liras, P. (1997) 'Arginine boxes and the *argR* gene in *Streptomyces clavuligerus*: evidence for a clear regulation of the arginine pathway', *Molecular microbiology*, 25(2), pp. 219–228.

Romero-Rodríguez, A., Rocha, D., Ruiz-Villafán, B., Guzmán-Trampe, S., Maldonado-Carmona, N., Vázquez-Hernández, M., Zelarayán, A., Rodríguez-Sanoja, R. and Sánchez, S. (2017) 'Carbon catabolite regulation in *Streptomyces*: new insights and lessons learned', *World Journal of Microbiology and Biotechnology*, 33(9), p. 162.

Romero-Rodríguez, A., Rocha, D., Ruiz-Villafan, B., Tierrafría, V., Rodríguez-Sanoja, R., Segura-González, D. and Sánchez, S. (2016) 'Transcriptomic analysis of a classical model of carbon catabolite regulation in *Streptomyces coelicolor*', *BMC microbiology*, 16(1), p. 77.

Romero, J., Liras, P. and Martín, J. F. (1984) 'Dissociation of cephamycin and clavulanic acid biosynthesis in *Streptomyces clavuligerus*', *Applied Microbiology and Biotechnology*, 20(5), pp. 318–325.

Romm, J. (2018) 'Climate Change: What Everyone Needs to Know. ', *Oxford University Press*.

Roseman, S. (1969) 'The transport of carbohydrates by a bacterial phosphotransferase system', *The Journal of general physiology*, 54(1), pp. 138–184.

Roy, A., Kucukural, A. and Zhang, Y. (2010) 'I-TASSER: a unified platform for automated protein structure and function prediction', *Nature protocols*, 5(4), p. 725.

Ruban-Ośmiałowska, B., Jakimowicz, D., Smulczyk-Krawczyszyn, A., Chater, K. F. and Zakrzewska-Czerwińska, J. (2006) 'Replisome localization in vegetative and aerial hyphae of *Streptomyces coelicolor*', *Journal of bacteriology*, 188(20), pp. 7311–7316.

Ryu, Y.-G., Butler, M. J., Chater, K. F. and Lee, K. J. (2006) 'Engineering of Primary Carbohydrate Metabolism for Increased Production of Actinorhodin in *Streptomyces coelicolor*', *Applied and Environmental Microbiology*, 72(11), pp. 7132 LP – 7139.

Sabater, B. and Asensio, C. (1973) 'Transport of hexoses in *Streptomyces violaceoruber*', *European journal of biochemistry*, 39(1), pp. 201–205.

Saier Jr, M. H., Reddy, V. S., Tamang, D. G. and Västermark, A. (2014) 'The transporter classification database', *Nucleic acids research*. 2013/11/12, 42(Database issue), pp. D251–D258.

Saier Jr, M. H. and Reizer, J. (1992) 'Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate: sugar phosphotransferase system.', *Journal of bacteriology*, 174(5), p. 1433.

Saier, M. H. (1989) 'Protein phosphorylation and allosteric control of inducer exclusion and catabolite repression by the bacterial phosphoenolpyruvate: sugar phosphotransferase system.', *Microbiology and Molecular Biology Reviews*, 53(1), pp. 109–120.

Saier, M. H. J., Reddy, V. S., Tamang, D. G. and Vastermark, A. (2014) 'The transporter classification database.', *Nucleic acids research*, 42(Database issue), pp. D251-8.

Saier, M. H. J., Reddy, V. S., Tsu, B. V, Ahmed, M. S., Li, C. and Moreno-Hagelsieb, G. (2016) 'The Transporter Classification Database (TCDB): recent advances.', *Nucleic acids research*, 44(D1), pp. D372-9.

Saier, M. H. J., Tran, C. V and Barabote, R. D. (2006) 'TCDB: the Transporter Classification Database for membrane transport protein analyses and information.', *Nucleic acids research*, 34(Database issue), pp. D181-6.

Saier, M. H. J., Yen, M. R., Noto, K., Tamang, D. G. and Elkan, C. (2009) 'The Transporter Classification Database: recent advances.', *Nucleic acids research*, 37(Database issue), pp. D274-8.

Saito, A., Shinya, T., Miyamoto, K., Yokoyama, T., Kaku, H., Minami, E., Shibuya, N., Tsujibo, H., Nagata, Y., Ando, A., Fujii, T. and Miyashita, K. (2007) 'The *dasABC* Gene Cluster, Adjacent to *dasR*, Encodes a Novel ABC Transporter for the Uptake of *N,N'*-Diacetylchitobiose in Streptomyces coelicolor A3(2)', *Applied and Environmental Microbiology*, 73(9), pp. 3000 LP – 3008.

Salamov, V. S. A., Solovyevand, A., Solovyev, V. and Salamov, A. (2011) 'Automatic annotation of microbial genomes and metagenomic sequences', *Metagenomics and its applications in agriculture. Nova Science Publishers, Hauppauge*, pp. 61–78.

Salas, J. A., Quiros, L. M. and Hardisson, C. (1984) 'Pathways of glucose catabolism during germination of *Streptomyces* spores', *FEMS microbiology letters*, 22(3), pp. 229–233.

Salehghamari, E., Hamedi, J., Elahi, E., Sepehrizadeh, Z., Sadeghi, M. and Muth, G. (2012) 'Prediction of the pho regulon in *Streptomyces clavuligerus* DSM 738.', *Microbiologica-Quarterly Journal of Microbiological Sciences*, 35(4), p. 447.

Salowe, S. P., Krol, W. J., Iwata-Reuyl, D. and Townsend, C. A. (1991) 'Elucidation of the order of oxidations and identification of an intermediate in the multistep clavaminate synthase reaction', *Biochemistry*, 30(8), pp. 2281–2292.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual.

Sambrook, Joseph, Russell, D. W. and Sambrook, J (2006) The condensed protocols: from molecular cloning: a laboratory manual.

Sánchez, L. and Brana, A. F. (1996) 'Cell density influences antibiotic biosynthesis in *Streptomyces clavuligerus*', *Microbiology*, 142(5), pp. 1209–1220.

Sanchez, S., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Sánchez, M., Rocha, D., Sánchez, B., Ávalos, M. and Guzmán-Trampe, S. (2010) 'Carbon source regulation of antibiotic production', *The Journal of antibiotics*, 63(8), pp. 442–459.

Sanno, Y., Wilson, T. H. and Lin, E. C. C. (1968) 'Control of permeation to glycerol in cells of *Escherichia coli*', *Biochemical and biophysical research communications*, 32(2), pp. 344–349.

Santamarta, I., López-García, M. T., Pérez-Redondo, R., Koekman, B., Martín, J. F. and Liras, P. (2007) 'Connecting primary and secondary metabolism: AreB, an IclR-like protein, binds the ARE *ccaR* sequence of *S. clavuligerus* and modulates leucine biosynthesis and cephamycin C and clavulanic acid production', *Molecular microbiology*, 66(2), pp. 511–524.

Santamarta, I., Pérez-Redondo, R., Lorenzana, L. M., Martìn, J. F. and Liras, P. (2005) 'Different proteins bind to the butyrolactone receptor protein ARE sequence located upstream of the regulatory *ccaR* gene of *Streptomyces clavuligerus*', *Molecular microbiology*, 56(3), pp. 824–835.

Santamarta, I., Rodríguez-García, A., Pérez-Redondo, R., Martín, J. F. and Liras, P. (2002) 'CcaR is an autoregulatory protein that binds to the *ccaR* and *cefD-cmcI* promoters of the cephamycin C-clavulanic acid cluster in *Streptomyces clavuligerus*', *Journal of bacteriology*, 184(11), pp. 3106–3113.

Saudagar, P. S., Survase, S. A. and Singhal, R. S. (2008) 'Clavulanic acid: a review', *Biotechnology advances*, 26(4), pp. 335–351.

Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., Funk, K., Ketter, A., Kim, S. and Kimchi, A. (2020) 'Database resources of the national center for biotechnology information', *Nucleic acids research*, 48(D1), p. D9.

Schatz, A., Bugle, E. and Waksman, S. A. (1944) 'Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.∗†', *Proceedings of the*

Society for Experimental Biology and Medicine, 55(1), pp. 66–69.

Schlatter, D., Fubuh, A., Xiao, K., Hernandez, D., Hobbie, S. and Kinkel, L. (2009) 'Resource amendments influence density and competitive phenotypes of *Streptomyces* in soil', *Microbial ecology*, 57(3), pp. 413–420.

Schniete, J. K. (2015) Systems Biology Approach for Metabolic Engineering in Streptomyces: The Phosphoenolpyruvate-Pyruvate-Oxaloacetate Node.

Schniete, J. K., Cruz-Morales, P., Selem-Mojica, N., Fernández-Martínez, L. T., Hunter, I. S., Barona-Gómez, F. and Hoskisson, P. A. (2018) 'Expanding primary metabolism helps generate the metabolic robustness to facilitate antibiotic biosynthesis in *Streptomyces*', *MBio*, 9(1), pp. e02283-17.

Schniete, J. K., Selem-Mojica, N., Birke, A. S., Cruz-Morales, P., Hunter, I. S., Barona-Gomez, F. and Hoskisson, P. A. (2021) 'ActDES–a curated actinobacterial database for evolutionary studies', *Microbial genomics*, p. 498.

Schrempf, H. (2001) 'Recognition and degradation of chitin by *streptomycetes*', *Antonie van Leeuwenhoek*, 79(3), pp. 285–289.

Segura, D. (1996) '*Streptomyces* mutants insensitive to glucose repression showed deregulation of primary and secondary metabolism', *ASPA J. Mol. Biol. Biotechnol.*, 4, pp. 30–36.

Sen, A., Daubin, V., Abrouk, D., Gifford, I., Berry, A. M. and Normand, P. (2014) 'Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders "Frankiales" and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. an', *International journal of systematic and evolutionary microbiology*, 64(Pt 11), pp. 3821–3832.

Seno, E. T., Bruton, C. J. and Chater, K. F. (1984) 'The glycerol utilization operon of *Streptomyces coelicolor*: genetic mapping of gyl mutations and the analysis of cloned gyl DNA', *Molecular and General Genetics MGG*, 193(1), pp. 119–128.

Seno, E. T. and Chater, K. F. (1983) 'Glycerol catabolic enzymes and their regulation in wild-type and mutant strains of *Streptomyces coelicolor* A3 (2)', *Microbiology*, 129(5), pp. 1403–

1413.

Seo, J.-W., Ohnishi, Y., Hirata, A. and Horinouchi, S. (2002) 'ATP-binding cassette transport system involved in regulation of morphological differentiation in response to glucose in *Streptomyces griseus*', *Journal of bacteriology*, 184(1), pp. 91–103.

Shi, Y., Zhang, X. and Lou, K. (2013) 'Isolation, characterization, and insecticidal activity of an endophyte of drunken horse grass, *Achnatherum inebrians*', *Journal of Insect Science*, 13(1), p. 151.

Shirling, E. B. T. and Gottlieb, D. (1966) 'Methods for characterization of *Streptomyces* species.', *International journal of systematic bacteriology*, 16(3), pp. 313–340.

Siebold, C., Flükiger, K., Beutler, R. and Erni, B. (2001) 'Carbohydrate transporters of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS)', *FEBS letters*, 504(3), pp. 104–111.

Smith, C. P. and Chater, K. F. (1988a) 'Cloning and transcription analysis of the entire glycerol utilization (*gylABX*) operon of *Streptomyces coelicolor* A3(2) and identification of a closely associated transcription unit', *Molecular and General Genetics MGG*, 211(1), pp. 129–137.

Smith, C. P. and Chater, K. F. (1988b) 'Structure and regulation of controlling sequences for the *Streptomyces coelicolor* glycerol operon', *Journal of molecular biology*, 204(3), pp. 569–580.

Smith, M. C. M. (2015) 'Phage-encoded Serine Integrases and Other Large Serine Recombinases', *Mobile DNA III*, pp. 253–272.

Soh, B. S., Loke, P. and Sim, T.-S. (2001) 'Cloning, heterologous expression and purification of an isocitrate lyase from *Streptomyces clavuligerus* NRRL 3585', *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1522(2), pp. 112–117.

Sola-Landa, A., Moura, R. S. and Martin, J. F. (2003) 'The two-component PhoR-PhoP system controls both primary metabolism and secondary metabolite biosynthesis in *Streptomyces lividans*', *Proceedings of the National Academy of Sciences*, 100(10), pp. 6133–6138.

Sola-Landa, A., Rodríguez-García, A., Apel, A. K. and Martín, J. F. (2008) 'Target genes and structure of the direct repeats in the DNA-binding sequences of the response regulator PhoP

in *Streptomyces coelicolor*', *Nucleic acids research*, 36(4), pp. 1358–1368.

Song, J. Y., Jensen, S. E. and Lee, K. J. (2010) 'Clavulanic acid biosynthesis and genetic manipulation for its overproduction', *Applied microbiology and biotechnology*, 88(3), pp. 659–669.

Song, J. Y., Jeong, H., Yu, D. S., Fischbach, M. A., Park, H.-S., Kim, J. J., Seo, J.-S., Jensen, S. E., Oh, T. K. and Lee, K. J. (2010) 'Draft genome sequence of *Streptomyces clavuligerus* NRRL 3585, a producer of diverse secondary metabolites', *Journal of bacteriology*, 192(23), pp. 6317–6318.

Stackebrandt, E., Rainey, F. A. and Ward-Rainey, N. L. (1997) 'Proposal for a New Hierarchic Classification System, Actinobacteria classis nov.', *International Journal of Systematic Bacteriology*, 47(2), pp. 479–491.

Stanier, R. Y. (1942) 'Agar-decomposing strains of the *Actinomyces coelicolor* species-group', *Journal of bacteriology*, 44(5), p. 555.

Stillwell, W. (2016) 'Chapter 19 - Membrane Transport', in Stillwell, W. B. T.-A. I. to B. M. (Second E. (ed.), pp. 423–451.

Stülke, J. and Hillen, W. (1999) 'Carbon catabolite repression in bacteria', *Current opinion in microbiology*, 2(2), pp. 195–201.

Sun, C.-F., Xu, W.-F., Zhao, Q.-W., Luo, S., Chen, X.-A., Li, Y.-Q. and Mao, X.-M. (2020) 'Crotonylation of key metabolic enzymes regulates carbon catabolite repression in *Streptomyces roseosporus*', *Communications Biology*, 3(1), pp. 1–14.

Süsstrunk, U., Pidoux, J., Taubert, S., Ullmann, A. and Thompson, C. J. (1998) 'Pleiotropic effects of cAMP on germination, antibiotic biosynthesis and morphological development in *Streptomyces coelicolor*', *Molecular microbiology*, 30(1), pp. 33–46.

Sweet, G., Gandor, C., Voegele, R., Wittekindt, N., Beuerle, J., Truniger, V., Lin, E. C. and Boos, W. (1990) 'Glycerol facilitator of *Escherichia coli*: cloning of *glpF* and identification of the *glpF* product.', *Journal of Bacteriology*, 172(1), pp. 424–430.

Świątek-Połatyńska, M. A., Bucca, G., Laing, E., Gubbens, J., Titgemeyer, F., Smith, C. P., Rigali, S. and van Wezel, G. P. (2015) 'Genome-wide analysis of *in vivo* binding of the master

regulator DasR in *Streptomyces coelicolor* identifies novel non-canonical targets', *PLoS One*, 10(4), p. e0122479.

Świątek, M. A., Gubbens, J., Bucca, G., Song, E., Yang, Y.-H., Laing, E., Kim, B.-G., Smith, C. P. and van Wezel, G. P. (2013) 'The ROK family regulator Rok7B7 pleiotropically affects xylose utilization, carbon catabolite repression, and antibiotic production in *Streptomyces coelicolor*', *Journal of bacteriology*, 195(6), pp. 1236–1248.

Świątek, M. A., Tenconi, E., Rigali, S. and van Wezel, G. P. (2012) 'Functional analysis of the N-acetylglucosamine metabolic genes of Streptomyces coelicolor and role in control of development anŚwiątek, M. A., Tenconi, E., Rigali, S. and van Wezel, G. P. (2012) 'Functional analysis of the N-acetylglucosamine metabolic ge', *Journal of bacteriology*, 194(5), pp. 1136–1144.

Takano, E., Tao, M., Long, F., Bibb, Maureen J, Wang, L., Li, W., Buttner, M. J., Bibb, Mervyn J, Deng, Z. X. and Chater, K. F. (2003) 'A rare leucine codon in *adpA* is implicated in the morphological defect of *bldA* mutants of *Streptomyces coelicolor*', *Molecular microbiology*, 50(2), pp. 475–486.

Tenconi, E., Urem, M., Świątek-Połatyńska, M. A., Titgemeyer, F., Muller, Y. A., van Wezel, G. P. and Rigali, S. (2015) 'Multiple allosteric effectors control the affinity of DasR for its target sites', *Biochemical and biophysical research communications*, 464(1), pp. 324–329.

Thompson, J., Cundliffe, E. and Stark, M. (1979) 'Binding of Thiostrepton to a Complex of 23-S rRNA with Ribosomal Protein L11', *European journal of biochemistry*, 98(1), pp. 261–265.

Thompson, J., Schmidt, F. and Cundliffe, E. (1982) 'Site of action of a ribosomal RNA methylase conferring resistance to thiostrepton.', *Journal of Biological Chemistry*, 257(14), pp. 7915–7917.

Thorpe, H. M. and Smith, M. C. M. (1998) 'In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family', *Proceedings of the National Academy of Sciences*, 95(10), pp. 5505–5510.

Tiffert, Y., Supra, P., Wurm, R., Wohlleben, W., Wagner, R. and Reuther, J. (2008) 'The *Streptomyces coelicolor* GlnR regulon: identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in actinomycetes', *Molecular microbiology*, 67(4),

pp. 861–880.

Titgemeyer, F., Reizer, J., Reizer, A. and Saier Jr, M. H. (1994) 'Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria', *Microbiology*, 140(9), pp. 2349–2354.

Titgemeyer, F., Walkenhorst, J., Reizer, J., Stuiver, M. H., Cui, X. and Saier Jr, M. H. (1995) 'Identification and characterization of phosphoenolpyruvate: fructose phosphotransferase systems in three *Streptomyces* species', *Microbiology*, 141(1), pp. 51–58.

Tomono, A., Tsai, Y., Ohnishi, Y. and Horinouchi, S. (2005) 'Three Chymotrypsin Genes Are Members of the AdpA Regulon in the A-Factor Regulatory Cascade in *Streptomyces griseus*', *Journal of Bacteriology*, 187(18), pp. 6341 LP – 6353.

Townsend, C. A. and Ho, M. F. (1985) 'Biosynthesis of clavulanic acid: origin of the C5 unit', *Journal of the American Chemical Society*, 107(4), pp. 1065–1066.

Treangen, T. J. and Rocha, E. P. C. (2011) 'Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes', *PLoS Genet*, 7(1), p. e1001284.

Trepanier, N. K., Jensen, S. E., Alexander, D. C. and Leskiw, B. K. (2002) 'The positive activator of cephamycin C and clavulanic acid production in *Streptomyces clavuligerus* is mistranslated in a *bldA* mutant', *Microbiology*, 148(3), pp. 643–656.

Tschowri, N., Schumacher, M. A., Schlimpert, S., babu Chinnam, N., Findlay, K. C., Brennan, R. G. and Buttner, M. J. (2014) 'Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development', *Cell*, 158(5), pp. 1136–1147.

Ueda, K., Oinuma, K.-I., Ikeda, G., Hosono, K., Ohnishi, Y., Horinouchi, S. and Beppu, T. (2002) 'AmfS, an extracellular peptidic morphogen in *Streptomyces griseus*', *Journal of Bacteriology*, 184(5), pp. 1488–1492.

Ueda, K., Takano, H., Nishimoto, M., Inaba, H. and Beppu, T. (2005) 'Dual transcriptional control of *amfTSBA*, which regulates the onset of cellular differentiation in *Streptomyces griseus*', *Journal of bacteriology*, 187(1), pp. 135–142.

Umeyama, T., Lee, P.-C. and Horinouchi, S. (2002) 'Protein serine/threonine kinases in signal transduction for secondary metabolism and morphogenesis in *Streptomyces*', *Applied*

microbiology and biotechnology, 59(4–5), pp. 419–425.

Ünsaldı, E., Kurt-Kızıldoğan, A., Voigt, B., Becher, D. and Özcengiz, G. (2017) 'Proteome-wide alterations in an industrial clavulanic acid producing strain of *Streptomyces clavuligerus*', *Synthetic and Systems Biotechnology*, 2(1), pp. 39–48.

Valentine, B. P., Bailey, C. R., Doherty, A., Morris, J., Elson, S. W., Baggaley, K. H. and Nicholson, N. H. (1993) 'Evidence that arginine is a later metabolic intermediate than ornithine in the biosynthesis of clavulanic acid by *Streptomyces clavuligerus*', *Journal of the Chemical Society, Chemical Communications*, (15), pp. 1210–1211.

Vasant Kumar, C. and Martín, J. F. (1994) 'Thiostrepton induced proteins in *Streptomyces*, *Amycolatopsis* and *Nocardia* species', *FEMS microbiology letters*, 118(1–2), pp. 107–111.

van Veluw, G. J., Petrus, M. L. C., Gubbens, J., de Graaf, R., de Jong, I. P., van Wezel, G. P., Wösten, H. A. B. and Claessen, D. (2012) 'Analysis of two distinct mycelial populations in liquid-grown Streptomyces cultures using a flow cytometry-based proteomics approach', *Applied Microbiology and Biotechnology*, 96(5), pp. 1301–1312.

Ventola, C. L. (2015) 'The antibiotic resistance crisis: part 1: causes and threats', *Pharmacy and therapeutics*, 40(4), p. 277.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F. and van Sinderen, D. (2007) 'Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum', *Microbiology and molecular biology reviews*, 71(3), pp. 495–548.

Vial, L. and Hommais, F. (2020) 'Plasmid-chromosome cross-talks', *Environmental microbiology*, 22(2), pp. 540–556.

Vobis, G. and Kämpfer, P. (2015) '*Pilimelia*', *Bergey's Manual of Systematics of Archaea and Bacteria*, pp. 1–10.

Vohradsky, J., Li, X.-M., Dale, G., Folcher, M., Nguyen, L., Viollier, P. H. and Thompson, C. J. (2000) 'Developmental Control of Stress Stimulons in *Streptomyces coelicolor* Revealed by Statistical Analyses of Global Gene Expression Patterns', *Journal of bacteriology*, 182(17), pp. 4979–4986.

Vohradsky, J., Li, X. and Thompson, C. J. (1997) 'Identification of procaryotic developmental

stages by statistical analyses of two-dimensional gel patterns', *Electrophoresis*, 18(8), pp. 1418–1428.

Vos, M., Quince, C., Pijl, A. S., de Hollander, M. and Kowalchuk, G. A. (2012) 'A comparison of *rpoB* and 16S rRNA as markers in pyrosequencing studies of bacterial diversity', *PloS one*, 7(2).

Wagner, A. (2008) 'Gene duplications, robustness and evolutionary innovations', *Bioessays*, 30(4), pp. 367–373.

Waksman, S. A. (1961) 'The role of antibiotics in nature', *Perspectives in biology and medicine*, 4(3), pp. 271–287.

Waksman, S. A. and Henrici, A. T. (1943) 'The Nomenclature and Classification of the Actinomycetes.', *Journal of Bacteriology*, 46(4), pp. 337–41.

Waksman, S. A., Robinson, H., Metzger, H. J. and Woodruff, H. B. (1941) 'Toxicity of actinomycin', *Proceedings of the Society for Experimental Biology and Medicine*, 47(2), pp. 261–263.

Waldvogel, E., Herbig, A., Battke, F., Amin, R., Nentwich, M., Nieselt, K., Ellingsen, T. E., Wentzel, A., Hodgson, D. A. and Wohlleben, W. (2011) 'The PII protein GlnK is a pleiotropic regulator for morphological differentiation and secondary metabolism in Streptomyces coelicolor', *Applied microbiology and biotechnology*, 92(6), pp. 1219–1236.

Walshaw, J., Gillespie, M. D. and Kelemen, G. H. (2010) 'A novel coiled-coil repeat variant in a class of bacterial cytoskeletal proteins', *Journal of structural biology*, 170(2), pp. 202–215.

Wang, F., Xiao, X., Saito, A. and Schrempf, H. (2002) '*Streptomyces olivaceoviridis* possesses a phosphotransferase system that mediates specific, phosphoenolpyruvate-dependent uptake of N-acetylglucosamine', *Molecular Genetics and Genomics*, 268(3), pp. 344–351.

Wang, S.-L., Fan, K.-Q., Yang, X., Lin, Z.-X., Xu, X.-P. and Yang, K.-Q. (2008) 'CabC, an EF-hand calcium-binding protein, is involved in Ca2+-mediated regulation of spore germination and aerial hypha formation in *Streptomyces coelicolor*', *Journal of bacteriology*, 190(11), pp. 4061–4068.

Wang, Y., Cen, X.-F., Zhao, G.-P. and Wang, J. (2012) 'Characterization of a new GlnR binding

box in the promoter of *amtB* in *Streptomyces coelicolor* inferred a PhoP/GlnR competitive binding mechanism for transcriptional regulation of *amtB*', *Journal of bacteriology*, 194(19), pp. 5237–5244.

Ward, J. M. and Hodgson, J. E. (1993) 'The biosynthetic genes for clavulanic acid and cephamycin production occur as a 'super-cluster'in three *Streptomyces*', *FEMS Microbiology letters*, 110(2), pp. 239–242.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009) 'Jalview Version 2—a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189–1191.

Wellington, E. M. H., Stackebrandt, E., Sanders, D., Wolstrup, J. and Jorgensen, N. O. G. (1992) 'Taxonomic status of *Kitasatosporia*, and proposed unification with *Streptomyces* on the basis of phenotypic and 16S rRNA analysis and emendation of *Streptomyces* Waksman and Henrici 1943, 339AL', *International Journal of Systematic and Evolutionary Microbiology*, 42(1), pp. 156–160.

van Wezel, G. P., König, M., Mahr, K., Nothaft, H., Thomae, A. W., Bibb, M. and Titgemeyer, F. (2007) 'A new piece of an old jigsaw: glucose kinase is activated posttranslationally in a glucose transport-dependent manner in *Streptomyces coelicolor* A3(2)', *Journal of molecular microbiology and biotechnology*, 12(1–2), pp. 67–74.

van Wezel, G. P., Mahr, K., König, M., Traag, B. A., Pimentel-Schmitt, E. F., Willimek, A. and Titgemeyer, F. (2005) 'GlcP constitutes the major glucose uptake system of *Streptomyces coelicolor* A3(2)', *Molecular microbiology*, 55(2), pp. 624–636.

van Wezel, G. P., van der Meulen, J., Kawamoto, S., Luiten, R. G. M., Koerten, H. K. and Kraal, B. (2000) 'ssgA Is Essential for Sporulation of *Streptomyces coelicolor* A3(2) and Affects Hyphal Development by Stimulating Septum Formation', *Journal of bacteriology*, 182(20), pp. 5653–5662.

van Wezel, G. P., White, J, Bibb, M. J. and Postma, P. W. (1997) 'The *malEFG* gene cluster of *Streptomyces coelicolor* A3(2): characterization, disruption and transcriptional analysis', *Molecular and General Genetics MGG*, 254(5), pp. 604–608.

van Wezel, G. P., White, Janet, Young, P., Postma, P. W. and Bibb, M. J. (1997) 'Substrate

induction and glucose repression of maltose utilization by *Streptomyces coelicolor* A3(2) is controlled by malR, a member of the lacI–galR family of regulatory genes', *Molecular microbiology*, 23(3), pp. 537–549.

Wi, K., Gyun, K. S., Seop, K., Kyu, L. B., Taik, R. Y. and Joon, L. (2006) 'Proteases and protease inhibitors produced in streptomycetes and their roles in morphological differentiation', *Journal of microbiology and biotechnology*, 16(1), pp. 5–14.

Wietzorrek, A. and Bibb, M. (1997) 'A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold', *Molecular microbiology*, 25(6), pp. 1181–1184.

Wildermuth, H. and Hopwood, D. A. (1970) 'Septation during sporulation in *Streptomyces coelicolor*', *Microbiology*, 60(1), pp. 51–59.

Willey, J. M., Willems, A., Kodani, S. and Nodwell, J. R. (2006) 'Morphogenetic surfactants and their role in the formation of aerial hyphae in *Streptomyces coelicolor*', *Molecular microbiology*, 59(3), pp. 731–742.

Willey, J., Santamaria, R., Guijarro, J., Geistlich, M. and Losick, R. (1991) 'Extracellular complementation of a developmental mutation implicates a small sporulation protein in aerial mycelium formation by *S. coelicolor*', *Cell*, 65(4), pp. 641–650.

Willey, J., Schwedock, J. and Losick, R. (1993) 'Multiple extracellular signals govern the production of a morphogenetic protein involved in aerial mycelium formation by *Streptomyces coelicolor*.', *Genes & development*, 7(5), pp. 895–903.

Wolański, M., Donczew, R., Kois-Ostrowska, A., Masiewicz, P., Jakimowicz, D. and Zakrzewska-Czerwińska, J. (2011) 'The level of AdpA directly affects expression of developmental genes in *Streptomyces coelicolor*', *Journal of bacteriology*, 193(22), pp. 6358–6365.

Wolański, M., Jakimowicz, D. and Zakrzewska-Czerwińska, J. (2014) 'Fifty years after the replicon hypothesis: cell-specific master regulators as new players in chromosome replication control', *Journal of bacteriology*, 196(16), pp. 2901–2911.

Wolánski, M., Wali, R., Tilley, E., Jakimowicz, D., Zakrzewska-Czerwińska, J. and Herron, P.

(2011) 'Replisome trafficking in growing vegetative hyphae of *Streptomyces coelicolor* A3(2)', *Journal of bacteriology*, 193(5), pp. 1273–1275.

Woodman, M. E. (2008) 'Direct PCR of Intact Bacteria (Colony PCR)', *Current Protocols in Microbiology*, 9(1), p. A.3D.1-A.3D.6.

Wu, X. and Roy, KENNETH 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.

Wu, X. and Roy, Kenneth 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.

Xiao, X., Wang, F., Saito, A., Majka, J., Schlösser, A. and Schrempf, H. (2002) 'The novel *Streptomyces olivaceoviridis* ABC transporter Ngc mediates uptake of N-acetylglucosamine and N, N'-diacetylchitobiose', *Molecular Genetics and Genomics*, 267(4), pp. 429–439.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015) 'The I-TASSER Suite: protein structure and function prediction', *Nature methods*, 12(1), p. 7.

Yang, J. and Zhang, Y. (2015) 'I-TASSER server: new development for protein structure and function predictions', *Nucleic acids research*, 43(W1), pp. W174–W181.

Yang, Y.-H., Song, E., Kim, E.-J., Lee, K., Kim, W.-S., Park, S.-S., Hahn, J.-S. and Kim, B.-G. (2009) 'NdgR, an IclR-like regulator involved in amino-acid-dependent growth, quorum sensing, and antibiotic production in *Streptomyces coelicolor*', *Applied microbiology and biotechnology*, 82(3), pp. 501–511.

Yim, G., Huimi Wang, H. and Davies Frs, J. (2007) 'Antibiotics as signalling molecules', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1483), pp. 1195–1200.

Yim, G., Wang, H. H. and Davies, J. (2006) 'The truth about antibiotics', *International Journal of Medical Microbiology*, 296(2–3), pp. 163–170.

Yin, H., Xiang, S., Zheng, J., Fan, K., Yu, T., Yang, X., Peng, Y., Wang, H., Feng, D. and Luo, Y. (2012) 'Induction of holomycin production and complex metabolic changes by the *argR*

mutation in *Streptomyces clavuligerus* NP1', *Applied and environmental microbiology*, 78(9), pp. 3431–3441.

Zacchetti, B., Smits, P. and Claessen, D. (2018) 'Dynamics of pellet fragmentation and aggregation in liquid-grown cultures of *Streptomyces lividans*', *Frontiers in Microbiology*, 9, p. 943.

Zhang, J. (2003) 'Evolution by gene duplication: an update', *Trends in ecology & evolution*, 18(6), pp. 292–298.

Zhang, Y. (2008) 'I-TASSER server for protein 3D structure prediction', *BMC bioinformatics*, 9(1), p. 40.

Zhang, Z., Wang, Y. and Ruan, J. (1997) 'A proposal to revive the genus *Kitasatospora* (Omura, Takahashi, Iwai, and Tanaka 1982)', *International Journal of Systematic and Evolutionary Microbiology*, 47(4), pp. 1048–1054.

Zheng, Y., Sun, C.-F., Fu, Y., Chen, X.-A., Li, Y.-Q. and Mao, X.-M. (2019) 'Dual regulation between the two-component system PhoRP and AdpA regulates antibiotic production in *Streptomyces*', *Journal of industrial microbiology & biotechnology*, 46(5), pp. 725–737.

Zhou, Z., Gu, J., Li, Y.-Q. and Wang, Y. (2012) 'Genome plasticity and systems evolution in *Streptomyces*', in *BMC bioinformatics*, p. S8.

Ziegelin, G., Fürste, J. P. and Lanka, E. (1989) 'TraJ protein of plasmid RP4 binds to a 19-base pair invert sequence repetition within the transfer origin.', *Journal of Biological Chemistry*, 264(20), pp. 11989–11994.

10   Appendix

Copy of

Schniete, J. K., Selem-Mojica, N., Birke, A. S., Cruz-Morales, P., Hunter, I. S., Barona-Gomez, F. and Hoskisson, P. A. (2021) 'ActDES–a curated actinobacterial database for evolutionary studies', *Microbial genomics*, p. 498.

A

ActDES – a curated Actinobacterial Database for Evolutionary Studies

Jana K. Schniete[1,2], Nelly Selem-Mojica[3], Anna S. Birke[2], Pablo Cruz-Morales[3], Iain S. Hunter[2], Francisco Barona-Gomez[3] and Paul A. Hoskisson[2]*

Abstract

Actinobacteria is a large and diverse phylum of bacteria that contains medically and ecologically relevant organisms. Many members are valuable sources of bioactive natural products and chemical precursors that are exploited in the clinic and made using the enzyme pathways encoded in their complex genomes. Whilst the number of sequenced genomes has increased rapidly in the last 20 years, the large size, complexity and high G+C content of many actinobacterial genomes means that the sequences remain incomplete and consist of large numbers of contigs with poor annotation, which hinders large-scale comparative genomic and evolutionary studies. To enable greater understanding and exploitation of actinobacterial genomes, specialized genomic databases must be linked to high-quality genome sequences. Here, we provide a curated database of 612 high-quality actinobacterial genomes from 80 genera, chosen to represent a broad phylogenetic group with equivalent genome re-annotation. Utilizing this database will provide researchers with a framework for evolutionary and metabolic studies, to enable a foundation for genome and metabolic engineering, to facilitate discovery of novel bioactive therapeutics and studies on gene family evolution. This article contains data hosted by Microreact.

DATA SUMMARY

1. All genome sequences used in this study can be found in the National Center for Biotechnology Information (NCBI) Taxonomy Browser (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi) and are summarized along with accession numbers in Table S1 (available on Figshare – https://doi.org/10.6084/m9.figshare.13143407.v1). Other data are available on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

2. Perl script files are available on GitHub (https://github.com/nselem/ActDES), including details of how to batch annotate genomes in RAST from the terminal (https://github.com/nselem/myrast).

3. Table S1 shows a list of genomes from the NCBI (actinobacteria database.xlsx) and is available on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

4. CVS genome annotation files including the FASTA files of nucleotide and amino acids sequences (individual .cvs files) are available on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

5. BLAST nucleotide database (.fasta file) information is available on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

6. BLAST protein database (.fasta file) information is available on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

7. Table S2 expansion table – genus level (expansion table.xlsx – tab genus level) is available on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

8. Table S2 expansion table – species level (expansion table.xlsx – tab species level) is available on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

9. All GlcP and Glk data – BLAST hits from ActDES, MUSCLE alignment files and .nwk tree files – can be found on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

10. Interactive trees in Microreact for Glk (https://microreact.org/project/w_KDfn1xA/5a178533) and associated files can be found on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

11. Interactive trees in Microreact for GlcP (https://microreact.org/project/VBUdiQ5_k/045c95e1) and associated files can be found on Figshare (https://doi.org/10.6084/m9.figshare.13143407.v1).

12. Jupyter Notebook for exploring ActDES in MyBinder can be found at https://github.com/nselem/ActDES.

INTRODUCTION

The increase in availability of bacterial whole-genome sequencing provides large amounts of data for evolutionary and phylogenetic analysis. However, there is great variation in the quality, annotation and phylogenetic skew of the data available in large universal databases, meaning that evolutionary and phylogenetic studies can be challenging. To address this variation, curated, high-level, taxa-specific, non-redundant sub-databases need to be assembled to aid detailed analysis. Given that there is a direct correlation between phylogenetic distance and the discovery of novel function [1–3], it is imperative that any derived databases must be phylogenetically representative and non-redundant to enable insight into the evolution of genes, proteins and pathways within a given group of taxa [1].

The phylum *Actinobacteria* is a major taxon amongst the *Bacteria*, which includes phenotypically and morphologically diverse organisms found on every continent and in virtually every ecological niche [4]. They are particularly common in soils, yet within their ranks are potential human and animal pathogens such as *Corynebacterium*, *Mycobacterium*, *Nocardia* and *Tropheryma*, inhabitants of the gastrointestinal tract (*Bifidobacterium* and *Scardovia*), as well as plant commensals and pathogens such as *Frankia*, *Leifsonia* and *Clavibacter* [4, 5]. Perhaps the most notable trait of the phylum is the renowned ability to produce bioactive natural products such as antibiotics, anti-cancer agents and immuno-suppressive agents, with genera such as *Amycolatopsis*, *Micromonospora* and *Streptomyces* being particularly prominent [6]. As a result, computational 'mining' of actinobacterial genomes has become an important part of the drug-discovery pipeline, with increasing numbers of online resources and software devoted to identification of natural-product biosynthetic gene clusters (BGCs) [7–9]. It is important to move beyond approaches that rely on similarity searches of known BGCs and to expand searches to identify hidden chemical diversity within the genomes [6, 7, 10–13].

A recent study of 830 actinobacterial genomes found >11000 BGCs comprising 4122 chemical families, indicating that there is a vast diversity of strains and chemistry to exploit

Significance as a Bioresource to the Community

The *Actinobacteria* is a large diverse phylum of bacteria, often with large, complex genomes with a high G+C content. Sequence databases have great variation in the quality of sequences, equivalence of annotation and phylogenetic representation, which makes it challenging to undertake evolutionary and phylogenetic studies. To address this, we have assembled a curated, taxa-specific, non-redundant database to aid detailed comparative analysis of *Actinobacteria*. ActDES (Actinobacterial Database for Evolutionary Studies) constitutes a novel resource for the community of actinobacterial researchers that will be useful primarily for two types of analyses: (i) comparative genomic studies, facilitated by reliable identification of orthologs across a set of defined phylogenetically representative genomes, and (ii) phylogenomic studies, which will be improved by identification of gene subsets at specified taxonomic level. These analyses can then act as a springboard for the studies of the evolution of virulence genes, the evolution of metabolism and identification of targets for metabolic engineering.

[14], yet within each of these strains there will be hidden diversity in the form of cryptic BGCs. To exploit this undiscovered diversity as the technology develops and databases expand, new biosynthetic logic will emerge, yet we know little of how natural selection shapes the evolution of BGCs and how biosynthetic precursors are supplied to gene products of BGCs from primary metabolism and to identify targets for metabolic engineering of industrially relevant strains. Such logic will expedite industrial strain improvement processes, enabling titre increases and development of novel molecules, as well as the engineering of strains to use more sustainable feedstocks.

To aid this process, we have created an actinobacterial metabolism database including functional annotations for enzymes from 612 species to enable phylum-wide interrogation of gene expansion events that may indicate adaptive evolution, help shape metabolic robustness for antibiotic production [15] or enable the identification of targets for metabolic engineering. Actinobacterial Database for Evolutionary Studies (ActDES) provides a curated list of high-quality, phylum-specific genomes and data to help users navigate the redundancy and inconsistency in sequence databases in a simplified format that enables researchers with little taxonomic knowledge to develop testable evolutionary hypotheses. To demonstrate the utility of ActDES, we have detailed its construction and used it to investigate the glucose permease/glucokinase system phylogeny across the *Actinobacteria*.

METHODS

We generated ActDES, a database for evolutionary analysis of actinobacterial genomes, in two formats: a database for
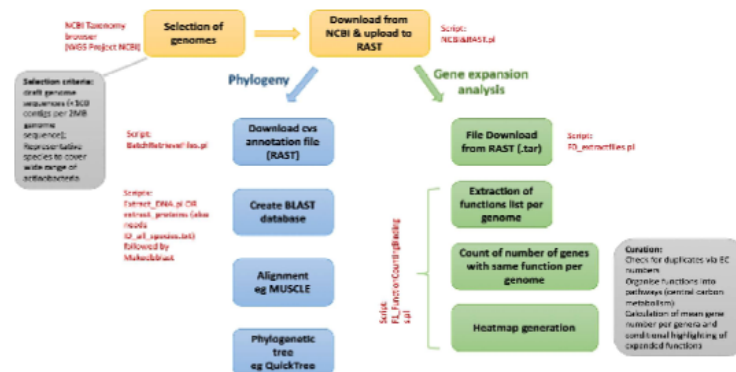
Fig. 1. Schematic workflow for the creation of ActDES. Genomes were selected from NCBI Taxonomy Browser and uploaded for annotation to RAST [38]. The annotated genomes were then processed for two different analyses. Firstly, the functional roles were downloaded and for each functional role the numbers of occurrences per genome were counted in order to obtain an expansion table (Table S2) by comparing the mean of each genus to the overall mean of all genera. Secondly, the genomes were used to extract all nucleotide and protein sequences in FASTA format, which could then be queried by sequence using BLAST [39]. The hits were aligned in MUSCLE [16] and after refinement the alignment was used to reconstruct phylogenetic trees in QuickTree [17].

interrogation by BLASTN or BLASTP for phylogenetic analysis, and a primary metabolic gene expansion table, which can be mined at different taxonomic levels (Tables S1 and S2) for specific metabolic functions from primary metabolism. A schematic overview of the generation of the dataset is shown in Fig. 1.

The database was generated via the National Center for Biotechnology Information (NCBI) Taxonomy Browser (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax. cgi) to identify actinobacterial genome sequences. The quality of the genome sequences was filtered by the number of contigs (<100 contigs per 2 Mb of genome sequence) and the genomes were downloaded from the NCBI WGS repository (https:// www.ncbi.nlm.nih.gov/Traces/wgs/). These genomes were then dereplicated to ensure that the database comprised a wide taxonomic range of the phylum, resulting in 612 species from 80 genera within 13 suborders of the *Actinobacteria* (Table S1).

Each of these 612 genomes was reannotated using RAST. Default settings were used to ensure equivalence of annotation across the database and the annotation files of each genome were downloaded (cvs files – https://doi.org/10.6084/ m9.figshare.13143407.v1). These annotation files were subsequently used to extract all protein and nucleotide sequences into two files. Each of these files was subsequently converted into BLAST databases (a protein database and a nucleotide database – https://doi.org/10.6084/m9.figshare.12167724) to facilitate phylogenetic analysis. Sequences of interest can be aligned using MUSCLE [16] and phylogenetic trees

reconstructed using a range of tree construction software such as QuickTree [17], IQ tree [18] or MrBayes [19]. Subsequent trees may be visualized in software such as FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

The RAST annotation files were also used to extract the functional roles of each coding sequence (CDS) per genome and the level of gene expansion was assessed for each genome by counting the number of genes per species per functional category (gene function annotation). The dataset was then curated manually for central carbon metabolism and amino acid biosynthesis pathways to create the gene expansion table (Table S2), with the organisms grouped according to their taxonomic position. The quality of the data was checked at each step for duplicates and inconsistencies, and was curated manually to exclude faulty entries. As the NCBI Taxonomy Browser database is overrepresented in *Streptomyces* genomes due to the number of species that have been sequenced relative to other *Actinobacteria*, this is also reflected in ActDES (288 *Streptomyces* genomes from a total of 612 genomes). However, this was addressed in the expansion table (Table S2) by calculating the mean occurrence of each functional category within each genus and then calculating an overall mean for the phylum to compensate. The mean occurrence of each functional category per genus plus the standard deviation was also calculated, and this was used to analyse the occurrence of each functional gene category per species within Table S2. A gene function annotation with a gene copy number value above the mean plus the standard deviation for each genus indicated that there had been a gene expansion event in that

species and this was noted. The gene expansion table (Table S2) enables researchers to identify groups of genes of interest for subsequent phylogenetic and evolutionary analysis, which can be performed with confidence due to the highly curated nature of the data included in the database.

As the NCBI Taxonomy Browser database is overrepresented in *Streptomyces* genomes due to the number of species that have been sequenced relative to other *Actinobacteria*, this is also reflected in ActDES (288 *Streptomyces* genomes from a total of 612 genomes). However, this was addressed in the expansion table (Table S2) by calculating the mean occurrence of each functional category within each genus and then calculating an overall mean for the phylum to compensate. The mean occurrence of each functional category per genus plus the standard deviation was also calculated, and this was used to analyse the occurrence of each functional gene category per species within Table S2. A gene function annotation with a gene copy number value above the mean plus the standard deviation for each genus indicated that there had been a gene expansion event in that species and this was noted. The gene expansion table (Table S2) enables researchers to identify groups of genes of interest for subsequent phylogenetic and evolutionary analysis, which can be performed with confidence due to the highly curated nature of the data included in the database.

RESULTS

The gene expansion table (Table S2) lists 612 species of 80 genera within the *Actinobacteria* with data that provides an extensive analysis at the phylum level, which is the starting point for detailed phylogenomic studies. Gene expansions were identified in separate datasets at the genus and species levels, along with details of the numbers of genes in each functional category per species and the mean numbers of genes in each functional category per genus expanded within the genomes. These data can be used subsequently in phylogenomic analyses to identify targets for metabolic engineering and gene function studies. Identification of expanded gene families may also facilitate the recognition of novel natural product BGCs, for which gene expansion events of primary metabolic genes have been classified to be associated within BGCs as biosynthetic enzymes or through provision of additional copies of antibiotic targets that may subsequently function as resistance mechanisms [6, 11, 20–24].

This database has found utility for studying primary metabolic gene expansions in *Streptomyces*. It enabled a detailed *in silico* analysis of the duplication event leading to the two pyruvate kinases in the genus of *Streptomyces*, subsequently enabling the functional characterization of the two isoenzymes to reveal how they contribute to metabolic robustness [15]. ActDES may also be useful for investigating the distribution of primary metabolic genes across the phylum to link phenotype to genotype and phylogenetic position. An initial RpoB phylogeny has been reconstructed previously using this database [15], which provided a robust universal phylogeny for comparison of individual protein trees [25].

To demonstrate the utility of ActDES, the glucose permease/glucokinase system of the *Actinobacteria* was investigated. The role of nutrient-sensing in regulation of antibiotic biosynthesis is well known [26], with the enzyme glucokinase (Glk) playing a central role in carbon-catabolite repression (CCR) in *Streptomyces* [27]. In most bacteria, CCR is mediated by the phosphoenolpyruvate-dependent phosphotransferase system (PTS), yet in *Streptomyces*, glucose uptake is mediated by the major-facilitator superfamily (MFS) transporter, glucose permease (GlcP), and there is evidence for direct interaction between Glk and GlcP, which may mediate CCR [28]. Understanding the nature and distribution of these enzymes will play a key role in developing industrial fermentations with glucose as major carbon source. Investigating the distribution of the glucose permease/glucokinase system across the phylum shows that GlcP and Glk have been the subject of gene expansion events in some members of the *Streptomycetales*, most notably the *Streptomyces*, with a patchy distribution of the Glk/GlcP system across the remainder of the phylum (Table S2; genus tab). However, where the Glk/GlcP system is found, the number of expansion events observed is greater for Glk than for GlcP (Fig. 2a, b). The phylogenetic trees (Fig. 2a, b) clearly show two clades for Glk and GlcP within the *Streptomycetales* (interactive trees are available via Microreact [29]: Glk – https://microreact.org/project/w_KDfn1xA/5a178533, and GlcP – https://microreact.org/project/VBUdiQ5_k/045c95e1). However, these clades differ in the number of sequences, with the Glk clades being equal in number, suggesting that a duplication event has occurred within the *Streptomycetales* (Fig. 2b). The is consistent throughout the order, with the patterns largely the same as observed for *Streptomyces coelicolor*. This species has two ROK-family ATP-dependent glucokinases, SCO2126 (*glkA*) and SCO6260, that share around 50% amino acid sequence identity, and each is found in one of the distinct clades (permease-associated kinases and orphan kinases) (Fig. 2b). Whilst SCO2126 is a GlcP-associated kinase, the gene encoding SCO6260 is located in an operon including genes encoding a putative carbohydrate ABC-transporter system, which has been reported elsewhere [30]. SCO6260 appears to be the only glucokinase in the database that is associated with an ABC-transporter. This may suggest that expansion of the Glk gene family in *Streptomycetales* might have occurred to extend the number of CCR-mediating kinases in the genome, adding increased regulatory complexity to carbohydrate metabolism in this group of organisms that use CCR as a major regulator of specialized metabolism.

The two clades for GlcP within the *Streptomycetales* differ in size, suggesting either gene duplication followed by gene loss or an expansion through horizontal gene transfer (HGT) has occurred. A detailed examination of these clades by species (Table S2; species tab) shows the presence of both scenarios. There are duplicated enzymes located within the same clade (as observed in *S. coelicolor*; group I) or additional copies of the permease that are located in a phylogenetically distinct clade, which lacks congruence with the RpoB tree [15], and remarkably consists entirely of sequences from the genus
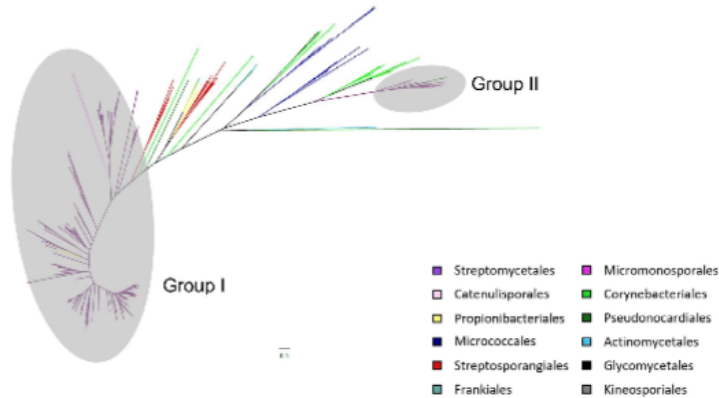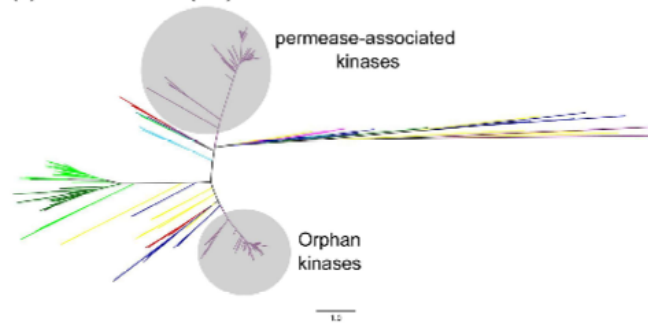
(a). Glucose permease (GlcP)

Group II

Group I

■ Streptomycetales	■ Micromonosporales
☐ Catenulisporales	■ Corynebacteriales
☐ Propionibacteriales	■ Pseudonocardiales
■ Micrococcales	■ Actinomycetales
■ Streptosporangiales	■ Glycomycetales
■ Frankiales	■ Kineosporiales

(b). Glucose kinase (Glk)

permease-associated kinases

Orphan kinases

Fig. 2. (a) Actinobacterial-wide phylogenetic tree of glucose permeases (GlcP) and (b) actinobacterial-wide phylogenetic tree of glucokinases (Glk). Trees are colour-coded according to the NCBI Taxonomy Browser (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi). Interactive trees are also available via Microreact [29]: GlcP –https://microreact.org/project/VBUdiQ5_k/045c95e1, and Glk – https://microreact.org/project/w_KDfn1xA/5a178533. Scale bar indicates branch lengths equivalent to one substitution per site.

Streptomyces (group II; Fig. 2a). This suggests that they may have been acquired via HGT. The expansive nature of the duplicated Glk enzymes compared to GlcP may be due to the role played in CCR by the GlkA enzymes [27], and the different transcriptional activities under glycolytic and gluconeogenic conditions [31], yet quite how these different Glk enzymes interact with the permease(s) under various conditions requires further experimental investigation to understand their exact physiological role, and how this may be translated into industrial strain improvement processes.

DISCUSSION

Large-scale whole-genome sequencing and phylogenomic analysis is increasingly used for identifying targets for genome and metabolic engineering, studies of metabolic capabilities, pathogen phylogenomics and evolutionary studies. These studies are often complicated by the large number of sequences in the databases, database redundancy and the poor quality of some genome sequence data. The development of the high-quality, curated ActDES, reported here, enables phylum-wide

taxonomic representation of the *Actinobacteria* coupled with quality-filtered genome data and equivalent annotation for each CDS.

The intended primary use of ActDES will be in the study of primary metabolism, but it is not limited. It can also inform the development and evolution of metabolism in strains that produce bioactive metabolites, given the high representation of genera renowned for their ability to produce natural products such as *Streptomyces* and *Micromonospora*. Due to a greater understanding of BGC evolution and genome organization in *Actinobacteria*, it is becoming increasingly clear that genes whose functions are in primary metabolism may actually contribute directly to the biosynthesis of specialized metabolites and, hence, the identification of duplicates may indicate the presence of cryptic BGCs [6, 11] or, when associated with precursor biosynthetic genes, provide the raw material for the enzymes across multiple BGCs [32–34].

ActDES may also find utility in evolutionary studies of expanded gene families across the actinobacterial phylum that contribute to virulence, such as the *mce* locus, which is known to facilitate host survival in mycobacteria [35], but also facilitates xenobiotic substrate uptake in *Rhodococcus* [36], and enables root colonization and survival in *Streptomyces* [37]. With phylum-wide taxonomic representation of established actinobacterial animal and plant pathogens, the scope for evolutionary studies using these data is enormous.

Usage notes

The CVS files of each genome contain the RAST annotation details in addition to the DNA and protein sequences for each annotated CDS (https://doi.org/10.6084/m9.figshare.12167880). The genome list contains the RAST ID (which is equivalent to the name of the .cvs file) along with the NCBI ID (sequence ID; Table S1) plus the species name, which are included in the dataset. Further details of annotating batches of genomes in RAST can be found at https://github.com/nselem/myrast.

The primary metabolism expansion tables (Table S2) are organized by metabolic pathway along the top row with the Enzyme Commission (EC) number and functional annotation, with the first column being the taxonomic assignment. The genus table shows the mean number of genes of the annotated function. Highlighted cells reflect gene expansion events, i.e. those genes that are present in a higher number than the overall mean across the database plus the standard deviation.

It is suggested that the gene expansion table (Table S2) is searched in the first instance (either by species or genus of interest or by a specific enzymatic function). This can be carried out by a simple text search. This will then allow the identification of a query sequence from a species or gene of interest (either nucleotide or amino acid sequence), which can then be searched against the curated BLAST database allowing a detailed phylogenetic analysis of a gene/protein of interest by using standard alignment and tree building software tools.

These data can also be used in detailed evolutionary analysis of selection, mutation rates, etc. We have set up a Jupyter Notebook through the MyBinder project (https://mybinder.org/) to enable ease of use of the code (https://github.com/nselem/ActDES) with a tutorial to enable use of the database (https://github.com/nselem/ActDES).

Author contributions
Conceptualization: J. K. S., N. S.-M., P. A. H. and F. B.-G. Data curation: J. K. S. and N. S.-M. Formal analysis: J. K. S., N. S.-M., P. C.-M. and A. S. B. Funding acquisition: P. A. H. and F. B.-G. Methodology: J. K. S., N. S.-M., P. C.-M., F. B.-G. and P. A. H. Project administration: P. A. H. and F. B.-G. Supervision: P. A. H., I. S. H. and F. B.-G. Writing – original draft: J. K. S., A. S. B. and P. A. H. Writing – review and editing: J. K. S., N.S.-M., P. C.-M., A. S. B., I. S. H., F. B.-G. and P. A. H.

References

1. Mukherjee S, Seshadri R, Varghese NJ, Eloe-Fadrosh EA, Meier-Kolthoff JP *et al.* 1003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat Biotechnol* 2017;35:676–683.

2. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E *et al.* A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 2009;462:1056–1060.

3. Kunin V, Cases I, Enright AJ, de Lorenzo V, Ouzounis CA. Myriads of protein families, and still counting. *Genome Biol* 2003;4:401.

4. Goodfellow M. *Bergey's Manual of Systematics of Archaea and Bacteria.* New York: Springer; 2015.

5. Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF *et al.* Genomics of actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* 2007;71:495–548.

6. Chevrette MG, Gutiérrez-García K, Selem-Mojica N, Aguilar-Martínez C, Yañez-Olvera A *et al.* Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat Prod Rep* 2020;37:566–599.

7. Adamek M, Alanjary M, Ziemert N. Applied evolution: phylogeny-based approaches in natural products research. *Nat Prod Rep* 2019;36:1295–1312.

8. Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes – a review. *Nat Prod Rep* 2016;33:988–1005.

9. Medema MH, Fischbach MA. Computational approaches to natural product discovery. *Nat Chem Biol* 2015;11:639–648.

10. Adamek M, Alanjary M, Sales-Ortells H, Goodfellow M, Bull AT *et al.* Comparative genomics reveals phylogenetic distribution patterns

of secondary metabolites in Amycolatopsis species. BMC Genomics 2018;19:426.

11. Cruz-Morales P, Kopp JF, Martínez-Guerrero C, Yáñez-Guerra LA, Selem-Mojica N et al. Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model streptomycetes. Genome Biol Evol 2016;8:1906–1916.

12. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH et al. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 2020;16:60–.

13. Sélem-Mojica N, Aguilar C, Gutiérrez-García K, Martínez-Guerrero CE, Barona-Gómez F. EvoMining reveals the origin and fate of natural product biosynthetic enzymes. Microb Genom 2019;5:e000260.

14. Doroghazi JR, Metcalf WW. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. BMC Genomics 2013;14:611.

15. Schniete JK, Cruz-Morales P, Selem-Mojica N, Fernández-Martínez LT, Hunter IS et al. Expanding primary metabolism helps generate the metabolic robustness to facilitate antibiotic biosynthesis in Streptomyces. mBio 2018;9:e02283-17.

16. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 2004;5:113.

17. Howe K, Bateman A, Durbin R. QuickTree: building huge neighbour-joining trees of protein sequences. Bioinformatics 2002;18:1546–1547.

18. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 2020;37:1530–1534.

19. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 2012;61:539–542.

20. Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC et al. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. ACS Chem Biol 2015;10:2841–2849.

21. Schmidt KL, Peterson ND, Kustusch RJ, Wissel MC, Graham B et al. A predicted ABC transporter, FtsEX, is needed for cell division in Escherichia coli. J Bacteriol 2004;186:785–793.

22. Steffensky M, Mühlenweg A, Wang Z-X, Li S-M, Heide L. Identification of the novobiocin biosynthetic gene cluster of Streptomyces spheroides NCIB 11891. Antimicrob Agents Chemother 2000;44:1214–1222.

23. Kling A, Lukat P, Almeida DV, Bauer A, Fontaine E et al. Targeting DnaN for tuberculosis therapy using novel griselimycins. Science 2015;348:1106–1112.

24. Peterson RM, Huang T, Rudolf JD, Smanski MJ, Shen B. Mechanisms of self-resistance in the platensimycin- and platencin-producing

Streptomyces platensis MA7327 and MA7339 strains. Chem Biol 2014;21:389–397.

25. Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF et al. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microb 2007;73:278–288.

26. Fernández-Martínez LT, Hoskisson PA. Expanding, integrating, sensing and responding: the role of primary metabolism in specialised metabolite production. Curr Opin Microbiol 2019;51:16–21.

27. Gubbens J, Janus M, Florea BI, Overkleeft HS, van Wezel GP. Identification of glucose kinase-dependent and -independent pathways for carbon control of primary metabolism, development and antibiotic production in Streptomyces coelicolor by quantitative proteomics. Mol Microbiol 2012;86:1490–1507.

28. van Wezel GP, König M, Mahr K, Nothaft H, Thomae AW et al. A new piece of an old jigsaw: glucose kinase is activated posttranslationally in a glucose transport-dependent manner in Streptomyces coelicolor A3(2). J Mol Microb Biotech 2007;12:67–74.

29. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom 2016;2:e000093.

30. Bertram R, Schlicht M, Mahr K, Nothaft H, Saier MH et al. In silico and transcriptional analysis of carbohydrate uptake systems of Streptomyces coelicolor A3(2). J Bacteriol 2004;186:1362–1373.

31. Schniete JK, Reumerman R, Kerr L, Tucker NP, Hunter IS et al. Differential transcription of expanded gene families in central carbon metabolism of Streptomyces coelicolor A3(2). Access Microbiol 2020;2:e000122.

32. Chan YA, Podevels AM, Kevany BM, Thomas MG. Biosynthesis of polyketide synthase extender units. Nat Prod Rep 2009;26:90–114.

33. Pfeifer BA, Khosla C. Biosynthesis of polyketides in heterologous hosts. Microbiol Mol Biol Rev. 2001;65:106–118.

34. Zhang G, Li Y, Fang L, Pfeifer BA. Tailoring pathway modularity in the biosynthesis of erythromycin analogs heterologously engineered in E. coli. Sci Adv 2015;1:e1500077.

35. Arruda S, Bomfim G, Knights R, Huima-Byron T, Riley L. Cloning of an M. tuberculosis DNA fragment associated with entry and survival inside cells. Science 1993;261:1454–1457.

36. Mohn WW, van der Geize R, Stewart GR, Okamoto S, Liu J et al. The actinobacterial mce4 locus encodes a steroid transporter. J Biol Chem 2008;283:35368–35374.

37. Clark LC, Seipke RF, Prieto P, Willemse J, van Wezel GP et al. Mammalian cell entry genes in Streptomyces may provide clues to the evolution of bacterial virulence. Sci Rep 2013;3:1109.

38. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T et al. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 2008;9:75.

39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.