Unlocking the potential of novel *taxa* – a study on *Actinoallomurus*

João Carlos Santos Cruz

Submitted for the degree of
Doctor of Philosophy

School of Life Sciences,
University of Warwick

March, 2016

TABLE OF CONTENTS

ACKNOWLEDGMENTS

To those who gave me the opportunity to have this wonderful experience, who went above and beyond to make me feel welcome and ensure I was getting the most out of this project. For their advices professionally, but also personal, I thank you Dr. Stefano Donadio and Dr. Margherita Sosio.

I also want to thank Professor Elizabeth Wellington, who has always made an effort to help, support and guide me on this journey, despite being based different countries.

None of this work would have been accomplished without the guidance of Marianna, the support of Matteo or the comments of Alice; the advices of Paolo, Arianna, or Sonia were equally important, and last, but certainly not least, Cristina, Eleonora, and Stefania and their ever present helping hand. NAICONS Research group, I thank you all.

Because these three years were spent also outside of the Lab, I have to thank Marcello, Maxime, and Thomas for welcoming me in Milan with open arms and creating a bond that I will always cherish. Also to Pedro for keeping my Portuguese side alive while spending all this time in foreign lands. For all the nights out exploring Milan, the handball games played together or the lazy evenings spent at home: I thank you all.

Finally, my thanks go to those who have helped me in this journey even before it started. To my Father, Mother and Sister a big thank you for their unwavering support and love. I would not have done it without you.

LIST OF FIGURES

LIST OF TABLES

DECLARATION

In accordance with the University of Warwick regulations for the degree of Doctor of Philosophy, I certify that this thesis has been written solely by me. The work contained in this thesis is my own unless otherwise stated. No aspect of this work has been submitted to any other institution for any other degree of award.

SUMMARY

The discovery of antibiotics has played a significant role in raising public health standards and improving the outcomes for patients with bacterial infections. Unfortunately, drug discovery efforts have not kept pace with emergence and spread of antibiotic resistance.

This project is part of a discovery programme designed to assess the metabolic potential of *Actinoallomurus*, a recently described genus within the *Actinobacteria*. The schedule of work included small-scale fermentations (from 200 strains) followed by bioactivity testing. The most promising positive samples were then subjected to a combination of bioassays with chemical analyses to determine the chemical identity of the active compounds, with the objective of finding novel antibiotics. The Results section is organized in four different but closely related chapters. These present *Actinoallomurus* as a potential new source for active metabolites. Firstly, by showing, in a sample of around 200 strains, over 40% inhibited antimicrobial growth in the conditions tested but also by adding to the elevated quantity of positive hits a high degree of diversity, with the major biosynthetic pathways being identified whether in the molecules screened or also in the genomes analysed.

To illustrate not only the type of metabolites produced by *Actinoallomurus*, but also the multidisciplinary of Drug Discovery expeditions three chapters characterizing new compounds are presented followed by a genomic characterization of four *Actinoallomurus* Strains. These three chapters show how the delivery of new potential drug candidates demands an always broader set of diverse tools to be applied. While the three new molecules support not only the ability of *Actinoallomurus* to produce interesting active metabolites, but also the importance that a proper strain library plays in drug discovery, the genomic analyses predict, as for many other actinomycetes, an even brighter future, since the genetic potential for producing bioactive molecules surpasses the compounds that can be detected under routine cultivation conditions.

ABBREVIATIONS

AcOH	Acetic Acid
ACP	Acyl Carrier Protein
AMP	Antimicrobial Peptides
BGC	Biosynthetic Gene Cluster
CBE_EtoAc	Clear Broth Extract Organic phase
CBE_Exh	Clear Broth Extract Aqueous phase
CDS	Coding DNA Sequences
CFU	Colony Forming Units
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High Pressure Liquid Chromatography
HR-MS	High Resolution Mass Spectrometry
KS	Ketosynthase
LA	Lost Activity
LAF	Lost Activity after Fractionation
LC-MS	Liquid Chromatography Mass Spectrometry
LOW	LOW activity

MeCN	Acetonitrile
MeOH	Methanol
MES	2-(*N*-morpholino)ethanesulfonic acid
MIBiG	Minimum Information on Biosynthtic Gene Cluster
MIC	Minimum Inhibitory Concentration
MS	Mass Spectrometry
MycExt	Mycelium Extract
NAICONS	New Anti-Infective Consortium
NGM	Nematode Growth Media
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser effect spectroscopy
NRPS	Non-Ribosomal Peptides Synthases
NC	Not Characterized Yet
ORF	Open Reading Frame
PCP	Peptidyl Carrier Protein
PCR	Polymerase Chain Reaction
PKS	Polyketide Synthase
PMV	Packed Mycelium Volume
R&D	Research and Development

RCF	Relative Centrifugal Force
RiPP	Ribosomally synthesized and Post-translationally modified Peptide
RNA	ribonucleic acid
Rpm	Revolution Per Minute
rRNA	ribosomal ribonucleic acid
Spp	Species
TFA	Trifluoroacetic Acid
TOCSY	Total Correlation Spectroscopy
UV	Ultraviolet
WIP	Work in Progress

1. INTRODUCTION

1.1. State of the art - antibiotics

Antibiotics are one of the most successful forms of chemotherapy in the history of medicine. The miscellaneous family of (usually) low molecular–weight compounds able to inhibit the growth of other microbes has been a key player in raising the society health standards, and setting the grounds for modern medicine. By the end of the 20th century's first half, antibiotics held a great promise regarding the eradication of infections as a major health threat and by the mid-1980s infectious diseases were considered virtually conquered as antibiotics kept being introduced into clinical practice. (Coates et al., 2011; Davies and Davies, 2010; Powers, 2004)

However, in this very success lies also the main issue: antibiotics are a unique class of medicines with characteristics unlike most others, complicating and restricting their use and long-term efficacy. The use of antibiotics inevitably creates a selective pressure for sensitive microorganisms to escape the antibiotic's toxic effect by becoming not susceptible to that same antibiotic. (Wise, 2002) Fig. 1.1 depicts the temporal succession of when the first instance of resistance to an antibiotic was observed along the year of introduction of that same antibiotic into the market. The striking observation from Fig. 1.1 is that the longest it took for resistance to be recorded in a given pathogen was 15 years in the case of erythromycin; in most cases,

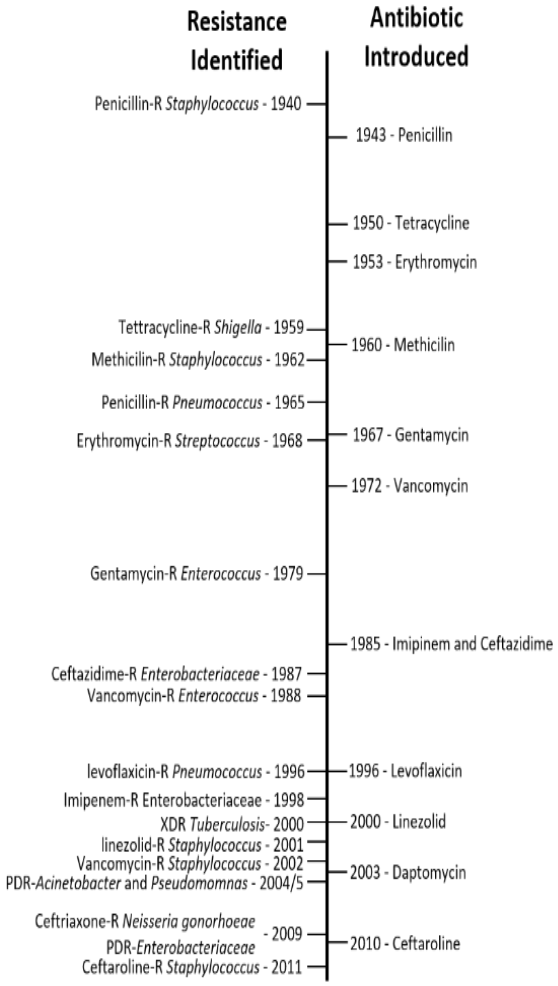


Figure 1.1 - Timeline correlating the introduction of an antibiotic into the market and the first resistance to that same compound was observed. (adapted from http://www.cdc.gov/drugresistance/about.html - October, 2015)

resistance was seen shortly after introduction of the corresponding antibiotic into the market or, as in the case of penicillin, even before.

The closeness of these two events (beginning of antibiotic use and observation of the first resistant bacterium) shows how the introduction of new antibiotics into medicine will never be a one-time event but a continuous effort will be necessary to keep bacterial infections under control by supplying the antibacterial arsenal in a continuous and sustainable manner. However, while we must accept that resistance will eventually develop to any antibiotic, the misuse of this precious class of compounds has played also a very important role in promoting the selection of particularly hard to kill bacteria. Self-medication approaches, failure to complete treatment, or erroneous prescription of antibiotics are just examples often encountered in our society that may lead to the proliferation of highly resistant bacteria. This ill-use of antimicrobials is not limited to human medicine: in animal husbandry, the prophylactic use of antibiotics to prevent infections or their use as growth promotants has represented the major use for some classes of antibiotics, providing a constant selective pressure for resistant strains and their spread from animals to humans, or vice-versa (Gould, 2009).

Antimicrobial resistance is not a recent event, since it naturally occurs in the environment and, recently, it has been documented in bacterial strains from at least 30,000 years (D/'Costa et al., 2011). Yet, now more than ever, we are witnessing the spreading and increase of drug-recalcitrant bacteria – for instance multidrug-resistant bacteria caused 25,000 deaths in the European Union in 2007, and antibiotic-resistant infections are estimated to result in excess health-care costs in the USA of US$20 billion per year, (Fauci and Morens, 2012) proving that a war against infectious agents is far from over and indicating that investment into new therapeutic options should not be disregarded (Donadio S., 2005). The increasing number of drug resistant infections with at least 480 000 new cases of multidrug resistant tuberculosis up to 2008 clearly portrays the landscape of drug resistant microbes and the risks that we, as a society, face in a near future (it is important to mention that similar figures are also observable beyond the antibacterial sector). Therefore, it seems straightforward to conclude that addressing antibiotic resistance is indeed an urgent matter.

Antibiotics can induce cell death (bactericidal drugs) or merely inhibit cell growth (bacteriostatic drugs). Both have been successfully used in the treatment of infection by targeting critical pathways such as: cell wall synthesis, DNA replication, RNA transcription,

protein synthesis. To overcome these hurdles, bacteria have come up with different resistance strategies (Hermann, 2000; Kohanski et al., 2010; Nikaido, 2009). Relevant resistance mechanisms include: drug modification (*e.g.*, β-lactamases that hydrolyse penicillin) (Bush, 1989) ; target modification (*e.g.*, mutations in the 30S ribosomal protein RpsL confer resistance to gentamycin) (Springer et al., 2001); restricted penetration and/or efflux of the drug (*e.g.*, efflux of linezolid by the AcrAB–TolC multidrug pump in *Escherichia coli*) (Eliopoulos et al., 2004); or reduction of the intracellular target concentration (*e.g.*, decreased expression of topoisomerase IV grants resistance to quinolones) (Ince and Hooper, 2003).

Targeting critical functions as antibiotics may lead to the belief that resistant strains suffer a decrease in biological fitness. Thus, a possible strategy to deal with antibiotic resistance could rely on a reduced use of that particular antibiotic, with the expectation that there will be a selective advantage for the resistant bacteria to lose the resistance mechanism and become more fit, thus outcompeting resistant mutants over time (Andersson and Levin, 1999). However, studies have been showing that reducing the usage of an antibiotic as the sole solution to antibiotic resistance is not a viable option, as evolution might not necessarily do what we would hope for. Indeed, not all resistance mechanisms come with a fitness cost, (Luo et al., 2005; Sander et al., 2002; Sundqvist et al., 2010) or compensatory mutations can occur that make reversion to sensitivity an unlikely event. Horizontal gene transfer of resistance determinants further complicates the scenario (Martinez et al., 2007).

For all these reasons above, improving our antibacterial arsenal is imperative in order to provide alternative options to the most commonly used antibiotics. Enhancing the already available antibiotics to make them active against resistant strains has been the favourite approach by the pharmaceutical industry. However, this approach is not generally applicable to all antibiotic classes and it will just buy some time in the war against bacterial infections. It is widely recognized that only novel chemical structures will bring truly novel mechanisms of action and, most likely, will effectively surpass known mechanisms of resistance (Donadio et al., 2010). This rationale was the trigger for the work developed in this PhD project – to discover new structural classes of antibiotics with novel mechanisms of action.

During the last 20 years of research and development, the pharmaceutical industry has not kept pace with the rise of multidrug resistant pathogens and the number of new drug applications in this area it has been steadily diminishing, as described in Fig. 1.2 (Butler *et al*., 2014). Different factors have contributed to the lack of new antimicrobials, but they can be reconducted to three main categories: biological, regulatory and economic factors. From the biological point of view, finding new antibiotics is not easy, since microorganisms have devised effective barriers (cell wall, membranes and efflux systems) to cope with chemical warfare in their natural environments. From the regulatory side, approval of new antibiotics requires meeting high clinical standards to enter the market, in which new compounds must show equal efficacy to existing drugs in a scenario where there is a lack of appropriate diagnostic tools that reflect the necessities of the healthcare system (e.g. in clinical trials it is very difficult to evaluate the drugs performance against multi-drug resistant pathogens). From an economical point of view, for an industry that covers the enormous costs of clinical development through sales of approved drugs, antibiotics against bacterial infections, an acute disease, are seen as a lower source of income compared to drugs that deal with chronic diseases, since the former will always have a smaller sale volume in respect to the latter. Furthermore,

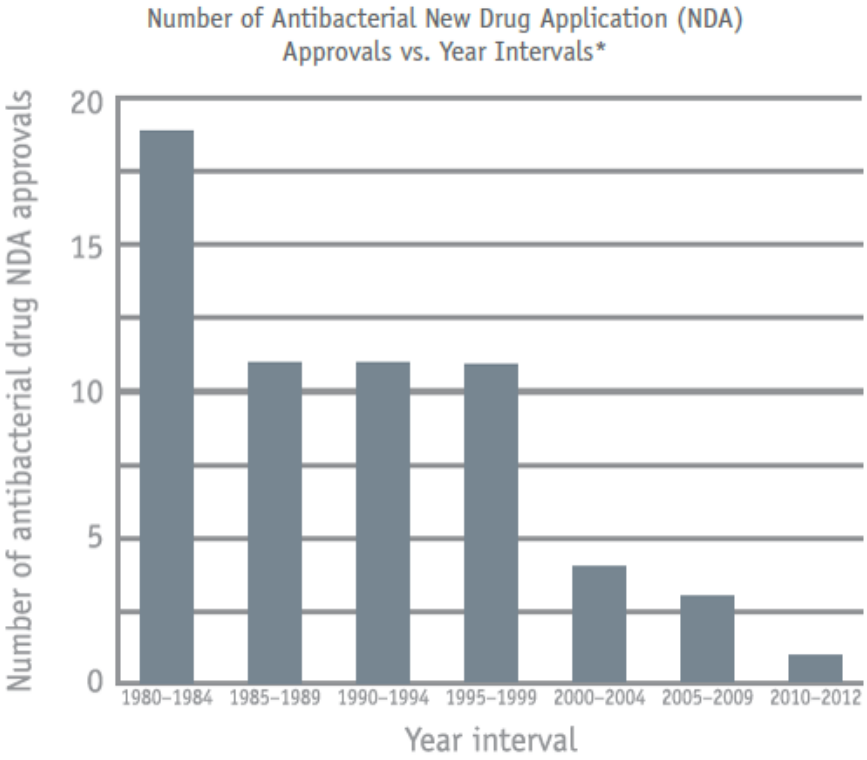


Figure 1.2 - Number of Antibacterial drugs approved in the United States from 1980 until 2012. * time intervals of 5 years except from 2010 to 2012. (Prevention, 2013)

in the case of antibiotics, a correct public health policy may curb their prophylactic use and reduce the revenue stream (Cassir et al., 2014; Gould; Michael et al., 2014; Shlaes, 2015; Tängdén and Giske, 2015). For large pharmaceutical companies that deal with many different therapeutic areas, antibiotic discovery and development represents a business with a low return on investment, with the result that largest pharma companies have abandoned antibiotic R&D altogether.

Still, before leaving the antibacterial discovery sector, many pharma companies heavily invested in a screening strategy that has essentially failed to delivered new scaffolds useful for further optimization. Indeed, the genomics era in the 1990's unveiled a plethora of new potential targets (essential genes conserved into many pathogens and not present in humans) that were explored by high throughput screening technologies on large collections of synthetic compounds derived from either historical (and thus biased) company collections or *de novo* generated through combinatorial chemistry. The underlying assumption to this approach was that any hit could be transformed into a lead compound by synthetic chemistry. Although this approach was successful in delivering potent inhibitors of selected targets, most of those hits turned out to be poor antibacterial agents, often because of inability to reach their targets within the cell cytoplasm. Although there is still a dearth of compounds that are able to keep their interactions with a predetermined target in more complex environments, these last years of research did deliver a better understanding of the microbial cell by identifying critical functions within the cell's metabolism (Arya et al., 2001; Collins and Donadio, 2011; Feng et al., 2007; Harvey, 2000; Hilpert et al., 2005; Pena et al., 2015; Scanlon et al., 2014).

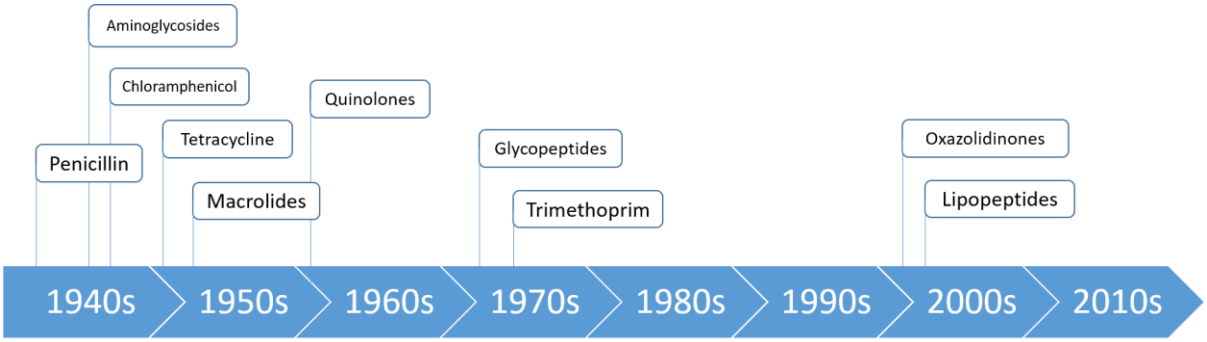


Figure 1.3 - Timeline of the introduction of new antimicrobial classes into the market (Adapted from Högberg, LD., *et al*. 2012) (Hogberg et al., 2010)

Figure 1.3 shows the decade in which the first representative of a new class of antibiotics was introduced into the market and how the 40's and 50's yielded as many new classes as the

forthcoming 60 years. Up to 2012, among the 13 new antibiotics introduced into the market there are only two new classes: linezolid and daptomycin, an Oxazolidinone and a Lipopetide, respectively (Fig. 1.3) (Jabes, 2011). Moreover, the actual discovery programs that led to those drugs date back many years. For instance, linezolid belongs to a class known since the 50's, the first molecules with antibacterial bacterial activity were described in the 70's (Barbachyn, 2012). Only in 1984 its potential for treatment in humans is registered in a patent. Then, E. I. du Pont de Nemours and Company - the original discoverer of the antimicrobial properties of this class - shuts down the development process due to toxicity issues (Moellering, 2003). Later on, in the 90's, Pharmacia & Upjohn (now part of Pfizer) reactivated the Oxazolidinone program and developed new derivatives, amongst which was linezolid, that eventually received regulatory approval in 2000 (Barbachyn and Ford, 2003). Daptomycin, on the other hand, was originally discovered by Eli Lilly but later licensed and registered by Cubist (now Merck) (Tally and DeBruin, 2000). More recent examples of approved antibiotics include fidaxomicin, a new chemical class approved in 2011 for treating *Clostridium difficile* infections, (Mullane, 2014) and dalbavancin, a second generation glycopeptide approved in 2014. The latter represents a dramatic example of a tortuous history, with six different companies contributing to its development (Maffioli et al., 2016).

Although the antibiotic landscape is not attractive to big pharma, it represents an interesting area for smaller companies. In fact, despite a smaller market, the cost of clinical development is generally smaller for an antibiotic than for a drug in other therapeutic areas. Furthermore, in contrast to many other therapeutic areas, animal models of infections are highly predictive of efficacy in humans, so that the "attrition rate" (i.e., the probability that a drug will fail during its costly clinical development path) for antibiotics is far smaller than for other therapeutic areas, such as cancer or cardiovascular diseases. Thus, smaller companies can be attracted by the prospect of substantially "de-risking" a new antibiotic after the less expensive early clinical trials (Phase I and Phase 2). Rather than competing with big pharma on the so called blockbuster drugs, smaller companies take a limited financial risk and can later partner in late developmental stages to bigger companies.

It should be noted most of the last approved antibiotics (daptomycin, fidaxomicin, dalbavancin) were discovered using traditional methods and represent either natural products or derivatives thereof (Harvey et al., 2015). This fact highlights not only the success of the classical approaches but also the potential of natural products to deliver promising leads, as discussed in the next section.

1.2. Importance of natural products

Nature has contributed to improve our health and overall quality of life since the dawn of times. We passed from patching moistures of plants over injuries to chemically mimic natural products and even improve their properties. Any compound produced by a living organism can be considered a natural product. In this work, I will restrict this definition to the so-called secondary or specialized metabolites, molecules that are not associated with most living organisms (i.e. primary metabolites) but that are usually specific to a particular group of organisms, often a single species or strain. Most of these compounds do present a detectable biological activity, although it depends on the type of tests we apply. Nonetheless, these compounds are important for the producers as they represent tools to interact with the outer world, be it friends or foes. Quorum sensing is the classical example of microorganisms communicating using specialized metabolites. However, not all metabolites aim only at communication: some are used to gain a competitive advantage, usually by impairing the viability of predators or competitors in a very challenging environment. It is exactly that kind of molecules that we, as humanity, have always used and are still interested in.

The importance of natural products to human health has been acknowledged over the past decades. A recent confirmation came in 2015, when the Nobel prize for medicine was awarded to William C. Campbell and Satoshi Ōmura for their discovery of avermectin, a novel therapy against infections caused by roundworm parasites, and to Youyou Tu for or her discoveries concerning a novel therapy against Malaria. Avermectin is a molecule produced by the actinomycete *Streptomyces avermitilis*, which was identified through a systematic screening for anthelminthic compounds. While artemisinin is a molecule produced by the plant *Artemisia annua*, which had been used in traditional Chinese medicine as a treatment for malaria. Natural compounds have found applications in virtually all therapeutic areas, from pain relief (morphine derived from poppies was the first isolated compound to enter the market back in 1817 (Luch, 2012)) to the treatment of patients with high cholesterol (mevastatin, a fungal metabolite, served as scaffold for the synthesis of atorvastatin (Endo et al., 1976)), and of course, penicillin, the first antibiotic introduced into the market (Glynn, 2005). Supporting these examples is the actual number of new chemical entities approved as drugs between 1981 and 2010, where approximately 68% are directly derived or inspired from natural products.

This success, despite the growing disinterest by pharmaceutical companies in natural products as a source of drug leads, underscores Nature's ability to deliver complex and high affinity chemical scaffolds that have no match in most synthetic libraries. One therapeutic area that has been strongly dependent on natural products has been anti-infectives where, in the last 30 years, out of the 104 new chemical entities introduced into the market to fight bacterial infections, 77 derived from natural products (Newman and Cragg, 2012).

Natural products as defined above are mainly produced by bacteria, fungi, and plants, however my focus will be on those of bacterial origin. As mentioned, secondary metabolites are believed to play a role in interactions and communication with the surrounding environment. This expansion in functions results in different specializations, hence an increase in diversity. However, not all bacteria are efficient at producing secondary metabolites. Some groups of bacteria have evolved to become effective producers of secondary metabolites while other groups did not. This is because production of secondary metabolites comes with a cost associated with the often complex pathways required to produce these molecules which can represent a competitive advantage in some environments and under some circumstances. For example, many *Actinobacteria* (the taxonomic group containing the large majority of secondary metabolite producers) gave up on a fast growth rate as their genome increased to harbour, among other functions, the genetic information for the biosynthesis of specialized metabolites. Their life cycle became complex, involving growth as a multicellular mycelial structure probably important for catabolizing complex organic polymers, an abundant component of soil, their typical habitat, and ultimately producing spores, which can then colonise new niches. Consequently, their life cycle typically involves four different stages: a spore germinates, leading to hyphae formation to create the mycelium that can fully colonize the solid support to which it is attached. Then, aerial hyphae starts differentiating from the substrate mycelium, eventually producing spores (Kalakoutskii and Agre, 1976). While this life cycle is best described for *Streptomyces* spp., it is believed to occur in similar ways in most filamentous actinomycetes.

Before endeavouring in a quest for new bioactive compounds among natural products, it is important to be aware of what kind of molecules one might encounter, which in turn depends on the types of biosynthesis machineries that are present in the group of bacteria under investigation, as discussed in the next section.

The polyketide natural products are produced by bacteria, fungi and plants. They represent a remarkable class of compounds not only for their chemical diversity, but also for the astounding range of important biological activities. This family of compounds has been able to deliver many clinically important drugs such as tetracycline and erythromycin (antibacterial), daunorubicin (anticancer), rapamycin (immunosuppressive), or lovastatin (cholesterol-lowering agent), making it the most successful and versatile group in terms of drug discovery. These compounds are assembled through the condensation of short chain fatty acids as Acyl-CoA precursors, in a process that has likely evolved from fatty acid synthesis. Nonetheless, important variations exist in the biosynthetic routes which have resulted in the classification of the responsible enzymes, the polyketide synthases or PKSs, into type I, II and III. (Shen, 2003)

Type I PKS, or PKS-I, were firstly described in the beginning of the 1990's with the discovery of erythromycin's biosynthetic pathway. (Cortes et al., 1990; Donadio et al., 1991) Probably the most studied system, PKS-I is organized in modules of pluri-functional enzymes, each one catalysing a cycle of polyketide chain elongation on an intermediate enzyme-bound through a thioester linkage. These modules contain, as core units, the domains required to load a new acyl-CoA unit, condense it to the growing carbon backbone and transferring it onto the next module. Other domains may be present in each module to process the growing polyketide, increasing the number of chemical functionalities present in the polyketide. At the end of the cycle, the enzyme-bound polyketide chain is released - and often cyclized - by a dedicated thioesterase. In addition to this general architecture, some modules lack a specific core domain (i.e. the acyltransferase unit to capture the acyl-CoA precursor) which is instead provided *in trans* by a separate, promiscuous acyltransferases capable of interacting with different modules (Nguyen et al., 2008). The PKS-I non-iterative system usually yields reduced and complex polyketides such as macrolides, polyethers or polyenes (August et al., 1998; Caffrey et al., 2001), however there are also PKS-I systems that act in iterative mode and yield aromatic polyketides, as type II PKS, or PKS-II, do (Gao and Thorson, 2008). Examples of molecules produced by PKS-I enzymes are shown in Fig. 1.4.

PKS-II were firstly described during the mid-80's with the analysis of the gene clusters responsible for the biosynthesis of tetracenomycin and actinorhodin (Malpartida and Hopwood, 1984; Motamedi and Hutchinson, 1987). PKS II relies on three distinct monofunctional enzymes, called a minimal PKS, formed by an acyl carrier protein (ACP), and a ketosynthase (KS) made of two subunits – alpha and beta – in which KSα makes the carbon-

carbon bond while KSβ contributes to specifying the length of the carbon chain and its appropriate folding for the subsequent cyclisation and aromatisation reactions (Daum et al., 2009; Wendt-Pienkowski et al., 2005). Examples of molecules produced by PKS-II enzymes are shown in Fig. 1.4.

Probably the least studied system, firstly reported in 1999, (Funa et al., 1999) is the type III PKS, or PKS-III, which lacks the ACP protein or domain required for covalently binding the growing intermediate. In PKS-III, chain elongation is performed on a CoA derivative, as with chalcone synthase, part of plant metabolism (Funa et al., 2002; Funa et al., 1999). PKS-III systems represent the structurally simplest member of the PKS superfamily and often provide specialized precursors to be used by other enzymes (e.g. non-ribosomal peptide synthases – see below) in a complex biosynthetic pathway (Pelzer et al., 1999). Examples of molecules produced by PKS-III enzymes are shown in Fig. 1.4.

In addition to the generation of secondary metabolites through the formation of carbon-carbon bonds as in polyketides, microorganisms can assemble biologically active compounds through the formation of C-N bonds, as in the synthesis of small peptides. Also this biosynthetic strategy can provide wide chemical diversity and, hence, different biological activities. Two main families of peptides are distinguished in bacteria according to their route of biosynthesis: the non-ribosomally synthesized peptides (NRPs) and the ribosomally synthesized and post translationally modified peptides (RIPPs) (Arnison et al., 2013; Finking and Marahiel, 2004).

NRPs are made by non-ribosomal peptide synthetases (NRPS), multi-enzyme complexes also organized in modules, where each module containing a set of enzymatic domains capable of recognizing and activating the substrate, delivering it to catalytic centres, where formation of the peptide bond occurs on an enzyme-bound intermediate. Therefore, in each module, three core enzymes are present: the A domain, which activates a specific amino acid and transfers it to the PCP (peptidyl carrier protein) domain, where it forms a thioester, and the C domain, which forms the peptide bond between the newly added amino acyl-ACP and the growing peptidyl unit. Additional domains may be present in each module for epimerisation, heterocyclisation or oxidation. NRPS biosynthesis usually yields macrocyclic structures, or dimers of identical structural elements. NRPS products have been key players in our antibacterial arsenal providing historical molecules as penicillin, our drug of last resort vancomycin, and one of the newest introductions, daptomycin. (Felnagle et al., 2008; Schwarzer et al., 2003) Examples of NRPS-produced compounds are shown in Fig. 1.5.

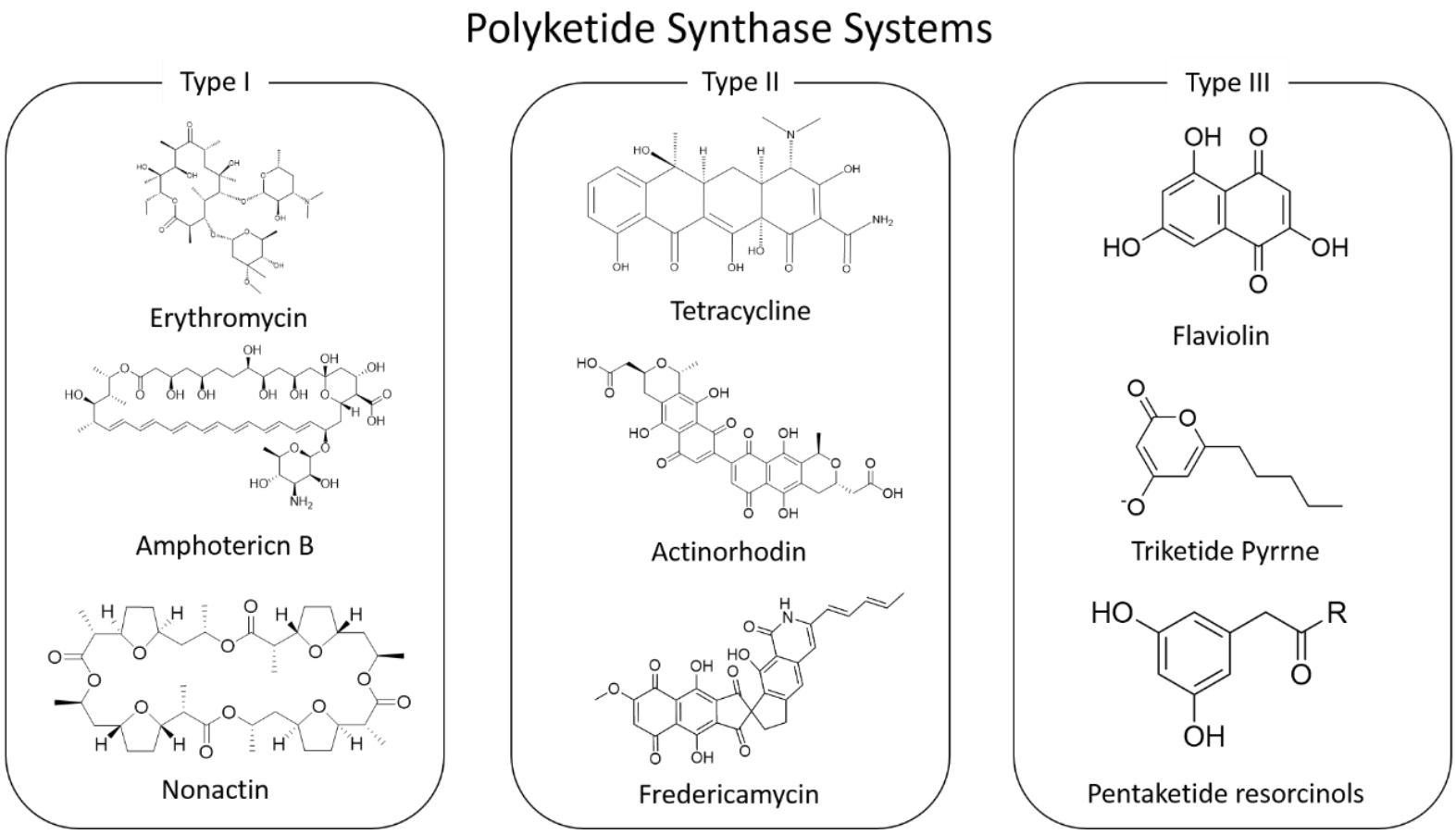


Figure 1.4 - Examples of PKS metabolites from microbial source.

Adding one more level of complexity are the hybrid systems that bring together a PKS-I with NRPS. Usually these hybrid systems operate by adding a few amino acids to a polyketide backbone or, conversely, a few acyl units to a peptide backbone. The biosynthetic logics of PKS-I and NRPS are comparable, since they both involve the progressive elongation and processing of enzyme-bound intermediates, which facilitates a transition between NRPS and PKS modules. This flexibility is underscored by the fact that ACP domains in PKS-I are biochemically equivalent to the PCP domains of NRPSs and are utilized as carriers of biosynthetic intermediates. The NRPS- and PKS-derived portions of the final molecule can vary, and there are examples of molecules constituted mainly by their peptidic part with few units added by the PKS, like bleomycin A2, but also the opposite can occur as in epothilone, a polyketide that incorporates one amino acid (Fig. 1.5) (Du et al., 2000; Molnar et al., 2000).

As mentioned, the other major class of peptide-based secondary metabolites is represented by the RiPPs. These molecules originate from a longer precursor peptide composed by an *N*-terminal leader peptide and a core peptide. The latter is post-translationally modified by dedicated enzymes, followed by proteolytic removal of the leader peptide and transport into its final, usually extracellular destination. According to their structural features, which in turn depend on the specific modifying enzymes, different classes of RiPPs exist, for example: the lanthipeptides (lanthionine containing peptides), with nisin, used in cheese manufacturing, as the most preeminent representative; or the Thiopeptides (characterized by several thiozole rings), for which thiostrepton, used in as a growth promotant, is a good example. It should be noted that seventeen additional classes of RiPPs have been documented (Breukink and de Kruijff, 1999). Examples of RiPPs are reported in Fig. 1.5.

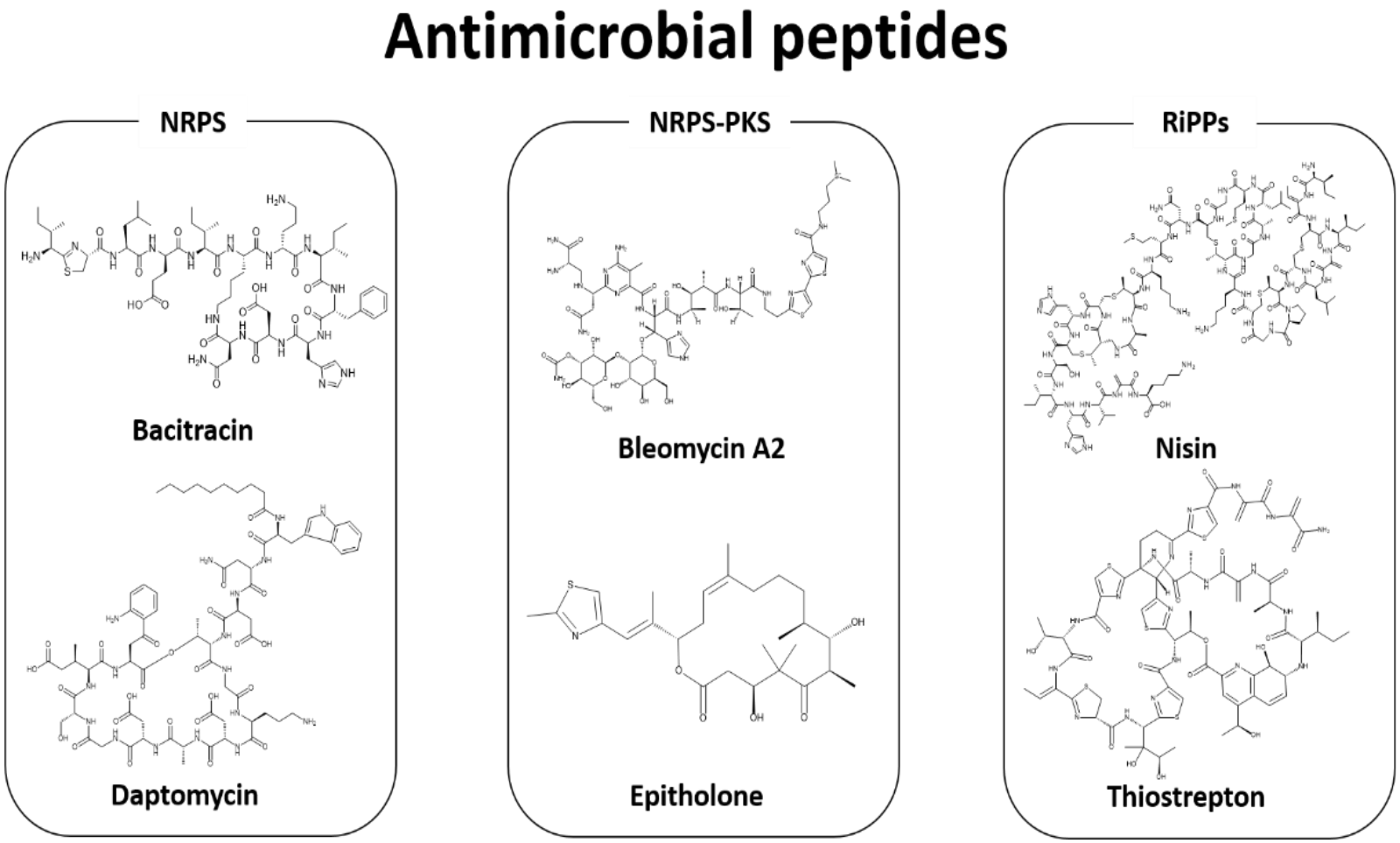


Figure 1.5 – Example of compounds with antimicrobial activity from NRPs (Actinomycin), RiPPs (Nisin), and hybrid systems (Bleomycin A2).

The terpenes, a subclass of prenylipids, are a large and diverse class of naturally occurring organic chemicals (Nowicka and Kruk, 2010). Their activities span from anti-parasitic and anti-microbial, to anti-inflammatory and immunomodulatory properties. These molecules are spread throughout a large portion of the animal kingdom with special emphasis in insects and plants. All together terpenes have become the largest group of natural products constituting about 60% of known natural products with artemisinin being a good example of terpenes' value to today's society (Ajikumar et al., 2008; Ishida, 2005; Tu, 2011). Most are multicyclic structures that differ from one another usually in functional groups but also in their basic carbon skeletons (de las Heras et al., 2003). Examples of terpenes are reported in Fig. 1.6.

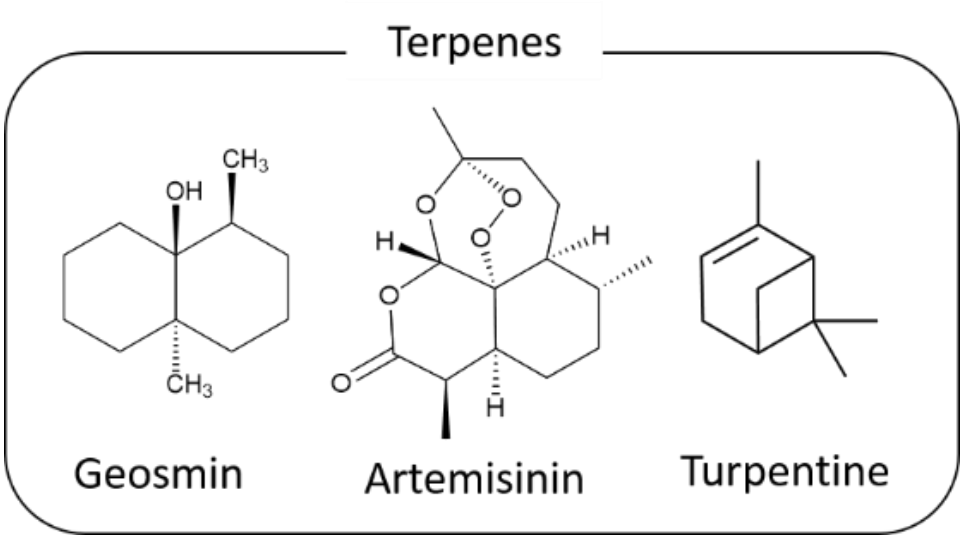


Figure 1.6 - Examples of biologically relevant terpenes

Finally, adding to the cosmic amount of molecules produced by the previously mentioned routes there are other groups that have contributed antibiotics in current or past clinical use such as (Fig. 1.7): aminoglycosides, such as streptomycin; amphenicols such as chloramphenicol; or lincosamides, such as the semi-synthetic clindamycin. These molecules originate from other, specialized biosynthetic pathways that fall outside the scope of this dissertation.

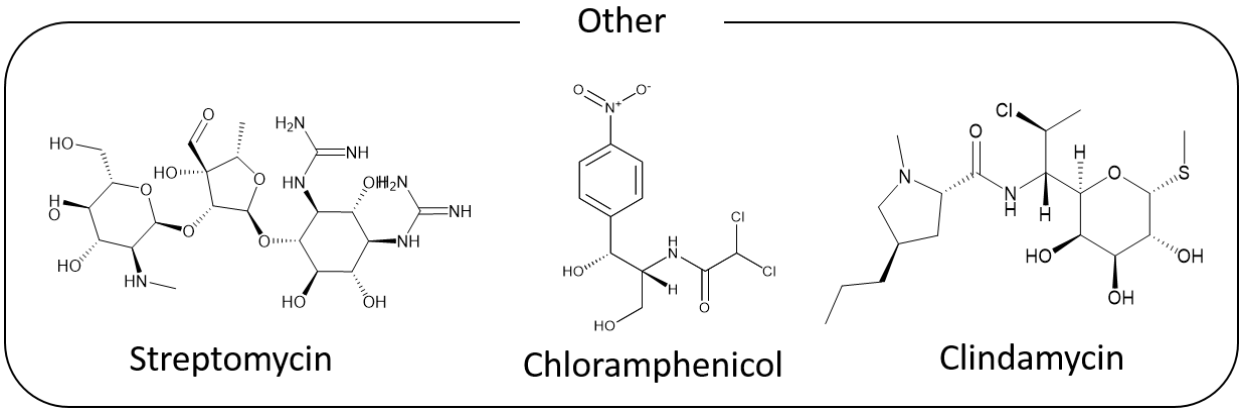


Figure 1.7 - Examples of biologically relevant aminoglycosides (Streptomycin), Amphenicols (Chloramphenicol) and Lincosamides (Clindamycin)

The mechanism of action of an antibiotic depends on the way it affects physiological processes within the target bacterial cell, which in turn results from which molecular target it binds to and how, a property that is intrinsically related to the chemical structure of the antibiotic. Resistance mechanisms that modify a cellular target prevent an antibiotic from binding to that target in a specific mode. These mechanisms confer to resistance to any antibiotic that binds to that same target in the same mode (e.g., in the same binding pocket) but not to other antibiotics that bind elsewhere on the same target. Similarly, resistance mechanisms that modify an antibiotic (for example, by hydrolysis or by acetylation) in order to be efficient, need to be highly selective for a particular chemical scaffold. As a consequence of the above, novel chemical classes of antibiotics, structurally unrelated to those in clinical use, have a high probability of bypassing current resistance mechanisms. [Resistance mechanisms based on export tend to be more promiscuous in terms of chemical class but less effective in the resistance level they confer] Hence, the search for new molecular structures is expected to unravel new mechanisms of inhibiting bacterial growth.

A strategy that relies on Nature's resources would seem the most reasonable approach to drug discovery. The richness and diversity of natural products has not yet ceased to amaze us (Harvey et al., 2015) (Walsh, 2003) (Strohl, 2000). Among plants, fungi, and bacteria, Nature has a cornucopia of metabolite producers waiting to be screened that we have just started to explore. In fact, the theoretical chemical space is estimated to be greater than 10^{60} molecules with 30 or fewer atoms other than Hydrogen (Bohacek et al., 1996). Which in turn could give some doubts on the probability of actually finding a new antimicrobial compound with the correct chemical features through this myriad of molecules. Furthermore, all natural

products are expected to have some biological activity, and indeed the high frequency in which these compounds are found when screening natural products shows that Nature is a rich and diverse source of compounds but also biased library towards biologically active compounds.

Although plants and fungi play an important role in healthcare, this work will focus on the potential of bacteria. Earth belongs to bacteria, not humans. Through their metabolic and catabolic processes, these microorganisms control and sustain Life on Earth. Despite their omnipresence we still know little about these small creatures. In fact, it is estimated that 99% of the microbial flora is as yet uncultured under laboratory conditions. This is understandable, since culturing in the laboratory requires devising appropriate growth conditions (pH, temperature, oxygen levels, nutrient concentration, essential cofactors, etc.) that can vary greatly for different bacterial taxa. Furthermore, some of our basic techniques like autoclaving produce compounds (e.g. Hydrogen Peroxide) that selectively kill certain taxonomic groups (Lewis, 2013; Tanaka et al., 2014). But perhaps the most limiting factor is that we like studying bacteria in axenic cultures, a condition rarely occurring in Nature. In 2004, around five thousand species of bacteria were described (the genus of this work – *Actinoallomurus* – was formally described in 2009, see below) (Bull, 2004; MacNeil et al., 2001; Tamura et al., 2009). Despite these limitations, culture-independent techniques and dedicated isolation efforts are continuously expanding our knowledge. However, estimates about the actual number of existing bacterial species vary greatly from 10^4 to as many as 10^7 (Whitman et al., 1998). Since different isolates from the same species can produce different secondary metabolites, there is ample scope for bacteria to deliver a large number of additional bioactive products. The challenge now is how to effectively access this natural chemical diversity.

1.3. Changing the screening paradigm

1.3.1. The screening strategies

Insanity has been defined as "doing the same thing over and over again and expecting different results" (Albert Einstein). This aphorism is certainly applicable to the most common approach employed in drug discovery, screening large collections of chemicals against a particular biological target, and especially to the approach universally employed during the golden era of antibiotic discovery: screening large collections of freshly isolated soil

microorganisms for their ability to inhibit growth of a relatively few target pathogens and keep expecting an infinite resource of novel chemical compounds. That is why it is imperative to change one or more of the screening conditions in order to increase the probability of finding new molecular structures (Donadio et al., 2010).

In the early days of pharmacology, therapeutic drugs were obtained from medicinal plants exploiting thousands of years of traditional medicine. This was the case for aspirin, which was derived from the bark of the willow tree. As analytical chemistry, biology and pharmacology progressed, the discoveries went from the occasional serendipity (i.e. the discovery of penicillin) to a rational approach (cephalosporin was discovered as a solution for penicillin resistant infections) (Clever, 1999).

At the beginning, drug discovery was essentially based on screening natural products, isolating the active ingredients and testing them for their ability to treat a given disease. Synthetic chemistry was mostly an aid in optimizing existing scaffolds against resistance mechanisms (e.g., analogues of penicillin that would be resistant to the action of known β-lactamases) or in conferring more drug-like properties (e.g., transforming nalidixic acid into the fluoroquinolone class of antibiotics). The continuous progress in spectroscopy (NMR and MS) and separation techniques (HPLC) enabled determining the chemicals structures using increasingly lower amounts of natural products, making this approach the pillar of drug discovery by the pharmaceutical industry.

In the early 1980s, the increases in computing power enabled a shift away from random searches to a "rational" computational approach to drug discovery, namely, computer-aided drug design. This shift was also possible thanks to significant advances in structural biology (and a growing number of solved 3D structures of biologically relevant proteins), an improved understanding of protein-protein interactions, and the identification of protein targets important for different pathologies. While there are a few successes of the "rational drug design" approach (e.g., dorzolamide reached the market as a antiglaucoma agent in 1995 (Greer et al., 1994)), this discovery approach performed surprisingly poorly and was supplanted by a return to the fundamentally empirical method of screening, albeit with important differences.

By mid-1990's a new strategy had taken over prompted, again, by new technological advances: combinatorial chemistry, laboratory automation and genomic sequencing. Combinatorial chemistry presented the opportunity of rapidly synthesising tens of thousands of compounds with a reduced effort, providing quick access to sensible libraries for screening.

Automation enabled the robotic handling of small amounts of liquids, enabling assays to be performed in small volumes in – originally – 96-well microtiter plates, which allowed increased throughput with a decreased use in reagents and chemicals. Finally, the ability to rapidly sequence whole bacterial genomes offered the possibility of rapidly accessing any target from any pathogen. At the same time, advances in the genetic manipulation of most bacterial (and fungal) pathogens enabled establishing the essentiality of any potential target, which had to be conserved in the desired spectrum of pathogens and absent, or significantly divergent, in humans. Such targets could then be expressed with a proper tag in a convenient host, single-step purified by affinity chromatography and turned into powerful and miniaturized enzymatic assays. The scene was then set for rapidly screening every essential target in a given pathogen against synthetic and combinatorial libraries in a relatively short time. As mentioned above, this approach also delivered below expectations (and the investment made!).

Science is a dynamic landscape that changes as knowledge is introduced and adapts to new discoveries. A great example of this feature is the case of the Lipinski's rule of five. In the late 1990s Christopher A. Lipinski created a set of rules of thumb to help choose molecules that could have higher potential in terms of drug-likeness and oral availability (Lipinski et al., 2001). These, shortly, consisted of:

- No more than five hydrogen bond donors
- No more than 10 hydrogen bond acceptors
- A molecular weight under 500
- A partition coefficient log P (a measure of lipophilicity) of less than 5

Although these are still valuable guidelines, we are aware that these rules do not apply to every drug in the market. In fact, antibiotics like erythromycin do not fit completely to these rules and Lipinski himself had warned about this caveat. This represents another example on the power of million years of evolution for natural products, which have thus reached a degree of complexity that cannot be restricted to a set of rules devised for synthetic drugs.

Therefore, a lack of truly effective screening alternatives, the availability of a large number of unexplored bacteria and the unusual chemical and biological of natural products make the latter source an attractive target for drug discovery.

1.3.2. Introducing elements of novelty

One of the main issues with antibiotic discovery expeditions based on natural products is the probability of rediscovering known metabolites. This requires introducing one or more elements of novelty into the screening endeavour, which in turn requires an understanding of the screening approaches employed in the past. Most of the past screening efforts involved isolating, among bacteria, strains belonging to the genus *Streptomyces* from a variety of sources, growing them in nutrient-rich media, and testing the resulting cultures for the ability to inhibit growth of one or more pathogens relevant at that time. The compounds responsible for bioactivity were then purified by bioassay-based fractionation, characterized for their antibacterial properties and structurally elucidated. This approach introduced a number of biases, namely: i) the type of microorganisms isolated and screened; ii) the conditions under which those strains were grown; iii) the type of pathogens that were screened against; and iv) the requirement for growth inhibition. Overall, this approach favoured the discovery of compounds that were potent and/or were produced in high amounts. Therefore, an element of novelty could address any of the above biases. For example, it could involve the use of previously unexplored bacterial group; or it could involve growing known strains under conditions that trigger the expression of the several biosynthetic gene clusters (BGCs) present in their genomes and not expressed under routine cultivation conditions; or it could involve screening against pathogens that are of medical relevance today but not considered important few decades ago (e.g., *Acinetobacter baumanii* or *C. difficile*); or it could employ assays more sensitive than growth inhibition of a standard strain.

An alternative, paradigm-shifting approach is the so called genome mining, in which genomic information is used as a starting point to identify metabolites predicted from the organization of the corresponding BCGs. If the strain does not normally produce the predicted metabolite, it can be engineered to do so by manipulating its regulatory circuits or the cluster can be moved and expressed in a heterologous host. While the genomic mining approach simultaneously addresses previous screening biases and it represents an enormous opportunity to unveil novel chemistry, awakening silent clusters is at the moment an ineffective trial and error procedure. For example, the most studied organism among actinomycetes, *Streptomyces coelicolor* is slowly delivering new chemical structures after the identification of the corresponding BGCs in its genome and thorough analyses of the metabolite profiles by different laboratories under a variety of conditions.

In the pursuit of my PhD program, it was decided to introduce an element of novelty by exploring different strains, namely a taxonomic group that had not witnessed extensive analyses of its secondary metabolites. Taxonomic diversity can be seen as a surrogate for chemical diversity (Jaspars and Challis, 2014; Kim et al., 2005). The main idea is that organisms that have been subjected to different evolutionary pressures have developed unique biology to survive and secondary metabolites are also part of their biology. While horizontal gene transfer can blur this picture and lead to the transfer of BGCs between unrelated bacterial taxa, as it does occasionally with housekeeping genes, one would expect that, on average, the BGCs present in a given taxon would be more related to each other than to BGCs present in an unrelated taxon. Thus, there is a higher probability of discovering novel BGCs, and hence novel chemistry, by exploring novel taxa.

Though we are aware of the universe of uncharacterized *bacteria* awaiting to be studied, we have also observed that production of secondary metabolites is not distributed equally amongst all species, with some groups – not surprisingly, those with large genomes – more prolific than others. Therefore, when endeavouring in a new expedition for bioactive natural products from new taxonomical groups it is important to select a taxon with a high potential to produce such compounds in order not to increase the probability of finding the desired compounds with a reasonable effort. A good example of a taxonomic group well known for the production of secondary metabolites is the phylum *Actinobacteria*, which stands out as being responsible for almost 9,000 molecules out of the 17,000 produced by microorganisms (Bull, 2004; Goodfellow and Fiedler, 2010). Among them is the genus *Streptomyces* which is the most studied and the one that has contributed with the largest amount of compounds (Bull, 2004; Goodfellow and Fiedler, 2010; Lazzarini et al., 2000; Pozzi et al., 2011; Stackebrandt et al., 1997; Tamura et al., 2009; Ventura et al., 2007; Watve et al., 2001). Previous reports have stressed the use of underexplored group of actinomycetes as an effective strategy to discover novel chemistry, an approach that has been recently reinvigorated by the results obtained from exploring unknown marine actinomycetes (Subramani and Aalbersberg, 2012). Moreover, drug discovery expeditions with new taxonomic groups have more than the obvious role of finding new drug candidates. The work goes beyond the screening itself, since it is also a characterization of the producer strains. For example, a significant amount of information is generated by simply recording which known metabolites are produced by the screened strains. This information will contribute to the better understanding of the chosen group, which can become an important factor for the success of posterior screening campaigns. In particular, this

work was designed to assess the metabolic potential of the genus *Actinoallomurus* while enriching the knowledge on this recently described group.

1.3.3. The Genus of *Actinoallomurus*

Associated to the *Actinobacteria* is the recently described genus *Actinoallomurus*, assigned to the family *Thermomonosporaceae*. This genus has been previously analysed in our laboratory and has showed a potential to produce different molecules (Pozzi et al., 2011). The genus *Actinoallomurus* was firstly described by Tamura *et al*. in 2009 while characterizing strains isolated from soil and dung samples collected in Japan (Tamura et al., 2009). Since its discovery at least eight new manuscripts reporting new entries in this genus were published, reporting the isolation of *Actinoallomurus* spp. as plant endophyte, from different soil samples and also from moss (Indananda et al., 2011; Koyama et al., 2012; Matsumoto et al., 2012; Pozzi et al., 2011; Thamchaipenet et al., 2010).

Actually, the first description of *Actinoallomurus* spp. resulted from a reclassification of strains originally assigned to the genus *Actinomadura*. The new genus was named trying to differentiate a group of actinomycetes (*Actino*) that possess a different (*allos*) wall (*murus*), hence *Actinoallomurus* (Tamura et al., 2009). In fact, analysis of the cell wall components of the reclassified strains identified the presence of lysine in the peptidoglycan as a characteristic of *Actinoallomurus*. However, the structure of the peptidoglycan, including the position of lysine, has not yet been determined (Tamura et al., 2009). Fig. 1.8 contains a selected portion of a phylogenetic tree with some relevant genera appertaining to the order *Actinobacteria*. On the basis of the branching pattern, *Actinoallomurus* is in the same family as *Actinomadura* – *Thermomonosporaceae (Bacteria; Actinobacteria; Actinobacteria; Streptosporangiales; Thermomonosporaceae*). This proximity might indicate that, as the fellow taxonomical groups, it might also contain the genetic information for secondary metabolism as *Actinomadura* does, but because it is still relatively far from extensively studied genera, such as *Streptomyces*, it might provide a different set of bioactive compounds (Mariani et al., 2005; Venkateswarlu et al., 2005). *Actinoallomurus* strains are slightly acidophilic, non-acid-fast, strictly aerobic actinomycetes that show extensive branching and non-fragmenting substrate hyphae with spore chains developing on the tips of the aerial mycelium (Inahashi et al., 2015; Mazzetti et al., 2012; Pozzi et al., 2011; Pruesse et al., 2012).

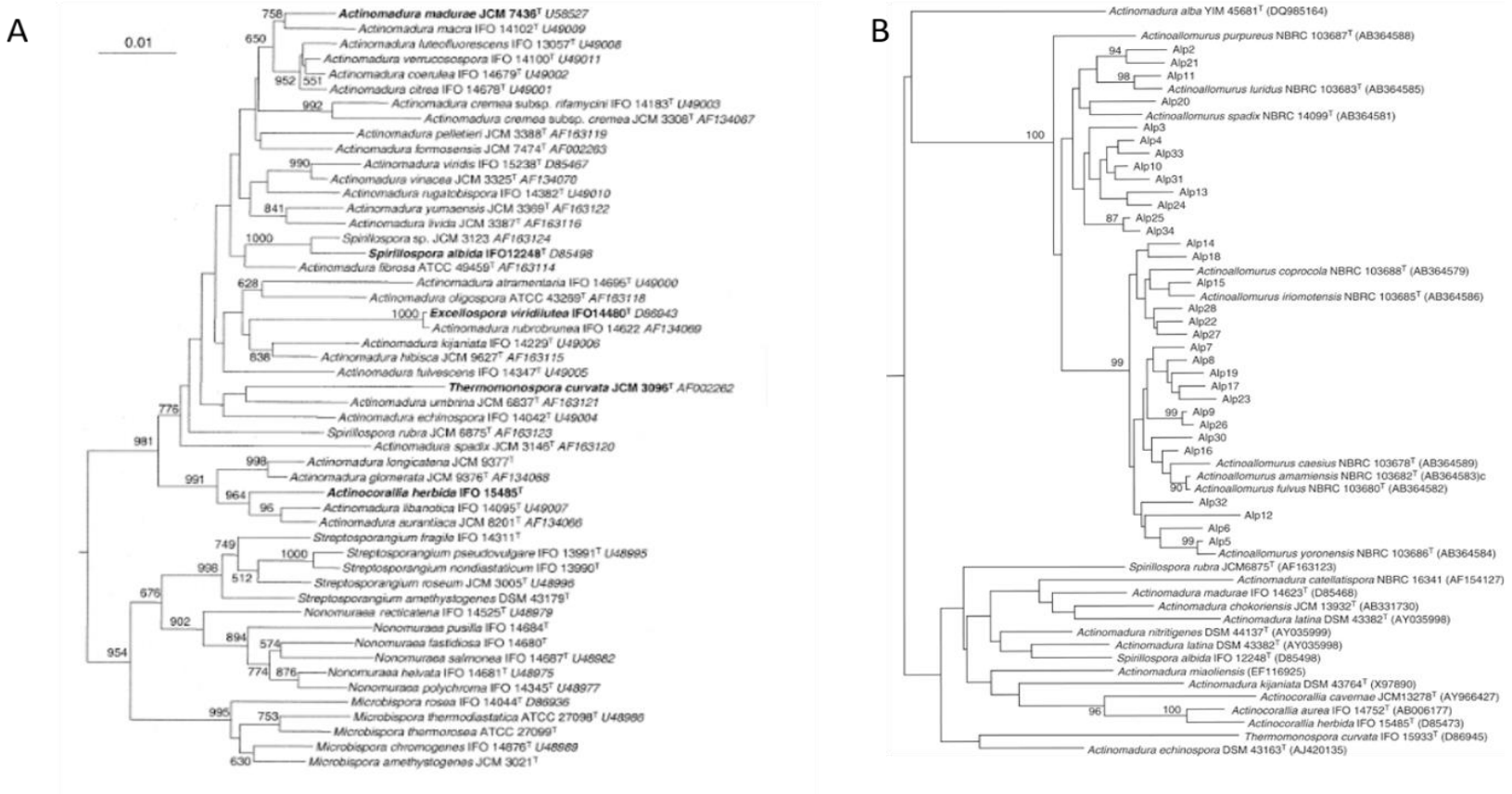


Figure 1.8 - Phylogenetic tree based on nearly complete sequences of the 16S rRNA gene. A – Phylogenetic analysis of the Thermomonosporaceae family. Adapted from (adapted from Wang, Y., et al., 2001)(Zhang et al., 2001) B - Phylogenetic analysis of the Actinoallomurus genus (Adapted from Pozzi, R. et al., 2011) (Hou et al., 2012). Note how *Actinoallomurus* closely relates with *Actinomadura* which in turn belongs to the *Thermomonosporaceae* family.

The *Actinoallomurus* strains analysed in this work derive from the NAICONS strain collection, which contains about 45,000 actinomycetes. The library was created with an emphasis on strain diversification. During the last years of library creation (in 2003-2005), a 16S-based isolation program was introduced, whereby morphologically different strains identified on isolation plates were systematically evaluated for phylogenetic novelty by 16S rRNA gene sequencing (Donadio et al., 2009; Monciardini et al., 2002). This endeavour rapidly yielded at five different groups, three of which were then formally described as *Catenulispora*, *Actinospica* and *Rugosimonospora*. Strains belonging to the groups originally dubbed "Alpha", and later established to coincide with *Actinoallomurus*, turned out to be relatively easy to isolate and, within one year, a collection of 1200 *Actinoallomurus* strains was isolated. Before the beginning of this study, an analysis of a small number of *Actinoallomurus* strains showed that at least 34 different phylotypes (Pozzi et al., 2011) were present within this taxonomic group, suggesting a substantial diversity in this genus (Busti et al., 2006; Donadio et al., 2010). Analysing the genus of *Actinoallomurus* thus brings one of the required elements of novelty previously mentioned for a successful screening program.

1.4. Hypothesis

As discussed in the Introduction, only new classes of antibiotics, different from those in clinical use, can ameliorate the problem of antibiotic resistance. Because antibiotic development is a lengthy and costly process, it is essential that intellectual property rights are secured once an interesting compound is identified, which requires structural novelty. The underlying hypothesis of this work is that, by systematically screening a previously unexplored genus, new chemistry could be found. In this work, 200 independent isolated strains belonging to the genus *Actinoallomurus* will be cultivated, and their extracts screened for antimicrobial activity in order to understand at which frequency antibiotics are produced. Will the frequency be comparable to that of other known and well exploited actinomycete producers, such as *Streptomyces*, *Amycolatopsis*, *Actinoplanes* and the like, or will production of antibiotics be a rare event for *Actinoallomurus*? If *Actinoallomurus* does produce antibiotics at high frequency, are some classes of molecules produced more frequently than others? Are there metabolites that are encountered in different *Actinoallomurus* strains? How frequently are new molecules encountered? Answering these questions will help to understand the potential of *Actinoallomurus* as a source of new drug leads as well as basic aspects of its physiology under laboratory conditions.

On the other side, the genome of *Actinoallomurus* will also be analysed to have insights into the number and types of BGCs it contains, and hence the type of chemistry that can be predicted. This analysis will thus complement the bioactivity-based screening and provide a complementary view of this genus.

Therefore, the goal of this work can be divided in to two main objectives:

- to verify the potential of a new taxon, specifically the genus *Actinoallomurus*, to deliver new chemical structures with antimicrobial activity and potential to become antibacterial drug candidates,
- Increase the knowledge on the poorly characterized genus of *Actinoallomurus*

2. SCREENING *ACTINOALLOMURUS*

2.1. Introduction

In this work it was decided to analyse the recently described genus *Actinoallomurus (Bacteria; Actinobacteria; Actinobacteria; Streptosporangiales; Thermomonosporaceae*) which has showed to be a potential producer of bioactive compounds (Inahashi et al., 2015; Mazzetti et al., 2012; Pozzi et al., 2011; Pruesse et al., 2012). The genus *Actinoallomurus* fulfils some of the requirements to be used in a productive antibiotic discovery effort: it is a recently described genus with few reports in the literature, making it unlikely that this genus has been systematically screened in the past; it is related to known producers, being in the same family as *Thermomonospora* and *Actinomadura*, which makes it probable that strains belonging to this genus might contain several BGCs for different metabolites; and, last but not least, the NAICONS' strain collection includes about 1,200 strains belonging to this genus (Pozzi et al., 2011) a sensible number for a screening endeavour. Other indirect but not exclusive evidence is its slow growth pace and the fact it presents a complex life cycle just like other known producers.

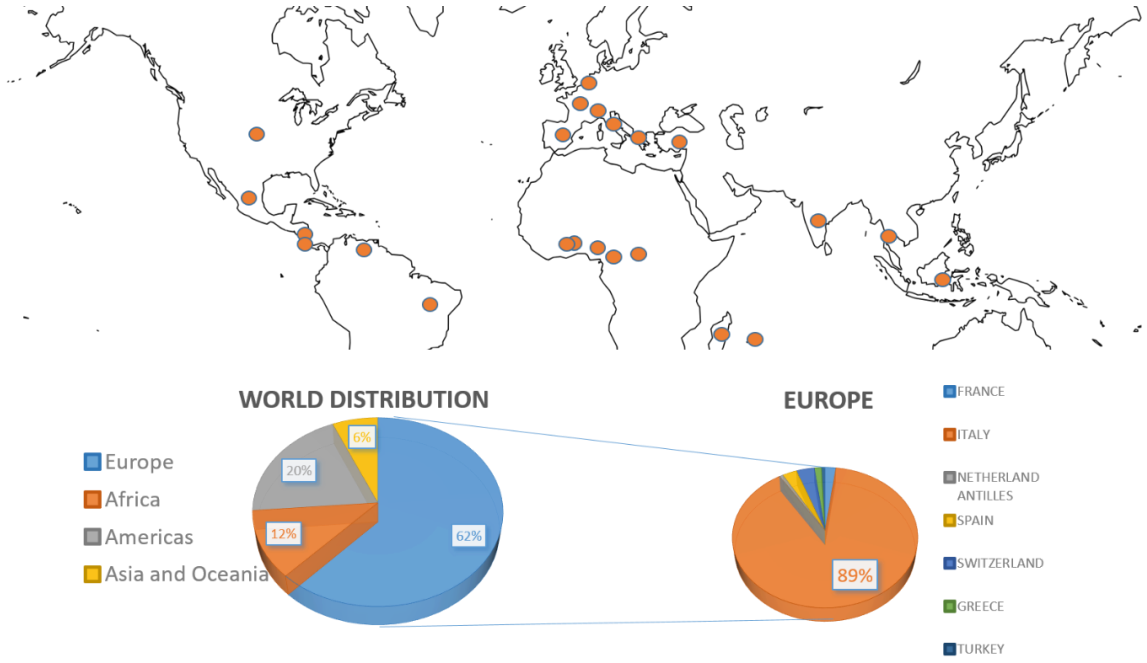


Figure 2. 1 - Geographic origin and distribution of the soil sample where the strains used in this work were isolated from.

For my PhD work, 200 *Actinoallomurus* strains were randomly selected. Although this number is not large enough to draw conclusions, it is interesting to present the origins of the strains. The soil samples from which these strains were isolated came from different geographic locations as described in Fig. 2.1, where it can be seen that there is a high percentage of strains coming from European soil samples, especially Italy. Nonetheless, a significant number of strains were isolated from other continents, making the 200 strains set globally distributed (Pozzi et al., 2011). Another analysed parameter is the type of environment from which these strains were isolated. Table 2.1 shows that the set under investigation includes strains isolated from a broad variety of environments, including densely vegetated areas (forest, semi-forest or woods), but also some harsh locations such as sulphur enriched craters of volcanic origin. It is also noteworthy the large amount of strains isolated from soils collected within the area that spawns the rhizosphere – one of the habitats from where the first strains were reported (Tamura et al., 2009). The data seems to indicate that *Actinoallomurus* is indeed a widespread genus able to colonize a broad range of environments. It would be interesting to confirm, in a future work, the relative abundance of this genus throughout the world.

Table 2.1–List and relative abundance of the type of environment from which the Actinoallomurus strains screened in this work were isolated from

High Altitude	*3%*
Deep Soil	*<1%*
Semi forest	*8%*
Woods	*22%*
Agricultural Land	*6%*
Forest	*10%*
Not Described	*15%*
Lawn	*2%*
Rhizosphere	*18%*
Inhospitable Terrain	*3%*
Unspecified Soil	*9%*
Freshwater	*4%*

2.2. Experimental Section

Actinomycete strains and media. *Actinoallomurus* strains, which were isolated from soil samples across the world, are from the NAICONS strain library. Each strain was propagated on S1-5.5 plates (60 g/L oatmeal, 18 g/L agar, 1mL/L Trace Elements Solution) at 30 °C for two to three weeks. From these plates, the grown mycelium was used to inoculate AF-A medium (10 g/L dextrose monohydrate, 4 g/L soybean meal, 1 g/L yeast extract, 0.5 g/L NaCl, 1.5 g/L 2-(N-morpholino)ethanesulfonic acid, pH adjusted to 5.6) in shake-flasks. After 8 days in a rotatory shaker (200 rpm) at 30 °C, cultures were harvested.

Antibacterial assays. Agar diffusion assays were prepared by inoculating $5x10^5$ CFU/mL in 15 mL of one of the two strains used in this work in Mueller-Hinton Agar media. After spotting 20 µl of the solution to test, plates were inoculated at 37ºC for 18-20 h. Strains used in this assay were *Staphylococcus aureus* ATCC 6538P and *Escherichia coli* L4242, an in-house ΔtolC mutant strain. All Minimal Inhibition Concentration (MICs) were determined by broth micro dilution in sterile 96-well microtiter plates according to CLSI guidelines, using Mueller Hinton broth (Difco Laboratories) containing 20 mg/L $CaCl_2$ and 10 mg/L $MgCl_2$ for all strains except for *Streptococcus* spp., which were grown in Todd Hewitt broth. Strains were inoculated at $5x10^5$ CFU/mL and incubated at 37 °C for 20−24 h.

Preparation of extracts. Three extracts were made throughout this work. **Mycelium Extract(MycExt)**: A culture was centrifuged at 16,000 rcf for 5 min. The pellet was resuspended in 0.4 volume of ethanol and incubated at 55ºC for 1 h. The suspension was centrifuged once more (16,000rcf, 5 min) and the clear liquid transferred to a new tube and, unless stated otherwise, resuspended in 10% DMSO at 5x the original culture concentration by adding firstly the DMSO and only after the required volume of water. **Clear Broth Extract**: A culture was centrifuged at 16,000 rcf for 5 min. The pellet was discarded and 0.5x volumes of ethyl acetate was added to the cleared broth, mixing and then waiting until complete separation of the liquid phases. The organic phase was transferred to a new tube and further 0.5 volumes of ethyl acetate was added to the exhausted broth, which was extracted as before. «e two organic phases were combined, dried and resuspended at 5x the original concentration in 10% DMSO adding it before the water. Creating two Extracts the organic phase, the Clear Broth Extract_EtoAc (**CBE_EtOAc**) and the aqueous phase, the Clear Broth Extract_Exhausted (**CBE_Exh**).

Biomass analysis. Biomass accumulation was measured as percent packed mycelial volume (PMV%) after centrifuging cultures for 10 min at 3000 rpm in a graduated tube.

Analytical procedures. For monitoring metabolites production analytical HPLC was performed on Shimadzu Series 10 spectrophotometer (Kyoto, Japan), equipped with a reverse-phase column, Lichrospher RP-18, 5µm, 4.6 x 125 mm (Merck, Darmstadt, Germany), phase A as 0.1% trifluoroacetic acid (TFA)in water, phase B as acetonitrile, and the flow rate was 1 mL/min. The program consisted of a linear gradient from 10% to 36% phase B in 5 min; from 36% to 50% phase B in 7 min; and from 50% to 80% phase B in 1 min; followed by a 4-min isocratic step at 80% phase B and column re-equilibration. UV detection was at 230 and 270 nm. LC-MS analysis were performed with an Agilent 1100 series liquid chromatographer equipped with a Ascentis express Supelco RP18, 2.7 µm, 50 x 4.6 mm column eluted at 1 mL/min at 40 °C. Elution was with a 10 min multistep program that consisted of 5%, 95%, 100%, and 5% phase B at 0, 6, 7, and 7.2 min, respectively (phases A and B were 0.05% TFA [v/v] in H_2O and CH_3CN, respectively) with detection at 220 nm. The effluent from the column was split 1:1 into a photodiode array detector and into the ESI interface of a Bruker Esquire 3000 plus ion trap mass spectrometer. Mass spectrometry analyses were performed using as sample inlet conditions: sheath gas (N_2) 50 psi, dry gas 10 L/min, capillary heater 365 °C; as sample inlet voltage settings: positive polarity, capillary voltage –4000 V, end plate offset –500 V; as scan conditions: maximum ion time 200 ms, ion time 5 ms, full micro scan 3; as segment: duration 10 min, scan events positive (100–2400 m/z). ^{1}H-NMR spectra were measured in CD_3CN-D_2O at 298K using an AMX 400 MHz spectrometer. Chemical shifts are reported relative to D_2O (δ 4.79 ppm).

Strains and Media for *C. elegans* work. The strain of *Caenorhabditis elegans* used throughout this work was the AU37 strain [glp-4(bn2) I; sek-1(km4)] which has a temperature sensitive defect in germ-line proliferation during larval development. Defect can be reversed by shifting worms from restrictive (25ºC) to permissive temperature (16ºC), plus this strain has also a knockout at the sek-1 gene which is encodes a mitogen-activated protein (MAP) kinase. This deletion disrupts part of the innate immune response of the strain which renders it more sensitive for pathogen infection(Kim et al., 2002). The *E. coli* strain OP50 was used a normal food source for the strains while the strains *S. aureus* ATCC 8325-4 and L100 (Naicons Collection) were used to perform pathogenic infections. The *C. elegans* was maintained on Nematode Growth Medium agar in Petri's plates (Before Autoclaving: NaCl 3g/L; Agar 17g/L;

Peptone 2.5 g/L. After Autoclaving add: 1ml of 1 M KPO_4 buffer pH 6; 1ml of 1M $MgSO_4$; 1ml of 5 mg/mL of Cholesterol in ethanol). For liquid Cultures S medium was used (5,85g/L NaCl; 1 g/L of K_2HPO_4; 6 g/L KH_2PO_4; 1ml of 5 mg/mL of Cholesterol in ethanol; 10ml of 1 M Potassium Citrate solution - 20 g/L citric acid monohydrate and 293.5 g/L of Tripotassium Citrate monohydrate; 10 mL of Trace Metals Solution; 3 ml of 1M $CaCO_2$). The transfer and cleaning process of the nematodes was done using the M9 buffer (3 g/L K_2HPO_4; 6 g/L Na_2HPO_4; 5 g/L NaCl; 1ml of 1M $MgSO_4$). **Growth of *C. elegans*.** The NGM media was previously inoculated with a layer of *E. coli* OP50 and then the nematodes were place onto it. Worms were transferred to a new plate every three to four days, depending on the quantity of *E. coli* still present in the plate. **Transfer of *C. elegans*.** A plate of fully 2nd and 3rd generation worms would be washed with M9 buffer, creating a suspension of bacteria and worms that was left to deposit for one minute. The clear broth was then replaced for new M9 buffer and the suspension dispensed through new NGM plates. **Synchronizing *C. elegans***. After 3 to 4 days of incubation of C. elegans at 16ºC a plate of fully 2nd and 3rd generation worms with the presence of eggs should be visible at a light microscopy. After adding 2 mL of M9 buffer and wash the worms onto to a 1,5 mL Eppendorf, centrifuge at 1300 g for 1 minute. Discard the supernatant and add 550 µL of water, 200 µL of Household bleach, and 250 µL of 2M NaOH. Mix gently and vortex occasionally until the worms disappeared. Centrifuge at 1300 g for 1 minute, Discard the Supernatant and add 1 mL of M9 buffer, repeat the last two steps 3 times. On the last wash leave about 250 µL of M9 buffer and incubate the Eppendorf at 16ºC for 30 hours. Disperse the suspension in new NGM plates with *E. coli*, incubate at 25ºC for 48 hours. At this point the worms are sterile, all at the same growth stage and ready to be used.

2.3. The screening strategy

Since *Actinoallomurus* is an unexploited genus, the simplest way to evaluate its potential to produce new antibiotics would be to use simple, phenotypic assays that simultaneously can detect an activity against many different targets: i.e., growth inhibition of one or more test strains. To this end, I decided to use *Staphylococcus aureus*, a Gram-positive pathogen, and *E. coli*, a Gram-negative. However, since Gram-negative bacteria are usually insensitive to many antibiotics because of the low permeability of their outer membrane, I selected as test strain a mutant lacking TolC, the major proteins involved in the expulsion of a number of toxic compounds (Wandersman and Delepelaire, 1990). I decided to test for

antibacterial activity three different types of culture extracts: Mycelium Extract – MycExt – and two extracts resulting from ethyl acetate extraction of the clear broth - CBE_EtOAc and CBE_Exh (see Materials and Methods). There have been reports of using prefractionated extracts for screening assays. These are obtained by resolving whole culture extracts in different fractions, in order to afford a more predictable (because of less complex samples) biological profiling before focusing on isolating and characterizing the active compounds (Wagenaar, 2008). However, since the purpose of my project was to rapidly identify the chemical nature of the active compounds, I preferred not to invest significant resources on preparing a prefractionated library and tested instead the extracts mentioned above.

Once a positive extract is identified, the next step is to understand the chemical nature of the active molecule(s). This is achieved by resolving the extract in 30 different HPLC fractions, concentrating each fraction to a testable volume and identifying the fraction(s) associated with the original antibacterial activity. While the extract is fractionated, information is acquired on the UV spectra and MS peaks of each component of the fractionated extract. Then, the active fraction(s) can be associated with the presence of *m/z* ratios, MS fragmentation patterns and specific chromophores that help in database searches for enabling structural hypotheses on the type of compound(s) responsible for the activity. This process, designated as "dereplication", is an essential part of natural product screening and requires access to dedicated databases. Indeed, many large companies involved in microbial product screening created their own databases covering products described in the scientific and patent literature over several decades. One such database was available for my project, which reports information about 27,000 compounds of microbial origin, including – when known – the identity of the producer strain, molecular weight, empirical formula, UV maxima and biological activity of the compound. If the properties of an active compound identified in the screening do not match any of those described in the database, the compound is likely to be novel and worth of further investigation. In my project, a "hit" was designated as any compound endowed with antibacterial activity against at least one of the tested strains which was likely to be novel. According to this definition, extracts containing known compounds of whose activity was lost upon fractionation were not further investigated.

From a hit to the beginning of a lead optimization process lies a very personalized path of chemical and biological characterization to better understand the molecule and its activity. Each molecule has its own set of properties that needs to be addressed in a particular way. Discovering new molecules is a partnership between chemists and biologists that correlates biological activity and chemical proprieties, associated with preliminary structure activity relationship studies. This part has only been marginally addressed during my project (see Chapter 1.3.1). The whole screening process is described in Fig 2.2. Although the flow chart represents useful guidelines for the discovery process, an effective progress requires often some iteration of those steps and/or exploration of alternative routes.

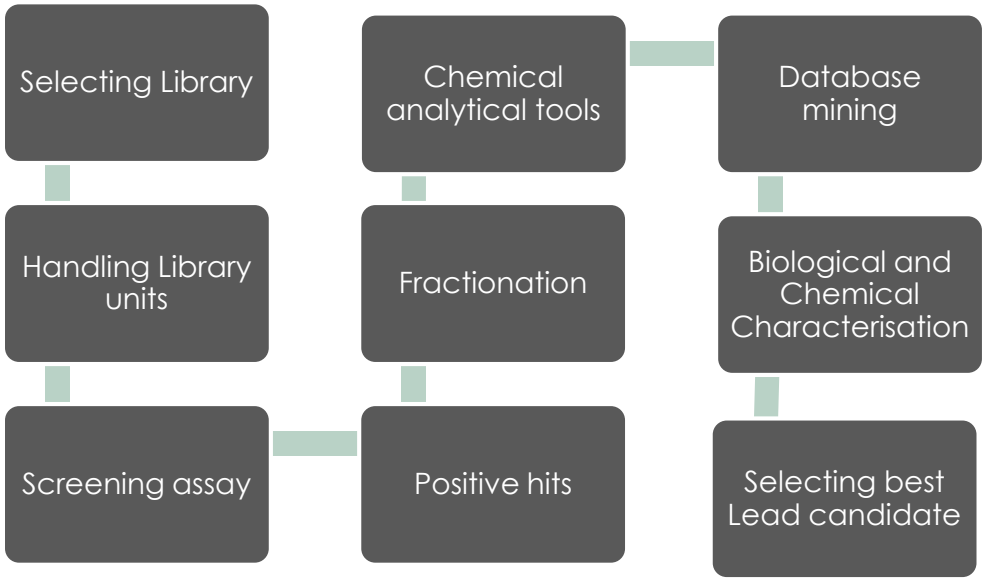


Figure 2.2 – Workflow of the High Quality Strategy adopted in this work.

2.4. Screening Overview

As mentioned earlier a pool of slightly over 200 strains was selected for antimicrobial screening. All strains were processed under the same conditions and yet, without any customization of the fermentation conditions, 104 out of 232 strains produced an extract that inhibited the microbial growth of *S. aureus* and, within those, 17 also prevented growth of the *ΔtolC E. coli* strain. As expected, the geographic origin of the active strains was scattered around the world in a manner that reflected their relative abundance (data not shown). Note that these 232 strains include 13 strains that were screened before the beginning of this project. While it is unrealistic to expect that 104 active strains will translate into 104 different and novel

active compounds, the high frequency of active strains supports part of the premise of this work, that taxonomical groups related to known producers have a high probability to produce active metabolites as well.

The other important point to address is the understanding of how many of these strains do present new structures and how many produce already documented metabolites. While screening 200 strains and characterizing some of the novel bioactive compounds they produced represented a realistic PhD project, the high frequency of positive extracts prevented analysis of every active metabolite from all active strains. However, to our understanding it was more important to characterise selected new compounds than to analyse superficially the metabolome of *Actinoallomurus*.

Table 2.2 list all the strains that upon activity testing inhibited antibacterial growth. In 22 cases the 16S rDNA gene was partially sequenced and a phylotype was given with nine of the previously 37 observed represented in this table. The third column list the Active Extract, as mentioned previously three different extracts were made and tested and the bioactivity was usually spread in more than one extract. However, for a matter of simplicity only the extract with the Higher inhibition halo was listed. On the forth column a qualitative evaluation of the intensity of the signal is given, which can be translated by the wideness of the inhibition halo just as explained in the footnotes of Table 2.2. Finally, a note on the type of metabolite identified from each strain is also given. When a compound was identified its name is provided (45 cases from which 12 were analysed previous to my arrival), when the identification of the compound was halted or not even started an acronym was given explaining the main reason for the interruption (See Chapter 2.4.1).

Table 2.2 – The set of *Actinoallomurus* strains screened at Ktedogen up to September, 2015. When available, information is also provided on phylotype, active extracts, signal intensity and identified metabolite

Strain ID	Phylotype[a]	Active Extract[b]	Signal intensity[c]	Compounds[d]
145113	Alp4	MycExt*	++	NAI-113[d1]
145114	Alp4			Pradimycin
145206	Alp35	CBE_EtOAc		NAI-113
145219				CycloPhe-Leu
145226	Alp24	CBE_EtOAc		Benanomycin
145250	Alp22	CBE_EtOAc		Coumermycin
145260	Alp22	CBE_EtOAc		Pyrrolosporin

145318				Pradimycin
145414	Alp6			NAI-414
145519	Alp22			Coumermycin
145527	Alp36	MycExt	++	LAF
145529	Alp37	CBE_EtOAc*	+	Allocyclinone
145530	Alp24	CBE_EtOAc	+	Allocyclinone
145531		CBE_EtOAc	+	LOW
145532		MycExt	+	LOW
145534	Alp24	CBE_EtOAc*	++	Allocyclinone
145535	Alp24	CBE_EtOAc*	+	Allocyclinone
145536		MycExt	+	NC
145537		MycExt	+	NC
145538	Alp24	MycExt*	++	Allocyclinone
145539		CBE_EtOAc	+	NC
145541	Alp24	MycExt	+	Allocyclinone
145542	Alp24	MycExt*	+	Allocyclinone
145543	Alp24	CBE_EtOAc*	+	Allocyclinone
145544	Alp24	MycExt*	+	Allocyclinone
145546	Alp24	CBE_EtOAc*	++	Allocyclinone
145547		MycExt	++	NC
145548		CBE_EtOAc	+	NC
145550		CBE_EtOAc	+	NC
145552		CBE_EtOAc	+	NC
145553		CBE_EtOAc	+	NC
145554		MycExt*	++	NAI-823
145555		CBE_EtOAc	+	LOW
145557		CBE_EtOAc	+	NC
145559		MycExt	++	NC
145610		CBE_EtOAc	+	NC
145613		MycExt	+	NC
145617		CBE_EtOAc	+	NC
145626		CBE_EtOAc	++	NC
145633		MycExt	+	New Macrolide [a1]
145635		CBE_EtOAc	+	NC
145638		MycExt	++	NC
145639		CBE_EtOAc	+++	NC
145640		MycExt	++	NAI-107
145641		CBE_EtOAc	++	NC
145644		MycExt	+	NC
145645		MycExt	+	NC
145646		MycExt	++	NYC
145649		CBE_EtOAc	+++	Polyether

145654		MycExt;	+	NC
145656		CBE_EtOAc	++	Polyether
145663		MycExt	+	NC
145665		CBE_EtOAc	+++	Polyether
145669		MycExt	+	LOW
145672		MycExt	+	NC
145673		MycExt	+	NC
145676		CBE_EtOAc	++	Polyether
145677		MycExt	++	LA
145689	Alp9	MycExt*	+++	Allocyclinone
145695		CBE_EtOAc	++	NC
145698	Alp9	CBE_EtOAc*	+++	Allocyclinone [d1]
145699	Alp7	MycExt	++	NAI-107 [d1]
145705		CBE_EtOAc*	+++	LAF
145707		MycExt	+	LA
145709		CBE_EtOAc	+	NC
145711		MycExt	++	New Halogenated quinone
145712		CBE_EtOAc*	++	LAF
145713		MycExt*	++	LA
145716		CBE_EtOAc	++	LA
145722		MycExt	+++	NC
145723		MycExt	+++	NC
145724		MycExt	+	NC
145726		MycExt	+	NC
145754		CBE_EtOAc*	++	NAI-113
145757		MycExt	+++	Polyether
145758		CBE_EtOAc	++	NC
145763		MycExt	++	NC
145770		MycExt		NAI-770
145772		CBE_EtOAc	+	NC
145778		MycExt	+++	Polyether
145779		MycExt	+++	Polyether
145802		MycExt	+++	NAI-823
145805		MycExt	+	LOW
145806		MycExt	+	LOW
145807		MycExt	++	NC
145808		CBE_EtOAc	++	WIP
145809		MycExt	++	LAF
145811		MycExt	+++	NAI-823
145814				Pyrrolosporin
145816				NC

145817			Octacyclomycin
145823	MycExt	+++	NAI-823[d1]
145824	CBE_EtOAc	++	NC
145828	CBE_EtOAc	++	LOW
145829	MycExt	++	NC
145830	MycExt	+	LOW
145831	MycExt	+++	Polyether
145832	MycExt	+++	NC
145833	MycExt	++	NC
145835	CBE_EtOAc	++	Coumermycin
145837	MycExt	+	LOW
145838	CBE_Exh	+++	NC
145841	MycExt*	+++	NAI-823
145846	MycExt	+++	NC

a - Phylotypes were assigned using the method described in Pozzi, R., et al. 2009

b - Only the most active extract is listed inthis table, which was the one ued for HPLC fractionations. Strains marked with * are active against *S. aureus* and *E.coli* ΔtolC otherwise activity was observed only against *S. aureus*

c - + - Growth inhibition halo < 15mm; ++ - Growth inhibition halo > 15mm; +++ - Growth inhibition halo > 25mm

d - LOW – Low Activity LA – Lost Activity LAF – Lost Activity upon Fractionation NC – Not yet Characterized; Metabolites were characterized by MS, UV; di – also confirmed with NMR analysis

Strains in between dotted lines were characterized before the beginning of this project.

As mentioned before, it would not be possible, within the time frame of this project, to ascertain every single active compound produced by 104 active strains. Actually, active compounds are often produced as a family of related molecules, the so-called complex, and characterization of the major congeners present in a complex further complicates this task. For example, identifying the active compounds from strain ID145698 involved structural elucidation of the four major components of the allocyclinone complex (see Chapter 4). Table 2.2 lists a number of cases where, although structural elucidation has not been completed, dereplication has failed to identify known compounds with a similar molecular weight, MS/MS fragmentation, UV maxima and activity spectrum, which provides a high degree of confidence about the novelty of the corresponding molecule(s). To increase the throughput, active extracts were prioritized on the basis of their potency (i.e. inhibition halo), antibacterial spectrum and ease of purification of the active compound(s). Extracts that did not reconfirm antibacterial activity or presented a profile typical of known classes (e.g. polyethers) were not further investigated. When a metabolite matched known molecules by at least two parameters, no NMR confirmation was sought. Although we are aware of the possibility that, by doing so, we might have ignored a few new compounds, we decided to focus on those molecules that had a high probability of being novel and structurally interesting. Some examples of *Actinoallomurus*

metabolites with a high probability of having the identities assigned in Table 2.2 are described herein.

2.4.1. The universe of the unknown

Within the universe of the 50 strains with uncharacterized metabolites there are different categories. An extract may have been deprioritized for different reasons: the activity was not confirmed on subsequent fermentations (LA in Table 2.2); no active fraction was observed after fractionation (LAF in Table 2.2); or the observed inhibition halo was smaller than 15 mm, an empirical threshold set after observing that extracts with lower inhibition halos were unlikely to yield activity after fractionation (e.g., activity might get dispersed in several fractions), and thus set to low priority (LOW in Table 2.2). Those extracts that did not fall into the LA, LAF or LOW categories are marked as not yet characterized (NC in Table 2.2), and they actually represent the majority of them. From the strains with NC extracts, ID145633 and ID145808 stand out as producers of two unrelated new macrolides whose characterization could not be completed before the end of this project.

2.4.2. Polyethers from *Actinoallomurus*

Throughout the work several strains presented large inhibition halos against *S. aureus*. As explained below, these compounds were dereplicated as polyether antibiotics. Polyether antibiotics are produced by PKS-I and show a broad set of bioactivities in addition to antimicrobial (e.g. coccidiostatic, anticancer, etc.) The antimicrobial activity of polyethers is strictly connected to their ability to insert into cellular membranes and alter transport of metal cations, which leads to changes in the osmotic pressure inside the cytoplasm and cell death. (Rutkowski and Brzezinski, 2013) Although there are over 120 reported polyethers in the literature, this mechanism of action has been studied on a limited number of molecules and little is known on how to make polyethers more selective towards a particular cell type. This diminishes the interest in pursuing polyethers as understanding their value requires complex testing. In fact, to our knowledge no polyether has been shown to have in vivo efficacy for treating bacterial infections. (Kevin Ii et al., 2009) These limitations led us to abandon further analysis of any molecule likely to be a polyether.

Table 2.3 - List of polyether producing strains, tentative mass value along with confirmation of activity by fractionation

	Tentative m/z^{+} [a]	Fractionation[b]
145554	953	Yes
145649	918	No
145656	936	No
145665	923	No
145676	904	No
145757	909	Yes
145770	953	No
145778	809	No
145779	909	No
145802	914	No
145811	909	No
145823[B]	914	Yes
145831	884	Yes
145841	914	Yes

a – PUTATIVE MASS VALUE WITH ON POSITIVE CHARGE

b – CONFIRMED BY NMR

Work performed previous to the beginning of this project led to the characterization of the new polyethers NAI-823 and NAI-770 (work by Dr. Marianna Iorio - data not published). This experience was instrumental in devising a simple way to tentatively recognizing polyethers on the basis of their chemical and biological profiles: large inhibition halos of *S. aureus* with no or little activity against *E. coli ΔtolC*; no or limited UV absorption; high retention times on reverse phase HPLC; and, diagnostically, formation of Na adducts upon MS analysis with a characteristic loss of 14 (from a leaving methyl group) and 62 amu (loss of a carboxylic group). Active compounds presenting all these characteristics were categorized as polyethers and their analysis interrupted. Table 2.3 presents the strains producing compounds that had these characteristics and the m/z^{+} value of the most abundant congener of each strain.

As an example, a large growth inhibition halo against *S. aureus* was identified from mycelium and ethyl acetate extracts of strain ID145757. The mycelium extract – the most active extract - was fractionated and each fraction tested for antibacterial activity. Growth inhibition was observed for fractions collected at high retention times, eluting in the presence of at least 60% of organic phase. The active fractions presented no UV absorption and m/z^{+} values between 909 and 938 (See Fig. 2.3). All these tentative masses presented the similar fragmentation patterns with characteristic losses of 14, 23, and 62.

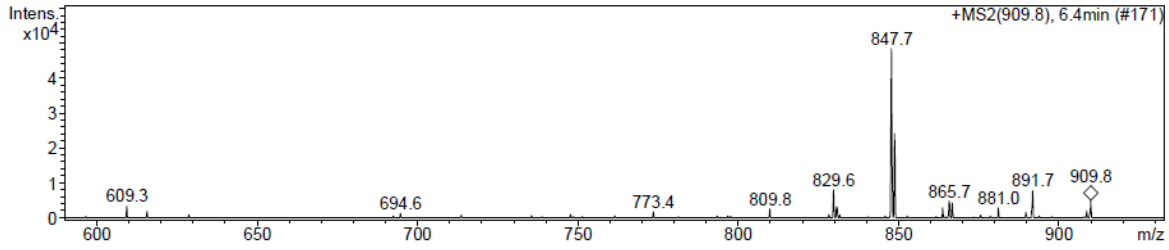


Figure 2.3 – LC-MS of the mycelium extract of strain ID145757. Mass fragmentation pattern of the m/z signal at 909 (M+H)+ corresponding to polyether**.**

The strains listed in Table 2.3 could potentially be producing a new entry in the polyether family (the *m/z* values did not match any of those present in the literature), although, without a confirmation through NMR, we cannot be sure be sure those molecules are actually polyethers and how different they are from known molecules. Nonetheless, those strains – and perhaps *Actinoallomurus* in general – may represent a valuable resource to explore new chemical space within the polyethers.

2.4.3. The case of coumermycin

The family of coumarin antibiotics is a successful member from the realm of natural products that actually made into the market as novobiocin, which was used to treat bacterial infections due to Gram-positive bacteria. It acts by targeting bacterial DNA gyrase (Montecalvo et al., 1995). Coumermycin is another member of the coumarin family that presents higher antibacterial activity than novobiocin.

During the screening exercise we came across with strain ID 145835, whose ethyl acetate extract presented an inhibition halo of around 20 mm against *S. aureus*. The corresponding active fractions showed an MS signal of 1108 m/z^+, associated with maxima at 220, 270, and 340 nm. The MS/MS pattern of this compound matched that of pure coumermycin A1 that, during negative electrospray ionisation, fragments at the glycosidic, amide and ether bonds, producing ions at m/z^- 1108, 1001, 620, 487 (Weinstein and Wagman, 1978; Wolpert et al., 2008). Fig. 2.4 summarizes these data.

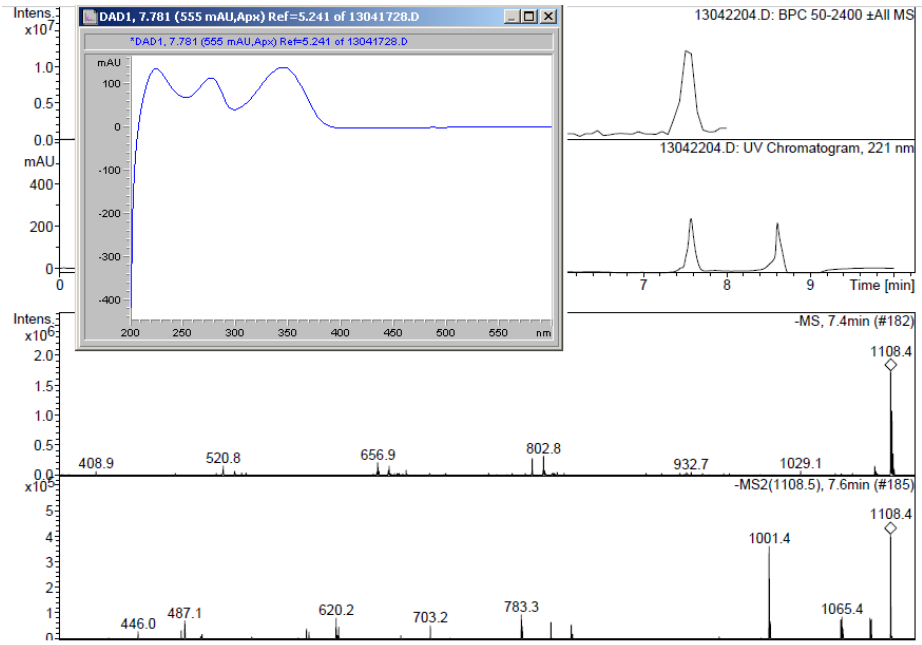


Figure 2.4 – UV and MS analysis of Mycelium extract from strain ID 145835 with chemical features equal to those of Coumermycin. See text for details.

Also in this case, previous experience helped. Indeed, previous to this work, another scientist from the team (Dr. Matteo Simone) had previously characterized coumermycin from other *Actinoallomurus* strains, confirming its identity by ^{1}H NMR and ^{13}C NMR analyses, (Veneziani, 2009-2010) while production of coumermycin A2 had been reported previously (Pozzi et al., 2011). Consequently, strain ID145835 was identified as a putative producer of coumermycin A1. The *Actinoallomurus* strains found to be producing coumermycin-related molecules are described in Table 3.4.

Table 2.4 –Summary of the analysis done on the strains producing coumermycin

STRAIN ID	Tentative m/z^{+}	Fractionation	NMR	Compound
145250	1083	Yes	Yes	Coumermycin A2
145519	1109	Yes	Yes	Coumermycin A1
145835	1109	Yes	No	Coumermycin A1

2.5. Another Screening Tool

2.5.1. Rationale

After realizing the enormous potential of *Actinoallomurus* for the production of bioactive compounds it was clear that it would not be possible to ascertain the bioactive compounds for all the positive extracts. Therefore, I decided to explore other techniques that would allow filtering the positive hits and focus only on those that showed a potential for further development. This requires extracts to be evaluated in a more complex environment than a Petri dish. Hence, the choice was on exploring a miniaturized infection model to establish potential efficacy of extracts. Several model organisms are available to develop these kinds of assays, and I decided to explore *Caenorhabditis elegans* as the host for the infection model. Since this expertise was not available at Ktedogen laboratories, I decided to resource to more experienced facilities such as Professor Anders Løbner-Olesen's research group at the University of Copenhagen, Department of Biology, another beneficiary in the TRAIN-ASAP project. To support my initiative in Copenhagen I successfully applied for support by FEMS, through a FEMS Research Grants application and spent two months [March-May 2014] in Copenhagen. One of the requirements of a granted fellowship is the delivery of a report describing the results achieved throughout that period of time. This report includes all the main achievements of that period and was added at this point to show the work performed in Copenhagen.

By the end of the first half of 20th century, antibiotics held a great premise regarding the eradication of infections as major health threat and by the mid-1980s infectious diseases were considered virtually conquered as antibiotics kept being introduced into clinical practice (Coates et al., 2011; Davies and Davies, 2010; Powers, 2004). Antimicrobial agents are one of the measures that have done more to improve public health in the last 50 years. However, in this very success lies also the major problem, the overuse of antibiotics has created a formidable selective pressure. We are now witnessing the increase of drug-recalcitrant bacteria, proving that, in these past few years, a war against infectious agents is far from over and showing that investment in this area should not be disregarded (Wise, 2002).

Despite substantial effort, the combination of genomics, high throughput screening and chemical libraries have not lived up to its promise: although molecules with high target affinity in vitro were found, they failed to deliver the same results in more complex environments, such as in vivo testing or even whole cell experiments. In contrast, microbial products have been extremely successful in providing existing antibiotics. In this case however, effective screening strategies must be implemented to reduce the rediscovery of known compounds while increasing the probability of finding new molecules with efficacy against bacterial cells but lacking undesired effects on eukaryotic systems (Donadio et al., 2010; Monciardini et al., 2014).

One possible source to enhance the antibacterial arsenal is to look on what Nature has been using to solve the very same problem. My PhD project is part of a discovery program designed to assess the metabolic potential of the recently described genus *Actinoallomurus*. The schedule of work includes small-scale fermentations (from 200 strains) followed by bioactivity testing. Positive samples are then subjected to a combination of a bioassay with chemical analysis to determine the novelty of the active compounds. However, the screening strategy plays an important role in selecting the most promising antimicrobial candidates. Whilst traditional *in vitro* and whole cell drug screens are acknowledged as the established paradigm for identifying antimicrobial molecules, the whole animal approach allows for early and direct assessment of *in vivo* drug efficacy, thus, eliminating compounds that are toxic to the host or with poor pharmacokinetic properties. It is for these reasons that complementing work done *in vitro* with an *in vivo* analysis could lead to a set of new antimicrobial candidates that do have the potential to actually succeed as new antibacterials.

In order to introduce a screening tool at an early stage of the discovery process, it is important to maintain a relatively high throughput and low costs. Hence, one would need the simplest differentiated organism, small enough to be handled in large numbers, easy to cultivate, with a short life cycle and with a relevant amount of biological processes conserved with other higher organisms (Segalat, 2007). *Caenorhabditis elegans* is a promising candidate fitting all the criteria mentioned above, the fact that a remarkably large number of human pathogens, including the Gram-negative *Pseudomonas aeruginosa* and the Gram-positive *Staphylococcus aureus*, can infect and kill this nematode, makes it a versatile platform for drug discovery (Tampakakis et al., 2008).

C. elegans model as a tool for drug discovery has been developed within the last decade with the first screen for antibacterial agents active against *Enterococcus faecalis* being reported in 2006 (Moy et al., 2009) and a subsequent screen that expanded the model also to Gram negatives (Durai et al., 2013). All those developments are transforming this nematode in a widely recognized assay amenable tool.

2.5.2. Evaluation of the tool

The main goal of this project was to learn the techniques for a further development of an assay to pinpoint the microbial extracts containing a compound able to rescue *C. elegans* previously infected with *S. aureus*. Hosted in the University of Copenhagen, Department of Biology, I had the opportunity to work with Professor Anders Løbner-Olesen's research group. Although mainly focused in cell cycle control of DNA replication in bacteria, they also perform small molecule screening *in vitro* and *in vivo*, using *C. elegans* as an animal model. Their last work on screening and analysing the effect of harmane in *C. elegans* reported the increase of life span of these nematodes when treated with this alkaloid (Jakobsen et al., 2013). The assays were all made in solid media using a pure compound instead of a mix of natural products such as the ones I work with at Ktedogen.

Starting from the screening assay already established, two objectives were set for this assay improvement. Firstly, to transfer the assay from the solid media platform into a liquid one. Hence it would be important to verify the best infection conditions as well as the effect of solvents in the nematode's life span. Secondly, to assess the feasibility of using complex mixtures, such as microbial extracts, as natural product libraries. Mainly by verifying if the microbial extracts could in fact rescue infected nematodes.

The Liquid media assay. In order to synchronize AU37 worms, eggs from gravid adults were isolated and then hatched overnight in M9 buffer, L1-stage worms were plated onto lawns of *E. coli* on nematode growth medium agar media. Worms were grown to sterile, young adults by incubation at 25°C for 48h, washed off the plates with M9 buffer and transferred to 96 well plate. The first part of this process was to verify that microbial broth extract prepared would not disturb the nematodes' viability.

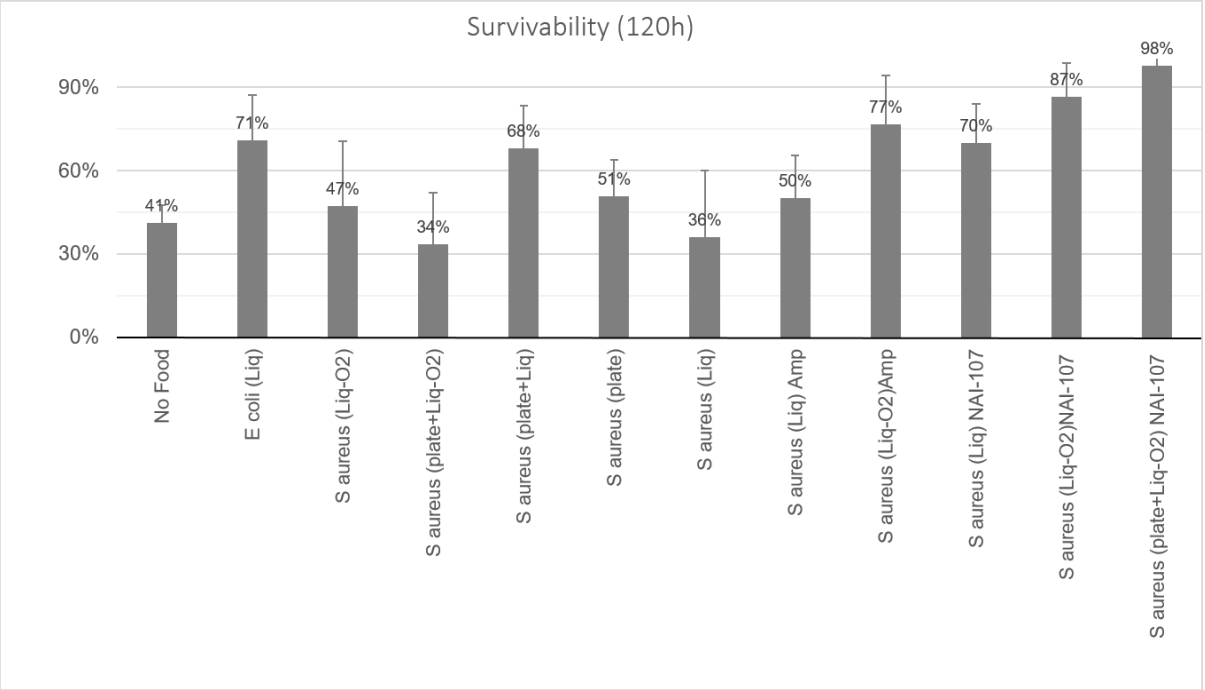


Figure 2.5 - % of nematodes' survival with different methods of infection. liq-02 – nematodes fed with an anaerobic overnight culture; plate-liq-o2 – nematodes infected in a s. aureus plate and then incubated in an anaerobic o/n culture; liq – nematodes fed with an aerobic overnight culture; plate-liq – nematodes infected in a s. aureus plate and then incubated in an aerobic o/n culture: ampicillin and nai-107 were added as positive controls at 4 µg/l.

To this point, 1% DMSO was confirmed as the best solvent to use to allow complete solubilisation of the complex mixture without interfering with the survivability of the nematodes. The infection with *S. aureus* was performed by feeding *S. aureus* on a solid media and then transferred into the liquid stage, or by directly feeding the nematodes with the pathogen on the liquid media. On this case, the liquid culture of *S. aureus* was prepared by fermenting an overnight culture aerobically and anaerobically. The presence/absence of oxygen could be a triggering factor for virulence factors and enhance the pathogenicity of *S. aureus*.

Figure 2.5 shows the most effective method, achieved by feeding the nematodes with *S. aureus* on plate first and then with an overnight culture of the same bacteria grown anaerobically. However, the difference between the procedure just mentioned and direct incubation with *S. aureus* from an aerobic overnight culture is not significant which led me to select the latter as the method of infection due to practical reasons. From this moment on, synchronized nematodes were infected with *S. aureus* by feeding, the previously prepared overnight culture (in aerobic conditions), directly in the liquid media. Finally, it was also

observed that the survival of the nematodes was greatly influenced by the shape of the well were the assay was performed. Graph on Fig. 2.7 shows how nematodes do not survive more than 72 hours in U-shaped wells while they were able to live for over 120 hours in flat bottom wells.

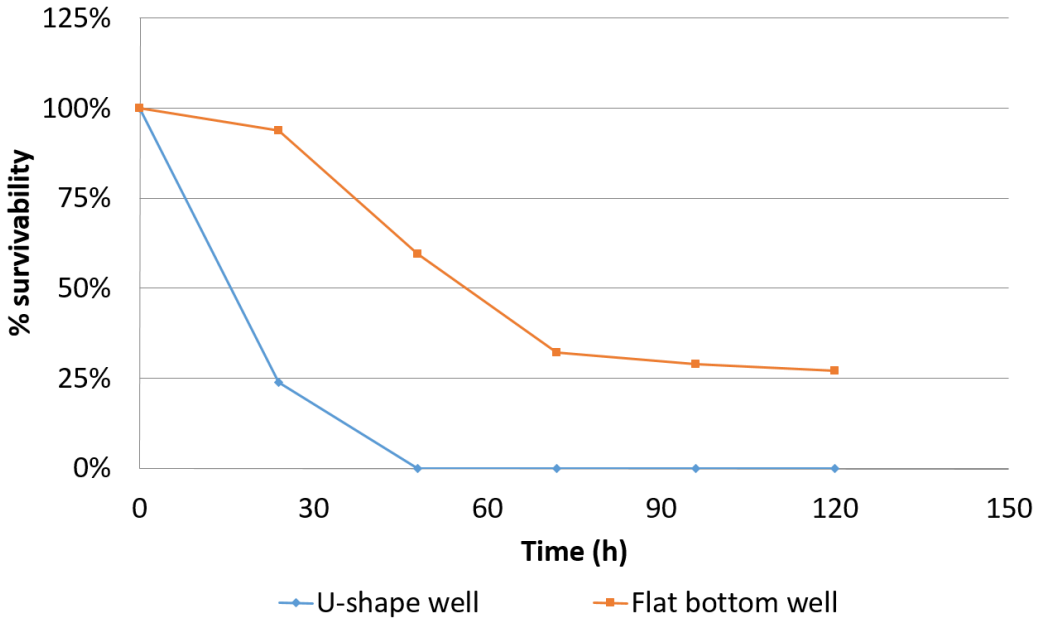


Figure 2.6 - % nematodes' survival with two different types of 96-well plates iwhen inoculated with the *S. aureus* strain

A possible explanation to the difference in surviving is the accumulation of bacteria and worm on the deepest point of U-shaped well and consequent death by asphyxiation. Although it was not ascertained the main reason for the nematode's death it is also worth mentioning that the assay while using U-shaped 96 wells plates has a higher reproducibility rate. At this point, it was decided to use U-shaped 96-well plates to assess the microbial extracts.

These extracts were prepared from 10-days cultures of 10 *Actinoallomurus* strains: from each strain an extract was obtained from the mycelium and two from the supernatant. All together 30 extracts were prepared and tested. These were previously tested for *in vitro* activity against *S. aureus* and selected in such way that I could test in this assay active and inactive samples from all sources. In addition, some of these metabolites are still unknown but others were already dereplicated. The goal here was to evaluate the feasibility of this assay more than the identification of the active compounds. Therefore, the focus was given more on pinpointing

extracts that promoted the survivability of the nematodes. The plot from Fig. 2.7 denotes the surviving after 72 hours of incubation: as shown Extract 1, 5 and 26 were able to promote worm survivability throughout the experiment (over 40% survival).

As mentioned before, because the *S. aureus* infection is not a persistent one, it is not possible to know if the natural extracts are killing the bacteria inside the nematodes or only outside and *C. elegans* has cleansed the infection by eating dead bacteria and expelling the living into the bactericidal liquid media. However, the fact that nematodes are surviving in the presence of these active compounds, present in the previously mentioned extracts, is an indication that these molecules are not toxic to a higher organism such as the nematodes. Although it will be necessary to reconfirm the results achieved it is also worth mentioning that some of the extracts marked as non-active on the *in vitro* assay, promoted *C. elegans* survival (Extract 2 and 16). This kind of data reflects other characteristic of this assay, molecules that do not directly inhibit bacterial growth but help the host's immune response may also be amenable to this screening exercise. The other interesting point is the fact the nematodes died so quickly in the presence of the Extract 3 and 27 (100% mortality after less than 12 hours), which indicates that those might contain nematocidal compounds. Although it was not the purpose of this experiment to screen for such molecules the nature of the assay allows also to identify the extracts that might have such activity. Although the design of the assay has to be completed and improved, the main goal of the project was achieved and it was possible to transfer the technology to the laboratory where I am based. In the following months I will set up and validate the screening assay to apply it to our natural product libraries.

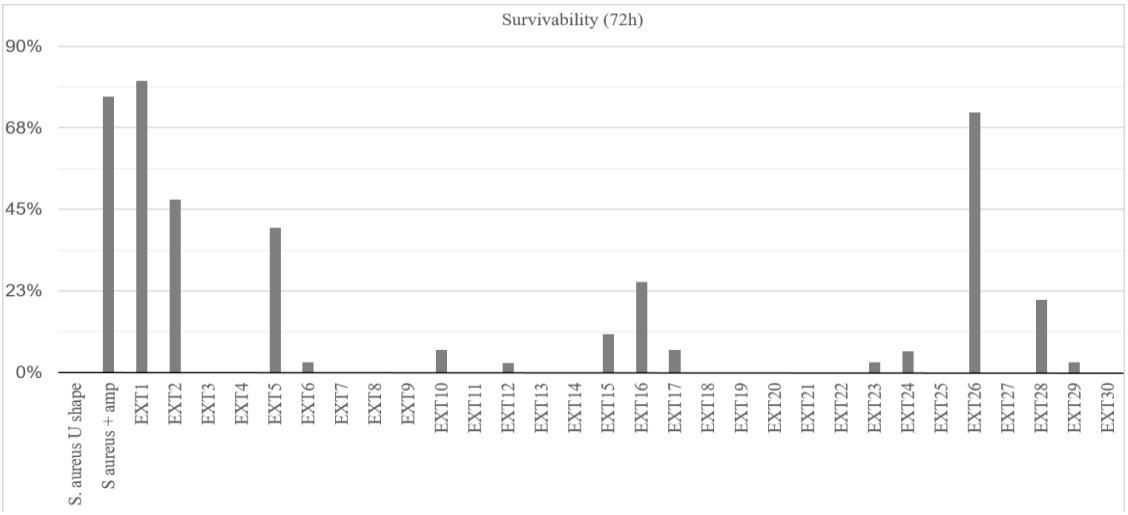


Figure 2.7 - % of nematodes' survival in the presence of *S. aureus* and different extracts

Even though several attempts to create a successful infection model were tried it was not possible to observe a clear killing curve of *C. elegans* due to the bacterial infection. The data obtained suggested that *S. aureus* 8325-4 kills the nematodes in liquid media by suffocation instead of bacterial infection, pointing to the need of extending the spectre of bacterial species to be used to convey the infection. Setting up methods appropriate for screening complex samples such as fermentation extracts would have required a significant amount of my time, with the risk of coming up with no useful results in time to implement the new assay for systematic screening. Therefore, the *C. elegans* assay was not further pursued during this project.

2.6. Discussion

Overall, with 104 active strains, only 50 had their active compounds ascertained, sometimes tentatively, with a further 54 strains still waiting a more detailed analysis. Nonetheless, it is possible to start drawing some conclusions from the strains that were characterized.

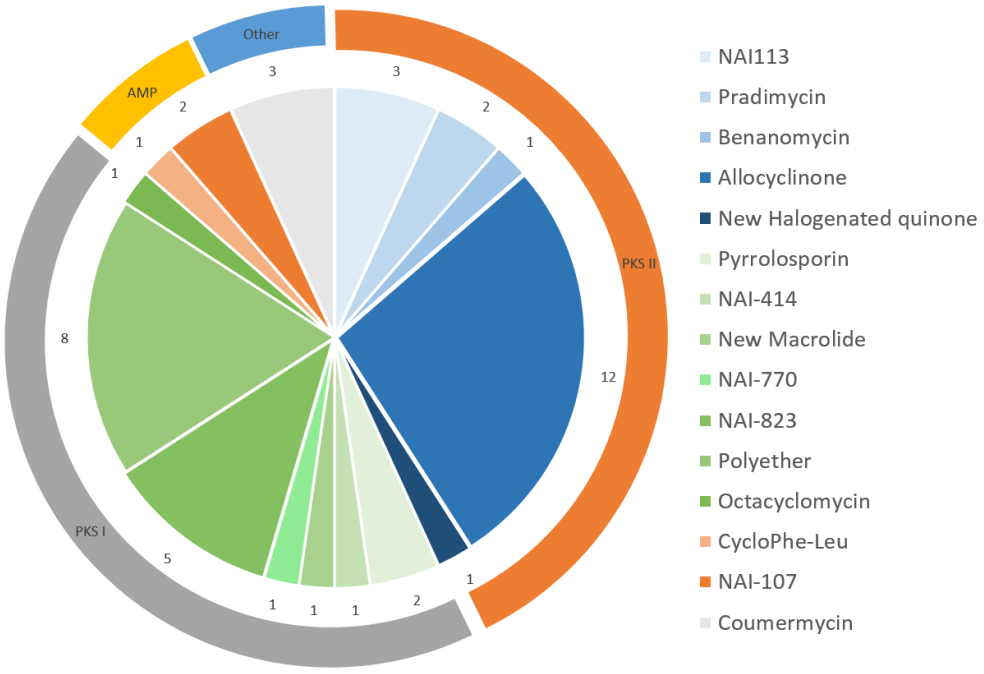


Figure 2.8 – Graphical analysis of the compounds discovered from screening the *Actinoallomurus* genus. Inner circle –Distribution of each specific identified compound from Table 3.2. Outer circle – biosynthetic classification according to the classification from Chapter 1.3. AMP – Antimicrobial peptides

Focusing on the compounds from the characterized extracts, it is possible to observe that there is a large variety of chemical classes present, as illustrated in the graph from Fig. 2.8. Six distinct family of metabolites were identified, including compounds adverting from PKS I, i.e. polyethers (see sub chapter 0), and the spirotetronates pyrrolosporin and NAI-414 (Mazzetti et al., 2012); from PKS-II, i.e. NAI-113 (see chapter 3), allocyclinone (see Chapter 4), and pradimycin; from NRPSs, i.e. Cyclo Phe-Leu (work performed previous to my arrival); RiPPs, i.e. NAI-107 (see chapter 5); as well as compounds biosynthesized outside the major pathways described in the Introduction, i.e. coumermycin (see subchapter 2.4.3). It is also noticeable the large number of strains producing allocyclinone (12 representatives, as discussed in Chapter 4) and polyethers (15 strains). While a little over 200 strains is a small number to draw strong

conclusions, the results obtained thus far suggest that angucyclinones and polyethers might be common encounters in *Actinoallomurus* and these compounds may play important roles in its physiology and be part of the so called 'core' metabolome.

During the screening exercise the major hurdle was the high number of positive hits. The selection criterion was based on the presence of antimicrobial activity against a Gram-positive representative and a hyper-sensitive Gram-negative bacterium. This turned out to be a quite loose criterion as a large number of *Actinoallomurus* strains produced bioactive compounds in the conditions used and not all of them could be structurally characterized within the timeframe of this PhD. This limitation prompted investigating the use of nematodes as a screening tool. The advantage of nematodes was in the possibility of using a small but complex organism at a relatively high throughput without any ethical constraints. A successful venture in this case would have allowed the selection of extracts active against bacteria without adverse effects on a complex organism. However, the resilience of the nematodes to a *S. aureus* infection did not let this endeavour arrive in safe harbour.

As mentioned before, the exploration of new taxa is an important tool to keep exploiting natural chemistry without a major investment on the early discovery process. Different evolutionary pressures and conditions may trigger diverse processes and alter metabolic routes which in turn may be reflected on a different panel of secondary metabolites produced. The results of this expedition on *Actinoallomurus* brought already an interesting set of new compounds reinforcing the idea that new taxa do offer unseen chemistry. However, the finding of new chemistry is only the beginning for new genera. To fully understand the potential of *Actinoallomurus*, it would be important to understand how vast is the chemical space offered by this genus, as already at this early stage different strains were found producing the same compound (see Chapter 4), being interesting to understand how far is *Actinoallomurus* from reaching an exhaustion scenario similar to that more studied genera seem to face.

3. CONTRIBUTIONS FOR THE DISCOVERY OF A NEW METABOLITE

3.1. Introduction

A drug discovery expedition amongst natural products is, when correctly designed, a fructiferous endeavour to unravel new molecular structures with applications that go beyond the boundaries to which it was initially created. For instance, the use of antimicrobial assays for the discovery of new antibacterial compounds may deliver new molecules that although endowed with antimicrobial activity are better suited for other biological activities. The unmatched structural complexity of natural products bequeaths an equally broad range of activities that have been quite exploited by our society.

The formation of aromatic rings is a common biosynthetic strategy from natural products, largely exploited by bacterial PKS II systems to deliver a renowned diversity of structures originating from different polyketide chain lengths and different folding/cyclization patterns, in addition to a large number of post-polyketide modifications. For instance, this biosynthetic logic has provided antibacterials such as tetracycline and antitumor agents like gilvocarcin (Hou et al., 2012; Wang et al., 2012). A family of compounds from the PKS II system that supports the notion of activity beyond the purpose for which it was initially screened for, is represented by the anthracyclines. Anthracyclines, firstly described in 1960's, are a diverse group of tetracyclic compounds that have become very successful in chemotherapy for cancer treatment (Minotti et al., 2004). Four mode-of-actions are known for this family that range from intercalating DNA strands to the generation of free oxygen radicals (Szulawska and Czyz, 2006). The high activity observed against eukaryotic cells make these compounds unsuitable as antibacterial candidates, but they do possess potent antibacterial activity and are readily picked up in such screens (Weiss, 1992).

The diversity of natural products presents quite often different challenges that need to be addressed by multidisciplinary teams. Such was the case for the molecule currently designated as NAI-113, which had been identified as an *Actinoallomurus* metabolite previous

to the start of this project. However, due to its paramagnetic characteristic, the compound showed no distinct NMR signals and its structure remained unsolved. Similar MS and UV pattern were observed when a new *Actinoallomurus* producing strain was identified. My contribution to this work was the use of biological tools to help structural elucidation, through the preparation of ^{13}C-labelled compound by inoculation in a specifically designed media containing ^{13}C glucose and the *in sillico* analysis of the corresponding gene cluster. These results were integrated with chemical and spectroscopic data to eventually establish the molecular structure of NAI-113.

3.2. Experimental Section

Strains and growth conditions. *Actinoallomurus* strains were from the NAICONS collection. Strains were maintained on S1-5.5 plates.(Pozzi et al., 2011) For the initial screening, a loopful of mycelium was used to inoculate shake-flasks containing AF-A medium (dextrose 10 g/L, soybean meal 4 g/L, yeast extract 1 g/L, NaCl 0.5 g/L, MES 1.5 g/L, adjusted to pH 5.6 before sterilization). Cultures were incubated on a rotatory shaker at 200 rpm and 30 ºC and harvested after eight days unless stated otherwise. All extracts were assayed by agar diffusion against *S. aureus* ATCC 6538P and *E. coli* K2020, a ΔtolC strain of MG1641 (from NAICONS collection). Strain ID 145113 was inoculated in AF-N medium (^{13}C-Glucose 5 g/L, Glutamine 1 g/L, yeast extract 0.5 g/L, NaCl 0.5 g/L, MES 1 g/L, adjusted to pH 5.6 before sterilization).

Biological Activity. The antibacterial assays were performed a described in chapter 2.2.

Analytical Procedures. LC-MS and UV data were recorded on a Thermo LCQ equipped with HESI ion source with a Dionex Ultimate 3000 HPLC Column: Atlantis T3, 3µm, 100Å, 4.6mm x 50mm. Phase A: HCOOH 0.1%. Phase B: MeCN. Flow rate: 0.8 ml/min. Gradient: 10% of phase B for 1 minute followed by a linear gradient from 10 to 95% of phase B in 6 minutes and additional 2 minutes at 95% of phase B. HPLC analyses and resolution were carried out using a Shimadzu LC-2010AHT equipped with a Merck LiChrosphere RP18 column, 5 µm (125 × 4 mm). Phase A: 0.1% TFA. Phase B: MeCN. Linear gradient of B from 10 to 90% in 30 minutes. Detection 270 nm.

Purification of NAI-113. A culture of strain ID145113 was harvested and centrifuged at 3000 rpm for 10 min to separate the mycelium from the broth. The latter (about 1.8 L) was concentrated under vacuum to about 1 L and acidified at pH 3.2 by using 7.2 mL HCl 6N. The suspension was kept at 4 °C overnight to allow total precipitation of the desired compounds. The pellet was then collected by centrifugation at 4000 rpm for 10 min, washed twice with acidic water and dried in a vacuum oven at 40 °C for 48 h. 595 mg of a black powder containing compounds 420 and 421 were obtained (named after their mz^+ value). The same downstream process was applied on 200 mL culture of strain 145113 grown in ^{13}C-labeled medium. 42 mg of 80% 13C-labeled 420-421 mixture were obtained. 56 mg of compound 393 were obtained using the same procedure starting from 200 mL of strain 145113 fermentation broth. After purification a 1 mg standard was created and used a reference for production analysis throughout this work.

Genomic sequencing. the methods used for the genome isolation are described in Chapter 6.3.

Bioinformatic analyses. the methods used for the bioinformatic analysis are described in Chapter 6.3.

3.3. Results

Discovery of NAI-113. During a screening program for new antibacterials, three strains presented the same antimicrobial compound, *Actinoallomurus* ID145113, ID145206, ID145754 (being the latter found during this screening exercise; see Table 2.2). From those three, the best producer was strain ID145113, which was used for further analyses. The 16S rDNA phylogenetic analysis confirmed it as a member of the *Actinoallomurus* genus (Fig. 3.1), more accurately belonging to phylotype Alp 4.

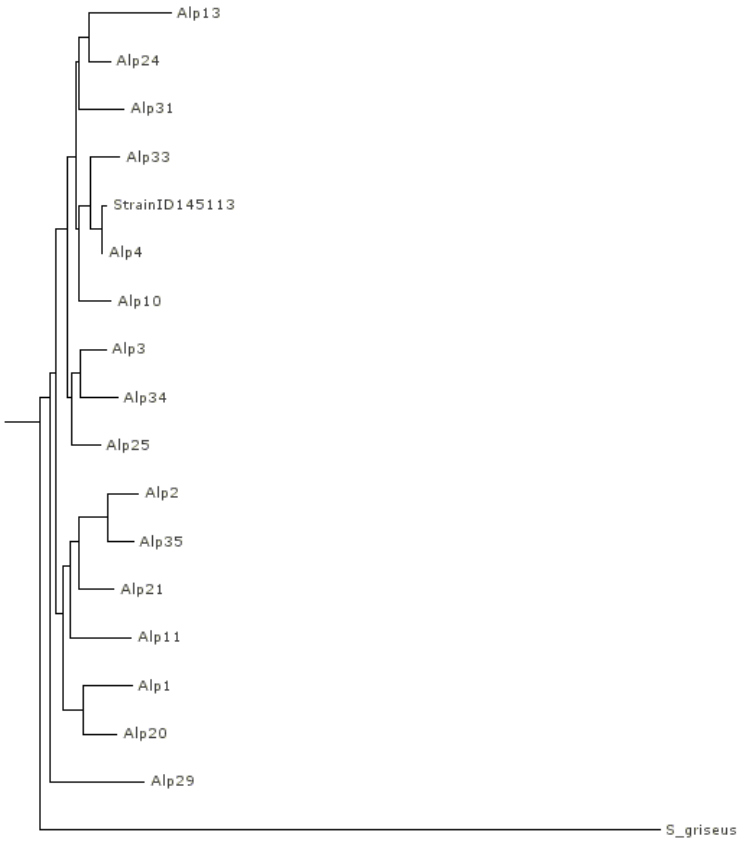


Figure 3.1- Close up on phylogenetic tree of 16S rRNA gene sequences from 134 *Actinoallomurus* strains from the whole collection from NAICONS. Note that strain ID 145113 clusters together with Alp4 phylotype.

The extract showing the largest activity against both Gram positive and Gram negative bacteria – obtained with ethyl acetate – was fractionated using reversed-phase HPLC and each fraction was tested for antibacterial activity, which was observed against *S. aureus* and *E .coli* ΔtolC spell out strains. The activity was correlated to two related peaks with m/z^+ values of 420 and 421, designated **1** and **2**, as described in Fig. 3.2.

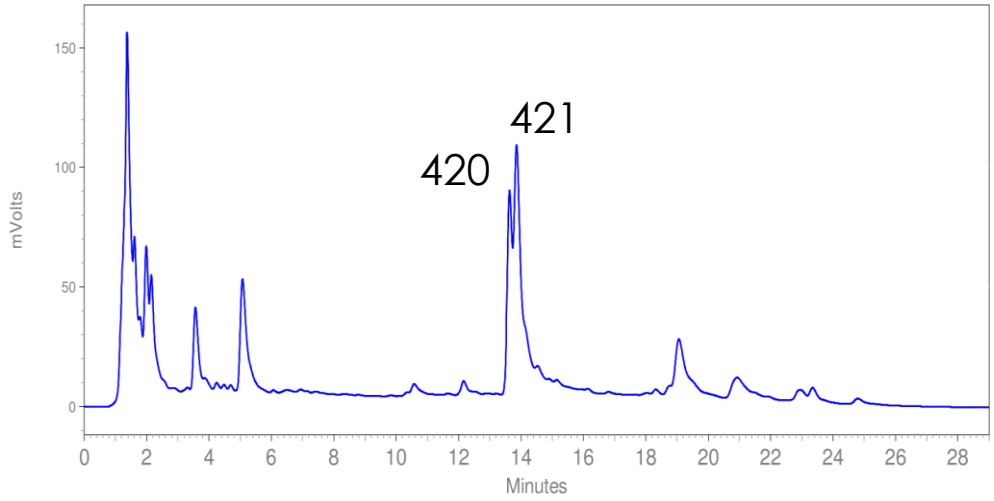


Figure 3.2 – HPLC trace of a clear broth of Strain ID145113 at 270 nm. Note the two peaks correspondent to the m/z+ values of NAI-113 active compounds

The compounds were recovered by precipitation of a concentrated and acidified cleared broth. Washing the resulting pellet with acidified water and drying the residue afforded a black powder that was practically insoluble in water, slightly soluble in organic solvents through the addition of drops of acid (yellow solution), soluble in DMSO, and highly soluble in alkaline aqueous conditions (dark purple solution). LC analysis showed two major peaks with retention times of 13 and 14 min (Fig. 3.2) in a 4:6 *ratio*. The UV-Vis spectrum of **1** showed absorption maxima at 245, 280 and 455 nm, while **2** showed maxima at 254, 275 and 455 nm, suggesting a close relatedness of the two molecules. Both of them exhibited a characteristic bathochromic shift from 455 nm at pH ≤ 7.0 to 550 nm at higher pH. Previous HR-MS analyses established their chemical formula as $C_{22}H_{14}NO_8$ for 420 $[M+H]^+$ (420.0714 found, 420.0719 calcd) and $C_{22}H_{13}O_9$ for 421 $[M+H]^+$ (421.0557 found, 421.0559 calcd).

The first NMR analyses of the **1-2** mixture showed total absence of signals, indicating interference between the sample and the instrument. This behaviour suggested the possible presence of a free radical as in the so-called "NMR-silent" compounds, paramagnetic compounds that can exist partially or totally as free radicals, interfering with the magnetic field of the NMR instrument, hence the missing NMR signals. This feature is usually associated with quinonic structures that undergo oxygen-mediated radical formation (Shin-ya et al., 1992). In the literature it has been possible to see and analyse NMR signals only in derivatives of such compounds, usually in their acetylated products.

Acid-catalysed acetylation was thus performed on the mixture **1**-**2** leading to the formation of a complex mixture of reduced acetylated products (Wilgus and Gates, 1967). In particular, **1** and **2** incorporated up to four and three acetyl groups, respectively, with the most abundant peaks consisting of a three-acetyl-2H-**1** and a diacetyl-2H-**2**. The two major acetylated molecules were resolved by preparative TLC and the purified compounds finally afforded sharp NMR signals. Selected data is reported on Table 3.1. In addition to the signals belonging to the introduced acetyls, whose positions were not investigated, the ^{1}H-NMR spectrum of diacetyl-2H-**1** showed a phenolic proton (147.2 ppm), three aromatic singlets (8.26, 7.85 and 7.67 ppm), a methylene (5.85 ppm) and two methoxy signals (4.11 and 3.98 ppm). Further 2D experiments highlighted a NOESY correlation between the aromatic CH at 7.85 ppm and the methoxy group at 4.11. HMBC correlations among the methyl groups at 3.98 and 4.11 ppm and the quaternary carbons at 147.2 and 158.4 ppm, respectively, revealed a phenolic nature for these methoxy groups while only one carbonyl signal was detected (167.4 ppm) correlating in HMBC with the methylene at $_{\delta}$H 5.85. This one, in turn, showed a long-range TOCSY correlation with the aromatic CH at $_{\delta}$H 7.67. HMBC data also showed the presence of two unreduced quinonic signals ($_{\delta}$C 181 and 180.6).

Table 3.1 - Selected NMR spectroscopic data (400 MHZ, $CDCL_3$) for compound 2AC-2H-421, 3AC-2H-420 AND 4AC-2H-393.

POSITION	2Ac-2H-421		3Ac-2H-420		4Ac-2H-393	
	δ_C, type	**δ_H**	**δ_C, type**	**δ_H**	**δ_C, type**	**δ_H**
1	167.4, C		166.4, C		166.1, C	
3	147.2, C		162.9, C			
4	122.4, CH	7.67, s	124,2, CH	7.49, s	122.1, CH	8.14, s
6	120.1, CH	8.26, s	n.d.	8.41, s	116.1, CH	8.35, s
8	158.4, C		n.d.			
9	108.7, CH	7.85, s	108.7, CH	7.91, s	119.9, CH	7.34, s
14	70.9, CH_2	5.85, s	70.6, CH_2	5.77, s	71.1, CH_2	5.81, s
15	61.3, CH_3	3.98, s	57.2, CH_3	4.23, s		
16	56.5, CH_3	4.11, s	36.1, CH_3	3.25, s		
OH-11		14.75, brd		n.d.		14.86, brd

^{1}H-NMR data of the triacetyl-2H-**2** showed a similar pattern (see Table 4.1) except for the presence of a single methoxy (4.23 ppm) and of an additional methyl (3.25 ppm) whose δC

at 36.1 was in agreement with an aromatic methyl-amino group. This suggested that the difference between **2** and **1** might consist in the substitution of a methoxy with a methyl-amino group, consistent with the HRMS-derived formulae. However, the paucity of protons together with the abundance of quaternary carbons prevented complete structure elucidation (Marianna Iorio, Personal communication). Therefore, new strategies to overcome this hurdle were designed from using biological tools.

Bioinformatic Analysis. A first strategy employed was a genomic approach. In this case, the genome of strain ID 145113 was sequenced and its 10.6 Mbp genome sequence (see Chapter 6 for further details on the genome sequence) was used to run the bioinformatic tool Antismash V3.0.0.4 (Weber et al., 2015), which is able to pinpoint the putative gene clusters for secondary metabolism. Looking through the generated list and aware of some of the structural features of **1** (quinonic structure, no halogens, high degree of decorative motifs across the core structure) it was possible to find the cluster most likely responsible for its production (Fig. 3.3).

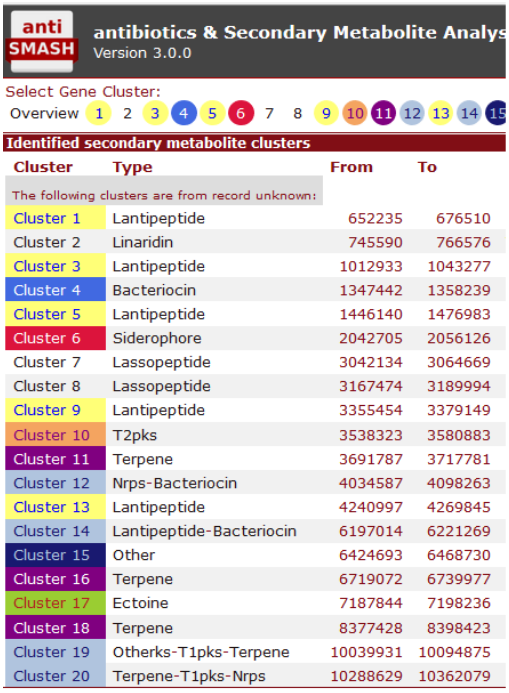


Figure 3.3 - List of the potential clusters for secondary metabolism from Strain ID 145113. Cluster 10 –candidate for NAI-113 biosynthesis

Cluster 10 was selected as the most likely candidate because it encodes a PKS II system, which usually produces this type of molecules. The ORFs present in cluster 10 were analysed

for similarity against the NCBI database but also with a recently introduced specialised database of BGCs responsible for compounds with known structure, available at MIBiG. (Medema et al., 2015) The results are described on Table 3.2.

Table 3.2- Deduced Function of each ORF in the NAI-113 Biosynthetic Gene Cluster

CDS	Length (aa)	Proposed function	Best PKS Match[a]	Best hit (Ac Number)	Identity
NAI-113_3480	330	ABC Transporter		WP_010356765.1	65%
NAI-113_3481	352	ABC Transporter		WP_041824945.1	77%
NAI-113_3482	299	ABC Transporter		WP_026413477.1	79%
NAI-113_3483	390	Myo-inositol 2-dehydrogenase		KND25197.1	65%
NAI-113_3484	175	Cation-binding protein [Streptomyces flavotricini]		WP_030036697.1	31%
NAI-113_3485	534	bilirubin oxidase		WP_012870914.1	56%
NAI-113_3486	151	Monoxygenase	RubQ	AAM97373.1	47%
NAI-113_3487	330	Malonyl CoA-acyl carrier protein	TstB	AGN11890.1	39%
NAI-113_3488	621	EmrB/QacA subfamily drug resistance transporter		YP_001866562.1	52%
NAI-113_3489	127	Glyoxalase		YP_003115770.1	66%
NAI-113_3490	431	E2_dihydrolipoyl_acyltransferase		ACR50769.1	38%
NAI-113_3491	344	Transketolase		AGD80611.1	53%
NAI-113_3492	327	Pyruvate_dehydrogenase_E1_alpha_subunit		ACR50771.1	66%
NAI-113_3493	130	Monoxygenase	PdmH	ABM21754.1	42%
NAI-113_3494	107	tetracenomycin F2 cyclase		WP_043504911.1	46%
NAI-113_3495	184	PadR Regulator		AAV48829.2	71%
NAI-113_3496	287	Trasncriptional Regulator - SARP family		CAA60570.1	62%
NAI-113_3497	123	Cyclase	Lcz30	ABX71147.1	52%
NAI-113_3498	421	Type II PKS ketoacyl synthase	PdmA	ABM21747.1	70%
NAI-113_3499	432	Chain Length Factor	SnoA2	CAA12018.1	61%
NAI-113_3500	84	Acyl Carrier Protein	SimA3	AAK06786.1	53%
NAI-113_3501	155	Cyclase	RubF	AAG03070.2	72%
NAI-113_3502	349	O-methyltransferase	PokMT3	ACN64847.1	50%
NAI-113_3503	576	Acetyl Carboxylase	PokAC1	ACN64837.1	58%
NAI-113_3504	167	Biotin Carboxyl carrier protein	PokAC3	ACN64838.1	45%
NAI-113_3505	452	Acyl-CoA carboxylase	PokAC2	ACN64839.1	63%
NAI-113_3506	53	Unknown Function		YP_001539692.1	51%
NAI-113_3507	357	O-methyltransferase	PokMT2	ACN64843.1	41%
NAI-113_3508	562	Hydroxylase	TcmB2	CAP12596.1	47%
NAI-113_3509	331	O-methyltransferase	PbtM4	AGY49596.1	53%
NAI-113_3510	299	Nucleoside-diphosphate sugar epimerase		ABP54642.1	47%
NAI-113_3511	53	AAA family ATPase		WP_055481033.1	74%
NAI-113_3512	282	LigA protein		EWM17234.1	50%

Best PKS matches are based on MiBIG Database.
Dashed lines define the limits of the putative gene clusters within the Antismash output

Looking into more detail at cluster **10** it was possible to locate the minimal PKS – the group of enzymes responsible for the construction of the initial backbone chain – but also other enzymes (such as cyclases or methyltransferases) important to the definition of the final structure. Previous studies have shown that the KSβ is directly related to the type of molecule its cluster synthesizes and it can even be used to predict novelty (Kang and Brady, 2014; Metsa-Ketela et al., 1999). Using this line of thought, a phylogenetic analysis was performed as described in Fig. 3.4. By aligning KSβ from cluster **10** with 35 different sequences from other PKS II systems, it was possible to establish that the enzyme responsible for the synthesis of the compound under analysis (dubbed "NAI-113") aligns best with the KSβs responsible for the synthesis of pradamycin-like compounds. These molecules derive from a carbon backbone made of 12 to 13 acetate units. Although the NAI-113 KSβ appears to diverge from other enzymes of the same pradimycin-like branch, it is nonetheless more related to these enzymes than to those from the angucycline or tetracycline branches. The lack of close sequences, while suggestive of novel structural features, prevented a tentative identification of a close structural homologue for NAI-113.

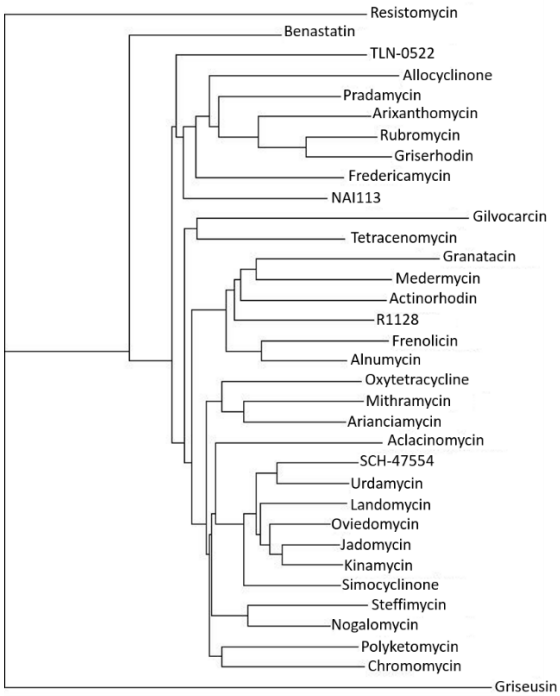


Figure 3.4 - Phylogenetic tree of the KSβ from 35 different biosynthetic clusters with known product. The KSβ from NAI-113 cluster is close to the Arixanthomycin KSβ within the Pradamycin-like cluster

Production of ^{13}C-labeled NAI-113. By increasing the frequency of 13 C present in the metabolite, the NMR response is also expected to be increased and also, the labelled compound can be used in oxidation reactions to enhance the detection of the numerous quaternary carbons and complete structural characterization. The first step to produce labelled compound was to design a minimal media convenient for *Actinoallomurus* growth without compromising its ability to produce the desired metabolite. The simple introduction of a labelled carbon source in the place of regular glucose did not yield satisfactory results (data not shown). The medium used as a starting point was the M9 minimal medium, with glucose substituted by its labelled form. In this case there was no production and growth was very limited, indicating that M9 was not appropriate. Dobrova *et al*., 2004, had reported a minimal medium for *Actinomadura* – actually, it was *Nonomurea* sp. producing the glycopeptide A40926 – able to support growth and where the nitrogen source played a key role for triggering secondary metabolism. (Technikova-Dobrova et al., 2004) This data was also corroborated with *Streptomyces coelicolor* by Wentzel et al., 2012. (Wentzel et al., 2012) Taking inspiration from this work, different nitrogen sources were used with M9 media, achieving a detectable production with glutamine, but still insufficient for further work. Therefore, a new strategy was used, starting with the AF-A medium (used for growth of *Actinoallomurus*) and reducing and

replacing individual components. Eventually, a new recipe was devised with a different nitrogen source, with reduced yeast extract and a new Nitrogen source. (Table 4.3).

Table 3.3 – Differences in components between the media used

Components	AF-A	AF-X	M9-X
Glucose	10	10	10
Soy flour	4	-	-
N-source	-	2	2
Yeast extract	1	0,5	-
MES	1,5	-	-
MOPS	-	1,5	1,5
NaCl	0,5	0,5	0,5

The new media were called AF-X and M9-X, where X indicates different nitrogen source. Fig. 3.5 compares NAI-113 production in the six best media. Both glutamine and nitrate allowed production of 15-20 µg/mL NAI-113 in AF-X media, whereas production levels in the presence of ammonium as a nitrogen source were much lower. Note the importance for NAI-113 production of the small amount of yeast extract, present in AF-X but not in M9-X media. Yeast extract usually contains elicitors important for bacterial growth and a higher amount of biomass may also lead to an increase of produced compound. As mentioned in the literature, the use of glutamine seems to enhance the production of secondary metabolites, so AF-Gln medium was used for further optimisation.

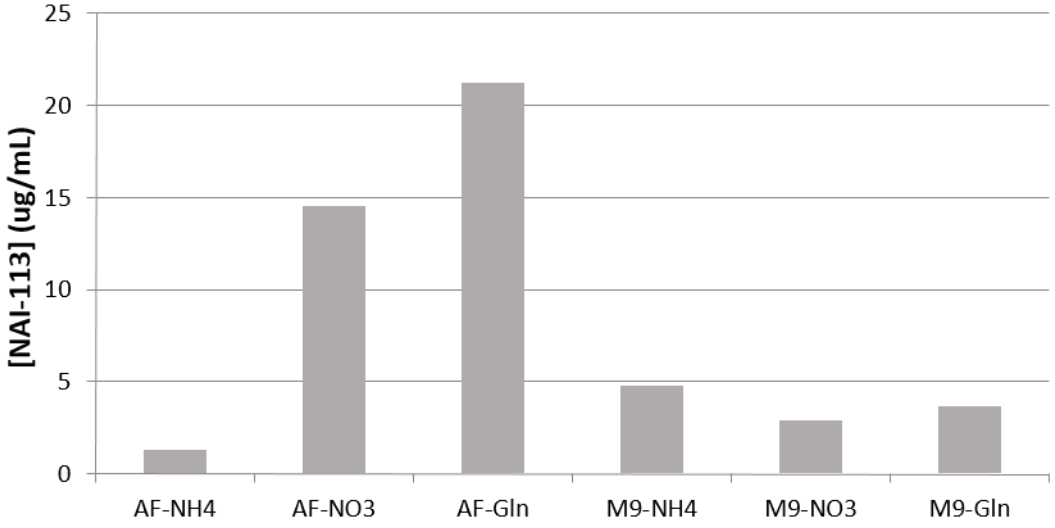


Figure 3.5 - Comparison of production in six different media. Media are named "AF-X" where "X" is replaced by the indicated Nitrogen source (n=3).

Figure 3.5 shows the impact on NAI-113 production of altering the concentrations of glucose and glutamine. Under the best conditions (5 g/L glucose and 1 g/L glutamine), it was possible to increase NAI-113 production up to 20 fold compared to AF-Gln. Although, it was noted that for glucose concentrations of 10 g/L the inverse was observed and the production was higher with the increase of glutamine for 2g/L, while in the case of 20g/L no trend was observed, further studies would be necessary to explain such behaviour from *Actinoallomurus* which was not the goal of this project but instead we were only interested in achieving a sustained production of labelled NA-113 which was accomplished. Finally, instead of using 8 eight days of cultivation (the standard procedure), it was accidentally observed that longer incubation times (10 days) yielded a culture with a darker colour, which was due to a higher quantity of NAI-113, reaching concentrations of 400 μg/mL. Note that there was no linear relationship between glucose concentration and NAI-113 production, nor there was a consistent effect of increasing glutamine concentration (Fig. 3.6). This phenomenon was not further investigated.

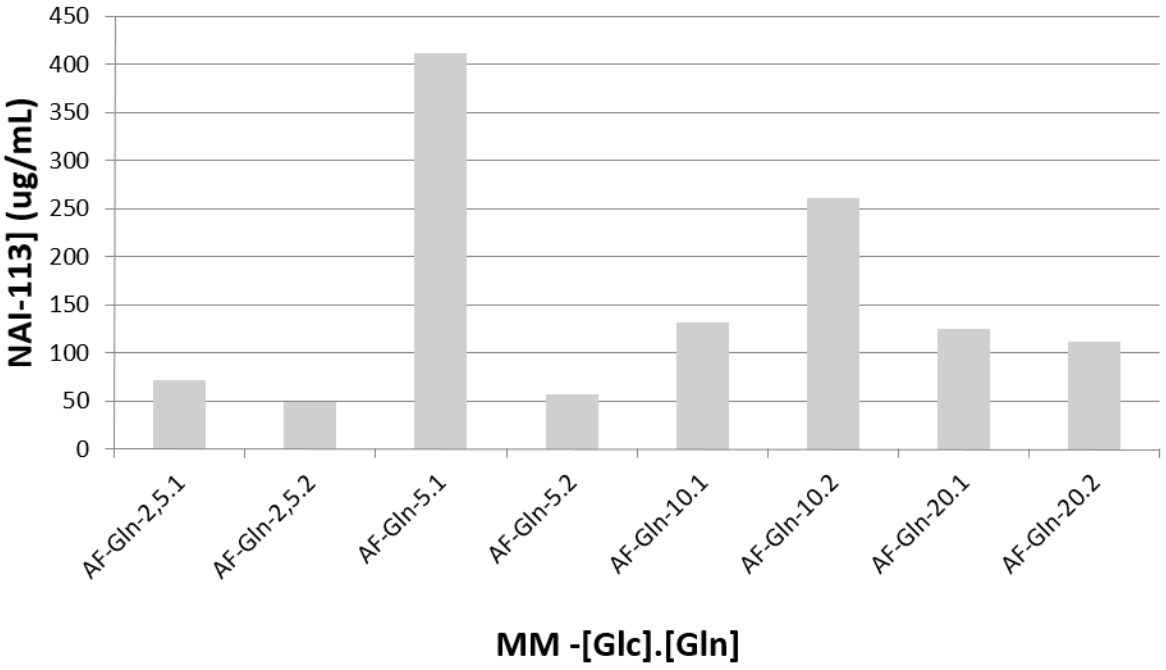


Figure 3.6 - Comparison of production in AF-based media, after eight days, containing different concentrations of glucose and glutamine. The concentrations of glucose, and glutamine (in g/L) are indicated by the first and second number, respectively.

According to mass analysis, the labelling reached, on average, up to 8 out of 10 of ^{13}C incorporation, showing mainly a shift of 18 mass units from the original value as described in Fig. 3.7 (according to High resolution mass, 22 carbon atoms are present). Although glutamine is known to be involved in the primary metabolism and, mainly in the amino acid synthesis, it still can be used as a carbon source for secondary metabolism (as it was already established by Kanehisa, *et al*., 2004 (Kanehisa and Goto, 2000)) which can be one of the reasons for the lack of a higher degree of incorporation. However, two other facts may have played a role: the presence unknown compounds present in Yeast Extract, and a carryover of unlabelled carbon from the inoculum. Anyhow, the labelled compound was sufficient for the chemical degradation studies described below.

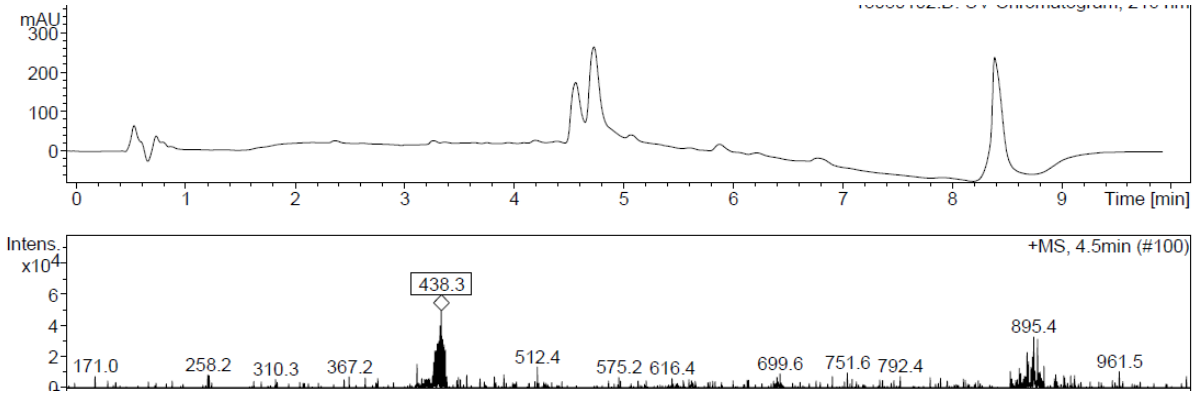
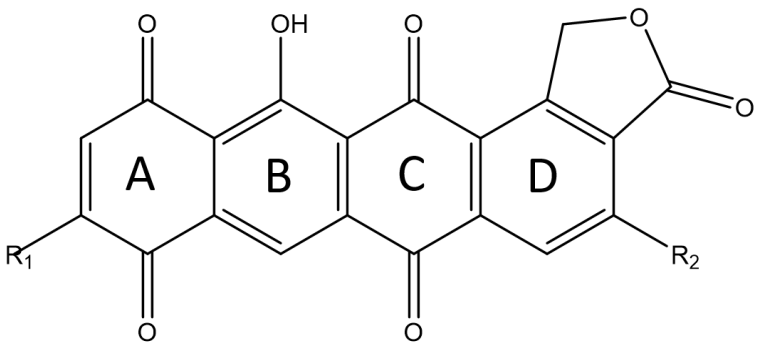


Figure 3.7 – m/z$^+$ values of ^{13}C labelled NAI-113

Degradation studies. In order to break the quinone system and thus eliminating the radical behaviour, the mixture **1-2** was treated with 30% H_2O_2 in water, which led to the formation of a more hydrophilic molecule with *m/z* of 399 $[M+H]^+$ and UV maxima at 230, 280 and 420 nm. This result indicated that oxidation had removed the portion responsible for the mass difference between 420 and 421. As predicted, the purified compound was no longer paramagnetic and clear ^{1}H-NMR signals were observed. Moreover, MS data combined with absence of one aromatic proton and one methoxy group in **2** (or the methyl-amino in **1**) were in agreement with oxidation of a terminal quinone moiety cleaved and replaced by two carboxylic acids as reported in the literature. (Ogawa and Natori, 1965) This oxidation reaction was thus repeated on a sample of ^{13}C enriched compound to enhance the detection of the numerous quaternary carbons and complete structural characterization. The resulting NMR analyses allowed determination of the structures of **1** and **2** as reported in Fig. 3.8 (M. Iorio et al, personal communication). All the data combined led to the conclusion that the compounds described in this work are new members of the anthracyclinone family, presenting an unusual lactone condensed to ring D. These structures are consistent with a biosynthetic route in which a longer polyketide is cleaved to afford the lactone system, consistent with the bioinformatic predictions of a 12- or 13-unit polyketide.



	R_1	R_2
A	$NHCH_3$	OCH_3
B	OCH_3	OCH_3
C	OH	OH

Figure 3.8 - Molecular Structure of NAI-113 complex

Related metabolites from a different *Actinoallomurus* strain. Compounds indistinguishable from **1** and **2** were also found as minor congeners in the fermentation broths from strains ID145206 and ID145754, belonging to a different *Actinoallomurus* phylotype from strain ID145113. The main compound produced by strain ID145206 is congener C (Fig. 3.8), which is also NMR silent. Its structure was elucidated following the same acetylation and degradation procedures applied to NAI-113.

Biological activity. As expected for anthracyclinones, NAI-113 was highly active against Gram-positive bacteria (M. Iorio, personal communications).

3.4. Discussion

The discovery of NAI-113 starts before my personal project. The first time this molecule was found was two years before the beginning of my project (Marianna Iorio, personal communication) with strains ID 145113 and ID145206. In a similar manner to what was done during this screening exercise, these strains were identified as producers of an unidentified compound that upon NMR analysis revealed silent and thus impossible to have its structure solved. Already at that time, the first attempts on designing a minimal medium for *Actinoallomurus* were performed with no usable results. The major hurdle when working with a poorly characterized genus, such as *Actinoallomurus*, is the lack of literature information that renders every experiment with many unknowns. Designing a new medium was a good example of such obstacles.

There is no doubt that it would have been of interest to actually pinpoint the compounds that actually are indispensable for the production of secondary metabolites in *Actinoallomurus*, however a *de novo* approach for media design might reveal a time consuming process with no guarantee of results in the desired timeframe. Since the main goal of this project is the characterization of the new compound, it was important to devise conditions usable for the efficient uptake of labelled carbon (^{13}C) to be incorporated in NAI-113. The use of chemically defined media such as M9, used successfully with *Streptomyces*, *Nonomurea* and many other organisms, was not possible with this *Actinoallomurus* strains. The presence of yeast extract, a rather complex component, in the media seems essential for NAI-113 production. On the other hand, soy flour, another complex ingredient, could be replaced by suitable nitrogen sources such as nitrate or glutamine without a significant decrease of NAI-113 production. Despite no visible effects were observed in biomass growth these changes altered significantly the production of the secondary metabolite. This discovery may have implications for future drug discovery expeditions as the use of soy flour can lead to bioactivity artefacts. Indeed, some *Actinoallomurus* strains can metabolize soy flour components into genistein and other soy saponins that present a detectable antibacterial activity (Pozzi et al. 2011), complicating the screening endeavours. In addition, soy saponins have little UV absorption but give strong MS and NMR signals, complicating purification and analysis of *Actinoallomurus* metabolites. These issues could be avoided if glutamine could be systematically used to replace or reduce the amount of soy flour in culture media. However, this remains an unexplored possibility.

In medium AF-Gln containing labelled glucose, NAI-113 contained around 80% ^{13}C, suggesting that additional carbon source(s) were used as substrates. Nonetheless it is quite remarkable that in a medium containing diverse sources of unlabelled carbon such as biomass, carry-over from previous inoculum, glutamine, or yeast extract still such a high percentage of ^{13}C uptake was already achieved. Although a complete labelling would be interesting, components such as Yeast extract or Glutamine will always render the task quite difficult. Still, it would be wise to verify if with the use of nitrate instead of Glutamine, the percentage of labelled carbons in NAI-113 would vary. Although a complete characterization of a chemically defined media for *Actinoallomurus* was not achieved. Important insights were made into this strain's metabolism that shall be of value in future drug discovery expedition.

The structural insights provided by analysis of the NAI-113 BGC were less incisive than those derived from ^{13}C-labeled compound. The bioinformatic analysis suggested possible polyketide scaffolds for NAI-113 while making others less likely. The study of the Minimal PKS, and namely of the KSβs, is becoming an important tool to predict with a high degree of confidence the type of aromatic polyketide scaffold produced by the PKS (Szu et al., 2011; Tang et al., 2003), with the tantalizing corollary that divergent KSβs might in turn lead to novel polyketide scaffolds. Hence, the bioinformatic techniques used in this and subsequent chapters might also be useful in genome mining approaches to select strains potentially producing unusual aromatic polyketides

4. ALLOCYCLINONES, HYPERCHLORINATED ANGUCYCLINONES FROM *ACTINOALLOMURUS*

4.1. Introduction

The importance of natural products to human health has been acknowledged over the past few decades and still is. According to recent analyses, among the new chemical entities approved as drugs between 1981 and 2010, approximately 68% are directly derived or inspired from natural products. This success, despite the growing disinterest by pharmaceutical companies in natural products as a source of drug leads, underscores Nature's ability to deliver complex and high affinity chemical scaffolds that have no match in most synthetic libraries (Ganesan, 2008; Newman and Cragg, 2012). One therapeutic area that has been strongly dependent on natural products has been the anti-infectives, particularly antibacterial agents that have derived almost exclusively from microbial products. The past few years have seen a dramatic change in the landscape of infectious diseases, with a steady emergence and spread of antibiotic resistance giving origin to pathogens that can escape most available therapeutic options (Wright, 2012).

The so-called golden era of antibiotics has yielded the vast majority of microbial metabolites, most of them presumably discovered because of their antibacterial potency, the amount produced by the corresponding strains, and/or the natural frequency of occurrence in nature (Baltz, 2008; Monciardini et al., 2014). As these 'low-hanging fruits' were discovered, random screening of microbial isolates led to progressively diminished returns in terms of chemical novelty. Therefore, different approaches have been proposed and implemented to increase the probability of making metabolite discovery from microbial sources a cost-effective endeavour (Baltz, 2007; Donadio et al., 2009; Gengenbacher and Dick, 2015; Ling et al., 2015; Pena et al., 2015).

One possible approach to identify new metabolites is to explore microbial taxa that have not been previously systematically evaluated for the production of bioactive metabolites. In this respect, screening for antimicrobial activity may be a simple surrogate for other biological effects, since over half of the antimicrobial compounds discovered so far can exhibit other bioactivities (Berdy, 2012). The main premise to the choice of using poorly characterized taxa is that taxonomic divergence can be a marker for chemical diversity. In particular, the genus

Actinoallomurus has proving to be a versatile producer of bioactive compounds, including novel ones (Inahashi et al., 2015; Mazzetti et al., 2012; Pozzi et al., 2011).

This chapter coincides with the manuscript submitted to the Journal of Antibiotics, the work reports new entries in the angucyclinone family and the first natural product bearing an aromatic trichloromethylene substitution. Allocyclinone, as the molecules have been dubbed, seems to be quite common amongst different *Actinoallomurus* isolates. The molecule possesses an interesting antibacterial activity and the first insights on structure-activity relationship have been surmised from an evaluation of the purified component of the allocyclinone complex. Along with influential insight of the biosynthesis of this molecule to which the gene cluster was putatively identified by *in sillico* analysis.

4.2. Experimental Section

General procedures. NMR spectra were measured in $CDCl_3$ on a Bruker 400 MHz instrument using tetramethylsilane as an internal reference. HR-MS analyses were performed on an Exactive benchtop mass spectrometer (Thermo Fisher Scientific), equipped with a NSI-ESI ion source by direct infusion of the mixture **1**, **2**, **3**, and **4** dissolved in MeOH:H_2O 8:2. LC-MS and UV data were recorded on a Thermo LCQ equipped with HESI ion source with a Dionex Ultimate 3000 HPLC Column: Atlantis T3, 3µm, 100Å, 4.6mm x 50mm. Phase A: HCOOH 0.1%. Phase B: MeCN. Flow rate: 0.8 ml/min. Gradient: 10% of phase B for 1 minute followed by a linear gradient from 10 to 95% of phase B in 6 minutes and additional 2 minutes at 95% of phase B. HPLC analyses and resolution were carried out using a Shimadzu LC-2010AHT equipped with a Merck LiChrosphere RP18 column, 5 µm (125 × 4 mm). Phase A: 0.1% TFA. Phase B: MeCN. Linear gradient of B from 10 to 90% in 30 minutes. Detection 270 nm.

Strains and growth conditions. *Actinoallomurus* strains were from the NAICONS collection. Strains were maintained on S1-5.5 plates.(Pozzi et al., 2011) For the initial screening, a loopful of mycelium was used to inoculate shake-flasks containing AF-A medium (dextrose 10 g/L, soybean meal 4 g/L, yeast extract 1 g/L, NaCl 0.5 g/L, MES 1.5 g/L, adjusted to pH 5.6 before sterilization). Cultures were incubated on a rotatory shaker at 200 rpm and 30 ºC and harvested after eight days. The mycelium was separated from the clear broth by centrifuging at 12000 rpm for 5 min. The clear broth was extracted with one volume of ethyl acetate and the two resulting phases were dried and resuspended in 10% DMSO in 1/5 the original volume using the same strategy described in Chapter 2.2. The mycelium was

resuspended in 0.4 vol EtOH and incubated at room temperature for 2 h. After separating the mycelium by centrifugation as above, the supernatant was dried and resuspended in 0.2 vol 10% DMSO. All three extracts were assayed by agar diffusion against *S. aureus* ATCC 6538P and *E. coli* K2020, a ΔtolC strain of MG1641 (from NAICONS collection).

For characterization of allocyclinones, frozen stocks of *Actinoallomurus* sp. ID145698 were used to inoculate AF-A medium and incubated for 72 h as above. Then, a 5% inoculum was transferred into fresh medium and incubated as above for the required amount of time. Biomass accumulation was measured as percent packed mycelial volume (PMV%) after centrifuging cultures for 10 min at 3000 rpm in a graduated tube. For metabolite analysis, culture samples were mixed with 2 vol MeOH-AcOH (92:8), incubated for 1 h at room temperature and analysed by HPLC as above after separating the residual biomass. Congener ratio was established assuming identical molar absorptions at 270 nm. Compounds concentrations were measured using pure **4** as reference standard.

Compound purification. After 192 h cultivation in medium AF-A, the entire culture (400 mL) was acidified to pH 5 with AcOH, and centrifuged (3000 rpm for 10 min) to separate the mycelium from the cleared broth. The mycelium was extracted with EtOH (100 mL), incubating in shaker, at room temperature for one hour, after centrifugation the resulting supernatant was separated from the mycelium and evaporated. The cleared broth was extracted with EtOAc (150 mL) and the organic phase was separated and evaporated. The two extracts showed similar congener distributions and were separately submitted to the following purification procedure with comparable results. Preparative TLC were performed in dichloromethane-MeOH = 9:1 containing 0.5% (*v/v*) AcOH. The main spot with 0.9 *Rf* was isolated, recovered by elution with MeOH resulting in a mixture of compounds **1**–**4** (13 mg). Additional spots with 0.2 and 0.5 R*f*, recovered in the same way, resulted in pure compound **5** (4mg) and compound **6** (1 mg).

An aliquot (500 μg) from the mixture **1-4** was then further purified by HPLC obtaining pure samples of **1** (65 μg), **2** (178 μg), **3** (100 μg) and **4** (176 μg). Quantities were estimated by integrating the area of the respective peak from the HPLC chromatogram.

Oxidation and stability analyses. Solutions of **2**, **3** and **4** in DMSO at a concentration of 10 mg/mL were analysed by LC-MS before and after treatment at 60°C overnight.

Antimicrobial Assays. All MICs were determined by broth microdilution in sterile 96-well microtiter plates, using Müller Hinton broth (Difco Laboratories) containing 20 mg/L $CaCl_2$ and 10 mg/L $MgCl_2$ for all strains except for *Streptococcus* spp., which were grown in Todd

Hewitt broth (Difco Laboratories). Strains were inoculated at $5x10^5$ cfu/mL and incubated at 37 °C for 20 h. For growth inhibition of *E. coli* K2020, the same procedure was used except the OD_{590} was continuously measured in the BioTek's plate reader, Synergy2. Tested compounds were dissolved in DMSO at 10 mg/mL and diluted with the culture medium immediately prior to testing. All strains were either from the ATCC or the NAICONS collection and the relevant resistance phenotypes are reported in Table 5.3.

Sequence analyses. PCR amplification of the 16S rRNA sequences was performed as previously described (Monciardini et al., 2002) and the amplified products were sequenced by Cogentech (Milan, Italy). Whole genome sequencing was performed by Universität Bielefeld Centrum für Biotechnologie (Bielefeld, Germany). Sequence alignments were made using ClustalW (for 16S rRNA) (Thompson et al., 1994)and strains were assigned to the same phylotype when their identity was higher than 99.5%. Phylogenetic trees were made with the neighbor-joining method (Saitou and Nei, 1987). Secondary metabolite gene clusters were searched with the ANTISMASH v3.0.4 tool (Weber et al., 2015). Blast analysis were done using data from DoBISCUIT and MIBiG as a reference (Ichikawa et al., 2013; Medema et al., 2015). Phylogenetic analysis of all Chain lengths factors was done using MUSCLE algorithm. All the bioinformatic work was done with the support of the workflow platform Armadillo. The putative allocyclinone gene cluster has been deposited in GenBank under Accession Number KT996129.

4.3. Results

Identification of new angucyclinones. The procedure for screening has been previously described, and can be summarized by the use of strain cultivation in shake-flasks, followed by testing the resulting extracts for antimicrobial activity. Active extracts are then evaluated for the novelty of the active compound(s) (Monciardini et al., 2014). With this approach, we identified *Actinoallomurus* strain ID145698 as a strain worth of further investigations. Extracts derived from this strain inhibited growth of *Staphylococcus aureus* and of a hyper-permeable *Escherichia coli* strain. Upon LC-MS analysis, the antimicrobial activity was associated to a complex of related halogenated molecules **1-4** (Appendix 1 - Figure 1).

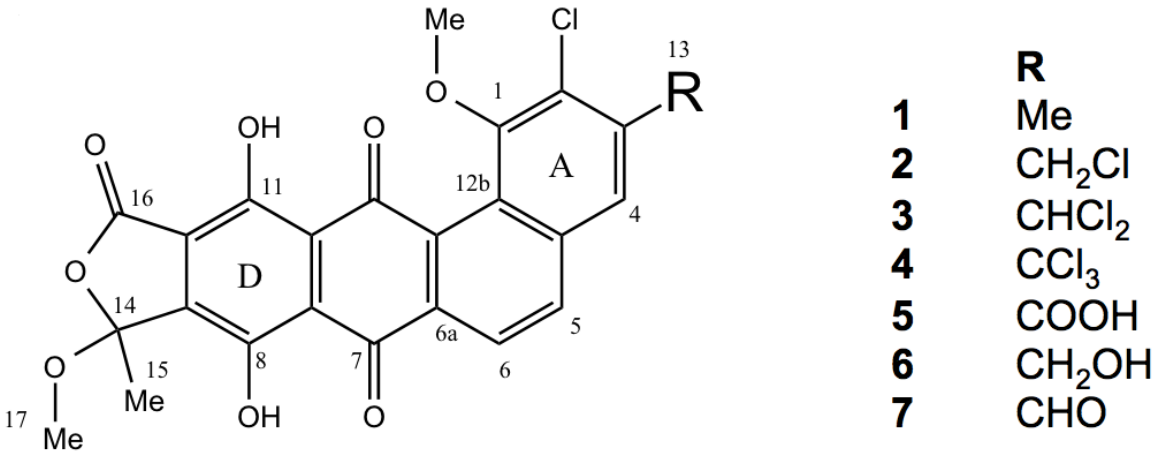


Figure 4.1 - Molecular structures of compounds **1**–**7**.

Structure determination. The structures of compounds **1-4** are shown in Figure 4.1 and were determined as follows. The UV-spectra of the purified molecules were very similar with absorption up to 650 nm and the presence of three maxima at 240, 300 and 480 nm, suggesting an angucyclinone chromophore (Potterat et al., 2007). The isotopic pattern in the ESI mass spectra clearly revealed the presence of at least one chlorine atoms in all congeners. The experimental isotopic patterns and those calculated from the HR-MS-derived molecular formulas (Table 4.1) were in full agreement (Appendix1 – Figure 2) and confirmed that compounds **1** to **4** contain one to four chlorines, respectively.

Table 4.1 - HR-MS data for congeners 1 to 4 of allocyclinone**.**

Compound	Observed mass[a]	Calculated[a]	Formula
1	469.0681	469.0690	$C_{24}H_{17}ClO_8$
2	503.0289	503.0300	$C_{24}H_{16}Cl_2O_8$
3	536.9899	536.9911	$C_{24}H_{15}Cl_3O_8$
4	570.9510	570.9443	$C_{24}H_{14}Cl_4O_8$

[a] for $[M + H]^+$ adducts

^{1}H-spectra were acquired on each single, purified compound **1**–**4** (Appendix 1 - Figure 3–6). However, there was insufficient material to obtain strong carbon signals in the 2D NMR spectra, which were instead obtained using the **1-4** mixture (Appendix 1 Figure 7–9). A comparison of the ^{1}H-spectra of the purified compounds with the 2D NMR spectra of the mixture enabled assigning signals from the latter spectra to each congener. This approach was possible thanks to the low complexity of the proton spectra originating from a single aromatic AB system and a few singlets.

When comparing the ^{1}H-spectra of compounds **1**–**4** we observed a downfield shift of low intensity peaks (which turned out to arise from protons at positions 5 and 6 in the aromatic AB system). A similar behaviour was observed for the third aromatic signal at position 4, whose chemical shifts ranged from 7.63 in **1** to 8.6 in **4,** reflecting the electron withdrawing power of the substituent at the vicinal position 3, where the molecules differ (Appendix 1 Figure 10). Specifically, the substituents at position 3 range from CH_3 in **1** to CCl_3 in **4** through a stepwise replacement of each hydrogen with a chlorine atom. The corresponding protons were clearly identifiable in each ^{1}H-spectrum (Appendix 1 – Figure 3 to 6) while the position of the C-13 groups was assigned through its HMBC correlation with H-4 (Appendix 1 – Figure 9). The TOCSY signals observed between CH_3, at 2.60 ppm, with the H-4, at 7.63 ppm, in compound **1** and the chloromethylene, at 4.90 ppm, with the H-4, at 7.92 ppm, in compound **2**, provided an additional confirmation (Appendix 1 – Figure 10). On the contrary, the signals arising from common groups further away from position 3 were isochronous in compounds **1-4**: two methoxy, one methyl, two phenols and a quinone moiety. Interestingly, the HMBC correlations between the quaternary carbon 14 and each of the methyl group at positions 17 and 15 (Appendix 1 – Figure 11) indicated the presence of a lactone ring. The position of the

additional chlorine atom present in all congeners at position 2 was also confirmed by HMBC correlations. The position of the methoxy group on ring A, indicated by the HMBC correlation between C-1 and H-4, was confirmed after reaction with 30% H_2O_2 thus observing by LC-MS the oxidative rupture of the internal quinone affording a degraded compound comprising the AB ring system and the methoxy moiety (data not shown). The observed oxidation products are also consistent with the presence of the quinone moiety in ring C. The ^{1}H and ^{13}C assignments are reported in Table 4.2.

Table 4.2 - ^{13}C and ^{1}H-NMR assignments for compounds 1, 2, 3, 4, 5 and 6.

	1	2	3	4	5	6
Position	δ_C					
1	155.5				155	Nd
2	130	128.2	125.7	nd		Nd
3	140.7	138.9	140.3	136.5	nd	Nd
4	124.8	125.1	123.7	123.2	127.4	Nd
5	135.1	135.2	135.9	136.2	136.1	Nd
6	122.3 - 122.7				122.4	Nd
7	187.8				187.8	Nd
7a	117.4				117.4	Nd
8	150.2				149.9	Nd
9	144				144	Nd
10	123.3					
11	152.4				154	Nd
11a	117.5				117.4	
13	21.3	43.3	67.1	94	169	62.7
14	107				107	Nd
15	23.5				23	23.2
17	52				52	51.8
MeO-C-1	61.8				61.5	61.6
Position	δ_H (m, J Hz)					
4	7.63 (s)	7.92 (s)	8.36 (s)	8.59 (s)	8.34 (s)	7.95 (s)
5	8.13 (d, 8.5)	8.22 (d, 8.5)	8.30 (d, 8.5)	8.33 (d, 8.5)	8.3 (d, 8.5)	8.23 (d, 8.4)
6	8.3 (d, 8.5)	8.35 (d, 8.5)	8.38 (d, 8.5)	8.42 (d, 8.5)	8.41 (d, 8.5)	8.33 (d, 8.4)
13	2.66 (s)	4.92 (s)	7.29 (s)	-	-	5.03 (s)
15	2.02 (s)	2.02 (s)	2.02 (s)	2.02 (s)	2.03 (s)	2.02 (s)
17	3.31 (s)	3.3 (s)	3.3 (s)	3.31 (s)	3.32 (s)	3.30 (s)
MeO-C-1	3.85 (s)	3.89 (s)	3.89 (s)	3.87 (s)	3.95 (s)	3.86 (s)
OH-8 and OH-11	12.28; 12.36	12.24; 12.36	12.22; 12.34	12.24; 12.34	12.36	12.4; 12.3

It should be noted that the fermentation broth of strain ID145698 contained additional molecules that, on the basis of their UV spectra and MS fragmentations, appeared related to the main components **1**–**4**. Among them, compounds **5** and **6** present a carboxylic acid (Appendix 1 – Figure 11–13) and a benzylic alcohol (Appendix 1 – Figure 14–15), respectively, at position 3, suggesting that they originate by formal "hydrolysis" of the halogenated congeners. Indeed, **2**, **3** and **4** converted into **7**, **6** and **5**, respectively, after storage for several weeks in DMSO at 4 ºC, or overnight at 60 ºC (Appendix 1 – Figure 16). In agreement with the literature e(Ogawa, 1965), compound **7** showed *m/z* values consistent with the presence of an aromatic aldehyde. With hindsight, compound **7** was also detected as a minor peak in the fermentation broth (Appendix 1 – Figure 17). Overall, these data suggest that compounds **5**–**7** are likely to be degradation products and not pathway intermediates.

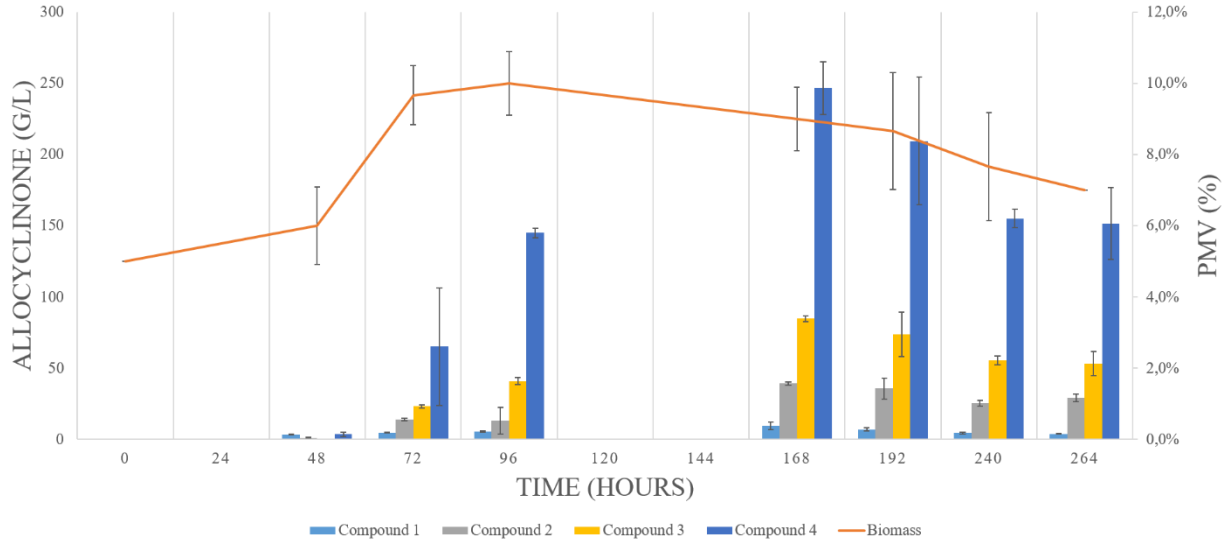


Figure 4.2 - Time course of biomass accumulation, and total angucyclinone production by *Actinoallomurus* sp. ID145698. The concentrations of the individual congeners 1–4 are also indicated. Values correspond to the average measurements of five flasks and the Error bars to the respective standard deviation.

Under our experimental conditions, *Actinoallomurus* sp. ID145698 reached stationary phase after about 72 h, while angucyclinone production started when growth ceased and proceeded at a steady rate up to 168 h, when 380 µg/mL (as the sum of compounds **1** through **4**) were observed. Thereafter, the angucyclinone content declined (Fig. 4.2). In the 72-168 h interval, the relative ratio among the four main congeners remained approximately constant – with **4** always as the main product. This observation suggests that **4** is the end product of the

biosynthetic pathway and that full chlorination of the aromatic methyl at position 3 is not a consequence of accumulation of incompletely chlorinated compounds.

Only a few chlorinated angucyclinones have been described so far: JBIR-88, chlorocyclinones, and marmycin B (Potterat, 2007). Among them, chlorocyclinone D is most related to the compounds described herein, sharing, in addition to the chlorine at position 2 and the methoxy at position 1, a lactone fused to ring D. However, while only one member of the chlorocyclinone family contains such a lactone ring, this moiety is a common feature of compounds **1–4** where it is further decorated with a methoxy group. The most remarkable feature, however, is that chlorination in compounds **1–4** is not limited to the aromatic rings, as found in other chlorinated angucyclinones, but involves also the methyl group at position 3. Indeed, to our knowledge **4** represents the first instance of a natural product containing an aromatic trichloromethylene substituent. Aliphatic trichloromethylene groups are however known, for example in barbamide, produced by a cyanobacterium, and muironolide, isolated from a marine sponge.

Biological Activity. The entire complex (compounds **1–4**) showed MIC values in the range of 0.25–1 µg/mL against all Gram-positive bacteria tested, with the exception of *Enterococcus faecium*, which was about one order of magnitude less sensitive (Table 4.3). As expected, none of the evaluated resistance mechanisms affected the activity of the compounds. It should be noted that there were no measurable MICs at 20 h against the hyper-permeable *E. coli* strain used in the initial screening. However, >50% growth inhibition was observed with 32 µg/mL **1**–**4** (Appendix 1 – Figure 18).

Table 4.3 – Antibacterial activities of angucyclinones from *Actinoallomurus* sp. ID145698

Strain	Strain code	resistance phenotype[a]	MIC (μg/mL)						
			Complex[b]	cpd **1**	cpd **2**	cpd **3**	cpd **4**	VAN[c]	GEN[c]
Staphylococcus aureus	ATCC6538P		0.5	32	2	1	0.25	1	0.5
Staphylococcus aureus	ATCC25922	MR	0.13					1	>64
Staphylococcus aureus	L3797	GI	0.25					16	>64
Streptococcus pyogenes	L49		0.5	4	16	2	0.25	0.5	8
Streptococcus pneumoniae	L44		2					0.5	8
Enterococcus faecalis	L559		0.25					2	4
Enterococcus faecalis	L560	VanA	0.5	64	16	1	0.5	>64	4
Enterococcus faecalis	A3650		1					2	>64
Enterococcus faecium	L568		16					4	16
Enterococcus faecium	L569	VanA	8	>64	>64	8	4	>64	8
Escherichia coli	ATCC25922		>64					>64	0.5
Escherichia coli	L4242		>64					>64	1

a Abbreviations: MR/MS-Methilicin Resistant/Susceptible; VanA/VanS-Vancomycin Resistant/Susceptible; GI- Glycopeptide Intermediate

b : relative complex composition, determined as described in Experimental Section

c 'VAN' and 'GEN' designate vancomycin and gentamicin, respectively.

The purified molecules **1–4** were also evaluated for their antibacterial activities. With one exception, the results indicate a direct relationship between number of chlorines atoms and antibacterial activity, with each additional chlorine contributing to a 2-4 fold decrease in the MIC (Table 4.3). Therefore, **4** is the most abundant and most active compound in the complex. In contrast, compounds **5-7** did not present a noticeable antibacterial activity.

Chlorinated angucyclinones possess different biological activities, ranging from antitumor as for JBIR-88 to modulators of a potential target in diabetes, like chlorocyclinones.[16,17] Further studies will be necessary to evaluate the full potential of compounds **1**–**4** as antibacterial agents or for other therapeutic applications.

Analysis of additional *Actinoallomurus* strains. We next analysed other *Actinoallomurus* strains for the presence of similar compounds. Remarkably, compounds **1**–**4** were a relatively common encounter in *Actinoallomurus*, since we identified eleven additional

independent strains (out of 200 analysed) producing this family of angucyclinones, as indicated by similar HPLC profiles (Fig. 4.3) and confirmed by LC-MS (Appendix 1 – Figure 19–21). The twelve *Actinoallomurus* strains were isolated from three distant geographic areas and belong to three different phylotypes (Table 4.4), with nine and two strains clustering with to the previously described Alp24 and Alp9 phylotype, respectively.(Pozzi et al., 2011) This analysis also identified the new phylotype Alp37 (Table 4.4).

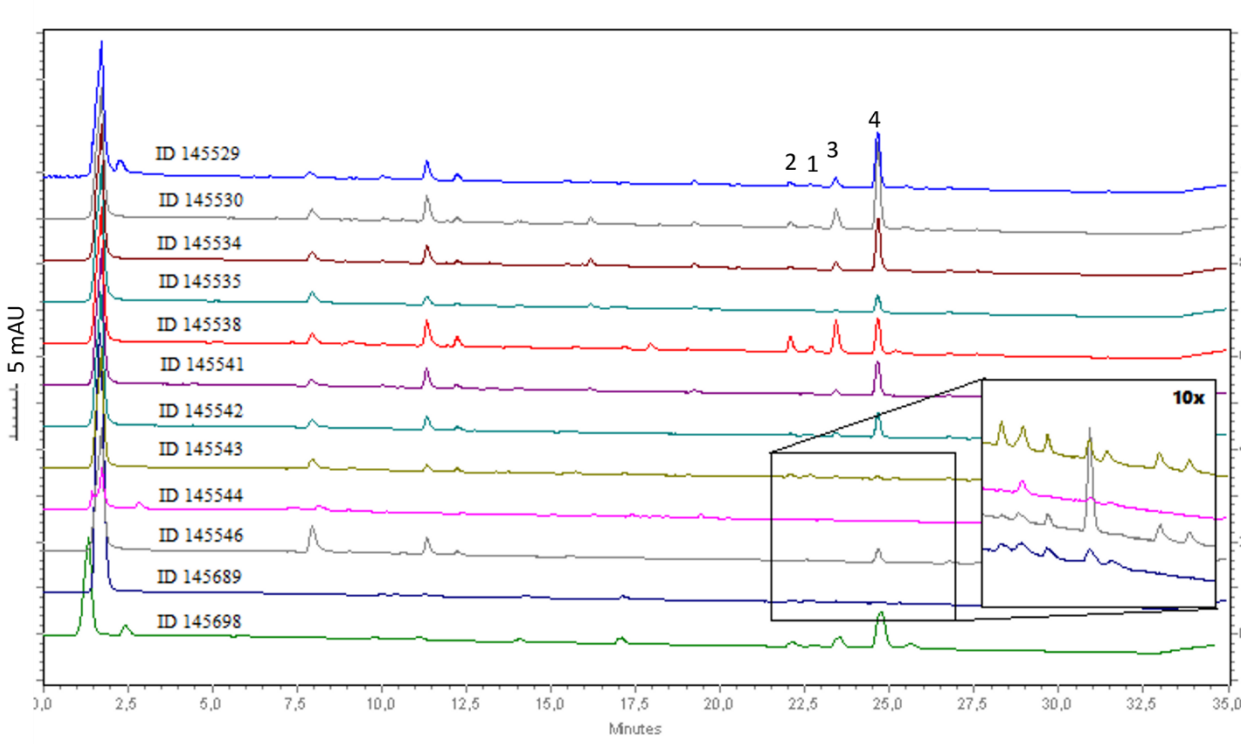


Figure 4.3 - HPLC at 270 nm traces of MeOH extracts from the twelve *Actinoallomurus* isolates identified as producers of allocyclinones (Table 4.4). Note the 10fold enlargement of the chromatograms for the extracts derived from four strains.

When analysed in a single medium and at a single time point, all the *Actinoallomurus* strains produced variable amounts of the complex. However, for all strains, **4** appeared to be the main congener, consistent with the hypothesis that this hyperchlorinated angucyclinone is the end product of the pathway. Overall, the relatively high frequency at which compounds **1**–**4** have been detected (i.e., in 12 out of 232 screened strains) and their production by three phylogenetically distinct strains suggests an important role for these angucyclinones in

Actinoallomurus physiology. We thus designated the compounds allocyclinones because of their association with and possible importance for *Actinoallomurus*. Thus, compounds **1**–**4** would be Allocyclinone A–D, respectively.

Table 4.4 - Allocyclinone-producing *Actinoallomurus* strains.

Strain ID	Country (area) of isolation	Phylotype	Accession No.
145529	Togo (unknown)	Alp37	KR781097
145530	Madagascar (Masoala)	Alp24	KR781088
145534	Italy (Tuscany)	Alp24	KR781089
145535	Italy (Tuscany)	Alp24	KR781090
145538	Italy (Tuscany)	Alp24	KR781091
145541	Italy (Tuscany)	Alp24	KR781092
145542	Italy (Tuscany)	Alp24	KR781093
145543	Italy (Tuscany)	Alp24	KR781094
145544	Italy (Tuscany)	Alp24	KR781095
145546	Italy (Tuscany)	Alp24	KR781096
145689	Madagascar (Masoala)	Alp9	KR781099
145698	Italy (Lombardy)	Alp9	KR781098

Identification of the putative Allocyclinone gene cluster. From a draft genome sequence of *Actinoallomurus* sp. ID145698, we identified a likely candidate for allocyclinone biosynthesis within a 43 kbp genomic segment (Table 4.5). This segment contains 34 Coding DNA Sequences (CDS) possibly involved in allocyclinone biosynthesis, regulation and perhaps resistance. In addition to CDSs for the minimal PKS, the cluster specifies polyketide cyclases, *O*-methyltransferases, hydroxylases, hydrolases, oxidoreductases and halogenases that can account for many of the structural features of allocyclinones (Table 4.5). From this preliminary analysis, the allocyclinone cluster presents two relevant features: the KSβ/Chain Length Factor of the minimal PKS is more related to the dodeca/tridecaketide (e.g. pradimycin) than to the decaketide (e.g. jadomycin) clade of type II PKSs (Fig. 4.4); and the cluster encodes two distinct, closely related halogenases (Table 4.5). These observations suggest that allocyclinones are biosynthesized as a dodecaketide precursor, with the fifth ring cleaved to

ultimately afford the lactone ring (presumably an identical biosynthetic pathway occurs for Chlorocyclinones); and that two distinct halogenases are responsible for chlorination of the aromatic ring and of the methyl moiety. Experimental data will be necessary to corroborate these bioinformatic predictions about the allocyclinone cluster.

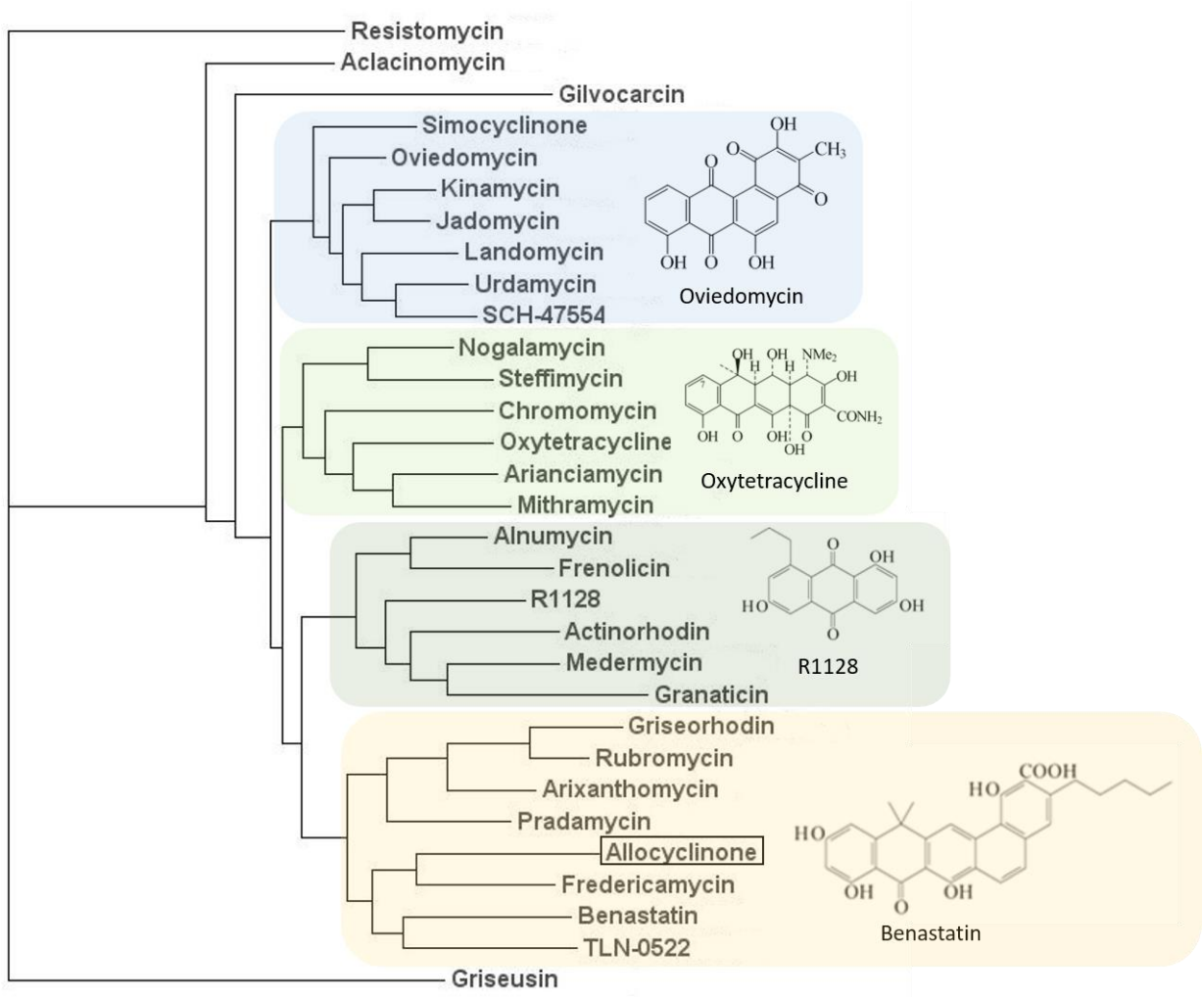


Figure 4.4 - Phylogenetic comparison of KS_{β} enzymes from characterized type II PKS clusters with NAI698_9173 from putative allocyclinone gene cluster (Table 4.5). The structures of one representative aglycone for each clade is also shown. Supporting Table 5.1 provides additional information on the KS_{β}'s used for the phylogenetic analysis.

Table 4.5 – Description of CDSs identified by Antismash v3.0.0.4 and proposed functions of gene products. Dotted line denotes the putative edges of the cluster

CDS	Length (aa)	Putative Function	Best PKS Match	Accession Number - Best PKS Match	Identity
NAI698_9155	369	Hypothetical Protein		CAD89764.1	36%

NAI698_9156	961	Lux family Transcriptional Regulator		AEK75494.1	30%
NAI698_9157	177	Regulator Protein		CAF32368.1	37%
NAI698_9159	218	Putative Cyclase	FR9Q	AIC32703.1	29%
NAI698_9160	301	Putative DNA-binding protein		CAF32370.1	63%
NAI698_9161	289	Putative DNA-binding protein		CAF32370.1	54%
NAI698_9162	172	Oxidoreductase	MmyF	CAC36752.1	44%
NAI698_9163	262	SARP-like transcriptional Regulator		AGH68913.1	45%
NAI698_9164	342	O-methyltransferase	Lcz35	ABX71152.1	47%
NAI698_9165	318	Hypothetical protein		CAI94709.1	33%
NAI698_9166	308	Transcriptional regulator		AHA12085.1	44%
NAI698_9167	294	Putative Hydrolase	PedQ	AAW33976.1	38%
NAI698_9168	211	Unknown Function	NapU4	ABS50471.1	42%
NAI698_9169	108	Polyketide synthesis cyclase		ABW11821.1	60%
NAI698_9170	537	Decarboxylase	SimA12	AEA30254.1	71%
NAI698_9171	138	Putative cyclase	PdmL	ABW11824.1	63%
NAI698_9172	424	Ketoacyl synthase - alpha	XanF	ADE22315.1	70%
NAI698_9173	407	Chain Length Factor	FdmG	ADG86316.1	60%
NAI698_9174	243	Polyketide cyclase	RubF	AAG03070.2	71%
NAI698_9175	67	Hypothetical protein		CAF05638.1	33%
NAI698_9176	245	Acyl Carrier Protein	GrhO10	AAM33668.1	54%
NAI698_9177	112	Monooxygenase	FdmJ	AAQ08920.1	35%
NAI698_9178	108	Monooxygenase	FdmQ	AAQ08928.1	35%
NAI698_9179	393	Oxidoreductases	SHC52	CDF96615.1	36%
NAI698_9180	153	Hydroxylase	GrhM	AAM33665.1	38%
NAI698_9181	583	Acetyl-CoA carboxylase biotin carboxylase		CAD55505.1	66%
NAI698_9182	168	MarR-like transcriptional regulator		AGS77345.1	34%
NAI698_9183	569	Drug resistance transporter		BAJ52670.1	43%
NAI698_9184	255	ACP-reductase	SanS	ADG86311.1	40%
NAI698_9185	276	O-methyltransferase	PieB2	AEZ54383.1	47%
NAI698_9186	555	Halogenase	PrlN	AGC24268.1	53%
NAI698_9187	145	Ester Cyclase	SnoL	ACB47065.1	32%
NAI698_9188	168	Cupin domain protein		CAQ64713.1	35%
NAI698_9189	146	Ester Cyclase	Lct51	ABX71134.1	31%
NAI698_9190	435	Membrane antiporter		ADU86006.1	61%
NAI698_9191	560	Halogenase	CtcP	AEI98659.1	52%
NAI698_9192	147	Hypothetical protein		AEH42476.1	29%
NAI698_9193	530	Monoxygenase	RdmE	AAA83424.1	45%
NAI698_9194	590	Peptidase		AAC26133.1	29%

4.4. Discussion

This work describes the first hyper-halogenated angucyclinone. They represent an additional example of halogenated metabolites produced by *Actinoallomurus*. Indeed, this genus has been previously reported to produce spirotetronates with multiple halogens in the pyrrole moiety and a lantibiotic with a halogenated tryptophan residue (Cruz, 2015; Mazzetti et al., 2012). While the presence of a third chlorine atom in the pyrrole moiety resulted in a lower antibacterial activity in the spirotetronate NAI-414, allocyclinones almost present a linear relationship between antibacterial activity and number of chlorine substituents on the methyl group. The increased number of chlorines results in progressively higher molecular weight and is expected to result in higher lipophilicity. The most halogenated compounds **3** and **4** possess the highest antibacterial activity (in µg/ml), which suggests that lipophilicity is important for antibacterial potency. However, compounds **1** and **2** show an inversion in their HPLC retention times and also in their relative potency against *Streptococcus pyogenes* (but not with the other strains), suggesting the interplay of different factors in antibacterial potency and HPLC retention times.

Other to the compounds **1-4** that were analysed in more detail, additional molecules were **5**, **6**, and **7**, all differing for the substituent at position 3 (Figure 4.1): a carboxylic acid (**5**), an alcohol (**6**), or an aldehyde (**7**). Above we presented evidence that these compounds can be formed by conversion of the halogenated counterparts **4**, **2** and **3**, respectively, which suggest they are not biosynthetic intermediates but form as the results of formal hydrolysis of the chlorinated metabolites. Indeed, the methyl at position 3 is likely to derive from the acetate starter unit of the polyketide chain, and to be subsequently converted into **2**, **3** and **4** by the progressive action of halogenase(s). In this respect, it is tempting to speculate that one of the two halogenases specified by the allocyclinone BGC is involved in halogenation of the aromatic ring, while the other modifies the methyl group.

Allocyclinones, although previously unreported, might not be uncommon metabolites within the genus *Actinoallomurus*, since we identified several producer strains. A retrospective analysis by Baltz highlighted the correlation between the frequency at which different antibiotics were encountered during screening programs and the history of antibiotic discovery during the so called 'golden era' (Baltz, 2007). It is thus not surprising that, when systematically evaluating a poorly explored taxon such as *Actinoallomurus*, the first identified metabolites

would represent the 'low-hanging fruits' for this genus. Indeed, the most abundant and most active congener, allocyclinone D, fits well within this definition: it is produced by different phylotypes (or by a frequent phylotype such as Alp24); it is produced in relatively high amounts by different isolates (>100 µg/mL); and it has good antimicrobial activity (MIC < 1 µg/mL against *S. aureus*). Compounds like this are thus hard to miss in a screening program. Nonetheless, the low-hanging fruits within an unexplored taxonomic groups may still provide interesting structural features and bioactivities.

This work was halted with the discovery of the molecular structure of allocyclinone and the partial characterization of the organism. Although it could have been interesting to deepen the characterization of both, the overall goal of this PhD project is to deliver new molecular structures that have the potential to become leads in a drug discovery pipeline. The angucyclinone family does not fall in this category. Angucyclinones are a very large family of compounds with a broad range of activities, however it is also true that none of the members has succeeded as an antibacterial agent, with lipophilicity and cytotoxicity as major hurdles. Moreover, completing the whole characterization of this molecules required time and techniques that went beyond the scope of this project.

5. A BROMINATED VARIANT OF THE LANTIBIOTIC NAI-107 WITH ENHANCED ANTIBACTERIAL POTENCY

5.1. Introduction

The threat posed by the increasing prevalence of antibiotic resistant bacterial pathogens is prompting a renewed interest in additional classes of antibacterial agents not affected by prevailing resistance mechanisms. One class that is attracting increasing attention is represented by the lantibiotics. These ribosomally-synthesized peptides, characterized by a variable number of lanthionine residues, bind to lipid II, a key intermediate in peptidoglycan biosynthesis that is also targeted by the glycopeptides (Arnison et al., 2013). While some bacterial pathogens can produce alternative lipid II intermediates that escape glycopeptide action, they remain equally sensitive to lantibiotics (Schneider and Sahl, 2010). This feature, and the increasing number of lantibiotics discovered by bioassay-based screening or genome mining, are among the driving forces behind this renewed interest in this class of antibacterial agents (Arnison et al., 2013; Zhang et al., 2015).

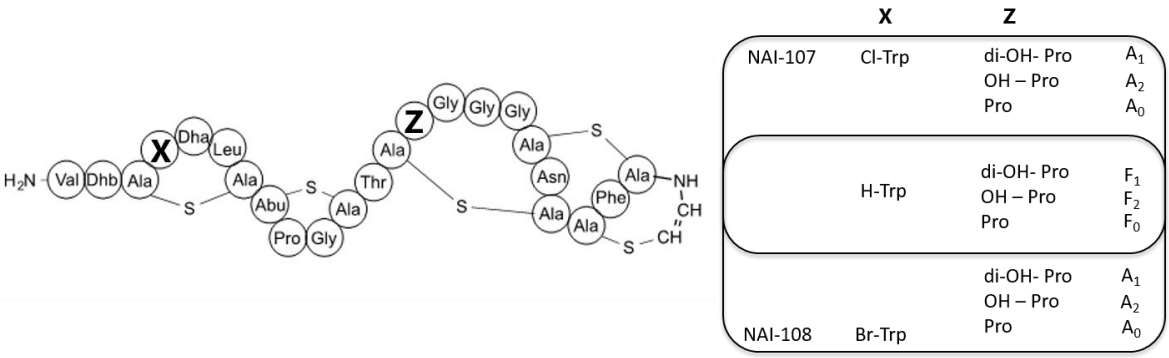


Figure 5.1 - Structure of NAI-107 and NAI-108 and of the corresponding major congeners.

The lantibiotic NAI-107 (Fig. 5.1) is among the most potent lantibiotics reported so far. It belongs to the same class as nisin, the most studied lantibiotic, with which it shares the

two N-terminal lanthionine rings. NAI-107 can form complexes with bactoprenol-pyrophosphate-coupled precursors of the bacterial cell wall, including lipid II, and interferes with membrane function, (Munch et al., 2014) consistent with its high potency and rapid bactericidal activity. These features, together with sustained concentrations observed in plasma, explain NAI-107's efficacy in different animal models of infection (Jabes et al., 2011; Lepak et al., 2015) Unusual among lantibiotics, NAI-107 is produced by *Microbispora* sp. ID107891 as a complex of related molecules, differing for the extent of hydroxylation at proline residue 14 (zero, one or two hydroxyl groups can be present) and for the presence of chlorine or hydrogen on tryptophan residue 4 (Fig. 5.1)(Maffioli et al., 2014).

This chapter contains a manuscript, published on *J Nat Prod* (DOI: 10.1021/acs.jnatprod.5b00576), the manuscript reports the substitution of chlorine with bromine in the lantibiotic NAI-107, which is a promising drug candidate for Gram-positive infections. This substitution was effectively performed firstly by a recently discovered new producer of NAI-107, *Actinoallomurus* strain ID145699 that was originated from the screening exercises performed in this project, and later by the unrelated actinomycete *Microbispora* sp. ID107891. In addition to representing the first example of a brominated lanthipeptide, the brominated NAI-107 variant shows a small but consistent increase in potency against bacterial pathogens, which are killed at equivalent rates as the parent lantibiotic but at lower concentrations. The brominated lantibiotic reported in the manuscript thus further expands chemical diversity in this family of lantibiotics and provide further insights into structural features that make these compounds highly potent against bacterial pathogens.

5.2. Experimental Section

General experimental procedures. For monitoring metabolites production analytical HPLC was performed on Shimadzu Series 10 spectrophotometer (Kyoto, Japan), equipped with a reverse-phase column, Lichrospher RP-18, 5µm, 4.6 x 125 mm (Merck, Darmstadt, Germany), phase A as 0.1% trifluoroacetic acid (TFA), phase B as acetonitrile, and the flow rate was 1 mL/min. The program consisted of a linear gradient from 10% to 36% phase B in 5 min; from 36% to 50% phase B in 7 min; and from 50% to 80% phase B in 1 min; followed by a 4-min isocratic step at 80% phase B and column re-equilibration. UV detection was at 230 nm.

LC-MS analyses were performed with a Dionex UltiMate 3000 liquid chromatographer equipped with an Atlantis T3 RP18, 3 µm, 50 x 4.6 mm column eluted at 0.8 mL/min at 40 °C. Elution was with a 11 min multistep program that consisted of 10%, 10%, 95%, 95%, and 10% phase B at 0, 1, 7 and 9 min, respectively [phase A was 0.05% TFA (*v/v*) in H_2O and phase B was CH_3CN]. The effluent from the column was split 1:1 into a photodiode array detector and into the ESI interface of a Thermo scientific LCQ Fleet mass spectrometer with a tridimensional ion trap in the positive ionization mode.

^{1}H-NMR spectra were measured in CD_3CN-D_2O at 298K using an AMX 400 MHz spectrometer. Chemical shifts are reported relative to D_2O (δ 4.79 ppm).

The concentrations of NAI-107 and of NAI-108 in microbial fermentation extracts or partially purified fractions were established by measuring areas of relevant HPLC peaks in comparison with an internal standard of pure NAI-107 obtained from *Microbispora* sp. ID107891 as described (Monciardini, P., *et al*. 2015).

Actinomycete strains and media. *Actinoallomurus* strain ID145699, which was isolated from a soil sample collected in Northern Italy (GPS: 45.633207, 8.964064), is from the NAICONS strain library. It was propagated on S1-5.5 plates at 30 °C. A loopful of mycelium was used to inoculate AF-A medium (10 g/L dextrose monohydrate, 4 g/L soybean meal, 1 g/L yeast extract, 0.5 g/L NaCl, 1.5 g/L 2-(N-morpholino)ethanesulfonic acid, pH adjusted to 5.6) in shake-flasks. After 8 days in a rotatory shaker (200 rpm) at 30 °C, cultures were harvested and stored at -80 ºC. Subsequent inocula were made using a 10% (v/v) inoculum from frozen stocks. When appropriate, NaCl was substituted with 1.03 g/L KBr, 1.4 g/L KI or 0.5 g/L KF in AF-A medium. It should be noted that yeast extract and soybean meal provide chloride at a calculated concentration of 19 mg/L, whereas a 10% inoculum from AF-A medium is expected to contribute further 30 mg/L of chloride. PCR amplification of the 16S rRNA gene sequence was performed as described and the resulting sequence from *Actinoallomurus* sp. ID145699 has been deposited under accession number KT225504.

Microbispora sp. ID107891 was maintained and grown as described previously(Monciardini, 2015). The GE82G10 medium used for *Microbispora* sp. ID107891 consisted of: 10 g/L dextrose monohydrate, 20 g/L maize dextrin, 15 g/L soymeal, 5 g/L yeast extract, 2 g/L NaCl, 1 g/L $CaCO_3$, 1/g L agar, 0.15 ml/L polypropylene glycol P2000, pH 7.3; when appropriate, NaCl was substituted with 2 g/L KBr.

Production and recovery of NAI-108. After a 5-days cultivation of *Microbispora* sp. ID107891 in KBr-supplemented GE82G10 medium, a 1.6-L culture (containing approximately

200 mg lantibiotic) was acidified to pH 3 with sulfuric acid and paper-filtered. The collected material (0.5-L) was treated with one volume EtOH-AcOH 92:8 *v/v* and shaken overnight at room temperature. This procedure was repeated twice. After centrifugation (3200 rcf, 10 min), the combined extracts (2-L) contained 113 mg lantibiotic. After evaporation under vacuum, the residue was resuspended in 5 mL H_2O-dimethylformamide 1:1 and purified by reverse-phase medium pressure liquid chromatography in a Combiflash system (Teledyne ISCO, Lincoln, NE, USA) using an RP18 60 g column, a 30-min linear gradient of 0.01 M AcOH - acetonitrile (from 20% to 90% of eluent B) at 30 mL/min with a 230-nm detection wavelength. Fractions eluting at 50% acetonitrile were combined and, after removing solvent under vacuum, afforded 265 mg of a solid residue containing 21% total lantibiotic, as established by HPLC. The residue was suspended in 25 mL H_2O-EtOH (2:3) and, after adjusting the pH to 3.5 with hydrochloric acid, it was resolved by ion exchange chromatography using 40 mL of activated Toyopearl SP 550C resin in a column (3x40 cm column) previously equilibrated with five bed volumes (BV) of 50% EtOH at 5 mL/min. The resin was washed with 8 BV of 0.04 M $AcONH_4$-EtOH 1:1 (adjusted to pH 5.2 with AcOH) and the lantibiotic was eluted using 20 BV of 0.05 M NH_4Cl-EtOH 1:1 (adjusted to pH 3.5 with diluted hydrochloric acid). Lantibiotic-containing fractions were pooled and, after drying under vacuum, the residue was washed twice with water and once with acetone. The product was recovered by centrifugation at 1800 rcf for 10 min and dried at 40 °C overnight under vacuum, yielding 30 mg of beige solid.

Antibacterial assays. MICs were determined as previously described (Monciardini, P., *et al*. 2015), using media supplemented with 0.02% bovine serum albumin. Time-kill assays were performed using 10 mL of *Staphylococcus aureus* ATCC 29213 (or *Enterococcus faecium* L568) exponential cultures in Müller-Hinton broth. Antibiotics were added as 0.1 mL solutions in 10% DMSO at appropriate concentrations to achieve the desired final concentration values after a 100 fold dilution. Controls were made adding an equivalent amount of 10% DMSO to the culture. At defined time points, aliquots were withdrawn and the number of viable colonies was determined by placing 0.1 mL from serial dilutions onto Müller-Hinton agar plates. Plates were incubated at 37 °C. All strains were either from the ATCC or from the NAICONS collection (the latter indicated by L- or ND- prefixes). The profile of both NAI-107 and Nai-108 complex are described in Table 6.1.

5.3. Results

Discovery of NAI-107 from *Actinoallomurus*. While screening 200 *Actinoallomurus* isolates for antimicrobial activity, an extract from strain ID145699 inhibited growth of Gram-positive bacteria. HPLC fractionation of a mycelium extract led to the identification of the active fractions (Fig. 5.2), whose UV spectra, masses and lack of fragmentation corresponded to those of NAI-107 purified from *Microbispora* sp. ID107891 (Appendix 2 - Figure 1) (Maffioli et al., 2014). Under the tested conditions, strain ID145699 produced approximately 5 µg/ml NAI-107.

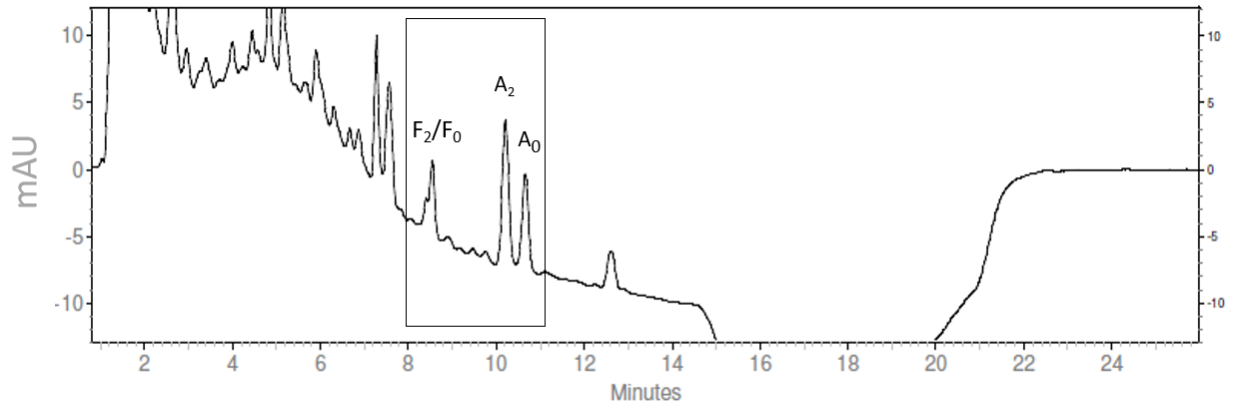


Figure 5.2 - HPLC profile at 230 nm of *Actinoallomurus* sp. ID145699. The boxed area highlights the HPLC fractions that inhibited bacterial growth with the main NAI-107 congeners labelled. Under these chromatographic conditions, F_2 and F_0 are only partially resolved.

On the basis of its 16S rRNA gene sequence, strain ID145699 was confirmed to belong to the genus *Actinoallomurus*. The finding of a NAI-107 producer among a few *Actinoallomurus* strains was somehow unexpected: this lantibiotic emerged after screening 120,000 microbial fermentation extracts derived from 40,000 actinomycetes(Monciardini et al., 2014), and has been identified only as a product of two strains belonging to the genus *Microbispora* (Castiglione et al., 2008; Foulston and Bibb, 2011). It should be noted that the genera *Actinoallomurus* and *Microbispora* belong to the distinct families: *Thermomonosporaceae* and *Streptosporangeacea* within the order *Streptosporangiales*. Within the *Firmicutes*, it has been reported that the same lantibiotic is generally produced by the same bacterial species or by members of the same genus but this general rule might not apply to actinomycete-derived lantibiotics. While this manuscript was in preparation, Carrano and co-workers reported NAI-107 production by an apparently independent *Actinoallomurus* isolate (Carrano et al., 2015).

Actinoallomurus strain ID145699 produced NAI-107 as a complex of the A_0 and A_2 congeners, with m/z^+ of 1108 and 1116, respectively, and of their dechlorinated counterparts F_0 and F_2, with m/z^+ of 1091 and 1099, all corresponding to their double-charged ions $[M+2H]^{2+}$ (Appendix 2 – Figure 1). Congeners A_2 and F_2 carry a mono-hydroxylated Pro14 residue (Fig. 5.1). Under the examined conditions, only traces were detected of A_1, carrying a di-hydroxylated Pro14 residue, and of the corresponding dehalogenated congener F_1. In terms of proline hydroxylation, the complex produced by the *Actinoallomurus* strain ID145699 thus appears to be intermediate between the complex produced by *Microbispora corallina* NRRL30420, which yields the non-hydroxylated A_0 as the main peak, and that of *Microbispora* sp. ID107891, which produces the mono- and di-hydroxylated Pro14 variants in equivalent amounts, with only traces of A_0 (Maffioli et al., 2014).

Incorporation of bromine into NAI-107 by *Actinoallomurus*. Previous work had indicated that *Actinoallomurus* frequently produces chlorine-containing metabolites (Mazzetti et al., 2012; Pozzi et al., 2011), including hyper-chlorinated angucyclinones (Cruz, 2015). Furthermore, we were able to show an effective substitution of two chlorine atoms with bromine in the pyrrole moiety of NAI-414, a recently described spirotetronate from *Actinoallomurus* (Mazzetti et al., 2012).These results prompted us to investigate whether strain ID145699 could also effectively incorporate other halogens into NAI-107.

The medium used for NAI-107 production (AF-A; see 5.2 Experimental Section) is a complex medium containing 0.5 g/L (8.6 mM) NaCl. Thus, experiments were carried using the same medium but substituting NaCl with KBr at an identical molar concentration. The addition of 8.6 mM KBr to the medium readily resulted in the formation of new peaks having retention times similar to those of the NAI-107 complex but slightly more lipophilic (Fig. 5.3A). The m/z^+ values of the corresponding molecules, 46 amu higher than the chlorinated congeners, were consistent with the replacement of the single chlorine atom by a bromine in the A_0 and A_2 congeners (Appendix 2 – Figure 1). Under these conditions, strain ID145699 produced around 7 µg/ml of brominated lantibiotic consisting mainly of the A_2 and A_0 congeners in a 3:2 ratio.

One of the observed consequences of adding bromide to the medium was the appearance of additional compounds (Fig. 5.3A) whose m/z^+ ratios were consistent with N-terminally extended variants of A_0 and A_2 carrying a Gly-Pro-Ala tripeptide preceding the N-terminal Val in mature NAI-107 (Fig. 5.1). These N-terminal extensions were previously detected in the non-hydroxylated NAI-107 congeners when *Microbispora* sp. ID107891 was grown in

minimal medium (Maffioli et al., 2014). It should be noted that, with an increase of the F congeners to approximately 1.5 µg/ml, we were occasionally able to detect small amounts of the GPA-extended variants of congeners F_0 and F_2 in the KBr-fed culture of *Actinoallomurus* sp. ID145699 and of *Microbispora* ID107891 (Appendix 2 – Figure 1).

Thus, at least the last three C-terminal amino acids in the leader peptide must be identical between the *Actinoallomurus* and the *Microbispora* producers. Furthermore, these results suggest that abnormal processing of the precursor peptide occurs at the same secondary location in two different *Actinobacteria* genera, despite the fact that the NAI-107 gene cluster does not encode a recognizable precursor-processing protease (Foulston and Bibb, 2010; Sosio et al., 2014). Indeed the lack of a protease LanP seems to be a common characteristic of clusters responsible for the biosynthesis of other lanthipetides, which can be heterologously expressed and processed by apparently promiscuous host proteases present in different *Actinobacteria* (Krawczyk et al., 2013).

We also investigated the effects of adding either KF or KI to the medium. Although no significant changes in biomass or culture pH were seen under these conditions, we could only detect traces of A and F congeners, and no evidence for incorporation of other halogens into the lantibiotic (data not shown). The presence of a small amount of chlorinated congeners might result from a carryover from the inoculum grown in medium AF-A, which contains NaCl or from residual chloride in medium components (see 5.2 Experimental Section). The lack of fluoride or iodide incorporation is consistent with the properties expected from halogenases which can readily use supplemented bromide but not fluoride or iodide (van Pee, 2012). Nonetheless, the overall low yield of lantibiotic in KF- and KI-supplemented media suggests that halogenation is necessary to trigger full production of NAI-107 in *Actinoallomurus* ID145699.

Incorporation of bromine into NAI-107 by *Microbispora*. We next wondered whether *Microbispora* sp. ID107891, which is a better characterized NAI-107 producer than *Actinoallomurus* sp. ID145699 (Maffioli et al., 2014), was also able to produce a Br-containing lantibiotic. When grown under the same conditions (i.e., AF-A medium with KBr substituting NaCl), *Microbispora* sp. ID107891 readily produced brominated compounds, even if overall titers were one third that in NaCl-supplemented medium (Fig. 5.3B). Nonetheless, strain ID107891 produced more Br-containing congeners than strain ID145699 (compare blue line in Fig. 5.3 A and B). Thus, once we achieved proof of concept that Br incorporation into NAI-107 was not confined to *Actinoallomurus*, we resorted to the better-characterized *Microbispora*

sp. ID107891 to obtain sufficient amounts of brominated lantibiotic for further characterization. It should be noted that in the AF-A based media *Microbispora* sp. ID107891 produced significant amounts of GPA-extended congeners (Fig. 5.3B).

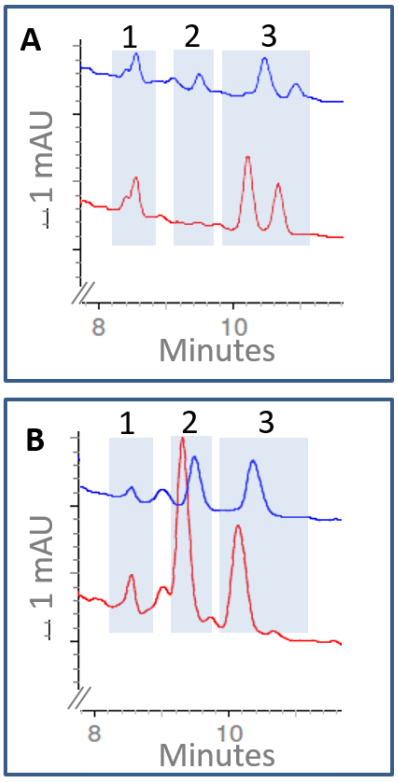


Figure 5.3 - Incorporation of Br into the lantibiotic. Selected portion of the HPLC profile of extracts from *Actinoallomurus* sp. ID145699 (panel A) or from *Microbispora* sp. ID107891 (panel B) grown in medium AF-A (red line) or AF-A containing KBr instead of NaCl (blue line). Numbers 1, 2 and 3 over the shaded areas denote the F, GPA-extended and A congeners, respectively. Note the differences in retention times of peaks *2* and *3* between red- and blue-line chromatograms, due to the different halogens present in the lantibiotic. Note also that the F congeners are F_2 and F_0 in panel A, and F_1 and F_2 in panel B, whereas the A congeners are A_2 and A_0 in panel A, and A_1 and A_2 in panel B.

In a suitable medium with acceptable concentrations of Cl-containing ingredients, supplemented with 2 g/L KBr (see 5.3. Experimental Section), *Microbispora* sp. ID107891 produced a complex consisting mostly of brominated congeners. The overall production after a 5-day cultivation was around 50 mg/L, comparable to the levels observed in KBr-free medium. As seen for the production of chlorinated lantibiotics (Monciardini, P., Unpublished experiments), addition of 1 g/L ammonium sulphate allowed a 3-fold increase in the titer of brominated molecules (data not shown).

Characterization of the brominated lantibiotic. The lantibiotic was recovered from a 2.4-L culture of KBr-supplemented medium (see 5.3. Experimental Section), leading to about 30 mg of purified complex with a relative congener distribution summarized in Table 6.1. The presence of a Br-Trp residue was confirmed by ^{1}H-NMR analysis that highlighted differences in the chemical shifts of protons belonging to the indole ring (Appendix 2 – Figure 2). Among them, the aromatic proton at position 4 was particularly diagnostic: the resonance at 7.35 ppm in the chlorinated congeners was shifted to 7.5 ppm in the brominated congeners, in agreement with the values expected for halogenated aromatic systems (Spiesecke and Schneider, 1961). All other chemical shifts were superimposable with those of NAI-107 (Appendix 2 – Figure 2). While the chemical shifts derived from the Pro14 residue fall in a complex region of the spectrum (Maffioli et al., 2014), the overall results are consistent with the presence of 4-OH or 3,4-di-OH Pro14 residue.

Table 5.1 - Relative congener distribution in the NAI-107 and NAI-108 batches purified from *Microbispora sp.* ID107891. [a]

Congener	RT	amu	Cl (%) [b]	Br (%) [c]
F_1	8.5	2232	<1	<1
F_2	8.5	2194	<1	<1
F_0	8.0	2178	<1	<1
NAI-107 A_1	10.0	2244	58.7	6.3
NAI-107 A_2	10.0	2228	39.0	6,6
NAI-107 A_0	10.5	2212	-	2.3
NAI-108 A_1	10.25	2290	-	34.7
NAI-108 A_2	10.25	2274	-	34.3
NAI-108 A_0	10.75	2262	-	6.3

[a] Established by integrating UV-absorbing HPLC peaks. [b] Internal NAI-107 reference standard. [c] Purified NAI-108 batch as described in this work.

Thus, these experiments confirmed that the Br-containing lantibiotic is identical to NAI-107 except for the substitution of 5-Cl- with 5-Br-Trp and is produced as a complex of closely related congeners, differing for the presence or absence of halogen at Trp4 and for the presence of zero, one or two hydroxyl groups at Pro14 (Maffioli et al., 2014). Further variation is brought by the – presumably spontaneous – sulfur oxidation of the first lanthionine (Fig. 5.1). For the sake of simplicity, we refer to the batches of lantibiotic consisting mostly of Cl- and Br-containing metabolites as "NAI-107" and "NAI-108", respectively. This artificial distinction allows maintaining the same nomenclature for the brominated variants as for the NAI-107 congeners (Monciardini, P., *et al.* 2015). Accordingly, NAI-108 A_1, A_2 and A_0 represent compounds carrying two, one and zero hydroxyl groups, respectively, at Pro14. Obviously, the F congeners, which are devoid of halogens, are identical minor congeners of both NAI-107 and NAI-108 preparations (Fig. 5.1).

Antibacterial activity of the brominated lantibiotic. The newly prepared NAI-108 batch consisted mostly of A_1 and A_2, and contained lower amounts of chlorinated or non-halogenated congeners (Table 5.1). We next investigated whether the presence of a Br-Trp residue had any impact on antibacterial activity. The results are reported in Table 5.2. Interestingly, on the basis of its minimum inhibitory concentrations (MICs), NAI-108 turned out to be about twice as active as NAI-107 against *Staphylococcus*, *Streptococcus* and *Enterococcus* spp.

Table 5.2 – Antibacterial activities of NAI-107 and NAI-108

Strain	Resistance type[a]	MIC (µg/mL) NAI-108	NAI-107	VAN[a]
Staphylococcus aureus ATCC2913	MSSA	0.25	0.5	0.5
Staphylococcus aureus L100	MSSA	≤0.06	≤0.06	0.5
Staphylococcus aureus ND523	MRSA	0.25	0.5	n.d.
Staphylococcus aureus ND605	MRSA	0.125	0.25	1
Staphylococcus aureus L1400	MRSA	0.5	0.5	2
Staphylococcus aureus L3797	GISA	2	4	8
Staphylococcus aureus L3798	GISA MRSA	2	2	16
Streptococcus pneumoniae ATCC49619		0.004	0.004	0.25
Streptococcus pneumoniae L44		0.03	0.06	0.5
Streptococcus pyogenes L49		0.001	0.002	0.5
Enterococcus faecalis L559		2	2	1
Enterococcus faecalis L560	VanA	0.25	1	>64
Enterococcus faecium L568		4	8	2
Enterococcus faecium L569	VanA	2	4	>64

[a] Abbreviations: MRSA/MSSA/GISA, methicillin-resistant/ methicillin-susceptible/ glycopeptide-intermediate *Staph. aureus*; VanA, vancomycin resistant; VAN, vancomycin. n.d., not determined.

One important feature of NAI-107 is its rapid *in vitro* bactericidal activity, which results in >99.9% killing of an exponential culture within 30 min after addition of the compound at 1xMIC (Munch et al., 2014). We thus investigated whether this effect could also be observed with the more potent variant NAI-108. Fig. 5.4A reports the results of an experiment in which the killing activity of identical concentrations of NAI-107 and NAI-108 was evaluated against *Staphylococcus aureus* ATCC29213. At the lowest concentration tested (1xMIC, or 0.25 µg/ml), NAI-108 rapidly killed the strain, leading to a 5-log reduction in viable cells within 30 min. Conversely, the same concentration of NAI-107 (equivalent to 0.5xMIC) resulted in only a 3-log reduction in viable counts, followed by strain growth at a rate

comparable to that in the absence of antibiotic (Fig. 5.4A). As expected, higher lantibiotic concentrations resulted in an equivalent extent of killing for both lantibiotics.

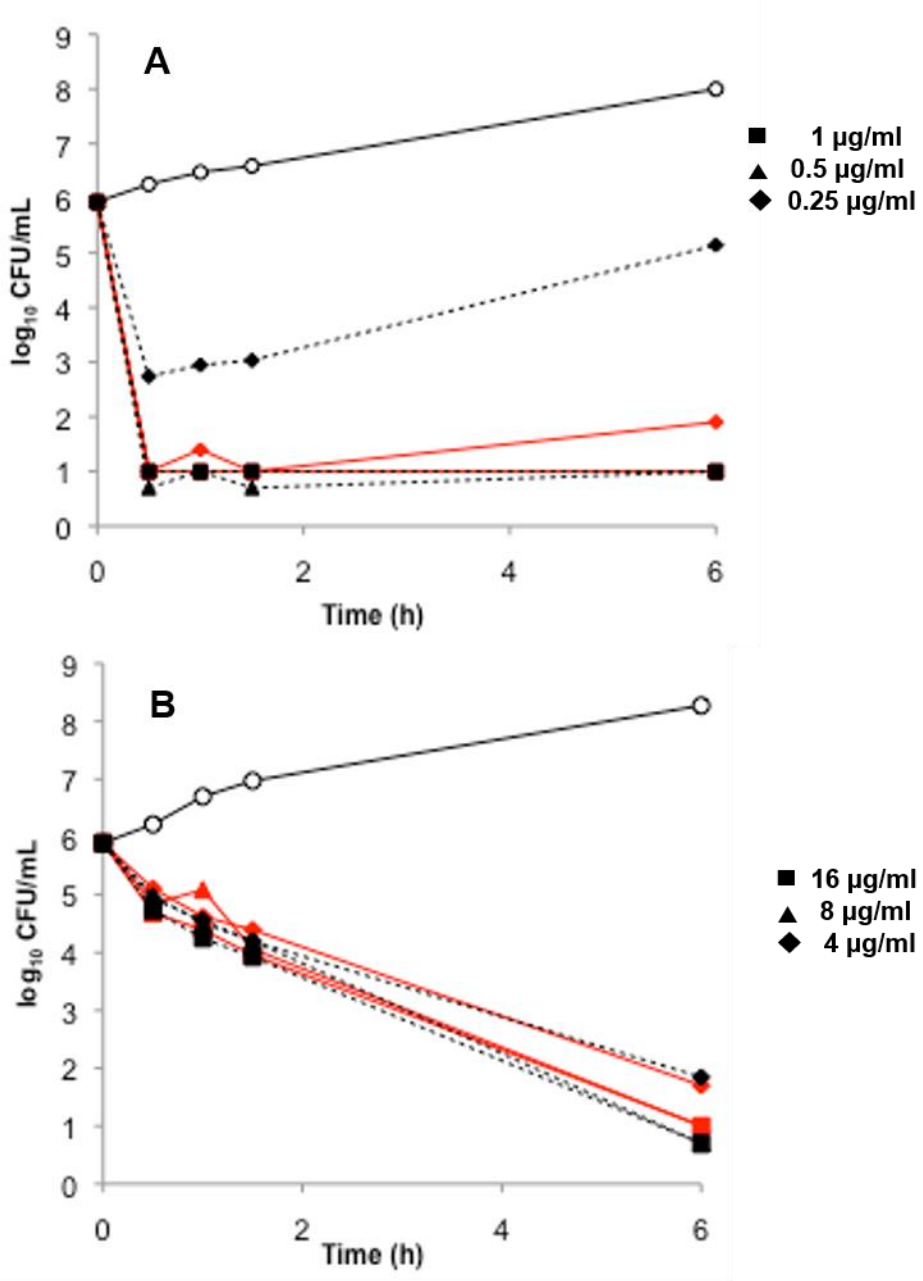


Figure 5.4 - Time-kill assays. Viability of an exponential culture of *Staph. aureus* ATCC 29213 (panel A) or *E. faecium* L568 (Panel B) exposed to NAI-107 (dashed line) or NAI-108 (red line) at the indicated concentrations. Empty symbols with solid black line refers to untreated controls.

When an equivalent experiment was performed with *Enterococcus faecium* (Fig. 5.4B), superimposable results were seen with identical concentrations of both lantibiotics. With this strain, killing occurred at a lower rate and 6 h were necessary to achieve a 5-log reduction in

viable counts, which required ≥8 µg/ml of each lantibiotic, equivalent to 2x and 1xMIC for NAI-108 and NAI-107, respectively. Interestingly, a significant killing effect of NAI-107 was also observed at 0.5xMIC, suggesting that, as with *Staph. aureus* (Fig. 5.4A), exponentially growing cultures are more sensitive to NAI-107 than the stationary phase cultures used for standard MIC determination.

In conclusion, the data reported indicate that NAI-108 is slightly but consistently more active than NAI-107. Furthermore, NAI-108 shows time-kill activity against two different bacterial pathogens at 1xMIC, as does NAI-107, suggesting the possibility of using lower therapeutic doses of the former compound.

5.4. Discussion

Variations in type I lantibiotics have been found within *Actinobacteria*. Indeed, strains belonging to the genera *Streptomyces*, *Streptosporangium* and *Planomonospora* have been recently reported to produce lantibiotic variants with one to three amino acid differences (Maffioli et al., 2015). These molecules, designated as "97-like", share the same amino acid length and lanthionine rings with NAI-107, but possess lower antibacterial activity than NAI-107, with interesting MICs observed only with semi-synthetic derivatives. In one case, the same lantibiotic was identified as the product of two strains belonging to the unrelated genera *Planomonospora* and *Streptomyces* (Maffioli et al., 2015). It should be noted that the 97-like lantibiotics present just the minimal set of post-translational modifications (i.e., dehydration and lanthionine formation) required for this family of RiPPs (Arnison et al. 2013), whereas NAI-107 undergoes additional modifications (i.e halogenation of Trp4, hydroxylation of Pro14 and decaboxylation of the C-terminal Cys). Thus, the finding that a lantibiotic with the same amino acid sequence and the same extensive post-translational modifications is produced by microorganisms belonging to the unrelated genera *Microbispora* and *Actinoallomurus* is *per se* an interesting event, which casts further shadow on the early assumption that lantibiotics are species-specific compounds. The occurrence of NAI-107 as a product of *Microbispora* and *Actinoallomurus* can result from one of two possible explanations: either the corresponding gene cluster was recently acquired by horizontal gene transfer, or NAI-107 offers little room for amino acid substitutions while maintaining all the post-translational modifications, and the corresponding gene cluster has been transmitted vertically within a common ancestor of the *Streptosporangiales* but lost in most of the progeny. Personally, I am more inclined to the former possibility, since variations in amino acid sequence of the 97-like molecules have been observed and it is likely that the further post-translational modifications of NAI-107 are the result of later gene acquisition events.

Several examples have been reported in the literature in which, under appropriate conditions, Br can substitute Cl in microbial metabolites. This feature relates to the higher reactivity of Br^- than Cl^- ions and is consistent with the mechanism of halogenating enzymes (van Pee, 2012). To our knowledge, NAI-107 represents the only reported example of a chlorinated lantibiotic (Arnison et al., 2013) and, hence, the work reported here represents the first description of a Br-containing lantibiotic, suggesting that the Trp halogenases specified by the *mib* clusters from *Microbispora* (Foulston and Bibb, 2010; Sosio et al., 2014) and by

the corresponding cluster from *Actinoallomurus* sp. ID145699 are also capable of efficiently incorporating Br when exposed to sufficient bromide concentrations.

It should be noted that most Br-containing microbial metabolites are usually less active than the corresponding chlorinated congeners with, sometimes, significant drops in antibacterial activity (Ajisaka et al., 1969; Bister et al., 2003; Castiglione et al., 2008; Hochlowski et al., 1997). There are also occasional reports of brominated metabolites showing higher antimicrobial activity (Ezaki et al., 1983) or higher toxicity toward tumour cells (Ogawa et al., 1989) than their chlorinated analogues. NAI-108 is slightly but consistently more active than NAI-107.

It has been previously shown that the F congeners, which contain a naked Trp4 residue, are less active than the chlorinated A congeners and that the former are more hydrophilic than the latter, suggesting that the hydrophobicity at Trp4 influences antibacterial potency (Maffioli et al., 2014). This observation is further corroborated by the results presented here, since substitution of Cl with Br is expected to enhance the hydrophobicity of this lantibiotic, consistent with the slightly higher retention time of the brominated versus chlorinated congeners (Table 5.1). It is worth mentioning that the peptide backbone of the first two N-terminal lanthionine rings of NAI-107 is likely to be involved in lipid II binding (Munch et al., 2014). This in turn suggests that Trp4 might directly interact with the bacterial membrane, explaining why a higher hydrophobicity at this residue leads to improved antibacterial potency.

NAI-107 has pharmacological properties that make it a suitable candidate for treating serious infections by Gram-positive pathogens (Lepak et al., 2015). The same lanthionine ring topology as in NAI-107 is present in a recently described actinomycete-produced lantibiotics, whose antibacterial activity can be substantially enhanced by a simple semi-synthetic derivatization (Maffioli et al., 2015). The work described here provides additional players in this promising family of antibacterial agents.

6. GENOMIC ANALYSIS OF *ACTINOALLOMURUS*

6.1. Introduction

The final part of my project involved the characterization of the genome sequences of four *Actinoallomurus* strains. Throughout these three years two strains were initially sequenced as part of the metabolite characterization by analysing the putative BGCs devoted to the synthesis of the compounds: strain ID145113, with analysis of the BGC providing insights for the structural elucidation of NAI-113 (Chapter 3); and strain ID145698, with the BGC revealing unusual biosynthetic features of allocyclinone (Chapter 4). By the end of the project, I decided to analyse two more *Actinoallomurus* strains in order to have a reasonable number of strains for some genomic analysis of the genus. The four *Actinoallomurus* genomes were compared amongst each other and against reference genomes, which were chosen on the basis of their phylogenetic relatedness to *Actinoallomurus* or their relevance as producers of secondary metabolites within the class *Actinobacteria*.

6.2. Experimental Section

For Genome sequencing: Five mL of a three days old culture was centrifuged at 1800 rcf for 10 minutes, the pellet was resuspended in 10 mL of 10.3% Sucrose. A new centrifugation was performed (1800rcf, 10 min) and the pellet was resuspended in 5 ml of SET buffer. pH was adjusted to 8 with NaOH. Lysozyme was added to a final concentration of 20mg/mL, and left to incubate at 37ºC for 24h, then 100 μg of ribonuclease were added and incubated for another 4 hours at 37ºC. Finally, 12mg of Proteinase K were added leaving it incubating for 2 more hours. At this point SDS was added to a final concentration of 1% and put to incubate at 55ºC for 20 hours. On the next day added 2 mL of 5M NaCl followed 2 mL of phenol-chloroform-isomylalcohol (pH8). Mixed gently and let the two liquid phases separate by centrifuging at 1800 rcf during 10 min. Retrieved the aqueous phase and added 2ml of fresh phenol-chloroform-isomylalcohol, mixed, centrifuged and discarded the heavy phase. Added one volume of ice cold isopropanol. Mixed gently until DNA precipitates. Using a sealed Pasteur pipette transferred the DNA to a new tube. Washed it with 70% ethanol and let

it dry. Resuspended in water. Sequencing of the amplified PCR products was done by Cogentech.

Bioinformatic analysis. The sequencing was conducted on an Illumina MiSeq, reading 2x250 bases of a paired-end and a mate-pair library. These libraries had pair sizes of 700 bp and 7.600 bp, respectively. Alignments of single genes or gene products were done through the workflow platform Armadillo v1.1. For DNA alignments, ClustalW2 was used with a gap opening penalty of 10; Extension penalty of 0.2. The alignment of protein sequences was performed using MUSCLE algorithm at a maximal number of iteration of 4. Distances were calculated based on Jones-Taylor-Thornton matrix which was then used to create the phylogenetic trees. The visualization of all phylogenetic trees was achieved with Archaeopteryx tool. (Edgar, 2004; Han and Zmasek, 2009; Thompson et al., 1994) Genome sequencing was performed by Universität Bielefeld Centrum für Biotechnologie – CeBiTec. For the orthology analysis, annotated proteins were subject to all-vs-all BLASTP. OrthoMCL package that provided all the tools was used with the following parameters: identity of 50% and evalue threshold at 0.00001. (Chen et al., 2006) For the synteny analysis the Symap tool was used with a cutoff of 85% identity. (Soderlund et al., 2006) The discovery of the putative gene clusters for secondary metabolism was performed using Antismash v3.0.1 tool at the default conditions. (Weber et al., 2015) Blast analysis of the genes clusters identified by Antismash were also blasted against MIBiG database using the evalue above 0.00001 as a cutoff.

Fosmid library preparation and screening. The DNA extraction was achieved using Promega's Wizard Genomic DNA Purification Kit which, although providing a smaller amount of DNA than the method used for genome sequencing, afforded a larger average fragment size. After the retrieving the DNA sample with the corrected characteristics the ligation reaction was performed with pCC2FOS vector and consequent packaging and transfection using the CopyControl HTP Fosmid Library Production Kit with pCC2FOS Vector and following the correspondent protocol but the clones were stored in a semiliquid media using SeaPrep agarose as described in Gurgui *et al.* (Gurgui and Piel, 2010) The screening for the inserts of interest was performed through PCR using primers designed to amplify fragments at the two ends of the cluster (Table 6.1).

Table 6.1 – PCR primers for the NAI-113 and Cin-like clusters in strain ID145113 genomic library

Lantipeptide primers		
Left flank	(5'->3')	Tm
Forward primer	CGGTCCAGTGTTCCGAGTC	60.08
Reverse primer	AATCGCTCGTGGAGGGGATA	60.47
Product length	588	
Right Flank	(5'->3')	Tm
Forward primer	GAGCATCGATCGTCTGGGAC	60.32
Reverse primer	CCGGACAGGAAGATCATCGG	59.97
Product length	322	
NAI-113		
Right Flank	Sequence (5'->3')	Tm
Forward primer	ATCTACAGTCTGGTCCCGCT	60.03
Reverse primer	CCCGTGATGACGAGGTTGAA	60.04
Product length	701	
Left Flank	Sequence (5'->3')	Tm
Forward primer	GGTGACACGTTCTGGGACAT	59.97
Reverse primer	GTCGATGCTCTTGTCCTGCT	60.11
Product length	510	

6.3. Results

The features of the four sequenced *Actinoallomurus* genomes are reported in Table 6.2. The genome sizes vary from 8.5Mbp for strain ID145716 to 10.9 for strain ID145698 The four genomes show a similar GC content (from 69.4 to 71.2%), coding density (from 89.8 to 85.6%), with a number of CDSs ranging from 8325 to 10174 (Table 6.2). In selecting the genomes to be used for comparative work, I picked two more representatives of the *Thermomonosporaceae* family (order *Streptosporangiales*): one *Actinomadura* sp. and one *Thermomonospora* sp. (Table 6.2). Unfortunately, no additional genomes were available from these genera, and the genome of *Thermomonospora curvata* is substantially smaller than the others (Table 6.2). Additional genomes were selected taking advantage of previous analyses performed on one strain, each from three distinct genera (*Microbispora*, *Planobispora* and *Streptosporangium*) within the *Streptosporangiaceae* family (order *Streptosporangiales*). (Tocchetti et al., 2015) As outgroup genomes, I selected *Streptomyces coelicolor* (order *Streptomycetales*) and *Actinoplanes friuliensis* (order *Micromonosporales*), both with a large genomic potential to produce secondary metabolites. Table 6.2 shows the general genomic features of all the strains used throughout this work.

Table 6.2 – Features of the genomes used throughout the analyses

Strains	Family	Order	Length (Mbp)	GC Content (%)	Gene Density (%)	CDS	Accession Number
Actinoallomurus ID145113	*Thermomonosporaceae*	*Streptosporangiales*	10.7	69.4	85.6	10174	This work
Actinoallomurus ID145698			10.9	69.9	87.3	10070	This work
Actinoallomurus ID145711			9.5	71.1	89.8	8899	This work
Actinoallomurus ID145716			8.47	71.2	86.7	8325	This work
Actinomadura oligospora			9.35	72.1	85.6	8178	NZ_JADG00000000.1
Thermomonospora curvata			5.6	71.6	84.8	4895	NC_013510.1
Microbispora sp.	*Streptosporangeacea*		8.7	71.3	89.0	7966	NZ_JNZQ00000000.1
Planobispora rosea			8.7	70.0	88.0	8071	SRP041970
Streptosporangium roseum			10.34	70.9	87.4	9043	NC_013595.1
Streptomyces coelicolor	*Streptomycetaceae*	*Streptomycetales*	8.67	72.1	83.3	8152	NC_003888.3
Actinoplanes friuliensis	*Micromonosporaceae*	*Micromonosporales*	9.38	70.4	90.4	8621	NC_022657.1

From the taxonomical point of view, Table 6.2 covers three orders and four different families, but the emphasis is on a comparative analysis of members of the *Streptosporangiales*. From the general traits (genome size, GC content and coding density), *Actinoallomurus* fits well with the features observed for filamentous actinomycetes (Table 6.2).

6.3.1. Whole-genome comparisons

The orthology analysis to the panel of strains is described on Table 6.3. That is an all-vs-all blast analysis of the protein sequences from the eleven strains, with every sequence being compared as subject or query. This fact may affect the scoring of the alignments, especially the Expect value (e-value) that takes into account also the length of the query sequence. As a consequence, there may be cases where the same alignment gives an e-value just below the threshold (0.00001) when one of the sequences is used as query and the other as subject, but when the roles are reversed the e-value goes above the threshold. As a consequence, the number of observed homologues when a proteome is used as a query cannot be identical to that observed when it is used as a subject. Nonetheless, the numbers of orthologues observed in each pairwise comparison when switching subject and query are very similar (Table 6.3). Note also that in Table 6.3 the values in the diagonal actually represent the total of homologs seen against all the remaining strains. Thus, these numbers represent the total number of CDSs present in any strain minus the singletons (Table 6.3) and those excluded from analysis (see below).

Table 6.3 – Number of homologous proteins among the 10 selected genomes. Top panel absolute values. Bottom panel relative values.

Absolute Values

Subject \ Query	ID145113	ID145698	ID145711	ID145716	Actinomadura	Thermomonospora	Microbispora	Planobispora	Streptosporangium	Streptomyces	Actinoplanes
ID145113	8110	5595	5422	5596	2976	3077	4120	3800	4299	3610	3317
ID145698	5473	8175	6236	4838	2911	3061	4197	3661	4169	3495	3231
ID145711	5387	6313	7706	4793	2900	3023	4144	3630	4044	3485	3248
ID145716	5580	4939	4825	6706	2720	2918	3756	3430	3707	3234	3011
Actinomadura	2967	2909	2904	2710	4768	2138	2838	2922	3013	3001	2892
Thermomonospora	3099	3113	3051	2932	2174	4019	2770	2821	2985	2465	2342
Microbispora	4138	4253	4159	3758	2845	2760	6586	4240	4617	3301	3135
Planobispora	3779	3696	3627	3412	2887	2789	4215	6473	5022	3171	3162
Streptosporangium	4287	4186	4040	3694	2998	2947	4567	5012	7357	3518	3259
Streptomyces	3588	3524	3464	3205	2998	2428	3275	3169	3493	5658	3067
Actinoplanes	3307	3257	3245	2993	2896	2313	3119	3213	3262	3074	5391

CoOrthologs	739	829	654	624	682	452	673	691	800	819	864
Paralogs	566	802	359	320	723	280	407	647	678	919	1077
Singletons	1683	1571	1041	1436	1757	744	1231	1373	1337	1674	2243
Total	9970	9971	8816	8247	6917	4891	7978	8072	8946	7768	8248

Relative Values

Query / Subject	ID145113	ID145698	ID145711	ID145716	Actinomadura	Thermomonospora	Microbispora	Planobispora	Streptosporangium	Streptomyces	Actinoplanes
ID145113	81%	56%	62%	68%	43%	63%	52%	47%	48%	46%	40%
ID145698	55%	82%	71%	59%	42%	63%	53%	45%	47%	45%	39%
ID145711	54%	63%	87%	58%	42%	62%	52%	45%	45%	45%	39%
ID145716	56%	50%	55%	81%	39%	60%	47%	42%	41%	42%	37%
Actinomadura	30%	29%	33%	33%	69%	44%	36%	36%	34%	39%	35%
Thermomonospora	31%	31%	35%	36%	31%	82%	35%	35%	33%	32%	28%
Microbispora	42%	43%	47%	46%	41%	56%	83%	53%	52%	42%	38%
Planobispora	38%	37%	41%	41%	42%	57%	53%	80%	56%	41%	38%
Streptosporangium	43%	42%	46%	45%	43%	60%	57%	62%	82%	45%	40%
Streptomyces	36%	35%	39%	39%	43%	50%	41%	39%	39%	73%	37%
Actinoplanes	33%	33%	37%	36%	42%	47%	39%	40%	36%	40%	65%
CoOrthologs	7%	8%	7%	8%	10%	9%	8%	9%	9%	11%	10%
Paralogs	6%	8%	4%	4%	10%	6%	5%	8%	8%	12%	13%
Singletons	17%	16%	12%	17%	25%	15%	15%	17%	15%	22%	27%

Values in diagonal correspond to the total number of homologous of that strain shared with the universe of 10 strains used. The Colouring of the cells ranges from dark red to dark green correspond from the lowest to the highest values.

Looking at Strain ID 145113, and going from the top cell onto the bottom of each panel, it is observable that 8110 sequences from the analysed 9970 proteins find a homologue with at least one other sequence from the set of sequences used for this analysis. On the second table, at the same position it is possible to see that this corresponds to 81%. Moving on to the next three cells below, a high number of homologues to its peers of the same genus (ID145698; ID145711; ID145716) is observed, with the number of sequence around 5500, corresponding to 54% to 56% of the strains proteomes. Dropping the reader's eyes for further two cells it is possible to see a comparison of Strain ID 145113 with *Actinomadura oligospora* and *Thermomonospora curvata*, the other genera belonging to the same family as *Actinoallomurus*. In this case, the number of homologous sequences drops to 2967 for *A. oligospora* and 3099 for *T./ curvata*, corresponding to 30 and 31% of the proteome, respectively. The next three cells contain the results from the *Streptosporangeaceae* family members (*Microbispora* sp., *Planobispora rosea*, Streptosporangium *roseum*). In this case the numbers of homologues increase, with *Microbispora* (4138 – 42%) and *S. roseum* (4287 – 43%) as the most similar organisms, while *P. roesa* (3779 – 38%) has a somewhat lower number. Finally, *Streptomyces* and *Actinoplanes*, which were chosen as outgroups, show lower number of homologues.

The last three lines of Table 6.3 report "Singletons" – protein sequences that have no match in the tested conditions and probably represent the proteins unique to this organism within the chosen comparators; "Paralogs" – sequences that find homology within the same genome; and Co-orthologs – proteins that find a match in the own proteome and in the set of proteomes tested. In the case of strain ID145113, Singletons, Paralogs and Co-orthologs are 1683, 566 and 739, respectively, which correspond to 17, 6 or 7% of the whole proteome. The number of Singletons is significant if one considers that three more genomes from the same genus were analysed. The very last line of the top panel shows the total of sequences analysed during this work. The total number of Table 6.3 does not match that presented in Table 6.2, since RNAs and proteins with less than 50 amino acids were removed from the analysis.

A comparison horizontally reveals the number of homologues found in ID145113 when the entire proteome from the other strains was used as query. This cross-comparison is useful because of the different numbers of homologues observed when inverting subject and query, as explained above, and for taking into account different genome sizes (Table 6.2). For example, while "only" 3077 *T. curvata* sequences find a match in ID145113, this number represents 63% of the *T. curvata* proteome.

Overall, the data in Table 6.3 provide the following picture. As expected, the four *Actinoallomurus* strains share the largest number of orthologues among themselves, with strains ID145698 and 145711 appearing as the most related by this criterion. Nonetheless, each strains contains 12-17% of Singletons, not very different from the numbers seen in the other strains used in the comparison, where one strain per genus was used. Unexpectedly, however, the four *Actinoallomurus* strains share the highest number of orthologues with other members of the *Streptosporangiacaea* and not with members of *Thermomonosporaceae*, as would be expected from their taxonomic position.

Although we have performed a comparison with the Cluster of Orthologous Groups (COG) database, the results showed no obvious trend (data not shown). However, it is important to mention that there seems to be a tendency to abandon this kind of analysis as the last published update of COG is over a decade old. Furthermore, the COG database includes only four *Actinobacteria* from the *Corynebacterium* and *Mycobacterium* genera, thus far away from the group of organisms analysed in this work (Tatusov et al., 2003).

The orthology analysis was taken to a more detailed level with a selected few organisms based on the data gathered from Table 6.3. An initial analysis was done amongst the four *Actinoallomurus* strains. Two main features can be easily observed from the Venn diagram (Fig. 6.1A): a significant number of orthologues (3942) common within the four organisms, which are likely to include the core *Actinoallomurus* genome; and an elevated number of singletons (between 2569 and 3756 CDSs in each genome). While the former number is expected from a number of conserved functions within *Actinoallomurus*, the high number of singletons provides grounds for the existence of an extended pangenome within this genus. It should be noted that a significant fraction of the singletons from this four-strain comparison find a match within at least one of the other analysed genomes (i.e., compare the "Singleton" line in Table 6.3 with the numbers reported in Fig. 6.1A).

The next question was to ask how many of the 3942 *Actinoallomurus* core CDSs (Fig. 6.1A) are shared with other organisms from the same order. This analysis was first applied to *A. oligospora* (the most similar species according to 16S-based phylogenetic analysis) and *T. curvata* (the other member of the *Thermomonosporaceae* family, but with a smaller genome). Fig. 6.1B shows that less than half of the shared *Actinoallomurus* orthologues (1688 out of 3942 sequences) find a match in both *A. oligospora* and *T. curvata*. Furthermore, despite its smaller genome, *T. curvata* shares a larger number of CDSs with the *Actinoallomurus* core than does *A. oligospora*. This result is consistent with the data reported in Table 6.3, and suggests that the overall closer relationship of *Actinoallomurus* spp. with *T. curvata* than with *A. oligospora* is due to sharing a larger number of core functions.

Finally, we extended this type of analysis by including two representatives from the *Streptosporangiaceae*, *S. roseum* and *Microbispora* sp., together with either *A. oligospora* (Fig. 6.1C) or *T. curvata* (Fig. 6.1D). [This type of analysis can be easily visualized with up to four different genomes.] Consistent with the previous observations, there is a larger number of shared orthologues (1969; Fig. 6.1D) when *T. curvata* is included in the comparison than when using *A. oligospora* (1669; Fig. 6.1C). Interestingly, the inclusion of two representatives from the *Streptosporangiaceae* does not alter significantly the number of core CDSs (compare Fig. 6.1B with either 6.1C or 6.1D), suggesting the core genome of the order *Streptosporangiales* might approximate to ~1600 sequences. Another interesting observation from the data shown in Fig. 6.1C is the significant number of orthologues shared by *Actinoallomurus* with both *S. roseum* and *Microbispora* sp. (818) but not with *A. oligospora*.

Since only one genome sequence from representatives of the genus *Actinomadura* was used, we cannot exclude the possibility that the genome of *A. oligospora* is particularly divergent, and similar analyses with other members of this genus would show a higher relatedness with *Actinoallomurus*. Nonetheless, the results from the analyses presented here are also consistent with the hypothesis that *Actinoallomurus* might actually be similarly related to members of the *Streptosporangiaceae* family as to members of the *Thermomonosporaceae* family. All together these data call for further phylogenetic analyses to resolve a possible discrepancy between 16S-based classification and whole genome comparisons. However, this type of analyses falls outside the scope of my PhD project and was not pursued any further.

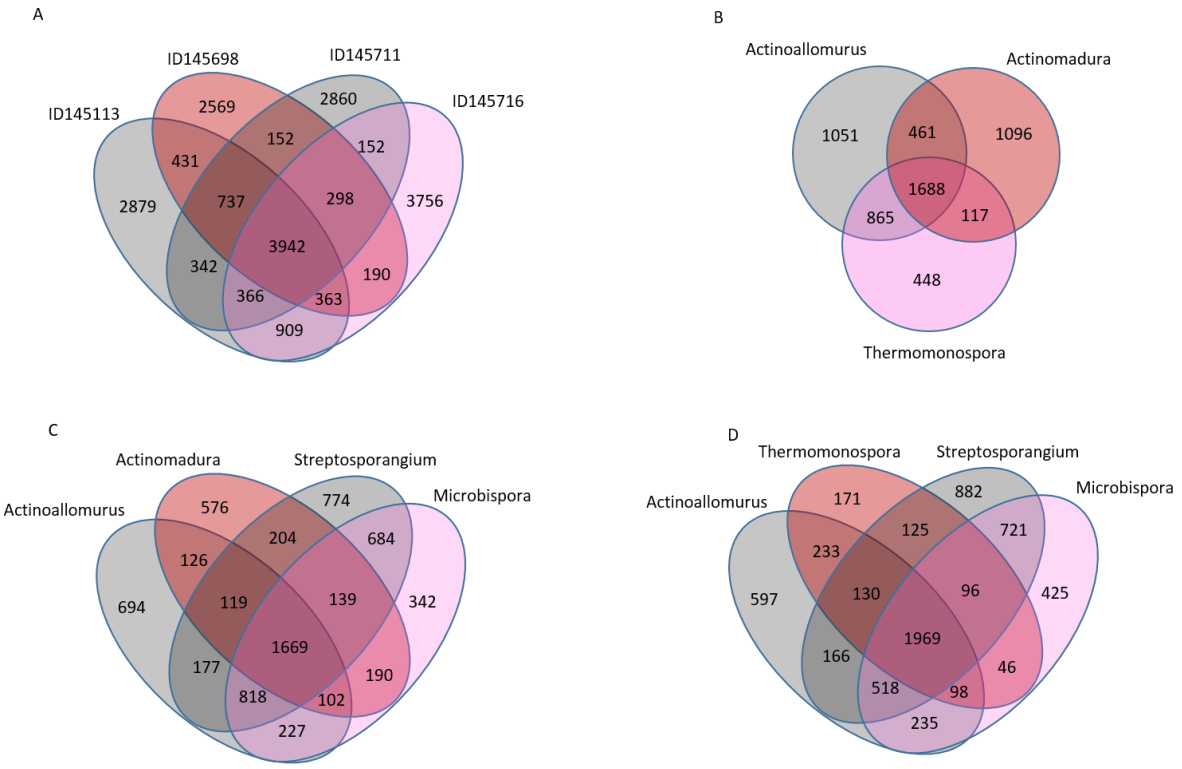


Figure 6.1 – Venn diagrams describing the number of shared orthologues between selected organisms. Panel A, comparison of the four *Actinoallomurus*; Panel B, comparison of the common *Actinoallomurus* core with other *Thermomonosporaceae*; Panel C, comparison of common *Actinoallomurus* core with *A. oligospora* and two members of the *Streptosporongaceae* family; Panel D comparison of common *Actinoallomurus* core with *T. curvata* and two members of the *Streptosporongaceae* family;

Finally, I concluded whole genome comparisons by analysing the extent of synteny existing among the *Actinoallomurus* genomes, including also *A. oligospora*, *T.curvata*, and *Microbispora* for comparison. A graphical representation of those results is shown Fig. 6.2.

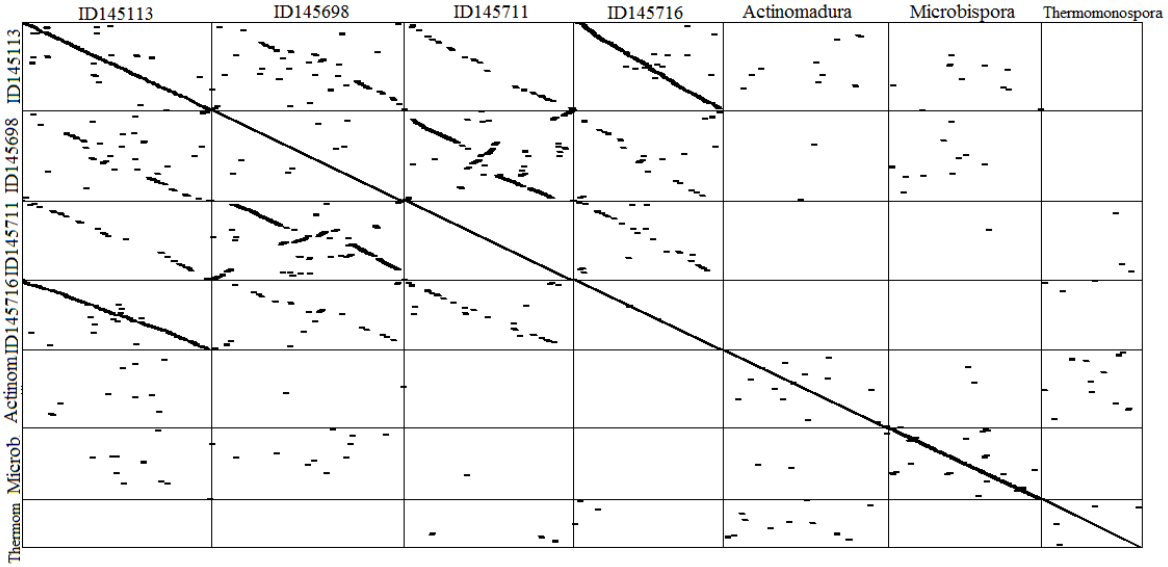


Figure 6.2 – Synteny analysis of the four *Actinoallomurus* genomes with those from *A. oligospora*, *S. roseum*, *T. curvata*. Each dot corresponds to homologous region. See Experimental Section for details.

This type of analysis shows that the four *Actinoallomurus* strains share a high degree of synteny. Specifically, the genomes of ID145113 and ID145716 appear highly related, consistent with the high number of shared orthologues (Table 6.3 and Fig. 6.1A). In contrast, strain ID145698 appears to be the most divergent of the four strains in terms of shared gene order and content. A comparison between the genomes of ID145698 and ID145711 points to the existence of large inverted segments. As expected from the previous analyses, little synteny was observed with genomes from outside the *Actinoallomurus* genus. This observation contrasts with the synteny observed between representatives from the genera *Streptosporangium*, *Planobispora* and *Microbispora* (Tocchetti. *et al*., 2015), warrantying further analyses of one of the lesser studied genera of the *Actinobacteria*.

6.3.2. Biosynthetic gene clusters

By taking advantage of the Antismash v3.0.4 tool, it was possible to pin point a list of all BCGs and summarize it in Table 6.4. Overall, the number of BGCs and the fraction of the

genome dedicated to secondary metabolism fall within the range of filamentous actinomycetes, including the most prolific producers of secondary metabolites, i.e. members of the *Streptomyces* genus.(Bentley et al., 2002; Harrison and Studholme, 2014; Ikeda et al., 2003)

Table 6.4 – List of all the gene clusters found by antiSMASH in the genomes of *Actinoallomurus* strains ID145113, ID145698, ID145711, and ID145716

Type of Cluster	Total	ID145113	ID145698	ID145711	ID145716
Terpene	16	3	4	6	3
Lantipeptide	15	5	2	3	5
T1PKS	9	0	4	1	4
Other	6	1	2	1	2
T2PKS	6	1	2	2	1
NRPS	5	0	1	1	3
Ectoine	4	1	1	1	1
Siderophore	4	1	1	1	1
Bacteriocin	3	1	0	1	1
Lassopeptide	3	2	0	0	1
Bacteriocin-lanthipeptide	2	0	1	1	0
Indole	2	0	1	1	0
Bottromycin	1	0	0	0	1
Butyrolactone-Ladderane-OtherKS	1	0	0	0	1
Lantipeptide-bacteriocin	1	1	0	0	0
Lantipeptide-OtherKS	1	0	0	1	0
NRPS-bacteriocin	1	1	0	0	0
NRPS-terpene	1	0	0	0	1
Otherks-lanthipeptide	1	0	1	0	0
Otherks-t1PKS-terpene	1	1	0	0	0
T3PKS-lassopeptide	1	0	1	0	0
Terpene-bacteriocin	1	0	0	1	0
Terpene-t1PKS-NRPS	1	0	0	0	0
Terpene-t3PKS	1	0	0	0	1
Thiopeptide-lanthipeptide-bacteriocin	1	0	1	0	0
Total number BGCs	88	18	22	21	26
Total size of BGCs (Mbp)	2.78	0.59	0.76	0.51	0.92
Genome fraction	7%	6%	7%	5%	11%

Overall, Antismash identified 88 BGCs. Those were then regrouped in sets according to their biosynthetic pathway, as to facilitate a comparison between whole genome analysis and the results from the screening exercise (Table 6.4). At this point it was also important to filter the Antismash results, highlighting which *Actinoallomurus* BGCs presented significant homology to BGCs involved in the synthesis of known compounds (see Experimental Section

for details). This analysis reduced the list to 55 BGCs that could be analysed in further detail (Table 6.5).

Table 6.5 – Summary of the BGCs from Table 6.4 grouped according to their biosynthetic process (second column). The number of BGCs with homology to known cluster is also indicated.

Biosynthetic class	BGCs	BGCs with homology*
Ectoine	4	3
Indole	2	2
Other	7	3
polyketide	15	11
hybrid PKS-NRPS	6	6
RiPP	26	13
non-ribosomal peptide	14	11
Siderophore	4	0
Terpene	16	12
Total	88	55

*** percentage of genes in the *Actinoallomurus* cluster with an identity ≥ 45% with the corresponding genes in the reference cluster.**

Overall, Table 6.5 shows how BGCs for RiPPs are found in relatively more abundance, followed by those for terpenes, polyketides and non-ribosomal peptides.

6.3.2.1. Analysis of PKS1 Clusters

One of the significant findings of screening *Actinoallomurus* was the relatively high frequency at which likely polyethers were detected (Chapter 2.42.). Therefore, I wondered if this feature was also reflected within the genomes. One strain, ID145716, presents two BGCs (clusters 13 and 17) with features consistent with polyether biosynthesis (data not shown): presence of PKS I with modules enriched with domains carrying out reductive steps; and presence of an epoxidase, reported as landmark enzyme for polyether biosynthesis (Jiang et al.,

2012). The other three genomes showed no obvious BGC for the production of these compounds.

6.3.2.2. Aromatic Polyketides

As mentioned earlier the sequenced genomes originate from strains that produced or showed the potential to produce new compounds. Three out of four of the selected strains (ID145113, ID145698 and ID145711) produced an aromatic polyketide with a quinone moiety and hence their genomes were expected to contain a PKS II BGC. The analysis of two of such BGCs have been reported in Chapters 4 and 5, including how the corresponding KSβ's have been used to predict the polyketide length. An additional four PKS II BGCs were identified in the four genomes and the KSβ analysis was extended to all them to infer the likely polyketide scaffold. The resulting phylogenetic tree is described in Fig. 6.3.

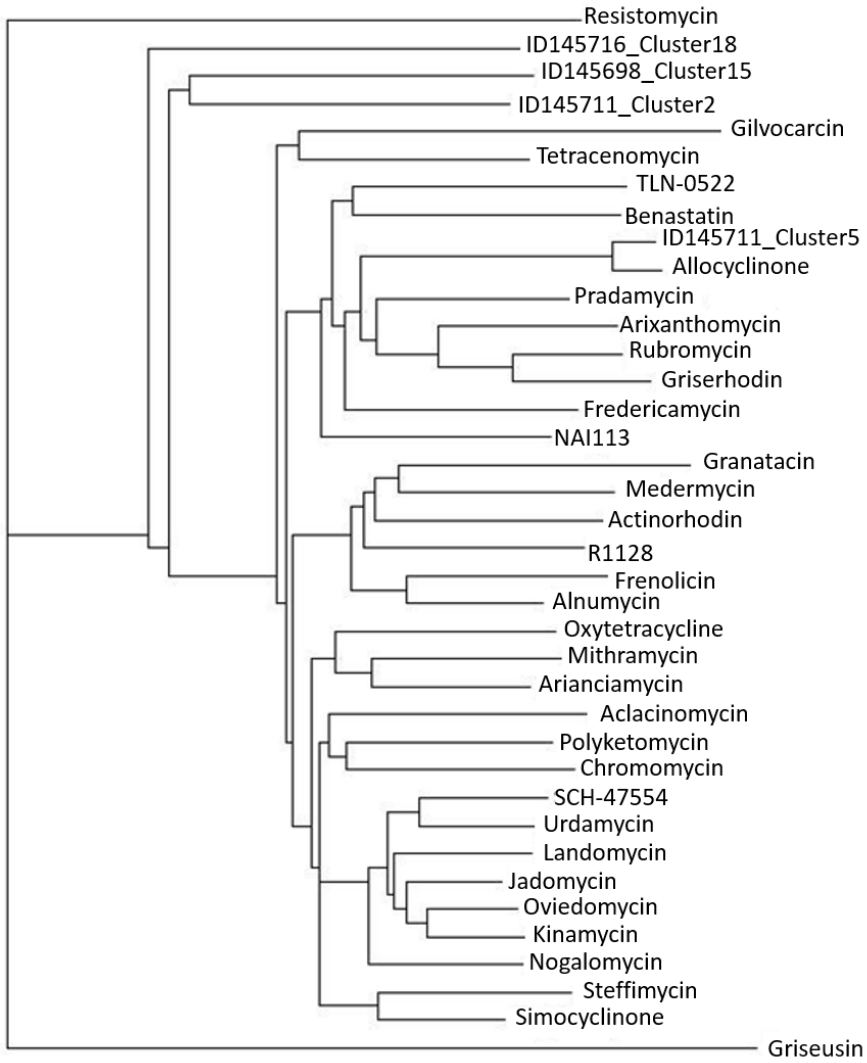


Figure 6.3 –Phylogenetic tree of the KSβ from 35 different biosynthetic clusters with known product. The KSβ from all Actinoallomurus BGCs are represented by the ID of the strain. Supporting Table 1 provides additional information on the KSβ's used for the phylogenetic analysis.

In Fig. 6.3 it is possible to see how the already mentioned KSβ s –those believed to be involved in the synthesis of allocyclinone and NAI-113 – cluster within the clade for dodecaketide or tridecaketide carbon backbones along with an enzyme from Strain ID 145711, suggesting that the as yet unidentified polyketide from this strain may derive from an equivalent number of acetate units. The second striking feature of Fig. 6.3 is the remaining three KSβs from *Actinoallomurus* (ID145716_Cluster18; ID145711_Cluster15; ID145698Cluster_2) that seem not to be correlated with any of the known KSβ clades. This may imply that these BGCs are involved in the synthesis of polyketides unrelated to the families used for phylogenetic analysis.

Figure 6.4 shows an alignment of the CLFs from the allocyclinone BGC and from Cluster 5 of strain ID145711. The two proteins share 91% identity with conservation of aa positions 109, 112, 116, and 133, which have been shown to play a key role in determining the length of the carbon backbone of aromatic polyketides (Tang et al., 2003).

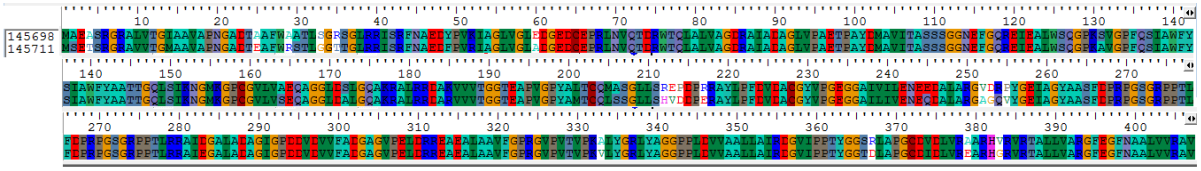


Figure 6.4 – Alignment of the KSβs from Allocyclinone cluster and from Cluster 5 of Strain 145711

Due to the high similarity of the two KSβs from Fig. 6.4, and having already analysed the allocyclinone BGC (see Chapter 4), I decided to have a closer look at both clusters comparing each CDS. The results, reported in Table 6.6, show that the BGC from strain ID145711 is highly related to the allocyclinone cluster, showing, with one exception, the same composition and order of CDSs, each with a high percent identity to the cognate protein from the other cluster. One noteworthy feature of the allocyclinone cluster is the presence of two halogenases, presumably involved in different regioselective halogenations (see Chapter 4). So does the BGC from strain ID145711, suggesting that this cluster might also be involved in the synthesis of an aromatic polyketide containing two halogenated positions. The only difference between the two BGCs is the presence of a four-gene insertion in the ID145711 cluster, coding for two putative regulators, a protein of unknown function, and strikingly, an aminotransferase. It should be pointed out that MS analysis of an extract prepared from strain ID145711 did not find m/z^+ values corresponding to any of the congeners in the allocyclinones complex (data not shown), but instead a compound with UV spectrum and MS isotopic pattern suggestive of a halogenated quinone (see Chapter 4). Further analyses will be necessary to characterize the metabolites produced by strain ID145711 and establish any association with Cluster 5.

Table 6.6 – Comparison of CDSs from Allocyclinone biosynthetic cluster with those from NAI711_Cluster 5.

Allocyclinone			NAI711_Cluster5		
CDS	Length (aa)	Putative Function[a]	Length (aa)	CDS	ID%
---	---	Hydrolase	363	NAI711_2242	---
NAI698_9155	369	Hypothetical Protein	---	---	---

NAI698_9156	961	Lux family Transcriptional Regulator	963	NAI711_2243	81
NAI698_9157	177	Regulator Protein	175	NAI711_2244	80
NAI698_9159	218	hypothetical protein	215	NAI711_2246	86
NAI698_9160	301	putative DNA-binding protein	301	NAI711_2247	95
NAI698_9161	289	putative DNA-binding protein	289	NAI711_2248	91
NAI698_9162	172	Oxidoreductase	172	NAI711_2249	93
NAI698_9163	262	SARP-like transcriptional regulator	262	NAI711_2250	93
NAI698_9164	342	O-methyltransferase	342	NAI711_2251	93
NAI698_9165	318	hypothetical protein	320	NAI711_2252	88
NAI698_9166	308	Transcriptional regulator	303	NAI711_2253	79
NAI698_9167	294	Putative Hydrolase	290	NAI711_2254	86
NAI698_9168	211	Unknown Function	211	NAI711_2255	86
NAI698_9169	108	Polyketide synthesis cyclase	108	NAI711_2256	80
NAI698_9170	537	Decarboxylase	537	NAI711_2257	95
NAI698_9171	138	Putative cyclase	136	NAI711_2258	89
NAI698_9172	424	Ketoacyl synthase - alpha	425	NAI711_2259	91
NAI698_9173	407	Ketoacyl Synthase - Beta	407	NAI711_2260	91
NAI698_9174	243	Polyketide cyclase	246	NAI711_2261	87
NAI698_9175	67	Hypothetical protein	68	NAI711_2262	84
NAI698_9176	245	Acyl Carrier Protein	245	NAI711_2263	87
NAI698_9177	112	Monooxygenase	108	NAI711_2264	88
NAI698_9178	108	Monooxygenase	108	NAI711_2265	95
NAI698_9179	393	Oxidoreductases	393	NAI711_2266	90
NAI698_9180	153	Hydroxylase	155	NAI711_2267	90
NAI698_9181	583	Acetyl-CoA carboxylase biotin carboxylase	584	NAI711_2268	84
---	---	hypothetical protein	221	NAI711_2269	---
---	---	Serine/threonine kinase	644	NAI711_2270	---
---	---	Aspartate transaminase	386	NAI711_2271	---
---	---	transcriptional_regulator, _XRE_family	185	NAI711_2272	---
NAI698_9182	168	MarR-like transcriptional regulator	159	NAI711_2273	84
NAI698_9183	569	Drug resistance transporter	568	NAI711_2274	89
NAI698_9184	255	ACP-reductase	257	NAI711_2275	87
NAI698_9185	276	O-methyltransferase	276	NAI711_2276	93
NAI698_9186	555	Halogenase	552	NAI711_2277	94
NAI698_9187	145	Ester Cyclase	145	NAI711_2278	94
NAI698_9188	168	Cupin domain protein	186	NAI711_2279	85
NAI698_9189	146	Ester Cyclase	146	NAI711_2280	88
NAI698_9190	435	Membrane antiporter	435	NAI711_2281	83
NAI698_9191	560	Halogenase	560	NAI711_2282	91
NAI698_9192	147	Hypothetical protein	148	NAI711_2283	85
NAI698_9193	530	Monoxygenase	530	NAI711_2284	88

NAI698_9194	590	Peptidase	---	---	---
---	---	Transcriptional Regulator	184	NAI711_2285	---

a – the putative function was already assigned on chapter 5, Table 5.5.

Dashed lines define the limits of the putative gene clusters within the Antismash output

6.3.2.3. Curious case of halogenases

As mentioned, each of the two clusters from Table 6.6 encodes two distinct halogenases. Halogenated metabolites are often found encountered as products from *Actinoallomurus* (see also Chapters 3, 4 and 5) (Cruz, 2015; Cruz et al., 2015; Mazzetti et al., 2012) and, hence, the corresponding genomes should encode the required halogenases. This feature was investigated and, within three of the four genomic sequences, six genes for halogenases were found, with only strain ID145113 not presenting any. Strain ID145716 presented one, strain ID145711 two, and strain ID145698 three halogenases. The protein sequences were then aligned and a phylogenetic tree built upon this data, as represented in Fig. 6.5.

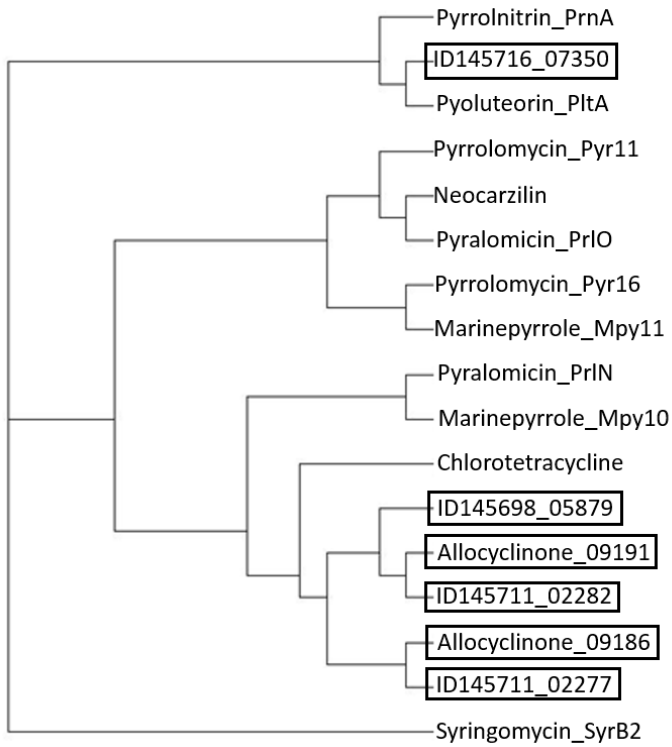


Figure 6.5 – Phylogenetic comparison of halogenase enzymes found in all *Actinoallomurus* genomes (square boxes) with halogenases of known function.

As shown in Table 6.6, four of the halogenases (two from the allocyclinone cluster and two from the related BGC from strain ID145711) represent two highly related pairs presumably involved in regiospecific chlorination of two positions in related aromatic polyketides. Closely related to these two pairs is a third halogenase (ID145698_05879) from the allocyclinone producer ID145698, which is associated with a cluster identified as “Other” by Antismash. These five halogenases find their best match with the halogenase from the chlortetracycline cluster suggesting halogenation of aromatic system(s). Finally, strain ID145716 has also a halogenase associated to a cluster classified by AntiSMASH as “Other" but encoding some enzymes similar to those involved in the biosynthesis of chloropeptin, a hyper-halogenated heptapeptide structurally related to the glycopeptide antibiotics (Garfunkle et al., 2009) (data not shown). This amino acid sequence clade best with those from pyoluteorin and pyrrolonitrin clusters, which might indicate that this halogenase uses a pyrrole moiety as a substrate. Note that the previously described *Actinoallomurus* metabolite NAI-414 carries a halogenated pyrrole (Mazzetti et al., 2012).

6.3.2.4. NRPS and PKS-NRPS hybrids

A total of 15 clusters were identified by Antismash, and six of them are PKS-NRPS hybrids. Only for ten clusters antiSMASH was able to assign a similar BGC for characterized compounds. However, the *Actinoallomurus* clusters were significantly divergent from the identified reference clusters as to suggest unrelated secondary metabolites as the products of the *Actinoallomurus* BGCs.

6.3.2.5. RiPPs

AntiSMASH yielded 24 BGCs involved in RiPP formation, but only three of them show a significant homology with clusters directing the synthesis of known molecules. Table 6.7 summarizes these findings.

Table 6.7 – Identity* Of *Actinoallomurus* RIPP BGCs with the most similar gene cluster for known peptides.

	ID145113	ID145698	ID145711	ID145716
SSV-2083	37%			
CINNAMYCIN	33%	33%	33%	19%
CATENULIPEPTIN				60%

* PERCENTAGE OF GENES IN THE *ACTINOALLOMURUS* CLUSTER WITH AN IDENTITY ≥ 45% WITH THE CORRESPONDING GENES IN THE REFERENCE CLUSTER.

The most striking case of Table 6.6 is the discovery that a cinnamycin-like cluster was present in all the four strains. A detailed analysis of the CDSs present in these four clusters was then undertaken. As it can be seen in Fig. 6.6, three of the clusters show a conserved set of eight genes, while only the cluster from strain ID145716 appears to be slightly divergent.

The CDSs from the four clusters were blasted against the MIBiG database. The resulting best hits are reported in Table 6.8. As it can be seen all four clusters present a high relatedness with the cinnamycin cluster. However, while the BGCs from strains ID145113, ID145698 and ID145711 encode all the proteins required for cinnamycin formation (i.e., the precursor peptide CinA, the bifunctional dehydratase/cyclase CinM, the Asp hydroxylase CinX, the transporters CinH and CinT, as well as the regulator CinR1 and a homologue of CinORF7), the cluster from strain ID145716 presents a lower percent of identity with the Cin proteins and, notably, lacks CinX and the transporters CinH and CinT, suggesting that the cluster in strain ID145716 might direct the synthesis of a lantibiotic more divergent from cinnamycin than those specified by the other three clusters.

Table 6.8 – Comparison of the CDSs from each of the *Actinoallomurus* Cinnamycin-like BGCs with the Cinnamycin CDSs

CDS	Function	145113		145698		145711		145716	
		CDS	Id%	CDS	Id%	CDS	Id%	CDS	Id%
CINORF7	Unknown Function	5960	66.7	3137	66.4	2356	68.1	3536	59.1
CINA	Cinnamycin prepropeptide	5961	56.4	3138	56.4	2355	56.4	3535	47.2
CINM	LanM family of proteins	5962	62.0	3139	62.2	2354	61.8	3534	57.2
CINX	hydroxylation of aspartate	5963	53.5	3140	53.3	2353	53.0		
CINT	ABC transporter	5964	79.2	3141	79.0	2352	79.7	3533	65.3
CINH	ABC transporter	5965	65.3	3142	65.0	2351	66.3		
CINR1	*SAR* Protein	5959	55.7	3136	54.7	2357	53.2		

.

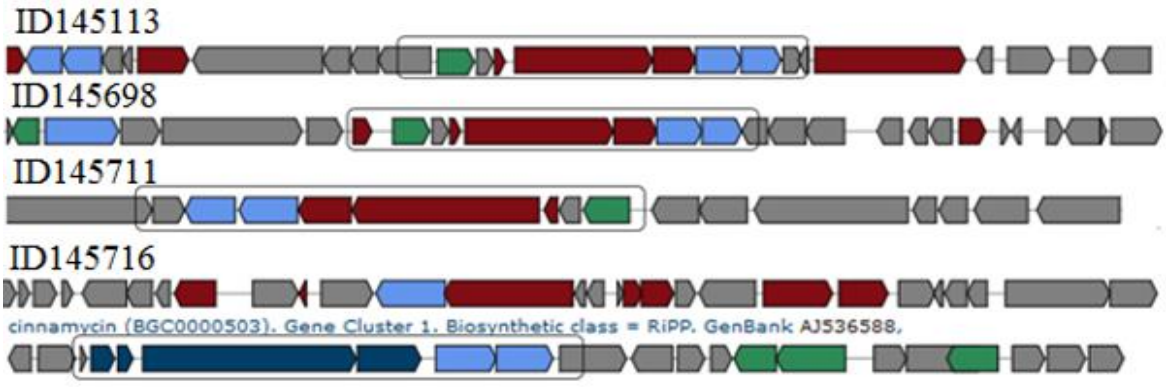


Figure 6.6 – Antismash output comparing the four Actinoallomurus RiPPs gene clusters with the Cinnamycin Cluster. Red/Dark blue – Putative biosynthetic genes; Light Blue – Transport-related genes; Green - Regulation-related genes; Grey – unassigned function

An alignment of the four precursor peptides with CinA is shown in Fig. 6.7. All four *Actinoallomurus* peptides contain the diagnostic Cys and Ser/Thr residues required for making the three (methyl)lanthionine bridges present in cinnamycin, as well as conserved Lys-Ser pair (required for the N-C bond typical of cinnamycin) and Asp residue (site of hydroxylation). However, the peptide from strain ID145716 is predicted to form only methyllanthionine bridges, in contrast to the two methyllanthionine and one lanthionine bridges present in cinnamycin and the other three predicted *Actinoallomurus* lantibiotics (Fig. 6.1, Panel B). It is also worth pointing out that, with respect to cinnamycin, the four *Actinoallomurus* peptides present variations at positions 2 (Gly, Gln or Ala in place of Arg), 3 (Thr in place of Gln), 7 (Trp or Ala in place of Phe), 10 (Ile in place of Phe), 12 (Ile or Ala in place of Phe) and 18 (Ala, Gln or Thr in place of Asn). Overall, the lanthipeptide predicted from strain ID145716 presents itself as the most divergent one with changes (S4T, and V13I) at otherwise invariant positions, plus one extra amino acid (Ala) at the C-terminus. The lack of a putative Asp-hydroxylating enzyme would also indicate that this lanthipeptide is not post-translationally modified at that position.

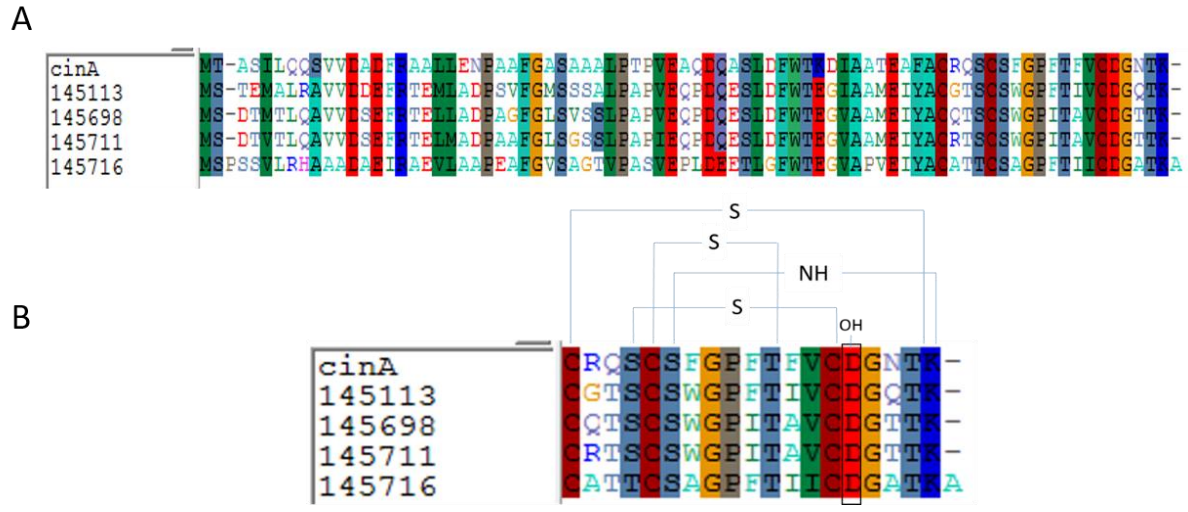


Figure 6.7 – Panel A - Comparison of Cinnamycin leader peptide with correspondent sequences from gene clusters from the four Actinoallomurus Strains. Panel B - Comparison of Cinnamycin Propeptide with correspondent sequences from gene clusters from the four Actinoallomurus Strains. Prediction of Lanthionine bridges is also described

When cultivated under standard conditions (see Chapter 2.4), none of the four strains produce a metabolite with properties compatible with those predicted from the lanthipeptides of Fig. 6.7. Further preliminary investigations are reported later in this Chapter (subchapter 6.3.3). Nonetheless, it is striking that all four sequenced strains contain a related cluster for a lanthipeptide with a predicted ring topology identical to that of cinnamycin, but presenting variations in amino acid sequences.

6.3.2.6. Siderophores

It was also noted that each strain possessed BGCs for a siderophore. Upon a closer look the BGCs from strains ID145698, ID145711 and ID145716 present a high degree of relatedness in gene content and order (Fig. 6.8). A similarity search of the CDS against the MIBiG database is reported in Table 6.8 Interestingly, all four clusters find the BGC for vibrioferrin synthesis as the best match and all encode the four biosynthetic enzymes required for vibrioferrin formation plus an ABC transporter. However, the regions flanking this core 5-gene segments diverge among the four genomes. The low percent identity between the *Actinoallomurus* sequences and those from the vibrioferrin BGC might suggest that the siderophore specified by *Actinoallomurus* is structurally different from vibrioferrin. Alternatively, vibrioferrin has been described as the product of marine *Proteobacteria* (Tanabe et al., 2003), so a low percent identity might derive from phylogenetic distance from *Actinobacteria*. Nonetheless, the finding is interesting and unveils another unexpected feature of the secondary metabolites from *Actinoallomurus*.

Figure 6.8 - Comparison of the gene cluster encoding the biosynthetic machinery for the synthesis Siderophores from Actinoallomurus strain ID145113, ID145698, ID145711, ID145716. Red– CDS Putative biosynthetic genes; Light Blue – Transport-related genes; Green - Regulation-related genes; Grey – un assigned function**.**

Table 6.9 - Description of the Siderophore gene clusters from *Actinoallomurus* strain ID145113, ID145698, ID145711, ID145716.

Most Similar Match | Accession Number	Identity %			
	ID145113	ID145698	ID145711	ID145716
Monensin_3-O-methyl_transferase|AAO65792.1	39,3	---	---	37,7
NapR8|ABS50498.1	---	---	---	36,5
EncO|AAF81734.1	---	---	---	43,9
Putative_GCN5				
-related_N-acetyltransferase|ABX24478.1	51,9	---	---	---
DNA_helicase_IV|CBJ79907.1	---	39,3	---	---
CalU19|AAM70361.1	48,9	---	---	51,1
CalR8|AAM70335.1	66,7	---	---	71,2
Vibrioferrin_biosynthesis_protein_PvsE|ABM30201.1	58,7	42,9	43,2	42,8
Vibrioferrin_biosynthesis_protein_PvsD|BAC16547.1	32,4	31,4	31,8	31,6
Vibrioferrin_biosynthesis_protein_PvsB|ABM30204.1	31,7	34,4	31,9	31,0
Vibrioferrin_biosynthesis_protein_PvsA|BAC16544.1	37,0	36,7	37,6	35,8
ATP-binding_phosphonate_ABC_transporter_PhnC|ACG70837.1	39,2	32,4	31,1	31,6
Putative_aminoacylase|BAE95562.1	---	---	36,3	37,2
Peptidase_M20|AEH42498.1	---	32,3	---	---
NTP_pyrophosphohydrolase|ACB47032.1	---	40,5	---	---

6.3.3. Creation of a genomic library from Strain ID145113

Temporally, the first *Actinoallomurus* genome was obtained from strain ID145113 (see Chapter 3). I was interested in learning techniques for efficiently creating genomic libraries with inserts large enough to contain some clusters identified in strain ID145113, such as the BGCs for the aromatic polyketide NAI-113 (see Chapter 3) and the cinnamycin-like lanthipeptide (see above). Thus, I spent my second secondment (April 2015) at the University of Warwick, where the expertise and adequate conditions exist for library preparation and screening.

The creation and screening of a genomic library consists of: preparation of high quality DNA from the chosen strain; construction of the genomic library in a vector suitable for further

steps; and screening the library for the desired clones. Since a well-made library can be an important asset to use over and over, appropriate conditions are important for storing and easily retrieving the individual clones. It was thus decided to adopt a technique that allows screening several clones at the same time and permits storing in a small amount of space. Libraries are usually constructed and stored in microtiter plates (96-well or more) where each clone is placed in a different well; screening is then performed going well by well until the desired clone is found. This library storage system allows a rational organization of the library but makes the process of library screening and storing time- and space-inefficient. This might not be an issue when making a single library, but can become a limiting factors when dealing with many different libraries or with metagenomic libraries, which require a large number of clones. Gurgui and Piel have employed a strategy in which bacterial clones are stored in a semi-liquid medium at low CFU/mL. (Gurgui and Piel, 2010) As shown in Fig. 6.9, this strategy can be easily adapted to *Actinoallomurus*, enabling the isolation of a single clone from a 400,000 clone library in significantly lower amount of time.

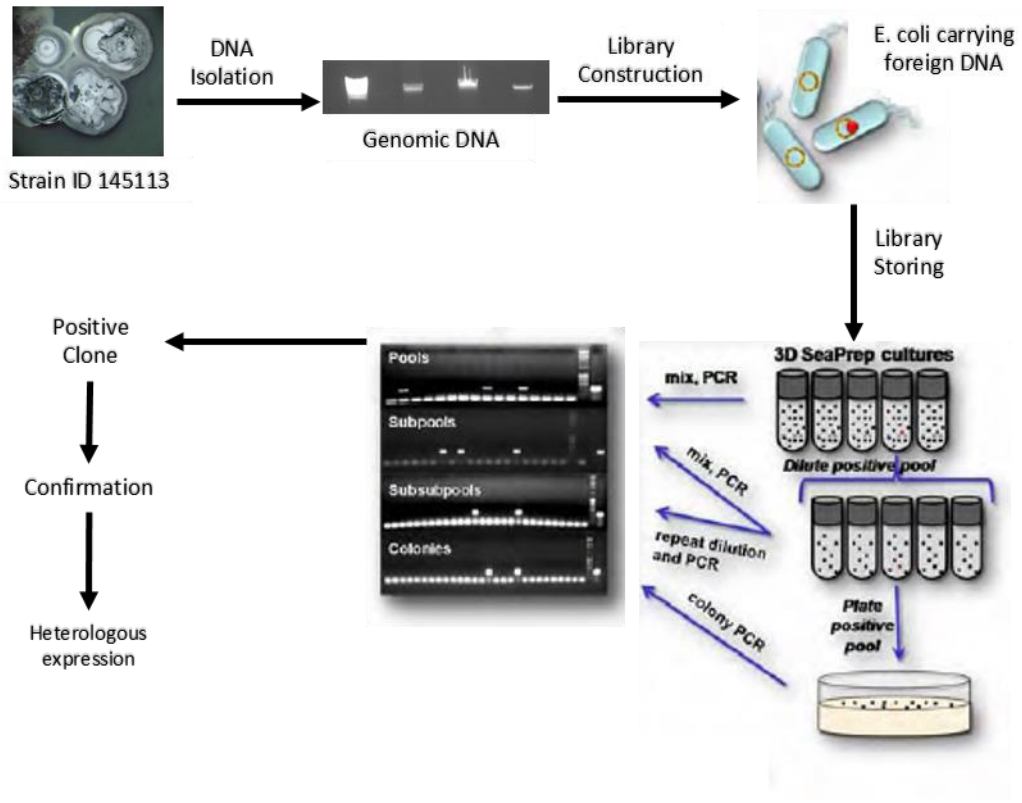


Figure 6.9 – Workflow used for obtaining heterologous production of gene clusters from *Actinoallomurus*. Adapted from Gurgui, C. and Piel, J., 2010. (Gurgui and Piel, 2010)

The technique involves storing the library in separate vials of roughly 10,000 clones per mL each. A small sample from each vial can then be used for PCR screening. From the positive pools, appropriate dilutions are made into separate vials, which are then incubated for regrowth and subjected to a second PCR analysis, and so forth until the number of different clones per vial is low enough to be plated. At this point a final PCR is made to screen individual colonies. This way it is possible to significantly reduce the amount of storage space (one cryo vial with 10,000 clones instead of 100 96-well microtiter plates) and rapidly finding the desired clones in a process taking four to five days (Hrvatin and Piel, 2007). Although this technique is particularly suited for large libraries requiring millions of clones for appropriate coverage, it was also adapted to serve the need of my project.

The number of clones needed for having a given probability of given DNA sequence from a genome can be calculated with according to Equation (1).

N=ln(1-P)/ln(1-f), N – Number of clones; P- Desired probability; f – insert size (1)

For this work, we used the fosmid vector pCCFOS, which combines all the stability and ability of harbouring larger inserts granted by the low-copy BAC vectors, but are endowed with a system that allows a high number of copies of the vector per host cell by induction with l-arabinose. For this particular work we used pCC2FOS whom has also a primer sequence to help the end sequencing after vector screening. (Wild et al., 2002) With a genome size of 10.6 Mbp for strain ID145113, 1400 clones would be sufficient to give a 99% probability of finding any gene according to formula (1). However, since we were looking for not one but two clusters (for NAI-113 and for the cinnamycin-like lanthipeptide) which each means that we were not looking just for one gene but to a set of genes, all in the same fragment. To be on the safe side, the library size was scaled up to 300,000 clones, which translates into a total 105Gbp of inserts, or roughly 10,000 genome equivalents. This was achieved as described under Experimental Section.

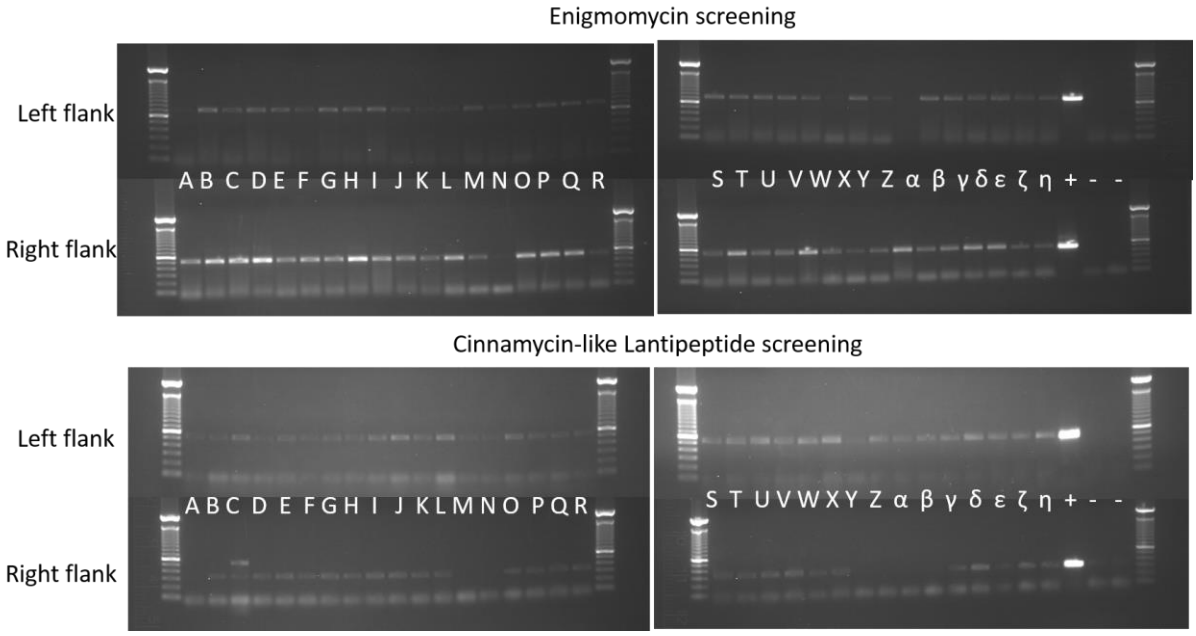


Figure 6.10 – Agarose gel with the PCR results from the screening of the super-pools. Top layer – Amplicons from Left and right flanks of the NAI-113 Cluster. Bottom layer Amplicons from Left and right flanks of the Cinnamycin-like Cluster. Positive Control (+) Total genomic DNA. Negative Controls (-) *S.coelicolor* total genomic DNA and Water.

The first round of PCR screening was made pooling 10 µL each from 10 separate vials, yielding about 10,000 clones. Needless to say that several positive pools were found at this stage (Fig. 7.10). On the basis of the observed amount of amplicon, super-pools D and H, and J were selected for the NAI-113 and lanthipeptide screening, respectively.

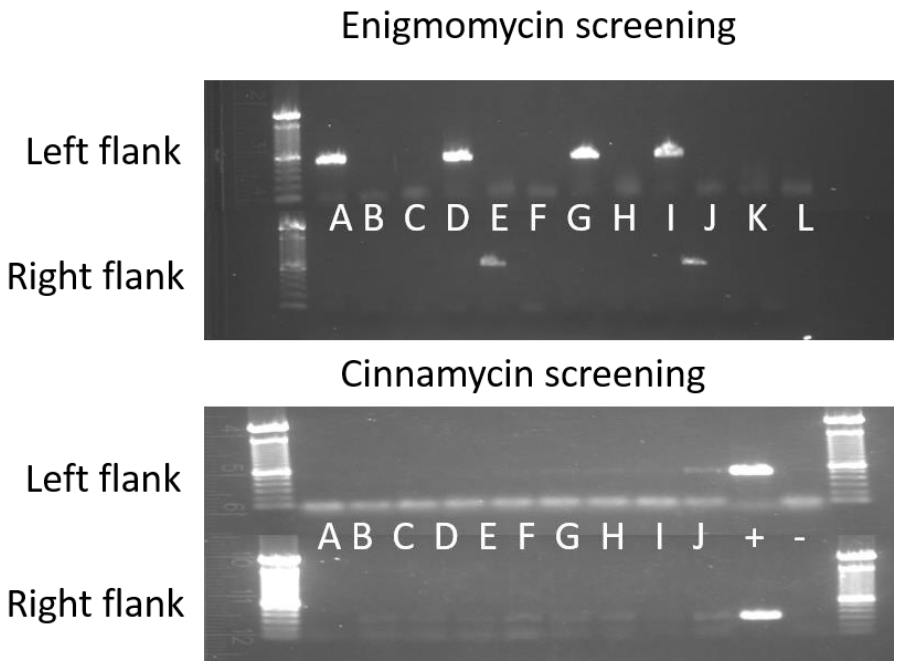


Figure 6.11 – Agarose gel with the PCR results from the screening of the pools. Top layer – Amplicons from Left and right flanks of the NAI-113 Cluster. Bottom layer Amplicons from Left and right flanks of the Cinnamycin-like Cluster. Positive Control (+) Total genomic DNA. Negative Control (-) *S. coelicolor* total genomic DNA

Upon dilution, regrowth and PCR screening (according to the scheme of Fig. 6.11), none of the two different super-pools yielded sub-pools containing both ends of the NAI-113 cluster (Fig. 6.12), implicating that a larger pool size was needed to identify a complete 33-kb cluster. However, due to time constraint, no further work was performed. In contrast, different subpools were PCR-positive for both ends of the Cin-like cluster (Fig. 6.11- vial F, G, H, and J). After one further round, four clones containing a DNA fragment encompassing the entire cluster were found (Fig. 6.11). End sequencing of the resulting fosmids confirmed that the insert contained the entire cluster, with flanking regions of 5 and 1 kbp for a calculated insert size of 33.5kbp.

Back to KtedoGen, the identified cosmid has been engineered for conjugation into different actinomycetes for heterologous expression, work that is currently ongoing (Dr. Arianna Tocchetti, personal communication).

6.4. Discussion

The genome characterization provided another interesting dimension to this work allowing to have a better picture of what *Actinoallomurus* can produce and where it falls within the grand scheme of microbial classification. This last part of the analysis was performed by looking at the predicted proteome and compare it to those of known organisms. As expected, the four *Actinoallomurus* strains share a large number of orthologues (Table 6.3). Strikingly, however, they share a similar number of orthologues with members of the *Streptosporangiaceae* (between 41 and 53%) and not with members of the *Thermomonosporaceae* (29-36%) which would be expected being all strains closely related, according to the phylogenetic analysis of the 16S rRNA gene. Moreover, *Actinoallomurus* strains seem to share a high homology with *T. curvata* (another member of the *Thermomonospora*ceae family) and with those of the *Streptosporangeacea* family. *S. coelicolor*, *A. oligospora* and *A. friulensis* seem to present less homologs and by consequence a higher percentage of singletons and paralogs. It is also noteworthy how *A. oligospora* – the

closest relative of *Actinoallomurus* – shows such a diminished set of homology with the remaining organisms.

Taxonomy suffers always from an original sin that is being an artificial creation from Man to organize Nature in a systematic way, which does not seem to work in such manner. The orthology analysis shows again this apparent conflict between classification based on phylogenetic analysis of 16S rDNA gene sequences and strain relatedness based on the number of shared orthologues. These results suggest that the assignment of *Actinoallomurus* to the *Thermomonosporaceae* family and in particular its closeness to *Actinomadura* is worth further investigations. An analysis based on a single taxonomic marker might not reveal the complete evolutionary history of this organism and it would be interesting to perform phylogenetic analyses using other known markers to secure the phylogenetic position of *Actinoallomurus*. At the same time, the same orthology analysis provided other insights, for instance the high number of unique proteins found in each *Actinoallomurus* strain suggests a large pangenome where different BGCs are likely to be found.

Regarding the novelty of the metabolites predicted from genome analysis, a few insights were obtained, for instance from the KSβs that clustered away from other known enzymes. Nevertheless, further investigations to actually isolate and correlate a compound with these BGCs are required.

One of common characteristics of the metabolites from *Actinoallomurus* so far seems to be the presence of halogens and hence the presence of halogenases (see Chapter 6.3.2.3.). A phylogenetic analysis of the six halogenases identified in the four genomes allowed to separate the enzymes into two groups. One consists of the halogenases associated with the allocyclinone cluster and the related cluster from strain ID145711, together with the additional halogenase from strain ID145698. These five enzymes cluster together with the halogenases from chlorinated aromatic polyketides such as allocyclinones and chlortetracycline. The second group includes the halogenase identified in strain ID145716, which cluster with pyrrole-modifying halogenases. A least an additional halogenase is predicted from strain ID145698, which produces a lantibiotic carrying a halogenated tryptophan (see Chapter 5) but whose genome sequence has not yet been determined. Thus, further analyses of *Actinoallomurus* strains may substantially expand the phylogenetic positioning of the halogenases present in this genus. In the end, it would be worth comparing the frequency of halogenated compounds found

in other actinomycete genera to evaluate the relative frequency with which they are encountered in different taxa.

The genetic analysis found two BGCs (one for a cinnamycin-related lanthipeptide and one for a siderophore) common to the four strains. Since none of the strains was known to produce such molecules, these "unsuspected" compounds might play a role in the *Actinoallomurus* life cycle, thus justifying their presence across all the strains. Siderophores are known to play an important role in metal chelation and acquisition, so the conservation of one such BGC comes as no surprise. However, the presence of a cin-like BGC in all four strains is intriguing, especially since each of the *cinA* structural genes has slightly diverged among the different strains, accumulating mutations at amino acid positions not involved in post-translational modifications. This finding, which contrasts with the conservation of the NAI-107 sequence observed between *Microbispora* and *Actinoallomurus* (see chapter 5), suggests that these BGCs were not recent acquisitions but were rather maintained throughout this genus speciation. However, the role of these lanthipeptides in the *Actinoallomurus* life cycle remains to be established. It should be noted that the detection of these potentially common metabolites most likely would require modification in the cultivation conditions. For instance, siderophore production is known to require the use of media with limiting iron concentrations (Dimkpa et al., 2008).

7. GENERAL DISCUSSION

7.1. Introduction

The end of the antibiotic era is coming at a rather unexpected pace. The inefficiency of the most powerful antimicrobial agents is increasing and we, as a society, have not been able to introduce effective therapeutic alternatives. Moreover, the struggle to achieve new efficacious antibiotics is uncanny: between the difficulty of finding novel molecular structures and the requirements of the regulatory entities, antimicrobial drug discovery and development has become a laborious endeavour with an extremely low output and a perceived low reward. Therefore, new successful projects in drug discovery must replenish the pipeline of antibacterial drug candidates.

Historically, natural products have represented one of the most valuable sources of compounds for drug discovery. However, the elevated frequency of re-discoveries of known antibiotics and the incrementally high difficulty to identify new structures led to the assumption that Nature was close to exhaustion as the most successful sources of therapeutic agents and alternative sources had to be sought. However, alternative sources have substantially failed to deliver drug candidates at a reasonable rate while we have systematically assessed the chemical repertoire of a small portion of living organisms, bacteria in particular, leaving a universe of novel sources to be analysed.

As the pharmaceutical industry was progressively abandoning natural product screening, scientific and technological advantages were profoundly changing the way that natural products could be identified, characterized and manipulated: the accessibility of genomic information and the ability to handle it; major improvements in analytical techniques; and an understanding of biosynthetic pathways involved in secondary metabolism. The –omics age made available a set of tools: from the production of synthetic cell factories (e.g., artemisinin is a plant metabolite that can now be produced in yeast through a reconstructed pathway (Ro et al., 2006)) to the direct expression of environmental DNA in a surrogate host (*e.g.*, production of the polyketide arixanthomycin by *Strep. albus* containing soil DNA)(Kang and Brady, 2014); from correlating metabolome profiles through MS networking (Yang et al.,

2013) to the production of rapamycin analogues through mutasynthesis (Seto, 2012). Complementing these advances are improved methods in synthetic chemistry, which enable the creation of an improved range of semi-synthetic derivatives and *de novo* synthesis of natural products (Lam, 2007; Ling et al., 2015).

These are all interesting approaches that are having an impact in the identification, production and manipulation of natural products. However, they are bound to our knowledge of existing compounds, which in turn derives from exploring a limited fraction of bacterial diversity. Hence, the exploration of novel (i.e., previously not systematically screened) *taxa* is likely to yield new biosynthetic enzymes, new combination thereof, thus new molecular structures and, possibly, new interesting bioactivities. Recent examples in this direction bear promise (Feling et al., 2003; Piel et al., 2004).

Among bacteria, some taxa have become versatile producers of bioactive metabolites, in particular many, but notably not all, taxa belonging to the *Actinobacteria*. Therefore, the main premise of this work was to use phylogeny as a predictive marker of the ability to produce secondary metabolites: unexplored taxa within the *Actinobacteria*, which are phylogenetically related to known producers, would have tuned and tweaked their physiology to better adapt to the particularities of the selective pressure to which have been subjected and, since secondary metabolism plays an important role in this evolutionary process, would have come up with new chemical space. For the reasons explained in Introduction, the focus of this project has been on the poorly explored genus *Actinoallomurus*.

7.2. Critical overview of the results

In this work 105 strains, out of 230 strains tested under one single condition (same medium and incubation parameters for all strains), produced an antibacterial activity against one single test strain, *Staphylococcus aureus*. This result indicates that *Actinoallomurus*, much as other *Actinobacteria*, is geared for the production of bioactive metabolites under artificial laboratory conditions, *i.e.*, shake-flasks.

When evaluating a potential new bacterial taxon for antibiotics, one can encounter two extreme scenarios: no positives are found against the chosen test organism, which implies that either the taxon does not produce antibiotics or that a different test organism must be used; or virtually all samples are positive, which implies that no screening has been made, since all

strains need still to be analysed through the next step in the process. In the latter case, the screening strategy needs a revision, especially if it is to be applied to a large number of samples. This is somehow the issue encountered in this project, in which about 50% of the strains possessed an antibiotic activity against *S. aureus*, so a large number of positives needed to be evaluated. However, this is a nice problem to have: now that *Actinoallomurus* has proven to be a prolific producer of antimicrobial compounds, additional tools can be implemented to allow prioritising the positive extracts.

While less than half (45 out of 105) of the positive extracts were actually evaluated for the chemical nature of the active compound(s), a reasonable diversity was observed among the characterized extracts. Regarding novelty, two new antibiotics (NAI-113 and allocyclinone) were identified, as novel representatives of known families possessing interesting structural features. In addition, the rediscovery of a known lantibiotic (NAI-107) spurred the production of a variant (NAI-108) with enhanced antibacterial activity. Furthermore, to those compounds, several new polyethers were pinpointed but, for the reasons explained in Chapter 0, were not further pursued.

Although the frequency of new discoveries is expected to lower as the number of screened strains increases, it is worth mentioning the NAICONS strain collection used in this work includes 1200 independently isolated *Actinoallomurus* strains, so the 45 strains analysed in this work still represent a modest fraction of that microbial diversity.

Among the molecules produced by strain ID145698, we identified the first hyper-halogenated aromatic polyketide. Named allocyclinone, these angucyclinones present also an uncommon lactone group fused to the chromophore and up to three chlorines on a single methyl group, with the number of chlorines being related to antibacterial potency. Allocyclinones, although previously unreported, might not be uncommon metabolites within the genus *Actinoallomurus*, since this family of metabolites was identified in phylogenetically distinct *Actinoallomurus* isolates. Allocyclinones represent an additional example of halogenated metabolites produced by *Actinoallomurus*. Indeed, this genus produces spirotetronates with multiple halogens in the pyrrole moiety (Mazzetti, C., *et* al, 2012) and a lantibiotic with a halogenated tryptophan residue (Chapter 5).

Sequentially next to strain ID145698 there is strain ID 145699, which was identified as the producer of NAI-107, a chlorinated lantibiotic effective against multi-drug resistant Gram-positive pathogens and previously reported from the distantly related genus *Microbispora*. Inclusion of KBr in the production medium of either the *Actinoallomurus* or the *Microbispora* producer readily afforded brominated variants of NAI-107, which were designated as NAI-108, while the other post-translational modifications naturally occurring in this lantibiotic family (i.e., hydroxylation of Pro-14 and C-terminal decarboxylation) were unaffected by the presence of a brominated tryptophan. In addition to being the first example of a bromine-containing lantibiotic, NAI-108 displayed a small but consistent improvement in antibacterial activity against all tested strains. It should be noted that while substitution of chloride with bromide has been successfully achieved in two instances, NAI-414 (Mazzetti et al., 2012)and NAI-108 (chapter 5), this is not a general rule and it did not yield any brominated metabolites in the case of strain ID145698, the allocyclinone producer (data not shown).

An additional contribution has been on the characterization of NAI-113. While the producer strain ID145113 had been previously marked as potentially interesting, significant hurdles in structural elucidation led the project to a halt. The major hindrance for structure elucidation was the paramagnetic nature of the molecule, which emitted no NMR signals even after simple derivatization. My work consisted in providing ^{13}C-labelled NAI-113 and insights into NAI-113 biosynthesis through analysis of the corresponding BGC. These results, along with a number of chemical degradation studies, ultimately led to establishing that NAI-113 is a new anthracyclinone, carrying an unusual lactone moiety. For both allocyclinone and NAI-113, bioinformatic predictions suggest that these compounds are synthesized as longer polyketide precursors.

The first genomic analysis of the *Actinoallomurus* genus confirmed expectations for filamentous actinomycetes: the four sequenced strains contain large genomes (8-10 Mbp) with approximately 70% GC and with a significant fraction dedicated to BGCs. The four sequenced genomes provided a total of 3.7 Mbp dedicated to secondary metabolism subdivided into 88 BGCs with a minimum of 19 clusters per strain. With some exceptions, each *Actinoallomurus* strain presents a diverse set of clusters, and every strain harbours at least one BGC for each of the major biosynthetic classes – PKS, NRPS, RiPPs and terpenes. In terms of abundance, terpenes appear to be the most common, followed by RiPPs and polyketides.

7.3. Insights into *Actinoallomurus*

As previously mentioned the *Actinoallomurus* strains used in this screening were isolated previously to beginning of this project. *Actinoallomurus* spp. are slightly acidophilic filamentous actinomycetes that can be differentiated by a different cell wall and 16S rRNA gene sequence. The first described species actually resulted from a reclassification of some strains that were previously assigned to the genus *Actinomadura.* (Tamura et al., 2009) *Actinoallomurus* has thus been assigned to the family *Thermomonosporaceae*, together with *Actinomadura* and *Thermomonospora* spp. This family belongs to the order *Streptosporangiales*.

Taking advantage of the four *Actinoallomurus* sequenced genomes, an orthology analysis was performed comparing their deduced proteomes with those derived from representatives of other *Actinobacteria*. This analysis has shown that, in terms of shared homologues, Actinoallomurus strains are as related to members of the *Streptosporangeaceae* as they are to *Thermomonospora curvata*. In contrast, the number of shared homologues with *Actinomadura oligospora* is lower. While these results must await confirmation through a comparison with additional genome sequences from *Actinomadura* (only one species was used), nonetheless they suggest that *Actinoallomurus* might be as related to members of the *Streptosporangiaceae* as of the *Thermomonosporaceae*.

As mentioned in the introduction a screening expedition into new taxonomical space goes beyond the obvious task of bringing to light new compounds. As compounds new and known are characterized, the work provides insights into the physiology of the new taxon. Only few metabolites were published from *Actinoallomurus* prior to the beginning of the work. (Mazzetti et al., 2012; Pozzi et al., 2011) The discovery of allocyclinones and NAI-113 confirms the existence of new chemical space within *Actinoallomurus*, particularly the presence of previously unreported features in relatively widespread classes of metabolites. These metabolites, in particular allocyclinones, are relatively common among *Actinoallomurus*. So are polyether-producing strains. Therefore, the work presented here provides a glimpse on some commonly encountered metabolites that, although not universally present, may constitute common encounters in this genus, as for example actinomycin and aminoglycosides are for *Streptomyces* and *Micromonospora*, respectively (Běhal, 2000; Mingeot-Leclercq et al., 1999).

Some of the possible core metabolites of *Actinoallomurus* have been unveiled by genomic analysis. Indeed, all four analysed genomes harbour a BGC for the synthesis of a vibrioferrin-like siderophore and one for a cinnamycin-like lanthipeptide. Siderophores play an important role in bacterial physiology, as they bind iron (or other heavy metals) with high affinity and facilitate the intracellular uptake of this essential micronutrient. While all actinomycete genomes harbour one or more siderophore BGC, only strains from the streptomycetaea family have been reported as possessing a BGC (but with no metabolite associated so far) for the biosynthesis of Vibrioferrin-like compounds which, in fact is a siderophore previously only found in some *Proteobacteri.* (Ichikawa et al., 2010; Nett et al., 2009).

Even more remarkable is the finding of the cinnamycin-like BGCs. Cinnamycin is a class II lanthipeptide previously described from *Streptomyces* spp. Lanthipeptides were long-believed to be species-specific, but this concept has already been challenged in this work by the production by *Actinoallomurus* of a highly modified lanthipeptide previously observed only in *Microbispora* spp. The finding that all four genomes harbour BGCs for producing cinnamycin-like molecules, with considerable variations in their amino acid sequences, suggest that these lanthipeptides play an important and yet unknown role in *Actinoallomurus* physiology. Although neither the vibrioferrin-like siderophore nor the cinnamycin-like compounds were identified in the screened extracts, particular conditions (e.g., Fe-limited media for the siderophore) might be required to trigger expression of the corresponding BGCs.

Halogenated metabolites are often encountered among the *Actinoallomurus* metabolites. Combining previous and current work, three structurally unrelated Cl-containing metabolites have been characterized: spirotetronates carrying halogens on a pyrrole moiety, (Mazzetti et al., 2012) hyper halogenated angucyclinones (Chapter 5), and NAI-107 (Chapter 6). Furthermore, genomic analyses led to the identification of two additional BGCs encoding halogenases: one for an allocyclinone-related metabolite in strain ID145711, and one for a metabolite of unknown function in strain ID 145716.

7.4. The concept of low-hanging fruits

The exploration of a new region will always start by registering the main, easier to be acknowledge features. The same experience is to be expected from a screening program in a

new taxonomical region. Just like during the initial screening of *Streptomyces* strains, the first encounters are those that strike the observer the most, whether because they are produced in high quantities or because they have a potent activity. Furthermore, those "striking encounters" must occur at a relatively high frequency to catch the observer's attention (and prevent him/her to switch to a different project!). Metabolites with these characteristics can be called the low hanging fruits in drug discovery from microbial products, in analogy to the easy-to-collect fruits out of a tree. As different trees bear different fruits, so one would expect different taxa to bear different metabolites. Since this work has been the first systematic evaluation of *Actinoallomurus*, it is not unexpected that, among the first molecules identified, are the low hanging fruits from this genus. The characteristics of allocyclinone fit well to those of a low hanging fruit: firstly, it is produced in high quantities and it has high antimicrobial activity, including against the *E. coli* mutant used in this work, a fairly rare event that made this compound stand out in the observer's eye with respect to the others; secondly, the high number of chlorines present in the active molecule was another feature that arose interest. Finally, several producers were found among the tested strains, hence augmenting the probability of allocyclinone producing strains being picked up randomly. This same reasoning could be applied to the putative polyethers identified but not further pursued during the screening.

The results presented here suggest that further screening of *Actinoallomurus* strains will not yield a single class of compounds but the genus can offer a broad set of different compounds, although it appears to be enriched in some types of molecules. From the results observed so far, polyketides are probably the most common class, especially polyethers through the PKS I system and aromatic polyketides through PKS II. However, it is striking that, for both NAI-113 and allocyclinone, genomic evidence suggests that these molecules are produced from longer dodeca- or tridecaketide backbones, which are then somehow shortened into the final product. This observation, together with the relatively high frequency of halogenated polyketides, suggests that, even within a class of microbial products with several known representatives, new chemical twists are to be found. The highly divergent chain length factors identified through genomic sequencing may represent another potential source for novel chemistry, although no metabolite has been associated to the corresponding BGCs.

Judging from the genomic sequences, additional low-hanging fruits might be represented by the vibrioferrin-like siderophores and the cinnamycin-like lanthipeptides.

However, the appropriate conditions necessary to trigger the expression of the corresponding BGCs currently represent *terra incognita*.

7.5. The genomic potential of *Actinoallomurus* *vs* the screening results

The characterization of the metabolites produced by 45 strains and the identification of 88 BGCs from genomic analysis of four strains provide the basis for a preliminary comparison of the frequency of metabolites encountered with the two approaches. Interestingly, the two types of data only partially match. Most noticeable is the case of terpenes, one of the most frequent type of BGCs observed but with no terpene detected in the analysed extracts. This discrepancy may be due to the fact that these molecules did not show a detectable antibacterial activity (actually, few terpenes from actinomycetes do), that they are often volatile molecules that can be lost during extract preparation or that the corresponding BGCs were not expressed under the employed cultivation conditions. The second most frequent class of metabolites predicted by Antismash is represented by RiPPs, a class from which wet-lab experiments yielded one compound (the lantibiotic NAI-107). The third most frequent class from genomic analysis is represented by the polyketides, and indeed several polyketides were actually identified during screening. Nevertheless, there is a discrepancy between screening data and genomic analysis: while the indicated high frequency of polyethers, and hence of PKS1 clusters, only one such BGC could tentatively be assigned to polyether production.

The above mentioned discrepancies should not come as a surprise. Screening of microbial fermentation extracts requires that compounds are produced under the chosen experimental conditions and that they are stable enough to be detected at the time extracts are actually prepared. If a bioassay is used to select extracts worthy of further analyses, an additional bias is introduced, i.e. that the metabolite is active. These limitations might help explain the lack of identified terpenes in the extracts and the bias for polyether and aromatic polyketides, which usually possess potent antibacterial activities. On the other hand, the BGCs identified from genomic analysis might not encode a functional biosynthetic pathway, might require special conditions for their expression or the cultivation medium might lack essential precursors or cofactors.

7.6. Finding new chemical classes and finding improved antibiotics

The probability of bringing to light a revolutionary discovery is usually inversely related to the amount of knowledge available in the area of interest. For instance, discovering an unprecedented chemical class during screening of microbial products reflects this axiom quite well, since this was a common feature in the golden era of antibiotic discovery and can now be expected a rare event. Furthermore, the major biosynthetic pathways (i.e. those for polyketides, non-ribosomal peptides, RiPPs and terpenes) are common to all bacterial phyla that harbour the corresponding BGCs. Hence, the discovery of unprecedented biosynthetic classes can be expected only from the lesser abundant classes of secondary metabolites.

As our knowledge increases, the features of a chemical class of secondary metabolites can be interpreted more diverse ways, whether in terms of the molecular target(s) it interacts with or of the potential applications. As mentioned, several compounds initially identified for their antibacterial or antifungal properties became successful drugs for treating cancer or as immunosuppressants in organ grafts. As another example, polyketide macrolactones can bind to the bacterial ribosome (e.g., erythromycin), to eukaryotic tubulin (e.g. epothilone) or they can inhibit bacterial and eukariotic tRNA synthase (e.g. borrelidin) (Goodin et al., 2004; Igarashi et al., 1969; Wakabayashi et al., 1997).

Not all discoveries are expected to impact our knowledge in the same manner, as this depends on their ability for surprising us. At first sight, the discovery of another angucyclinone might attract as little interest as the discovery of a new polyether, since both events would add a small extent of incremental knowledge to large families of compounds with a well-established biosynthetic pathway. However, some unique features, such as the presence of a trichloromethyl group linked to an aromatic ring or the existence of two halogenases in the allocyclinone cluster might change this perspective.

One of the advantages in screening microbial products is that one might encounter improved variants of chemical families with a limited number of representatives. Such was the case that led to the discovery of NAI-108, the brominated variant of the potent lantibiotic NAI-107, which could be made by either *Actinoallomurus* or *Microbispora* in KBr-supplemented media. This finding indicates that the halogenases specified by the *mib* clusters from

Microbispora and by the corresponding cluster from *Actinoallomurus* sp. ID145699 are capable of efficiently incorporating Br when exposed to sufficient bromide concentrations, leading to the first example of a Br-containing lantibiotic. Although this substitution is interesting from the biochemical point of view, it also resulted in a compound slightly but consistently more active than the chlorinated counterpart against different bacterial pathogens. While this result can be explained by the importance of the hydrophobicity in the lipid II-binding lanthionine bridges of class I lantibiotics (Maffioli et al., 2015), many brominated metabolites are usually less active than their chlorinated homologues (Cruz et al., 2015).

The theoretical chemical space is estimated to be immense, and we have information on an infinitesimal fraction. Nevertheless, this fraction contains millions of compounds, with just few of them used as therapeutics. Expanding on this set to bring benefits to our society requires introducing new molecules that present enhanced or unexpected bioactivity properties within the corresponding chemical class.

7.7. Outlook

The results presented here present only the first insights on the screening of 230 strains from a library of 1200 *Actinoallomurus* isolates. If the activity results observed herein are to be extrapolated, then under the same conditions around 600 strains will produce an extract with activity against *S. aureus*. Identifying the active compound(s) within each of those theoretical 600 active extracts requires implementing a robust screening pipeline with strict threshold criteria for further progress. At the same time, such a high number of positives might require introducing more stringent assay criteria, for examples tests able to filter out frequently encountered metabolites. As additional strains are evaluated, the probability of finding new compounds is expected to decrease, eventually reaching a point of saturation in which a large screening effort is required to identify new molecules. So far, out of 45 extracts analysed, 11 different sets of active compounds were identified (9 polyethers and two aromatic polyketides), with each set often represented by a complex of related molecules. This high frequency of novelty events does encourage further investigations within this genus and suggests that the point of low return of *Actinoallomurus* screening is unlikely to be just around the corner. This assumption is corroborated by the genomic analysis, with several predicted classes of metabolites that have not been identified during the screening exercise.

To keep improving our health standards and to fight antibiotic resistance, we need to identify new drug candidates. Nature has provided a large number of them and all the available evidence indicates that it can continue to do so, provided that judicious choices are made in the source of chemical diversity to be explored and in the approach taken to explore it. This requires a continuous adjustment of conditions, especially dealing with microbial strains, since we have little information on the role these compounds play in the environment for the producing strains. In the case of *Actinoallomurus*, future efforts may require adjustments of the screening assay, to make it more selective, and exploration of different cultivation conditions, to identify some genome-predicted metabolites. Overall, I am confident that *Actinoallomurus* can play an important role in bringing to light and, possibly, into clinical use new antibiotics and other bioactive compounds.

APPENDIX 1

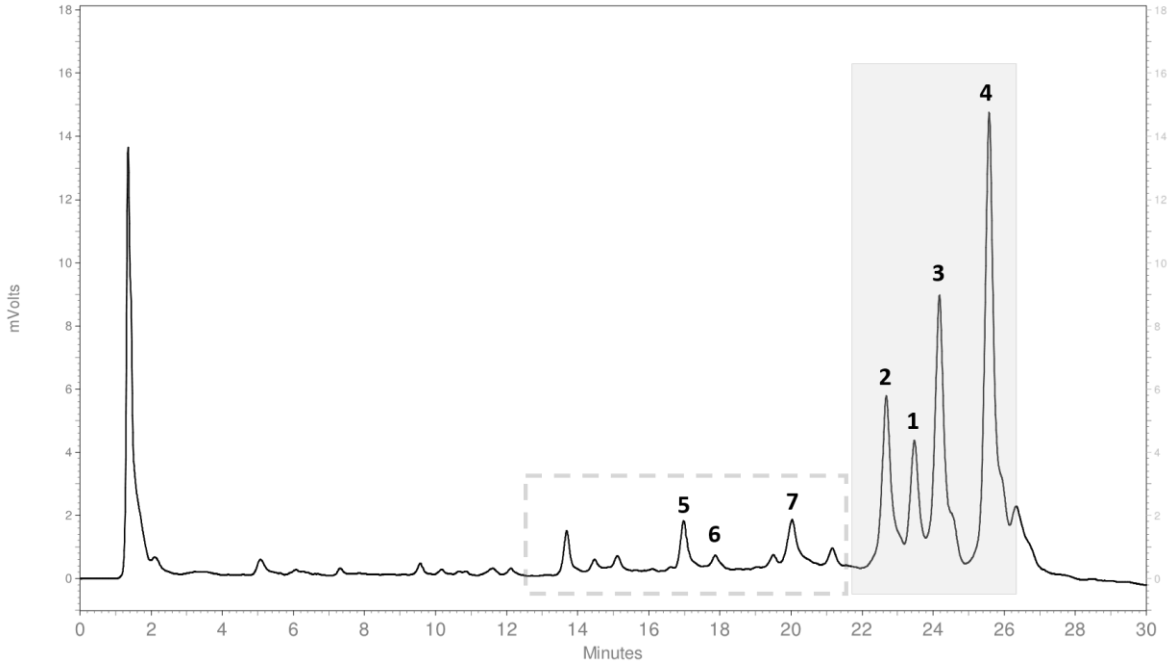


Figure 1 - HPLC trace at 270 nm of the mycelium extract from a culture of *Actinoallomurus* sp. ID145968. The shaded are indicates peaks associated with antibacterial activity, where peaks corresponding to compounds **1**, **2**, **3** and **4** are labelled accordingly. The dashed box highlights the hydrophilic part of the chromatogram containing the partially characterized compounds **5**, **6** and **7**, as well as other uncharacterized species related to **1**–**4**.

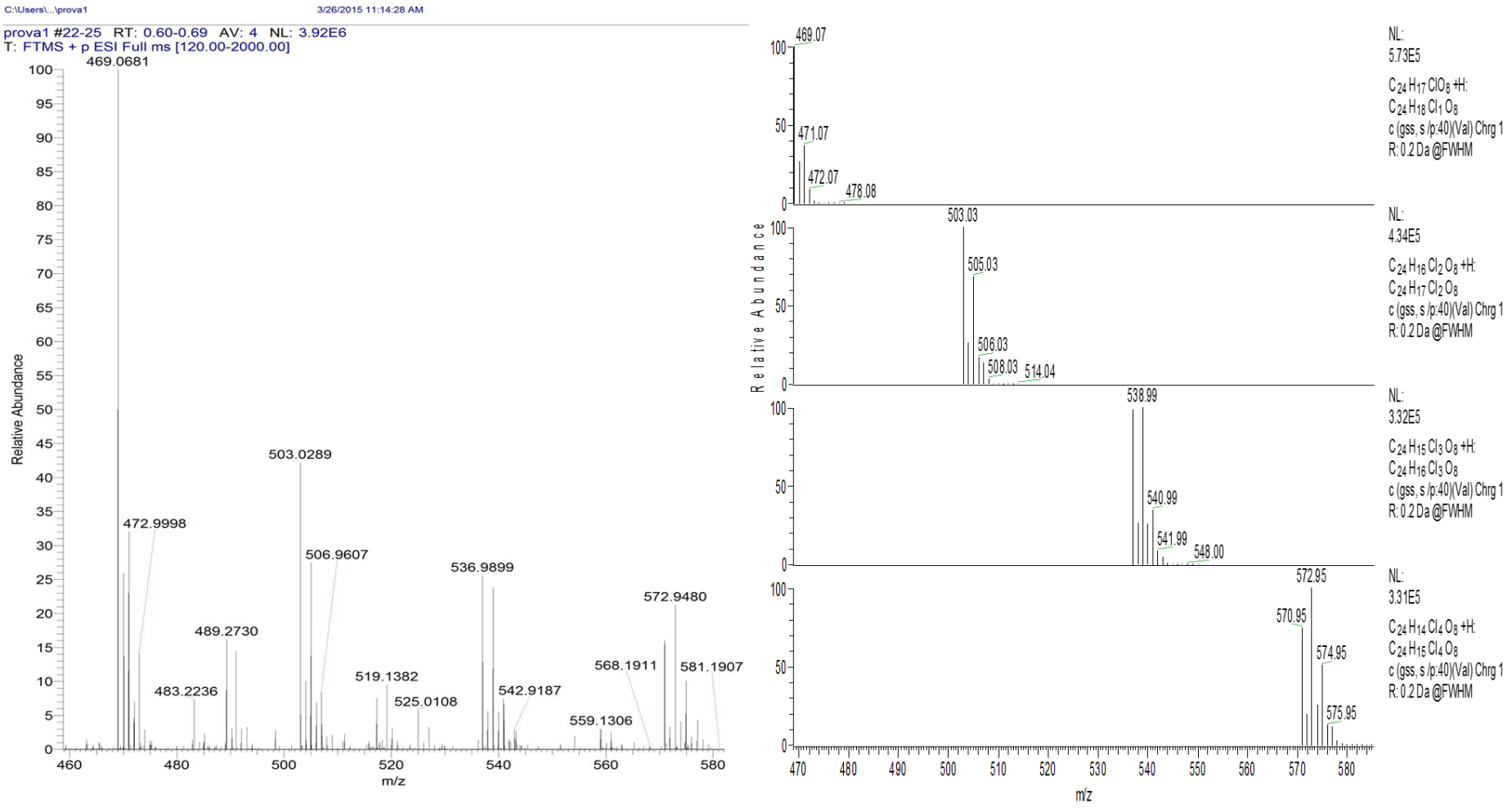


Figure 2 - Left: Experimental HR-MS of compounds **1**-**4** (from left to right**).** Right: calculated isotopic patterns for **1**-**4** (top to bottom).

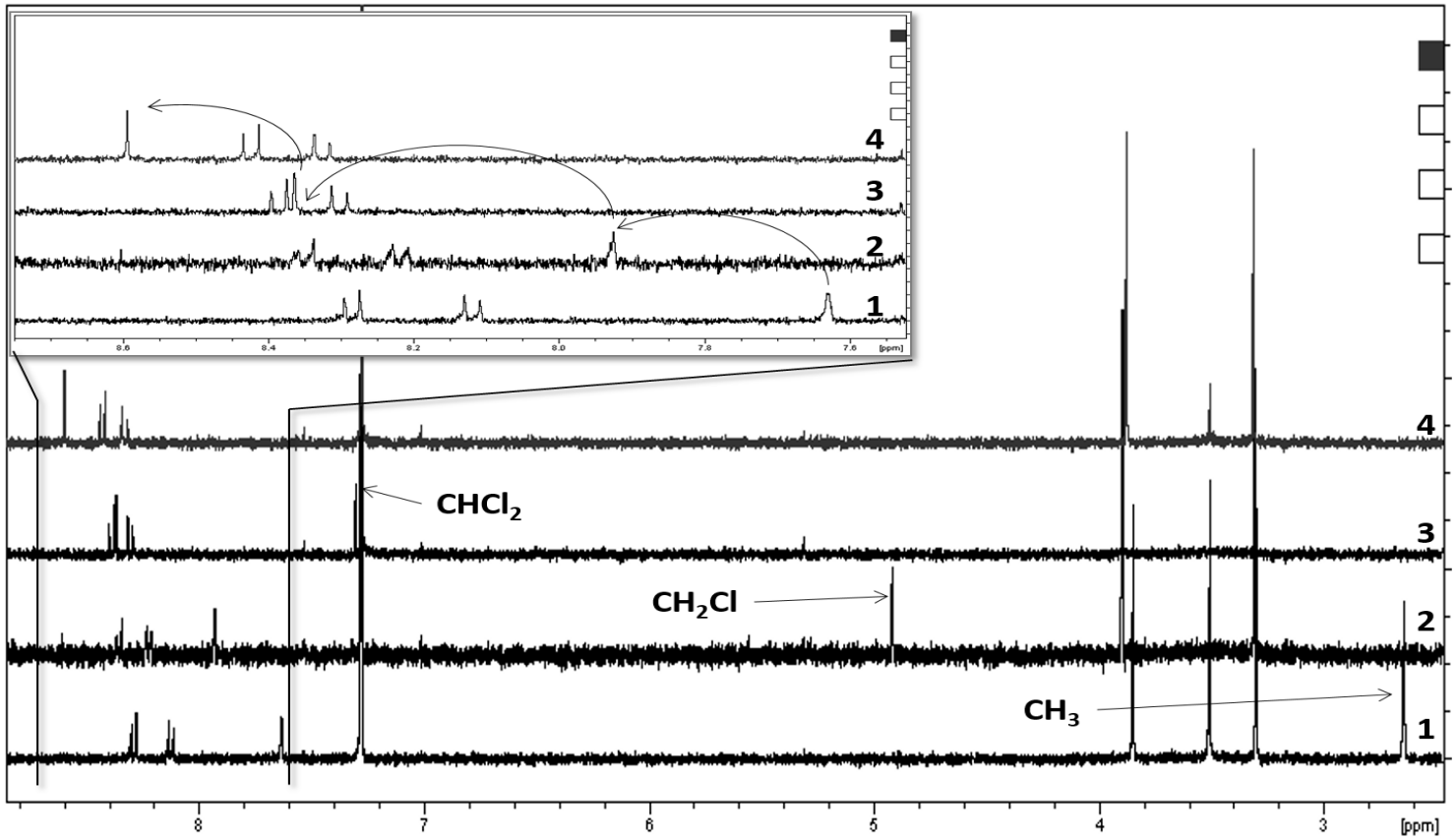


Figure 3: Comparison of the ^{1}H-NMR spectra of Allocyclinone **1**,**2**,**3**, and **4** in $CDCl_3$ at 298K. The variation at position 3 is clearly visible: the methyl group of Allocyclinone **1** at 2.66 ppm becomes CH_2Cl at 4.92 ppm in **2** and $CHCl_2$ at 7.29 (near the solvent peak) in **3**. In the upper expanded panel the arrows indicate the downfield shift for the aromatic proton in position 4 due to the increasing deshielding effect of the vicinal substituent. In the same panel the less dramatic shift of the AB system is also visible.

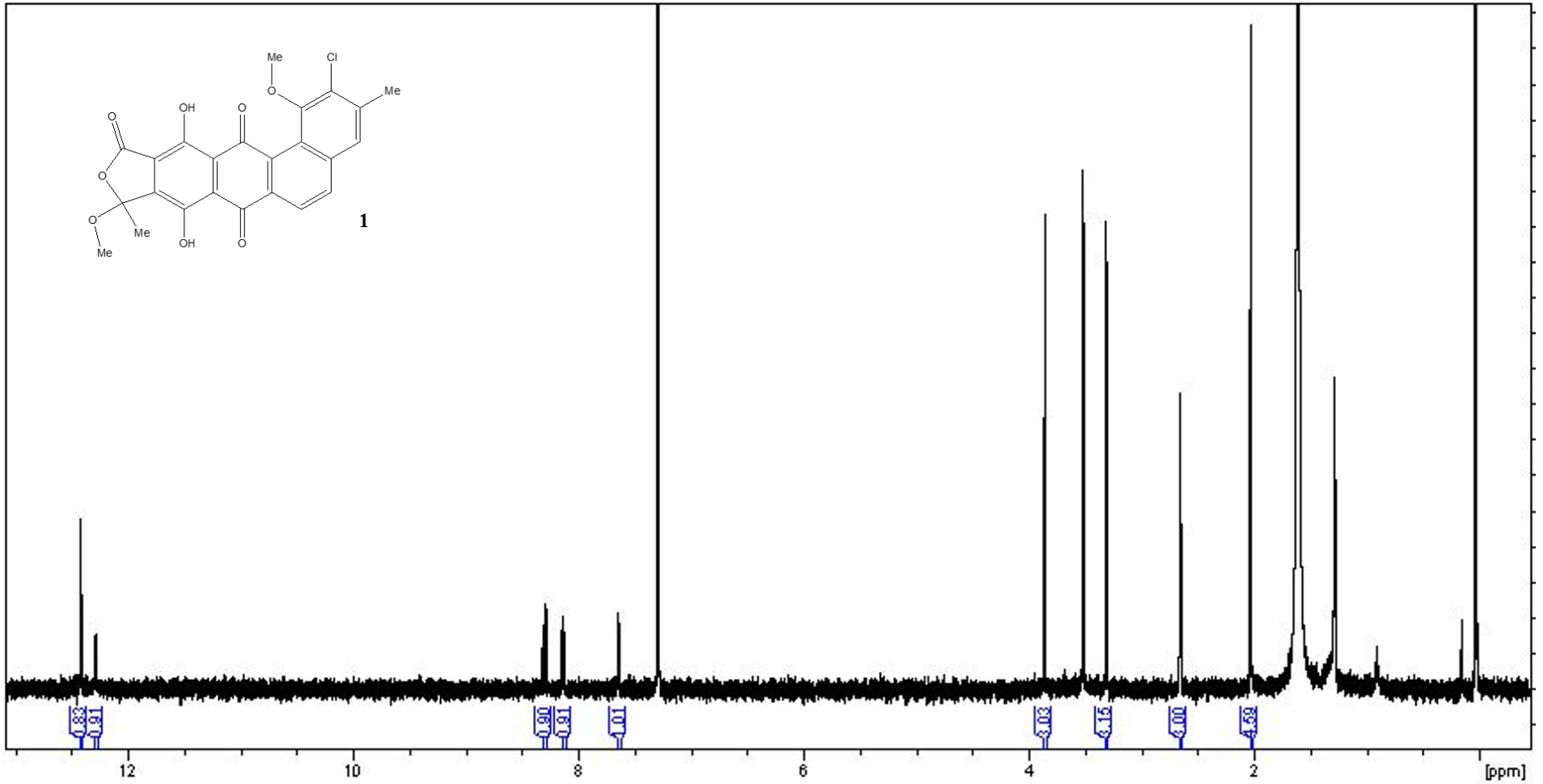


Figure 4. ^{1}H-NMR of **1** in $CDCl_3$ at 298K.

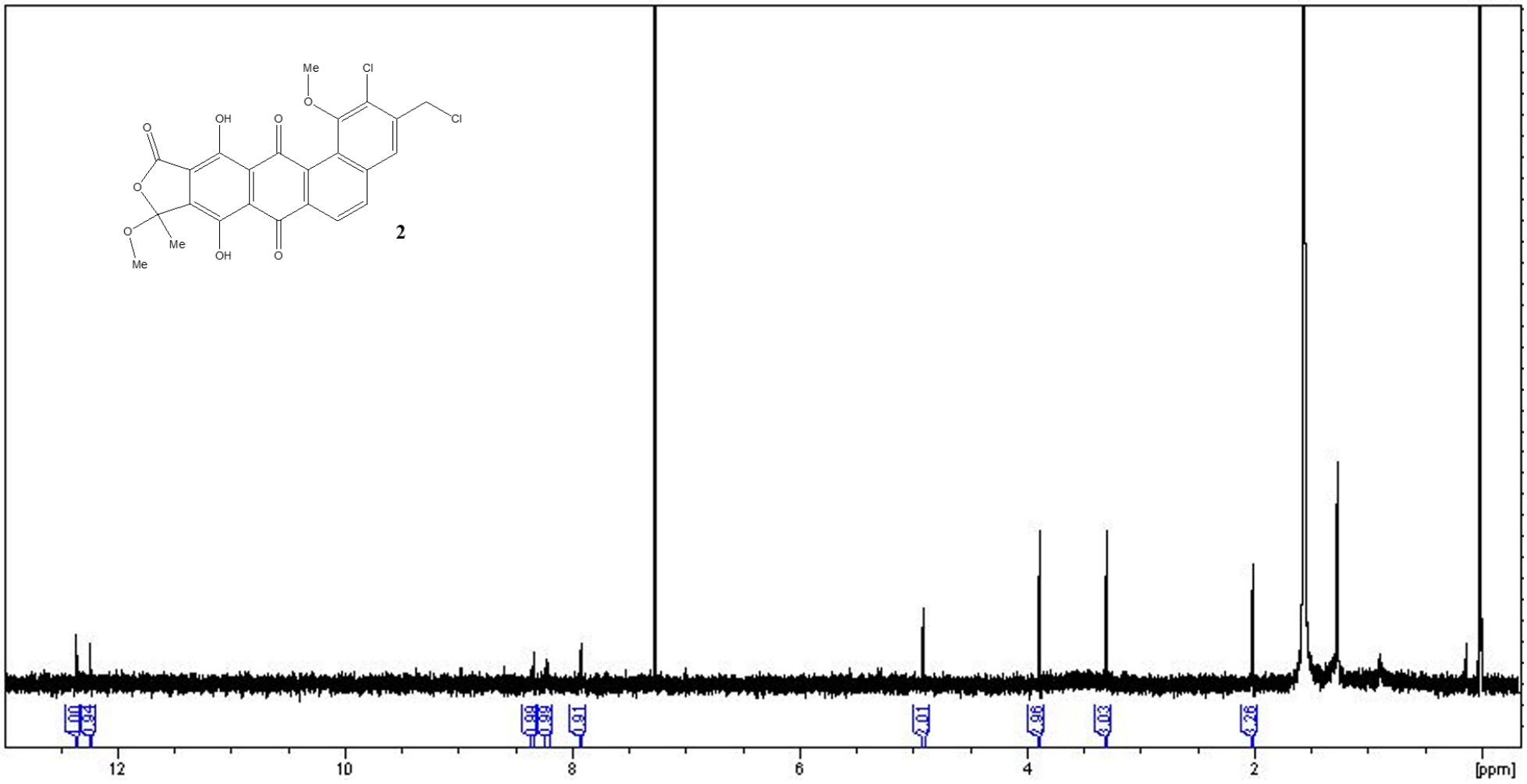


Figure 5. ^{1}H-NMR of **2** in $CDCl_3$ at 298K.

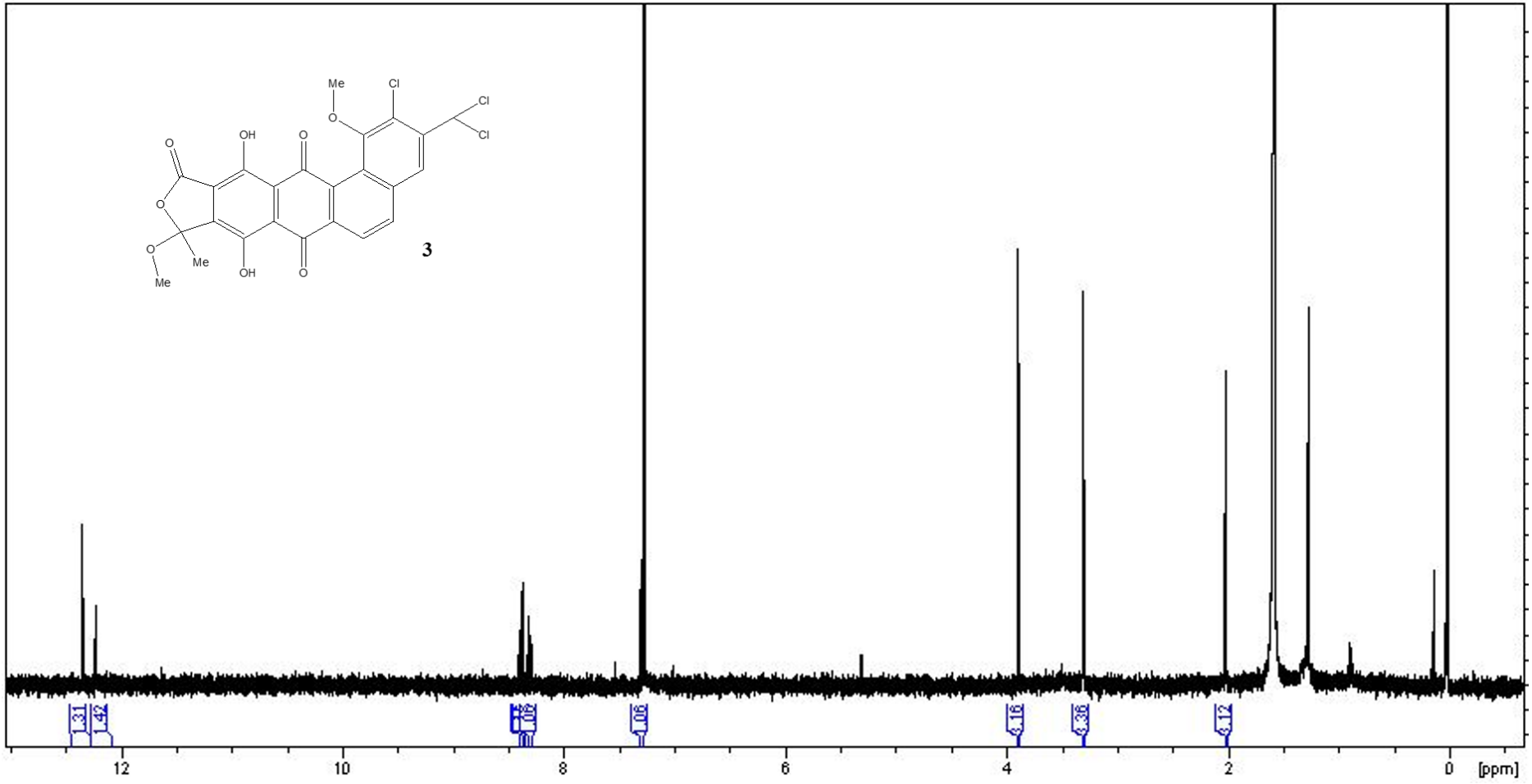


Figure 6. ^{1}H-NMR of **3** in $CDCl_3$ at 298K.

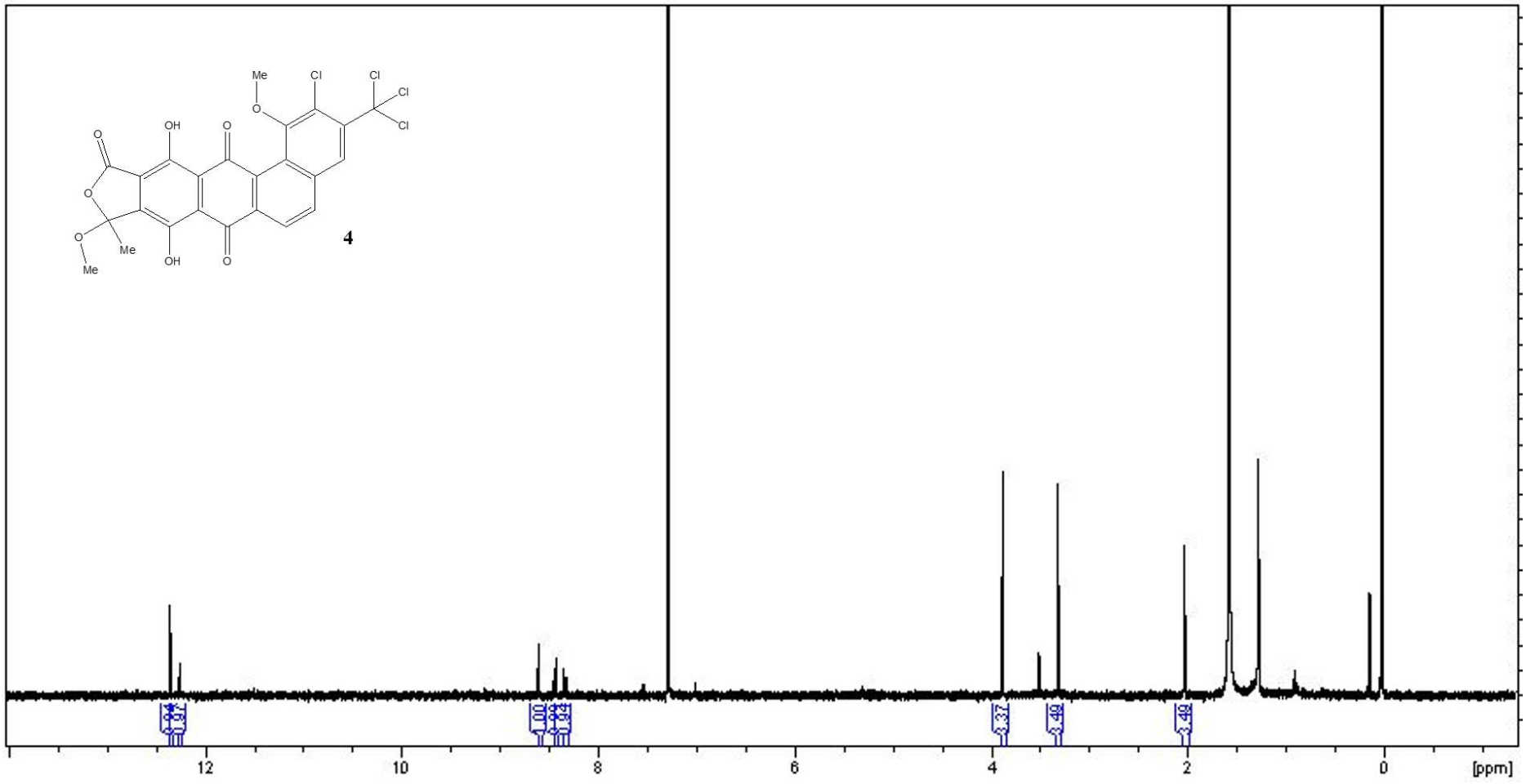


Figure 7: ^{1}H-NMR of **4** in $CDCl_3$ at 298K.

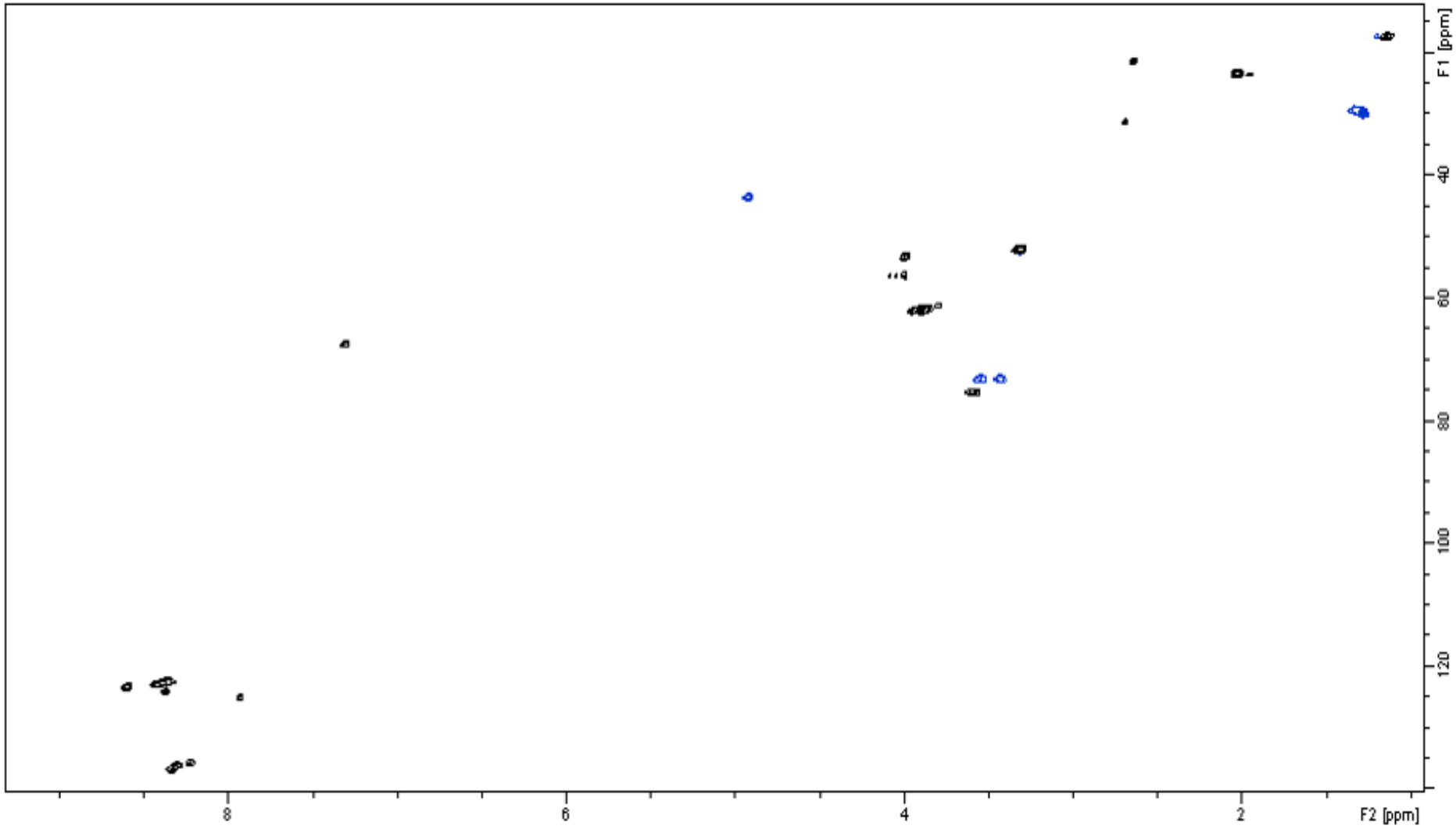


Figure 8. HSQC NMR of the **1**–**4** mixture in $CDCl_3$ at 298K.

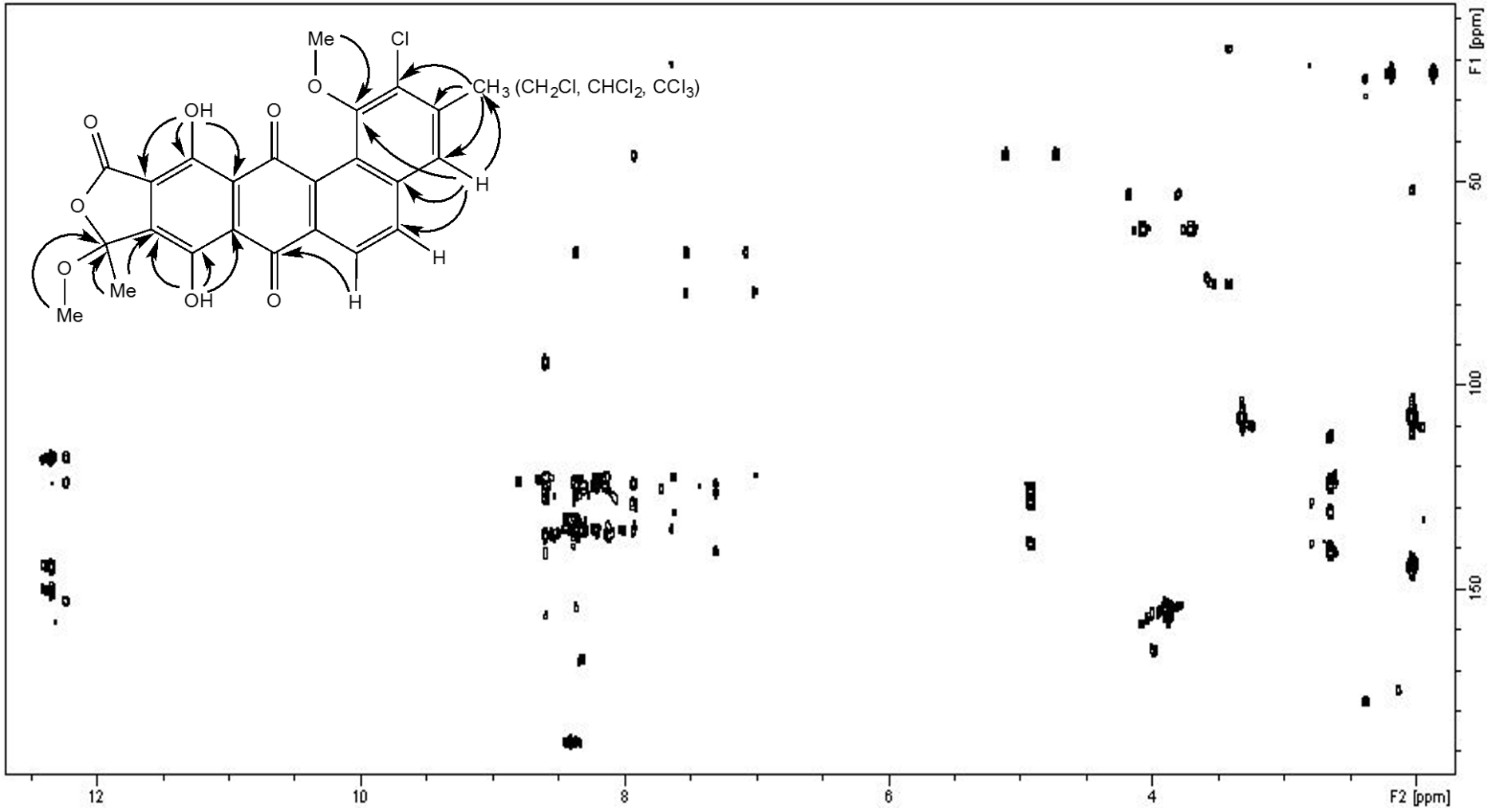


Figure 9: HMBC NMR of the **1**–**4** mixture in $CDCl_3$ at 298K.

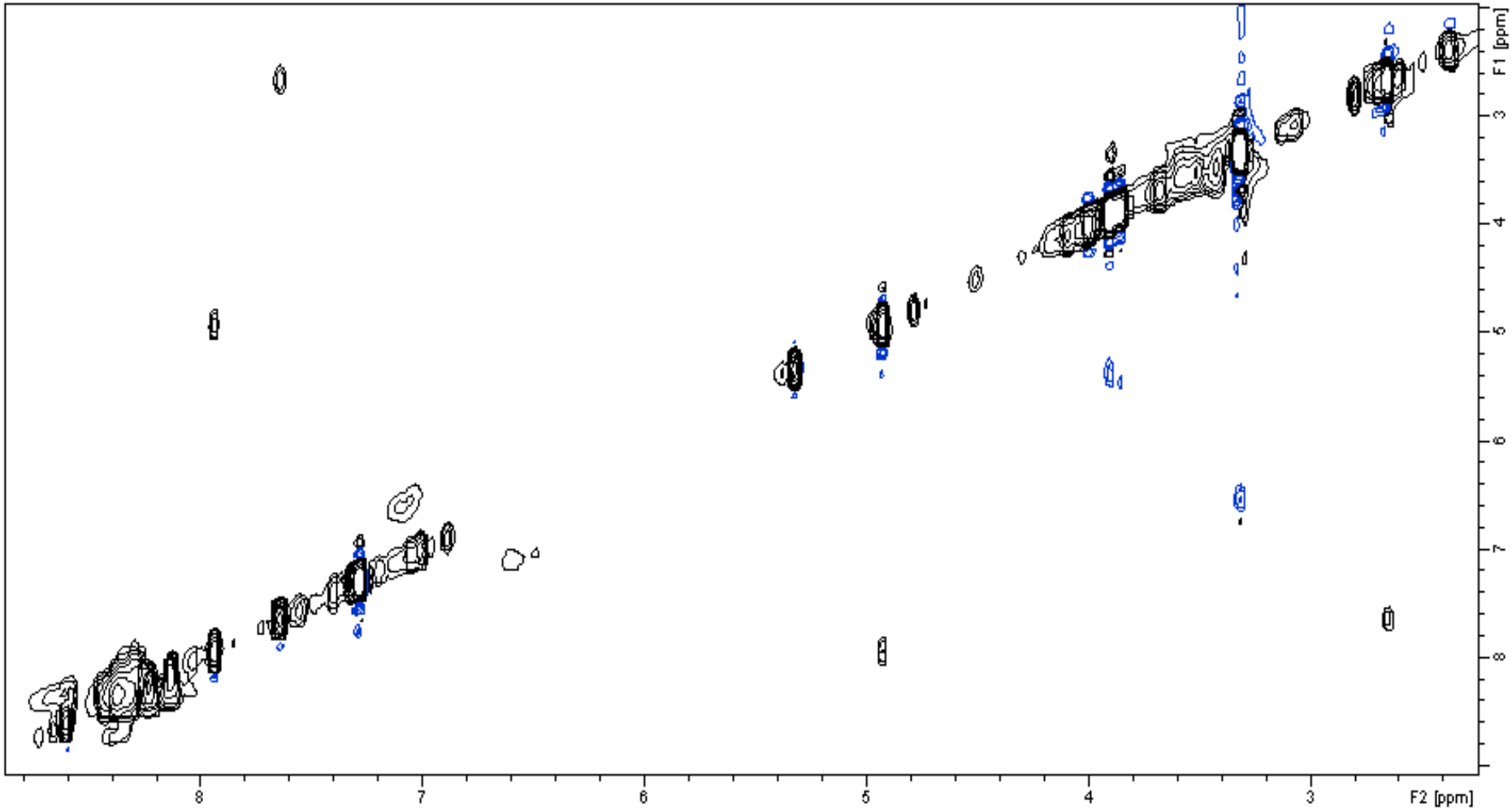


Figure 10: TOCSY NMR of the **1**–**4** mixture in $CDCl_3$ at 298K.

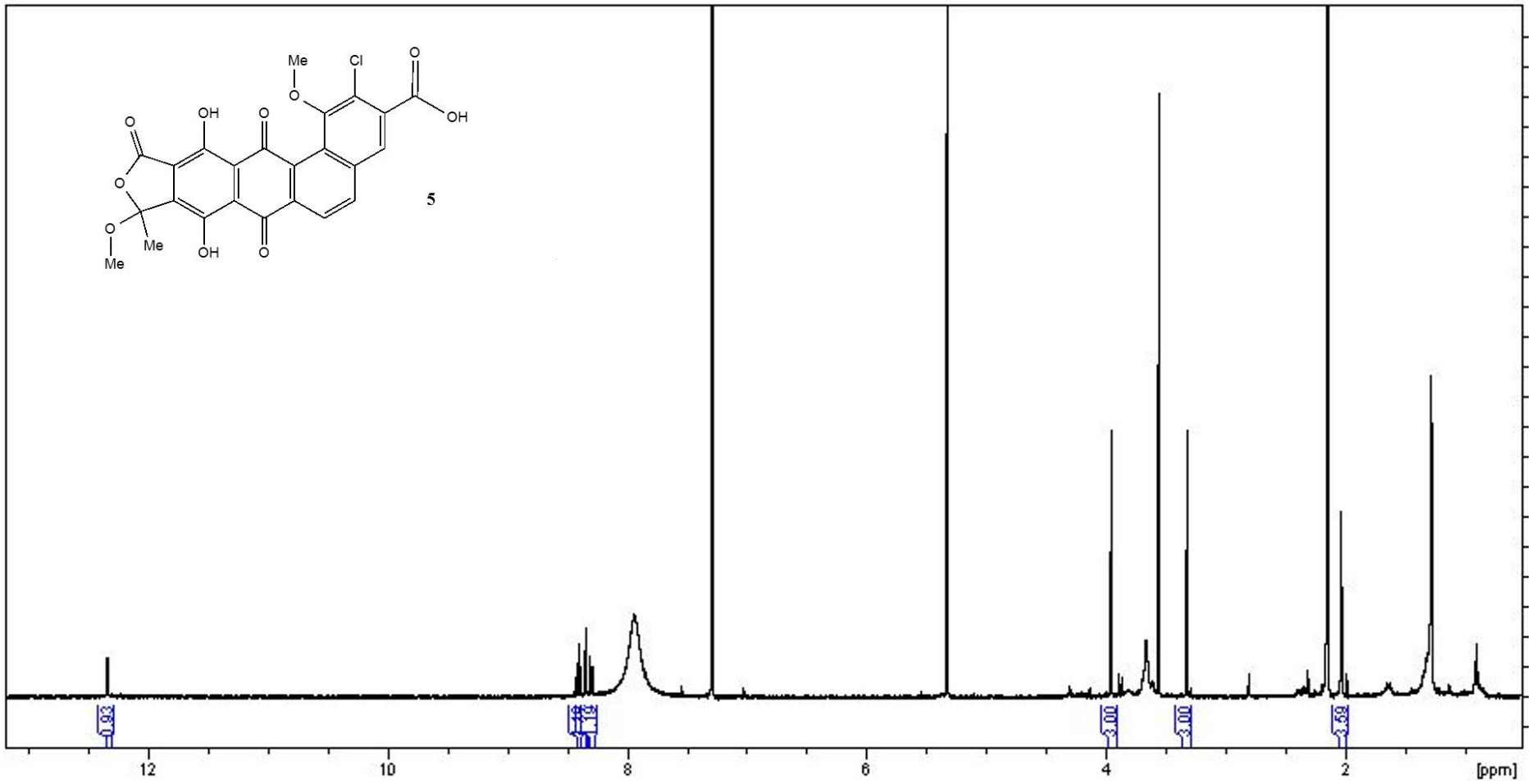


Figure 11: ^{1}H-MMR of **5** in $CDCl_3$ with addition of 10µL TFA at 298K.

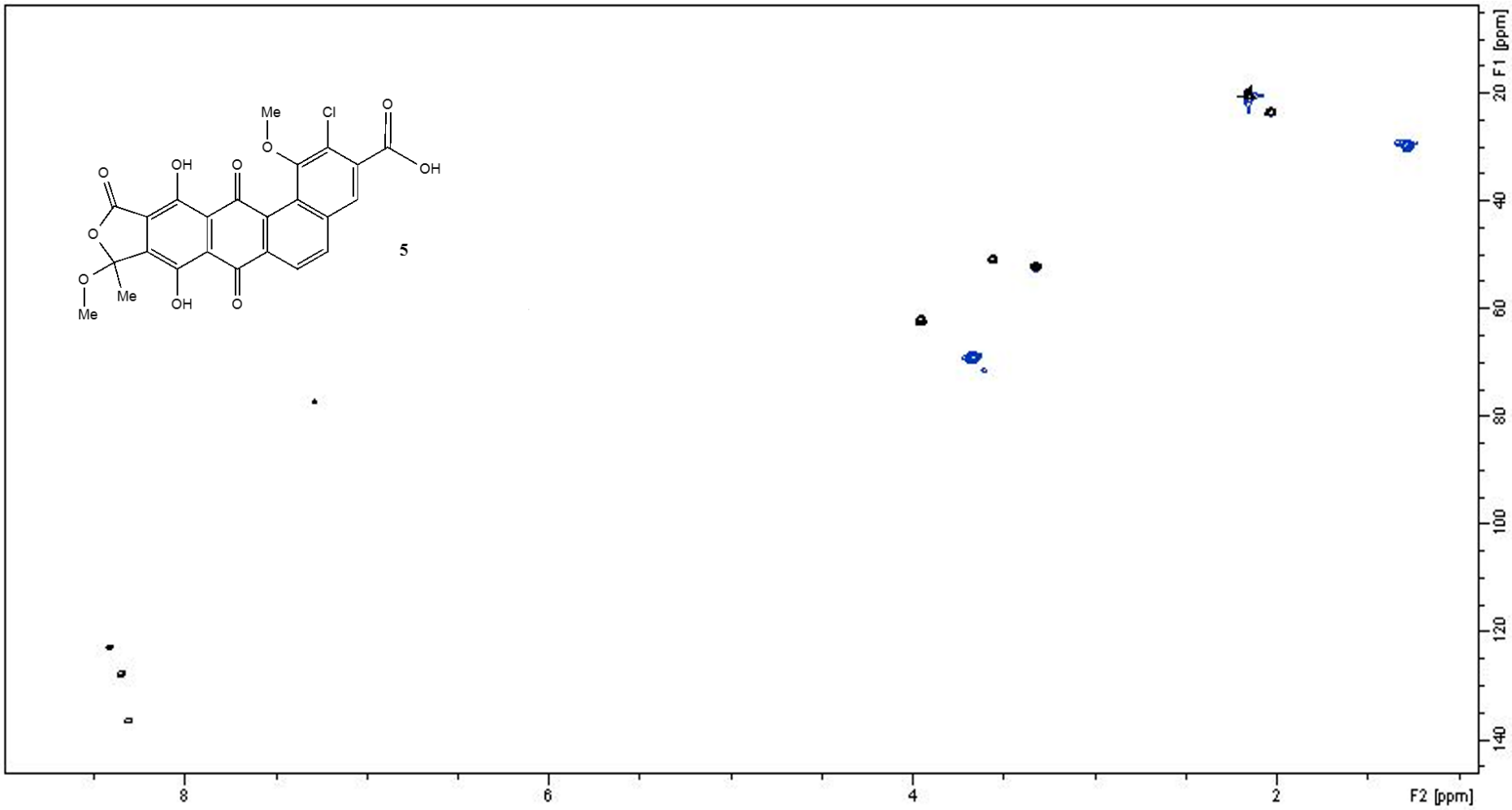


Figure 12: HSQC NMR of **5** in $CDCl_3$ with addition of 10µL TFA at 298K

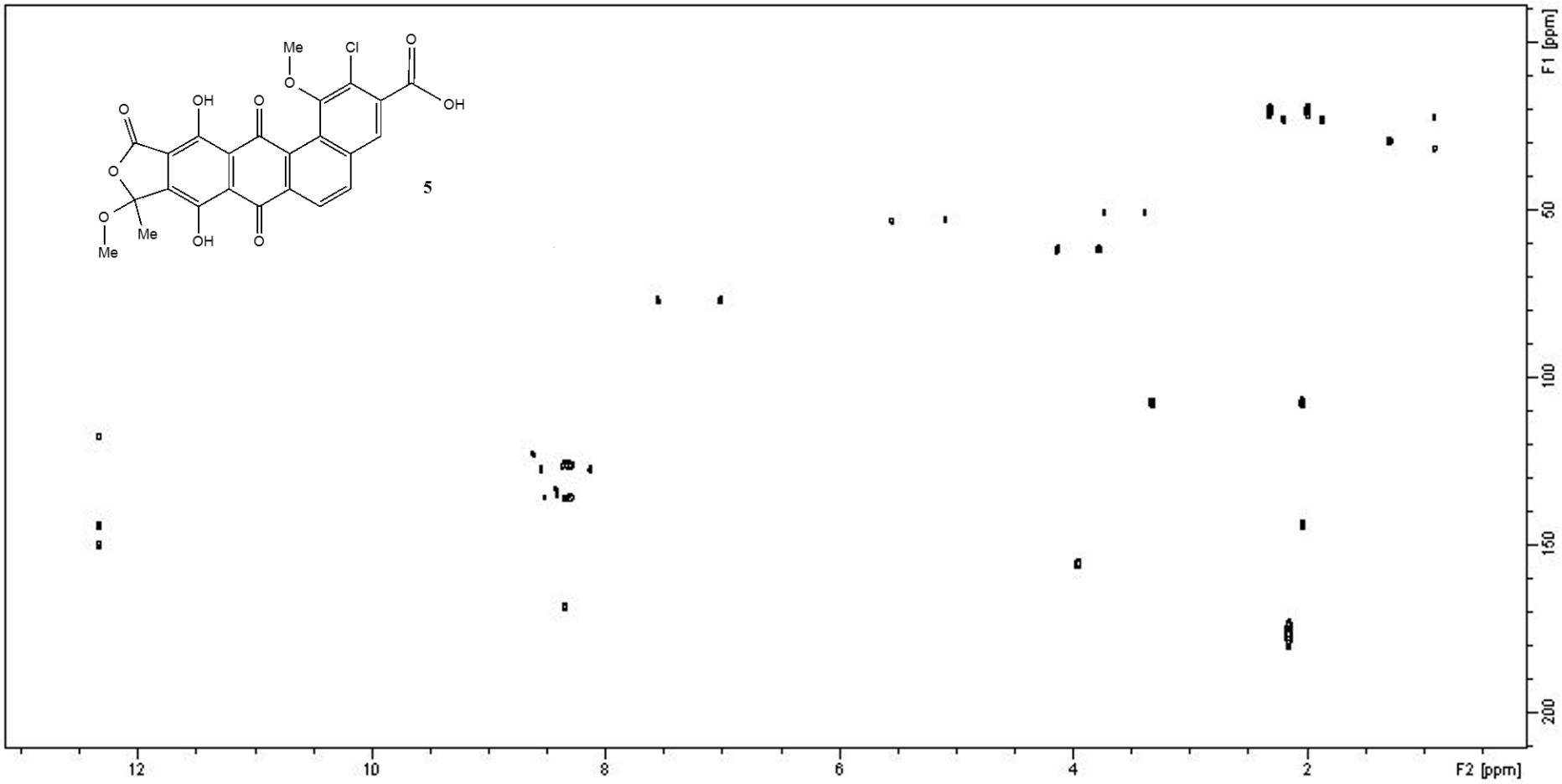


Figure 13: HMBC NMR of **5** in $CDCl_3$ with addition of 10µL TFA at 298K.

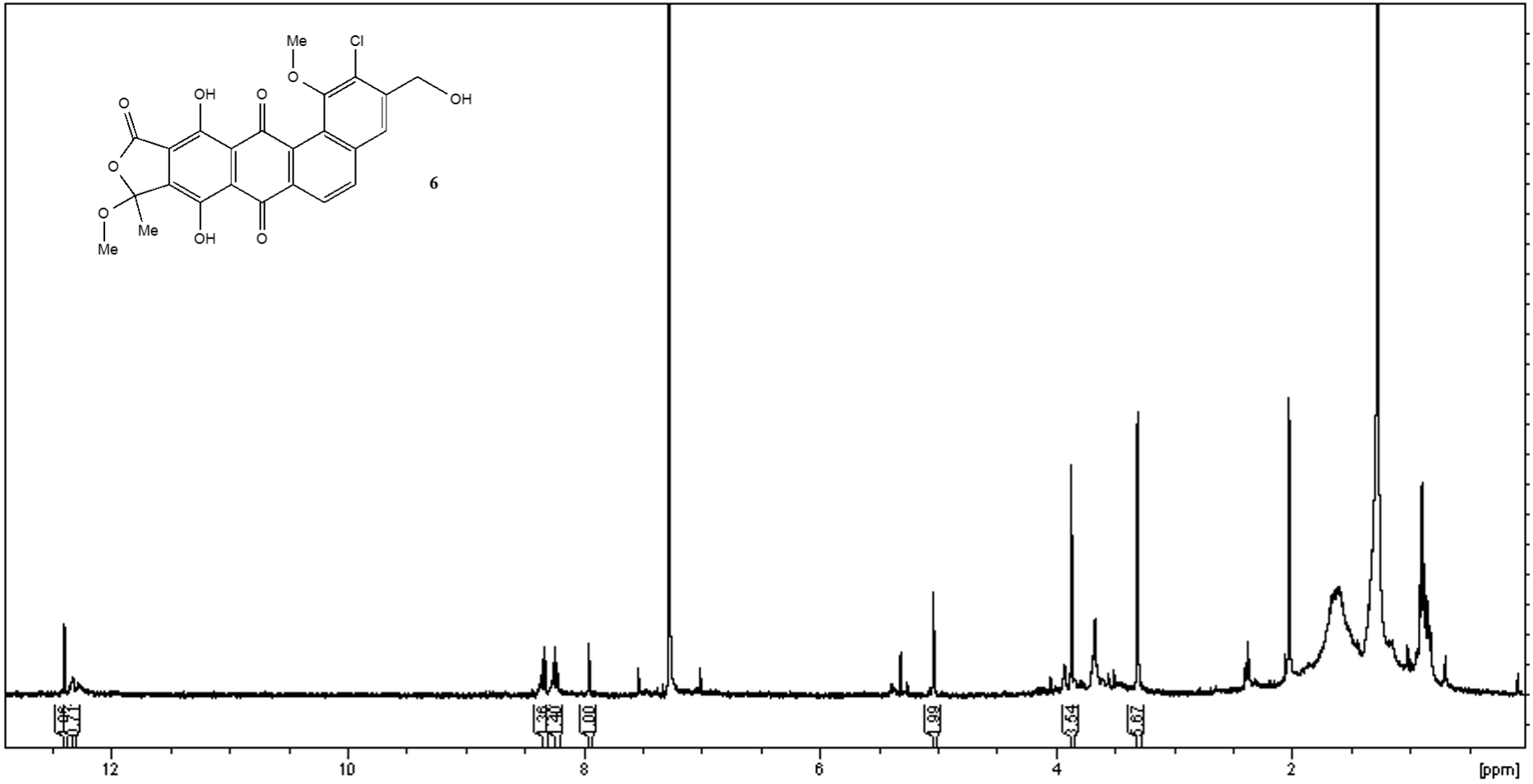


Figure 14: ^{1}H-NMR of **6** in $CDCl_3$ at 298K.

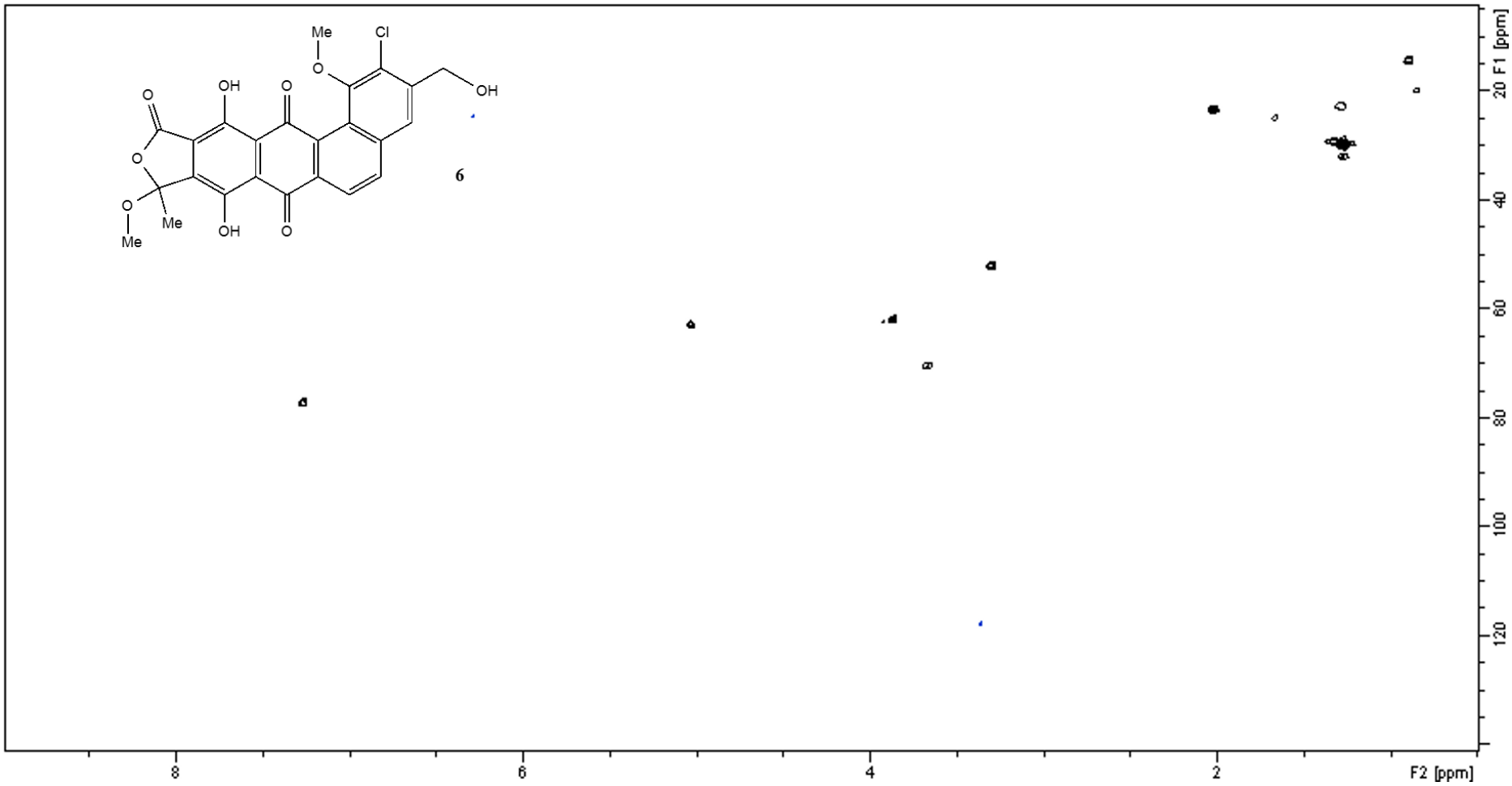


Figure 15. HSQC NMR of **6** in $CDCl_3$ at 298K.

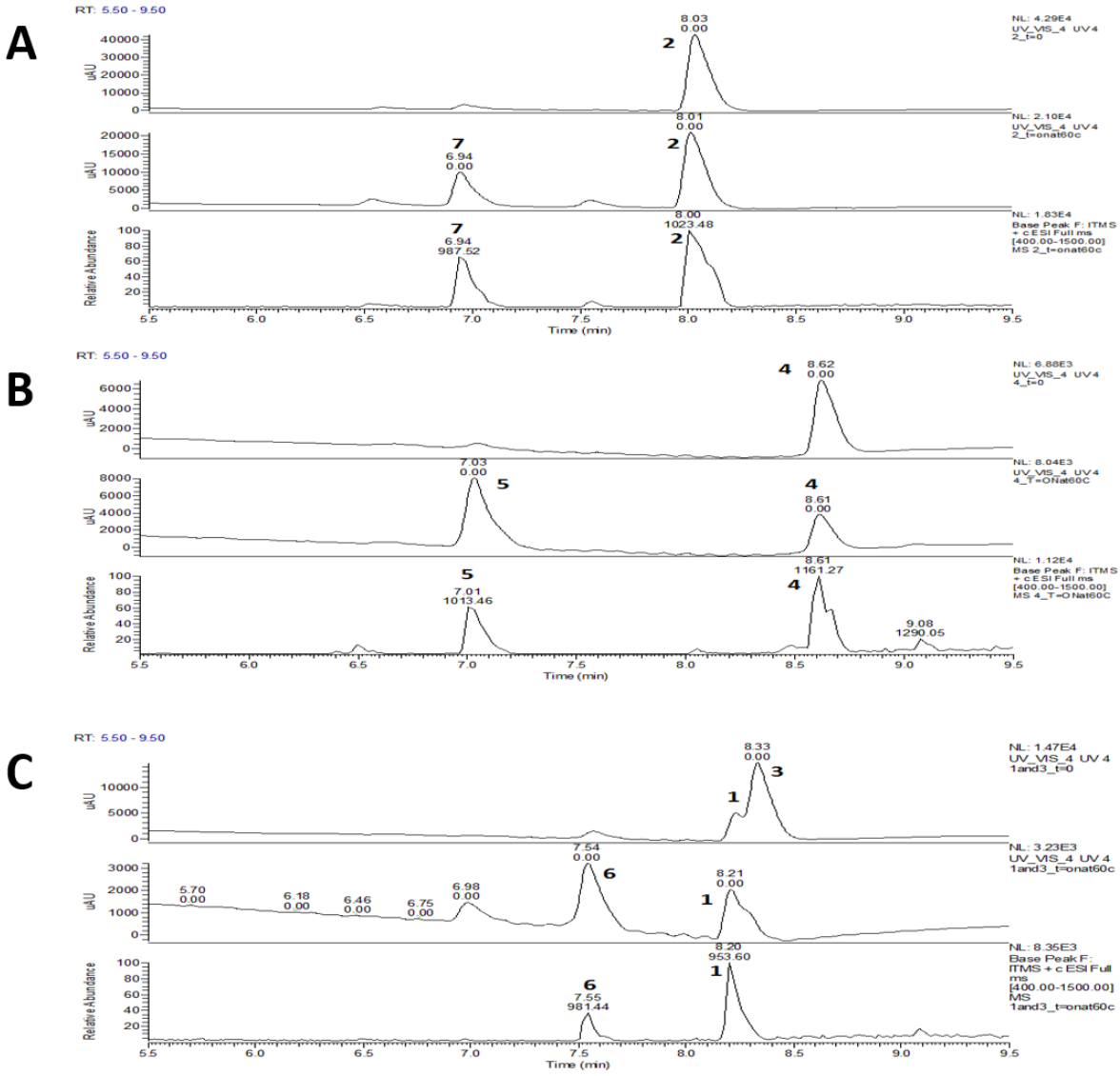


Figure 16. Stability of Allocyclinones **2** (Panel A), **4** (Panel B) and **3** in mixture with **1** (Panel C) in DMSO (60°C overnight). In each panel the following traces are reported: UV at 400 nm of the starting solution (upper trace), UV at 400 nm (middle trace) and MS (lower trace) after 16h at 60°C in DMSO. Peaks were detected 17as $[2M + NH_4]^+$ ions. Allocyclinone **1** resulted stable in these conditions (see panel C)

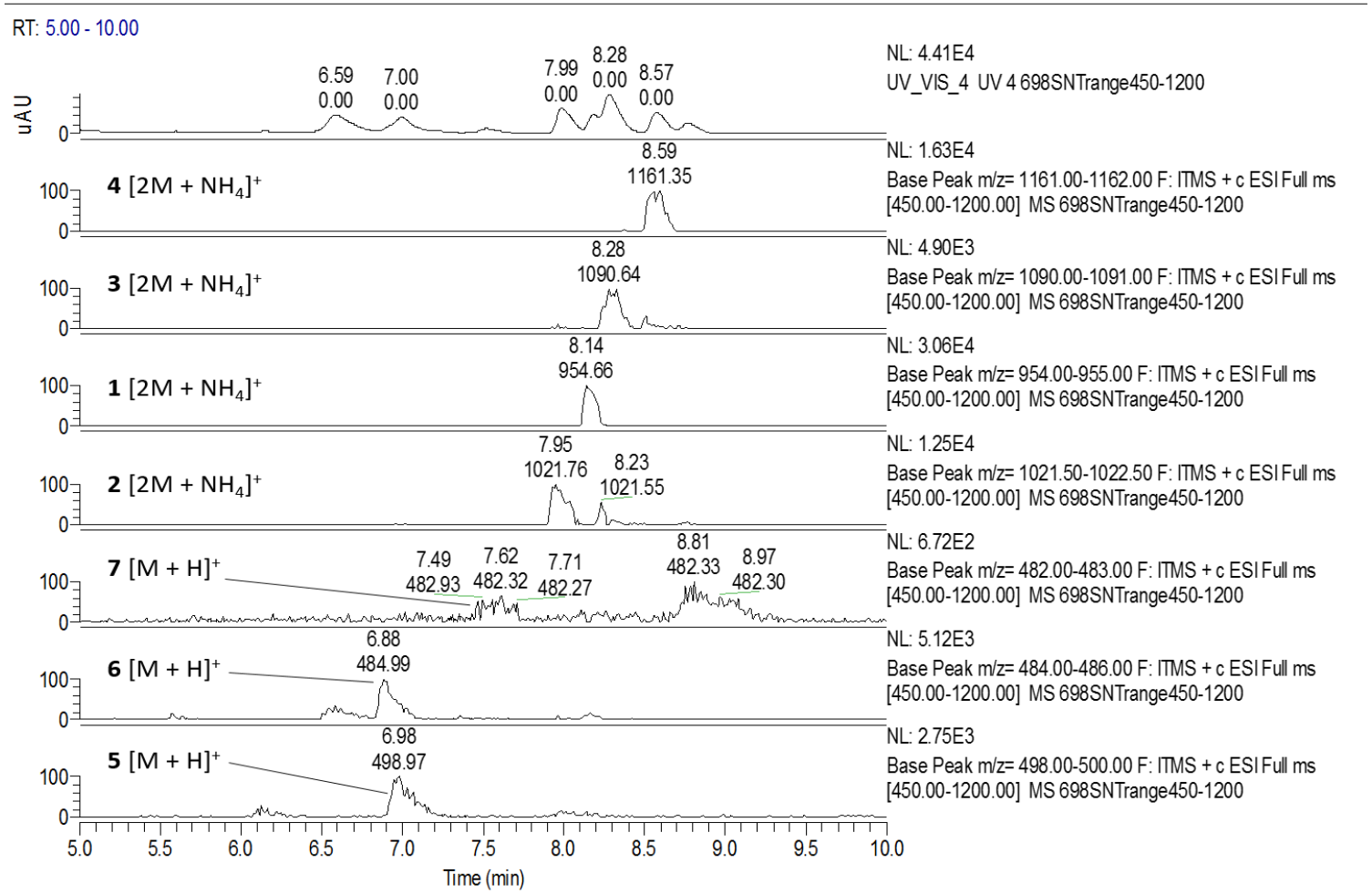


Figure 17: LC-MS analysis a fermentation broth showing pH 8.26 at the end of fermentation. In the left panel the UV trace at 400 nm (upper trace) and the extracted MS peaks for the allocyclinones **1**, **2**, **3** and **4** as $[2M + NH_4]^+$; in the right panel: the same UV trace at 400 nm (up) and the extracted MS peaks of the degraded Allocyclinones **5**, **6** and **7**.

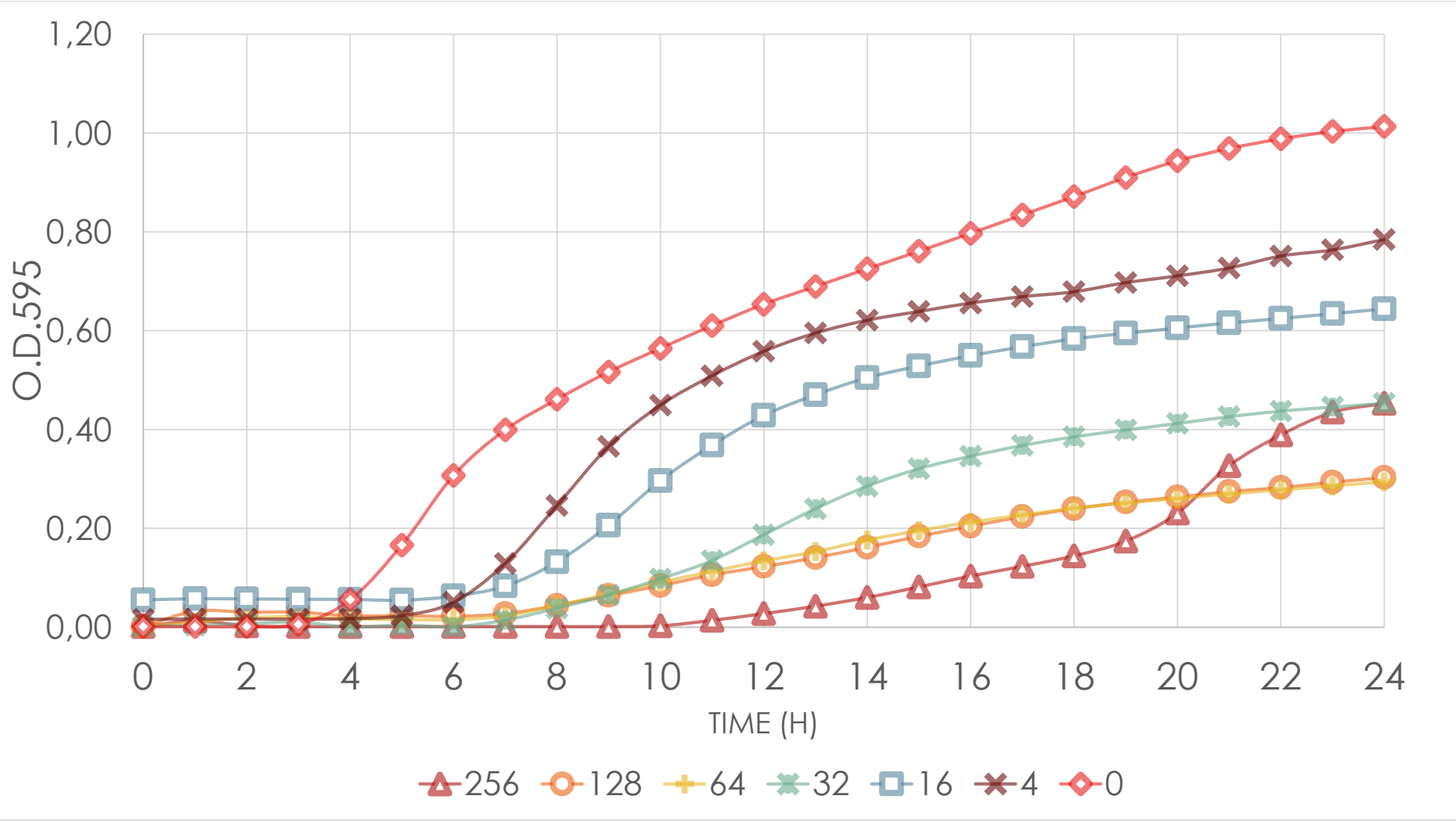


Figure 18. Growth kinetics of *E. coli* L4242 in the presence of different concentrations (µg/mL) of the **1**–**4** complex (same batch used for the experiments reported in Table 5.3)

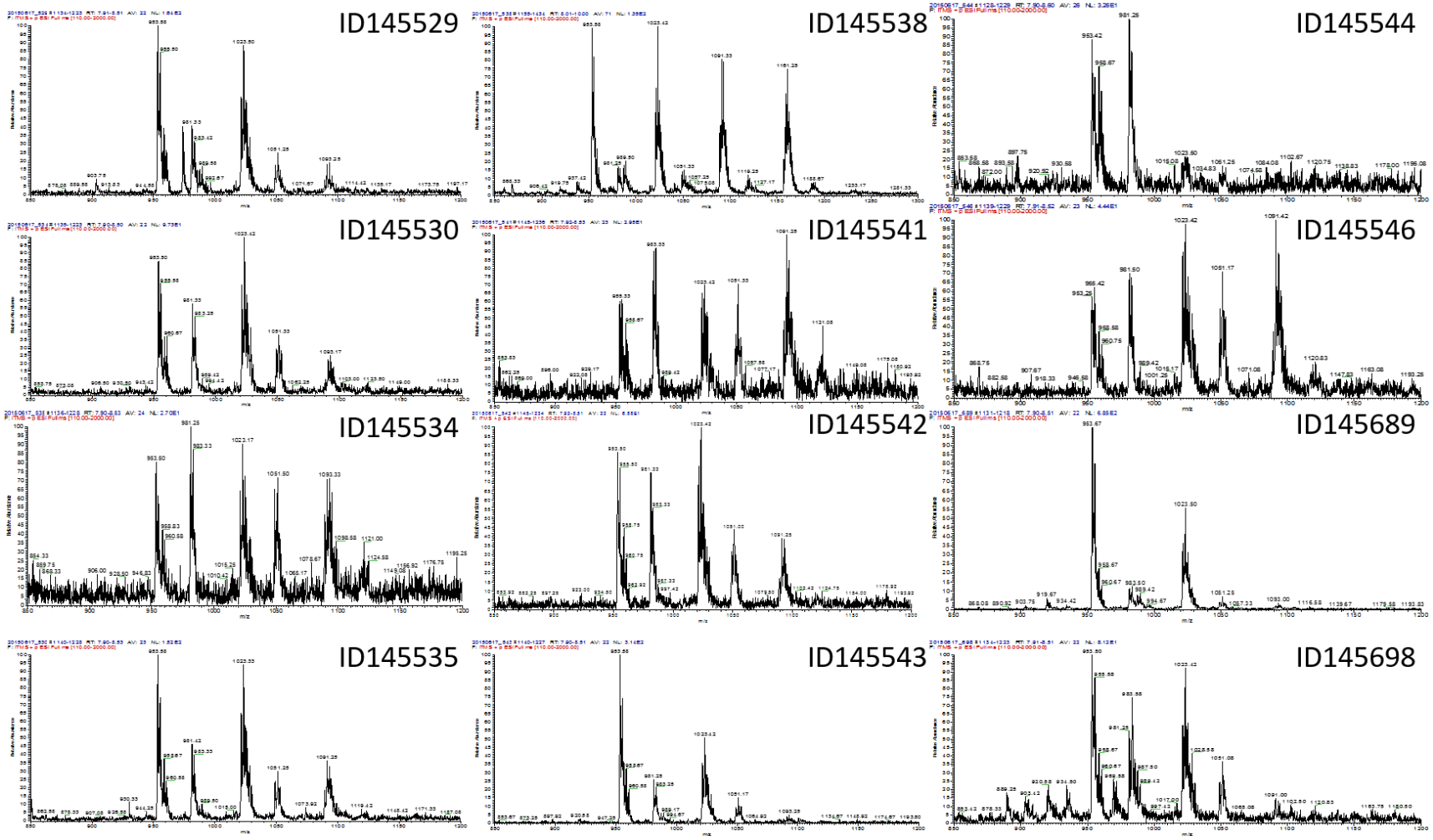
1. 

Figure 19: LC-MS analyses of extracts from the additional eleven allocyclinone producers listed in Table 5.4. Note the MS peaks for **1**, **2**, **3** and **4** as $[2M + NH_4]^+$.

Table 1 - Additional information on the CLFs used for the phylogenetic analysis of NAI698_9173 shown on Figure 4.4.

Compound	Classification	Starter-Extender	Length (aa)	Accession Number
Aclacinomycin	Anthracycline	Propionyl-9	407	AAF70107.1
Gilvocarcin	Angucycline	Propionyl-9	391	AAP69574.1
Simocyclinone	Angucycline	Acetyil-9	404	AAB36563,1
Oviedomycin	Angucycline	Acetyil-9	407	AAB36563.1
Kinamycin	Diazobenzofluorene	Acetyil-9	403	AAO65347.1
Jadomycin	Angucycline	Acetyil-9	404	AAB36563.1
Urdamycin	Angucycline	Acetyil-9	408	CAA60570.1
SCH-4755	Angucycline	Acetyil-9	408	CAH10116.1
Nogalamycin	Anthracycline	Acetyil-9	409	CAA12018.1
Steffimycin	Anthracycline	Acetyil-9	410	CAJ42319.1
Chromomycin	Aureolic acid	Acetyil-9	422	CAE17526.1
Oxytetracycline	Tetracycline	Malonyl-8	422	AAZ78326.1
Aranciamycin	Anthracycline	Acetyil-9	408	ABL09958.1
Mithramycin	Aureolic acid	Acetyil-9	408	CAA61990.1
Alnumycin	Naphthoquinone	Butyryl-7	428	ACI88862.1
Frenolicin	Benzoisochromanequinone	Butyryl-7	426	AAC18108.1
R1128	Anthraquinone	Butyryl-7	415	AAG30188.1
Actinorhodin	Benzoisochromanequinone	Acetyil-7	407	CAC44201.1.
Medermycin	Benzoisochromanequinone	Acetyil-7	408	BAC79045.1
Granatacin	Benzoisochromanequinone	Acetyil-7	415	CAA34370.1
Griseorhodin	Pentangular polyphenols	Acetyil-12	420	AAM33654.1
Rubromycin	Pentangular polyphenols	Acetyil-12	422	AAG03068.1
Arixanthamycin	Pentangular polyphenols	n.d.	412	AHX24702.1
Pradamycin	Pentangular polyphenols	Acetyil-11	392	ABM21748.1
Fredericamycin	Pentangular polyphenols	Hexanoyl-12	403	AAQ08917.1
Benastatin	Pentangular polyphenols	Hexanoyl-11	409	CAM58799.1

TLN-0522	Pentangular polyphenols	Methylbutyryl-12	423	ADB23392.1
Griseusin	Benzoisochromanequinone	Acetyil-9	412	CAA54859.1
Resistomycin	-	Malonyl-9	412	CAE51175.1
Allocyclinone	Angucycline	Unknown	407	This work

Inspired from the work of Ogasawara, Y., et al. 2015 [3](Ogasawara et al., 2015)

APPENDIX 2

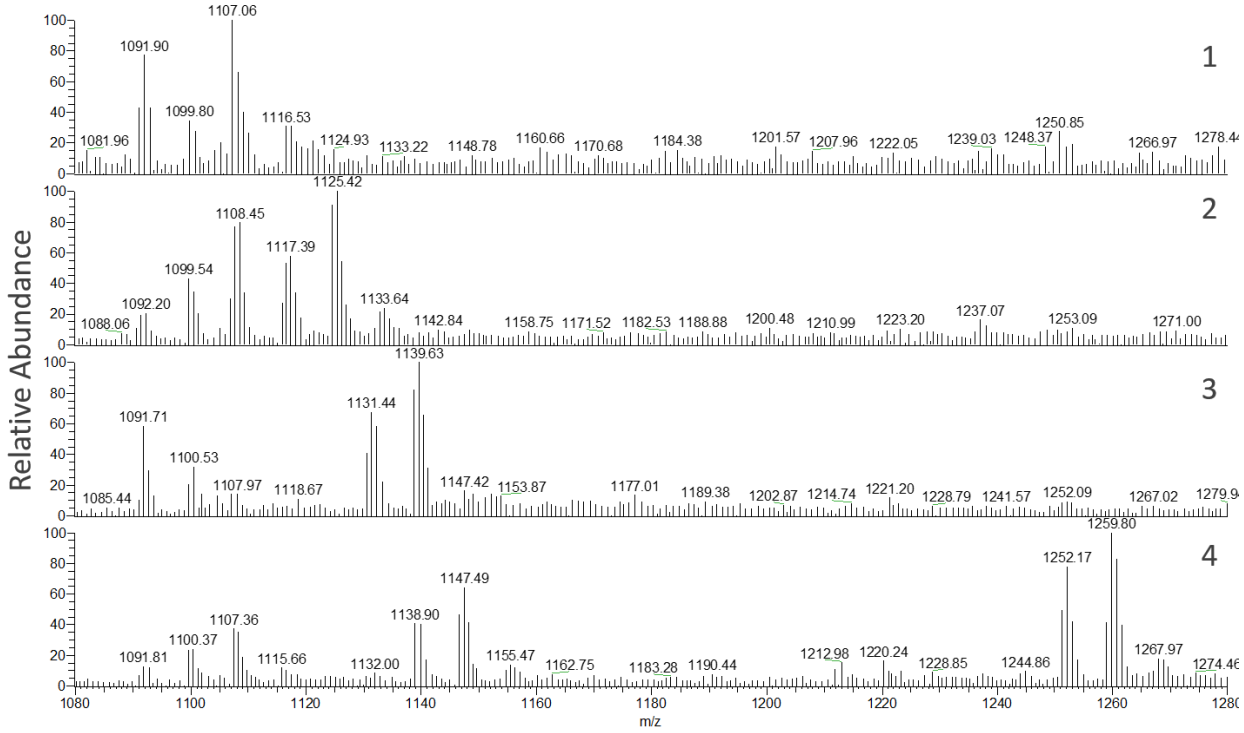


Figure 1. MS analyses of extracts from *Actinoallomurus* sp. ID145699 grown in medium AF-A with NaCl (trace 1) or KBr (trace 3) and from *Microbispora* sp. ID107841 grown in medium AF-A with NaCl (trace 2) or KBr (trace 4). Note the m/z$^+$ values corresponding to the GPA congeners ranging from 1212 to 1259.

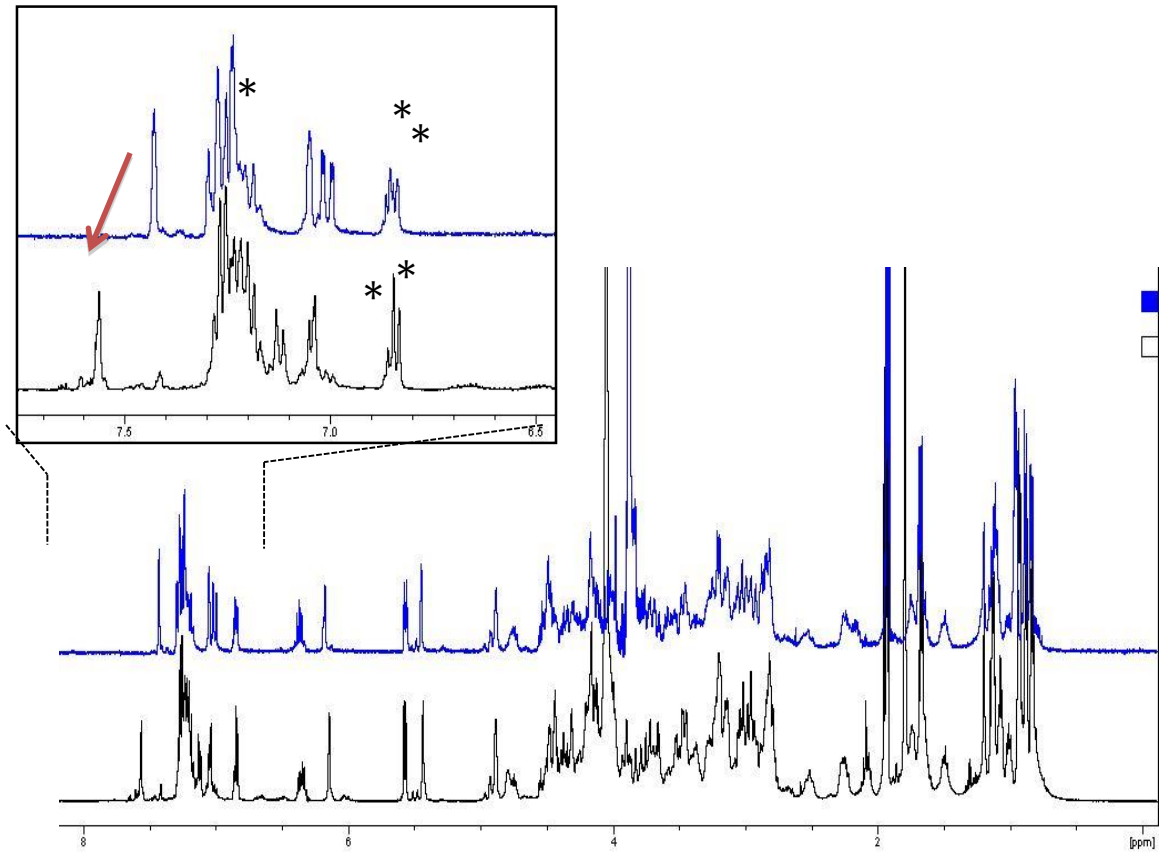


Figure 2. ^{1}H-NMR analysis of NAI-108 (black line) and NAI-107 (blue line) in CD_3CN:D_2O 6:4 at 298K. The enlargement shows the aromatic portion of the spectrum, with the asterisks denoting three selected indolic protons. The red arrow indicates the diagnostic change in the chemical shift observed for proton 4 in NAI-108.

APPENDIX 3

Table 1 - Additional information on the KSβs found in the four genomes of *Actinoallomurus* analysed

Compound	Classification	Starter-Extender	Length (aa)	Accession Number
Aclacinomycin	Anthracycline	Propionyl-9	407	AAF70107.1
Actinorhodin	Benzoisochromanequinone	Acetyil-7	407	CAC44201.1.
Allocyclinone	Angucycline	Unknown	407	This work
Alnumycin	Naphthoquinone	Butyryl-7	428	ACI88862.1
Aranciamycin	Anthracycline	Acetyil-9	408	ABL09958.1
Arixanthamycin	Pentangular polyphenols	n.d.	412	AHX24702.1
Benastatin	Pentangular polyphenols	Hexanoyl-11	409	CAM58799.1
Chromomycin	Aureolic acid	Acetyil-9	422	CAE17526.1
Fredericamycin	Pentangular polyphenols	Hexanoyl-12	403	AAQ08917.1
Frenolicin	Benzoisochromanequinone	Butyryl-7	426	AAC18108.1
Gilvocarcin	Angucycline	Propionyl-9	391	AAP69574.1
Granatacin	Benzoisochromanequinone	Acetyil-7	415	CAA34370.1
Griseorhodin	Pentangular polyphenols	Acetyil-12	420	AAM33654.1
Griseusin	Benzoisochromanequinone	Acetyil-9	412	CAA54859.1
Jadomycin	Angucycline	Acetyil-9	404	AAB36563.1
Kinamycin	Diazobenzofluorene	Acetyil-9	403	AAO65347.1
Medermycin	Benzoisochromanequinone	Acetyil-7	408	BAC79045.1
Mithramycin	Aureolic acid	Acetyil-9	408	CAA61990.1
NAI-113	Tetracenomycin	Unknown	432	This work
NAI-698_Cluster15	Unknown	Unknown	438	This work
NAI-711_Cluster2	Unknown	Unknown	424	This work
NAI-711_Cluster5	Unknown	Unknown	407	This work
NAI-716_CLuster18	Unknown	Unknown	418	This work
Nogalamycin	Anthracycline	Acetyil-9	409	CAA12018.1
Oviedomycin	Angucycline	Acetyil-9	407	AAB36563.1
Oxytetracycline	Tetracycline	Malonyl-8	422	AAZ78326.1
Pradamycin	Pentangular polyphenols	Acetyil-11	392	ABM21748.1
R1128	Anthraquinone	Butyryl-7	415	AAG30188.1
Resistomycin	-	Malonyl-9	412	CAE51175.1
Rubromycin	Pentangular polyphenols	Acetyil-12	422	AAG03068.1
SCH-4755	Angucycline	Acetyil-9	408	CAH10116.1
Simocyclinone	Angucycline	Acetyil-9	404	AAB36563,1
Steffimycin	Anthracycline	Acetyil-9	410	CAJ42319.1
TLN-0522	Pentangular polyphenols	Methylbutyryl-12	423	ADB23392.1
Urdamycin	Angucycline	Acetyil-9	408	CAA60570.1

8. REFERENCES

Ajikumar, P.K., Tyo, K., Carlsen, S., Mucha, O., Phon, T.H., and Stephanopoulos, G. (2008). Terpenoids: Opportunities for Biosynthesis of Natural Product Drugs Using Engineered Microorganisms. Molecular Pharmaceutics *5*, 167-190.

Ajisaka, M., Kariyone, K., Jomon, K., Yazawa, H., and Arima, K. (1969). Isolation of the bromo analogues of pyrrolnitrin from Pseudomonas pyrrolnitrica. Agricultural and Biological Chemistry *33*, 294-295.

Andersson, D.I., and Levin, B.R. (1999). The biological cost of antibiotic resistance. Curr Opin Microbiol *2*, 489-493.

Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., *et al.* (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Natural product reports *30*, 108-160.

Arya, P., Chou, D.T.H., and Baek, M.-G. (2001). Diversity-Based Organic Synthesis in the Era of Genomics and Proteomics. Angewandte Chemie International Edition *40*, 339-346.

August, P.R., Tang, L., Yoon, Y.J., Ning, S., Muller, R., Yu, T.W., Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X., *et al.* (1998). Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of Amycolatopsis mediterranei S699. Chemistry & biology *5*, 69-79.

Baltz, R.H. (2007). Antimicrobials from actinomycetes: back to the future. Microbe-American Society for Microbiology *2*, 125.

Baltz, R.H. (2008). Renaissance in antibacterial discovery from actinomycetes. Current opinion in pharmacology *8*, 557-563.

Barbachyn, M. (2012). Oxazolidinone Antibacterial Agents. In Antibiotic Discovery and Development, T.J. Dougherty, and M.J. Pucci, eds. (Springer US), pp. 271-299.

Barbachyn, M.R., and Ford, C.W. (2003). Oxazolidinone Structure–Activity Relationships Leading to Linezolid. Angewandte Chemie International Edition *42*, 2010-2023.

Běhal, V. (2000). Bioactive products from streptomyces. In Advances in Applied Microbiology (Academic Press), pp. 113-156.

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

Berdy, J. (2012). Thoughts and facts about antibiotics: where we are now and where we are heading. The Journal of antibiotics *65*, 385-395.

Bister, B., Bischoff, D., Nicholson, G.J., Stockert, S., Wink, J., Brunati, C., Donadio, S., Pelzer, S., Wohlleben, W., and Süssmuth, R.D. (2003). Bromobalhimycin and Chlorobromobalhimycins—Illuminating the Potential of Halogenases in Glycopeptide Antibiotic Biosyntheses. Chembiochem : a European journal of chemical biology *4*, 658-662.

Bohacek, R.S., McMartin, C., and Guida, W.C. (1996). The art and practice of structure-based drug design: A molecular modeling perspective. Medicinal Research Reviews *16*, 3-50.

Breukink, E., and de Kruijff, B. (1999). The lantibiotic nisin, a special case or not? Biochimica et Biophysica Acta (BBA) - Biomembranes *1462*, 223-234.

Bull, A.T.S., J.E.M. (2004). Microbial diversity and Bioprospecting (ASM Press).

Bush, K. (1989). Characterization of beta-lactamases. Antimicrobial Agents and Chemotherapy *33*, 259-263.

Busti, E., Monciardini, P., Cavaletti, L., Bamonte, R., Lazzarini, A., Sosio, M., and Donadio, S. (2006). Antibiotic-producing ability by representatives of a newly discovered lineage of actinomycetes. Microbiology *152*, 675-683.

Butler, M.S., Robertson, A.A.B., and Cooper, M.A. (2014). Natural product and natural product derived drugs in clinical trials. Natural product reports *31*, 1612-1661.

Caffrey, P., Lynch, S., Flood, E., Finnan, S., and Oliynyk, M. (2001). Amphotericin biosynthesis in Streptomyces nodosus: deductions from analysis of polyketide synthase and late genes. Chemistry & biology *8*, 713-723.

Carrano, L., Abbondi, M., Turconi, P., Candiani, G., and Marinelli, F. (2015). A Novel Microbisporicin Producer Identified by Early Dereplication during Lantibiotic Screening. BioMed Research International *2015*, 10.

Cassir, N., Rolain, J.-M., and Brouqui, P. (2014). A new strategy to fight antimicrobial resistance: the revival of old antibiotics. Frontiers in Microbiology *5*.

Castiglione, F., Lazzarini, A., Carrano, L., Corti, E., Ciciliato, I., Gastaldo, L., Candiani, P., Losi, D., Marinelli, F., Selva, E., *et al.* (2008). Determining the structure and mode of action of microbisporicin, a potent lantibiotic active against multiresistant pathogens. Chemistry & biology *15*, 22-31.

Chen, F., Mackey, A.J., Stoeckert, C.J., Jr., and Roos, D.S. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res *34*, D363-368.

Clever, L.H. (1999). Book Review. New England Journal of Medicine *340*, 659-660.

Coates, A.R., Halls, G., and Hu, Y. (2011). Novel classes of antibiotics or more of the same? Br J Pharmacol *163*, 184-194.

Collins, J.J., and Donadio, S. (2011). Expanding and enhancing our antibacterial arsenal. Current Opinion in Microbiology *14*, 511-512.

Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J., and Leadlay, P.F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea.

Cruz, J.C., S; Maffioli, S.; Bernasconi, A.; Brunati, C.; Gaspari, E.; Sosio, M.; Wellington, E.; Donadio S. (2015). Allocyclinones, hyperchlorinated angucyclinones and common metabolites from Actinoallomurus. Unpublished.

Cruz, J.C.S., Iorio, M., Monciardini, P., Simone, M., Brunati, C., Gaspari, E., Maffioli, S.I., Wellington, E., Sosio, M., and Donadio, S. (2015). Brominated Variant of the Lantibiotic NAI-107 with Enhanced Antibacterial Potency. Journal of natural products.

D/'Costa, V.M., King, C.E., Kalan, L., Morar, M., Sung, W.W.L., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., *et al.* (2011). Antibiotic resistance is ancient. Nature *477*, 457-461.

Daum, M., Peintner, I., Linnenbrink, A., Frerich, A., Weber, M., Paululat, T., and Bechthold, A. (2009). Organisation of the biosynthetic gene cluster and tailoring enzymes in the biosynthesis of the tetracyclic quinone glycoside antibiotic polyketomycin. Chembiochem : a European journal of chemical biology *10*, 1073-1083.

Davies, J., and Davies, D. (2010). Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev *74*, 417-433.

de las Heras, B., Rodriguez, B., Bosca, L., and Villar, A.M. (2003). Terpenoids: sources, structure elucidation and therapeutic potential in inflammation. Curr Top Med Chem *3*, 171-185.

Dimkpa, C.O., Svatoš, A., Dabrowska, P., Schmidt, A., Boland, W., and Kothe, E. (2008). Involvement of siderophores in the reduction of metal-induced inhibition of auxin synthesis in Streptomyces spp. Chemosphere *74*, 19-25.

Donadio, S., Maffioli, S., Monciardini, P., Sosio, M., and Jabes, D. (2010). Sources of novel antibiotics--aside the common roads. Appl Microbiol Biotechnol *88*, 1261-1267.

Donadio, S., Monciardini, P., and Sosio, M. (2009). Chapter 1. Approaches to discovering novel antibacterial and antifungal agents. Methods in enzymology *458*, 3-28.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. Science *252*, 675-679.

Donadio S., B.L., Serina S., Sosio M., Stinchi S. (2005). Discovering Novel Antibacterial Agents by High Throughput Screening. Frontiers in Drug Design & Discovery *1*, 14.

Du, L., Sánchez, C., Chen, M., Edwards, D.J., and Shen, B. (2000). The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. Chemistry & biology *7*, 623-642.

Durai, S., Vigneshwari, L., and Balamurugan, K. (2013). Caenorhabditis elegans-based in vivo screening of bioactives from marine sponge-associated bacteria against Vibrio alginolyticus. Journal of Applied Microbiology *115*, 1329-1342.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research *32*, 1792-1797.

Eliopoulos, G.M., Meka, V.G., and Gold, H.S. (2004). Antimicrobial Resistance to Linezolid. Clinical Infectious Diseases *39*, 1010-1015.

Endo, A., Kuroda, M., and Tsujita, Y. (1976). ML-236A, ML-236B, and ML-236C, new inhibitors of cholesterogenesis produced by Penicillium citrinium. The Journal of antibiotics *29*, 1346-1348.

Ezaki, N., KOYAMA, M., SHOMURA, T., TSURUOKA, T., and INOUYE, S. (1983). Pyrrolomycins C, D and E, new members of pyrrolomycins. The Journal of antibiotics *36*, 1263-1267.

Fauci, A.S., and Morens, D.M. (2012). The Perpetual Challenge of Infectious Diseases. New England Journal of Medicine *366*, 454-461.

Feling, R.H., Buchanan, G.O., Mincer, T.J., Kauffman, C.A., Jensen, P.R., and Fenical, W. (2003). Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus Salinospora. Angewandte Chemie International Edition *42*, 355-357.

Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D., and Thomas, M.G. (2008). Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. Molecular pharmaceutics *5*, 191-211.

Feng, B.Y., Simeonov, A., Jadhav, A., Babaoglu, K., Inglese, J., Shoichet, B.K., and Austin, C.P. (2007). A High-Throughput Screen for Aggregation-Based Inhibition in a Large Compound Library. Journal of Medicinal Chemistry *50*, 2385-2390.

Finking, R., and Marahiel, M.A. (2004). Biosynthesis of Nonribosomal Peptides. Annual Review of Microbiology *58*, 453-488.

Foulston, L., and Bibb, M. (2011). Feed-forward regulation of microbisporicin biosynthesis in Microbispora corallina. Journal of bacteriology *193*, 3064-3071.

Foulston, L.C., and Bibb, M.J. (2010). Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. Proceedings of the National Academy of Sciences of the United States of America *107*, 13461-13466.

Funa, N., Ohnishi, Y., Ebizuka, Y., and Horinouchi, S. (2002). Alteration of reaction and substrate specificity of a bacterial type III polyketide synthase by site-directed mutagenesis. Biochem J *367*, 781-789.

Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y., and Horinouchi, S. (1999). A new pathway for polyketide synthesis in microorganisms. Nature *400*, 897-899.

Ganesan, A. (2008). The impact of natural products upon modern drug discovery. Current opinion in chemical biology *12*, 306-317.

Gao, Q., and Thorson, J.S. (2008). The biosynthetic genes encoding for the production of the dynemicin enediyne core in Micromonospora chersina ATCC53710. FEMS Microbiol Lett *282*, 105-114.

Garfunkle, J., Kimball, F.S., Trzupek, J.D., Takizawa, S., Shimamura, H., Tomishima, M., and Boger, D.L. (2009). Total Synthesis of Chloropeptin II (Complestatin) and Chloropeptin I. Journal of the American Chemical Society *131*, 16036-16038.

Gengenbacher, M., and Dick, T. (2015). Antibacterial drug discovery: doing it right. Chemistry & biology *22*, 5-6.

Glynn, A.A. (2005). Penicillin Man: Alexander Fleming and the Antibiotic Revolution K. Brown Sutton Publishing, 2004. ISBN 0-7509-3152-3. £20.00, 320 pp. Journal of Antimicrobial Chemotherapy *56*, 444-445.

Goodfellow, M., and Fiedler, H.P. (2010). A guide to successful bioprospecting: informed by actinobacterial systematics. Antonie van Leeuwenhoek *98*, 119-142.

Goodin, S., Kane, M.P., and Rubin, E.H. (2004). Epothilones: Mechanism of Action and Biologic Activity. Journal of Clinical Oncology *22*, 2015-2025.

Gould, I.M. Antibiotic resistance: the perfect storm. International Journal of Antimicrobial Agents *34*, S2-S5.

Greer, J., Erickson, J.W., Baldwin, J.J., and Varney, M.D. (1994). Application of the three-dimensional structures of protein target molecules in structure-based drug design. J Med Chem *37*, 1035-1054.

Gurgui, C., and Piel, J. (2010). Metagenomic approaches to identify and isolate bioactive natural products from microbiota of marine sponges. Methods Mol Biol *668*, 247-264.

Han, M.V., and Zmasek, C.M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics *10*, 356.

Harrison, J., and Studholme, D.J. (2014). Recently published Streptomyces genome sequences. Microbial biotechnology *7*, 373-380.

Harvey, A. (2000). Strategies for discovering drugs from previously unexplored natural products. Drug Discov Today *5*, 294-300.

Harvey, A.L., Edrada-Ebel, R., and Quinn, R.J. (2015). The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discov *14*, 111-129.

Hermann, T. (2000). Strategies for the Design of Drugs Targeting RNA and RNA-Protein Complexes. Angew Chem Int Ed Engl *39*, 1890-1904.

Hilpert, K., Volkmer-Engert, R., Walter, T., and Hancock, R.E.W. (2005). High-throughput generation of small antibacterial peptides with improved activity. Nat Biotech *23*, 1008-1012.

Hochlowski, J.E., Jackson, M., Rasmussen, R.R., Buko, A.M., Clement, J.J., Whittern, D.N., and McAlpine, J.B. (1997). Production of brominated tiacumicin derivatives. The Journal of antibiotics *50*, 201-205.

Hogberg, L.D., Heddini, A., and Cars, O. (2010). The global need for effective antibiotics: challenges and recent advances. Trends Pharmacol Sci *31*, 509-515.

Hou, J., Liu, P., Qu, H., Fu, P., Wang, Y., Wang, Z., Li, Y., Teng, X., and Zhu, W. (2012). Gilvocarcin HE: a new polyketide glycoside from Streptomyces sp. J Antibiot *65*, 523-526.

Hrvatin, S., and Piel, J. (2007). Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. J Microbiol Methods *68*, 434-436.

Ichikawa, N., Oguchi, A., Ikeda, H., Ishikawa, J., Kitani, S., Watanabe, Y., Nakamura, S., Katano, Y., Kishi, E., Sasagawa, M., *et al.* (2010). Genome Sequence of Kitasatospora setae NBRC 14216(T): An Evolutionary Snapshot of the Family Streptomycetaceae. DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes *17*, 393-406.

Ichikawa, N., Sasagawa, M., Yamamoto, M., Komaki, H., Yoshida, Y., Yamazaki, S., and Fujita, N. (2013). DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucleic Acids Res *41*, D408-414.

Igarashi, K., Ishitsuka, H., and Kaji, A. (1969). Comparative studies on the mechanism of action of lincomycin, streptomycin, and erythromycin. Biochemical and Biophysical Research Communications *37*, 499-504.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., and Omura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. Nat Biotech *21*, 526-531.

Inahashi, Y., Iwatsuki, M., Ishiyama, A., Matsumoto, A., Hirose, T., Oshita, J., Sunazuka, T., Panbangred, W., Takahashi, Y., Kaiser, M., *et al.* (2015). Actinoallolides A-E, New Anti-trypanosomal Macrolides, Produced by an Endophytic Actinomycete, Actinoallomurus fulvus MK10-036. Organic letters *17*, 864-867.

Ince, D., and Hooper, D.C. (2003). Quinolone Resistance Due to Reduced Target Enzyme Expression. Journal of bacteriology *185*, 6883-6892.

Indananda, C., Thamchaipenet, A., Matsumoto, A., Inahashi, Y., Duangmal, K., and Takahashi, Y. (2011). Actinoallomurus oryzae sp. nov., an endophytic actinomycete isolated from roots of a Thai jasmine rice plant. Int J Syst Evol Microbiol *61*, 737-741.

Ishida, T. (2005). Biotransformation of terpenoids by mammals, microorganisms, and plant-cultured cells. Chem Biodivers *2*, 569-590.

Jabes, D. (2011). The antibiotic R&D pipeline: an update. Current Opinion in Microbiology *14*, 564-569.

Jabes, D., Brunati, C., Candiani, G., Riva, S., Romano, G., and Donadio, S. (2011). Efficacy of the new lantibiotic NAI-107 in experimental infections induced by multidrug-resistant Gram-positive pathogens. Antimicrob Agents Chemother *55*, 1671-1676.

Jakobsen, H., Bojer, M.S., Marinus, M.G., Xu, T., Struve, C., Krogfelt, K.A., and Løbner-Olesen, A. (2013). The Alkaloid Compound Harmane Increases the Lifespan of <italic>Caenorhabditis elegans</italic> during Bacterial Infection, by Modulating the Nematode's Innate Immune Response. PLoS ONE *8*, e60519.

Jaspars, M., and Challis, G. (2014). Microbiology: A talented genus. Nature *506*, 38-39.

Jiang, C., Wang, H., Kang, Q., Liu, J., and Bai, L. (2012). Cloning and Characterization of the Polyether Salinomycin Biosynthesis Gene Cluster of Streptomyces albus XM211. Applied and Environmental Microbiology *78*, 994-1003.

Kalakoutskii, L.V., and Agre, N.S. (1976). Comparative aspects of development and differentiation in actinomycetes. Bacteriological Reviews *40*, 469-524.

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

Kang, H.S., and Brady, S.F. (2014). Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols. ACS chemical biology *9*, 1267-1272.

Kevin Ii, D.A., Meujo, D.A., and Hamann, M.T. (2009). Polyether ionophores: broad-spectrum and promising biologically active molecules for the control of drug-resistant bacteria and parasites. Expert Opin Drug Discov *4*, 109-146.

Kim, D.H., Feinbaum, R., Alloing, G., Emerson, F.E., Garsin, D.A., Inoue, H., Tanaka-Hino, M., Hisamoto, N., Matsumoto, K., Tan, M.-W., *et al.* (2002). A Conserved p38 MAP Kinase Pathway in Caenorhabditis elegans Innate Immunity. Science *297*, 623-626.

Kim, T.K., Garson, M.J., and Fuerst, J.A. (2005). Marine actinomycetes related to the "Salinospora" group from the Great Barrier Reef sponge Pseudoceratina clavata. In Environ Microbiol (England), pp. 509-518.

Kohanski, M.A., Dwyer, D.J., and Collins, J.J. (2010). How antibiotics kill bacteria: from targets to networks. Nature reviews Microbiology *8*, 423-435.

Koyama, R., Matsumoto, A., Inahashi, Y., Omura, S., and Takahashi, Y. (2012). Isolation of actinomycetes from the root of the plant, Ophiopogon japonicus, and proposal of two new species, Actinoallomurus liliacearum sp. nov. and Actinoallomurus vinaceus sp. nov. The Journal of antibiotics *65*, 335-340.

Krawczyk, J.M., Voller, G.H., Krawczyk, B., Kretz, J., Bronstrup, M., and Sussmuth, R.D. (2013). Heterologous expression and engineering studies of labyrinthopeptins, class III lantibiotics from Actinomadura namibiensis. Chemistry & biology *20*, 111-122.

Lam, K.S. (2007). New aspects of natural products in drug discovery. Trends in Microbiology *15*, 279-289.

Lazzarini, A., Cavaletti, L., Toppo, G., and Marinelli, F. (2000). Rare genera of actinomycetes as potential producers of new antibiotics. Antonie van Leeuwenhoek *78*, 399-405.

Lepak, A.J., Marchillo, K., Craig, W.A., and Andes, D.R. (2015). In vivo pharmacokinetics and pharmacodynamics of the lantibiotic NAI-107 in a neutropenic murine thigh infection model. Antimicrob Agents Chemother *59*, 1258-1264.

Lewis, K. (2013). Platforms for antibiotic discovery. Nat Rev Drug Discov *12*, 371-387.

Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L., Engels, I., Conlon, B.P., Mueller, A., Schaberle, T.F., Hughes, D.E., Epstein, S., *et al.* (2015). A new antibiotic kills pathogens without detectable resistance. Nature *517*, 455-459.

Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev *46*, 3-26.

Luch, A. (2012). Molecular, Clinical and Environmental Toxicology: Volume 3: Environmental Toxicology, Vol 101 (Springer Science & Business Media).

Luo, N., Pereira, S., Sahin, O., Lin, J., Huang, S., Michel, L., and Zhang, Q. (2005). Enhanced in vivo fitness of fluoroquinolone-resistant Campylobacter jejuni in the absence of antibiotic selection pressure. Proc Natl Acad Sci U S A *102*, 541-546.

MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H., Loiacono, K.A., Lynch, B.A., Phillips, T., Narula, S., Sundaramoorthi, R., *et al.* (2001). Expression and isolation of antimicrobial small molecules from soil DNA libraries. J Mol Microbiol Biotechnol *3*, 301-308.

Maffioli, S.I., Cruz, J.C., Monciardini, P., Sosio, M., and Donadio, S. (2016). Advancing cell wall inhibitors towards clinical applications. Journal of industrial microbiology & biotechnology *43*, 177-184.

Maffioli, S.I., Iorio, M., Sosio, M., Monciardini, P., Gaspari, E., and Donadio, S. (2014). Characterization of the congeners in the lantibiotic NAI-107 complex. Journal of natural products *77*, 79-84.

Maffioli, S.I., Monciardini, P., Catacchio, B., Mazzetti, C., Munch, D., Brunati, C., Sahl, H.G., and Donadio, S. (2015). Family of class I lantibiotics from actinomycetes and improvement of their antibacterial activities. ACS chemical biology *10*, 1034-1042.

Malpartida, F., and Hopwood, D.A. (1984). Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. Nature *309*, 462-464.

Mariani, R., Granata, G., Maffioli, S.I., Serina, S., Brunati, C., Sosio, M., Marazzi, A., Vannini, A., Patel, D., White, R., *et al.* (2005). Antibiotics GE23077, novel inhibitors of bacterial RNA polymerase. Part 3: Chemical derivatization. Bioorg Med Chem Lett *15*, 3748-3752.

Martinez, J.L., Baquero, F., and Andersson, D.I. (2007). Predicting antibiotic resistance. Nat Rev Micro *5*, 958-965.

Matsumoto, A., Fukuda, A., Inahashi, Y., Omura, S., and Takahashi, Y. (2012). Actinoallomurus radicium sp. nov., isolated from the roots of two plant species. Int J Syst Evol Microbiol *62*, 295-298.

Mazzetti, C., Ornaghi, M., Gaspari, E., Parapini, S., Maffioli, S., Sosio, M., and Donadio, S. (2012). Halogenated spirotetronates from Actinoallomurus. Journal of natural products *75*, 1044-1050.

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., *et al.* (2015). Minimum Information about a Biosynthetic Gene cluster. Nature chemical biology *11*, 625-631.

Metsa-Ketela, M., Salo, V., Halo, L., Hautala, A., Hakala, J., Mantsala, P., and Ylihonko, K. (1999). An efficient approach for screening minimal PKS genes from Streptomyces. FEMS Microbiol Lett *180*, 1-6.

Michael, C.A., Dominey-Howes, D., and Labbate, M. (2014). The Antimicrobial Resistance Crisis: Causes, Consequences, and Management. Frontiers in Public Health *2*, 145.

Mingeot-Leclercq, M.-P., Glupczynski, Y., and Tulkens, P.M. (1999). Aminoglycosides: Activity and Resistance. Antimicrobial Agents and Chemotherapy *43*, 727-737.

Minotti, G., Menna, P., Salvatorelli, E., Cairo, G., and Gianni, L. (2004). Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. Pharmacol Rev *56*, 185-229.

Moellering, J.R.C. (2003). Linezolid: The First Oxazolidinone Antimicrobial. Annals of Internal Medicine *138*, 135-142.

Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D.D., *et al.* (2000). The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90. Chemistry & biology *7*, 97-109.

Monciardini, P., Iorio, M., Maffioli, S., Sosio, M., and Donadio, S. (2014). Discovering new bioactive molecules from microbial sources. Microbial biotechnology *7*, 209-220.

Monciardini, P., Sosio, M., Cavaletti, L., Chiocchini, C., and Donadio, S. (2002). New PCR primers for the selective amplification of 16S rDNA from different groups of actinomycetes. FEMS Microbiol Ecol *42*, 419-429.

Monciardini, P.S., S.; Gallo, G.; Mazzei, E.; Maffioli, S; Sosio, M.; Donadio, S. (2015). Enhanced production of the lantibiotic NAI-107 by Microbispora through strain selection and medium optimization. Submitted for Publication.

Montecalvo, M.A., Horowitz, H., Wormser, G.P., Seiter, K., and Carbonaro, C.A. (1995). Effect of novobiocin-containing antimicrobial regimens on infection and colonization with vancomycin-resistant Enterococcus faecium. Antimicrob Agents Chemother *39*, 794.

Motamedi, H., and Hutchinson, C.R. (1987). Cloning and heterologous expression of a gene cluster for the biosynthesis of tetracenomycin C, the anthracycline antitumor antibiotic of Streptomyces glaucescens. Proceedings of the National Academy of Sciences of the United States of America *84*, 4445-4449.

Moy, T.I., Conery, A.L., Larkins-Ford, J., Wu, G., Mazitschek, R., Casadei, G., Lewis, K., Carpenter, A.E., and Ausubel, F.M. (2009). High Throughput Screen for Novel Antimicrobials using a Whole Animal Infection Model. ACS chemical biology *4*, 527-533.

Mullane, K. (2014). Fidaxomicin in Clostridium difficile infection: latest evidence and clinical guidance. Therapeutic Advances in Chronic Disease *5*, 69-84.

Munch, D., Muller, A., Schneider, T., Kohl, B., Wenzel, M., Bandow, J.E., Maffioli, S., Sosio, M., Donadio, S., Wimmer, R., *et al.* (2014). The lantibiotic NAI-107 binds to bactoprenol-bound cell wall precursors and impairs membrane functions. J Biol Chem *289*, 12063-12076.

Nett, M., Ikeda, H., and Moore, B.S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. Natural product reports *26*, 1362-1384.

Newman, D.J., and Cragg, G.M. (2012). Natural products as sources of new drugs over the 30 years from 1981 to 2010. Journal of natural products *75*, 311-335.

Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C., and Piel, J. (2008). Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nature biotechnology *26*, 225-233.

Nikaido, H. (2009). Multidrug resistance in bacteria. Annu Rev Biochem *78*, 119-146.

Nowicka, B., and Kruk, J. (2010). Occurrence, biosynthesis and function of isoprenoid quinones. Biochimica et Biophysica Acta (BBA) - Bioenergetics *1797*, 1587-1605.

Ogasawara, Y., Yackley, B.J., Greenberg, J.A., Rogelj, S., and Melançon, C.E. (2015). Expanding our Understanding of Sequence-Function Relationships of Type II Polyketide Biosynthetic Gene Clusters: Bioinformatics-Guided Identification of Frankiamicin A from Frankia sp. EAN1pec. PLoS ONE *10*, e0121505.

Ogawa, H., and Natori, S. (1965). HYDROXYBENZOQUINONES FROM MYRSINACEAE PLANTS. I. RECONFIRMATION OF THE STRUCTURE OF MAESAQUINONE AND ISOLATION OF ACETYLMAESAQUINONE FROM MAESA JAPONICA MORITZI. Chem Pharm Bull (Tokyo) *13*, 511-516.

Ogawa, T., Ichimura, M., KATSUMATA, S., MORIMOTO, M., and TAKAHASHI, K. (1989). New antitumor antibiotics, duocarmycins B1 and B2. Journal of antibiotics *42*, 1299-1301.

Pelzer, S., Sussmuth, R., Heckmann, D., Recktenwald, J., Huber, P., Jung, G., and Wohlleben, W. (1999). Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908. Antimicrob Agents Chemother *43*, 1565-1573.

Pena, I., Pilar Manzano, M., Cantizani, J., Kessler, A., Alonso-Padilla, J., Bardera, A.I., Alvarez, E., Colmenarejo, G., Cotillo, I., Roquero, I., *et al.* (2015). New compound sets identified from high throughput phenotypic screening against three kinetoplastid parasites: an open resource. Scientific reports *5*, 8771.

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., and Matsunaga, S. (2004). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. Proceedings of the National Academy of Sciences of the United States of America *101*, 16222-16227.

Potterat, O., Puder, C., Wagner, K., Bolek, W., Vettermann, R., and Kauschke, S.G. (2007). Chlorocyclinones A-D, chlorinated angucyclinones from Streptomyces sp. strongly antagonizing rosiglitazone-induced PPAR-gamma activation. Journal of natural products *70*, 1934-1938.

Powers, J.H. (2004). Antimicrobial drug development--the past, the present, and the future. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases *10 Suppl 4*, 23-31.

Pozzi, R., Simone, M., Mazzetti, C., Maffioli, S., Monciardini, P., Cavaletti, L., Bamonte, R., Sosio, M., and Donadio, S. (2011). The genus Actinoallomurus and some of its metabolites. The Journal of antibiotics *64*, 133-139.

Prevention, C.f.D.C.a. (2013). ANTIBIOTIC RESISTANCE THREATS in the United States. CDC.

Pruesse, E., Peplies, J., and Glockner, F.O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics (Oxford, England) *28*, 1823-1829.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., *et al.* (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature *440*, 940-943.

Rutkowski, J., and Brzezinski, B. (2013). Structures and Properties of Naturally Occurring Polyether Antibiotics. BioMed Research International *2013*, 31.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution *4*, 406-425.

Sander, P., Springer, B., Prammananan, T., Sturmfels, A., Kappler, M., Pletschette, M., and Bottger, E.C. (2002). Fitness cost of chromosomal drug resistance-conferring mutations. Antimicrobial agents and chemotherapy *46*, 1204-1211.

Scanlon, T.C., Dostal, S.M., and Griswold, K.E. (2014). A High-Throughput Screen for Antibiotic Drug Discovery. Biotechnology and bioengineering *111*, 232-243.

Schneider, T., and Sahl, H.G. (2010). An oldie but a goodie - cell wall biosynthesis as antibiotic target pathway. Int J Med Microbiol *300*, 161-169.

Schwarzer, D., Finking, R., and Marahiel, M.A. (2003). Nonribosomal peptides: from genes to products. Natural product reports *20*, 275-287.

Segalat, L. (2007). Invertebrate animal models of diseases as screening tools in drug discovery. ACS Chem Biol *2*, 231-236.

Seto, B. (2012). Rapamycin and mTOR: a serendipitous discovery and implications for breast cancer. Clin Trans Med *1*, 1-7.

Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Current opinion in chemical biology *7*, 285-295.

Shin-ya, K., Furihata, K., Teshima, Y., Hayakawa, Y., and Seto, H. (1992). Structures of stealthins A and B, new free radical scavengers of microbial origin. Tetrahedron Letters *33*, 7025-7028.

Shlaes, D.M. (2015). Research and Development of Antibiotics: The Next Battleground. ACS Infectious Diseases.

Soderlund, C., Nelson, W., Shoemaker, A., and Paterson, A. (2006). SyMAP: A system for discovering and viewing syntenic regions of FPC maps. Genome Research *16*, 1159-1168.

Sosio, M., Gallo, G., Pozzi, R., Serina, S., Monciardini, P., Bera, A., Stegmann, E., and Weber, T. (2014). Draft Genome Sequence of the Microbispora sp. Strain ATCC-PTA-5024, Producing the Lantibiotic NAI-107. Genome Announc *2*.

Spiesecke, H., and Schneider, W.G. (1961). Effect of Electronegativity and Magnetic Anisotropy of Substituents on C13 and H1 Chemical Shifts in CH3X and CH3CH2X Compounds. The Journal of Chemical Physics *35*, 722-731.

Springer, B., Kidan, Y.G., Prammananan, T., Ellrott, K., Böttger, E.C., and Sander, P. (2001). Mechanisms of Streptomycin Resistance: Selection of Mutations in the 16S rRNA Gene Conferring Resistance. Antimicrobial Agents and Chemotherapy *45*, 2877-2884.

Stackebrandt, E., Rainey, F.A., and Ward-Rainey, N.L. (1997). Proposal for a New Hierarchic Classification System, Actinobacteria classis nov. International Journal of Systematic Bacteriology *47*, 479-491.

Strohl, W.R. (2000). The role of natural products in a modern drug discovery program. Drug discovery today *5*, 39-41.

Subramani, R., and Aalbersberg, W. (2012). Marine actinomycetes: An ongoing source of novel bioactive metabolites. Microbiological Research *167*, 571-580.

Sundqvist, M., Geli, P., Andersson, D.I., Sjölund-Karlsson, M., Runehagen, A., Cars, H., Abelson-Storby, K., Cars, O., and Kahlmeter, G. (2010). Little evidence for reversibility of trimethoprim resistance after a drastic reduction in trimethoprim use. Journal of Antimicrobial Chemotherapy *65*, 350-360.

Szu, P.-H., Govindarajan, S., Meehan, M.J., Das, A., Nguyen, D.D., Dorrestein, P.C., Minshull, J., and Khosla, C. (2011). Analysis of the Ketosynthase-Chain Length Factor Heterodimer from the Fredericamycin Polyketide Synthase. Chemistry & biology *18*, 1021-1031.

Szulawska, A., and Czyz, M. (2006). [Molecular mechanisms of anthracyclines action]. Postepy Hig Med Dosw (Online) *60*, 78-100.

Tally, F.P., and DeBruin, M.F. (2000). Development of daptomycin for Gram-positive infections. Journal of Antimicrobial Chemotherapy *46*, 523-526.

Tampakakis, E., Okoli, I., and Mylonakis, E. (2008). A C. elegans-based, whole animal, in vivo screen for the identification of antifungal compounds. Nat Protoc *3*, 1925-1931.

Tamura, T., Ishida, Y., Nozawa, Y., Otoguro, M., and Suzuki, K. (2009). Transfer of Actinomadura spadix Nonomura and Ohara 1971 to Actinoallomurus spadix gen. nov., comb. nov., and description of Actinoallomurus amamiensis sp. nov., Actinoallomurus caesius sp. nov., Actinoallomurus coprocola sp. nov., Actinoallomurus fulvus sp. nov., Actinoallomurus iriomotensis sp. nov., Actinoallomurus luridus sp. nov., Actinoallomurus purpureus sp. nov. and Actinoallomurus yoronensis sp. nov. Int J Syst Evol Microbiol *59*, 1867-1874.

Tanabe, T., Funahashi, T., Nakao, H., Miyoshi, S.-I., Shinoda, S., and Yamamoto, S. (2003). Identification and Characterization of Genes Required for Biosynthesis and Transport of the Siderophore Vibrioferrin in Vibrio parahaemolyticus. Journal of bacteriology *185*, 6938-6949.

Tanaka, T., Kawasaki, K., Daimon, S., Kitagawa, W., Yamamoto, K., Tamaki, H., Tanaka, M., Nakatsu, C.H., and Kamagata, Y. (2014). A Hidden Pitfall in the Preparation of Agar Media Undermines Microorganism Cultivability. Applied and Environmental Microbiology *80*, 7659-7666.

Tang, Y., Tsai, S.-C., and Khosla, C. (2003). Polyketide Chain Length Control by Chain Length Factor. Journal of the American Chemical Society *125*, 12708-12709.

Tängdén, T., and Giske, C.G. (2015). Global dissemination of extensively drug-resistant carbapenemase-producing Enterobacteriaceae: clinical perspectives on detection, treatment and infection control. Journal of Internal Medicine *277*, 501-512.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., *et al.* (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics *4*, 41-41.

Technikova-Dobrova, Z., Damiano, F., Tredici, S.M., Vigliotta, G., di Summa, R., Palese, L., Abbrescia, A., Labonia, N., Gnoni, G.V., and Alifano, P. (2004). Design of mineral medium for growth of Actinomadura sp. ATCC 39727, producer of the glycopeptide A40926: effects of calcium ions and nitrogen sources. Appl Microbiol Biotechnol *65*, 671-677.

Thamchaipenet, A., Indananda, C., Bunyoo, C., Duangmal, K., Matsumoto, A., and Takahashi, Y. (2010). Actinoallomurus acaciae sp. nov., an endophytic actinomycete isolated from Acacia auriculiformis A. Cunn. ex Benth. Int J Syst Evol Microbiol *60*, 554-559.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research *22*, 4673-4680.

Tocchetti, A., Bordoni, R., Gallo, G., Petiti, L., Corti, G., Alt, S., Cruz, J.C.S., Salzano, A.M., Scaloni, A., Puglia, A.M., *et al.* (2015). A Genomic, Transcriptomic and Proteomic Look at the GE2270 Producer <italic>Planobispora rosea</italic>, an Uncommon Actinomycete. PLoS ONE *10*, e0133705.

Tu, Y. (2011). The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. Nat Med *17*, 1217-1220.

van Pee, K.H. (2012). Enzymatic chlorination and bromination. Methods in enzymology *516*, 237-257.

Veneziani, M. (2009-2010). Isolamento e Caratterizzazione di Metaboliti Prodotti da un Nuovo Genere de Attinomiceti. Universita degli studi di Milano - Corso di laurea Triennale in Chimica.

Venkateswarlu, S., Panchagnula, G.K., Guraiah, M.B., and Subbaraju, G.V. (2005). Isoaurostatin: total synthesis and structural revision. Tetrahedron *61*, 3013-3017.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., and van Sinderen, D. (2007). Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. Microbiology and Molecular Biology Reviews : MMBR *71*, 495-548.

Wagenaar, M.M. (2008). Pre-fractionated microbial samples--the second generation natural products library at Wyeth. Molecules *13*, 1406-1426.

Wakabayashi, T., Kageyama, R., Naruse, N., Tsukahara, N., Funahashi, Y., Kitoh, K., and WATANABET, Y. (1997). Borrelidin is an angiogenesis inhibitor; disruption of angiogenic capillary vessels in a rat aorta matrix culture model. The Journal of antibiotics *50*, 671-676.

Walsh, C. (2003). Where will new antibiotics come from? Nature reviews Microbiology *1*, 65-70.

Wandersman, C., and Delepelaire, P. (1990). TolC, an Escherichia coli outer membrane protein required for hemolysin secretion. Proceedings of the National Academy of Sciences *87*, 4776-4780.

Wang, P., Kim, W., Pickens, L.B., Gao, X., and Tang, Y. (2012). Heterologous Expression and Manipulation of Three Tetracycline Biosynthetic Pathways(). Angewandte Chemie (International ed in English) *51*, 11136-11140.

Watve, M.G., Tickoo, R., Jog, M.M., and Bhole, B.D. (2001). How many antibiotics are produced by the genus Streptomyces? Arch Microbiol *176*, 386-390.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., *et al.* (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Research *43*, W237-W243.

Weinstein, M.J., and Wagman, G.H. (1978). Antibiotics, isolation, separation, and purification (Elsevier Scientific Pub. Co.).

Weiss, R.B. (1992). The anthracyclines: will we ever find a better doxorubicin? Semin Oncol *19*, 670-686.

Wendt-Pienkowski, E., Huang, Y., Zhang, J., Li, B., Jiang, H., Kwon, H., Hutchinson, C.R., and Shen, B. (2005). Cloning, sequencing, analysis, and heterologous expression of the fredericamycin biosynthetic gene cluster from Streptomyces griseus. Journal of the American Chemical Society *127*, 16442-16452.

Wentzel, A., Sletta, H., Ellingsen, T.E., and Bruheim, P. (2012). Intracellular metabolite pool changes in response to nutrient depletion induced metabolic switching in Streptomyces coelicolor. Metabolites *2*, 178-194.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: The unseen majority. Proceedings of the National Academy of Sciences of the United States of America *95*, 6578-6583.

Wild, J., Hradecna, Z., and Szybalski, W. (2002). Conditionally Amplifiable BACs: Switching From Single-Copy to High-Copy Vectors and Genomic Clones. Genome Research *12*, 1434-1444.

Wilgus, H.S., and Gates, J.J.W. (1967). Thiele acetylation of substituted benzoquinones. Canadian Journal of Chemistry *45*, 1975-1980.

Wise, R. (2002). Antimicrobial resistance: priorities for action. J Antimicrob Chemother *49*, 585-586.

Wolpert, M., Heide, L., Kammerer, B., and Gust, B. (2008). Assembly and heterologous expression of the coumermycin A1 gene cluster and production of new derivatives by genetic engineering. Chembiochem : a European journal of chemical biology *9*, 603-612.

Wright, G.D. (2012). Antibiotics: a new hope. Chemistry & biology *19*, 3-10.

Yang, J.Y., Sanchez, L.M., Rath, C.M., Liu, X., Boudreau, P.D., Bruns, N., Glukhov, E., Wodtke, A., de Felicio, R., Fenner, A., *et al.* (2013). Molecular Networking as a Dereplication Strategy. Journal of natural products *76*, 1686-1699.

Zhang, Q., Doroghazi, J.R., Zhao, X., Walker, M.C., and van der Donk, W.A. (2015). Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in actinobacteria. Appl Environ Microbiol *81*, 4339-4350.

Zhang, Z., Kudo, T., Nakajima, Y., and Wang, Y. (2001). Clarification of the relationship between the members of the family Thermomonosporaceae on the basis of 16S rDNA, 16S-23S rRNA internal transcribed spacer and 23S rDNA sequences and chemotaxonomic analyses. Int J Syst Evol Microbiol *51*, 373-383.