

INDEX

Abstract.....	5
Introduction	6
The urgent need.....	6
Actinobacteria & <i>Streptomyces</i>	8
<i>Streptomyces</i> secondary metabolism	12
<i>Streptomyces</i> produce a range of structurally diverse bioactive metabolites	14
Non Ribosomal Peptide Synthetases (NRPS)	16
Polyketide Synthases (PKS)	24
The Pyrrole Amides congocidine, distamycin and other minor groove binding molecules.....	25
Mutasyntesis.....	31
Combinatorial biosynthesis	34
Synthetic biology	36
Resistance mechanism in bacteria.....	38
Resistance to pyrrole amides-efflux pumps	39
Scope of this PhD thesis	42
Specific aims.....	42
Materials and Methods.....	43
Strains, Plasmids and cultivation	43
Media preparation	45
Maintenance of plasmid/cosmid selection in strains	47
Cultivation of <i>Streptomyces</i> strains	49
Liquid culture growth curves-dry weight.....	49
Liquid culture growth curves-OD	49
Preparation of <i>Streptomyces</i> spore stocks	49
Cultivation of <i>E. coli</i> strains.....	49
Storage of <i>E. coli</i> stocks	50
Culture supernatant sample preparation for HPLC	50

Congocidine and distamycin standards.....	50
Liquid-Liquid extraction	51
Biomass extraction.....	51
Reconstitution of dried samples for HPLC	52
Thin layer chromatography.....	52
Antimicrobial susceptibility assay	52
Minor groove binding susceptibility assay.....	53
Preparation of <i>E. coli</i> chemically competent cells	54
Preparation of <i>E. coli</i> electro-competent cells (for transformation of DNA constructs ≥ 11 kB)	55
Transformation of <i>E. coli</i> chemically competent cells	55
Transformation of <i>E. coli</i> electro-competent cells.....	56
Isolation of plasmid and cosmid DNA from <i>E. coli</i> (alkaline lysis method)	56
Isolation of plasmid DNA from <i>E. coli</i> using commercial kits.....	57
Intergeneric conjugation of DNA from <i>E. coli</i> to <i>Streptomyces</i>	57
Isolation of genomic DNA from <i>Streptomyces</i>	57
Agarose gel electrophoresis.....	58
Agarose gel DNA extraction and purification	58
Cultivation of yeast strains	59
Storage of yeast stocks	59
Lithium Acetate/SS DNA/PEG transformation of yeast	59
Electroporation of yeast (for up to 20 μ g of DNA).....	60
Use of a haemocytometer	61
Yeast recombineering	61
Yeast Zymolase plasmid DNA extraction protocol.....	61
Primers design for PCR and cloning techniques	62
Diagnostic PCR	66
Cloning PCR.....	66
Blunt end cloning.....	66
Restriction digest cloning.....	67
Ligation	67
Polymerase cycling assembly cloning	67
Overlapping PCR ligation	68
Gene disruption by λ -Red recombination.....	69
Strain identification by 16S rDNA clonal DNA	70
Chemical synthesis of novel heterocycles for feeding to congocidine producing strains.....	70
4-acetamido-1-methyl-pyrrole-2-carboxylic acid.....	71
4-acetamido-1-(methyl-d3)-pyrrole-2-carboxylic acid.....	72

4-acetamido-1-methyl-imidazole-2-carboxylic acid.....	72
2-acetamidothiazole-4-carboxylic acid.....	73
Chemical synthesis of two minor groove binders used in the chapter four assays (compounds 24 and 26).....	75
Synthesis of MGB-6/ Compound 24.....	75
Synthesis of MGB-7/Compound 26.....	77
Annex:Compound characterisation.....	241
Chapter 1: distamycin-congocidine combinatorial biosynthesis	80
Introduction	80
Validating congocidine biosynthesis.....	81
Congocidine-distamycin combinatorial biosynthesis	87
Conclusions	108
Chapter 2: Synthetic assembly of the congocidine cluster.....	109
Introduction	109
Organisation of the congocidine cluster	109
Yeast recombination based assembly.....	113
Gibson cloning to assemble the congocidine biosynthetic gene cluster.	135
Conclusions	143
Chapter 3: Mutasynthesis of congocidine	144
Introduction.....	139
Competitive inhibition assay	147
Non-natural substrate incorporation	150
Strain construction	152
Production of novel congocidine derivatices by mutasynthesis.....	153
Conclusions	160
Chapter 4: Pyrrole-amide ABC transporter study	161
Introduction.....	161
<i>In silico</i> analysis of the putative transporters	163
Strain construction.....	176
Congocidine resistant strain TK23RC01	176

Distamycin resistant strain TK23RD01	180
Control strain TK23pIJ6902	180
Compound screening of the pump mutants	184
Pump inhibition	187
Non-natural pyrrole amide assay	189
Conclusions	206
Discussion	207
Combinatorial biosynthesis between distamycin and congocidine clusters	207
Synthetic assembly of the congocidine cluster	210
Mutasynthesis of congocidine	211
Pyrrole-amide ABC transporter study	212
Conclusions and Future work	215
References	216

Abstract

Novel antibiotics are desperately needed. We are on the verge of the pre-antibiotic era, with bacteria resistant to last resort drugs spreading quickly worldwide. Natural products, or chemical molecules produced by living organisms, evolved over time to interact with biological targets following rules that researchers have not yet outlined fully. However microorganisms, once the source of most of the drugs we use today, were abandoned in favor of synthetic chemistry, under the false premise that natural sources were depleted.

When pursuing novel antimicrobials, two options are in hand, either drug discovery or drug functionalisation. In this thesis, approaches that were left on the golden era of antibiotic discovery due to technological limitations are revisited to produce derivatives of concocidine and distamycin. These two natural products are pyrrole-amide peptide antibiotics, characterised by antibiotic, antifungal, antiviral and antitumoral properties, which arise from their non-covalent DNA-binding properties. Since both antibiotics are chemically very close, this thesis explores the combinatorial biosynthetic opportunities to exploit between the pathways that encodes them. Their chemical outline enables them to bind rich AT sequences in the minor groove of the DNA, with different preferred motifs according to specific differences in peptide length and functional groups. Research on synthetic compounds imitating these two natural products, have shown that the DNA binding affinity can be modulated by peptide length and heterocycle content. As synthetic chemistry represents an inefficient approach for the production of these derivatised compounds based on the natural scaffold of concocidine, the in-vivo substitution of the pyrrole rings for other heterocycles such as imidazole or thiazole is pursued through a mutasynthetic approach. The last part of this study focuses on the resistance mechanism present in both clusters. Since both antibiotics are closely related, the study of their resistance mechanism can help understand how natural antimicrobial evolution shapes natural antibiotic resistance, and the structure to activity relationship of novel compounds are explored.

Introduction

The urgent need

There is an urgent need to develop or discover new antimicrobial chemotherapeutics (Wise, 2011) due to the global crisis we are facing in antimicrobial resistant infections. This is a result of over 40 years of poor antimicrobial drug stewardship. The first resistance plasmids found in bacteria due to human antibiotic use were reported in the late 1950's and the fluoroquinolone resistance reported in the late 1990's (Davies and Davies, 2010). Very recently, bacteria has been isolated with a resistance to one of the 'last resort antibiotics' (colistin). Because of the enzymes allowing resistance are encoded in plasmid DNA, there is the danger of ending in global dissemination (Liu *et al.*, 2016). Consequently, we have the potential to return to the pre-antibiotic era. Passivity of big pharma due to profit-risk of development and patenting issues (Spellberg *et al.*, 2007) was one of the reasons that fuelled the exponential rise of multi drug resistant strains.

There are two obvious paths we can take to address this problem: drug development and drug discovery. Drug discovery focuses on compounds that have bioactivity. Those compounds could be synthetic, but there is a weight of experience that points towards natural products as the biggest pool of drug molecules (Baltz, 2008). The phylum Actinobacteria is by far the most prolific source of antimicrobials, with about two thirds of known bioactive compounds coming from this source (Baltz, 2008). It would therefore seem sensible to continue screening this phylum of Gram positive bacteria; however, discovering novel strains does not guarantee novel drugs, as many newly described strains produce previously identified natural products – the so called 'chemical dereplication problem' (Cortes-Sanchez and Hoskisson, 2015). The solution for this 'dereplication' of metabolites was initially to search in novel and previously unexplored ecological niches (Buchanan *et al.*, 2005), with marine bacteria being a known source of useful bioactive compounds (Bull and Stach, 2007). However, the discovery rate of novel bioactive metabolites is relatively low as the ones that have been already isolated, are the ones with a bigger appearance frequency. While streptothricin is present in 10 % of isolated streptomycetes, only 0.1 % produce tetracycline. Using statistical logic, it can be hypothesised that the compounds already isolated are common, while the drugs that may be isolated in the future present a lower occurrence in strains across the ecosystems. At the same time, is often that molecules

discovered are not suitable for therapeutic use due to safety regulations or pharmacokinetics (Lewis, 2013). Nevertheless, the biggest issue of this approach is the inability to culture the vast majority of microorganisms in a laboratory setting, with culturability rates ranging 0.1-0.3% (Amann *et al.*, 1995), and although new culturing techniques such as the iChip (Kaeberlein *et al.*, 2002) are helping address this problem, it is only with moderate success, i.e. the isolation of a completely novel antibiotic (Ling *et al.*, 2015) after 13 years and 10,000 novel strains isolated. A deeper knowledge on microbial ecology is needed as it may provide insights that will improve these poor cultivability rates.

Perhaps the solution to natural product discovery is not to discover more strains, but to exploit the plethora of biosynthetic pathways found within the genomes of already isolated organisms. The discovery of a new group of antibiotics, the turbomycins, was the result of heterologous expression of a natural product biosynthetic gene cluster obtained through a metagenomics screen (Gillespie *et al.*, 2002).

In 2002 researchers at the John Innes Centre first noticed the discordance between known and isolated natural products produced by a strain and those that were encoded within the genome. The model Actinobacterium *Streptomyces coelicolor* was known to produce three specialised metabolites however the genome sequence revealed 20 biosynthetic gene clusters putatively encoding novel metabolites (Bentley *et al.*, 2002). This proved not to be an exception, as later genome sequencing efforts in *Streptomyces avermitilis* confirmed, as it had 30 specialised metabolite biosynthetic clusters coding for potentially bioactive molecules (Ikeda *et al.*, 2003), albeit while expressing just the macrolide avermectin under laboratory conditions. The expression of these so called 'silent' or 'cryptic' biosynthetic clusters has proven challenging, with some successes achieved via heterologous expression with engineered repressors or activators (Laureti *et al.*, 2011; Olano *et al.*, 2014), by pathway specific regulatory gene disruption (Sidda *et al.*, 2014), cultivation under stress (Yoon and Nodwell, 2014) or exploitation of quorum sensing factors (Nodwell, 2014). The exploitation of these techniques depends on knowledge of *Streptomyces* metabolism, and their metabolic switches in response to signalling molecules. An example of such molecules would be the γ -butyrolactones family of compounds. The first characterised γ -butyrolactone was called A-factor (Horinuchi and Beppu, 2007) and along with other molecules such as 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (Corre *et al.*, 2008), they often play a role in secondary metabolism and morphological differentiation (Sidda and Corre, 2012). However,

falling sequencing costs (Check-Hayden, 2014) and the advances in synthetic biology are radically changing the scientific landscape, with the creation of completely new synthetic tools, i.e. promoters leading to expression of this clusters (Siegl *et al.*, 2013) or the use of pattern based genome mining to link metabolites to gene clusters (Duncan *et al.*, 2014).

The second avenue, drug development is often based on natural products too, as semi-synthesis relies in a natural product drug core, since the production of natural products is generally cheaper and more efficient than any analogous synthetic counterpart. While bacteria can be fed on cheap nitrogen and carbon sources to industrially produce modified compounds, their chemical synthesis is energy intensive and would require pure starting materials, contaminating solvents, and a highly specialised and skilled workforce. An example of the difficulty of chemically synthesising natural product is erythromycin, which was discovered in 1951, but with a full chemical synthesis reported 67 years later (Seiple *et al.*, 2016). However, biosynthesis of novel compounds also presents challenges. Natural product scaffolds are not optimised for human use; oral bio-viability, chemical stability and human toxicity were not evolutionary pressures for bacteria. Because of these problems, their functionalisation is often performed to improve pharmacokinetics or to concede new action mechanisms. Research from the end of the “golden era of antibiotic discovery” such as combinatorial biosynthesis (Hopwood *et al.*, 1985) proved possible “biological functionalisation of drugs”. Before that, another approach, mutasynthesis (Shier *et al.*, 1969) exploited enzymatic promiscuity to produce the *in vivo* functionalisation of natural products based on simple chemical “building blocks” fed to culture. The recent revival on these two approaches, relies on the current potential of recent advances in DNA manipulation and synthesis and a greater understanding of biosynthetic pathways and their biochemistry. At the same time, the development of more sensitive detection methods, permits the detection of molecules that may have been overlooked in the past.

Actinobacteria & *Streptomyces*

Actinobacteria are currently the largest described bacterial phylum, as a result of their capacity of produce bioactive molecules, despite not being the most abundant bacteria in the soil, and are found in most habitats globally (Goodfellow *et al.*, 2012). They were first described as a distinct group from other bacteria or fungi, based on their unique and complex morphologies, such as branching mycelium, hyphae and spores (Waksman, 1939), later the

genus *Streptomyces* was proposed by Waksman and Henrici (Waksman and Henrici, 1943). Despite this, over a long time they were thought to be a fungi, or the missing link between fungi and bacteria. That was because of their filamentous morphology and apical growth habit (Flardh and Buttner, 2009). The *Streptomyces* genus is the largest among the actinobacteria, comprising around 600 described species (Kämpfer *et al.*, 2008), with many novel species being described each year, i.e. (Hozzein and Goodfellow, 2007; Xiao *et al.*, 2009).

Most of the known species have been isolated from soil where they are primarily saprophytes within the ecosystem. *Streptomyces* have the capability to degrade carbohydrate polymers such as cellulose or lignin through secretion of hydrolytic enzymes into the medium (Pometto and Crawford, 1986). This is likely an adaptation to soil, as it is a rich carbohydrate environment with low levels of nitrogen and phosphate (Hodgson, 2000). Soil is a complex and highly variable environment due to its physical, nutritional and biological characteristics. Considering this oligotrophic environment, *Streptomyces* is able to grow in both poor and rich nutrient conditions, a quality that makes them facultative oligotrophs (Hodgson, 2000). This is perhaps the reason that *Streptomyces* is a widespread bacteria in a range of soils (Hodgson, 2000).

These Gram-positive soil bacteria have a complex life cycle which facilitates their survival in the complex soil environment and is summarized in Fig. 1; a single spore germinates and the resulting hyphae grows by tip extension and branching to form a dense mat of vegetative mycelium. Once nutrients are exhausted or in response to stress, aerial hyphae emerge from the substrate mycelium and grow upwards, curl and septate, to form uniform, unigenomic spores that can act as dispersal agents (Flardh and Buttner, 2009). It is at the onset of aerial hyphae formation that antibiotics are often produced and there is a wealth of genetic evidence that demonstrates an intimate link between the regulation of secondary metabolite production and morphological development (Chakraborty and Bibb, 1997).

Streptomyces coelicolor A3(2) is the model species for the genus given its long history of laboratory investigation (Chater, 1998). It has an 8,667,507 bp genome arranged in one single linear chromosome with an origin of replication situated in the centre at 4,269,853bp (Bentley *et al.*, 2002) and a circular plasmid with 31.3 kb called SC2 (Haug *et al.*, 2003). The largest described chromosome among *Streptomyces* is 10.1 Mbp belonging to *S. scabies* (Bignell *et al.*, 2010) and the smallest is *S. somaliensis* at 5 Mbp (Kirby *et al.*, 2012). The genus

is characterized by a high GC percentage (69-78%) within its DNA (Madigan , 2005). Most streptomycetes show extensive synteny (conservation in the gene order) in the central core region of the genome (Hopwood, 2007).

Coding density is uniform over the whole chromosome, with essential genes appearing to be located in this conserved core. The non-essential loci coding for functions such as inducible carbon metabolism and secondary metabolism are largely found in the arms of the chromosome (Bentley *et al.*, 2002) which possesses terminal ends (telomeres) which are formed by inverted repeats of diverse length that also have specific binding proteins in their 5' end which facilitate stabilisation of the chromosome ends (Bao and Cohen, 2003).

One remarkable feature of *Streptomyces* is their genetic instability, i.e. either of the ends of the chromosome can suffer deletions or amplifications up to two Mbp without noticeable changes to its metabolism under laboratory cultivation (Volf and Altenbuchner, 1998). Spontaneous deletions into the arms of the chromosome occur at a rate 0.1% per spores and can target the inverted repeats in the telomeres, leading to circularization of the chromosome.

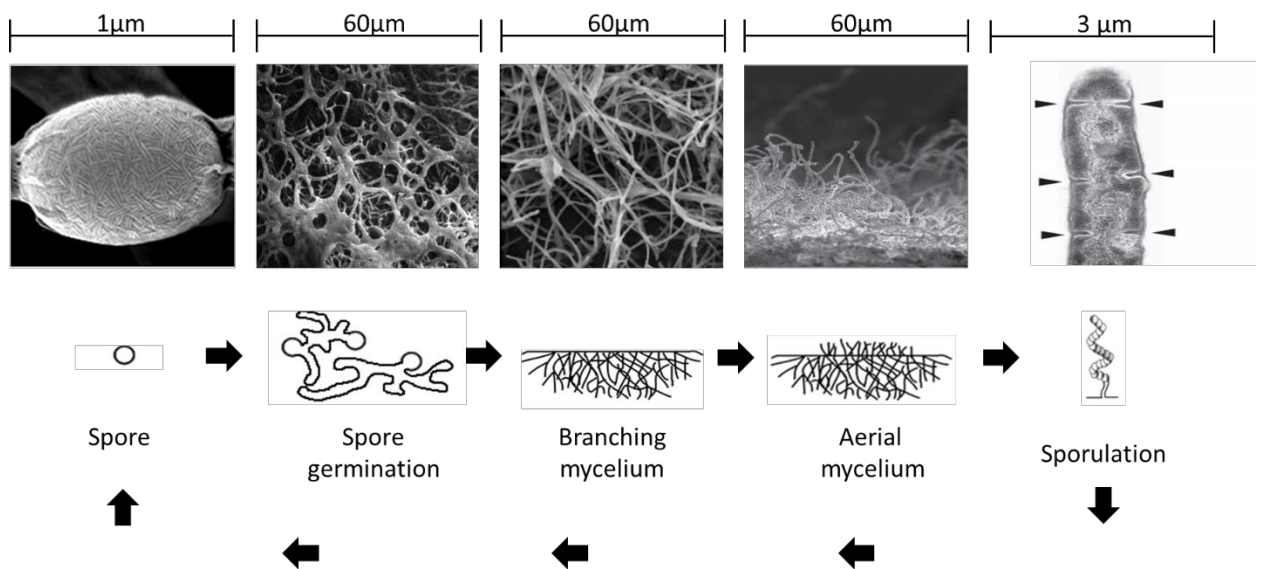


Figure 1. Life cycle of *Streptomyces*.

A single spore germinates and the resulting hyphae grows by tip extension and branching to form a dense mat of vegetative mycelium. Once nutrients are exhausted or in response to stress, aerial hyphae emerge from the substrate mycelium and grow upwards, curl and septate, to form uniform, unigenomic spores. Adapted from (Berardo *et al.*, 2008; Elliot *et al.*, 2008; Trujillo, 2008; Luzhetskyy, 2016)

Another characteristic of *Streptomyces* is that they are predisposed to form aggregates or pellets when cultivated in liquid media. This creates a myriad of different physiological states among cells in culture, which consequently creates different metabolic states with cells active on the pellet margins but oxygen and nutrient depleted on the centre (Meyerhoff and Bellgardt, 1995; Meyerhoff and Tiller, 1995). To address this problem various methods have been tested, such the inclusion of moving disrupting artefacts (i.e. steel springs or glass beads), cultivation in vessels with baffles to increase turbulence and addition of polymers to inhibit cell adhesion, however this only minimises the problem. Differentiated multicellular growth in streptomycetes is a particular problem during their industrial fermentation (Nieminen *et al.*, 2013), and a phenomenon that requires further research as it is intimately related to secondary metabolism and often involved in antibiotic production signalling (Claessen *et al.*, 2014).

***Streptomyces* secondary metabolism**

As the discovery of bioactive molecules from *Streptomyces* species gives a better knowledge in the details of their biochemistry, the interplay between “primary and secondary” metabolism becomes clearer.

The distinction between primary and secondary metabolism is inherited from the ideas of the Nobel prize Albrecht Kossel, who pioneered the concept of a primary or basic metabolism, universal and focused on basic cell cycle functions, whilst the presence of organism specific metabolism focused on non-essential metabolites called secondary metabolites (Hartmann, 2008; Firn and Jones, 2009). Such distinction was adopted in the biological sciences for a long time (Hiltner *et al.*, 2015). The precursor supply for a given specialised metabolite is provided from primary metabolic pathways, i.e. acetyl-CoA and malonyl-CoA (Olano *et al.*, 2010), or amino acids metabolism in the case of polyketide synthesis (Stirrett *et al.*, 2009). These relationships may represent metabolic conflicts affecting specialised metabolite synthesis (Rodriguez *et al.*, 2012), and knowledge on metabolite flux may provide insight on avenues to improve secondary metabolism productivity.

The evolution of primary metabolism is thought to be the product of gene expansion (Hiltner *et al.*, 2015), events in which gene duplications allow genes to become divergent in their function. Examples of these phenomena in *Streptomyces* are well documented, such as *devA* neo-functionalisation, or the changes in SsgA/SsgB adding complexity to morphological

differentiation (Clark and Hoskisson, 2011; Girard *et al.*, 2013). Perhaps natural selection favoured the survival of cells able to thrive in constantly changing environments, which rely on robust metabolism due to genetic redundancy (Kim and Copley, 2007). The presence of isoenzymes and closely related genes leads to gene or protein rearrangements resulting in enzymatic promiscuity, also called moonlighting (Copley, 2003), tending to preserve catalytic activities that provide a better fitness (Copley, 2012), which is a consequence of natural selection. These phenomena are not only due to random gene mutations. Horizontal gene transfer (especially important in microorganisms) may be playing a capital role in acquisition of novel metabolic capacities (Noda-García and Barona-Gómez, 2013).

The secondary metabolism in *Streptomyces* species is complex as its regulation depends on pleiotropic regulators (Madigan, 2005). Such complexity results in the production of a plethora of structurally different secondary metabolites, reflecting the abundance of different pathways. The isoprenoid geosmin is perhaps its most famous metabolite due to a characteristic earthy smell (Gerber and Lechevalier, 1965), and is often associated with an "after rain" smell. However other more useful secondary metabolites which they produce are antifungals, antivirals, anticancer drugs and antibiotics. In fact *Streptomyces* accounts for around two-thirds of the drugs that are used extensively in the clinic (Baltz, 2008). Industrial production relies on several filamentous microorganisms including *Streptomyces* (Wezel *et al.*, 2006). For example, a commonly prescribed drug for the co-treatment of β -lactamase producing pathogens, clavulanic acid, is industrially produced by *Streptomyces clavuligerus* (Townsend, 2002) by GSK in the UK.

The model species *Streptomyces coelicolor* A3(2) has been reported to produce four compounds with antibiotic activity (Gomez-Escribano and Bibb, 2011), actinorhodin or ACT, a blue pigmented polyketide compound (Wright and Hopwood, 1976), a calcium dependent antibiotic (Lane *et al.*, 1983), the hybrid NRPS/PKS derived prodiginines and the type I polyketide CPK (Pawlik *et al.*, 2007). This metabolic output contrasts with the presence of up to 20 secondary metabolite biosynthetic clusters (inferred from bioinformatic analysis) within its genome (Bentley *et al.*, 2002). The complexity of antibiotic production in actinobacteria relies on a complex network of cross-regulation coupled to morphological development illustrated by a number of pleiotropic regulators within the organism (Huang, *et al.*, 2005) and also a number of pathway specific regulatory mechanisms that exert a tight control on regulation of production. There is evidence of metabolic switches that induce

rearrangements to metabolism prior to the onset of secondary metabolite production (Nieselt *et al.*, 2010) which also affect cell differentiation. These switches are intimately related to primary metabolism (Alam *et al.*, 2010) and this 'nutritional sensing' exerts key temporal control over metabolic switches (Wentzel *et al.*, 2012). Some signalling molecules involved in these transitions have been identified, i.e. ppGpp or γ -butyrolactone (Bibb, 2005) which are associated with physiological stress (Rigali *et al.*, 2008) or quorum sensing. More recent studies at the transcriptional level show that the strongest switches in metabolism can happen during phosphate (D'Alia *et al.*, 2011) or nitrogen (Waldvogel *et al.*, 2011) limiting conditions, which confirms earlier genetic and physiological studies (Bibb, 1997). As of 2016 the factors controlling antibiotic production in *Streptomyces* are not completely understood, however it is clear that multiple levels of regulation occur for each metabolite which provides a multiplex of checkpoints to ensure that specialised metabolites are produced at the appropriate points within the organism's lifecycle.

***Streptomyces* produce a range of structurally diverse bioactive metabolites**

Genome mining and metabolite isolation show how prolific and diverse the secondary metabolism of *Streptomyces* are, with reports on non-ribosomal peptides encoded by non ribosomal peptide synthetases (NRPS), polyketides encoded by polyketide syntases (PKS), terpenoids, bacteriocins, shikimate-derived metabolites, aminoglycosides and siderophores (Nett *et al.*, 2009). Much of the metabolic capacity of streptomycetes is owed to the presence of PKS, NRPS and hybrid NRPS/PKS pathways. When looking at specialised databases such as StreptomeDB (Lucas *et al.*, 2013), which compiles more than 4000 compounds produced by more than 2500 listed strains (Klementz *et al.*, 2015), the most abundant biosynthetic pathway by class are PKS, NRPS and hybrid PKS-NRPS in that order, with an astonishing margin over shikimate, terpene, aromatic polyketide-terpene hybrid, glycoside, peptidyl-nucleoside or shikimate-polyketide pathways. Examples of secondary metabolites produced by these pathways are displayed in Fig.2. While studying the genes responsible for these pathways, it was shown that they can be horizontally acquired (Stinear *et al.*, 2004), which pointed towards the organisation of these genes into clusters (Paulsen *et al.*, 2005). Such clustering may facilitate the transmission of pathways, but also helped to clone and study such clusters.

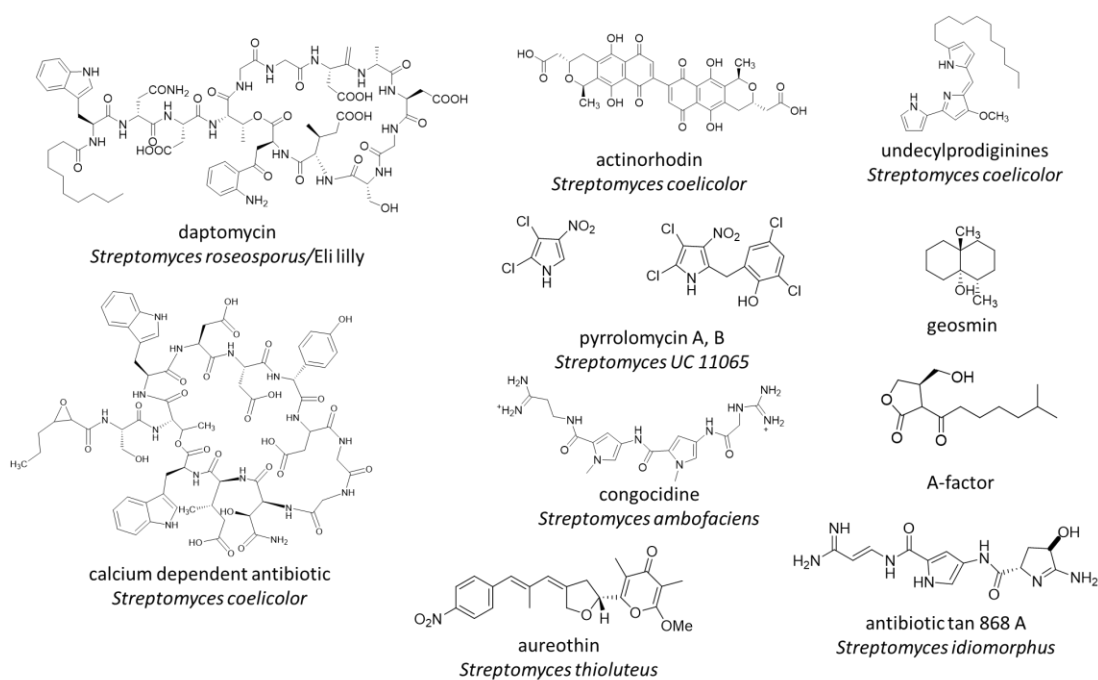


Figure 2. Examples of secondary metabolites produced by *Streptomyces* sp.

Daptomycin is a lipopeptide (Steenbergen *et al.*, 2005) and one of the last antibiotics brought to the clinic in 2003 (Charles and Grayson, 2004), it shares part of the scaffold with the Calcium Dependant Antibiotic isolated from *S.coelicolor* (Lane *et al.*, 1983). Actinohorhodin (Wright and Hopwood, 1976) and the pyrrolomycins (Norio *et al.*, 1981; Zhang and Parry, 2007) are an example of polyketide metabolites. congocidine is a non ribosomal peptide (Juguet *et al.*, 2009) and undecylprodigenines are PKS-NRPS hybrids (Pawlik *et al.*, 2007). Aureothin (Nakata *et al.*, 1961) is an polyketide-shikimate hybrid (Ziehl *et al.*, 2005). Geosmin (Gerber and Lechevalier, 1965) is of terpene origin, and the signalling A factor (Sidda and Corre, 2012) a lipopeptide (Cane and Haruo, 2013). Antibiotic tan 868 A (Takizawa *et al.*, 1987) biosynthetic pathway has not currently been elucidated.

The similarity at the chemical level between PKS and NRPS pathways may be a hint of why their occurrence in nature is higher than that of other pathways. Both use a set of acyl monomer units (this will be further explained for each pathway class) that follow a template based condensation assembly into oligomers (Fischbach and Walsh, 2006). In both cases the monomers follow an iterative cycle that involves A-activation, T-thiolation and C-condensation. The activation step uses ATP to activate the carboxyl group to be captured by a nucleophile in the case of the NRPS. This nucleophile is a thiol group in the thiolation/carrier protein domains (Stein *et al.*, 1996; Keating and Walsh, 1999) which drives chain elongation into the condensation domains in which a peptide C-N bond is formed in the case of NRPS or C-C bond in a Claisen condensation in the case of PKS (Fischbach and Walsh, 2006). The similarity of both systems can be appreciated in Fig.3 in which the A-T-C domains are in the same colour for both pathways, but labelled according to their nomenclature.

Non Ribosomal Peptide Synthetases (NRPS).

Non Ribosomal Peptides (NRP's), are a large class of natural peptides from bacteria or fungi that are synthesized by Non-Ribosomal Peptide Synthetases (NRPS), and have tremendous importance in terms of their pharmacological applications (Schwarzer *et al.*, 2003). The importance of these kind of peptides lies in their chemical diversity; while ribosomal encoded peptides use only proteinogenic amino acids, NRP's are capable of using non-proteinogenic amino acids, lactones, heterocyclic rings and multiple other precursors (Marahiel *et al.*, 1997) to assemble their peptide chains. In fact more than 300 different precursors have been identified in the past century (Kleinkauf and von Döhren, 1990) which leads to the synthesis of a myriad of structurally different natural products, i.e. bacitracin as shown in Fig.3.

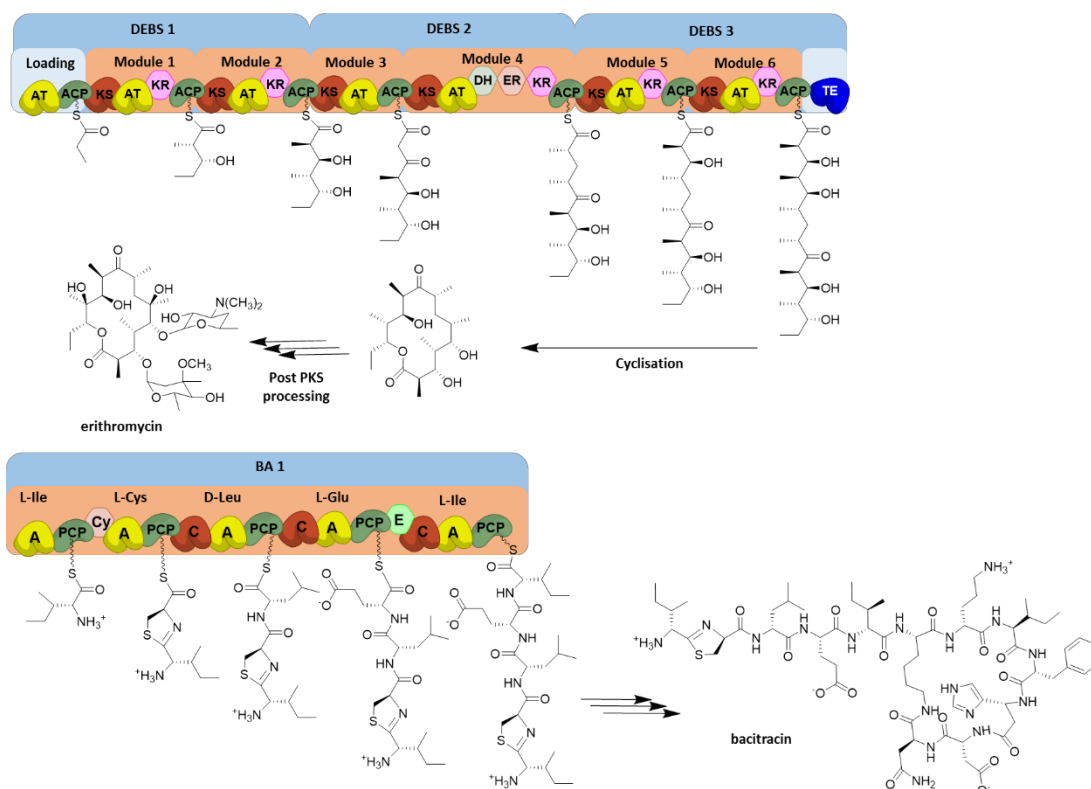


Figure 3. PKS and NRPS pathways are of modular nature.

Erythromycin biosynthesis illustrates modular PKS function. It represents a non iterative type I PKS pathway, in which a linear assembly applies. The molecule is biosynthesised with a starting loading module that uptakes propionyl-CoA, and it finishes by the action of a thioesterase. This PKS called DEBS (6-deoxyerythronolide B synthase) is comprised of six chain extension modules that adds carboxylated propionyl-CoA, expanding across three large proteins (DEBS 1-3). Each one of the extension modules contains an acyl transferase (AT), a ketosynthase (KS) and an acyl carrier protein (ACP) domain as well as variable numbers of other catalytic domains (DH, dehydratase; ER, Enoyl reductase; KR, ketoreductase) which are the responsible of introducing structural diversity on PKS. Adapted from (Shen, 2003; Weissman and Leadlay, 2005).

Bacitracin is assembled by a PKS-NRPS hybrid pathway. In the figure the five NRPS modules of the bacitracin synthetase are shown. Each module is constituted of three catalytic domains (A, adenylation; PCP, peptidyl carrier domain; C, condensation domain) as well as additional domains (E, epimerisation; Cy, cyclisation). Each of the aminoacids activated by the A domains are loaded onto the PCP protein which are in bold on the image. Adapted from (Konz *et al.*, 1997; Fischbach and Walsh, 2006).

Although NRPs are structurally very diverse, their synthesis follows a conserved pattern of assembly which is either linear, iterative or non-linear. The assembly pattern followed depends on the type of NRPS that encodes their biosynthesis. The biosynthetic polypeptides are comprised of enzymatic modules, each encoding a separate biosynthetic reaction which enables the biosynthetic process. There are three different NRPS classes:

Type I: the constituent modules are arranged in to a single polypeptide and are generally found in fungal species (Weber *et al.*, 1994; Jirakkakul *et al.*, 2008). The number and sequence of the amino acids in the final polypeptide matches the number of modules in the NRPS (Hur *et al.*, 2012). Type I NRPS produce molecules such as β -lactams, daptomycin, cyclosporin A, etc. (Felnagle *et al.*, 2011)

Type II: the modules are arranged in single, standalone proteins which interact between each other multiple times in a linear fashion, creating repeated sequences of amino acids, and are characteristically found in bacteria (Guenzi *et al.*, 1998; Keating *et al.*, 2000). Type II NRPS produce molecules such as quinoxalines, i.e. echinomycin and thiocoraline (Felnagle *et al.*, 2011).

Type III: the modules are single, standalone proteins that interact in an iterative or non-linear fashion generating peptides that do not match the putative order of the NRPS template (Hur *et al.*, 2012). Type III NRPS's produce molecules such as capreomycin and bleomycins (Felnagle *et al.*, 2011).

Regardless of the type, NRPSs are made up of modules containing a range of catalytic centers, which work coordinately to direct the synthesis of the final peptide product (Walsh and Fischbach, 2010). Characteristically there are three core motifs which always are present within the catalytic centers (Hur *et al.*, 2012):

First the Adenylating domain (A), is responsible for the recognition and activation of the building block by adenylation through the consumption of ATP.

Secondly, the Peptidyl Carrier Protein (PCP) which is a small peptide (80-100 aa's) acting in a manner analogous to a tRNA, transports the activated amino acids between catalytic centers through a thioester bond.

Thirdly, the Condensation domain (C) catalyzes the peptide bond formation with the downstream amino acyl unit that is tethered to the PCP of the adjacent module.

The so called 'A domain' is the first domain of the NRPS production line, and is often thought of as the "gatekeeper". It is responsible for the recognition and activation of its cognate amino acid in a two-step reaction (Fig.4A). This adenylation consumes one molecule of ATP, as AMP is incorporated to the amino acid. Once the amino acid is adenylated, it donates an electron pair via a nucleophilic attack from a thiol domain to the phosphopantetheine moiety of the PCP domain, creating an aminoacyl-S-PCP with a thioester bound (Lambalot *et al.*, 1996).

According to previous studies (Lee *et al.*, 2010) the main structure of an adenylation domain consists of a small C-terminal domain and a larger N-terminal domain, with the catalytic centre at the junction of these two regions (Hur *et al.*, 2012). This adenylation binding motif is conserved among the family of adenylate forming enzymes such the 4-coumarate-CoA ligases, acyl-CoA ligases and the oxidoreductases. This module catalyses the same reaction as the aminoacyl-tRNA-synthetases although no structural homology has been observed (May *et al.*, 2002), (Fig.4B).

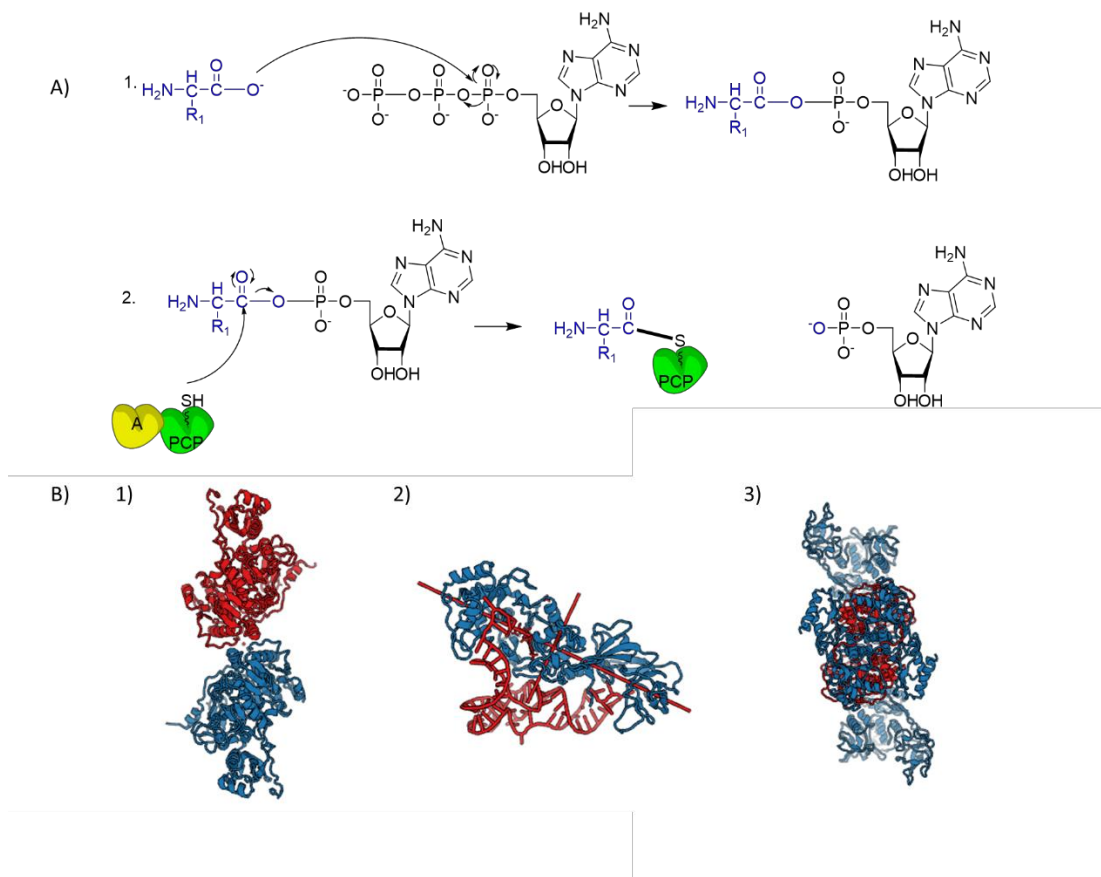


Figure 4. A) Two step adenylation mechanism. B) No structural homology between t-RNA aminoacyl synthetases and NRPS adenylation domains.

Adapted from (Clugston *et al.*, 2003) figure 4 A shows the 2 step mechanism in which the adenylation domain drives the creation of an aminoacyl-S-PCP bond. Figure 4 B) is adapted from (Weber and Marahiel, 2001), and in 1) shows the structure of the protein responsible for adenylation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S (Conti *et al.*, 1997), 2) shows the structure of a glutamyl-tRNA (Rath *et al.*, 1998) and 3) shows the structure of a phenylalanyl-tRNA (Reshetnikova *et al.*, 1999).

Within NRPS, the A domains share an identity ranging from 30 to 60 % which results in a core of conserved residues. Among those, eight to ten residues within the active sites of the adenylation core exhibit a degree of degeneracy and variation depending on substrate specificity. This has been exploited for the synthesis of new antibiotics (Eppelmann *et al.*, 2002), through the use of directed mutagenesis to modify selectivity in adenylation of non-cognate amino acids. Those experiments have helped to develop predictive algorithms that study the effect of point mutations in altering substrate specificity (Cheng-Yu Chena, *et al.*, 2009) based on solved crystal structures of existing A domains.

Additional studies on this family of enzymes has showed that at least two conformational changes are required to recognise and activate NRP substrates (Hur *et al.*, 2012). This discovery lead to the creation of bioinformatics tools that are based on the structure of adenylation domains allowing the prediction of substrate specificity of novel NRPS clusters. Examples of this tools are the NRPS predictor2 (Röttig *et al.*, 2011) integrated into the online web-tool ANTISMASH (Blin *et al.*, 2013) or the NaPDoS tool (Ziemert *et al.*, 2012).

The peptidyl carrier protein (PCP) carries the growing product chain across the NRP catalytic steps. This enzyme is first synthesized as an inactive Apo form. The PCP is then post-translationally modified by a phosphopantetheinyl transferase (PPTase) which covalently adds the 4' arm of a CoA molecule onto a conserved serine residue within the catalytic centre of the PCP (Conti *et al.*, 1997). After that the thiol moiety is added and the PCP is considered in its holo or active form, being able to create a thioester bond among the phosphopantetheine and its cognate aminoacyl substrate. Following the formation of the bond, the peptide is transferred from one PCP to the next, located in the adjacent module throughout the NRP assembly. In the NRPS modification the PPTase involved is a type II thioesterase which is somewhat promiscuous, leading to some mis-priming events (Hur *et al.*, 2012) which may be useful in biotechnological applications and probably the basis of evolution of NRP's through diversification.

The C domain is the largest of the catalytic modules with an average size of 450 amino acids. It is usually fused to an N-terminal domain that accepts acyl groups from the previous module (Stachelhaus, 1998). The catalytic center is situated in the junction between the N and C termini and is usually arranged in a V-shape. It was thought that a conserved histidine motif was responsible for the catalysis (HHXXXDG), but mutational studies shown that catalysis in C domains is based on electrostatic interactions rather than an acid/base mechanism

(Gordon *et al.*, 2005; Keating *et al.*, 2002). The V-shape of the domain facilitates the extension of the pantetheinyl arm up and downstream of the PCP. It also been shown that the C-terminal face of the V is non-selective against the substrate whilst the N-face shows strict stereo-selectivity (Clugston *et al.*, 2003). However, the C domain does not only accept aminoacyl and peptidyl substrates but is also able to condense polyketide intermediates and non-proteinogenic amino acids. Given the mechanism of NRP biosynthesis, it is useful to consider polyketides biosynthesis as they are similar in mechanism as its shown in (Fig.3), as well as PKS knowledge is necessary to understand hybrid polyketide/NRP pathways. The modular nature of NRPS, PKS and hybrid pathways makes them suitable and interesting for bioengineering; it was theorized that making changes in the catalytic modules could facilitate the incorporation of alternative substrates (Menzella and Reeves, 2007) . This proved to be true with reported successful incorporation of bio-ortogonal handles such as alkynil and azide groups that can undergo Huisgen cycloaddition "click" reactions (Winter and Tang, 2014), both in PKS (Koryakina *et al.*, 2013; Winter *et al.*, 2013; Bravo-Rodriguez *et al.*, 2015) and NRPS systems (Kries *et al.*, 2014) as displayed in Fig.5.

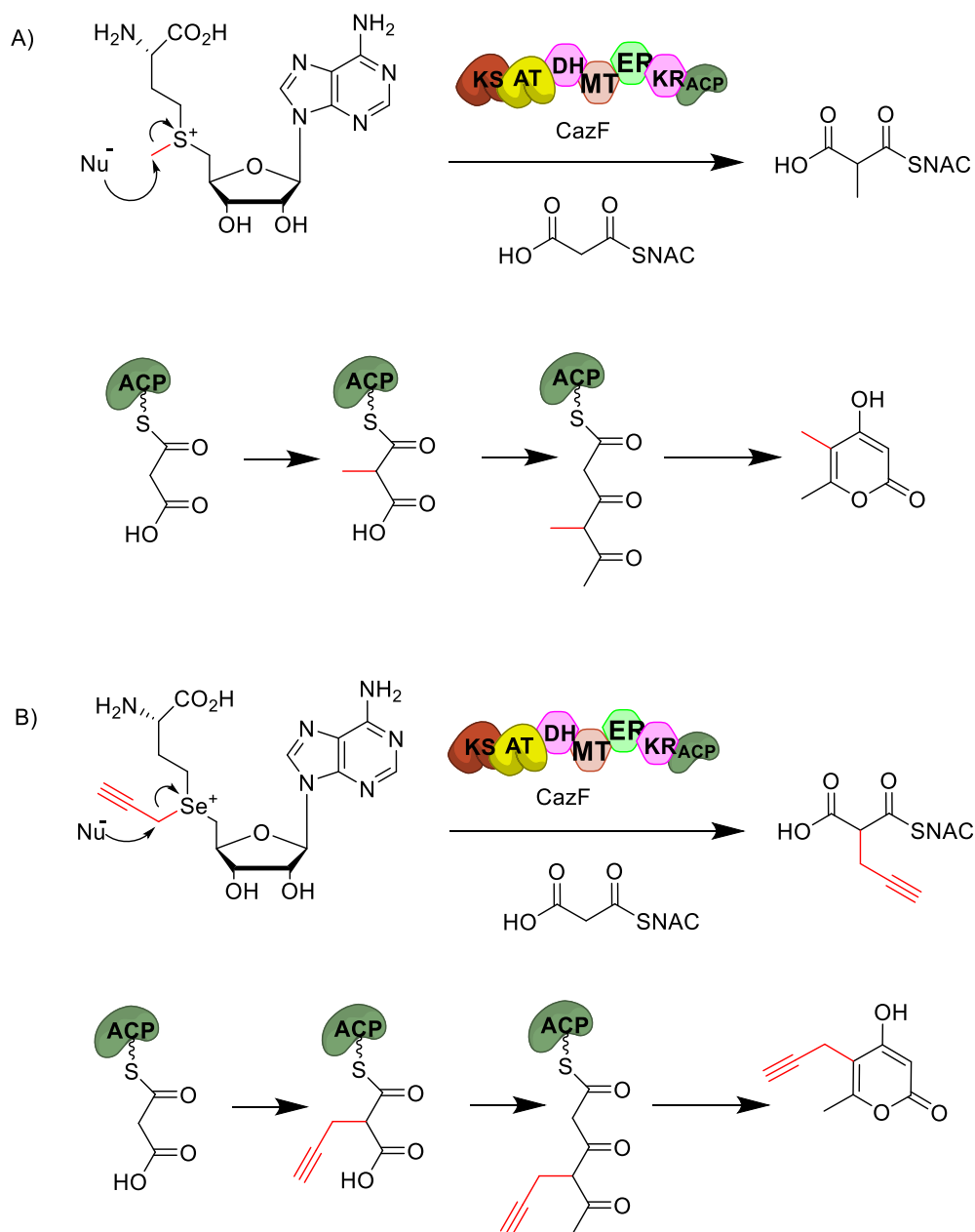


Figure 5. Incorporation of a propynyl "handle" with the fungal PKS chaetoviridin CazF.

Adapted from (Winter *et al.*, 2013), the figure A) shows the activity of the PKS CazF (Winter *et al.*, 2012) while incubated with the natural substrate. In B) the promiscuity of the methyl transferase domain is exploited to incorporate a propynyl group when incubating CazF with Se-adenosyl-L-methionine. The growing polyketide chain cyclizes on to a propargyl pyrone or can be fully processed into an alkyne containing 4'-propargyl-chaetoviridin A.

Polyketide Synthases (PKS).

Polyketides are molecules synthesised by polyketide synthases (PKS) which are also a family of multimodular enzymes forming large complexes. The synthesis of polyketides is similar to fatty acid biosynthesis (Jenke-kodama *et al.*, 2005) in the sense of using acyl and malonyl building blocks to catalyze two-carbon linear extensions in a Claisen type reaction (Khosla *et al.*, 2007). Presence of different reductase enzymes results in the vast structural diversity that is characteristic of this family of molecules (Meier and Burkart, 2012). To date, there are two major classes of these synthases, the *cis* and *trans* PKS type I, type II (Hopwood, 1997; Staunton and Weissman, 2001), and a third group or type III that are acyl carrier protein (ACP)-independent and work in an iterative fashion (Shen, 2003). The first PKS discovered and the most studied are the type I *cis*-PKS which are subdivided in modular (bacteria) or iterative (fungi) type I PKS. The canonical *cis*-PKS will have an acyltransferase domain (AT) that usually loads malonyl units into an acyl carrier protein (ACP). However other malonyl molecules are loaded, and sometimes discrimination within ethyl-malonyl CoA, methyl-malonyl CoA and malonyl CoA determines incorporation order i.e. in kirromycin (Musiol *et al.*, 2013). After loading, the ACP domain brings the unit to a ketosynthase (KS) module which catalyses a condensation reaction with another unit. All these reactions will follow a template in which the molecules are linked to the domains by thioester bounds, and the polyketide will be elongated until the synthesis is terminated by a thioesterase (TE) domain, that cleaves the thioester linkage of the polyketide from the enzyme (Hertweck, 2015). The diversity of the polyketide will be determined by the order and presence (or lack) of the 'tailoring' domains within the ACP and KS modules: ketoreductases (KR) which will reduce a β -keto group to a β -hydroxy group, dehydratases (DH) which will produce insaturations by H₂O removal, and enoylreductases (ER) which will reduce an α - β -double-bond to a single-bond.

The *trans*-PKS is less characterized although some researchers believe that their AT and ER domains present interesting modification opportunities (because are found on freestanding domains and follow an iterative assembly). More over, this type of PKS are responsible for the assembly of non-aromatic polyketides (Piel, 2010).

The Pyrrole Amides - congoicidin, distamycin and other minor groove binding molecules.

Deoxyribonucleic acid (DNA) was known since the 1940's, but only in 1952 did researchers first propose a structure for it (Pauling and Corey, 1953). It was a year later however, when the biological role and structure of DNA was accurately unravelled by Watson and Crick; a double helix formed by two antiparallel strands made up of a sugar-phosphate backbone. In such structure the sugar (2'-deoxyribose) is linked to the phosphates in the positions 3' and 5', and the helices are held together by interaction between purine and pyrimidine bases, which are covalently attached to the C1 of the 2'-deoxyribose through a N-Glycosidic bond (Crick et al., 1953). Such base interaction followed the so-called Watson-Crick rule, establishing Adenine hydrogen bonding with Thymine, and Guanine with Cytosine (Fig.6a); this rule was a pairing model based on the work by Chargaff (Elson and Chargaff, 1952). The double helix is flexible and because of that present several polymorphisms that emerge from differences in the orientation and distance between the bases and the helix axis of symmetry; the most common DNA geometries are the A-form (Fernandez et al., 1997), the B-form (Drew et al., 1981) and the Z-form (Wang et al., 1979), although the B-form is believed to be the most common among cells (Richmond and Davey, 2003). In B-conformation, the antiparallel arrangement of the strands results in an asymmetrical distribution of the bases, leading to the formation of two differently sized grooves (Fig.6b); the major groove that is 22Å wide and the minor groove 12Å wide (Wing et al., 1980). This structural difference has a biological consequence in transcription processes since the wider groove can better accommodate different regulatory proteins (Pabo and Sauer, 1984). Interestingly, almost all the natural DNA binding peptides, specifically bind rich AT regions in the minor groove (Kopka et al., 1997). The chemical features that make pyrrole amides bind specifically in the minor groove are discussed below.

congoicidin (also known as sinanomycin/netropsin) was isolated from an uncharacterised streptomycete (Finlay. *et al*, 1951). It was named *Streptomyces netropsis* at first, but later classified as *Streptomyces distallicus* by the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). Distamycin was isolated 12 years later by Italian researchers (Arcamone *et al.*, 1964) from a *Streptomyces netropsis* strain which was not producing congoicidin. These pyrrole amides proved to be effective antimicrobial, antifungal, antiviral and antitumor drugs, however their promising potential was not exploited as they proved to be too toxic

for human therapeutic use (Foye, 1999). Nevertheless, attention was brought to them as novel DNA binding peptides, since classic synthetic DNA binding drugs were either intercalating or alkylating agents, which generally showed low DNA sequence selectivity (Ren and Chaires, 1999) and very few of them bind to the minor groove (Bose *et al.*, 1992).

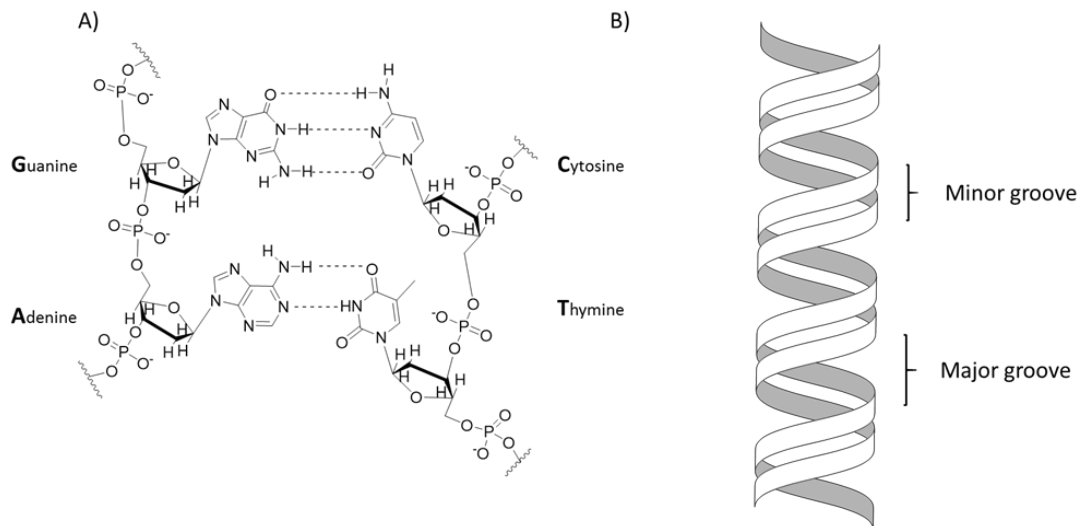


Figure 6. DNA purine-pyrimidine base interactions and helix geometry in B-form DNA.

Figure 6 A) shows the pairing between DNA bases, by hydrogen bonding with three interactions between bases Guanine and Cytosine or two interactions between Adenine and Thymine. Adapted from (Guerra et al., 1999). Figure 6 B) show the major and minor groove in the B-DNA. Adapted from (Wing et al., 1980).

Both natural products share a methyl pyrrole amide chain (two rings in congocidine, three rings in distamycin) that terminates with an ethylenamidino tail. Apart from the pyrrole amide chain length, the major difference between the compounds is in the 'head' region of the molecule, with an N-formyl group in the case of distamycin versus an N-carbamimidoylglycyl amino moiety in the case of congocidine (Fig.7A-B).

These features makes them able to establish DNA interactions in the minor groove of the DNA helix; the amide groups establish Van de Waals with the N-3 in adenine or O-2 when the nucleotide is thymine, in rich AT sequences (Kopka *et al.*, 1985). The different length and moieties in congocidine and distamycin impact in the manner to which they bind to DNA. congocidine binds to AATT motifs (Fig.7C) forming at least two different 1:1 complexes (Lewis *et al.*, 2011). Distamycin was also thought to bind in 1:1 complexes, although with an extended binding sequence (five nucleotide AT rich sequences) following the same amide-base interactions, as an early crystallography (Coll *et al.*, 1987) study reflected. However, the formation of 2:1 complexes (Fig.7D) was discovered by NMR not long afterwards (Pelton and Wemmer, 1989), indicating possible discrimination between end base pairs reversals (TA for AT) and potentially explaining the bigger binding footprint of distamycin. Those facts along with their chemical composition made them become lead compounds of a class of molecules that bind to dsDNA, and inspired chemical synthesis of compounds, namely lexitropsins (Neidle, 2001), which module gene expression, i.e. human glioma cell lines (Bobola *et al.*, 2007).

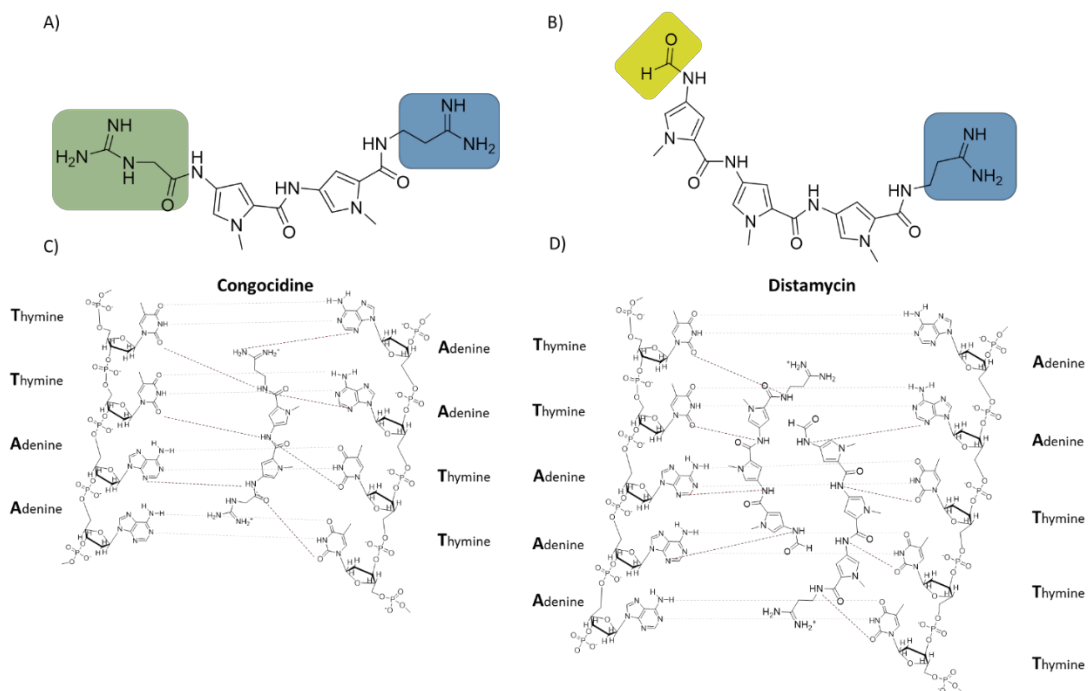


Figure 7. Differences between congocidine and distamycin molecules and their binding to DNA.

Figure 7A) and B) shows the different head and tail moieties presents in the methyl bi-pyrrole amide chain (congocidine) and the three methyl pyrrole amide chain (distamycin). The shared ethylenamidino tail is highlighted in a blue background where as the different heads are highlighted in green for the N-carbamimidoylglycyl amino congocidine moiety, and in yellow for the N-formyl group in the distamycin molecule.

Figure 7C) shows the interaction of a 1:1 complex of congocidine and a -TTAA- DNA sequence, whereas Figure 7D) shows how distamycin forms 2:1 complexes while binding a -TTAAA- DNA sequence. Figure adapted from (Dervan, 2001).

Once the bases of peptide-DNA interaction for these compounds were clear, researchers realised the potential of altering the DNA binding sequence by replacement of the pyrrole moieties for imidazole rings (Kopka et al., 1985). The pursuit of this hypothesis led to artificial minor groove binders, the so-called lexitropsins, able to exchange A-T for G-C binding selectivity due to imidazole N-3 hydrogen bonding with the guanine amine (Kopka et al., 1997). Subsequently, research on these natural-based compounds split into two trends. One was the making of large synthetic ligands based on pyrrole-imidazole amide chains that bind with high selectivity long (up to 16bp) DNA sequences (Yang et al., 2013). The second trend focused on keeping a small molecule but introducing several structural changes, such as tail-head substitution (to modulate the binding affinity and polarity), changing the N-alkyl moiety (to modulate polarity) or heterocycle variation from pyrrole to imidazole, thiazole, furan, oxazole or thiophene (Khalaf, 2009). Those changes were done as it was known that certain heterocycles such as thiazoles and oxazoles play a key structural role in several natural products (Walsh and Nolan, 2008). Early research using the latter approach produced a congoic acid derivative called thiazotropsin. The inclusion of an isopropyl substituted thiazole in the amide chain instead of an N-alkyl imidazole, favoured selectivity against C or G while increasing the lipophilicity of the molecule. Such compounds were found to bind several sequences, although the highest selectivity was found to match 5'-ACTAGT-3' (James et al., 2004), binding DNA in a 2:1 complex that slightly deforms (widens) the minor groove (Anthony et al., 2004). Later research demonstrated that sequence specificity of such molecules could be completely changed. The replacement of a methyl pyrrole by a methyl imidazole on thiazotropsin led to thiazotropsin B that had a 5'-(A/T)CGCG(A/T)-3' sequence specificity (Hampshire et al., 2006). This knowledge, and a continuing synthetic minor groove binder screening campaign focused on antimicrobial hits, led to the development of clinically relevant drugs (Parkinson et al., 2013) and eventually to a novel drug, BP3 which showed antibiotic, antifungal and anti-parasitic activities. This compound shows low toxicity to mammalian cells, perhaps due to uptake/efflux kinetics, although research has to be done in that area. A remarkable feature of this compound is that it acts on several bacterial targets, as RNA-seq studies on *Staphylococcus aureus* have shown (Suckling, 2015). As several adaptations should occur simultaneously, this drug has the potential to elude common resistance mechanisms, thus expanding the time until resistance appears.

This suggests that the pyrrole amide molecules have great potential for therapeutic use, as chemical synthesis has proven. Biosynthesis of these molecules, or modification of other

natural occurring pyrrole amides could result in the development of novel molecules with therapeutic utility.

Mutasynthesis

The two biggest problems associated with antibiotics are toxicity and resistance. In the late 1960's the chemical functionalisation of kanamycin was explored in order to overcome some of these toxicity and resistance problems (Akio and Hamao, 1968). Kanamycin is a member of the deoxystreptamine antibiotics, a group composed by more than 20 different antibiotics with subtle differences in terms of their functional groups and chemical decorations covalently linked to a common core, which represents an example of natural derivatisation (Rinehart, 1969). That observation and the inability of either total synthesis or partial modification of such core, not to mention more complex antibiotic such as neomycin, inspired the search for an alternative to chemical synthesis. Since the discovery of penicillin it was known that media composition could affect the metabolites produced (Behrens, 1949). These observations led researchers to pursue a precursor directed biosynthesis. In the canonical precursor directed biosynthesis experiment, intermediate "building blocks" of the natural product are identified, synthetic analogues so-called "mutasyntons" (Weist and Sussmuth, 2005) are synthesised and then fed to the producing strain. Unless the synthetic surrogate completely outcompetes the natural precursor, a mixture of the original and derivatised compound will be obtained in the final fermentation. However, when a mutation blocks the supply of the natural precursor, the enzymes only can use the synthetic analogues fed to the strain in the fermentation.

The first example of this approach focused on *Streptomyces fradiae* which produces neomycin. Through the use of a deoxystreptamine biosynthetic mutant, cultivation of the strain in the presence of 1,3-diamino-1,3-dideoxy-scylo-inositol or 1,3-diamino-1,3-dideoxy-myoinositol instead of the naturally biosynthesised deoxystreptamine resulted in the production of hybrimycins A1 and A2 and B1 and B2 respectively (Shier *et al.*, 1969). Although these novel compounds were not as biologically active as the original, diversity through a novel approach was achieved, and the basis of modern mutasynthesis set.

Since then, this approach has been used to modify a range of molecules, including aminocoumarins (Galm, 2004), non-ribosomal peptides (Weist, 2002; Weist *et al.*, 2004; Stegmann, 2005), nucleosides (Christiane and Attila, 1999), polyketides (Boddy *et al.*, 2004; Gregory, 2005; Moss, 2006) and siderophores (Ankenbauer *et al.*, 1991).

The revival of this technique not only relies on the need for novel antibiotics, as it is also used on other applications such as bioremediation (Yu *et al.*, 2014) and to produce other useful natural products. Developments in the chemical and biological sciences, have also play a fundamental role in continuing this technique, which was not particularly fruitful in the past. Researchers have now a better insight on the range of metabolic capabilities of microorganisms through advances in molecular biology and analytical chemistry. Cheap sequencing and comparative bioinformatics have simplified the work-flow to characterise a natural product biosynthetic pathway, allowing quicker access to details of assembly of compounds that had previously been obscured. At the same time, novel DNA manipulation techniques allow for easier and faster gene knock outs or modifications. In addition, libraries of thousands of small molecules that were synthesised in short scale for particular assays, are now commercially available, and the synthesis of intermediates have also been simplified (Weissman, 2007). All of the above can be exemplified with the antibiotic and antifungal aureothin. It was discovered in 1960 from a culture of *Streptomyces thioluteus* (Nakata *et al.*, 1961), however its biosynthetic pathway was obscure until 2003, when researchers were able to identify the gene cluster responsible for assembly of the peptide (He *et al.*, 2003). It was found to encode several PKS enzymes, with a set of them dedicated to para-amino-benzoic-acid (PABA) synthesis early in the assembly of the compound. This knowledge was used by the researchers to produce mutants unable to biosynthesise PABA. These mutants were unable to produce aureothin, however, when fed with para-cyano-benzoic acid (PCBA) they were able to produce a novel molecule, aureonitrile (Ziehl *et al.*, 2005).

Beyond the interest in relation to functionalisation of molecules (in the case of aureothin - aryl starting unit, oxygenated backbone and pyrone head group), this work provided to the field of mutasynthesis with a very useful information. It established that incorporation of intermediates into a final compound not only depended on enzymatic engineering but on the choice of heterologous expression strains. This appreciation was done as some non natural substrates which were not incorporated in the 2005 study (Ziehl *et al.*, 2005), were succesfully incorporated by the same enzymatic pathway when changing the host (Werneburg *et al.*, 2010).

This highlighted one of the biggest challenges that this technique faces, the ability to discern which molecules can be incorporated to which pathways, as well as understanding the

bioavailability and transport of the mutasyntons to the cells. Perhaps a systematic study of related pathways will outline some of the rules governing substrate uptake.

Although the rules for substrate recognition by adenylating enzymes are known due to comparative studies on residues within the phenyl-alanine binding pocket (Stachelhaus *et al.*, 1999), some researchers create 'jump start' mutasyntons, such as N-acetylcysteamine thioesters (SNAC) derivatives, since they mimic acyl-carrier-protein (ACP) thioester and are able to diffuse through the cell wall (Staunton and Sutkowski, 1991). However, in some clusters containing polyketide synthases or non-ribosomal peptide synthetases, the adenylating domains are promiscuous enough such that feeding the free carboxylic acid or the SNAC derivative of the mutasynton does not affect the incorporation rate (Werneburg *et al.*, 2010; Lautru *et al.*, 2012).

Only adenylating (A) and acyl transferase (AT) specificity can be circumvented. Escaping the selectivity of these enzymes has can be done by creating SNAC derivatives of their natural substrate. However, the selectivity of enzymes downstream may be problematic, especially when there is little known about the rules governing their chemistry. Although research has shed light on NRPS systems, the most extensive is on polyketide synthesis (Nguyen *et al.*, 2008; Piel, 2010). Although research on the enzymes within these clusters has been done, only a few observations have been reported. A clear one is that enzyme specificity is tied to the electronic configuration of the substrates (Brendel *et al.*, 2011), but general guidelines are not established, although that could change with novel ELISA based probing technologies (Ishikawa *et al.*, 2015) that can detect active-site phenotypes with desired activities. There is a distinct lack of published research on mutagenesis of enzymes to push forward the uptake of mutasyntons especially where the biosynthetic pathways fail to incorporate fed materials. This is likely due to the complex arrangement of the large multimodular enzymes within these biosynthetic clusters, especially considering the ample array of different substrates, but may also reflect uptake dynamics. So far the efforts in this area had been largely put into random mutagenesis screening, although the output of such approach often renders completely inactive proteins (Bershtein and Tawfik, 2008) or semi-rational screening of single enzymes with some successes (Chen *et al.*, 2012), although structure guided mutagenesis is quickly advancing the field (Hoffmann *et al.*, 2011; Moretti *et al.*, 2011). In the absence of a structural knowledge, a combination of error-prone PCR and site-saturation mutagenesis can identify mutations that improve catalytic activities. This "enzyme-directed

mutasynthesis” is undoubtedly able to complement the efforts on natural product derivatisation, as researchers proved possible on PKS forcing the incorporation of a propargylated malonic acid thioester into erythromycin through a single point mutation (Sundermann *et al.*, 2013), and on NRPS by inclusion of O-propargyl-L-Tyr in gramicidin S by Trp to Ser mutation in one of the adenylation domains (Kries *et al.*, 2014) providing orthogonal handles on the natural products for easy ‘click’ reactions. Nevertheless in nature, genetic diversification usually does not rely on single point mutations, and often includes global gene rearrangements (Goldsmith and Tawfik, 2012). It is likely that the future of mutasynthesis will be complementary to other bioengineering approaches such as combinatorial biosynthesis.

Combinatorial biosynthesis

Combinatorial biosynthesis, like mutasynthesis, was developed to potentially answer the research community need to diversify the range of known natural products. The first example of combinatorial biosynthesis emerged in 1985, from the Hopwood laboratory at the John Innes Centre, in which researchers pioneered the idea of combining biosynthetic gene clusters from different antibiotic-producing strains in the same cell. They were able to create new hybrid compounds, starting a new route to, ‘unnatural natural products’ (Hopwood, et al 1985). At this time the chemical structures of the isochromanequinone antibiotics were known but their biosynthetic pathways were poorly characterised. However, researchers narrowed down a minimal set of genes that was necessary for the synthesis of one of these antibiotics, actinorhodin. Introducing the actinorhodin cluster into the madermycin producing *Streptomyces* sp. enabled cross talk between the two biosynthetic pathways and the production of a novel molecule, maderrhodin (Hopwood et al., 1985). The same approach was applied with success (Mcalpine *et al.*, 1987) to a more commercially relevant class of antibiotics, the macrolides.

Later on, sequencing brought genetic knowledge on the pathways responsible for antibiotic assembly. Elucidating enzymatic activity from the genes that encode them led to studies on gene disruption. An example of this was the prevention of hydroxylation on certain positions on erythromycin (Weber *et al.*, 1991). More insights into the modular architecture of the polyketide synthases in the erythromycin biosynthesis, allowed the production of the first point mutations to inactivate specific catalytic centres without disrupting a whole module,

which resulted in the introduction of saturations in to the lactone ring of erythromycin (Donadio *et al.*, 1993).

Insights into polyketide synthesis enabled the construction of the first structure-function rules (Hutchinson, 1995) including the study of reduction, oxidation and methylation mechanisms as well as targeted modification of type I PKS and shuffling of type II PKS. Much of this early work focussed on PKS rather than in non-ribosomal peptide synthesis. While polyketide synthesis is not completely understood, it allows for production of derivative libraries (Lenik *et al.*, 2015)..

Peptide antibiotics synthesis were subdivided in two classes, either peptides that were synthesised by the ribosome following amino acid incorporation and then modified post-translationally (Lipmann, 1980), or peptides synthesised with multidomain enzymes following a thio-template that incorporated acyl-adenylates of amino and hydroxyacids with the aid of another enzymes instead of tRNAs (Marahiel, 1992). The enzymatic layout of non-ribosomal peptides only began to be defined towards the end of 1994 when the minimal size of functioning domains was determined by examining the gramicidin S synthetase number 1 or GrsA (Stachelhaus and Marahiel, 1994). These findings set the basis to explore the conserved motifs of the enzymes responsible for the amino acid recognition and activation, which resulted in the first combinatorial experiments in non-ribosomal peptides. A knock out mutant of lipopeptide surfactin was created in *Bacillus subtilis* and complemented with an operon coding for a fusion protein of homolog proteins in *Bacillus brevis* and *Penicillium chrysogenum* (Stachelhaus *et al.*, 1995), which resulted in production of several new surfactins, although none of them with improved bioactivity.

The current knowledge of the enzymatic structure of both PKS and NRPS led to four main combinatorial approaches: domain swapping, site-specific mutagenesis, directed evolution and pathway recombination, with some authors considering mutasynthetic experiments a branch of combinatorial biosynthesis (Sun *et al.*, 2015). Examples of these approaches for engineering of PKS pathways are numerous, but as in the early research on combinatorial biosynthesis, NRPS studies are far less reported despite being one of the biggest sources of secondary metabolites (Doroghazi *et al.*, 2014). Since polyamide biosynthesis (i.e. congocidine, distamycin) is governed by NRPS, the current developments of combinatorial biosynthesis will focus in such literature for the rest of the text. The classification of mutasynthesis as combinatorial biosynthesis by some, is due to the tailored precursor supply

concept. While exogenous compounds fed to culture are clearly a form of precursor directed biosynthesis, producing those *in vivo* by altering enzymes that supply those precursors could be classified as pathway engineering, although not necessarily combinatorial biosynthesis; this would depend upon which genetic modifications were made.

An example of this would be the work on pacidamycin, which is an uridyl antibiotic made by *Streptomyces coeruleorubidus*. The NPRS cluster responsible for its synthesis was expressed alongside a tryptophan-7-halogenase from *Pseudomonas fluorescens Pf-5*, which resulted in incorporation of a halogenated tryptophan (Roy *et al.*, 2010), this then allows cross coupling synthetic modifications afterwards. While the yield was low (around one mg per litre), it was at the same level as the productivity achieved by a previous study on precursor directed biosynthesis with chemically made 7-chlorotryptophan (Grüschow *et al.*, 2009). One of the main attributes of these systems is that by using such a combinatorial biosynthesis approach, no specialised metabolite had to be synthesised for feeding experiments. However, this kind of approach is more challenging than traditional mutasynthesis and should be focused on non-proteinogenic amino acids, since introducing modifications in the amino acid metabolism could compromise strain growth and the creation of auxotrophic strains would be required (Micklefield *et al.*, 2008).

Synthetic biology

Although synthetic biology seems to be a modern term, it has been in use since the late 1970's (Waclaw, 1978). Several descriptions of synthetic biology are made by chemists, bioinformaticians, engineers and biologists. However the BBSRC defines it as 'the design and engineering of biologically based parts, novel devices and systems as well as the redesign of existing, natural biological systems' (BBSRC, 2016). Although biological systems are hard to model and manipulate, current technology is starting to enable the application of true engineering principles. Some of the key technologies which are responsible for producing an 'engineering view' of biological systems include DNA synthesis and manipulation, DNA sequencing, and digital modelling.

The first chemical synthesis of a gene was reported by the Nobel laureate Khorana and his colleagues, who synthesised a 77 bp long alanine tRNA in yeast (Khorana *et al.*, 1972). Later research suggested a technical limit for synthetic assembly of nucleotides at 200 bps due to cumulative errors and coupling efficiencies (Beaucage and Iyer, 1991), which turned out to

be very accurate, given the fact that the current technology allows for reliable chemical synthesis of around 150 bps nucleotides (Leproust *et al.*, 2010). However, DNA synthesis benefits from these chemically made oligomers that can be assembled together to create bigger fragments. The true revolution in the field happened in 2010 when the Craig Venter Institute made public the assembly of the complete mouse mitochondrial genome (Gibson *et al.*, 2010), and the breakthrough feat of the creation of a bacterial cell controlled by a 1.08 mega base chromosome built entirely synthetically (Gibson *et al.*, 2010). Both achievements were supported by a technique reported a year before, which later on became the commercially available 'Gibson' cloning kit (Gibson *et al.*, 2009). The technique combined three enzymes (exonuclease, ligase, and polymerase) to produce DNA assembly in a single isothermal reaction. This innovation was not the only methodology followed, since several other approaches were pursued. Those were based on the PCR overlap extension of DNA (Higuchi *et al.*, 1988), the ligase cycling assembly (Wiedmann *et al.*, 1994) or in yeast based methods, such as the transformation-associated recombination (TAR) cloning (Burke *et al.*, 1985). Despite success with these techniques, there are still technical limitations in DNA synthesis since all of the current methods rely on single strand alignment. The biggest mishaps in DNA assembly are due to secondary structures, inverted repeats or repetitive sequences, with sequences with extremely low or high GC content being the most problematic.

A second pillar of synthetic biology is DNA sequencing. The sequencing technologies experienced a revolution in the last ten years with the emergence of next generation sequencing (NGS). During the sequencing of the human genome (International Human Genome Sequencing Consortium, 2001), it became clear that high throughput, fast and cheaper sequencing technologies were needed. It was then that an initiative within the research community was set, to achieve the \$ 1,000 genome (Schloss, 2008). That premise helped to develop and commercialise the second generation of sequencers. The three major innovations from the sanger sequencing were the cell free approach to produce DNA libraries, increased parallel sequencing reactions and direct detection of the sequencing output without need for electrophoresis (Jaszczyszyn *et al.*, 2014). Several technologies are now able to produce a \$ 1,000 human genome, and are chosen depending on application due to their cost, read length and sequencing error rate. The latest achievement on the field is the creation of a hand sized sequencer, the Min-ION, able to produce genome reads up to a bacterial chromosome in the field (Check Hayden, 2015). However this portable

sequencer is based on protein pore technology (Jaszczyszyn *et al.*, 2014), which is error prone although it produces long reads, which allow for algorithm correction of some of those errors (Madoui *et al.*, 2015).

With the availability of laboratories to access the DNA of different species, as well as the need to assign function to the genes sequenced, the bioinformatics and modelling of data has become the third mainstay of synthetic biology. As in any other engineering discipline, synthetic biology needs the creation of models that can simulate the behaviour of certain synthetic gene networks (Kaznessis, 2007), so that the research community can start using computer aid design (CAD) to produce engineered microorganisms. The development of bioinformatics is also helping to address the antimicrobial crisis, i.e. creating database homology based predictors such as AntiSMash (Medema *et al.*, 2011; Blin *et al.*, 2013) or NaPDoS (Ziemert *et al.*, 2012) that search for natural products clusters among un-annotated DNA.

Resistance mechanisms in bacteria.

Antibiotic resistance is a mechanism of survival that has existed since the first antibiotic was biosynthesised, so that it protects the producer itself. Knowing this, it is easy to conclude that resistance will always be the menace of the perfect antibiotic as a consequence of natural selection. But perhaps resistance can be studied to create 'better' antibiotics, so that the time frame between drug creation and the appearance of resistance is long enough to make it a useful drug in clinic.

There is a distinction to be made regarding how bacteria become resistant antibiotics. Usually reports of antimicrobial resistance focus on resistance mechanisms of pathogenic bacteria or bacteria to which the antibiotic tested is not endogenous. Resistance mechanisms are sometimes shared by pathogens and antibiotic producers alike (and will be discussed later), but some are intrinsically self defence mechanisms. One example of the latter can be seen with platensimycin and platencin - two potent fatty acid inhibitors (selectively targeting FabF, an enzyme involved in fatty acid biosynthesis; Wang *et al.*, 2006), which had no obvious resistance mechanism within the genetic cluster containing its biosynthetic enzymes. Bioinformatic analysis of the sequence focused on genes that had no known role in antibiotic or fatty acid synthesis. In doing so, researchers identified the ptmP3 / ptnP3 gene, which

encode an enzyme able to alter the FabH and FabF fatty acid synthases, making them insensitive to the drugs by target modification and replacement (Peterson *et al.*, 2014).

Other less sophisticated but equally effective resistance mechanisms are the classically reported ones: drug enzymatic degradation, alteration of the drug target or efflux of the drug (Dever and Dermody, 1991). While the most reported resistance mechanism in the clinic is enzymatic hydrolysis, especially by β -lactamases (Wright, 2005), perhaps the most prevailing, especially among multidrug resistant strains, is drug efflux (Bambeke *et al.*, 2000). Since efflux mechanisms are omnipresent in all forms of life, it is easy to conclude that their overexpression is an easy way of removing toxic products from the cells. These mechanisms are either single or multi drug resistant proteins (SDR/MDR) generally associated with an ATP binding cassette (ABC type transporter). When not connected to an ABC, these mechanism are mediated by the Major Facilitator superfamily proteins (MFS type) (Prasad *et al.*, 2015), and other families such as the multi antimicrobial extrusion protein family (MATE), the small multidrug resistance family (SMR) and the resistance-nodulation-cell division superfamily (RND). Their role as antimicrobial and anticancer drug resistance mechanisms is now widely known (Boumendjel *et al.*, 2009), and they are suspected to also play a role in the pathogenicity of bacteria (Piddock, 2006). In eukaryotic cells, import of molecules play a role in resistance, since cells can modulate the amount of drug uptake (White and Marr, 1998) but there are no reports in bacteria.

Resistance to pyrrole amides - efflux pumps

The resistance mechanism displayed in congocidine producing *Streptomyces* (Stumpp *et al.*, 2005) and perhaps in strains producing distamycin is an ABC type efflux pump. This integral membrane protein forms the largest and oldest family of transporters (Jones and George, 2007) found in all domains of life. Despite the variability of the substrates transported, the ABC transporters are comprised of a core structure consisting of two transmembrane domains (TMD) and two cytosolic nucleotide binding domains (NBD) (Higgins *et al.*, 1986), which is also known as ABC, standing for ATP binding cassette (Hyde *et al.*, 1990). This arrangement can be seen in Fig.8A. These four domains are usually arranged in homo or hetero dimers, in which the NBD is C terminal to the TMD (Daus *et al.*, 2007), although the opposite arrangement has also been reported (Jones and George, 2007). The TMD presents a variable region that contrast with some consensus regions conserved in the NBD domain, such as the consensus Walker regions (Walker *et al.*, 1982) and the LSGGQ signature motif

(Annereau *et al.*, 1997) which are displayed in Fig.8B. The TMD spans across the membrane as seen in Fig.8C forming a channel that is thought to contain the substrate binding sites. The original study proposing the channel theory (Croop and Housmant, 1986) proposed the presence of six transmembrane helices per domain arranged in two arcs surrounding a central pore. Alternative models have been proposed such as the formation of a transmembrane β -barrel (Holland *et al.*, 2003) and two substrate binding sites (Jones and George, 2000). The unidirectional efflux of molecules seems to be motivated by a switch from inward to outward facing of the TMD domains. A conserved glutamine residue in the Walker B motif is responsible for a nucleophilic attack on ATP, which drives the NBD to close triggering the flip of the TMD domains (Hollenstein *et al.*, 2007) as shown in Fig.8A.

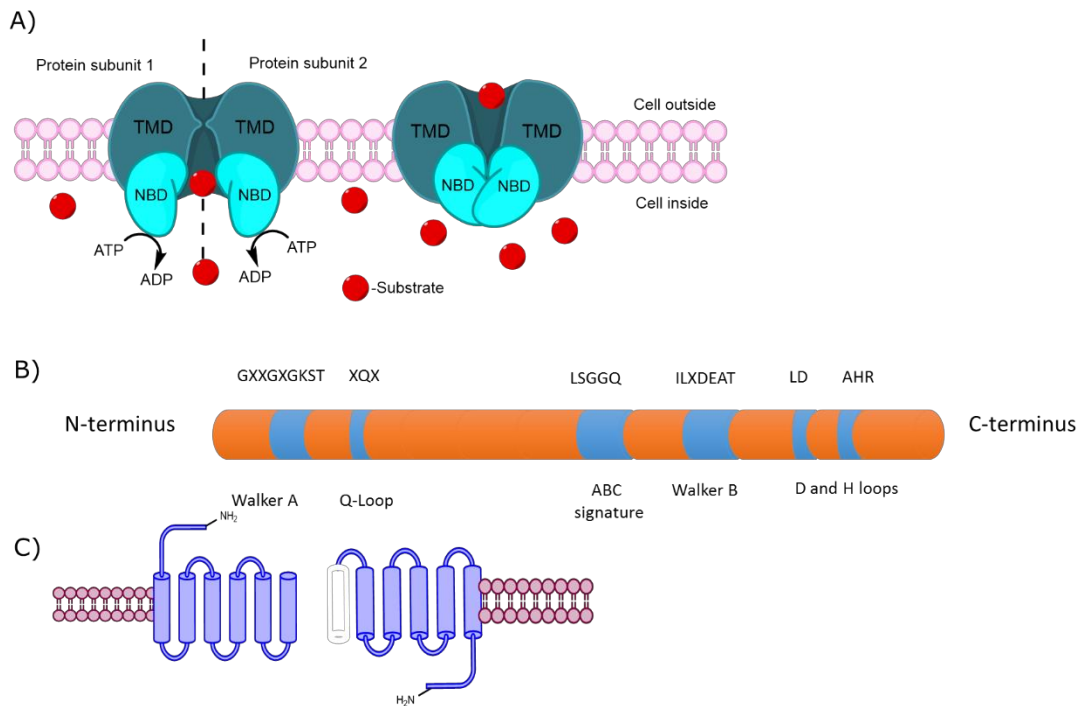


Figure 8. Structural features of ABC-type transporters

Figure 8A) presents a schematic view of the arrangement of the TMD and NBD dimers on the ABC transporter and how the binding of ATP produces the translocation of a molecule associated to the flipping in orientation of the TMD domain. Figure 8B) displays the conserved core domains on ABC transporters drawn in the orientation that the congocidine and distamycin transporter have. Figure 8C) illustrates the spanning transmembrane helices of the congocidine transporter.

Scope of this PhD thesis

This thesis describes a multiplex approach to drug discovery. Novel peptide-antibiotic development, based in the bioactive compounds congocidine and distamycin is pursued. This was done by the systematic study of combinatorial biosynthesis and mutasynthesis, exploitation of recent DNA assembly technologies, and the mechanism of resistance that could be developed by pathogenic bacteria to such novel drugs, by studying the resistance mechanism present in those biosynthetic clusters.

Specific Aims

The work carried out with a view to producing novel drugs has been conducted following four main aims that correspond with each of the main result chapters:

1. Produce a novel pyrrole amide antibiotic by combinatorial biosynthesis of two known pyrrole amide antibiotics, congocidine and distamycin.
2. Establish a DNA synthetic platform able to produce quick modifications within the congocidine cluster to help expedite aims one and three.
3. Introduce alternative heterocycles on congocidine by mutasynthesis, intending to produce peptides with different DNA binding affinity.
4. Study the resistance mechanism within the two clusters, their evolutionary relationship and implications for drug synthesis and heterologous host design.

Materials and Methods

Strains, plasmids and cultivation

The organisms used in this study are summarised in Table 1. Unless otherwise stated, *E. coli* strains were cultured at 37°C, *Streptomyces* strains were cultured at 30°C and yeast strains were cultured at 30°C. Strains carrying plasmids with selection markers were cultured with the required selection for the marker (antibiotic or auxotrophic) in order to maintain selection. Details of the particular requirements are stated for each plasmid/strain. A list of the plasmid and cosmids used in this study are provided in Table 2. Media recipes and selection procedures for markers are provided in Table 3 & 4.

Table 1. Strains used in this study

Strain name	Description	Genotype	Reference
DH5α	<i>E. coli</i> K12 derivative	F-, endA1, hsdR17 (rK- mK+), supE44, thi-1 λ-, recA1, gyrA96, relA1, deoR, Δ(lacZYA-araF)U169, Φ80dlacZ ΔM15	(Grant <i>et al.</i> , 1990a)
BW25113	<i>E. coli</i> K12 derivative	lacI+, rrnBT14, ΔlacZWI6, hsdR514, ΔaraBADAH33, ΔrhaBADLD78	(Datsenko and Wanner, 2000)
ET12567pUZ8002	<i>E. coli</i> K12 derivative	dam13::Tn9, dcm6, hsdM, hsdR, recF143, zij201::Tn10, galK2, galT22, ara14, lacY1, xylS, leuB6, thi-1, tonA31, rpsL136, hisG4, tsx78, mtli, glnV44, F-	(Macneil <i>et al.</i> , 1992)
TOP10	<i>E. coli</i> MC1061 derivative	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139	(Grant <i>et al.</i> 1990, Invitrogen)

		$\Delta(\text{ara-leu})7697$ galE15 galK16 rpsL(StrR) endA1 λ -	
BL21	<i>E. coli</i> K12 derivative	dcm, ompT, hsdS (rB- mB-), gal	
BY4741	S288C-derivative	MATa his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0	(Brachmann <i>et al.</i> , 1998)
BY4742	S288C-derivative	MAT α his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0	(Brachmann <i>et al.</i> , 1998)
BY4743	S288C-derivative	MATa/ α his3 Δ 1/his3 Δ 1 leu2 Δ 0/leu2 Δ 0 LYS2/lys2 Δ 0 met15 Δ 0/MET15 ura3 Δ 0/ura3 Δ 0	(Brachmann <i>et al.</i> , 1998)
M145	<i>S. coelicolor</i> A(3)2	scp1 Δ scp2 Δ	
M1146	<i>S. coelicolor</i>	act Δ red Δ cpk Δ cda Δ	(Gomez-Escribano and Bibb, 2011)
M1152	<i>S. coelicolor</i>	act Δ red Δ cpk Δ cda Δ rpoB[C1298T]	(Gomez-Escribano and Bibb, 2011)
M1156	<i>S. coelicolor</i>	act Δ red Δ cpk Δ cda Δ rpoB[C1298T] rpsL[A262G]	(Gomez-Escribano and Bibb, 2011)
TK23	<i>S. lividans</i>		(Kieser <i>et al.</i> , 2000)
DSM40846	<i>Streptomyces netropsis</i>		(Witt and Stackebrandt, 1990)
Aurodox strain	<i>S. K06-0806</i>		(Kimura <i>et al.</i> , 2011)

Table 2. Plasmid / Cosmid used in this study

Plasmid/Cosmid	Description	Resistance marker	Reference
pCS-BRC1	Derivative from pCR-Blunt [®] containing netropsin resistance protein subunit 1&2 from ATCC 23877	kanamycin,	This study
pCS-BRD1	Derivative from pCR-Blunt [®] containing distamycin resistance	kanamycin,	This study

	protein subunit 1&2 from DSM40846		
pCS-RC1	Derivative from pIJ6902 containing netropsin resistance protein subunit 1&2 from ATCC 23877	apramycin, thiostrepton,	This study
pCS-RD1	Derivative from pIJ6902 containing distamycin resistance protein subunit 1&2 from DSM40846	apramycin, thiostrepton,	This study
pCG002	Bac containing the entire congocidine cluster from ATCC 23877	chloramphenicol, hygromycin	(Juguet <i>et al.</i> , 2009)
pIJ790	λ -RED (gam, bet, exo), cat, araC, rep101 ^{ts}	chloramphenicol	(Gust <i>et al.</i> , 2003)
pIJ773	oriT, aac(3)-IV	apramycin	(Gust <i>et al.</i> , 2003)
pIJ6902	int ϕ C31, attP, oriT RK2, ori pUC18, tipA, tsr	apramycin	(Huang <i>et al.</i> , 2005)
pRS416	M13 ori, f1, ColE, URA3, T7, AmpR, lac, CEN6	ampicillin	(Sikorski and Hieter, 1989)
pCS-RS31	<i>E. coli</i> /yeast/ <i>Streptomyces</i> shuttle from pRS416/pIJ6902	apramycin, ampicillin, thiostrepton	This study
pCS-RS32	pCS-RS31 Δ tsr	apramycin, ampicillin	This study
pCS-RS31TL	pCS-RS31 with homology arms to congocidine cluster	apramycin, ampicillin, thiostrepton	This study
pCS-RS32TL	pCS-RS32 with homology arms to congocidine cluster	apramycin, ampicillin	This study
pCS-AO	part1.1+6.2 into pCS- RS31	apramycin, ampicillin	This study

Media preparation

Components for all media are listed in Table 3. All components were mixed in Duran bottles to 50% of the bottle total volume for solid media, and to 80% of the total volume for liquid media. This was done to avoid media boiling over during autoclaving. All media were sterilised by autoclaving at 121°C for 15 min.

Table 3- Media composition

Media	Composition per Litre of deionised water
Lysogenic broth (LB) (Bertani, 1951)	10 g tryptone 10 g NaCl 5 g yeast extract 20 g agar (added for agar medium) pH to 7.0
Mannitol soy flour (MS) (Hobbs <i>et al.</i>, 1989)	16 g soya bean flour 16 g d-mannitol 20 g agar (added for agar medium) Use of tap water pH to 7.0
Minimal medium (MM) (Hopwood, 1967)	0.5 g L-asparagine/ NH ₂ SO ₄ 0.5 g K ₂ HPO ₄ 0.2 g MgSO ₄ .7H ₂ O 0.01 g FeSO ₄ .7H ₂ O 10 g agar (added for agar medium) After autoclaving add 10 g glucose (or alternative carbon source)
Yeast extract malt extract (YEME) (Kieser <i>et al.</i>, 2000)	3 g yeast extract 5 g peptone 3 g malt extract 10 g glucose 340 g sucrose After autoclaving add 2 mL filter sterile 2.5 M MgCl ₂ *6H ₂ O
MP5 (Pernodet <i>et al.</i>, 1993)	7 g yeast extract 5 g NaCl 1 g NaNO ₃ 36 mL Glycerol 20.9 g 3-(N-morpholino) propanesulfonic acid pH to 7.5
GYM/ DSMZ65 (DSZM, n.d.)	4 g glucose 4 g yeast extract 10 g malt extract 2 g CaCO ₃ (not added to liquid medium) 10 g agar (for solid medium) pH to 7.2 before adding agar
2xYT (Kieser <i>et al.</i>, 2000)	16 g tryptone 5 g NaCl 10 g yeast extract
SOC (Hanahan, 1983)	20 g tryptone 5 g yeast extract 0.5 g NaCl 10 mL KCl (250 mM)

	pH to 7.0 with NaOH after autoclaving add 20 mL sterile glucose solution (1 M) just before use add 5 mL sterile MgCl ₂ (2 M)
SOB (Hanahan, 1983)	20 g tryptone 5 g yeast extract 0.5 g NaCl 10 mL KCl (250 mM) pH to 7.0 with NaOH just before use add 5 mL sterile MgCl ₂ (2 M)
YPAD (Robert H. Schiestl, n.d.)	10 g yeast extract 20 g peptone 20 g glucose (added filtered after autoclaving) 100 mg adenine hemisulphate Bacto-agar 20 g (added for agar medium)
SC synthetic media (SC drop out) (Robert H. Schiestl, n.d.)	6.60 g yeast nitrogen base (w/o amino acids) 20 g glucose (added filtered after autoclaving) 0.83 g Synthetic Complete Drop Out Mix 20 g Agar (add for solid medium) pH to 5.6 with 10 N NaOH.
Synthetic Complete Drop Out Mix *Used as Auxotroph marker	2.0 g adenine hemisulfate 2.0 g arginine HCl 2.0 g histidine HCl 2.0 g isoleucine 2.0 g leucine 2.0 g lysine HCl 2.0 g methionine 3.0 g phenylalanine 6.0 g homoserine 3.0 g tryptophan 2.0 g tyrosine 1.2 g uracil* 9.0 g valine

Maintenance of plasmid/cosmid selection in strains

To maintain selection of specific vectors in the strains or to select strains following transformation or conjugation, antibiotics were added to media at the concentrations indicated in Table 4. Yeast strains were selected using auxotrophic markers

Table 4- Selective markers used.

Substance	Class	Stock concentration (mg/mL)	Working concentration (µg/mL)	Stock solution solvent
kanamycin	Aminoglycoside antibiotic	50	50	H ₂ O
apramycin	Aminoglycoside antibiotic	50	50	H ₂ O
ampicillin	β-lactam antibiotic	50	50	H ₂ O
carbenicillin	β-lactam antibiotic	100	100	H ₂ O
chloramphenicol	N-dichloroacyl phenylpropanoid antibiotic	25	25	Ethanol 96%
nalixidic acid	Gyrase inhibitor antibiotic	50	50	H ₂ O
congoicidine	Pyrrole amide antibiotic	100	100	Dimethyl sulfoxide 100%
hygromycin B	Substituted aminoglycoside antibiotic	50	50	H ₂ O
uracil	Nucleotide	Prepared fresh	1.2	H ₂ O

Cultivation of *Streptomyces* strains

Maintenance and propagation of *Streptomyces* was carried out on Mannitol Soy (MS) agar (Table 3), with the appropriate marker where necessary to select for specific exconjugants or transformants (Table 4). Phenotypic screening was carried out in a range of solid and liquid media (see Table 3).

When cultured in liquid media, the air to liquid volume ratio was at least 5:1 in culture vessels, usually in conical Erlenmeyer flasks containing steel springs to avoid mycelial pelleting. The inoculum in this case was generated from an overnight culture grown from spores in the same media and inoculated to a final OD₆₀₀ of 0.05-0.08 for survival assays or OD₆₀₀ of 0.1 for routine growth. Overnight cultures (16 h) were grown either in Erlenmeyer flasks containing steel springs or in 50 mL Falcon tubes with sterile glass beads (4 mm diameter) with the same liquid/volume ratio.

All cultures were incubated at 30°C. Liquid cultures were grown with agitation at 250 rpm.

Liquid culture growth curves – Dry weight

Growth curves were made by harvesting 50 mL culture triplicates of *Streptomyces* grown in 250 mL conical flasks. At each time point, the liquid culture was filtered through a Whatman glass microfiber filter (diameter = 0.22 μM) (weight previously determinate after humidity removal) in a reduced pressure Büchner funnel apparatus. Following filtration the filter disks were recovered from the funnel and dried in a microwave until constant weight (800 W for 5 min, 1 minute pause, and 800 W for 5 min). Final mass of the filters was determined gravimetrically and biomass calculated by subtraction from the original filter weight.

Liquid culture growth curves – OD

Streptomyces cultures were grown in multi-well plates placed in a Synergy HTX (BioTek) temperature controlled-shaking incubator provided with OD measurement. To avoid excessive absorbance due to the pigments produced by particular *Streptomyces* strains, OD was set to 600 and 450 nm, and recorded every 30 min for a total elapse of 48 h. The data was then curated to cover the linear phase of growth and calculate the specific growth.

Preparation of *Streptomyces* spore stocks

Adapted from (Kieser *et al.*, 2000).

The desired strain was streaked evenly on an MS plate and incubated at 30°C until confluent growth and spore formation was visible, determined by the grey powdery surface of cultures. To harvest spores, aliquots (5 mL) of sterile 20% aqueous glycerol was added to the surface of the plate and the spores were dislodged from the mycelium with the help of a sterile cotton bud. The resulting spore suspension was then transferred to bijoux bottles and stored at -20°C for up to 4 years without viability loss.

Cultivation of *E. coli* strains

Single colonies from an agar plate were inoculated into 5-10 mL of LB liquid media in a 20 mL universal bottle, with addition of the appropriate selection. Cultures were then incubated at 37°C (or 30°C for strains containing temperature sensitive *ori*) in a shaking incubator at 220 rpm overnight.

When cultivating larger volumes of *E. coli*, the inoculum from an overnight culture was prepared as above and transferred to the larger volume (normally Erlenmeyer flasks) in a 1:100 ratio unless otherwise stated.

Storage of *E. coli* stocks

For long term storage of *E. coli* strains overnight cultures prepared as described above were mixed in a 1:1 ratio with a solution of 50% sterile glycerol in water, (final glycerol concentration of 25%). These glycerol stocks were stored frozen at -80°C.

Culture supernatant sample preparation for HPLC/LCMS

Submerged mycelial cultures were centrifuged at 4000 x *g* to separate biomass from supernatant. From the resulting liquid, 2 mL were then filtered with a 0.22µm filter (Millipore) to ensure no insoluble material was carried on to the chromatography column. 1 mL of these samples were then diluted with 1mL of a 1:1 mixture of the eluent buffers (Water /0.1% Formic acid: Acetonitrile/0.1% Formic acid) to ensure no component would precipitate in the column. If precipitation was observed further 0.22 µm filtration was applied before injecting the sample.

Congocidine and distamycin standards

The congocidine used in the assay was purchased from Enzo lifesciences ≥90% (HPLC). As distamycin could not be purchased from vendors, and its chemical synthesis was not achieved, it was isolated from *Streptomyces netropsis* DSM40846. Spores were inoculated in YEME medium for 48 h in a 250 mL Erlenmeyer flask containing a spring. The preculture was inoculated into the final YEME culture at a 4% (v/v) rate. The cultures were grown for four days. The resulting culture was centrifuged at 4200 x *g* for 10 min at room temperature and the supernatant discarded. The resulting mycelial pellet was resuspended in 50% of the original culture volume in methanol. Cells were subsequently disrupted in a Branson Sonicator 450 (with 1/8 microtip, settings: intensity=7, continuous cycle) for 1.5 min in batches of 30 mL. All batches were pooled and centrifuged for 1 min at 4200x*g* to obtain a cell pellet and a methanol supernatant containing a mixture of metabolites containing distamycin, some aureothin and traces of congocidine and disgocidine. To remove aureothin, a liquid extraction with chloroform was performed.

As standards for distamycin, disgocidine and aureothin were not available, an estimate purity of 93% was estimated using area under the peak on its UV-LCMS chromatogram. To obtain pure distamycin the methanol was removed by rotary evaporation and residue resuspended in DMF. The resulting solution was then injected into a semi preparative HPLC column (Luna 5u C18(2) 100A, 50 x 21.20 mm 5µm) and distamycin was separated from the other analytes using AcN/H₂O in 0.1% trifluoro acetic acid (TFA) with the procedure represented in the Table 5. The elution was monitored at 298 nm and fractions collected freeze dried and resuspended. Purity was confirmed by H-NMR (see annex).

Table 5. Semipreparative HPLC purification of Distamycin

Time (min)	Flow rate (mL/min)	% Water (with 0.1% TFA)	% MeCN (with 0.1% TFA)
0	6	90	10
25	6	50	50
35	6	50	50
40	6	30	70
45	6	90	10

Liquid-liquid extraction

Supernatant from a *Streptomyces* culture was mixed by vigorous shaking in a separation funnel in a ratio 2:1 with a water immiscible organic solvent which selection depended upon application (chloroform, dichloromethane, ethyl acetate). Filtration and/or addition of NaCl was used when the formation of emulsions avoided the separation of both phases. After phase differentiation, the organic layer was collected and set aside. The exhausted phase was extracted twice again. Subsequently, the collected phases were pooled and evaporated in a rotary evaporator to a small volume, transferred to a small vial and evaporated the remaining solvent with an inert gas. If the layer had a high melting point, instead of evaporation in a normal rotary evaporator, it was evaporated under high vacuum rotary evaporator, or freeze dried.

Biomass extraction

Biomass extraction was performed to extract compounds that were not present in the supernatant but presumed to accumulate in the cells. Biomass of the desired culture was pelleted in a centrifuge to separate it from the culture media. The pellet then was resuspended in methanol (70%) and lysated with a Branson Sonifier 450 (with 1/8 microtip, settings: intensity=7, continuous cycle) for 1.5 min in batches of 30 mL. The resulting mixture was centrifuged to give a washed cell pellet and a supernatant that was freeze dried.

Reconstitution of dried samples for HPLC/LCMS

Freeze dried samples were resuspended in a mixture 2:1 of the (Water /0.1% Formic acid: Acetonitrile/0.1% Formic acid) eluting buffers used in the analytical system, then filtrated with a 0.22 µm filter.

Thin Layer Chromatography (TLC)

Organic synthesis was monitored by the use of TLC plates (Macherey-Nagel, 0.25mm Kieselgel 60/ Merck Kieselgel 60 F254 aluminium sheets). Intermediate reaction crudes were spotted on silica plates and eluted with methanol, DCM, ethyl acetate, hexane or a mixture of them according to expected polarity of compound. All the analytes were visualised with the internal UV reactive dye after UV illumination.

Antimicrobial susceptibility assay

Broth Microdilution minimum inhibition concentration (MIC) Panel-Adapted from (American Society for Microbiology, 2005)

Broth microdilution MIC testing is performed in 96 well plates containing Mueller-Hinton broth at pH 7.2-7.4.

An overnight culture of the strain subject to the test was grown in LB media, then diluted 1:100 in fresh LB media and incubated at 35°C until the resulting suspension reaches $1-2 \times 10^8$ CFU/mL. The resulting dilution is further diluted with Mueller-Hinton broth to 5×10^5 CFU/mL. The wells in the plate are filled with 200 µL of this suspension. The drugs to be tested are added to each well in triplicates, and the positive control is the suspension containing the carrier solvent used to dissolve the drug tested. Growth is monitored by OD at 650nm during incubation at 35°C for 16-24 h.

Minor groove binding susceptibility assay

Prior to screening the spore stocks, each strain was inoculated into YEME and grown for 16 h or when a reading of OD₆₀₀=0.4 was achieved. The inoculation ratio was 1:100 for a total volume of 20 mL of medium on a 50 mL Falcon tube which was incubated at 30°C and 220 rpm.

These cultures were used to make the inoculum for the assay. The inoculation ratio was variable according to the particular OD₆₀₀ in such an amount that 20 mL of YEME at OD₆₀₀ = [0.04-0.08] (the variability is due to pelleting on the strain, which makes difficult OD reads) was achieved for each strain. This 20 mL were then used to fill 198 µL of the 360 µL total volume of a 96 well plate (TPP). In order to establish a working concentration to assay the different compounds, a preliminary study was carried out with congocidine and distamycin.

The assays were performed over a 48 h period and OD₆₀₀ monitored at intervals of 15 min. The data sets were then curated to include only the exponential phase of growth of the strain, time in which bacterial growth resembles a first order chemical reaction:

$$\frac{\partial X}{\partial t} = kX,$$

X the bacterial population, k specific growth rate and t the time

Regrouping terms the equation can be seen as:

$$\frac{\partial X}{X} = k \partial t$$

Which can be integrated:

$$\int_0^X \frac{\partial X}{X} = k \int_0^t \partial t$$

To result:

$$\ln\left(\frac{X}{0}\right) = kt + C$$

$$\ln X = kt + C$$

Equation that describes a straight line:

$$y = ax + b$$

So then, the OD values on the exponential phase were taken natural logarithms and plotted against the time, allowing a linear regression and specific growth rate calculation by least squares method.

Some data points show negative specific growth rate which makes no physical sense, however this happens because a diminishing trend on OD values in a narrow absolute value. Those are displayed as 0 values but taken in account to calculate the error of the measurement. For clarity of interpretation, the specific growth rates were then normalised to a coefficient of survival defined as:

$$\text{Survival rate} = \frac{\text{Compound A treated Specific growth rate}}{\text{Carrier solvent treated Specific growth rate}} \cdot 100$$

Preparation of *E. coli* chemically competent cells (for transformation of DNA constructs up to 11kB)

Adapted from (Sambrook and Russell, 2000)

An overnight culture of the *E.coli* strain was cultured as described above. 150 μL of this culture was used to inoculate 50 mL of fresh LB media in a 250 mL erlenmeyer flask and then incubated at 37°C (or 30°C for strains containing temperature sensitive *ori*) in a shaking incubator at 220 rpm until the $\text{OD}_{600} = 0.4-0.5$. The culture was then centrifuged at 4200 $\times g$ at 4°C for 10 min. The supernatant was discarded and cells washed in 12.5 mL of ice cold 100 mM CaCl_2 . Cells were recovered by centrifugation for 10 min at 4200 $\times g$ and 4°C. Supernatant was discarded, and the step repeated once. Supernatant was discarded and cells were centrifuged again for 10 min at 4200 $\times g$ and 4°C. The resulting cell pellet was resuspended in 1 mL of ice cold 100 mM CaCl_2 in 20% glycerol. The resulting solution was then aliquoted in 50 μL volumes in precooled Eppendorf tubes and left on ice for 30 min. Aliquots were used right away or snap frozen with liquid nitrogen for storage.

Preparation of *E. coli* electro-competent cells (for transformation of DNA constructs ≥ 11 kB)

Adapted from (Untergasser, 2008).

LB media (Without NaCl) was used to prepare an overnight culture of the required *E.coli* strain with the appropriate antibiotic selection. This overnight culture was used to inoculate 50 mL of LB media (without NaCl) in a 250 mL Erlenmeyer flask and incubated at 37°C and 200 rpm until OD₆₀₀ reached 0.4-0.5. As this step limits the quality of the cells, OD was measured every 30 min, and if the cells were discarded if the OD was over 0.5. Cultures were transferred to an ice bath for 15 min and harvested by centrifugation at 3500 x g at 4°C for 20 min. The supernatant was discarded and the cell pellet was washed in 50 mL of ice cold deionised water of which 1.5 mL was added first to resuspend the pellet, and the rest added subsequently. Cells were recovered by pelleting at 4200 x g at 4°C for 20 min. The supernatant was discarded and pellet resuspended in 50 mL of ice cold deionised water. This step was repeated twice and finally the cells were re-suspended in 0.5 mL 10% glycerol (v/v in deionised water), aliquoted in to 50 µL aliquots and used immediately or snap frozen in liquid nitrogen and stored at -80C, for six months.

Transformation of *E. coli* chemically competent cells

Adapted from (Sambrook *et al.*, 1989)

An aliquot of cells prepared as indicated above was thawed on ice and mixed carefully with a pipette tip (rather than pipetting up and down) with up to 5 µL of the DNA to be transformed (15-1000 ng) and incubated on ice for 30 min. The cell/DNA mixture was heat shocked at 42°C for 50 s followed by two min incubation on ice. Immediately afterwards, 1 mL of SOC medium (Super Optimal broth with Catabolite repression) was added and cells allowed to recover in a shaking incubator set at 220 rpm and 37 °C (or 30 °C for strains containing a temperature sensitive *ori*) for one hour. Following recovery, cells were centrifuged for 5 s at 6000 x g and resuspended in 300 µL of SOC before plating on LB agar with the appropriate selection marker(s).

Transformation of *E. coli* electro-competent cells

Adapted from (Untergasser, 2008).

Electro competent *E. coli* cells were thawed on ice, and the DNA used for the transformation was dissolved in desalted water. A maximum of 2 µg of DNA in 2 µL of water was mixed with 50 µL cells. Cells and DNA were mixed carefully with circular moves of a pipette tip (rather than pipetting up and down) for one min while on ice. The mixture was then transferred to a 1 mm electro cuvette and stored at -80°C (amounts can be doubled if using 2 mm, ie 100 µL cells + 4 µL-4 µg DNA) and immediately electroporated in a Biorad Gene Pulser set up at 1.25 kV and 5 ms. Following electroporation, cells were recovered as quickly as possible with 1 mL of SOC medium. After the medium was added to the cuvette, it was recovered and placed in a 20 mL flask and cells were allowed to recover at 37 °C (or 30 °C for strains containing temperature sensitive *ori*) in a shaking incubator (220 rpm) for a period of 1-2 h depending on the size of the construct and *E. coli* strain. Cells were then pelleted by centrifugation for 5 seconds at 6000 x g and resuspended in 300 µL of SOC before plating them in a LB agar plate containing the appropriate selection markers. After overnight incubation colonies were screened for the appropriate construct.

Isolation of plasmid and cosmid DNA from *E. coli* (Alkaline lysis method)

Adapted from (Bimboim and Doly, 1979).

E. coli strains containing the desired plasmid/cosmid were grown overnight with appropriate selection as detailed above. To harvest cells, the culture was centrifuged in a bench top centrifuge at 16000 x g for 5 min. The cell pellet was resuspended in 100 µL of ice cold solution 1 (50 mM glucose, 25mM Tris.Cl (pH 8.0), 10 mM EDTA (pH 8.0)) and transferred to a 1.5 mL microcentrifuge tube. Following resuspension, 200 µL of solution 2 (0.2 M NaOH, 1% SDS) was added and the resulting solution was mixed by inverting the tube 6-8 times. After 5 min incubation on ice, 150 µL of ice cold solution 3 (60 mL 5 M potassium acetate, 11.5 mL glacial acetic acid, 28.5 mL deionised water) was added and the tube vortexed. The tube was incubated on ice for 5 min before centrifuging for 10 min at 13,000 rpm in a microcentrifuge. The supernatant was transferred to a fresh 1.5 mL microcentrifuge tube, mixed by inversion with 2 x volumes of ethanol or isopropanol and incubated for 5 min at room temperature. Subsequently, the tube was centrifuged (with the hinge pointing

outwards) for 10 min at 13,000 rpm in a microcentrifuge. The supernatant carefully removed and the pellet of DNA rinsed with 1 mL 70 % ethanol. Excess ethanol was removed under vacuum at 37°C in a speedy vac (DNA concentrator, no brand listed). The resulting DNA pellet was then resuspended in deionised water (20-100 µL).

Isolation of plasmid DNA from *E. coli* using commercial kits

Performed according to manufacturer's (BIOLINE) protocol.

Intergeneric conjugation of DNA from *E.coli* to *Streptomyces*

Adapted from (Kieser *et al.*, 2000)

To avoid the methyl specific restriction system in *Streptomyces*, the non-methylating *E.coli* strain ET12567/pUZ8008 was transformed with the required plasmid to be introduced in to the desired *Streptomyces* strain.

An overnight culture of the ET12567/pUZ8008 transformant was grown as described above and used as inoculum for a 50 mL culture in a 1-3:100 ratio and incubated at 37°C at 220 rpm in a shaking incubator until the OD₆₀₀ reached 0.4-0.6. Cells were harvested by centrifugation at 4200 x g for 10 min, washed twice with LB medium to remove antibiotics and resuspended in a final volume of 500 µL LB medium. In parallel a suspension of *Streptomyces* spores was prepared to a concentration of 2 x 10⁸ spores/mL in 2YT medium and heat shocked at 50°C for 10 min. The spore suspension was mixed with the *E.coli* cells (500 µL + 500 µL) and plated on MS medium. Plates were incubated at 30°C for 16-18 h, prior to plates being overlaid with a 1 mL of a solution containing 500 µg of nalixidic acid (to remove *E.coli* cells) and 1 mg of the appropriate selection marker for the incoming plasmid/cosmid. The overlay was left to dry in a laminar flow hood and further incubated until primary exconjugant *Streptomyces* colonies were observed. These colonies were subsequently patched onto MS agar or NA to screen for double crossovers.

Isolation of genomic DNA from *Streptomyces*

Adapted from (Kieser *et al.*, 2000).

A 250 mL Erlenmeyer flask containing a spring to reduce pelleting, with 50 mL of YEME medium was inoculated with approximately 1 x 10⁶ *Streptomyces* spores and incubated at 30°C in a shaking incubator (220 rpm) for 48 h. The culture was centrifuged for 10 mins at 16000 x g and the biomass resuspended in 500 µL of Solution 1 (50mM glucose, 25mM Tris.Cl

(pH 8.0), 10mM EDTA (pH 8.0)) in a 1.5 mL microcentrifuge cap tube. To this 10 μ L of 30 mg/mL lysozyme and 5mg/mL RNase was added and the cells incubated at 37°C for 60 min. Subsequently 10 μ L of a 10% solution of SDS was added and the tube mixed by inversion. Thereafter, 500 μ L of phenol-chloroform-isoamyl alcohol (25:24:1) saturated with 10 mM Tris, pH 8.0, 1 mM EDTA, was added and the mixture was vortexed for 5 min, before centrifugation at 16000 \times g for 10 min. The aqueous (upper) layer was transferred to a fresh tube leaving behind as much as possible as the protein layer formed in the interphase. The transferred phase was reextracted with phenol-chloroform-isoamyl alcohol. The new purified layer was washed twice with chloroform to remove phenol dissolved in the water. The resulting supernatant was transferred into a fresh 1.5 mL tube and mixed by inversion with 2 x volumes of isopropanol to precipitate the DNA. Following 5 min incubation at room temperature, the tube was centrifuged 13,000 rpm in a microcentrifuge and the supernatant carefully removed. The resulting DNA pellet was rinsed with 1 mL 70 % ethanol. Excess ethanol was removed under vacuum at 37°C in a speedy vac (DNA concentrator, no brand listed). The resulting DNA pellet was then resuspended in deionised water (20-100 μ L) and stored at -20°C.

Agarose gel electrophoresis

Agarose gels were made at different percentages according to the size of DNA that was separated on them; 0.5% to 2.5 % w/v of agarose in 1 x TAE buffer (40 mM Tris-acetate, 1 mM EDTA) was dissolved using a microwave. Following cooling of the mixture to approximately 60°C, ethidium bromide (final concentration of 100 μ g/mL) was added and the gel poured in to an appropriate casting tray.

Gels were run at a constant current of 100 volts for around 1 hr in Biorad electrophoresis tanks filled with 1 x TAE buffer. DNA samples were mixed with 6X loading buffer (60 mL of 100% glycerol, 6 mL of 1M Tris HCl, pH 8.0, 1.2 mL of 0.5 M EDTA, pH 8.0, 32.8 mL of molecular biology grade water, 60 mg of Bromophenol Blue and/or 100 mg Orange G). Gels were visualised on a Syngene Gel Doc system.

Agarose gel DNA extraction and purification

When DNA purification was required, the gel was visualised in a low energy UV trans-illuminator at the lowest setting, the desired bands were excised from the gel with a freshly opened disposable scalpel and processed according to the extraction kit manufacturer instructions (Bioline).

Cultivation of yeast strains

Adapted from personal communication with Dr. Megan Lenardon (University of Aberdeen). Approximately 2mm diameter single yeast colonies were selected and grown in YPAD broth at 30°C, with shaking at 220 rpm. Cultures were maintained on plates for up to 2 weeks and re-streaked as required.

Storage of yeast stocks

Adapted from personal communication with Dr. Megan Lenardon (University of Aberdeen). Strains were grown at 30°C with shaking at 220 rpm until a stationary phase culture was obtained. An aliquot of 500 µL was mixed with 250 µL of YPAD medium and 250 µL of a 50% glycerol solution. This was followed by snap freezing in liquid nitrogen and storage at -80°C. To recover the yeast, strains were streaked on to fresh YPAD agar and given 6 days at 30 °C to recover.

Lithium Acetate/SS DNA/PEG transformation of yeast

Adapted from (Gietz and Schiestl, 2007).

YPAD broth or aminoacid complemented synthetic media SC was inoculated with a single yeast colony according to strain/plasmid requirement and grown for 18-26 h at 30°C on a shaking incubator at 220 rpm.

A 250 mL Erlenmeyer flask containing 50 mL of YPAD broth was inoculated to a final cell density of 5×10^6 cells/mL with the aid of a haemocytometer. The culture was grown until a cell density of 1×10^8 cells/mL was reached and the yeast cells were harvested by centrifugation at $3000 \times g$ for 3 min. Cells were washed once with 50 mL of sterile ice cold deionised water, and then resuspended in 10 mL of 100 mM lithium acetate (LiAc). The suspension was aliquoted into sterile microcentrifuge tubes to harvest the cells at $11000 \times g$ in a microcentrifuge. The supernatant was discarded and cells resuspended into 200 µL of 100 mM LiAc. Cells were centrifuged again at $11,000 \times g$, the supernatant discarded and the pellet resuspended in freshly prepared transformation buffer (per transformation 36 µL 1 M LiAc, 240 µL 50% PEG 3500 (w/v), 50 µL ssDNA 2 mg/mL and 29 µL ddH₂O) and 5 µL of the DNA to transform with. The mixture was incubated at 42°C for 30 min with occasional agitation, before rapidly cooling by the addition of 1 mL of sterile deionised water. Cells were harvested and washed twice with 1 mL of deionised water and plated into amino acid

complemented synthetic media SC complemented agar with the appropriate aminoacids and marker.

Electroporation of yeast (for up to 20 µg of DNA)

Adapted from (Benatuil *et al.*, 2010).

All the yeast electroporation in this study was carried out with freshly made competent cells. A single yeast colony was inoculated into YPAD or complemented SC according to strain/plasmid requirement and grown for 18-26 h at 30°C on a shaking incubator set to 220 rpm. With the aid of a haemocytometer, 50 mL of the same medium was inoculated to a cell density of 5×10^6 cells/mL. When this culture cell density reached 1×10^8 cells/mL, the yeast was harvested by centrifugation at $3000 \times g$ for 3 min. Cells were washed twice with 50 mL of ice cold deionised water and once with 50 mL of ice cold electroporation buffer (1 M sorbitol / 1 mM CaCl_2 in water). Another centrifugation was carried out and cells resuspended in 20 mL of 0.1M LiAc/10 mM DTT. The resulting suspension was incubated at 30°C on a shaking incubator set to 220 rpm for 30 min. After this, cells were harvested by centrifugation at $3000 \times g$ for 3 min, washed once with 50 mL of electroporation buffer and the pellet finally resuspended in ice cold electroporation buffer to a final volume of 1 mL, sufficient for two 400 µL reactions. Cells were kept on ice before electroporation.

Electroporation of yeast cells was carried out in pre chilled Bio-Rad 2 mm electro cuvette. Yeast cells and DNA were mixed by stirring with a pipette tip rather than pipetting up and down and stored on ice for 5 min. Electroporation was carried out with a 2.5 kV pulse , 25 µF and 3 ms in a Bio Rad MicroPulser electroporator. After electroporation, cells were resuspended in 8 mL of a 1:1 mixture of YPAD and 1 M sorbitol. This mixture was incubated for 1 hour at 30°C on a shaking incubator at 220 rpm. The cells were then washed with SC complemented media ΔURA or appropriate marker and plated on to the same agar media and incubated at 30°C.

Use of a haemocytometer

10 µL of the cell suspension to be measured was carefully placed between the cover slip and the base of the haemocytometer. Cells were left to settle onto the haemocytometer grid for a few mins. The grid area is typically 1 mm^2 , divided into 25 equal-sized squares, and the volume measured is 10^{-4} mL .Cell count per millilitre was calculated by using the following equation:

$$\frac{\sum_{Cell 6}^{Cell 1} \text{Cells per cell}}{6} \times \text{dilution factor} \times 10^6 \frac{\text{cells}}{\text{mL}} =$$

The efficiency of the transformation was calculated as follows:

$$\text{Transformation efficiency} = \frac{\frac{\text{Cells}}{\text{mL}}_{\text{experiment}} - \frac{\text{Cells}}{\text{mL}}_{\text{negative control}}}{\frac{\text{Cells}}{\text{mL}}_{\text{URA plates}}}$$

Yeast recombineering

Adapted from personal communication with Dr. Paul Rowley and (Zhao and Shao, 2009). Primers were designed to produce overlapping DNA fragments by PCR that would assemble large DNA fragments. Templates with no homology were attempted to be “stitched” together with 80 bp long oligos (40 bp complementary to each non homologous DNA segment). The incorporation of vectors was pursued either by PCR amplification or digestion with two different restriction enzymes and dephosphorylation –recombineered vectors contained both a bacterial and yeast origin of replication. All DNA fragments were gel purified prior to recombineering. Yeast competent cells were then transformed with 200-1000 ng DNA of each of the overlapping fragments. Colonies resulting from the transformation were harvested from the plate with the help of a spreader and 1 mL of deionised water. Depending on cell density, the recombineered plasmid could be extracted directly using the Zymolase protocol (see below) or by growing the yeast strain overnight in broth culture. After obtaining the recombineered plasmid (usually in low yield), *E.coli* was transformed in order to enrich it and verify the construct.

Yeast Zymolase plasmid DNA extraction protocol

Adapted from personal communication with Dr. Paul Rowley.(Gietz and Schiestl, 2007). Approximately 1.5×10^9 cells were harvested from a suitable transformation plate. Cells were washed with 1 mL of sterile deionised water and centrifuged at $13000 \times g$ in a microcentrifuge for 30 s. The pellet was resuspended in 200 μL spheroblasting buffer (1 M sorbitol, 0.1 M EDTA, 5 mM β -mercaptoethanol [added fresh 0.3 μL in 1mL]) and 5 μL of 10 mg/mL Zymolase.

The suspension was incubated at 37°C for 1 hour. The resulting spheroblasts were then centrifuged at 13,000 x g in a microcentrifuge for 10 s. At this point spheroblasts were treated in the same manner as a bacterial pellet to extract the desired plasmid DNA.

Primers design for PCR and cloning techniques

All primers were designed using the online tool primer Blast (Ye *et al.*, 2012), which uses the latest algorithm of Primer3 (Koressaar and Remm, 2007) to design the primers within user defined constraints and then searches the primers using basic local alignment search tool, BLAST (Altschup *et al.*, 1990) against an user defined database to avoid selection of primer pairs that could result in unspecific amplification or self-priming. SnapGene (GSL Biotech) was used to visualize binding sites and modification of the DNA templates. The oligonucleotides used in this study are listed in Table 6:

Table 6. Oligonucleotides used in this study.

Primer Name	Sequence (5'-3')	Description
RD1fusion	ATGCGCCACCTTCTGG	Overlap of congocidine and distamycin efflux genes
C1fusion	TCAGTGGCTCTCCTCCCTG	Overlap of congocidine and distamycin pump
D1	TCAGGGGGATCCCTCCT	Overlap of congocidine and distamycin pump
RC1FUSION	ATGGCGCACTTGCTGG	Overlap of congocidine and distamycin pump
DC1LINKER	CCAGAAGGTGCGCCATTCAGTGGCTC TCCTCCCTG	Overlap of congocidine and distamycin pump
CD2LINKER	CCAGCAAGTGCAGCCATTCACCCCATC CCTCCT	Overlap of congocidine and distamycin pump
DPRV2	GTCATGTACCGCCCTTGACA	distamycin pump cloning
DPFV2	GAAGGAATGCAGCAGGGGAA	distamycin pump cloning
FORWARD TO CGC3*	ATGGCCACCGAGTCCGCC	Cloning of SAMR0921 for protein expression
CHAMPION KIT F	CACCATGGCCACCGAGTCCGCC	Directional cloning of SAMR0921 for protein expression
LESSCGREVERSECGC3*	ACAAGCGGTACGGCGCTC	Cloning of SAMR0921 for protein expression
REVERSECGC3*	GGCGCTCTACCCGCCG	Cloning of SAMR0921 for protein expression
DISGOPUMPF	CCGATCTTCGGCAGC	Distamycin pump cloning
DISGOPUMPR	ATCATGAGTGAGGAGAGCCG	Distamycin pump cloning

REVERSEDPPROMOTER	CTCACTTCTTTCGCTGCAAC	Distamycin pump cloning
CGC10OUT-F	ACGCGCAGGACATCGCGCCCCGCGGA ACCGGGGCGTGAGCATTCCGGGGATC CGTCGACC	λ-Red knock out of SAMR0909 from pCG002(Juguet <i>et al.</i> , 2009)
CGC10OUT-R	AGTCGGCGTCCCTCGCCTCCGGCCAGT ACCACTGCTCTCATGTAGGCTGGAGCT GCTTC	λ-Red knock out of SAMR0909 from pCG002 (Juguet <i>et al.</i> , 2009)
CGC10CHECKF	CAGCTGCGTGAGGATG	PCR check of knock out of SAMR0909
CGC10CHECKR	GGTTGCGGGAAGTAC	PCR check of knock out of SAMR0909
RSTF	CGAATCGAGTATGTCCGAGATCCACG ATCACGATCGTCCCTGACGTCCCGCCTT GCGGAGTTGCCGTCATGGGGTCCGTC CGGATCGGTCTTGCCTTGCT	In vitro synthesis of congocidine cluster
RSTR	CTGGAACGGGACTTCGAGTACGCCGT CCTGTCCATGAGCGGCGAGGCCACG TCGACGGCGTGCCCTGGTACCGGGC TCATACGTGCGGGTGAAGTTCAG	In vitro synthesis of congocidine cluster
Pij6902f	GGAGAGGCGGTTTTCGATTTATGTC CGCTCCCTTCTCTG	In vitro synthesis of congocidine cluster- Delivery vector
Pij6902r	AGGACGATCGTGATCGTGGATCTCCG ACATACTCGATTTCGATTCGGTGCCTG AGGCTTG	In vitro synthesis of congocidine cluster- Delivery vector
pRS416F	CGCTCATGGACAGGACGGCGTACTCG AAGTCCCCTTCCAGCAGGGGATAACG CAGGAAAG	In vitro synthesis of congocidine cluster- yeast vector
pRS416R	CAGAGAAGGGAGCGGACATAAATAC GCAAACCGCCTCTCC	In vitro synthesis of congocidine cluster- yeast vector
Start1	CGAATCGAGTATGTCCGAGATCCACG ATCACGATCGTCCCTTACTAACGTCTG GAAAGACGACAAAACCTTAGATCTGG GG	In vitro synthesis of congocidine cluster
Stvect1	TATGTCCGCTCCCTTCTCTGAAGCCGT CCACGCTGCCTCCCTAGAGCGGCCGCC ACCGCGGTGGAGCTCCAGCTTTTGTT	In vitro synthesis of congocidine cluster
stend1	GCCGCTCATGGACAGGACGGCGTACT CGAAGTCCCCTTCCACATTGGGTACCG GGCCCCCCTCGAGGTGACGGTATC GATA	In vitro synthesis of congocidine cluster
phiC31F	TTCGAACGCATCCTGAACGA	In vitro synthesis of congocidine cluster- Delivery vector

phiC31R	TCTTGATCTCACGCCACCAC	In vitro synthesis of congocidine cluster- Delivery vector
PartACGCF	TTGGTGAACAGCCCGATCAT	In vitro synthesis of congocidine cluster
PartACGCR	TGTCCTCTTCCGTGAACTGC	In vitro synthesis of congocidine cluster
PartBCGCF	GGAGCAGATGGGCGTAGAAG	In vitro synthesis of congocidine cluster
PartBCGCR	CGTTCCACATCGAGTCCGTC	In vitro synthesis of congocidine cluster
PartCCGCF	TCGTCTTCGTCTGTCTGTTT	In vitro synthesis of congocidine cluster
PartCCGCR	ACGACCTCTTTGATCCGCTG	In vitro synthesis of congocidine cluster
URA F	GCCCAGGTATTGTTAGCGGT	In vitro synthesis of congocidine cluster
URA R	TTCCAGCCTGCTTTTCTGT	In vitro synthesis of congocidine cluster
R12Fcongo	TTTTCACCTACTCGCCGCAT	In vitro synthesis of congocidine cluster
R1Fcongo	TTCACCTACTCGCCGCATTT	In vitro synthesis of congocidine cluster
R1RCONGO	CCGAGGTAGTTCGTGAACAT	In vitro synthesis of congocidine cluster
FRAGMENT1FOR	TGGAGCTCCACCGCGGTGGCCGAATC GAGTATGTCGGAG	In vitro synthesis of congocidine cluster
FRAGMENT2REV	TGCATGCCTGCAGGTCGACTGCGTAG CCGGATGACCAT	In vitro synthesis of congocidine cluster
FRAGMENT1REV	GACCGTACTCTGAACTGCCGGAGCTCT A	In vitro synthesis of congocidine cluster
FRAGMENT2FOR	CGGCAGTTCAGAGTACGGTCTGCTC AC	In vitro synthesis of congocidine cluster
FRAGMENTREV	TGCATGCCTGCAGGTCGACTTGAAGT CCGGAGCTCTA	In vitro synthesis of congocidine cluster
ACB AF	TGGAGCTCCACCGCGGTGGCCGAATC GAGTATGTCGGAGATC	In vitro synthesis of congocidine cluster
ACB AR	TTCATCGAGGGCAACGTCATGCACCA GATA	In vitro synthesis of congocidine cluster
ABC BF	ATGACGTTGCCCTCGATGAACACCGA GTAC	In vitro synthesis of congocidine cluster
ABC BR	GAGTTCGTCGCGTAGCCGGATGACC ATAG	In vitro synthesis of congocidine cluster
ABC CF	CCGGCTACGCGACGGAAGTCCGGCTGA TCGG	In vitro synthesis of congocidine cluster
ABC CR	TGCATGCCTGCAGGTCGACTCTGGAA CGGGACTTCGAGTAC	In vitro synthesis of congocidine cluster
FRAGMENT1FORAO	TGGAGCTCCACCGCGGTGGCCGAATC GAGTATGTCGGAGATC	In vitro synthesis of congocidine cluster

FRAGMENT1REVAO	TGTCCTGTCGTCTAGATGAACTGCCGG AGCTCTATC	In vitro synthesis of congocidine cluster
FRAGMENT2FORAO	CGGCAGTTCATCTAGACGACAGGACA TGACGTTCCA	In vitro synthesis of congocidine cluster
FRAGMENT2REVAO	TGCATGCCTGCAGGTCGACTCTGGAA CGGGACTTCGAGTA	In vitro synthesis of congocidine cluster
PEPAPTF	TCATTCATATGNNKNNKNNKNNKNNK NNKNNKNNKNNKNNKNNKNNKNNK NNKNNKNNKTGAGAATTTCGCAT	Aptamer study
PEPAPTR	ATGCGAATTCTCANNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNCATATGAATGA	Aptamer study
1.1f	CGAATCGAGTATGTCCGAGA	In vitro synthesis of congocidine cluster
1.1r	TGAACTGCCGGAGCTCTATC	In vitro synthesis of congocidine cluster
1.2f	AGCTGCCGAAAGGTGATGTG	In vitro synthesis of congocidine cluster
1.2r	GCAACGTCATGCACCAGATAC	In vitro synthesis of congocidine cluster
2f	CCTCGATGAACACCGAGTACC	In vitro synthesis of congocidine cluster
2r	CTGGAACTGGACTTCTCGCT	In vitro synthesis of congocidine cluster
3f	GTACTIONGCCCTCGAAGTCCT	In vitro synthesis of congocidine cluster
3r	TCAACGGGCACTACGAGAAC	In vitro synthesis of congocidine cluster
4f	GAGTACGGTCCTGCTCAGC	In vitro synthesis of congocidine cluster
4r	GCGTAGCCGGATGACCATAG	In vitro synthesis of congocidine cluster
5f	GACGGAACTCGGCTGATCG	In vitro synthesis of congocidine cluster
5r	TTGCCTCTCGAATTTTGC GG	In vitro synthesis of congocidine cluster
6.1f	TTTCACCTACTCGCCGCATT	In vitro synthesis of congocidine cluster
6.1r	GTTGACCTCGTCGAGTCCTG	In vitro synthesis of congocidine cluster
6.2f	CGACAGGACATGACGTTCCA	In vitro synthesis of congocidine cluster
6.2r	CTGGAACGGGACTTCGAGTA	In vitro synthesis of congocidine cluster

Diagnostic PCR

For routine diagnostic use in checking for DNA integrity and for sequencing. MyTaq from BIOLINE was used according to manufacturer guidelines in 200 μ L mastermixes, with reaction volumes of 10 μ L. Thermocycling conditions were individually determined and are detailed in the Results section.

Cloning PCR

The blunt ended high fidelity KOD Xtreme Hot Start DNA Polymerase (TOBOYO) was used for amplicons that would subsequently be cloned according to supplier recommendations. Thermocycling conditions were individually determined and are detailed in the Results section. To amplify large products or GC rich amplicons special step down thermocycling conditions were used, as it can be seen in Table 7.

Table 7. Cycling conditions of KOD Xtreme polymerase used to amplify GC rich

Cycle	Temperature ($^{\circ}$ C)	Time (min: seconds)
1. Polymerase activation	94	2:00
2. Denature	98	0:10
3. Annealing/Extension	72	1 min/kbp
Steps 2 and 3 for 5 cycles		
4. Denature	98	0:10
5. Annealing/Extension	70	1 min/kbp
Steps 4 and 5 for 5 cycles		
6. Denature	98	0:10
7. Annealing/Extension	68	1:10 min/kbp
Steps 6 and 7 for 25 cycles		

Blunt end cloning

The Invitrogen Zero Blunt[®] PCR cloning kit was used for blunt end cloning of PCR products. It was used according to the manufacturer's protocol and the resulting ligation mixtures were chemically transformed with competent TOP10 *E.coli*.

Restriction digest cloning

Adapted from (Sambrook *et al.*, 1989).

DNA to be cloned was obtained by PCR or restriction digest performed as per manufacturers guidelines (Promega, New England Biolabs). The DNA was examined by electrophoresis, and if correct, insert and vector were ligated.

Ligation

DNA inserts were obtained by PCR or restriction digest and purified from an agarose gel using an ISOLATE II purification kit (BIOLINE). DNA obtained through restriction digest was dephosphorylated with 1 μ L of thermosensitive alkaline phosphatase (TSAP, Promega) for 15 min at 37°C (this was done along with the restriction digest when the enzymes were from Promega). The fragments were mixed with different insert to vector ratios depending on size of fragments, typically 2:1, 3:1, and 5:1 or 10:1 and ligations performed with T4 Quick-StickLigase (BIOLINE) according to the manufacturer protocol.

Polymerase cycling assembly cloning

Adapted from the work of (Quan and Tian, 2014).

Vectors were linearised by either restriction digest or PCR amplification and the gel products purified to avoid carry over of the whole vector. The insert was produced by PCR with primers that included vector homology to the insert (required for the assembly), or with homology between each fragment, and subsequently gel purified. The parameters used to determine optimal length were the melting temperature (T_m) of the overlapping region and secondary structures that could arise were studied independently. This was designed to be as close to 60°C as possible for consistency between fragments. Once the vector and insert fragment(s) were ready, the assembly PCR was set up (see reagents in Table 8a/8b)

Table 8a. Reagents used in high fidelity cloning amplification of vector/insert

Component	Volume	Final Concentration
2X Xtreme buffer	25 μ L	1X
dNTPs (2 mM each)	10 μ L	0.4 mM (each)
PCR grade water	X μ L	
KOD Xtreme® Hot Start DNA Polymerase (1 U/ μ L)	1 μ L	0.02 U/ μ L

Template DNA	Y μ L	0.2 ng/ μ L from plasmid DNA 2 ng/ μ L from genomic DNA
Sense (5') primer (10 μM)	1.5 μ L	0.3 μ M
Anti-Sense (3') primer (10 μM)	1.5 μ L	0.3 μ M
Final volume	50 μ L	

Table 8b. Reagents in the assembly PCR reaction on Polymerase cycling assembly

Component	Volume	Final Concentration
2X Xtreme buffer	25 μ L	1X
dNTPs (2 mM each)	10 μ L	0.4 mM (each)
PCR grade water	X μ L	
KOD Xtreme[®] Hot Start DNA polymerase (1 U/μL)	1 μ L	0.02 U/ μ L
Vector	Y μ L	5-10 ng/ μ L
Insert	Z μ L	10-20 ng/ μ L
Final volume	50 μ L	

Overlapping PCR ligation

Adapted from (Heckman and Pease, 2007).

To achieve directional ligation of non-homologous blunt ended fragments an approach similar to polymerase cycling assembly cloning was used. A linker primer was designed to overlap 15-25 nucleotides at both ends of fragments to be joined, making up a single primer of 30-50 nucleotides. The two PCR fragments to be joined were gel purified and mixed in the reaction mix with the linker primer and can be seen in Table 9. Cycling conditions can be seen in Table 10.

Table 9. Overlapping PCR ligation reagents assembly.

Component	Volume	Final Concentration
2X Xtreme Buffer	25 μ L	1X
dNTPs (2 mM each)	10 μ L	0.4 mM (each)
PCR Grade Water	X μ L	
KOD Xtreme® Hot Start DNA Polymerase (1 U/ μ L)	1 μ L	0.02 U/ μ L
DNA part A	Y μ L	5-10 ng/ μ L
DNA part B	Z μ L	10-20 ng/ μ L
A-B primer (10 μ M)	4 μ L	0.8 μ M
Final volume	50 μ L	

Table 10. Cycling conditions of KOD Xtreme polymerase overlapping reaction.

Cycle	Number of cycles	Temperature ($^{\circ}$ C)	Time (min: seconds)
Polymerase activation	1	94	2:00
3 step cycling :			
Denaturation		98	00:10
Annealing	30	Linker Tm-3	00:30
Extension		68	1:00/kbp
Final extension	1	72	10:00

Gene disruption by λ -Red recombination

Adapted from (Gust *et al.*, 2006).

The apramycin cassette flanked by FRT recognition sites was excised from plasmid pIJ773 with a double restriction digest reaction using *EcoRI* and *HindIII* enzymes. The resulting DNA fragment was gel purified to avoid plasmid contamination. To further check the absence of the plasmid, DH5 α competent cells were transformed with the resulting cassette to ensure no carry-over of vector.

Two sets of primers were designed to target the DNA sequence to disrupt (gene *cgC10*): the sense primer would match 59 nucleotides upstream of the region to disrupt followed by a 20 nucleotide sequence present in the apramycin cassette (FRT sequence). The antisense primer would match 19 nucleotides in the FRT 3' end of the apramycin cassette and then 58 nucleotides downstream the region to disrupt. The apramycin cassette, was amplified using KOD Xtreme high fidelity polymerase and the desired PCR product was gel purified.

Electrocompetent cells of *E. coli* BW25113/pIJ790 were prepared as detailed above (cultivation temperature was changed to 30°C due to sensitive *OriT*) to introduce the *Streptomyces* cosmid to be disrupted. An overnight culture of the newly produced BW25113/pIJ790/cosmid of interest was inoculated to 10 mL of SOB without magnesium chloride containing the appropriate antibiotics and was induced with L-arabinose to a final concentration of 10 mM. When cells reached an OD₆₀₀ of approximately 0.5, electrocompetent cells were prepared. The resulting *E. coli* was transformed with the gel purified apramycin cassette and cells were then plated on to LB media containing the appropriate antibiotics and incubated for 12-16 h at 37°C to promote the loss of the temperature sensitive pIJ790. The resulting colonies were grown overnight, the cosmids extracted and confirmed by restriction digest and PCR. The correct cosmid was conjugated into *Streptomyces* to create the allelic replacement mutant.

Strain identification by 16S rDNA clonal DNA

Adapted from (Woese et al., 1990).

Strains were cultivated and their DNA extracted. The 16S rDNA clonal DNA was amplified with universal primers (Jiang et al., 2006) 27 F and 1492 R and submitted for sequencing.

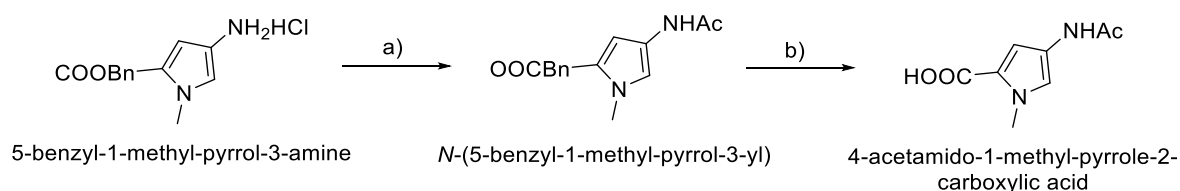
Chemical synthesis of novel heterocycles for feeding to congocidine producing strains

Synthesis of all compounds was carried out in the Dr. Glenn Burley Laboratory (Strathclyde Organic Chemistry department) with assistance from Giacomo Padroni (PhD student in the Burley laboratory), unless stated otherwise.

4-acetamido-1-methyl-pyrrole-2-carboxylic acid

Adapted from (Tao *et al.*, 1999).

The compound 4-acetamido-1-methyl-pyrrole-2-carboxylic acid was obtained by acetylation followed by saponification from the commercially available benzyl amino methyl pyrrole.



a) AcCl (1.5 eq.), TEA (3 eq), 0°C reaction time, 16 h, 98% yield.

b) NaOH 1N (10 eq), reflux, reaction time, 16 h, 94% yield.

a) A solution of benzyl 4-amino-1-methyl-pyrrole-2-carboxylate hydrochloride (0.4 g, 1.5 mmol, 1 eq.) in dry dichloromethane (DCM) (20 mL) was mixed with triethylamine (TEA) (0.8 mL, 4.5 mmol, 3 eq.). Acetyl chloride was slowly added at 0°C and the mixture was left stirring overnight at room temperature. DCM (10 mL) was added and the organic phase extracted with water (3x 10 mL), then dried over Na₂SO₄. The resulting oil was purified through silica gel chromatography using 5% MeOH in DCM to yield benzyl 4-acetamido-1-methyl-1H-pyrrole-2-carboxylate (0.4 g, 98%) as a brown oil.

¹H-NMR (CDCl₃ 400 MHz): 2.11 (s, 3H, Ac-H), 3.89 (s, 3H, N-C H₃), 5.26 (s, 2H, CH₂-Bn), 6.71 (d, 1H, Py-H, J= 2 Hz), 7.02 (s, 1H, CON H), 7.37 (m, 6H, Ph-H Py-H).

b) Benzyl 4-acetamido-1-methyl-pyrrole-2-carboxylate (0.4 g, 1.47 mmol, 1 eq.) was dissolved in tetrahydrofuran (THF) (10 mL), mixed with NaOH 1N (10 mL) and heated with reflux overnight. THF was removed and the aqueous phase acidified to pH 2 using HCl (1N), followed by extraction with ethyl acetate (3x 10 mL). The organic phases were pooled and washed with brine, dried over Na₂SO₄ and filtered to give 4-acetamido-1-methyl-1H-pyrrole-2-carboxylic acid (0.250 g, 94%) as a brown solid.

1 H-NMR (d6-DMSO 400 MHz): 1.94 (s, 3H, Ac-H), 3.80 (s, 3H, N-CH₃), 6.64 (d, 1H, Py-H, J= 2 Hz) 7.31 (d, 1H, Py-H, J= 2 Hz), 9.80 (s, 1H, CONH), 12.16 (b, 1H, COOH).

LC-MS (+ve mode) = m/z 183 [M+H] +

NMR and Mass data in Annex.

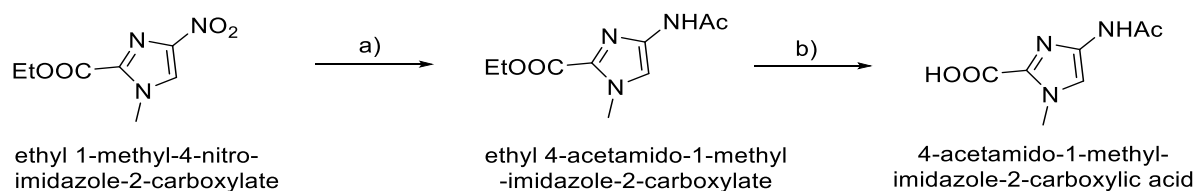
4-acetamido-1-(methyl-d3)-pyrrole-2-carboxylic

The deuterated compound 4-acetamido-1-(methyl-d3)-pyrrole-2-carboxylic acid was prepared from commercially available methyl 4-nitro-1H-pyrrole-2-carboxylate by alkylation and subsequent acylation and saponification by Dr. Abedawn Khalaf. NMR and Mass data in Annex.

4-acetamido-1-methyl-imidazole-2-carboxylic acid

Adapted from (Tao *et al.*, 1999).

Synthesis of 4-acetamido-1-methyl-imidazole-2-carboxylic acid started from the commercially available nitro ester, which was acetylated and reduced in a one pot reaction with a 66% yield. The final compound was obtained upon saponification with a 42% yield.



a) Pd / C (10% w/w), Ac₂O (1.5 eq.), H₂, reaction time, 16 h, 66% yield.

b) KOH 1N (5 eq.), reaction time, 16h, 42% yield.

A solution of ethyl 1-methyl-4-nitro-imidazole-2-carboxylate (0.5g, 2.5mmol, 1mmol) in MeOH/ EtOAc (1:1, 20 mL) was mixed with acetic anhydride (0.35 mL, 3.7 mmol, 1.5 eq.) and

Pd/C (10% w/w), and left to stir overnight at room temperature in a hydrogen atmosphere. Subsequently, the catalyst was filtered off and the mixture taken to dryness and purified through silica gel chromatography (MeOH 2% in DCM) to give ethyl 4-acetamido-1-methylimidazole-2-carboxylate (0.352g, 66%).

1 H-NMR (CDCl₃ 400 MHz) δ : 1.44 (t, 3H, COOCH₂CH₃), 1.58 (s, 3H, Ac-H), 4.08 (s, 3H, N-CH₃), 4.42 (q, 2H, COOCH₂CH₃), 7.83 (d, 1H, Im-H), 7.83 (s, 1H, CONH).

4-acetamido-1-methylimidazole-2-carboxylate (0.3 g, 1.4mmol, 1 eq.) was dissolved in THF and mixed with NaOH 1N (4 mL). The mixture was stirred at room temperature for 4 h. THF was removed and the water phase acidified to pH 2 with HCl 1N to form an orange precipitate, 4-acetamido-1-methyl-1H-imidazole-2-carboxylic acid that was collected after centrifugation (0.11 g, 42%).

1 H-NMR (d₆-DMSO 400 MHz): 1.92 (s, 3H, Ac-H), 3.89 (s, 3H, N-CH₃), 7.47 (d, 1H, Im-H), 10.54 (s, 1H, CONH).

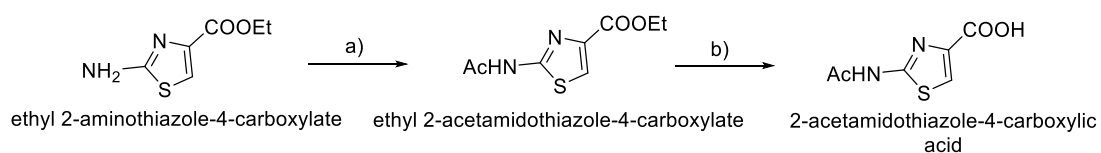
LC-MS (+ve mode) = m/z 184 [M+H] +

NMR and Mass data in Annex.

2-acetamidothiazole-4-carboxylic acid

Adapted from (Kobayashi *et al.*, 2011).

The synthesis of 2-acetamidothiazole-4-carboxylic acid started from an amine, kindly provided by Giacomo Padroni, who works in the Glenn Burley laboratory (Strathclyde University, Organic Chemistry department). Alkylation followed by saponification provided the pure compound for an overall yield of 82%.



a) AcCl (1.5 eq.), TEA (3 eq.), reaction time, 16 h.

b) KOH 1N (5 eq.), reaction time, 6h, 82% yield.

Ethyl 2-acetamidothiazole-4-carboxylate (0.26 g, 1.5 mmol, 1eq) was dissolved in dry DCM (20 mL) and mixed with TEA (0.8 mL, 4.5 mmol, 3 eq.). Acetyl chloride was slowly added at 0°C and the mixture was left stirring overnight at room temperature. DCM (10 mL) was added and the organic phase extracted with water (3x 10 mL), then dried with Na₂SO₄. The resulting oil was purified through silica gel chromatography using 5% MeOH in DCM to give ethyl 2-aminothiazole-4-carboxylate.

Ethyl 2-aminothiazole-4-carboxylate (0.7 g, 3.27 mmol, 1 eq.) was dissolved in THF (15 mL) and mixed with KOH 1N (15 mL) and stirred overnight at room temperature. THF was removed by vacuum and the water phase was acidified to pH 2 using HCl 1N, then extracted with EtOAc (3x 10 mL). The organic layers were combined, washed with brine, dried over Na₂SO₄ and taken to dryness to give the desired compound (33) (0.5 g, 82%) as a white solid.

¹H-NMR (d₆-DMSO 400 MHz): 2.15 (s, 3H, Ac-H), 7.94 (s, 1H, Th-H), 12.40 (s, 1H, CONH).

LC-MS (-ve mode) =m/z 187 [M+H]⁻

NMR and Mass data in Annex.

For clarity the compounds synthesised and used in the feeding experiments are summarised in Table 13, Chapter 2, along with the predicted congocidine-like compounds that may result from incorporation of the non-natural heterocycles in the congocidine peptide.

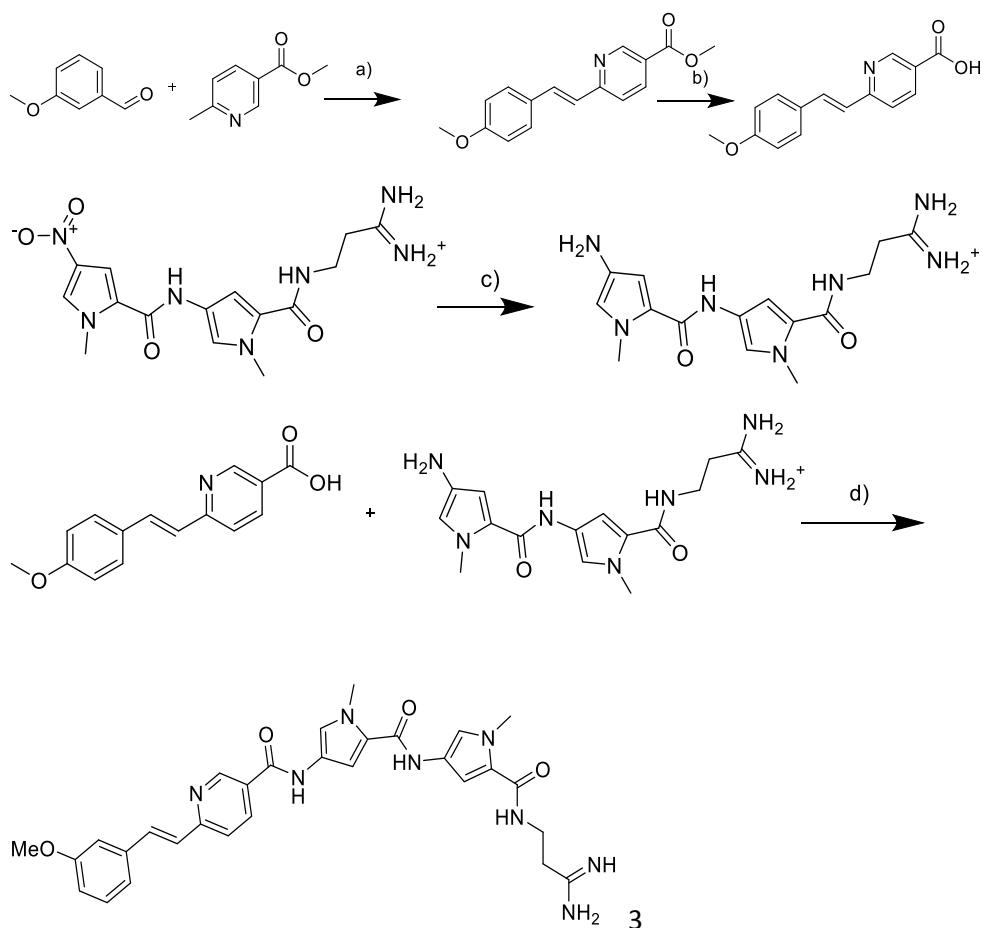
Chemical synthesis of two minor groove binders used in the chapter four assays (compounds 24 and 26)

Synthesis of ECS6 and ECS7 were carried out in the Colin Suckling laboratory (Strathclyde Organic Chemistry department) with guidance provided by Dr. Abedawn Khalaf and Dr. Fraser Scott.

Synthesis of MGB-6/ Compound 24

Reaction scheme:

- a) Ac_2O , ZnCl_2 , heated at 140°C , reaction time 48 h.
- b) $\text{EtOH}/\text{H}_2\text{O}$, NaOH , reflux heated, reaction time 3 h.
- c) MeOH , Pd/C -10 %, hydrogenated at room temperature for 3 h.
- d) HBTU , TEA , room temperature, reaction time 18 h.



3-Methoxybenzaldehyde (5.64 g) was dissolved in acetic anhydride (8 g) to which (6.2 g) of methyl-6-methylnicotinate was added in the presence of a catalytic amount of zinc chloride. The mixture was left stirring in a 100 mL round flask heated at 140° C for 48 h. The product formed was precipitated by adding ethyl acetate and small amounts of brine and methanol. The precipitate was recovered by filtration, and dried under vacuum. Methyl [(E)-6-(3-methoxystyryl)] nicotinic acid (592 mg) was dissolved in ethanol (5 mL) and 40 mL of water containing (0.8 g) NaOH was added. The mixture was reflux-heated for 3 h. The product was precipitated by acidification with HCl and filtered (313 mg).

1-amino-3-(1-methyl-4-(1-methyl-4-nitro-1H-pyrrole-2-carboxamido)-1H-pyrrole-2-carboxamido) propan-1-iminium (50 mg, 0.15 mmol) was dissolved in methanol (30 mL). Pd/C-10 % (80 mg) was added portionwise with stirring under nitrogen at 0°C. The reaction mixture was hydrogenated for 3 h at room temperature and atmospheric pressure. The catalyst was removed over kieselguhr and the solvent was removed under reduced pressure

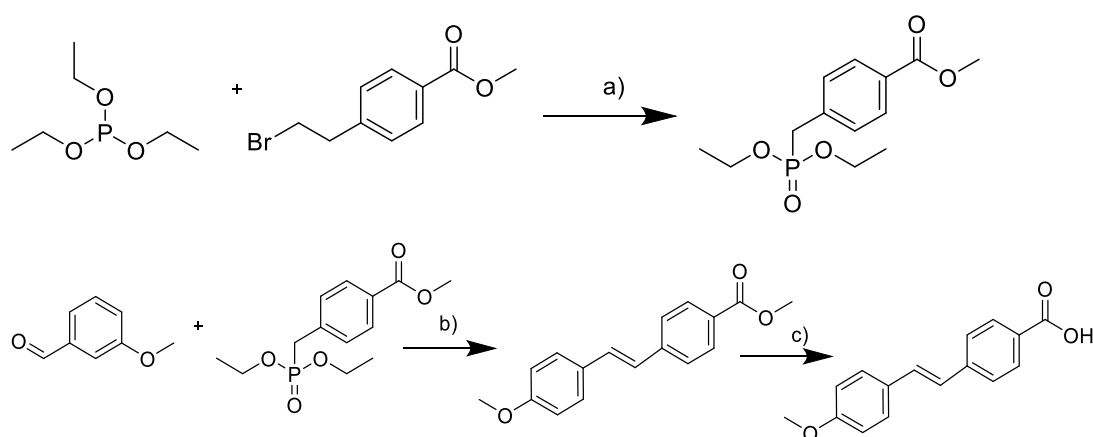
to give the amine which was dissolved in DMF (2 mL, dry). (E)-6-(3-methoxystyryl) nicotinic acid (39 mg, 0.15 mmol) and the coupling agent (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) (HBTU) (90 mg, 0.24 mmol) were added with a few drops of TEA. The reaction mixture was left stirring at room temperature overnight. DMF was removed under reduced pressure and the crude product was treated with ethyl acetate containing 5% methanol and a solution of sodium hydroxide (60 mg, 1.484 mmol in water 10 mL). The organic layer was collected, dried (Na_2SO_4), filtered and the solvent removed under reduced pressure. The crude product purified was purified by semi preparative HPLC. NMR and Mass data in Annex.

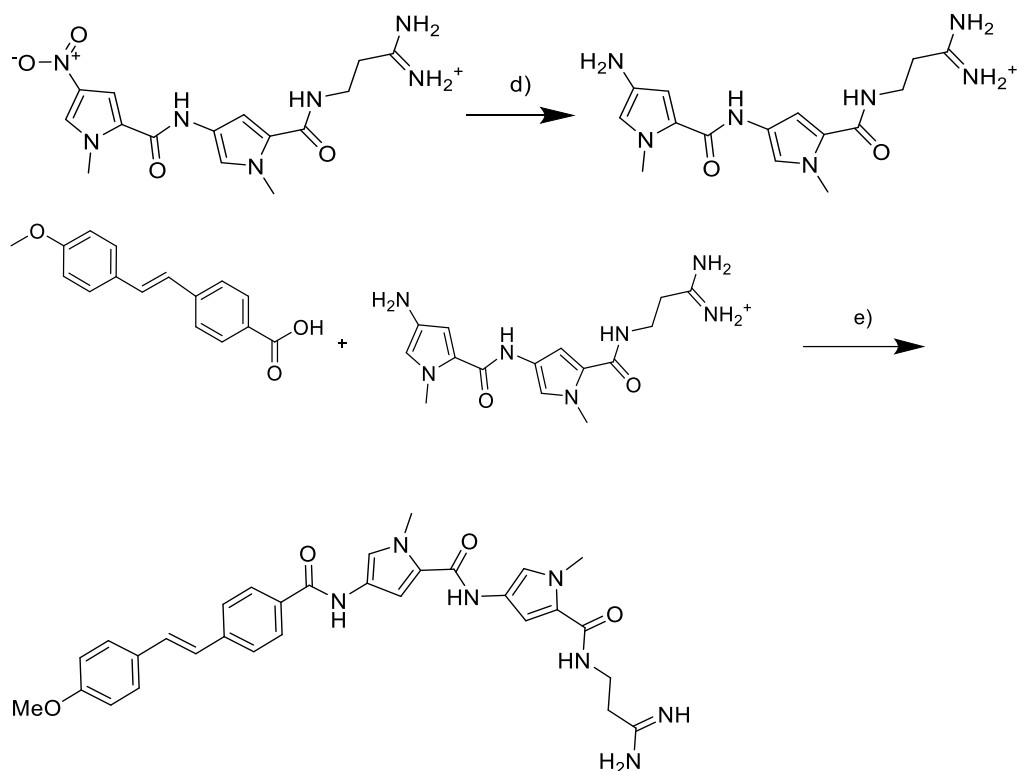
Synthesis of MGB-7/ compound 26

Reaction scheme:

- Heated at 160°C , reaction time 3 h.
- THF, NaH on ice, reaction time 18 h.
- EtOH/ H_2O , NaOH, heated to reflux, reaction time 2.5 h.
- MeOH, Pd/C-10 %, hydrogenated at room temperature for 3 h.
- HBTU, TEA, room temperature, reaction time 18 h.

Reaction scheme:





Triethylphosphite (4.3g) and methyl 4-(bromoethyl)-benzene (2.97g) was heated at 160°C for 2 h, after which, the resulting oil was dried in a rotary evaporator for 4 h. 4-[(diethoxyphosphoryl) methyl]-benzoate (3 g) was dissolved in (10 mL) dry THF under N₂ and mixed with (0.6 g) of NaH on ice, to which (1.6 g) m-anisaldehyde was added drop wise. The mixture was left stirring overnight and neutralised with HCl. The organic layer was kept, and the aqueous extracted with ethyl acetate. Both organic layers were combined and taken to dryness, washed with hexane and filtered (2.794 g). The resulting methyl (E)-4-(4-methoxystyryl)benzoate (1.976 g) was dissolved in 30 mL of ethanol and 80 mL of water containing (4.3g) NaOH and left stirring heated to reflux for 2.5 h. The product was precipitated by addition of HCl and filtered.

1-amino-3-(1-methyl-4-(1-methyl-4-nitro-1H-pyrrole-2-carboxamido)-1H-pyrrole-2-carboxamido) propan-1-iminium (50 mg, 0.15 mmol) was dissolved in methanol (30 mL). Pd/C-10 % (80 mg) was added portion wise with stirring under nitrogen at 0 °C. The reaction mixture was hydrogenated for 3 h at room temperature and atmospheric pressure. The catalyst was removed over kieselguhr and the solvent was removed under reduced pressure to give the amine which was dissolved in DMF (2 mL, dry). (E)-4-(4-methoxystyryl) benzoic

acid (39 mg, 0.154 mmol) and HBTU (90 mg, 0.238 mmol) were added with a few drops of TEA. The reaction mixture was left stirring at room temperature overnight. DMF was removed under reduced pressure and the crude product was treated with ethyl acetate containing 5% methanol and a solution of sodium hydroxide (60 mg, 1.484 mmol in water 10 mL). The organic layer was collected, dried (Na_2SO_4), filtered and the solvent removed under reduced pressure. The crude product purified was purified by semi preparative HPLC. NMR and Mass data in Annex.

Chapter 1: distamycin-congocidine combinatorial biosynthesis.

Introduction

Society needs new antibiotics to fight pathogenic micro-organisms. Paradoxically the source of these drugs is mostly micro-organisms according to a report indicating that 64% of novel compounds are natural products or natural product derivatives (Cragg and Newman, 2013).

Several approaches are used now in the effort to produce novel drugs. The ideas of those pioneers of antibiotic discovery back in the 'golden age' of antibiotic discovery (Cortes-sanchez and Hoskisson, 2015) are being revisited and expanded thanks to novel technologies. The research that led to the early characterisation of the congocidine pathway, narrowed down the set of genes necessary for its expression to 24 genes by gene knock out studies. Those genes were cloned into a cosmid (pCG002) containing 43.4 kb DNA from *S.ambofaciens*, which comprised the congocidine biosynthetic cluster, as well as a pBeloBAC11 backbone, including a Φ C31 integrase gene and *attP* and *oriT* with hygromycin as a selection marker (Juguet et al., 2009). Through this thesis, all work involving congocidine uses pCG002 which was a kind gift from Dr. Pernodet.

On the other hand distamycin was first isolated in 1964 (Arcamone *et al.*, 1964), but until recently its biosynthetic pathway was unknown. Recently two groups reported putative mechanisms of distamycin biosynthesis almost simultaneously (Hao *et al.*, 2014; Vingadassalon *et al.*, 2014), however when this study started, the enzymatic machinery remained unpublished. To overcome this lack of knowledge at the pathway level, an alternative to single gene combinatorial biosynthesis was envisioned. Combinatorial biosynthesis (Hopwood *et al.*, 1985), is utilised to engineer biosynthetic clusters, usually by mutating enzymes or module swapping (being this modules from the same or a different cluster) (Menzella and Reeves, 2007), however here we describe a whole cluster recombination strategy between the congocidine cluster (encoded in the pCG002 cosmid) and the distamycin cluster (present in the distamycin producing strain *Streptomyces netropsis* DSM40846). Since the chemical structures of congocidine and distamycin are very

similar, it was theorised that the genes comprising each biosynthetic cluster would also be very similar. This could allow for a recombination event between highly similar genes leading to the assembly of a hybrid cluster resulting in the assembly of a congocidine-distamycin hybrid peptide.

Validating congocidine biosynthesis

The original congocidine biosynthetic pathway was characterised by Dr. Pernodet lab (Juguet *et al.*, 2009), and a later -more in depth- biochemical characterisation of the pathway was achieved by Dr. Lautru (Lautru *et al.*, 2012) using *Streptomyces lividans* TK23 as the chosen heterologous host for congocidine production. At the same time, three novel *Streptomyces coelicolor* strains, named M1146, M1152, M1154 were reported to be more productive superhost strains (Gomez-Escribano and Bibb, 2011), which could be used for heterologous expression of several metabolite clusters.

These strains as well as another *S.coelicolor* engineered strain for heterologous production, (CH999) (McDaniel *et al.*, 1993), were used to ensure the cosmid pCG002 was functional and allowed production of congocidine. The cosmid was first transformed into *E coli* ET12567/pUZ8002, and then mixed with spores of each of the *Streptomyces* strains to facilitate conjugation, to finally screen for double cross overs. The phiC31 integrase carried on the cosmid facilitates integration of the biosynthetic cluster in to the genome of most *Streptomyces* strains.

The strains obtained were named CH999-002, M1146-002, M1152-002 and M1154-002. Such strains exhibited delayed sporulation (one day later when comparing with the parental strains).

Following sporulation, a glycerol stock solution was created and used to store the strains and to submit a liquid culture in YEME and GYM. In these experiments, strains M1146-002, M1152-002 and M1154-002 cultured in GYM medium (Fig.9) or YEME medium (Fig.10) showed a peak with retention time equal to that of congocidine (4.1 min) in the HPLC UV signal, and in the case of M1146-002 with the expected *m/z* mass (431). Both culturing media allowed production of the secondary metabolites. However, we never explored the levels of congocidine production, as strains were used to confirm congocidine pathway integrity. An unexpected result was that only strain M1146+pCG002 was able to produce congocidine as measured by LCMS. The reason why the other strains failed to produce congocidine was not

explored. Overexpression of some metabolites in M1146+pCG002 vs M1146 was analysed, but as the data in Fig.9 A) shows for the GYM media experiment, or Fig.10 A) for the YEME media experiment, there are no production of novel metabolites other than congoic acid between the strains. This is confirmed by MS data, and the overexpression of certain metabolites may be a result of increased metabolite flux, since the same masses are present in both strains but at different relative abundance levels.

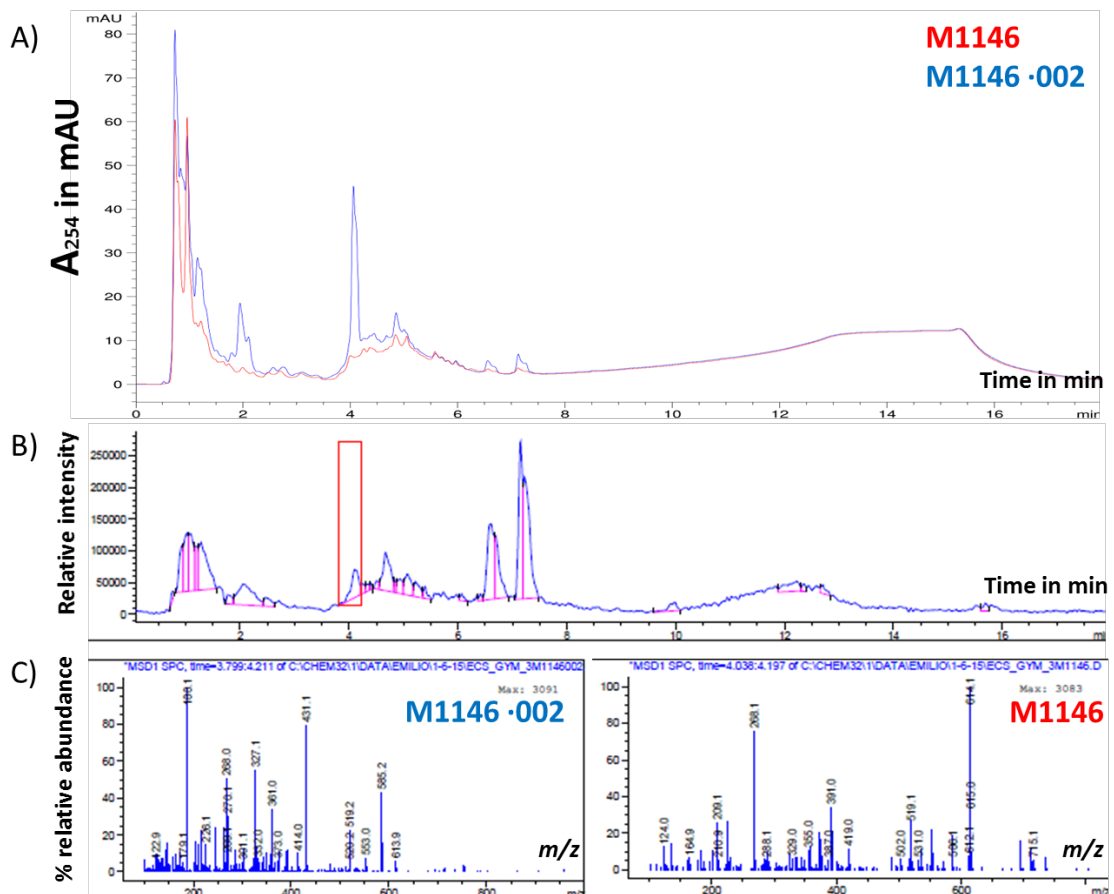


Figure 9. HPLC UV (254 nm) overlay of culture supernatants of M1146/M1146-002 strains cultured in GYM medium and MS of congoicidine. Congoicidine production is validated by mass spectrum. A) Signal of the strain M1146-002 (blue) overlaid to the signal of the parental strain M1146 (red). B) Associated to A) is the signal of the mass spectrum expressed in arbitrary intensity units. Congoicidine mass of 430 is seen in as $m/z + H$ (431) in C) for the case of M1146+pCG002 while is absent in M1146.

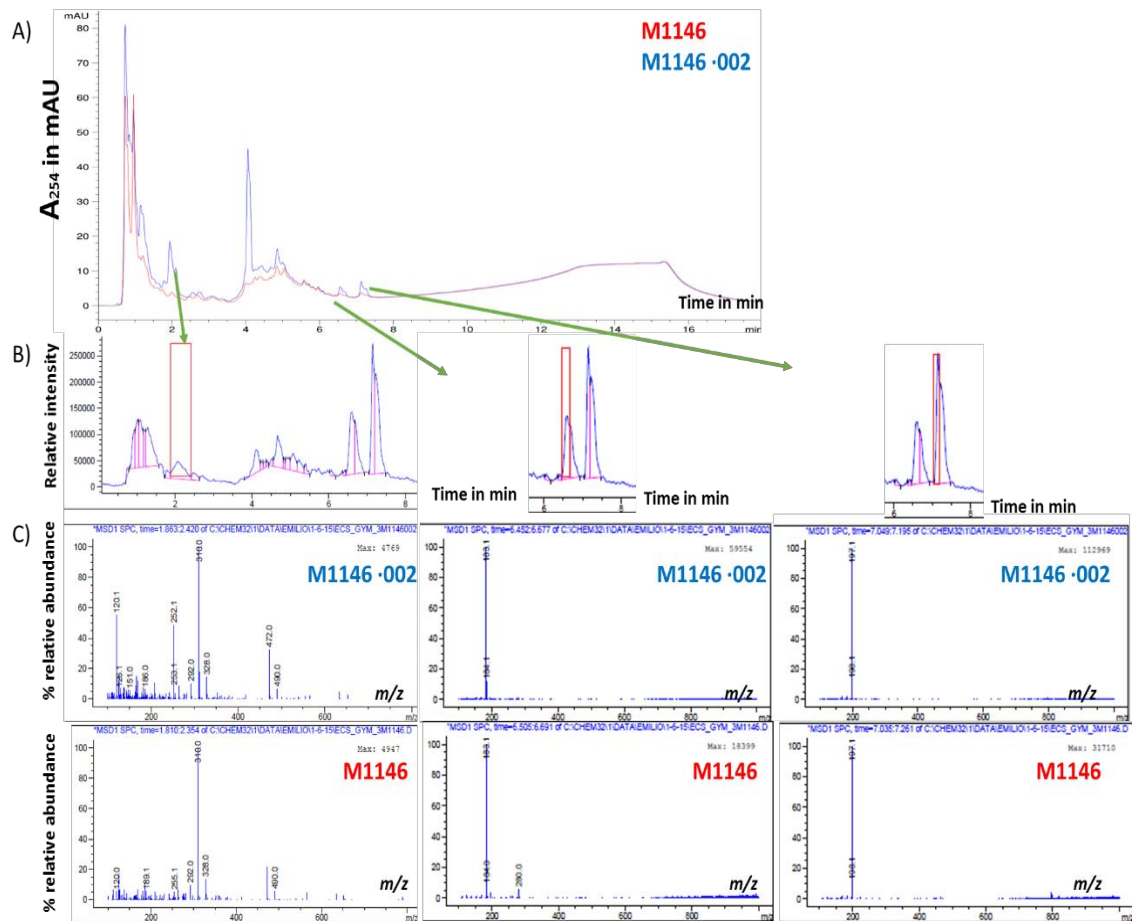


Figure 9 A. HPLC UV (254 nm) overlay of culture supernatants of M1146/M1146-002 strains cultured in GYM medium, with MS data showing no differences other than congoicidine production

A) Signal of the strain M1146-002 (blue) overlaid to the signal of the parental strain M1146 (red). B) Associated to A) is the signal of the mass spectrum expressed in arbitrary intensity units. C) Display the integrated MS signal of the peaks that differs in the overlay (signalled with green arrows), showing no differences between strains.

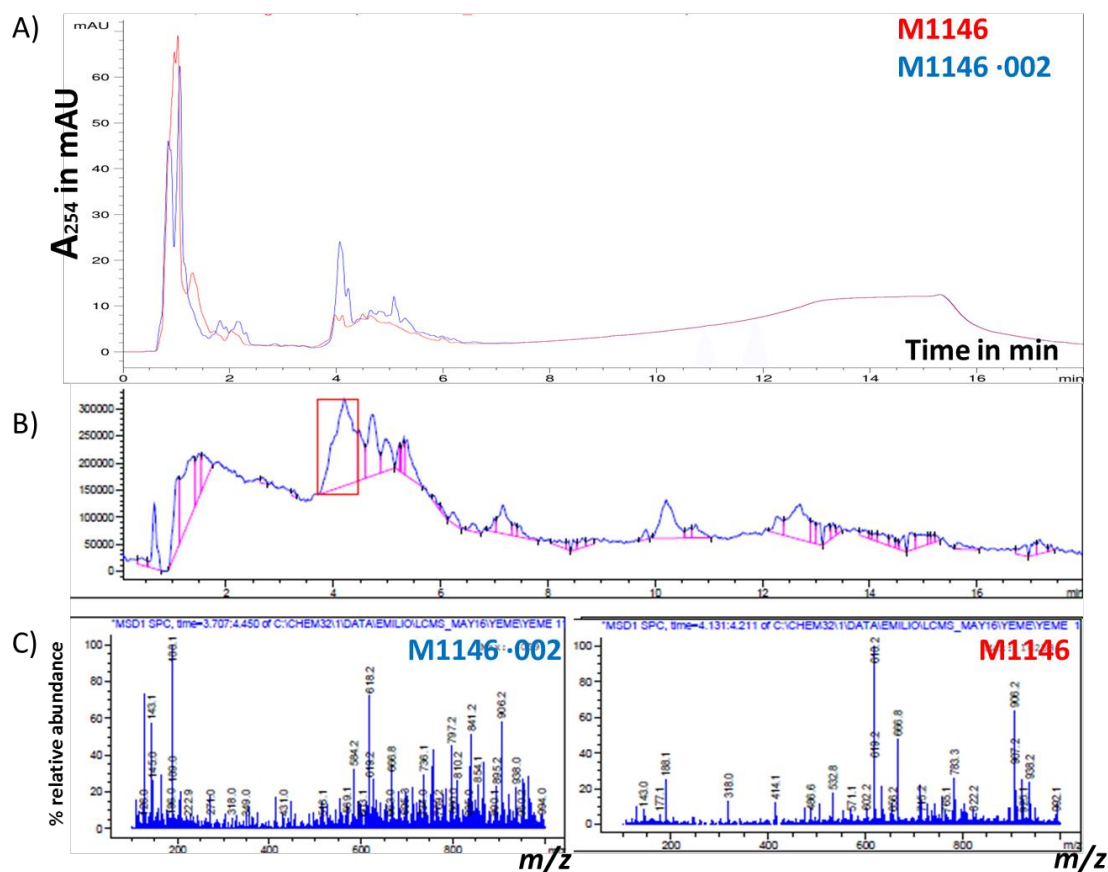


Figure 10. HPLC UV (254 nm) overlay of culture supernatants of M1146/M1146-002 strains cultured in YEME medium.

A) Signal of the strain M1146-002 (blue) overlaid to the signal of the parental strain M1146 (red), congoicidine mass of 430 is seen in as $m/z + H$. B) Associated to A) is the signal of the mass spectrum expressed in arbitrary intensity units. C) Congoicidine mass of 430 is seen in as $m/z + H$ (431) in C) for the case of M1146+pCG002 while is absent in M1146.

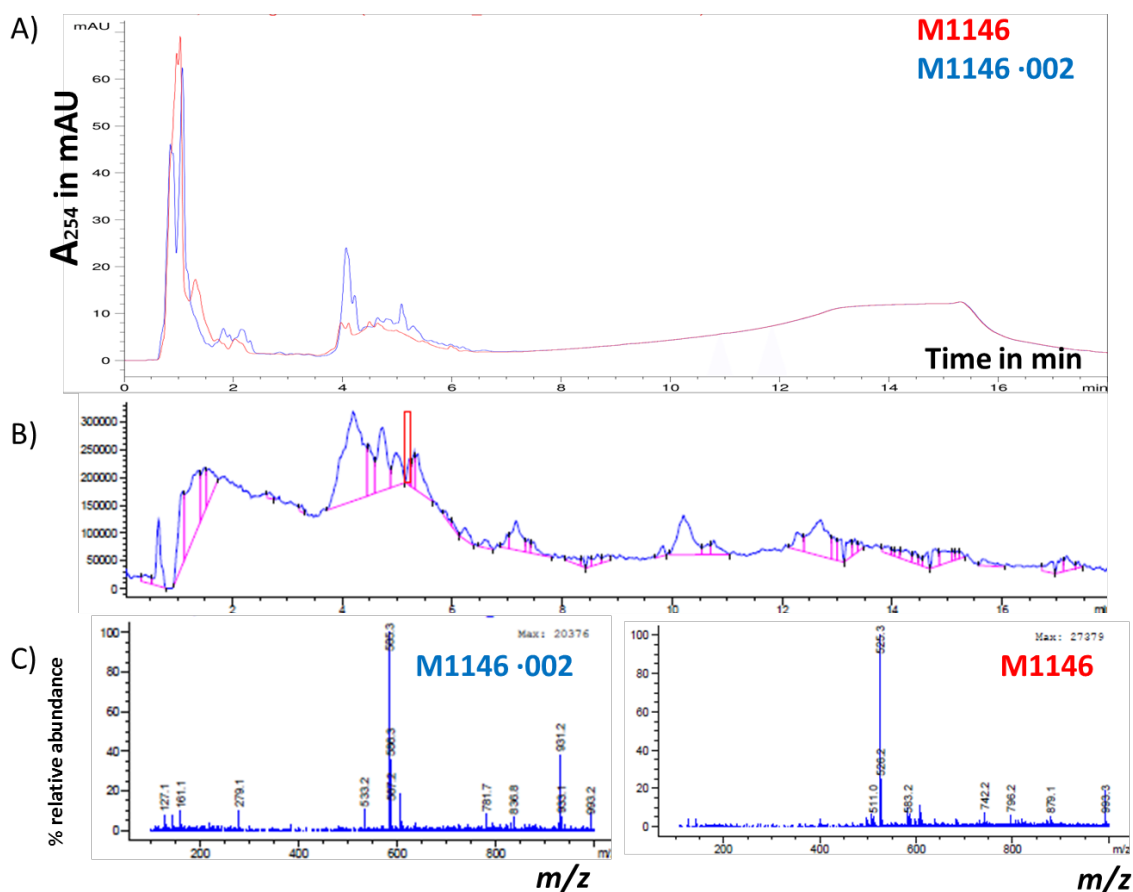


Figure 10 A. HPLC UV (254 nm) overlay of culture supernatants of M1146/M1146-002 strains cultured in YEME medium with MS data showing no differences other than congolectin production.

A) Signal of the strain M1146-002 (blue) overlaid to the signal of the parental strain M1146 (red). B) Associated to A) is the signal of the mass spectrum expressed in arbitrary intensity units. C) Display the integrated MS signal of the peaks that differs in the overlay, showing no differences between strains.

After validation that pCG002 by integration in the genome of some *Streptomyces* strains lead to the production of congocidine, the integration in the genome of a distamycin producer strain was pursued with a view to potential native combinatorial biosynthesis.

Congocidine-distamycin combinatorial biosynthesis.

A strain with reported distamycin production was purchased from DSMZ (*Streptomyces netropsis* DSM40846). Spores of this strain were prepared for storage in glycerol stocks but also to allow conjugational transfer of the cosmid pCG002 containing the congocidine cluster. The cosmid was first transformed into *E. coli* ET12567/ pUZ8002, and then the cosmid was transferred by conjugation in to *Streptomyces netropsis* DSM40846. The resulting strain was named *S.net002*, which possessed a delayed sporulation phenotype (one extra day) when compared to the wildtype strain on MS agar. More evident morphological variations were obvious when cultured in solid YEME such as a change in colony elevation from crateriform to umbonate and a more filiform margin, with production of a red pigment on the spores, as shown in Fig.11.

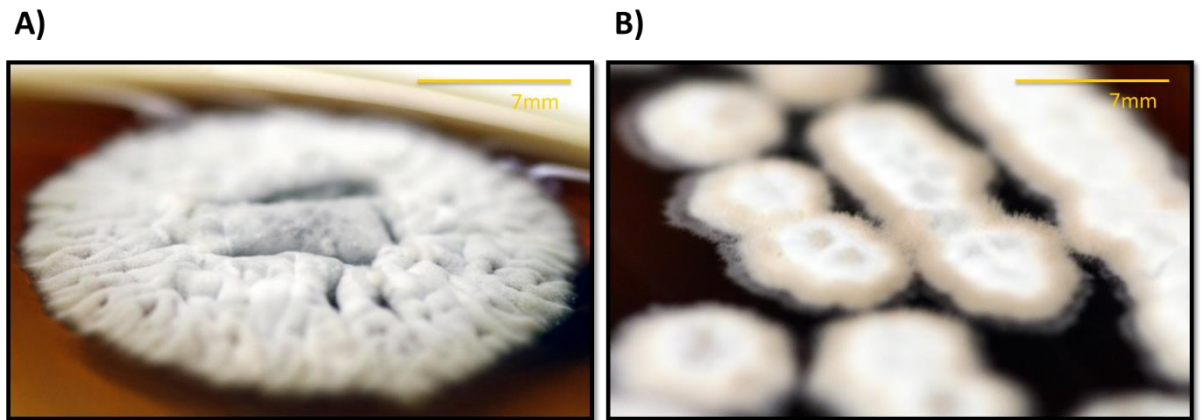


Figure 11. Morphological differences between *S.netropsis* (A) and *S.net002* (B) when culturing in solid YEME.

Change in colony elevation from crateriform in A) *S.netropsis* to umbonate in B) *S.net002*, as well as a more pronounced filiform margin, with production of a red pigment on the spores.

The spores obtained from *S.net002* were used to start a 50 mL YEME culture in parallel with *Streptomyces netropsis* in conical flasks containing springs. The cultures were sampled at 96 h. An overlay of the UV signal for the supernatants of both cultures at that time point showed no distinctive production of novel metabolites. Presence of congocidine is proved by comparison of the MS data and LC retention time of these samples to the authentic standard for the same conditions (see annex). For the authentication of distamycin production, as no standard could be purchased, the putative metabolite was purified by semipreparative HPLC and structure and purity confirmed by proton NMR (see annex). Production of congocidine is detected in both strains (retention time of 4.1 min, detected m/z+H, 431.2), although increased (in terms of UV area) in the strain containing the pCG002 congocidine pathway (Fig.12 A). Distamycin is detected this time point only in the strain containing the pCG002 congocidine pathway (Fig.12 B), (retention time of 5.3 min, detected m/z+H, 482.1), although the MS data for the same retention time indicates the presence of a compound with mass 445, which could be the distamycin backbone prior to N-pyrrole methylation.

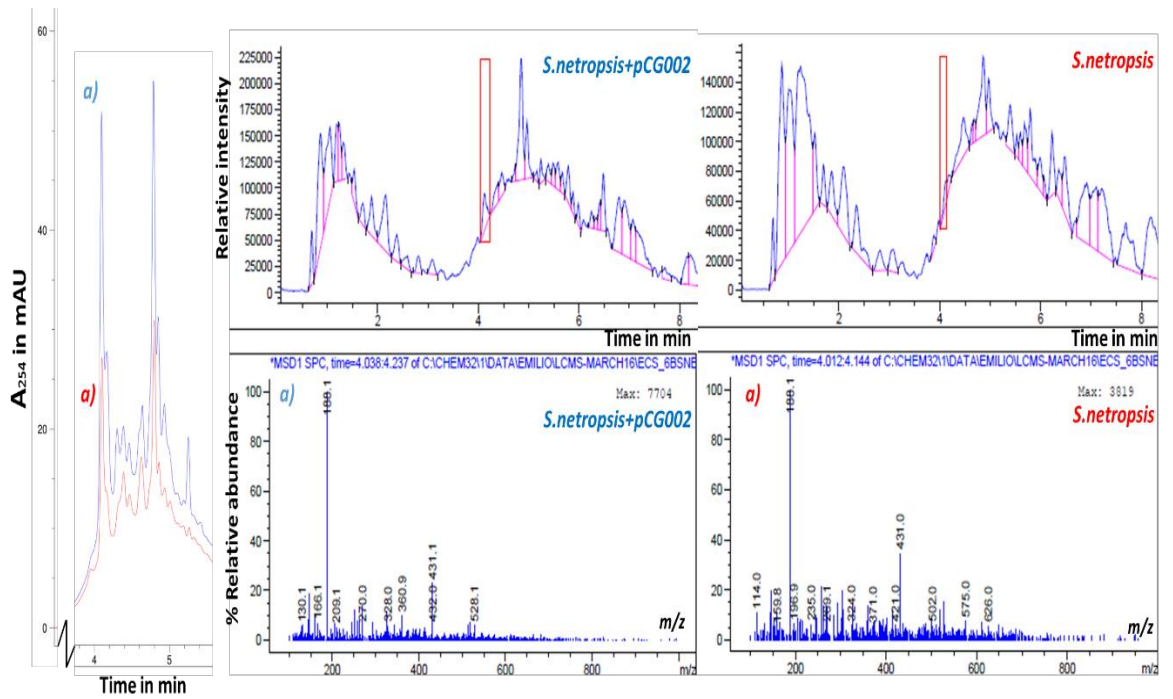


Figure 12 A. LCMS chromatogram signal overlay of two supernatants samples from *Streptomyces netropsis* and *S.net002*, with detection at 254 nm, and associated MS data.

The signal from *S.net002* (blue) indicates an increased production of congocidine ($t=4.1$, qualitative, detected $m/z+H$, 431.1) over the wildtype producer *S.netropsis* (red).

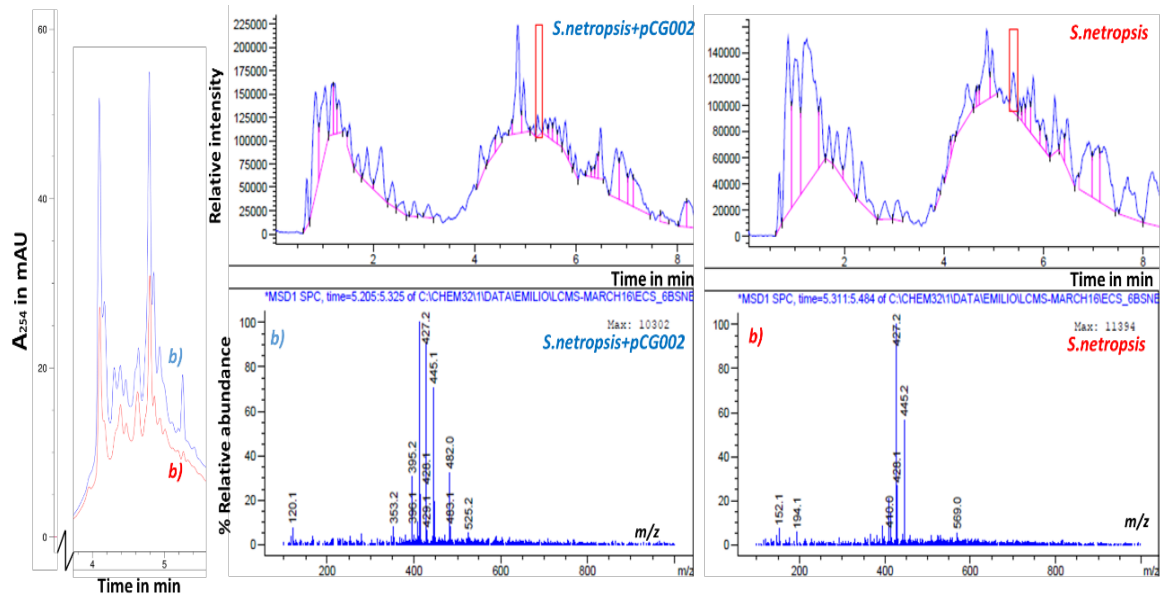


Figure 12 B. LCMS chromatogram signal overlay of two supernatants samples from *Streptomyces netropsis* and *S.net002*, with detection at 254 nm, and associated MS data.

The signal from *S.net002* (blue) indicates production of distamycin ($t=5.273$, qualitative, detected $m/z+H$, 482.1) over the wildtype producer *S.netropsis* (red).

This result was highly unexpected as that *Streptomyces netropsis* strain was not reported to produce congocidine. In light of the results and to further check the validity of this result both strains were grown again in YEME medium, but also in the recommended DSMZ propagation medium, GYM medium. At the same time, the m/z of congocidine and distamycin was scanned among the signal of the mass spectrum of the LCMS in positive mode, confirming the production of both metabolites by the wildtype *Streptomyces netropsis* DSM40846 when culturing on YEME (Fig.13 A), but with no production of either metabolite when cultured in GYM (Fig.14 A). The culture was grown for longer periods of time to see if at some point, congocidine disappeared and distamycin increased its yield, so that congocidine could be incorporated in distamycin biosynthesis, which could explain the production of congocidine not reported for the DSM4086 strain. However, the analysis performed on the supernatant always showed presence of congocidine, although at 120h the intensity of distamycin was greater (Fig.14).

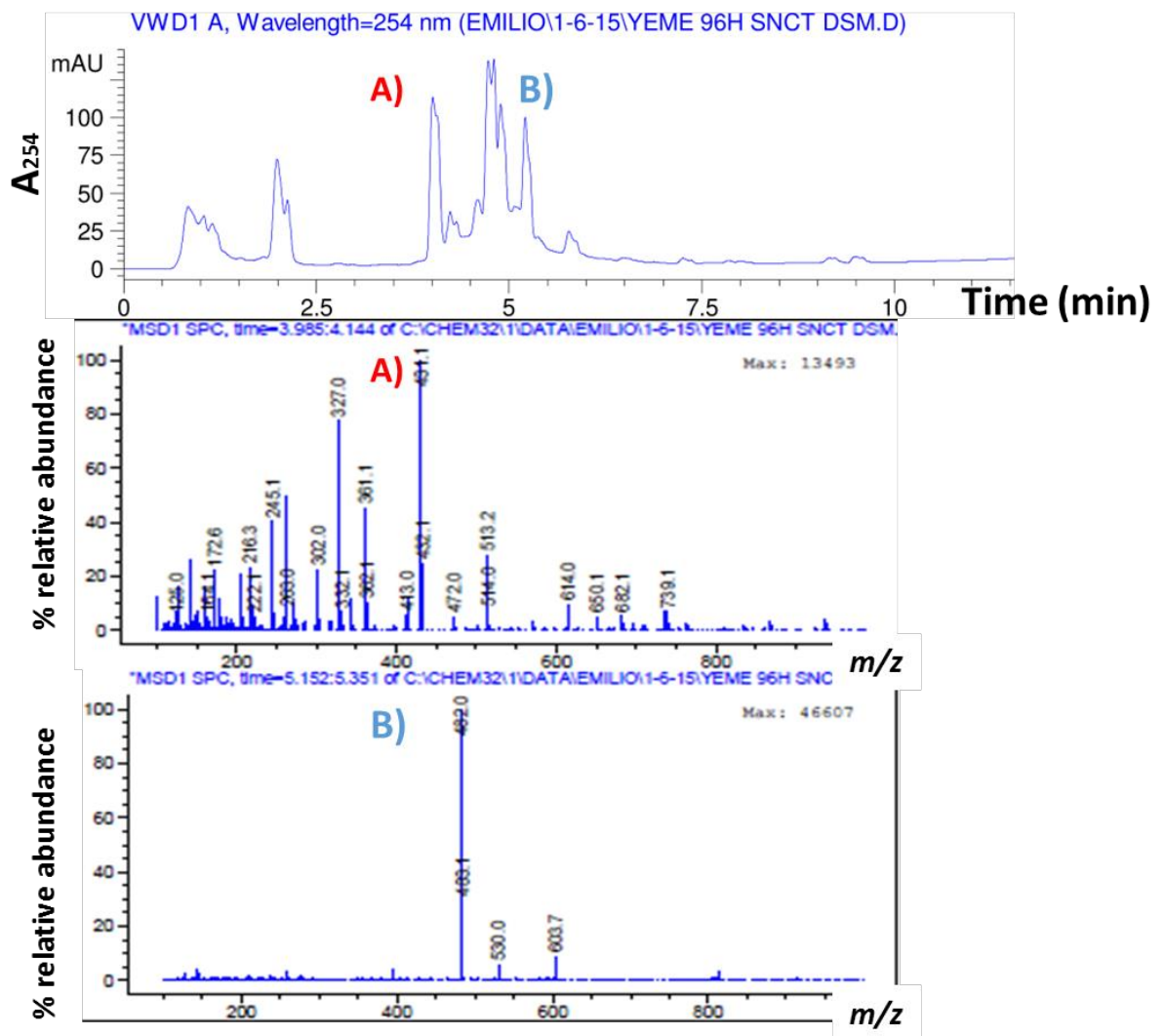


Figure 13. LCMS analysis of *S.netropsis* DSM40846 supernatant in YEME media.

UV signal of the metabolic profile in an LCMS of *S.netropsis* DSM40846 supernatant when cultured in YEME at 96h. In A) (retention time 4.041 min) the mass of congocidine (430), is detected as a molecular ion of 431.2 m/z (+H). In B) presenting a retention time of 5.229 min, he mass of distamycin+H (481+1) is observed, showing a molecular ion with 482.1 m/z .

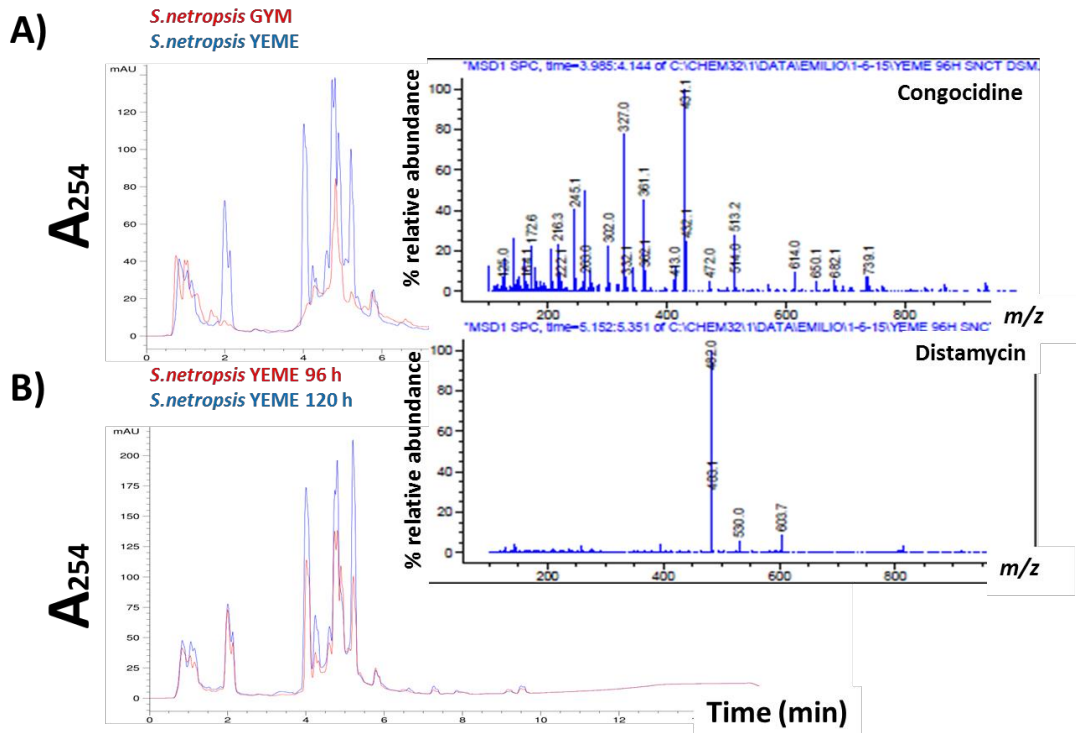


Figure 14. LCMS analysis of *S.netropsis* DSM40846 supernatant in YEME vs GYM medium, and in YEME at two different time points.

Figure 14A) is an overlay of the chromatogram of an LCMS of *S.netropsis* DSM40846 supernatant when cultured in GYM over the one when cultured in YEME. Absence of congocidine (4.05 min) and minuscule amounts of distamycin (5.3 min) are obtained when culturing in GYM as appreciated in the UV. Identity of peaks confirmed by mass spectrum, 431.2 m/z for congocidine, 482.1 m/z for distamycin.

In B) the red UV signal of an LCMS of *S.netropsis* DSM40846 supernatant when cultured in YEME at 96h and overlaid to it in blue is the UV signal of *S.netropsis* DSM40846 supernatant when cultured in YEME at 120h. Overall increase of metabolites is observed (qualitative), but especially with distamycin that at this time point is secreted in larger amounts than congocidine (in relation to UV absorbance). Identity of peaks confirmed by mass spectrum, 431.2 m/z for congocidine, 482.1 m/z for distamycin.

Rather than focusing in the presence of congocidine in the supernatant of the distamycin producer DSM40846, the research efforts were directed to isolate putative assembled novel hybrid compounds between congocidine and distamycin. It was theorised that such compounds may accumulated on the cell pellet, due to the inability of the mutant to secrete such compound, since the native resistance mechanisms is based on the efflux of distamycin. Since no differences between wildtype and engineered strain were detected in the supernatant, the extraction of the supernatant and the biomass with organic solvents of varied polarities (chloroform, ethyl acetate), was performed to concentrate the metabolites that may be unappreciable when analysing the supernatant and to assure that compounds inside the cells were also screened. For these extractions, larger volumes of culture were used. Culturing conditions remained constant, but the media was split on 30 erlenmeyers per strain, such as at every time point sampled, three replicates were extracted. To ensure homogeneous distribution of the inoculum, this was done in a two litre flask containing 1500mL of YEME media which was shaken evenly and distributed into the 30 flasks.

Due to the large number of samples generated, the biomass and bioactivity of the supernatant was monitored in liquid culture against *Bacillus subtilis* following the BSAC guidelines in a 96 well plate to check for any difference that could focus the solvent extraction work. As expected, biomass yield increased with time, to reach stationary phase at around 100 h. Both cultures grew without significant differences in terms of specific growth rate. The bioactivity of the supernatant against *Bacillus subtilis* remained low (10-15% inhibition rate) and constant through all the samples, indicating a low concentration of bioactive metabolites in culture. These results are shown in Fig.15.

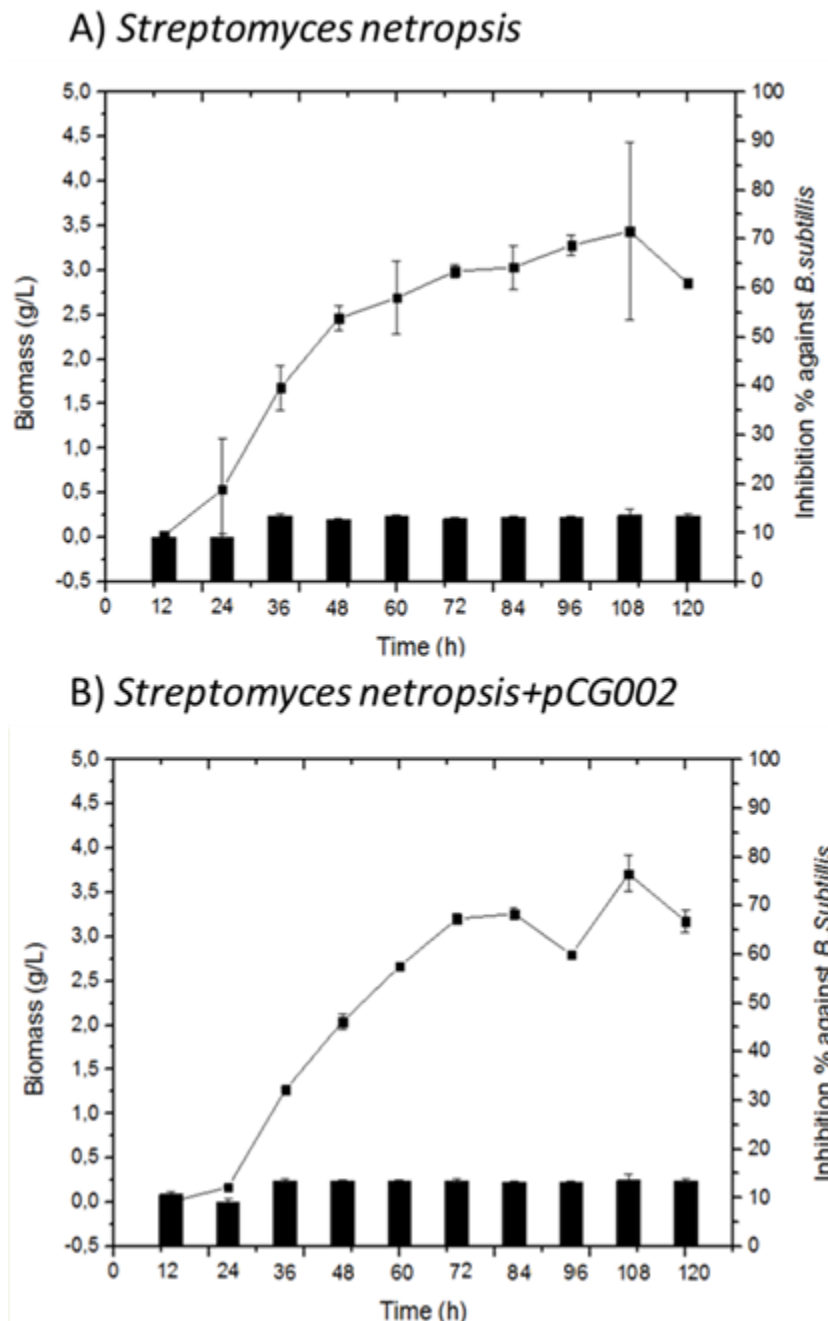


Figure 15. Biomass and bioactivity against *Bacillus subtilis* of two cultures of *Streptomyces* at different time points (12-120 h).

Each graph shows the amount of biomass in grams per litre at every time point and the bioactivity against *Bacillus subtilis* in liquid culture of the supernatant of the cultures of A) *Streptomyces netropsis* and B) *Streptomyces netropsis*+pCG002.

Since no significant differences were observed, it was decided to extract all the samples individually. In the fractioning of the culture and biomass, the first solvent used was chloroform, then ethyl acetate sequentially. The organic layers were concentrated to dryness in each case and resuspended in one millilitre of dimethyl sulfoxide, and analysed by HPLC (sample made up of 10% DMSO extract in water). While the ethyl acetate extraction did not present differences among strains, the extraction performed with chloroform rendered a different metabolite profile between *Streptomyces netropsis* and *S.net002* at time point 72 h (Fig.16). *S.net002* clearly overproduced some peaks present in both cultures but also produced peaks that were not in the sample of *Streptomyces netropsis*. These novel peaks were purified by semi-preparative HPLC, and their bioactivity screened against *B. subtilis* (Fig.16). The purified compounds were also submitted to High Resolution Mass Spectrum Direct Infusion Electrospray analysis, and although the mass for peak ten, indicated the putative assembly of a six pyrrole amide compound of molecular formulae of $C_{40}H_{45}N_{15}O_7$, with molecular mass of 847.36, hybrid of congocidine and distamycin, it was not possible to find a decomposition pattern from the proposed adduct, although the molecular ion was found [M+ACN+H]. As the reproducibility of this novel peaks had to be proved and with views on to isolate a larger amount of compound to produce two dimensional nuclear magnetic resonance spectroscopy data that would help identifying the molecule, another culture of the strains was performed. This time the strains were cultivated in six replicates made up of 400mL YEME medium in two litre flask, also containing a steel spring, at 30°C and 220rpm. The whole culture was extracted with chloroform at the same time point (72 h), all the organic layers pooled, concentrated and resuspended in Dimethyl Sulfoxide. However, this time the metabolite profile between both strains was identical, with only small deviations regarding productivity of some metabolites. The process was repeated twice a month for half a year without success. At this point, it was thought that, perhaps the strain *S.net002* was not stable and that lost or mutated the genes from pCG002 over time. In order to discard this eventuality, the strain *S.net002* was streaked on nutrient agar plates containing hygromycin to subsequently propagate into MS agar and create new spore stocks to perform new cultures. The spores arising from this plates were then used to produce the inoculum of new cultures, yet the HPLC results were again the same for both strains *Streptomyces netropsis* and *S.net002* after three attempts. To discard mutations in the stocks of *S.net002*, the strain was made again through a fresh conjugation, although the novel peaks were never reproduced, even when looking at earlier and later time points, and analysis of the

supernatants, biomass and supernatant extraction individually that were performed for an additional half year without constant reproducibility.

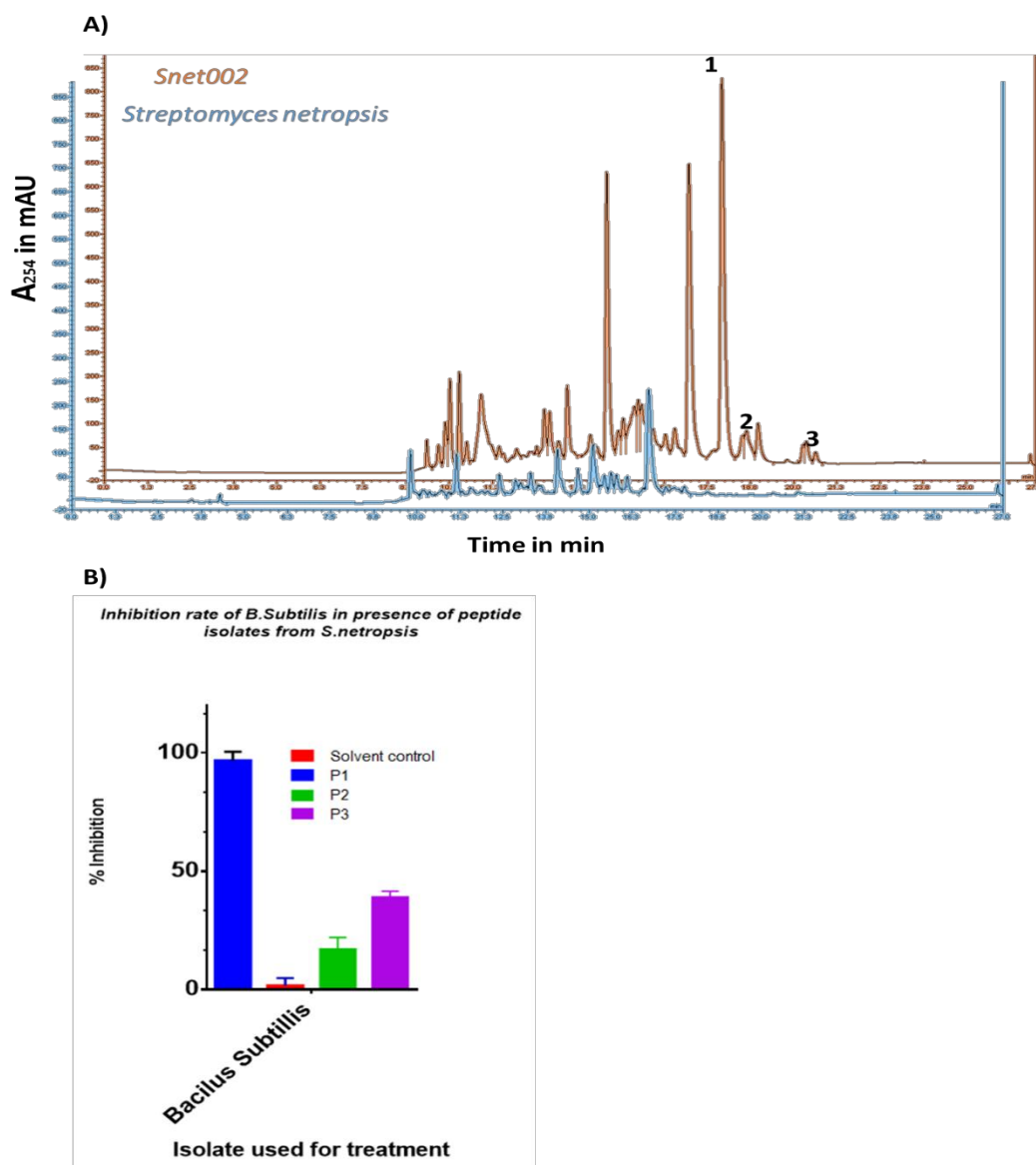


Figure 16. A) HPLC UV signal overlay of the chloroform extract of *Streptomyces netropsis* and *Snet002* fermentations. B) bioactivity of some of the metabolites isolated from *Streptomyces netropsis*+pCG002 on *Bacillus subtilis*.

A) Is the HPLC UV signal overlay of apolar metabolites present in the culture of *S.netropsis* (Blue) / *S.netropsis*+pCG002 (red). Putative *de novo* production of three metabolites (peaks 1-3). Such metabolites were purified by preparative HPCL and assayed against *Bacillus subtilis* following BSAC guidelines, with results on B) as percentage of inhibition, with 0% indicating no effect and 100% no measurable growth, negative control was done with carrier solvent (DMSO).

Repeated analysis could not detect the peaks 1-3 produced by *S.net002*, although a consistent difference between both cultures was observed. Several metabolites had their yield increased in the *S.net002* culture up to two fold (in relative terms, as increasing UV absorbance of certain peaks between strain samples), but one particularly was overproduced to a six fold (in terms of area of the UV peak at 254nm). UV overlay and molecular ion results are shown in Fig.17. This metabolite was identified by LCMS as aureothin, a known antifungal also known as mycolutein, distacin or antibiotic 74A, first isolated from *Streptomyces thioluteus* (Nakata *et al.*, 1961), that was also reported to be produced by some strains of *S.netropsis* (Akhunov *et al.*, 1976). Details of their biosynthesis, chemical synthesis and characterisation are known (He *et al.*, 2003; Jacobsen *et al.*, 2005)

Since the production of aureothin was already reported in the *Streptomyces netropsis* strain, its detection was not a product of introducing the congocidine cluster. The increase might be a regulatory response that could rely on genetic elements from the congocidine cluster or just a response to stress.

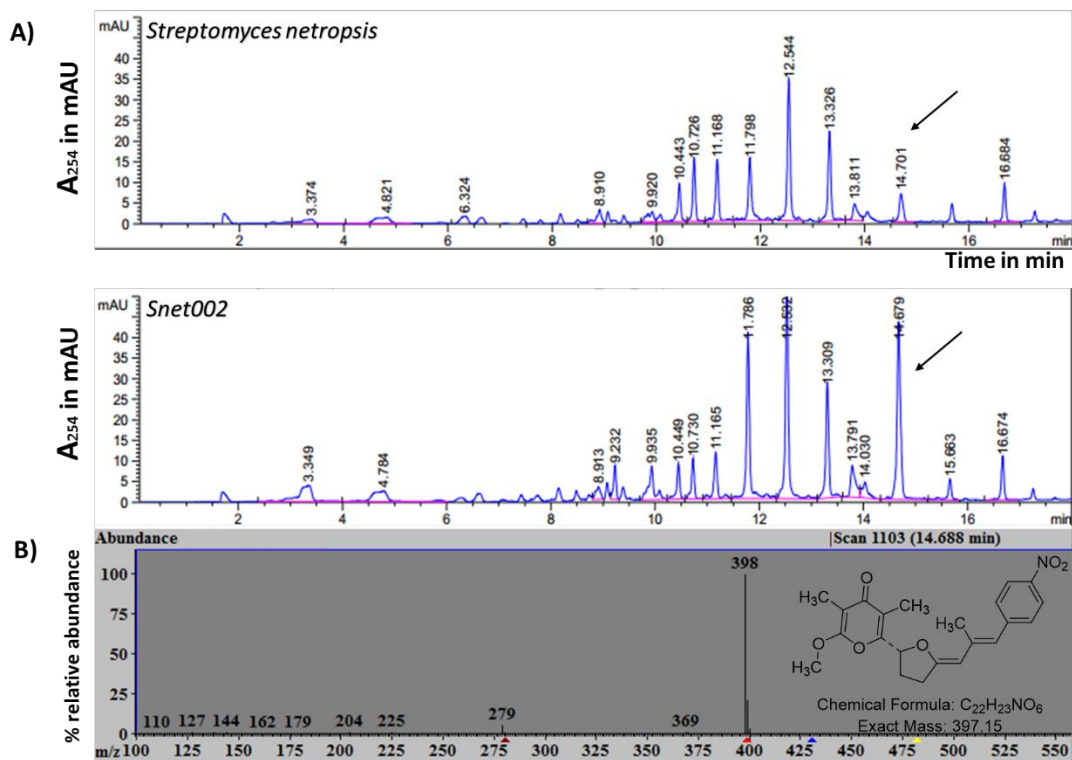


Figure 17. UV trace from LCMS data of a chloroform extraction of *S.netropsis* and *S.net002* cultures. Aureothin increased production is indicated with an arrow.

UV trace of apolar metabolites present in the culture of *S.netropsis* / *S.net002* (A). The increase of a peak at approx. $t=14.7$ min is due to overproduction of aureothin (indicated with an arrow), which structure is indicated in B), along with the molecular ion with m/z of 398.0 (M+H) as reported in (Jacobsen *et al.*, 2005).

While trying to replicate the results, big deviations in culture colour (red, black or intense yellow) from culture to culture of *S.net002* were observed, even between replicates inoculated with the same pre-culture. To check for these differences, a small aliquot (500µL) as well as a 1:4 dilution of the pre-culture originated from *S.net002* spores, were spread into a MS plate and incubated at 30°C for four days; after this time, several different *Streptomyces*-like phenotypes were observed, from which four different morphotypes could be isolated (Fig.18). The morphotypes were named *S.nety1*, *S.nety2*, *S.nety3*, *S.nety4* and significantly differed from each other when cultivated in MS agar. To the naked eye, *S.nety4* was the one that showed closest similarity to the colonies formed by *S.net002* (originating inoculum); *S.nety3* presented an umbonate elevation with undulate margin and irregular form, but the most remarkable phenotypical differences was production of an intense yellow pigment and complete loss of sporulation; *S.nety2* had a circular form with raised elevation and entire margin, with a dark grey pigment production and fast sporulation; *S.nety1* was similar to the colonies of *S.netropsis* (Fig.18).

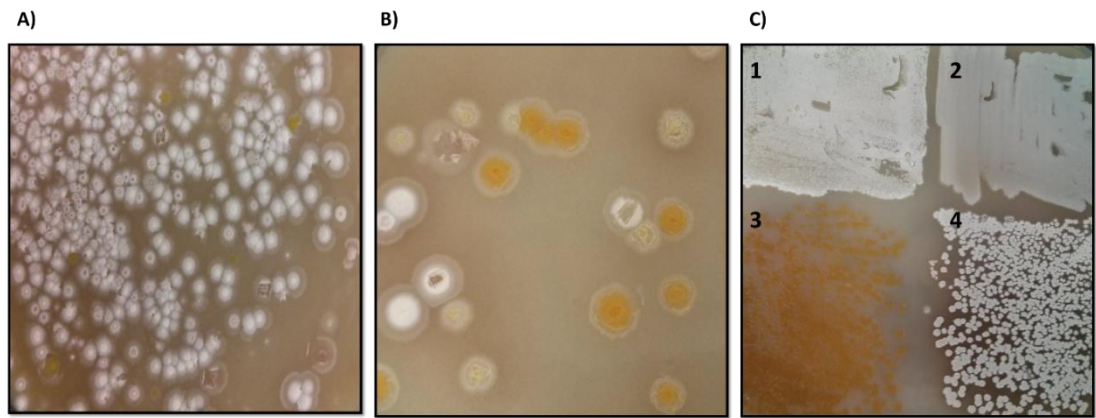


Figure 18. Different morphotypes isolated from a culture of *S.net002*.

The figure show the original MS agar plate in which the preculture was spreaded with A) supernatant, B) a 5x dilution with water. C) Shows the plate in which four stable morphotypes were isolated with and were randomly named *S.nety1*, *S.nety2*, *S.nety3*, *S.nety4*.

To disregard cross-contamination from other *Streptomyces* spores in the laboratory, the morphotypes were cultivated alone and their genomic DNA extracted to perform 16S rDNA sequencing (Woese *et al.*, 1990). Analysis of the sequences confirmed that all the four morphotypes belonged to phenotype variations of the *S.netropsis* strain, as the phylogenetic tree on Fig.19a illustrate (built upon the alignment displayed on ANNEX 1).

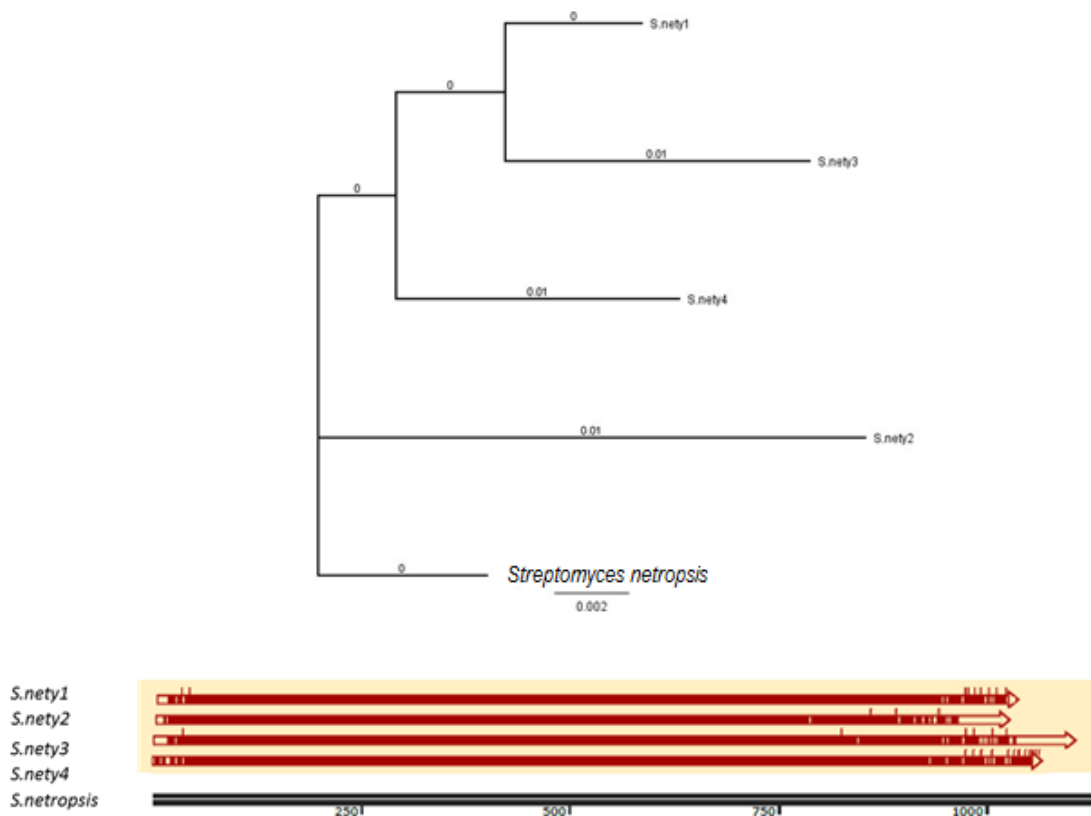


Figure 19a. Phylogenetic tree showing differences between *Streptomyces netropsis* and the four phenotypes originated after conjugational transfer of pCG002 and 16rDNA alignment overview.

Phylogenetic tree built with the Neighbor-Joining method (Saitou N, 1987) using the Jukes-Cantor or JC69 (Jukes and Cantor, 1969) model for genetic distance showing the morphotype relationships and expressing their distances in number of amino acids substitution per site of each branch. Alignment of the sequences (unclipped sequence) are graphically indicated below. Figure 19.b show the alignment used for the tree.

Due to the marked morphological differences in the four isolated strains, new cultures in YEME medium were performed in order to check if the metabolite variability observed in previous cultivations of *S.net002* were the product of different morphotypes in culture. This time the metabolite extraction was performed by disruption of the mycelium by sonication in methanol, providing a way to monitor all the metabolites present in the ethyl acetate and chloroform extractions with one sample analysis. LCMS analysis of such extractions from the four morphotypes cultured in parallel with *Streptomyces netropsis* are shown in Fig.20. Qualitative differences were found (based on UV absorbance), such as increased production of aureothin by *S.nety1* and different levels of congocidine and distamycin production between all the strains. When comparing the UV signals presented in Fig.20, the parental strain *Streptomyces netropsis* produces distamycin, congocidine and aureothin (in order of UV absorbance), *S.nety1* presents diminished congocidine and distamycin UV absorbance which is detected in an 1:1 ratio, however it doubles the aureothin UV peak area; in *S.nety2* and *S.nety3* congocidine, distamycin and aureothin are almost not detected; *S.nety4* presents the same levels of congocidine and aureothin UV absorbance but no detectable distamycin.

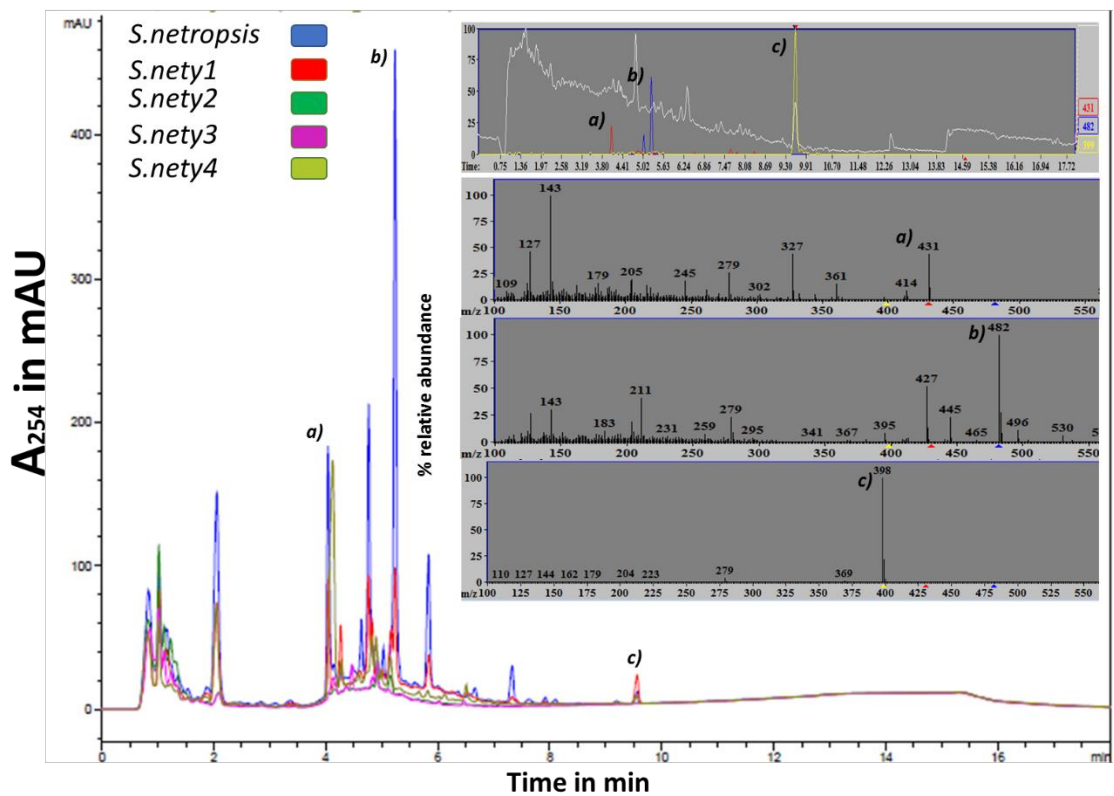


Figure 20. LCMS chromatogram signal overlay of *S.netropsis* and different morphotypes of *S.net002* (methanolic extract). Detection at 254 nm.

Congocidine (a), distamycin (b), aureothin (c) are identified by m/z . Metabolite profile differ greatly between morphotypes. When comparing UV absorption levels with the parental strain *Streptomyces netropsis*, only *Snety4* maintain the congocidine production, while *Snety1* halves it and the other two morphotypes hardly produce any. Distamycin production is greatly diminished in *Snety1* and non existing on the remaining morphotypes. Interestingly *Snety1* doubles aurothin production (in terms of UV detection area).

Conclusions

Integrating the whole congocidine cluster on a cosmid into *Streptomyces netropsis* resulted in genomic instability. This maybe the result of pleiotropic effects from genetic elements within the cluster or several recombination events due to high cluster similarity between that of congocidine on the cosmid and the native genome encoding the distamycin biosynthetic clusters. Further study should be carried out to reach any of these conclusions. Combinatorial biosynthesis is a complex technique that is not completely understood, and thus every pathway should be studied individually, without making generalisations regarding techniques to use. Perhaps the approach followed in this chapter was too ambitious, as the *S.netropsis* strain is not well known and the knowledge of its genomic sequence should be pursued. Study of combinatorial biosynthesis among congocidine and distamycin could be studied in known heterologous hosts in which more variables can be set constant. Perhaps combinatorial experiments may be more successful by synthetically constructing congocidine-distamycin genetic cluster hybrids or smaller modifications such as homologous module swapping between clusters.

Chapter 2: Synthetic assembly of the congocidine cluster.

Introduction

The study of specialised metabolite biosynthetic genetic clusters and their engineering is restricted by the ability to rapidly manipulate the DNA that encodes their enzymes. Exploring new DNA editing or assembly methods may allow precise control over assembly of large DNA fragments in a rapid and robust manner (Wright, 2014). Developing methods that are independent of restriction enzymes have been sought in an attempt to produce tuneable platforms for rapid mutasynthetic and combinatorial biosynthetic experiments. This chapter shows how three recent DNA assembly techniques were applied to attempt the synthetic assembly of the congocidine biosynthetic gene cluster. This was done to establish a proof of concept for the technology, and to provide a platform for rapid manipulation via gene replacement or exchange within the cluster. If successful, future work may enable rapid alterations of biosynthetic gene clusters to increase the biosynthetic capability of natural gene clusters.

Organisation of the congocidine cluster

The congocidine pathway is contiguously organised in a 31.127 bp gene cluster (Fig. 21) within the genome of *Streptomyces ambofaciens* (Juguet *et al.*, 2009). The gene cluster was previously cloned in a cosmid vector, pCG002 (Juguet *et al.*, 2009), which was kindly provided by Dr. Jean Luc Pernodet (University Paris Sud). Gene putative function are indicated in table 11. This cosmid was used as a template to amplify the DNA parts used to perform the assembly experiments.

Table 11. Organisation of the congocidine gene cluster

SAM number	ORF		Putative function	Proposed role in congocidine biosynthesis
	Cgc annotation	Size (aa's)		
SAMR0922		321	Pirin-like protein	Not involved
SAMR0921	cgc3*	532	Acyl-CoA synthetase	Congocidine assembly
SAMR0920	cgc2*	611	ACB transporter, ATP binding protein	Resistance
SAMR0919	cgc1*	619	ACB transporter, ATP binding protein	Resistance
SAMR0918	cgc1	223	Two-component response regulator	Regulation
SAMR0917	cgc2	470	NRPS C domain	Congocidine assembly
SAMR0916	cgc3	485	Aldehyde dehydrogenase	Precursor biosynthesis
SAMR0915	cgc4	181	Nucleoside 2-deoxyribosyltransferase	Precursor biosynthesis
SAMR0914	cgc5	297	Dihydroorotate dehydrogenase	Precursor biosynthesis
SAMR0913	cgc6	276	Creatininase	Precursor biosynthesis
SAMR0912	cgc7	377	Hypothetical protein	Precursor biosynthesis
SAMR0911	cgc8	436	Nucleotide sugar dehydrogenase	Precursor biosynthesis
SAMR0910	cgc9	327	Nucleoside diphosphate sugar epimerase	Precursor biosynthesis
SAMR0909	cgc10	360	Glycosyltransferase	Precursor biosynthesis
SAMR0908	cgc11	254	Sugar nucleotidyltransferase	Precursor biosynthesis
SAMR0907	cgc12	381	Nucleotide-sugar aminotransferase	Precursor biosynthesis
SAMR0906	cgc13	641	Glycoside hydrolase	Precursor biosynthesis
SAMR0905	cgc14	302	Amidohydrolase	Precursor biosynthesis
SAMR0904	cgc15	268	Methyltransferase	Methylation of pyrrole
SAMR0903	cgc16	445	NRPS, C domain	Congocidine assembly

SAMR0902	cgc17	348	Alcohol dehydrogenase	Precursor biosynthesis
SAMR0901	cgc18	1067	NRPS, A-PCP-C domains	Congocidine assembly
SAMR0900	cgc19	121	NRPS, PCP domain	Congocidine assembly
SAMR0899	cgc20	221	RNA polymerase ECF subfamily factor	Not involved
SAMR0898	cgc21	217	Hypothetical protein	Not involved
SAMR0897	cgc22	100	Membrane protein	Not involved

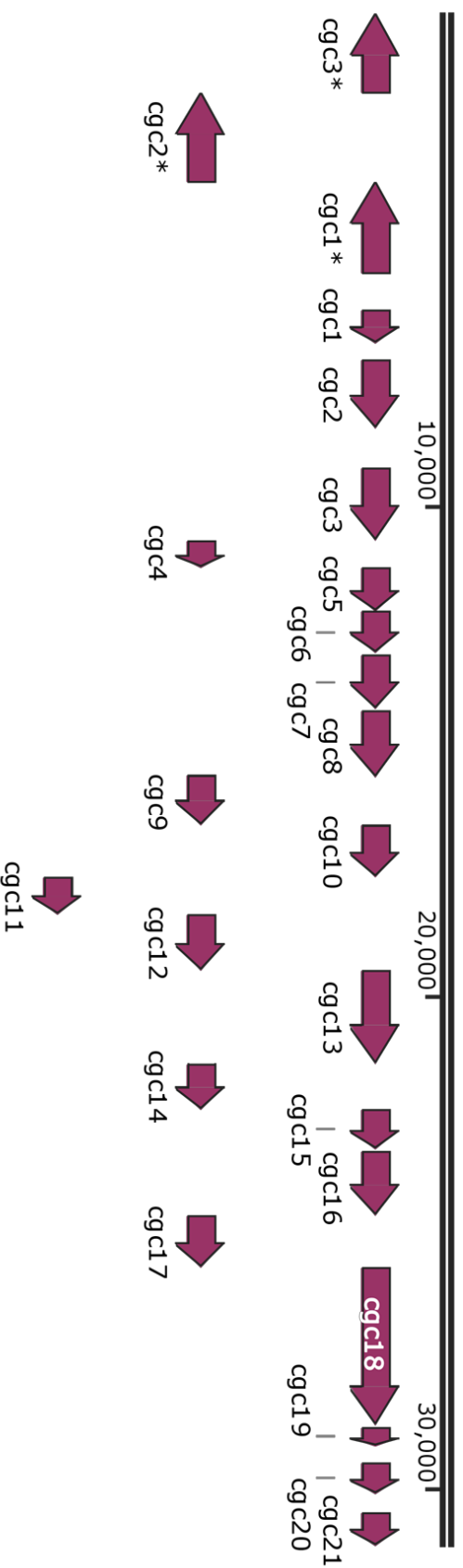


Figure 21. Annotated diagram representing the minimal set of genes required for biosynthesis of congocidine.

The DNA is found on *Streptomyces ambofaciens* and the purple features refer to the original annotation of the coding DNA sequences (Juguet *et al.*, 2009). In the named study, the authors cloned the genes in a cosmid, pCG002, proving this are the minimal set of genes coding for the congocidine pathway, as congocidine heterologous biosynthesis was achieved.

Yeast recombination based assembly.

Transformation-associated recombination (TAR) cloning, was first described by Burke and colleagues (Burke *et al.*, 1985), and was later refined by removing the need for *in vitro* enzymatic reactions (Larionov *et al.*, 1996). Yeast based cloning methods were widely used for sub-cloning of large DNA fragments, (specially from human sources) specially at the final stages of the human genome sequencing project (Raymond *et al.*, 2002). However, the use of this technology for the assembly of complex bacterial DNA clusters, was reported six years later (Shao *et al.*, 2009). At that time, only one report of the reconstruction of a streptomycete biosynthetic cluster with yeast recombineering technology was published (Zhao and Shao, 2009). This DNA manipulation technique for the assembly of large autonomously replicating vectors is based on native recombinatorial activity in yeast (Kunesf-*et al.*, 1985) and requires homology between the assembled fragments (Hua *et al.*, 1997). Based on the information provided on those studies, the initial experimental design had a homology between pathway DNA fragments of 350-670 bp (overlaps used in the only report of *Streptomyces* pathway refactor were 450 bp) and 40 bp between vectors (following the methodology of (Zhao and Shao, 2009)). The number of fragments and particular overlaps were determined within the particular constraints of the congocidine cluster. *In silico* sequence analysis of the congocidine biosynthetic gene cluster sequence reveals highly repetitive GC rich DNA sequences, which can be difficult to amplify. This precludes the use of short oligonucleotides which would anneal to multiple locations within the sequence, however longer, more specific oligonucleotides were difficult to design due to self-annealing secondary structures. When taking all of these factors into consideration, a strategy could be designed to divide the pathway into modules (DNA Fragments) which span several biosynthetic genes (i.e. one part coding for resistance, one part coding for acyl carrier enzymes, etc). Eventually the DNA fragments were designed to avoid high melting temperature regions, to facilitate PCR amplification. The first assembly design strategy was comprised of six fragments, obtained by polymerase chain reaction as shown in Fig. 22.

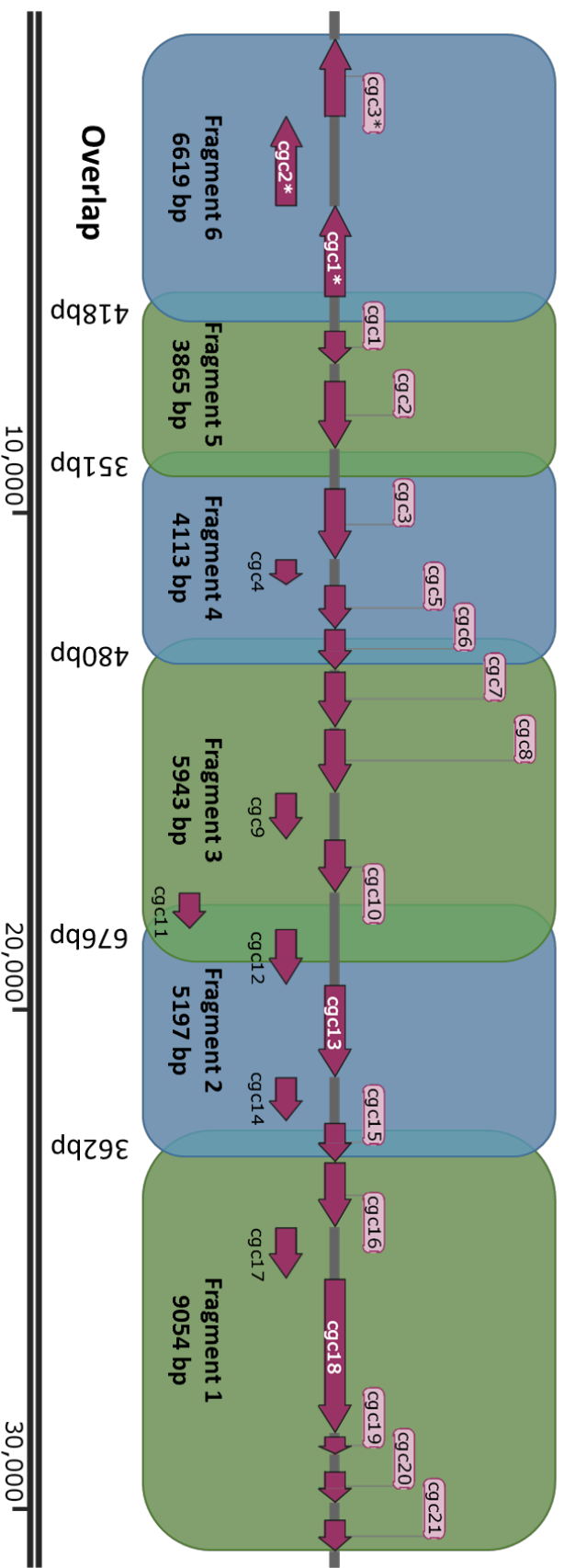


Figure 22. Diagram of the congocidine DNA biosynthetic cluster divided into overlapping fragments.

Purple arrows represent the coding DNA sequence in the congocidine biosynthetic cluster which are embedded into coloured rectangles symbolizing the PCR amplicons; the alternate colour is to portray the end and beginning of the different fragments, which are also represented by their name and size at the bottom of each feature. Overlaps between the amplicons are represented by superposition of the rectangles, which have indicated the overlap numerically below.

Amplification of the high GC content DNA fragments from *Streptomyces* by PCR was found to be challenging. Amplification of the fragments were initially attempted using Q5 polymerase (New England Biolabs). After initial failure to obtain discrete bands of the correct size for each PCR amplicon, several PCR enhancing additives were used, including Betaine (Henke *et al.*, 1997), DMSO alone and DMSO with Betaine (Kang *et al.*, 2005), dimethyl formamide (Chakrabarti and Schutt, 2001), 1,2-propanediol and ethylene glycol (Zhang *et al.*, 2009). However, little to no change in performance was observed in the PCR reactions (data not shown). Following reagent optimisation, iterative thermo-cycling conditions (changing annealing temperatures and extension times) were used in addition to the use of different polymerases (myTaq, One Taq and Q5 Hot start, Phusion) was performed. Using Taq98 polymerase (Lucigen) fragments two, three, four and five were successfully amplified (Fig. 23), while fragments one and six still proved to be difficult to amplify.

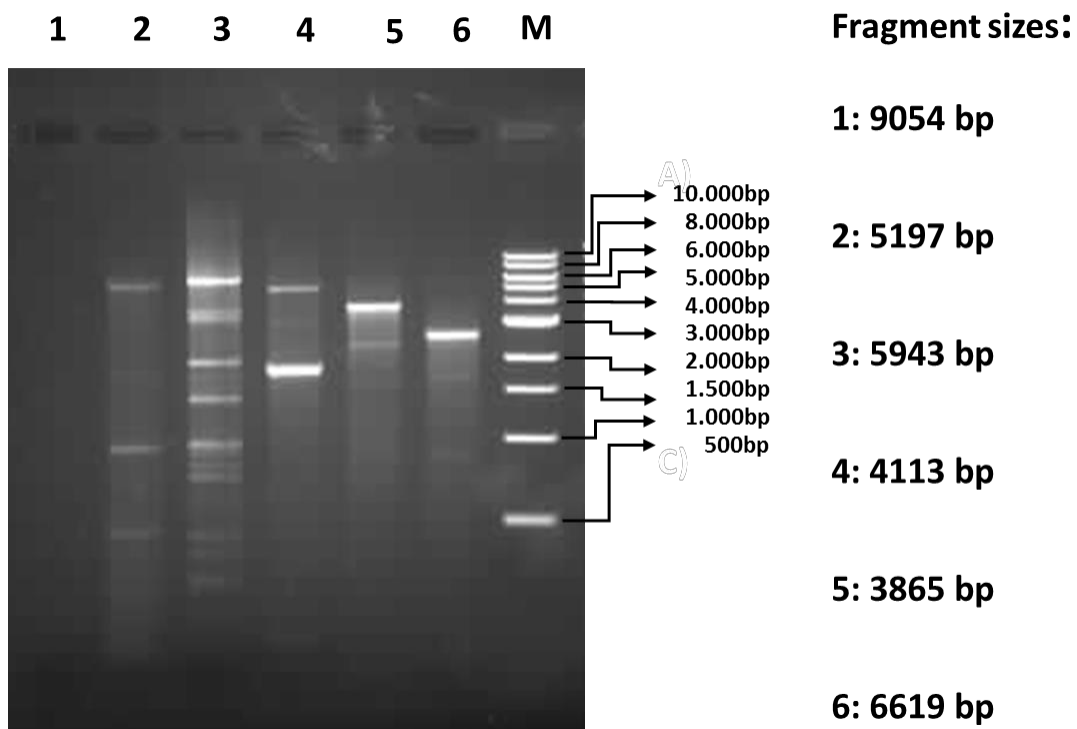


Figure 23. PCR-amplified DNA (Taq 98, Lucigen) for the use in the assemble of the congocidine biosynthetic cluster

Agarose gel of PCR amplicons as design in **Fig. 23** shows, for use in yeast recombination assembly experiment. Fragments two, three, four and five showing the correct size, lanes one and six on the contrary do not match the expected size.

The two fragments that were not obtained (one and six), were significantly larger (9 Kbp and 6.6 Kbp) than the successfully amplified PCR fragments (3.8-5.9 Kbp). To discard a size limitation, it was decided to split each amplicon into two subparts (Fig. 24). Each of these smaller parts was amplified successfully using Taq98 polymerase, based on the experience gained in the previous optimisation process.

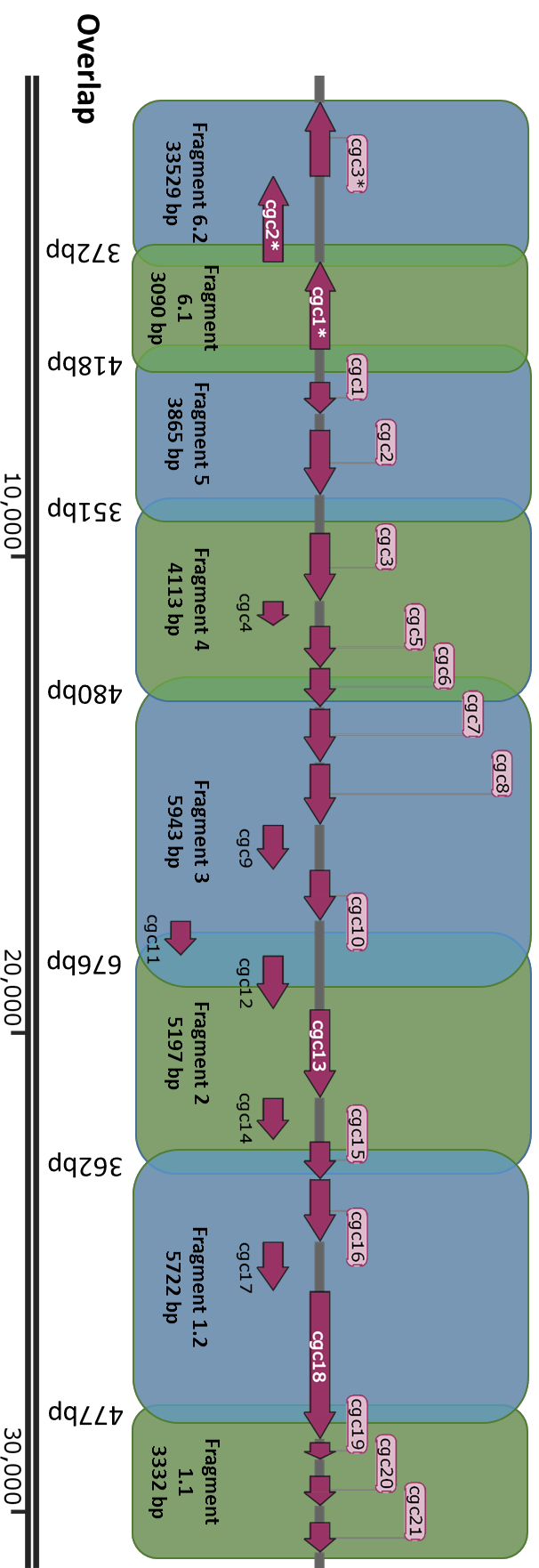
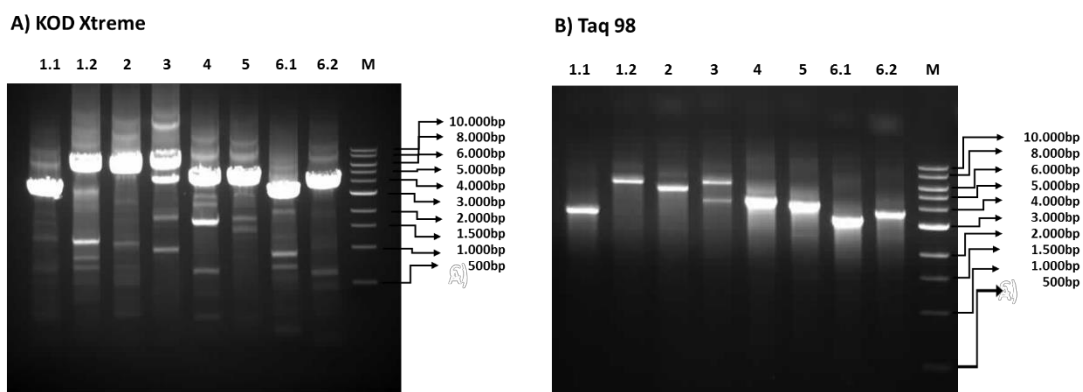


Figure 24. Subdivision of parts one and six of the DNA fragments to be amplified by PCR from the congoicidine biosynthetic cluster.

Purple arrows represent the coding DNA sequence in the congoicidine biosynthetic cluster which are embedded into coloured rectangles representing the PCR amplicons (fragments one and six in the previous design are now split into two new amplicons each); the alternate colour is to illustrate the end and beginning of the different fragments, which are also represented by their name and size at the bottom of each feature. Overlap between the amplicons is represented by superposition of the rectangles, which have indicated the overlap numerically below.

Due to the large amount of DNA required for the assembly experiments and with a view to increase the yield of amplicons, the use of KOD XTreme polymerase (Toboyo) was used for amplifying large fragments. Whilst it was impossible to eliminate the formation of non-specific amplicons using this polymerase, despite attempts to optimise additives (DMSO etc), annealing and extension temperatures, the yield of product per unit of polymerase was higher with the same number of cycles than observed for the other polymerases used in the study, allowing for better cost per assembly and lower mutation rate. The amplification of all fragments and the comparison of KOD XTreme with Taq98 polymerase is shown in Fig. 25, showing increase of amplicon yield by the KOD XTreme, even when presenting more unspecific amplicons.



Fragment sizes:

1.1: 3332 bp	2: 5197 bp	4: 4113 bp	6.1: 3090 bp
1.2: 5722 bp	3: 5943 bp	5: 3865 bp	6.2: 3529 bp

Figure 25. Congocidine biosynthetic cluster DNA parts PCR-amplified to perform its assembly.

Agarose gels showing the amplicons designed for the assembly of the congocidine pathway in yeast by recombination. A) KOD XTreme polymerase. B) Taq98 polymerase.

Once the correct amplification conditions were established for all gene fragments comprising the congocidine cluster, the PCR amplification of each fragment was scaled up from 25 μ L to 500 μ L reactions (no more than 50 μ L per tube) and run on agarose gels to purify the desired bands in larger quantities (an average of 3 μ g per amplicon). Each DNA band was purified to a 260/280 ratio of 1.75-1.83 and the correct sequence was confirmed by sequencing 500 bp at each end of the parts (Eurofins genomics).

Based on the experimental design of (Zhao and Shao, 2009), the yeast recombination experiments were designed to create constructs that could undergo recombination in yeast. Consequently the inclusion of a yeast origin of replication and suitable yeast selection markers were needed. The vector was also constructed with a view to verifying the construct in *E. coli*, and required markers and a suitable origin of replication. Finally, the vector was delivered to the chromosome of *Streptomyces* via bacteriophage integration through intergeneric conjugation, which required the inclusion of an origin of transfer. Yeast recombination experiments were set-up as shown in Fig. 26. The *Streptomyces* vector pIJ6902 (Huang *et al.*, 2005) contained an apramycin resistance cassette, phiC31 integrase and *attB* site and pRS416 (Sikorski *et al.* 1989; yeast centromere, URA3 marker, ampicillin resistance and an MCS). Both vectors were linearised with *EcoRI/NdeI* and *XbaI/HindIII* enzymes respectively. To ensure high quality material for recombination all vector DNA fragments were dephosphorylated and gel purified. To allow joining between the vectors and the PCR amplicons generated, there was a requirement to introduce 40 bp of homology. Based in the idea of linker mediated recombination (Raymond *et al.*, 2002), we designed three linking or "stitching" 80 bp oligomers. One introducing homology between the two different cloning vectors (STVECT), and two more (START, STEND) were designed to facilitate the stitching of the vectors and the biosynthetic gene fragments. The START oligo introduced a 40 bp homologous region between pIJ6902 and fragment 1.1 (see Fig. 25), and STEND introduced 40 bp homology to pRS416 and fragment 6.2 (See Fig. 25). Once ready, the DNA components of the assembly were mixed according to the conditions specified by (Zhao and Shao, 2009) (200 ng of vector DNA and biosynthetic gene fragments and 625 ng for the stitching oligos). The resulting mixture was concentrated to 80 μ L, which was achieved by taking it to dryness in a DNA centrifugal concentrator, followed by resuspension in water. Subsequently the DNA solution was used to transform yeast chemically.

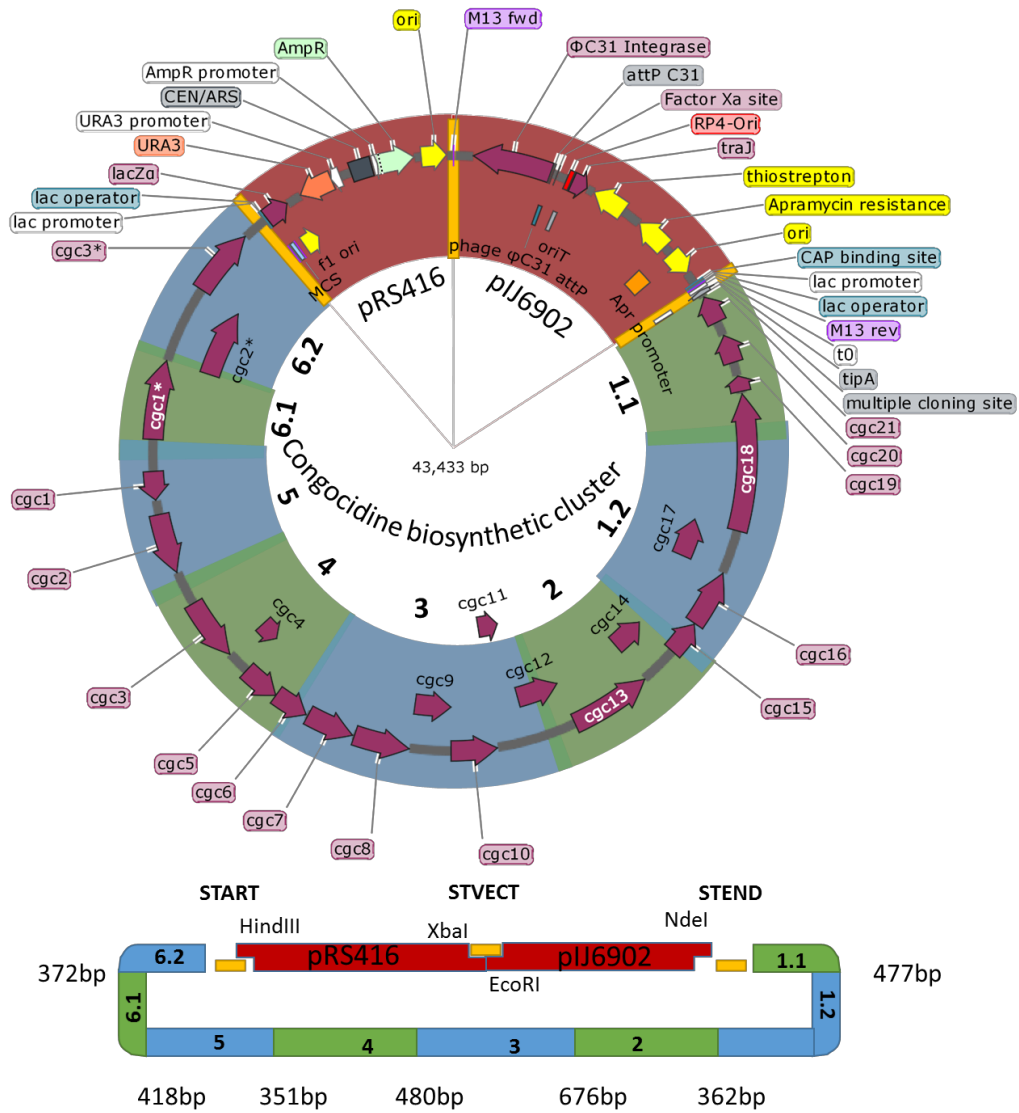


Figure 26. Diagram of the congocidine biosynthetic pathway assembly as designed for yeast recombination.

The *Streptomyces* integrative vector pIJ6902 (red) and the yeast vector pRS416 (red) were linearised with the indicated enzymes. The linking oligomers (yellow) START and STEND stitching the vectors to fragment 1.1 and 6.2 respectively by inclusion of 40 bp of homology. Another oligomer (yellow), STVECT, introduces 40bp homology between both vectors. Green and blue blocks represent the DNA fragments that contains the DNA from the cluster. The overlap between fragments is represented by graphic superposition (numeric bp overlap indicated in diagram below).

The negative control for the assembly was the transformation mixture with water (to check for background yeast growth) and the positive control the same mixture containing a total of 200 ng of uncut pRS416 (to check transformation efficiency). The yeast cells were then transformed and plated into synthetic dropout media complemented with leucine, valine, isoleucine, histidine and lysine. An additional positive control was set up by plating the negative control in the same media with additional URA (this would determine viability of cells). After four days, the plates and cells counted with a haemocytometer (Table 12).

Table 12. Cell count in yeast assembly.

The number of cells (averaged from a total of 6 counts in a haemocytometer) per millilitre in each of the four samples, and the efficiency of transformation for the experimental samples and positive control.

Experiment	Cells per mL (from Haemocytometer count n=6)	Efficiency
Water control	2.50×10^6	-
Water control in URA plates	1.64×10^8	-
congocidine assembly	2.14×10^7	11%
pRS416 control	1.56×10^7	8%

To establish if the yeast cells in the congocidine pathway recombination experiment had successfully assembled the correct construct, the yeast cells were harvested and grown in synthetic media until a total cell count of 5×10^9 was achieved. At this point the plasmid DNA was extracted using a Qiagen QIAprep Spin miniprep kit. This kit allows purification of constructs up to 50,000 bp in size. The procedure yielded 65 μ L of DNA at 500 ng/ μ L and 280/260 ratio of 1.80 as measured by NanoDrop spectrophotometer. To check that the DNA isolated corresponded with the correct assembly of the congocidine biosynthetic cluster, a restriction digest with *Bam*HI was performed along with a PCR to check for the DNA fragment 1.1 that was used in the assembly. The restriction digest mixture was electrophoresed on a 1% agarose gel (**Fig. 27 A**) along with undigested DNA, (**Fig. 27 B**). The gel showed no defined bands in any of the lanes of the restriction digest gel, suggesting DNA may have been degraded. The PCR analysis with primers 1.1F and 1.1R however, showed presence of

fragment 1.1, (Fig. 27 C). This however may indicate either the presence of the correct assembly or carry-over DNA.

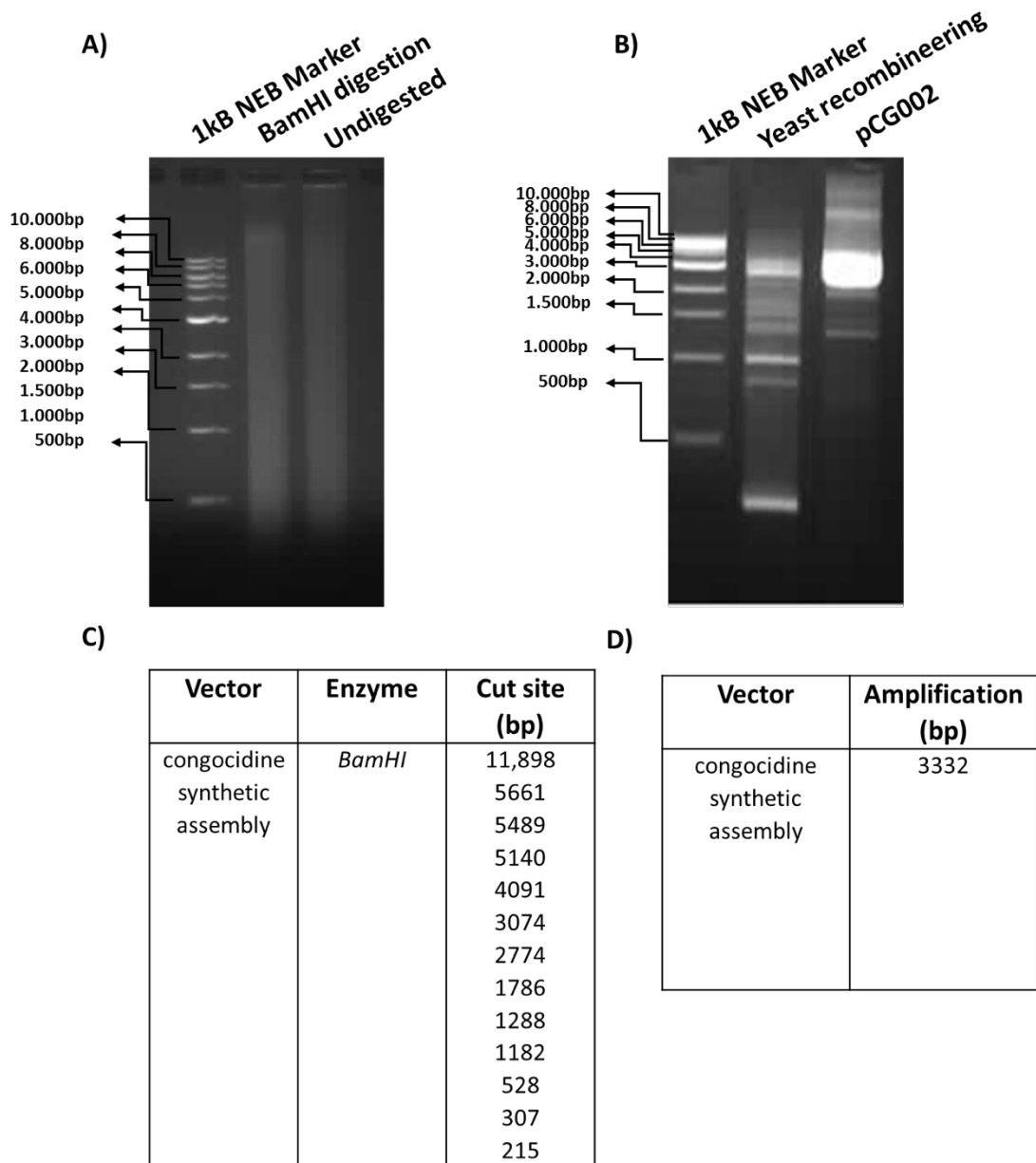


Figure 27. Restriction digest with *Bam*HI and PCR analysis to check for successfully assembly of the congocidine biosynthetic cluster

A) Restriction digest of the congocidine biosynthetic gene cluster with *Bam*HI with DNA extracted from yeast (lane 1) and un-digested DNA (lane 2). B) PCR amplification, using 1.1F and 1.1R primers to amplify DNA isolated from the yeast in lane 1 or pCG002 in lane 2 used as a control.

Although the DNA extracted from yeast looked degraded (Fig.27A), because of the amplification of fragment 1.1 by PCR (Fig.27B), *E.coli* DH5 α was transformed both by chemical transformation and electroporation with it. By doing so, the cells containing at least the two vector assembled should survive. Controls were performed with water (negative) and 500 ng of pIJ6902 (positive). After three attempts of recovery on SOC plates containing apramycin, the *E. coli* cells transformed with the DNA from yeast (500ng) never grew, indicating absence of the apramycin marker.

Other homology based methods for synthetic assembly of large DNA fragments such as Gibson cloning (Gibson *et al.*, 2009), have reported that efficiency decreases with the number of fragments in the assembly reaction (Kok *et al.*, 2014). Therefore, we attempted to redesign the experiment with a lower number of amplicons.

Reducing the number of amplicons, implied the amplification of bigger ones. KOD Xtreme polymerase allowed the amplification of the congocidine biosynthetic gene cluster in three DNA parts. Part A (size 8577 bp) using primers 1.1f/1.2r, B (size 15095 bp) using primers 2f/4r and C (size 9636 bp) using primers 5f/6.2r. The new amplicons had 353 bp overlaps between fragment A and B, and 352 bp overlap between fragment B and C. This enabled the removal of the stitching oligonucleotides from the design. Homology between these fragments and the vectors was introduced by PCR amplification of the vectors with primers -Pij6902F/Pij6902R and pRS416F/pRS416R- (20 bp between both cloning vectors and 40bp between those and the biosynthetic cluster parts). A diagram with the assembly design is shown in Fig. 28. Agarose gels of the PCR amplicons of the newly designed fragments are also shown in Fig. 28.

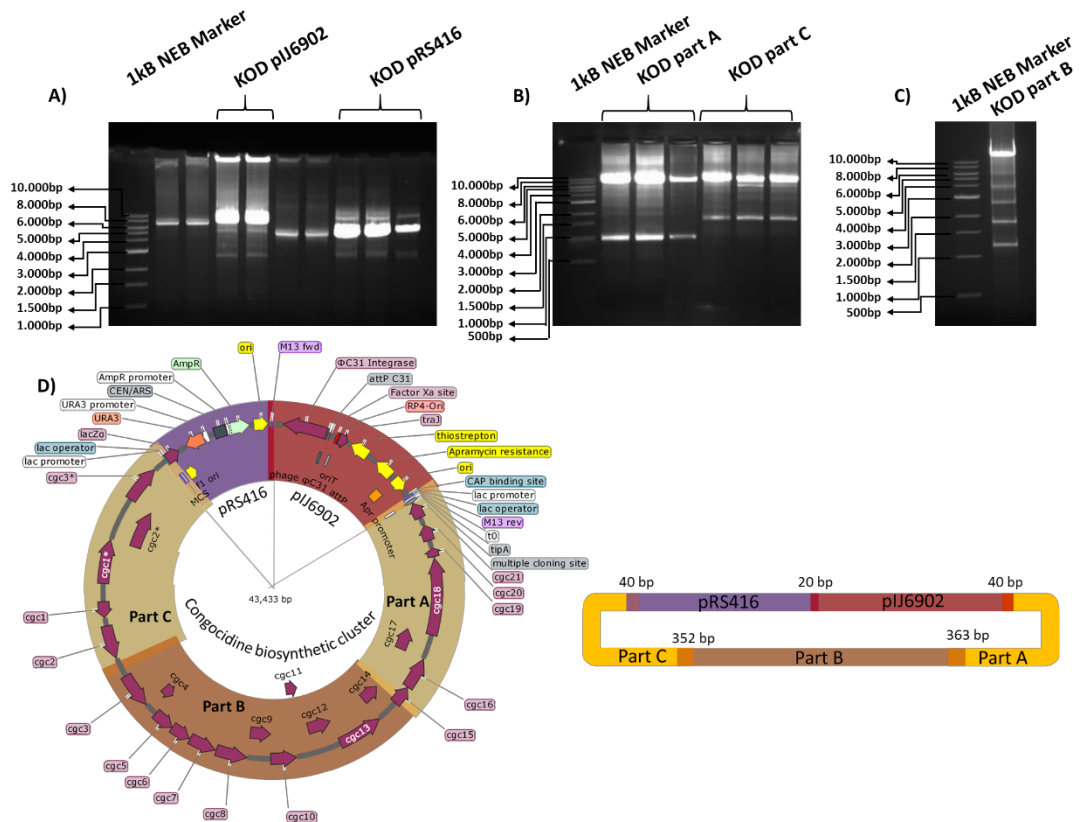


Figure 28. Alternative strategy for congenidine biosynthetic gene cluster assembly in yeast.

Three agarose gels showing DNA fragments, A) Cloning vectors pIJ6902 and pRS416 amplified with KOD Xtreme and Q5 polymerases B) 'Part' A (size 8577 bp), B (size 15095 bp) amplified with KOD Xtreme polymerase, C) Part C (size 9636 bp) amplified with KOD Xtreme polymerase and D) diagram providing an overview of the cloning strategy, stating the overlaps and the features contained in each DNA fragment in the assembly.

Amplicons were gel purified and verified by sequencing of the 5' and 3' ends before proceeding to the recombination assay. In this assembly method, the DNA added to the recombination mix was calculated in picomols rather than in nanograms of DNA; this was carried out to keep the vector-fragment ratio constant, due to molecular size differences in the fragments (significantly larger than the initial amplicons and those described by Zhao and collaborators (Zhao and Shao, 2009) (4-6 kb vs 10-15 kb)). The initial DNA amount was set to 0.35 pmol of each of the three fragments and vector which was a direct translation from the amounts employed in the previous attempt with smaller parts. Controls were set up as previously stated (water for negative, 500ng pRS416 as positive). Unfortunately, none of the three yeast strains used in the assay (BY4741, BY4742, BY4743) grew when transformed with uncut pRS416, either when chemically transformed or by using electroporation. After several iterations it was concluded that cells were not viable after a thawing cycle after being stored at -80°C, since the competency was non-existent. From this point onwards, all transformations carried out in yeast, were done with freshly made competent yeast cells. Subsequent recombination assays were also unsuccessful. To combat this, amounts of DNA in each synthetic DNA assembly reaction were varied in ranging from 0.1-0.8 pmol per reaction per fragment. Nevertheless, no successful assembly was achieved.

Reviewing previous designs and strategies it was hypothesised that the weakest part of the strategy was the homology between the cloning vectors. To address this, a new shuttle vector was created that allowed transfer between yeast-*E.coli-Streptomyces* and reduced the number of fragments in the assembly by one. The shuttle vector was constructed by restriction digest of pIJ6902 and pRS416 with *EcoRI* and *XbaI* yielding DNA fragments of 7313bp and 5055bp respectively. These two fragments were ligated together to create pCS-RS31 (Fig. 29), of which the restriction digest pattern is shown in Fig.30.

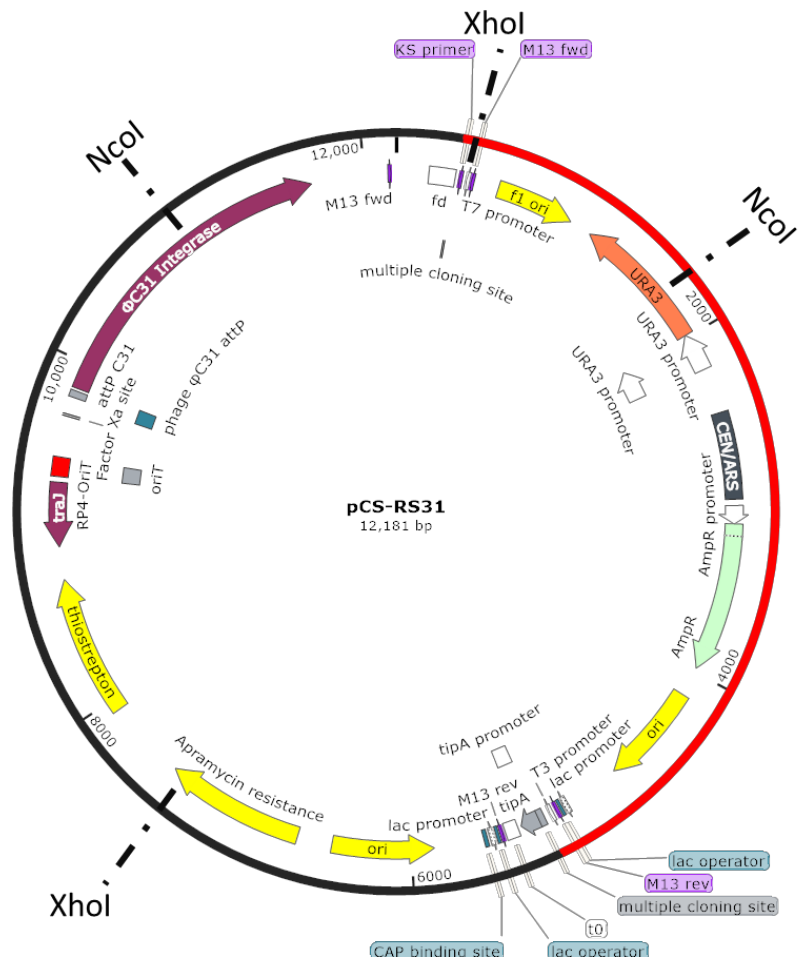
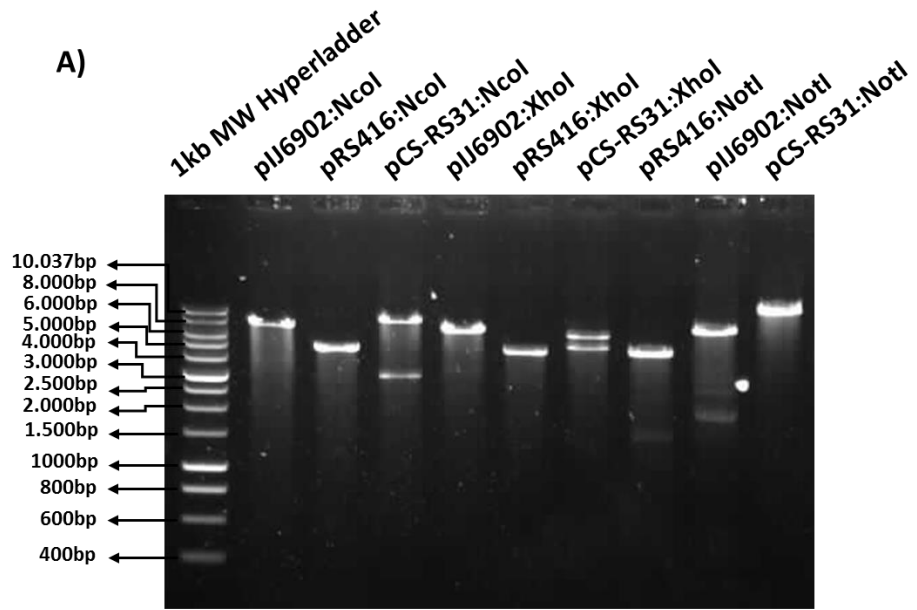


Figure 29. Plasmid map of pCS-RS31

Yeast-*Ecoli-Streptomyces* shuttle vector used in yeast assembly experiments, derivate from two plasmids (pIJ6902, pRS416) after linearisation with *EcoRI* and *XbaI* and subsequent ligation between them: the pRS416 part of the plasmid is highlighted in red and the pIJ6902 part in black. Restriction digest *NcoI* and *XhoI* sites are represented in the map as they were used to make a restriction digest to check the assembly was correct, which can be seen in Fig.30.



B)

Vector	Enzyme	Cut site (bp)
pIJ6902	<i>NcoI</i>	7340
pRS416	<i>NcoI</i>	4898
pCS-RS31	<i>NcoI</i>	9179,3002
pIJ6902	<i>XhoI</i>	7340
pRS416	<i>XhoI</i>	4898
pCS-RS31	<i>XhoI</i>	6861, 5320
pRS416	<i>NotI</i>	4898
pIJ6902	<i>NotI</i>	Non cutter
pCS-RS31	<i>NotI</i>	12181

Figure 30. Restriction digest of plasmid pIJ6902, pRS416 and pCS-RS31.

A) Agarose gel showing the restriction digest of plasmid pIJ6902, pRS416 and pCS-RS31 with enzymes *NcoI*, *XhoI* and *NotI* (represented in the plasmid map of pCS-RS31 in Fig.29). B) Expected sizes of the fragments arising after cutting plasmids pIJ6902, pRS416 and pCS-RS31 with enzymes *NcoI*, *XhoI* and *NotI*.

This new shuttle vector (pCS-RS301) was used to attempt the cloning of another design of the synthetic version of the congocidine biosynthetic cluster. The amplicons A, B and C were obtained as described before. The homology to the new vector was added by PCR with two 100 bp oligomers (RSTF, RSTR) that would allow 20 bp annealing to the vector for amplification of an 11.261 bp linear fragment and addition of 80 bp of homology to parts A-C. The design of the assembly as well as the agarose gel showing the amplification of pCS-RS31 to a linearised "tailored" fragment is shown in Fig. 31 A), 31 B). The recombination mixture was made with amplified A, B, C fragments and PCR amplified pCS-RS31 at 0.35 pmol/fragment/ μ L. The mix was used to transform freshly make chemically competent yeast. This approach yielded no colonies, however the positive control (500 ng of pRS416) produced a lawn of cells, which led to think that chemical transformation of fragments larger than pRS416 were not efficient. To address this, new prepared mixture at 0.1 pmol/fragment/ μ L and freshly made electrocompetent yeast cells were transformed. This approach produced no colonies in the negative control and again a lawn of cells in the positive control (positive control 2 μ L of 250ng/ μ L pRS416, negative control 2 μ L of water). Yeast cells grew in the plates containing the yeast electroporated with the genetic construct, (1 μ L transformation mix, n=17; 3 μ L transformation mix, n=confluent grow; 5 μ L transformation mix, n=confluent grow) although the cell count was too low to extract plasmid DNA. The three plates (1,3,5 μ L transformation mix used) were flooded with 1mL YEPD medium and grown until the approx. cell count reached 1×10^9 cells, at which point they were harvested and spheroplasted to extract plasmid DNA as previously described. This DNA was used to transform *E.coli*, which was plated on to SOC media plates selecting for ampicillin (resistance on pCS-RS31) with 100 μ g. mL⁻¹ carbenicillin. No colonies were recovered however, and the same DNA was used to transform *E.coli*, this time with selection using apramycin 50 μ g. mL⁻¹, however no colonies were obtained either, although a faint halo of growth could be seen on the plates. To analyse the reaction mixture, analysis by PCR of the DNA extracted from the yeast cells was undertaken (Fig. 32).

This analysis consisted in the screening of five different elements in the desired construct; the phiC31 integrase (using primers phiC31F and phiC31R), the URA auxotrophic marker (using primers URA F and URA R), and the 3 parts of the assembly (using primers PartACGCF/PartACGCR, PartBCGCF/PartBCGCR and PartCCGCF/PartCCGCR). For a positive control of the vector backbone pCS-RS31 was used as template. For the positive control of the congocidine biosynthetic gene cluster, pCG002 was used as a template. The results

suggested that the assembly was not done as designed. After the failure of the different designs, and without a comprehensive knowledge of the pitfalls of the process, the yeast recombineering approach was abandoned.

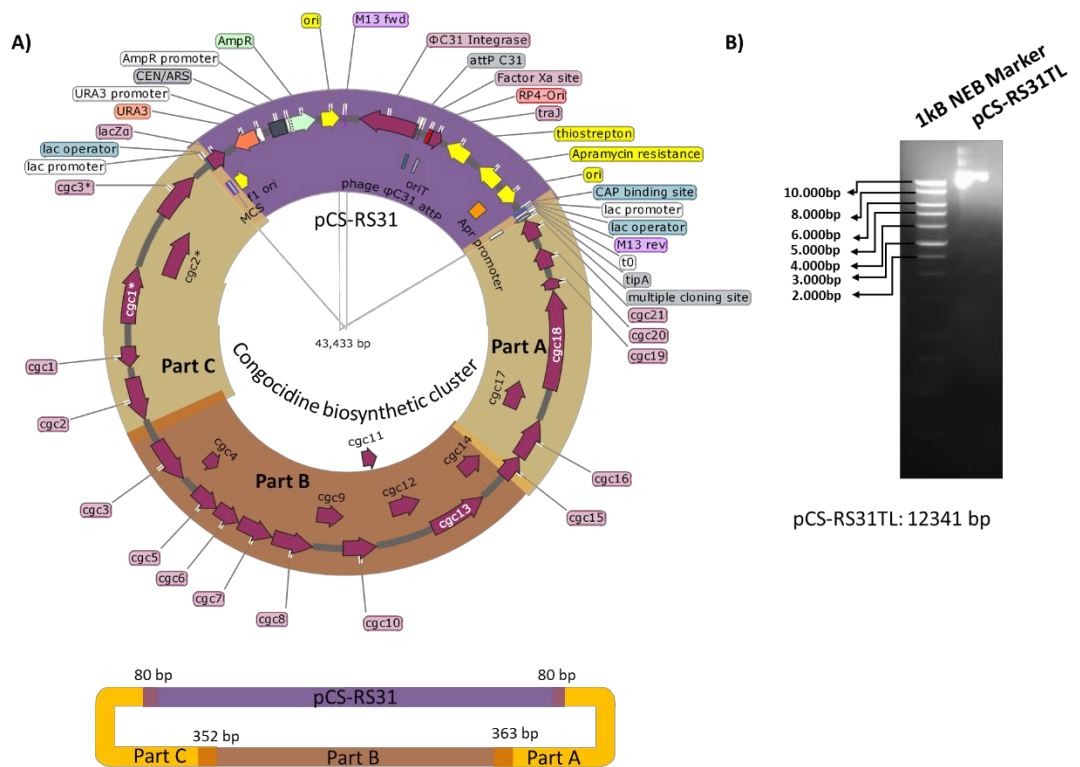


Figure 31. Alternative strategy for congenidine biosynthetic gene cluster assembly in yeast with the new shuttle cloning vector pCS-RS31

A) Design used in the assembly of congenidine pathway by yeast recombination, with pCS-RS31. A, B and C parts as in the previous experiment, but single vector backbone with increased homology.

B) Agarose gel showing the amplification of pCS-RS31 with primers to add homology to the congenidine pathway.

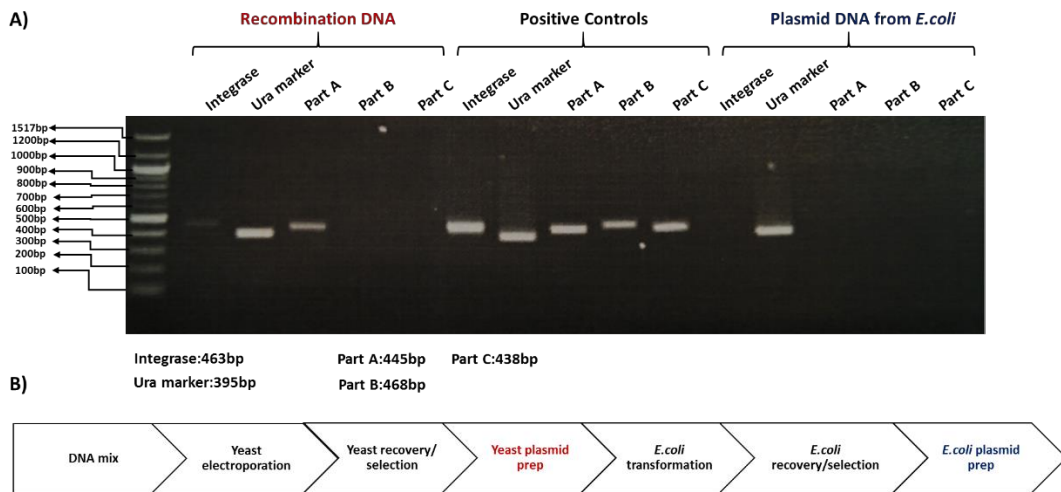


Figure 32. PCR based screening to diagnose correct congocidine assembly

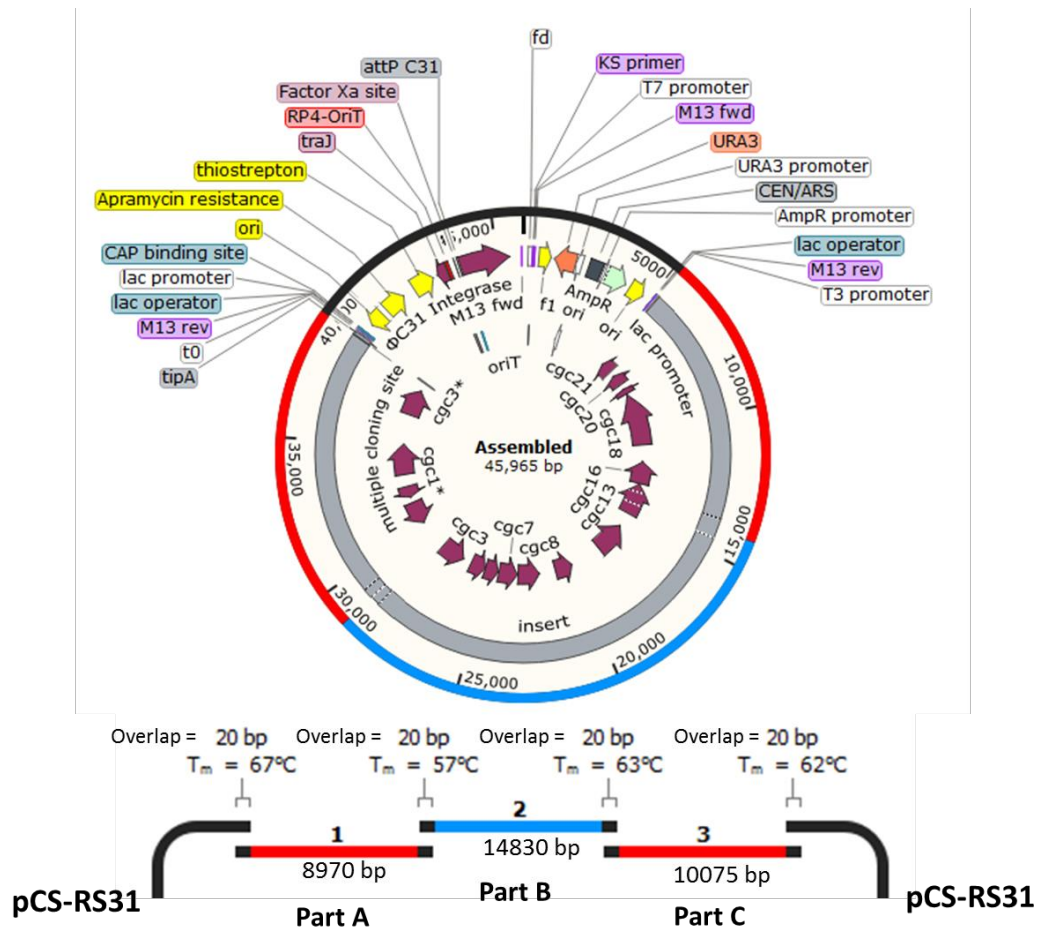
A) Agarose gel with the amplicons of five markers with three different templates (phiC31 integrase (phiC31F/ phiC31R), URA auxotrophic marker (URA F/URA R), and the 3 parts of the assembly (PartACGCF/PartACGCR,PartBCGCF/PartBCGCR,PartCCGCF/PartCCGCR)).

B) Overview of the plasmid isolation, highlighting the plasmid preps used as template for the check in the PCR screen performed in A). In red the plasmid isolated from the yeast assay, in black the positive controls (pCS-RS31 for Integrase and Ura marker, pCG002 for the amplicons A, B, C and in blue the plasmid isolated after transforming *E. coli* cells with the plasmid library obtained from yeast.

Gibson cloning to assemble the congocidine biosynthetic gene cluster.

The assembly of the congocidine pathway was also attempted with the Gibson cloning technology. Due to the requirements of this methodology, such as, overlaps that are no longer than 80 bps (due to the limitation of the exonuclease activity in the reaction), it was necessary to redesign the congocidine biosynthetic gene cluster parts to reduce the overlapping regions to an optimal of 20 bps. Six oligonucleotides were then designed to amplify the cluster in a similar fashion as described for the yeast recombination approach (ACB AF/ACB AR, ACB BF/ACB BR, ACB CF/ACB CR).

These new three fragments were also named A, B and C with sizes 8970 bp, 14830 bp and 10075 bp respectively. These fragments included an overlapping span of 20 bp between each other and to the plasmid backbone, that would be pCS-RS31 restriction digested (*NotI-XbaI*). The use of pCS-RS31 instead of pIJ6902 was justified because the latter could decrease assembly efficiency due to size (small vector comparing with the final assembly). An overview of the strategy is shown in Fig. 33. The DNA fragments were amplified and gel purified as previously described and the vector was digested and gel purified. All parts were mixed with the assembly kit as per manufacturer guidelines (0.35 pmol of each part taken to dryness and mixed with the master mix), including the control reaction provided with the kit. The assembly mixture was incubated at 50°C for 60 min and subsequent transformation of *E.coli* was performed using 1µL of the Gibson reaction of each of the mixtures by electroporation. While the positive (control reaction provided in kit) and negative (cells transformed with water) control reactions were successful, the assembly was not. Several additional attempts were made changing the total DNA content used in the assembly (0.30 pmol parts/0.1 pmol vector, 0.30 parts pmol / 0.2 pmol vector, 0.6 pmol parts/ 0.1 pmol vector), extending the incubation time or by using up to 5µL of the reaction mixture in the electroporation after desalting. However any of these conditions worked.



A)

Figure 33. Diagram showing a Gibson assembly on the congocidine cluster, using the plasmid pCS-RS31 as a backbone.

A) *In silico* assembly of congocidine pathway by Gibson cloning between plasmid backbone pCS-RS31 and three DNA fragments encoding genes of congocidine pathway. Overlaps are illustrated at the bottom.

With a view to producing smaller assemblies as a proof of concept, and to check the compatibility of the Gibson cloning approach with large (>1kb), high GC DNA fragments, the design was simplified and three new assemblies were attempted:

One assembly of three fragments (parts 1.1 and 4 of congocidine cluster, plus vector pCS-RS31) as designed in the yeast assembly but with redesigned primers to adapt the overlapping homologous regions.

One assembly of three DNA fragments including a unique restriction site between them (modules 1.1 and 6.2 of congocidine cluster plus vector pCS-RS31). This construct was used as a tool to re-attempt a yeast recombineering experiment. Modules 1.1 and 6.2 are at the beginning and end of the congocidine biosynthetic cluster, and having cloned them into a shuttle yeast-bacteria vector that incorporates a unique restriction site between the congocidine biosynthetic gene parts would allow the construct to be linearised. This would produce big 3 kb overlaps to each side of the pathway, thus eliminating the weakest point in the design (the vector homology to the congocidine biosynthetic gene cluster parts).

The last assembly was the vector and one module (Fragment 1.1 of the congocidine biosynthetic gene cluster) to simplify the experiment to its minimal point.

Figures illustrating the three assembly strategies are shown below (Fig. 34, Fig. 35, Fig. 36).

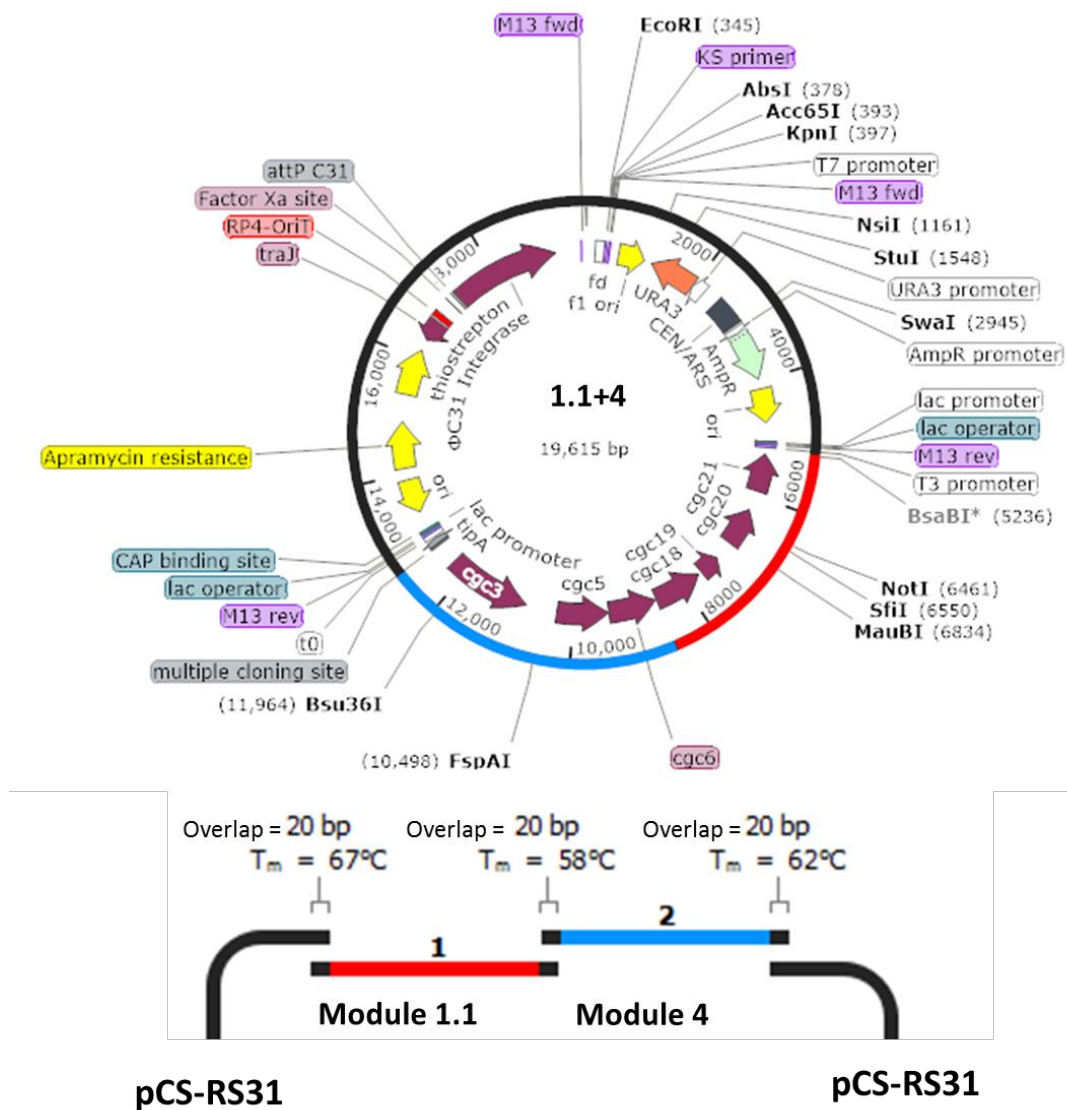


Figure 34. Diagram of a Gibson assembly experiment developed to join the vector pCS-RS31 and the fragments 1.1 and 4 of the congocidine biosynthetic gene cluster.

The modules are shown in red and blue with the pCS-RS31 backbone in black. Overlaps between modules and vector are illustrated in the bottom.

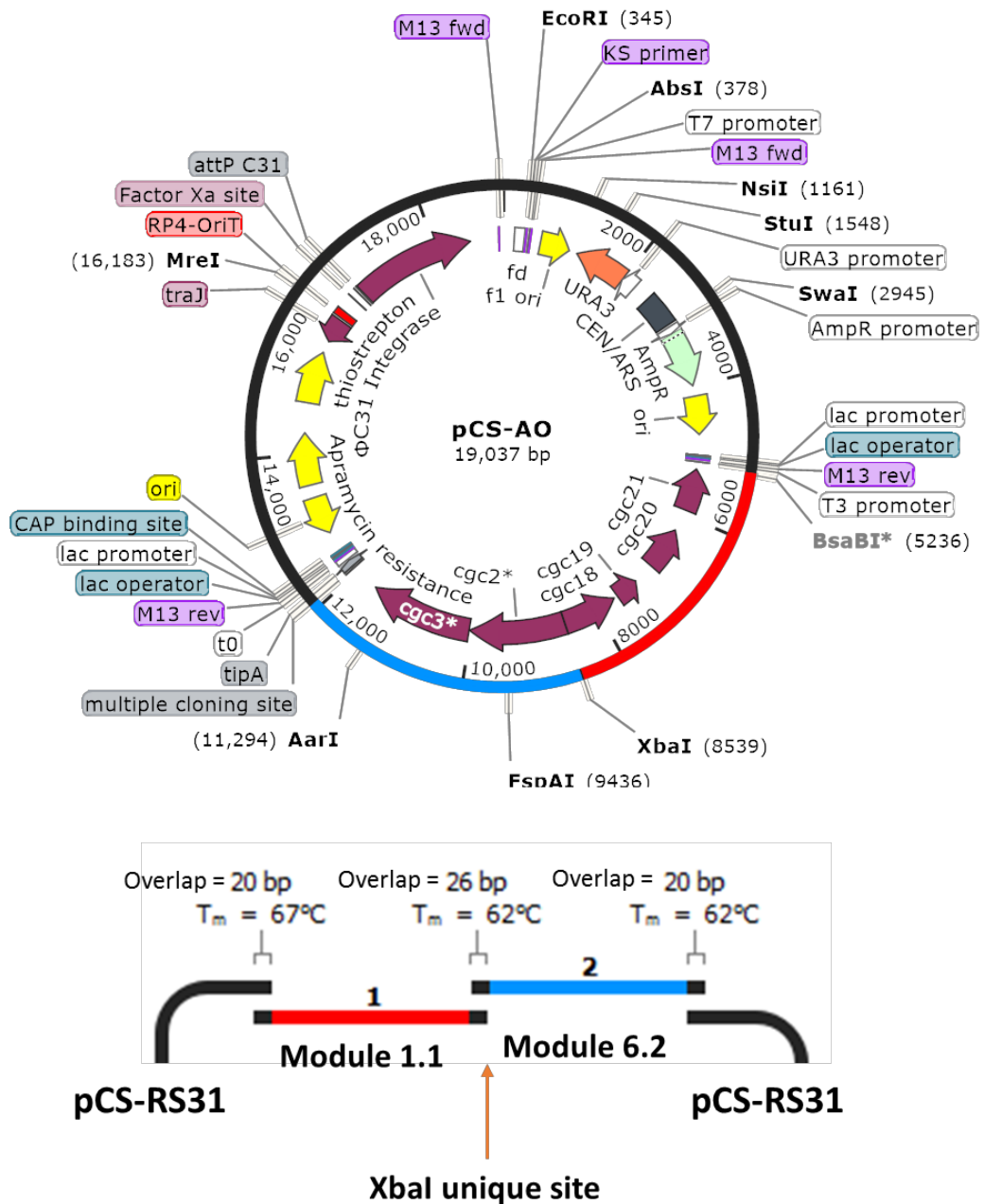


Figure 35. Diagram of a Gibson assembly ideated to join the vector pCS-RS31 and the fragments 1.1 and 6.2 of the congocidine biosynthetic gene cluster.

The parts are shown in red and blue with the pCS-RS31 backbone in black. The modules are shown in red and blue with the pCS-RS31 backbone in black. Overlaps between modules and vector are illustrated in the bottom. Unique restriction digest sites shown in bold.

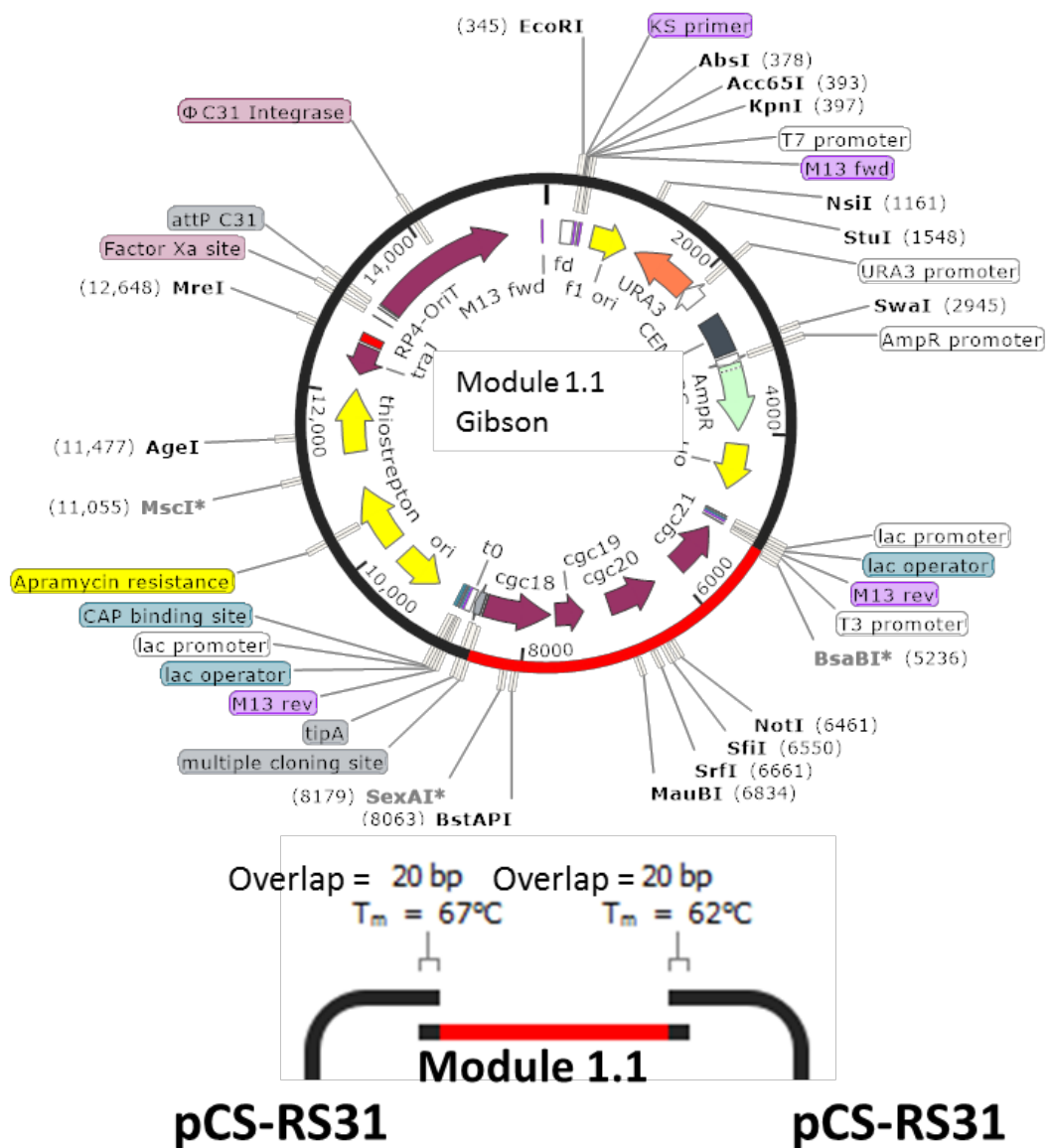


Figure 36.Diagram of a Gibson assembly to join the vector pCS-RS31 and the fragment 1.1 congocidine biosynthetic gene cluster.

The module is displayed in red with the pCS-RS31 backbone in black. Overlaps between module and vector are illustrated in the bottom.

However after several attempts none of the assembly strategies were successful. Analysis of the plasmid DNA by restriction digest, showed either vector circularisation (colonies presenting three bands) or incorrect assembly (unexpected band pattern) (data shown for the assembly pCS-AO (Fig. 37)).

At this point Gibson cloning approach seemed to be inadequate for the assembly of the congocidine biosynthetic gene cluster and it was abandoned to pursue other means for synthetic assembly of biosynthetic gene cluster.

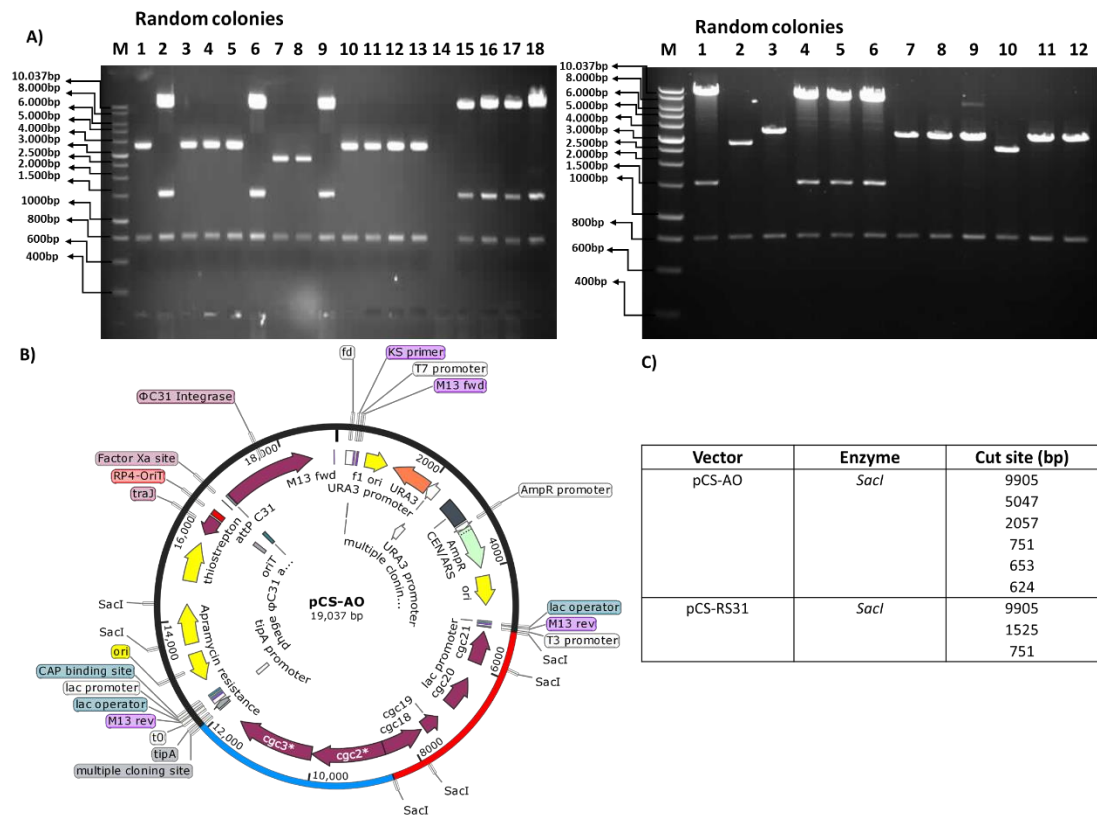


Figure 37. Restriction digest with the restriction digest enzyme *SacI* of plasmid DNA

A) Agarose gels showing the restriction digest pattern of plasmid DNA extracted from colonies formed after transformation with a Gibson cloning assembly of vector pCS-RS31 and fragments 1.1 and 6.2. B) Map of the plasmid pCS-AO (parent vector pCS-RS31 in black, fragments 1.1 and 6.2 in blue and red) and a table indicating the restriction fragments expected after digestion with *SacI*.

Conclusions

One biosynthetic pathway from *Streptomyces* was reconstructed with the use of the yeast assembling techniques described in this chapter before (Shao *et al.*, 2009; Zhao and Shao, 2009). However after using these approaches to the congocidine platform, and exploring alternatives that were reported during the completion of this PhD work (Yamanaka *et al.*, 2014),(Montiel *et al.*, 2015), it was hypothesised that these techniques does not work with all pathways equally, and are highly DNA template-dependant. This impression was discussed with Dr.Huimin Zhao and Dr.Jeff Boeke to further theorise that the use of each technique depends on the particular DNA to assemble, and general guidelines cannot be given at the current state of DNA assembly technology. The best assembling methodology for a particular construct relies hugely on individual customisation, and often the reports in literature are isolated successes. Despite ongoing research, the hypothesis derived from this chapter is that assembly of high G+C content and repetitive templates, such as NRPS containing gene clusters is hard to achieve.

Chapter 3: Mutasynthesis of congocidine

Introduction

Mutasynthetic experiments were first described in work on *Streptomyces fradiae* in the late 1960s (Shier *et al.*, 1969). The aim of these experiments was to produce novel hybrid metabolites derived from the original chemical structures produced by the native organism, but through the incorporation of non-native moieties into the chemical backbone. Synthetic chemistry provides the “building blocks” which resemble the natural precursor in the biosynthetic pathway. Mutations in the original strain are performed to block the synthesis of endogenous precursors in order to maximize uptake of the non-natural precursor, without altering the catalytic enzyme, and/or through modifying the biosynthetic enzymes towards increased affinity for the heterologously fed non-natural precursor (Cortes-sanchez and Hoskisson, 2015). This chapter describes how the previous research deciphering the putative congocidine assembly pathway inspired a strategy to perform mutasynthetic experiments, with a view to produce Congocidine-like compounds (with different hetero-rings and N-substitutions) *in vivo*, pursuing the assembly of molecules with different DNA binding properties.

The congocidine biosynthetic assembly pathway was first proposed when the minimal biosynthetic gene cluster for its synthesis was determined (Juguet *et al.*, 2009). However, not all the steps were clear and most of the putative enzymatic functions were based on bioinformatics analysis of gene homology with other previously characterised biosynthetic enzymes. Subsequent work from the same group described in more detail the chemistry of the pathway (Lautru *et al.*, 2012). Most recently, an unrelated research group proposed an alternative pathway based on enzyme activity assays (Al-Mestarihi *et al.*, 2015). Nevertheless, all three studies conclude that the biosynthesis of 4-acetamidopyrrole-2-carboxylate is the direct precursor of congocidine. This was confirmed by the work of Sylvie

Lautru and colleagues, (Lautru *et al.*, 2012) which showed that a *cgc10* null mutant in the congocidine biosynthetic cluster was unable to make congocidine but exogenous supplementation of 4-acetamidopyrrole-2-carboxylate could restore production in the mutant. The gene *cgc10* does not encode for the enzyme catalysing the final step of pyrrole synthesis, but does block production of 4-acetamidopyrrole-2-carboxylate. Moreover, these studies indicate that the pyrrole molecules used in biosynthesis are N-methylated in a final tailoring step by a methyltransferase enzymes following synthesis of the polyamide (Fig. 38).

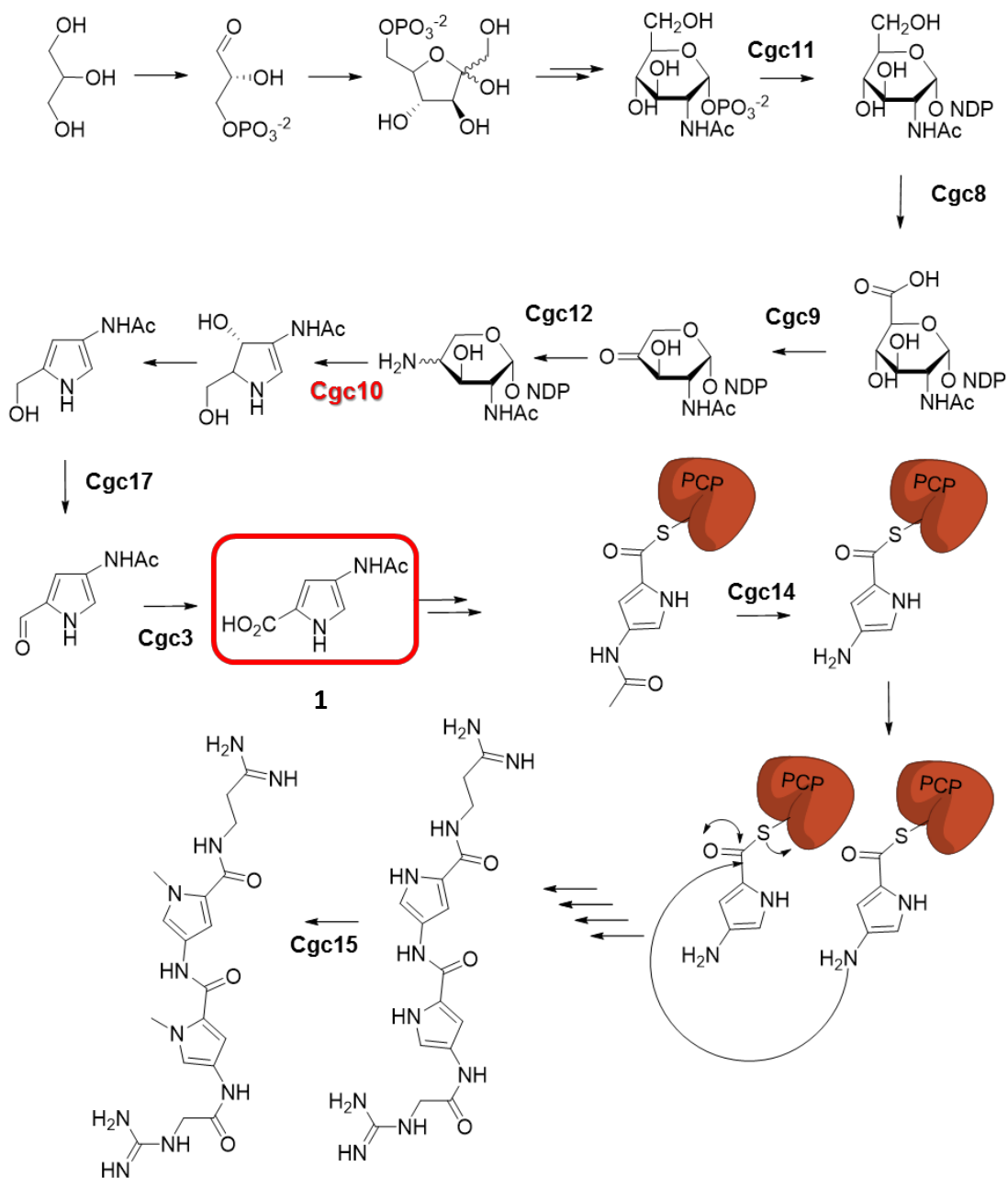


Figure 38. Putative congoicidin assembly pathway

Adapted from (Juguet *et al.*, 2009; Lautru *et al.*, 2012; Al-Mestarihi *et al.*, 2015). Putative biochemical assembly of congoicidin. In bold are the genes encoding for enzymes in which there is a consensus about their biosynthetic role. Highlighted is the 4-acetamidopyrrole-2-carboxylate precursor (red box, compound 1) which synthesis is abolished in the *cgc10* (red, bold) null mutant. PCP = Peptidyl Carrier Protein

Competitive inhibition assay

Given the potential promiscuity of the Non-Ribosomal Peptide Synthetases in biosynthetic pathways (Kim *et al.*, 2015; See also Chapter 1), the possibility of incorporating diverse heterocycles in to the congocidine backbone was explored. A preliminary experiment was performed in order to provide some insight in to the catalytic plasticity of the pathway. The affinity towards substrates other than 4-acetamido-1-methyl-pyrrole-2-carboxylic acid (1) (Fig.38) was assayed by competitive inhibition.

It was hypothesised that feeding of [1H]-imidazole-2-carboxylic acid (2) and 1-methyl-imidazole-2-carboxylic acid (3) (Fig.39) into cultures with the natural producer of congocidine, *S.netropsis* would stop the polymerisation of the growing peptide chain, due to the lack of the acetamido group. Moreover, if these substrates were to bind irreversibly to the enzymes, they would compete with the natural precursor. This would prevent congocidine biosynthesis or proportionally diminish production through non-natural substrates blocking natural substrates.

Cultures of *S. coelicolor* M1146+pCG002 were grown in 50 mL triplicates of MP5 media and the chemicals added at 1 mM final concentration after 48 h, with sampling at 96 h. It is likely that the blockage in biosynthesis occurs because the carboxylic moiety is still able to be thiolated at the pocket of peptidyl carrier proteins, but once there, the lack of the acetamido moiety prevents the condensation reaction proceeding after enzymatic de-acetylation. This mechanism is represented in Fig. 39. In Fig.40, HPLC chromatograms reveal diminished production of congocidine (measured by decrease of congocidine peak area in a UV chromatogram, qualitative data) when 1H-imidazole-5-carboxylic acid (2) and 1-methyl-1H-imidazole-5-carboxylic acid (3) are in culture respectively. The reduction observed in congocidine production when cultivated with 1H-imidazole-2-carboxylic (2) acid (1mM), appeared to be less pronounced than when 1-methyl-1H-imidazole-2-carboxylic acid (3) (1mM) was present in the culture. This qualitative data suggest that the enzyme coded by *cgc18*, which according to (Juguet *et al.*, 2009) catalyses the reaction between two pyrroles, may have a higher affinity for the 1-methyl-1H-imidazole-2-carboxylic acid (3).

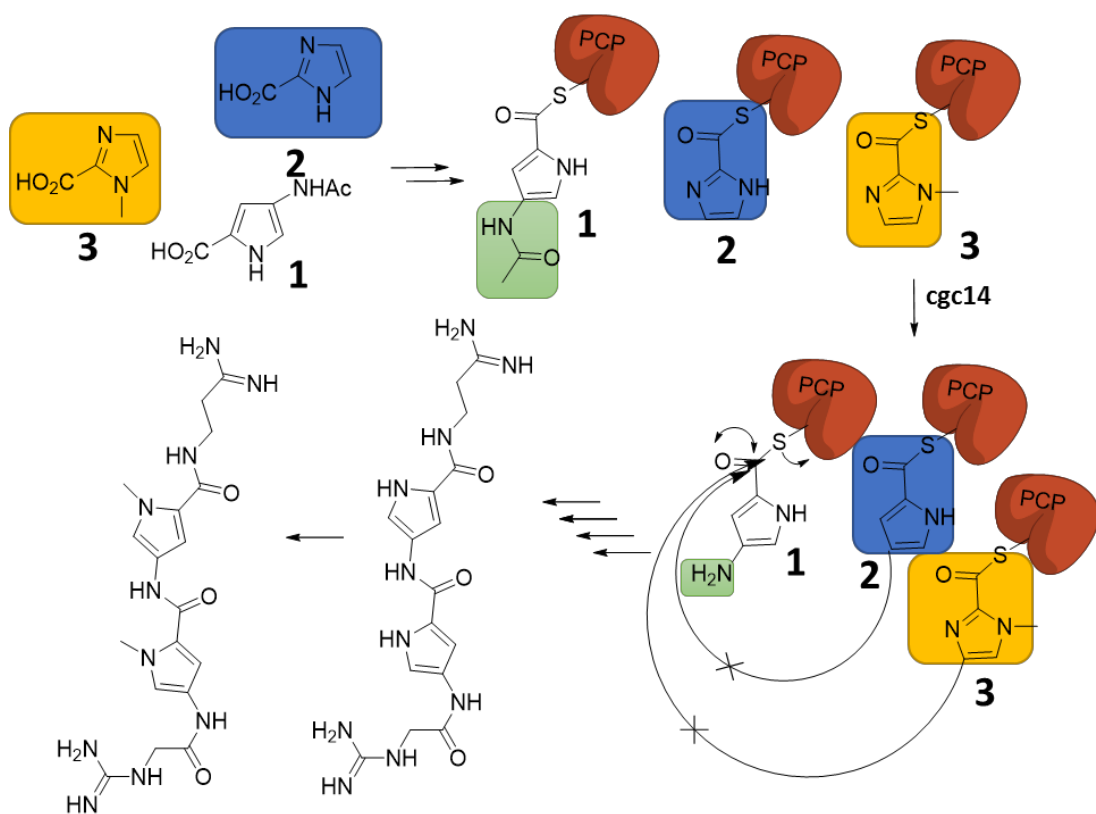
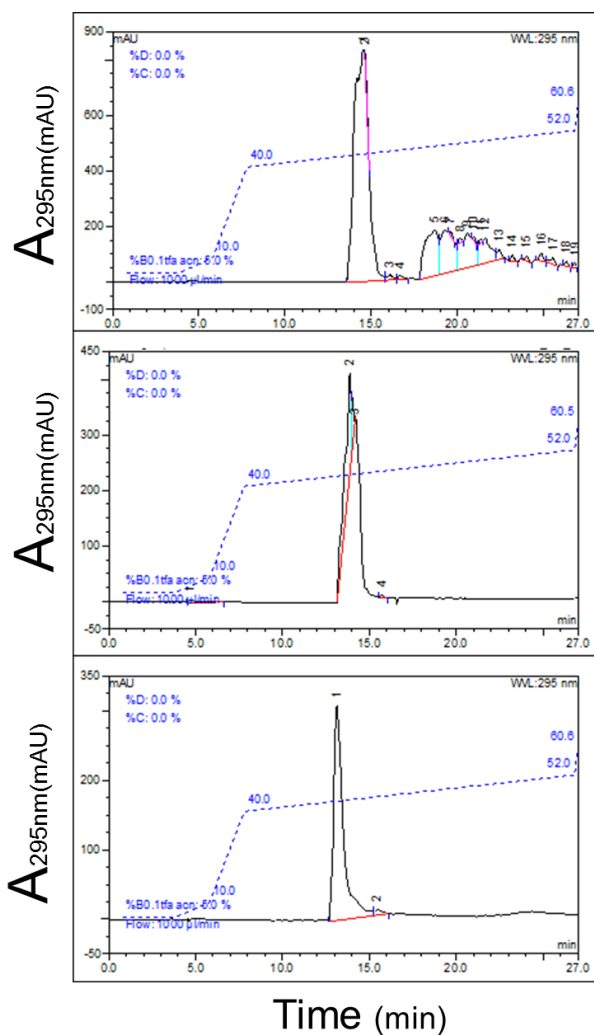


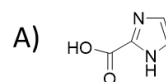
Figure 39. Inhibition assay diagram and chromatograms

The diagram shows how the lacking of the acetamido moiety (in green) on the natural precursor blocks the Peptidyl Carrier Protein (PCP) to continue polymerisation. As the acetamido moiety is missing, the compound cannot be deacetylated, exposing an amino moiety that forms the amide bond.



Carrier solvent fed,
natural level of congocidine production

1*H*-imidazole-2-carboxylic acid **2**



B) 1-methyl-1*H*-imidazole-2-carboxylic acid **3**

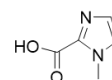


Figure 40. Inhibition assay chromatograms

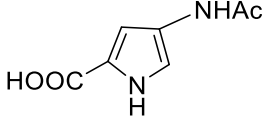
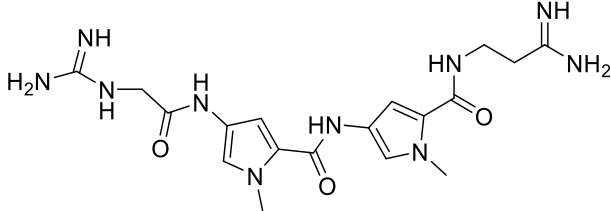
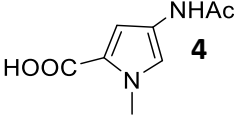
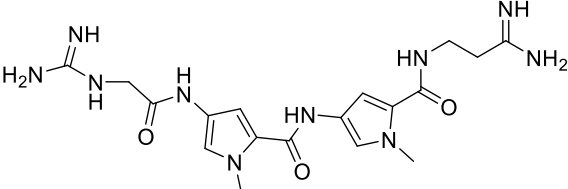
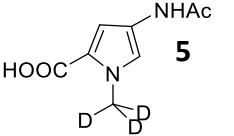
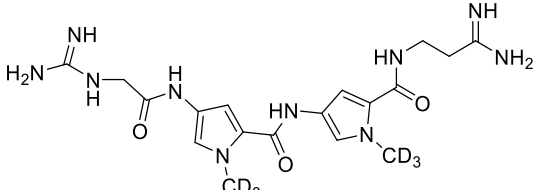
HPLC UV chromatograms (295 nm) that quantitatively show decrease in congocidine detection when feeding the cultures with A) 1*H*-imidazole-5-carboxylic acid (**2**) and B) 1-methyl-imidazole-2-carboxylic acid (**3**), at final concentration of 1 mM.

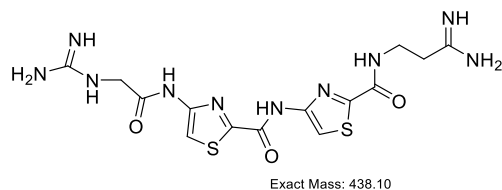
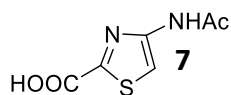
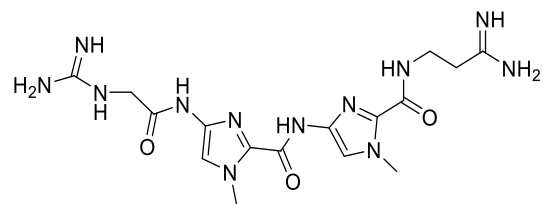
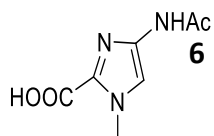
Non-natural substrate incorporation

The inhibition assay allowed insight into the possibility of broad catalytic activity of the congocidine cluster. Sylvie Lautru and colleagues (Lautru *et al.*, 2012), did work on chemical complementation of a null *cgc10* mutant in the cosmid carrying the congocidine biosynthetic cluster to restore congocidine production. Inspired by the complementation assay, the same mutant was made and heterologously expressed in *S. lividans*. Building on this work, it was proposed that an already methylated analogue of the natural precursor used in the 2012 study (1), 4-acetamido-1-methyl-pyrrole-2-carboxylic acid (4), may be a suitable substrate for the enzymatic machinery to incorporate an N-alkylated pyrrole during congocidine biosynthesis. The natural mechanism involves obtaining the methyl group on the congocidine molecule by the action of a SAM methyl transferase as (Juguet *et al.*, 2009). If incorporation of a methyl group was achieved, the possibility of feeding alternative N-alkylated pyrroles in the pyrrole could be explored.

To exclude potential de-methylation of the intermediate in the feeding experiments as a result its uptake or by another processes that may be present in the heterologous host, the cultures were fed a deuterated methyl precursor, 4-acetamido-1-(methyl-d₃)-1H-pyrrole-2-carboxylic acid (4) as well. To further broaden the production of potential congocidine derivatives, the synthesis of different heterocycles, based on the natural precursor, were carried out and fed. Synthesis of all the precursors used for feeding are described in the materials and methods chapter. For clarity the compounds synthesised and used in the feeding experiments are summarised in Table 13, along with the predicted congocidine-like compounds that may result from incorporation of the non-natural heterocycles in the congocidine peptide. To determine congocidine restoration and production of methyl-deuterated congocidine, pyrrole-thiazole (py-tz) or pyrrole-imidazole (py-im) congocidine derivatives, LCMS was employed to analyse samples from the feeding experiments as discussed later.

Table 13. Alternative heterocycles used in the feeding experiments

Building block	Expected Compound
 <p>1</p> <p>(Lautru <i>et al.</i>, 2012)</p>	 <p>Exact Mass: 430.22</p>
 <p>4</p>	 <p>Exact Mass: 430.22</p>
 <p>5</p>	 <p>Exact Mass: 436.26</p>



Strain construction of a Δ *cgc10* congocidine mutant

After three attempts to produce the strain as described in materials and methods, and due to time constrains, the Δ *cgc10* mutant already integrated in *Streptomyces lividans* TK23 was kindly provided by Dr. Sylvie Lautru.

Production of novel congocidine derivatives by mutasynthesis

Due to lack of initial availability and means of production of 4-Acetamidopyrrole-2-carboxylate (1) that experimental procedures could focused on feeding the carboxylic acids of the acetamido methyl-pyrrole (4), methyl-imidazole (6) and thiazole (7) directly. All experiments were performed in triplicate. A negative control of each experiment was also conducted which consisted of the bacterial strain, with no substrate fed, but the culture was extracted and processed as described for the fed strains in parallel. After feeding to 1 mM at 48h, when total cultivation time reached 96 h, the supernatant was sampled for congocidine production according to what is described in the protocol of Dr.Lautru (Lautru *et al.*, 2012). A UV-LCMS signal overlay of the supernatant from each each feeding experiment (Fig. 40) shows that congocidine production is restored when feeding the unnatural, methylated acetyl amino pyrrole carboxylic acid. This shows that the biosynthetic enzymes are capable of accepting methylated precursors, whereas the natural biosynthetic pathway incorporates unmethylated substrates, with methylation occurring after the polyamide has been synthesised. When the 4-acetamido-1-methyl-imidazole-2-carboxylic acid (6) and the 4-acetamidothiazole-2-carboxylic (7) acid are fed, a polyamide-like structure is not synthesised suggesting that the biosynthetic machinery is incapable of incorporating these structures in to the extending polyamide chain. This hypothesis could be further tested by individual enzymatic assays in another study.

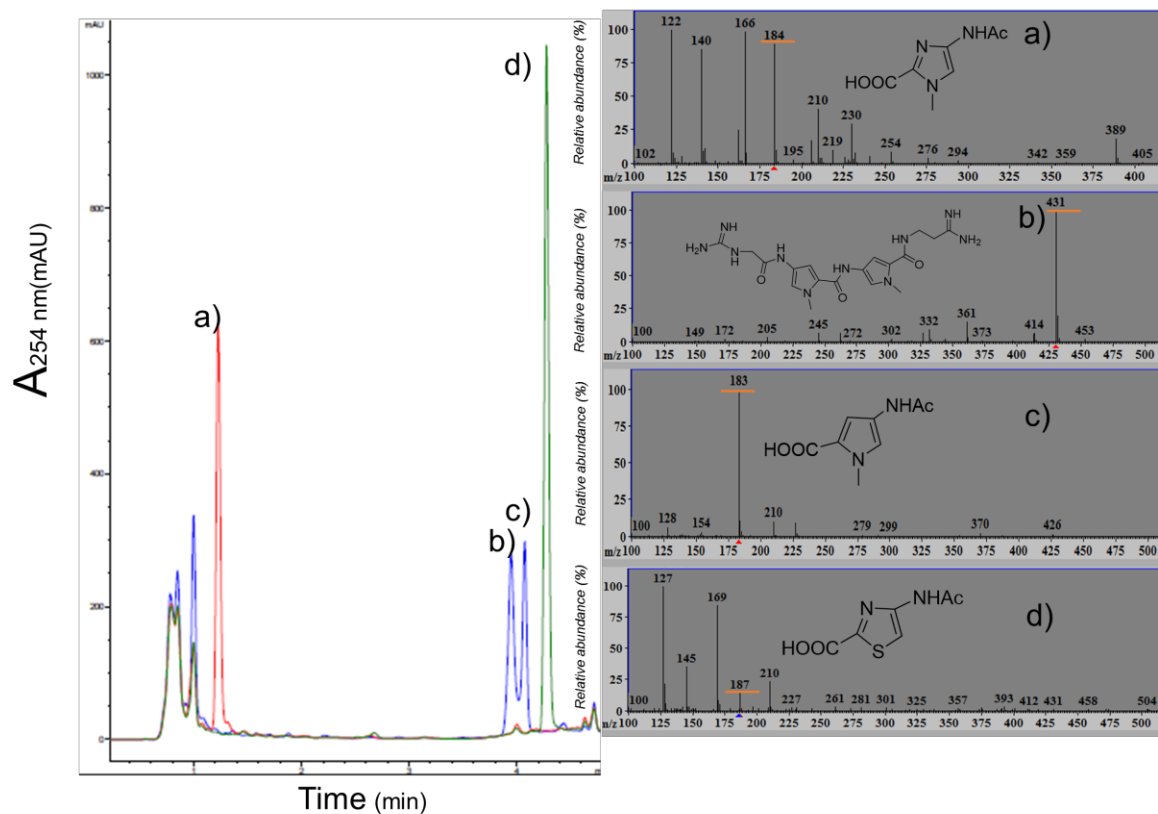


Figure 41. UV signal overlay of the LCMS analysis of the supernatants of the *S.lividans* congocidine $\Delta cgc10$ mutant after 96 h in culture with different chemical precursor addition at 48 h.

Each colour represents a different feeding experiment; 4-acetamido-1-methyl-pyrrole-2-carboxylic acid in blue, 4-acetamido-1-methyl-imidazole-2-carboxylic acid in red, 4-acetamidothiazole-2-carboxylic acid in green). The structures of the compounds identified by m/z are indicated.

Given several enzymes would be required to accept these non-natural substrates, it was hypothesised that only very small amounts (not detected by UV) of the novel compounds may have been produced. This hypothesis was tested by m/z scanning of the masses indicated in Table 13 on the LCMS spectral data. This would allow identification of the putative compounds rather than identifying the masses of only the major peaks on the UV. This approach helped to identify the molecules fed into culture and also those visible in the UV spectra. Mass scanning shows presence of a molecular ion with mass of a thiazole-congocidine hybrid compound (437), of which the retention time is slightly increased compared to congocidine (Fig. 42). However finding the mass does not guarantee assembly of any compound, and purification of this compound was not achieved for further characterisation.

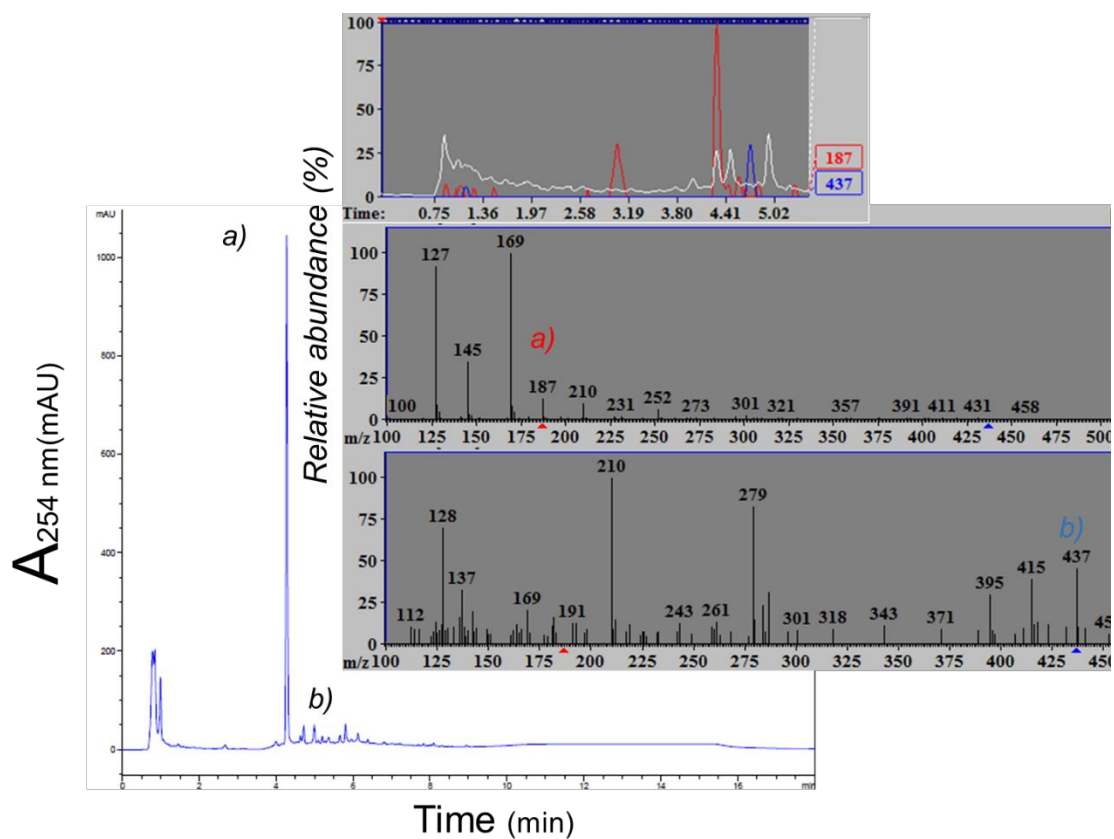


Figure 42. LCMS data on the supernatant culture of 4-acetamidothiazole-2-carboxylic fed batch of the $\Delta cgc10$ congocidine production mutant. Detection at 254 nm.

UV chromatogram and m/z scans for the mass of, the chemical fed into culture (7) (a) and the putative novel compound (b). 4-acetamidothiazole-2-carboxylic (7) (a) elutes at retention time = 4.277min. The putative thiazole derivate of congocidine (437) is eluting at retention time = 4.722min.

Since the secretion of congocidine in to the medium is facilitated through an ABC type transporter/efflux pump (Juguet *et al.*, 2009), it was postulated that any novel congocidine-like compounds may not be transported by the cognate efflux/resistance system to the outside of the cell and the novel compounds may be trapped in the inside of the cells due this pump-substrate incompatibility. One further possibility is that the novel substrates are not taken up by the strains and thus cannot be incorporated in to the biosynthetic reaction. To explore this the whole cell contents were extracted to check the intracellular contents and also to allow an estimation of the amount of non-natural precursor available for the enzymes. However, no assembly of novel compounds was detected either (data not shown). In light of the putative success with the feeding of a methylated pyrrole in the restoration of congocidine, the experiment was repeated with a deuterated methyl pyrrole, in order to check that the pathway was still functioning. The suggestion is that this experiment would confirm the successful feeding of an alkylated pyrrole, since de-acetylase activity in *S.lividans* may had reversed the N-methylated pyrrole to the natural precursor. Following the synthesis of 4-acetamido-1-(methyl-d3)-1H-pyrrole-2-carboxylic acid (5), it was fed to *S. lividans* $\Delta cgc10$ as previously described. After several assays were performed without success, only trace amounts of the deuterated congocidine were putatively obtained (Fig. 43). Purification of these components for characterisation, supernatant concentration and semi preparative HPLC were performed, although only the monomer was able to be recovered from culture supernatant. The small change in steric hindrance due to a deuterated methyl group against a non-deuterated methyl group is unlikely to produce any change in precursor uptake. To further understand such a decrease in congocidine production when the complementation was done with 4-acetamido-1-(methyl-d3)-1H-pyrrole-2-carboxylic acid (5) instead of 4-acetamido-1-(methyl)-1H-pyrrole-2-carboxylic acid (4) the feeding experiments were repeated for both methyl and deuterated methyl pyrrole in parallel, although without success. The original conditions were followed as well as earlier and later time points were sampled. Some of the *m/z* scans performed in consecutive batches showed the expected mass, although with negligible UV peaks and congocidine recovery was not achieved.

To ensure that the mutant was still metabolically able to produce congocidine under complementation with the natural substrate, a positive control (feeding of the reported 4-acetamido-1H-pyrrole-2-carboxylic (3)) was performed. The LCMS data proved the assay failed to produce any congocidine after four attempts following the exact steps as described

in the work by (Lautru *et al.*, 2012) nor with small deviations from the reported assay, such as prolonged incubation or earlier addition of the precursor or slight changes in media pH or temperature incubation to name some. At this point, the strain was thought to be the problem, so it was tested for the presence of the congocidine cluster and the antibiotic marker used in the conjugation. The presence of the pathway in the strain was tested by PCR as well as by ability to grow in presence of hygromycin.

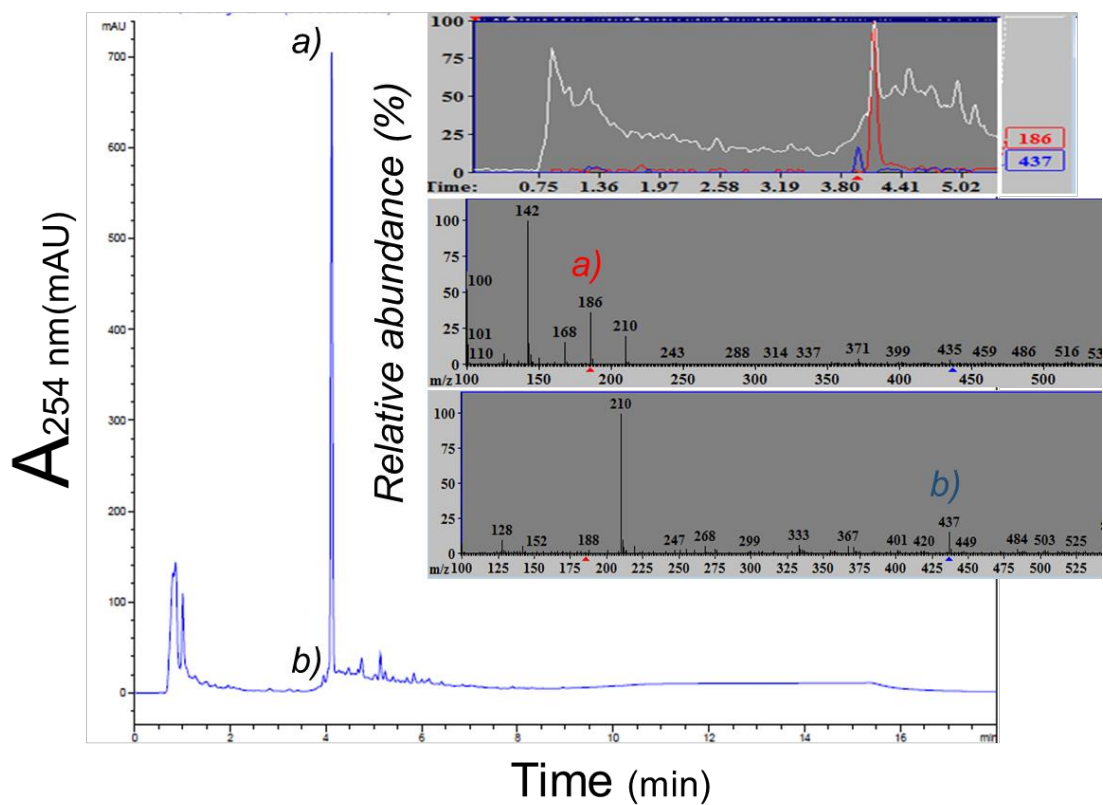


Figure 43. UV signal (LCMS) and m/z scan for 4-acetamido-1-(methyl-d₃)-pyrrole-2-carboxylic acid and deuterated congocidine. Detection at 254 nm.

UV chromatogram showing a (a) peak corresponding to 4-acetamido-1-(methyl-d₃)-1H-pyrrole-2-carboxylic acid (Retention time=4.138) . Putatively at retention time = 3.974 there is formation of minute amounts of deuterated congocidine as indicated by the (b) peak showing the m/z of M+H 437.1.

Conclusions

The incorporation of 4-acetamido-1-(methyl)-1H-pyrrole-2-carboxylic acid (4) and to some extent 4-acetamido-1-(methyl-d3)-1H-pyrrole-2-carboxylic (5) is putatively achieved, but not reproducible, and the positive control as described in the work by Dr. Lautru (Lautru *et al.*, 2012) fails to restore congocidine production. It is thought that the strain containing the congocidine cluster lacking the *cgc10* gene is not functional. Pathway deletion should be discarded as the strain is able to grow in presence of hygromycin, the selection marker used in the cosmid containing the pathway. Perhaps a spontaneous mutation compromised the expression of the pathway. Future work should include creation of the same strain, making a knock out in a congocidine producing strain. Once the work by Dr. Lautru and colleagues is reproducible, feedings could be pursued again.

Chapter 4: Pyrrole-amide ABC transporter study

Introduction

ABC transporters represent the largest class of all transporter protein families (Zhang *et al.*, 2015) and efflux pumps belong to this protein family. Their role in antibiotic resistance has been extensively reported (Webber and Piddock, 2003). The efflux mechanism does not alter the drug, but reduces its cellular concentration due to continuous pumping (efflux) of it to the external environment, thus increasing the Minimum Inhibitory Concentration (MIC) of the drug.

Early studies indicated large sections of the bacterial chromosome (5-10%) were dedicated to genes encoding efflux pumps. A closer look into their phylogenetic distribution shows multiple functions for the ABC superfamily, mostly transporting different molecules such as siderophores (Saier and Paulsen, 2001). Their presence not only in bacteria but also eukaryotic cells (Bambeke *et al.*, 2000) points towards a universal and conserved resistance mechanism. Any antibiotic producing bacteria will be self-protected by a resistance mechanism, usually clustered within the pathway genes responsible for the antimicrobial assembly (Cundliffe, 1989). Knowing that soil bacteria are the biggest source of antimicrobials to the date (Baltz, 2008), it is easy to conclude that soil is the most likely ecosystem in which resistance to any new or old antibiotic will arise due to mobile genetic elements from producers to neighbouring bacteria. This idea was proposed back in 1998 when a study established the presence of paralogs of the vancomycin cluster in glycopeptide antibiotic resistant bacteria (Marshall *et al.*, 1998), and supported by the shocking fact that the average soil bacterium is resistant to between 15 and 20 different antimicrobials. In another study, 8% of the isolates studied showed resistance to daptomycin shortly after its FDA approval (D'Costa, 2006). It also has been found that the vast majority of resistant bacteria used non-modifying drug resistance mechanisms. Efflux pumps are not the only mechanisms involved in resistance that do not alter drugs, but the abundance of DNA sequences encoding efflux pump genes, may indicate that efflux is the predominant mechanism for resistance in the bacteria used in the D'Costa study.

ABC type efflux pumps play a role not only in resistance, but also in regulation and antibiotic production. In *Streptomyces sp.* Regulation, this has been recently studied and the data leans towards secondary metabolites being used as signalling molecules (Lee *et al.*, 2012). In the era of antibiotic resistance huge efforts are put into the understanding of resistance mechanisms, with current research on efflux systems focused on multidrug efflux systems of pathogenic bacteria. However, integral protein structural determination is still difficult, since creating crystals of membrane proteins suitable for X-ray crystallography remains challenging (Moraes *et al.*, 2014). Although the basic mechanistic details on drug efflux are known (Sun *et al.*, 2014), these are focused on pathogenic bacteria, and the profiling is done using common antibiotics (Kumar and Worobec, 2005). So far, little attention is focused on the mono-resistant efflux pumps clustered into antibiotic producing pathways. Their study may provide an easier avenue to establish SAR between drugs and efflux systems, that could be used to improve productivity on industrial heterologous hosts or guide the development of 'non-pump-able' drugs, instead of following the trend of using antibiotics synergistically with pump inhibitor drugs (Tegos *et al.*, 2011). This study focus on two ABC type efflux system paralogs that likely evolved from the same common ancestor but represent a different chemical sensitivity profile. The systematic study of this putative resistance mechanism for congoicidin and distamycin with functionalised synthetic compounds intends to gain insights into SAR at the efflux level and to establish an iterative platform to anticipate resistance and generate novel drugs for pathogenic bacteria. Furthermore, this knowledge will also be fundamental to our understanding on how to produce laboratory strains able to export non-natural metabolites that are assembled *in vivo* due to mutasynthetic experiments (see chapter 3) in the quest for novel drugs.

Since one of the aims of this thesis is the production of novel pyrrole amide peptides, the substrate specificity of the putative congoicidin and distamycin ABC transporters were characterised with their natural substrates, and with derivatives of those. A systematic approach was taken in the study; the rationale on the compound synthesis for the screenings was focused in fixing the charge and steric impediment in the molecules, while making single moiety changes subdividing these changes into families or sets. The total number of compounds assayed as a result of this rationale was 54 compounds that were synthesised by Professor Colin Suckling's laboratory, plus two novel compounds synthesised for this study, compiled in 3 sets according to their structural similarity to distamycin, congoicidin or the

new pyrrole amide BP3, an antibacterial compound assembled by the Suckling group, now in phase 1 of clinical trials,. The modifications of the two natural products were:

- Same C-ends as in the natural product but different N- alkylations in the pyrrole
- Substituted amidine moiety
- Substituted guanidinium moiety
- Substituted amidine moiety and different N- alkylations in the pyrrole
- Substituted guanidinium moiety and different N- alkylations in the pyrrole
- Substituted guanidinium and amidine moieties
- Substituted guanidinium and amidine moieties and different N- alkylations in the pyrrole
- Introduction of a thiazole instead of a pyrrole with N- alkylations and substituted guanidinium and amidine moieties

***In silico* analysis of the putative transporters**

The congocidine resistance genes from *S.ambofaciens* are SAMR0919 and SAMR0920 or *cgc1** and *cgc2** to follow the nomenclature used in the study that unravelled the congocidine production (Juguet *et al.*, 2009), and the putative distamycin resistance genes are AAU04844 and AAU04845 or Dst20 and Dst21 as named in the paper depicting distamycin biosynthesis (Vingadassalon *et al.*, 2014b). These were analysed using bioinformatics tools (BLAST, TMpred, Geneious tree builder) before their substrate tolerance profiling.

The distribution of the resistance genes within the two antibiotic clusters are identical. Common structural features are the clustering with a putative Acyl-CoA transferase in an operon with opposite transcription direction of the antibiotic biosynthetic genes and with 1 bp overlap with them. This distribution, results in an overlap of 17 bp to a putative Acyl-CoA transferase with Dst21 in the distamycin cluster, and only 3 bp distance between *cgc2** and the other putative Acyl-CoA transferase in the congocidine cluster, which points towards a single promoter for the three genes in each case (SAMR0919/ SAMR0920/Acyl transferase gene; AAU04844/ AAU04845/ Acyl transferase gene). An alternative to this single promoter theory, would be the presence of additional promoters within the genes, since the Shine-Dalgarno consensus sequence GGAGG is present in the both -10 region of both *dst21* and

*cgc2**, however not present before the acyl-CoA transferase coding gene in either cluster. If the hypothesis of a single promoter controlling the resistance genes as well as a biosynthetic gene were to be true, this arrangement could be a fail-safe type regulatory system, in which only after the resistance mechanism is in place, the biosynthetic pathway can move forward, and it would call for a revision on the distamycin biosynthetic pathway as some other authors already proposed with the revision of the congocidine pathway (Al-Mestarihi *et al.*, 2015). However both biosynthetic clusters possess an obvious regulatory system. One gene downstream of the ABC transporters and starting in the opposite biosynthetic operon, codes for a two component response regulator, Dst1 (AAU04843) in the distamycin cluster 1 and *cgc1*(SAMR0918) in the congocidine cluster, with no more obvious regulatory elements in the neighbouring DNA up or downstream of both ABC type transporters.

In order to compare the putative resistance genes of both clusters structurally, and confirm their role as ABC transporters, an alignment search of their aminoacidic sequences was performed using the online tool BLAST (Tatusova and Madden, 1999) which uses the Clustal algorithm (Larkin *et al.*, 2007). Highly similar sequences were identified using a cut off rate of 50% homology for a query cover bigger than 90% were downloaded and aligned by the CLUSTAL W algorithm (Larkin *et al.*, 2007) on the software tool geneious. The results are displayed in Fig.44, where a clear differentiation within the sequence can be observed, including a highly conserved ATP binding domain in which the core motifs have been highlighted and a more variable TMD region, which surprisingly shows a higher degree of similarity towards a multidrug efflux pump from *Streptomyces hygroscopicus* rather than to the protein coded by Dst20 in *Streptomyces netropsis*. This is illustrated using a Neighbor-Joining (Saitou, 1987) phylogenetic tree built using the Jukes-Cantor or JC69 (Jukes and Cantor, 1969) model for genetic distance. It shows the evolutionary relationships among similar proteins from other strains and express their distances in the number of amino acids substitution per site in each branch, as seen in Fig.45. This same process was carried out for *cgc2**, *dst20* and *dst21* as seen in Figs.44-49.

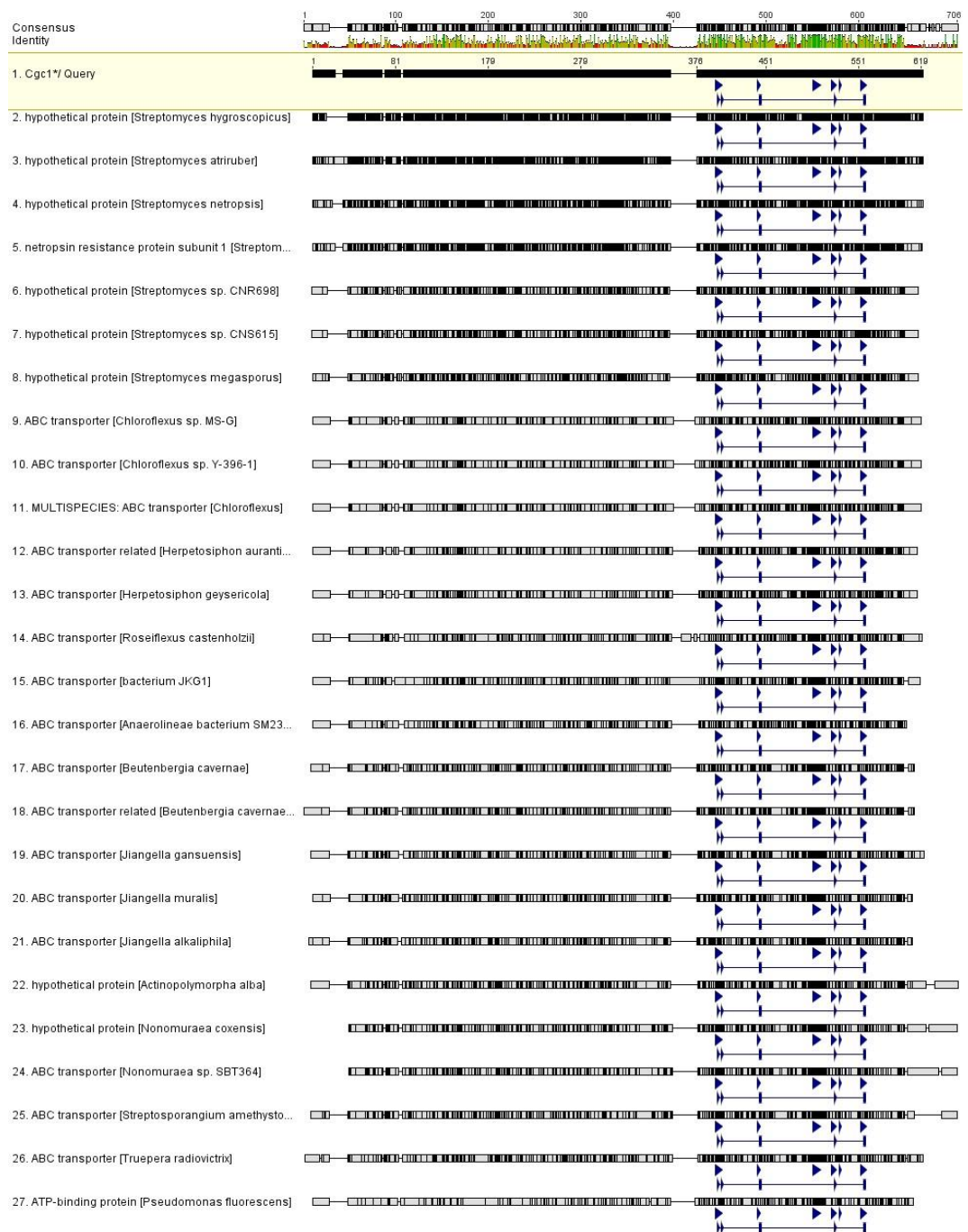


Figure 44. Protein BLAST data search of *cgc1**

Sequences aligned following the CLUSTAL W (Larkin *et al.*, 2007) algorithm. *cgc1** sequence is highlighted at the top of the figure, right below a sequence chromatogram showing a consensus sequence (green tall bar for highly conserved residues, short red for non conserved). Marked in blue with triangles are signature motifs for ABC proteins. The first 27 sequences (homology above 50 % and query cover of at least 90%) are displayed.

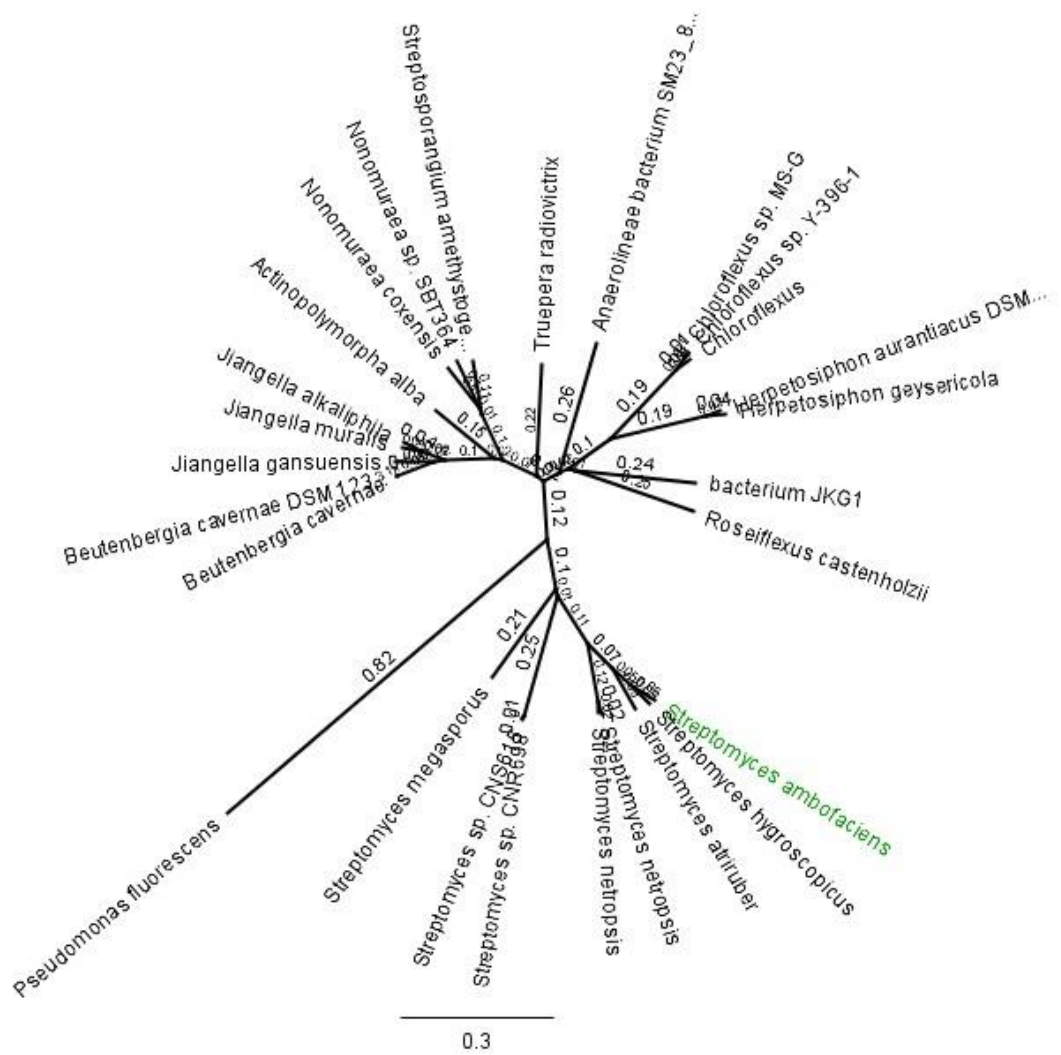


Figure 45. Phylogenetic tree of *cgc1.**

Evolutionary relationships between resistance proteins that show high homology to the protein coded by *cgc1** (in green). The distance is shown as number of aminoacids substitutions per site and calculated by JC69, to then build the tree with the Neighbour-Joining method.

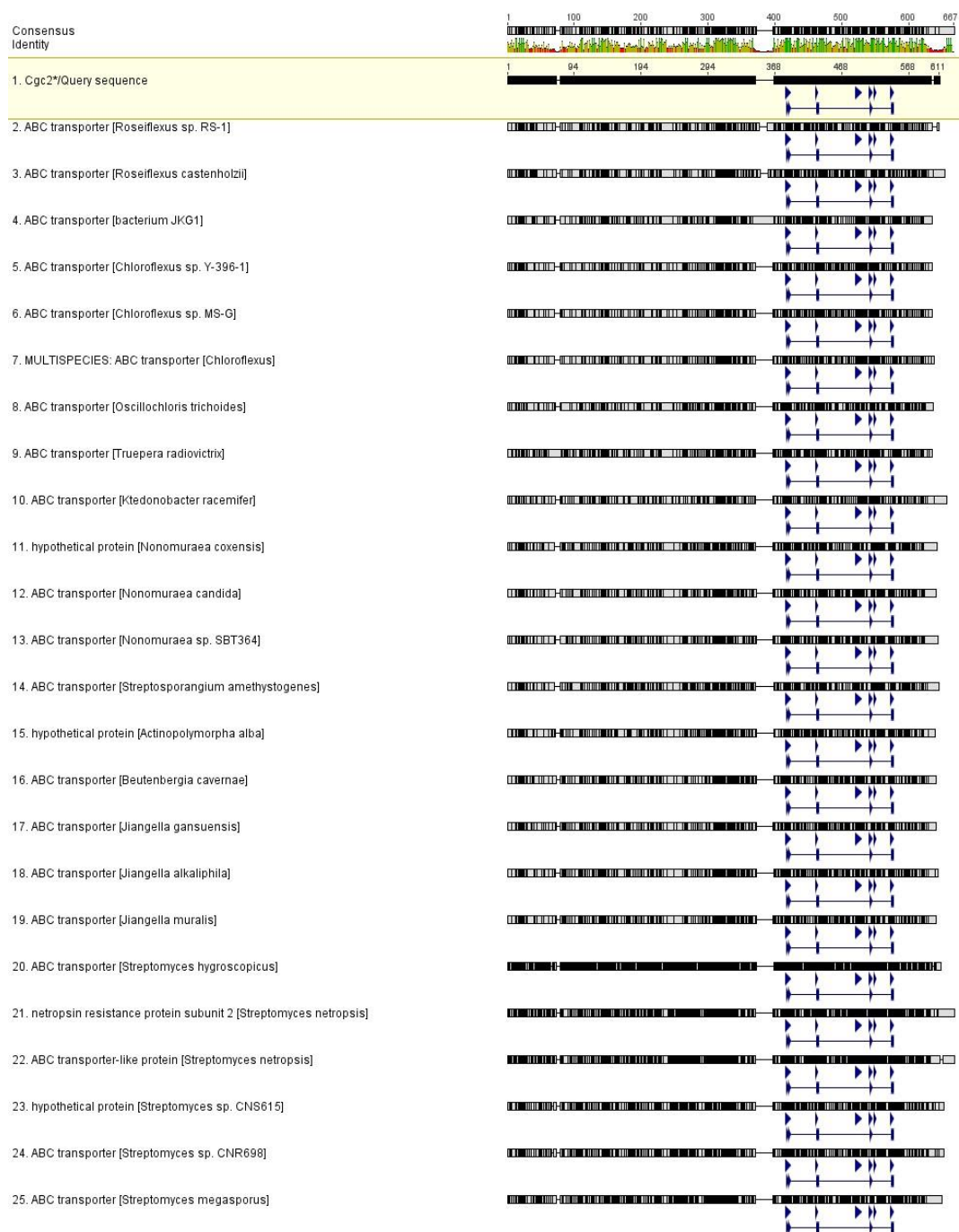


Figure 46. Protein BLAST data search of *cgc2.**

Sequences aligned following the CLUSTAL W (Larkin et al., 2007) algorithm. *cgc2** sequence is highlighted at the top of the figure, right below a sequence chromatogram showing a consensus sequence (green tall bar for highly conserved residues, short red for non conserved). Marked in blue with triangles are signature motifs for ABC proteins. The first 25 sequences (homology above 50 % and query cover of at least 90%) are displayed.

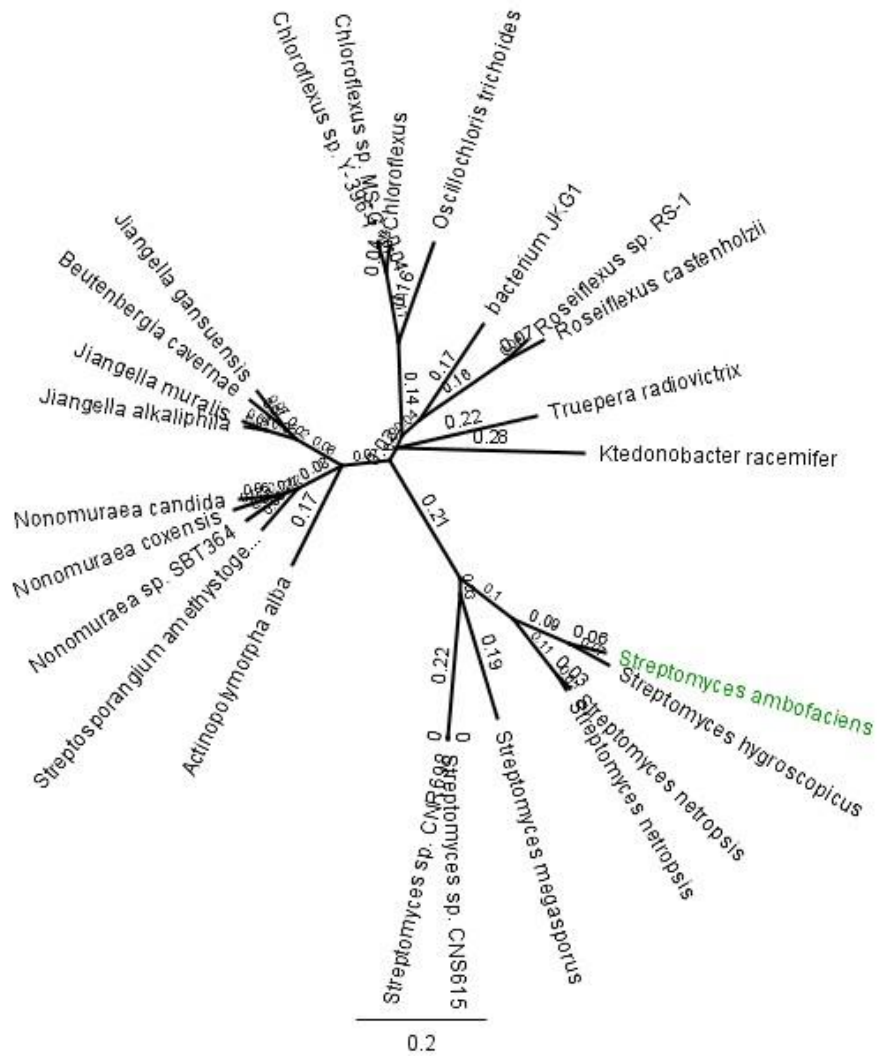


Figure 47. Phylogenetic tree of *cg2.**

Evolutionary relationships between proteins that show high homology to the protein coded by *cg2** (in green). The distance is shown as number of amino acids substitutions per site and calculated by JC69, to then build the tree with the Neighbour-Joining method.

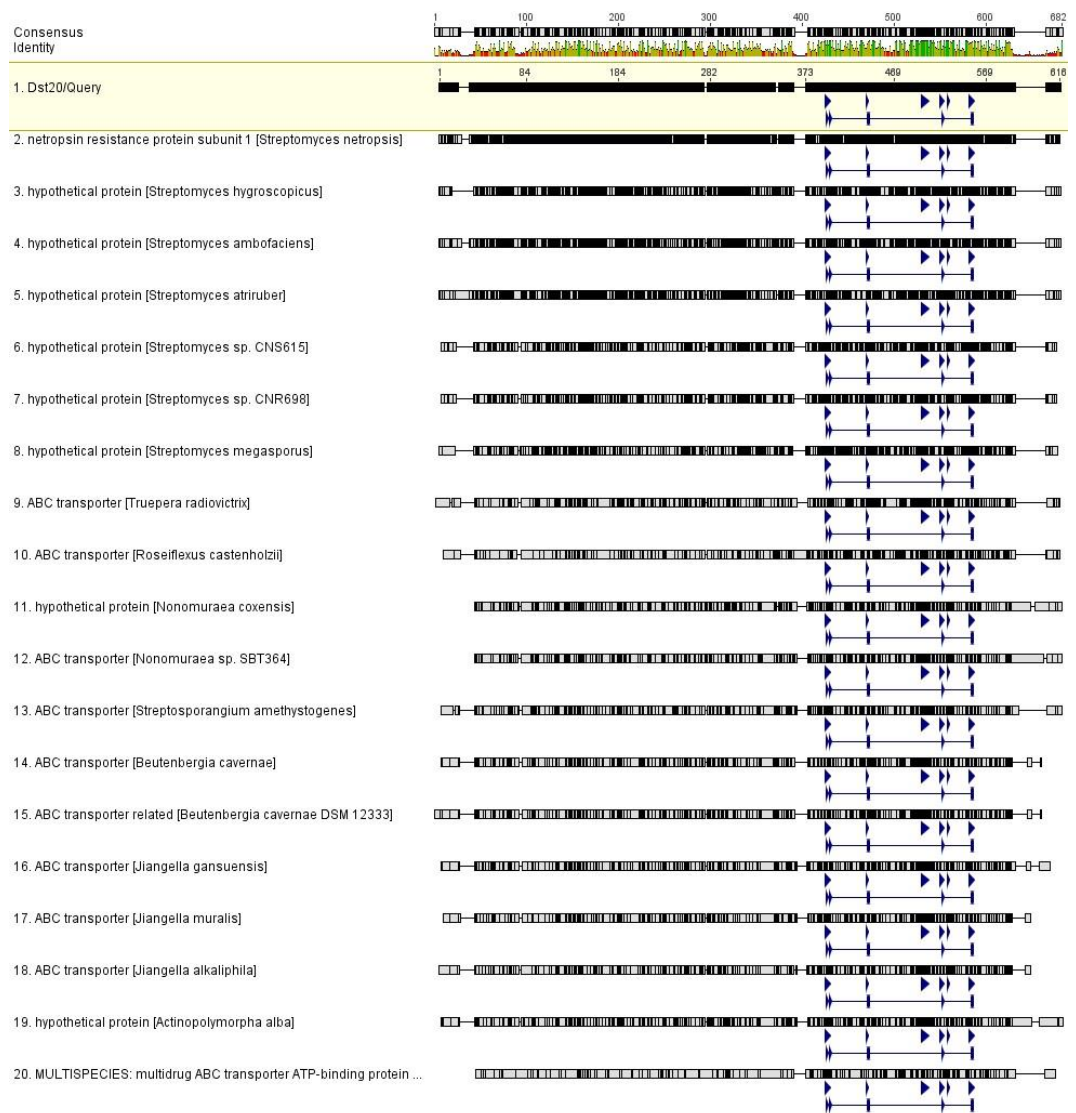


Figure 48. Protein BLAST data search of Dst20

Sequences aligned following the CLUSTAL W (Larkin *et al.*, 2007) algorithm. Dst20 sequence is highlighted at the top of the figure, right below right below a sequence chromatogram showing a consensus sequence (green tall bar for highly conserved residues, short red for non conserved). Marked in blue with triangles are signature motifs for ABC proteins. The first 20 sequences (homology above 50 % and query cover of at least 90%) are displayed.

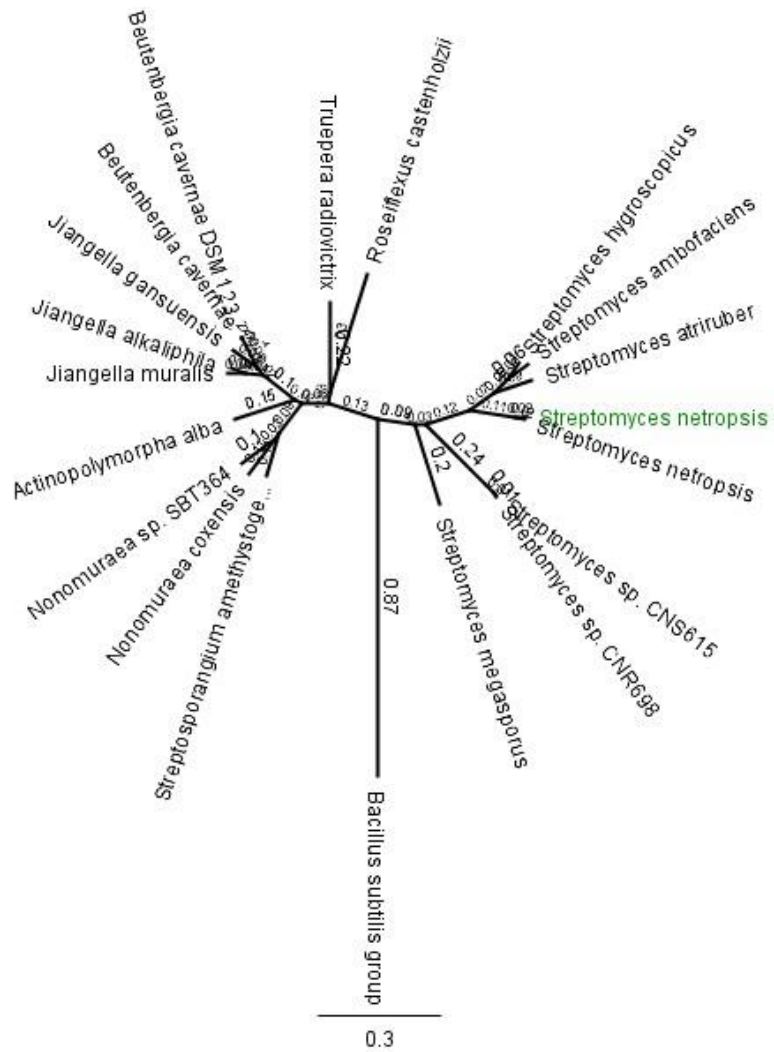


Figure 49. Phylogenetic tree of Dst20.

Evolutionary relationships between proteins that show high homology to the protein coded by Dst20 (in green). The distance is shown as number of aminoacids substitutions per site and calculated by JC69, to then build the tree with the Neighbour-Joining method.

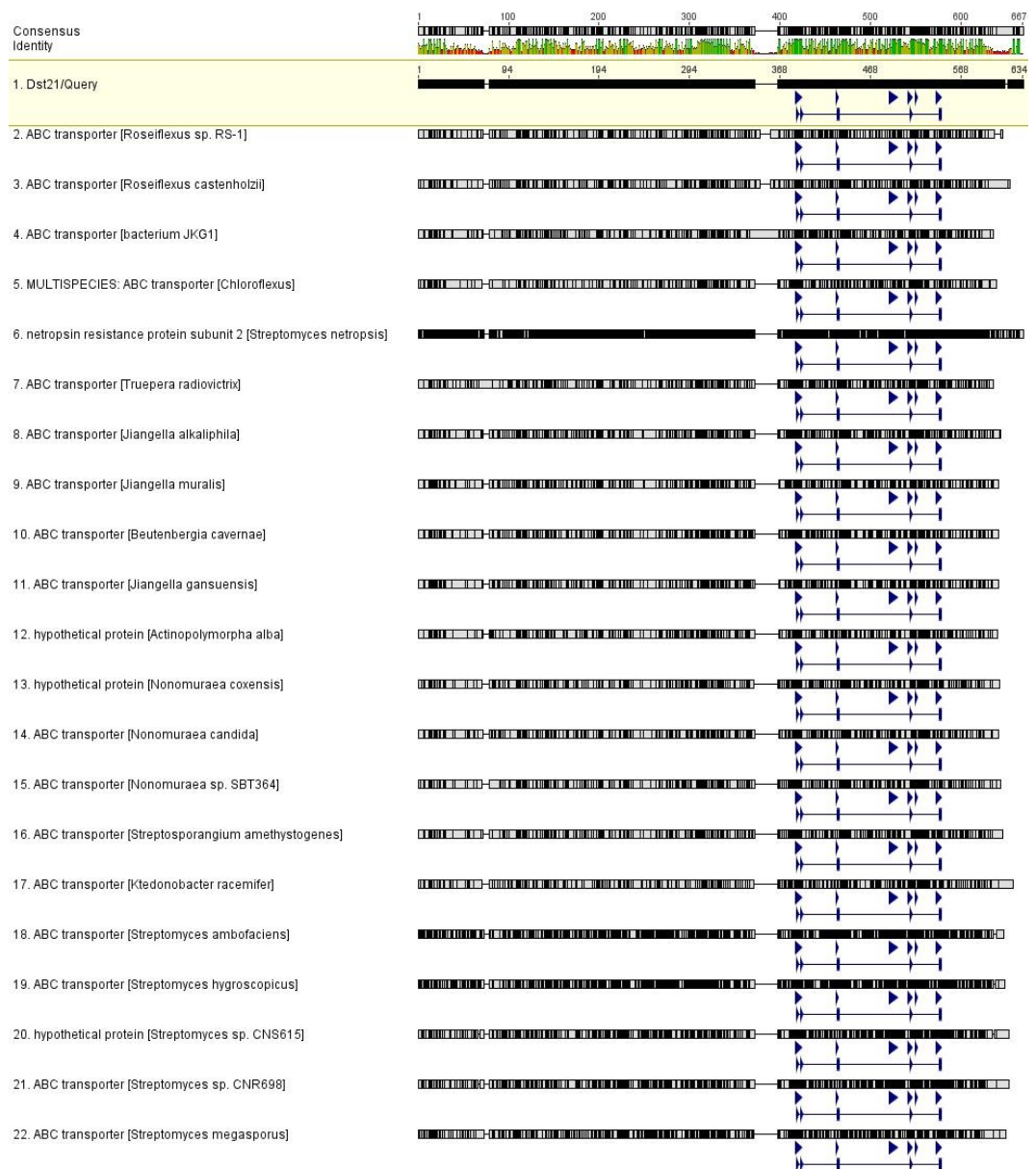


Figure 50. Protein BLAST data search of Dst21

Sequences aligned following the CLUSTAL W (Larkin *et al.*, 2007) algorithm. Dst21 sequence is highlighted at the top of the figure, right below a sequence chromatogram showing a consensus sequence (green tall bar for highly conserved residues, short red for non-conserved). Marked in blue with triangles are signature motifs for ABC proteins. The first 22 sequences (homology above 50 % and query cover of at least 90%) are displayed.

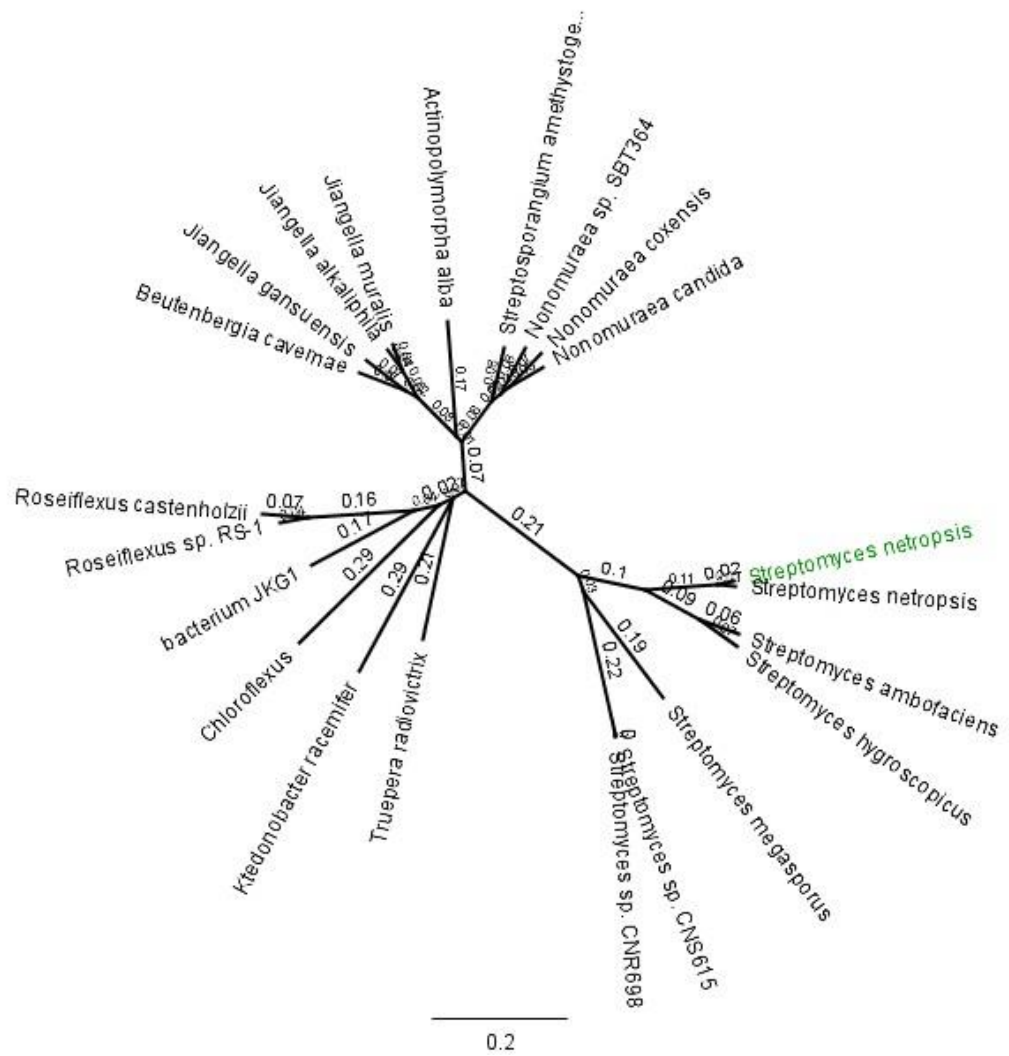


Figure 51. Phylogenetic tree of dst20.

Evolutionary relationships between proteins that show high homology to the protein coded by Dst20 (in green). The distance is shown as number of aminoacids substitutions per site and calculated by JC69, to then build the tree with the Neighbour-Joining method.

As the amino acid alignments Fig.45, 47, 49, 51 show, all four genes are closely related, coding for proteins comprising highly conserved ABC domains and slightly divergent TMD domains. The expression of these genes, *cgc1**+*cgc2** / *Dst20*+*Dst21* points towards the formation of hetero dimer ABC transporters. By sequence homology, *cgc1** relates to *dst20* and *cgc2** to *dst21*. However when running a prediction with the transmembrane spanning algorithm TMbase (Hofmann and Stoffel, 1993) (TMpred, Swiss protein database ExPASy), six transmembrane helices for both *Dst21* and *cgc1** are estimated, although with opposite N-terminal orientation for each one (N terminus inside for *cgc1**), and five helices for *Dst20* and *cgc2** but again with opposed N orientations (N terminus outside for *cgc2**). The predictions can be seen in Fig.53 for the *cgc1*/cgc2** dimer and in Fig.54 for *dst20/dst21*. In the case of *cgc1*/cgc2** the model looks plausible. However, in the case of *dst20/dst21*, the extensive length of the peptides, including the C terminus outside the cell, makes no biological sense, since the ATP binding domain is located at residues 489-498 on *dst21* and 499-508 on *dst20*, which would indicate the presence of the ATPase on the exterior of the cell.

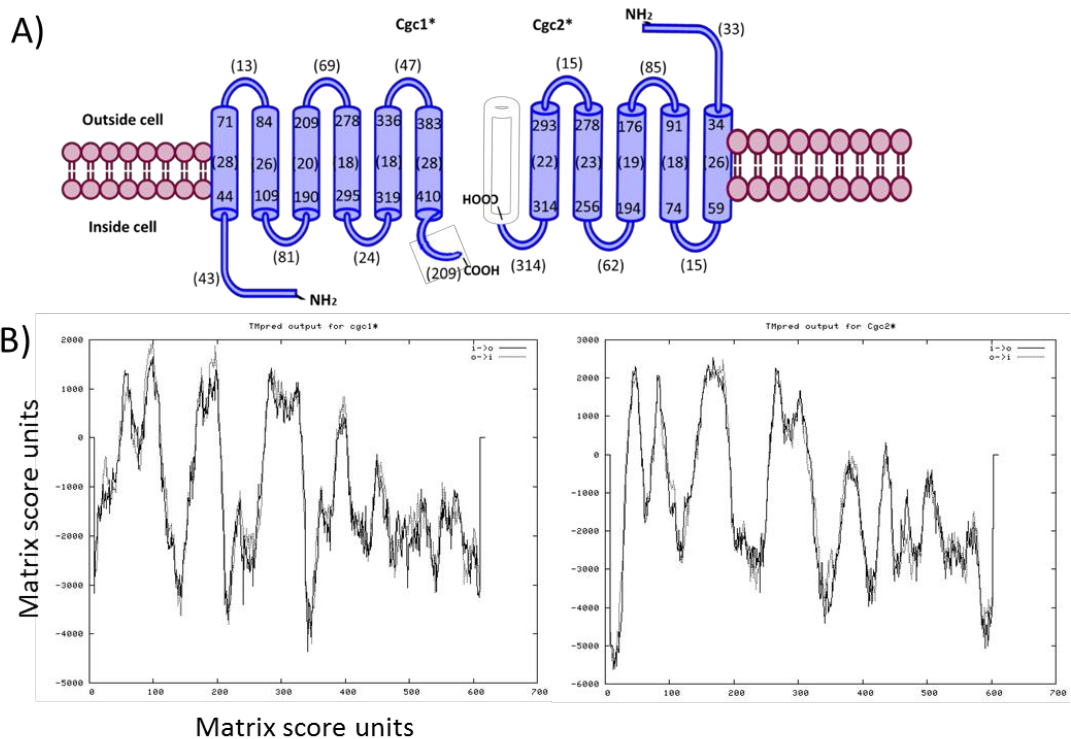


Figure 53. TMpred model and graphic outline of results for congoidine ABC transporter.

Transmembrane domains distribution (A) as predicted by the online tool TMpred (B, Axis represent scoring values calculated with amino acid residues (Hofmann and Stoffel, 1993)) of proteins corresponding to genes *cgc1** and *cgc2**. Numbers between parentheses indicate length of amino acids transmembrane or on the spanning helices; numbers outside parenthesis indicate the beginning and end of those regions.

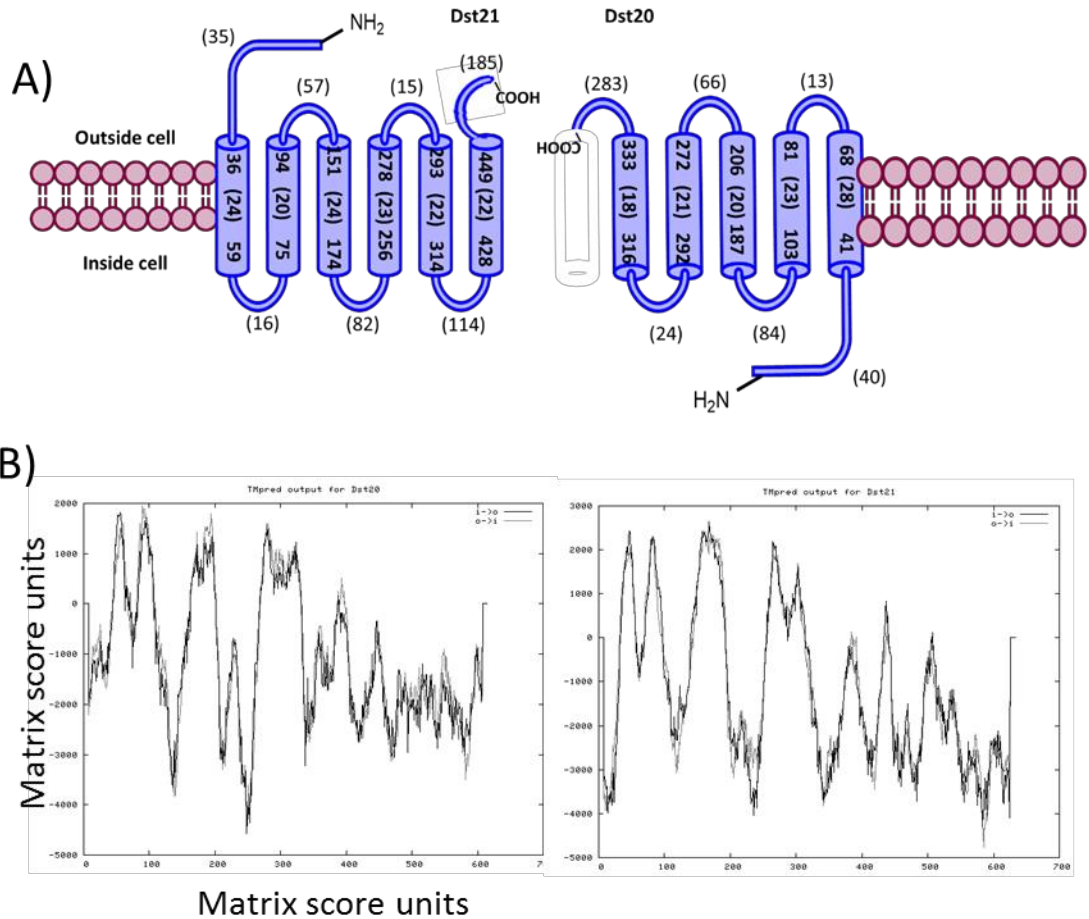


Figure 54. TMpred model and graphic outline of results for distamycin ABC transporter.

Transmembrane domains distribution (A) as predicted by the online tool TMpred (B, Axis represent scoring values calculated with amino acid residues (Hofmann and Stoffel, 1993)) of proteins corresponding to genes *Dst20* and *Dst21*. Numbers between parenthesis indicate length of amino acids transmembrane or on the spanning helices; numbers outside parenthesis indicate the beginning and end of those regions.

Strain construction

While some properties can be inferred from *in silico* analysis, only physical characterisation can determine the substrate specificity of the pumps. To that end, a congocidine and distamycin sensitive host was selected, in which the putative function of the genes *cgc1*/cgc2** and *dst20/dst21* would be tested.

Streptomyces lividans TK23, a *Streptomyces* strain characterised by a quick growth and easy and stable genetic manipulation, was picked as the parental strain in which the putative ABC transporters would be integrated to yield the antibiotic resistant strains. Following the creation of *Streptomyces lividans* TK23RC01 (TK23 resistant to congocidine) and *Streptomyces lividans* TK23RD01 (TK23 resistant to distamycin) from *Streptomyces lividans* TK23, the library of pyrrole-amides was assayed.

Congocidine resistant strain TK23RC01

An overview of the cloning strategy is shown in Fig.55. The putative congocidine resistance genes coding for an ABC type transmembrane transporter (Juguet *et al.*, 2009) were PCR amplified using the high fidelity, blunt end yielding polymerase, KOD XL from the congocidine containing cluster pCG002 (Juguet *et al.*, 2009). A region upstream of the genes of interest (around 400 bp) was included to assure the inclusion of the native promoter. The desired DNA fragment was obtained using the primer pair R12F/R1Rcongo, but the amplification was difficult and several unspecific amplicons were obtained. The problem was solved after PCR optimization and subsequently the target band (5379bp, Fig.55) was gel purified and cloned into the vector pCR-Blunt. The intermediate plasmid generated was named pCS-BC01 (Fig.56) which was used to sub-clone the genes into the *Streptomyces* integrative plasmid pIJ6902 by double restriction digest with the enzymes *KpnI/XbaI*. The resulting plasmid was named pCS-RC01 (Fig.57) which was tested by double restriction digest with the enzymes *KpnI/XbaI* (Fig.57) and verified by sequencing with M13 universal primers, then introduced into *Streptomyces lividans* TK23 via conjugation with a non-methylating *E.coli* strain (see materials and methods) yielding a *Streptomyces lividans* congocidine resistant strain.

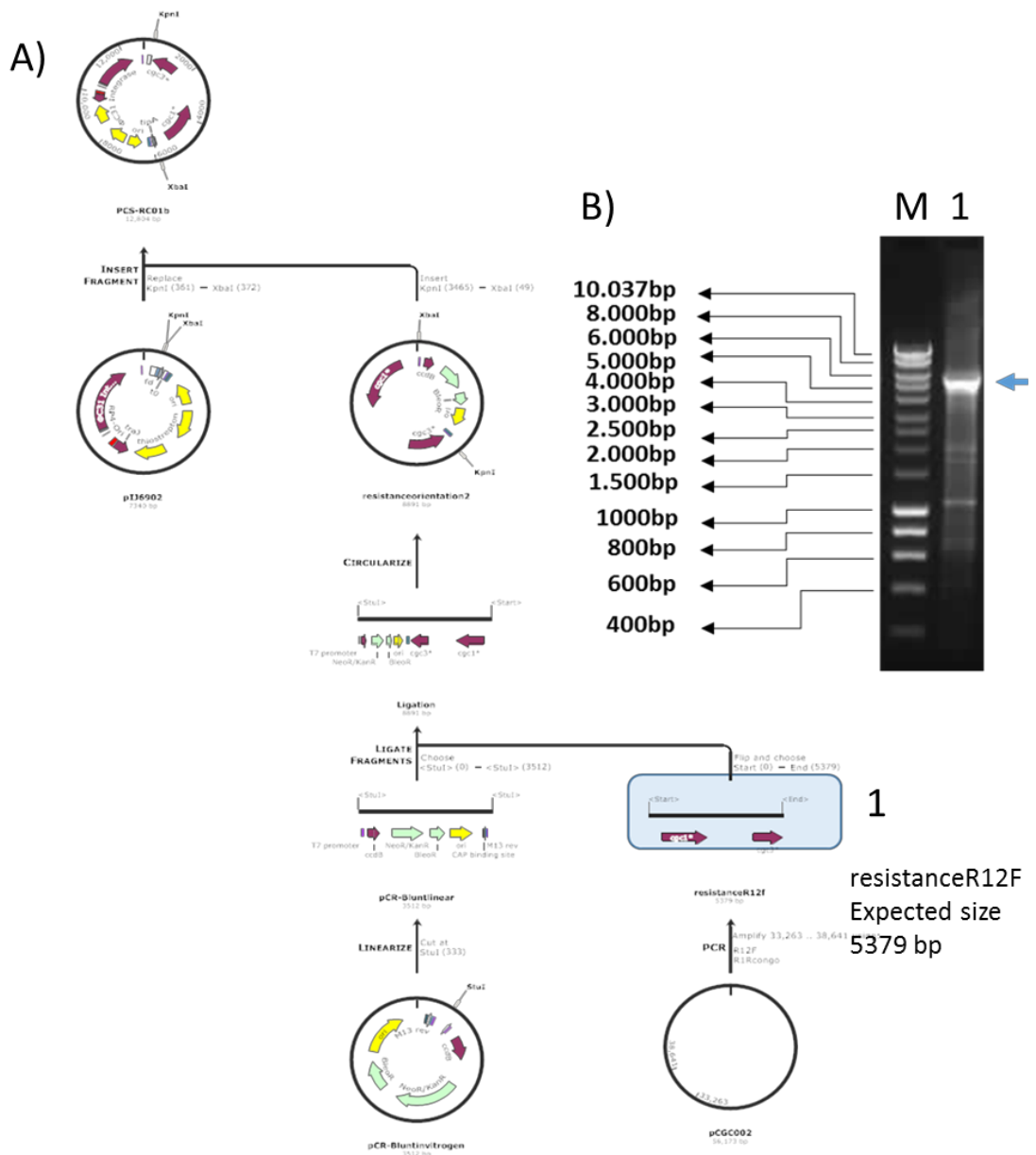


Figure 55. Cloning strategy to obtain the plasmid pCS-RC01.

A) Overview of the cloning process for the strain containing the congocidine transport genes. Those were PCR amplified using pCG002 as template and blunt end cloned into pCRblunt. The intermediate produced pCS-BC01, was then digested with *XbaI* and *KpnI* RE and subcloned into pIJ6902 on those sites. B) PCR product is highlighted in blue.

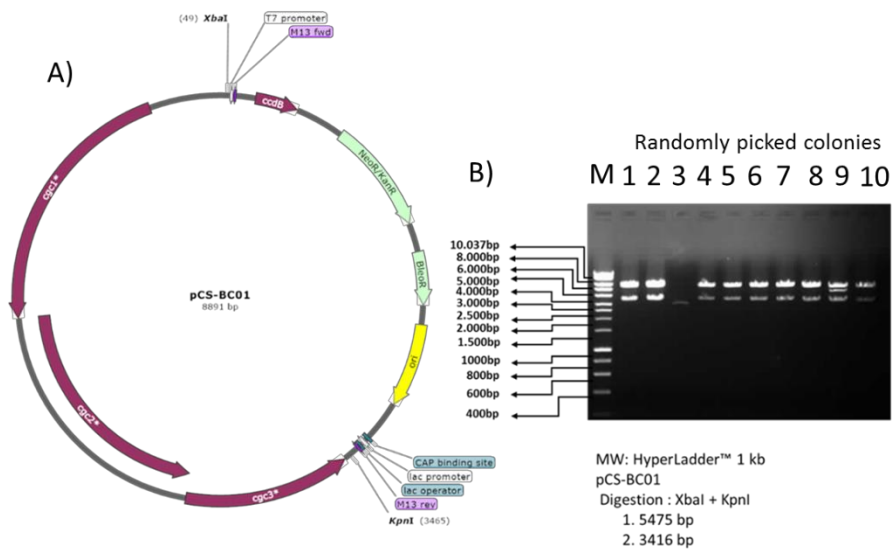


Figure 56. Subcloning vector pCS-BC01.

A) Plasmid map of the subcloning intermediate pCS-BC01, resulting from cloning the congocidine resistance genes on the plasmid pBLUNT. B) The DNA size marker (HyperLadder adapted from www.BIOLINE.com), has the bands labelled. An agarose gel (1.5% agarose/Ethidium Bromide) shows the digestion of a plasmid prep with *XbaI/KpnI* obtained from ten randomly picked colonies from the cloning of pBC01. Expected sizes are below the gel, and the restriction sites are highlighted in the plasmid map.

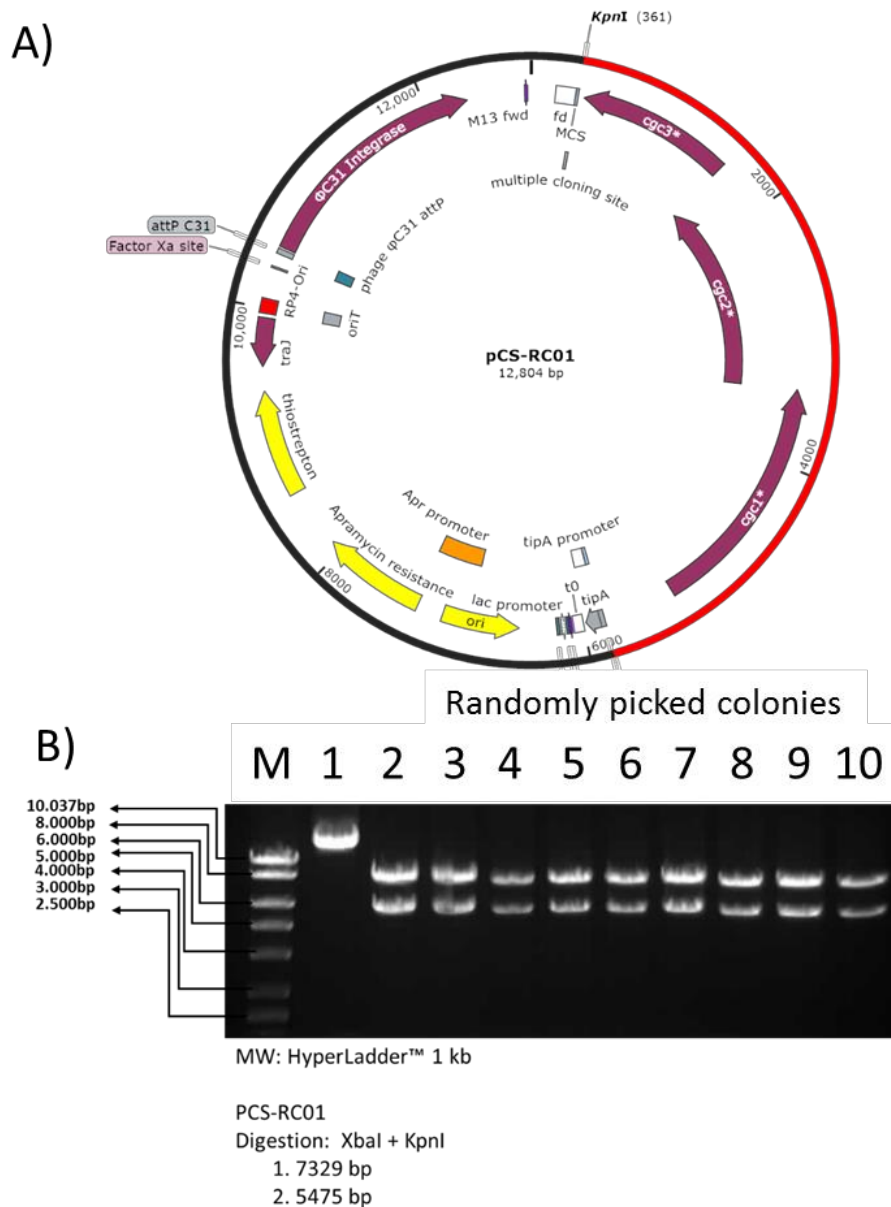


Figure 57. The integrative plasmid pCS-RC01, containing the congocidine efflux system

A) Plasmid map of pCS-RSC01, derivative of pIJ6902 containing the two putative genes for congocidine resistance *cgc1** and *cgc2** (Juguet *et al.*, 2009), the DNA inserted is pictured in red, and the pIJ6902 backbone in black. B) Agarose gel (1.5% agarose/Ethidium Bromide) shows the digestion of a plasmid prep with *XbaI/KpnI* obtained from ten randomly picked colonies from the cloning of pCS-RC01. Expected sizes are below the gel, and the restriction sites are highlighted in the plasmid map.

Distamycin resistant strain TK23RD01

An overview of the cloning strategy is visible in Fig.58. The putative distamycin resistance genes coding for an ABC type transmembrane transporter (Vingadassalon *et al.*, 2014b) were PCR amplified with an high fidelity, blunt ending polymerase KOD XL (Toboyo) from *Streptomyces netropsis* genomic DNA. In order to include the native promoter a region upstream the genes of interest (400 bp) was included. After several attempts the correct band was obtained using the primer pair DPFV/DPRV (Fig.58). Following gel purification the amplicon was cloned into the vector pCR-Blunt (Fig.59) to obtain the intermediate plasmid pBdP. This construct was analysed by restriction digest to ensure correct cloning of the DNA fragments (Fig.59) and subsequently subcloned into the *Streptomyces* integrative plasmid pIJ6902. The resulting plasmid was named pCS-RD01 (Fig.60) and was confirmed by sequencing with M13 universal primers and then introduced into *S.lividans* TK23 via conjugation, yielding an *S.lividans* distamycin resistant strain.

Control strain TK23pIJ6902

To remove the possibility of increased antimicrobial resistance of the chosen host due to the inclusion of the apramycin acyl transferase or thiostrepton resistance in its genome, a novel strain TK23pIJ6902 containing the empty vector pIJ6902 was also made by conjugative transfer.

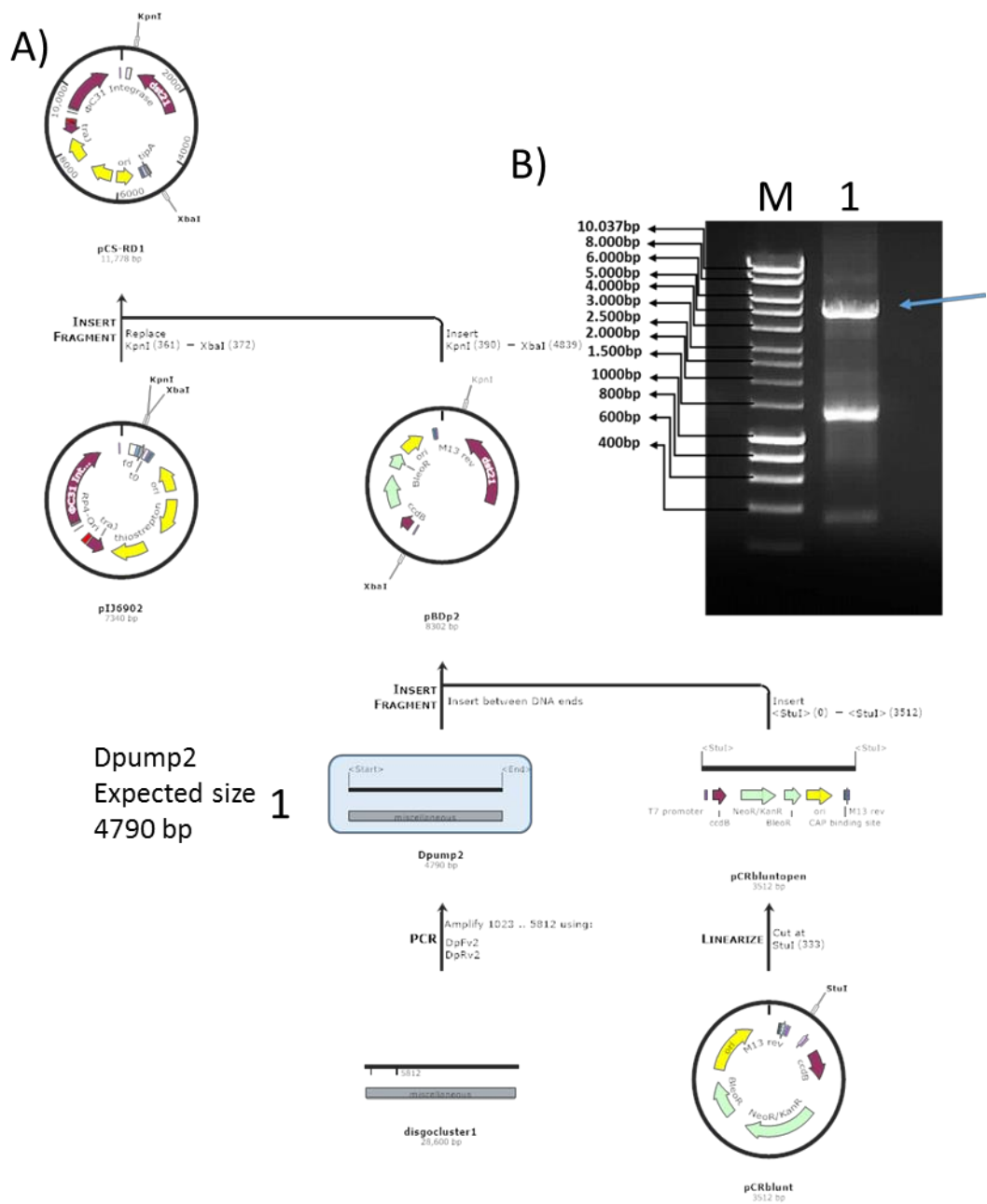


Figure 58. Cloning strategy to obtain the plasmid pCS-RC01.

A) Cloning strategy to obtain the plasmid pCS-RD01. B) Distamycin transporter genes were PCR amplified using *Streptomyces netropsis* genomic DNA as template and blunt end cloned into pCRblunt. The intermediate obtained pBD2, was then digested with *XbaI* and *KpnI* RE and sub-cloned into pI16902 on those sites. The agarose gel shows the distamycin efflux system, marked in blue in the diagram.

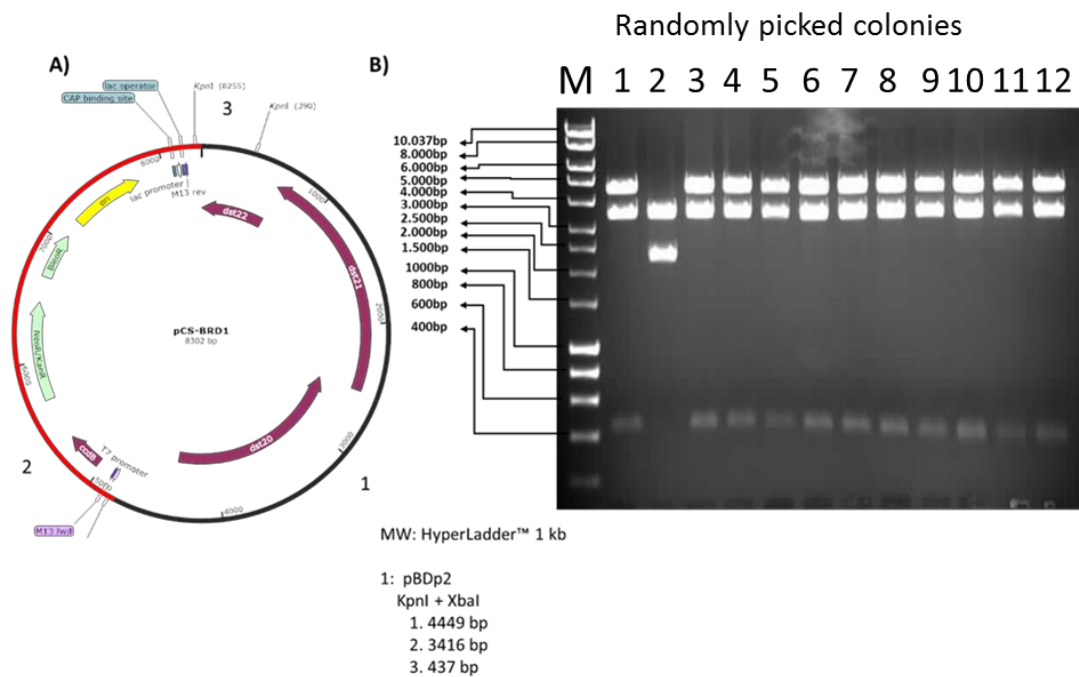


Figure 59. Map and restriction digest of pBdP containing the two putative genes for Distamycin resistance

A) The two putative genes for distamycin resistance *dst20* and *dst21* (Hao *et al.*, 2014) (Vingadassalon *et al.*, 2014b), created by blunt end cloning with pCRBlunt. The plasmid map has the vector backbone in red, and in black the DNA cloned. B) The agarose gel shows restriction pattern of ten randomly picked *E. coli* colonies from the pBdP cloning experiment, and the expected sizes after a digestion with *KpnI*+*XbaI*.

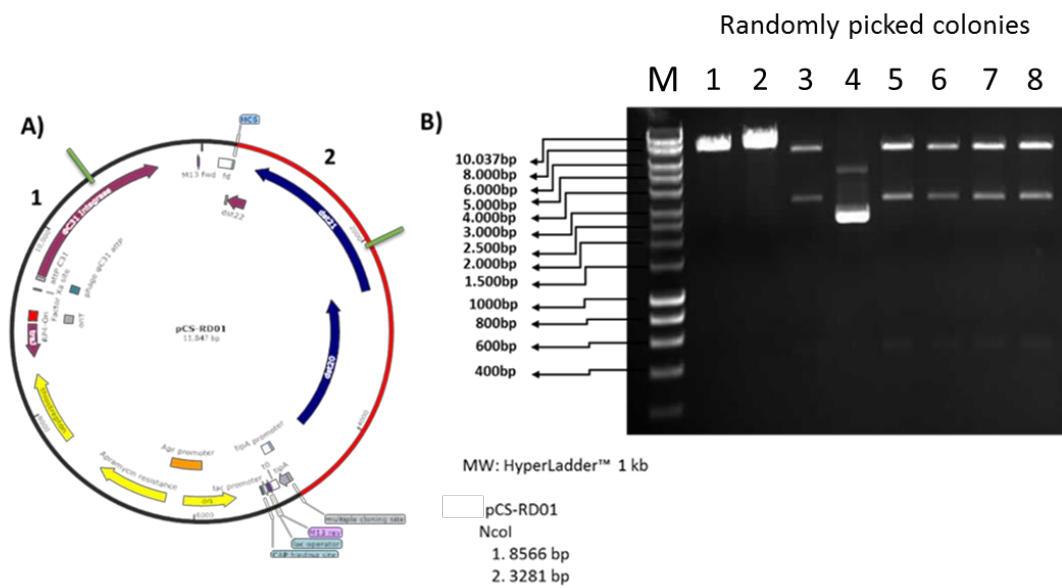


Figure 60. The integrative plasmid pCS-RD01, containing the distamycin efflux system

A) Plasmid map of pCS-RDC01, derivative of pIJ6902 containing the two putative genes for Distamycin resistance *dst20* and *dst21* (Hao *et al.*, 2014; Gerbaud, 2014), the map has two colours, black for the pIJ6902 backbone and red for the DNA cloned in; at the same time it is numbered in two fragments split by green bars corresponding to the expected restriction digest by *Nco*I. B) Agarose gel showing the digestion with *Nco*I (expected sizes below gel) of eight plasmid preparations from randomly picked *E. coli* colonies from the cloning of pCS-RD01 experiment.

Compound screening in the pump mutants

To determine the congocidine and distamycin MIC for the *S. lividans* strains, the carrier solvent DMSO was used as a control and varying concentrations of the compounds were assayed. Since there is no method developed to monitor *Streptomyces* sensitivity, the procedure presented in this work is novel. The approach used is adapted from the British Society for Antimicrobial Chemotherapy MIC determination in clinical settings (Andrews, 2001) and a high throughput method published in nature protocols (Wiegand *et al.*, 2008) suited for the growth requirements and measurement limitations of *Streptomyces*.

Due to the natural occurring pelleting of *Streptomyces* during liquid culture, some of the data present big variations. To give a better appreciation of these phenomena, pictures of the MIC determination plates are displayed on Fig.61A. For this reason statistical analysis were designed to avoid false discovery rates (Benjamini and Hochberg, 1995), as in every significant difference arising from the two way analysis of variance (ANOVA; Yates, 1934)(Gelman and Hill, 2006) was tested for one way ANOVA individually. Nevertheless as visible in Fig.61B, sometimes growth was not detected. This is especially pertinent in the case of the strain TKRD01 which demonstrates some limited congocidine resistance, although in some measurements its growth is not detected due to pelleting or adherence to the sides of the wells in the plate.

Along with a visual outlook of the plate, Fig.61 show a congocidine MIC₉₀ for *Streptomyces lividans* TK23/pIJ6902 of 50 μ M (survival rate after 48h < 10%). The mutant *S.lividans* TK23RC01 containing the congocidine transporter shows a survival rate of 40% with a congocidine concentration of 200 μ M, which roughly represents its MIC₅₀, showing at least a 4-fold increase in congocidine resistance. In the case of TK23RD01, the congocidine MIC was no different from the control strain, however growth is visible in the wells. This is later appreciated with data produced while assaying other compounds and using congocidine as a control for variability (Figs. 61, 62, 63). This data suggest that the strain has a congocidine MIC₅₀ of 100 μ M. In parallel the distamycin MIC was determined with the same methodology as for congocidine. However even assayed up to 400 μ M, distamycin did not affect the growth of any strain (Fig.61).

In light of these results and knowing the MICs of these compounds on pathogenic bacteria (results not published, data courtesy of MGB biopharma), it was decided to use 100 μ M for the remaining assay compounds. In order to do that, each well containing 198 μ L of the

culture was filled with 2 μ L containing 100 mM of the compound to assay dissolved in pure DMSO (final concentration of DMSO is 1%). The carrier solvent control were 3 wells with 2 μ L of pure DMSO per strain. The blank was performed with YEME without inoculum.

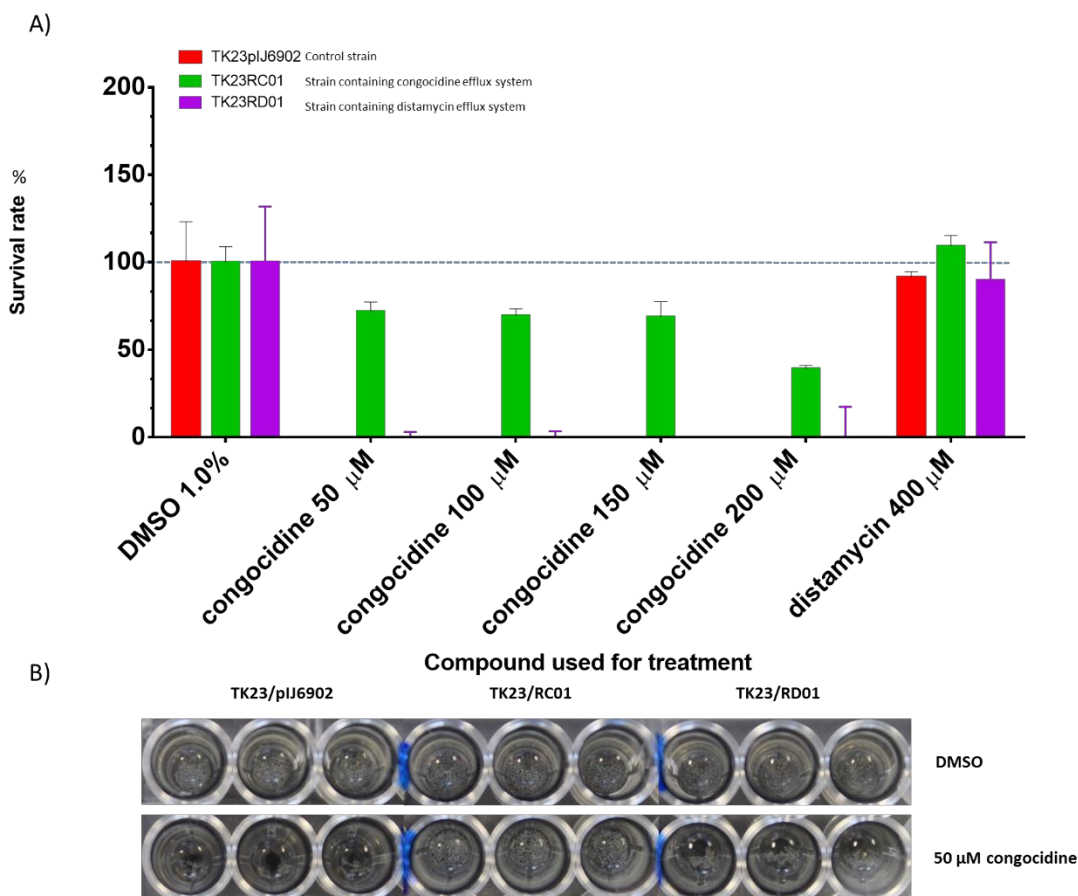


Figure 61. Congocidine MIC assay on three *Streptomyces* mutants

A) Survival rate (defined as the percentile coefficient between the specific growth rate of the strain treated with drug and carrier solvent (DMSO)) for the strains *Streptomyces lividans* TK23/pIJ6902, *Streptomyces lividans* TK23RC01 and *Streptomyces lividans* TK23RD01 after treatment with different concentrations (50-200 μM) of congocidine and distamycin (400 μM). Data normalised for DMSO at 1%.

B) Picture of the 96 well plate showing the pelleting of the strains.

Pump inhibition

To ensure that the resistance mechanism in the strains assayed was due to the effect of the ABC transporters introduced in them, two different pump inhibitors were tested in the presence of congocidine. The efflux pump inhibitors (EPI's) tested were 1-(1-naphthylmethyl)-piperazine (NMP) at 200 mg/L and phenylalanine-arginine beta-naphthylamide (PAβN) at 100 mg/L based on previous work on Gram negative pathogens (Kern *et al.*, 2006). Both EPI's show slight bacterial inhibition when applied alone but a synergetic effect was observed when added in presence of congocidine at 100 μM. Inclusion of the EPI's completely cancels the ability of the strain TK23RC01 to survive to congocidine at 100 μM, while without the presence of inhibitors it was able to grow at 200 μM. Results are visible in Fig.62.

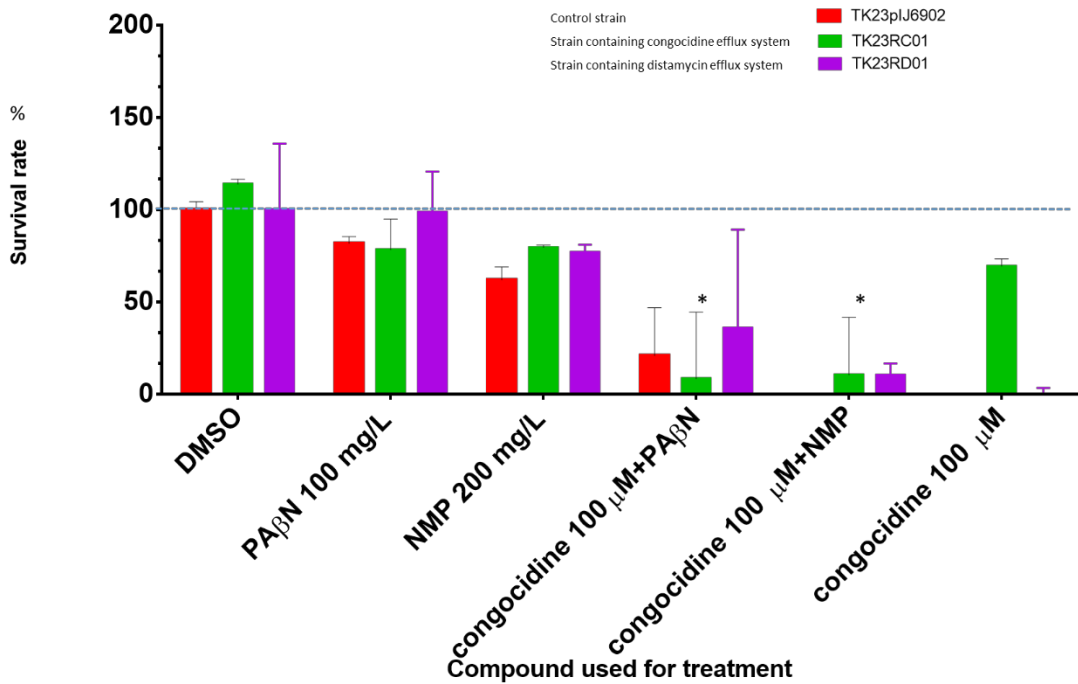


Figure 62. Effect of efflux pump inhibitors on three *Streptomyces* strains

Survival rate (defined as the percentual coefficient between the specific growth rate of the strain treated with drug and carrier solvent (DMSO)) for the strains *Streptomyces lividans* TK23/pJ6902, *Streptomyces lividans* TK23RC01 and *Streptomyces lividans* TK23RD01, after treatment with different concentrations of NMP, PAβN and Congocidine.

Non-natural pyrrole amide assay

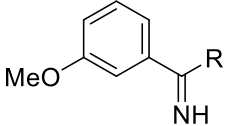
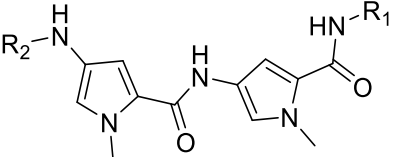
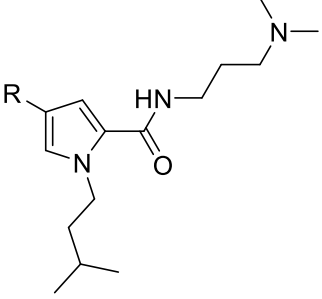
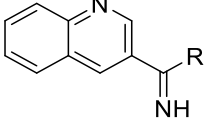
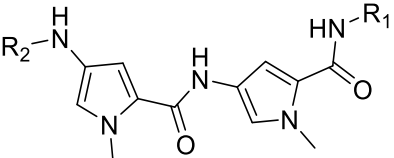
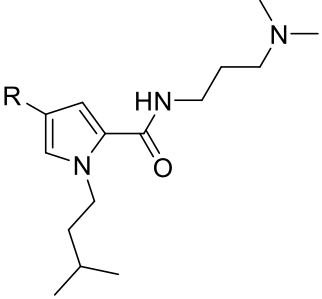

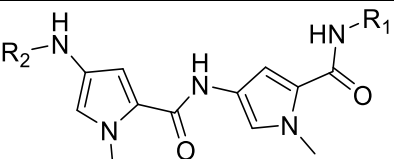
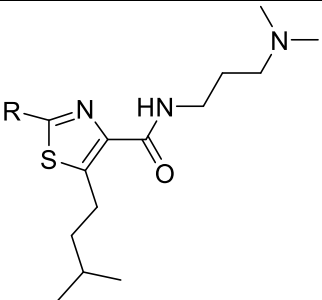
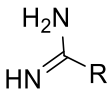
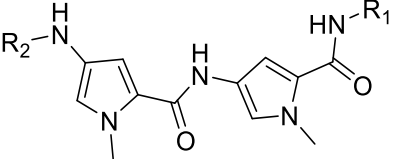
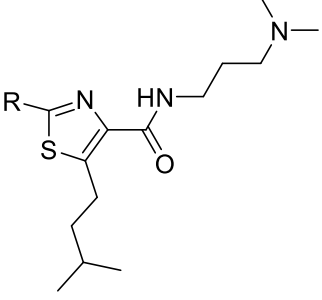
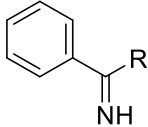
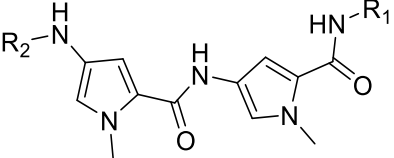
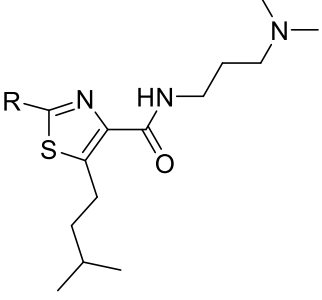
The compounds assayed were distributed in three major groups. The first is a group of distamycin like compounds, which structures are in the Table 14. The first five compounds have the same pyrrole amide core as distamycin, and differs from the parent compound by the substitution of the amidine group by a trimethylamine, a change that is constant among all compounds, and variable head groups. Compounds six to ten contains a substitution of the N-methyl in the pyrrole tail for an N-pentil-2-methyl moiety, presenting the same changes in the head groups as compounds one to five. Compounds 11 to 15, follow the same logic as compounds six to ten, but instead of substitution of the N-methyl in the pyrrole tail, the pentil-2-methyl is attached to a thiazole ring.

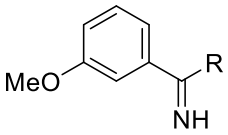
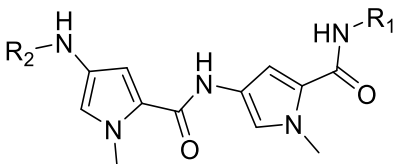
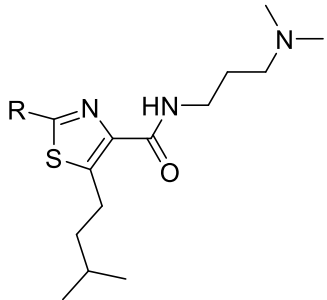
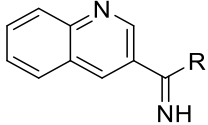
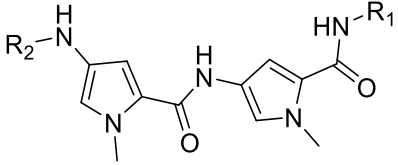
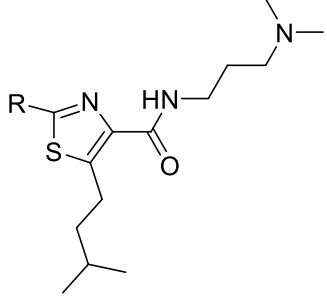
These substitutions explore the effect of increased lipophilicity on a set of head groups on a three ring amide. Results of the screening are visible in Fig.63.

Table 14. Distamycin-like functionalised compounds.

COMPOUND	R2	Minor Groove Binding region	R1
Netropsin/ Congocidine			
Distamycin			
1			
2			

3			
4			
5			
6			
7			
8			

9			
10			
11			
12			
13			

14			
15			

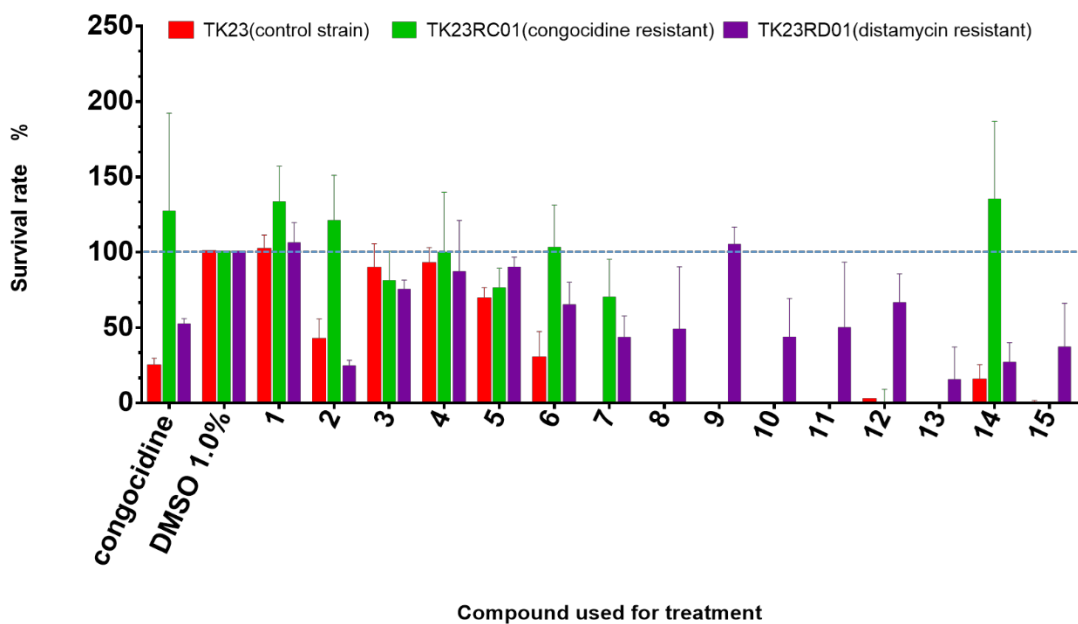



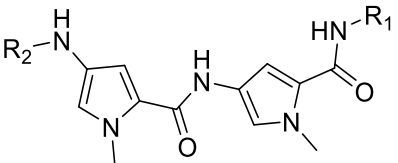
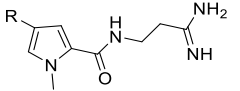
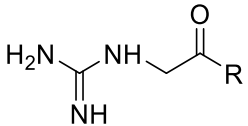
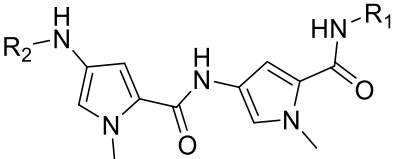
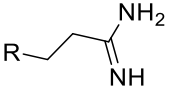
Figure 63. Response of congocidine and distamycin ABC efflux resistant strains in the presence of functionalised minor groove binders.

Survival rate (defined as the percent coefficient between the specific growth rate of the strain treated with drug and carrier solvent (DMSO)) for the strains TK23 or control strain (red), TK23RC01 or resistant to congocidine (green) and TK23RD01 or resistant to distamycin (purple), after treatment with a library of distamycin like compounds. Congocidine included as an internal control.

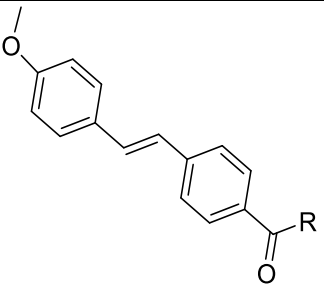
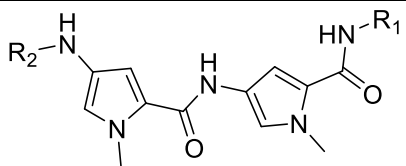
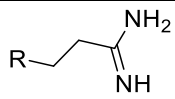
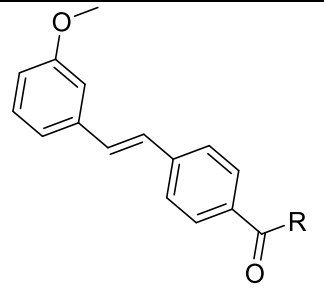
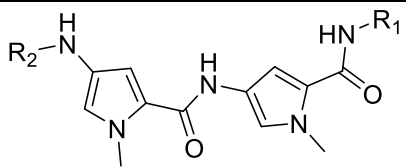
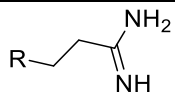
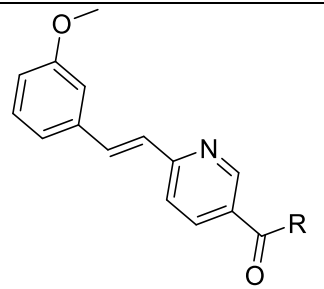
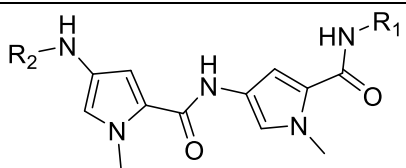
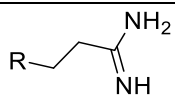
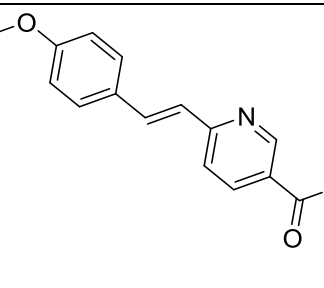
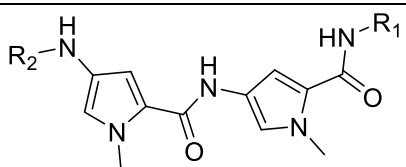
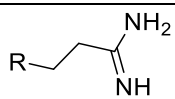
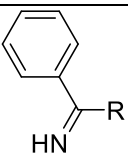
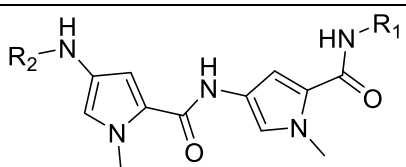
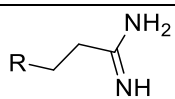
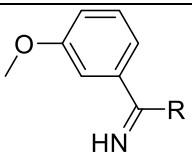
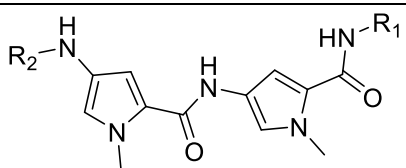
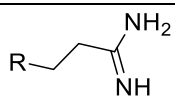
Interestingly compound number one, (the most similar to distamycin) behaves like distamycin in terms of bacterial inhibition, with no observable differences between the two ABC transporter strains and the control strain TK23+pIJ6902. In stark contrast to this, the inclusion of a guanidinium in the head group on compound two, (as it is present in the congocidine molecule), affects severely the control strain TK23+pIJ6902 and the distamycin resistant strain (still presenting minimum growth) , while the congocidine resistant strain grows without any impairment (comparing to solvent treated). While examining the compounds that possess the same moiety in the head group AIK-30/41 and AIK-30/26/1 a similar appreciation can be done. As the lipophilicity of the compounds increases, the control strain is completely inhibited, while the congocidine mutant strain is able to survive AIK-30/41 but not the most lipophilic compound AIK-30/26/1. However the distamycin strain performs almost identically with all three drugs. The trends emerging from this set are that increasing the lipophilicity of the compounds impacts negatively on the ability of survival of all three strains, while the distamycin strain adapts better to it.

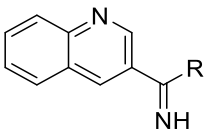
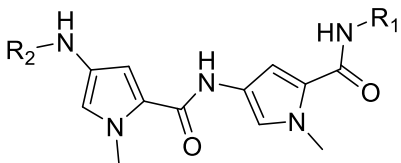
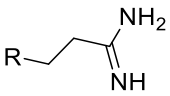
The second set of compounds are congocidine like functionalised compounds (although because of the size of some head groups, they could also be seen as three ring substituted compounds). For this set, two novel compounds not present in the library of MGB biopharma were synthesized. In this set, the pyrrole amide core as well as the tail group remain unchanged in the series, while varying the head groups. Results of the screening with this set are on Fig.64 and structures in Table 15.

Table 15. Congocidine-like functionalised compounds used.

COMPOUND	R2	Minor Groove Binding region	R1
Distamycin			
Netropsin/ Congocidine			

16			
17			
18			
19			
20			
21			
22			
23			

24			
25			
26			
27			
28			
29			

30			
----	---	--	---

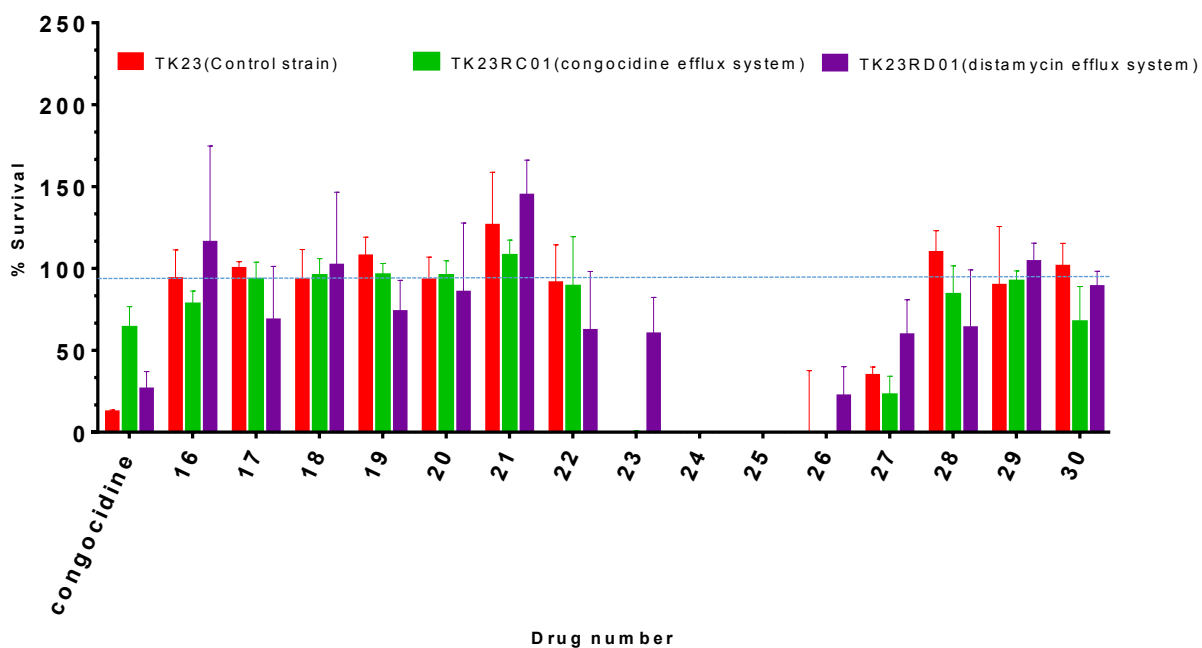


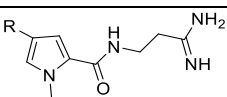
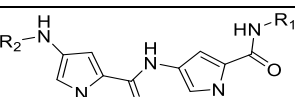

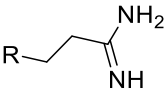
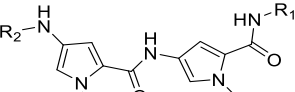
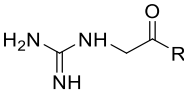
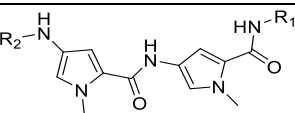
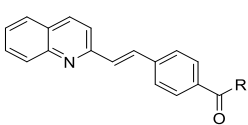
Figure 64. Response of congocidine and distamycin ABC efflux mutants in presence of functionalised minor groove binders.

Survival rate (defined as the percentual coefficient between the specific growth rate of the strain treated with drug and carrier solvent (DMSO)) for the strains TK23 or control strain (red), TK23RC01 or resistant to congocidine (green) and TK23RD01 or resistant to distamycin (purple), after treatment with a library of congocidine like compounds.

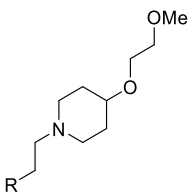
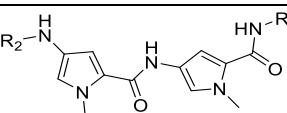
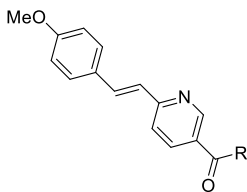
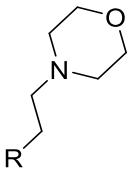
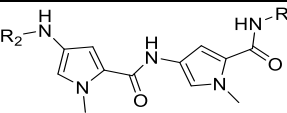
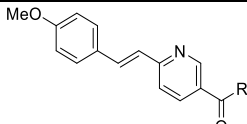
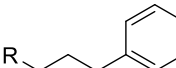
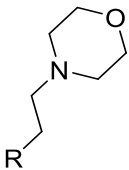
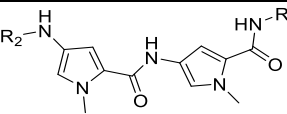
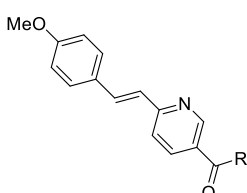
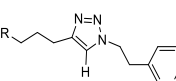
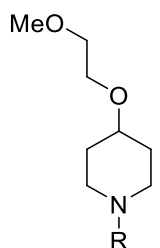
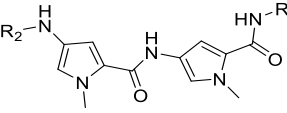
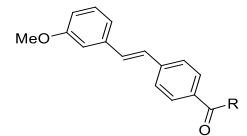
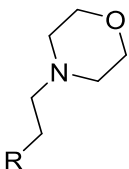
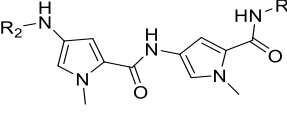
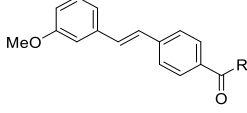
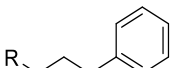
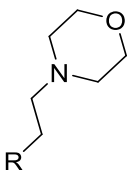
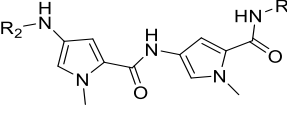
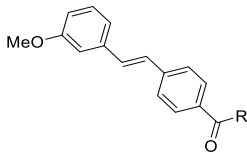
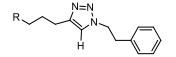
In this series, compounds 16-22 and 28-30 share the same pyrrole amide core as congocidine, but possess different head groups other than the guanidinium moiety present in the natural product. None of these compounds seem to be ineffective to kill any of the three strains. Only when the head group is bigger (compounds 23-27), and the compound turn into a three ring substituted amide is that the bacteria is affected. It is compelling though that the strain containing the distamycin efflux system survive better than the other two strains to treatment with compounds 23, 26 and 27. While compounds 26 and 25 completely kill all three strains (both compounds share structure apart from a meta / para methyl substitution in the terminal benzyl of the head group), 26 and 27 allows for a limited survival of the strain containing the distamycin efflux system, putatively indicating that a pyridine ring on the head group decreases the potency of the compound. At the same time, 50% survival of the strain containing the distamycin efflux system to compound 23 but not to compound 27 points towards the importance of the presence of methoxybenzene, even when the quinoline moiety is known to be present in several bioactive molecules (Prajapati *et al.*, 2014).

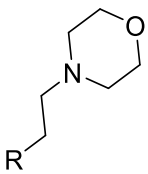
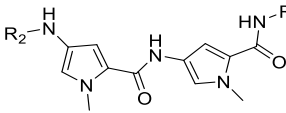
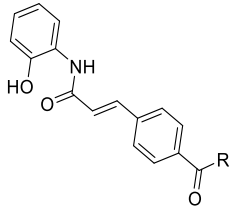
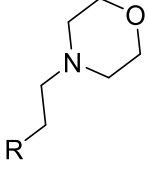
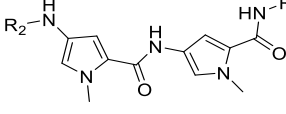
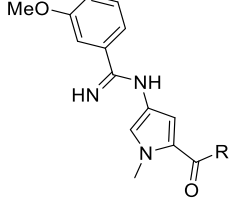

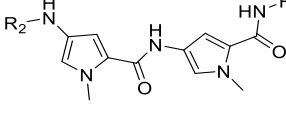
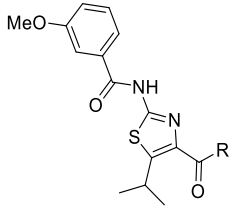
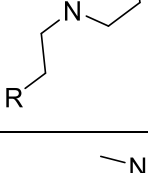
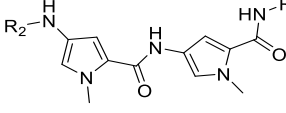
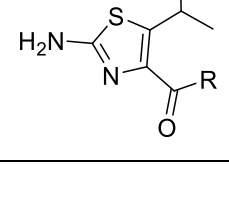
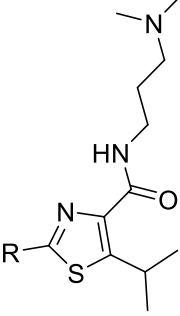
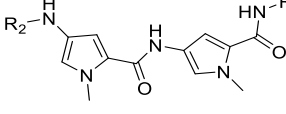
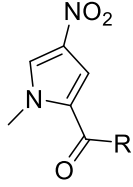
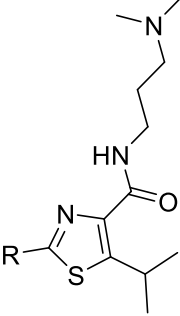
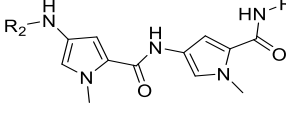
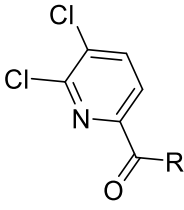
The last set is the most heterogeneous of all three, containing structures sharing a pyrrole amide core with congocidine but changing the tail and head group, as well as substitutions on the N-methyl pyrrole, creating bulkier compounds. At the same time smaller compounds with only one pyrrole ring are included. The structure of these compounds are represented in Table 16, and results of the screening are shown in Fig.65.

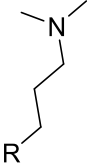
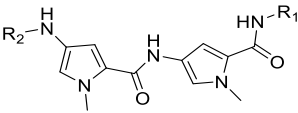
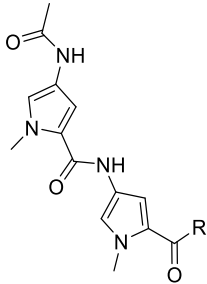

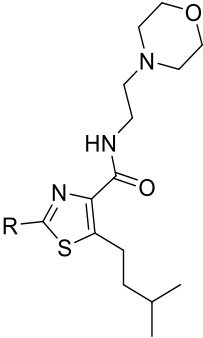
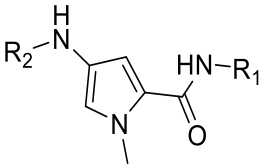
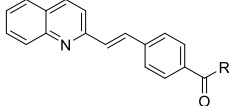
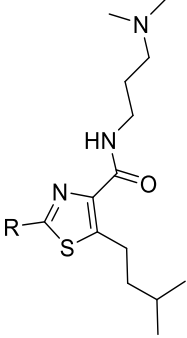
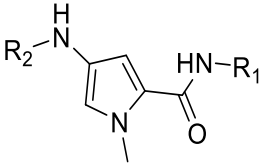
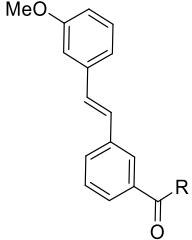
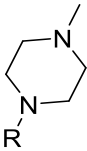
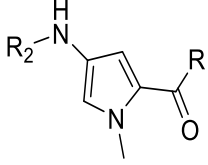
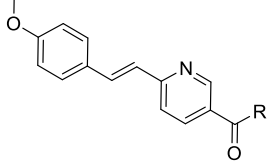

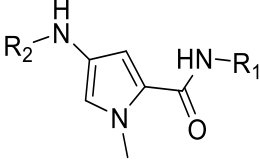
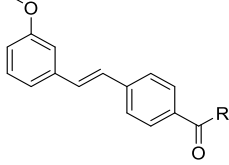
Table 16. Polyamides used for the treatment of *Streptomyces* strains.

Compound	R1	Minor Groove Binding	R2	R3
Distamycin				
Netropsin/ Congocidine				
31				

32				
33				
34				
35				
36				
37				

38				
39				
40				
41				
42				
43				

44			
45			
46			
47			
48			
49			

50				
51				
52				
53				
54				

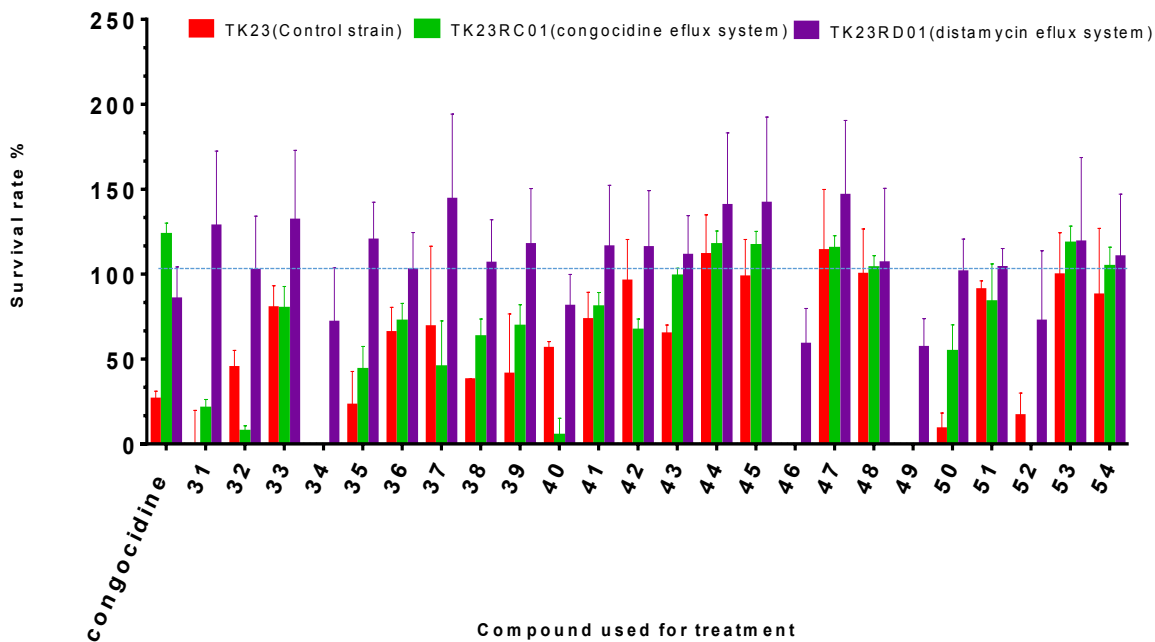


Figure 65. Response of congoicidine and distamycin ABC efflux mutants in presence of functionalised minor groove binders.

Survival rate (defined as the percentual coefficient between the specific growth rate of the strain treated with drug and carrier solvent (DMSO)) for the strains TK23 or control strain (red), TK23RC01 or resistant to congoicidine (green) and TK23RD01 or resistant to distamycin (purple), after treatment with a library of diverse polyamides.

Although repeated on several occasions, the assay should be taken with care as the variability due to the pelleting nature of *Streptomyces* allows for conflicting results such as the control strain growing significantly better than the strain containing the congoicidine resistance genes for compounds 32 and 40. Apart from these two compounds, the strains containing the efflux genes of congoicidine or distamycin perform better than the control strain for compounds having an adverse effect on growth. As in the previous two sets, the strain containing the distamycin efflux genes performs better overall.

Some of the compounds that present three substitutions seem not to affect the growth of any strain, perhaps reflecting uptake dynamics, due to their size. The compounds that significantly inhibit the growth of all three strains but allow the strain containing the distamycin efflux system to grow better are 31, 32, 34, 46, 49 and 52. Compounds 31 and 32 contains a quinoline head group in addition to a morpholino tail not present in the second set. Compounds containing quinoline in set two, inhibited the control and congoicidine mutant strain but not TKRD01. Intriguingly compound 33 being very similar to compound 34 (different alkene linking chain to the morpholino tail) do not affect any of the strains while the latter compound only allows strain TK23RD to grow. For the remaining compounds, since the group is so heterogeneous, proposing more SAR is difficult. While compounds 52 and 54 present features seen in compounds from the first and second group of is not possible to conclude if differences in growth after treatment are due to the different tail groups or the overall lipophilicity of the molecules.

Conclusions

While the strain containing the congocidine efflux system is better at surviving with congocidine, the data suggest that the strain containing the distamycin efflux system is able to tolerate a broader range of substrates. It appears that the congocidine ABC transporter is better suited to transporting specific molecules, and the guanidinium head group plays a fundamental role in congocidine efflux. On the other hand, the distamycin efflux system is able to expel bulkier and more lipophilic compounds. This along with the fact that distamycin is not toxic up to 400 μM and that the distamycin producing clusters also produces two other pyrrole amides, suggest that the efflux system in the distamycin cluster evolved towards a (multidrug resistance system) MDR. While some SAR have been outlined from the data sets one and two, a bigger library is needed to establish definitive conclusions over the results on the third set of compounds.

Discussion

The work described in this thesis was dedicated to exploring combinatorial biosynthesis and mutasynthesis as an approach to antibiotic discovery and functionalisation of the peptide congoicidine.

Although both approaches were described during the pioneering years of antibiotic discovery (Shier *et al.*, 1969; Hopwood *et al.*, 1985), these concepts have been revisited, applying current technologies such as DNA recombineering (Shao *et al.*, 2009) or Gibson cloning (Gibson *et al.*, 2009; Gibson *et al.*, 2010).

Combinatorial biosynthesis between distamycin and congoicidine clusters.

While current combinatorial approaches focus on the replacement of one gene with a mutagenized version of itself in a given cluster (Ru *et al.*, 2015) or swapping of analogue modules from similar pathways (Kakule *et al.*, 2014), these approaches require precise information about the pathway and the genes coding for it. However the first reported success of combinatorial biosynthesis (D.A Hopwood *et al.*, 1985) was the product of a “blind” whole cluster approach. In this first study, due to genetic manipulation limitations, a big part of the actinorhodin cluster, of which the size was arbitrarily determined by the presence of a *BamHI* site, was cloned into a medermycin producing strain. Because of a similar logic of assembly and the use of a common core compound, but with different tailoring enzymes, as Fig.66 show, the experiment brought to fruition novel derivatives. While exploring the expansion of natural minor groove binders due to their potential for therapeutic use, a putative pathway for congoicidine was already proposed (Juguet *et al.*, 2009), which later would be revised (Al-Mestarihi *et al.*, 2015), but the gene cluster responsible for distamycin biosynthesis was not proposed. However, following the same logic that Hopwood used, it was thought that due to the compound similarity, (Fig.66) congoicidine and distamycin may also share a core of biosynthetic enzymes, with additional enzymes governing the polymerisation of the pyrrole chain as well as the head-tail functionalisation. We hypothesised that mixing of the two gene sets would lead to assembly of a novel peptide. The distamycin producing strain was at the time not characterised other than its production of distamycin (Finlay. *et al.*, 1951) and aurothin (Akhunov and Otroshchenko, 1976), with protocols to achieve conjugative transfer being reported (Wang and Jin, 2014) two years after the first combinatorial biosynthetic experiments in this lab were performed.

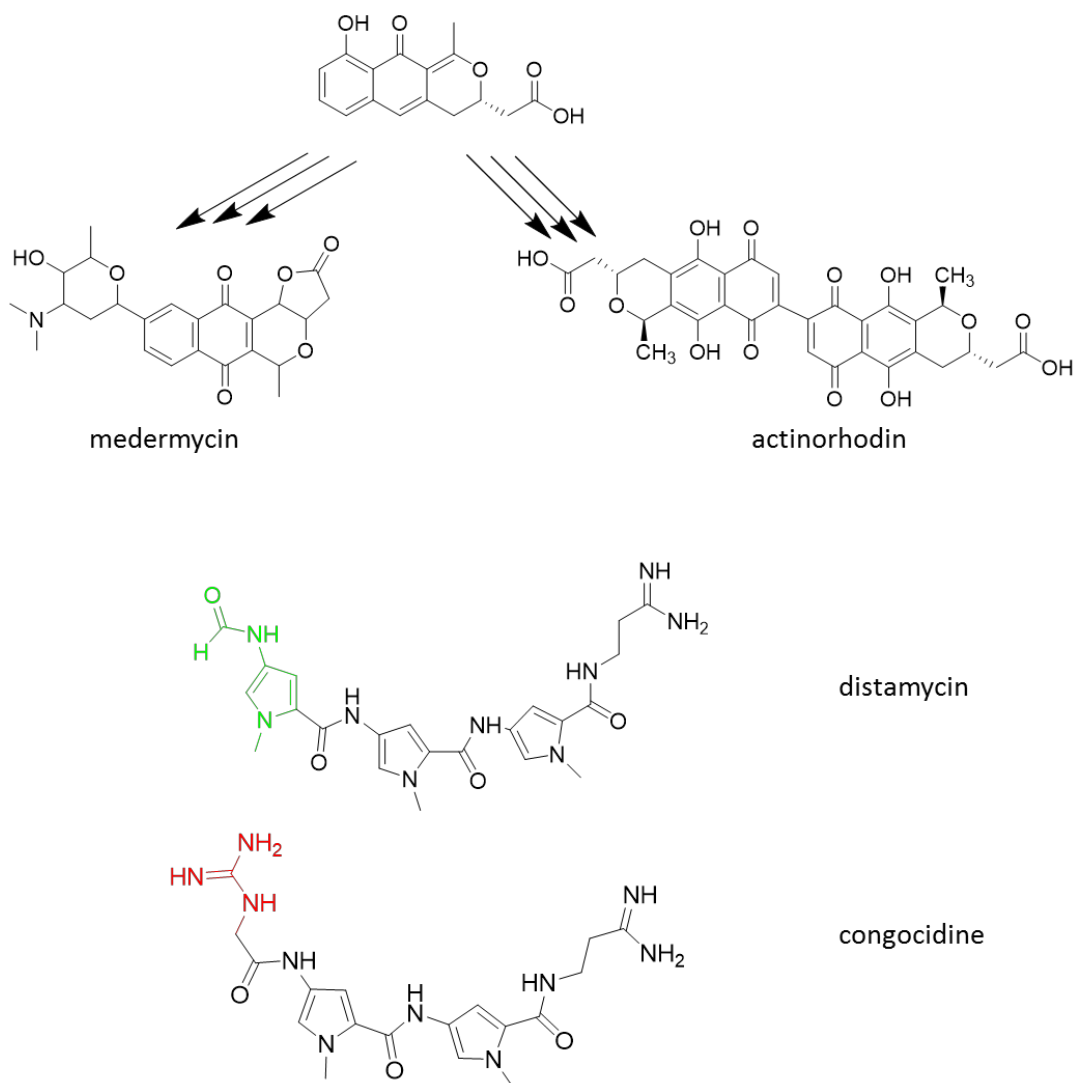


Figure 66. Medermycin and actinorhodin common precursor. Distamycin and congocidine similarities.

Common biosynthetic precursor of medermycin and actinorhodin (He *et al.*, 2015) leads to the production of different antibiotics. Peptides congocidine and distamycin also share structural core, with different chemical entities highlighted in green and red.

While it was possible to transfer 43.4 kb DNA from *S. ambofaciens* comprising the congocidine biosynthetic cluster including a pBeloBAC11 backbone used to integrate the pathway on the *attP* sites through Φ C31 recombinase activity, this caused genetic instability of the strain.

At first glance, the production of three novel compounds was achieved in sufficient amounts to purify and test their antimicrobial activity against *Bacillus subtilis*. However, in order to propose a chemical structure for this compounds through 2D NMR experiments, more compound was trying to be obtained, but the strain stopped production. Several factors were taken in consideration while changing the experimental design since some parameters may have changed but without success. To discard a deletion of the pathway, the strain was made again. However, the metabolic profile, while changing randomly, never re-produced the production of the three peaks that were observed once.

As the metabolic profile variability was investigated, isolation of four different phenotypes of the strain containing the congocidine pathway were isolated. At first they were thought to be contaminants, but 16S rDNA sequencing showed that all were *S. netropsis* strains.

It is thought that gene similarity and repetitive GC sequences on the clusters may have led to multiple recombination events between pathways. External DNA which is related to pathways within the cell genome may have triggered recombinase active in *Streptomyces netropsis* originating multiple hybrid pathways. Different prevalence rates, could explain the isolation of different morphotypes as well as the different metabolic profile. Perhaps some of the spores coming from the conjugative transfer plate contained one combination leading to the expression of these putatively novel compounds (or awakened pathway), while the average population does not. It would be possible to trace back the genetic elements accountable for the synthesis of these novel compounds, shall they be observed back in culture, through strain recovery and sequencing, but reproducing the result seems to lay on random parameters.

As it turns out, later research on the biosynthesis of distamycin (Hao *et al.*, 2014; Vingadassalon *et al.*, 2014a), has shown that in fact the production of distamycin is governed by not one but two biosynthetic clusters. While antibiotic pathways in bacteria are generally clustered, in the case of distamycin, the two clusters necessary for peptide production are separated by 58 Kb. Interestingly, in the surroundings of the genetic cluster most similar to congocidine from these two clusters, there is a putative kirromycin cluster and a region rich in phage-related proteins (Vingadassalon *et al.*, 2014a). These facts along with the

phylogenetic trees produced for the analysis of the resistance mechanisms in both clusters (chapter 4) points toward a horizontal acquisition of a congocidine-like biosynthetic cluster which led to natural combinatorial biosynthesis, keeping the novel genes and deleting the redundant ones, leading to the synthesis of different pyrrole amides.

After these results, trying to induce a unique controlled combinatorial event in *S.netropsis* is very complicated, as genetic elements or processes present the strain are not known. The compounds that were produced could may be the product of pleiotropic changes on the strain leading to activation and expression of silent metabolite clusters on it.

Perhaps, the key for obtaining these compounds again, requires sequencing of the strain to understand its metabolic capabilities. An overview of the biosynthetic clusters on the genome could provide cues of pathways that can be 'awakened'. Knowledge of other genes could also allow understanding of the observed genetic instability. Another approach would be using the congocidine and distamycin clusters in combinatorial experiments generated *in vitro* and expressed in different known heterologous host strains.

Synthetic assembly of the congocidine cluster

Since the whole cluster combinatorial biosynthetic approach was too variable, novel synthetic DNA assembling techniques were applied to pursue a synthetically built congocidine cluster, in which the genes constituting the cluster could be refactored on DNA modules. This design would allow for quick combinatorial biosynthetic experiments, or gene knockouts for mutasynthetic experiments. At the same time, possessing a library of genes encoding interesting enzymes such as N-methyl SAM methyl transferase or a standalone pyrrole carrier protein could serve for combinatorial experiments with unrelated clusters.

Refactoring of genetic clusters into plug and play scaffolds is currently a hot topic due its possibilities (Shao *et al.*, 2013; Smanski *et al.*, 2016), and while the success has been reported in previously characterised well-known pathways, the novel assembling techniques pitfalls are not reported. In a perspective article in nature chemical biology (Kim *et al.*, 2015) techniques such as linear-plus-circular homologous recombination; LCHR, linear-plus-linear homologous recombination; LLHR, sequence and ligation independent cloning; SLIC, transformation associated recombination; TAR, and DNA assembler (or recombineering) are presented and praised as a platform for antibiotic discovery.

While successful use of some of these assembling techniques ((Shao *et al.*, 2009; Zhao and Shao, 2009)) were inspiration to produce a congocidine refactored biosynthetic cluster, and other pathways had been cloned and refactored or modified (Yamanaka *et al.*, 2014; Montiel *et al.*, 2015) with the same yeast technology, after several attempts with Gibson assembly, DNA assembler and polymerase cycling assembling approaches, results have shown that these techniques are template-dependant, as there are no universal guidelines for DNA assembly (Huimin Zhao and Jeff Boeke personal communication) and the use of each one depends on the assembly being attempted, and needs to be assessed individually by trial-error. At the same time, these methods relies hugely on individual customisation, and are not yet universal. Unfortunately, often the reports in literature appear to be the exceptions of the norm. While constant advances in the field are achieved, it is thought that NRPS containing gene clusters (usually containing high G+C content and repetitive sequences) are difficult to assemble.

Mutasynthesis of congocidine.

Incorporation of non-native precursors on natural product biosynthesis have been exploited to generate diversity. Pathways containing PKS are often more successful in incorporating non-natural moieties (Winter and Tang, 2014), perhaps due to better enzymatic knowledge of these pathways. Nevertheless, NRPS or NRPS-PKS systems have also successfully incorporated synthetically made precursors, with reports on incorporation of propargylic click reaction handles (Kries *et al.*, 2014). While on some of these works, mutagenesis of the acceptor enzyme is needed, some NRPS and PKS pathways contains highly promiscuous enzymes (Li *et al.*, 2010; Werneburg *et al.*, 2010; Crawford *et al.*, 2011; Winter *et al.*, 2013; Xie *et al.*, 2014; Sato *et al.*, 2016), perhaps as a natural way to achieve diversity, and were exploited by scientists to produce novel molecules.

The congocidine assembly, also pointed towards enzymatic promiscuity on the congocidine cluster (Al-Mestarihi *et al.*, 2015). Based on the precursor study of congocidine biosynthesis, (Lautru *et al.*, 2012), we proposed the use of surrogate heterocycles instead of the natural pyrrole substrate for feeding to the strain containing the pyrrole producing enzyme knock out.

The natural precursor 4-acetamidopyrrole-2-carboxylate, used to complement the knock-out mutant in the previous study (Lautru *et al.*, 2012), was used as a positive control. Since congocidine is N-methylated after the peptide is formed, we explored incorporation of the closest non-natural pyrrole substrate to congocidine, by feeding of the 4-acetamido-1-methyl-1H-pyrrole-2-carboxylic acid and the deuterium labelled 4-acetamido-1-(methyl-d3)-1H-pyrrole-2-carboxylic acid. However results are not conclusive, and although some enzymes within the cluster may appear to be accepting surrogate precursors, more research has to be done.

Feeding of heterocycles with different electronic configuration such as imidazole and thiazole did not result on incorporation. Whether if this was due to a catalytic affinity problem due to electrostatic forces and/or steric impediment or impossibility of deacetylation later in the biosynthetic cascade is unknown. While trying to perform the feeding of the imidazole and thiazole building blocks again, the positive control feeding of the natural precursor 4-acetamidopyrrole-2-carboxylate ceased to work, so it is plausible that incorporation of these heterocycles is possible, but as congocidine production was not restored upon feeding of the positive control, the work should be repeated.

Pyrrole-amide ABC transporter study.

Crystallisation of the bacterial multidrug exporter AcrB from *E.coli* in 2002, greatly advanced the mechanistic knowledge of drug efflux (Murakami *et al.*, 2002). Since then, several pathogenic bacteria had been reported to possess this resistance mechanisms and their tolerance profile has been determined by screening against several antibiotics (Nishino *et al.*, 2009; Kosmidis *et al.*, 2012). As some studies suggested that drug efflux plays a key role in antibiotic efficiency (Lomovskaya and Bostian, 2006) (for new and old drugs), the use of efflux pump inhibitors (EPI's) was extensively studied. Phenyl-arginine beta-naphthylamide (PA β N) was the first identified EPI that inhibits all four clinically relevant *P. aeruginosa* efflux systems (Kourtesi *et al.*, 2013). However the use of this molecule in the clinic remains questionable, as PA β N affects the membrane integrity of bacteria and thus can lead to modification of its lipopolysaccharides (Lomovskaya *et al.*, 2001), creating undesired resistance profiles by reducing drug penetration. Another extensively used EPI is naphthylpiperazine, which has been shown to increase the intracellular accumulation of several antibiotics although it is too toxic for clinical use due to serotonin agonistic properties (Pagès and Amaral, 2009).

While genetic overexpression of these efflux systems has been studied, there is a lack of understanding of how EPI's can block them, or the ability to predict which drugs would be subject to efflux. The research done in this thesis, presents an innovative approach to study efflux systems. By profiling efflux systems of antibiotic producing bacteria with compounds related to the antibiotics they produce, but with a rational approach to their functionalisation, the importance of certain moieties in drug efflux can be understood. However the approach used to measure the activity of the drugs presents some difficulties experimental challenges. As the ABC systems to characterise were from *Streptomyces* they were cloned in *S.lividans* TK23, and cultured in 96 wells plates, in which their growth was measured. The problem of this approach lies on the tendency of *Streptomyces* to form cell pellets while in liquid culture, producing highly variable growth data. This phenomenon was in part avoided by spacing the measurements over long periods of time, and calculating the specific growth. The initial characterisation with congocidine showed that the control strain was unable to survive to concentrations of 50 μM , while the congocidine ABC transporter containing strain did survive to at least four-fold the concentration. Addition of the aforementioned EPI's showed a decreased capacity of survival in all strains, but only when co-cultivated with the peptide congocidine, proving that the strains prepared for the study were surviving due to efflux mechanisms. Interestingly, distamycin while reporting antibiotic activity against some bacteria (Sanfilipo *et al.*, 1966), did not show any effect on the growth of the control strain in concentrations as high as 400 μM . The presence of a efflux ABC transporter in the distamycin cluster might seem unjustified then, but as research has shown, the bi-cluster that produces distamycin, also assembles other pyrrole peptides (Vingadassalon *et al.*, 2014a). It could be just an exporting system for the bacteria to expel those peptides to the surroundings, in which competitive bacteria could be affected, while sparing symbiotic bacteria.

At the same time, some compounds that were predicted to be toxic for the cells (due to key moieties seen in bioactive compounds), were not bio-active, but this was perhaps due to inability of the drug to permeate to the strain. Future work can be done to test this hypothesis, either by displacement fluorescence microscopy (using the natural fluorescence of these compounds), or by cell extraction and quantification. Another not expected result was that compound 33 which is very similar to compound 34, did not affect any of the strains while the latter compound only allowed the distamycin efflux overexpressing strain to grow. Nevertheless a high number of compounds inhibited the control strain as well as the

congoicidine overexpressing strain, while not causing a growth alteration the strain possessing the distamycin ABC transporter. At the same time, the inclusion of a head guanidinium group seemed to bring potency to the compound, while the strain bearing the congoicidine resistance genes, would be able to expell it. These results points toward a distinct evolutionary divergence on both transporters, with the distamycin one, evolving towards a MDR efflux system.

Conclusions and Future work

While synthetic DNA manipulation technologies have improved significantly, and novel methods are reported frequently, rich G+C DNA templates are still difficult to edit precisely perhaps due to stronger affinity and tendency of form secondary structures due to frequent repeats. With the current technology, perhaps the best pathway capturing approach is based on yeast recombining (and novel TAR cloning approaches), and high G+C cluster editing should rely on CRISP/CAS and λ -Red systems. For the assembly of small cassettes destined to produce mutations in particular enzymes, the use of polymerase cycling assembly could be used. Independently of the DNA manipulation technique used, combinatorial biosynthesis can be performed in biosynthetic clusters. While it can be argued that introducing whole antimicrobial clusters in *Streptomyces* strains could lead to accelerated evolution of the strain, and perhaps creation of novel pathways or activation of dormant clusters from pleiotropic regulatory effects, those experiments should be performed when knowing the genetic imprint of the parental strain, and studying the arising transcriptomic changes. In this study, although compounds not present in the parental *S.netropsis* were detected. Since this detection could not be repeated, no proof of these hypothesis was achieved, and thus the recommended approach for combinatorial biosynthesis is to perform the experiments on known heterologous hosts which are easier to work with, after years of experience and optimisation.

On the other hand, the mutasynthetic experiment showed tendency of the enzymes in the congoicidine pathway to incorporate non-natural pyrroles. However the irreproducibility of the results and the control experiment calls to further explore these experiments. Perhaps saturation enzymatic assays could be done to every enzyme in the pathway to assess the affinity for non-natural substrates and further characterise the congoicidine pathway. At the same time, these essays could be done while studying the effect of point mutations and the effect of those in the affinity for novel substrates.

The profiling of the efflux systems on congoicidine and distamycin showed some clear structure-activity relationships, although to fully understand the mechanisms governing the efflux of these pumps, a more systematic and bigger set of compounds should be made and assayed.

References

- Akhunov, A., Tokhtamuratov, T., Otroshchenko, O.S., and Sagandykov, R. (1975) *Actinomyces netropsis*. *Khimiya Prir Soedin* **615.33**: 114–115.
- Akio, F., and Hamao, U. (1968) SYNTHESIS OF TETRA-N-PHENYLALKYLKANAMYCINS. *J Antibiot (Tokyo)* **XXI**: 340–349.
- Alam, M.T., Merlo, M.E., Hodgson, D. a, Wellington, E.M.H., Takano, E., and Breitling, R. (2010) Metabolic modeling and analysis of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics* **11**: 202.
- Al-Mestarihi, A.H., Garzan, A., Kim, J.M., and Garneau-Tsodikova, S. (2015) Enzymatic Evidence for a Revised Congocidine Biosynthetic Pathway. *ChemBioChem* **16**: 1307–1313.
- Altschup, S.F., Gish, W., Pennsylvania, T., and Park, U. (1990) Basic Local Alignment Search Tool. *J Mol Biol* **215**: 403–410.
- Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–169.
- Andrews, J.M. (2001) Review Activity of phenothiazines against antibiotic-resistant. *J Antimicrob Chemother* **48**: 5–16.
- Ankenbauer, R.G., Staley, A.L., Rinehart, K.L., and Cox, C.D. (1991) Mutasynthesis of siderophore analogues by *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* **88**: 1878–1882.
- Annereau, J., Wulbrand, U., Vankeerberghen, A., Cuppens, H., Bontems, F., Tümmeler, B., *et al.* (1997) A novel model for the first nucleotide binding domain of the cystic fibrosis transmembrane conductance regulator. *FEBS Lett* **407**: 303–308.
- Anthony, N.G., Johnston, B.F., Khalaf, A.I., Mackay, S.P., Parkinson, J.A., Suckling, C.J., and Waigh, R.D. (2004) Short Lexitropsin that Recognizes the DNA Minor Groove at 5' -ACTAGT-3' : Understanding the Role of Isopropyl-thiazole. *JAmChem Soc* **126**: 11338–11349.
- Arcamone, F., Penco, S., Orezzi, P., Nicoletta, V., and Pirelli, A. (1964) Structure and Synthesis of Distamycin A. *Nature* **203**: 1064–1065.
- Baltz, R.H. (2008) Renaissance in antibacterial discovery from actinomycetes. *Curr Opin Pharmacol* **8**: 557–63.
- Bambeke, F. Van, Balzi, E., and Tulkens, M. (2000) Antibiotic Efflux Pumps Franc. *Biochem Pharmacol* **60**: 457–470.
- Bao, K., and Cohen, S.N. (2003) Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. *GENES Dev* **17**: 774–785.
- BBSRC (2016) Synthetic Biology. <http://www.bbsrc.ac.uk/news/topic/corporate-synthe>.

- Beaucage, S.L., and Iyer, P. (1991) Advances in the synthesis of oligonucleotides by the phosphoramidite approach. *Tetrahedron* **48**: 2223–2311.
- Behrens, O. (1949) *The Chemistry of Penicillin*. Princeton University Press, .
- Benatuil, L., Perez, J.M., Belk, J., and Hsieh, C.M. (2010) An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel* **23**: 155–159.
- Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* **57**: 289 – 300.
- Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, a-M., Challis, G.L., Thomson, N.R., James, K.D., et al. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141–147.
- Berardo, C. Di, Capstick, D.S., Bibb, M.J., Findlay, K.C., Buttner, M.J., and Elliot, M.A. (2008) Function and Redundancy of the Chaplin Cell Surface Proteins in Aerial Hypha Formation , Rodlet Assembly , and Viability in. *J Bacteriol* **190**: 5879–5889.
- Bershtein, S., and Tawfik, D.S. (2008) Ohno ' s Model Revisited : Measuring the Frequency of Potentially Adaptive Mutations under Various Mutational Drifts. *Mol Biol Evol* **11**: 2311–8.
- Bertani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* **62**: 293–300.
- Bibb, M.J. (2005) Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol* **8**: 208–215.
- Bignell, D.R.D., Seipke, R.F., Huguet-Tapia, J.C., Chambers, A.H., Parry, R.J., and Loria, R. (2010) *Streptomyces scabies* 87-22 contains a coronafacic acid-like biosynthetic cluster that contributes to plant-microbe interactions. *Mol Plant Microbe Interact* **23**: 161–75.
- Bimboim, H.C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**: 1513–1523.
- Blin, K., Medema, M.H., Kazempour, D., Fischbach, M. a, Breitling, R., Takano, E., and Weber, T. (2013) antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* **41**: W204–W212.
- Bobola, M.S., Smith, N.W., Goff, R.D., Kolstoe, D.D., Blank, A., Gold, B., and Silber, J.R. (2007) Cancer Therapy : Preclinical Human Glioma Cell Sensitivity to the Sequence-Specific Alkylating Agent Methyl-Lexitropsin. *Cancer Ther Preclin* **13**: 612–620.
- Boddy, C.N., Hotta, K., Tse, M.L., Watts, R.E., and Khosla, C. (2004) Precursor-Directed Biosynthesis of Epothilone in *Escherichia coli*. *JAmChem Soc* **126**: 7436–7437.
- Bose, D.S., Hartley, J.A., and Yeidle, S. (1992) Rational Design of a highly efficient irreversible DNA interstrand cross-linking agent based on the pyrrolobenzodiazepine ring system.

JAmChem Soc **7**: 4939–4941.

Boumendjel, A., Boutonnat, J., and Robert, J. (2009) *ABC transporters and multidrug resistance*. JOHN WILEY & SONS, INC., PUBLICATION, .

Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D. (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**: 115–132.

Bravo-Rodriguez, K., Klopries, S., Koopmans, K.R.M., Sundermann, U., Yahiaoui, S., Arens, J., *et al.* (2015) Substrate Flexibility of a Mutated Acyltransferase Domain and Implications for Polyketide Biosynthesis. *Chem Biol* **22**: 1425–1430.

Brendel, N., Kirchner, H., Dahse, H., and Hertweck, C. (2011) Assessing Oxazole Bioisosteres as Mutasynthons on the Rhizoxin Assembly. *ChemBioChem* **12**: 2284–2288.

Buchanan, G.O., Williams, P.G., Feling, R.H., Kauffman, C. a., Jensen, P.R., and Fenical, W. (2005) Sporolides A and B: Structurally unprecedented halogenated macrolides from the marine actinomycete *Salinispora tropica*. *Org Lett* **7**: 2731–2734.

Bull, A.T., and Stach, J.E.M. (2007) Marine actinobacteria : new opportunities for natural product search and discovery. *Trends Microbiol* **15**: 491–499.

Burke, D.T., Carle, G.F., and Olson, M. V (1985) Large Segments of Exogenous DNA into Yeast by Means of Artificial Chromosome Vectors. *Science (80-)* **236**: 806–836.

Cane, D.E., and Haruo, I. (2013) Exploration and Mining of the Bacterial Terpenome. *Acc Chem Res* **10**: 54–56.

Chakrabarti, R., and Schutt, C.E. (2001) The enhancement of PCR amplification by low molecular weight amides. *Nucleic Acids Res* **29**: 2377–2381.

Chakraburttty, 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. *J Bacteriol* **3**: 5854–5861.

Charles, P.G.P., and Grayson, L. (2004) Worried and What We Can Do About It. *Med J Aust* **181**: 549–553.

Chater, K.F. (1998) Taking a genetic scalpel to the *Streptomyces* colony. *Microbiology* **3**: 1465–1478.

Check Hayden, E. (2015) Pint-sized DNA sequencer impresses first users. *Nature* 2–3.

Check-Hayden, E. (2014) The \$1,000 genome. *Nature* 4–5.

Chen, M.M.Y., Snow, C.D., Vizcarra, L., Mayo, S.L., and Arnold, F.H. (2012) Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Eng Des Sel* **25**: 171–178.

- Chena, C.-Y., Georgiev, I., Anderson, A.C., and Bruce, R., D. (2009) Computational structure-based redesign of enzyme activity. *Proc Natl Acad Sci* **106**: 7678–7678.
- Christiane, B., and Attila, K. (1999) Production of Nikkomycins Bx and Bz by Mutasynthesis with Genetically Engineered *Streptomyces tendae* TU901. *J Antibiot (Tokyo)* **52**: 102–108.
- Claessen, D., Rozen, D.E., Kuipers, O.P., and Sørensen, L. (2014) Bacterial solutions to multicellularity : a tale of biofilms , filaments and fruiting bodies. *Nat Publ Gr* 1–10.
- Clark, L.C., and Hoskisson, P. a. (2011) Duplication and evolution of devA-like genes in *Streptomyces* has resulted in distinct developmental roles. *PLoS One* **6**: 1–10.
- Clugston, S.L., Sieber, S. a, Marahiel, M. a, and Walsh, C.T. (2003) Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst. *Biochemistry* **42**: 12095–104.
- Coll, M., Frederick, C.A., Wang, A.H., and Rich, A. (1987) A bifurcated hydrogen-bonded conformation in the d (AT) base pairs of the DNA dodecamer d (CGCAAATTTGCG) and its complex with distamycin. *Proc Natl Acad Sci U S A* **84**: 8385–8389.
- Conti, E., Stachelhaus, T., Marahiel, M. a, and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo J* **16**: 4174–4183.
- Copley, S.D. (2003) Enzymes with extra talents: Moonlighting functions and catalytic promiscuity. *Curr Opin Chem Biol* **7**: 265–272.
- Copley, S.D. (2012) Moonlighting is mainstream: Paradigm adjustment required. *BioEssays* **34**: 578–588.
- Corre, C., Song, L., Rourke, S.O., Chater, K.F., and Challis, G.L. (2008) 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *coelicolor* genome mining. *Proc Natl Acad Sci U S A* **105**: 17510–17515.
- Cortes-sanchez, E., and Hoskisson, P.A. (2015) New approaches for new antibiotics. *Biochem (Lond)* 2–5.
- Coyle B M. (2005) *Manual of antimicrobial susceptibility testing* .
- Cragg, G.M., and Newman, D.J. (2013) Natural products: A continuing source of novel drug leads. *Biochim Biophys Acta - Gen Subj* **1830**: 3670–3695.
- Crawford, J.M., Portmann, C., Kontnik, R., Walsh, C.T., and Clardy, J. (2011) NRPS substrate promiscuity diversifies the xenomatides. *Org Lett* **13**: 5144–5147.
- Croop, J., and Housmant, D. (1986) Mammalian Multidrug Resistance Gene : Complete cDNA Sequence Indicates Strong to Bacterial Transport Proteins. *Cell* **47**: 371–380.
- Cundliffe, E. (1989) HOW ANTIBIOTIC-PRODUCING ORGANISM AVOID SUICIDE. *Annu Rev Genet* **43**: 207–233.

- D'Alia, D., Eggle, D., Nieselt, K., Hu, W.S., Breitling, R., and Takano, E. (2011) Deletion of the signalling molecule synthase ScbA has pleiotropic effects on secondary metabolite biosynthesis, morphological differentiation and primary metabolism in *Streptomyces coelicolor* A3(2). *Microb Biotechnol* **4**: 239–251.
- D'Costa, V.M. (2006) Sampling the Antibiotic Resistome. *Science* (80-) **311**: 374–377.
- Datsenko, K. a, and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640–6645.
- Daus, M.L., Berendt, S., Wuttge, S., and Schneider, E. (2007) Maltose binding protein (MalE) interacts with periplasmic loops P2 and P1 respectively of the MalFG subunits of the maltose ATP binding cassette transporter (MalFGK 2) from *Escherichia coli* / *Salmonella* during the transport cycle. *Mol Microbiol* **66**: 1107–1122.
- Davies, J., and Davies, D. (2010) Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**: 417–433.
- Dervan, P.B. (2001) Molecular Recognition of DNA by Small Molecules. *Bioorg Med Chem* **9**: 2215–2235.
- Dever, L.A., and Dermody, T.S. (1991) Mechanisms of Bacterial Resistance to Antibiotics. *Jama Intern Med* **151**: 886–895.
- Donadio, S., Mcalpine, J.B., Sheldont, P.J., Jackson, M., and Katz, L. (1993) An erythromycin analog produced by reprogramming of polyketide synthesis. *Proc Natl Acad Sci* **90**: 7119–7123.
- Doroghazi, J.R., Albright, J.C., Goering, A.W., Ju, K.-S., Haines, R.R., Tchalukov, K. a, *et al.* (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* 1–8.
- DSZM 65. Gym Streptomyces Medium. 1.
- Duncan, K.R., Crusemann, M., Lechner, A., Sarkar, A., Li, J., Ziemert, N., *et al.* (2014) Molecular Networking and Pattern-Based Genome Mining Improves Discovery of Biosynthetic Gene Clusters and their Products from *Salinispora* Species. *Chem Biol* 460–471.
- Elliot, M.A., Buttner, M.J., and Nodwell, J.R. (2008) *Multicellular Development in Streptomyces*. ASM Press .
- Eppelmann, K., Stachelhaus, T., and Marahiel, M. a (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**: 9718–26.
- Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, D., McMahon, M.D., and Thomas, M.G. (2011) Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *mol Pharm* **5**: 191–211.

- FINLAY., A., HOCHSTEINB, A., AND, A.S., and MURPHY, F.X. (1951) Netropsin, a New Antibiotic Produced by a *Streptomyces*. *JAmChem Soc* **73**: 341–343.
- Firn, R.D., and Jones, C.G. (2009) A Darwinian view of metabolism : molecular properties determine fitness. *J Exp Bot* **60**: 719–726.
- Fischbach, M.A., and Walsh, C.T. (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. *Chem Rev* **106**: 3468–3496.
- Flardh, K., and Buttner, M.J. (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Micro* **7**: 36–49.
- Foye, W.O. (1999) *Cancer Chemotherapeutic Agents* .
- Galm, U. et al (2004) In vitro and in vivo production of new aminocoumarins by a combined biochemical, genetic and synthetic approach. *Chem Biol* **11**: 173–183.
- Gelman, A., and Hill, J. (2006) Data analysis using regression and multilevel/hierarchical models. *Journal of Applied Statistics* **35**: 3-351-356.
- Gerber, N.N., and Lechevalier, H. a (1965) Geosmin, an earthy-smelling substance isolated from actinomycetes. *Appl Microbiol* **13**: 935–8.
- Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**: 52–56.
- Gibson, D.G., Smith, H.O., Iii, C.A.H., Venter, J.C., and Merryman, C. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nature* **7**.
- Gibson, D.G., Young, L., Chuang, R., Venter, J.C., Iii, C.A.H., and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature* **6**: 12–16.
- Gietz, R.D., and Schiestl, R.H. (2007) Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**: 38–41.
- Gillespie, E., Brady, S., Bettermann, A., Cianciotto, N., Mark, R., Rondon, M., et al. (2002) Isolation of antibiotic Turbomycin A and Turbomycin B from a metagenomic library. *Appl Environ Microbiol* **68**: 4301–4306.
- Girard, G., Traag, B. a, Sangal, V., Mascini, N., Hoskisson, P. a, Goodfellow, M., and Wezel, G.P. van (2013) A novel taxonomic marker that discriminates between morphologically complex actinomycetes. *Open Biol* **3**: 2–13.
- Goldsmith, M., and Tawfik, D.S. (2012) Directed enzyme evolution: beyond the low-hanging fruit. *Curr Opin Struct Biol* **22**: 406–12.
- Gomez-Escribano, J.P., and Bibb, M.J. (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol* **4**: 207–

Goodfellow, M., Kämpfer, P., Busse, H.-J., Trujillo, M.E., Suzuki, K., Ludwig, W., *et al.* (2012) *BERGEY'S MANUAL of Systematic Bacteriology*.

Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S., and Onufriev, A. (2005) H++: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **33**: W368–71.

Grant, S.G., Jessee, J., Bloom, F.R., and Hanahan, D. (1990a) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci U S A* **87**: 4645–4649.

Grant, S.G., Jessee, J., Bloom, F.R., and Hanahan, D. (1990b) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci U S A* **87**: 4645–4649.

Gregory, M., Petkovic, H., Lill, R., Moss, S., Wilkinson, B., Gaisser, S., *et al.* (2005) Mutasynthesis of rapamycin analogues through the manipulation of a gene governing starter unit biosynthesis. *Angew Chem Int Ed* **44**: 4757–4760.

Grüschow, S., Rackham, E.J., Elkins, B., Newill, P.L.A., Hill, L.M., and Goss, R.J.M. (2009) New Pacidamycin Antibiotics Through Precursor-Directed Biosynthesis. *ChemBioChem* **10**: 355–360.

Guenzi, E., Galli, G., Grgurina, I., Gross, D.C., Grandi, G., and Sapienza, I.G. (1998) Characterization of the Syringomycin Synthetase Gene Cluster. *J Biol Chem* **273**: 32857–32863.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* **100**: 1541–1546.

Gust, B., O'Rourke, S., Bird, N., Kieser, T., and Chater, K. (2006) Recombineering in *Streptomyces coelicolor*. *FEMS Online Protoc* 1–22.

Hampshire, A.J., Khairallah, H., Khalaf, A.I., Ebrahimabadi, A.H., Waigh, R.D., Suckling, C.J., *et al.* (2006) DNA sequence recognition by an imidazole-containing isopropyl-substituted thiazole polyamide (thiazotropsin B). *Bioorg Med Chem Lett* **16**: 3469–3474.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**: 557–580.

Hao, C., Huang, S., Deng, Z., Zhao, C., and Yu, Y. (2014) Mining of the Pyrrolamide Antibiotics Analogs in *Streptomyces netropsis* Reveals the Amidohydrolase-Dependent “Iterative Strategy” Underlying the Pyrrole Polymerization. *PLoS One* **9**: 1–9.

Hartmann, T. (2008) The lost origin of chemical ecology in the late 19th century. *Proc Natl Acad Sci U S A* **105**: 4541–4546.

- Haug, I., Weissenborn, A., Brolle, D., Bentley, S., Kieser, T., and Altenbuchner, J. (2003) *Streptomyces coelicolor* A3(2) plasmid SCP2*: deductions from the complete sequencedeductions from the complete sequence. *Microbiology* **3**: 505–513.
- He, J., Hertweck, C., and Jena, D.- (2003) Iteration as Programmed Event during Polyketide Assembly ; Molecular Analysis of the Aureothin Biosynthesis Gene Cluster. *Chem Biol* **10**: 1225–1232.
- He, Q., Li, L., Yang, T., Li, R., and Li, A. (2015) Functional characterization of a ketoreductase-encoding gene med-ORF12 involved in the formation of a stereospecific pyran ring during the biosynthesis of an antitumor antibiotic medermycin. *PLoS One* **10**: 1–17.
- Heckman, K.L., and Pease, L.R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* **2**: 924–932.
- Henke, W., Herdel, K., Jung, K., Schnorr, D., and Loening, S.A. (1997) Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res* **25**: 3957–3958.
- Hertweck, C. (2015) Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology. *Trends Biochem Sci* **40**: 189–199.
- Higgins, C.F., Hiles, I.D., Salmond, G.P.C., Gill, D.R., Downie, J.A., Evans, I.J., *et al.* (1986) A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature* **323**: 448–450.
- Higuchi, R., Krummel, B., and Saiki.k, R. (1988) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Nucleic Acids Res* **16**: 7351–7367.
- Hiltner, J.K., Hunter, I.S., and Hoskisson, P.A. (2015) Tailoring Specialized Metabolite Production in *Streptomyces*. *Adv Appl Microbiol* **91**: 237–255.
- Hobbs, G., Frazer, C.M., Gardner, D.C.J., Cullum, J. a., and Oliver, S.G. (1989) Dispersed growth of *Streptomyces* in liquid culture. *Appl Microbiol Biotechnol* **31**: 272–277.
- Hodgson, D.A. (2000) Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv Microb Physiol* **42**: 47–238.
- Hoffmann, G., Bo, K., and Greiner-sto, T. (2011) Changing the substrate specificity of P450cam towards diphenylmethane by semi-rational enzyme engineering. *Protein Eng Des Sel* **24**: 439–446.
- Hofmann, K., and Stoffel, W. (1993) TMbase: A Database of Membrane Spanning Protein Segments. *Biol Chem* **374**: 166.
- Holland, I., Cole, S.P.C., Kuchler, K., and Higgins, C.F. (2003) *ABC proteins: from bacteria to man*. Academic Press, San Diego, CA.
- Hollenstein, K., Dawson, R.J.P., and Locher, K.P. (2007) Structure and mechanism of ABC transporter proteins. *Curr Opin Struct Biol* **17**: 412–418.

- Hopwood, D. a (1967) Genetic analysis and genome structure in *Streptomyces coelicolor*. *Bacteriol Rev* **31**: 373–403.
- Hopwood, D., Malpartida, F., Kieser, H., Ikeda, H., Duncan, J., I, *et al.* (1985) Production of “hybrid” antibiotics by genetic engineering. *Nature*, **314**: pp. 642–644.
- Hopwood, D.A. (1997) Genetic Contributions to Understanding Polyketide Synthases. *Chem Rev* **97**: 2465–2498.
- Hopwood, D.A. (2007) *Streptomyces in Nature and Medicine: The Antibiotic Makers*. .
- Hopwood, D.A., Malpartida, F., Kieser, H.M., Ikeda, H., Duncan, J., Fujii, I., *et al.* (1985) Production of “hybrid” antibiotics by genetic engineering. *Nature* **314**: 642–644.
- Horinuchi, S., and Beppu, T. (2007) Hormonal control by A-factor of morphological development and secondary metabolism in *Streptomyces*. *Proc Jpn Acad* **83**: 277–295.
- Hozzein, W.N., and Goodfellow, M. (2007) *Streptomyces synnematoformans* sp . nov ., a novel actinomycete isolated from a sand dune soil in Egypt. *Int J Syst Evol Microbiol* **57**: 2009–2013.
- Hua, S., Qiu, M., Chan, E., Zhu, L., and Luo, Y. (1997) Minimum Length of Sequence Homology Required for in Vivo Cloning by Homologous Recombination in Yeast. *Plasmid* **38**: 91–96.
- Huang, J., Shi, J., Molle, V., Sohlberg, B., Weaver, D., Bibb, M.J., Karoonuthaisiri, N., Lih, C.-J., *et al.* (2005) Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Mol Microbiol* **58**: 1276–1287.
- Huang, J., Shi, J., Molle, V., Sohlberg, B., Weaver, D., Bibb, M.J., Karoonuthaisiri, N., Lih, C.J., *et al.* (2005) Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Mol Microbiol* **58**: 1276–1287.
- Hur, G.H., Vickery, C.R., and Burkart, M.D. (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* **29**: 1074–98.
- Hutchinson, C.R. (1995) POL YKETIDE SYNTHASE GENE MANIPULATION : A Structure-Function Approach in Engineering Novel Antibiotics. *Annu Rev Microbiol* 202–231.
- Hyde, S.C., Emsley, P., Hartshorn, M.J., Mimmack, M.M., Gileadi, U., Pearce, S.R., *et al.* (1990) Structural model of ATP-binding proteing associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature* **346**: 362–365.
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., *et al.* (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**: 526–531.
- International Human Genome Sequencing Consortium (2001) Finishing the euchromatic sequence of the human genome. *Nature* 931–945.
- Ishikawa, F., Miyamoto, K., Kasai, S., and Takeya, H. (2015) Accurate Detection of Adenylation

Domain Functions in Nonribosomal Peptide Synthetases by an Enzyme-linked Immunosorbent Assay System Using Active Site-directed Probes for Adenylation Domains. *ACS Chem Biol* .

Jacobsen, M.F., Moses, J.E., Adlington, R.M., and Baldwin, J.E. (2005) A Short Total Synthesis of Aureothin and N-Acetylaureothamine. *Org Biomol Chem* **9**: 9–12.

James, P.L., Merkina, E.E., Khalaf, A.I., Suckling, C.J., Waigh, R.D., Brown, T., and Fox, K.R. (2004) DNA sequence recognition by an isopropyl substituted thiazole polyamide. *Nucleic Acids Res* **32**: 3410–3417.

Jaszczyszyn, Y., Thermes, C., and Dijk, E.L. Van (2014) Ten years of next-generation sequencing technology. *Trends Genet* **30**: 418–426.

Jenke-kodama, H., Sandmann, A., Mu, R., and Dittmann, E. (2005) Evolutionary Implications of Bacterial Polyketide Synthases. *Mol Biol Evol* **10**: 2027–39.

Jirakkakul, J., Punya, J., Pongpattanakitsote, S., Paungmoung, P., Vorapreeda, N., Tachaleat, A., *et al.* (2008) Identification of the nonribosomal peptide synthetase gene responsible for bassianolide synthesis in wood-decaying fungus *Xylaria* sp. BCC1067. *Microbiology* **154**: 995–1006.

Jones, P.M., and George, A.M. (2000) Symmetry and structure in P-glycoprotein and ABC transporters What goes around comes around. *Eur J Biochem* **267**: 5298–5305.

Jones, P.M., and George, A.M. (2007) Cellular and Molecular Life Sciences The ABC transporter structure and mechanism : perspectives on recent research. *C Cell Mol Life Sci* **61**: 682–699.

Juguet, M., Lautru, S., Francou, F.X., Nezbedová, Š., Leblond, P., Gondry, M., and Pernodet, J.L. (2009) An Iterative Nonribosomal Peptide Synthetase Assembles the Pyrrole-Amide Antibiotic Congocidine in *Streptomyces ambofaciens*. *Chem Biol* **16**: 421–431.

Jukes, T.H., and Cantor, C.R. (1969) Evolution of protein molecules. *Mamm Protein Metab* **21**: 123.

Kaeberlein, T., Lewis, K., and Epstein, S.S. (2002) Isolating “uncultivable” microorganisms in pure culture in a simulated natural environment. *Science* **296**: 1127–1129.

Kakule, T.B., Lin, Z., and Schmidt, E.W. (2014) Combinatorialization of Fungal Polyketide Synthase – Peptide Synthetase Hybrid Proteins. *J Mol Biol* **136**: 17882–17890.

Kämpfer, P., Huber, B., Buczolits, S., Thummes, K., Grün-Wollny, I., and Busse, H.-J. (2008) *Streptomyces specialis* sp. nov. *Int J Syst Evol Microbiol* **58**: 2602–6.

Kang, J., Myung, S.L., and Gorenstein, D.G. (2005) The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: Application to in vitro combinatorial selection of aptamers. *J Biochem Biophys Methods* **64**: 147–151.

- Kaznessis, Y.N. (2007) Models for systems Biology. *BMC Syst Biol* **4**: 1–4.
- Keating, T. a, Marshall, C.G., and Walsh, C.T. (2000) Vibriobactin biosynthesis in *Vibrio cholerae*: VibH is an amide synthase homologous to nonribosomal peptide synthetase condensation domains. *Biochemistry* **39**: 15513–21.
- Keating, T. a, Marshall, C.G., Walsh, C.T., and Keating, A.E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol* **9**: 522–6.
- Keating, T.A., and Walsh, C.T. (1999) Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Curr Opin Chem Biol* **3**: 598–606.
- Kern, W. V, Steinke, P., Schumacher, A., Schuster, S., Baum, H. von, and Bohnert, J.A. (2006) Effect of 1-(1-naphthylmethyl)-piperazine, a novel putative efflux pump inhibitor, on antimicrobial drug susceptibility in clinical isolates of *Escherichia coli*. *J Antimicrob Chemother* **57**: 339–343.
- Khalaf, A.I. (2009) Minor groove binders : Some recent research in drug development. *Curr Trends Medical Chem* 53–63.
- Khorana, H.G., Buchi, H., Caruthers, M.H., and Gupta, N. k (1972) Total Synthesis of the Structural Gene for an Alanine Transfer Ribonucleic Acid from Yeast. *J Mol Biol* 209–217.
- Khosla, C., Tang, Y., Chen, A.Y., Schnarr, N.A., and Cane, D.E. (2007) Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annu Rev Biochem* **76**: 195–221.
- Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000) *Practical Streptomyces Genetics*. John Innes Centre, .
- Kim, E., Moore, B.S., and Joon Yoon, Y. (2015) Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat Chem Biol* **11**: 649–659.
- Kim, J., and Copley, S.D. (2007) Why metabolic enzymes are essential or nonessential for growth of *Escherichia coli* K12 on glucose. *Biochemistry* **46**: 12501–12511.
- Kimura, K., Iwatsuki, M., Nagai, T., Matsumoto, A., Takahashi, Y., Shiomi, K., *et al.* (2011) A small-molecule inhibitor of the bacterial type III secretion system protects against in vivo infection with *Citrobacter rodentium*. *J Antibiot (Tokyo)* **64**: 197–203.
- Kirby, R., Sangal, V., Tucker, N.P., and Zakrzewska-czerwin, J. (2012) Draft Genome Sequence of the Human Pathogen *Streptomyces somaliensis* , a Significant Cause of Actinomycetoma. **194**: 3544–3545.
- Kleinkauf, H., and Döhren, H. von (1990) Nonribosomal biosynthesis of peptide antibiotics. *Eur J Biochem* **192**: 1–15.
- Klementz, D., Döring, K., Lucas, X., Telukunta, K.K., Erxleben, A., Deubel, D., *et al.* (2015) StreptomeDB 2.0-an extended resource of natural products produced by *streptomyces*.

Nucleic Acids Res **44**: D509–14.

Kobayashi, Y., Daido, H., Katsuta, H., Nomura, M., Tsukada, H., Hirabayashi, A., *et al.* (2011) Amide derivative, pest control agent containing the amide derivative and use of the pest control agent.

Kok, S. De, Stanton, L.H., Slaby, T., Durot, M., Holmes, V.F., Patel, K.G., *et al.* (2014) Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth Biol* **3**: 97–106.

Konz, D., Klens, a, Schörgendorfer, K., and Marahiel, M. a (1997) The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol* **4**: 927–937.

Kopka, M.L., Goodsell, D.S., Han, G.W., Chiu, T.K., Lown, J.W., and Dickerson, R.E. (1997) Defining GC-specificity in the minor groove : side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. *Structure* 1033–1046.

Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R.E. (1985) The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc Natl Acad Sci U S A* **82**: 1376–80.

Koressaar, T., and Remm, M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**: 1289–1291.

Koryakina, I., McArthur, J.B., Draelos, M.M., and Williams, G.J. (2013) Promiscuity of a modular polyketide synthase towards natural and non-natural extender units. *Org Biomol Chem* **11**: 4449.

Kosmidis, C., Schindler, B.D., Jacinto, P.L., Patel, D., Bains, K., Seo, S.M., and Kaatz, G.W. (2012) Expression of multidrug resistance efflux pump genes in clinical and environmental isolates of *Staphylococcus aureus*. *Int J Antimicrob Agents* **40**: 204–209.

Kourtesi, C., Ball, A.R., Huang, Y.-Y., Jachak, S.M., Vera, D.M. a, Khondkar, P., *et al.* (2013) Microbial efflux systems and inhibitors: approaches to drug discovery and the challenge of clinical implementation. *Open Microbiol J* **7**: 34–52.

Kries, H., Wachtel, R., Pabst, A., Wanner, B., Niquille, D., and Hilvert, D. (2014) Reprogramming nonribosomal peptide synthetases for “clickable” amino acids. *Angew Chemie - Int Ed* **53**: 10105–10108.

Kumar, A., and Worobec, E.A. (2005) Cloning, sequencing, and characterization of the SdeAB multidrug efflux pump of *Serratia marcescens*. *Antimicrob Agents Chemother* **49**: 1495–1501.

Kunesf, S., Botstein, D., and Fox, S. (1985) Transformation of Yeast with Linearized Plasmid DNA Formation of Inverted Dimers and Recombinant Plasmid Products repair. *J Mol Biol* **184**: 375–387.

Lambalot, R.H., Gehring, a M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M. a, *et al.* (1996) A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* **3**: 923–936.

- Lane, C., Nr, N., and Innes, J. (1983) CDA is a new chromosomally-determined antibiotic from *Streptomyces coelicolor* A3(2). *J Gen Microbiol* **3**: 3575–3579.
- Larionov, V., Kouprina, N., Graves, J., Chen, X.N., Korenberg, J.R., and Resnick, M.A. (1996) Specific cloning of human DNA. *Proc Natl Acad Sci* **93**: 491–496.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., McWilliam, H., *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
- Laureti, L., Song, L., Huang, S., Corre, C., Leblond, P., and Challis, G.L. (2011) Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc Natl Acad Sci* **108**.
- Lautru, S., Song, L., Demange, L., Lombès, T., Galons, H., Challis, G.L., and Pernodet, J.-L. (2012) A Sweet Origin for the Key Congocidine Precursor 4-Acetamidopyrrole-2-carboxylate. *Angew Chemie Int Ed* **51**: 7454–7458.
- Lee, S.K., Mo, S., and Suh, J.W. (2012) An ABC transporter complex containing S-adenosylmethionine (SAM)-induced ATP-binding protein is involved in antibiotics production and SAM signaling in *Streptomyces coelicolor* M145. *Biotechnol Lett* 1–8.
- Lee, T.V., Johnson, L.J., Johnson, R.D., Koulman, A., Lane, G. a, Lott, J.S., and Arcus, V.L. (2010) Structure of a eukaryotic nonribosomal peptide synthetase adenylation domain that activates a large hydroxamate amino acid in siderophore biosynthesis. *J Biol Chem* **285**: 2415–27.
- Lenik, U., Lukezic, T., Podgorsek, A., Horvat, J., Polak, T., Sala, M., *et al.* (2015) Construction of a new class of tetracycline lead structures with potent antibacterial activity through biosynthetic engineering. *Angew Chemie - Int Ed* **54**: 3937–3940.
- Leproust, E.M., Peck, B.J., Spirin, K., Mccuen, H.B., Moore, B., Namsaraev, E., and Caruthers, M.H. (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res* **38**: 2522–2540.
- Lewis, E. a, Munde, M., Wang, S., Rettig, M., Le, V., Machha, V., and Wilson, W.D. (2011) Complexity in the binding of minor groove agents: netropsin has two thermodynamically different DNA binding modes at a single site. *Nucleic Acids Res* **39**: 9649–58.
- Lewis, K. (2013) Platforms for antibiotic discovery. *Nat Rev Drug Discov* **12**: 371–87.
- Li, B., Sher, D., Kelly, L., Shi, Y., Huang, K., Knerr, P.J., *et al.* (2010) Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc Natl Acad Sci U S A* **107**: 10430–10435.
- Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L., Engels, I., Conlon, B.P., *et al.* (2015) A new antibiotic kills pathogens without detectable resistance. *Nature* **517**: 455–9.
- Lipmann, F. (1980) Bacterial production of antibiotic polypeptides by thiol-linked synthesis on protein templates. *Adv Microb Physiol* **21**: 227–66.

- Liu, Y.-Y., Wang, Y., Walsh, T.R., Yi, L.-X., Zhang, R., Spencer, J., *et al.* (2016) Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *Lancet Infect Dis* **16**: 161–168.
- Lomovskaya, O., and Bostian, K.A. (2006) Practical applications and feasibility of efflux pump inhibitors in the clinic - A vision for applied use. *Biochem Pharmacol* **71**: 910–918.
- Lomovskaya, O., Warren, M.S., Lee, A., Fronko, R., Lee, M., Blais, J., *et al.* (2001) Identification and Characterization of Inhibitors of Multidrug Resistance Efflux Pumps in *Pseudomonas aeruginosa* : Novel Agents for Combination Therapy Identification and Characterization of Inhibitors of Multidrug Resistance Efflux Pumps in *Pseudomonas a.* *Antimicrob Agents Chemother* **45**: 105–116.
- Lucas, X., Senger, C., Erxleben, A., Grüning, B.A., Döring, K., Mosch, J., *et al.* (2013) StreptomeDB: A resource for natural compounds isolated from *Streptomyces* species. *Nucleic Acids Res* **41**: 1130–1136.
- Luzhetskyy, A. (2016) Actinobacteria Metabolic Engineering Helmholtz Centre. https://www.helmholtz-hzi.de/en/research/research_topics/anti_infectives/actinobacteria_metabolic_engineering/projects/.
- Macneil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and Maeneil, T. (1992) *Streptomyces avermitilis*. *Plasmid* **111**: 61–68.
- Madigan M, M.J. (2005) *Brock Biology of Microorganisms (11th ed.)*. Prentice Hall., .
- Madoui, M.-A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., *et al.* (2015) Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics* **16**: 327.
- Marahiel, M. (1992) Multidomain enzymes involved in peptide synthesis. *FEBS Lett* **307**: 40–43.
- Marahiel, M. a., Stachelhaus, T., and Mootz, H.D. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev* **97**: 2651–2674.
- Marshall, C.G., Lessard, I.A.D., and Park, I. (1998) Glycopeptide Antibiotic Resistance Genes in Glycopeptide-Producing Organisms. *Antimicrob Agents Chemother* **42**: 2215–2220.
- May, J.J., Kessler, N., Marahiel, M. a, and Stubbs, M.T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A* **99**: 12120–5.
- Mcalpine, J.B., Tuan, J.S., Whittern, D.N., and Buko, A. (1987) New antibiotics from genetically engineered actinomycetes. *J Antibiot (Tokyo)* **8**: 1115–1122.
- McDaniel, R., Ebert-Khosla, S., Hopwood, D. a, and Khosla, C. (1993) Engineered biosynthesis of novel polyketides. *Science* **262**: 1546–1550.

- Medema, M.H., Blin, K., Cimermancic, P., Jager, V. De, Zakrzewski, P., Fischbach, M. a., *et al.* (2011) AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* **39**: 339–346.
- Meier, J.L., and Burkart, M.D. (2012) The chemical biology of modular biosynthetic enzymes. *Chem Soc Rev* **38**: 2012–2045.
- Menzella, H.G., and Reeves, C.D. (2007) Combinatorial biosynthesis for drug development. *Curr Opin Microbiol* **10**: 238–45.
- Meyerhoff, J., and Bellgardt, K. (1995) A morphology-based model for fed-batch cultivations of *Penicillium chrysogenum* growing in pellet form. *J Biotechnol* **38**: 201–217.
- Meyerhoff, J., and Tiller, V. (1995) Two mathematical models for the development of a single microbial pellet. *Bioprocess Eng* **12**: 305–313.
- Micklefield, J., Amir-heidari, B., Thirlway, J., and Micklefield, J. (2008) Auxotrophic-precursor directed biosynthesis of nonribosomal lipopeptides with modified tryptophan residues. *Org Biomol Chem* **6**: 2–6.
- Montiel, D., Kang, H.-S., Chang, F.-Y., Charlop-Powers, Z., and Brady, S.F. (2015) Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *Proc Natl Acad Sci* 1–6.
- Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S., and Stewart, P.D. (2014) Membrane protein structure determination - the next generation. *Biochim Biophys Acta* **1838**: 78–87.
- Moretti, R., Chang, A., Peltier-pain, P., Bingman, C.A., Phillips, G.N., and Thorson, J.S. (2011) Expanding the Nucleotide and Sugar 1-Phosphate Promiscuity of Nucleotidyltransferase RmlA via Directed Evolution. *J Biol Chem VOL* **286**: 13235–13243.
- Moss, S.J. et al. (2006) Biosynthesis of the angiogenesis inhibitor borrelidin: directed biosynthesis of novel analogues. *Chem Commun* **22**: 2341–2343.
- Murakami, S., Nakashima, R., Yamashita, E., and Yamaguchi, A. (2002) Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**: 587–593.
- Musiol, E.M., Greule, A., Härtner, T., Kulik, A., Wohlleben, W., and Weber, T. (2013) The AT2 domain of KirCI loads malonyl extender units to the ACPs of the kirromycin PKS. *ChemBioChem* **14**: 1343–1352.
- Nakata, H., Yamada, K., Okuhara, K., and Naito, T. (1961) The structure of aureothin, a nitro compound obtained from *Streptomyces thioluteus*. *Tetrahedron* **14**: 252–274.
- Neidle, S. (2001) DNA minor-groove recognition by small molecules. *Nat Prod Rep* **18**: 291–309.

- Nett, M., Ikeda, H., and Moore, B.S. (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* **26**: 1362–1384.
- Nguyen, T., Ishida, K., Jenke-kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., *et al.* (2008) Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat Biotechnol* **26**: 225–233.
- 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* **8**: 1–14.
- Nieselt, K., Battke, F., Herbig, A., Bruheim, P., Wentzel, A., Jakobsen, Ø.M., *et al.* (2010) The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics* **11**: 1–10.
- Nishino, K., Nikaido, E., and Yamaguchi, A. (2009) Regulation and physiological function of multidrug efflux pumps in *Escherichia coli* and *Salmonella*. *Biochim Biophys Acta* **1794**: 834–43.
- Noda-García, L., and Barona-Gómez, F. (2013) Enzyme evolution beyond gene duplication: A model for incorporating horizontal gene transfer. *Mob Genet Elements* **3**: e26439–1 e26439–4.
- Nodwell, J.R. (2014) Are you talking to me? A possible role for γ -butyrolactones in interspecies signalling. *Mol Microbiol* **94**: 483–485.
- Norio, E., Takashi, S., Masao, K., Tomizo, N., and Taro, N. (1981) NEW CHLORINATED NITRO-PYRROLE ANTIBIOTICS, PYRROLOMYCIN A AND B (SF-2080 A AND B) Sir: *J Antibiot (Tokyo)* **XXXIV**: 5–7.
- Olano, C., García, I., González, A., Rodríguez, M., Rozas, D., Rubio, J., *et al.* (2014) Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb Biotechnol* **7**: 242–256.
- Olano, C., Méndez, C., and Salas, J. a (2010) Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat Prod Rep* **27**: 571–616.
- Pagès, J.M., and Amaral, L. (2009) Mechanisms of drug efflux and strategies to combat them: Challenging the efflux pump of Gram-negative bacteria. *Biochim Biophys Acta - Proteins Proteomics* **1794**: 826–833.
- Parkinson, J.A., Scott, F.J., Suckling, C.J., and Wilson, G. (2013) Exceptionally strong intermolecular association in hydrophobic DNA minor groove binders and their potential therapeutic consequences. *Medchemcomm* **4**: 1105–1108.
- Paulsen, I.T., Press, C.M., Ravel, J., Kobayashi, D.Y., Myers, G.S. a, Mavrodi, D. V, *et al.* (2005) Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat*

Biotechnol **23**: 873–878.

Pawlik, K., Kotowska, M., Chater, K.F., Kuczek, K., and Takano, E. (2007) A cryptic type I polyketide synthase (cpk) gene cluster in *Streptomyces coelicolor* A3(2). *Arch Microbiol* **187**: 87–99.

Pelton, J.G., and Wemmer, D.E. (1989) Structural characterization of a 2 : 1 distamycin A d (CGCAAATI' GGC) complex by two-dimensional NMR. *Proc Natl Acad Sci USA* **86**: 5723–5727.

Pernodet, J.-L., Alegre, M.-T., Blondelet-Rouault, M.-H., and Guérineau, M. (1993) Resistance to spiramycin in *Streptomyces ambofaciens*, the producer organism, involves at least two different mechanisms. *J Gen Microbiol* **139**: 1003–1011.

Peterson, R.M., Huang, T., Rudolf, J.D., Smanski, M.J., and Shen, B. (2014) Mechanisms of Self-Resistance in the *Streptomyces platensis* MA7327 and MA7339 Strains. *Chem Biol* 389–397.

Piddock, L.J. V (2006) Multidrug-resistance efflux pumps - not just for resistance. *Nat Rev Microbiol* **4**: 629–636.

Piel, J. (2010) Biosynthesis of polyketides by trans -AT polyketide synthases. *Nat Prod Rep* 996–1047.

Pometto, a L., and Crawford, D.L. (1986) Effects of pH on Lignin and Cellulose Degradation by *Streptomyces viridosporus*. *Appl Environ Microbiol* **52**: 246–50.

Prajapati, S.M., Patel, K.D., Vekariya, R.H., Panchal, S.N., and Patel, H.D. (2014) Recent advances in the synthesis of 2-substituted benzothiazoles: a review. *RSC Adv* **4**: 24463–24476.

Prasad, R., Banerjee, A., Khandelwal, N.K., and Dhamgaye, S. (2015) The ABCs of *Candida albicans* multidrug transporter Cdr1. *Eukaryot Cell* EC.00137–15.

Quan, J., and Tian, J. (2014) Circular polymerase extension cloning. *Methods Mol Biol* **1116**: 103–117.

Rath, V.L., Silvian, L.F., Beijer, B., Sproat, B.S., and Steitz, T. a (1998) How glutaminyl-tRNA synthetase selects glutamine. *Structure* **6**: 439–449.

Raymond, C.K., Sims, E.H., and Olson, M. V (2002) Linker-Mediated Recombinational Subcloning of Large DNA Fragments Using Yeast. *Genome Res* **12**: 190–197.

Ren, J., and Chaires, J.B. (1999) Sequence and Structural Selectivity of Nucleic Acid Binding Ligands. *Biochemistry* **38**: 16067–16075.

Reshetnikova, L., Moor, N., Lavrik, O., and Vassilyev, D.G. (1999) Crystal structures of phenylalanyl-tRNA synthetase complexed with phenylalanine and a phenylalanyl-adenylate analogue. *J Mol Biol* **287**: 555–68.

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A.W., Hopwood, D. a, and Wezel,

- G.P. van (2008) Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep* **9**: 670–675.
- Rinehart, K. (1969) The Neomycins and Related Antibiotics. *J Infect Dis* **119**: 1–4.
- Robert H. Schiestl, R.D.G. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet* **Volume 16**: pp 339–346.
- Rodriguez, E., Navone, L., Casati, P., and Gramajo, H. (2012) Impact of malic enzymes on antibiotic and triacylglycerol: Production in *Streptomyces coelicolor*. *Appl Environ Microbiol* **78**: 4571–4579.
- Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res* **39**: W362–7.
- Roy, A.D., Gru, S., Cairns, N., and Goss, R.J.M. (2010) Gene Expression Enabling Synthetic Diversification of Natural Products : Chemogenetic Generation of Pacidamycin Analogs. *JAmChem Soc* **132**: 12243–12245.
- Ru, D.E., Schmidt, E.W., and Heemstra, J.R. (2015) Assessing the Combinatorial Potential of the RiPP Cyanobactin tru Pathway. *ACS Synth Biol* **4**: 482–492.
- Saier, M.H., and Paulsen, I.T. (2001) Phylogeny of multidrug transporters. *Semin Cell Dev Biol* **12**: 205–213.
- Saitou N, N.M. (1987) The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol Biol Evol* **4**: 406–425.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning: A laboratory manual: 2nd ed.* Cold Spring Harbor Laboratory Press, .
- Sambrook, J., and Russell, D. (2000) *Molecular Cloning.* .
- Sanfilippo, A., Enza, M., and Ghione, M. (1966) Activity of Distamycin A on the Induction of Adaptive Enzymes in *Escherichia coli*. *J.gen Microbiol* **43**: 369–374.
- Sato, M., Winter, J.M., Kishimoto, S., Noguchi, H., Tang, Y., and Watanabe, K. (2016) Combinatorial Generation of Chemical Diversity by Redox Enzymes in Chaetoviridin Biosynthesis. *Org Lett* **18**: 1446–1449.
- Schloss, J.A. (2008) How to get genomes at one ten-thousandth the cost. *Nat Biotechnol* **26**: 1113–1115.
- Schwarzer, D., Finking, R., and Marahiel, M. a. (2003) Nonribosomal peptides: from genes to products. *Nat Prod Rep* **20**: 275.
- Seiple, I.B., Zhang, Z., Jakubec, P., Langlois-Mercier, A., Wright, P.M., Hog, D.T., et al. (2016) A platform for the discovery of new macrolide antibiotics. *Nature* **533**: 338–345.

- Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y., and Zhao, H. (2013) Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth Biol* **2**: 662–669.
- Shao, Z., Zhao, H., and Zhao, H. (2009) DNA assembler , an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* **37**: 1–10.
- Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* **7**: 285–295.
- Shier, W.T., Rinehart, K.L., and Gottlieb, D. (1969) Preparation of four new antibiotics from a mutant of *Streptomyces fradiae*. *Proc Natl Acad Sci U S A* **63**: 198–204.
- Sidda, J.D., and Corre, C. (2012) Gamma-butyrolactone and furan signaling systems in *streptomyces*. *Methods Enzymol* **517**: 71–87.
- Sidda, J.D., Song, L., Poon, V., Al-Bassam, M., Lazos, O., Buttner, M.J., *et al.* (2014) Discovery of a family of g-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. *Chem Sci* **5**: 86–89.
- Siegl, T., Tokovenko, B., Myronovskiy, M., and Luzhetskyy, A. (2013) Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metab Eng* **19**: 98–106.
- Sikorski, R.S., and Hieter, P. (1989) A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in *Saccharomyces cerevisiae*. *Genet Soc Am A* **122**: 19–27.
- Smanski, M.J., Zhou, H., Claesen, J., Shen, B., Fischbach, M.A., and Voigt, C.A. (2016) Synthetic biology to access and expand nature’s chemical diversity. *Nat Rev Microbiol* **14**: 135–149.
- Spellberg, B., Miller, L.G., Kuo, M.N., Bradley, J., Scheld, W.M., Jr, J.E.E., *et al.* (2007) Societal Costs Versus Savings from Wild-Card Patent Extension Legislation to Spur Critically Needed Antibiotic Development. *Infection* 167–174.
- Stachelhaus, T. (1998) Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. CATALYTIC ROLE OF THE CONDENSATION DOMAIN. *J Biol Chem* **273**: 22773–22781.
- Stachelhaus, T., Marahiel, A., and Mootz, D. (1999) The specificity-conferring code of adenylation nonribosomal peptide synthetases A Marahiel in. *Chem Biol* 493–505.
- Stachelhaus, T., and Marahiel, M.A. (1994) Modular Structure of Peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J Biol Chem* **270**: 6163–6169.
- Stachelhaus, T., Schneider, A., and Marahiel, M.A. (1995) Rational Design of Peptide Antibiotics by Targeted Replacement of Bacterial and Fungal Domains Peptide secondary. *Science (80-)* **269**: 69–71.
- Staunton, J., and Sutkowski, A.C. (1991) The Polyketide Synthase (PKS) of Aspyrone Biosynthesis : Evidence for the Enzyme Bound Intermediates from Incorporation Studies with

N-Acetylcysteamine Thioesters in Intact Cells of *Aspergillus melleus*. *J CHEM SOC, CHEM COMMUN* 1110–1112.

Staunton, J., and Weissman, K.J. (2001) Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **18**: 380–416.

Steenbergen, J.N., Alder, J., Thorne, G.M., and Tally, F.P. (2005) Daptomycin: A lipopeptide antibiotic for the treatment of serious Gram-positive infections. *J Antimicrob Chemother* **55**: 283–288.

Stegmann, E. et al. (2005) Precursor-directed biosynthesis for the generation of novel glycopeptides. *Ernst Scher Res Found Work* 215–232.

Stein, T., Vater, J., Kruff, V., Otto, A., Wittmann-Liebold, B., Franke, P., et al. (1996) The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J Biol Chem* **271**: 15428–15435.

Stinear, T.P., Mve-Obiang, A., Small, P.L.C., Frigui, W., Pryor, M.J., Brosch, R., et al. (2004) Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci U S A* **101**: 1345–1349.

Stirrett, K., Denoya, C., and Westpheling, J. (2009) Branched-chain amino acid catabolism provides precursors for the Type II polyketide antibiotic, actinorhodin, via pathways that are nutrient dependent. *J Ind Microbiol Biotechnol* **36**: 129–137.

Stumpp, T., Himbert, S., and Altenbuchner, J. (2005) Cloning of the netropsin resistance genes from *Streptomyces flavopersicus* NRRL 2820. *J Basic Microbiol* **45**: 355–362.

Suckling, C.J. (2015) The Antibacterial Drug MGB-BP3: from discovery to clinical trial. *CHEMISTRY & BIOLOGY INTERFACE*. **5**: 166–174.

Sun, H., Liu, Z., Zhao, H., and Ang, E.L. (2015) Recent advances in combinatorial biosynthesis for drug discovery. *Drug Des Devel Ther* **9**: 823–833.

Sun, J., Deng, Z., and Yan, A. (2014) Bacterial multidrug efflux pumps: Mechanisms, physiology and pharmacological exploitations. *Biochem Biophys Res Commun* **453**: 254–267.

Sundermann, U., Bravo-Rodriguez, K., Klopries, S., Kushnir, S., Gomez, H., Sanchez-Garcia, E., and Schulz, F. (2013) Enzyme-directed mutasynthesis: A combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS Chem Biol* **8**: 443–450.

Takizawa, M., Shigetoshi, T., Seiichi, T., Setsuo, H., and Toru, H. (1987) A NEW PYRROLE-AMIDINE ANTIBIOTIC TAN-868 A. *J Antibiot (Tokyo)* 1220–1230.

Tao, Z.F., Fujiwara, T., Saito, I., and Sugiyama, H. (1999) Sequence-specific DNA alkylation by hybrid molecules between segment A of duocarmycin A and pyrrole/imidazole diamide. *Angew Chemie - Int Ed* **38**: 650–653.

- Tatusova, T.A., and Madden, T.L. (1999) Erratum to `` BLAST 2 SEQUENCES , a new tool for comparing protein and nucleotide sequences [FEMS Microbiol . 174 (1999) 247 ^ 250] 1. *FEMS Microbiol Lett* **177**: 187–188.
- Tegos, G.P., Haynes, M., Strouse, J.J., Khan, M.M.T., Bologna, C.G., Oprea, T.I., and Sklar, L.A. (2011) Microbial efflux pump inhibition: tactics and strategies. *Curr Pharm Des* **17**: 1291–302.
- Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Bentley, S.D., Harper, D., *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces*. *Nature* **3,417**:141-147.
- Townsend, C.A. (2002) New reactions in clavulanic acid biosynthesis. *Curr Opin Chem Biol* **6**: 583–589.
- Trujillo, M.E. (2008) Actinobacteria. In *eLS*. John Wiley & Sons, Ltd, .
- Untergasser, A. (2008) Preparation of Electro-Competent Cells. *Winter 2008* http://www.untergasser.de/lab/protocols/competent_cells_electro_v1_0.htm.
- Vingadassalon, A., Lorieux, F., Juguet, M., Goff, L., Gerbaud, C., Pernodet, J., and Lautru, S. (2014a) Natural combinatorial biosynthesis involving two clusters for the synthesis of three pyrrolamides in *Streptomyces netropsis*.
- Vingadassalon, A., Lorieux, F., Juguet, M., Goff, L., Gerbaud, C., Pernodet, J., and Lautru, S. (2014b) Natural Combinatorial Biosynthesis Involving Two Clusters for the Synthesis of Three Pyrrolamides in *Streptomyces netropsis*. *ACS Chem Biol* **10**: 601–610.
- Volff, J., and Altenbuchner, J. (1998) MicroReview Genetic instability of the *Streptomyces* chromosome. **27**: 239–246.
- Waclaw, S. (1978) NOBEL PRIZES AND RESTRICTION ENZYMES. *Gene* **4**: 181–182.
- Waksman, A., and Henrici, A. (1943) The nomenclature and classification of Actinomycetes. *J Bacteriol* **4**: 337–341.
- Waksman, S. (1939) Classification of actinomycetes. *Exp Agric* 549–558.
- Waldvogel, E., Herbig, A., Battke, F., Amin, R., Nentwich, M., Nieselt, K., *et al.* (2011) The PII protein GlnK is a pleiotropic regulator for morphological differentiation and secondary metabolism in *Streptomyces coelicolor*. *Appl Microbiol Biotechnol* **92**: 1219–1236.
- Walker, J.E., Saraste, M., Runswick, M., and Gay, N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**: 945–951.
- Walsh, C.T., and Fischbach, M. a (2010) Natural products version 2.0: connecting genes to molecules. *J Am Chem Soc* **132**: 2469–93.
- Walsh, C.T., and Nolan, E.M. (2008) Morphing peptide backbones into heterocycles. *Proc Natl Acad Sci USA* **105**: 5655–5656.

- Wang, J., Soisson, S.M., Young, K., Shoop, W., Kodali, S., Galgoci, A., *et al.* (2006) Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature* **441**: 358–361.
- Wang, X.K., and Jin, J.L. (2014) Crucial factor for increasing the conjugation frequency in *Streptomyces netropsis* SD-07 and other strains. *FEMS Microbiol Lett* **357**: 99–103.
- Webber, M. a., and Piddock, L.J. V (2003) The importance of efflux pumps in bacterial antibiotic resistance. *J Antimicrob Chemother* **51**: 9–11.
- Weber, G., Schörgendorfer, K., Schneider-Scherzer, E., and Leitner, E. (1994) The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr Genet* **26**: 120–5.
- Weber, J.M., Leung, J., Swanson, S.J., and Idler, K.B. (1991) An Erythromycin Derivative Produced by Targeted Gene Disruption in *Saccharopolyspora erythraea*. *Science (80-)* **252**: 114–117.
- Weber, T., and Marahiel, M.A. (2001) Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure* **9**: 3–9.
- Weissman, K.J. (2007) Mutasynthesis - uniting chemistry and genetics for drug discovery. *Trends Biotechnol* **25**: 139–142.
- Weissman, K.J., and Leadlay, P.F. (2005) Combinatorial biosynthesis of reduced polyketides. *Nat Rev Micro* **3**: 925–936.
- Weist, S. et al. (2002) Fluorobalhimycin – a new chapter in glycopeptide antibiotic research. *Angew Chem Int Ed Engl* 3383–3385.
- Weist, S., Kittel, C., Bischoff, D., Bister, B., Pfeifer, V., Nicholson, G.J., *et al.* (2004) Mutasynthesis of Glycopeptide Antibiotics : Variations of Vancomycin ' s. *JAmChem Soc* **126**: 5942–5943.
- Weist, S., and Sussmuth, R. (2005) Mutational biosynthesis — a tool for the generation of structural diversity in the biosynthesis of antibiotics. *Appl Microbiol Biotechnol* 141–150.
- Wentzel, A., Bruheim, P., Øverby, A., Jakobsen, Ø.M., Sletta, H., Omara, W. a M., *et al.* (2012) Optimized submerged batch fermentation strategy for systems scale studies of metabolic switching in *Streptomyces coelicolor* A3(2). *BMC Syst Biol* **6**: 59.
- Werneburg, M., Busch, B., He, J., Richter, M.E. a, Xiang, L., Moore, B.S., *et al.* (2010) Exploiting enzymatic promiscuity to engineer a focused library of highly selective antifungal and antiproliferative aureothin analogues. *J Am Chem Soc* **132**: 10407–10413.
- Wezel, G.P. Van, Krabben, P., Traag, B.A., Keijser, B.J.F., Kerste, R., Vijgenboom, E., *et al.* (2006) Unlocking *Streptomyces spp* . for Use as Sustainable Industrial Production Platforms by Morphological Engineering. *Appl Environ Microbiol* **72**: 5283–5288.

- White, T.C., and Marr, K.A. (1998) Clinical , Cellular , and Molecular Factors That Contribute to Antifungal Drug Resistance. **11**: 382–402.
- Wiedmann, M., Wilson, W.I., Luo, J., Barany, F., and Batt, A. (1994) Ligase Chain Reaction Applications. *Genome Res* **3**: S51–S64.
- Wiegand, I., Hilpert, K., and Hancock, R.E.W. (2008) Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc* **3**: 163–175.
- Winter, J.M., Chiou, G., Bothwell, I.R., Xu, W., Garg, N.K., Luo, M., and Tang, Y. (2013) Expanding the structural diversity of polyketides by exploring the cofactor tolerance of an inline methyltransferase domain. *Org Lett* **15**: 3774–3777.
- Winter, J.M., Sato, M., Sugimoto, S., Chiou, G., Garg, N.K., Tang, Y., and Watanabe, K. (2012) Identification and characterization of the chaetoviridin and chaetomugilin gene cluster in chaetomium globosum reveal dual functions of an iterative highly-reducing polyketide synthase. *J Am Chem Soc* **134**: 17900–17903.
- Winter, J.M., and Tang, Y. (2014) Natural products: Getting a handle on peptides. *Nat Chem* **6**: 1037–8.
- Wise, R. (2011) The urgent need for new antibacterial agents. *J Antimicrob Chemother* **66**: 1939–40.
- Witt, D., and Stackebrandt, E. (1990) Unification of the genera *Streptoverticillum* and *Streptomyces*, and amendment of *Streptomyces* Waksman and Henrici 1943, 339 AL. *Syst Appl Microbiol* **13**: 361–371.
- Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms : Proposal for the domains. **87**: 4576–4579.
- Wright, G. (2014) Perspective: Synthetic biology revives antibiotics. *Nature* **509**: S13–S13.
- Wright, G.D. (2005) Bacterial resistance to antibiotics : Enzymatic degradation and modification. *Adv Drug Deliv Rev* **57**: 1451–1470.
- Wright, L.F., and Hopwood, D. a (1976) Actinorhodin is a Chromosomally-determined Antibiotic in *Streptomyces coelicolor* A3(2). *J Gen Microbiol* **3**: 289–297.
- Xiao, J., Wang, Y., Luo, Y., Xie, S., and Ruan, J. (2009) *Streptomyces avicenniae* sp . nov ., a novel actinomycete isolated from the rhizosphere of the mangrove plant *Avicennia mariana*. *Xiao, J, Wang, Y, Luo, Y, Xie, S, Ruan, J Streptomyces avicenniae* sp nov , a Nov actinomycete Isol from Rhizosph mangrove plant *Avicennia Marian* 2624–2628 2624–2628.
- Xie, Y., Cai, Q., Ren, H., Wang, L., Xu, H., Hong, B., et al. (2014) NRPS substrate promiscuity leads to more potent antitubercular sansanmycin analogues. *J Nat Prod* **77**: 1744–1748.
- Yamanaka, K., Reynolds, K. a, Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., et al. (2014)

- Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci U S A* **111**: 1957–62.
- Yang, F., Nickols, N.G., Li, B.C., Marinov, G.K., Said, J.W., and Dervan, P.B. (2013) Antitumor activity of a pyrrole-imidazole polyamide. *Proc Natl Acad Sci USA* **110**: 1863–1868.
- Yates, F. (1934) The analysis of multiple classifications with unequal numbers in the different classes. *J Am Stat Assoc* **29**: 51–66.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 134.
- Yoon, V., and Nodwell, J.R. (2014) Activating secondary metabolism with stress and chemicals. *J Ind Microbiol Biotechnol* **41**: 415–424.
- Yu, H., Tang, H., and Xu, P. (2014) Green strategy from waste to value-added-chemical production: efficient biosynthesis of 6-hydroxy-3-succinoyl-pyridine by an engineered biocatalyst. *Sci Rep* **4**: 1–8.
- Zhang, X., and Parry, R.J. (2007) Cloning and characterization of the pyrrolomycin biosynthetic gene clusters from *Actinosporangium vitaminophilum* ATCC 31673 and *Streptomyces* sp. strain UC 11065. *Antimicrob Agents Chemother* **51**: 946–957.
- Zhang, X.C., Han, L., and Zhao, Y. (2015) Thermodynamics of ABC transporters. *Protein Cell* 1–11.
- Zhang, Z., Yang, X., Meng, L., Liu, F., Shen, C., and Yang, W. (2009) Enhanced amplification of GC-rich DNA with two organic reagents. *Biotechniques* **47**: 775–779.
- Zhao, H., and Shao, Z. (2009) Manipulating Natural Product Biosynthetic Pathways via DNA Assembler Zengyi. *Curr Protoc Chem Biol* **5**: 251–68.
- Ziehl, M., He, J., Dahse, H., and Hertweck, C. (2005) Mutasythesis of Aureonitrile: An Aureothin Derivative with Significantly Improved Cytostatic Effect. *Angew Chem Int Ed* **44**: 1202–1205.
- Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., and Jensen, P.R. (2012) The Natural Product Domain Seeker NaPDoS : A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS One* **7**: 1–9.

Annex-Figure 19

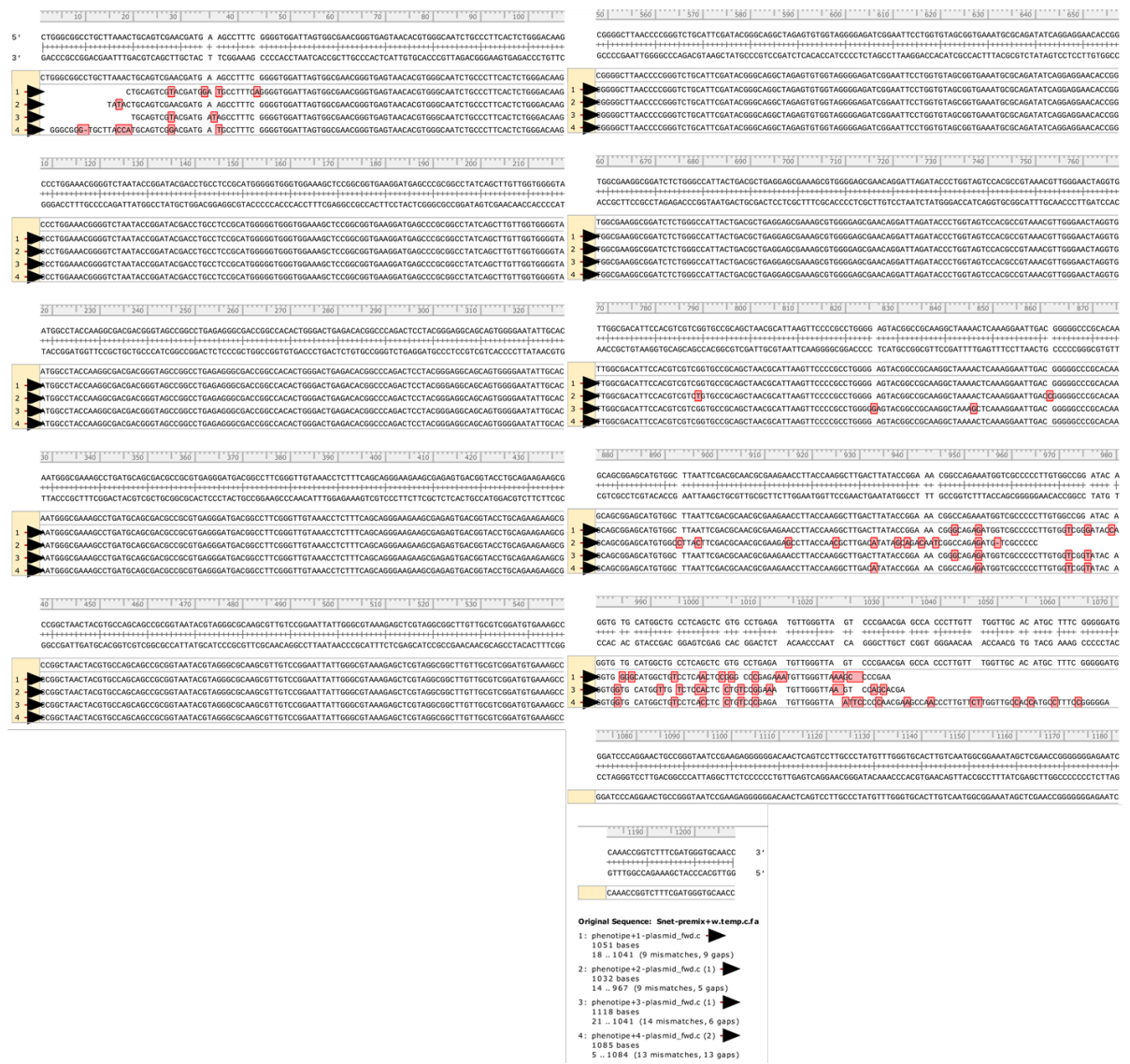


Figure 19b. Alignment used to produce the phylogenetic tree between *Streptomyces netropsis* and the four phenotypes originated after conjugational transfer of pCG002.

Reference sequence is the 16rDNA of *S.netropsis*. Sequences 1-4 corresponding to phenotypes Snety1-4 are aligned with the ClustalW (Larkin *et al.*, 2007) algorithm and displayed used Snap Gene interface. Mismatches are highlighted in red and insertion/deletions with spaces. The mismatches are in the clipping DNA regions

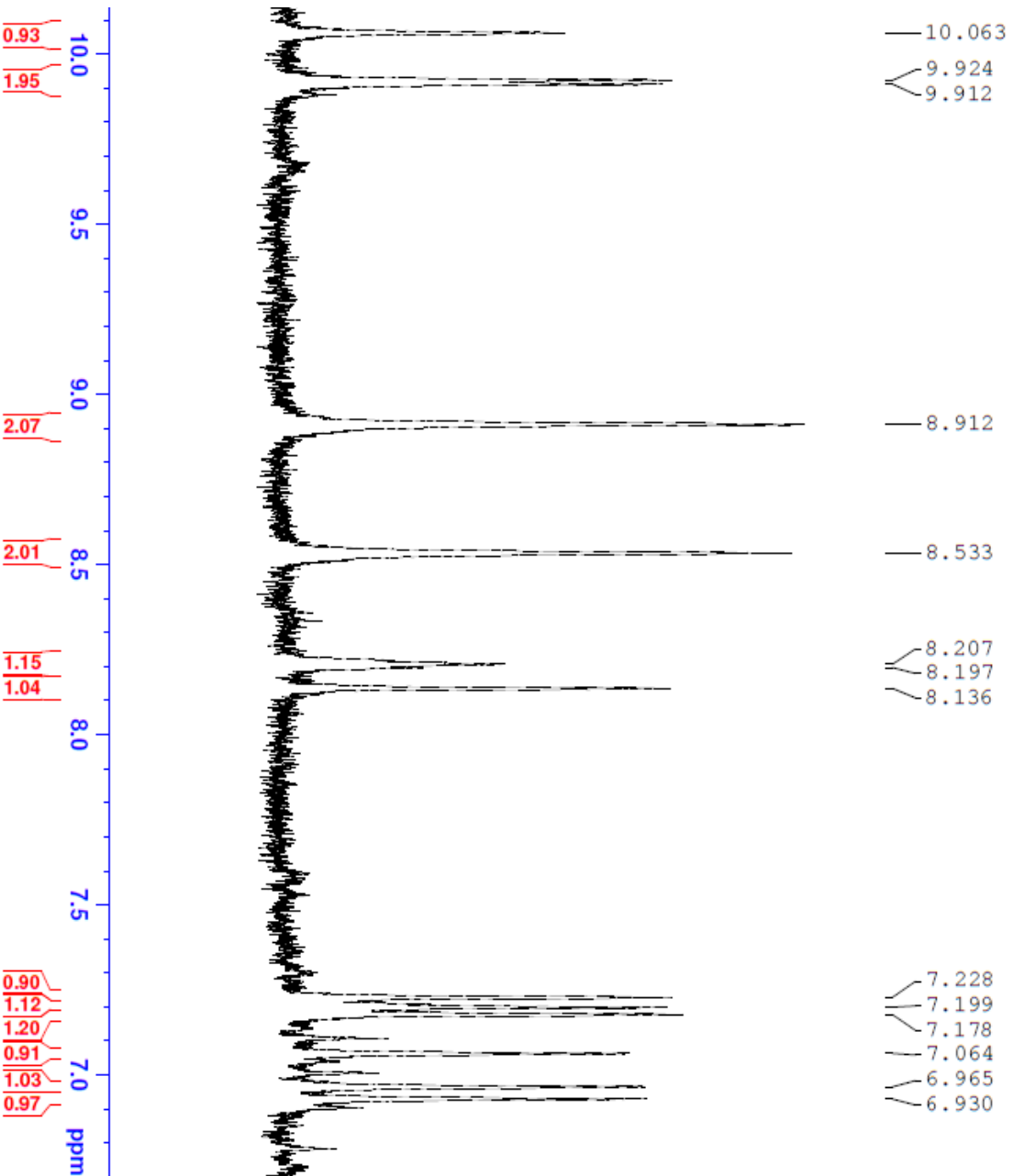
Annex-1 Distamycin characterisation data H-NMR



Current Data Parameters
 NAME E14760
 EXPNO 1
 PROCNO 1

F2 - Acquisition Parameters
 Date_ 20160606
 Time 9.21
 INSTRUM spect
 PROBHD 5 mm PABBO BB/
 PULPROG zg30
 TD 39578
 SOLVENT DMSO
 NS 16
 DS 2
 SWH 10000.000 Hz
 FIDRES 0.252666 Hz
 AQ 1.9789000 sec
 RG 198.22
 DW 50.000 usec
 DE 6.50 usec
 TE 300.1 K
 DI 2.00000000 sec
 TDO 1

==== CHANNEL f1 =====
 SF01 500.1330885 MHz
 NUC1 1H
 P1 10.00 usec
 PLW1 20.00000000 W
 F2 - Processing parameters
 SI 65536
 SF 500.1300000 MHz
 WDM EM
 SSB 0
 LB 0.30 Hz
 GB 0
 PC 1.00



Person 3-14
 FS-Distamycin

Annex-2 Distamycin characterisation data H-NMR

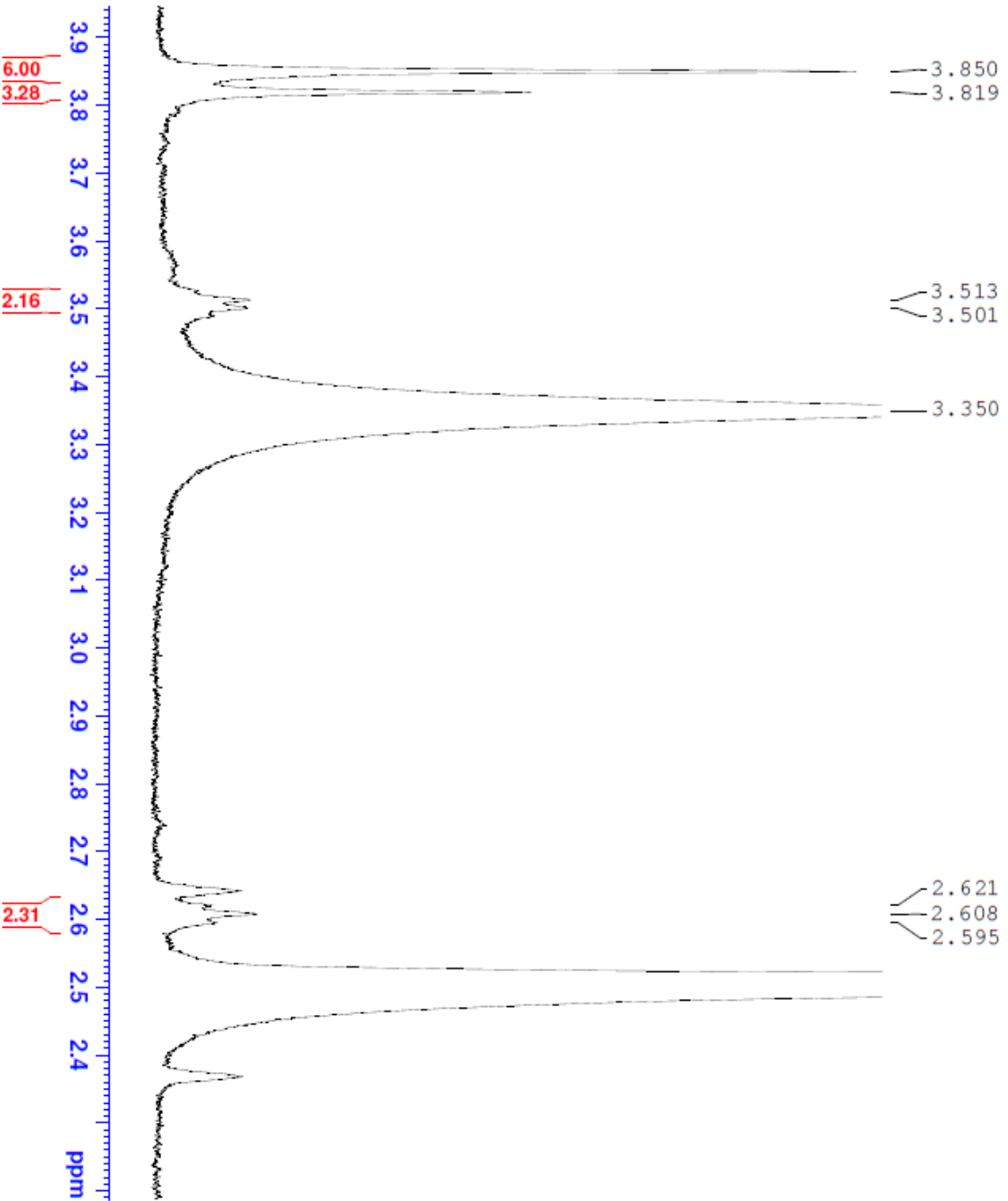


Current Data Parameters
 NAME: E14760
 EXPNO: 1
 PROCNO: 1

F2 - Acquisition Parameters
 Date_: 20160606
 Time: 9.21
 INSTRUM: spect
 PROBHD: 5 mm PABBO BB/
 PULPROG: zg30
 TD: 39578
 SOLVENT: DMSO
 NS: 16
 DS: 2
 SMH: 10000.000 Hz
 FIDRES: 0.252666 Hz
 AQ: 1.9789000 sec
 RG: 198.22
 DW: 50.000 usec
 DE: 6.50 usec
 TE: 300.1 K
 D1: 2.00000000 sec
 TDO: 1

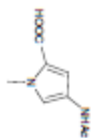
==== CHANNEL f1 =====
 SFO1: 500.1330885 MHz
 NUC1: 1H
 P1: 10.00 usec
 PLW1: 20.00000000 W

F2 - Processing parameters
 SI: 65536
 SF: 500.1300000 MHz
 MIDW: EM
 SSB: 0
 LB: 0.30 Hz
 GB: 0
 PC: 1.00



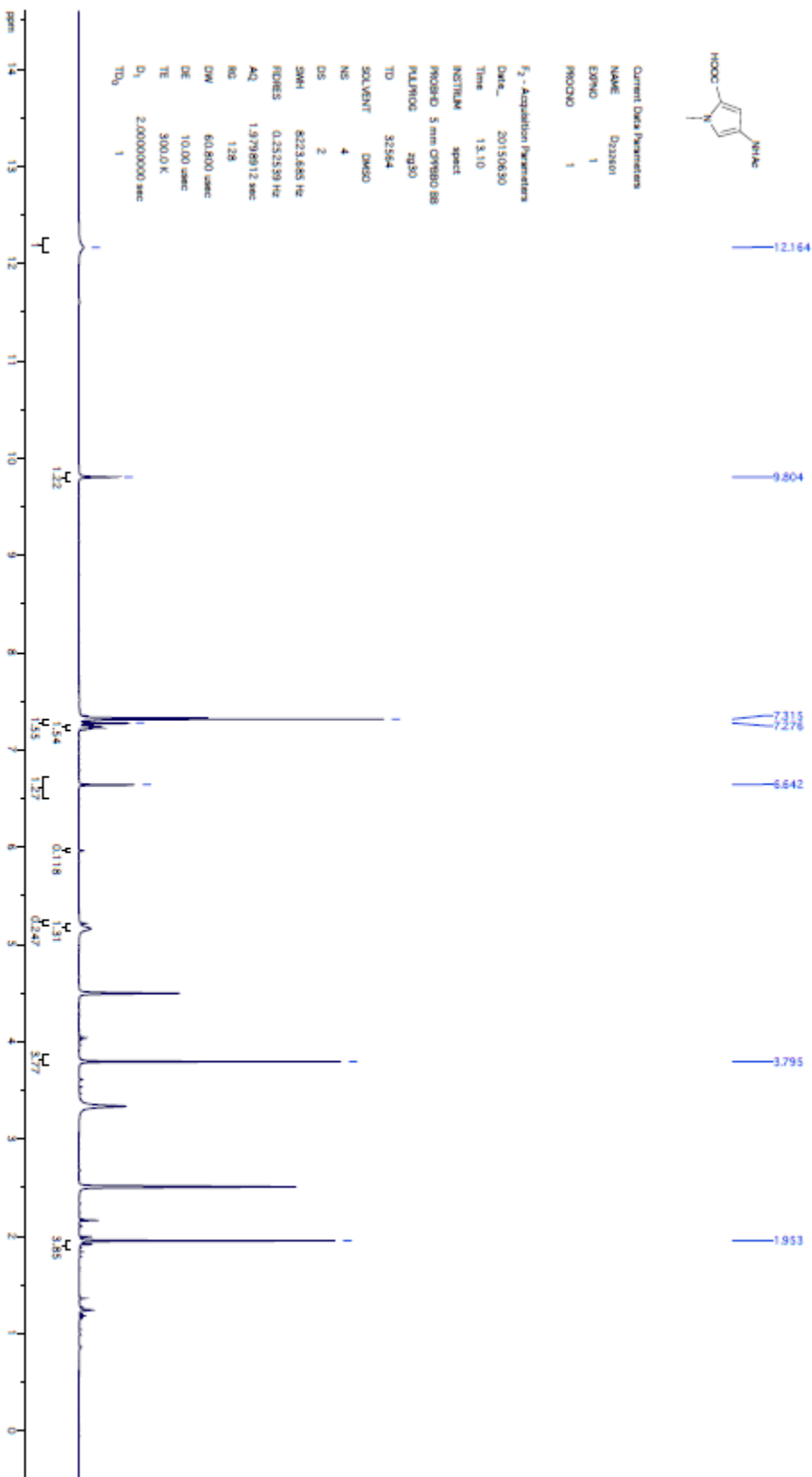
Person 3-14
 FS-Distamycin

Annex-3 H NMR 4-acetamido-1-methyl-pyrrole-2-carboxylic acid



Current Data Parameters
 NAME: D212101
 EXPNO: 1
 PROCNO: 1

f_2 - Acquisition Parameters
 Date_: 20150630
 Time: 13.10
 INSTRUM: spect
 PROBHD: 5 mm QNP80 BB
 PULPROG: zg30
 TD: 32764
 SOLVENT: DMSO
 NS: 4
 DS: 2
 SWH: 8223.885 Hz
 FIDRES: 0.252530 Hz
 AQ: 1.9798912 sec
 RG: 128
 DW: 60.800 usec
 DE: 10.000 usec
 TE: 300.0 K
 D1: 2.00000000 sec
 TDO: 1



Annex-4 LCMS characterisation data of 4-acetamido-1-methyl-pyrrole-2-carboxylic acid

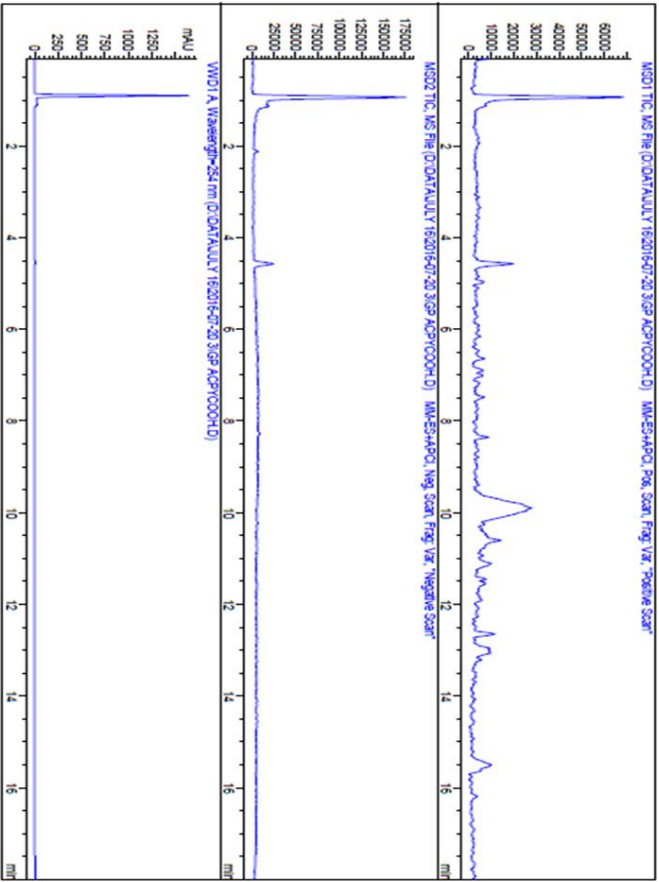
Print of window 38: Current Chromatogram(s)

Data File : D:\DATA\UTILITY 16\2016-07-20 9\GP_ABP0000.D
 Sample Name : gp_ABP0000

Acq. Operator :
 Acq. Instrument : Instrument 1
 Injection Date : 7/20/2016 5:30:07 PM
 Inj Volume : 10.000 µl

Acq. Method : D:\DATA\UTILITY 16\2016-07-20 9\FORO_ABPACCT_254.M
 Last changed : 2/12/2015 2:09:30 PM

Analysis Method : C:\CHEM32\1\METHODS\REPORTING_P000.M
 Last changed : 7/20/2016 12:52:36 PM
 (modified after loading)



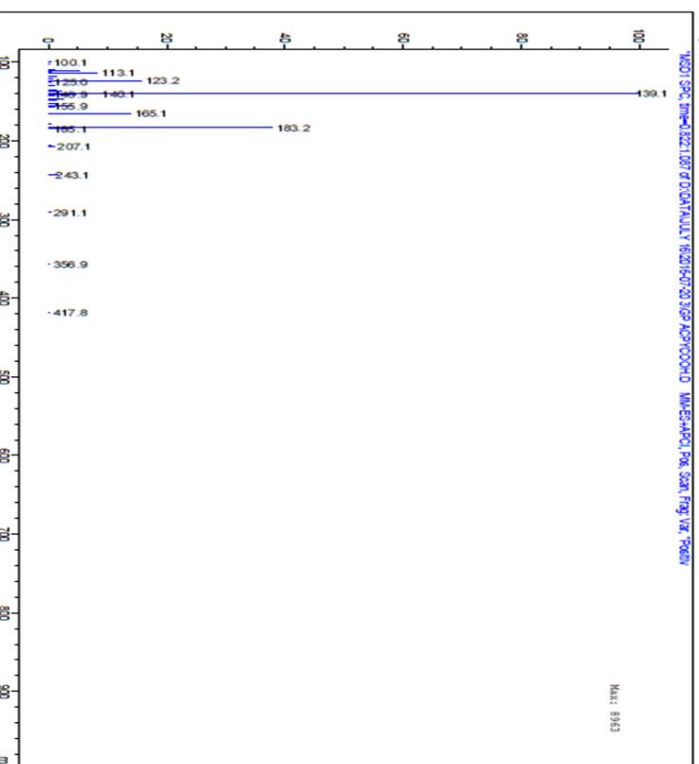
Print of window 50: MS Spectrum

Data File : D:\DATA\UTILITY 16\2016-07-20 9\GP_ABP0000.D
 Sample Name : gp_ABP0000

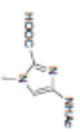
Acq. Operator :
 Acq. Instrument : Instrument 1
 Injection Date : 7/20/2016 5:30:07 PM
 Inj Volume : 10.000 µl

Acq. Method : D:\DATA\UTILITY 16\2016-07-20 9\FORO_ABPACCT_254.M
 Last changed : 2/12/2015 2:09:30 PM

Analysis Method : C:\CHEM32\1\METHODS\REPORTING_P000.M
 Last changed : 7/20/2016 12:52:36 PM
 (modified after loading)

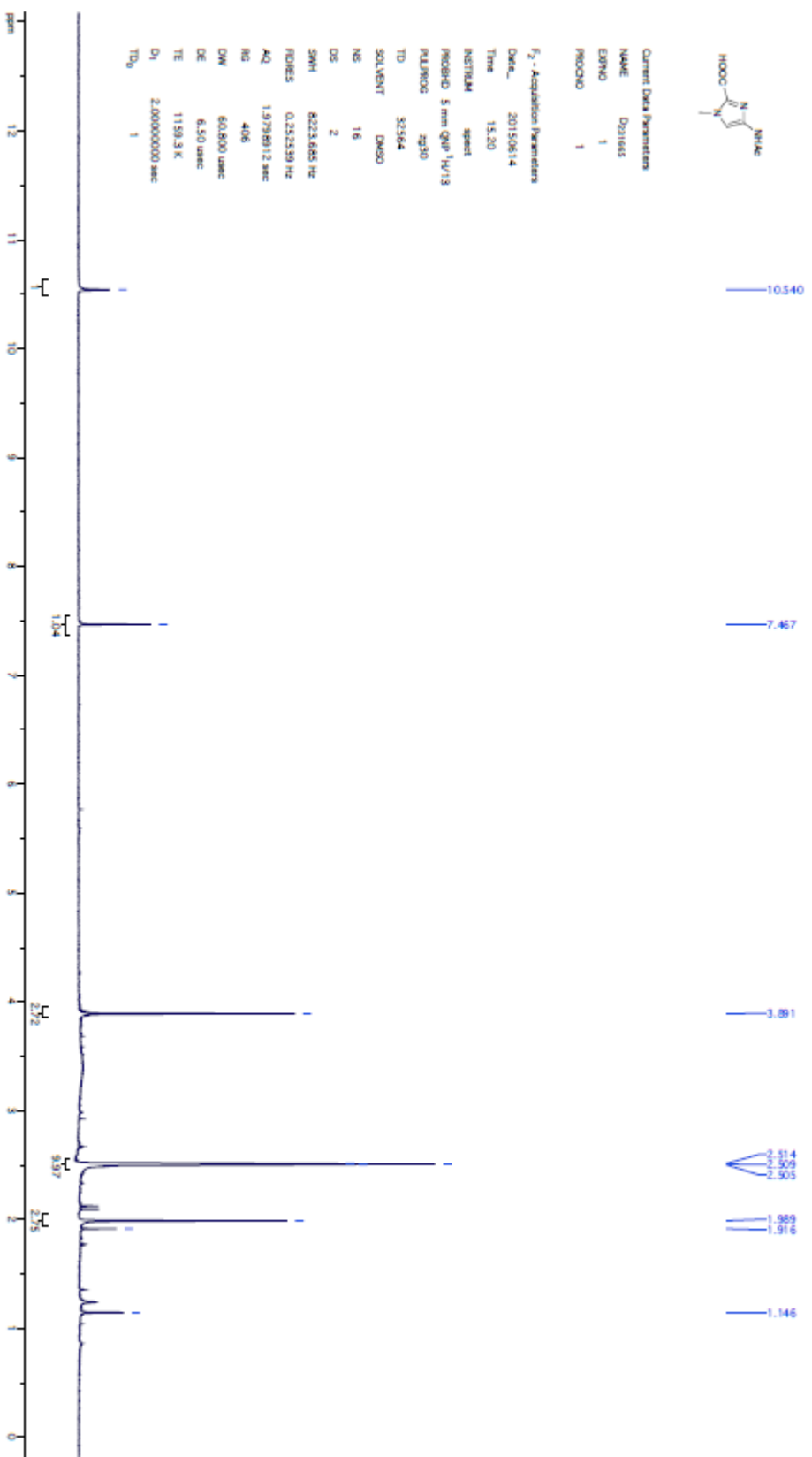


Annex-5 H NMR characterisation data of 4-acetamido-1-methylimidazole-2-carboxylic acid



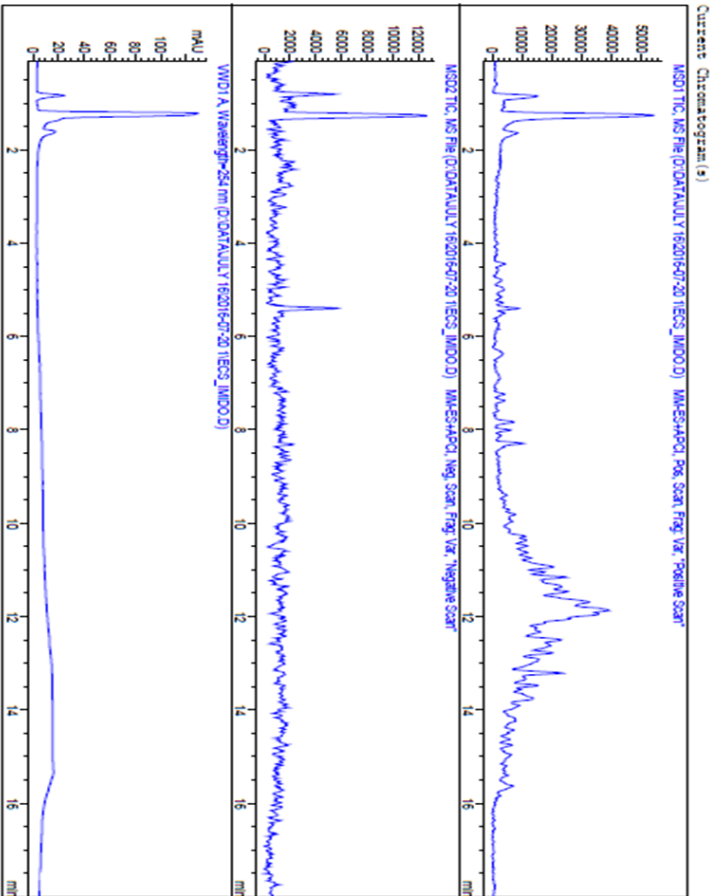
Current Data Parameters
 NAME D211615
 EXNO 1
 PROCNO 1

F₂ - Acquisition Parameters
 Date_ 20150614
 Time 15.20
 INSTRUM spect
 PROBHD 5 mm QNP 1H/13
 PULPROG zg30
 TD 32768
 SOLVENT DMSO
 NS 18
 DS 2
 SWH 8223.885 Hz
 FIDRES 0.252339 Hz
 AQ 1.9798912 sec
 RG 406
 DW 60.800 usec
 DE 6.50 usec
 TE 1159.3 K
 D1 2.00000000 sec
 TDB 1



Annex-6 LCMS characterisation data of 4-acetamido-1-methyl-imidazole-2-carboxylic acid

Acq. Method : D:\DATA\WVU 16\2016-07-20\1\TEST_F08M1C_254_11.M
 Last changed : 11/6/2015 8:58:31 AM
 Analysis Method : C:\CHEM32\1\METHODS\REPORTING_F080.M
 Last changed : 7/20/2016 12:52:36 PM
 (modified after loading)

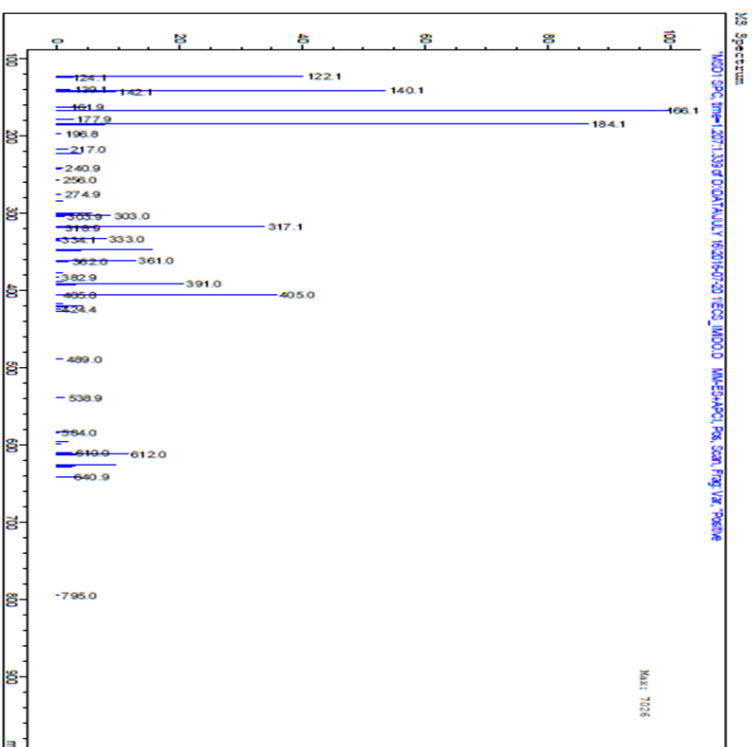


Instrument 1 7/21/2016 2:48:48 PM

Print of Window 80: MS Spectrum
 Data File : D:\DATA\WVU 16\2016-07-20\1\TEST_F08M1C_254_11.M
 Sample Name : F08_1.mz

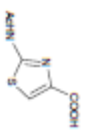
Acq. Operator :
 Acq. Instrument : Instrument 1
 Injection Date : 7/20/2016 10:36:56 AM
 Inj Volume : 10.000 µl

Acq. Method : D:\DATA\WVU 16\2016-07-20\1\TEST_F08M1C_254_11.M
 Last changed : 11/6/2015 8:58:31 AM
 Analysis Method : C:\CHEM32\1\METHODS\REPORTING_F080.M
 Last changed : 7/20/2016 12:52:36 PM
 (modified after loading)



Page 1 of 1 Instrument 1 7/21/2016 2:48:48 PM

Page 1 of 1

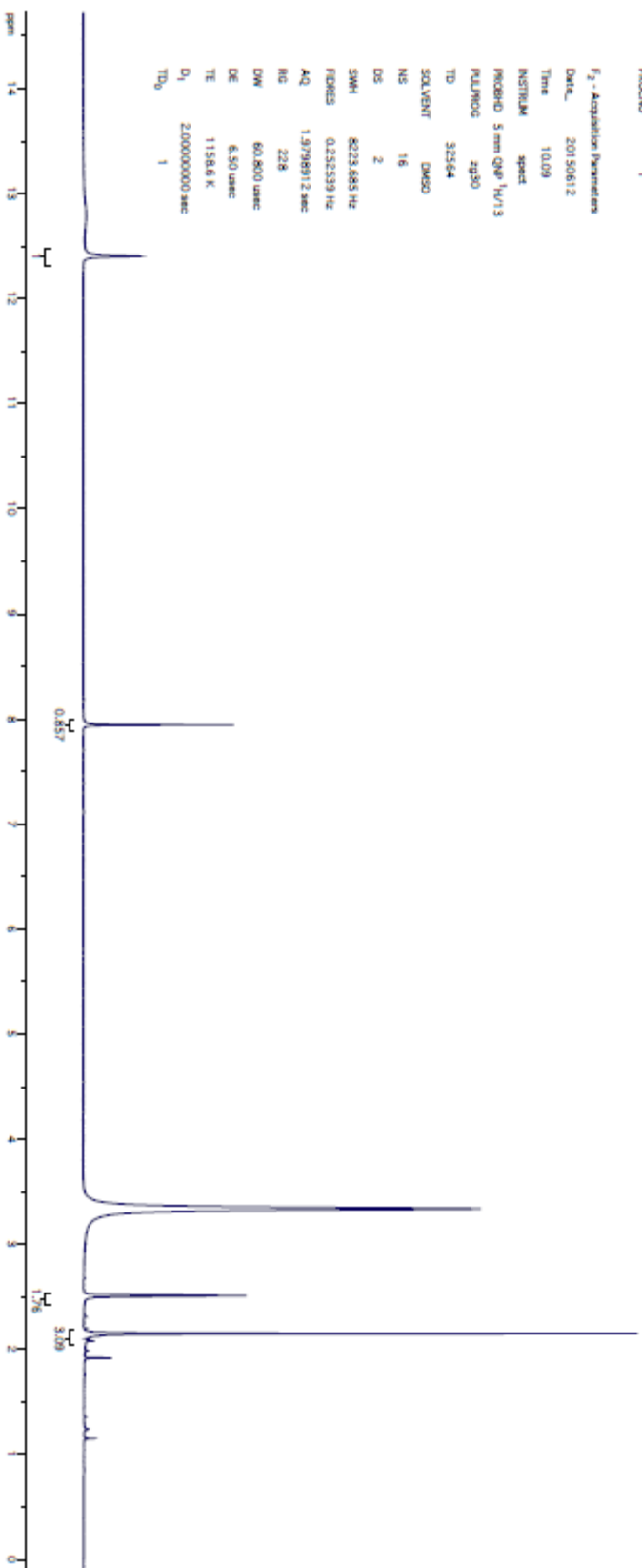


Annex-7 ¹H NMR characterisation data of 2-acetamidothiazole-4-carboxylic acid

Current Data Parameters
 NAME D_211449
 EXPNO 1
 PROCNO 1

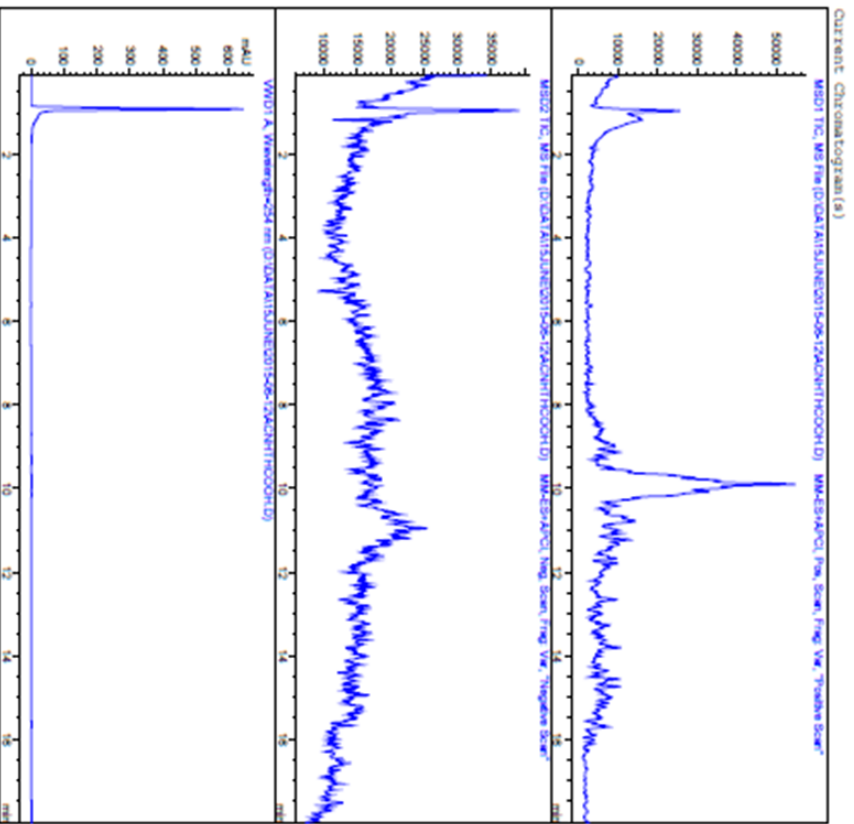
F₂ - Acquisition Parameters

Date_ 20150612
 Time 10.09
 INSTRUM spect
 PROBRD 5 mm QNP 1H/13
 PULPROG zg30
 TD 32768
 SOLVENT DMSO
 NS 16
 DS 2
 SWH 823.685 Hz
 FIDRES 0.252539 Hz
 AQ 1.9798972 sec
 RG 228
 DW 60.800 usec
 DE 6.50 usec
 TE 115.8 K
 D₁ 2.00000000 sec
 TD₀ 1



Annex-8 H NMR characterisation data of 2-acetamidothiazole-4-carboxylic acid

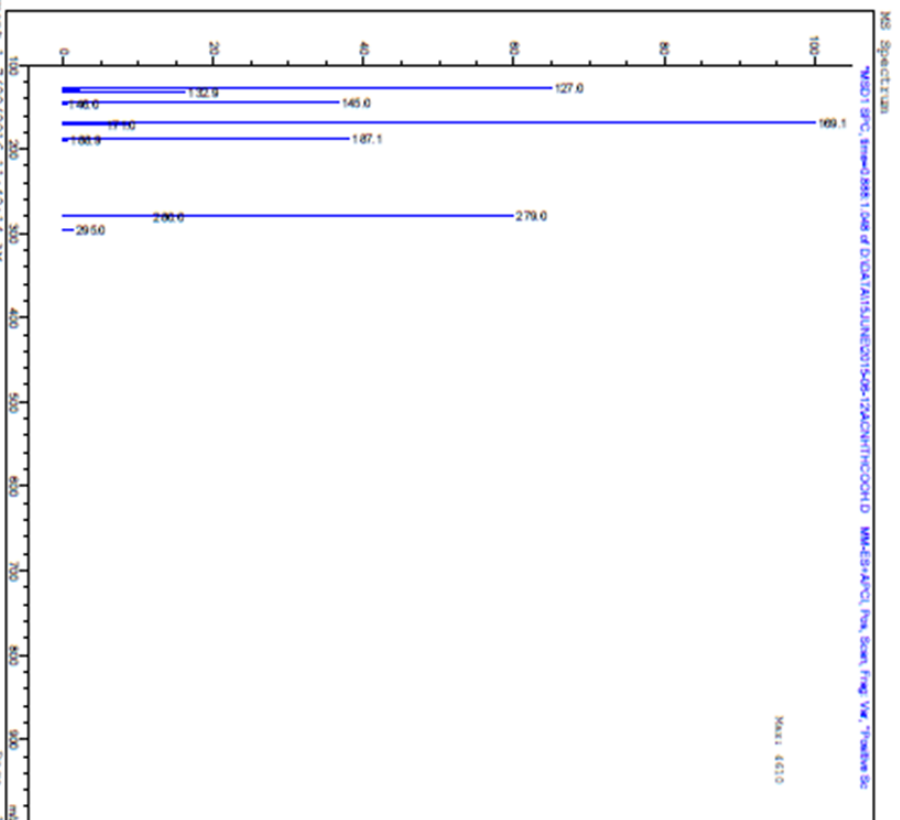
Acq. Method : D:\DATA\15JUNE\2015-06-12\PORO_A999CE_254.M
 Last changed : 2/12/2015 2:09:30 PM
 Analysis Method : C:\CHEM32\1\METHODS\REPORTING_PORO.M
 Last changed : 7/19/2016 10:58:43 AM
 (modified after loading)



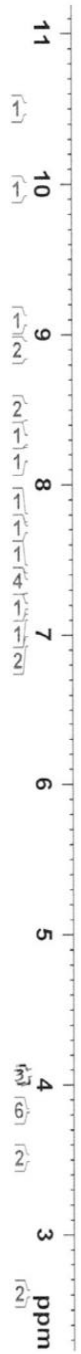
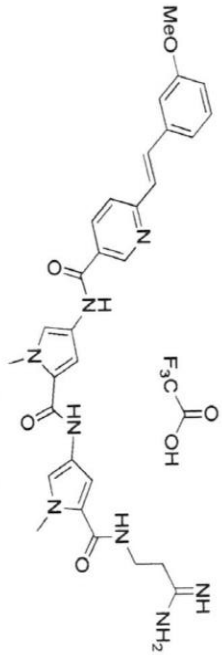
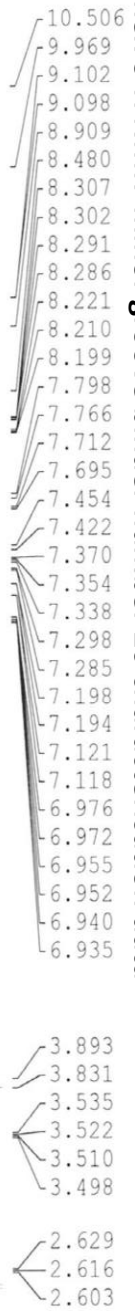
Print of Window 80: MS Spectrum
 Data File : D:\DATA\15JUNE\2015-06-12\ACNTRH008.D
 Sample Name : ACNTRH008

Acq. Operator :
 Acq. Instrument : Instrument 1
 Injection Date : 6/12/2015 12:24:10 PM
 Inj Volume : 10.000 µl
 Seq. Line : 11
 Location : VIA1 25
 Inj : 1

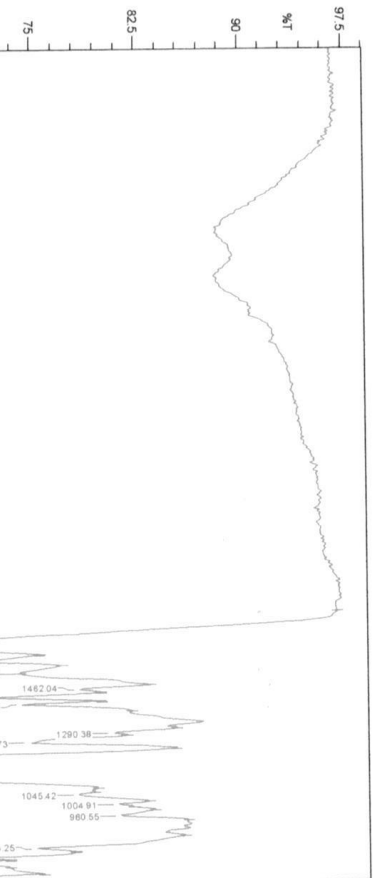
Acq. Method : D:\DATA\15JUNE\2015-06-12\PORO_A999CE_254.M
 Last changed : 2/12/2015 2:09:30 PM
 Analysis Method : C:\CHEM32\1\METHODS\REPORTING_PORO.M
 Last changed : 7/19/2016 10:58:43 AM
 (modified after loading)



Annex-9 H NMR minor groove binder number 26 characterisation data



Annex-10 IR characterisation data minor groove binder number 26



Current Data Parameters
NAME E14108
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20160520
Time 10.04

INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 32730
SOLVENT DMSO
NS 64

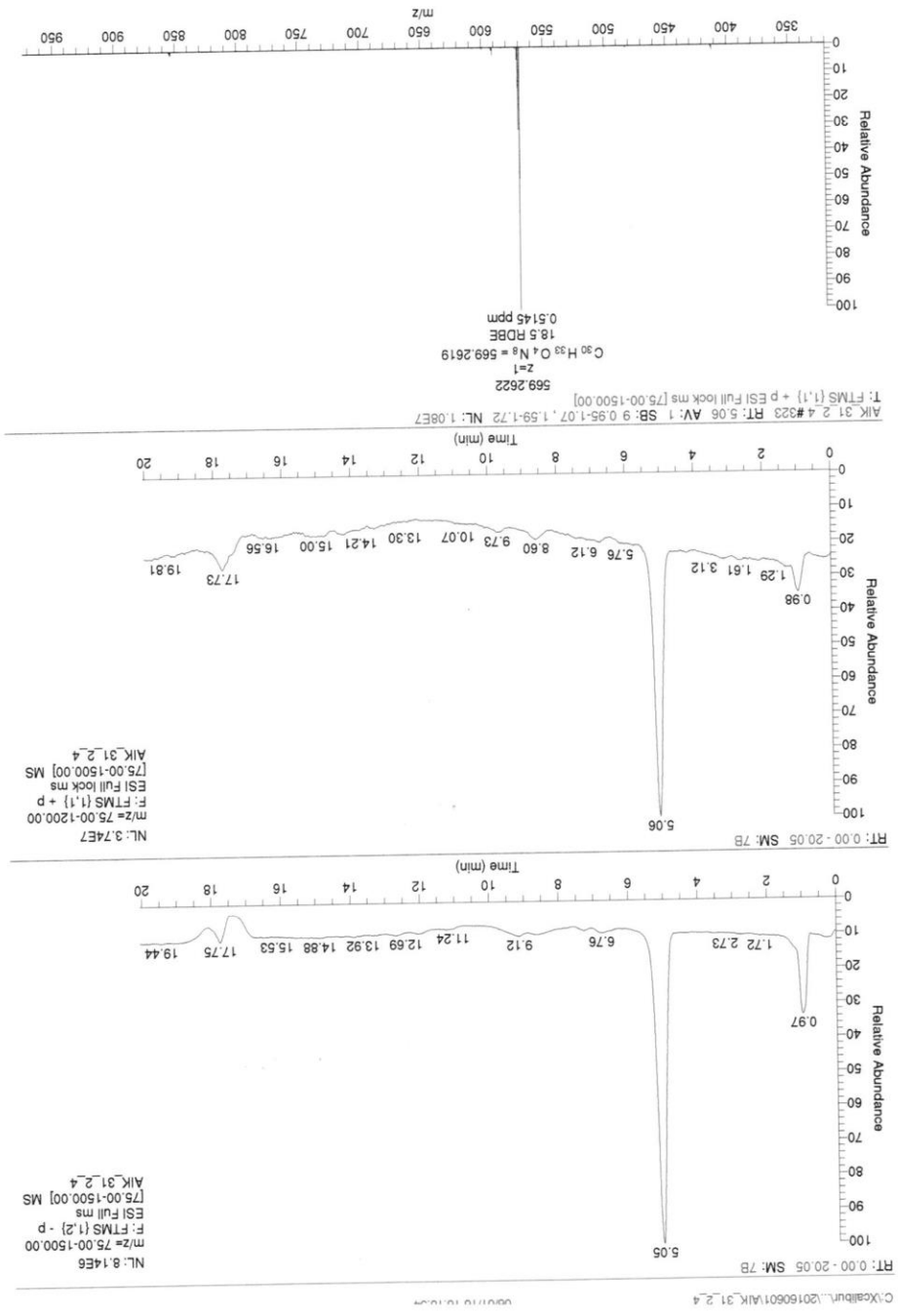
DS 2
SWH 10000.000 Hz
FIDRES 0.252666 Hz
AQ 1.9789000 sec
RG 198.22
DW 50.000 usec
DE 6.50 usec
TE 300.0 K
D1 2.00000000 sec
TD0 1

CHANNEL f1
SFO1 500.130885 MHz
NUC1 1H
P1 10.00 usec
PLM1 20.00000000 W

F2 - Processing parameters
SI 65536
SF 500.130000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

Person 3-16
AIK_31_2_4

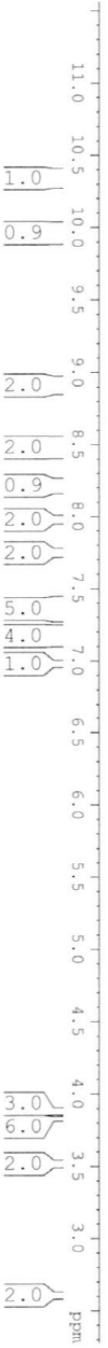
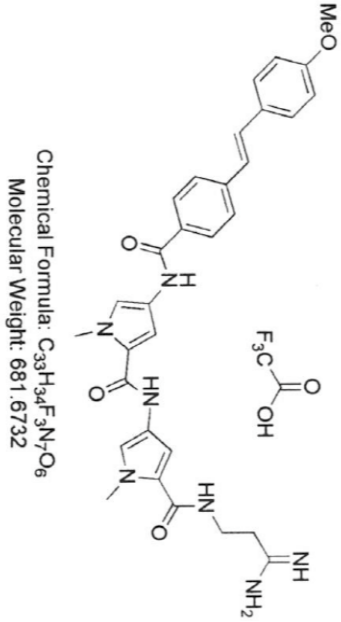
Annex-11 HRMS characterisation data minor grove binder number 26



Annex-12 H NMR characterisation data minor grove binder number 24



—	10.3277
—	9.9568
—	8.9085
—	8.4689
—	8.2178
—	8.2069
—	8.1956
—	7.9849
—	7.9685
—	7.7580
—	7.7415
—	7.3772
—	7.3733
—	7.3393
—	7.3302
—	7.3142
—	7.2250
—	7.1921
—	7.1237
—	6.9764
—	3.8856
—	3.8307
—	3.8226
—	3.5338
—	3.5223
—	3.5101
—	3.4984
—	2.6287
—	2.6154
—	2.6026



Person 3-16
 ATK_31_2_3

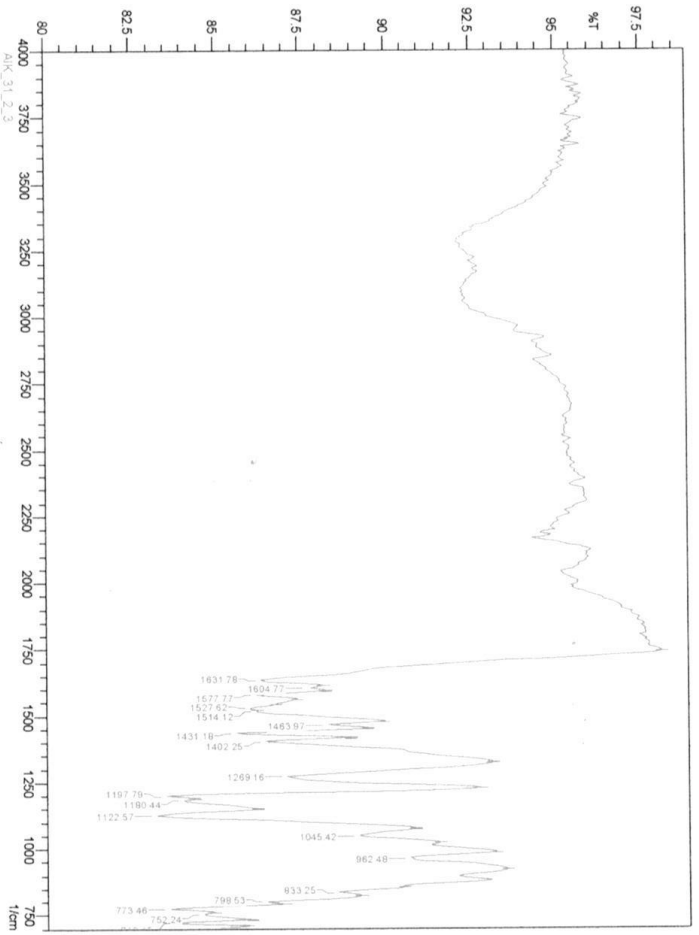
```

===== CHANNEL F1 =====
SFO1 500.1330885 MHz
NUC1 1H
P1 10.00 usec
SI 65536
SF 500.1300000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00
    
```

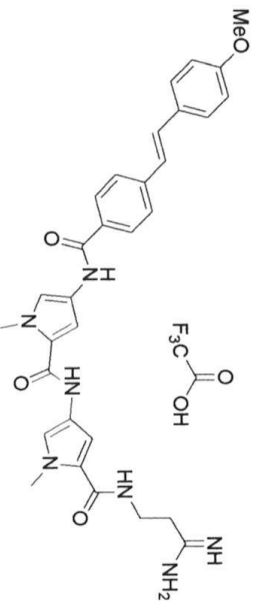
```

NAME E14013
EXPNO 1
PROCNO 1
Date_ 20160518
Time 15.33
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 2930
SOLVENT DMSO
NS 64
DS 2
SWH 10000.000 Hz
AQ 0.252666 Hz
FIDRES 1.9789500 sec
RG 198.22
DM 50.000 usec
DE 6.50 usec
TE 300.0 K
D1 2.00000000 sec
TD0 1
    
```


Annex-13 IR characterisation data minor groove binder compound 26



Peaks: 720, 752, 774, 799, 833, 963, 1045, 1123, 1180, 1198, 1269, 1402, 1431, 1464, 1514, 1528, 1578, 1605, 1632 cm^{-1}



Annex-14 HRMS characterisation data minor groove binder compound 26

