Nancy-Université
Université Henri Poincaré

Faculté des Sciences- UFR Sciences et Techniques Biologiques

Ecole Doctorale BioSE : Biologie, Santé, Environnement

Thèse

présentée et soutenue publiquement pour l'obtention du titre de

Docteur de l'Université Henri Poincaré

en Sciences de la Vie et de la Santé

par

Luisa LAURETI

Activation of a silent type I polyketide synthase gene cluster in *Streptomyces ambofaciens* ATCC23877: isolation and characterization of a novel giant macrolide

---

Activation d'un cluster de gènes de polycétide synthase de type I chez *Streptomyces ambofaciens* ATCC23877 : isolation et caractérisation d'un nouveau macrolide géant

Soutenance publique prévue le 7 janvier 2010 devant la commission d'examen :

Membres du jury :

Rapporteurs :	M. Mervyn BIBB	Professeur, John Innes Center, Norwich
	M. Jean-Luc PERNODET	Directeur de Recherche CNRS, Université Orsay
Examinateurs :	M. Frédéric BOURGAUD	Professeur, Nancy-Université
	M. Bernard DECARIS	Professeur, Nancy-Université
	M. Pierre LEBLOND	Professeur, Nancy-Université (Directeur de thèse)
	M. Bertrand AIGLE	Maître de Conférences, Nancy-Université

Laboratoire de Génétique et Microbiologie UMR INRA-UHP 1128
Faculté des Sciences et Techniques – 54506 Vandœuvre-lès-Nancy

The Fishermen know that the sea is dangerous and the storm terrible,
but they have never found these dangers sufficient reason for remaining ashore.
(Vincent van Gogh)

ACKNOWLEDGEMENTS/REMERCIEMENTS

Je remercie le Professeur **Pierre Leblond** pour m'avoir accueillie au sein du Laboratoire de Génétique et Microbiologie et pour m'avoir permis d'entreprendre cette thèse.

I would like to thank Professor **Mervyn Bibb** and Professor **Jean-Luc Pernodet** for accepting to be my referees. For me it has been a great honour and I hope for them it will be a pleasure to read this thesis.

Je voudrais aussi remercier le Professeur **Frédéric Bourgaud** et le Professeur **Bernard Decaris** pour avoir accepté d'être mes examinateurs.

I would like to thank Professor **Greg Challis** and the Dr **Lijiang Song** and Dr **Christophe Corre** for welcoming me in the lab and for all the precious help in the discovery and the characterization of sambomycin, our "little baby". I really learned a lot from them and I enjoyed my staying in England very much.

Je tiens surtout à remercier mon "dear boss" (☺) le Docteur **Bertrand Aigle** qui a toujours cru en moi et dans ce projet, même quand moi je n'y croyais plus ! Merci pour la confiance que tu m'as accordée tout au long de cette thèse et pour tous tes conseils professionnels, mais aussi personnels. Merci d'avoir accepté mon « sacré caractère » de petite Calimero et d'avoir été toujours très compréhensible et disponible avec moi ! J'ai beaucoup appris de toi. Ton enthousiasme et ta rigueur scientifique resteront toujours un exemple pour moi. Enfin, travailler côte à côte avec toi a été un vrai plaisir et je sais que tout ça va beaucoup me manquer.

Cette thèse n'aurait jamais été achevée sans l'aide et le soutien de toutes les personnes que j'ai eu la chance de rencontrer à Nancy et de celles qui faisaient déjà partie de ma vie. Je suis devenue celle que je suis aussi grâce à eux !

Tout d'abord, je tiens à remercier de tout mon coeur l'équipe *Streptomyces* ☺ :
Laurence, pour sa gentillesse infinie et pour tous les petits moments de bonheur et de pause thé partagés avec moi. Je me suis sentie en famille avec toi, Christophe et les enfants ! **Robert**, un grand merci à toi pour ton aide à la paillasse et pour ta patience à répondre à mes questions et à mes doutes. Un grand merci également pour avoir été mon partenaire de badminton et de danse, on s'est beaucoup éclatés ensemble. Sans toi le labo n'a plus été pareil ! Merci à **Caths**, toujours pleine de bonne humeur, d'énergie et toujours disponible envers les autres… promets-moi d'apprendre à dire non de temps en temps ! ;-) Thanks a lot to **Isolde**, the other foreign English speaking in the lab ! Even if we met at the end of this adventure, we had the chance to share a

lot of nice and funny moments and I wish you all the best for your future! Merci à **Annabelle**, pour son accueil chaleureux et pour son aide à chercher mon petit coin douillet, sans même me connaître auparavant ! Merci pour m'avoir laissée faire du babysitter au petit Daniel, j'ai vraiment aimé !

Je tiens aussi à remercier toutes les autres personnes du laboratoire présentes, passées ou venant juste d'arriver, avec qui j'ai eu le plaisir de partager ces trois années de thèse et qui m'ont permis de ne pas trop sentir la solitude et la distance avec mon pays et mes proches.
En particulier, merci à **Catherine**, **Virginie**, **Séverine**, **Céline, Xavier** et **Stéphane** pour toutes les petites soirées gourmandes et arrosées passées ensemble. Merci à mes « collègues » thésards **Nico**, **Ludo** et **Romain** à qui je souhaite très bon courage ! Merci aux « petits » du labo (comme je les appelle amicalement) **Julien**, **Aurore**, **Emilie**, **Fabien**, **Justine** et **Maxime** car ils contribuent à rendre le labo très sympa et rigolo! Merci à **Sophie** pour son super talent d'organisation ! Merci enfin à **Fred** et **Julie**, mes deux amis en dehors du labo, avec qui j'ai passé des moments très sympas, oubliant pour quelques heures mon boulot !

A special thank to my dear friend **Bjørn**, who has always been there for me for better or for worse, even being across the ocean as proof that nothing can ruin our friendship! Thank you very much for reading and correcting my manuscript, you gave me very useful advices and comments, but most of all you always encouraged me in my work because you know how silly I can be when an experiment doesn't work as I want!! Waar ik ook heen zal gaan, ik zal je altijd in mijn hart dragen.

Un grazie infinito ai miei cari amici italiani **Gi**, **Valentina** e **Laura** che sono sempre riusciti a infondermi coraggio e fiducia e che per dimostrarmi il loro affetto sono venuti fino a Nancy! ☺

E infine voglio ringraziare la mia **famiglia** che mi ha sostenuto per tutto questi anni, standomi sempre vicino e facendomi sentire il loro affetto incondizionato. Se sono quella che sono, lo devo soprattutto a voi e all'esempio di amore, rispetto, altruismo e professionalità che mi avete sempre dato.

INDEX

ABBREVIATIONS

A: Adenylation domain
ABC: ATP-binding cassette
ACP: Acyl carrier protein domain
ADS: Amplified DNA sequence
AHL: Acylated homoserine lactone
AT: Acyltransferase domain
ATP: Adenosine 5′-triphosphate
AUD: Amplifiable unit of DNA
BAC: Bacterial artificial chromosome
C: Condensation domain
CLF: Chain length factor
Cy: Cyclization domain
CYP: Cytochrome P450
DAD-HPLC: Diode array-high performance liquid chromatography
DEBS: Deoxyerythronolide B synthase
DH: Dehydratase domaine
E: Epimerization domain
ER: Enoylreductase domain
FAS: Fatty acid biosynthesis
FDA: Food and Drug Administration
GlcNAc: N-acetylglucosamine
GT: Glycosyltransferase
HPLC: High performance liquid chromatography
HR: Histidine kinase sensor
kb: Kilobases
KR: Ketoreductase domaine
KS: β-ketoacyl synthase
KSQ: β-ketoacyl synthase with a glutamine instead of a cysteine in the active site
LAL: Large ATP-binding regulators of the LuxR family
LC-MS: Liquid chromatography coupled with mass spectrometry
MRSA: Methicillin-resistant *Staphylococcus aureus*
MW: Molecular weight
NADH: Nicotinamide adenine dinucleotide hydrogenated
NMet: N-methyltransferase
NRPS: Non-ribosomal peptide synthase
OSMAC: One strain-many compounds
PCP: Peptidyl carrier protein domain
PKS: Polyketide synthase
PKS/NRPS: Hybrid polyketide/non ribosomal peptide synthase

ppGpp : Highly phosphorylated guanosine nucleotide
PPTase: Phosphopantetheinyl transferase
RR: Response regulator
SAM: *S*- adenosyl-L-methionine
SARP: *Streptomyces* antibiotic regulatory protein
SDS-PAGE: Sodium dodecylsulfate polyacrilamide gel electrophoresis
TCS: Two component system
TDP: Thymidine diphosphate
TE: Thioesterase
TEII: Type II thioesterase
TIR: Terminal inverted repeated region
TRP: tetratricopeptide repeat

INTRODUCTION

1. GENERAL CONTEXT

Infectious diseases are the leading cause of human morbidity and mortality worldwide; in 1998 the World Health Organization reported 13.3 million deaths. In the last two decades, new infections (e.g., Lyme and Legionnaire's disease, HIV, ebola), compounded by the reemergence of old ones (e.g., cholera, tuberculosis, malaria) and the rising of multi-drug resistant pathogens highlighted dramatically the need for new and more effective antimicrobial agents. The problem is not only from a health point of view, but it also has a significant impact on the politics and economy of all the countries, as demonstrated nowadays by the spreading of H1N1 virus all over the world. To give an example, the estimated annual cost of antimicrobial resistance in hospitals is $4.5 billion and $122 million for the sole *Staphylococcus aureus* (Cassell and Mekalanos, 2001). Except for an oxazolidone antibacterial drug, approved in 2000 by the Food and Drug Administration (FDA), for more than 25 years ago no other new chemical classes were introduced in the market, leaving many infectious diseases still untreatable (Cassell and Mekalanos, 2001). Therefore, the challenge now is to discover novel and effective anti-infective agents possessing pharmaceutical properties, such as high target specificity combined with a good bioavailability and low or no toxicity, as required more and more by the Food and Drug Administration. Besides, the new drugs should have low cost of production to be appealing to the pharmaceutical companies.

At the end of the 19th century, Cohn, Pasteur and Koch demonstrated that microorganisms (bacteria, fungi and virus) were the causative agents of infectious diseases, leading the way for a new microbiologist era. The first natural antibacterial agent, discovered by Alexander Fleming in 1928, was penicillin, a secondary metabolite produced by *Penicillum notatum* (now called *P. chrysogenum*). Few years later, Gerhard Domagk synthesized the first sulfonamide used in the struggle against bacterial infections. It was only in 1942 that penicillin was introduced in human therapy, a few years before Waksman isolated streptomycin from *Streptomyces griseus*, which proved to be at that time a "miraculous" active agent against tuberculosis.

Since the early breakthroughs of Fleming and Waksman, the discovery of novel natural antibacterial and antifungal compounds, produced mostly by fungi and bacteria, experienced an exponential phase, especially between 1950s and 1960s, the so called "Golden Age". Many of these compounds found important pharmaceutical applications, revolutionizing medicine (see Table 1), but a considerable proportion was also discarded because of their toxicity. However lately, these natural products have been revaluated for other medical applications, for example as anticancer drug due to their ability to inhibit proliferation of eukaryotic cells (e.g., doxorubicin). Rapamycin and mycophenolic acid were "rediscovered" as agents for organ transplantation due to their immunomodulation property. Fungal statins are one of the best selling cholesterol-lowering agents (Marinelli, 2009).

Antibiotic	Year deployed	Resistance observed
Sulfonamides*	1930s	1940s
Penicillin	1942	1946
Streptomycin	1943	1959
Chloramphenicol	1947	1959
Tetracycline	1948	1953
Erythromycin	1952	1988
Vancomycin	1956	1988
Methicillin	1960	1961
Ampicillin	1961	1973
Cephalosporins	1960s	late 1960s

Table 1 List of representative antibiotics discovered in the Golden Age and therapeutically in use today, correlated to the appearance of the first resistance strain (from Palumbi, 2001).*the first synthetic antibiotic.

Economically and, more importantly, therapeutically, the repercussions of these microbial compounds seemed to be endless and in 1969 the US surgeon general claimed that "the war against infectious disease had been won". The rapid and successful development of these drugs and the general confidence in the existing compounds were followed by a lag in the search for natural antimicrobial agents by the pharmaceutical companies. For the economical incomes, pharmaceutical companies favored the high-throughput screening of synthetic libraries; investigation of natural sources instead presents more difficulties, in particular accessing the untapped biological resources. Besides, as a matter of fact, companies are more interested in producing and selling "quality of life" drugs for chronic diseases, such as cholesterol-lowering or hypertensive agents, than in selling antibiotics used only for a short period of time (Li and Vederas, 2009). Infections are a huge human scourge for those countries that cannot afford the cost of these medicines, such as the developing and third-world countries. However, the fast spreading of multi-drug resistance strains (e.g., methicillin-resistant *Staphylococcus aureus* strains), together with the reemergence of old diseases and the problems with treatment of fungal infections in immuno-compromised patients, gave a new driving force to the search of novel natural antimicrobial agents.

The potential problem of antibiotic resistance was unfortunately underestimated at first, despite resistant strains appeared immediately after the introduction of antibiotics in human therapy; already in 1955 the first multi-drug resistance strain appears. At that time, the belief was that resistance was specific and mutations occurred with a low frequency, thus the development of antibiotic-resistant strains would be an unlikely and rare event (reviewed by Davies, 1995). Later on, it became apparent that bacterial resistance can be natural or acquired through mutations or, more commonly, through horizontal transfer. The first evidences came from works

in Japan at the end of 1950s since the country was stroke by serious outbreaks of dysentery caused by *Shigella*. Numerous isolates showed to be soon resistant to four different antibiotics (Kitamoto *et al*., 1956) and several laboratories identified a plasmid associated with the mechanism of resistance-gene transfer (Harada *et al.*, 1960; Nakaya *et al*., 1960).

To cope with antibiotic resistance, the initial approach was to chemically modify the pre-existent antibiotics, creating a novel generation of known antibiotics which was successful for a limited period of time, as the case for the methicillin and related β-lactams (as reported by Davies, 2007). Bacterial resistance is inevitable (selection favours resistant strain), but it can be managed with proper and regulated medical treatments. Indeed, the increasing number of resistance strains is mostly due to the over-use and indiscriminate prescription and dosage of antibiotics in medical therapy, but also in agriculture (as growth factor promoters) and aquaculture, which worsen the spreading out. Therefore, previous diseases, which were thought to be under control, now represent a huge mankind problem (Davies, 2007).

Alternatives to antimicrobial agents in order to fight infections are under development and principally involved vaccines or other therapeutical molecules (Demain and Adrio, 2007), but antibiotics discovery can not be completely replaced. The search is not over and the fact that more than half of the drugs approved, either antibacterial or others, are based on natural compounds proves that the nature has still to disclose plenty of bioactive molecules [see Newman and Cragg (2007), for a comprehensive review on the subject]. Hence, new scientific approaches together with the economical support of governments must aim to exploit microbial natural resources to find novel powerful anti-infective compounds.

This PhD project inserts into this biological context. It endeavoured to contribute to the research for a novel molecule with interesting antibacterial or other biological properties. The natural reservoir used was *Streptomyces ambofaciens*, which like all the actinomycetes is an important producer of secondary metabolites.

2. SECONDARY METABOLISM AND MICROBIAL NATURAL PRODUCTS

The German plant physiologist Bu'Lock introduced for the first time the term "secondary metabolite" to define the metabolic products which derive from the morphological differentiation of restricted taxonomic groups, mainly from bacteria of the order of *Actinomycetales* and from fungi (Bu'Lock, 1961). They are generally produced at the end of the vegetative growth phase, or in nutrient limiting conditions, probably triggered by specific cellular signallings. For instance, in *Streptomyces*, the production occurs in a growth-phase dependent manner; generally it starts at the early stationary phase for liquid cultures and coincides with morphological differentiation in agar-grown cultures (Bibb, 2005). Secondary

metabolite pathways are considered dispensable under several growth conditions and their products are not essential for the survival of the microorganisms, at least in the laboratory conditions.

Microbial secondary metabolites are generally low molecular weight molecules (<3000 Da) and their chemical structures are extremely varied and complex. Most of them endow a biological activity which makes them very interesting from a medical point of view, thus having a tremendous impact on the economy (Marinelli, 2009). More than 20,000 bioactive molecules were identified so far and some of them are exploited, for example, as anti-infective agents, antivirals, anticancer, antiparasitics, bioinsecticides and herbicides (Fig.1). Indeed, in the last decades, the use of secondary metabolites in veterinary medicine and agriculture to fight against worm infections or parasites in farm animals has significantly increased (Berdy, 2005).

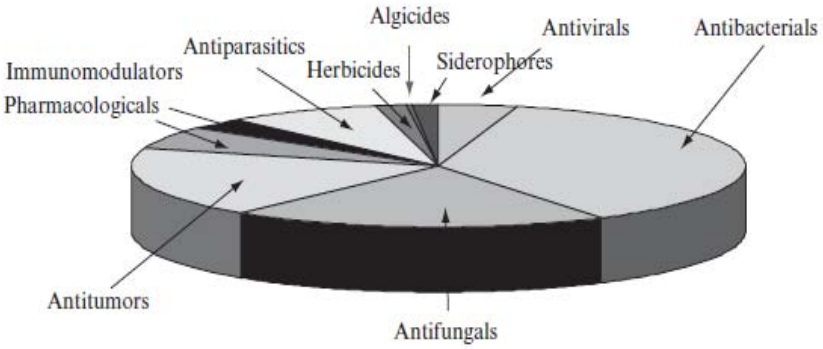


Fig. 1 Distribution of secondary metabolites according to the biological property. The term pharmacologicals refers to the substances that act on the physiology of higher animals (Marinelli, 2009).

Several hypotheses are formulated on the possible natural role of these molecules. Most probably, they are produced under stress conditions, such as nutrient starvation, to protect and preserve the producer from the other competitors present in the same environment (Marinelli, 2009). Indeed, filamentous fungi and actinomycetes have a no motile and saprophytic life cycle in a complex habitat such as terrestrial soil, where competition with other inhabitants is high.

Another proposed function is their involvement in signalling mechanisms. It has been well demonstrated that at a very low concentration (the so called subinhibitory concentrations) antibiotics induce changes in expression profiles on a wide range of genes, not necessary involved with the target function (Davies *et al.*, 2006). These studies correlate antibiotics with the phenomenon of hormesis: low concentrations of antibiotics would represent the real natural environmental function, while high concentrations represent the therapeutic function (Davies *et al.*, 2006; Mesak and Davies, 2009). Along with the idea that antibiotics are signalling molecules to control and influence the homeostasis of bacterial community, Martinez and co-workers (Linares *et al.*, 2006) proposed that antibiotics, at a low concentration, can induce biofilm formation in *Pseudomonas aeruginosa*. Biofilms are a good example of mechanism

used by microbial community to exchange “information” by transfer of genes (Parsek and Fuqua, 2004). Hence, the natural role of antibiotics might be to trigger changes that are advantageous for the survival of the community or for the microbial cooperation and adaptation in the ecological niche.

2.1 Interconnection between primary and secondary metabolism

Secondary metabolite biosynthesis is closely related to the so called “primary” metabolism, which is responsible for the production of the essential biological components, such as amino acids, nucleotides, vitamins and small organic acids. These molecules are produced by all living cells and are intermediates or end products of the metabolic pathways, or are the building blocks for essential macromolecules (Demain and Adrio, 2007).

It clearly appears that primary and secondary metabolisms are deeply interconnected, both in terms of precursor supply and through nutrient regulation (chapter 4 of introduction). For instance, antibiotics are synthesized from 12 precursor metabolites which are all used for the synthesis of the key cellular constituents (Rokem *et al.*, 2007). These precursors come in turn from different carbon sources used by the cell, as well described by the bow-tie structure of Figure 2. Another link between primary and secondary metabolisms is represented by the common use of cofactors: ATP for the energy input to drive secondary metabolite biosynthesis, NADH and NADPH as electron acceptors (see also Fig. 2). In particular, ATP was shown to be a key cofactor of polyphosphate kinase Ppk, which is involved in the repression of antibiotic production in *Streptomyces lividans* (Ghorbel *et al*., 2006).

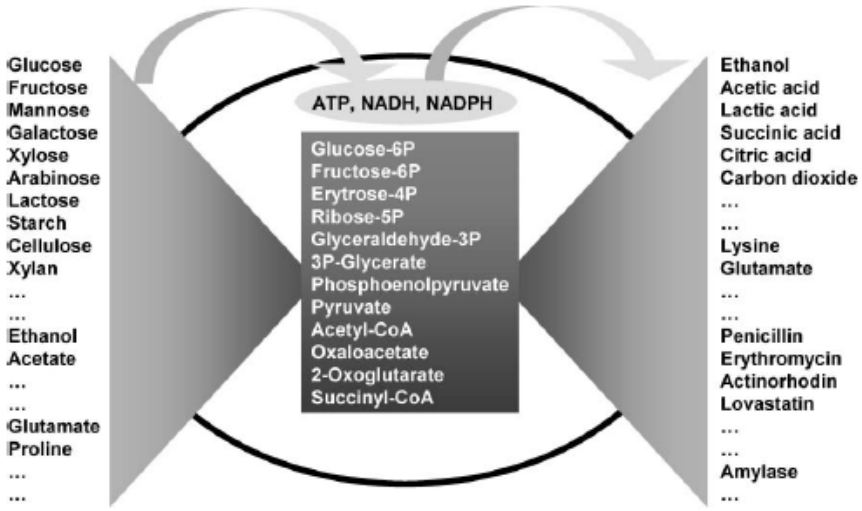


Fig. 2 Schematic representation of the bow-tie structure of metabolism. On the left several carbon sources which, once degraded, supply the 12 metabolic precursors necessary for the biosynthesis of amino acids, lipids, nucleotides and also secondary metabolites (on the right). In the same time the degradation of these products restore the metabolic precursor pool creating an open metabolic flux (from Rokem *et al.*, 2007). In particular, acetyl-CoA is the most used precursors to synthesize secondary metabolites like polyketides and isoprenoids.

Thereby, a way to positively affect secondary metabolite production could be to engineer primary metabolism overbalancing the flux of precursors and cofactors, improving their availability. Manipulation of central carbohydrate metabolism has already been shown to lead to an increased production of the antibiotic actinorhodin in *Streptomyces coelicolor* and *Streptomyces lividans* (Ryu *et al.,* 2006; Butler *et al.,* 2002). In both cases, the researchers tried to channel carbon flux, in particular that of acetyl-CoA and malonyl-CoA precursors, towards secondary metabolite biosynthetic enzymes.

2.2 The antibiotics

From an industrial and economical point of view, molecules showing antibacterial properties are the most important and exploited group of secondary metabolites, as underlined in the diagram of Figure 1. Consequently, the research has been predominantly focused in the discovery and the characterization of these particular compounds.

Waksman defined the word antibiotic ("against life") as "a natural chemical substance, derived from living microorganisms, which has the capability, at a low concentration, to inhibit growth or even destroy other microorganisms without harming the eukaryotic host". Antibiotics are mainly produced by fungi (30%) and actinomycetes (53%), especially the genus *Streptomyces* (40%), but also by other bacteria (17%), such as *Bacillus, Pseudomonas*, Myxobacteria and Cyanobacteria (Berdy, 2005).

The definition of antibiotics encloses natural products as well as natural products which undergo several chemical modifications to improve their efficacy and pharmacokinetic properties, the so called semisynthetic antibiotics. Synthetic antibacterial agents are instead entirely chemical synthesised (e.g. quinolones, sulphonamides, oxazolidinones).

Antibiotics and synthetic antibacterials can be grouped on the basis of their chemical structure, as listed in Tables 2 and 3, or of their mechanism of action (Mascaretti, 2003). The targets of antibacterial drugs might be either the cellular structures or a target enzyme. The most common mechanisms of action involve the inhibition of the bacterial cell wall biosynthesis (e.g., β-lactams, glycopeptides); the inhibition of protein, RNA or DNA synthesis (respectively, macrolides, ansamycins, quinolones); and the damage of cell membranes (polymyxins).

Antibiotics	Examples
β-lactams	(penicillins, cephalosporins, clavulanic acid,…)
Aminoglycosides	(streptomycin, neomycin, kanamycin,…)
Macrolides	(erythromycin, oleandomycin, spiramycin,…)
Ketolides	(telithromycin)
Tetracyclines	(tetracycline, demeclocycline, doxycycline…)
Lincosamides	(clindamycin)
Ansamycins	(rifampicin)
Glycopeptides	(vancomycin, teicoplanin)
Chloramphenicol	
Cyclic peptides	(bacitracin, polymyxin B)
Streptogramins	(quinupristin, dalfopristin)
Mupirocin	
D-cycloserine	
Fosfomycin	

Table 2 Classification of antibiotics (natural or semisynthetic) according to the chemical structure (Mascaretti, 2003).

Synthetic antibacterials	Examples
Oxazolidinones	(linezolid)
Quinolones	(nalidix acid, ciprofloxacin, …)
Sulfonamides	(sulfamethoxazole)
Diaminopyrimidines	(trimethoprim)
5-Nitroimidazoles and 5-Nitrofurans	(metronidazole, furazolidone)

Table 3 Classification of synthetic antibacterials according to the chemical structure (Mascaretti, 2003).

However in the last decade, antibiotics displaying a novel mechanism of action have been discovered through a specific screening for new antibacterial targets. In particular, platensimycin was shown to block the bacterial type II fatty acid biosynthesis (Wang *et al*., 2006), while IXP1, a cyclic peptide, and several acyldepsipeptides inhibit the ClpPX protease complex and the ClpP component, respectively (Cheng *et al*., 2007; Brötz-Oesterhelt *et al*., 2005). Since these compounds showed no cross-resistance to other known antibiotics, they are an interesting and suited alternative in the treatment of multidrug resistance pathogens.

The chemical diversity observed for antibiotics and more generally for secondary metabolites results from the different and complex post-modification reactions that these compounds can undergo, such as methylation, glycosylation or oxidation (Fig 3). For example, the macrolactone ring 6-dEB, precursor of erythromycin, undergoes two glycosylation and two hydroxylation reactions before generating the active antibiotic erythromycin A.

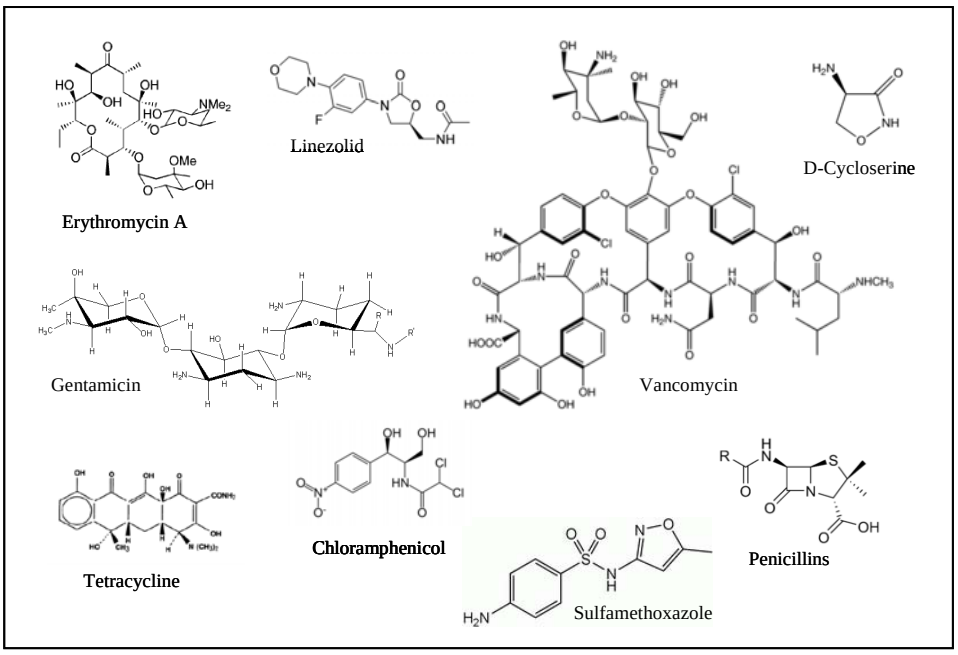


Fig. 3 Structures of some natural and synthetic antibacterial drugs.

2.3 Mechanisms of bacterial resistance

Microorganisms that produce bioactive compounds show no sensitivity to their own product; therefore they must contain a self-protection system, which probably co-evolved together with the biosynthetic genes. Indeed, genes encoding enzymes for bacterial resistance are often identified inside secondary metabolite clusters. In the same time, non-producing microbes, to be able to compete and survive in their environment, need to have a bacterial resistance system (Hopwood, 2007). Three general strategies have been described and characterized, which might sometimes overlap and be used for self-protection and resistance at one time.

2.3.1 Active-efflux pumps

Levy and co-workers (McMurry *et al.*, 1980) were the first to demonstrate that an active efflux pump, mediated by a transmembrane protein, named TET, was responsible for tetracycline resistance in *E. coli*. In gram-positive bacteria, these transmembrane proteins are located in the cytoplasmic membranes, while in gram-negative bacteria they are also found in the outer membranes (Walsh, 2003). The pumps can be driven by proton motive force or by ATP hydrolysis, in order to export antibiotics or other bioactive compounds, thus decreasing their concentration inside the cell. The pumps can be specific for one chemical class, like the TET pump for tetracyclines, but more often they have a broader range of substrate specificity, typically found in multi-drug pathogens (Van Bambeke *et al.*, 2000).

Bacterial efflux pumps can be divided in five protein families, according to the amino acid sequence identity and mode of energy coupling: the major facilitator subfamily (MFS), the small multidrug regulator family (SMR), the resistance-nodulation-cell division family (RND) and the multidrug and toxic compound extrusion family (MATE) use an electrochemical gradient of cations (H^+ or Na^+) for drug transport; whereas the ATP-binding cassette family (ABC) hydrolyzes ATP (Mascaretti, 2003).

Some ABC transporters, whose genes were found in secondary metabolite clusters, have been reported to have a resistance-independent role. OleB is an ABC transporter from *S. antibioticus* responsible for the secretion of the glycosylated inactive form of the antibiotic oleandomycin (Hernandez *et al*., 1993; Quiros *et al*., 1998). Recently, Menges and co-workers (2007) have characterized the ABC transporter Tba from *Amycolatopsis balhimycin*, which exports the glycopeptide balhimycin, without taking part in any resistance mechanism to the antibiotic. Intracellular accumulation of balhimycin was observed in the *tba* mutant strain.

2.3.2 Enzymatic inactivation of the antibiotic

Another strategy that bacteria can employ is to inactivate the secondary metabolite changing its chemical structure. Synthetic antibacterials, as fluoroquinolones or oxazolidinones, are the only class not affected by this resistance mechanism (Walsh, 2003). β-lactamases are well characterized bacterial enzymes that hydrolyze the amide bond of the four-membered β-lactam ring of penicillins, cephalosporins and carbapenems, which are no longer able to bind and inactivate their cellular target, the transpeptidases (Wilke *et al*., 2005). With an analogous mechanism the reactive epoxide ring of fosfomycin is opened by a glutathione S-transferase (Arca *et al*., 1990). Three types of enzymatic modification are known to inactivate aminoglycoside antibiotics which specifically interact with the 16S rRNA, inhibiting protein biosynthesis (Kotra *et al*., 2000). The -OH groups can undergo phosphorylation or adenylation, the $-NH_2$ group an acetylation; all these reactions are irreversible.

2.3.3 Modification of the target site

Bacteria are also able to modify the drug target, without affecting any cellular function. These modifications often occur to the cell wall components or the ribosomal subunits (Walsh, 2003). Methicillin-resistant *Staphylococcus aureus* (MRSA) strains have acquired, through horizontal transfer, a 30-40 kb mobile DNA element containing the *mecA* gene which encodes a novel type of transpeptidase, insensitive to all β-lactam antibiotics (Hiramatsu *et al*., 2001). Macrolides interact with the 23S rRNA in the 50S ribosomal subunit: mono or di-methylation of the amino group of the adenine residue A_{2058} decrease the affinity of the antibiotic for the RNA. This mechanism was identified and characterized in erythromycin and tylosin producer strains (Zalacain and Cundliffe, 1989).

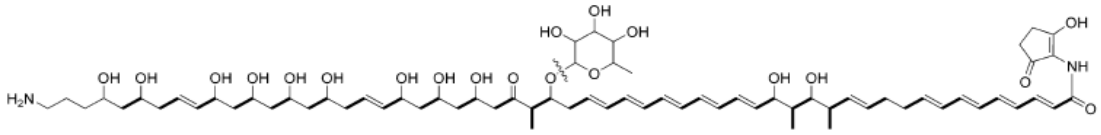
3. STRATEGIES TO DISCOVER NOVEL NATURAL BIOACTIVE MOLECULES

The constant need for new antimicrobial drugs focuses the attention on the huge and last to be completely explored potential coming from microbes and also plants. The traditional methods, used to discover most of the natural pharmaceuticals up to date, were mainly based on the phenotypic and bioassay-guided screening, but their limits showed up soon. The processes were long, laborious and iterative, and even though they were successful in the isolation of the potent antibiotic platensimycin (Wang *et al*., 2006) more often they lead to the "discovery" of already known compounds. In order to avoid this inconvenience, it is important to introduce elements of novelty in the screening programme, as exploring untapped microbial reservoirs, improving the bioassay sensitivity or finding new bacterial targets (Donadio *et al*., 2009). The advent of new chemical genetic tools and high-throughput screening platforms has already led and hopefully will continue to lead in accessing the natural product diversity of microorganisms, since at the moment 99% of the microorganisms can not still be cultured (Schloss and Handelsman, 2005).
Recently, genomic-guided approaches came to the fore to help in the discovery of new natural products thanks to the increasing number of completed genome sequence projects (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). Thus an enormous amount of data is now available and ready to be exploited: a new genomics era just begins. Indeed, after the genome sequencing project of several microorganisms, especially from the order of actinomycetes, fungi and also plants, it became evident that their potential as secondary metabolite producers was underestimated and that the reservoir of new natural products was far from being exhausted. For example, the genome of *Streptomyces coelicolor* A3(2), which was thought to produce only six compounds, revealed the presence of 18 additional orphan clusters (Bentley *et al.* 2002). Likewise, *Streptomyces avermitilis* genome contains a total of 30 secondary metabolite clusters, much more than the three antibiotics so far known (Ikeda *et al.*, 2003). Of relevant importance are also the genome analysis of other actinomycetes, such as *Rhodococcus* sp. RHA1 (McLeod *et al*., 2006) and *Salinispora tropica* (Udwary *et al*., 2007), and of fungi (Nierman *et al*., 2005; Pel *et al*., 2007), which demonstrated to have a wide reservoir of secondary metabolite pathways.
The major aim of the genomic-guided strategies is to identify biosynthetic pathways responsible for the production of novel bioactive molecules. A secondary metabolite cluster is defined as a locus in which the genes for the biosynthesis of a specific product, as well as genes for the regulation, the transport or the resistance mechanism are clustered all together. A cluster is said "orphan" (or cryptic) when the corresponding metabolite is unknown (Gross, 2007). Because the genes are all clustered together and the biosynthetic genes are usually associated to a known chemical class, such as terpene cyclases, polyketide and non-ribosomal peptide synthases (PKS and NRPS), the identification of these pathways and the isolation of the natural product is facilitated.

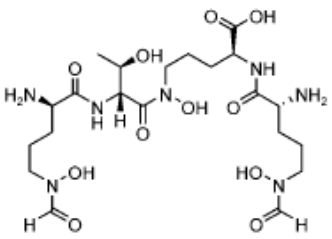
3.1 Genome mining

The bio-informatics search for secondary metabolite clusters is generally referred to as genome mining; the whole genome sequence of a microorganism is scanned *in silico* looking for consensus motifs, typical of secondary metabolite pathways. Once a cluster is identified and when is possible, the following and equally important step is the *in silico* prediction of the putative chemical structure derived from it. The current paradigms obtained from the studies of PKS and NRPS clusters allowed a better understanding of the biosynthetic mechanisms, leading to an exhaustive structural prediction (Fischbach and Walsh, 2006). At this time, it is possible to predict with a high degree of confidence the type and number of substrate precursors, the stereochemistry, the order of enzymatic assembly and possible post-modifications reactions operated by tailoring enzymes, in particular for the modular PKS clusters. Based on the chemical structure prediction, physico-chemical properties can also be deduced, which facilitate the isolation and detection of the compound. Therefore, targeting the new metabolite is significantly simplified, but other genetic and technical approaches (see after in the text) are usually combined to isolate the natural product.

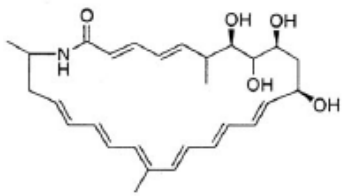
Ecopia BioSciences Company was spearheaded of this innovative strategy, proving immediately effective in several cases, as shown by the work of McAlpine and co-workers (2005) who discovered and elucidated the structure of a novel antifungal agent (ECO-02301) from *Streptomyces aizunensis* NRRL B-11277 (see Fig. 4). The fruitful uses of genome mining and structural prediction have also been exemplified by the discovery of the siderophore coelichelin (Fig. 4) in *S. coelicolor* (Lautru *et al*., 2005) or the polyene salinilactam A (Fig.4), produced by the marine actinomycetes *Salinispora tropica*, in which the polyene structure absorbs UV light at a characteristic wavelength (Udwary *et al*., 2007).



ECO-02301



Coelichelin



Salinilactam A

Fig. 4 Chemical structures of ECO-02301, coelichelin and salinilactam A, identified by genome mining, in particular by structural prediction from which physico-chemical properties can derive.

3.1.1 OSMAC approach

The majority of these orphan clusters are "silent" or expressed at a very low level, at least under laboratory conditions, even if the enzymatic genes appear functional. In order to induce the expression of the biosynthetic genes, different culture and growth conditions need to be tested. The "One strain-many compounds" (OSMAC) approach (Bode *et al*., 2002) showed that simple variations in cultural parameters, as media composition, temperature, aeration rate etc, can allow isolation of several products, which were otherwise hidden. At the base of this phenomenon, there are probably effects on the transcriptional, translational or enzymatic level. Bode and co-workers (2002) showed that by changing the fermentor conditions, the media composition and the pH value they were able to isolate some of the metabolites produced by *Streptomyces* sp. Gö40/14.

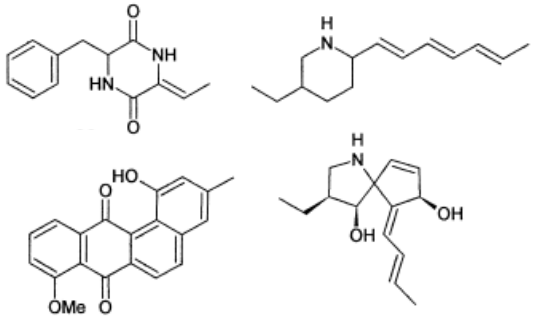


Fig. 5 Some of the secondary metabolites isolated from *Streptomyces* sp. Gö40/14 using OSMAC approach (Bode *et al*., 2002).

3.1.2 Genomisotopic approach

The genomisotopic approach has first been applied to NRPSs or hybrid NRPS/PKSs products (the product is synthesised by a NRPS and a PKS cluster, combined together), but extensions to other secondary metabolites classes can be envisioned in the future. This technique combines genomic sequence analysis, in particular an accurate prediction of substrate incorporation, with an isotope-guided fractionation to identify cryptic metabolites (Gross *et al*., 2007). If the bio-informatic prediction revealed a specific precursor incorporated in the metabolite structure, it can be isotopically labelled and used to feed the producing-strain. The precursor can easily be tracked though NMR isolation process. It is indispensable to check if the orphan cluster is expressed in the conditions tested, otherwise other approaches can be integrated (e.g. OSMAC), and to choose a highly specific precursor, to avoid incorporation in the other cell metabolites.

The genome of *Pseudomonas fluorescens* Pf-5 revealed a NRPS gene cluster whose product was unknown. From the *in silico* analyses, the product was predicted to be a lipopeptide, containing four leucines. Feeding experiments, using ^{15}N-labelled leucine, together with isotope-guided fractionation, using ^{1}H-^{15}N-HMBC NMR, led to the identification of a new class of bioactive cyclic lipopetides, orfamide A, B and C (Gross *et al*., 2007).

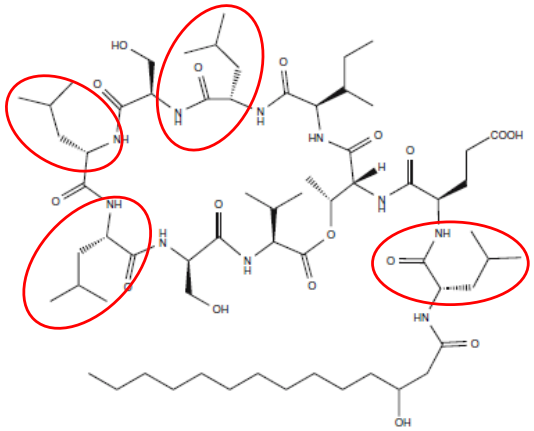
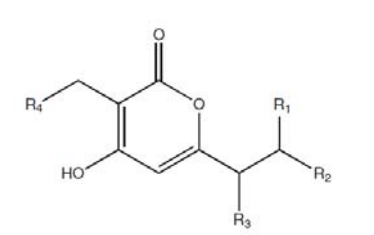


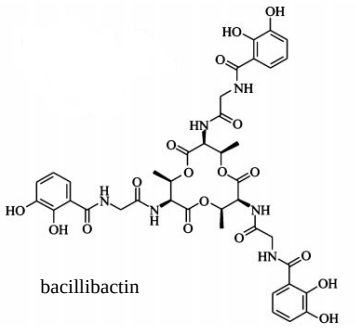
Fig. 6 Structure of the lipopeptide orfamide A (from Gross H *et al*., 2007). The four leucines are circled in red.

3.1.3 Gene inactivation and comparative metabolic profiling approach

This alternative method does not necessarily require structural predictions of the expected metabolite and directly indicates the link between the metabolite and the orphan biosynthetic cluster, identified by genome mining. A gene of interest in the cluster is disrupted and the metabolic profile of the mutant strain is compared with that one of the wild type strain, employing analytical techniques such as LC-MS or DAD-HPLC to identify the potential natural product. The limits can derive from the level of expression, which is the case for silent clusters, and from the construction of the mutant strain, since for many organisms genetic tools are still unavailable. This method was applied successfully for the discovery of the cathecolic siderophore bacillibactin, from *Bacillus subtilis* (May *et al*., 2001) and for the discovery of three new metabolites of *S. coelicolor*, isogermicidin A, B and germicidin C (Song *et al*., 2006) (See Fig. 7).



R1=R2=R4=CH_3; R3=H (isogermicidin A)
R1=R3=H; R2=R4=CH_3 (isogermicidin B)
R1=R4=H; R2=R3=CH_3 (germicidin C)



bacillibactin

Fig. 7 Structures of isogermicidin A, B, germicidin C and bacillibactin identified by gene inactivation/comparative metabolic profiling approach.

3.1.4 Heterologous expression

An orphan secondary metabolite cluster can be cloned and transferred in a heterologous host, modified to optimise the expression and the production. This method offers several practical advantages, especially when the cluster is found into a strain difficult to cultivate, manipulate or with a low level of metabolite production, when no possible prediction is available or when the cluster is silent. Therefore, heterologous expression allows analyses of the natural product pathways of a wide range of organisms, including plants. Matsuda and colleagues (Fazio *et al*., 2004) were able to produce triterpens in *Saccharomyces cerevisiae*, which was engineered to express several oxidosqualene cyclases, originally derived from *Arabidopsis thaliana.*

Heterologous expression is used in combination with comparative metabolic profiling: the wild type host is compared to the host containing the heterologous cluster. The presence of the heterologous metabolite can also provide good evidence that all the genes required for the biosynthesis, the regulation and the resistance have been cloned all together, as demonstrated for the congocidine gene cluster in *Streptomyces ambofaciens* (Juguet *et al*., 2009).

Escherichia coli can prove to be a suitable host, after adequate modifications, for heterologous expression due to its faster growth rate and the more efficient fermentation parameters, already well developed. The feasibility of engineering *E. coli* was demonstrated by the production of the complex macrolactone ring of erythromycin (Pfeifer *et al*, 2001). However, subsequently other hosts, equally easy to manipulate, come to the fore. *Streptomyces albus* demonstrated to be a good host for fredericamycin production (Wendt-Pienkowski *et al*., 2005), as well as *Streptomyces venezuelae* YJ003 for the production of kanamycin and spectinomycin (Thapa *et al.*, 2007; Thapa *et al*., 2008). To generate novel derivatives of the aminocoumarin novobiocin or clorobiocin, *S. coelicolor* M512 (lacking of actinorhodin, undecylprodigiosin and methylenomycin production) was used as heterologous host. The advantages came from the complete genome sequence which allows modifying rationally the primary metabolism supplies or the regulation network (Heide, 2009). A more efficient *S. coelicolor* host have been recently obtained, the strain M1154 is also deleted in the cluster *cpk* and *cda*, and it carries single mutation in the genes coding for the β-subunit of the RNA polymerase and the ribosomal protein S12 (Juan Pablo Gomez-Escribano, pers. comm.). However, no universal host strain is likely to be available; each case needs an optimized and specific host.

Despite its numerous advantages, a considerable drawback of heterologous expression is the difficulty of cloning large biosynthetic clusters (usually up to 40 kb) in a single vector, which is the case for most of the modular PKS and the NRPS gene clusters. However, recently Wenzel *et al.* (2005) elaborated a valuable strategy to express large biosynthetic gene clusters.

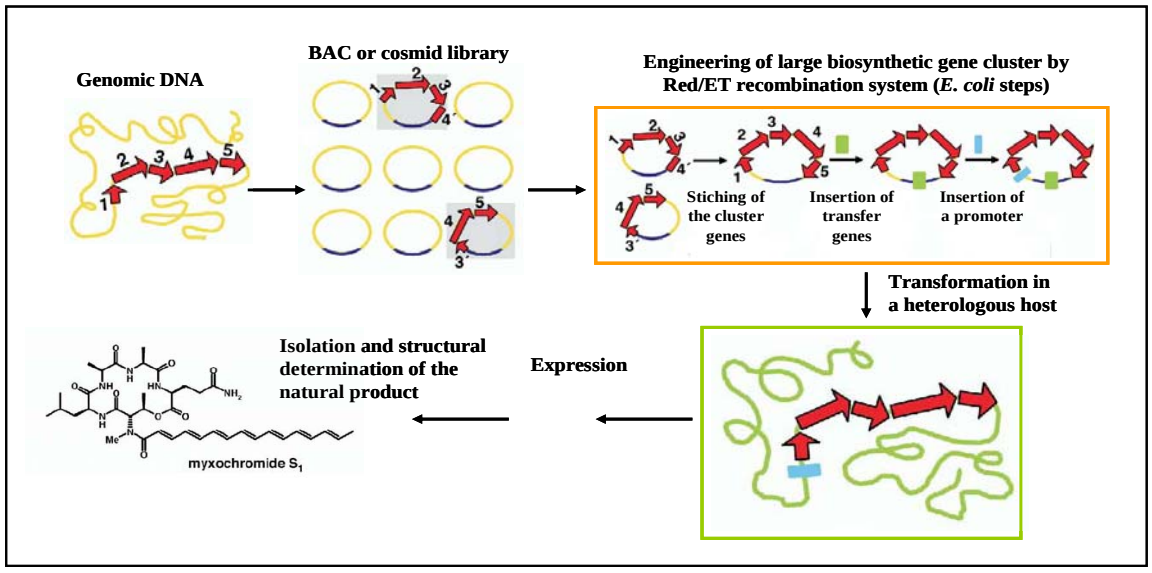


Fig. 8 General strategy for the cloning and heterologous expression of large biosynthetic cluster (adapted from Wenzel *et al.*, 2005). As example, the heterologous expression of myxochromide S is represented.

They were able to obtain a construct harbouring the entire myxochromide S gene cluster from *Stigmatella aurantiaca*, whose size is about 60 kb. Two cosmid clones, containing overlapping sequences of the cluster, were assembled together in *E. coli*, using Red/ET recombination system and the vector has been transferred into *Pseudomonas putida* for heterologous expression (see Fig 8). Nevertheless, the efficacy of this homologous recombination strategy needs to be still demonstrated for larger clusters.

3.1.5 Modification of the regulatory network

It is possible that none of the above strategies proves to be either effective or applicable and that the cluster still remains silent or poorly expressed. In this case, genetic manipulations of the regulatory network can be envisioned in order to activate the orphan cluster (Zerikly and Challis, 2009). Subsequently, comparative metabolic profiling, between the wild type and the mutant strains, can be used to detect the natural product. Potentially, this approach has the advantage to be faster and more rational, once a putative regulator has been identified. The manipulation can involve pleiotropic regulators or pathway-specific regulators, as demonstrated for example by Bergmann *et al.* (2007) who unravelled new PKS-NRPS hybrid products from *Aspergillus nidulans* enhancing the expression of a positive regulator within the cluster. Nevertheless, the efficiency of this method still needs to be confirmed in other microorganisms.
The deletion of the pleiotropic regulator DasR in *S. coelicolor* (paragraph 4.2 introduction), but also in other *streptomyces*, might be a potential alternative to trigger the activation of silent secondary metabolite clusters, which might be also achieved by addition of N-

acetylglucosamine to minimal media (Rigali *et al*., 2008). Another strategy to manipulate productively the regulatory network employs decoy oligonucleotides (McArthur and Bibb, 2008). These oligonucleotides can mimic binding sites for repressor transcriptional factors, increasing the production of the target secondary metabolite cluster. However, at this moment these approaches have not led to the discovery of novel natural products.

3.2 Metagenomics

Less than 1% of the microorganisms can be cultured in pure cultures by conventional methods (Schloss and Handelsman, 2005) and this limits tremendously our knowledge about microbial physiology, genetics and ecology. In particular, this data highlights the endless resources, all around us, which need to be explored. "Metagenomics" is the culture-independent analysis of a mixture of microbial genomes (termed the metagenome) from a given ecosystem using an approach based either on the expression (functional analysis) or on the sequencing (Schloss and Handelsman, 2005). DNA is extracted directly from environmental samples and it can be used to produce clone libraries, for sequencing or phenotype screening (Fig. 9). However nowadays, with the new 454 Sequencer method the creation of a DNA library is going to be unnecessary (Rothberg and Leamon, 2008).

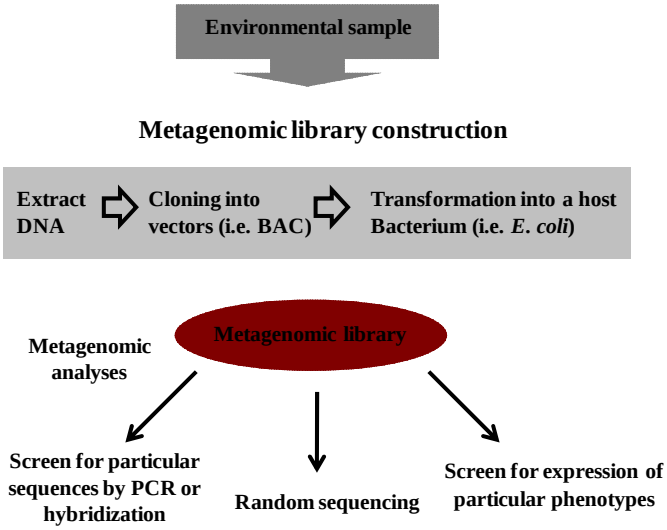


Fig. 9 Metagenomics approach involves the construction of a DNA library which can be subsequently analysed by expression or sequencing (modified from Riesenfeld *et al*., 2004).

Metagenomics approach has already proved to be very useful in the exploitation of marine and terrestrial reservoirs still unknown, which can be an amazing source of secondary metabolite products. Therefore, heterologous expression in a cultivable host becomes a necessary tool in

order to express the identified secondary metabolite clusters and eventually produce therapeutical drugs, derived from these unculturable and untapped strains. Examples of the efficacy of this approach came from the discovery of two antitumoral compounds, pederin from *Paederus* species rove beetles extracts (Piel, 2002) and bryostatin, from the marine protozoan *Bugula neritina* (Hildebrand *et al.,* 2004).

In the same time, metagenomics often contributes in enlarging our understanding of the molecular mechanisms beyond secondary metabolite biosynthesis through the pathways identified in uncultivable strains. For example, in the pederin cluster, isolated from a bacterial symbiont, a modular type I PKS was found lacking of its cognate AT domain, encoded instead by a discrete gene (Piel, 2002). The use of metagenomics is also very promising to search for regulatory genes involved in global or pathway-specific secondary metabolite regulation. A DNA library from an environmental sample can be transferred in a heterologous host and the production of secondary metabolites is analyzed. In this way, the *rep* gene,encoding for a transcriptional regulator of the ROK family, was identified as an activator of antibiotic production when transformed in *Streptomyces* species (Martinez *et al*., 2005) and others regulators might been soon identified.

3.3 Combinatorial biosynthesis

In the new genomic era, the production of non-natural products is no longer a pipe dream and it emerged as a potential powerful tool in generating novel diverse bioactive molecules. Nowadays it is possible to manipulate and combine biosynthetic machinery of different enzymatic pathways in order to produce hybrid products with specific chemical and pharmaceutical properties; this strategy is referred to as combinatorial biosynthesis (Wilkinson and Micklefield, 2007). A complete and comprehensive knowledge of modular PKS and NRPS molecular paradigms (see paragraph 5 of introduction) and the reprogramming know-how of these enzymatic machineries are the milestones of this method, but better insights on how enzymatic domains operate and interact need to be obtained. Engineering approaches have also been useful in the improvement of heterologous expression, optimizing for example the production of a key compound as a precursor (Gross *et al.,* 2006), or generating analogous of known effective drugs, by modifying some structural characteristic.

One of the first attempts for combinatorial biosynthesis of polyketides and genetic manipulation was carried out by Menzella *et al*. (2005), using as host strain *E. coli*. They were able to construct synthetic building blocks (enzymatic domains) flanked by unique sites for restriction enzymes which allowed to engineer and assemble any desired PKS gene, providing an efficient modular system. Expression of 154 bimodular combinations, flanked by the loading module and the thioesterase domain of erythromycin, gave rise to several different triketide lactone products. This was a pioneering attempt to demonstrate the efficacy and the great potential of

combinatorial biosynthesis which was followed by the research of Baltz and colleagues who generate a combinatorial library of novel lipopetides structurally related to daptomycin in *Streptomyces roseosporus*, the producing strain (Nguyen *et al*., 2006). One of derivative showed a stronger activity against *E. coli*. In the future, combinatorial biosynthesis will probably be a great and fascinating strategy to supply novel antimicrobial agents.

3.4 Screening of rare genera of actinomycetes and other untapped sources

In order to enlarge and increase the chance to discover novel secondary metabolites, in the last few years, industrial and academic attention was focused on improving the traditional screening approach aiming to exploit rare and untapped sources of microbial diversity.

Actinomycetes are responsible for the production of more than half of the discovered bioactive molecules (Berdy, 2005), especially those isolated from terrestrial environments. More than 10% of these secondary metabolites are isolated from the so called "rare" actinomycetes, which belong to the families of *Micromonosporaceae* (e.g. *Micromonospora* and *Actinoplanes*), *Pseudonocardiaceae* (e.g. *Amycolaptopsis* and *Saccharopolyspora*), *Thermomonosporaceae* (e.g. *Actinomadura*), *Nocardiaceae* (e.g. *Nocardia*) and *Streptosporangiaceae* (e.g. *Streptosporangium*) (Fig. 10). The term "rare" does not refer to their abundance in the environment, but mostly to their isolation frequency using conventional methods, which is much lower compared to the genus of *Streptomyces*. The importance of these strains is demonstrated by the discovery of many successful antibacterial agents derived from them, such as erythromycin from *Saccharopolyspora erythraea* or vancomycin from *Amycolatopsis orientalis* (Lazzarini *et al*., 2001).

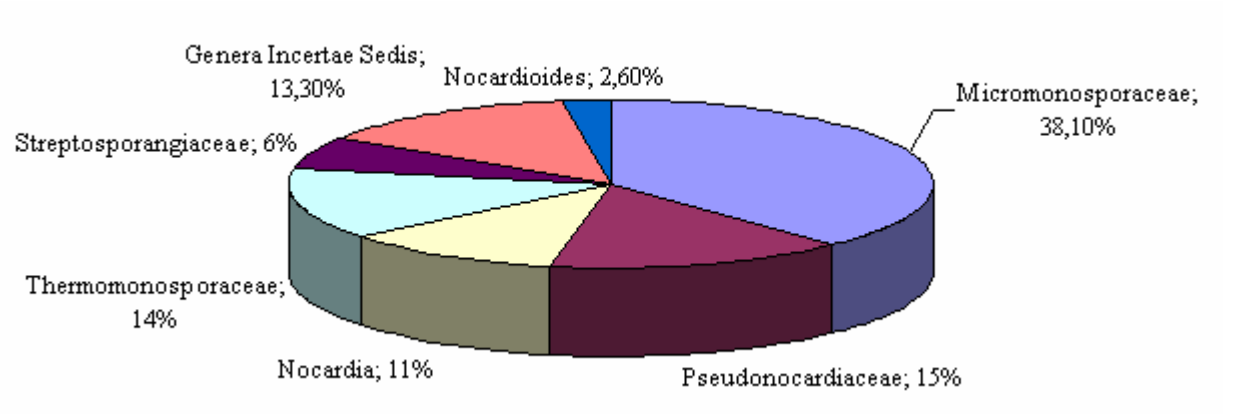


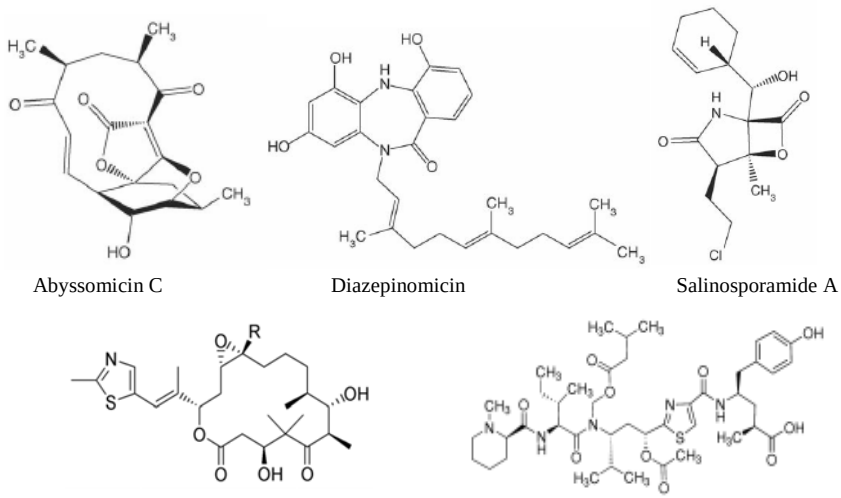
Fig. 10 Relative distribution of producing strains among rare actinomycetes (adapted from Lazzarini *et al*., 2001).

Culture-dependent approaches together with metagenomics brought to light the greatest biodiversity of the underexploited marine ecosystem, as well as the presence of indigenous

marine actinomycetes which would be a novel and untapped source for novel secondary metabolites. The first marine actinomycetes to be characterized was *Rhodococcus marinonascene* (Helmke and Weyland, 1984), followed later by the discovery of members of the genera *Streptomyces, Salinispora,* the first obligate new marine actinomycetes genus (Mincer *et al*., 2005), *Marinophilus*, *Salinibacterium* and others (Lam, 2006).

The exploitation of marine actinomycetes is only at the beginning, but it proved to be very promising. From 2003 to 2005, several metabolites were isolated showing interesting biological activities (Fig. 11A), in particular as anticancer agent, along with a broad range of chemical structures (Lam, 2006). Abyssomycin C is a polycyclic polyketide from a marine *Verrucosispora* strain which possesses potent antibacterial activity against Gram-positive and remarkably against multi-drug resistant pathogens (Riedlinger *et al*., 2004; Bister *et al*., 2004). Diazepinomicin (ECO-4601) from a *Micromonospora* strain and salinosporamide A (NPI-0052) from *Salinispora tropica* are two potential anticancer drugs, already going through clinical trials (Charan *et al*., 2004; Feling *et al*., 2003).

Another unexploited microbial source that came to the fore in the last few years is represented by the Gram-negative myxobacteria. To date these bacteria already yielded more than 100 natural structures and the interesting feature is that most of these compounds exhibit unique modes of action, against both prokaryotes and eukaryotes (Garcia *et al*., 2009). For example, epothilone and tubulysin, isolated respectively from *Sorangium cellulosum* and *Archangium geophyra*, act on the cytoskeleton of eukaryotic cells, in particular on the tubulin components (Fig. 11B).



Abyssomicin C Diazepinomicin Salinosporamide A

Epothilone Tubulysin

Fig. 11 A) Chemical structure of the marine secondary metabolites abyssomicin C, diazepinomicin and salinosporamide A. B) Chemical structure of the myxobacteria secondary metabolites epothilone and tubulysin.

4 INSIGHTS IN THE REGULATION OF SECONDARY METABOLISM

The onset of secondary metabolism coincides with a specific phase of the bacterial growth, more often when the nutrients are going to be exhausted (condition of "famine") and morphological differentiation takes place. The link between primary and secondary metabolism, together with the complex biosynthetic pathways of secondary metabolites, suggested a precise multilevel mechanism of regulation, which can vary from nutritional signals, including extracellular signals, to global or pathway-specific regulators. A better understanding of these mechanisms is essential in order to improve the production of already known natural drugs, but especially to facilitate the discovery of novel bioactive compounds, for example activating silent clusters (Bergmann *et al.*, 2007). In this thesis introduction, we will focus on the regulation of secondary metabolites in *Streptomyces*, the major producing genus. A particular attention will be given to the molecular mechanisms that have a significant effect on metabolite biosynthesis and that might be exploited in the future to awakening the numerous silent clusters so far identified in microorganisms.

4.1 Nutritional regulation

Nutrient sources play a very important role in the onset of secondary metabolite production, which proves once again that primary and secondary metabolism are tightly interconnected. In particular nutrient limitation seems to be a "signal" able to trigger antibiotic biosynthesis.

The carbon source is one of the most important medium components, being responsible for growth, energy driving force and secondary metabolism. The choice of carbon source can be fundamental to find the optimal conditions for secondary metabolite production and it changes according to the microorganisms. Glucose is not always the preferable choice, since it showed in some cases to prevent antibiotics formation, such as penicillin or erythromycin (Rokem *et al*., 2007). This phenomenon is known as carbon catabolite repression (ccr) and can involve other carbon sources, like glycerol, sucrose or fructose (Rokem *et al*., 2007). The mechanism is still not completely understood, but Langley and colleagues (Ramos *et al*., 2004; Guzman *et al*., 2005) demonstrated two independent systems for ccr in *Streptomyces peucetius*.

Likewise, nitrogen source and its concentration in the medium are crucial parameters for secondary metabolite production. Ammonium is the preferred nitrogen source for bacterial growth because of its easy assimilation, but similarly to glucose in most cases it represses antibiotics formation (Rokem *et al*., 2007). In *Streptomyces* sp., highly phosphorylated guanosine nucleotide (ppGpp) synthesis has been correlated with the onset of antibiotic production (Chakraburtty *et al*., 1996). ppGpp is synthesized from ATP and GTP by the ribosome-bound ppGpp synthetase RelA which is activated under nitrogen starvation, specifically amino acid depletion (Chakraburtty *et al*., 1996). In *S. coelicolor*, *relA* null mutants

showed to significantly reduce or prevent the expression of the pathway-specific regulators *actII-ORF4* and *redD*, respectively the transcriptional activators of the antibiotics actinorhodin and undecylprodigiosin (Chakraburtty and Bibb, 1997). Hesketh *et al*. (2001) demonstrated with a modified *relA* gene, independent of nitrogen limitation, that ppGpp is the direct activator of antibiotic production.

Phosphate concentration plays also an important role in the regulation of secondary metabolism. High level of inorganic phosphate exerts a negative control at the transcriptional and posttranscriptional levels on a wide and diverse range of secondary metabolite clusters, while phosphate-limiting condition triggers antibiotic production (Martin, 2004). This phenomenon is not only common to *Streptomyces* species, but it was also observed in Gram-negative bacteria, such as *Serratia* (Gristwood *et al*., 2009). The two component system PhoR-PhoP, well characterized in *S. coelicolor* and *S. lividans*, is involved in phosphate metabolism (Sola-Landa *et al.*, 2003). The sensor kinase PhoR, when activated by low concentration of phosphate, phosphorylates the response regulator PhoP, which in turn activates the expression of the *pho* regulon genes by binding to a consensus sequence, named PHO boxes (see Fig. 12). Recently, these boxes have been discovered to be formed by several direct repeated units (DRu) close to the -35 promoter region (Sola-Landa *et al*. 2008). In *S. lividans*, the deletion of *phoP* or both *phoR* and *phoP* genes induces higher production of actinorhodin and undecylprodigiosin, while failing to synthesize alkaline phosphatase (Sola-Landa *et al.,* 2003). However, no PHO boxes have been identified in promoter regions of the pathway-specific regulators of these antibiotics, suggesting a different mechanism of regulation (Martin, 2004). Very recently the group of Juan Martin showed that PhoP is able to bind to a sequence upstream *afsS*, thus being in competition with the global regulator AfsR. This latter is a SARP-like pleiotropic regulator (see paragraph 4.4), which regulates multiple antibiotic pathways in *S. coelicolor* (Floriano and Bibb, 1996). Its effect on antibiotic production is not direct, rather in its phosphorylated form AfsR promotes transcription of *afsS*, encoding a small protein which in turn activates *actII-ORF4* and *redD* through a yet unknown mechanism (see Fig. 12; Lee *et al*., 2002). Therefore, PhoP has a negative role in antibiotic biosynthesis, downregulating the transcription of *afsS*, while AfsR induces secondary metabolite production, activating *afsS* expression (Santos-Beneit *et al*., 2009). These last findings proved the interconnection between phosphate concentration and secondary metabolism, in a complex signalling transduction cascade (see Fig. 12).

Along with the idea that phosphate exerts a negative control on antibiotic production, Virolle and co-workers demonstrated that inactivation of the polyphosphate kinase gene *ppk* in *S. lividans* enhanced the expression of *actII-ORF4, redD* and *cdaR*, the latter being the transcriptional activator of the calcium-dependent antibiotic (Chouayekh and Virolle, 2002). Ppk is involved in the generation of polyphosphate from ATP, as well as regeneration of ATP from poly(P) and ADP. Poly(P) is used as energy storage and phosphate reservoir, its degradation release inorganic phosphate which caused the repression on secondary metabolism.

Even if the mechanisms beyond antibiotic repression due to the carbon, nitrogen or phosphate source remain still to be completely clarified, it emerged that these nutritional regulations are dependent on each other. In *S. coelicolor*, PHO boxes have been found in the promoter sequence of *glnR*, *glnA*, *glnII* and *amtB* genes, which are involved in nitrogen assimilation and transport, and the binding of PhoP prevents the transcription of these genes, indicating a phosphate control over nitrogen metabolism (see Fig. 12; Rodriguez-Garcia *et al*., 2009).

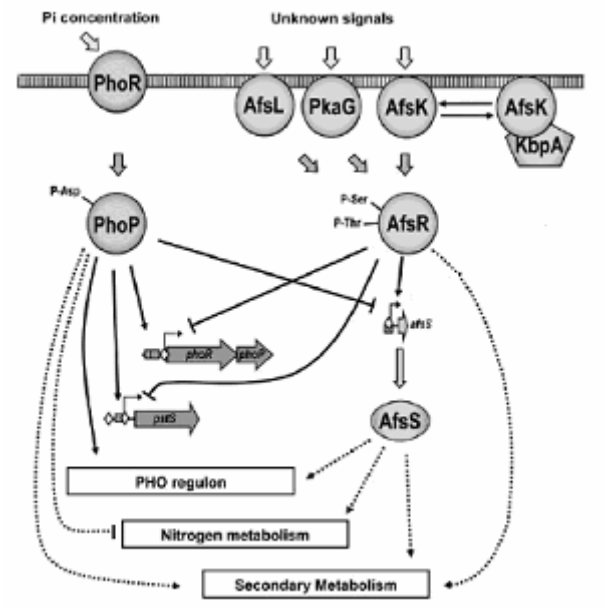


Fig. 12 Schematic model for the cross-regulation between the two component system PhoR-PhoP and AfsR (from Santos-Beneit *et al*., 2009). The sensor kinase PhoR in response of a condition of phosphate limitation undergoes self-phosphorylation. In turn, PhoR-P phosphorylates the response regulator PhoP which as dimer binds to the PHO boxes and controls gene expression. Besides the Pho regulon, PhoP has an effect on nitrogen metabolism and secondary metabolism, binding to the promoter regions of *afsS*, a small transcriptional factor whose transcriptional is positively regulate by AfsR. To bind to *afsS* promoter, AfsR needs to be phosphorylated by AfsK, or other serine-threonine kinases, like AfsL or PkaG. KbpA inhibits AfsK.

4.2 Global regulation

Several genes in the genome of *Streptomyces* species were identified and characterized as pleiotropic regulators, since their mutation affects multiple and different metabolic pathways, including those of secondary metabolism. The effect on antibiotic production seems to be dependent on certain growth and nutritional conditions (i.e., starvation); suggesting pleiotropic regulators play a role in sensing different environmental changes (Bibb and Hesketh, 2009). One example is given by the *dasR* regulon of *S. coelicolor*. DasR is a DNA binding protein which controls both morphological differentiation, through N-acetylglucosamine (GlcNAc) regulon, and secondary metabolism, through repression of *actII-ORF4* and *red* genes (Fig. 13; Rigali *et al*., 2006, Rigali *et al*., 2008). GlcNAc is a carbon and nitrogen source derived from the polymer chitin or the peptidoglycan, whose degradation coincides with morphological differentiation. It

has been reported for several *streptomyces* that GlcNAc, supplemented in rich media, blocks morphological differentiation and antibiotic production. On the contrary, when GlcNAc is added at high levels (at least 10mM) in minimal medium, it triggers antibiotic biosynthesis (Rigali *et al*., 2008). The transcriptional repression by DasR is in fact inhibited by the binding of glucosamine-6-phosphate, a degradation product of GlcNAc (see Fig. 13). The DasR protein and *dasR* regulon network is not only present in the genome of *S. coelicolor*, but is highly conserved in other *streptomyces* (van Wezel *et al*., 2009).

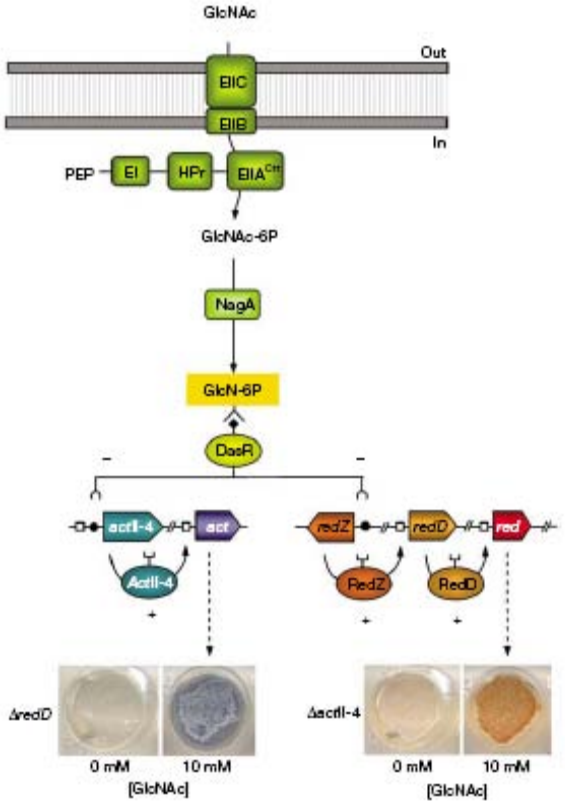


Fig. 13 Schematic model for the GlcNAc-dependent signalling pathway according to Rigali *et al*., 2008. The GlcNAc is internalized and phosphorylated (GlcNAc-6P) by a sugar phosphotransferase system (EI, HPr, EIIA, EIIB and EIIC), followed by deacetylation operated by NagA. The resulting glucosamin-6-phosphate (GlcN-6P) is able to bind to DasR, thus activating the expression of the pathway-specific regulators actII-ORF4 and RedZ. In the absence of GlcNAc in the medium, antibiotic production is prevented.

The *bld* genes, whose expression is phase-dependent, are also a good example of pleiotropic regulation, being involved both in morphological differentiation and secondary metabolism. In particular, the gene *bldA* in *S. coelicolor* encodes the only tRNA for the rare leucin codon UUA (Lawlor *et al*., 1987). This tRNA is not essential for bacterial growth, but is required for morphological development and secondary metabolite production. *bld*A mutants indeed failed to produce both aerial hyphae and antibiotics. The mechanism behind this repression is probably exerted at the translational level, since the mRNA transcript is not translated correctly because of the lack of the tRNA (Guthrie and Chater, 1990; Bruton *et al*., 1991). This phenomenon is not

exclusive to *S. coelicolor*, TTA codons were found in several pleiotropic and pathway-specific regulators for secondary metabolite clusters from streptomycetes. Chandra and Chater (2008) compiled a comprehensive table in which 143 secondary metabolite pathways from *Streptomyces* species contain a TTA codon.

A way to avoid *bldA*-dependence and probably improve antibiotic production is to change the TTA codon to an alternative leucin codon, as confirmed by Fernandez-Moreno *et al*. (1991) for actinorhodin biosynthesis or by O'Rourke *et al*., (2008) for methylenomicin biosynthesis. Otherwise, multiple copies of *actII-ORF4* gene or *redZ* have also shown to induce actinorhodin and undecylprodigiosin production, respectively, even in a *bldA* mutant (Passantino *et al*., 1991; White and Bibb, 1997). This is probably due to a mistranslation of the TTA codon by another tRNA specific for the amino acid leucine.

4.3 Extracellular signals

Bacterial cell-to-cell communication is mediated by small signalling molecules that are produced at a very low concentration to sense the extracellular changes or the physiological status and respond with a specific gene expression (Camilli and Bassler, 2006). These molecules have been intensively studied in proteobacteria, but in the 1960s the first signalling molecule, a γ-butyrolactone, was isolated in *Streptomyces griseus* by Khokhlov *et al*. (1967). To date, several γ-butyrolactones and other types of extracellular signals have been identified and they showed to be involved in the onset of secondary metabolite biosynthesis.

4.3.1 γ-butyrolactones

γ-butyrolactones are synthesised in response to a particular environmental or nutritional signal (Takano, 2006). They have been identified mostly in *Streptomyces*, where they act as positive regulators, but they are present in other actinobacteria, as demonstrated by Choi *et al*. (2003). So far, fourteen structures of γ-butyrolactones have been elucidated from seven *Streptomyces* species (see Fig 14). The chemical structure of γ-butyrolactones is similar to that of acylated homoserine lactone (AHL), the small signalling molecules of Gram-negative bacteria (Venturi, 2006), but they differ for the carbon side chain. AHL are considered as bacterial hormones and they are involved in the quorum sensing mechanisms, while γ-butyrolactones have been associated mainly with secondary metabolism, being involved in the onset of antibiotic biosynthesis.

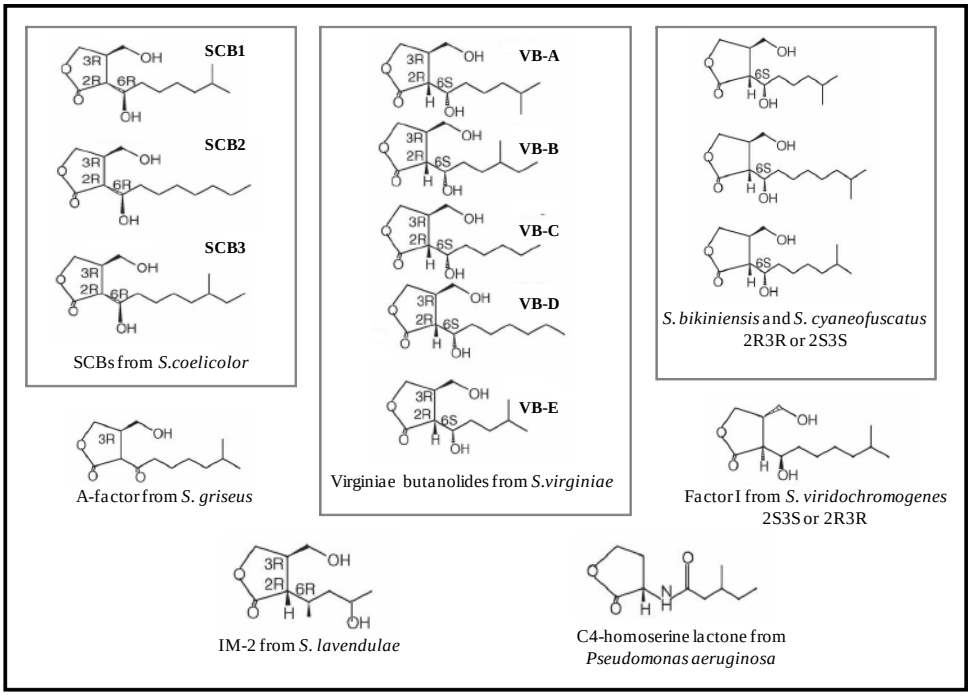


Fig. 14 Chemical structure of the fourteen γ-butyrolactones identified in seven *Streptomyces* species (adapted from Takano, 2006). SCBs are responsible for the production of actinorhodin and undecylprodigiosin; A-factor for streptomycin, grixazone and for sporulation; virginiae butanolides for virginiamycin; IM-2 for showdomycin and minimycin; factor I for anthracyclines and for sporulation in *S. griseus*; the γ-butyrolactones from *S. bikiniensis* and *cyaneofuscatus* for antracyclines in *S. griseus*. The AHL of *Pseudomonas aeruginosa* is also presented to compare its chemical structure to that of γ-butyrolactones.

A-factor is an exception since it is also involved in morphological differentiation (Horinouchi and Beppu, 1992), which underlines again the direct link between primary and secondary metabolism. These signalling molecules are active at nanomolar concentrations, the threshold necessary to bind to a specific cytoplasmic receptor protein. This γ-butyrolactones receptor acts as a repressor, preventing transcription of target genes, such as genes for secondary metabolite biosynthesis; the only exception is the receptor SpbR, which seems to have a positive role in pristinamycin production in *Streptomyces pristinaespiralis* (Folcher *et al*., 2001). The binding of the cognate γ-butyrolactones to the receptor produces a conformational change of the protein which is no longer able to bind to the DNA target, thus triggering gene expression.

A tremendous number of γ-butyrolactone receptor proteins has been identified in many streptomycetes, especially nearside secondary metabolite clusters, suggesting a role as pathway-specific regulators, which has been confirmed for some of them (Pang *et al*., 2004; Nakano *et al*., 1998; Bate *et al*., 1999; Takano *et al*., 2006), despite the fact that the corresponding γ-butyrolactone molecules are currently unidentified.

The first and best characterized γ-butyrolactone is the A-factor from *S. griseus* (Horinouchi and Beppu, 1992), which is required for both secondary metabolism (i.e. streptomycin and grixazone biosynthesis) and morphological differentiation (i.e. sporulation). The receptor of A-factor is

ArpA which represses the transcription of *adpA*, encoding a global regulator of *S. griseus* (see Fig. 15). Very recently, it was found that AfsA is a key enzyme in the biosynthesis of the A-factor (Kato *et al*., 2007).

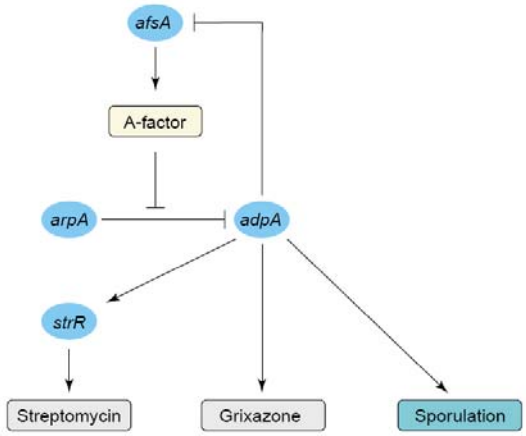


Fig. 15 The A-factor regulatory cascade in *S. griseus* (from Bibb, 2005). StrR is the transcriptional activator of streptomycin cluster and is one of the target of AdpA.

4.3.2 Furans

Challis and co-workers have recently discovered a novel class of antibiotic production inducers involved in the biosynthesis of methylenomycin in *S. coelicolor* (Corre *et al*., 2008). They elucidated the structure of five members of this family of 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCAs), named also Mm furans (Fig. 164). The structure resembles that of γ-butyrolactones, except for the lactone ring which is substituted by a furan ring. This structural change confers to furans the property of alkali-treatment resistance, absent in γ-butyrolactones, which allowed distinguishing these two types of inducers. Nevertheless, both signalling molecules share the same biosynthetic precursors, β-ketoacyl thioesters and dyhydroxyacetone phosphate, and the same biosynthetic enzymes (Corre *et al*., 2008). The biosynthesis diverges at the last step. Besides this, both molecules bind to the same family of receptor protein (i.e. TetR family), suggesting that other *Streptomyces* can produce this kind of inducers. An extracellular molecule was shown to bind to the TetR receptor protein AlpZ, triggering kinamycin (also known as alpomycin) biosynthesis in *Streptomyces ambofaciens*, even after alkali treatment (Bunet *et al*., 2008).

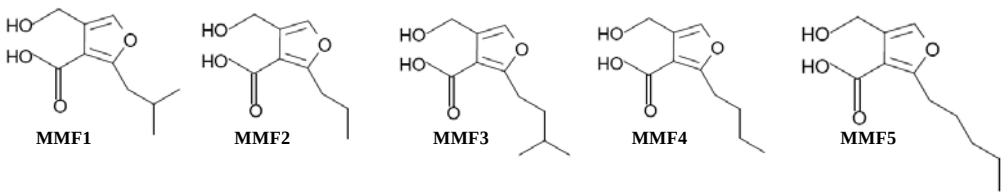


Fig. 16 Chemical structure of the five Mm furans isolated from *S. coelicolor*.

4.3.3 PI factor

A novel chemical class of inducers was isolated from *Streptomyces natalensis* and the first member has been named PI factor (Recio *et al*., 2004; Fig 17). This signalling molecule is active at nanomolar concentrations, as γ-butyrolactones, and is needed to induce pimaricin biosynthesis. A strain lacking PI factor was unable to produce pimaricin, which was restored only when the PI factor or A-factor from *S. griseus* were added, suggesting a certain flexibility toward signalling molecules (Recio *et al*., 2004).

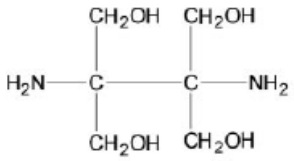


Fig. 17 PI factor [2,3-diamino-2,3-bis(hydroxymethyl)-1,4-butanediol] structure.

4.4 Pathway-specific regulators

As already quoted, secondary metabolite clusters contains genes coding for pathway-specific regulators, mostly transcriptional activators, which induce the expression of the biosynthetic genes in the cluster. Two protein families have been described in literature so far. However, γ-butyrolactone receptors need to be considered as pathway-specific regulators that negatively regulate metabolite production.

4.4.1 SARP regulators

The best characterized family of regulators in *Streptomyces* belongs to the SARP family (*Streptomyces* antibiotic regulatory proteins), described firstly by Wietzorrek and Bibb (1997). SARPs can be associated with different types of secondary metabolite clusters, such as PKSs or NRPSs, and they were also identified in other actinomycetes, as EmbR in *Mycobacterium avium* (Belanger *et al*., 1996). Members of this family are, for examples, ActII-ORF4, RedD and CdaR from *S. coelicolor*, which regulate the production of actinorhodin, undecylprodigiosin and the calcium-dependent antibiotic (Arias *et al*., 1999; Takano *et al*., 1992); DnrI from *S. peucetius* which regulates daunorubicin production (Sheldon *et al*., 2002); and CcaR from *S. clavuligerus* which controls cephamycin and clavulanic acid biosynthesis (Perez-Llarena *et al*., 1997).

SARPs are considered as essential transcriptional activators; inactivation of a SARP-like regulatory gene prevents antibiotic production. SARPs contain a winged helix-turn-helix OmpR-like domain at the N-terminus (Mizuno and Tanaka, 1997), to bind to a target DNA, followed by a bacterial transcriptional activation (BTA) domain and a diverse domain organisation at the C-terminus. The DNA targets of SARPs are direct heptameter repeats that can overlap the -35 recognition region of the target gene promoter (Wietzorrek and Bibb, 1997).

Among SARP family, some members are unique, since they present peculiar additional domains. EmbR contains a FHA (forkheaded associated) domain at the C-terminus which might be the target of a Ser/Thr kinase in a regulatory cascade mechanism (Alderwick *et al*., 2006). AfsR instead has a tetratricopeptide repeat (TRP) domain at the C-terminus, but the function is not clarified yet (Tanaka *et al*., 2007). The role of the C-terminus is not understood yet, but for AfsR it does not seem to be involved in the binding to DNA (Lee *et al*, 2002) or in initiation of transcription (Tanaka *et al*., 2007).

4.4.2 LAL regulators

Recently, a subfamily of regulatory proteins has been identified among the LuxR regulatory family (De Schrijver and De Mot, 1999) and has been named LAL for "Large ATP-binding regulators of the LuxR family". LAL regulators have a winged helix-turn-helix domain at the C-terminus, similar to the LuxR-like regulators; in addition they contain an ATP-binding domain, identified by Walker A and B domains (Walker *et al*., 1982), at the N-terminus. Their size can vary between 800 and 1160 aa, whilst LuxR-like regulators size is about 250 aa.

LAL regulators are one of the first kinds of pathway-specific regulator identified in modular type I PKS clusters. In particular, they have been associated to the synthesis of macrolides, such as pikromycin (Wilson *et al*., 2001), pimaricin (Anton *et al*., 2004), rapamycin (Kuscer *et al*., 2007); of polyene macrolides, such as candicidin (Chen *et al*., 2003) and nystatin (Sekurova *et* al., 2004) and more in general of other cyclic compounds.

The prototype member of this new family is MalT, an ATP-dependent transcriptional activator of the maltose regulon in *E. coli* (Richet and Raibaud, 1989). MalT needs the cAMP receptor protein to interact with the RNA polymerase and it is able to activate transcription only in the presence of ATP and maltotriose, suggesting that the mechanism of regulation of LAL proteins is rather complex. LAL proteins were identified first in proteobacteria (De Schrijver and De Mot, 1999), but they seem to be more widespread in actinomycetes, where they play a positive role in the regulation of secondary metabolite clusters. Deletion of the genes encoding this kind of regulator abolishes antibiotic production (Wilson *et al*., 2001; Anton *et al*., 2004). Unlike the SARPs for which a binding consensus site was identified and characterized, the DNA targets of LAL proteins are still unknown. Indirect evidences propose a direct regulation on the biosynthetic genes as well as for the resistance genes (Wilson *et al*., 2001; Kuscev *et al*., 2007; Anton *et al*., 2004).

Table 4 lists all the LAL proteins described in literature so far from *Streptomyces* species. The percentage of homology is not very high among LAL proteins, only the N and C-termini appeared to be well conserved. Probably the middle part of the gene is involved in a protein-protein interaction mechanism which might modulate the activation and the binding of the regulator to the DNA target. Indeed, near the C-terminus of AveR, GdmRI and GdmRII a tetratricopeptide repeat motif was identified (He *et al*., 2008; Kitani *et al*., 2009) which is

supposed to be a module to facilitate protein-protein interaction (Marck *et al*., 1993) and which is more likely present in all the LAL members.

Species	Protein	Function	Reference
S. venezuelae	PikD*	Pikromycin biosynthesis	Wilson *et al*. 2001
S. hygroscopicus	RapH*	Rapamycin biosynthesis	Kuscer *et al*. 2007
S. hygroscopicus	Orf6	Type I PKS cluster	Ruan *et al*. 1997
S. noursei	NysRI, NysRII, NysRIII*	Nystatin biosynthesis	Sekurova *et al*. 2004
S. cinnamonensis	MonH	Monensis biosynthesis	Oliynyk *et al*. 2003
S. avermitilis	AveR*	Avermectin biosynthesis	Kitani *et al*. 2009
S. sp. CK4412	TmcN*	Tautomycetin biosynthesis	Hur *et al*. 2008
S. albus	SalRI, SalRII	Salinomycin biosynthesis	Knirschova *et al*. 2007
S. natalensis	PimR*	Pimaricin biosynthesis	Anton *et al*. 2004
S. hygroscopicus 17997	GdmRI, GdmRII*	Geldanamycin biosynthesis	He *et al.* 2008
S. hygroscopicus var. *ascomyceticus*	FkbN	FK520 biosynthesis	Wu *et al*. 2000
S. sp. FR-008	FscRII, FscRIII, FscRIV	FR-008/candicidin biosynthesis	Chen *et al*. 2003
S. nodosus	AmphRI, AmphRII, AmphRIII	Amphotericin biosynthesis	Carmody *et al*. 2004
S. platensis Mer-11107	PldR	Pladienolide biosynthesis	Machida *et al*., 2008
S. ambofaciens	SAMR0484	Sambomycin biosynthesis	This thesis

Table 4 List of the LAL proteins from *Streptomyces* sp. found in literature involved in secondary metabolite biosynthesis. In the table, the regulatory protein SAMR0484 identified in the type I PKS cluster of *S. ambofaciens* and discussed later in this thesis is also listed. *LAL regulators that have been characterized experimentally.

5. POLYKETIDE SYNTHASES VERSUS NON-RIBOSOMAL PEPTIDE SYNTHASES

The great majority of natural products belongs to the class of polyketides (PK) and non-ribosomal peptides (NPR) or hybrids PK-NRP, which are synthesized by complex multienzymatic machineries, respectively named polyketide synthases (PKSs) and non-ribosomal peptide synthases (NRPSs) (Fischbach and Walsh, 2006). PKS and NRPS gene clusters are characteristic of microorganisms and fungi, but they were also identified in plants. Examples of these products are erythromycin, tetracyclines, vancomycin and penicillins, which found a therapeutical application (see Fig. 3). PKSs and NRPSs are capable of producing a complex and diverse range of chemical structures due to the combination of different building blocks and due to the presence of modifying tailoring enzymes. Therefore these biosynthetic machineries have become a fascinating and valuable (for the economical repercussions) subject of research.

Genes encoding for PKS or NRPS are found in secondary metabolite clusters and they are characterized as polyfunctional megasynthetases (200-2000 kDa), suggesting an organized and coordinated enzymatic assembly line (Cane and Walsh, 1999). The majority of the knowledge of PKS paradigm comes from the modular type (see next paragraph), but it can be applied to the other types too. To date, only a type of NRPS is known, which might present sometimes modifications to the principle paradigm.

5.1 The core domains

A modular PKS and the NRPSs are organized in several multifunctional modules, responsible for the chain elongation of the polyketide or the polypeptide. Each module contains at least a set of three enzymatic domains, "the core domains", which are involved in the activation, the transfer and the condensation of specific precursors, the acyl-CoA thioesters (see Fig. 18 and 19). In the core domains of modular PKSs, an acyltransferase (AT) catalyzes the transfer of a selected activated substrate to an acyl carrier protein (ACP) domain and the β-ketoacyl synthase (KS) promotes the condensation of two substrates, by loss of carbon dioxide (Fig. 18; Cane and Walsh, 1999). The domains are active as dimers, forming a sort of channel for the polyketide chain elongation. The order in which these domains are found in a module is KS-AT-ACP. The mechanism of polyketide biosynthesis has strong homologies to that of fatty acid biosynthesis (FAS), both share the same precursor supply and the same type of catalytic domains (Staunton and Weissman, 2001; Cronan and Thomas, 2009); and in either case the final product released is a linear β-keto ester.

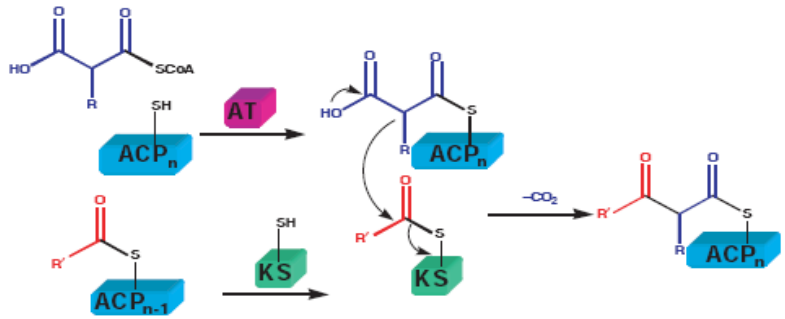


Fig. 18 Enzymatic mechanism of the core domains of a modular type PKS (from Cane and Walsh, 1999). The AT domain loads an acyl-CoA on the thiol residue of ACP domain, generating an acyl-S-enzyme intermediates. The KS domain is transiently acylated on the active-site cysteine by an acyl-CoA starter or chain elongation unit. The KS catalyzes the decarboxylative acylation which produces a C-C bond, via Claisen condensation (Walsh, 2003; Cane and Walsh, 1999).

In functional analogy to modular PKSs, the core domains of NRPSs are represented by a peptidyl carrier protein (PCP), tethering the precursors, by an adenylation (A) domain, which selects, activates and transfers the amino acid monomer to PCP, and by a condensation (C)

domain that has the same function of the KS domain in the PKSs (see Fig. 19). The order of these domains in a module is C-A-PCP.

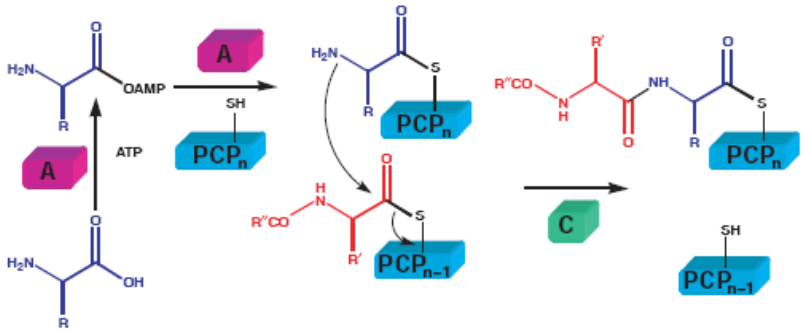


Fig. 19 Enzymatic mechanism of the core domains of a NRPS (from Cane and Walsh, 1999). The A domain activates the monomer in an aminoacyl-AMP and transfers it to the phosphopantetheinyl arm of PCP domain. The C domain catalyzes the condensation reaction between a nucleophilic amino group and an electrophilic carbonyl group both loaded into a PCP domain, giving rise to a C-N bond as an amide link.

Initially, the ACP and PCP domains are in the form of apo enzymes. They are converted in holo enzymes only after a posttranslational modification operated by a phosphopantetheinyl transferase (PPTase) which transfers a 20-Å-long phosphopantetheine prosthetic group from coenzyme A to a conserved serine in the ACP (or PCP) domain (Lambalot *et al*., 1996).
The building blocks of PKSs are made up of small organic acids, such as acetate or propionate, originated from fatty acid biosynthesis, while the 20 common amino acids and a variety of nonproteinogenic amino acids are the building blocks for NRPSs (Fischbach and Walsh, 2006). The precursors need to be activated before being recruited by AT (or A) domain and then covalently bound to the flexible phosphopantetheinyl arm of ACP (or PCP) domain. The activated PKS substrates are acyl-CoA thioesters and are formed by an acyl-CoA synthetase through a two-step reaction, which involves ATP and co-enzyme A and which is the same needed for the activation of fatty acid precursors. The NRPS monomers are aminoacyl-AMP mixed anhydrides, activated by ATP and the adenylation domain (Fischbach and Walsh, 2006).

5.2 The auxiliary domains

A part from the core domains, PKSs and NRPSs modules can contain other enzymatic domains, involved in post-modifications of the β-keto group or of the amino acid chain, respectively. These auxiliary domains and all their possible combinations are responsible for the tremendous chemical structural variety of polyketide and polypeptide products.
Modular PKS and FAS also share the same set of additional domains: the ketoreductase (KR) domain uses NADPH to reduce the β-keto group to hydroxyl residue; the dehydratase (DH) domain dehydrates the hydroxyl to a double bound and the enoylreductase (ER) domain can

reduce the double bond to a saturated carbon (Walsh, 2003). When present all together, the three domains operate in the sequence KR→DH→ER (see Fig. 20A), but a module can contain only an additional KR domain or a KR+DH domain.

NRPS modules might additionally contain an epimerization (E) domain, present when an L-amino acid needs to be incorporated and epimerized (D-amino acids are indeed very rare in microbial cells); in this case the E domain is located downstream of the PCP domain (Fischbach and Walsh, 2006). A cyclization (Cy) domain catalyzes the formation of a thiazoline or oxazoline ring, when a cysteine, a serine or a threonine residue are present in the chain elongation. It is possible to find a tandem Cy-Cy domain, which gives rise to a bithiazole moiety, as shown by bleomycin structure (Du *et al*., 2000). The N-methyltransferase (NMet) domain transfers a methyl group, from S-adenosylmethionine, to the amino group of an aminoacyl intermediate (see Fig. 20B).

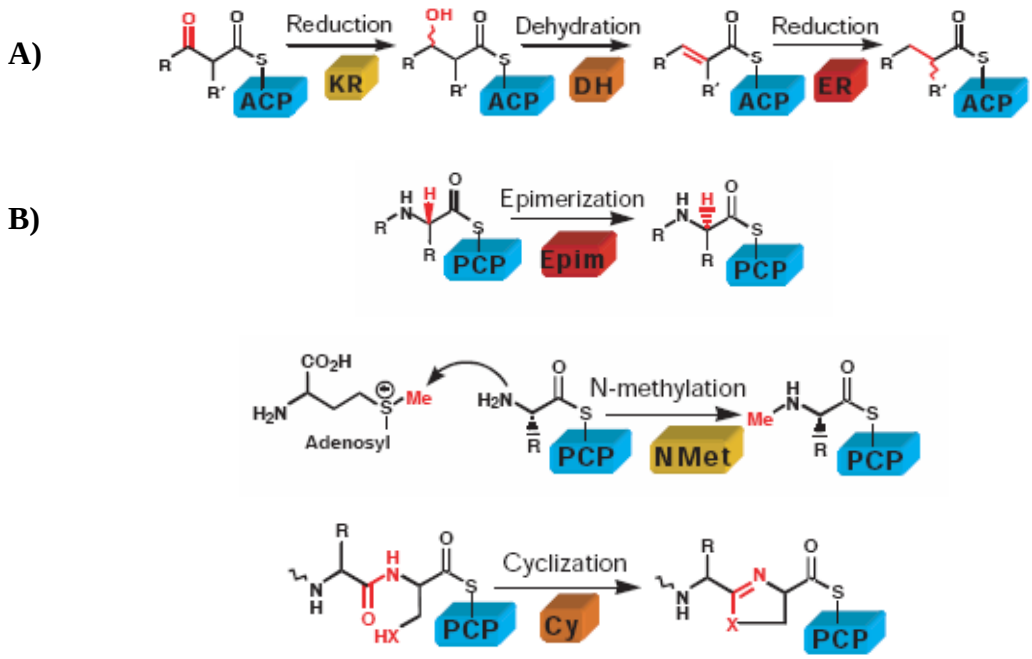


Fig. 20 Additional set of enzymatic domains in PKSs (a) and in NRPSs (b). All these modification reactions are carried out before loading the growing polyketide or polypeptide chain to the following module (from Cane and Walsh, 1999). Heterocyclization can occur when a serine, a threonine or a cysteine residue is present (X=O or S).

5.3 The initiation and termination domains

The core domains and the auxiliary domains are mostly involved in the elongation steps of the natural product biosynthesis. Therefore, the enzymatic assembly of PKSs and NRPSs is usually flanked by an initiation and a termination module. In the initiation module of a modular PKS (also named loading module), an AT and ACP domain are always present (Fig. 21A), as

demonstrated for avermectin cluster in *S. avermitilis* (Marsden *et al*., 1998) or for erythromycin cluster in *S. erythraea* (Aparicio *et al*., 1994). However, it can occur that a KS-like domain, catalytically inactive, is also part of the module. This domain has been named KSQ because it contains a glutamine instead of a cysteine in the active site and therefore is no longer able to catalyse carbon condensation (Bisang *et al*., 1999). KSQ domain acts as a (methyl)malonyl-CoA decarboxylase, producing decarboxylation of CoA-esters of dicarboxylic acids, as demonstrated for monensin loading module (Bisang *et al*., 1999). The decarboxylation takes place *in situ* to produce a propionyl or an acetyl-ACP intermediate, as shown in Fig. 21B.
Typically, the initiation module of NRPSs is an A+PCP didomain; a C domain is present when N-acylation of the first amino acid is required (Fischbach and Walsh, 2006).

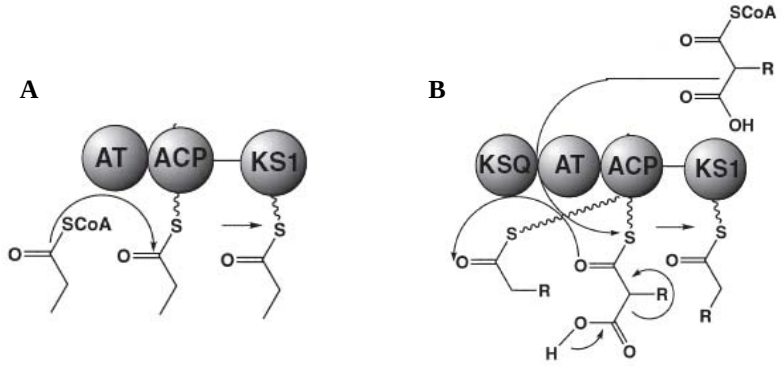


Fig. 21 Two different models for PKS initiation module (from Long *et al*., 2002). Tethering of an acyl-CoA starter unit (a propionate) to the ACP domain in a didomain loading module (**A**). Acylation of a (methyl)malonyl-CoA to the ACP of a KSQ loading domain, followed by decarboxylation *in situ* (**B**). R= H or CH_3.

Common to polyketides and polypeptides, the last module generally contains a thioesterase (TE) domain which is responsible for the release of the linear backbone of the molecule by hydrolysis. This reaction can generate a free acid, as for β-lactam biosynthesis, or the TE can promote the intramolecular cyclization of the molecule to give rise to a lactone ring, as for macrolide biosynthesis (Fischbach and Walsh, 2006).

5.3.1 Type II thioesterase

Several PKS biosynthetic clusters, such as pikromycin (Xue *et al*., 1998), tylosin (Merson-Davies and Cundliffe, 1994) or spiramycin (Karray *et al*., 2007) contain additional discrete thioesterase enzymes, named type II TE (TEII) to distinguish them from the TE present in the terminal module of PKSs. TEII were also found to be associated with NRPS clusters (Schneider and Marahiel, 1998).
The proposed role for TEII is to edit through hydrolysis aberrant groups attached to the ACP domain, resolving the problem of a stalled chain elongation, due, for example, to a substrate misincorporation (Fischbach and Walsh, 2006). Gene inactivation produced a significant

decrease in polyketide biosynthesis, showing that even if not essential, TEII plays an important role which might be useful to improve the yields of production (Kim *et al*., 2002).

6. POLYKETIDE SYNTHASES

6.1 Typology of polyketide synthases

Based on the catalytic domain organization, polyketide products can be originated by four different types of PKSs: modular type I and iterative type I, type II and type III.

6.1.1 Modular type I PKS

As described in the previous chapter, a modular type I PKS is a polyfunctional megasynthetase, composed of several modules responsible for a single step of the polyketide elongation (Staunton and Weissman, 2001). Each elongation module contains the core domains and often additional domains; a loading module and a TE domain are also generally present. The modules interact to each other head-to-tail by specific linkers (also named "docking domain"), thus determining the order of action of the polypeptides (Staunton and Weissman, 2001).

The first modular type I PKS gene cluster to be identified and characterized was that for erythromycin biosynthesis from *S. erythraea* (Cortes *et al*., 1990; Caffrey *et al*., 1992; Donadio and Katz, 1992). Erythromycin cluster contains three biosynthetic genes (*eryAI-III*), which in turn are divided in six modules plus the loading module (see Fig. 22). The starter unit is a propionate and six malonyl-CoA substrates are the extender units used to synthesize the 6-deoxyerythronolide B (6-dEB), the 14-membered macrolactone backbone of erythromycin.

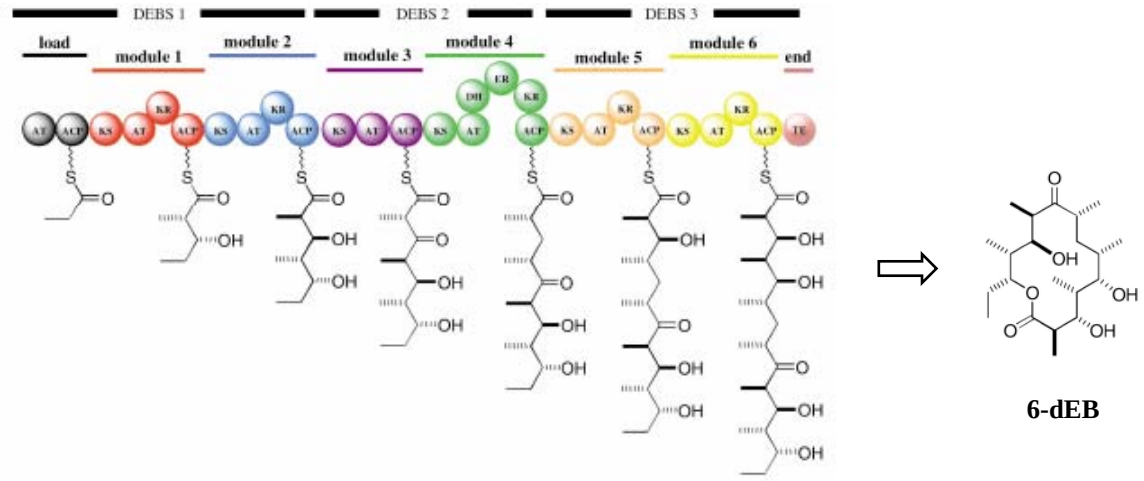


Fig. 22 Module and domain organization of deoxyerythronolide B synthase (DEBS) which is responsible for the biosynthesis of the macrolactone 6-dEB (from Staunton and Weissman, 2001). DEBS1, 2 and 3 are respectively coded by the gene *eryAI-III*. The initiation domain is a loading didomain.

An exception to the modular type I paradigm came from the discovery of leinamycin PKS genes from *Streptomyces atroolivaceus* lacking the cognate AT domain (Cheng *et al*., 2003). Leinamycin is synthesized by a hybrid NRPS-PKS cluster which contains a discrete AT protein encoded by *lnmG*. LnmG works iteratively *in trans* to load malonyl-CoA extender units to the different ACP domains of the PKS modules. Similarly, Piel (2002) reported that the PKS genes of pederin cluster lack the AT domain and proposed that both *pedC* and *pedD* encode a discrete AT domain.

6.1.2 Iterative type I PKS

Iterative type I PKSs have been discovered by Bechtold and co-workers who identified the gene *aviM* from *Streptomyces viridochromogens* as a type I PKS biosynthetic gene, which interestingly acted in an iterative way to produce the orsellinic acid moiety of avilamycin (Gaisser *et al*., 1997). Other examples of this new subgroup of PKSs were later found in *Micromonospora echinospora* (Ahlert *et al*., 2002), in which CalO5 also synthesizes the orsellinic moiety of calicheamicin, and in *Streptomyces carzinostaticus* subsp. *neocarzinostaticus*, where NcsB catalyses the biosynthesis of the naphthalinic acid moiety of neocarzinostatin (Shen, 2003). So far, all the products synthesized by these iterative type I PKSs are mono or bicyclic aromatic compounds.

As the modular type I, the iterative type is structural and functional related to fatty acid biosynthesis. A unique module contains the typical core domains (KS+AT+ACP) and other additional domains can also be present, as demonstrated for AviM, CalO5 and NcsB (see Fig. 23). The lack of an initiation and a termination domains, as well as the iteratively action of the module, are the differences with the modular type I PKS.

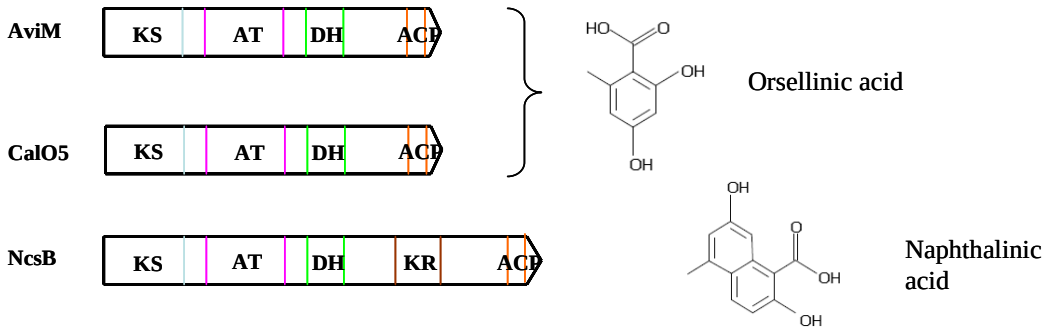


Fig. 23 Module organization of iterative type I PKS from avilamycin (AviM), calicheamicin (CalO5) and neocarzinostatin (NcsB) biosynthetic clusters (adapted from Shen, 2003).

6.1.3 Type II PKS

Actinomycetes are the only producer strains using type II PKS systems to produce a wide range of bioactive aromatic polyketide compounds, which in some cases have found an application in human therapy as anticancer or antibacterial drugs (Hertweck *et al*., 2007). Members of this

group are anthracyclines (e.g. doxorubicin), tetracyclines, angucyclines (e.g. landomycin A) or aureolic acids (e.g. mithramycin).
In contrast to modular type I PKS, types II are composed of distinct genes encoding for different enzymes which act iteratively to generate a polyketide backbone (see Fig. 24). The minimal set of PKS genes, generally clustered together, includes a KS domain (KS_α), containing an active-site cysteine, a KS domain (KS_β), catalytically inactive, and an ACP domain (Zhang and Tang, 2009). All together these enzymes catalyse the condensation of an acyl-CoA starter unit and of several malonyl extender units, the only type of precursor used by the type II PKS.

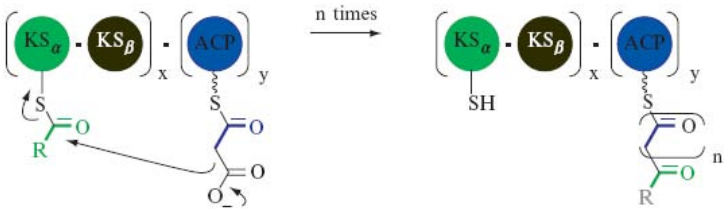


Fig. 24 Type II PKS mechanism: discrete catalytic enzymes are responsible for the chain elongation (from Weissman, 2009). KS_α and KS_β domains form a heterodimer, in analogy to the modular KS homodimer. The ACP domain is used as anchor for the growing chain (Hertweck *et al*., 2007).

The KS_β has a glutamine residue instead of the cysteine in the active site, as already observed for the KSQ domain in the modular type I PKS (Bisang *et al*., 1999), which implies that this domain is probably involved in the decarboxylation of the malonyl extender units. Another function assigned to is the control of the carbon chain elongation, for this reason the domain is also named CLF for chain length factor. However, KS_β is not the only determinant, the entire PKS complex seems to be involved in the length of the polyketide, in particular cyclases were shown to influence the number of chain elongation (Hertweck *et al*., 2007).
Additional monofunctional domains, which include ketoreductases, cyclases and aromatases, were identified along with the minimal set of biosynthetic genes, in order to modify the nascent polyketide chain (Hertweck *et al*., 2007) and generate aromatic compounds.

6.1.4 Type III PKS

Type III PKSs belong to the family of chalcone and stilbene synthases (CHS and STS), which previously were found only in plants where they play a role in flavonoid biosynthesis (Moore and Hopke, 2001). This kind of PKSs is phylogenetically unrelated to the other types of bacterial PKSs; probably the genes were acquired through horizontal transfer between plant and symbiotic bacteria (Moore and Hopke, 2001). Unlike type I and II PKSs, type III consists of a single KS-like domain, which catalyzes repetitive condensations of free CoA thioesters (see Fig. 25) and it is independent from an ACP domain. The conserved active-site cysteine of KS domain is involved in all the enzymatic reactions (Moore and Hopke, 2001). The chain

extension is often followed by cyclization or decarboxylation, yielding cyclic aromatic compounds. The mechanism, by which the polyketide chain length is controlled, is not yet fully understood.

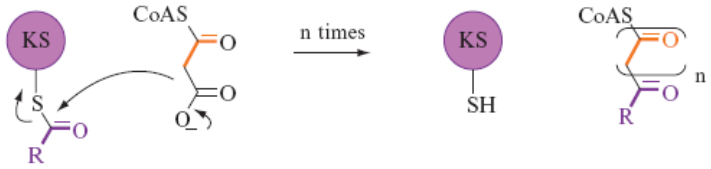


Fig. 25 Bacterial type III PKS mechanism: a single multifunctional KS domain, in form of homodimer, is responsible for the polyketide biosynthesis (from Weissman, 2009). The thiol residue of KS domain recruits CoA thioesters and catalyzes the chain extension by a Claisen condensation.

The first bacterial type III was identified and characterized in *S. griseus* (Funa *et al*., 1999): *rrpA* codes for a CHS-like protein which is responsible for the biosynthesis of a tetrahydroxynaphtalene (THN). RrpA condensates five malonyl-CoA extender units to generate a pentaketide which then cyclizes in a THN. This end product is probably instable and undergoes oxidation to form flaviolin, which in turn polymerizes to produce several coloured compounds (Funa *et al*., 1999).

Subsequently, several type III PKSs were identified in other microorganisms, such as *Pseudomonas fluorescens* or *S. coelicolor* (Bangera and Thomashow, 1999; Song *et al*., 2006), suggesting that this enzymatic system is not only widespread in plants.

6.2 Polyketide-tailoring genes

After being released by the TE, the polyketides can undergo several post-modifications that yield an incredible variety of bioactive compounds. These “tailoring” genes are generally clustered together with the biosynthetic genes and are probably submitted to the same regulatory mechanism (Staunton and Weissman, 2001).

6.2.1 Glycosyltransferases

Many secondary metabolite structures display a sugar moiety or even an oligosaccharide chain, like landomycin A, which contains up to six sugars (Zhu *et al*., 2007). The glycosylation is responsible for the biological properties of the molecules; indeed the sugar residue seems to interact directly with the cellular target, for example the ribosome subunits for the macrolides (Walsh, 2003). The type and the numbers of sugars attached to the aglycone significantly affect the biological activity and therefore glycosylation can be a powerful tool in engineering novel bioactive natural products.

Glycosyltransferases (GTs) are the enzymes that catalyze the transfer of an activated sugar donor to an acceptor molecule, in this case a polyketide backbone. GTs are not only involved in

secondary metabolite biosynthesis, but also in the biosynthesis of glycopeptides, glycolipids and polysaccharides (Härle and Bechthold, 2009). According to the CAZy (Carbohydrates-active enzymes) classification (Campbell *et al*., 1997; Coutinho *et al*., 2003), the majority of GTs from actinomycetes and associated with secondary metabolite clusters belongs to the GT-1 family, in turn a subgroup of the GT-B superfamily. Usually deoxysugars, in particular 6-deoxyhexoses, are the substrates used by GTs, which can attach them to an OH-group, or alternatively to a NH_2 or CH-group. A GT is generally specific for a unique donor and acceptor substrates; nevertheless several GTs proved to have a flexible specificity towards both the sugar donor and the aglycones (Zhang *et al.*, 2006), highlighting the potential to create new unnatural products by combinatorial biosynthesis.

The sugar moieties originate from α-D-glucose-1-phosphate, derived from primary metabolism. The deoxysugar needs to be activated by a nucleotidylyltransferases, in most of the cases a thymidylyltransferase which generates TDP-glucose (Fig. 26). Before being attached by the GT to the aglycone acceptor, the TDP-glucose can undergo several modifications, through methyltransferases, oxygenases, ketoreductases or amino transferases, to generate a wide range of unusual sugar residues, as desosamine, mycarose, rhamanose, mycaminose (White-Phillip *et al*., 2009). The genes involved in the activation and in the modification of the donor sugar are clustered together, as mentioned in the definition of secondary metabolite cluster.

In certain cases, glycosylation is associated with a mechanism of resistance, whereby the glycosylated product is inactive (Vilches *et al*., 1992; Gourmelen *et al*., 1998).

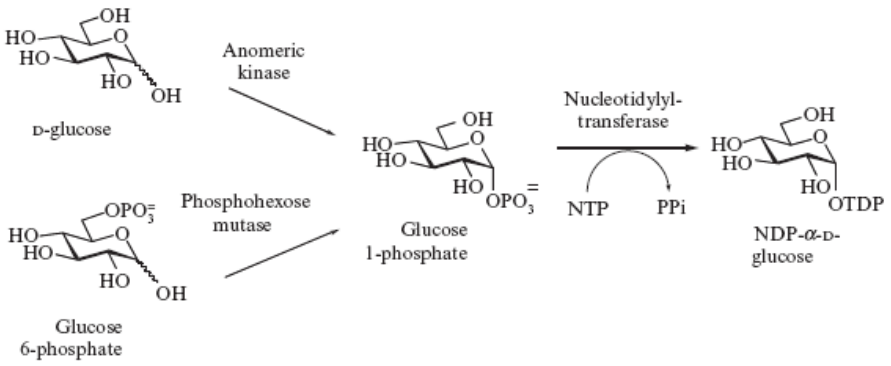


Fig. 26 Activation by a nucleotidylyltransferase of α-D-glucose-1-phosphate derived from phosphorylation of α-D-glucose or from epimerisation of glucose-6-phosphate (adapted from White-Phillip *et al*., 2009).

6.2.2 Methyltransferases and oxygenases

Methylation systems are widespread among all types of PKSs and NRPSs. Methyltransferases use as cofactor *S*-adenosyl-L-methionine (SAM) in order to transfer an activated methyl group on the carbon, nitrogen or oxygen atom of the polyketide backbone (Walsh, 2003). The recent crystal structure of DnrK (Jansson *et al*., 2004), a methyltransferase identified in the

daunorubicin cluster of *S. peuceticus*, clearly confirmed that DnrK in form of homodimer binds the SAM cofactor and the substrate at the same time. O-methylation of deoxysugars ring is also a common mechanism observed in nature and it proved to improve the pharmacokinetics properties of the molecules. Li *et al*. (2009) demonstrated that MycE and MycF methylate respectively deoxyallose and javose, producing mycinamicin, an antibacterial macrolide from *Micromonospora griseorubida*.

Secondary metabolite clusters can also contain genes coding for oxygenases, which are involved in oxidative reactions. Oxygenases are divided in mono and di-oxygenases, according to the number of oxygen atoms incorporated in the substrate (Hertweck *et al*., 2007). These enzymes can catalyze hydrozylation, epoxidation, quinone formation or oxidative rearrangement reactions, using different type of mechanisms and cofactors. The cytochrome P450 monooxygenases are the most common enzymes found in type I PKS clusters, while anthrone and flavin-dependent oxygenases are commonly used for aromatic polyketide biosynthesis (Hertweck *et al*., 2007).

Similarly to glycosyltransferases, a better knowledge of these tailoring genes could be useful in the production of more powerful and efficient bioactive compounds.

7. *STREPTOMYCES*, A PROLIFIC PRODUCING GENUS

Actinomycetes, fungi and some other bacteria are the major responsible for the production of microbial secondary metabolites; according to Berdy (2005), more than 20000 molecules have been discovered. In particular, 7600 natural products have been isolated from *Streptomyces* species, thus being responsible of about 38% of all secondary metabolites and representing the most prolific genus of microbial producers. Therefore, *Streptomyces* have been and they continued to be the most exploited genus in the academic and industrial field.

7.1 General characteristics

The genus *Streptomyces* encompasses aerobic sporulating Gram-positive soil-dwelling bacteria, belonging to the order of *Actinomycetales*. They are ubiquitous saprophytic non-motile multicellular organisms that can colonize different ecosystems, either marine or more often terrestrial, thus they are constantly in competition with other bacteria and fungi (Goodfellow *et al*., 1983). *Streptomyces* have a complex lifecycle characterized by a filamentous vegetative growth, resembling that of fungi (Fig. 27). The vegetative mycelium extends in branching aerial hyphae which can undergo synchronous fragmentation and differentiate into unigenomique spores, a semi-dormant phase which allow resisting to starvation or drought, but also allow

multiplying and spreading out. In the presence of nutrient signals, single spores germinate, restarting a new cycle (Kieser *et al*., 2000).

Morphological differentiation coincides with secondary metabolite production (see Fig. 27); indeed both mechanisms are triggered by a nutrient limiting condition. From an industrial and economical point of view, *Streptomyces* are very interesting organisms and they are greatly exploited as secondary metabolites producers. Along with antibiotics and other therapeutical compounds, they also secrete a wide range of enzymes (e.g., hydrolases, isomerases) which found an application as biocatalysts.

Streptomyces own a linear chromosome with a high G+C content (typically more than 70%), whose size can vary from 8 to 11Mbp, two times more than the well studied bacteria *Escherichia coli* and *Bacillus subtilis* (both approximately 4Mbp). Certainly, their relatively large genomes reflect the complex lifecycle of these microorganisms, but also the vast arsenal of secondary metabolites produced by *Streptomyces*. Recently, several genomes of *Streptomyces* were completely sequenced (Bentley *et al*., 2002; Ikeda *et al*., 2003; Ohnishi *et al*., 2008), enlarging our knowledge of these fascinating microorganisms.

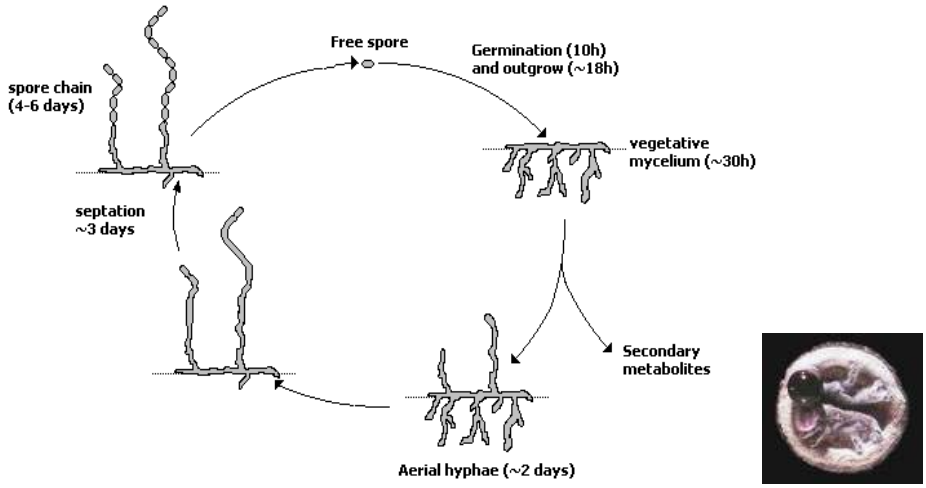


Fig. 27 The complete *Streptomyces* lifecycle usually lasts 5-6 days. The morphological differentiation, with aerial hyphae and spore production, occurs in most of the cases only on solid medium, while on liquid culture it usually stops at the vegetative state. Aerial hyphae and spores are hydrophobic and pigmented. The picture of a single *S. coelicolor* colony with a drop containing probably secondary metabolites is also represented.

7.2 *Streptomyces ambofaciens*

Similarly to other streptomycetes, *Streptomyces ambofaciens* ATCC23877 has a large linear chromosome of about 8Mbp and in addition it contains two plasmids, pSAM1 and pSAM2 (Pernodet *et al*., 1984; Leblond *et al*., 1990). At the extremities, the chromosome presents 200Kb of terminal inverted repeated regions (TIR) (Leblond *et al*., 1996).

When *S. ambofaciens* ATCC23877 was isolated in 1954 in Peronne, France (Pinnert-Sindico, 1954), it was originally described to produce only two antibiotics: the 16-membered macrolide

spiramycin, which is used in human therapy, and the pyrrole-amide congocidine, which due to its toxicity has not found medical application. Through genome sequencing, the biosynthetic gene cluster of spiramycin has been completed characterized, as well as that of congocidine (Richardson *et al*., 1990; Karray *et al*., 2007; Juguet *et al*., 2009).

S. ambofaciens is a well studied model organism for genetic instability phenomenon, typical of *Streptomyces* species. This instability is limited to chromosomal rearrangements, especially large deletions (up to 2Mbp) in the terminal regions (Leblond *et al*., 1991), thus named unstable regions. DNA amplifications of specific regions (Amplifiable Unit of DNA, AUD) have been often observed in proximity of DNA deletions. All these mutations can affect morphological differentiation, such as sporulation or antibiotic biosynthesis. For example, Dary *et al*. (1992) showed that an iterative repetition of a particular AUD prevented spiramycin production in *S. ambofaciens*.

Together with the Genoscope (Evry, France), the laboratory of Jean-Luc Pernodet (Orsay, France) and our laboratory, a genome sequencing project started in 2000 and at the beginning of my thesis the content of the dispensable regions (approximately 3Mbp) of the chromosome were completely sequenced (Choulet *et al*., 2006). This project aimed to delve into the mechanisms of genetic instability, focusing in particular in the unstable regions of the chromosome, but also to investigate the potential of this microorganism in producing secondary metabolites. The extremities of the chromosome have revealed the presence of 11 putative secondary metabolite clusters (Fig. 28 and Table 5), coding for polyketides, polypeptides, siderophores or terpenes (Choulet, thesis 2006).

One of these clusters was identified in the TIR, thus occurring in duplicate, and was demonstrated to be responsible for the biosynthesis of an angucycline-like compound, named kinamycin, and an orange pigment (Pang *et al*., 2004; Bunet *et al*., 2008). Two NRPS clusters were assigned to produce, respectively, the siderophore coelichelin (Lautru *et al*., 2005) and the antibiotic congocidine (Juguet *et al*., 2009), while another cluster is responsible for the biosynthesis of the second siderophore, desferrioxamine. For the other clusters the corresponding product still remains unknown. Interestingly, all the PKS clusters showed a novel type of organization and a very low percentage of identity (head to tail less than 50%) with other PKSs identified in *Streptomyces* sp., suggesting that the clusters might synthesize new bioactive molecules. Besides this, upon completion of the genome sequencing project, it is likely that other secondary metabolite clusters would emerge. Spiramycin cluster is indeed located in the core region of the chromosome, as well as the desferrioxamine gene cluster (Barona-Gomez *et al*., 2006).

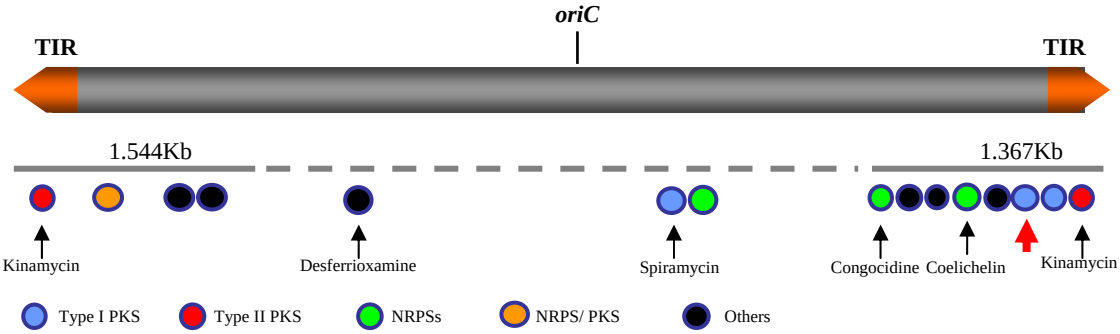


Fig. 28 Schematic representation of *S. ambofaciens* chromosome. At the moment, only the extremities are completely sequenced (solid lines) and the 11 secondary metabolite clusters identified are represented; for some of them the corresponding products have been isolated and characterized. From partial sequences of the core region other secondary metabolite clusters have already emerged, in particular that of spiramycin and desferrioxamine. The red arrow indicates the type I PKS cluster, which is the object of this thesis work.

	Type of cluster	**Limits of the cluster**	**Size (kb)**
TIR	Type II PKS (kinamycin)	SAMT0158-SAML/R0185	33
Left arm	NRPS/PKS	SAML0370-SAML0383	24
	Phytoene synthase	SAML0729-SAML0735	8
	Terpene synthase	SAML0739-SAML0744	5
Right arm	Type I PKS	SAMR0265-SAMR0278	25
	Type I PKS	SAMR0454-SAMR0485	156
	Lycopene cyclase	SAMR0510-SAMR0513	5
	NRPS (coelichelin)	SAMR0548-SAMR0559	24
	ACP synthase	SAMR0594-SAMR0609	18
	Terpene synthase	SAMR0831-SAMR0836	6
	NRPS (congocidine)	SAMR0894-SAMR0921	34

Table 5 List of putative secondary metabolite clusters identified in the extremities of *S. ambofaciens* chromosome. For some of them, the corresponding product has been already characterized (in brackets). SAMT indicates that the cluster is located in the TIR region, SAML that is in the left arm and SAMR in the right arm.

8. OBJECTIVES OF THE THESIS

This thesis was in line with a European programme (ActinoGEN, FP6) that aimed to exploit actinomycete genomes to discover novel natural antibacterial agents in order to address the ever growing problem of multi-drug resistant pathogens and the emergence of new infectious diseases. The genome sequence project of *S. ambofaciens* ATCC23877 revealed several orphan secondary metabolite clusters, whose products are unknown. In particular, we decided to focus our attention on a large modular type I PKS cluster, located on the right arm of the chromosome. Indeed, previous *in silico* analyses showed that the percentage of identity (head to tail) with other known type I PKS clusters was quite low, as well as the type of organization found (Choulet, thesis 2006), leading to the goal to identify and potentially isolate a novel compound. Therefore, the main objective of my thesis was to isolate this natural product, by first verifying the potential functionality of the genes in the cluster and by predicting the linear structure of the polyketide backbone. Another important objective of my work was also to investigate in the mechanisms of regulation that control this cluster, with a particular interest in a possible cross regulation with the spiramycin cluster. Indeed, in our laboratory an amplification of the locus containing the type I PKS cluster was observed to be associated with the loss of spiramycin production (Dary *et al*., 1992; Aigle *et al*., 1996).

RESULTS

An article that includes part of the results presented in chapter 1 and 2 is in preparation for the journal *Nature*. Moreover, the results of chapter 2 have been valorised and protected by a patent demand done the 24 July 2009 by the University Henri Poincaré (Nancy, France) jointly with the University of Warwick (UK).

Laureti L., Song L, Corre C., Huang S., Leblond P., Challis G. and Aigle B.
"Discovery of a novel macrolide in *Streptomyces ambofaciens* by awaking a sleeping giant"

Patent No: 09290587.6-2101
"Sambomycin and derivatives, their production and their use as drugs"

CHAPTER 1

In silico characterization of a large modular type I PKS cluster

The sequencing of the extremities of *S. ambofaciens* ATCC23877 chromosome has unveiled the presence of several orphan biosynthetic clusters for putative secondary metabolites (Fig. 28) (Choulet, thesis 2006). Among these, a large type I PKS was identified on the right arm and according to our previous annotation (see the SAMDB server at http://www.weblgm.scbiol.uhp-nancy.fr/ambofaciens/) the cluster included 32 ORFs, from *sam*R0454 to *sam*R0485 (Choulet, thesis 2006). Nevertheless, based on the sequence analyses and considering the putative function attributed to the genes (Table 6), the boundaries of the cluster were approximately fixed from ORF *sam*R0465 to *sam*R0487. Indeed, the gene *sam*R0464 presumably codes for an endoribonuclease and the gene *sam*R0488 for a "cold-shock" DNA binding protein, neither function are related to the biosynthesis of a secondary metabolite. Hence, the cluster is presumptively composed of 25 genes (Fig 29), that covering almost 150 kb put this cluster among the largest modular PKS gene clusters. Examples of other large secondary metabolite clusters are the 128 kb daptomycin gene cluster, a NRPS cluster in *S. roseosporus* (Miao *et al*. 2005), and the 137 kb candicidin gene cluster, a modular PKS cluster in *S.* sp. FR-008 (Chen *et al*., 2003).

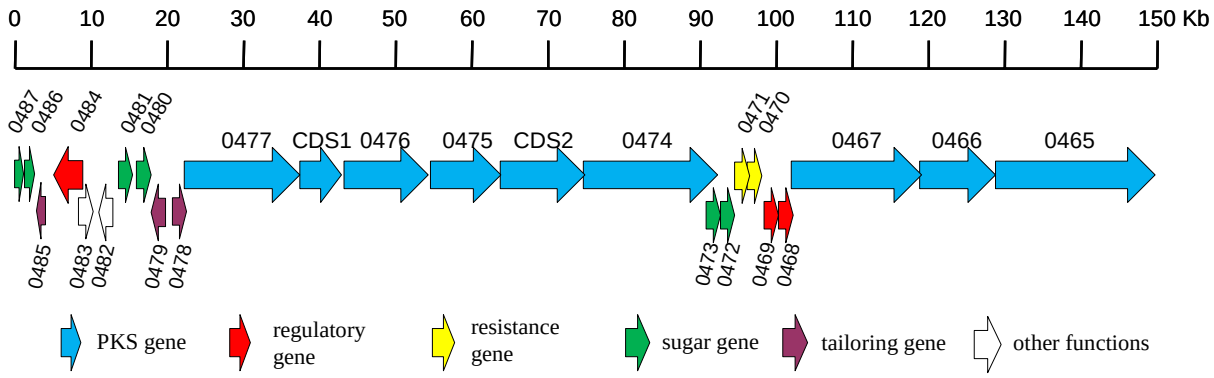


Fig 29 In scale representation of the type I PKS cluster. The genes are colored according to their putative function.

The most important goal of my thesis was to isolate the cryptic product synthesised by this large modular PKS. Therefore, we started with the analysis and the characterization of the biosynthetic genes, as well as the other genes in the cluster, to verify their functionality and to predict the possible final structure of the expected metabolite.

ORF	Product size (aa)	% identity / Similarity	Species	Putative function	Proposed function
***sam*R0465**	8154			Type I PKS	Polyketide biosynthesis
***sam*R0466**	3661			Type I PKS	Polyketide biosynthesis
***sam*R0467**	5771			Type I PKS	Polyketide biosynthesis
***sam*R0468**	217	84/88	*S. griseus* (SGR874)	Response regulator	Regulation
***sam*R0469**	442	77/84	*S. griseus* (SGR875)	Histidine kinase	Regulation
***sam*R0470**	261	87/93	*S. griseus* (SGR876)	Putative permease protein	Resistance
***sam*R0471**	312	88/93	*S. griseus* (SGR877)	Putative ABC transporter ATP-binding protein	Resistance
***sam*R0472**	244	55/68	*S. erythraea* (EryCVI)	N-dimethyltransferase	Sugar biosynthesis
***sam*R0473**	185	46/59	*S. fradiae*	Isomerase	Sugar biosynthesis
***sam*R0474**	6333			Type I PKS	Polyketide biosynthesis
***sam*RCDS1**	3556			Type I PKS	Polyketide biosynthesis
***sam*R0475**	3157			Type I PKS	Polyketide biosynthesis
***sam*R0476**	3565			Type I PKS	Polyketide biosynthesis
***sam*RCDS2**	1569			Type I PKS	Polyketide biosynthesis
***sam*R0477**	5447			Type I PKS	Polyketide biosynthesis
***sam*R0478**	414	43/61	*Nocardiopsis dassonvillei*	Cytochrome P450	Lactone ring modification
***sam*R0479**	401	41/56	*Roseiflexus castenholzii*	Cytochrome P450	Lactone ring modification
***sam*R0480**	369	74/82	*Streptomyces* sp. TP-A0274	Aminotransferase	Sugar biosynthesis
***sam*R0481**	418	48/63	*Micromonospora griseorubida* (MycB)	Glycosyltransferase	Sugar attachment
***sam*R0482**	595	64/75	*S. hygroscopicus*	Acyl-CoA synthetase	Unknown
***sam*R0483**	532	62/75	*Micromonopora* sp. ATCC39149	Carboxyl transferase	Unknown
***sam*R0484**	958	35/47	*S. venezuelae* (PikD)	Transcriptional activator (LAL)	Regulation
***sam*R0485**	255	87/94	*S. griseus* (SGR200)	Type II thioesterase	PKS editing
***sam*R0486**	329	77/84	*Stremptomyces tenebrarius* (AprE)	dTDP-glucose-4,6-dehydratase	Sugar biosynthesis
***sam*R0487**	290	74/85	*S. avermitilis* (AveBIII)	Glucose-1-phosphate thymidyltransferase	Sugar biosynthesis

Table 6 Principle characteristics of each gene belonging to the type I PKS cluster. For the putative functions, the results were obtained with BlastP program and the best hits are indicated, taking into account a "head to tail" alignment (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

The cluster is composed of nine large PKS genes: *sam*R0465, 0466, 0467, 0474, CDS2, 0475, 0476, CDS1 and 0477, which belong to the modular type I family, characterized by several polyfunctional modules. The program SEARCHPKS (http://www.nii.res.in/searchpks.html,

Yadav *et al*., 2003a) allowed us to identify the number of modules in each PKS gene and the type of enzymatic domains; in total, 25 modules, including a loading module, and 112 catalytic domains, typically of modular PKS, were recognized in the cluster (see Fig 30).

The PKS genes are not arranged straight forwardly, but they are separated by six genes of the cluster (see Fig. 29). The order of the biosynthetic genes does not always reflect the order in which the PKS genes interact into an assembly line to channel all the precursors. For a correct prediction of the structure, it is important to identify the order of interaction of the PKS proteins. For this cluster, the order was established according to the position of the initiation and termination module, located respectively at the beginning of the gene *sam*R0467 and at the end of *sam*R0474. The first module of SAMR0467 contains along with a AT and ACP domains a particular KS domain, named KS^{Q}, which has been described to have a role in carboxylating the starter unit (Bisang *et al*., 1999), loaded by the initiation domain. The last domain of SAMR0474 is a thioesterase, known to be involved in the release of the polyketide backbone (Staunton and Weissman, 2001). In addition, preliminary observations of the intergenic regions suggested that the PKS genes *sam*R0467, 0466 and 0465 are cotranscribed, as well as *sam*R0477, CDS1, 0476, 0475, CDS2 and 0474. Taken together all this data, the order of involvement of the PKS gene products has been proposed to be SAMR0467→0466→0465→0477→CDS1→0476→0475→CDS2→0474.

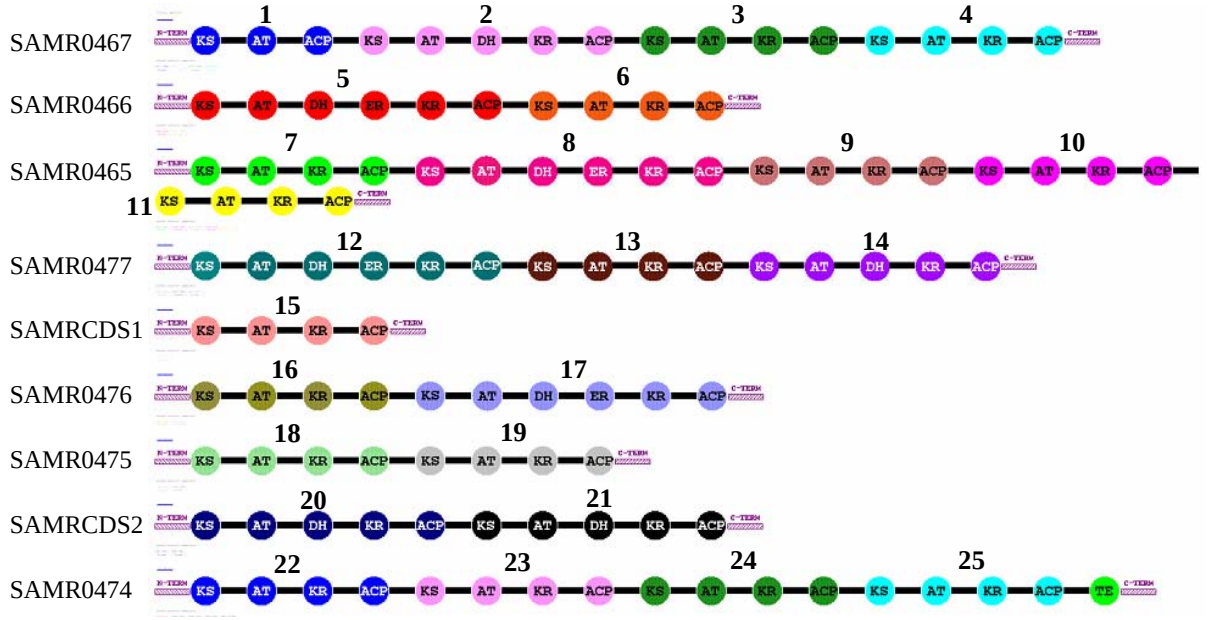


Fig. 30 Modular organization of the biosynthetic genes in the type I PKS cluster (Yadav *et al*., 2003a; http://www.nii.res.in/searchpks.html). Each module is numbered and represented in a different colour. It contains the typical core domains of a modular type I PKS (KS+AT+ACP), together with auxiliary domains (KR, DH, ER). The first module of SAMR0467 contains the loading module (named module 1), whereas the last module of SAMR0474 the thioesterase domain (named module 25).

1.1 Sequence analysis of the enzymatic domains

To verify the potential functionality of the PKSs, each enzymatic domain was analysed, by sequence comparison with already well characterized domains of other PKSs, looking for the conserved motifs and the active sites.

1.1.1 Ketosynthase domains

A functional KS domain contains at the N-terminus a motif similar to EPIAIVGMACR and at the C-terminus a consensus of GTNAHV$^{V/}_{I}{}^{L}/_{V}$E (Donadio and Katz, 1992). In the central region of the gene is situated the active site, a conserved cysteine followed by three serine residues (marked with an asterisk in Fig. 31). All the elongation modules (2-25) showed to have a potential functional KS domain (Fig 31). The KS domains, as well all the other enzymatic domains, have been characterized to be active as dimers to form a channel for the polyketide elongation (Staunton and Weissman, 2001).

```
KS1     EAFAIIGYACRLPG-APGPAPLWE-127-PSLTVDAGQSSSLVSVHLACES-215-FVGVSSFGMGGTNCHLVLGPAP
KS2     DPVVIVSMACRFPGGIDSPAALWE-134-PAVTVDTACSSSLVALHLAAQS-220-RAGVSSFGISGTNAHAVIEAAP
KS3     EPIAVVAMSCRYPGGVDGPEELWR-139-PAVTVDTACSSSLVALHLAARS-220-RAAVSSFGISGTNAHTVIEQAP
KS4     DPVAVVAMGCRFPG-AATPEEFWD-131-PAVTVDTACSSSLVALHLAAQS-219-RAAVSSFGISGTNAHVVLEQAD
KS5     EPIAIVSMACRFPGGVRTPEQLWH-140-PALTVDTACSASLVALHLAVQS-220-RAAVSSFGVSGTNAHTVLEQAP
KS6     EPIAIIGMSCRLPGGVRSPADLWR-131-PAVTVDTACSSSLVAVHLAVQA-220-RAGVSSFGMSGTNAHVIIEQAE
KS7     EPIAIVAMSCRLPGGVGAPEDLWR-134-PAVTVDTACSSSLVALHLAVQA-220-RIGVSAFGVSGTNAHLILEEAP
KS8     EPIAIVGMSCRYPGGANSAEQLWD-130-PAVTVDTACSSSLVAVHLAVQA-220-RAGVSSFGMSGTNAHVIIEQAE
KS9     EPIAIVGMACRLPGGVASPADLWP-130-PAVTVDTACSSSLVALHLAVQA-221-RAGVSSFGISGTNAHVILEAAG
KS10    EPIAIVGMSARYPGGVRNAEQLWE-130-PAVTVDTACSSSLVALHWAAQA-220-RAGVSSFGISGTNAHILLESAP
KS11    EPLALVGMACRLPGGVGSPDELWD-133-PAVTVDTACSSSLVALHLAAQA-220-RAGVSSFGISGTNAHVILEAEP
KS12    EPIAIIGMSCRYPGDVHTPEDLWR-139-PAVTLDTACSSSLVALHLACRS-220-RAGVSSFGVSGTNGHVILEEAP
KS13    DPIAIVGMACRFPGGADTPEALWR-139-PAVTVDTACSSSLVALHLAIGA-220-RAAVSSFGMSGTNAHTILEQAL
KS14    EPLAVIGMSCRYAGGVDSPEDLWR-140-PAATVDTACSSSAVAIHLACQA-221-RAGVSSFGGSGTNAHLVLEQYV
KS15    EPIAVIGMACRYPGGVRSPEDLWK-128-PAVTVDTACSSSLVALHLAAQA-218-RAGVSSFGVSGTNAHVILEGVE
KS16    EPIAIVGMACRLPGGVDSPASLWD-139-PAVTVDTACSSSLVALHLAVRA-218-RAGVSSFGVSGTNAHVILEGVE
KS17    EPVAIIGMACRYPGGVTSPDELWQ-139-PTLTVDTACSSSLVALHLAVRA-218-RAGVSSFGVSGTNAHVVLEAAP
KS18    EPVAVVGMGCRYPGGVHSPEDLWH-139-LAVTVDTACSSSLVATHLAVQA-221-RAGVSSFSISGTNAHLVLEGAD
KS19    EPLAIVGMACRYPGGASSPEELWR-135-PTLTVDTACSSSLVALHLAARS-220-RAGVSAFGISGTNAHVVLEHAP
KS20    EPMAVVSVACTLPGGVRSPEDLWR-138-PAVSVDTACSSSLTALHLAAQS-218-RAAVSSFGISGTNAHVILEEAP
KS21    EPIAIVAMSCRYPGEVNSPEDLWR-140-PAMTVDTACSSSLVTLHLACQA-220-RAAVSSFGISGTNAHVVLEEYR
KS22    EPIAIIGMSCRFPGGADSPEALWR-138-PALTVDTACSSSLVALHLAMRS-220-RAAVSSFGVSGTNAHVVLEEYR
KS23    DPIVIVGMACRFPGGVRSPEDLWK-132-PAVTVDTACSSSLVALHLACQS-218-RAGVSSFGISGTNAHTILEQAQ
KS24    EPVAIIGMSCRFPGGVDSPEALWR-136-PALSVDTACSSSLVALHLAARS-218-RAGVSSFGMSGTNAHVILESAP
KS25    DPVVIVSMSCRYPGGADSPENLWE-130-PAVTVDTACSSSLVALHLAAQS-218-RAGVSSFGISGTNAHVLIEEAP
                                      ****
```

Fig. 31 Alignment of all the KS domains of the type I PKS cluster in *S. ambofaciens*. The characteristic motifs at the N and C termini are underlined in yellow and light blue, respectively. The amino acids marked with an asterisk correspond to the active site. In the KS1, the cysteine (C) is replaced by a glutamine (Q) underlined in red, typical of a loading module with a KS^{Q} domain.

The initiation module of this modular PKS is not composed of the usual AT+ACP didomain, but it presents an additional KS^{Q} domain. The alignment with the other KS domains of the cluster (see Fig 31) showed indeed that in the active site the conserved cysteine is replaced by a glutamine, which is no longer able to catalyse a condensation reaction (Bisang *et al*., 1999). The presence of the KS^{Q} domain is probably required to decarboxylate the starter unit. Except for

this mutation in the active site, the KS^Q domain presents the conserved sequences of the KS domains.

1.1.2 Acyltransferase domains

All the modules contain a putative functional AT domain, responsible for the transfer of an activated precursor to the ACP domain (Fig 32). The active site is composed of several conserved amino acids QQGHSQGRSHV, spread out all long the protein sequence (Yadav *et al*., 2003b). The serine residue, underlined in blue in Fig. 32, is directly involved in the catalytic attachment of the substrate; the arginine (in yellow) can discriminate between a mono and a dicarboxylic acid, and in the latter case it interacts with the carboxyl group of the precursor to stabilise the binding (Yadav *et al*., 2003b). The limits of the AT domain were fixed according to Donadio and Katz (1992).

Fig. 32 Amino acid alignment of all the AT domains. The amino acids of the active site highly conserved are marked with an asterisk (Yadav *et al.*, 2003b). The amino acids boxed are involved in the specificity of the substrates; in particular the amino acid at position 200 (the number referred to the position in the *E. coli* FAS AT domain) is underlined in green. The consensus sequences for malonyl-CoA and methylmalonyl-CoA substrate are also given. The arginine underlined in yellow interacts with the carboxyl group of the precursors. The substrate for AT13 is supposed to be an uncommon one because the amino acid at position 200 is different from the ones predicted for malonyl or methylmalonyl-CoA.

The specificity for a particular starter unit and for the extender units can be determined by the amino acid sequence of the AT domains, in particular by those in the active site region (Haydock *et al*., 1995; Yadav *et al*., 2003b). The amino acid at position 200 (underlined in green in Fig. 32) plays an important role in the choice of substrate: a phenylalanine or a proline

residue is specific for malonyl-CoA, whereas a serine is observed when a methylmalonyl-CoA is incorporated in the polyketide chain. When a different amino acid is present, as the case of the AT13 of this cluster (Fig. 32), the precursor is an uncommon one and cannot be predicted. The AT present in a loading domain containing a KS^{Q} domain is usually specific for a malonate or a methylmalonate, since the decarboxylation will generate an acetate or a propionate as starter unit. In our case, the first precursor to be incorporated is a methylmalonyl-CoA, thus a propionate; then 16 malonyl-CoA, 7 methylmalonyl-CoA and an unusual substrate are the extender units used for the biosynthesis of the polyketide chain (see Fig. 32).

1.1.3 Acyl carrier protein domains

The ACP domain is the shortest enzymatic domains of PKSs, since it contains between 60 and 80 aa. The ACP domain presents a conserved motif LGFDS, in which the highly conserved serine residue is the phosphopantetheine attachment site (Donadio and Katz, 1992), target for the phosphopantetheinyl transferase that activates the ACP domain. All the ACP domains of the cluster showed to be potentially functional (Fig. 33).

```
                                    *****
ACP1    ELIRSHAAVVGGFAAAEEVDPEQTFKALGIASLTLVELRTRLATATGLSLPPTFLFDHPTPAAAARHL
ACP2    ELVRTTVAAVLGHSSGTGVDTGRTFKDLGFDSLTGVELRNRLAAATGLRLAATLVFSHPSPAALTAHL
ACP3    ELVRTLAAFVLGHGTAATVAPDKAFRDLGFDSLTAVELRNALRAATGLPLPAGLIFDHPTPAALATRL
ACP4    DLVVSRTAAVLGHGTPAAIDPDRQFRDLGTDSVTAVELRNLLDVATGRTLPATLVFDHPTPAALATHL
ACP5    GLVCAEAATVLGHTGADAVAPDRVFKELGFDSLTSVELRNRINAA
ACP6    DLVRTEAAAVLRYREGQVVEAAQAFKDLGFDSLTAVELRNRLSRTTGLRLPVSLAFNHRTPTGVAAYL
ACP7    DAVRTHTAQVLGHTNVAVVEPRRAFRDLGFDSLTAVELRNLLTSETGLPLPTTLVFDHPTPTALTDHL
ACP8    DLVRGNAATVLGHARADAVGARQAFRDLGFDSLTAVELRNRLNTATGLRLPATLVFDHPTPTVLAEHL
ACP9    DLIRARAAVSLGHASAEAIDPDRSFRDAGFDSLTAIELRNLLGADTGLTLPATLVFDHPSPTALAEHL
ACP10   AAVRSWAATVLGHGSAEAVGAGQAFRDLGFDSLAAVEFRNLAGVRTGLRLPASLVFDRPTPAELARYL
ACP11   ELIRGALVVVLRYDAHEQIDGARPFRDLGFDSLTAVEFRDVLARECGTPLPSTIVFDYPTPAALVDHL
ACP12   DLVRGLVASVLGFASPDAVDPSRPFKDLGFDSLTSVELRNRIGRTAGRRLPATLVFDHPTPEALATHL
ACP13   DLVREHAAAALGYADTAAIEAGRAFRDLGFDSLMAVDVRNRLQAATGRKLPPTLVFDHPTPADLAGHL
ACP14   DLVRGQVAATLGFATPGEVDVDRGFLELGMDSLTGVELRNRLAAS
ACP15   ELVRAQTALVLGHAHGGDIKPDRAFRDLGFDSLTAVEMRNRLTAVSGLSLTATVVFDYPTPAELAEHL
ACP16   ESVRARAAMVLGHHDADAIEPRRAFRDLGFDSLTAVELRNLLAEVTGFTLPATVVFDHPTPAALAEHL
ACP17   DWVRSEAATVLGFAGAEAVERDRAFKDLGFDSLTAVELRNRLGAATGTRLPATLVFDYPTPDALAERL
ACP18   DLVRRETAAVLGYA--EPVDADRPFKDLGVDSLTAVTVRNRVAGALDLRLPTTLVFDHPTPSDVSRHV
ACP19   ELVRTHAAVALGHARAEDIEPERTFRDLGADSLTAVALRDGLASA
ACP20   DVVRSEVAGVLGHTGAEPIEDHRPFKDLGFDSLTAVELRNRLGAVTGLRLPAGLVFDHPSPQALVAHL
ACP21   DLVRGHAAAVLGHASAEAVDPDTGFVDLGFSSLSAIEIRNQLSAATGLRLTTTLVFDHPTASRLADHL
ACP22   DLVRDQVALVLGYESGSEVEPGRALGELGLTSAGAVELRNRLTFTTGLRISATVVFDHPSPLALARFL
ACP23   HLVLEAVAGVLGHGSPADVDRERAFHDMGFDSLTALELRNLLAARTGLTLHAGLVFDFPTPGALAEHL
ACP24   DLVRTVVATVLGHPGPRSVDPDRVFTELGIDSLTALRVRDGLADATGLRLPATLVFDHPTARVAATEL
ACP25   DLVRGTAATVLGHTDAASVAPDRQFLELGVDSLAALEIRGTLERATGLSLPATVVFDHPTPAALARHL
```

Fig. 33 Alignment of all the ACP domains. The amino acids marked with an asterisk correspond to the active site, in particular the catalytic serine, underlined in yellow, is highly conserved. The limits of the ACP domain are based on Donadio and Katz (1992).

1.1.4 Ketoreductase domains

Among the auxiliary domains, a KR domain was identified in all the extension modules (2-25). At the beginning of a KR domain, the motif GxGxxGxxxA is characteristic for the binding to NADP(H), the cofactor used by the enzyme (Aparicio *et al*., 1996). A serine, a tyrosine and an asparagine residue constitute the catalytic triade (Reid *et al*., 2003). In particular, the Tyr is

responsible for the proton transfer to the carboxyl oxygen of the substrate; therefore, mutation of this residue would affect dramatically the activity of the KR domain. All the KR domains harbour a putative functional active site, but one, KR25, present in the PKS gene *sam*R0474, in which the Tyr is replaced by an arginine and the Asn is replaced by a serine; thus the domain is considered to be nonfunctional (Fig 34).

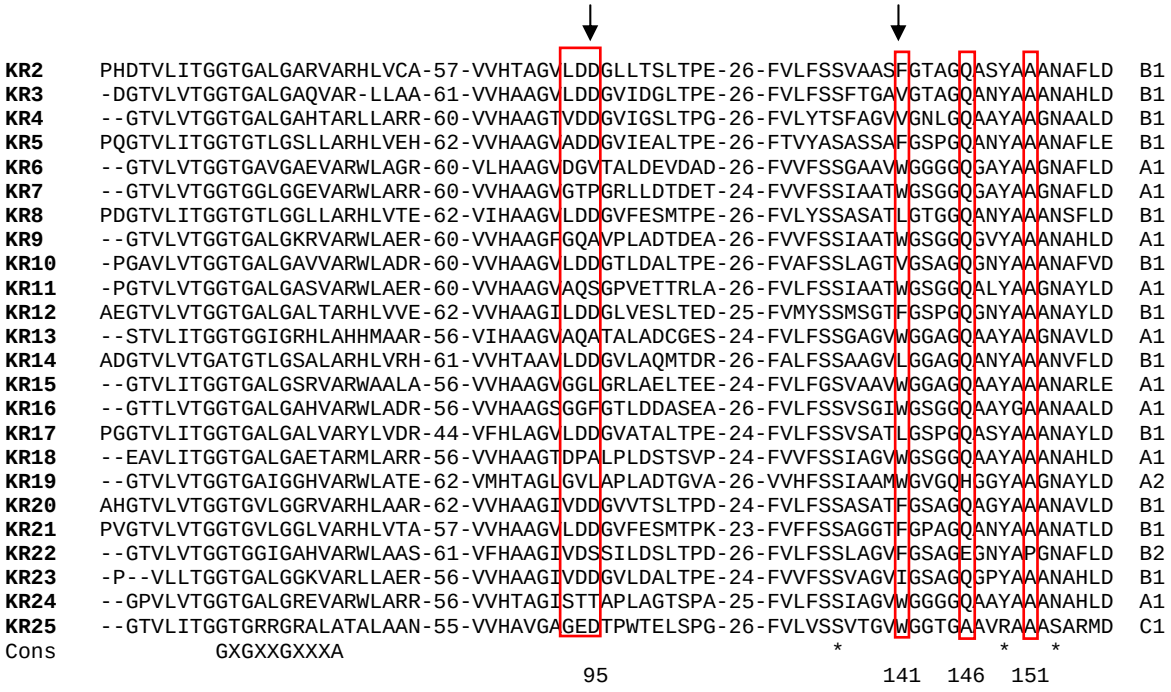


Fig. 34 Amino acid alignments of the KR domains. The consensus sequence at the N-termini was proposed to bind to the NAD(P)H (Aparicio *et al*., 1996). The amino acids of the catalytic triad are marked with an asterisk. The boxed amino acids are involved in the stereochemistry of the polyketide. The arrows indicated the residue 95 and 141 specific for the B and A-type alcohol stereochemistry, respectively. At the bottom the amino acid position referred to the KR domain in the DEBS genes are indicated. On the right of the alignment, for each KR domain, the configuration of α-substituents and β-hydroxyl groups derived from each precursor are given, according to the nomenclature proposed by Keatinge-Clay (2007). A1= 2R, 3S; A2=2S, 3S; B1=2R, 3R; B2= 2S, 3R; C1=2R.

The KR domain plays an important role in the control of the stereochemistry of the β-hydroxyl group and of the α-substituent derived from each precursor. Six different stereo combinations of the α-substituents and β-hydroxyl groups have been currently described. Two types of alcohol stereochemistry can be generated by a KR domain: the A-type produces a (*3S*)-3-hydroxylacyl-ACP intermediate, while the B-type produces a (*3R*)-3-hydroxylacyl-ACP intermediate (Fig. 35; Caffrey, 2003). The motif LDD and in particular the strictly conserved second Asp residue are typical of B-type, whereas a Trp in position 141 (the position referred to the KR domain of DEBS genes) is present only in the A-type (see Fig 34). When the KR domain is followed by a dehydratase activity, it was observed that the second Asp residue of the LDD motif is always present, suggesting that the substrate for the DH domain is specifically a B-type (Reid *et al*., 2003; Caffrey, 2003).

Similarly to the β-hydroxyl groups, two types of stereochemistry can be obtained for the α-substituents. According to the nomenclature given by Keatinge-Clay (2007), the appending 1 indicates an α-substituent with a R stereochemistry, while the appending 2 an α-substituent with a S stereochemistry (Fig. 35). A Gln residue in position 146 is typical of A1 and B1 types, while a His residue is typical of A2 type; a Pro residue in position 151 is observed in B2 type (Fig. 34). The nonfunctional KR domains belong to the C-type; C1 lacks the active site tyrosine, while C2 lacks the active site asparagine. In the cluster A1 and B1 types are predominant, with the exception of one A2 (KR19) and B2 type (KR22).

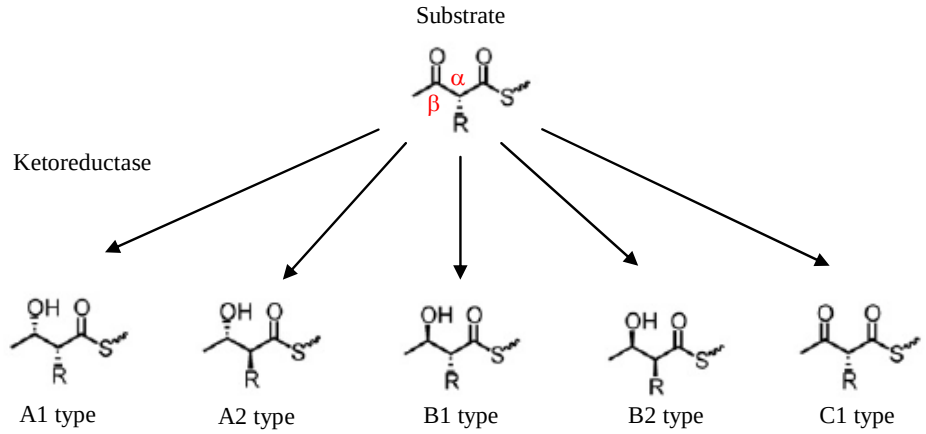


Fig. 35 The KR domain processes the same precursor, controlling the stereochemistry of the α-substituents and β-htdroxyl groups. According to its amino acid residue a KR domain can give rise to different kind of products, A1, A2, B1, B2 and C1 type (from Keatinge-Clay, 2007).

1.1.5 Dehydratase domains

In the cluster, eight DH domains were identified and they all appeared to be functional (see Fig. 36). Indeed, all sequences contain the characteristic motif HxxxGxxxxP (Donadio and Katz, 1992), in which the His residue is the directly implicated in the catalysis.

The order of interaction of the enzymatic domains does not correspond to their position in the module. In the sequence, the DH domains are found downstream of the AT domain and they can be followed by an ER domain before the KR domain, but they always interact as KR→DH→ER.

```
                                                          *    *    *
DH2     TS-SA--TALAAAGVDTAGHPLLGAVVPLAGDGGLVWTGVVGLATHPWLADHMVHGRIVLPGTALVDLALHA
DH5     VRLSA--GDLTSAGLGATGHPLLGAAVSMAAGGGALLTGRLSAATHGWLTDHTILGGVLVPGTAFVELAWQA
DH8     ----------------GLHHAILGATVALPS-GGAVLTGRLSPSSHPWLQDHAVRGQVLVPGTAFAELALRA
DH12    --------DVASAGLGVTDHPLLGAGVALAEEDGYLFTARISTATQPWLAEHRVHGRIVVPGTAFVDLAVRA
DH14    ------------AGLSGTGHPLLTATLTLADGASTVFTGRLSARTQPWLTDHAVLGAVLLPGTAFVEMALHA
DH17    VPASAHPGDVVSAGLGETGHPVLAAGVDLADDGGLLFTGRISLRTHPWLAEHRVHGRIVVPGTVFVDLAVRA
DH20    --------------LSGAGHPLLGALVSVAGDGGLVLTGRLSRATHPWLDDHRVLGAVLVPGTALVDLALHA
DH21    ------------GLDRTDHPMLGAAVPVAGTDGVLLTGRIAADTHPWIADHVLLGTVLLPGTALVELAV-R
```

Fig. 36 Alignment of all the DH domains present in the cluster. The characteristic motif is underlined in yellow and the amino acids of the active site are marked by an asterisk. The number assigned to the DH domain corresponds to the number of the module in which they were identified.

1.1.6 Enoyl reductase domains

Four ER domains were identified in the cluster, in particular in the module 5 (SAMR0466), module 8 (SAMR0465), module 12 (SAMR0477) and module 17 (SAMR0476). A typical ER sequence contains the motif LxHxAAxGGVG (Donadio and Katz, 1992), which is specific for NADPH binding, the cofactor used by this domain (Fig. 37). Therefore, all the ER domains are presumably functional and they would generate 4 saturated β-carbons. Very recently, Kwan *et al*. (2008) showed the contribution of ER domain in the stereochemical control of a methylmalonyl-CoA precursor, in particular of the methyl substituent (the α-substituent). The presence of a conserved Tyr produces a (2*S*)-methyl branch, while another amino acid residue at this position (i.e. Val, Ala or Phe) produces the opposite configuration. The module 8 loads a methylmalonyl-CoA and it contains an ER domain (ER2), which presents a Tyr residue (see Fig. 37), thus predicted to generate a (2*S*)-methyl branch.

```
ER5     AAGVNFRDVLIALGQYPDPTALMGSEAA……YYALVDLAGLSAGESVLVHAAAGGVGMAAVQVARHLGAEVYGTASPA……
ER8     AAGVNFRDVLIALGMYPDR-AQMGAEAA……YYGLVDVARARPGESVLVHSAAGGVGMAAVQLGRHLGLEVFGTAKPS……
ER12    AAGVNFRDVLNVLGMYPGE-VLVGGEAA……YYGLVDLAGLSAGESVLVHAAAGGVGMAAVQVARHLGAEVYGTASPA……
ER17    AAGMNFRDVLNVLGMYPGE-VELGGEAA……YYGLRDVGGLAAGESVLVHAAAGGVGMAAVQIARHVGAEVYGTASPG……
```

Fig. 37 Alignment of the four ER domains in the cluster. The amino acids boxed in red are involved in the NADPH binding. The Tyr residue involved in the control of the stereochemistry of a methylmalonyl-CoA precursor is underlined in yellow. The number assigned to the ER domain corresponds to the number of the module in which they were identified.

1.1.7 Thioesterase domain

The last module of SAMR0474 contains the termination TE domain. The sequence contains a GxSxG motif, common to the AT domains, as well as the GDH motif in the C-terminus of the protein, which is essential for the activity of the enzyme (Fig. 38; Donadio and Katz, 1992).

```
TE      LLVCVSGLTAGGG-51-DTPVVLFGHSGGAILAHRLAVRL-87-DWRSSWKAARDIVDVPGNHFTMMAEHAGTTARAVHRWI
```

Fig. 38 Amino acid sequence of the TE domain of the cluster. The characteristic motifs of the domain are underlined in yellow.

Based on the TE sequence and by sequence comparison analyses, it was possible to predict whether the TE is responsible only for the release of the polyketide backbone or also for the intramolecular cyclization. In the first case it will generate a linear compound; in the latter case it will give rise to a lactone ring structure. In the phylogenetic tree of Fig. 39, the TE of our cluster shows to be more closely related to the TE domains that generate lactone compounds, such as pikromycin, than to those generating the linear polyketides, as ECO-2301. Therefore, the most plausible hypothesis is that the final product undergoes cyclization to produce a lactone ring.

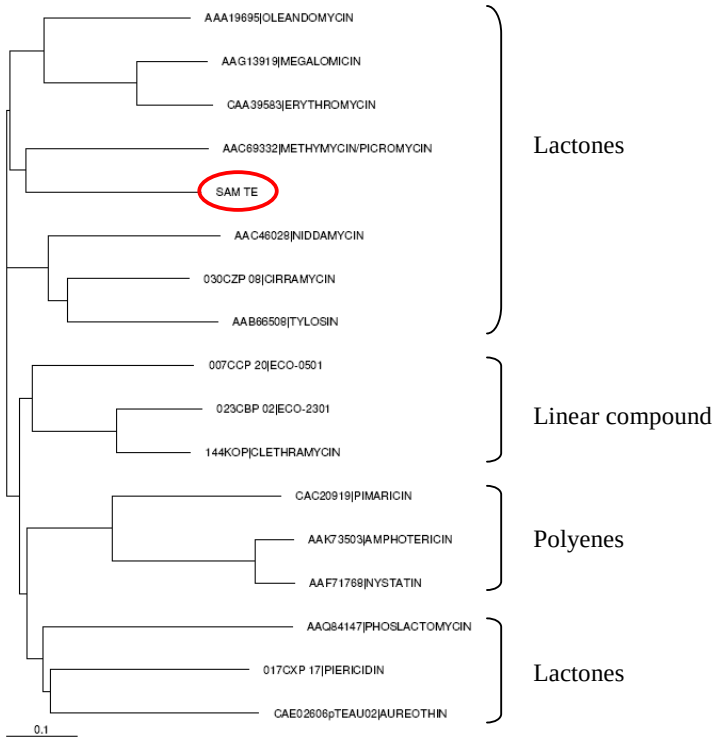


Fig. 39 Phylogenetic tree of some of TE domains characteristic of the biosynthesis of lactone compounds, linear polyketides and polyenes (James McAlpine, pers. comm.). The TE domain of the cluster is circled in red. The accession number of each known compound is given and is accessible on PubMed, except those of the linear compounds (Ecopia database).

1.2 Prediction of the linear polyketide structure

The *in silico* analyses on the enzymatic domains provided the information necessary to predict with high accuracy all the intermediates of the polyketide chain and the final linear structure of the polyketide backbone, after being realised by the TE domain (Fig. 40 A and B). Except for one precursor, loaded by the AT13 (Fig. 32), all the other 24 substrates were predictable, as well as their stereochemistry. The auxiliary domains are all presumably functional, but one (KR25); therefore the modifications on the β-keto group are also predictable.

The linear molecule contains fifteen hydroxyl groups and only one non-reduced β-keto group. The structure also presents a diene, which could absorb at a certain wavelength, helping in the isolation of the compound, two other double bounds and four saturated carbons.

According to the prediction, the linear polyketide backbone has a mass of approximately 1127Da with $C_{61}H_{107}O_{18}$ as its estimated molecular formula, which can vary based on the nature of the precursor loaded by the AT13 domain.

Although the TE domain is proposed to operate the cyclization, it was not possible to establish where it will occur and which kind of lactone ring it would generate; this can be unveiled only by spectrochemical analysis (i.e. Nuclear Magnetic Resonance, NMR).

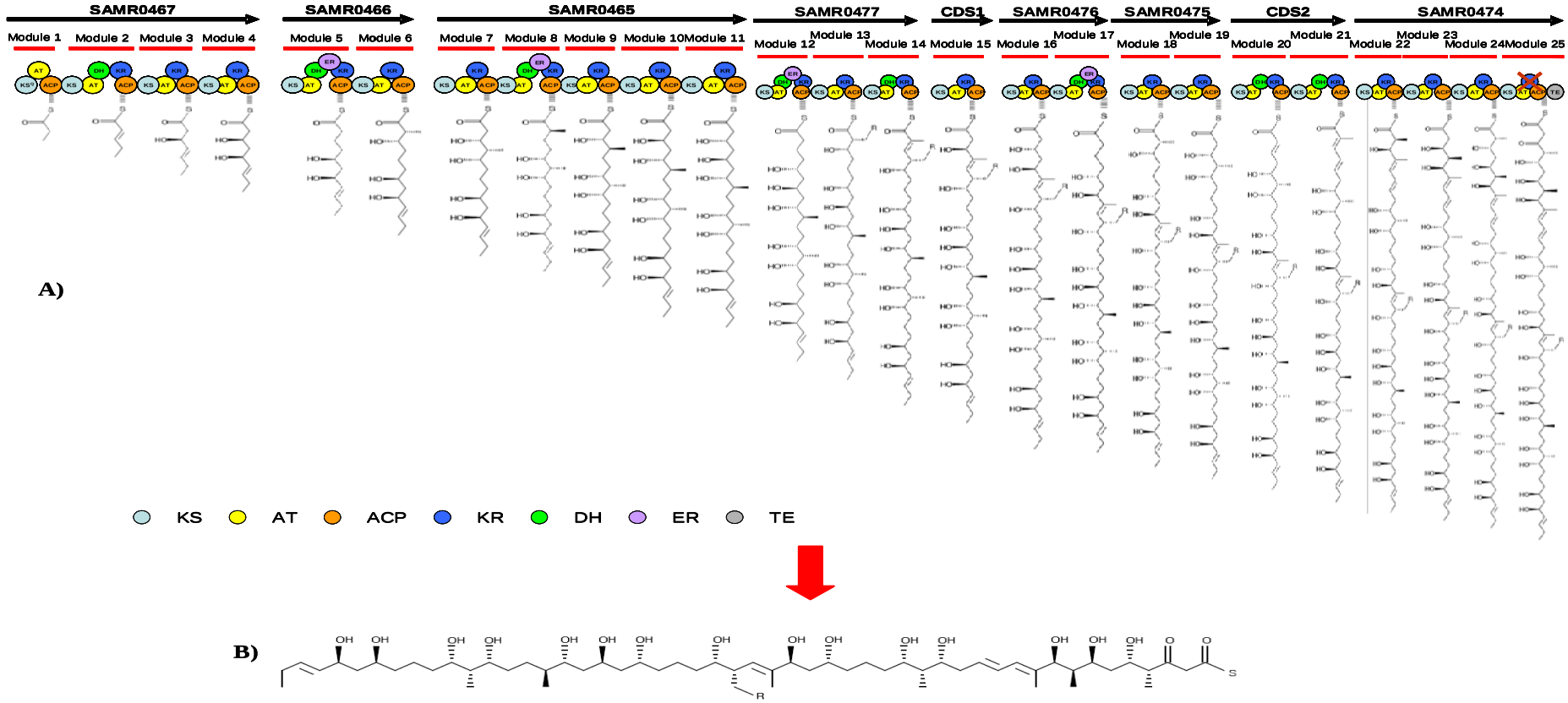


Fig. 40 A) Domain organization of the modules of the nine PKS genes. For each module the chemical intermediate is given, with the β-keto modification and the stereochemistry, according to our prediction. The nonfunctional KR24 is crossed in red. **B)** Predicted linear polyketide structure. The R group represents the unknown precursor loaded by the AT13 domain.

.

The cyclization often occurs between the first hydroxyl group and the last keto group of the linear structure. In our case it would produce a giant 48-membered lactone ring. Only the macrolide monazomycin have been described with such large lactone ring (Kuo *et al*., 1990).

1.3 Analysis of the other genes putatively involved in the biosynthesis

Along with the PKS genes, the cluster probably contains 16 additional genes; some of them are putatively involved in the biosynthesis of the molecule. According to the sequence analysis, they were divided in different functional groups.

1.3.1 Glycosyltransferase and sugar biosynthetic genes

The gene *sam*R0481 encodes a 418 aa protein which showed to belong to the glycosyltransferase GT-1 family. The best hit is with MycB, a GT identified in the gene cluster for the macrolide mycinamicin in *Micromonospora griseorubida* (see Table 6). Glycosyltransferases are characterized as enzymes that transfer an activated sugar donor to an aglycone structure; usually they are specific to transfer only one residue. The presence in the cluster of this gene along with additional sugar biosynthetic genes clearly indicates that the final product would be a macrolide. As for the cyclization, it is not possible to predict where the sugar would be attached.

The program SEARCHGTr (http://linux1.nii.res.in/~pankaj/gt/gt_DB/html_files/searchgtr.html, Kamra *et al.,* 2005) allows the prediction of the substrate specificity of a glycosyltransferase. For SAMR0481, the program suggested the incorporation of an amino 6-deoxyhexose, most probably a desosamine or a mycaminose residue.

The genes involved in the biosynthesis of these two deoxysugars have been already characterized (Melançon *et al*., 2007). Additionally, the structure of spiramycin, another macrolide synthesized by *S. ambofaciens* ATCC23877, contains two amino sugar moieties, a forosamine and a mycaminose (Karray *et al*., 2007). By sequence comparison, five genes inside the cluster were identified to be homologous to the spiramycin genes involved in the biosynthesis of the deoxysugar mycaminose (see Fig. 41).

The gene *sam*R0487 (homologous of *orf4* in spiramycin gene cluster) encodes a thymidyl transferase, which uses thymidine triphosphate to generate a TDP-α-D-glucose from an α-D-glucose-1-phosphate (White-Phillip *et al*., 2009). This step is necessary to activate the deoxysugar in order to be used by the glycosyltransferase. A 4,6-dehydratase, encoded by *sam*R0486 (homologous to *orf5*), produces a TDP-4-keto-6-deoxy-α-D-glucose, a common intermediate for sugar biosynthesis pathways (White-Phillip *et al*., 2009).

TDP-4-keto-6-deoxy-α-D-glucose is converted in TDP-3-keto-6-deoxy-α-D-glucose by a 3,4-isomerase (SAMR0473), which in turn is converted in a TDP-3-amino-3,6-dideoxy-α-D-glucose by an amino transferase (SAMR0480). A *N*-dimethyltransferase (SAMR0472) is the last

enzyme that takes part in the biosynthesis of a TDP-D-mycaminose (White-Phillip *et al*., 2009) (see Fig. 41).

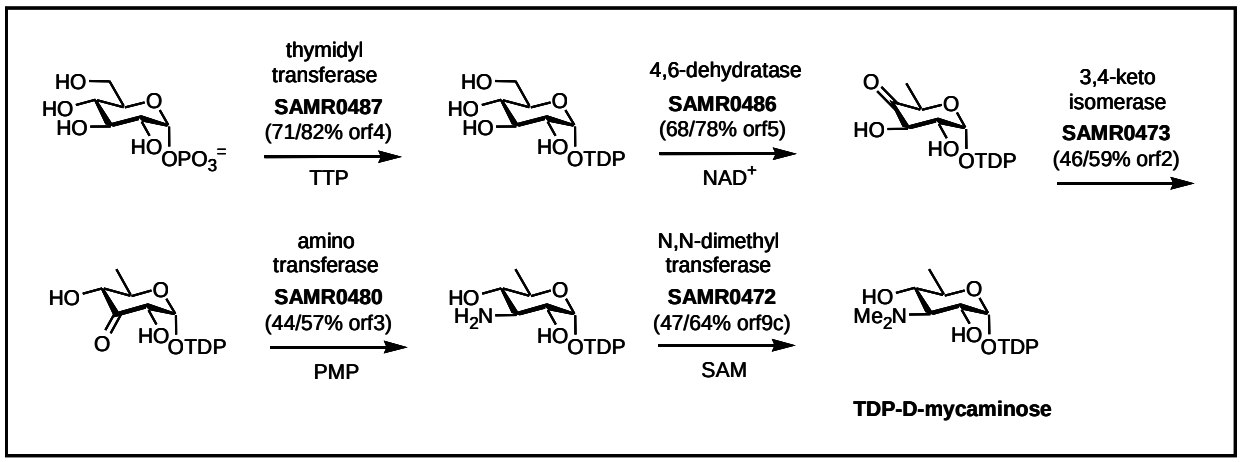


Fig. 41 Schematic representation of TDP-D-mycaminose biosynthetic pathway from an α-D-glucose-1-phosphate. For each enzymatic step, the gene product is indicated in bold. In brackets, the percentages of identity/similarity with the spiramycin enzymes are indicated (Karray *et al.*, 2007). Abbreviations: TTP, thymidine triphosphate; TDP, thymidine diphosphate; NAD^+, nicotinamide adenine dinucleotide; PMP, pyridoxamine 5´-phosphate; SAM, *S*-adenosyl-L-methionine.

The molecular formula of a mycaminose residue is $C_8H_{17}NO_4$ which corresponds to a mass of 191Da. Taking into account this information, the predictable molecular weight of the macrolide synthesized by this cluster would be around 1318Da ($C_{69}H_{124}NO_{22}$). In the chemical databases, no molecule was found with this mass and molecular formula, confirming once more that the expected compound produced by this large modular PKS has a novel interesting structure.

1.3.2 Cytochrome P450

By sequence analysis, two genes, *sam*R0478 and *sam*R0479, showed to belong to the cytochrome P450 monooxygenase family (CYPs). P450 are ubiquitous hemoproteins able to activate a dioxygen molecule to catalyze the insertion of a single oxygen atom into allylic positions, double bonds or even into non-activated C-H bonds, according to the following reaction (Hamdane *et al*., 2008):

$$R\text{-}H + O_2 + NAD(P)H + H^+ \quad \rightarrow \quad R\text{-}OH + H_20 + NAD(P)^+$$

These enzymes are commonly found in secondary metabolite clusters, where they are usually responsible for the stereo and region-specific hydroxylation reactions of the metabolite (Parajuli *et al*., 2004). Both SAMR0478 and SAMR0479 contain the heme-binding motif GxxxCxG (Parajuli *et al*., 2004), the highly conserved ion-pair ExxR, which is probably involved in both the redox partner interaction and heme binding, and the tetrapeptide motif Gx[DEH]T, the oxygen-binding site (Shrestha *et al*., 2008) (Fig. 42). Therefore, both cytochromes are most

probably functional and the final structure would present two additional hydroxyl groups, but it was not possible to predict their position.

```
SAMR0478  EMVPLCMFLPLAGTETTVNLIGNGLLALLEHPEQWDMLVADPSLADAVVRETLRYDPSVQQY-RRIAHTD
SAMR0479  EVVANILLIFNAGHETTVNLIGNGMLALLRQPEALEALRADPGLMATAVDELSRFDPPVTLS-SRIATAE
CYP105F2  EVAGIGVLMLIAGHETSANMLGIGTYTLLQNPGQWDLLRNDIGLIDQAVEELLRHQTIVQQGLPRGVTED
                    *
SAMR0478  LQMEGASIAAGEEVAICAGGANRDPEVYPDPGRFDITRDPGPENLAFSAGIHFCLGAALARMEAETALAA
SAMR0479  MEFGGKVIPPGSHVIGFLDAAGRDPERYPDPDRLDLSRTEP-KTLAFSAGPHFCLGAVLGRLEAATVFSK
CYP105F2  MEIAGHAIKPGETLLASLPAANRDPEVFPDPDRLDITREHNP-HLAFGHGIHLCLGMELARVEMRCAWRG
                                                      *
```

Fig. 42 Amino acid sequence alignment of SAMR0478, SAMR0479 and CYP105F2, a cytochrome P450 from *S. peucetius*, already characterized (Shrestha *et al*., 2008). The conserved motifs are underlined in yellow. The amino acids marked with an asterisk are directly involved in the catalysis.

1.3.3 Thioesterase type II

The product of *sam*R0485 shares a percentage of identity/similarity (87%/94%) with a putative thioesterase of *S. griseus* (SGR200) and also with the gene product of *aveG* (65%/79%), a thioesterase from the avermectin cluster of *S. avermitilis*. These thioesterases belong to the type II (TEII), since the genes code for a discrete protein and they are not located in the termination module of a PKS gene. The role attributed to TEII in the biosynthesis of polyketide is to remove aberrant groups attached to the ACP, which block the chain elongation (Fischbach and Walsh, 2006).

TEII contains a motif GxSxG, in which the Ser residue is the active site, and on the C-terminus a GxH motif, where the His is necessary to increase the nucleophilic activity of the Ser (Witkowski *et al*., 1992). SAMR0485 presents all the characteristics of a type II thioesterase (Fig. 43).

```
SAMR0485  MATSSDASDLWVRRFHPAPQAPVRLLCLPHAGGSASYYFPVSQRLAPRVETLAVQYPGRQDRRNESCLESVRDLADRIVEVLG
          PWQDKPLALFGHSLGATVAFEVALRLEARGSLPLALIASGRRAPSAPRDERVHLRDDEGLLAEIRRLDGTQSQLLEDDEIRRM
          VLPSIRADYKAAETYRYEPGPPLSTPVYALTGDSDPKATTDEVRAWADHTTGRFEMNVYPGGHFYLNAHAPAVTEEIAAWLVP
          AATAGR
```

Fig. 43 Amino acid sequence of SAMR0485. The conserved motifs are highlighted in yellow and green and the active site residues are underlined.

1.3.4 Additional genes

For the genes *sam*R0482 and *sam*R0483, we could not propose a precise function linked to the biosynthesis of the polyketide. SAMR0482 has the highest identity/similarity (64%/75%) with a NRPS module of *S. hygroscopicus*, probably being an acyl-CoA synthetase, the enzyme usually involved in the activation of fatty acid and PKS precursors. SAMR0483 is homologous to a

carboxyl transferase of *Micromonospora* sp. (62%/75% identity/similarity). The elucidation of the metabolite structure had given insights in the role of these two genes (see Chapter 3).

1.4 The resistance genes

In silico analysis highlighted the presence of two genes, *sam*R0470 and *sam*R0471, probably cotranscribed since the end of one gene overlaps with the start of the other (CATG), whose products might be responsible for the secretion and the resistance mechanism of the produced metabolite. Indeed, SAMR0470 and SAMR0471 showed a very high percentage of identity (87% and 88%, respectively) with SGR876 and SGR877 of *S. griseus* which were proposed to form an ABC transporter, a transmembrane pore, able to pump out cellular metabolites. However, the role of this transporter in *S. griseus* has not been clarified yet and the genes are not localised in a putative secondary metabolite gene cluster.

Three types of ABC transporters, associated with secondary metabolism, have been described based on the number and organization of the nucleotide-binding domains, and on the composition of the transporter system (Mendez and Salas, 2001). The gene *sam*R0470 and *sam*R0471 putatively encode the permease subunit and the ATP-binding subunit of a type I ABC transporter, respectively, as also observed for daunorubicin and oleandomycin gene clusters in *S. peucetius* and *S. antibioticus*, respectively (Mendez and Salas, 2001). The type I system is composed of two discrete proteins: an ATP-binding subunit which contains a unique nucleotide-binding domain, represented by the Walker A and B motifs, and a permease subunit, a hydrophobic protein with six transmembrane domains. The Fig. 44 shows the conserved Walker A and B motifs identified in SAMR0471 and SAMR0470 was analysed by the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) to confirm the presence of the six transmembrane domains (data not shown).

```
                                          #                                  #
SAMR0471  ……GGRPVVDGVSLAVEEGEIFGVLGPNGAGKTTTVESILGLRVP……AIALALVGSPRVVILDELTTGLDP……
OleC      ……GETKALDGVDLDVREGTVMGVLGPNGAGKTTLVRILSTLITP……DLAASMIGRPAVLYLDEPTTGLDP……
MtrA      ……GPVEAVAGVDFTVHTGEIVGFLGPNGAGKTTTMRMLTTLLPP……DMALGLMHQPELLFLDEPTANLDP……
DrrA      ……NGTRAVDGLDLNVPAGLVYGILGPNGAGKSTTIRMLATLLRP……DIAASIVVTPDLLFLDEPTTGLDP……
```

Fig. 44 Amino acid alignment of SAMR0471, OleC from *S. antibioticus*, MtrA from *Streptomyces argillaceus*, and DrrA from *S. peucetius*. All proteins form an ATP-binding domain of type I ABC transporters. The Walker A and B motifs (Walker *et al*., 1982) are underlined in yellow and the catalytic Lys and Asp residues are marked with a hash.

1.5 The regulatory genes

1.5.1 Two component system

Between the two sets of PKS genes, *sam*R0468 and *sam*R0469 were identified to form a two component signal transduction system (TCS), where the first gene codes for the response

regulator (RR) component and the second for the histidine kinase sensor (HR). Indeed, SAMR0468 contains the required catalytic Asp residue, which is the target for the phosphorylation operated by the HR; likewise, SAMR0469 contains the required catalytic His residue (Laub and Goulian, 2007).

Some TCSs have been characterised to regulate secondary metabolism, by sensing an environmental change that they convey by binding to specific DNA targets and modulating gene expression. As transcriptional regulator, the RR usually plays a negative role in the onset of the biosynthesis. For example, two component system *absA* of *S. coelicolor* is a global regulatory system which represses the production of four antibiotics: actinorhodin, undecylprodigiosin, methylenomycin and calcium-dependent antibiotic (Adamidis *et al*., 1990; McKenzie and Nodwell, 2007). However, in some cases, RRs have shown to have a positive control on secondary metabolism, such as AfsQ1 from *S. coelicolor* (Ishizuka *et al*., 1992). Similarly, two genes coding for a TCS were found nearside the clavulanic acid gene cluster in *S. clavuligerus* NRRL3585 and their overexpression increased antibiotic production, thus playing a positive role (Jnawali *et al*., 2008).

Both SAMR0468 and SAMR0469 showed a high percentage of identity with a two component system (SGR874 and SGR875) of *S. griseus* (Table 6), not located in a secondary metabolite gene cluster. It is interestingly to observe that these genes are adjacent and probably cotranscribed with the resistance genes *sam*R0470 and *sam*R0471 (the stop codon of one gene overlaps with the start codon of the following gene), which were also found to have homologous in *S. griseus* (SGR876 and SGR877). It might be possible that these four genes, coding for an ABC transporter and a TCS, originated from a horizontal transfer event.

1.5.2 LAL regulator

A third gene, *sam*R0484, was found to be putatively involved in the regulation of the cluster. Its product belongs to the LAL regulatory family. Similarly to other LAL regulators, SAMR0484 is a large protein (958aa) that contains at the N-terminus a Walker A and B domain, able to bind ATP or GTP, and in the C-terminus a helix-turn-helix (HTH) domain, to bind to DNA targets (Fig. 45). A TRP domain (from 389 aa to 852 aa) was also identified in the middle region of SAMR0484, probably involved in a protein-protein interaction (Fig. 45). The gene *sam*R0484 contains a single TTA codon at the beginning of the sequence (114aa), found also in other LAL, indicating that its translation would be dependent on the *bldA* tRNA gene (Leskiw *et al*., 1991). These kinds of regulators have been shown to be positive transcriptional activators of secondary metabolite clusters. For example, PikD controls the expression of the pikromycin biosynthetic genes in *S. venezuelae* (Wilson *et al*., 2001), as well as RapH regulates rapamycin production in *S. hygroscopicus* (Kuscer *et al*., 2007). The deletion of these genes prevents metabolite production, while their overexpression enhances the biosynthesis.

```
SAMR0484    ……KHGHGQVALVTGAVGSGKTQVLET……TLAGRGPLVLAVDDLQHTDAASLHCLL……
PikD        ……AAGDGTLLLVSGPAGSGKTELLRS……SVSRRTPFLVAVDDLTHADTASLRFLL……
RapH        ……TSGTGSIALIEGAASTGKTELLHS……ALVNRSPALITVDDVQHADPESVLWLL……
AveR        ……AAGTVKLLVAEGGMGCGKSTFLGE……RLAAGAPLAIGIDDVQDADPESLHCLM……
NysRI       ……RDGRAGLVLLHGPAGMGKTSLLRS……RLMAQRPLVLVLDDVHWCDERSLAWID……

SAMR0484    ……RDQDGLLSEAERRVAVLAARGRTNRQISNELYITVSTVEQHLTRVYRKLDVKSRTDL
PikD        ……GPDVGLLSEAERRVAALAARGLTNRQIARRLCVTASTVEQHLTRVYRKLNVTRRADL
RapH        ……DTALTFLSRAERRVATLASAGYTNRAIAEKLFVTESTVEQHLTKVYRKLGVKHREDL
AveR        ……AATEALLSESEMRVATLAAGGNTNREIAGRLCVTVSTVEQHLTRVYRKLNITRRREL
NysRI       ……ASPLDMLTGMERTVADLAVTGASNRAIAEALFVTVRTIETHLTSVYRKLGVGGRAEL
```

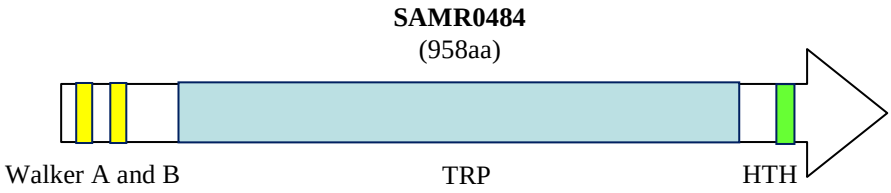


Fig. 45 Amino acids alignment of the N and C terminus regions of SAMR0484 and of other LAL regulators, already characterized. PikD comes from the pikromycin cluster of *S. venezuelae*, RapH from the rapamycin cluster of *S. hygroscopicus*, AveR from the avermectin cluster of *S. avermitilis* and NysRI from the nystatin cluster of *S. noursei*. The Walker A and B domain are underlined in yellow, the TRP domain in light blue and the HTH domain in green.

CHAPTER 2

Activation of a silent cluster and isolation of a novel bioactive macrolide of *Streptomyces ambofaciens* ATCC23877

The main objective of this project was to isolate and characterize the product of the type I PKS gene cluster in *S. ambofaciens*. The *in silico* analyses of the PKS gene products and of all the additional gene products showed that the cluster was potentially functional. In addition, we were able to predict the structure and the physicochemical properties of the polyketide backbone, which most likely would cyclize in a novel giant macrolide. However, to date this natural product had never been detected and isolated in the culture conditions used in the laboratory and it suggested that the cluster might be silent or that the level of expression was insufficient to synthesise the compound. By traditional phenotypical and biological screenings, only spiramycin and congocidine were characterized in *S. ambofaciens* ATCC23877, and very recently also the antibiotic kinamycin, using specific growth conditions (Pang *et al*., 2004). A hypothesis might be that the other secondary metabolite clusters are silent in the standard laboratory conditions or that their biological activity is hidden by the other three. In this later case, it would be interesting to obtain a mutant lacking of spiramycin, congocidine and kinamycin production to screen for other bioactive molecules.

To confirm the hypothesis of a silent cluster, the expression of the PKS genes was verified by RT-PCR in the wild type strain. Since the expected product was a macrolide, as spiramycin, we decided to choose media in which the production of spiramycin has already been observed experimentally. For this reason, cultures of the wt *S. ambofaciens* strain were grown in MP5 and HT (Hickey-Tresner) (see Methods in the article) and RNAs were extracted at different points of the growth to analyse the level of transcription of the PKS genes *sam*R0467 and *sam*R0477 (Fig. 46).

As shown in Fig. 46, for both PKS genes no expression, or at a very low level in the late stationary phase, was detected when the wt was grown in MP5. The same results were obtained from RNA samples extracted from cultures grown in HT. Therefore, this data strongly suggested that the cluster was silent at least under this laboratory conditions.

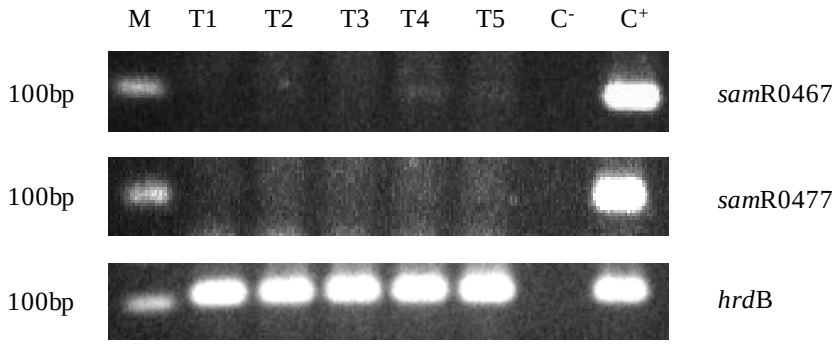


Fig. 46 Transcriptional analysis by RT-PCR of the PKS genes *sam*R0467 and *sam*R0477 in the wt *S. ambofaciens* ATCC23877. The strain was grown in MP5 medium for 84h. The expression of the PKS genes *sam*R0467 and *sam*R0477 was verified after 28 cycles of PCR at different points of the growth curve: exponential phase (T1 and T2), transition phase (T3), stationary phase (T4) and late stationary phase (T5). The gene *hrdB* encodes the major sigma factor and it is used as internal control. The positive control C^+ is obtained using the wt genomic DNA as template. The negative control C^- does not contain any template. Absence of DNA contamination was previously checked. cDNA were obtained as described in Methods in the article.

2.1 Activation of a silent cluster

In order to detect and isolate the predicted natural product of the cluster, the biosynthetic genes needed to be activated. Heterologous expression was not a feasible approach in this case because of the large size of the cluster (almost 150 kb). So far no such large cluster has been transferred in a host strain. Only recently, Wenzel *et al.* (2005) showed to be able to clone and heterologously express a large biosynthetic gene cluster of about 60 kb. At the same time Miao *et al*. (2005) used a modified BAC to integrate and express daptomycin cluster (128 kb) in *S. lividans*. Nevertheless, to obtain a vector containing our entire cluster would have been time consuming. Another possible strategy could have been to screen several media and growth conditions, as suggested by the OSMAC approach and which proved successful for kinamycin discovery, but except for an industrial company this approach is random, and also time and cost consuming.

Therefore, our strategy to trigger the expression of the PKS genes of the cluster was based on the manipulation of the regulatory genes, previously identified in the cluster and encoding a two component system and a LAL regulator. In fact, the already characterized LAL regulators were described as positive transcriptional regulators of the onset of the biosynthesis (Wilson *et al*., 2001) and overexpression of these genes results in increase of metabolite production (Kuscer *et al*., 2007). On the contrary, the two component system might play a positive or a negative role in the activation of biosynthetic genes. As positive regulators, their overexpression showed to enhance antibiotic production (Jnawali *et al*., 2008), likewise the LAL regulators.

On the basis of these observations, we decided to increase the expression of the regulatory genes, using a conjugative and integrative plasmid, pIB139 (Wilkinson *et al*. 2002) that allows constitutive gene expression (Fig. 47). The CDS of the genes were cloned under the control of

the strong and constitutive promoter *ermEp**. The ribosome binding site (RBS) has been previously modified *ad hoc* to be optimal for *Streptomyces* (Fig. 47). The vector exploits site specific recombination between the *att*P site, derived from the phage ΦC31 and carried by the vector, and the *att*B site of the chromosome to integrate stably in the genome.
As illustrated in the section Methods in the article, the constructs for the overexpression of *sam*R0468-469 and *sam*R0484 were obtained and introduced in the wt strain of *S. ambofaciens* by conjugation, thus obtaining the two mutant strains ATCC/OE468-9 and ATCC/OE484. As a control, to be sure that the vector does not influence secondary metabolite production, the wild type strain was conjugated with the vector alone, generating the mutant ATCC/pIB139.

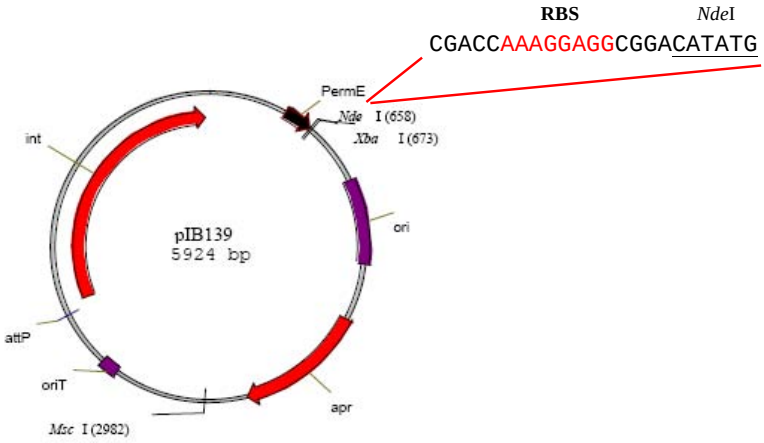


Fig. 47 Map of the plasmid pIB139. The RBS of the original plasmid was previously modified in this laboratory to have a typical *Streptomyces* RBS sequence (in red in the figure) upstream the ATG start codon. The unique restriction sites *Nde*I and *Xba*I are used for cloning the gene of interest.
Abbreviations: int for integrase; apr for apramycin resistance gene; oriT for origin of transfer; ori for origin of replication in *E. coli*.

To investigate the effect of the overexpression of the regulatory genes on the expression of biosynthetic genes of the cluster, transcriptional analyses by RT-PCR were carried out from samples obtained from the mutant and the control strains, grown in MP5 medium. Different points of the growth were chosen to analyse the expression of the biosynthetic genes *sam*R0467, *sam*R0465, *sam*R0477 and *sam*R0474. While no effect on the gene expression was observed for the control strain, as expected, only the constitutive expression of the LAL regulator SAMR0484 triggered the expression of the PKS genes (Fig. 2A in the article). Finally this silent cluster was activated, confirming the positive role of the LAL regulator in the onset of the biosynthesis. As expected, the figure shows that the level of expression of the gene *sam*R0484 is significantly higher in the mutant strain ATCC/OE484 than in the control strain. Also the overexpression of SAMR0468-9 functioned (data not shown), however no effect on the biosynthetic genes was obtained when the two component system was overexpressed, indicating a different role in the regulatory cascade of the cluster or that a specific signal is still required (data not shown).

The gene *sam*R0484 is expressed as well as in the control strain (Fig. 2A in of the article), even if at a lower level compared to the mutant strain, thus most likely this level of transcription is not sufficient to activate the biosynthetic genes and to allow the production of the unknown metabolite.

2.2 Detection and isolation of a novel metabolite

The transcriptional analyses demonstrated that the cluster might be activated by manipulating the regulatory system of the cluster. Biological assays of the mutant strain ATCC/OE484 against *B. subtilis* or *Micrococcus luteus*, as indicator strains, did not highlight any new activity associated to the activation of the cluster. For this reason, in order to identify the product synthesized by this cluster, the overexpression strategy needed to be combined to a comparative metabolic profiling approach. These analyses and the following structural elucidation of the molecules identified were performed through collaboration with the laboratory of Prof. Greg Challis, from the University of Warwick (UK), in accordance with ActinoGen Program (FP6). I went to Warwick to perform the first preliminary experiments that allowed identifying the novel compound.

The mutant ATCC/OE484 and the control strain ATCC/pIB139 were grown in MP5, in which the transcriptional analyses have been carried out, and the metabolic profiles of both strains were analyzed by Liquid Chromatography coupled to Mass Spectrometry (LC-MS). The majority of secondary metabolites are secreted in the medium, but since some compounds have been observed to accumulate intracellularly (e.g. undecylprodigiosin), both possibilities were checked for the mutant and the control strain. As described in Methods in the article, the mycelium and the supernatant have been extracted with methanol and analysed by LC-MC.

Interestingly, the extract of the ATCC/OE484 strain from the mycelium revealed the presence of two peaks, double charged, at *m/z* 673 and 680, not identified either in the control strain or in the samples from the supernatant (see Fig. 2B in the article). The real mass of these peaks, 1362 and 1376, and the molecular formula of the corresponding molecules, $C_{72}H_{132}NO_{22}$ and $C_{73}H_{134}NO_{22}$, are fairly consistent with our prediction made according to the *in silico* analyses (see Chapter 1). In addition, the molecules detected absorb at characteristic wavelength of 240 nm, which might be due to the presence of a diene, previously predicted too. The fragmentation pattern of these peaks pointed out the presence of several hydroxyl groups as well as a fragment of about 174 Da, which might correspond to the mycaminose residue (data not shown). Taken together, these results clearly indicated that the modular type I PKS cluster is responsible to synthesize at least two related compounds that differ from each other by a CH_2 group (14 Da of difference) and that so far seemed to match with our prediction. The compounds were identified only in the mycelium extracts, suggesting that they are accumulated intracellularly probably because of the important size of the molecules or because of their hydrophilic nature (fifthteen

hydroxyl groups) which can make difficult the crossing through the cell wall. A similar case has been observed for the linear polyene ECO-02301, containing also several hydroxyl groups, which was isolated from the mycelium of *S. aizunensis* NRRL B-11277 (McAlpine *et al*., 2005).

2.3 The detected metabolites are synthesised by the type I PKS cluster

From the preliminary results obtained by LC-MS analysis, the detected molecules clearly seemed to correspond to the compound predicted from the *in silico* analysis. To confirm that the molecules were directly linked to the cluster, a deletion of the first biosynthetic gene *sam*R0467 was made in the overexpressed mutant ATCC/OE484. The gene *sam*R0467 is 17 kb long and along with the loading module it contains three extension modules (Fig. 30), thus the deletion should dramatically affect the biosynthesis of the polyketide, not leading to the same final product, or it might completely prevent production of any compounds. In both cases, the two peaks at *m/z* 673 and 680 should no longer be detected.

The gene *sam*R0467 was replaced by a kanamycin resistance marker using the PCR-targeting protocol (Gust *et al*., 2003), as described in the section Methods in the article, obtaining the new strain ATCC/OE484/Δ467::kan. The mycelium of two independent double mutants was extracted with methanol and analysed by LC-MS, in parallel with the overexpressed mutant. The figure 48 shows the LC-MS chromatogram of the three mutant strains. The double charged peaks at *m/z* 673 and 680 are present only in the ATCC/OE484 strain, while both clones ATCC/OE484/Δ467::kan lack these compounds. This data provided clear evidences that the compounds directly derived from the analyzed type I PKS gene cluster.

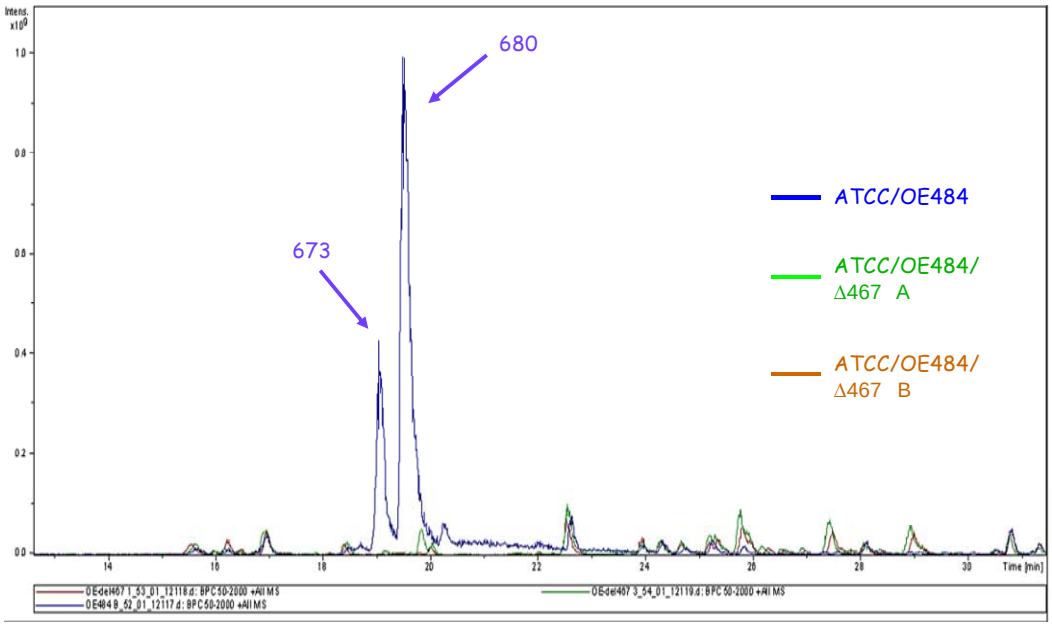


Fig. 48 LC-MS chromatogram of the overexpressed mutant ATCC/OE484 (in blue) and two independent clones ATCC/OE484/Δ467::kan (A in green and B in brown). The x axis corresponds to the retention time, between 19-20 min for the peaks 673 and 680. The y axis represents the value of absorbance.

The molecule synthesised by this new modular type I PKS cluster was predicted to be a macrolide containing as sugar moiety a mycaminose residue. The preliminary results from the fragmentation pattern of the compounds showed a peak whose mass is consistent with a mycaminose residue. Nevertheless, to confirm the presence of the sugar, the gene *sam*R0481, coding for the glycosyltransferase, was deleted in the mutant strain ATCC/OE484. The gene was replaced by an apramycin resistance gene, using the PCR-targeting protocol (see Methods in the article). Usually, glycosyltransferases transfer the activated sugar after the polyketide backbone has been released by the thioesterase. Hence, we were attending to detect a peak corresponding to the aglycone structure, a macrolactone ring.

The extracts of the mycelium of the new strain ATCC/OE484/Δ481::scar were analysed by LC-MS. Two new peaks were detected at the same retention time of the previous peaks (around 19-20 min), but at different *m/z*, 1212 and 1226, both being mono charged (Fig. 49). The two compounds also differ from each other for 14 Da, which corresponds to a CH_2 group. The real mass of the new compounds was calculated to be 1189 and 1203 Da (see Fig. 49). These masses correspond to the mass of the aglycone structure. Indeed, the difference between the final product masses (1362 and 1376) and the aglycone masses (1189 and 1203) is 173, which corresponds to the mycaminose moiety.

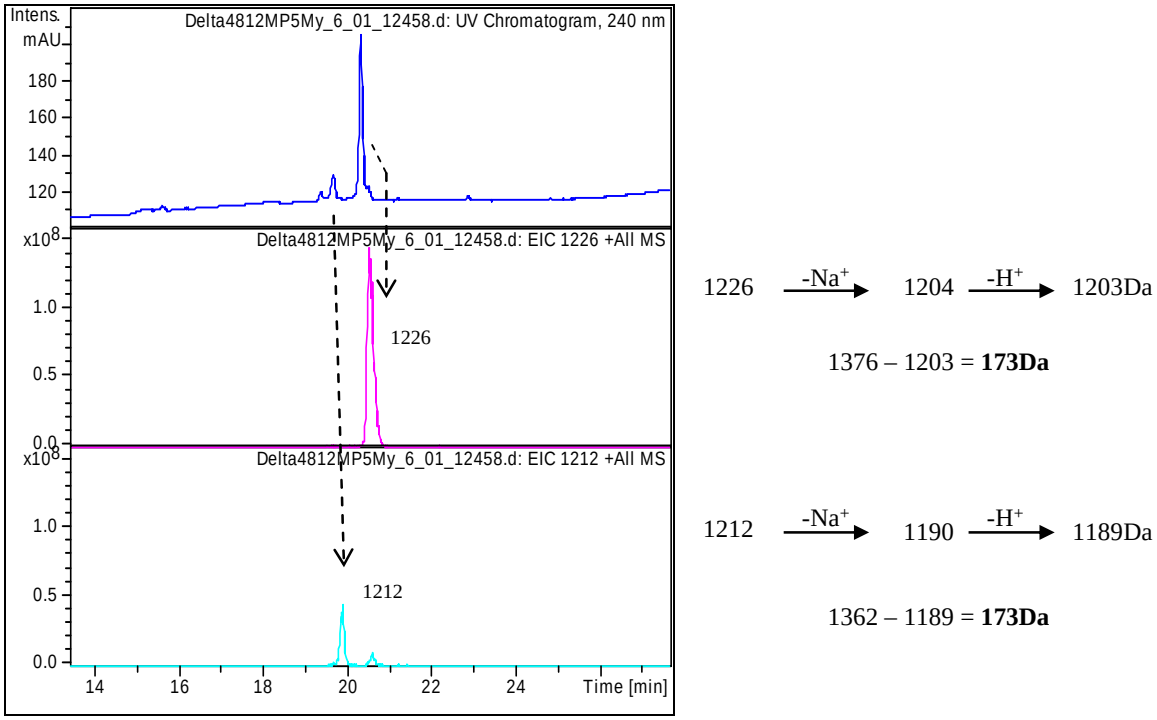


Fig. 49 LC-MS chromatogram of the double mutant ATCC/OE484/Δ481::scar. On the top, the UV spectrum at 240 nm showed the two new peaks. On the bottom, the MS spectrum of the peak 1226 and the peak 1212, which correspond to the masses of the aglycone structure of the final products.

2.4 Structural elucidation of the novel compounds

The mutants in the biosynthetic and the glycosyltransferase genes definitely confirmed that the two molecules identified by comparative metabolic profiling in the strain ATCC/OE484 were synthesized by the modular PKS cluster. Moreover, the preliminary results seemed to validate our predictions, for example the presence of a sugar moiety. Consequently, the following step was to isolate and purify those compounds by semi-preparative HPLC and to elucidate their structure by NMR. The later analysis was performed by Prof. Greg Challis and Dr. Lijiang Song.

The strain ATCC/OE484 was grown on MP5 medium and the mycelia were extracted by methanol (see Methods in the article). The choice of solid medium depended on the experimental observation that secondary metabolite production is higher than in liquid conditions. The extracts were injected in a reverse phase semi-preparative HPLC column and the peaks corresponding to the two compounds were collected separately (Fig. 50). The samples were then lyophilised and analysed by NMR techniques using different solvent (see Fig. 1-13 of Appendix).

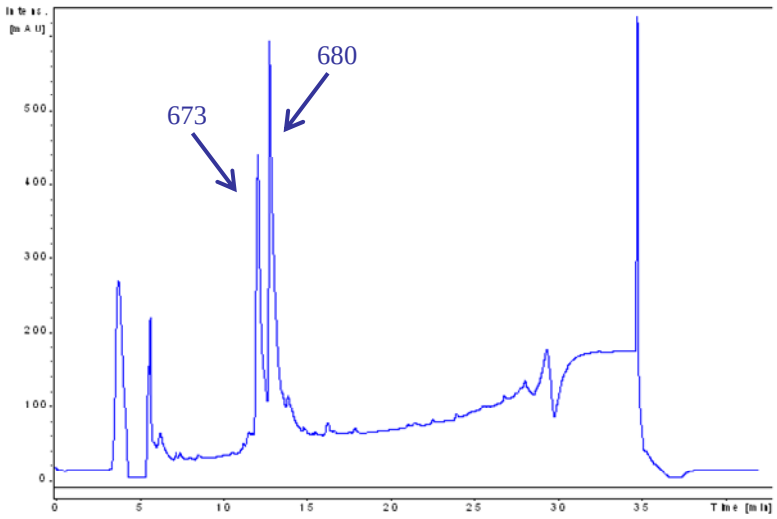


Fig. 50 HPLC chromatogram of the mycelium extracts of ATCC/OE484 grown on MP5. The absorbance was monitored at 240 nm. The peaks 673 and 680 are highlighted with an arrow.

From the NMR analysis it emerged that the cluster was responsible to produce four different forms of an unexpected 50-membered macrolide, hereafter referred as sambomycin A, B, C and D (Fig. 51). These kinds of metabolites have been isolated for the first time in a microorganism, to the best of our knowledge. No previous characterized macrolides own such a giant lactone ring, indeed the cyclization occurs between the first and the last precursor of the polyketide backbone, generating a 50-membered.

Sambomycin A and B are isomers; they share the same mass (1376 Da) and the same molecular formula ($C_{73}H_{134}NO_{22}$), even though they do not have the same structural formula. Likewise, sambomycin C and D share the same mass (1362 Da) and the same molecular formula ($C_{72}H_{132}NO_{22}$). The sambomycin A/B isomers corresponded to the previous double charged peak 680, identified by LC-MS, while the sambomycin C/D isomers to the peak 673.

The absolute stereochemistry of the stereocenters was assigned based on the bio-informatics analysis; however for some of them it was not possible to determine it (see Fig. 51). The structural elucidation confirmed most of our previous predictions, providing also new additional information on the structure and on its biosynthesis. A spontaneous cyclization occurred inside the macrolactone ring between the keto group of C-3 and the hydroxyl group of C-7 (Fig. 51). This reaction is quite common for these kinds of metabolites [e.g., the polyene macrolides pimaricin or candicidin (Martin and Aparicio, 2009)]. A hydroxyl group is inserted at the position C-28, certainly due to the action of a cytochrome P450. The sugar moiety, which confirmed to be a mycaminose residue, is instead attached by the glycosyltransferase to the C-5 of the macrolactone ring.

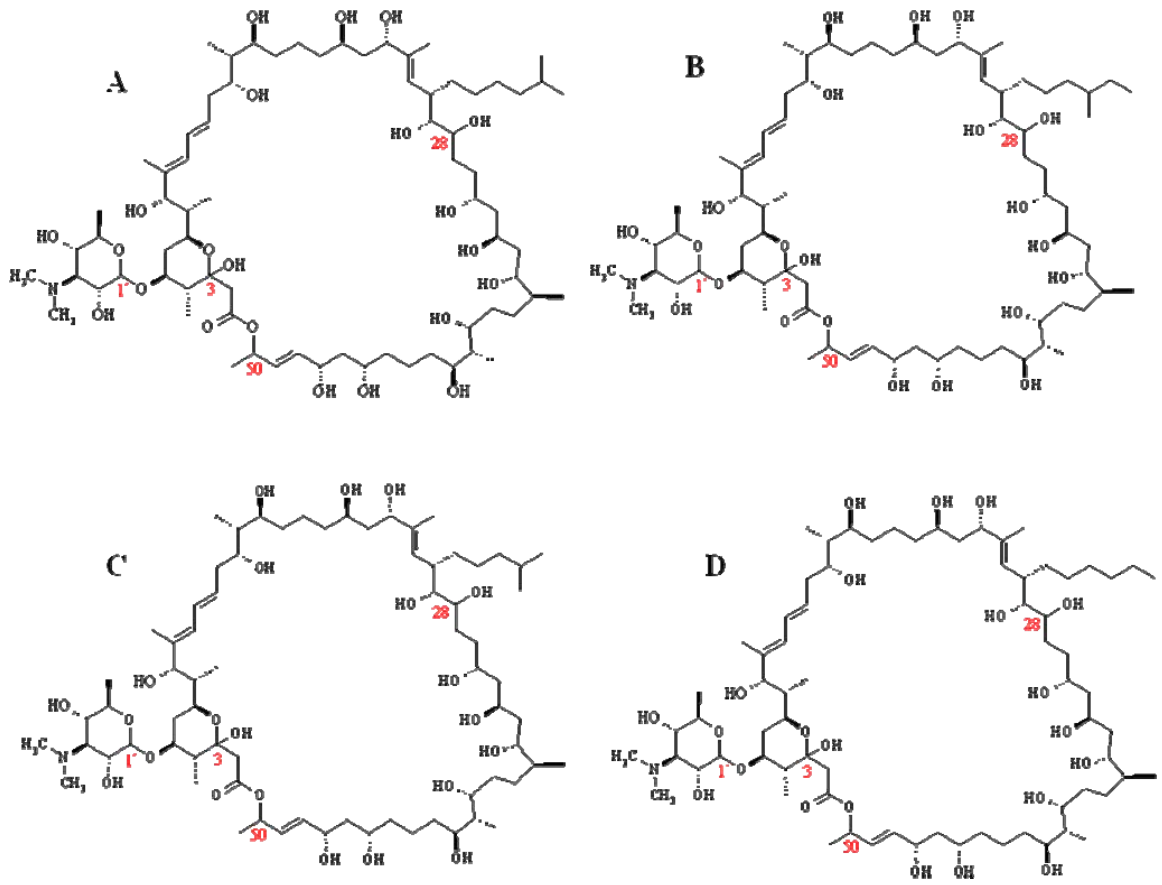
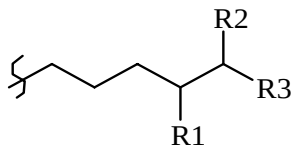


Fig. 51 Structure of the four sambomycin macrolides (A, B, C and D). The absolute stereochemistry of the stereocenters in the structure is shown, except for C-3, C-28, C-50 and the C-1′ of the sugar residue.

More interestingly, the structural determination gave new insights on the nature of the precursor loaded by the AT13 domain and attached on the C-26, which was unpredictable by the *in silico* analyses (Chapter 1). This uncommon extender unit is an alkyl chain represented in Fig. 52 and according to the substituents of the R groups it gives rise to the four sambomycin derivatives.



R1=H ; R2=R3=Me → sambomycin A
R1=R3=Me ; R2=H → sambomycin B
R1=Me ; R2=R3= H → sambomycin C
R1=R2=H ; R3=Me → sambomycin D

Fig. 52 Structure of the alkyl chain loaded by the AT13 domain and the different R substituents, based on the NMR analysis.

2.5 Biological properties of the macrolide sambomycin

We isolated and structurally characterized a novel group of compounds, four forms of a 50-membered macrolide, synthesized by *S. ambofaciens* ATCC23877. Macrolides are usually used as antibacterial agents, in particular against Gram-positive bacteria, for example erythromycin and spiramycin, but they are also known to be effective in antifungal therapy (e.g. the polyene macrolides nystatin or candicidin). My thesis project aimed to discover a novel bioactive molecule; therefore investigation of possible interesting biological properties associated to sambomycin was another important objective.

Sambomycin A/B and C/D were initially tested against *B. subtilis* and *M. luteus*. Both indicator strains were sensitive to the macrolides, when at least 1µg of the molecule was used (Fig. 53).

Below this quantity (0.25µg), no activity was clearly observed. In addition, sambomycin C/D appeared more effective than sambomycin A/B at the same concentration, which indicates that the substituent of the alkyl chain may have some influences on the biological properties. However for both form of isomers, the inhibition zones are not significantly large considering the quantity used for the assay. This could be explained by a low diffusion in the medium or by a low efficiency of the biological activity.

In order to extend the spectrum of action of the molecules, we performed other bioassays against several Gram-positive (*Enterococcus faecalis*, *Staphylococcus aureus*, *Mycobacterium smegmatis*) and Gram-negative bacteria (*E. coli*, *P. aeruginosa*), as well against fungi (*Fusarium oxysporum* and *Aspergillus fumigatus*) and pathogenic yeast (*Candida albicans*). Only the sambomycin C/D isomers were tested and proved to be active against Gram-positive bacteria, except for *M. smegmatis*, as generally observed for other macrolides, probably due the wall composition of *Mycobaterium*. The IC90 of sambomycin C/D for *B. subtilis*, *E. faecalis* and *S. aureus* was determined (see Table 1 in the article).

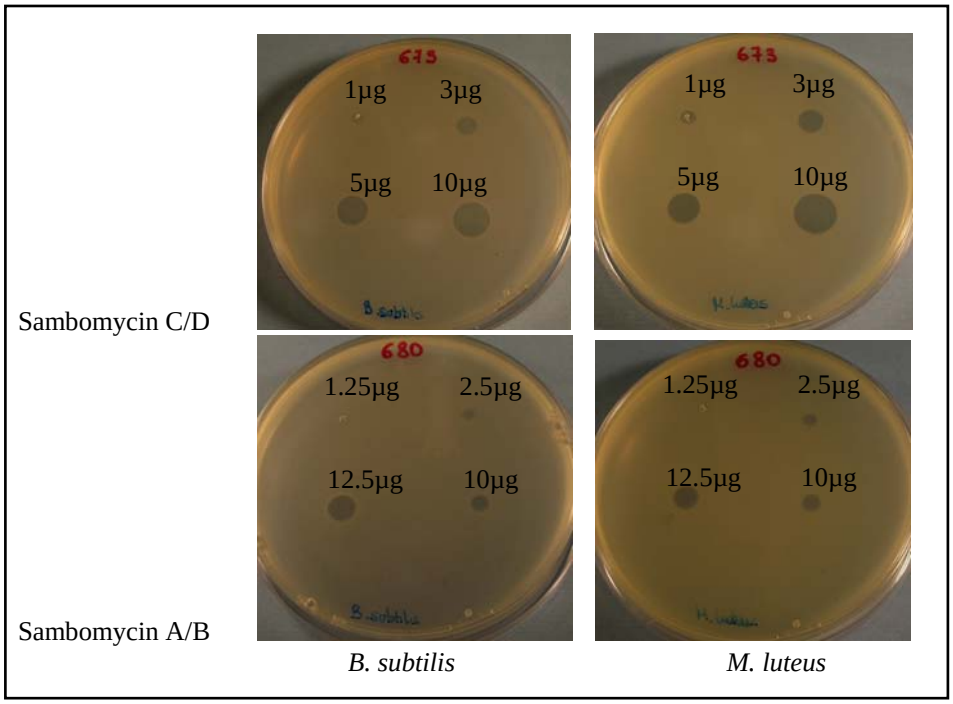


Fig. 53 Bioassay against *B. subtilis* and *M. luteus* using different concentration of sambomycin A/B and C/D.

Marine macrolides, often characterized by a gigantic structure as sambomycin, have been recently discovered to endow potent antitumor properties, for example IB-96212, a 26-membered macrolide from *Micromonospora* sp. L-25-ES25-008 (Olano *et al.*, 2009). Interestingly sambomycin macrolides showed too to have a relevant activity against the human colon adenocarcinoma cell line HT29 comparable to that of doxorubicine, an anticancer agent already used in human therapy (Table 1 in the article). The toxicity of sambomycin compounds was tested in comparison with the control doxorubicine: the macrolides showed an even lover cytotoxicity than the control (Table 1 in the article) which makes hoping for future medical exploitations.

Letter in preparation for *Nature*

"Discovery of a novel macrolide in *Streptomyces ambofaciens* by awaking a sleeping giant"

Laureti L., Lijiang S., Corre C., Huang S., Leblond P., Challis G. and Aigle B.

MANUSCRIPT TEXT

Natural products are of utmost importance because of their broad spectrum of applications in human and veterinary therapy and in agriculture (1,2). Genome mining approach appears to be a novel and powerful avenue to new natural bioactive metabolite discovery. Indeed, analysis of microbial genome sequences has revealed the presence of numerous secondary metabolite gene clusters. This disclosed the tremendous untapped potential of microorganisms in producing bioactive molecules even for the actinomycetes which are responsible for nearly 70% of the known microbial natural products (3,4,5). However, most of the clusters are dormant in standardized laboratory conditions because the appropriate chemical or physical signals necessary for their induction are unknown. Their activation thus represents the bottleneck for isolating the corresponding products.

Here we report the activation of a silent type I polyketide synthase (PKS) gene cluster identified from the genome mining of *Streptomyces ambofaciens* leading to the discovery of sambomycin, one of the largest macrolides ever described. The expression of the biosynthetic genes have been triggered by manipulating a pathway-specific regulatory gene belonging to the Large ATP-binding of the LuxR (LAL) family and located within the cluster. Subsequently, based on the structural prediction, we carried out comparative metabolic profiling using a mutant overexpressing the LAL regulator encoding gene allowing detecting the expected metabolite. The structure elucidation revealed a 50-membered macrolide with several striking features (i.e. an atypical extender unit and a peculiar cyclisation reaction). Interestingly, sambomycin compounds showed promising antiproliferative activities.

Streptomyces ambofaciens ATCC23877 was known to produce two antibiotics, the macrolide spiramycin, used for a long time in human therapy, and the pyrrole-amide congocidine (6). Mining its linear partial sequenced chromosome has unveiled several secondary metabolite clusters located principally in the arms (7) and allowed the identification of the congocidine cluster (8). For few other clusters, the corresponding product has been isolated and/or characterized (9,10), but for most of the other clusters the product remains unknown. Our attention had been focused on a large type I PKS gene cluster. Polyketides are one of the major classes of natural products derived from secondary metabolite clusters, endowing a wide range of biological properties (e.g., antimicrobial, anticancer, immunosuppressant) (11). The cluster, located in the right arm at 500 kb of the end of the chromosome, is proposed to contain 25 genes including 9 PKS encoding genes (**Fig. 1a** and **Supplementary Table 1**). Its size is about 150Kb

(124 kb for the PKS genes), making this cluster one of the largest type I PKS gene clusters ever described. *In silico* analysis of the PKS gene with SEARCHPKS (12) revealed a total of 25 modules and of 112 enzymatic domains and allowed to establish the order of the chain assembly line (from *sam*R0467 to *sam*R0465, and from *sam*R0477 to *sam*R0474, **Fig. 1b**). All the enzymatic domains, but one, harbour a putative functional active site. Based on the acyltransferase (AT) domain sequences (13), we forecasted the substrates incorporated in the polyketide: the starter unit, a propionate, and the extender units, 16 malonyl-CoA, 7 methylmalonyl-CoA and an unknown precursor, loaded by the AT13 domain (**Supplementary Fig. 1**). The chirality of α-substituent and β-hydroxyl group of each precursor was established according to the conserved motifs inside the ketoreductase (KR) and when necessary inside the enoylreductase (ER) domain (14,15). The chirality of the polyketide produced by this cluster is mostly A1 type and B1 type, with just two exceptions for KR18, which is an A2 type, and KR21, a B2 type (**Supplementary Fig. 2**). The last ketoreductase domain (KR25) is predicted to be non-functional as the tyrosine and asparagine of the catalytic triad are not conserved (16). Taken together, these data enabled to predict the linear backbone of the product (**Fig. 1c**) and consequently the approximate mass (~1127Da) and the estimated molecular formula ($C_{61}H_{107}O_{18}$). Phylogenetic analyses of the thioesterase domain predicted a cyclized PKS product, giving rise to a lactone ring (James McAlpine, pers. com.). However, it was not possible to predict where the cyclization occurs.

A glycosyltransferase gene (*sam*R0481) has been identified inside the cluster, suggesting that the final structure would be a macrolide. The analysis of *sam*R0481 with SEARCHGTr (17) proposed an amino sugar, as desosamine or mycaminose, for the sugar incorporation. In addition, a set of five genes of the cluster (**Fig. 1b**) were predicted to be potentially involved in the conversion of an α-D-glucose-1-phosphate into a TDP-D-mycaminose residue (**Supplementary Fig. 3**). Therefore, the possible mass and molecular formula of the macrolide would be ~1318Da and $C_{69}H_{124}NO_{22}$, respectively. Searches in the chemical databases indicated that this structure would be novel.

Most of the clusters identified from the genome analysis are cryptic *i.e.* not associated with the production of a known metabolite. A likely explanation can be the absence or poor expression of these clusters in the laboratory conditions (18). Therefore, we first checked the expression of the biosynthetic genes in the wt strain of *S. ambofaciens* under two growth conditions known for the production of the macrolide spiramycin (19). RT-PCR analysis showed an absence of expression (or a very low level of expression) of the biosynthetic genes all along the growth (**Fig. 2a**). We noted that the gene cluster contained three regulatory genes: *sam*R0468 and *sam*R0469 whose

deduced products form a two component system and *sam*R0484 which encodes a putative LAL regulatory protein (**Supplementary Table 1**). Therefore, to activate this silent metabolic pathway, the strategy was based on the manipulation of these regulatory genes. Members of the LAL family were previously described as positive regulators of polyketide biosynthesis, such as PikD for the pikromycin production or RapH for rapamycin production (20,21,22). We thus decided to overexpress *sam*R0484 in the wt strain. The gene was cloned under the strong and constitutive promoter *ermEp** within the conjugative and integrative vector pIB139 (23). *S. ambofaciens* was transformed with the recombinant vectors and integration took place at the *att*B site of the chromosome. As control, the wt strain was also transformed with the vector alone. Comparative transcriptional analyses by RT-PCR confirmed the positive role of the LAL protein in triggering the expression of the cluster. Indeed, a significant level of transcription of the PKS genes was detected all along the growth in the clone overexpressing *sam*R0484 (**Fig. 2b**). Conversely, no or weak transcription was detected in the control strain. The same approach was applied for the two component system encoding genes. However, no effect was observed on the expression of the biosynthetic genes (data not shown).

In order to detect the predicted metabolite, comparative metabolic profiling by LC-MS was performed between the control (ATCC/pIB139) and the LAL regulator overexpressing strain (ATCC/OE484). The LC-MS spectrum of the mycelium extract of ATCC/OE484 revealed two MS peaks, double charged, at *m/z* of 673 and 680 (positive ion mode) with a molecular formula of $C_{72}H_{132}NO_{22}$ and $C_{73}H_{134}NO_{22}$, respectively, which were not present in the control strain (**Fig. 2b**). The mass of the two peaks (1362 and 1376) and their molecular formulas were consistent with the predictions from the *in silico* analyses. The mass difference could be due to a $-CH_2$ group. These peaks were absent in the supernatant extracts. To confirm that these metabolites were directly linked to cluster, the first PKS gene, *sam*R0467, was deleted in the mutant ATCC/OE484. By LC-MS analysis, the previous peaks were no longer detected in the new mutant strain (data not shown), thus confirming that the two compounds derived from the cluster.

The two metabolites were partially purified from the concentrated methanol extract using semi-preparative HPLC on a reverse phase column. Both fractions were evaporated under reduced pressure and lyophilised. High resolution ESI-TOF-MS analyses established their molecular formulae as $C_{73}H_{134}NO_{22}$ (for the compounds named sambomycins A and B) and $C_{72}H_{132}NO_{22}$ (for the compounds named sambomycins C and D) ($[M+H]^+$ calculated for $C_{73}H_{134}NO_{22}$: 1376.9392, found: 1376.9391; $[M+H]^+$ calculated for $C_{72}H_{132}NO_{22}$: 1362.9236, found: 1362.9232).

^{1}H, ^{13}C, COSY, NOESY, TOCSY, HSQC and HMBC NMR spectra (in d_4-MeOH and d_6-DMSO) established the planar structures of sambomycin A, B, C and D (**Figure 3**). The absolute stereochemistry of the stereocentres in the macrolide ring of sambomycin A, B, C and D was predicted by bioinformatics analyses using the method of (14,15). The stereochemistry at C-3, C-28 and C-50, as well as the R^1 group in the C-26 alkyl chain, could not be predicted using these analyses. The absolute stereochemistry of the stereocentres in the mycaminose residue of sambomycin A, B, C and D was predicted using bionformatics analyses based on the known pathway for TDP-D-mycaminose biosynthesis from D-glucose-1-phosphate (24). The stereochemistry of C-1′ could not be predicted using these analyses.

Macrolides show a broad range of biological properties such as antibacterial, antifungal and antiproliferative. The latter has been mainly found associated with large macrolides. The sambomycin compounds showed to endow antimicrobial activities against Gram positive bacteria except *Mycobacterium smegmatis* (**Table 1a**) but not against Gram negative bacteria and fungi. More interestingly, they presented an antiproliferative activity against the human adenocarcinoma (HT29) cell lines with an IC50 value comparable to the therapeutic drug doxorubicin (**Table 1b**). In addition, when tested on ovary sane cell line (CHO-K1) from an adult Chinese hamster, they showed to be less cytotoxic than doxorubicin (**Table 1c**).

In conclusion, we have isolated and characterized the largest macrolide ever described in *Streptomyces*. The structural characterization has revealed an unusual alkyl chain precursor probably derived from fatty acid metabolism. At least four forms of sambomycin are produced according to the type of alkyl chain incorporated. The structure also revealed a novel lactonization mechanism. Interestingly, in addition to antibacterial activities, promising antiproliferative activities have been associated to sambomycin.

The discovery of sambomycin has been successfully achieved by ectopic expression of a pathway-specific LAL regulator allowing awaking a sleeping type I PKS gene cluster. A similar approach was described in *Aspergillus nidulans* for a PKS-NRPS hybrid gene cluster showed to be responsible for the production of aspyridone (25) and we proved for the first time its feasibility and efficacy in *Streptomyces*. Hence, this represents a promising strategy to activate the numerous cryptic pathways identified in actinomycetes, the main producers of microbial natural products.

REFERENCES

[1]Berdy J. (2005). "Bioactive microbial metabolites." *J Antibiot (Tokyo)* **58**(1): 1-26.

[2]Newman D. J. and Cragg G. M. (2007). "Natural products as sources of new drugs over the last 25 years." *J Nat Prod* **70**(3): 461-77.

[3]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., Bateman A., Brown S., Chandra G., Chen C. W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C. H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S., Rabbinowitsch E., Rajandream M. A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B. G., Parkhill J. and Hopwood D. A. (2002). "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)." *Nature* **417**(6885): 141-7.

[4]Oliynyk M., Samborskyy M., Lester J. B., Mironenko T., Scott N., Dickens S., Haydock S. F. and Leadlay P. F. (2007). "Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338." *Nat Biotechnol* **25**(4): 447-53.

[5]Udwary D. W., Zeigler L., Asolkar R. N., Singan V., Lapidus A., Fenical W., Jensen P. R. and Moore B. S. (2007). "Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica." *Proc Natl Acad Sci U S A* **104**(25): 10376-81.

[6]Pinnert-Sindico S. (1954). "[Not Available]." *Ann Inst Pasteur (Paris)* **87**(6): 702-7.

[7]Choulet F., Aigle B., Gallois A., Mangenot S., Gerbaud C., Truong C., Francou F. X., Fourrier C., Guerineau M., Decaris B., Barbe V., Pernodet J. L. and Leblond P. (2006). "Evolution of the terminal regions of the Streptomyces linear chromosome." *Mol Biol Evol* **23**(12): 2361-9.

[8]Juguet M., Lautru S., Francou F. X., Nezbedova S., Leblond P., Gondry M. and Pernodet J. L. (2009). "An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in Streptomyces ambofaciens." *Chem Biol* **16**(4): 421-31

[9]Pang X., Aigle B., Girardet J. M., Mangenot S., Pernodet J. L., Decaris B. and Leblond P. (2004). "Functional angucycline-like antibiotic gene cluster in the terminal inverted repeats of the Streptomyces ambofaciens linear chromosome." *Antimicrob Agents Chemother* **48**(2): 575-88.

[10]Barona-Gomez F., Lautru S., Francou F. X., Leblond P., Pernodet J. L. and Challis G. L. (2006). "Multiple biosynthetic and uptake systems mediate siderophore-dependent iron acquisition in Streptomyces coelicolor A3(2) and Streptomyces ambofaciens ATCC 23877." *Microbiology* **152**(Pt 11): 3355-66.

[11]Staunton J. and Weissman K. J. (2001). "Polyketide biosynthesis: a millennium review." *Nat Prod Rep* **18**(4): 380-416.

[12]Yadav G., Gokhale R. S. and Mohanty D. (2003a). "SEARCHPKS: A program for detection and analysis of polyketide synthase domains." *Nucleic Acids Res* **31**(13): 3654-8.

[13]Yadav G., Gokhale R. S. and Mohanty D. (2003b). "Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases." *J Mol Biol* **328**(2): 335-63

[14]Keatinge-Clay A. T. (2007). "A tylosin ketoreductase reveals how chirality is determined in polyketides." *Chem Biol* **14**(8): 898-908

[15]Kwan D. H., Sun Y., Schulz F., Hong H., Popovic B., Sim-Stark J. C., Haydock S. F. and Leadlay P. F. (2008). "Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases." *Chem Biol* **15**(11): 1231-40.

[16]Reid R., Piagentini M., Rodriguez E., Ashley G., Viswanathan N., Carney J., Santi D. V., Hutchinson C. R. and McDaniel R. (2003). "A model of structure and catalysis for ketoreductase domains in modular polyketide synthases." *Biochemistry* **42**(1): 72-9.

[17]Kamra P., Gokhale R. S. and Mohanty D. (2005). "SEARCHGTr: a program for analysis of glycosyltransferases involved in glycosylation of secondary metabolites." *Nucleic Acids Res* **33**(Web Server issue): W220-5.

[18]Challis G. L. (2008) "Mining microbial genome for new natural products and biosynthetic pathways." *Microbiology* **154**(Pt 6): 1555-69.

[19]Pernodet J. L., Alegre M. T., Blondelet-Rouault M. H. And Guérineau M. (1993) "Resistance to spiramycin in *Streptomyces ambofaciens*, the producer organism, involves at least two different mechanisms." *J Gen Microbiol* **139**(5): 1003-11.

[20]Wilson D. J., Xue Y., Reynolds K. A. and Sherman D. H. (2001). "Characterization and analysis of the PikD regulatory factor in the pikromycin biosynthetic pathway of Streptomyces venezuelae." *J Bacteriol* **183**(11): 3468-75.

[21]Anton N., Mendes M. V., Martin J. F. and Aparicio J. F. (2004). "Identification of PimR as a positive regulator of pimaricin biosynthesis in *Streptomyces natalensis*" *J Bacteriol* **186**(9): 2567-75.

[22]Kuscer E., Coates N., Challis I., Gregory M., Wilkinson B., Sheridan R. and Petkovic H. (2007). "Roles of rapH and rapG in positive regulation of rapamycin biosynthesis in Streptomyces hygroscopicus." *J Bacteriol* **189**(13): 4756-63.

[23]Wilkinson C. J., Hughes-Thomas Z. A., Martin C. J., Bohm I., Mironenko T., Deacon M., Wheatcroft M., Wirtz G., Staunton J. and Leadlay P. F. (2002). "Increasing the efficiency of heterologous promoters in actinomycetes." *J Mol Microbiol Biotechnol* **4**(4): 417-26.

[24]Melançon C. E. III, Yu W. L. and Liu H. W. (2005). "TDP-mycaminose biosynthetic pathway revised and conversion of the desosamine pathway to mycaminose pathway with one gene." *J Am Chem Soc* **127**(35): 12240-1.

[25]Bergmann S., Schumann J., Scherlach K., Lange C., Brakhage A. A. and Hertweck C. (2007). "Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans." *Nat Chem Biol* **3**(4): 213-7.

[26]Hanahan D. (1983). "Studies on transformation of *Escherichia coli* with plasmids." *J Mol Biol* **166**: 557-580.

[27]MacNeil D. J., Gewain K. M., Ruby C. L., Dezeny G., Gibbons P. H. and MacNeil T. (1992). "Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector." *Gene* **111**: 61-68.

[28]Datsenko K. A. and Wanner B. L. (2000). "One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products." *Proc Natl Acad Sci USA* **97**: 6640-45.

[29]Gust B., Challis G. L., Fowler K., Kieser T. and Chater K. F. (2003). "PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin." *Proc Natl Acad Sci USA* **100**: 1541-46.

FIGURE LEGENDS:

Figure 1: The sambomycin cluster
A) Organization of the sambomycin cluster. The limits of the cluster have been fixed based on the putative function of the genes. **B)** Modular structure of the PKS genes and domain organization of the 25 modules, according to the SEARCHPKS prediction (Yadav *et al*., 2003a). The order of the genes corresponds to the order of interaction. Module 1 corresponds to the loading module. KS^{Q}, β-ketoacyl synthase in which the active site cysteine residue is replaced by glutamine; KS, β-ketoacyl synthase; AT, acyltransferase; ACP, acyl carrier protein; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; TE, thioesterase. The inactive KR25 domain is crossed. **C)** Prediction of the linear structure of the polyketide backbone. The R substituent of the precursor loaded by AT13 is unknown.

Figure 2: Activation of the silent cluster and detection of the predicted metabolite.
A) Transcriptional analysis by RT-PCR of the control ATCC/pIB139 and the mutant strain ATCC/OE484, grown in MP5 medium. The expression of four biosynthetic genes (*sam*R0467, 0465, 0477, 0474), together with the expression of the regulatory gene *sam*R0484, were analysed by RT-PCR using 4µg of total RNA. The gene *hrdB*, coding for the major sigma factor, has been used as internal control. The experiments have been repeated more than three times. T1= exponential phase; T2= transition phase; T3= stationary phase. **B)** LC-MS profile of the control ATCC/pIB139 and the mutant strain ATCC/OE484. The two framed peaks, with a retention time of 19 and 20 min, are double charged with a *m/z* of 673 and 680.

Figure 3: Sambomycin structure

Table 1: Biological activities of the sambomycin compounds.
IC90 indicates the concentration needed to inhibit the growth of 90% of cell population. HT29, human colon adenocarcinoma cell line; CHO-K1, adult Chinese hamster ovary sane cell line. ND, not determined the IC90 for sambomycin A/B, but it showed to have antibacterial activity. Vancomycin and doxorubicin are the control references.

Supplementary Table 1: List of the gene within the sambomycin cluster.
Principle characteristics of each gene belonging to the cluster. For the putative functions, the results were obtained with BlastP program and the best hits are indicated, taking into account a "head to tail" alignment (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Supplementary Figure 1: Amino acid alignment of the AT domains.
The amino acids of the active site highly conserved are marked with an asterisk (Yadav *et al*., 2003b). The amino acids boxed are involved in the specificity of the substrates; in particular the amino acid underlined in green. The arginine underlined in yellow interacts with the carboxyl group of the precursors. The serine residue, underlined in blue, is directly involved in the catalytic attachment of the substrate. The consensus sequences for malonyl-CoA and methylmalonyl-CoA substrate are also given. The substrate for AT13 is supposed to be an uncommon one because the one of conserved amino acid (circled) is different from the ones predicted for malonyl or methylmalonyl-CoA.

Supplementary Figure 2: Amino acid alignments of the KR domains.
The consensus sequence at the N-termini was proposed to bind to the NAD(P)H (Aparicio *et al*., 1996). The amino acids of the catalytic triad are marked with an asterisk. The arrows indicate the residue specific for the B and A-type alcohol stereochemistry, respectively. On the right of the alignment, for each KR domain the configuration of α-substitute and β-hydroxyl groups derived from the precursors are given, according to the nomenclature proposed by Keatinge-Clay (2007). A1= 2R, 3S; A2=2S, 3S; B1=2R, 3R; B2= 2S, 3R; C1=2R.

Supplementary Figure 3: TDP-D-mycaminose biosynthetic pathway.
Schematic representation of TDP-D-mycaminose biosynthetic pathway from an α-D-glucose-1-phosphate. For each enzymatic step, the gene product is indicated in bold. In brackets, the percentages of identity/similarity with the spiramycin enzymes are indicated (Karray *et al.*, 2007). TTP, thymidine triphosphate; TDP, thymidine diphosphate; NAD^+, nicotinamide adenine dinucleotide; PMP, pyridoxamine 5´-phosphate; SAM, *S*-adenosyl-L-methionine.

Supplementary Table 2: List of strains, plasmids and BAC used in this work
[a] *bla*, ampicillin resistance gene; *neo*, kanamycin resistance gene; *aac(3)IV*, apramycin resistance gene; *oriT*, origin of transfer; *aadA*, spectinomycin and streptomycin resistance gene; *gam*, inhibitor of the host exonuclease V; *bet*, single-stranded DNA binding protein; *exo*, exonuclease promoting recombination along with *bet*; *cat*, chloramphenicol resistance gene; $attP_{\phi C31}$, ϕC31 attachment site from the ϕC31 phage; $int_{\phi C31}$, integrase gene of ϕC31.

Supplementary Table 3: Primers used in this work
The sequences underlined correspond to the *Xba*I or *Nde*I site. The sequences in bold are homologous to the extremities of the gene to be deleted (*sam*R0467).

FIGURES

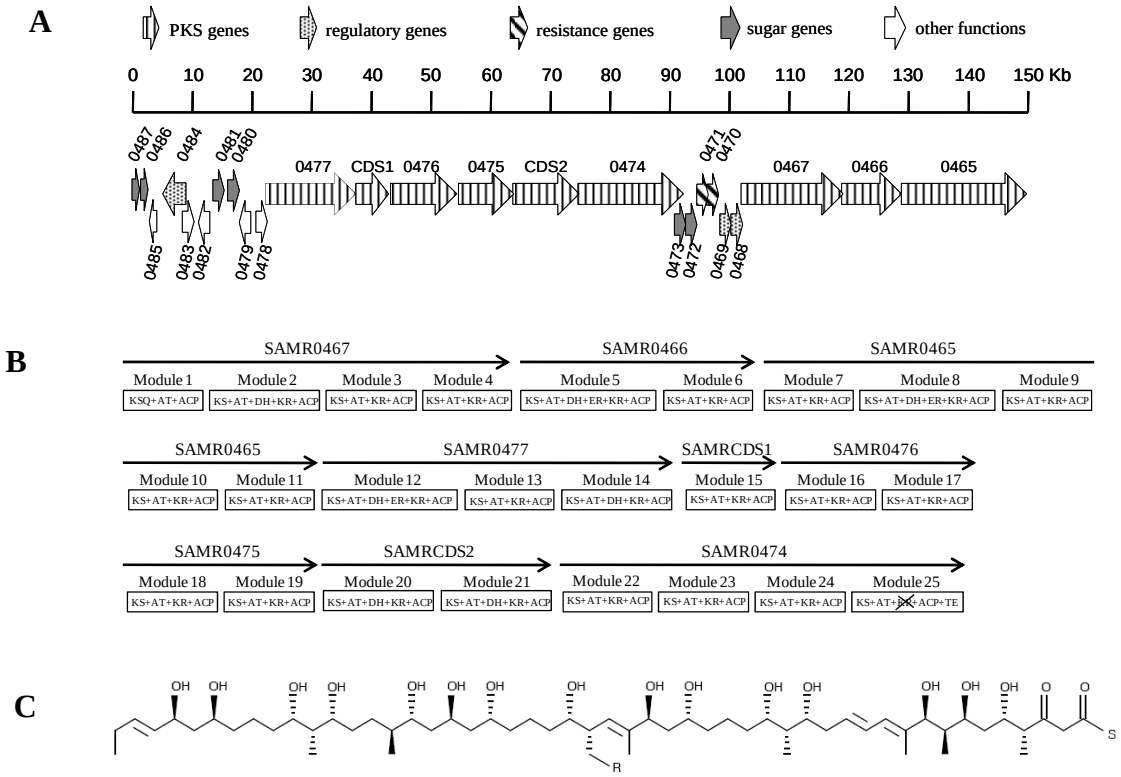


Figure 1
The sambomycin cluster

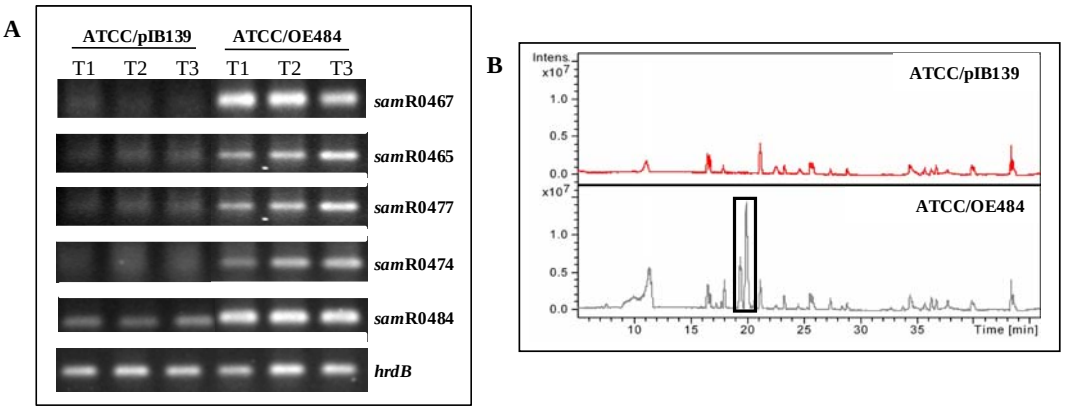


Figure 2
Activation of the silent cluster and detection of the predicted metabolite.

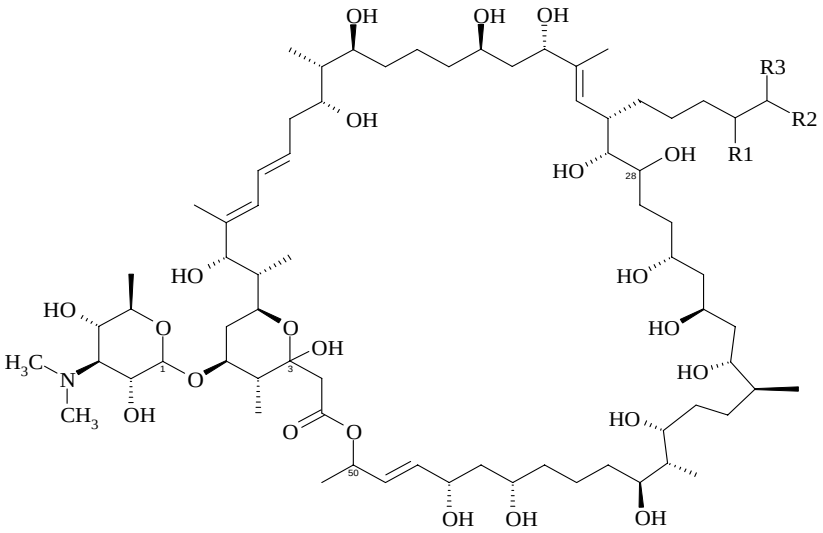


Figure 3
The structure of sambomycin

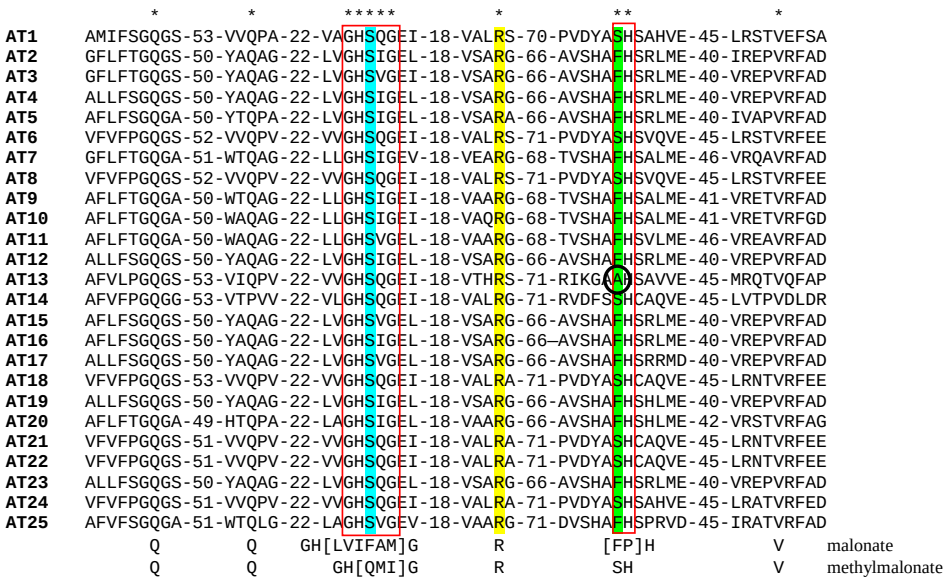
Organism	Sambomycin A/B	Sambomycin C/D	Vancomycin	Doxorubicin
Antibacterial activity (IC90, µg/ml):				
Bacillus subtilis BGSC 1A72	ND	33,53 +/- 0,86	0,25-1	
Enterococcus faecalis LG40	ND	8,65 +/- 0,49	0,25-1	
Staphylococcus aureus LG21	ND	33,95 +/- 0,16	0,25-1	
Antiproliferative activity (IC50, µM):				
HT29	1.77 +/- 0.04	1.74 +/- 0.04		1.32 +/- 0.08
Cytotoxicity (IC50, µM):				
CHO-K1	8.47 +/- 0,67	8.46 +/- 0,52		1.99 +/-0.25

Table 1
Biological activities of the sambomycin compounds

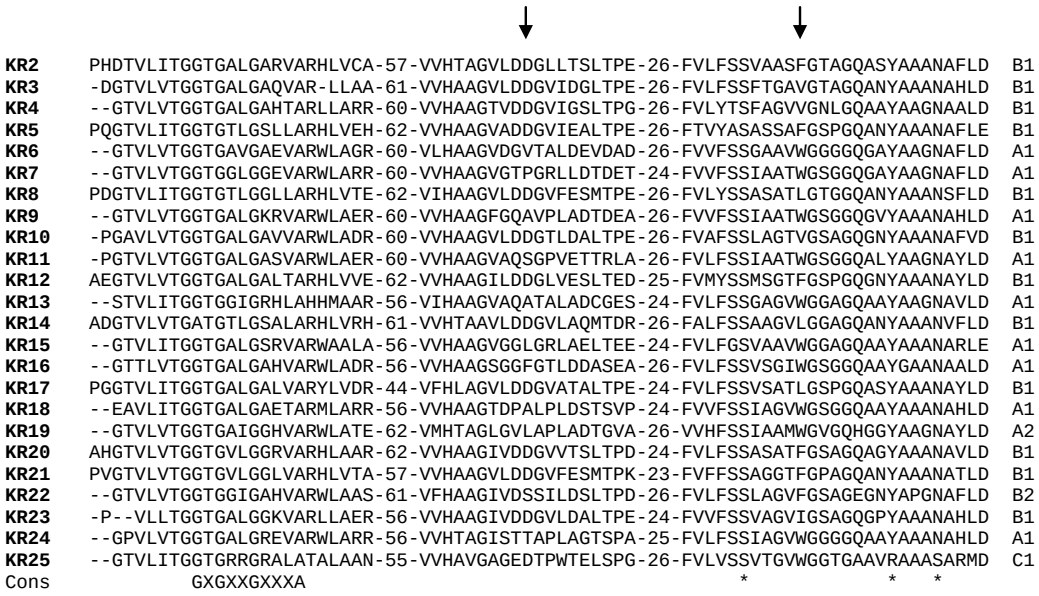
SUPPLEMENTARY FIGURES

ORF	Product size (aa)	% identity / similarity	Species	Putative function	Proposed function
***sam*R0465**	8154			Type I PKS	Polyketide biosynthesis
***sam*R0466**	3661			Type I PKS	Polyketide biosynthesis
***sam*R0467**	5771			Type I PKS	Polyketide biosynthesis
***sam*R0468**	217	84/88	*S. griseus* (SGR874)	Response regulator	Regulation
***sam*R0469**	442	77/84	*S. griseus* (SGR875)	Histidine kinase	Regulation
***sam*R0470**	261	87/93	*S. griseus* (SGR876)	Putative permease protein	Resistance
***sam*R0471**	312	88/93	*S. griseus* (SGR877)	Putative ABC transporter ATP-binding protein	Resistance
***sam*R0472**	244	55/68	*S. erythraea* (EryCVI)	N-dimethyltransferase	Sugar biosynthesis
***sam*R0473**	185	46/59	*S. fradiae*	Isomerase	Sugar biosynthesis
***sam*R0474**	6333			Type I PKS	Polyketide biosynthesis
***sam*RCDS1**	3556			Type I PKS	Polyketide biosynthesis
***sam*R0475**	3157			Type I PKS	Polyketide biosynthesis
***sam*R0476**	3565			Type I PKS	Polyketide biosynthesis
***sam*RCDS2**	1569			Type I PKS	Polyketide biosynthesis
***sam*R0477**	5447			Type I PKS	Polyketide biosynthesis
***sam*R0478**	414	43/61	*Nocardiopsis dassonvillei*	Cytochrome P450	Lactone ring modification
***sam*R0479**	401	41/56	*R.oseiflexus castenholzii*	Cytochrome P450	Lactone ring modification
***sam*R0480**	369	74/82	*Streptomyces* sp. TP-A0274	Aminotransferase	Sugar biosynthesis
***sam*R0481**	418	48/63	*Micromonospora griseorubida* (MycB)	Glycosyltransferase	Sugar attachment
***sam*R0482**	595	64/75	*S. hygroscopicus*	Acyl-CoA synthetase	Unknown
***sam*R0483**	532	62/75	*Micromonopora* sp. ATCC39149	Carboxyl transferase	Unknown
***sam*R0484**	958	35/47	*S. venezuelae* (PikD)	Transcriptional activator (LAL)	Regulation
***sam*R0485**	255	87/94	*S. griseus* (SGR200)	Type II thioesterase	PKS editing
***sam*R0486**	329	77/84	*S. tenebrarius* (AprE)	dTDP-glucose-4,6-dehydratase	Sugar biosynthesis
***sam*R0487**	290	74/85	*S. avermitilis* (AveBIII)	Glucose-1-phosphate thymidyltransferase	Sugar biosynthesis

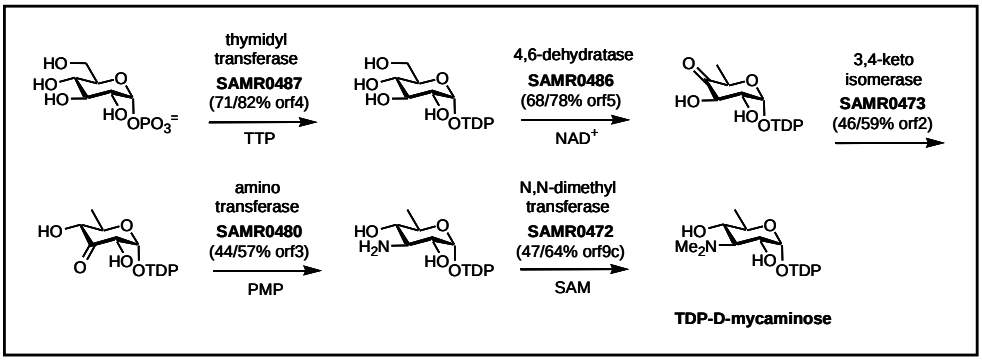
Supplementary Table 1



Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3

Strains, BAC or plasmid	Principle characteristics[a]	Reference
Strains:		
S. ambofaciens		
ATCC23877	wildtype	6
ATCC/pIB139	empty vector integrated in the *att*B site	This work
ATCC/OE484	Overexpression of the LAL regulator	This work
ATCC/OE484/Δ467	Overexpression of the LAL regulator and the gene *sam*R0467 replaced by a kanamycin resistance cassette	This work
E.coli		
DH5α	General cloning strain	26
ET12567/pUZ8002	Nonmethylating strain with mobilization plasmid for conjugation with *Streptomyces*	27
BW25113/pKD20	Strain used for the PCR-targeting mutagenesis (*gam, bet, exo, bla*)	28
BAC or plasmids:		
BBB	BAC from the genomic library of *S. ambofaciens* (*cat*)	7
BBB/Δ467::*neo*+*oriT*	*sam*R0467 replaced by a neomycin cassette in BBB (*cat, neo*)	This work
BBBspec/Δ467::*neo*+*oriT*	*cat* of BBB/Δ467::*neo*+*oriT* replaced by a *aadA* cassette	This work
pIJ776	*oriT, neo*	29
pIJ778	*oriT, aadA*	29
pGEMT-easy	PCR cloning vector, *bla*	Promega
pGEMT-0484	pGEMTeasy + *sam*R0484 without promoter region	This work
pIB139	Conjugative and integrative plasmid (*oriT attP*$_{\phi C31}$ *int* $_{\phi C31}$ *aac*(3)*IV ermEp**)	23
pOE-0484	pIB139 + *sam*R0484	This work

Supplementary Table 2

Primers	Nucleotide sequence (5′→3′)
Overexpression:	
OE484-F	CATATGCTGGTCCATCGAGACGAAC
OE484-R	TCTAGACTCTGCTCTCTCCAAGGCT
Deletion:	
D467-F	**GGCGATCTCGCGCCGTCAAGTGATTTCGGGGCATTCATG**TGTAGGCTGGAGCTGCTTC
D467-R	**CAGAGCGACTCCCAGCGGGCAGGACCGTACATGGCGTCA**ATTCCGGGGATCCGTCGACC
Transcriptional analysis:	
RT-467-F	GTCGCCGGATCACCGAGGAA
RT-467-R	AGGTCGCGGAACGCCTTGTC
RT-465-F	TGCCTGCGGTGCTCCACCAA
RT-465-R	CGTCGTCTTCTCCTCCATCG
RT-477-F	GGAACAGCTCGCCGTACTCC
RT-477-R	CCGAACTCGTCGGCGTATGG
RT-474-F	ACCGCGCCGGAGGTGAGACA
RT-474-R	GCTGCTCGCCTGCGTGGACA
RT-484-F	CTGGAGACCTTCGGGGAGTG
RT-484-R	TGCCCGAGCACTCCGAAATG
HrdB-F	CGCGGCATGCTCTTCCT
HrdB-R	AGGTGGCGTACGTGGAGAAC

Supplementary Table 3

METHODS SUMMARY

Bacterial strains, plasmids and BACs. All are described in **Supplementary Table 2** and their manipulation in Supplementary information.

Overexpression of the LAL regulator. For the expression plasmid pOE-0484, the *sam*R0484-encoding sequence was amplified by PCR from the wt genomic DNA with the primers OE484-F and OE484-R (see **Supplementary Table 3**). The PCR product was cloned in the pGEMT-easy (Promega) and then checked by sequencing. After restriction digestion with *Nde*I and *Xba*I, the insert was cloned into the pIB139 vector under the control of the *ermEp** promoter modified to have a typical *Streptomyces* RBS (Bunet *et al,* 2008). The recombinant plasmid was introduced in *S. ambofaciens* by conjugation and integrated in the chromosomal *att*$B_{\phi C31}$ site by site-specific recombination.

Transcriptional analysis. Total RNA of *Streptomyces* wildtype and mutant strains was isolated from MP5 liquid cultures and treated as described (Kieser *et al.,* 2000). cDNAs were prepared as described (Bunet *et al*., 2008). The primers used for the analysis are listed in **Supplementary Table 3**. PCR conditions are described in supplementary information.

Isolation and structural elucidation of sambomycin. See supplementary data.

METHODS (Supplementary data)

Bacterial strains, plasmids and growth conditions. *Streptomyces* strains were cultivated on or SFM, in YEME and on or in HT media and manipulated as described previously (Kieser *et al,* 2000). Growth curves, RNA extractions, LC-MS and HPLC analysis were obtained from cultures grown in or on MP5 medium (Pernodet *et al*., 1993). *Escherichia coli* strains were cultivated in LB and SOB liquid medium (Sambrook, 1989). When needed these antibiotics were added into the cultures: ampicillin, apramycin, kanamycin, spectinomycin at 50µg/ml, chloramphenicol and nalidixic acid at 25µg/ml. For λ *red* genes induction, 10mM of filtered L-arabinose was added to the culture.

DNA manipulation. Plasmid and BAC DNA were extracted from *E. coli* by alkaline lysis method (Sambrook, 1989). Genomic DNA extraction of *Streptomyces* and pulsed-field gel electrophoresis were performed as already described (Leblond 1996). Southern blots were performed with a Hybond-N nylon membrane (Amersham-Pharmacia) and a vacuum transfer system (BioRad), as previously described (Pang *et al*., 2004). Amplification of DNA fragments by PCR was performed with *Taq* DNA polymerase (NEB) or Takara polymerase (Fermentas), according to the manufacturer's instructions. PCR products and restriction fragments were purified from agarose gel with the High Pure PCR product purification kit (Roche).

Overexpression of the LAL regulator. To amplify *sam*R0467 encoding gene we used the Takara polymerase (Fermentas). The PCR conditions were: 2 min at 94°C, 30 cycles of 10 sec at 98°C and 15 min at 68°C, followed by a final extension of 10 min at 72°C.
The mutant strain ATCC/OE484 was checked by PCR, Southern blot and PFGE.

Transcriptional analysis. The PCR conditions for the transcriptional analysis were: 4 min at 94°C, 28 cycles of 30s at 94°C; 30s at 60°C, and 30s at 72°C, followed by a final extension of 5 min at 72°C. To check possible contamination of genomic DNA, the same PCR programme (35 cycles instead of 28) was applied to RNA samples, after DNase treatment, using as control the primers to amplify *hrdB* gene, which encodes a major sigma factor, considered to be constitutively expressed.

Construction of *S. ambofaciens* mutant strain. The REDIRECT system (Gust et al., 2003) was used to make an in-frame deletion of *sam*R0467 in *S. ambofaciens* ATCC 23877 as described in previous works for other genes (Pang et al., 2004; Aigle et al., 2005).
The *neo+ oriT* mutagenesis cassette, derived from the plasmid pIJ776 (Gust *et al.*, 2003), was used as template in the PCR reaction with the primers D467-F and D467-R. The *E. coli* strain BW25113/pKD20, providing the λ *red* system, was transformed, first with the BAC of interest (BBB), and then with the PCR product (~1400bp) to replace the biosynthetic gene *sam*R0467 by homologous recombination. The *cat* resistant gene of the vector pBELO was replaced by a $Spec^R$ gene, using the same strategy. The strain *E. coli* ET12567/pUZ8002 was transformed with the mutated BAC (BBBspec/Δ467::*neo*+*oriT*) in order to carry out the conjugation with the strain ATCC/OE484. The clones Kan^R and $Spec^S$ were retained and checked by PCR, Southern Blot and PFGE.

Isolation and structural elucidation of sambomycin. MP5 medium was used for production of sambomycin. Twenty 10cm x 10cm square Petri dishes each containing 50 mL of medium were used. A sterile cellophane membrane was placed on the agar in each plate and 20µl of spores of the strain ATCC/OE484 were spread on top of each cellophane membrane. The plates were incubated for 4 days at 30°C and the cellophane membranes were lifted off the plates. The mycelia were scraped off the cellophane membranes and combined. The mycelia were extracted with 3 x 200ml of methanol. The combined extracts were concentrated to 200 mL under reduced pressure and passed through a 0.2 micron filter.

Sambomycins were partially purified from the concentrated methanol extract using semi-preparative HPLC on a reverse phase column (C18, 100x21 mm, fitted with C18 pre-column 10x21 mm). The mobile phases used were water (A) and acetonitrile (B), both with 0.1% formic acid added. The gradient employed was as follows: 0 minutes, 80% A / 20% B; 10 minutes, 80% A / 20% B; 20 minutes, 100% B; 35 minutes 100% B. Absorbance was monitored at a wavelength of 240nm. Fractions containing Sambomycins were identified using ESI-MS and combined. The combined fractions were evaporated under reduced pressure and re-suspended in small volume of 50% aqueous methanol. Sambomycins were purified from the combined fractions by semi-preparative HPLC on the same column using the following gradient: 0 minutes, 60% A / 40% B; 15 minutes, 5% A / 95% B; 20 minutes, 100% B; 25 minutes, 100% B. Fractions containing a mixture of Sambomycins A and B were collected, combined and lyophilised as well as fractions containing a mixture of Sambomycins C and D.

Antimicrobial and antiproliferative tests. The antibacterial activities of the purified sambomycin compounds were first analysed by loading a drop of compounds (mixture of sambomycin A and B or sambomycin C and D; from 1 µg up to 12.5 µg) onto a lawn of *Bacillus subtilis* ATCC6633 and *Micrococcus luteus* used as indicator strains. The IC50 and IC90 values were then determined only for the mixture of sambomycin C/D (the most active fraction) against *Bacillus subtilis* BGSC 1A72, *Mycobacterium smegmatis* ATCC700084 and against clinical isolates of *Enterococcus faecalis* and *Staphylococcus aureus* (strains from Libragen, Toulouse, France). Analyses were also performed against Gram-negative bacteria (*Escherichia coli* and *Pseudomonas aeruginosa*, two clinical isolates from Libragen), against yeast (*Candida albicans* IHEM8060) and filamentous fungi (*Aspergillus fumigatus* GASP4707 and *Fusarium oxysporum* DSM2018).

For the antiproliferative activities and cytotoxity, the IC50 values were determined for sambomycin A/B and sambomycin C/D using human adenocarcinoma (HT29) cell line and

ovary sane cell line from an adult Chinese hamster (CHO-K1), respectively. The cell viability was determined by measuring the cellular concentration of ATP using the CellTiter-Glo® Luminescent Cell Viability Assay (Promega, G7573) according to the recommendation of the manufacturer.

CHAPTER 3

Characterization of sambomycin biosynthesis

The activation of the sambomycin cluster by the overexpression of the LAL regulatory gene led to the isolation and characterization of four forms of a novel giant macrolide endowing a 50-membered lactone ring. The structural elucidation confirmed most of our *in silico* predictions, but it also brought to highlight remarkable features associated with the biosynthesis of these molecules. Another important aspect of my work project was to investigate in more details the biosynthetic mechanisms behind the production of sambomycin compounds. Better insights into this peculiar large type I PKS cluster might prove useful for a combinatorial biosynthesis approach, towards the generation of other non natural bioactive drugs.

3.1 Limits of the sambomycin cluster

Initially, the boundaries of sambomycin cluster had been fixed from the gene *sam*R0465 to *sam*R0487 according to the putative function of the genes nearside the PKS genes, as described at the beginning of chapter 1. However, it was interesting to determine more precisely which and how many genes were directly responsible for the biosynthesis of sambomycin macrolides.

Our approach aimed to delete from each side of the borders of the cluster a large DNA region (about 7 kb), comprising several genes. If the deletions would affect the production of sambomycin, a strategy to delete each single gene would have been planned afterwards. Two groups of genes, from *sam*R0493 to *sam*R0488 on the left border and from *sam*R0457 to *sam*R0464 on the right border, were interrupted by a spectinomycin resistance cassette, using the PCR-Targeting strategy. The constructs were separately introduced in the overexpression mutant ATCC/OE484, generating two new mutants, ATCC/OE484/Δ457-464::spec and ATCC/OE484/Δ488-493::spec.

The new mutants were grown in MP5 medium and the mycelia and the supernatant were extracted and analysed by LC-MS. The results pointed out that both mutants were still able to produce sambomycin compounds, which were found only in the mycelium extracts, as expected. The MS chromatogram showed the typical double charged peaks at *m/z* 673 and 680, corresponding to sambomycin C/D and A/B, respectively (see Fig. 54). Therefore, the deletion of a significant number of genes flanking the cluster clearly did not affect the biosynthesis of sambomycin, which indicates that all the genes necessary for the biosynthesis seemed to be clustered between *sam*R0465 and *sam*R0487 (see Table 6). However, we cannot exclude that other genes involved in the biosynthesis of the product, in the regulation of the cluster or in the resistance mechanism are located somewhere else on the chromosome. For example, the genes responsible for kinamycin production are not found all clustered together (Bertrand Aigle, pers.

comm.). The expression of the cluster in a heterologous host, although difficult due to the size of the cluster, might better clarify this aspect.

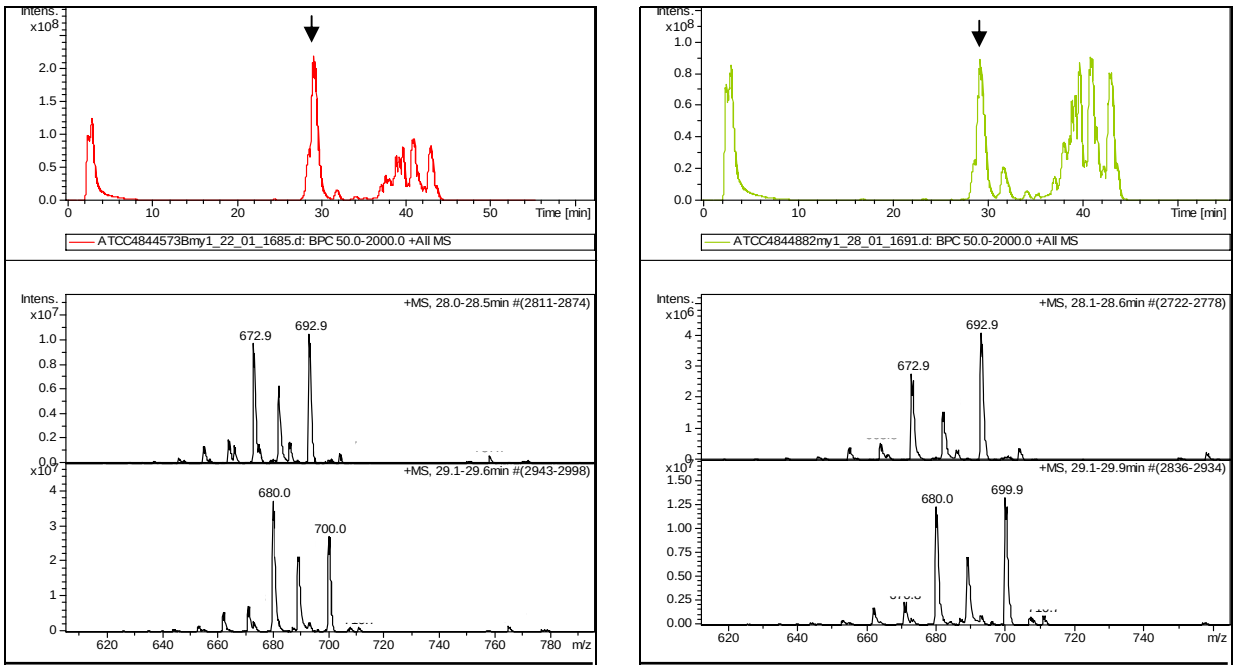


Fig. 54 LC-MS spectra of the mycelia extracts of the double mutants ATCC/OE484/Δ457-464::spec (on the left) and ATCC/OE484/Δ488-493::spec (on the right) grown in MP5 medium. The different retention time of the peaks 673 and 680 (indicated with an arrow) is due to the machine. On the MS chromatogram, under the LC spectrum, the double charged peaks 673 (672.9) and 680, as well as their sodiated forms 693 (692.9) and 700 (699.9) are detected.

3.2 The phosphopantetheinyl transferase

As anticipated above, some genes involved in the biosynthesis of sambomycin might be located elsewhere on the chromosome. Indeed, inside the sambomycin cluster no gene was identified to encode a 4′-phosphopantetheinyl transferase (PPTase), the enzyme responsible for activating the apo form of the ACP domains of the PKSs (Lambalot *et al.* 1996). Similarly, the spiramycin cluster as well as the congocidine cluster lack this gene (Karray *et al.* 2007; Juguet *et al.*, 2009), suggesting that perhaps the three clusters share the same PPTase. On the contrary, in the kinamycin cluster, another secondary metabolite gene cluster present on the chromosome of *S. ambofaciens*, the gene *alpN* encodes a PPTase (Pang *et al.*, 2004). Deletion of this gene prevents kinamycin biosynthesis, but not that of spiramycin (Robert Bunet, pers. comm.). PPTase have a broad range of specificity towards the active site serine of the ACP domains, the PCP domains or the fatty acids. In this latter case PPTases are essential enzymes (Sunbul *et al.*, 2009). Hence, we wanted to verify if AlpN might be involved in the biosynthesis of sambomycin. The construct for the overexpression of the LAL regulator, pOE-0484 (see Methods in the article), was introduced in the mutant ATCC/ΔΔ*alpN* by intergeneric conjugation. The mutant strain was also conjugated with the vector pIB139 to use as control.

The two strains, ATCC/ΔΔ*alpN*/OE484 and ATCC/ΔΔ*alpN*/pIB139, were grown in MP5 medium and the mycelia, extracted with methanol, were analysed by LC-MS. The deletion of the PPTase gene did not affect the biosynthesis of sambomycin when the regulatory gene is overexpressed (see Fig. 55), thus AlpN does not activate the ACP domain of the sambomycin cluster, and therefore it is probably specific for the kinamycin cluster. However, other genes coding for a putative PPTase have been identified in the chromosome and they are possibly involved in the production of sambomycin.

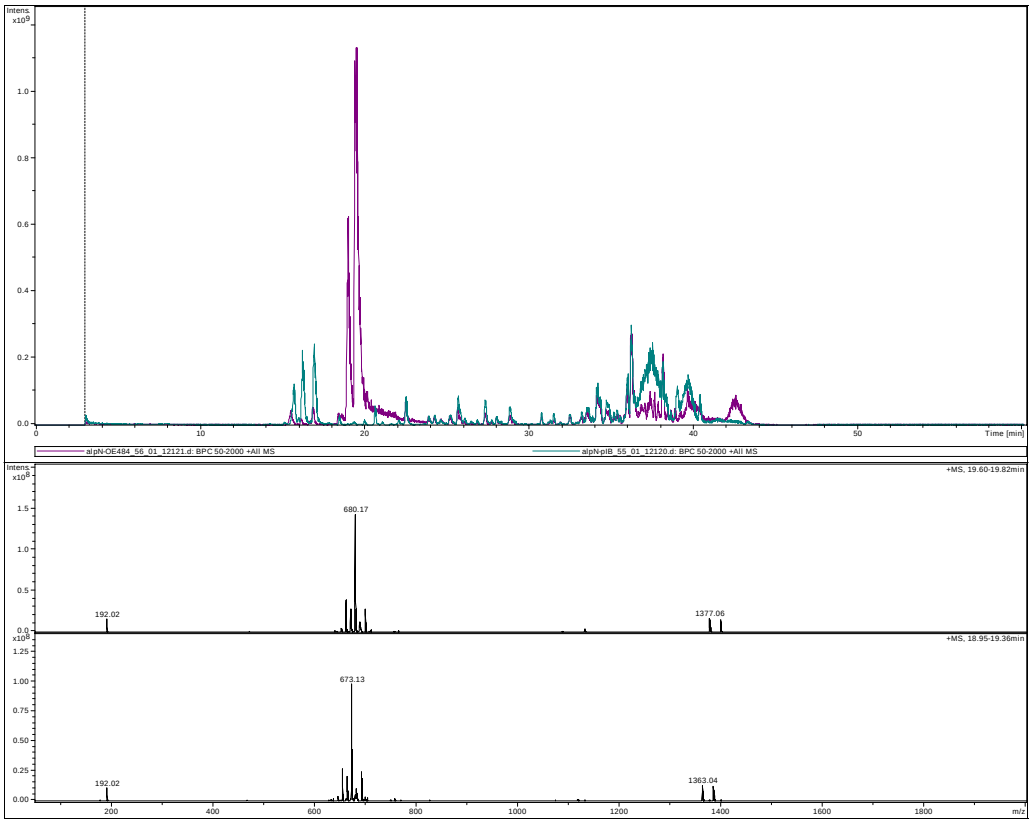


Fig. 55 LC-MS spectrum of the mutants ATCC/ΔΔ*alpN*/OE484 (in violet) and ATCC/ΔΔ*alpN*/pIB139 (in green). On the bottom of the figure, the MS chromatogram showed the mass of the double charged peaks 673 and 680, and the mass of the molecules mono charged (1363 and 1377).

3.3 The cytochromes P450

The sambomycin structure unveiled the presence of an additional hydroxyl group, previously predicted, inserted at the C-28 position of the lactone ring (see Fig 56). More important the elucidation of the structure enabled us to identify where and how the cyclization takes place. The reaction occurs between the first and the last precursor, generating a gigantic 50-membered macrolactone ring (see Fig. 56). Usually cyclization, operated by the TE domain, occurs between a hydroxyl and a keto group, already present in the polyketide backbone. However, the starter unit of sambomycin, a propionate, lacks a hydroxyl group, which indicates that it was unexpectedly introduced just before the cyclization reaction. The addition of these free oxygen

atoms was verified by incorporation experiments with isotopically-labelled ^{18}O, which confirm the presence of both oxygens (see Fig. 14 and 15 of Appendix).

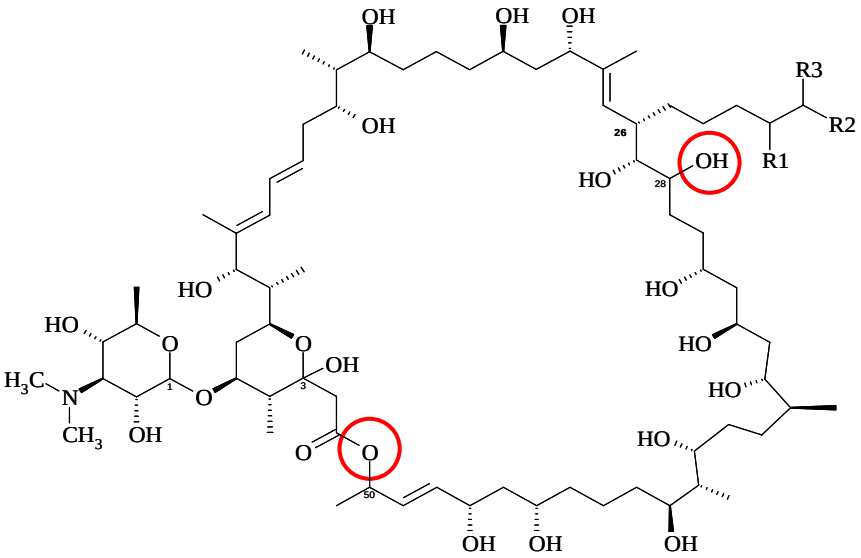


Fig. 56 General structure for sambomycin macrolides. The two oxygen atoms added by the cytochromes P450 are circled in red. The substituents R1, R2 and R3 give rise to the different sambomycin compounds. The numbers 1, 3, 26, 28 and 50 highlight specific structural characteristics. The labelling experiments have been performed by Dr. Christophe Corre at the University of Warwick.

The cytochromes P450 are the enzymes responsible for the hydroxylation reactions, a post-modification usually observed in the structure of secondary metabolites (Hertweck *et al*., 2007). The sambomycin cluster contains two genes encoding a putative cytochrome P450, *sam*R0478 and *sam*R479. To investigate their role and to establish which of them is probably involved in the cyclization reaction, the deletion of the two genes, independently and together, was achieved, in the overexpression strain. The genes were replaced by a streptomycin resistance cassette, thus generating three distinct mutant strains, ATCC/OE484/Δ478::spec, ATCC/OE484/Δ479::spec and ATCC/OE484/Δ478-9::spec.

The mycelia of the mutants, grown in MP5 medium, were extracted and analysed by LC-MS. The gene *sam*R0478 appeared to be involved in the hydroxylation reaction at the C-28 (Fig. 56). As shown in the Fig. 57, the MS chromatogram of the mutant strain ATCC/OE484/Δ478::spec presented two double charged peaks at *m/z* 656 and 663, and their mono charged peaks at *m/z* 1347 and 1361 (with molecular formula $C_{72}H_{132}NO_{21}$ and $C_{73}H_{134}NO_{21}$, respectively). The compound with mass 1347 corresponds to the mix sambomycin C/D without an oxygen atom (16 of difference); likewise, the compound with mass 1361 to the mix sambomycin A/B lacking an oxygen atom. Besides this, the two peaks differ by 14Da, corresponding to a CH_2 group, as sambomycin A/B and sambomycin C/D. The mutant has also been confirmed by NMR analysis, in which the signal corresponding to this hydroxyl group disappeared completely (data not shown), and by incorporation experiments with isotopically-labelled ^{18}O which highlighted the

presence of only one free oxygen atom (see Fig. 16 and 17 of Appendix). Taken together, these results clearly point out that the gene *sam*R0478 encodes the cytochrome P450 responsible for this oxidation reaction.

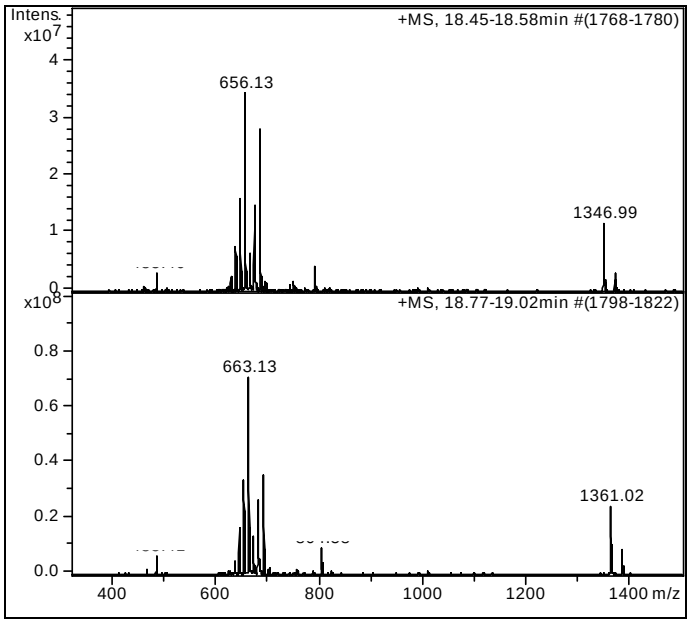


Fig. 57 MS chromatogram of the mycelium of the mutant ATCC/OE484/Δ478::spec. On the top the mass of the double peak 656 and of its mono charged peak 1347. On the bottom the other double peak 663, with its mono charged peak 1361. These experiments were performed by Dr. Lijiang Song at the University of Warwick.

The deletion of the gene *sam*R0479, and of both cytochrome genes, appeared to affect more the biosynthesis of sambomycin, most likely because the cyclization is prevented. In fact, the LC-MS analyses of ATCC/OE484/Δ479::spec highlighted the presence of several peaks with mass 1214, 1228, 1242 (see Fig 58 A, B, C), which might correspond to some linear structures associated to sambomycin compounds. At the same time, the mutant ATCC/OE484/Δ478-9::spec also presented several peaks, 1198, 1212, and 1226 (see Fig. 58 D, E, F), which are 16Da less than the previous one, corresponding to the loss of an oxygen atom, presumably the one added on by the cytochrome SAMR0478. This cytochrome seems to operate even when the polyketide is not cyclised yet. On the other hand, no sugar moiety has been identified in the linear structure; thus it is probable that the glycosyltransferase is strictly specific only for a macrolactone ring receiver, as also observed for other GTs (Härle and Bechthold, 2009).
Taking into account these results, the gene *sam*R0479 encodes a cytochrome P450 that for the first time is described as a key enzyme in the cyclization reaction of a polyketide product.

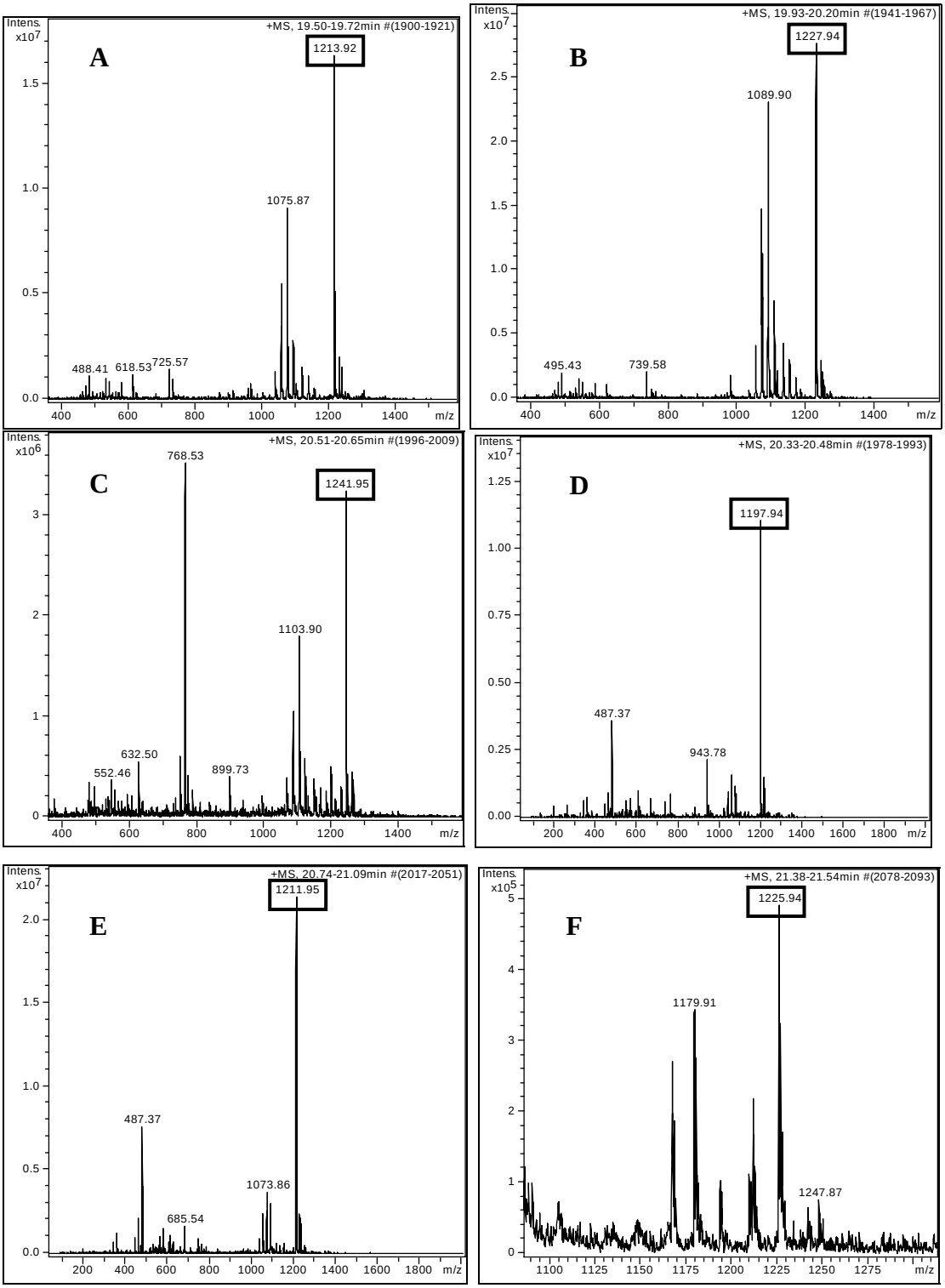


Fig. 58 MS spectrum of the peaks 1214, 1228 and 1242 detected in the mutant ATCC/OE484/Δ479::spec (A, B, C) and the peaks 1189, 1212 and 1226 detected in the mutant strain ATCC/OE484/Δ478-9::spec (D, E, F). The experiments have been performed by Dr. Lijiang Song at the University of Warwick.

3.4 The acyl-CoA synthetase and the acyl-CoA carboxylase

The structure of the sambomycin compounds presents a peculiar alkyl chain at the position C-26 (Fig. 56), loaded by the AT13 domain. From the *in silico* analysis it had not been possible to predict this type of precursor and NMR analysis was necessary to clarify it. Normally, polyketides are synthesised from simple small organic acids, such as acetate or propionate. When uncommon extender units are incorporated in the polyketide structure, a specific group of genes responsible of their biosynthesis are identified inside the cluster (Chan *et al*., 2009).

Looking at the sambomycin cluster, only two genes have not been assigned with a putative function linked to the biosynthesis, *sam*R0482 and *sam*R0483. The first gene is homologous to an acyl-CoA synthetase, the enzyme responsible for the activation of the precursors of fatty acid biosynthesis, as well as of polyketides. FAS and PKS pathways share this enzyme, whose gene was rarely identified in secondary metabolite clusters, more often linked to fatty acid metabolism. Therefore, it could be possible that the presence of this gene in the cluster is correlated to the presence of the atypical precursors, used by the AT13 domain.

The product of the gene *sam*R0483 is homologous to a propionyl-CoA carboxylase (PPCase) in *Micromonospora* sp. ATCC39149. This kind of enzyme catalyzes the carboxylation of a propionyl-CoA substrate to provide methylmalonyl-CoA, the substrates incorporated in polyketide biosynthesis, as described for PccB in *S. coelicolor* (Diacovich *et al*., 2004). The β-subunit is thought to be directly involved in the specificity of the substrate, which in this case might be for the uncommon alkyl chain.

In order to investigate in the role of these genes, we decided to construct two deletion mutants of *sam*R0482 and *sam*R0483, respectively, employing the PCR-Targeting strategy. Nonetheless, only the mutant strain ATCC/OE484/Δ482::spec was obtained, introducing in the gene a spectinomycin resistance cassette. It was not possible to replace the gene *sam*R0483 in the strain ATCC/OE484 because the region of homology to obtain a double crossing-over for not sufficient. Unfortunately, we have a unique BAC of the library (BBC, http://www.weblgm.scbiol.uhp-nancy.fr/ambofaciens/index.html) which contains the entire gene *sam*R0483 for the mutagenesis.

The mycelium extracted from the mutant ATCC/OE484/Δ482::spec was analysed by LC-MS. Two double charged peaks at *m/z* 673 and 680, characteristic of sambomycin C/D and A/B, respectively were still detected (Fig. 59). Therefore, the gene *sam*R0482 does not seem to be essential for the biosynthesis of the compounds, and in particular no structural changes are associated with the alkyl chain. The loss-of-function could be compensated by another acyl-CoA synthetase located elsewhere on the chromosome.

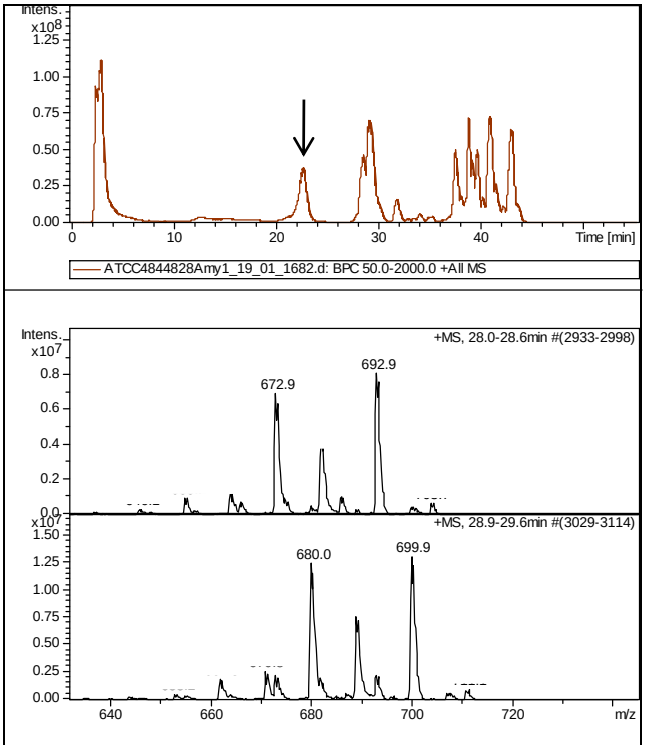


Fig. 59 The LC-MS spectrum of the mycelium extract of the mutant strain ATCC/OE484/Δ482::spec. On the top, the arrow pointed out the peaks 673 and 680. On the bottom the MS chromatograms of the double charged peaks 673 (672.9) and 680 are reported. The sodiated form for each peak is also represented, 693 (692.9) and 700 (699.9).

3.5 Secretion of sambomycin

Sambomycin macrolides have always been detected and purified by extraction of the mycelium and have previously not been detected in the extracellular fraction. Surprisingly very recently, we have identified a growth condition in which the sambomycin products are finally secreted.

The overexpression mutant ATCC/OE484 was grown on HT supplemented with 15mM $MgCl_2$, since magnesium experimentally showed to stimulate antibiotic production (Jean-Luc Pernodet, pers. comm.). The biological activity of the strain was tested against *M. luteus* and *B. subtilis*, as indicator strains, since sambomycin (the pure compounds) is active against both microorganisms. A clean inhibiting activity, never detected before, was observed against *M. luteus* (Fig. 60 C) but not *B. subtilis* (data not shown), which could possibly be associated to the product of the sambomycin cluster.

S. ambofaciens ATCC23877 produces other bioactive molecules: the macrolide spiramycin, the pyrrole amide congocidine and the angucyclinone kinamycin. Only the first two antibiotics are effective against *M. luteus* (also kinamycin at high concentrations), however the wt strain used in our laboratory showed to lack congocidine production for unknown reasons (Sarka Nezbedova, pers. comm.). In addition, we have observed that the inhibition zone produced by spiramycin against *M. luteus* is quite diffused on the border (Fig. 60 E), compared to this new

clean zone (Fig. 60 C). Spiramycin activity was observed in the wt and the control strain ATCC/pIB139 (Fig. 60 A and B), thus it is probable that the new activity hides spiramycin one. To confirm our hypothesis, that the clean biological activity could be due to sambomycin, the mutant ATCC/OE484/Δ467::kan, deficient for sambomycin production, was tested in the same conditions. Any antibacterial activity was longer detected (Fig. 60 D), either associated to sambomycin or to spiramycin. This last result suggested, and we later confirmed (Chapter 4) that the overexpression of the LAL regulator not only has a positive effect on sambomycin cluster, but it might have a negative effect on spiramycin cluster, preventing or delaying the antibiotic biosynthesis.

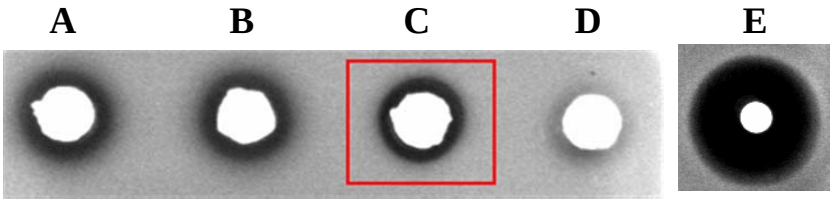


Fig. 60 Bioassay on LB soft mixed with cultures of *M. luteus*. Agar plugs of the strains ATCC (A), ATCC/pIB139 (B), ATCC/OE484 (C) and ATCC/OE484/Δ467 (D) grown on HT+$MgCl_2$ for 3 days at 30°C were used. The bioassay plates were incubated at 4°C for four hours to allow the molecules to diffuse in the medium and then the plates were incubated at 37°C O/N. 0.5µg of spiramycin were also tested (E) as control.

To confirm that the antimicrobial activity observed was linked to sambomycin, the experiment was repeated with a mutant strain unable to produce both spiramycin and kinamycin [spi^-, alp^-], named SPM701A (constructed in collaboration with Jean-Luc Pernodet). Two new mutants were obtained: SPM701/OE484 and SPM701/OE484/Δ467::kan, which were verified by PCR and Southern analysis. We first confirmed by LC-MS analysis that the mutant SPM701/OE484 was able to produce sambomycin macrolides. Then, the new mutant strains were grown on HT supplemented with 15mM $MgCl_2$ and their activity tested against *M. luteus*. As expected, a clean inhibition zone was observed only for the strain 701/OE484, while in the same strain deleted for the PKS gene *sam*R0467 the activity disappears (Fig. 61).

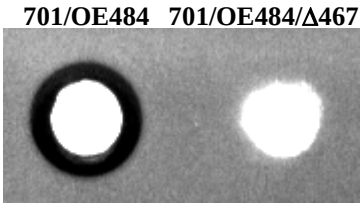


Fig. 61 Biological assays of the mutant strains 701/OE484 and 701/OE484/Δ467::kan, grown three days on HT+$MgCl_2$. The conditions used were the same indicated in Fig. 60.

These additional data clearly indicated that sambomycin or a modified form of the macrolides can be secreted in particular growth conditions. To validate this hypothesis, we decided to

extract and analyse by LC-MS the secretome of the strain SPM701/OE484 and in parallel that of SPM701/OE484/Δ467::kan. Thus, a sterile cellophane membrane was placed on the surface of the HT+$MgCl_2$ plates to separate the mycelium from the medium and the spores of both mutants were[2] spread on top of it. After three days of incubation at 30°C, the membranes were removed and the medium was extracted with the same volume of methanol. The samples were first tested for the biological activity against *M. luteus* and also *B. subtilis* (Fig. 62 A). This time both indicator strains showed to be sensitive to the extract obtained from SPM701/OE484 probably because the concentration of sambomycin in the extracts is higher. As expected, no antibacterial activity was detected for SPM701/OE484/Δ467::kan.
In parallel, to verify if sambomycin is also accumulated intracellularly, the mycelium attached to the membranes was scrapped off, recovered and extracted by methanol, as described in Methods in the article. Biological assays assessed that only the extract of the strain SPM701/OE484 contains a bioactive molecule, most probably the sambomycin, able to inhibit the growth of *M. luteus* (Fig. 62 B).

Fig. 62 A) Bioassay of the medium extracts of SPM701/OE484 (OE1 circled in red) and SPM701/OE484/Δ467::kan against *M. luteus* (OEΔ1on the left) and *B. subtilis* (on the right). 10µl of the methanol extract were dropped directly on the LB soft mixed with the indicator strains. The plates were incubated 2 hours at 4°C and then incubated O/N at 30°C for *B. subtilis* and 37°C for *M. luteus*. **B)** Bioassay of the mycelium extracts of SPM701/OE484 and SPM701/OE484/Δ467::kan against *M. luteus*, obtained as described above.

Subsequently, the secretome extracts of the two mutant strains were analyzed by LC-MS. The chromatogram of SPM701/OE484 showed the typical two double charged peaks at *m/z* 680 and 673, with a retention time of 18-19 min. The MS analysis confirmed that the peaks have a fragmentation profile corresponding to sambomycin A/B and C/D, respectively (Fig. 63). The sambomycin characteristic peaks were not detected in the strain SPM701/OE484/Δ467::kan (Fig. 63). This analysis definitely confirmed that sambomycin compounds can be secreted in particular growth condition without any changes in their chemical structure. How magnesium induces the secretion of sambomycin is still a mechanism to clarify.

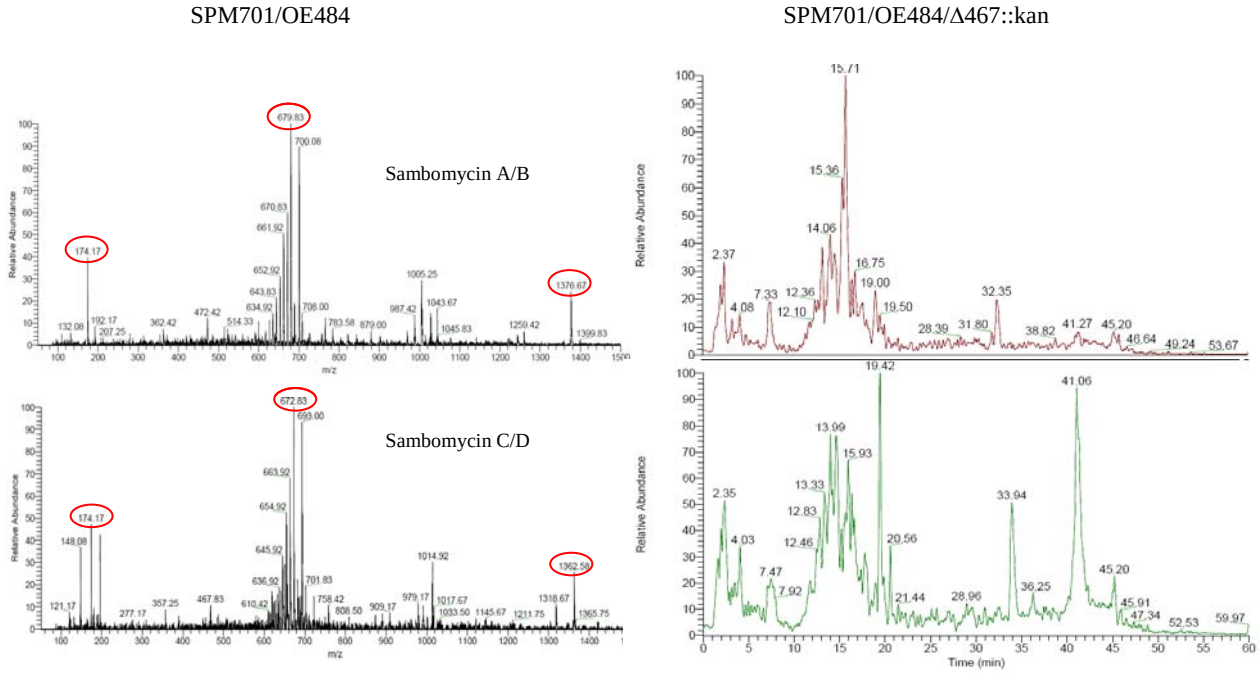


Fig. 63 MS spectra of the medium extracts of the mutant strain SPM701/OE484 and SPM701/OE484/Δ467::kan. Sambomycin compounds have been identified only in the strain SPM701/OE484 (on the left). On the top left the MS profile of sambomycin A/B in which the double peak 680, the mono charged peak 1376 and the sugar fragment 174 are circled in red. On the bottom left the MS profile of sambomycin C/D with the double peak 673, the mono charged peak 1362 and the sugar fragment 174 are circled in red. The LC-MS analyses were performed at the University of Nancy (ENSAIA).

In summary, we found a growth condition which enables to directly and visually analyse the mutant strains. We can thus verify when sambomycin is produced, secreted and which mutation can significantly affect its biological properties. All the mutant strains previously obtained, in particular those of the tailoring genes, were grown on HT medium supplemented with 15mM $MgCl_2$ and their antibacterial activity was assessed against *M. luteus* (see Fig. 64).

The mutants for the limits of the cluster, ATCC/OE484/Δ457-464::spec and ATCC/OE484/Δ488-493::spec, still produced sambomycin (Fig. 64), in good accordance with the results from the LC-MS analysis; likewise the mutant in the acyl-CoA synthetase ATCC/OE484/Δ482::spec still showed activity (Fig. 64), probably due a complementation of another homologous functional gene. On the contrary, the deletion of the glycosyltransferase SAMR0481 and of the cytochrome P450 SAMR0479 (responsible for the cyclization) showed to abolish the antibacterial activity of sambomycin. It is likely that the sugar moiety is necessary for the biological properties of the compounds, as usually observed for the glycosylated metabolites (Härle and Bechthold, 2009). Interestingly, the mutant ATCC/OE484/Δ478::spec, which only lacks a hydroxyl group, is still effective against *M. luteus* (Fig. 64), suggesting that this product can be considered as a bioactive derivative of sambomycin.

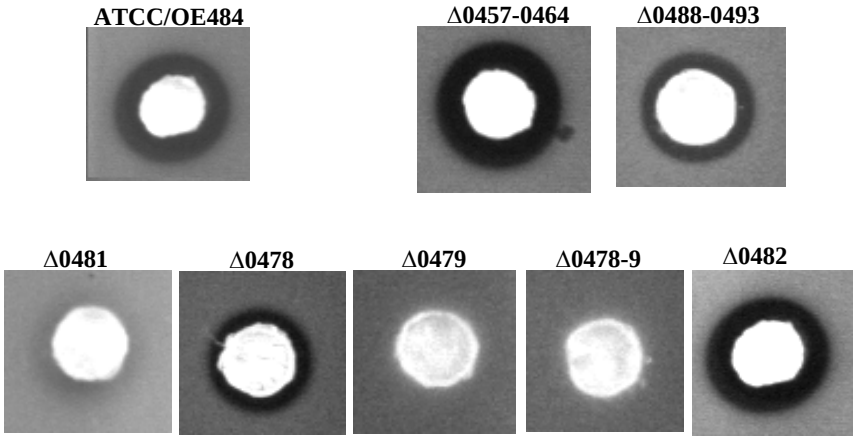


Fig. 64 Bioassays against *M. luteus*. The agar plugs were taken from the mutant strains grown on HT+$MgCl_2$ (15mM) after three days of incubation at 30°C. The deletion mutants all derived from the overexpression mutant ATCC/OE484, which is represented for comparison to the other inhibition zones. *sam*R0481, glycosyltransferase gene; *sam*R0478 and *sam*R0479, cytochrome P450; *sam*R0482, acyl-CoA synthetase.

3.6 The resistance genes

The sambomycin cluster contains two genes, *sam*R0470 and *sam*R0471, encoding a putative ABC transporter. When present in secondary metabolite gene clusters, this kind of transporters are usually associated with a resistance mechanism used by the producing strain to self-protect against the bioactive molecules (Mendez and Salas, 2001). More rarely, they were described as active transmembrane pores to transport extracellularly the product of the cluster out of the cell, without being involved in the resistance mechanism, as shown for the macrolide oleandomycin or the glycopeptide balhimycin (Hernandez *et al*., 1993; Quiros *et al*., 1998; Menges *et al*., 2007).

To investigate the role of *sam*R0470 and *sam*R0471, both genes were replaced together by an apramycin resistance gene in the spiramycin and kinamycin deficient strain SPM701A [spi^-, alp^-], using the PCR-Targeting strategy, and generating the mutant SPM701/Δ470-1::apr. Before inserting into the chromosome the construct for the overexpression of the LAL regulator (pOE-0484) sharing the same resistance marker, the apramycin cassette used to delete the genes *sam*R0470 and *sam*R0471 was removed by expressing an exogenous Flip recombinase, adapted for *Streptomyces* genome (Fedoryshyn *et al*., 2008). At the end the mutant strain SPM701/OE484/Δ470-1::scar was obtained and checked by PCR and Southern analysis.

The new mutant strain was grown on HT supplemented with 15mM $MgCl_2$ and the biological activity was tested against *M. luteus*. Very surprisingly, a clean inhibition zone characteristic of sambomycin products was observed for the two independent clones SPM701/OE484/Δ470-1::scar (Fig. 65 C and D), even if smaller compared to the one produced by SPM701/OE484 (Fig. 65 A).

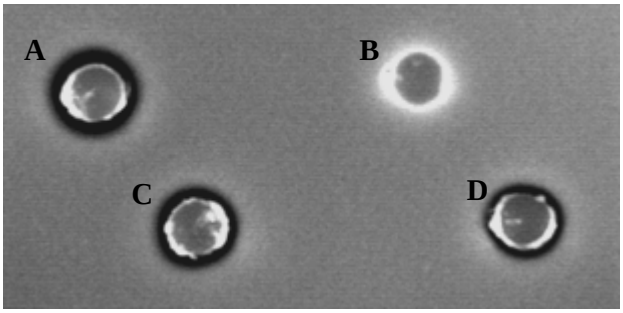


Fig. 65 Bioassay against *M. luteus* using agar plugs from the medium used to grow SPM701/OE484 (A), SPM701/OE484/Δ467::kan (B) and two independent clones SPM701/OE484/Δ470-1::scar (C and D). The strains were grown on HT+$MgCl_2$ (15mM) and were incubated for three days at 30°C. On the plates a cellophane membrane was positioned to separate the mycelium from the medium.

To corroborate that the activity observed is linked to sambomycin, the secretome and the mycelium of the mutant strains were extracted with methanol and analysed by LC-MS; in the same time the extracts were also tested against *M. luteus*. As obtained for the agar plugs, a biological activity was observed in the secretome and in the mycelium of 701/OE484/Δ470-1::scar (data not shown). The LC-MS analysis showed the presence of the sambomycin compounds in both the mycelium and the medium extract (data not shown). The deletion of the putative resistance genes did not affect sambomycin production or its secretion; hence another resistance mechanism must be responsible for the self-protection of the strain as well as another transport system should be involved in the secretion of this giant metabolite.

CHAPTER 4

Regulation of the sambomycin cluster

Another major aspect of this thesis consisted of characterizing the regulatory mechanism that controls the expression of the sambomycin gene cluster. In particular, our attention has focused on the *sam*R0484 gene, encoding a LAL regulator, which was shown to act positively in the production of the macrolide compounds (see Chapter 2). A better insight in this mechanism might prove useful in a future improvement of the metabolite production. Moreover, we also analyzed more in detail a mechanism of cross regulation observed between the spiramycin and the sambomycin gene cluster.

4.1 The LAL regulator is an activator of the sambomycin gene cluster

In the strain ATCC/OE484 the overexpression of the LAL regulator allowed inducing the transcription of the biosynthetic genes of the type I PKS cluster (see Fig. 2A in the article), thus leading to the isolation of its cryptic product.

In literature, LAL regulators have been described to have a positive effect on other genes of the cluster they control, in particular on the resistance genes or the glycosyltransferase gene (Wilson *et al*., 2001; Sekurova *et al*., 2004). Therefore, transcriptional analyses by RT-PCR were carried out comparing the control strain ATCC/pIB139 and the overexpression mutant ATCC/OE484 in order to investigate the expression of all the genes of the sambomycin cluster.

Both strains were grown in MP5 medium and RNA samples were extracted at different points of the growth curve. The cDNAs were prepared as described in Methods in the article. The results of the comparative transcriptional analyses are illustrated in Fig. 66. Two time points of the curve were chosen to exemplify, one from the exponential phase and another from the stationary phase. With the exception of four genes (*sam*R0485, 0484, 0482 and 0479), all the sambomycin genes are transcribed in the same direction. Besides, some of them are terminally overlapping, such as the genes *sam*R0468, 0469, 0470, 0471 and *sam*R0472-0473. Therefore, the putative transcriptional starting points have been established when the intergenic regions were longer than 100bp, showing that most of the genes in the cluster are probably cotranscribed, in particular the PKS genes, as well as the resistance genes, the two component system and some of the sugar genes. In order to confirm these transcriptional units, experiments with terminator sequence cassettes inserted at appropriate locations inside the cluster and transcriptional analysis on the resulting mutants strains, as previously described (Dangel *et al*., 2009), will be accomplished in the future.

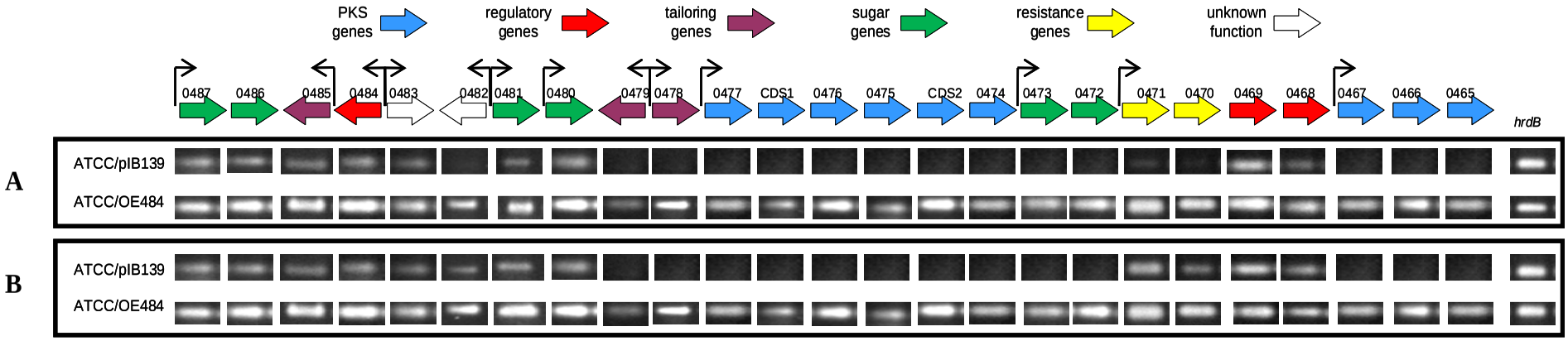


Fig. 62 Comparative transcriptional analyses of the control strain ATCC/pIB139 and the mutant strain ATCC/OE484, grown in MP5 medium. The expression of all the genes of the sambomycin cluster was analysed during the exponential phase (A) and during the stationary phase (B) of growth. The gene *hrdB* is used as internal control. The genes are not represented in scale. The hypothetical transcriptional starting points are marked by an arrow.

The overexpression of the LAL regulator triggers the expression of all the PKS encoding genes, as already expected (see Fig. 66). In addition, it activates the transcription of the genes *sam*R0470 and *sam*R0471, coding for the putative resistance system, and the genes *sam*R0478 and *sam*R0479, coding for a putative cytochrome P450, as well as the tailoring genes *sam*R0472, *sam*R0473 and *sam*R0482.

The constitutive expression of the LAL regulator seems to have a positive effect on almost all the other genes of the cluster (e.g., the glycosyltransferase *sam*R0481 and the sugar genes *sam*R0480, 0486 and 0487) in which the level of transcription seems to increase, compared to the control strain (Fig. 66). On the contrary, the regulatory genes *sam*R0468 and *sam*R0469, forming the two component system, do not seem to be significantly affected by the overexpression of the LAL regulator (Fig. 66). However, to better quantify the difference in the level of expression of sambomycin gene cluster between the control and the overexpression strain an experiment of real time RT-PCR would be necessary.

4.2 Searching for the targets of the LAL regulator

The transcriptional analyses of the overexpressed mutant proved that the LAL regulator controls, either directly or indirectly, the expression of the sambomycin cluster. The regulation most probably occurs at the transcriptional level, since the LAL regulator contains an HTH domain at the C-terminus (Fig. 45), typical of DNA binding proteins. In addition, the LAL prototype, MalT, was described as a positive transcriptional regulator controlling the maltose regulon in *E. coli* (Richet and Raibaud, 1989). It was shown that MalT dimers bind to a direct repeat decanucleotide sequence ($GGGGA^{T}/_{G}GAGG$) in the promoter region of its targets, promoting the recruitment of the RNA polymerase (Boos and Shuman, 1998).

In an attempt to identify the direct targets of the LAL regulator, a serie of promoter-probe constructs fused to the *redD* gene were engineered. RedD is the specific transcriptional activator of the pigmented antibiotic undecylprodigiosin in *S. coelicolor* and *S. lividans* (Narva and Feitelson, 1990). The constructs were based on the reporter system described by van Wezel *et al*. (2000). For these experiments, the host strain *S. coelicolor* M512 was used, which contains deletions in the regulatory genes *actII-ORF4* (transcriptional activator of the actinorhodin gene cluster) and *redD*, rendering the strain colourless since the production of both pigmented antibiotics actinorhodin and undecylprodigiosin is prevented (Floriano and Bibb, 1996). No homologous genes of the LAL regulator were found inside the genome of *S. coelicolor*, indicating that the strain would be a good candidate to test the transcriptional activity of SAMR0484. For this reason, the recombinant vector pOE-0484 was inserted in the chromosome of *S. coelicolor* M512 by site specific recombination, generating the mutant M512/OE484. In parallel, the empty vector pIB139 was also integrated to obtain the control strain M512/pIB139. The presence of the recombinant vectors was verified by PCR and Southern analysis.

For the reporter system, we chose the low-copy-number plasmid pIJ2587, which is able to replicate autonomously in *Streptomyces* and contains a promoterless copy of the *redD* gene. Initially, the intergenic regions tested were those of the PKS genes *sam*R0467 (from +210 to -538 with respect to the translational start point) and *sam*R0477 (from +82 to -538), and of the resistance gene *sam*R0471 (from +197 to -324). As previously mentioned, it is likely that the PKS genes are cotranscribed (from *sam*R0467 to *sam*R0465 and from *sam*R0477 to *sam*R0474), which is possibly true for the resistance genes as well (*sam*R0471-0471)(see Fig. 66). The choice of these regions depended on the results from the transcriptional analysis that indicated a significant effect of the LAL regulator on the PKS genes and the resistant genes (see Fig. 66).
The introduction of the recombinant vectors pIJ-R0467 and pIJ-R0477 in the M512 strain overexpressing the LAL regulator produced colonies with a red pigmentation, due to the production of undecylprodigiosin (Fig. 67). On the contrary, in the control strain M512/pIB139 the colonies remained colourless, even after several days of incubation (see Fig. 67). These results suggested that SAMR0484 most probably has a direct effect on the transcription of the PKS genes, binding to a specific sequence within their promoter regions and recruiting the RNA polymerase. The prototype of LAL regulator, MalT, was shown to be able to bind to its target only in the presence of ATP (the regulator contains a Walker A and B domain) and maltotriose (Richet and Raibaud, 1989). With this reporter system, it is not possible to establish if a cofactor like ATP or another substrate contribute to the binding of SAMR0484 to the DNA targets. However, if a particular substrate is needed to promote SAMR0484 regulation, *S. coelicolor* does produce it. Moreover, the control strain suggested that no regulator of *S. coelicolor* is capable of specifically activating the PKS genes, at least not to a detectable level.

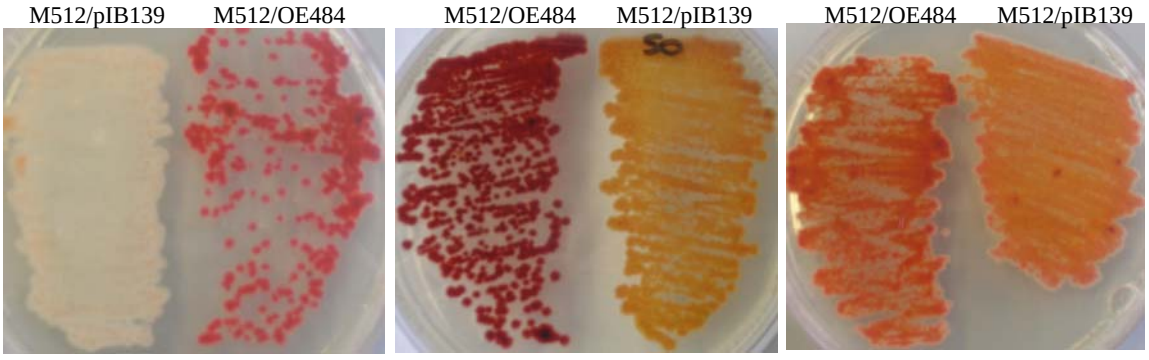


Fig. 67 Undecylprodigiosin-producing colonies by the transformants M512/OE484 containing the recombinants vectors pIJ-R0467, pIJ-R0477 and pIJ-R04701, compared to the transformants M512/pIB139 containing the same plasmids. The transformants are plated on R2YE+25µg/ml thiostrepton. The colour of the colonies can sometimes appear darker because the plates were incubated at 30°C for a longer time.

A preliminary analysis of the intergenic regions of *sam*R0467 and *sam*R0477 did not reveal any palindromic repeat sequences (CRISPR finder, http://crispr.u-psud.fr/Server/CRISPRfinder.php) as well as no significant direct repeats (REPFIND, http://zlab.bu.edu/repfind/form.html). However, further analysis, also comparing the two sequences, will be done in the future.
On the other hand, the presence of the vector pIJ-R04701, containing the intergenic region of the putative ABC transporter genes, produced red colonies both in the overexpression mutant M512/OE484 and in the control strain M512/pIB139 (Fig. 67). This suggests that a transcriptional factor encoded by *S. coelicolor* is able to bind to this region and promote transcription. Indeed, the genome of *S. coelicolor* contains homologous of the genes *sam*R0470 and *sam*R0471, namely SCO5453 and SCO5452, respectively, and it might be as well for the promoter regions. Nevertheless, we can not exclude that the LAL regulator bind to the same region; indeed red colour in the mutant strain M512/OE484 containing pIJ-R04701 seems more intense than in the control strain (Fig. 67). A more direct way to prove interaction would be by using electrophoretic mobility shift assay (EMSA), with the recombinant SAMR0484 purified protein and the labelled DNA targets.

4.3 *In vitro* characterization of the LAL regulator

In support of the transcriptional analyses and the reporter system data, electrophoretic mobility shift assays could provide insight into the direct role of SAMR0484 on the transcription, since it is possible to detect the *in vitro* binding of a protein to its cognate target DNA sequence. To functionally characterize SAMR0484, it was necessary to express and purify the protein.
The coding sequence of the LAL regulator was cloned in the expression vector pET15b (see Methods in the Appendix), containing an N-terminal His-Tag sequence fused to the coding sequence to facilitate the detection and the protein purification. The recombinant vector was introduced in the *E. coli* BL21(DE3) strain. Since, the sequence of SAMR0484 contains several codons for arginine, AGA and AGG, rare codons in *E. coli* BL21(DE3)/pET15+*sam*R0484 was transformed with the plasmid pSBET, which has an arginine tRNA gene, specific to, the AGA and AGG codons to improve the level of protein expression (Schenk *et al*., 1995). The protein profile 3h after IPTG induction was visualised on a 10% SDS-PAGE electrophoresis gel (Fig 68 A). The gene *sam*R0484 codes for a 958 aa protein with a predicted molecular weight (MW) of 102 kDa (which becomes around 105 kDa considering the His-Tag). A band consistent with this MW was detected on the gel, in the total, the soluble and insoluble fraction (Fig. 68 A). Western blot analysis using antibodies against the hexahistidine-tag confirmed that the band observed was actually the LAL regulator (Fig. 68 B).
The protein seemed to accumulate principally in inclusion bodies, whereas only a small percentage was present in the soluble fraction. In either case, the level of expression was very low, thus affecting protein purification. We have tested several growth temperatures for protein

expression and found that the highest expression occurred at 24°C (Fig. 68). This limiting protein production could be explained by the induction conditions, which might not be optimal, by the large size of the protein or by the low efficiency of translation in *E. coli*. Indeed, it was experimentally observed that the type of nucleotides in the second codon can significantly influence the efficiency of translation in *E. coli* (Looman *et al*., 1987). Gene expression can vary by a factor of 15, depending on the codon following the AUG start codon. In SAMR0484 sequence, the second codon (CUG) encodes a Leu, which has an apparent efficiency of 0.1% (Looman *et al*. 1987).

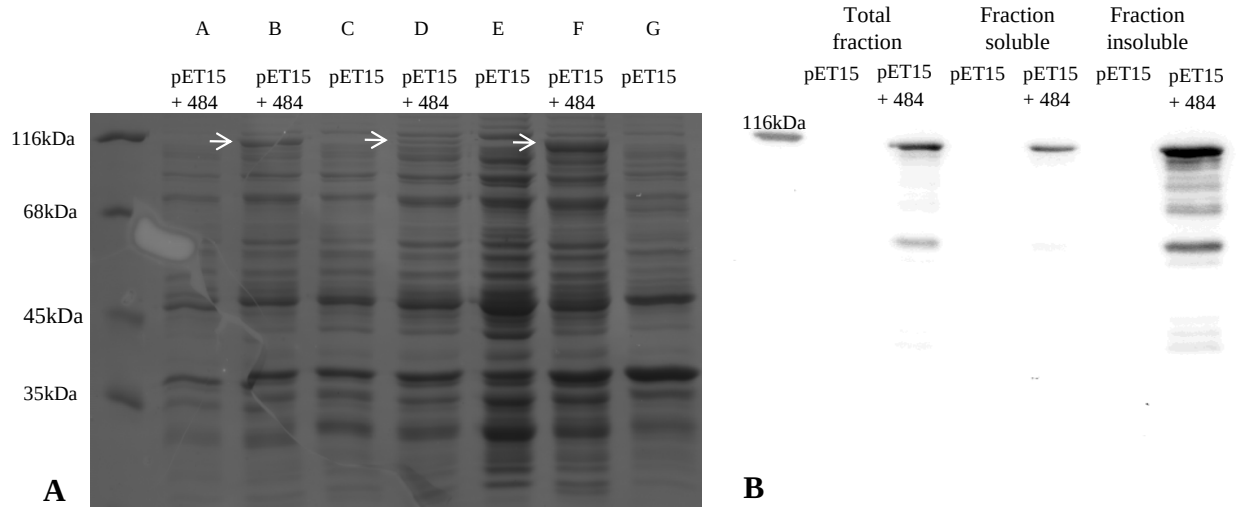


Fig. 68 A 10% SDS-PAGE gel of the protein extracts of *E. coli* BL21DE(3)/pSBET containing pET15+*sam*R484 or pET15 grown on LB and induced with 0.5 mM IPTG. Lane A corresponds to the sample before the induction; lanes B and C to the samples after 3h of induction at 24°C; lanes D and E to the soluble fractions and lanes F and G to the insoluble fractions. The recombinant SAMR0484 protein is indicated with an arrow in the different fractions. **B** Western analysis using antibodies anti HisTag on different fractions containing the heterologous expression of SAMR0484. The band at 116 kDa of the marker was also detected by the same antibodies.

A possible alternative to functionally characterize SAMR0484 was to express only the HTH domain. Indeed, it has been demonstrated that this domain might be sufficient to obtain DNA binding, as described for example for the SARP-like regulator AfsR (Tanaka *et al*., 2007). Indeed, a truncated form of AfsR, containing only the HTH domain, was shown to be sufficient to bind to a specific sequence within the promoter region of *afsS*. Besides, we would avoid the need for any cofactor or substrate necessary for the binding. We chose to amplify a sequence encoding the HTH domain of SAMR0484 and interrupting the protein in a loop region, to avoid affecting the tertiary structure (see Appendix). Moreover, the truncated sequence was chosen to have as second initiation codon an Ile residue, ATC, which is predicted to have a factor of efficiency of 0.9% (Looman *et al*. 1987). In this way, a small truncated protein (T484) of 220 aa was expected.

E. coli BL21(DE3) was transformed with the new recombinant vector pET15+T484 and protein expression was obtained as described (see Methods of this chapter). The protein samples, after

induction with IPTG at 24°C, were loaded on a 12% SDS-PAGE gel (Fig. 69 A). The truncated protein SAMR0484 has a predicted molecular weight of 26kDa and a band corresponding to this size was detected. As before for the recombinant SAMR0484, Western blot analysis confirmed that the protein was expressed in the soluble fraction (Fig 69 B). However, this time the level of expression is higher than that obtained with the full length recombinant SAMR0484 (Fig. 68 B).

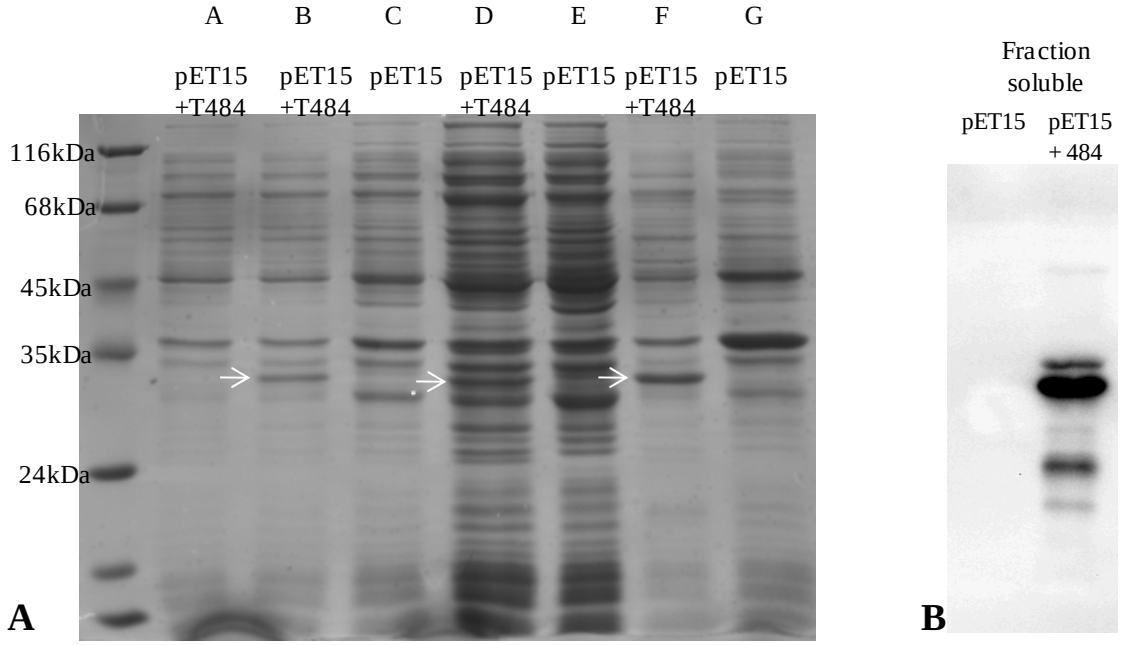
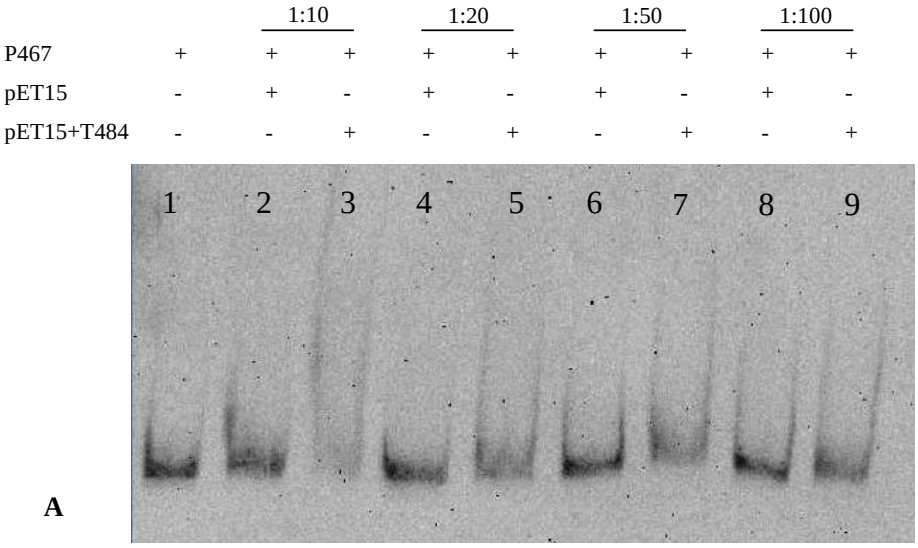


Fig. 69 A 12% SDS-PAGE gel of the protein extract of *E. coli* BL21DE(3) containing pET 15+T484 or pET15 grown on LB and induced with 0.5 mM IPTG. The lane A corresponds to the sample before the induction; the lanes B and C to the samples after 3h of induction at 28°C; lanes D and E to the soluble fractions and lanes F and G to the insoluble fractions. The recombinant truncated form of SAMR0484 (T484) is indicated with an arrow. **B** Western analysis using antibodies anti-HisTag detects the presence of the truncated protein T484 in the soluble fraction.

Gel retardation assays were carried out with the crude extracts of the soluble fraction of strains containing pET+T484 and the empty vector pET15 as control. The DNA target sequences chosen to assay the binding of the HTH domain of SAMR0484 were the intergenic regions of the PKS genes *sam*R0467 and *sam*R0477, since these regions gave positive results with the reporter system (Fig. 67). However, the DNA regions were shorter compared to those used for the reporter system: 215bp (from +1 to -214) for *sam*R0467; 194bp (from +16 to -178) for *sam*R0477. The fragments, indicated as P467 and P477, were labelled as described in Methods and they were incubated with different dilutions of the crude extracts of pET15+T484 and pET15.

No shifts were observed for the crude extract bearing the insert T484 when it was incubated with the probe P467 (see Fig. 70). The same result was obtained for the probe P477 (data not shown). Nevertheless, a decrease in the intensity of the DNA probe P467 mixed with the crude extract pET15+T484 is visible (Fig. 70 lanes 3, 5, 7, 9), compared to the same probe mixed with

the control pET15 (Fig. 70 lanes 2, 4, 6, 8), and also compared to the intensity of the probe alone (Fig. 70 lane 1). This decrease might indicate the binding of SAMR0484 to the target sequence.



A

Fig. 70 EMSA assay with the crude extracts of the soluble fraction of *E. coli* BL21DE(3) containing pET15 and pET15+T484 and using as DNA target the intergenic region of the *sam*R0467, named P467 (A), and *sam*R0477, named P477 (B). Several dilutions, from 1:10 up to 1:100, of the extracts were used. 2ng of labelled probes P467 and P477 were used to detect the possible binding of the truncated protein.

These results seem to suggest that the HTH domain can bind to the DNA targets, but with a low affinity. Further experiments improving the binding conditions will have to be done to provide conclusive answer to this.

4.4 Sambomycin production in R2 medium

The overexpression of the LAL regulatory gene seemed to be the only condition to trigger the expression of the biosynthetic genes and to allow the production of sambomycin. Therefore, we decided to screen other growth conditions to test whether the cluster could be activated in the wt strain. Since, the production of kinamycin, another secondary metabolite of *S. ambofaciens*, was observed only in R2 medium (Pang *et al*., 2004), we wanted to test whether the sambomycin cluster could be expressed in the wt strain grown in this medium. Transcriptional analyses by RT-PCR were carried out on the wt strain to investigate the expression of the biosynthetic gene *sam*R0467. The PKS gene showed to be expressed (Fig. 72 A), while in MP5 or HT media it was never the case unless by overexpressing *sam*R0484. Hence, the sambomycin gene cluster seemed to be activated in this condition and it was likely that the macrolide compounds were produced. However, the biological assays against *M. luteus* did not point out any activity related to sambomycin production, presumably because the product is not secreted. In order to detect

sambomycin macrolides, the mycelium of the wt strain grown on R2 was extracted by methanol and analysed by LC-MS. The typical double charged peaks at *m/z* 673 and 680 of sambomycin were thus finally detected (Fig. 71).

To confirm the positive role of SAMR0484 in the regulation of the sambomycin cluster, the gene *sam*R0484 was deleted in the wt strain by replacing it with an apramycin resistance cassette using the PCR-targeting strategy (adapted from the Methods in the article).In parallel, to confirm that the two detected peaks were linked to sambomycin biosynthesis, the PKS gene *sam*R0467 was also deleted in the wt strain using the same strategy.

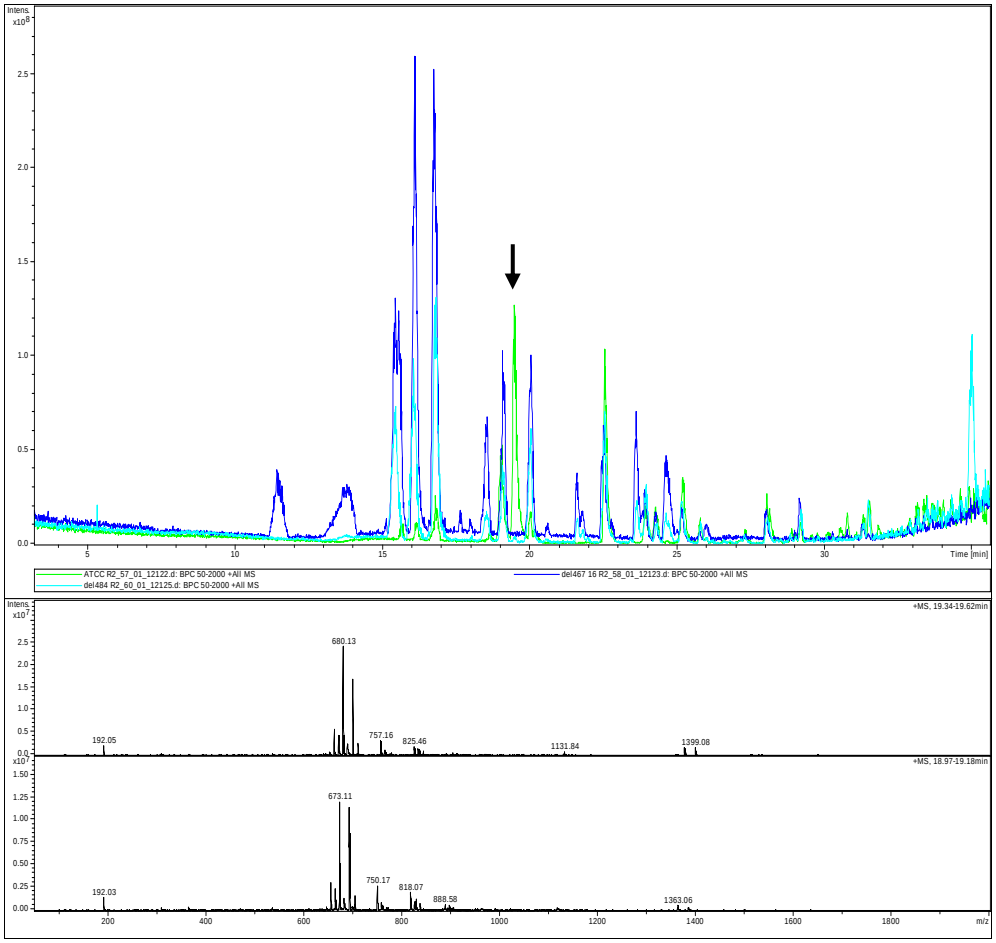


Fig. 71 LC-MS spectrum of the mycelium extracts of the ATCC23877 strain (green line) and the mutant strains ATCC/Δ484::apr (light blue line) and ATCC/Δ467::kan (blue line). The peak indicated with an arrow come out around 19-20 min, and its MS profile (underneath) confirmed that it corresponds to sambomycin A/B (peak at *m/z* 680) and C/D (peak at *m/z* 673).

The two mutants, ATCC/Δ484::apr and ATCC/Δ467::kan, were grown in R2 medium using the same conditions as for the wt strain (incubation for three days at 30°C under agitation). The strains showed the typical orange pigmentation, due to kinamycin production (Pang *et al.*, 2004). The mycelia of the mutants were extracted after three days of incubation (see Methods in

the article) and analysed by LC-MS. The chromatograms showed that both mutants were unable to produce sambomycin (see Fig. 71). This data indicated that the molecules isolated were really associated to the cluster and it proved for the first time that the LAL regulator is an essential pathway-specific activator of sambomycin cluster.

In order to delve into the production of sambomycin in R2 medium, comparative transcriptional analysis by RT-PCR were carried out between the wild type strain and the mutant ATCC/Δ484::apr. The RNA was extracted at different time points of the growth curve and the cDNA were prepared as described in Methods in the article. The expression of the PKS gene *sam*R0467 and of the regulatory gene *sam*R0484 was examined after 30 cycles of PCR. As already observed, the biosynthetic gene is expressed in the wt strain, while its transcription is blocked in the strain deleted for the regulatory gene *sam*R0484 (Fig. 72). This additional result confirmed the role of essential pathway-specific regulator of SAMR0484.

Interestingly, the PKS gene showed a sort of cyclic pattern of expression in the wt strain, as well as the LAL regulator (see Fig. 72). Indeed, the genes seem to be well expressed in the exponential phase (points 1 and 2), then the expression decreases during the transition phase and the early stationary phase (points 3 to 6), to restart in the late stationary phase (points 7 and 8). Since the expression of both *sam*R0467 and *sam*R0484 seems to be growth-phase dependent, it will be interesting to analyse sambomycin production by LC-MS at different time of the curve, when the genes seem more activated and when the level of transcription drops off.

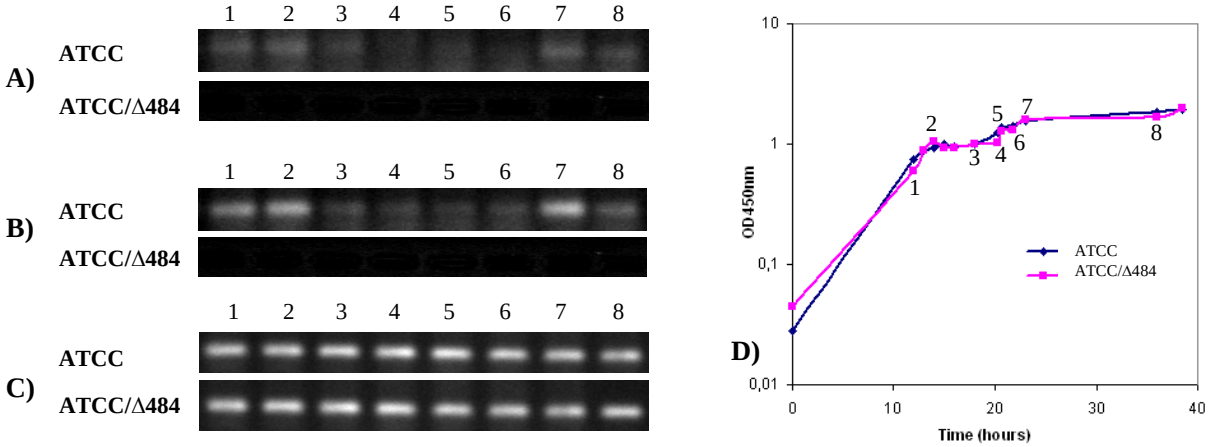


Fig. 72 Transcriptional analyses by RT-PCR of the PKS gene *sam*R0467 **(A)**, the regulatory gene *sam*R0484 **(B)** and the internal control *hrdB* **(C)**. RNA samples were obtained from the wt strain and the mutant ATCC/Δ484::apr, grown in R2 medium. The growth curve of both strains is represented **(D)** and the time points chosen for the analyses are marked.

4.5 Effect of SAMR0484 on spiramycin production

We had observed that the introduction of the recombinant vector pOE-0484 in *S. ambofaciens* strains, which leads to the production of sambomycin, inhibited or decreased significantly spiramycin production according to the growth conditions used. For example, the strains

ATCC/pIB139 and ATCC/OE484 were grown on MP5, the medium for spiramycin production, and R2 and the biological properties were tested against *M. luteus*. In these conditions, sambomycin is not secreted, thus spiramycin production can be easily detected. An antibacterial activity was clearly observed for the control strain ATCC/pIB139 in both media (Fig. 73A) which is likely to be spiramycin, since kinamycin is not effectively active against *M. luteus* and the *S. ambofaciens* strain in our laboratory was shown to be defective in congocidine production (Sarka Nezbedova, pers. comm.). In the mutant overexpressing the LAL regulator, grown in liquid or solid R2, this activity disappeared (Fig. 73 A on the top), while when grown in liquid or solid MP5 the activity decreased significantly (Fig. 73 A on the bottom).

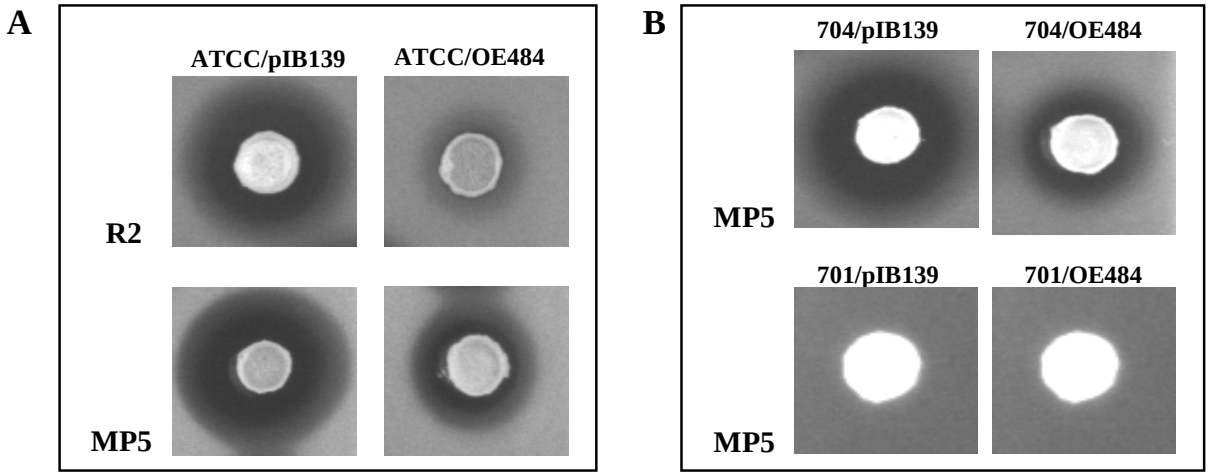


Fig. 73 A) Biological assays against *M. luteus* using agar plugs of the mutant strains ATCC/pIB139 and ATCC/OE484 grown on R2 or MP5 media. The same results were obtained in liquid cultures. **B)** Biological assays against *M. luteus* of the mutant strains SPM704/pIB139 and SPM704/OE484, and SPM701/pIB139 and SPM701/OE484 grown on MP5 medium. All strains were incubated at 30°C for at least three days.

Similarly, we tested the mutant strains SPM704/pIB139 and SPM704/OE484 and SPM701/pIB139 and SPM701/OE484, grown on MP5. The strain SPM704 is defective in kinamycin and congocidine production [alp$^-$ and cgc$^-$], and the strain SPM701 is defective in spiramycin and kinamycin production (constructed in collaboration with Jean-Luc Pernodet). The use of these strains enables us to verify that the effect on spiramycin cluster was not due to the other known secondary metabolites produced by *S. ambofaciens*. The strain SPM704/pIB139 presented an antibacterial activity against *M. luteus*, certainly due to spiramycin, since this activity is undetectable in the strain SPM701 lacking spiramycin production (see Fig. 73B). The activity decreased significantly when the recombinant vector pOE-0484 is integrated in the chromosome, clearly suggesting a direct or indirect effect of the LAL regulator on spiramycin cluster.

To confirm that the activity observed was due to spiramycin, the supernatants of ATCC/pIB139 and ATCC/OE484 grown in R2 and MP5 liquid cultures were extracted with ethyl acetate. The

samples were analysed by HPLC, as described by Gourmelen *et al*. (1998). The chromatograms of the overexpression mutant showed that the peak of spiramycin disappeared, when the samples are grown in R2, or markedly decreased, in MP5 cultures (Fig. 74).

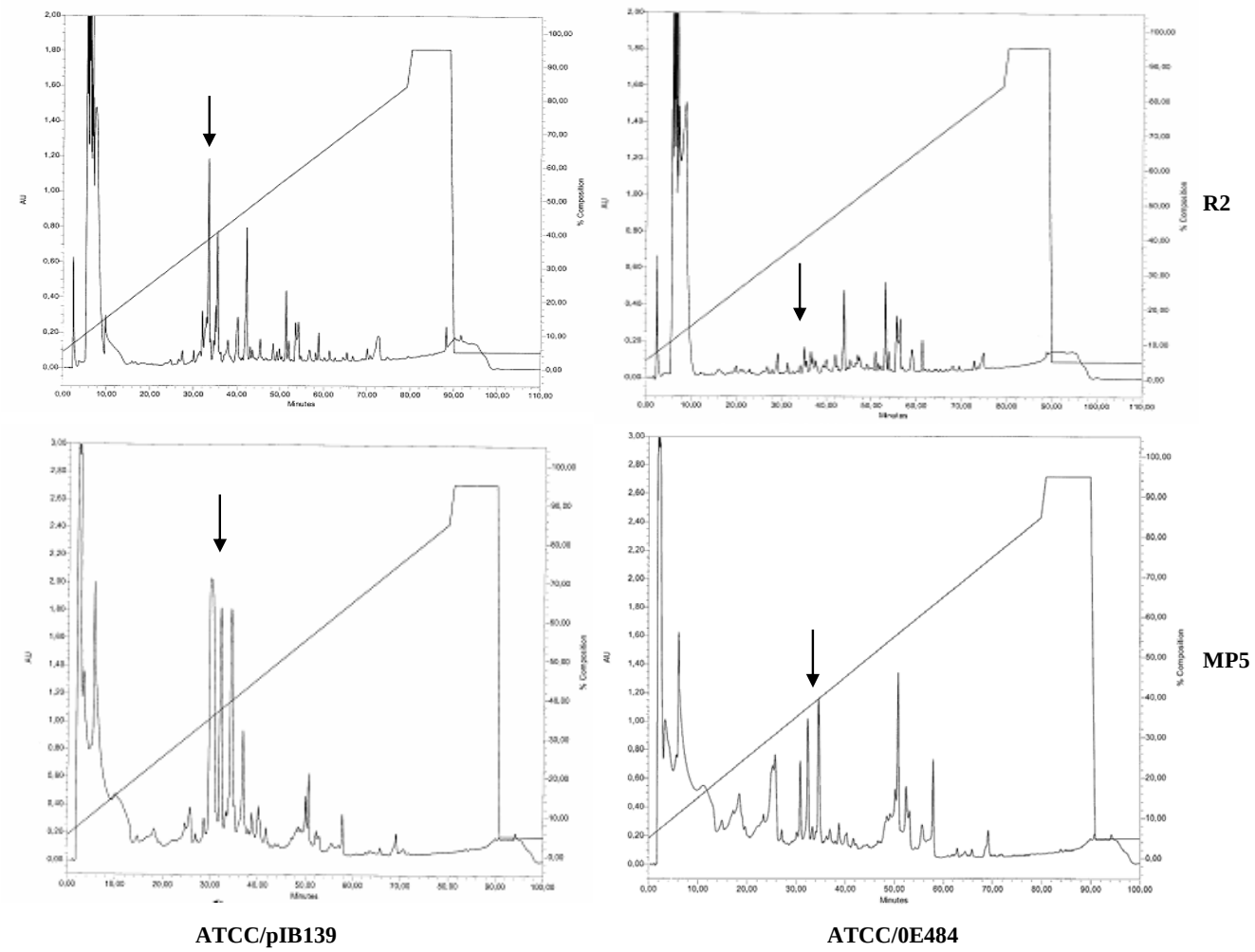


Fig. 74 HPLC chromatograms of the supernatant extracts of the strains ATCC/pIB139 and ATCC/OE484, grown in R2 (on the top) or in MP5 (on the bottom). The arrows indicated the peaks of spiramycin, which is produced in three different forms. Spiramycin absorbs at a characteristic wavelength of 232nm.

One of the hypotheses formulated to explain the effect on spiramycin production was a shunt of metabolites. In literature this phenomenon has been already described for other secondary metabolite clusters (Sun *et al*., 2002). Indeed, both spiramycin and sambomycin clusters belong to the modular type I PKS genes and they share the same pool of precursors (i.e., malonyl and methylmalonyl-CoA thioesters) for the biosynthesis, as well as the sugar residue mycaminose. Sambomycin production required more substrates than spiramycin and it is possible that the flux of precursors is directed towards the biosynthesis of this giant macrolide. If this was the case, when sambomycin biosynthesis is prevented, spiramycin production should be restored. Nevertheless, the strain ATCC/OE484/Δ467::kan, which is no longer able to produce sambomycin, did not present any antibacterial activity when grown on MP5 or R2 medium,

except after more then four days of incubation at 30°C. This phenomenon was as well observed on HT supplemented with 15mM $MgCl_2$. Indeed, the wt and the control ATCC/pIB139 strains, as expected, are able to synthesize spiramycin, as judged by its activity against *M. luteus* (Fig. 60 A and B) and the mutant ATCC/OE484 secretes sambomycin (Fig. 60 C). When the PKS gene *sam*R0467 is deleted, no biological activity is detected, not even associated to spiramycin (Fig. 60 D). This suggests that another mechanism, perhaps involving a regulatory factor, e.g. the LAL regulator, controls the spiramycin cluster.

4.6 The two component system

The products of the genes *sam*R0468 and *sam*R0469 form a putative two component system, another regulatory factor that might control sambomycin cluster, by sensing a specific environmental signal. In chapter 2, we already demonstrated that the constitutive expression of these genes did not activate the expression of PKS genes (data not shown), suggesting that they must not have a positive role in the biosynthesis of sambomycin metabolites. Moreover, no important effects were detected on the transcription of the LAL regulator, which probably is not a target of the response regulator. Similarly, the transcriptional analyses of the overexpression of SAMR0484 showed no significant increment of expression of both *sam*R0468 and *sam*R0469 (Fig. 66).

In order to establish whether the two component system was a negative regulator of the cluster, both genes, presumably co-transcribed, were deleted in the wt strain by PCR-targeting, using a spectinomycin resistance gene cassette. An easy way to verify if the deletion triggered sambomycin production was by biological assays. Hence, two independent clones of the mutant strain ATCC/Δ468-9::spec were grown on HT supplemented with 15 mM $MgCl_2$ and during several days of incubation at 30°C agar plugs were tested against *M. luteus*. The inhibition zone observed is most likely spiramycin, considering the biological activity obtained with the wt strain in the same conditions (Fig. 60 A). However, we still need to confirm this by HPLC analysis. In the same time, no clear activity was associated with the secretion of sambomycin, but also in this case we need a confirmation through LC-MS analysis (Fig. 75).

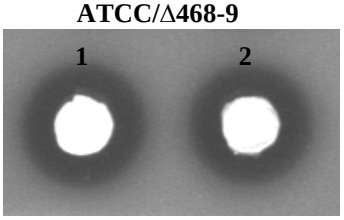


Fig. 75 Bioassay against *M. luteus* of two independent clones of the mutant strain ATCC/Δ468-9::spec grown on HT+$MgCl_2$ (15mM) and incubated for three days at 30°C.

CONCLUSIONS AND PERSPECTIVES

This project started with the sequencing of the *S. ambofaciens* chromosome, which unveiled several putative secondary metabolite clusters located at the extremities of the arms (Choulet, thesis 2006). The interest for novel bioactive compounds has always been high, considering the constant spreading of old and new infectious diseases, a sore for humankind. In the search for new drugs, the genome mining approach proved to be a powerful tool in the identification of cryptic secondary metabolite biosynthetic pathways, as exemplified by the isolation of the siderophore coelichelin in *S. coelicolor* (Lautru *et al*., 2005). Here, we reported the discovery of a novel bioactive natural product of *S. ambofaciens* ATCC23877, obtained by a successful combination of different genomic-guided strategies.

A genome mining approach firstly led to focus our attention on a large type I PKS gene cluster, identified in the right arm of the chromosome of *S. ambofaciens*. The cluster contains 9 large PKS genes, composed of 25 modules and covering in total 124 kb. The domain organization revealed to be unique. In literature and genome databases no other clusters presented the same type of organization or this high number of enzymatic domains. Before discovering the sambomycin cluster, the clusters for the polyene macrolides were thought to be the largest modular PKS gene clusters (Aparicio *et al*., 2003). Subsequently, the *in silico* analyses of the PKS genes enabled to predict the polyketide backbone of the expected metabolite and the tailoring genes within the cluster suggested the incorporation of a sugar residue.

The major drawback, the lack of expression of the cluster, was successfully overcome, for the first time in *Streptomyces*, by manipulating a pathway-specific regulator encoded by *sam*R0484. Eventually, a comparative metabolic profiling approach, in combination with spectroscopic techniques, led to the isolation of four forms of a 50-membered macrolide, named sambomycin A, B, C and D. The novel products of *S. ambofaciens* endowed antibacterial and interesting antiproliferative activities.

The structural elucidation allowed delving into the mechanisms of sambomycin biosynthesis, proposing a role for each gene of the cluster. In parallel, we have tried to partially decipher the regulation system that controls the expression of the cluster. The LAL regulator was demonstrated to be an essential transcriptional activator.

The results obtained during this work contributed to demonstrate the potential of the *Streptomyces* genus, in our case of *S. ambofaciens*, in producing novel metabolites of interest. Besides, a better understanding of the mechanisms beyond sambomycin biosynthesis would be useful for a possible combinatorial approach, in order to generate sambomycin derivatives or new non-natural bioactive compounds.

1. Sambomycin, a peculiar 50-membered macrolide

Sambomycin A, B, C and D represent a novelty in the class of macrolides, since they possess a 50-membered lactone ring that has never been described. In literature, the largest macrolides characterized to date were the 44-membered swinholide A, isolated from the marine sponge *Theonella swinhoei* (Kobayashi *et al*., 1990), and the 48-membered monazomycin, from *Streptomyces mashuensis* (Kuo *et al*., 1990) (see Fig. 76). Swinholide A is an unusual dimeric dilactones with a 2-fold axis of symmetry. Both macrolides endow biological properties, swinholide A showed antitumoral activity, while monazomycin showed antimicrobial activity.

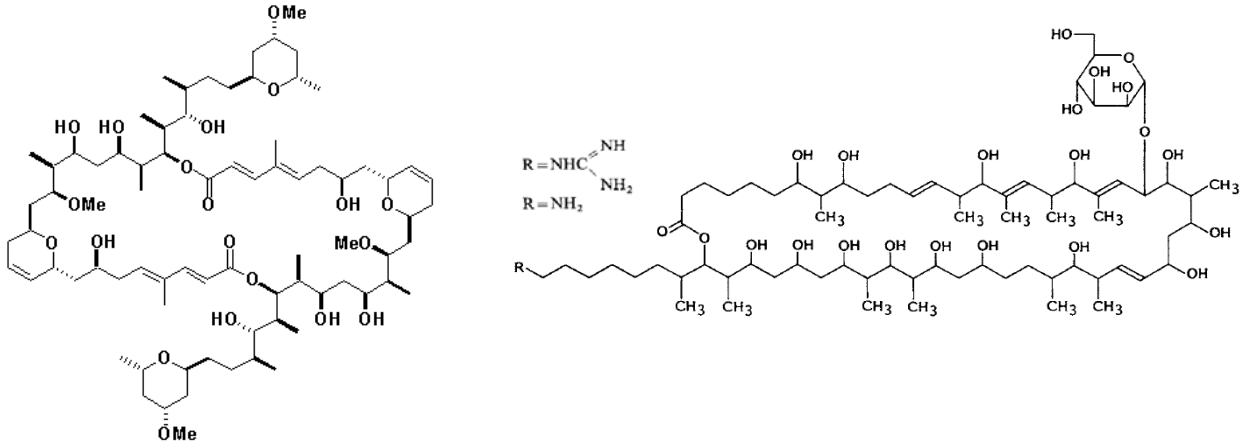


Fig. 76 The structures of swinholide A (on the left) and monazomycin (on the right), isolated from a marine and a soil environment.

The best characterized macrolides, used especially in human therapy, belong to the 14 or the 16-membered class, such as erythromycin or spiramycin, respectively. In the last few decades, novel classes of macrolides with a large macrolactone ring (from 22 up to 48-membered) have been discovered (Yeung and Paterson, 2002; Olano *et al*., 2009). These compounds were principally isolated from marine microorganisms, especially from the order of *Actinomycetes*, with only a few exceptions, for example scytophycins compounds produced by a terrestrial cyanobacterium (Lu *et al*., 2008) or monazomycin compounds from *S. mashuensis* (Kuo *et al*., 1990).

1.1 A new cyclization mechanism for polyketides

The structural elucidation of sambomycin macrolides revealed that the cyclization occurs between the last keto group of the chain and a hydroxyl group located on the starter unit, a methylmalonyl-CoA, thus generating a 50-membered lactone ring.

The last module of a modular type I PKS contains in most of the cases a thioesterase domain, which is responsible for the region and stereospecific formation of the macrolactone ring

between a hydroxyl group, already present in the polyketide chain, and the enzyme-bound oxoester (Kopp and Marahiel, 2007). Alternatives to this mechanism are very rare. They have been described for the ansamycin antibiotics rifamycin and geldanamycin, in which the amide bond forming the macrolactam ring is catalyzed by a discrete amide synthetase (Kopp and Marahiel, 2007).

The first precursor was not supposed to present a hydroxyl group and experiments with isotopically-labelled ^{18}O confirmed that the oxygen involved in the cyclization has been incorporated after the chain elongation (see Chapter 3 and Appendix). This data clearly suggested that a hydroxylation reaction, probably operated by a cytochrome P450, took place to promote the macrocyclization. As proposed by Prof. Greg Challis, a cytochrome P450 would hydroxylate the C-2 of the starter unit when the polyketide chain is still attached to the last ACP domain (module 25). Subsequently, the thioesterase will catalyse the cyclization, giving rise to the aglycone structure of sambomycin (Fig. 77).

We have demonstrated that the gene *sam*R0479 codes for this cytochrome P450 (see Chapter 3). Therefore, SAMR0479 proved to be a key enzyme in the biosynthesis of sambomycin macrolides, proposing a novel mechanism for the macrocyclization of polyketides.

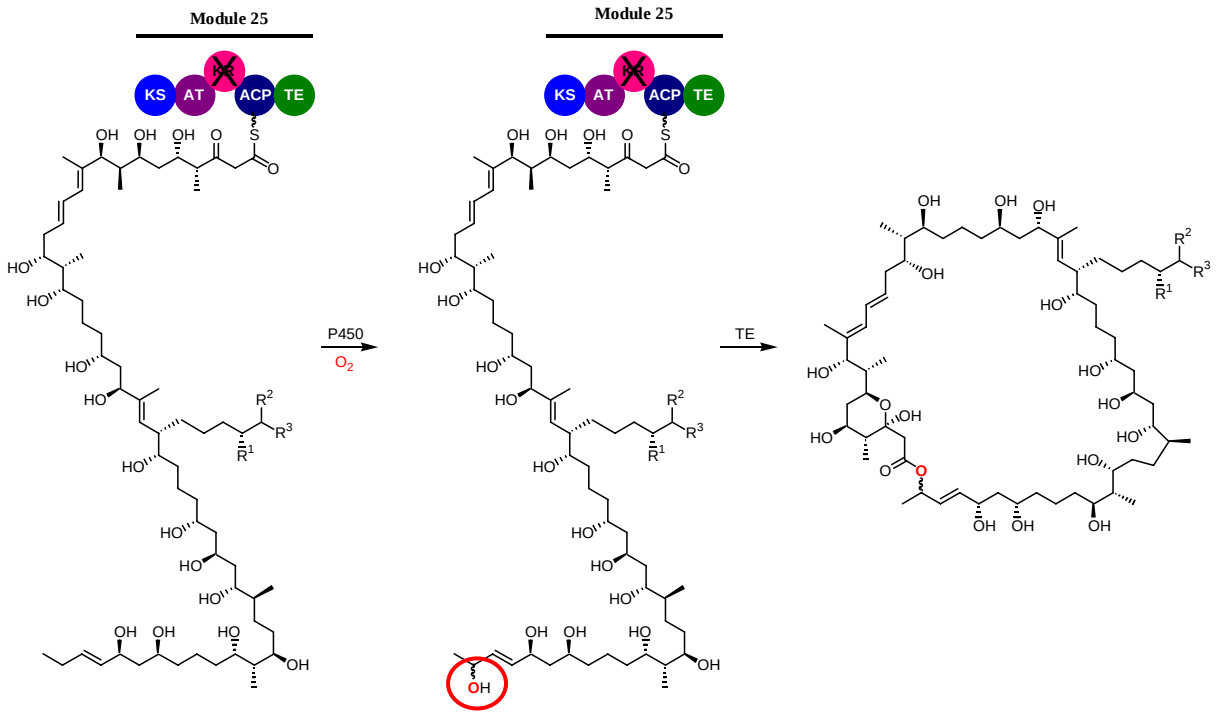


Fig. 77 Hypothetical macrocylization reaction to give rise to the aglycone structure, as proposed by Prof. Greg Challis. The cytochrome P450 would hydroxylate the starter unit, a propionate, to generate the hydroxyl group involved in the intramolecular cyclization catalysed by the thioesterase (TE).

Recently, a fungal cytochrome P450 was found to be responsible to promote a cyclization reaction that generates the diketopiperazine mycotoxin fumitremorgin C, isolated from

Aspergillus fumigatus (Kato *et al*., 2009). However, the *ftm* (for fumitremorgin) cluster belongs to the NRPS class and the hydroxylation necessary for the cyclization does not occur when the polypeptide chain is still attached to the PCP domain. The gene *ftmE*, encoding a cytochrome P450, is indeed involved in the C-N bound formation between the C-3 and the N-4, as shown in Fig. 78. The gene product is essential for the biosynthesis as *sam*R0479 for sambomycin biosynthesis.

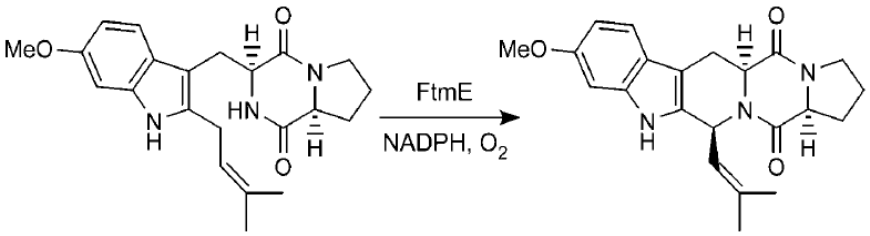


Fig. 78 Representation of the cyclization reaction operated by the cytochrome P450 FtmE of *A. fumigatus* which generates a C-N bound and gives rise to the mycotoxin fumitremorgin C (Kato *et al*., 2009).

However, to the best of our knowledge, SAMR0479 is the first example of bacterial cytochrome P450 involved in the macrocyclization of a polyketide product and it would be interesting to investigate if this mechanism might be exploited for the biosynthesis of other natural products and whether other cytochromes P450 possess the same capability.

Bacterial cytochromes P450 (CYPs) belong to the superfamily of heme-thiolate containing enzymes, and they are often found associated to macrolide biosynthesis. CYPs follow a specific nomenclature, according to the amino acid sequence, the phylogenetic criteria and gene organization (Werck-Reichhart and Feyereisen, 2000). It will be interesting to establish the family and subfamily of SAMR0479 and to check if there are other members with the same characteristic mechanism. However, it is already noteworthy that SAMR0479 has a very low identity/homology with other CYPs involved in secondary metabolite biosynthesis in *Streptomyces* (Table 6), such as EryF for erythromycin, NysL for nystatin or PikC for pikromycin. This finding seems to reinforce the idea that the gene *sam*R0479 codes for a very peculiar cytochrome P450.

1.2 Biosynthesis of an unusual extender unit

The novelties of sambomycin biosynthesis do not end with the cyclization. Structural elucidation also unveiled the nature of the precursor loaded by the AT13 domain, an unusual branched alkyl chain, (see Fig. 52), which has not previously been described in other polyketide structures. Generally, malonyl-CoA, methylmalonyl-CoA and less frequently ethylmalonyl-CoA are the most used substrates in polyketide biosynthesis (Fig. 79A). However, in the last few years, several uncommon extender units have been identified in polyketide structures (Fig. 79A and B) and their biosynthesis has already been characterized (Chan *et al*., 2009). For example, a

novel CoA-linked extender unit was found in the structure of salinisporamide A, containing a chloroethyl moiety, which derives directly from the incorporation of a chloroethylmalonyl-CoA precursor (Eustaquio *et al*., 2009). It was also speculated that many other CoA-linked precursors might be originated from the intermediates of the β-oxydation pathway for straight and branched fatty acid degradation. For example, the structure of polyoxypeptin A contains a peculiar alkyl side chain, probably derived from a 2-(2-methyl-butyl)-malonyl-CoA (Chan *et al*., 2009).
Another class of extender units, recently unravelled, has the peculiarity to be covalently attached to an ACP-like domain (see Fig. 79B). To date, they have only been identified in modular type I PKS products. Members of this new class of precursors are the methoxymalonyl-ACP, as well as the hydroxymalonyl-ACP, the aminomalonyl-ACP and the glyceryl-ACP, which more precisely adds three-carbon unit to the polyketide chain (Chan *et al*., 2009). Several macrolide structures showed to incorporate these kinds of extender units, such as geldanamycin, niddamycin and spiramycin.

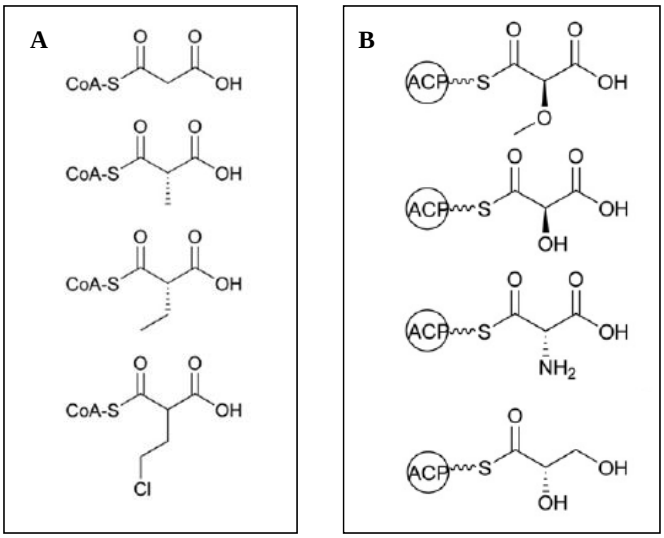


Fig. 79 A) Structure of the CoA-linked extender units, from the top to the bottom: malonyl-CoA, (2*S*)-methylmalonyl-CoA, (2*S*)-ethylmalonyl-CoA and chloroethylmalonyl-CoA. B) Structures of the ACP-linked extender units, from the top to the bottom: (2*R*)-methoxymalonyl-ACP, (2*R*)-hydroxymalonyl-ACP, (2*S*)-aminomalonyl-ACP and glyceryl-ACP (Chan *et al*., 2009).

Unlike the "classical" extender units that are shared between primary and secondary metabolism, all the atypical CoA or ACP-linker precursors are used solely for polyketide production. Therefore, a specific set of genes for their biosynthesis has been found inside the clusters (Chan *et al*., 2009). The ACP-linked extender units need for instance a discrete gene coding for an ACP domain, which tethers the specific substrate. The spiramycin cluster contains five genes involved in the synthesis of the uncommon extender unit methoxymalonyl-ACP; one of these genes encodes an ACP-like domain (Karray *et al.*, 2007).

The only genes in the sambomycin cluster without a putative functional role related to the biosynthesis of sambomycin are *sam*R0482 and *sam*R0483, whose products are homologous to an acyl-CoA synthetase and an acyl-CoA carboxylase, respectively. It has been proposed that an intermediate of the fatty acid pathway is activated to an acyl-CoA precursor by SAMR0482 and it is subsequently carboxylated by SAMR0483 to generate the branched alkyl-CoA chain (see Fig. 80). The AT13 domain would recognize the four types of extender units and loads them on the ACP domain. Hence, both the enzymes for the biosynthesis of the extender unit and the AT13 domain need to have certain flexibility toward different types of branched alkyl chains, giving rise to four forms of sambomycin. Based on this hypothesis and because no genes coding for a discrete ACP-like domain have been identified close to the sambomycin cluster, the substrates produced by these two enzymes clearly seem to belong to the class of CoA-linker extender units.

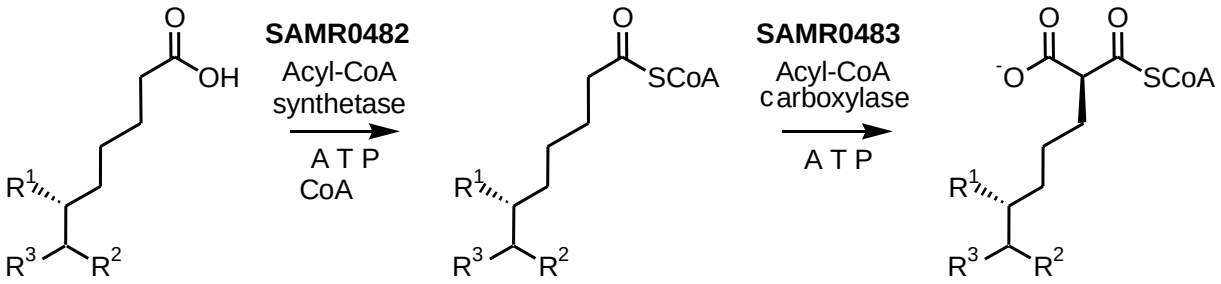


Fig. 80 Hypothetical mechanism for the biosynthesis of the uncommon extender unit of sambomycin macrolides, according to Prof. Greg Challis. SAMR0482 catalyses the activation of the precursor, using as cofactor ATP and the coenzyme A, and SAMR0483 carboxylates the activated precursor, using ATP.

To investigate the role of these two genes and confirm our hypothesis, deletion of the genes was carried out. The deletion of the gene *sam*R0482 did not affect sambomycin production or the morphological growth (Chapter 3, paragraph 3.4). Acyl-CoA synthetase is the enzyme responsible for activation of the precursors of fatty acid biosynthesis, but also of secondary metabolism. Therefore, SAMR0482 seems to be non essential, neither for secondary metabolism in the production of the uncommon extender unit or for primary metabolism in the fatty acid biosynthesis. It is likely that another gene, coding for a protein with the same function, in the genome of *S. ambofaciens* supplies to the lack of SAMR0482. Indeed, several paralogous genes of *sam*R0482 have been identified on the chromosome, some of them associated to other secondary metabolite clusters, for example *sam*L0371 (29/39% of identity/similarity) in the hybrid PKS/NRPS (Table 5 in the introduction) or *sam*R0921 (27/37%) in the congocidine cluster. In particular, SAMR0921 (corresponding to the gene *cgc3**) was recently proposed to be an essential key enzyme in the activation of congocidine precursor since the deletion of the gene impairs congocidine production (Juguet *et al*. 2009).

The replacement of the gene *sam*R0483, coding for a putative acyl-CoA carboxylase, by a resistance cassette was unsuccessful so far. In some cases, it was demonstrated that disruption of these genes does not necessarily prevent metabolite production, but only reduces its yield, as demonstrated for jadomycin biosynthesis in *S. venezuelae* (Han *et al*., 2000). In other cases the deletion prevents antibiotic production, as for the kinamycin cluster (Bertrand Aigle, pers. comm.). Indeed, the product of the gene *alpX* (SAMT0183), in the kinamycin cluster, is a putative functional homologue of SAMR0483 (51/66% of identity/similarity), but its role has not been completely clarified yet. However, both jadomycin and kinamycin are synthesised by a type II PKS, while sambomycin is the product of a modular type I PKS.

SAMR0483 shares a high percentage of identity/similarity (59%/72%) with the product of the gene *pccB* of *S. coelicolor* (see Fig. 81), coding for the β-subunit of a propionylCoA-carboxylase, which has been recently crystallized to analyse the basis of its substrate specificity (Diacovich *et al*., 2004). Indeed, it was established that the β-subunit is responsible for discriminating among different extender units.

```
PccB            MSEPEEQQP---DIHTTAGKLADLRRRIEEATHAGSARAVEKQHAKGKLTARERIDLLLD
SAMR0483        MSLQEPVSPVEESPTSTADRIADLAARHEEAVVLAEKKAADRQHLKGKLTARARIDLLLD

PccB            EGSFVELDEFARHRSTNFGLDANRPYGDGVVTGYGTVDGRPVAVFSQDFTVFGGALGEVY
SAMR0483        PGSFVELDEFVRHRTVEAGI--PRPYGDGVVTGHGTIDGRQVCVFSHDFTTLGGSMGEAF

PccB            GQKIVKVMDFALKTGCPVVGINDSGGARIQEGVASLGAYGEIFRRNTHASGVIPQISLVV
SAMR0483        GSKVVKIYDFAMSVGCPVIGINDSGGARIQEGVMSIAYYTELGVRNVHSSGVIPQISLIM

PccB            GPCAGGAVYSPAITDFTVMVDQTSHMFITGPDVIKTVTGEDVGFEELGGARTHNSTSGVA
SAMR0483        GPCAGGSVYSPALTDFTVMVKDISYMFVTGPEVVSAVMGEQVTAEQLGGPAVHAEVSGNA

PccB            HHMAGDEKDAVEYVKQLLSYLPSNNLSEPPAFPEEADLAVTDEDAELDTIVPDSANQPYD
SAMR0483        HYVGDDEQDAISWVQTLLGYLPPNNLDPAPVYDHDCAPGITEADLALDTVIPDSEQQVYD

PccB            MHSVIEHVLDDAEFFETQPLFAPNILTGFGRVEGRPVGIVANQPMQFAGCLDITASEKAA
SAMR0483        MADVITAVLDDGDYLEIHPDFARNIICALGRVEGHSVAVVANQPRHLAGVLDIDASEKAA

PccB            RFVRTCDAFNVPVLTFVDVPGFLPGVDQEHDGIIRRGAKLIFAYAEATVPLITVITRKAF
SAMR0483        RFIRFCDSFNIPVLTFMDVPGYLPGVGQEHQGIIRRGIKLFYAYAESTVPKITVITRKAY

                     *
PccB            GGAYDVMGSKHLGADLNLAWPTAQIAVMGAQGAVNILHRRTIADAG-DDAEATRARLIQE
SAMR0483        GGGYAVMGSRQIGADRVMAWPTAEIAVMGANSAVPILHRRELAAVPPEERAAVKENLVDD

PccB            YEDALLNPYTAAERGYVDAVIMPSDTRRHIVRGLRQLRTKRESLPPKKHGNIPL
SAMR0483        YRRRFGNPYEAAAHGYVDMVISPSRTRYEVARALASLRNKRQARPARKHGNIPL
```

Fig. 81 Sequence alignment of PccB from *S. coelicolor* and SAMR0483. The amino acids important for the acyl-CoA binding pocket are underlined in green. The residue marked with an asterisk determines the shape of the acyl-CoA pocket, thus the substrate specificity, according to Diacovich *et al*. (2004).

Diacovich and co-workers (2004) identified the amino acid residues involved in the binding of the acyl-CoA (Fig. 81) as well as those implicated in the binding of the biotin cofactor. More important, they identified a residue, at position 422 with regard to PccB (indicated with an asterisk in Fig. 81), which would define the shape of the acyl-CoA pocket, thus determining the choice of the substrate for the carboxylation. This position is occupied by a small amino acid,

for example an asparagine for PccB, when an extender unit, different than an acetate, is used; otherwise the position is occupied by a larger hydrophobic amino acid. The sequence of SAMR0483 contains an alanine residue at position 422, an even smaller amino acid than Asp, which would suggest the binding of an extender unit with a particular size as for example a branched alkyl-CoA chain.

To confirm our hypothesis we could study the *in vitro* activity of both enzymes, as well as the *in vivo* activity by heterologous expression in a host, naturally lacking this metabolic capability. This latter strategy was successfully applied to produce (2*S*)-methylmalonyl-CoA in *E. coli* (Pfeifer *et al*., 2001) or *S. cerevisiae* (Mutka *et al*., 2006). In both cases, the genes coding for a propionyl-CoA synthase along with the β-subunit of a propionyl-CoA carboxylase were introduced and expressed in the host strain, grown on a medium supplemented with propionate.

1.3 Glycosylation of the sambomycin aglycone

The *in silico* analysis of the tailoring genes predicted the incorporation of a mycaminose residue in the expected metabolite and the structural determination confirmed our prediction (see Fig. 51). We demonstrated that the gene *sam*R0481 encodes the glycosyltransferase responsible for the sugar attachment (see Chapter 2, paragraph 2.3).

It has been observed that GTs often need auxiliary proteins to glycosylate their acceptors, especially when the donor substrate is an amino sugar, like desosamine or mycaminose (Hong *et al*., 2007). The presence of the genes coding for these enzymes has been confirmed for pikromycin (DesVIII), tylosin (TylMIII), erythromycin (EryCII) and spiramycin cluster (Orf2*c and Orf16), all containing an amino sugar residue (Borisova *et al*., 2004; Melançon *et al*., 2004; Yuan *et al*., 2005; Karray *et al*., 2007). However, no homologous gene has been identified in the sambomycin cluster. The auxiliary protein Orf16 in the spiramycin cluster is responsible for the attachment of the mycaminose residue. It might be possible that the activity of SAMR0481, the GT in the sambomycin cluster, is also regulated by Orf16 in a sort of cross-regulation between the two clusters. A deletion of the gene *orf16* in the mutant ATCC/OE484 and the analysis of the metabolite product will presumably answer this question. Likewise, the analysis of the expression of *orf16* in the overexpressing strain might indicate a more direct cross-regulation between the two clusters.

Five genes, namely *sam*R0487, *sam*R0486, *sam*R0473, *sam*R0480 and *sam*R0472, in this order, have been proposed to synthesise the sugar donor mycaminose of sambomycin, on the basis of sequence similarities with protein databases. Paralogous of these genes have been identified in the spiramycin cluster (*orf2*, *orf3*, *orf4*, *orf5* and *orf9c*; Fig. 41). However, the order of the genes in the two clusters is different to hypothesise just a duplication event (see schema in Karray *et al*., 2007), unless perhaps for the genes *sam*R0487 and *sam*R0486, corresponding to *orf4* and *orf5* in the spiramycin cluster. Therefore, the other sugar genes have probably been acquired through different events of horizontal gene transfer. It is quite surprising that both spiramycin

and sambomycin clusters need a specific set of genes for the biosynthesis of the sugar, instead of sharing one. Therefore, the question arises whether the deletion of one of these genes in one of the cluster might be complemented by the gene in the other cluster. We already know that deletion of *orf2* and *orf3* abolished spiramycin production, thus *sam*R0480 and *sam*R0473, respectively, can not complement for the loss of function (Karray *et al*., 2007). This might be due to a problem of gene expression, since sambomycin cluster is silent or low expressed in certain conditions. To confirm this result, it will be necessary to delete *orf2* or *orf3* in the strain overexpressing the LAL regulator.

2. Macrolides: biological properties and resistance mechanisms

2.1 Biological properties

Sambomycin compounds showed to endow biological activity against Gram-positive bacteria. It is noteworthy that the IC_{90} value of sambomycin C/D is not remarkable comparing to the references (i.e. vancomycin), hence exploitation of these compounds as antibacterial agents is unlikely, unless further structural modification may improve the biological activity.

Macrolides, especially the 14 and 16-membered, have been extensively used as antibacterial agents for the treatment of respiratory infections. They interfere at the translational level, blocking the protein synthesis (Douthwaite and Champney, 2001). In particular, they bind the 23S rRNA component of the 50S ribosome, preventing the loading of a new amino acid. The reactive groups of the sugar moiety and of the lactone ring (i.e. OH groups) form strong hydrogen bounds with the peptidyltransferase cavity, as shown by the crystal structure of a 50S ribosome binding several 15 or 16-membered macrolides (Hansen *et al*., 2002). Macrolides are also capable of inhibiting the 50S ribosomal subunit assembly, with the same effect on cell viability (Mascaretti, 2003). However, there is no evidence that this mode of action is the same for larger macrolide; it might be that the significant size of the lactone ring makes them unable to bind to the ribosome and/or enter in the peptidyltransferase cavity.

More interestingly, sambomycin showed to have antiproliferative activity, as described for the majority of the large marine macrolides (Yeung and Paterson, 2002; Olano *et al*., 2009); even though also small macrolides can endow potent antitumoral activity, like the 12-membered macrolide pladienolide (Machida *et al*., 2008). Sambomycin products were less toxic than the control agent, doxorubicin (Table 4 in the article) and their IC_{50} are comparable to that of doxorubicin, which is an antitumoral drug already used in human therapy. These findings are very promising and in the future further tests on other cancer cell lines are planned. Besides, it will be interesting to investigate in more details the mechanism of action of sambomycin compounds as anticancer, but also as antibacterial agents. The mode of action of some antitumoral macrolides has been already characterized and the targets are mostly actin and

tubulin, the fundamental components of the eukaryotic cytoskeleton (Yeung and Paterson, 2002). Actin can be present in a monomeric globular form (G-actin) or can polymerize in a filamentous shape (F-actin). Swinholide A prevents actin polymerization, while the 24-membered aplyronine A depolymerises F-actin (Yeung and Paterson, 2002). Instead, other macrolides destabilize tubulin provoking cell cycle arrest and apoptosis, such as the 16-membered epothilone (Marinelli, 2009). Nevertheless, not all of the macrolides targets are known and if sambomycin compounds will disclose a novel mechanism of action, due to the original structure, their exploitation will become even more interested.

2.2 Resistence mechanisms

Three mechanisms of self-resistance have been described for macrolide producer strains: modification of the target, by methylation of the adenine residue A2058 in the 23S rRNA, inactivation of the macrolide by glycosylation and use of active efflux pumps, belonging to the ABC transporter family (Walsh, 2003). Two genes in the sambomycin cluster, *sam*R0470 and *sam*R0471, have been proposed to form an ABC transporter, where SAMR0470 is the permease subunit and SAMR0471 the ATP-binding subunit of a type I ABC transporter. This class has been identified in several clusters producing antitumoral agents, such as daunorubicin and mithramycin (Mendez and Salas, 2001), in which the genes were essential for conferring resistance.

However, neither sambomycin production nor its export seemed to be affected by the deletion of *sam*R0470 and *sam*R0471. This result indicates that these genes are dispensable and suggests the participation of another mechanism of self-protection or of transport located elsewhere on the chromosome of *S. ambofaciens*. Several genes conferring resistance to macrolides have been identified in *S. ambofaciens*, in particular associated to spiramycin cluster which contains an ABC transporter and a methyltransferase encoding genes (Richardson *et al*., 1987; Gourmelen *et al*., 1998; Karray *et al*., 2007). It is likely that one of these mechanisms can take part in the resistance to sambomycin when the ABC transporter SAMR0470-1 is not functional. Indeed, similarly, when the gene *srmA*, coding for a methyltransferase in the spiramycin cluster, is deleted, the strain is still able to produce the macrolide, suggesting the presence of another resistance system (Pernodet *et al*., 1999).

3. The LAL regulator is a positive pathway-specific regulator of the sambomycin cluster

This thesis work proved that a rational manipulation of the regulatory network, in our case involving a pathway-specific regulator, can be a successful strategy to activate a silent cluster. This strategy has been employed for the first time in bacteria and in particular in *Streptomyces* and it seems to be a promising solution to awakening other cryptic and silent clusters. The

sambomycin cluster contains three putative regulatory genes, *sam*R0468, *sam*R0469 and *sam*R0484. Overexpression of the latter enabled us to promote the expression of the biosynthetic genes in the cluster. Thus, we have tried to characterize in more details the function of SAMR0484 and its role in the regulation of the biosynthesis of sambomycin macrolides.

Based on sequence comparison, SAMR0484 belongs to the LAL regulatory family and here we have shown that it acts as an essential positive transcriptional regulator, similarly to other members of this family already described in literature. Moreover, this work demonstrated that SAMR0484 is a pathway-specific regulator of the sambomycin cluster (Chapter 4). Specifically, SAMR0484 controls, directly or indirectly, the expression of nearly all the genes in sambomycin cluster; it activates the PKS genes and the cytochrome P450 genes, while it enhances the transcription of the other genes (Fig. 66). We have not been able to verify yet if SAMR0484 regulates its own expression, but most probably another global regulator controls its transcription, as suggested by Hur *et al*. (2008) for the LAL regulator TmcN. However, no evident consensus recognition sequences for pleiotropic regulators such as PhoP or DasR have been identified in the upstream region of *sam*R0484. Indeed, pathway-specific regulators are part of a more complex regulatory network that might involve nutritional signals or pleiotropic regulators. For example, the regulators for actinorhodin and undecylprodigiosin biosynthesis are tightly regulated by AfsR, PhoP and DasR (Santos-Beneit *et al*., 2009; Rigali *et al*., 2006).

LAL regulators are one of the first class of pathway-specific regulators identified in modular type I PKS clusters and to date their direct targets have not been identified and/or characterized, thereby our interest in investigating SAMR0484. LAL proteins contain a HTH domain in the C-terminus, typical of transcriptional regulators (De Schrijver and De Mot, 1999). To identify the DNA targets of SAMR0484, we have used a reporter system based on *redD* of *S. coelicolor* as reporter gene (van Wezel *et al*., 2000). SAMR0484 was shown to recognize specifically a yet unknown sequence in the intergenic region upstream the PKS genes *sam*R0467 and *sam*R0477. The intergenic region of the resistance gene *sam*R0471 was also tested, but the result obtained was ambiguous. It would be interesting in the future, to test other intergenic regions, especially those upstream the cytochrome P450 gene *sam*R0479, responsible for the cyclization, and of *sam*R0482 and *sam*R0483, which seem to be involved in the biosynthesis of the uncommon extender unit. Transcriptional analyses showed that these genes were the most affected by the overexpression of the LAL regulator, along with the PKS genes.

In support to these results, *in vitro* studies of the LAL regulator by gel retardation assays using the intergenic regions of the PKS genes *sam*R0467 and *sam*R0477 along with the truncated form of SAMR0484 were carried out. We were unable to obtain a clear shift that would have decisively demonstrated the role of SAMR0484 (Fig. 70). However, preliminary data with the truncated form of SAMR0484 seems to suggest a possible interaction with these DNA sequences. For the future, the conditions of protein expression and purification, as well as those for the *in vitro* binding need to be significantly improved. Otherwise, we could attempt DNA-

protein cross linking experiments to prove the interaction of SAMR0484 with these sequences. Eventually, a consensus binding sequence for the LAL regulator can be identified by footprinting experiences and the genome of *S. ambofaciens* and of other organisms coding for a LAL regulator could be scanned in search for other LAL-DNA targets using the PREDector software system, for instance (Hiard *et al*., 2007).

It is also possible that a cofactor or another substrate is missing, thus explaining the lack of binding. Indeed, the prototype of LAL regulator, MalT from *E. coli*, has been shown to need ATP and maltotriose to bind to its DNA targets and to regulate the maltose regulon (Richet and Raibaud, 1989). Maltotriose is an intermediate of maltose metabolism and we could hypothesise that an intermediate of sambomycin biosynthesis might also interact and regulate SAMR0484, thus providing a positive feedback on gene expression. A similar mechanism has been observed for ActR, a TetR-like repressor in the actinorhodin cluster (Tahlan *et al*., 2008). However, this would not explain the positive result obtained with the reporter system in *S. coelicolor*, in which the sambomycin cluster is absent. Therefore, it is likely that another yet unknown substrate, also produced by *S. coelicolor*, is involved in the regulation of SAMR0484. Some of the LAL regulators have been described to contain a TPR domain near their C-terminus (Rascher *et al*., 2003; Kitani *et al*., 2009), typical of protein-protein interaction (Marck *et al*., 1993) which can modulate the binding of the regulator to its target. The analysis of SAMR0484 also highlighted the presence of this particular domain, which reinforces the hypothesis of an unknown factor participating in the regulation of the sambomycin cluster. A preliminary EMSA assay, using the intergenic region of *sam*R477 and the crude extracts of *S. coelicolor* M512/OE484 and M512/pIB139, produced a shift only in the presence of the LAL regulator (data not shown), reinforcing our hypothesis. Eventually, further experiments will confirm this result and will provide us better insights in the mechanism of regulation of SAMR0484.

Another common feature of LAL regulator is the presence of one or more TTA codons in their coding sequence. Some pathway-specific regulator encoding genes have shown to contain this rare codon (Chandra and Chater, 2008), indicating that a post-transcriptional regulation operated by *bldA* is at the basis of secondary metabolite production (Leskiw *et al*., 1991). The overexpression of *sam*R0484 would overcome the control of *bldA*, as already demonstrated by Passantino *et al*. (1991) and White and Bibb (1997), and as it has been successfully demonstrated with the discovery of sambomycin. A strategy to verify if SAMR0484 is really subjected to *bldA* control or whether another mechanism is involved could be to substitute the TTA codon present at the beginning of *sam*R0484 by an alternative leucine codon (Fernandez-Moreno *et al*., 1991) and analyse the expression of the cluster in MP5 and HT media, in which the cluster is silent.

It is interesting to highlight that we have observed a sort of cross-regulation between the spiramycin and the sambomycin cluster which might depend on SAMR0484 (see Chapter 4, paragraph 4.5). Indeed, the overexepression of the LAL regulator showed to significantly delay

and in some cases to abolish spiramycin production, even when sambomycin is not synthesised as in the mutant ATCC/OE484/Δ467::kan. A similar phenotype was previously observed in our laboratory by analysing several mutants of *S. ambofaciens* derived from genome instability. The amplification of a DNA locus of 89 kb, named AUD205, prevented the production of the macrolide spiramycin (Dary *et al*., 1992). Antibiotic production was restored when the strain lost this amplification. A 3 kb sequence of this AUD205 was sequenced and showed to be homologous to PKS genes such as those found in erythromycin and rapamycin clusters (Aigle *et al*., 1996). More precisely, the sequence revealed to be one of the modules of the PKS gene *sam*R0475 of the sambomycin cluster. The *Bam*HI digestion profile of the ADS205 (Dary *et al*., 1992) suggests that a large part of the sambomycin cluster is amplified, in particular most of the PKS genes. The hypotheses beyond this phenomenon formulated at that moment were a shunt of metabolites, towards one of the clusters, or the involvement of a regulatory gene, acting negatively on spiramycin biosynthesis. We have shown that the cause is not a shunt of metabolites, but most probably the involvement of a regulatory factor, such as SAMR0484. However, the amplification of the AUD205 does not involve the gene *sam*R0484. We can thus think that two different causes are the origin of the same phenomenon: in the mutant ATCC/OE484 it might be due to the LAL regulator, while for the mutant ADS205 it can be due to another unknown regulatory factor.

4. Combinatorial biosynthesis of the sambomycin gene cluster

Combinatorial biosynthesis is a new exciting approach to generate modified forms of known compounds, with the aim to improve their biological or pharmacokinetics properties, or in order to generate structural diversity by synthesising novel non-natural products, which might find an application in human therapy or in other fields. A better understanding of the structure-bioactivity relationship of the known compounds will provide very useful in the choice of the target modifications. Combinatorial biosynthesis is not as successful as nature in producing potent biomolecules, because natural products are the result of a selectable evolution process in response to a specific environmental role. However, engineering known compounds to improve their therapeutic use is still an important step in drug development (Floss, 2006).

Therefore, on the one hand we could modify the natural structure of sambomycin to obtain a more effective antibacterial agent or a more specific and less toxic antitumoral agent; on the other hand we could exploit the biosynthetic novelties identified in the sambomycin cluster to create novel non natural compounds.

Most of the therapeutic drugs are glycosylated compounds in which the sugar component is essential for their biological activity. The typology of the glycosyl moiety significantly affects the pharmacokinetic properties, the biological targets and the mechanism of action of the

molecules, thus glycosylation have become a fascinating tool in the development of novel bioactive derivatives (Härle and Bechthold, 2009). Hence, analogous of sambomycin can be obtained by engineering the glycosylation pattern. After being attached to the aglycone by the glycosyltransferase, the sugar moiety can undergo further modifications, mainly methylation and acylation on the OH or NH_2 groups (Salas and Mendez, 2007). Heterologous methyltransferases, as well as acyltransferases, aminotransferases might be expressed together with the sambomycin cluster. Another approach implies a combination of gene inactivation and gene expression, with the intent of obtaining a novel product (Salas and Mendez, 2007). For example, the genes involved in the mycaminose biosynthesis can be deleted and substituted by genes responsible for the biosynthesis of a different sugar. In the same time, we could verify if the glycosyltransferase SAMR0481 owns flexibility towards different sugar residues. This flexibility was already shown for other GTs, on the other hand the specificity towards the receptor (the aglycone) is more strict (Salas and Mendez, 2007). Indeed, in the mutant strain ATCC/OE484/Δ479::spec, unable to catalyse the cyclization of sambomycin polyketide chain, the linear product turned out to be not glycosylated. This indicates both that glycosylation is the last step in the biosynthesis of sambomycin and that SAMR0481 specifically recognized a cyclic structure as its receptor molecule to attach the mycaminose moiety.

In the second case, i.e. to generate novel non-natural products, the peculiarities disclosed by sambomycin structure can be exploited for other secondary metabolite pathways, especially if these structural characteristics are related to the mode of action of sambomycin. For example, the gene *sam*R0479, coding for the cytochrome P450 responsible for the atypical macrocylization, might be expressed in a macrocyclic-producing strain and observe if the cyclization would occur in a different way than expected, requiring the incorporation of an additional oxygen atom. In this way, we would also be able to establish if there is another factor influencing the cyclization in sambomycin cluster until now unidentified.

Another approach might be “reprogramming” the modular PKS by altering the substrate specificity of a known type I PKS cluster. The AT13 domain, which recognize a branched alkyl chain, can be exchanged (“module swapping”) with an AT domain already present in a PKS gene, thus modifying the final structure. A similar strategy was successfully applied to produce a derivative of erythromycin (Petkovic *et al*., 2003). In case it would be necessary, the genes *sam*R0482 and *sam*R0483, probably involved in the activation of the atypical precursor, could be also expressed in the host producing strain.

5. The role of sambomycin in nature

In the classical view, microorganisms produce antimicrobial agents to fend off other organisms living in the same microenvironment. Typically, streptomycetes secrete secondary metabolites

during the transition from vegetative growth to morphological differentiation. Some of these compounds probably act as bacterial weapons to protect against other predators, since at this stage of growth streptomycetes are more vulnerable. According to the microorganism that threatens the survival of *Streptomyces*, a different kind of metabolite is probably produced and secreted. This might explain the mechanism of antagonism or cross-regulation that we have observed between the spiramycin and the sambomycin clusters (see Chapter 4, paragraph 4.5). When sambomycin is synthesised, it seems that the production of the other macrolide is delayed or completely blocked; vice versa, when spiramycin is normally produced sambomycin is not detectable. The basis of this phenomenon is not well clarified yet. It might be due to a shunt of metabolites, since both molecules share the same pool of precursors, or a mechanism to avoid energy waste because of sambomycin size. Another hypothesis is the involvement of a regulatory gene, such as SAMR0484 that can control positively the sambomycin cluster and negatively the spiramycin cluster. Preliminary results seem to support the latter. Therefore, it could be that spiramycin is used as a powerful antibacterial agent in certain conditions, while sambomycin has another effect on microbes or another functional role in response to different environmental signals. This is also the reason why it will be interesting to unravel in details the mechanisms of the LAL regulator and in particular to identify what protein or environmental factor controls SAMR0484.

Our anthropocentric point of view on the use of microbial secondary metabolites, i.e. antibacterial, antifungal and antitumoral agents, does not always reflect the real role of these molecules in nature. The term antibiotic principally refers to the therapeutic use of the compound which implies a higher concentration than the environmental one (Aminov, 2009). In the last few years, scientists have focused on unravelling the environmental function of these extraordinary compounds that are not only exploited as bacterial weapons.

Nowadays, it became clear that secondary metabolites are considered as signalling and regulatory molecules, used to communicate inter or intra species in response to a particular environmental and physiological state. On the other hand, resistance genes probably originated and evolved as a mechanism to control and regulate the cross-communication operated by these molecules (Aminov, 2009).

Subinhibitory concentration of antibiotics showed to have a significant effect on the expression of a wide range of genes, involved for example in virulence, biofilm formations, cell development or immunomodulation (Davies *et al*., 2006). In particular, different class of antibiotics, such as macrolides or ansamycins, showed to modulate the transcription of 5% of the promoters in *Salmonella typhimurium* (Goh *et al*., 2002) and more generally, most of the antibiotics interact with the ribosome, which being a universal macromolecule is a suitable multiligand sensor (Aminov, 2009). Another interesting effect of subinhibitory concentration is the increase of antibiotic resistance, either through horizontal gene transfer or point mutations

(Aminov, 2009). This finding might partially explain the spreading of resistance genes among bacteria, but it has also important implication in human therapy.
With the intent to investigate into the environmental role of sambomycin, a simple promoter-*lux* reporter construct would be suitable to analyse the effect of this giant macrolide on the transcriptome of some other bacteria present in its environment. Unravelling its real targets might also provide useful to understand is mode of action as an antitumoral agent. Similarly, it might be interesting to test whether sambomycin can be a factor inducing or enhancing conjugal transfer of resistance or virulence genes.

6. *Streptomyces ambofaciens* and its fascinating secondary metabolites

Streptomyces ambofaciens ATCC23877 was known to produce only two antibiotics, the macrolide spiramycin and the pyrrole amide congocidine. The sequencing of its chromosome in combination with several genomic-guided approaches revealed the presence of other natural products, interesting for their biological properties, but also for their peculiar biosynthetic pathways, enlarging our knowledge on secondary metabolism.
The antibiotic kinamycin, synthesised by a type II PKS cluster, was discovered through an OSMAC approach (Pang *et al*., 2004; Bunet *et al*., 2008). The cluster is regulated by a novel kind of extracellular molecule, most probably a furan-like ligand. The siderophores desferrioxamine E and coelichelin are the products of an NRPS-independent pathway and of an NRPS cluster, respectively, and have been identified by a genome mining approach (Barona-Gomez *et al*., 2006). Coelichelin cluster is one of the first examples of non-linear enzymatic logic in NRPS, since the first module acts iteratively to incorporate the same molecule twice (Lautru *et al*., 2005). Also the congocidine cluster presents an atypical NRPS assembly with discrete genes instead of a modular gene and uncommon precursors, whose biosynthesis still needs to be elucidated (Juguet *et al*., 2009). At last, this thesis work discovered and characterized the sixth natural product of *S. ambofaciens*, a novel 50-membered macrolide. Sambomycin biosynthesis revealed two interesting aspects, concerning the incorporation of an uncommon extender unit and the cyclization reaction, mediated by a cytochrome P450.
Even with this last discovery, the potential and the exploitation of *S. ambofaciens* genome have not reached an end. In the terminal unstable regions of the chromosome the product of a hybrid NRPS/PKS as well as another type I PKS might be interesting to be detected and isolated (Table 5). In the meantime the preliminary information from the core sequence of the chromosome pointed out additional clusters, for a total of 16 secondary metabolite pathways. One of the strategies to identify these cryptic metabolites is based on the construction of a *S. ambofaciens* strain deficient in producing spiramycin, congocidine, kinamycin and sambomycin, whose biological properties can hide a novel one, thus also removing possible competitors for the substrate uptake. Alternatively, the manipulation of the global or pathway-specific regulatory

network can also prove to be a successful tool in the activation of silent clusters. It is more than likely that very soon other metabolites synthesized by *S. ambofaciens* will be isolated and characterized.

APPENDIX

Additional figures

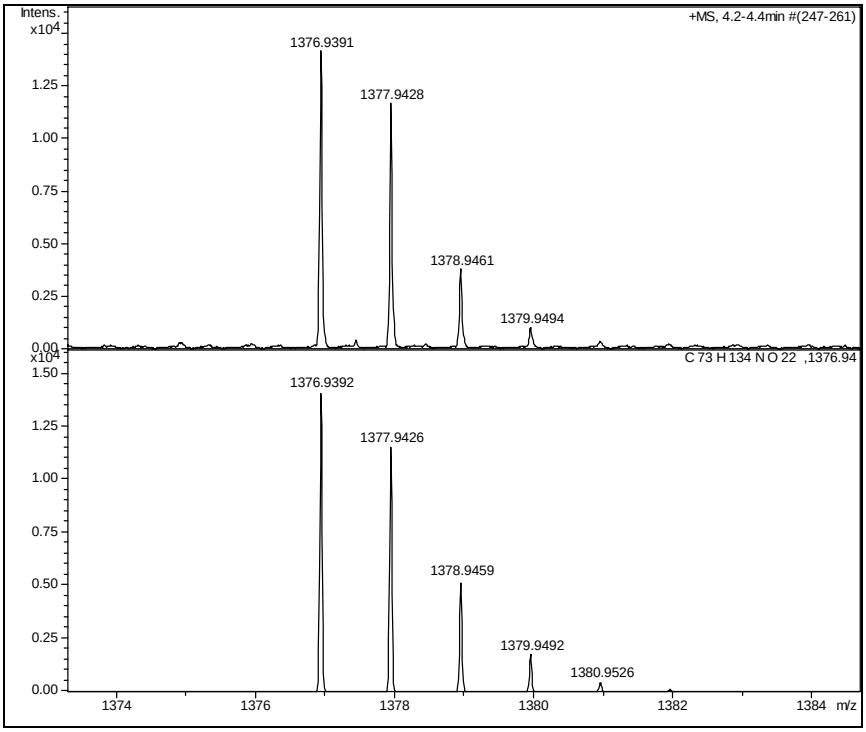


Fig. 1 ESI-TOF mass spectrum of sambomycin A/B (top) and the simulated mass spectrum for the $C_{73}H_{134}NO_{22}$+ ion (bottom).

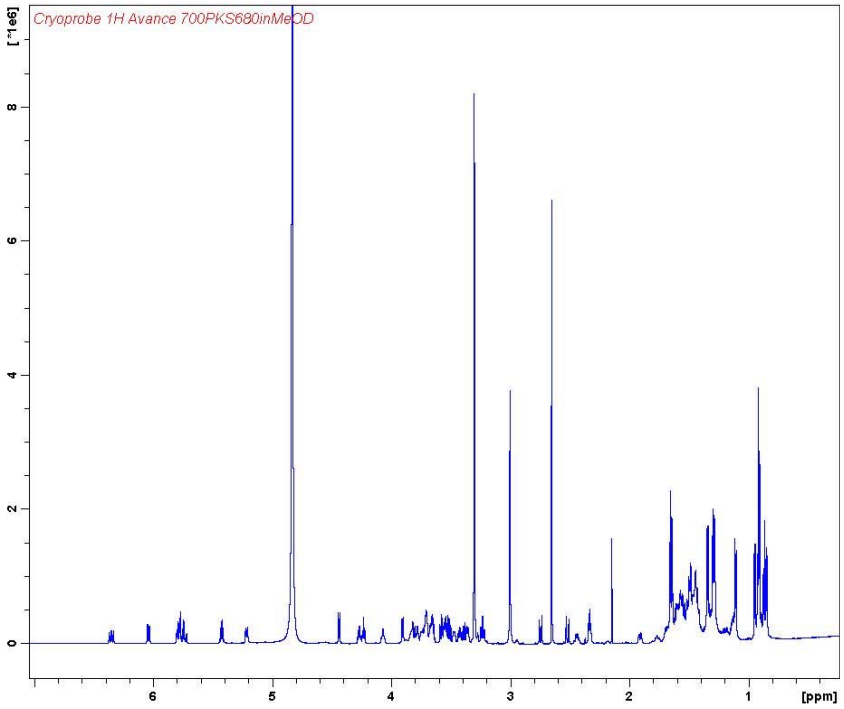


Fig. 2 1H-NMR spectrum (700 MHz) of sambomycin A/B in d_4-MeOH.

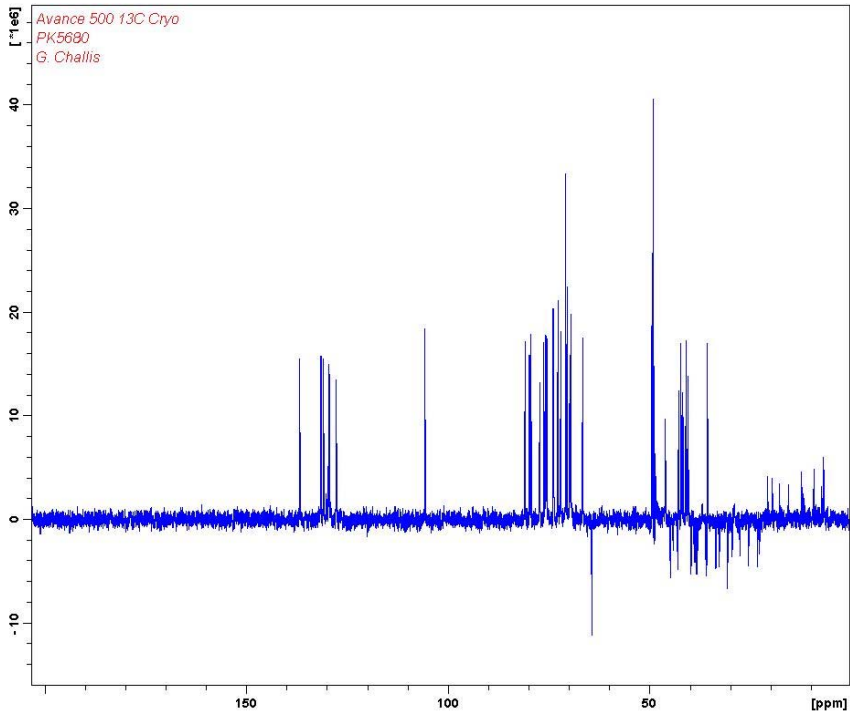


Fig. 3 13C-NMR spectrum (125 MHz) of sambomycin A/B in d_4-MeOH.

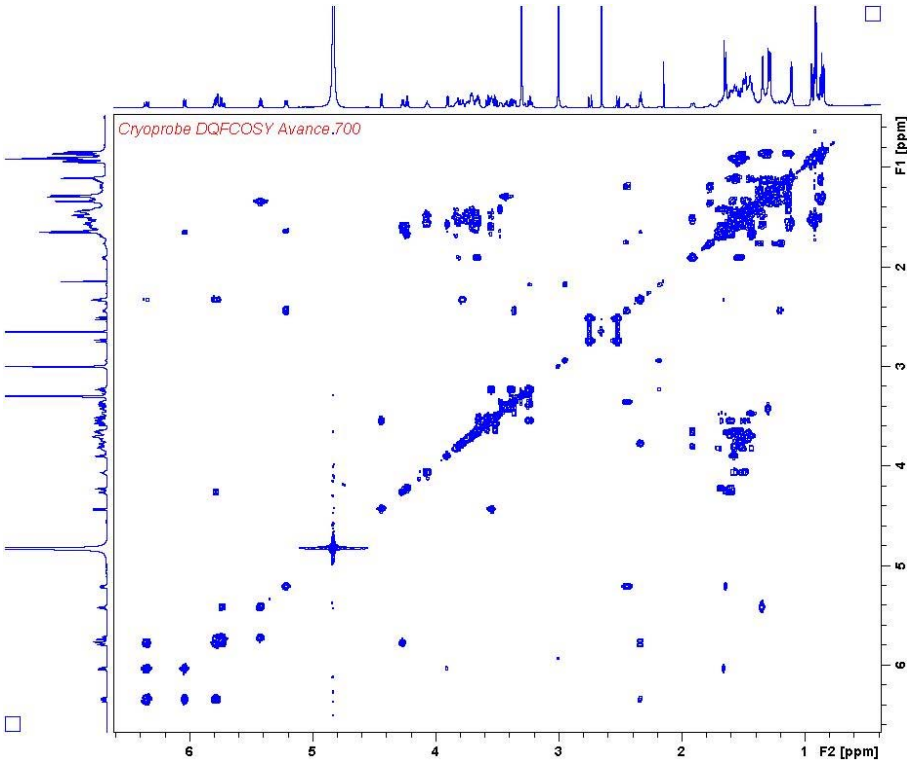


Fig. 4 DQF-COSY spectrum (700 MHz) of sambomycin A/B, in d_4-MeOH.

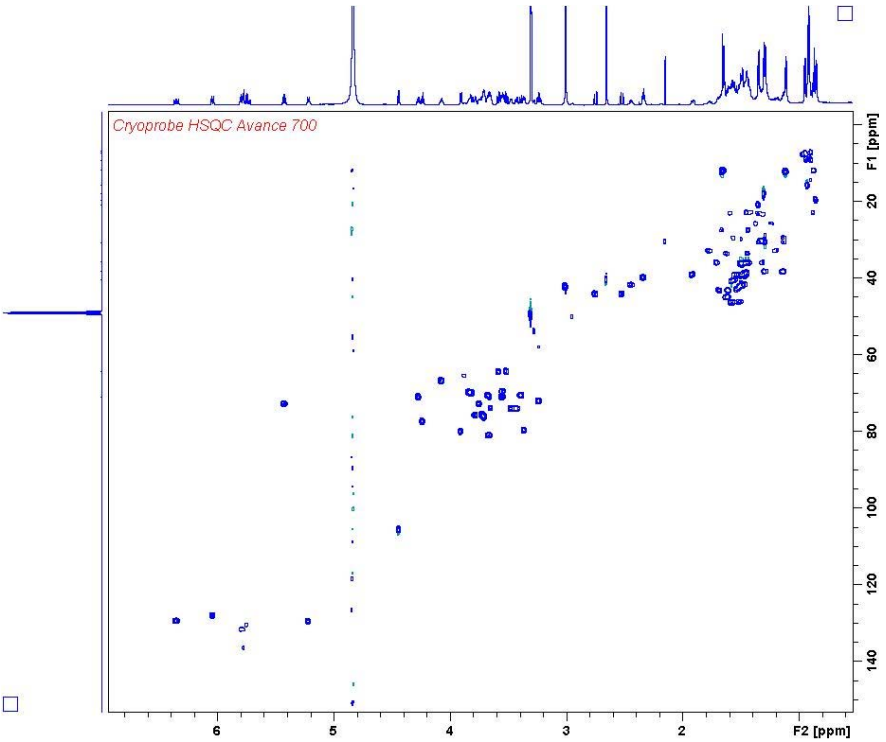


Fig. 5 HSQC spectrum (700/175 MHz) of sambomycin A/B in d_4-MeOH.

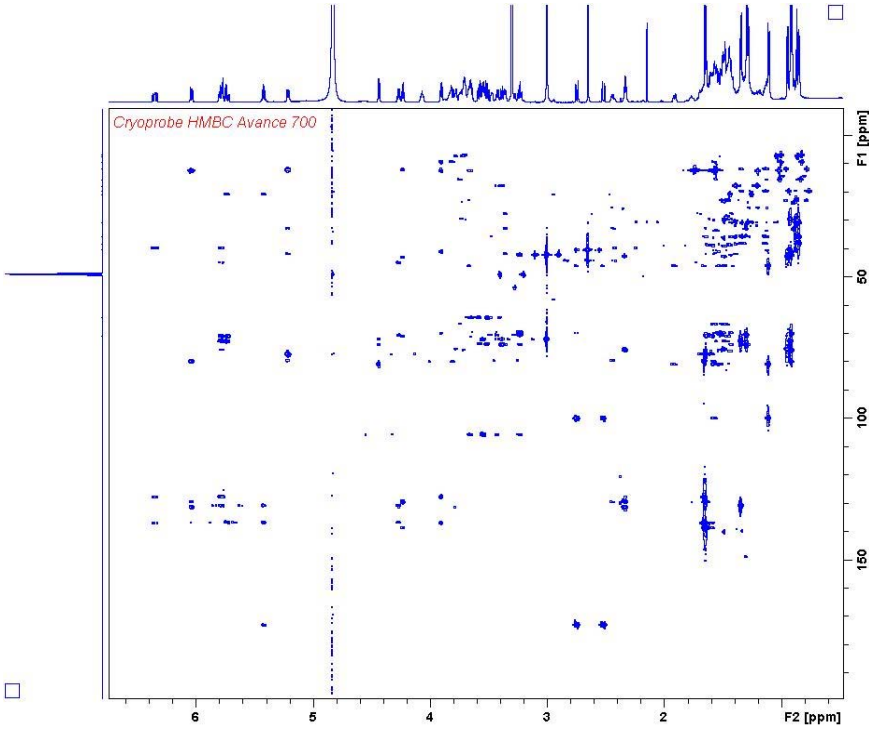


Fig. 6 HMBC spectrum (700/175 MHz) of sambomycin A/B in d_4-MeOH.

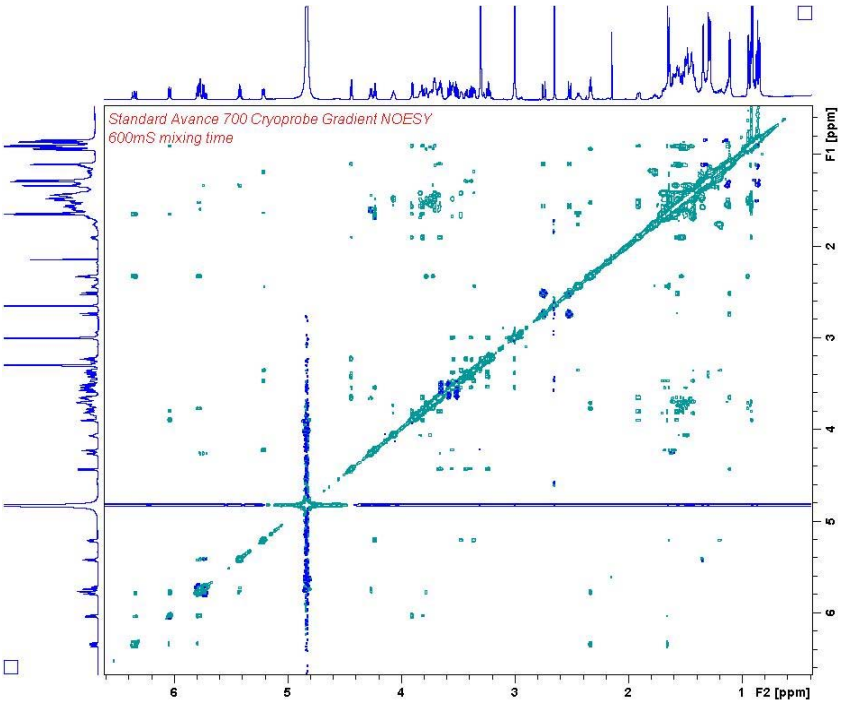


Fig. 7 NOESY spectrum (700 MHz) of sambomycin A/B in d_4-MeOH.

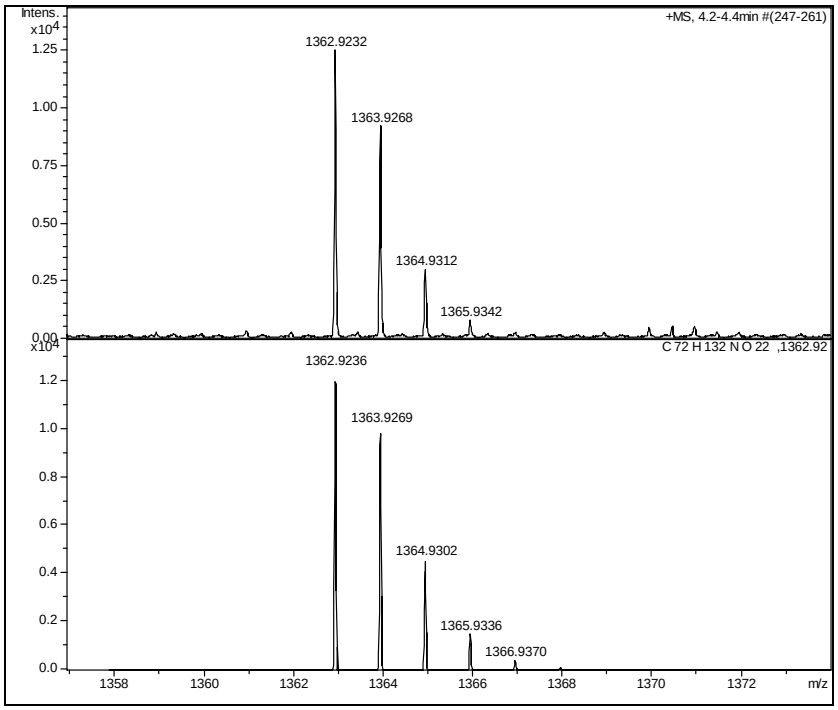


Fig. 8 ESI-TOF mass spectrum of sambomycin C/D (top) and the simulated mass spectrum for the $C_{72}H_{132}NO_{22}^{+}$ ion (bottom).

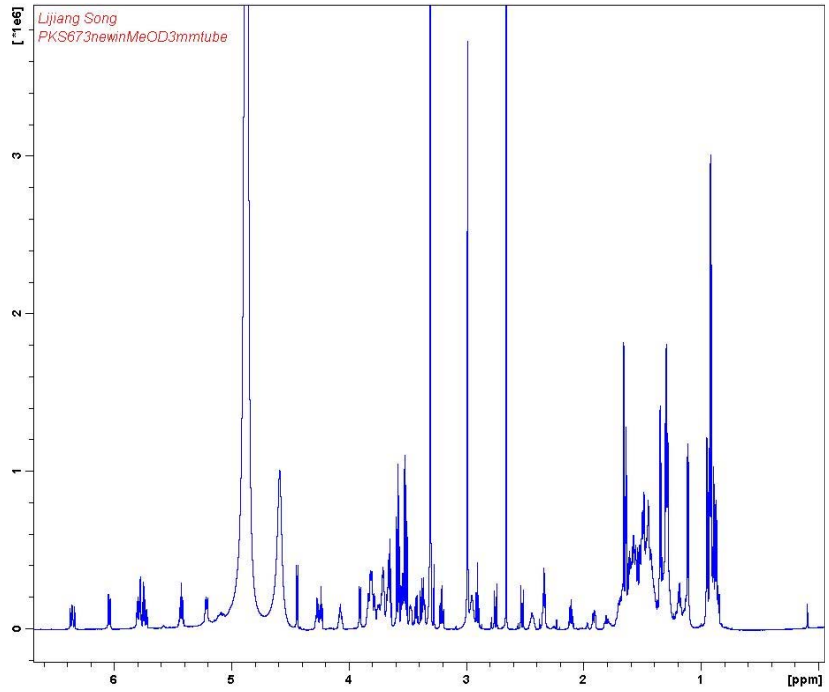


Fig. 9 ^{1}H-NMR spectrum (700 MHz) of sambomycin C/D in d_4-MeOH.

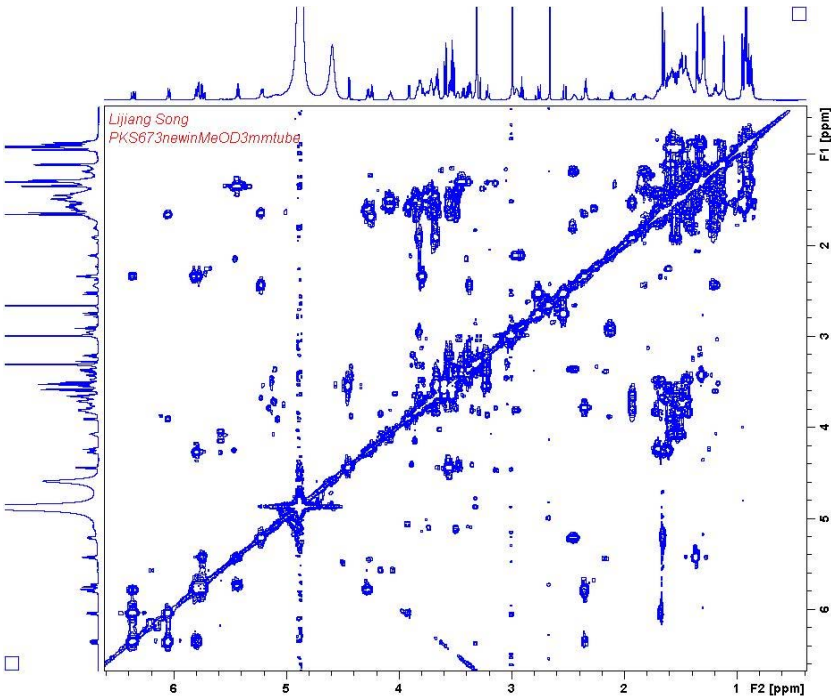


Fig. 10 DQF-COSY spectrum (700 MHz) of sambomycin C/D, in d_4-MeOH.

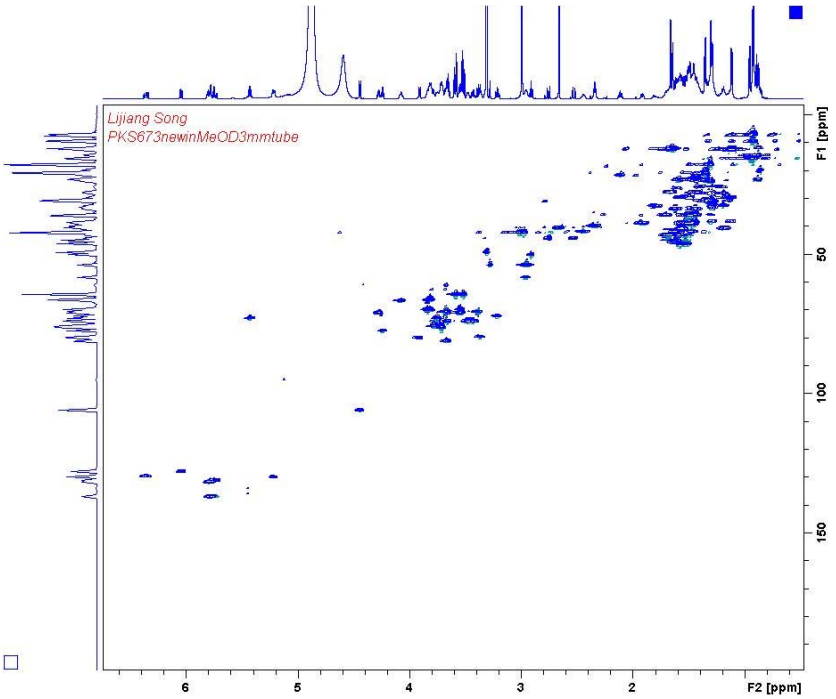


Fig. 11 HSQC spectrum (700/175 MHz) of sambomycin C/D in d_4-MeOH.

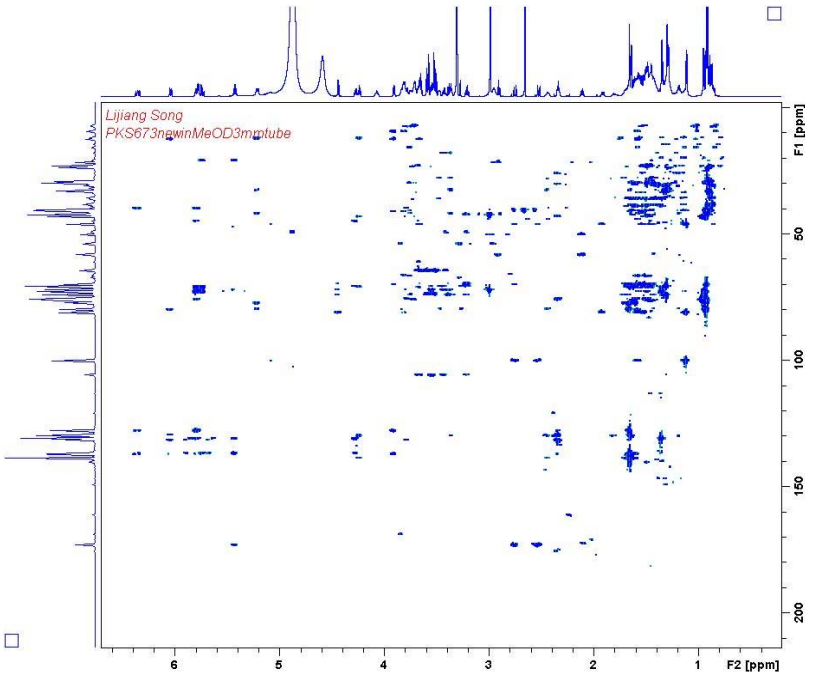


Fig. 12 HMBC spectrum (700/175 MHz) of sambomycin C/D in d_4-MeOH.

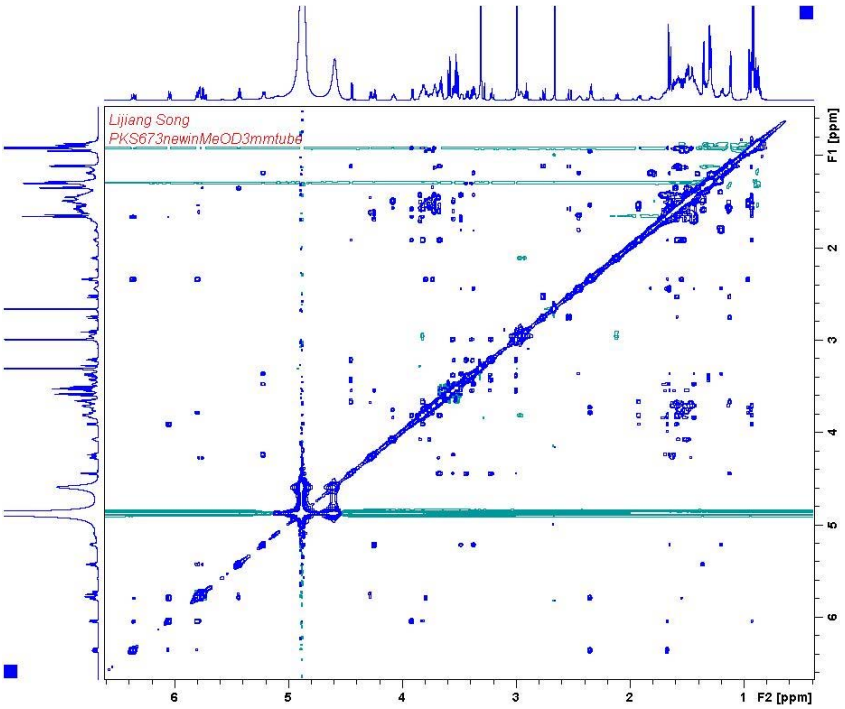


Fig. 13 NOESY spectrum (700 MHz) of sambomycin C/D in d_4-MeOH.

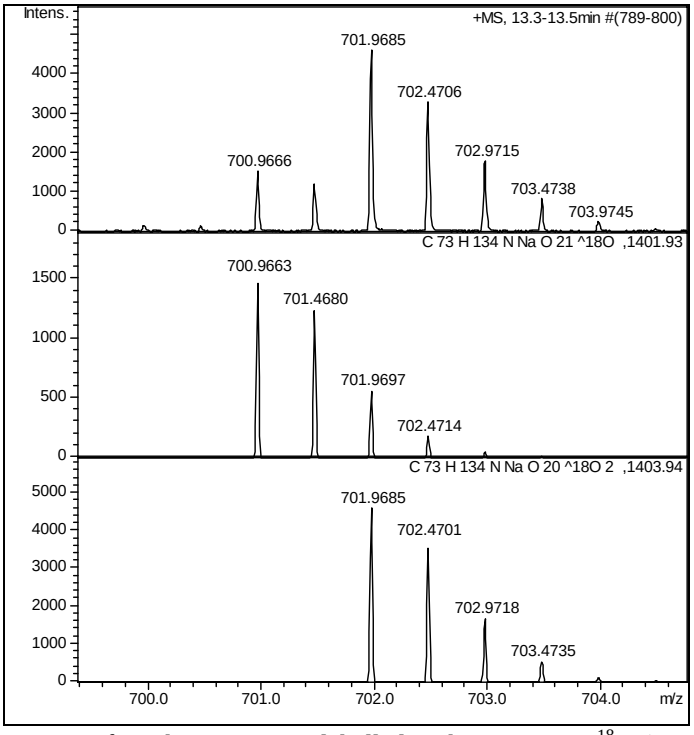


Fig. 14 ESI-TOF mass spectrum of sambomycin A/B labelled with one or two ^{18}O (on top), corresponding to the sodiated form, and their simulated mass spectra (on bottom).

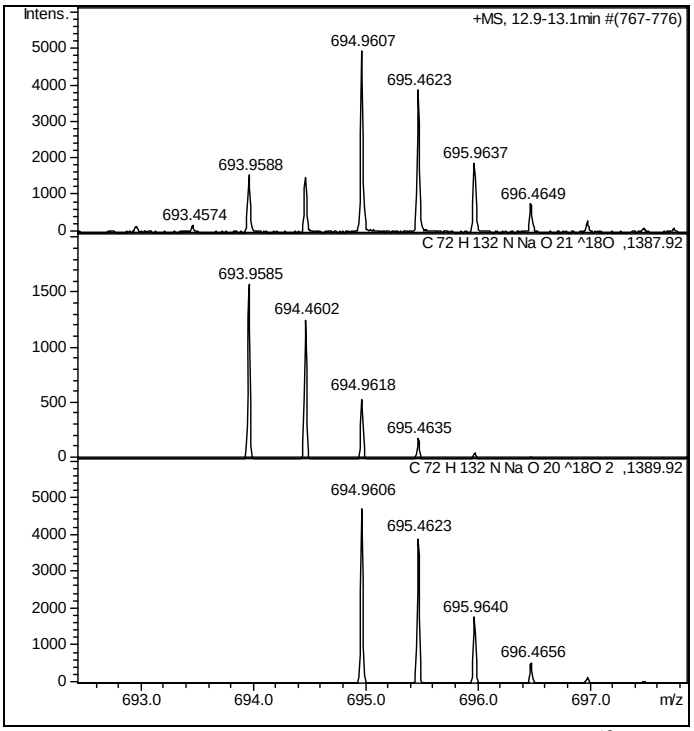


Fig. 15 ESI-TOF mass spectrum of sambomycin C/D labelled with one or two ^{18}O (on top), corresponding to the sodiated form, and their simulated mass spectra (on bottom).

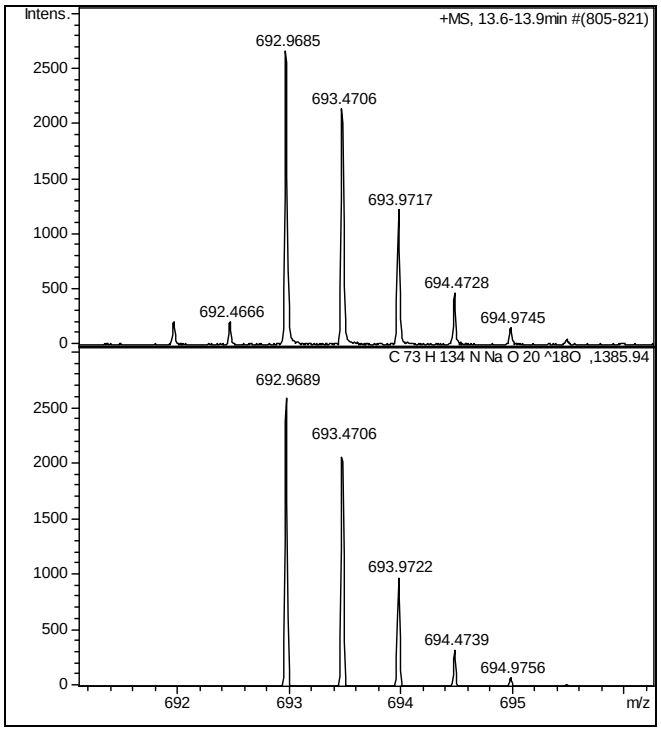


Fig. 16 ESI-TOF mass spectrum of a modified form of sambomycin A/B labelled with one ^{18}O (on top), generating from the mutant strain ATCC/Δ478, and the simulated mass spectra (on bottom), lacking on oxygen atom.

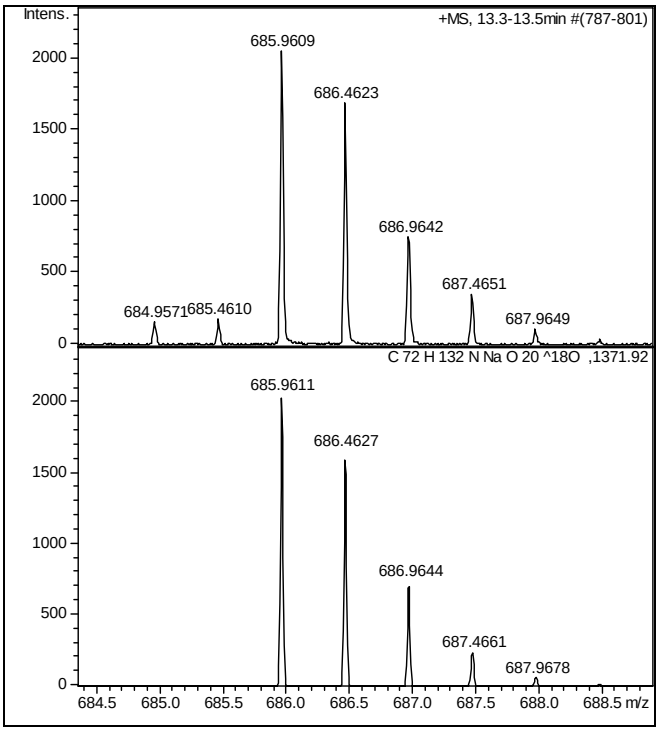


Fig. 17 ESI-TOF mass spectrum of a modified form of sambomycin C/D labelled with one ^{18}O (on top), generating from the mutant strain ATCC/Δ478, and the simulated mass spectra (on bottom), lacking on oxygen atom.

Methods

Extraction of the secretome

A sterile cellophane membrane was placed on the top of the agar plate (HT supplemented with 15 mM $MgCl_2$) and the spores of the strain were spread over it. After several days of incubation at 30°C and according to biological activity time course, the membrane was removed and the agar was scrapped off and extracted in a glass flask with the same volume of methanol (usually 25 ml), under shaking for at least one hour. The methanol extract was collected in a falcon and centrifuged at max speed at 4°C for 15 min to remove any agar traces. Then, it was concentrated to a volume of about 1000-500 µL at the rotavapor and analysed by LC-MS

Reporter system for the transcriptional activity of the LAL regulator

The intergenic regions upstream the genes *sam*R0467, *sam*R0477 and *sam*R0471 were amplified by PCR using primers containing at the 5′ the sites for the restriction enzymes *Bam*HI and *Eco*RI (see Table in the Appendix), unique sites used for the cloning in the vector pIJ2587 (van Wezel *et al.*, 2000). The PCR fragments, named 467redD (748bp), 477redD (620bp) and 4701redD (521bp) were digested by *Bam*HI and *Eco*RI and they were ligated together with the plasmid pIJ2587, digested with the same restriction enzymes. Three recombinant vectors pIJ-R0467, pIJ-R0477 and pIJ-R04701 have been obtained. The constructs were verified by

sequencing. *E. coli* ET12567 cells have been transformed with the recombinant vectors in order to obtain non-methylated DNA. This condition is recommended when protoplast transformation is used for the insertion of exogenous DNA (Oh and Chater, 1997).
Protoplasts of M512/pIB139 and M512/OE484 were prepared as described by Kieser *et al.* (2000), as well as protoplast transformation, which was carried out with the vectors pIJ-R0467, pIJ-R0477 and pIJ-R04701 extracted from *E. coli* ET12567. The transformants were plated on R2YE medium and incubated at 30°C. As indicated in the protocol (Kieser *et al.*, 2000), after 20h of incubation, the plates were flooded with 2.5 ml of SNA supplemented with 12µg/ml of thiostrepton, the selective antibiotic for pIJ2587.

Heterologous expression of SAMR0484 and the truncated form of SAMR0484

The coding sequence of *sam*R0484 was isolated from the plasmid pGEMT-0484 digested by *Nde*I and it was cloned in the expression vector pET15b (Novagen), digested with the same enzyme. The correct orientation of the insert was verified by digestion profile. The accuracy of the construct was verified by sequencing. The expression vector pET15b, as control, and the recombinant vector pET15+484 were used to transform *E. coli* BL21(DE3) competent cells. To improve the level of protein expression, *E. coli* BL21(DE3)/pET15+484 cells, as well as the control, were transformed with the plasmid pSBET, which has an arginine tRNA gene, specific for the AGA and AGG codons (Schenk *et al.*, 1995). Subsequently, *E. coli* BL21(DE3)/pET15+484/pSBET and the control were grown in LB medium supplemented with ampicillin (100µg/ml) and kanamycin (50µg/ml) at 37°C until the optical density measured at 600 nm reached 1. Then the cultures were induced with 0.5 mM of IPTG and incubated for other 3h testing different temperatures, 37°C, 30°C and 24°C, to optimize protein expression. Samples were collected before and after 3h of induction. At the end, the cells were harvested, resuspended in a protein buffer (50 mM Tris HCl,1mM EDTA) and disrupted by sonication. After 20 min of centrifugation at 4°C, samples of the soluble and insoluble protein extract were collected. Aliquots of the samples before and after induction, as well as the soluble and insoluble fractions, were boiled for 5 min together with the same volume of loading dye, before loading them on a 10% SDS-PAGE electrophoresis gel.
The truncated sequence of 484 (T484) was amplified by PCR using the primers T484-F and OE484-R (see Table in the Appendix). After terminal dATP addition, the PCR product (766bp) was ligated with pGEM-Teasy (Promega) and checked by sequencing. The insert was removed by *Nde*I digestion into the expression vector pET15b, digested by the same enzyme. The good orientation of the insert was verified by digestion profile. The accuracy of the construct was verified by sequencing. The protein expression was achieved as described for pET15+484, except that the plasmid pSBET was not used. Protein samples were loaded on a 12% SDS-PAGE gel.

Western Blot

For Western blotting, proteins were transferred onto PVDF membrane (Roche) using a semi-dry blotting system (Bio-Rad) following standard protocols (Towbin *et al.*, 1992). To check for the presence of the hexahistidine-tag on the fusion proteins, immunodetection assays were performed using rabbit IgG against RGS-His tag (Qiagen, at dilution 1:2000). The Western blotting detection was done according to the manufacturer's instructions (Roche) and visualized by FluorS'max (Bio-Rad).

Gel retardation assays

Labelling of the probes (PCR products), gel retardation assays, and chemiluminescence detection were carried out as described in the DIG Gel Shift kit, 2nd generation, following the manufacturer's instructions (Roche). The primers P467-F/P467-R and P477-F/P477-R were used to amplify the DNA probes P467 (215bp) and P477 (194bp), respectively. Probes (20 fmol) end labelled with 3′ digoxigenin (DIG)-11-ddUTP were incubated at 30°C for 15 min with crude protein extract, pET15 or pET15+T484, in binding buffer containing poly (dI-dC), according to the manufacturer's protocols (Roche). Reaction mixtures were analyzed by native PAGE using 5% acrylamide buffered and run in 0.5X TBE (Tris-borate-EDTA) buffer (Sambrook *et al.*, 1989). DNA was then transferred onto a positively charged nylon membrane (Amersham Hybond-N^+) by electroblotting (400 mA for 30 min in 0.5X TBE buffer) using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad). Light emission was recorded with a Fluor-S MultImager (Bio-Rad)

Table with the primers used in this thesis

Primer function and name	**Nucleotide sequence 5′→3′**
Construction of the mutant strains	
Δ457-F	GACCGTCGGCGTCGGACCGGACGCGTGAAGAGGGCGATGATTCCGGGGATCCGTCGACC
Δ464-R	CCGACGAGCACGTGCCCGCCTCGCTCCTCGACCGCCTCATGTAGGCTGGAGCTGCTTC
Δ488-F	CTGTCGGTGCGGGCCTGCCTCGGAACGGGTCCGGTGTCATGTAGGCTGGAGCTGCTTC
Δ493-R	CCGTCCGCCCACGTGTCCCCTCCGCGGAGGTCCCCCATGATTCCGGGGATCCGTCGACC
Δ478-F	TGCAGACGGGCTGCTCGACGCGCGGGGAGTGACGGCATGTGTAGGCTGGAGCTGCTTC
Δ478-R	GTCCCGGGCTCGGGCCGCGCCTTTGCCGCTGCCGGACTCATTCCGGGGATCCGTCGACC
Δ479-F	GAGCGGCCGGGAGGCGGGCGCGGCGCAGTCGGCGTTCTGTAGGCTGGAGCTGCTTC
Δ479-R	CGCCCCGCAGGACACGGCAACCACCGACGTAGAGGGATGATTCCGGGGATCCGTCGACC
Δ482-F	CGGCACCGGCACAAGGACAGCAAGGAGACCTGACCCATGATTCCGGGGATCCGTCGACC
Δ482-R	CTGTACGGGGGCCCACCGGTCGCGCGGGCCGCTCCGTCATGTAGGCTGGAGCTGCTTC
Δ4701-F	CCGATCGGTTGATCCCCCTGACGAGCGTGATGGGCGATGATTCCGGGGATCCGTCGACC
Δ4701-R	TGTCGGCCATGGCACCCGGCGCGCGTGACACGTTCGTCATGTAGGCTGGAGCTGCTTC
Δ484-F	CGTTCGGAACTGCTAACGGGGAGGCAGACGCCAGTCATGTGTAGGCTGGAGCTGCTTC
Δ484-R	GTGCGCGGCCCCGGCCGGACGCCGTGCTCGGTCGGCCTAATTCCGGGGATCCGTCGACC
Δ468-R	GCTCCGCGGTCCGGCCCAGGGGTGTCCGGGGCGCGGTCATGTAGGCTGGAGCTGCTTC
Δ469-F	CGCGTCCGCCACGCGCTGGTTCCGCTGGGAGTGACGAACATTCCGGGGATCCGTCGACC

Transcriptional analysis	
RT-466-F	CGGCCGTCAGTTCGTGCAGG
RT-466-R	TCGCCGACGACGGTGTCATC
RT-468-F	GTGGGTCAGGTGCGTCTTGAC
RT-468-R	ACTGCAAGTGCTGGAGCACG
RT-469-F	GAGGACGGGAGGGCTGAAGC
RT-469-R	GCGTGGCATCCGACGCGACCC
RT-470-F	GGCATCCTGCGCCGGATGTC
RT-470-R	GCACGGCGGAGATCAGCGAG
RT-471-F	CGCCGACCGTGCGAAGACGA
RT-471-R	CCGGCGTAGAGCTGGAGTGC
RT-472-F	AGGTGCGTGACACCGTCCTC
RT-472-R	GTACCTACGTCCGCGTGTCC
RT-473-F	CGGCGAACTCGGTCCACACT
RT-473-R	CGTCGTGGACGACGGCTTCT
RT-CDS2-F	GGTCGGCGAGTTCGAGCGAG
RT-CDS2-R	TTCAGCTGGTCCGGCGTCAC
RT-475-F	GGTAGCGCCAGGAGTTCACG
RT-475-R	GCGCATGGACTCACATGACG
RT-476-F	TCCGGCGGAGACCACGTCAC
RT-476-R	GAGGCCGGCGCGTCGAACTA
RT-CDS1-F	AGCACGAAGGCGTCGAGGTC
RT-CDS1-R	CTGGCCGAACTCACCGAGGA
RT-478-F	GCTCGGCGATGATCTCGTGG
RT-478-R	GCATGGCCTACCTGCTCGAC
RT-479-F	CATGCGCAGCGTCAGCAAGT
RT-479-R	TCGATGTAGCGCGGCGTGAA
RT-480-F	CGCCGTCGACCAGGTGTTCC
RT-480-R	GTGCGCCAGTCCGAGGTAGG
RT-481-F	CGGCAGCAGCGTGGCGATCA
RT-481-R	GGCATCTCGGTCCGCAGCTA
RT-482-F	CGCCGTGCTCATCTATCCGT
RT-482-R	TTGAGGTGGCGCGGTGTCTG
RT-483-F	TCGGTGTAGTAGGCGATGGA
RT-483-R	TCTACGACTTCGCCATGAGC
RT-485-F	GCGGAGATCAGGCGACTGGA
RT-485-R	AGGTCTCGGCGGCCTTGTAG
RT-486-F	GCTGATCGAGCGGTCCACAT
RT-486-R	GTACGCGGCGACATCTGTGA
RT-487-F	CGGTTGACGTCCGTGATCTC
RT-487-R	ACCTGGCTGTCACCGGACTT
RT-srmGI-F	GCTCCGGCTGAAGGTGCAGA
RT-srmGI-R	TGCCACCGATGCCGAACGAA
RT-orf3*c-F	GCGCGTCGAGCGGTACGAAG
RT-orf3*c-R	GCCGGCTGGACATCGAGGTG
Reporter system	
467redD-F	AAGGATCCAGAAGCCGGCGTGGAAGTCC
467redD-R	TCGAATTCCGAGCTGTTCGCGGCCATCC
477redD-F	CAGGATCCCGGTCACGCAGCTTCTGCTT
477redD-R	CTGAATTCGTCCAGCAGTACCGCCGGAT
4701redD-F	AAGGATCCAACGCGCAGTCCGAGGATGG
4701redD-R	TCGAATTCCGGCGGGGTCAGCCACATGG
Protein expression and EMSA probes	
T484-F	CATATGATCGACGTGGCCGG
P467-F	GAATGCCCCGAAATCACTTG
P467-R	CCCTGGGCCGGACCGCGGAG

P477-F	GTCTCGTCCGTGGCCATCTG
P477-R	AGGCGCGGCCCGAGCCCGGG

Prediction of the secondary structure of SAMR0484, using the program PSIPRED Protein Structure Prediction Server (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html)

Conf: Confidence (0=low, 9=high)
Pred: Predicted secondary structure (H=helix, E=strand, C=coil)
AA: Target sequence (SAMR0484)

The HTH domain in the C-terminus is underlined in yellow. The red arrow indicates where the truncated sequence T484 starts. In blue light the second amino acid of the truncated form, a Ile residue, is also underlined.

```
Conf: 935709999999999999986599769998659988899999999999985897799985434577865
Pred: CCCCCHHHHHHHHHHHHHHHHCCCCEEEEEECCCCCCHHHHHHHHHHHHHHCCCEEEEEECCCCCCCCC
  AA: MLVHRDEQLAHLRWAFHACEKHGHGQVALVTGAVGSGKTQVLETFGEWAATAGGQVLSAAGSRAEQGLHF
              10        20        30        40        50        60        70

Conf: 999999987411665311245544676632541233456776644455683567889999999999984
Pred: HHHHHHHHHHHCCCCCCHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHH
  AA: GVLGQLLHSARLKPEGIREVEELIREVALAVPTAEARSGAAPETLAEALTADGLWAPLLRTAFDTFLTLA
              80        90       100       110       120       130       140

Conf: 899879998660427879999999997178782899998679877653156788852777426447999
Pred: CCCCEEEEEECCCCCCHHHHHHHHHHHHCCCCCCEEEEEEECCCCCCCCHHHHHHHHCCCCCEEEECCCC
  AA: GRGPLVLAVDDLQHTDAASLHCLLYVTRRLRNARIMVLLSEATTLRPAHPLFHAELRSLPRFTRVTLPLL
             150       160       170       180       190       200       210

Conf: 999999999987187212332035666789999999968993799999999866876555543467 08
Pred: CHHHHHHHHHHHCCCCHHHHHCCCCCCCHHHHHHHHHHHCCCHHHHHHHHHHHHHCCCCCCCCCCCCCH
  AA: TADSVSRLLGEEAGTPDVRGDGAGAPDVREEAGRILRVTGGNPLLSQALVDERAHRGTDSAPDTGARAGD
             220       230       240       250       260       270       280

Conf: 999999999871999999999998638878999999886889899999999987788532021579 99
Pred: HHHHHHHHHHHCCCHHHHHHHHHHHHCCCCCCHHHHHHHHCCCHHHHHHHHHHHHHCCCCCCCCCCHHH
  AA: TFERAVLDCLYRHEPGVRLVAQALAALDRPASPELLGQLLDVLPDSTVPAVRVLEGAGLVHAGRLRHPRI
             290       300       310       320       330       340       350

Conf: 999987499999999999999986338986999999986677673789999999999997779899 99
Pred: HHHHHHHCCHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHCCCCCCHHHHHHHHHHHHHHHCCCHHHHH
  AA: VLAVRSDMPAEERRRLHQRAAEVLHENGAEASVVAEHLVASAWTDGAWVVPVLRDAAAHALSTGRPDHAA
             360       370       380       390       400       410       420

Conf: 999999860766378999999888987523737789999999974547123289999999998736667
Pred: HHHHHHHHCCCCHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHCCCHH
  AA: ACLRLGGRAETDKDRRNSLMAMLITARWQVNPLAVTGQVSQLVEAARADESSPSTAVSAVPYLLWQGRAE
             430       440       450       460       470       480       490

Conf: 899999752255666542577899999986272022213343332100346666412799999999 9876
Pred: HHHHHHHHHCCCCCCCCCHHHHHHHHHHHHHHCCHHHHHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHHH
  AA: EAAEAISGCGAGDHGPASNELRLTRLLIALSHPDYLGTVREDPSSWTRAATAPDSASPLLQAVSVLGNAL
             500       510       520       530       540       550       560
```

```
Conf: 0476189999999997325774123799999999998678789999999999999735697389999999
Pred: HCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHCCCCHHHHHHHHH
  AA: MPTGGVDTVAAAEQLLERHHTDSGSLGLLTAPLLALLCAGPADRVAVWGERLLARQGVQHTPAWRGVIRA
            570       580       590       600       610       620       630

Conf: 99999998799999999999999876348705799999999999977999999999988764300 14
Pred: HHHHHHHHCCCHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHCC
  AA: IHAEATLRLGDMDAAERGARLALEDLPAPAWGVAVGGPLGTLITCTTESGRLFEAERWLSQPVPAGVFRT
            640       650       660       670       680       690       700

Conf: 35899999999999689999999999999999865897527888899999999977992589999999
Pred: HHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHCCCHHHHHHHHH
  AA: PVGAHYLIARGRHHLAMGQHQAAAADLHRSGELVRSWGIDVAGLVPWRLELARVQLSLGHRTHAAQLLQE
            710       720       730       740       750       760       770

Conf: 998753257508999999886158402588999999998858998999998899999987799267 99
Pred: HHHHHHCCCCCHHHHHHHHHHHHHCCHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHCCCHHHHH
  AA: QLQVSRGLDDRTRGRALRLLASTATPVLRRKLLSKAVTVLQACSDQQELARALSDTSQTLPQTGDAPRAR
            780       790       800       810       820       830       840

Conf: 99999999998758644569998403221001133455566666677877766799899999999858
Pred: HHHHHHHHHHHHCCCCHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHCC
  AA: LFVRRAGRLAQASGRAPALPQQVQRRSAPSDAAPTGGSSSDASSDGFESRDQDGLLSEAERRVAVLAARG
            850       860       870       880       890       900       910

Conf: 98799999858889999999999987499999999999986168999
Pred: CCHHHHHHHHCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHCCCCCC
  AA: RTNRQISNELYITVSTVEQHLTRVYRKLDVKSRTDLPNRLMALAEPMA
            920       930       940       950
```

REFERENCES

Adamidis T., Riggle P. and Champness W. (1990). "Mutations in a new Streptomyces coelicolor locus which globally block antibiotic biosynthesis but not sporulation." *J Bacteriol* **172**(6): 2962-9.

Ahlert J., Shepard E., Lomovskaya N., Zazopoulos E., Staffa A., Bachmann B. O., Huang K., Fonstein L., Czisny A., Whitwam R. E., Farnet C. M. and Thorson J. S. (2002). "The calicheamicin gene cluster and its iterative type I enediyne PKS." *Science* **297**(5584): 1173-6.

Aigle B., Schneider D., Morilhat C., Vandewiele D., Dary A., Holl A. C., Simonet J. M. and Decaris B. (1996). "An amplifiable and deletable locus of Streptomyces ambofaciens RP181110 contains a very large gene homologous to polyketide synthase genes." *Microbiology* **142** (Pt 10): 2815-24.

Aigle B., Pang X., Decaris B. and Leblond P. (2005). "Involvement of AlpV, a new member of the Streptomyces antibiotic regulatory protein family, in regulation of the duplicated type II polyketide synthase alp gene cluster in Streptomyces ambofaciens." *J Bacteriol* **187**(7): 2491-500.

Alderwick L. J., Molle V., Kremer L., Cozzone A. J., Dafforn T. R., Besra G. S. and Futterer K. (2006). "Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in Mycobacterium tuberculosis." *Proc Natl Acad Sci USA* **103**(8): 2558-63.

Aminov R. I. (2009). "The role of antibiotics and antibiotic resistance in nature" *Environ Microbiol* [Epub ahead of print].

Anton N., Mendes M. V., Martin J. F. and Aparicio J. F. (2004). "Identification of PimR as a positive regulator of pimaricin biosynthesis in *Streptomyces natalensis*" *J Bacteriol* **186**(9): 2567-75.

Aparicio J. F., Caffrey P., Marsden A. F., Staunton J. and Leadlay P. F. (1994). "Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase." *J Biol Chem* **269**(11): 8524-8.

Aparicio J. F., Molnar I., Schwecke T., Konig A., Haydock S. F., Khaw L. E., Staunton J. and Leadlay P. F. (1996). "Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: analysis of the enzymatic domains in the modular polyketide synthase." *Gene* **169**(1): 9-16.

Aparicio J. F., Caffrey P., Gil J. A. and Zotchev S. B. (2003). "Polyene antibiotic biosynthesis gene clusters." *Appl Microbiol Biotechnol* **61**(3): 179-88.

Arca P., Hardisson C. and Suarez J. E. (1990). "Purification of a glutathione S-transferase that mediates fosfomycin resistance in bacteria." *Antimicrob Agents Chemother* **34**(5): 844-8.

Arias P., Fernandez-Moreno M. A. and Malpartida F. (1999). "Characterization of the pathway-specific positive transcriptional regulator for actinorhodin biosynthesis in Streptomyces coelicolor A3(2) as a DNA-binding protein." *J Bacteriol* **181**(22): 6958-68.

Bangera M. G. and Thomashow L. S. (1999). "Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from Pseudomonas fluorescens Q2-87." *J Bacteriol* **181**(10): 3155-63.

Barona-Gomez F., Lautru S., Francou F. X., Leblond P., Pernodet J. L. and Challis G. L. (2006). "Multiple biosynthetic and uptake systems mediate siderophore-dependent iron acquisition in Streptomyces coelicolor A3(2) and Streptomyces ambofaciens ATCC 23877." *Microbiology* **152**(Pt 11): 3355-66.

Belanger A. E., Besra G. S., Ford M. E., Mikusova K., Belisle J. T., Brennan P. J. and Inamine J. M. (1996). "The embAB genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol." *Proc Natl Acad Sci U S A* **93**(21): 11919-24.

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., Bateman A., Brown S., Chandra G., Chen C. W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C. H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S., Rabbinowitsch E., Rajandream M. A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B. G., Parkhill J. and Hopwood D. A. (2002). "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)." *Nature* **417**(6885): 141-7.

Berdy J. (2005). "Bioactive microbial metabolites." *J Antibiot (Tokyo)* **58**(1): 1-26.

Bergmann S., Schumann J., Scherlach K., Lange C., Brakhage A. A. and Hertweck C. (2007). "Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans." *Nat Chem Biol* **3**(4): 213-7.

Bibb M. J. (2005). "Regulation of secondary metabolism in streptomycetes." *Curr Opin Microbiol* **8**(2): 208-15.

Bibb M. and Hesketh A. (2009). "Chapter 4. Analyzing the regulation of antibiotic production in streptomycetes." *Methods Enzymol* **458**: 93-116.

Bisang C., Long P. F., Cortes J., Westcott J., Crosby J., Matharu A. L., Cox R. J., Simpson T. J., Staunton J. and Leadlay P. F. (1999). "A chain initiation factor common to both modular and aromatic polyketide synthases." *Nature* **401**(6752): 502-5.

Bister B., Bischoff D., Strobele M., Riedlinger J., Reicke A., Wolter F., Bull A. T., Zahner H., Fiedler H. P. and Sussmuth R. D. (2004). "Abyssomicin C-A polycyclic antibiotic from a marine Verrucosispora strain as an inhibitor of the p-aminobenzoic acid/tetrahydrofolate biosynthesis pathway." *Angew Chem Int Ed Engl* **43**(19): 2574-6.

Bode H. B., Bethe B., Hofs R. and Zeeck A. (2002). "Big effects from small changes: possible ways to explore nature's chemical diversity." *Chembiochem* **3**(7): 619-27.

Boos W. and Shuman H. (1998). "Maltose/Maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation." *Microbiol and Mol Biol Reviews* **62**(1): 204-229.

Borisova S. A., Zhao L., Melancon I. C., Kao C. L. and Liu H. W. (2004). "Characterization of the glycosyltransferase activity of desVII: analysis of and implications for the biosynthesis of macrolide antibiotics." *J Am Chem Soc* **126**(21): 6534-5.

Brotz-Oesterhelt H., Beyer D., Kroll H. P., Endermann R., Ladel C., Schroeder W., Hinzen B., Raddatz S., Paulsen H., Henninger K., Bandow J. E., Sahl H. G. and Labischinski H. (2005). "Dysregulation of bacterial proteolytic machinery by a new class of antibiotics." *Nat Med* **11**(10): 1082-7.

Bruton C. J., Guthrie E. P. and Chater K. F. (1991). "Phage vectors that allow monitoring of transcription of secondary metabolism genes in Streptomyces." *Biotechnology (N Y)* **9**(7): 652-6.

Bu'Lock J. D. (1961). "Intermediary metabolism and antibiotic synthesis." *Adv Appl Microbiol* **3**: 293-342.

Bunet R., Mendes M. V., Rouhier N., Pang X., Hotel L., Leblond P. and Aigle B. (2008). "Regulation of the synthesis of the angucyclinone antibiotic alpomycin in Streptomyces ambofaciens by the autoregulator receptor AlpZ and its specific ligand." *J Bacteriol* **190**(9): 3293-305.

Butler M. J., Bruheim P., Jovetic S., Marinelli F., Postma P. W. and Bibb M. J. (2002). "Engineering of primary carbon metabolism for improved antibiotic production in Streptomyces lividans." *Appl Environ Microbiol* **68**(10): 4731-9.

Caffrey P., Bevitt D. J., Staunton J. and Leadlay P. F. (1992). "Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from Saccharopolyspora erythraea." *FEBS Lett* **304**(2-3): 225-8.

Caffrey P. (2003). "Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases." *Chembiochem* **4**(7): 654-7.

Camilli A. and Bassler B. L. (2006). "Bacterial small-molecule signaling pathways." *Science* **311**(5764): 1113-6.

Campbell J. A., Davies G. J., Bulone V. and Henrissat B. (1997). "A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities." *Biochem J* **326** (Pt 3): 929-39.

Cane D. E. and Walsh C. T. (1999). "The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases." *Chem Biol* **6**(12): R319-25.

Carmody M., Byrne B., Murphy B., Breen C., Lynch S., Flood E., Finnan S. and Caffrey P. (2004). "Analysis and manipulation of amphotericin biosynthetic genes by means of modified phage KC515 transduction techniques." *Gene* **343**(1): 107-15.

Cassell G. H. and Mekalanos J. (2001). "Development of antimicrobial agents in the era of new and reemerging infectious diseases and increasing antibiotic resistance." *Jama* **285**(5): 601-5.

Chakraburtty R., White J., Takano E. and Bibb M. (1996). "Cloning, characterization and disruption of a (p)ppGpp synthetase gene (relA) of Streptomyces coelicolor A3(2)." *Mol Microbiol* **19**(2): 357-368.

Chakraburtty R. and Bibb M. (1997). "The ppGpp synthetase gene (relA) of Streptomyces coelicolor A3(2) plays a conditional role in antibiotic production and morphological differentiation." *J Bacteriol* **179**(18): 5854-61.

Chan Y. A., Podevels A. M., Kevany B. M. and Thomas M. G. (2009). "Biosynthesis of polyketide synthase extender units." *Nat Prod Rep* **26**(1): 90-114.

Chandra G. and Chater K. F. (2008). "Evolutionary flux of potentially bldA-dependent Streptomyces genes containing the rare leucine codon TTA." *Antonie Van Leeuwenhoek* **94**(1): 111-26.

Charan R. D., Schlingmann G., Janso J., Bernan V., Feng X. and Carter G. T. (2004). "Diazepinomicin, a new antimicrobial alkaloid from a marine Micromonospora sp." *J Nat Prod* **67**(8): 1431-3.

Chater K. F. and Chandra G. (2008). "The use of the rare UUA codon to define "expression space" for genes involved in secondary metabolism, development and environmental adaptation in streptomyces." *J Microbiol* **46**(1): 1-11.

Chen S., Huang X., Zhou X., Bai L., He J., Jeong K. J., Lee S. Y. and Deng Z. (2003). "Organizational and mutational analysis of a complete FR-008/candicidin gene cluster encoding a structurally related polyene complex." *Chem Biol* **10**(11): 1065-76.

Cheng Y. Q., Tang G. L. and Shen B. (2003). "Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis." *Proc Natl Acad Sci U S A* **100**(6): 3149-54.

Cheng L., Naumann T. A., Horswill A. R., Hong S. J., Venters B. J., Tomsho J. W., Benkovic S. J. and Keiler K. C. (2007). "Discovery of antibacterial cyclic peptides that inhibit the ClpXP protease." *Protein Sci* **16**(8): 1535-42.

Choi S. U., Lee C. K., Hwang Y. I., Kinosita H. and Nihira T. (2003). "Gamma-butyrolactone autoregulators and receptor proteins in non- Streptomyces actinomycetes producing commercially important secondary metabolites." *Arch Microbiol* **180**(4): 303-7.

Chouayekh H. and Virolle M. J. (2002). "The polyphosphate kinase plays a negative role in the control of antibiotic production in Streptomyces lividans." *Mol Microbiol* **43**(4): 919-30.

Choulet F., Aigle B., Gallois A., Mangenot S., Gerbaud C., Truong C., Francou F. X., Fourrier C., Guerineau M., Decaris B., Barbe V., Pernodet J. L. and Leblond P. (2006). "Evolution of the terminal regions of the Streptomyces linear chromosome." *Mol Biol Evol* **23**(12): 2361-9.

Choulet F. (2006). Evolution du génome des *Streptomyces*: transfert horizontal et variabilité des extrémités chromosomiques. PhD Thesis.

Corre C., Song L., O'Rourke S., Chater K. F. and Challis G. L. (2008). "2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by Streptomyces coelicolor genome mining." *Proc Natl Acad Sci U S A* **105**(45): 17510-5.

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." *Nature* **348**(6297): 176-8.

Coutinho P. M., Deleury E., Davies G. J. and Henrissat B. (2003). "An evolving hierarchical family classification for glycosyltransferases." *J Mol Biol* **328**(2): 307-17.

Cronan J. E. and Thomas J. (2009). "Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways." *Methods Enzymol* **459**: 395-433.

Dangel V., Harle J., Goerke C., Wolz C., Gust B., Pernodet J. L. and Heide L. (2009). "Transcriptional regulation of the novobiocin biosynthetic gene cluster." *Microbiology* [Epub ahead of print].

Dary A., Bourget N., Girard N., Simonet J. M. and Decaris B. (1992). "Amplification of a particular DNA sequence in Streptomyces ambofaciens RP181110 reversibly prevents spiramycin production." *Res Microbiol* **143**(1): 99-112.

Davies J. (1995). "Vicious circles: looking back on resistance plasmids." *Genetics* **139**(4): 1465-8.

Davies J. (2006). "Where have All the Antibiotics Gone?" *Can J Infect Dis Med Microbiol* **17**(5): 287-90.

Davies J., Spiegelman G. B. and Yim G. (2006). "The world of subinhibitory antibiotic concentrations." *Curr Opin Microbiol* **9**(5): 445-53.

Davies J. (2007). "Microbes have the last word. A drastic re-evaluation of antimicrobial treatment is needed to overcome the threat of antibiotic-resistant bacteria." *EMBO Rep* **8**(7): 616-21.

Demain (2007). "Contributions of microorganisms to industrial biology." *Molecular biotechnology* **38**: 41-55.

Diacovich L., Mitchell D. L., Pham H., Gago G., Melgar M. M., Khosla C., Gramajo H. and Tsai S. C. (2004). "Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity." *Biochemistry* **43**(44): 14027-36.

Donadio S. and Katz L. (1992). "Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in Saccharopolyspora erythraea." *Gene* **111**(1): 51-60.

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

Douthwaite S. and Champney W. S. (2001). "Structures of ketolides and macrolides determine their mode of interaction with the ribosomal target site." *J Antimicrob Chemother* **48 (**Suppl T1**)**: 1-8.

Eustaquio A. S., McGlinchey R. P., Liu Y., Hazzard C., Beer L. L., Florova G., Alhamadsheh M. M., Lechner A., Kale A. J., Kobayashi Y., Reynolds K. A. and Moore B. S. (2009). "Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine." *Proc Natl Acad Sci U S A* **106**(30): 12295-300.

Fazio G. C., Xu R. and Matsuda S. P. (2004). "Genome mining to identify new plant triterpenoids." *J Am Chem Soc* **126**(18): 5678-9.

Fedoryshyn M., Petzke L., Welle E., Bechthold A. and Luzhetskyy A. (2008). "Marker removal from actinomycetes genome using Flp recombinase." *Gene* **419**(1-2): 43-7.

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." *Angew Chem Int Ed Engl* **42**(3): 355-7.

Fernandez-Moreno M. A., Caballero J. L., Hopwood D. A. and Malpartida F. (1991). "The act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the bldA tRNA gene of Streptomyces." *Cell* **66**(4): 769-80.

Fischbach M. A. and Walsh C. T. (2006). "Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms." *Chem Rev* **106**(8): 3468-96.

Floriano B. and Bibb M. (1996). "afsR is a pleiotropic but conditionally required regulatory gene for antibiotic production in Streptomyces coelicolor A3(2)." *Mol Microbiol* **21**(2): 385-96.

Folcher M., Gaillard H., Nguyen L. T., Nguyen K. T., Lacroix P., Bamas-Jacques N., Rinkel M. and Thompson C. J. (2001). "Pleiotropic functions of a Streptomyces pristinaespiralis autoregulator receptor in development, antibiotic biosynthesis, and expression of a superoxide dismutase." *J Biol Chem* **276**(47): 44297-306.

Floss H. G. (2006). "Combinatorial biosynthesis. Potentail and problems." *J Biotechnol* **124**(1): 242-57.

Funa N., Ohnishi Y., Fujii I., Shibuya M., Ebizuka Y. and Horinouchi S. (1999). "A new pathway for polyketide synthesis in microorganisms." *Nature* **400**(6747): 897-9.

Gaisser S., Trefzer A., Stockert S., Kirschning A. and Bechthold A. (1997). "Cloning of an avilamycin biosynthetic gene cluster from Streptomyces viridochromogenes Tu57." *J Bacteriol* **179**(20): 6271-8.

Garcia R. O., Krug D. and Muller R. (2009). "Chapter 3. Discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods." *Methods Enzymol* **458**: 59-91.

Ghorbel S., Smirnov A., Chouayekh H., Sperandio B., Esnault C., Kormanec J. and Virolle M. J. (2006). "Regulation of ppk expression and in vivo function of Ppk in Streptomyces lividans TK24." *J Bacteriol* **188**(17): 6269-76.

Goh E., Yim G., Tsui W., McClure J., Surette M. G. and Davies J. (2002) "Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. " *Proc Natl Acad Sci U S A* **99**(26): 17025-30.

Goodfellow M., Mordarski M. and Williams S. T. (1984). "The biology of actinomycetes." Academic Press Inc. (London)

Gourmelen A., Blondelet-Rouault M. H. and Pernodet J. L. (1998). "Characterization of a glycosyl transferase inactivating macrolides, encoded by gimA from Streptomyces ambofaciens." *Antimicrob Agents Chemother* **42**(10): 2612-9.

Gristwood T., Fineran P. C., Everson L., Williamson N. R. and Salmond G. P. (2009). "The PhoBR two-component system regulates antibiotic biosynthesis in Serratia in response to phosphate." *BMC Microbiol* **9**: 112.

Gross F., Luniak N., Perlova O., Gaitatzis N., Jenke-Kodama H., Gerth K., Gottschalk D., Dittmann E. and Muller R. (2006). "Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads." *Arch Microbiol* **185**(1): 28-38.

Gross H. (2007). "Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects." *Appl Microbiol Biotechnol* **75**(2): 267-77.

Gross H., Stockwell V. O., Henkels M. D., Nowak-Thompson B., Loper J. E. and Gerwick W. H. (2007). "The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters." *Chem Biol* **14**(1): 53-63.

Gust B., Challis G. L., Fowler K., Kieser T. and Chater K. F. (2003). "PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin." *Proc Natl Acad Sci U S A* **100**(4): 1541-6.

Guthrie E. P. and Chater K. F. (1990). "The level of a transcript required for production of a Streptomyces coelicolor antibiotic is conditionally dependent on a tRNA gene." *J Bacteriol* **172**(11): 6189-93.

Guzman S., Ramos I., Moreno E., Ruiz B., Rodriguez-Sanoja R., Escalante L., Langley E. and Sanchez S. (2005). "Sugar uptake and sensitivity to carbon catabolite regulation in Streptomyces peucetius var. caesius." *Appl Microbiol Biotechnol* **69**(2): 200-6.

Hamdane D., Zhang H. and Hollenberg P. (2008). "Oxygen activation by cytochrome P450 monooxygenase." *Photosynth Res* **98**(1-3): 657-66.

Han L., Yang K., Kulowski K., Wendt-Pienkowski E., Hutchinson C. R. and Vining L. C. (2000). "An acyl-coenzyme A carboxylase encoding gene associated with jadomycin biosynthesis in Streptomyces venezuelae ISP5230." *Microbiology* **146** (Pt 4): 903-10.

Harada K., Suzuki M., Kameda M. and Mitsuhashi S. (1960). "On the drug-resistance of enteric bacteria. 2) Transmission of the drug-resistance among Enterobacteriaceae." *Jpn J Exp Med* **30**: 289-99.

Harle J. and Bechthold A. (2009). "Chapter 12. The power of glycosyltransferases to generate bioactive natural compounds." *Methods Enzymol* **458**: 309-33.

He W., Lei J., Liu Y. and Wang Y. (2008). "The LuxR family members GdmRI and GdmRII are positive regulators of geldanamycin biosynthesis in Streptomyces hygroscopicus 17997." *Arch Microbiol* **189**(5): 501-10.

Heide L. (2009). "Aminocoumarins mutasynthesis, chemoenzymatic synthesis, and metabolic engineering." *Methods Enzymol* **459**: 437-55.

Helmke (1984). "Rhodococcus marinonascens sp. nov., an actinomycete from the sea." *Int J Syst Bacteriol* **34**: 127-138.

Hernandez C., Olano C., Mendez C. and Salas J. A. (1993). "Characterization of a Streptomyces antibioticus gene cluster encoding a glycosyltransferase involved in oleandomycin inactivation." *Gene* **134**(1): 139-40.

Hertweck C., Luzhetskyy A., Rebets Y. and Bechthold A. (2007). "Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork." *Nat Prod Rep* **24**(1): 162-90.

Hesketh A., Sun J. and Bibb M. (2001). "Induction of ppGpp synthesis in Streptomyces coelicolor A3(2) grown under conditions of nutritional sufficiency elicits actII-ORF4 transcription and actinorhodin biosynthesis." *Mol Microbiol* **39**(1): 136-44.

Hiard S., Maree R., Colson S., Hoskisson P. A., Titgemeyer F., van Wezel G. P., Joris B., Wehenkel L. and Rigali S. (2007). "PREDetector: a new tool to identify regulatory elements in bacterial genomes." *Biochem Biophys Res Commun* **357**(4): 861-4.

Hiramatsu K., Cui L., Kuroda M. and Ito T. (2001). "The emergence and evolution of methicillin-resistant Staphylococcus aureus." *Trends Microbiol* **9**(10): 486-93.

Hong J. S., Park S. J., Parajuli N., Park S. R., Koh H. S., Jung W. S., Choi C. Y. and Yoon Y. J. (2007). "Functional analysis of desVIII homologues involved in glycosylation of macrolide antibiotics by interspecies complementation." *Gene* **386**(1-2): 123-30.

Hopwood D. A. (2007). "How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them?" *Mol Microbiol* **63**(4): 937-40.

Horinouchi S. and Beppu T. (1992). "Regulation of secondary metabolism and cell differentiation in Streptomyces: A-factor as a microbial hormone and the AfsR protein as a component of a two-component regulatory system." *Gene* **115**(1-2): 167-72.

Huang J., Shi J., Molle V., Sohlberg B., Weaver D., Bibb M. J., Karoonuthaisiri N., Lih C. J., Kao C. M., Buttner M. J. and Cohen S. N. (2005). "Cross-regulation among disparate antibiotic biosynthetic pathways of Streptomyces coelicolor." *Mol Microbiol* **58**(5): 1276-87.

Hulett F. M. (1996). "The signal-transduction network for Pho regulation in Bacillus subtilis." *Mol Microbiol* **19**(5): 933-9.

Hur Y. A., Choi S. S., Sherman D. H. and Kim E. S. (2008). "Identification of TmcN as a pathway-specific positive regulator of tautomycetin biosynthesis in Streptomyces sp. CK4412." *Microbiology* **154**(Pt 10): 2912-9.

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 Biotechnol* **21**(5): 526-31.

Ishizuka H., Horinouchi S., Kieser H. M., Hopwood D. A. and Beppu T. (1992). "A putative two-component regulatory system involved in secondary metabolism in Streptomyces spp." *J Bacteriol* **174**(23): 7585-94.

Jnawali H. N., Oh T. J., Liou K., Park B. C. and Sohng J. K. (2008). "A two-component regulatory system involved in clavulanic acid production." *J Antibiot (Tokyo)* **61**(11): 651-9.

Juguet M., Lautru S., Francou F. X., Nezbedova S., Leblond P., Gondry M. and Pernodet J. L. (2009). "An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in Streptomyces ambofaciens." *Chem Biol* **16**(4): 421-31.

Kamra P., Gokhale R. S. and Mohanty D. (2005). "SEARCHGTr: a program for analysis of glycosyltransferases involved in glycosylation of secondary metabolites." *Nucleic Acids Res* **33**(Web Server issue): W220-5.

Karray F., Darbon E., Oestreicher N., Dominguez H., Tuphile K., Gagnat J., Blondelet-Rouault M. H., Gerbaud C. and Pernodet J. L. (2007). "Organization of the biosynthetic gene cluster for the macrolide antibiotic spiramycin in Streptomyces ambofaciens." *Microbiology* **153**(Pt 12): 4111-22.

Kato J. Y., Funa N., Watanabe H., Ohnishi Y. and Horinouchi S. (2007). "Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in Streptomyces." *Proc Natl Acad Sci U S A* **104**(7): 2378-83.

Kato N., Suzuki H., Takagi H., Asami Y., Kakeya H., Uramoto M., Usui T., Takahashi S., Sugimoto Y. and Osada H. (2009). "Identification of cytochrome P450s required for fumitremorgin biosynthesis in Aspergillus fumigatus." *Chembiochem* **10**(5): 920-8.

Keatinge-Clay A. T. (2007). "A tylosin ketoreductase reveals how chirality is determined in polyketides." *Chem Biol* **14**(8): 898-908.

Khokhlov A. S., Tovarova, II, Borisova L. N., Pliner S. A., Shevchenko L. N., Kornitskaia E., Ivkina N. S. and Rapoport I. A. (1967). "[The A-factor, responsible for streptomycin biosynthesis by mutant strains of Actinomyces streptomycini]." *Dokl Akad Nauk SSSR* **177**(1): 232-5.

Kieser T., Bibb M. J., Buttner M. J., Chater K. F. and Hopwood D. A. (2000). "Practical streptomyces genetics." The John Innes Foundation (Norwich).

Kim B. S., Cropp T. A., Beck B. J., Sherman D. H. and Reynolds K. A. (2002). "Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin." *J Biol Chem* **277**(50): 48028-34.

Kitamoto O., Kasai N., Fukaya K. and Kawashima A. (1956). "Drug sensitivity of the *Shigella* strains isolated in 1955." *J. Jpn. Assoc. Infect. Dis.* **30**: 403-404.

Kitani S., Ikeda H., Sakamoto T., Noguchi S. and Nihira T. (2009). "Characterization of a regulatory gene, aveR, for the biosynthesis of avermectin in Streptomyces avermitilis." *Appl Microbiol Biotechnol* **82**(6): 1089-96.

Knirschova R., Novakova R., Feckova L., Timko J., Turna J., Bistakova J. and Kormanec J. (2007). "Multiple regulatory genes in the salinomycin biosynthetic gene cluster of Streptomyces albus CCM 4719." *Folia Microbiol (Praha)* **52**(4): 359-65.

Kobayashi M., Tanaka J., Katori T. and Kitagawa I. (1990). "Marine natural products. XXIII. Three new cytotoxic dimeric macrolides, swinholides B and C and isoswinholide A, congeners of swinholide A, from the Okinawan marine sponge Theonella swinhoei." *Chem Pharm Bull (Tokyo)* **38**(11): 2960-6.

Kopp F. and Marahiel M. A. (2007). "Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis." *Nat Prod Rep* **24**(4): 735-49.

Kotra L. P., Haddad J. and Mobashery S. (2000). "Aminoglycosides: perspectives on mechanisms of action and resistance and strategies to counter resistance." *Antimicrob Agents Chemother* **44**(12): 3249-56.

Kuo M. S., Yurek D. A., Laborde A. L., Truesdell S. E., Nielsen J. W. Argoudelis A. D. and Baczynskyj L. (1990). "Monazomycin B, a new macrolide antibiotic of the monazomycin family." *J Antibiot (Tokyo)* **43**(4): 438-40.

Kuscer E., Coates N., Challis I., Gregory M., Wilkinson B., Sheridan R. and Petkovic H. (2007). "Roles of rapH and rapG in positive regulation of rapamycin biosynthesis in Streptomyces hygroscopicus." *J Bacteriol* **189**(13): 4756-63.

Kwan D. H., Sun Y., Schulz F., Hong H., Popovic B., Sim-Stark J. C., Haydock S. F. and Leadlay P. F. (2008). "Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases." *Chem Biol* **15**(11): 1231-40.

Lam K. S. (2006). "Discovery of novel metabolites from marine actinomycetes." *Curr Opin Microbiol* **9**(3): 245-51.

Lambalot R. H., Gehring A. M., Flugel R. S., Zuber P., LaCelle M., Marahiel M. A., Reid R., Khosla C. and Walsh C. T. (1996). "A new enzyme superfamily - the phosphopantetheinyl transferases." *Chem Biol* **3**(11): 923-36.

Laub M. T. and Goulian M. (2007). "Specificity in two-component signal transduction pathways." *Annu Rev Genet* **41**: 121-45.

Lautru S., Deeth R. J., Bailey L. M. and Challis G. L. (2005). "Discovery of a new peptide natural product by Streptomyces coelicolor genome mining." *Nat Chem Biol* **1**(5): 265-9.

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

Lazzarini A., Cavaletti L., Toppo G. and Marinelli F. (2001). "Rare genera of actinomycetes as potential producers of new antibiotics." *Antonie Van Leeuwenhoek* **79**(3-4): 399-405.

Leblond P., Demuyter P., Moutier L., Laakel M., Decaris B. and Simonet J. M. (1989). "Hypervariability, a new phenomenon of genetic instability, related to DNA amplification in Streptomyces ambofaciens." *J Bacteriol* **171**(1): 419-23.

Leblond P., Francou F. X., Simonet J. M. and Decaris B. (1990). "Pulsed-field gel electrophoresis analysis of the genome of Streptomyces ambofaciens strains." *FEMS Microbiol Lett* **60**(1-2): 79-88.

Leblond P., Demuyter P., Simonet J. M. and Decaris B. (1991). "Genetic instability and associated genome plasticity in Streptomyces ambofaciens: pulsed-field gel electrophoresis evidence for large DNA alterations in a limited genomic region." *J Bacteriol* **173**(13): 4229-33.

Leblond P., Fischer G., Francou F. X., Berger F., Guerineau M. and Decaris B. (1996). "The unstable region of Streptomyces ambofaciens includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA." *Mol Microbiol* **19**(2): 261-71.

Lee P. C., Umeyama T. and Horinouchi S. (2002). "afsS is a target of AfsR, a transcriptional factor with ATPase activity that globally controls secondary metabolism in Streptomyces coelicolor A3(2)." *Mol Microbiol* **43**(6): 1413-30.

Linares J. F., Gustafsson I., Baquero F. and Martinez J. L. (2006). "Antibiotics as intermicrobial signaling agents instead of weapons." *Proc Natl Acad Sci U S A* **103**(51): 19484-9.

Long P. F., Wilkinson C. J., Bisang C. P., Cortes J., Dunster N., Oliynyk M., McCormick E., McArthur H., Mendez C., Salas J. A., Staunton J. and Leadlay P. F. (2002). "Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase." *Mol Microbiol* **43**(5): 1215-25.

Looman A. C., Bodlaender J., Comstock L. J., Eaton D., Jhurani P., de Boer H. A. and van Knippenberg P. H. (1987). "Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in Escherichia coli." *Embo J* **6**(8): 2489-92.

Lu X. L., Xu Q. Z., Liu X. Y., Cao X., Ni K. Y. and Jiao B. H. (2008). "Marine drugs-macrolactins." *Chem Biodivers* **5**(9): 1669-74.

Machida K., Arisawa A., Takeda S., Tsuchida T., Aritoku Y., Yoshida M. and Ikeda H. (2008). "Organization of the biosynthetic gene cluster for the polyketide antitumor macrolide, pladienolide, in Streptomyces platensis Mer-11107." *Biosci Biotechnol Biochem* **72**(11): 2946-52.

Marck C., Lefebvre O., Carles C., Riva M., Chaussivert N., Ruet A. and Sentenac A. (1993). "The TFIIIB-assembling subunit of yeast transcription factor TFIIIC has both tetratricopeptide repeats and basic helix-loop-helix motifs." *Proc Natl Acad Sci U S A* **90**(9): 4027-31.

Marinelli F. (2009). "Chapter 2. From microbial products to novel drugs that target a multitude of disease indications." *Methods Enzymol* **458**: 29-58.

Marsden A. F., Wilkinson B., Cortes J., Dunster N. J., Staunton J. and Leadlay P. F. (1998). "Engineering broader specificity into an antibiotic-producing polyketide synthase." *Science* **279**(5348): 199-202.

Martin J. F. (2004). "Phosphate control of the biosynthesis of antibiotics and other secondary metabolites is mediated by the PhoR-PhoP system: an unfinished story." *J Bacteriol* **186**(16): 5197-201.

Martin J. F. and Aparicio J. F. (2009). "Enzymology of the polyenes pimaricin and candicidin biosynthesis." *Methods Enzymol* **459**: 215-42.

Martinez A., Kolvek S. J., Hopke J., Yip C. L. and Osburne M. S. (2005). "Environmental DNA fragment conferring early and increased sporulation and antibiotic production in Streptomyces species." *Appl Environ Microbiol* **71**(3): 1638-41.

Mascaretti O. A. (2003). "Bacteria versus antibacterial agents. An integrated approach."ASM press (Washington, D.C.).

May J. J., Wendrich T. M. and Marahiel M. A. (2001). "The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin." *J Biol Chem* **276**(10): 7209-17.

McAlpine J. B., Bachmann B. O., Piraee M., Tremblay S., Alarco A. M., Zazopoulos E. and Farnet C. M. (2005). "Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example." *J Nat Prod* **68**(4): 493-6.

McArthur M. and Bibb M. J. (2008). "Manipulating and understanding antibiotic production in Streptomyces coelicolor A3(2) with decoy oligonucleotides." *Proc Natl Acad Sci U S A* **105**(3): 1020-5.

McKenzie N. L. and Nodwell J. R. (2007). "Phosphorylated AbsA2 negatively regulates antibiotic production in Streptomyces coelicolor through interactions with pathway-specific regulatory gene promoters." *J Bacteriol* **189**(14): 5284-92.

McLeod M. P., Warren R. L., Hsiao W. W., Araki N., Myhre M., Fernandes C., Miyazawa D., Wong W., Lillquist A. L., Wang D., Dosanjh M., Hara H., Petrescu A., Morin R. D., Yang G., Stott J. M., Schein J. E., Shin H., Smailus D., Siddiqui A. S., Marra M. A., Jones S. J., Holt R., Brinkman F. S., Miyauchi K., Fukuda M., Davies J. E., Mohn W. W. and Eltis L. D. (2006). "The complete genome of Rhodococcus sp. RHA1 provides insights into a catabolic powerhouse." *Proc Natl Acad Sci U S A* **103**(42): 15582-7.

McMurry L., Petrucci R. E., Jr. and Levy S. B. (1980). "Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in Escherichia coli." *Proc Natl Acad Sci U S A* **77**(7): 3974-7.

Melancon C. E., 3rd, Takahashi H. and Liu H. W. (2004). "Characterization of tylM3/tylM2 and mydC/mycB pairs required for efficient glycosyltransfer in macrolide antibiotic biosynthesis." *J Am Chem Soc* **126**(51): 16726-7.

Melancon C. E., 3rd, Hong L., White J. A., Liu Y. N. and Liu H. W. (2007). "Characterization of TDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase from the D-mycaminose biosynthetic pathway of Streptomyces fradiae: in vitro activity and substrate specificity studies." *Biochemistry* **46**(2): 577-90.

Mendez C. and Salas J. A. (2001). "The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanisms." *Res Microbiol* **152**(3-4): 341-50.

Menges R., Muth G., Wohlleben W. and Stegmann E. (2007). "The ABC transporter Tba of Amycolatopsis balhimycina is required for efficient export of the glycopeptide antibiotic balhimycin." *Appl Microbiol Biotechnol* **77**(1): 125-34.

Menzella H. G., Reid R., Carney J. R., Chandran S. S., Reisinger S. J., Patel K. G., Hopwood D. A. and Santi D. V. (2005). "Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes." *Nat Biotechnol* **23**(9): 1171-6.

Merson-Davies L. A. and Cundliffe E. (1994). "Analysis of five tylosin biosynthetic genes from the tyllBA region of the Streptomyces fradiae genome." *Mol Microbiol* **13**(2): 349-55.

Mesak L. R. and Davies J. (2009). "Phenotypic changes in ciprofloxacin-resistant Staphylococcus aureus." *Res Microbiol* [Epub ahead of print].

Miao V., Coeffet-Legal M. F., Brian P., Brost R., Penn J., Whiting A., Martin S., Ford R., Parr I., Bouchard M., Silva C. J., Wrigley S. K. and Baltz R. H. (2005). "Daptomycin biosynthesis in Streptomyces roseosporus: cloning and analysis of the gene cluster and revision of peptide stereochemistry." *Microbiology* **151**(Pt 5): 1507-23.

Mincer T. J., Fenical W. and Jensen P. R. (2005). "Culture-dependent and culture-independent diversity within the obligate marine actinomycete genus Salinispora." *Appl Environ Microbiol* **71**(11): 7019-28.

Mizuno T. and Tanaka I. (1997). "Structure of the DNA-binding domain of the OmpR family of response regulators." *Mol Microbiol* **24**(3): 665-7.

Moore B. S. and Hopke J. N. (2001). "Discovery of a new bacterial polyketide biosynthetic pathway." *Chembiochem* **2**(1): 35-8.

Mutka S. C., Bondi S. M., Carney J. R., Da Silva N. A. and Kealey J. T. (2006). "Metabolic pathway engineering for complex polyketide biosynthesis in Saccharomyces cerevisiae." *FEMS Yeast Res* **6**(1): 40-7.

Nakano H., Takehara E., Nihira T. and Yamada Y. (1998). "Gene replacement analysis of the Streptomyces virginiae barA gene encoding the butyrolactone autoregulator receptor reveals that BarA acts as a repressor in virginiamycin biosynthesis." *J Bacteriol* **180**(13): 3317-22.

Nakaya R., Nakamura A. and Murata Y. (1960). "Resistance transfer agents in Shigella." *Biochem Biophys Res Commun* **3**: 654-9.

Narva K. E. and Feitelson J. S. (1990). "Nucleotide sequence and transcriptional analysis of the redD locus of Streptomyces coelicolor A3(2)." *J Bacteriol* **172**(1): 326-33.

Newman D. J. and Cragg G. M. (2007). "Natural products as sources of new drugs over the last 25 years." *J Nat Prod* **70**(3): 461-77.

Nguyen K. T., Ritz D., Gu J. Q., Alexander D., Chu M., Miao V., Brian P. and Baltz R. H. (2006). "Combinatorial biosynthesis of novel antibiotics related to daptomycin." *Proc Natl Acad Sci U S A* **103**(46): 17462-7.

Nierman W. C., Pain A., Anderson M. J., Wortman J. R., Kim H. S., Arroyo J., Berriman M., Abe K., Archer D. B., Bermejo C., Bennett J., Bowyer P., Chen D., Collins M., Coulsen R., Davies R., Dyer P. S., Farman M., Fedorova N., Feldblyum T. V., Fischer R., Fosker N., Fraser A., Garcia J. L., Garcia M. J., Goble A., Goldman G. H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B., Haas H., Harris D., Horiuchi H., Huang J., Humphray S., Jimenez J., Keller N., Khouri H., Kitamoto K., Kobayashi T., Konzack S., Kulkarni R., Kumagai T., Lafon A., Latge J. P., Li W., Lord A., Lu C., Majoros W. H., May G. S., Miller B. L., Mohamoud Y., Molina M., Monod M., Mouyna I., Mulligan S., Murphy L., O'Neil S., Paulsen I., Penalva M. A., Pertea M., Price C., Pritchard B. L., Quail M. A., Rabbinowitsch E., Rawlins N., Rajandream M. A., Reichard U., Renauld H., Robson G. D., Rodriguez de Cordoba S., Rodriguez-Pena J. M., Ronning C. M., Rutter S., Salzberg S. L., Sanchez M., Sanchez-Ferrero J. C., Saunders D., Seeger K., Squares R., Squares S., Takeuchi M., Tekaia F., Turner G., Vazquez de Aldana C. R., Weidman J., White O., Woodward J., Yu J. H., Fraser C., Galagan J. E., Asai K., Machida M., Hall N., Barrell B. and Denning D. W. (2005). "Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus." *Nature* **438**(7071): 1151-6.

O'Rourke S., Wietzorrek A., Fowler K., Corre C., Challis G. L. and Chater K. F. (2009). "Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in Streptomyces coelicolor." *Mol Microbiol* **71**(3): 763-78.

Oh S. H. and Chater K. F. (1997). "Denaturation of circular or linear DNA facilitates targeted integrative transformation of Streptomyces coelicolor A3(2): possible relevance to other organisms." *J Bacteriol* **179**(1): 122-7.

Ohnishi Y., Ishikawa J., Hara H., Suzuki H., Ikenoya M., Ikeda H., Yamashita A., Hattori M. and Horinouchi S. (2008). "Genome sequence of the streptomycin-producing microorganism Streptomyces griseus IFO 13350." *J Bacteriol* **190**(11): 4050-60.

Olano C., Mendez C. and Salas J. A. (2009). "Antitumor compounds from marine actinomycetes." *Mar Drugs* **7**(2): 210-48.

Oliynyk M., Stark C. B., Bhatt A., Jones M. A., Hughes-Thomas Z. A., Wilkinson C., Oliynyk Z., Demydchuk Y., Staunton J. and Leadlay P. F. (2003). "Analysis of the biosynthetic gene cluster for the polyether antibiotic monensin in Streptomyces cinnamonensis and evidence for the role of monB and monC genes in oxidative cyclization." *Mol Microbiol* **49**(5): 1179-90.

Palumbi S. R. (2001). "Humans as the world's greatest evolutionary force." *Science* **293**(5536): 1786-90.

Pang X., Aigle B., Girardet J. M., Mangenot S., Pernodet J. L., Decaris B. and Leblond P. (2004). "Functional angucycline-like antibiotic gene cluster in the terminal inverted repeats of the Streptomyces ambofaciens linear chromosome." *Antimicrob Agents Chemother* **48**(2): 575-88.

Parajuli N., Basnet D. B., Chan Lee H., Sohng J. K. and Liou K. (2004). "Genome analyses of Streptomyces peucetius ATCC 27952 for the identification and comparison of cytochrome P450 complement with other Streptomyces." *Arch Biochem Biophys* **425**(2): 233-41.

Parsek M. R. and Fuqua C. (2004). "Biofilms 2003: emerging themes and challenges in studies of surface-associated microbial life." *J Bacteriol* **186**(14): 4427-40.

Passantino R., Puglia A. M. and Chater K. (1991). "Additional copies of the actII regulatory gene induce actinorhodin production in pleiotropic bld mutants in Streptomyces coelicolor A3(2)." *J. Gen. Microbiol* **137**: 2059-2064.

Pel H. J., de Winde J. H., Archer D. B., Dyer P. S., Hofmann G., Schaap P. J., Turner G., de Vries R. P., Albang R., Albermann K., Andersen M. R., Bendtsen J. D., Benen J. A., van den Berg M., Breestraat S., Caddick M. X., Contreras R., Cornell M., Coutinho P. M., Danchin E. G., Debets A. J., Dekker P., van Dijck P. W., van Dijk A., Dijkhuizen L., Driessen A. J., d'Enfert C., Geysens S., Goosen C., Groot G. S., de Groot P. W., Guillemette T., Henrissat B., Herweijer M., van den Hombergh J. P., van den Hondel C. A., van der Heijden R. T., van der Kaaij R. M., Klis F. M., Kools H. J., Kubicek C. P., van Kuyk P. A., Lauber J., Lu X., van der Maarel M. J., Meulenberg R., Menke H., Mortimer M. A., Nielsen J., Oliver S. G., Olsthoorn M., Pal K., van Peij N. N., Ram A. F., Rinas U., Roubos J. A., Sagt C. M., Schmoll M., Sun J., Ussery D., Varga J., Vervecken W., van de Vondervoort P. J., Wedler H., Wosten H. A., Zeng A. P., van Ooyen A. J., Visser J. and Stam H. (2007). "Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88." *Nat Biotechnol* **25**(2): 221-31.

Perez-Llarena F. J., Liras P., Rodriguez-Garcia A. and Martin J. F. (1997). "A regulatory gene (ccaR) required for cephamycin and clavulanic acid production in Streptomyces clavuligerus: amplification results in overproduction of both beta-lactam compounds." *J Bacteriol* **179**(6): 2053-9.

Pernodet J. L., Simonet J. M. and Guerineau M. (1984). "Plasmids in different strains of Streptomyces ambofaciens: free and integrated form of plasmid pSAM2." *Mol Gen Genet* **198**(1): 35-41.

Pernodet J. L., Gourmelen A., Blondelet-Rouault M. H. and Cundliffe E. (1999). "Dispensable ribosomal resistance to spiramycin conferred by srmA in the spiramycin producer Streptomyces ambofaciens." *Microbiology* **145** (Pt 9):2355-64.

Petkovic H., Lill R. E., Sheridan R. M., Wilkinson B., McCormick E. L., McArthur H. A., Staunton J., Leadlay P. F, Kendrew S. G. (2003). "A novel erythromycin, 6-desmethyl erythromycin D, made by substituting an acyltransferase domain of the erythromycin polyketide synthase." *J Antibiot (Tokyo)* **56**(6): 543-51.

Pfeifer B. A., Admiraal S. J., Gramajo H., Cane D. E. and Khosla C. (2001). "Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli." *Science* **291**(5509): 1790-2.

Piel J. (2002). "A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles." *Proc Natl Acad Sci U S A* **99**(22): 14002-7.

Pinnert-Sindico S. (1954). "[Not Available]." *Ann Inst Pasteur (Paris)* **87**(6): 702-7.

Quiros L. M., Aguirrezabalaga I., Olano C., Mendez C. and Salas J. A. (1998). "Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by Streptomyces antibioticus." *Mol Microbiol* **28**(6): 1177-85.

Ramos I., Guzman S., Escalante L., Imriskova I., Rodriguez-Sanoja R., Sanchez S. and Langley E. (2004). "Glucose kinase alone cannot be responsible for carbon source regulation in Streptomyces peucetius var. caesius." *Res Microbiol* **155**(4): 267-74.

Recio E., Colinas A., Rumbero A., Aparicio J. F. and Martin J. F. (2004). "PI factor, a novel type quorum-sensing inducer elicits pimaricin production in Streptomyces natalensis." *J Biol Chem* **279**(40): 41586-93.

Reid R., Piagentini M., Rodriguez E., Ashley G., Viswanathan N., Carney J., Santi D. V., Hutchinson C. R. and McDaniel R. (2003). "A model of structure and catalysis for ketoreductase domains in modular polyketide synthases." *Biochemistry* **42**(1): 72-9.

Richardson M. A., Kuhstoss S., Solenberg P., Schaus N. A. and Rao R. N. (1987). "A new shuttle cosmid vector, pKC505, for streptomycetes: its use in the cloning of three different spiramycin-resistance genes from a Streptomyces ambofaciens library." *Gene* **61**(3): 231-41.

Richardson M. A., Kuhstoss S., Huber M. L., Ford L., Godfrey O., Turner J. R. and Rao R. N. (1990). "Cloning of spiramycin biosynthetic genes and their use in constructing Streptomyces ambofaciens mutants defective in spiramycin biosynthesis." *J Bacteriol* **172**(7): 3790-8.

Richet E. and Raibaud O. (1989). "MalT, the regulatory protein of the Escherichia coli maltose system, is an ATP-dependent transcriptional activator." *Embo J* **8**(3): 981-7.

Riedlinger J., Reicke A., Zahner H., Krismer B., Bull A. T., Maldonado L. A., Ward A. C., Goodfellow M., Bister B., Bischoff D., Sussmuth R. D. and Fiedler H. P. (2004). "Abyssomicins, inhibitors of the para-aminobenzoic acid pathway produced by the marine Verrucosispora strain AB-18-032." *J Antibiot (Tokyo)* **57**(4): 271-9.

Rigali S., Nothaft H., Noens E. E., Schlicht M., Colson S., Muller M., Joris B., Koerten H. K., Hopwood D. A., Titgemeyer F. and van Wezel G. P. (2006). "The sugar phosphotransferase system of Streptomyces coelicolor is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development." *Mol Microbiol* **61**(5): 1237-51.

Rigali S., Titgemeyer F., Barends S., Mulder S., Thomae A. W., Hopwood D. A. and van Wezel G. P. (2008). "Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by Streptomyces." *EMBO Rep* **9**(7): 670-5.

Rodriguez E., Menzella H. G., Gramajo H. (2009). "Heterologous expression of polyketides in bacteria." *Methods Enzymol.* **459**: 339-65.

Rodriguez-Garcia A., Sola-Landa A., Apel K., Santos-Beneit F. and Martin J. F. (2009). "Phosphate control over nitrogen metabolism in Streptomyces coelicolor: direct and indirect negative control of glnR, glnA, glnII and amtB expression by the response regulator PhoP." *Nucleic Acids Res* **37**(10): 3230-42.

Rokem J. S., Lantz A. E. and Nielsen J. (2007). "Systems biology of antibiotic production by microorganisms." *Nat Prod Rep* **24**(6): 1262-87.

Rothberg J. M. and Leamon J. H. (2008). "The development and impact of 454 sequencing." *Nat Biotechnol* **26**(10): 1117-24.

Ruan X., Stassi D., Lax S. A. and Katz L. (1997). "A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC 29253, a rapamycin-producing strain." *Gene* **203**(1): 1-9.

Ryu Y. G., Butler M. J., Chater K. F. and Lee K. J. (2006). "Engineering of primary carbohydrate metabolism for increased production of actinorhodin in Streptomyces coelicolor." *Appl Environ Microbiol* **72**(11): 7132-9.

Salas J. A. and Méndez C. (2007). "Engineering the glycosylation of natural products in actinomycetes." *Trends Microbiol.* **15**(5): 219-32.

Sambrook J., Fritsch E. F. and Maniatis T. (1989). "Molecular cloning: a laboratory manual." Cold Spring Harbor Laboratory Press (2nd ed.), Cold Spring Harbor (NY).

Santos-Beneit F., Rodriguez-Garcia A., Sola-Landa A. and Martin J. F. (2009). "Cross-talk between two global regulators in Streptomyces: PhoP and AfsR interact in the control of afsS, pstS and phoRP transcription." *Mol Microbiol* **72**(1): 53-68.

Schenk P. M., Baumann S., Mattes R. and Steinbiss H. H. (1995). "Improved high-level expression system for eukaryotic genes in Escherichia coli using T7 RNA polymerase and rare ArgtRNAs." *Biotechniques* **19**(2): 196-8, 200.

Schloss P. D. and Handelsman J. (2005). "Metagenomics for studying unculturable microorganisms: cutting the Gordian knot." *Genome Biol* **6**(8): 229.

Schneider A. and Marahiel M. A. (1998). "Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in Bacillus subtilis." *Arch Microbiol* **169**(5): 404-10.

Schrijver D. (1999). "A subfamily of MalT related ATP-dependent regulators in the LuxR family." *Microbiology* **145**: 1287-1288.

Sekurova O. N., Brautaset T., Sletta H., Borgos S. E., Jakobsen M. O., Ellingsen T. E., Strom A. R., Valla S. and Zotchev S. B. (2004). "In vivo analysis of the regulatory genes in the nystatin biosynthetic gene cluster of Streptomyces noursei ATCC 11455 reveals their differential control over antibiotic biosynthesis." *J Bacteriol* **186**(5): 1345-54.

Sheldon P. J., Busarow S. B. and Hutchinson C. R. (2002). "Mapping the DNA-binding domain and target sequences of the Streptomyces peucetius daunorubicin biosynthesis regulatory protein, DnrI." *Mol Microbiol* **44**(2): 449-60.

Shen B. (2003). "Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms." *Curr Opin Chem Biol* **7**(2): 285-95.

Shrestha P., Oh T. J., Liou K. and Sohng J. K. (2008). "Cytochrome P450 (CYP105F2) from Streptomyces peucetius and its activity with oleandomycin." *Appl Microbiol Biotechnol* **79**(4): 555-62.

Sola-Landa A., Moura R. S. and Martin J. F. (2003). "The two-component PhoR-PhoP system controls both primary metabolism and secondary metabolite biosynthesis in Streptomyces lividans." *Proc Natl Acad Sci U S A* **100**(10): 6133-8.

Sola-Landa A., Rodriguez-Garcia A., Apel A. K. and Martin J. F. (2008). "Target genes and structure of the direct repeats in the DNA-binding sequences of the response regulator PhoP in Streptomyces coelicolor." *Nucleic Acids Res* **36**(4): 1358-68.

Song L., Barona-Gomez F., Corre C., Xiang L., Udwary D. W., Austin M. B., Noel J. P., Moore B. S. and Challis G. L. (2006). "Type III polyketide synthase beta-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by Streptomyces coelicolor genome mining." *J Am Chem Soc* **128**(46): 14754-5.

Staunton J. and Weissman K. J. (2001). "Polyketide biosynthesis: a millennium review." *Nat Prod Rep* **18**(4): 380-416.

Sun Y., Zhou X., Liu J., Bao K., Zhang G., Tu G., Kieser T. and Deng Z. (2002). "'Streptomyces nanchangensis', a producer of the insecticidal polyether antibiotic nanchangmycin and the antiparasitic macrolide meilingmycin, contains multiple polyketide gene clusters." *Microbiology* **148**(Pt 2): 361-71.

Sunbul M., Zhang K. and Yin J. (2009). "Chapter 10 using phosphopantetheinyl transferases for enzyme posttranslational activation, site specific protein labeling and identification of natural product biosynthetic gene clusters from bacterial genomes." *Methods Enzymol* **458**: 255-75.

Tahlan K., Yu Z., Xu Y., Davidson A. R. and Nodwell J. R. (2008). "Ligand recognition by ActR, a TetR-like regulator of actinorhodin export." *J Mol Biol* **383**(4): 753-67.

Takano E., Gramajo H. C., Strauch E., Andres N., White J. and Bibb M. J. (1992). "Transcriptional regulation of the redD transcriptional activator gene accounts for growth-phase-dependent production of the antibiotic undecylprodigiosin in Streptomyces coelicolor A3(2)." *Mol Microbiol* **6**(19): 2797-804.

Takano E. (2006). "Gamma-butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation." *Curr Opin Microbiol* **9**(3): 287-94.

Tanaka A., Takano Y., Ohnishi Y. and Horinouchi S. (2007). "AfsR recruits RNA polymerase to the afsS promoter: a model for transcriptional activation by SARPs." *J Mol Biol* **369**(2): 322-33.

Thapa L. P., Oh T. J., Lee H. C., Liou K., Park J. W., Yoon Y. J. and Sohng J. K. (2007). "Heterologous expression of the kanamycin biosynthetic gene cluster (pSKC2) in Streptomyces venezuelae YJ003." *Appl Microbiol Biotechnol* **76**(6): 1357-64.

Thapa L. P., Oh T. J., Liou K. and Sohng J. K. (2008). "Biosynthesis of spectinomycin: heterologous production of spectinomycin and spectinamine in an aminoglycoside-deficient host, Streptomyces venezuelae YJ003." *J Appl Microbiol* **105**(1): 300-8.

Torriani A. (1990). "From cell membrane to nucleotides: the phosphate regulon in Escherichia coli." *Bioessays* **12**(8): 371-6.

Towbin H., Staehelin T. and Gordon J. (1992). "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. 1979." *Biotechnology* **24**: 145-9.

Udwary D. W., Zeigler L., Asolkar R. N., Singan V., Lapidus A., Fenical W., Jensen P. R. and Moore B. S. (2007). "Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica." *Proc Natl Acad Sci U S A* **104**(25): 10376-81.

Van Bambeke F., Balzi E. and Tulkens P. M. (2000). "Antibiotic efflux pumps." *Biochem Pharmacol* **60**(4): 457-70.

van Wezel G. P., White J., Hoogvliet G. and Bibb M. J. (2000). "Application of redD, the transcriptional activator gene of the undecylprodigiosin biosynthetic pathway, as a reporter for transcriptional activity in Streptomyces coelicolor A3(2) and Streptomyces lividans." *J Mol Microbiol Biotechnol* **2**(4): 551-6.

van Wezel G. P., McKenzie N. L. and Nodwell J. R. (2009). "Chapter 5. Applying the genetics of secondary metabolism in model actinomycetes to the discovery of new antibiotics." *Methods Enzymol* **458**: 117-41.

Venturi V. (2006). "Regulation of quorum sensing in Pseudomonas." *FEMS Microbiol Rev* **30**(2): 274-91.

Vilches C., Hernandez C., Mendez C. and Salas J. A. (1992). "Role of glycosylation and deglycosylation in biosynthesis of and resistance to oleandomycin in the producer organism, Streptomyces antibioticus." *J Bacteriol* **174**(1): 161-5.

Walker J. E., Saraste M., Runswick M. J. and Gay N. J. (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." *Embo J* **1**(8): 945-51.

Walsh C. T. (2003). "Antibiotics. Actions, origins, resistance," ASM Press (Washington, D.C.).

Wang J., Soisson S. M., Young K., Shoop W., Kodali S., Galgoci A., Painter R., Parthasarathy G., Tang Y. S., Cummings R., Ha S., Dorso K., Motyl M., Jayasuriya H., Ondeyka J., Herath K., Zhang C., Hernandez L., Allocco J., Basilio A., Tormo J. R., Genilloud O., Vicente F., Pelaez F., Colwell L., Lee S. H., Michael B., Felcetto T., Gill C., Silver L. L., Hermes J. D., Bartizal K., Barrett J., Schmatz D., Becker J. W., Cully D. and Singh S. B. (2006). "Platensimycin is a selective FabF inhibitor with potent antibiotic properties." *Nature* **441**(7091): 358-61.

Weissman K. J. (2009). "Introduction to polyketide biosynthesis." *Methods Enzymol* **459**: 3-16.

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." *J Am Chem Soc* **127**(47): 16442-52.

Wenzel S. C., Gross F., Zhang Y., Fu J., Stewart A. F. and Muller R. (2005). "Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering." *Chem Biol* **12**(3): 349-56.

Werck-Reichhart D. and Feyereisen R. (2000). "Cytochromes P450: a success story." *Genome Biol* **1**(6): REVIEWS3003.

White J. and Bibb M. (1997). "bldA dependence of undecylprodigiosin production in Streptomyces coelicolor A3(2) involves a pathway-specific regulatory cascade." *J Bacteriol* **179**(3): 627-33.

White-Phillip J., Thibodeaux C. J. and Liu H. W. (2009). "Enzymatic synthesis of TDP-deoxysugars." *Methods Enzymol* **459**: 521-44.

Wietzorrek A. and Bibb M. (1997). "A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold." *Mol Microbiol* **25**(6): 1181-4.

Wilke M. S., Lovering A. L. and Strynadka N. C. (2005). "Beta-lactam antibiotic resistance: a current structural perspective." *Curr Opin Microbiol* **8**(5): 525-33.

Wilkinson C. J., Hughes-Thomas Z. A., Martin C. J., Bohm I., Mironenko T., Deacon M., Wheatcroft M., Wirtz G., Staunton J. and Leadlay P. F. (2002). "Increasing the efficiency of heterologous promoters in actinomycetes." *J Mol Microbiol Biotechnol* **4**(4): 417-26.

Wilkinson B. and Micklefield J. (2007). "Mining and engineering natural-product biosynthetic pathways." *Nat Chem Biol* **3**(7): 379-86.

Wilson D. J., Xue Y., Reynolds K. A. and Sherman D. H. (2001). "Characterization and analysis of the PikD regulatory factor in the pikromycin biosynthetic pathway of Streptomyces venezuelae." *J Bacteriol* **183**(11): 3468-75.

Witkowski A., Naggert J., Witkowska H. E., Randhawa Z. I. and Smith S. (1992). "Utilization of an active serine 101----cysteine mutant to demonstrate the proximity of the catalytic serine 101 and histidine 237 residues in thioesterase II." *J Biol Chem* **267**(26): 18488-92.

Wu K., Chung L., Revill W. P., Katz L. and Reeves C. D. (2000). "The FK520 gene cluster of Streptomyces hygroscopicus var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units." *Gene* **251**(1): 81-90.

Xue Y., Zhao L., Liu H. W. and Sherman D. H. (1998). "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity." *Proc Natl Acad Sci U S A* **95**(21): 12111-6.

Yadav G., Gokhale R. S. and Mohanty D. (2003a). "SEARCHPKS: A program for detection and analysis of polyketide synthase domains." *Nucleic Acids Res* **31**(13): 3654-8.

Yadav G., Gokhale R. S. and Mohanty D. (2003b). "Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases." *J Mol Biol* **328**(2): 335-63.

Yadav G., Gokhale R. S. and Mohanty D. (2009). "Towards prediction of metabolic products of polyketide synthases: an in silico analysis." *PLoS Comput Biol* **5**(4): e1000351.

Yeung K. S. and Paterson I. (2002). "Actin-binding marine macrolides: total synthesis and biological importance." *Angew Chem Int Ed Engl* **41**(24): 4632-53.

Yuan Y., Chung H. S., Leimkuhler C., Walsh C. T., Kahne D. and Walker S. (2005). "In vitro reconstitution of EryCIII activity for the preparation of unnatural macrolides." *J Am Chem Soc* **127**(41): 14128-9.

Zalacain M. and Cundliffe E. (1989). "Methylation of 23S rRNA caused by tlrA (ermSF), a tylosin resistance determinant from Streptomyces fradiae." *J Bacteriol* **171**(8): 4254-60.

Zerikly M. and Challis G. L. (2009). "Strategies for the discovery of new natural products by genome mining." *Chembiochem* **10**(4): 625-33.

Zhang C., Griffith B. R., Fu Q., Albermann C., Fu X., Lee I. K., Li L. and Thorson J. S. (2006). "Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions." *Science* **313**(5791): 1291-4.

Zhang W. and Tang Y. (2009). "In vitro analysis of type II polyketide synthase." *Methods Enzymol* **459**: 367-93.

Zhu L., Luzhetskyy A., Luzhetska M., Mattingly C., Adams V., Bechthold A. and Rohr J. (2007). "Generation of new landomycins with altered saccharide patterns through over-expression of the glycosyltransferase gene lanGT3 in the biosynthetic gene cluster of landomycin A in Streptomyces cyanogenus S-136." *Chembiochem* **8**(1): 83-8.

Zotchev S. B., Stepanchikova A. V., Sergeyko A. P., Sobolev B. N., Filimonov D. A. and Poroikov V. V. (2006). "Rational design of macrolides by virtual screening of combinatorial libraries generated through in silico manipulation of polyketide synthases." *J Med Chem* **49**(6): 2077-87.

RESUME

INTRODUCTION

La recherche de nouvelles molécules bioactives est plus que jamais un sujet d'actualité ces dernières années. En effet, le développement de souches pathogènes multi-résistantes (telles que les souches MRSA –« methicilin-resistant *Staphylococcus aureus* »-), ainsi que l'apparition de nouvelles maladies infectieuses (par exemple la maladie de Lyme ou le SIDA), mais également de maladies infectieuses anciennes comme la tuberculose ou la malaria soulignent le besoin urgent de trouver des nouvelles drogues.

Plusieurs stratégies ont été développées pour isoler de nouveaux composés d'intérêt médical, comme, par exemple, l'amélioration des méthodes de criblages ou encore la production par chimie combinatoire de banques de nouvelles molécules. Cependant, ces techniques se sont révélées peu fructueuses jusqu'à présent. L'exploitation des données génomiques (ou « genome mining ») apparaît comme une approche prometteuse pour la recherche de nouveaux métabolites. En effet, le séquençage de génomes de différents microorganismes a révélé que la capacité de synthèse de métabolites secondaires (comprenant les antibiotiques) d'une espèce donnée excède largement celle révélée par des méthodes de criblage traditionnelles.

La recherche de gènes impliqués dans la biosynthèse de ces molécules est facilitée par le fait que les gènes participant à la synthèse d'un composé donné sont groupés sur le chromosome (notion de cluster de gènes). Ces clusters comprennent également les gènes de régulation et de résistance. Mais la plupart de clusters identifiés sont considérés comme cryptiques puisque le métabolite produit par un cluster donné n'a jamais été détecté auparavant. Différents approches ont été décrites afin d'identifier ces molécules comme, par exemple, l'expression chez un hôte hétérologue ou encore l'interruption d'un gène de biosynthèse associé à une comparaison de profil métabolique entre les souches sauvage et mutante par des méthodes chromatographiques. La détection et l'isolation du composé recherché peuvent être facilitées grâce à la prédiction de la structure du produit.

Le caractère cryptique de certains clusters de gènes de métabolites secondaires peut être lié au fait que les gènes ne soient pas ou très peu exprimés en condition standard de laboratoire. Il est en effet possible que ces clusters ne soient exprimés que dans des conditions très particulières. Par conséquent, l'activation de l'expression de gènes de biosynthèse peut s'avérer être une étape cruciale. Cette activation peut être obtenue par modification de conditions de culture. Mais trouver les conditions de culture dans lesquelles le microorganisme exprime ces clusters peut s'avérer très difficile. La manipulation de gènes de régulation constitue une approche prometteuse. Il est possible d'agir à des niveaux différents. En effet, la régulation de l'expression des gènes de biosynthèse de métabolites secondaires et en particulier

d‘antibiotiques est un processus complexe impliquant à la fois des régulateurs spécifiques et pléïotropiques. Il est également possible de jouer sur des mécanismes plus centraux telle que la machinerie transcriptionnelle et traductionnelle.

La majorité de métabolites secondaires appartiennent à la famille des polycétides ou des peptides à synthèse non ribosomique, qui sont synthétisés par de complexes multienzymatiques nommés respectivement polycétides synthases (PKS) ou synthases de peptides non ribosomiques (NRPS). Les produits de ces deux familles possèdent de structures chimiques très différentes et complexes, qui sont le résultat de condensations successives de petites molécules d’acides carboxyliques, pour les polycétides, ou d’acides aminés atypiques, pour les peptides non ribosomiques. Les chaînes polycétidiques ou peptidiques qui vont se former peuvent subir d’autres modifications grâce à l’action d’enzymes de modification (ou « tailoring » enzymes), telles que des glycosyltransférases, de méthyltransférases ou encore des oxygénases, augmentant ainsi la diversité structurale des composés produits. Ces modifications confèrent des propriétés biologiques intéressantes.

Les actinomycètes constituent les principaux producteurs de métabolites secondaires. En fait, près des deux tiers des antibiotiques d’origine microbienne utilisés en thérapie humaine sont produits par les bactéries du genre *Streptomyces*, des bactéries filamenteuses du sol. La plupart de ces composés ont des applications non seulement comme agents antibactériens, mais également comme antifongiques, immunosuppresseurs ou encore antitumoraux. Certains d’entre eux sont également utilisés en agriculture pour la protection des plantes notamment grâce à leurs activités fongicides ou insecticides. Dans notre laboratoire, le modèle d’étude est la bactérie *Streptomyces ambofaciens* ATCC23877, une souche connue pour produire la spiramycine, un macrolide utilisé en thérapie humaine, et la congocidine, un antibiotique appartenant à la classe de pyrrolamides. Le séquençage du génome, en particulier des extrémités de son chromosome linéaire, a mise en évidence douze clusters de gènes potentiellement impliqués dans la synthèse d’autres métabolites secondaires indiquant que *S. ambofaciens* représente une source intéressante pour la recherche de nouveaux métabolites.
Mon projet de thèse s’inscrivait dans ce contexte. Il faisait partie intégrante du projet européen ActinoGEN ($6^{ème}$ PCRD) dont l’objectif portait sur la recherche de nouveaux antibiotiques basée sur l’exploitation des actinomycetes. Nous nous sommes intéressés à un cluster de gènes de polycétides synthases (PKS) de type I, localisé dans le bras droit du chromosome. La grande taille de gènes de biosynthèse ainsi que leur organisation originelle suggérait la production d’un nouveau métabolite. Les objectifs de ma thèse étaient donc d’identifier puis de caractériser le métabolite associé à ce cluster et de déterminer ses propriétés biologiques. En parallèle, nous nous sommes intéressés à la régulation de l’expression du cluster.

RESULTATS

1. Analyse *in silico* du cluster de gènes PKS de type I

Le cluster est composé de neuf gènes codant des PKS (*sam*R0465, 0466, 0467, 0474, CDS2, 0475, 0476, CDS1 et 0477) et de plusieurs gènes impliqués dans des réaction de modification de la structure polycétidique. Il contient également des gènes potentiellement impliqués dans la régulation et dans des mécanismes de résistance (Fig. 1). Les limites du cluster ont été établies entre le gène *sam*R0465 et le gène *sam*R0487, selon la fonction putative de chaque gène. Le cluster comprendrait donc 25 gènes.

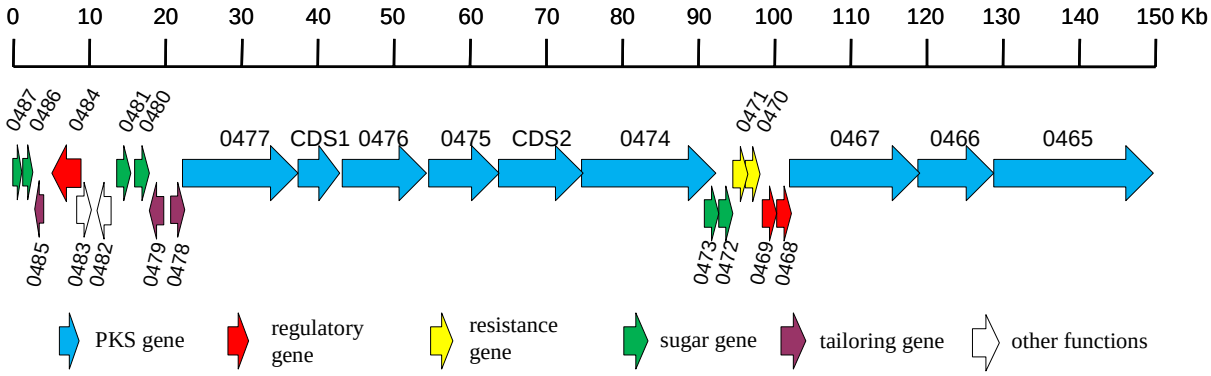


Figure 1. Organisation génétique du cluster de gènes de PKS de type I localisé dans le bras chromosomique droit de *Streptomyces ambofaciens.*

Les gènes de PKS de type I sont également appelés PKS modulaires. Il s'agit de très grandes protéines multifonctionnelles organisées en unités répétées appelées modules (Donadio *et al.*, 1991; Donadio and Katz, 1992). Ces enzymes catalysent la condensation successive de molécules d'acides carboxyliques (acétate, propionate, butyrate, …) en une longue chaîne carbonée. Par analyse *in silico* des modules, il est possible de prédire la taille et la structure du squelette polycétidique du métabolite attendu.

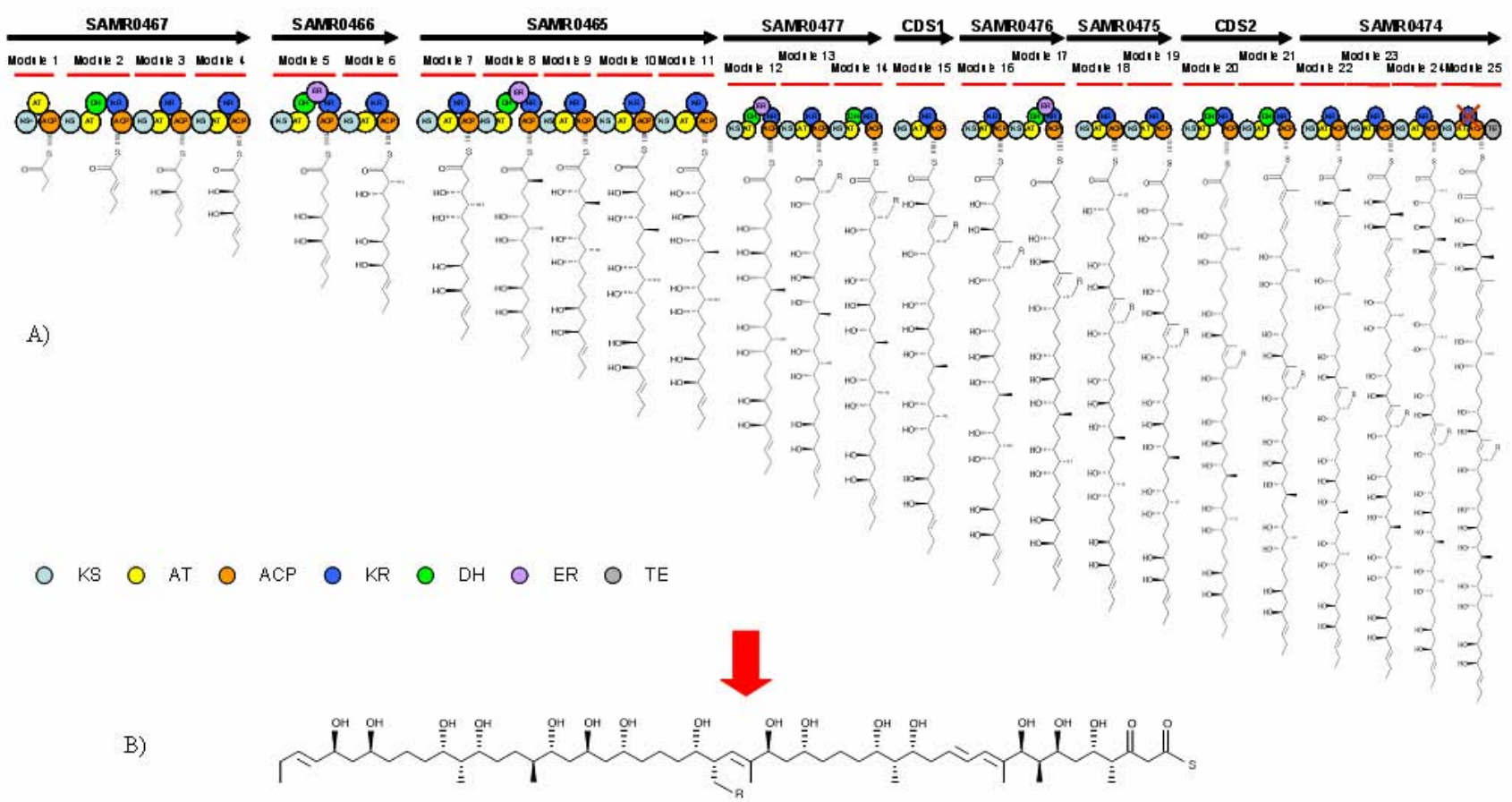


Figure 2 A) Organisation modulaire de neuf gènes de PKS. Pour chaque module l'intermédiaire a été dessiné, avec les modifications attendues pour le β-keto et avec la stéréochimie prédite. Le domaine KR non fonctionnelle est marque par une croix **B)** Prédiction du squelette polycétidique. Le substituent R représentes le précurseur inconnu incorporé par le domaine AT13.
KS, Cétoacylsynthase ; AT, Acyltransférase ; ACP, Acyl carrier protein ; KR, Cétoréductase ; DH, Déshydratase ; ER, Enoyl réductase.

L'utilisation du programme SEARCHPKS (Yadav *et al.*, 2003) a permis de déterminer le nombre total de modules portés par les neuf gènes de PKS de notre cluster et de définir les différents domaines catalytiques présents dans chacun des modules identifiés (au total 112) (Fig. 2 A). Le gène *sam*R0467 code la première sous-unité du complexe de PKS de type I puisqu'il contient le module d'initiation alors que *sam*R0474 code la dernière sous-unité intervenant dans la synthèse. En effet, le dernier domaine catalytique du produit de *sam*R0474 correspond à un domaine thioestérase (TE). Ce domaine est impliqué dans la libération de la chaîne polycétidique par hydrolyse de la liaison thioester qui la lie au dernier domaine ACP (Acyl carrier protein). Parfois le domaine TE peut aussi être impliqué dans la cyclisation de la molécule polycétidique. Les neufs gènes codant les PKS sont répartis en deux loci distincts (*sam*R0467-0465 et *sam*R0477-0474) séparés par un groupe de six gènes (Fig. 1). Ainsi, il n'y a pas une relation directe stricte entre l'ordre des gènes codant les sous-unités du complexe PKS et celui dans lequel ces sous-unités interviendraient dans la synthèse à savoir *sam*R0467-466-465-477-CDS1-476-475-CDS2-474. Cet ordre est établi à partir de l'organisation des gènes par rapport à *sam*R0467 et *sam*R0474.

L'analyse de certains domaines enzymatiques permet de prédire la nature des précurseurs incorporés dans la structure, mais aussi leur stéréochimie. En plus, le type de domaines enzymatiques retrouvés dans chaque module permet d'établir les modifications que les unités carbonées vont subir pendant l'assemblage du polycétide.

La nature des acides carboxyliques incorporés dans la chaîne polycétidique a été prédite à partir des domaines acyltransférases (AT) responsables du choix de l'unité carbonée utilisée par chaque module. En effet, plusieurs études ont montré que la présence de certains motifs conservés à l'intérieur des domaines AT et de certaines caractéristiques structurales du site actif permettait de déterminer le type d'acide carboxylique incorporé. Ainsi, nous avons pu prédire que la première unité permettrait l'incorporation d'un propionate et les unités d'extension l'incorporation de 16 acétates, 6 propionates et 2 butyrates. Le niveau de réduction des groupes cétone incorporés dans la molécule en formation a été prédit en fonction de la présence dans chaque module des domaines accessoires : le domaine KR, déshydratase (DH) et énoyl-réductase (ER). Enfin, la stéréochimie de chaque précurseur, en particulier celle associé au C α et β, a été déterminée par l'analyse des domaines cétoréductase (KR ; Keatinge-Clay, 2007). Il est à noter que tous les domaines enzymatiques apparaissent être fonctionnels, déduction basée sur la conservation des acides aminés catalytiques, à l'exception du domaine KR du module 25. En effet, la tyrosine, appartenant à la triade catalytique Tyr-Ser-Asn est remplacée par un résidu arginine (KR4.0474). Or, il a été montré que le remplacement de cet acide aminé par un résidu Phe dans le domaine KR6 de la protéine DEBS3, l'une des trois PKS intervenant dans la synthèse de l'érythromycine, abolissait toute activité réductase par perte du groupement hydroxyle donneur de proton (Reid *et al.*, 2003). L'acide aminé Asn de la triade est également absent dans le domaine KR du module 24 (substitution par un résidu sérine). Enfin, le résidu

conservé Lys, qui est également important pour l'activité réductase (Reid *et al.*, 2003), est remplacé par une alanine.
D'après ces analyses, nous avons proposé que le squelette polycétidique produit par les PKS aurait la structure présentée en figure 2 B. Il s'agit d'une longue chaîne contenant quinze groupes hydroxyles, un groupe céto, quatre carbones saturés et un diène. La masse prédite de cette molécule était de 1127 Da et sa formule moléculaire $C_{61}H_{107}O_{18}$. Il était très probable que cette molécule soit cyclisée. En effet, le domaine TE présent dans le dernier module est plus proche des domaines TE présents dans les clusters de PKS de type I impliqués dans la synthèse de molécules cycliques que de ceux impliqués dans la synthèse de dérivés polycétidiques linéaires (James McAlpine, communication personnelle). Cependant, il n'est pas possible de prédire quel groupement hydroxyle serait impliqué dans cette étape de cyclisation. Le noyau macrolactone le plus important qui puisse être généré après cyclisation contiendrait 48 atomes.
L'analyse du cluster a mis en évidence la présence d'un gène, *sam*R0481, codant pour une glycosyltransférase qui serait responsable de la fixation d'un sucre sur le cycle lactone. Donc, la structure finale du métabolite synthétisé serait probablement un macrolide. L'analyse de cette glycosyltransférase à l'aide du programme SEARCHGTr (Kamra *et al*., 2005)suggère que le sucre reconnu et attaché au noyau aglycone par le produit de *sam*R0481 serait un sucre aminé, soit un résidu désosamine ou un résidu mycaminose. Cinq ORF présentes dans le cluster (*sam*R0472, 0473, 0480, 0486 et 0487) sont paralogues (ou xénologues) aux cinq ORF du cluster spiramycine de *S. ambofaciens* qui sont responsables de la synthèse de la mycaminose, l'un des trois sucres attachés au noyau aglycone de l'antibiotique.

L'ensemble de ces analyses a donc permis de prédire que la molécule produite (vraisemblablement un macrolide) répondrait à la formule moléculaire $C_{69}H_{127}N_1O_{22}$ et que sa masse moléculaire serait d'environ 1322Da. Cette prédiction représente une étape essentielle afin d'identifier le composé produit par le cluster de PKS de type I.

Le cluster contient aussi d'autres gènes dont la fonction a pu être prédite. Les gènes *sam*R0478 et *sam*R0479 codent potentiellement pour des cytochromes P450, impliqués dans de réactions d'hydroxylation de la structure polycétidique. Le produit du gène *sam*R0485 est homologue à une thioestérase de type II, probablement responsable de l'élimination de produits aberrants attachés aux PKS. En revanche, nous n'avons pas pu proposer une fonction spécifique aux deux gènes *sam*R0482 et *sam*R0483, codant respectivement pour un acyl-CoA synthase et un acyl-CoA carboxylase.
Les produits de gènes *sam*R0470 et *sam*R0471 formeraient un ABC (ATP-binding cassette) transporteur, souvent caractérisé comme système de transport et de résistance de métabolites secondaires.

Enfin, trois régulateurs potentiels ont été identifiés dans le cluster et ils pourraient être impliqués dans la régulation de la biosynthèse du métabolite. Les gènes *sam*R0468 et *sam*R0469 codent les sous unités d'un système à deux composants (SAMR0468, le régulateur et SAMR0469, l'histidine kinase). Le troisième gène, *sam*R0484, code un régulateur de la famille LAL (pour « Large ATP-binding regulators of the LuxR family »).

2. Activation d'un cluster silencieux et isolation d'un nouveau macrolide produit par *Streptomyces ambofaciens* ATCC23877

Le macrolide prédit par les analyses *in silico* n'avait jamais été détecté ou isolé auparavant, probablement parce que le cluster responsable de sa synthèse est silencieux dans les conditions testées ou que son expression est trop faible pour permettre la production du métabolite. Pour confirmer notre hypothèse, des analyses transcriptionelles par RT-PCR à partir d'ARN extrait de la souche de *S. ambofaciens* ATCC23877 ont été réalisées. Les résultats obtenus ont montré que les gènes codant les PKS SAMR0467 et SAMR0477 n'étaient pas transcrits dans les conditions utilisées (cultures réalisées en milieu HT ou en milieu MP5, milieux utilisés pour la production du macrolide spiramycine) et ce quelque soit la phase de croissance testée. Le cluster était donc silencieux. Afin d'identifier le macrolide produit par ce cluster, une approche aurait été de tester différentes conditions de croissance afin de déterminer celles permettant l'expression des gènes de PKS. Une autre approche consistait à activer de façon « artificielle » l'expression de ces gènes en jouant sur la régulation de l'expression du cluster.
Nous avons opté pour cette dernière stratégie. Il s'agissait de manipuler de gènes de régulations, précédemment identifiés dans le cluster, c'est-à-dire les gènes *sam*R0468-0469, codant un système à deux composants, et le gène *sam*R0484, codant un régulateur de la famille LAL. Le choix de *sam*R0484 s'avérait d'autant plus judicieux qu'il avait été démontré que deux membres de cette famille, PikD et RapH, agissent comme activateurs spécifiques des gènes de biosynthèse des antibiotiques macrolides pikromycine et rapamycine respectivement chez *S. venezuelae* et *S. hygroscopicus* (Wilson *et al.*, 2001; Kuscer *et al.*, 2007). Sur la base de ces observations, nous avons décidé de surexprimer ces trois gènes de régulations.
Les gènes *sam*R0468-9 et *sam*R0484 ont donc été clonés dans le vecteur conjugatif et intégratif pIB139 sous le contrôle de *ermE*p*, un promoteur constitutif fort. Les constructions ont alors été introduites chez la souche *S. ambofaciens* ATCC23877 dans laquelle les plasmides sont intégrés (recombinaison site-spécifique) dans le chromosome au niveau du site d'attachement *attB* du phage ϕC31. Deux souches mutantes ont été obtenues, ATCC/OE468-9 et ATCC/OE484. En parallèle, le vecteur seul a été introduit chez la souche sauvage, et le clone obtenu (ATCC/pIB139) a servi de contrôle.

Des analyses transcriptionelles comparatives menées par RT-PCR à partir d'ARN extrait des souches mutantes et du contrôle ont été réalisées. Chez les exconjugants surexprimant le régulateur LAL, l'expression des gènes codant les PKS était induite et ce quelque soit la phase de croissance (d'exponentielle à stationnaire), confirmant que SAMR0484 agit comme un activateur transcriptionnel (Figure 3).

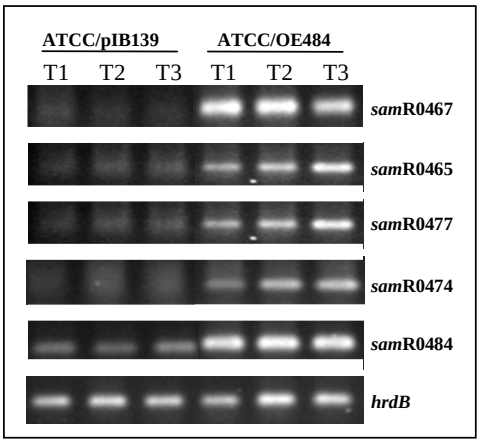


Figure 3 Analyses transcriptionelles par RT-PCR de la souche mutant ATCC/OE484 et du control ATCC/pIB139, en culture MP5. L'expression de gènes de PKS *sam*R0467, 0465, 0477, 0474 a été testé. Le gène *hrdB* code pour le sigma facteur majeur en *Stremptomyces* et il est utilisé comme control interne.

Nous avons également testé l'effet de la surexpression simultanée des gènes *sam*R0468 et *sam*R0469. Dans ce cas, aucun effet sur la transcription des gènes de biosynthèse n'a pu être observé.

La souche mutante ATCC/OE484 a donc été utilisée pour la détection du nouveau macrolide qui serait produit par *S. ambofaciens*. Afin d'identifier ce composé produit par le cluster de type I, nous avons opté pour une approche de « profiling » en chromatographie liquide couplée à la spectrométrie de masse (LC-MS) en collaboration avec l'équipe de Greg Challis de l'université de Warwick (GB) dans le cadre du programme européen ActinoGEN ($6^{ème}$ PCRD).
D'après la structure prédite à partir de l'analyse des PKS, la molécule recherchée apparaît être relativement hydrophile, propriété due à la présence de nombreux groupes hydroxyles. Par conséquent, des extractions à l'aide de méthanol ont été réalisées à partir de mycélium obtenu après culture liquide en milieu MP5. Les extraits du mutant (ATCC/OE484) et du contrôle (ATCC/pIB139) ont été analysés par LC-MS. La comparaison des profils obtenus pour la souche surexprimant le régulateur LAL et le témoin négatif contenant le vecteur pIB139 a permis de mettre en évidence deux pics (*m/z* 673 et 680) présents uniquement dans le profil de la première souche (Fig. 4). La masse réelle de deux pics e dest 1362 et 1376, et leur formule moléculaire, $C_{72}H_{132}NO_{22}$ et $C_{73}H_{134}NO_{22}$. Malgré quelques différences, ces valeurs sont en

accord avec celles prédites par analyse *in silico*. La différence entre les deux molécules (14 Da) pourrait être due à la présence d'un groupement éthyle au lieu d'un groupement méthyle.

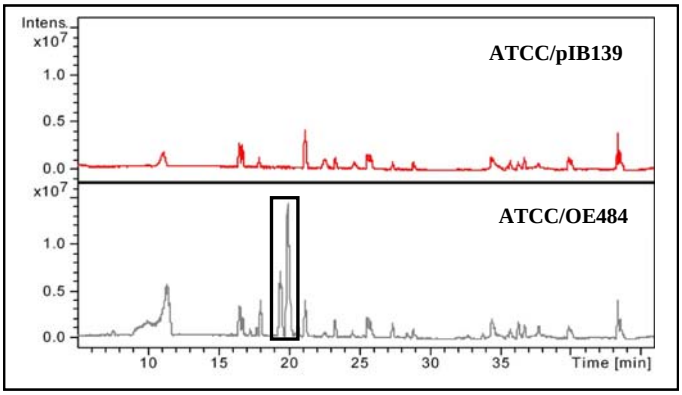


Figure 4. Chromatogramme de LC-MS d'un extrait du mycélium des souches ATCC/pIB139 et ATCC/OE484. Les deux pics encadrés sont double chargés avec un *m/z* of 673 et 680.

Pour confirmer que les métabolites identifiés dans la souche ATCC/OE484 étaient produits par le cluster, nous avons analysé un mutant surexprimant le régulateur LAL en parallèle avec une souche dérivée délétée en plus pour le premier gène de PKS (*sam*R0467). Dans ce dernier mutant, les pics 673 et 680 disparaissent, confirmant que les deux métabolites sont associés au cluster de gènes de PKS de type I.

Les molécules correspondant aux pics ont pu être purifiées par HPLC. Apres lyophilisation, les échantillons ont été analysés par résonance magnétique nucléaire (RMN) dans différents solvants. Nous avons ainsi découverts quatre nouvelles formes d'un macrolide à 50 atomes, nommé sambomycine A, B, C et D (Fig. 5).

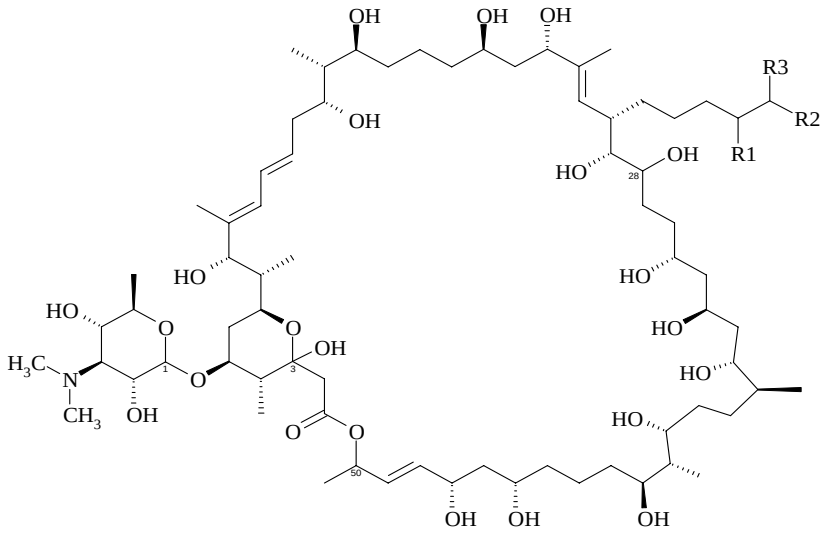


Figure 5. Structure de sambomycin produits. Selon la nature de substitues R1, R2 et R3 on peut obtenir la sambomycin A et B (R1=H ; R2=R3=Me) ou la sambomycin C et D (R1=R2=H ; R3=Me).

Les molécules sambomycines A et B sont des isomères ayant comme formule/masse moléculaires $C_{73}H_{134}NO_{22}$/1376Da. Les molécules sambomycines C et D sont aussi des isomères avec la formule/masse moléculaires $C_{72}H_{132}NO_{22}$/1363Da.

Les molécules purifiées ont été testées pour chercher d'éventuelles propriétés biologiques intéressantes. Les deux isomères sont actifs contre les bactéries Gram-positives, telles que *Bacillus subtilis*, *Micrococcus luteus*, *Enterococcus faecalis* et *Staphylococcus aureus* (Table 1). En particulier, le mélange sambomycine C/D apparaît plus efficace à la même concentration. En revanche, aucune activité bactérienne contre les bactéries Gram-négatives et aucune activité antifongique n'ont pu être observées.

Organisme	**Sambomycine**		**Vancomycin**	**Doxorubicin**
	A/B	**C/D**		
Activité antibactérienne (IC90, µg/ml):				
Bacillus subtilis BGSC 1A72	ND	33,53 +/- 0,86	0,25-1	
Enterococcus faecalis LG40	ND	8,65 +/- 0,49	0,25-1	
Staphylococcus aureus LG21	ND	33,95 +/- 0,16	0,25-1	
Activité antiproliférative (IC50, µM):				
HT29	1.77 +/- 0.04	1.74 +/- 0.04		1.32 +/- 0.08
cytotoxicité (IC50, µM):				
CHO-K1	8.47 +/- 0,67	8.46 +/- 0,52		1.99 +/-0.25

Table 1. Activités biologiques liées à la sambomcyine. La valeur IC90 indique la concentration nécessaire à inhiber la croissance de 90% de la population cellulaire. HT29, cellules tumorigènes d'adénocarcinome de colon humain; CHO-K1, cellules saines d'ovaire de hamster chinois. ND, la valeur IC90 n'a pas été déterminée pour la sambomycine A/B, qui montre cependant une activité antibactérienne. La doxorubicine a été utilisée comme référence dans les tests d'activité antiproliférative et de cytotoxicité.

Des tests de cytotoxicité ont été réalisés sur des cellules animales (cellules saines CHO-K1, c'est-à-dire cellules saines d'ovaire de hamster chinois). Ils ont révélé que les macrolides sont moins toxiques que la doxorubicine, utilisée comme référence (Table 1). La doxorubicine est un dérivé polycétidique aromatique produit par *Streptomyces peucetius* et qui est utilisé en thérapie pour le traitement de certains cancers, en particulier de cancers du sein.
Des tests ont alors été réalisés afin de déterminer les propriétés antiprolifératives des composés sambomycines. L'activité des sambomycines a été mesurée contre la lignée cellulaire HT29, une lignée correspondant à des cellules tumorigènes d'adénocarcinome de colon humain. De façon très intéressante, les mélanges sambomycine A/B et C/D ont montré une activité antiproliferative du même ordre de grandeur que celle observée avec la doxorubicine (Table 1).

3. Etude de la biosynthèse de la sambomycine

La détermination de la structure de la sambomycine a confirmé la plupart de nos prédictions, comme, par exemple, la présence d'un sucre aminé, la mycaminose. Elle a aussi mis en évidence des aspects très intéressants liés à la biosynthèse de cette molécule. Une connaissance plus approfondie des mécanismes conduisant à la production de la sambomycine peut être très utile dans le cadre de la biologie combinatoire, en vue d'obtenir des nouveaux composés avec des propriétés biologiques plus performantes ou plus spécifiques.

La sambomycine contient un noyau lactone à 50 atomes, une taille jamais décrite dans la littérature (Fig. 5). La cyclisation doit se produire entre le premier et le dernier précurseur incorporé dans le squelette polycétidique. Le premier substrat est un propionate qui, après incorporation, ne contient de groupe hydroxyle. Des expériences d'incorporation d'isotope marqué ^{18}O ont permis de démontrer la présence de deux oxygènes incorporés après la formation du squelette polycétidique. L'un de ces deux oxygènes est impliqué dans la cyclisation.

Les cytochromes P450 sont des enzymes catalysant des réactions d'hydroxylation et des gènes codant de cytochromes P450 sont souvent présents dans les clusters de métabolites secondaires. Dans le cluster responsable de la synthèse de sambomycine, deux gènes, *sam*R0478 et *sam*R0479, codent ce type d'enzyme. Afin d'établir leur rôle dans la biosynthèse du polycétide, les deux gènes ont été interrompus indépendamment dans la souche ATCC/OE484. Les extraits de mycélia de ces nouvelles souches ont été analysés par LC-MS.

Les résultats obtenus ont montré que la délétion du gène *sam*R0479 produit un effet drastique sur la biosynthèse du macrolide. Uniquement des formes linéaires du polycétide ont été identifiées mais pas la forme cyclique, indiquant que SAMR0479 serait un enzyme clé dans la cyclisation en catalysant l'hydroxylation du premier précurseur incorporé.

En revanche, l'analyse du mutant délété pour samR0478 montre que la souche produit toujours une forme cyclique du polycétide mais dépourvue d'un groupe hydroxyle par rapport à la sambomycine.

Une autre caractéristique intéressante de la sambomycine est le précurseur incorporé par le domaine AT13. Les analyses RMN ont en effet révélé la nature de ce précurseur, une chaîne alkyle ramifiée (Fig. 5). Généralement, les unités utilisées pour la synthèse de polycétides sont les acétates et les propionates et très rarement d'autres substrats. Dans ce dernier cas, des gènes candidats pour la synthèse du précurseur ont été identifiés dans le cluster. Il s'agit des gènes *sam*R0482 et *sam*R0483. Les produits de ces gènes sont homologues à des enzymes intervenant dans l'activation et la carboxylation des précurseurs du métabolisme des acides gras, mais aussi des précurseurs du métabolisme secondaire. Le professeur Greg Challis a proposé donc que SAMR0482 agisse comme un acyl-CoA synthétase qui activerait une chaîne alkyle ramifié

dérivant du métabolisme des acides gras, et que SAMR0483 soit une carboxylase spécifique pour le précurseur activé, selon le modèle présenté Figure 6.

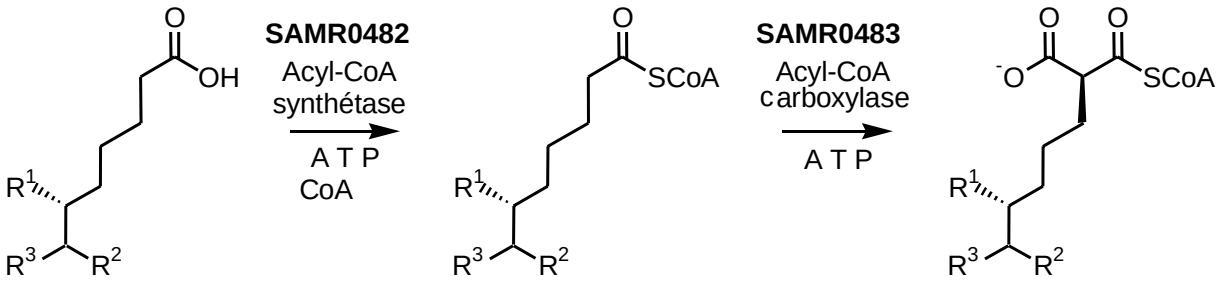


Figure 6. Hypothèse sur le mécanisme qui conduirait à la synthèse du précurseur incorporé par le domaine AT13, comme proposé par le Prof. Greg Challis. SAMR0482 catalyse l'activation du précurseur en utilisant le cofacteur ATP et le coenzyme A; SAMR0483 carboxyle le précurseur activé en utilisant comme cofacteur l'ATP.

Afin d'étudier le rôle de ces deux gènes et de confirmer notre hypothèse, l'interruption de deux gènes a été envisagée dans la souche qui surexprime le régulateur LAL. Cependant, seul le mutant délété pour le gène *sam*R0482 (ATCC/OE484/Δ482 ::spec) a été obtenu et l'analyse par LC-MS de l'extrait du mycélium a montré la présence de deux pics 673 et 680 de la sambomycine. Il semblerait donc que le gène *sam*R0482 ne soit pas essentiel pour la biosynthèse du macrolide. Il est possible qu'un autre gène, codant la même fonction et localisé ailleurs dans le chromosome, intervienne dans la synthèse du précurseur alkylé.

4. Régulation du cluster sambomycine

Rôle du régulateur LAL

Une partie de ma thèse s'est focalisée sur la caractérisation des mécanismes de régulation qui gouvernent la biosynthèse de la sambomycine. En particulier, nous avons essayé d'élucider le rôle du régulateur LAL, codé par le gène *sam*R0484.

Comme décrit précédemment, nous avons montré que SAMR0484 est un activateur du cluster *samb* (nom du cluster responsable de la production de sambomycine). En effet, sa surexpression induit l'expression de gènes de PKS (Fig. 3). Nous avons étudié si la surexpression affectait également les autres gènes du cluster. Les analyses transcriptionnelles de la souche ATCC/OE484 en parallèle avec le contrôle ATCC/pIB139 ont mis en évidence que le régulateur LAL contrôlait également l'expression des gènes codant les cytochromes P450 (*sam*R0478, 0479), de deux gènes impliqués dans la synthèse du sucre (*sam*R0472 et 0473, qui forment très

probablement un opéron) et des gènes codant le transporteur ABC (*sam*R0470, 0471, également organisés en opéron). En fait, la surexpression du LAL semble affecter la transcription de l'ensemble des gènes du cluster *samb*.

Les régulateurs de la famille LAL contiennent un motif hélice-coude-hélice (HTH) dans leur parti C-terminale, indiquant SAMR0484 est très probablement un activateur transcriptionnel. Les cibles de LAL ne sont pas encore connues, mais des expériences indirectes avec des autres membres de cette famille suggèrent que ces cibles seraient les gènes de PKS, ainsi que les gènes du système de résistance et le gène codant la glycosyltransférase.
Afin de démontrer que SAMR0484 se fixe sur les régions promotrices de gènes de PKS et de résistance, nous avons opté pour l'utilisation d'un système rapporteur utilisant un plasmide portant le gène *redD* de *Streptomyces coelicolor* dépourvu de son promoteur (pIJ2587; van Wezel et al., 2000). Le gène *redD* code l'activateur spécifique et essentiel de la biosynthèse de l'antibiotique undécylprodigiosine, une molécule qui présente une pigmentation rouge caractéristique. Les régions intergéniques des gènes *sam*R0467, *sam*R0477 et *sam*R0471 ont été amplifiées par PCR puis clonées dans le plasmide pIJ2587 immédiatement en amont de la l'ORF *redD*. Les plasmides recombinants ont été introduits dans les souches de *S. coelicolor* M512/OE484 et M512/pIB139. La souche M512 (Floriano and Bibb, 1996) est une souche de *S. coelicolor* délétée pour les gènes *act*IIORF4 et *redD*, donc une souche incapable de produire les antibiotiques pigmentés actinorhodine (pigment bleu) et undécylprodigiosine (le gène *act*IIORF4 code un activateur transcriptionnel indispensable pour activer les gènes de biosynthèse de l'actinorhodine). Le choix de la souche M512 pour ces tests s'explique également par le fait qu'aucun gène codant un régulateur LAL n'est présent dans le génome de *S. coelicolor*.
Les clones obtenus ont alors été testés pour leur capacité à produire l'undécylprodigiosine. Les résultats montrent que SAMR0484 active la transcription du gène *redD* quand ce dernier est sous le contrôle des régions promotrices de gènes de PKS *sam*R0467 et *sam*R0477. En effet, les colonies deviennent rouges, contrairement aux colonies obtenues après transformation de la souche contrôle M512/pIB139 qui, elles, restent blanches (Fig. 7). Cependant, pour la région intergénique du gène *sam*R0471 les colonies deviennent rouges que ce soit pour les transformants de la souche M512/OE484 ou de la souche contrôle. Il est possible que *S. coelicolor* contienne un facteur de transcription capable de se fixer sur cette séquence. Nous ne pouvons cependant pas exclure que SAMR0484 reconnaisse tout de même cette région.

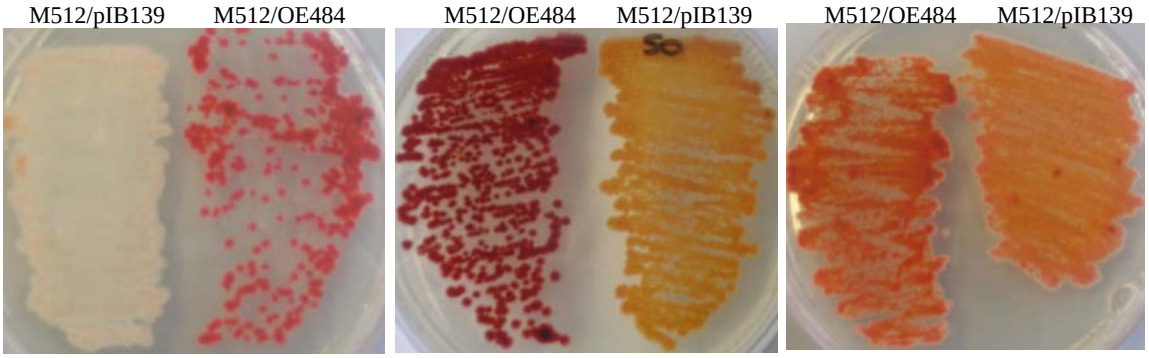


Figure 7. Colonies produisant la pigmentation rouge typique de l'antibiotique undécylprodigiosine obtenues après transformation des souches M512/pIB139 et M512/OE484 contenant les vecteurs recombinants pIJ-467, pIJ-477 et pIJ-4701. Les transformants ont été étalés sur milieu R2YE+25µg/ml thiostrepton. La couleur des colonies peut quelque fois apparaître plus foncé car les boites ont été incubées pour une période plus longue à 30°C.

Afin de confirmer que les régions promotrices des gènes *sam*R0467 et *sam*R0477 sont les cibles directes de SAMR0484, des expériences de retard sur gel ont été réalisées avec des extraits bruts d'*Escherichia coli* exprimant soit la protéine SAMR0484 entière soit une forme tronquée contenant uniquement le motif HTH. Pour cela, le gène *sam*R0484 ou la forme tronquée du gène ont été clonés dans le vecteur d'expression pET15a (Novagen) puis les constructions ont été introduites dans la souche *E. coli* BL21(DE3), une souche dédiée pour l'expression de protéines hétérologues (Novagen).
Les résultats préliminaires de retard sur gel ont montré que la forme tronquée de SAMR0484 se fixerait à la séquence intergénique des gènes *sam*R0467 et *sam*R0477, ce qui est en accord avec les résultats obtenus avec le système rapporteur.

Effet du régulateur LAL sur la production du macrolide spiramycine

Lors de l'étude des mutants surexprimant la protéine LAL SALR0484, des tests de bioactivité ont été réalisés à partir de cultures en milieu liquide ou solide (milieux MP5 et/ou R2) contre *Micrococcus luteus* et *Bacillus subtilis* utilisées comme souches indicatrices (ces bactéries sont sensibles aux macrolides). La souche ATCC/OE484 montrait une activité réduite contre *M. luteus*, par rapport à la souche de référence ATCC/pIB139. Ce phénotype a été également observé avec le mutant SPM704 surexprimant le gène *sam*R0484. SPM704 est une souche non-productrice de congocidine et de kinamycine mais synthétisant la spiramycine (construite en collaboration avec J-L Pernodet, IGM, Paris XI, Orsay). L'activité observée contre *M. luteus* ne correspondait pas à la sambomycine, cette dernière n'étant pas secrétée dans le milieu (au moins dans les conditions utilisées). Le phénotype semblait corrélé à un effet négatif de la surexpression de *sam*R0484 sur la production de spiramycine.

En effet, des analyses par HPLC nous ont permis de confirmer la diminution (voire la perte) de production de spiramycine chez les clones issus de la souche sauvage ou du mutant SPM704 et contenant le gène *sam*R0484 sous la dépendance du promoteur *ermE*p*.

Cet effet sur la production de spiramycine pouvait être du à une action du régulateur LAL sur l'expression des gènes de biosynthèse de la spiramycine ou la production du nouveau macrolide pouvait se faire au détriment de celle de la spiramycine par un shunt des précurseurs. Le mutant ATCC/OE484/Δ467::kan dans lequel le 1er gène de PKS (*sam*R0467) est interrompu a été analyse pour la production de spiramycine. Cette dernière ne retournait pas à la valeur observée chez la souche sauvage sauf après plusieurs jours d'incubation. Ce résultat indiquerait donc que l'effet observé avec la surexpression du gène LAL serait due à une régulation opérée directement ou indirectement par le régulateur SAMR0484 sur les gènes du cluster responsables de la synthèse de la spiramycine.

CONCLUSIONS

Les résultats obtenus au cours de cette thèse contribuent à démontrer que le potentiel de *Streptomyces* et en particulier de *Streptomyces ambofaciens* dans la production de métabolites secondaires reste encore à exploiter.
Nous avons découvert un nouveau macrolide présentant un noyau macrolactone à 50 atomes de C, nommé sambomycine, produit par *S. ambofaciens* ATCC23877. Il s'agit du plus gros noyau macrolactone décrit jusque présent. En fait, quatre formes différentes de la sambomycine ont été identifiées (les isomères A, B et les isomères C, D).
La détermination de la structure de la sambomycine a dévoilé des caractéristiques de synthèse très intéressantes, telle que la cyclisation faisant intervenir une étape d'hydroxylation du squelette polycétidique par un enzyme cytochrome P450 ou encore la présence d'une unité carboxylique atypique (une chaîne alkyle ramifiée). Ces caractéristiques pourraient être exploitées dans le futur pour synthétiser de nouveaux composés dans le cadre de la biologie combinatoire.
Nous avons montré que la sambomycine possède des activités antibactériennes contre des bactéries Gram-positives et des activités antiprolifératives très prometteuses.

La découverte de la sambomycine a été possible grâce à une combinaison de différentes approches à savoir l'exploitation du génome qui a permis d'identifier le cluster *samb,* l'analyse *in silico* des produits des gènes du cluster qui a permis de prédire le squelette polycétidique du composé recherché et ses caractéristiques (masse moléculaire, hydrophilicité…) et enfin la

comparaison de profils chromatographiques des métabolites produits. Mais l'originalité de notre travail porte essentiellement sur le réveil d'un cluster silencieux dans les conditions standard de laboratoire grâce à la surexpression d'un régulateur spécifique appartenant à la famille LAL. Une approche similaire avait été décrite uniquement chez le champignon *Aspergillus nidulans* et avait permis d'isoler l'aspyridone, un composé produit par un cluster PKS/NRPS (Bergmann *et al.*, 2007). Par conséquence, nous avons démontré que cette stratégie est applicable aux *Streptomyces*. Ceci est particulièrement intéressant puisque le séquençage du génome de plusieurs espèces de *Streptomyces* a mis en évidence la présence de très nombreux clusters cryptiques de métabolites secondaires chez ces organismes, clusters qui représentent une source potentielle importante de nouveaux composés pouvant avoir des applications intéressantes en médecine mais également en agriculture.

Notre étude a permis de montrer que le régulateur LAL SAMR0484 est un activateur transcriptionnel essentiel pour la biosynthèse de la sambomycine. La surexpression du gène s'accompagne d'un effet positif sur l'ensemble de gènes du cluster. Nos résultats montrent que SAMR0484 agirait directement sur les régions promotrices des gènes de PKS. Aucune cible directe des régulateurs LAL n'avait été identifiée jusqu'à présent.
La compréhension de la régulation de la production de la sambomycine est essentielle pour optimiser les conditions de production du composé chez *S. ambofaciens*. Une stratégie d'expression hétérologue du cluster dans une souche hôte dédiée à la production de métabolites secondaires n'est guère envisageable en tant donné la taille très importante du cluster (> 150 kb).

Les expériences de surexpression de *samR0484* ont montré que ce régulateur a également un effet pléiotrope. En effet, sa surexpression s'accompagne d'une diminution/perte de la production du macrolide spiramycine. L'origine de ce phénotype reste pour le moment encore inconnue. Toutefois, il est possible que le phénotype observé soit dû à une régulation croisée directe de SAMR0484 sur l'expression des gènes impliqués dans la synthèse de la spiramycine. Une analyse préliminaire de l'expression de gènes *srmGI* et *orf3*c* de la spiramycine dans la souche qui surexprime le gène *samR0484* a montré une diminution au niveau de transcrit dans la phase stationnaire. Si le régulateur LAL a une action directe sur l'expression de ces gènes, il est intéressant de noter que SAMR0484 jouerait un rôle positif dans le cas de la sambomycine et négatif pour la spiramycine. Cela pose la question de compétition entre les deux clusters et de l'intérêt de chacun d'eux pour *S. ambofaciens* dans son habitat naturel.

Europäisches Patentamt
European Patent Office
Office européen des brevets

European Patent Office
80298 MUNICH
GERMANY
Tel. +49 (0)89 2399 - 0
Fax +49 (0)89 2399 - 4465

For any questions about this communication:
Tel.:+31 (0)70 340 45 00

Laureti, Luisa
9 Rue Gilbert
54000 Nancy
FRANCE

Date
23.09.09

Reference	Application No./Patent No.
	09290567.6 - 2101

Applicant/Proprietor
Université Henri Poincaré - Nancy 1, et al

Designation as inventor - communication under Rule 19(3) EPC

You have been designated as inventor in the above-mentioned European patent application. Below you will find the data contained in the designation of inventor and further data mentioned in Rule 143(1) EPC:

DATE OF FILING : 24.07.09

PRIORITY : / /

TITLE : Sambomycin and derivatives, their production and their use as drugs

DESIGNATED STATES : AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO SE SI SK SM TR

INVENTOR (PUBLISHED = 1, NOT PUBLISHED = 0):

1/Aigle, Bertrand/47 Rue Paul Bert/54520 Laxou/FR
1/Challis, Gregory/13 Abbotts Green Burbage/Hinckley, Leics LE10 2QZ/GB
1/Laureti, Luisa/9 Rue Gilbert/54000 Nancy/FR
1/Song, Lijiang/46 Lyndale Road/Conventry CV5 8AQ/GB
1/Leblond, Pierre/62, Rue de Mirecourt/54630 Flavigny-Sur-Moselle/FR

DECLARATION UNDER ARTICLE 81 EPC:
The applicant(s) has (have) acquired the right to the European patent as employer(s).

Receiving Section

Europäisches Patentamt • European Patent Office • Office européen des brevets

EPO Form 1048 12.07

Résumé

La recherche de nouveaux métabolites d'intérêt médical est toujours d'actualité, surtout si l'on considère que d'anciennes, mais aussi des nouvelles infections bactériennes ou virales apparaissent régulièrement. Les actinomycètes, et plus particulièrement les *Streptomyces*, sont les principaux producteurs de molécules anti-microbiennes. En effet, ils produisent presque 60-70% des produits naturels d'origine microbienne. La plupart de ces métabolites appartient à la classe des polycétides, qui sont synthétisés par des complexes multienzymatiques, les polycétide-synthétases (PKS). Les PKS utilisent des acides gras simples pour assembler des structures polycétidiques très diversifiées. Une approche très prometteuse pour identifier des nouvelles voies de biosynthèse de métabolites secondaires est basée sur une approche génomique ou « genome mining ». Le séquençage du chromosome linéaire de *Streptomyces ambofaciens* ATCC23877 a, entre autre, révélé sur le bras chromosomique droit, un cluster cryptique de gènes de PKS de type I de grande taille.

Ce cluster contient 25 gènes, dont 9 gènes de biosynthèse, pour un total de 25 modules, tous fonctionnels. Les analyses *in silico* des gènes de PKS ont permis de prédire que le cluster serait responsable de la synthèse d'une molécule appartenant à la famille des macrolides. Dans des conditions standard de laboratoire, le cluster était silencieux. Afin d'activer l'expression du cluster, un gène de régulation, *sam*R0484, présent dans le cluster et codant un régulateur de la famille LAL (Large ATP binding protein of the LuxR family), a été surexprimé dans la souche sauvage. Les analyses transcriptionelles ont montré que cela se traduisait par l'induction de l'expression des gènes de biosynthèse. Par conséquence, une stratégie de « comparative metabolic profiling » a été menée entre la souche sauvage et la souche mutante afin d'identifier le nouveau métabolite. Quatre formes différentes d'un macrolide avec un cycle lactonique de 50 carbones, ont été isolées et caractérisées. Ces composants, nommés sambomycine A, B, C et D, ont montré une activité antibactérienne et une activité antiproliférative intéressante.

La détermination de la structure de la sambomycine a révélé des caractéristiques uniques et intéressantes, concernant la réaction de cyclisation et la synthèse d'un précurseur atypique. Ces mécanismes de biosynthèse ont fait l'objet d'une étude plus approfondie.

Nous nous sommes également intéressés à la régulation de ce cluster. Le régulateur LAL agit comme un activateur transcriptionnel essentiel. Des analyses préliminaires indiquent que ce régulateur se fixe aux régions promotrices de certains gènes, notamment celles des gènes de biosynthèse ainsi que celles d'autres gènes de post-modification, activant ainsi leur transcription.

Mots Clés: *Streptomyces ambofaciens*, cluster silencieux, activation, régulateur spécifique, macrolide

Summary

The constant and urgent need of novel bioactive compounds is the result of the emergence in the last decades of new and old infectious diseases, a sore for humankind. Actinomycetes and especially the genus *Streptomyces* are the principal producers of microbial drugs producing nearly 60-70% of the natural products. The majority of secondary metabolites belong to the class of polyketides that are synthesised by multienzymatic complexes named polyketides synthases (PKS). PKSs condensate simple small carboxylic acids to generate a wide range of complex polyketide structures. In the search for new drugs, the genome mining approach proved to be a powerful tool in the identification of cryptic secondary metabolite pathways.

The sequencing and the analysis of *Streptomyces ambofaciens* ATCC23877 genome has revealed a large type I PKS cluster, on the right arm of the chromosome. The cluster contains 9 large PKS genes, composed of 25 functional modules. *In silico* analysis of the PKS and of the tailoring genes enabled to predict the structure of the polyketide backbone. In the laboratory standard conditions, the cluster showed to be silent. Therefore, to promote the expression of the cluster, the regulatory gene *sam*R0484, encoding a LAL regulator (Large ATP binding protein of the LuxR family) was overexpressed in the wt strain. Transcriptional analyses showed that the PKS genes were expressed. Subsequently, by comparative metabolic profiling between the mutant strain and the wt, we were able to detect the novel metabolite produced by *S. ambofaciens*. Structural elucidation revealed four forms of a 50-membered macrolide, named sambomycin. The compounds endow antibacterial and antitumoral activities.

The structure of sambomycin unveiled unique and interesting characteristics, i.e. the cyclization reaction and the presence of an atypical extender unit. The mechanisms of biosynthesis have been analysed more in details in this work.

We also investigated in the regulation of the cluster. The LAL regulator was shown to be an essential transcriptional activator, binding to the promoter regions of the PKS genes and probably to other genes in the cluster.

Key Words: *Streptomyces ambofaciens*, silent cluster, activation, pathway-specific regulator, macrolide