Hanne Jørgensen

Analysis of genes for the biosynthesis of polyene macrocyclic compounds in streptomycetes isolated from the Trondheimsfjord

Thesis for the degree of Philosophiae Doctor

Trondheim, May 2009

Norwegian University of Science and Technology
Faculty of Natural Sciences and Technology
Department of Biotechnology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Natural Sciences and Technology
Department of Biotechnology

ISBN 978-82-471-1550-3 (printed ver.)
ISBN 978-82-471-1551-0 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2009:86

Printed by NTNU-trykk

Acknowledgements

The work presented in this thesis has been carried out at the Department of Biotechnology at the Norwegian University of Science and Technology (NTNU) and has been supported by funding from the Research Council of Norway, NTNU and SINTEF.

The work has been supervised by Professor Sergey B. Zotchev and I would like to thank him for all help and support, his skillful supervision of the work and for sharing his extensive knowledge of *Streptomyces* genetics with me.

In addition, I wish to thank the following present and former colleagues at NTNU and SINTEF: Harald Bredholt, Per Bruheim, Kristin F. Degnes, Alexander Dikiy, Espen Fjærvik, Sigrid Hakvåg, Geir Klinkenberg and Håvard Sletta for performing parts of the analytical work. Harald Bredholt and Sigrid Hakvåg established the strain libraries from sediment and neuston layer samples, Harald Bredholt, Espen Fjærvik, Geir Klinkenberg and Per Bruheim participated in the screening of the strain libraries and the subsequent characterization of leads, Kristin F. Degnes and Håvard Sletta performed the fermentations, LC-MS-TOF analysis of mutants and the feeding studies, while Alexander Dikiy clarified the structures of BE-14106 and ML-449 by NMR spectroscopy.

I also wish to thank our collaborators at the Department of Biomedicine at the University in Bergen, Lars Herfindal, Gyrid Nygård and Stein Ove Døskeland, for performing the cytotoxic screen against cancer cell lines.

All colleagues at the Department of Biotechnology, SINTEF and Biosergen receive my thanks for creating a nice working atmosphere and for many helpful advices, particularly my good friends Kerstin, Torunn, Mali, Marianne and Rahmi. You always manage to brighten up my day. I thank my parents and sister for their love and support

and Jon Erik for always being there for me and for solving all and any computer related issues.

Trondheim, March 2009
Hanne Jørgensen

Abstract

The isolation of microbial producers of bioactive natural products from environmental samples has historically been a great success and laid the foundation for the modern medical science we enjoy the fruits of today. To sustain and further improve the treatment of diseases, we rely upon the continued discovery and development of new bioactive compounds.

The work presented in this thesis describes the screening of more than 4000 actinomycete isolates recovered from sediment and neuston layer samples collected in the Trondheimsfjord. The objective was to discover producers of new compounds with antifungal or cytotoxic activity. The primary screening approach based on assays with *Candida* strains uncovered a large number of isolates producing bioactive compounds, however, spectroscopic analyses of extracts from these isolates revealed that a high percentage of the isolates were potentially producing the same compound. LC-MS-TOF analysis of the extracts identified the compound in question as the polyene macrolide candicidin. A genetic analysis of eight isolates showed that they all contained the candicidin biosynthetic gene cluster and that the cluster was present on a large plasmid in one of the isolates. The plasmid's ability for transfer to other *Streptomyces* strains was investigated, but interspecific transfer could not be detected. A "cured" strain unable to produce candicidin was obtained by incubation of the plasmid-containing isolate at a high temperature and reintroduction of the plasmid restored the candicidin production, thus indicating that the plasmid is transmissible by conjugation. It is possible that the plasmid may have been responsible for the dissemination of the candicidin biosynthetic gene cluster among actinomycetes in sediments and neuston layer of the Trondheimsfjord, although the results from this study were not conclusive.

Candidates selected based on the primary screening against *Candida* strains were further evaluated in assays with different cancer cell lines. A compound displaying good cytotoxic activity was identified as the previously described macrolactam antibiotic BE-14106 by LC-MS-TOF analysis and NMR spectroscopy. A genomic library was

constructed for the BE-14106 producer and screened with a molecular probe targeting polyketide synthase genes. The biosynthetic gene cluster was successfully identified and sequenced and the biosynthetic pathway leading to production of BE-14106 was elucidated. The proposal for the biosynthetic pathway is supported by results from gene inactivation experiments, enzyme assays with heterologously expressed proteins and feeding studies with isotope labeled components.

A second macrolactam, ML-449, was identified in the primary screening against the *Candida* strains. LC-MS-TOF analysis indicated structural resemblance to BE-14106. The complete structure of ML-449 was obtained by NMR spectroscopy, showing that BE-14106 and ML-449 only differ in the length of the acyl side chain. The ML-449 biosynthetic gene cluster was subsequently cloned, sequenced and compared to the BE-14106 biosynthetic gene cluster. The two clusters were found to be remarkably similar, differing only in the genes encoding the polyketide synthases synthesizing the acyl side chain. Phylogenetic analyses pointed to common ancestry for the two clusters as well as an evolutionary relationship with other macrolactam biosynthetic gene clusters.

Abbreviations

A	Adenylation
aa	Amino acid(s)
ABC	ATP-binding cassette
ACP	Acyl carrier protein
AMP	Adenosine monophosphate
Ala	Alanine
aLRT	Approximate likelihood ratio test
Asn	Asparagine
AT	Acyltransferase
BLAST	Basic local alignment search tool
bp	Base pair(s)
BSA	Bovine serum albumin
C	Condensation
CFU	Colony-forming unit
CIAP	Calf intestinal alkaline phosphatase
CoA	Coenzyme A
Da	Dalton
DAD	Diode array detector
DH	Dehydratase
DIG	Digoxygenin
DMSO	Dimethyl sulfoxide
DNP	Dictionary of natural products
ER	Enoyl reductase
ESI	Electrospray ionization
Gln	Glutamine
Gly	Glycine
HGT	Horizontal gene transfer
His	Histidine
HMBC	Heteronuclear multiple bond correlation
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum coherence
HTH	Helix-turn-helix
IC-50	Half maximal inhibitory concentration
JTT	Jones-Taylor-Thornton
kb	Kilo base pairs
kD	Kilo dalton
KR	Ketoreductase
KS	Ketosynthase
LC-50	Lethal concentration 50
LC-MS	Liquid chromatography-mass spectrometry
Lys	Lysine
Mb	Mega base pairs
MFS	Major facilitator superfamiliy

MIC50	Minimum inhibitory concentration (50 % inhibition of growth of test organism)
ML	Maximum likelihood
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase
nt	Nucleotide(s)
NTP	Nucleoside triphosphate
orf	Open reading frame
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PK	Polyketide
PKS	Polyketide synthase
PFGE	Pulsed field gel electrophoresis
ppm	Parts per million
Pro	Proline
RDP	Ribosomal Database Project
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TE	Thioesterase
TOF	Time-of-flight
UV/VIS	Ultraviolet-visible
WAG	Whelan and Goldman

Table of contents

1 Introduction

One cannot overestimate the importance of natural products in modern medicine. Without them, all of us would be at risk of succumbing to infectious diseases, organ transplantations would not be feasible and cancer chemotherapy would be lacking some of its most potent and important agents. The number of bioactive natural products is difficult to estimate, but it is thought to exceed at least 200.000, including contributions from prokaryotes and the animal, fungi and plant kingdoms (Bérdy, 2005).

1.1 A brief history of actinomycetes and antibiotic discovery

The first actinomycete was discovered in 1874 when a Norwegian doctor described the causal agent of leprosy, *Mycobacterium leprae*. The name *Actinomyces* stems from 1877, when a German botanist described the microorganism responsible for lumpy jaw in cattle as a fungus and gave it the name *Actinomyces bovis* (Hopwood, 1999). In the following decades, many more actinomycetes were discovered, but the true nature of these microorganisms remained a mystery for some time. Often they were thought to be fungi or an intermediate of bacteria and fungi, and this is not surprising as many actinomycetes produce mycelium and spores superficially similar to that of moulds. The phylogenetic confusion is well reflected by the names given to the microorganisms (*Actinomyces* = ray fungus, *Mycobacterium* = fungus bacterium, *Streptomyces* = twisted fungus) (Hopwood, 2007). The issue was finally settled in the 1970s after Carl Woese developed the method of ribosomal RNA sequencing to establish evolutionary relationships (Woese and Fox, 1977). The actinomycetes were found to be true bacteria and the resemblance to fungi seems to be a matter of convergent evolution perhaps due to similar habitats.

At the beginning of the 20th century, research on actinomycetes was scarce, with only a few scientists working with this group of peculiar microorganisms. This was soon to change. One scientist interested in the actinomycetes, although initially not as a source

for antimicrobial agents, was Selman A. Waksman (Hopwood, 2007). The discovery of penicillin in 1928 by Alexander Fleming (Fleming, 1929) and René Dubos' isolation of tyrocidin/gramicidin in 1939 (Hotchkiss and Dubos, 1940), was probably Waksman's inspiration for initiating a search for antibiotic producers among soil microbes (Hopwood, 2007). His former research interest, the actinomycetes, soon proved superior to the other bacteria and fungi tested with respect to production of antimicrobial agents. The breakthrough for Waksman's group came in 1943 with the discovery of streptomycin production in a *Streptomyces griseus* strain (Schatz and Waksman, 1944). Streptomycin turned out to be the first effective treatment for tuberculosis and Waksman was awarded the Nobel Prize for Physiology or Medicine in 1952. With the pioneering work of Fleming, Dubos and Waksman, the interest in antibiotic discovery was sparked. During the 1950s and 60s, screening programs were initiated by all major pharmaceutical companies as well as many academic groups around the world (Hopwood, 2007). During these two decades, hundreds of new antibiotics were discovered every year, with streptomycetes being the major contributors. This period is often called the golden age of antibiotic discovery. By the end of the 1970s, however, the number of new antibiotics discovered each year was already declining, and during the 1980s and 1990s, most of the screening programs were discontinued, partly because of diminishing success in discovering new types of antibiotics, but also due to the increasing R&D (research and development) expenditure. Instead, the pharmaceutical companies turned their attention to the more profitable drug market of lifestyle problems and chronic diseases. Such agents often require daily use for the rest of a patient's life and the short one-time cures of antibiotics cannot compete with respect to justifying the R&D expenditure for new drug candidates (Finch and Hunter, 2006; Overbye and Barrett, 2005). The use of antibiotics has not declined, however, and has even found a new niche of use as a feed-additive in farming, selecting for an ever increasing amount of resistant strains. By now, the problems with antibiotic resistant pathogens have become quite evident (Boucher et al., 2009; Finch and Hunter, 2006) and contrary to earlier beliefs, the fight against infectious diseases is far from over.

1.2 Natural products

Since the discovery of penicillin, the value of microorganisms as producers of bioactive natural products has truly been revealed. Natural products find their use in a wide range of applications, such as antibacterial and antifungal antibiotics, anticancer agents, immunosuppressants, hypolipidemic agents and anthelminthics (Table 1. 1).

Table 1. 1 Examples of bioactive natural products. Adapted from Singh and Pelaez (2008), Gredičak and Jerić (2007), Amyes (2001).

Application	Example	Producing organism
Antibacterial	Penicillins	*Penicillium* spp., *Aspergillus* spp.
	Cephalosporins	*Acremonium* spp., *Amycolatopsis* spp. *Streptomyces* spp.
	Tetracyclines	*Streptomyces* spp., *Actinomadura* spp.
	Aminoglycosides	*Streptomyces* spp., *Micromonospora* spp.
	Macrolides	*Saccharopolyspora* spp.
	Vancomycin	*Streptomyces* spp.
Antifungal	Polyene macrolides	*Streptomyces* spp.
	Echinocandins	*Glarea* spp.
Anticancer	Bleomycins	*Streptomyces* spp.
	Anthracyclines	*Streptomyces* spp.
	Enediynes	*Streptomyces* spp., *Micromonospora* spp., *Actinomadura* spp.
Immunosuppressant	Cyclosporin A	*Tolypocladium* spp.
	Rapamycin	*Streptomyces* spp.
Hypolipidemic	Lovastatin	*Aspergillus* spp.
	Mevastatin	*Penicillium* spp.
Anthelminthic	Avermectin	*Streptomyces* spp.

Among the most prolific producers of bioactive compounds are the actinomycetes. The total number of bioactive metabolites isolated from actinomycetes is estimated to 10100, with antibiotics as the major group of 8700 compounds. One genus, *Streptomyces*, accounts for over 75 % of both numbers (Bérdy, 2005). With the total number of antibiotics discovered estimated to around 25.000-30.000 (Bérdy, 2005), this makes *Streptomyces* the biggest single contributing genus (20-25 %). When considering the antibiotics in practical use, the impact of the streptomycetes becomes even clearer: of a total of 140-160 antibiotics in use, approximately 100-120 is of actinomycete (mostly *Streptomyces*) origin (Bérdy, 2005). The reason for this may be that pharmaceutical companies have targeted actinomycetes, and in particular *Streptomyces*, in their screening programs. On the other hand, it seems clear that the streptomycetes constitute one of the most versatile genera of all the bacteria, biochemically speaking. Genome sequencing has been completed for three streptomycetes, *S. avermitilis*, *S. coelicolor*, and *S. griseus* (Õmura et al., 2001; Bentley et al., 2002; Ohnishi et al., 2008) and has revealed that these streptomycetes have a much greater potential for secondary metabolite production than first assumed as they contain many more secondary metabolite biosynthetic gene clusters than the number of actually identified metabolites would suggest.

1.2.1 Antibacterials

The kind of bioactive natural product that people most commonly encounter is the antibacterial antibiotics, which include the huge group of β-lactams. Since the discovery in 1928, penicillins have been semi-synthetically modified into many different derivatives. In contrast to the acid-unstable original penicillin, which Fleming struggled to isolate from the producing organism and with treatments consisting of intravenous injections several times a day (Amyes, 2001), penicillins are now prescribed in tablet form and able to pass through our digestive system. This revolution has come about through semi-synthetic modifications of the penicillin side chain (Figure 1. 1). Another feat accomplished in this way by natural drug scientists is an increased spectrum of

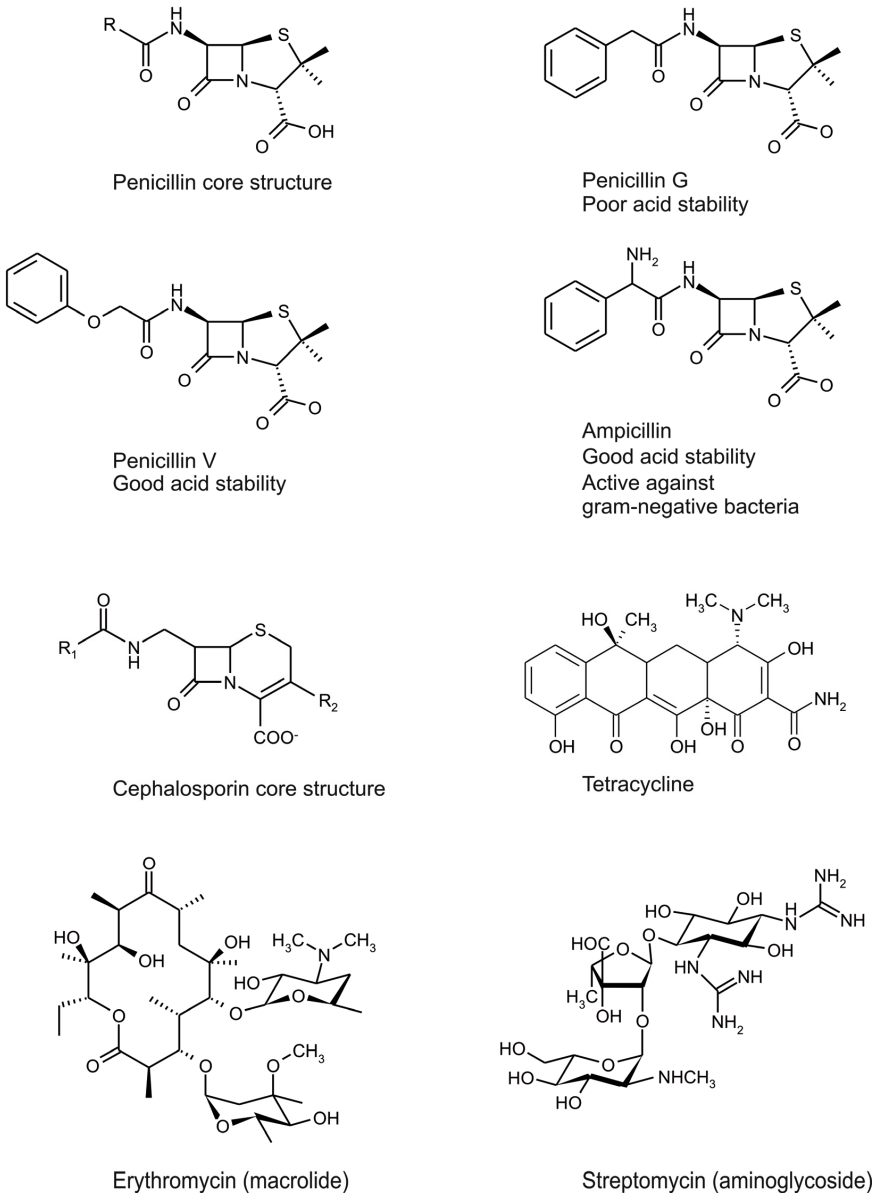


Figure 1. 1 Chemical structures of important antibacterials and modifications of the penicillin side chain. The penicillin core structure is shown on top. Penicillin G with a phenylacetyl chain is acid-unstable, while penicillin V with a phenoxyacetyl chain has better acid stability. Ampicillin with a phenylglycyl chain is acid stable and active against gram-negatives. Adapted from Walsh (2003).

activity for penicillins. Naturally occurring penicillins have little or no effect on gram-negative bacteria as the molecules often cannot pass through their outer cell membrane, but by introducing new side chains, the permeability can be increased (as for e.g. ampicillin, amoxicillin) (Walsh, 2003). The β-lactams also include the cephalosporins, another hugely successful group of antibacterials, which has been subjected to semi-synthetic modifications analogous to the penicillins, as well as the carbapenems, monobactams and clavams (Walsh, 2003). Although β-lactams is the biggest and most widely used group of antibacterials, there are also several other important groups. The tetracyclines, aminoglycosides, macrolides, streptogramins and lincomycins all target the bacterial protein synthesis, while rifampicin inhibits RNA synthesis. The glycopeptide vancomycin interferes with bacterial cell wall synthesis, but in a different way than β-lactams and vancomycin is therefore vastly important in the treatment of infections caused by certain resistant gram-positive bacteria (Amyes, 2001). Synthetic antibacterials include the sulphonamides, which target folic acid synthesis, and the quinolones, which affect DNA synthesis. Actinomycetes represent important producers of antibacterial antibiotics and e.g. cephalosporins, tetracyclines, aminoglycosides, macrolides and vancomycin are among the bioactive metabolites produced by this group (Table 1. 1).

Almost all of the antibacterial antibiotics were discovered during the 1950s, 60s and 70s and new additions have mostly come about by semi-synthetic modifications of the original molecules. The only truly new antibacterial antibiotics launched in the last 35 years are linezolid and daptomycin (Overbye and Barrett, 2005). The uprise and dissemination of antibiotic resistant pathogens such as methicillin/vancomycin resistant *Staphylococcus aureus* (MRSA/VRSA) and multiresistant *Mycobacterium tuberculosis* in later years, makes this fact particularly worrying and necessitates a continued search for new natural products possessing antibacterial activity (Jones, 2008; Zumla and Grange, 2001).

1.2.2 Antifungals

While the antibacterials are (for the most part) benign molecules with mild side-effects, the antifungals are another story. The eukaryotic fungal cells resemble our own cells so much that finding a selective target has proved difficult. Amphotericin B, the "golden standard" of antifungal treatment, has such adverse side-effects as nephrotoxicity and thrombophlebitis (Chen and Sorell, 2007). The seriousness of a fungal infection depends on its nature, as infections can be superficial or invasive. Superficial infections, which involve the skin and mucous membranes, are generally easier to cure than the invasive type. Invasive infections are potentially life-threatening, and although they are rare in healthy individuals, different types of immunocompromised patients (HIV infected, undergoing organ transplantations or chemotherapy) are at great risk. As the number of such patients steadily increase (Gupte et al., 2002), the need for better antifungal antibiotics is becoming ever more urgent. There are four main classes of antifungal antibiotics currently in use, polyene macrolides, azoles, allylamines and echinocandins (Chen and Sorell, 2007). The polyene macrolides and the echinocandins represent natural products synthesized by actinomycetes and fungi (Figure 1. 2), while azoles and allylamines are synthesized chemically.

The prospects for antifungal treatment look better now than in a long time, with new lipid formulations of amphotericin B and the safer, less toxic echinocandins (Chen and Sorell, 2007). However, the agents used in antifungal therapy are rather few and as with the antibacterials, the development of antibiotic resistant strains always represents a serious threat.

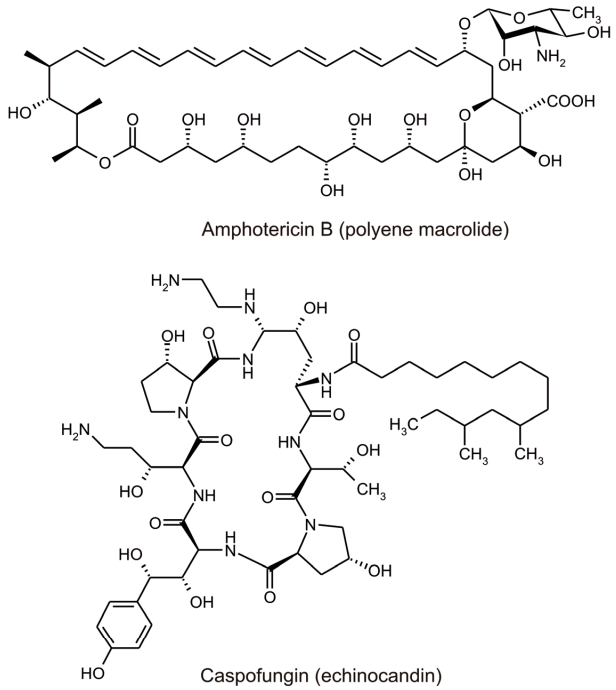
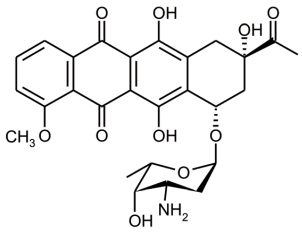


Figure 1. 2 Chemical structures of antifungal antibiotics. Adapted from Caffrey et al. (2008), Letscher-Bru and Herbrecht (2003).

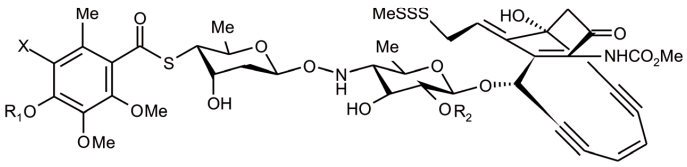
1.2.3 Anticancer agents

Many of the most important anticancer agents are natural products of microbial origin and chemical analogs of the molecules we usually think of as antibiotics. The very first antibiotic identified by Selman Waksman's group (Waksman and Woodruff, 1940) had in fact anticancer activity, although this was not realized at the time of discovery. It was given the name actinomycin, but to Waksman's despair it proved to be too toxic for antibacterial treatment and was not pursued further (Hopwood, 2007). Decades later, it found its right place among the agents approved for anticancer therapy. Actinomycin remains today an important drug in pediatric oncology (Willems et al., 2006). Among the natural products of anticancer treatment, the bleomycin glycopeptides are perhaps the best known (Figure 1. 3). They find their application in the treatment of several

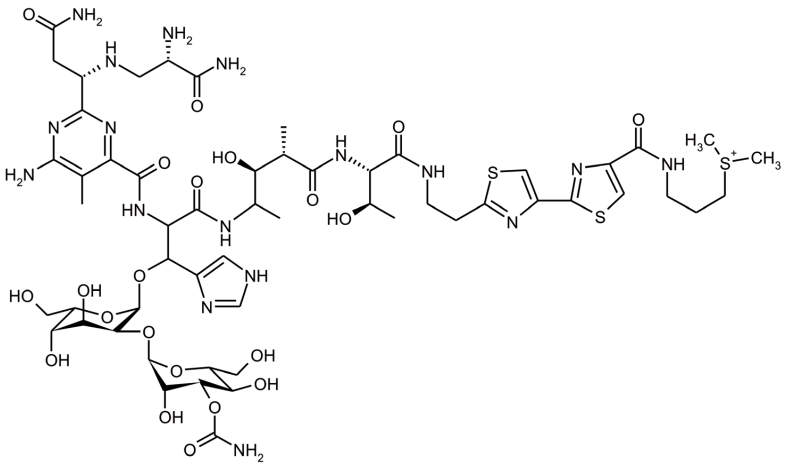
types of cancer, e.g. Hodgkin's disease and testicular cancer (Lazo and Chabner, 2006). Another clinically important group are the anthracyclines (Figure 1. 3), mainly daunorubicin (daunomycin) and doxorubicin, which are among the most widely used



Daunorubicin (anthracycline)



Calicheamicin (enediyne)



Bleomycin A2

Figure 1. 3 Chemical structures of anticancer agents. Adapted from Fujii and Ebizuka (1997), Gredičak and Jerić (2007), Chen and Stubbe (2004).

antineoplastic agents today (Doroshow, 2006). Actinomycin, bleomycins and anthracyclines are all natural products of actinomycetes and so are the new exciting molecules of the enediyne group of anticancer agents. The characteristic structural component of the enediyne compounds is the enediyne moiety (Z-hexa-1,5-diyn-3-ene), which is often described as a "warhead" (Gredičak and Jerić, 2007). An example of the enediyne structure is shown in Figure 1. 3 for calicheamicin. The enediyne moiety and its interactions with DNA make enediynes extremely potent molecules and therefore generally quite toxic (Gredičak and Jerić, 2007). Coupling of the enediyne to a monoclonal antibody has proven to be a successful approach for calicheamicin by keeping the molecule inactive until it is delivered at the right place (the cancer cells). Such coupling of active molecules to cancer cell-specific antibodies may represent the future of anticancer therapy, although the cost of making monoclonal antibodies complicates the issue (Harris, 2004).

1.2.4 Other bioactive natural products

There are important natural products that find their use in areas other than anti-infective and cancer treatment, but it is beyond the scope of this thesis to give a thorough description of those. Among the most important metabolites are the immunosuppressants, such as cyclosporin A and rapamycin, which make organ transplantations possible, and the hypolipidemic agents, which are important cardiovascular agents.

1.3 Discovering new bioactive natural products

For decades, natural products were the most important source for new drugs and drug leads, but from the early 1990s things began to change (Harvey, 2008; Rouhi, 2003). With new technical advancements such as automation and use of robotics, the rate of screening was no longer the limiting step, but rather finding new chemical entities to be tested. To meet this new demand, and to avoid the problems of intellectual property

issues associated with natural products, many pharmaceutical companies switched to combinatorial chemistry and in the process often terminating all efforts in natural product screening (Rouhi, 2003; Ortholand and Ganesan, 2004). In retrospect, this turned out to be a rather unfortunate decision as combinatorial chemistry has not been able to live up to the expectations. For the period 1981-2006, combinatorial chemistry produced only one *de novo* chemical entity that has been approved by the U.S. Food and Drug Administration (Newman and Cragg, 2007). In general, combinatorial synthesis seem to produce less variety and cannot compete with respect to structural complexity, e.g. regarding the number of chiral centers (Feher and Schmidt, 2003; Grabowski and Schneider, 2007). In later years, there has therefore been a renewed interest in natural product screening, as well as exploiting the complexity of natural products as scaffolds in combinatorial chemistry (Ortholand and Ganesan, 2004; Harvey, 2008).

1.3.1 Prospects for discovering new bioactive natural products

Culture-independent methods have demonstrated that a large number of bacterial species are present in soil and marine environments and that successful cultivation has only been achieved for a low percentage of the species present. (Amann et al., 1995; Torsvik et al., 1996; Venter et al., 2004; Yooseph et al., 2007). These bacteria represent a huge resource with respect to production of natural products if their biosynthetic capacities can be exploited.

As actinomycetes are known to be prolific secondary metabolite producers, natural product screening has often targeted this particular order. With an estimated 8700 antibiotics of actinomycete origin (Bérdy, 2005), one might wonder if this group, and particularly the streptomycetes, are exhausted as a source for new drug candidates? In an effort to calculate the antibiotic producing potential of the streptomycetes, it was estimated that only 3 % of the compounds produced by this genus has been characterized (Watve et al., 2001). So in that respect, even streptomycetes can be worth pursuing. The screening programs from the 1950s and up until recently have mostly focused on actinomycetes of terrestrial origin and the existence of purely marine

actinomycetes has not been widely accepted (Jensen and Fenical, 2005). More recent isolations of marine-derived actinomycetes that are phylogenetically unique compared to their soil-derived relatives (Magarvey et al., 2004; Jensen et al., 2005; Pathom-aree et al., 2006), have shaken this perception. In addition, the new taxa have turned out to be good sources for compounds of novel, unique structures, such as the salinosporamides and sporolides from *Salinispora* spp. (Williams et al., 2005; Buchanan et al., 2005) and the marinomycins from *Marinispora* spp. (Kwon et al., 2006) (Figure 1. 4). Several of these compounds have anticancer activity and salinosporamide A has made it into preclinical development as an anticancer agent (Jensen and Fenical, 2005). Several studies have come to the conclusion that there is a great diversity of actinobacteria in marine environments and with the world's oceans covering 70 % of the earth's surface, the prospects for finding new drug leads among compounds produced by marine actinomycetes look promising (Fenical and Jensen, 2006; Bull and Stach, 2007).

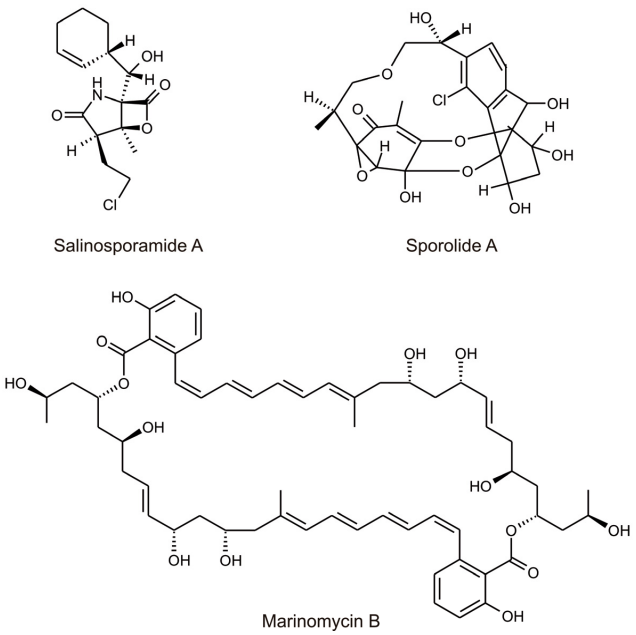


Figure 1. 4 Novel structures from *Salinispora* spp. and *Marinispora* spp. Adapted from Jensen and Fenical (2005).

1.3.2 Methods for discovering new bioactive natural products

Several approaches have been proposed and utilized to harness the "unculturable" microbes present in soil and marine environments. Through application of new methods for isolation and cultivation, bacterial species previously assumed to be unculturable have been recovered from environmental samples (Zengler et al., 2002; Joseph et al., 2003; Maldonado et al., 2005). The cloning of "environmental DNA" (so-called metagenomics) can also be used to exploit microbes that have so far not been cultivated (Handelsman et al., 1998; Gillespie et al., 2002). The isolation of microbial organisms from different types of ecosystems and locations is still considered a very important approach in the search for new bioactive molecules (Zhang, 2005), and the time of sampling may be of significance as there appear to be temporal differences in the diversity of species (Morris et al., 2005). The presence of viruses in samples may also influence the outcome and removal of such factors can greatly increase yield (Bouvier and del Giorgio, 2007).

Although the outcome from combinatorial chemistry in number of drug leads has been rather poor, the technical advancements achieved in the last 20 years have had important consequences for natural product screening. Development of high throughput methods for isolation, cultivation and screening of microorganisms have greatly reduced the time and effort needed to be invested in such measurements (Bruns et al., 2003; Gich et al., 2005; Casey et al., 2004). Combined with new sensitive methods, such as liquid chromatography mass spectrometry (LC-MS) and two-dimensional NMR spectra arrays, active compounds can be identified more rapidly (Schroeder et al., 2007; Koehn, 2008). Such methods may also be used for metabolic profiling and can thus be an aid in dereplication. Genome sequencing has opened up a new way for discovering bioactive natural products. Actinomycete genome sequences studied so far contain far more secondary metabolite clusters than the number of identified metabolites would suggest (Bentley et al., 2002; Udwary et al., 2007). Activation of pathways for such orphan clusters, either in the original strain or in a heterologous host, may lead to isolation of new interesting metabolites (Zazopoulos et al., 2003; McAlpine et al., 2005).

1.4 Biosynthesis of polyketide and nonribosomal peptide natural products

Thousands of bioactive natural products are polyketides (PK), nonribosomal peptides (NRP) or hybrids of polyketides and nonribosomal peptides. These include for example the polyketides amphotericin B and nystatin (antifungals) (Caffrey et al., 2001; Brautaset et al., 2000), the nonribosomal peptides penicillin and vancomycin (antibacterials) (Martín, 1998; Süssmuth and Wohlleben, 2004), and the hybrid PK-NRPs bleomycin (anticancer) (Shen et al, 2002) and cyclosporin A (immunosuppressant) (Weber et al., 1994).

Both polyketides and nonribosomal peptides are synthesized by large enzyme complexes through condensation of specific building blocks (Fischbach and Walsh, 2006). The identity and order of domains in these enzymes determine which building blocks are incorporated and in what order, and what chemical modifications the blocks are subjected to.

1.4.1 Polyketide synthases (PKS)

There are three types of PKS (polyketide synthase) systems (type I, II and III), but only the modular type I PKS will be described in detail. Type II and III will be considered briefly below. Both PKS and NRPS (nonribosomal peptide synthetase) systems make use of thioesters to perform condensation reactions (Fischbach and Walsh, 2006). PKS systems are based on the condensation of short chain organic acids to form a long polyketide chain (Donadio et al., 1991). The simplest unit that can be utilized as a starter is acetyl-CoA, but also longer chain acids such as malonyl-CoA or methylmalonyl-CoA can serve as starter units (Brautaset et al., 2000; Leadlay et al., 2001; Waldron et al., 2001). As the condensation reaction releases CO_2 in the process, the simplest extender unit that can be incorporated is malonyl-CoA. Other possible extender units are e.g. methylmalonyl-, ethylmalonyl-, and methoxymalonyl-CoA

(Haydock et al., 1995; Reeves et al., 2001; Haydock et al., 2005). Units to be incorporated are selected by an acyltransferase (AT) domain and transferred to a prosthetic phosphopantetheinyl group of an acyl carrier protein (ACP) domain, generating the activated thioester (Fischbach and Walsh, 2006). The ACP domain does not itself possess any catalytic activity, but represents a platform that brings the substrates and the catalytic centers together. A Claisen condensation is carried out by a ketosynthase (KS) domain, linking together a new extender unit with the growing polyketide chain. This reaction starts with transfer of the acyl chain from the upstream ACP domain to the conserved cysteine of the KS domain active site (Fischbach and Walsh, 2006). The KS decarboxylates the downstream extender unit to form a thioester enolate, which then performs a nucleophilic attack on the upstream acyl thioester (Figure 1. 5 A).

The KS, AT and ACP domains constitute a minimal extender module, but other domains may also be present (Khosla and Harbury, 2001). A ketoreductase (KR) domain will modify the ketogroup of the extender unit incorporated by the preceding module. The hydroxyl group resulting from the ketoreduction by the KR domain can be further modified by a dehydratase (DH) domain and the presence of a third domain, enoyl reductase (ER), will result in formation of a fully saturated α-β bond (Figure 1. 5 B).

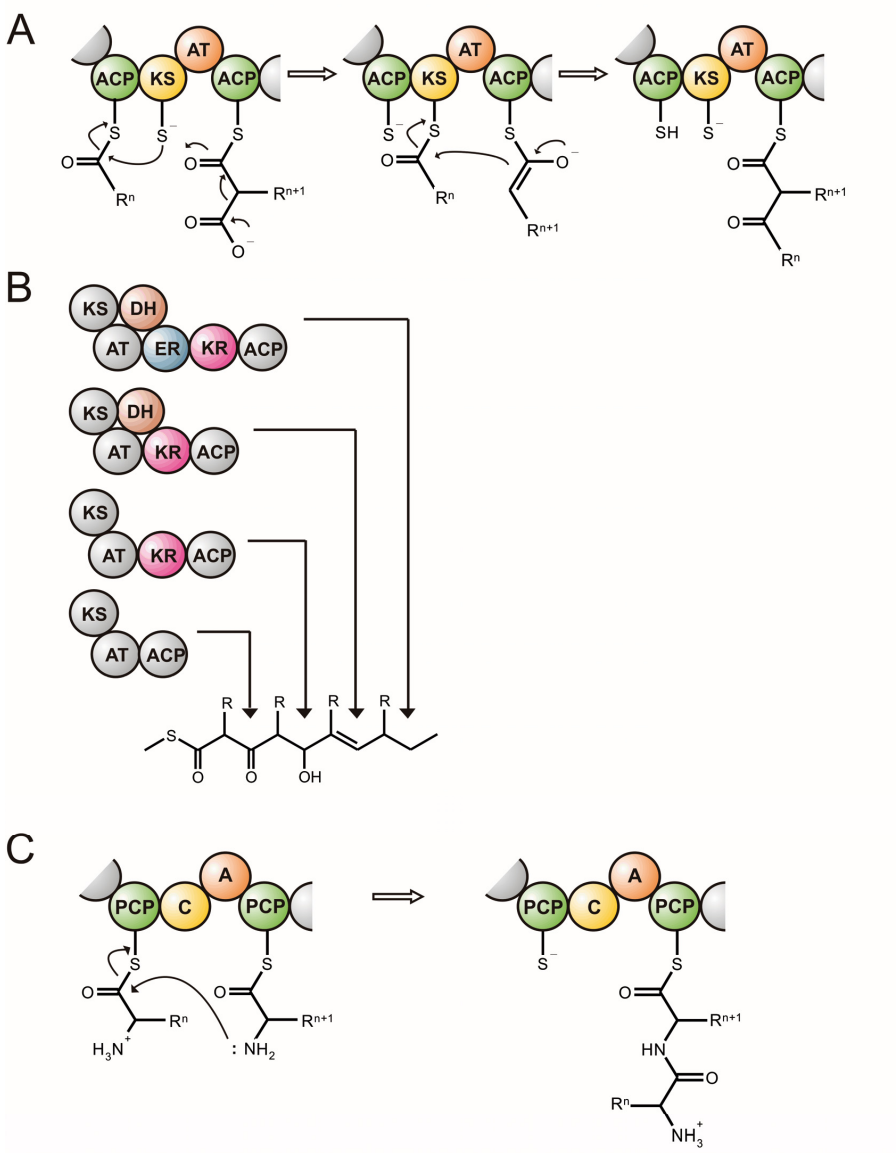


Figure 1. 5 A) The condensation reaction in polyketide synthesis. The upstream acyl chain is transferred to the conserved cysteine of the KS domain. The KS domain decarboxylates the downstream extender unit to form the thioester enolate, which attacks the upstream acyl thioester. Adapted from Fischbach and Walsh (2006). B) Modifications by the KR, DH and ER domains. The minimal KS-AT-ACP module will leave a ketogroup. An additional KR domain will result in a hydroxyl group, a KR and a DH domain will produce a double bond and a KR, a DH and an ER domain will yield a fully saturated α-β bond. Adapted from Jenke-Kodama et al. (2006). C) The condensation reaction in nonribosomal peptide synthesis. The C domain catalyzes a C-N bond formation between the downstream nucleophilic aminoacyl amine group and the upstream electrophilic peptidyl thioester. Adapted from Fischbach and Walsh (2006).

Initiation and termination modules are somewhat different from the elongation modules. Initiation modules often lack a KS domain and the starter unit is simply selected by the AT domain and loaded on the first ACP domain. Initiation modules may, however, contain KS-like domains lacking condensing, but retaining decarboxylating activity for processing dicarboxylic acid starters prior to the subsequent chain elongation (Bisang et al., 1999, Long et al., 2002). The last module in the polyketide synthesis is the termination module, which is responsible for releasing the mature polyketide chain. The termination module contains a thioesterase (TE) domain, which hydrolyzes the thioester bond between the last ACP domain and the mature polyketide chain, and is also responsible for the chain cyclization (Fischbach and Walsh, 2006).

In type II PKS systems, individual domains are separate proteins and interact in trans. The growing polyketide chain remains tethered to the same ACP domain and the KS and AT domains act iteratively. Chain length is determined by the chain length factor (CLF), which forms a heterodimer with the active subunit KS_{α} (Fischbach and Walsh, 2006). Type III PKS systems also work iteratively, but in contrast to type I and II systems, there are no ACP domains and acyl-CoA substrates are used directly (Shen, 2003).

1.4.2 Nonribosomal peptide synthetases (NRPS)

NRPS systems perform condensation of amino acids to form peptide chains (Sieber and Marahiel, 2005), and are involved in the biosynthesis of many important antibiotics (e.g. vancomycin, bleomycin). Available loading and extender units constitute the common 20 amino acids in both L-and D-form as well as other amino acids and carboxylic acids. More than 300 different units have been described (Kleinkauf and von Döhren, 1990). In analogy to the PKS systems, NRPSs can also be divided into modules and domains, where each module is responsible for incorporation and modification of one specific amino acid (Schwarzer et al., 2003). Instead of AT domains, the NRPSs will have adenylation (A) domains responsible for selecting and activating the amino acid units to be incorporated into the peptide chain. The amino acids are activated

through adenylation by the A domain and transferred to the prosthetic phosphopantetheinyl group on a peptidyl carrier protein (PCP) domain, which is analogous to the ACP domain of the PKS (Sieber and Marahiel, 2005). Transpeptidation of the amino acid substrates is carried out by the condensation (C) domain (Figure 1. 5 C). As for PKS systems, there are also other domains present in NRPS systems in addition to the core C-A-PCP module. Peptide antibiotics often contain D-amino acid residues and since D-amino acids are generally not exploited in the microbial antibiotic producers' primary metabolism, NRPSs often contain epimerization domains (Stein et al., 2005). Some peptide antibiotics have *N*-methylated peptide bonds generated by *N*-methylation domains (Nishizawa et al., 1999; Schauwecker et al., 2000). Additional NRPS domains include e.g. specialized condensation domains with heterocyclization activity, oxidation domains and reduction domains (Gehring et al., 1998; Du et al., 2000; Reimmann et al., 2001).

1.4.3 Hybrid PKS/NRPS systems

PKS and NRPS systems are also compatible with each other and several hybrid PKS-NRPS assembly lines have been reported (Shen et al., 2002; O'Connor et al., 2002; Wu et al., 2000). While most hybrid PKS-NRPS systems involve direct contact between PKS and NRPS modules, there are also systems that are not true hybrids and where the PKS and NRPS parts are synthesized separately (Du et al., 2001). Such a system operates in the coronatine biosynthesis, where the polyketide and amino acid-derived parts are presumably joined together by a discrete ligase (Bender et al., 1999). In true hybrids, there will be transfer of the growing chain across PKS-NRPS and/or NRPS-PKS interfaces (Figure 1. 6). At an NRPS-PKS interface, the downstream KS domain must then accept a peptidyl chain rather than an acyl chain and vice versa for an NRPS-PKS interface (Fischbach and Walsh, 2006).

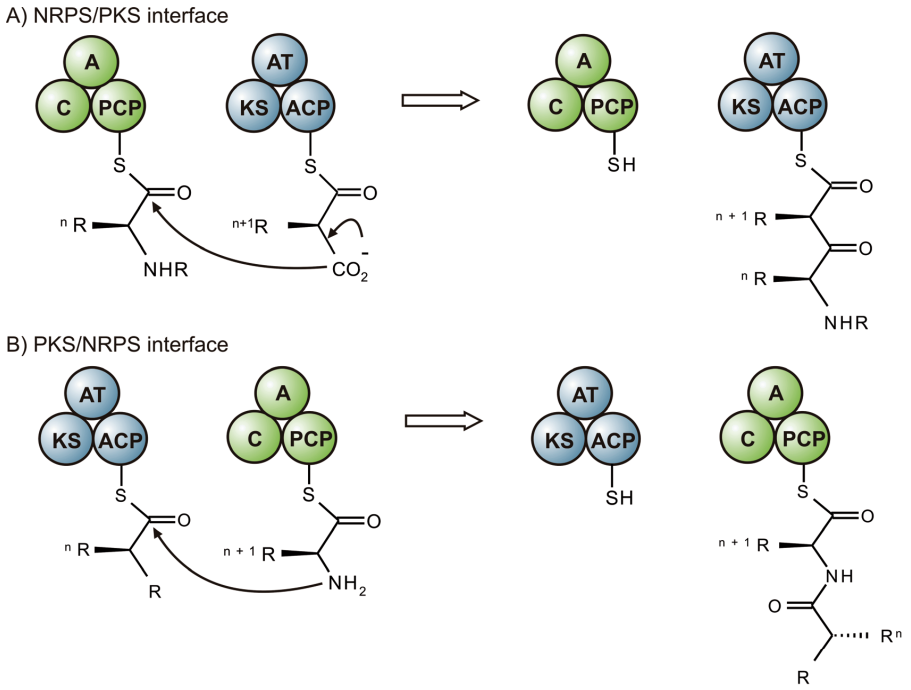


Figure 1. 6 Hybrid polyketide and nonribosomal peptide synthesis. A) NRPS/PKS interface: KS domain catalyzed C-C bond formation. B) PKS/NRPS interface: C domain catalyzed C-N bond formation. Adapted from Shen et al. (2001).

1.4.4 Precursors and post-PKS/NRPS modifications

Both PK and NRP synthesis make use of special monomer units that are not part of the common biosynthetic machinery of the cell. Synthesis of such precursors requires dedicated enzyme systems and genes encoding such enzymes are usually clustered with the PKS and NRPS genes (Fischbach and Walsh, 2006). In the biosynthesis of chloroeremomycin, five out of seven amino acids are nonproteinogenic and 11 additional enzymes are required to convert precursors into the corresponding substrates for the NRPS assembly line (van Wageningen et al., 1998; Hubbard and Walsh, 2003). Many PKSs use amino acids, derivatives of amino acids or other amines as starter units, e.g. in the biosynthesis of the ansamycin antibiotic rifamycin, 3-amino-5-hydroxybenzoic acid (3,5-AHBA) is utilized as a starter (Moore and Hertweck, 2002;

Admiraal et al., 2001), while 3-methylaspartate is incorporated in the biosynthesis of the macrolactam vicenistatin (Ogasawara et al., 2004).

PKs and NRPs often require post-assembly modification to achieve full activity. The most common post-PKS/NRPS modifications are hydroxylation, glycosylation and methylation and genes encoding enzymes performing these modifications are usually found clustered with the PKS/NRPS genes. Hydroxylations are often performed by P450 monooxygenases and P450 monooxygenases are commonly found encoded within PKS biosynthetic gene clusters (Weber et al. 1991; Betlach et al., 1998; Byrne et al., 2003). PKs and NRPs can be decorated with sugar moieties and genes involved in the biosynthesis of these sugar moieties, as well as the glycosyltransferases responsible for attaching the sugar molecules to the aglycone scaffolds, are clustered within the PKS/NRPS biosynthetic gene clusters (Losey et al., 2001; Nedal and Zotchev, 2004). NRPSs often contain methyltransferase (MT) domains, but MTs can also be found as separate enzymes performing methylation in trans, such as the *N*-MT in the chloroeremomycin biosynthesis and the *O*-MT in the saframycin Mx1 biosynthesis (O'Brien et al., 2000; Pospiech et al., 1996).

1.4.5 Other aspects of PKS and NRPS systems

Recent literature indicates that polyketide synthesis and nonribosomal peptide synthesis cannot be confined to the strict categories of type I, II and III systems. Variations on the theme seem to be abundant (Shen, 2003; Wenzel and Müller, 2005). As described above, the conventional minimal or core extender module consists of three domains (KS-AT-ACP or C-A-PCP), but several cases have been described were a domain is missing from this trio. In the biosynthesis of yersiniabactin, three NRPS modules are all supplied with substrate by the same A domain present in only one of the modules (Keating et al., 2000) and in the biosynthesis of leinamycin and disorazols, the PKS modules lack the AT domain. Extender units are supplied in trans by a discreet AT (Cheng et al., 2003; Carvalho et al., 2005). Domains may be skipped, as in the biosynthesis of pikromycin and myxochromides (Beck et al., 2002; Wenzel et al.,

2005), or used more than once, so called stuttering (He and Hertweck, 2003; Gaitatzis et al., 2002). Modules may be split so that domains belonging to the same module are present on separate proteins (Kopp et al., 2005; Silakowski et al., 2001). As more and more PKS and NRPS clusters are being discovered, these heretofore unusual systems seem to represent more of a norm than an exception and the well-known "textbook examples" no longer seem to adequately portrait the full extent of PK and NRP synthesis.

1.4.6 Genetic engineering of PKS/NRPS and combinatorial biosynthesis

The realization that PKS systems are amenable to modifications in a way that facilitates production of new types of molecules followed shortly after the isolation of the first PKS gene cluster (Malpartida and Hopwood 1984; Hopwood et al., 1985; Õmura et al., 1986). The sequencing and characterization of the first gene cluster for a modular type I PKS, established erythromycin biosynthesis in *Saccharopolyspora erythraea* as a model system for polyketide synthesis and represented the first example of genetic modification of a complex PKS (Cortes et al., 1990; Donadio et al., 1991). Since then, genetic engineering has produced numerous derivatives of well-known antibiotics through manipulation of both PKS and NRPS systems (Baltz, 2006).

The modular nature of many PKS and NRPS systems make them particularly suitable for genetic engineering as predicted structural changes in a compound can be introduced by removal of specific domain functionalities or introduction of new domains, e.g. modification of KR, DH and ER domains allow for changes in the reduction of the β-carbonyl at particular sites in the structure of a compound (Donadio et al., 1991; Donadio et al., 1993; Tang et al., 2005; Borgos et al., 2006) and swapping of domains or site-specific mutations can alter the substrate specificity of both A and AT domains (Ruan et al., 1997; Schauwecker et al., 2000; Reeves et al., 2001; Eppelmann et al., 2002). Knowledge about how PKS subunits interact through docking domains, as well as the presence of intrapolypeptide linkers, have improved the possibilities for combining individual domains and modules of different origin into new functional

enzyme complexes (Menzella et al., 2005; Weissman, 2006; Menzella et al., 2007). This bodes well for the future generation of large compound libraries through combinatorial biosynthesis, however, at the moment it is difficult to combine just two or three modules efficiently and the generation of large, complex systems resides somewhat further down the line (Menzella and Reeves, 2007).

1.5 Horizontal gene transfer

Organisms can obtain their genetic material in two ways; by vertical gene transfer or horizontal/lateral gene transfer (HGT/LGT). Vertical gene transfer is the passing on of genes from ancestors to descendants, while horizontal gene transfer is the direct passage of genetic material from one organism to another, independent of descent.

Horizontal gene transfer in bacteria was first described in 1928, when transfer of virulence determinants between pneumococci in infected mice was detected (Griffith, 1928). In the 1960s, it became clear that antibiotic resistance genes could be transferred between pathogenic strains of bacteria (Watanabe and Fukasawa, 1961; Watanabe, 1963). The extent of HGT in bacteria and archaea and what role it has played/plays in the evolution of prokaryotes has been heavily debated since the first bacterial genomes were sequenced in the 1990s and comparison of whole genomes was made possible. Complete sequencing of several *Streptomyces* genomes and subsequent genome comparisons, as well as the species specific distribution of secondary metabolite biosynthetic gene clusters among actinomycetes point to a possible role for HGT in the evolution and dissemination of such clusters (Chater and Kinashi, 2007).

1.5.1 Ways of acquiring foreign DNA

There are essentially three ways a cell can acquire foreign DNA; transformation, conjugation and transduction. All three processes are well documented. Uptake of DNA by natural transformation has been described for many bacterial strains and while some

strains are perpetually competent, others can acquire competence at certain stages in their life cycle (Ochman et al., 2000). Conjugation usually involves plasmids or transposable elements. Such elements may also integrate into chromosomal DNA and can in that way mobilize chromosomal segments (Thomas and Nielsen, 2005). While conjugation requires direct contact between the donor and recipient cells, transduction is not dependent on direct contact and the donor and recipient cells may even be separated in time. Phages can package random DNA fragments (general transduction) or include DNA adjacent to their chromosomal attachment site (specialized transduction) (Ochman et al., 2000). However, functional phages usually package their own genome and more seldom host chromosomal DNA fragments. Derivatives of tailed bacteriophages, called gene transfer agents (GTAs), have recently been discovered in several groups of bacteria and archaea (Stanton, 2007; Lang and Beatty, 2007). GTAs preferentially package random fragments of chromosomal DNA rather than parts of their own genome and may represent specialized HGT devices (Stanton, 2007; Koonin and Wolf, 2008).

Having entered the recipient cell, there are more obstacles to overcome for a foreign DNA fragment. In order not to be eliminated from the new host, it must either be able to replicate autonomously (plasmids) or it must integrate into the chromosome of the host. Integration into chromosomal DNA may be mediated by e.g. homologous recombination. A homologous region of 25-200 bp is necessary for homologous recombination to occur and the process is therefore most successful when the donor and recipient are quite closely related (Thomas and Nielsen, 2005). Insertion sequence (IS) elements and other repeated sequences can mediate homologous recombination.

1.5.2 How to detect HGT

There are several methods that can be applied to detect HGT. One common approach is to look for atypical phylogenetic distributions (Ragan, 2001). If two distantly related species both contain highly similar nucleotide regions, invocation of HGT as the cause is usually unproblematic, such as in the extreme case of *Pyrococcus furiosus* and *Thermococcus litoralis*, which share a transposon-like sequence of 16 kb that is 99 %

similar (DiRuggiero et al., 2000). However, for less conserved DNA regions or protein sequences, establishing the right set of homologs can be difficult. The term “homology” implies common ancestry and a certain level of sequence identity between two protein sequences does not ensure a common ancestral relationship (Ragan, 2001). Another possibility is to look for regions of deviant composition, such as differences in GC content, different patterns of codon usage or dissimilar frequencies of oligonucleotides (Ragan, 2001). However, one cannot rule out the possibility that such regions may arise by other processes than HGT. Association of factors common in mobile genetic elements with an atypical genomic region would strengthen the suspicion that HGT has occurred. Such factors can be insertion sequence elements, transfer origins of plasmid or attachment sites of phage integrases. A more recent approach to detecting HGT, particularly among closely related species, is comparison of whole genomes. That aspect will be covered in more detail below. Incongruence among phylogenetic trees is another indication of HGT. If one can construct well-resolved trees for two or more gene families and the trees are incongruent, then HGT is considered the most likely explanation. Caution should be taken, however, as inference of phylogenetic trees are subject to bias and methodical artifacts (Ragan and Charlebois, 2002).

What can be considered the best approach will vary with each individual case, e.g. differences in GC content would not be applicable for investigating HGT between two distantly related GC-rich species. In addition, one can say that ancient transfer events are difficult to detect as well as transfer between closely related species.

1.5.3 Comparison of genomes

Sequencing of the first bacterial genome (*Haemophilus influenzae*) was completed in 1995 (Fleischmann et al., 1995). Since then, the genomes of hundreds of other prokaryotes, as well as eukaryotes, have been sequenced. At the moment, 788 bacterial and 56 archaeal genomes have been completed while 1538 more are in the pipeline (http://www.ncbi.nlm.nih.gov, March 2009). Ever since the first genomes became available for comparison, a debate on the extent of horizontal gene transfer has raged.

Opinions have ranged from HGTs being so frequent that phylogeny is rendered useless to the view that HGTs are rare and account for only minor noise when inferring phylogeny (Philippe and Douady, 2003). As more and more genome sequences have been published, the picture has become more clear, but also much more complex.

One early study compared the genomes of non-pathogenic *Escherichia coli* K12 and its pathogenic relative O157:H7 (Perna et al., 2001). The two strains diverged from a common ancestor some 4.5 million years ago. O157:H7 contains 1387 genes distributed among several different pathogenicity islands that are not present in K12, while K12 has 528 genes not found in O157:H7. This means that as much as 30 % of the pathogen's genome may have been acquired by recent HGT. Further studies on *E. coli* has shown that recent additions to the metabolic network are mostly due to HGT rather than gene duplications and that genes encoding coupled enzyme pairs are often transferred together as parts of operons (Pál et al., 2005).

A recent study of 312 bacterial and 26 archaeal genomes found that the distribution of orthologous genes among prokaryotes points to the major trends in prokaryotic evolution being "extensive horizontal transfer of genes, pervasive gene loss and functional plasticity of many cellular systems" (Koonin and Wolf, 2008). Another study of 181 prokaryotic genome sequences found that approximately 80 % of the genes in each genome at some point in time had been involved in HGT (Dagan et al., 2008). This necessarily has consequences for the Tree of Life paradigm, as the tree would only fit for a small part of the prokaryotic genome. Thus, the Tree of Life is no longer considered enough to truly represent the evolution of prokaryotes (Doolittle and Bapteste, 2007). Instead, a dynamic network including both vertical and horizontal gene transfer would represent a more accurate picture (Koonin and Wolf, 2008).

1.5.4 Horizontal gene transfer in *Streptomyces*

Comparison of *Streptomyces* genomes has revealed a central conserved region with high synteny and highly variable terminal regions (Ikeda et al., 2003). The terminal regions vary considerably even between closely related species such as *S. ambofaciens* and *S.*

coelicolor and the central conserved region decreases in size with increasing phylogenetic distance (Bentley et al., 2002; Choulet et al., 2006). Moreover, the terminal regions contain few essential genes and a much higher proportion of transposable elements than the rest of the chromosome (Bentley et al., 2002; Chen et al., 2002), suggesting that the variability of the terminal regions is a result of HGT events and DNA rearrangements (Choulet et al., 2006). In addition, gene clusters related to secondary metabolism are often found to be species specific, e.g. the spiramycin biosynthetic gene cluster in *S. ambofaciens* is not present in *S. coelicolor*, while the actinorhodin, undecylprodigiosin and *cda* gene clusters have no counterparts in *S. ambofaciens* (Choulet et al., 2006). Comparison of the 23 gene clusters related to secondary metabolism in *S. coelicolor* with the 30 gene clusters found in *S. avermitilis* show that many are present only in one genome and not in the other (Ikeda et al., 2003; Ventura et al., 2007). The terminal regions are usually rich in gene clusters for secondary metabolism, especially clusters that are species specific. Gene clusters that are present in several species are more often found in the core region, such as those for the biosynthesis of siderophores and geosmin (Ikeda et al., 2003; Ventura et al., 2007). The sporadic occurrence of gene clusters for secondary metabolite production in *Streptomyces* suggests that some of them may have been acquired through HGT (Chater and Kinashi, 2007), although evidence seem to have been provided only for the streptomycin biosynthetic gene cluster (Egan et al., 1998; Egan et al., 2001). The mechanism of transfer was, however, not addressed.

There seem to be no reason to doubt that HGT occurs among streptomycetes, however, the extent of their promiscuity and what consequences it has for their evolutionary development, remains to be elucidated. The sequencing and comparison of more *Streptomyces* genomes may provide better answers in the following years.

1.5.5 Linear plasmids in *Streptomyces*

Linear as well as circular plasmids are abundant in *Streptomyces* (Hopwood and Kieser, 1993). While the size of circular plasmids is usually quite small, linear plasmids up to 1

Mb in size has been described (Gravius et al., 1994). Where studied, most linear plasmids appear to be transmissible by conjugation (Vivian, 1971; Hopwood et al., 1983), but since not all carry genetic markers, the phenomenon cannot be studied in all cases. Linear plasmids can mediate transfer of chromosomal DNA by integration into the chromosome or possibly by protein-protein interactions between the terminally bound proteins (Hopwood and Wright, 1973; Hopwood and Wright, 1976; Hopwood, 2006), but as this has only been shown for laboratory-grown strains, it is uncertain if this is a naturally occurring mechanism.

Perhaps the best studied linear plasmid is SCP1 from *S. coelicolor*. The early work of D. A. Hopwood and his coworkers demonstrated that *S. coelicolor* carries a plasmid fertility factor (SCP1) capable of integrating into the chromosome and giving rise to high-fertility variants analogous to the *E. coli* F-factor (Hopwood, 1967; Hopwood et al., 1969; Vivian and Hopwood, 1970; Vivian, 1971; Vivian and Hopwood, 1973; Hopwood and Wright, 1973; Hopwood and Wright, 1976). SCP1 resisted isolation for many years as it was thought to be circular, but the invention of pulsed field gel electrophoresis in the 1980s demonstrated its linear nature (Kinashi et al., 1987). SCP1 was, however, not the first linear plasmid to be discovered as Hayakawa et al. described the linear plasmid pSLA2 from *S. rochei* already in 1979 (Hayakawa et al., 1979).

In the 1970s, it was speculated that antibiotic biosynthetic gene clusters might in general be carried by plasmids (Okanishi et al., 1970), and the discovery that SCP1 harbors the gene cluster for methylenomycin biosynthesis further fueled the speculations (Kirby et al., 1975). The supposition was later disproved and only a few other examples of plasmid encoded gene clusters related to antibiotic biosynthesis have been found (Kinashi et al., 1987; Kinashi et al., 1988; Mochizuki et al., 2003; Jia et al., 2006; Migita et al., 2009). Rather interesting is the fact that the methylenomycin gene cluster has also been found on a large circular plasmid, pSV1, from *S. violaceoruber* SANK95770 (Okanishi et al., 1980). The two gene clusters are 99 % identical, but the rest of the plasmid sequences share little homology, suggesting recent horizontal transfer of the cluster (Yamasaki et al., 2003).

Another well-studied linear plasmid is pSLA2-L from *S. rochei*, carrying several clusters for secondary metabolite production, including the antibiotics lankamycin, lankacidin and an unknown aromatic polyketide (Mochizuki et al., 2003). The plasmid sequence itself did not reveal any explanation for the dense concentration of such clusters on pSLA2-L, however, it has been speculated that linear plasmids might provide a platform for bringing several gene clusters together for a natural type of combinatorial biosynthesis or for "selection of synergistic cooperations" (Chater and Kinashi, 2007).

Besides biosynthetic gene clusters for secondary metabolites, *Streptomyces* linear plasmids may provide the ability for growth in toxic environments. Ravel et al. isolated several *Streptomyces* strains carrying large linear plasmids conferring resistance against mercuric compounds (Ravel et al. 1998; Ravel et al., 2000). Two of the plasmids were shown to be transmissible to another *Streptomyces* strain in sterile soil microcosms, indicating that mercury resistance may spread among streptomycetes via conjugative linear plasmids (Ravel et al., 2000).

Both the linear plasmids and the linear chromosomes of *Streptomyces* have terminal inverted repeats (TIRs) and terminal proteins attached to the ends. The plasmid and chromosome ends appear to be rich in transposable elements, which may facilitate recombinatorial exchanges between the plasmids and chromosomes (Chater and Kinashi, 2007). Exchange between a linear plasmid and the chromosome has been reported for *S. rimosus*, where the linear plasmid apparently had acquired the oxytetracycline biosynthetic gene cluster from *S. rimosus*' chromosome (Gravius et al., 1994). It is possible that exchanges can also occur between plasmids. SCP1 has internal regions that resemble plasmid ends, indicating earlier integration of other linear replicons, and the methylenomycin gene cluster may have been acquired from a precursor of the circular plasmid pSV1 of *S. violaceoruber* (Bentley et al., 2004). Combined with the highly transmissible nature of at least some linear plasmids, it is possible that such processes have played an important part in the evolution of *Streptomyces* genomes, perhaps particularly in relation to the distribution of gene

clusters for secondary metabolite production, and may account for the streptomycetes great success in colonizing the soil (Chater and Kinashi, 2007).

2 Aims of the study

The initial aim of this study was to discover new antifungal agents through bioprospecting of marine actinomycete bacteria. As described in the introduction, there is a critical need for new antifungal antibiotics as there are few options available to treat systemic fungal infections, and there are problems with toxicity and resistance. Isolation of bioactive natural products from actinomycete bacteria has historically been a great success and marine environments have remained largely unexplored as a source of actinomycete isolates. To achieve the goal of discovering new and preferably less toxic antifungal antibiotics from actinomycete isolates, several strategies were employed:

- Use of various production media to induce secondary metabolite production in a majority of isolates.
- Screening against two different clinical isolates of *Candida* differing in the resistance pattern to known antifungal agents.
- Spectroscopic screening of extracts for comparison of the UV-profiles.

In parallel with the screening for antifungal compounds, a selection of extracts from some of the isolates were also screened against several cancer cell lines with the aim of identifying new cytotoxic compounds.

Discovery of a potentially new, or in an other way interesting, compound, would make the candidate a subject for further studies aimed at isolating the biosynthetic gene cluster involved in production of the compound as well as elucidation of the biosynthetic pathway. A full genetic and biosynthetic overview is important if a compound is to be considered as a new drug candidate, but even if the compound would not be further pursued, the study could provide important information about enzymes involved in biosynthetic pathways and contribute to the accumulating knowledge about biosynthesis of natural products in general.

3 List of papers

Paper I: Jørgensen, H., Fjærvik, E., Hakvåg, S., Bruheim, P., Bredholt, H., Klinkenberg, G., Ellingsen, T. E., Zotchev, S. B. (2009). Candicidin biosynthetic gene cluster is widely distributed among *Streptomyces* spp. isolated from the sediments and the neuston layer in the Trondheimsfjord. Appl. Environ. Microbiol. In press.

Paper II: Jørgensen, H., Degnes, K. F., Sletta, H., Fjærvik, E., Dikiy, A., Herfindal, L., Bruheim, P., Klinkenberg, G., Bredholt, H., Nygård, G., Døskeland, S. O., Ellingsen, T. E., and Zotchev, S. B. Biosynthetic pathway for the cytotoxic macrocyclic lactam BE-14106 involves two distinct PKS systems and amino acid processing enzymes for generation of the aminoacyl starter unit. Submitted to Chem. Biol.

Paper III: Jørgensen, H., Degnes, K. F., Dikiy, A., Fjærvik, E., Klinkenberg, G., and Zotchev, S. B. Insights into the evolution of macrolactam biosynthesis: cloning and comparative analysis of the biosynthetic gene cluster for a novel macrocyclic lactam ML-449. Submitted to Appl. Environ. Microbiol.

4 Summary of results and discussion

4.1 Screening for producers of antifungal antibiotics and limited cytotoxic screen

(Basis for paper I, II and III)

At the time this study was initiated, a strain collection of actinomycete bacteria isolated from the Trondheimsfjord had already been established in our research group. A screening procedure for identifying producers of antifungal antibiotics was established with the prospect of discovering potentially new compounds. Extracts made from all isolates were assayed using two *Candida* species, one sensitive for polyene macrolides (*C. albicans*) and one resistant to amphotericin B (*C. glabrata*). Inhibition of *C. albicans*, but not *C. glabrata*, could therefore indicate the presence of polyene macrolides in the extract. In addition, all extracts were subjected to UV/VIS scans, as polyenes usually produce characteristic UV spectra due to having several conjugated double bonds (Õmura and Tanaka, 1984). As polyene macrolides are in general quite toxic, discovery of new (or old) polyene macrolides was seen as undesirable. Designing the screen in this way would hopefully allow for a quick way of detecting extracts containing such compounds. Cultivation of strains in a 96 well plate format and robotic handling of extraction samples, made it possible to screen more than 4000 isolates.

Evaluation of the UV/VIS spectra in connection with the inhibition pattern in the bioassay revealed a great number of putative polyene producers as well as several promising candidates presumed to be synthesizing non-polyenes. Of the latter, the candidates displaying the highest level of inhibition for both strains were chosen for further studies. A secondary screen was performed for 32 promising candidates to verify the bioactivity observed in the primary screen and the most interesting extracts were subjected to fractionation by HPLC. LC-MS-TOF performed on the bioactive fractions made it possible to identify the accurate mass of the putative active compound in several cases. Searching the Dictionary of Natural Products (http://dnp.chemnetbase.com/) using the accurate mass made it possible to assess whether the compound was new or

already described. In addition, extracts made from all 32 candidates were sent to the University in Bergen for a cytotoxic screen at the Department of Biomedicine. Results obtained in the antifungal screen, the cytotoxic screen and the LC-MS-TOF analysis formed the basis for selection of candidates for further studies.

4.2 Candicidin biosynthetic gene cluster is widely distributed among *Streptomyces* spp. isolated from the sediments and the neuston layer in the Trondheimsfjord

(Paper I)

As mentioned above, a large number of putative polyene producers were identified in the initial antifungal screen. A closer inspection of the UV/VIS spectra, revealed that over 70 % of the spectra indicating polyene production were of the same pattern and displayed exactly the same absorption maxima. To investigate if all these extracts contained the same compound, an LC-DAD-TOF-MS analysis was performed on 51 randomly chosen extracts of this kind. 11 extracts displaying different spectra were also included in the analysis. A compound with the accurate mass of 1108.5703 Da was shown to be present in 52 of the extracts and a search in the Dictionary of Natural Products (DNP) using a ± 5 ppm range returned the result of candicidin D, an aromatic polyene macrolide with antifungal activity first isolated in 1953 from a *Streptomyces griseus* strain (Lechevalier et al., 1953). In addition, two other extracts were shown to contain potentially new polyene macrolides, as none of the identified accurate masses returned a hit in DNP. Further work with these two compounds was not pursued, however.

4.2.1 A group of streptomycetes isolated from the Trondheimsfjord harbors the candicidin biosynthetic gene cluster

Such a high frequency of potential candicidin producers among the isolates prompted speculation about horizontal gene transfer and 8 morphologically diverse candidates

were chosen for a genetic study (*Streptomyces* spp. MP47-06, MP47-91, MP18-04, MP15-36, MPS08-73, MPS08-39, MPS05-43 and MPS05-34). Partial 16S rDNA sequences were obtained for the 8 isolates and *S. griseus* IMRU 3570 (original candicidin-producer). *S. griseus* and 7 out of the 8 isolates turned out to have very similar 16S rDNA sequences, indicating that they are probably closely related, whereas one isolate (*Streptomyces* sp. MPS05-43) was significantly different from the rest. To verify that the isolates were indeed harboring the candicidin biosynthetic gene cluster (*can*), 3 different gene fragments (from *canRA* [ABC transporter], *pabAB* [p-aminobenzoic acid synthetase] and *canP3* [PKS]) were amplified from 2 of the isolates with primers designed for the *can* cluster from *S. griseus* IMRU 3570 (Campelo and Gil, 2002). PCR products were obtained for all three primer sets and two of the PCR products (from *pabAB* and *canP3*) were cloned and sequenced to verify that they belonged to the *can* cluster. Both fragments were found to be 99 % similar to the corresponding *S. griseus* sequences. To check for the presence of a *can* cluster in the remaining 6 isolates, a Southern blot analysis was performed using total DNA isolated from all 8 candicidin-producing isolates, *S. griseus*, a putative non-polyene producer (*Streptomyces* sp. MPS07-63) and a putative producer of a different polyene (*Streptomyces* sp. MPS07-67). Labeled versions of the sequenced PCR products were used as probes (*pabAB* and *canP3*). The results are presented in Figure 4. 1.

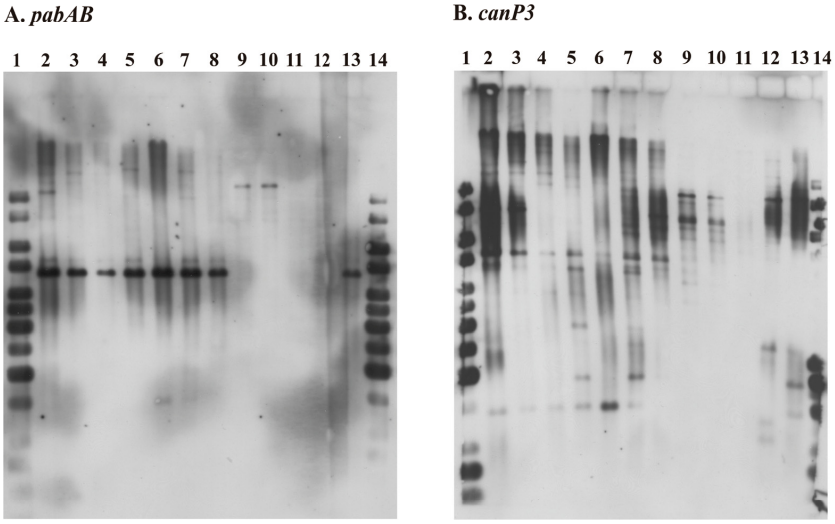


Figure 4. 1 Southern blot analyses with probes *pabAB* (A) and *canP3* (B). Both: Lane 1 and 14, Fermentas GeneRuler™ 1 kb DNA Ladder, lane 2, *S. griseus*, lane 3, *Streptomyces* sp. MP47-06; lane 4, *Streptomyces* sp. MP47-91; lane 5, *Streptomyces* sp. MP18-04; lane 6, *Streptomyces* sp. MP15-36; lane 7, *Streptomyces* sp. MPS08-73; lane 8, *Streptomyces* sp. MPS08-39; lanes 9 and 10, *Streptomyces* sp. MPS05-43; lane 11, *Streptomyces* sp. MPS07-63; lane 12, *Streptomyces* sp. MPS07-67; lane 13, *Streptomyces* sp. MPS05-34. DNA was digested with *Bam*H I.

For the *pabAB* probe, hybridization to a fragment of the same size was obtained for 7 of the 8 candicidin-producing isolates (lanes 3-8 and 13) as well as for *S.griseus* (lane 2). The hybridization pattern for *Streptomyces* sp. MPS05-43 (lane 9 and 10, same isolate included twice) was different, indicating a different organization of the gene cluster, at least in the *pabAB*-containing region. Interestingly, *Streptomyces* sp. MPS05-43 differed phylogenetically from the rest (see above). The pattern of hybridization was a bit more varied for the *canP3* probe, perhaps indicating the presence of more than one PKS cluster in several of the isolates. As the *pabAB* probe appeared to be specific for the *can* cluster, the presence of this gene cluster in the 8 isolates could be said to have been verified.

4.2.2 Candicidin biosynthetic gene cluster is located on a linear plasmid in one of the isolates

To examine the possibility that the candicidin gene cluster may have been transferred among isolates by the means of a mobile genetic element, the same isolates included in the Southern blot analysis were subjected to a pulsed field gel electrophoresis (PFGE). Such an experiment would show the presence of large plasmids in the chosen isolates. Pulsed field gels were blotted to allow hybridization with the *pabAB* and *canP3* probes. The results are presented in Figure 4. 2.

Large plasmids, presumably linear, were identified in 6 of the candicidin producing isolates and hybridization to one of the plasmids (lane 7) was obtained for both probes. The *can*-specific probes did not hybridize to plasmids present in the other isolates and hybridization to chromosomal DNA was observed for these isolates as well as *S. griseus*. Integration of plasmids into chromosomes has been described in at least two cases for *Streptomyces* strains (Kinashi et al., 1992; Gravius et al., 1994), and it is possible that the *can*-plasmid may be integrated into the chromosomes for some or all of the isolates.

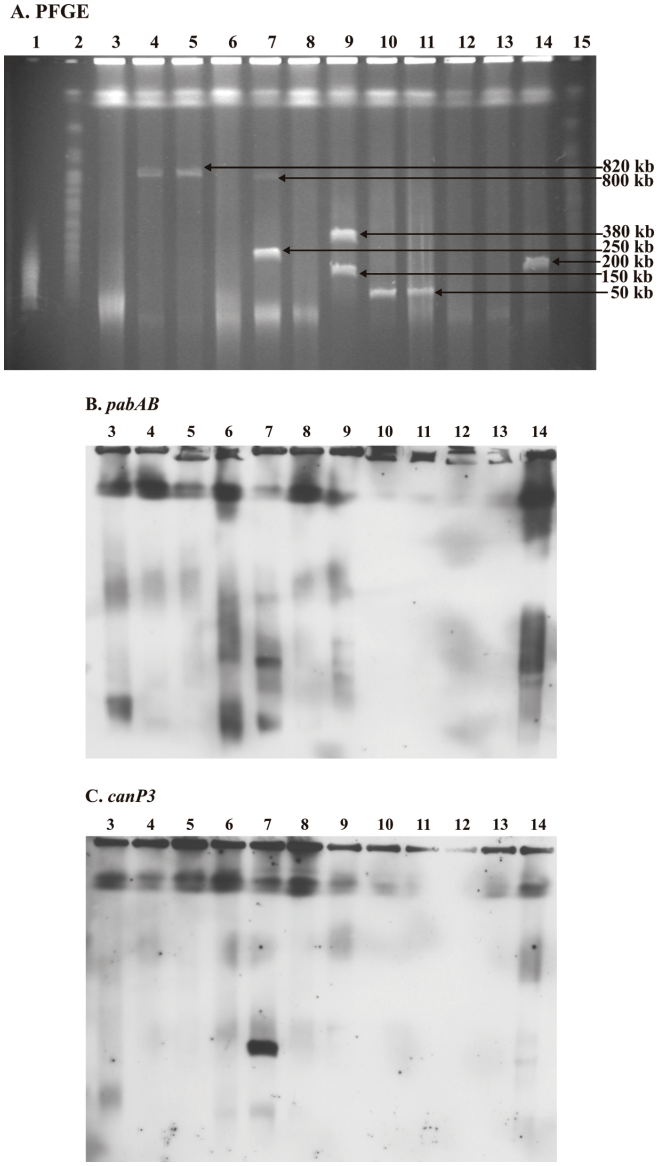


Figure 4. 2 (A) PFGE, size of linear plasmids indicated with arrows (B) Southern hybridization analysis with probe *pabAB* and (C) Southern hybridization analysis with probe *canP3*. Lane 1, MidRange II PFG Marker; lanes 2 and 15, Yeast Chromosome PFG Marker (markers not shown on Southern hybridization); lane 3, *S. griseus*; lane 4, *Streptomyces* sp. MP47-06; lane 5, *Streptomyces* sp. MP47-91; lane 6, *Streptomyces* sp. MP18-04; lane 7, *Streptomyces* sp. MP15-36; lane 8, *Streptomyces* sp. MPS08-73; lane 9, *Streptomyces* sp. MPS08-39; lanes 10 and 11, *Streptomyces* sp. MPS05-43; lane 12, *Streptomyces* sp. MPS07-63; lane 13, *Streptomyces* sp. MPS07-67; lane 14, *Streptomyces* sp. MPS05-34.

4.2.3 Mating and plasmid curing experiments

Many *Streptomyces* plasmids are known to be conjugative (Hopwood and Kieser, 1993) and mating is facilitated by mixing spores from the donor with that of the recipient and allowing them to develop mycelium and sporulate. To investigate the *can*-plasmid's ability for conjugative transfer to other *Streptomyces* strains, several mating experiments were undertaken. In a first attempt, *S. lividans* TK64 (pSET152) was chosen as a recipient, but transconjugants could not be detected. A second experiment was attempted with *S. noursei* (unpublished), but the results were again negative. In a third attempt, a Trondheimsfjord isolate was used as the recipient, considering the possibility that the "native" strain might be more suitable as a host for the plasmid. Screening for candicidin production among 100 random colonies of the recipient, resulted in two putative transconjugants. A Southern blot analysis of the putative transconjugants could, however, not confirm any transfer of the plasmid.

Next, a "curing" experiment of the *can*-plasmid-containing strain was attempted. It has been shown that growth at elevated temperatures can induce loss of linear plasmids in *Streptomyces* (Pang et al., 2002). The *can*-plasmid strain was incubated at several different temperatures and candidates with no candicidin production was observed after growth at 37°C. A subsequent Southern blot analysis showed no hybridization to the *pabAB* probe for one of the candidates, thus indicating loss of the plasmid. A mating experiment with the original *can*-plasmid strain and the "cured" strain resulted in reintroduction of the plasmid in the plasmid-free strain and restoration of the candicidin production. This experiment indicated that the *can*-plasmid is transmissible by conjugation, although transfer to strains other than the "cured" strain could not be achieved. The failure to demonstrate transfer to other strains may be due to several factors, such as instability of the plasmid in the chosen hosts or a transfer frequency below the detection limit of the chosen screening method. However, the study demonstrated that biosynthetic gene clusters for candicidin production seem to be wide-spread among marine sediment derived streptomycetes and the presence of such a gene

cluster on a transmissible plasmid in one of the isolates suggests that the *can* cluster can be spread to other strains via horizontal gene transfer.

4.3 Biosynthetic pathway for the cytotoxic macrocyclic lactam BE-14106 involves two distinct PKS systems and amino acid processing enzymes for generation of the aminoacyl starter unit

(Paper II)

From the antifungal screen described in section 4.1, 6 producers of a compound identified as Antibiotic BE-14106 in the Dictionary of Natural Products were recognized. As the compound displayed particularly good activity in the cytotoxic screen performed at the University of Bergen, it was decided that the compound should be further pursued. The 6 producers were found to be morphologically similar and 16S rDNA sequencing of 3 of the producers showed that they most likely represented replicates of the same strain. The isolate denoted *Streptomyces* sp. DSM 21069 displayed the highest level of production and was chosen for further characterization. To verify the identity of the isolated compound, structure determination by NMR was undertaken and confirmed that the compound was indeed BE-14106. The structure of BE-14106 is shown in Figure 4. 3.

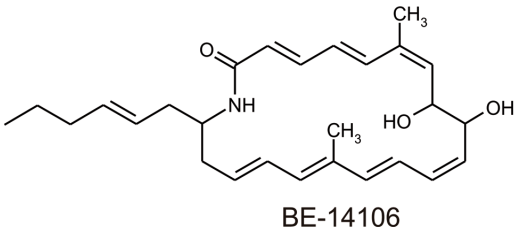


Figure 4. 3 The structure of BE-14106.

4.3.1 Cloning and sequencing of the BE-14106 biosynthetic gene cluster

A genomic cosmid library was constructed for *Streptomyces* sp. DSM 21069 and screened with a probe targeting PKS genes. The probe was generated using degenerate primers targeting the conserved KS domain-encoding regions of PKS genes. DNA fragments obtained by PCR were cloned, sequenced and compared to known PKSs by BLAST analyses. A fragment encoding part of a KS domain with strong homology to VinP1 involved in the biosynthesis of the macrolactam vicenistatin was judged to be the most promising as the structure of vicenistatin resembles that of BE-14106. The fragment was used for a gene inactivation experiment to verify involvement in the BE-14106 biosynthesis and the production of BE-14106 in the mutant was found to be severely affected. 2304 cosmid clones from the genomic library were screened with the KS probe, resulting in isolation of 3 positive clones, which were end-sequenced. Cross-sequencing with primers designed for the end-sequences established the degree of overlap between the 3 cosmids and a new probe was generated for one of the end-sequences to facilitate a search for the missing part of the biosynthetic gene cluster. Hybridization with the new probe resulted in isolation of two new cosmid clones, and one of them was found to cover the missing part. Before full sequencing of the cosmids was initiated, involvement of the gene cluster contained in the four cosmids in the biosynthesis of BE-14106 was confirmed by gene inactivation experiments using PCR generated fragments from two of the cosmids. All tested transconjugants were found to be deficient in BE-14106 biosynthesis and the four cosmids were fully sequenced.

4.3.2 Sequence analysis and gene inactivation experiments

The gene cluster contained in the four cosmids was annotated using Frameplot and BLAST analyses and 27 complete and 1 incomplete open reading frames (*orfs*) were identified in the 85 kb DNA region (Figure 4. 4 and Table 4. 1). Based on the sequence

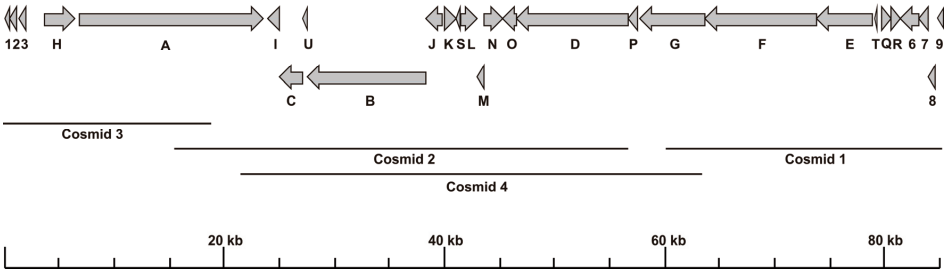


Figure 4. 4 Organization of the BE-14106 biosynthetic gene cluster in *Streptomyces* sp. DSM 21069 and the coverage of the four sequenced cosmids.

Table 4. 1 Description of *orfs* identified in and around the BE-14106 biosynthetic gene cluster.

Gene designation	Product (aa)	Putative function
orf1	135	Putative secreted metal-binding protein
orf2	199	Putative lipoprotein
orf3	144	Hypothetical protein
becH	951	LuxR-type transcriptional regulator
becA	5582	Polyketide synthase type I
becI	362	Glycine oxidase/FAD-dependent oxidoreductase
becC	694	Polyketide synthase type I
becU	187	Putative NRPS accessory protein
becB	3527	Polyketide synthase type I
becJ	532	AMP-dependent acyl-CoA synthetase/ligase
becK	323	Acyltransferase
becS	78	Peptidyl carrier protein
becL	505	NRPS adenylation domain
becM	198	TetR-type transcriptional regulator
becN	524	MFS-type efflux pump
becO	411	P450 monooxygenase
becD	3372	Polyketide synthase type I
becP	313	Putative L-amino acid amidase/proline iminopeptidase
becG	1986	Polyketide synthase type I
becF	3377	Polyketide synthase type I
becE	1631	Polyketide synthase type I
becT	95	Hypothetical protein, SimX2-like protein
becQ	257	Thioesterase type II
becR	237	PlsC-type phospholipid/glycerol acyltransferase
orf6	537	Putative tripeptidylaminopeptidase
orf7	256	Putative urease accessory protein
orf8	231	Putative urease accessory protein
orf9	incomplete	Putative urease accessory protein

analysis and functional assignment of domains for the PKS enzymes, a hypothesis for the biosynthetic pathway was proposed involving 21 of the *orfs* present in the gene cluster (region from *becH* to *becR*). In the proposed scheme, the putative PlsC-type phospholipid/glycerol acyltransferase, BecR, would be linking the C20-C25 acyl side chain to the macrolactam ring. To test this hypothesis, a mutant containing a deletion in *becR* was created and verified by a Southern blot analysis. The mutant was tested for BE-14106 production by fermentation and subsequent LC-MS-TOF analysis of fermentation extracts. The production of BE-14106 was found not to be influenced by the deletion in the *becR* gene and the proposed pathway for the BE-14106 biosynthesis was revised. Following the new hypothesis, the C20-C25 acyl side chain would be linked to an amino acid or an amino acid derivative at an early point in the biosynthesis, forming an aminoacyl starter for the macrolactam ring biosynthesis. The C20-C25 acyl chain was presumed to be made by the BecA PKS enzyme, while the macrolactam ring was assumed to be synthesized by the BecB, BecD, BecE, BecF and BecG PKS enzymes. The gene inactivation experiments described above involving PCR amplified fragments from two of the cosmids, where upon comparison of sequences found to affect BecA, and thus verifying a role for BecA in the biosynthesis. Of the remaining genes in the cluster, there were uncertainty about the role of *becI*, *becP* and *becC*, encoding a putative glycine oxidase/FAD-dependent oxidoreductase, L-amino acid amidase/proline iminopeptidase and a truncated PKS, respectively. In addition, the role of BecO as a P450 monooxygenase modifying the macrolactam ring by hydroxylation needed to be verified. Δ*becI*, Δ*becP*, Δ*becC* and Δ*becO* mutants were constructed, verified by Southern blot analyses and tested for BE-14106 production. The Δ*becI*, Δ*becP* and Δ*becC* mutants were found to be deficient in BE-14106 production, while the Δ*becO* mutant produced the expected deoxy-BE-14106. The remaining genes in the cluster were not subjected to inactivation experiments and their putative role will be addressed below in connection with the proposed biosynthetic pathway for BE-14106.

4.3.3 Heterologous expression of BecI and BecP and enzyme assays

To explore further the role of BecI and BecP in the biosynthesis of BE-14106, both enzymes were heterologously expressed in *Escherichia coli* as His-tag fusion proteins, purified and used for enzyme assays. BecI's role in the biosynthesis was presumed to be modification of an amino acid substrate, which would subsequently be joined together with the acyl side chain. The activity of BecI-6xHis was tested in an assay designed for D-amino acid oxidases using several different L-and D-amino acid substrates. Highest activity was achieved with glycine and D-asparagine, however, the overall activity level was rather low. BecP-6xHis was tested in an assay designed for proline iminopeptidases using L-proline-pNA (p-nitroanilide) as a substrate. High activity was achieved with this substrate and to verify that proline was BecP's preferred substrate the assay was repeated with L-alanine-pNA and L-lysine-pNA as negative controls. Surprisingly, L-ala-pNA turned out to be a better substrate. The assay was repeated also with glycine-pNA, but L-ala-pNA was found to be the best substrate of those tested. In addition, purified BE-14106 was incubated with BecP-6xHis and subjected to an LC-MS-TOF analysis to detect potential enzymatic degradation of the molecule, but such an effect could not be detected. The latter experiment was performed to evaluate the possibility that BecP might have a role in the BE-14106 resistance mechanism by causing an enzymatic cleavage of the macrolactam ring. Since no degradation of BE-14106 could be demonstrated, BecP most likely has a role to play in the biosynthesis of the macrolactam. This is also consistent with the observation that the Δ*becP* mutant does not produce BE-14106.

4.3.4 Feeding experiments

Judging from the results of the enzyme assay for BecI, glycine or D-asparagine were considered the most likely candidates for the starter unit of the biosynthesis and feeding experiments were initiated to investigate the potential incorporation of these amino acids. A setup involving ^{15}N isotope labeled medium components and feeding with ^{14}N-containing amino acids were chosen, as ^{15}N isotope labeled D-asparagine could not be

obtained. L-glutamate and no addition of amino acid were included as controls. LC-MS characterization of fermentation extracts showed an increased incorporation of ^{14}N in BE-14106 when glycine was added, thus indicating that the nitrogen in BE-14106 originates from the amino group of glycine. A second feeding experiment was performed with ^{13}C isotope labeled medium components and feeding with ^{12}C-containing amino acids to determine if the complete glycine molecule is incorporated into BE-14106. There was no enrichment of ^{12}C in BE-14106 when feeding with glycine compared to the control, thus demonstrating that only the amino group from glycine is incorporated into BE-14106.

4.3.5 Proposed biosynthetic pathway for BE-14106 in *Streptomyces* sp. DSM 21069

Based on the results from the gene inactivation experiments and enzyme assays, several hypothetical pathways for synthesis of the aminoacyl starter were evaluated that involved incorporation of glycine, however, no pathway could be constructed that resulted in the synthesis of the correct starter. Instead, a scheme was proposed involving the truncated PKS BecC and recruitment of the glycine amino group. Inactivation of *becC* abolished the production of BE-14106 in *Streptomyces* sp. DSM 21069, thus confirming a role for BecC in the biosynthesis. Subsequent feeding experiments showed that only the nitrogen atom, and not the carbons, from glycine is incorporated into BE-14106, supporting the hypothesis that the glycine amino group is recruited to generate the starter unit. The proposed biosynthetic pathway for BE-14106 is presented in Figure 4. 5. The pathway is thought to start with the synthesis of the acyl chain by BecA and BecC. The third module of BecA is truncated and lacks the ACP domain necessary for completion of the third extension. BecC, which constitutes a KR and an ACP domain, is presumed to supply the terminal ACP. The resulting acyl chain is assumed to be further modified while still tethered to the ACP domain of BecC. The discrete adenylation domain, BecL, presumably activates glycine through adenylation and subsequent loading on the discrete PCP domain, BecS. The BecU protein of unknown function may mediate the interaction between the PCP domain and the BecC ACP domain. Putative glycine oxidase, BecI, is thought to catalyze oxidative deamination of glycine, releasing

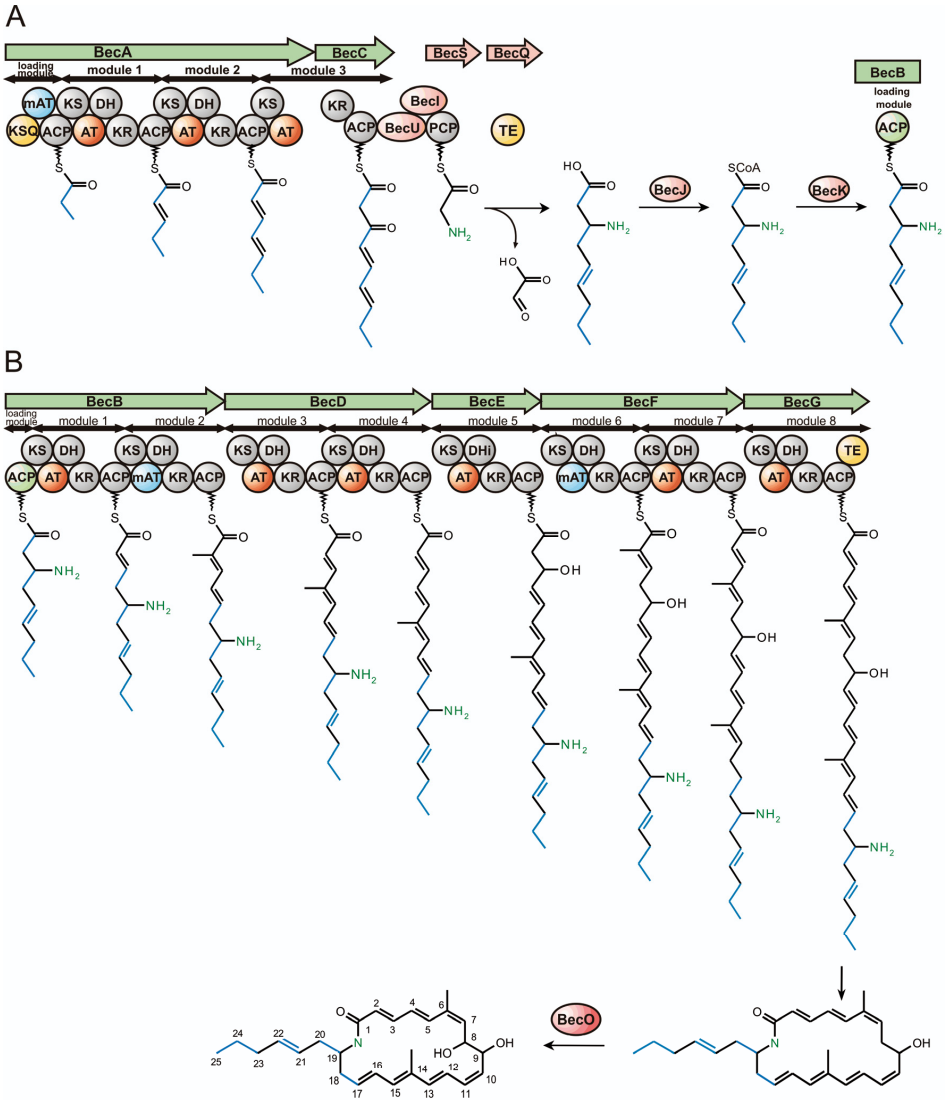


Figure 4. 5 Proposed pathway for the biosynthesis of BE-14106. A) Synthesis of the aminoacyl starter. B) Synthesis of the macrolactam ring. Certain domains are indicated by use of color: malonyl-CoA specific AT domains: orange, methylmalonyl-CoA specific AT domains: blue, loading module ACP domain: green, KS^Q and TE domains: yellow.

ammonium, which then performs a nucleophilic attack on the acyl C-19 carbonyl, forming a C-19 imino group. The deaminated glycine residue is assumed to be released as glyoxylate. Reduction of the C-19 imino to an amino group may cause the migration/elimination of the acyl double bonds. The aminoacyl is presumably released by the thioesterase type II, BecQ. Putative acyl-CoA synthetase/ligase BecJ presumably activates the aminoacyl through ligation with CoA, making the starter an acceptable substrate for loading on BecB by the discrete acyltransferase BecK. After loading of the aminoacyl starter on the first ACP domain of BecB, the polyketide synthesis is assumed to proceed through all modules of the BecB-BecG PKS enzymes with subsequent release and cyclization of the macrolactam by the terminal TE domain of BecG. The putative L-amino acid amidase/proline iminopeptidase BecP may aid in this process, but no experimental results have been obtained that can verify this role. The resulting deoxy-BE-14106 is hydroxylated by the P450 monooxygenase BecO, presumably at the C-8 position. The biosynthesis of BE-14106 is thought to be regulated by the two regulators BecH and BecM and the efflux/resistance mechanism is presumed to involve the MFS-type efflux pump BecN.

4.4 Insights into the evolution of macrolactam biosynthesis: cloning and comparative analysis of the biosynthetic gene cluster for a novel macrocyclic lactam ML-449

(Paper III)

The antifungal screen described in section 4.1 resulted in detection of another “hit” compound exhibiting antifungal activity. Fractionation and LC-MS-TOF analysis of fermentation extracts from the producer *Streptomyces* sp. MP39-85, revealed a putative active compound with the accurate mass of 449. The compound subsequently dubbed ML-449 (ML = macrolactam), was subjected to NMR spectroscopy for structure elucidation and found to be remarkably similar to BE-14106, differing only in the length of the acyl side chain (Figure 4. 6).

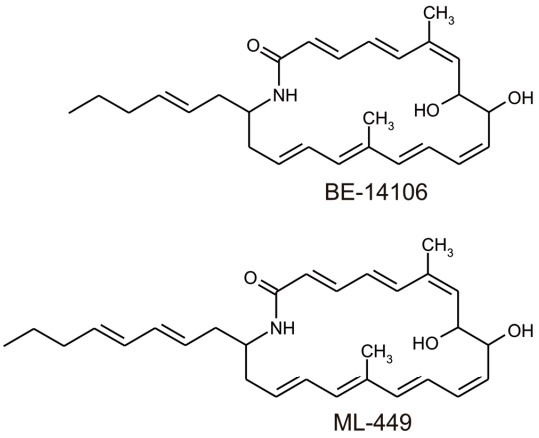


Figure 4. 6 The structures of BE-14106 and ML-449.

MIC50 was determined for ML-449, but the compound was found to be less active against *Candida* than BE-14106. The ML-449 biosynthetic gene cluster (*mla*) was identified and sequenced following the same procedure as for the BE-14106 gene cluster (*bec*).

4.4.1 Comparison of the ML-449 and the BE-14106 biosynthetic gene clusters

Annotation of the sequenced gene cluster revealed a high similarity with the BE-14106 biosynthetic gene cluster described in section 4.3. Homologs of all genes present in the *bec* cluster were identified in the *mla* cluster and the organization of the two gene clusters was found to be identical (Table 4. 2 and Figure 4. 7). Similarity on both protein and nucleotide level was in the range of 80-90 % for most of the *orfs*.

Table 4. 2 Description of proteins encoded by the *mla* cluster and corresponding homologs from the *bec* cluster.

Protein	Size (aa)	Homolog from BE-14106 cluster	% positives/ % identity	Proposed function
MlaH	951	BecH	94/88	LuxR-type transcriptional regulator
MlaA1	6383	BecA	91/86	Polyketide synthase type I
MlaA2	1043	-	-	Polyketide synthase type I
MlaI	363	BecI	89/79	Glycine oxidase/FAD-dependent oxidoreductase
MlaC	695	BecC	90/84	Polyketide synthase type I
MlaU	187	BecU	95/88	Putative NRPS accessory protein
MlaB	3530	BecB	92/87	Polyketide synthase type I
MlaJ	532	BecJ	91/86	AMP-dependent acyl-CoA synthetase/ligase
MlaK	313	BecK	89/83	Acyltransferase
MlaS	78	BecS	94/88	Peptidyl carrier protein
MlaL	504	BecL	90/83	NRPS adenylation domain
MlaM	198	BecM	89/86	TetR-type transcriptional regulator
MlaN	523	BecN	94/87	MFS-type efflux pump
MlaO	411	BecO	97/92	P450 monooxygenase
MlaD	3365	BecD	92/88	Polyketide synthase type I
MlaP	311	BecP	91/84	Putative L-amino acid amidase/proline iminopeptidase
MlaG	1992	BecG	90/85	Polyketide synthase type I
MlaF	3373	BecF	93/88	Polyketide synthase type I
MlaE	1637	BecE	92/87	Polyketide synthase type I
MlaT	88	BecT	82/74	Hypothetical protein, SimX2-like protein
MlaQ	256	BecQ	93/86	Thioesterase type II

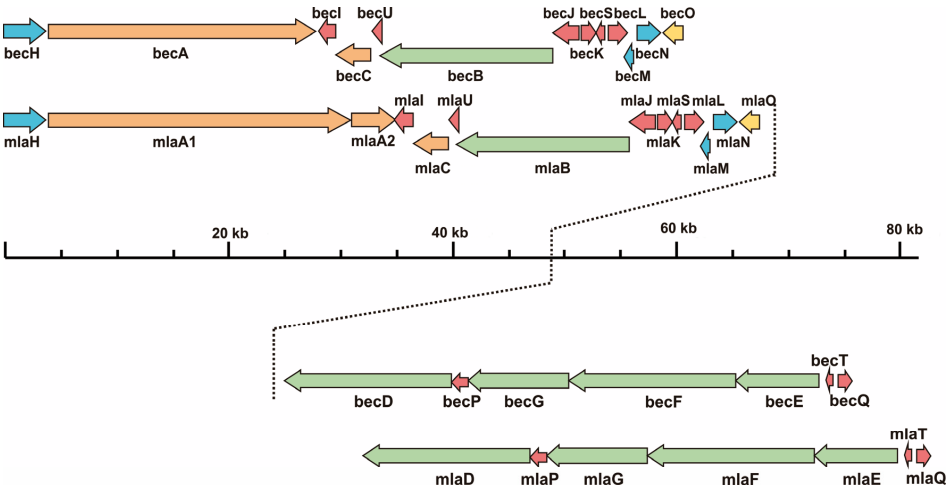


Figure 4. 7 Comparison of the *bec* and *mla* clusters. Genes encoding PKS enzymes involved in synthesis of the acyl side chain is shown in orange, remaining PKS genes in green. Genes encoding regulators and an efflux pump are shown in blue. The P450 monooxygenase-encoding gene is shown in yellow, while genes encoding enzymes presumed to be involved in activation/modification of the aminoacyl starter are shown in red, as are the cryptic *bec/mlaT* and the iminopeptidase-encoding *bec/mlaP*.

The only difference between the two clusters was represented by the *mlaA1* and *mlaA2* genes. MlaA1 appeared to be the complete version of BecA, with a full third module, while MlaA2 is truncated in the same way as BecA and lacks the terminal ACP domain. All in all, MlaA1, MlaA2 and MlaC constitute four complete modules, while BecA and BecC represents three modules, and thus the former presumably incorporates one extra acetate unit into the acyl chain. The elucidated structure of ML-449 is in agreement with this proposal. All other aspects of the ML-449 biosynthesis are assumed to be in correspondence with the proposed pathway for the biosynthesis of BE-14106 (described in section 4.3.5). To confirm the involvement of the sequenced gene cluster in the ML-449 biosynthesis, a gene inactivation experiment was performed using a DNA fragment covering the KR2-, ACP2-, KS3- and AT3-encoding part of *mlaA1*. The mutant was found to produce less than 1 % of wildtype level, thus confirming the involvement of the gene cluster in the biosynthesis of ML-449.

4.4.2 Evolutionary analysis

A phylogenetic analysis was undertaken for the KS domains of the PKS enzymes encoded in the two gene clusters. For comparison, KS domains from 9 other characterized PKS clusters were included in the analysis, including the biosynthetic gene clusters for the structurally related macrolactams vicenistatin and salinilactam. The phylogenetic tree generated for the 151 KS domains demonstrated a split between the KS domains involved in the synthesis of the acyl side chain and the rest of the KS domains from the *mla* and *bec* clusters (Figure 4. 8). KS domains belonging to one particular gene cluster usually form a separate group, indicating that the individual modules were generated by duplication of a single ancestor module (Lopez, 2003; Ginolhac et al., 2005; Jenke-Kodama et al., 2006). The splitting of the *mla*/*bec* KS domains into two separate groups thus indicates a different origin for the *mlaA1*/*becA* and *mlaA2* genes than the remaining PKS genes in the cluster. As to the process by which *mlaA1*/*becA* and *mlaA2* ended up in the *mla/bec* cluster, one can only speculate, but a recombinatorial event involving a different PKS gene cluster could be a plausible explanation. All Mla/BecB-G KS domains (except KS1 from Mla/BecB) formed a subclade together with all KS domains from the vicenistatin biosynthetic gene cluster (except KS1 from VinP1) and some KS domains from the salinilactam biosynthetic gene cluster, indicating a common origin for at least some of the PKS genes from these macrolactam biosynthetic gene clusters. KS1 domains from Mla/BecB, VinP1 and Strop_2768 formed a separate group from the other KS domains included in the study. Earlier studies have shown that KS domains accepting amino acid derived substrates usually form a separate clade and do not group with the KS domains from their respective gene clusters (Moffitt and Neilan, 2003; Ginolhac et al., 2005). As KS1 from Mla/BecB, VinP1 and Strop_2768 presumably accept amino acid derived/aminoacyl substrates, the formation of a separate group by these KS domains is consistent with the above-mentioned observation. All KS^{Q} domains formed a separate group and this is consistent with earlier studies (Moffitt and Neilan, 2003; Ginolhac et al., 2005). The analysis also revealed that the KS domains from the *mla* cluster were always direct

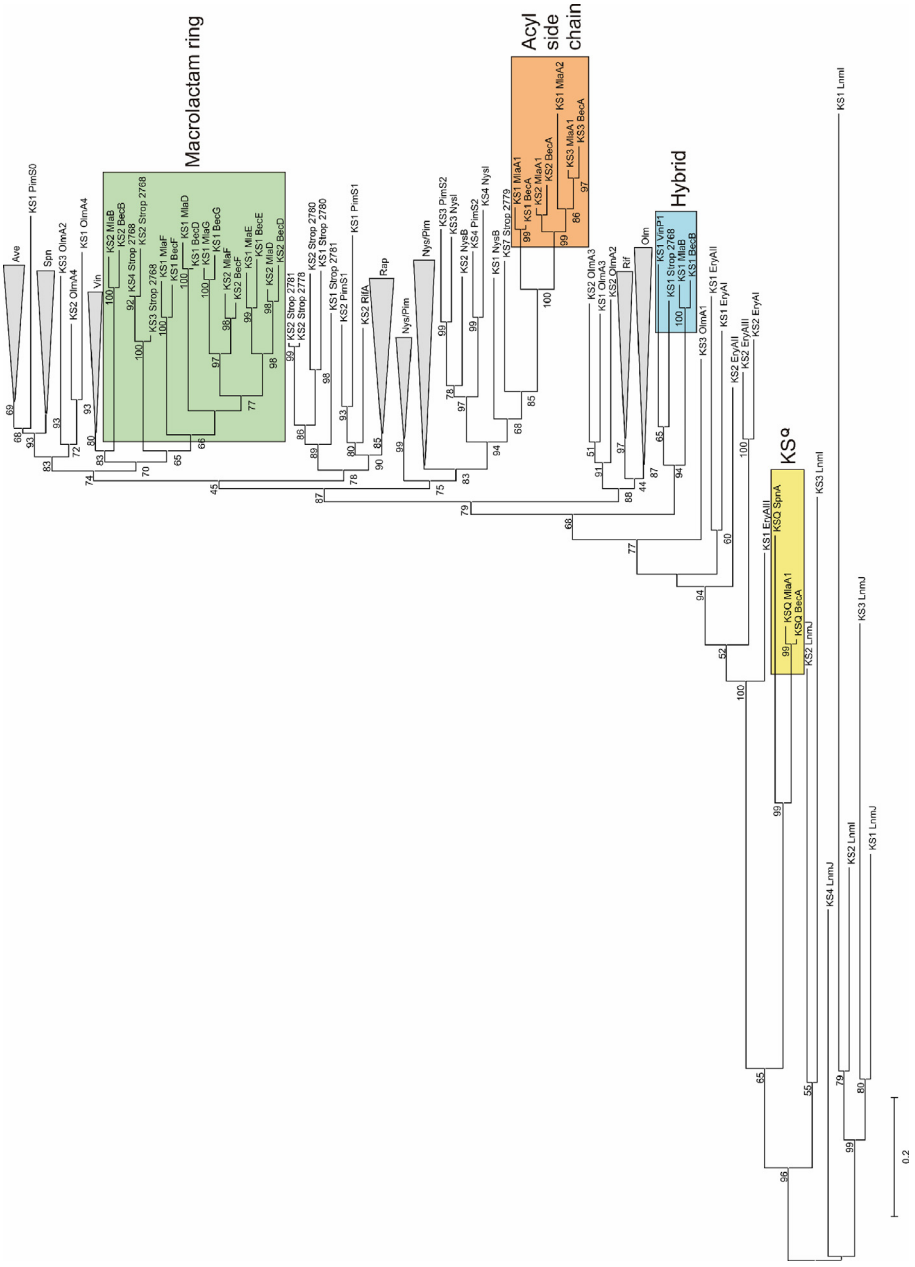


Figure 4. 8 Phylogenetic analysis of ketosynthase (KS) domains. The tree was reconstructed using maximum likelihood. aLRT statistic values are indicated at each node. The scale bar indicates 0.2 substitutions per amino acid position. The tree is unrooted. Green box: KS domains involved in macrolactam ring synthesis; orange box: KS domains involved in acyl side chain synthesis; blue box: KS domains accepting amino acid/aminoacyl; yellow box: KS^Q domains.

neighbors of their *bec* homologs from the same module, and this strongly suggests a common origin for the two gene clusters.

4.4.3 Enzymes presumed to be involved in activation/modification of the aminoacyl starter

In the central region of the *mla*/*bec* clusters is a small subcluster of four genes and similar subclusters of putative homologs can be found in the *vin* and *slm* clusters, although the organization of the genes is different (Figure 4. 9 A). The four genes presumably encode a discrete NRPS adenylation domain (*mla/becL*), a discrete PCP domain (*mla/becS*), an AMP-dependent acyl-CoA synthetase/ligase (*mla/becJ*) and a discrete acyltransferase (*mla/becK*). In contrast to the *mla/bec*, *vin* and *slm* clusters, the *lnm* cluster contains only two of the putative homologs.

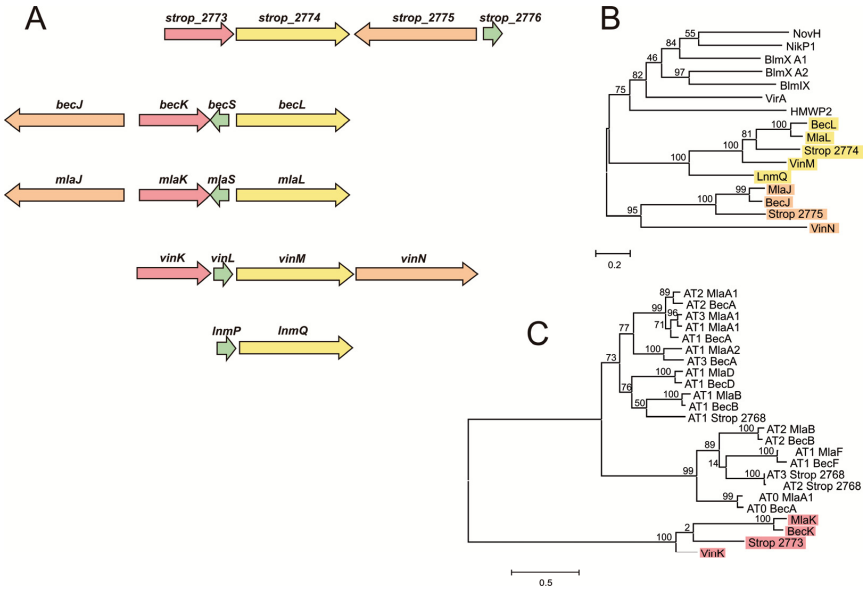


Figure 4. 9 A) The subclusters of putative homologs from the *mla*, *bec*, *slm*, *vin* and *lnm* clusters. Potential homologs are indicated by use of the same color: encoding putative NRPS adenylation domains: yellow, AMP-dependent synthetases/ligases: orange, acyltransferases: red, PCP domains: green. Phylogenetic trees for adenylation domains/AMP-dependent acyl-CoA synthetases/ligases (B) and acyltransferases (C) are shown on the right. Phylogenetic trees were generated using PhyML (maximum likelihood). aLRT statistic values are indicated at each node. Trees are unrooted.

To establish the evolutionary relationship between these potential homologs, phylogenetic trees were reconstructed for the putative NRPS adenylation (A) domains, AMP-dependent acyl-CoA synthetases/ligases (Figure 4. 9 B) and acyltransferases (AT) (Figure 4. 9 C). The NRPS A domains and AMP-dependent acyl-CoA synthetases/ligases were included in the same tree reconstruction as they all contain the 10 core motifs of the adenylate-forming superfamily of enzymes and are presumed to represent homologs. The analysis showed that the discrete A domains and AMP-dependent acyl-CoA synthetases/ligases both formed separate groups distinct from the other A domains included in the study, thus indicating a closer relationship between these enzymes than other A domains. The same phenomenon was observed for the discrete ATs, as they formed a subclade separate from the other AT domains included in the analysis.

In addition to these four genes, there are several other putative homologs present in the *vin* and *slm* clusters. The *slm* cluster contains a total of 9 potential homologs (not counting PKS genes) of genes from the *mla/bec* clusters, although the overall organization of the *slm* cluster is quite different from the *mla/bec* clusters. The phylogenetic analyses of the KS, A and AT domains suggest common ancestry for these four gene clusters, however, several rearrangements of the clusters as well as deletions/insertions must be postulated if all should stem from the same ancestor cluster. Earlier studies have shown that individual modules and domains in PKS and NRPS genes may have been exchanged through a natural type of biocombinatorics (Jenke-Kodama et al., 2006; Fewer et al., 2007). In that respect, it does not seem unlikely that also other genes in such gene clusters are exchangeable and that new gene clusters for secondary metabolites are generated through recombinational events with other clusters or within the same cluster, loss or acquisition of genes as well as gene duplications.

5 Concluding remarks

The work presented in this thesis has focused on discovering new compounds with antifungal or cytotoxic activity, genes for their biosynthesis and elucidation of biosynthetic pathways. In addition, potential mechanisms behind distribution and evolution of these biosynthetic gene clusters have been suggested.

Screening of more than 4000 actinomycete isolates from sediment and neuston layer samples, revealed a high number of putative heptaene producers and LC-MS-TOF analysis identified the heptaene in question as the polyene macrolide candicidin. A gene cluster for the biosynthesis of candicidin was located on a linear plasmid, suggesting involvement of horizontal gene transfer in the dissemination of the gene cluster among actinomycetes in marine sediments and the neuston layer. The plasmid was, however, only discovered in one isolate and more work remains to be done before a full understanding of the phenomenon can be obtained. Further study involving a larger number of isolates as well as full sequencing of selected genes from the *can* cluster could potentially provide a more defined conclusion.

Following the screening of actinomycete isolates, several compounds exhibiting both antifungal and cytotoxic activity were identified, including the previously described macrolactam antibiotic BE-14106. Cloning and sequencing of the BE-14106 biosynthetic gene cluster (*bec*) has been accomplished and revealed a rather unusual cluster consisting of genes encoding both PKS and NRPS-related enzymes. Two of the genes were found to encode truncated PKSs lacking some of the core domains usually present in such enzymes. The proposed pathway for the production of BE-14106 involves synthesis of the acyl side chain by a separate PKS system and recruitment of an amino group from the amino acid glycine to generate the aminoacyl starter unit for the macrolactam ring synthesis. Such a mechanism for synthesizing a starter unit for macrolactam biosynthesis has to our knowledge never been described before. Although a defined role for most of the enzymes encoded in the gene cluster could be suggested, experimental work to verify some of these functions remains to be carried out. In that

respect, the role of the putative L-amino acid amidase/proline iminopeptidase BecP may be the most puzzling and in dire need of verification.

Sequencing of the biosynthetic gene cluster for a second macrolactam, ML-449, revealed a gene cluster with high similarity to the *bec* cluster. The two compounds, ML-449 and BE-14106, were shown to be structurally similar, only differing in the length of the acyl side chain. The structural differences appear to be due to the incorporation of one extra acetate unit during the synthesis of the ML-449 acyl side chain as the PKS system performing this synthesis contains one extra module compared to the enzymes encoded by the *bec* cluster. Phylogenetic analyses of KS, A and AT domains involved in the biosynthesis of the two compounds pointed to common ancestry for the two clusters as well as an evolutionary relationship with the vicenistatin and salinilactam biosynthetic gene clusters.

The sequencing of the BE-14106 and ML-449 biosynthetic gene clusters provides an important starting point for biosynthetic engineering of these compounds into a possible anticancer drug candidate. Derivatives with improved activity or better pharmacological properties can potentially be generated through genetic manipulation of the genes found within these clusters. In that respect, the elucidation of the biosynthetic pathways provides vitally important information for predicting the structural changes resulting from such genetic engineering. In addition, the work has provided more knowledge about macrolactam biosynthesis in general as well as insight into the underlying processes governing the evolution of macrolactam biosynthetic gene clusters.

6 References

Admiraal, S. J., Walsh, C. T., and Khosla, C. (2001) The loading module of rifamycin synthetase is an adenylation-thiolation didomain with substrate tolerance for substituted benzoates. Biochemistry. 40, 6116-6123.

Amann, R. I., Ludwig, W., and Schleifer, K.-H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev. 59, 143–169.

Amyes, S. G. B. (2001) Magic bullets lost horizons. The rise and fall of antibiotics. 1st ed. Taylor & Francis, London.

Baltz, R. H. (2006) Molecular engineering approaches to peptide, polyketide and other antibiotics. Nat. Biotechnol. 24, 1533-1540.

Beck, B. J., Yoon, Y. J., Reynolds, K. A., and Sherman, D. H. (2002) The hidden steps of domain skipping: macrolactone ring size determination in the pikromycin modular polyketide synthase. Chem. Biol. 9, 575-583.

Bender, C. L., Alarcón-Chaidez, F., and Gross, D. C. (1999) *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol. Mol. Biol. Rev. 63, 266–292.

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, 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, 141-147.

Bérdy, J. (2005) Bioactive microbial metabolites. J. Antibiot. (Tokyo). 58, 1-26.

Betlach, M. C., Kealey, J. T., Betlach, M. C., Ashley, G. W., and McDaniel, R. (1998) Characterization of the macrolide P-450 hydroxylase from *Streptomyces venezuelae* which converts narbomycin to picromycin. Biochemistry. 37, 14937–14942.

Bisang, C., Long, P. F., Cortés, 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, 502-505.

Borgos, S. E. F., Tsan, P., Sletta, H., Ellingsen, T. E., Lancelin, J.-M., and Zotchev, S. B. (2006) Probing the structure-function relationship of polyene macrolides: engineered biosynthesis of soluble nystatin analogues. J. Med. Chem. 49, 2431-2439.

Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards Jr., J. E., Gilbert, D., Rice, L. B. Scheld, M., Spellberg, B., and Bartlett, J. (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. Clin. Infect. Dis. 48, 1–12.

Bouvier, T., and del Giorgio, P.A. (2007) Key role of selective viralinduced mortality in determining marine bacterial community composition. Environ. Microbiol. 9, 287–297.

Brautaset, T., Sekurova, O. N., Sletta, H., Ellingsen, T. E., Strøm, A. R., Valla, S., and Zotchev, S. B. (2000) Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. Chem. Biol. 7, 395-403.

Bruns, A., Hoffelner, H., and Overmann, J. (2003) A novel approach for high throughput cultivation assays and the isolation of planktonic bacteria. FEMS Microbiol. Ecol. 45, 161-171.

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

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

Byrne, B., Carmody, M., Gibson, E., Rawlings, B., and Caffrey, P. (2003) Biosynthesis of deoxyamphotericins and deoxyamphoteronolides by engineered strains of *Streptomyces nodosus*. Chem. Biol. 10, 1215-1224.

Caffrey, P., Aparicio, J. F., Malpartida, F., and Zotchev, S. B. (2008) Biosynthetic engineering of polyene macrolides towards generation of improved antifungal and antiparasitic agents. Curr. Top. Med. Chem. 8, 639-653.

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

Campelo, A. B., and Gil, J. A. (2002) The candicidin gene cluster from *Streptomyces griseus* IMRU 3570. Microbiol. 148, 51-59.

Carvalho, R., Reid, R., Viswanathan, N., Gramajo, H., and Julien, B. (2005) The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation. Gene. 359, 91-98.

Casey, J. T., O'Cleirigh, C., Walsh, P. K., and O'Shea, D. G. (2004) Development of a robust microtiter plate-based assay method for assessment of bioactivity. J. Microbiol. Methods. 58, 327–334.

Chater, K. F., and Kinashi, H. (2007). *Streptomyces* linear plasmids: their discovery, functions, interactions with other replicons, and evolutionary significance, pp. 1–31. *In* F. Meinhardt and R. Klassen (eds.) Microbial linear plasmids. Springer-Verlag, Berlin, Germany.

Chen, C. W., Huang, C.-H., Lee, H.-H., Tsai, H.-H., and Kirby, R. (2002) Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. Trends Genet. 18, 522-529.

Chen, J., and Stubbe, J. (2004) Bleomycins: new methods will allow reinvestigation of old issues. Curr. Opin. Chem. Biol. 8, 175-181.

Chen, S. C. A., and Sorrell, T. C. (2007) Antifungal agents. Med. J. Aust. 187, 404-409.

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. USA. 100, 3149-3154.

Choulet, F., Aigle, B., Gallois, A., Mangenot, S., Gerbaud, C., Truong, C., Francou, F.-X., Fourrier, C., Guérineau, 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, 2361-2369.

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, 176-178.

Dagan, T., Artzy-Randrup, Y., and Martin, W. (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. Proc. Natl. Acad. Sci. USA. 105, 10039-10044.

DiRuggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F. T., Boos, W., and Weiss, R. B. (2000) Evidence of recent lateral gene transfer among hyperthermophilic Archaea. Mol. Microbiol. 38, 684-693.

Donadio, S., McAlpine, J. B., Sheldon, P. J., Jackson, M., and Katz, L. (1993) An erythromycin analog produced by reprogramming of polyketide synthesis. Proc. Natl. Acad. Sci. USA. 90, 7119-7123.

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

Doolittle, W. F., and Bapteste, E. (2007) Pattern pluralism and the Tree of Life hypothesis. Proc. Natl. Acad. Sci. USA. 104, 2043-2049.

Doroshow, J. H. (2006) Anthracyclines and anthracenediones, pp. 414-450. *In* Chabner, B. A and Longo, D. L. (eds.) Cancer chemotherapy & biotherapy. Principles and practice. 4th ed., Lippincott Williams & Wilkins, Philadelphia, PA.

Du, L., Chen, M., Sánchez, C., and Shen, B. (2000) An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *Streptomyces verticillus* ATCC15003. FEMS Microbiol. Lett. 189, 171-175.

Du, L., Sánchez, C., and Shen, B. (2001) Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. Metabol. Eng. 3, 78-95.

Egan, S., Wiener, P., Kallifidas, D., and Wellington, E. M. H. (1998) Transfer of streptomycin biosynthesis gene clusters within streptomycetes isolated from soil. Appl. Environ. Microbiol. 64, 5061-5063.

Egan, S., Wiener, P., Kallifidas, D., and Wellington, E. M. H. (2001) Phylogeny of *Streptomyces* species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. Antonie Van Leeuwenhoek. 79, 127-133.

Eppelmann, K., Stachelhaus, T., and Marahiel, M. A. (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. Biochemistry. 41, 9718-9726.

Feher, M., and Schmidt, J. M. (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. J. Chem. Inf. Comput. Sci. 43, 218-227.

Fenical, W., and Jensen, P. R. (2006) Developing a new resource for drug discovery: marine actinomycete bacteria. Nat. Chem. Biol. 2, 666-673.

Fewer, D. P., Rouhiainen, L., Jokela, J., Wahlsten, M., Laakso, K., Wang, H., and Sivonen, K. (2007) Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. BMC Evol. Biol. 7, 183.

Finch, R., and Hunter, P. A. (2006) Antibiotic resistance-action to promote new technologies: report of an EU Intergovernmental Conference held in Birmingham, UK, 12-13 December 2005. J. Antimicrob Chemother. 58, Suppl. S1, i3-i22.

Fischbach, M. A. and Walsh, C. T. (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. 106, 3468-3496.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocyne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. (1995) Whole-Genome Random Sequencing and Assembly of *Haemophilus Influenzae* Rd. Science. 269, 496-512.

Fleming, A. (1929) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. Br. J. Exp. Path. 10, 226-236.

Fujii, I., and Ebizuka, Y. (1997) Anthracycline biosynthesis in *Streptomyces galilaeus*. Chem. Rev. 97, 2511-2523.

Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blöcker, H., Höfle, G., and Müller, R. (2002) The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. J. Biol. Chem. 277, 13082-13090.

Gehring, A. M., Mori, I., Perry, R. D., and Walsh, C. T. (1998) The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. Biochemistry. 37, 11637-11650.

Gich, F., Schubert, K., Bruns, A., Hoffelner, H., and Overmann, J. (2005) Specific detection, isolation, and characterization of selected, previously uncultured members of the freshwater bacterioplankton community. Appl. Environ. Microbiol. 71, 5908-5919.

Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., Clardy, J., Goodman, R. M., and Handelsman, J. (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl. Environ. Microbiol. 68, 4301-4306.

Ginolhac, A., Jarrin, C., Robe, P., Perrière, G., Vogel, T. M., Simonet, P., and Nalin, R. (2005) Type I polyketide synthases may have evolved through horizontal gene transfer. J. Mol. Evol. 60, 716-725.

Grabowski, K., and Schneider, G. (2007) Properties and architecture of drugs and natural products revisited. Curr. Chem. Biol. 1, 115–127.

Gravius, B., Glocker, D. Pigac, J. Pandža, K. Hranueli, D., and Cullum, J. (1994) The 387 kb linear plasmid pPZG101 of *Streptomyces rimosus* and its interactions with the chromosome. Microbiol. 140, 2271-2277.

Grdičak, M., and Jerić, I. (2007) Enediyne compounds-new promises in anticancer therapy. Acta Pharm. 57, 133-150.

Griffith, F. (1928) The significance of pneumococcal types. J. Hyg. 27, 113–159.

Gupte, M., Kulkarni P., and Ganguli, B. N. (2002) Antifungal antibiotics. Appl. Microbiol. Biotechnol. 58, 46–57.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. 5, R245-R249.

Harris, M. (2004) Monoclonal antibodies as therapeutic agents for cancer. Lancet Oncol. 5, 292–302.

Harvey, A. L. (2008) Natural products in drug discovery. Drug Discov. Today. 13, 894-901.

Hayakawa, T., Tanaka, T., Sakaguchi, K., Õtake, N., and Yonehara, H. (1979) A linear plasmid-like DNA in *Streptomyces* sp. producing lankacidin group antibiotics. J. Gen. Appl. Microbiol. 25**,** 255-260.

Haydock, S. F., Aparicio, J. F., Molnár, I., Schwecke, T., Khaw, L. E., König, A., Marsden, A. F. A., Galloway, I. S., Staunton, J., and Leadlay, P. F. (1995) Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. FEBS Lett. 374, 246-248.

Haydock, S. F., Appleyard, A. N., Mironenko, T., Lester, J., Scott, N., and Leadlay, P. F. (2005) Organization of the biosynthetic gene cluster for the macrolide concanamycin A in *Streptomyces neyagawaensis* ATCC 27449. Microbiol. 151, 3161-3169.

He, J., and Hertweck, C. (2003) Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. Chem. Biol. 10, 1225-1232.

Hopwood, D. A. (1967) Genetic analysis and genome structure in *Streptomyces coelicolor*. Bacteriol. Rev. 31, 373–403.

Hopwood, D. A. (1999) Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. Microbiol. 145, 2183-2202.

Hopwood, D. A. (2006) Soil to genomics: the *Streptomyces* chromosome. Annu. Rev. Genet. 40, 1-23.

Hopwood, D. A. (2007) *Streptomyces* in nature and medicine, the antibiotic makers. 1st ed., Oxford University Press, Inc., New York, NY.

Hopwood, D. A., Harold, R. J., Vivian, A., and Ferguson, H. M. (1969) A new kind of fertility variant in *Streptomyces coelicolor*. Genetics 62, 461–477.

Hopwood, D. A., and Kieser, T. (1993) Conjugative plasmids of *Streptomyces*, pp. 293-311. *In* D. B. Clewell (ed.) Bacterial Conjugation, 1st ed. Plenum Press, New York, NY.

Hopwood, D. A., Kieser, T., Wright, H. M., and Bibb, M. J. (1983) Plasmids, recombination and chromosome mapping in *Streptomyces lividans* 66. J. Gen. Microbiol. 129, 2257-2269.

Hopwood, D. A., Malpartida, F., Kieser, H. M., Ikeda, H., Duncan, J., Fujii, I., Rudd, B. A. M., Floss, H. G., and Õmura, S. (1985) Production of 'hybrid' antibiotics by genetic engineering. Nature. 314, 642-644.

Hopwood, D. A., and Wright, H.M. (1973) A plasmid of *Streptomyces coelicolor* carrying a chromosomal locus and its inter-specific transfer. J. Gen. Microbiol. 79, 331–342.

Hopwood, D. A., and Wright, H. M. (1976) Genetic studies on SCP1-prime strains of *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. 95, 107–120.

Hotchkiss, R. D., and Dubos, R. J. (1940) Fractionation of the bactericidal agent from cultures of a soil bacillus. J. Biol. Chem. 132, 791-792.

Hubbard, B. K., and Walsh, C. T. (2003) Vancomycin assembly: nature's way. Angew. Chem. Int. Ed. 42, 730-765.

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

Jenke-Kodama, H., Börner, T., and Dittman, E. (2006) Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. PLoS Comput. Biol. 2, e132.

Jensen, P. R. and Fenical, W. (2005) New natural-product diversity from marine actinomycetes. *In* Zhang, L. and Demain, A. L. (eds.) Natural products: drug discovery and therapeutic medicine, 1st ed., Humana Press, Totowa, NJ.

Jensen, P. R., Mincer, T. J., Williams, P. G., and Fenical, W. (2005) Marine actinomycete diversity and natural product discovery. Antonie van Leeuwenhoek. 87, 43–48.

Jia, X.-Y., Tian, Z.-H., Shao, L., Qu, X.-D., Zhao, Q.-F., Tang, J., Tang, G.-L., and Liu, W. (2006) Genetic characterization of the chlorothricin gene cluster as a model for spirotetronate antibiotic biosynthesis. Chem. Biol. 13, 575-585.

Jones, R. N. (2008) Key considerations in the treatment of complicated staphylococcal infections. Clin. Microbiol. Infect. 14, Suppl. 2, 3-9.

Joseph, S. J., Hugenholtz, P., Sangwan, P., Osborne, C. A., and Janssen, P. H. (2003) Laboratory cultivation of widespread and previously uncultured soil bacteria. Appl. Environ. Microbiol. 69, 7210–7215.

Keating, T. A., Suo, Z., Ehmann, D. E., and Walsh, C. T. (2000) Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain. Biochemistry. 39, 2297-2306.

Khosla, C., and Harbury, P. B. (2001) Modular enzymes. Nature. 409, 247-252.

Kinashi, H., Otten, S. L., Duncan, J. S., and Hutchinson, C. R. (1988) Frequent loss and restoration of antibiotic production by *Streptomyces lasaliensis*. J. Antibiotics (Tokyo). 41, 624-637.

Kinashi, H., Shimaji, M., and Sakai, A. (1987) Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. Nature. 328, 454-456.

Kinashi, H., Shimaji-Murayama, M., and Hanafusa, T. (1992) Integration of SCP1, a giant linear plasmid, into the *Streptomyces coelicolor* chromosome. Gene. 115, 35-41.

Kirby, R., Wright, L. F., and Hopwood, D. A. (1975) Plasmid-determined antibiotic synthesis and resistance in *Streptomyces coelicolor*. Nature. 254, 265–267.

Kleinkauf, H., and von Döhren, H. (1990) Nonribosomal biosynthesis of peptide antibiotics. Eur. J. Biochem. 192, 1-15.

Koehn, F. E. (2008) High impact technologies for natural products screening, pp. 177-210. *In* Petersen, F. and Amstutz, R. (eds.) Progress in drug research, vol. 65, Birkhäuser, Stuttgart.

Koonin, E. V., and Wolf, Y. I. (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36, 6688-6719.

Kopp, M., Irschik, H., Pradella, S., and Müller, R. (2005) Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes. ChemBioChem, 6, 1277-1286.

Kwon, H. C., Kauffman, C. A., Jensen, P. R., and Fenical, W. (2006) Marinomycins A-D, antitumor-antibiotics of a new structure class from a marine actinomycete of the recently discovered genus "*Marinispora*". J. Am. Chem. Soc. 128, 1622-1632.

Lang, A. S., and Beatty, J. T. (2007) Importance of widespread gene transfer agent genes in α-proteobacteria. Trends Microbiol. 15, 54-62.

Lazo, J. S., and Chabner, B. A. (2006) Bleomycin, pp. 344-358. *In* Chabner, B. A and Longo, D. L. (eds.) Cancer chemotherapy & biotherapy. Principles and practice. 4th ed., Lippincott Williams & Wilkins, Philadelphia, PA.

Leadlay, P. F., Staunton, J., Oliynyk, M., Bisang, C., Cortés, J., Frost, E., Hughes-Thomas, Z. A., Jones, M. A., Kendrew, S. G., Lester, J. B., Long, P. F., McArthur, H. A. I., McCormick, E. L., Oliynyk, Z., Stark, C. B. W., and Wilkinson, C. J. (2001) Engineering of complex polyketide biosynthesis--insights from sequencing of the monensin biosynthetic gene cluster. J. Ind. Microbiol. Biotechnol. 27, 360-367.

Lechevalier, H., Acker, R. F., Corke, C. T., Haenseler, C. M., and Waksman, S. A. (1953) Candicidin, a new antifungal antibiotic. Mycologia. 45, 155-171.

Letscher-Bru, V., and Herbrecht, R. (2003) Caspofungin: the first representative of a new antifungal class. J. Antimicrob. Chemother. 51, 513-521.

Long, P.F., Wilkinson, C. J., Bisang, C. P., Cortés, 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, 1215–1225.

Lopez, J. V. (2003) Naturally mosaic operons for secondary metabolite biosynthesis: variability and putative horizontal transfer of discrete catalytic domains of the epothilone polyketide synthase locus. Mol. Genet. Genomics. 270, 420-431.

Losey, H. C., Peczuh, M. W., Chen, Z., Eggert, U. S., Dong, S. D., Pelczer, I., Kahne, D., and Walsh, C. T. (2001) Tandem action of glycosyltransferases in the maturation of vancomycin and teicoplanin aglycones: novel glycopeptides. Biochemistry. 40, 4745-4755.

Magarvey, N. A., Keller, J. M., Bernan, V., Dworkin, M., and Sherman, D. H. (2004) Isolation and characterization of novel marine-derived actinomycete taxa rich in bioactive metabolites. Appl. Environ. Microbiol. 70, 7520–7529.

Maldonado, L. A., Stach, J. E. M., Pathom-aree, W., Ward, A. C., Bull, A. T., and Goodfellow, M. (2005) Diversity of cultivable actinobacteria in geographically widespread marine sediments. Antonie Van Leeuwenhoek. 87, 11-18.

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

Martín, J. F. (1998) New aspects of genes and enzymes for β-lactam antibiotic biosynthesis. Appl. Microbiol. Biotechnol. 50, 1-15.

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, 493-496.

Menzella, H. G., Carney, J. R., and Santi, D. V. (2007) Rational design and assembly of synthetic trimodular polyketide synthases. Chem. Biol. 14, 143-151.

Menzella, H. G., and Reeves, C. D. (2007) Combinatorial biosynthesis for drug development. Curr. Opin. Microbiol. 10, 238-245.

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, 1171-1176.

Migita, A., Watanabe, M., Hirose, Y., Watanabe, K., Tokiwano, T., Kinashi, H., and Oikawa, H. (2009) Identification of a gene cluster of polyether antibiotic lasalocid from *Streptomyces lasaliensis*. Biosci. Biotechnol. Biochem. 73, 169-176.

Mochizuki, S., Hiratsu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K., and Kinashi, H. (2003) The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. Mol. Microbiol. 48, 1501-1510.

Moffitt, M. C., and Neilan, B. A. (2003) Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. J. Mol. Evol. 56, 446-457.

Moore, B. S., and Hertweck, C. (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units. Nat. Prod. Rep. 19, 70-99.

Morris, R. M., Vergin, K. L., Cho, J.-C., Rappé, M. S., Carlson, C. A., and Giovannoni, S. J. (2005) Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic time-series study site. Limnol. Oceanogr. 50, 1687–1696.

Nedal, A., and Zotchev, S. B. (2004) Biosynthesis of deoxyaminosugars in antibiotic-producing bacteria. Appl. Microbiol. Biotechnol. 64, 7-15.

Newman, D. J., and Cragg, G. M. (2007) Natural products as sources of new drugs over the last 25 years. J Nat Prod. 70, 461-477.

Nishizawa, T., Asayama, M., Fujii, K., Harada, K., and Shirai, M. (1999) Genetic analysis of the peptide synthetase genes for a cyclic heptapeptide microcystin in *Microcystis* spp. J. Biochem. 126, 520-529.

O'Brien, D. P., Kirkpatrick, P. N., O'Brien, S. W., Staroske, T., Richardson, T. I., Evans, D. A., Hopkinson, A., Spencer, J. B., and Williams, D. H. (2000) Expression and assay of an N-methyltransferase involved in the biosynthesis of a vancomycin group antibiotic. Chem. Commun. 2000, 103-104.

Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. Nature. 405, 299-304.

O'Connor, S. E., Chen, H., and Walsh, C. T. (2002) Enzymatic assembly of epothilones: The EpoC subunit and reconstitution of the EpoA-ACP/B/C polyketide and nonribosomal peptide interfaces. Biochemistry. 41, 5685-5694.

Ogasawara, Y., Katayama, K., Minami, A., Otsuka, M., Eguchi, T., and Kakinuma, K. (2004) Cloning, sequencing, and functional analysis of the biosynthetic gene cluster of macrolactam antibiotic vicenistatin in *Streptomyces halstedii*. Chem. Biol. 11, 79-86.

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, 4050-4060.

Okanishi, M., Manome, T., and Umezawa, H. (1980) Isolation and characterization of plasmid DNAs in actinomycetes. J. Antibiot. (Tokyo). 33, 88-91.

Okanishi, M., Ohta, T., and Umezawa, H. (1970) Possible control of formation of aerial mycelium and antibiotic production in *Streptomyces* by episomic factors. J. Antibiot. (Tokyo). 23, 45–47.

Õmura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M. (2001) Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc. Nat. Acad. USA. 98, 12215-12220.

Õmura, S., Ikeda, H., Malpartida, F., Kieser, H. M., and Hopwood, D. A. (1986) Production of new hybrid antibiotics, mederrhodins A and B, by a genetically engineered strain. Antimicrob. Agents Chemother. 29, 13-19.

Õmura, S., and Tanaka, H. (1984) Production, structure, and antifungal activity of polyene macrolides, p. 351-404. In S. Õmura (ed.) Macrolide Antibiotics. Chemistry, Biology and Practice, 1st ed. Academic Press, Inc., Orlando, FL.

Ortholand, J.-Y., and Ganesan, A. (2004) Natural products and combinatorial chemistry: back to the future. Curr. Opin. Chem. Biol. 8, 271-280.

Overbye, K. M., and Barrett, J. F. (2005) Antibiotics: where did we go wrong? Drug Disc. Today. 10, 45-52.

Pál, C., Papp, B., and Lercher, M. J. (2005) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat Genet. 37, 1372-1375.

Pang, X., Sun, Y., Liu, J., Zhou, X., and Deng, Z. (2002) A linear plasmid temperature-sensitive for replication in *Streptomyces hygroscopicus* 10-22. FEMS Microbiol. Lett., 208, 25-28.

Pathom-aree, W., Nogi, Y., Ward, A. C., Horikoshi, K., Bull, A. T., and Goodfellow, M. (2006) *Dermacoccus barathri* sp. nov. and *Dermacoccus profundi* sp. nov., novel actinomycetes isolated from deep-sea mud of the Mariana Trench. Int. J. Syst. Evol. Microbiol. 56, 2303-2307.

Perna, N. T., Plunkett, G. 3rd, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Pósfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A., and Blattner, F. R. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature. 409, 529-533.

Philippe, H., and Douady, C. J. (2003) Horizontal gene transfer and phylogenetics. Curr. Opin. Microbiol. 6, 498-505.

Pospiech, A., Bietenhader, J., and Schupp, T. (1996) Two multifunctional peptide synthetases and an *O*-methyltransferase are involved in the biosynthesis of the DNA-binding antibiotic and antitumour agent saframycin Mx1 from *Myxococcus xanthus*. Microbiol. 142, 741-746.

Ragan, M. A. (2001) Detection of lateral gene transfer among microbial genomes. Curr. Opin. Genet. Dev. 11, 620-626.

Ragan, M. A., and Charlebois, R. L. (2002) Distributional profiles of homologous open reading frames among bacterial phyla: implications for vertical and lateral transmission. Int. J. Syst. Evol. Microbiol. 52, 777-787.

Ravel, J., Schrempf, H., and Hill, R. T. (1998) Mercury resistance is encoded by transferable giant linear plasmids in two Chesapeake bay *Streptomyces* strains. Appl. Environ. Microbiol. 64, 3383-3388.

Ravel, J., Wellington, E. M. H., and Hill, R. T. (2000) Interspecific transfer of *Streptomyces* giant linear plasmids in sterile amended soil microcosms. Appl. Environ. Microbiol. 66, 529-534.

Reeves, C. D., Murli, S., Ashley, G. W., Piagentini, M., Hutchinson, C. R., and McDaniel, R. (2001) Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. Biochemistry. 40, 15464-15470.

Reimmann, C., Patel, H. M., Serino, L., Barone, M., Walsh, C. T., and Haas, D. (2001) Essential PchG-dependent reduction in pyochelin biosynthesis of *Pseudomonas aeruginosa*. J. Bacteriol. 183, 813-820.

Rouhi, A. M. (2003) Rediscovering natural products. Chem. Eng. News. 81, pp. 77-78, 82-83, 86, 88-91.

Ruan, X., Pereda, A., Stassi, D. L., Zeidner, D., Summers, R. G., Jackson, M., Shivakumar, A., Kakavas, S., Staver, M. J., Donadio, S., and Katz, L. (1997) Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. J. Bacteriol. 179, 6416-6425.

Schatz, A., and Waksman, S. A. (1944) Effect of streptomycin and other antibiotic substances upon *Mycobacterium tuberculosis* and related organisms. Proc. Soc. Exptl. Biol. Med. 57, 244-248.

Schauwecker, F., Pfennig, F., Grammel, N., and Keller, U. (2000) Construction and *in vitro* analysis of a new bi-modular polypeptide synthetase for synthesis of *N*-methylated acyl peptides. Chem. Biol. 7, 287-297.

Schroeder, F. C., Gibson, D. M., Churchill, A. C. L., Sojikul, P., Wursthorn, E. J., Krasnoff, S. B., and Clardy, J. (2007) Differential analysis of 2D NMR spectra: new natural products from a pilot-scale fungal extract library. Angew. Chem. Int. Ed. 46, 901-904.

Schwarzer, D., Finking, R., and Marahiel, M. A. (2003) Nonribosomal peptides: from genes to products. Nat. Prod. Rep. 20, 275-287.

Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr. Opin. Chem. Biol. 7, 285–295.

Shen, B., Du, L., Sanchez, C., Edwards, D. J., Chen, M., and Murrell, J. M. (2001) The biosynthetic gene cluster for the anticancer drug bleomycin from *Streptomyces verticillus* ATCC15003 as a model for hybrid peptide-polyketide natural product biosynthesis. J. Ind. Microbiol. Biotechnol. 27, 378-385.

Shen, B., Du, L., Sanchez, C., Edwards, D. J., Chen, M., and Murrell, J. M. (2002) Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003. J. Nat. Prod., 65, 422–431.

Sieber, S. A., and Marahiel, M. A. (2005) Molecular mechanisms underlying non-ribosomal peptide synthesis: approaches to new antibiotics. Chem. Rev. 105, 715-738.

Silakowski, B., Nordsiek, G., Kunze, B., Blöcker, H., and Müller, R. (2001) Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga15. Chem. Biol. 8, 59-69.

Singh, S. B., and Pelaez, F. (2008) Biodiversity, chemical diversity and drug discovery, pp. 141-174. *In* Petersen, F. and Amstutz, R. (eds.) Progress in drug research, vol. 65, Birkhäuser Verlag, Basel.

Stanton, T. B. (2007) Prophage-like gene transfer agents-novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. Anaerobe. 13, 43-49.

Stein, D. B., Linne, U., and Marahiel, M. A. (2005) Utility of epimerization domains for the redesign of nonribosomal peptide synthetases. FEBS J. 272, 4506-4520.

Süssmuth, R. D., and Wohlleben W. (2004) The biosynthesis of glycopeptide antibiotics-a model for complex, non-ribosomally synthesized, peptidic secondary metabolites. Appl. Microbiol. Biotechnol., 63, 344–350.

Tang, L., Chung, L., Carney, J. R., Starks, C. M., Licari, P., and Katz, L. (2005) Generation of new epothilones by genetic engineering of a polyketide synthase in *Myxococcus xanthus*. J. Antibiot. (Tokyo). 58, 178-184.

Thomas, C. M., and Nielsen, K. M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat. Rev. Microbiol. 3, 711-721.

Torsvik, V., Sørheim, R., and Goksøyr, J. (1996) Total bacterial diversity in soil and sediment communities-a review. J. Indust. Microbiol. 17, 170-178.

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. USA. 104, 10376-10381.

van Wageningen, A. M. A., Kirkpatrick, P. N., Williams, D. H., Harris, B. R., Kershaw, J. K., Lennard, N. J., Jones, M., Jones, S. J. M., and Solenberg, P. J. (1998) Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. Chem. Biol. 5, 155-162.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H., and Smith, H. O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science. 304, 66-74.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., and van Sinderen, D. (2007) Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. Microbiol. Mol. Biol. Rev. 71, 495-548.

Vivian, A. (1971) Genetic control of fertility in *Streptomyces coelicolor* A3(2): plasmid involvement in the interconversion of UF and IF strains. J. Gen. Microbiol. 69, 353–364.

Vivian, A., and Hopwood, D. A. (1970) Genetic control of fertility in *Streptomyces coelicolor* A3(2): the IF fertility type. J. Gen. Microbiol. 64, 101–117.

Vivian, A., and Hopwood, D. A. (1973) Genetic control of fertility in *Streptomyces coelicolor* A3(2): new kinds of donor strains. J. Gen. Microbiol. 76, 147–162.

Waksman, S. A. and Woodruff, H. B. (1940) Bacteriostatic and bactericidal substances produced by soil *Actinomyces*. Proc. Soc. Biol. Med. 45, 609-614.

Waldron, C., Matsushima, P., Rosteck, P. R. Jr., Broughton, M. C., Turner, J., Madduri, K., Crawford, K. P., Merlo, D. J., and Baltz, R. H. (2001) Cloning and analysis of the spinosad biosynthetic gene cluster of *Saccharopolyspora spinosa*. Chem. Biol. 8, 487-499.

Walsh, C. (2003) Antibiotics-Actions, Origins, Resistance. 1st ed. ASM Press, Washington, DC.

Watanabe, T. (1963) Episome-mediated transfer of drug resistance in *Enterobacteriaceae*. VI. High-frequency resistance transfer system in *Escherichia coli.* J. Bacteriol. 85, 788-794.

Watanabe, T., and Fukasawa, T. (1961) Episome-mediated transfer of drug resistance in *Enterobacteriaceae*. I. Transfer of resistance factors by conjugation. J. Bacteriol. 81, 669-678.

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

Weber, G., Schörgendorfer, K., Schneider-Scherzer, E., and Leitner, E. (1994) The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. Curr. Genet., 26, 120-125.

Weber, J. M., Leung, J. O., Swanson, S. J., Idler, K. B., and McAlpine, J. B. (1991) An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. Science. 252, 114-117.

Weissman, K. J. (2006) The structural basis for docking in modular polyketide biosynthesis. Chembiochem. 7, 485-494.

Wenzel, S. C., Kunze, B., Höfle, G., Silakowski, B., Scharfe, M., Blöcker, H., and Müller, R. (2005) Structure and biosynthesis of myxochromides S_{1-3} in *Stigmatella aurantiaca*: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. ChemBioChem. 6, 375-385.

Wenzel, S. C., and Müller, R. (2005) Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. Curr. Opin. Chem. Biol. 9, 447–458.

Willems, E. W., Nooter, K., and Verweij, J. (2006). Antitumor antibiotics, pp. 359-370. *In* Chabner, B. A and Longo, D. L. (eds.) Cancer chemotherapy & biotherapy. Principles and practice. 4th ed., Lippincott Williams & Wilkins, Philadelphia, PA.

Williams, P. G., Buchanan, G. O., Feling, R. H., Kauffman, C. A., Jensen, P. R., and Fenical, W. (2005) New cytotoxic salinosporamides from the marine actinomycete *Salinispora tropica*. J. Org. Chem. 70, 6196-6203.

Woese, C. R., and Fox G. E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA. 74, 5088-5090.

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, 81-90.

Yamasaki, M., Ikuto, Y., Ohira, A., Chater, K., and Kinashi, H. (2003) Limited regions of homology between linear and circular plasmids encoding methylenomycin biosynthesis in two independently isolated streptomycetes. Microbiol. 149, 1351-1356.

Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J., Remington, K., Eisen, J. A., Heidelberg, K. B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P., Miller, C. S., Li, H., Mashiyama, S. T., Joachimiak, M. P., van Belle, C., Chandonia, J.-M., Soergel, D. A., Zhai, Y., Natarajan, K., Lee, S., Raphael, B. J., Bafna, V., Friedman, R., Brenner, S.E., Godzik, A., Eisenberg, D., Dixon, J. E., Taylor, S. S., Strausberg, R. L., Frazier, M., and Venter, J. C. (2007) The *Sorcerer II* Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol. 5, e16, 0432-0466.

Zazopoulos, E., Huang, K., Staffa, A., Liu, W., Bachmann, B. O., Nonaka, K., Ahlert, J., Thorson, J. S., Shen, B., and Farnet, C. M. (2003) A genomics-guided approach for discovering and expressing cryptic metabolic pathways. Nat. Biotechnol. 21, 187-190.

Zengler, K., Toledo, G., Rappé, M., Elkins, J., Mathur, E. J., Short, J. M., and Keller, M. (2002) Cultivating the uncultured. Proc. Natl. Acad. Sci. USA. 99, 15681-15686.

Zhang, L. (2005) Integrated approaches for discovering novel drugs from microbial natural products. *In* Zhang, L. and Demain, A. L. (eds.) Natural products: drug discovery and therapeutic medicine. Humana Press, Totowa, NJ.

Zumla, A., and Grange, J. M. (2001) Multidrug-resistant tuberculosis-can the tide be turned? Lancet Infect. Dis. 1, 199-202.

REFERENCES

Paper I

Is not included due to copyright

Paper II

Is not included due to copyright

Paper III

Is not included due to copyright