

**Analysis of the genetic potential of the sponge-derived fungus *Penicillium chrysogenum* E01-10/3 for polyketide production**

Dissertation  
zur  
Erlangung des Doktorgrades (Dr. rer. nat.)  
der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von  
Marija Avramović  
aus  
Belgrad

Bonn 2010

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen  
Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Jörn Piel
2. Gutachterin: Priv.-Doz. Dr. Gerhild van Echten-Deckert

Tag der Mündlichen Prüfung: 16.02.2011

Erscheinungsjahr: 2011

## ***To the unconditional love of my parents***

*“In every job that must be done  
There is an element of fun  
You find the fun and snap!  
The job’s a game  
And every task you undertake  
Becomes a piece of cake  
A lark! A spree! It’s very clear to see that...  
A Spoonful of sugar helps the medicine go down  
The medicine go down  
The medicine go down  
Just a spoonful of sugar helps the medicine go down  
In a most delightful way...  
The honey bees that fetch the nectar  
From the flowers to the comb  
Never tire of ever buzzing to and fro  
Because they take a little nip  
From every flower that they sip  
And hence,  
They find  
Their task is not a grind ...”*

***“A spoon full of sugar”, Disney (Mary Poppins)***

## **Acknowledgements**

I want to thank my supervisor, Prof. Dr. Jörn Piel for giving me the opportunity to work in his lab and for introducing me into the puzzling world of polyketides. I am grateful for all suggestions that I got from him during the practical work in the lab, as well as for the writing of my thesis. Moreover, I want to thank him for his trust and unique experience in building up a laboratory together with other “pioneers” of his group. This helped me to face easier other challenges in the life.

Special thanks go to Priv. Doz. Dr. Gerhild van Echten-Deckert for very efficient and dedicated review of my thesis as the second reviewer. Moreover, I appreciate a lot her kind encouragement for the finalization of my dissertation.

I want to thank all members of the examination committee for their participation and taking their time for reading of my thesis.

Furthermore, I would like to thank the collaborating groups of Prof. Dr. Bringmann (University of Würzburg) and Prof. Dr. Imhoff (University of Kiel) for providing us with the laboratory samples and valuable information. Special thank in this aspect goes to Dr. Tobias Gulder and Dr. Rüdiger Stöhr.

I would like to thank all members of the lab for creating a pleasant and exciting environment for scientific research.

Especially, I want to say thanks to my colleagues Dr. Cristian Gurgui, Zeynep Yunt, Dr. Katrin Zimmerman, Dr. Jing He, Sinisa Hrvatin and Dr. Daniel Butzke for introducing me into the diverse methods in molecular biology and biochemistry and for their helpful discussions. This brought me always a step forward in my research.

Special thanks go to Nina Heycke for the help on preparation and screening of the genomic library. Zeynep Yunt I thank for helping me to deal with the HPLC experiments. Dr. Katja Fisch I thank for her assistance in designing degenerate primers for screening of the genomic library.

Moreover, without fast and selfless assistance of Dr. Katrin Zimmerman and Dr. Kathrin Reinhardt in administrative questions at the institute, I would not be able to proceed with my work. I also thank them for encouraging me to speak the German language.

Thanks to all of my friends in Germany and Serbia who gave me support during all these years. Especially to Zeynep Yunt and Cristian Gurgui who encouraged me in critical phases of my research and life, and were always there when I had homesickness.

Members of my YES Medical Device Services GmbH team I want to thank for their encouragement and support.

I want to express my deepest thanks to my parents for their absolute support and their belief in me. I want to thank them for giving me strength in crucial phases of my life, also in critical times when I was supposed to help and encourage them. I am grateful for their love.

The last but not least, I want to thank my husband Dr. Oliver Wingerter without whose inexhaustible support and belief in me I would never be writing these acknowledgements. I want to thank him not only for proof-reading of my dissertation and for giving his critical opinion and suggestions, but also for all his motivation, love and care in the last few years, including his patience.

<b>SUMMARY .....</b>	<b>- 1 -</b>
<b>INTRODUCTION.....</b>	<b>- 4 -</b>
1 FUNGI AS PRODUCERS OF BIOLOGICALLY ACTIVE SECONDARY METABOLITES .....	- 4 -
1.1 Fungal mycotoxins .....	- 5 -
1.2 Fungal secondary metabolites as approved pharmaceuticals.....	- 8 -
1.3 Marine-derived fungal polyketide metabolites .....	- 13 -
1.4 The diversity of polyketides produced by fungus <i>P. chrysogenum</i> .....	- 20 -
2 POLYKETIDE BIOSYNTHESIS.....	- 27 -
2.1 The molecular background to understand polyketide biosynthesis.....	- 28 -
2.2 Types of polyketide synthases.....	- 29 -
2.3 Fungal polyketide synthases.....	- 31 -
2.4 Current challenges in the fungal polyketide research area.....	- 36 -
2.5 The perspectives in fungal polyketide research area .....	- 44 -
3 BIOSYNTHESIS OF SORBICILLACTONE A .....	- 45 -
4 RESEARCH GOALS.....	- 48 -
<b>RESULTS AND DISCUSSION .....</b>	<b>- 50 -</b>
5 DETECTION OF SORBICILLACTONE A FROM <i>P. CHRYSOGENUM</i> E01-10/3 LIQUID CULTURES.....	- 50 -
6 GENETIC POTENTIAL OF <i>P. CHRYSOGENUM</i> STRAIN E01-10/3 FOR POLYKETIDE PRODUCTION .....	- 51 -
6.1 Design of PCR primers and cloning of putative PKS gene fragments .....	- 51 -
6.2 Phylogenetic analysis of amino acid sequences from putative PKS gene fragments .....	- 85 -
7 SCREENING OF A GENOMIC LIBRARY FOR THE PUTATIVE SORBICILLACTONE GENE CLUSTER .....	- 92 -
7.1 Construction of <i>P. chrysogenum</i> genomic fosmid library .....	- 92 -
7.2 Screening of the genomic library via hybridization .....	- 92 -
7.3 Screening of the genomic library via PCR.....	- 101 -
7.4 Summary of <i>P. chrysogenum</i> E01-10/3 genomic library screening .....	- 117 -
8 DNA SEQUENCE ANALYSIS OF THE PUTATIVE SORBICILLACTONE GENE CLUSTER .....	- 119 -
8.1 Shotgun sequencing of the putative sorbicillactone gene cluster .....	- 120 -
8.2 Identification of ORFs and prediction of enzyme functions .....	- 120 -
8.3 Subcloning of fragment-2 from 49C8 fosmid clone.....	- 125 -
8.4 KS domain phylogenetic analysis of cloned whole-length PKS genes.....	- 136 -
8.5 Domain analysis of cloned whole-length PKS genes .....	- 140 -
8.6 Proposed PKS biosynthetic route of sorbicillactones biosynthesis.....	- 149 -
9 ISOTOPE-LABELLED FEEDING EXPERIMENTS .....	- 154 -
10 GENERAL DISCUSSION AND FURTHER PROSPECTIVES.....	- 157 -
<b>MATERIALS AND METHODS .....</b>	<b>- 165 -</b>
11 MATERIALS .....	- 165 -
11.1 Media .....	- 165 -
11.2 Buffers and solutions.....	- 165 -
11.3 Strains, vectors and plasmids.....	- 169 -
11.4 Antibiotics and enzymes .....	- 171 -
11.5 PCR primers.....	- 172 -
11.6 Equipment and expendable materials .....	- 175 -
11.7 Chemicals and solvents .....	- 177 -
11.8 Commercial kits .....	- 178 -
11.9 Special computer programs and internet resources.....	- 178 -
12 METHODS .....	- 180 -
12.1 Centrifugation .....	- 180 -
12.2 Sterilization .....	- 180 -
12.3 Cultivation of fungus <i>P. chrysogenum</i> E01-10/3.....	- 180 -
12.4 Cultivation of <i>E. coli</i> cells.....	- 180 -
12.5 Isolation of chromosomal DNA from <i>P. chrysogenum</i> .....	- 181 -
12.6 Determination of DNA concentration and purity.....	- 181 -
12.7 Restriction enzyme digestion of DNA.....	- 182 -
12.8 Amplification of DNA using polymerase chain reaction (PCR).....	- 182 -
12.9 Agarose-gel-electrophoresis .....	- 186 -
12.10 Purification of DNA fragments from solutions or agarose gel.....	- 186 -

12.11	<i>Preparation of high quality plasmid DNA from E. coli</i> .....	- 187 -
12.12	<i>Cloning of PCR products with the pGEM-T Easy Vector System</i> .....	- 187 -
12.13	<i>Preparation of the TA cloning vector</i> .....	- 188 -
12.14	<i>Introduction of DNA into E. coli</i> .....	- 189 -
12.15	<i>Preparation of genomic library of P. chrysogenum</i> .....	- 190 -
12.16	<i>Screening the genomic P. chrysogenum fosmid library via PCR</i> .....	- 194 -
12.17	<i>Screening of the genomic fosmid library via hybridization</i> .....	- 195 -
12.18	<i>Automated DNA sequencing</i> .....	- 199 -
12.19	<i>In silico sequence analysis</i> .....	- 200 -
12.20	<i>Detection of sorbicillactone A from P. chrysogenum liquid cultures</i> .....	- 201 -
<b>REFERENCES</b> .....		<b>- 203 -</b>
<b>APPENDIX</b> .....		<b>- 218 -</b>
13	UV SPECTRUM OF SORBICILLACTONE A .....	- 218 -
14	DEGENERATE PRIMER DESIGN .....	- 219 -
15	NUCLEOTIDE SEQUENCES OF AMPLIFIED PKS GENE FRAGMENTS .....	- 226 -
16	NUCLEOTIDE SEQUENCE OF SHOT-GUN SEQUENCED 42H12 FOSMID CLONE .....	- 228 -
17	FGENESH IDENTIFICATION OF ORFS WITHIN CONTING-7 DNA SEQUENCES .....	- 239 -
18	LISTS OF FIGURES, TABLES AND STRUCTURES .....	- 242 -
18.1	<i>List of figures</i> .....	- 242 -
18.2	<i>List of tables</i> .....	- 242 -
18.3	<i>List of structures</i> .....	- 243 -
19	CURRICULUM VITAE .....	- 246 -

## Abbreviations

%	Percent	F	Fragmen
°C	Degree	FAD	Flavin adenine dinucleotide
μ	Micro	FAS	Fatty acid synthases(s)
6-MSA	6-Methylsalicylic acid	FASTA	DNA and protein sequence alignment software package
A	Adenosin	FDA	Food and Drug Administration
A	Ampere	FGSC	Fungal Genomic Stock Center
<i>A.</i>	<i>Aspergillus</i>	FMN	Flavin mononucleotid
aa	Amino acid(s)	FUMS	Fumonisin synthase
ACV	δ(L-a-aminoadipyl)-L-cysteinyl-D-valane	FUSS	Fusarin synthase
ad.	Adapted	h	Hour
AMP	Adenylation	HCl	Hydrochloric acid
approx.	Approximately	HMG-CoA	3-Hydroxy-3-methylglutaryl-coenzyme A
ARO	Aromatase(s)	HMM	Hidden Markov model
ARSs	Autonomously replicating sequences	HMW	High molecular weight
ATP	Adenosine triphosphate	HR	Highly reducing
BAC	Bacterial Artificial Chromosome	HR	Highly-reduced
BBE	Berberine bridge enzyme	I	Inosine
BLAST	Basic Local Alignment Search Tool	i.e.	That is
BMBF	Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research – Germany)	IARC	International Agency for Research on Cancer
bp	Base pair	IDH	Isoepoxydon dehydrogenase
BSA	Bovin Serum Albumin	IPTG	Isopropylthio-β-D-galactoside
C	Cytosin	k	Kilo
C/CON	Condensation (domain of PKS/NRPS)	kDa	Kilo Daltons
CCC	Covalently closed circular DNA	KS	ketosynthase
CDKS	Compactin diketide synthase	l	Liter
cDNA	Complementary DNA	LDKS	Lovastatin diketide synthase
CHS	Chalcone synthase	LDL	Low-density lipoprotein
CitS	Citrinin synthase	LED	Light emitting diode
CLC-TE	Claisen-cyclase–thioesterase(s)	LMP	Low melting point
CNKS	Compactin nonaketide synthase	LNKS	Lovastatin nonaketide synthase
CNS	Central nervous system	m	Mili
CoA	Coenzyme A	m	Meter
conc.	Concentration	M	Molar
CYC	Cyclase(s)	MAT	Malonyl-CoA:ACP acyl transferase(s)
deg.	Degenerate	max.	Maximum
DNA	Desoxyribonucleic acid	Mb	Megabases
DPPH	1,1-Diphenyl-2-picrylhydrazyl	ME	Minimum evolution
ds	Double strand	MEGA	Molecular Evolutionary Genetics Analysis
e.g.	Example given	MEM	Malt extract media
EA	Ergot alkaloids	MeTs	Methyl transferases
EDTA	Ethylendiamine tetraacetic acid	MeTs	Methyltransferases
EH	EPS15 homology	min	Minute
EMBL-EBI	European Molecular Biology Laboratory- European Bioinformatics Institute	<i>M.</i>	<i>Monascus</i>
EMBnet	European Molecular Biology network	MOPS	3-(N-Morpholino) propane sulphonic acid
EQS	Equisetin synthase	MOS	3-Methylorcinaledehyde synthase
es	End sequence	MP	Microtiter plate
EST	Expressed sequence tag	<i>MSAS</i>	6-Methylsalicylic acid synthase
etc.	<i>Et cetera</i>	MT/ME	Methyl transferase(s)
		MWCO	Molecular weight cut off
		n	Nano
		NA	Norsolorinic acid
		NADP	Nicotinamide-adenine



	dinucleotide phosphate		integration
NCBI	National Center for Biotechnology Information	RNA	Ribonucleic acid
NDA	New Drug Application	ROS	Reactive oxygen species
NEB	New England Biolabs	rpm	Revolutions per minute
NJ	Neighbour-joining	RT	Room temperature
NLS	Nuclear localization signal	<i>S.</i>	<i>Saccharomyces</i>
NR	Non-reducing	SAM	<i>S</i> -adenosyl methionine
NRP	Non-ribosomal peptide	SDS	Sodium dodecylsulfate
NRPS	Non-ribosomal peptide synthetase(s)	sec	Second
NSAS	Norsolorinic acid synthase	SQTKS	Squalestatin tetraketide synthase
ORF	Open reading frame(s)	SSH-PCR	Subtractive hybridization PCR- based approach
OSAS	Orsollinic acid synthase	T	Thymin
<i>P.</i>	<i>Penicillium</i>	TE	Tris-EDTA solution
PBS	Phosphate-buffered saline	TENS	Tenellin synthase
PEG	Polyethylene glycol	THNS	Tetrahydroxynaphthalene synthase
PKS	Polyketide synthase(s)	Tm	Melting temperature
PLP	Pyridoxal phosphate	TNF- $\alpha$	Tumor necrosis factor- $\alpha$
Pol	Polymerase	Tris	Tris-(hydroxymethyl)- aminomethane
PP	Phosphopantetheine	TTS1	T-toxin synthase 1
PPTase	4-Phosphopantetheinyl transferase	TTS2	T-toxin synthase 2
PR	Partially reducing	USA	United States of America
PR	Partially-reduced	UV	Ultra violet
PSI-BLAST	Position-Specific Iterative Basic Local Alignment Search Tool	V	Volt
PSSM	Position specific score matrix	WA	Naphthopyrone
PT	Product template	WAS	Naphthopyrone synthase
PTGS	Post-transcriptional gene silencing	WG	Working group
R	Reductase	WH2	Wiskott-Aldrich homology 2
R PKS	Reducing polyketide synthase	YWA1	Naphthopyrone
R/RED	Reducing/ reductase	ZS-A	Zearalenone synthase A
rDNA	Ribosomal desoxyribonucleic acid	ZS-B	Zearalenone synthase B
REMI	Restriction-enzyme-mediated		

## SUMMARY

The main goal of the presented dissertation was to evaluate the genetic potential of *P. chrysogenum* E01-10/3 strain for the production of polyketides.

This marine-derived *P. chrysogenum* strain was isolated from the Mediterranean sponge *Ircinia fasciculata* in the course of a research program focused on the discovery and characterization of novel natural products. This led to the isolation and characterisation of two novel and structurally highly similar polyketides: sorbicillactone A and sorbicillactone B. Consistent with their structural similarity, it was proposed that the biosynthesis of both compounds might be encoded by the same gene cluster. Consequently, the identification of the sorbicillactone A gene cluster was in the focus of this dissertation as this compound was previously shown to possess promising antileukaemic, antiviral and neuroprotective properties. The iterative type I PKSs, which have only one multidomain protein with all the enzyme activities covalently bound together, are responsible for the biosynthesis of fungal polyketides. The single multifunctional protein is used to iteratively catalyze multiple rounds of chain elongation and appropriate  $\beta$ -keto processing of a polyketide.

Since all of the fungal PKSs belong to the iterative type I PKS enzymes, degenerate primers and hybridization probes fitting to this type of fungal PKS systems were used in PCR and hybridization experiments. Of special help for the present study was the possibility to differentiate between subtypes of fungal iterative type I PKSs on the amino acid level: Nonreducing (NR), partially reducing (PR), and highly reducing (HR), in respect to level of reduction of their polyketide products. Accordingly, PCR and hybridization experiments were set up in order to take advantage of this fact.

During the course of this study, use of PCR enabled the amplification of partial PKS sequences from nine putatively distinct fungal type I PKS gene loci from *P. chrysogenum* E01-10/3. Six partial KS domain DNA sequences were used to reconstruct evolutionary relationships in respect to other iterative type I PKSs. The results of the phylogenetic analysis for KS domains illustrated that *P. chrysogenum* E01-10/3 strain has the genetic potential to produce all three main categories of fungal polyketides – HR, PR (i.e. 6-MSA type) and NR. The results of the phylogenetic analysis of PCR-amplified partial KS domains were valuable to judge on good candidates for the screening of genomic library for putative sorbicillactone gene cluster: Three partial PKS sequences shown to be putative members of the NR clade III

(KHKS1, KHKS32 and AT-9-11). For each of these three sequences specific primer pairs were designed and used in a complex genomic library screening procedure.

The subcloning and sequencing of the AT-9-11 genomic region revealed that one putative reducing PKS and one non-reducing PKS were located next to each other. The genomic insert anticipated to carry the core of the putative AT-9-11 PKS gene cluster was completely sequenced and analysed.

The sequence analysis of the putative sorbicillactone gene cluster identified eight putative ORFs matching fungal protein sequences from public databases. The first identified gene encodes for a putative transcriptional regulator (*slr* or *orf1*), which might be capable to coordinate expression of the structural genes in the cluster. This putative regulatory gene is followed by the gene encoding for one putative monooxygenase (*slmox* or *orf2*) that may be responsible for one post-PKS hydroxylation reaction during the sorbicillactones biosynthesis. The core of the identified gene cluster contains two PKS genes (*PKSSL1/orf3* and *PKSSL2/orf4*) located next to each other in opposite directions. The second putative transcriptional regulatory gene (*sltr* or *orf5*) appears as a fifth ORF and is placed next to a putative gene for a MFS transporter protein (*slMFS* or *orf6*). The product encoded from the seventh ORF (*slox* or *orf7*) could be an oxidoreductase. The last identified ORF that showed putative homology to fungal proteins was *orf8* that shared significant similarity with fungal actin cytoskeleton-regulatory complex proteins. This ORF was excluded from the predicted sorbicillactone gene cluster model, since its involvement in cytoskeleton dynamics is not likely to be required for the biosynthesis nor transport of sorbicillactones.

The analysis of the exact domain organisation of the identified *P. chrysogenum* PKS genes – *pksSL1* and *pksSL2*, enabled the prediction of their putative routes in the sorbicillactone A and B biosynthesis. The performed protein sequence analysis showed that the domain organisation of PKSSl1 is consistent with the one expected for the members of NR clade III: SAT-KS-AT-PT-ACP-MT-RED. All known PKS active site motifs were conserved and the analysed domains were not fragmented. These facts supported the prediction that they are active in *P. chrysogenum* PKSSl1. Additionally, the phylogenetic analysis of the PT domain gave clear hint that PKSSl1 could work as a tetraketide synthase.

Based on the domain analysis it was confirmed that PKSSl2 has the typical domain structure of a R clade I PKS: KS-AT-DH-(MT)-ER-KR-ACP. Thus, PKSSl2 could be sufficient for biosynthesis of the advanced triketide starter unit, since it contains all domains required for condensation and the complete processing of a  $\beta$ -carbon. All PKSSl2 domains have conserved consensus sequences within active sites and are not fragmented. However, there is a high

chance that the detected PKSs12 MT domain has no activity since the consensus sequence within its active site is not preserved. Such nonmethylated, reduced triketide starter unit could be produced by the PKSs12 within two successive condensation and  $\beta$ -processing rounds. This advanced starter unit may be then accepted by the SAT domain of PKSs11 that probably makes three additional nonreducing extension rounds. At the end of the discussed putative sorbicillactone polyketide routes, it was proposed that hexaketide thiolester intermediates of sorbicillactone biosynthesis are released in an aldehyde form by the RED domain encoded reductase followed by the release of the free *holo*-ACP thiol of PKSs11.

Taking into account the number and potential bioactivities of polyketides isolated from marine-derived fungi, the number of compounds reported up to date from different *P. chrysogenum* strains as well as the genetic potential of *P. chrysogenum* E01-10/3 for polyketide production identified in this study, it becomes apparent that this particular marine-derived strain may be a valuable source for pharmacologically useful polyketides.

## INTRODUCTION

### 1 Fungi as producers of biologically active secondary metabolites

The fungal kingdom includes many species with unique and unusual biochemical pathways [1]. The production of secondary metabolites in fungi is a complex process often coupled with morphological development [2]. Secondary metabolites often have obscure or unknown functions in organisms but have considerable importance for mankind due to their broad range of useful antibiotic, pharmaceutical as well as toxic activities [3]. The products of these pathways include important pharmaceuticals such as penicillin (**6**), cyclosporin (**12**) and statins (**10** and **11**), as well as potent poisons including aflatoxins (e.g. 1 aflatoxin B<sub>1</sub>) and trichothecenes [1]. The systematic study of fungal secondary metabolites began in 1922 by Harold Raistrick, who characterised more than 200 fungal metabolites [1]. However, it was the discovery and development of penicillin that have awakened scientific interest in fungal metabolites. Penicillin has been the first broad-spectrum antibiotic discovered from fungus *Penicillium notatum* (alias *Penicillium chrysogenum*) in 1929 by Alexander Fleming [4, 5]. The discovery of penicillin and its clinical use encouraged pharmaceutical companies to establish extensive screening programs for microbial bioactive metabolites [1]. A recent literature survey of fungal metabolites found 1500 compounds isolated and characterised between 1993 and 2001. More than half of these molecules show antibacterial, antifungal or antitumour activity [6] thereby illustrating that natural products are the most important source of anti-cancer and anti-infective agents [7]. Besides, it has been estimated that over 60% of approved and pre-New Drug Application (NDA) candidates are either natural products or related to them [7, 8]. The most prolific sources of fungal secondary metabolites are members of Aspergilli and Penicillia with lots of salt tolerant, fast growing species that are easily obtained from many substrates. Numerous compounds have been detected as promising candidates for investigation of secondary metabolites with pharmacological activities [9]. Up to date, there are dozens of fungal secondary metabolites that are used as antibiotic, antitumor, immunosuppressive, hypocholesterolaemic, antimigraine, and antiparasitic agents. A number of these products were first discovered as antibiotics that failed in their development, or as mycotoxins [10]. In addition new active fungal secondary metabolites have been discovered with pharmacological activities both in terrestrial [11, 12] and marine environment [13, 14] and some of them have already entered clinical testing.

## 1.1 Fungal mycotoxins

In the early 1960s thousands of turkey poultts mysteriously died in hatcheries in and around London. The investigations on suspicious fungi led to the isolation of toxic metabolites called aflatoxins (name derived from the species *A. flavus*) [1, 5].

Mycotoxins are known as low-molecular-weight natural products (i.e. small molecules) produced as secondary metabolites in filamentous fungi. These metabolites represent a chemically heterogeneous group, but they all exhibit high toxicity. By definition mycotoxins are synthesised in filamentous fungi and are toxic to vertebrates and other animal groups in extremely low (i.e. microgram) concentrations [15]. When present in food, these fungal metabolites can have toxic effects with acute (e.g. liver or kidney deterioration) and chronic (e.g. liver cancer) symptoms. Due to their mutagenic and teratogenic profile, exposure to fungal metabolites may result in skin irritation, immunosuppression, birth defects, neurotoxicity and death [16]

Since the natural fungal flora in foods is dominated by three genera — *Aspergillus*, *Fusarium*, and *Penicillium*, many toxicology studies are focussed on metabolites produced by these genera [17-21]. It is important to note that as many as 300-400 compounds are now recognised as mycotoxins [22]. Despite their toxicity, the fact that mycotoxins are associated with strong biological activities has initiated the search of new drug candidates among these secondary metabolites.

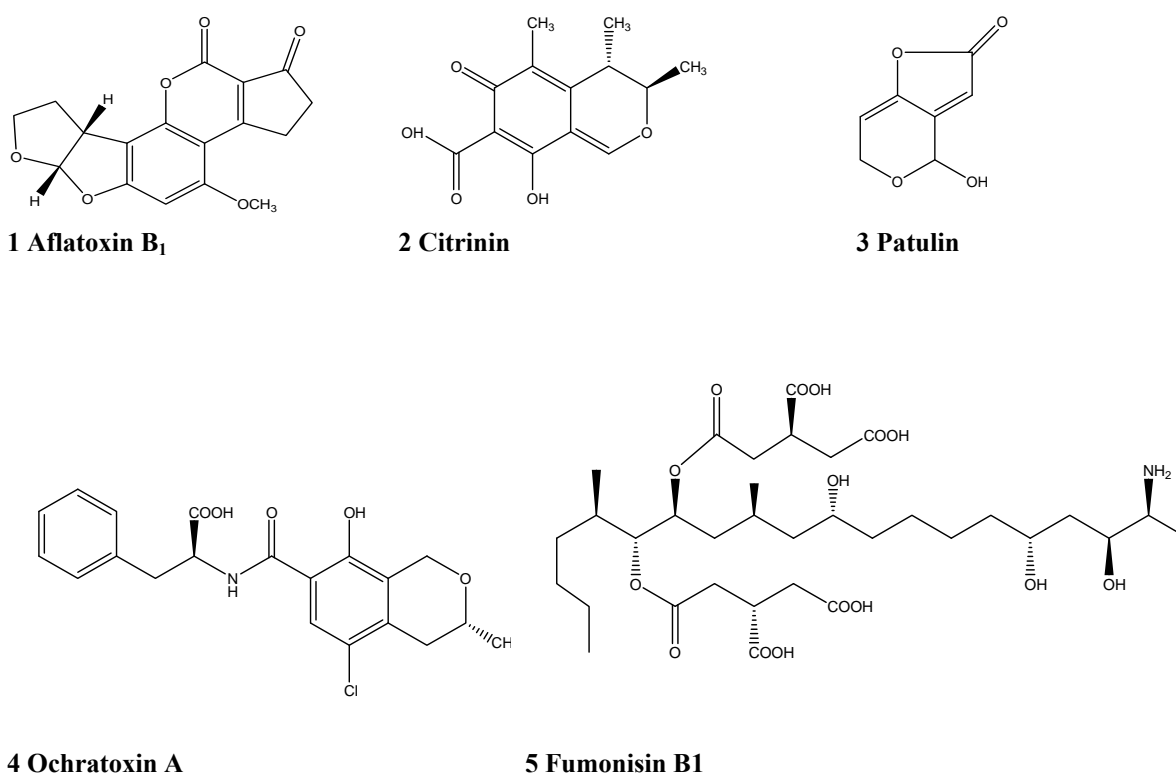
### 1.1.1 Polyketide-derived mycotoxins

The biosynthesis of many compounds characterised as mycotoxins has shown to be associated with the polyketide pathway. Since this study deals with the genetic potential of marine-derived *P. chrysogenum* strain E01-10/3 (IFM-GEOMAR collection) to produce polyketide-derived compounds, it is essential to present details on polyketide-derived mycotoxins, their bioactivities and biosynthesis in the following sections.

Aflatoxin and sterigmatocystin are potent, polyketide-derived, carcinogenic mycotoxins produced by fungi of the genus *Aspergillus* and constitute a major agricultural problem [23]. Among at least 16 structurally related aflatoxins, there are only four major aflatoxins, B<sub>1</sub>, B<sub>2</sub>, G<sub>1</sub> and G<sub>2</sub>, that contaminate agricultural commodities and are regarded as potential risk to livestock and human health [24]. Among them, aflatoxin B<sub>1</sub> (**1**) is recognised as the most potent (known) natural hepatocarcinogen and is usually the major aflatoxin produced by toxigenic strains [22, 25, 26]. These biosynthetically related compounds share the same polyketide precursor, norsolorinic acid (NA) [23]. The cloning of the norsolorinic synthase

(NAS) gene from both the sterigmatocystin pathway of *A. nidulans* [27] and the aflatoxin pathway of *A. parasiticus* [28, 29] was achieved in 1995. Up to date 25 open reading frames (ORF), or genes involved in the yet well-defined aflatoxin pathway have been identified and shown to be clustered on one chromosome. Similarly, the homologous genes of sterigmatocystin synthesis in *A. nidulans* and their involvement in the common part of biochemical pathway have been analyzed as well [30].

**Figure 1: Fungal polyketide-derived mycotoxins.**



Citrinin (**2**) was first isolated from *Penicillium citrinum* before World War II; subsequently, it was identified in over a dozen species of *Penicillium* and several species of *Aspergillus* [22]. More recently, citrinin has also been isolated from *Monascus ruber* and *Monascus purpureus*, *i.e.* industrial species used to produce red pigments [19]. Citrinin is bactericidal against Gram-positive bacteria, but due to its nephrotoxicity in mammalian systems it was not used in the clinical setting [31]. Besides, it shows embryotoxic, teratogenic, and genotoxic activity in short-term tests [32]. The biosynthesis of citrinin is another example of a mycotoxin polyketide. The working group of Nihira *et al.* identified genomic DNA fragment in *M. purpureus* suspected to be citrinin PKS (*pksCT* [33]). Later on, the same group identified additional five ORF within 21 kb region flanking *pksCT* from both

sides. Among these ORFs, one was proven to be an activator-gene involved in citrinin biosynthesis – *ctnA* [34].

Fumonisin were first described and characterised in 1988 [22]. They are primary amines thought to be synthesised by condensation of the amino acid alanine into an acetate-derived precursor [35]. Unlike other mycotoxins, they do not have an aromatic structure or a unique chromophore for easy analytical detection. Fumonisin are produced by a number of *Fusarium* species [36]. The most abundantly produced member of the family is fumonisin B<sub>1</sub> (**5**) [22]. Fumonisin affect animals in different ways by interfering with the sphingolipid metabolism [21, 37-39]. In humans fumonisin are supposed to be linked with oesophageal cancer [40]. The genes involved in fumonisin-biosynthesis are clustered in around 45 kb stretch of DNA. An expression analysis in *Fusarium verticillioides* indicated that 15 genes (ORF1 and ORF6-19) are co-regulated and exhibit patterns of expression that correlate with fumonisin-production. These ORF are designated as *FUM* genes [41]. *FUM5* encodes the polyketide-synthase-gene that was shown to be required for the fumonisin biosynthesis [42].

Patulin (**3**) is produced by many different fungal species but was first isolated as an antimicrobial active principle from *Penicillium patulum* (now *Penicillium griseofulvum*) during the 1940s. In early studies patulin was assessed for its antibiotic activity. However, during the 1950s and 1960s, it became apparent that besides its antibacterial, antiviral, and antiprotozoal activity, patulin is toxic to both plants and animals and was then re-classified as a mycotoxin [22]. Most of the patulin toxicity research conducted since 1995 has been focused on genotoxicity [43]. The patulin biosynthetic pathway is complex and contains at least ten enzymes involved in the direct biosynthetic pathway and as many as 19 enzymes involved in side pathways [44]. The patulin biosynthetic pathway played an important role in the study of the classical biochemistry of polyketide-biosynthesis. In 1990 the first fungal PKS-gene coding 6-methylsalicylic acid synthasae (*MSAS*) was isolated by Beck *et al.* from the *P. patulum* [45]. Actually, 6-methylsalicylic acid (6MSA) is the polyketide that serves as a precursor of patulin and genetic homologues of *MSAS* have since been isolated from numerous fungi [46, 47]. Isoepoxydon dehydrogenase (IDH) is an enzyme involved in the terminal portion of the patulin biosynthetic pathway [44]. Recently, the full length *idh* gene has been cloned from *P. expansum* and shown to be useful for detection of patulin-producers via PCR analysis [48]. Up to now sequences of only these two genes (*MSAS* and *idh*) of the patulin pathway enzymes are present in GenBank [49]. It is interesting to note that even though numbers of genes for partially reduced (PR) PKS (see 2.3.1) are now known from



fungal genomesequencing projects, the *MSAS* gene has been the only gene so far with a matched chemical product [50] (see Table 3).

Ochratoxin A (**4**) was discovered as a metabolite of *Aspergillus ochraceus* in 1965 during a large screening-procedure of fungal metabolites intended to identify new mycotoxins [22]. Subsequently, members of the ochratoxin family have been also found as metabolites in many different *Aspergillus* species [17, 51]. Although some early reports suggested their presence in several *Penicillium* species, it is now established that *P. verrucosum* and *P. nordicum* are the only confirmed ochratoxin producers in this genus [20, 52]. It has been shown that ochratoxin A is nephrotoxic, hepatotoxic, immune-suppressive, a potent teratogen, and carcinogen as well [53]. The PKS gene required for ochratoxin A biosynthesis in *A. ochraceus* was detected by a suppression-subtractive-hybridization PCR-based approach (SSH-PCR) [54]. In addition, it was reported by Karolewicz and Geisen [55] that a polyketide-synthase gene, i.e. *otapks*PN from *P. nordicum*, is essential for ochratoxin biosynthesis.

## 1.2 Fungal secondary metabolites as approved pharmaceuticals

Certainly, this section has to start with the penicillin story. Fleming's discovery led to the first successful chemotherapeutic produced by microbes that has initiated the golden age of antibiotics. This discovery opened the way for the development of many other antibiotics, and up to date penicillin has still remained the most active and the least toxic compound among many others [11].

Penicillins together with cephalosporins belong to the group of  $\beta$ -lactam compounds. The biosynthesis of the penicillin/cephalosporin antibiotics involves a common pathway with core enzyme activities conserved in all producer microorganisms. These producer species include a number of filamentous fungi, most notably members of the genera *Penicillium*, *Aspergillus* and *Cephalosporium* but also bacteria including *Streptomyces* and *Nocardia* species harbour genes for production of  $\beta$ -lactam compounds [56]. There are several types of penicillins, e.g. F, G, K, N and V. Penicillin V and G are active against most aerobic Gram-positive organisms. Penicillin G (**6**) is one of the most widely-used antibiotic agents today and is used against streptococcal, staphylococcal and meningococcal infections. The bactericidal effect of penicillins is based on the inhibition of bacterial cell-wall-synthesis [57]. In general, penicillins are well-tolerable [Food and Drug Administration (FDA), see Table 31], however, many cases of allergic reactions have been reported for penicillins [58, 59]. The penam nucleus of penicillins, as well as the cephem nucleus of

cephalosporins, is formed by condensation of the three precursor amino acids L- $\alpha$ -aminoadipic acid, L-cysteine and L-valine into tripeptide  $\delta$ (L- $\alpha$ -aminoadipyl)-L-cysteinyl-D-valane (ACV) by a mechanism designated as non-ribosomal peptide (NRP) synthesis [60]. In the penicillin and cephalosporin pathways, the initial NRP product, ACV, undergoes extensive modifications via transacylases, epimerases, and other enzymes [61]. Both enzymes and corresponding genes of these pathways were extensively studied in filamentous fungi [60, 62]. In the industrial scale penicillin G (**6**) is produced by fermentation of *P. chrysogenum* [63].

The ergot alkaloids are a complex family of indole alkaloids that are classified as mycotoxins [64] and are derived from a tetracyclic ergolane ring system through nonribosomal peptide synthesis [65]. Lysergic acid, with a structure common to all ergot alkaloids, was first isolated in 1934 [22]. Probably the most famous member of the family is ergotamine (**8**). Some ergot alkaloids show structural similarity to the neurotransmitters serotonin and dopamine and have affinity to the cognate receptors in the central nervous system. Therefore, ergot alkaloids have a long history in treatment of a variety of disorders of the central and peripheral nervous systems. In addition, several ergot alkaloids induce smooth muscle contractions [66]. These compounds are produced as a toxic cocktail of alkaloids in the sclerotia of *Claviceps* species that are common pathogens of various grass species [22]. Besides, the rather distantly related fungus *A. fumigatus* was also shown to be producer of ergot alkaloids [65]. The ingestion of these sclerotia, or ergots, has been associated with diseases such as “St. Anthony’s fire” since antiquity [10].

Sometimes the line between toxin and drug is defined with the shift of a decimal point of compound amount, or a change in a small chemical moiety with ergot alkaloids being a good example [22]. It is amazing that these “poisons” are now used to treat angina pectoris, hypertonia, migraine headache, cerebral circulatory disorder, uterine contraction, hypertension, serotonin-related disturbances, inhibition of prolactin-release in galactorrhoea, to reduce bleeding after child birth, and to prevent implantation in early pregnancy. Some of the ergot alkaloids also have antibiotic activity [10]. The roles of several genes from the ergot alkaloids synthesis gene cluster in *Claviceps purpurea* have been characterised [66] and their functions have been analyzed by heterologous expression or gene replacement approaches [67-69]. The non-ribosomal peptide synthetase (NRPS) that catalyzes the assembly of ergot-alkaloids is unique among fungal synthetases and consists of two separately encoded polypeptides, Lps1 and Lps2. Lps1 catalyzes the adenylation and thiolation of amino acids, while Lps2 activates D-lysergic acid [61]. The cluster also contains

genes predicted to encode oxidoreductases, a catalase and a gene for an additional NRPS, i.e. LpsC for which the function is still not known [70, 71].

Zearalenone (**9**) is another example of a mycotoxin with pharmacologically useful properties. It is a polyketide that is synthesised entirely from acetate-malonate units [72]. Zearalenone is produced by several *Fusarium* species [73]. It resembles 17 $\beta$ -estradiol, the principal hormone produced by the human ovary, to allow binding to the estrogen receptors in mammalian target cells. Thus, zearalenone is better classified as a nonsteroidal estrogen or mycoestrogen. It causes hyperestrogenism, especially in pigs, and reproductive disorders in experimental studies with animals and livestock [74]. The reduced form of zearalenone, i.e.  $\alpha$ -zearalenol, has revealed to increase estrogenic activity [75]. *Alpha*-zearalenol was recently shown to inhibit atherogenesis, lowering plasma low-density lipoprotein (LDL), cholesterol and limiting aortic plaque formation in ovariectomised rabbits fed with high doses of cholesterol [76]. Although this estrogenic compound showed no mutagenicity in Ames tests, this substance induced chromosomal anomalies in some lymphocyte, oocyte, and kidney cell cultures [77]. A synthetic commercial formulation called zeranol has been successfully marketed for use as an anabolic agent for both sheep and cattle [78]. In 1989, zeranol was banned by the European Union, but is still in use in other parts of the world [79]. Zearalenone has also been applied for the treatment of postmenopausal symptoms in women [80], and both zearalenol and zearalenone have been patented as oral contraceptives [81]. In summary, the zearalenone family of metabolites is an example of both potentially harmful metabolites and promising pharmaceutical candidate. Recently, Graffoor *et al.* located two PKS genes (*PKS4* and *PKS13*) [23] within the complete genome-sequence of *Gibberella zea* (alias *Fusarium graminearum*) and proved their involvement in zearalenone-biosynthesis via gene disruption experiments [82].

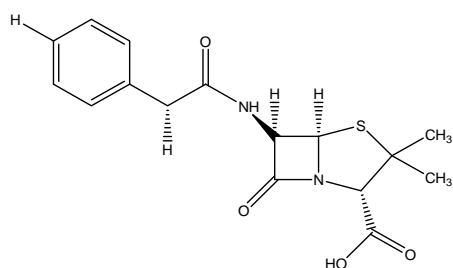
Another fungal secondary metabolite, cyclosporine A (**12**), was originally discovered as a narrow spectrum antifungal metabolite produced by the fungus *Tolypocladium infatum* [83]. It is a lipophilic cyclic polypeptide that produces specific calcium-dependent reversible inhibition of transcription of interleukin-2 and several other cytokines, predominantly in T-helperlymphocytes. This fungal secondary metabolite is associated with reduction of cytokine formation and inhibition of activation and/or maturation of various cell types. This includes as well those cells involved in cell-mediated immunity, thus, cyclosporine A is used as an immunosuppressant in human transplantation surgery and the treatment of autoimmune diseases [84]. The entire peptide biosynthesis of cyclosporine A is catalyzed by the

cyclosporin synthetase non-ribosomally. The gene encoding the cyclosporin synthetase of *Tolytocladium infatum* (*simA*) has been cloned and sequenced [85].

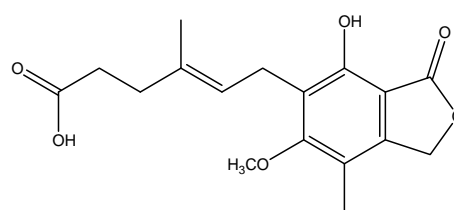
A very old broad-spectrum compound, mycophenolic acid (**7**), first discovered in 1896 and never commercialised as an antibiotic, has recently been developed as a new immunosuppressant [86]. Before being developed for an approved immunosuppressant, this organic acid was used to treat psoriasis [10]. 5-Methylorsellinic acid, but not orsellinic acid, is a precursor of mycophenolic acid in *P. brevicompactum* [72]. PKS genes neither for 5-methylorsellinic acid synthase, nor orsellinic acid synthase have been found yet, although they are coding for simplest tetraketides that require no  $\beta$ -keto-reductions (see 2.1) during their biosynthesis [50].

The members of the statin family of secondary metabolites are potent inhibitors of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, the key enzyme in cholesterol biosynthesis in humans [87]. Beside their main cholesterol-lowering effect, members of the statin family have also strong antifungal activities, especially against yeasts [10]. Brown *et al.* discovered in 1976 the first member of this group - compactin (**10**) (i.e. ML-236B), as an antibiotic product of *P. brevicompactum*. Independently, in the same year, Endo *et al.* discovered compactin in broths of *P. citrinum* as an inhibitor of HMG-CoA. Few years later Endo and Alberts independently discovered the more active methylated form of compactin known as lovastatin (**11**) (monacolin K or mevinolin) in broths of *Monascus ruber* and *A. terreus*, respectively [88]. These compounds are synthesised via the polyketide pathway and are composed of a conjugated decene ring system and 2-methylbutyryl side chain joined with an ester linkage [89]. The identification of a ~250 kDa polypeptide correlated with a production of lovastatin in *A. terreus*, and led to the cloning of the *lovB* gene that encodes lovastatin nonaketide synthase (LNKS) [90]. The lovastatin diketide synthase (LDKS) gene – *lovF*, was identified in the lovastatin biosynthetic gene cluster adjacent to the *lovB* gene [91]. The compactin (ML-236B) biosynthetic gene cluster was cloned from *P. citrinum*. The presence of *mlcA* and *mlcB* genes that show high homology with *lovB* and *lovF* genes, respectively, were identified in the gene cluster [92]. It is important to emphasise that natural statins and their derivatives are an example of multibillion-dollar drugs arising from fungal secondary metabolites [93]. One such derivative is pravastatin, which is produced by bioconversion of compactin [94].

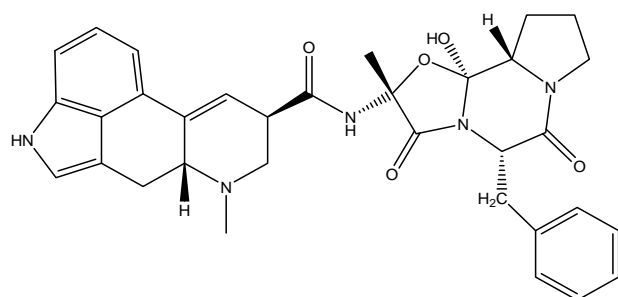
Figure 2: Pharmaceutically approved fungal secondary metabolites.



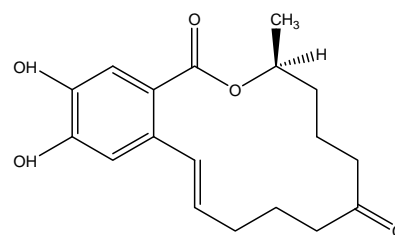
6 Penicillin G



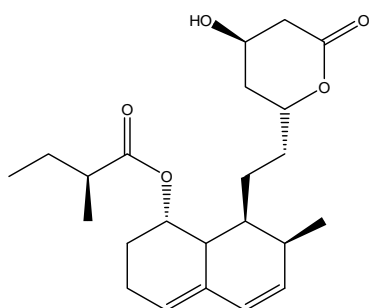
7 Mycophenolic acid



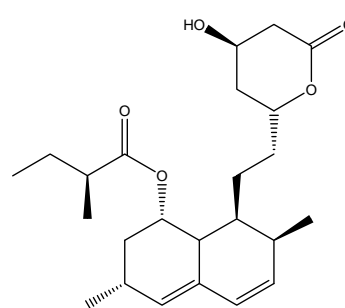
8 Ergotamine



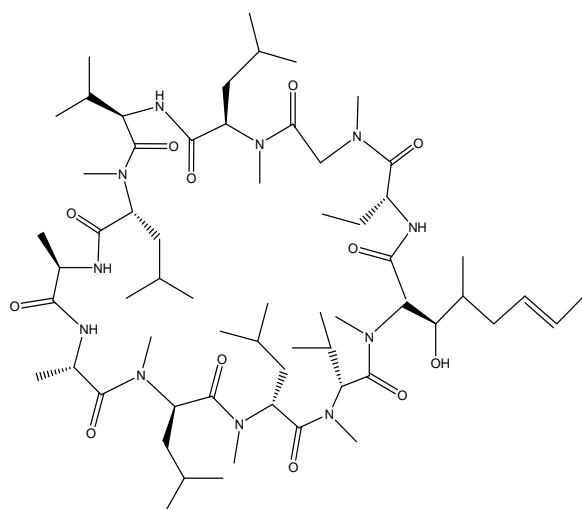
9 Zearalenon



10 Compactin



11 Lovastatin



12 Cyclosporin A

### 1.3 Marine-derived fungal polyketide metabolites

In search for novel and bioactive molecules for drug development, marine-derived natural resources have become an important research area. Although almost three quarters of the Earth's surface is occupied by seas and oceans, the isolation from soil was rather a common method to get fungal isolates. However, since fungal strains from terrestrial sources often yielded already known secondary metabolites, it was obvious that new sources are needed and therefore marine fungi have become an important source for isolation of pharmacologically active metabolites [9, 95, 96].

The group of marine-derived fungi include obligate marine fungi, which grow and sporulate exclusively in the marine or estuarine habitat, as well as facultative marine fungi, which grow equally well in marine, freshwater, and terrestrial environment [97]. Interestingly, all known marine fungal products have been isolated from cultured organisms, though, up to now it is estimated that fewer than 1% of all microorganisms, including fungi, have been successfully cultured [98]. As illustrated in section 1.2, terrestrial fungi produce many therapeutically significant molecules. Since marine organisms live in an environment significantly different from those of terrestrial organisms, it is reasonable to expect that their secondary metabolites will differ considerably. Although these natural resources have only recently been explored for natural products, there are currently over 15 fungal marine-derived secondary metabolites in clinical trials [13]. Moreover, it has been shown that among the nearly 300 new natural products isolated from marine-derived fungi [9], many of them are polyketides [13, 99]. Therefore, metabolites of this class might be the largest part of the secondary metabolites derived from marine as well as terrestrial microorganisms.

Since the *P. chrysogenum* E01-10/3 (IFM-GEOMAR collection) strain being subject of this work was cultured from a sample of the Mediterranean sponge *Ircinia fasciculata* (Elba, Italy) and has been shown to be capable for production of polyketides with pharmacologically interesting features [100], it is of interest to illustrate in the following the potential bioactivities of fungal polyketide metabolites isolated from marine environments. These polyketides are presented in the following section dealing with polyketides isolated from the fungus *P. chrysogenum* (see 1.4, Table 2).

All other polyketides that were up to date isolated from marine-derived fungi other than *P. chrysogenum* and have shown promising bioactivities are listed in the Table 1. Bioactive polyketides that were originally isolated from terrestrial fungi and later on re-isolated from marine-derived fungi, as well as marine-derived fungal polyketides without significant

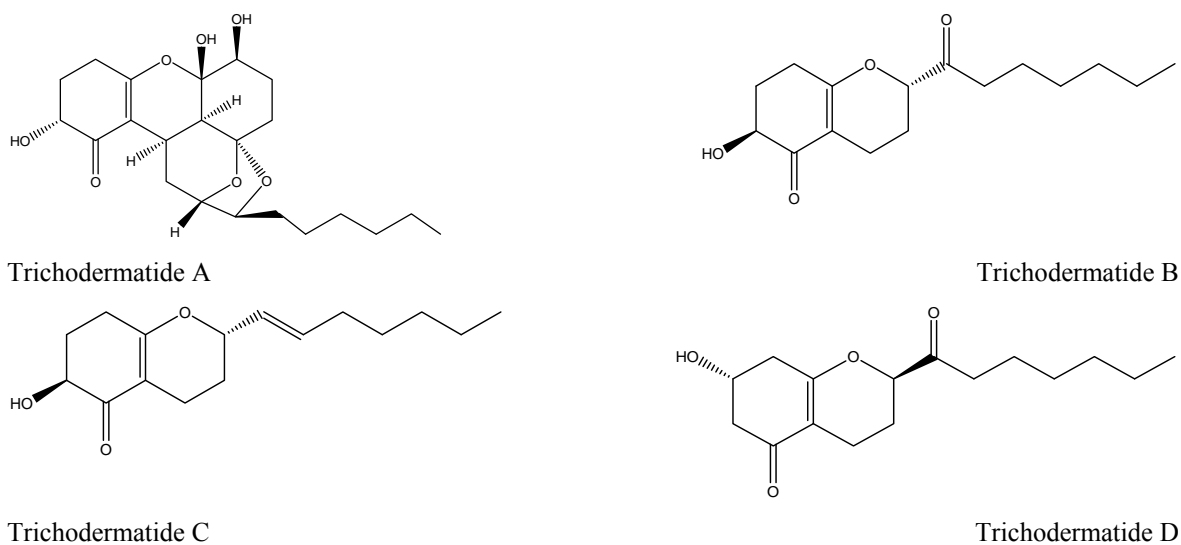
bioactivity are not presented in this table. This list is covering reports from 1989 until 2008 (February).

**Table 1: Bioactive polyketide metabolites from marine-derived fungi.**

Isolated fungus	Source/ symbiont of	Isolated polyketide(s)	Reported bioactivity	References
<i>Trichoderma reesei</i>	Marine mud	Trichodermatide A-D ( <b>13</b> )	Cytotoxicity	Sun <i>et al.</i> , 2008 [101]
<i>Penicillium</i> sp.	<i>Ulva pertusa</i> (marine green alga)	Redoxicitrinin ( <b>14</b> )	Antioxidant activity	Zhang <i>et al.</i> , 2007 [102]
<i>Gliocladium</i> sp.	<i>Syringodium Isoetifolium</i> (sea grass)	Cladionol A ( <b>15</b> )	Cytotoxicity	Kasai <i>et al.</i> 2005 [103]
<i>Emericella varicolor</i>	Marine sediment	Shimalactone A ( <b>16</b> )	Neurotrophic factor	Wei <i>et al.</i> , 2005 [104]
<i>Penicillium terrestre</i>	Marine sediment	Penicillones A and B ( <b>20</b> )	Cytotoxicity	Liu <i>et al.</i> , 2005 [105]
<i>Cladosporium</i> sp.	<i>Actinotrichia fragilis</i> (marine brown alga)	Sporiolides A and B ( <b>21</b> )	Cytotoxicity	Shigemori <i>et al.</i> , 2004 [106]
<i>Aspergillus</i> sp.	<i>Mytilus edulis</i> (marine mussel)	Aspermytin A ( <b>22</b> )	Neurotrophic factor	Tsukamoto <i>et al.</i> , 2004 [107]
<i>Penicillium terrestre</i>	Marine sediment	Dihydrobisvertinolone and Tetrahydrobisvertinolone ( <b>17</b> ) 2-(2',3-dihydrosorbyl)-3,6-dimethyl-5-hydroxy-1,4-benzoquinone ( <b>18</b> ) 3-acetonyl-2,6-dimethyl-5-hydroxy-1,4-benzoquinone ( <b>19</b> )	Cytotoxicity	Liu <i>et al.</i> , 2005 [108]
<i>Phoma</i> sp.	-	Phomoxin and Phomoxide ( <b>23</b> ), ( <b>24</b> )	Antioxidant activity	Liu <i>et al.</i> , 2003 [109]
<i>Epicoccum</i> sp.	<i>Fucus vesiculosus</i> (marine brown alga)	Epicoccone ( <b>25</b> )	Antioxidant activity	Abdel-Lateff <i>et al.</i> , 2003 [110]
<i>Emericella varicolor</i>	<i>Haliclona valliculata</i> (marine sponge)	Evariquinone and Isoemicellin ( <b>26</b> ), ( <b>27</b> )	Antiproliferative activity	Bringmann <i>et al.</i> 2003 [111]
<i>Wardomyces anomalus</i>	<i>Enteromorpha</i> sp. (marine green alga)	2,3,6,8-tetrahydroxy-1-methylxanthone( <b>37</b> ) 2,3,4,6,8-pentahydroxy-1-methylxanthone ( <b>37</b> )	Antioxidant activity	Abdel-Lateff <i>et al.</i> 2003 [112]
<i>Aspergillus ostianus</i>	Marine sponge of Pohnpei	8-chloro-9-hydroxy-8,9-deoxyasperlactone ( <b>38</b> ) 9-chloro-8-hydroxy-8,9-deoxyasperlactone ( <b>38</b> ) 9-chloro-8-hydroxy-8,9-deoxyaspyrone ( <b>39</b> )	Antibiotic activity	Namikoshi <i>et al.</i> , 2003 [113]

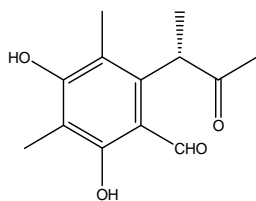
Isolated fungus	Source/ symbiont of	Isolated polyketide(s)	Reported bioactivity	References
<i>Acremonium</i> sp.	<i>Cladostephus spongus</i> (marine brown alga)	7-isopropenylbicyclo [4.2.0]octa-1,3,5-triene-2,5-diol ( <b>36</b> ) 7-isopropenylbicyclo [4.2.0]octa-1,3,5-triene-2,5-diol-5-b-D-glucopyranoside ( <b>36</b> )	Antioxidant activity	Abdel-Lateff <i>et al.</i> , 2002 [114]
<i>Aspergillus parasiticus</i>	<i>Carpopeltis cornea</i> (marine red alga)	Parasitenone ( <b>28</b> )	Antioxidant activity	Son <i>et al.</i> , 2002 [115]
<i>Emericella varicolor</i>	Spong of the Caribbean Sea	Varitriol , varioxirane and varixanthone ( <b>29</b> )	Cytotoxic and antibiotic properties	Malmstrom <i>et al.</i> 2002 [116]
<i>Paecilomyces</i> sp.	Coral reef	Paecilospirone ( <b>30</b> )	inhibitor of microtubule assembly	Namikoshi <i>et al.</i> 2000 [117]
<i>Penicillium</i> sp.	<i>Mytilus coruscus</i> (marine bivalve)	Sculezonones A and B ( <b>32</b> )	Inhibition of DNA polymerase	Komatsu <i>et al.</i> 2000 [118]
<i>Penicillium</i> sp.	<i>Enteromorpha intestinalis</i> (marine alga)	Penochalasin A-C ( <b>33</b> )	Cytotoxicity	Numata <i>et al.</i> 1996 [119]
Unidentified fungus	<i>Spirastrella vagabunda</i> (sponge)	14,15-Secocurvularin ( <b>31</b> )	Antibiotic activity	Abrell <i>et al.</i> 1996 [120]
<i>Penicillium</i> sp.	Sea sediment	Epolactaene ( <b>34</b> )	Neuritogenic activity	Takeya <i>et al.</i> 1995 [121]
<i>Passeriniella obiones</i>	<i>Spartina alterniflora</i> (salt marsh grass)	Obionin A ( <b>35</b> )	CNS activity	Poch <i>et al.</i> 1989 [122]

Figure 3: Bioactive polyketide metabolites from marine-derived fungi.

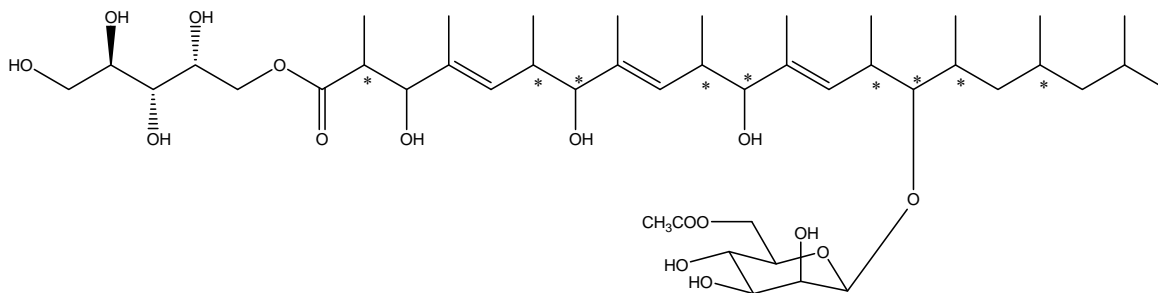


### 13 Trichodermatide A-D

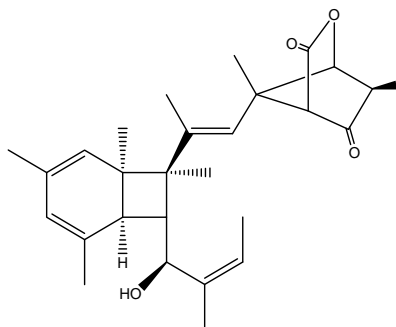




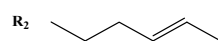
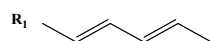
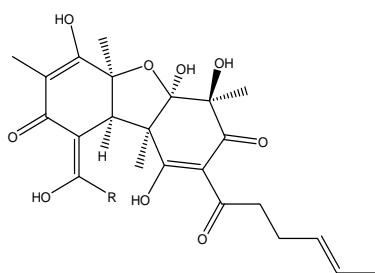
**14 Redoxitrinin**



**15 Cladionol A**



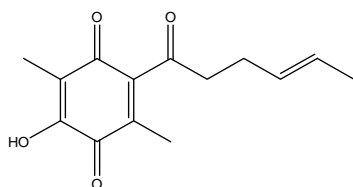
**16 Shimalactone A**



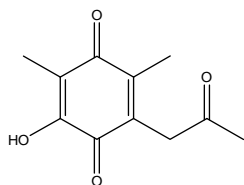
R<sub>1</sub>: Dihydrobisvertinolone

R<sub>2</sub>: Tetrahydrobisvertinolone

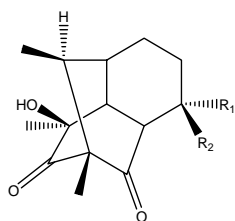
**17 Dihydrobisvertinolone and Tetrahydrobisvertinolone**



**18 2-(2',3-dihydrosorbyl)-3,6-dimethyl-5-hydroxy-1,4-benzoquinone**



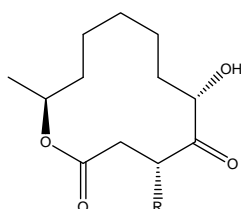
**19 3-acetyl-2,6-dimethyl-5-hydroxy-1,4-benzoquinone**



**20 Penicillones**

Penicillones A  $R_1, R_2: =O$

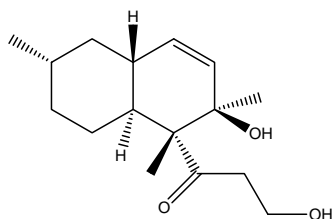
Penicillones B  $R_1: H, R_2: OH$



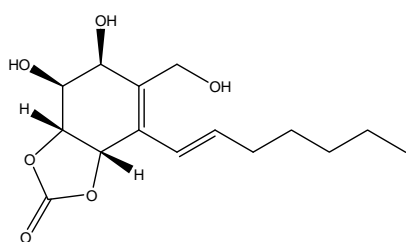
**21 Sporiolides**

Sporiolides A R: COPH

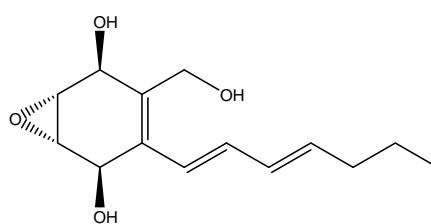
Sporiolides B R:  $CH_3$



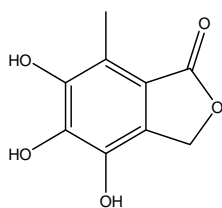
**22 Aspermytin A**



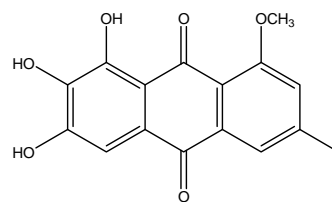
**23 Phomoxin**



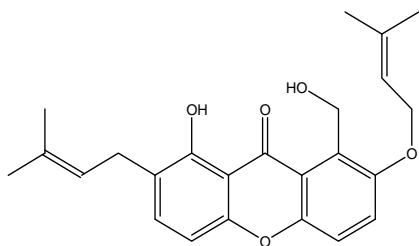
**24 Phomoxide**



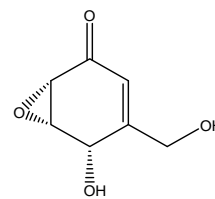
**25 Epicoccone**



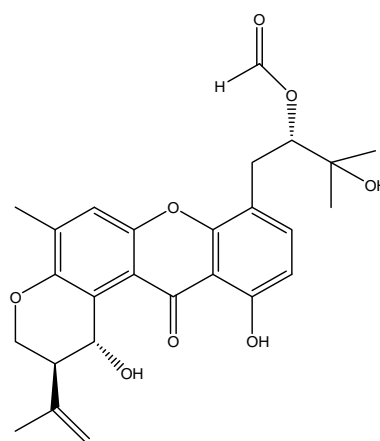
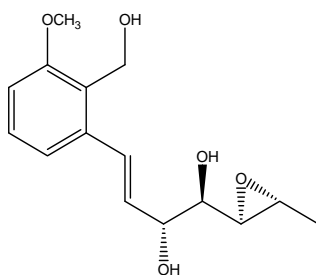
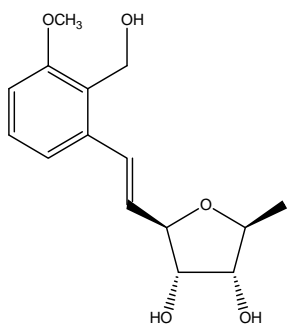
**26 Evariquinone**



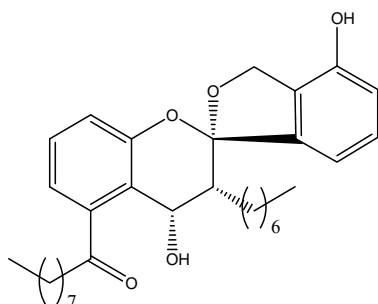
27 Isoemicicellin



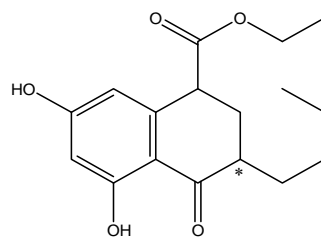
28 Parasitenone



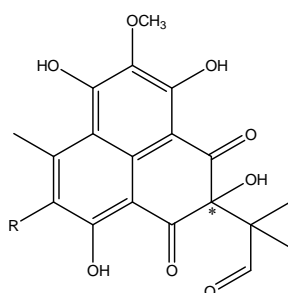
29 Varitriol, Varioxirane, Varixanthone



30 Paecilospirone



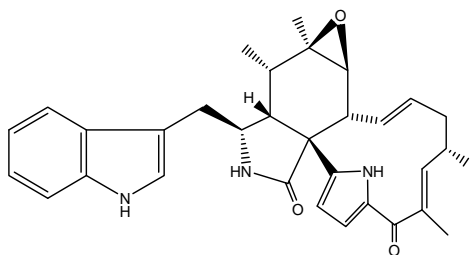
31 14, 15-Secocurvularin



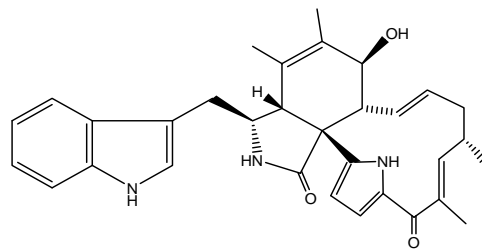
32 Sculezonone

Sculezonone A R: H

Sculezonone B R: OH

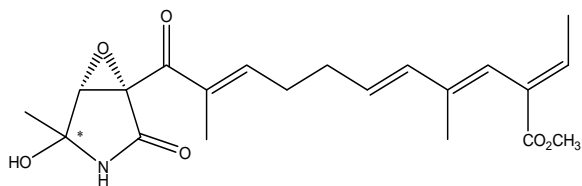


Penochalasin A

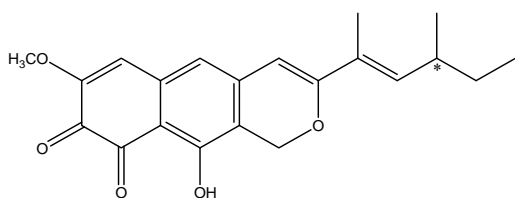


Penochalasin B

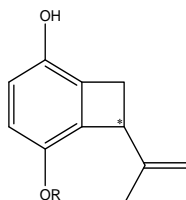
**33 Penochalasin**



**34 Epolactaene**



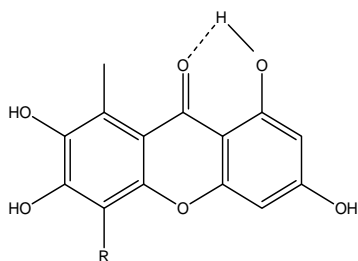
**35 Obionin A**



R<sub>1</sub>: H: 7-isopropenylbicyclo[4.2.0]octa-1,3,5-triene-2,5-diol

R<sub>2</sub>: Glc: 7-isopropenylbicyclo[4.2.0]octa-1,3,5-triene-2,5-diol-5-b-D-glucopyranoside

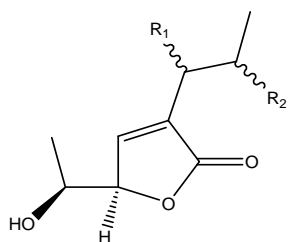
**36 7-isopropenylbicyclo[4.2.0]octa-1,3,5-triene-2,5-diol**



2,3,6,8-tetrahydroxy-1-methylxanthone R: H

2,3,4,6,8-pentahydroxy-1-methylxanthone R:OH

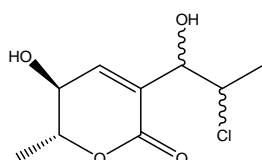
**37 methylxanthone**



**38 deoxyasperlactone**

8-chloro-9-hydroxy-8,9-deoxyasperlactone R<sub>1</sub>: Cl, R<sub>2</sub>: OH

9-chloro-8-hydroxy-8,9-deoxyasperlactone R<sub>1</sub>: OH, R<sub>2</sub>: Cl



**39 9-chloro-8-hydroxy-8,9-deoxyaspyrone**

Evidentially, polyketides isolated from marine-derived fungi represent an interesting group of bioactive substances that display myriads of effects on cell systems such as antioxidant, antibiotic, cytotoxic, neurotrophic and antiproliferative activities. Furthermore, polyketides can act as inhibitors of microtubule assembly, be responsible for inhibition of DNA polymerase or have antagonistic effect on receptors in the central nervous system (CNS). Their structural complexity can be quite impressive which complicates the establishment of commercially viable synthesis [13]. In addition, it is extremely difficult to provide sufficient amounts of active substances from fungi due to their limited levels of biosynthesis. Biosynthesis of bioactive marine-derived polyketides is dependent on many factors, one of which is certainly the rare occurrence of fungi themselves, especially those from marine environment. The establishment of fermentation procedures for such microorganisms would be an option, even though good knowledge on the biosynthetic as well as the genetic background of a particular strain is an essential prerequisite. Therefore, the research on marine fungi and the use of biotransformation and biotechnological methods may help to obtain potent candidates among polyketides for clinical use.

#### **1.4 The diversity of polyketides produced by fungus *P. chrysogenum***

*P. chrysogenum* is an ascomycotina fungus belonging to the genus *Penicillium* that together with other members of subgenus *Penicillium* has a characteristic terverticillate penicill. [123]. It is a ubiquitous fungus, among the most common eukaryotic life forms on earth. It occupies a wide range of habitats including soils, decaying vegetation and foods, and several other habitats from marine environment. Thus, the distribution of the fungus appears to be universal, throughout all biologically accessible regions and climates [124].

In 1927 Fleming incorrectly identified the producer of the  $\beta$ -lactam antibiotic penicillin (**6**) as *P. rubrum*. This view was corrected by Clutterbuck and collaborators in 1932 proving that Fleming's penicillin producing strain is actually *P. chrysogenum* [125]. It has been reported that *P. chrysogenum* strains produce a surprisingly high number of antibiotic families - penicillins, sorbicillins, secalonic acids, roquefortines, xanthocillins, questiomycins, negapillin, notatin and the antifungal protein PAF. The reason for this high number of identified compounds is based on the fact that *P. chrysogenum* has been intensively investigated since 1940 for antibiotics other than penicillin [126].

Moreover, several polyketides were up to date isolated from different *P. chrysogenum* strains. Among them are the teratogenic mycotoxins secalonic acid (**60**) that possesses antibiotic, antifungal, antitumour and antiulcer activity [126-128]. Two oktaketides, emodic acid (**61**) and  $\omega$ -hydroxy emodin (**62**) were isolated from this fungus in 1940s. Since that time several studies were conducted to investigate potential bioactivities of these two compounds [129-131]. There is evidence that emodic acid, due to its capability to mediate generation of reactive oxygen species (ROS), could be employed in targeted chemotherapy of tumours by coupling with receptors that are preferentially expressed by tumour cells [132]. The xanthoviridicatin E (**58**) and F (**59**) are two novel polyketide natural products isolated from a fermentation broth of an endophytic strain of *P. chrysogenum* from living leaves collected in Peru. The HIV-1 integrase inhibitory activity of these two compounds was described by the Merck Research Laboratories in 2003 [133].

The yellow pigment sorbicillin (**40**) has been observed previously as a minor metabolite of *P. chrysogenum* (formerly known as *P. notatum*) by Cram *et al.* in 1948 [134]. Later it was shown by Abe and collaborators that sorbicillin exhibits scavenger activity against 1,1-diphenyl-2-picrylhydrazyl (DPPH) free radical [135]. The typical carbon skeleton of sorbicillin is found in a wide variety of fungal metabolites, which accordingly are called sorbicillinoid or, in the case of dimers, bisorbicillinoid natural products [136]. Several fungal genera, including *Penicillium*, *Trichoderma*, *Verticillium*, *Aspergillus*, and *Paecilomyces*, produce sorbicillinoids and bisorbicillinoids [137].

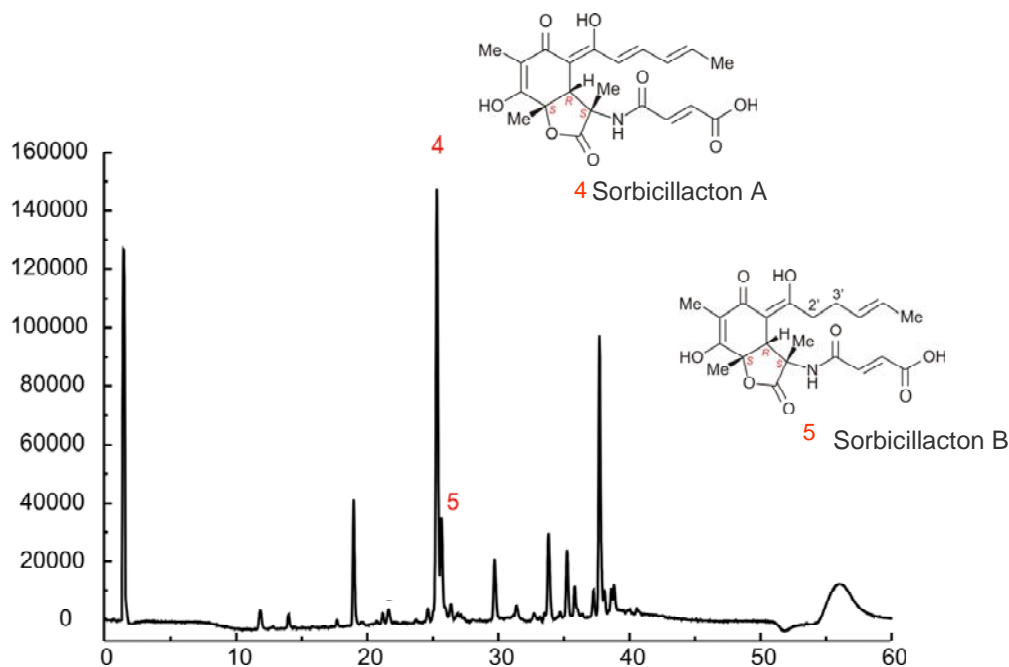
During the course of time, numbers of sorbicillinoid and bisorbicillinoid compounds were reported from different *P. chrysogenum* strains [100, 136, 138, 139]. However, the majority of strains sequester the bisorbicillinoid yellow pigment trichodimerol (**56**) [136, 140], which have recently been prepared in biomimetic syntheses [141]. It has been shown that trichodimerol suppresses the formation of tumor necrosis factor alpha (TNF- $\alpha$ ) [142]. In the year 1995, Miller and Huang isolated the new sorbicillinoid monomer that was named

sorrentanone (**42**) [139] due to its distinctive sorrel color. Within the same study, this compound showed activity against both Gram-positive and Gram-negative bacteria. They also isolated bisvertinolone (**57**), dimeric sorbicillin-related natural product [143], that was originally isolated from fungus *Verticillium intertestum* [144]. Bisvertinolone is reported to be a  $\beta$ -1,6-glucan biosynthesis inhibitor [145] and a potent DPPH radical scavenger [135].

The group of Laatsch from Göttingen isolated and characterised from a *P. chrysogenum* strain that was actually a bench top contaminant some sorbicillin analogues and related dimeric compounds [138]. Among isolated compounds were the monomers 2',3'-dihydrosorbicillin (**41**), the new sohirrones A-C (**43-45**), oxosorbicillinol (**46**; a stable tautomer of sorbicillin) and 7-deacetoxyyanuthone (**49**). This group also isolated a complex mixture of dimeric compounds containing the sorbicillin skeleton, i.e., the Diels-Alder dimers bisorbicillinol (**52**), its dihydro derivative bisvertinoquinol (**53**), bisorbibutenolide (**54**) and the new adducts rezishanones A-D (**55**). All compounds were tested for potential antimicrobial activity in the agar diffusion tests against several bacterial, fungal and algal species. The (**41**), (**43**), (**44**), (**46**), (**49**), (**52**), (**53**), and (**55**) exhibited weak activity against *Staphylococcus aureus* and *Bacillus subtilis*. Fungi and algae were not inhibited by any of the isolated compounds [138]. In one prior study it was shown that 7-deacetoxyyanuthone exhibits weak cytotoxicity against human solid tumor cells [146]. In addition, DPPH radical scavenging activity was previously reported for bisorbicillinol, bisvertinoquinol and bisorbibutenolide by Abe *et al.* [135].

In the search for novel bioactive compounds from sponge-derived microorganisms, research groups within the BMBF funded "BIOTECmarin" consortium identified three novel sorbicillin-derived compounds from the *P. chrysogenum* strain cultured from a sample of the Mediterranean sponge *Ircinia fasciculata* (Elba, Italy). Among these compounds there are two structurally and biosynthetically unprecedented fungal metabolites, the novel-type alkaloids sorbicillactone A (**50**) and its 2',3'-dihydroanalog sorbicillactone B (**51**) [136, 147]. Moreover, sorbicillactone A exhibits highly selective activity against the murine leukaemic lymphoblast cell line L5178y, and has antiviral and neuroprotective properties [136, 148]. Despite its almost identical molecular structure, sorbicillactone B is significantly less active than sorbicillactone A (by factor 10) [136, 148]. The third novel sorbicillin-derived compound sorbivinetone (**48**) might be an artifact derived from sorbicillinol by Diels-Alder reaction with ethyl vinyl ether [136].

A short overview on bioactivities of mentioned polyketide metabolites from different *P. chrysogenum* strains is given in Table 2, while chemical structures of these metabolites are shown in Figure 5.



**Figure 4: HPLC chromatogram of metabolites isolated from *P. chrysogenum* strain E01-10/3.** Newly isolated substances during the course of this study [136] are given in red (personal communication) [149].

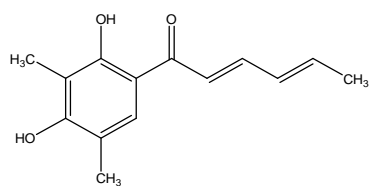
**Table 2: Reported diversity of polyketides produced by *P. chrysogenum*.**

<i>P. chrysogenum</i> polyketide(s)	Reported bioactivity	References
2', 3'-Dihydrosorbicillin ( <b>41</b> )	Antibiotic activity	Maskey <i>et al.</i> , 2005 [138]
7-Deacetoxyanuthone ( <b>49</b> )	Antibiotic and cytotoxic activity	Maskey <i>et al.</i> , 2005 [138]
Bisorbibutenolide ( <b>54</b> )	Antioxidant activity	Maskey <i>et al.</i> , 2005 [138]
Bisorbicillinol ( <b>52</b> )	Antibiotic and antioxidant activity	Maskey <i>et al.</i> , 2005 [138]
Bisvertinolone ( <b>57</b> )	Antioxidant activity and $\beta$ -1,6-glucan biosynthesis inhibitor	Trifonov <i>et al.</i> , 1986 [144]
Bisvertinoquinol ( <b>53</b> )	Antibiotic and antioxidant activity	Maskey <i>et al.</i> , 2005 [138]
Emodic acid ( <b>61</b> ) and $\omega$ -hydroxy emodin ( <b>62</b> )	Antitumour	Rahimipour <i>et al.</i> , 2001 [132]
Oxosorbicillinol ( <b>46</b> ) (tautomer)	Antibiotic activity	Maskey <i>et al.</i> , 2005 [138]
Rezishanones A-D ( <b>55</b> )	Antibiotic activity	Maskey <i>et al.</i> , 2005 [138]
Secalonic acid ( <b>60</b> )	Antibiotic, antifungal, antitumour and antiulcer activity	Frisvad <i>et al.</i> , 2004 [126]
Sohirnonones A ( <b>43</b> ) and B ( <b>44</b> )	Antibiotic activity	Maskey <i>et al.</i> , 2005 [138]
Sorbicillactone A ( <b>50</b> ) and B ( <b>51</b> )	Cytotoxicity, antiviral and neuroprotective activities	Bringmann <i>et al.</i> , 2005 [136]

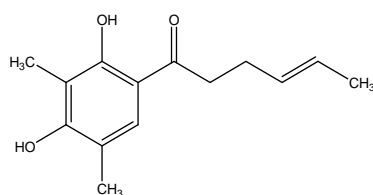


<i>P. chrysogenum</i> polyketide(s)	Reported bioactivity	References
Sorbicillin ( <b>40</b> )	Antioxidant activity	Cram <i>et al.</i> , 1948 [134]
Sorbivinetone ( <b>48</b> )	-	Bringmann <i>et al.</i> , 2005 [136]
Sorrentanone ( <b>42</b> )	Antibiotic activity	Miller <i>et al.</i> , 1995 [139]
Trichodimerol ( <b>56</b> )	Anti-inflammatory activity (suppression of the formation of TNF- $\alpha$ )	Andrade <i>et al.</i> , 1999 [140]
Xanthoviridicatin E ( <b>59</b> ) and F ( <b>59</b> )	Inhibition of HIV-1 integrase	Singh <i>et al.</i> , 2003 [133]

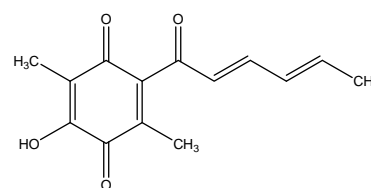
Figure 5: Polyketides produced by different *P. chrysogenum* strains.



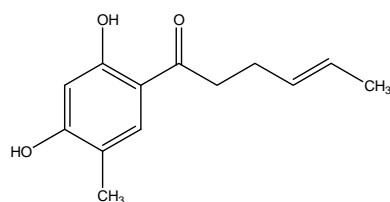
**40 Sorbicillin**



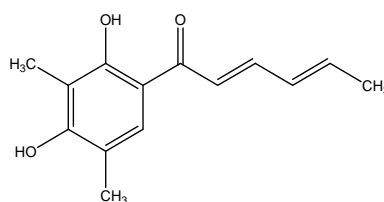
**41 2', 3'-dihydrosorbicillin**



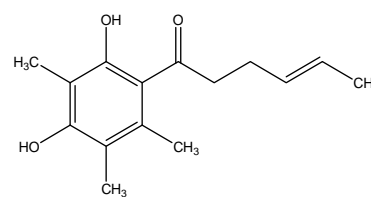
**42 Sorretanone**



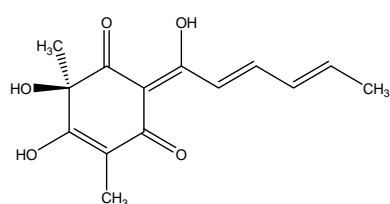
**43 Sohirnonone A**



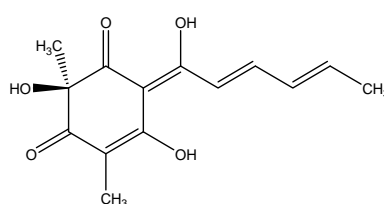
**44 Sohirnonone B**



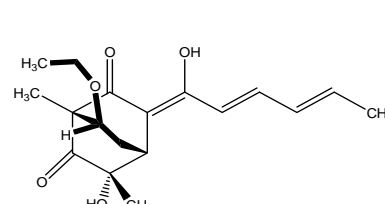
**45 Sohirnonone C**



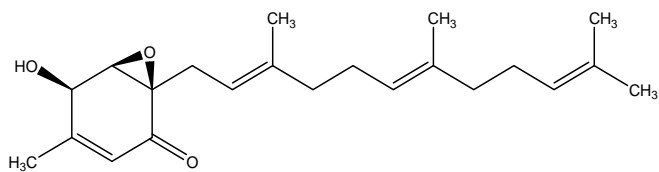
**46 Oxosorbicillinol**



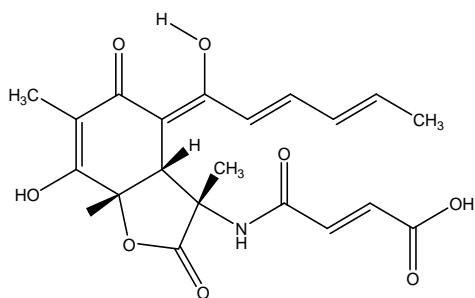
**47 Tautomer of oxosorbicillinol**



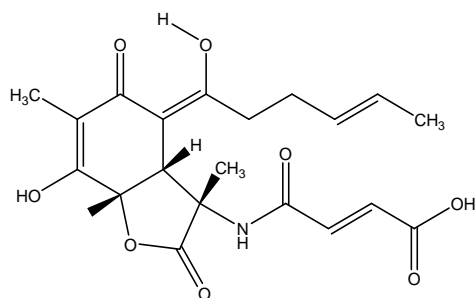
**48 Sorbivinetone**



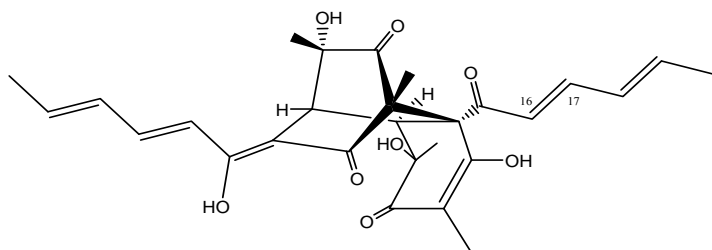
**49 7-Deacetoxyanuthone**



**50 Sorbicillactone A**

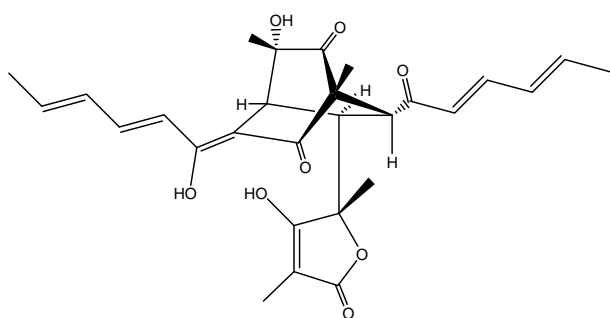


**51 Sorbicillactone B**



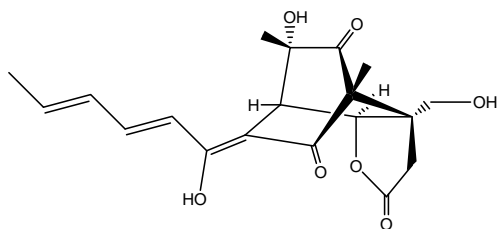
**52 Bisorbicillinol**

**53 Bisvertinoquinol (16, 17-dihydro)**

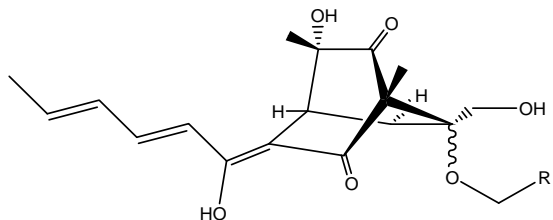


**54 Bisorbibutenolide**

55 Rezishanones A-D



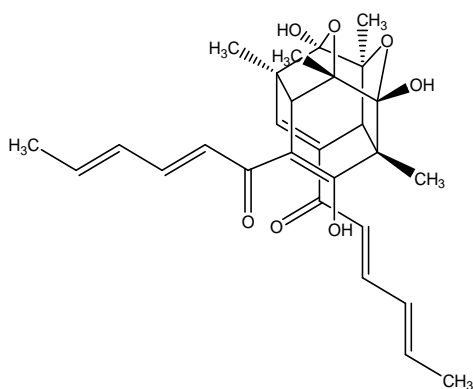
Rezishanones A



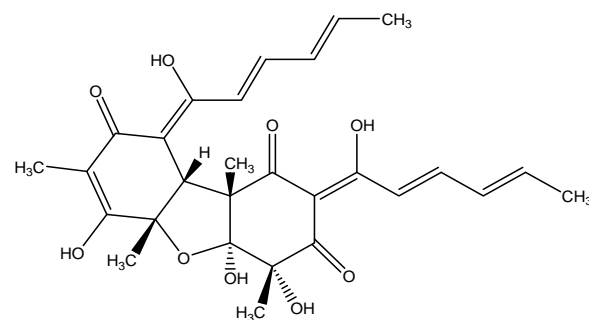
Rezishanones B R: n-C<sub>3</sub>H<sub>7</sub>

Rezishanones C R: CH<sub>3</sub>

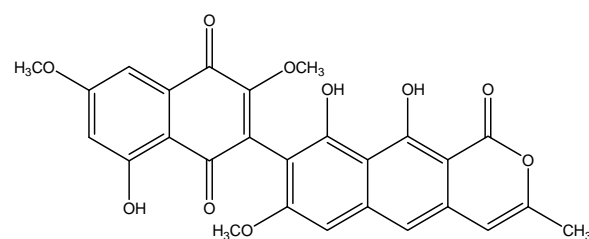
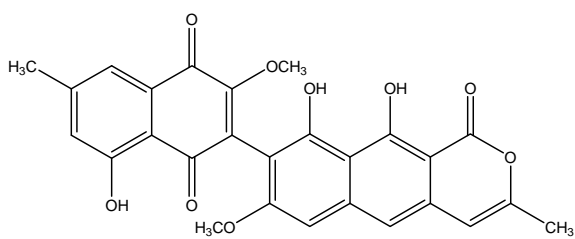
Rezishanones D R: CH<sub>3</sub>, Δ<sup>10</sup>, 13 saturated

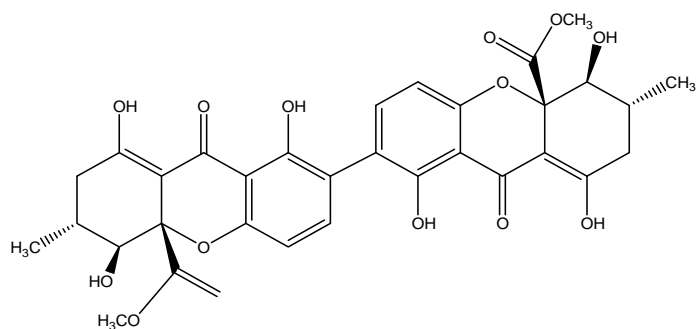
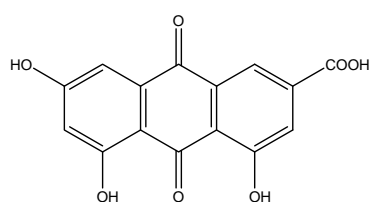
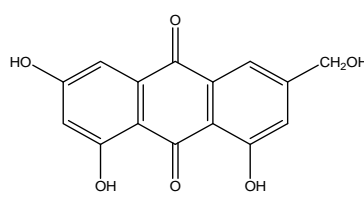


56 Trichodimerol



57 Bisvertinolone



**58 Xanthoviridicatin E****59 Xanthoviridicatin F****60 Secalonic acid****61 Emodic acid****62 ω-hydroxy emodin**

As shown above, many polyketides have been isolated from *P. chrysogenum*. In this context, a surprisingly high number of sorbicillinoid and bisorbicillinoid compounds are found among these polyketides. To the best of my knowledge, none of these metabolites has been up to date associated with certain PKS genes. In particular, in light of the diverse structures and interesting bioactivities it is of scientific interest to identify genes involved in biosynthesis of these compounds. Especially for sorbicillinoid and bisorbicillinoid compounds, it would be challenging for researchers to disclose the genetics behind the biosynthesis of these bioactive and structurally related fungal secondary metabolites.

## 2 Polyketide biosynthesis

More than hundred years ago, Collie coined the term “polyketide” for natural products derived from simple two-carbon acetate building blocks [89]. This proposal was later proven experimentally by Birch who used isotopically labelled acetate in the study of 6MSA biosynthesis in fungi and showed that it was formed from four acetate units. Then, Lynen and his coworkers succeeded in detecting MSAS activity in a cell-free extract of *P. patulum*, the first demonstration of polyketide synthase function *in vitro*. These chemical and biochemical experiments with fungi established the concepts of “polyketide biosynthesis” and “polyketide synthase” (PKS) [89]. Nowadays it is obvious that polyketides represent the largest family of structurally diverse secondary metabolites synthesised in both prokaryotic and eukaryotic

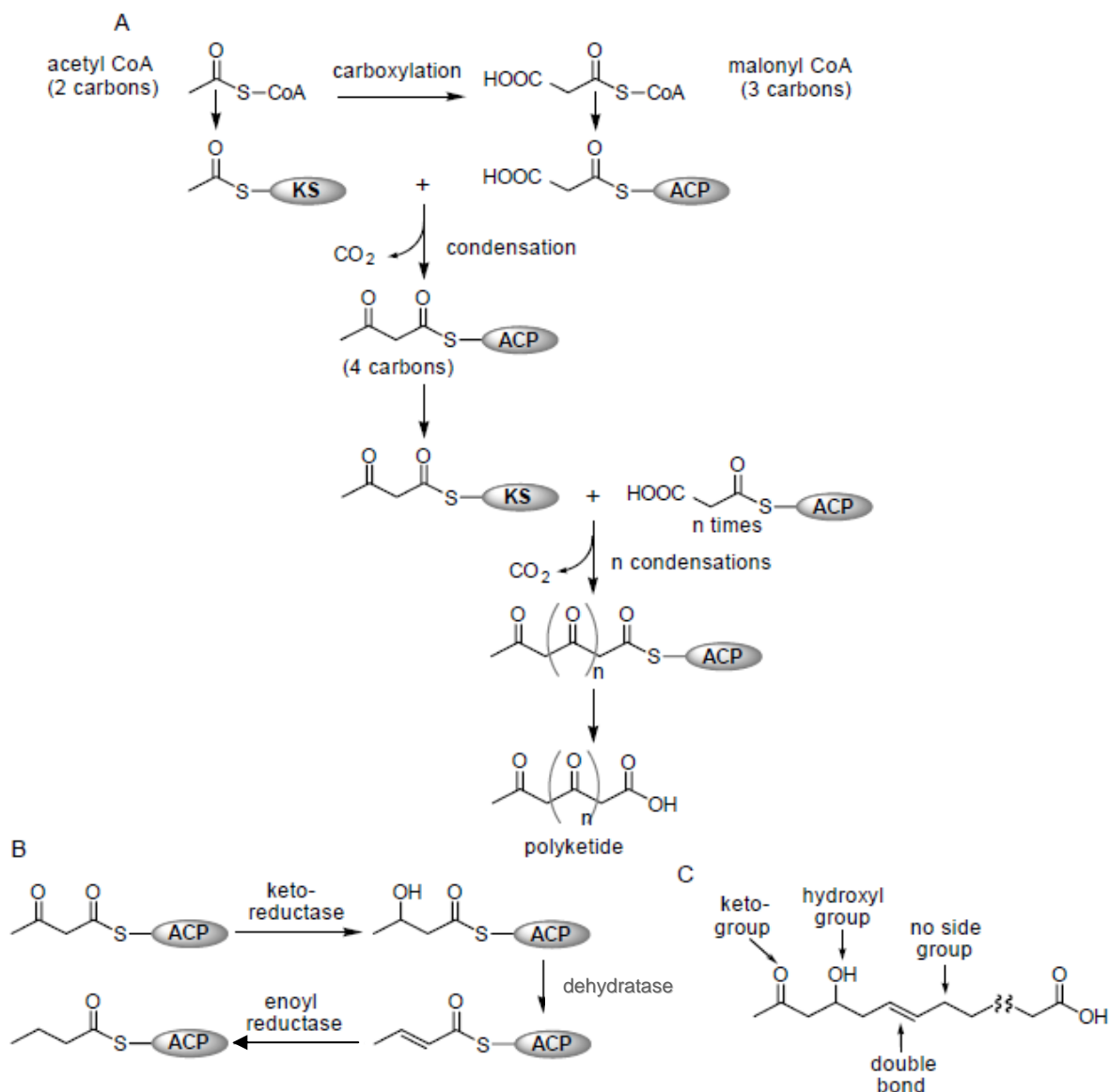
organisms [150]. The biological activities associated with polyketides encompass e.g. antibacterial, antiviral, antitumor, antihypertensive activities, as well as immunosuppressant and mycotoxin compounds.

## **2.1 The molecular background to understand polyketide biosynthesis**

Independent of their structural diversity all polyketides have a common biosynthetic origin. They are derived from highly functionalised carbon chains whose assembly mechanism has close resemblance to the fatty acid biosynthetic pathway [151]. The assembly process is controlled by multifunctional enzyme complexes called PKS [150]. The core of the PKS function is the synthesis of long chains of carbon atoms through repetitive Claisen condensation reactions of small organic acids (such as acetic and malonic acid) via a ketosynthase (KS) enzyme activity. The building units, acetate, propionate, malonate or methylmalonate, are activated units in form of coenzyme A (CoA) esters, such as acetyl-CoA and malonyl-CoA, before involvement in the assembly of the polyketide chain. The most common starter-unit acetyl-CoA with two carbon atoms is condensed with a malonyl-CoA, with three carbons, to give a chain of four carbon atoms with loss of one carbon dioxide. Only two carbons are included into the chain in each round of condensation with malonyl-CoA (see Figure 6). If the extender unit is methylmalonyl-CoA, the “extra” carbon forms a methyl side branch to keep the original extension speed in the main chain [152]. Each condensation is followed by a cycle of optional modifying reactions that involve the enzymes ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) in the subsequent reduction steps. At this stage, a major difference between fatty acid and polyketide biosynthesis becomes apparent. Fatty acid synthases (FAS) catalyze the full reduction of each  $\beta$ -keto moiety prior to further chain extension in every cycle. The polyketide biosynthesis, however, shows a higher degree of complexity due to full or partial omission of reduction steps following condensation [23] and thus affecting function:  $\beta$ -keto (no reduction),  $\beta$ -hydroxy (keto reduction), enoyl (keto reduction and dehydration), to alkyl (keto reduction, dehydration and enoyl reduction).

This control of  $\beta$ -keto reduction is the key feature of the reducing PKS (R PKS) that differentiates these enzymes from FAS and that leads to a great structural diversity among polyketide compounds [153, 154]. The ability to use different chain starter-units (such as acetate, benzoate, cinnamate, and/or amino acids) and alternate extender units (malonate, methylmalonate and ethylmalonate) by bacterial synthases gives rise to further structural diversity among the polyketides. The assembled polyketide chain can also undergo further

modifications such as cyclization, reduction or oxidation, alkylation, and rearrangements after release from a PKS [23].



**Figure 6. The chemistry of polyketide chain assembly.**

- After activation as the corresponding coenzyme A (CoA) esters, acetyl and malonyl units are attached by specific acyl transferases to components of the polyketide synthase (PKS). The starter-unit acetyl-CoA is attached to the active site (cysteine thiol) of the ketosynthase (KS), which catalyzes condensation, and the extender unit malonyl-CoA to a thiol residue of the acyl carrier protein (ACP). One carbon from malonyl-CoA is lost as carbon dioxide during the condensation to yield a four-carbon chain attached to the ACP. After transfer back of the saturated chain from the ACP to the KS, the cycle is then repeated to produce a polyketide chain.
- The three-step reductive cycle that converts a keto group to a hydroxyl, then to a double bond, and finally to a fully saturated carbon.
- A complex polyketide containing keto groups, hydroxyl groups, double bonds and fully saturated carbons at different positions. (Hopwood, 2004) [152].

## 2.2 Types of polyketide synthases

In analogy to the classification of FASs, PKSs have traditionally been subdivided into two main categories [155]. The first category encompasses multifunctional modular systems that

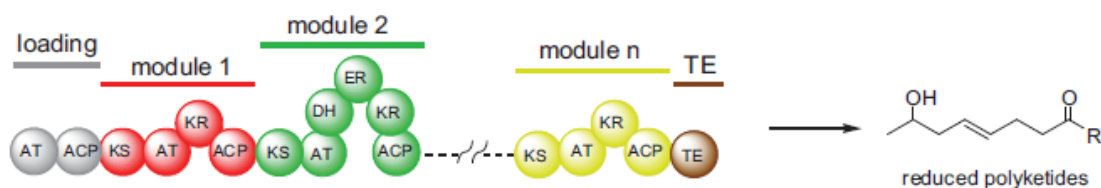
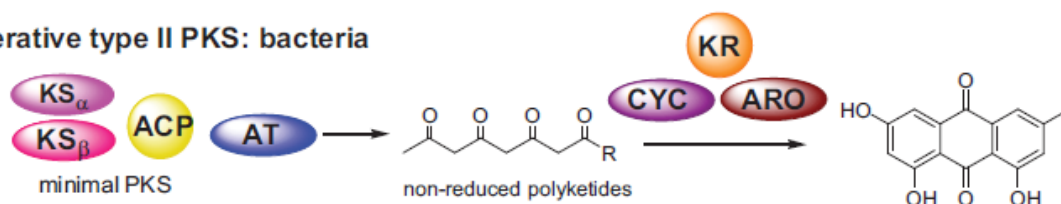
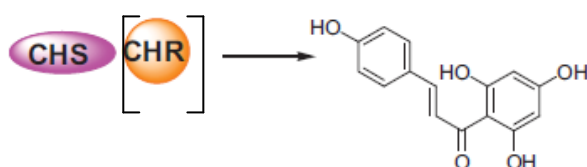
are responsible for the biosynthesis of macrolactones, polyenes and polyethers and are designated as type I PKS [156]. The fully dissociable complex of small, discrete monofunctional proteins that catalyze the biosynthesis of bacterial aromatic polyketides is termed type II PKS [157]. In the past decade, as cloning and sequencing of PKS genes were advancing especially since the discovery of fungal and plant PKSs, these categories of PKSs were redefined and enriched by type III PKS and the expansion of type I PKS into two subclasses, modular type I PKSs and iterative type I PKSs (see Figure 7) [158].

The modular type I PKSs representing bacterial systems are large multifunctional polypeptides arranged in a modular fashion with each module being responsible for one round of chain extension and subsequent  $\beta$ -keto processing. Particularly, each active site in modular type I PKS is used only once during polyketide biosynthesis [156].

The iterative type I PKSs are responsible for the biosynthesis of fungal metabolites such as 6-methylsalicylic acid [45] and lovastatin (**11**) [90, 91]. The iterative type I PKS has only one multidomain protein, in which all the enzyme activities are covalently bound together. The single multifunctional protein is used iteratively to catalyze multiple rounds of chain elongation and appropriate  $\beta$ -keto processing [156].

In the iterative type II PKSs, the active site for each biosynthetic step is encoded in a single gene. There is only one set of a heterodimeric ketosynthase ( $\text{KS}\alpha$ - $\text{KS}\beta$ ) and an acyl carrier protein (ACP), that have to operate a specific number of times to build a polyketide chain in correct length and subsequent cyclisation, reduction and aromatization are performed by cyclase (CYC), KR and aromatase (ARO), respectively (see Figure 7). In some type II PKSs, the malonyl-CoA ACP acyl transferase (MAT), which catalyzes acyl transfer between malonyl-CoA and the ACP, is missing and is possibly shared between the PKS and the housekeeping FAS [159]. The type II PKSs usually catalyze the biosynthesis of a broad range of polyfunctional aromatic natural products and are so far restricted to bacteria [157].

In contrast to the type I and II PKSs that are composed of ketosynthases and accessory enzymes, the type III PKSs are dimers of KS-like enzymes (more precisely homodimers) that accomplish a complex set of reactions, such as priming of a starter-unit, decarboxylative condensation of extender units, ring closure, and aromatization of the polyketide chain, in a multifunctional active site pocket [160] (see Figure 7). Chalcone synthases, the most well-known representatives of this family, are ubiquitous in higher plants and provide the starting material for a diverse set of biologically important phenylpropanoid metabolites [161]. Type III PKSs were traditionally associated with plants but recently discovered in a number of bacteria [162] as well as in fungi [160, 163].

**Modular type I PKS: bacteria****Iterative type I PKS: fungi****Iterative type II PKS: bacteria****CHS-like type III PKS: plants and bacteria****Figure 7 : Different classes of polyketide synthases.**

Modular type I PKSs consist of multi-domain proteins forming a modular unit for each condensation cycle. In iterative type I PKSs, one copy of each active domain, ketosynthase (KS), malonyl-acetyl transferase (MAT), acyl carrier protein (ACP) and optional activities for reduction, ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) are assembled in one protein and iteratively used during the biosynthesis. In contrast, active sites of type II PKSs are encoded in different genes and act in an iterative fashion. Type III CHS-like PKS have a simple architecture like CHS (a homodimer of identical KS monomeric domains) with an optional chalcone reductase (CHR). TE, thioesterase; CYC, cyclase; ARO, aromatase [164].

**2.3 Fungal polyketide synthases**

Up to date, only few fungal PKS genes have been isolated as compared to the large number of isolated bacterial PKS genes. In general, the fungal PKSs are iterative type I enzymes [23]. The non-iterative fungal type I PKS perform only one condensation cycle and result in a production of a diketide (e.g. LDKS). Each one that has been characterised so far is encoded by a gene that resides in a gene cluster, along with a PKS gene encoding an iterative PKS [90, 92]. Recent genome projects for *Neurospora crassa* [165] and *A. oryzae* [166] predict the presence of type III PKS genes in these filamentous fungi [160, 163]. On the other side, fungal modular type I and fungal type II systems have not yet been observed [23]. Unfortunately, the nomenclature in the area is rather confusing with many fungal PKS genes



being named *pks1* or similar. Cox *et al.* brought some order in this area and established a system based on the function of genes (see Table 3).

The minimal domain structure of fungal PKSs consists of KS, AT and ACP domains. The KS domain is the most highly conserved domain in type I PKSs and FASs [167]. The optional  $\beta$ -keto processing reactions may be catalysed by KR, DH, and ER domains, in a stepwise fashion. Further accessory domains are represented by CYC and methyl transferase (MT) activities [168]. The C-methylation takes place during polyketide chain formation due to activity of this intrinsic MT domain, thus it is not a tailoring reaction as it can be in the case of O- and N-methylation reactions governed by distinct enzymes following polyketide assembly [169]. It is important to note that in the fungal polyketide biosynthesis there are no methylmalonyl elongation units employed, in contrary to the numerous examples of the bacterial type I PKSs. As previously mentioned, the fungal iterative PKS can use each active site in an iterative way during chain assembly and determine the degree of reduction and C-methylation within each elongation round. It is fascinating, but a still not resolved mystery, how a single set of active domains determines chain length, degree of reduction, and timing of C-methylation at a particular step in the pathway [168].

**Table 3. Fungal PKS genes.**

Organism	Gene	Synthase component	Final product	PKS type	Reference
<i>Acremonium strictum</i>	<i>PKS1</i>	MOS	3-methylorcinaldehyde	NR PKS	Bailey <i>et al.</i> 2007 [170]
<i>A. fumigatus</i>	<i>alb1</i>	alb1p	YWA1	NR PKS	Watanabe <i>et al.</i> 2000 [171]
<i>A. nidulans</i>	<i>PKST</i>	NSAS	Sterigmatocystin	NR PKS	Minto and Townsend 1997 [172]
<i>A. nidulans</i>	<i>wA</i>	WAS	YWA1	NR PKS	Watanabe <i>et al.</i> 1999 [173]
<i>A. parasiticus</i>	<i>pksA</i>	NSAS	Aflatoxin B1 (1)	NR PKS	Minto and Townsend 1997 [172]
<i>A. terreus</i>	<i>atX</i>	MSAS	6-MSA	PR PKS	Fuji <i>et al.</i> 1996 [46]
<i>A. terreus</i>	<i>lovB</i>	LNKS	Lovastatin (11)	HR PKS	Kennedy <i>et al.</i> 1999 [91]
<i>A. terreus</i>	<i>lovF</i>	LDKS	Lovastatin (11)	HR PKS	Kennedy <i>et al.</i> 1999 [91]
<i>Beauveria bassiana</i>	<i>ORF4</i>	TENS	Tenellin	HR PKS	Eley <i>et al.</i> 2007 [174]

## Introduction

Organism	Gene	Synthase component	Final product	PKS type	Reference
<i>Cochliobolus heterostrophus</i>	<i>pks1</i>	TTS1	T-toxin	HR PKS	Yang <i>et al.</i> 1996 [175]
<i>Cochliobolus heterostrophus</i>	<i>pks2</i>	TTS2	T-toxin	HR PKS	Baker <i>et al.</i> 2006 [176]
<i>Colletotrichum lagenarium</i>	<i>PKS1</i>	THNS	Tetrahydroxy naphthalene	NR PKS	Takano <i>et al.</i> 1995 [177]
<i>Dothistroma septosporum</i>	<i>pksA</i>	NSAS	Dothistromin	NR PKS	Bradshaw <i>et al.</i> 2006 [178]
<i>Fusarium heterosporum</i>	<i>eqiS</i>	EQS	Equisetin	HR PKS	Sims <i>et al.</i> 2005 [179]
<i>Fusarium moniliforme</i>	<i>ORF3</i>	FUSS	Fusarin C	HR PKS	Song <i>et al.</i> 2004 [180]
<i>Gibberella zeae</i>	<i>PKS13</i>	ZS-B	Zearalenone (9)	NR PKS	Gaffoor and Trail 2006 [82]
<i>Gibberella fujikuroi</i>	<i>fum1</i>	FUMS	Fumonisin B1 (5)	HR PKS	Proctor <i>et al.</i> 1999 [42]
<i>Gibberella zeae</i>	<i>PKS4</i>	ZS-A	Zearalenone (9)	HR PKS	Gaffoor and Trail 2006 [82]
<i>Glarea lozoyensis</i>	<i>pks2</i>	MSAS	6-MSA	PR PKS	Lu <i>et al.</i> 2005 [47]
<i>Monascus purpureus</i>	<i>pksCT</i>	CitS	Citrinin (2)	NR PKS	Shimizu <i>et al.</i> 2005 [33]
<i>P. citrinum</i>	<i>mlcA</i>	CNKS	Compactin (10)	HR PKS	Abe <i>et al.</i> 2002 [92]
<i>P. citrinum</i>	<i>mlcB</i>	CDKS	Compactin (10)	HR PKS	Abe <i>et al.</i> 2002 [92]
<i>P. patulum</i>	<i>MSAS</i>	MSAS	6-MSA	PR PKS	Beck <i>et al.</i> 1990 [45]
<i>Phoma</i> sp.	<i>PhPKS1</i>	SQTKS	Squalestatin tetraketide	HR PKS	Cox <i>et al.</i> 2004 [181]
<i>Wangiella dermatitidis</i>	<i>WdPKS1</i>	THNS	Tetrahydroxy naphthalene	NR PKS	Feng <i>et al.</i> 2001 [29]

PKS, polyketide synthase; NR, non-reduced; PR, partially-reduced; HR, highly-reduced; NSAS, norsolorinic acid synthase; WAS, naphthopyrone synthase; YWA1, naphthopyrone; THNS, tetrahydroxynaphthalene synthase; ZS-A, zearalenone synthase A; ZS-B, zearalenone synthase B; CitS, citrinin synthase; MOS, 3-methylorcinaldehyde synthase; MSAS, 6-methylsalicylic acid synthase; 6-MSA, 6-methylsalicylic acid; LNKS, lovastatin nonaketide synthase; LDKS, lovastatin diketide synthase; CNKS, compactin nonaketide synthase; CDKS, compactin diketide synthase; SQTKS, squalestatin tetraketide synthase; TTS1, T-toxin synthase 1; TTS2, T-toxin synthase 2; FUMS, fumonisin synthase; FUSS, fusarin synthase; EQS, equisetin synthase; TENS, tenellin synthase [23, 50].

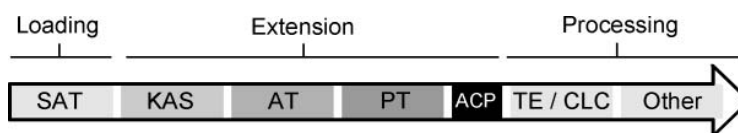
### 2.3.1 Classification of fungal polyketide synthases

According to their architecture and the presence or absence of additional  $\beta$ -keto-processing domains, fungal PKSs are grouped into the non-reducing (NR PKS), partially-reducing (PR PKS), and highly-reducing PKSs (HR PKS) [168]. Recent phylogenetic studies on the basis

of KS amino acid sequences have provided valuable insights into the evolutionary relationship between different types of fungal PKS [169, 182]. Kroken *et al.* showed that amino acid sequences of fungal KS domains cluster according to the degree of reduction of their products into reducing ( $\beta$ -keto reductive domains: KR, ER, DH) and non-reducing PKSs (no  $\beta$ -keto reduction), each type being further divided into four subclades [167].

### ***Non-reducing PKSs***

NR PKS are shown to be responsible for the biosynthesis of non-reduced polyketides such as 1,3,6,8-tetrahydroxynaphthalene, norsolorinic acid (NA) and naphthopyrone (YWA1), that require no  $\beta$ -keto reductive steps during their biosynthesis [23]. In all cases, known genes for these synthases encode type I iterative PKS proteins [50]. The main characteristic of NR PKSs is that they do not contain  $\beta$ -keto processing domains in their multidomain organization. At the N-terminus, a domain is present that appears to mediate the loading of a starter-unit and is thus named starter unit-ACP transacylase (SAT) component (see Figure 8). It is assumed that the starter unit is derived from a corresponding FAS, another PKS or an acyl-CoA. The SAT domain is followed by typical KS and AT domains responsible for chain extension and malonate loading. Beyond the AT there is a conserved domain designated as a product template (PT) with a not yet proven function. Nevertheless, a sequence analysis of this domain suggested that it may be involved in the control of chain-length. The PT domain is followed by one or more ACP domains. Some NR PKS appear to terminate after the ACP, but many feature a diverse range of different domains including Claisen-cyclase-thioesterases (CLC-TE), MT and reductases (R). Although not described in the literature, a sequence analysis of the *M. purpureus pksCT* sequence [33] showed that it has a C-terminal thioester reductase domain. Similar domains were found in the NRPS systems with reductase domains as chain release mechanisms resulting in an aldehyde or primary alcohol. Very recently, by joint efforts Cox and Simpsons demonstrated the role of the terminal reductase domain in product release via heterologous expression of MOS in *A. oryzae* [170]. In sum, it appears that these synthases are equipped with an N-terminal loading component, a central chain extension component consisting of KS, AT and ACP domains with a possible control over a number of extensions, and a C-terminal processing component as shown in Figure 8.



**Figure 8. General domain organisation of NR PKSs.**

The general architecture of NR PKS genes in fungi: SAT, starter unit ACP transacylase; KAS,  $\beta$ -ketoacylsynthase; AT, acyl transferase; PT, product template; ACP, acyl carrier protein; TE, thiolesterase; CLC, Claisen cyclase. There is a possibility that other domains can be included as well at the C-terminus after TE/CLC (e.g. C-MeT; see next figure). According to Cox 2007 [50].

### *Partially reducing PKSs*

Less is known about the enzymology of the PR PKS [50]. The domain structure is much closer to mammalian FASs, with an N-terminal KS followed by AT, and DH domains. A so-called “core” domain follows the DH, and this is followed by a KR domain. A typical PR PKS terminates with an ACP domain as e.g. for MSAS (see Figure 9). The domain structure differs considerably from the NR PKS in such a way that there is no SAT or PT domain, and the PKS terminates after the ACP with obviously no requirement for a CLC–TE domain responsible for offloading of the product. Although a number of PR PKS genes are known from genome sequencing projects, only three genes have been matched to their chemical products - in all cases the tetraketide 6MSA (e.g., a single round of KR and DH) [23].

The first MSAS to be discovered was from *P. patulum*, encoded by *MSAS* [45]. The Ebizuka group have worked with the *atX* gene from *A. terreus* [46] and most recently Tkacz *et al.* have described an MSAS gene (*pks2*) isolated from *Glarea lozoyensis* [47]. Both *P. patulum* and *A. terreus* MSAS form homo-tetramers [50]. A short region of the core domain was identified by Fujii *et al.*; the presence of this region proved to be essential for successful complementation among diverse deletion mutants of *atX* gene. It was hypothesised that this region of 122 amino acids probably forms a motif required for subunit–subunit interaction. Interestingly, this core sequence is present in other fungal PR PKS, and in the bacterial PKS such as *CalO5* from calicheamicin biosynthesis [183, 184].



**Figure 9. Domain organisation of MSAS.**

Domain architecture of MSAS encoded by *A. terreus at X*, *G.lozoyensis pks2* and *P. patulum MSAS*: KAS,  $\beta$ -ketoacylsynthase; AT, acyl transferase; dehydratase (DH); “Core” domain; ketoreductase (KR), and ACP, acyl carrier protein enoylreductase (ER). According to Cox 2007 [50].

### *Highly reducing PKSs*

The HR PKSs is the third class of fungal PKSs that produce complex, highly reduced compounds such as lovastatin (**11**), T-toxin, fumonisins B1 (**5**) and squalestatin. These PKSs

have an N-terminal KS domain, followed by AT and DH domains. In many cases, the DH is followed by a MT domain. Some HR PKSs possess an ER domain, in others there is a roughly equivalent length of sequence without known function. An ER domain is succeeded by a KR domain, and finally the PKS often terminates with an ACP (see Figure 10). The *lovB*, gene that encodes for LNKS involved in lovastatin biosynthesis, appears to encode one part of an NRPS condensation (C) domain immediately downstream of the ACP. It was proposed that this domain plays role in product release. In general, in HR PKS there seems to be no domains that are similar to the PT or SAT domains of the NR PKS, as well as no “core domain” of the PR PKS [50]. PKS genes of all three classes identified up to now are presented in Table 3.



**Figure 10. General domain organisation of HR PKSs.**

General domain architecture of HR PKSs: KAS,  $\beta$ -ketoacylsynthase; AT, acyl transferase; dehydratase (DH); C-MeT, methyltransferase; enoylreductase (ER) as optional; ketoreductase (KR), and ACP, acyl carrier protein. According to Cox 2007 [50].

## 2.4 Current challenges in the fungal polyketide research area

The number of genes and chromosomes, as well as sizes of fungal genomes are not yet known exactly [185]. The variability of intraspecies ploidy levels is an important consideration in fungi, while it can lead to variations in estimation of genome size and chromosomes number. Compared with bacteria, fungi have larger genomes: ~90% of the available fungal data are within the range of 10–60 megabases (Mb), with an average of ~37 Mb [186]. Generally, fungal genomes are compact, showing a high gene density, i.e. one gene in every 2.1 – 3.6 kb of a genome [187-189]. Introns are usually small in fungal genes, ranging from 50 to 300 bp in the majority of cases. In *Saccharomyces cerevisiae*, introns are rare; only 4% of genes are interrupted by introns [187]. The frequency of introns in other fungi is considerably higher; as much as 43% among *S. pombe* genes [188]. The splice sites resemble general eukaryotic consensus sequences [190]. Thus, the determination of gene structure is relatively straightforward in *S. cerevisiae* due to its high gene density and low frequency of introns. In contrast, the identification of genes within the genomic sequences of other fungi has proven to be much more challenging [185].

### 2.4.1 Accessing DNA sequences of fungal genomes

Over 40 fungal genomes sequences are currently publicly available with over 40 additional projects underway. These genomes represent important human pathogens, plant pathogens,

saprophytes, and model organisms from all four major fungal groups i.e., ascomycetes, basidiomycetes, zygomycetes, and chytrids [191].

At the very beginning it is important to note that extraction of genomic DNA represents an essential step for the molecular analyses of fungi. The standard method to prepare fungal DNA consists of lyophilization of mycelia, disruption of cell wall by grinding, extraction of DNA in buffers containing detergents like sodium dodecyl sulfate, removal of proteins with a mixture of phenol and chloroform, and precipitation of DNA with 2-propanol [192]. Such a method is particularly suitable to obtain a large amount of pure DNA usually needed for the preparation of genomic libraries within the course of sequencing of fungal genomes. However, it is time consuming and is associated with labor intensive method, furthermore, it generates toxic phenol and chloroform waste [193]. Therefore, developing rapid and simple methods for DNA preparation on a small scale are goals of researchers in this area ever since. Although many current fungal DNA extraction procedures eliminate the contaminants that commonly inhibit the PCR, the resulting protocols are still laborious and costly [194]. Thus, the fungal research community still needs to work on protocols for genomic DNA isolation that meet with efficient procedure and purity of isolated DNA, both of which is needed for PCR-based applications, screening transformants to obtain isolates with a targeted gene modification or for the preparation of fungal genomic libraries.

Initially, eukaryotic genomes were sequenced based on mapped, large-insert clones, e.g., cosmids and Bacterial Artificial Chromosomes (BACs). Later on, whole-genome shotgun-sequencing has proved to be a more efficient approach for sequencing of fungal genomes [185]. Despite this improvement, such projects still leave behind hundreds of clone gaps even after multiple coverage of the genome [165, 185]. Some gaps may result from DNA sequences that are not clonable in *E. coli* or present difficulties for conventional sequencing reactions. A sequencing of additional clones from different genomic libraries with different technologies will be necessary to improve genome assemblies. However, all fungal genome-sequencing projects usually exclude reads that cannot be assembled. Many of these unassembled reads are repetitive sequences often associated with telomeres, centromeres, and ribosomal DNA (rDNA) repeats. Robust automated methods are needed to include non-mitochondrial excluded reads into genome assemblies [195]. These facts indicate that the sequencing of fungal genomes itself represents a challenging point in this area of research. In addition, the data of many fungal genome projects are not yet publicly accessible [185].

### 2.4.2 Targeting PKS genes from fungal genome DNA sequence

Besides whole genome sequencing there are other methods to access genes of interest within a particular fungal genome. A partial sequencing approach can be used as well in order to target PKS genes in microorganisms. In this approach, randomly chosen small fragments (~500 bp) of uniformly sized random genomic library (~1 kb fragments) are subject to sequencing in order to identify PKS genes within that genomic library. It is clear that this approach is only effective when the target PKS gene cluster comprises a significant portion of the genome (0.4-1%) as in the case of modular PKS clusters (40-150 kb each) within relatively small-sized bacterial genomes (average 4-10 Mb). In opposite to this, the iterative fungal type I PKS genes are only 6-8 kb in size, and the fungal genomes are at least twice that of an actinomycete, making it unlikely that this library-sequencing method could be efficient [23].

Thus, for finding fungal PKS genes it is the usual approach to establish and screen the genomic (cosmid or BAC) as well as the complementary (cDNA; e.g. phage) DNA libraries via Southern hybridization with heterologous or homologous probes and/or by PCR with degenerated primers [168]. These methods are facing several challenges, at first due to the quite low sequence conservation among fungal PKS genes. Consequently, the use of PKS probes derived from bacterial sources is not generally successful [23]. Even more important, fungal PKS genes which are not functional homologues are rather highly divergent to serve as probes for cloning novel PKS genes [182]. The strategy of designing degenerate PCR primers based upon the extent of reduction required for the formation of the fungal polyketide product (see 2.3.1), as well as the use of primers based on the methylation domain, appears to overcome some of the previously mentioned limitations [23, 169].

Moreover, when targeting a gene cluster responsible for the production of a particular fungal polyketide, many other PKS genes present in a producer strain may complicate the isolation of a wanted cluster. Up to date many PCR methods have sampled different templates including fungal genomic DNA and cDNA generated at different growth stages. A typical filamentous fungus has the genetic potential to produce as many as 10 structurally diverse polyketides [23]. Besides, more putative PKS gene clusters may be present than expected based on the number of known products, and some of these clusters might not be expressed under laboratory conditions [1]. A good example is the fungus *Neurospora crassa*, which was thought to produce only one polyketide, i.e. the polyketide-derived pigment melanin. Surprisingly, seven putative PKS genes were identified via the whole genome sequencing project, but only one of which was a melanin-type PKS gene [165].

The use of automated and manual annotations of fungal genomes provided with predictions of gene functions based on the homology to already characterised genes and their products. This procedure has revealed that many fungal genomes possess even larger numbers of putative *pks* genes, e.g. 15 in *Gibberella moniliformis* and *Gibberella zeae*, 20 in *Botriotynia fuckeliana*, 25 in *Cochliobolus heterostrophus* [167], 23 in *Magnaporthe grisea* [196], 30 in *A. oryzae*, 14 in *A. fumigatus*, and 27 in *A. nidulans* [1] (etc.). Up to date, the experimental settings have shown that it is not manageable to identify all PKS gene clusters from a particular fungus solely by use of PCR and/or Southern hybridization methods.

### **2.4.3 Bioinformatic analysis of accessed PKS genes**

Once a DNA sequence has been obtained - whether the sequence of a single cloned fragment or an entire chromosome - various methods can be employed to locate the genes that are present. The computational approach involves the inspection of the sequence by the use of specialised computer programs to look for special sequence features in genes. The second approach involves those methods that locate genes by experimental analysis of the DNA sequence. The sequence inspection via bioinformatic tools can be used to locate genes because genes are not random series of nucleotides but have distinctive features. This is a powerful tool and is usually the first method applied to analyse a new sequence from a certain genome [197], since the identification of gene coding sequences is an immediate goal in every genome sequencing project.

Several software tools have been used for the prediction of genes in fungi. However, in particular programs for the prediction of genes in filamentous fungi are not yet widespread. Among the often used programs are GeneID [198], FGenesh and FGenesh+ [199], SNAP [200] and GeneWise [165]. Although nowadays bioinformatic programs trained on eukaryotic organisms rarely miss completely one fungal gene, occasionally these programs artificially fuse two closely linked genes. For such programs it is especially difficult to match very small fungal genes [185]. However, many small genes have been shown to be transcribed in *S. cerevisiae* [201], and several are already known in *Neurospora crassa* as well [185, 202]. Additionally, the presence of introns within a genome can be a major challenge for screening, heterologous expression, and functional studies of fungal genes [168]. In brief, given the significant differences in the characteristics of exons and introns between fungi, the training of gene prediction tools on organism-specific data would be of great importance [191].



Annotated databases provide with data on proteins predicted by gene sequence analysis. The functional properties of proteins are deducible from a variety of database searches. The most reliable prediction of protein function is based on a high sequence identity to known proteins covering the entire sequence length [185]. In contrast to bacteria, only insufficient information on functions and cellular locations of fungal proteins are available in public databases. For illustration purpose, only one third of genes identified in the genomic sequence of *Neurospora crassa* represent novel genes lacking readily detectable relatives in accessible databases. This rather large portion of orphan genes is an indication for the small sampling of known genes from filamentous fungi [185]. As a consequence, gene prediction in fungi relies considerably on *de novo* gene prediction that is rather not reliable taking into account that significant discrepancy exists between different automated gene prediction systems used even on a same genome [195].

As with many other eukaryotes, an additional factor that complicates gene annotation in fungi, as well as experiments with fungal DNA, is the occurrence of alternative splicing [191]. The manual annotation could improve the accuracy of automated annotation, but this is time-consuming and labour-intensive with regard to the amount of data generated by a single genome-sequencing project. Apparently, sequencing and analysis of more fungal genomes, especially of closely related species, would open the door for more intensive comparative gene prediction that could correct automated annotation as well [195].

In general, the prediction of PKS gene clusters is based on protein domain analysis. Within proteins of a particular biosynthetic pathway, such domains are well conserved and provide with information on protein function as well as on types of produced metabolites. In the case of modular type I bacterial PKS systems, one can literally “read” the sequence of domains in a deduced protein sequence (from a gene DNA sequence) and infer the structure of the produced polyketide [203]. On the contrary, it is not possible to predict the structure of any polyketide of an iterative type I fungal PKS by analysis of domain and motif structures alone. However, on then basis of solely protein domain structure analysis, numerous PKS genes have been identified from fungal genomic DNA sequence data so far [166]. The rapid progress in this area was facilitated by what is now considered as attribute of secondary metabolic biosynthetic pathways – the grouping of pathway genes in a contiguous cluster [1]. Despite this fact, there are many PKS and associated genes of certain metabolites that are still not found. An interesting example is orsellinic acid, the simplest tetraketide that requires no reduction steps during its biosynthesis and that is produced by many fungi in the genera

*Aspergillus* and *Penicillium* [50]. Despite the early work (1968) [204] with the protein – orsellinic acid synthase (OSAS), a corresponding PKS gene has not been discovered yet [50]. Furthermore, the homology search for oxidoreductase, methylase, acetylase and esterase genes as putative members of such clusters is challenging, since genes for these so called “tailoring” enzymes are not exclusive to secondary metabolism [1]. Even though bioinformatics nowadays can give first information on putative protein function, the functional study of putative gene clusters is necessary in order that their functions do not remain speculative.

### **2.4.4 Functional studies of putative genes**

At this point, research obviously turns again towards the experimental analysis of fungal DNA sequence. This process of identifying a gene’s function subsequent to describing its DNA sequence is known as reverse genetics [205]. The comparative analysis using the GenBank database can designate putative functions by association to earlier identified genes, but only the characterization of mutants can offer a definitive answer to gene functions as mentioned previously [206].

The classical fungal genetic approaches for gene identification rely on the inactivation of genes in order to receive recognizable phenotypes. A number of methods have been explored for this purpose including homologous recombination, insertional mutagenesis, and gene silencing. Up to date, in most fungal systems knockout mutations are generated by promoting the homologous recombination of mutated gene construct with the genomic wild-type sequence by using an appropriate transformation system [205]. One of the major problems for functional gene analyses is that the integration of transformed DNA can occur in a non-homologous manner as well [168]. The frequency of homologous integration varies greatly between different organisms and depends on the length of homologous DNA. The insertional mutagenesis is an alternative method to gene knock out and is used for large-scale reverse genetics. It is based on the insertion of foreign DNA into the gene of interest by using mobile elements of various origins *Agrobacterium tumefaciens* T-DNA, transposable elements, retrotransposons, restriction-enzyme-mediated integration (REMI), etc. [205]. In addition to gene disruption or deletion, RNA-mediated gene silencing represents another genetic tool that can be helpful for exploring gene functions in fungi [168]. This method represents a post-transcriptional gene silencing (PTGS), in which synthesised double-stranded RNA triggers degradation of sequence-homologous mRNA. A disadvantage of this method may be that homologous genes are silenced simultaneously [207-209].

The success of such methods relies on a set of tools to facilitate these genetic manipulations. Such a tool set normally comprises of fungal vectors and transposons, transformation technology, means of mutagenesis, selection markers, controllable promoters, etc. It is important to point out that rates of protoplast regeneration, transformation frequency, and the degree of gene targeting are all variable depending on the fungal species and even the genetic background of a particular strain [206]. The application of molecular genetics has been adopted in fungal research quite slowly because of the need to develop protoplasting techniques, transformation systems, and appropriate vectors. These are still the most important restrictions for transformation of many fungal species [205].

Traditionally, fungal transformation relies largely on variations of the same basic protocol: Use of protoplasts, cell wall degradation with lytic enzymes and permeabilisation of cell membranes with polyethylene glycol (PEG) and calcium chloride [210]. It is worth to note that the main obstacle in any transformation system is the fungal cell wall [205]. Beside, the process of protoplast generation could be very tedious mainly due to the fact that cell-wall-free fungal cells are osmotically sensitive and therefore must be handled with care. Furthermore, independent of the transformation method used, transgenic line selection is one of the most important limitations. Commonly used selection markers for transformants are either nutritional markers, which complement an auxotrophic requirement (e.g. *pyrG*, *amdS*, *argB*, *trpC*, *niaA*) or dominant antibiotic resistance markers against hygromycin B, phleomycin, bialaphos (BASTA), sulfonyleurea, and benomyl [168]. The main advantage in the use of selection markers (that confer drug resistance) is that it is not necessary to have prior knowledge of the genotype of the fungus subjected to transformation [205]. Nevertheless, in some cases, fungi show natural drug resistance and consequently, a high drug concentration is required and has to be defined.

An important aspect of fungal transformation is the fate of DNA introduced into the fungal cells. In fact, in order to obtain stable transgenic lines, the transforming DNA has to be transferred to the daughter cells after mitotic and/or meiotic divisions [205]. There are two alternatives in order to achieve this condition: Either exogenous DNA (transforming) should replicate in an autonomous way, so vectors containing an origin of replication are required; or stable integration of the exogenous DNA into the chromosomes should occur via homologous recombination as briefly mentioned before.

Autonomously replicating vectors are commonly derived from mitochondrial plasmids or contain autonomously replicating sequences (ARSs). Vectors with ARSs promote a high transformation frequency when used in *Saccharomyces cerevisiae*; for this reason they

represent powerful tools in the isolation and cloning of yeast genes [211]. However, ARSs arising from yeast do not seem to work in filamentous fungi [205]. Even though many efforts have been made toward generation of truly autonomously replicating vectors in fungi, none of these attempts were clearly conclusive [212-215]. Partly this may be due to the limited occurrence of ARSs among filamentous fungi. An alternative explanation could be that the known plasmids used with ARSs do not contain telomeric and centromeric sequences of adequate size, so they can not behave as true chromosomes [205]. Furthermore, centromeric sequences that provide stability to the replicating vectors in filamentous fungi have not yet been isolated to best of my knowledge. Thus, the fungal vector technology remains relatively unchanged with integrative vectors being prevalent for filamentous fungi [206].

### **2.4.5 Industrial production of fungal polyketides**

Since many of fungal polyketides show impressive bioactivities (see sections 0-1.4) and could serve for drug development, there is of course interest to produce these polyketide metabolites in large-scale fermentation processes [216]. Even though the functional analysis of PKSs cluster and a confirmation of gene (or genes) identity could be successfully accomplished, a production of polyketide by the original host might be still hampered by the requirement for sophisticated growth conditions, inefficient production or/and large biosynthetic background of the host [168]. These drawbacks are obstacles for the isolation and work on particular metabolite. Consequently, some universal fungal PKS gene expression systems are highly desirable.

In sum, at this point the fungal research community is facing many additional challenges. Explicitly, the reconstruction of polyketide biosynthesis in heterologous hosts demands that:

- Large multienzyme assemblies be functionally expressed,
- their posttranslation modification needs to be adequately met,
- their substrates be available *in vivo* in sufficient quantities at adequate time point, and
- the producer cells be protected against the toxicity of the biosynthetic product [216].

The very first challenge is undoubtedly the proper assembly of all enzymatic functions on the DNA level, and this can be tedious work due to the presence of introns and strain-specific splicing mechanisms [168]. The formation of a biologically active polyketide often requires the activity of various “tailoring” enzymes acting on a PKS-derived intermediate in order to yield the final natural product. In most cases, heterologous expression of these monofunctional enzymes is relatively straightforward, although co-substrate availability can be an issue [216]. In addition, the ACP domain itself requires posttranslational

4'-phosphopantetheinylation catalysed by 4'-phosphopantetheinyl transferase (PPTase). This enzyme can be quite specific, which makes it necessary to co-express a species-specific PPTase together with a PKS in order to obtain a functional holo-ACP [168]. Besides, the pathway specific regulatory genes (i.e. narrow transcriptional factors) that positively regulate expression of genes in a cluster are also embedded in PKS gene clusters and have to be heterologously expressed as well [1].

Given the potential cytotoxicity of many bioactive polyketides, transmembrane proteins are required for their export out of the cells. Although putative export proteins are often found associated with PKS gene clusters, very little is known about their mechanism or selectivity [216]. If the goal is heterologous production of one biologically active polyketide, then a resistance mechanism capable to abolish effect of a natural product on the heterologous host has to be considered as well. Such self-resistance genes have been already found within the natural PKS gene clusters [217, 218]. However, coexpression of these genes adds another level of complexity to the heterologous production of polyketides and can be particularly challenging in cases where novel polyketides are engineered for which no known resistance mechanisms have been identified. Consequently, a large size of the gene locus encoding particular biosynthetic pathway in question can lead to instability of expression construct with so many genes [168].

In the end, it is important to emphasise that beside the already mentioned narrow (i.e. specific) pathway regulators, broad regulatory factors are necessary in order to ensure that polyketide pathway can respond to the demands of general cellular metabolism and the presence of specific pathway inducers [216]. Broad transcriptional factors are usually located outside the gene cluster and can either positively or negatively regulate the metabolite production. They are conserved in all fungi and other eukaryotes. Apparently, closer knowledge on these factors could increase chance for successful heterologous expression of one PKS-derived metabolite.

### **2.5 The perspectives in fungal polyketide research area**

As a consequence of recent fungal genome sequencing projects, the data on as-yet unexplored putative PKS gene clusters has been considerably enlarged [167, 168, 219]. The opportunities that have been opened for work in the PKS research area by means of “combinatorial biosynthesis” and “metabolic engineering” approaches are more than impressive. Combinatorial biosynthesis makes use of more than one biosynthetic pathway and is able to generate hybrid polyketide products with improved or even novel bioactivities.

Additionally, metabolic pathway engineering is an approach to optimise production of a single polyketide in a suitable heterologous host [220]. Based on recent pioneering work it seems to be manageable to engineer fungal polyketide biosynthesis pathways as well [168]. The first experiments that involved domain swapping have been recently done by Du *et al.* in order to closely investigate programming in HR PKS systems [221]. Very recently, Hertweck and collaborators managed to express a silent PKS-NRPS gene cluster from the genome of *A. nidulans* via ectopic expression of a specific regulatory gene [222]. Nevertheless, very few conclusions have been drawn from such experiments and many more questions have been arisen so far.

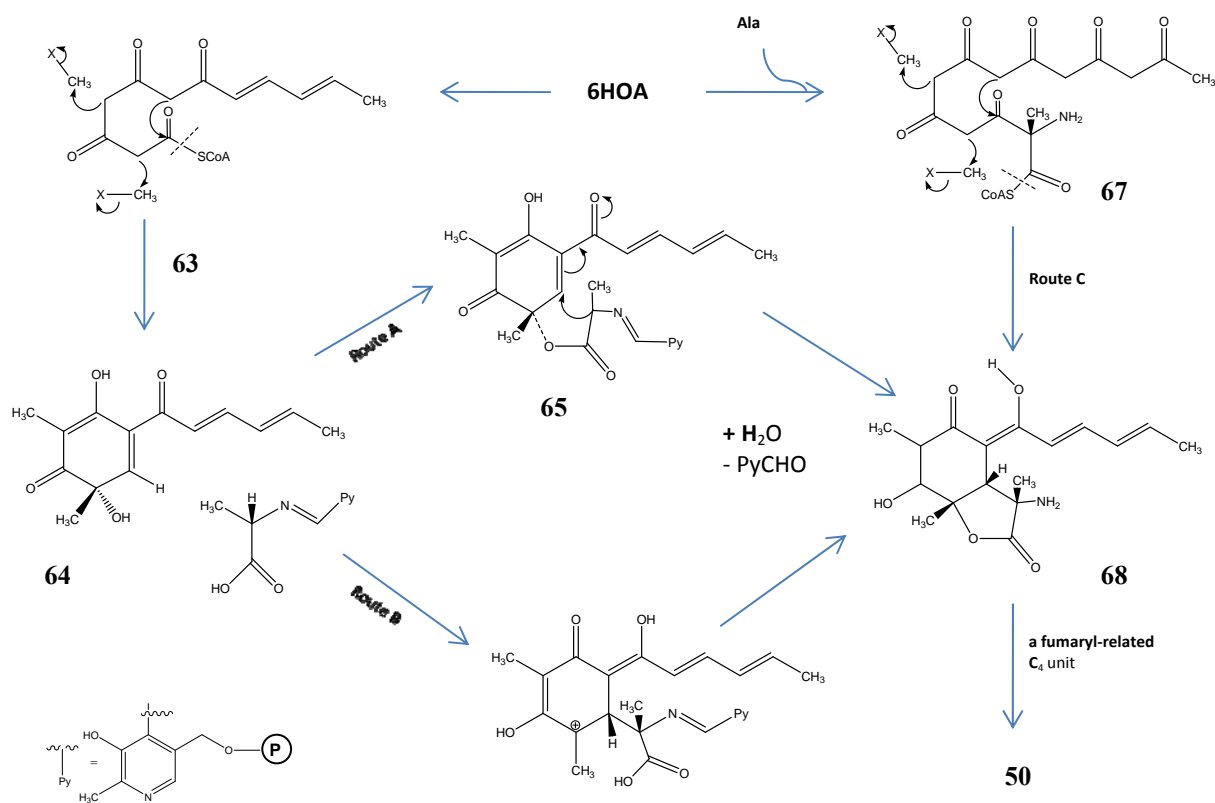
Even though many molecular tool kits for fungi have been developed over the last few decades, it will still take a while until fungal PKS research can follow up the speed at which achievements are currently accumulating for bacterial PKS systems. Apparently, the monomodular-iterative nature of fungal PKS systems, together with absence of sexual systems for many producer fungal species, is regarded as constantly challenging genetic work on these gene cluster systems. In the end, taking in account the moderate size of the fungal research community, tool development and improvement remain a constant challenge in order to clarify many fundamental aspects of fungal PKS systems.

### 3 Biosynthesis of sorbicillactone A

As a part of our collaboration with “Center of Competence *BIOTEC*marin” we were provided with the *P. chrysogenum* strain E01-10/3 (IFM-GEOMAR collection). The interest of our group was the identification of the genetic background for sorbicillactone A (**50**) production for which the knowledge on sorbicillactone A biosynthesis is essential. The working groups of Prof. G. Bringmann (Institut für Organische Chemie, Würzburg) and Prof. J. F. Imhoff (IFM Geomar, Kiel) conducted a number of feeding experiments after optimizing the growth conditions of the fungus for the production of sorbicillactone A as the major secondary metabolite [136]. Based on these feeding experiments with [<sup>13</sup>C<sub>2</sub>]-acetate, [<sup>13</sup>C<sub>3</sub>]-L-alanine, and [methyl-<sup>13</sup>C]-L-methionine, a biosynthetic route for sorbicillactone A was postulated (see Figure 11). A key intermediate of this presumed biosynthesis is a twice C-methylated and, thus, branched hexaketide chain (**63**), which is reduced at C-1, cyclised, and oxidatively dearomatised to give sorbicillinol (**64**). To this highly reactive compound, the amino acid alanine is introduced by esterification with the hydroxyl group at C-5 to give the intermediate (**68**). After  $\alpha$ -deprotonation of the alanine portion, the 5-ring lactone could be closed by intramolecular Michael addition (route A, intermediate **65**). Alternatively, the Michael

addition could take place first, leading to the intermediate (**66**), followed by the ring-closing step now through an SN1 type substitution (route B). Both routes seem imaginable, but the cis-fused annulated ring system can be seen as an argument in favour of route A, which leads to this configuration “automatically” as a consequence of the intramolecular C, C-bond formation. In the last step of the proposed pathway, the bicyclic amino lactone (**68**) is converted to (**50**) by *N*-acylation with a fumaric acid related, yet unsymmetric precursor.

Sorbicillactone A does not only possess a unique structure, but is also the first member of a novel class of amino acid derived sorbicillinoid natural products, the “sorbicillinoid alkaloids” [136]. All other sorbicillinoid and bisorbicillinoid metabolites, as yet known, are derived from one or two sorbicillinol molecules with no other precursors involved in their biosynthesis. Therefore, identification of the gene cluster responsible for biosynthesis of bioactive sorbicillactone A, involving introduction of amino acid alanine and one extra acyl moiety, represents a great scientific challenge. In addition, high structural similarity of sorbicillactone A to pharmacologically unwanted compound sorbicillactone B (**51**) leads to the proposal that this two metabolites could originate from the same biosynthetic gene cluster. This further implicates that probably one distinctive enzyme activity (i.e. an oxidoreductase) within a cluster may be responsible for the conversion of sorbicillactone A to sorbicillactone B. This even more increases the interest for identification of the cluster, since it would give opportunity to preclude biosynthesis of sorbicillactone B via knock out of a distinctive enzyme activity within the gene cluster itself.



**Figure 11: Possible biosynthetic routes to sorbicillactone A [136]**

All three putative biosynthetic routes are in more details discussed in the text above. Legend for the chemical structures within Figure 11:

63 Sorbicillactone A hexaketide intermediate, 64 Sorbicillinol, 65 Intermediate of sorbicollactone A, biosynthesis, 66 Intermediate of sorbicollactone A biosynthesis, 67 Intermediate of sorbicollactone A biosynthesis, 68 Intermediate of sorbicollactone A biosynthesis, 50 Sorbicillactone A, Py Pyridoxal phosphate moiety.



## 4 Research goals

As already illustrated throughout the introduction section, fungal secondary metabolites play an important role as drugs and drug leads. Although a number of compounds have been discovered in recent years, the rate of discovery has declined drastically while the need for novel compounds has grown considerably. In recent years the marine environment has been brought into focus as a novel source for bioactive compounds.

The present evaluation is based on the effort to access the potential of marine-derived fungi for production of bioactive compounds. Our investigation focused on the marine-derived *P. chrysogenum* strain E01-10/3 (FM-GEOMAR collection) that was isolated from the Mediterranean sponge *Ircinia fasciculata* within a collaborative project of “BIOTECmarin”. The unusual structure and promising bioactivities of the sorbicillactone A has prompted molecular analysis of the requisite gene cluster.

The main goals of this study are the identification of the gene cluster responsible for biosynthesis of sorbicillactone A and the assessment of the genetic potential of this particular *P. chrysogenum* strain for polyketide production. Several tasks, which are given below, had to be elaborated in order to achieve these goals:

- Designing primers for amplification of partial PKS sequences of different classes of fungal type I iterative PKS gene clusters.
- Sequence analysis of partially amplified PKS sequences via homology search in gene databases in order to predict the putative functions of certain PKS.
- Disclosure of the evolutionary relationships of partially amplified KS domain sequences and evaluation of promising candidates in order to screen *P. chrysogenum* E01-10/3 genomic library for the presence of the sorbicillactone A gene cluster.
- Construction and screening of the genomic library of the sorbicillactone A producer via hybridization and PCR methods.
- Shot-gun sequencing of the fosmid harbouring the potential candidate for gene cluster involved in biosynthesis of sorbicillactone A.
- Detailed sequence analysis of the sequenced genomic region in order to reveal the putative function of the deduced gene products.
- Analysis of the domain composition of the sequenced PKS genes and evaluation of active site motifs for each domain.

- Defining evolutionary relationships of the sequenced *P. chrysogenum* PKS gene(s) in respect to other members of fungal type I PKS systems.
- Proposal for the polyketide routes of sorbicillactone A and sorbicillactone B biosynthesis.
- Feeding experiments with radioactively labelled putative precursors.

The identification of the sorbicillactone A gene cluster would help to specify the biosynthesis of sorbicillinoid compounds and to open the way for future strain optimization through genetic manipulations.

## RESULTS AND DISCUSSION

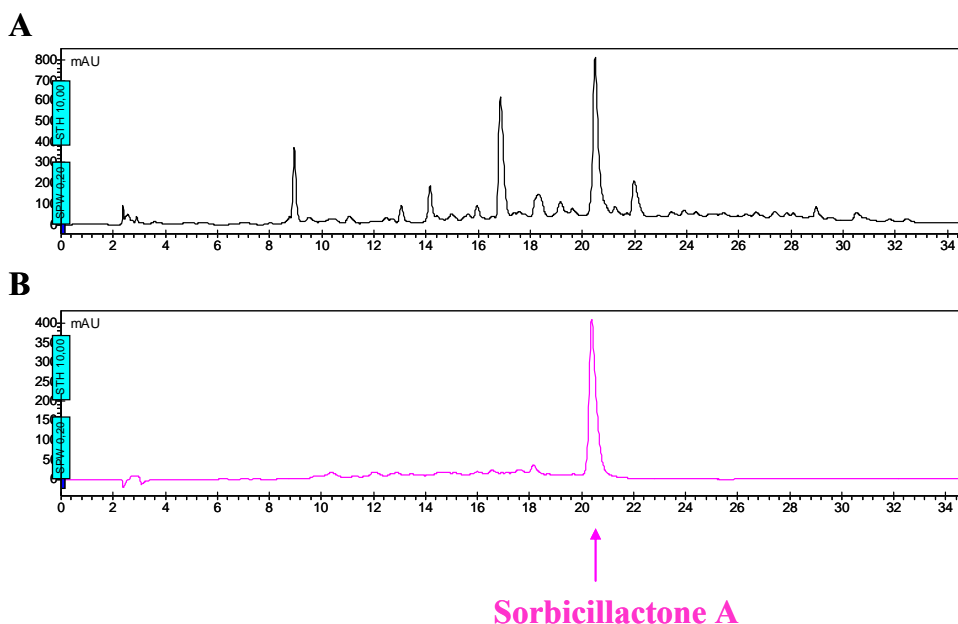
### 5 Detection of sorbicillactone A from *P. chrysogenum* E01-10/3 liquid cultures

In order to confirm that *P. chrysogenum* strain E01-10/3 retained its [135] capability for sorbicillactone A production (see 1.4) it was necessary to re-isolate this compound from the liquid cultures of the fungus. A general rule is that secondary metabolite genes are not constitutively expressed, [223] since they are not required for growth or development of producing organism under laboratory conditions [224]. Although in most cases the production of metabolite is consistent from isolate to isolate in a species [126], to the unfortune of the scientists in this area of research – some formerly active secondary metabolite genes become silent or lost with repeated culturing [223]. The abolishment of metabolite production via inactivation of the gene involved in its biosynthesis and successive restoring of biosynthesis via recovery of the gene by introduction of a wild type gene copy is a common way to prove function of the gene in question (see 2.4.4). Taking these facts into account, production of sorbicillactone A by *P. chrysogenum* strain E01-10/3 is a prerequisite for the localization of the gene cluster in its genome, as well as for future functional studies on the cluster itself.

In order to check for production of sorbicillactone A, malt extract agar plates were inoculated with frozen *P. chrysogenum* spores. Agar cultures were grown 5-7 days before a small piece of mycelium was transferred to a malt extract liquid medium [100]. After 10-14 days of growth, the mycelium facing the liquid culture became yellow which is an indicative for sorbicillactone A production (even though some other metabolites of *P. chrysogenum* have yellowish colour as well, see 1.4). The liquid culture medium was collected via pipetting and the pH was changed to a slightly acidic before extraction with ethyl acetate (for details see 12.20). The extraction was repeated two more times in the same manner before the upper ethyl acetate yellowish phase was collected into a fresh flask. Subsequently, this phase was filtrated through round filter papers and dried in rotary evaporator. Before high-performance liquid chromatography (HPLC) runs, both the sorbicillactone A standard and the culture extract were dissolved in acetonitrile. Both samples were submitted to HPLC under conditions published by Bringmann et al 2003 [100] (see 12.20).

In Figure 12 it is shown that extract from liquid culture of *P. chrysogenum* contained a compound whose retention time was ~21 min in HPLC runs, which correlated to the

retention time of the reference compound sorbicillactone A. Moreover, the UV spectrum of the detected compound correlated to the UV spectrum of the reference compound (see Appendix Figure 42). This was the proof that the fungus is still capable of sorbicillactone A production. Other peaks that can be seen in Figure 12 (under A), may correspond to other metabolites that are produced by the strain (see 1.4) or to their degradation products.



**Figure 12: Production of sorbicillactone A by *P. chrysogenum* E01-10/3 liquid cultures.** **A:** HPLC chromatogram of *P. chrysogenum* E01-10/3. **B:** HPLC chromatogram of pure sorbicillactone A compound that served as a reference.

## 6 Genetic potential of *P. chrysogenum* strain E01-10/3 for polyketide production

### 6.1 Design of PCR primers and cloning of putative PKS gene fragments

The PCR-based molecular biological studies were conducted in order to determine the biosynthetic potential of *P. chrysogenum* strain E01-10/3 for polyketide biosynthesis. The published methods to have access to PKS genes often employ consensus or degenerate PCR with primers designed on conserved sequences found in proteins from specific PKS families. The choice of approach is dependent on the desired outcome: Either selective cloning of a PKS associated with a unique compound, or overall assessment of the total PKS biosynthetic potential of an organism [23]. The structure and incorporation pattern of sorbicillactone A suggested involvement of a fungal type I PKS in the biosynthesis of the polyketide skeleton (see 3). In this study, PCR screenings were applied in order to locate the gene cluster responsible for the biosynthesis of sorbicillactone A, as well as to estimate the number of the PKS genes present in the analysed *P. chrysogenum* strain. The genomic DNA was extracted

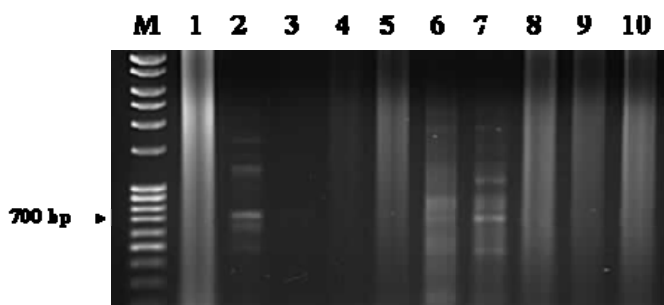
from *P. chrysogenum* (12.5) and used as a template in a number of PCRs with diverse conditions and primer pairs. The PCR parameters like annealing temperature, MgCl<sub>2</sub> concentration and primer concentration were optimised (12.8). To create optimal PCR conditions for the primer annealing step, usually a temperature gradient PCR was set up in accordance to the given or calculated T<sub>m</sub> temperature of primers (12.8.2). The MgCl<sub>2</sub> concentration varied between 1.5 and 4.5 mM (12.8.1). Furthermore, the primer concentration of the degenerate primers ranged from 50-100 pmol according to the degree of degeneration. Specific amplification products of expected sizes were detected by agarose gel electrophoresis, extracted and purified from the agarose gel (12.9 and 12.10, respectively). They were directly ligated into the pGEM-T Easy® or pBluescript vectors (12.12) and transformed into *E. coli* (12.14). The purified plasmid DNA (12.11) was sequenced and cleaned from cloning vector residues using the VecScreen tool available on the web site of the National Center for Biotechnology Information (NCBI; Table 31). All sequences were checked for sequence homologies via Basic Local Alignment Search Tool (BLAST; Table 31), an additional publicly available program on NCBI Homepage. This program compares nucleotide or protein sequences with sequences in a database and calculates the statistical significance of the matches. Particularly, the “BLASTX” variant of the BLAST search was employed for the analysis of PCR-amplified genomic DNA fragments. Via the “BLASTX” nucleotide sequences are translated into their corresponding protein sequences according to a specified genetic code, thus allowing cross-comparisons between nucleotide and protein sequences. The translations were performed in the three forward as well as the three reverse reading frames, so that no possible translation was missed. This approach enabled detection of a number of putative PKS partial sequences from the genome of *P. chrysogenum* as it is illustrated in the following text.

### **6.1.1 KS domain amplification primers**

The KS domain is the most conserved domain among fungal type I PKS [154]. Therefore primers deduced from conserved amino acid motifs within this domain were first employed to amplify homologous KS regions in the genome of *P. chrysogenum*. The primary choice was the LC series of primers published by Bingle *et al.* [182]. The primer pair LC1 and LC2 was designed for the naphthopyrone synthase (WA) subclass of fungal NR type I PKS, while the primer pair LC3 and LC5c was created to suit the MSAS subclass of PR PKSs. The positions of both primer pairs within the KS domain of typical fungal type I PKS and amino acid motifs from which they originate are shown in Figure 15. A temperature gradient PCR

was performed for both pairs of degenerate KS-specific primers in order to determine the optimal temperature for the expected size of fragments. For the LC1/LC2 primer pair, the expected 720 bp PCR amplified fragments (see Figure 13) were electrophoretically separated and cloned into the pGEM-T Easy Vector in order to be sequenced. The optimal annealing temperature for LC1/LC2 primers was 52 °C (see Figure 13, lane 2). The sequence analysis of four single colonies revealed three insert DNA sequences (KSnr-7, KSnr-8 and KSnr-10) with putative homology to the WA class of fungal type I PKS (see Table 5). A closer analysis of these sequences via alignment in the BioEdit program (see Table 31) and comparison of sequences chromatograms via the FinchTV program (see Table 31) suggested that they may be amplified from the same gene locus of the analysed *P. chrysogenum* strain. The minor sequence differences were 3-5% (data not shown). These differences may be due to errors that occurred during automated DNA sequencing, or to differences in the length of analysed DNA. The sequences shared 96-97% identity with the corresponding fragment of the putative WA type PKS of *P. patulum* (GeneBank: Q9Y832). Additionally, they were 88-89% identical to the naphthopyrone synthase (WAS) of *A. nidulans* (GeneBank: Q03149) - the PKS proven to be involved in biosynthesis of naphthopyrone (YWA1) [173].

In the case of the PCR products generated with the LC3/LC5c primer pair, the 680 bp amplicons were purified and cloned in the vector in the same manner as for LC1/LC2 primer pair. The permissive primer annealing temperature for LC3/LC5c primer pair was 57 °C (see Figure 13, lane 7). Only one sequence (out of four sequenced) showed homology to fungal type I PKS sequences. The KSpr-MI sequence (see Table 5) was 94% similar to the putative PKS of *P. patulum* (PKS2; GenBank: P87001) and shared 71% identity with the deduced sequence of the characterised gene for MSAS of *A. terreus* (GenBank: P87162) [46].



**Figure 13: PCR amplification with LC series KS-specific primers - ethidium bromide agarose gel.**

Lane M contains 100 bp ladder as a fragment size standard. Lanes 1-5 are showing temperature gradient PCR obtained with the LC1/LC2 primer pair, with expected size fragment being amplified in the lane 2. The temperature gradient PCR with the LC3/LC5 primer pair is shown in lanes 6-10, and the correct size PCR product is visible in the lane 7. In PCRs for both primer pairs temperature gradient was spanning from 46.0 – 64.0 °C with individual gradient steps being automatically generated by the PCR cycler for the given range of temperatures.

Based on a phylogenetic analysis of KS domain amino acid sequences, Kroken and coworkers have identified so-called NR clade III fungal PKS genes whose activity involves two rare PKS domains: MT and the C-terminal thiolester reductase (R) domain (see 2.3.1) [167]. The research group of Cox designed a pair of degenerate primers, KHKS2 and KHKS3c, based on conserved KS regions belonging to this clade [170]. The authors carefully chose clade-specific positions in order to design primers unlikely to amplify members of other clades. The relative positions of these primers within the typical clade III members are given in Figure 15.

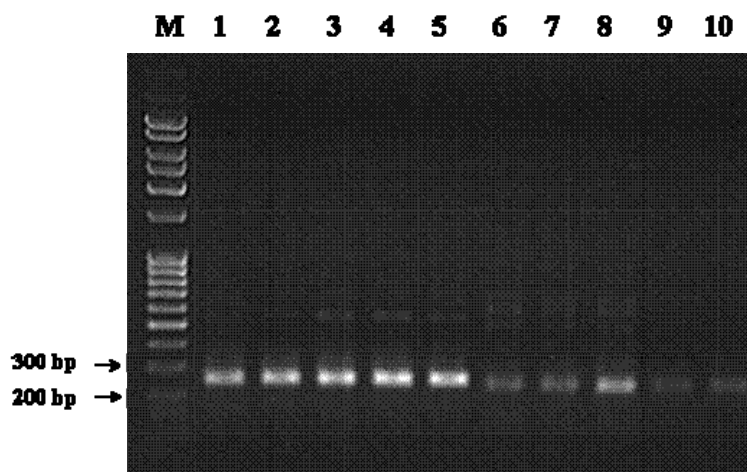
This primer pair was used in order to screen for putative NR PKS clade III members in the genome of *P. chrysogenum*. PCRs afforded the expected 245 bp products that were separated on stained ethidium bromide agarose gel (see Figure 14, lanes 1-5). The PCR products were collected, purified and cloned in a standard manner. As many as 50 clones containing the correct insert size were sequenced. All sequences appeared to correspond to putative PKS genes. They grouped into 10 distinctive KS fragments, initially suspected to originate from 10 different genomic loci. However, after detailed inspection of all DNA sequences (BioEdit and FinchTV programs) it appeared that 4 discrete PKS gene loci have been amplified. Moreover, some of the amplified sequences appeared not to belong to the NR PKS clade III, for which the PCR primer pair was designed. The closest homologues of PKS sequences amplified with KHKS primer pairs are presented in Table 5.

The KHKS1, KHKS9 and KHKS37 sequences shared homology to putative PKS of *A. terreus* (GeneBank: Q0C8A4) with identity values ranging between 72-78% (see Table 5). In case of these 3 sequences, no characterised fungal homologue appeared within the first 50-60 BLASTX matches. In line with detailed sequence analysis in the BioEdit program and the BLAST search results, it was concluded that all three sequences originate from a same locus of the *P. chrysogenum* genome.

Another three sequences (KHKS32, KHKS42 and KHKS46) shared homology to the same fungal type I PKS sequences. They displayed homology to putative PKS of *A. nidulans* (GeneBank: Q5AUX7) with 63-69% identity (see Table 5). Besides, the closest characterised homologue of all three sequences was the citrinin PKS [33] (PKSct; GeneBank: Q65Z23). Further on, the closest homologue of KHKS3 and KHKS44 was putative PKS of *Chaetomium globosum* (GeneBank: Q2GPL2), while KHKS27 sequence shared homology with a putative PKS of *Magnaporthe grisea* (Syn8; GeneBank: Q5XF88). The closest characterised homologue of these three sequences was the CNKS of *P. citrinum* [92] (67-73%; GeneBank: Q8J0F7). Thus, it could be concluded that they were amplified from a

single locus of *P. chrysogenum* genome. Furthermore, the KHKS51 sequence showed 55% and 51% similarity to a putative PKS of *Phaeosphaeria odorum* (GeneBank: Q0UAE2) and *Gibberella moniliformis* (PKS14; GeneBank: Q6RKK1), respectively (see Table 5). No characterised homologue was identified for this sequence in public gene databases.

It is important to note that two *E. coli* fosmid clones from a genomic library of *A. nidulans* F6SCA4 were purchased from Fungal Genetic Stock Center (FGSC), each containing one putative PKS gene of NR clade III. The AN8352 F2 fosmid containing the Q5BEJ6 PKS (later labelled as FOS AN1 F2 fosmid) and the AN8459 B5 fosmid with the Q5B7U4 PKS (later labelled as FOS AN3 B5) were used in a protein alignment of putative NR PKS clade III members (see Figure 15). Moreover, the isolated plasmids were used in PCR with the KHKS series of primers as positive controls for amplification of NR clade III PKS (illustrated for FOS AN1 F2 in the Figure 14). The identity of both controls was confirmed by sequencing of PCR-amplified KS fragments from these two loci.



**Figure 14:** - PCR with the KHKS series of KS-specific primers - ethidium bromide stained agarose gel. Lane M contains 100 bp ladder as a fragment size standard. Lanes 1-5 show temperature gradient PCR obtained with KHKS2/KHKS3c primer pair and genomic DNA from *P. chrysogenum* as a template. Temperature gradient PCR with KHKS2/KHKS3c primer pair is shown in lanes 6-10 with plasmid DNA from FOS AN1 F2 fosmid as a positive control for PKS fragment amplification. In PCRs employing both primer pairs, a temperature gradient was chosen from 44.0 – 54.0 °C with individual gradient steps being automatically generated by PCR cycler for the given rang of temperatures.

Besides, the newly designed KS domain specific primers were made with the intention to increase the chance for amplification of a partial PKS fragment from sorbicillactone A gene cluster and to further examine PKS diversity in the fungus *P. chrysogenum*. These are based on conserved amino acid motifs obtained from an alignment of deduced protein sequences from the putative members of NR PKS clade III genes. The sequences chosen for alignment were those that had similar PKS domain organisation [167] and high sequence similarities to deduced amino acid sequence of the citrinin polyketide synthase (PKSct) of *M. purpureus*, i.e.



the only characterised member of clade III up to that time point. The sequences to be aligned with the PKSct were chosen via standard protein-protein BLAST (BLASTP) search, i.e. a tool to find similar sequences in protein databases. Thus, besides PKSct of *M. purpureus*, eighteen additional sequences from diverse fungal species that fulfilled the given criteria were used in alignments via BioEdit program. One part of this alignment that encompasses the KS domain and was used to design KS-specific primers is shown in Figure 15. Detailed information on used sequences including species origin, locus designations, accession numbers and references are to be found in Table 4.

**Table 4: Sequences used in alignment for members of NR PKSclade III - 19 sequences alignment.**

Organism	Gene/ORF <sup>a</sup>	Protein	Accession number <sup>b</sup>	Shortened name <sup>c</sup>	References
<i>A. nidulans</i> F6SC A4	AN0523.2	Hypothetical protein similar to PKS	Q5BG07	Q5BG07 An	Galagan <i>et al.</i> , 2005 [219]
	AN1034.2		Q5BEJ6	Q5BEJ6 An	
	AN2032.2		Q5BBP8	Q5BBP8 An	
	AN3230.2		Q5B8A0	Q5B8A0 An	
	AN3386.2		Q5B7U4	Q5B7U4 An	
<i>A. fumigatus</i>	AFUA_3602570	Hypothetical protein similar to PKS	Q4WED7	Q4WED7 Af	Nierman <i>et al.</i> , 2005 [225]
<i>A. oryzae</i>	AO090010000114	Hypothetical protein similar to PKS	Q2TXJ8	Q2TXJ8 Ao	Machida <i>et al.</i> , 2005 [166]
	AO09070100831		Q2U7I0	Q2U7I0 Ao	
	AO090001000402		Q2UNE1	Q2UNE1 Ao	
<i>A. terreus</i> NIH 2624	ATEG_03432	Hypothetical protein similar to PKS	Q0CSA2	Q0CSA2 At	Birren <i>et al.</i> , 2005 [226]
	ATEG_08662		Q0CCC2	Q0CCC2 At	
	ATEG_07661		Q0CF73	Q0CF73 At	
<i>Botryotinia fuckeliana</i> B05.10	<i>pks17</i>	Hypothetical protein similar to PKS	Q6RKI3	PKS17 Bf	Kroken <i>et al.</i> , 2003 [167]
<i>Chaetomium globosum</i> CBS	CHGG_10027	Hypothetical protein similar to PKS	Q2GPS7	Q2GPS7 Cg	Birren <i>et al.</i> , 2005 [227]
	CHGG_09586		Q2GR18	Q2GR18 Cg	
<i>Coccidioides immitis</i> RS	CIMG_07081	Hypothetical protein similar to PKS	Q1DRI2	Q1DRI2 Ci	Birren <i>et al.</i> , 2005 [228]
	CIMG_05571		Q1DVU2	Q1DVU2 Ci	
<i>Gibberella moniliforme</i>	<i>pks3</i>	Hypothetical protein similar to PKS	AY495593	PKS3 Gm	Kroken <i>et al.</i> , 2003 [167]
<i>M. purpureus</i>	<i>pksCT</i>	Citrinin polyketide synthase	Q65Z23	PKSct Mp	Shimizu <i>et al.</i> , 2005 [33]

<sup>a</sup> Gene or ORF designation in GenBank; <sup>b</sup> TrEMBL Accession Number; <sup>c</sup> Shortened name of deduced proteins that are used in alignment (Figure 1) and Table 32 – Table 36.

Within three regions of high amino acid similarity (homology regions I-III; see Figure 15), conserved motifs were chosen for degenerate primer design. The forward sorb1-FFK-forw primer was designed from the (N/D)YDTFDHKFFK conserved motif of the homology region

I. Two other primers, sorb2-PTG-rev and sorb3-KPF-forv originate from the PTGQCKPF motif of the KS homology region II. Additional reverse primer, sorb4-GAS-rev, was constructed from the conserved GASGSNA amino acid motif within KS homology region III. The exact positions of designed primers, as well as corresponding amino acid motifs are shown in Figure 15. More details on the conserved protein motifs are given in Table 32 of appendix section. Despite of efforts taken, no cloned fragments from the PCR reactions with these newly design primers correspond to any known PKS or PKS-related proteins in publicly available databases (data not shown).

In sum, 14 sequences from KS domain were amplified with three different degenerate primer pairs (LC1/LC2c, LC3/LC5c [182] and KHKS2/KHKS3c [170]). However, detailed sequence analysis in the BioEdit program and mentioned BLAST search results showed that these 14 sequences actually originate from 6 different loci of the analyzed *P. chrysogenum* genome.

According to the presented BLASTX results, 3 partially amplified KS domains belong to the non-reducing (NR) fungal type I PKSs. The KSnr-7, KSnr-8 and KSnr-10 (i.e. locus KSnr-7) showed putative homology to members of the fungal NR clade I. The KHKS1, KHKS9 and KHKS37 sequences (i.e. KHKS1 locus) as well as the KHKS32, KHKS42 and KHKS46 sequences (i.e. locus KHKS32) shared homology with members of the fungal NR clade III.

The KSpr-MI sequence showed similarity to members of joined bacterial and fungal type I PKS systems with a partially reducing (PR) character. The KHKS51 sequence, showed putative homology to the reducing (R) clade IV in respect to closest match of a BLASTX analysis (see Table 5). Last but not least, the KHKS3, KHKS27 and KHKS44 (i.e. locus KHKS3) shared similarity with members of the R clade II.

Thus, the PCR amplification from the KS domain illustrated that the genome of the analyzed *P. chrysogenum* strain contains 3 main fungal type I PKS systems – NR, PR and R.

## Results and Discussion

**Figure 15: Alignment for members of NR PKS clade III - KS domain.**

19 deduced amino acid sequences of putative members of NR PKS clade III: *M. purpureus* (Mp: PKSct ), *Botryotinia fuckeliana* (Bf: PKS17), *A. nidulans* (An: Q5BG07, Q5BEJ6, Q5BBP8, Q5B8A0 and Q5B7U4), *Gibberella moniliformis* (Gm: PKS3), *A. terreus* (At: Q0CSA2, Q0CCC2 and Q0CF73), *Coccidioides immitis* (Ci: Q1DRI2 and Q1DVU2), *Chaetomium globosum* (Cg: Q2GPS7 and Q2GR18), *A. oryzae* (Ao: Q2TXJ8, Q2U7I0 and Q2UNE1) and *A. fumigatus* (Af: Q4WFD7). The most conserved amino acids in all compared sequences are shaded red; similar amino acid sequences are shaded yellow. Threshold for both identity and similarity shading is 80%, meaning for at least fifteen sequences out of nineteen. The arrows indicate the primer position and direction. The exact amino acid motif is given for each primer from which it was designed. The newly designed primers based on here presented alignment are shown in red. LC and KHKS series of primers for NR PKS are given in green, while LC primers for PR PKS amplification are blue. The KS domain active site motif (DTACSSS) is given in black. Numbers on the right indicate the position of the displayed amino acid sequence within the complete sequence published. The ruler above indicates the position of the extracted alignment in the overall alignment.

		450	460	470	480	490	500	510	520	530	540	550	
		. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	
<b>PKSct</b>	<b>Mp</b>	PGQLLG	GTQKLN	LNTDL	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>PKS17</b>	<b>Bf</b>	PGELLG	THRTNA	FANI	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q5BG07</b>	<b>An</b>	PAPLY	-----	TLQDSG	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>PKS3</b>	<b>Gm</b>	PTSVPS	VPDAG	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q0CSA2</b>	<b>At</b>	S----	AVKTPG	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q1DRI2</b>	<b>Ci</b>	EG---	LQKDQ	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q2GPS7</b>	<b>Cg</b>	SPKDGAL	NPPGSY	TAWSAAR	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q0CCC2</b>	<b>At</b>	VSP--	LLQTPR	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q2TXJ8</b>	<b>Ao</b>	TPP--	LYQTPR	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q4WFD7</b>	<b>Af</b>	DAHQ-	-----	SSPRPL	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q0CF73</b>	<b>At</b>	--SSWMP	-NAPHG	PQQQLQ	---	QRQLP	VEVHTK	PAFDVS	-----	-----	-----	-----	-----
<b>Q1DVU2</b>	<b>Ci</b>	DHNNH	NHNH	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q5BEJ6</b>	<b>An</b>	--SFWL	TPQSS	PPQPQL	QPVLQ	LQQTTR	VEPVM	VPSP	-----	-----	-----	-----	-----
<b>Q2U7I0</b>	<b>Ao</b>	DDDLV	NDNNP	VDDKNP	INDNNT	INGNT	TPVNS	DKPVN	NDDPVN	DDAIA	IVGMS	IKVAG	ADDL
<b>Q5BBP8</b>	<b>An</b>	--RQYDK	GKEP	VHDTESVQ	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q5B8A0</b>	<b>An</b>	-----	-----	QENM	IAVGM	SIKTA	GADD	VDEFA	QMLRT	TGTSQH	QAIM	EDD	INF
<b>Q2GR18</b>	<b>Cg</b>	GGDSNP	DARPVVS	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q5B7U4</b>	<b>An</b>	RDH--	ELATRP	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<b>Q2UNE1</b>	<b>Ao</b>	PPALG	-----	ALLGPI	-----	-----	-----	-----	-----	-----	-----	-----	-----

## Results and Discussion

### Homology region I (KS domen)

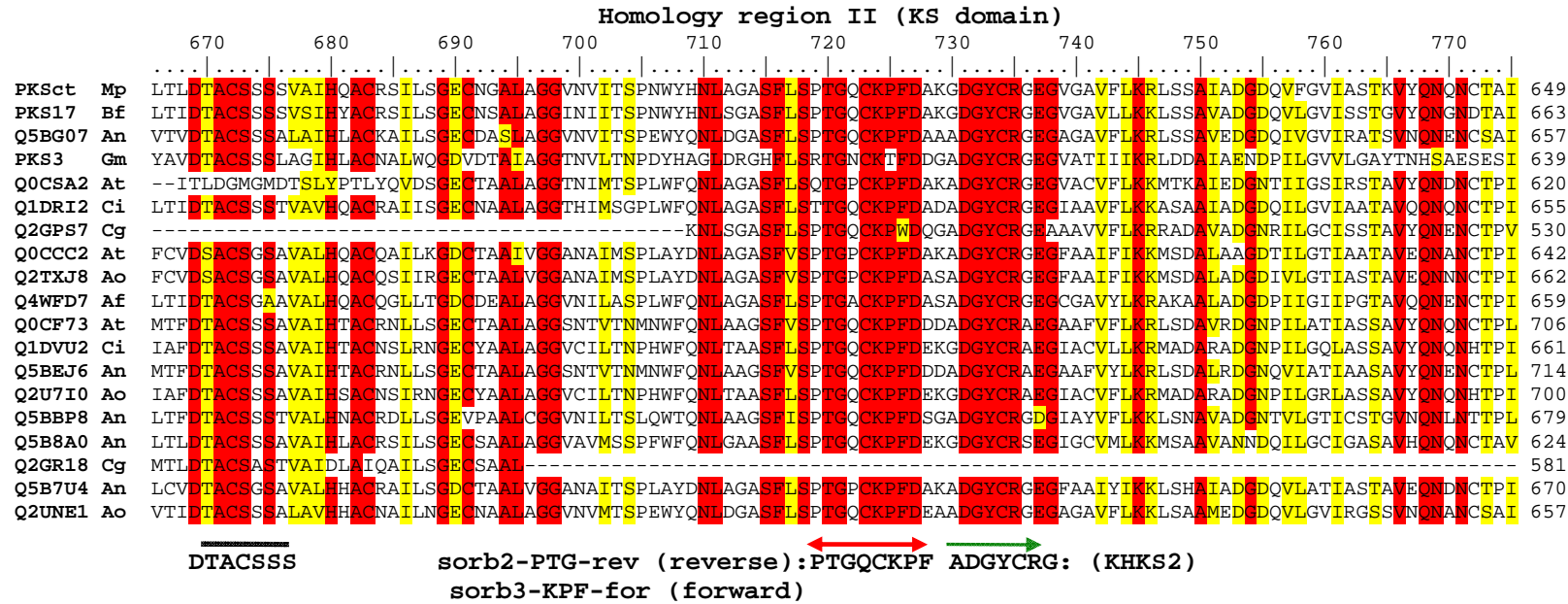
		560	570	580	590	600	610	620	630	640	650	660		
PKSct	Mp	GNFIDNYDTFDHKFFKKS	SPREMASTDPQHR	LMLOVAYQAVEQSGY	FRN--NGTNR	-----	ICCFM	GVGNVDYEDNI	ACYPANA	YSATGNL	KSF	LAGKISHH	FGWTG	539
PKS17	Bf	GNFVENYNTFDHKFFKKS	SPREMASTDPQHR	LMLOIAYQTVEQSGY	FGS--LNADKH	-----	ICCFL	GVGNVDYEANV	ACYPATA	YSATGNL	KSF	VAGKISHY	FGWTG	553
Q5BG07	An	GNFIREYDTFDHKFFKKS	SPREMISTDPQHR	IMLOVAYQAVQSGY	FRN--PGRSKH	-----	VGCYV	GIGVTDYENNV	ACHPPTA	YTATGNL	KSFAAGK	ISHH	FGWSG	547
PKS3	Gm	GCWLDDPSEFDARFFNI	SPREAPQIDPAQR	LALMTAYEAEIQAGI	VDPATPSTRPD	----	RVCIF	YGVTSNDW	METNSAQ	NIDTYFIP	GNRAE	IPGRIN	YFKFS	529
Q0CSA2	At	GNFVRDHDAFDHKFFKKS	SPREATTQDPQR	IFLOSAYQAVEQSGY	FRNSLHADRN	-----	VGVSV							512
Q1DRI2	Ci	GNFVNGHDMFDHKFFKKS	SPRESATMDPQR	HLLOIAYQAVEQSGY	FRHSANPEKN	-----	ICCYM	GVCATDYENNI	IACHAPN	AFSATGNL	QGF	IAGKVSHY	FGWTG	545
Q2GPS7	Cg	GNFIRDHDTFDHKFFKKS	SPREMTSSDPQR	LMLOCAVQAVEQSGY	FHQAKVDT	-----								463
Q0CCC2	At	GNFVRDIDAFDHKFFR	KVPREAMSQDPQR	LLVQVAYQALEQAGY	FQQDDIDKN	-----	ICCYI	ASCTVDYEHNV	NCHPAS	YAATGL	LRSF	LAGKLSHY	FGWRG	532
Q2TXJ8	Ao	GNFVRDIDAFDHKFFR	KVPREAMSQDPQR	LLVQVAYQAVQSGY	FHRPNINRN	-----	ICCYI	ASCTVDYEHNV	NCHPAS	YAATGL	LRSF	MAGKLSHY	FGWRG	552
Q4WFD7	Af	GNFIDDDYDAFDHRFFKKS	SPREAAAMDQQR	LMLOIAYQAVQAGHF	VQDPSRRTRR	----	VGCYI	GVSNVDYENNV	ACHPAN	YSATGL	LRSF	VAGKLSHY	FGWTG	549
Q0CF73	At	CCFFRDGDAFDHKFFKKS	SPREAAAMDQQR	IVLQIAYQALEQSGY	FQEDHTGYTPDGR	DKAHV	VYLG	SCVDYEHNI	SCHDPN	AFTATG	LRSF	ITGRVSHL	FGWTG	596
Q1DVU2	Ci	CNFVRDIDAFDHKFFKKS	SPREAAAMDQQR	LLVQVAYQAVQAGHF	TEATRAGQED	----	RDRSH	VGVYIGSP	SVDYEHNV	STHPLN	ALMATGN	LQSYLP	GRVANH	551
Q5BEJ6	An	CCFFRDGDAFDHKFFKKS	SPREAAAMDQQR	IVLQIAYQALEQSGY	FVEDHNGYTPDGR	DKMHV	VYLG	SCVDYEHNI	SCYDPN	AFTATG	LRSF	ITGRVSHH	FGWTG	604
Q2U7I0	Ao	CNFVRDIDAFDHKFFKKS	SPREAAAMDQQR	LLVQVAYQAVQAGHF	TEATRAGQED	----	KDRNH	VGVYIGSP	SVDYEHNV	ATHPLN	AFMATGN	LQSYLP	GRVANY	590
Q5BBP8	An	GNFVRDPDAFDHKFFKKS	SPRESMAIDPQR	LSLEAAYQALEQSGY	FNELTMTSAAEQER	KKHV	VYV	GVCSY	EYDSNV	HCHPPS	AFTTTG	ELRSF	IPGRISHY	569
Q5B8A0	An	SNLIRDSDQFDHSFFKKS	SPREAAAMDQQR	HLVQVAYQAVEQSGY	FLEPGHN	----	ETKLP	EDSVGV	VFH	-----	VSCYDP	NAYTATGN	LRAFLAG	514
Q2GR18	Cg	GNFLDDTAVFDHKFFKKS	SPREALHMDPQR	LILOIAYQAVQAGY	FQPKGNKSSD	----	RRIC	YIGAVTNDY	EYNI	SHAIPN	AFSATG	ALRSY	IAGKVSHH	551
Q5B7U4	An	GNFVRDIDAFDHKFFR	KVPREAMSQDPQR	LMLOVAYQALEQSGY	FQSPSPGKD	-----	ICCFI	ATCTVDYEHNV	NCHPAS	YAATGL	LRSF	LAGKLSHY	FGWRG	560
Q2UNE1	Ao	GNFVRDYDAFDHKFFKKS	SPRETVSSDPQR	LMLOVAYQAVQSGY	FRSK--AAIDQH	-----	VGCYV	GIGVVDYENNV	ACHAPT	YTATGNL	KSFAAGK	ISHH	FGWSG	547

(N/D)YDTFDHKFFK: sorb1-FFK-for

DPRFFNM: LC1

AEQMDPQ: LC3

## Results and Discussion



## Results and Discussion

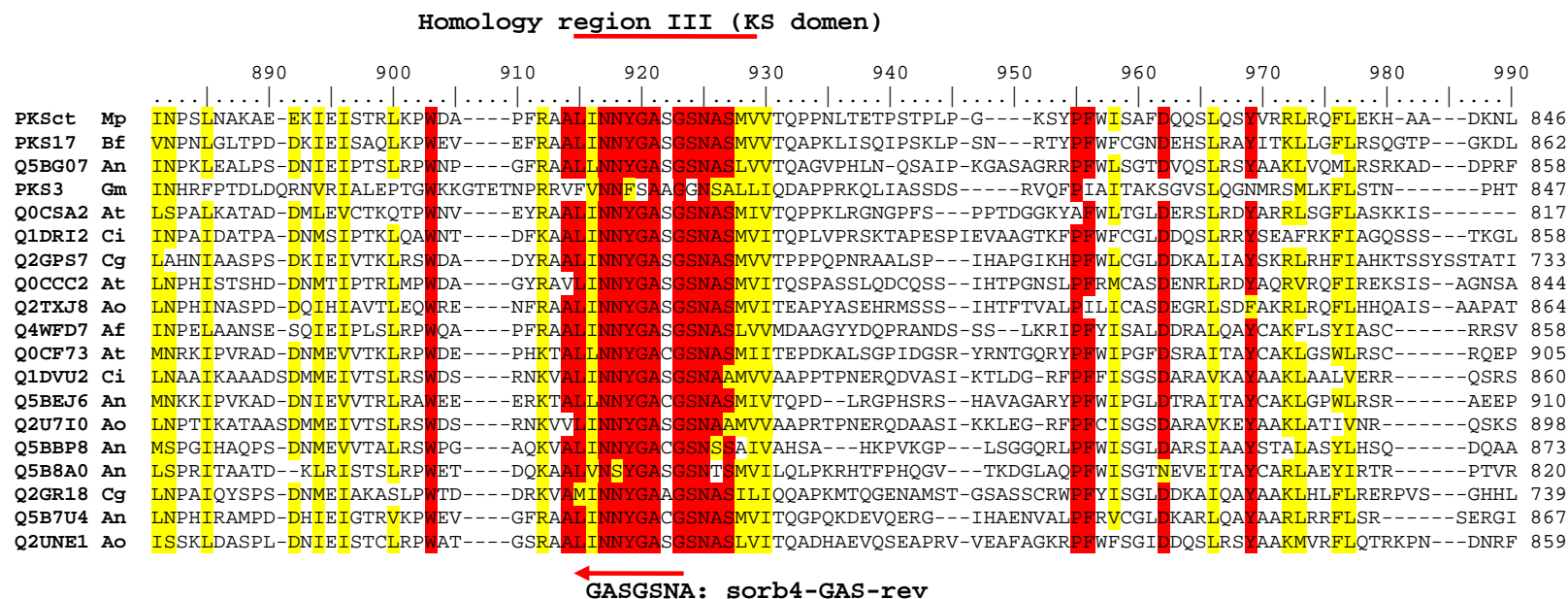
	780	790	800	810	820	830	840	850	860	870	880												
PKSct Mp	TVPNAI	SLSELE	FTDVRQ	ARLEPKD	ITLVEAHGTG	TAVG	DPAEYDGI	RAVFGG	-----	PIRS--	DVI	SLG	SVKGLVGHTE	CASGVVSL	IKTLLMI	QQGF	I	PPQASFS	-S	750			
PKS17 Bf	TVPI	ISDLS	DLFLH	HLHAK	LEPKDISV	VEAHGTG	TPVGD	PAEYKGI	Q	RVFGG	-----	PDSS--	HKVSL	TSVKGLIGHTE	CASGVASL	LKVVILM	VQEGF	I	PPQASFT	-S	764		
Q5BG07 An	TAPSVR	SLANV	ENGIRK	ARVDPKQ	ISVVEAHGTG	TQVGD	RAEYDSI	RTVLGG	-----	PGRA--	YPI	SLG	SVKGLIGHLE	CASGIAAL	IKVLLMV	QNGI	I	PPQPGFS	-K	758			
PKS3 Gm	TRPHVGA	QRVIF	NKILNEA	AVDPYS	VSYVMHGTG	TQAGD	AEMSSV	LETFAPPLAEGK	VARPDS	QKLYIG	S	A	KANIGHG	EAA	SCVSVIK	VLMKK	DTI	VPHCGIK	KNK	749			
Q0CSA2 At	FVPNAP	SLSGL	FDDV	RKSGL	SPKDITL	VEAHGTG	TPVGD	PAEWD	SIRKTLG	GR-----	SVRSV--	PMP	VS	VKGLIGHTE	CTSGVVA	LKVVLLLI	HYGI	I	PPQASFQ	-T	722		
Q1DRI2 Ci	FVPNVPS	LSSLF	CAVRN	QARIK	PAEISV	VEAHGTG	TAVGD	PAEYDS	SIRKALG	G-----	PARTG--	PLMLG	SVKGLVGHTE	CTSGIVS	IKVLLMI	QKSMI	I	PPQASFN	-T	756			
Q2GPS7 Cg	FVPNSP	SLTYL	FKDVL	DRAKLE	PRDISV	VEAHGTG	TPVGD	PAEYLS	VLQALG	G-----	SLRDK--	PLALG	SVKGLIGHTE	GASGVVSL	IKIILMM	QHGY	V	PPQASFQ	-S	631			
Q0CCX2 At	VVPDTQ	SLAGL	FKKAT	ARAHIL	PRDISV	VEAHGTG	TQAGD	PAEYAS	VQRSLG	G-----	PQRKA--	PLFLG	SVKGLVGHTE	GVSGLIA	LKVVLLM	LQEGK	I	PPQPNFQ	-T	743			
Q2TXJ8 Ao	VVPDAS	SLAGL	FTQV	TKKAHL	HPRDISV	VEAHGTG	TQAGD	PAEYVS	IRQTLA	G-----	PRRTS--	PLSLG	SVKGLVGHTE	GVSGLIA	LKILLMI	NEG	I	PPQPNFQ	-T	763			
Q4WFD7 Af	TVPNSV	SLTDL	FRRV	LDKSQL	SPDQISV	VEAHGTG	TAVGD	PAEYTSI	QQVLG	G-----	PKRPR	HNPV	NLGS	VKGLVGHTE	CASGIVS	LKILLMI	LKARTI	PPHVGL	E-T	762			
Q0CF73 At	FVPNSP	LSHL	FKDVM	HQAKIT	ANDVSL	VEAHGTG	TPVGD	PAEYES	IRVALG	G-----	PIRK--	KTLP	I	GSVKGHI	GHTE	GASGIAI	LVKI	IMMREG	F	PPQASF	-K	807	
Q1DVU2 Ci	FVPNSP	SLAQL	FGDV	IKKAHV	SPRDIAL	VEAHGTG	TAVGD	PAEWAAL	RKAVG	G-----	PIRP--	QPLP	V	SVKGHIGHTE	GTSGVIS	LKVLIM	MHEGY	I	PPQASFN	-Q	762		
Q5BEJ6 An	FVPNSP	LSHL	FKDVM	RQAKV	TANDVSL	VEAHGTG	TPVGD	PAEYES	LALALG	G-----	PSRK--	KKLP	I	GSVKGHI	GHTE	GASGIAI	LVKI	IMMREG	F	PPQASF	-T	815	
Q2U7I0 Ao	FVPNSP	SLAQL	FGDV	VKKAHL	SPRDIAL	VEAHGTG	TAVGD	PAEWAAL	RKAVG	G-----	PIRP--	QPLP	V	SVKGHVGHTE	GASGII	SLKVLMI	HEGF	I	PPQASFN	-K	801		
Q5BBP8 An	FVPNVPS	SLSTL	FNEV	IRKAR	IARRISL	VEAHGTG	TPVGD	PAEWQ	SIRNAVA	G-----	PRRD--	TVLP	I	GSVKGHVGHTE	GASGLV	SLKVLMM	MRGNF	I	PPQASFN	-T	780		
Q5B8A0 An	VVPNKP	SLAPL	FKQV	ISKSG	LTPADISL	VEAHGTG	TAVGD	PAEYAS	IADALNV	-----	ASRP--	SPLI	L	GSVKGHI	GHTE	AAA	GVVSL	LKVLMM	NEGF	I	PPQASHT	-R	725
Q2GR18 Cg	-----	-----	-----	-----	-----	-----	-----	IRQIFG	-----	SARAG	MKPLQ	I	GS	AKGLVGHTE	GASGIVAL	LKVVLLM	ILESRI	P	PPQASFN	-T	638		
Q5B7U4 An	VVPDTAS	LAGL	FKKVT	QRAHL	HSRDISI	VEAHGTG	TQAGD	PAEYES	VRDVLG	G-----	PRRVG--	NLALG	SVKGLVGHTE	GVSGIIA	LKVVLM	MILNG	I	PPQPGFH	-S	771			
Q2UNE1 Ao	TAPSVQ	SLTGV	FNTV	LRKARL	DPKQISV	VEAHGTG	TQVGD	RAEYDS	VRRVLG	G-----	PGRA--	EPL	SLG	SVKGLIGHLE	CASGVAAL	IKVLLMI	QNGAI	PPQ	SSFR	-T	758		

← LC2c :EMHGTG, KHKS3c :VEAHGTG

← LC5c :EAHATST

## Results and Discussion



**Table 5: Sequence similarities of the deduced amino acid sequences of the amplified partial KS fragments from *P. chrysogenum* E01-10/3.**

Sequence name <sup>(a)</sup>	Primer pair <sup>(b)</sup>	Sequence size	Sequence homologues <sup>(c)</sup>	Identities (%) <sup>(d)</sup>	Positives (%) <sup>(d)</sup>	References <sup>(e)</sup>
KSnr-7 (pMA4)	LC1,LC2c	707 bp	▪ Putative WA type PKS (fragment), <i>P. patulum</i>	97	99	Q9Y832; Bingle <i>et al.</i> , 1999
			▪ Naphthopyrone PKS (WA), <i>A. nidulans</i>	89	96	Q03149; Watanabe <i>et al.</i> , 1999
KSnr-8 (pMA5)	LC1,LC2c	703 bp	▪ Putative WA type PKS (fragment), <i>P. patulum</i>	97	99	Q9Y832; Bingle <i>et al.</i> , 1999
			▪ Naphthopyrone PKS (WA), <i>A. nidulans</i>	88	95	Q03149; Watanabe <i>et al.</i> , 1999
KSnr-10 (pMA6)	LC1,LC2c	707 bp	▪ Putative WA type PKS (fragment), <i>P. patulum</i>	96	98	Q9Y832; Bingle <i>et al.</i> , 1999

## Results and Discussion

Sequence name <sup>(a)</sup>	Primer pair <sup>(b)</sup>	Sequence size	Sequence homologues <sup>(c)</sup>	Identities (%) <sup>(d)</sup>	Positives (%) <sup>(d)</sup>	References <sup>(e)</sup>
			▪ Naphthopyrone PKS (WA), <i>A. nidulans</i>	88	94	Q03149; Watanabe <i>et al.</i> , 1999
KSpr-MI (pMA1)	LC3,LC5c	686 bp	▪ Putative PKS (PKS2), <i>P. patulum</i>	94	97	P87001; Bingle <i>et al.</i> , 1997
			▪ 6-methylsalicylic acid synthase (MSAS), <i>A. terreus</i>	71	83	P87162; Fuji <i>et al.</i> , 1996
KHKS1 (pMA9)	KHKS2, KHKS3c	248 bp	▪ Putative PKS, <i>A. terreus</i>	78	87	Q0C8A4; Birren <i>et al.</i> , 2005
			▪ Myxalamid PKS (Mxac), <i>Stigmatella aurantiaca</i> (bacterium)	56	73	Q93TW9; Silakoski <i>et al.</i> , 2001
KHKS3 (pMA10)	KHKS2, KHKS3c	259 bp	▪ Putative PKS, <i>Chaetomium globosum</i>	73	87	Q2GPL2; Birren <i>et al.</i> , 2005
			▪ Compactin nonaketide synthase (CNKS), <i>P. citrinum</i>	72	82	Q8J0F7; Abe <i>et al.</i> , 2002
KHKS9 (pMA11)	KHKS2, KHKS3c	290 bp	▪ Putative PKS, <i>A. terreus</i>	72	82	Q0C8A4; Birren <i>et al.</i> , 2005
			▪ Curacin A PKS (CurL), <i>Lyngbya majuscula</i> (bacterium)	50	69	Q6DNE1; Chang <i>et al.</i> , 2004
KHKS27 (pMA12)	KHKS2, KHKS3c	290 bp	▪ Putative PKS (Syn8), <i>Magnaporthe grisea</i>	73	82	Q5XF88; Bohnert <i>et al.</i> , 2004
			▪ Compactin nonaketide synthase (CNKS), <i>P. citrinum</i>	67	77	Q8J0F7; Abe <i>et al.</i> , 200
KHKS32 (pMA13)	KHKS2, KHKS3c	278 bp	▪ Putative PKS ; <i>A. nidulans</i>	65	80	Q5AUX7; Galagan <i>et al.</i> , 2005
			▪ Citrinin PKS (PKSct), <i>M. purpureus</i>	61	76	Q65Z23; Shimizu <i>et al.</i> , 2005
KHKS37 (pMA14)	KHKS2, KHKS3c	259 bp	▪ Putative PKS, <i>A. terreus</i>	78	87	Q0C8A4; Birren <i>et al.</i> , 2005
			▪ Myxalamid PKS (Mxac), <i>Stigmatella aurantiaca</i> (bacterium)	56	73	Q93TW9; Silakoski <i>et al.</i> , 2001
KHKS42 (pMA15)	KHKS2, KHKS3c	286 bp	▪ Putative PKS , <i>A. nidulans</i>	69	85	Q5AUX7; Galagan <i>et al.</i> , 2005
			▪ Citrinin PKS (PKSct), <i>M. purpureus</i>	65	81	Q65Z23; Shimizu <i>et al.</i> , 2005
KHKS44 (pMA16)	KHKS2, KHKS3c	310 bp	▪ Putative PKS; <i>Chaetomium globosum</i>	73	87	Q2GPL2; Birren <i>et al.</i> , 2005
			▪ Compactin nonaketide synthase (CNKS); <i>P. citrinum</i>	69	82	Q8J0F7; Abe <i>et al.</i> , 2002
KHKS46 (pMA17)	KHKS2, KHKS3c	247 bp	▪ Putative PKS , <i>A. nidulans</i>	63	82	Q5AUX7; Galagan <i>et al.</i> , 2005
			▪ Citrinin PKS (PKSct), <i>M. purpureus</i>	59	79	Q65Z23; Shimizu <i>et al.</i> , 2005
KHKS51 (pMA18)	KHKS2, KHKS3c	350 bp	▪ Putative PKS, <i>Phaeosphaeria nodorum</i>	55	62	Q0UAE2; Birren <i>et al.</i> , 2005
			▪ Putative PKS (PKS14), <i>Gibberella moniliformis</i>	51	68	Q6RKK1; Kroken <i>et al.</i> , 2003

<sup>(a)</sup> Working name of a sequence with plasmid name in brackets; <sup>(b)</sup> Primer pair used for amplification of the sequence in question; <sup>(c)</sup> Putative sequence homologues retrieved via BLASTX search showing the first match for homology search and the closest characterised homologue; names for corresponding genes are given in brackets and species in italics <sup>(d)</sup> Identity/similarity (%) of the deduced amino acid sequences. <sup>(e)</sup> Reference to TrEMBL accession number, author and the year for first deposition of a sequence in GenBank database.



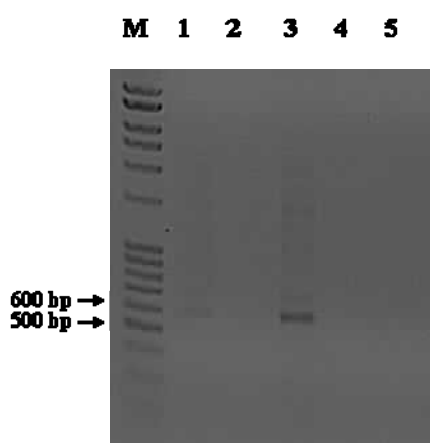
### 6.1.2 AT domain amplification primers

The AT domain was also used for design of degenerate primers suitable for detecting members of NR PKS clade III. The same alignment of 19 sequences was used as in the case of KS specific degenerate primers design (see Table 4). Five homology regions (I–V) were chosen within the AT domain. The part of the alignment corresponding to the AT domain is shown in Figure 17. The homology region I (PVILCFGGQV) and the homology region II (VVGHSFGE) each gave rise to one forward and one reverse primer (sorb7-GQV-for and sorb6-PVI-rev; sorb9-FGE-for and sorb8-VVG-rev, respectively). From the homology region III one forward primer was generated (NVTNAFH: sorb10-NAFH-for), while homology region V gave rise to one reverse primer (PPYQFEK: sorb13-PPY-rev). Within the homology region IV two variations of conserved motifs were used for the creation of two reverse primers: IWLEAGSN for sorb11-IWL-rev and IFLEAGSN for sorb12-IFL-rev. In addition, homology region III, which corresponds to the end of the KS domain, was used to design sorb5-ASG-for as the forward primer for amplification from the AT domain (see Figure 15). Corresponding amino acid residues from these sixth homology regions of all deduced protein sequence used in alignments are shown in Table 33. The number of PCRs with different combinations of forward and reverse primers (data not shown) as well as variation of PCR conditions (see 12.8) were performed in order to obtain the amplification from AT domains from *P. chrysogenum* that correspond to putative members of NR clade III PKS. After optimization of PCR conditions and utilization of a hot start PCR protocol (12.8.1), a combination of sorb9-FGE-for and sorb11-IWL-rev primer pair gave rise to a PCR product of expected size (587 bp) (see Figure 16). One of two sequenced PCR products showed 59% identity to a putative PKS of *Chaetomium globosum* (GeneBank: Q2GR18) as well as 52% identity to the characterised PKSct (citrinin) of *M. purpureus* (GeneBank: Q65Z23); see Table 6.

Although 5 forward and 6 reverse newly designed primers were used for the amplification of the AT domain of the fungal NR PKS clade III (see Table 27), only one cloned sequence corresponded to a member of this subclade (i.e. AT-9-11). With regard to the 6 fungal partial KS domain sequences amplified with the aid of 3 different KS domain primers (see section 6.1.1), the amplification from the AT domain appears rather less efficient. These KS domain amplification primers were designed in a manner to be able to amplify both members of the fungal PR PKSs (LC3/LC5c) as well as members of the fungal NR PKSs (LC1/LC2 and KHKS2/KHKS3c), which might contribute to the diversity of amplified KS domain

sequences. Secondly, the KS domain of fungal type I PKSs is regarded as the most conserved domain.

However, the combination of sorb9-FGE-for and sorb11-IWL-rev AT domain primer pair gave rise to a dominant band with the PCR product of expected size (587 bp), as illustrated in Figure 16. It is of interest to mention that from all the AT domain primers used, this primer pair was the only one which was not degenerated. It might be that this fact enhanced the amplification of the partial sequence that corresponds to a member of the targeted fungal NR PKS clade III.



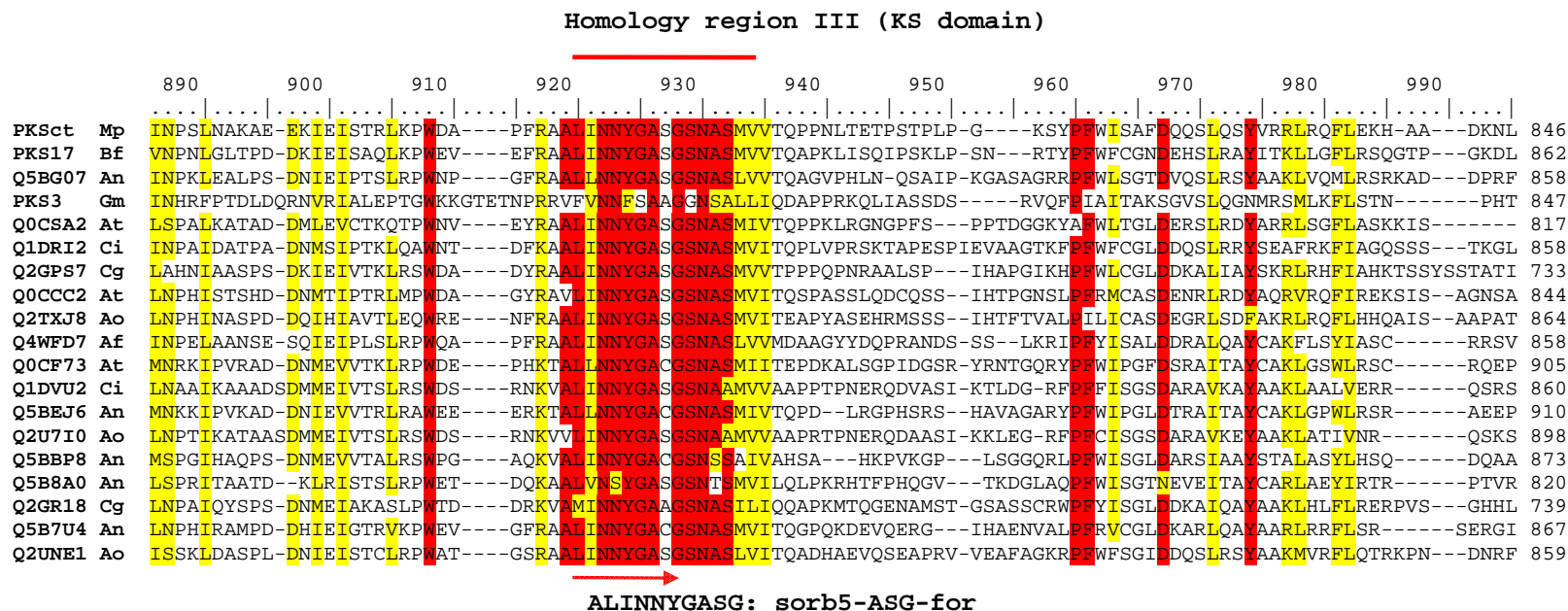
**Figure 16: PCR amplification with AT-specific primers - ethidium bromide agarose gel.**

The sorb9-FGE-for and sorb11-IWL-rev primer pair was used. Lane M contains 100 bp ladder as a fragment size standard. Lanes 1 and 2 are showing hot start PCR for two different temperatures (55 °C and 59 °C, respectively) obtained with 1 ng (1:10 dilution) of genomic DNA from *P. chrysogenum* that served as a template. Lanes 3 and 4 represent hot start PCR for the same two temperatures (55 °C and 59 °C, respectively) obtained with non diluted genomic DNA preparation from *P. chrysogenum* (10 ng/reaction). In lane 5 a negative control was run in parallel with samples. Lane 3 shows clear amplification of PCR product of the expected size (587 bp).

## Results and Discussion

**Figure 17: Alignment for members of NR PKSb clade III - AT domain.**

The extracted part of the alignment corresponds to the end of KS domain and the whole AT domain. Nineteen deduced amino acid sequences of putative members of the NR PKS clade III: *M. purpureus* (Mp: PKSct ), *Botryotinia fuckeliana* (Bf: PKS17), *A. nidulans* (An: Q5BG07, Q5BEJ6, Q5BBP8, Q5B8A0 and Q5B7U4), *Gibberella moniliformis* (Gm: PKS3), *A. terreus* (At: Q0CSA2, Q0CCC2 and Q0CF73), *Coccidioides immitis* (Ci: Q1DRI2 and Q1DVU2), *C. globosum* (Cg: Q2GPS7 and Q2GR18), *A. oryzae* (Ao: Q2TXJ8, Q2U7I0 and Q2UNE1) and *A. fumigatus* (Af: Q4WFD7). The most conserved amino acids in all compared sequences are shaded red; similar amino acid sequences are shaded yellow. The threshold for both identity and similarity shading is 80%, meaning for at least fifteen sequences out of nineteen. The arrows indicate the primer position and direction. The exact amino acid motif is given for each primer from which it was designed with the corresponding region of homology and direction of given primer. The newly designed primers are shown in red. In sum, six regions of homology are shown in alignment: The region of homology III from KS domain, and five regions of homology with AT origin (I-V). The numbers on the right indicate the position of the displayed amino acid sequence within the complete sequence published. The ruler above indicates the position of the extracted alignment in the overall alignment.



## Results and Discussion

**Homology region I (AT domain)**

	1000	1010	1020	1030	1040	1050	1060	1070	1080	1090	1100	
<b>PKSct Mp</b>	SVANLSFQVAC-QSNWLPQALVFSASTKEELNRALASFE-----KGSTDFPSVQLP-DPKVPVILCFGGQVSTVYVGLDQEVYNSTAILRHLYDQCDA MCLSLG-LQSIY	947										
<b>PKS17 Bf</b>	STSNLSFQVSR-QSNRSLDWALIFNCNSSSELQDKLAAHL-----KGEKSITAIRTP-STRPVILCFGGQVSTVYVGLNQEIYDTIPLLRYLDQCDAIFISMD-LGSIY	963										
<b>Q5BG07 An</b>	TVANLSFQLAR-QSNRNLGQALIFSCASVDELEAKLADFA-----SGGNTLTSVPRPDSSRPVILCFGGQVSTVYVGLDREAFDSFKLLKSHLTHCHETSALG-LGGIL	960										
<b>PKS3 Gm</b>	SLAELSYTTTA--RRIHQHRLVPGATPEELICSKLETALQN-----NTGVTRPKAAPKVVFTFTTGGQAQVPGMGKQLFEENEFFVNEILLSLDRIAQN LGFTSMLS	946										
<b>Q0CSA2 At</b>	SLSSLSFNAYR-QSNRALPHGLIFSSSSMEEIQTQLTDFAKG----NGKLSATTRKPVV---PVLVLCFQQISRFBVGLDKAVYESIGTLETHLDRCDATLQSLG-LDGLY	918										
<b>Q1DRI2 Ci</b>	SLSNLSFNLAR-QSNRSLERSLMFVSRVDELLQKLEMEFEK----DMSVASAATTAQP--RPVVLVCFGGQVSTFVGLDPNVIYKRVAILQKHLDTVDAAAARSLG-AGSIF	960										
<b>Q2GPS7 Cg</b>	SLANLSFNVCR-QSNRHLDRALMFSCSLGELDEKLALFDG----DAKLGSTSTKTIKRRPVLVLCFQQVSTFVGLDRQVYDSVKIVFDYLDQCNAVFLSIG-LEGIF	837										
<b>Q0CCC2 At</b>	SLANLSYNNCR-QSNPALEAQIAFCASPEELDNNLTAIDG----RGPQITKLRKSSR---PVLVLCFQQVSTFVGLDGGVYDAFPLLRQHLLDCDITLKEMG-HEGLY	945										
<b>Q2TXJ8 Ao</b>	DLGNLSFNICR-QSNPTLDSQVAFTCCSKSELSSKLSIIDG----ETSHIFRFKISPR---SIVLVCFQQVSTFVGLDRTVYDCISLLRYWLDQCDSLAQSGF-YGSLF	965										
<b>Q4WFD7 Af</b>	SLDALAFNLSR-QSNRDLGCVLTTLGCTSIEDLEKQLEVVQ----AGRAWRSRVPQ---RPVILVCFGGQVSTFVGLNDRVDFISNPLLRMHLLDCDRMCRSMA-LDGIY	957										
<b>Q0CF73 At</b>	TLADVLSFNVR-QSNRSLTQGFIFNCRSMTLEHEKLEQAAAAG---KDAANAGITPVKAERPVLVLCFQQVSRFBVGLDRNLFESVAILLRQHLLHVDVAVVTSQG-LGSIY	1010										
<b>Q1DVU2 Ci</b>	ALADLSFNINR-QSNPDLQRFLLRCSRVDLKER-----LASVSEHMDTDTQAEERPVLVLCFQQVSTFVGLDRKLYDSIAVFRHHLDECDEIVQAEGLGPSIF	958										
<b>Q5BEJ6 An</b>	TLADLSFNLSR-QSNRGLPQGFIFNARSLAELEHEKLEQAVAAAPSSKDAASVGIAPVKAERPVLVLCFQQISRFBVGLDRGLFDAVALFKKHLDAVDTVVKAQG-LVSIY	1018										
<b>Q2U7I0 Ao</b>	TLADLSFNINR-QSNPDLQRLLLRCSRVDLKR-----LASVSEELTNTQAEERPVLVLCFQQVSTFVGLDRKLYDNIAVFRHHLDEC DNVLQTELGPSIF	996										
<b>Q5BBP8 An</b>	SLADAKLALAA-KAT-----KDTAPSVGLVPVKPERPVLVLCFQQVSTFVGLDRAVYEGASVFRHHLTCNAAITSHS-LESYI	950										
<b>Q5B8A0 An</b>	LDDLSFNLSR-QCNRTLARQLVFSRCRSMSELQAMLS-----RGSAEGSSKVVPTEERPVLVLCFQQVSTFVGLNRLVYENVGLLRYLDEC DMAMTNMG-LESYI	918										
<b>Q2GR18 Cg</b>	DIENVSFNVR-QSMNGSLGRAAMFAAGSIDELEQQLG-----SLETAATPVSTRPVLVAFGGQVGVKVVGLDRVFDKSTILLRHLDCDRACKSIQ-AGSIF	835										
<b>Q5B7U4 An</b>	SFANIAFNLTR-KSNPALECCQVVFQTRSELEKDIITGLEEG---DNKYI IQVKKPKR---PVLVLCFQQVGRSISGLDRTFYNAFPLFKHHLSCDDILKANG-DSSYI	968										
<b>Q2UNE1 Ao</b>	TIDNLSYQLAR-QSNRSLGQSLIFSCASVALEAKLASFA-----DGWSELKSTQRQSSRPVILVLCFQQKSNFVGLDR EAYDHFKVLETHLDCHEICLSLG-LGGII	961										

↔  
**PVILCFGGQV: sorb6-PVI-rev, sorb7-GQV-for**

**Homology region (AT domain)**

	1110	1120	1130	1140	1150	1160	1170	1180	1190	1200	1210	
<b>PKSct Mp</b>	--PAIFQRSFIEDIVQLQTFALFAMQYSCAKAWIISGLK--VAASVGHSGFELIALCVSNVAVSLKDAVKMISGRARLIKERNGADKGSMIAVEADLSDVEALLAKVKSQM-	1052										
<b>PKS17 Bf</b>	--PDIFQKSFVQDIVKLSILFALQYSTAKCVIDCGIK--VAAAIGHSGFELTALCVSNALPKHAIEMI TGRAHLIQKKTAEKGCMMAVEADAANVNALLAKSK---	1065										
<b>Q5BG07 An</b>	--PAIFDRTFRSSIVELQLMQFALQYSCAKTWIESGII--VAALVGHSGFELTAMCVSSTLSLQDTLRMIAGRARIIEEKWGPDRGSMMAVDGELQVLRLLHNAHEAS-	1065										
<b>PKS3 Gm</b>	FIQSDPEPDKFAPSFLVQLASVCLQITLSKLMATWGIT--PTAVVGHSLGEYAAINVAQVLSDDTTLFVLVGGRAQLLEKQCTRGTHAMLVVKGSGEEIANALKGEN---	1050										
<b>Q0CSA2 At</b>	--PGIFERSQVDDPVKLTQTFALQYSCAKCWLDCGLQ--PVAVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1062										
<b>Q1DRI2 Ci</b>	--PSIFERSFIQDVTVKLQLILFAMQYACAHSMWISGIIQ--PAAVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1062										
<b>Q2GPS7 Cg</b>	--PAVHQREPIEDPVRLQTFALFAVQYACAKSMMDSGIIQ--AEAVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	943										
<b>Q0CCC2 At</b>	--PGIFEHAPINDQVKLQTKLFAVQYACARSWISGILS--VAAVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1051										
<b>Q2TXJ8 Ao</b>	--PGIFEHNPIIDQIQHLIQFLSLOYSACRCWISGLA--VEAVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1071										
<b>Q4WFD7 Af</b>	--PEIFQKSEYQDITKLQTFALFALQYACATWIDCGAP--PAAVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1063										
<b>Q0CF73 At</b>	--PEIFEREVVRTVKLQTFALFAMQYACAKSMMDSGIIQ--AEAVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1116										
<b>Q1DVU2 Ci</b>	--PDIFSRTELGDPVRLQVALFAMQYASARSWIDCGISLGVVSVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1065										
<b>Q5BEJ6 An</b>	AAPDLFSREPIEDTVKLQTFALFAMQYACATWIDCGISLGVVSVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1128										
<b>Q2U7I0 Ao</b>	--PDIFSRTELGDPVRLQVALFAMQYASARSWIDCGISLGVVSVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1104										
<b>Q5BBP8 An</b>	--PDIFSSEFYQDITKLQTFALFAMQYASAKTWIDCGIAEKVSVVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1056										
<b>Q5B8A0 An</b>	--PEIFSTSFPQDITVKLQTFALFAMQYSCARCWLDCGVSTRIAAVVGHSGFELTALSISGLVSVQDTIKLVAGRARLIRDQWSDRGAMMAVEGDESTIQQLLEEASQQTP	1026										
<b>Q2GR18 Cg</b>	--PTIFQREPIINDPSVLQPVLSLOYSACARSWIDCGVEP--AALVGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	941										
<b>Q5B7U4 An</b>	--PGIFATAPVLDIVQLHTQTFALQYACARSWIDCGVE--VTAVIGHSGFELTALCVSGLVSLDASLDMVVRVKKQLSRGDSQHCLLQRTQELHYCWYFSVYRSRR---	1070										
<b>Q2UNE1 Ao</b>	--PAIFERTFRSSIVELQTMQFALQYSCAKSWIISVQ--VAALVGHSGFELTAMCVSSTLSLQDTLRMIAGRARLLEEKWGPDRGSMMAVDGELQVLRLLHNAHEAS-	1066										

↔  
**VVGHSGFE: sorb8-VVG-rev, sorb9-FGE-for**

## Results and Discussion

### Homology region III (AT domain)

		1220	1230	1240	1250	1260	1270	1280	1290	1300	1310	1320													
PKSct	Mp	---GSETGLA	IACYNASKS	FLLAGPTKDV	DHAENLLKN	---DPDFSG	IRYKRLN	VNTNAFHS	LVLDALIDDES	LGGQIRF	KEPT	IKLERATEQES	---TSTLNANY	VAT	1152										
PKS17	Bf	-----NVA	IACYNGPRS	FLLAGTGKA	IDSLEEMAKS	---DPTFSG	VKLLKLN	VNTNAFH	CHLVDPLMED	LEKFGMKL	VEREDFPL	TATEYEF	---TDKRDEK	FVAR	1161										
Q5BG07	An	---PNEPAVN	IACFNGPRL	FLLAGTTKAM	RVVREVL	SK---DNRLSS	IKVKS	LETSHAFHS	TLVEPLIPD	LEELGEE	SIFRKPV	VVHERATQNSV	---AGPPPPS	SIFAS	1165										
PKS3	Gm	-----YETAC	INSPIETV	LQGPNEQ	LAKVKEQLTA	-----ASF	KTLK	VPYAFHS	SQLEPVVFD	IEKLAKK	VAFSEPR	IPVLCPLEGT	VIENENPF	NASYLAR	1144										
Q0CSA2	At	----EETPAT	VACFNGPTT	FLLAGSTKA	IDAVSECIS	-----SVSV	VRGKKS	VSNAFHS	TLVEPLKDQ	LGGLAEG	MIERGEAK	IRWERATENQTS	--ANIGP	EFFAS	1120										
Q1DRI2	Ci	-----CFYLH	TSFID	-----E	SEETECYQ	CISL	CFA	RPLLEQL	EQGSKNL	NGEPI	IPMERATE	FPHQ	--EKLTP	RFVSD	1131										
Q2GPS7	Cg	----DDEL	PVTIAC	YNGPRS	FLLAGPTK	SMDKFVET	LTGSP	--MYS	PVMR	VKRLN	VNTNAFHS	LVLEPLIDEL	DAIGRT	LTFR	REPS	ILLERSTKELN	---LETTS	EFVAQ	1043						
Q0CCC2	At	----IGGF	PASIA	CYNGPRS	FLLAGSTSS	IDKIQDV	VRTN	---LRF	SHIR	TKRLS	VNTNAFH	SSLVDPL	VLTELEH	VDGL	TLNQPS	IHLHATEN	PSN--CTIPT	TFVAD	1151						
Q2TXJ8	Ao	----TNVAP	ATIAC	FNGPSS	FLLAGTTAA	LDGIQAT	LKT	---PSY	ISL	KAKRL	VNTNAFHS	DLVDPL	LLPALED	VMCG	IHPQEP	IIPC	KATENACT	--GTVTS	DMVAK	1170					
Q4WFD7	Af	F---PEGAG	ASIA	CYNGPRS	FLLAGPTK	AMDAIEA	IVSEPAEQ	QLGAL	RMKRLN	VNTNAFHS	TLVQPL	TRDLE	LLGQSL	SFYPAR	IRVER	STECAS	---PPPST	GRYVAD	1167						
Q0CF73	At	----S-DGS	ASIA	CYNGPRS	FLLAGSTGA	IDAVQQT	MGS	---NSK	FSG	IKSKRL	VNTNAFHS	ALVDKI	SDGLER	IGKTL	TFHRPI	IPVERATE	MPFD--MDN	LDG	SFVSQ	1216					
Q1DVU2	Ci	-----DNP	ATIAC	YNGPRS	FLLAGSVEA	IDA	VAEAIANN	--SDK	FSA	LSKRLN	VNTNAFHS	SLVDPL	LDRL	LEQV	GNDLE	FHEPS	IPLERATE	TAF---TG	PLTPKF	VDP	1164				
Q5BEJ6	An	----S-DGS	ASIA	CYNGPRS	FLLAGSTSA	VDQVQ	TIS	---RPE	FSG	IKGKRLN	VNTNAFHS	SLVDKI	SDGL	SI	GKTL	TNSPL	IPVERATE	VAS---AR	ATDAS	FVSQ	1226				
Q2U7I0	Ao	----GLD	SASIA	CYNGPRS	FLLAGSVK	AIDAVAE	AI	SN---GDK	FSA	VRSKRLN	VNTNAFHS	SLVDPL	LDRL	LEDV	GKQLE	FHEPS	IPLERATE	TAF---AG	PLTPKF	VPN	1204				
Q5BBP8	An	----S-DGT	AGIAC	YNGPRS	FLLAGSTKA	IDAF	AATLP	-----G	KEV	SKRLN	VNTNAFHS	ALVES	IVDRL	GEV	KEV	TFHDA	VPIERATE	HSG--DAT	LDWTF	VGS	1151				
Q5B8A0	An	----S-DHP	ATIAC	YNGPRS	FLLAGSTAA	IEG	VKETLLS	----KYP	YIR	HKRLN	VNTNAFH	C	SLVGN	IVE	PLNQL	GEEL	TNSPV	HIHSS	STPST--GAR	PSSTF	VAD	1124			
Q2GR18	Cg	NKGDGR	HMACTI	CYNGPRS	FLLAGSVA	ACDAV	QHHI	EARDAD	SIRPT	IKHKRI	HVNTNAFHS	GLVEPL	KPELLA	VGSQL	TRQPR	IPLERE	TEGYR	---KCP	SDASY	VAE	1048				
Q5B7U4	An	-----AN	IACF	NGPRS	FLLAGPTAV	IDFLQ	EELGA	-----D	SA	FLKRL	VNTNAFHS	TLVDPL	LLPAL	ASA	IDGL	ALNTAT	IPIERATE	HQAA--DTI	PL	SIVAD	1162				
Q2UNE1	Ao	----TTE	PPVNI	IACFNG	PRQFLL	ACTSKAM	AVVKQ	TLS	---NPF	FSS	IKAKQ	LDTHAFHS	VLVDPL	VPEL	EKL	GEDLI	IRSPV	I	PHERATE	QEI	---RD	PPAF	NV	FAS	1166

→  
NVTNAFH: sorb10-NAFH-for

Homology region IV (AT domain) Homology region V (AT domain)

## Results and Discussion

		1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430	
PKSct	Mp	HMRKPVFFAQAVKRLSDK	---FPVAIWLEAGSNSTITAMASRALGTS	---	nss---	FQAVNITS	-EGAFRFICDTTVKLNKEG	-KVSFWAHHRLQTPMYTPVLLPPYQ	1250				
PKS17	Bf	HLRSPVFFHHAVQRIAKK	---HPDATWLEAGSNSTVTITMASRALGSS	---	GASSSHFQPINITS	-DNSFNFLAEATSR	LWKEGL-RVLFWAHHPVQASSYTPVILPPYQ	1262					
Q5BG07	An	HMRDPVYFDRAVQRLANR	---YPSIWLLEAGSGSGVTINLASRAAGSR	---	AMA---	FQSNINITS	-SSAVQNVADATLNLWKEGL-QVMFWEHVR-PSPKFPLLLPPYQ	1262					
PKS3	Gm	HSRQPVNMLTALTTAYRDGYLSDRSMVLEVGPHPAVSGMVKPTLGQQ	-----	ITCVASLQSRAPWDMLSAALKSYDAGA	-SINWADYQSNFPGAHTVVDLPAMS	1244							
Q0CSA2	At	HMRDPVYANHALQRLHRE	---FPSAIWLEAGSNSTITRMANKALGF	---	PAESYFADINITS	-DSASNTITSEFVNLWKEGL-AIPHWAHCTQAPAYSTLLPPYQ	1219						
Q1DRI2	Ci	HIRSPVFFHHAVERLSKR	---HLSCIFLEAGSNSTITSMASRALGN	---	PGSSHFAVNIITC	-DNGINNLDTTILSLWRAGL-KVDHWPHQTAQTKHTPPLPPYQ	1230						
Q2GPS7	Cg	HMRQPVFFEHAVKRLAEQ	---YPDAAWI EAGSASTITNMA SRALGS	---	PSTYHFQSLNITC	-DHGMKSLADATLALWKAGV-ETAYWPHHPGOTYEFDHLFLPPYQ	1142						
Q0CCC2	At	HLRRPVYFNQAVERLAEK	---YKSCVWLEAGSNSTITLMARALGS	---	SGEHHFQAVNITS	-NAGLQNLTDRTISLWREGV-FVRFWGHHCQSPEYGPVLLPPYQ	1250						
Q2TXJ8	Ao	HLRQPVYFHHAVQRLAEK	---YGPCVWLEAGSNSTITYVNRALVP	---	SSGHHCQAVNITT	-ERGMQSLDATVSLWKASV-SVAFWGHHSQAQKYVPMFLPPYQ	1269						
Q4WFD7	Af	HMRNPVYFHQAVQRLAHE	---YPSAIWLEAGSNSTITNMA SRALDNP	---	ASS--HFQPVNLTN	-DSSSVTFADVFIKLDQGLSFVSWPHHPSQGHRYEMLLPPYQ	1267						
Q0CF73	At	HMRQPVYFNHALQRLVVKR	---YPQAIWLEAGSSSTITIMASRAIAQS	---	QASSSDAHHFQAMSITS	-DTAFDSLTDATMALWKQGL-RVSFWAHHAVQARDYAQLLLPPYQ	1320						
Q1DVU2	Ci	HMRQPVFFSHAMQRLAKK	---HPSAIFLEAGSSSTITVMASRALAG	---	QASKSHHFESVSVIN	-EKGLDGLDITLISLWKQGL-RVAFWAHHTLOTNDYTYILLPPYQ	1265						
Q5BEJ6	An	HMRQPVFFFNHAVQRLAKR	---HPQAIWLEAGSSSTITVMAGRAIAQG	---	QASS-ESHYFQAVSITN	-ETALDLSLADTTALWKQGL-RVTFWAHDVQTAEYAHLLPPYQ	1329						
Q2U7I0	Ao	HMRNPVFFSHAIQRLAKK	---YPSAIFLEAGSSSTITVMARRALAG	---	QVSKSQHFESVSVIN	-EKGLDGLDITVSLWEQGL-RVAFWAHHTLOTKDYTHILLPPYQ	1305						
Q5BBP8	An	HMRRPVFFFNHAVQRLAEK	---YPD AIFLEAGSNSTITVMASRALAN	---	PKSTHHFQSIISITNTNKGIARLTDATVDLWKQGL	-RVSFWAHHRFQKDEYAQLLLPPYQ	1252						
Q5B8A0	An	HLRKPVYFYHAIERLARA	---YPDSIWLEAGSNSTITMAGRALRPTSDIVGQSHHFQALNI	---	HGGRGSDSLTETLTLWKEGL	-RVSFWQHHPVKDQYEHLLPPYQ	1230						
Q2GR18	Cg	HMRDPVYWLOAVERLASK	---YPD AIWLEAGSNSTITNMASKALGMP	---	RSATFLPVNITGDDRCQLHLDITMGLWRAGV	-HVAFWPHSRACTHQYAPIMLPPYQ	1148						
Q5B7U4	An	HLRQPVYFNNAVQRLAAR	---HGPAIWLEAGSNSTITSLARRALGLG	---	VSGNTFHSVNVTS	-TSALMNLTDVTVGLWSDNV-PCTFWGYHARCTREYAPLLPPYQ	1262						
Q2UNE1	Ao	HMRDPVYFDQAVQRLAQK	---YPSIWLLEAGSGSGVTALASRAAGSC	---	GMV---	FQSNINITG-SGAVQNVADATFNLWKEGL-NVSFWGHVG-CTVKS-LLLPPYQ	1262						

I(W/F)LEAGSN: sorb11-IWL-rev, sorb12-IFL-rev

PPYQFEK: sorb13-PPY-rev

## Results and Discussion

### Region of homology V (AT domain)

```

...
      1440      1450      1460      1470      1480      1490      1500      1510      1520      1530      1540
PKSct Mp FEKS-RHWMDLKVPKPEAS---VQVAEQTAIIE-----APKGLTTFVGYQDAS--QRSVFRVNVTTTEKFNRLLSGHIMANAIAVCPGMFQVETALDAL 1339
PKS17 Bf FEKS-KHWMDLKPKKLEP-----LVVKTQTKE-----LPGKLTTFVGYQDEA--NHSALFQVNTKTERFARLVSGHVMASVAAVCPGIFQMEALDAL 1348
Q5BG07 An FAKS-RHWLEERRKLLTKVF-----VPASPVQE-----AQKGLWTFVGYQDSD--RCQARFQIHITSDEFQNYVSAHVIAQTAPICPSMFQQVIARDAL 1347
PKS3 Gm WDLK-EYWIQVWVNDWSLRKGDPLVINNVSKLEST-----TIHSVVEESGNS-KKTHIVVEADIARKDLSPLVQGHEVDG-IPLCTPSVYADIALSLG 1334
Q0CSA2 At FEKS-RHWLELKKPQ--AVQ---LIELPSKAQQEE-----LPTKLYTFVGYQDEG--KRQARFRVNTMIKAYEEFVSGHLIVQTAPICPATLEVDITIEAL 1307
Q1DRI2 Ci EDKV-SHWIELRGGP--KVS---STPTLPKEEEE-----VPDQLLTFVGYQDGR--KRLAKFQINTMISKYDRLIVGHVIAETETPCPSTVQLDLVVEAI 1318
Q2GPS7 Cg EEKV-KHWMDLKPKKGLVE---LPAAEAGNVEAR-----VPVGLWSFLGYDDSK--EQTARFVVNTDSKKYHDFVSGHLIANTAPICPATLEVDMAIEAL 1232
Q0CCC2 At FEKS-RHWMDNKKLPS-----TTSHP TSAQA-----REDDLICFMGYEGRN--KETAKFRINTAHPRYQSAVTGHIVAHTPVAPASII LNTVLEAL 1334
Q2TXJ8 Ao FEKS-RHWLSNKKLPG-----PAHEAGATPA-----VSASALRFVGYQDTQ--QLVARFSIDTAHPQYQESIAGHVVSHTSPVAPASFMLDYVVEAL 1353
Q4WFD7 Af FELF-RHLLPMTPREEQDVQ---GKVGSA S TKPG-----SEQDLSWFSVFSHN---AQS RFQIHMDSNKIQEYVTGHTVAGSAPLCPSTLQLELAIETI 1354
Q0CF73 At EDTSSRHWLPKMSPLEEVKKAAMVAAGDVG TG--QHQQNDALQDPRLQSLWNVEVFQDGD--NKKPRFRINTGSDKYNRFLVLSHVIAQTAPICPGTLECDIVIEAL 1425
Q1DVU2 Ci FEKA-RHWMDFK SATDLVAG---FLTQNAPPVSD-----PKTMPLEWTFVGYQGN D--TKHPRFRVNLVSDKFKKLVSGHVFHAHTAPVLSGTVQSDMAVEAL 1355
Q5BEJ6 An EDTSSRHWLPKMSPVKEKVEAALALIAANGGSLAGALQQAGTPQDPRPLPVWTFVGYQDDE--TRQARFRVNTSADKYNRYVLSHVIAQTAPICPGTLECDIVIEAL 1437
Q2U7I0 Ao FEKA-RHWM DYK SASDLVAG---MTQR--APVSD-----SKSMPLWTFVGYQNN D--TKHPRFRVNIISDKFKKLVSGHIFHAHTAPVLSGTVQSDMAVEAL 1393
Q5BBP8 An FEKT-RHWLELKSPIEQAMN--IAKVQDGI AAQKG-----HVNNKSLIWTFLGFQKQKKTKLARFRINTSSDKYQRLFATHVIAKTAPIAEATLEIDMAIETL 1349
Q5B8A0 An FART-RHGLDI IAP-----KTDQGAQQTG-----VPKPDGLWTFVGYEDNN--RSRPRFLINNSSEKYLAMVEGHVMARTAPICPAGLEVDMAIEAL 1315
Q2GR18 Cg FERN-RHWLDFKPPKQVQGETQPSEQAKS GAEGG-----FLPPSGPYTFVGYKDNK--TKKESRFLINNSIKSYVGVVSGHVIAKQAPVLEVPFAIDLAIQAI 1244
Q5B7U4 An FERT-RHWMENKPLP-----LKYNQAVMEVSGHTAAKTAP IAPATLLLDYAI ELL 1313
Q2UNE1 Ao FAKT-RHWLEERRKPEVKEA-----APVAPWPK-----SPKGLWTFMGFQDSG--NTHARFQIHSTSDFEKKYVGAHLVIAQTAPICPSMFQHVMAREAL 1347
...

```

PPYQFEK: sorb13-PPY-rev

**Table 6: Sequence similarities of the deduced amino acid sequences of the amplified partial MT fragments from *P. chrysogenum* E01-10/3.**

Sequence name <sup>(a)</sup>	Primer pair <sup>(b)</sup>	Fragment size	Sequence homologues <sup>(c)</sup>	Identities (%) <sup>(d)</sup>	Positives (%) <sup>(d)</sup>	References <sup>(e)</sup>
AT-9-11	sorb9-FGE-for/ sorb11-IWL-rev	592 bp	<ul style="list-style-type: none"> <li>▪ Putative PKS, <i>Chaetomium globosum</i></li> <li>▪ Citrinin PKS (PKSct/CitS), <i>M. purpureus</i></li> </ul>	59 52	72 68	Q2GR18; Birren <i>et al.</i> , 2005 Q65Z23; Shimizu <i>et al.</i> , 2005

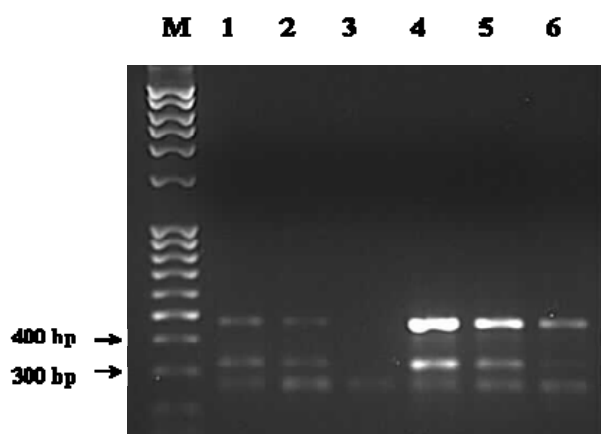
<sup>(a)</sup> Working name of a sequence with plasmid name in brackets; <sup>(b)</sup> Primer pair used for amplification of the sequence in question; <sup>(c)</sup> Putative sequence homologues retrieved via BLASTX search: showing the first match for homology search and the closest characterised homologue; <sup>(d)</sup> Identity/similarity (%) of the deduced amino acid sequences. NR stands for non-reducing fungal PKS. <sup>(e)</sup> Reference combines sequence TrEMBL accession number, referent author and the year for first deposition of a sequence in GenBank database.

### 6.1.3 MT domain amplification primers

The PCR primers designed to amplify C-MT domain regions could provide additional specificity on the cloning approach for gene clusters of methylated polyketides like sorbicillactone A. The C-MT domain is thought to add a methyl group, derived from S-adenosyl methionine (SAM), to  $\beta$ -ketoacyl PKS intermediates (see 2.3.1 for NR PKS). This strategy for introducing branch points is used as well by bacterial modular synthases, with exception of actinomycetes where side-chain groups derive from the incorporation of modified chain extender (e.g. methyl-malonyl) unit [23]. Despite the overall low sequence similarity generally observed for non-DNA methyltransferases, Kagan and Clarke (1994) [229] were the first to identify three conserved sequence motifs (motifs I, II, III see Figure 19). The motif I is involved in SAM binding, while the motif III is part of the active site or is thought to participate in binding of the magnesium cofactor. The motifs I and III were found in all known MT domains, while the motif II is conserved in approx. 80% of the enzymes analyzed (e.g. not found in LDKS nor LNKS) [169]. In addition, the I-IV consensus motifs were described by Miller et al. in 2003 [230]. Their positions in respect to motifs described by Kagan and Clarke are also shown in Figure 19.

Some sets of primers composed of one forward primer (CMeT1) derived from Kagan and Clarke motif I and two reverse primers (CMeT2c, CMeT3c) derived from Kagan and Clarke motif III were designed by Nicholson *et al.* (2001) [169]. The CMeT1 is a general SAM binding site primer, where reverse primers (CMeT2c, CMeT3c) contain MT domain active site and correspond to either LDKS or LNKS, respectively. The relative positions of these primers together with conserved amino acid motifs are shown in Figure 19. By use of a protocol for hot start PCR (see 12.8.1), PCR products of expected size (320 bp) could be amplified with both primer pairs (see Figure 18). From several sequenced amplicones, two showed homology to the fungal type I PKS. The closest putative homologues of these two partial MT domain sequences are shown in Table 8. The deduced amino acid sequence of CMeT-3 was 56% identical with a putative PKS of *A. terreus* (GeneBank: Q0CGB3) while the closest characterised homologue was *A. terreus* LNKS synthase (52% identity; GeneBank: Q0C8M3). The CMeT-9 sequence shared 54% identity with putative PKS of *Neruospora crassa* (GeneBank: Q7S380) and the closest characterised homologue of this sequence was LDKS synthase of *A. terreus* (49% identity; GeneBank: Q8JOF5).





**Figure 18: PCR amplification with CMeT series of MT-specific primers - ethidium bromide agarose gel.**

Lane M contains 100 bp ladder as a fragment size standard. Lanes 1-3 show temperature gradient PCR obtained with CMeT1/CMeT2c primer pair and genomic DNA from *P. chrysogenum* as a template. Temperature-gradient PCR with the CMeT1/CMeT3c primer pair is shown in lanes 4-6. In PCRs employing both primer pairs temperature hot start conditions were employed with a preselected narrow range of temperatures (46.0 – 52.0 °C) and individual gradient steps automatically generated by the PCR cyclor for the chosen rang of temperatures.

Additionally, new MT primers that correspond to members of the NR PKS clade III were designed. Initially, only eight conceptually translated MT domain containing PKSs were used for the alignment, of which six were already classified by Kroken *et al.* [167] as putative members of NR clade III. Two others PKS actually originate from AN1034.2 and AN3386.2 loci of *A. nidulans* F6SC A4 as it was already described in section 6.1.1. Detailed information on the used sequences including species origin, locus designations, accession numbers, their shortened names and references are listed in Table 7.

The amino acid sequence within motif I (SAM binding site) [229] was used as the first region of homology for design of forward primers. The goal was to modify this often used position within the MT domain in order to better apply them/it for members of NR PKS clade III (see Figure 19). Besides, a second region of high amino acid sequence conservation that correlates with positions of motif II (Kagan and Clarke, 1994) [229] and motif III (Miller *et al.*, 2003) [230] was chosen for the construction of reverse primers (see Figure 19). The extracted amino acid motifs of each sequence from alignment are given in Table 34. The degeneracy of primers was decreased by choosing two variations of amino acids within each conserved motif (see Figure 19 region of homology Ia and II). As a result, two forward (MTnr III\_Fa and MTnr III\_Fb) and two reverse (MTnr III\_R1a and MTnr III\_R1b) primers were made. The combination of these forward and reverse primers was expected to amplify PCR products between 220-230 bp accorging to differences in the length of amino acid stretches between these two regions of homology for compared protein sequences. Although a number of PCRs were performed with variations of PCR conditions (temperature, primer and template concentration, magnesium concentration and other; see 12.8), no cloned PCR product showed homology to any known PKS gene deposited in publicly available databases.

As a result of the performed PCRs with 2 pairs of the MT domain-specific primers, two different partial PKS sequences were amplified. The CMeT-3 sequence was amplified with CMeT1/CMeT2c primer pair (Nicholson *et al.*) [169], while the CMeT1/CMeT3c primer pair amplified the CMeT-9 sequence. By yielding two different fungal type I PKS sequences, MT domain based PCR strategy appeared to be more efficient than the AT domain amplification approach in regard to assessment of the biosynthetic potential of *P. chrysogenum* E01-10/3 for polyketide biosynthesis.

Furthermore, the closest classified putative homologue of CMeT-3 sequence is the *C. heterostrophus* PKS5 (GenBank: Q6RKF9) grouped as member of R clade I. Besides, this sequence showed homology to the LNKS of *A. terreus* (GenBank: Q0C8M3) as the first characterised match that otherwise belongs to the R clade II. In the case of the CMeT-9 sequence, both – the first classified putative homologue (*Botryotinia fuckeliana* PKS2, Q6RKJ8) and the first characterised putative homologue (*A. terreus* LDKS, Q8JOF5; Table 8) - belonged to the R clade I. Nevertheless, among other putative homologues of CMeT-9 sequence there are also putative members of R clade II and R clade IV. Based on such dispersed grouping of BLASTX matches for these two MT domain amplified sequences, it is difficult to infer a nature of the full-length PKS.

**Table 7: Sequences used for the deduced protein alignment for members of NR PKS clade III for design of MT and RED domain degenerate primers.**

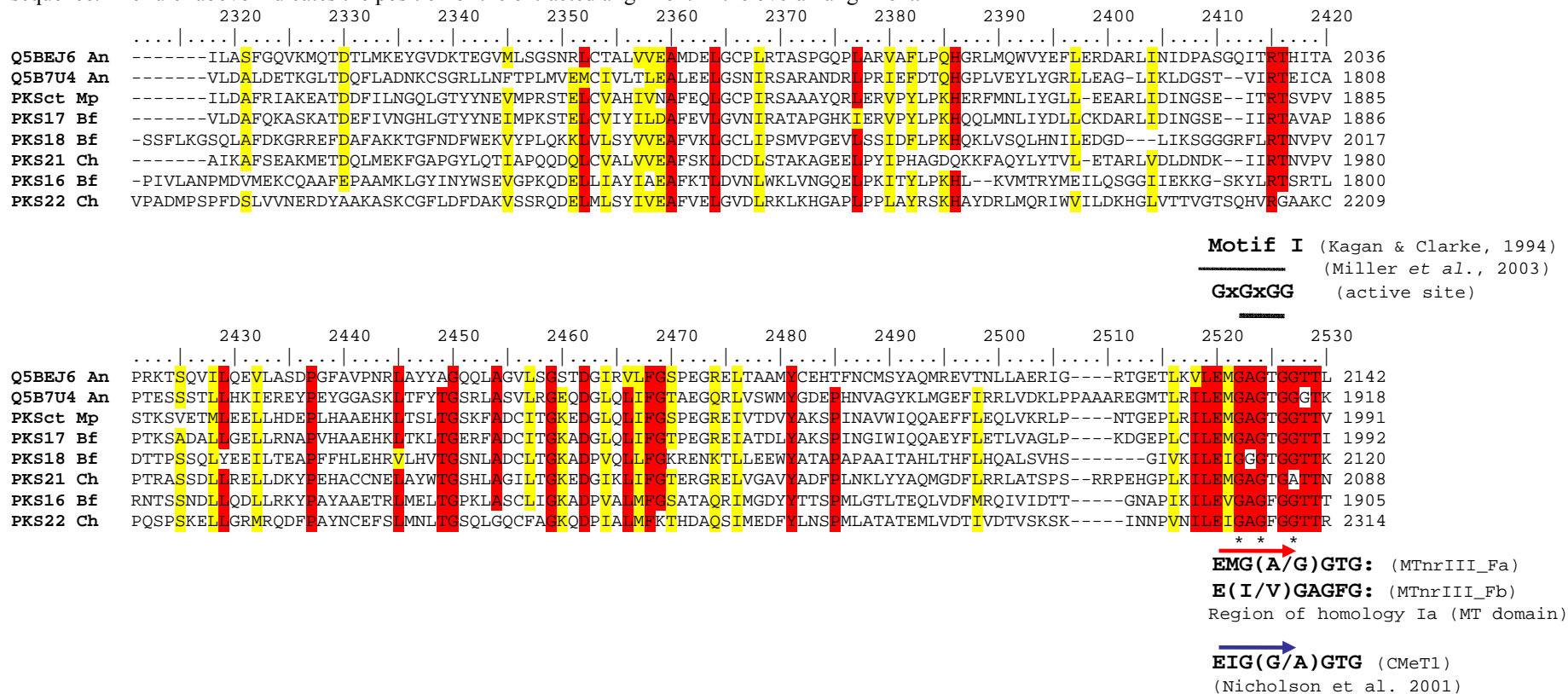
Organism	Gene/ORF*	Protein	Accession number**	Shortened name***	References
<i>A. nidulans</i> F6SC A4	AN1034.2 AN3386.2	Hypothetical protein similar to PKS	Q5BEJ6 Q5B7U4	Q5BEJ6 An Q5B7U4 An	Galagan <i>et al.</i> , 2005 [219]
<i>Botryotinia fuckeliana</i> B05.10	<i>pks16</i> <i>pks17</i> <i>pks18</i>	Hypothetical protein similar to PKS	Q6RKI4 Q6RKI3 Q6RKI2	PKS16 Bf PKS17 Bf PKS18 Bf	Kroken <i>et al.</i> 2003 [167]
<i>Cochliobolus heterostrophus</i> C4	<i>pks21</i> <i>pks22</i>	Hypothetical protein similar to PKS	Q6RKE4 Q6RKE3	PKS21 Ch PKS22 Ch	Kroken <i>et al.</i> 2003 [167]
<i>M. purpureus</i>	<i>pksCT</i>	Citrinin polyketide synthase	Q65Z23	PKSct Mp	Shimizu <i>et al.</i> , 2005 [33]

\* Gene or ORF designation in GenBank; \*\* TrEMBL Accession Number; \*\*\* Shortened names of deduced proteins that are used in alignment (see Figure 19) and Table 34 - Table 35 (appendix).

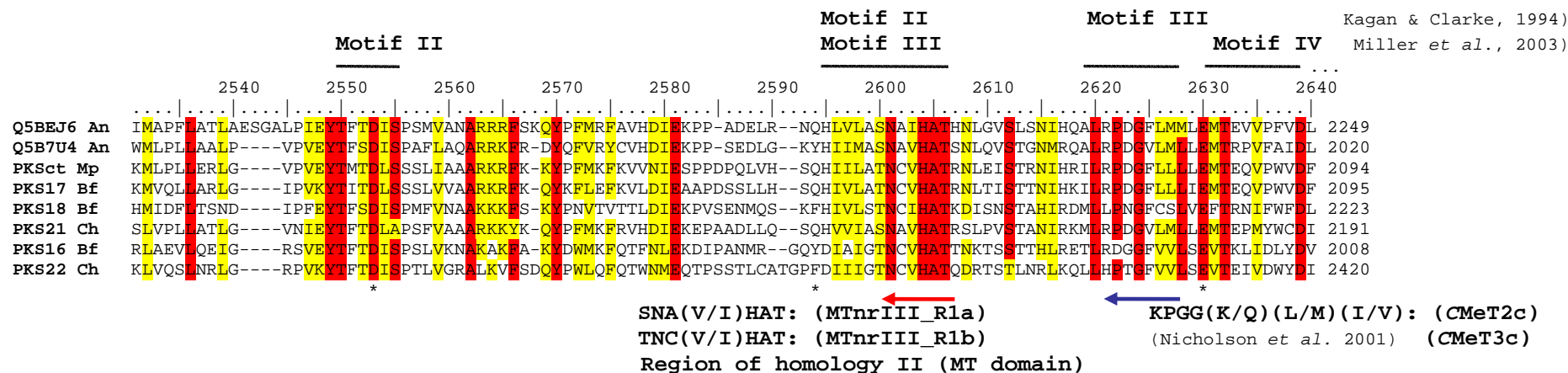
## Results and Discussion

**Figure 19: Alignment for members of NR PKS clade III - MT domain.**

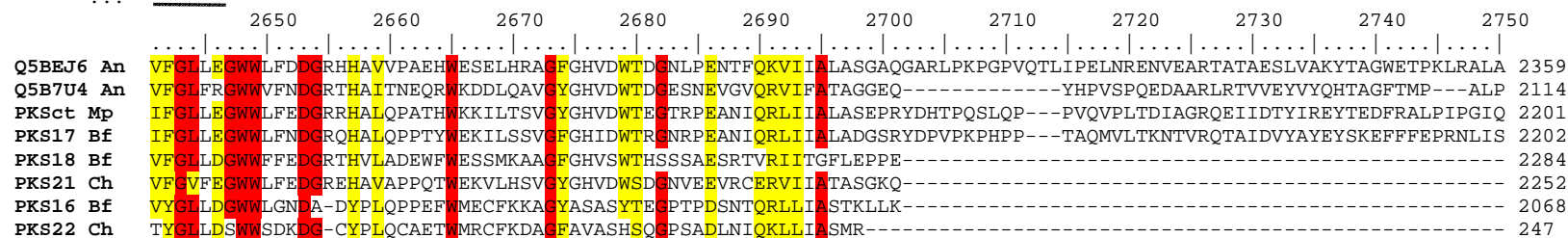
Eight deduced amino acid sequences of putative members of the NR PKS clade III: *A. nidulans* (An: Q5BG07, Q5BEJ6, Q5BBP8, Q5B8A0 and Q5B7U4), *M. purpureus* (Mp: PKSct), *B. fuckeliana* (Bf: PKS16, PKS17 and PKS18) and *Cochliobolus heterostrophus* (Ch: PKS21 and PKS22). The most conserved amino acids in all compared sequences are shaded red; similar amino acid sequences are shaded yellow. Threshold for both identity and similarity shading is 80%, meaning for at least six sequences out of eight. Arrows indicate the primer position and direction. Exact amino acid motif is given for each primer from which it originates. Newly designed primers based are shown in red. Forward and reverse primers of CMeT series are given in blue. Motif I that has the same position in both cases (Kagan and Clarke, 1994 and Miller et al., 2003) is given in black. Motifs II and III (Kagan and Clarke, 1994) are given in green while motifs II-IV (Miller et al., 2003) are in violet colour. The asterisk is indicating the key amino acid residue that interacts with the SAM (Miller et al., 2003). Numbers on the right indicate the position of the displayed amino acid sequence within the completely published sequence. The ruler above indicates the position of the extracted alignment in the overall alignment.



## Results and Discussion



### Motif IV continuing (Miller et al., 2003)



## Results and Discussion

(V/I)(L/I)VTGATG RED domain primer design: region of homology Ib (MT domain) (see)

```

                2760      2770      2780      2790      2800      2810      2820      2830      2840      2850      2860
Q5BEJ6 An  SRAEKESGKTQAPHAAPGRRRAHEAVVIVTGATGSLGSHIVQRLAETPSVATVVCLNRRSSSTTPEKROQAALTARGITLSPGARAKLRVLETDTSKPQLGLPPLYEGWLL 2469
Q5B7U4 An  PRIR-----AP-----ANHACILVTGATGSLGSHLVARLVQLSNVQAVICLN-RVSRMGPRVRQKEAVAARGLSLESKEETKLMVLETDTANDRMGLSVEQCRYLQ 2209
PKSct Mp   QAVMPA-----PTGHCVLVTGATGSLGSHVVGYSRLPNVHTVVCLN-RRSTVPATIRQEEALKVIRGISLDDNSRSKLVVLEEVETAKPLLGLPVETTYQKLV 2296
PKS17 Bf   YPVHPS-----NEKYCVLVTGTTGSLGSHIVAYFAQLPEVETVVCLN-RLSMVDATLRQQSFALRGISLTKISMSKLRVIETDTAKSMLGLPKSTFQYLT 2297
PKS18 Bf   -----RSTFVPKRPRKSFKYESLEYKITGECKLFADVYYDSSAAKNDIKRPVALMI 2335
PKS21 Ch   -----LERLTVPSKPVQEQVCMIPISVD----- 2275
PKS16 Bf   -----TPNREIATKQKSKLCHKIDTVTYKIVDGVVEVLADVFLPEVAPINPMPIALMI 2120
PKS22 Ch   -----NDIEAPKRAPLRPKLETVVYKTIIDGVVDIHADVFFPQQPPATAMPLGTNS 2526

```

**Table 8: Sequence similarities of the deduced amino acid sequences of the amplified partial MT fragments from *P. chrysogenum* E01-10/3**

Sequence name <sup>(a)</sup>	Primer pair <sup>(b)</sup>	Fragment size	Sequence homologues <sup>(c)</sup>	Identities (%) <sup>(d)</sup>	Positives (%) <sup>(d)</sup>	References <sup>(e)</sup>
CMeT-3 (pMA2)	CMeT1/ CMeT2c	326 bp	▪ Putative PKS, <i>A. terreus</i>	56	74	Q0CGB3; Birren <i>et al.</i> , 2005
			▪ LNKS synthase, <i>A. terreus</i>	52	69	Q0C8M3; Birren <i>et al.</i> , 2005
CMeT-9 (pMA3)	CMeT1/ CMeT3c	327 bp	▪ Putative PKS, <i>Neurospora crassa</i>	54	70	Q7S380; Galagan <i>et al.</i> , 2003
			▪ LDKS synthase, <i>A. terreus</i>	49	69	Q8JOF5; Abe <i>et al.</i> , 2002

<sup>(a)</sup> Working name of a sequence with plasmid name in brackets; <sup>(b)</sup> Primer pair used for amplification of the sequence in question; <sup>(c)</sup> Putative sequence homologues retrieved via BLASTX search: showing the first match for homology search and the closest characterised homologue; <sup>(d)</sup> Identity/similarity (%) of the deduced amino acid sequences. <sup>(e)</sup> Reference combines sequence TrEMBL accession number, referent author and the year for first deposition of a sequence in GenBank database. RED stands for reducing fungal PKS.

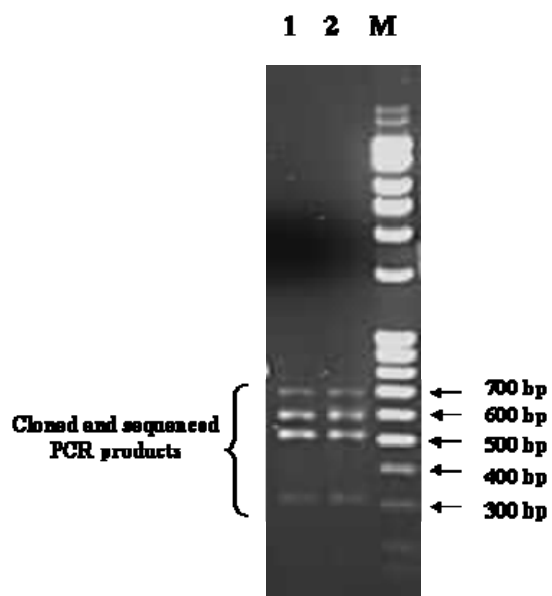
#### 6.1.4 RED domain amplification primers

As already mentioned in the introduction section (see 2.3.1 ), some of the NR type fungal PKS possess at the C-terminus, after the MT domain, a specific domain corresponding to a NAD(P)H-dependent thiolester reductase (RED). Thus, the design of primers that are capable to amplify this specific domain could facilitate an amplification of NR PKS sequences from the genome of *P. chrysogenum*. With this goal in mind, several primers were designed from conserved motifs found within the RED domain itself and the last part of the MT domain. At that time point only five deduced amino acid PKS sequences containing both domains were available in GenBank: The citrinin PKS [33] from *M. purpureus* and four additional putative NR PKS sequences arising from whole genome sequencing projects of *A. fumigatus*, *A. nidulans*, *Botryotinia fuckeliana* and *Gibberella moniliformis* [167]. Approximately 500 C-terminal amino acids corresponding to the RED domain were chosen from each protein sequence and subsequently aligned in the BioEdit program as given in Figure 21. Detailed information on used sequences, along with their accession numbers, are given in Table 9. In this alignment three homology regions were chosen for primer design. From homology region Ib, located at the very end of MT domain, the VLVTGATG consensus motif was chosen for design of two forward (RED-F1a, RED-F1b) and two reverse primers (RED-R1a, RED-R1b, see Table 35). The GYWNPVEH consensus sequence from homology region III of RED domain served for the creation of one forward and one reverse primer, namely, RED-F2 and RED-R2, respectively. Finally, homology region IV (RED domain) containing the MSCGGLI consensus amino acid stretch has been used for the design of two additional reverse primers: RED-R3a and RED-R3b. The exact positions of all three conserved motifs within the RED domain alignment are illustrated in Figure 21, while primer sequences are to be found in Table 27.

Additionally, the reverse primers CMeT2c and CMeT3c (Nicholson *et al.*, 2001) [169] for amplification of the MT domain (see 6.1.3) were modified to serve as forward primers for the RED domain amplification. The CMeT1 forward primer of the same series was used unchanged. Numerous combinations of primers (see Table 37) were applied under various PCR conditions (12.8) and many amplified PCR products of expected sizes were cloned and sequenced.

Figure 20 illustrate one such attempt for amplification of the region between the MT and the RED domain using previously described CMeT1 and newly designed RED-R1b primer. Due to differences in size of amino acid stretches between two conserved motifs, these primers may amplify PCR products ranging from 375 bp as expected PKSct of *M. purpureus* and PKS17 of *B. fuckeliana*, or up to 696 bp for a putative PKS sequence from *A. nidulans*. Beside this, two remaining sequences from alignment (PKS3 *M. purpureus* and Q4WFD7 of *A. fumigatus*; see Table 9) did not contain any of the conserved MT motifs previously described by Kagan and Clarke (1994) [229] or Miller *et al* (2003) [230].

After extensive optimization of PCR via protocol for hot start PCR (12.8.1) it was possible to restrict PCR products sizes to expected range (375-696 bp; see Figure 20). Despite extensive efforts for RED domain amplification, no sequenced PCR product showed homology to any reported PKS sequence in publicly available databases. Other performed combinations of PCR primers for the MT and RED domains are shown in Table 37.



**Figure 20: Ethidium bromide-stained agarose gel for RED domain amplification via CMeT1 and RED-R1b primer pair.**

Lane M contains 100 bp ladder as a fragment size standard. Lanes 1 and 2 show temperature optimised hot start PCR (49 °C and 50 °C, respectively) obtained with CMeT1/RED-R1b primer pair and genomic DNA from *P. chrysogenum* as a template. The expected size of PCR product ranged from 375-695 bp.

**Table 9: Sequences used in deduced protein alignment for members of NR PKS clade III for design of RED domain specific degenerate primers**

Organism	Gene/ORF*	Protein	Accession number**	Shortened name***	References
<i>A. fumigatus</i>	AFUA_3602570	Hypothetical protein similar to PKS	Q4WFD7	Q4WFD7 Af	Nierman <i>et al.</i> , 2005
<i>A. nidulans</i> F6SC A4	AN0523.2	Hypothetical protein similar to PKS	Q5BG07	Q5BG07An	Galagan <i>et al.</i> , 2005

## Results and Discussion

<b>Organism</b>	<b>Gene/ORF*</b>	<b>Protein</b>	<b>Accession number**</b>	<b>Shortened name***</b>	<b>References</b>
<i>Botryotinia fuckeliana</i> B05.10	<i>pks17</i>	Hypothetical protein similar to PKS	Q6RKI3	PKS17 Bf	Kroken <i>et al.</i> 2003
<i>Gibberella moniliformis</i>	<i>pks3</i>	Hypothetical protein similar to PKS	AY495593	PKS3 Gm	Kroken <i>et al.</i> 2003
<i>M. purpureus</i>	<i>pksCT</i>	Citrinin polyketide synthase	Q65Z23	PKSct Mp	Shimizu <i>et al.</i> , 2005

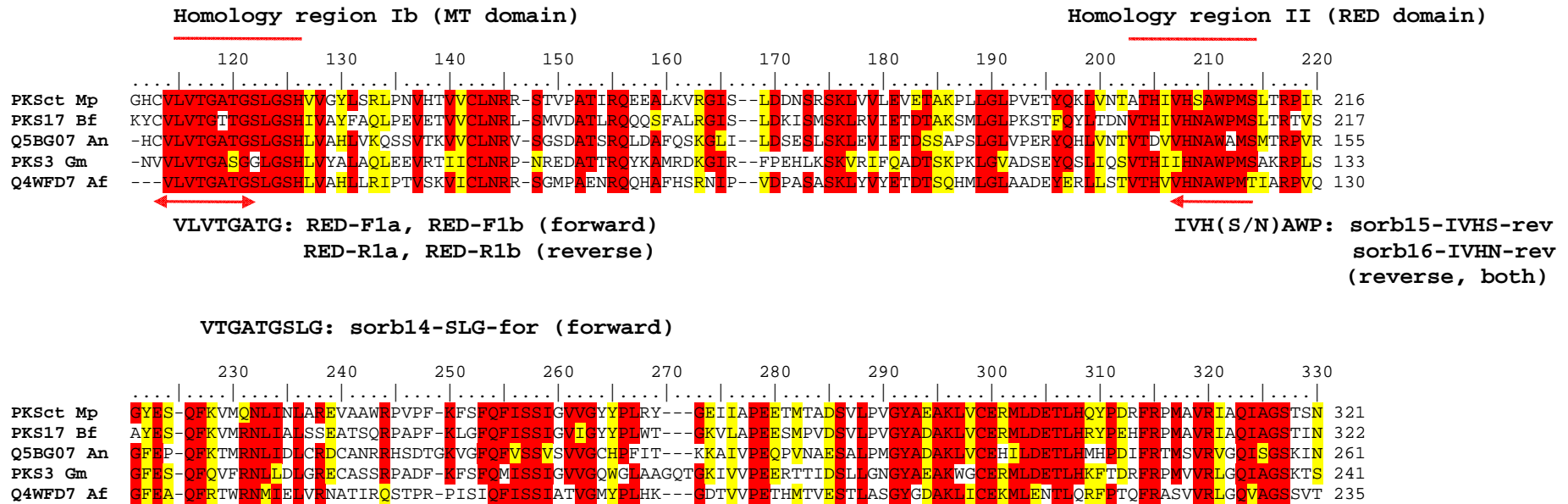
\* Gene or ORF designation in GenBank; \*\* TrEMBL Accession Number; \*\*\* Shortened names of deduced proteins that are used in alignment (Figure 21) and Table 35 (appendix).



## Results and Discussion

**Figure 21: Alignment for members of NR PKS clade III - RED domain.**


The extracted part of the alignment that corresponds to very end of MT domain and RED domain. 8 deduced amino acid sequences of putative members of the NR PKS clade III *A. fumigatus* (Af: Q4WFD7), *A. nidulans* (An: Q5BG07), *M. purpureus* (Mp: PKSct ), *Botryotinia fuckeliana* (Bf: PKS17) and *Gibberella moniliformis* (Gm: PKS3). The most conserved amino acids in all compared sequences are shaded red; similar amino acid sequences are shaded yellow. Threshold for both identity and similarity shading is 80%, meaning for at least four sequences out of five. Arrows indicate the primer position and direction. exact amino acid motif is given for each primer from which it originate. Newly designed primers based on here presented alignment are shown in red. Regions of homology from which the primers are derived, designed primers and their orientations are shown in red. Numbers on the right indicate the position of the displayed amino acid sequence within the completely published sequence. The ruler above indicates position of amino acids within the alignment for C-terminal part of these five sequences.



## Results and Discussion


### Homology region III (RED domain)

	340	350	360	370	380	390	400	410	420	430	440											
PKSct Mp	CHWNPVEHF	AFIKSSQ	TKALP	DFDGS	LSWCP	VDVSA	TIGELLIS	N-----	TPVYSI	VHIENS	SRQWR	KMVKT	LAQSL	DTL	PRDGI	IPFDQ	WIERV	NSSASI	--ND	423		
PKS17 Bf	GYWNPVEH	FASFLK	SSQTLN	ILPDL	KGTLS	WCTVQ	VAATIG	ELLIS	N-----	TPPYPI	VHIENS	SRQWP	DMIMV	LANEL	GLPSN	NI	VPFDE	WLSRV	RFPGST	--ND	424	
Q5BG07 An	GYWNPVEH	LVLHIK	SSKTLN	VLPDL	EGVLS	WCPVD	VAAAG	DDL	LTN-----	KPAYSV	VHIENS	VRQWP	DMLTI	LADAL	DI	ERTNA	VPFKE	WLRVR	HFPPSL	GFSE	365	
PKS3 Gm	GYWNPMEH	HFGFLI	KSSQTL	NALPD	VGNLN	WTPVN	DIAD	SETD	LILSD	N-----	RTPYPI	YHIDN	PIGQNR	DVNNI	LSDTL	RIP--	NKVPF	KQWLD	MVRKAP	Q----	339	
Q4WFD7 Af	GYWNPVEH	LAFIFK	SAKTLN	ILNLP	GE	LSWCP	VNEAA	AITD	LILCP	N-----	DPVPP	VFHV	ENEVR	QPNAD	MIAAL	ARELD	TRIGNI	VPFED	WLNRM	NFVGDE	--AA	337


  
**GYWNPVEH: RED-F2 (forward), RED-R2 (reverse)**  
**WNPVEHFAF: sorb17-WNP1-rev, sorb18-WNP2-rev (reverse, both)**

### Homology region IV (RED domain)

	450	460	470	480	490	500	510	520																
PKSct Mp	NPARQ	LEFFD	QHFI	R-MSC	NLIL	LDTTK	TRHSAT	LRER	GPV	GPLV	EKYI	SANK	TMG	FELD-----	484									
PKS17 Bf	NPAGQ	LVEFF	DKHF	V-R	MSC	GGLI	LDTVK	SQEH	SKTIR	STGP	VSID	LVRKY	LSK	SSG	FLS-----	485								
Q5BG07 An	NPAAR	LADFF	ETDF	LR-M	SCG	MILD	TTRS	REHS	ATLRS	SLGP	IDQD	LVM	ITFK	RIYAI	AKS	SEVK	IKKSS	-----	436					
PKS3 Gm	NPAAL	LADFL	EDTY	LR-	MAC	GGLV	LVKH	SL	EH	SKSL	SAV	GPV	SET	VVRKY	HI	NKEI	GFLK	TTA	EDK	AGFE	ERL	KLWG	PRV	421
Q4WFD7 Af	NPAKR	VDDLE	EHF	V-R	MAC	GGV	LD	TQ	TARL	SAR	FRA	LKA	VDP	GMV	KLF	VEY	WRK	T	G	MN-----	398			


  
**MSCGGLI: RED-R3a, RED-R3b (reverse)**

Within the course of time, the number of members of the NR PKS clade III increased due to the extensive programs for sequencing of fungal genomes. Thus, 19 sequences were aligned together in order to further optimise the design of the RED domain primers for NR PKS clade III. The alignment itself is not shown, but conserved motifs from chosen regions of homology that were used for degenerate primer construction, as well as conserved amino acid stretches from each protein sequence in alignment that correspond to certain motif are given in Table 36 (see Appendix). The region of homology Ib, which is placed within the MT domain, and the region of homology III from RED domain were again used for primer design (sorb14-SLG-for; sorb17-WNP1-rev and sorb18-WNP2-rev, respectively). In addition, a new region of homology within the RED domain, designated as region of homology II was used for primer design. Two variations of conserved motifs gave rise to two reverse primers: sorb15-IVHS-rev (from IVHSAWP) and sorb16-IVHN-rev (from IVHNAWP). The position of the region of homology II with respect to other chosen regions is illustrated in previously shown alignment for the RED domain in Figure 21. Some more information on used sequences with accession numbers for all deduced protein sequences are given in Table 4. Newly designed primers from this extended alignment were used in numerous PCR reactions that employed different PCR conditions and primer pairs (data not shown). Despite the efforts taken to optimise PCRs, all attempts failed to amplify DNA regions homologous to any PKS sequence so far deposited in GenBank database.

### **6.1.5 Summary of PCR-amplified PKS gene fragments from *P. chrysogenum* E01-10/3 strain**

During the course of present study, the application of a PCR-based strategy enabled the amplification of partial PKS sequences from 9 putatively distinct fungal type I PKS gene loci from *P. chrysogenum* E01-10/3. The use of various primer pairs designed for different PKS domains (Table 27), as well as variation of PCR conditions (see 12.8), contributed to the diversity of amplified partial PKS sequences.

Most of the amplified sequences originate from the KS domain and were amplified with 3 different degenerate primer pairs as shown in Table 5. Three same partial KS sequences (KSnr-7, KSnr-8 and KSnr-10, later in text as KSnr-7) amplified with the LC1/LC2c primer pair (Bingle *et al*) [182] appeared to originate from the same gene locus encoding for putative WA type PKS. Besides, the application of another primer pair of the same series, i.e. LC3/LC5c, resulted in amplification of the unique sequence KSpr-MI that displayed putative homology to MSAS of *A. terreus* [46].

Further on, 10 partial PKS sequences that were amplified with the KHKS2/KHKS3c primer pair (Bailey *et al*) [170] after detailed analysis were shown to represent 4 distinct putative PKS genes. The KHKS1, KHKS9 and KHKS37 sequences, later in this study designated as KHKS1, showed approximately the same level of homology to the putative PKS of *A. terreus* (Q0C8A4; see Table 5). Another three sequences (KHKS3, KHKS27 and KHKS44 ) had CNKS of *Penicillium citrinum* [92] as the common putative PKS homologue (later in text as KHKS3). Furthermore, an already characterised member of the NR clade III – citrinin synthase (PKSct) [33] was BLASTX-match of three additional partial KS sequences (KHKS32, KHKS42 and KHKS46, later in text as KHKS32). The last sequence amplified with KHKS2/KHKS3c primer pair was KHKS51 that showed the highest homology to a putative PKS of the fungus *Phaeosphaeria nodorum* (GeneBank: Q0UAE2).

In addition to numerous partial KS sequences, the AT-9-11 partial PKS sequence was amplified with newly synthesised primers for the AT domain (see Table 6). According to BLASTX analysis, this sequence shares homology to PKSct as it was the case for KHKS32 set of sequences (see in text before). Moreover, two different partial PKS sequences were amplified with two different combinations of degenerate primers for MT domains (see Table 8). The CMeT-3 sequence that showed homology to the characterised LNKS of *A. terreus* (GeneBank: Q0C8M3) was amplified with CMeT1/CMeT2c primer pair (Nicholson *et al.*) [169]. Further on, the CMeT1/CMeT3c primer pair of the same series yielded the CMeT-9 sequence that was similar to LDKS synthase of *A. terreus* (GeneBank: Q0C8M3) as it was shown previously (see 6.1.3).

Despite efforts for the RED domain amplification, no sequenced PCR product shared homology to any known fungal PKS gene up to date deposited in GenBank database. Accordingly, it is possible that the chosen PCR parameters or the degenerate primer sequences may not be optimal for the amplification of this particular PKS domain from the analysed *P. chrysogenum* strain. The fact that the RED domains appear relatively rare among fungal type I PKS systems and that they are in general weakly conserved may have influenced the degenerate primer design and thus the outcome of PCR screening study for this particular domain.

As it was illustrated in the introductory section (2.4.2), different fungal species can differ in the number of putative PKS genes that are present in their genomes: from 7 PKS in the genome of *Neurospora crassa* [165] to even 30 PKS genes in the genome of *A. oryzae* [166]. Apart from the whole genome sequencing projects, the PCR screening studies employing degenerate primers have given clear evidence that most fungal strains have the genetic

capacity to produce a considerable variety of polyketides. Nevertheless, only a subset of PKS genes that are present in a genome of particular fungal strain may be detected via PCR screening. As an illustration, 4-5 PKS genes were localised in producer of squalestatin (*Phoma sp.*) [23] or 7 distinct PKS genes from the genome of the lovastatin producer (*A. terreus*) were amplified by use of degenerate PKS primers. Further on, 3 PKS; including one that correlates to melanin biosynthesis, were identified in the genome of *Glarea lozoyensis* [231] with various sets of KS primers developed by Lee *et al.* [232]. On the other side, the screening the genome of *P. citrinum* with different degenerate primers for various domains resulted in detection of 4 PKS genes, one of which was a PKS involved in compactin biosynthesis [33]. Among other polyketide biosynthesis genes that were located in genomes of their producers by usage of degenerate PCR approach were also the fumonisin polyketide synthase gene (*fum5*) from *Gibberella fujikuroi* [42] and PKS gene encoding the citrinin synthase (*pksCT*) in *M. purpureus* [33]. Moreover, one big PCR screening study performed by Wright and collaborators [99] employed LC series degenerate primers [182] in order to estimate the genetic potential of marine fungal isolates for polyketide production. They analyzed genomic DNA from 160 cultured isolates representing 142 distinct marine fungal species. In overall, 99 isolates yielded products in the expected range, most of which generated only a single PCR product. Although this study was able to demonstrate that marine-derived fungi possess an astonishing diversity of putative PKS genes, it should be pointed out again that the PCR-based approach is likely to underestimate the number of present PKS genes in a single fungal species.

Despite the fact that only a subset of PKS genes harboured in one fungal genome can be identified by application of degenerate PCR, this method stays a powerful tool for the assessment of the polyketide-producing potential of both pathogenic and pharmaceutically interesting fungi and very often a start point for the isolation of particular PKS clusters. The detection of 9 distinct partial PKS sequences from the genome of *P. chrysogenum* E01-10/3 strain, which was analysed in the course of present study, does support this evidence as well. Besides, we expect that among the amplified PKS fragments there could be also sequences that correspond to sorbicillacton A polyketide synthase. With this goal in mind and to get more reliable information on the reducing character and clade distribution of targeted partial PKS sequences, efforts were made to perform phylogenetic analysis of cloned KS fragments as presented in the following section (see 6.2). The results of this analysis were important to identify good candidates for the screening of a genomic library to clone the putative

sorbicillactone A and B gene cluster via the hybridization method (see 7.2) and PCR based screening of the library (see 7.3).

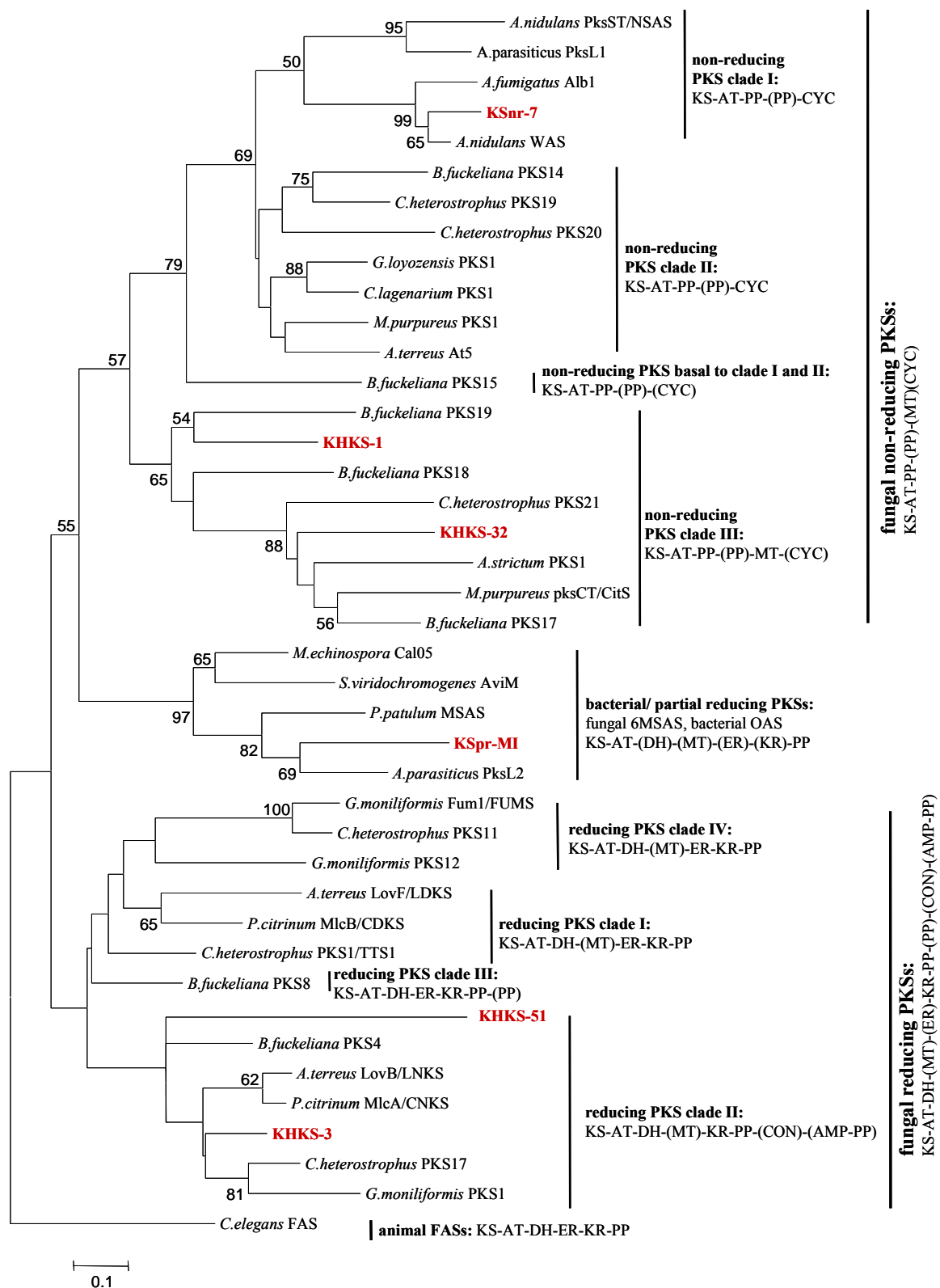
## **6.2 Phylogenetic analysis of amino acid sequences from putative PKS gene fragments**

The phylogenetic reconstruction was performed in order to disclose the evolutionary relationships of partially amplified *P. chrysogenum* PKS genes (6.1) compared to other members of the fungal type I PKS genes. As already mentioned in the introductory section (2.3.1) the fungal PKS were grouped into NR PKS, PR PKS and HR PKS [168] according to their presence of  $\beta$ -keto reducing domains. Moreover, Kroken *et al.* [167] have shown that similar PKS domain organization and similar chemical characteristics of the resulting polyketide products are reflected in evolutionary relationships between the amino acid sequences of their KS domain. The KS domain was chosen by Kroken *et al.* for analysis because it is the most highly conserved domain in type I PKSs and FASs. The resulting KS domain genealogy were interpreted by defining the major clades and subclades of PKS enzymes (see Kroken *et al.* 2003) [167].

The same strategy was applied for 6 cloned *P. chrysogenum* E01-10/3 partial KS sequences in order to reconstruct an evolutionary tree aimed at unravelling its ancestry. For this purpose KS domains from the closest BLASTX matches, as well as known members from defined PKS clades [167] were extracted from published full-length PKS genes, deduced from amino acid sequences and multiple-aligned in the BioEdit program (Table 31). The FAS of *Caenorhabditis elegans* (GenBank: NP\_492417), i.e. an animal fatty acid synthases, was used to root the resulting phylogenetic tree according to Kroken *et al.* [167] and Chooi *et al.* [233]. This step was important since pre-knowledge on direction of changes is crucial for a proper reconstruction of fungal PKS evolutionary processes [167, 234]. Importantly, members of the NR clade III were included into the alignment, since the putative sorbicillactone A PKS may belong to this clade.

In particular, some time was spent to look for regions of conservation and to manually adjust the alignment to the phylogenetic tree. In order to calculate an evolutionary relationships among the aligned KS sequences, the alignment was transferred into FASTA format to the Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 [235] (Table 31). A minimum evolution (ME) method was used to calculate the pairwise distances between the aligned sequences and to group sequences that are most similar. The resulting bootstrapped consensus tree inferred after 1000 replications is shown in Figure 11. Principally, the bootstrap value for a clade is the percentage of replicates supporting that clade. The labeling

of clades and subclades, as well as domain organisation, is presented in accordance to the published classifications [167, 233]. The scale bar given in this figure represents 0.1 substitutions per amino acid site. In Figure 22, KS domain sequences amplified from the genomic DNA of *P. chrysogenum* are shaded red, while reference sequences from GenBank are given in black. Additionally, GenBank accession numbers of 35 reference sequences are shown in Table 10.



**Figure 22: Phylogeny of the PCR amplified partial KS domains from *P. chrysogenum* E01-10/3 strain.** Bootstrapped consensus tree in MEGA 4.0 inferred by the ME method. Bootstrap values greater than 50 are shown at the nodes and were calculated from 1000 replications. The scale bar represents 0.1 substitutions per amino acid site. KS domain sequences that were amplified from genomic DNA of *P. chrysogenum* are shown in red, while reference sequences from GenBank are given in black. Classification of PKS based on Kroken et al. 2003 [167].



For 5 of 6 amplified KS domains from *P. chrysogenum* E01-10/3 the resulting genealogy presented in Figure 22 is supported by the BLASTX results. This means that for each amplified KS fragment the closest full-length classified homologue is in the same clade and subclade as indicated by the BLASTX results. The only exception among the analysed sequences is KHKS51, which groups into the reducing (R) clade II in the tree but has PKS14 of *Gibberella moniliformis* (Q6RKK1) from R clade IV as the closest match according to a BLASTX analysis (see Table 5).

The phylogenetic analysis showed that two PCR-amplified KS domains, i.e. KHKS3 and KHKS51, are grouped within the R clade II. In principal, members of R clade type I fungal PKSs are featuring  $\beta$ -keto processing domains [167]. The only exception among reducing PKSs is the absence of an ER domain from R clade II. The PKS of this subclade were also found to have a condensation (CON) domain typical for NRPS, or even an entire NRPS module consisting of a CON domain, an adenylation (AMP) domain and a PP (phosphopantetheine) as shown in Figure 22. Moreover, the reducing clade III was characterised by the lack of a MT, while members of three remaining reducing subclades may or may not have a conserved MT domain. In addition, most of the reducing PKS have only one PP domain, although some members of R clade III or IV may have two PP domains at once.

As it is illustrated in phylogenetic tree (see Figure 22) three analyzed KS domains belong to the non-reducing (NR) fungal type I PKS clade. The key characteristic of this clade is the absence of  $\beta$ -keto reducing domains – DH; ER and KR [167]. The general domain organisation of NR members is mainly the same for all 4 subclades with a cyclization (CYC) domain being obligatory for members of the NR clade I and II, but rather facultative for the NR clade basal to I and II, and NR clade III (see Figure 22). Thus, polyketides synthesised by NR PKSs are unreduced and usually cyclic (i. e. aromatic) [167]. The members of NR clade III may have a MT domain located directly after PP domain (synonym ACP). An additional peculiarity in this subclade is the presence of a C-terminal RED domain as already mentioned in the introductory section (2.3.1). The resulting phylogenetic analysis showed that the sequence KSnr-7 clusters with the NR clade I, while KHKS1 and KHKS32 are members of NR clade III.

The last PCR-amplified sequence (KSpr-MI) clustered with members of joined bacterial and fungal type I PKS systems that together share a partially reducing (PR) character (Figure 22).

The partial reduction implies that various members may or may not possess certain  $\beta$ -keto reducing domains anticipated as general organisation of this clade: KS-AT-(DH)-(MT)-(ER)-(KR)-PP. Likewise, they may or may not contain MT domain.

An especially challenging point for this phylogenetic analysis was the fact that only very short and size-different KS domain sequences (from 259-707 bp, see Table 5) were compared with the whole KS domains of reference sequences from the GenBank (see Table 10). A complete deletion of gaps was used as an option in the MEGA 4.0 phylogeny reconstruction program (see Table 31) in order to compensate for these differences. As a result of this site-by-site sequence comparison only 87 amino acid positions were considered for the phylogenetic analysis. Nevertheless, in the resulting phylogenetic tree (see Figure 22) all the reference KS domain sequences were clustered in proper clades as in the reviewed literature [167, 233]. The fact that these 87 analyzed amino acid positions did not encompass the KS active site indicates that even such KS fragments might be useful for disclosing evolutionary relationships among fungal type I PKS (Figure 15 for KHKS2/KHKS3c primer pair amplified sequence). Up to now, several studies have shown that short KS regions including active sites are valuable for phylogenetic reconstruction and investigation of the reducing character of the full-length fungal type I PKS [99, 233].

On the contrary, other fungal type I PKS domains have not been approved yet as suitable for such phylogenetic reconstruction. Consequently, only BLASTX results can be used to judge upon a putative reducing character and subclade distribution of the remaining 3 PCR-amplified PKS sequences from *P. chrysogenum*.

Accordingly, all the closest putative homologues of amplicon covering the AT domain (AT-9-11) are members of the NR clade III, inclusive well characterised clade member, i.e. the citrinin PKS of *M. purpureus* [33] (see Table 6). Furthermore, the closest classified putative homologue of CMeT-3 sequence (see 6.1.3) is *C. heterostrophus* PKS5 (GenBank: Q6RKF9) grouped as a member of R clade I. Besides, the same sequence showed homology to LNKS of *A. terreus* (GenBank: Q0C8M3) as the first characterised match that otherwise belongs to the R clade II (see Table 8). On the other hand, in the case of CMeT-9 sequence, both the first classified putative homologue (*Botryotinia fuckeliana* PKS2, Q6RKJ8) and the first characterised putative homologue (*A. terreus* LDKS, Q8JOF5; Table 8) belonged to R clade I. Nevertheless, among other putative homologues of CMeT-9 sequence are the putative members of R clade II and R clade IV as well. Based on such dispersed grouping of the BLASTX matches for these two MT domain amplified sequences, it is difficult to infer the nature of the full-length PKS.

Within the course of our PCR-based study on genetic potential of *P. chrysogenum* E01-10/3 for polyketide production, 9 putatively distinct PKS gene loci were detected by use of 6 different degenerate primer pairs. The phylogenetic analysis of KS domains illustrated that the analyzed *P. chrysogenum* strain has genetic potential to produce all three main classes of fungal polyketides – reducing, partially reducing (i.e. 6-MSA type) and non-reducing. Taking in account the numbers of reported polyketides produced by different *P. chrysogenum* strains (1.4), as well as the relatively low degeneracy of early PCR primers (e.g. Bingle *et al.* and Nicholson *et al.*) [169, 182] that were used in this study, it can be anticipated that the actual number of putative PKS genes within analyzed strain might be much higher.

As mentioned previously (see section 4), the identification of the sorbicillactone A gene cluster is within the focus of the current study. Thus, among the amplified partial PKS sequences the candidates that might originate from this cluster were considered for screening of the *P. chrysogenum* genomic library (see sections 7).

The chemical structure of sorbicillactone A, as well as of sorbicillactone B, indicated  $\beta$ -keto-reductive activities during the early PKS steps and no  $\beta$ -keto reduction in later steps of the biosynthesis (see 3). In line with the proposed biosynthesis of the sorbicillactone A (Bringmann *et al.*), our working group proposed that hexaketide-thiolester-intermediate of sorbicillactone A (and B) is off-loaded from the PKS in the aldehyde-form (Figure 11, intermediate 63) by action of C-terminal-reductase (i.e. RED domain) of fungal NR clade III. Some other fungal polyketides like citrinin, dehydrocurvularin, monocerin and zearelenone share this structural feature as well [50]. Furthermore, specific incorporations of isotope-labelled acetate-precursors supported the model of reducing/non-reducing biosynthesis of these polyketides.

The PCR studies on genetic potential of the fungus *P. chrysogenum* E01-10/3 for polyketide production revealed the presence of three partial PKS sequences as putative members of the fungal NR clade III. The KHKS1 and KHKS32 sequences derived from KS domain (Table 5), while AT-9-11 was amplified with primers specific for AT domain of fungal PKS genes (Table 6). According to anticipated involvement of the fungal NR clade III in the biosynthesis of sorbicillactones, these three partial PKS sequences were considered as potential cluster candidates for sorbicillactone A (and B) production.

**Table 10: Protein sequences included in the multiple alignment and phylogenetic analysis of PCR amplified *P. chrysogenum* partial KS domains.**

<b>Organism</b>	<b>Gene/Protein*</b>	<b>Accession number**</b>
<i>Acremonium strictum</i>	PKS1/MOS	gb: CAN87161
<i>Aspergillus fumigatus</i>	<i>alb1</i>	gb:AAC39471
<i>Aspergillus nidulans</i>	PKST7NSAS	tr: Q12397
<i>Aspergillus nidulans</i>	wA/WAS	tr: Q03149
<i>Aspergillus parasiticus</i>	PKSs11	tr: Q12053
<i>Aspergillus parasiticus</i>	PKSs12	gb: AAC23536
<i>Aspergillus terreus</i>	<i>at5</i>	gb: BAB88752
<i>Aspergillus terreus</i>	<i>lovB/LNKS</i>	gb: AAD39830
<i>Aspergillus terreus</i>	<i>lovF/LDKS</i>	gb: AAD34559
<i>Botryotinia fuckeliana</i>	PKS4	gb: AAR90240
<i>Botryotinia fuckeliana</i>	PKS8	gb: AAR90244
<i>Botryotinia fuckeliana</i>	PKS14	gb: AAR90250
<i>Botryotinia fuckeliana</i>	PKS15	gb: AAR90251
<i>Botryotinia fuckeliana</i>	PKS17	gb: AAR90253
<i>Botryotinia fuckeliana</i>	PKS18	gb: AAR90254
<i>Botryotinia fuckeliana</i>	PKS19	gb: AAR90254
<i>Caenorhabditis elegans</i>	FAS	gb: NP_492417
<i>Cochliobolus heterostrophus</i>	PKS1/TTS1	gb: AAB08104
<i>Cochliobolus heterostrophus</i>	PKS11	gb: AAR90266
<i>Cochliobolus heterostrophus</i>	PKS17	gb: AAR90271
<i>Cochliobolus heterostrophus</i>	PKS19	gb: AAR90271
<i>Cochliobolus heterostrophus</i>	PKS20	gb: AAR90274
<i>Cochliobolus heterostrophus</i>	PKS21	gb: AAR90275
<i>Colletotrichum lagenarium</i>	PKS1	gb: BAA18956
<i>Gibberella moniliformis</i>	<i>fum1/FUMS</i>	gb: AAD43562
<i>Gibberella moniliformis</i>	PKS1	gb: AAR92208
<i>Gibberella moniliformis</i>	PKS12	gb: AAR92219
<i>Glarea loyozensis</i>	PKS1	gb: AAN59953
<i>Micromonospora echinospora</i>	<i>calO5</i>	gb: AAM70355
<i>M. purpureus</i>	PKS1	gb: CAC94008
<i>M. purpureus</i>	<i>pksCT/CitS</i>	gb: BAD44749
<i>Penicillium citrinum</i>	<i>mlcA/CNKS</i>	gb: BAC20564
<i>Penicillium citrinum</i>	<i>mlcB/CDKS</i>	gb: BAC20566
<i>Penicillium patulum</i>	6MSAS/MSAS	gb: CAA39295
<i>Streptomyces viridochromogenes</i>	<i>aviM/AviM</i>	gb: AAK83194

\*Gene or protein designation in GenBank; \*\* GenBank (gb) or TrEMBL (tr) Accession Number

## 7 Screening of a genomic library for the putative sorbicillactone gene cluster

In order to clone the sorbicillactone A biosynthesis gene cluster, the genomic DNA of *P. chrysogenum* E01-10/3 was used for construction of a fosmid genomic library.

### 7.1 Construction of *P. chrysogenum* genomic fosmid library

The genomic DNA of *P. chrysogenum* E01-10/3 was size-fractionated by gel electrophoresis yielding fragments with an average size of 40 kb (12.15.1 and 12.15.2) and blunted using the end-repair mix (12.15.3). After size selection by electrophoresis the approximate 40 kb DNA fragments were ligated into the supplied blunt-ended pCC1FOS fosmid vector (12.15.4). Thereupon, the ligation product was subject to an ultra high-efficiency MaxPlax<sup>TM</sup> Lambda Packaging Extract and plated on phage T1-resistant *E. coli* EPI300-T1<sup>TM</sup> cells (Epicentre<sup>®</sup>; see 12.15.5). The obtained fosmid clones were selected via antibiotic resistance by the chloramphenicol resistance gene (*chl*) located on the fosmid vector. In order to determine the titer of the packaged fosmids, serial dilutions were made and the fosmid library size was estimated to be approximately 400 000 clones. After random detection of insert sizes within 10 single colonies (~40-50 kb), the 4800 fosmid clones were inoculated into 50 ninety-six-well plates for preservation of the library at -80 °C. The overall size of the genomic DNA deposited in genomic library represents a 5 fold coverage of *P. chrysogenum* E01-10/3 genome (~34 Mb), therefore, the created fosmid genomic library should comprise all genes of this eukaryote including the sorbicillactone A biosynthesis gene cluster.

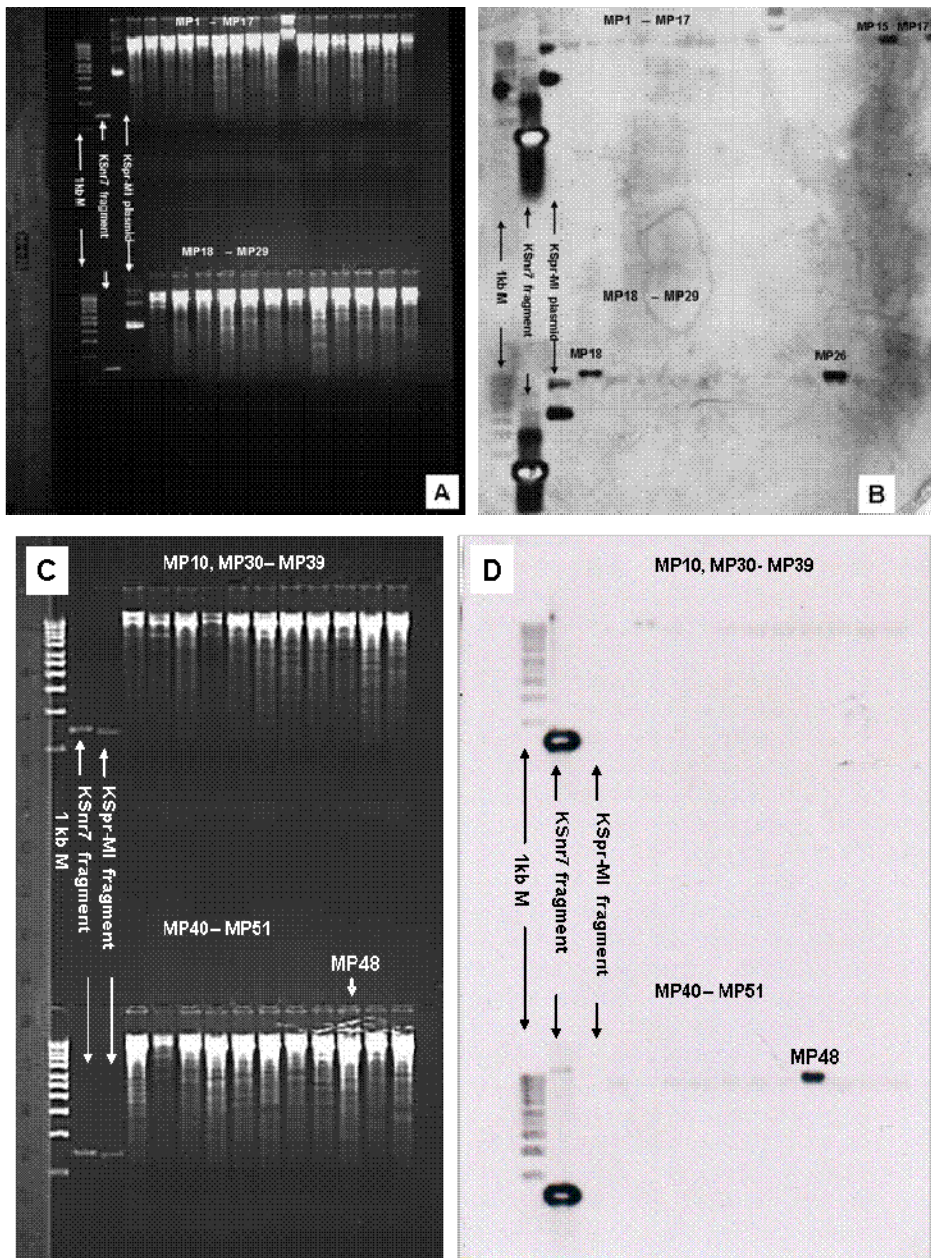
### 7.2 Screening of the genomic library via hybridization

During this project, in parallel with PCR experiments for the amplification of partial PKS fragments (6.1), the *P. chrysogenum* E01-10/3 genomic library was screened via hybridization with homologous and heterologous DNA probes for the presence of NR PKS gene clusters. The KSnr7 partial PKS sequence that showed high homology to the WA type of fungal NR PKS clade I (see 6.2) was used as a homologous probe to screen the genomic library for fosmid clones containing KS domain of the NR type. As already explained (6.1.1), two *E. coli* fosmid clones from genomic library of *A. nidulans* F6SCA4 strain were purchased from FGSC, each containing one putative PKS gene of NR clade III. The isolated plasmids of these clones were used in the PCR with the KHKS series of primers as the positive controls for amplification of the NR clade III PKSs.

Moreover, being aware of the fact that reductive domains are quite rare among fungal PKSs, specific primers that originate from these two *A. nidulans* F6SCA4 PKS genes (Q5BEJ6 and Q5B7U4, see Table 4) were used for amplification of two RED domains. The RED-F.AN1 and RED-R.AN1 primer pair was used for RED-AN1 domain probe amplification from Q5BEJ6 PKS. Besides, the RED-AN3 domain probe from Q5B7U4 PKS was amplified with the RED-F.AN3/ RED-R.AN3 primer pair. For primer sequences see Table 27. These two heterologous DNA probes were cloned and sequenced in order to confirm their identities, and thereupon combined together for screening the *P. chrysogenum* genomic library for presence of the NR-RED domain fosmid clones.

Different hybridization techniques were employed to screen the genomic library: Southern hybridization (12.17.1), colony hybridization (12.17.2) and dot blot hybridization (12.17.3). A variety of parameters crucial for the outcome of hybridization procedure were optimised: The procedure for lysing and fixing colonies to the nylon membrane in the protocol for colony hybridization (e.g. treatment of cell debris after fixation of DNA to the membrane), the amount of DNA probe that was used per hybridization reaction in tubes (100-600 ng/ reaction tube), pre-hybridization and hybridization incubation times (e.g. reducing or prolonging), incubation times for each washing step, detection reaction time, X-ray films exposure and developing times for all three used hybridization techniques, as well as the type of hybridization buffer that gives most acceptable intensity of colony background (12.17.2).

*Bam*HI-restricted plasmid DNA preparations, each corresponding to one microtiter plate from *P.chrysogenum* E01-10/3 library, were electrophoretically separated and transferred via Southern blotting procedure to the nylon membrane and further hybridised with the homologous KSnr7 probe (see Figure 23). As a result, five microtiter plates (MPs) gave strong hybridization signals to KSnr7 probe, namely: MP15, MP17, MP18, MP26 and MP48, as shown in the Figure 23.

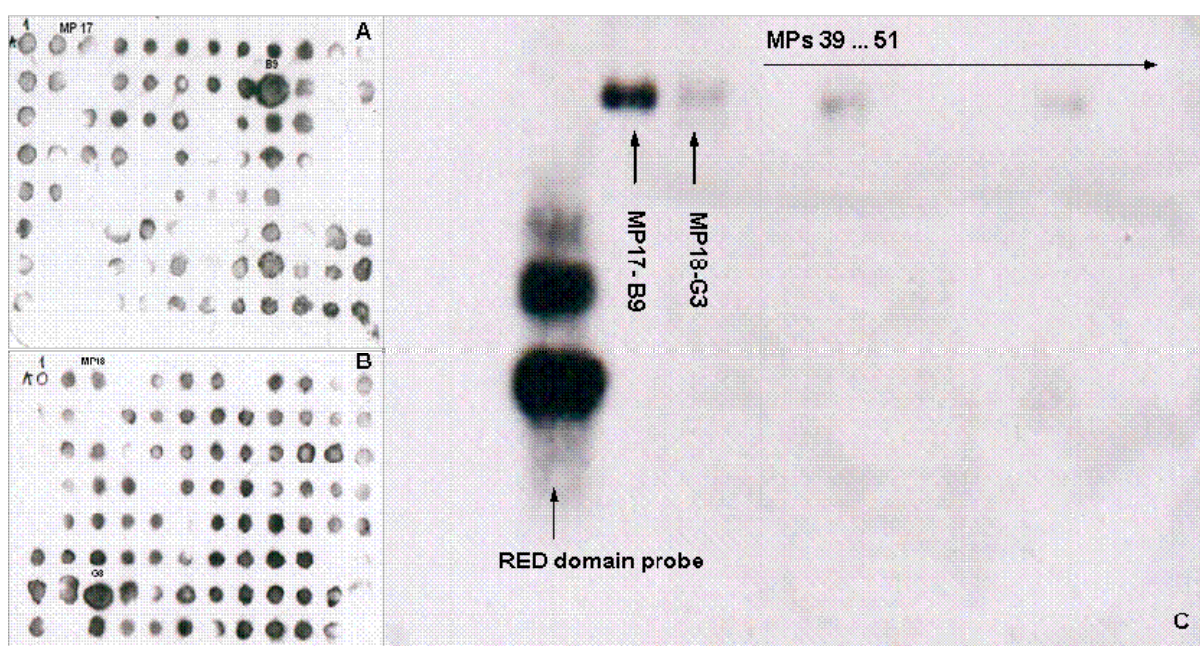


**Figure 23: Southern hybridization with KSnr7 probe.**

A and C: Plasmid DNA preparations from 50 microtiter plates (MP) enzymatically digested with *Bam*HI and electrophoretically separated. A shows MP1-MP29 divided in two rows. C displays MP30-MP51 also divided into two rows; in addition, run for MP10 from the first gel was repeated (first sample lane of the first row). Each agarose gel was photographed with a ruler on the left side in order to enable estimation of fragments sizes in the case of weak signal from marker (1 kb, first lane in each row). Second lane of each row contains KSnr7 fragment as a control for the positive signal generated with hybridization probe itself. KSpr-MI probe, used as a plasmid in A and as a fragment in C, served as a control for stringency of hybridization procedure. B and D: After gel electrophoresis samples were transferred onto a nylon membrane via Southern blot. A labelled KSnr7 probe was used for hybridization. B is showing positive signals coming from four MP preparations: MP15, MP17, MP18 and MP26. The fifth positive signal is illustrated in D and is coming from MP48.

These five MPs were then hybridised with the combined probe for the RED domain (RED-AN1 and RED-AN3) via the colony hybridization protocol. Hybridization, washing and

detection procedures were all done in the same day. The goal was to shorten the hybridization procedure from overnight to just 4-5 h of incubation and thus get reduced colony background signals from cell debris bound to a nylon membrane. Also washing steps were done under low stringency conditions (11.2.5) in order to enable heterologous hybridization probe to stay bound to the potentially homologous fosmid DNA. These colony hybridization experiments resulted in two clones from *P. chrysogenum* library as putative hosts of PKS gene clusters with the RED domain: The 17B9 and the 18G3 fosmids as shown in Figure 24 (A and B, respectively). In order to verify these results, Southern hybridization with *Bam*HI-restricted plasmid DNA from these two clones was performed with the combined RED domain probe. This time a strong signal came only from 17B9 candidate clone (see Figure 24, C).



**Figure 24: Hybridization with RED domain probe for fungal NR PKS.**

A and B representing developed autoradiographs from colony hybridization for MP17 and MP18 microtiter plates, with two positive signals coming from clones MP17-B9 and MP18-G3 respectively. C show autoradiography after Southern blot hybridization with *Bam*HI-restricted plasmid DNA of MP17-B9 and MP18-G3. On the right side are *Bam*HI plasmid preparations from MP39-MP51; not shown in complete.

Further on, plasmid DNA of 17B9 clone candidate was digested with numerous enzymes to look for a restriction pattern that enables the best separation of fragments via agarose gel electrophoresis. Among 12 restriction patterns, *Hind*III digestion was chosen since it gave rise to 12 optically easily detectable bands (i.e. fragments between 0.75-18 kb). The plasmid DNA from clone 18G3 was also restricted with the same enzymes, giving clear information that we are dealing with two clones from *P. chrysogenum* genomic library that carry different genomic inserts (data not shown). The *Hind*III-digested 17B9 plasmid DNA was hybridised with the RED domain probe in Southern blotting experiments aimed to reduce efforts and



costs required to prove the identity of the cloned *P. chrysogenum* genomic fragment, i.e. cloning and sequencing of only RED domain positive *Hind*III restriction fragments.

Despite repetition of these experiments, no fragment showed prominent hybridization signal with the RED domain probe (data not shown). In order to get more 17B9 plasmid DNA for transfer to the nylon membrane, of special importance for small restriction fragments, the Epicentre CopyControl<sup>®</sup> protocol for induction of high copy fosmid number per bacterial cell (12.17.2) was employed and hybridization experiments were repeated. The experiments encountered difficulties in aspect of transfer such DNA to a nylon membrane. A significant amount of sample-plasmid DNA stayed in slots of agarose gel, thus reducing the overall amount of DNA that could be transferred to the nylon membrane. Even after several repetition of the experiment under optimised conditions (higher amount of hybridization probe, prolonged hybridization and DNA transfer incubation times etc.) no significant signals could be observed.

The subcloning of *Hind*III fragments of 17B9 was an option for the previously described hybridization experiments with the RED domain probe. In this case 17B9 carrying an insert of interest was not induced to a high copy number prior to plasmid isolation. Twelve restriction fragments were separated via agarose gel, cloned in the pBluescript<sup>®</sup> vector and end-sequenced (from both sides of genomic insert if insert size was more than 1.2 kb). The sequencing of cloned fragments was realised by GATC Biotech (Konstanz, Germany) with the T7 and T3 primer pair. The average size of end sequences was 500-800 bp. The end-sequencing revealed that at least 15 different *Hind*III restricted fragments were generated from 17B9. The sum of all released fragments led to the estimation of ~ 50 kb size for this genomic insert (see Table 11). The BLAST results of deduced amino acid sequences for all 15 end sequenced inserts with the highest similarities to sequences in the GenBank are summarised in Table 11. The band designated as fragment 10 (F10, see Table 17) contained actually three different fragments (17B9F10a, 17B9F10b and 17B9F10c; Table 11), while band named as F12 (not shown in Figure 25) contained two individual fragments (17B9F12a and 17B9F12b; see in Table 11). Three end sequences: 17B9F2-T3, 17B9F2-T7 and 17B9F7-T7, were of rather very low quality with numerous ambiguous letters (i.e. not representing nucleotides) over the whole sequence length. Even repetition of sequencing with freshly made plasmid preparation of the subclones carrying F2 and F7 fragments did not improve the sequence quality. Thus, no homology data are available for these three end sequences. Beside that, three end sequences showed no homology to any known fungal sequences in database: 17B9F1-T7, 17B9F5-T7 and 17B9F12b-T7. They showed homology to mostly bacterial

proteins that are either not associated with PKS biosynthesis clusters, or not exclusively associated with these type of clusters (e.g. 17B912b-T7: Cytochrome c oxidase subunit I *Ehrlichia canis*; Q3YQV6). For more details see Table 11.

All homologies to bacterial, plant or animal sequences span relatively small amino acid region of subject sequence (from 20-100 aa). However, if the similarity value is good, the inferred gene function should be correct. These similarity values are shown within Table 11.

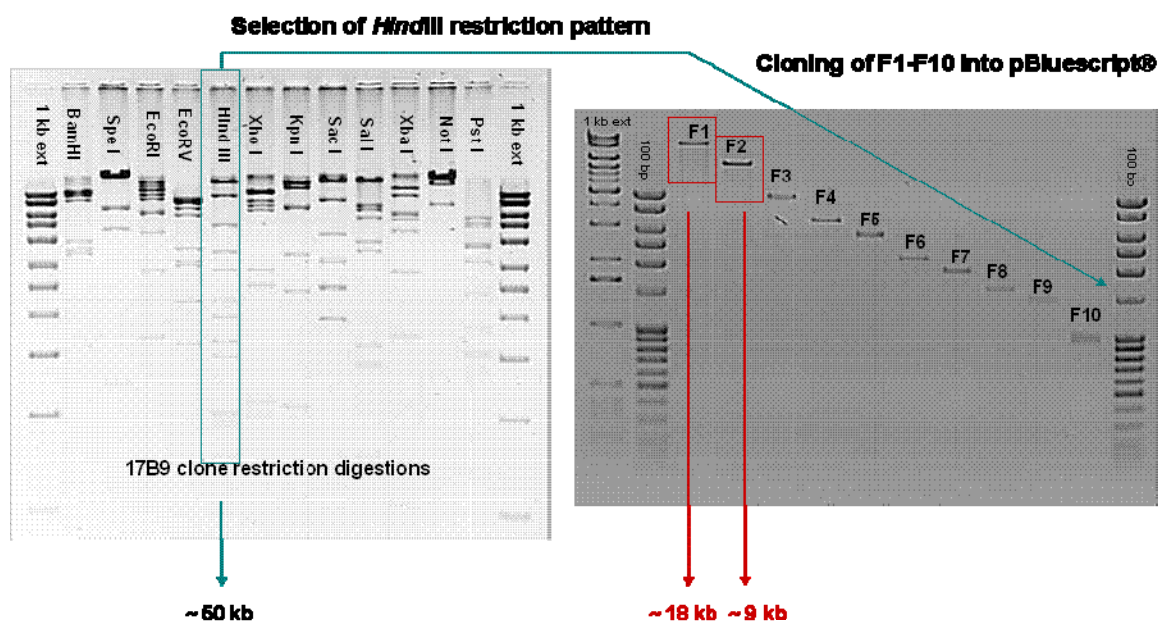
Several putative fungal protein homologues, with homology regions of 30-60 aa and the identity value between ~40-75% aa were found for the following sequences: 17B9F3-T3 end sequence (es), 17B9F3-T7 es, 17B9F4-T3 es, 17B9F4-T7 es, 17B9F6-T3 es, 17B9F9-T3 es and 17B9F11-T7 es. For the information on their inferred gene functions see Table 11.

Moreover, up to 85% identity to putative flavin binding monooxygenase of *A. clavatus* (GenBank: A1CGT2) and *Neosartorya fischeri* (GenBank: A1CXZ8) was found over more than 220 aa homology stretch for 17B9F1-T3 es. Similarity to putative C6 zinc finger domain protein (i.e. putative function in transcriptional regulation) of *A. clavatus* (GenBank: A1CBQ9) and *A. fumigatus* (GenBank: B0XR17) is shared between three end sequences: 17B9F5-T3 17B9F10a-T7 and 17B9F12a-T7. A homology to *A. niger* putative proline permease (GenBank: A2QHN9) was detected for both end sequences of 17B9F7 subclone (T3: 72 aa, 61% identity; T7: 191 aa, 80% identity). Likewise, a homology to putative amino acid permease of *Coccidioides immitis* (Q1DZQ4) was found within 17B9F10c-T7 sharing 72% identity over the 43 amino acid stretch. The putative JmjC domain protein, as demethylase signature motif [236], was identified within 17B9F8-T7 end sequence and showed homology to *A. niger* protein (GenBank: A2QHN7; 85 aa, 34% identity), while 17B9F10b-T7 end sequence showed significantly higher homology to *A. fumigatus* protein (B0Y5G6; 197 aa, 69% identity) for the same predicted function.

In sum, analysis of end sequences of all 15 *Hind*III 17B9 fragments revealed no homology to deduced amino acid sequences of known fungal PKS genes. Flavin-binding monooxygenases are frequently associated with the PKS biosynthesis clusters and are involved in tailoring reactions that often follow the release of nascent polyketide. The similarity to these enzyme functions was high (up to 85% identity). Moreover, the 17B9F9-T3 deduced amino acid sequence of F9 fragment showed 41% identity to acyltransferase (AT) domain of bovine fatty acid synthase (FAS; GenBank: Q71SP7) over homology stretch of 39 aa residues. This sequence was presumably the only fatty acid synthase (FAS) among numerous homology matches that were found over a rather small homology stretch within F9-T3 end sequence (not all shown in Table 11).

At this point, it is important to note that the homology search for new fungal genomic sequences are encountered with considerable problem due to the scarce availability of fungal sequences in gene databases as compared to bacterial ones (see 2.4.3). In addition, 18 kb size of F1, the largest fragment that was subcloned from the 17B9 clone (see Figure 25), leaves enough space for even two PKS genes next to each other, when keeping in mind that the average size of fungal PKS gene is 6-8 kb [23]. Beside all that, there is no information on sequence content for the second largest fragment F2 that was estimated to be 9 kb large. In accordance to previous statement for an average size of fungal PKS gene, it might be possible to place one fungal PKS gene cluster within the F2 fragment as well.

Thus, PCRs with F1 and F2 plasmid DNA and KS domain specific primers of LC and KHKS series were employed under conditions that previously enabled amplification of the partial KS sequences using the *P. chrysogenum* genomic DNA as a template. No fragment was cloned that correspond to any PKS biosynthesis cluster or to associated genes (data not shown). In order to investigate the eventual presence of PKS gene clusters within these two fragments, more sequence data are required from both fragments. According to data available up to this time point, it is not possible to anticipate the presence of fungal Type I PKS gene cluster within the 17B9 clone from *P. chrysogenum* E01-10/3 genomic library.



**Figure 25: Cloning *HindIII* restriction fragments of 17B9 clone candidate.**

On the left side are shown fragment patterns generated via 12 different restriction endonucleases. Enzymatic digestion with *HindIII* was chosen for subcloning of 17B9 clone candidate into the pBluescript® vector (on the right). Overall size of genomic insert cloned within 17B9 clone is estimated to be ~50 kb, while the size of the largest cloned fragment was identified to be ~18 kb.

## Results and Discussion

**Table 11: Sequences homologies of deduced amino acid sequences from 17B9 subcloned fragments.**

Sequence name <sup>(a)</sup>	Fragment size <sup>(b)</sup>	Sequence homologues <sup>(c)</sup>	Identities (%) <sup>(d)</sup>	Positives (%) <sup>(d)</sup>	Hom. str. (aa) <sup>(e)</sup>	References <sup>(f)</sup>
17B9F1 (T3)	18.0 Kb	▪ Putative flavin binding monooxygenase, <i>A. clavatus</i>	66	80	223	A1CGT2; Fedorova <i>et al.</i> , 2008;
		▪ Putative flavin binding monooxygenase, <i>Neosartorya fischeri</i>	64	78	223	A1CXZ8; Fedorova <i>et al.</i> , 2008
17B9F1 (T7)		▪ Putative uncharacterised protein, <i>Desulfovibrio vulgaris</i> (bacterium)	35	49	59	Q728B0; Heidelberg <i>et al.</i> ,
		▪ Elongation factor G1 ( <i>fusA1</i> ), <i>Hahella chejuensis</i> (bacterium)	29	48	74	2004; Q2S909; Jeong <i>et al.</i> , 2005
17B9F2 (T3)	9.0 kb	▪ No significant similarity to any sequence (low sequence quality)	-	-	-	-
17B9F2 (T7)		▪ No significant similarity to any sequence (low sequence quality)	-	-	-	-
17B9F3 (T3)	5.0 Kb	▪ Putative mating type switch (DNA repair) protein, <i>A. fumigatus</i>	50	61	138	Q4WNMA4; Nierman <i>et al.</i> ,
		▪ Putative dihydrofolate reductase, <i>Neosartorya fischeri</i>	38	54	59	2005; A1DGI8; Fedorova <i>et al.</i> , 2008
17B9F3 (T7)		▪ Putative progesterone binding protein, <i>A. clavatus</i>	75	86	47	A1CGT6; Fedorova <i>et al.</i> , 2008;
		▪ Putative helicase IV protein, <i>Vibrio cholerae</i> (bacterium)	33	48	80	Q9KLM6; Heidelberg <i>et al.</i> , 2000
17B9F4 (T3)	3.2 Kb	▪ Putative DNA binding protein B7F21.040, <i>Neurospora crassa</i>	37	62	29	Q9P3J0; Schulte <i>at al.</i> , 2000;
		▪ Putative membrane protein MembB, <i>Bifidobacterium longum</i> (bacterium)	33	48	68	Q5FBB1; Tanaka <i>at al.</i> , 2005
17B9F4 (T7)		▪ Putative sugar kinase, <i>Alcanivarax borkumensis</i> (bacterium)	51	59	27	Q0VPP3; Schneiker <i>at al.</i> , 2006;
		▪ Putative MFS transporter protein, <i>A. nidulans</i>	51	63	33	Q5B8H0; Galagan <i>et al.</i> , 2005
17B9F5 (T3)	2.8 Kb	▪ Putative C6 zinc finger domain protein (transcriptional regulation), <i>A. clavatus</i>	54	61	219	A1CBQ9; Fedorova <i>et al.</i> , 2008
17B9F5 (T7)		▪ Putative transposase family protein, <i>Shigella boydii</i> (bacterium)	36	54	91	B2TST7; Rasko <i>et al.</i> , 2008
		▪ Putative uncharacterised protein (unknown function), <i>Oryza sativa</i>	30	52	96	Q7XVR8; Feng <i>et al.</i> , 2002
17B9F6 (T3)	2.0 Kb	▪ Putative uncharacterised protein, <i>Tetrahymena thermophila</i> (bacterium)	29	41	85	Q22YM3; Eisen <i>et al.</i> , 2006 ;
		▪ Putative protein kinase, <i>A. fumigatus</i>	52	64	34	BOXZK9; Fedorova <i>et al.</i> , 2008
17B9F6 (T7)		▪ Putative cell division protein, <i>A. nidulans</i>	51	67	218	Q5B8H1; Galagan <i>et al.</i> , 2005
17B9F7 (T3)	1.7 Kb	▪ Putative prolane permease, <i>A. niger</i>	61	75	72	A2QHN9; Pel <i>et al.</i> , 2007

## Results and Discussion

Sequence name <sup>(a)</sup>	Fragment size <sup>(b)</sup>	Sequence homologues <sup>(c)</sup>	Identities (%) <sup>(d)</sup>	Positives (%) <sup>(d)</sup>	Hom. str. (aa) <sup>(e)</sup>	References <sup>(f)</sup>
17B9F7 (T7)		▪ Putative prolane permease, <i>A. niger</i>	80	90	191	A2QHN9; Pel <i>et al.</i> , 2007
17B9F8 (T3)	1.5 Kb	▪ Putative JmjC domain protein (demethylase signature motif), <i>A. niger</i>	34	55	85	A2QHN7; Pel <i>et al.</i> , 2007
17B9F8 (T7)		▪ Putative flavin binding monooxygenase, <i>Neosartorya fischeri</i>	85	93	74	A1CXZ8; Fedorova <i>et al.</i> , 2008;
		▪ Putative flavin binding monooxygenase, <i>A. clavatus</i>	79	86	74	A1CGT2; Fedorova <i>et al.</i> , 2008
17B9F9 (T3)	1.4 Kb	▪ EST an_1834, <i>A. niger</i>	58	83	31	A2QHP1; Pel <i>et al.</i> , 2007
		▪ AT domain of fatty acid synthase (from transcription), <i>Bos taurus</i>	41	53	39	Q71SP7; Roy <i>et al.</i> , 2003
17B9F9 (T7)		▪ No significant similarity to any sequence (low sequence quality)	-	-	-	-
17B9F10a (T7)	1.0 Kb	▪ Putative C6 zinc finger domain protein <i>A. fumigatus</i>	65	79	163	B0XR17; Fedorova <i>et al.</i> , 2008
17B9F10b (T7)	950 bp	▪ Putative JmjC domain protein (demethylase signature motif), <i>A. fumigatus</i>	69	77	197	B0Y5G6; Fedorova <i>et al.</i> , 2008
17B9F10c (T7)	900 bp	▪ Putative amino acid permease, <i>Coccidioides immitis</i>	72	90	43	Q1DZQ4; Birren <i>et al.</i> , 2005
17B9F11 (T7)	850 bp	▪ Putative transporter protein, <i>Streptomyces avermitilis</i> (bacterium)	34	52	69	Q82JP3; Ikeda <i>et al.</i> , 2003
		▪ Putative starch-binding protein (phosphodiesterase activity), <i>A. oryzae</i>	42	56	50	Q2UED0; Machida <i>et al.</i> , 2005
17B9F12a (T7)	800 bp	▪ Putative C6 zinc finger domain protein (transcriptional regulation), <i>A. clavatus</i>	47	61	150	A1CBQ9; Fedorova <i>et al.</i> , 2008
17B9F12b (T7)	750 bp	▪ Hypothetical protein Ccur92 <i>Campylobacter curvus</i> (bacterium)	30	54	50	A7H056; Fouts <i>et al.</i> , 2007,
		▪ Cytochrome c oxidase, subunit I <i>Ehrlichia canis</i> str. Jake (bacterium)	27	43	97	Q3YQV6; Mavromatis <i>et al.</i> , 2006

<sup>(a)</sup> Fragment name; T7 or T3 in brackets are standing for primer used for sequencing. <sup>(b)</sup> Fragment size estimated by comparison with referent fragment sizes via agarose gel electrophoresis <sup>(c)</sup> Putative sequence homologues retrieved via BLASTX search showing the first match for homology search. In addition, second putative homologue, if given, represents the closest fungal protein that is available: If the first match is not a fungal protein, second match that differs in predicted function from a first match in BLAST search, or a first characterised putative gene homologue available in TrEMBL database I do not understand, please shorten sentence <sup>(d)</sup> Identity/similarity (%) of the deduced amino acid sequences. <sup>(e)</sup> Homology stretch represents a number of amino acid residues within deduced subject sequence from TrEMBL database which was found to share homology with a given query sequence. <sup>(f)</sup> Reference combines sequence TrEMBL accession number, reference to author and the year for first deposition of a sequence in GenBank database.

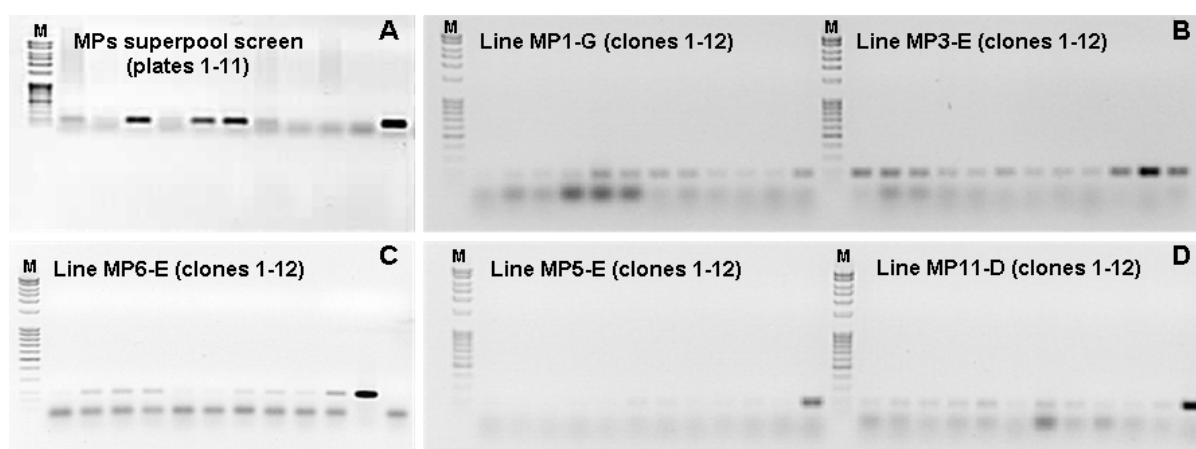
### 7.3 Screening of the genomic library via PCR

PCR studies on the genetic potential of *P. chrysogenum* E01-10/3 for polyketide production revealed the presence of three partial PKS sequences as putative members of fungal NR clade III (see 6.2). The KHKS1 and KHKS32 sequences derived from KS domains (Table 5), while AT-9-11 was amplified with primers specific for AT domains of fungal PKS genes (Table 6). Since the structure of sorbicillactone A indicated involvement of the fungal NR PKS clade III in its biosynthesis (see 3), these three partial PKS sequences qualified themselves as potential cluster candidates for the sorbicillactone A production. For all three PKS sequences, specific primers that originate exactly from the amplified DNA stretches were designed, and PCR screening reactions were optimised to enable amplification of specific PCR products. Each screening primer pair was designed in such a manner that amino acid sequences, from which they originate, were inside the sequence spanned by degenerate primers, since it can always be some mismatch pairing between degenerate primers and the DNA template. The PCR protocols with either complex superpool plasmid DNA as a template or a whole cell PCR (12.8.1) were employed during the screening procedure. As a result, individual fosmid clones that corresponded to each cluster candidate were localised within the genomic library of *P. chrysogenum*. The identity of genomic inserts cloned within fosmids was investigated by end sequencing and further subcloning of endonuclease-released fragments shared between fosmid clones that belong to a particular gene cluster. The aim was to get detailed information on PKS domain organization of these three clusters and to evaluate which cluster might be a probable candidate for the sorbicillactone A biosynthesis.

#### 7.3.1 Screening for KHKS1 cluster

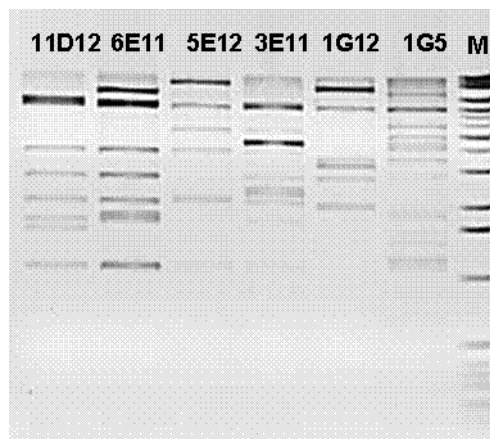
A screening for the KHKS1 cluster candidate was performed with sorb-spec2-for and sorb-spec2-rev primer pair in order to amplify an expected 193 bp PCR product (Table 27). Initially, all superpool plasmid DNAs corresponding to MP1-MP51 (excluding MP44 that was lost during library preparation) were used as templates in PCRs with this primer pair. The amplification of correctly sized PCR product was detected for five MPs: MP1, MP3, MP5, MP6 and MP11 (Figure 26A). Later on, eight colony pools were made from every microtiter plate, each corresponding to 12 individual fosmid clones of one MP-row (12.16). In sum, 60 pools were created for 5 microtiter plates. One  $\mu$ l aliquot of the cell suspension from each pool was used as a template in protocols for the whole-cell PCR (12.8.1). Rows 1G, 3E, 5E, 6E and 11D were positive for amplification, and thus all individual clones belonging to these

rows were further screened in the whole-cell PCR setups. Apart from lane 1G that gave two clone candidates (1G5 and 1G12), exactly one positive clone was found for each of four remaining lanes: 3E11, 5E12, 6E11 and 11D12 (Figure 26: B, C and D). The distinction between real and false positives was rather challenging, since the correct PCR products and primer dimers overlapped on agarose gel (see Figure 26). The plasmid DNA of these 6 individual fosmid clone candidates was restricted with several endonucleases in order to check whether they contain the common part of *P. chrysogenum* genome. *Bam*HI digestion of fosmid candidates clearly showed that they share numerous fragments of different sizes within their restriction profiles (see Figure 27) and thus supported the PCR screening results.



**Figure 26: PCR screening for KHKS1 gene cluster.**

M stands for 100 bp ladder as a fragment size standard. The strongest bands are considered as positive signals for amplification of specific 193 bp PCR product with KHKS1 screening primer pair (sorb-spec2-for and sorb-spec2-rev). **A** displays results of MPs (microtiter plates) superpool screening: five positive signals from five MPs: MP1, MP3, MP5, MP6 and MP11. **B**: On the left side screening results for lane MP1-G; positive signals for two fosmids - 1G5 and 1G12. On the right side are results for lane MP3-E with the strongest signal correlating with fosmid 3E11. **C** shows positive signal from 6E11 fosmid clone of MP6-E lane. **D** represents PCR screening results for two lanes: on the left side MP5-E with 5E12 positive fosmid clone and on the right positive signal from 11D12 within lane MP11-D.



**Figure 27: BamHI restriction profiling of KHKS1 gene cluster.**

M stands for 1 kb ladder as a fragment size standard. *Bam*HI digestion showing that 11D12, 6E11, 5E12, 3E11 and 1G12 fosmids share numerous fragments of different size within their restriction profiles. *Bam*HI digestion profile of 1G5 is rather alternate compare to other restricted DNAs.

The position of the cloned KHKS1 cluster within these five fosmid was investigated via end sequencing of each fosmid (data shown in Table 12), while the first information on PKS cluster organisation came from further subcloning and sequencing of the selected fragments shared among analysed fosmids (data shown in Table 13). The pCC1<sup>TM</sup>/pEpiFOS<sup>TM</sup> forward primer (FP) and pCC1<sup>TM</sup>/pEpiFOS<sup>TM</sup> reverse primer (RP) were used for end-sequencing from the CopyControl<sup>TM</sup> pCC1FOS<sup>TM</sup> genomic library preparation vector (Table 27). The end-sequencing provided information that the 1G5 fosmid could contain the intact core of the KHKS1 cluster, since both terminal sequences did not correspond to any putative PKS sequence. It could not be estimated if this is the case for 1G12, 3E11 and 5E12 fosmids as well, since 1G12-RP, 3E11-FP and 5E12-FP end sequences were of very low quality and gave no significant sequence similarities via BLAST search. Beside, the *Bam*HI restriction profile of 1G5 fosmid was rather puzzling, generating a number of fragments that were not present in restriction profiles of other KHKS1 fosmid candidates (Figure 27), thus indicating that this fosmid may be a false positive. Two KHKS1 fosmid candidates, 6E11 and 11E12, contained at one end of their inserts the same partial PKS sequences resembling the AT domain of a putative PKS of *A. terreus* (GenBank: Q0C8G5; see Table 12), therefore supporting the presence of a PKS on KHKS1 cluster candidate. For further subcloning of the KHKS1 cluster, the *Bam*HI-fragments F7, F10, F13, F15 and F17 were chosen to be sequenced from 11D12 fosmid, since they were suspected to be relevant for this PKS gene cluster. For the construction of the subclone library, gel purified DNA of these five fragments was ligated into the pBluescript<sup>®</sup> II phagemide vector (2.9 kb in size). In addition, fragment F16 was co-cloned with fragment 15 (similar in size, see Table 13). The BLAST results for the deduced amino acid T7/T3 end sequences of these six subclones, with the highest similarities to sequences in GenBank, are summarised in Table 13. The deduced amino acid F7-T3 sequence showed 74% identity to the last part of the SAT domain and the beginning of KS domain from *A. nidulans* putative PKS (GenBank, Q5ATJ7). In addition, the same sequence shared around 50% identity with the SAT-KS-amino acid-stretch of the citrinin PKS from *M. purpureus* (GenBank: Q65Z23, citrinin biosynthesis), the first genetically characterised member of the fungal NR PKS clade III [33], as it was already mentioned. Therefore, one side of ~4 kb large F7 most probably encodes the part of the SAT and the KS domains of putative fungal PKS. In addition, the whole fragment F15 (~1.8 kb) was similar to the part of the KS and the AT domains from PKSs of *A. nidulans* and *M. purpureus* (GenBank: Q5ATJ7 and Q65Z23, respectively) meaning that this genomic



fragment of *P. chrysogenum* could be part of an extension component of the NR fungal type I PKS gene [167].

The second end sequence of F7 (F7-T7), when translated to protein sequence, shares moderate homology to a putative ABC transporter protein of *A. niger* (GenBank: A2QM91). In addition, F16 (~1.6 kb) exhibits significant homology (56-58% identity) to the same transporter protein of *A. niger* over both sequenced parts of the insert, providing further support that one gene encoding putative transporter protein could be in close environment of the PKS core of this gene cluster. At least 1 kb DNA (~314 aa) of F10 (T7 and T3 es) shows significant homology (62% identity) to a putative flavin binding monooxygenase of *A. terreus* (GenBank: Q0CGP4). The whole F17 (~1.2 kb) exert high homology (71-87% identity) to the already mentioned putative flavin binding monooxygenase of *A. terreus* (GenBank: Q0CGP4; see before in text). Thus, it is probable that the KHKS1 gene cluster could encode one flavin binding monooxygenase enzyme.

The second end sequence of F10 (F10-T3) shares high similarity (86% identity) with putative C2H2-zinc-finger domain protein of *A. terreus* (GenBank: Q0CGP3). The F13-T3 es showed less homology (62% identity) to another putative zinc finger domain protein of *A. fumigatus* (GenBank: Q4WD01). Beside, this sequence shared 32% identity with one characterised C2H2 zinc finger domain DNA-binding regulatory protein (AmdX) from *A. nidulans* (GenBank: P79045). AmdX is an activator of *amdS* gene expression [237]. The *amdS* gene of *A. nidulans* encodes an acetamidase enzyme and is required for acetamide catabolism. The *amdS* gene expression is one of the best-studied systems for eukaryotic structural gene expression. C2H2-zinc-binding-motifs recognise DNA sequences by binding to the major groove of DNA but can also bind to RNA and protein targets (see Table 31). As a result, proteins that contain this motif are deemed to have diverse functions which complicate the function prediction for putative proteins that display homology to this motif. A rather moderate similarity of F13-T3 amino acid sequence (32% identity) to the AmdX transcriptional regulator can be suggestive for a putative protein function.

On the other side, F13-T7 has displayed high homology (76% identity) to another putative FAD oxidoreductase of *A. terreus* (GenBank: Q0D021). The same deduced sequence also shared 62% identity over the 74 aa with putative gene for sulphhydryl oxidase (*soxA*) from *A. niger* (GenBank: Q68CM8). Although this homology is involved over relatively short amino acid stretch, this could be an indication for the potential presence of one additional oxidoreductase in this gene cluster.

## Results and Discussion

**Table 12: Sequences homologies of deduced amino acid sequences for KHKS1 gene cluster.**

Fosmid name	Sequencing primers <sup>(a)</sup>	Sequence homologues <sup>(b)</sup>	Identities (%) <sup>(c)</sup>	Positives (%) <sup>(c)</sup>	References <sup>(d)</sup>
1G5 fosmid	FP	▪ Putative RTA1 domain protein, <i>A. fumigatus</i>	41	50	Q4WXU9; Nierman <i>et al.</i> , 2005
	RP	▪ Putative serine/threonine-protein kinase BUR1, <i>Kluyveromyces lactis</i>	32	46	Q6CRA9; Dujon <i>et al.</i> , 2004
1G12 fosmid	FP	▪ Putative carboxylesterase type B protein, <i>A. oryzae</i>	65	79	Q2UC28; Machida <i>et al.</i> , 2005
	RP	▪ No significant similarity to any sequence (low sequence quality)	-	-	-
3E11 fosmid	FP	▪ No significant similarity to any sequence (low sequence quality)	-	-	-
	RP	▪ Similar to extensin cyc 17 protein, <i>Catharanthus roseus</i>	33	39	Q39600; Ito <i>et al.</i> , 1998
5E12 fosmid	FP	▪ No significant similarity to any sequence (low sequence quality)	-	-	-
	RP	▪ Similar to extensin cyc 17 protein, <i>Catharanthus roseus</i>	83	91	Q1DHX7; Birren <i>et al.</i> , 2005
6E11 fosmid	FP	▪ Putative mitochondrial dihydroxy acid dehydratase, <i>A. nidulans</i>	69	80	Q5AZD4; Galagan <i>et al.</i> , 2005
	RP	▪ Putative PKS (AT domain), <i>A. terreus</i> NIH 2624	47	65	Q0C8G5; Birren <i>et al.</i> , 2005
11E12 fosmid	FP	▪ Putative extracellular serine-rich protein, <i>A. terreus</i> NIH 2624	36	45	Q0C8G4; Birren <i>et al.</i> , 2005
	RP	▪ Putative PKS (AT domain), <i>A. terreus</i> NIH 2624	51	72	Q0C8G5; Birren <i>et al.</i> , 2005

<sup>(a)</sup> Primers that were used for end sequencing from CopyControl™ pCC1FOS™ library preparation vector: FP stands for pCC1™/pEpiFOS™ forward primer, while RP represents pCC1™/pEpiFOS™ reverse primer. <sup>(b)</sup> Putative sequence homologues retrieved via BLASTX search showing the first match for homology search. <sup>(c)</sup> Identity/similarity (%) of the deduced amino acid sequences. <sup>(d)</sup> Reference combines sequence accession number in GenBank database, authors and the year for first deposition of a sequence in GenBank database.

**Table 13: Results of end sequencing of subcloned *Bam*HI restricted fragments from KHKS1 gene cluster.**

Sequence name <sup>(a)</sup>	Fragment size <sup>(b)</sup>	Sequence homologues <sup>(c)</sup>	Identity (%) <sup>(d)</sup>	Positive (%) <sup>(d)</sup>	Hom. str. (aa) <sup>(e)</sup>	References <sup>(f)</sup>
Fragment 7 (T7)	4.0 kb	▪ Putative transporter protein (ABC family; ATP binding), <i>A. niger</i>	32	47	94	A2QM91; Pel <i>et al.</i> , 2000
Fragment 7 (T3)		▪ Putative PKS (SAT + KS domain), <i>A. nidulans</i>	74	84	186	Q5ATJ7; Galagan <i>et al.</i> , 2005 ;
		▪ Citrinin PKS (PKSct) (SAT + KS domain), <i>M. purpureus</i>	48	67	176	Q65Z23; Shimizu <i>et al.</i> , 2005

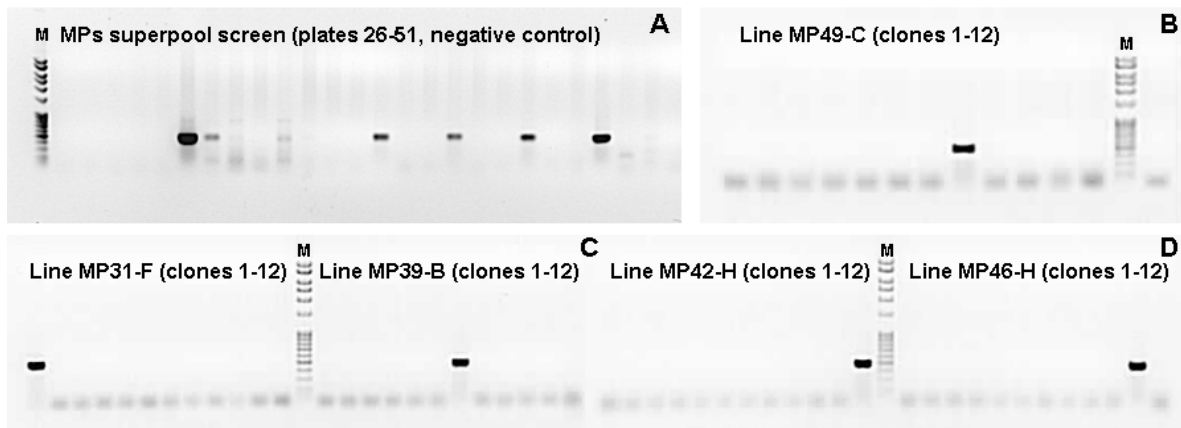
## Results and Discussion

Sequence name <sup>(a)</sup>	Fragment size <sup>(b)</sup>	Sequence homologues <sup>(c)</sup>	Identity (%) <sup>(d)</sup>	Positive (%) <sup>(d)</sup>	Hom. str. (aa) <sup>(e)</sup>	References <sup>(f)</sup>
Fragment 10 (T7)	2.0 kb	▪ Putative flavin binding monooxygenase, <i>A. terreus</i>	62	74	315	Q0CGP4; Birren <i>et al.</i> , 2005
Fragment 10 (T3)		▪ Putative flavin binding monooxygenase, <i>A. terreus</i>	59	63	88	Q0CGP4; Birren <i>et al.</i> , 2005
		▪ Putative C2H2 zinc finger domain protein, <i>A. terreus</i>	86	91	96	Q0CGP3; Birren <i>et al.</i> , 2005
Fragment 13 (T7)	2.2 kb	▪ Putative FAD oxidoreductase, <i>A. terreus</i>	76	87	63	Q0D021; Birren <i>et al.</i> , 2005
		▪ Sulphydryl oxidase (SoxA), <i>A. niger</i>	62	77	74	Q68CM8; Schaap <i>et al.</i> , 1995
Fragment 13 (T3)		▪ C2H2 zinc finger domain protein, <i>A. fumigatus</i>	62	75	126	Q4WD01; Nierman <i>et al.</i> , 2005 ;
		▪ C2H2 zinc finger domain DNA binding regulatory protein (AmdX), <i>A. nidulans</i>	32	52	46	P79045; Murphy <i>et al.</i> , 1997
Fragment 15 (T7)	1.8 kb	▪ Putative PKS (KS-AT domain), <i>A. nidulans</i>	51	69	276	Q5ATJ7; Galagan <i>et al.</i> , 2005 ;
		▪ Citrinin PKS (PKSct) (KS-AT domain), <i>M. purpureus</i>	32	51	301	Q65Z23; Shimizu <i>et al.</i> , 2005
Fragment 15 (T3)		▪ Putative PKS (KS-AT domain), <i>A. nidulans</i>	54	70	189	Q5ATJ7; Galagan <i>et al.</i> , 2005 ;
		▪ Citrinin PKS (PKSct) (KS-AT domain), <i>M. purpureus</i>	40	64	190	Q65Z23; Shimizu <i>et al.</i> , 2005
Fragment 16 (T7)	1.6 kb	▪ Putative transporter protein (ABC family; ATP binding), <i>A. niger</i>	56	75	339	A2QM91; Pel <i>et al.</i> , 2007
Fragment 16 (T3)		▪ Putative transporter protein (ABC family; ATP binding), <i>A. niger</i>	58	73	180	A2QM91; Pel <i>et al.</i> , 2007
Fragment 17 (T7)	1.2 kb	▪ Putative flavin binding monooxygenase, <i>A. terreus</i>	71	82	70	Q0CGP4; Birren <i>et al.</i> 2005
Fragment 17 (T3)		▪ Putative flavin binding monooxygenase, <i>A. terreus</i>	87	92	123	Q0CGP4; Birren <i>et al.</i> 2005

<sup>(a)</sup> Fragment name; T7 or T3 in parentheses are standing for primer used for sequencing. <sup>(b)</sup> Fragment size estimated by comparison with reference fragment sizes via agarose gel electrophoresis <sup>(c)</sup> Putative sequence homologues retrieved via BLASTX search showing the first match for homology search. In addition, a second putative homologue, if given, represents the closest fungal protein that is available if the first match is not a fungal protein, a second match that differs in predicted function from a first match in BLAST search, or a first characterised putative gene homologue available in GenBank database. I do not understand Please shorten sentence <sup>(d)</sup> Identity/similarity (%) of the deduced amino acid sequences. <sup>(e)</sup> Homology stretch represents a number of amino acid residues within deduced subject sequence from GenBank database which was found to share homology with a given query sequence. <sup>(f)</sup> Reference combines the sequence TrEMBL accession number, the reference author and the year for first deposition of a sequence in GenBank.

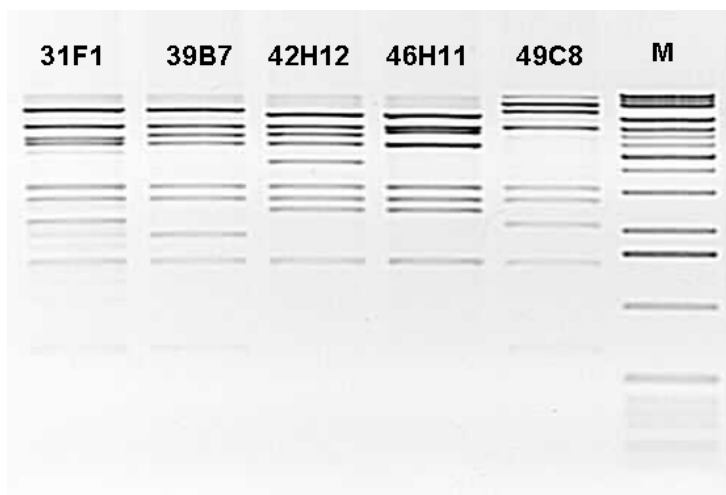
### 7.3.2 Screening for the AT-9-11 cluster

The screening for fosmid clones that contain the AT-9-11 cluster was performed analogously to the strategy applied for the KHKS1 cluster. The sorb 9-11 spec f1 and sorb 9-11 spec r1 primer pair were employed for the amplification of an expected 587 bp PCR product from the AT sequence previously amplified with degenerate primers (see B 6.1.2). Initially, all 50 pool plasmid DNA preparations, each one composed of plasmid DNA from 96 individual fosmid clones from one genomic library plate, were used as templates in PCR screening. The expected PCR product was amplified when plasmid preparations from MP31, MP39, MP42, MP46 and MP49 were used as PCR templates (Figure 28A). From these five positive plates, 60 lane pools were made and screened by whole-cell PCR (12.8). Further on, all individual fosmid clones from detected positive rows were used as the source of DNA for new whole-cell PCRs. Five putative AT-9-11 fosmid clones were finally identified: 31F1, 39B7, 42H12, 46H11 and 49C8 library clones (Figure 28: B, C and D). The plasmid DNA of these 5 individual fosmid clones were enzymatically restricted in order to investigate whether they belong to the same genomic region of *P. chrysogenum*. The *Bam*HI restriction profiles supported the assumption that these five fosmid clones share one genomic region (Figure 29).



**Figure 28: PCR screening for AT-9-11 gene cluster.**

M stands for 100 bp ladder as a fragment size standard. The strongest bands are considered as positive signals for amplification of specific 589 bp PCR product with AT-9-11 screening primer pair (sorb 9-11 spec f1 and sorb 9-11 spec r1). **A** displays results of microtiter plates (MPs) superpool screening - five positive signals from five MPs: MP31, MP39, MP42, MP46 and MP49. On **B** is to find positive signal from 49C8 fosmid clone of MP49-C lane. On the left side of **C** are screening results for row MP31-F with positive signal from 31F1 fosmid clone. On the right side of **C** are results for row MP39-F with the strongest signal correlating with fosmid 39B7. **D** represents PCR screening results for two row s: on the left side MP42-H with 42H12 positive fosmid clone and right side of **D** shows positive signal from 46H11 within row MP46-H.



**Figure 29: *Bam*HI restriction profiling of AT-9-11 cluster fosmid clones**

M stands for 1 kb ladder as a fragment size standard. *Bam*HI digestion shows that 31F1, 39B7, 42H12 and 49C8 share numerous fragments of different size within their restriction profiles.

As in the case of the KHKS1 cluster, the position of the cloned AT-9-11 cluster within fosmid-candidates was investigated via end sequencing of inserts (data shown in Table 14). The preliminary information on cluster organisation was provided via subsequent subcloning and sequencing of fragments that were released by *Bam*HI enzymatic digestion of individual fosmid inserts (data shown in Table 15).

Two fosmids, 42H12 and 46H11, had no end sequences that corresponded to any putative PKS cluster sequences. Therefore, they were considered as candidates for the cluster-containing fosmids. Fosmids 31F1 and 39B7 contain at one side of their inserts approximately the same part of an MT domain that belonged to a PKS most similar to the putative PKS R clade I from *Chaetomium globosum* (GenBank: Q2GR19) according to the phylogenetic analysis performed by Kroken *et al.* [167]. In addition, the reverse sequenced side of the 49C8 insert displayed homology to the SAT domain of putative PKS NR clade III (due to Kroken *et al.*) [167] also from *Chaetomium globosum* (GenBank: Q2GR18). Thus, the end sequencing provided information that the gene cluster covered by these fosmids could encode two different classes of fungal type I PKSs. Nevertheless, it is important to emphasise that domain analysis and phylogenetic studies on larger portions of these PKS genes may provide more reliable information upon the reducing type and the clade they belong to (see 8.4).

## Results and Discussion

**Table 14: End sequencing of fosmid clones from *P. chrysogenum* genomic library candidates for AT-9-11 gene cluster.**

Fosmid name	Sequencing primers <sup>(a)</sup>	Sequence homologues <sup>(b)</sup>	Identities (%) <sup>(c)</sup>	Positives (%) <sup>(c)</sup>	References <sup>(d)</sup>
31F1 fosmid	FP	▪ Putative PKS (MT domain), <i>Chaetomium globosum</i>	49	63	Q2GR19; Birren <i>et al.</i> , 2005
	RP	▪ Putative transcriptional regulator, RpiR family; <i>Pseudomonas syringae</i> (bacterium)	32	50	Q48BM3; Joardar <i>et al.</i> , 2005
39B7 fosmid	FP	▪ Putative protoporphyrinogeno oxidase, <i>Streptomyces avermitilis</i>	33	48	Q82KY9; Ikeda <i>et al.</i> , 2003
	RP	▪ Putative PKS (MT domain), <i>Chaetomium globosum</i>	51	65	Q2GR19; Birren <i>et al.</i> , 2005
42H12 fosmid	FP	▪ Putative actin cortical patch assembly protein (similar to Pan1), <i>A. fumigatus</i>	60	66	Q4WG58; Nierman <i>et al.</i> , 2005
	RP	▪ Putative uncharacterised protein (unknown function), <i>A. terreus</i>	32	47	Q0CBE5; Birren <i>et al.</i> , 2005
46H11 fosmid	FP	▪ Putative uncharacterised protein (unknown function) <i>Borrelia burgdorferi</i> (bacterium)	33	53	Q9S042; Fraser <i>et al.</i> , 1997
	RP	▪ MOUSE-5 (Nat13) isoform 5 of Q6PGB6, <i>Mus musculus</i> (mouse)	55	75	Q6PGB6-5; Carninci <i>et al.</i> , 2005
49C8 fosmid	FP	▪ Putative uncharacterised ASPCL type protein, <i>A. fumigatus</i>	47	62	Q4WG63; Nierman <i>et al.</i> 2005
	RP	▪ Putative PKS (SAT domain), <i>Chaetomium globosum</i>	53	69	Q2GR18; Birren <i>et al.</i> 2005

<sup>(a)</sup> Primers that were used for end sequencing from CopyControl<sup>TM</sup> pCC1FOS<sup>TM</sup> library preparation vector: FP stands for pCC1<sup>TM</sup>/pEpiFOS<sup>TM</sup> forward primer, while RP represents pCC1<sup>TM</sup>/pEpiFOS<sup>TM</sup> reverse primer. <sup>(b)</sup> Putative protein homologues retrieved via BLASTX search showing the first match for homology search. <sup>(c)</sup> Identity/similarity (%) of the deduced amino acid sequences. <sup>(d)</sup> Reference combines sequence TrEMBL accession number, authors and the year for first deposition of a sequence in GenBank database.

A subclone library was constructed in order to gain more information on the AT-9-11 cluster organisation. In overall, 10 *Bam*HI fragments (out of 15 initially recognised among all fosmid clones) were estimated to cover the core of putative PKS gene cluster were subcloned into pBluescript<sup>®</sup> II. The smallest fragments: F13 (~1.9 kb, present only within 39B7 clone), F14 (~1.5 kb, common for all clones) and F15 (~ 0.7 kb, shared among 31F1, 39B7 and 49C8 clones) were not cloned. Similarly, F4 (~ 8.0 kb) presumably be present only within the 46H11 clone and F8 (~ 4.5 kb), postulated to be unique for the 42H11 clone, was excluded from this subcloning. In addition, F9 (~3.2 kb) was repetitively cloned instead of cloning F10 that was estimated to be very similar in size (~2.8 kb). In a similar manner, attempts to clone F5 (~7.2 kb) failed. Instead, F7 (~6.0 kb) was repetitively cloned. Nevertheless, cloning of F6, initially recognised as a single ~7.0 kb fragment, yielded two fragments encoding different putative fungal proteins. The BLAST results for the deduced amino acid T7 and T3 es of 10 cloned fragments with the highest homologies to sequences deposited in GenBank are summarised in Table 15.

Numerous fragments showed homology to different domains of fungal type I PKS as shown in Table 15. The sequence F12-T3 (~2.1 kb) showed 64% identity to the beginning of a putative PKS from *Chaetomium globosum* (GenBank: Q2GR18) [167]. Additional homology to the citrinin PKS of *M. purpureus* (GenBank: Q65Z23; 31% identity) [33], revealed that homology to the SAT domain may be in question (Table 15). As previously mentioned (2.3.1), the SAT domain, located at the N-terminus of the fungal NR clade III PKS systems, is thought to be responsible for a transfer of  $\beta$ -reduced “advanced” starter unit (made by R PKS or FAS) onto an ACP domain of the same NR clade III PKS [50]. Moreover, the chemical structures of sorbicillactones support the model of reducing/non-reducing biosynthesis of these polyketides.

Concerning the estimated size of F12 (~2.1 kb) it is probable that it encodes for the whole SAT domain and a part of the KS domain of a putative NR PKS. The T3 end sequence of the much larger F7 (~6.0 kb) covers exactly the same SAT-KS part like the F12-T7 sequence. Since these two end sequences are almost identical, it is possible that F12 represent a shortened version of F7.

Moreover, both ends of F9 (~3.2 kb), when translated to protein sequences, show homology to the already mentioned two putative NR clade III PKS. It has to be emphasised that the identity is significantly higher [48% (T3)-58% (T7)] for the putative PKS from *Chaetomium globosum* (GenBank: Q2GR18) than as it is for the citrinin PKS of *M. purpureus* [GenBank: Q65Z23; 32% (T3)-40% (T7)]. The F9-T3 is homologous to the last part of the AT domain

and probably the beginning of the PT domain, while F9-T7 end sequence exerts homology to the ending part of the MT domain and the first half of the RED domain of both mentioned NR clade III PKS. This would mean that the whole fragment F9 probably spans the biggest part of the putative NR PKS: AT-PT-ACP-MT-RED domains. The last sequence that corresponds to the putative AT-9-11 PKS cluster, namely F14-T3, shows 70% and 28% identity to RED domains of the putative PKS from *Chaetomium globosum* (GenBank: Q2GR18) and the citrinin PKS from *M. purpureus* (GenBank: Q65Z23), respectively.

It is important to emphasise that a detailed analysis of PKS domain organization through the phylogenetic study of the whole PKS sequence and identification of active domain sites is necessary in order to make the correct prediction of the class and the clade of fungal type I PKSs. Furthermore, the subcloning of *Bam*HI restricted AT-9-11-fragments supported the results of fosmid end sequencing: Several sequences were found to show homology to putative members of NR clade III and R clade I PKS.

The fragment F7 showed homology to a putative NR clade III PKS over its T3 es, and shared significant homology with the putative members of R clade I PKS systems over its T7-sequenced end. The highest similarity (52% identity) F7-T7 sequence showed to be a part of the AT domain of the putative PKS of *A. oryzae* (GenBank: Q2U7I5). In addition, for the same sequence, significant similarity (40% identity) was found for the AT domain of the well-characterised gene for LDKS [91] from *A. terreus* that is involved in biosynthesis of the cholesterol-lowering drug lovastatin (GenBank: Q9Y7D5). The second part of the AT domain was identified within the T7 end sequence, while part of the MT domain was localised on the T3 end sequence of F11. These two DNA sequences, when translated to protein sequences, were the most similar to the putative PKS from *Chaetomium globosum* (GenBank: Q2GR19; 53% and 50% identity, respectively). The closest genetically characterised homologue of these two sequences was CDKS [92] of *P. citrinum* (GenBank: Q8JOF5; 37% identity for F11-T7 and 33% identity for F11-T3) responsible for biosynthesis of the methylbutyryl side chain during compactin biosynthesis. The last end sequence that corresponded to one part of a putatively R-clade I PKS from the analysed *P. chrysogenum* strain was the F3-T7 sequence. This sequence showed similarity to the part of the ER domain of the putative *Chaetomium globosum* PKS (GenBank: Q2GR19; 49% identity) as first match, and the CDKS of *Penicillium citrinum* (GenBank: Q8JOF5; 35% identity) as the first characterised putative homologue. Nevertheless, a more detailed analysis is required in order to judge upon the subclassification of the cloned putative fungal type I PKS from *P. chrysogenum*.



The remaining end sequences showed homologies to some non-PKS genes. For example, a significant similarity (46% identity) to a putative flavin-binding monooxygenase of *Coccidioides immitis* (GenBank: Q1DRI4), over a large homology stretch (170 aa), was identified for F2-T3 amino acid (see Table 15). In addition, F2-T3 sequence showed homology to the protein of *A. nidulans* that is similar to the bacterial salicylate 1-monooxygenase (SalA) (GenBank: Q9HFQ8; 39% identity, 191 aa). SalA has been demonstrated to be involved in resistance to terbinafine [238]. Beside that, homology to a putative flavin oxidoreductase gene was found over a very small part of the F14-T7. The first match in the homology search was the putative Zn-binding dehydrogenase of *Chaetomium globosum* (GenBank: Q2GZ40; 62% identity, 58 aa). Further on, in accordance to the BLASTX results, one characterised gene for the salicylate hydroxylase (*salA*) from the bacterium *Pseudomonas pseudoalcaligenes* (GenBank: Q4F8I6) shared 48% identity to F14-T7 as well. The *salA* gene is part of a gene cluster that is responsible for the conversion of salicylate to acetyl CoA [239]. This observation may indicate that at least one flavin-binding oxidoreductase gene could be associated with two different types of PKS genes within putative AT-9-11 cluster candidate that might be involved in the sorbicillactone A production. Furthermore, the F2-T7 translated sequence shared high (72%) identity with the putative Barren family protein of *A. clavatus* (GenBank: A1CD76) over a large amino acid stretch (223 aa). Besides, more than 1 kb of the deduced F1-T7 sequence exhibited the same level of homology (72% identity) to the putative Barren family protein of *A. terreus* (GenBank: Q0CPW2; over 339 aa). Similar homology (71% identity) to the same putative Barren family protein of *A. terreus* was found for T3 end sequence of F6a over the 334 aa as it is shown in Table 15. The deduced F1-T7 and F6a-T3 sequences covered the same homology stretch within putative Barren family protein of *A. terreus* (positions 532-872 aa in subject sequence) and displayed 96% identity to each other in the BioEdit pairwise alignment which allowed ends of sequences to slide over each other in order to perform an optimal alignment (data not shown). Since size of F1 is estimated to be much larger (~15.0 kb) compared to F6a (~7.0 kb), it is probable that F6a represents a shortened version of F1 that was found at one end of 31F1 fosmid insert, while F1 was found only in the *Bam*HI restriction profile of 49C8 fosmid clone. In contrast, the F2-T7 sequence did not pair with any of these two sequences, and showed homology to beginning of putative Barren family protein of *A. clavatus* (positions 19-242 aa). These results indicate that one protein of the Barren family, required for the correct sister-chromatid segregation during mitosis, is located relatively near to the core of the analysed PKS cluster of *P. chrysogenum*.

Two putative enzymatic functions involved in carbohydrate metabolic processes were detected within the course of subsequencing of the AT-9-11 cluster. A putative glycosyltransferase function was inferred based on homology that F6b-T7 shared with putative protein of this type from *Magnaporthe grisea* (GenBank: A4RLP8; over 128 aa). Additionally, homology to a putative glycosyl hydrolase from *A. terreus* (GenBank: Q0CXJ2; over 340 aa), an enzymatic function that hydrolyses *O*-glycosyl compounds, was identified for the aa sequence at F6b-T3 sequenced end. Since both putative functions were identified as end sequences of the ~7.6 kb F6b, it is adequate to expect that a part of the fragment F6b encodes for enzymes involved in carbohydrate metabolism.

Finally, it was not possible to distinctively attribute a putative function of the F1-T3 deduced aa sequence to only one gene product. This sequence shared the highest similarity (47% identity) with the putative adenosine triphosphate (ATP) synthase of *Neosartorya fischeri* (GenBank: A1DC56). This sequence also appeared to be a putative homologue of the *veA* gene of *A. parasiticus* (GenBank: Q69B22; 30% identity). The protein of the *veA* gene represents one broad transcriptional regulator involved in mycotoxin production and fungal development [240].

The PCR screening for the AT-9-11 PKS-cluster and further subsequencing of five AT domain-encoding fosmids identified two different PKSs. One PKS was similar to the NR clade III PKS members, including the well characterised citrinin PKS of *M. purpureus* [33]. The second PKS showed high similarity to members of the R clade I (LDKS [91] and CDKS [92]). Concerning the reducing/non-reducing character of the sorbicillactones biosynthesis, the results presented here implicate that the AT-9-11 cluster might be responsible for the biosynthesis of these polyketides.

Moreover, subsequencing of the AT-9-11 genome identified several non-PKS protein functions in the vicinity of two PKS genes: a putative fungal flavin-binding oxidoreductase, a putative fungal Barren family protein, as well as homology to two putative proteins involved in fungal carbohydrate metabolic processes.

In regard to their diversity, the results of the PCR screening for the AT-9-11 cluster are summarised in section 7.4.

## Results and Discussion

**Table 15: Results of end sequencing of subcloned *Bam*HI restricted fragments from AT-9-11 gene cluster.**

Sequence name <sup>(a)</sup>	Fragment size <sup>(b)</sup>	Sequence homologues <sup>(c)</sup>	Identities (%) <sup>(d)</sup>	Positives (%) <sup>(d)</sup>	Hom. str. (aa) <sup>(e)</sup>	References <sup>(f)</sup>
Fragment 1 (T7)	~ 15.0 kb	▪ Putative Barren family protein, (mitosis), <i>A. terreus</i>	72	81	339	Q0CPW2; Birren <i>et al.</i> , 2005
Fragment 1 (T3)		▪ Putative ATPase (alfa and beta subunits), <i>Neosartorya fischeri</i> ▪ Velvet global gene regulator (VeA), <i>A. parasiticus</i>	47 30	62 41	142 101	A1DC56; Nierman <i>et al.</i> , 2006; Q69B22; Calvo <i>et al.</i> , 2004
Fragment 2 (T7)	~ 12.5 kb	▪ Putative Barren family protein, (in mitosis), <i>A. clavatus</i>	72	76	223	A1CD76; Nierman <i>et al.</i> , 2006
Fragment 2 (T3)		▪ Putative flavin binding monooxygenase, <i>Coccidioides immitis</i> ▪ putative salicylate 1-monooxygenase, <i>Aspergillus nidulans</i>	46 39	62 57	170 191	Q1DR14; Birren <i>et al.</i> , 2005 Q9HFQ8; Graminha <i>et al.</i> , 2004
Fragment 3 (T7)	~ 10.0 kb	▪ Putative PKS (ER domain), <i>Chaetomium globosum</i> ▪ Compactine diketide synthase (ER domain), <i>Penicillium citrinum</i>	49 35	62 54	153 142	Q2GR19; Birren <i>et al.</i> , 2005 Q8JOF5; Abe <i>et al.</i> , 2002
Fragment 3 (T3)		▪ Putative RNA polymerase; <i>A. terreus</i> Putative actin cortical patch assembly protein, <i>A. fumigatus</i>	66 63	70 67	172 180	Q0CPW4; Birren <i>et al.</i> , 2005 Q4WG58; Nierman <i>et al.</i> , 2005
Fragment 6a (T7)	~ 7.0 kb	▪ Putative tRNA modification enzyme, <i>Botryotinia fuckeliana</i>	26	44	98	A6SM10; Birren <i>et al.</i> , 2005
Fragment 6a (T3)		▪ Putative Barren family protein, (mitosis), <i>A. terreus</i>	71	81	343	Q0CPW2; Birren <i>et al.</i> , 2005
Fragment 6b (T7)	~ 7.6 kb	▪ Putative glycosyltransferase, <i>Magnaporthe grisea</i>	68	78	128	A4RLP8; Dean <i>et al.</i> , 2005
Fragment 6b (T3)		▪ Putative glycosyl hydrolase (hydrolyzing O-glycosyl compounds), <i>A. terreus</i>	77	87	304	Q0CXJ2; Birren <i>et al.</i> , 2005
Fragment 7 (T7)	~ 6.0 kb	▪ Putative PKS (AT domain), <i>A. oryzae</i> ▪ Lovastatin diketide synthase (AT domain), <i>A. terreus</i>	52 40	68 58	332 348	Q2U715; Machida <i>et al.</i> , 2005 ; Q9Y7D5; Kennedy <i>et al.</i> , 1999
Fragment 7 (T3)		▪ Putative PKS NR clade III ( SAT+ KS domains), <i>Chaetomium globosum</i> ▪ Citrinin PKS (PKSct) (SAT+ KS domains), <i>Monascus purpureus</i>	54 38	73 55	261 294	Q2GR18; Birren <i>et al.</i> , 2005 Q65Z23; Shimizu <i>et al.</i> , 2005
Fragment 9 (T7)	~ 3.2 kb	▪ Putative PKS ( MT+ RED domains), <i>Chaetomium globosum</i> ▪ Citrinin PKS (PKSct) (MT+ RED domains), <i>Monascus purpureus</i>	58 40	68 57	351 261	Q2GR18; Birren <i>et al.</i> , 2005 Q65Z23; Shimizu <i>et al.</i> , 2005
Fragment 9 (T3)		▪ Putative PKS ( AT+ PT domains), <i>Chaetomium globosum</i> ▪ Citrinin PKS (PKSct) (AT+ PT domains), <i>Monascus purpureus</i>	48 32	65 56	229 233	Q2GR18; Birren <i>et al.</i> , 2005 Q65Z23; Shimizu <i>et al.</i> , 2005
Fragment 11 (T7)	~ 2.5 kb	▪ Putative PKS (AT domain), <i>Chaetomium globosum</i> ▪ Compactine diketide synthase (AT domain), <i>P. citrinum</i>	53 37	68 56	320 326	Q2GR19; Birren <i>et al.</i> , 2005 Q8JOF5; Abe <i>et al.</i> , 2002
Fragment 11 (T3)		▪ Putative PKS (MT domain), <i>Chaetomium globosum</i> ▪ Compactine diketide synthase (MT domain), <i>P. citrinum</i>	50 33	66 52	350 363	Q2GR19; Birren <i>et al.</i> , 2005 Q8JOF5; Abe <i>et al.</i> , 2002

## Results and Discussion

Sequence name <sup>(a)</sup>	Fragment size <sup>(b)</sup>	Sequence homologues <sup>(c)</sup>	Identities (%) <sup>(d)</sup>	Positives (%) <sup>(d)</sup>	Hom. str. (aa) <sup>(e)</sup>	References <sup>(f)</sup>
Fragment 12 (T7)	~ 2.1 kb	▪ Putative PKS ( SAT+ KS domains), <i>Chaetomium globosum</i>	54	73	262	Q2GR18; Birren et al., 2005
		▪ Citrinin PKS (PKSct) (SAT+ KS domains), <i>Monascus purpureus</i>	37	53	313	Q65Z23; Shimizu et al., 2005
Fragment 12 (T3)		▪ Putative PKS ( SAT domain), <i>Chaetomium globosum</i>	64	77	104	Q2GR18; Birren et al., 2005
		▪ Citrinin PKS (PKSct) (SATdomain), <i>Monascus purpureus</i>	31	50	95	Q65Z23; Shimizu et al., 2005
Fragment 14 (T7)	~ 1.5 kb	▪ Putative Zn binding dehydrogenase (FAD-oxidoreductase; not as part of PKS), <i>Chaetomium globosum</i>	62	79	58	Q2GZ40; Birren et al., 2005
		▪ Salicylate hydroxylase (SalA), <i>Pseudomonas pseudoalcaligenes</i>	48	68	64	Q4F8I6; Fujihara et al., 2006
Fragment 14 (T3)		▪ Putative PKS ( RED domain), <i>Chaetomium globosum</i>	70	82	168	Q2GR18; Birren et al., 2005
		▪ Citrinin PKS (PKSct) (RED domain), <i>Monascus purpureus</i>	28	43	165	Q65Z23; Shimizu et al., 2005

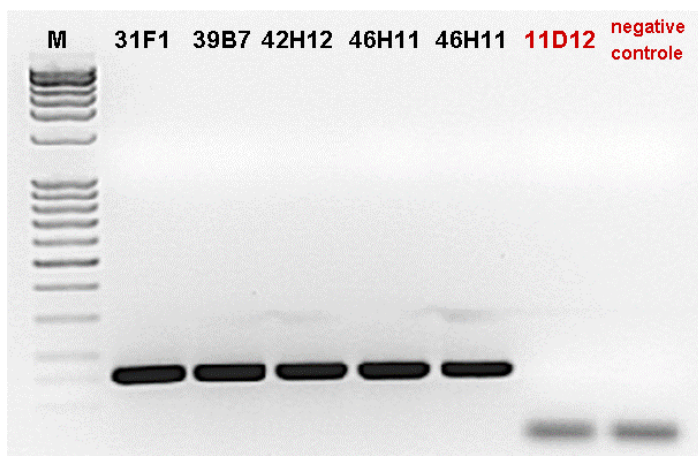
<sup>(a)</sup> Fragment name; T7 or T3 in brackets stand for primer used for sequencing. <sup>(b)</sup> Fragment size estimated by comparison with referent fragment sizes via agarose gel electrophoresis

<sup>(c)</sup> Putative protein homologues retrieved via BLASTX search showing the first match for homology search. In addition, the second homologue, if given, represents the closest putative fungal homologue that is available if the first match is not a fungal protein, the second match that differs in predicted function from a first match in BLAST search, or a first characterised putative gene homologue available in GenBank database <sup>(d)</sup> Identity/similarity (%) of the deduced amino acid sequences; <sup>(e)</sup> Homology stretch represents the number of amino acid residues within the deduced subject sequence from GenBank database which was found to share homology with the given query sequence. <sup>(f)</sup> Reference combines sequence TrEMBL accession number, referent author and the year for first deposition of a sequence in GenBank.

### 7.3.3 Screening for KHKS32 cluster

The first step to disclose the identity of KHKS32 cluster candidate was to check whether the KHKS32 amplified partial KS sequence of the *P. chrysogenum* genome originates from the same PKS cluster as AT-9-11 sequence amplified with degenerate primers for the AT domain. An indication that this may be the case was the fact that both partial PKS sequences showed similarity to the citrinin PKS from *M. purpureus* (GenBank: Q65Z23, citrinin biosynthesis) in BLAST search: 61% identity for KHKS32 and 52% identity for AT-9-11 (Table 5 and Table 6 respectively). On the contrary, the KHKS1 PKS sequence, amplified with the same degenerate primer pair as KHKS32 sequence (see 6.1.1) displayed rather different homology to members of the NR clade III (see Table 5) and thus was considered to be likely amplified from another PKS gene.

After KHKS32, KHKS42 and KHKS46 partial KHKS sequences were aligned in the BioEdit (seem to originate from the same PKS - 6.1.1), the specific primer pair was designed to cover the sequence regions of 100% DNA identity. The expected size of the PCR product with the 32 triple f1 and 32 triple r1 primer pair was 160 bp. The plasmid DNAs from all five clones, previously identified to carry parts of AT-9-11 cluster candidate (see 7.3.2), were used in PCRs with this KS-KHKS32 amplification primer pair. In addition, the plasmid DNA of clone 11D12 that was proven to carry parts of the KHKS1 cluster candidate (see B:7.3.1), was used as a negative control. The KHKS32 specific primer pair was able to amplify the expected size PCR product from all five AT-9-11 fosmid inserts, while the PCR was negative when the 11D12 fosmid DNA served as a template (Figure 30). This assessment proved the KHKS32 and the AT-9-11 partial PKS sequences to be originally amplified from two different PKS domains (KS and AT, respectively) of the same *P. chrysogenum* putative PKS cluster.



**Figure 30: PCR for the KHKS32 cluster.**

“M” stands for 100 bp ladder as a marker of fragments sizes 31F1, 39B7, 42H12, 46H11 and 46H11 are fosmid clones that carry the AT-9-11 cluster candidate from the *P. chrysogenum* genome (shown in black). The 11D12 fosmid represents the KHKS1 cluster (shown in red). The negative control represents PCR that was carried out without a DNA template (last lane, shown in red). The KS-KHKS32 primer pair amplified 160 bp large PCR products from all five fosmids of the AT-9-11 PKS cluster. The PCR was negative for fosmid 11D12 as well as for the negative control (absence of specific product amplification).

#### 7.4 Summary of *P. chrysogenum* E01-10/3 genomic library screening

Despite efforts taken to screen the genomic library by the use of different hybridization protocols (12.17), no fosmid clone was found to carry the genomic region involved in the biosynthesis of sorbicillactone A. On the other side, performed studies on PKS genetic potential of the *P. chrysogenum* E01-10/3 strain resulted in amplification of three partial PKS sequences presumably belonging to the NR clade III based on BLAST search results (see 6.1) and subsequent phylogenetic studies (see 6.2). Two of these sequences, namely KHKS1 and KHKS32, were amplified from the KS domain of the fungal type I PKS (see Table 5), while AT-9-11 derived from an AT domain (see Table 6). All sequences were initially amplified with degenerate primer sets as described previously (see 6.1). In the course of the PCR screening of the genomic library, specific primer pairs that encompass each sequence were designed and applied in complex screening procedures (see 7.3.1, 7.3.2 and 7.3.3).

In the case of the putative KHKS1 PKS cluster (see 7.3.1) six library clones appeared to carry a part of the KS domain for which the sorb-spec2-for and sorb-spec2-rev primer pair was designed. Further subsequencing of *Bam*HI-digested 11D12 fosmids generated several DNA sequences with similarity to the fungal type I PKS genes. Several end sequences from 11D12 genomic insert corresponded to putative SAT, KS and/or AT PKS domains. These end sequenced fragments showed high similarity to the citrinin PKS of *M. purpureus* [33] which was not the case for the partial KS sequence initially amplified with degenerate primers (see

6.1.1). In addition, subcloning and subsequencing of 11D12 insert indicated the presence of one putative fungal ABC transporter gene and one putative fungal flavin-binding monooxygenase gene that could be part of this cluster as well (see 7.3.1). Moreover, the BLAST-search for KHKS1 PKS cluster sequences, indicated similarity to the class of putative C2H2 zinc finger domain proteins (see 7.3.1).

The PCR screening procedure for the AT-9-11 PKS cluster candidate resulted in the identification of five AT domain-encoding fosmids within the genomic library (see 7.3.2). The end sequencing of the five *P. chrysogenum* genomic regions indicated the presence of two apparently different fungal type I PKSs that may differ in respect to their PKS-reducing potential. In order to obtain more DNA sequence information, ten *Bam*HI-restricted fragments from the putative AT-9-11 gene cluster, originating from different fosmid clones, were subcloned and sequenced as mentioned before (see 7.3.2). The subsequencing results (Table 15) support the assumption of the co-existence of two different PKS systems.

One of two PKS showed consistent homology to the NR clade III PKS members, including the well characterised citrinin PKS of *M. purpureus* [33]. The BLAST search revealed that several AT-9-11 PKS cluster sequences together span a typical domain organization for one fungal NR class III PKS, along with the optional MT domain and the relatively rare RED domain (see 2.3.1 for NR PKS): SAT-KS-AT-PT-ACP-MT-RED (Table 15). On the other hand, several other AT-9-11 PKS cluster sequences indicated the presence of one reducing PKS as well. Among these sequences, high similarity to members of the R clade I (LDKS [91] and CDKS [92]) was found for putative AT, MT and ER domains (Table 15).

Further on, subsequencing of the AT-9-11 genomic region identified several putative non-PKS protein functions that could be nearby the two previously described putative PKS genes of *P. chrysogenum*. First of all, one putative fungal flavin-binding oxidoreductase that could be part of the PKS gene cluster was discovered in this genomic region (Table 15). Besides, one putative fungal Barren family protein with probable function in mitosis, as well as homology to two putative enzyme functions involved in fungal carbohydrate metabolic processes, were detected within the course of BLAST analysis. Putative functions of other non-PKS genes recognised within the large AT-9-11 genomic region of *P. chrysogenum* (i.e. more than 50 kb) are still ambiguous, as for the F1-T3 sequene that showed similarity to both, the ATPase of *Neosartorya fischeri* and the VeA regulatory protein of *A. parasiticus* (Table 15).

In order to identify the complete AT-9-11 biosynthesis gene cluster, as well as the KHKS1 gene cluster, further sequence data are required that may be provided via shot-gun sequencing (see 12.18.2) of fosmid inserts containing a whole cluster in question.

As it was shown via PCR amplification of the specific KS-KHKS32 sequence from AT-9-11 fosmids (7.3.3) the KHKS32 and the AT-9-11 genomic regions harbour the same PKS biosynthesis gene cluster (Figure 30). This simplified the decision on the choice of PKS gene cluster for the shot-gun-sequencing, by reducing options from three to only two.

The decision was made taking into account the chemical structure of sorbicillactone A (**50**) which indicated presence of one R PKS and one NR PKS (with involvement of RED domain), similarly as postulated for citrinin (**2**) and zearalenone (**9**) [50]. The structures of the compounds indicate high levels of reduction during early polyketide biosynthesis and no reductions during later steps. In fact, the chemical structure of sorbicillactone A (**50**) implicated the involvement of  $\beta$ -keto reductive PKS domains only in the early steps of its biosynthesis (i.e. sorbyl side chain). Beside, C-methylation that take place during later non-reducing steps of polyketide chain formation indicated the necessity of one intrinsic MT domain. Moreover, the presence of one terminal RED domain responsible for reductive release of the polyketide intermediate to the aldehyde level [170] has been anticipated for the sorbicillactone A biosynthesis as well (see 3). This aldehyde could play a role in the ring closure (either spontaneous or enzymatically catalyzed) after PKS product release.

In the light of these postulations and according to our findings that one putative reducing PKS (homology over ER domain) and one non-reducing PKS (homology over SAT, MT and RED required domains) are encoded next to each other, our choice lead sequencing of AT-9-11 putative PKS gene cluster. Moreover, the fact that during subcloning and sequencing of KHKS1 genomic region no similarity was found to any  $\beta$ -keto reducing domain, neither to MT and/or RED domains supported this decision (Table 12 and Table 13).

## **8 DNA sequence analysis of the putative sorbicillactone gene cluster**

In order to construct a random shotgun-library for sequencing the putative sorbicillactone gene cluster from the AT-9-11 genomic region of *P. chrysogenum* E01-10/3, lengths of fosmid inserts carried by the 42H12 and 46H11 clones, supposed to contain the core of this cluster, were calculated from their *Bam*HI restriction profiles (see Figure 29). The fosmid insert of 42H12 was estimated to be ~37.7 kb, while the fosmid insert of 46H11 was presumed to be ~34 kb in size. Since it appeared that the fosmid DNA of clone 42H12



contains a larger portion of the putative gene cluster, this clone was chosen for whole fosmid shotgun sequencing.

### **8.1 Shotgun sequencing of the putative sorbicillacone gene cluster**

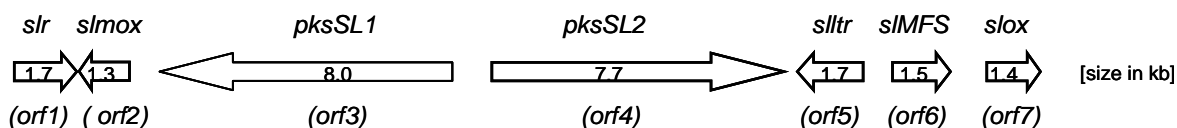
Shotgun sequencing and generation of continuous DNA sequence from 42H12 insert was performed by GATC Biotech (Konstanz, Germany). For this purpose, the fosmid DNA of 42H12 was dissolved in sterile purified water and sonicated under suitable conditions in order to generate most of DNA fragments ranging between 1-3 kb (see 12.18.2). The size of sheared DNA fragments was monitored by gel electrophoresis. The cloning procedure of the shotgun fragments was performed with the pCR4 blunt Topo<sup>®</sup> vector system. For the sequencing reactions, the primers M13-RP and M13-FP were used (Table 27). The 42H12 shotgun library in size of one 384-well plate was sequenced forwards and reverses. After evaluation of initial 768 DNA sequences for quality and eventual presence of vector sequence, 625 trimmed sequences were left. With those DNA sequences an amount of 50.3520 bases in total was generated that represented 10.4 coverage for the estimated size of cloned *P. chrysogenum* genomic insert (40-50 kb). The GATC used complex SeqMan<sup>®</sup> software (Lasergene<sup>™</sup>) to compile these 625 trimmed sequences into one single DNA contig of 45.93 kb. This continuous DNA sequence, namely contig-4, was subsequently analysed for the presence of putative fungal open reading frames (ORFs).

### **8.2 Identification of ORFs and prediction of enzyme functions**

The analysis of putative ORFs within contig-4 (i.e. the GATC generated contig) was managed via a combination of two online programs: “Fgenesh” [199], software for hidden Markov model (HMM) based gene structure prediction in eukaryota and “FramePlot 2.3.2” [241], a web based tool for analysing protein coding regions in bacterial DNA with a high GC content, i.e. similar as for fungal genomes (see Table 31). However, the subsequent analysis of PKS active sites (see 8.5.1) indicated that the putative NR PKS gene from contig-4 misses the AT-active-site motif (Figure 33) that implies inactivity of such PKS. Thus, all 768 original DNA sequences were checked again for the presence of cloning vector traces and after trimming used for the creation of new contigs via the SeqMan<sup>®</sup> program (Lasergene<sup>™</sup>; see Table 31) under different program settings. The longest generated continuous DNA sequence, named as contig-7, comprised of 48.83 kb in total and included the intact AT site (Figure 33) that qualified this contig for further ORF analysis. The Fgenesh was able to detect 10 putative ORF with 31 predicted exons in both orientations of 46.91 kb coding sequence (Figure 43).

The protein sequences correspond from each ORF were generated by Fgenesh program and were subsequently analysed via BLAST search for the presence of putative homologues in protein databases. The protein-to-protein sequence searches were performed using original member of the BLAST suite of programs, known as “BLASTP”. The results of this analysis are shown in Table 16. ORFs with the same deduced protein homologies were also inferred by FramePlot 2.3.2, (data not shown). As shown in Table 16, the deduced amino acid sequences of eight ORFs showed similarity to fungal sequences in gene databases.

Among the identified ORFs there were two putative PKS genes: *orf3* that encodes for the putative fungal NR PKS, and *orf4* that implies to be a reducing fungal PKS. As illustrated in Figure 31, these two PKS genes are of opposite orientation. The highest similarity of *orf3* is to the citrinin PKS (CitS/PKSct; GenBank: Q65Z23; 42% identity) [33] responsible for biosynthesis of citrinin (**2**), i.e. the polyketide mycotoxin of *M. purpureus*. The closest characterised homologue of the *orf4*-derived protein was the compactin diketide synthase (CDKS or MlcB; GenBank: Q8JOF5; 41% identity) [92] that is responsible for synthesis of 2-methylbutyryl side chain during compactin biosynthesis pathway.



**Figure 31: Proposed gene organization of putative sorbicillactone gene cluster.**

Arrows represent relative positions and orientations of seven identified ORFs postulated to be part of putative sorbicillactone gene cluster. Additionally, each arrow contains an approximate size (in kb) for the ORF in question. Below arrows names of ORFs are given. Above arrows are given working names of identified putative genes (ORF) with hints to their putative functions. This chart contains the following genes according to their appearance in predicted sorbicillactone gene cluster (orf1-orf7): putative transcriptional regulatory gene (*slr*), gene for putative monooxygenase (*slmox*), gene for putative sorbicillactone PKS1 (PKSSL1), gene for putative sorbicillactone PKS2 (PKSSL2), gene for putative transcriptional regulatory gene (*sltr*), gene for putative MFS transporter protein (*slMFS*) and gene for putative oxidoreductase (*slox*).

## Results and Discussion

**Table 16: FgenesH inferred ORFs within the shotgun sequenced fosmid 42H12.**

<sup>(a)</sup> The predicted amino acid sequences for the ORFs were deduced following removal of the putative intron regions <sup>(b)</sup> (+) for ORFs identified in direct DNA strand and (-) for ORF identified in complementary DNA strand. <sup>(c)</sup> The function of each encoded protein was deduced with its putative functional domains and from the predicted function of the first characterised protein sequence or protein sequence showing the highest degree of similarity if no characterised gene was available. <sup>(d)</sup> Identity/similarity (%) of the deduced amino acid sequences. <sup>(e)</sup>Reference combines sequence accession number in the GenBank database, referent author and the year of first deposition of a sequence in TrEMBL database.

ORF	Gene name	Size (bp/aa) <sup>(a)</sup>	Orientation (+/-) <sup>(b)</sup>	Plausable function <sup>(c)</sup>	Protein homology <sup>(c)</sup>	Identity/ Similarity (%) <sup>(e)</sup>	Reference <sup>(d)</sup>
<i>orf1</i>	<i>slr</i>	1662/553	+	Fungal specific transcriptional regulator	CtnR, citrinin biosynthesis transcriptional activator, <i>M. purpureus</i>	32/50	Q1ERI1; Shimizu <i>et al.</i> , 2007
<i>orf2</i>	<i>slmox</i>	1338/445	-	Monooxygenase (hydroxylase)	SalA, salicylate 1-monooxygenase, <i>A. nidulans</i>	34/52	Q9HFQ8; Graminha <i>et al.</i> , 2004
<i>orf3</i>	<i>PKSSL1</i>	7995/2664	-	Non reducing class polyketide synthase	CitS, citrinin polyketide synthase, <i>M. purpureus</i>	42/61	Q65Z23; Shimizu <i>et al.</i> , 2005
<i>orf4</i>	<i>PKSSL2</i>	7746/2581	+	Reducing class polyketide synthase	MlcB, compactin diketide synthase, <i>Penicillium citrinum</i>	41/59	Q8JOF5; Abe <i>et al.</i> , 2002
<i>orf5</i>	<i>sltr</i>	1743/580	-	Fungal specific transcriptional factor	Putative transcriptional factor, <i>Neosartorya fischeri</i>	56/67	A1DJQ4; Fedorova <i>et al.</i> , 2008
<i>orf6</i>	<i>slMFS</i>	1539/512	+	Transporter	Putative Major Facilitator Superfamily (MFS) transporter, <i>A. nidulans</i>	75/81	Q5B860; Galagan <i>et al.</i> , 2005
<i>orf7</i>	<i>slox</i>	1416/471	+	FAD dependant oxidoreductase (oxidase)	Putative FAD dependant oxidoreductase, <i>A. nidulans</i>	63/73	Q5B862; Galagan <i>et al.</i> , 2005
<i>orf8</i>	/	4176/1391	-	Component of the PAN1 actin cytoskeleton-regulatory complex	Putative actin cytoskeleton-regulatory complex protein, <i>A. terreus</i>	55/63	Q0CPW4; Birren <i>et al.</i> , 2005

Along with these PKS genes, two genes for putative fungal flavin-dependent oxidoreductases were detected on the cluster. *orf2* displays similarity to *sala* of *A. nidulans* (GenBank: Q9HFQ8; 34% identity) which is deemed to be involved in the resistance development towards terbinafine. This is probably managed through degradative hydroxylation of the naphthalene nucleus of this aromatic compound, as mentioned before [238]. Consequently, FAD-dependent oxidoreductase domain was detected within *orf2*. This could mean that the putative flavin monooxygenase encoded by *orf2* may perform a hydroxylation reaction in the sorbicillactone A and sorbicillactone B biosynthesis.

The second oxidoreductase was inferred from the deduced amino acid sequence of *orf7*, and shared the highest similarity with the putative FAD-dependent oxidoreductase of *A. nidulans* (GenBank: Q5B862; 63% identity). However, no characterised oxidoreductase gene was among the matches for *orf7* searched via BLASTP. Based on these results of the “BLASTP” analysis it is likely that *orf7* encodes an FAD-dependent oxidoreductase. The precise role of such an oxidoreductase during sorbicillactone biosynthesis has to be further investigated.

Additionally, two fungal specific transcriptional regulators were identified (Table 16), each nearby one PKS gene (see Figure 31). The BLASTP analysis showed that both ORFs contained the full length (180 aa) fungal specific transcriptional regulator domain. The transcriptional regulator encoded by *orf1* is placed next to the putative NR PKS gene *orf3*, and shares 32% identity with citrinin biosynthesis transcriptional activator [34] of *M. purpureus* (CtnR, GenBank: Q1ERI1) that regulates biosynthesis of citrinin. The second transcriptional regulatory gene (*orf5*) is next to the putative reducing class PKS (*orf4*) and has significant similarity to the putative transcriptional regulator of *Neosartorya fischeri* (GenBank: A1DJQ4; 56% identity). The relative positions of mentioned ORF are shown in Figure 31.

Further on, one of the identified ORFs, namely *orf6*, exhibits 75% identity to the putative Major Facilitator Superfamily (MFS) transporter protein of *A. nidulans* (GenBank: Q5B860). The MFS transporters are single-polypeptide secondary carriers capable only of transporting small solutes in response to chemiosmotic ion gradients. It is interesting that among putative homologues of *orf6* was also the putative citrinin biosynthesis transporter [34] of *M. purpureus* (GenBank: Q1ERH8; 36% identity) that was recently identified adjacent to the *pksCT* gene.

The last identified ORF that showed similarity to fungal genes was *orf8* that shared significant similarity with the putative actin cytoskeleton-regulatory complex from *A. terreus* (GenBank: Q0CPW4; 55% identity). BLASTP found several domains within *orf8*: precisely two EPS15 homology (EH) domains and one calcium-binding domain of the EF-hand type within each of the EH domains. Besides, one Wiskott-Aldrich homology 2 (WH2) domain was identified as well. The EH protein-protein interaction module was found in several proteins involved in endocytosis, vesicle transport and signal transduction in organisms ranging from yeast to mammals. This domain is often implicated in the regulation of protein transport/sorting and membrane-trafficking. The WH2 actin-binding motif (~18 aa) was found as a modular part of larger proteins. It binds actin monomers and can facilitate the assembly of actin monomers into newly forming actin filaments. The WH2 domains occur in eukaryotes from yeast to mammals, in insect viruses, and in some bacteria.

### 8.2.1 Summary of the ORF prediction

Based upon the highest similarities that the deduced products of the identified ORFs displayed via BLASTP database search, putative functions were postulated along with the gene organization of identified cluster (see Figure 31). It was presumed that seven ORFs (*orf1-orf7*, see Table 16) could be part of the same gene cluster. *orf8* was excluded from the cluster model since its involvement in cytoskeleton dynamics does not suggest the biosynthesis or transport of sorbicillactones .

As it is shown in the Figure 31, one gene for the putative transcriptional regulator (*slr* or *orf1*), that could coordinate expression of the structural genes in the cluster, is followed by the gene encoding for one putative monooxygenase (*slmox* or *orf2*) that could be responsible for one post-PKS hydroxylation reaction during sorbicillactone A/B biosynthesis. The core of the identified putative gene cluster contains two PKS genes (*PKSSL1/orf3* and *PKSSL2/orf4*) located next to each other in opposite orientations. The second putative transcriptional regulatory gene (*sltr* or *orf5*) appears as a fifth ORF in a row and is located before putative gene for one MFS transporter protein (*slMFS* or *orf6*). As mentioned before (see 2.4.5), transporter genes are usually associated with gene clusters responsible for biosynthesis of active secondary metabolites and play a role in excretion of potentially harmful intracellular metabolites to a surrounding environment as form of self-defense mechanism. The product of the last identified putative ORF (*slox* or *orf7*) may be an oxidoreductase.

An additional confidence to putative PKS functions detected within analyzed gene cluster and a precise identification of the domain architecture were made through the phylogenetic

analysis (see 8.4) and domain active sites analysis (see 8.5) of the deduced protein sequences from the two presented PKS genes (*PKSSL1* and *PKSSL2*, see Figure 31). Taking into account the chemical structure of sorbicillactone A (**50**) and its proposed biosynthetic route (see 3) more protein functions need to be considered for this compound to be made. The Introduction of amino acid alanine into the molecule via esterification with the hydroxyl group of sorbicillinol-intermediate could be performed by the action of one pyridoxal phosphate (PLP)-dependent aminotransferase. Besides, the *N*-acylation step after lactone ring closure requires the presence of one gene encoding for a post-PKS acyl-transferase activity. In order to search for the presence of these additional protein functions within the assumed sorbicillactone cluster, further subcloning was made from AT-9-11 genomic region of *P. chrysogenum* (see 8.3).

### 8.3 Subcloning of fragment-2 from 49C8 fosmid clone

The shotgun-sequencing of the fosmid clone 42H12 that was presented in the previous subsection (see 8.2) revealed the presence of seven putative ORFs as candidates for genes involved in the sorbicillactone A biosynthesis. In order to identify additional putative protein functions being part of the cluster, larger fragments generated by *Bam*HI profiling of AT-9-11 genomic-region (see 7.3.2) were considered for further subcloning. The detailed sequence analysis of F1 (~15.0 kb) and F2 (~12.5 kb) of the AT-9-11 genomic region (Figure 29) in alignment with the contig-7 DNA sequence (BioEdit program, alignment not shown) pointed out that the reverse complement of F2-T3 sequence shares one *Bam*HI restriction site, as well as surrounding nucleotide sequence, with contig-7. Consequently, these two sequences also share homology to a putative flavin monooxygenase protein function as previously mentioned (see for F2-T3 in Table 15 and for *orf2* in Table 16). This indicates that F2 could contain additional genes of the identified cluster, upstream of the transcriptional regulatory gene (*orf1*, Figure 31). Since the last identified ORF (*orf8*, Figure 31) from sequencing of 42H12-fosmid was a gene whose putative product is probably involved in cell cytoskeleton dynamics, it was regarded to be pivotal to focus on the upstream region of the putative sorbicillactone cluster.

F2 was subsequently restricted with several endonucleases. The *Pst*I restriction that released six fragments (see Table 18: F2a-F2f) was chosen for further subcloning of F2 subfragments in pBluescript<sup>®</sup> II. The cloned subfragments were end sequenced with T7 and T3 primers (Table 27). The closest homologues for all retrieved end-sequences are shown in Table 17. Several putative non-PKS protein functions were detected via BLASTX search. BLAST

search results for subcloned F2 fragments (i.e. F2a-F2f), as well as other putative non-PKS protein functions identified from analyzed AT-9-11 region, are all presented in Table 18. Therefore, additional identified non-PKS protein functions (apart from ones already described in 8.2) will be discussed in the following sections.

One of the putative non-PKS protein functions detected in a relative neighbourhood of two detected PKS genes was the Barren-family-protein-function assigned to the *A. terreus* protein (GenBank: Q0CPW2, 72% identity). As already mentioned before, this putative function was found in large portions of F1-T7 and F6a-T3 sequences (see 7.3.2). This family consists of several Barren-protein homologues from eukaryotic organisms. In *Drosophila*, The Barren (encoded by *barr* gene) is required for sister-chromatid segregation in mitosis and has homology to yeast and humans. The Barren-protein is localised in chromatin throughout mitosis [242]. Thus, there is a relatively high probability that one protein with a role in chromosomal segregation during mitosis or chromatin-condensation is encoded on the genomic fragment of *P. chrysogenum* that was analyzed within this study. Further on, the F1-T3 sequence exerted homology to two putative functions: a putative ATPase of *Neosartorya fischeri* (GenBank: A1DC56, 47% identity) and VeA protein of *A. parasiticus* (GenBank: Q69B22, 30% identity). Both homologies are displayed over a relatively short homology stretch as shown in Table 18. A putative ATPase function indicates probability that one protein involved in transport of protons across a membrane could be in the proximity of two PKS genes. On the other hand, it is interesting that VeA, the second putative homologue, is being produced exclusively by fungal species. This global regulator is conserved in numerous fungal species although it was not strictly found in yeast organisms such as *S. cerevisiae* or *Schizosaccharomyces pombe* [243]. Besides the role of VeA in regulation of morphogenesis, VeA also functions as a key global metabolic regulator in the biosynthesis of secondary metabolites. Among them are carcinogenic mycotoxins like aflatoxin and sterigmatocistin (*A. nidulans* [244], *A. parasiticus* [240] and *A. flavis* [245]), and antibiotics as penicillin (*A. nidulans* [244]) and cephalosporin C (*Acremonium chrysogenum* [246]). Within the herein cited studies it was found that the expression of genes involved in the synthesis of secondary metabolites commonly found in clusters was affected by VeA. The VeA N-terminal region contains two putative nuclear localization signal (NLS) motifs: predicted pat7 motif and bipartite NLS motif [243]. However, F1-T3 sequence shared homology with less conserved C-terminal part of protein that did not carry these NLS-motifs. No VeA homologue was reported from *P. Chrysogenum* in public data bases. It is certain that more sequence data

from F1 is necessary in order to estimate whether such putative fungal protein function is present.

The first putative protein function discovered via subcloning of F2 was an ion exchanger protein function. The deduced amino acid sequences of F2b-T7, F2f-T7 and F2f-T3 shared high similarity to putative integral membrane sodium/calcium exchanger protein of *A. niger* (GenBank: A2R181; 48-84% identity, see Table 18). Probably the significantly lower quality of the F2b-T7 end-sequence, compared to the sequenced F2f, decreased the otherwise high homology to this putative ion exchanger. Nevertheless, together these three sequences covered approximately 600 aa homology stretch within the match protein and supported the likelihood for the presence of a Ca<sup>2+</sup> signalling protein about 15-16 kb away from the flavin monooxygenase of the putative sorbicillactone gene cluster. In filamentous fungi, in which growth patterns and development are complex, there is evidence for the involvement of Ca<sup>2+</sup> in many physiological processes, including the cell cycle, sporulation, spore germination, hyphal tip growth, hyphal orientation, hyphal branching, and circadian rhythms [247].

Further, the subcloned F2d-T7 derived amino acid sequence shared modest homology to two different putative bacterial protein functions. Although the quality of this ~0.9 kb DNA sequence was sufficiently good, surprisingly no fungal protein matches were found via BLASTX search. Among the discovered bacterial putative homologues, the highest similarity, F2d-T7 shared with putative RNA methyltransferase of *Desulfotalea psychrophila* (GenBank: Q6AQE8; 45% identity over ~30 aa). This is an enzyme that is potentially involved in the modification of nucleotides during ribosomal RNA maturation via transfer of methyl group from the ubiquitous SAM (*S*-adenosyl methionine). Besides, the same conceptually translated sequence (F2d-T7) shows homology to a putative radical SAM superfamily domain protein of *Staphylococcus epidermidis* (GenBank: Q5HN57; 27% identity over ~60 aa). The radical SAM proteins catalyze diverse reactions, including unusual methylations, isomerizations, sulphur insertion, ring formation. They function in DNA precursor, vitamin, cofactor, antibiotic and herbicide biosynthesis and in biodegradation pathways [248]. Due to their wide repertoire of functions these enzymes are not easily recognizable via bioinformatic tools for sequence inspection.

On the other hand, the T3 sequenced end of F2d showed relatively low similarity to even two putative bacterial proteins. The first deduced function exerts similarity toward the putative TonB-like protein of *Xanthomonas oryzae pv. oryzae* (GenBank: Q2P6I6; 24% identity over ~70 aa within putative peptidase M56 domain). This is a putative cell membrane protein involved in iron ion and protein transport. Beside that, the presence of the peptidase domain



may imply the involvement in cell signalling pathways through site-specific proteolytic cleavage of a target protein [249]. The second deduced homology for F2d-T3 sequence matches the protein sequence annotated as non-ribosomal peptide synthase modules and related proteins from *Hahella chejuensis* (GenBank: Q2SHZ4; 27% identity over ~100 aa within the first condensation domain). One of the largest and most important groups of microbial secondary metabolites comprises peptides that are synthesised by enzymes without ribosome function. These enzymes are known as non-ribosomal peptide synthetases (NRPSs). Similarly like PKSs, NRPSs are large, multifunctional enzymes typically comprised of numerous semiautonomous catalytic domains in a linear series [61]. There are also known complexes that include both enzyme systems, so called NRPS/PKSs. Although NRPSs from fungi share many important characteristics with bacterial NRPSs, they also differ from their bacterial counterparts in some key traits. Thus, the low similarity, detected over a small homology stretch of condensation domain, implies that the identified match might not be a homologue (proteins are either homologous or not homologue).

As it was expected on the basis of results presented for subcloning of the whole AT-9-11 cluster candidate (Table 15), one part of F2 shares homology with the putative Barren family of proteins. Further subcloning of this fragment (Table 17) pointed out that the F2e-T7-derived protein sequence is similar to a putative protein from *A. clavatus* (GenBank: A1CD76; 58% identity over ~250 aa), annotated as condensing complex component Cnd2. This protein carries the same predicted Barren domain as in the previously described cases of F1-T7 and F6a-T3 sequences and has the same putative function as it was emphasised before for the protein of *A. terreus* (GenBank: Q0CPW2). Onwards, one side of T3-sequenced F2e subfragment shares homology to the same protein of *A. clavatus* over a significantly shorter homology stretch (GenBank: A1CD76; 83% identity ~71 aa).

However, another side of derived F2e-T3 sequence is not as uniform concerning demonstrated homology. First homology referred to putative DNA-binding-protein of *A. nidulans* (GenBank: Q5B5S4; 24% identity over ~150 aa), while the second putative homologue is the predicted PLP-dependent aminotransferase from the bacterium *Methanococcoides burtonii* (GenBank: Q12VN3; 36% identity over ~70 aa). As illustrated in Table 18, both homologies span relatively short parts of their protein matches. In addition, it is not easy to directly infer a concrete function for such putative DNA binding protein since the Myb domain that has been predicted for this protein may be found in a large number of proteins and could be involved in different stages of chromatin dynamics. Maddox and collaborators [250], that studied Myb domain containing proteins, underlined that it was not

straightforward to make a clear link between such proteins from different organisms during their initial bioinformatic analysis. Concerning the capability of bioinformatic tools to identify PLP-dependant enzymes, the situation is not less demanding, taking into account that PLP (pyridoxal phosphate, a vitamin B<sub>6</sub> derivate) represents one of the most versatile organic cofactors in biology, and is used by a variety of enzymes in all organisms [251] having more than 140 distinct enzymatic activities which are catalogued by the Enzyme Commission (EC; <http://www.chem.qmul.ac.uk/iubmb/enzyme/>). Almost all PLP-dependent enzymes are associated with biochemical pathways that involve amino compounds, mainly amino acids. Despite their functional variety, all structurally characterised PLP-dependent enzymes belong to just five distinct structural groups. Although the limited structural diversity facilitates the identification of PLP-dependent enzymes from genomic sequences, there are at least two limitations that are inherent to homology-search for PLP enzymes. Structurally similar enzymes may escape detection if their sequence-similarity has become negligible [252]. Furthermore, homology searches fail to identify PLP-dependent genes if encoding enzymes do not fall into the five fold-type categories [253, 254]. Generally, such limitations are not exclusive for PLP-dependent enzymes and can be extrapolated into other classes of enzymes as well.

In order to reveal which match is a true homologue, the use of the PSI-BLAST (position specific iterated BLAST) was valuable. PSI-BLAST provides with a tool for detecting distant relationships between proteins, or in other words – it enables searching for distantly related homologues that are in danger to be missed if a search would be solely based on similarity between compared sequences (like it is for “BLASTX” and “BLASTP”). The PSI-BLAST tool utilises a profile called a Position Specific Score Matrix (PSSM) as a model of sequence alignment and provides with values on the amino acids at certain positions. The iterative nature of the procedure further improves sensitivity by incorporation of increasingly distant functional homologues in the profile resulting in increasingly wider searches among protein sequences.

After four PSI-BLAST iterations for F2e-T3 protein sequence, the most similar putative homologues were annotated as condensing complex component Cnd2 proteins. However, no protein similar to a PLP-dependent enzyme produced alignments with E-value better than the preset threshold (data not shown).

Among additional non-PKS homologies that were detected during subcloning of a putative AT-9-11-cluster is a homologue of putative RNA polymerase of *A. terreus* (GenBank: Q0CPW4; 66% identity over ~150 aa). This homologue was identified from the F3-T3-

deduced amino acid sequence. Besides, the F6a-T3-derived sequence displayed homology to one putative tRNA modification enzyme of *Botryotinia fuckeliana* (GenBank: A6SM10; 26% identity over 100 aa) that may be involved in biosynthesis, transport and/or utilisation of methionine as it has been annotated. The last two non-PKS putative protein functions are referred to enzymes of carbohydrate metabolism as it was already discussed in 7.3.2. A putative glycosyltransferase function was inferred based on similarity that the F6b-T7 sequence shares with a putative protein of this class from *Magnaporthe grisea* (GenBank: A4RLP8; over 128 aa homology stretch). The homology to a putative glycosyl hydrolase from *A. terreus* (GenBank: Q0CXJ2; over 340 aa homology stretch), an enzymatic function that hydrolyses *O*-glycosyl compounds, was identified for the deduced amino acid sequence of F6b-T3.

## Results and Discussion

**Table 17: Results of subcloning and subsequent sequencing of fragment 2 from the AT-9-11 cluster**

Sequence name <sup>(a)</sup>	Fragment size <sup>(b)</sup>	Sequence homologues <sup>(c)</sup>	Identities (%) <sup>(d)</sup>	Positives (%) <sup>(d)</sup>	References <sup>(e)</sup>
Fragment F2a (T7)	3.3 kb	• Putative flavoprotein monooxygenase; <i>A. terreus</i>	40	58	Q0CCB8; Birren <i>et al.</i> 2005
Fragment F2a (T3)		• Putative flavoprotein monooxygenase; <i>Coccidioides immitis</i>	43	61	Q1DRI4; Birren <i>et al.</i> 2005
Fragment F2b (T7)	2.8 kb	• Putative sodium/calcium exchanger protein; <i>A. niger</i>	48	57	A2R181; Pel <i>et al.</i> 2007
Fragment F2b (T3)		• Predicted protein (no function proposed); <i>Botryotinia fuckeliana</i>	66	83	A6RVH5; Birren <i>et al.</i> 2005
Fragment F2c (T7)	2.6 kb	• Putative fungal specific transcription factor; <i>Chaetomium globosum</i>	74	88	Q2GR15; Birren <i>et al.</i> 2005
Fragment F2c (T3)		• Putative fungal specific transcription factor; <i>A. terreus</i>	61	78	Q0CF68; Birren <i>et al.</i> 2005
Fragment F2d (T7)	1.4 kb	• Putative RNA methyltransferase (Mtase); <i>Desulfotalea psychrophila</i>	45	67	Q6AQE8; Rabus <i>et al.</i> 2004
		• Radical SAM superfamily domain protein; <i>Staphylococcus epidermidis</i>	27	56	Q5HN57; Gill <i>et al.</i> 2005
Fragment F2d (T3)		• TonB-like protein; <i>Xanthomonas oryzae pv. oryzae</i> (bacterium)	24	44	Q2P616; Ochiai <i>et al.</i> 2005
		• Non-ribosomal peptide synthetase (condensation domain); <i>Hahella chejuensis</i> (bacterium)	27	44	Q2SHZ4; Jeong <i>et al.</i> 200
Fragment F2e (T7)	1.3 kb	• Condensin complex component cnd2 (Barren protein family, mitosis); <i>A. clavatus</i>	58	6	A1CD76; Nierman 2006
Fragment F2e (T3)		• Condensin complex component cnd2 (Barren protein family, mitosis); <i>A. clavatus</i>	83	87	A1CD76; Nierman 2006
		• Putative DNA-binding protein; <i>A. nidulans</i>	24	36	Q5B5S4; Galagan <i>et al.</i> 2005
		• Putative DegT/DnrJ/EryC1/StrS PLP dependant aminotransferase; <i>Methanococcoides burtonii</i>	36	46	Q12VN3; Copeland <i>et al.</i> 2006
Fragment F2f (T7)	1.1 kb	• Putative sodium/calcium exchanger protein; <i>A. niger</i>	84	93	A2R181; Pel <i>et al.</i> 2007
Fragment F2f (T3)		• Putative sodium/calcium exchanger protein; <i>A. niger</i>	82	89	A2R181; Pel <i>et al.</i> 2007

(a) Fragment name; T7 or T3 in brackets are standing for primer used for sequencing. (b) Fragment size estimated by comparison with referent fragment sizes via agarose gel electrophoresis (c) Putative homologues retrieved via BLASTX search showing the first match for the sequence similarity search. In addition, the second match, if given, represents the closest fungal protein that is available if the first sequene match is not a fungal protein, the second match that differs in predicted function from a first match in BLAST search, or a first characterised putative gene homologue available in GenBank database (d) Identity/similarity (%) of the deduced amino acid sequences. (e) Reference combines sequence accession number in TrEMBL database, referent author and the year for first deposition of a sequence in GenBank.

## Results and Discussion

**Table 18: Inferred non-PKS gene homologies retrieved via subcloning of putative AT-9-11 gene cluster.**

Data in this table contain information on putative homologous protein functions and their organisms of origin. In addition, protein cellular localization and putative protein domains/motifs are given if stated in reviewed literature. For each protein function one or more query DNA sequences (fragments from AT-9-11 genomic region), for which the homology search was made, are also presented in the table. Besides, the lengths of amino acid stretch for which a certain level of homology [identity (I)/similarity (P)] was identified are also listed. The last column of this table contains reference to publications or internet databases that served as source of information for this analysis.

<b>Protein(s)</b>	<b>Organism / Cellular localization</b>	<b>Predicted molecular function</b>	<b>Protein domains / motifs</b>	<b>Query sequences<sup>(a)</sup></b>	<b>Homology stretch<sup>(b)</sup></b>	<b>I, P (%)<sup>(c)</sup></b>	<b>Reference/ Source<sup>(d)</sup></b>
Putative Barren family protein	• <i>A. terreus</i> (Q0CPW2) / Nucleus	• Chromosomal segregation during mitosis	• Barren domain	F1 (T7) F6a (T3)	~ 340 aa out of 874 aa ~ 340 aa out of 874 aa	72, 81 71, 81	• Pfam* • Bhat et al., 1996
Putative ATPase	• <i>Neosartorya fischeri</i> (A1DC56) / Cellular membrane	• ATP synthesis and/or hydrolysis • Transport of protons across a membrane.	• alpha and beta subunits	F1 (T3)	~ 140 aa out of 308 aa (at the start of protein)	47, 62	• InterPro**
VeA (velvet) global regulator	• <i>A. parasiticus</i> (Q69B22) / Cytoplasm or nucleus	• Morphological development • Mycotoxin biosynthesis	Two putative nuclear localization signals (NLS): • bipartite NLS • pat7 motif	F1 (T3)	~ 100 aa out of 574 aa (toward end of the protein)	30, 41	• InterPro • Stinnett et al. 2007 • Calvo et al., 2008
Putative flavoprotein monooxygenase	• <i>A. terreus</i> (Q0CCB8) • <i>Coccidioides immitis</i> (Q1DR14) • <i>Chaetomium globosum</i> (Q2GZ40) / Substrate dependant	• Oxygenation reaction: incorporation of hydroxyl group into substrate	• FAD binding domain • Aromatic-ring hydroxylases domain	F2a (T7) F2a (T3) F14 (T7)	~ 280 aa out of 466 aa (toward end of the protein) ~ 190 aa out of 482 aa ~ 60 aa out of 498 aa	40, 58 43, 61 62, 79	• InterPro
Putative sodium/calcium exchanger protein	• <i>A. niger</i> (A2R181) / Cell membrane	• Regulation of intracellular Ca <sup>2+</sup> concentration	• Integral membrane regions	F2b (T7) F2f (T7) F2f (T3)	~ 600 aa overall out of 994 aa	48, 57 84, 93 82, 89	• Pfam • Zelter et al., 2004

## Results and Discussion

<b>Protein(s)</b>	<b>Organism / Cellular localization</b>	<b>Predicted molecular function</b>	<b>Protein domains / motifs</b>	<b>Query sequences<sup>(a)</sup></b>	<b>Homology stretch<sup>(b)</sup></b>	<b>I, P (%)<sup>(c)</sup></b>	<b>Reference/ Source<sup>(d)</sup></b>
Putative RNA methyltransferase	• <i>Desulfotalea psychrophila</i> (Q6AQE8) / Nucleolus	• Modification of nucleotides during ribosomal RNA maturation in a site-specific manner	Not specified	F2d (T7)	~ 30 aa out of 158 aa (in the middle of the protein)	45, 67	• Pfam
Radical SAM superfamily domain protein	• <i>Staphylococcus epidermidis</i> (Q5HN57) / Substrate-dependent	Catalyze diverse reactions: • unusual methylations • isomerization • ring formation, etc.	• Radical SAM domain • YfkB-like domain, potentially interacts with radical SAM domain	F2d (T7)	~ 60 aa out of 380 aa (within radical SAM domain)	27, 56	• Pfam • Sofia et al., 2001
TonB-like protein	• <i>Xanthomonas oryzae pv. oryzae</i> (Q2P6I6) / Cell membrane	• Iron ion and protein transport • Proteolytic cleavage	• Peptidase M56 domain • TonB energy-transducer domain	F2d (T3)	~ 70 aa , out of 438 aa within peptidase M56 domain	24, 44	• InterPro • Pfam • Zhang et al., 2001
Non-ribosomal peptide synthetase modules and related protein	• <i>Hahella chejuensis</i> (Q2SHZ4) / Subcellular: organelle-like membrane-associated complex	• Generates polypeptides sans ribosome	• two AMP binding domains • three condensation domains • two PP domains	F2d (T3)	~ 100 aa within the first condensation domain, out of 2426 aa	27, 44	• Pfam
Condensin complex component cnd2	• <i>A. clavatus</i> (A1CD76) / Nucleus	• Chromosomal segregation during mitosis	• Barren domain	F2e (T7) F2e (T3)	~ 250 aa out of 881 aa ~ 70 aa out of 881 aa	58, 63	• Pfam
Putative DNA-binding protein	• <i>A. nidulans</i> (Q5B5S4) / Nucleus	• DNA binding	• Myb-like DNA-binding region	F2e (T3)	~ 150 aa out of 427 aa	24, 36	• InterPro

## Results and Discussion

<b>Protein(s)</b>	<b>Organism / Cellular localization</b>	<b>Predicted molecular function</b>	<b>Protein domains / motifs</b>	<b>Query sequences<sup>(a)</sup></b>	<b>Homology stretch<sup>(b)</sup></b>	<b>I, P (%)<sup>(c)</sup></b>	<b>Reference/ Source<sup>(d)</sup></b>
Putative DegT/DnrJ/EryC1/S trS PLP-dependant aminotransferase	• <i>Methanococcoides burtonii</i> (Q12VN3) / Substrate-dependent	• Members of this family have variety of molecular functions	• Domain referring to the DegT/DnrJ/EryC1/Str S PLP-dependant aminotransferase family	F2e (T3)	~ 70 aa out of 370 aa	36,46	• Pfam
Putative RNA polymerase	• <i>A. terreus</i> (Q0CPW4) / Nucleus	• DNA- directed RNA synthesis • Calcium ion binding	• RNA polymerase Rpb3/Rpb11 dimerisation domain • Domain of unknown function (DUF1720); often engaged with EF hand domain and thus likely involved in cytoskeletal processes • EF hand domain (calcium ion binding) • WH2 actin-binding motif	F3 (T3)	~ 150 aa at the end of protein containing WH2 actin-binding motif	66, 70	• Pfam
Putative tRNA modification enzyme	• <i>Botryotinia fuckeliana</i> (A6SM10) / Cytoplasm	• May be involved in methionine biosynthesis, transport and/or utilisation	• Met-10 like-protein domain	F6a (T3)	~ 100 aa out of 295 aa	26, 44	• Pfam
Putative glycosyltransferase	• <i>Magnaporthe grisea</i> (A4RLP8) / Substrate-dependent	• Transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds	• Glycosyltransferase family 28 N-terminal domain with acceptor binding site and likely membrane association site	F6b (T7)	~ 120 aa out of 801 aa	68, 78	• Pfam
Putative glycosyl hydrolase	• <i>A. terreus</i> (Q0CXJ2) / Substrate-dependent	• Hydrolyse the glycosidic bond (O-glycosyl)	• Glycosyl hydrolase family 5 domain	F6b (T7)	~ 300 aa out of 761 aa	77, 87	• Pfam

## Results and Discussion

---

<sup>(a)</sup>Query sequence is a T7/T3 sequenced end of a fragment (F) for which a BLAST homology search was performed. <sup>(b)</sup> Homology stretch represents a number of amino acid residues within deduced subject sequence from TrEMBL database which was found to share homology with a given query sequence. <sup>(c)</sup> Identity (I) or similarity (P) of the deduced amino acid sequences shown off in %. <sup>(d)</sup>Reference describes used online database resources like Pfam and InterPro (see beneath) or a publication as a source of information. \* Pfam: Protein domain database; \*\* InterPro: Integrated resource of protein families, domains and functional sites; AMP – adenosine monophosphate; ATP - adenosine triphosphate; DNA - deoxyribonucleic acid; FAD - flavin adenine dinucleotide; PLP - Pyridoxal phosphate-dependent; PP- Phosphopantetheine binding domain; RNA - ribonucleic acid; SAM - S-adenosyl methionine:



### 8.3.1 Summary of F2-49C8 subsequencing

After subsequencing of the F2 fragment, no putative post-PKS acyl transferase activity needed for addition of acyl moiety to sorbicillactone A and B molecules was detected among the analyzed end-sequences (Table 17). Beside a putative condensing complex component Cnd2 protein function, the BLASTX results for F2e-T3 end-sequence matched a putative PLP-dependent aminotransferase activity as well. A rather low similarity value with the PLP-dependent aminotransferase (i.e. 36% identity) was found over a very short homology stretch (see Table 18) implying that this is probably not a real homologue. The PSI-BLAST with the F2e-T3 protein sequence showed that the most similar matches were annotated as condensing complex component Cnd2 proteins. Further sequencing of F2 fragment and/or functional studies might completely resolve this issue.

A subcloning and sequencing of bigger fragments that correlate to the end part of the analyzed AT-9-11 region [e.g. F4 (~8.0 kb) and F5 (~7.2 kb)] would be an option in order to screen for putative downstream members of predicted sorbicillactone gene cluster.

The analysis of sequence similarity with regard to putative sorbicillactone gene cluster is indicative of underlying structural similarity, which may be associated with similar function. However, it is not adequate to assume that all similar sequences share a common function. Thus, a reconstruction of evolutionary relationships of cloned PKS genes as well as investigation of the domain architecture within the deduced protein sequences is highly recommended. As already illustrated in section 6.2, the KS domain genealogy is successfully used to predict reducing character and thus to disclose the classification of the whole-length fungal PKSs. On the other hand, domains can be regarded as construction frames of protein functionality, i.e. the presence or lack of a protein domain may help to assess homology predictions based on BLAST search. Therefore, the following two sections (8.4 and 8.5) will attend to these two aspects of analysis of two cloned PKS genes from *P. chrysogenum*.

### 8.4 KS domain phylogenetic analysis of cloned whole-length PKS genes

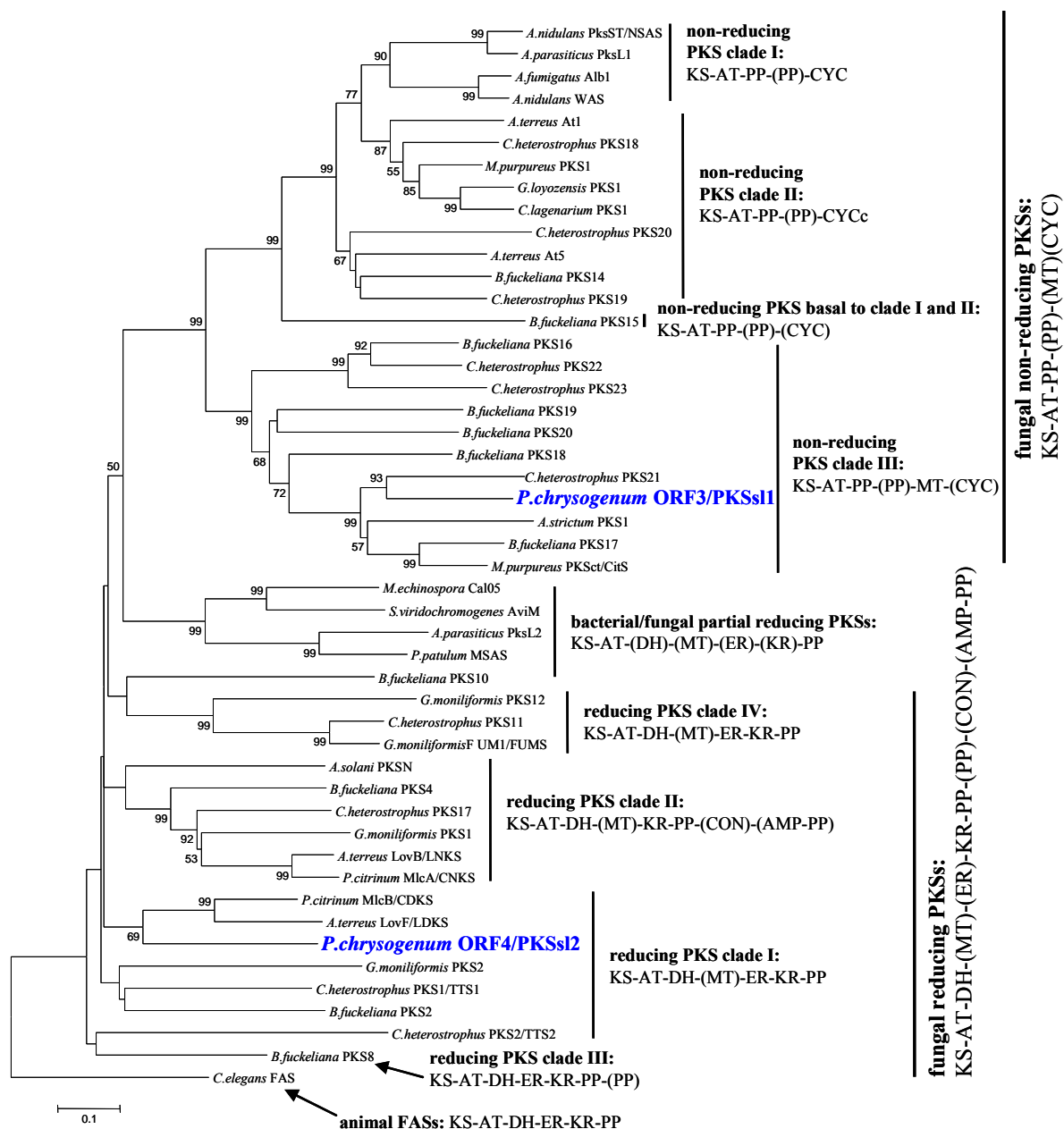
In order to confirm the reducing characters of the two cloned whole-length *P. chrysogenum* PKS genes (*PKSSL1* and *PKSSL2*, Figure 31), KS domains of 48 selected amino acid sequences available from Kroken *et al.* [167] were used for multiple alignments in the BioEdit program (see Table 16).

A neighbour-joining (NJ)-distance method from MEGA 4.0 (Table 31) was used to infer the phylogenetic tree after 1000 replications. The resulting bootstrapped NJ consensus tree is

shown in Figure 32. The KS domains from the cloned *P. chrysogenum* PKS are shown in blue (ORF3/PKSs11 and ORF4/PKSs12), while reference sequences are shown in black. Both, PKS classification and domain organisation in this figure are presented according to Kroken *et al.* [167]. The GenBank accession numbers of the protein sequences used for this alignment and subsequent phylogenetic analysis are listed in Table 19.

The resulting KS genealogy supported predictions made by previous BLAST analysis (see 8.2). In accordance to it, *P. chrysogenum* ORF3/PKSs11 KS domain grouped with members of the NR clade III, together with the characterised *M. purpureus* PKSct. As it can be seen in Figure 32, members of this subclade have in common the following domain organisation: KS-AT-PP-(PP)-MT-(CYC) [167]. As mentioned previously (see 2.3.1 and 6.2) they can alternatively contain the C-terminal reductase known as reductive domain (RED).

The second analysed *P. chrysogenum* KS domain (ORF4/PKSs12) clustered with the R clade I known to include two well characterised diketide synthases: LDKS [91] of *A. terreus* and CDKS [92] of *Penicillium citrinum*. The deduced general domain organisation of this subclade consists of KS-AT-DH-(MT)-ER-KR-PP domains [167]. However, many of the predicted PKS in this subclade have highly divergent and often non-functional MT domains.



**Figure 32: Phylogeny of the cloned KS domains from putative PKS of *P. chrysogenum* E01-10/3 strain.** Bootstrapped consensus tree with MEGA 4.0 inferred by the NJ method. Bootstrap values greater than 50 are shown at the nodes and were calculated from 1000 replications. The scale bar represents 0.1 substitution per amino acid site. KS domain sequences that were amplified from the genomic DNA of *P. chrysogenum* are shown in red, while reference sequences from GenBank are given in black. Classification of PKS based on Kroken *et al.* 2003 [167]. The accession numbers of sequences used in alignment are shown in Table 19.

**Table 19: Protein sequences for multiple alignment and phylogenetic analysis of KS domains from the whole-length PKS .**

Organism	Gene/Protein*	Accession number**
<i>Altenaria solani</i>	PKSN	gb:BAD83684
<i>Acremonium strictum</i>	PKS1/MOS	gb:CAN87161
<i>Aspergillus fumigatus</i>	<i>alb1</i>	gb: AAC39471

## Results and Discussion

<b>Organism</b>	<b>Gene/Protein*</b>	<b>Accession number**</b>
<i>Aspergillus nidulans</i>	PKST7NSAS	tr: Q12397
<i>Aspergillus nidulans</i>	wA/WAS	tr: Q03149
<i>Aspergillus parasiticus</i>	PKSs11( <i>pksA</i> )/ NSAS	tr: Q12053
<i>Aspergillus parasiticus</i>	PKSs12	gb: AAC23536
<i>Aspergillus terreus</i>	<i>at1</i>	gb: BAB88688
<i>Aspergillus terreus</i>	<i>at5</i>	gb: BAB88752
<i>Aspergillus terreus</i>	<i>lovB/LNKS</i>	gb: AAD39830
<i>Aspergillus terreus</i>	<i>lovF/LDKS</i>	gb: AAD34559
<i>Botryotinia fuckeliana</i>	PKS2	gb: AAR90238
<i>Botryotinia fuckeliana</i>	PKS4	gb: AAR90240
<i>Botryotinia fuckeliana</i>	PKS10	gb: AAR90246
<i>Botryotinia fuckeliana</i>	PKS8	gb: AAR90244
<i>Botryotinia fuckeliana</i>	PKS14	gb: AAR90250
<i>Botryotinia fuckeliana</i>	PKS15	gb: AAR90251
<i>Botryotinia fuckeliana</i>	PKS16	gb: AAR90252
<i>Botryotinia fuckeliana</i>	PKS17	gb: AAR90253
<i>Botryotinia fuckeliana</i>	PKS18	gb: AAR90254
<i>Botryotinia fuckeliana</i>	PKS19	gb: AAR90254
<i>Botryotinia fuckeliana</i>	PKS20	gb: AAR90256
<i>Caenorhabditis elegans</i>	FAS	gb: NP_492417
<i>Cochliobolus heterostrophus</i>	PKS1/TTS1	gb: AAB08104
<i>Cochliobolus heterostrophus</i>	PKS2/TTS2	gb: ABB76806
<i>Cochliobolus heterostrophus</i>	PKS11	gb: AAR90266
<i>Cochliobolus heterostrophus</i>	PKS17	gb: AAR90271
<i>Cochliobolus heterostrophus</i>	PKS18	gb: AAR90272
<i>Cochliobolus heterostrophus</i>	PKS19	gb: AAR90271
<i>Cochliobolus heterostrophus</i>	PKS20	gb: AAR90274
<i>Cochliobolus heterostrophus</i>	PKS21	gb: AAR90275
<i>Cochliobolus heterostrophus</i>	PKS22	gb: AAR90276
<i>Cochliobolus heterostrophus</i>	PKS23	gb: AAR90277
<i>Colletotrichum lagenarium</i>	PKS1/THSN	gb: BAA18956
<i>Gibberella moniliformis</i> ( <i>fujikuori</i> )	<i>fum1</i> /FUMS	gb: AAD43562
<i>Gibberella moniliformis</i> ( <i>fujikuori</i> )	PKS1	gb: AAR92208
<i>Gibberella moniliformis</i> ( <i>fujikuori</i> )	PKS2	gb: AAR92209
<i>Gibberella moniliformis</i> ( <i>fujikuori</i> )	PKS12	gb: AAR92219
<i>Glarea loyozensis</i>	PKS1	gb: AAN59953
<i>Micromonospora echinospora</i>	<i>calO5</i>	gb: AAM70355
<i>Monascus purpureus</i>	PKS1	gb: CAC94008
<i>Monascus purpureus</i>	<i>pksCT</i> /CitS	gb: BAD44749
<i>Penicillium citrinum</i>	<i>mlcA</i> /CNKS	gb: BAC20564
<i>Penicillium citrinum</i>	<i>mlcB</i> /CDKS	gb: BAC20566
<i>Penicillium patulum</i>	6MSAS/MSAS	gb: CAA39295
<i>Streptomyces viridochromogenes</i>	<i>aviM</i> /AviM	gb: AAK83194

\*Gene or protein designation in GenBank; \*\* GenBank (gb) or TrEMBL (tr) Accession Number

## 8.5 Domain analysis of cloned whole-length PKS genes

The results of the phylogenetic analysis enabled prediction of the reducing character for each of two cloned whole-length PKS genes from *P. chrysogenum* E01-10/3 strain. Based on the proposed subclade classification (see 8.4), a putative domain organisation was inferred. A detailed analysis on the domain architecture and conservation of amino acid residues within the PKS domain active sites was necessary in order to gain additional confidence for the results generated by the phylogenetic analysis (see Figure 22). Therefore, deduced amino acid sequences of these two cloned *P. chrysogenum* PKS genes were aligned in the BioEdit program with members of the corresponding fungal type I PKS clades from the phylogenetic tree (i.e. reducing/non-reducing). The active site residues were analysed for each putative domain within these two PKS proteins and consequently predictions were made with respect to their activity status. A supportive assistance for this analysis were the internet tools Pfam and InterPro (see Table 31) that are able to recognise domains within protein sequences (data not shown). The active site motifs analysis enabled valuable information upon preservation of PKS domains and their probable activity, and provided a platform for prediction of putative sorbicillactone A and sorbicillactone B PKS biosynthetic routes (see 8.6).

### 8.5.1 Non-reducing PKS

#### *Analysis of domain organisation and active sites motifs*

Based on the results of the BLASTX analysis (see 6.1.1) and KS domain genealogy (see 8.4), the identified *pksSL1* gene (i.e. *orf3*, Table 16) is supposed to be member of the NR clade III [167]. In order to check the existence of subclade characteristic domains [50, 167], the deduced PKSSl1 sequences from two generated contigs (4 and 7) were aligned with PKS members of NR clade III. The reference sequences from alignments are shown in Figure 33 and are detailed presented in Table 4. In the following text letter “x” stands for a variable amino acid residue within the active site of PKS domain in question.

At the very beginning of the deduced PKSSl1-protein-sequence, a SAT domain with the conserved GxCxG motif [255] was identified as shown in Figure 33-A. As previously mentioned (2.3.1), this N-terminal starter unit ACP transacylase domain is thought to be responsible for a transfer of advanced starter-unit (reduced) onto an ACP domain of fungal NR clade III PKS systems. The chemical structures of sorbicillactones support the model of reducing/non-reducing biosynthesis of these polyketides requiring of two different PKSs. In a recent study Crowford *et al.* [255] proved that SAT of NSAS catalyses the selective transfer

of hexanoate from CoA onto ACP. The conservation of this active-site-motif is an indicative that the identified SAT domain is able to perform its activity.

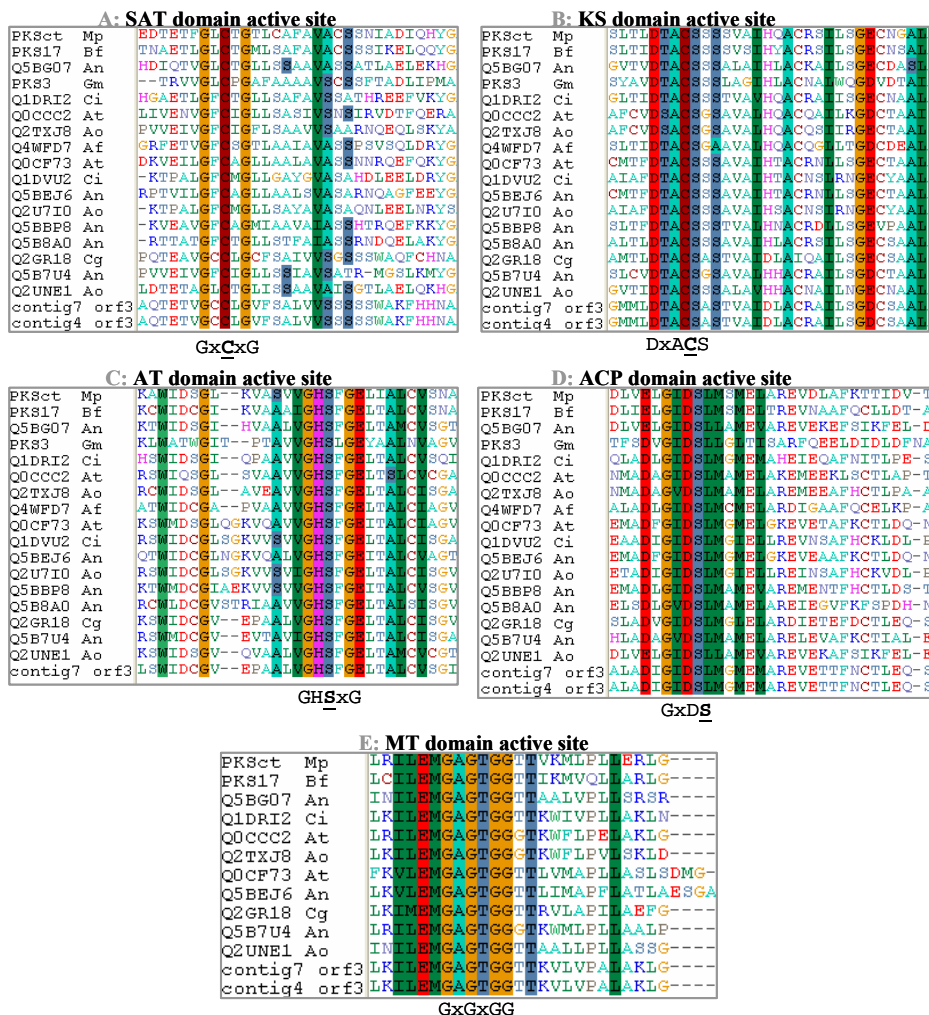
Following up, a KS domain with the conserved DxACS active site motif [46] was identified. It is important to mention that the position of acyl binding cysteine (C) was intact as shown in Figure 33-B. Further on, CHSxG conserved motif [46] of AT domain active site, with preserved pantetheine binding serine (S) was found as well (see Figure 33-C). According to the conservation of amino acid residues crucial for enzymatic activity, both domains should be active within PKSs11.

It is expected that a PT domain (i.e. Product Template) is expected to follow up the AT domain in typical NR clade III member. However, this domain is still not well characterised and an active site motif is not yet defined. Thus, its function is not established although there are assumptions that this domain could be involved in chain-length control [50, 256]. Consequently, in the absence of a defined active-site-motif, the whole size PKSs11 PT domain from contig7 was compared with PT domains of other NR clade III PKS whose polyketide products were already known. This phylogenetic analysis of PKSs11 PT domain, that was performed according to published analysis of Cox and collaborators [50, 170], is shown in the next subsection.

Moreover, the conserved motif of ACP domain – GxDS [46], with the preserved phosphopantotheine binding serine (S) was detected within PKSs11 as well (see Figure 33-D). The next located domain was a MT domain with the conserved glycine (G) loop (GxGxGG) [229, 230] that is involved in binding of *S*-adenosylmethionine (Figure 33-E). All designated glycines are preserved within the MT domain of PKSs11. In alignment for this domain less reference sequences were used since not all of the NR clade III members used for phylogenetic analysis contained the MT domain. At this level of sequence analysis, it seems that both domains could be functional in PKSs11.

Finally, at the C-terminus of the deduced sequence a domain corresponding to a NAD(P)H-dependent thioester-reductase (i.e. RED domain) was localised. As it was recently demonstrated by Bailey *et al.* [170], the RED domain of 3-methylorcinaldehyde synthase (MOS) is involved in reductive release of nascent polypeptide chain in form of an aldehyde. In addition, the position of a RED domain after an ACP domain supports the claim that this domain might not be involved in reductive modifications during PK-biosynthesis itself, but rather is involved in its release. To the best of my knowledge, the active site motif for this domain has not been defined yet. Nevertheless, the above analysed PKSs11 of *P. chrysogenum* shared several conserved motifs within this domain when aligned to other members of NR

clade III (alignment not shown). As noted before, the presumed path of the sorbicillactone A and B biosynthesis requires one RED domain to be active in the release of nascent polyketide



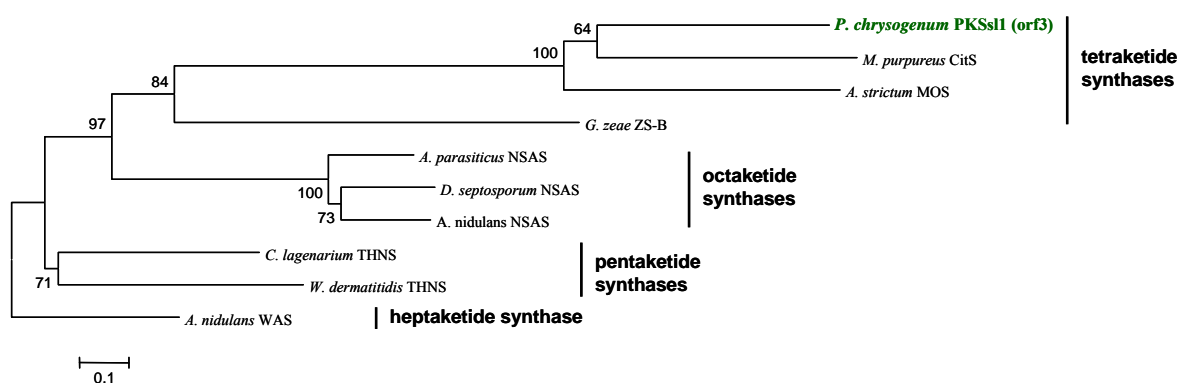
chains (see 3).

**Figure 33: PKS domain consensus motifs within *P. chrysogenum* PKSs11/orf3.**  
The deduced protein sequences that were used in alignments are shown in Table 19.

### Phylogenetic analysis of PT domain

The PT domain has been shown to have a low primary sequence similarity to other known proteins or domains in published in databases [257]. Nevertheless, significant similarity has been found among PT domains of putative fungal NR type I PKS systems. Accordingly, Cox and collaborators [50] have shown that the geneology of the PT domain correlates with the product length in the case of several characterised fungal NR PKSs. In consequence that would mean that disclosure of evolutionary relationships between PKSs11 PT domain of *P. chrysogenum* and PT domains of known fungal NR PKS could be helpful to anticipate the number of extension rounds performed by PKSs11.

For this purpose nine PT domain amino acid sequences, along with the *P. chrysogenum* PKSs11-PT sequence (in Figure 34 as “PT orf3”), were aligned together in the BioEdit program. Therefore, approximately 450-550 amino acids were extracted between AT and ACP domain of each sequence from the whole-length protein sequences (accordingly to alignment; data not shown). A neighbour-joining (NJ) distance method from MEGA 4.0 was used to infer the phylogenetic tree. The resulting bootstrapped NJ consensus tree is shown in Figure 34. The *P. chrysogenum* PT domain is shown in green colour, while reference sequences are shown in black. The GenBank accession numbers of the protein sequences used in this alignment and for the subsequent phylogenetic analysis, as well as respective polyketides are listed in Table 20.



**Figure 34: Phylogeny of PT domain from PKSs11 (orf3) of *P. chrysogenum* E01-10/3 strain.**

The bootstrapped consensus tree in MEGA 4.0 was inferred by the NJ method. Bootstrap values greater than 50 are shown at the nodes and were calculated from 1000 replications. The scale bar represents 0.1 substitution per amino acid site. PT domain of *P. chrysogenum* is shown in green, while reference sequences from GenBank are given in black. Classification of PKSs is based on Kroken et al. 2003 [167]. Accession numbers of sequences used in the alignment are shown in Table 20.



**Table 20: Reference sequences used for PT domain phylogenetic analysis.**

Organism	Gene*	Synthase component	Intermediate	Final product	Accession number**
<i>Acremonium strictum</i>	<i>PKS1</i>	MOS	Tetraketide	3-methylorcinaldehyde	gb:CAN87161
<i>Aspergillus nidulans</i>	<i>PKST</i>	NSAS	Oktaketide	Sterigmatocystin	tr: Q12397
<i>Aspergillus nidulans</i>	<i>wA</i>	WAS	Heptaketide	Naphthopyrone (YWA1)	tr: Q03149
<i>Aspergillus parasiticus</i>	<i>pksA</i> ( <i>PKSs11</i> )	NSAS	Oktaketide	Aflatoxin B1	tr: Q12053
<i>Colletotrichum lagenarium</i>	<i>PKS1</i>	THNS	Pentaketide	Tetrahydroxy naphthalene	gb: BAA18956
<i>Gibberella zeae</i>	<i>PKS13</i>	ZS-B	Tetraketide	Zearalenone	gb:AAS57297
<i>Dothistroma septosporum</i>	<i>pksA</i>	NSAS	Oktaketide	Dothistromin	gb:AAZ95017
<i>Monascus purpureus</i>	<i>pksCT</i>	CitS	Tetraketide	Citrinin	gb: BAD44749
<i>Wangiella dermatitidis</i>	<i>WdPKS1</i>	THNS	Pentaketide	Tetrahydroxy naphthalene	gb:AAD31436

\*Gene designation in GenBank; \*\* GenBank (gb) or TrEMBL (tr) Accession Number

In the calculated phylogenetic tree, the *P. chrysogenum* PT domain is most similar to the PT domains of *M. purpureus* PKSct/CitS, *Acremonium strictum* PKS1 and PKS13 of *Gibberella zeae*. These three PKS catalyze biosynthesis of tetraketide polyketides and thus were designated as tetraketide synthases [50]. Beside this, in Figure 34 all PKS are grouped analogue to the length of their biosynthesised products (i.e. number of extension rounds): Tetraketide-, pentaketide-, heptaketide- and oktaketide synthases.

In summary, the performed protein sequence analysis showed that the domain organisation of PKSs11 (Figure 35) is consistent with the one expected for members of NR clade III: SAT-KS-AT-PT-ACP-MT-RED (see 2.3.1). All known PKS active site motifs are conserved, and the analysed domains are not fragmented. These facts are supporting the evidence that they all may be active in *P. chrysogenum* PKSs11. Additionally, the PT domain phylogenetic analysis gave a clear hint that PKSs11 could be involved in the biosynthesis of a non-reduced tetraketide chain, as it is expected for the sorbicillactones according to their late biosynthesis steps.

**Figure 35: Domain organisation of *P. chrysogenum* PKSs11.**

The figure shows a putative non-reducing type I fungal PKS with domain organisation characteristic for NR clade III.

### 8.5.2 Reducing PKS

The results of the KS domain geneology in section 8.4 led to assumption that the *pksSL2*-encoded protein sequence (*orf4*, Table 16) is a member of R clade I. The subclade domain architecture was inferred from Kroken *et al.* [167] (see Figure 32) and active motifs were analysed through alignment of PKSsl2 protein-sequence (contig4 and contig7, see Figure 36) with characterised members of fungal reducing PKS clade. For each of the sequences used in the alignment, the respective subclades are shown in Figure 36 as well. The accession numbers for all reference sequences from the alignment are listed in Table 21.

#### *Analysis of domain organisation and active sites motifs*

The first identified N-terminal PKS domain is a KS domain with the conserved DxACS [46] active site motif shared by the majority of aligned protein sequences. An acyl-binding-cysteine (C) is preserved within the KS domain of PKSsl2. At the first position of the motif, an aspartic acid (D) residue is substituted by a glutamic acid (E) residue. Nevertheless, this substitution should not influence the KS domain activity since a similar substitution [aspartic acid (D) is substituted with histidine (H)] has been detected in PKS4 (i.e. ZS-A) that has been proven to be necessary for biosynthesis of mycotoxin zearalenone in *Gibberella zeae* [73]. Besides, the FUMS polyketide synthase involved in biosynthesis of fumonisine B1 in *Gibberella fujikuroi* [42] is also active despite the substitution of aspartic acid with lysine (K) [ see Figure 36-A].

As expected, directly after the KS domain, the AT domain's GH $\underline{S}$ xG [46] conserved motif along with conserved pantetheine-binding-serine (S) was identified. The amino acid glycine (G) at the first position in this motif is changed to serine (S) [Figure 36-B]. The same substitution is seen in LDKS (LovF, *A. terreus*) [90] and CDKS (MlcB, *Penicillium citrinum*) [92], two PKSs that are involved in the biosynthesis of lovastatin and compactin (respectively). Thus, there is evidence that this domain is fully active in PKSsl2.

Following up, a dehydrase (DH) active site conserved motif HxxxGxxxxP [258] was detected. Based on mutagenesis studies, it is up to now confirmed that the replacement of histidine (H) completely eliminates DH activity [259]. Therefore, this residue should be crucial for the activity of this domain. Less is known about the importance of preservation of other residues of this motif. The *P. chrysogenum orf4* possesses a conserved histidine (H) and a proline (P) residue, but glycine (G) is changed for serine (S) as shown in Figure 36-C. Nevertheless, other fungal R PKS from this alignment have been proven to be involved in the biosynthesis

of products like lovastatin (LDKS) [91], compactin (CDKS) [92], T-toxin (PKS1 and PKS2 of *Cochliobolus heterostrophus*) [175], and have exactly the same substitution at the same position (see Figure 36-C).

The next identified domain is a MT domain that is rather optional for members of the R clade I. As mentioned before, many members of fungal reducing clades feature MT domains that are either inactive due to changes in active site motif or domains that are completely absent. The typical MT active site motif is represented by conserved glycine (G) loop – GxGxGG, in which glycines form the pocket for SAM-binding [229, 230]. Within the putative R clade I PKS protein sequence from *P. chrysogenum* two glycines are missing in the glycine loop (GAATGA). In Figure 36-D the alignment with characterised PKSs that contain active MT domains was made, while in Figure 36-E inactive MT domains of otherwise active PKS are used for alignment with PKSs12-MT. As it is illustrated in Figure 36-D, active MT domains have conserved glycines at positions that are missing in PKSs12; while aligned MT domains thought to be inactive miss glycines at the same positions like in PKSs12 (Figure 36-E). According to structure and biosynthesis paths of produced polyketides (T-toxin and compactin), none of the two PKSs used in alignment in Figure 36-E requires functional MT domain (TTS1 [175] and CNKS [92], respectively). Based on this motif analysis it can be suspected that the MT motif of putative R clade I PKS encoded by *P. chrysogenum orf4* is inactive as well. This would be coherent to the proposed early biosynthesis steps of sorbicillactones that do not require C-methylation of a  $\beta$ -reduced “advanced” starter-unit (i.e. sorbyl moiety). The methylations of the unreduced part of sorbicillactones [see structure (**50**) and (**51**)] are presumably performed by the activity of the MT domain that is part of the previously described PKSs11 (see 8.5.1)

The nicotinamide-adenine dinucleotide phosphate (NADP)-binding sites of ER and KR are commonly recognised by the presence of the characteristic glycine-rich GxGxxG/A motif [258, 259]. The first two glycine residues in the motif are believed to be necessary to permit access of the pyridine nucleotide, since the presence of side chains would block this step. As experimentally shown [260], substitutions of these two glycines severely compromised reductive activities of these domains. However, this NADP-binding-motif is conserved within both – ER and KR domains of *P. chrysogenum* PKSs12 as it is shown in Figure 36-F and -G. Thus, most likely these two domains are active within this PKS.

At this point, it is interesting to mention that the absence of a PKS domain sometimes can be compensated by externally encoded enzyme function. Although biosynthesis of dihydro-monacolin L (i.e. intermediate of lovastatin) requires enoyl-reduction [90], a putative ER

domain of LNKS that is responsible for biosynthesis of this nonaketide does not contain a full NADP-binding consensus sequence as shown in Figure 36-F and has low similarity with the known ERs [23]. Nevertheless, the *lovC* gene [91], coding for an enzyme that shows similarity to ER domains of PKS, was found to be adjacent to the previously identified gene for LNKS (*lovB*) of *A. terreus*. Subsequent disruption of *lovC* blocked the production of dihydromonacolin L. Therefore, the conducted experiments suggested that at least two enzymes, LNKS and the *lovC*-encoded ER, are needed for proper biosynthesis of lovastatin's nonaketide intermediate dihydromonacolin L.

The last detected domain at the C-terminus of PKSs12-deduced protein sequence was an ACP domain. The GxDS<sub>2</sub> active site with conserved phosphopantetheine-binding serine (S) was found within this sequence as shown in Figure 36-H, thus, its activity was postulated within PKSs12.

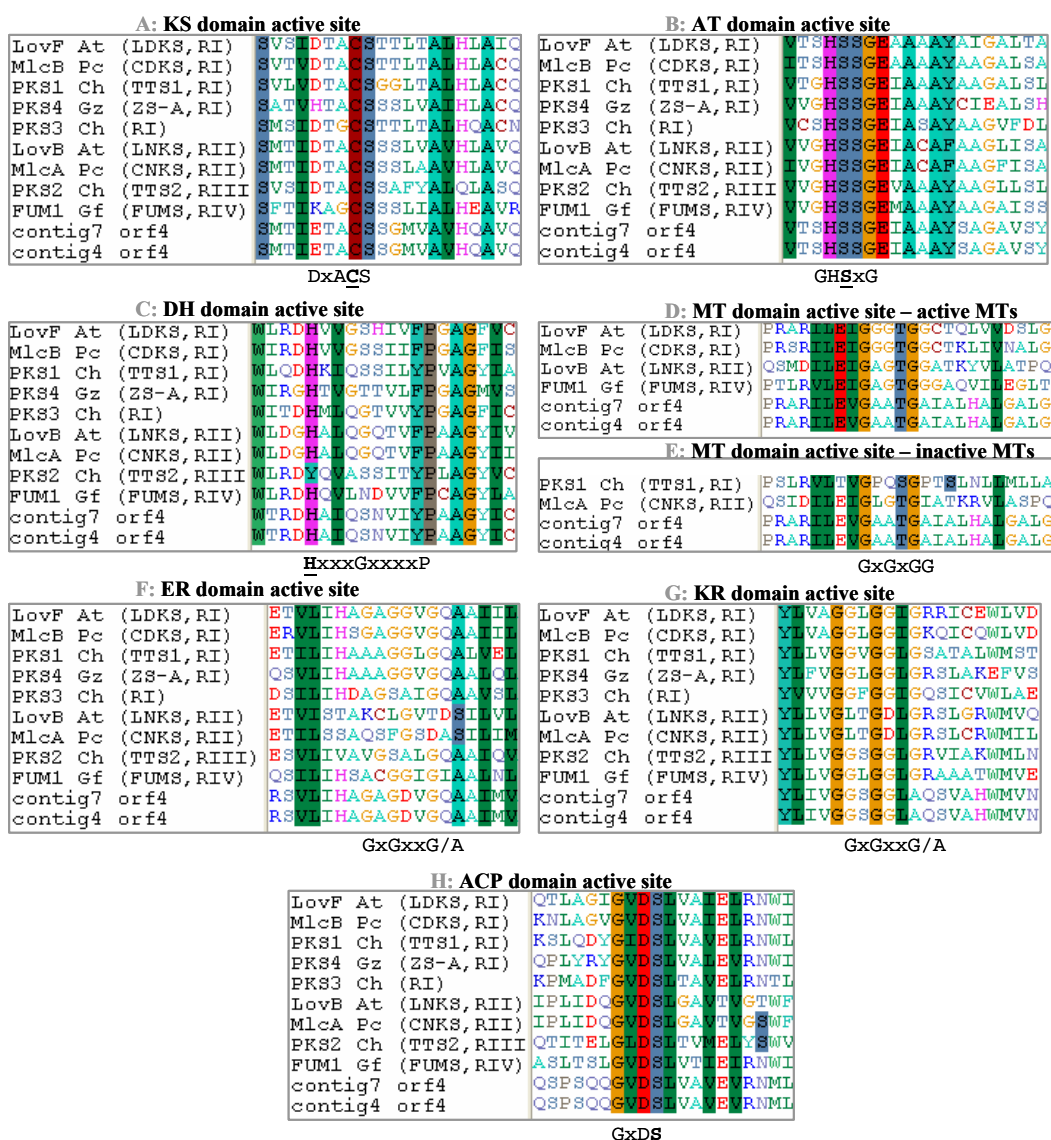


Figure 36: PKS domain consensus motifs within *P. chrysogenum* PKSS12/orf4.

Deduced protein sequences that were used in alignments are shown in Table 21. **A** represents part of alignment containing KS domain active site (DxACS); In **B** AT domain active site is shown (GHSxG), while DH domain active site (HxxxGxxxxP) is presented in **C**. **D** shows part of alignment of active MT domains containing consensus sequence GxGxGG, while in **E** inactive MT domains are align. In **F** and **G** active motifs of ER and KR are shown (respectively). Part of alignment encompassing ACP consensus sequence is shown in **H**. Conserved amino acid residues that are known to be crucial for domain-activity are shown in bold letters. The small “x” letter stands for variable amino acid residues within conserved motif. Contig7-orf4 and contig4-orf4 stand for deduced protein sequence of NR clade III PKSS1 that was identified within orf3 of compiled DNA sequences generated via the SeqMan® software of Lasergene™ during the course of this study.

Table 21: Reference sequences used for *P. chrysogenum* PKSS12 domain active site analysis.

Organism	Gene/Protein*	Reducing PKS subclade	Accession number**
<i>Aspergillus terreus</i>	<i>lovB/LNKS</i>	R II	gb: AAD39830
<i>Aspergillus terreus</i>	<i>lovF/LDKS</i>	R I	gb: AAD34559

Organism	Gene/Protein*	Reducing PKS subclade	Accession number**
<i>Cochliobolus heterostrophus</i>	<i>pks1/TTS1</i>	R I	gb: ABB76806
<i>Cochliobolus heterostrophus</i>	<i>pks2/TTS2</i>	R III	gb: ABB76806
<i>Cochliobolus heterostrophus</i>	<i>pks3</i>	RI	gb: AAR90258
<i>Gibberella fujikuroi</i>	<i>fum1/ FUMS</i>	R IV	gb: AAD43562
<i>Gibberella zeae</i>	<i>PKS4/ ZS-A</i>	R I	AAS57289
<i>Penicillium citrinum</i>	<i>mlcA/CNKS</i>	R II	gb: BAC20564
<i>Penicillium citrinum</i>	<i>mlcB/CDKS</i>	R I	gb: BAC20566

Based on this domain architecture analysis we confirmed that PKSs12 has the typical domain structure of R clade I PKSs : KS-AT-DH-(MT)-ER-KR-ACP [167]. Therefore, PKSs12 itself could be sufficient for biosynthesis of an advanced triketide starter unit, since it contains all of the domains required for condensation (KS, AT and ACP) and complete processing of the  $\beta$ -carbon (KR, DH and ER). Additionally, the mentioned domains have conserved consensus sequences within active sites and were not fragmented. However, there is a high chance that the detected PKSs12-MT domain is inactive, since the consensus sequence within its active site is not preserved. Such a non-methylated, reduced triketide starter unit generated by PKSs12 could be further accepted by the SAT domain of PKSs11 that is proposed to make three additional non-reducing extension rounds, presumably resulting in biosynthesis of hexaketide intermediates (see 8.6) of sorbicillactone A and sorbicillacton B.



**Figure 37: Domain organisation of *P. chrysogenum* PKSs12.**

The figure shows a putative reducing fungal type I PKS with the domain organisation characteristic for fungal R clade I PKSs. The MT domain is shown in red to point out its putative inactivity.

The putative biosynthetic routes for the two sorbicillactones presumably associated with the two *P. chrysogenum* PKS, are discussed in detail in the following section (see 8.6).

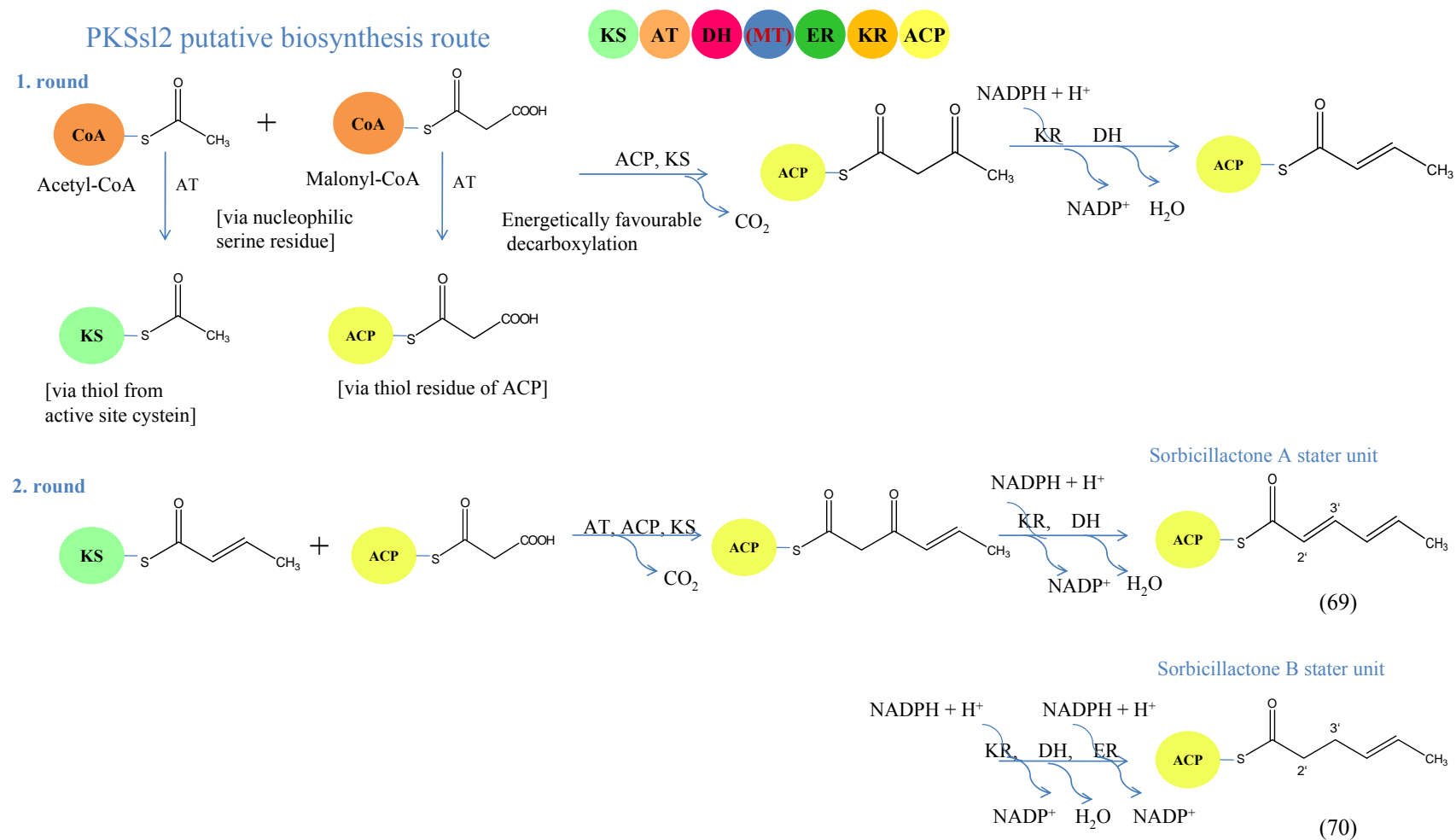
## 8.6 Proposed PKS biosynthetic route of sorbicillactones biosynthesis

The structures of sorbicillactone A (**50**) and B (**51**), two highly similar metabolites from the analysed *P. chrysogenum* E01-10/3 strain, are shown in introduction section (see 1.4). It is likely that the same PKS gene cluster could be responsible for biosynthesis of both metabolites. Besides, the chemical structures of both compounds indicated  $\beta$ -keto-reductive activities during the early PKS steps and no  $\beta$ -keto reduction in the remaining steps. Some other fungal polyketides like citrinin, dehydrocurvularin, monocerin and zearelenone share this structural feature as well [50]. Furthermore, specific incorporations of isotope-labelled

acetate supported the model of reducing/non-reducing biosynthesis of these polyketides. It was argued that such compounds may be formed by more than one PKS [50]. This would mean that a first PKS (or FAS) makes a  $\beta$ -reduced “advanced” starter unit, which is then passed to a second PKS for further extension in a non-reducing manner. This assumption is substantiated by the experiments of Watanabe *et al.* who isolated and characterised this kind of protein complex being involved in the aflatoxin biosynthesis of *A. parasiticus* [261]. Additionally, Crawford and coworkers have shown that the SAT domain of NSAS catalyses the transfer of the advanced starter unit (i.e. hexanoate) from CoA onto ACP and thus confirmed, based on sequence analysis the acyl transferase activity of this domain [255].

As described in section 8, two PKSs (PKSsl1 and PKSsl2) were found to be encoded next to each other in the genome of the analysed *P. chrysogenum* strain. The domain organisation and active sites analysis confirmed predictions about the reducing character of these two PKSs initially made according to the performed phylogenetic analysis (see Figure 32). The putative domain organisation of the reducing PKSsl2 and its probable role in the biosynthesis of advanced starter units for sorbicillactone A and B are drafted in Figure 38. Similarly, the putative domain organization and predicted biosynthesis routes for sorbicillactone A and B performed by the NR PKSsl1 are shown in Figure 39.

## Results and Discussion



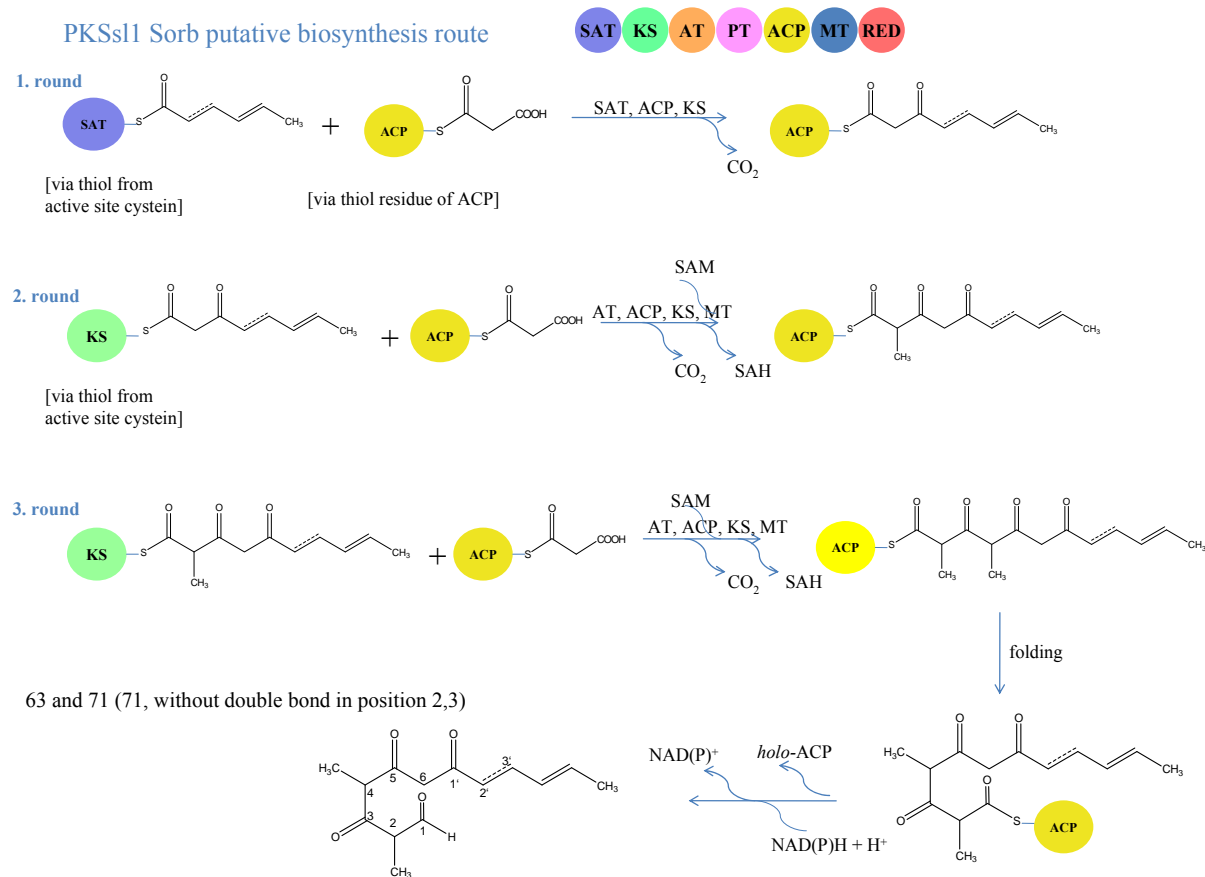
**Figure 38: PKSs12 putative biosynthesis route.**

The second round of putative PKSs12 biosynthesis route is different for sorbicillactone A and sorbicillactone B advanced starter units. NADP\* is oxidised, and NADPH reduced form of NADP coenzyme;

69 sorbicillactone A starter unit; 70 sorbicillactone B starter unit



## Results and Discussion



**Figure 39: PKSs11 – Sorbicillactone A biosynthesis route.**

SAM; *S*-adenosylmethionine; SAH, *S*-adenosyl-L-homocysteine; NAD(P) stands for cofactor that may be NAD or NADP; NAD(P)<sup>\*</sup>, oxidised form of cofactor; NAD(P)H, reduced form of cofactor.

*Dashed line is for optional double bond*

**63 Sorbicillactone A hexaketide intermediate**

**71 Sorbicillactone B hexaketide intermediate**

The biosynthetic routes in Figure 38-Figure 39 are presented in accordance to the general rules of polyketide biosynthesis that are already illustrated in section 2.1. The individual active site residues that are important for condensation of starter and extender units are considered in these routes and were previously found to be conserved in both *P. chrysogenum* PKS (see 8.5). As it was discussed in the same subsection, all other active site motifs within the PKS domains of both multifunctional enzymes were found to be conserved as well. The only exception is the MT of PKSs12 that was regarded to be inactive (see 8.5.2).

According to the structures of these two “sorbicillinoid alkaloids” and previous feeding experiments [136] (see 3), we expect that PKSs12 catalyzes two elongation rounds in order to create the non-methylated reduced triketide starter unit that correlates to the sorbyl sidechain of both compounds (see sorbicillactone A **50** and B **51**). Since sorbicillactone B represents the 2',3'-dihydroderivate of sorbicillactone A, the  $\beta$ -keto reducing steps within the second PKSs12 round are slightly different for these two compounds: That means that the creation of sorbicillactone A starter unit (**69**) uses only ketoreduction and dehydration steps (KR and DH domains, respectively), while the creation of the sorbicillactone B starter unit (**70**) additionally requires action of the ER in order to reduce the enoyl moiety to the saturated acyl moiety (i.e. 2',3'-dihydro, Figure 38). These advanced starter units are then transferred to the SAT domain of PKSs11 that is responsible for further three non-reducing elongation rounds leading to the doubly methylated hexaketide intermediates of sorbicillactone A (**63**) and sorbicillactone B (**71**) biosynthesis. The Figure 39 shows that the MT domain of PKSs11 is responsible for two C-methylations in each of two sorbicillactones by use of *S*-adenosylmethionine as a donor of a methyl-group. The first methylation takes place in the second round, while the second methyl group is added in the third round of the PKSs11 route. In line with the above way, we propose that the hexaketide thiolester intermediates are released in the aldehyde form (intermediates **63** and **71**) by action of the C-terminal reductase (i.e. RED domain), accompanied by the release of the free *holo*-ACP thiol of PKSs11 (Figure 39).

The predictions of the post-PKS biosynthesis steps were already made by Bringmann and collaborators based on the feeding experiments [136] and are presented in section 3. These experiments along with a recent study on biosynthesis of sorbicillinoids in *Trichoderma* sp [137] strongly support the assumption that the closure of the six-membered ring could be coupled with an oxidative process which may involve a putative monooxygenase activity (i.e. Slmox/ ORF2, see Table 16) that has been detected within the putative sorbicillactone cluster (see Figure 31).

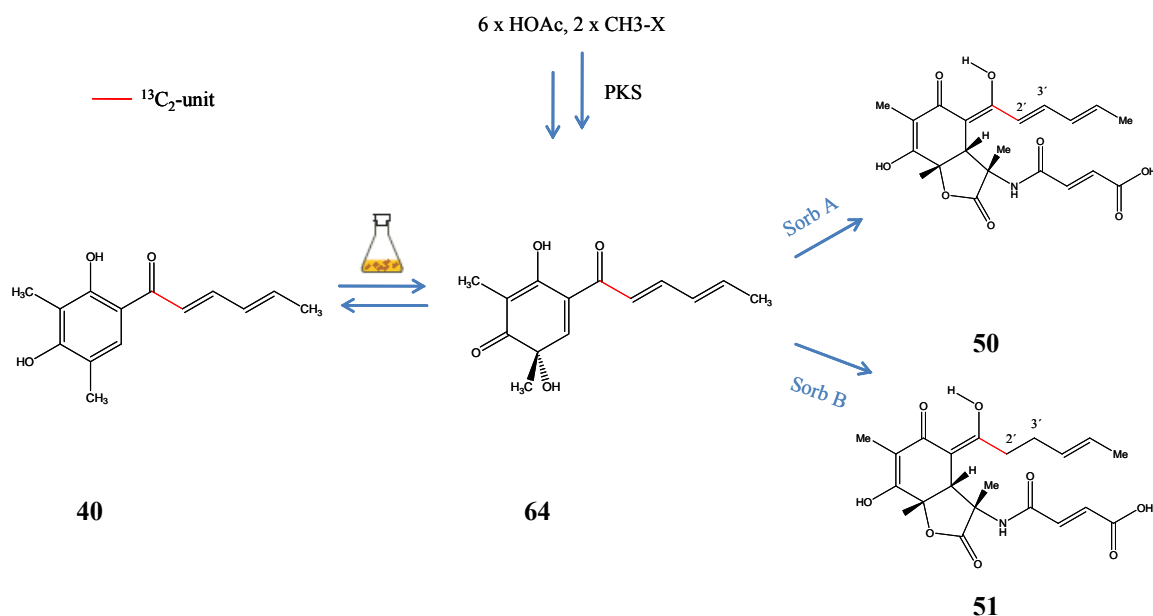
Up to now, there is only a hint that one PLP-dependent aminotransferase activity, which might be required for introduction of alanine, could be encoded from the sequenced genomic region of the *P. chrysogenum* E01-10/3 strain (see Table 18). Only small parts of deduced protein sequence (~70 aa) from F2 of 49C8 fosmid clone exhibit 36% identity to the PLP-dependent aminotransferase of bacteria *Methanococcoides burtonii* (GeneBank: Q12VN3). As emphasised in the introduction section 2.4.3, the most reliable *in situ* prediction of the protein function implies high sequence identity over the entire sequence. Similarly, no gene encoding for external AT activity was identified near the analysed gene cluster. However, such AT enzymatic activity is necessary in order to form post-PKS fumaryl-side-chain that is attached to the lactone ring of both sorbicillactone molecules.

## 9 Isotope-labelled feeding experiments

As already mentioned above, the production of sorbicillactone A in a marine-derived *P. chrysogenum* strain is ineffectively low. Therefore, it is crucial to know how to improve the production of sorbicillactone A and especially to clarify the time point of unwanted hydration of the C2'/C3' double bond and thus the moment of divergency in biosynthesis of sorbicillactone A and B. The feeding experiments aimed to check the incorporation of the <sup>13</sup>C<sub>2</sub>-labeled sorbicillin in *P. chrysogenum* and should demonstrate if the reduction of C-2'/C-3'-double bond occurs earlier or later in the course of biosynthesis of these natural products. The incorporation of <sup>13</sup>C<sub>2</sub>-labeled sorbicillin (**40**) into both sorbicillactone A and B would indicate the hydration at C-2'/C-3' at a later time point of biosynthesis, i.e. after the complete formation of the basic polyketide intermediate (see Figure 40). However, in case that the labeled precursor (**40**) would be incorporated only in sorbicillactone A and conversely <sup>13</sup>C<sub>2</sub>-2',3'-dihydrosorbicillin only in sorbicillactone B, then the reduction of the double bond for the formation of sorbicillactone B would be expected very early, e.g. at the stage of the linear poly-β-ketoesters like (**63**) and (**71**) (see Figure 39).

The feeding experiment of <sup>13</sup>C<sub>2</sub>-labelled sorbicillin in *P. chrysogenum* was conducted by the working group (WG) of Prof. Imhoff. The <sup>13</sup>C<sub>2</sub>-labelled sorbicillin, as well as unlabelled synthetic material used to optimise condition of the experiment, were prepared by the WG of Prof. Bringmann. Even though experimentators were able to resolve the initial problem concerning solubility of the substrate, it was not possible to exclude the considerable lowering of sorbicillactone A production with sometimes complete absence of the desired secondary metabolite. Furthermore, irrespective of the above-mentioned low production performances, it

was astonishing to find out that all fungal cultures completely metabolised applied  $^{13}\text{C}_2$ -labelled sorbicillin. Only in one set of experiments it was possible to isolate very low amounts ( $\ll 1$  mg) of sorbicillactone A. However, by means of NMR spectroscopy it was not possible to definitely prove the incorporation of  $^{13}\text{C}_2$ -labelled sorbicillin into sorbicillactone A.



**Figure 40: Feeding of  $^{13}\text{C}_2$ -labelled sorbicillin in *P. chrysogenum*.**

Determination of the time point of the unwanted hydration of the 2'-3' double bond in the biosynthesis of sorbicillactone A and B [262].

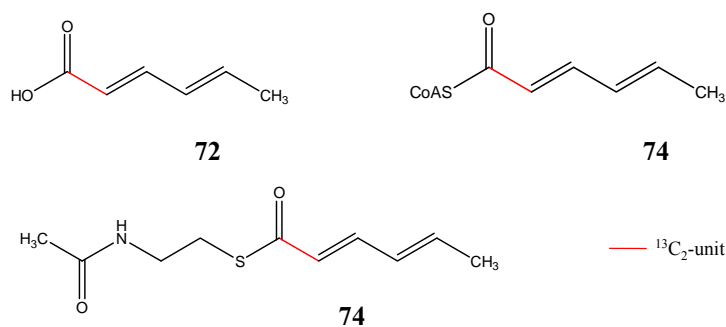
Furthermore, in repeated experiments with a larger amount of labelled substance in a higher number of cultures the production of the secondary metabolites was completely absent. Thus, the feeding experiments could not prove nor disprove the incorporation of  $^{13}\text{C}_2$ -labelled sorbicillin (40). According to Gulder T. A. [262], the future feeding experiments should be done with a higher number of independently inoculated small single cultures, instead of small numbers of big cultures, in order to increase probability of sorbicillactone A production.

On the basis of molecular genetic analysis of the putative sorbicillactone A gene cluster that was conducted in our group, it was proposed that the starter unit for the production of sorbicillactones is not an acetate but a sorbic acid-related precursor (see 8.6). A proof of this assumption would offer the possibility to increase the production of sorbicillactone A by means of feeding fungal cultures with higher amounts of sorbic acid as a precursor compound.

If the biosynthesis of sorbicillactone B (51) begins directly with 2',3'-dihydrosorbic acid, then a massive feeding with sorbic acid would be associated in the optimal case with a

complete collapse of sorbicillactone B production. This would be a simple alternative for the biotechnological production of sorbicillactone A. Thus, new serials of feeding experiments with  $^{13}\text{C}_2$ -labelled sorbic acid (**72**) were performed in working group of Professor Imhoff and Professor Bringmann. First, feeding experiments were optimised (WG Imhoff) with unlabelled sorbic acid with results thereof showing non-toxicity for the fungus. This initial administration of unlabelled sorbic acid was neither associated with a significant increase of sorbicillactone A and B production, nor a shift in the production ratio of sorbicillactone A and B. Under the same conditions feeding experiments were repeated with a  $^{13}\text{C}_2$ -labelled sorbic acid (**72**) synthesised by WG Bringmann. The fermentation products from these experiments were sent to WG Bringmann where sorbicillactone A and B were extracted and isolated. Subsequently, a potential incorporation of  $^{13}\text{C}_2$ -label into these compounds was determined with  $^{13}\text{C}$ -NMR spectroscopy; however, in both cases no incorporation of  $^{13}\text{C}_2$ -units could be detected.

One possible explanation according to the feeding experiments with sorbic acid is that the substance itself may not serve as a starter molecule for the sorbicillactone A biosynthesis. But it may also be possible that the compound is too polar to penetrate into the cell membrane of the fungus and therefore would not be available for the biosynthesis of (**50**) and (**51**). However, this is contrary to the observation that reference sorbic acid could not be detected in the culture incubations and extracts neither with HPLC-MS nor HPLC-UV and co-elution. Therefore, it seems that the substance was metabolised by the fungus, or catabolised for the gain of energy, but it was not used for the biosynthesis of (**50**) and (**51**). Another assumption is that sorbic acid in its free form is not used for sorbicillactone biosynthesis. This open question could be elucidated in future work by the synthesis of activated thioesters, e.g. (**73**) or (**74**) and their feeding to *P. chrysogenum* [262]. Additionally, a potential catabolic degradation may be suppressed by concomitant feeding with  $\beta$ -oxidation inhibitors.



**Figure 41: Putative substrates for sorbicillactone PKS: sorbic acid and its thioesters.**

72 sorbic acid, 73 thioester precursor, 74 thioester precursor. According to Gulder 2008 [262].

## 10 General discussion and further prospectives

The main goal of the present dissertation was to evaluate the genetic potential of *P. chrysogenum* E01-10/3 strain for the production of polyketides. This marine-derived *P. chrysogenum* strain was isolated from the Mediterranean sponge *Ircinia fasciculata* in the course of a research program focused on the discovery and characterization of novel natural products [147]. Within the course of that program two novel and structurally highly similar polyketides were isolated and characterised: sorbicillactone A (**50**) and sorbicillactone B (**51**) [136]. Consistent to this structural similarity, we proposed that biosynthesis of both compounds could be encoded by the same gene cluster. Consequently, we put our attention at the identification of the sorbicillactone A gene cluster as this compound was previously shown to possess promising antileukaemic, antiviral and neuroprotective properties [136].

Since all of fungal PKSs belong to the iterative type I PKS enzymes, degenerate primers and hybridization probes (6.1 and 7.2, respectively) fitting to this type of fungal PKS systems were used in PCR and hybridization experiments. Of special help for the present study was the possibility to differentiate between subtypes of fungal iterative type I PKSs on the amino acid level: Nonreducing (NR), partially reducing (PR), and highly reducing (HR) (see 2.3.1) in respect to level of reduction of their polyketide products. Accordingly, PCR and hybridization experiments were set up in order to take advantage of this fact.

During the course of this study, use of PCR enabled the amplification of partial PKS sequences from nine putatively distinct fungal type I PKS gene loci from *P. chrysogenum* E01-10/3. Six amplified sequences derived from the KS domain and were amplified with three different degenerate primer pairs (LC1/LC2c, LC3/LC5c [182] and KHKS2/KHKS3c [170], see 6.1.1). These six partial KS domain DNA sequences were used to reconstruct evolutionary relationships in respect to other iterative type I PKS (accordingly to Kroken *et al* [167]). The disclosed evolutionary relationships were then presented as the bootstrapped ME consensus tree created in MEGA 4.0 software from the alignment of deduced amino acid sequences in BioEdit (Figure 22, in 6.2). The results of the phylogenetic analysis for KS domains illustrated that *P. chrysogenum* E01-10/3 strain has the genetic potential to produce all three main categories of fungal polyketides – HR, PR (i.e. 6-MSA type) and NR. For five of six amplified KS domains results from the genealogy analysis supported the predicted subclade distribution according to previous BLASTX results (see 6.1.1 and 6.1.5). Three analyzed KS domains belonged to the NR fungal type I PKS clade. The resulting phylogenetic analysis showed that the KS<sub>nr-7</sub> sequence clusters to the NR clade<sup>o</sup>I, while

KHKS1 and KHKS32 are members of NR clade III. Two other PCR-amplified KS domains, KHKS3 and KHKS51, are grouped within R clade II. Last of all, the KSpr-MI sequence was grouped with members of joined bacterial and fungal type I PKSs that feature PR character of their polyketide products (see 6.2).

In addition, three non-KS partial sequences were amplified from the genomic DNA of the analysed *P. chrysogenum* strain. However, fungal PKS domains other than KSs have not yet been approved as suitable for the phylogenetic reconstruction that should give information upon reducing character and subclade distribution of a whole protein. Consequently, only the BLASTX-analysis was used to judge on the character of these three putative PKS (see 6.1.5). Among these non-KS sequences, one sequence was amplified with newly synthesised primers for the AT domain (sorb9-FGE-for/ sorb11-IWL-rev). The BLASTX-analysis indicated that the AT-9-11 sequence could be a member of the NR clade III (see 6.1.2). Besides, two different MT domain-derived partial PKS sequences were PCR-amplified as well. The CMeT-3 sequence was amplified with the CMeT1/CMeT2c primer pair (Nicholson *et al.*) [169], while for the amplification of the CMeT-9 sequence the CMeT1/CMeT3c primer pair was used (see 6.1.3). Each of the MT domain sequences showed homology to putative members of several fungal R-subclades (see Table 8). Based on such dispersed grouping of BLASTX matches it was difficult to infer a putative nature of the full-length PKSs. In order to predict the character of these two putative PKSs, analysis of other member domains would be necessary which implies localization of these two gene clusters in the genome of the analysed *P. chrysogenum* strain. Regarding time and funding limitations such analysis was not possible.

Despite the efforts to amplify the RED domain (see 6.1.4), no sequenced PCR product shared homology to any known fungal PKS gene. It is possible that the chosen PCR parameters or degenerate primer sequences were not optimal for the amplification of this particular PKS domain from this fungal strain. The fact that the RED domain appears relatively rare among fungal type I PKS systems and its generally low conservation may have influenced degenerate primer design and thus the outcome of the PCR screening study for this particular domain. The identification and sequencing of more fungal NR PKSs that harbour the RED domain and their deposition in public databases would improve chances to design degenerate PCR primers able to amplify RED domains from the analysed *P. chrysogenum* strain and many other fungal species as well.

Taking in account the range of PKS genes usually identified in one fungal strain by genome sequencing (see 2.4.2), the number of reported polyketides produced by different *P.*

*chrysogenum* strains (see 1.4), a relatively low degeneracy of PCR primers used in this study (e.g. Bingle *et al.* and Nicholson *et al.*), as well as the inherent limitations of PCR-based technology itself (see 2.4.2), it is to anticipate that the actual number of putative PKS genes within the analyzed *P. chrysogenum* strain could be much higher.

Further on, it is important to note that the results of the phylogenetic analysis of PCR-amplified partial KS domains were valuable to judge on good candidates for the screening of genomic library for the presence of the putative sorbicillactone A and B gene cluster via hybridization method (see 7.2) and PCR-based screening of the library (see 7.3). Moreover, the characteristic reducing-nonreducing dynamic of polyketide biosynthesis is reflected in the structures of both compounds and led to the anticipation that two differentially reducing PKS could be involved in the biosynthesis of these sorbicillactone metabolites. This would mean that a first PKS makes a  $\beta$ -reduced “advanced” starter-unit (corresponding to the sorbyl moiety of both sorbicillactones), which is then passed to a second PKS for further extension in the non-reducing manner according to Cox *et al.* [50].

The next assumption that was crucial for further experiments within the current study, and subsequent to the sorbicillactone A biosynthetic route proposed by Bringmann *et al.*, is that the hexaketide thiolester intermediates of sorbicillactone A and B are probably released in an aldehyde form by a C-terminal reductase encoded by a RED domain. Such polyketide release via the RED domain of MOS synthase was already proven in the case of 3-methylorcinaldehyde biosynthesis in *Acremonium strictum* [170]. Taking in account the rarity of the RED domain in NR PKSs, the current study focused on isolation of such fungal type I PKS systems from the genome of *P. chrysogenum* E01-10/3 strain. For this purpose, the genomic library of analysed strain with about 4800 individual clones was constructed (i.e. five time coverage of *P. chrysogenum* genome; see 7.1). This genomic library was screened for the presence of the sorbicillactone gene cluster by aid of hybridization (see 7.2) and PCR (see 7.3).

Despite efforts to screen the genomic library using different protocols for hybridization, no fosmid clone was detected to carry the genomic region potentially involved in the biosynthesis of sorbicillactone A and B (see 7.4). On the other hand, the PCR studies with the genomic DNA of *P. chrysogenum* E01-10/3 amplified three partial PKS sequences shown to be putative members of NR clade III according to the BLAST search results (see 6.1) and the subsequent phylogenetic study (see 6.2). Some members of the NR clade III, like PKSct [33] and MOS [170], have been reported to possess a RED domain, which makes this subclade particularly interesting for the search of putative sorbicillactone cluster. Thus, for each of



these three sequences (KHKS1, KHKS32 and AT-9-11) specific primer pairs were designed and used in complex genomic library screening procedure (see 7.3.1, 7.3.2 and 7.3.3).

However, during subcloning and sequencing of KHKS1 cluster candidates no homology was found to  $\beta$ -keto reducing domains, neither for MT and/or RED domains that are all prerequisite for the creation of hexaketide intermediates of both sorbicillactones biosynthesis. Moreover, via clear amplification of KHKS32-specific DNA sequence from AT-9-11 fosmid candidates it was proven that these two partial PKS sequences (KS and AT, respectively) belong to the same putative NR clade III PKS. Further on, subcloning and sequencing of the AT-9-11 genomic region revealed that one putative reducing PKS (homology over ER domain) and one non-reducing PKS (homology over SAT, MT and RED domains) were next to each other. Thus, the insert of the 42H12 fosmid clone, that has been anticipated to carry the core of the putative AT-9-11 PKS gene cluster, was completely sequenced and analysed (see 8.1).

By use of the “Fgenesh” [199] free online software, ten putative ORFs were detected within 46.91 kb of the recognised coding DNA sequence. All ORFs were analysed via BLASTP in order to detect putative protein homologues from public databases. Two of these ORFs (*orf9* and *orf10*) were found to be cloning vector sequences and were thus excluded from further analysis. Deduced amino acid sequences of other eight putative ORF matched fungal protein sequences from public databases and were further analyzed as shown in section 8.2.

The first identified gene encodes for a putative transcriptional regulator (*slr* or *orf1*), which could coordinate expression of the structural genes in the cluster. This putative regulatory gene is followed up by the gene encoding for one putative monooxygenase (*slmox* or *orf2*) that may be responsible for one post-PKS hydroxylation reaction during sorbicillactone A and B biosynthesis. The core of the identified gene cluster contains two PKS genes (*PKSSL1/orf3* and *PKSSL2/orf4*) located next to each other in opposite orientations. The second putative transcriptional regulatory gene (*sltr* or *orf5*) appears as a fifth ORF in the lane and is placed next to a putative gene for a MFS transporter protein (*slMFS* or *orf6*). As mentioned before, transporter genes are usually associated with gene clusters responsible for the biosynthesis of active secondary metabolites and play a role in excretion of potentially harmful intracellular metabolites to a surrounding environment as a kind of self-defence mechanism. The product encoded from the seventh ORF (*slox* or *orf7*) could be an oxidoreductase whose putative role in biosynthesis of sorbicillactones still cannot be predicted with certainty, although InterProScan online software recognised the FAD-linked oxidase domain that is characteristic for oxygen-dependent oxidoreductases. The last identified ORF that showed putative

homology to fungal proteins was *orf8* that shared significant similarity with fungal actin cytoskeleton-regulatory complex proteins. This ORF was excluded from the predicted sorbicillactone gene cluster model, since its involvement in cytoskeleton dynamics is not likely to be required for biosynthesis nor transport of sorbicillactones.

In order to screen for the missing PLP-dependent aminotransferase and *N*-acyl-transferase (i.e. external AT) protein functions expected to be encoded on the sorbicillactone gene cluster for the proposed biosynthetic route [136], a 12.5 kb of F2 from the 49C8 fosmid clone were additionally subcloned and sequenced (see 8.3). Despite efforts, no sequence was found to unambiguously support the presence of such two functions within the analysed AT-9-11 genomic region of the studied *P. chrysogenum* strain. It is possible that the required protein functions are encoded on a genomic region placed downstream from *orf8* that was up to now not analysed due to time limitation matter. Of course, it can not be that additional protein functions could be encoded from genes localised somewhere else in the genome of the fungus. Nevertheless, the localization of aminotransferase and acyltransferase activities and bringing them in connection to biosynthesis of the two sorbicillactones is necessary in order to disclose these biosynthetic routes.

In order to revise the BLASTP-inferred reducing characters of the whole-length PKS genes, the whole-length KS domains of these two PKSs were used to construct the phylogenetic tree with whole-length KS domains of the selected forty-eight protein sequences available from online supporting material by Kroken *et al.* [167] (see 8.4). The resulting bootstrapped NJ consensus tree constructed from MEGA 4.0, supported the predictions made by the previous BLASTP analysis (see 8.2). Accordingly, the *P. chrysogenum* ORF3/PKSs11 KS domain grouped with members of NR clade III, while the KS domain of ORF4/PKSs12 clustered to the R clade I.

The next task within current study was to analyse exact domain organisation of the identified *P. chrysogenum* PKS genes – *pksSL1* and *pksSL2*, in order to be able to predict their putative routes in sorbicillactone A and B biosynthesis. This was mainly done through alignment of the PKSs11 and PKSs12 protein sequences with members of corresponding fungal PKS clades (R/ NR) and subclades (I-IV). The results of BioEdit alignment and additional InterProScan protein domain analysis supported the subclade distribution and putative domain organisation of these two PKS genes. Moreover, the performed protein sequence analysis showed that domain organisation of PKSs11 (see Figure 35) is consistent with the one expected for the members of NR clade III: SAT-KS-AT-PT-ACP-MT-RED. All known PKS active site motifs were conserved and the analysed domains were not fragmented. These facts supported the

prediction that they are active in *P. chrysogenum* PKSs11. Additionally, the phylogenetic analysis of the PT domain gave clear hint that PKSs11 could work as a tetraketide synthase (Figure 34).

Based on the domain analysis it was confirmed that PKSs12 has the typical domain structure of a R clade I PKS: KS-AT-DH-(MT)-ER-KR-ACP [167]. Thus, PKSs12 could be sufficient for biosynthesis of the advanced triketide starter unit, since it contains all domains required for condensation (KS, AT and ACP) and the complete processing of a  $\beta$ -carbon (KR, DH and ER). Additionally, the domains had conserved consensus sequences within active sites and were not fragmented. On the other hand, there is a high chance that the detected PKSs12 MT domain has no activity since the consensus sequence within its active site is not preserved (Figure 37). Such non-methylated, reduced triketide starter unit could be produced by PKSs12 within two successive condensation and  $\beta$ -processing rounds. This advanced starter unit may be then accepted by the SAT domain of PKSs11 that probably makes three additional nonreducing extension rounds, since it was predicted to be a putative tetraketide synthase as mentioned before. At the end of the discussed putative sorbicillactone polyketide routes (see 8.6), it was proposed that hexaketide thiolester intermediates of sorbicillactone biosynthesis (**63** and **71**) are released in an aldehyde form by the RED domain encoded reductase, followed by the release of the free *holo*-ACP thiol of PKSs11 (see Figure 39).

The comparative analysis of deduced amino acid sequences from the identified putative sorbicillactone gene cluster enabled prediction of protein functions by means of previously identified genes deposited in the GeneBank database. Nevertheless, functional studies of these putative genes are necessary in order that their functions do not remain speculative. The fact that the analysed *P. chrysogenum* strain retained its capability for sorbicillactone A production is encouraging for future functional studies of the identified cluster. As illustrated in the introduction section, such studies focused on the generation and characterization of mutants for a gene of interest can be highly complex and time consuming. Especially, steps of DNA introduction and its integration into a fungal host represent obstacles for positive outcome of functional studies and can not be only species-specific but also dependent on the genetic background of particular fungal strain. Despite possible difficulties, encouraging for future functional studies on the putative sorbicillactone gene cluster is that several labs reported a successful transformation of *P. chrysogenum* strains [263-271] as well as a gene knock-out in *P. chrysogenum* [272, 273]. As it was shown in examples of NSAS of *A. parasiticus*, WAS from *A. nidulans* and THNS of *C. lagenarium* active site mutations or even complete removal of particular PKS domain are manageable in the fungal type I PKS systems

and provide valuable information on a single domain function [50]. Moreover, domain swap experiments were proven feasible in the case of the fungal type I PKS as well: The KS domain from *Cochliobolus heterostrophus* TTS1 was inserted into FUMS of *Gibberella fujikuroi* by Du and coworkers [221]. However, up to now, such experiments were not performed in our lab due to time and resource limitations.

Nevertheless, of special interest for the future investigation of the sorbicillactone pathway would be to elucidate the exact functions of two putative oxidoreductases (Slmox/ ORF2 and Slox/ORF7) encoded on the localised putative sorbicillactone gene cluster. More light on the biosynthetic routes of the sorbicillactones could be put by analyzing roles of particular domains from both identified putative PKS (PKSsl1/ORF3 and PKSsl2/ORF4) through functional protein studies. Especially, it would be interesting to point out if the ER domain of PKSsl2 is responsible for “unwanted” hydration of the C2'/C3' double bond in the early polyketide biosynthesis of both sorbicillactones. This is the only chemical difference between sorbicillacton A and its 2',3'-dihydro derivate (sorbicillactone B) that consequently “decides” upon their bioactivities [136, 262]. Otherwise, an external (non-PKS) oxidoreductase, as a part of the cluster, could be responsible for the hydration of the double bond at the later time point.

In general, feeding experiments with radioactively labelled precursor molecules can be used to approve proposed biosynthetic pathways [50]. Similarly, *P. chrysogenum* E01-10/3 strain cultures were fed with <sup>13</sup>C<sub>2</sub>-labelled sorbicillin and <sup>13</sup>C<sub>2</sub>-labelled sorbic acid in order to check time point of nondesirable hydration at C-2'/C-3'. Besides, feeding of fungal cultures with marked sorbic acid was aimed to test the hypothesis that two different (sorbyl-related) triketide precursors (**63** and **71**) of sorbicillactone A and B biosynthesis, could serve as advance starter-units that are successively accepted by PKSsl1. These experiments were performed by the working groups of Professor Imhoff (Kiel) and Professor Bringmann (Würzburg). Unlabelled precursors were used for an initial optimization of culture conditions. Afterwards, radioactively marked precursors were used in optimised experimental settings with bigger cultures (see section 9). Despite efforts, the feeding experiments could not prove nor disprove the incorporation of <sup>13</sup>C<sub>2</sub>-labelled sorbicillin. In the future, the main problem concerning the isolation of sufficient amount of radioactively labeled sorbicillin could be overcome by the use of a higher number of independently grown small single cultures, instead of small numbers of big cultures, in order to increase the probability of sorbicillactone A production.

Similarly, feeding experiments with sorbic acid were not able to show that sorbic acid itself is a precursor molecule for the biosynthesis of sorbicillactone A. The interesting point of this experiment was that labelled sorbic acid could not be detected within the produced sorbicillactone A and B, neither it could be identified in fungal cultures itself. This led to the proposal [262] that labelled sorbic acid could be metabolised by the fungus as energy source or its free form is not suitable as precursor of sorbicillactone A biosynthesis. According to Gulder T. A., the use of  $\beta$ -oxidation inhibitors could prevent that sorbic acid as putative precursor gets metabolised within the fungus for the reason of energy gain. Besides for the future experiments, the use of activated sorbic acid thioesters might answer the question of the precursor molecule for sorbicillactone A biosynthesis.

The nine putatively different fungal type I PKS genes of *P. chrysogenum* E01-10/3 represent the result of intensive PCR, hybridization, sequencing and bioinformatics studies that were performed within the course of the presented dissertation. Taking in account the number and potential bioactivities of polyketides isolated from marine-derived fungi (see 1.3 and 3), the number of compounds reported up to date from different *P. chrysogenum* strains (see 1.4), as well as in this study identified genetic potential of *P. chrysogenum* E01-10/3 for polyketide production, it becomes apparent that this particular marine-derived strain may be a valuable source of pharmacologically useful polyketides. Besides, dynamic and engagement within this research field (see 2.4) assure that concomitant with motivation and effort the fungal society will be soon able to answer many more questions related to these fascinating multi-enzymatic proteins and their polyketide products.

## MATERIALS AND METHODS

### 11 Materials

#### 11.1 Media

##### 11.1.1 Media for cultivation of *Penicillium crysogenum*

Malt extract medium (MEM)

Yeast extract 3 g, malt extract 3 g, peptone 5 g, glucose 10 g, NaCl 5g, purified water 1000 ml. Set up pH 6.7 before autoclaving ( with HCl).

##### 11.1.2 Media for the cultivation of *E. coli* Strains

**LB Medium** (Sambrook *et al.*, 2001) [274]

Tryptone 10.0 g, yeast extract 5.0 g, NaCl 10.0 g, agar 15.0 g (optional), purified water 1000 ml.

2 × YT Medium

Tryptone 16.0 g, yeast extract 10.0 g, NaCl 5.0 g, purified water 1000 ml.

**SOC Medium** (Sambrook *et al.*, 2001)

Bacto tryptone 20 g, yeast extract 5 g, NaCl 0.5 g, KCl 0.186 g, MgCl<sub>2</sub> 0.95 g, glucose 3.6 g, purified water 1000 ml, pH 7.0 (adjust with 5N NaOH). After autoclaving and cooling the medium up to 60 °C (at least) MgCl<sub>2</sub> and glucose were added to the medium from sterile stock solutions.

**For LB agar plates**, 0.9% (w/v) agar was added to the LB medium and the medium was then autoclaved. For the antibiotic selection of *E. coli* transformants, ampicillin (100 mg/l) or chloramphenicol (12.5 mg/l) were added to the autoclaved medium after cooling it to approximately up to 50 °C. For blue/white selection of *E. coli* transformants, 40 µl IPTG (20 mg/ml) and 40µl X-gal solution (20 mg/ml in 2% indimethylformamide) was plated on a 90 mm plate and the plate was incubated at 37 °C for at least 30 min before use.

#### 11.2 Buffers and solutions

##### 11.2.1 Buffers for plasmid DNA preparation from *E. coli*

**Solution I** (*E. coli* cell suspension buffer) (Sambrook *et al.*, 2001) [274]: 50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0). The solution was stored at 4 °C after autoclaving.

**Solution II** (*E. coli* lysis buffer) (Sambrook et al., 2001) [274]: 0.2 M NaOH (fresh diluted from a 10 M stock), 1% (w/v) SDS The solution was prepared freshly and used at room temperature (RT).

**Solution III** (neutralizing buffer) (Sambrook et al., 2001):

Potassium acetate (5 M)	60.0 ml
Glacial acetic acid	11.5 ml
Purified water	28.5 ml

The resulting solution was 3 M with respect to potassium and 5 M with respect to acetate and stored at 4 °C.

TE Buffer:

Tris-HCl (2 M, pH 8.0)	5 ml
Sodium EDTA (0.25 M, pH 8.0)	4 ml
Purified water	991 ml

### 11.2.2 Buffers for electrophoresis

**50× TAE Buffer** (Sambrook *et al.*, 2001) [274]: Tris base 242 g, glacial acetic acid 57.1 ml, EDTA ( 0.5M, pH 8.0) 100 ml. The concentrated buffer was usually diluted into 1× TAE as working solution.

**6× Gel-loading Buffer** (Sambrook *et al.*, 2001) [274]: Sucrose 40 g, purified water 100 ml, bromophenol blue 0.25 g.

### 11.2.3 Buffers and solutions for isolation of genomic DNA from *P. chrysogenum*

This method was adapted from Mutasa et al. [192].

**Lysis buffer:** The following sterile solutions were mixed together:

EDTA (0.5M, pH 7.5)	4 ml
20% sarkosyl	1 ml
Proteinase K stock solution (20 mg/ml)	200 µl
Purified water	14.8 ml

**20x SSC:** NaCl 175.32 g, tri-sodium citrate 2-hydrate 88.23 g, purified water 1000 ml 20× SSC has 3M NaCl and 0.3 M sodium citrate. The concentrated buffer was usually diluted for use as e.g. 1× SSC.

**10x TEN:** The following sterile solutions were mixed together:

Tris-Cl (1M, pH 7.5)	100 ml
EDTA (0.5M, pH 7.5)	20 ml
NaCl (5M)	200 ml
Purified water	680 ml

**Proteinase K Stock Solution (20 mg/ml):** The following sterile solutions were mixed together:

Proteinase K	40 mg
Ca(C <sub>2</sub> H <sub>3</sub> O <sub>2</sub> ) <sub>2</sub> (30mM)	100 µl
Tris-Cl (1M, pH 8.0)	100 µl
Purified water	1.8 ml

**20% Sarkosyl:** 20 g of sarkosyl (N-lauroylsarcosine sodium salt) were dissolved in 100 ml of purified water and mixed thoroughly.

#### 11.2.4 Solutions for preparation of competent *E. coli* cells by chemical method

**TFB I:** The following sterile solutions were mixed together:

Potassium acetate (5 M)	0.6 ml
CaCl <sub>2</sub> (1 M)	1 ml
KCl (1 M)	10 ml
MnCl <sub>2</sub> (1 M)	5 ml
Glycerol (100%)	15 ml
Purified water	68.4 ml

**TFB II:** The following sterile solutions were mixed together:

MOPS (1 M)	1 ml
CaCl <sub>2</sub> (1 M)	7.5 ml
KCl (1 M)	1 ml
Glycerol (100%)	15 ml
Purified water	75.5 ml

#### 11.2.5 Buffers and solutions for hybridization

**Denaturation Solution** (0.5M NaOH, 1.5M NaCl; for neutral DNA transfer, double-stranded DNA targets only): NaOH 20 g, NaCl 87.66 g, purified water 1000 ml

**Neutralization Buffer** (0.5M Tris HCL, 1.5M NaCl): Tris base 60.57 g, NaCl 58.44 g.

The pH of the buffer was adjusted to 7.4 with HCl.

**20× SSC:** NaCl 175.32 g, tri-sodium citrate 2-hydrate 88.23 g, purified water 1000 ml

20× SSC has 3M NaCl and 0.3 M sodium citrate. The concentrated buffer was usually diluted for use as e.g. 2× SSC, 5× SSC.

**10% SDS:** Sodium lauryl sulphate 100g, purified water 1000 ml. Heated to 68 °C for better dissolution. The pH was adjusted to 7.2 by addition of a few drops of concentrated HCl.

**Primary wash buffer without urea** (0.4% SDS, 0.5/0.1x SSC): 4g of sodium lauryl sulfate, 25/5 ml of 20x SSC, purified water 1000 ml. A stringent washing is increased by the use of a lower final concentration of SSC. The colony hybridization wash buffer without urea (0.1%



SDS; 2x SSC) was used as a final step for the fixation of DNA before the exposure to the UV light in the crosslinker.

**ECL Golden Hybridization Buffer:**

Composition of buffer in	100 ml
Gold hybridization buffer	100 ml
Blocking reagent	5 g (5%)
Sodium chloride (0.5 M)	2.922 g

Both gold hybridization buffer and blocking reagent were provided by the manufacturer. All content of one bottle of the gold hybridization buffer (500 ml) from ECL direct kit was transferred into a 1000 ml glass bottle. While vigorously stirring on a magnetic stirrer, 25 g of blocking reagent were added. 14.61 g of sodium chloride were added to the buffer as well. Solution was heated on the magnetic stirrer and stirred continuously for 2-3 hours. Aliquots of the buffer were dispensed into a 50 ml falcon tubes and stored in a freezer at -20 °C before usage.

**Church Buffer preparation:** The following sterile solutions were mixed together:

Sodium phosphate buffer (0.5M pH 7.2)	500 ml
EDTA (0.5M)	2 ml
BSA	10 g
SDA	70 g
Purified water	450 ml

Sodium phosphate buffer preparation:

The following sterile solutions were mixed together:

Solution A	342 ml
Solution B	158 ml
Purified water	500 ml
Set up pH to 7.2 before autoclaving	

**Solution A:** 177.9 g of  $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$  was dissolved in 1000 ml of purified water. Final concentration was 1 M.

**Solution B:** 137.99g of  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$  was dissolved in 1000 ml of purified water. Final concentration was 1 M.

### 11.3 Strains, vectors and plasmids

**Table 22: Used strains and their characteristics**

Strain	Characteristics	Reference or source
<i>E. coli</i> EPI300™-T1R	F <sup>-</sup> <i>mcrA</i> D( <i>mrr-hsdRMS-mcrBC</i> ) f80 <i>dlacZ</i> ΔM15 <i>DlacX74 recA1 endA1 araD139 D(ara, leu)7697</i> <i>galU galK l- rpsL nupG trfA tonA dhfr</i>	Epicentre Biotechnologies (Madison, USA)
<i>E. coli</i> DH5α	<i>SupE44 ΔlacU169 (Φ80 lacZ</i> ΔM15) <i> hsdR17 recA1</i> <i>endA1 gyrA96 thi-1 relA1</i>	Hanahan, 1983
<i>E. coli</i> XL1blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac[F'</i> <i>proAB lacIqZ</i> ΔM15 Tn10 (Tetr)]	Stratagene, Amsterdam (NL)
<i>P. chrysogenum</i> E01-10/3	Sorbicillactone A producer	IFM-GEOMAR collection

**Table 23: Used vectors and their characteristics.**

Vector	Size (kb)	Replicon	Origin	References
pBluescript KS-	3.0	pMB1	<i>lacZ orif1 bla</i>	Stratagene, Amsterdam (NL)
pCC1FOS	8.0	RepE	<i>lacZ oriV cam</i>	Epicentre Biotech (Madison, USA)
pGEM-T Easy	3.0	---	<i>lacZ orif1 amp</i>	Promega (Mannheim, Germany)

**Table 24: Used vectors and their characteristics**

Construct	Size (kb)	Vector	Insert	Resistance
pMA1	3.7	pGEM-T Easy	0.7 kb PCR fragment (LC3/LC5c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA2	3.3	pGEM-T Easy	0.3 kb PCR fragment (CMeT1/ CMeT2c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA3	3.3	pGEM-T Easy	0.3 kb PCR fragment (CMeT1/ CMeT2c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA4	3.7	pGEM-T Easy	0.7 kb PCR fragment (LC1/LC2c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA5	3.7	pGEM-T Easy	0.7 kb PCR fragment (LC1/LC2c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA6	3.7	pGEM-T Easy	0.7 kb PCR fragment (LC1/LC2c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA7	3.7	pGEM-T Easy	0.7 kb PCR fragment (LC1/LC2c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA8	3.7	pGEM-T Easy	0.7 kb PCR fragment (LC3/LC5c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>

## Materials and Methods

<b>Construct</b>	<b>Size (kb)</b>	<b>Vector</b>	<b>Insert</b>	<b>Resistance</b>
pMA9	3.3	pBluescript	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA10	3.3	pBluescript -	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA11	3.3	pBluescript	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA12	3.3	pBluescript	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA13	3.3	pBluescript	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA14	3.3	pBluescript	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA15	3.3	pBluescript	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA16	3.3	pBluescript	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA17	3.3	pBluescript	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA18	3.3	pBluescript	0.3 kb PCR fragment (KHKS2/KHKS3c primers) from <i>P. chrysogenum</i> (template)	Amp <sup>R</sup>
pMA19	~18	pBluescript	15 kb <i>Bam</i> HI fragment from fosmid 49C8	Amp <sup>R</sup>
pMA20	~15.5	pBluescript	12.5 kb <i>Bam</i> HI fragment from fosmid 49C8	Amp <sup>R</sup>
pMA21	~13	pBluescript	10 kb <i>Bam</i> HI fragment from fosmid 42H12	Amp <sup>R</sup>
pMA22	~10.2	pBluescript	7.2 kb <i>Bam</i> HI fragment from fosmid 42H12	Amp <sup>R</sup>
pMA23	~10	pBluescript	7 kb <i>Bam</i> HI fragment from fosmid 31F1	Amp <sup>R</sup>
pMA24	~10	pBluescript	7 kb <i>Bam</i> HI fragment from fosmid 31F1	Amp <sup>R</sup>
pMA25	~9	pBluescript	6 kb <i>Bam</i> HI fragment from fosmid 46H11	Amp <sup>R</sup>
pMA26	~ 6.2	pBluescript	3.2 kb <i>Bam</i> HI fragment from fosmid 46H11	Amp <sup>R</sup>
pMA27	~5.5	pBluescript	2.5 kb <i>Bam</i> HI fragment from fosmid 46H11	Amp <sup>R</sup>
pMA28	~ 5.1	pBluescript	2.1 kb <i>Bam</i> HI fragment from fosmid 49C8	Amp <sup>R</sup>
pMA29	~ 4.5	pBluescript	1.5 kb <i>Bam</i> HI fragment from fosmid 49C8	Amp <sup>R</sup>
pMA30	~ 7	pBluescript	4 kb <i>Pst</i> I subfragment of fragment 2 from fosmid 49C8	Amp <sup>R</sup>
pMA31	~ 6.2	pBluescript	3.2 kb <i>Pst</i> I subfragment of fragment 2 from fosmid 49C8	Amp <sup>R</sup>
pMA32	~ 5.7	pBluescript	2.7 kb <i>Pst</i> I subfragment of fragment 2 from fosmid 49C8	Amp <sup>R</sup>
pMA33	~ 4.8	pBluescript	1.8 kb <i>Pst</i> I subfragment of fragment 2 from fosmid 49C8	Amp <sup>R</sup>

## Materials and Methods

Construct	Size (kb)	Vector	Insert	Resistance
pMA34	~ 4.4	pBluescript	1.4 kb <i>Pst</i> I subfragment of fragment 2 from fosmid 49C8	Amp <sup>R</sup>
pMA35	~ 4.2	pBluescript	1.2 kb <i>Pst</i> I subfragment of fragment 2 from fosmid 49C8	Amp <sup>R</sup>
pMA36	~ 7	pBluescript	4 kb <i>Bam</i> HI fragment from fosmid 11D12	Amp <sup>R</sup>
pMA37	~ 5.8	pBluescript	2.8 kb <i>Bam</i> HI fragment from fosmid 11D12	Amp <sup>R</sup>
pMA38	~ 5.2	pBluescript	2.2 kb <i>Bam</i> HI fragment from fosmid 11D12	Amp <sup>R</sup>
pMA39	~ 4.8	pBluescript	1.8 kb <i>Bam</i> HI fragment from fosmid 11D12	Amp <sup>R</sup>
pMA40	~ 4.6	pBluescript	1.6 kb <i>Bam</i> HI fragment from fosmid 11D12	Amp <sup>R</sup>
pMA41	~ 4.2	pBluescript	1.2 kb <i>Bam</i> HI fragment from fosmid 11D12	Amp <sup>R</sup>
pMA42	~ 21	pBluescript -	18 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA43	~ 12	pBluescript	9 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA44	~ 8	pBluescript	5 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA45	~ 6.2	pBluescript	3.2 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA46	~ 5.8	pBluescript	2.8 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA47	~5	pBluescript	2 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA48	~ 4.7	pBluescript	1.7 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA49	~ 4.5	pBluescript	1.5 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA50	~ 4.4	pBluescript	1.4 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA51	~ 4	pBluescript	1kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA52	~ 3.95	pBluescript	0.95 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA53	~ 3.85	pBluescript	0.85 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA54	~ 3.8	pBluescript	0.80 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
pMA55	~ 3.7	pBluescript	0.75 kb <i>Hind</i> III fragment from fosmid 17B9	Amp <sup>R</sup>
FOS AN1 F2	~ 47.5	pEpiFOS <sup>TM</sup> -5	AN1034.2 locus <i>A. nidulans</i> FGSCA4	Chl <sup>R</sup>
FOS AN1 G11	~ 47.5	pEpiFOS <sup>TM</sup> -5	AN1034.2 locus <i>A. nidulans</i> FGSCA4	Chl <sup>R</sup>
FOS AN3 B5	~ 47.5	pEpiFOS <sup>TM</sup> -5	AN3386.2 locus <i>A. nidulans</i> FGSCA4	Chl <sup>R</sup>
FOS AN3 G8	~ 47.5	pEpiFOS <sup>TM</sup> -5	AN3386.2 locus <i>A. nidulans</i> FGSCA4	Chl <sup>R</sup>

### 11.4 Antibiotics and enzymes

**Table 25: Used antibiotics and their concentrations.**

Antibiotic	Working concentration	Company
Ampicilin	50-100 µg/ml	Roth (Karlsruhe, Germany)
Chloramphenicol	12.5 µg/ml	Fluka (Buchs, Switzerland)

**Table 26: Used enzymes and their buffers.**

<b>Enzyme</b>	<b>Reaction buffer</b>	<b>Company</b>
<i>Bam</i> HI	BamH I NEBuffer + BSA	NEB (Frankfurt/M, Germany)
<i>Cla</i> I	NEBuffer 4 + BSA	NEB (Frankfurt/M, Germany)
<i>Eag</i> I	NEBuffer 3	NEB (Frankfurt/M, Germany)
<i>Eco</i> RI	EcoR I NEBuffer	NEB (Frankfurt/M, Germany)
<i>Eco</i> RV	NEBuffer 3 + BSA	NEB (Frankfurt/M, Germany)
<i>Hind</i> III	NEBuffer 2	NEB (Frankfurt/M, Germany)
<i>Kpn</i> I	NEBuffer 1 + BSA	NEB (Frankfurt/M, Germany)
<i>Not</i> I	NEBuffer 3 + BSA	NEB (Frankfurt/M, Germany)
<i>Pst</i> I	NEBuffer 3 + BSA	NEB (Frankfurt/M, Germany)
<i>Rsa</i> I	NEBuffer 1	NEB (Frankfurt/M, Germany)
<i>Sac</i> I	NEBuffer 1 + BSA	NEB (Frankfurt/M, Germany)
<i>Sal</i> I	NEBuffer 3 + BSA	NEB (Frankfurt/M, Germany)
<i>Sma</i> I	NEBuffer 4	NEB (Frankfurt/M, Germany)
<i>Spe</i> I	NEBuffer 2 + BSA	NEB (Frankfurt/M, Germany)
<i>Xba</i> I	NEBuffer 2 + BSA	NEB (Frankfurt/M, Germany)
<i>Xho</i> I	NEBuffer 2 + BSA	NEB (Frankfurt/M, Germany)
T4 DNA Ligase	1 x T4 DNA Ligase Buffer	NEB (Frankfurt/M, Germany)
Shrimp Alkaline Phosphatase (SAP)	1 x SAP Buffer	Fermentas (St. Leon-Rot, Germany)
Proteinase K	Lysis buffer	Merck (Darmstadt, Germany)
Ribonuclease A	Plasmid preparation solution I	Roth (Karlsruhe, Germany)
Taq DNA Polymerase	1 x PCR Buffer	NEB (Frankfurt/M, Germany)
TaqBead Hot Start Polymerase	1 x PCR Buffer	Promega (Mannheim, Germany)

## 11.5 PCR primers

**Table 27: Used primers and their characteristics.**

<b>Primer</b>	<b>Function</b>	<b>Sequence (5'– 3')</b>	<b>Reference</b>
CMeT1	Amplification of MT domain of fungal <i>pks</i>	GARATIGGIGSIGGIACIGG	Bingle <i>et al.</i> , 1999
CMeT2c	Amplification of MT domain of fungal <i>pks</i>	ATIARYTTICCCICGGYTT	Bingle <i>et al.</i> , 1999
CMeT3c	Amplification of MT domain of fungal <i>pks</i>	ACCATYTGICCCICGGYTT	Bingle <i>et al.</i> , 1999
LC1	Amplification of KS domain of NR fungal <i>pks</i>	GAYCCNMGNTTYTTYAAYATG	Nicholson <i>et al.</i> , 2001

## Materials and Methods

Primer	Function	Sequence (5'– 3')	Reference
LC2c	Amplification of KS domain of NR fungal <i>pks</i>	GTNCCNGTNCRTGCATYTC	Nicholson <i>et al.</i> , 2001
LC3	Amplification of KS domain of PR fungal <i>pks</i>	GCNGARCARATGGAYCCNCA	Nicholson <i>et al.</i> , 2001
LC5c	Amplification of KS domain of PR fungal <i>pks</i>	GTNGANGTNGCRTGNGCYTC	Nicholson <i>et al.</i> , 2001
CMeT2c-F1	Amplification of MT domain of fungal <i>pks</i>	AARCCIGGIGGIAARYTIAT	Modification of Bingle <i>et al.</i> , 1999
CMeT3c-F2	Amplification of MT domain of fungal <i>pks</i>	AARCCIGGIGGICARATGGT	Modification of Bingle <i>et al.</i> , 1999
RED-F1a	Amplification of RED domain of funga <i>pks</i>	GTRGAIGGNGCNCARGG	This study
RED-F1b	Amplification of RED domain of fungal <i>pks</i>	GTYGAIGGNGCNCAAYGG	This study
RED-R1a	Amplification of RED domain of fungal <i>pks</i>	CCYTGIGCNCCNTCYAC	This study
RED-R1b	Amplification of RED domain of fungal <i>pks</i>	CCRTGIGCNCCNTCRAC	This study
RED-R2	Amplification of RED domain of fungal <i>pks</i>	TGYTCNACNGGRITCCARTAICC	This study
RED-R3a	Amplification of RED domain of fungal <i>pks</i>	ATNAGNCCNCCRCANGACAT	This study
RED-R3b	Amplification of RED domain of fungal <i>pks</i>	ATYAANCCNCCRCARCTCAT	This study
RED-F.AN1	Amplification of RED domain of <i>A. nidulans</i>	GTCACTGGTGC GACTGGCAGC	This study
RED-R.AN1	(FGSC A4) locus AN1034.2 <i>pks</i>	GTGCTCGACCGGGTTCCAGAAAACC	This study
RED-F.AN3	Amplification of RED domain of <i>A. nidulans</i>	GTCACTGGGGCCACAGGTAGC	This study
RED-R.AN3	(FGSC A4) locus AN3386.2 <i>pks</i>	ATGCTCGGCCGTGTTCCAGTACCC	This study
KS-F.AN1	Amplification of KS domain of <i>A. nidulans</i>	GATACGGCGTGTTCTCCTCGGCGGTG	This study
KS-R.AN1	(FGSC A4) locus AN1034.2 <i>pks</i>	TGCATTGACCCGCAAGCGCCGTAGTT	This study
KS-F.AN3	Amplification of KS domain of <i>A. nidulans</i>	GATACAGCGTGTTCTGGCTCTGCTGTA	This study
KS-R.AN3	(FGSC A4) locus AN3386.2 <i>pks</i>	AGCATTAGACCCACAAGCTCCATAATT	This study
MTnrIII_Fa		GARATGGGNGSIGGNCAIGG	This study
MTnrIII_Fb	Amplification of MT domain of fungal NR clade III <i>pks</i>	GARRTIGGNGCIGGNTTYGG	This study
MTnrIII_R1a		SSAAYGCIRTICAYGCNCA	This study
MTnrIII_R1b		CANAAYTYRTICAYGCICA	This study

## Materials and Methods

Primer	Function	Sequence (5'– 3')	Reference
KHKS2	Amplification of KS domain of fungal NR clade III <i>pks</i>	GCIGAYGGITAYTGYMGIGG	Bailey <i>et al.</i> , 2007
KHKS3c		GTICCICTICCRTGIGCYTC	Bailey <i>et al.</i> , 2007
Sorb1-FFK-for	Amplification of KS domain of fungal NR clade III <i>pks</i>	AYTAYGAYACNTTYGAYCAYAARTTYT TYAA	This study
Sorb2-PTG-rev		RAANGGYTTRCAYTGNCCNGTNGG	This study
Sorb3-KPF-for	Amplification of KS domain of fungal NR clade III <i>pks</i>	NCCNACNGGNCARTGYAARCCNTT	This study
Sorb4-GAS-rev		GCRTTNSWNCCNSWNGCNCC	This study
Sorb5-ASG-for	Amplification of AT domain of fungal NR clade III <i>pks</i>	CNYTNATHAAYAAYTAYGGNGCNWSNG G	This study
Sorb6-PVI-rev		ACYTGNCCNCCRAARCANARDATNACN G	This study
Sorb7-GQV-for	Amplification of AT domain of fungal NR clade III <i>pks</i>	CNGTNATHYTNTGYTTYGGNGGNCARG T	This study
Sorb8-VVG-rev		TCNCCRAANSWRTGNCCNACNAC	This study
Sorb9-FGE-for	Amplification of AT domain of fungal NR clade III <i>pks</i>	GTNGTNGGNCARWSNTTYGGNGA	This study
Sorb10-AFH-for		AAYG TNACNAAYGCNTTYCA	This study
Sorb11-IWL-rev	Amplification of AT domain of fungal NR clade III <i>pks</i>	TTNSWNCCNGCYTCNARCCADAT	This study
Sorb12-IFL-rev		TTNSWNCCNGCYTCNARYAADAT	This study
Sorb13-PPY-rev	Amplification of AT domain of fungal NR clade III <i>pks</i>	WYTTYTCRAAYTGRTANGGNGG	This study
Sorb14-SLG-for		GTNACNGGNGCNACNGGNWSNYTNGG	This study
Sorb15-IVHS-rev	Amplification of RED domain of fungal NR clade III <i>pks</i>	GGCCANGCNSWRTGNACDAT	This study
Sorb16-IVHN-rev		GGCCANGCRTTRTGNACDAT	This study
Sorb17-WNP1-rev	Amplification of RED domain of fungal NR clade III <i>pks</i>	AANGCNAARTGYTCNACNGGRTTC	This study
Sorb18-WNP2-rev		AANGCNAARTGYTCNACNGGRTTCCA	This study
Sorb-spec1-for	Screening genomic library for KS domain of KHKS1 <i>pks</i> sequence	GGGAGAGGGAGCAGGACTA	This study
Sorb-spec2-for		GGACTAGTTGTTCTACGACCCTTAGCAG	This study
Sorb-spec1-rev		CCAGCAGGGGCCTTCCGATAA	This study
Sorb-spec2-rev		CGTAAAGTGRCGTCTTCTGGAGGAATC	This study
Sorb 9-11 spec fl	Screening genomic library for AT domain of Sorb 9-11 <i>pks</i> sequence	GTCCATGGCCGGTCGAAGATTAT	This study
Sorb 9-11 spec r1		CCAAATGGCCTCCGGATACTGAC	This study

## Materials and Methods

Primer	Function	Sequence (5'– 3')	Reference
32 triple f1	Screening genomic library for KS domain of KHKS32 <i>pks</i> sequence	GCCGTGTTCTTGAAGAAATTATCC	This study
32 triple r1		CCTTGCCGACAACGTTCTGG	This study
pCC1 <sup>TM</sup> /pEpiF OST <sup>TM</sup> (FP)	End sequencing of fosmid clones from <i>P. chrysogenum</i> library	GGATGTGCTGCAAGGCGATTAAGTTGG	From CopyControl <sup>TM</sup> pCC1FOS <sup>TM</sup> vector (Epicentre Biotechnologies, Madison, USA)
pCC1 <sup>TM</sup> /pEpiF OST <sup>TM</sup> (RP)		CTCGTATGTTGTGTGGAATTGTGAGC	
M13-FP	Shotgun sequencing of putative sorbicillactone gene cluster	TGTA AACGACGGCCAGT	From pCR4 blunt Topo <sup>®</sup> (Invitrogen; Karlsruhe, Germany)
M13-RP		CAGGAAACAGCTATGACC	

### 11.6 Equipment and expendable materials

**Table 28: Used equipment and material.**

Equipment/Material	Model, Company
Agarose gel electrophoresis chambers	Sub-cell GT, BioRad (Munich, Germany)
Balances	440-47N (max 2000 g), Kern & Sohn (Balingen, Germany) BP 110 (max 10 g), Sartorius (Göttingen, Germany)
Autoclave	V 65, Systec (Wettenberg, Germany)
Centrifuges	5417R, Eppendorf (Hamburg, Germany) Micro 200, Hettich (Tuttlingen, Germany) Rotina 35 R, Hettich (Tuttlingen, Germany) Z 513 K, Hermle (Wehingen, Germany)
Disposable reagent reservoir 25/50/100 ml	Matrix Technologies Corporation
Serological plastic pipettes 5/10/25 ml	Sarstedt (Nümbrecht, Germany)
Hybridization transfer membrane	Hybond-N+, Amersham Biosciences
Centrifuge tubes, 15 ml	Sarstedt (Nümbrecht, Germany)
Centrifuge tubes, 50 ml	Sarstedt (Nümbrecht, Germany)
Dialysis membrane	Cellulose MWCO1000; Roth (Karlsruhe, Germany)
Microfuge tubes, 0.2/0.5/1.5/2.0 ml	Sarstedt (Nümbrecht, Germany)
Flasks 0.1-2.0 l	Simax (Sázava, Czech Republic)
Miracloth filtre paper Ø22-25µm	Calbiochem (Darmstadt, Germany)
Gel blotting paper 460 x 570 mm	Whatman (Dassel, Germany)
Agarose gel imaging system	Gene Genius, Syngene
Heating agitator	Combimag RCT, Ika-Labortechnik (Staufen, Germany)
Hybridization oven	OV 5, Biometra (Göttingen, Germany)
Hybridization glass tubes	Biometra (Göttingen, Germany)
Incubator	Heraeus (Hanau, Germany)



## Materials and Methods

<b>Equipment/Material</b>	<b>Model, Company</b>
Incubation shaker	Certomat BS-1, Sartorius (Göttingen, Germany)
Dark room lamp (660 nm, LED)	Cluster 50 RT, Kingbright (Issum, Germany)
Laminar flow (Biological Safety Cabinet)	Biowizard, Kojair (Trendelburg-Langenthal, Germany)
Microtiter plates	Sarstedt (Nümbrecht, Germany)
Parafilm	Alcan (Neuhausen, Switzerland)
Petri dishes, 8 cm(2)	Sarstedt (Nümbrecht, Germany)
pH-Meter	pH 211, Hanna instruments
Pipetts 0.5-10, 2-20, 10-200, 100-1000 µl (Pipetman)	Gilson (Limburg-Offheim, Germany)
Multipipett 10-200 µl	Eppendorf (Hamburg, Germany)
Pipettes tips	Sarstedt (Nümbrecht, Germany)
Rotational evaporator	Concentrator 5301, Eppendorf (Hamburg, Germany)
Serological glass tubes	Roth (Karlsruhe, Germany)
Thermal PCR cycler	T-gradient, Biometra (Göttingen, Germany)
Thermo paper for gel documentation	UPP 110HD, Sony (Berlin, Germany)
Tooth sticks	Roth (Karlsruhe, Germany)
Vortex	Mixer UZUSIO VTX-3000L, LMS Consult (Brigachtal, Germany)
Water baths	GFL Labortechnik (Burgwedel, Germany) JULABO Labortechnik (Seelbach, Germany)
Voltage supply sources	Standard Power Pack P25, Biometra (Göttingen, Germany) BioRad (Munich, Germany)
X-ray cassette	Rego (Augsburg, Germany)
X-ray film(18x24 cm)	Hyperfilm ECL; Amersham Bioscience
HPLC device	AS-2057 Plus, Jasco (Groß-Umstadt, Germany)
Thermo-mixer	Comfort, Eppendorf (Hamburg, Germany)
Shaker	Rotmax 120, Heidolph (Schwalbach, Germany)
Sterile filters (MWCO of 0.22 µm)	Millipore GmbH (Schwalbach, Germany)
Ultraviolet crosslinker	CL-1000, UVP (Cambridge, United Kingdom)
Microwave	Lifetec (Mülheim/Ruhr, Germany)
Replicator (96-well plates)	Boekel Scientific (Feasterville, USA; Key distributor in Germany: Fisher scientific, Schwerte)
Photometer	Biophotometer, Eppendorf (Hamburg, Germany)
Timer	Oregon Scientific (Neu-Isenburg, Germany)
Micropulser	BioRad (Munich, Germany)
Hybridization glass trays	Pyrex (Staffordshire, United Kingdom)
Gene Pulser Cuvette 0.2 cm electrode	BioRad (Munich, Germany)
Reagent glass bottles 30-2000 ml	Simax (Sázava, Czech Republic)
Digital graphic printer (gel documentation)	BioRad (Munich, Germany)

## 11.7 Chemicals and solvents

**Table 29: Used chemicals and solvents.**

Chemical/Solvent	Company
Acetic acid	Merck (Darmstadt, Germany)
Acetone	Merck (Darmstadt, Germany)
Acetonitrile	Roth (Karlsruhe, Germany)
Agar-agar	Roth (Karlsruhe, Germany)
Agarose	Roth (Karlsruhe, Germany)
Agarose Low Melt	Roth (Karlsruhe, Germany)
Bromophenol blue sodium salt	AppliChem (Darmstadt, Germany)
BSA (albumin fraction V),protease free	Roth (Karlsruhe, Germany)
Calcium chloride	Roth (Karlsruhe, Germany)
Chloroform	Roth (Karlsruhe, Germany)
Dimethylsulfoxide (DMSO)	Roth (Karlsruhe, Germany)
DNA ladder,100 bp,extended	Roth (Karlsruhe, Germany)
DNA ladder, 1k bp	Roth (Karlsruhe, Germany)
dNTP's mix	Invitrogen (Karlsruhe, Germany)
Ethanol	Roth (Karlsruhe, Germany)
Ethidium bromide	Roth (Karlsruhe, Germany)
Ethylacetate	Sigma-Aldrich (Steinheim, Germany)
Fosmid Control DNA	Epicentre Biotechnologies (Madison, USA)
Glycerol	Merck (Darmstadt, Germany)
Glucose	Merck (Darmstadt, Germany)
Hydrochloric acid	Merck (Darmstadt, Germany)
IPTG	Roth (Karlsruhe, Germany)
Isoamylalcohol	Roth (Karlsruhe, Germany)
Kodak GBX developer and replenisher	Sigma-Aldrich (Steinheim, Germany)
Kodak GBX fixer and replenisher	Sigma-Aldrich (Steinheim, Germany)
Malt extract	Difco (Lawrence, USA)
Magnesium chloride hexahydrate	Roth (Karlsruhe, Germany)
Magnesium sulphate dried	Roth (Karlsruhe, Germany)
3-(N-morpholino) propane sulphonic acid (MOPS)	Roth (Karlsruhe, Germany)
Orange G	Sigma-Aldrich (Steinheim, Germany)
Phenol/chloroform/isoamylalcohol (25:24:1)	Roth (Karlsruhe, Germany)
Potassium acetate	Roth (Karlsruhe, Germany)
Potassium chlorate	Roth (Karlsruhe, Germany)
2-Propanol	Merck (Darmstadt, Germany)
Sarkosyl	Merck (Darmstadt, Germany)
Sodium chloride	Roth (Karlsruhe, Germany)
tri-Sodium citrate dihydrate	Roth (Karlsruhe, Germany)
Sodium dihydrogen phosphate dihydrate	Roth (Karlsruhe, Germany)
Sodium dodecyl sulphate (SDS)	Roth (Karlsruhe, Germany)
Sodium hydrogen carbonate	Roth (Karlsruhe, Germany)
di-Sodium hydrogen phosphate dihydrate	Roth (Karlsruhe, Germany)
Sodium hydroxide	Roth (Karlsruhe, Germany)

Chemical/Solvent	Company
Tetramethylammonium chloride	Sigma-Aldrich (Steinheim, Germany)
Tris	Roth (Karlsruhe, Germany)
X-Gal	AppliChem (Darmstadt, Germany)
Yeast extract	Difco (Lawrence, USA)

## 11.8 Commercial kits

Table 30: Used commercial kits and their origin.

Kit	Company
ECL direct nucleic acid labelling and detection	Amersham Biosciences (Braunschweig, Germany)
pGEM Teasy kit	Promega (Mannheim, Germany)
QIAquick PCR Purification Kit	Qiagen GmbH (Hilden, Germany)
QIAquick Gel Extraction Kit	Qiagen GmbH (Hilden, Germany)
CopyControl™ Fosmid Library Production Kit	Epicentre Biotechnologies (Madison, USA)

## 11.9 Special computer programs and internet resources

Table 31: Used programs and internet sources for analysis of biological sequence information.

Program/ Resource	Function	Web address*	Source (available at)
FGENESH	HMM-based gene structure prediction in Eukaryota (multiple genes, both chains)	<a href="http://linux1.softberry.com/berry.phtml?topic=fgenes&amp;group=programs&amp;subgroup=gfind">http://linux1.softberry.com/berry.phtml?topic=fgenes&amp;group=programs&amp;subgroup=gfind</a>	At SoftBerry web site
BLAST	Finds regions of local similarity between protein or nucleotide sequences	<a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi">http://blast.ncbi.nlm.nih.gov/Blast.cgi</a>	NCBI web site
PSI-BLAST	Provides means of detecting distant relationships between proteins	<a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi">http://blast.ncbi.nlm.nih.gov/Blast.cgi</a>	NCBI web site
Advanced BLAST	A local alignment search tool similar to NCBI-BLAST	<a href="http://www.ch.embnet.org/software/aBLAST.html">http://www.ch.embnet.org/software/aBLAST.html</a>	At EMBnet (European Molecular Biology network) web site
InterProScan	Web available algorithm that identifies protein signatures from the InterPro member databases	<a href="http://www.ebi.ac.uk/Tools/InterProScan/">http://www.ebi.ac.uk/Tools/InterProScan/</a>	EMBL-EBI (European Molecular Biology Laboratory-European Bioinformatics Institute) web site
Pfam	The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and HMMs	<a href="http://pfam.sanger.ac.uk/search">http://pfam.sanger.ac.uk/search</a>	Sanger Institute web site

## Materials and Methods

<b>Program/ Resource</b>	<b>Function</b>	<b>Web address*</b>	<b>Source (available at)</b>
VecScreen	System for quickly identifying segments of a nucleic acid sequence that may be of vector origin	<a href="http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html">http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html</a>	NCBI web site
SEARCHPKS	Program for detection and analysis of PKS domains	<a href="http://www.nii.res.in/search/pks.html">http://www.nii.res.in/search/pks.html</a>	National Institute of Immunology New Delhi, India
Lasergene Version 7	Software package, which comprises several modules (see 12.19), is suitable for alignment of sequences, contigs assembly, primers design, and restriction mapping.	<a href="http://www.dnastar.com/">http://www.dnastar.com/</a> (only information)	DNASTAR, Inc, USA
FinchTV (Finch Trace Viewer) Version 1.4.0	A graphic viewer for DNA chromatogram files	<a href="http://www.geospiza.com/Products/finchtv.shtml">http://www.geospiza.com/Products/finchtv.shtml</a>	Geospiza web site
MEGA 4.0	Integrated tool for conducting automatic and manual sequence alignment, inferring phylogenetic trees, mining web-based databases, estimating rates of molecular evolution, and testing evolutionary hypotheses.	<a href="http://www.megasoftware.net/index.html">http://www.megasoftware.net/index.html</a>	MEGA web site
FramePlot Version 2.3.2	Web based tool for prediction of protein-coding regions in bacterial DNA with high GC content**	<a href="http://www0.nih.go.jp/~jun/cgi-bin/frameplot.pl">http://www0.nih.go.jp/~jun/cgi-bin/frameplot.pl</a>	Japanese Institute of Infectious Diseases web site
TreeView Version 1.6.6	Program for displaying phylogenies on Apple Macintosh and Windows personal computers	<a href="http://taxonomy.zoology.gla.ac.uk/rod/treeview.html">http://taxonomy.zoology.gla.ac.uk/rod/treeview.html</a>	Taxonomy and Systematics web site of the Glasgow University
BioEdit Version 7.0.0	Protein and nucleic acids sequence alignment editor and sequence analysis program	<a href="http://www.mbio.ncsu.edu/BioEdit/page2.html">http://www.mbio.ncsu.edu/BioEdit/page2.html</a>	Brown Lab web site
ClustalX Version 1.8	Multiple protein and nucleic acids sequence alignment computer program	<a href="http://www.clustal.org/">http://www.clustal.org/</a>	Clustal web site
FDA	It is an agency of the United States Department for Health and Human Services. It has responsibility to regulates and supervises the safety of foods, dietary supplements, drugs, vaccines, biological medical products etc.	<a href="http://www.fda.gov/Drugs/EmergencyPreparedness/BioterrorismandDrugPreparedness/ucm072755.htm">http://www.fda.gov/Drugs/EmergencyPreparedness/BioterrorismandDrugPreparedness/ucm072755.htm</a>	FDA web site

\* if applicable

\*\* guanine-cytosine content

## **12 Methods**

### **12.1 Centrifugation**

Centrifugation of small sample volumes in Eppendorff tubes was performed using the 5417R centrifuge model (Eppendorf; Hamburg, Germany) and the Micro 200 (Hettich; Tuttlingen, Germany), while centrifugation of larger sample volumes in falcon tubes was performed using the Rotina 35 R centrifuge model (Hettich; Tuttlingen, Germany) and the Z 513 K model (Hermle; Wehingen, Germany) centrifuge models.

### **12.2 Sterilization**

Heat-stable media, buffers, other solutions and glassware were sterilised in a steam autoclave Autoklav V65System, Wetzlar, at least 121 °C for 60-120 minutes (min). Millipore filters (Millipore GmbH, Schwalbach) with an MWCO of 0.22 µm were used to sterilise heat sensitive solutions. Purest water was obtained from deionised water applying a Milli-Q Reagent-Water system. It was sterilised under the given condition. If not stated differently this water was used for all media, buffers and solutions in molecular biological methods.

### **12.3 Cultivation of fungus *P. chrysogenum* E01-10/3**

Agar plates with malt extract media (MEM) were inoculated with fungal spores. After 5-6 days of incubation at 30 °C fungal spores were collected and stored in glycerol (up to 20%) or used for inoculation of liquid cultures. The fungus was grown in flasks containing 40-400 ml of liquid MEM depending on the purpose of the experiment - isolation of chromosomal DNA or detection sorbicillactone A production. Liquid cultures were incubated 3-10 days at temperature 28-30 °C with shaking (150 rpm; submerged cultures; collection of mycelium for DNA isolation) or without shaking (surface cultures; detection of sorbicillactone A production). In case of genomic DNA isolation, young mycelial cultures were preferred.

### **12.4 Cultivation of *E. coli* cells**

A single colony was inoculated onto 3 ml of LB liquid medium or a LB agar plate and incubated at 37 °C overnight for isolation of plasmid DNA or detection of covalently closed circular (CCC) DNA plasmid size. For selection of correct clones, different antibiotics were added to media at the appropriate concentration (see Table 25). *E. coli* DH5α and X11blue strains served as hosts for routine subcloning of plasmids. *E. coli* cells were concentrated by centrifugation and resuspended up to 20% glycerol for preservation at -80 °C.

### 12.5 Isolation of chromosomal DNA from *P. chrysogenum*

This method was established analogue to Mutasa *et al.* [192]. For the purpose of running PCRs and the creation of genomic library of *P. chrysogenum* E01-10/3, high molecular weight (HMW) DNA was isolated from this strain. Flasks containing 40 ml liquid MEM were inoculated with *P. chrysogenum* spores and liquid cultures were grown for 2-3 days at 30 °C with shaking conditions of 200 rpm. The fungal mycelium was harvested by filtration through thin filter (Miracloth filter paper Ø22-25 µm, see Table 28). The harvested mycelium was freeze dried by submersion into liquid nitrogen and then immediately grinded or stored at -80 °C prior to use. Approximately 2 g of freeze-dried mycelium was grinded in pre-chilled mortar (-80 °C, for 0.5-1 h) by addition of some sterile sand. The grinded powder was dissolved in 20 ml of 1x SSC (0.15 M NaCl, 0.015 sodium citrate, pH 7.4). Subsequently an equal volume (20 ml) of lysis buffer was added to the mixture. The mixture was incubated for 2 h at 55 °C in waterbath. The fungal mycelium was separated via centrifugation at 11,000 rpm for 10 min. After removal of mycelium, the mixture was extracted two times with phenol/chloroform/isoamylalcohol (IAA: 25:24:1) and once with chloroform/IAA (24:1) respectively. All centrifugation steps were run at the speed of 11,000 rpm for 10 min. Approximately 2-5 ml of recovered aqueous phase was dialysed (dialysis membrane, cellulose MWCO1000; Roth) against 2-4 ml of 1x TEN buffer (10 mM Tris pH 7.5, 1 mM EDTA, 100 mM NaCl) in a fridge at 4 °C. Samples were dialysed overnight (12-17 h) with continuous stirring. The content of dialysis membranes containing purified DNA was recovered in fresh falcon tubes. The DNA was precipitated by addition of one volume of isopropanol and centrifugation at 10,000 rpm for 5 min. The precipitated DNA was dissolved overnight at 4 °C in 1 ml of sterile water.

**Lysis buffer:** 0.1 M EDTA, 1% sarkosyl, 200 µg/ml proteinase K

### 12.6 Determination of DNA concentration and purity

The DNA concentration and purity were measured by means of a spectrophotometer (Biophotometer, Eppendorf). The absorption of 1.0 at 260 nm (1 cm thickness) corresponds to a concentration of 50 µg/ml of double stranded DNA. The degree of purity was determined by use of the  $A_{260\text{ nm}}$  to  $A_{280\text{ nm}}$  ration. This quotient was aspired to a value of about 2.0 (Sambrook *et al.*, 2001) [274].

Alternatively, the concentration and the quality of DNA was estimated performing agarose gel electrophoresis via comparison with appropriate DNA ladder that was applied as a DNA size and concentration standard.

### **12.7 Restriction enzyme digestion of DNA**

The restriction endonucleases cleave double stranded DNA into discrete fragments. The enzymes were used in a 0.5-10 U/ $\mu$ g DNA ration. The incubation occurred according to the manufacturer's instructions (1-3 h, 25 °C/37 °C) utilizing the appropriate reaction buffer. The list of endonucleases that were used within this study is given in Table 26.

### **12.8 Amplification of DNA using polymerase chain reaction (PCR)**

The polymerase chain reaction (PCR) is the procedure for the *in vitro* enzymatic amplification of a specific segment of DNA. The basis of this method is formed by a segment of double-stranded DNA to be amplified. Two single-stranded oligonucleotide primers flank this segment. The primers hybridise the opposite strands of the DNA in the manner that they are oriented with their 3' ends facing each other. A thermostabile DNA polymerase, catalyzing growth of new strands in 5'-3' direction, extends the DNA strand starting from the 3' ends of the primers. Thus, a new DNA strand is synthesised across the segments of DNA between the primers. The first cycle of synthesis results in new DNA strands of indeterminate length. The second cycle produces two single stranded products that together compose a discrete double-stranded DNA segment, which is exactly the length between the primer ends. In this manner, the discrete product accumulates exponentially with each successive cycle of amplification. Within each cycle of amplification the following steps occur; denaturation of the double stranded DNA, annealing of primers to the template DNA and primer extension by means of DNA polymerase (Goblet *et al.*, 1989) [275]. For PCR a thermocycler (T-gradient, Biometra; Göttingen, Germany) was used. In all PCRs, thermostabile *Taq* polymerase (*Thermus aquaticus*) was applied. This polymerase does not exhibit proofreading activity, a 3'-5-exonuclease activity causing an adjustment of incorrect nucleotides.

#### **12.8.1 PCR protocols**

The optimization of particular reaction condition was necessary to improve the yield of the desired product.

**Final concentration of MgCl<sub>2</sub>:** Usually an appropriate 10x PCR buffer was used to guarantee the optimal ion concentration and pH-value for the *Taq* polymerase. Final concentration of MgCl<sub>2</sub> varied between 1.5 and 4.5 mM.

**DMSO:** Dimethylsulfoxide is used to reduce unacceptably high levels of mispriming and to increase the efficiency of amplification of G + C rich templates [86]. Thus, PCR conditions were modified concerning the presence or absence of DMSO as cosolvent.

**BSA:** Bovine serum albumin (BSA) was used in the PCR solution, since BSA, which has no direct effect on the enzymatic reaction itself, can stabilise enzymes and neutralise inhibitory contaminants that may be present in the DNA template preparation or in the reaction buffers [276]. BSA also prevents adhesion of the enzyme to reaction tubes and pipette surfaces (New England Biolabs). The utilization of BSA was helpful for optimization of the PCR protocol for screening the genomic library of *P. chrysogenum*. Usually, the concentration of BSA in PCR reaction was 50 ng/μl.

**Primer concentration:** The primer concentration was modulated depending on the degree of primer degeneracy. Primers were used in concentrations between 20 to 100 pmol.

**Hot-Start PCR:** The hot-start PCR is a common technique to reduce non-specific amplification due to the assembly of amplification reactions at room temperature or on ice. At these lower temperatures, PCR primers can anneal to template sequences that are not perfectly complementary. Since thermostable DNA polymerases have activity at these low temperatures (although in most cases the activity is less than 25%) the polymerase can extend misannealed primers. This newly synthesised region is perfectly complementary to the DNA template, allowing primer extension and the synthesis of undesired amplification products. However, if the reaction is heated to temperatures >60 °C before polymerization begins, the stringency of primer annealing is increased, and the subsequent synthesis of undesired PCR products is avoided or reduced. The hot-start PCR can also reduce the amount of primer-dimer synthesised by increasing the stringency of primer annealing. The formation of nonspecific products and primer-dimers can compete for reagent availability with the amplification of the desired product. Thus, hot-start PCR can improve the yield of the specific PCR products [277]. DNA polymerase can be sequestered in a wax bead, which melts as the reaction is heated to 94 °C during the denaturation step, releasing the component only at higher temperatures [278-280]. In this study, *TaqBead*<sup>™</sup> Hot Start Polymerase (Promega Mannheim, Germany) was used to increase the specificity of the PCR with genomic DNA of *P. chrysogenum*.

Genomic DNA and plasmid DNA PCR:



## Materials and Methods

### Basic PCR:

Forward primer (50 $\mu$ M)	0.5 $\mu$ l	final conc., 0.5 $\mu$ M
Reverse primer (50 $\mu$ M)	0.5 $\mu$ l	final conc., 0.5 $\mu$ M
10x PCR buffer	5 $\mu$ l	final conc., 1x
dNTPs mix (10 mM for each)	0.25 $\mu$ l	final conc., 0.2 mM
Taq DNA polymerase (5 U/ $\mu$ l)	0.25 $\mu$ l	final conc., 1.25 U
Template DNA	1 $\mu$ l	10 – 100 ng
sterile purified water	42.5 $\mu$ l	
Total reaction volume:	50 $\mu$ l	

### Hot start PCR:

Forward primer (50 $\mu$ M)	1 $\mu$ l	final conc., 1.0 $\mu$ M
Reverse primer (50 $\mu$ M)	1 $\mu$ l	final conc., 1.0 $\mu$ M
10x PCR buffer (Mg-free)	5 $\mu$ l	final conc., 1x
MgCl <sub>2</sub> solution (25 mM)	3 – 6 $\mu$ l	final conc., 1.5 – 3.0 mM
dNTPs mix (10 mM for each)	1 $\mu$ l	final conc., 0.8 mM
Taq DNA polymerase (1.25 U/bead)	1 bead	final conc., 1.25 U
Template DNA	1 $\mu$ l	10 – 100 ng
Sterile purified water	37 $\mu$ l	
Total reaction volume:	50 $\mu$ l	

### Genomic library screening PCR:

The main difference between the following two PCR protocols is the material that is used as a DNA template through the screening procedure of clones containing particular PKS genes. The plasmid pool PCR protocol uses plasmid DNA that contains DNA from 96 clones from one microtiter plate. The whole-cell PCR protocol utilises a bacterial cell suspension that represents lane pools, or alternatively single clones from each lane, that originated from microtiter plates from genomic library of *P. chrysogenum*.

### Plasmid pool PCR:

Forward primer (50 $\mu$ M)	0.25 $\mu$ l	final conc., 0.5 $\mu$ M
Reverse primer (50 $\mu$ M)	0.25 $\mu$ l	final conc., 0.5 $\mu$ M
10x PCR buffer	2.5 $\mu$ l	final conc., 1x
dNTPs mix (10 mM for each)	0.25 $\mu$ l	final conc., 0.4 mM
BSA (50 ng/ $\mu$ l)	0.125 $\mu$ l	6.25 ng
Taq DNA polymerase (5 U/ $\mu$ l)	0.0625 $\mu$ l	final conc., 0.3 U
Template: superpool plasmid DNA	1 $\mu$ l	50 - 100 ng
Sterile purified water	20.56 $\mu$ l	
Total reaction volume:	25 $\mu$ l	

### Whole-cell PCR:

Forward primer (50 $\mu$ M)	0.25 $\mu$ l	final conc., 0.5 $\mu$ M
Reverse primer (50 $\mu$ M)	0.25 $\mu$ l	final conc., 0.5 $\mu$ M
10x PCR buffer	2.5 $\mu$ l	final conc., 1x

dNTPs mix (10 mM for each)	0.25 $\mu$ l	final conc., 0.4 mM
BSA (50 ng/ $\mu$ l)	0.125 $\mu$ l	6.25 ng
Taq DNA polymerase (5 U/ $\mu$ l)	0.0625 $\mu$ l	final conc., 0.3 U
Template: cell suspension	0.5 -1 $\mu$ l	
Sterile purified water	20.56 $\mu$ l	
Total reaction volume:	25 $\mu$ l	

### 12.8.2 PCR programs

In order to improve the yield of the desired product, the following conditions were modified:

**Extension time:** The extension time depended on the expected size of the amplicon. In general, for the amplification of gene fragments size up to 1 kb and extension time of 1 min was used. 74 °C or 75 °C were used as extension temperatures for PCR programs within this study.

**Annealing temperature:** The annealing temperature is significantly responsible for the product specificity of the polymerase chain reaction. To calculate the melting temperature of oligonucleotides, the Wallace-rule (Thein *et al.*, 1986) [281] was utilised:

$$T_m [^{\circ}\text{C}] = 2(\text{A}+\text{T}) + 4(\text{G}+\text{C})$$

**Temperature gradient:** For the purpose of working with the degenerate primers, the temperature gradient for primer annealing step was set up in range of 22 °C with the middle point temperature of 53 °C in the programs using the genomic DNA of *P. chrysogenum*. In the PCR programs used for screening the genomic library of *P. chrysogenum* for the presence of fosmids containing PKS gene clusters, the annealing step was set up in range of 5 °C with the middle point temperature of 57.5 °C that was in accordance to the predicted melting temperatures of specific primers used.

For *pks* screening with genomic DNA of *P. chrysogenum*

Denaturation:	94°C, 2 min	} 34 cycles
Denaturation:	94°C, 30 sec	
Primer annealing:	53°C, 1 min	
Extension:	75°C, 1 min	
Final extension:	75°C, 10 min	

For *pks* screening of the library with plasmid superpool DNA PCR protocol

Denaturation:	95°C, 2 min	} 35 cycles
Denaturation:	95°C, 1 min	
Primer annealing:	58/59°C, 1 min	
Extension:	74°C, 1 min	
Final extension:	74°C, 10 min	

For *pks* screening of the library with whole-cell PCR protocol

Denaturation:	95°C, 5 min	
Denaturation:	95°C, 1 min	} 39 cycles
Primer annealing:	58/59°C, 1 min	
Extension:	74°C, 1 min	
Final extension:	74°C, 10 min	

The primer annealing temperatures were slightly different dependent on the specific primers used. Generally, the specific primers were designed in such a way that their annealing temperatures were higher in order to prevent the unspecific annealing of the primers and the generation of false positive results.

### 12.9 Agarose-gel-electrophoresis

For the gel electrophoresis of DNA, 1-2% (w/v) agarose in 1x TAE buffer was dissolved in the microwave and then supplemented with 0.1 µg/ml ethidium bromide. After pouring and hardening of the gel, prior to the samples the gel pockets were loaded with loading buffer. The choice of loading buffer was dependent on the size of the expected fragments. Only for very small fragments a 6x Xylencyanol loading buffer was used, otherwise a 6x Bromphenol blue loading buffer was required. Alternatively, a 1.5% (w/v) agarose gel in 1x TAE buffer was applied for gel electrophoresis. Either a 0.6 – 1 µl of a 1 kb- or a 100 bp DNA marker was used for labeling. After filling the gel pockets the gel electrophoresis was started with a 90 – 100 V in 1x TAE buffer. The bands were then analyzed by UV light gel-documentation.

### 12.10 Purification of DNA fragments from solutions or agarose gel

The GFX PCR DNA and the Gel Band Purification Kit (Amersham Biosciences; Braunschweig, Germany) was used to purify DNA (e.g. PCR products or restriction fragments) from solution and from TAE and TBE agarose gel bands. For purification of DNA from gel bands, the gel slice containing the desired DNA fragment was measured and treated with the same volume of capture buffer. For example, 300 mg of gel slice afforded 300 µl of capture buffer. The mixture was incubated at 60 °C until the agarose was completely dissolved. After brief centrifugation to collect the sample at the bottom of the tube, the sample was transferred to a fresh GFX column. After centrifugation in a microcentrifuge at full speed for 30 seconds, the flow-through was discarded and the column was washed with 500 µl of wash buffer by centrifugation at full speed for 30 seconds. The GFX-column was transferred to a fresh 1.5 ml microcentrifuge tube and the DNA was eluted

with 10-50  $\mu\text{l}$  of purified water at room temperature for 1 min. Finally, the purified DNA was recovered by centrifugation at full speed for 1 min. For purification of DNA from solution, the DNA solution (up to 100  $\mu\text{l}$ ) was mixed with 500  $\mu\text{l}$  of capture buffer and transferred to the GFX column for centrifugation at full speed for 30 seconds. The following procedures for washing and elution are the same as described above.

### 12.11 Preparation of high quality plasmid DNA from *E. coli*

For the preparation of high quality plasmid DNA 1.5 ml of the overnight culture was decanted into an Eppendorf tube and centrifuged in a benchtop centrifuge at 13,000 rpm for 1 min. After complete removal of the supernatant, the cell pellet was carefully resuspended in 200  $\mu\text{l}$  of solution I (see 11.2.1). 200  $\mu\text{l}$  of freshly prepared solution II (see 11.2.1) was added, and the tube was inverted several times. Then 200  $\mu\text{l}$  of pre-cooled solution III (see 11.2.1) was added to the mixture, and the sample was inverted several times without vortexing. After centrifugation at 13,000 rpm for 3 min, the supernatant was transferred to a clean tube containing 500  $\mu\text{l}$  of chloroform, which was added in order to remove proteins. After mixing and centrifugation, the aqueous supernatant was transferred into a new tube containing 350  $\mu\text{l}$  of isopropanol, mixed by inversion and centrifuged for 20 min. The DNA pellet was additionally washed with 70% ethanol, centrifuged at 13,000 rpm for 10 min and dried under reduced pressure. At the end of procedure DNA pellet was dissolved in 40  $\mu\text{l}$  TE buffer.

### 12.12 Cloning of PCR products with the pGEM-T Easy Vector System

The GFX PCR DNA and Gel Band Purification Kit (Amersham Bioscience) was used to purify DNA from solution or from TAE and TBE agarose gel bands as described in 12.10. The purified DNA fragments were then ligated into the pGEM-T Easy Vector (Promega GmbH) by using 0.5  $\mu\text{l}$  of vector (50 ng/ $\mu\text{l}$ ).

Ligation reaction:

Rapid ligation buffer (2x)	5 $\mu\text{l}$
pGEM-T Easy Vector (50 ng)	0.5 $\mu\text{l}$
PCR product (~200 ng/ $\mu\text{l}$ )	3 $\mu\text{l}$
T4 DNA Ligase (3 units/ $\mu\text{l}$ )	1 $\mu\text{l}$
Total volume	10 $\mu\text{l}$

The reaction was mixed by pipetting and incubated for 1 hour at room temperature or at 4 °C overnight to get the maximum number of transformants. Half of the volume of the reaction incubate was used for the transformation of competent *E. coli* cells (XL1 Blue or DH5 $\alpha$ , see

Table 22), and transformants were selected by blue/white screening. 40  $\mu$ l X-gal (20mg/ml) and 40  $\mu$ l IPTG (200 mg/ml) solutions were spread on the agar LB plate containing appropriate antibiotics (e.g. ampicilin or chloramphenicol) prior to transformation. After overnight incubation at 37 °C, the desired clones, which did not contain active  $\beta$ -galactosidase due to insertion of foreign DNA, appeared as creamy-white. Transformants that contained only vector (without insert) were pale blue. Bacterial clones that contained inserts were transferred on a fresh LB agar plate.

### 12.13 Preparation of the TA cloning vector

The attempts to clone PCR products as blunt-ended fragments were very inefficient due to the template-independent terminal transferase activity of *Taq* polymerase. This activity results in the addition of a single nucleotide at the 3' end of the fragment, which is almost exclusively an adenosine due to the strong preference of the polymerase for dATP. Thus, cloning the products as blunt-ended fragments requires enzymatic processing to remove of the 3' overhang using an enzyme with 3' to 5' exonuclease activity. This template-independent activity of *Taq* polymerase can be exploited to create a cloning scheme, which has the efficiency of sticky end cloning, but requires no additional enzymatic modification of the PCR product. The 3' overhangs of PCR products have been exploited in a commercially prepared cloning system like the previously described pGEM T Easy Vector System (Promega GmbH, see 12.12). The simple procedure outlined here can be readily adapted to any plasmid or viral DNA based vector with a unique blunt-ended restriction endonuclease site. The T-vector generated in this protocol contains a single deoxythymidine (T) residue at the 3' ends. This T residue on the vector will hybridise to the single A overhang on the PCR product and increase the efficiency of ligation(according to Marchuk *et al.* 1990) [282].

#### 12.13.1 Enzymatic digestion of the pBluescript plasmid – introduction of blunt ends

The plasmid preparation was made from overnight cultures of the X11 Blue *E. coli* strain. The pBluescript plasmid was digested with *EcoRV* enzyme to introduce blunt ends into a plasmid. The set up of the restriction digestion was as follows below.

Digestion reaction:

pBluescript miniprep (100-200 ng/ $\mu$ l)	10 $\mu$ l
<i>EcoRV</i> enzyme (20 U/ $\mu$ l)	2.5 $\mu$ l
Buffer 3	3 $\mu$ l
BSA (50 ng/ $\mu$ l)	0.5 $\mu$ l
Purified water	32 $\mu$ l
Total volume	48 $\mu$ l

The digestion reaction was incubated for 2 h at 37 °C. *EcoRV* endonuclease was heat inactivated 20 min at 80 °C.

### **12.13.2 Enzymatic introduction of 3' T overhangs into the pBluescript plasmid**

To 48 µl of inactivated reaction from the previous step 0.5 µl of *Taq* Polymerase (20U/µl) was added (1 U/µg of plasmid/ 20 µl volume) together with 10 µl of 10 mM dTTPs (final concentration of 2 mM). The reaction was incubated 2 h at 70 °C.

### **12.13.3 Chloroform extraction and precipitation of pBluescript plasmid with 3'T overhangs**

The pBluescript plasmid was extracted with chloroform to remove rest enzyme activity and to generate purified vector that is ready to be directly ligated with PCR product.

100 µl of chloroform was added to the 50 µl of the reaction incubate from the previous step. The mixture was vigorously mixed and centrifuged for 5 min at 13,000 rpm. The supernatant was transferred to a fresh microfuge tube that contained 70 µl of isopropanol. A sample was firstly mixed via inversion and then centrifuged for 20 min at 13,000 rpm and 4 °C to precipitate plasmid DNA. The supernatant was carefully removed via pipetting and washed with 200 µl of ice-cold 70% ethanol. After short centrifugation at 13,000 rpm ethanol was removed via pipetting and pellet was dried under reduced pressure and dissolved in 10 µl of purified water. The T-vector was ready for ligation reaction with PCR product. 0.5-1 µl of vector was used in each ligation reaction.

## **12.14 Introduction of DNA into *E. coli***

### **12.14.1 Preparation and transformation of competent cells by the chemical method**

A single colony was inoculated in 3 ml of LB liquid medium for overnight culturing. On the next day, 1 ml of the overnight culture was transferred into 100 ml fresh LB broth and incubated at 37 °C until the OD<sub>560</sub> reached 0.4-0.5 value. *E. coli* cells were harvested by centrifugation at 5,000 rpm for 6 min at 4 °C. The pellet was gently resuspended in 25 ml of ice-cooled TFB I buffer (see 11.2.4). After centrifugation at 5,000 rpm for 5 min at 4 °C, cells were resuspended in 4 ml of TFB II buffer (see 11.2.4). The aliquots of 100 µl competent cells were dispensed into sterile microfuge tubes and stored at -80 °C. Up to 20 µl DNA solution was added to one tube containing thawed competent cells. After incubation on ice for ca. 20 min cells were heat shocked at 37 °C for 2 min (or 42 °C for 90 sec). The tube

was cooled down on ice and after 2 min 0.9 ml of fresh LB broth was added to cell suspension. After incubation at 37 °C for about 1 hour cells were spread on selective plates for overnight incubation at 37 °C.

#### **12.14.2 Transformation of *E. coli* cells by electroporation**

This method was adapted from Chassy et al. [283]. A single colony of *E. coli* was inoculated from a fresh agar plate into 50 ml of LB medium and cultured overnight at 37 °C with vigorous aeration (200 rpm). 1000 ml of pre-warmed LB medium was inoculated with 25 ml of the overnight bacterial culture and incubated at 37 °C with agitation until OD<sub>600</sub> reached ~0.4. The culture was transferred into several ice-cold centrifuge tubes (50 ml) and cooled down on ice for 15-30 min. The cells were harvested by centrifugation at 2,500 rpm for 15 min at 4 °C. The supernatants were decanted and the cell pellets were resuspended in 500 ml of ice-cold pure water. After centrifugation at 2,500 rpm for 20 min at 4 °C, the cells were washed three times with 500 ml, 250 ml and 10 ml of ice-cold 10% glycerol in turn. The cells were harvested by centrifugation at 2,500 rpm for 20 min at 4 °C and all liquids were carefully removed. Finally, the pellet was resuspended in 2 ml of ice-cold 10% glycerol. Aliquots of 50 µl of the suspension were dispensed into sterile Eppendorf tubes and stored at -80 °C. For the transformation reaction a volume of 1-2 µl DNA solution was added to 50 µl of the freshly thawed electrocompetent cells and mixed by pipetting for several times. The mixture was transferred to the bottom of an ice-cold electroporation cuvette while avoiding bubbles. The outer side of the cuvette was dried with paper tissues before electroporation. A pulse of electricity was delivered to the cells at 25 µF capacitance, 2.5 kV and 200 Ohm resistance. As quickly as possible after the pulse, 1 ml of LB medium was added at RT and the mixture was transferred to a sterile eppendorf tube and incubated by gentle rotation for 1 hour at 37 °C. Different volumes (accordingly to expected transformation rates) of the culture were spread onto a LB agar plate containing appropriate antibiotics for selection of clones.

#### **12.15 Preparation of genomic library of *P. chrysogenum***

This protocol was modified from “EPICENTRE” CopyControl Fosmid Library Production Kit protocol. However, all components of the kit were used for the preparation of the genomic library.

### **12.15.1 Preparative gel purification of genomic DNA**

Isolated genomic DNA of *P. chrysogenum* was size-selected through a 20 cm long 1% LMPagarose gel. 100 µl of genomic DNA preparation containing approximately 2 µg DNA (20 ng/µl) was loaded into a 8-10 cm wide genomic DNA loading lane together with some loading dye. 100 ng of 40 kb control DNA was loaded into each of the outside lanes to be used as a marker for the isolation of correctly sized genomic DNA. Samples were resolved via gel electrophoresis: initial run 10-30 min at 60-70 V enabled that samples fast enter gel and thus prevented loss of DNA. After initial run gel was run overnight (12-14h) at 35-45 V. On the next day, using a sharp razor blade, 2 cm gel slices from both sides of the gel were made. These gel slices contained approximately 1 cm (each) of the well into which the genomic DNA was loaded. Both slices were stained with ethidium bromide for 20-30 min. UV light was used to locate the band of high molecular weight (HMW) genomic DNA (~ 40 kb) in the stained edge slices of the gel. By using a clean razor blade 5 mm wide gel slices from the central part of the non-stained gel was cut out. This slice contained *P. chrysogenum* genomic DNA corresponding to 40 kb of control DNA.

### **12.15.2 Recovery of the size-fractionated genomic DNA**

The gel slice containing the genomic DNA was placed in a pre-weighed 15 ml falcon tube. LMP agarose was melted by incubation at 70 °C for 10-15 min. The tube was transferred at 45 °C, and pre-warmed 50x GELase Buffer (Epicentre Biotechnologies) was added to a tube (final concentration 1x). 1 U (i.e.1µl) of GELase enzyme stock was added to the tube for each 500 µl of melted agarose. The reaction was incubated at 45 °C for at least one hour. The GELase enzyme was inactivated at 70 °C for 10 min and the mixture was transferred to an ice bath for 5 min to cool down. The sample was centrifuged at a speed of 10,000 rpm for 20 min to a pellet of insoluble oligosaccharides. The DNA was precipitated by the addition of 2.5 volume of ethanol and 0.1 volume of sodium acetate (pH 7.0). The sample was incubated for 10 min at RT and then centrifuged at 16,000 rpm for 20 min. The supernatant was carefully removed from the pelleted DNA. The pellet was washed two times with ice cold 70% ethanol with subsequent centrifugation and removal of supernatant. A pellet was then air-dried for 5-10 min under the clean bench and dissolved in 60 µl of TE buffer. The dissolving of DNA was eased by an incubation of 1-2 h at 50 °C or an overnight incubation at 4 °C. The DNA concentration and quality was determined by running an aliquot (0.5 µl) of the DNA on an agarose gel using dilutions of known amounts of the 40kb control DNA as standard.



### 12.15.3 End repairing of purified size-fractionated genomic DNA

This step generates blunt-ended, 5'-phosphorylated DNA. The end-repair reaction can be scaled up or scaled down by the amount of DNA available. The set up of the end repair reaction is as follows below.

End repair reaction:

Purified sterile water	0 $\mu$ l
10x End-repair buffer (Epicentre)	8 $\mu$ l
dNTP mix (2.5 mM)	8 $\mu$ l
ATP (10 mM)	8 $\mu$ l
HMW genomic DNA (~1 $\mu$ g )	52 $\mu$ l
End-repair enzyme mix (Epicentre)	4 $\mu$ l
Final reaction volume	80 $\mu$ l

The reaction was incubated at the RT for 45 min. The end-repair enzyme mixture was heat inactivated for 10 min at 70 °C. End repaired genomic DNA was precipitated by addition of the following components directly to the 80  $\mu$ l of reaction.

Precipitation of end repaired genomic DNA:

End repair reaction volume	80 $\mu$ l
Purified sterile water	140 $\mu$ l
Sodium acetate (pH 5.0)	20 $\mu$ l
Isopropanol	120 $\mu$ l
Final reaction volume	340 $\mu$ l

Further, the sample was gently mixed via inversion and incubated for 30 min at RT. The DNA was precipitated via centrifugation at maximum speed (16,000 rpm), and the supernatant was removed carefully by pipetting. The pellet was washed by the addition of 70% ice cold ethanol and centrifuged for 5 min (max. speed). The supernatant was removed and the pellet was air-dried under a clean bench for 20-30 min. 20  $\mu$ l of TE buffer was added to the pellet, and the sample was incubated 1-2 h at 50 °C prior to quality and concentration check via gel electrophoresis.

### 12.15.4 Ligation reaction

In this step the CopyControl pCC1FOS vector provided by the kit was ligated to size-selected and end-repaired genomic DNA of the fungus *P. chrysogenum*. A 10:1 molar ratio of the CopyControl pCC1FOS vector to insert DNA was proven to be optimal.

In a new tube at RT, the following reagents were combined in the order listed and mixed thoroughly after each addition:

Purified sterile water	0 $\mu$ l
------------------------	-----------

10x Fast-Link Ligation Buffer (Epicentre)	1 $\mu$ l
ATP (10 mM)	1 $\mu$ l
CopyControl pCC1FOS Vector (0.5 $\mu$ g/ $\mu$ l)	0.5 $\mu$ l
Concentrated insert DNA (20 ng/ $\mu$ l; 0.13 $\mu$ g overall)	6.5 $\mu$ l
Fast-Link DNA Ligase (Epicentre)	1 $\mu$ l
Total reaction volume	10 $\mu$ l

The reaction was incubated at RT for 2 hours. Subsequently, the sample was incubated for 10 min at 70 °C in order to inactivate the Fast-Link DNA ligase (Epicentre). At this point the sample could be directly used for packaging reaction or stored at -20 °C prior to use.

### 12.15.5 *In vitro* packaging and titering the packaged fosmid clones

The day before performing the packaging reactions, a single colony of EPI300-T1R cells was inoculated into 50 ml of LB broth supplemented with 10 mM MgSO<sub>4</sub> and shaken overnight at 37°C. On the next day, 5 ml of the overnight culture was transferred into 50 ml of supplemented fresh LB broth and shaken at 37°C until the OD<sub>600</sub> reached 0.8-1.0. The cells were then stored at 4°C up to 72 hours.

10  $\mu$ l of the ligated fosmid DNA was pipetted into a tube containing 25  $\mu$ l of thawed MaxPlax Packaging (Epicentre Biotechnologies) extract and incubated at 30°C for 90 min. At the end of the incubation time, additional 25  $\mu$ l of thawed MaxPlax Packaging extract was added to the mixture, which was further incubated at 30°C for the additional 90 min. Subsequently, phage dilution buffer [10 mM Tris-HCl (pH 8.3), 100 mM NaCl, 10 mM MgCl<sub>2</sub>] was added to 1 ml volume. At the end, 25  $\mu$ l of chloroform were added at the top of the mixture and homogenised by gentle vortexing.

To determine the titer of the packaged fosmids, a 1:10 dilution of packaged fosmids was made by adding 90  $\mu$ l of phage dilution buffer to 10  $\mu$ l of originally packaged fosmids. 10  $\mu$ l of the 1:10 dilution was added to 100  $\mu$ l of prepared EPI300-T1R host cells at RT for 20 min and spread on LB-chloramphenicol selection plates at 37°C overnight. The colonies were counted and titer was calculated.

$$\frac{(\text{x of colonies}) (\text{dilution factor}) (1000 \mu\text{l/ml})}{(\text{volume of phage plated } (\mu\text{l}))} = \frac{(400 \text{ cfu}) (10) (1000 \mu\text{l/ml})}{(10 \mu\text{l})} = 400\,000 \text{ cfu/ml}$$

The total number of clones that could be obtained with the cosmid library was calculated by multiplying the titer (cfu/ml). The number of clones required to ensure that any given DNA sequence will be found in the cosmid library varies with the size of the genome.

$$N = \ln(1-P) / \ln(1-f)$$

Where P is the desired probability (expressed as a fraction); *f* is the proportion of the genome contained in a single clone; and N is the required number of cosmid clones. For example, the number of clones required to ensure a 99.9% probability of a given DNA sequence of *Penicillium chrysogenum* is contained within a cosmid library composed of 40 kb inserts:

$$N = \ln(1-0.999) / \ln(1 - (4 \times 10^4 \text{ bases} / 34 \times 10^6)) = -6.9077 / -0.0012 = 5750 \text{ number of clones}$$

Finally, a sufficient number (~5000) of clones were inoculated into 51 ninety-six-well plates with 100  $\mu$ l of LB medium containing chloramphenicol (12.5  $\mu$ g/ $\mu$ l). After shaking at 37°C overnight, 100  $\mu$ l of 40% glycerol was added to the 100  $\mu$ l culture of each clone for storing at -80°C.

### 12.16 Screening the genomic *P. chrysogenum* fosmid library via PCR

For screening the genomic library of *P. chrysogenum* via PCR, fosmid pools from genomic library were made in 50 ninety-six well plates. Large (15 cm) Petri dishes were made with LB agar medium supplemented with 12.5  $\mu$ g/ml chloramphenicol. Each of the ninety-six well plates from the library was replicated onto such an agar plate. The overnight cultures were grown at 37°C and on the next day fosmid clones were striped off from the agar with sterile inoculation loop and by the addition of 1-2 ml fresh liquid LB medium containing chloramphenicol at the same concentration as LB agar plate. 750  $\mu$ l of the fosmid clones suspension was mixed with an equal amount of 40% glycerol. Superpools were stored in a freezer at -80°C. Plasmid preparations were made from each superpool after overnight growth of LB liquid cultures at 37°C and 200 rpm. These superpool plasmid preparations were used for screening for PKS positive plates in the first round of PCR screening. One  $\mu$ l of diluted (1:2, 1:4) superpool plasmid DNA was used for screening according to the plasmid superpool PCR protocol (see 12.8.1). The primers used for screening of genomic library via the PCR method are presented in Table 27 (see under “function”).

These plate plasmid superpools were also used for screening library via the Southern hybridization method (see 12.17).

Lane pools were made for each PKS positive ninety-six-well plate by taking out 20  $\mu$ l from 12 fosmid glycerol stocks that belong to one lane. In that manner, eight pools were made per each ninety-six-well plate and were screened with whole-cell PCR protocol (1  $\mu$ l of cell suspension per PCR reaction). From each positive lane 10  $\mu$ l of each clone were aliquoted to a fresh PCR microfuge tube in order to make a lane pool. From each of lane pools, a 0.5-1  $\mu$ l

was used directly in the whole-cell PCR (see 12.8.1). PKS positive fosmids that corresponded to the same gene cluster were digested with the set of same restriction enzymes. One restriction pattern was chosen for subcloning of fragments into the pBluescript KS(-) vector. Subcloned fragments were end-sequenced via GATC Biotech (Konstanz, Germany). In some cases, for the purpose of gaining more sequence data, fragments were cut with the second restriction enzyme, subcloned into pBluescript vector and end-sequenced.

## **12.17 Screening of the genomic fosmid library via hybridization**

### **12.17.1 Southern hybridization**

#### ***Capillary transfer and fixation of DNA***

This method was adapted from Sambrook et al. [274]. As mentioned in 12.16, the prepared ninety-six well plate's plasmid superpools were used for screening the library via Southern hybridization method as well. All 50 plate's plasmid preparations were electrophoretically separated (12.9). After completion of electrophoresis, the agarose gel containing DNA samples was stained with ethidium bromide and photographed under UV light. Subsequently, a gel was immersed in 0.25 M HCl solution for two times 15 min at RT with constant gentle agitation until bromophenol blue indicator turned yellow. After washing the gel with purified water, it was soaked in denaturation solution for two times 15 min by gentle agitation. Afterwards, the gel was briefly rinsed with purified water and soaked into neutralization buffer two times for 20 min with gentle agitation. Just before transfer, the gel was soaked in transfer buffer (10-20x SSC; see 11.2.5) for several min by gentle agitation. A nylon membrane (Hybond-N+, Amersham) was cut in such a way that the membrane was approximately 1 mm larger than the gel. Two sheets of thick blotting paper were cut to the same size as the membrane. The membrane was floated on the surface of a dish of purified water until it was completely wet, and then immersed in 2x SSC for at least 5 min before use. The following items were put on a smooth, flat surface in the order listed:

1. A 3 mm Whatman paper saturated with transfer buffer (to serve as a bridge soaked on both sides into a transfer buffer).
2. Two 3 mm Whatman papers larger 2 cm larger than a gel from each side.
3. Nylon membrane prewet and saturated with transfer buffer.
4. Three pieces of Whatman papers same size as gel (or smaller) saturated with transfer buffer.
5. Paper towels

6. A light weight to keep all layers compressed.

The gel was surrounded with a plastic wrap or parafilm in order to prevent two layers of gel blotting papers getting in contact with each other. The transfer of DNA was usually proceeded within 12-48 h. The DNA was fixed to the membrane by UV-crosslinking (2 min, 120.000 microjoules per cm<sup>2</sup>).

***Labeling of the probes (according to ECL Direct-labeling and detection system, “Amersham Bioscience”)***

The DNA hybridization probes were derived from cloned PCR fragments that were cleaned from cloning vector used by enzymatic restriction (see 12.7). Alternatively, the purified PCR fragments (see 12.10) were directly used as hybridization probes. Prior to labelling the DNA concentration of samples was adjusted to 10 ng/μl. The amount of DNA for a single hybridization reaction (per a glass tube) was 600 ng (for Southern blottings; see previous subsection) and 300 ng (for colony hybridization; see 12.17.2). The DNA was denaturated for 5 min in the heating block at 99 °C. The samples were incubated on ice for 5 min. Furthermore, samples were shortly centrifuged in order to collect the content of the tubes. After that, for each 10 μl of the sample (100 ng DNA) the same volume of labelling reagent was added and gently mixed by pipetting. In the following step, glutaraldehyd solution was added in a volume equivalent to the volume of the sample (i.e. 1:1) and thoroughly mixed by pipetting. The samples were incubated at 37 °C for 10 min in a water bath and once more shortly centrifugated to collect the content of the tubes. The labelled probes were combined with the small aliquot of pre-warmed ECL hybridization buffer (warmed up to hybridization temperature) and added to the prehybridised membranes.

***Hybridization and stringency washing in the tubes (ECL Amersham)***

The ECL golden hybridization buffer (see 11.2.5) was preheated to 42 °C in a water bath. In a suitable container, nylon membranes were pre-wetted in 5xSSC (see 11.2.5). The 5xSSC was added to the glass tubes up to the 1/3 volume of the tube. The dry membranes were rolled up in order to ease their introduction into hybridization tubes. After introduction in the hybridization tubes membranes were gently un-rolled by addition of 5xSSC solution. Nylon membranes were washed in a hybridization oven for 5 min. The 5xSSC solution was decanted and an appropriate volume of hybridization buffer (0.0625-0.125 ml/cm<sup>2</sup>) was added to the tubes. The membranes were pre-hybridised in a hybridization oven for 30-60 min. In the meantime, as described before, nucleic acid probe was labelled according to ECL

protocol. After pre-hybridization, labelled probe was added to the small aliquot of pre-warmed hybridization buffer (42 °C) and mixed with the hybridization buffer in the tubes. The samples were hybridised overnight in a hybridization oven at 42 °C. On the next day, appropriate volume of primary wash buffer without urea (see 11.2.5) was pre-warmed to 42 °C. The hybridization buffer was discarded and replaced with the 50-100 ml of 5x SSC (not necessary to be pre-warmed) and the membranes were washed for 5 min. The 5x SSC solution was discarded and replaced with the primary wash buffer up to 1/3 volume of the tube (around 100 ml). The membranes were washed 1-2 times with primary wash buffer (each time 15-20 min). The rate of the primary wash buffer (0.1/0.5x SSC; see 11.2.5) was adjusted to the type of the probe used for hybridization (e.g. homologous or non-homologous DNA probes). After discarding the primary wash buffer the membranes were washed 1-2 times for 5 min with excess of the secondary wash buffer – 2x SSC (see 11.2.5) .

#### ***Detection of chemiluminescence with X-ray films***

After all wash out steps were completed the membranes were taken out from the hybridization glass tubes and the excess of the secondary washing buffer was drained off from the membranes. Furthermore, the membranes were put in a container with the 3MM Whatman paper soaked into fresh 2x SSC buffer (see 11.2.5). The membranes were kept on this paper until detection reagent was prepared. For detection reaction equal volumes (depending on the size of a blot) of ECL detection reagent 1 and ECL detection reagent 2 (provided by the manufacturer; composition not given) were mixed (0.125 ml/ cm<sup>2</sup> recommended), and the mixture was put on a membrane (side carrying the DNA up). The reaction was incubated for 1 min at RT. Excess of detection reagent was drained off from the blots, and they were wrapped into transparent plastic wrap by avoiding air pockets. The wrapped membranes were placed into film cassettes (see Table 28) and the membranes were then exposed to the autoradiography film (see Table 28) for a initial time of 10-30 min. After first evaluation of the intensity of the signal, the new film was put on the membranes for the next 1-6 h. The developing procedure was performed with Kodak GBX developer and fixer system (see Table 28). The films were soaked into developing reagent for 1-5 min, while the duration of the fixing procedure that followed was 2 min. Between these two steps, films were shortly rinsed with the distilled water. At the end of the detection procedure, films were washed in excess of water and dried out. Kodak GBX developer and fixer were prepared from stock solutions in accordance to the instructions of the manufacturer.

### 12.17.2 Colony hybridization

#### ***Induction of the fosmid clones of the *P. chrysogenum* genomic library (EPICENTRE, procedure adapted for 96-well plates)***

The fosmid clones were induced to a high copy number in order to obtain a large amount of DNA on the colony blots (i.e. nylon membranes). For that purpose 10 µl of chloramphenicol stock solution (25 mg/ml, in ethanol) were added to 20 ml of LB liquid medium. Sterile 96-well microtiter plates were prepared, and 200 µl of LB medium (12.5 µg/ml chloramphenicol) were added to each vial. The plates were replicated with a sterilised 96-replicator and shaken at 100 rpm overnight at 37 °C. The aeration (i.e. shaking) was the crucial factor for receive a high copy number of plasmids per clone. On the next day chloramphenicol was added to 10 ml of fresh LB liquid medium to a final concentration of 12.5 µg/ml. The 20 µl of 1000x Copy Control Induction solution (provided by the manufacturer; see Table 30) was added to the LB medium (amount calculated for 96 clones). The 100 µl of overnight cultures from microtiter plate were added to the corresponding clone positions of the fresh 96-well plate by use of a multipipetor (see Table 28). The microtiter plates were incubated at 37 °C with 100 rpm for at least 5 h. The high copy number induced plates were replicated on a nylon membrane immediately or were frozen at -80 °C.

#### ***Lysing colonies and binding of DNA to the filters (Sambrook et al., 2001) [274]***

Four pieces of Whatman 3MM paper were cut to a size to fit bottoms of four plastic trays. Papers were saturated with one of the following solutions:

1. 10% SDS (3 min)
2. denaturizing solution (5 min)
3. neutralizing solution (5 min)
4. 2xSSC (5 min)

A blunt-ended forceps was used to peel the nylon membranes from the agar plates. Further on, the membranes were placed with the colony side up first on a SDS-impregnated 3MM paper and then subjected to the order of solutions and incubation times as mentioned before. The membranes were always transferred in the same order in which they were removed from their agar plates. After exposure of membranes to the final solution (i.e. 2X SSC), the membranes were dried on a fresh Whatman 3MM paper for about 20-30 min. The DNA was then fixed to the membranes by UV crosslinking (2 min, 120,000 microjoules per cm<sup>2</sup>). Optionally, cell debris was removed via submerging the membranes into washing solution (2x SSC, 0.1% SDS) and mechanical removal of colony debris with cotton pads. The

membranes were transferred to a Whatman 3MM paper freshly saturated with 2xSSC for 2 min. Afterwards, the membranes were dried on a clean and dry 3MM Whatman paper for about 20-30 min and sandwiched between papers until future use. The membranes were then hybridised according to the ECL direct nucleic acid labelling and detection systems (Amersham Biosciences; Braunschweig, Germany) presented before in section 12.17. Optionally, for the colony hybridization, the hybridization time was reduced from overnight (12-16 h) to just 4-5 h in order to reduce the colony background that may appear as a result of unspecific binding of the labelled DNA probe to the nylon membrane.

### **12.17.3 Dot blot hybridization**

Optionally, the plasmid microtiter plate (MP) pools preparations were directly used to check for the presence of homology with PKS derived DNA probes. The plasmid DNA was directly applied to a nylon membrane (Hybond-N+, Amersham) in steps of 1 µl, thus allowing DNA sample to dry between two applications. In total, 2-5 µl of each plasmid MP superpool preparation were applied onto nylon membrane. The DNA was fixed according to the protocol used for lysing and fixation of colony blots (see 12.17.1). Hybridization and detection procedures were performed in accordance to the manufacture protocol of the ECL direct nucleic acid labelling and detection systems (Amersham Biosciences; Braunschweig, Germany) as described before.

## **12.18 Automated DNA sequencing**

All samples within this study were sequenced via the service of GATC Biotech (Konstanz, Germany). The DNA sequencing was performed according to the dideoxy-mediated chain termination method (Sanger et al., 1977) [284]. This cycle sequencing technique combines *in vitro* amplification of the target DNA via polymerase chain reaction with specific base chain termination.

### **12.18.1 Sequencing of cloned PCR products and restriction fragments**

PCR products or restriction fragments were routinely cloned into the pGEM-T Easy (Promega; Mannheim, Germany) or pBluescript KS (-)(see 12.13). 30 µl of each plasmid preparation (100-200 ng DNA/µl) was sent to GATC Biotech (Konstanz, Germany) for sequencing with T7 or/and T3 primers. The sequences were further analyzed as outlined in 12.19 section (i.e. *in silico* sequence analysis).



### **12.18.2 Shotgun sequencing of fosmid DNA**

The shotgun sequence assembly approach is based on building up the master sequence directly from the short sequences obtained from individual sequencing experiments, simply by examining the sequences for overlaps (Brown 2002) [197].

The plasmid DNA of fosmid 42H12, estimated to contain approx. 40 kb of genomic fragment from *P. chrysogenum* E01-10/3, was sent for complete sequencing by the shotgun approach. The first step was a breakage of the plasmid DNA into fragments by sonication, a technique that uses high-frequency sound waves to make random cuts in DNA molecules. Following this procedure, the fragments were electrophoresed and those fragments in the range 1.5–2.0 kb were removed from the agarose gel (Brown 2002) [197]. The cloning of the shotgun fragments was performed with the vector system pCR4 blunt Topo<sup>®</sup>. For the sequencing reactions the primers M13-RP and M13-FP were used. All clones from one 384-well plate were sequenced in forward and reverse direction. After evaluation of all sequences, 625 trimmed sequences were left. With those sequences the total amount of 503,520 bases was reached representing 10.4 times of the sequenced fosmid insert and thus enabling the generation of a single contig of cloned *P.chrysogenum* fosmid 42H12 without any finishing step.

### **12.19 In silico sequence analysis**

The DNA and protein sequences were analyzed by the Lasergene 7 software package (DNASTAR, Inc, USA). This software, which comprises several modules, is suitable for alignment of sequences, contigs assembly, primers design, and restriction mapping. The module “EditSeq” allows translating, back-translating and reverse complementing of sequences. For assembling sequences and managing contigs the module “SeqMan” was applied. “PrimerSelect” is a module applied for designing of primers or comparison of self-designed primers against a sequence template. The restriction maps of plasmids were created with the help of the module “MapDraw”. Prior to analysis, all sequences were checked for cloning vector sequences by “VecScreen” system at the NCBI homepage. The algorithm named “FinchTV” (Geospiza web site) served as graphic viewer for DNA chromatogram files. The “BioEdit Version 7.0.0” was used for the alignment of nucleic acids and protein sequences (available at “Brown Lab” web site). Alternatively, “ClustalX Version 1.8” was used for the same purpose of sequence alignment (“Clustal” web site).

For the identification of open reading frames, two internet programs were used: “Fgenesh” that is specialised for eukaryotic gene recognition, and “FramePlot 2.3.2.” for bacterial gene

recognition. The homology searches of nucleotide and protein sequences were performed online ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)) with the help of the “BLAST” program (“BLASTX”, “BLASTP” or “PSI-BLAST” modules) at the National Center for Biotechnology Information (NCBI, USA). Besides, the “Advanced BLAST” program at EMBnet (European Molecular Biology network) web site was used with its extended features for sequence homology search. The protein domains within the deduced amino acid sequence were identified by “InterProScan” search algorithm at EMBL-EBI (European Molecular Biology Laboratory-European Bioinformatics Institute) or by search algorithm within the “Pfam” database (Sanger Institute homepage). Specifically, PKS domains were analyzed by the software “SEARCHPKS” at the homepage of the National Institute of Immunology - New Delhi, India. The “MEGA 4.0” internet-software (MEGA web site) was used to infer the phylogenetic trees from cloned PKS sequences. Furthermore, the freeware “TreeView Version 1.6.6” was used to display generated phylogenetic trees (Taxonomy and Systematics web site of the Glasgow University). For details and sources (i.e. web addresses) for used programs see Table 31.

### **12.20 Detection of sorbicillactone A from *P. chrysogenum* liquid cultures**

The malt extract agar medium plates were inoculated with 1 µl of frozen *P. chrysogenum* spores. The agar cultures were grown 5-7 days at 29 °C before a small piece of mycelium was transferred to 300 ml of malt extract liquid medium which was adjusted to pH 6.0 – 6.5 proved to be optimal for the sorbicillacton A production (Bringmann et al 2002). After 10-14 days of growth at 29 °C the mycelium started to become yellow indicating the production of sorbicillacton. The liquid culture medium was collected via pipetting, and the pH was adjusted to slightly acidic dropwise with 1 M HCL. The liquid medium was then extracted with ethyl acetate in a ratio of 2:1. The extraction was repeated two more times in the same manner before upper ethyl acetate yellowish phase was collected into a fresh flask. This phase was than filtrated through round filter papers (150 mm, Schleicher & Schüll) and dried in an evaporator for 10-15 min at 4 °C. Before HPLC both sorbicillactone A standard and culture extract were dissolved in acetonitrile and then subjected to HPLC with the following parameters (Bringmann et al. 2002):

Type of HPLC:	analytical
Column:	PerfectSil Target ODS-3-C18 250 x 4.0 mm x 5 µm
Eluent:	acetonitrile + 0.05% TFA, Water + 0.05% TFA
Gradient:	from 10% acetonitrile to 90% acetonitrile in 30 min
Flow:	1 ml/min

---

Detection: 254 nm

---

The sorbicillactone A standard were eluted between 20 and 21 min. The same peak with the same absorption characteristics ( $\lambda_{\text{sub.max}}$  215, 271, 379 nm) was detected at the culture extract chromatogram.

## REFERENCES

- [1] Keller NP, Turner G, Bennett JW. Fungal secondary metabolism - From biochemistry to genomics. *Nat Rev Microbiol* 2005;3(12):937-47.
- [2] Calvo AM, Wilson RA, Bok JW, Keller NP. Relationship between secondary metabolism and fungal development. *Microbiol Mol Biol Rev* 2002;66(3):447-59, table of contents.
- [3] Yu JH, Keller N. Regulation of secondary metabolism in filamentous fungi. *Annu Rev Phytopathol* 2005;43:437-58.
- [4] Fleming A. On the bacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *Br J Exp Pathol* 1929;10:226-36.
- [5] Bennett J. Alexander Fleming and the discovery of penicilin. *Adv Appl Microbiol* 2001;49:163-84.
- [6] Pelaez F. Biological activities of fungal metabolites. *Handbook of Industrial Mycology* (ed. An, Z) 2005;(Marcel Dekker, New York):49-92.
- [7] da Rocha AB, Lopes RM, Schwartzmann G. Natural products in anticancer therapy. *Curr Opin Pharmacol* 2001;1(4):364-9.
- [8] Cragg GM, Newman DJ, Snader KM. Natural products in drug discovery and development. *J Nat Prod* 1997;60(1):52-60.
- [9] Bugni TS, Ireland CM. Marine-derived fungi: a chemically and biologically diverse group of microorganisms. *Nat Prod Rep* 2004;21(1):143-63.
- [10] Demain AL. Pharmaceutically active secondary metabolites of microorganisms. *Appl Microbiol Biotechnol* 1999;52(4):455-63.
- [11] Demain AL. From natural products discovery to commercialization: a success story. *J Ind Microbiol Biotechnol* 2006;33(7):486-95.
- [12] Bentley R. Microbial secondary metabolites play important roles in medicine; prospects for discovery of new drugs. *Perspect Biol Med* 1997;40(3):364-94.
- [13] Saleem M, Ali MS, Hussain S, Jabbar A, Ashraf M, Lee YS. Marine natural products of fungal origin. *Nat Prod Rep* 2007;24(5):1142-52.
- [14] Blunt JW, Copp BR, HU W-P, Munro MH, Northcote PT, Prinsep MR. Marine natural products. *Nat Prod Rep* 2008;25(1):35-94.
- [15] Bennett JW. Mycotoxins, mycotoxicoses, mycotoxicology and Mycopathologia. *Mycopathologia* 1987;100(1):3-5.
- [16] ICMSF. Microorganisms in foods 5: microbiological specification of food pathogens. London, UK: Blackie Academic and Professional, 1996.
- [17] Abarca ML, Bragulat MR, Castella G, Cabanes FJ. Ochratoxin A production by strains of *Aspergillus niger* var. *niger*. *Appl Environ Microbiol* 1994;60(7):2650-2.
- [18] Bhat RV, Shetty PH, Amruth RP, Sudershan RV. A foodborne disease outbreak due to the consumption of moldy sorghum and maize containing fumonisin mycotoxins. *J Toxicol Clin Toxicol* 1997;35(3):249-55.
- [19] Blanc PJ, Loret MO, Goma G. Production of citrinin by various species of *Monascus*. *Biotechnol Lett* 1995;17:291-94.
- [20] Chu FS. Studies on ochratoxins. *CRC Crit Rev Toxicol* 1974;2(4):499-524.
- [21] Dutton MF. Fumonisin, mycotoxins of increasing importance: their nature and their effects. *Pharmacol Ther* 1996;70(2):137-61.
- [22] Bennett JW, Klich M. Mycotoxins. *Clin Microbiol Rev* 2003;16(3):497-516.
- [23] Cox R, Glod F. Fungal polyketide synthases in the information age. In: Tkacz JS, Lange L, editors. *Advances in fungal biotechnology for industry, agriculture and medicine* New York

Kluwer Academic/Plenum Publishers, 2004: 69-96.

[24] IARC. Aflatoxins: naturally occurring aflatoxins (Group 1), aflatoxins M1 (Group 2B). *Int Agency Res Cancer* 1993 (a);56:245.

[25] Goto T, Wicklow DT, Ito Y. Aflatoxin and cyclopiazonic acid production by a sclerotium-producing *Aspergillus tamaris* strain. *Appl Environ Microbiol* 1996;62(11):4036-8.

[26] Klich MA, Mullaney EJ, Daly CB, Cary JW. Molecular and physiological aspects of aflatoxin and sterigmatocystin biosynthesis by *Aspergillus tamaris* and *A. ochraceoroseus*. *Appl Microbiol Biotechnol* 2000;53(5):605-9.

[27] Yu JH, Leonard TJ. Sterigmatocystin biosynthesis in *Aspergillus nidulans* requires a novel type I polyketide synthase. *J Bacteriol* 1995;177(16):4792-800.

[28] Chang PK, Cary JW, Yu J, Bhatnagar D, Cleveland TE. The *Aspergillus parasiticus* polyketide synthase gene pksA, a homolog of *Aspergillus nidulans* wA, is required for aflatoxin B1 biosynthesis. *Mol Gen Genet* 1995;248(3):270-7.

[29] Feng GH, Leonard TJ. Characterization of the polyketide synthase gene (pksL1) required for aflatoxin biosynthesis in *Aspergillus parasiticus*. *J Bacteriol* 1995;177(21):6246-54.

[30] Yu J, Chang PK, Ehrlich KC, Cary JW, Bhatnagar D, Cleveland TE, et al. Clustered pathway genes in aflatoxin biosynthesis. *Appl Environ Microbiol* 2004;70(3):1253-62.

[31] Sabater-Vilar M, Maas RF, Fink-Gremmels J. Mutagenicity of commercial *Monascus* fermentation products and the role of citrinin contamination. *Mutat Res* 1999;444(1):7-16.

[32] Scott PM. Other mycotoxins (Chapter 17). In: Mangan N, Olsen M, editors. *Mycotoxins in food - Detection and control*, Woodhead Publishing in Food Science and Technology, 2004: 406-40.

[33] Shimizu T, Kinoshita H, Ishihara S, Sakai K, Nagai S, Nihira T. Polyketide synthase gene responsible for citrinin biosynthesis in *Monascus purpureus*. *Appl Environ Microbiol* 2005;71(7):3453-57.

[34] Shimizu T, Kinoshita H, Nihira T. Identification and in vivo functional analysis by gene disruption of ctnA, an activator gene involved in citrinin biosynthesis in *Monascus purpureus*. *Appl Environ Microbiol* 2007;73(16):5097-103.

[35] Sweeney MJ, Dobson ADW. Molecular biology of mycotoxin production. *FEMS Microbiol Lett* 1999;175:149-63.

[36] Rheeder JP, Marasas WF, Vismer HF. Production of fumonisins analogs by *Fusarium* species. *Appl Environ Microbiol* 2002;68:2102-05.

[37] Marasas WF. Fumonisins: their implications for human and animal health. *Nat Toxins* 1995;3(4):193-8; discussion 221.

[38] Merrill AH, Jr., Sullards MC, Wang E, Voss KA, Riley RT. Sphingolipid metabolism: roles in signal transduction and disruption by fumonisins. *Environ Health Perspect* 2001;109 Suppl 2:283-9.

[39] Wang E, Norred WP, Bacon CW, Riley RT, Merrill AH, Jr. Inhibition of sphingolipid biosynthesis by fumonisins. Implications for diseases associated with *Fusarium moniliforme*. *J Biol Chem* 1991;266(22):14486-90.

[40] Sydenham EW, Shephard GS, Thiel PG, Marasas WFO, Stockenstrom S. Fumonisin contamination of commercial corn-based human foodstuffs. *J Agric Food Chem* 1991;39:235-44.

[41] Proctor RH, Brown DW, Plattner RD, Desjardins AE. Co-expression of 15 contiguous genes delineates a fumonisin biosynthetic gene cluster in *Gibberella moniliformis*. *Fungal Genet Biol* 2003;38(2):237-49.

- [42] Proctor RH, Desjardins AE, Plattner RD, Hohn TM. A polyketide synthase gene required for biosynthesis of fumonisin mycotoxins in *Gibberella fujikuroi* mating population A. *Fungal Genet Biol* 1999;27(1):100-12.
- [43] Murphy PA, Hendrich S, Landgren C, Bryant CM. Food mycotoxins: an update (Scientific Status Summary). *J Food Sci* 2006;71(51-65).
- [44] Dombink-Kurtzman MA. The sequence of the isoeopoxydon dehydrogenase gene of the patulin biosynthetic pathway in *Penicillium* species. *Antonie Leeuwenhoek* 2007;91(2):179-89.
- [45] Beck J, Ripka S, Siegner A, Schiltz E, Schweiser E. The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide synthases. *Eur J Biochem* 1990;192(2):487-98.
- [46] Fujii I, Ono Y, Tada H, Gomi K, Ebizuka Y, Sankawa U. Cloning of the polyketide synthase gene *atX* from *Aspergillus terreus* and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. *Mol Gen Genet* 1996;253(1-2):1-10.
- [47] Lu P, Zhang A, Dennis LM, Dahl-Roshak AM, Xia YQ, Arison B, et al. A gene (*pks2*) encoding a putative 6-methylsalicylic acid synthase from *Glarea lozoyensis*. *Mol Genet Genomics* 2005;273(2):207-16.
- [48] White S, O'Callaghan J, Dobson AD. Cloning and molecular characterization of *Penicillium expansum* genes upregulated under conditions permissive for patulin biosynthesis. *FEMS Microbiol Lett* 2006;255(1):17-26.
- [49] NCBI. National Center for Biotechnology Information. In: <http://www.ncbi.nlm.nih.gov/Genbank>.
- [50] Cox RJ. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org Biomol Chem* 2007;5(13):2010-26.
- [51] Bayman P, Baker JL, Doster MA, Michailides TJ, Mahoney NE. Ochratoxin production by the *Aspergillus ochraceus* group and *Aspergillus alliaceus*. *Appl Environ Microbiol* 2002;68(5):2326-9.
- [52] Pitt JI. *Penicillium viridicatum*, *Penicillium verrucosum*, and production of ochratoxin A. *Appl Environ Microbiol* 1987;53(2):266-9.
- [53] Kuiper-Goodman T, Scott PM. Risk assessment of the mycotoxin ochratoxin A. *Biomed Environ Sci* 1989;2(3):179-248.
- [54] O'Callaghan J, Caddick MX, Dobson AD. A polyketide synthase gene required for ochratoxin A biosynthesis in *Aspergillus ochraceus*. *Microbiology* 2003;149(Pt 12):3485-91.
- [55] Karolewicz A, Geisen R. Cloning a part of the ochratoxin A biosynthetic gene cluster of *Penicillium nordicum* and characterization of the ochratoxin polyketide synthase gene. *Syst Appl Microbiol* 2005;28(7):588-95.
- [56] Paradkar AS, Jensen SE, Mosher RH. Comparative genetics and molecular biology of  $\beta$ -lactam biosynthesis. In: Strohl WR, editor. *Biotechnology of antibiotics*. 2nd ed. Rahway, New Jersey, Merck Research Labs, 1997: 241-79.
- [57] Bycroft BW, Shute RE. Chemistry and biosynthesis of penicillins and cephalosporins. In: Peberdy JF, editor. *Penicillium and Acremonium*, Springer, 1987: 113-26.
- [58] Rosario NA, Grumach AS. Allergy to beta-lactams in pediatrics: a practical approach. *J Pediatr (Rio J)* 2006;82(5 Suppl):S181-8.
- [59] Weltzien HU, Padovan E. Molecular features of penicillin allergy. *J Invest Dermatol* 1998;110(3):203-6.
- [60] Martin JF. New aspects of genes and enzymes for beta-lactam antibiotic biosynthesis. *Appl Microbiol Biotechnol* 1998;50(1):1-15.
- [61] Walton JD, Panaccione DG, Hallen HE. Peptide synthesis without ribosomes. In: Tkacz JS, Lange L, editors. *Advances in fungal biotechnology for industry, agriculture and medicine*. New York, Kluwer Academic/Plenum Publishers, 2004: 127-62.

- [62] Gutierrez S, Fierro F, Casqueiro J, Martin JF. Gene organization and plasticity of the beta-lactam genes in different filamentous fungi. *Antonie Van Leeuwenhoek* 1999;75(1-2):81-94.
- [63] Laich F, Fierro F, Martin JF. Production of penicillin by fungi growing on food products: identification of a complete penicillin gene cluster in *Penicillium griseofulvum* and a truncated cluster in *Penicillium verrucosum*. *Appl Environ Microbiol* 2002;68(3):1211-9.
- [64] Panaccione DG, Coyle CM. Abundant respirable ergot alkaloids from the common airborne fungus *Aspergillus fumigatus*. *Appl Environ Microbiol* 2005;71(6):3106-11.
- [65] Coyle CM, Panaccione DG. An ergot alkaloid biosynthesis gene and clustered hypothetical genes from *Aspergillus fumigatus*. *Appl Environ Microbiol* 2005;71(6):3112-8.
- [66] Lorenz N, Wilson EV, Machado C, Schardl CL, Tudzynski P. Comparison of ergot alkaloid biosynthesis gene clusters in *Claviceps* species indicates loss of late pathway steps in evolution of *C. fusiformis*. *Appl Environ Microbiol* 2007;73(22):7185-91.
- [67] Correia T, Grammel N, Ortel I, Keller U, Tudzynski P. Molecular cloning and analysis of the ergopeptine assembly system in the ergot fungus *Claviceps purpurea*. *Chem Biol* 2003;10(12):1281-92.
- [68] Haarmann T, Ortel I, Tudzynski P, Keller U. Identification of the cytochrome P450 monooxygenase that bridges the clavine and ergoline alkaloid pathways. *Chembiochem* 2006;7(4):645-52.
- [69] Tsai HF, Wang H, Gebler JC, Poulter CD, Schardl CL. The *Claviceps purpurea* gene encoding dimethylallyltryptophan synthase, the committed step for ergot alkaloid biosynthesis. *Biochem Biophys Res Commun* 1995;216(1):119-25.
- [70] Haarmann T, Machado C, Lubbe Y, Correia T, Schardl CL, Panaccione DG, et al. The ergot alkaloid gene cluster in *Claviceps purpurea*: extension of the cluster sequence and intra species evolution. *Phytochemistry* 2005;66(11):1312-20.
- [71] Tudzynski P, Holter K, Correia T, Arntz C, Grammel N, Keller U. Evidence for an ergot alkaloid gene cluster in *Claviceps purpurea*. *Mol Gen Genet* 1999;261(1):133-41.
- [72] Dewick PN. The acetate pathway: fatty acids and polyketides. In: Dewick PN, editor. *Medicinal natural products: A biosynthetic approach*. 2nd ed. West Sussex, England, John Wiley & Sons Ltd, 2001: 35-117.
- [73] Lysoe E, Klemsdal SS, Bone KR, Frandsen RJ, Johansen T, Thrane U, et al. The *PKS4* gene of *Fusarium graminearum* is essential for zearalenone production. *Appl Environ Microbiol* 2006;72(6):3924-32.
- [74] Desjardins AE, Proctor RH. *Biochemistry and genetics of Fusarium toxins*. St. Paul, Minn: American Phytopathological Society Press, 2001.
- [75] Shier WT, Shier AC, Xie W, Mirocha CJ. Structure-activity relationships for human estrogenic activity in zearalenone mycotoxins. *Toxicol* 2001;39(9):1435-8.
- [76] Dai S, Duan J, Lu Y, Zhang Y, Cheng J, Ren J, et al. Phytoestrogen  $\alpha$ -zearalenol inhibits atherogenesis and improves lipid profile in ovariectomised cholesterol-fed rabbits. *Endocrine* 2004;25(12):121-9.
- [77] Stopper H, Schmitt E, Kobras K. Genotoxicity of phytoestrogens. *Mutat Res* 2005;574(1-2):139-55.
- [78] Hodge EG, Hidy PH, Wehrmeister HJ, inventors; Estrogenic compounds and animal growth promoters. U.S. Patent 3239345. 1966.
- [79] Hagler WM, Jr. N. R. Towers, C. J. Mirocha, R. M. Eppley, Bryden WL. Zearalenone: mycotoxin or mycoestrogen? In: B. A. Summerell, J. F. Leslie, D. Backhouse, W. L. Bryden, Burgess LW, editors. *Fusarium*. Paul E. Nelson Memorial Symposium; 2001; St. Paul, Minn.: APS Press; 2001.
- [80] Utian WH. Comparative trial of P1496, a new non-steroidal oestrogen analogue. *Br Med J* 1973;1(5853):579-81.

- [81] Hidy PH, Baldwin RS, Greasham RL, Keith CL, McMullen JR. Zearalenone and some derivatives: production and biological activities. *Adv Appl Microbiol* 1977;22:59-82.
- [82] Gaffoor I, Trail F. Characterization of two polyketide synthase genes involved in zearalenone biosynthesis in *Gibberella zeae*. *Appl Environ Microbiol* 2006;72(3):1793-9.
- [83] Borel JF, Feurer C, Gubler HU, Stahelin H. Biological effects of cyclosporin A: a new antilymphocytic agent. *Agents Actions* 1976;6(4):468-75.
- [84] Faulds D, Goa KL, Benfield P. Cyclosporin. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic use in immunoregulatory disorders. *Drugs* 1993;45(6):953-1040.
- [85] Weber G, Leitner E. Disruption of the cyclosporin synthetase gene of *Tolypocladium niveum*. *Curr Genet* 1994;26(5-6):461-7.
- [86] Bookstein R, Lai CC, To H, Lee WH. PCR-based detection of a polymorphic BamHI site in intron 1 of the human retinoblastoma (RB) gene. *Nucleic Acids Res* 1990;18(6):1666.
- [87] Royer JC, Madden KT, C. NT, LoBuglio KF. *Penicillium* Genomics. In: Khachatouraus GG, Arora DK, editors., 2003: 285-95.
- [88] Moore RN, Bigam G, Chan JK, Hogg AM, Nakashima TT, Vederas JC. Biosynthesis of the Hypocholesterolemic Agent Mevinolin by *Aspergillus terreus*. Determination of the Origin of Carbon, Hydrogen, and Oxygen Atoms by <sup>13</sup>C NMR and Mass Spectrometry. *J Am Chem Soc* 1985;107:3694-701.
- [89] Fujii I, Watanabe A, Ebizuka Y. More functions for multifunctional polyketide synthases. In: Tkacz JS, Lange L, editors. New York, Kluwer Academic/Plenum Publishers, 2004: 97-125.
- [90] Hendrickson L, Davis CR, Roach C, Nguyen DK, Aldrich T, McAda PC, et al. Lovastatin biosynthesis in *Aspergillus terreus*: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene. *Chem Biol* 1999;6(7):429-39.
- [91] Kennedy J, Auclair K, Kendrew SG, Park C, Vederas JC, Hutchinson CR. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* 1999;284(5418):1368-72.
- [92] Abe Y, Suzuki T, Ono C, Iwamoto K, Hosobuchi M, Yoshikawa H. Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in *Penicillium citrinum*. *Mol Genet Genomics* 2002;267(5):636-46.
- [93] McAlpine J. Unnatural natural products by genetic manipulation. In: Sapienza DM, Savage LM, editors. Natural products II: New technologies to increase efficiency and speed. Southborough, Mass, International Business Communications, 1998: 251-78.
- [94] Peng YL, Demain AL. A new hydroxylase system in *Actinomadura* sp cells converting compactin to pravastatin. *J Industr Microbiol Biotechnol* 1998;20:373-75.
- [95] Jensen PR FW. *Drugs from the Sea*. Basel: Karger 2000.
- [96] Haefner B. Drugs from the deep: marine natural products as drug candidates. *Drug Discov Today* 2003;8(12):536-44.
- [97] Kohlmeyer J, Volkmann-Kohlmeyer B, Newell SY. *Marine and estaurine mycelial Eumycota and Oomycota*. Amsterdam: Elsevier, 2004.
- [98] Torsvik V, Salte K, Sorheim R, Goksoyr J. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. *Appl Environ Microbiol* 1990;56(3):776-81.
- [99] Mayer KM, Ford J, Macpherson GR. Exploring the diversity of marine-derived fungal polyketide synthases. *Can J Microbiol* 2007;53:291-302.
- [100] Bringmann G, Lang G, Mühlbacher J, Schaumann K, Steffens S, Rytik PG, et al. Sorbicillactone A, a structurally unprecedented bioactive novel-type alkaloid from a sponge-derived fungus. In: Müller WEG, editor. *Marine Molecular Biotechnology*. Berlin-Heidelberg, Springer-Press, 2003: 231-53.



- [101] Sun Y, Tian L, Huang J, Ma HY, Zheng Z, Lv AL, et al. Trichodermatides A-D, novel polyketides from the marine-derived fungus *Trichoderma reesei*. *Org Lett* 2008;10(3):393-6.
- [102] Zhang D, Li X, Kang JS, Choi HD, Jung JH, Son BW. Redoxcitinin, a biogenetic precursor of citrinin from marine isolate of fungus *Penicillium sp.* *J Microbiol Biotechnol* 2007;17(5):865-7.
- [103] Kasai Y, Komatsu K, Shigemori H, Tsuda M, Mikami Y, Kobayashi J. Cladionol A, a polyketide glycoside from marine-derived fungus *Gliocladium* species. *J Nat Prod* 2005;68(5):777-9.
- [104] Wei H, Itoh T, Kinoshita M, Kotoku N, Aoki S, Kobayashi M. Shimalactone A, a novel polyketide, from marine-derived fungus *Emericella variegata* GF10. *Tetrahedron* 2005;61:8054-58.
- [105] Liu WZ, Gu Q, Zhu W, Cui CB, Fan GT, Zhu T, et al. Penicillones A and B, two novel polyketides with tricyclo [5.3.1.0<sup>3,8</sup>] undecane skeleton, from a marine-derived fungus *Penicillium terrestre*. *Tetrahedron Letters* 2005;46:4993-96.
- [106] Shigemori H, Kasai Y, Komatsu K, Tsuda M, Mikami Y, Kobayashi H. Sporiolides A and B, new cytotoxic twelve-membered macrolides from a marine-derived fungus *Cladosporium* species. *Mar. Drugs* 2004;2:164-69.
- [107] Tsukamoto S, Miura S, Yamashita Y, Ohta T. Aspermytin A: a new neurotrophic polyketide isolated from a marine-derived fungus of the genus *Aspergillus*. *Bioorg Med Chem Lett* 2004;14(2):417-20.
- [108] Liu WZ, Gu QQ, Zhu WM, Cui CB, Fan GT. Two new benzoquinone derivatives and two new bisorbicillinoids were isolated from a marine-derived fungus *Penicillium terrestre*. *Journal of Antibiotics* 2005;58(7):441-46.
- [109] Liu Z, Jensen PR, Fenical W. A cyclic carbonate and related polyketides from a marine-derived fungus of the genus *Phoma*. *Phytochemistry* 2003;64(2):571-4.
- [110] Abdel-Lateff A, Fisch KM, Wright AD, König GM. A new antioxidant isobenzofuranone derivative from the algicolous marine fungus *Epicoccum sp.* *Planta Med* 2003;69(9):831-4.
- [111] Bringmann G, Lang G, Steffens S, Gunther E, Schaumann K. Evariquinone, isoemicellin, and stromemycin from a sponge derived strain of the fungus *Emericella variegata*. *Phytochemistry* 2003;63(4):437-43.
- [112] Abdel-Lateff A, Klemke C, König GM, Wright AD. Two new xanthone derivatives from the algicolous marine fungus *Wardomyces anomalus*. *J Nat Prod* 2003;66(5):706-8.
- [113] Namikoshi M, Negishi R, Nagai H, Dmitrenok A, Kobayashi H. Three new chlorine containing antibiotics from a marine-derived fungus *Aspergillus ostianus* collected in Pohnpei. *J Antibiot (Tokyo)* 2003;56(9):755-61.
- [114] Abdel-Lateff A, König GM, Fisch KM, Holler U, Jones PG, Wright AD. New antioxidant hydroquinone derivatives from the algicolous marine fungus *Acremonium sp.* *J Nat Prod* 2002;65(11):1605-11.
- [115] Son BW, Choi JS, Kim JC, Nam KW, Kim DS, Chung HY, et al. Parasitenone, a new epoxy-cyclohexenone related to gabosine from the marine-derived fungus *Aspergillus parasiticus*. *Journal of Natural Products* 2002;65:794-95.
- [116] Malmstrom J, Christophersen C, Barrero AF, Oltra JE, Justicia J, Rosales A. Bioactive metabolites from a marine-derived strain of the fungus *Emericella variegata*. *J Nat Prod* 2002;65(3):364-7.
- [117] Namikoshi M, Kobayashi H, Yoshimoto T, Meguro S, Akano K. Isolation and characterization of bioactive metabolites from marine-derived filamentous fungi collected from tropical and sub-tropical coral reefs. *Chem Pharm Bull* 2000;48(10):1452-7.

- [118] Komatsu K, Shigemori H, Mikami Y, Kobayashi J. Sculezonones A and B, two metabolites possessing a phenalenone skeleton from a marine-derived fungus *Penicillium* species. *J Nat Prod* 2000;63(3):408-9.
- [119] Numata A, Takahashi C, Ito Y, Minoura K, Yamada T, Matsuda C, et al. Penochalasin, a novel class of cytotoxic cytochalasins from a *Penicillium* species separated from a marine alga: structure determination and solution conformation. *J Chem Soc Perkin Trans I* 1996:239-45.
- [120] Abrell LM, Borgeson B, Crews P. A new polyketide, secocurvularin, from the salt water culture of a sponge derived fungus. *Tetrahedron Lett* 1996;37:8983-84.
- [121] Kakeya H, Takahashi I, Okada G, Isono K, Osada H. Epolactaene, a novel neurotogenic compound in human neuroblastoma cells, produced by a marine fungus. *J Antibiot (Tokyo)* 1995;48(7):733-5.
- [122] Poch G, Gloer J. Obionin A: a new polyketide metabolite from the marine fungus *Leptosphaeria obiones*. *Tetrahedron Lett* 1989;30:3483-86.
- [123] Frisvad JC, and Samson RA. Polyphasic taxonomy of *Penicillium* subgenus *Penicillium*: A guide to identification of food and air-borne terverticillate *Penicillia* and their mycotoxins. In: Samson RA, and Frisvad JC, editors. *Penicillium* subgenus *Penicillium*: new taxonomic schemes, mycotoxins and other extrolites. Utrecht, The Netherlands, Centraalbureau voor Schimmelcultures, 2004.
- [124] Pitt JI. A laboratory guide to common *Penicillium* species. North Ryde: Food Science Australia, 2000.
- [125] Clutterbuck PW, Lovell R, Raistrick H. Studies in the biochemistry of microorganisms: The formation from glucose by members of the *Penicillium chrysogenum* series of a pigment, an alkali-soluble protein and penicillin - the antibacterial substance of Fleming. *BiochemJ* 1932;26(6):1907-18.
- [126] Frisvad JC, Smedsgaard J, Larsen TO, and Samson RA. Mycotoxins, drugs and other extrolites produced by species in *Penicillium* subgenus *Penicillium*. In: Samson RA, and Frisvad JC, editors. *Penicillium* subgenus *Penicillium*: new taxonomic schemes, mycotoxins and other extrolites Utrecht, The Netherlands Centraalbureau voor Schimmelcultures 2004.
- [127] Steyn PS. The isolation, structure and absolute configuration of secalonic acid D, the toxic metabolite of *Penicillium oxalicum*. *Tetrahedron Lett* 1970;26(1):51-7.
- [128] Reddy CS, Reddy RV, Hayes AW, Ciegler A. Teratogenicity of secalonic acid D in mice. *J Toxicol Environ Health* 1981;7(3-4):445-55.
- [129] Koyama M, Takahashi K, Chou TC, Darzynkiewicz Z, Kapuscinski J, Kelly TR, et al. Intercalating agents with covalent bond forming capability. A novel type of potential anticancer agents. 2. Derivatives of chrysophanol and emodin. *J Med Chem* 1989;32(7):1594-9.
- [130] Westendorf J, Marquardt H, Poginsky B, Dominiak M, Schmidt J, Marquardt H. Genotoxicity of naturally occurring hydroxyanthraquinones. *Mutat Res* 1990;240(1):1-12.
- [131] Zhang L, Lau YK, Xi L, Hong RL, Kim DS, Chen CF, et al. Tyrosine kinase inhibitors, emodin and its derivative repress HER-2/neu-induced cellular transformation and metastasis-associated properties. *Oncogene* 1998;16(22):2855-63.
- [132] Rahimipour S, Bilkis I, Peron V, Gescheidt G, Barbosa F, Mazur Y, et al. Generation of free radicals by emodic acid and its [D-Lys6]GnRH-conjugate. *Photochem Photobiol* 2001;74(2):226-36.

- [133] Singh SB, Zink DL, Guan Z, Collado J, Pelaez F, Felock PJ, et al. Isolation, structure, and HIV-1 integrase inhibitory activity of xanthoviridicatin E and F, two novel fungal metabolites produced by *Penicillium chrysogenum*. *Helvetica Chimica Acta* 2003;86:3380-85.
- [134] Cram DJ. Mold metabolites; the structure of sorbicillin, a pigment produced by the mold *Penicillium notatum*. *J Am Chem Soc* 1948;70(12):4240-3.
- [135] Abe N, Murata T, Hirota A. Novel oxidised sorbicillin dimers with 1,1-diphenyl-2-picrylhydrazyl-radical scavenging activity from a fungus. *Biosci Biotechnol Biochem* 1998;62(11):2120-26.
- [136] Bringmann G, Lang G, Gulder TAM, Tsuruta H, Mühlbacher J, Maksimenka K, et al. The first sorbicillinoid alkaloids, the antileukemic sorbicillactones A and B, from a sponge-derived *Penicillium chrysogenum* strain. *Tetrahedron* 2005;61:7252-65.
- [137] Sugaya K, Koshino H, Hongo Y, Yasunaga K, Onose J, Yoshikawa K, et al. The biosynthesis of sorbicillinoids in *Trichoderma sp* USF-2690: prospect for the existence of a common precursor to sorbicillinol and 5-epihydroxyvertinolide, a new sorbicillinoid member. *Tetrahedron Lett* 2008;49(4):654-57.
- [138] Maskey RP, Grun-Wollny I, Laatsch H. Sorbicillin Analogues and Related Dimeric Compounds from *Penicillium notatum*. *J. Nat. Prod.* 2005;68(6):865-70.
- [139] Miller RF, Huang S. Isolation and structure of sorrentanone: a new tetrasubstituted quinone from *Penicillium chrysogenum*. *J Antibiot (Tokyo)* 1995;48(6):520-1.
- [140] Andrade R, Ayer WA, Mebe PP. The Metabolites of *Trichoderma-Longibrachiatum*. Part 1. Isolation of the metabolites and the structure of trichodimerol. *Can J Chem* 1992;70(10):2526-35.
- [141] Barnes-Seeman D, Corey EJ. A two-step total synthesis of the natural pentacycle trichodimerol, a novel inhibitor of TNF-alpha production. *Org Lett* 1999;1(9):1503-4.
- [142] Warr GA, Veitch JA, Walsh AW, Hesler GA, Pirnik DM, Leet JE, et al. BMS-182123, a fungal metabolite that inhibits the production of TNF-alpha by macrophages and monocytes. *J Antibiot (Tokyo)* 1996;49(3):234-40.
- [143] Abe N, Arakawa T, Hirota A. The biosynthesis of bisvertinolone: evidence for oxosorbicillinol as a direct precursor. *Chem Commun (Camb)* 2002(3):204-5.
- [144] Trifonov LS, Hilpert H, Floersheim P, Dreiding AS, Rast DM, Skrivanova R, et al. Bisvertinols: a new group of dimeric vertinoids from *Verticillium intertextum*. *Tetrahedron Lett* 1986;42(12):3157-79.
- [145] Kontani M, Sakagami Y, Marumo S. First [beta]-1,6-glucan biosynthesis inhibitor, bisvertinolone isolated from fungus, *Acremonium strictum* and its absolute stereochemistry. *Tetrahedron Lett* 1994;35(16):2577-80.
- [146] Li X, Choi HD, Kang JS, Lee CO, Son BW. New polyoxygenated farnesylcyclohexenones, deacetoxyyanuthone A and its hydro derivative from the marine-derived fungus *Penicillium sp.* *J Nat Prod* 2003;66(11):1499-500.
- [147] Bringmann G, Gulder TA, Lang G, Schmitt S, Stohr R, Wiese J, et al. Large-scale biotechnological production of the antileukemic marine natural product sorbicillactone a. *Mar Drugs* 2007;5(2):23-30.
- [148] Bringmann G, Lang G, Mühlbacher J, Schaumann K, Steffens S, Müller WEG, inventors; Sorbicillactone A derivatives for the treatment of tumor and viral diseases. 2004.
- [149] Lang G, Stohr R. HPLC chromatogram of metabolites isolated from *P. chrysogenum* strain E01 10/3 (IFM-GEOMAR collection). In.
- [150] Simpson TJ. Polyketide biosynthesis. *Chemistry and industry* 1995:407-11.
- [151] O' Hagan D. The polyketide metabolites. Chichester: Ellis Horwood Limited, 1991.
- [152] Hopwood DA. Cracking the polyketide code. *PLoS Biol.* 2004;2(2):0166-69.
- [153] Hutchinson CR, Fujii I. Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. *Annu Rev Microbiol* 1995;49:201-38.

- [154] Hopwood DA. Genetic Contributions to Understanding Polyketide Synthases. *Chem Rev* 1997;97(7):2465-98.
- [155] Hopwood DA, Sherman DH. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* 1990;24:37-66.
- [156] Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. *Nat Prod Rep* 2001;18(4):380-416.
- [157] Shen B. Biosynthesis of aromatic polyketides. *Top Curr Chem* 2000;209:1-51.
- [158] Shen B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* 2003;7(2):285-95.
- [159] Revill WP, Bibb MJ, Hopwood DA. Purification of a malonyltransferase from *Streptomyces coelicolor* A3(2) and analysis of its genetic determinant. *J Bacteriol* 1995;177(14):3946-52.
- [160] Funo N, Awakawa T, Horinouchi S. Pentaketide resorcylic acid synthesis by type III polyketide synthase from *Neurospora crassa*. *J Biol Chem* 2007;282(19):14476-81.
- [161] Schroder J. Probing plant polyketide biosynthesis. *Nat Struct Biol* 1999;6(8):714-6.
- [162] Austin MB, Noel JP. The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* 2003;20(1):79-110.
- [163] Seshime Y, Juvvadi PR, Fujii I, Kitamoto K. Discovery of a novel superfamily of type III polyketide synthases in *Aspergillus oryzae*. *Biochem Biophys Res Commun* 2005;331(1):253-60.
- [164] He J. Molecular analysis of the aureothin biosynthesis gene cluster from *Streptomyces thioluteus* HKI-227; new insights into polyketide assembly. Jena: Friedrich-Schiller-Universität Jena; 2005.
- [165] Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 2003;422(6934):859-68.
- [166] Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, et al. Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* 2005;438(7071):1157-61.
- [167] Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc Natl Acad Sci U S A* 2003;100(26):15670-5.
- [168] Schumann J, Hertweck C. Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes. *J Biotechnol* 2006;124(4):690-703.
- [169] Nicholson TP, Rudd BA, Dawson M, Lazarus CM, Simpson TJ, Cox RJ. Design and utility of oligonucleotide gene probes for fungal polyketide synthases. *Chem Biol* 2001;8(2):157-78.
- [170] Bailey AM, Cox RJ, Harley K, Lazarus CM, Simpson TJ, Skellam E. Characterisation of 3-methylorcinolaldehyde synthase (MOS) in *Acremonium strictum*: first observation of a reductive release mechanism during polyketide biosynthesis. *Chem Commun (Camb)* 2007(39):4053-5.
- [171] Watanabe A, Fujii I, Tsai H, Chang YC, Kwon-Chung KJ, Ebizuka Y. *Aspergillus fumigatus* *alb1* encodes naphthopyrone synthase when expressed in *Aspergillus oryzae*. *FEMS Microbiol Lett* 2000;192(1):39-44.
- [172] Minto RE, Townsend CA. Enzymology and Molecular Biology of Aflatoxin Biosynthesis. *Chem Rev* 1997;97(7):2537-56.
- [173] Watanabe A, Fujii I, Sankawa U, Mayorga ME, Timberlake WW, Ebizuka Y. Re-identification of *Aspergillus nidulans* *wA* gene to code for a polyketide synthase of naphthopyrone. *Tetrahedron Lett* 1999;40:91-94.

- [174] Eley KL, Halo LM, Song Z, Powles H, Cox RJ, Bailey AM, et al. Biosynthesis of the 2-Pyridone Tenellin in the Insect Pathogenic Fungus *Beauveria bassiana*. *Chembiochem* 2007;8(3):289-97.
- [175] Yang G, Rose MS, Turgeon BG, Yoder OC. A polyketide synthase is required for fungal virulence and production of the polyketide T-toxin. *Plant Cell* 1996;8(11):2139-50.
- [176] Baker SE, Kroken S, Inderbitzin P, Asvarak T, Li BY, Shi L, et al. Two polyketide synthase-encoding genes are required for biosynthesis of the polyketide virulence factor, T-toxin, by *Cochliobolus heterostrophus*. *Mol Plant Microbe Interact* 2006;19(2):139-49.
- [177] Takano Y, Kubo Y, Shimizu K, Mise K, Okuno T, Furusawa I. Structural analysis of PKS1, a polyketide synthase gene involved in melanin biosynthesis in *Colletotrichum lagenarium*. *Mol Gen Genet* 1995;249(2):162-7.
- [178] Bradshaw RE, Jin H, Morgan BS, Schwelm A, Teddy OR, Young CA, et al. A polyketide synthase gene required for biosynthesis of the aflatoxin-like toxin, dothistromin. *Mycopathologia* 2006;161(5):283-94.
- [179] Sims JW, Fillmore JP, Warner DD, Schmidt EW. Equisetin biosynthesis in *Fusarium heterosporum*. *Chem Commun* 2005(2):186-88.
- [180] Song Z, Cox RJ, Lazarus CM, Simpson TT. Fusarin C biosynthesis in *Fusarium moniliforme* and *Fusarium venenatum*. *Chembiochem* 2004;5(9):1196-203.
- [181] Cox RJ, Glod F, Hurley D, Lazarus CM, Nicholson TP, Rudd BA, et al. Rapid cloning and expression of a fungal polyketide synthase gene involved in squalestatin biosynthesis. *Chem Commun (Camb)* 2004(20):2260-1.
- [182] Bingle LE, Simpson TJ, Lazarus CM. Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. *Fungal Genet Biol* 1999;26(3):209-23.
- [183] Moriguchi T, Ebizuka Y, Fujii I. Domain-domain interactions in the iterative type I polyketide synthase ATX from *Aspergillus terreus*. *Chembiochem* 2008;9(8):1207-12.
- [184] Moriguchi T, Ebizuka Y, Fujii I. Analysis of subunit interactions in the iterative type I polyketide synthase ATX from *Aspergillus terreus*. *Chembiochem* 2006;7(12):1869-74.
- [185] Schulte U. Genomics of filamentous fungi. In: Tkacz JS, Lange L, editors. *Advances in fungal biotechnology for industry, agriculture and medicine*. New York, Kluwer Academic/Plenum Publishers, 2004: 15-29.
- [186] Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, et al. Eukaryotic genome size databases. *Nucleic Acids Res* 2007;35(Database issue):D332-8.
- [187] Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. *Science* 1996;274(5287):546, 63-7.
- [188] Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature* 2002;415(6874):871-80.
- [189] Mannhaupt G, Montrone C, Haase D, Mewes HW, Aign V, Hoheisel JD, et al. What's in the genome of a filamentous fungus? Analysis of the *Neurospora* genome sequence. *Nucleic Acids Res* 2003;31(7):1944-54.
- [190] Spingola M, Grate L, Haussler D, Ares M, Jr. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *Rna* 1999;5(2):221-34.
- [191] Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res* 2005;15(12):1620-31.
- [192] Mutasa E, Tymon A, Goettgens B, Mellon F, Little P, Casselton L. Molecular organisation of an A mating type factor of the basidiomycete fungus *Coprinus cinereus*. *Curr Genet* 1990;18:223-29.
- [193] Saitoh K, Togashi K, Arie T. A simple method for a mini-preparation of fungal DNA. *J Gen Plant Pathol* 2006;72:348-50.
- [194] Borman AM, Linton CJ, Miles SJ, Johnson EM. Molecular identification of pathogenic fungi. *J Antimicrob Chemother* 2008;61 Suppl 1:7-12.

- [195] Xu JR, Peng YL, Dickman MB, Sharon A. The dawn of fungal pathogen genomics. *Annu Rev Phytopathol* 2006;44:337-66.
- [196] Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 2005;434(7036):980-6.
- [197] Brown TA. Studing genomes. In: *Genomes*. second ed. Oxford, BIOS Scientific Publishers Ltd, 2002: 93-187.
- [198] Guigo R, Knudsen S, Drake N, Smith T. Prediction of gene structure. *J Mol Biol* 1992;226(1):141-57.
- [199] Salamov AA, Solovyev VV. Ab initio gene finding in Drosophila genomic DNA. *Genome Res* 2000;10(4):516-22.
- [200] Korf I. Gene finding in novel genomes. *BMC Bioinformatics* 2004;5:59.
- [201] Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Jr., et al. Characterization of the yeast transcriptome. *Cell* 1997;88(2):243-51.
- [202] Schulte U, Becker I, Mewes HW, Mannhaupt G. Large scale analysis of sequences from *Neurospora crassa*. *J Biotechnol* 2002;94(1):3-13.
- [203] Hildebrand M, Waggoner LE, Lim GE, Sharp KH, Ridley CP, Haygood MG. Approaches to identify, clone, and express symbiont bioactive metabolite genes. *Nat Prod Rep* 2004;21(1):122-42.
- [204] Gaucher GM, Shepherd MG. Isolation of orsellinic acid synthase. *Biochem Biophys Res Commun* 1968;32(4):664-71.
- [205] Monfil VO, Penagos CC, Estrella AH. Three decades of fungal transformation: key concepts and applications. In: Balbas P, Lorence A, editors. *Methods in molecular biology: recombinant gene expression: reviews and protocols*. Second ed. Totowa, NJ, Humana Press Inc., 2004: 297-313.
- [206] Hamer JE. A molecular tool kit for fungal biotechnology. In: Tkacz JS, Lange L, editors. *Advances in fungal biotechnology for industry, agriculture and medicine*. New York, Kluwer Academic/Plenum Publishers, 2004: 31-39.
- [207] Thierry D, Vaucheret H. Sequence homology requirements for transcriptional silencing of 35S transgenes and post-transcriptional silencing of nitrite reductase (trans)genes by the tobacco 271 locus. *Plant Mol Biol* 1996;32(6):1075-83.
- [208] Baulcombe DC. Gene silencing: RNA makes RNA makes no protein. *Curr Biol* 1999;9(16):R599-601.
- [209] Nakayashiki H. RNA silencing in fungi: mechanisms and applications. *FEBS Lett* 2005;579(26):5950-7.
- [210] Brakhage AA, Langfelder K. Menacing mold: the molecular biology of *Aspergillus fumigatus*. *Annu Rev Microbiol* 2002;56:433-55.
- [211] Fincham JR. Transformation in fungi. *Microbiol Rev* 1989;53(1):148-70.
- [212] Gems D, Johnstone IL, Clutterbuck AJ. An autonomously replicating plasmid transforms *Aspergillus nidulans* at high frequency. *Gene* 1991;98(1):61-7.
- [213] Roncero MI, Jepsen LP, Stroman P, van Heeswijk R. Characterization of a leuA gene and an ARS element from *Mucor circinelloides*. *Gene* 1989;84(2):335-43.
- [214] Tsukuda T, Carleton S, Fotheringham S, Holloman WK. Isolation and characterization of an autonomously replicating sequence from *Ustilago maydis*. *Mol Cell Biol* 1988;8(9):3703-9.
- [215] Fierro F, Kosalkova K, Gutierrez S, Martin JF. Autonomously replicating plasmids carrying the AMA1 region in *Penicillium chrysogenum*. *Curr Genet* 1996;29(5):482-9.
- [216] Pfeifer BA, Khosla C. Biosynthesis of polyketides in heterologous hosts. *Microbiol Mol Biol Rev* 2001;65(1):106-18.
- [217] Bochar DA, Stauffacher CV, Rodwell VW. Sequence comparisons reveal two classes of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Mol Genet Metab* 1999;66(2):122-7.

- [218] Fernandez-Moreno MA, Caballero JL, Hopwood DA, Malpartida F. The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. *Cell* 1991;66(4):769-80.
- [219] Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 2005;438(7071):1105-15.
- [220] Wilkinson B, Micklefield J. Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol* 2007;3(7):379-86.
- [221] Zhu X, Yu F, Bojja RS, Zaleta-Rivera K, Du L. Functional replacement of the ketosynthase domain of FUM1 for the biosynthesis of fumonisins, a group of fungal reduced polyketides. *J Ind Microbiol Biotechnol* 2006;33(10):859-68.
- [222] Bergmann S, Schumann J, Scherlach K, Lange C, Brakhage AA, Hertweck C. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nat Chem Biol* 2007;3(4):213-7.
- [223] Hoffmeister D, Keller NP. Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat Prod Rep* 2007;24(2):393-416.
- [224] Shwab EK, Keller NP. Regulation of secondary metabolite production in filamentous ascomycetes. *Mycol Res* 2008;112(Pt 2):225-30.
- [225] Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, et al. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 2005;438(7071):1151-6.
- [226] Birren B, Lander ES, Galagan JE, Devon K, Nusbaum C, Ma LJ, et al. Annotation of the *Aspergillus terreus* NIH 2624 genome. In, UniProt, 2005.
- [227] Birren B, Lander ES, Galagan JE, Devon K, Nusbaum C, Ma LJ, et al. Annotation of the *Chaetomium globosum* CBS 148.51 genome. In, UniProt, 2005.
- [228] Birren B, Lander ES, Galagan JE, Devon K, Nusbaum C, Ma LJ, et al. Annotation of the *Coccidioides immitis* RS genome. In, UniProt, 2005.
- [229] Kagan RM, Clarke S. Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Arch Biochem Biophys* 1994;310(2):417-27.
- [230] Miller DJ, Ouellette N, Evdokimova E, Savchenko A, Edwards A, Anderson WF. Crystal complexes of a predicted S-adenosylmethionine-dependent methyltransferase reveal a typical AdoMet binding domain and a substrate recognition domain. *Protein Sci* 2003;12(7):1432-42.
- [231] Zhang A, Lu P, Dahl-Roshak AM, Paress PS, Kennedy S, Tkacz JS, et al. Efficient disruption of a polyketide synthase gene (*pksI*) required for melanin synthesis through *Agrobacterium*-mediated transformation of *Glarea lozoyensis*. *Mol Genet Genomics* 2003;268(5):645-55.
- [232] Lee T, Yun SH, Hodge KT, Humber RA, Krasnoff SB, Turgeon GB, et al. Polyketide synthase genes in insect- and nematode-associated fungi. *Appl Microbiol Biotechnol* 2001;56(1-2):181-7.
- [233] Chooi YH, Stalker DM, Davis MA, Fujii I, Elix JA, Louwhoff SH, et al. Cloning and sequence characterization of a non-reducing polyketide synthase gene from the lichen *Xanthoparmelia semiviridis*. *Mycol Res* 2008;112(Pt 2):147-61.
- [234] Harrison CJ, Langdale JA. A step by step guide to phylogeny reconstruction. *Plant J* 2006;45(4):561-72.
- [235] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24(8):1596-9.

- [236] Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, et al. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 2006;439(7078):811-6.
- [237] Murphy RL, Andrianopoulos A, Davis MA, Hynes MJ. Identification of amdX, a new Cys-2-His-2 (C2H2) zinc-finger gene involved in the regulation of the amdS gene of *Aspergillus nidulans*. *Mol Microbiol* 1997;23(3):591-602.
- [238] Graminha MA, Rocha EM, Prade RA, Martinez-Rossi NM. Terbinafine resistance mediated by salicylate 1-monooxygenase in *Aspergillus nidulans*. *Antimicrob Agents Chemother* 2004;48(9):3530-5.
- [239] Fujihara H, Yoshida H, Matsunaga T, Goto M, Furukawa K. Cross-regulation of biphenyl- and salicylate-catabolic genes by two regulatory systems in *Pseudomonas pseudoalcaligenes KF707*. *J Bacteriol* 2006;188(13):4690-7.
- [240] Calvo AM, Bok J, Brooks W, Keller NP. veA is required for toxin and sclerotial production in *Aspergillus parasiticus*. *Appl Environ Microbiol* 2004;70(8):4733-9.
- [241] Ishikawa J, Hotta K. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. *FEMS Microbiol Lett* 1999;174(2):251-3.
- [242] Bhat MA, Philp AV, Glover DM, Bellen HJ. Chromatid segregation at anaphase requires the barren product, a novel chromosome-associated protein that interacts with Topoisomerase II. *Cell* 1996;87(6):1103-14.
- [243] Stinnett SM, Espeso EA, Cobeno L, Araujo-Bazan L, Calvo AM. *Aspergillus nidulans* VeA subcellular localization is dependent on the importin alpha carrier and on light. *Mol Microbiol* 2007;63(1):242-55.
- [244] Kato N, Brooks W, Calvo AM. The expression of sterigmatocystin and penicillin genes in *Aspergillus nidulans* is controlled by veA, a gene required for sexual development. *Eukaryot Cell* 2003;2(6):1178-86.
- [245] Cary JW, GR OB, Nielsen DM, Nierman W, Harris-Coward P, Yu J, et al. Elucidation of veA-dependent genes associated with aflatoxin and sclerotial production in *Aspergillus flavus* by functional genomics. *Appl Microbiol Biotechnol* 2007;76(5):1107-18.
- [246] Dreyer J, Eichhorn H, Friedlin E, Kurnsteiner H, Kuck U. A homologue of the *Aspergillus velvet* gene regulates both cephalosporin C biosynthesis and hyphal fragmentation in *Acremonium chrysogenum*. *Appl Environ Microbiol* 2007;73(10):3412-22.
- [247] Zelter A, Bencina M, Bowman BJ, Yarden O, Read ND. A comparative genomic analysis of the calcium signaling machinery in *Neurospora crassa*, *Magnaporthe grisea*, and *Saccharomyces cerevisiae*. *Fungal Genet Biol* 2004;41(9):827-41.
- [248] Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res* 2001;29(5):1097-106.
- [249] Zhang HZ, Hackbarth CJ, Chansky KM, Chambers HF. A proteolytic transmembrane signaling pathway and resistance to beta-lactams in staphylococci. *Science* 2001;291(5510):1962-5.
- [250] Maddox PS, Hyndman F, Monen J, Oegema K, Desai A. Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. *J Cell Biol* 2007;176(6):757-63.
- [251] Percudani R, Peracchi A. A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep* 2003;4(9):850-4.
- [252] Eswaramoorthy S, Gerchman S, Graziano V, Kycia H, Studier FW, Swaminathan S. Structure of a yeast hypothetical protein selected by a structural genomics approach. *Acta Crystallogr D Biol Crystallogr* 2003;59(Pt 1):127-35.



- [253] Chang CH, Frey PA. Cloning, sequencing, heterologous expression, purification, and characterization of adenosylcobalamin-dependent D-lysine 5, 6-aminomutase from *Clostridium sticklandii*. J Biol Chem 2000;275(1):106-14.
- [254] Chen HP, Wu SH, Lin YL, Chen CM, Tsay SS. Cloning, sequencing, heterologous expression, purification, and characterization of adenosylcobalamin-dependent D-ornithine aminomutase from *Clostridium sticklandii*. J Biol Chem 2001;276(48):44744-50.
- [255] Crawford JM, Dancy BC, Hill EA, Udvary DW, Townsend CA. Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. Proc Natl Acad Sci U S A 2006;103(45):16728-33.
- [256] Udvary DW, Merski M, Townsend CA. A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. J Mol Biol 2002;323(3):585-98.
- [257] Crawford JM, Thomas PM, Scheerer JR, Vagstad AL, Kelleher NL, Townsend CA. Deconstruction of iterative multidomain polyketide synthase function. Science 2008;320(5873):243-6.
- [258] Pazoutova S, Linka M, Storkova S, Schwab H. Polyketide synthase gene *pksM* from *Aspergillus terreus* expressed during growth phase. Folia Microbiol (Praha) 1997;42(5):419-30.
- [259] Smith S, Witkowski A, Joshi AK. Structural and functional organization of the animal fatty acid synthase. Prog Lipid Res 2003;42(4):289-317.
- [260] Rangan VS, Joshi AK, Smith S. Mapping the functional topology of the animal fatty acid synthase by mutant complementation in vitro. Biochemistry 2001;40(36):10792-9.
- [261] Watanabe CM, Townsend CA. Initial characterization of a type I fatty acid synthase and polyketide synthase multienzyme complex NorS in the biosynthesis of aflatoxin B(1). Chem Biol 2002;9(9):981-8.
- [262] Gulder TA. Neue Bioaktive Naturstoffe: Strukturaufklärung, Biosynthese und Synthese, sowie Stereochemische Analyse von Naturstoffen und Synthetischen Verbindungen durch HPLC-CD. Würzburg: Julius-Maximilians-Universität Würzburg; 2008.
- [263] Banuelos O, Casqueiro J, Gutierrez S, Martin JF. Intrachromosomal recombination after targeted monocopy integration in *Penicillium chrysogenum*: stabilization of the direct repeats to prevent loss of the inserted gene. Curr Genet 2001;39(4):231-6.
- [264] Cantwell CA, Beckmann RJ, Dotzlaw JE, Fisher DL, Skatrud PL, Yeh WK, et al. Cloning and expression of a hybrid *Streptomyces clavuligerus* *cefE* gene in *Penicillium chrysogenum*. Curr Genet 1990;17(3):213-21.
- [265] Carramolino L, Lozano M, Perez-Aranda A, Rubio V, Sanchez F. Transformation of *Penicillium chrysogenum* to sulfonamide resistance. Gene 1989;77(1):31-8.
- [266] Gouka RJ, van Hartingsveldt W, Bovenberg RA, van Zeijl CM, van den Hondel CA, van Gorcom RF. Development of a new transformant selection system for *Penicillium chrysogenum*: isolation and characterization of the *P. chrysogenum* acetyl-coenzyme A synthetase gene (*facA*) and its use as a homologous selection marker. Appl Microbiol Biotechnol 1993;38(4):514-9.
- [267] Kolar M, Punt PJ, van den Hondel CA, Schwab H. Transformation of *Penicillium chrysogenum* using dominant selection markers and expression of an *Escherichia coli lacZ* fusion gene. Gene 1988;62(1):127-34.
- [268] Picknett TM, Saunders G. Transformation of *Penicillium chrysogenum* with selection for increased resistance to benomyl. FEMS Microbiol Lett 1989;51(1):165-8.
- [269] Renno DV, Saunders G, Bull AT, Holt G. The genetic stability of *Penicillium chrysogenum* transformants in a fermentor. Appl Microbiol Biotechnol 1990;34(3):364-7.
- [270] Sanchez F, Lozano M, Rubio V, Penalva MA. Transformation in *Penicillium chrysogenum*. Gene 1987;51(1):97-102.

- [271] Sun CB, Kong QL, Xu WS. Efficient transformation of *Penicillium chrysogenum* mediated by *Agrobacterium tumefaciens* LBA4404 for cloning of *Vitreoscilla* hemoglobin gene. *Electronic Journal of Biotechnology* 2002;5(1):21-28.
- [272] Casqueiro J, Gutierrez S, Banuelos O, Hijarrubia MJ, Martin JF. Gene targeting in *Penicillium chrysogenum*: disruption of the *lys2* gene leads to penicillin overproduction. *J Bacteriol* 1999;181(4):1181-8.
- [273] Lamas-Maceiras M, Vaca I, Rodriguez E, Casqueiro J, Martin JF. Amplification and disruption of the phenylacetyl-CoA ligase gene of *Penicillium chrysogenum* encoding an aryl-capping enzyme that supplies phenylacetic acid to the isopenicillin N-acyltransferase. *Biochem J* 2006;395(1):147-55.
- [274] Sambrook J, and Russell DW. *Molecular Cloning: A Laboratory Manual*. third edition ed: CSHL Press, 2001.
- [275] Goblet C, Prost E, Whalen RG. One-step amplification of transcripts in total RNA using the polymerase chain reaction. *Nucleic Acids Res* 1989;17(5):2144.
- [276] Ralser M, Querfurth R, Warnatz HJ, Lehrach H, Yaspo ML, Krobitch S. An efficient and economic enhancer mix for PCR. *Biochem Biophys Res Commun* 2006;347(3):747-51.
- [277] Promega G. *Nucleic Acid Amplification*. In: *Protocols & Applications Guide*, 2006.
- [278] Clark JM. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res* 1988;16(20):9677-86.
- [279] Carothers AM, Urlaub G, Mucha J, Grunberger D, Chasin LA. Point mutation analysis in a mammalian gene: rapid preparation of total RNA, PCR amplification of cDNA, and Taq sequencing by a novel method. *BioTechniques* 1989;7(5):494-6, 98-9.
- [280] Krishnan BR, Kersulyte D, Brikun I, Berg CM, Berg DE. Direct and crossover PCR amplification to facilitate Tn5supF-based sequencing of lambda phage clones. *Nucleic Acids Res* 1991;19(22):6177-82.
- [281] Thein SL, Wallace RB. *Human genetic diseases: A practical approach*. Oxford, United Kingdom: IRL Press, 1986.
- [282] Marchuk D, Drumm M, Saulino A, Collins FS. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res* 1990;19(5):1154.
- [283] Chassy BM, Mercenier A, Flickinger J. Transformation of bacteria by electroporation. *Trends. Biotechnol.* 1988;6:303-09.
- [284] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 1977;74:5436-67.

## APPENDIX

## 13 UV spectrum of sorbicillactone A

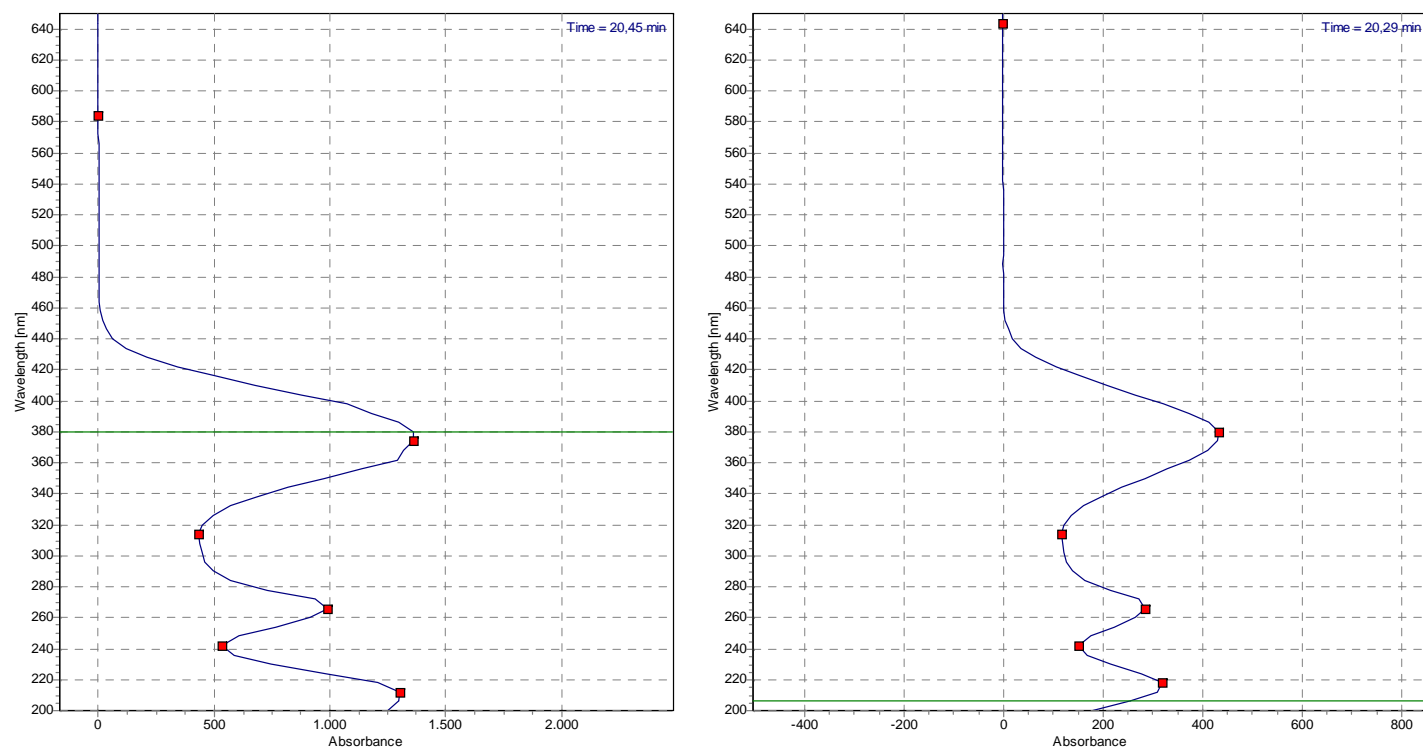


Figure 42: UV spectrum of sorbicillactone A. Links: extracted compound from the liquid culture of *P. chrysogenum*. Rechts: reference compound sorbicillactone A.

## 14 Degenerate primer design

**Table 32: NR PKS clade III specific KS domain degenerate primer design:**

Sequence name*	Conserved amino acid sequences		
	Region of homology I (KS)	Region of homology II (KS)	Region of homology III (KS)
PKSct Mp, <i>M. purpureus</i>	<u>NYDTFDHKFFKKS</u> PR	SFLSPTGQCKPFD	ALINNYGASGSNASMVV
PKS17 Bf, <i>Botryotinia fuckeliana</i>	<u>NYNTFDHKFFKKS</u> PR	SFLSPTGQCKPFD	ALINNYGASGSNASMVV
Q5BG07 An, <i>A. nidulans</i>	EYDTFDHKFFQKSPR	SFLSPTGQCKPFD	ALLNNYGASGSNASLTV
PKS3 Gm, <i>Gibberella moniliformis</i>	DPSEFDARFFNISPR	HFLSRTGNCKTFD	VFVNNFSAAGNSALLI
Q0CSA2 At, <i>A. terreus</i>	<u>DHDAFDHKFFQK</u> SPR	SFLSQTGPCKPFD	ALINNYGASGSNASMIV
Q1DRI2 Ci, <i>Coccidioides immitis</i>	<u>GHDMPFDHKFFKKS</u> PR	SFLSTTGQCKPFD	ALINNYGASGSNASMVI
Q2GPS7 Cg, <i>Chaetomium globosum</i>	<u>DHDTFDHKFFKKS</u> PR	SFLSPTGQCKPFD	ALINNYGASGSNASMVV
Q0CCC2 At, <i>A. terreus</i>	<u>DIDAFDHKFFRK</u> VPR	SFVSPTGPCKPFD	VLINNYGASGSNASMVI
Q2TXJ8 Ao, <i>A. oryzae</i>	<u>DIDAFDHKFFK</u> KVPR	SFVSPTGPCKPFD	ALINNYGASGSNASMVI
Q4WFD7 Af, <i>A. fumigatus</i>	<u>DYDAFDHRFFKKS</u> PR	SFLSPTGACKPFD	ALINNYGASGSNASLTV
Q0CF73 At, <i>A. terreus</i>	<u>DGDAFDHKFFKRS</u> PR	SFVSPTGQCKPFD	ALLNNYGACGSNASMII
Q1DVU2 Ci, <i>Coccidioides immitis</i>	<u>DVDAFDHRFFKRS</u> PR	SFLSPTGQCKPFD	ALINNYGASGSNAAMVV
Q5BEJ6 An, <i>A. nidulans</i>	<u>DGDAFDHKFFKRS</u> PR	SFVSPTGQCKPFD	ALLNNYGACGSNASMIV
Q2U710 Ao, <i>A. oryzae</i>	<u>DVDAFDHKFFKRS</u> PR	SFLSPTGQCKPFD	VLINNYGASGSNAAMVV
Q5BBP8 An, <i>A. nidulans</i>	<u>DPDAFDHKFFKKS</u> PR	SFISPTGQCKPFD	ALINNYGACGSNSAIV
Q5B8A0A An, <i>A. nidulans</i>	<u>DRSQFDHSFFKMS</u> PR	SFLSPTGQCKPFD	ALVNSYGASGSNTSMVI
Q2GR18 Cg, <i>Chaetomium globosum</i>	<u>DTAVFDHKFFKKS</u> PR	-----	AMINNYGAAGSNASILI
Q5B7U4 An, <i>A. nidulans</i>	<u>DIDAFDHKFFRKS</u> PR	SFLSPTGPCKPFD	ALINNYGACGSNASMVI
Q2UNE1 Ao, <i>A. oryzae</i>	<u>DYDAFDHKFFQK</u> SPR	SFLSPTGQCKPFD	ALINNYGASGSNASLVI
<b>Conserved motif(s)**</b>	(N/D)YDTFDHKFFK	PTGQCKPF	GASGSNA
<b>Designed primers</b>	sorb1-FFK-for	sorb2-PTG-rev, sorb3-KPF-forv	sorb4-GAS-rev

\* Shortened name of each sequence with full name of fungus that it originates from; for sequence legend see table (.). \*\* Conserved amino acid motif that was chosen from displayed conserved amino acid sequences from alignment; chosen motif was used for design of degenerate primers that are shown in the same column below the motif they correspond to. Position of conserved motif within the chosen region of homology is shown underlined. Dashed lane indicates that no conserved sequence was found within region of homology.

Appendix

**Table 33: NR PKS clade III specific AT domain degenerate primer design.**

Sequence name*	Conserved amino acid sequences		
	Region of homology III (KS)	Region of homology I (AT)	Region of homology II (AT)
PKSct Mp, <i>M. purpureus</i>	ALINNYGASGSNASMVV	PVILCFGQVSTYVGL	SVVGHSEFGELIALCV
PKS17 Bf, <i>Botryotinia fuckeliana</i>	ALINNYGASGSNASMVV	PVILCFGQVSTYVGL	AAIGHSEFGELTALCV
Q5BG07 An, <i>A. nidulans</i>	ALLNNYGASGSNASLVV	PVILCFGQVSTYVGL	ALVGHSEFGELTAMCV
PKS3 Gm, <i>Gibberella moniliformis</i>	VFVNNFSAAGGSNALLI	KVVFTFTGQGAQYPGM	AVVGHSLGEYAALNV
Q0CSA2 At, <i>A. terreus</i>	ALINNYGASGSNASMIV	PVILCFGQVSTYVGL	AVVGHSEFGELTALCI
Q1DRI2 Ci, <i>Coccidioides immitis</i>	ALINNYGASGSNASMVI	PVILCFGQVSTYVGL	AVVGHSEFGELTALCV
Q2GPS7 Cg, <i>Chaetomium globosum</i>	ALINNYGASGSNASMVV	PVIMCFGQVSTYVGL	AVVGHSEFGELTALCV
Q0CCC2 At, <i>A. terreus</i>	VLINNYGASGSNASMVI	PVILCFGQVSTYVGL	AVVGHSEFGELTSLCV
Q2TXJ8 Ao, <i>A. oryzae</i>	ALINNYGASGSNASMVI	SIVLCFGQVSTYVGL	AVVGHSEFGELTALCI
Q4WFD7 Af, <i>A. fumigatus</i>	ALINNYGASGSNASLVV	PVILCFGQVSTYVGL	AVVGHSEFGELTALCI
Q0CF73 At, <i>A. terreus</i>	ALLNNYGACGSNASMII	PVILCFGQVSTYVGL	AVVGHSEFGELTALCI
Q1DVU2 Ci, <i>Coccidioides immitis</i>	ALINNYGASGSNAAMVV	PVILCFGQVSTYVGL	SVVGHSEFGELTALCI
Q5BEJ6 An, <i>A. nidulans</i>	ALLNNYGACGSNASMIV	PVILCFGQVSTYVGL	ALVGHSEFGELTALCV
Q2U710 Ao, <i>A. oryzae</i>	VLINNYGASGSNAAMVV	PVILCFGQVSTYVGL	SVIGHSEFGELTALCI
Q5BBP8 An, <i>A. nidulans</i>	ALINNYGACGSNSAIV	PVILCFGQVSTYVGL	SVVGHSEFGELTALCV
Q5B8A0A An, <i>A. nidulans</i>	ALVNSYGASGSNTSMVI	PVILCFGQVSTYVGL	AVVGHSEFGELTALS I
Q2GR18 Cg, <i>Chaetomium globosum</i>	AMINNYGAAGSNASILI	PVILAFGGQVGVKVVGL	ALVGHSEFGELTALCI
Q5B7U4 An, <i>A. nidulans</i>	ALINNYGACGSNASMVI	PLVLCFGQVGRSIGL	AVIGHSEFGELTALCI
Q2UNE1 Ao, <i>A. oryzae</i>	ALINNYGASGSNASLVI	PVILCFGQVSTYVGL	ALVGHSEFGELTAMCV
<b>Conserved motif(s)**</b>	ALINNYGASG	PVILCFGQV	VVGHSEFGE
<b>Designed primers</b>	sorb5-ASG-for	sorb6-PVI-rev, sorb7-GQV-for	sorb8-VVG-rev, sorb9-FGE-for

## Appendix

Sequence name*	Conserved amino acid sequences			
	Region of homology III(AT)	Region of homology IV (AT)		Region of homology V (AT)
PKSct Mp, <i>M. purpureus</i>	KRLNVTNAFHSLV	AIWLEAGSNSTITAM		VLLPPYQFEKS
PKS17 Bf, <i>Botryotinia fuckeliana</i>	KKLNVTNAFHCHLV	ATWLEAGSNSTVTTM		VLLPPYQFEKS
Q5BG07 An, <i>A. nidulans</i>	KSLETSHAFHSTLV	SIWLEAGSGSGVTNL		LLLPPYQFAKS
PKS3 Gm, <i>Gibberella moniliformis</i>	TLLKVPYAFHSSQL	SMVLEVGPHPAVSGM		VDLPAYSWDLK
Q0CSA2 At, <i>A. terreus</i>	KKLSVSNFAFHSTLV	AIWLEAGSNSTITRM		LLLPPYQFEKS
Q1DRI2 Ci, <i>Coccidioides immitis</i>	EETECYQCISLCSFA	CIFLEAGSNSTITSM		LLLPPYQFDKV
Q2GPS7 Cg, <i>Chaetomium globosum</i>	KRLNVTHAFHSLV	AAWLEAGSASTITNM		LFLPPYQFEKV
Q0CCC2 At, <i>A. terreus</i>	KRLSVTNAFHSSLV	CVWLEAGSNSTITLM		VLLPPYQFEKS
Q2TXJ8 Ao, <i>A. oryzae</i>	KRLFVTNAFHSDLV	CVWLEAGSNSTITYM		MFLPPYQFEKS
Q4WFD7 Af, <i>A. fumigatus</i>	KRLNVSNAFHSTLV	AIWLEAGSNSTITNM		LLLPPYQFELF
Q0CF73 At, <i>A. terreus</i>	KRLSVTNAFHSSLV	AIFLEAGSSSTITIM		LLLPPYQFDTS
Q1DVU2 Ci, <i>Coccidioides immitis</i>	KRLNVTNAFHSSLV	AIFLEAGSSSTITVM		ILLPPYQFEKA
Q5BEJ6 An, <i>A. nidulans</i>	KRLNVTNAFHSSLV	AIFLEAGSSSTITVM		LLLPPYQFDTS
Q2U710 Ao, <i>A. oryzae</i>	KRLNVTNAFHSSLV	AIFLEAGSSSTITVM		ILLPPYQFEKA
Q5BBP8 An, <i>A. nidulans</i>	KRLNVTNAFHSSLV	AIFLEAGSNSTITVM		LLLPPYQFEKT
Q5B8A0A An, <i>A. nidulans</i>	KRLNVTNAFHCSLV	SIWLEAGSNSTITTM		LILPPRQFART
Q2GR18 Cg, <i>Chaetomium globosum</i>	KRIHVTNAFHSGLV	AIWLEAGSNSTITNM		IMLPPYQFERN
Q5B7U4 An, <i>A. nidulans</i>	KRLEVTNAFHSTLV	AIWLEAGSNSTITSL		LLLPPYQFERT
Q2UNE1 Ao, <i>A. oryzae</i>	KQLDTTHAFHSLV	SIWLEAGSGSGVTAL		LLLPPYQFAKT
<b>Conserved motif(s)**</b>	<u>NVTNAFH</u>	IWLEAGSN	IFLEAGSN	PPYQFEK
<b>Designed primers</b>	sorb10-NAFH-for	sorb11-IWL-rev	sorb12-IFL-rev	sorb13-PPY-rev

\* Sequence shortened name with name of fungus it originates from; for sequence legend see table ( ). \*\* Conserved amino acid motif that was chosen from displayed conserved amino acid sequences from alignment; chosen motif was used for design of degenerate primers that are shown in the same column below the motif they correspond to. Position of conserved motif within the chosen region of homology is shown underlined.

## Appendix

**Table 34: NR PKS clade III specific MT domain degenerate primer design.**

Sequence name *	Conserved amino acid sequences			
	Region of homology Ia (MT)		Region of homology II (MT)	
Q5BEJ6 An, <i>A. nidulans</i>	VLEMGAGTGG		SNAIHAT	
Q5B7U4 An, <i>A. nidulans</i>	ILEMGAGTGG		SNAVHAT	
PKSct Mp, <i>M. purpureus</i>	ILEMGAGTGG		TNCVHAT	
PKS17 Bf, <i>Botryotinia fuckeliana</i>	ILEMGAGTGG		TNCVHAT	
PKS18 Bf, <i>Botryotinia fuckeliana</i>	ILEIGGGTGG		TNCIHAT	
PKS21 Ch, <i>Cochliobolus heterostrophus</i>	ILEMGAGTGA		SNAVHAT	
PKS16 Bf, <i>Botryotinia fuckeliana</i>	ILEVGAGFGG		TNCVHAT	
PKS22 Ch, <i>Cochliobolus heterostrophus</i>	ILEIGAGFGG		TNCVHAT	
<b>Conserved motif(s)**</b>	EMG(A/G)GTG	E(I/V)GAGFG	SNA(V/I)HAT	TNC(V/I)HAT
<b>Designed primers</b>	MTnr III_Fa	MTnr III_Fb	MTnr III_R1a	MTnr III_R1b

\* Sequence shortened name with name of fungus it originates from; for sequence legend see table ( ). \*\* Conserved amino acid motif that was chosen from displayed conserved amino acid sequences from alignment; chosen motif was used for design of degenerate primers that are shown in the same column below the motif they correspond to. Position of conserved motif within the chosen region of homology is shown underlined.

## Appendix

**Table 35: NR PKS clade III specific RED domain degenerate primer design. The five sequences alingment.**

Sequence name*	Conserved amino acid sequences		
	Region of homology Ib (MT)	Region of homology III (RED)	Region of homology IV (RED)
PKSct Mp, <i>M. purpureus</i> PKS17 Bf, <i>Botryotinia fuckeliana</i> Q5BG07 An, <i>A. nidulans</i> PKS3 Gm, <i>Gibberella moniliformis</i> Q4WFD7 Af, <i>A. fumigatus</i>	<u>VLVTGATGSLGSHV</u> <u>VLVTGTTGSLGSHIV</u> <u>VLVTGATGSLGSHLV</u> <u>VLVTGASGGLGSHLV</u> <u>VLVTGATGSLGSHLV</u>	<u>GHWNPVEH</u> <u>GYWNPVEH</u> <u>GYWNPVEH</u> <u>GYWNPMEH</u> <u>GYWNPVEH</u>	<u>MSCGNLILD</u> <u>MSCGGLILD</u> <u>MSCGMILD</u> <u>MACGGLVLDV</u> <u>MACGGVVLDT</u>
<b>Conserved motif(s)**</b>	<u>VLVTGATG</u>	<u>GYWNPVEH</u>	<u>MSCGGLI</u>
<b>Designed primers</b>	RED-F1a, RED-F1b RED-R1a, RED-R1b	RED-F2, RED-R2	RED-R3a, RED-R3b

\* Sequence shortened name with name of fungus it originates from; for sequence legend see table (.). \*\* Conserved amino acid motif that was chosen from displayed conserved amino acid sequences from alignment; chosen motif was used for design of degenerate primers that are shown in the same column below the motif they correspond to. Position of conserved motif within the chosen region of homology is shown underlined.



## Appendix

**Table 36: NR PKS clade III specific RED domain degenerate primer design.** The five sequences alignment.

Sequence name*	Conserved amino acid sequences			
	Region of homology Ib (MT)	Region of homology II (RED)		Region of homology III (RED)
PKSct Mp, <i>M. purpureus</i>	VLVTGATGSLGSHVV	THIVHSAWPMS		GHWNPVEHFAF
PKS17 Bf, <i>Botryotinia fuckeliana</i>	VLVTGTTGSLGSHIV	THIVHNAWPMS		GYWNPVEHFFAS
Q5BG07 An, <i>A. nidulans</i>	VLVTGATGSLGSHLV	TDVVHNAWAMS		GYWNPVEHLVH
PKS3 Gm, <i>Gibberella moniliformis</i>	VLVTGASGGLGSHLV	THIIHNAWPMS		GYWNPMEHFGF
Q0CSA2 At, <i>A. terreus</i>	VLVTGATGSLGAHLV	TAIVHNAWPMS		GYWNSLEHLSF
Q1DRI2 Ci, <i>Coccidioides immitis</i>	VLVTGATGSLGSHII	TDIIHNAWLMN		GYWNPMEHLSF
Q2GPS7 Cg, <i>Chaetomium globosum</i>	VLITGATGSLGTHLV	TGILHNAWPMS		GYWNPVEHFCF
Q0CCC2 At, <i>A. terreus</i>	VLVTGTTGSLGSHIA	THIIHNAWPMN		GCWNPSEHLPA
Q2TXJ8 Ao, <i>A. oryzae</i>	VLLTGATGSLGSHIV	THIIHNAWPMN		GCWNTAEHFFA
Q4WFD7 Af, <i>A. fumigatus</i>	VLVTGATGSLGSHLV	THVVHNAWPMT		GYWNPVEHLAF
Q0CF73 At, <i>A. terreus</i>	VLVTGATGSLGSHLV	TDIVHNAWPMS		GFWNPVEHFAF
Q1DVU2 Ci, <i>Coccidioides immitis</i>	IIVTGATGSLGAHVV	SHIVHSAWPLS		GVWNSVEQIPF
Q5BEJ6 An, <i>A. nidulans</i>	VIVTGATGSLGSHIV	TDIIHNAWPMS		GFWNPVEHFAF
Q2U710 Ao, <i>A. oryzae</i>	VIVTGATGSLGVHVV	SHIVHSAWPLS		GVWNSVEQVPF
Q5BBP8 An, <i>A. nidulans</i>	VIVTGASGSLGSHLV	THIIHNAWPMS		GFWNPVEHFPF
Q5B8A0A An, <i>A. nidulans</i>	VIIITGATGSLGSHLV	THIVHNAWPMS		GCWPHTEHMPL
Q2GR18 Cg, <i>Chaetomium globosum</i>	VLVTGATGSLGSHIV	THIIHNAFPVN		GYWNHVEVLAF
Q5B7U4 An, <i>A. nidulans</i>	ILVTGATGSLGSHLV	THIIHNAWPMN		GYWNTAEHFFA
Q2UNE1 Ao, <i>A. oryzae</i>	VLVTGATGSLGSHLV	TDVAHNAWAMS		GYWNPVEHLVH
<b>Conserved motif(s)**</b>	<u>VTGATGSLG</u>	<u>IVHSAWP</u>	<u>IVHNAWP</u>	<u>WNPVEHFAF</u>
<b>Designed primers</b>	sorb14-SLG-for	sorb15-IVHS-rev	sorb16-IVHN-rev	sorb17-WNP1-rev sorb18-WNP2-rev

\* Shortened name of each sequence with full name of fungus that it originates from; for sequence legend see table ( ). \*\* Conserved amino acid motif that was chosen from displayed conserved amino acid sequences from alignment; chosen motif was used for design of degenerate primers that are shown in the same column below the motif they correspond to. Position of conserved motif within the chosen region of homology is shown underlined.

Appendix

**Table 37: Predicted size of PCR products for different combinations of forward and reverse primers - amplification of RED domain.**

Primers	CMeT1c	CMeT2c-F1	CMeT3c-F2	RED-F1a	RED-F1b	RED-R1a	RED-R1b	RED-F2	RED-R2	RED-R3a	RED-R3b
CMeT1c	-	-	-	-	-	375-696 bp	375-696 bp	-	1035-1356 bp	1425-1746 bp	1425-1746 bp
CMeT2c-F1	-	-	-	-	-	414 bp	414 bp	-	1074 bp	1464 bp	1464 bp
CMeT3c-F2	-	-	-	-	-	414 bp	414 bp	-	1074 bp	1464 bp	1464 bp
RED-F1a	-	-	-	-	-	-	-	-	660 bp	1050 bp	1050 bp
RED-F1b	-	-	-	-	-	-	-	-	660 bp	1050 bp	1050 bp
RED-R1a	*	*	*	-	-	-	-	-	-	-	-
RED-R1b	*	*	*	-	-	-	-	-	-	-	-
RED-F2	-	-	-	-	-	-	-	-	-	390 bp	390 bp
RED-R2	*	*	*	*	*	-	-	-	-	-	-
RED-R3a	*	*	*	*	*	-	-	*	-	-	-
RED-R3b	*	*	*	*	*	-	-	*	-	-	-

Given numbers represent PCR product length in base pairs (bp). A dash (-) represents absence of PCR product due to incompatibility of the given primer pair; such PCRs were not performed. Asterisk represents the combination of primers for the PCR products shown in Figure 20.

## 15 Nucleotide sequences of amplified PKS gene fragments

### KSpr-MI seq

CCGAGCAGATGGATCCTCAGCAGCGAATTGCCCTCGAAGTTGCCTGGGAGGCCTTGAAG  
ATGCTGGAATCTCTCCTTCCCCTTTCGGGGTCAAACACGTCCTGATACATGGGTGTCA  
ACTCCGATGACTATGCCAACTGGTCTTGGAAAGACCTTCCCAATGTCGGTGCATATGG  
GCGTGGGCACAGCCTACTGCGGCATTCCCAGCCGTATATCCTATCTGCTGGATTTGATGG  
GTCTAGCGTTGCACTGGATGCGGCGTGTGCATCATCCCTGGTGCCTGTACACCATGCC  
GTCAGGCTATCCGTGCTGGCGAGACTGACCTAGCTATCGCAGGTGGTGTCAACGCACCTCC  
TGGGACCAGGTTTGACCCGGGTGCTGGACGAAGCTGGGGCTATCTCCGCGGATGGAAAGT  
GCCGTTCCCTTCGATGACTCCGCTAACGGGTACGGTAGAGGCGAAGGAGCGGGTGTGGTCA  
TCTTAAAGCGTCTGGATAAGGCCTTCGCGGACGGTGACCAAGTCTAGCCGTACTCAAGG  
GCAGTGCCTGGCCTCAGACGGCAAGACACTTGGGATCATGGCCCCAATGCGCAGGCAC  
AGTTCTCGTTGGCCAAAAGCGCTGAAAGAGGCAAAAGTACGCTCTGACTCCATCAGTT  
ATATCGAGGCTCATGCCACATCGACA

### KSnr-7

CCCCGTGAAGCGCTTCAAGCAGACCCCTGCTCAACGCTCTGGCACTTCTTACGGCTTATGAGGCGTTGAAATGGCGGGATT  
CATCCAGACAGCAGCCCTCTACGCAAAAAGAACCGTGTGGGTGTCTTATGGAATGACCAGCGATGATTACCGTGAGA  
TCAACAGTGGTTCAGGACATTGACACCTACTTTATCCCGGGTGGTAACCGTGCCTTCACTCCTGGTGCATCAACTACTAT  
TTCAAGTTTCAAGCGCCCTAGTGTGAGTGTGGATACAGCTTGTCTCAGTCTTGCAGCAATTCATGTCGCTGTAATTC  
CCTGTGGAGGAATGAATGTGACTCTGCCGTGCGCGGTGGCGTCAATATTTGACAAAATCCTGACAACCATGCGGGCCTTG  
ACCGTGGACACTTCTCTCAAGGACTGGAACTGTACCACCTTTGATGATGGGGCTGATGGCTATTGCAGAGCAGATGGA  
ATTGGCTCCATTGCTAATAAGAGACTCGAAGATGCTCAGGCAGACAACGATCCCATCTACGGTGTATTGGAGGAGCTTA  
CACCAACCACTCCGACAGAGGCTGTATCAATTACGCGCCCTCATGTAGGAGCACAGTCTTCACTTTGACAAGCTTCTCA  
ACGAGTCCAACAGCGACCCGAAGGAGATCAGCTATATTGAAATGCACGGTACAGGAACATCACTAN

### KSnr-8

GATCCCCGATTTTCAACATGTCTCCCCGTGAAGCGCTTCAAGCAGACCCCTGCTCAACGCTCTGGCACTTCTTACGGCTTA  
TGAGGCGTTGAAATGGCGGGATTTCATCCAGACAGCAGCCCTCTACGCAAGAAGAACCGTGTGGGTGTCTTCTATGGAA  
TGACCAGCGATGATTACCGTGAGATCAACAGTGGTTCAGGACATTGACACCTACTTTATCCCGGGTGGTAACCGTGTCTTC  
ACTCCTGGTGCATCAACTACTATTTCAAGTTTCAAGCGCCCTAGTGTGCTAGTGTGGATACAGCTTGTCTCAGAGTCTTGC  
AGCAATTCATGTCGCCTGTAATTCCTGTGGAGGAATGAATGTGACTCTGCCGTGCGCGGTGGCGTCAATATTTGACAA  
ATCCTGACAACCATGCGGGCCTTGACCGTGGACACTTCTCTCAAGGACTGGAACCTGTACCACCTTTGATGATGGGGCT  
GATGGCTATTGCAGAGCAGATGGAATTTGGCTCCATTGTGCTAATAAGAGACTCGAAGATGCTCAGGCAGACAACGATCCCAT  
CTACGGTGTATTGGAGGAGCTTACACCAACCACTCCGACAGAGCTGTACCAATTACGCGGCCCTCATGTAGGAGCACAGT  
CCTTCACTTTGACAAGCTTCTCAACGAGTCCAACAGCGACCCGAAGGAGATCAGCTATATTG

### KSnr-10

GACCCAAGTTTCTTAAACATGTCTCCCCGTGAAGCGCTTCAAGCAGACCCCTGCTCAACGCTCTGGCACTTCTTACGGCTTA  
TGAGGCGTTGAAATGGCGGGATTTCATCCAGACAGCAGCCCTCTACGCAAGAAGAACCGTGTGGGTGTCTTCTATGGAA  
TGACCAGCGATGATTACCGTGAGATCAACAGTGGTTCAGGACATTGACACCTACTTTATCCCGGGTGGTAACCGTGTCTTC  
ACTCCTGGTGCATCAACTACTATTTCAAGTTTCAAGCGCCCTAGTGTGCTAGTGTGGATACAGCTTGTCTCAGAGTCTTGC  
AGCAATTCATGTCGCCTGTAATTCCTGTGGAGGAATGAATGTGACTCTGCCGTGCGCGGTGGCGTCAATATTTGACAA  
ATCCTGACAACCATGCGGGCCTTGACCGTGGACACTTCTCTCAAGGACTGGAACCTGTACCACCTTTGATGATGGGGCT  
GATGGCTATTGCAGAGCAGATGGAATTTGGCTCCATTGTGCTAATAAGAGACTCGAAGATGCTCAGGCAGACAACGATCCCAT  
CTACGGTGTATTGGAGGAGCTTACACCAACCACTCCGACAGAGCTGTATCAATTACGCGGCCCTCATGTAGGAGCACAGT  
CCTTCACTTTGACAAGCTTCTCAACGAGTCCAACAGCGACCCGAAGGAGATCAGCTATATTGAAAT

### KHKS1

GCGGACGGGTATTGTGCGGGGAGAGGGAGCAGGACTAGTTGTTCTACGACCCCTTAGCAGAC  
GCAATTGCAACGGAGACCCCATTTGCGGTAATTGGCGGGTCTGCGGTCAACCGGGA  
TCCAACGTCTCCCGATCACTGTGCCAGACTCGAACTCCAGAGATCCCTTTATCGGAAG  
GCCCTGCTGGCATCTGGGATTCTCCAGAAGACGTCACTTACGTTGAGGCCACGGCACC  
GGCACAAT

### KHKS3

GCGGATGGGTATTGTGCGGGTGAAGGGATTGCGGCCATTTTCTAAAGACTCTTTCTCGT  
GCCCTGGCAGATGGAGACCACATTGAGGGCATCATCCGAGAACCGGCGTGAACCTCGGAT  
GGTCTACTCGAGGCATTACAATGCCTAGTGTGACGCTGACGCTTCTTTGATCAGAGCC  
ACTTACAAGAAAGCTGGGTTGGATCCCTGAATCCAGATCAACGCTGCCAGTACTTCGAG  
GCCACGGCACCGGCACAA

### KHKS9

AGCGGCCCCAGAGTCGAGGTATCGATAGCTTGATGCGGACGGGTATTGCGGGGAGAC  
GTTTCAGGACTAGTTGTTCTACGACCCCTTAGCAGACGCAATTTCAACGGAGACCCCATTC  
TTGCCGTAATTGGCGGGTCTGCGGTCAACCGGATCCAACGCTCCCGATCACTGTGC  
CAGACTCGAACTCCAGAGATCCCTTTATCGGAAGGCCCTGCTGGCATCTGGGATTCTCT

## Appendix

CAGAAGACGCTCACTTACGTTGAAGCCCACGGCACCGGCACAATCGAATTC

### KHKS27

GATTGTGCCGGTGCCGTGGGCCCTCGAAGTACTGGCAGCGTTGATCTGGATTTCAGGGGATC  
CAACCCAGCTTTCTGTAAAGTGGCTCTGATCAAAGAAAAGCTGAGCGTCAGCACTAGGCAT  
TGTAATGCCTCGAGTACGACCATCCGAGTTCACGCCGGTTTCTCGGATGATGCCCTCAAT  
GTGGTCTCCATCTGACAGGGCACGAGAAAGAGTCTTTAGAAAAGATGGCCGCAATAACTTC  
ACCCCGACAATACCCGTCGCAATCAAGCTATCGTACCTCGACTCCAGGG

### KHKS32

AGGTGCAGCTATCTATAGCTTGATTGTGCCGGTGCCGTGGGCCCTCGACCATTATATATCA  
TTAACTTCCAACCTCGGCCTTGCCGACAACGTTCTGGAACACGATGGTCAGGGGGGACGGG  
TTGGGAACAAAAGATCGGAGTATCGTTTTGATTCTGGTTGATGGCGGTTGCCGAAATCACA  
CCCAGTATCTGATCGCCGTCGGCAATAGCATCGGATAATTTCTTCAAGAACACGGCGCCA  
ATAGCCTCACCCGGCAATACCCATCCGCAATCGAATT

### KHKS37

GATTGCGGATGGGTACTGCCGGGAGAGGGAGCAGGACTAGTTGTTCTACGACCCTTAGC  
AGACGCAATTCGCAACGGAGACCCCATTTCTGCCGTAATTGGCGGGTCTGCGGTCAACCA  
GGGATCCAACCTGCTCCCGATCACTGTGCCAGACTCGAACTCCAGAGATCCCTTTATCG  
GAAGGCCCTGCTGGCATCTGGGATTCCTCCAGAAGACGTCACTTACGTTGAGGCCACGG  
CACCGGCACAATCGAATTC

### KHKS42

TTGATTGTGCCGGTGCCGTGGGCCCTCGACGACCGAGATATCATTAACCTCCAACCCGGC  
CTTGCCGACAACGTTCTGGAACACGTTGGTCAGGGAAGACGGGTTGGGAACAAAAGATCGG  
AGTATCGTTTTGATTCTGGTTGATGGCGGTTGCCGAAATCACACCCAGTATCTGATCGCC  
GTCGGCAATAGCATTTGGATAATTTCTTCAAGAACACGGCGCCAATAACCTCACCCCGACA  
ATACCCGTCGCAATCAAGCTATCGATACCTCGACTCGTGGGGCCG

### KHKS44

GGATCCCCGGGGTGCAGGAATTCGATTGCGGATGGGTACTGCAGGGGTGAAGGGATTGC  
GGCCATCTTTCTAAAGACTCTTTCTCGTGCCCTGGCAGATGGAGACCACATTGAGGGCAT  
CATCCGAGAAAACCGCGTGAACCTCGGATGGTCTGACTCGAGGCATTACAATGCCCTAGTGC  
TGACGCTCAGCTTTCTTTGATCAGAGCCACTTACAAGAAAAGCTGGGTTGGATCCCTGAA  
TCCAGATCAACGCTGCCAGTACTTCGAGGCCACGGCACCGGCACAATCAAGCTATCGAT  
ACCTCGACCC

### KHKS46

GAGGCCCTCAGGTGAGGTATCGATAGCTTGATTGCGGATGGGTACTGCCGGGGTGCTAC  
CCTGGGCGCCGTGTTCTTGAAGAAAATTCAATGCTATCTCTGAGGGCGATCAGATACT  
GGGTGTGATTTCCGCAACCGCCATCAACCAGAAATCAAAACGATACTCCGATCTTTGTTCC  
CAACCCGTCCTCCCTGACCAACGTGTTCCAGAACGTTGTCGGCAAGGCCGGTTGGAAAT  
CGAATTC

### KHKS51

GGGCCCTGAGTCGAGGTTTCGATAGCTTGATTGCGGATGGGTACTGCCGGGGAGAATTTT  
CAACGTTCTGTACCTGAAACGTTGTGCGGATGCAATTCGGACCGGTGATCCTATCCGTGG  
AGTTATTGAGGATCTTCAATTAAGTACTGCTAAGTACTCATATCTGCCATTATGCCCCT  
TTGGGCGTATGCTGAGACACAAATGCATACAGCAACGGCAAAAATAACGGAATCACTCTC  
CCAGTGTCTTTGCTCAAGAGCTGTCTATCCGAAAGGCATACGAGCATTGAGGCCCTCTT  
GACTATAATTCAGTTGGTTACGTGGAGGCCACGGCACCGGCACAATCGA

### CMeT-3

CCATCTGGCCGCCGGGCTTCAACAGCGATCGAAGGTTTCGATAATGCGTTATGCAGGTTTGCAGTAGCGTGAATGACATTG  
CCAGCAAATATCAGGTATATGTCCCCAGCTGGAATCCCTGGTCGACGGGGCTTCGCTCAATATTGAGCGATTGGAAAGTT  
CATGATATCGGACCATGGCTGGAACGCTGTTTTGCTTTCTCGAAGAAGCCGGATGAAATGTCAGTGAATGTGTAGCTGG  
CAATGTTTCCCTTGGTCTTTGCAAGGTCTGGCGAAAGGACGGCCAGGACTTCAGCAGTGAAGCTTCCCGTCCCCCCCC  
ATCTCA

### CMeT-9

GAGATGGGGGGGGGACGGGTGGTGCCACCTTGCCCTATTTTGCAGAGGTTGGGCGGAGGCGAAGGAGGATCTACGCCCTCG  
GTTTCGGCCACTACATCTACACCGACATCTCTCCGCTTCTTTGAAAAGGCAAAGACAAAAGTTTCGAGAGCTGGGAAATC  
TCATGACGTACCAGGCACTAGATGTGTGCGGATGACCCACTGCCAAGGCTTCAGTAATGGCACATACGACGTTGTGGTT  
GCTTGCAACGTGTGCATGCTACTCCCAGACATCAGCCAGACAATGTCAAACATCCGAGGCTTTTGAAGCCCGGGCCCA  
GATGGTA

### AT-9-11

TGTAGTCGGGACAGGTTTCGGGAACTCACCGCACTTTGCGTATCAGGCATCCTCAGCCT  
GGAGGATGCGCTCAAGCTCGTCCATGGCCGGTTCGAAGATTATCAAAGAGAGCTGGGGCC

GGAGAAGGGTCCATGATGTCAGTTGAAGCGGACCAGAAACGATGTGAAAAGCTCTTGGT  
TGCTTCAAATGCGCGTCTTGGTGAACCTGAACGGGCTGGACATGCGACAATTGCCTGTTT  
CAATGGCCCGAAGAGCTTACCAATCGCTGGTTACGCCGCTGCAATTGATGCTGTGCAGCA  
GACAGTTTCAACCCTTGATATCCCAATTAACACAAGAGGCTAGACGTGACTAATGCATT  
CCATTCGACCCTAGTTGAGCAGCTTGGAGCCACAACCTGGAGGCCCTGGGCCGAGTCTGAG  
CTTTGAAATGCCACATACCTCTCGAAAGGGCAACCGAGCAGCGGGAGACCGGTCCGAT  
CTCACCTGCTTACGTTGCTGAGCACATGAGGAATCCTGTCTACTTCGACCATGCCGTCCA  
AAGACTTGCAGTCAATTCGGAGGCCATTTGGCTGGAGGCAGCGAGAAA

## 16 Nucleotide sequence of shot-gun sequenced 42H12 fosmid clone

TCACACCCTCAGCAGCAGCGGGCAACCGCCTCCCCATCCAGGTCCTGACCGTTCTGT  
CCGTCACCTCCAGATCCGCGCTTTCTCTGTCTTCTGTGCGACGGTTACGCCGCTCCA  
TGAGCTTATCGCGAATAAATACCTGTGACGGAAGATCACTTCGAGAATAAATAAATCCT  
GGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGAT  
CGGCACGTAAGAGGTTCCAACCTTACCCATAATGAAATAAGATCACTACCGGGCGTATTT  
TTTGAGTTATCGAGATTTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAATCACTGGA  
TATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGGAGCATTTCAGTCA  
GTTGCTCAATGTACCTAYRRCMGACCSYTYAGCTGGATAATTACGGCCTTTTTAAAGACC  
GTAAGAAAATAAGACAAGTTTATCCGCGCTTATTCACATTTCTTCCCGCCTGATG  
AATGCTCATCCGGAATTTTCGTATGGCAATGAAAGACGGTGAGCTGGTGTATGGGATAGT  
GTTACCCCTTGTACACCGTTTTCCATGAGCAAACCTGAAACGTTTTTCATCGCTCTGGAGT  
GAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTAC  
GGTGAAGAACTGGCCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTCTGCTCAGCC  
AATCCCTGGGTGAGTTTACCAGTTTGTATTTAAACGTGGCCAAATATGGACAACCTTCTC  
GCCCCGTTTTCCACCATGGGCAATATTTATACGCAAGGCGACAAGGTGCTGATGCCGCTG  
GCGATTCAGGTTTCATCATGCCGTTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAA  
TTACAACAGTACTGCGATGAGTGGCAGGCGGGCGTAATTTTTTAAGGCAGTTATTGG  
TGCCCTTAAACGCTGGTTGCTACGCCCTGAATAAGTGATAATAAGCGGATGAATGGCAGA  
AATTCGATGATAAGCTGTCAAACATGAGAATTGGTCGACGCGCCGGGGCCGCAAGGGG  
GTTGCGGTTGGCCGATTCATTAATGACGCTGGCAGCAGGTTTCCCGACTGGAAAGCGG  
GCAGTGAGCGCAACGCAATTAATGTGAGTTAGTCACTCATTAGGCACCCAGGCTTTAC  
ACTTTATGCTTCCGGCTCGTATGTTGTGTTGAAATTTGTGAGCGGATAACAATTTACACAG  
GAAACAGCTATGACCATGATTACGCCAAGCTATTTAGGTGAGACTATAGAATACTCAAGC  
TTGCATGCCTGCAGGTCGACTCTAGAGGATCCCACGTCCGAGGATGAACGACACAATCT  
CAAATGACGAGTATTTGGTGGTCTTTCTTTTGGTTAGTTTGGTGTCCGTTAGGCAT  
TATGGGGATGTTCTTTTTCTGTTTCAATGTATAGCTCATCGCATACTGCTAGTACAGT  
ATTTCTGCATATGAAGTAATTTGATCACCCTACTGAACATAGCAATATGCTTTGTAA  
GGCCATTAAGGTTATAGGGATGGACCCGCTAGCTCACCCTCATTGTCGTTGTCGTTGCT  
TGTCGTCAGTATCAGCCTCAGTGGTTTTAGTCCAGTTCTTCTCCTCTGCTAGTTTCTTC  
GTTTCTCCTTCTTGCAGTTTTCAGTAGCCTTAGCAATTCATCGTCCACCCTTGTCAA  
ATCTAGCCACTTGCCCTTCCAGAAGCTACATCGGTCACTTGGGCATTGCTCCCTCTTAA  
GCTGCGTAAGCCAATCATATTAGCATGCAGTCTATCTCGGCCTTTCTCTTCTTGGTA  
TCCCGATCTTGGGGAACCCCTACCCTCCGATATACCCAAACAGAATACGTTTATATA  
AGGATACCAGATCTAGAATGTCTATTTTTTTAAACCCGTATATAGATAGAGAAATTCATT  
GGTCTATATATCTCCTTATAAAGGAAATTTAGTTTATGTAACCTCTATCCACTAGTTATT  
GCCTAAGGGCAGGGAAGTGAATTTCTCACACTTAGGACCAAAAGACCATAGCAACAATCG  
GAACCCCGATTTACGCTGTAACGGTGTGTATATCAGGCCCCGATCTCCATTAGGGATC  
TTAAGCAGTTATGGCGTAATCCATGTGGGAGCTGACCACCCAAAAACCGTAATAAGCG  
ACATATTTACGCTTAAATCTTAGAAGGTTCTTATCTTGTTCATTGTTGTGCGATAGTT  
CAGCCTATATGTTAGTTTGGGACCTGGAAGCTATATTGCCACGATACCAGCTAGAAAATA  
TATATCTTAATCTCAAATTAGAAAAGAAAAGAACATCTTAAATTTCCAAAGGAGTTA  
GATTTTGGCGTGAAGACCATTAGGCTGTAAGGAACAGTTTCTATAGGCAACCGTTGCTGCG  
ATATCATGGTGTGATATATCGCATGTGAATTACGTTATGGTATTACATCGGTCCATATG  
CGCCTCGAACCTGTTTGGGAAGTTTTTCGTATACGACATTAATCAGCCTTCTACTCCCA  
AATATTGGGTAGCATGTGGATGGATGGATGGATGGATGGCTTGGAGGATGCTGATTCTG  
CCCAATTTGGTAAATATCTTTATGCCAGAGCGGATCACATTTAGGGTCTCGCAACTGTGA  
TGCGGGTACTGGAGATTAACACATACATGGGCCTGCAGTAAGAGTGCCTGAACTATT  
GAGTCTGATAGGCCCCCTTAAAGCATCTCAGCCCTATTACTCCGTTGGGCTCTTTTCAT  
TTCCCAAACCTCGCTTTCTGTTCTCCGAATATGTCGATTCACTTCCGTTCACTCCT  
GGGGGCCAATAGTCAAGATGGTGACCCAGGAGCCGTACAAAACCTCGGTAGTTTCGTACA  
TGTGCTTGTGATGCTCAGCAGCAAGCCGTTGAGTGAATGCTATCGTATCAATGG  
GATGGAATTTCTGAGAGCGGGAAGTTAAGAGATGGGCGGGACGGAAGCCGTGTAAGATA  
GATTAACGCTGAGATTTTAGCGTATTCTGATTCTGAAATTTGCTGACATGCAAGTACTG  
CTGAACCCCGCTCGGAGTCGGATAGCCAGAGAGTCAGAGAGACCAGAGAGACCATAGC  
GGGTCAGTTATGCCCAATGTGACTCTTTCTCGGAAGAGCTTAGTTTCCCAAGTTTGG  
CCTCTAAGACTTGGGATTTCCCATCTCCCATGGGGTATTACACTGAAGGCCGTATA  
AGTGTAGCCCTTTGCCGCGGATACCAGGATGGCGAAAAATAAGGTGCCAAGAAGGGGCGC  
TGAGAGGATGATAGGGTGTAGCCGCAATCCAACGGGACGCAAGCAAGGATCAAGGCTTGT  
TATAGTGTGATCAGTCAATCTCAAACCGTGAAGGCTCGAGCGGAGGCGCCTTGAGT  
CTTGCCGTCGCTTGTTTAGGAAAATCTACAAGCCACCCTTGGGTGGCATTTTTCGTCA  
GCCAAGAAGCCAAACCTAACGAGAGGATGATCGTTGCACTCACTTTCTGGGATTACGGG  
AAGATTTCCCGTCTCGTCCAAACCTGTGGCTTCCAATGTTGTAACAGTTTGTATAC  
TAGGTGTGGTCTTGGCAAGTTAGTGGCCACTGCATATGATAGGCTTGGAGCTTCACTTAT  
TTCGGAGAAAGCACGGCAATCGATAGTGGCATCGAATGTTGGTGACTTCTCCGTTGAA

Appendix

TCAATCGCAACCTCCGGAGTACAGAACAGCTGGAATAAGTGGCTCCATGCCCTCGGCTA  
TGCCAAAGGGTTCCGCCTCCCTGCATAGCGAGCATATCTAGAGTGTCTAAGAACCATA  
CATCGCTTACCTTTTCCCTGTTCCAGCTAATAGAGAATAGGGGACTGGCATCGTTGGAAC  
ACATTATCATCCCTGTGAGCCAGTTTGGCATTGATTTCGGCCCTGGCGGAGCGCCGAGAAT  
CATGCACAGGAATAGGGCGTCCAGATATATCCCTGTCTGCCAGGGTCCCTCCCCACGGAC  
CTCCCAAACCTGGGGCCGTCGACAGCCGCTGAGGAGGATTCAACTTACGCTTGAAGCTTGA  
AGAGAGCAATTCACATGTAAATTGATGGTATAACTATTGGAACATGCTCTCGCTTCGCTC  
CGTTGCCAAGTCTTCAAGATAAAGGATGCAACATCGTAATTGCTTGGTTGCGCAAGCCAA  
GTGGCTAAAATATACCTTTGGCAATACGAGGCATTGAGGGGATTTTTGTTCGGAGGCA  
GAACAGCCACCAGGAAACAATGGACCATGAGGAGGGTCTCCAAGGTTGTCCATTGGA  
TGCCAAAGGGACCTGGCGATGACCAAGATGGGGCCGATGCGTATTGTCCGAACACCTGGCG  
CCGCAAAAGGGCTAGGATAACAACAACACACGAGGTCTTCAAGGTTCCATAAAGGGTTCGCGA  
TCCTCGATCCATTGCAAGGTTTGGTTGTTTTTTCTTTTTTTCTTTTTTGGCCTGAGCGTC  
CAATACGAGACTGGGGTTTTCGCGATTGTAGTTTGTATTACTATAAAAATCTCCCGCTCGAGC  
CTGCAGTCCCCATCTGCATATCAGGTAGCCAGAGACTTGATATAACCTCAGCAATCGA  
CTAACCCCTCGCACACTTCGAGAAAAAAGAAAAAGAAAGATGAGAAGGCAACAGTCTGGG  
CTGGCTTTCGAGGAATGTCTGACGGAAGCAGATGCGATCGTGTGCGTCCCTAGTGC  
GGTATATCGCCGATGCAAGGCGTACTTGCATCGTCTGATAAAAAGATCTCCACGAGGC  
CCCAAGAAGGGACAATTAAGGATTTGCGATATCGGTTAGGTGAGTCCGGTCTAGTTGAA  
GAGCTCTGGCAATAGTGTCCACTGACCTAGACGGGCTACCACGCGGTGATCGAACACA  
ACTGATCAATCAGAACCAGTCCCTTGCAGTAGTCACATATTACCAGACTCAAATAGAGG  
AAGCCTGGAACGAGAAGCGTCAAATGAACTTCACTGTGAAGATGGCACGATAGGATGCGA  
CATGGATGAGTTGCAAAATGTTGAACTTTGCTTTCATCTGAAAGTATCTCCGCGTCTGACGG  
GCAGGAGGCAATGACTGGACGTTGGCGTATGATTGTGCGGAAGAGAGATCGGGTACTTG  
GAATCATCTCGATATAAGCAACCCTCGCCATGTATCATCACTCCTATTTTCGCCATTCA  
AACTTCTGCAAGCGAGATGACGCCCAAGGGGAGCTGGATATGTCCGACTCATGCAAGC  
TGATCTGTATGGAGTTATGGTCCCTAAAACACAGGGTCTGAAAGTGATATTACTGCAT  
CTTTATAGGGATCTACTATATTTTGGAGGAGTCCATCCAATTTGCCGATGATACACAAA  
CGACGCTATCTTTCATGGGCCAACGAGAAAACGGTCTCACCTGCCGAGCTTGTCTGCGC  
TCAGCCATGCGCACCATCGCTGCCGCCATGTCCGCTCAGTCTTCCGCGTTTTCCGATAAA  
CTCTATGCCTGTACCCGAAGTATGCTGGAGATGCAAGACGTACAGGGCGAGAACGGTTTG  
CCATGGATGCAACACCAGAGAGCGTCTCGCCGCGGATTGAACACGAGATGATTAGGCA  
TGGCTCTTACTTGGCCATTGCGAGTCTTACGCAAGCCTGAGCAGGACGCCCTCCTTGGC  
TCAACCAGGGCGGTTAGGCTTCTCCAGCTCTCTCGGCTATTTGATATTGACATGCACGAT  
GATGAGACTCGCCGAATGAGAACTCAGGGTCTGTCCGTGAGTTTCTCCTTCCACGACG  
CAGAATCTCCAGATGAAGCGTGGATAGAGACCGGAAGAGAAGCGGCAACGCTGTGGACA  
GCCTTTGCTCTGACTGTTTATCCAGTATGCTTAGCGATCGACCGTCAATGTTGCACGAA  
GAGATGGTCCGGTGAAGTCTCCTCTCCTTCTTCTGTGACAATAGACAGTTTCAGCTAACAA  
GATTGACCCTACAGATAAACACACGCTTTCGATGCCGAGAGGGATTTTCAAGGTGGCC  
AGCGACCAACCCTATGGGCTTCCCTACCCGAAACAATGGGCAAGACCGGTGACTGCGAGA  
CACTGTCTCTTTTGGCCAGTGCCTTGTGCTCGCAACCTTTTCCGGCGTTGCATAGCAC  
ACCGCGCGCTCGCACAGTCAATATCCTTCCGGAGTCTGGCTCCGAGTCAAAATCGCGGC  
AGTTTTGGATGAGGCACGAATGGCTGGCCGCGGAGCCGCCATGCAACAAGGACGATGC  
CACCGCACAAAGCCCAATGAGTGAATCTGAGACCACAAAGTGTGATCCATTTGGCTG  
CTTTCAATCGTATCCTCGCTACAGTGCATGCATATCCCTTAGCGAGACGGCCGAAGCAA  
GGGCATGGGAGACCCCTCGACGATCACACGCTGGCTTTGTGATATAGACAGGTAGCCAGCC  
AAGCCGCATACGAAATCGCTCTACTTATTCAAAAGGCGCTCCGGATAGCTTTCTTTAAGA  
TGATCCCTGTCCCTTCCCTAATGCAATATACTTGGCTGCCAGGTTTCTCAGAACTACGACGC  
CACACTTCGCGACTCCAGCAGAATGACCACAACAGCATTACCATCTACTCGTCCGCC  
TGGGTTATCTTAGCAGTGTCAATAATCTGGCCCGAGACCTGCTGGTCAAAGCCGAAGCAG  
ACGTTGGGAAAAGTGCACGGATAGCAACTGAAGTGACGGGGTCAAATTTGGGATGCGTTAA  
TGGCGGATACATCTAGACATATAGGCATGATGGAACAATCTATATCAGTGTAAAATCCAA  
ATAAACTCAAAGCCATTACCGCGTATGTCTCCCGGTCGCCAAAGCTGCCTTTCAATTTG  
TCTGAGGTAGATGTGAAATATAATATATACGTTATCATCGTGTCTTTCGGCTTCCCTG  
GCAGGTATAGATCTAGATGCAGTTATCTGAACGTGCTTTCGGGATTCAGAGGTGCGAAA  
TAAACCCGAACCGCAATGAAGGCCCTTACCCTGCAGATCCAAATCGACGATAGGATGAC  
AGCCGCGTCTCCGATAAATCCCCAGAGGCTCAGAATGCAGTGTCTGCGCCCCCTAACTCG  
CTCTCCCAAGATAGCCAAAGCGTCCGTTACCATAACCAGAGTATCGTGTATCCATAATTG  
ATGACTACGGCTAATTATATCACGACTCAGCTCTTACGCTGCCAGCCCCAGTACGATC  
CTTCCAAGTACACAGTTACCTGTTTCCCTGCTACTCCGAACCCAGCTGTGACCTCTC  
GTACCGTACTCTACTGTACACGGCAATGCTTCCGATGCCACAATACCTGAGACATTCCG  
TGCTTCCGCCAACACAGCAATACCTCCGCGAGAACCAGAGCGTCTTCTATCCCAAATCC  
AGCTCCGCTCCTTGGTTTGGCGTTGAGGCATGTGCCGATCCCCAGCTAAGCAAACAG  
GCCCTTGGCAAACGTCGGTACCGGTGCTTCTAGCATATCAAACACGGCCATTTTCTAG  
CTTCTCCGGGAACATCGACACAGCTGACCCGACAGTTGGACAAAATTCGGCAAATGCTTT  
GGTTGCCCTCGTTCCGCGTTGCCGGGAGTACATATCTCTTGTATCGCTACTTGTTTGAAC  
CTCAGGCCATTTCTTCTGTCCATGATGAAAGCCTCGATGTGCATTGCATGCACCTCGGC  
GAGGGGATAGTGAATACAAATGCGCTGACCCAGTATGAAGTACCCGACTGAGTGTCTT  
TTCTGGGCGGAGCACAGCGGTTGCTTGGCGTAAAGGCACCATAACCAGGTATCCTAGTTG  
GTGTGAGTACTGGGCCCGAGGACAAAGTTCATCGGTTCCGAACATGGACGCGCGGACTTG  
AGATCTAATCCATCACAGCCATGACTATTGTATTATTAGTACCGCCATGTTATTCAA  
CAACCCGACGCGCTGATCCATTAACTCACAACATCAGCATGGGCAATTTCCGCATCC  
TGAACCTTAGCACTGCCCTCCCGGATTCATGATCTGTCTCAATCGAATCTAGCCATTCC  
GGTATTGAACGAGTCTTTCGGGATATGTTGGACAATTTGGTGCAGGAAGTCCGACCCGCC  
GACACGCTTGTAGCTCCCTTTACGTGAAGTTGAAACAACAATGCTGCGGTTGATGTGC  
GAGGATCTTCCGCTGCTTGTCTCGAACCCGCTCCACAAAATTTGATCTCCACCTGACG  
GCGCACTCGCCACATTGGTAAAGCTTTGAAAGCAGCATGGATCCAGCGCTCCATGGCGC

Appendix

GCTGCGCTTGGGTGTGAAGGTGATGCCGACGCCAACCTACCAAATTTCTCGGCTCGTT  
CATAAATGGTGAAGTAAACATTCGCTTCAACAGCCCGACAGCGAGGGCTAGGCCAGTGA  
TGCCCGCCGACGATGGCAACCTCGAATGGGCTGTTGGCGGATCGAGTCATATTTAGGT  
CGGGTCGCGGATAAACCCGATGGGTGGATTTGAGGAATAACGACCTGGGCTAGATGAGTC  
TTATCATCCTCGATATTTGCCACATCAATAATCTTCTAGACGGAGTTTCTGCCGCTTCAA  
TGACGTGGTTGGCTTTGTATCGGCGGACTCCGGAGTTTTTGTGCAACAAACCCGGAGCCAC  
ATCTACTCCGAAGTACGTTGTTTTGAAACTTGGCGGGGCCATAGAACAAGAACAACAGC  
TGATATGCGTTTTCGAGATGGCGTCAGGTCAACCTCGTGTACGGTGGGTCTGAGTTATTGG  
AACAGTATAACCTTACATAACAGATCTCCTCTGCCGGAGCTTTCCTATCTCAGTCACCA  
CTGTGAGCTCAAAGTCAGGGCTGCCAGTGCACGTGATAAACAGTGTGGACTATGGTC  
GAGAATATAGGCAAAATTCACATTTTCTAAGGTATTTGAGTGTAGCGCCTGCCTGGTTG  
ATAATTAACCACCGAGGTAGATTCTAGGGTGTAGTGTAGATATGCTGCAAGAA  
ATGGGATCGCATCTTCTTCTATCTTCTCAGCCTGTCTGGTAAATAAGTCAGCCTTTT  
CACAGCCAAATCAGGTGGAAGTTACCGAGGAATACATCAGAGCAGTTTTACAAAGATCA  
AACGAAATGAAGCTTGAAAATACAGATTGTAGGCGAGAATATATCTAGAGTAGAGCTAT  
CATTCTATATAATCCATGGAGGTAGACAAATGTGCAACTAGGATTAAGGAGGAAAGAAACC  
ATCAACGACGAAACCCAGTGTCTTCCAAACCTGGAAGTATTTATAACTACCGCGTCAG  
CCACCGGCTGCACGCCTCTCAGTGTCCGCGAATGCTCCACGGCATTTATTCGTGGCCATGG  
TGACCCACCCGCAAGACATGTGCTGGAACCTGTGCTCAAGAAATCGATCGCCTTGCCCG  
CGGGTTATCCACGGGCTCCACAGGGAATGCTTCCACCGTCTCAGCCAGTCATCTA  
GTGGGACAATCCCTTTCTCTGGGATGCCCTAGAGCCTGAGCTAGTACAGGGACGATCTCTG  
CCCAAGGCTTCCGACCCGGGTTATCCACATGGTAAACGGGATGGCAGTCCGGCGCATCAC  
GTAGTAGTGGTCTGTAGCGTAGCCGAGGCTTGTCTGAGGGCAGCCAGGTTAAGATGC  
CTTCCACCGCTGGGAAAGAACGACGAGGCTGAGCGGACTTGAACAGGAACCCGAGAACTT  
CCATGTGGTTCCAGTACCCAGTCTCCATGGATCCTGATAGCTGACCCAATCGGACTGTCA  
TTGCGCGAAACGCTCCGGGTACTGACCAATGTCTCGTGGAGGATCCGCTCACATATGA  
CTTTCGCCCCCGTAGCCGTTCCGAAGAGCACTGTGATATTCAGATGTTCTTCCGGAA  
CTTGAATCTCGCCGCGTGGACGGATGGGTATTTACCCACCGGACAGGGAAGAGATGA  
ATTGGAAGCGAACTTGAAGTCGGTGGCCTTGGCGGAGCTGAGATCTCTGCTGCAAGGT  
CCACAAGATTGCCATGATCGTGAACCTGGGCTCATTTTGGCTTGAAGAGCGCAGGCCGT  
TGACAGGAAGGCATTTGGATAATGTGAGTGCAGTGGCAAAGTAAATGTTTGTACTCCT  
CTGTGTCTAGACCAACTGTGCTTTGGAGGAGTCACTCTCGATTACTTTCAATTTGGCGA  
GTTGTGAGGCATCTAGGGCAATGGACTTTGACTCCAGCGATTGAAGTGGCGGTGCAATG  
GGTCACGAGGGGTAGCGTCTTACTCTCGCACCTCCGATCGCCGGTTCGATTCAGGCAGT  
AGACCCGCTCGACCGAAGGAAGAGAGACTAAATGGGCAATAATGTGGCTTCCAAGACTCC  
CGCTGCTCTCGCCGCGTGGACGGACGCCCTTGCCTACTCAGAAGAGTCACTGAGATCTCCGG  
AGTAGGCAGGGATTGTGAAGCCTCGGATAGTTTCTTTGACATAGTCATCAGCAACCTGCT  
TTTTCTCGTTCAATTCCTCCTCGGATAGGTCATGGTCTCAACTTGTGTCTGGCGACAG  
GGGAAGACGCAACCGTTCTGCTCGACATCGTCCGCCAGAGCGATTATGACCCGTTGGA  
CAGCACTTCCGGTACTTCCCATCGGTCACACTCAACATGTTTTGATGCTCGAGTACGAGTA  
GCTCCTTTTCCATTGGCGTTCGTCTACAACGGCGTGGTGGCTCCGCTTCAAAGAGCC  
ACCAGCCTTCCAGAGTACCCCAACGACGCTCCACCCAATGCAGCGTGTCTTCAATTTCCA  
GCAACATCAGAAAGCCATCTGGGCGCAGGAATTTGGGATATTTTGGCTCGAGACTTGGG  
GTGAGTGTGATGATGCACAGCCTTACTAGCAATCACAATGTTTTGCGAGCCATCAGTA  
GCGGGTCGGACGGAGGTTGTTCAATGTCTGTCACAGCGAACTTCAATGAAGGGTACTGTT  
TGAATCTTCTTGGCTTGGCCACAAGAGACGGCGAAAGGTCGGTAAAGGTGTACTCGA  
CGGGGATTTCAAGCTTGGCCAGCGCGGGAACGAGAACCTTGGTTGTCCCGCCAGTCCCAG  
CACCCATCTCAAGGATCTTCAAGGGACACGATCTGAGCGCAAGTCTCAAGCTTTCTG  
CAATACGGGAGAGGAAATCAGCCATCAGCTGGAATAACAGCTTGTGAACGGCAGCTCGC  
CATAGAAAGCGGCAACTAGTTACGATTTTGGCATCTCCAAAATGAGTTGGGGACCAT  
CGGCCTTTCCGCAAAAACGCTTGCATCCGAGAACCTATATTTGATACGTAAGCTGGTGGG  
AAGGGCCATCTGCTCCGGTGGTGGGACATGAGGTCGTCCAATATGGCTTTGGGAGGACTGTG  
TAGGGAGAGGAAGGGCGGTGCGAGTAATGATGCCTTCGTCCACGTCATGATTTCTGCTCT  
CTTCGAGCATCTTGTACAGTATTCATGGAAGCGGTGATGGCGAGGAACAAGGGTACGG  
GCTGCAGCACTTCCCGGCTTGGCGGCCACCAGGTCGAGCCAGCTGCTTGAAGGCAT  
CGCTTGTAGACGACGAGGACAGTCTGGTTGCTTCTGGGAGGCACCATCAAGATAGCCGG  
CGCATTTCATTTCTTCAAGTAGGCATCGGCTGCTCATTAGCTGCGCGAAGGCCTCGA  
TAACGGCAAAATGCCGGCAGGCCACCTCGTTGCCAAGTTCGATATAAGATCTAGAAATGT  
CCTCTTCTCCAGCTTGGGGCTTGGCGTGGCAAGAATCCCGTCTGTGCTGGATGGAGGAG  
TGTTTCGGGACGAAGAAGCTGCATCCGAAGACTGCGAGGCATCATCTCACCCCTCAAGCC  
CAAGTGTAGACTGAAGGAAAGCGAGAATGCCTGGTACGTCAAAGATACTCATGAGTTCAG  
ACTGTTCCAAGGTGCAATTAAGGTGGTTTCGACCTCGCGGGCATCTCCATTCCCATCA  
GAGAATCAATCCCAATGTCTGCCAGTGCATCGGTTTCAATTGATTTCTCAGGCTCGAGTC  
CAGAGATATCTGCCAGGACGGAAGAAGCTTCCGCCATAGCTCAGTCTTGGTGTCTCTCT  
TTTGAACCTGAGCACGAGCGGGCGGTGATGTCTCATAGATGAATACTGGTTCTCAGTTG  
GCGCAGAACTGATGGGAGCAATTTAGCGGGGAACGGTACAGGCGGTGCGGCAATAACAA  
CGTTCTGTGAAAAGTTCACCTCTTAAAGCTATATCAGAATCGTGATTCGTAGGCGGAATAA  
AATACTACATGATCTAGCACTGAAATTTGACGCAAGTGAAGGCTCGTCAAGTACGCTA  
GATGCCGAATCGAACACAAAATGTCTGTTATAAAGGTGTCTCCTTCTGTGCGCTTGTGT  
GTCGCAAGGATTTGCCATTCGCTTTGATGGTCCGCGTAAGATCCTTCGCTGATTTCCCGC  
AGGAGACTCGGTGACCTCATCCATTGCTCAATGCCATCGGCAATATAGACGGTGTCTACT  
GCGGTGTTCCGGCCCGGAGCAAGACAGTTCACCCAAATGCTGCCAACCTGGGAAAAGTC  
TCCCGAGGGCGAAGTCAAGCCAGGAGTCCCGAGACCGTCTCACAGCTCGGCCAGCT  
GACTCGCTCGGTGCGCCAAACAGCCTTTGCAGGCGCGGAATTTGGGGCGTAGCTCACG  
AGGTCCGAGTAACTTTATATATGCTCTGCCCTTGAATCACTTCGTCAGCATCTGCAGCA  
GACTCCAACCGCCGAGGCATCGTTTCATGCTTACAAGCGCTCCAGCCTACTAAACTCG  
AAGTTTGAAGCGGGCTCATCGGCACGGTGAACCTAAGCTGTCCGCTCAGATGCACAGTC

Appendix

TTGCTAGAGTCGTCGGCTTCACTGGTCAACTTGAACCTTCCAACCTTCAAGGGGCATGTCCA  
GAACGTTTCGAATTAAGAAAATACTGCCCTTGTGGATCCATCACAAGCGGGAGGTGATTC  
AAGACGTTATAAATGTGGGATGACAGCTATTGGCCGAATGACTTCCAGTCGAACGCTT  
GTGATTGCCTGGATTGCGGTGTCAATCCCGAATATGGGCGACAGGCCTGGACAGTTTTT  
CCCAATGTGTATCCTGAGACAATATCGACATATGATTTCTGGTGTATTTGATGCGGAAC  
CGACAATCGGTCTCAGTCTTGTGCGCGTATCCCATGAAAGTGTAGAGGCTGGTGCCGGA  
GCTTCTTGGTCCACACCACCGTTTTCATACACCAGTCGCTCGATTACCACCGGCTTTGGG  
GGCGGCTTAAACTCGAGCCAATGGCGATACTTCTCGAATTGGTACGGTGGGAGCATGATG  
GGTGCATATCCATAGGTTTGCGCCGGGAGTGTGGCCAGAACGAGCAGGGCAACCAGCA  
TTCCACAGACTCATGGTACGTCGGCGAGTGTGCGGGTCCCCTGAGTGGTGCAGTGACA  
CTGACTGGCTGGAAGGTGGAACCTTTCGGCATCCCCAACGCGCTGCCTGTCATTGTCGTT  
ATGGTGGAAATTAGAGCCAGCCTCCAACCAGATGGCCTCCGGATACTGACTCGCAAGTCTT  
TGGACGGCATGGTCAAGTAGACAGGATTCCTCATGTGCTCAGCAACGTAAGCAGGTGAG  
ATCGGACCGGCTTCCCGTGTCTCGGTTGCCCTTTCGAGAGGTATGTGGCATTTCCAAAG  
CTCAGACTGCGCCAGGCCTCCAGTTGTGGCCTCAAGTGTCACTAGGGTCAAGTGGTGAATGG  
AATGCATTAGTACAGTCTAGCCTTGTGTTTAAATGGGATATCAAGGGTTGAAACTGTC  
TGCTGCACAGCATCAATTGCAGCGGCTGAACCAGCGATGGTGAAGCTCTTCGGGCCATTG  
AAACAGGCAATTGTCGCATGTCCAGCCGTTCAAGTTTCAACAAGACGCGCATTGGAAGCA  
ACCAAGAGCTTTTCCACATCGTTTTCGGTCCGCTTCAACTGCAATCATAGACCCCTTCC  
GGCCCCAGCTCTCTTTGATAAATCTTCGACCCGGCCATGGACGAGCTTGAGCGCATCCTCC  
AGGCTGAGGATGCCTGATACGCAAAGTGCAGGTGAGTTCTCCGAATGAGTGGCCAAACCAAG  
GCTGCTGGTTCAACGCCCAATCGATCCAGCTCAACGCACATGAATACTGCATCGGAAAGG  
AGCAGCGGCTGGAGTACGGCCGGTCAAGGATCGGCTCGCTTGGGAATATTCCCGGGTAG  
ATACTACCCGCCCCGATCGACTTGCAGACACTGTGCGATCGATCGAGGTATTTACGTAGC  
ACTGTGGCTTTGTCAAAGACTCCTCGGTGAGGCCAACGGATTTGGACACCTGGCCCCCG  
AAGCACAAGATAACAGGTCGTACGGAAGGCACCGCAAAATGTTTCGAAAGACGCCAGCTTT  
TCTTCGAGTTCAGTGTGACTCGGCGCGAACACGAAGCCCCGGCTAAGCGACACAGTTA  
GATTGCGGATTGACGTTGAATGACAAGTTCCTCAATTCCTAGTAGCTCTCGAGATGACC  
TTGTTCTTGATGAATTGGCGCAGTCTGTGGCATAGGCCCGAATGGCTTTGTGCTGAGC  
CCAGAAATATAAATGGACATCTGAAAGTGAAGTTCGGCGATGTTAGATCCCGCCTTCCG  
TGACCGACTGCTTCACTTCTGAGGGGTATTTCGGTGCCTGCTTAATCACCATGGAAGCA  
TTTGGCCAGCTGCGCCATAAATGTTGATGAGTGTACCTGGACTCGTCTTCCAAAGGG  
AGGGCTGCTTTCGTGATCTCCATATTATCCGCAGGTGAAGCCTTGATAGACGCGCTCATG  
GAGGTAAGCTGGCCTGTGGGGTATGCGACTTCTTGTGATCATCAACAAACATCTTGATC  
AGTGCTACAACCCAGAGGCACCTCAGTATGACCAATCAGCCCTTCCACAGAGCCGAGT  
TGAAGAGGCTTTAGCCCCGCGCGGACCGAGCCACCAAGACCTGGCCGAATGCTGCTG  
TCTGCTGGGTACCAACCGGAGTTCCTGTCCCATGAGCCTCGACGACCGAGATATCATT  
ACTTCCAACCCGGCCTTCCCGACAACGTTCTGGAACACGTTGGTCAAGGAAGACGGGTTG  
GGAACAAGATCGGAGTATCGTTTGTGTTGATGCGGTTGCGGAAATCACACC  
AGTATCTGATCGCCTGCGCAATAGCATTTGATAAATTTCTCAAGAACCGGCCCAATA  
GCCTCACCTCGACAATATCCATCAGCCTTTGCATCAAACGGCTTGCAATGCCCAGTAGGA  
CTGAGGAACGATCCTGCAGCCAGGTTTGGAAAGACATCGGGGTGCTATAAAAGTTTCGTG  
CCACCCGCAAGGCAGCAGAACAATCACCGCTCAGGATAGCCCGGAGGCTAGATCTATG  
CGACAGTTGTGATGCTGACACCGGGTATCGAGCATCATTCCTGGGCCCTCGTTCGAACCAAG  
TAATGACTAACCTTCCAGCGATGTAGCTTCAAGGGCACCTGTGGCGGAAAAGGCCGTC  
GGGGAAGTATGAGAGATATTGTTTTCATAGTCAATGGCGACGCAAGCAATATAGCATCCA  
ATTCGGCGATCTGCACCAGGCTGTGATAGTAGCCGACTGTGCCACGGCCTGATAGGCA  
GTCTGCAAAAATCAGTCTGTGCTGTGATCCATGAGGAGACTCGCGAGGAGACTTCTGTG  
AAAACTTGTAGTCGAAAGCGTGTAGTCACTGATGAAGTTTCCGTACCATTTCCTGTCC  
TCGCCGCTTGGCCAGGTGCGAATACAGTCTCCATTGCAAAACGCTCGTTAGGGACAAGG  
TTCTTGTGCTGCGATCGGCCTTCCAACAGGATATTCCAGTACTGCTCCAAATCCTGTGCC  
CCGGCAGCTGGCATGACATGCCAATTAACAGCGATATCGTTGTGATACATCCTGACGGC  
ATATCTGGATTTGAATGGCGTTGGCCAGTACTGAAATTGCTAGTGGGTGACTTGGCTG  
TTTAGTCTTCGACAGGAGGTGGGCGGGACACATCGCTCCTGGCCAAATCAATGACCTTA  
GAAGTCTATCTTGAATGAACTGGACACTGCGGAACGGAAGGTTTCAACCCAGTTAAAT  
TGTTCCACAGGAACGCACGCGAAGCAACTTCTAGTAGACTGTCAATGTTCTGCTAGGATC  
TTCTCGGAGTTGACCCGTGTGCGAAGAAATGAGAGAAGAAGCATCGGGCAATTGGAAAAGG  
GGATCTTTCTACAGAAGCTAAACAGGGCTTCAAGATCGTTATTTGATAGCTCGCCTGCA  
TGAAAGCGTCCATGAAACTCGGTTTTCGCTAGCTGTAACCCCTGCTCGTGAAGATGTCCC  
TTCAAGTTCGGAGGCAATTTCGGGAAGGCTTGTACCCGTAGCTCGATTATCGTCAATATAGG  
ACAGATATGTAAGCCTGTGCGTGTGGCTTGTAGACTGAATCAGAAAAGCGAGAAATGGA  
GACTCACGCCCGGACAAATTTCCAGCACTTCTTTAAGTTCGGACAAAGATTGCCACCTC  
TCCAGAAAGCGATCAATGACACTGATGATCCAGTACATCAGAGATGCTTGTGCGTCAAG  
AGAGAGCTCCTAATACAAAGACTTTGCGAAGCACCCGCGGCAATGTTGGTGGAACTTTG  
CCCAGGATGAGCTGCTCGATAACCACAGGGCACTGAAAACGCCTAGGCAGCAGCCACTG  
TCTCTGTTTGGGCGGTAGAAGGTAATTGAAACAGCTTGCATCTCCATTGCTGACTGTG  
ATGACTCGACATATTGTAGGTATTCACAAGTTGTGCGATGATGACAAGTGGCCCAAGAA  
TGGCGTTGGGCAAAATTTGCAACCGTATGTTGAGGAGTAATGACTCTGTGTTGCAAGCCATT  
CTGCTAGCCTACGGGCGTGGTCTGCGCCGGGAGTCTGTTGTAGTTTGGGAATTGATTCAAG  
ACAGAGCATCCCAGCCACTCTCGATGTCCTCGATAGCGGGACAGCCATTGGCTGGCAG  
CATCATCTTTCGCAATGACAGAATCCGACTGAAATACGTTCTGGTCAAGGCTCATGGCCC  
CAGGCCAAACAAAAGCAATGTCTTAGCACTCGGCATTGCCATAGTTGATGTGGTATCAG  
GTTGGGAAGCTGGACTGACGATCAAGTGGCAGTGTGAGTACCACAAGTCTTACTTCAAG  
ATGATTTTCGACTACTCAGGTCAAGAGTCCATGGGATGAGGGTTTAGAAGTAACCATAT  
AGAAGCCCGGAGAAACCATCTCCTTCTGTGGCCGCCATGACAGACTCAGACAACATAACC  
TGCAATTAGGAATATCGTGGCCGTTATTGCGAGACTGCGGCGCGGACTTGGAGAAAAACAC  
GTCCCTGTGAACCGGAGTTGATACCTCGGCCGGAGTTTTCGGCCATATCCGGGGTCCA



Appendix

TCGTCCTGGAGATAGCCATGGATGTAATAATGATGATATTTCTGGGGCTGACCATCGGAGAT  
GACCAACGGGAGAGGCAAAAAGTAACTAGATACAGTGGGATCGAAGTATTGGGCTAAC  
TACCTAGGTAGTTAACCCCTGTACTCTAGCTAGGTGTACCGCCTTATATACCGCCTCGCTT  
TTAATATACTGCTATGAAAAAACAAGTAAAGCCGCTTATGCGACCGTGTACTTCTCCGGT  
AGTATTCTCCCATAGTCTTTTATGTATATTTTATAGTGTCTTTTGTAGCAGTCTCCCT  
TTTTACATACCTTTGACTGCTTAGTGTGTCTAGATCATGTGGCTAGGTACATACTAACTG  
CGAACCTTAACTAGAGTTTATGTCGGCAGAAGCGGCAGAATTTGGAGAGCGGTATATC  
TAGGTACATCTTTGCGAGGTCCACATACGGCACATCGTCCAAGGAGACATGATAGAGAG  
ATTTTACAATGAAATTGATTGCTCGACTATGTGTACGGCAAATAGCTTCCAGGTCACTAA  
CCTGGCGCGTTGAGTGGTTCACTTGGATAGGAAAATATATATATCTGGTGAATATCTTCA  
AAATAAGTCTGGTTTACCCCACTCAGATTGAGTAGTCTGTTTCCACACCACAGGATAATC  
CAGAGACTTGGGCCGAATACGGGAACTAATCTGCCTGCATGGGCTTTCATCCAGGACATG  
TCGTGATGTGGCTCATTGTTGTGGTTTCTCCGACGACTTGGATAACTCCAACCTCCGC  
CTCACAGAATCACAGCATGTACCCGACACCACTGATTACTCCGATCCCAATGTCCATT  
GTCCGGGAACAGATAATGAATACTGAGCCATCCGATGCCCTTAATTTACCCACAGAA  
GCCCTGGGTATCTAAGAGGCAAGGGCATAGATCTGGATTCCCACTGACAGTTCCGCCAAA  
TCCCATGTTATCTATCCACGGCACGAAGCCCTATCCTTCAACCTAGTGTGCGACATCT  
AAACAACAATATGGGAAGCATCGATAACACGGCACGGGCTCATCCGCCAGCGAGCCG  
ATCGCCATCATTTGGCATGAGCGCAAATTTGCTGGCGACGCAACCAACCCGACAATCTC  
TGGCGCATGTTAATTTAGGGGAAGAAGCGGTTGGTCCCTTCCCGACTCTAGATTTCGA  
TCAGAAGGCGTCTACCATCCAAATAATGAACGACTCAATAGTGTGAGGCTCACCCGCA  
CTCCTACAGAAAGAGTCTTAAACAAAAGTCAATGCTTAGACACATGTGAAGGGTGCACAT  
TTTCTCGCAGAGGATGTCGGACTGTTGACGCGGCATTTTGGTTACTCGGGTGAACA  
CGCGCATGTTAGTCTTTGCTGGTTCAGCTTCTTGGACTTGGACGCTACTGACCATGACTT  
TAAGTCAATGGACCCGACGTACAGACTCCAGCTCGAGTCCGTCTACGAGGCATTGGAAAA  
TGGTGAGAGAAAGCTTCCCTTCCAGCCCTTTTCTTTGAAAAGTCTTAAACCAAGCATCA  
CAGCCGCTGCTGCCATTGACGAAGATCGCTGGTTCCAACACCTCCGTTGTTACAGGAGTGT  
TTGTACACGATTTATAGAGATGGCTTACTCCGCGATGCCGACAATTTACCCCGGTTGATGG  
CCACTGGCACGGGTGTTCCCATGATGGCGAACCGTGTATCACATTTCTTTGACCTGCGCG  
GGCCAGTATGACAATAGAGACGGCGTGTCTCCGGGAATGGTGGCAGTGCATCAAGCCG  
TTCAAAGCTTGGAGACTGGAGAGGCGGATATGTCCATTTGTTGGCGGTGCCAACCTGACGC  
TCAACCCAGATATGTTCAAGCCCTAGGTTCTGCTGGTAAAGTCAAGTACCCTAACTACC  
GTAATAGAACGAATGGCTGAATTTGTTTATGAAAGGTTTCTTCTGCCGACGGCAAGTCTC  
ATGCATTTGATTCCCGCCAGTGGATATGGCCCGGGCGAAGGTTTGGAAACACTAGTCCG  
TGAAGCGCTTGTTCAGACGCCCTTGGCCGAGGATCCGATTAGGGCTGTGATTCCGAGAAT  
CGATGCTCAACCAAGATGGCAAAACTGAAAACAATTACGTTCCCGAGTCTAGAGGCACAAG  
AAGCCTTGGTGGCGGATGCTATCAAAAAGCAGGTCTCGACCCCTCGAGAAAACGCAATATT  
TTGAGGCACATGGCACTGGTACGCGAGCCGGGATACTATTGAGGCACAGGGCATTGCAA  
ACACAGAGGCGCGAAGCGGACTTGAAGCATCATCAAAAAGTGCCTAGCCATGGAAAAATG  
GAGTCATCCCGCTTCTATCAACTTTCGAGAAGCCTAACCCGAAGATCAGCTTGGATGATT  
GGAATCTGAAGCTTGTTCGGGAAGTGGAAAACATGGCCAGCGGGCCCATCAGACGCGCAT  
CAATCAACAACCTCGGATATGGAGGAAGCAATGCGCACATAATCTTAGAAGATAGCGCTT  
CGTGGGTCAAGGCTATTGGTGGCCAGAATGGACGTACCAATGGGTTTCCCGGATGGACATT  
CGAACGGACCAACGCAATGGTCAACCTCCACGCTGGACCCACATGTGCAAGAAAGCC  
AAGTTATCTCAAAGGCTCTTGTATTGAGTGGGAAGGACAAGCAGGCGTCCGAGAAAATGA  
CAGCGAACCTWTCGGACTACCTGAGACAAAACCCAGTCAACAACTCCAATCCACGAGAGC  
TCCTCGACAGTTTGTATACGCTAGTTCACCGCGCAGCCGCTTCCCATGGGTATAG  
CACATCCAATACCAGTTACGGAGGGGTATGAAACCGTAGTTTACAGACTCTCCAGTCCGCCA  
AATTCAAACCAACACGCACTTCCGCTCGACCTCGGATCCGATGGTGTTCACAGGCCAGG  
GGGCACAGTGGAAATGCCATGGGAAGGGAGCTCATCGAGGCTTATCCCGTATTCAAAGCAT  
CTCTTCAAAGGCGCCAGGATATCTTGAACAGTTTGGCCCGCAATGGTCTATTGATGGATG  
AGTTAATGCGAGATGCCGAAAAAGCCGCATCAATGAAGTCCGCTTGGTACTCCGATCT  
GCGTGGCAGTACAGATCTCGCTTGTGCGCTTGTACGGCTTGGGGAATCGTTCTGTGCG  
CTGTTACCAGCCATTCGAGTGGAGAGATTGCGCGCGTTACAGTGGGGTGGCCGTAAGTT  
ACAAAACAGCTATGGCCTTTTCTTACTACTACCGTGGGTGCTGGCCGACAGACAAGAGCCTAC  
GCGGGCCAGTCAAGGGCGGCATGATTGCCGTCGGACTTGGATTAGAAGAGACGGAAATCCT  
ATCTTCCGCGGCTGAGCTCAGAGGGCCAAGCTGCCATAGCTTGCATCAACAGCCCGTCTA  
GCATAACGGTCTCCGGTACCTTTCGGCAGTGGTAGAGCTGGAGGATCTGGCCAAATGCAG  
ATGGTGTATTTGCTGCTGCTGAAGGTGGACACGGCCTGGCACTCGCATCATATGACTC  
CAATTGCGAATGTCTATTGCGAAGCCTTGGAGAATACACGAGCTGAAAAGATTGACCGAG  
ATGCTCTGACCACCGTTGCATCTCATCTCCAGTAACTGGAGGTCGTATTACAGATGCTC  
AACAGATCGCGCGCCCGGAGCACTGGGTTGAAAGCTTGGTACAGCCTGTGAGTGTGTCG  
CCGTTTCAACGATATGTACTCGCGGCTCGGATCTGTTGGTCTAACGTGGATGTGG  
TCGTTGAGGTGGGCCCGCATACAGCGCTGGGAGGCCGATCCAGGAGATCCTTGGACTGC  
CCGAGTTCAAAGATTGAAACATTCCATATTATGAACTCTCGTTCCGAAATTAGACGCC  
GGGACAGCATGCATGCACTTGTCTTAGTCTTCTACGAGAAGGCTATCCTGTTAATATGG  
GAGCAGTGAATTTTGCACATGGGCGGGGACAGTACGTCAAAGTACTGACCAACCTACCAT  
CGTACCCCTGGAACACCAGGCAAAAGCACTGGGCTGAGCCACGGCTAAAATCGGGCCATAC  
GTGAACGATCCCAGCCTCTCATGACTTGTCTCGGATCCATCGTGAAGGCTCAAATCCAA  
ATGCACCGTCTTGGCGACACATCCTTCGAATGTCCGAGTACCCGTGGACCAGAGATCAGC  
CTATTCAATCCAACGCTCATCTATCCAGCCGCTGGGTACATTTGCTGGCTATCGAGGCCAA  
GCCGTCAGCTTCAATGTGCTCAATCAAACGGCCGGAGAGATTGGTGGATACCGGCTTCGCG  
ACGTTGATTTCTTACAGGCCCTCATGATCCGGATAGCTCAGACGGCATCGAGATTCAA  
CGACGATACGTCAGTACGCGAGAAGGACATTCCTCGCAAGGATGGAGGCAATTTGAGG  
TCTGGTCCGTTACAACAGACAACCCGCTGGACCCAACACGCAAAAAGGTTGGTCTGTGTTG  
AACTTGGAGAGTCTTCTGTCGGATGTCCCGACAGCTAGGAAGAACATTACTGGCTACA

Appendix

CGGGCGAATTCTCCTGCTGATCTATTGCCAAGTGGGAACTCTGGGGATTACACACG  
GGCCGGTTTTCCAGAATATGGACAGCATCATCCAGTCTGGTCTGAAATGCGAAGTGTGG  
TGAGATGACTTTGGCCGACGTCTCTGTTCCTAATGACCTTCCCGAACCACATCTGTC  
ATCCCGTACGTTGGACTCGGTAATCACGGCCCTTACTCAGCAGTCCCTGGAGCTGCTG  
CCCGTGAAATTAATGCAAGGTGCCAGGTCTGTTGAGAGATTCTGGGTATCCAGCAAGA  
TAAGCCACGATGCGGGACATTCTGTTGGAGGCGACACGACGCTCATCCGCGATGATGATC  
AAGGAATGGCGGACAGATGTCTGGTTCCTGATCATGACACCGGAAATATATGCTCGAAA  
TGAATGGCTTTTCTTACCAGTCTTTGGGACGGAGCACGTCACCTACAGAAATCCGAATCTT  
GGCGAATGAGCTGTGCAACAAGGTAGTTTGGTGGCTCGATATTTCCAGCCCTTGCCTG  
CTACCTTGGCTGCAGTGAGAAATGAATTTGGCCTGTACCCTCAATCTGTGAAATGTGACA  
CTACAAAAGCTACATTGGGTGCTGTATCTACTTTATGCAACTGGCTCTCGTCTGCTGCTGG  
ACTCGCACGACATAGCTGAGATGGAGCAACATAATGCGTCATACTATACATGGATGAAGG  
ACACTGTGAGCTGGCCAGCTCGGGAAAAGTGTTCGAAGGTAGCGCCGAATGGTTATACC  
ATTCAGAGAATGAGAGGACGCTTCAATATGAAACAGGTTTACAGCAGATTGGATGGGGAGA  
TCGTGTGTCGGTGGGAACTCAGCTGGTAGACATATTCGCGGACATCCCGGAGCACTCG  
ACCTGGTTCATGCAAGACAATCTGCTATCTCGTTTCTACAGCTATGCTCCAGGTTGAAAC  
GGCGGGGACGACAGATCGCAGGACTTCTTCGCCATCTCTCCACAAGAAATCCCGGTGCTC  
GCATTTTGGAGGTGGGTGCGGCCACAGGCGCATTGCACTCCATGCCCTTGGAGCCCTAG  
GCACGCTGACTCGGGTGGTCCCAATGCCTCCATGTACCCTTTACGGACACTTCTACGG  
CTTTGTTTCGAGACAGCGAGAGAAAGCCCTGCAGCCCTGGGCTGATCTGCTGTCTGATG  
AACTCGACATTGAGCATGATCCAGCATCGCAGGGGTATACACCCGGGACCTACGATATAG  
TGATCGCCTCAAATATCCGATCTATCTCTGAGTCCACATCGCAAGCGCTGAGCAACATCA  
GATCCCTGCTAAAGCCCGGCGCACCTCTTGTGTTGGAACCTTTGAAATACGAGGTGCG  
ATGTTCACTTTTCCGCTGCTACTTCTTGGCCGGTGGTGGGACGATAGCACAGAGCTGA  
AGGCAAACCTATGTCTGGATATGCCATCTGGGAAAATCAACTCTAAGCGCCGGTTTCA  
CAGGTGTTGAACTCGAGTTGCTGGATCGTGAAGACCCCAAGAAGCCGCTTGGTGACTT  
TCATGTCCACTGTGCAATTAACACAGCCACCAAAATCAAATGTGGATGCGGACCAAGTGG  
TCATTTGCACAAGTCAAGACGGATGTCTTCCAGCTGCTTGGGTGAAGGGCCCTCAAGGATG  
CCATCGCTGCCTACACCGTCAAGTGAAGGAAACTAGGTCCCATTTGTTGAGATTAGAAAT  
CCTTAGCTGCAACAGCTGCGTCTTATGCGGACAGATCTGTATCTTCTTGGCGAGGTGG  
ATGAAGGATCCCTATACAACCTGAATTCGACATTATTGGAGGGAATTCGTTCAATGAGCA  
CCAACCTAAAGGATTTGATCTGGGTGACGCGTGGTGGTCTGTGGACTGTGAAAGGCCAG  
AAATGAGTTTGGCGACCCGGTTCATTCCGTTCCCTGCGCAATGAATATGTCTGGTGCAGC  
TGCTGACCTCGACCTGGATCCCAAGGGAACACCGTGGTCTGACGTTAGTATGGCTGCAA  
TTGCCAAGATTCTGGGTACGGTCAATGGGAATTCGCCCGGTGGCTCTATGGTAGAGAAAG  
GTGCCCTGGAGCTCGAGTACGCCGAACGAGATGGCGTTATCTTGATCCCGCAACTTACC  
ATGATGTGACGAGAAACCGAATGCTTTCCCGGATGCATCAGATGCCGCCATGGAGAAAA  
TCTCAATTGAGAAATTTCTACCAACCAACCCGCCCTTTGTGTTTAAACCGGATTGCTAG  
TCTTCCGTGACGATGACTTCTCTGCCGATTATCTTGAACATCTCCACCCGGCATCCCTGG  
AAGTGCACACTAAGCCGATGGTGTACTACTGAACAGTGTCCGGTGCATCATCTGCTGGCT  
TTGAGTGTGCCGGAATAATTACGCAAGTTGGGGAAGAAGCAGCAGCCCAAGGCTATGCAG  
TCGGTGTGCGCTTCTCTCAGTCTTGCACATTCTCTTTCCGAGCCGGGCTGCTGCTCG  
ACTGGAACCTGACGACCGCATGCCAATGACATGACCTTCCAGGAAGGAGCATCACTTC  
CGTGTCTTTCTCCTCAGTGCATACTTTGCCCTGGTGCAGAAATCGCCGCACTGCAGCGTTCTC  
GGTCACTTGTATTACGCTGGTGTGGAGATGTTGGGCAAGCTGCAATCATGGTTGCC  
AGCATCTCGGGCGGAGGTATATGTGACAGTTGGTAGCCCTGCAGAGCGTGGCCCTGCTCA  
TACTGAAATATGGTCTGCCGGCGGATCATACTTCACTTGTACAGACTTGTCACTTGCATG  
ATGCGGTAGTACTGCGACACAAGGCCGTTGGAGTTGACGTGGTCTCAACTCACTGACTG  
GCCCGCTCTTTCAAGAAAGTCTTAACTTGTGGCCCGCTTGGCCACTTTGTGGAGATTG  
GCAGGCGCAATACCCAGACAAATGGCTATATGCACATGCGGCCATTTCGATCGTGGCATT  
CATTCGCGACCTTGATATACCTAGCTTGTGGAGTATCGGGCGATGGATGTTACCCGTT  
GCCTCGCTGAGTTGACACGCTCTCATCGAGTTAAAGCCGTTGACACTGTCCACCAATCA  
CCTTCCATGCCATAGGAGAGATCGCGGAGGCATCTCGTCTTTAAAGCGGGAGACCAGA  
TTGGCAAAGTGGTCTTGTGCGGTGATGACGATTCAACGTTACTGCCGTTGCCATCCAAGC  
CGGCTGCAAAGCTCTCTTCCGAGGTCTCGTACTTGTGATCGTGGTGGCAGTGGCGGCTTAG  
CCAGTCTGTGGCGCACTGGATGGTCAACCGTGGAGCAAGAAATCTGGTCTTCTATCTC  
GGAGTGTGGGACGAGCGAGAAGACCGCTGCATTTGCCGAGGATCTCCGCCAGGCAGGAT  
GTCGTGAGTTTTGCCCATCAGCTGCGATGTTGCCAATGAGGAAAGTTTGGGCGACGCCA  
TCAATCAATGTGCTCAAGAGGGCTTACCCCATCAGAGGTATCATTCATGCTGCGTTTGT  
TTCTTCGTGTAAGTGTGCAAAACTCCAACCGTAATCCTGAAATAAAACAATAACCGCGT  
ATCTAACAGGATGCCCTTCGTGGAGAAAAAGACCCTTGTGACTGGACATACACTATTCAG  
AGCAAGTTCGCCGACCTGGAACCTGCACAACAGTTTAAATTTGCCCGGCGACCTCGAC  
TTCTTCGTCTTGTCTCTTATCAATGGAATCTTGGATATGCCAGCCAGTCTGCATAT  
TCCGAGCTGGTGCATACGAAGACGCCCTCGCCCACTGGCGAGTCAAGCATTGTGGCCCT  
CCCGCTGTGCAATTGACCTCTCCGTCGTCAACCGGTCGGCTACGTTGCCGAGGCAAAC  
GCATCCGAAACACTACGCGGCTCTCTCCTCAGAGCTGGCCGACAGTCACTGATGAAGAT  
CATGTTCTTGGCTCGCTAGAGTCTGCCATTCTATCGCCCTTCGACCCACAATTCGTGGT  
GGTGGTATAAAGTCCGGCCAGGTCCCATTTGGGATCTTGACGGCGATCTAGGCCGTGAC  
ATGCGTGTCTTGGCACTCAAGTATCGCCCTCCGGCTGTAACCGGACAGAGTCAAGGACGAT  
GATTTAGCAGGACTCTCTCGCCGCAAAAATGATCGCCTGCGAGTCAAGGGCGATGCC  
GTTCTGTGTTGGGACTGCCATCGCAGAAATGCTGGCTGAAATGTTCTTGTCTTCTATT  
GAGGATGTCGATCTAGGCGAGTCCCTCCGACGCAAGGAGTTGACTCTTGTGAGCAGTT  
GAGGTCCGGAACATGCTTTTTCAGCCAGGCGGCTGTAAGTTTCCATCTTCAATATCATG  
CAAAGCCCGACCTGACGCAACTAGCGATTGATGTTGTGGATCGCAGTGCACGCTCAAG  
CTTGGCCGTTGATCATATATTGGTTCATGTCGAGAGCTGTCAAATTTTGTATAGACCA  
CGGGAGCAGTCAAGCTGCATCGGCGAGCTAGTCTTGTCTTATGGGTGATCAATATT  
ATTGGAGAGCCGAGTTGGGTTTATATTGACTACATACTGAGGATTGATGGGATTCTC

Appendix

AACATTTTCGTGGGCATCTTTCTCAAGTCTCCTCTGGAGCTTTTATATGAATAAAATTCGTA  
CAACTTCATGGAATGAAGTGGATAGCACTTCTAGGACAGGTAGATTTACATTTGCTTTGA  
ATTTCTACACATTACGGCTTGTACCAGTCTCGACAATAGGGCAAAATGTCAGTTGCCACTA  
AACCAACCCCTTCTCCAAAGTCTCAGGCCCAAATATAAAAATTGTCAAAGTCGAAATTTGAT  
ATCCTTGAAAATGTCGTGGAATGAAACTCCCGAAGACCCGTCTACCATCAAACACACCC  
ACCGCCATCCATACTGGGGTAAGGCTTGCCCGTGTATCCTGGCTGCGAGGGTCTCAAACGT  
CCACAGGCACTGCTGGACATAAGGACTCAAGTGTCTCATGAGCGGGAGCGCCAGCATAAC  
ACTTTCCCAAGACTGGGAAAACAGAATCTGTGAATAGCTCCGATCGAAACATAGCTGCGAG  
GAAGTTGGCACC CGCGATGTGAAGGTAGTATATTTCTGTACCACCACGGAAGGAGTCCAAT  
CGGCTCATCCGGTCAAGCGTCTCGTTAACTAATGACGCCACCTGTTGCGCGGCTTCGAT  
GCACATGCTGGCGTTTTACGGAGGAGTCCGTGGCTTAGGCTGGGTGATTTGTGTAAAGG  
CTGGGTGTCTGGTTTCATCGCGTAAAATCGGGCGAGCATGGGTCTGTATAGAAAATCCG  
GTGGTGGAGGTAGCTACATAGAACGATAAAAATCGATTTTGGTGGAGATAAAGGGATTGT  
TGGCTTACCGGAGATGAAGCAAGTATCTCTCAGCTCTAGACGACCTGTCCGGTCAACATTC  
TGAGGTTCCGAGGTGGCAATCGTGGGATACTATTTTCCCAATCTGAAGCATGCAT  
CTAGTTGCAATGCCCGTTATGGTACTCGTCTTGTGAGATAGGAGTGGAGGGCGGAATC  
TGGCTGCGAGACTAGTTGATTTCTGTATCTGTGCCAGTTGAATTTGGTTCGGATTTCTGT  
GGAGTCTTAATCCCAAGGTATCATGCATCGCGTTTTCACTAGTTGTTGGGGGCTACTCG  
GTGGTGGGACGGAATTAGCACCATGCGGACGCTTACCCAATGACCAGGAGATGCATC  
TACCTGAGGTCACTGGAGAATACTATAGAAGAGCGATCACTCGATGCATACCGATCGA  
GCGCAACACAGCTAGCCATACTTTCTGTTTCAAGGCTTGTCTGCTCACTGGAGCCTTTCA  
ACAAGGGCGTCTCGGTGAGATGGCAGCAGATGCTTGGCCAATTCGCATCGCCAGACCTG  
CTGTCTCCATGTCTTCTGTGGTTATTGGTGCAGTGGAGGTATCGATTCAATCAGCATCA  
GACACTGTACGAGTCCGAGTATCCAGGTTTCCATAGCATAGACTCAGCTGGAAGCAGGA  
CCCATGCGCGTGAAAAAATCGATTTGCTTCTCGTTTTCTTGGTGCAGCGGATTGAGTT  
CCTGTCTTTGAAGTCCAGAGCAAATACTACATGAGGATACCAAGCCAGAGGTGCTAGT  
CTGCACAGAGTGGTGTATGGTGTGAATATGCTTTTTCATAGTTTTCAAGAAAACGCT  
GGCGATCTAGAACTGGCTCCATAGGGTCAACGTGTTGCCAGTATATGTCAATTAATGAT  
CGGCATATGCCCGGGGAGGCAATTCGGTGCAGCAAGACAAGGCACAAGATTGAGAAGGCA  
GGTTGAGTCTCCAAATAGCGGCGCATCTACCAGTGGGACCTGATTACAGGTGGCCGGAA  
TGAGACCTGCCCTTTCTGTTTATGGCAGCGGCGACCTGGCCGGCAAACCGCCATGTGCGG  
TATAGTATGCCCGTTTTCTCCAGGTCAAGGCTCTGCCTCTGGATCTTCAATTCATGATGGT  
GAGACACAGGTATTGAATCGACCGTGGGAGGTGACGACCCACTGATTTGGCGGAGTCT  
CCAATGAAGGTGCCGAAGATATGGGAGGCACCAGCATAGGAACAGGCATGTGAAGTATG  
ATGCTTGGCCACGAGGTTTATGCACTGTCTGCTTCTCGACGGCACCTTGGCCATAGGCAC  
ACGCTTTCCTTCAAGTCCATCTGGCGCATGCTTTATGGTATGTGTGCTCAGCCTTATGCA  
CGAACTCTGCAGGTCAATGAAACAATGAAGGGAGAGATACCTGGCCGAGCACCATCACAT  
TTGATCTTCTCGAGCGCAGATCTTGGCAAGCGCTTGTACTCTTAGCCTCTGCCTGCGT  
AAAGGCTTCGAGTGGTATGTGATCAATATCACTTCAATCATCACGCTTTTATCGAAC  
AGTTGGGGATGGAGAGTGGAGTGGCCCGTCAATCAGGGCCCGGATCTACCTGATTGATA  
ATCAGCAGTTCCCAATCAGATAACTCGCCGTTTTTCAATAGACCCGACGGGAAAGCT  
CTGTTATTCGCTAGTCTGCTTTCACTTCCCGCAATCTCGAAAGCAGCCAATGAGAAGTG  
CATCCATTTTCCATGTGTCAACCAATGAAATAGCAGTCCGCTTTGATGGCTTCCGCAGT  
AAATAACTCCCGCGTCCGGTAAGCCGTTATATAGATCGAACGTTTTGAGAAAAAAGAAG  
AGTCTCCTGCTACACAGCCTAGAATTTAAGTCTATGCACTTAATATTTCTCCAATTTTTTA  
GACGGAGCTTCCCTCCCTTCCCTTCTTCTTTTTCAGCCTTGTCCATAGTCAATACAGTGGT  
ATTGATCTCTCCCTTCCCTCACTTGCACCGCTAAGACAAATCCATGTCTCATAACAGAGCC  
AAAGGCTCCTGTCAACACAGGCGAGGTTGAAAATGGTCACTTATACGATGGCTCCGGCAC  
CGAAGATGACCCATTCATCGTGGAGTTCAAAAAGACGACCCCGCAATCCGATGAACTG  
GGGTCACTCCCGTAAAGTGGTTCATTGCAGCCATTCGACCCCTCGGCTTTTGGCGTTAC  
TTTTACCTCCTCTGCATACTCGGTATCGGCAAAATGAAGTCTTCAAGGACTTTGACATCAG  
CACCGAGTCTTCAATTTGCGGGCTTTCTCTCTTTGTGTGCGGATTTGGGATTGGTCTGCG  
CGTATGGGGTCTTTGGTAAGCCATTAACCATTTTGCATTTTCAATGGAGCCAGCTAAC  
CATACGTTGCTTACAGTACGTTTCGATGGTGCATCCTATTACATACTATAACATTC  
AGCTAACATTTGGCTTAAGGTCCGAAGTGTACGGAAGACAGATACTGTGGATTATCACTCA  
CATCCCATGCTCGCTTCTTAGGAGGTTCCCGCAGGCAGCCAAAAACGTGGCCACGCTCTCT  
CATCCTGAGATTTCTTGGCCGACATTTGGCGGCTCTCCACTCGTCAACTCTGGCGGAAC  
AATGTCTGATCTCTCCACCTGCTCAGCGTGGTCTGGCATTGACCATCTATTGCGTTGCG  
GCCCTTCTCGGCCCCATCTTGGGGCAATCGTGGGCGGATTCTGTCTGAGAGCGTCCG  
GTGGAGATGGGTCCAGGTTGTCTGCGTGTATCTTATTGGCGTGGTCCGGATTTCTGGGAAT  
CGTCTTCACTTCTGAGACATATGGCCCGTATTGCTTCCAGCGACGGACACATCAACTGGC  
CAAAGCTGATGGCAAGATCTACGTGAGCGTTTTGGAGAAGAACCAGGGAAAGAAGCTGCC  
ATCGGAAGTCTTCAAGCGTGTCTTTGTTCCGTCCTGGATCTTCTTGTCTCTTGGAGCCAT  
TGCTTTGATAGCGTCAAGTTACATGGCTATCATTTACGGCACAGGCTCTATATGTTTATGGG  
TGCCATGCCCATCGTGTACAACGAGGACCGTGGTTGGAGCGTGGGCATCGGCGGACTGGC  
GTTCTTGGGAATTGCTGTGGCATCATCTTTGGCTTAGTTTATGCCATCTGGGACAACAA  
CGTCCGCTACATGAAGCTTTTTGCGGCAAAAATCTGCAAAACCCCGAATCTCGCCTTCCACC  
TGCAATTTGTTGGAGGTTGTCCTTCCCTTCCCATTTGGCATGTTTCGCAATTTGCTGGACCACTA  
CCCCAGCATACACTGGTCTGTCAAGTATAATACTGTCTGCACCGTTTGGATTGGCTGCGT  
GCTGGTCACTCCTGCTATCATGAAGTATTTGATCGACACTTACACCATCTATGGCCCTC  
TGTCTGGCTGACAGTGCATCTTCCGCTCAGTCTGGGCGCTGTGTTTCTCTTTTTCAC  
GACACAGATGTACCACAATTTAGGAATCACTGGGCTTCTCTATCCACAGCATTTTGTGAC  
TCTTCTGTGATGCCATTTCCCGTTGATCATGTATCGCTACGGTGAAGCGGTCCGGATGAA  
GTGCAAGTACTCATTGGAAGCGGAGAGATGATGAGGAAGATGCAATTTGAGCAAAACAGC  
TGCTGCTACTACTACAGAGAAGGACAAGGACTCTTCTCTGAGTGGAGCTGTTAGAAAT  
GGGCCCTTACTCCGGCATCAATGGAGGGCAAGGCTTGGGGTCTATCATTTGGAGGATC  
GCCGATATAAATATCAATATCTGCCCTAGAAAATAGTTAATTAAGTGTAAATCAACTC

Appendix

AACTCATCGGAGTCCCTAGAAGTTGATTATCACCTCTCTAATGTGCATGTTCTCCTTACA  
GACTAGCAATAATATTTATGTCTATCAAAGTGTCTTAGTAGTTCCTCCGGCTCTAGTCTT  
TGAGATACAGTATCTGTGTACATAGTTGATATTGGGGTATCAAGCCAAGGTGTGATAGGG  
TAAAGGAATCCTGTATGTGCGAACTTGCTTAAGATTGCACTTTCGGGTTAGTCTGTAAATA  
TCCGTGGACATATTTATTGATTCACTCTCGTGTCCGGAGTTGAGAATACTCCGTAGTGAGGC  
TGTTTTATAGTTCCGGAATGTGCGTGCCAATGAGAATCGTGAACCTGTTCTTCGAGCG  
GAGAAGAGTAGATCAACTGTCTGGATATTTTTGGTCCCGTTGACCTCGGAGTTGTCC  
CAAGGCAAGACCAGGAGTTCCGAGTTATAAGTAGCTATACAATGGTCTCGGAACTCTCC  
CGCGCACAGAGCTCATATATAACCAACGAGTTCAATGACGAAATAATCCTAGCTTTA  
TAGAATTAACAATCTAAAATGCAGGCCGCGCAGTGCATTTGCGACCTGTCTCTTGCTTCA  
GTGGGTGGCAACAGCAGCGCCGTGCGCTTTCCAACCAAGCAACTATTCTACTCTTGTCT  
GCGCCGTACAATTTGATCTTCTCACCAACCCCTTCGCGCATTGTCTGGCCCCAGGATACA  
CAGCAGGTAGCAGCCGCGTCAAATGTGCGGTGATTGACATCAAAGTTCAACCGAAA  
AGCGGCGGACACAACATATGGCACTATGGATCAACCACCGCGAGCTGTCTGTGAATCTG  
GACAATCTGACGACTCAGCATGGATGAAACAGCTGGACTGCCAGATGGGACCAGGT  
AACCCGCTAGGCGAGTCCAGGCTCATGTATAACAACGGTGGTTCGACATGTTCCACAT  
GGGACGACCTTACGGTGGGACTCGGTGGACACGCAACTGTTGGAGGTGCGGAGCGGCA  
TCACGAATGCACGGACTGTTGCTCGACTATGTGGAGGAGGTAGAGGTTGTTCTGGCCAAC  
TCATCCATCGTCCGAGCATCAAAGTCTCACACGAGGATCTTTCTTCGCGGTGCGTGGT  
GCGCTTCAAGCGTTGGCATTGTGACCGACTTTTCTATACGACCGGACCGCTTCCCGTA  
TCCAGCGTCACTTACTCTTATATCTGGGAGGGACAGATCCAGCAGCCCGCGCAGAAGTA  
TTCTTGACTTGGCAGTCAATGCTCGCCGTTGGCTTTTCCCAACACATGGCTTACGAT  
CTGTTGGCAGCGGCAATAGCATGATACTGGGTGGGGCTTACTTTGGCAGTCAAGGAGGAC  
TTGAGGGCTTTTAACTCAGCAGCCACTTCAAAGTAGCGCAGACGTGCGACATATAAAA  
ACATATAACCAACTTCTTGCATCTCTCCGCGCGCAAGCGCTCAGACTAAAGCAGCTGGG  
ATTGCTTCCCGTGCATTTCTATGCCAAATCTCTCGTCTCAATCAACAGACTCTCATA  
CCTGATGATGCTGCTGAGGAGGTTCTCAAGTACTTGGCCACGACCAAGAATGGTACTGAT  
TTGTACGCGTCAATTTCCGCGCATTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
ACAGCTTCTATACCGTGTATGCTTCACTTTATGTTCTCTTTGGAAGGACTAGTGGA  
GACCTGACTGACACGACGGTGCAGTTCTGGATGGGCTGAGCGAAGTTCTGACTAGTGGG  
CAACCCGATGCCTACTACGGCCAGTACGTGGGGAATGTGATCCCAGACAATCAACTGAC  
AAGCGATTGACGGGTATTTATGGCAAGAATCTGCATCGCTGCAGCAGATCAAGTCTGCG  
GTAGACCCGAATGATGTTTTCCATAACCAAGCAGAGTATCCACCTCTGTCTTAGTCTGTCT  
GGTGGCGGATTTCTGTTAGTCTTAAGATTGTCTGTTGATCGTGCACAATGAACGCCATTT  
GTACTCATTGGGCAGAAATCACAAATACTATGGAGTATAACAGCCAACACACAAGAATT  
TGAATGGAATAAAAATCACGGCATTATTCCATTGCAACCAAAAAAACAAGGTTACTG  
TACATGGCCTAACGGAGACCATAATCCCGTTTGACAATTTGGTTTGTGACAAAAATATCC  
GACATAGGGATTCTGTAGGTAACCTATGCAAGTTGTTTCGGGTAGTCCATGAGCCCTGGCA  
GGTGTATATGTACCTAAGCTACGTAGGTAGATAGTAATGTTAGTTGACACAGACTGGAGT  
AGCCCCAGCAGTATCTTTGCAAGTGTCAATGAATCAGACGCTGTTTAAAGGCTCCAGGGGGT  
TGAAACAGCCCTGGAATGAAGCATTGGAAAGATTGGGACAGGTTAGGGTATGATGAAA  
ATAAACAAAACCGATGAGTATTTCAAGAAAAGTACGGAAAACAGCTAACTACTCGCTAACT  
ACCCTCCCATTTACCGCTGACCTCGTCACTTACCCTCAATCATGTAGGTACGGTCACTG  
TCTTTCCCGTCAACTCTGCATCTGTACAGTATGTAGTCTCAAAAACGAAAATCAAAAAGAT  
GCCAATTTCCATCTTACAGCACGCGTCCAGCCGATGAGCTGGTGTACGGTCAATCACT  
TGAGTCTTCTCAGGGACTTGCCTTGTGGATAGACGCAAGGAGGGCGCTCGGCCAGCG  
CCACCGGACGAGCTGGGGGAGCAGGTGGAGGCGGGATGCCAGCTGGAGGGGGTGGCGGG  
GGAGCAGAGGGGGCTCCGCCAGGGGGCGGGGGCGGAGGCATAGCAGGACCAACAGCTCC  
CCTGGTGGGGGAGGGGAGGTGGGGGAGGGGAGGTACTGCAGATGGAGCGGCTGGTGGG  
GAGGCGGAGTCTGTAGGCTTGGTATAGCAGGAACAGGAGGCAGAGTCTGGGGGTGGG  
GGCGGAGGTGCAACAGGGCTTCTGTACCGGGGTAGTGTATTTGAGGGGAGACAGCTCT  
TCGTTCCATGCACTCAGAGGACGTGGAGGCGCCATAGTCCGCAACAGGATGCTGGCAAGT  
TGTTTGGCGTGCACACAGTTGGACGGTCTCTCATCATCAGACGAATCAACCTCGGAT  
CCAGCAGCGGACAGTCACTGCTTCTCGGGGCGAGCAGAGTCTTGTGTTCCAGGGGG  
GCTGCACCGGTGAAGTGGGCTTGTAGGTTCTGGCTGGGTCAAACCGGTGGAAGGGGTTCT  
GTAGACTGAGCTGTGGTTGGCGCGAGGTGCGTGGGTGGGTTGGAAGGCTGACCTGTG  
TGCTTGAAGTATGGGTTCTTGGACTCAGCATCGGGCGGACCGCAGGAACACCACGGCTG  
CTTTCGGGGGAGCTCACTTGCCTCGGCACACTTTCGGGGCGGCGARGKACAGRAATMGCK  
RYCGTGGRCACGGGAGGCGGGGTAGGGAGAATCTGGCTCTGCGTGGTGTGCTGTACTGG  
CGGTAATGTCCGACAGGCTTCTGCTCATCGGACGATGACTCGTCAAGGCCCTCGAGTTCA  
CGCTGCAGCTGTCTCAGTTCCTTAGCCGCTCCAGTTCAACCGGTTGGGCTTCAAGC  
ATAGCTTCTGCTCTTTGGCCTGTGCGGTAGCCTTCTCCTGCGGCGCTTCTCTTCTGCTG  
TTCTTAATCTTGGCCCTGGCGGACTTGTCTCAAGAGCACGAAGTCTGCTCGGCTTGGGCA  
TCCTTCTCCGTTGGAATTCAGCTCCGTTGGCTTGCCTCACCCCTGTAATGATTAG  
CCAAGCCCACTACTTTTGTCTTTATCCGCTGTGGATTACTCACTCTAGACGCTTAGTCTC  
TGCTTCTGCACTCTTGTCTTCTTGTATAGCCTGCTCACGGGACGCTGGTCCGCTTT  
GGATCTCTTCTTCTTGGCATCGGCTTGGCCGCACTGTCCGTCCGAGCCCTGCGAGT  
GGGCGGCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
CCGTTCTCGGCGAGGCGACGCTGTCTCTCTTCTTCTCAGCTCTTGGGTTGGCAAGGGTCTC  
AGTAGCGTCAATGGGCTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC  
ACGCTCTCGAGCAGCTGCCACACGATCTTCATGCGTGTACCGGGAAGAGAGCTTGAAGGA  
GTCCACGACGGGGGATCGAAGGGCGCGCAGCGGAGATTACAGCCGCTGGAGATGCTCG  
AGATTGCACATCCGGGGAAGGTCTCTCTCGCCTATGGATTGTCAAGTAATGGACACTCTTT  
TTGATTAAGGGGAGAWTMRCKMCKCTWMTTCTCSCKWATATTCGCTGTTCGCGCAATG  
CGCTCAGGTCGTAGATGAAGTCCGCGGATGACGCTCTCACTCCAATGCTCTTCCCAA  
CGACGCTTCTCAYGCTCAGTGTGGAGTTCTGGCCTTCTTCCCGGAGACTGTCTCCAAA

Appendix

CTGCGAGCGAAGTACCGGACACTTTCCTCAACGTCGCGGGTTCATGGTGTTCATTGCGCTCC  
CGCTCCATCTTACAGTTAGTGTCTCAGACTCCAGACGACGAGCTGCTTGGCCATCGTCA  
TCTTCCGTCGCGGGTGCAGGGCCAGCAAGCTCGGCAGCCGGGCTTGCATACGTGCA  
CGCGCGGAGCCTTATTTCGATCTGATTACAGTACAGTACCGCCAGGGCCAGTGCATG  
ATGTTAGAGGCCACCACCGGGGTGTCCTTGGCATCCTTGGAGACGGAAGAGTTCGAGCTTC  
GCATCGGCGATTTACCGTCAAGACGACGCACATCAGATGCAACCTGGGGAACCTTGATCT  
TCAAAGACTGCAATTGGCGACGAAGAGCCTGCGATCTGCTCCACTGTCCACTTGGCGG  
AAGGCAGCATTGGGGTGGGTATCGAGATCGTCTGCACACGCCGATACGGTCCATGAGA  
CTCTCTGCTCAGGGCGATCCTGGCGGTCTAGAGCATCATCATCTCGGCTCGGTTCTCG  
TCCTCGAAGTCCGCGAGCTCAATCATAACTTTGGCTTCGCGGATCTTCTTGTGACGCTGT  
TCCACGAATACTCCTCATCTGACGCGCCGAGGCGGGTGGTGACGGGGTACGAGCCTCG  
TTACCAACCGCAGCAGCTGCACTAGAGCGGTATCCACCGCGGAGTCAATTGTTCTTGAAC  
ATCGTTGCATCTTGGCGCCAGCCCAGAACGCGCTGCTGCACCACCACGGAATGAGTGA  
TCTTTAAGGTAGTGTGACACCACTTCTTGGGGTGCAGGAACGCCCGGTAGCCTTGGCA  
GTTTACAGATCCTGAGAGAGCAGAGACTTGACGGTACCGATAGAGTCAATTAAGTGTCTC  
GTAGAAGGAGGGGCAAGCTCGGGTGGTAGACGGCTGGGAACCTGGGTACCCATTAGAGCA  
CGTAGATCAAGTGCATAGCCACGGCGAATCATCCATGTTTAGTCCGGCCGCGGTTATGG  
GGATCAGCCAGGGTCCAGATCCGTTCCAGATCTTTTCTATTAGACCACTTTGGCCCATG  
ATCTCGATGGCAGTCTCGCCACTGATGAAGCCCTTGGGAAACCGTCCCATGCGCGGAAA  
AGTTTCATCGTAAATCTTCTTCTCCTCTCGTAATTGCCAGGCAACTTGGCATTTTCTCT  
GAAAGGCCAACAGCACTGAACCCGCTTACGCCCAGGCTGCGGATGAGCTGCTGCTTTC  
ATAGCATCGATGTTGGGCGAGCCCTGCGCTGGGGTGTGATGAATCCCATTTGACCTGGC  
ATGCCCGTAGGTTGCGCTGCCAGACCTGTAGGCTGCACACTCAGGCTTGCATTTCCACC  
GTCTGGGCGGGGCTGAGGTTGACCCGTAACCAGTAGGCATAGGCGGCATCGGAGGCCGA  
GGGCGGATATCCAGTAGCCTGCGGGTATGTCATGAAACCTGCTGCTGAGGCTGCAGG  
AATTGAGACTGGTCTGCGCCGGGAAGCCCGTCTGGTGGGCTGCTGGATACCACTGCTGC  
TGGGGGAAGAACCAGTAGGCTGTGCGGTGAGTGCAGACGAGGCTGGCATTGTTAGGT  
TGCTGCGGAAGCGGTTGCTGAACAACCTGGGGTGTGCACTTGTCTCTATGATGGCGCA  
TCGAAGTTGGGACGTTTGTTCGTTGTTGGACCGCGCGGCTGTGTATCGGGAACATCG  
AATGAGATGATGCAACCATACTAGAAACCTCGTCTTGTATCGTCTCGGTTAGGGAAGAA  
GGAAGATCGGCTCCGCTCAGACGAATGTTGACAGGTTACATAGCCAATGCGAATTCGGGG  
AATGCAACTGCTCCGATTTAGTGGTATCAGACAAGACCCAGATCTTAGAAAAGATCAGCG  
CCCCAACACTGGAGCGCAACAAAAGATCTCTTGTCTTGTGCGCTATTGTGACCAGTCCAG  
TACTACCATCCAAGCCATAACATGGGGCCCACTGGGTGTACGGAAACCTACCATCCATTG  
TCTTGTATCGCAACCCGAGACTTGAAGAGCTGCTCGAACCTCGCTGATCTGCGCCG  
TGATAAAGAAAGCCGAATGCTCGGGATCTTTGATCCTGTCTTTGGCGGCACCTGAGGAG  
GCGCCCCAGATCCAGCAGCATCTTGAAGGAGTTTGGCATATCGCTAGAGGTCTTGAATC  
GAGTTGATCTGCGGCACCGGAGGAGGCTGCTGTGCTTGGTTAGAGGAGGGTATCCAG  
TGTATTGTTGCTGGAAGCTTTGAGCTTGGCGCTGCTGAGGTTGCTGGGGCGGGCACCCG  
GGAAGCCAGTGAATTTGCGCTGAAGCTGCTGCGGGGAATCCGGTGGCTTGAAGGCTGA  
GCTGCGAGCTACCGAACGGGGAGGGCTGGCCGGCAATCCGGTCCGCTGTGGAGCAATAC  
CAGTGGGCTGCGGAGCAATCCAGTCCGTTGAGGTTGTTGGATTGATGGCGGCTGCTGGC  
CTTGGGGAAATTCGAATAGGGGGTGTGTTGTTGCTGTTGCATGAATGGTGCCTGGCCG  
GACGGCCAGTGGCAGCACCCCGAGGAAATATTCGACGAGGAAAACATCAGAGACACCG  
CGCTTGATGAGCAAAAGGATGCAGTAAAGCCAAAGAAATTGACAAGAAAAACGAATG  
CAAAGTAGTAAGTGAAGTTGAGTCTCAGGAGGATGCGAATGACGCGGTTGCTGGACGGGGA  
CGCACCTGCGCTAACCTCATCGGGCGCTTAGTCATGTGCTCGGGATATTTTGGATCTAC  
GCTCGAAATAGTGGGTTGTGATACAAAGTGAATGATGATGATTTAAAGCTTTTATCTGAAAT  
TTGTGGCATCCGCTGAACGATGGGGATAACTGAGCCTCTCAGAACTGCAAAAAGCAAA  
AAGAAAAGAAAACGCGCCGTAAAAAAAGGGGTGGGATCCCGGGTACCGAGCTC  
GAATTCGCCCTATAGTGAAGTTCGATTAACAATTCAGTGNNNNNNNNNNNNNNNNNNNNN  
NN  
CTGGCGTAATAGCGAAGAGGCCCGCACCAGTCCGCTTCCCAACAGTTGCGCAGCTGAAT  
GGCGAATGGCGCTGATGCGGATTTTCTCCTTACGCATCTGTGCGGATTTTACACCCG  
ATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGATAGTTAAGCCAGCCCCGACAC  
CCGCAACACCCCGCTGACGCGAACCCCTTGGCGCCGCATCGAATAAATCTCGTATAATG  
TATGCTATACGAAGTTATTAGCGATGAGCTCGGACTTCCATTGTTTATTCCACGGACAAA  
AACAGAGAAAGGAAACGACAGAGGCCAAAAGCTCGCTTTCAGCACCTGCTGTTTCTTT  
CTTTTCAGAGGGTATTTTAAATAAAAACATTAAGTTATGACGAAGAAGAACGGAAACGCC  
TTAAACCGGAAAATTTTATAAATAGCGAAAACCCGCGAGGCTGCGCCGCTAACCTGT  
CGGATCACCGGAAAGGACCCGTAAGTGAATAATGATTATCATCTACATATCACAACGTGC  
GTGGAGGCCATCAAAACCGTCAATAATCAATATGACGCGAGTATCGTATTAAKTGRT  
MKGCAATCAAMTYMACGTAATAAKCAAYTTGACACAWTAMWWATCWGCRWCACTKAATACRR  
RGCAACYTCATGWCMSARMTCCGASCTCGWMKACRGCRCACACYTGCATCGGATSCAG  
CYCGGYKARKYGYCGRCASSGMCWSRSTWRCMWSGKATKYWGYCCRSWTAACCGTGCGS  
MAMRKS YTGTTGATAASCAGGACACWGMRCATCCACAGCAGMATASWWSYGSAYAMC  
SAGGWTACTSCGTTTACAGGTTACGACGACAYGWSRATACTTGCSTTKACAGGCATTG  
ATGGAATCGWARTSTCACGMTRATWGTCTSWTCGACAR YAYWRRTGGRAYCGRTRGTCYCA  
SRCYGATARTCWGAYCGACAAYACRAGTGGGAYCGTGGTCCCAGACYRATAATCAGACCG  
ACRAYACGAGTGGGAYCGTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACC  
GTGGTTCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATA  
ATCAGACCGACGATACGAGTGGGACCATGGTCCAGACTAATAATCAGACCGACGATACG  
AGTGGGACCGTGGTCCCAGTCTGATTATCAGACCGACGATACGAGTGGGACCGTGGTCCC  
AGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACC  
GACGATACGAGTGGGACCGTGGTCCCAGTCTGATTATCAGACCGACGATACAAGTGGAA  
AGTGGGCCAGAGAAATATTCAGGCGATTATGCTTTCTGGCCTGTAACAAAGGACATT  
AAGTAAAGACAGATAAACGTAGACTAAAACGTGGTGCATCAGGGTGTGCTGCTTTTCAAG

Appendix

TTCTTAAGAATGGCCTCAATTTCTCTATACACTCAGTTGGAACACGAGACCTGTCCAG  
GTTAAGCACCATTTTATCGCCCTTATACAATACTGTGCGCTCCAGGAGCAAACCTGATGTCG  
TGAGCTTAAACTAGTTCTTGATGCAGATGACGTTTTAAGCACAGAAGTTAAAAGAGTGAT  
AACTTCTTCAGCTTCAAATATCACCCAGCTTTTTTCTGCTCATGAAGGTTAGATGCCTG  
CTGCTTAAAGTAATTCCTCTTTATCTGTAAGGCTTTTTGAAAGTGCATCACCTGACCGGGC  
AGATAGTTACCCGGGTGAGAAAAAGAGCAACAACCTGATTTAGGCAATTTGGCGGTGTT  
GATACAGCGGGTAATAATCTTACGTGAAATATTTCCGCATCAGCCAGCGCAGAATATTT  
TCCAGCAAATTCATTCTGCAATCGGCTTGACATAACGCTGACCAGCTTCATAAGCACTTGT  
TGGGCGATAATCGTTACCCAATCTGGATAATGCAGCCATCTGCTCATCATCCAGCTCGCC  
AACCAGAACACGATAATCACTTTCCGGTAAGTGACGAGCTTTACGACGGCGACTCCCATC  
GGCAATTTCTATGACACCAGATACTCTTCGACCGAACCAGCGGTGTCTGTGACAGTCAG  
TAGAAAAGAAGGGATGAGATCATCCAGTGCCTCAGTAAGCAGCTCCTGGTCACGTTT  
ATTACCTGACCATACCCGAGAGGCTTTCTCAACACTATCACCCGGAGCACTTCAAGAGT  
AAACTTACATCCCGACACATACAGGCAAAGTAATGGCATTACCGCGAGCCATTACTCC  
TACGCGCGCAATTAACGAATCCACCATCGGGGAGCTGGTGTGCATAACGAAGTATCTTC  
AACCGGTTGAGTATTGAGCGTATGTTTTGGAAACAGGCGCAGCTTATTATCTAATC  
TCCAGCGTGGTTAATCAGACGATCGAAAATTCATTGCAGACAGGTTCCCAAATAGAA  
AGAGCATTTCTCCAGGCACAGTTGAAGAGCGTTGATCAATGGCTGTTCAAAAAACAGTT  
CTCATCCGGATCGACCTTACCAACTTCATCCGTTTACGTACAACATTTTTAGAAC  
ATGCTTCCCGAGCATCCGAATTTGCTCCTCCATCCAGGGGACTGAGAGCCATTACTA  
TTGCTGTATTTGGTAAGCAAAATACGTACATCAGGCTCGAACCTTTAAGATCAACGTTT  
TTGAGCAGATCAGGAAGCATATCGAAAACTGCAGTGCAGGAGGTGATGCAAAACAACTCA  
GCAGGCGTGGGAACAATCAGCACATCAGCAGCACATACGACATTAATCGTGCAGGATCC  
AGGTTAGCGCGCTGTCAATAACTATGACATCATAGTCATGAGCAACAGTTTCAATGGCC  
AGTCGGAGCATCAGGTGTGGATCGGTGGCAGTTTACCTTCATCAAATTTGCCATTAA  
TCAGTTTCAATACGGTGCAGAGCCAGACAGGAAGGAATATGTCAAGCCCGGCCAGCAA  
GTGGCTTTTATGATAAGTGACATCGTCTTTTCCCAAGATAGAAAGGCAGGAGAGTG  
TCTTCTGCATGAATATGAAAGATCTGGTACCCATCCGTGATACATGAGGCTGTTCCCTGG  
GGTCTGTTACCTTCCAGGACAAAACACGTAGCCCTTCCAGGCCAGATCCTGAGCAAGA  
TGAACAGAAACTGAGTTTTGTAAACGCCACCTTTATGGGCAGCAACCCCGATCACCGGT  
GGAATACGTCTTCCAGCAGTCGCAATCGCGTACCAACACATCACGCATATGATTAAT  
TGTTCAATTTGATAAACCACACGTTGCTCAACCCGTCCTCGAATTTCCATATCCGGGTGC  
GGTAGTCGCCCTGCTTTCTCGGCATCTCTGATAGCCTGAGAAGAAACCCCACTAAATCC  
GCTGCTTACCTATTCTCCAGCGCCGGTTATTTTCTCGCTTCCGGGCTGTCTCATTA  
AACTGTCAATGGCGATAGCCTTCGTCATTTTATGACCAGCGTTTATGCACTGGTTAAGT  
GTTTCCATGAGTTTTCAATCTGAACATCTTTAATCATTTGCTTTTCTGTTTTTTTAAAT  
CTTGCAATTTACTGCAAAGCAACAACAAAATCGCAAAGTCATCAAAAAACCGCAAAGTTG  
TTTAAAATAAGAGCAACACTACAAAAGGAGATAAGAAGAGCACATACCTCAGTCACTTAT  
TATCACTAGCGCTCGCCGAGCGGTGTAACCGAGCATAGCGAGCGAAGTGGCGAGGAAGC  
AAAGAAGAACTGTTCTGTGATAGCTTACGCTCAGCGCAAGAAGAAATATCCACCGT  
GGGAAAACCTCCAGGTAGAGGTACACACCGGATAGCCAAATTCAGAGTAATAAACTGTGA  
TAATCAACCCCTCATCAATGATGACGAACTAACCCCGATATCAAGTCACATGACGAAGGG  
AAAGAGAAGGAATCAACTGTGACAACTGCCCTCAAATTTGGCTTCTTAAAAATTACA  
GTTCAAAAATAGTAGAAAAATCCATGCAGGCTGAAGGAAAACAGCAAAACTGTGACAAAT  
ACCCTCAGTAGGTGAGAACAAATGTGACGAACCCCTCAAATCTGTGACAGATAACCCCT  
CAGACTATCCTGTGTCATGGAAGTGATATCGCGGAAGGAAAATACGATATGAGTCGTCT  
GGCGGCTTTCTTTTCTCAATGATGAGAGGCGCATTTGGAGTTCTGCTGTTGATCTCAT  
TAACACAGACCTGCAGGAAGCGCGGGAAGTCAAGCATACGCTGTTAACTTTGAGGCA  
GCTGGTAACGCTCTATGATCCAGTGCATTTTTCAGAGAGACGATGCCTGAGCCATCCGGCT  
TAGCATACTGACACAGGATTCGTATAAACGCATGGCATAACGGATTGGTGATTTCTTTG  
TTTCACTAAGCCGAAACTGCGTAAACCGGTTCTGTAAACCGGATAAAGAAGGGAATGAGAT  
ATGGGTTGATATGATACACTGTAAGCCCTCTGGATGGACTGTGCGCAGCTTTTGATAAACC  
AAGGAAAAGATTCATAGCCTTTTTTCTATCGCCGGCATCCTTCCAGGGCGATAAAAAACCA  
CTTCTTCCCGGAAACTCTTCAATGCCTGCCGTATATCCTTACTGGCTTCCGCGAGAGG  
TCAATCCGAATATTTAGCATATTTAGCAACATGGATCTCGCAGATACCGTCAATGTTCTCT  
GTAGGGTCCATCAGATTTTCTGATCTGGTCAACGAACAGATAACAGCATACGTTTGTGAT  
CCCGGGAGAGACTATATGCCCTCAGTGGGTCGTTTACTGGACGATTCGCGGGCTAT  
TTTACGTTTCTGTGATGATAACCGCTGTTTCCGCCATGACAGATCCATGTGAAGTGT  
GACAAGTTTTTAGATTGTCACTAAATAAAAAAGAGTCAATAAGCAGGGATAAATTTGT  
GAAAAAACAGCTTCTTCTGAGGGCAATTTGTACAGGGTTAAGGGCAATTTGTACAGAC  
AGGACTGTCAATTTGAGGGTGTATTGTACACTGAAAGGGCAATTTGTACAAACACCTTCT  
CTAGAACCAGCATGGATAAAGGCTTACAGGCGCTTAAAAAAGAAGATCTAAAAACTAT  
AAAAAAATAATATAAAAAATATCCCGTGGATAAGTGGATAACCCCAAGGGAAGTTTTT  
TCAGGCATCGTGTGTAAGCAGAATATATAAGTGCTGTTCCCTGGTCTTCCGCTCACT  
CGACCGGGAGGGTTGAGAAAGGGGGGACCCCTTCCGGCTGCGGGTACGCGCACA  
GGGCGCAGCCCTGGTTAAAAACAAGGTTTATAAATATTGGTTTAAAAGCAGGTTAAAAGA  
CAGGTTAGCGGTGGCCGAAAAACGGGCGGAAACCCCTGCAATGCTGGATTTTCTGCCTG  
TGGACGCCCTCAAATGTCAATAGGTGCGCCCTCATCTGTACAGCACTTCCCTCAA  
GTGTCAAGGATCGCGCCCTCATCTGTGATGATGCGCCCTCAAGTGTCAATACCGCA  
GGGCACCTATCCCGAGGCTTGTCCACATCATCTGTGGGAAACTCGCGTAAATCAGGCGT  
TTTCGCCGATTTGCGAGGCTGGCCAGCTCCACGTCGCCGGCGAAATCGAGCCTGCCCT  
CATCTGTCAACGCGCGCGCGGGTGAATGCTGCGCCCTCAAGTGTCAACGTCGCCCTCATC  
TGTGATGAGGGCAAGTTTTCCGCGAGGTATCCACAACGCGCGGGCGGGCGGGTGT  
CTGCGCACCGGCTTGCAGGCGTTTCTGGCGCTTTCAGGGCCATAGACGGCCGCCAGC  
CCAGCGCGGAGGGCAACAGCCGAGGGCTTCCGCTGTGCTGACTGCGGGCAGCACTA  
CTGGCTGTAAGAAGGACAGACACATCATGTTTCTGTGTTTCAATAGGTTTCTGTCTGCAAT  
GCTGACATAATCCGCTCCACTTCAACGTAACCCGACGAAGATTTCTATGTTCTGAA

## Appendix

---

GGCATATTCAAATCGTTTTTCGTTACCGCTTGCAGGCATCATGACAGAACACTACTTCCTA  
TAAACGCTACACAGGCTCCTGAGATTAATAATGCGGATCTCTACGATAATGGGAGATTTT  
CCCGACTGTTTCGTTTCGTTTCTCAGTGGATAACAGCCAGCTTCTCTGTTTAAACAGACAAA  
AACAGCATATCCACTCAGTTCACATTTCCATATAAAGGCCAAGGCATTTATTCTCAGGA  
TAATTGTTTCAGCATCGCAACCGCATCAGACTCCGGCATCGCAAACCTGCACCCGGTGCCG  
GGCAGCCACATCCAGCGCAAAAACCTTCGTGTAGACTTCCGTTGAACTGATGGACTTATG  
TCCCATCAGGCTTTCAGAACTTTCAGCGGTATACCGGCATACAGCATGTGCATCGCATA  
GGAATGGCGGAACGTATGTGGTGTGACCGGAACAGAGAACGTACACCCGTCAGCAGCAGC  
GGCGGCAACCGCTCCCAATCCAGGCTCCTGACCGTTCGTCCGTCACCTCCAGATCCG  
CGCTTCTCTGTCTTCCCTGTGCGACGGTTACGCCGCTCCATGAGCTTATCGCGAATAAA  
TACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCG  
GGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCA  
ACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTT  
CAGGAGCTAAGGAAGCTAAAAATGAGAAAAAATCACTGGATAATACCACCGTTGATATAT  
CCCAATGGCATCGTAAAGAACAATTTTGAGGCATTTCACTCAGTTGCTCAATGTACCTATM  
ACSWGWYCATGCACCGGATCTGTTGCCGACAGCAGTGGTAGATCCAGATATACGCAACCG  
CTGTTGGGATGACAAAAAGGTCGATGCGCACCCAGCCATCATTCCGACCGCACGGAGTTC  
TGCATCAACCTGACGGAGAACGAAGCGAAGGTCATATAACCTGATTGCCCGTCAGTATCT  
GATGCAATTCGCCCGGATGCGGTGTTCCGCAAGTGTGTTATCGAACTGGACATTGCCAA  
AGGCAAAATTTGTCGCTAAAGCGGTTTTCTTGTGAAGCAGGCTGGCGCACGCTGTTAGG  
CAGCAAAGAGCGCGATGAAGAAAACGACGGCACGCCACTGCCGTGGTGGCGAAAGGCGA  
TGAGTTGCTGTGTGAAAAAGGTGAAGTGGTAGAGCGGCAAAACCCAGCCGCCGCCATTT  
TACCGATGCAACACTGCTTTCGGCGATGACCGGATCGCGCGCTTTGTGCAGGATAAAGA  
TCTGAAAAAGATCCTCCGTGCGACCGATGGTCTGGGGACAGAGGCAACGCGTGCCGGGAT  
TATTGAACTGTTGTTCAAGCGTGGTTTCTGACCAAAAAGGGCGCTATATCC

## 17 Fgenesh identification of ORFs within contig-7 DNA sequences

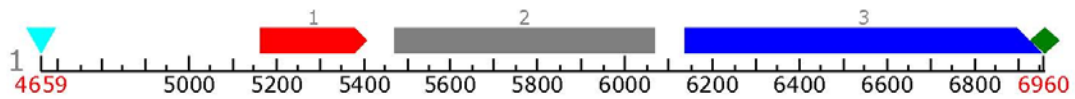
Length of sequence: 46913

Number of predicted genes 10: in +chain 4, in -chain 6.

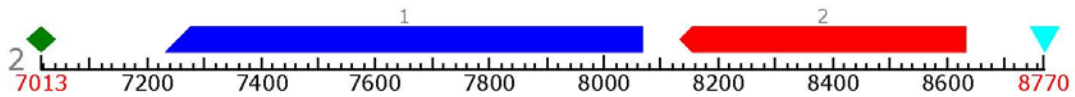
Number of predicted exons 31: in +chain 12, in -chain 19.

Positions of predicted genes and exons: Variant 1 from 1, Score:1554.453516

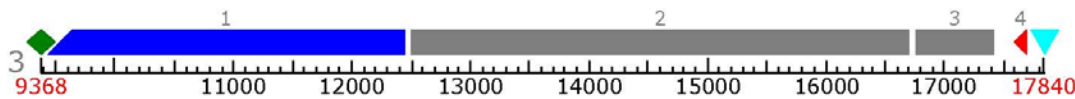
▶ CDSf   
  CDSi   
 ▶ CDSl   
  CDSo   
 ◆ PoLA   
 ▼ TSS



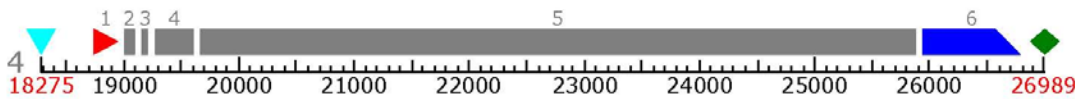
1 +	TSS	4659		-0.02			
1 +	1 CDSf	5162 -	5406	5.56	5162 -	5404	243
1 +	2 CDSi	5469 -	6066	20.09	5470 -	6066	597
1 +	3 CDSl	6135 -	6953	23.81	6135 -	6953	819
1 +	PoLA	6960		3.25			



2 -	PoLA	7013		-3.65			
2 -	1 CDSl	7231 -	8066	30.77	7231 -	8064	834
2 -	2 CDSf	8131 -	8632	31.06	8132 -	8632	501
2 -	TSS	8770		-6.62			



3 -	PoLA	9368		-4.85			
3 -	1 CDSl	9422 -	12440	193.90	9422 -	12439	3018
3 -	2 CDSi	12490 -	16694	217.31	12492 -	16694	4203
3 -	3 CDSi	16748 -	17409	28.21	16748 -	17407	660
3 -	4 CDSf	17580 -	17688	6.48	17581 -	17688	108
3 -	TSS	17840		0.88			

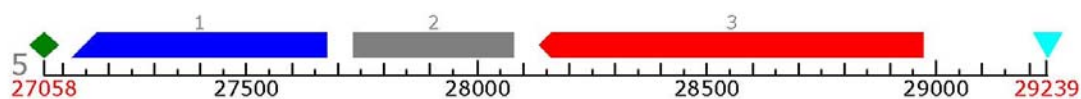


4 +	TSS	18275		-4.12			
4 +	1 CDSf	18733 -	18942	7.74	18733 -	18942	210
4 +	2 CDSi	19000 -	19086	7.81	19000 -	19086	87
4 +	3 CDSi	19145 -	19202	3.74	19145 -	19201	57

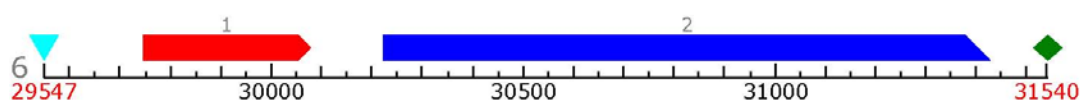


## Appendix

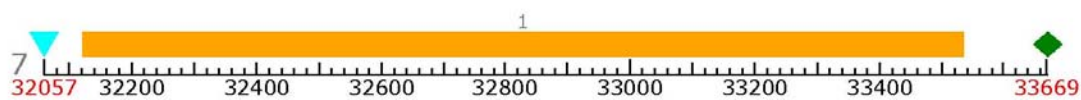
4 +	4	CDSi	19264 -	19597	23.50	19266 -	19595	330
4 +	5	CDSi	19655 -	25868	305.29	19656 -	25868	6213
4 +	6	CDSL	25930 -	26772	63.65	25930 -	26772	843
4 +		PolA	26989		3.25			



5 -		PolA	27058		-4.35			
5 -	1	CDSL	27118 -	27673	18.01	27118 -	27672	555
5 -	2	CDSi	27729 -	28079	1.21	27731 -	28078	348
5 -	3	CDSf	28134 -	28969	26.52	28136 -	28969	834
5 -		TSS	29239		-4.61			



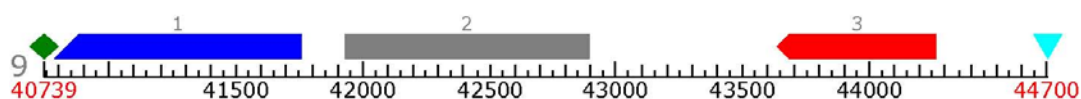
6 +		TSS	29547		-5.62			
6 +	1	CDSf	29744 -	30076	30.02	29744 -	30076	333
6 +	2	CDSL	30221 -	31426	84.22	30221 -	31426	1206
6 +		PolA	31540		0.46			

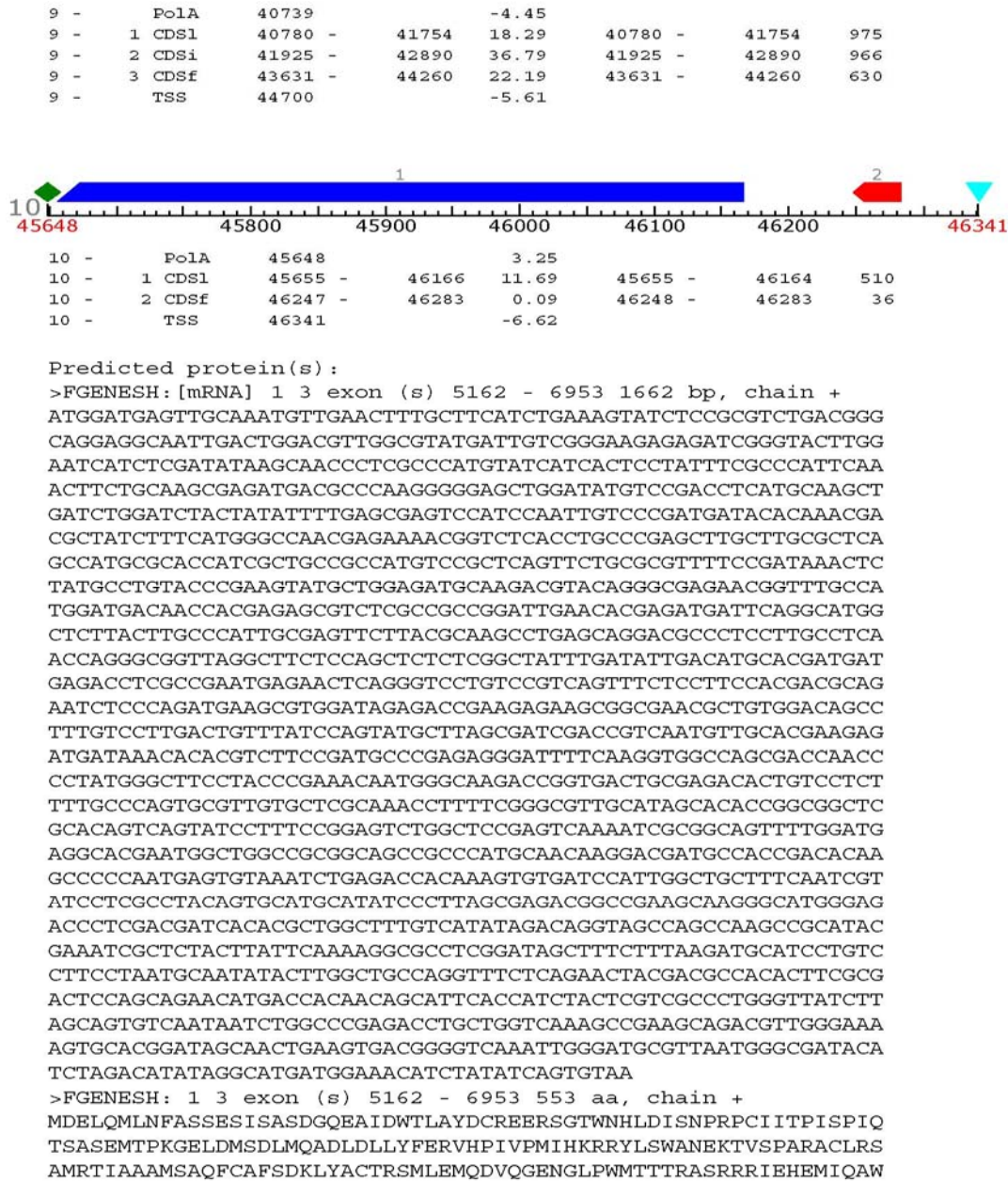


7 +		TSS	32057		-3.01			
7 +	1	CDSo	32119 -	33534	84.34	32119 -	33534	1416
7 +		PolA	33669		3.25			



8 -		PolA	34172		-6.44			
8 -	1	CDSL	34215 -	34977	53.11	34215 -	34976	762
8 -	2	CDSi	35098 -	35385	31.45	35100 -	35384	285
8 -	3	CDSi	35444 -	35966	53.22	35446 -	35964	519
8 -	4	CDSi	36140 -	38143	182.21	36141 -	38141	2001
8 -	5	CDSf	38212 -	38809	39.95	38213 -	38809	597
8 -		TSS	39019		-5.12			





**Figure 43: Identification of ORFs within contig 4 (or 7) with Fgenesh program.**

10 ORF were identified as shown in grey on the left side of each ruler lane (1-10). The number of putative exons that were identified for each ORF is also shown in grey color above each exon. Different types of exons are shown in different colors: CDSf (in red) stands for first exon in an ORF that starts with a start codon; CDSi (in grey) represents internal exon(s) that was identified within an ORF; CDSl (in blue) is a signature for the last coding segment ending with a stop codon. CDSo (in orange) stands for a singular exon within a particular ORF. (+) or (-) on the left side of the exon numeration represents orientation of the detected exon (ORF): direct or opposite DNA chain. TSS (in turquoise) represents position of transcriptional start (TATA-box position). PolA (in green) is a symbol for polyadenylation signal sequence.

## 18 Lists of figures, tables and structures

### 18.1 List of figures

FIGURE 1: FUNGAL POLYKETIDE-DERIVED MYCOTOXINS.....	- 6 -
FIGURE 2: PHARMACEUTICALLY APPROVED FUNGAL SECONDARY METABOLITES.....	- 12 -
FIGURE 3: BIOACTIVE POLYKETIDE METABOLITES FROM MARINE-DERIVED FUNGI.....	- 15 -
FIGURE 4: HPLC CHROMATOGRAM OF METABOLITES ISOLATED FROM <i>P. CHRYSOGENUM</i> STRAIN E01-10/3.....	- 23 -
FIGURE 5: POLYKETIDES PRODUCED BY DIFFERENT <i>P. CHRYSOGENUM</i> STRAINS.....	- 24 -
FIGURE 6. THE CHEMISTRY OF POLYKETIDE CHAIN ASSEMBLY.....	- 29 -
FIGURE 7 : DIFFERENT CLASSES OF POLYKETIDE SYNTHASES.....	- 31 -
FIGURE 8. GENERAL DOMAIN ORGANISATION OF NR PKSS.....	- 35 -
FIGURE 9. DOMAIN ORGANISATION OF MSAS.....	- 35 -
FIGURE 10. GENERAL DOMAIN ORGANISATION OF HR PKSS.....	- 36 -
FIGURE 11: POSSIBLE BIOSYNTHETIC ROUTES TO SORBICILLACTONE A [136].....	- 47 -
FIGURE 12: PRODUCTION OF SORBICILLACTONE A BY <i>P. CHRYSOGENUM</i> E01-10/3 LIQUID CULTURES.....	- 51 -
FIGURE 13: PCR AMPLIFICATION WITH LC SERIES KS-SPECIFIC PRIMERS - ETHIDIUM BROMIDE AGAROSE GEL.....	- 53 -
FIGURE 14: - PCR WITH THE KHKS SERIES OF KS-SPECIFIC PRIMERS - ETHIDIUM BROMIDE STAINED AGAROSE GEL.....	- 55 -
FIGURE 15: ALINGMENT FOR MEMBERS OF NR PKS CLADE III - KS DOMAIN.....	- 58 -
FIGURE 16: PCR AMPLIFICATION WITH AT-SPECIFIC PRIMERS - ETHIDIUM BROMIDE AGAROSE GEL.....	- 65 -
FIGURE 17: ALINGMENT FOR MEMBERS OF NR PKSB CLADE III - AT DOMAIN.....	- 66 -
FIGURE 18: PCR AMPLIFICATION WITH CMeT SERIES OF MT-SPECIFIC PRIMERS - ETHIDIUM BROMIDE AGAROSE GEL.....	- 72 -
FIGURE 19: ALINGMENT FOR MEMBERS OF NR PKS CLADE III - MT DOMAIN.....	- 74 -
FIGURE 20: ETHIDIUM BROMIDE-STAINED AGAROSE GEL FOR RED DOMAIN AMPLIFICATION VIA CMeT1 AND RED-R1B PRIMER PAIR.....	- 78 -
FIGURE 21: ALINGMENT FOR MEMBERS OF NR PKS CLADE III - RED DOMAIN.....	- 80 -
FIGURE 22: PHYLOGENY OF THE PCR AMPLIFIED PARTIAL KS DOMAINS FROM <i>P. CHRYSOGENUM</i> E01-10/3 STRAIN.....	- 87 -
FIGURE 23: SOUTHERN HYBRIDIZATION WITH KSNR7 PROBE.....	- 94 -
FIGURE 24: HYBRIDIZATION WITH RED DOMAIN PROBE FOR FUNGAL NR PKS.....	- 95 -
FIGURE 25: CLONING <i>HIND</i> III RESTRICTION FRAGMENTS OF 17B9 CLONE CANDIDATE.....	- 98 -
FIGURE 26: PCR SCREENING FOR KHKS1 GENE CLUSTER.....	- 102 -
FIGURE 27: <i>BAM</i> HI RESTRICTION PROFILING OF KHKS1 GENE CLUSTER.....	- 102 -
FIGURE 28: PCR SCREENING FOR AT-9-11 GENE CLUSTER.....	- 107 -
FIGURE 29: <i>BAM</i> HI RESTRICTION PROFILING OF AT-9-11 CLUSTER FOSMID CLONES.....	- 108 -
FIGURE 30: PCR FOR THE KHKS32 CLUSTER.....	- 117 -
FIGURE 31: PROPOSED GENE ORGANIZATION OF PUTATIVE SORBICILLACTONE GENE CLUSTER.....	- 121 -
FIGURE 32: PHYLOGENY OF THE CLONED KS DOMAINS FROM PUTATIVE PKS OF <i>P. CHRYSOGENUM</i> E01-10/3 STRAIN.....	- 138 -
FIGURE 33: PKS DOMAIN CONSENSUS MOTIFS WITHIN <i>P. CHRYSOGENUM</i> PKSSL1/ORF3.....	- 142 -
FIGURE 34: PHYLOGENY OF PT DOMAIN FROM PKSSL1 (ORF3) OF <i>P. CHRYSOGENUM</i> E01-10/3 STRAIN.....	- 143 -
FIGURE 35: DOMAIN ORGANISATION OF <i>P. CHRYSOGENUM</i> PKSSL1.....	- 144 -
FIGURE 36: PKS DOMAIN CONSENSUS MOTIFS WITHIN <i>P. CHRYSOGENUM</i> PKSSL2/ORF4.....	- 148 -
FIGURE 37: DOMAIN ORGANISATION OF <i>P. CHRYSOGENUM</i> PKSSL2.....	- 149 -
FIGURE 38: PKSSL2 PUTATIVE BIOSYNTHESIS ROUTE.....	- 151 -
FIGURE 39: PKSSL1 – SORBICILLACTONE A BIOSYNTHESIS ROUTE.....	- 152 -
FIGURE 40: FEEDING OF <sup>13</sup> C <sub>2</sub> -LABELLED SORBICILLIN IN <i>P. CHRYSOGENUM</i> .....	- 155 -
FIGURE 41: PUTATIVE SUBSTRATES FOR SORBICILLACTONE PKS: SORBIC ACID AND ITS THIOESTERS.....	- 156 -
FIGURE 42: UV SPECTRUM OF SORBICILLACTONE A. LINKS: EXTRACTED COMPOUND FROM THE LIQUID CULTURE OF <i>P. CHRYSOGENUM</i> . RECHTS: REFERENCE COMPOUND SORBICILLACTONE A.....	- 218 -
FIGURE 43: IDENTIFICATION OF ORFs WITHIN CONTIG 4 (OR 7) WITH FGENESH PROGRAM.....	- 241 -

### 18.2 List of tables

TABLE 1: BIOACTIVE POLYKETIDE METABOLITES FROM MARINE-DERIVED FUNGI.....	- 14 -
--	--------

TABLE 2: REPORTED DIVERSITY OF POLYKETIDES PRODUCED BY <i>P. CHRYSOGENUM</i> .....	- 23 -
TABLE 3: FUNGAL PKS GENES.....	- 32 -
TABLE 4: SEQUENCES USED IN ALIGNMENT FOR MEMBERS OF NR PKS CLADE III - 19 SEQUENCES ALIGNMENT.....	- 56 -
TABLE 5: SEQUENCE SIMILARITIES OF THE DEDUCED AMINO ACID SEQUENCES OF THE AMPLIFIED PARTIAL KS FRAGMENTS FROM <i>P. CHRYSOGENUM</i> E01-10/3.....	- 62 -
TABLE 6: SEQUENCE SIMILARITIES OF THE DEDUCED AMINO ACID SEQUENCES OF THE AMPLIFIED PARTIAL MT FRAGMENTS FROM <i>P. CHRYSOGENUM</i> E01-10/3.....	- 70 -
TABLE 7: SEQUENCES USED FOR THE DEDUCED PROTEIN ALIGNMENT FOR MEMBERS OF NR PKS CLADE III FOR DESIGN OF MT AND RED DOMAIN DEGENERATE PRIMERS.....	- 73 -
TABLE 8: SEQUENCE SIMILARITIES OF THE DEDUCED AMINO ACID SEQUENCES OF THE AMPLIFIED PARTIAL MT FRAGMENTS FROM <i>P. CHRYSOGENUM</i> E01-10/3.....	- 76 -
TABLE 9: SEQUENCES USED IN DEDUCED PROTEIN ALIGNMENT FOR MEMBERS OF NR PKS CLADE III FOR DESIGN OF RED DOMAIN SPECIFIC DEGENERATE PRIMERS.....	- 78 -
TABLE 10: PROTEIN SEQUENCES INCLUDED IN THE MULTIPLE ALIGNMENT AND PHYLOGENETIC ANALYSIS OF PCR AMPLIFIED <i>P. CHRYSOGENUM</i> PARTIAL KS DOMAINS.....	- 91 -
TABLE 11: SEQUENCES HOMOLOGIES OF DEDUCED AMINO ACID SEQUENCES FROM 17B9 SUBCLONED FRAGMENTS.....	- 99 -
TABLE 12: SEQUENCES HOMOLOGIES OF DEDUCED AMINO ACID SEQUENCES FOR KHKS1 GENE CLUSTER.....	- 105 -
TABLE 13: RESULTS OF END SEQUENCING OF SUBCLONED <i>BAM</i> HI RESTRICTED FRAGMENTS FROM KHKS1 GENE CLUSTER.....	- 105 -
TABLE 14: END SEQUENCING OF FOSMID CLONES FROM <i>P. CHRYSOGENUM</i> GENOMIC LIBRARY CANDIDATES FOR AT-9-11 GENE CLUSTER.....	- 109 -
TABLE 15: RESULTS OF END SEQUENCING OF SUBCLONED <i>BAM</i> HI RESTRICTED FRAGMENTS FROM AT-9-11 GENE CLUSTER.....	- 114 -
TABLE 16: FGENESH INFERRED ORFS WITHIN THE SHOTGUN SEQUENCED FOSMID 42H12.....	- 122 -
TABLE 17: RESULTS OF SUBCLONING AND SUBSEQUENT SEQUENCING OF FRAGMENT 2 FROM THE AT-9-11 CLUSTER.....	- 131 -
TABLE 18: INFERRED NON-PKS GENE HOMOLOGIES RETRIEVED VIA SUBCLONING OF PUTATIVE AT-9-11 GENE CLUSTER.....	- 132 -
TABLE 19: PROTEIN SEQUENCES FOR MULTIPLE ALIGNMENT AND PHYLOGENETIC ANALYSIS OF KS DOMAINS FROM THE WHOLE-LENGTH PKS.....	- 138 -
TABLE 20: REFERENCE SEQUENCES USED FOR PT DOMAIN PHYLOGENETIC ANALYSIS.....	- 144 -
TABLE 21: REFERENCE SEQUENCES USED FOR <i>P. CHRYSOGENUM</i> PKSSL2 DOMAIN ACTIVE SITE ANALYSIS.....	- 148 -
TABLE 22: USED STRAINS AND THEIR CHARACTERISTICS.....	- 169 -
TABLE 23: USED VECTORS AND THEIR CHARACTERISTICS.....	- 169 -
TABLE 24: USED VECTORS AND THEIR CHARACTERISTICS.....	- 169 -
TABLE 25: USED ANTIBIOTICS AND THEIR CONCENTRATIONS.....	- 171 -
TABLE 26: USED ENZYMES AND THEIR BUFFERS.....	- 172 -
TABLE 27: USED PRIMERS AND THEIR CHARACTERISTICS.....	- 172 -
TABLE 28: USED EQUIPMENT AND MATERIAL.....	- 175 -
TABLE 29: USED CHEMICALS AND SOLVENTS.....	- 177 -
TABLE 30: USED COMMERCIAL KITS AND THEIR ORIGIN.....	- 178 -
TABLE 31: USED PROGRAMS AND INTERNET SOURCES FOR ANALYSIS OF BIOLOGICAL SEQUENCE INFORMATION.....	- 178 -
TABLE 32: NR PKS CLADE III SPECIFIC KS DOMAIN DEGENERATE PRIMER DESIGN:.....	- 219 -
TABLE 33: NR PKS CLADE III SPECIFIC AT DOMAIN DEGENERATE PRIMER DESIGN.....	- 220 -
TABLE 34: NR PKS CLADE III SPECIFIC MT DOMAIN DEGENERATE PRIMER DESIGN.....	- 222 -
TABLE 35: NR PKS CLADE III SPECIFIC RED DOMAIN DEGENERATE PRIMER DESIGN. THE FIVE SEQUENCES ALIGNMENT.....	- 223 -
TABLE 36: NR PKS CLADE III SPECIFIC RED DOMAIN DEGENERATE PRIMER DESIGN. THE FIVE SEQUENCES ALIGNMENT.....	- 224 -
TABLE 37: PREDICTED SIZE OF PCR PRODUCTS FOR DIFFERENT COMBINATIONS OF FORWARD AND REVERSE PRIMERS - AMPLIFICATION OF RED DOMAIN.....	- 225 -

### 18.3 List of structures

1 AFLATOXIN B <sub>1</sub> .....	- 6 -
2 CITRININ.....	- 6 -
3 PATULIN.....	- 6 -

## Appendix

4 OCHRATOXIN A .....	- 6 -
5 FUMONISIN B1 .....	- 6 -
6 PENICILLIN G .....	- 12 -
7 MYCOPHENOLIC ACID .....	- 12 -
8 ERGOTAMINE .....	- 12 -
9 ZEARALENON .....	- 12 -
10 COMPACTIN .....	- 12 -
11 LOVASTATIN .....	- 12 -
12 CYCLOSPORIN A .....	- 12 -
13 TRICHODERMATIDE A-D .....	- 15 -
14 REDOXICITRININ .....	- 16 -
15 CLADIONOL A .....	- 16 -
16 SHIMALACTONE A .....	- 16 -
17 DIHYDROBISVERTINOLONE AND TETRAHYDROBISVERTINOLONE .....	- 16 -
18 2-(2',3-DIHYDROSORBYL)-3,6-DIMETHYL-5-HYDROXY-1,4-BENZOQUINONE .....	- 16 -
19 3-ACETONYL-2,6-DIMETHYL-5-HYDROXY-1,4-BENZOQUINONE .....	- 17 -
20 PENICILLONES .....	- 17 -
21 SPORIOLIDES .....	- 17 -
22 ASPERMYTIN A .....	- 17 -
23 PHOMOXIN .....	- 17 -
24 PHOMOXIDE .....	- 17 -
25 EPICOCONE .....	- 17 -
26 EVARIQUINONE .....	- 17 -
27 ISOEMERICELLIN .....	- 18 -
28 PARASITENONE .....	- 18 -
29 VARITRIOL, VARIOXIRANE, VARIXANTHONE .....	- 18 -
30 PAECILOSPIRONE .....	- 18 -
31 14, 15-SECOCURVULARIN .....	- 18 -
32 SCULEZONONE .....	- 18 -
33 PENOCHALASINS .....	- 19 -
34 EPOLACTAENE .....	- 19 -
35 OBIONIN A .....	- 19 -
36 7-ISOPROPENYLBI-CYCLO[4.2.0]OCTA-1,3,5-TRIENE-2,5-DIOL .....	- 19 -
37 METHYLXANTHONE .....	- 19 -
38 DEOXYASPERLACTONE .....	- 20 -
39 9-CHLORO-8-HYDROXY-8,9-DEOXYASPYRONE .....	- 20 -
40 SORBICILLIN .....	- 24 -
41 2', 3'-DIHYDROSORBICILLIN .....	- 24 -
42 SORRETANONE .....	- 24 -
43 SOHIRNONES A .....	- 24 -
44 SOHIRNONES B .....	- 24 -
45 SOHIRNONES C .....	- 24 -
46 OXOSORBICILLINOL .....	- 24 -
47 TAUTOMER OF OXOSORBICILLINOL .....	- 24 -
48 SORBIVINETONE .....	- 24 -
49 7-DEACETOXYANUTHONE .....	- 25 -
50 SORBICILLACTONE A .....	- 25 -
51 SORBICILLACTONE B .....	- 25 -
52 BISORBICILLINOL .....	- 25 -
53 BISVERTINOQUINOL (16, 17-DIHYDRO) .....	- 25 -
54 BISORBIBUTENOLIDE .....	- 25 -
55 REZISHANONES A-D .....	- 26 -
56 TRICHODIMEROL .....	- 26 -
57 BISVERTINOLONE .....	- 26 -
58 XANTHOVIRIDICATIN E .....	- 27 -
59 XANTHOVIRIDICATIN F .....	- 27 -
60 SECALONIC ACID .....	- 27 -
61 EMODIC ACID .....	- 27 -
62 $\Omega$ -HYDROXY EMODIN .....	- 27 -
63 SORBICILLACTONE A HEXAKETIDE INTERMEDIATE .....	- 47 -

## Appendix

---

64 SORBICILLINOL .....	- 47 -
65 INTERMEDIATE OF SORBICOLLACTONE A BIOSYNTHESIS.....	- 47 -
66 INTERMEDIATE OF SORBICOLLACTONE A BIOSYNTHESIS.....	- 47 -
67 INTERMEDIATE OF SORBICOLLACTONE A BIOSYNTHESIS.....	- 47 -
68 INTERMEDIATE OF SORBICOLLACTONE A BIOSYNTHESIS.....	- 47 -
69 SORBICILLACTONE A STARTER UNIT .....	- 151 -
70 SORBICILLACTONE B STARTER UNIT.....	- 151 -
71 SORBICILLACTONE B HEXAKETIDE INTERMEDIATE .....	- 152 -
72 SORBIC ACID.....	- 156 -
73 THIOESTER PRECURSOR .....	- 156 -
74 THIOESTER PRECURSOR .....	- 156 -

## 19 Curriculum Vitae

### Marija Avramović

Address: Leonardo-da-Vinci-Allee 5  
60486 Frankfurt/Main

Date of birth: April 27<sup>th</sup> 1974

Place of birth: Belgrade, Serbia

Profession: Molecular Biologist and Physiologist

Email: marija\_avramovic@yahoo.co.uk

EDUCATION / QUALIFICATION	<p>02. 2005—present, Rheinischen Friedrich-Wilhelms-Universität-Bonn, Mathematical-Natural Science Faculty, Kekulé-Institut for Organic Chemistry and Biochemistry, PhD program</p> <p>1993—2003, Faculty of Biology, University of Belgrade Study group: Molecular Biology and Physiology, B.Sc. program Section: Experimental Biomedicine Awarded the title: <b>Graduate Molecular Biologist and Physiologist</b> Average grade: 9.49 (of 10) Graduate thesis: “Detection of Genes for Stx1 and Stx2 Toxins of Shiga Toxin-Producing <i>Escherichia coli</i> (STEC) Strains”, graded with mark: 10</p> <p>1989—1993, “Mihajlo Pupin” Gymnasium in Belgrade Awarded the title: <b>Associate for Natural Sciences</b></p>
PROFESSIONAL EXPERIENCE	<p>04.2009—present, <b>Scientific Advisor – Device Technology</b>, YES Medical Device Services GmbH, Friedrichsdorf (Frankfurt am Main)</p> <p>02. 2005—03.2008, RFW-Universität, Kekulé-Institut for Organic Chemistry and Biochemistry</p> <p>07. —10. 2004, Johann Wolfgang Goethe-University, Clinic of Neurology (Molecular Neurogenetics) Frankfurt am Main. Visiting scientist on the project: “Functional Studies of Spinocerebellar Ataxia Type 2 Disease Protein using Knock-out Gene Technology”.</p> <p>2002—2003, Military Medical Academy in Belgrade, Institute of Epidemiology Experimental graduate thesis work: “Detection of Genes for Shiga Toxins in enterohemorrhagic <i>Escherichia coli</i> strains using Polymerase Chain Reaction (PCR) method”.</p>
ATTENDED COURSES	<p>19—25. 08. 2007 and 21—26. 08. 2006, BIOTECmarin Summer School: Lectures and practical training in different aspects of marine organisms investigation (ecological, genetic, natural product production); diving and snorkeling excursions.</p> <p>27—28. 07. 2004, EUROSCA Core Facility Training in Microarrays for Expression Analysis, University of Tübingen, Department of Medicinal Genetics.</p>