Ambigols A-C and tjipanazole D: bioinformatic analysis of their putative biosynthetic gene clusters

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Mustafa El Omari

aus

Münster, Deutschland

Bonn 2011

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

1. Gutachterin : Prof. Dr. G. M. König
2. Gutachter : Priv.-Doz. Dr. habil. W. Knöss

Tag der Promotion : 25.07.2011

Erscheinungsjahr : 2011

I Publications

Research papers

Wagner C, El Omari M, Konig GM. 2009. Biohalogenation: nature's way to synthesize halogenated metabolites. *J.Nat Prod* **72:** 540-553.

Erol Ö, Schäberle TF, Schmitz A, Rachid S, Gurgui C, El Omari M, *et al*., 2010. Biosynthesis of the myxobacterial antibiotic corallopyronin A. *Chembiochem* **11**: 1253-1265.

Research presentations

El Omari M & König GM. Halogenases in the cyanobacterium *Fischerella ambigua*. Poster presentation. Annual conference of the Deutsche Pharmazeutische Gesellschaft (DPHG), Bonn, Germany, October 8-11, 2008.

El Omari M, Wagner C, König GM. Halogenases in the cyanobacterium *Fischerella ambigua*. Poster presentation. 6th European Conference on Marine Natural Products (ECMNP), Porto, Portugal, July 19-23, 2009.

El Omari M & König GM. New insights into the biosynthesis of ambigols A-C and tjipanazole D. Poster presentation. Annual Conference of the Association for General and Applied Microbiology (VAAM), Karlsruhe, Germany, April 3-6, 2011.

II Acknowledgements

My sincere gratitude goes to Prof. Dr. G.M. König for giving me the opportunity to deal with a highly comprehensive topic involving various branches of biosynthesis and very interesting methods of molecular biology. I would like to thank her for supervising my project with expert guidance, encouragement and inspiring ideas as well as for careful proof-reading of my work. I would further like to appreciate her trust in me to take on further responsibility as assistant appointee for biological safety.

I would like to express my gratitude to Priv.-Doz. Dr. habil. W. Knöss for accepting the co-examination of this project.

My appreciation also goes to Prof. Dr. G. Bendas and Prof. Dr. E.A. Galinski for agreeing to participate in my examination committee.

This study involved many different and difficult tasks and problems to be solved. This was extremely facilitated by the nice work atmosphere and the support of members of our group. I would especially thank:

Dr. C. Wagner for introducing me to this project and for guiding me on cultivation of the cyanobacterial strain and DNA-extraction. I also thank her for providing partial halogenase sequences to me for genomic library screening.

Ö. Erol for familiarising me with the basic molecular biological laboratory work and always being a helpful and advising companion and special friend during the whole study. Furthermore, I would like to thank her for help with the construction of the genomic library and her good ideas concerning problems of the laboratory work.

Sarah Bouhired for organising the 454 sequencing of the *Fischerella ambigua* strain and explaining the CLUSEAN software.

Ms. E. Neu and Ms. E. Eguereva for technical and administrative assistance during the course of the project. Special thanks go to Ms. E. Neu and Ms. E. Eguereva for support with the cultivation of the cyanobacterial strain and ambigol-isolation.

S. Kehraus for being always helpful concerning general laboratory questions and for his humurous comments.

M. Elsebai for being always a helpful and funny fellow and good friend during this study.

Further, my thanks go to Prof. J. Piel for good advice on biosynthesis questions concerning my project in our collaborative group seminars.

I also would like to address special thanks to Dr. H. Groß for enriching my experience in teaching students and the nice and familiar atmosphere during the plant anatomy courses. Furthermore, I acknowledge his precious ideas in discussions.

My appreciation goes to Ms. E. Gassen and Dr. M. Koch for help with administrative issues and to Mr. T. Kögler for technical support.

Finally, I express my cordial gratitude to my family, especially to my parents Samir and Irmgard El Omari for their permanent encouragement and trust in me. I would like to thank my whole family for their extensive support during the course of this project.

III Table of content

IV Abbreviations

AAA-ATPase-like proteins	ATPases associated with diverse cellular activities
ACP	Acyl carrier protein
A domain	Adenylation domain
ArCP	Aryl carrier protein
Asn	Asparagine
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
bp	Base pairs
CoA	Coenzyme A
CPA	Chromopyrrolic acid
CPAS	Chromopyrrolic acid synthase
CTAB	Cetyl trimethyl ammonium bromide
CYP 450	Cytochrome P 450
DHB	Dihydroxybenzoic acid
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DMSO	Dimethyl sulfoxide
dNTP	Deoxynucleotide triphosphates
dATP	Deoxyadenosine triphosphate
EDTA	Ethylenedinitrolotetraacetic acid
FAD	Flavin adenine dinucleotide
$FADH_2$	Flavin adenine dinucleotide (reduced form)
FAS	Fatty acid synthase
4-HBA	4-Hydroxybenzoic acid
ICL	Isochorismate lyase

kb	Kilo base pairs (= kbp = 1,000 bp)
Lys	Lysine
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide (reduced form)
NCBI	National Center of Biotechnlology Information
NRPS	Nonribosomal peptide synthetases
OD	Optical density
ORF	Open reading frame
PCI	Phenol:Chloroform:Isoamyl alcohol
PCP domain	Peptidyl carrier protein domain
RNA	Ribonucleic acid
RNase	Ribonuclease
SAM	S-adenosyl-L-methionine
SDS	Sodium dodecyl sulfate
sp./spp.	Species /species (plural)
TE domain	Thioesterase domain
Trp	Tryptophan
UV	Ultraviolet light
WD-repeat containing proteins	Tryptophan-aspartate-repeat containing proteins

One letter code for amino acids

A	Alanine
C	Cysteine
D	Aspartic acid
E	Glutamic acid
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
W	Tryptophan
Y	Tyrosine

1. Introduction

1.1 Cyanobacterial secondary metabolites

The current study investigated the biosynthesis of the halogenated secondary metabolites ambigol A, B, C and tjipanazole D of the cyanobacterium *Fischerella ambigua*.

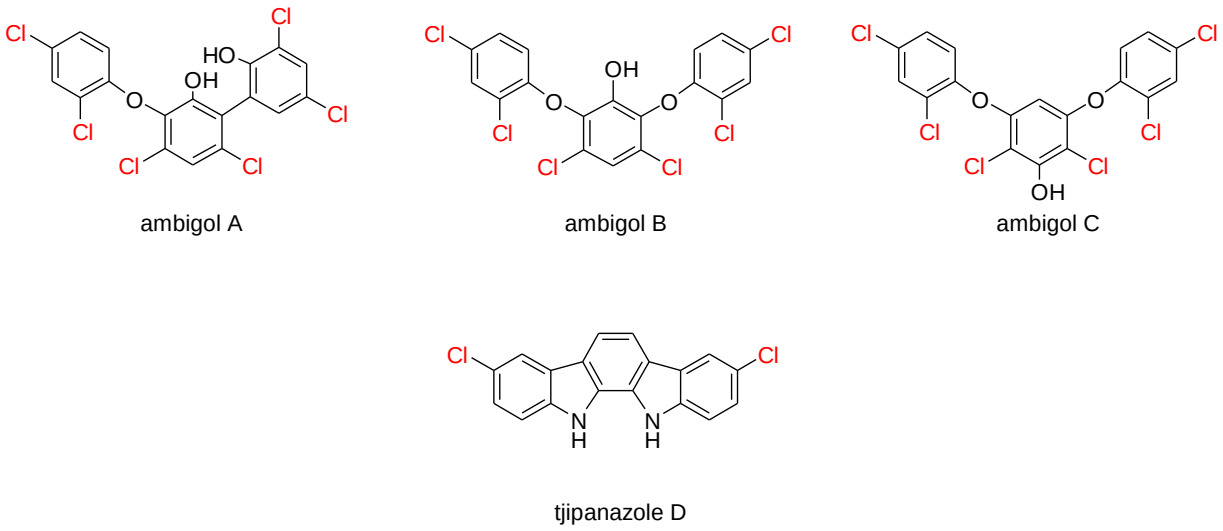


Figure 1.1-1: Structure of the polychlorinated phenolic ethers ambigols A-C and the indolocarbazole compound tjipanazole D, produced by the terrestrial cyanobacterium *Fischerella ambigua*.

Cyanobacteria represent an extremely diverse group of Gram-negative photoautotrophic prokaryotes. This concerns especially their morphological and physiological attributes (Gugger & Hoffmann, 2004; Teaumroong *et al.*, 2002) but also their production of a wide range of secondary metabolites (Van Wagoner *et al.*, 2007; Nett & Koenig, 2007). These metabolites play an ecological role in that they support cyanobacterial blooms or enable cyanobacteria to defend their habitat against other organisms like fungi, bacteria and vertebrates (Leao *et al.*, 2010; Nagle & Paul, 1998).

Cyanobacteria form symbiotic associations with different organisms (Svenning *et al.*, 2005; Meeks & Elhai, 2002), e.g. in lichen (Luecking *et al.*, 2009), sponges (Hentschel *et al.*, 2006; Hentschel & Bringmann, 2010) and plants (Adams & Duggan,

2008). Three major groups of cyanobacterial symbionts have been reported from marine sponges: the filamentous *Oscillatoria spongeliae*, the unicellular "*Candidatus Synechococcus spongiarum*" and *Synechocystis* species (Lemloh *et al.*, 2009). Since cyanobacterial symbionts contribute to the spectrum of sponge-derived compounds, structural characteristics of cyanobacterial chemistry often debunk the photosymbiont as being the true source of secondary metabolites (Flatt *et al.*, 2005; Moore, 1996).

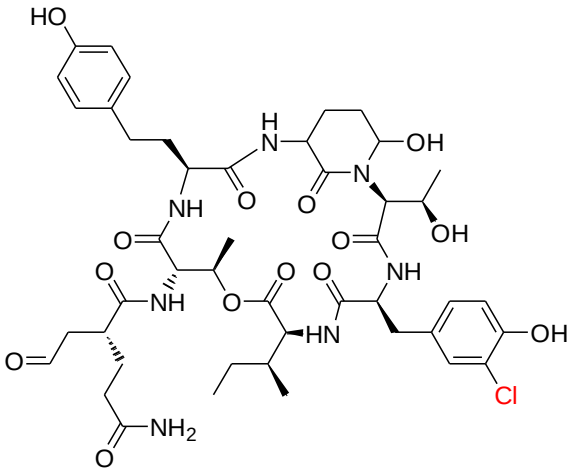
Natural products from cyanobacteria include toxins (Wiegand & Pflugmacher, 2005) like hepatotoxins (mentioned below), neurotoxins, e.g. jamaicamides (Edwards *et al.*, 2004) as well as dermatotoxins, e.g. lyngbyatoxin A (Cardellina *et al.*, 1979). The most prevalent class of hepatotoxins is represented by monocyclic oligopeptides, the microcystins and nodularins. They are produced by a hybrid nonribosomal-polyketide megasynthase (NRPS/PKS) (Dittmann & Wiegand, 2006; Moffitt & Neilan, 2004; Rouhiainen *et al.*, 2004). Microcystins and nodularins are potent inhibitors of serine/threonine protein phosphatases 1 and 2A (Yoshizawa *et al.*, 1990; Nishiwaki-Matsushima *et al.*, 1991). By contrast, numerous cyanobacterial secondary metabolites have promising biological activities regarding their therapeutic benefit (Liu & Rein, 2010; Jones *et al.*, 2009; Tan, 2007; Singh *et al.*, 2005). In this context the cyanobactins shall be mentioned, which form a new class of ribosomally biosynthesised, low-molecular-weight cyclic peptides (Sivonen *et al.*, 2010). The cyanobactin synthetase gene is wide-spread among cyanobacterial genera and congeners of these compounds show high structural diversity (Leikoski *et al.*, 2010; Leikoski *et al.*, 2009; Donia *et al.*, 2008). They include patellamides B, C, and D that were reported to reverse multidrug resistance (Jones *et al.*, 2009; Schmidt *et al.*, 2005).

Among many other pharmacological properties, cyanobacterial metabolites show activity against pathogenic protozoa, e.g. the antileishmanial lipopeptides almiramides B-C and the closely related dragonamide E, which all were obtained from Panamanian strains of *Lyngbya majuscula* (Sánchez *et al.*, 2010; Balunas *et al.*, 2010). Another lipopeptide, dragonamide A exhibited good activity against *Plasmodium falciparum* (McPhail *et al.*, 2007). The knowledge of involved genes in the biosynthesis of these compounds may open up affordable alternatives to produce effective drugs for the treatment of orphan diseases.

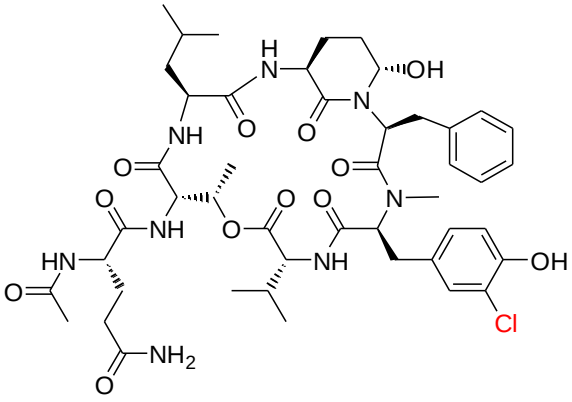
1.2 Halogenated secondary metabolites from cyanobacteria

The most frequently isolated cyanobacterial metabolites are peptides and depsipeptides, e.g. aeruginosins, microginins, cyanopeptolins and anabaenopeptins (Silva-Stenico *et al.*, 2010; Mehner *et al.*, 2008; Welker & von Doehren, 2006). Most structural congeners display inhibitory activity towards serine proteases, e.g. trypsin, human leucocyte elastase (cyanopeptolins) and thrombin (aeruginosins) (Sisay *et al.*, 2009; Ishida *et al.*, 2009; Cadel-Six *et al.*, 2008). The structural diversity of these oligopeptides and other cyanobacterial metabolites is due to variability of amino acid residues and tailoring reactions like glycosylation, sulfatation, methylation, or halogenation (Jones *et al.*, 2010; Ishida *et al.*, 2007; Rouhiainen *et al.*, 2000; von Elert *et al.*, 2005). These post-NRPS modifications explain the broad variety of derivatives in each of these peptide classes (Welker *et al.*, 2006). Halogenated compounds are frequently found in cyanobacteria, and were described, e.g. among aeruginosin-, cyanopeptolin- and microginin-type peptides (Rounge *et al.*, 2007; Welker & von Doehren, 2006). In cyanopeptolin- and aeruginosin-like oligopeptides, chlorination is performed on aromatic moieties by $FADH_2$-dependent halogenases (Vaillancourt *et al.*, 2006; van Pée & Patallo, 2006).

Genes encoding for enzymes involved in the biosynthesis of cyanopeptolins have been isolated and characterised as NRPS-gene clusters for *Anabaena* strain 90 (*apd*) (Rouhiainen *et al.*, 2000), the producer of anabaenopeptilide 90 B, and for *Planktothrix agardhii* NIVA CYA 116 (*oci*) (Tooming-Klunderud *et al.*, 2007), a source of cyanopeptolin 954 [figure 1.2-1].



anabaenopeptilide 90B
(Rouhiainen *et al.*, 2000)



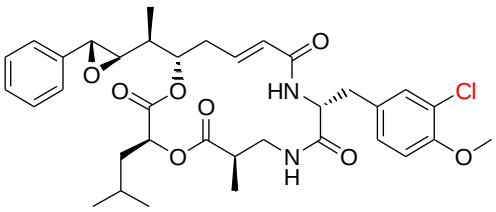
cyanopeptolin 954
(Tooming-Klunderud *et al.*, 2007)

Figure 1.2-1: Two cyanobacterial cyclic peptides, which belong to the large class of cyanopeptolins. Anabaenopeptilide 90B from *Anabaena* strain 90 and cyanopeptolin 954 from *Microcystis* NIVA-CYA 172/5 contain a chlorinated tyrosyl moiety.

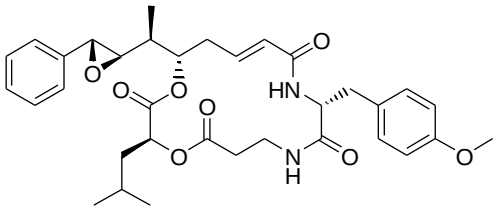
Several cyanopeptolin variants are known to occur as halogenated compounds along with non-halogenated ones. For instance, from *Anabaena* strain 90, anabaenopeptilide 90B and its non-chlorinated analogue anabaenopeptilide 90A were isolated (Gkelis *et al.*, 2005; Rouhiainen *et al.*, 2000; Fujii *et al.*, 1996). The *oci* biosynthetic gene cluster does not contain a phenolic halogenase gene, in contrast to the highly similar cyanopeptolin 984 synthetase operon (*mcn*) (Rounge *et al.*, 2007). It is supposed that NRPS gene clusters lacking halogenase genes arise from more ancient pathways by gene deletion and that chlorination is presumbly not necessary for the biological function of these cyclic peptides (Jones *et al.*, 2010; Tooming-Klunderud *et al.*, 2007).

In the biosynthesis of anabaenopeptilide 90B and cyanopeptolin 954 [figure 1.2-1], respectively, the flavin-dependent halogenases ApdC and McnD perform tyrosyl moiety chlorination during the NRPS assembly line (Wagner *et al.*, 2009). Other cyclic depsipeptides are the cryptophycins, which have been isolated exclusively from *Nostoc* spp. and not, as common for other peptide classes, from a broad taxonomic range of cyanobacteria (Welker & von Doehren, 2006). The antimitotic cryptophycin 1 [figure 1.2-2], which was first discovered in the lichen cyanobacterial symbiont *Nostoc* sp. ATCC 53789, represents the most important member of this

large class of cyanobacterial secondary metabolites. It possesses tubulin-depolymerising properties (Smith & Zhang, 1996; Bai *et al.*, 1996; Smith *et al.*, 1994), and thus served as a lead structure for efforts to design improved anticancer drugs (Liu *et al.*, 2009; Beck *et al.*, 2005). The biosynthetic gene cluster revealed the presence of a mixed PKS-NRPS-system (Du & Shen, 2001; Du *et al.*, 2001) producing the depsipeptide backbone. Similar to anabaenopeptilide 90B and cyanopeptolin 954, chlorination occurs at the tyrosine moiety, likely catalysed by the $FADH_2$-dependent halogenase CrpH. It is proposed that CrpH is also responsible for the production of dichlorinated cryptophycin congeners (Magarvey *et al.*, 2006). The closely related non-halogenated arenastatin A [figure 1.2-2] was isolated from the marine sponge *Dysidea arenaria* (Kobayashi & Kitagawa, 1999; Kobayashi *et al.*, 1995). Like cryptophycin 1, arenastatin A was used as a lead structure to design analogues aiming the development of new anticancer agents (Murakami *et al.*, 2004). The high structural similarity to cryptophycin 1 suggests that a cyanobacterial symbiont in the sponge may be the producing strain of this compound. This idea is supported by the fact that cryptophycin-24, which was isolated from *Nostoc* sp GSV 224 proved to be identical with arenastatin A (Moore, 1996).



cryptophycin 1



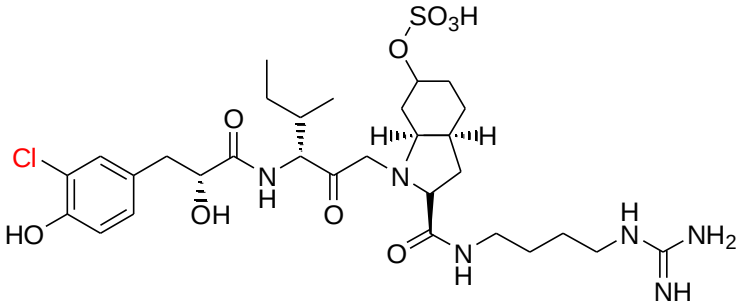
arenastatin A

Figure 1.2-2: Chemical structures of the cyanobacterial depsipeptide cryptophycin 1, which was first isolated from the lichen cyanobacterial symbiont *Nostoc* sp. ATCC 53789 and the structurally closely related arenastatin A, which was obtained from the marine sponge *Dysidea arenaria*.

Besides the prevalent production of halogenated cyclic peptides, cyanobacteria are also capable to biosynthesise linear peptides with halide substituents. Aeruginosins are a group of such linear tetrapeptides chlorinated either at the 4-hydroxyphenyllactic acid (4-Hpla) moiety, as found in *Microcystis* sp., or at the 2-carboxy-6-hydroxyoctahydroindole (Choi) residue, as described for such metabolites

from *Oscillatoria* sp. (Ishida *et al.*, 2009). Like cyanopeptolins, also aeruginosins revealed inhibition of serine proteases, especially thrombin (Ishida *et al.*, 1999).

As presumed for the chlorination step in anabaenopeptilide 90B [figure 1.2-1], aromatic moiety halogenations (van Pee *et al.*, 2006) during the biosynthesis of cyanopeptolins and aeruginosins are suggested to occur integrated in the NRPS biosynthetic pathway (Wagner *et al.*, 2009; Cadel-Six *et al.*, 2008; Tooming-Klunderud *et al.*, 2007). Figure 1.2-3 shows exemplarily the chemical structure of a chlorinated acyclic peptide, i.e. aeruginosin 98-A from *Microcystis aeruginosa* NIES-98.



aeruginosin 98-A

Figure 1.2-3: Chemical structure of the trypsin inhibitor aeruginosin 98-A from *Microcystis aeruginosa* NIES-98. It represents one congener of the large class of aeruginosins, which are linear, cyanobacterial tetrapeptides (Ishida *et al.*, 2009; Murakami *et al.*, 1995).

Several cyanobacterial metabolites that arise from mixed PKS/NRPS assembly lines include halogenation of alkyl moieties in their biosynthetic pathways (Jones *et al.*, 2009). These halogenations are accomplished by α-ketoglutarate-dependent non-heme Fe (II) halogenases. In the biosynthesis of the anticancer compound curacin A, produced by a Curaçao strain of *Lyngbya majuscula*, the halogenase CurA catalyses a cryptic chlorination to facilitate the formation of the cyclopropane ring (Khare *et al.*, 2010; Chang *et al.*, 2004).

Other chlorinated lipopetides are the neurotoxic jamaicamides (Edwards *et al.*, 2004) and the molluscicidal barbamide [figure 1.2-4] (Orjala & Gerwick, 1996; Jones *et al.*, 2010). The closely related barbaleucamides A and B were derived from a Philippine sponge of the genus *Lamellodysidea* (formerly *Dysidea)* (Harrigan *et al*., 2001).

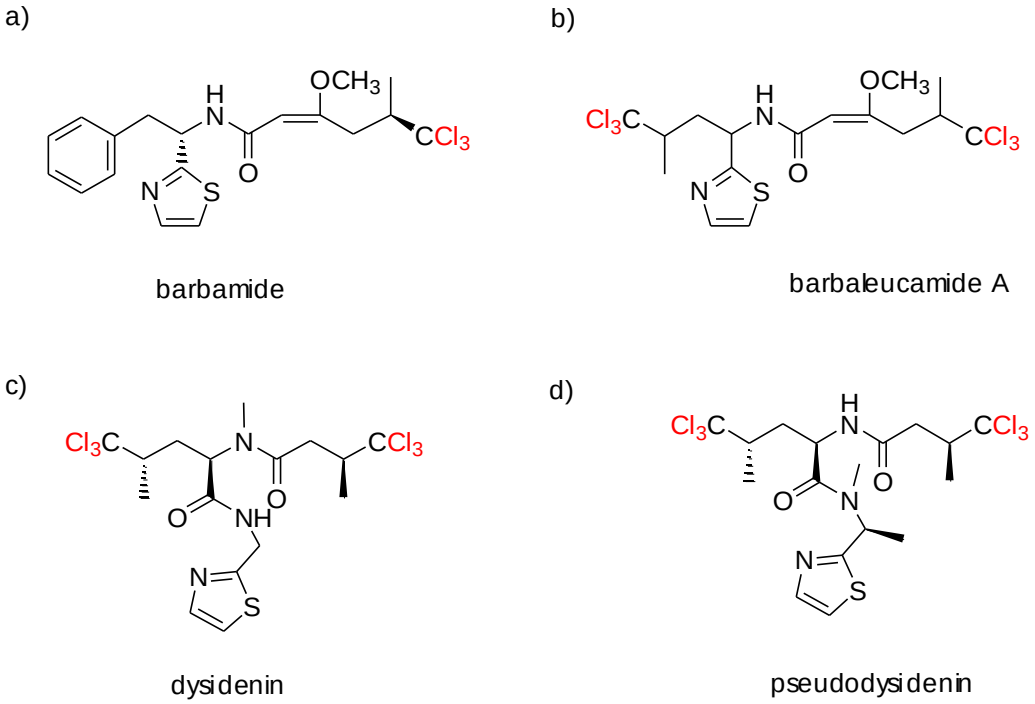


Figure 1.2-4: The molluscicidal barbamide (a) obtained from *Lyngbya majuscula* resembles the sponge-derived barbaleucamide A (b). The polychlorinated amino acid derivative dysidenin from the marine sponge *Lamellodysidea herbacea* (c) is structurally similar to the lipopeptide pseudodysidenin (d) from the marine cyanobacterium *L. majuscula* (Jones *et al.*, 2009; Jiménez & Scheuer, 2001). Halogenation of alkyl moieties of these compounds is carried out by α-ketoglutarate-dependent non-heme Fe (II) halogenases.

Dysidenin [figure 1.2-4], an inhibitor of iodide transport in thyroid cells has been isolated from the marine sponge *Lamellodysidea herbacea* (formerly known as *Dysidea herbacea*) (Van Sande *et al.*, 1990; Kazlauskas *et al.*, 1977). It carries a trichloromethyl residue and shows apparent structural similarity to pseudodysidenin, produced by the cyanobacterium *L. majuscula* (Jiménez & Scheuer, 2001). For the derivative 13-demethylisodysidenin, it was shown by fluorescence activated cell separation that only *O. spongeliae* contained the chlorinated metabolite, and therefore was likely the true producer of this compound (Unson *et al.*, 1994; Unson & Faulkner, 1993).

In a recent study presented by Flatt (2005), a genetic approach was applied to confirm the filamentous cyanobacterial symbiont *O. spongeliae* to be the source of polychlorinated peptides from *Lamellodysidea* (formerly *Dysidea*) herbacea. Sequence information available for the barbamide biosynthetic gene cluster (Chang

et al., 2002; Orjala & Gerwick, 1996) enabled to amplify a homologous gene (*dysB1*) from *Dysidea* sponge material. Using the CARD-FISH method, it was shown that *dysB1* probes did only hybridise to sequences in cyanobacterial cells (Flatt *et al.*, 2005).

1.3 Chlorinated secondary metabolites in *Fischerella ambigua*

The current project focused on *F. ambigua* and the biosynthesis of its chlorinated natural products. The latter did not belong to the group of nonribosomal peptides but were polychlorinated phenolic ethers and a bisindol alkaloid. These compounds, named ambigols A-C and tjipanazole D were found in a previous study (Falch *et al.*, 1995).

1.3.1 Ambigols and related structures

Ambigols represent a highly intriguing class of cyanobacterial compounds formed by a simple phenolic building block that is linked by either ether-bridges or aryl-aryl-bonds to give a trimeric basic structure. This skeleton is furthermore characterised by the presence of not less than six chloride atoms on the phenyl rings [figure 1.3.1-1].

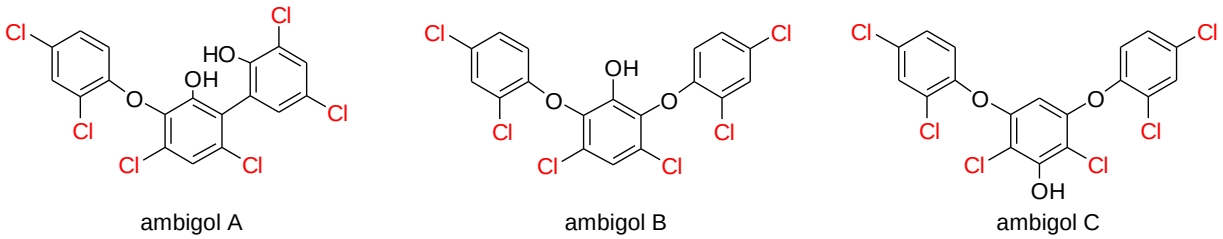


Figure 1.3.1-1: Polychlorinated phenolic ethers, ambigols A-C, produced by the terrestrial cyanobacterium *Fischerella ambigua*.

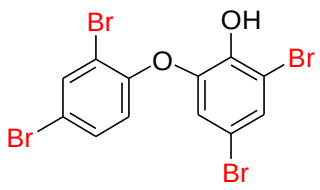
Polyhalogenated phenolic ethers feature a broad range of biological activities (Gribble, 2010). The polychlorinated ambigols A and B were found to exhibit antibacterial effects, whereby ambigol A was significantly more active than ambigol B. In a bioautographic assay, ambigol A was tested against *Bacillus subtilis* inhibiting its growth at the same order of magnitude as chloramphenicol. Furthermore, both, ambigol A and B, have been proven to show mild cytotoxic and antiviral properties as

well as a strong inhibitory effect on cyclooxygenase (Falch *et al.*, 1995). A significant cytotoxic effect was described for ambigol A, whereas ambigol B only revealed a weak lethal effect in a brine shrimp assay. Ambigol C possesses antiplasmodial activity, which could not been proven for ambigol A and B though (Wright *et al.*, 2005).

1.3.1.1 Related structures to the ambigols from the marine environment

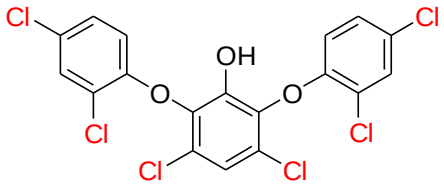
From the tropical marine sponge *L. herbacea* two chemotypes are known (Faulkner *et al.*, 1994), one of which produces polychlorinated amino acid derivatives (Unson *et al.*, 1993) like dysidenin and barbaleucamides A and B [figure 1.2-4], whereas the second exclusively produces polybrominated diphenyl ethers [figure 1.3.1.1-1a] (Handayani *et al.*, 1997).

a)



2-(2',4'-dibromophenoxy)-4,6-dibromophenol

b)



ambigol B

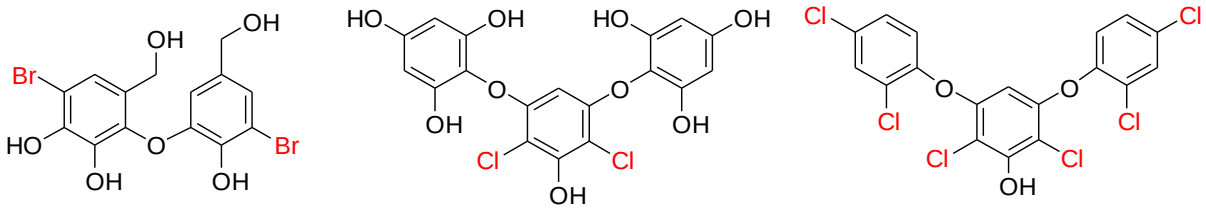
Figure 1.3.1.1-1: Polybrominated diphenyl ether (a) isolated from the marine sponge *L. herbacea* (formerly known as *D. herbacea*). It is structurally related to the polychlorinated phenolic ethers from *F. ambigua*, e.g. ambigol B (b).

The most prevalent prokaryotic endosymbiont in the mesohyl of the sponge *L. herbacea* is the host-specific filamentous cyanobacterium *O. spongeliae* (Thacker & Starnes, 2003), which was identified as the possibly responsible strain for the production of brominated diphenyl ethers (Unson *et al.*, 1994). A great variety of new polybrominated diphenyl ethers have been frequently encountered in *Lamellodysidea* (formerly *Dysidea)* spp. exhibiting a broad range of biological effects in bioassays. These include antibacterial, antifungal, and cytotoxic activity (Gribble, 2010). Recently, 2-(2',4'-dibromophenoxy)-4,6-dibromophenol [figure 1.3.1.1-1 a] previously isolated from *L. herbacea* (Carté & Faulkner, 1981) was obtained from a new source, the sponge *D. granulosa* and was found to display potent broad-spectrum in vitro

activity towards Gram-positive bacteria. Further, a potent activity against antibiotic-resistant bacteria including MRSA, erythromycin-resistant *S. aureus* and vancomycin-resistant *Enterococcus* has been reported (Shridhar *et al.*, 2009).

1.3.1.2 Related structures to the ambigols found in red and brown algae

Red algae like *Odonthalia corymbifera* belong to the family Rhodomelaceae and represent a rich source of bromophenols, which possess cytotoxic, anti-inflammatory and antimicrobial effects (Oh *et al.*, 2008). Studies on the structure-activity relationship of antimicrobial bromophenols from natural sources and of synthetic derivatives revealed that a di-phenolic backbone, the presence of free hydroxyl groups and two or more bromine substituents on the phenol ring are critical for antimicrobial attributes (Oh *et al.*, 2009; Oh *et al.*, 2008).



1 **2** **3**

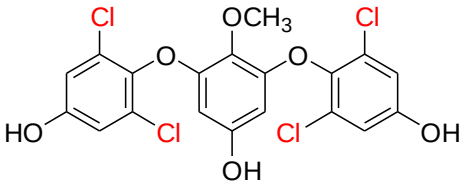
Figure 1.3.1.2-1: Phenolic metabolites from red and brown algae (**1** and **2**) and the structurally similar ambigol C (**3**) isolated from *F. ambigua*. Compound **1** belongs to the lanosol-type dimers obtained from the red alga *Odonthalia corymbifera*. Compound **2** represents a dichlorinated phloroglucinol derivative isolated from the brown alga *Laminaria ochroleuca*.

The lanosol-type dimer **1** [figure 1.3.1.2-1], also obtained from *Odonthalia corymbifera*, inactivates α-glucosidase, and thus could be a potential candidate for adjuvant Diabetes mellitus treatment (Kurihara *et al.*, 1999b; Kurihara *et al.*, 1999a). The phlorotannin triphlorethol was obtained along with chlorinated analogues, exemplified by compound **2** from the brown alga *Laminaria ochroleuca* (La Barre *et al.*, 2010; Glombitza *et al.*, 1977). They bear remarkable structural similarity to the ambigols from *F. ambigua* [figure 1.3.1.2-1 and 1.3.1.1-1].

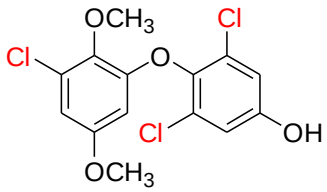
Phloroglucinol natural compounds including phlorotannins with ether bridges and phenyl linkages have been reviewed recently (Singh *et al.*, 2010; Gribble, 2010). They exhibit a broad range of biological activities such as anti-inflammatory, anticancer, neuro-regenerative and antioxidant effects (Singh *et al.*, 2009).

1.3.1.3 Ambigol-related structures from the terrestrial environment

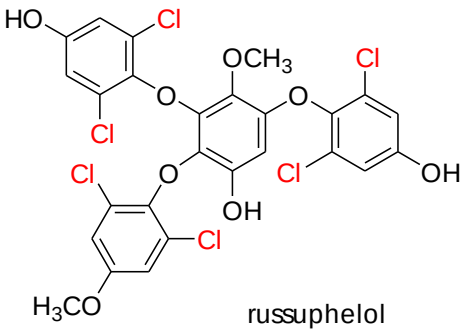
Terrestrial organisms usually utilise chlorine instead of bromine due to its prevalence in soil. The toxic mushroom *Russula subnigrans* (Russulaceae) produces interesting polychlorinated phenyl ethers, i.e. russuphelins A-F, which are reminiscent of the chlorinated phenyl ether structure of the ambigols from *F. ambigua.* Russuphelin B-D [figure 1.3.1.3-1] show cytotoxicity against P388 leukemia cells (Takahashi *et al.*, 1992). In the course of further studies on this mushroom, the optically active russuphelol was identified (Ohta *et al.*, 1995).



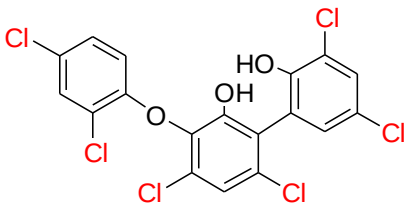
russuphelin B



russuphelin F



russuphelol

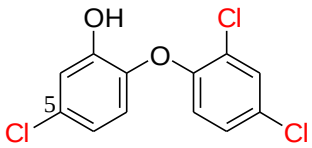


ambigol A

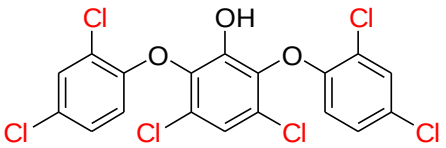
Figure 1.3.1.3-1: A series of polychlorinated phenyl ethers, exemplified by russuphelin B and F, were isolated from the toxic mushroom *Russula subnigrans*. It also produces the tetrameric ether russuphelol. All compounds structurally resemble the ambigols from *F. ambigua*.

1.3.1.4 Structural similarity of ambigols to triclosan

Triclosan [figure 1.3.1.4-1] is a synthetic bisphenol with a broad-spectrum activity against Gram-positive and Gram-negative bacteria. It is topically used for disinfecting ancillary treatment of neurodermitis, acne and ulcus cruris. Its antimicrobial effect is due to its inhibitory activity against bacterial fatty acid synthase (FAS) (Heath *et al.*, 2001). Recently, a suppressive effect of triclosan on rat mammary carcinogenesis has been discovered. Again, this effect is related to inhibition of FAS suggesting the latter as a promising molecular target for breast cancer chemoprevention (Lu & Archer, 2005). By analogy, the uncompetitive inhibition of type II fatty acid biosynthesis explains the potency of triclosan against in vitro cultured *Plasmodium falciparum* parasites as well as against *P. berghei* in vivo infections in mice (Freundlich *et al.*, 2007; Perozzo *et al.*, 2002).



triclosan



ambigol B

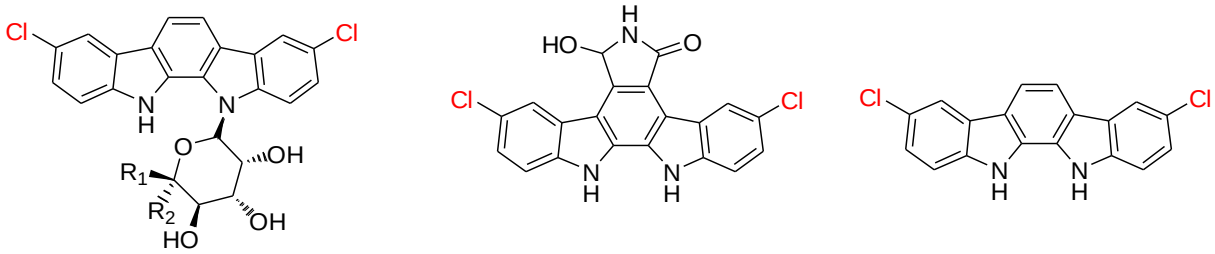
Figure 1.3.1.4-1: Chemical structure of the synthetic, antimicrobial bisphenol ether triclosan and the natural product ambigol A from *F. ambigua*.

A structure-based drug design approach led to the identification of triclosan derivatives containing alkyl and aryl substituents in 5-position. These compounds possessed significantly enhanced activity against purified InhA, an enoyl acyl carrier protein reductase of *Mycobacterium tuberculosis* (Freundlich *et al.*, 2009). The broad use of triclosan in many daily applied products, e.g. toothpaste, cosmetics, detergents increases the development of triclosan-resistant bacteria.

1.3.2 Tjipanazole D and related structures

The bisindole natural product tjipanazole D was also isolated from the cyanobacterium *F. ambigua* (Falch *et al.*, 1995) and is the second class of compounds targeted in this study. It was previously reported from the cyanobacterial

species *Tolypothrix tjipanasensis* together with 14 further tjipanazoles (Bonjouklian *et al.*, 1991). Of these compounds, only tjipanazole J possesses the pyrrolo[3,4-*c*] ring, which is a typical chemical attribute of indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazoles from actinomycetes, e.g. rebeccamycin from cultures of *Lechevalieria aerocolonigenes* (Bush *et al.*, 1987) and staurosporine from *Streptomyces* sp. TP-A0274 (Tamaoki *et al.*, 1986). Other tjipanazoles obtained from *T. tjipanasensis* were chlorinated and non-chlorinated indolo[2,3-*a*]carbazole *N*-glycosides or simple chlorinated indolo[2,3-*a*]carbazoles, such as tjipanazole D. All tjipanazoles possess only weak cytotoxicity against tumor cell lines and are inactive towards protein kinase C. Tjipanazoles A1 and A2 [figure 1.3.2-1] are described as antifungal agents (Bonjouklian *et al.*, 1991). Tjipanazole D, when isolated from *F. ambigua* showed only a moderate antibacterial activity against the Gram-positive strains of *Bacillus subtilis* and *Micrococcus luteus* and displayed no antifungal activity (Falch *et al.*, 1995).



tjipanazole A_1 $R_1 = CH_3$, $R_2 = H$
tjipanazole A_2 $R_1 = H$, $R_2 = CH_3$

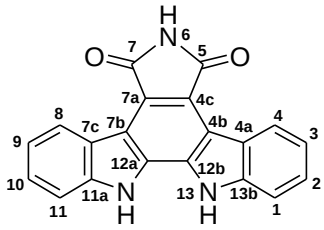
tjipanazole J

tjipanazole D

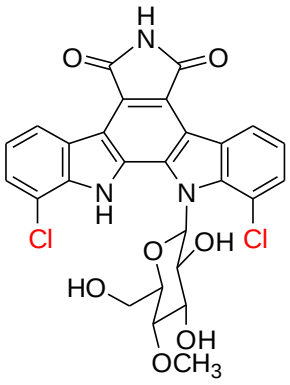
Figure 1.3.2-1: Chlorinated tjipanazoles obtained from the cyanobacterium *Tolypothrix tjipanasensis*. They were isolated among other tjipanazoles, which except tjipanazole J all have the same indolocarbazole framework and vary in their tailoring modifications, i.e. glycosylation and halogenation. Tjipanazole D was also isolated from the terrestrial cyanobacterium *F. ambigua*.

A great number of structurally diverse bisindole secondary metabolites have been isolated from a variety of microbial species (Ryan & Drennan, 2009). Structurally related indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazoles are divided into two classes, i.e. DNA-topoisomerase I inhibitors and protein kinase C inhibitors (Hyun *et al.*, 2003). The DNA-topoisomerase I inhibitor rebeccamycin and the structurally related AT2433-A1 are chlorinated indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazole derivatives that bear a fully

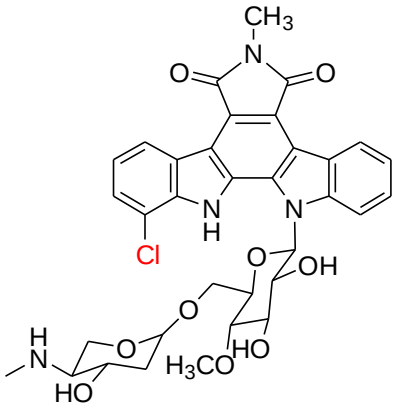
oxidised C-7 carbon [figure 1.3.2-2]. Except *N*-methylation of the pyrrole ring in AT2433-A1 and an additional chlorination in rebeccamycin, both compounds have the same aglycone. The single methoxy-β-glucoside residue in rebeccamycin was shown to be essential for interaction with DNA topoisomerase (Bailly *et al.*, 1999). The rebeccamycin structure was a target for semisynthetic design of a great number of analogues (Marminon *et al.*, 2008; Moreau *et al.*, 2005; Anizon *et al.*, 2003). Various substituents can be introduced to the lactam nitrogen [figure 1.3.2-2] without loss of the topoisomerase I inhibition activity (Prudhomme, 2003). A semisynthetic approach even led to analogues of the tjipanazoles, which lack an imide heterocycle but despite that exhibit strong inhibition of DNA topoisomerase I. For instance, a derivative containing the sugar moiety of rebeccamycin and two nitro substituents at the indolo[2,3-*a*]carbazole scaffold revealed the same profile of cytotoxicity as rebeccamycin (Voldoire *et al.*, 2004).



Indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazole-scaffold



rebeccamycin



AT2433-A1

Figure 1.3.2-2: Antitumor compounds rebeccamycin and AT2433-A1 represent a class of bisindoles with a fully oxidised imide heterocycle and a sugar moiety attached only to one indole nitrogen. The *N*-attached sugar is the same in both, rebeccamycin and AT2433-A1. However, in the indolocarbazole diglycoside AT2433-A1 a second pyranosyl-unit, i.e. 2,4-dideoxy-4-methylamino-L-xylose is attached to the *O*-methylated glucosyl moiety.

A second group of indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazoles is exemplified by the protein kinase C inhibitors staurosporine and K252a, which contain an identical aglycone with a non-oxidised C-7 carbon and a sugar that is linked to both indole nitrogens [figure 1.3.2-3]. The crystal structures of staurosporine and UCN-01 in complex with checkpoint kinase (Chk) 1 revealed that a free NH of the lactam heterocycle interacts with the ATP-binding pocket of Chk1 by hydrogen-bonding, and thus substitution at the imide nitrogen turned out to decrease Chk1 inhibitory potency of derivatives (Zhao *et al.*, 2002). A semisynthetically produced rebeccamycin derivative having a sugar moiety linked to both indole-nitrogens like in staurosporine exhibited strong Chk1 inhibition activity (Anizon *et al.*, 2009).

staurosporine K252a

Figure 1.3.2-3: Antitumor compounds staurosporine and K252a with an indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazole-skeleton, in which only one carbon, C5 is oxidised and the sugar is linked to both indole nitrogens. Staurosporine contains the sugar residue 2,3,6-trideoxy-3-aminoaldohexose, whereas K252a is substituted with a dihydrostreptose moiety.

A large number of semisynthetic rebeccamycin and staurosporine analogues have been developed to investigate structure-activity relationships and to generate synthetic drugs, which combine structurally essential elements of rebeccamycin- and staurosporine-like compounds, respectively, to display a dual effect on cancer cells (Prudhomme, 2004; Marminon *et al.*, 2008).

The biosynthetic gene clusters of staurosporine and rebeccamycin as well as those of the related indolocarbazoles K252a and At2433-A1 have been cloned and characterised (Onaka *et al.*, 2003; Onaka *et al.*, 2002; Sánchez *et al.*, 2002; Chiu *et al.*, 2009b; Gao *et al.*, 2006). Genes involved in the production of these compounds encode for highly conserved proteins, which generate the indolocarbazole core. They

even turned out to be exchangeable in combinatorial biosynthesis experiments (Salas & Mendez, 2009; Sánchez *et al.*, 2006a; Sánchez *et al.*, 2005). Tjipanazole D, bears an indolo[2,3-*a*]carbazole core structure instead of an indolo[2,3-*a*]pyrrolo[3,4-*c*] carbazole as found for the above mentioned natural products. This divergence consequently implies differences in the biosynthetic assembly of tjipanazole D. At this point of time, no published data exist on the biosynthesis of the indolocarbazole scaffold as present in tjipanazole D.

1.4 Proposed building blocks and reactions in ambigol and tjipanazole D biosynthesis

The ambigol core structure and its regioselective halogenation pattern as well as obtained sequence information on probable biosynthesis proteins encouraged to hypothesise on the involvement of specific enzymes in the endogenous assembly of these highly intriguing phenolic ethers [figure 1.1-1]. These enzymes are assumed to be phenolic halogenases (1.4.1.1), a chorismate lyase (1.4.2), a CoA synthetase (1.4.3), and a pair of two CYP 450 enzymes (1.4.4). The phenolic building blocks, which form the ambigol carbon scaffold is supposed to arise from chorismic acid via formation of 4-hydroxybenzoic acid (4-HBA). Phenolic halogenases usually act on peptidyl carrier-attached substrates (Buedenbender *et al.*, 2009; van Pée & Patallo, 2006). Based on sequence data obtained in this study, this PCP domain is part of an NRPS-like module carrying a starter C domain (1.4.5). To date, no published data exist on the biosynthesis of this type of natural products.

The indolocarbazole tjipanazole D [figure 1.1-1] is composed of two indole units, which are chlorinated in 5-position. As described in 1.3.2, the molecule shares remarkable structural similarity with other indolocarbazole compounds (Nakano & Omura, 2009), whose biosyntheses are well-characterised in literature (Ryan & Drennan, 2009), e.g. rebeccamycin [figure 1.3.2-2] from *Lechevalieria aerocolonigenes* (1.3.2). On the basis of published information on the biosynthesis of these natural products and their high structural resemblance to tjipanazole D, the latter most likely arises from the simple building block L-tryptophan, which is known to be a typical free diffusible substrate for tryptophan halogenases (1.4.1.2) (Zhu *et al.*, 2009; Sánchez *et al.*, 2006b). By analogy to rebeccamycin production (Onaka, 2009; Sánchez *et al.*, 2005; Onaka *et al.*, 2003), the biosynthesis of the tjipanazole D

carbon framework is proposed to include a dimerisation of 5-chlorinated L-tryptophan and an intramolecular aryl-aryl coupling between the pyrrole rings of the two indole units. Thereby, oxidative processes are suggested to be catalysed by three enzymes, a chromopyrrolic acid synthase (CPAS)-like protein, an FAD-dependent monooxygenase and a CYP 450 enzyme (1.4.4; 5.7.5).

Both, ambigols and tjipanazole D are characterised by aromatic halogenation, and thus it was hypothesised that the biosynthesis of these secondary metabolites involves $FADH_2$-dependent halogenases (1.4.1). Both, phenolic and tryptophan halogenases, are thoroughly described and characterised in literature (Neumann *et al.*, 2008; Flecks *et al.*, 2008; van Pée & Patallo, 2006). They are typically involved in the halogenation of aromatic moieties, in particular aromatic amino acids, i.e. chorismic acid derivatives (Wagner *et al.*, 2009). In case of ambigols and tjipanazole D, results of the current study suggest chorismate to be the origin of their aromatic scaffolds (5.5).

1.4.1 Flavin-dependent halogenases

$FADH_2$-dependent halogenases form a class of biosynthetic enzymes that typically halogenate aromatic substrates with strict regioselectivity, but also a few aliphatic substrates are known. The involvement of halogenases in the biosynthetic process has been thoroughly reviewed recently (Wagner *et al.*, 2009; Neumann *et al.*, 2008; Fujimori & Walsh, 2007). In regard to their substrate specifity, they may be divided into three subgroups [figure 1.4.1-1]. One of them mediates the halogenation of phenolic or pyrrole moieties, while a second one is responsible for the halogenation of tryptophan derivatives, and a third one acts on aliphatic substrates (van Pée & Patallo, 2006). However, concerning their substrate utilisation, i.e. carrier-bound or as free substrates, a classification into two groups is supported by crystal structures obtained from several $FADH_2$-dependent halogenases (Buedenbender *et al.*, 2009; Neumann *et al.*, 2008; Dong *et al.*, 2004): variant A acts on small substrates while completely embracing them. This group includes tryptophan halogenases and the aliphatic halogenase CmlS from chloramphenicol biosynthesis. Halogenating enzymes of variant B catalyse halogenations while the substrate is attached to a carrier protein (Podzelinska *et al.*, 2010).

Formerly, phenolic halogenases were supposed to act exclusively on carrier-bound substrates (van Pée & Patallo, 2006). However, very recently the halogenation steps in the biosynthesis of differentiation-inducing factor 1 (Dif-1), a polyketide-derived, dichlorinated signal molecule from *Dictyostelium discoideum* (Morris *et al.*, 1987) has been studied in vivo and in vitro. Although ChlA, a flavin-dependent halogenase identified in the genome of *D. discoideum*, clearly acts on a phenolic substrate, no carrier protein is required for halogenation (Neumann *et al.*, 2010; Austin *et al.*, 2006). It should be noted that this halogenase was proven to catalyse both chlorination steps in the biosynthesis of DIF-1.

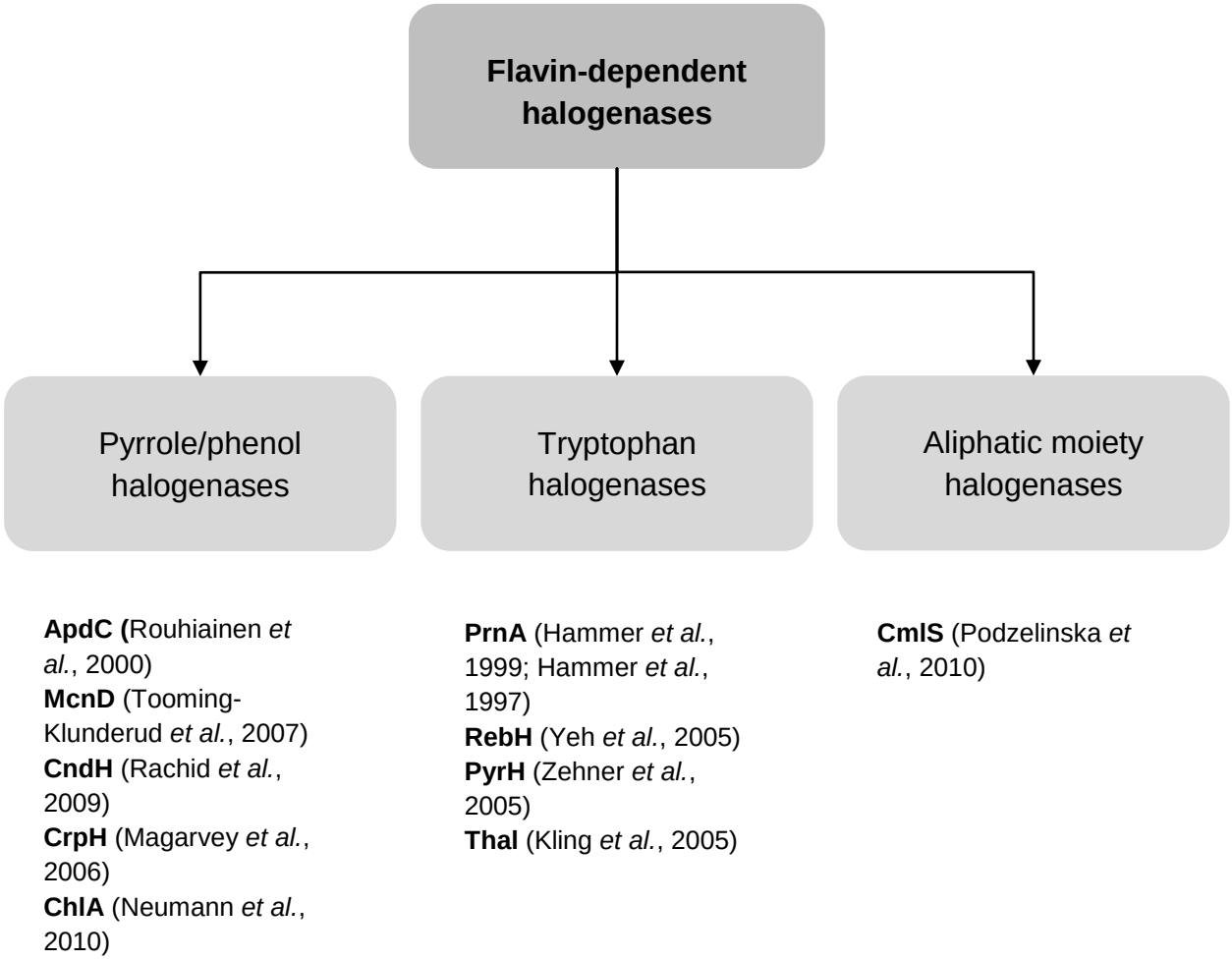
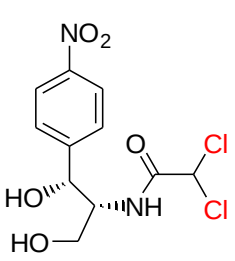
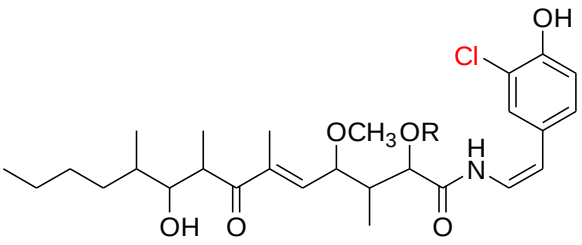


Figure 1.4.1-1: Chart illustrating different types of flavin-dependent halogenases with regard to their substrate specifity. Pyrrole and phenol halogenating enzymes often act on peptidyl carrier-tethered substrates. Similarly, halogenation of an aliphatic moiety, as in case of chloramphenicol biosynthesis, is supposed to be carried out on a CoA-bound intermediate (acetoacetyl-CoA). By contrast, tryptophan is usually utilised as a free diffusible substrate.

The crystal structure of the phenolic halogenase CndH, which is involved in the biosynthesis of chondrochloren in the myxobacterium *Chondromyces crocatus* [figure 1.4.1-2] was compared to previously published crystal structures of the tryptophan halogenases PrnA (Dong *et al.*, 2005) and RebH (Yeh *et al.*, 2007). This suggested a protein interaction between the phenolic halogenase CndH and the carrier protein during catalytic turnover (Buedenbender *et al.*, 2009). The enzymatic mechanism of $FADH_2$-dependent halogenases has been proposed to depend on three absolutely conserved regions, which mediate the formation of an HOCl molecule and its controlled nucleophilic attack on the aromatic substrate (Butler & Sandy, 2009; Buedenbender *et al.*, 2009; Yeh *et al.*, 2007; Anderson and Chapman, 2006). The highly conserved overall organisation of $FADH_2$-dependent halogenases has been recently confirmed for the crystal structure of CmlS, a flavin-dependent halogenase that is responsible for aliphatic moiety chlorination during chloramphenicol biosynthesis in *Streptomyces venezuelae* ISP5230 [figure 1.4.1-2] (Podzelinska *et al.*, 2010).



chloramphenicol



chondrochloren A/B (R= CH_3/ C_2H_6)

Figure 1.4.1-2: The two $FADH_2$-dependent halogenases CmlS from chloramphenicol biosynthesis and CndH from chondrochloren production perform chlorination on different substrates. Despite that, the crystal structures of these two enzymes reveal a highly conserved overall structure suggesting that all $FADH_2$-dependent halogenases use a conserved calaytic mechanism but differ in their substrate recognition and binding.

Halogenases of variant B, like the chondrochloren halogenase CndH, show a more open conformation of their active site, being accessible for a carrier-bound intermediate by sterical adaption including protein-protein interactions with a carrier domain (Podzelinska *et al.*, 2010). This could also explain the ability of some

phenolic halogenases to perform multiple halogenations with regioselectivity, which may be due to their higher conformational flexibility in their substrate recognition.

All aromatic moiety halogenations performed by $FADH_2$-dependent halogenases reveal excellent regiospecifity (Wagner *et al.*, 2009; van Pée & Patallo, 2006). Enzymes catalysing the halogenation of pyrrole and phenol moieties may be capable of regioselective dichlorinations [figure 1.4.1.1-1], whereas multiple halogenations of tryptophan derivatives require distinct halogenating enzymes for each halogenation step [figure 1.4.1.2-2]. Halogenase genes, which are closely associated to particular NRPS gene clusters are according to the current state of knowledge likely to encode for enzymes elaborating carrier-tethered intermediates (1.4.1.1). By contrast, tryptophan halogenases utilise their substrates without the need of carrier units (1.4.1.2).

1.4.1.1 Carrier proteins and halogenases in the biosynthetic process

Halogentions are typical tailoring modifications of cyanobacterial metabolites and very often associated with NRPS, PKS or mixed NRPS/PKS assembly lines (Walsh, 2008; Walsh, 2004). Peptidyl carrier proteins (PCP) and acyl/aryl carrier proteins (ACP/ArCP) are common attachment sites to present the substrate to a tailoring enzyme (Walsh, 2008; van Pée & Patallo, 2006; Walsh, 2004; Thomas *et al.*, 2002; Walsh *et al.*, 2001). A carrier-halogenase collaboration has been described for the biosynthesis of numerous cyanobacterial metabolites (Wagner *et al.*, 2009). For the chlorinated cyclic peptide anabaeopeptilide 90B [figure 1.2-1], it is proposed that the phenolic halogenase ApdC acts on a tyrosine residue while the growing peptide is bound to the NRPS protein ApdB (Rouhiainen *et al.*, 2000). Similar considerations were suggested for chlorinated cyanopeptolin and aeruginosin peptides (Ishida *et al.*, 2009; Cadel-Six *et al.*, 2008; Tooming-Klunderud *et al.*, 2007). Carrier proteins occur either freestanding or embedded in multifunctional proteins. They exist as three variants (Zhou *et al.*, 2007): acyl carrier proteins (ACP) found in polyketide synthetases (PKS) as well as fatty acid synthetases (FAS), peptidyl carrier proteins (PCP) found in NRPS systems, and aryl carrier proteins (ArCP) usually found in siderophore NRPS synthetases (Qiao *et al.*, 2007).

Not all carrier-depending halogenases do function in a multienzymatic biosynthetic environment. The hybrid PKS/NRPS molecule pyoluteorin from *Pseudomonas*

fluorescens Pf-5 consists of an NRPS-derived pyrrole moiety and a PKS-derived resorcinol ring. In the biosynthesis of this antifungal compound [figure 1.4.1.1-1], a simple freestanding carrier protein PltL, which contains a phosphopantetheinyl attachment site, is involved (Hutchinson, 2003). A prolyl-AMP ligase (1.4.3) PltF activates the substrate proline by adenylation to prolyl-AMP for transfer to the thiolation (PCP) domain PltL. A following FAD-dependent four-electron oxidation that is likely performed by PltE yields pyrrolyl-S-PltL, which represents a carrier protein-attached intermediate as the substrate for dichlorination (Walsh *et al.*, 2006; Thomas *et al.*, 2002). Interestingly, the $FADH_2$-dependent halogenase PltA, performs both chlorination steps on the protein-attached acyl moiety, pyrrolyl-S-PltL. This reaction has been proven experimentally by in vitro assay (Walsh *et al.*, 2006; Dorrestein *et al.*, 2005; Nowak-Thompson *et al.*, 1999).

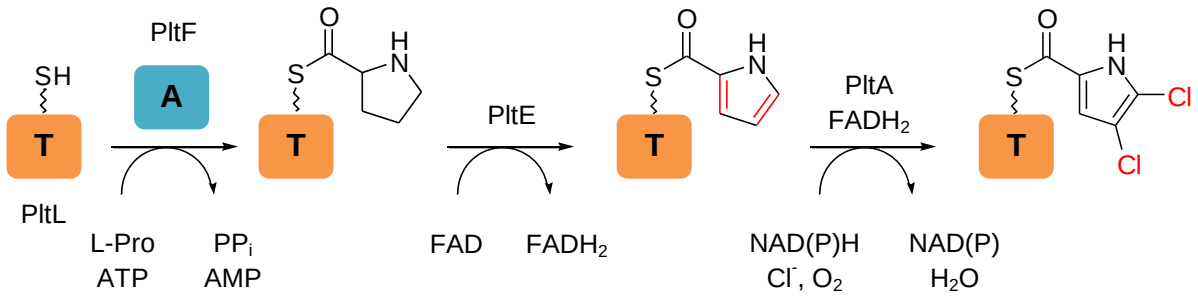


Figure 1.4.1.1-1: Formation of dichloropyrrole in the biosynthesis of pyoluteorin through the flavin-dependent halogenase PltA, which catalyses the dichlorination of the carrier-bound pyrrole moiety. T = thiolation domain; A = adenylation domain; PltF = prolyl-AMP ligase; PltE = FAD-dependent dehydrogenase

In complex NRPS- and PKS-systems, a modular organisation of the involved enzymes is favoured due to facilitated substrate channeling of large peptide or polyketide chains (Huang *et al.*, 2001). By contrast, the example of pyoluteorin biosynthesis demonstrates that intermolecular transfer of small substrates from freestanding adenylation (A) domains to separate carrier proteins can occur. Similarly, in the biosynthesis of enediyne C-1027 [figure 1.4.1.1-2 and 1.4.1.1-3], a PCP-tethered substrate undergoes regiospecific chlorination. Enediyne C-1027, an antitumor antibiotic was isolated from *Streptomyces globisporus* (Hu *et al.*, 1988) and is composed of an enediyne core, which is covalently bound to three distinct

moieties: a deoxy aminosugar, a β-amino acid, and a benzoxazolinate residue. For the biosynthesis of the (*S*)-3-chloro-4,5-dihydroxy-β-phenylalanine moiety, β-tyrosine is activated by an isolated A domain giving β-tyrosyl-AMP, which is subsequently loaded onto the freestanding PCP domain SgcC2 [figure 1.4.1.1-2]. The resulting 3-hydroxy-β-tyrosyl-S-SgcC2 intermediate is the substrate for halogenation catalysed by the $FADH_2$-dependent halogenase SgcC3. This halogenation step has been proven by generating the β-aminoacyl-S-SgcC2 substrate in vitro and subsequent incubation with the recombinant halogenase SgcC3 (Lin *et al.*, 2007). Again, the A domain encoded by *sgcC1* and the PCP domain SgcC2 are monofunctional proteins, i.e. not part of a modular architecture (Hutchinson, 2003). The A domains of both, enediyne biosynthesis (SgcC1) and pyoluteorin assembly (PltF) are responsible for aminoacyl-AMP formation and loading of this adenylate onto a carrier protein.

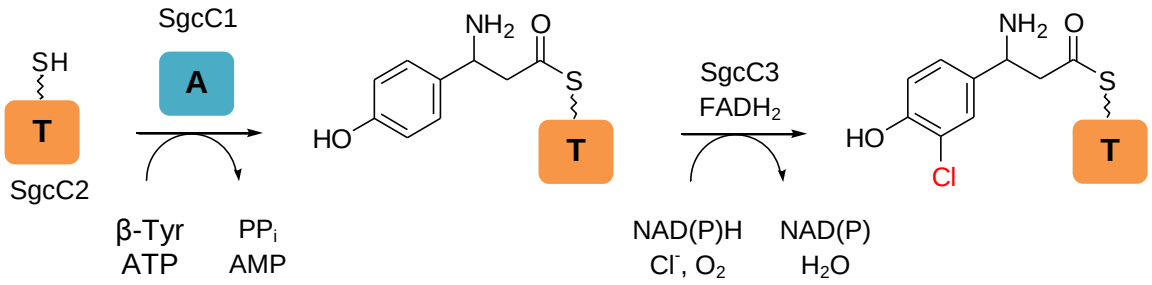


Figure 1.4.1.1-2: Formation of the (*S*)-3-chloro-4,5-dihydroxy-β-phenylalanine moiety catalysed by the flavin-dependent halogenase SgcC3, while the substrate is bound to the carrier protein SgcC2 in the biosynthesis of the antitumor antibiotic enediyne C-1027. T = thiolation domain; A = adenylation domain.

Also coenzyme A is a possible carrier for substrates to be halogenated. For instance, during the biosynthesis of chloramphenicol [figure 1.4.1.1-3] the acetoacetyl group is supposed to be linked to CoA by CmlK, which shares sequence homology to acyl CoA synthetases. The CoA-ester is suggested as a possible substrate for the $FADH_2$-dependent halogenase CmlS (Podzelinska *et al.*, 2010).

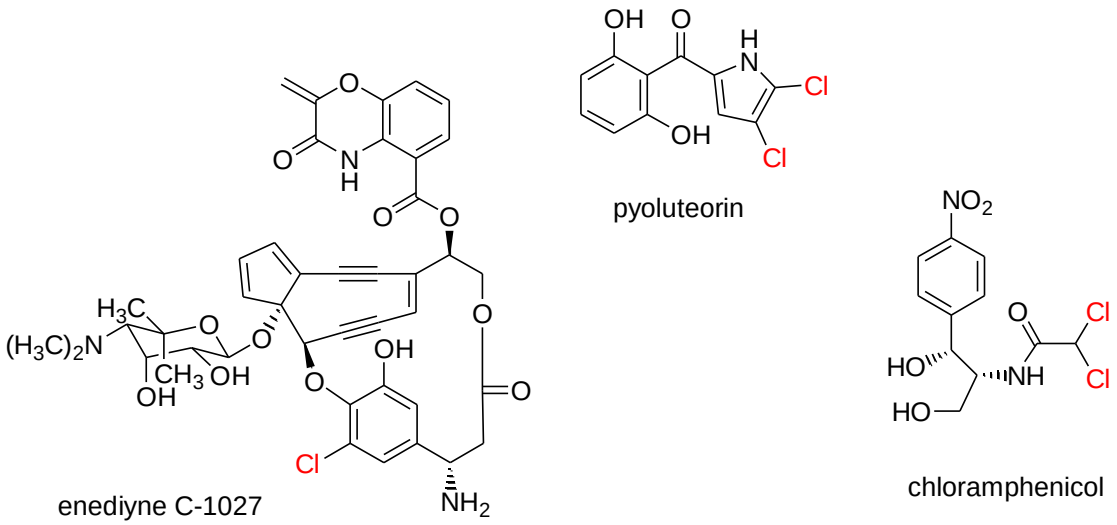


Figure 1.4.1.1-3: Chemical structures of natural compounds containing a chlorinated aromatic or aliphatic residue, respectively, which is biosynthesised by carrier-bound halogention.

1.4.1.2 Halogenases acting on free diffusible substrates

The most famous group of halogenases, which utilise free diffusible substrates is formed by enzymes responsible for the halogenation of tryptophan [figure 1.4.1.2-1] (Wagner *et al.*, 2009). They were first discovered during investigations towards the biosynthesis of pyrrolnitrin. It could be demonstrated that chlorination of tryptophan at position 7, mediated by PrnA, is the first step in pyrrolnitrin biosynthesis (Keller *et al.*, 2000; Hammer *et al.*, 1999; Kirner *et al.*, 1998; Hammer *et al.*, 1997). By analogy, the conversion of free tryptophan to 7-chlorotryptophan was shown to be catalysed by RebH as the initial step in the biosynthesis of the indolocarbazole rebeccamycin in *L. aerocolonigenes* (Yeh *et al.*, 2005). The biosynthesis of pyrrole containing compounds has been recently reviewed (Walsh *et al.*, 2006). To date, several other tryptophan halogenases have been characterised, some of which are a tryptophan 5-halogenase (PyrH) from *Streptomyces rugosporus* (Zehner *et al.*, 2005; Ding *et al.*, 1994) and a tryptophan 6-halogenase (Thal) from *Streptomyces albogriseolus* (Kling *et al.*, 2005). The mechanism of tryptophan halogenases has been studied extensively (Flecks *et al.*, 2008; Yeh *et al.*, 2007; Dong *et al.*, 2005). Recently performed mutagenesis experiments concerning the tryptophan binding site of the tryptophan-5-halogenase PyrH revealed that the latter binds tryptophan to present the C5 atom to the chlorinating species while blocking other potential reactive sites (Zhu *et al.*, 2009; Flecks *et al*., 2008).

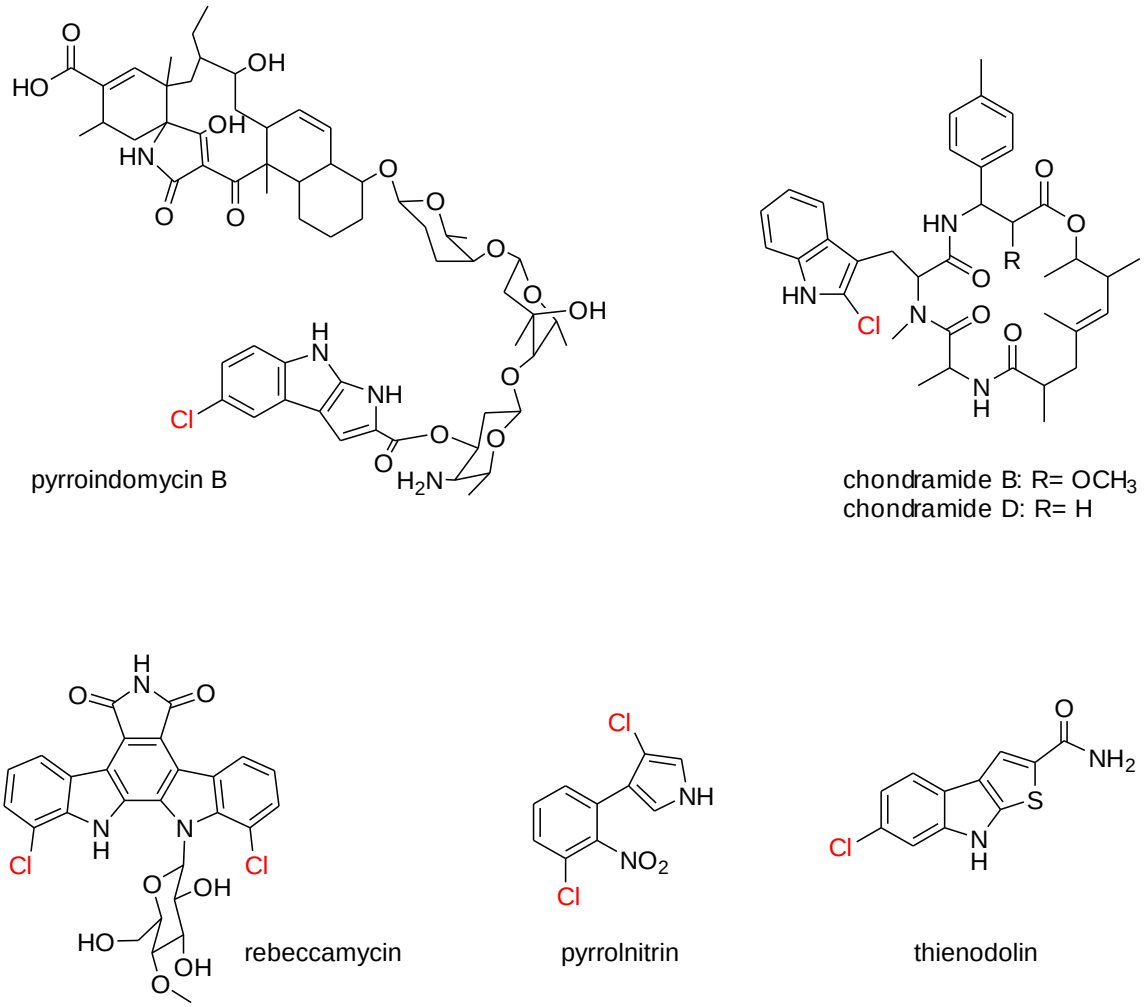


Figure 1.4.1.2-1: Bacterial secondary metabolites bearing a chlorinated tryptophan unit. All responsible halogenases of the respective biosynthetic gene clusters utilise free tryptophan, except in chondramides biosynthesis (see text below). Tryptophan halogenases act with distinct regioselectivity. Whereas in the biosynthesis of pyrrolnitrin and rebeccamycin halogenation occurs at position 7 of the indole ring, the halogenase Thal of thienodolin biosynthesis chlorinates at position 6, and PyrH at position 5 during pyrroindomycin B assembly.

It is noteworthy that the tryptophan 2-halogenase CmdE acts on a carrier-attached tryptophanyl-moiety in the chondramide B and D [figure 1.4.1.2-1] assembly line. An inactivation mutant of *cmdE* did not accept the corresponding 2-chloro-tryptophan as a substrate for the biosynthesis, which in conclusion would exclude free tryptophan as the substrate of CmdE (Rachid *et al.*, 2006). However, since CmdE rather chlorinates a pyrrole moiety as described for pyoluteorin [figure 1.4.1.1-1], it is not surprising that this reaction is performed carrier-bound.

Although the majority of tryptophan halogenases utilise the same substrate, i.e. free tryptophan, they are known to generate only a single chlorotryptophan isomer, for example RebH of rebeccamycin biosynthesis, which mediates chlorination at position 7 of the indole ring, whereas Thal chlorinates at position 6 in the production of thienodolin (Zhu *et al.*, 2009; Vaillancourt *et al.*, 2006; Kling *et al.*, 2005). As well, KtzQ and KtzR are two distinct $FADH_2$-dependent halogenases involved in the formation of 6,7-dichloro-L-tryptophan in the biosynthesis of kutznerides (Strieker *et al.*, 2009). Dichlorination occurs by successive tandem action of the two halogenases on free tryptophan and its monochlorinated derivative, respectively [figure 1.4.1.2-2]. Thus, for the production of the dichlorinated tryptophan derivative 6,7-dichloro-L-tryptophan, two distinct enzymes are required (Fujimori *et al.*, 2007; Heemstra, Jr. & Walsh, 2008).

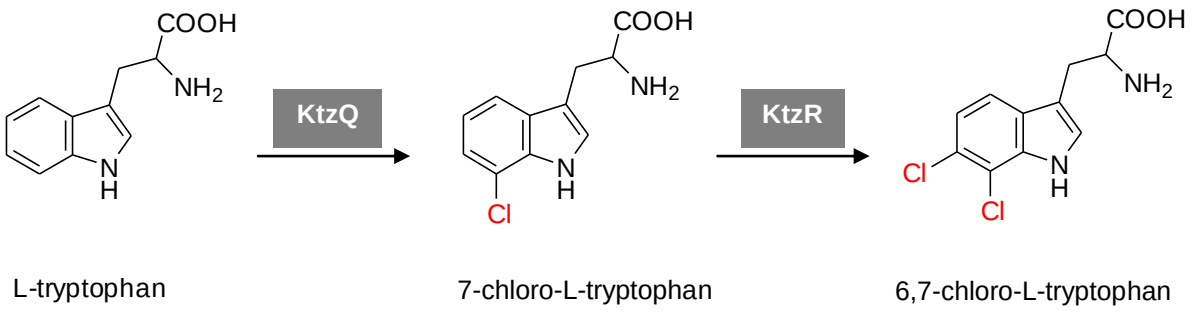


Figure 1.4.1.2-2: The dichlorination of L-tryptophan in the biosynthesis of kutznerides is performed by two different $FADH_2$-dependent halogenases.

1.4.2 Chorismate and halogenases in the biosynthetic process

As mentioned in 1.4, the biosynthesis of ambigols is proposed to involve a halogenase, which utilises a chorismate derived aromatic substrate. Chorismic acid is an important branch point intermediate that arises from the shikimate pathway (Knaggs, 2003) and is the precursor of the aromatic amino acids tryptophan, phenylalanine and tyrosine (Herrmann & Weaver, 1999; Arcuri *et al.*, 2010). It plays an important role for downstream biosynthetic assembly lines, in particular for the biosynthesis of nonribosomal peptides as well as mixed NRPS/PKS compounds (Du & Shen, 2001), alkaloids and also other amino acid derived frameworks like the amino coumarin scaffold of novobiocin and its chlorinated analogue clorobiocin

(Walsh *et al.*, 2006; Pojer *et al.*, 2003a; Pojer *et al.*, 2003b; Eustaquio *et al.*, 2003). As well, simple hydroxybenzoic acid derivatives have their origin in chorismate, for instance salicylic acid or 2,3-dihydroxybenzoic acid, which are precursors of siderophores, e.g. vibriobactin or mycobactin [figure 1.4.2-1] (Neres *et al.*, 2008). Nitrogen-containing compounds like the antitumor antibiotic enediyne C-1027 and the antibiotic metabolite chloramphenicol [figure 1.4.1.1-3] contain partial structures that are chorismate-derived (Van Lanen *et al.*, 2008).

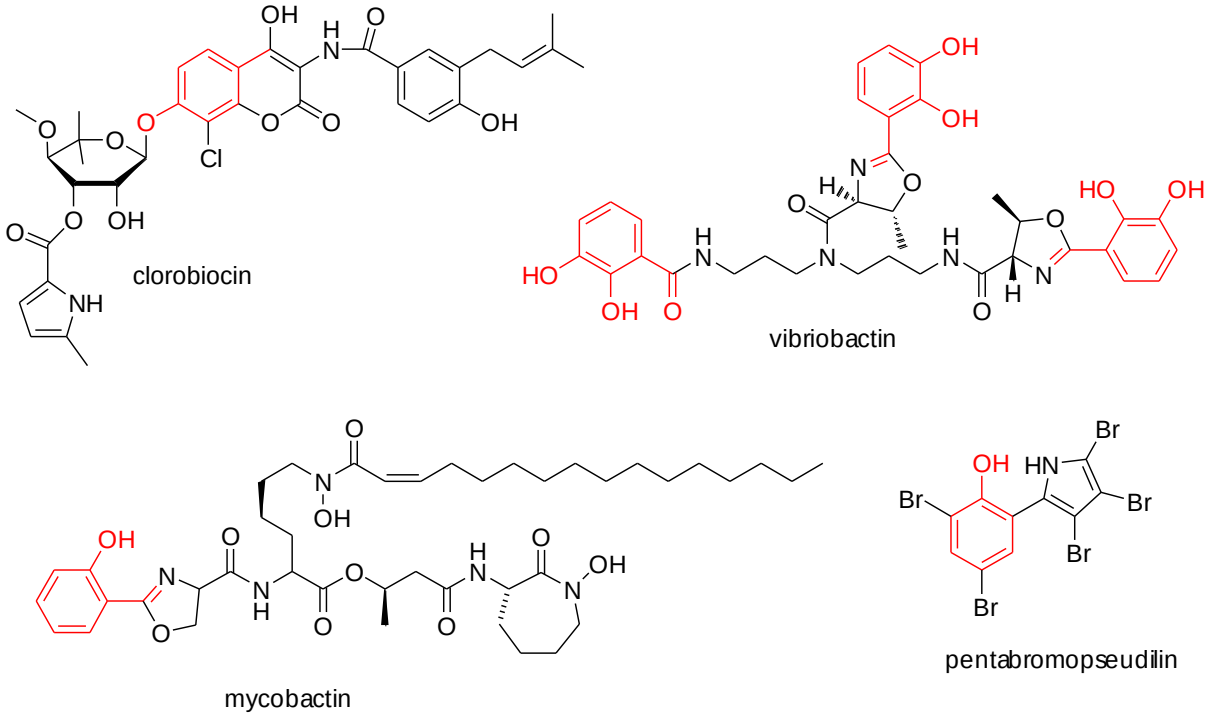


Figure 1.4.2-1: Bacterial metabolites with a phenolic building unit, which arises from chorismate. The amino coumarin framework of clorobiocin is derived from tyrosine, whereas vibriobactin, mycobactin or pentabromopseudilin contain a phenolic reisidue that results from a hydroxycarboxylic acid precursor.

Chorismate-derived metabolites with halide substituents as tailoring modifications are no rareness. The most typical substrates for $FADH_2$-dependent halogenases are amino acids like tyrosine or derivatives thereof as well as tryptophan (1.4.1.1 and 1.4.1.2). In this regard, it is very common that phenolic halogenases are often associated with NRPS gene clusters (Wagner *et al.*, 2009). However, also a simple building block like 4-hydroxybenzoic acid may be a substrate for halogenating enzymes, for instance in the biosynthesis of the marine antibiotic pentabromopseudilin [figure 1.4.2-1] that was obtained from different bacterial strains

including *Alteromonas luteoviolaceus* (Peschke *et al.*, 2005; Hanefeld *et al.*, 1994; Laatsch & Pudleiner, 1989).

1.4.3 AMP-dependent enzymes in the biosynthetic process

In chapter 1.4.1.1, it was outlined that some halogenases recognise their substrates exclusively when these are attached to peptidyl carrier domains. In order to link a PCP domain to a substrate, the latter has to be activated to its corresponding adenylate. This adenylation reaction is mediated by a class of enzymes refered to as AMP ligases.

Adenylation is a smart biological way to chemically activate carboxylate substrates by condensing them with ATP under pyrophosphate release. The superfamily of adenylate-forming enzymes comprises three subfamilies:

a. adenylation domains of NRPS (Linne *et al.*, 2007; Marahiel *et al.*, 1997), including freestanding A domains as described in 1.4.1.1
b. acyl and aryl CoA synthetases or ligases (Ingram-Smith *et al.*, 2006b)
c. luciferase oxidoreductase (White & Branchini, 1975)

All adenylate-forming enzymes catalyse a two step ping-pong reaction (Horswill & Escalante-Semerena, 2002), in which a carboxylic acid is adenylated under pyrophosphate liberation in a first half-reaction. The resulting carboxylate adenylate is a very reactive species, and thus the enzyme catalyses the second step as well, which is the reaction of the intermediate with a nucleophile. For most enzymes, this nucleophile is a phosphopantetheinyl moiety of either CoA, an aryl carrier protein (ArCP), or a peptidyl carrier protein (PCP), respectively [figure 1.4.3-1] (Ingram-Smith *et al.*, 2006b).

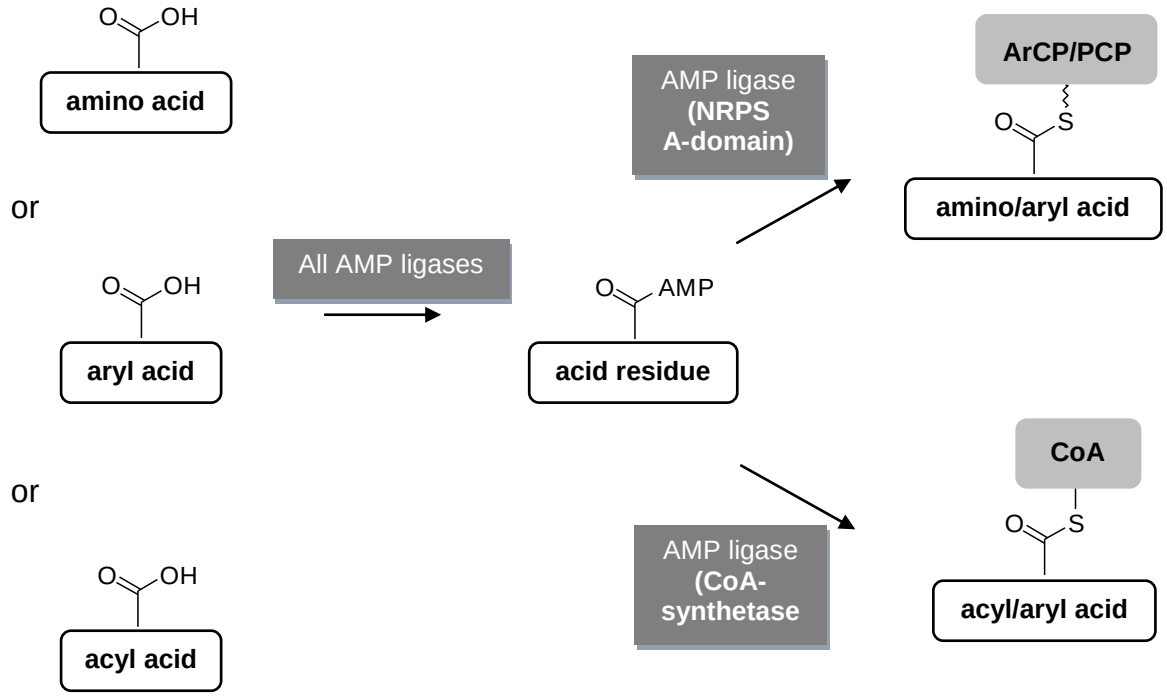


Figure 1.4.3-1: Activation of amino acids and aryl/acyl acids by AMP-ligases. Aryl acids may either be attached to an ArCP or a PCP domain. Alternatively, a CoA-ester may be formed by a group of AMP-ligases, the CoA synthetases, known from polyketide biosynthesis (Walsh, 2004; Townsend, 1997).

Besides their involvement in primary lipid metabolism, AMP-ligases are important substrate-activating proteins in secondary metabolism, for example as a part of modular NRPS, but also lone standing A domains are reported (Miethke *et al.*, 2006). Such independent aryl acid adenylation (AAA) domains give rise to adenylated carboxylic acids in the formation of several aryl-capped siderophores, e.g. enterobactin from *E. coli* or yersiniabactin from *Yersinia pestis* (Neres *et al.*, 2008).

EntB, one of six enzymes (EntA-EntF) of the enterobactin biosynthesis, has been described as a bifunctional protein with isochorismate lyase activity. EntC, EntB and EntA act in tandem to yield 2,3-dihydroxybenzoic acid, which is thioesterified to the ArCP domain of the bifunctional enzyme EntB. Subsequently, a 2,3-dihydroxybenzoic acid-AMP ligase, EntE is responsible for ATP-dependent acylation of EntB with 2,3-dihydroxybenzoic acid via adenylation [figure 1.4.3-2] (Gehring *et al.*, 1998a; Gehring *et al.*, 1997). Very recently, the two-step adenylation-ligation reaction of EntE was analysed for kinetic parameters with focus on possible siderophore inhibition strategies against pathogenic bacteria (Sikora *et al.*, 2010).

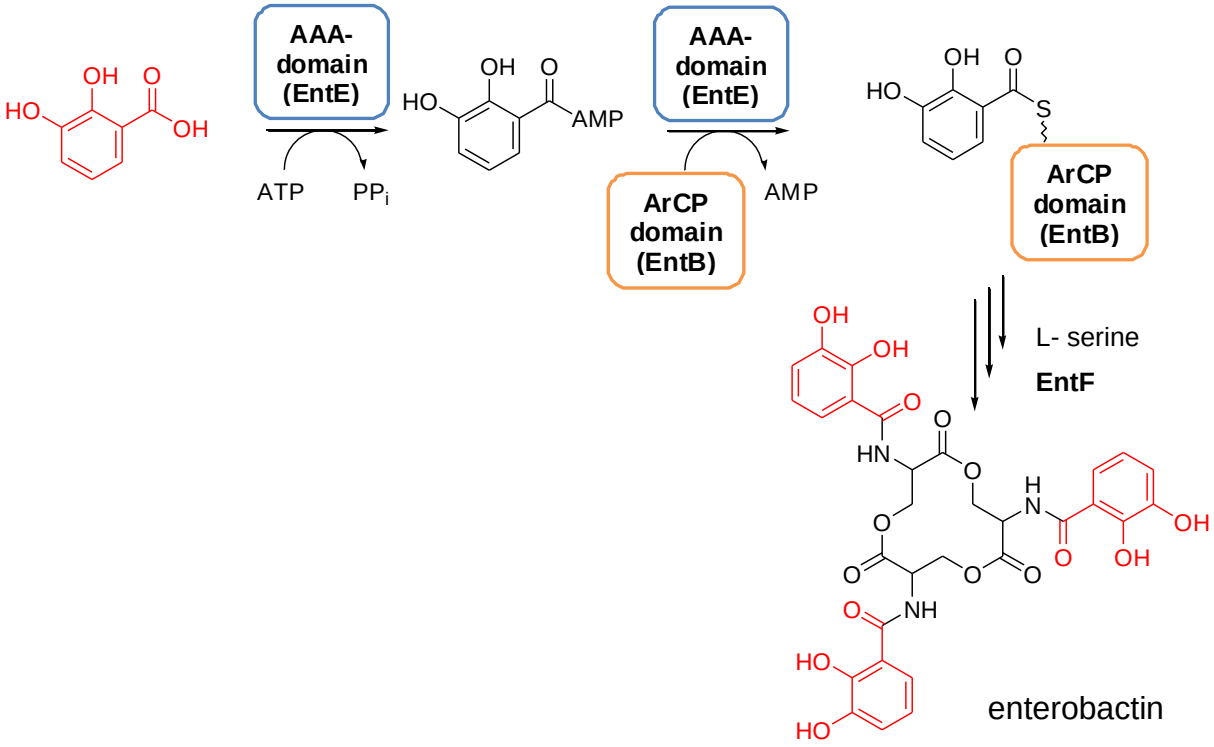


Figure 1.4.3-2: Biosynthesis of enterobactin, in which the aryl acid, 2,3-dihydroxybenzoic acid is activated by a 2,3-dihydroxy-AMP ligase EntE, which represents an aryl acid adenylation (AAA) domain. The adenylated substrate is subsequently loaded onto the ArCP domain of EntB, in order to be linked to L-serine by action of the starter C domain of the NRPS EntF (see figure 1.4.5-1). Three of these dimers are coupled through an ester bond. A final cyclisation step catalysed by the thioesterase domain of EntF leads to the trilacton enterobactin (Roche & Walsh, 2003).

The aryl carrier proteins EntB and VibB, respectively, are famous examples for freestanding representatives of usually integrated thiolation domains. As well, their substrates are activated by the lone standing A domains EntE and VibE, respectively (Gehring *et al.*, 1998b; Keating *et al.*, 2000a; Keating *et al.*, 2000b; Hutchinson, 2003).

1.4.4 Cytochrome P450-dependent enzymes in biosynthesis

The overall ambigol core structure [figure 1.1-1] is congruent with parts of the heptapeptide backbone of glycopeptide antibiotics, e.g. balhimycin or vancomycin [figure 1.4.4-1], whose aromatic residues are linked via phenyl ether and biphenyl bridges that arise via crosslinking reactions, mediated by CYP 450 enzymes (Woithe *et al.*, 2007; Stegmann *et al.*, 2006; Pylypenko *et al.*, 2003; Zerbe *et al.*, 2002).

Based on this common structural feature, it was concluded that the biosynthesis of the ambigol scaffold likewise involves this class of oxidative enzymes (1.4).

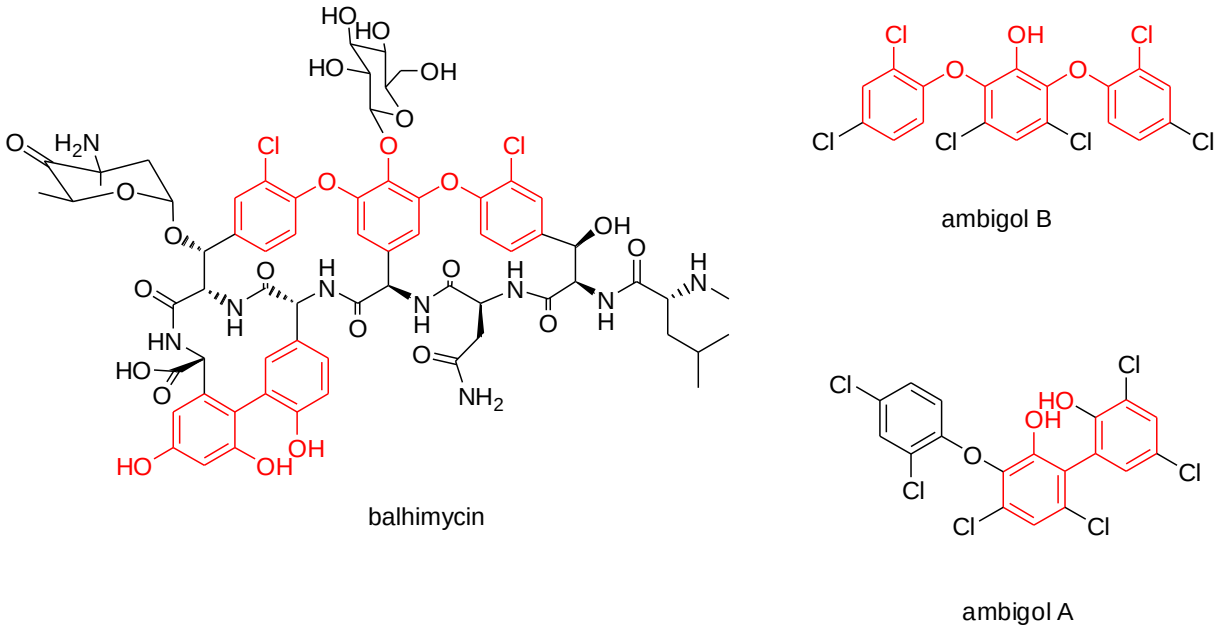


Figure 1.4.4-1: Ambigols A and B share structural similarity with parts of the balhimycin molecule.

Likewise, the production of the indolocarbazole scaffold of tjipanazole D [figure 1.1-1] is hypothesised to be maintained through oxidative dimerisation of 5-chlorotryptophan and intramolecular aryl-aryl coupling of the bisindole dimer between carbons 12a and 12b [figure 1.3.2-2]. This possible biosynthetic pathway, which is similar to that of rebeccamycin and related indolocarbazole compounds (Nakano & Omura, 2009; Sánchez *et al.*, 2006b) would then involve a chromopyrrolic acid synthase-like protein, an FAD-dependent monooxygenase and a CYP 450 enzyme (Ryan & Drennan, 2009; Howard-Jones & Walsh, 2007; Ryan *et al.*, 2007; Howard-Jones & Walsh, 2006). In this study, the collection of sequence data on tjipanazole D biosynthesis genes aimed to confirm this hypothesis.

CYP 450 monooxygenases are a superfamily of more than 8,500 heme-thiolate enzymes (Lewis & Wiseman, 2005; Werck-Reichhart & Feyereisen, 2000) containing an enzyme-bound prosthetic group (iron-protoporphyrin). These proteins comprise about 400 amino acids and represent a highly versatile class of oxidative enzymes catalysing reactions of primary metabolism, e.g. biodegradation pathways as well as secondary metabolism, i.e. biosynthesis pathways. Most of these reactions are

hydroxylations of aliphatic and aromatic carbons, however a large spectrum of other reactions is carried out as well (Guengerich, 2001). Some of them are *N*-oxidation, reductive dehalogenation (Li & Wackett, 1993), *N*-, *S*-, and *O*-dealkylation or epoxidation, e.g. in cryptophycin biosynthesis (Ding *et al.*, 2008; Magarvey *et al*., 2006). Furthermore, they can be involved in the formation of C-C- and C-O-bonds between aromatic rings by a radical mechanism [figure 1.4.4-2]. A free radical is generated by one-electron oxidation, and mesomeric structures arise by delocalisation of the free electron to positions *ortho* and *para* to the oxygen functionality. Enolised final products reveal carbon-carbon bonds either at *ortho*- or *para*-positions, but also ether linkages may be formed (Dewick, 2009).

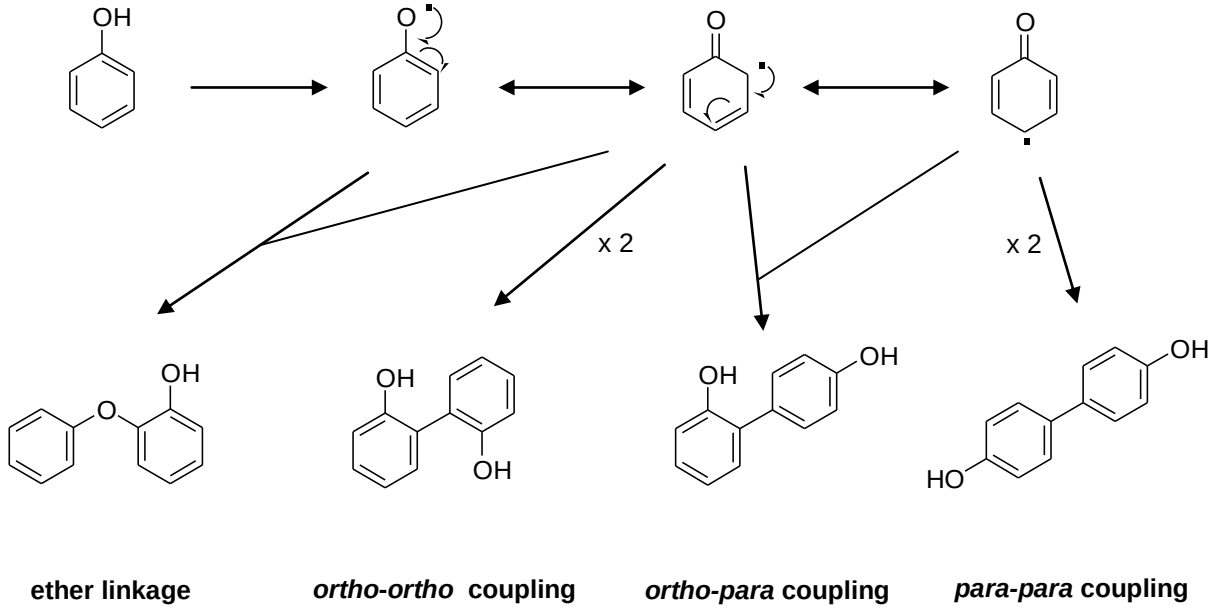


Figure 1.4.4-2: Range of dimeric systems generated by phenolic oxidative coupling reactions according to Dewick, 2009.

As described for halogenases (1.4.1.1), several CYP 450's were reported, which instead of binding free ligands, rely on an interaction with PCP domains. For instance, in the assembly lines of vancomycin-type antibiotics, the heptapeptide backbone is linked to a PCP domain when crosslinking occurs (Pylypenko *et al.*, 2003; Zerbe *et al.*, 2002). As well, NovI and CumD are likely responsible for carrier-bound β-hydroxylation of tyrosine, in order to produce the amino coumarin moiety in the biosynthesis of the antibiotic compounds novobiocin (Steffensky *et al.*, 2000a; Steffensky *et al.*, 2000b) and coumermycin A (Wang *et al.*, 2000), respectively.

The antibiotic novobiocin is produced by *Streptomyces spheroides* and contains a tyrosine-derived 3-amino-4-hydroxycoumarin moiety as its structural core. In the biosynthetic gene cluster of novobiocin, two genes were found, *novH* and *novI* that encode an NRPS-type didomain and a CYP 450 hydroxylase, respectively [figure 1.4.4-3]. The NRPS didomain consists of an L-tyrosine specific A domain and a C-terminal PCP domain. The biosynthesis of the coumarin scaffold starts with L-tyrosine, which is tethered to the PCP domain of NovH by action of an A domain, located upstream. Hydroxylation of the amino acid side chain is then performed by the CYP 450 enzyme NovI on the PCP-tethered substrate, L-tyrosyl-S-NovH. Following steps are also performed on the carrier-attached substrate and include *ortho*-hydroxylation of the phenyl ring and intramolecular cyclisation with coincident release of the coumarin derivative (Chen & Walsh, 2001). This example shows possible interactions between CYP 450 enzymes and PCP domains in biosynthetic pathways.

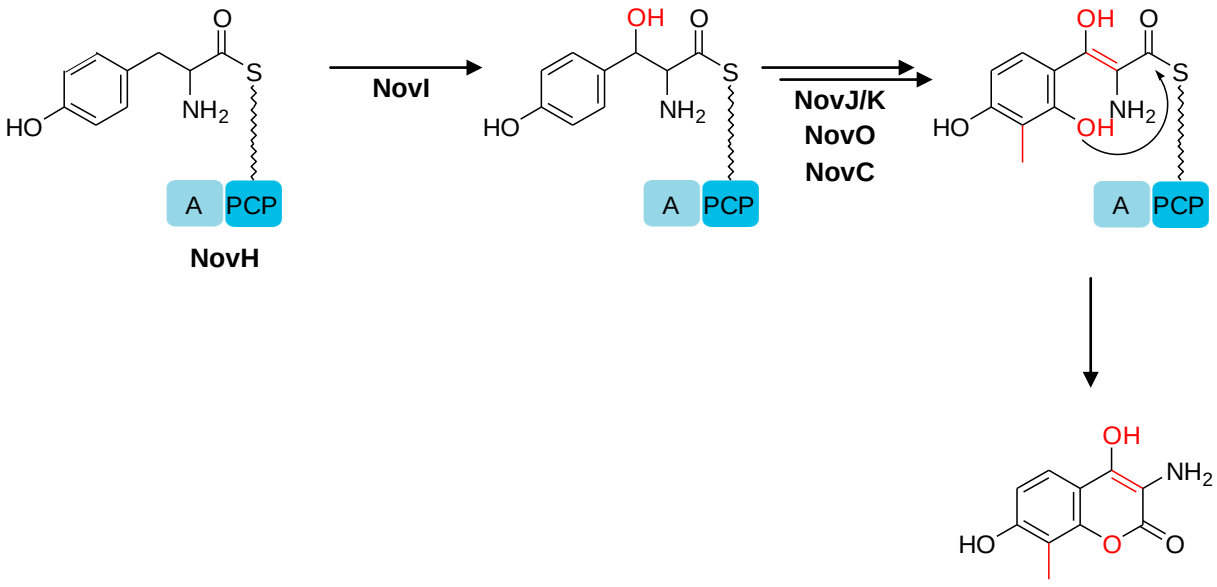


Figure 1.4.4-3: PCP-bound hydroxylation of tyrosine in the formation of the coumarin ring during novobiocin production. NovI = CYP 450 enzyme; NovJ/K = similar to 3-keto-ACP reductases; NovO = similarity to quinone C-methyltransferases; NovC = putative flavin-dependent monooxygenase. Spontaneous ring closure releases the final product

Nearly the same hydroxylation procedure was found for the production of a β-*R*-hydroxytyrosine moiety, which is an important precursor in vancomycin biosynthesis. The CYP 450 enzyme OxyD, whose crystal structure was solved recently, turned out to hydroxylate its substrate while this is bound to the PCP domain of the didomain BpsD rather than free in solution. As described for NovH, BpsD consists of an A and a PCP domain only (Cryle *et al.*, 2010).

Also ACP-bound fatty acids can be substrates for CYP 450's. $P450_{BioI}$ (CYP107H1) from biotin biosynthesis in *Bacillus subtilis* is responsible for the formation of a seven-carbon diacid via oxidative cleavage of a fatty acid C-C bond. This reaction includes several reaction steps starting with hydroxylation of a fatty acid linked to an ACP (Cryle & De Voss, 2004; Cryle & Schlichting, 2008).

1.4.5 Starter C domains in NRPS

In the presented study, an NRPS-module bearing a starter C domain was identified. The NRPS module is believed to play a crucial role in the enzymatic action of ambigol biosynthesis proteins.

NRPS are multienzyme complexes that are composed of highly specific catalytic domains, which are arranged in a modular fashion fascilitating intramolecular transfer of peptide chains from one domain to the next (Strieker *et al.*, 2010). A typical NRPS module comprises at least three protein domains, which are an A domain (~550 aa), a PCP (T) domain (~80 aa) and a C domain (~550 aa) (Finking & Marahiel, 2004; Walsh, 2003; Stachelhaus *et al.*, 1996b). These domains are usually found as the typical C-A-T module. By contrast, also lone standing A, PCP and C domains were reported, e.g. from the biosynthesis of iron(III)-chelating siderophores (Hutchinson, 2003; Sharma & Johri, 2003).

NRPS usually start with an A domain activating the first amino acid to the corresponding aminoacyl-AMP. Subsequently, the adenylated amino acid is loaded onto the contiguous PCP domain (Walsh *et al.*, 1997; Stachelhaus *et al.*, 1996a). The catalysed reaction of an adjacent C domain involves nucleophilic attack by the amino group of the downstream PCP-tethered amino acid on the acyl group of the growing chain while this is attached to the upstream PCP domain (Koglin & Walsh, 2009; Lautru & Challis, 2004).

Divergently, also an aryl acid, can be recognised by a freestanding A domain (1.4.3). This was reported for the production of siderophores, e.g. the biosynthesis of enterobactin [figure 1.4.3-2], in which 2,3-dihydroxybenzoic acid (DHB) is tethered to a PCP domain, in order to be linked to an amino acid, i.e. L-serine by the action of a so called starter C domain (Rausch *et al.*, 2007) [figure 1.4.5-1].

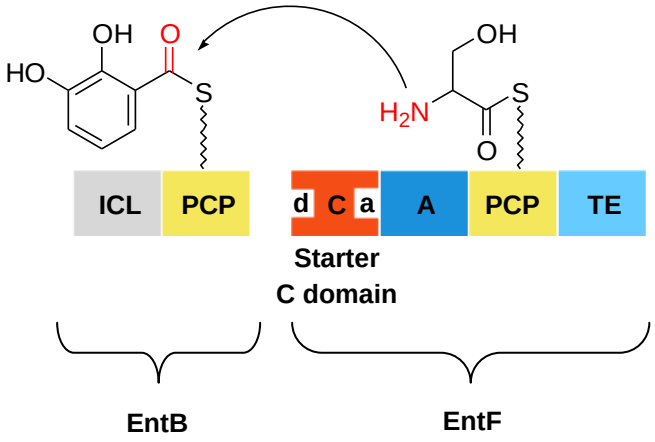


Figure 1.4.5-1: Enterobactin biosynthesis: peptide bond formation between an aryl acid and an amino acid, catalysed by an N-terminal (starter) C domain located on the NRPS module EntF. ICL = isochorismate lyse; PCP = peptidyl carrier protein; C = C domain with d (donor site) and a (acceptor site); A = A domain; TE = thioesterase domain.

EntE, a dissected A domain activates 2,3-DHB to give the corresponding acyl-AMP derivative, which is transferred also by EntE to the ArCP-domain of the bifunctional protein *holo*-EntB. The serine activating NRPS EntF contains a starter C domain, which receives 2,3-DHB at its donor site. The downstream located PCP domain of EntF binds L-serine, which is linked to 2,3-DHB to form 2,3-dihydroxybenzoylserine. Three such molecules are connected through ester bonds by iterative action of the TE domain of EntF giving the final trilacton enterbactin (Roche & Walsh, 2003).

Starter C domains are also reported for several lipopeptides e.g. surfactin from *B. subtilis* or lichenysins from *Bacillus licheniformis*. These compounds contain a β-hydroxyl fatty acid, which is connected to the carboxy group of the C-terminal amino acid, e.g. L-glutamic acid of the peptide (Konz *et al.*, 1999).

2. Intent of the presented study

The cyanobacterium *F. ambigua* produces highly interesting compounds, i.e. ambigols A-C and tjipanazole D (1.3). These compounds bear aromatic frameworks, whose biosynthesis most likely involves oxidative processes, i.e. aromatic coupling reactions and halogenations. The enzymatic control of these reactions enables strict regioselectivity paired with high effectiveness of catalysed steps. Thus, the knowledge of involved genes encoding the respective enzymes is of enormous interest with regard to their possible use for in vitro catalysis of biosynthesis reactions.

At the start of the current project, a suitable DNA extraction method for *F. ambigua* was sought, in order to obtain highly pure DNA for the establishment of a fosmid DNA library. In a preceding study, partial halogenase sequences were obtained applying PCR techniques (Wagner *et al.*, 2009). These sequences were designated to be used for a screening of the genomic library.

Thus, the here presented study aimed the construction and screening of a genomic library of the *F. ambigua* genome and the identification of the ambigols and tjipanazole D biosynthetic gene clusters as well as the bioinformatic analysis of the latter towards their possible involvement in secondary metabolites biosynthesis. This analysis is facilitated by the use of online databases (3.1.13) (Walsh & Fischbach, 2010), which contain collected information on genes and encoded proteins with proven or putative function. Based on sequence homology, proteins may be assigned to possible catalysed reactions. The most specific enzymes of the here investigated biosyntheses seemed to be the halogenases. In particular, the tryptophan halogenase of the tjipanazole D biosynthesis appeared to be a suitable enzyme for heterologous expression, based on compareable approaches in literature (1.3.2; 5.7.2). Since tryptophan halogenases utilise the simple substrate tryptophan as a free diffusible molecule, they are excellent candidates for in vitro halogenations. Therefore, a target of this study was the recombinant production of the putative tryptophan halogenase of the tjipanazole D biosynthesis and its experimental use. As an ambitious objective, the manipulation of the substrate susceptibility of this halogenase was planned by site-directed mutagenesis (Lang *et al*., 2011), in order to establish a compound library.

3. Materials and methods

3.1 Materials

3.1.1 Chemicals and other materials

Table 3.1.1-1 gives a list of chemicals used in this study. All chemicals listed were purchased in research grade/*pro analysi* quality. Those not mentioned below were supplied by Fluka Chemie GmbH (Buchs, Switzerland), Merck KGaA (Darmstadt, Germany), Roth Chemie GmbH (Karlsruhe, Germany) or Sigma-Aldrich Chemie GmbH (Steinheim, Germany).

Table 3.1.1-1: Purchased chemical substances and solutions used in this work.

Substance	Manufacturer
Acetic acid	KMF Laborchemie Handels GmbH (Darmstadt, Germany)
Acetone-d 99.8% D	Deutero GmbH (Kastellaun, Germany)
Agar	Fluka Chemie GmbH (Buchs, Switzerland)
Ammonium acetate	Roth Chemie GmbH (Karlsruhe, Germany)
Ampicillin	Roth Chemie GmbH (Karlsruhe, Germany)
BG-11 stock solution	Sigma-Aldrich Chemie GmbH (Steinheim, Germany)
Boric acid	Roth Chemie GmbH (Karlsruhe, Germany)
Bromophenol blue	Merck KGaA (Darmstadt, Germany)
BSA	Bio-Rad Laboratories GmbH (Munich, Germany)
$CaCl_2 \times 2\ H_2O$	Merck KGaA (Darmstadt, Germany)
Chloramphenicol	Fluka Chemie GmbH (Buchs, Switzerland)
Chloroform	Roth Chemie GmbH (Karlsruhe, Germany)
CTAB	Roth Chemie GmbH (Karlsruhe, Germany)
DMSO	Roth Chemie GmbH (Karlsruhe, Germany)
dATP	Fermentas GmbH (St. Leon Rot, Germany)

Substance	**Manufacturer**
dNTP	Promega GmbH (Mannheim, Germany) Fermentas GmbH (St. Leon Rot, Germany)
Ethanol 99,8% p.a.	Roth Chemie GmbH (Karlsruhe, Germany)
Ethidium bromide	Roth Chemie GmbH (Karlsruhe, Germany)
Glycerol	Roth Chemie GmbH (Karlsruhe, Germany)
Gel Loading Dye	Fermentas GmbH (St. Leon Rot, Germany)
GoTaq® Flexi Buffer (10 x)	Promega GmbH (Mannheim, Germany)
Isoamyl alcohol (3-methyl-1-butanol)	Roth Chemie GmbH (Karlsruhe, Germany)
Isopropanol	Roth Chemie GmbH (Karlsruhe, Germany) Merck KGaA (Darmstadt, Germany)
D(+)-Maltose x H_2O	Roth Chemie GmbH (Karlsruhe, Germany)
β-Mercaptoethanol	Roth Chemie GmbH (Karlsruhe, Germany)
$MgCl_2$ × 6 H_2O	Merck KGaA (Darmstadt, Germany)
$MgSO_4$ x 7 H_2O	Merck KGaA (Darmstadt, Germany)
Na-acetate	Merck KGaA (Darmstadt, Germany)
NaCl	Merck KGaA (Darmstadt, Germany)
Na_2-EDTA	Roth Chemie GmbH (Karlsruhe, Germany)
NaOH	Merck KGaA (Darmstadt, Germany)
peqGOLD Agarose	PEQLAB Biotechnologie GMBH (Erlangen, Germany)
peqGOLD Low Melt-Agarose	PEQLAB Biotechnologie GMBH (Erlangen, Germany)
Phenol	Merck KGaA (Darmstadt, Germany)
Sarcosyl (*N*-Lauroylsarcosin)	Sigma-Aldrich Chemie GmbH (Steinheim, Germany)
SDS	Roth Chemie GmbH (Karlsruhe, Germany)
D (+)-Saccharose (Sucrose)	Roth Chemie GmbH (Karlsruhe, Germany)
Tris	Roth Chemie GmbH (Karlsruhe, Germany)
Tris-HCl	Roth Chemie GmbH (Karlsruhe, Germany)
Tryptone/Peptone from Caseine	Roth Chemie GmbH (Karlsruhe, Germany)
Yeast extract	Fluka Chemie GmbH (Buchs, Switzerland)

Other materials used in this study are listed in table 3.1.1-2

Table 3.1.1-2: Technical equipment and other materials used in this study.

Material	**Manufacturer**
Autoclave	Varioklav®, H+P Labortechnik AG (Oberschleißheim, Germany)
Amicon Ultra Centrifugal Filters	Millipore GmbH (Schwalbach, Germany)
Bandelin Sonorex RK 31	Bandelin electronic GmbH & Co. KG (Berlin, Germany)
Biometra T3000 Thermocycler	Biometra GmbH (Goettingen, Germany)
BioRad PowerPac™ Basic	Bio-Rad Laboratories GmbH (Hercules, U.S.A.)
Boekel Replicator	Boekel Scientific (Feasterville, U.S.A.)
Centrifuge Heraeus Biofuge *fresco*	Thermo Fisher Scientific (Waltham, U.S.A.)
Centrifuge Heraeus Contifuge Stratos	Thermo Fisher Scientific (Waltham, U.S.A.)
Centrifuge Heraeus Fresco 17	Thermo Fisher Scientific (Waltham, U.S.A.)
Centrifuge tubes (15/50 ml)	TPP AG (Trasadingen, Germany)
CopyControl™ Induction solution	Epicentre (Madison, U.S.A.)
Eppendorf Centrifuge 5415 D	Eppendorf (Hamburg, Germany)
Eppendorf tubes (0.5, 1.5, 2 ml)	Eppendorf (Hamburg, Germany)
Gel chambers Horizon 58 and Horizon 11.14	Life technologies (Karlsruhe, Germany)
Incubator	Memmert GmbH + Co. KG (Schwalbach, Germany)
Inolab pH meter	WTW GmbH (Weilheim, Germany)
Intas iX Imager	Intas Science Imaging Instruments GmbH (Göttingen, Germany)
Kodak DC290	Kodak GmbH (Stuttgart, Germany)
Laminar Airflow Clean Bench BSB 4A (Hera Safe, Class II)	Heraeus (Hanau, Germany)
Magnetic stirrer (IKA® RH basic)	IKA® Werke GmbH & Co. KG (Staufen, Germany)

Material	Manufacturer
Milli-Q® Water System	Millipore (Eschborn, Germany)
Multittron incubation shaker	Infors AG (Bottmingen, Switzerland)
MS2 Minishaker	IKA® Works Inc. (Wilmington, U.S.A.)
Nalgene cryogenic vials	Nalgene Nunc International (Rochester, U.S.A.)
Olympus BX 51	Olympus America Inc. (Center Valley, U.S.A.)
Overhead mixer (IKA® RW 20)	IKA® Werke GmbH & Co. KG (Staufen, Germany)
Parafilm®	Pechiney Plastic Packaging Company (Chicago, U.S.A.)
Potter S Homogenisator	Sartorius AG (Göttingen, Germany)
Scale (Sartorius BL 3100)	Sartorius AG (Göttingen, Germany)
Scale (Sartorius BP 221S)	Sartorius AG (Göttingen, Germany)
Sterile filter (0.2 µm)	Renner GmbH (Dannstadt, Germany)
Thermomixer Eppendorf	Eppendorf (Hamburg, Germany)
Transferpette®-8	Brand GmbH + Co. KG. (Wertheim, Germany)
UV mini 1240 UV/Vis spectro-photometer	Shimadzu (Kyoto, Japan)
UV cuvettes	Ratiolab GmbH (Dreieich, Germany)
Water bath (Haake DC 10)	Thermo Haake GmbH (Karlsruhe, Germany)

3.1.2 Enzymes

Enzymes used in this study were supplied as listed in table 3.1.2-1. Appropriate enzyme reaction buffers, which provide optimal reaction conditions as for polymerase, restriction and ligation were purchased together with the enzymes.

Table 3.1.2-1: Enzymes that were used in this study for molecular biological experiments.

Enzyme	Manufacturer
Agarase (0.5 u/µl)	Fermentas GmbH (St. Leon Rot, Germany)
GoTaq® Flexi DNA Polymerase (5 u/µl)	Promega (Mannheim, Germany)
Lysozyme	Roth (Karlsruhe, Germany)
Proteinase K	Roth (Karlsruhe, Germany)
Restriction enzymes	Fermentas GmbH (St. Leon Rot, Germany)
RNase (DNase free)	Roth Promega (Mannheim, Germany)
Fast-Link™ DNA Ligase	Epicentre (Madison, U.S.A.)
T4-ligase	Fermentas GmbH (St. Leon Rot, Germany)

All enzymes were applied following the respective manufacturer's recommendations for use. Lysozyme and Proteinase K were used as stock solutions with concentrations of 100 mg/ml and 20 mg/ml, respectively. Restriction enzymes were purchased together with the appropriate reaction buffers and were used according to the provided company's protocols.

3.1.3 Molecular weight markers

The following DNA standards listed in table 3.1.3-1 were used for gel electrophoresis.

Table 3.1.3-1: Molecular weight markers applied for size estimation of DNA in gel electrophoresis

Molecular weight marker	Manufacturer
Gene Ruler™ DNA Ladder Mix	Fermentas GmbH (St. Leon Rot, Germany)
Lambda Mix Marker	Fermentas GmbH (St. Leon Rot, Germany)
36 kb Fosmid Control DNA (100ng/ul)	Epicentre Biotechnologies (Madison, U.S.A.)

3.1.4 Molecular biological kits

Commercial molecular kits that were utilised in this work are listed in table 3.1.4-1. The application of the kits was performed according to the manufacturer's instruction manuals.

Table 3.1.4-1: Commercial molecular biological kits used in this work for different applications.

Molecular biological kit	Manufacturer
QIAquick PCR Purification Kit	Qiagen GmbH (Hilden, Germany)
pGEM®-T Vector System I	Promega (Mannheim)
QIAquick Gel Extraction Kit	Qiagen GmbH (Hilden, Germany)
QIAGEN Plasmid Midi Kit	Qiagen GmbH (Hilden, Germany)
QIAGEN Blood & Cell Culture DNA Mini Kit	Qiagen GmbH (Hilden, Germany)
QIAprep®Spin Miniprep Kit	Qiagen GmbH (Hilden, Germany)
GeneJET™ Plasmid Miniprep Kit	Fermentas GmbH (St. Leon Rot, Germany)
CopyControl™ Fosmid Library Production Kit	Epicentre Biotechnologies (Madison, U.S.A.)
Fast-Link™ DNA Ligation Kit	Epicentre Biotechnologies (Madison, U.S.A.)

Molecular biological kit	Manufacturer
Quick Blunting Kit	New England Biolabs (Frankfurt, Germany)
TOPO® TA Cloning® Kit	Invitrogen (Karlsruhe, Germany)

3.1.5 Bacterial strains

Within this work, different *E. coli* host strains were utilised for transformation of various insert-vector constructs (3.4.9.3; 3.4.14.4). The *E. coli* strains, their genotype and manufacturers are listed in table 3.1.5-1.

Table 3.1.5-1: *E. coli* host strains applied in the presented study.

Strain	Genotype	Manufacturer
XL1-Blue competent *E.coli* cells	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*^{q}*ZΔM15* Tn*10* (Tet^{r})].	Agilent Technologies Deutschland GmbH (Böblingen, Germany)
TOP10 competent *E.coli* cells	F^{-} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (Str^{R}) *end*A1 *nup*G	Invitrogen (Karlsruhe, Germany)
Phage T1-Resistant TransforMax™ EPI-300™-T1^{R} chemically competent *E. coli*	F^{-} *mcr*A D(*mrr-hsd*RMS-*mcr*BC) f80d*lac*ZDM15 D*lac*X74 *rec*A1 *end*A1 *ara*D139 D(*ara*, *leu*)7697 *gal*U *gal*K λ^{-} *rps*L *nup*G *trf*A *ton*A *dhfr*	Epicentre Biotechnologies (Madison, U.S.A.)

3.1.6 Vectors

For phenolic halogenase screening as well as for fosmid subcloning techniques, DNA fragments were produced by PCR and ultrasonic treatment, respectively. These fragments were made available on plasmids via cloning procedures (3.4.9.3; 3.4.14.4; 3.4.14.10). The following vectors listed in table 3.1.6-1 were used in this work.

Table 3.1.6-1: Plasmids utilised in the presented study.

Vector	Size/Selectable marker	Manufacturer
pGEM®-T Vector	3.0 kb/ampicillin	Promega (Mannheim, Germany)
pCR®2.1-TOPO®	3.9 kb/ampicillin	Invitrogen (Karlsruhe, Germany)
pCC1FOS™ Vector (fosmid vector)	8.1 kb/chloramphenicol	Epicentre Biotechnologies (Madison, U.S.A.)

Recombinant plasmids generated within this work are listed in table 3.1.6-2.

Table 3.1.6-2: Vector-insert constructs produced in the presented study.

Construct	Applied vector	Insert
gem-phal-990	pGEM®-T	990 bp PCR product amplified from *F. ambigua* (template)
top-phal-700	pCR®2.1-TOPO®	700 bp PCR product amplified from fosmid E8 (4.3.4)
E8-3	pGEM®-T	830 bp long DNA fragment produced by sonication
E8-4	pGEM®-T	940 long DNA fragment produced by sonication
E8-5	pGEM®-T	856 long DNA fragment produced by sonication
E8-6	pGEM®-T	931 long DNA fragment produced by sonication

Construct	Applied vector	Insert
E8-7	pGEM®-T	913 long DNA fragment produced by sonication
E-8-8	pGEM®-T	928 bp long DNA fragment produced by sonication
E8-F4-M13	pCR®2.1-TOPO®	1017 bp long DNA fragment produced by sonication
E8-F61-M13	pCR®2.1-TOPO®	851 bp long DNA fragment produced by sonication
E8-F62-M13	pCR®2.1-TOPO®	962 bp long DNA fragment produced by sonication
E8-F71-M13	pCR®2.1-TOPO®	1000 bp long DNA fragment produced by sonication
E8-13	pGEM®-T	452 bp long DNA fragment produced by sonication
E8-15	pGEM®-T	999 bp long DNA fragment produced by sonication
E8-17	pGEM®-T	821 bp long DNA fragment produced by sonication
E8-18	pGEM®-T	824 bp long DNA fragment produced by sonication
E8-22	pGEM®-T	1009 bp long DNA fragment produced by sonication
E8-33	pGEM®-T	967 bp long DNA fragment produced by sonication

Construct	Applied vector	Insert
E8-34	pGEM®-T	902 bp long DNA fragment produced by sonication
E8-62	pGEM®-T	911 bp long DNA fragment produced by sonication
E8-69	pGEM®-T	966 bp long DNA fragment produced by sonication
Meo-6	pGEM®-T	966 bp long DNA fragment produced by sonication
FA16S clones:		
FA16S1-T7-4.3 FA16S2-T7-4.3 FA16S3-T7-4.3 FA16S2-T7-27.2 FA16S4-T7-27.2 FA16S5-T7-27.2	pGEM®-T	Approx. 1.5 kb long 16S rDNA fragments amplified from *F. ambigua* genomic DNA using the primers Cyan-16S-fw and Cyan-16S-rev [table 3.1.9-1]
Pstu clones:		
Pstu-1-T7 Pstu-3-T7 Pstu-4-T7 Pstu-5-T7 Pstu-7-T7 Pstu-8-T7 Pstu-9-T7	pGEM®-T	Approx. 1.5 kb long 16S rDNA fragments amplified from *P. stutzeri* genomic DNA using the primers Cyan-16S-fw and Cyan-16S-rev [table 3.1.9-1]

3.1.7 Fosmids

About 2,800 recombinant fosmids were generated, each containing approximately 40 kb of cyanobacterial recombinant DNA. The fosmid library was established using the CopyControl™ Fosmid Library Production Kit (3.1.4). Fosmid E8 listed in table 3.1.7-1 was analysed in detail within this project (4.4.4).

Table 3.1.7-1: Fosmid E8 was generated in the presented study and further analysed.

Fosmid	Vector	Insert
E8	pCC1FOS™ Vector	~ 45 kb of *F. ambigua* DNA

3.1.8 Phages

Phages used in this work for packaging of concatemer DNA (3.4.14.6) are presented in table 3.1.8-1.

Table 3.1.8-1: λ-phage applied in this work.

Phage	Type	Manufacturer
MaxPlax™ Lambda Packaging Extracts	λ-phage	Epicentre Biotechnologies (Madison, U.S.A.)

3.1.9 Oligonucleotides

In the presented work, several oligonucleotides were used as PCR primers (3.4.9), either specific or degenerate. They were designed from multiple sequence alignments (3.1.13). Table 3.1.9-1 shows the primers utilised in this project and the corresponding melting temperatures (T_M) as stated by the manufacturers. The primers were provided as lyophilised powders from Eurofins MWG Operon (Ebersberg, Germany) and were dissolved in TE buffer, adjusted to a concentration of 100 pmol/µl and stored at -20 °C. In standard PCR reaction mixtures, the primers were used at working concentrations of 10-20 pmol/µl.

Table 3.1.9-1: Oligonucleotides that were used as primers for PCR reactions.

Primer	Sequence	T_M
T7	TAATACGACTCACTATA	53
SP6	TATTTAGGTGACACTATAG	48
Epi-RP	CTCGTATGTTGTGTGGAATTGTGAGC	63
M13-fw	GTAAAACGACGGCCAG	52
M13-rev	CAGGAAACAGCTATGAC	50
T-Hal fw	GAAGCAACCTTCAGTACCATCAAG	63
T-Hal rev	TTCCAGATCCAGCCAGAACTTAAAG	63
C-Hal fw	GTCGTTTTGACGTGAGCG	61
C-Hal rev	CAACGATTGATCGAAAATACCTG	60
Chal-rev2	CCAIGCRTTIARCCA	48
Ckons-fw	GGT TCT GGT TTA GCT GGG	60
Ckons-rev	AGG AAT GAG CCA GAG CCA	60
T-Neu-fw	TAC AAA ACG GCA ATT AAG TTC AGC	58
T-Neu-rev	CTC TGT CAC ACA ACA GTG AGT CAG	63
Cyp-fw	GCT ATG CAT ACC TTT GTG	51
Cyp-rev	AGC TGC ATC AAA GAT CC	50
Kina-fw	TGA TCA CTC AGC CAT CAG	54
Kina-rev	TCG CAT ATC GTT AGA TAT ACC	54
Cyan-16S-fw	TCA GAW YGA ACG CTG GCG G (Fiore *et al*., 2000)	60
Cyan-16S-rev	AAG GAG GTG ATC CAG CC (Fiore *et al*., 2000)	55

The base abbreviations were used according to the IUPAC nucleotide code listed in table 3.1.9-2. Inosine (I) was integrated as a spacer base to reduce the degradation rate of primers.

Table 3.1.9-2: Base abbreviations according to IUPAC.

IUPAC nucleotide code	Base	IUPAC nucleotide code	Base
A	Adenine	W	A or T
C	Cytosine	K	G or T
G	Guanine	M	A or C
T	Thymine	B	C or G or T
U	Uracil	D	A or G or T
R	A or G	H	A or C or T
Y	C or T	V	A or C or G
S	G or C	N	any base

3.1.10 Water

For the preparation of culture media, demineralised water was provided by a reverse osmosis system (IMB, Germany). A Milli-Q® Academic Water Purification System (Millipore GmbH, Germany) was used to generate ultra-pure water prepared from demineralised water. Milli-Q® water was used for all applications, if not specified otherwise. For PCR application, autoclaved Milli-Q® water was applied.

3.1.11 Culture Media

Culture media were prepared with demineralised water prior to steam sterilization. Their composition was as listed in table 3.1.11-1.

Table 3.1.11-1: Culture media and their composition as they were used during this project.

Culture medium	Composition
BG-11 medium (blue green medium):	20 ml of 50 x BG-11 ready stock solution (purchased from Sigma-Aldrich Chemie GmbH, Karlsruhe, Germany) ad 1 l H_2O
BG-11 agar:	15 g Agar ad 1 l BG-11 medium
LB-medium (according to Luria-Bertani):	10 g Tryptone, 5 g Yeast extract, 10 g NaCl, ad 1 l H_2O, pH 7.5
LB-agar:	10 g Tryptone, 5 g Yeast extract, 5 g NaCl, 15 g Agar, ad 1 l H_2O, pH 7.5
LB-antibiotic-agar:	10 g Tryptone, 5 g Yeast extract, 5 g NaCl, 15 g Agar, ad 1 l H_2O, pH 7.5 sterile filtered ampicillin (final concentration of 100 µg/ml) or chloramphenicol (final concentration of 12.5 µg/ml) was added before casting the agar plates
2 x YT-medium	16 g Tryptone, 10 g Yeast extract, 5 g NaCl, ad 1 l H_2O, pH 7.5

3.1.12 Buffers and solutions

All buffers and solutions were prepared with ultra-pure water [table 3.1.12-1]. Stock solutions not containing any organic solvents or thermolabile compenents were usually steam sterilised prior to use (3.4.1).

Table 3.1.12-1: Buffers and solutions used within this study.

Buffers for DNA extraction:

Wash buffer to remove polysaccharides (Fiore *et al.*, 2000)	50 mM Tris-HCl (pH 8.0), 5 mM EDTA and 50 mM NaCl
Resuspension buffer (Fiore *et al.*, 2000)	50 mM Tris-HCl (pH 8.0), 50 mM EDTA
CTAB extraction buffer (Fiore *et al.*, 2000)	3% (w/v) CTAB, 1% (w/v) Sarkosyl, 20 mM EDTA, 1.4 M NaCl, 0.1 M Tris–HCl, pH 8.0, 1% (v/v) 2-mercaptoethanol
SET buffer (Sambrook & Russell, 2001)	75 mM NaCl, 25 mM EDTA, 10 mM Tris-HCl (pH 7.5)
TE buffer	10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)
RNase A stock solution (DNase-free)	50 mg lyophilised RNase A, 5 ml Tris-HCl (10 mM, pH 7.5), 15 mM NaCl, and storage at –20 °C
Buffer G2	800 mM Guanidine-HCl, 30 mM Tris-HCl (pH 8.0), 30 mM EDTA (pH 8.0), 5 % Tween-20; 0.5 % Triton X-100

Buffers for library construction:

Phage dilution buffer (PDB)	10 mM Tris-HCl (pH 8.3), 100 mM NaCl, 10 mM $MgCl_2$

Buffers for plasmid and fosmid purification:

Buffer P1	50 mM Tris-HCl (pH 8), 10 mM EDTA, 100 µg/mL RNase A
Buffer P2	200 mM NaOH, 1% SDS
Buffer P3	3M potassium acetate (pH 5.5)
Buffer QBT (Equilibration buffer)	750 mM NaCl, 50 mM MOPS (pH 7.0), including 15 % isopropanol and 15 ml of 10% Triton X-100 solution (v/v)
Buffer QC (Wash buffer)	1 M NaCl, 50 mM MOPS (pH 7.0) including 15 % isopropanol
Buffer QF (Elution Buffer)	1.25 M NaCl, 50 mM Tris (pH 8.5) including 15 % isopropanol

Buffers QBT, QC and QF were supplied with the QIAGEN Plasmid Midi Kit and the QIAGEN Blood & Cell Culture DNA Mini Kit (3.1.4). They were used for equilibration (QBT) and washing (QC) of the provided DNA purification columns as well as for the elution of purified DNA (QF). The procedure was carried out following the manufacturer's protocol.

Buffers for gel electrophoresis:

10 x TBE buffer	0.89 M Tris base, 0.02 M EDTA, 0.87 M H_3BO_3, purified water ad 1000 ml
Modified 50 x TAE buffer (according to Millipore protocol; 3.4.14.2)	2 M Tris-acetate, 50 mM EDTA (pH 8)
6 x Gel loading buffer (Sambrook & Russell, 2001)	Sucrose 40 g, bromophenol blue 0.25 g, purified water ad 100 ml. (400-500 bp)

3.1.13 Software and Databases

Basic Local Alignment Search Tool (BLAST) is provided by the National Center for Biotechnology Information (NCBI) [http://www.ncbi.nlm.nih.gov/pubmed]. This software was applied to analyse obtained nucleotide data for sequence similarities using blastx (nucleotide sequence is translated and compared to the amino acid sequences database). In order to analyse 16S rDNA, the database was searched using blastn (nucleotide query vs. nucleotide databases)

Entrez Nucleotide Database is maintained by the National Center for Biotechnology Information (NCBI) [http://www.ncbi.nlm.nih.gov/gene]. This source is linked to numerous databases, e.g. GenBank, NCBI Reference Sequences (RefSeq), and the Protein Data Bank of the Research Collaboratory for Structural Bioinformatics (RCSB PDB)**.** Therefore, it lends itself for detailed analysis of nucleotide and amino acid sequences and putative anticipation on their possible functions in secondary metabolites biosynthesis.

ClustalW2 sequence analysis tool version 2.1 is provided by the European Bioinformatics Institute (EBI), (EMBL) [http://www.ebi.ac.uk/Tools/clustalw]. This tool was used to generate multiple sequence alignments based on the Nucleotide Sequence Database, part of the European Molecular Biology Laboratory (EMBL).

DNASTAR Lasergene 8: This software was purchased from DNASTAR, Inc. (Madison, U.S.A.). It includes several sequence analysis tools, e.g. SeqMan Pro, GeneQuest, MegAlign, and PrimerSelect. In particular **SeqMan Pro** was extensively used to analyse fosmid assembly files.

CLUSEAN (CLUster SEquence ANalyzer): This software enables fast access to sequence data from established databases like BLAST and HMMER (Weber *et al.*, 2009). It can be used to identify functional domains and conserved motifs in a given nucleotide or amino acid sequence. Furthermore, the software allowed searching whole genome assembly files as well as assembled contigs of the 454 sequencing of *Fischerella ambigua*.

Kodak 1D software version 3.5.4 and **iX Imager software**: These programs were provided together with the respective gel documentation system (Kodak Scientific Imaging Systems and Intas Gel iX imager) and were applied to edit digital photos of electrophoresis gels.

3.2 Cultivation of cyanobacterial organism

3.2.1 Origin of cyanobacterial *Fischerella ambigua*

F. ambigua (Näg.) Gomont, designated strain number 108b, was obtained from the Culture Collection of Algae of the Swiss Federal Institute for Water Resources and Water Pollution Control (EAWAG), Dübendorf, Switzerland. This cyanobacterium was originally isolated from a shallow hollow in Mellingen, Kanton Aargau, Switzerland, in 1965.

3.2.2 Small-scale cultivation for DNA isolation

In order to obtain fresh cyanobacterial cells for DNA extraction, 10-20 300 ml flasks, each containing 100 ml BG-11 medium, were inoculated with 1 ml of a 21-day-old preculture and grown under gentle shaking at 120 rpm and constant cool white illumination at 25 °C. Cyanobacterial cells were harvested in late exponential phase (usually 21 days) for isolation of genomic DNA.

3.2.3 Large-scale cultivation for analytical purposes

F. ambigua was routinely cultivated in a 25 l glass-tube photobioreactor (Planctotec, Type: Pluto) for possible feeding experiments, but cultures were also used for DNA extraction. The photobioreactor was successively filled with BG-11 medium and the culture was grown at 25 °C under constant irradiation with white fluorescent light (Osram lamps L 58W/25) and aeration by sterile filtered ventilation. Biomass was harvested after 3-4 weeks by continuous flow centrifugation using the Heraeus Contifuge Stratos (3.1.1). The obtained cell material was stored at -20 °C while cultivation was continued by replenishing the remaining culture with fresh BG-11 medium.

3.3 Obtaining axenic cyanobacterial cultures

In order to remove heterotrophic contaminants from the cyanobacterial cultures, different attempts have been made using phototaxis, resistence to lysozyme (3.1.2) and antibiotic selection as well as resistence to UV light.

3.3.1 Phototaxis experiments

Gliding cyanobacteria like *F. ambigua* show phototactic migration in response to light (Moon *et al.*, 2004). This behaviour was expected to allow a light-dependent separation from an associated heterotrophic *P. stutzeri* strain present in the cyanobacterial cultures. One half of an agar plate containing solid BG-11 medium was wrapped up with alu foil. At the shielded half of the plate, a spheric cyanobacterial colony was placed. The illuminant was adjusted to shine on the second half of the agar plate while being locatable for the cyanobacterial cells. Incubation was then performed for 2 weeks.

3.3.2 UV treatment of cyanobacterial colonies

Due to their production of photoprotective UV-absorbers (Ehling-Schulz *et al.*, 1997), *F. ambigua* was also tried to be separated from the *P. stutzeri* strain using well-dosed UV light. For this purpose, a cyanobacterial colony was dispersed on a petri dish that should consist of translucent quartz glass. The petri dish is completely surrounded by alu foil sheets for effective light reflexion leaving an aperture at the top for the light source. The whole procedure was carried out within a laminar air flow clean bench [table 3.1.1-2] using an integrated UV lamp. The culture was exposed to UV light for 30-60 min.

3.3.3 Cyanobacterial resistence to lysozyme

During DNA extraction experiments (3.4.2), it was observed that *F. ambigua* showed resistence to the sole exposition to lysozyme (3.1.2) in a final concentration of 1mg/ml. Based on this observation, it was tried to cultivate the cyanobacterial cells in BG-11 medium containing lysozyme at 100 µg/ml and 1 mg/ml. The cultures were grown as described in 3.2.2.

3.3.4 Antibiotic selection with ampicillin

Due to their slow growth rate, i.e. cell division rate, cyanobacteria can be treated overnight with cell wall antibiotics (β-Lactam antibiotics) without being significantly affected. By contrary, accompanying strains like *Pseudomonas* sp. have a high growth rate when cultured overnight at 37 °C in nutrient media. One spheric cyanobacterial colony was placed in a 1.5 ml Eppendorf cap and dispersed in 500 µl

steam sterilised Milli-Q® water (table 3.1.1-2). The cell suspension was then added to a glass tube containing 5 ml liquid LB medium supplemented with 500 µg/ml ampicillin. The culture was incubated overnight at 37 °C while shaking at 180 rpm. The next day, the cyanobacterial cells were recovered by centrifugation at 13,000 rpm for 5 min, resuspended in 100 ml BG-11 medium and grown at 25 °C with shaking at 120 rpm.

3.4 Molecular biological methods

Molecular biological methods were conducted with the aim to identify halogenase nucleotide sequences and flanking genes, which may be involved in the biosynthesis of ambigols and tjipanazole D.

3.4.1 Sterilization of solutions and equipment

For elimination of potential foreign organisms, i.e. contaminations, all solutions, buffers and media that were not heat sensitive, were autoclaved at 121 °C and 2 bar for 20 min in a Varioklav® steam steriliser (table 3.1.1-2). Solutions of heat sensitive components were sterilised by filtration through 0.22 µm membrane filters (table 3.1.1-2).

Steam heat sterilization was also applied for decontamination of glass and plastic ware as well as for inactivation of genetically modified organisms.

3.4.2 Isolation of chromosomal DNA from *F. ambigua*

For DNA extraction from *F. ambigua*, two different protocols were tested for their effectiveness. Cell material was obtained from both, fresh grow cultures and from deep-frozen cell pellets. All incubation steps mentioned below were performed under constant gentle shaking. Centrifugation of probes up to 2 ml was carried out using the Heraeus Biofuge *fresco* or the Eppendorf Centrifuge 5415 D. For centrifugation of bigger volumes up to 50 ml, the Heraeus Contifuge Stratos was utilised [table 3.1.1-2].

3.4.2.1 DNA extraction with CTAB containing extraction buffer (Method A)

Buffers used in this protocol were prepared according to Fiore *et al*. (2000). 1-2 g of filamentous conglomerated cyanobacterial cells were pelleted (8,500 rpm for 5 min) and resuspended in 5 ml wash buffer (3.1.12) in a 50 ml collection tube. In order to rupture aggregated colonies and seperate filaments, the cell suspension was exposed to sonic shock for 10 min. After centrifugation at 8,500 rpm for again 5 min, the washing procedure was repeated twice. Cell pellets were then resuspended in 5 ml of resuspension buffer (3.1.12). 15 ml CTAB extraction buffer (3.1.12) were added and the tube was gently inverted for several times. The suspension was incubated at 55°C in a water bath for 1 hour and mixed every 10 min. After cooling down to room temperature, extraction with an equal volume of chloroform:isoamylalcohol (24:1) was performed several times until the interface layer was clear. Subsequently, centrifugation was carried out at 12,000 rpm for 30 min at 25°C to allow phase separation. The aquos phase (ca. 20 ml) was splitted into two 50 ml tubes followed by addition of 0.1 volumes (1 ml) Natrium acetate and 2.5 volumes (25 ml) of ice-cold ethanol. Subsequently, the tubes were centrifuged at 8,500 rpm for 1 h at 4°C or at room temperature to precipitate the nucleic acids. The supernatant was discarded and the derived pellets were washed twice with ethanol 70%. Air dried pellets were resuspended in 100 µl ultra purified water or TE buffer (3.1.12).

3.4.2.2 DNA extraction using SDS containing extraction buffer (Method B)

1 g of cyanobacterial cells were harvested by centrifugation (8,500 rpm, 10 min) after 2-3 weeks of growth at 25 °C, vigorous shaking at 120 rpm and constant illumination. The pelleted cells were washed three times with sterile, ultra purified water (Milli-Q® water) (3.1.10) and centrifuged at 8,500 rpm for 5 min. Cells were resuspended in 20 ml SET buffer (3.1.12) and the suspension was decanted into a Potter homogenisator (3.1.1) for filament breakage and cell separation. In order to remove or at least reduce associated organisms, the homogenised and washed cell suspension was treated with lysozyme (3.1.2) in a final concentration of 1mg/ml and incubation for 30 min at 55 °C prior to DNA extraction. Cyanobacterial cells were then washed and vortexed three times with Milli-Q® water (3.1.10) including necessary centrifugation steps, each performed for 5 min at 13,000 rpm. This step was integrated to remove released DNA of associated contaminants. The washed cyanobacterial cells were pelleted by centrifugation at 13,000 rpm for 2 min and resuspended in 20 ml SET

buffer. The following components were added in the given final concentrations: SDS 0,5 %, Proteinase K 500 µg/ml and lysozyme 2,5 mg/ml. The cell suspension was incubated at 55 °C for 2 h. Subsequently, an extraction with one volume of Phenol:Chloroform:Isoamyl alcohol (PCI; 25:24:1) was performed in a 50 ml tube by centrifugation at 8,500 rpm for 30 min. This step was repeated until no precipitated proteins could be found between the upper aqueous and the lower organic phase. The supernatant was then pipetted in a new 50 ml tube and 2.5 volumes of ice cold ethanol and 0.1 volume of 3 M sodium acetate were added, followed by a centrifugation step at 8,500 rpm for 1 hour at 4 °C. The derived pellet was washed twice with ethanol 70%. Air dried pellets were resuspended in 100 µl ultra purified water or TE buffer (3.1.12). Alternatively, the aqeous supernatant from the PCI extraction was purified using the QIAGEN Genomic-tips, which were supplied with QIAGEN Blood & Cell Culture DNA Mini Kit (3.1.4). In this case, the DNA solution was purified based on the manufacturer's instruction manual.

3.4.3 Isolation of genomic DNA from the associated *Pseudomonas* sp.

Aerobic heterotrophic bacteria were isolated by inoculation of 3 ml LB medium with a cyanobacterial spheric colony followed by incubation at 37 °C and shaking at 160 rpm overnight under light exclusion. 2 ml of the supernatant containing cultured heterotrophs were pipetted into a sterile 2 ml Eppendorf cap and centrifuged for 5 min at 13,000 rpm. The cell pellet was resuspended in 5 ml of SET-buffer, followed by addition of 500 µl SDS-solution (10 %), 275 µl Proteinase K solution (20mg/ml) and 500 µl lysozyme solution (40mg/ml) (Sambrook & Russell, 2001). The mixture was incubated at 55 °C for 30 min. Subsequently, an extraction with one volume of PCI (25:24:1) was performed in a 15 ml tube by centrifugation at 8,500 rpm for 30 min. This step was repeated until no precipitated proteins could be found between the upper aqueous and the lower organic phase. The genomic DNA was precipitated from the supernatant according to 3.4.4. RNase (3.1.2) was added in a final concentration of 20 µg/ml, to remove RNA from the DNA solution.

3.4.4 DNA precipitation

As a first step to recover nucleic acids from aqueous solutions, the ion concentration was adjusted by addition of 0.1 volume of 3 M sodium acetate (pH 5.2). For

precipitation of nucleic acids, two volumes of ice cold ethanol (99,8% p.a.) were added prior to centrifugation, which was performed at 4 °C or at room temperature for 30 min. The supernatant was discarded and the derived pellet was washed twice with 70 % ethanol, to remove potential remains of salts. After another centrifugation step at 4 °C or at room temperature for 15 min, the pellet was air dried at 37 °C until all ethanol residues were evaporated. Rehydration of the nucleic acids was accomplished in TE-buffer pH 8 or Milli-Q® water (3.1.10; 3.1.12)

3.4.5 Determination of nucleic acid concentration and purity of DNA

To quantify obtained nucleic acids and determine the quality of DNA solutions, the OD was measured at the wavelengths 260 nm (A_{260}) and 280 nm (A_{280}) using a UV/VIS spectrophotometer (table 3.1.1-2). The OD value at 260 nm (A_{260}) refers to the amount of DNA or RNA in the solutions to be quantified. The Lambert-Beer law describes the linear relation between the concentration of an absorbing material and its light absorption at a given wavelength. This linear correlation is only valid for absorption values between 0.1 and 1.5. Therefore, highly concentrated DNA solutions had to be diluted to an appropriate measurable concentration. The true concentration of the original nucleotide solution was then extrapolated by including the dilution factor (DF) in calculations. The absorption coefficient is a material-dependent parameter that describes the specific absorption of the respective material. For solutions containing double stranded DNA, an average extinction coefficient of 50 $(\mu g/ml)^{-1}\,cm^{-1}$ was applied (Sambrook & Russell, 2001). The concentration of nucleic acids was calculated according to the following equation:

$$conc._{nucleic\ acid} = DF \times A_{260} \times 50\ \mu g/ml$$

Aromatic amino acids of proteins show significant absorption at 280 nm and therefore the ratio of A_{260}/A_{280} is a common value to specify the purity of a DNA solution. A ratio value of 2.00 indicates 100 % nucleic acids and 0 % proteins. A_{260}/A_{280} values between 1.8 and 2 are favorable for DNA used in downstream applications.

3.4.6 16S rDNA analysis

A standard PCR reaction (3.4.9.2) was performed on isolated DNA applying the two primers Cyan-16S-fw and Cyan-16S-rev (3.1.9) (Fiore *et al.*, 2000). The obtained

PCR product at approximately 1,500 bp was excised from an agarose gel and purified using the QIAquick Gel Extraction Kit (3.1.4). Subsequently, the purified amplificate was ligated into the pGEMT®-vector (3.4.9.3) for sequencing with the T7-primer (3.4.9.5).

3.4.7 Agarose gel electrophoresis

Agarose gel electrophoresis was used to survey DNA manipulations such as preparative DNA isolation, restriction digests or PCR amplification. Standard gels were prepared by dissolving 1 % peqGOLD Agarose in 1 x TBE buffer [table 3.1.1-1] (3.1.12). Small gels were usually run in a Life Technologies Horizon® 58 chamber [table 3.1.1-2] for 40 min at a voltage of 120 V. Nucleic acid probes were mixed with gel loading dye before they were filled into the slots. DNA standards were applied to each gel for size estimation. To reduce ethidium bromide contamination in the laboratory, gels were stained subsequently to the separation run by immersing them in 100 ml ethidium bromide solution (10 mg/ml) for 5-10 min. This was followed by a washing step in water for another 5-10 min. The ethidium bromide solution was saved to be reused several times. Nucleic acids with intercalated ethidium bromide could be visualised by UV illumintion at 254 nm. Gels were digitally pictured using the Kodak DC 290 Zoom Digital Camera System, or alternatively the Intas Gel iX Imager [table 3.1.1-2]. Short-wave UV light exposure should be kept to a minimum, since it may cause photochemical damage of the DNA, which could affect downstream applications, in particular DNA sequencing.

3.4.8 Recovery of DNA from agarose gels

For cloning procedures (3.4.9.3), stained PCR fragments of interest were directly cut out from electrophoresis-gels and purified using the QIAquick gel extraction kit following the manufacturer's protocol. Purified PCR amplificates were usually cloned in sequencing vectors (pGEMT® or pCR®2.1-TOPO®) to analyse their nucleotide sequences by BLAST (3.1.13).

3.4.9 Polymerase chain reaction

In contrast to in vivo amplification of DNA inside host cells (3.4.12), for instance by cloning techniques, high yield propagation of specific DNA-sections with a moderate

size is often accomplished in vitro by polymerase chain reaction (PCR). This reaction requires the knowledge of flanking regions of the sequence to be amplified. Oligonucleotide primers with a length of typically 18-22 bp are complementary to these two opposite ends of the desired region. In a repititive manner, denaturation, annealing and elongation are carried out in usually up to 30 cycles. In the denaturation step, double stranded DNA is melted by heating. In this state, the single stranded DNA is ready for primer annealing and nucleic acid elongation, which is mediated by a Taq polymerase (3.1.2). This procedure enables exponential in vitro reproduction of a DNA-section between two primer annealing sites, and therefore is the method of choice for rapid amplification of nucleic acid fragments. Within this work, PCR was performed applying a T3 or a T-gradient thermocycler [table 3.1.1-2] according to the parameters and protocols described in sections 3.4.9.1and 3.4.9.2.

3.4.9.1 PCR parameters

For PCR amplification of template DNA, PCR-buffers supplied with the GoTaq® Flexi DNA Polymerase (3.1.2) were used to achieve optimal reaction conditions. To ensure hybridisation of PCR-primers with their complementary sequence after denaturation, the temperature was adjusted to the calculated annealing temperatures of the primers. Elongation by Taq-polymerase was accomplished at 72 °C, whereby the required elongation time was estimated from the length of the target region (about 1 min/kb). For DNA-extension, the polymerase recognises the primers annealed to the DNA-template and starts elongation in 3’-direction using dNTPs provided in the reaction mixture.

Fine tuning of PCR amplifications is depending on the $MgCl_2$ concentration, which affects polymerase activity and furthermore relies on specific hybridisation of the primers to the DNA-template. Primer annealing occurs with less specifity at lower temperatures, which was beneficial when using homologous, degenerate primers designed from sequence alignments (3.1.13). Optimisation of the procedure was achieved for each PCR by adjusting the annealing temperature as well as the $MgCl_2$ concentration. Self-complementarity of template DNA is usually less likely due to molar excess of oligonucleotides. Despite that, additives like DMSO are useful when working with GC-rich templates avoiding the formation of secondary structures in the polynucleotide molecules (Hung *et al.*, 1990).

3.4.9.2 PCR protocol

A standard PCR reaction mixture was composed as follows:

components		volume	final concentration
5x Tag-Puffer ($MgCl_2$ contained)	(7.5 mM)	4.0 µl	1.5 mM
extra $MgCl_2$	(25 mM)	0 / 1.2 / 2.4 µl	1.5 / 3.0 / 4.5 mM
dNTP	(each 10 mM)	0.4 µl	each 0.2 mM
Tag-Polymerase	(5 unit/µl)	0.2 µl	0.05 unit/µl
Forward-Primer	(20 µM)	1.0 µl	1 µM
Reverse-Primer	(20 µM)	1.0 µl	1 µM
DMSO		1.0 µl	0.5 %
Template		1.0 µl	
H_2O		**ad 20.0 µl**	

The typical PCR thermocycling was performed as described below:

		temperature	time
1.	initial denaturation step	95 °C	5 min
2.	primer annealing	45 - 60 °C	0.5 min
3.	elongation	72 °C	1 - 1.5 min
4.	denaturation	95 °C	0.5 min
5.	final primer annealing	45 - 60 °C	0.5 min
6.	final elongation	72 °C	4 min
7.	end	4 °C	hold

Steps 2 to 4 were repeated in a loop for 30 times before proceeding with step 5. Annealing temperatures were estimated from the lower T_m-value (3.1.9) of the two used primers.

3.4.9.3 Ligation of PCR Products

PCR fragments produced by Taq-polymerases had a 3'-poly-A nucleotide overhang. Some customary cloning kits, like the pGEM®-T Vector System I and the TOPO TA

Cloning® kit (3.1.4) provide vectors, which are compatible to PCR products synthesised by Taq polymerases. These cloning kits were utilised to directly ligate PCR fragments extracted from electrophoresis gels. A vector to insert ratio of 3:1 was favored in a total volume of 10 µl. Usually, an incubation time of 2 h at room temperature was sufficient before proceeding with the transformation step. The following general pipetting scheme was applied for standard ligations:

Component	pGEM®-T Vector System I	TOPO TA Cloning®
Buffer	5 µl (2 x Rapid Ligation Buffer, T4 DNA Ligase)	1 µl Salt Solution
Vector	1 µl (pGEM®-T)	1 µl pCR®2.1-TOPO®
PCR product	3 µl	0.5-3 µl
Ligase	1 µl T4 DNA Ligase	1 µl T4 DNA Ligase
Nuclease-free water	ad 10 µl	ad 6 µl

3.4.9.4 PCR with sequencing primers

To check the plasmids for correct integration of inserts, PCR reactions (3.4.9.2) were performed using the appropriate sequencing primers. The M13-fw/M13-rev primer pair was used for the pCR®2.1-TOPO® vector. For pGEM®-T vector, the T7/SP6 combination was applied.

3.4.9.5 Sequencing of vector-insert constructs

To characterise and verify the gene fragments inserted in PCR vectors or fosmids (3.1.9), the nucleic acid sequence had to be determined. Sequencing was performed by GATC Biotech AG (Konstanz, Germany) according to the chain termination method invented by Sanger *et al*. (1977). In this PCR based procedure, a single primer anneals to a single stranded DNA template next to the sequence of interest. Either standard vector sequence-specific primers (3.4.9.4) were used, or alternatively oligonucleotides designed from the insert sequence were applied for sequencing. Fluorescently labeled chain-terminating nucleotides (ddNTPs), which are included in

the PCR reaction, lack a 3'-OH group, which is necessary for phosphodiester bond formation, i.e. for chain elongation. Through occasional incorporation of these nucleotides, chain elongation is terminated. During the thermocycling procedure, PCR amplificates of any length are produced and are separable by capillary gel electrophoresis. Thereby, each fragment that reaches the detector has been extended by exactly one nucleotide (Metzker, 2010; Sanger *et al.*, 1977). PCR probes for sequencing were usually delivered as cloned fragments in suitable plasmid vectors (3.4.9.3). For end sequencing of fosmid E8, the T7 primer served as a forward primer to sequence in 5’3’-direction, whereas Epi-RP was applied as a reverse primer for sequencing in 3’5’-direction.

3.4.10 Restriction digestion

Endonucleases were applied during this study to achieve specific fragmentation of double stranded nucleic acids. These enzymes originally belong to the bacterial repertoire of defense mechanisms. They recognise certain DNA sequences, which are mostly palindromes enabling a cleavage of phosphodiester bonds at these specific restriction sites. Because of their palindromic nature, cleavage sites may have an increased chance of reassociation, which is an important advantage for cloning procedures, i.e. ligation of inserts into vectors, because different orientations for the reannealing of two restriction fragments are possible. Most of the restriction endonucleases employed in this work produce sticky ends, while *SmaI* for example generates blunt ends. However, ligation reactions are more reliable when carried out with sticky ended restriction fragments.

For optimal conditions, the buffers provided with the restriction enzymes were included in the reaction mixtures. The duration of the restriction digest was depending on the nature and quality of template DNA as well as on the type of restriction enzyme applied. Generally, a reaction mixture included 10 units of enzyme per µg DNA in a total volume of 20 µl and was incubated overnight at 37 °C.

3.4.11 Preparation of cells competent for DNA-transformation

In this work, *E. coli* strain XL-1 Blue (Stratagene, La Jolla, CA, USA) (3.1.5) with the following genotype was used for general cloning:

recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ *proAB lacI*^{q}*ZΔM15* Tn*10* (Tet^{r})].

For the production of chemically competent cells, an overnight preculture was prepared by inoculating 3 ml LB-medium with a single *E. coli* colony followed by incubation at 37 °C with constant shaking at 180 rpm. This preculture was transferred into an Erlenmeyer flask containing 70 ml of 2 x YT-medium. The mixture was again grown at 37 °C and shaking at 180 rpm until an OD_{600} of 0.3-0.4 was reached. At this point the cells were harvested by centrifugation for 10 min at 8,000 rpm and 4 °C (Heraeus Contifuge Stratos) [table 3.1.1-2]. To make the cells transformable, they were treated with ice cold $CaCl_2$/$MgSO_4$-solution (70 mM $CaCl_2$/20 mM $MgSO_4$) twice after resuspension. The first treatment was performed with 10 ml and the second one with 3.5 ml of the $CaCl_2$/$MgSO_4$-solution. Each time, the cell suspension was kept on ice for 30 min. It was possible to store the competent *E. coli* cells deep-frozen for later use without discernible loss of quality. After addition of 875 µl steam sterilised glycerol, the cell suspension was aliquoted to 100 µl probes for storage at -80 °C.

3.4.12 Transformation of host strains

In vivo amplification of DNA sequences was accomplished by introducing the insert-containing vector to competent *E. coli* XL1-Blue cells (3.1.5) for multiplication of the vector construct. An aliquot of competent cells (3.4.11) was thawed on ice and 5 µl of the ligation mixture (3.4.9.3) was added. After incubation on ice for 30 min, the probe was heat-shocked for 1.5 min at 42 °C to enable DNA uptake. In order to ensure that the delivered plasmid DNA was kept inside the *E. coli* transformants, the cell suspension was again placed on ice for 2 min. To initiate cell growth, 600 µl of sterile 2 x YT-medium were added and a following incubation was carried out for 1 h at 37 °C using the Thermomixer Eppendorf [table 3.1.1-2]. Depending on the applied vector, an antibiotic selection was performed by spreading 300 µl of this preculture on LB agar plates, which were supplemented with an appropriate antibiotic. The plates were incubated overnight at 37 °C using an incubator (3.1.1). Subsequently to plasmid isolation (3.4.13), the insert was confirmed by PCR (3.4.9.4).

3.4.13 Plasmid isolation from transformed *E. coli*

The vector construct was isolated from transformed *E. coli* cells (3.4.12) using the GeneJET™ Plasmid Miniprep Kit (3.1.4). For the isolation of plasmid DNA from PCR positive clones, cultures in a scale of 3 ml were sufficient. After inoculation of LB-medium supplemented with an appropriate antibiotic, the culture was grown overnight in a glass tube at 37 °C under constant shaking at 180 rpm. Cells were then harvested by centrifugation at 13,000 rpm for 2 min in 1.5 ml tubes (3.1.1). Subsequently, the pellets were resuspended in buffer 1 (3.1.12) by vigorous vortexing applying the MS2 Minishaker [table 3.1.1-2]. Cell lysis was achieved by addition of buffer 2 (3.1.12), followed by incubation for not more than 2 min while gently inverting the tube several times. After addition of the neutralization solution (buffer 3) (3.1.12), precipitated proteins and non-circular DNA were pelletted by centrifugation at 13,000 rpm for 15 min. The supernatant was purified with supplied columns following the manufacturer's instructions.

3.4.14 Methods for the establishment of a genomic library

3.4.14.1 LMP agarose gel electrophoresis

For preparative gel extraction of high molecular genomic DNA, peqGOLD Low Melt Agarose was applied. According to the manufacturer's manual, gels were prepared with TAE electrophoresis buffer. For separation of 35-40 kb of DNA, a 0.9 % TAE LMP agarose gel was chosen. The 50 x TAE stock solution was diluted to 1 x concentration. 1.8 g of LMP agarose was suspended in 200 ml of 1 x TAE buffer using a magnetic stirrer [table 3.1.1-2]. The suspension was weighed and subsequently heated to ebullition using a microwave until the agarose was completely dissolved. Afterwards, the solution was weighed and boiling losses were supplemented by adding distilled water. The LMP agarose solution was allowed to cool for 20 min to a temperature of about 50 °C. In a Horizon® 11.14 chamber, the LMP gel was casted and cooled at 4 °C for 1 hour to achieve optimal consistence. The nucleic acid solution was mixed with gel loading dye [table 3.1.1-1] and pipetted into the gel slots. For size estimation, the GeneRuler™ DNA Ladder Mix (3.1.3) and the 36 kb Control DNA supplied with the CopyControl™ Fosmid Library Production Kit (3.1.4) was used. In a first turn, the gel was run for 10 min at 100 V to assure that

the DNA was transferred to the gel matrix. Subsequently, the voltage was lowered to 15-30 V and the gel electrophoresis was carried out overnight. For preparative isolation of the desired band, only lanes, which contain marker DNA were stained with ethidium bromide, in order to prevent damages to the genomic DNA by intercalation.

3.4.14.2 DNA extraction from LMP agarose

For purification of DNA from LMP agaose, a modified protocol provided by Millipore (Millipore GmbH, Schwalbuch) allowed the application of protein purification columns (Amicon Ultra Centrifugal Filters 100K) [table 3.1.1-2] for DNA extraction from LMP agarose. Guided by the stained band of the Fosmid Control DNA (3.4.14.1), the desired band of about 36 kb had to be excised from the gel using a clean scalpel. To minimise UV light exposure to the band of interest, the latter was masked with a sheet of alu foil. Cut out gel slices were molten at 70 °C for 15 min and the temperature was subsequently equilibrated to 42 °C for additional 10 min. According to the manufacturer's protocol, 1 unit of Agarase (3.1.2) was added per 100 mg (approximately 100 µl) of molten LMP agarose. After repeated pipetting, the mixture was incubated in a water bath at 42 °C for 30 min. In order to remove undigested gel remnants, ammonium acetate was added to a final concentration of 2.5 M and the solution was chilled on ice for 10 min. This was followed by centrifugation at 12,000 rpm for 10 min. 90 % of the supernatant was pipetted in a new, sterile 2 ml Eppendorf tube for further purification steps. Up to 200 µl of digested agarose were pipetted into an Amicon Ultra Centrifugal Filter with a pore size retaining molecules of about 100kDa. The volume was filled up to 500 µl with 20 % isopropanol. A subsequent centrifugation was performed at 2,500 rpm for 14 min. This step was repeated until the whole DNA solution had passed the filter unit. A washing step using 450 µl of TE or Milli-Q® water was followed by another centrifugation step at 2,500 rpm. Finally, the purified and concentrated DNA solution could be removed from the vial by pipetting.

3.4.14.3 End-repair of size selected DNA

For cloning into the pCC1FOS vector (3.1.6), genomic DNA with a size of 35-40 kb had to undergo an end-repair for generating blunt ends. For this purpose the Quick

Blunting Kit (3.1.4) was used according to the manufacturer's protocol. A standard reaction mixture contained the following components:

Purified DNA (up to 5 µg)	1-19 µl
10 x Blunting Buffer	2.5 µl
1 mM dNTP Mix	2.5 µl
Blunt Enzyme Mix	1.0 µl
Sterile distilled water	ad 25 µl

Since the DNA fragments were sheared during DNA extraction, an incubation time of 30 min at room temperature was chosen following the manufacturer's recommendations. Subsequently, the enzyme mix was inactivated by heating for 10 min at 70 °. The reaction mixture containing the blunt end DNA could be directly used for ligation into the pCC1FOS fosmid vector.

3.4.14.4 Ligation of end-repaired DNA into the pCC1Fos vector

For the construction of a genomic library, the CopyControl™ Fosmid Library Production Kit (3.1.4) was applied with special regard to the manufacturer's instructions. For blunt end ligation, the molar ratio of the cloning vector to the desired insert DNA has to be taken into account, whereby a 10:1 molar ratio of the CopyControl™ pCC1FOS vector (3.1.6) to insert DNA is recommended.

A 10 µl reaction volume was pipetted in a 0.5 ml Eppendorf tube according to the following scheme:

concentrated insert DNA (~ 0.25 mg of approx. 40 Kb DNA)	6 µl
10 x Fast-Link Ligation Buffer	1 µl
10 mM ATP	1 µl
CopyControl pCC1FOS vector (0.5 mg/ml)	1 µl
Fast-Link DNA Ligase	1 µl

The reaction mixture may be scaled-up, if the ligation effectiveness is limited due to low quality of insert DNA. To calculate the number of clones (N) necessary to cover

the whole *F. ambigua* genome by 40 kb fragments with a given probability (P), the following equation was applied:

N = ln (1-P) / ln (1-f)	P= 99.9 %
N = ln (1-0.99) / ln (1-0.0073) = 630	f = 40 kb/5.5 x 10^3 kb = 0.0073
N = ln (1-0.99) / ln (1-0.0053) = 865	f = 40 kb/7.5 x 10^3 kb = 0.0053

In this formula, f is the proportion of the genome present in one clone (Clarke & Carbon, 1976). The size of cyanobacterial genomes can vary from 5.5 to 7.5 Mb (Herdman *et al.*, 1979), and therefore the number of required clones was estimated to lie between 630 and 865 clones. Based on empirical experiences, this value was tripled, and thus 2,600 clones would be necessary to represent the whole *F. ambigua* genome. According to manufacturer's data, a single ligation reaction is imputed to produce 10^3 - 10^6 clones. Therefore, only one reaction mixture was set up for ligation.

The mixture was incubated for 2 hours at room temperature or overnight at 4 °C. Subsequently, the Fast-Link™ DNA Ligase was inactivated for 10 min at 70 °C. The concatemer formation during ligation is essential for the following lambda packaging step into phage coats (3.4.14.6).

3.4.14.5 Preparation of EPI300-T1^R competent *E. coli* cells

As a preparation step, the *E.coli* strain, which was supplied as a glycerol stock solution was plated on LB agar plates, which did not contain an antibiotic. The colonies were grown overnight at 37 °C in an incubator [table 3.1.1-2]. For a preculture, an Erlenmeyer flask was prepared containing 5 ml of sterile 100 mM $MgSO_4$ solution and 45 ml of sterile LB medium. After inoculation with a single colony of the plating bacterial strain, the culture was shaken overnight at 160 rpm and 37 °C. A second Erlenmeyer flask contained 40 ml of autoclaved LB medium supplemented with 5 ml of 100 mM $MgSO_4$ solution and 5 ml of sterile filtrated 2 % maltose solution. The medium was inoculated with 5 ml of the EPI300-T1^R overnight preculture and shaken at 180 rpm at 37 °C until an OD_{600} of 0.8 was reached. The competent cells could be stored for up to three days.

3.4.14.6 Packaging of fosmid clones

Like cosmids, also fosmids contain at least one phage lambda cos site, which is recognised by λ phage particles, and therefore DNA fragments ligated into such fosmids may be packaged in vitro. The infective phage particles were used to deliver linear, concatemerised phage DNA to special *E. coli* cells (EPI300-T1^R Plating Strain; 3.1.5) by transduction. For the packaging reactions, the MaxPlax™ Lambda Packaging Extracts (3.1.8) were applied following the manufacturer's protocol. For one ligation reaction, one tube of the MaxPlax™ Lambda Packaging Extracts was thawed on ice. Immediately, 25 µl were transferred to a 1.5 ml Eppendorf tube, which was subsequently stored at -80 °C until use. The ligation reaction (10 µl) was added to the remaining 25 µl of the MaxPlax™ Packaging Extract, which was kept on ice for the whole time. The solution was then mixed by pipetting without introducing air bubbles, because the latter could interfere with the packaging process. The mixture was held at 30 °C for 1.5 hours in an incubator [table 3.1.1-2]. This incubation step was repeated after thawing and adding the second half of the MaxPlax™ Packaging Extract, which was stored at -80 °C priorly. The mixture was then diluted in Phage Dilution Buffer (PDB) (3.1.12) to a final volume of 1 ml. 25 ml of chloroform were added to prevent bacterial contaminations and the tube was inverted gently several times. The final solution was then stored at 4 °C until use. For long term storage, sterile glycerol was added to a final concentration of 20% to the remaining phage solution, which could then be stored at -80 °C.

To determine an appropriate titer for easier clone picking, a 1:10, 1:20, 1:100 and a 1:10^4 dilution of the packaging reaction were made in PDB. 100 µl of each dilution were mixed with 100 µl of competent EPI300-T1^R cells (3.4.14.5). The mixture was incubated for 30 min at 37 °C to allow the phage particles to infect the *E. coli* cells and to transfer the linear fosmid DNA to the cell interior by transduction. The infected EPI300-T1^R cells were spread on LB agar plates containing the selective marker chloramphenicol in a final concentration of 12.5 mg/ml. The plates were incubated at 37 °C overnight. Inside the *E. coli* cells, the cohesive ends (cos sites) enable the circulization of the linear DNA to form a plasmid bearing a 40 kb fragment of the genomic *F. ambigua* DNA. In contrast to cosmids, the pCC1FOS vector does not have the drawback of partial deletion of the insert DNA, because it contains a single copy origin of replication based on the F plasmid. A second origin of replication *ori*V

requires induction of desired clones to high copy number to obtain high yield of DNA. This procedure is described in 3.4.14.9.

3.4.14.7 Plating the genomic library

For each dilution prepared in 3.4.14.6, the colonies per LB plate were counted to determine the appropriate amount of host cell suspension for plating. The 1:20 dilution yielded about 200 colony forming units (cfu) per plate, and therefore this dilution was chosen to plate the fosmid library. Altogether, 2,800 colonies were cultivated on LB plates containing chloramphenicol (12.5 µg/ml). 29 microtiter plates, each having 98 wells, were prepared to contain 100 µl of LB liquid medium and 12.5 µg/ml chloramphenicol in every well using the Transferpette®-8 multipipette [table 3.1.1-2]. The clones were picked and suspended in the wells of the microtiter plates using tooth picks. Afterwards, the plates were incubated overnight at 37 °C. Copies of the produced cultures were made using the Boekel Replicator [table 3.1.1-2]. For long term storage, an equal volume of 100 % glycerol was added to the bacterial culture and mixed by pipetting applying the Transferpette®-8 [table 3.1.1-2] prior to freezing at -80° C.

3.4.14.8 Screening of the fosmid gene library

The fosmid gene library was screened by whole cell PCR for the presence of halogenase positive clones using specific primers designed from PCR fragments, obtained in a previous work (Wagner, 2008). In order to make the fosmid DNA accessible for the Taq-polymerase, a heating step has to be integrated in the standard PCR program described in 3.4.9.2. This denaturation step at the beginning of the PCR procedure was performed at 97 °C for 15 min to initiate cell lysis (whole cell PCR). The primer pairs used to identify phenolic halogenase genes were C-Hal fw/C-Hal rev and Ckons-fw/Ckons-rev (3.1.9). Screening for tryptophan halogenases was performed applying the primers T-Hal fw/T-Hal rev and T-Neu-fw/T-Neu-rev (3.1.9). To facilitate fast screening of the library, all 98 clones of each microtiter plate were pooled in 0.5 ml Eppendorf tubes. Standard PCR reactions (3.4.9) were set and performed with 1 µl of each pool at the appropriate annealing temperature including an initial heating step for cell breakage as described above. PCR probes were subsequently analysed by agarose gel electrophoresis (3.4.7). Localisation of single

clones in positive pools was achieved by splitting the latter into 8 x 12 clone pools (8 pools each containing the clones of one line of the microtiter plate). Subsequently, the right clone was localised in one of the 8 pools by PCR of the 12 single clones (one line).

3.4.14.9 Induction of identified positive clones

For the isolation of great amounts of fosmid DNA from PCR positive clones, it was necessary to induce a promoter (*ori*V) by addition of the CopyControl™ induction solution, which was supplied with the CopyControl™ Fosmid Library Production Kit (3.1.4). This induction step initiates expression of the *trfA* gene product, which enables the amplification of fosmid clones at high copy number. A preculture was set by supplementing 5 ml of LB medium with 12.5 mg/ml chloramphenicol in a 15 ml tube. After inoculation with a single fosmid clone, the culture was shaken overnight at 180 rpm and 37 °C. The next day, 45 ml of fresh, sterile LB medium containing 12.5 mg/ml chloramphenicol were mixed with 5 ml of the overnight culture. 50 µl of the 1000 x CopyControl™ induction solution (Epicentre, Madison, U.S.A.) were added to the mixture, which was then shaken under sufficient aeration at 37 °C and 180 rpm for 5 hours. Fosmid isolation was accomplished by applying the QIAgen Plasmid Midi Kit following the manufacturer's instruction manual (3.1.4).

3.4.14.10 Subcloning procedure for fosmid E8

Fosmid DNA was fragmented by ultrasonic treatment. Sheared DNA was size selected (range 500 bp to 2 kb) and reisolated by gel extraction using the QIAquick® Gel Extraction Kit (3.1.4) as described in the QIAquick® Spin Handbook supplied by the manufacturer. After elution of the size selected and purified DNA with Milli-Q® water (3.1.10), an end-repair reaction was set to obtain blunt end DNA fragments (3.4.14.3).

The mixture was incubated at room temperature for 1 hour and the reaction was stopped by heating at 70 °C for 20 min. The end-repaired DNA was purified with the QIAquick PCR Purification Kit (3.1.4) for ligation into the plasmid cloning vector. The applied pGEM®-T Vector System I (3.1.4) contains a 3'-T overhang that fit the 3'-A overhangs of PCR amplicons, generated by Taq-polymerases. To generate DNA-fragments that are suitable for ligation into the pGEM-T® vector, a tailing reaction had

to be performed in order to add an adenine to the PCR products. The reaction mixture contained the following components:

Reaction mix:	30 µl end-repaired template DNA
	10 µl 10 x GoTaq® Flexi Buffer
	1.5 µl GoTaq® Flexi DNA Polymerase
	5 µl $MgCl_2$
	2.5 µl dATP
	1 µl sterile water
	50 µl total reaction volume

The mixture was then incubated for 2 h at 72 °C for optimal polymerase activity. Subsequently, the A-tailed DNA fragments were purified using the QIAquick PCR Purification Kit (3.1.4). A ligation into the T-tailed plasmid vector (pGEM®-T Vector or pCR®2.1-TOPO® (3.1.6) was carried out applying the T4 DNA Ligase (3.1.2) according to the above mentioned protocol (3.4.9.3). Typically, a vector to insert ratio of 1:3 was utilised for these ligations and the reactions were preceded at room temparature preferably overnight.

3.4.15 Genome sequencing using the 454 sequencing procedure

Within this study, results of a 454 sequencing performed for the *F. ambigua* genome were analysed (4.4.4). The sequencing was carried out by GATC Biotech AG (Konstanz, Germany). In this method, the genomic DNA is fragmented, ligated to universal adapter primers and subsequently linked to beads at a dilution that promotes one DNA molecule per bead. These bead-DNA complexes are then encapsulated into emulsion based droplets. By emulsion PCR, the DNA fragments are enriched within these droplets. The beads are then allotted to individual picotiterplate wells, in which a bioluminescence method, referred to as pyrosequencing is performed. In this process, nucleotides are not added coincidently as done in dideoxy sequencing methods (Sanger method and variants thereof) but successively. If a matching nucleotide is added, a cascade of enzymatic reactions

visualises the formation of ATP from pyrophosphate via a light emitting conversion of luciferin into oxyluciferin (Metzker, 2010; Grody *et al.*, 2010; Sanger *et al.*, 1977).

4. Results

Cyanobacteria are a rich source of structurally diverse secondary metabolites. The occurrence of these bacteria in a wide range of environments demonstrates their capability to defend themselves probably by the production of chemical weapons. A filamentous cyanobacterial strain of *Fischerella ambigua* was found to produce the polyhalogenated natural products ambigol A, B, C and tjipanazole D under various culture conditions (Wright *et al.*, 2005; Falch *et al.*, 1995). Although to date numerous polyhalogenated phenolic ethers were found in nature (Gribble, 2010), little is known about their assembly and the involved biosynthetic enzymes.

The current project thus focussed on a genetic approach to elucidate the putative sequence information involved in the production of chlorinated natural secondary metabolites of *F. ambigua*. Preliminary information on several genes that are putatively involved in ambigol biosynthesis was obtained by PCR screening of a fosmid library with primers derived from halogenase genes. Further sequence data were obtained from a 454 genome sequencing effort of *F. ambigua*. Results of these experiments allowed hypothesising on the biosynthesis of the ambigols and also on the indole derivative tjipanazole D.

The sole information on sequence data probably relating to the biosynthesis of a natural product, however does not clarify the biochemical mechanisms and the succession of the catalysed reactions. The hypothesis presented here remains to be proven by gene knockout studies and heterologous expression of putatively involved biosynthesis genes as well as by further biochemical experiments, i.e. in vitro assays.

4.1 Morphological and molecular characteristics of *F. ambigua*

The cyanobacterial strain investigated in this work was obtained from the Culture Collection of Algae of the Swiss Federal Institute for Water Ressources and Water Pollution Control (Dübendorf, Switzerland). Taxonomic studies on this strain have already been done by Wagner (2008) in collaboration with the Culture Collection of Algae and Protozoa, Scottish Association for Marine Science (SAMS, Oban, United Kingdom).

Cyanobacteria are taxonomically divided into five subsections (Garrity *et al.*, 2001; Rippka, 1988; Rippka *et al.*, 1979): subsection I and II include unicellular strains that divide by binary or multiple fission, respectively. Filamentous, non-heterocystous cyanobacteria are classified to subsection III. Subsection IV and V include filamentous strains that form heterocysts in non-branching and branching filaments, respectively. Apart from this morphological classification, recent efforts were made to establish a key map for a molecular identification of cyanobacteria using PCR fingerprinting profiles (Valerio *et al.*, 2009).

Cyanobacteria of the genus *Fischerella* are classified to subsection V according to Rippka *et al*. (1979) and are characterised by true branching and filamentous growth [figure 4.1-2]. Their filamentous nature is mainly due to sheath material formed by exopolysaccharides (Hoiczyk & Hansel, 2000; Hoiczyk & Baumeister, 1995). Early studies on the fine structure of *F. ambigua* revealed that during cell division the middle and outer wall layer as well as the sheath invaginate separating the two daughter cells. This kind of septal division is normally found in most unicellular organisms (Thurston & Ingram, 1971). In fact, it was observed that the *F. ambigua* strain investigated in this work is able to change to unicellular morphology [figure 4.1-1] (Wagner, 2008).

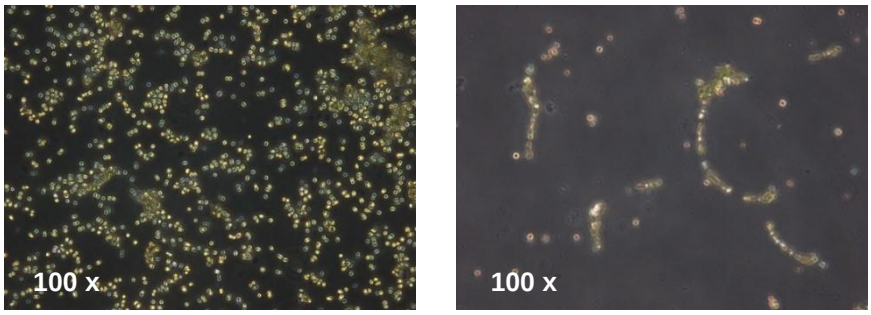


Figure 4.1-1: Unicellular (left) and filamentous (right) growth of *F. ambigua* (Näg.) Gomont (strain number 108b) shown with a magnification of 100 times.

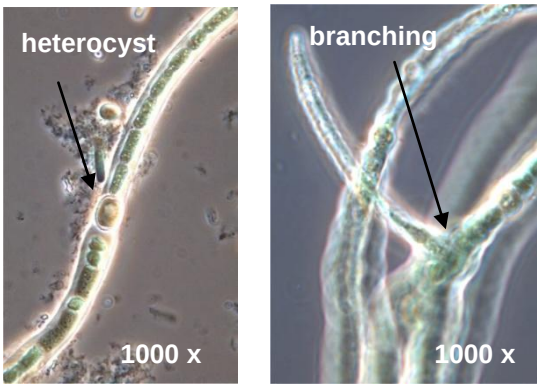


Figure 4.1-2: Phase-contrast micrograph showing heterocysts and true lateral branching of *F.ambigua* (magnification of 1000 times).

In the work of Wagner (2008), 16S rDNA sequences of the present strain revealed high homology with nucleotide sequences from *Fischerella* species available in GenBank. In order to assure that the strain used during the current study is still identical to that employed by Wagner (2008), 16S rDNA sequences were amplified using the primers Cyan-16S-fw and Cyan-16S-rev [table 3.1.9-1]. The amplified PCR product of approximately 1,500 bp was cloned into the pGEMT-vector and transformed into XL1-Blue competent *E. coli* cells (3.4.9.3; 3.4.12). 16S rDNA sequences obtained from six clones (see clones FA16S in table 3.1.6-2) were aligned with those previously found (Wagner, 2008). The compared sequences turned out to be nearly identical.

4.2 Identification and removal of associated bacteria

4.2.1 16S rDNA analysis of associated bacteria

In order to accomplish a 16S analysis (3.4.6) for the identification of the *Fischerella* associated heterotrophic bacterial strain, genomic DNA was isolated according to 3.4.3. Fresh LB medium (3 ml) was inoculated with a spheric cyanobacterial colony and grown overnight at 37°C under light exclusion. Cultured cells were then plated on solid LB medium [figure 4.2.1-1 a] and incubated overnight. The bacterial cells were subsequently grown in liquid LB medium for DNA extraction (3.4.3) [figure 4.2.1-1 b].

a) b)

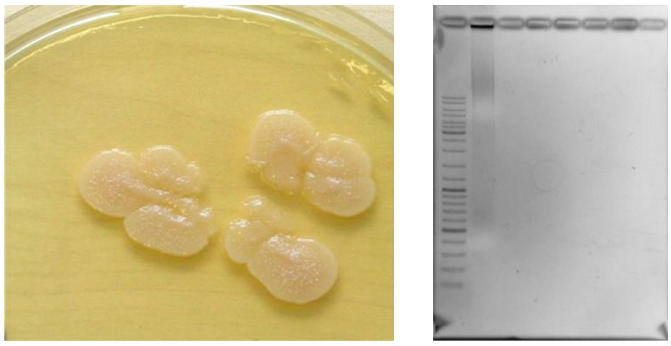


Figure 4.2.1-1: a) LB agar plate with the *Pseudomonas* sp. associated with *F. ambigua*; b) genomic DNA from the isolated *Pseudomonas* strain.

The primers Cyan-16S-fw and Cyan-16S-rev (3.1.9) were used to amplify a 16S rDNA fragment (~1.5 kb) of the unknown strain. The obtained PCR product was cloned in *E. coli* XL-1 Blue cells for sequencing (3.4.9.3; 3.4.12). The sequences contained in seven different clones (see clones Pstu; table 3.1.6-2) were analysed using the Basic Local Alignment Search Tool (3.1.13). They all led to the conclusion that the isolated bacterium is related to a partial 16S rDNA gene of *Pseudomonas stutzeri* strain Gr46 (GenBank accession: FR667910) with 99 % sequence identity.

4.2.2 Axenic cultures

Heterotrophic bacteria are able to grow within the mucilaginous cyanobacterial envelope of filamentous strains like *Fischerella* spp. (Simmons *et al.*, 2008). In this work, the addition of ampicillin in a final concentration of 500 µg/ml and shaking overnight at 37 °C led to axenic cultures (3.3.4). However, recultivation of these pretreated cells in BGII medium revealed a significantly slowed growth when compared to a non-treated culture. The decreased growth rate was judged by the diameter of the spheric colonies [figure 4.2.2-1]. Thus, cultivation of ampicillin-pretreated cyanobacteria turned out to be rather inefficient for isolation of genomic DNA due to poor amounts of available cell material.

Figure 4.2.2-1: On the left, a 300 ml culture flask is shown containing *F. ambigua* cells, which were not pretreated with ampicillin. The right picture shows an Erlenmeyer flask with cyanobacterial cells that were pretreated with ampicillin in a concentration of 500 µg/ml overnight. Both cultures were grown for 2 weeks at 25 °C and shaking at 120 rpm.

Attempts to separate the here investigated *Fischerella* strain from associates by phototaxis-induced gliding (3.3.1) were not successful, because the polysaccharide sheath, which harbours the heterotrophic bacteria is strongly involved in the gliding process (Adams & Duggan, 2008; Adams, 2001; Hoiczyk & Baumeister, 1995).

Exposure to UV light (3.3.2) in order to eliminate contaminants also did not lead to axenic cultures due to light induced production of UV-absorbing substances by the cyanobacterial cells. These compounds, e.g. carotenoids usually serve as light-harvesting pigments (5.2) facilitating photosynthesis and also protect against photooxidative damage (Liang *et al.*, 2006). During cultivation under constant illumination, a permanent release of carotenoids was observed, as judged by the colouring of the culture medium [figure 4.2.2-2]. These and other light-absorbing pigments did probably protect the associated heterotrophic bacteria from UV-light (see 5.2).

Figure 4.2.2-2: 300 ml culture flasks containing cyanobacterial cultures of *F. ambigua* in BG-11 after 3-4 weeks (left), 6 weeks (middle) and 2 months (right).

For the current project it was thus accepted that no suitable axenic culture of *F. ambigua* was available.

4.3 Construction and screening of a fosmid library

4.3.1 DNA extraction from filamentous cyanobacteria

4.3.1.1 Pretreatment of cyanobacterial cells for filament breakage

Problems in DNA extraction of cyanobacteria are well known and often caused by morphological and chemical characteristics, i.e. filamentous growth and polysaccharides, respectively. Furthermore, accompanying heterotrophic bacteria secluded in the thick cyanobacterial polysaccharide envelope may interfere with downstream applications [figure 4.3.1.1-1] (Simmons *et al.*, 2008; De Philippis *et al.*, 2001; Hoiczyk & Baumeister, 1995). The filamentous structure of the bacterial cells required suitable measurements for cell breakage, like sonication or application of a Potter homogenisator (3.4.2), in order to conduct cell lysis to the resulting homogenous cell suspension. For this purpose, the latter was centrifuged, the supernatant discarded and the pelleted cells were washed several times under vigorous vortexing to remove most of the polysaccharide sheath (3.4.2).

a) b)

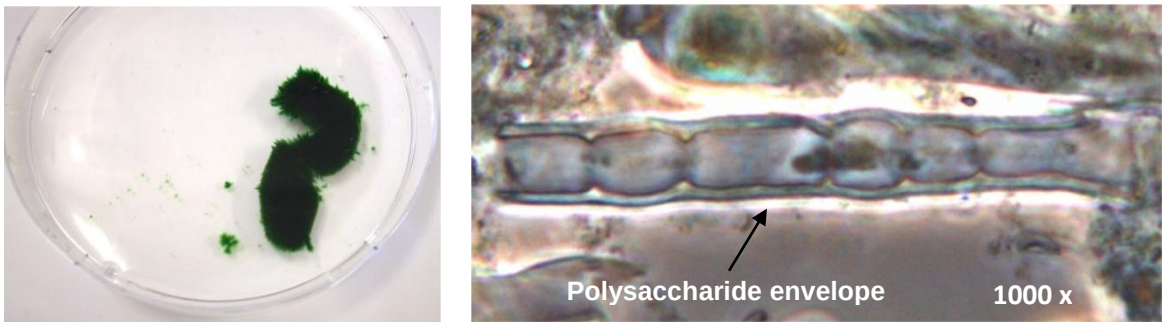


Figure 4.3.1.1-1: a) Petri dish containing a cyanobacterial colony of *F. ambigua* consisting of aggregated filaments; b) Phase-contrast micrograph showing a cyanobacterial filament with the typical thick polysaccharide envelope

4.3.1.2 Removal of heterotrophic bacteria to obtain axenic cyanobacterial DNA

The pretreatment to break cyanobacterial filaments does not remove contaminating and associated bacterial cells, which was tested by inoculation of sterile, liquid nutrient medium (LB medium) with pretreated cells and incubation at 37°C overnight and shaking at 180 rpm. In order to remove the associated *Pseudomonas* cells, the homogenised cell suspension containing both, cyanobacterial and *Pseudomonas* cells, was exposed to lysozyme as described in 3.4.2.2. *Pseudomomas* DNA was then removed by washing and centrifugation steps, which were performed three times with ultrapurified water. A similar method was described for the desert cyanobacterium *Chroococcidiopsis* (Billi *et al.*, 1998).

The applied procedure was suitable, because *F. ambigua* cells were observed to show resistance to the sole application of lysozyme, whereas the contaminants, i.e. *Pseudomonas* cells could be easily cracked under these conditions (3.4.2.2). However, the exposition of *F. ambigua* cells to lysozyme should not exceed 30 min, because after this long time of incubation, a release of a greenish color was observed revealing that some cyanobacterial cells were already lysed.

4.3.1.3 Comparison of two methods for DNA extraction

Two methods were investigated for the effective isolation of DNA. Method A uses a CTAB/sarkosyl buffer for cell breakage (3.4.2.1). In method B, cells lysis is achieved through an SDS-containing buffer together with Proteinase K and lysozyme (3.4.2.1). Especially concentrations higher than 0.4 % SDS were observed to cause effective cell lysis. From 1-2 g cells, an average amount of 9.3 µg almost axenic DNA was obtained with method B [figure 4.3.1.3-1], whereas method A usually gave amounts of 3-5 µg DNA per gram cells. Apart from the inferior yield of genomic DNA, method A does not include the removal of *Pseudomonas* cells, because this contaminating strain was discovered in a later period of this study. Therefore, method B was preferred for DNA isolation from *F. ambigua*.

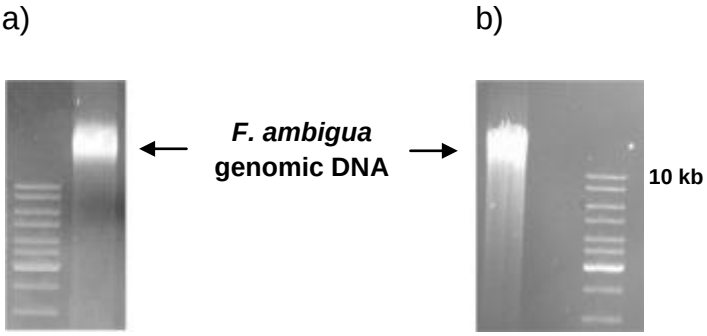


Figure 4.3.1.3-1: Comparison of two different DNA isolation strategies. The figure shows 1 % agarose gels with 1 µl of clear DNA solution in each lane: a) isolated genomic DNA by method A; b) isolated genomic DNA obtained by method B

For the extraction of good quality DNA, it is favorable to extract cyanobacterial cells in an early phase of growth (usually after 21 days). This turned out to be critical for the quality of isolated DNA, because in a later phase of growth, high amounts of polysaccharides were observed that are partially released into the culture medium, which becomes more and more viscous (De Philippis *et al.*, 2001). In this case, these polysaccharides did also increase the viscosity of the resulting DNA solution.

The cell lysate could be directly purified using DNA purification columns (Genomic tips) (3.4.2.2) and led to high amounts of DNA. However, a phenol:chloroform:isoamyl alcohol extraction prior to column use was preferred, in order to reduce proteins that may affect downstream applications, e.g. PCR reactions or ligation into the fosmid vector. The obtained DNA was tested successfully for applicability in PCR techniques (3.4.9).

4.3.2 PCR for phenolic halogenase genes

Using a homology based approach, a previous study on halogenases in *F. ambigua* had successfully obtained a sequence for approximately half of a putative phenolic halogenase (Wagner, 2008). Further design of reverse primers downstream of a conserved motif, i.e. W-X-W-X-I-P (van Pée & Patallo, 2006; 4.7.4), was managed using multiple sequence alignments (3.1.13) of published halogenase sequences [figure 4.3.2-1]. The reverse primer Chal-rev2 (3.1.9) was designed from the W-L-N-A-W-A motif near the C-terminus of the phenolic halogenase Ab7 (4.7.10). It was used together with the forward primer C-Hal fw (3.1.9) from a previous work (Wagner, 2008). A PCR reaction was set aiming to obtain missing genetic information on the *Fischerella* phenolic halogenase gene *ab7* (4.7.10).

McnD	M	P	E	M	L	P	K	V	S	S	N	S	W	L	N	A	W	A	I	S	L	N	593
ApdC	M	P	D	M	L	S	K	V	N	S	N	S	W	L	N	A	W	A	I	S	L	D	593
AerJ	R	P	E	L	L	P	K	I	N	E	H	P	W	L	N	A	W	A	I	S	L	D	593
Ab7	M	P	E	K	S	A	D	F	P	S	P	V	W	L	N	A	W	V	V	S	L	D	530

Figure 4.3.2-1: Multiple sequence alignment used to design degenerate primers from the consensus motif W-L-N-A-W-A. The lowest line shows new sequence information obtained for Ab7. McnD (GenBank: AAZ03553.1: phenolic halogenase from the cyanopeptolin 984 gene cluster; ApdC (GenBank: CAC01605.1): phenolic halogenase involved in the biosynthesis of anabaenopeptilide 90B; AerJ (GenBank: ACM68683.1): tyrosine moiety halogenation in aeruginosin biosynthesis. Ab7 (4.7.10): putative phenolic halogenase from *F. ambigua*. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

PCR conditions for primer annealing were optimised according to manufacturer's recommendations on the melting temperature (3.1.9). For the use of degenerate primers, it is recommended to choose a low annealing temperature to enable restricted unspecific binding to the target sequence, at least 5 °C below the melt temperature. In case of the primer combination C-Hal fw and Chal-rev2 the annealing temperature was adjusted to 43 °C yielding amplificates of different size. Halogenases are very similar concerning their general genetic constitution (Zhu *et al*., 2009; Flecks *et al*., 2008), and thus the expected size of the amplified fragment could be estimated and expected to be approximately 990 bp long.

The obtained PCR product was excised and purified from the agarose gel (3.4.8) for ligation into the pGEM-T® vector (3.4.9.3) and subsequent transformation into *E. coli* (3.4.12). The purified plasmid DNA was sequenced and identified as a fragment of a halogenase gene. The 990 bp long PCR product contained approximately 600 bp of new sequence information, as compared to the partial sequence of Wagner (2008).

4.3.3 Fosmid library production

In order to obtain the complete sequence of biosynthetic gene clusters in *F. ambigua*, the construction of a genomic library was essential (3.4.14). The number of recombinants needed in such a library depends on the genome size of the respective organism and on the size of fosmid inserts (3.4.14.4). *Fischerella* genomes can reach up to 7 Mb (Herdman *et al.*, 1979), and thus it was estimated that 2,800 clones were sufficient for a complete coverage of the genome (Sambrook & Russell, 2001; Clarke & Carbon, 1976).

In fosmids the origin of replication comes from the F plasmid that is known to be present in low copy number. This has the advantage that each clone of the fosmid library keeps a single copy of the inserted DNA fragment. Thereby, degradation or elimination of fosmids during high copy processes is avoided, and thus clone stability is assured. "On-demand" induction of positive clones to high copy number is mediated by a high copy origin of replication, which is also present in the pCC1FOS fosmid vector (3.4.14.9).

Chromosomal DNA, isolated for genomic library construction, was usually sheared during the extraction procedure and revealed a size of approximately 35-40 kb. This was verified by agarose gel analysis (3.4.7) using fosmid control DNA (3.1.3), supplied with the CopyControl™ Fosmid Library Production Kit (3.1.4). DNA fragments smaller than 25 kb are undesirable, because they increase the possibility to be ligated together in one fosmid. This would lead to chimeric clones, which complicate subsequent screening and gene localization processes (see manufacturer’s manual). Small DNA inserts may also not be recognised by phages, and thus would bypass the "head-full" mechanism, in which phages usually segment host DNA into smaller pieces to package them into the phage particle.

Therefore, the sheared DNA was size selected by excising from LMP agarose [figure 4.3.3-1] (3.4.14.2) and subsequently purified through Microcon YM-100 centrifugal filter devices according to a modified protocol provided by the manufacturer (Millipore, Schwalbach). Purification by this procedure is beneficial, because it enables to remove traces of electrophoresis buffer and agarose, in order to obtain highly purified DNA.

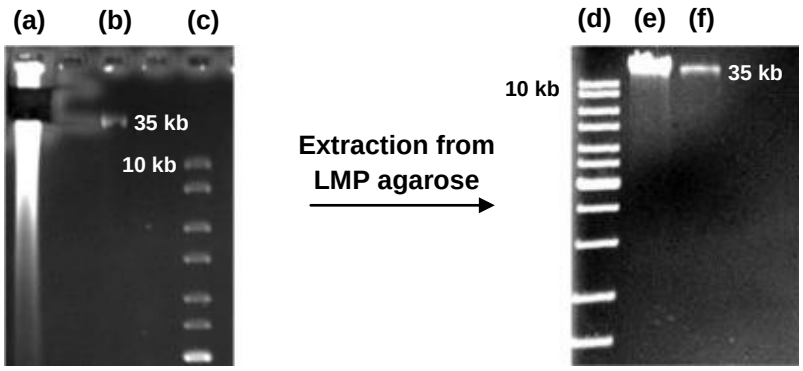
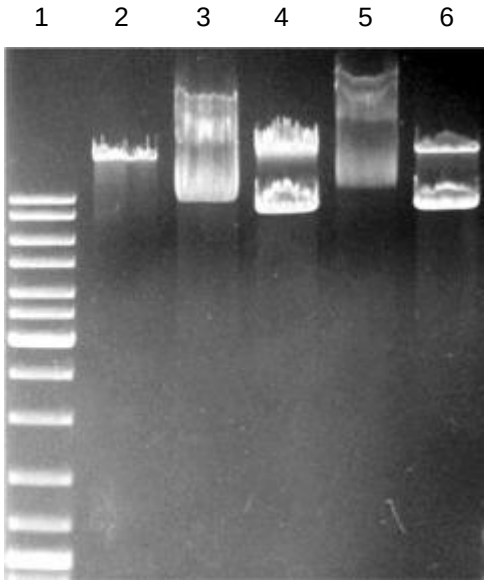


Figure 4.3.3-1: Gel extraction from LMP agarose: 100 µl of DNA solution (a) obtained from 1-2 g of *F. ambigua* cells were size selected using the fosmid control DNA as marker (left picture). The right gel picture shows genomic DNA with an average size of 40 kb that was purified with Microcon YM-100 centrifugal filter devices, see lane (e). Lane (b) and (f): fosmid control DNA (~36 kb). Lanes (c) and (d): GenRuler™ DNA ladder mix (largest band = 10 kb).

Several attempts were necessary to ligate the purified DNA (~90 ng/µl) into the fosmid vector by blunt end ligation (3.4.14.4). The finally successful molar ratio of vector to insert was 9:1 [figure 4.3.3-2].



lane 1: GenRuler™ DNA ladder mix (largest band=10 kb)

lane 2: fosmid control (36 kb)

lane 3: template ligated in pCC1FOS vector

lane 4: negative control: template and pCC1FOS (without FastLink™ DNA ligase)

lane 5: fosmid control ligated in pCC1FOS

lane 6: negative control: fosmid control and pCC1FOS (without FastLink™ DNA ligase)

Figure 4.3.3-2: Result of ligation of *F. ambigua* DNA into the pCC1FOS fosmid vector (see lane 3). The ligation was compared to a positive control applying the fosmid control DNA as insert (see lane 5).

Subsequently, the ligation reaction was packaged into phage heads according to the manufacturer's protocol using MaxPlax™ Lambda Packaging Extracts (3.4.14.6). Before plating the library (3.4.14.7), the titer of the phage particles had to be

determined. For this purpose, several dilutions of the packaged phage particles were prepared (3.4.14.6). Transfection of competent EPI300™-T1R *E. coli* cells was achieved by incubation for 30 min after adding the undiluted phage particles and the prepared dilutions, respectively. Infected *E. coli* cells were spread on LB plates and selected by an antibiotic resistance encoded by a *cat* gene located on the fosmid vector (3.1.7). With the $1:10^4$ dilution, only single clones were found, whereas the $1:10^2$ dilution yielded approximately 30-40 clones per LB agar plate. A 1:20 dilution finally led to about 200 clones per LB plate facilitating clone picking. 2,800 fosmid clones with an average insert length of 40 kb were generated from the *F. ambigua* genome.

Assuming that traces of genomic DNA of the accompanying *Pseudomonas* strain (4.2) were present in the ligation reaction mix, the required number of clones for the coverage of the *F. ambigua* genome would be significantly higher. The size of *P. stutzeri* genomes can reach 4.6 Mb (Ginard *et al.*, 1997), and thus additional clones would have been necessary to assure that every 40 kb DNA-fragment of the *F. ambigua* genome would be present in the genomic library (3.4.14.4).

4.3.4 Screening of the fosmid library for phenolic halogenase genes

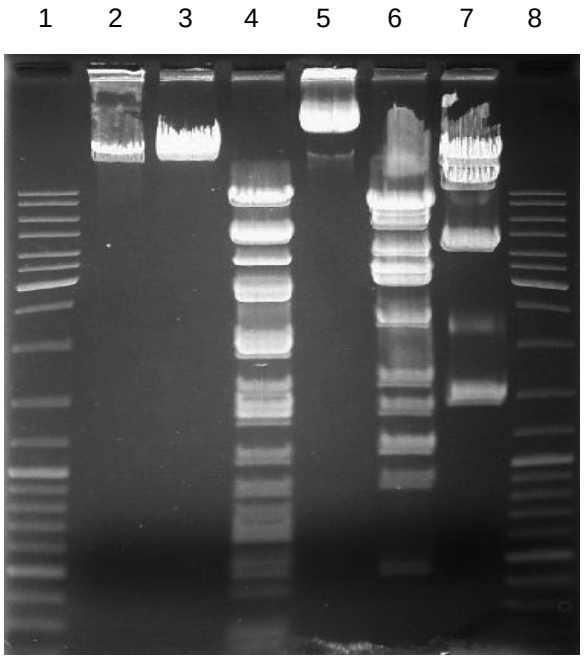
$FADH_2$-dependent halogenases are enzymes involved in regioselective halogenation of aromatic moieties (Vaillancourt *et al*. 2006). They contain highly specific signature motifs, which facilitate their recognition (van Pée & Patallo, 2006). Homologous primers for library screening were designed from sequence data obtained in a previous work by PCR strategies (Wagner, 2008) and from the PCR approach performed in the current study (4.3.2). To enable rapid fosmid library screening, 29 pools were made, each containing 98 clones (3.4.14.8). Using the primer pair C-Hal fw and Chal-rev2 (4.3.2), PCR amplification yielded several unspecific products. In particular, all pools gave a false-positive band when compared to a negative control without template DNA. This band was most likely caused by traces of host DNA (3.4.14.6). Therefore, attempts were made to recognise the right clone by evaluation of band intensity, which was possible. However, this preliminary result had to be assured by a modified, specific PCR. Consequently, several different primers were designed from the known PCR amplificates and tested until the final primer combination, i.e. Ckons-fw and Ckons-rev was found (3.4.14.8). This primer pair was

designed from two conserved motifs, G-S-G-L-A-G and W-V-W-L-I-P and gave a distinct band at approximately 700 bp for fosmid E8, which was localised in pool 15. Sequence analysis of construct top-phal-700 [table 3.1.6-2], which contained the 700 bp amplificate verified the presence of a phenolic halogenase gene on fosmid E8 (3.4.9.5; 3.1.13).

4.4 Sequence analysis of fosmid E 8

4.4.1 High yield fosmid isolation from clone E8 and restriction digestion

For sequencing and subcloning purposes high amounts of fosmid E8 had to be obtained. Therefore, the high copy origin of replication was induced by addition of a special induction solution activating the *trf*A gene product in the EPI300™-T1R *E. coli* strain. This allowed initiation of replication from *ori*V (3.4.14.9) and resulted in a high copy number of fosmid E8 per cell, and thus enhanced the amount of isolated fosmid DNA (3.4.14.9). The latter was analysed by restriction digest to ensure a sufficient insert size for further sequence investigations. Results are shown in figure 4.4.1-1.



lane 1: GeneRuler™ DNA ladder mix (3.1.3; largest band = 10 kb)
lane 2: restriction digest of fosmid E8 with SacII
lane 3: restriction digest of fosmid E8 with ApaI
lane 4: restriction digest of fosmid E8 with HindIII
lane 5: restriction digest of fosmid E8 with XhoI
lane 6: restriction digest of fosmid E8 with XbaI
lane 7: restriction digest of fosmid E8 with SpeI
lane 8: GenRuler™ DNA ladder mix (largest band = 10 kb)

Figure 4.4.1-1: Restriction digest of fosmid E8 using different restriction enzymes.

4.4.2 End sequencing of fosmid E8

End sequencing of fosmid E8 was performed by GATC Biotech AG (3.4.9.5) using two primers, a forward primer that binds to a T7 promoter and a reverse sequencing primer pEpiFOS™ RP-2 (Epi-RP) (3.1.9). Two sequences having 950 bp (T7 primer) and 980 bp (Epi-RP), respectively, were obtained. Their deduced peptide sequences were analysed using BLAST (3.1.13). Table 4.4.2-1 shows their highest identity with proteins listed in GenBank.

Table 4.4.2-1: BLAST search results for end-sequences of fosmid E8 with the vector-specific forward and reverse primers T7 and Epi-RP, respectively. The deduced amino acid sequences were analysed using BLAST. Proteins with highest identity are shown in column two. In column three, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets).

Primer	Enzyme name	Identity of aligned amino acids	GenBank accession number
T7	Acetate kinase [*Nostoc punctiforme* PCC 73102]	218/288 (76%)	YP_001868163
Epi-RP	CYP 450 2C44 [*Mus musculus*]	73/272 (27%)	NP_001161377

Acetate kinase is involved in the formation of acetyl-CoA from acetate via an ATP-dependent phosphorylating step, in which acetate is activated to acetyl phosphate. The following conversion of acetyl phosphate to acetyl-CoA is mediated by phosphotransacetylase (Ingram-Smith *et al.*, 2006a). No coherence to the ambigol biosynthetic gene cluster could be deduced for the acetate kinase, which likely participates in primary metabolism (Campos-Bermudez *et al.*, 2010). Hence, the conclusion was justified that probably one boundary of the cluster was covered by fosmid E8. The second sequence revealed homology to vertebrate CYP 450 enzymes, which represent a class of highly versatile proteins involved in primary metabolism but also in secondary metabolic pathways, for instance in phenolic coupling reactions (Dewick, 2009).

4.4.3 Subcloning of fosmid E8

4.4.3.1 Construction of a shotgun library

In order to characterise fosmid E8 further, a shotgun subcloning was accomplished by sonication of fosmid DNA and recovery of fragmented DNA to be cloned into the pGEMT®-vector (3.4.14.10). A 100 ml culture of the E8 clone was induced to high copy number and the fosmid was isolated as described in 3.4.14.9. Subsequently, the purified fosmid solution was sonicated and the resulting DNA fragments were size selected by gel extraction within a range of 0.5-2 kb (3.4.8) [figure 4.4.3.1-1 a]. To make the fragmented DNA suitable for the pGEMT® PCR vector (3.1.9), a blunt ending and subsequent A-tailing reaction was performed with the DNA fragments (3.4.14.10). The A-tailing reaction is necessary to make the ends of the inserts compatible for ligation into the target PCR vector. The ligation reaction was transformed to *E. coli* XL1 Blue cells (3.1.5) that were plated on ampicillin containing solid LB medium for selection of recombinants (3.4.12). Randomly picked clones were grown overnight and purified using the GeneJET™ Plasmid Miniprep Kit (3.1.4). Several purified clones were checked for inserts by PCR (3.4.9.4) [figure 4.4.3.1-1 b] and sent for sequencing (3.4.9.5). Sequencing results are presented in section 4.4.3.2.

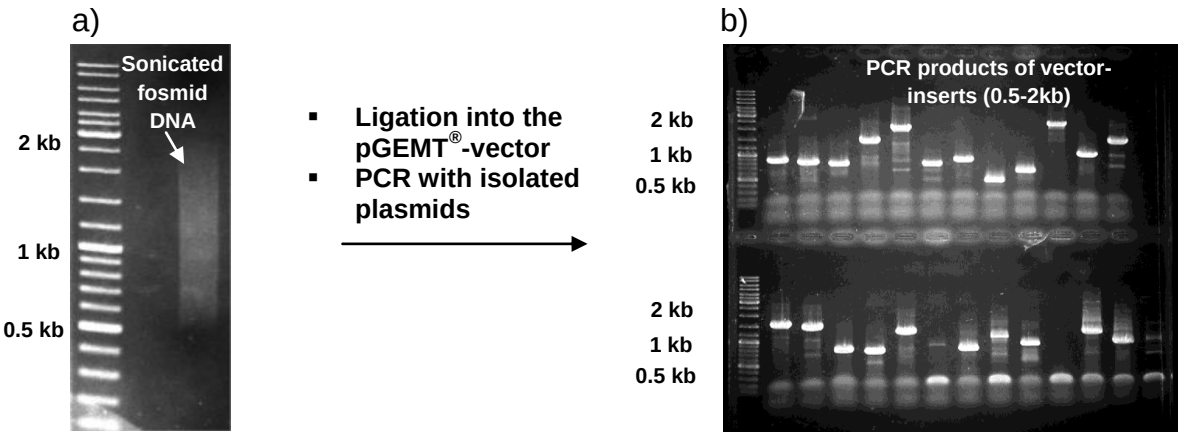


Figure 4.4.3.1-1: Fosmid E8 DNA was fragmented by ultrasonic shock and recovered from an agarose gel in a range from 0.5-2 kb (picture a). Fragmented DNA was then end-repaired and A-tailed for ligation into the pGEMT®-vector (3.4.14.10). Obtained constructs were verified by PCR (picture b) using the two vector sequence-specific primers T7 and SP6 (3.4.9.4). The gel photo on the right side shows PCR amplificates obtained from different plasmids, each containing one fragment of fosmid E8. The first lane of all gels was filled with the GeneRuler™ DNA ladder mix (3.1.3).

4.4.3.2 Sequencing results for E8 subclones

Sequences obtained from subclones generated in this work [table 3.1.6-2] were analysed using BLAST (3.1.13). Search results are shown in table 4.4.3.2-1. One subclone, E8-8 contained a putative halogenase gene with homology to *aerJ* from the aeruginosin biosynthetic pathway (Cadel-Six *et al.*, 2008; Welker & von Doehren, 2006). Adjacent to the putative halogenase gene on subclone E8-8, a sequence probably related to a TE domain was found. Its deduced amino acid sequence shows highest identity to the TE domain of HctF, an NRPS involved in hectochlorin biosynthesis (Ramaswamy *et al.*, 2007). Subclone E8-69 revealed a sequence related to a putative chorismate lyase that catalyses the formation of *p*-hydroxybenzoic acid from chorismic acid by removal of pyruvate (Gallagher *et al.*, 2001). Subclone MEO-6 contained partial sequence data of a gene that probably codifies for a phospho-2-dehydro-3-deoxy-heptonate aldolase. This enzyme is involved in an early step of the shikimate pathway, which is an entry point to the biosynthesis of aromatic secondary metabolites (Knaggs, 2003; Herrmann & Weaver, 1999).

Table 4.4.3.2-1: BLAST search results for several subclones, which were randomly picked from the shotgun subclone library of fosmid E8. The deduced amino acid sequences of E8 subclones (first column) were aligned with homologous proteins. In column four, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Forward primer (fw); Reverse primer (rev).

Subclone (Insert size)	Primer (fw/rev)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
E-8-8 (928 bp)	SP6 (rev)	Halogenase AerJ [*M. aeruginosa*]	121/202 (59%)	CAO82151
E-8-8 (928 bp)	T7 (fw)	Thioesterase domain of HctF [*Lyngbya majuscula*]	36/82 (43%)	AAY42398
E-8-69 (966 bp)	T7 (fw)	Conserved hypothetical protein (chorimate lyase) [*Microcoleus chthonoplastes* PCC 7420]	113/181 (62%)	ZP_05024769
MEO-6 (966 bp)	T7 (fw)	Phospho-2-dehydro-3-deoxy-heptonate aldolase *[Anabaena variabilis* ATCC 29413]	88/133 (66%)	YP_320855

Table 4.4.3.2-1 (continued):

Subclone (Insert size)	Primer (fw/rev)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
E8F4-M13 (1017 bp)	M13 (fw)	Condensation domain-containing protein [*Nostoc punctiforme* PCC 73102]	52/81 (65%)	YP_001866792
E8-15 (999 bp)	T7 (fw)	CYP 450 family 1 subfamily D polypeptide 1 [*Fundulus heteroclitus*]	42/103 (41%)	ACO51073
E8-18 (824 bp)	T7 (fw)	Acetate kinase [*Nostoc* sp. PCC 7120]	106/147 (73%)	NP_486601

The aromatic scaffold of the ambigols and its regioselective chlorination seemed to be consistent with the identification of a putative $FADH_2$-dependent halogenase as well as enzymes associated with the shikimate pathway. Thus, it was concluded that important genetic information on the ambigol biosynthesis was located on fosmid E8.

4.4.4 Complete sequencing of fosmid E8

Complete sequencing of fosmid E8 was performed by GATC-Biotech AG (Konstanz, Germany) through construction of a shotgun library, which was then sequenced by the Sanger method (3.4.9.5) (Sanger *et al.*, 1977). Sequence data previously found during subcloning of E8 (4.4.3) were identified and completed. Additional genes could be recognised upstream of a transposase gene (*orf25*), however the insert of E8 is boardered by an incomplete CYP 450 gene, *ab2*. Table 4.7.2-2 shows all genetic information of E8, i.e. sequences located between the incomplete gene *ab2* and a downstream located transposase gene *orf25*. Approximately 10 kb sequence information belonging to a putative ambigol biosynthetic gene cluster was obtained. Of this, a phenolic halogenase gene with a size of 1.67 kb proved to be identical with the partial sequence information found previously (4.3.2). Further sequences were analysed as a putative CYP 450 enzyme, a chorismate lyase, an AMP-ligase, a DAHP synthetase and an NRPS-like module. Details are presented in section 4.7.2 [table 4.7.2-2].

4.5 Screening for fosmids overlapping with E8

Since the CYP450 gene *ab2* was not completely covered by fosmid E8, the genomic library was screened by PCR (3.4.9.2) for overlapping fosmids applying primers Cyp-fw and Cyp-rev (3.1.9). The latter were designed from the respective boarder sequences of fosmid E8. The annealing temperature was adjusted according to manufacturer's recommendations (Eurofins MWG Operon, Ebersberg, Germany). A large spectrum of unspecific amplificates was observed, however no band was found at the expected size of 524 bp with the primers Cyp-fw and Cyp-rev. Thus, it was concluded that the CYP 450 gene was not completely present in the fosmid library.

4.6 Whole genome sequencing of *F. ambigua*

To obtain missing information on the putative ambigol biosynthetic gene cluster and for further studies on *F. ambigua*, a 454 whole genome sequencing was performed by GATC-Biotech (3.4.15) applying the Roche GS FLX Titanium sequencer (F. Hoffmann-La Roche AG, Basel, Switzerland). To assure that the sequencing will only include *F. ambigua* genomic DNA sequences, the cyanobacterial cells had to be pretreated, in order to remove the accompanying *Pseudomonas* strain DNA (4.2).

An assembly of obtained sequences led to 14280 contigs with an average contig size of 1661 bp. The largest contig assembled had a size of 237 kb. The number of aligned bases was 185,592,864 bp. These would offer at least a 26.5-fold coverage of the *F. ambigua* genome with the imputed size of maximal 7 Mb (4.3.3). Two contigs were identified by searching the genome assembly files for sequences of fosmid E8 and partial sequence information on a putative tryptophan halogenase from the work of Wagner (2008). On contig 00522, genes putatively involved in ambigol biosynthesis were identified (4.7). Contig 15287 harbours a putative tryptophan halogenase and further genes, which may participate in the formation of the indolocarbazole scaffold of tjipanazole D (4.8).

4.7 Elucidation of the putative ambigol biosynthetic gene cluster

4.7.1 Searching the genome assembly for sequences of the E8 fosmid

The genome assembly was searched for contigs containing sequences for the already deciphered phenolic halogenase gene *ab7* (4.4) [table 4.7.2-2]. For the identification of this halogenase, sequences obtained from fosmid E8 and the software CLUSEAN (3.1.13) (Weber *et al.*, 2009) were used. Contig 00522 with a size of 123 kb was identified as bearing all sequence information of fosmid E8 and additional genetic data upstream of the CYP 450-dependent monooxygenase gene *ab2*, which was incomplete on fosmid E8 (4.4.4).

4.7.2 Overall sequence analysis of contig 00522

The sequences were analysed applying BLAST (3.1.13) on contig 00522. Table 4.7.2-2 shows the BLAST search results of identified genes located on contig 00522, some of which are suggested to be involved in ambigol biosynthesis. In table 4.7.2-1, genes found upstream of the putative ambigol biosynthetic gene cluster [table 4.7.2-2] are listed. They are most likely not involved in the biosynthesis of the ambigols but probably in primary metabolism and resistence mechanisms.

Table 4.7.2-1: BLAST search results for genes on contig 00522, which were found upstream of the putative ambigol biosynthetic gene cluster [table 4.7.2-2] and are boardered by a transposase gene (*orf1*). They are likely involved in primary metabolism, resistance mechanisms, or fulfil unknown functions. The deduced amino acid sequences of *orf*s found on contig 00522 (first column) were aligned with homologous proteins by BLAST (3.1.13). In column four, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). In the first column, the orientation of *orf*s identified on contig 00522 is given: > (5',3'-direction); < (3',5'-direction).

Gene		Size (kb)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
orf1	**<**	0.7	Transposase IS4 family protein [*Cyanothece* sp. PCC 8801]	145/187 (77%)	YP_002364719
orf2	**>**	0.58	Conserved hypothetical protein [*Bacteroides* sp. 3_1_23]	43/196 (21%)	ZP_07037929
orf3	**>**	0.2	Partial sequence of ABC transporter gene [*Pseudomonas savastanoi* pv. *savastanoi* NCPPB 3335]	22/63 (34%)	ZP_07007141
orf4	**>**	0.25	Hypothetical protein MAE_10500 [*Microcystis aeruginosa* NIES-843]	45/82 (54%)	YP_001656064
orf5	**>**	0.35	Unnamed protein product [*Microcystis aeruginosa* PCC 7806]	83/120 (69%)	CAO89065
orf6	**>**	0.47	Translation-associated GTPase [*Paenibacillus larvae* ssp. *larvae*]	46/177 (25%)	ZP_02329628
orf7	**>**	0.41	Hypothetical protein MAE_47070 [*Microcystis aeruginosa* NIES-843]	88/137 (64%)	YP_001659721
orf8	**>**	0.9	Unnamed protein product [*Microcystis aeruginosa* PCC 7806]	206/303 (67%)	CAO90918

Table 4.7.2-1 (continued):

Gene	Size (kb)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
orf9 **>**	0.35	Hypothetical protein CwatDRAFT_4464 [*Crocosphaera watsonii* WH 8501]	58/117 (49%)	ZP_00515402
orf10 **>**	0.34	XisI protein-like [*Anabaena variabilis* ATCC 29413]	97/114 (85%)	YP_322782
orf11 **<**	0.98	Nucleotide-diphospho-sugar transferase protein [*Cyanothece* sp. PCC 8801]	172/277 (62%)	YP_002373870
orf12 **<**	0.93	Probable glycosyl transferase [*Nodularia spumigena* CCY9414]	216/311 (69%)	ZP_01629594
orf13 **<**	1.2	Glycosyl transferase, group 1 [*Nostoc punctiforme* PCC 73102]	309/406 (76%)	YP_001866301
orf14 **<**	2.24	ABC transporter related [*Nostoc punctiforme* PCC 73102]	549/774 (70%)	YP_001866302
orf15 **<**	1.16	Glycosyl transferase, group 1 [*Nostoc punctiforme* PCC 73102]	280/389 (71%)	YP_001870304
orf16 **<**	0.23	Hypothetical protein (glycosyltransferase family A; GT-A) [*Nostoc punctiforme* PCC 73102]	93/116 (80%)	YP_001866304
orf17 **<**	0.92	Hypothetical protein (glycosyl transferase family 11) [*Oscillatoria* sp. PCC 6506]	188/307 (61%)	CBN58369

Table 4.7.2-1 (continued):

Gene	Size (kb)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
orf18 >	0.77	Beta-lactamase-like protein [*Nodularia spumigena* CCY9414]	203/257 (78%)	ZP_01629590
orf19 <	0.15	50S ribosomal protein L9 [*Anabaena variabilis* ATCC 29413]	116/151 (76%)	YP_320850
orf20 <	0.92	Hypothetical protein [*Nostoc* sp. PCC 7120]	239/308 (77%)	NP_485051

In conclusion, the putative ambigol biosynthetic gene cluster is possibly boardered by *orf21* and *orf24* [figure 4.7.2-1] [table 4.7.2-2].

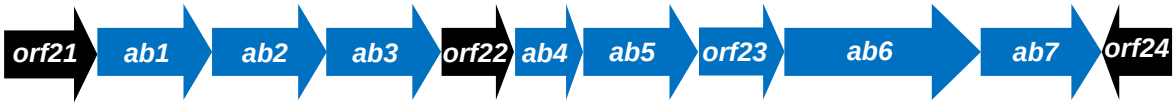


Figure 4.7.2-1: Organisation of genes putatively involved in the biosynthesis of the ambigols. Genes encoding Orfs with possible regulative functions (i.e. *orf21, 22*) as well as *orf24*, which encodes a protein of unknown function are coloured in black. Genes probably related to the biosynthesis of ambigols are shown in blue. A putative DAHP synthetase is encoded by *orf23*. This class of enzymes is typically involved in the output control of shikimate-dependent biosynthetic pathways (Thykaer *et al.*, 2010).

DNA sequences that are probably related to the production of ambigols are illustrated in figure 4.7.2-1. These enzymes are probably a phenolic halogenase (gene *ab1*), two CYP 450 enzymes (genes *ab2* and *ab3*), a chorismate lyase (gene *ab4*), an AMP-ligase (gene *ab5*), an NRPS-like module (gene *ab6*) and a second phenolic halogenase (gene *ab7*) [table 4.7.2-2].

Table 4.7.2-2: BLAST search results for the transposase-boardered assembly of genes on contig 00522, which are putatively involved in ambigol biosynthesis. The deduced amino acid sequences of *orf*s found on contig 00522 (first column) were aligned with homologous proteins by BLAST (3.1.13). In column four, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). In the first column, the orientation of *orf*s identified on contig 00522 is given: > (5',3'-direction); < (3',5'-direction).

Gene	Size (kb)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
orf21 **>**	1.45	Conserved hypothetical protein [*Cyanothece* sp. PCC 7424]	275/466 (59%)	YP_002380103
ab1 **>**	1.45	Phenolic halogenase CrpH [*Nostoc* sp. ATCC 53789]	283/485 (58%)	ABM21576
ab2 **>**	1.48	CYP 1D1 [*Danio rerio*]	168/479 (35%)	NP_001007311
ab3 **>**	1.46	CYP 1C2 [*Oncorhynchus mykiss*]	168/479 (35%)	NP_001171961
orf22 **>**	0.62	Cyclase/dehydrase-like protein [*Nodularia spumigena* CCY9414]; Polyketide_cyc2 superfamily	87/206 (42%)	ZP_01629447
ab4 **>**	0,62	Conserved hypothetical protein (Chorismate lyase superfamily); [*Microcoleus chthonoplastes* PCC 7420]	112/193 (58%)	ZP_05024769

Table 4.7.2-2 (continued):

Gene	Size (kb)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
ab5 **>**	1.53	AMP-dependent synthetase and ligase (Acyl CoA synthetases (AMP-forming)/AMP-acid ligases II) [*Herpetosiphon aurantiacus*]	259/511 (50%)	YP_001546718
orf23 **>**	1.13	3-deoxy-7-phosphoheptulonate synthase (NeuB Superfamily); [*Nodularia spumigena* CCY9414]	233/363 (64%)	ZP_01628888
ab6 **>**	2.63	a) Condensation domain (~900 bp); [*Nostoc punctiforme* PCC 73102]	248/483 (51%)	YP_001866789
		b) NRPS sequence (~510 bp) partially identified as the protein domain TIGR01720; [*Tolypothrix* sp. PCC 7601/1]	74/149 (49%)	CAC60280
		c) PP-binding domain (~192 bp) [*Anabaena variabilis* ATCC 29413]	70/117 (59%)	YP_325330
		d) Thioesterase domain (~771 bp); [*Nostoc punctiforme* PCC 73102]	133/286 (46%)	YP_001869788

Table 4.7.2-2 (continued):

Gene	Size (kb)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
ab7 **>**	1.67	Phenolic halogenase McnD [*Microcystis aeruginosa*]	367/556 (66%)	CAO82181
orf24 **<**	0.23	Conserved hypothetical protein Aazo_2377 ['*Nostoc azollae*' 0708]	56/78 (71%)	YP_003721441
orf25 **<**	1.4	Transposase, IS605 family [*Microcoleus* sp.] 62 %	324/441 (73%)	ZP_01629916

Genetic information identified on fosmid E8 (4.4.4) were identical to those found on contig 00522, whereby the sequence of *ab2* could be completed and the additional halogenase gene *ab1* was found [table 4.7.2-2].

4.7.3 Orf21 shows homology to regulative proteins

Orf21 shows homology to AAA-ATPase-like proteins (i.e. ATPases associated with diverse cellular activities) [table 4.7.3-1], which are known to participate in several cellular processes including the regulation of gene expression (Wilcox & Laney, 2009; Snider *et al.*, 2008; Snider & Houry, 2008; White & Lauring, 2007). Orf21 also resembles WD (tryptophan-aspartate)-repeat containing proteins, which fulfil cellular functions and may be involved in the regulation of gene transcription (Neer *et al.*, 1994).

Table 4.7.3-1: BLAST search results for the deduced peptide sequence of *orf21* found on contig 00522 obtained from 454 sequencing of *F. ambigua*. Homologous proteins are shown in column one. The deduced amino acid sequence of *orf21* was aligned with the respective homologous proteins by BLAST (3.1.13). In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number
Hypothetical protein [*Cyanothece* sp. PCC 7424]	275/466 (59%)	344/466 (73%)	YP_002380103
Hypothetical protein [*Trichodesmium erythraeum* IMS101	250/462 (54%)	333/462 (72%)	YP_723374
Hypothetical protein (AAA-ATPase-like) [*Microcystis aeruginosa* NIES-843]	220/346 (63%)	277/346 (80%)	YP_001657291
Hypothetical protein (WD-repeat containing protein) [*Cyanothece* sp. PCC 7822]	214/343 (62%)	273/343 (79%)	ZP_03155193
GGDEF domain protein (Diguanylate-cyclase) [*Microcoleus chthonoplastes* PCC 7420	204/437 (46%)	287/437 (65%)	ZP_05024949

4.7.4 Ab1, an $FADH_2$-dependent halogenase

The first protein identified in the putative ambigol biosynthetic gene cluster consists of 483 amino acids. A BLAST search performed with the amino acid sequence of Ab1 revealed homology to $FADH_2$-dependent halogenases (1.4.1) [table 4.7.4-1].

Table 4.7.4-1: BLAST search results for the deduced amino acid sequence of *ab1* found on contig 00522, obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
CrpH [*Nostoc* sp. ATCC 53789]	283/485 (58%)	377/485 (77%)	ABM21576 (Magarvey *et al.*, 2006)
Putative tryptophan halogenase [*Burkholderia pseudomallei* 576]	224/484 (46%)	331/484 (68%)	ZP_03453428
CmdE [*Chondromyces crocatus*]	221/498 (44%)	322/498 (64%)	CAJ46693 (Rachid *et al.*, 2006)
ORF8* halogenase [*Actinoplanes teichomyceticus*]	155/414 (37%)	243/414 (58%)	CAG15020 (Li *et al.*, 2004)
RadH [*Chaetomium chiversii*]	148/382 (38%)	217/382 (56%)	ACM42402 (Wang *et al.*, 2008)
Halogenase BhaA [*Amycolatopsis balhimycina*]	145/400 (36%)	234/400 (58%)	CAA76550 (Puk *et al.*, 2002)

The amino acid sequence of Ab1 resembles published $FADH_2$-dependent halogenases involved in the biosynthesis of cryptophycin (CrpH), chondramide B and D (CmdE), teicoplanin (ORF8*), radicicol (RadH) and balhimycin (BhaA) (for literature see table 4.7.4-1).

Figure 4.7.4-1 shows a multiple sequence alignment of the deduced amino acid sequence of gene *ab1* and the first three homologous proteins listed in table 4.7.4-1. Two motifs are highly conserved in $FADH_2$-dependent halogenases (van Pée & Patallo, 2006): the G-X-G-X-X-G, present as G-G-G-P-A-G in Ab1, located in the flavin-binding site near the amino terminal end (van Pée & Zehner, 2003). Near the middle of the sequence, a second absolutely conserved region W-X-W-X-I-P is

present as W-L-W-G-I-P in Ab1 (Dong *et al.*, 2005). In addition, a lysine moiety that is involved in the mechanistic action of the halogenation reaction was recognised at position 75 (Buedenbender *et al.*, 2009; Flecks *et al.*, 2008; Yeh *et al.*, 2007).

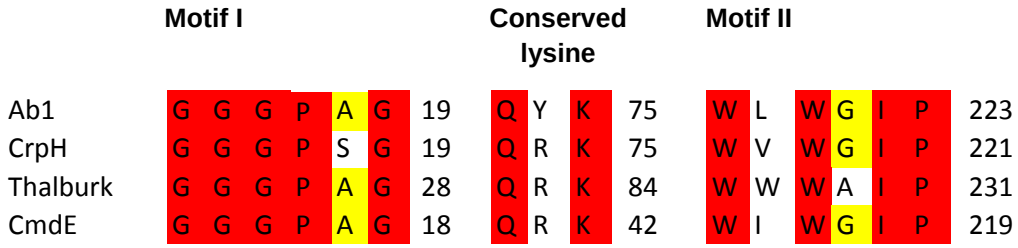


Figure 4.7.4-1: Multiple sequence alignment of Ab1 with putative halogenases from *Nostoc* sp. ATCC 53789 (CrpH), *Burkholderia pseudomallei* 576 (Thalburk) and *Chondromyces crocatus* (CmdE). For GenBank accession numbers, see table 4.7.4-1. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

4.7.5 Two adjacent CYP 450 enzymes

Ab2 and Ab3 were identified in the putative ambigol biosynthetic gene cluster and were analysed by BLAST search (3.1.13). They reveal significant homology to vertebrate CYP 450-dependent monooxygenases. For Ab2, highest similarity to those of subfamilies D and A were determined [table 4.7.5-1].

Table 4.7.5-1: BLAST search results for the deduced peptide sequence of *ab2*, partially found on fosmid E8. It was confirmed and completed by contig 00522, obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
CYP 1D1 [*Danio rerio*]	165/486 (33%)	255/486 (52%)	NP_001007311 (Goldstone *et al.*, 2009)
CYP 1D1 [*Fundulus heteroclitus*]	156/506 (30%)	252/506 (49%)	ACO51073 (Zanette *et al.*, 2009)
CYP 1A1 [*Xenopus laevis*]	156/507 (30%)	255/507 (50%)	NP_001090541 (Laub *et al.*, 2010)

A BLAST search performed for Ab3 showed similar results as found for Ab2, while it possesses highest homology to CYP 450 enzymes of subfamilies C and A [table 4.7.5-2].

Table 4.7.5-2: BLAST search results for the deduced amino acid sequence of *ab3*, found on fosmid E8. It was also confirmed by contig 00522, obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
CYP 1C2 [*Oncorhynchus mykiss*]	168/479 (35%)	259/479 (54%)	NP_001171961 (Jonsson *et al.*, 2010b)
CYP 1A1 [*Xenopus laevis*]	169/486 (34%)	264/486 (54%)	NP_001090541 (Laub *et al.*, 2010)
CYP 1C3 [*Oncorhynchus mykiss*]	168/482 (34%)	257/482 (53%)	NP_001171962 (Jonsson *et al.*, 2010a)

Ab2 and Ab3 harbour the typical three conserved regions found in CYP 450-dependent monooxygenases (1.4.4). These motifs were recognized in a multiple sequence alignment of Ab2 and Ab3, respectively, with similar CYP 450 enzymes [figure 4.7.5-1]. First, an absolutely conserved cysteine ligand in the heme binding motif F-X-X-G-X-X-X-C-X-G of the active center was found at position 433 in Ab2 and at position 437 in Ab3. In the second typical consensus sequence, the E-X-X-R motif is located in helix K, in which glutamic acid and arginine are generally but not absolutely conserved (Rupasinghe *et al.*, 2006). The A-(A,G)-X-(E,D)-T motif of helix I contains a generally conserved threonine residue present in most but not in all CYP 450s (Lewis & Wiseman, 2005). It is presumably involved in the binding and activation of dioxygen. The latter motif (motif I) is not completely conserved in Ab2 and Ab3 revealing a lysine residue instead of threonine.

	Motif I						Motif II				
CYP1D1[*Danio rerio*]	A	G	F	D	T	322	E	V	F	R	378
CYP1A1[*Xenopus laevis*]	A	G	F	D	T	329	E	M	F	R	385
CYP1C2[*Oncorhynchus mykiss*]	A	G	L	D	T	328	E	T	M	R	384
Ab2 [*F. ambigua*]	A	G	A	G	L	296	E	I	F	R	352
Ab3 [*F. ambigua*]	A	G	T	E	L	300	E	V	F	R	356

	Motif III										
CYP1D1[*Danio rerio*]	F	G	M	G	I	R	R	C	L	G	460
CYP1A1[*Xenopus laevis*]	F	G	L	G	K	R	R	C	V	G	468
CYP1C2[*Oncorhynchus mykiss*]	F	S	A	G	K	R	R	C	I	G	466
Ab2 [*F. ambigua*]	F	G	M	G	S	R	R	C	I	G	435
Ab3 [*F. ambigua*]	F	G	I	G	S	R	R	C	I	G	439

Figure 4.7.5-1: Multiple sequence alignment of Ab2 and Ab3 with amino acid sequences of CYP 450-dependent monooxygenases from the zebrafish (*Danio rerio*), the african clawed frog (*Xenopus laevis*) and the rainbow trout (*Oncorhynchus mykiss*; GenBank accession: NP_001171961). For GenBank accession numbers, see table 4.7.5-1 and 4.7.5-2. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to four compared sequences are coloured in yellow. Green shaded amino acids are only identical in Ab2 and Ab3.

Based on the presence of conserved motifs, Ab2 and Ab3 most likely belong to the P450 superfamily (GenBank accession: cl12078). A BLAST search comparing the sequences of genes *ab2* and *ab3* gave an identity of 1181/1397 (84%).

4.7.6 Orf22 belongs to the polyketide_cyc2 superfamily

Orf22 exhibits homology to proteins from the polyketide_cyc2 superfamily that comprises polyketide cyclases/dehydrases (GenBank accession: cl10449) and proteins involved in lipid transport [table 4.7.6-1].

Table 4.7.6-1: BLAST search results for the deduced amino acid sequence of *orf22*, found on fosmid E8. It was also confirmed by contig 00522, obtained from 454 sequencing of *F. ambigua.* In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number
Cyclase/dehydrase-like protein [*Nodularia spumigena* CCY9414]	87/206 (42%)	126/206 (61%)	ZP_01629447
Cyclase/dehydrase [*Nostoc punctiforme* PCC 73102]	78/174 (44%)	112/174 (64%)	YP_001866258
Cyclase/dehydrase [*Anabaena variabilis* ATCC 29413]	85/191 (44%)	116/191 (60%)	YP_324949

4.7.7 Ab4, a putative chorismate lyase

Ab4 belongs to the chorismate lyase superfamily (GenBank accession: cl01230) and shares 58 % identity with a putative 4-hydroxybenzoate (4-HBA) synthetase from *Microcoleus chthonoplastes* PCC 7420 and 55% identity with a putative chorismate lyase from *Beggiatoa* sp. PS [table 4.7.7-1]. The enzyme chorismate lyase is well-known from the ubiquinone pathway catalysing the removal of pyruvate from chorismate (1.4.2) to generate 4-hydroxybenzoic acid. The crystal structure of chorismate lyase from *E. coli* has been solved (Gallagher *et al.*, 2001), and furthermore studies on ligand binding and catalysis have been recently published (Smith *et al.*, 2006).

Table 4.7.7-1: BLAST search results for the deduced peptide sequence of *ab4*, found on fosmid E8. It was also confirmed by contig 00522, obtained from 454 sequencing of *F. ambigua.* In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number
Conserved hypothetical protein (putative 4-hydroxybenzoate synthetase); [*Microcoleus chthonoplastes* PCC 7420]	112/193 (58%)	151/193 (78%)	ZP_05024769
Conserved hypothetical protein (putative chorismate lyase); [*Beggiatoa* sp. PS]	98/177 (55%)	134/177 (75%)	ZP_02000334
Hypothetical protein (putative chorismate lyase); [*Herpetosiphon aurantiacus* ATCC 23779]	73/184 (39%)	118/184 (64%)	YP_001546717
4-Hydroxybenzoate synthetase (chorismate lyase); [*Hahella chejuensis* KCTC 2396]	59/193 (30%)	98/193 (50%)	YP_437233

4.7.8 Ab5, a putative CoA synthetase

The amino acid sequence of Ab5 was analysed by BLAST search and was found to resemble putative AMP-dependent synthetases/ligases of *Herpetosiphon aurantiacus* ATCC23779 with high homology. Detailed results are presented in table 4.7.8-1. AMP-forming acyl CoA synthetases belong to the adenylate-forming enzyme superfamily (1.4.3) (Linne *et al.*, 2007).

Table 4.7.8-1: BLAST search results for the deduced amino acid sequence of *ab5*, found on fosmid E8. It was also confirmed by contig 00522, obtained from 454 sequencing of *F. ambigua.* In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number
AMP-dependent synthetase and ligase [*Herpetosiphon aurantiacus* ATCC 23779]	259/511 (50%)	352/511 (68%)	YP_001546718
AMP-binding enzyme [*Teredinibacter turnerae* T7901]	213/510 (41%)	302/510 (59%)	YP_003075760
Acyl CoA synthetase (AMP-forming)/AMP-acid ligases II [*Hahella chejuensis* KCTC 2396]	184/511 (36%)	284/511 (55%)	YP_437234
Acyl CoA synthetase [*Geobacillus thermodenitrificans* NG80-2]	182/522 (34%)	288/522 (55%)	YP_001124702

A multiple sequence alignment of Ab5 with similar amino acid sequences [table 4.7.8-1] reveals the presence of three signature motifs [figure 4.7.8-1].

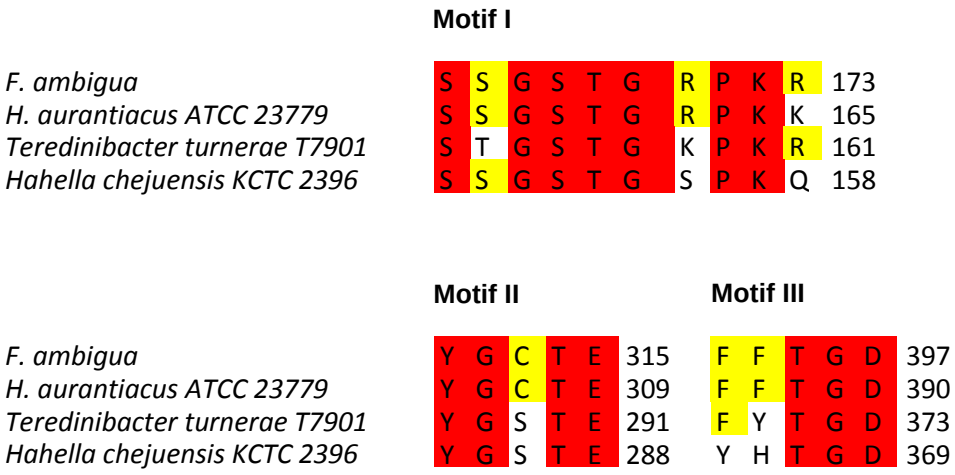


Figure 4.7.8-1: Multiple sequence alignment of Ab5 with amino acid sequences of adenylate-forming enzymes from *Herpetosiphon aurantiacus* ATCC 23779, *Teredinibacter turnerae* T7901, and *Hahella chejuensis* KCTC 2396. For GenBank accession numbers, see table 4.7.8-1. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

The first highly conserved consensus region **T**[SG]-**S**[G]-**G-**[ST]-**T**[SE]-**G**[S]-X-**P**[M]-**K**-**G**[LF] as described by Chang *et al*. (1997) was identified in Ab5 as S-S-G-S-T-G-R-P-K (Chang *et al.*, 1997). Instead of glycine, leucine or phenylalanine (G, L, F) an arginine was recognised as the last amino acid of the motif. Motifs II and III, which are **Y**[LWF]-**G**[SMW]-X-**T**[A]-**E** and **Y**[FL]-**R**[KX]-**T**[SV]-**G**-**D** could also be identified.

Based on crystal structures of firefly luciferase and other members of the superfamily, an involvement of these three motifs in substrate binding and enzymatic turnover is supported (Ingram-Smith *et al.*, 2006b).

4.7.9 Ab6, an NRPS-like module

Adjacent to the AMP-dependent CoA synthetase, an isolated NRPS-like module was found by BLAST search of the deduced amino acid sequence of *ab6*. As a starter NRPS-module, it reveals an unusual composition beginning with a C domain followed by a phophopantheteine (PP) binding site and a TE domain but lacking an A domain for activation and loading of an amino acid [figure 4.7.9-1].

Figure 4.7.9-1: NRPS-like module with relative sizes of its domains found on fosmid E8 and on contig 00522 adjacent to the adenylate-forming enzyme Ab5. Standard starter NRPS-modules begin with an A domain followed by a PCP and C domain (ATC-modules). Ab6 lacks an A domain in front of the C domain. The C domain is followed by an NRPS sequence that was partially identified as TIGR01720 (~ 40 amino acids), which is analysed below. Its function in NRPS-modules is currently unclear (4.7.9.2).

4.7.9.1 C domain of Ab6

The C domain of Ab6 resembles C domains (1.4.5) of the cyanopeptolin synthetase OciA from *Planktothrix rubescens* NIVA-CYA 98, the NRPS NosA and NosD from the nostopeptolide A biosynthetic gene cluster of *Nostoc* sp. GSV224 and CrpD, which is involved in cryprophycin assembly in *Nostoc* sp. ATCC 53789. Table 4.7.9.1-1 shows identity of Ab4 with these proteins.

Table 4.7.9.1-1: BLAST search results for the C domain of Ab6. The gene *ab6* was found on fosmid E8. It was also confirmed by contig 00522, obtained from 454 sequencing of *F. ambigua.* In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
Condensation domain-containing protein [*Nostoc punctiforme* PCC 73102]	232/434 (53%)	305/434 (70%)	YP_001866789
peptide synthetase [*Cyanothece* sp. ATCC 51142]	231/458 (50%)	311/458 (67%)	YP_001804456 (Welsh *et al.*, 2008)
OciA protein [*Planktothrix rubescens* NIVA-CYA 98]	213/417 (51%)	291/417 (69%)	CAQ48254 (Rounge *et al.*, 2009)
NosD [*Nostoc* sp. GSV224]	216/440 (49%)	298/440 (67%)	AAF17281 (Hoffmann *et al.*, 2003)
NosA [*Nostoc* sp. GSV224]	208/418 (49%)	288/418 (68%)	AAF17281 (Hoffmann *et al.*, 2003)
CrpD [*Nostoc* sp. ATCC 53789]	209/416 (50%)	278/416 (66%)	ABM21572 (Magarvey *et al.*, 2006)

The amino acid sequence of the C domain of Ab6 was aligned with related C domains of OciA, NosD and CrpD [figure 4.7.9.1-1] revealing the presence of the typical conserved core motif H-H-X-X-X-D-G [table 4.7.9.1-2]. Additional six signature sequences, which were described by Konz and Marahiel (1999), were recognised with some variations (Konz & Marahiel, 1999). Table 4.7.9.1-2 shows the general formulation of seven canonical core motifs of C domains and the corresponding sequence found in Ab6 of *F. ambigua*.

Table 4.7.9.1-2: Consensus sequences of C domains as described by Konz and Marahiel (1999) in the left column and conserved motifs for Ab6 are listed in the right column. Amino acids that are different in the compared sequences are emphasised by underlined, bold letters.

conserved motif of C-domains	**motif found in *F. ambigua***
C1: S-X-**A**-Q-X-R-(L,M)-(W,Y)-X-L	S-F-**S**-Q-E-R-L-W-F-L
C2: R-H-E-X-L-R-T-X-F	R-H-E-A-L-R-T-T-F
C3: **M**-H-H-X-**I**-**S**-D-**G**-(W,V)-S	**V**-H-H-I-**V**-**I**-D-**F**-W-S
C4: Y-X-D-(F,Y)-A-V-W	Y-A-D-F-A-V-W
C5: (I,V)-G-X-F-V-N-**T**-(Q,L)-(~)-X-R	I-G-Y-F-V-N-**L**-L-I-L-R
C6: (**HN**)-Q-D-(**Y**,**V**)-P-**F**-**E**	**Y**-Q-D-**L**-P-**V**-**Q**
C7: **R**-**D**-**X**-**S**-R-N-P-L	**Y**-**V**-**A**-**P**-R-N-P-L

Multiple sequence alignments with similar proteins found by BLAST search [table 4.7.9.1-1] led to the identification of the seven core motifs C1-C7 in the peptide sequence of Ab6 [figure 4.7.9.1-1].

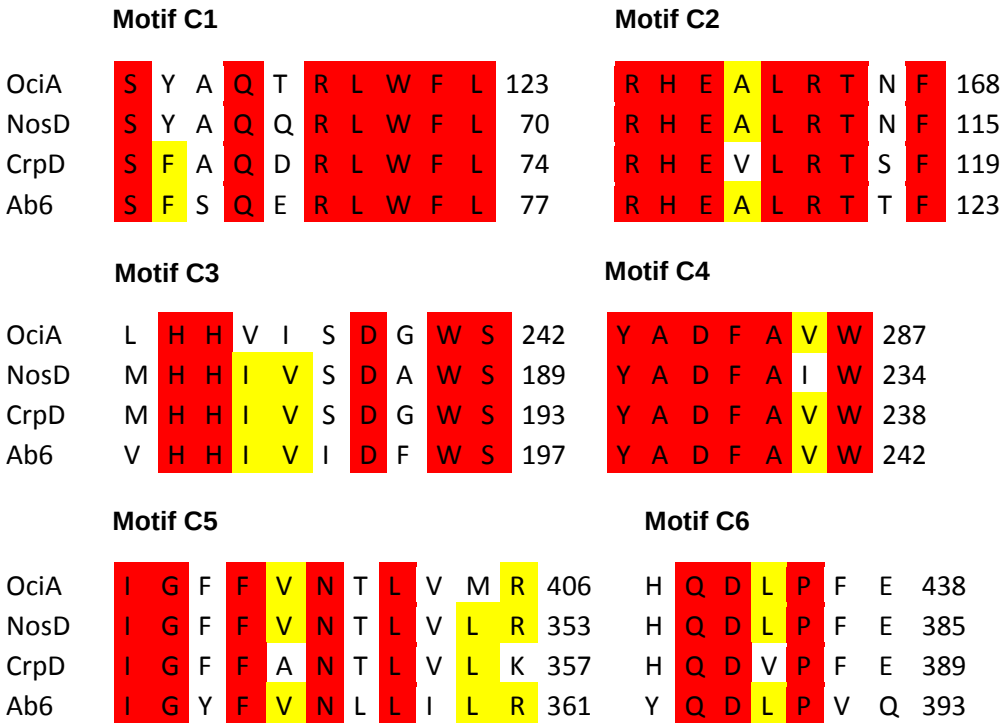


Figure 4.7.9.1-1: Multiple sequence alignment of Ab6 with amino acid sequences of C domains from OciA from *Planktothrix rubescens* NIVA-CYA 98, NosA and NosD from *Nostoc* sp. GSV224 and CrpD from *Nostoc* sp. ATCC 53789. For GenBank accession numbers, see table 4.7.9.1-1. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

Motif C7

OciA	I	H	Q	L	F	E	E	Q	V	E	R	T	P	N	A	V	A	V	V	F
NosD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CrpD	V	H	F	L	S	S	N	-	-	-	-	-	-	N	L	Q	I	Y	I	L
Ab6	V	Y	V	A	P	R	N	-	-	-	-	-	-	-	-	-	-	-	P	L

Figure 4.7.9.1-1 (continued)

Motif C1 is not completely conserved in Ab6 containing a phenylalanine residue instead of an alanine as the third amino acid of the motif. Motif C3 as it is present in the deduced peptide sequence of *ab6* differs from the usual motif found in C domains [table 4.7.9.1-2], however the essential second histidine residue is present (Sieber & Marahiel, 2005).

C domains belong to the superfamily of CoA-dependent acyltransferases, which carry the typical consensus sequence H-H-X-X-X-D-G. In case of C domains, the latter is responsible for the aminoacyl transfer to generate amide bonds in a growing peptide chain. The histidine and aspartate moieties of the motif are considered to be essential for acyltransferase activity (Bergendahl *et al.*, 2002).

Motif C5 is conserved completely except the 7th amino acid being leucine instead of threonine. However, striking difference in motifs C6 and C7 from those postulated by Konz and Marahiel (1999) were found for the C domain of Ab6.

4.7.9.2 NRPS domain TIGR01720

BLAST search performed on Ab6 revealed that the C domain is followed by a peptide synthase domain, which partially was recognised as TIGR01720 (NCBI Conserved Domains Database: TIGR01720). A classification to the NRPS-para261-superfamily (Genbank accession: cl11771) was suggested by BLAST. However, the conserved residues, which are the following three could not be recognised in the sequence: a highly conserved lysine residue at the N-terminus (position 11), which is followed by R-X-X-P-X-X-G-X-G-Y-G with an invariant proline and a conserved first glycine. A second motif is F-N-Y-L-G located about 22 residues later. Near the C-terminus of the domain, a third sequence T-X-S-D is usually conserved and carries a nearly invariant serine and aspartate (NCBI Conserved Domains Database: TIGR01720) (Marchler-Bauer *et al.*, 2009).

Furthermore, the peptide synthase domain TIGR01720 usually has a size of about 171 amino acids, whereas in Ab6 approximately 40 of 170 amino acids were recognised as TIGR01720 [table 4.7.2-2]. Therefore, the sequence is probably an artefact and not related to ambigol biosynthesis. Its possible function in ambigol biosynthesis could not be deduced.

4.7.9.3 PP-binding domain of Ab6

The PP-binding domain of Ab6 was analysed by BLAST search and resulted in the identification of the conserved motif D-X-F-F-X-X-L-G-G-(H,D)-S-(L,I) according to Konz, D. and Marahiel, M. (1999), see figure 4.7.9.3-1.

Table 4.7.9.3-1: BLAST search results for the PCP domain of Ab6. The gene *ab6* was found on fosmid E8. It was also confirmed by contig 00522, obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
Phosphopantetheine attachment site [*Anabaena variabilis* ATCC 29413]	49/83 (59%)	63/83 (75%)	YP_325330
Phosphopantetheine attachment site ['*Nostoc azollae*' 0708]	47/81 (58%)	59/81 (72%)	YP_003721147
PP-binding domain of MicD [*Planktothrix rubescens* NIVA-CYA 98]	47/83 (56%)	62/83 (74%)	CAQ48261 (Rounge *et al.*, 2009)
PP-binding domain of NosD [*Nostoc* sp. GSV224]	46/83 (55%)	62/83 (74%)	AAF17281 (Hoffmann *et al.*, 2003)

The PCP domain of Ab6 belongs to the superfamily of PP-binding proteins (GenBank accession: cl09936) containing a phosphopantetheine attachment site that is recognised by the signature motif D-X-F-F-X-X-L-G-G-(H,D)-S-(L,I) with an invariant serine residue according to Konz, D. and Marahiel, M. (1999). Figure 4.7.9.3-1 shows a multiple sequence alignment of the PP-binding domain with similar domains from strains of *Anabaena*, *Nostoc* and *Planktothrix*, which allowed specifying the conserved motif in Ab6, present as D-N-F-F-E-L-G-G-E-S-L.

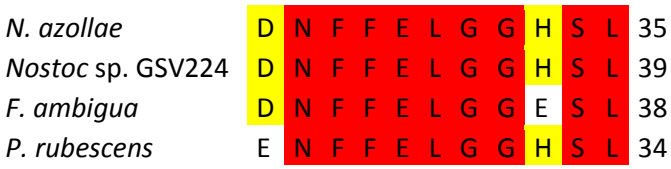


Figure 4.7.9.3-1: Multiple sequence alignment of Ab6 from *F. ambigua* with amino acid sequences of PP-domains from *Anabaena variabilis* ATCC 29413, '*Nostoc azollae*' 0708, MicD from *Planktothrix rubescens* NIVA-CYA 98 and NosD from *Nostoc* sp. GSV224. For GenBank accession numbers, see table 4.7.9.3-1. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

4.7.9.4 TE domain of Ab6

A BLAST search performed on the TE domain of Ab6 classifies it to the superfamily of esterases and lipases acting on carboxylic esters (GenBank accession: cl12031) (Marchler-Bauer *et al.*, 2009). It also belongs to the thioesterase protein family, i.e. pfam00975 (Schneider & Marahiel, 1998). In table 4.7.9.4-1, TE domains with highest identity to that of Ab6 are listed.

Table 4.7.9.4-1: BLAST search results for the TE domain of Ab6. The gene *ab6* was found on fosmid E8. It was also confirmed by contig 00522, obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
TE domain [*Nostoc punctiforme* PCC 73102]	130/276 (47%)	181/276 (65%)	YP_001869788
TE domain [*Anabaena variabilis* ATCC 29413]	126/287 (43%)	169/287 (58%)	YP_325330
TE domain of HctF [*Lyngbya majuscula*]	122/307 (39%)	168/307 (54%)	AAY42398 (Ramaswamy *et al.*, 2007)
TE domain of JamP [*Lyngbya majuscula*]	108/271 (39%)	153/271 (56%)	AAS98787 (Edwards *et al.*, 2004)

Thioesterases carry two consensus motifs. The first one is G-X-S-X-G, which is also common in the active sites of serine proteases, lipases and acyltransferases. The second conserved sequence, i.e. G-X-H-F is located approximately 140 amino acids downstream of the first. It contains a histidine residue which is essential for catalytic

activity (Reimmann *et al.*, 2004). The TE domain of Ab6 was aligned with sequences from similar TE domains, listed in table 4.7.9.4-1, in order to identify the two conserved motifs. As a result, the latter were present in Ab6 with some mutations in the second one, which however contains the critical histidine moiety [figure 4.7.9.4-1].

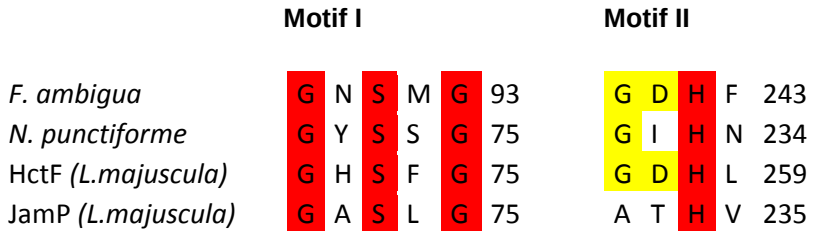


Figure 4.7.9.4-1: Multiple sequence alignment of Ab6 from *F. ambigua* with amino acid sequences of TE domains from *Nostoc punctiforme* PCC 73102, HctF and JamP identified in the cyanobacterium *Lyngbya majuscula*. For GenBank accession numbers, see table 4.7.9.4-1. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

4.7.10 Ab7, a second $FADH_2$-depedent halogenase

Ab7 shares significant homology with published $FADH_2$-dependent halogenases. BLAST search results for the peptide sequence of Ab7 are listed in table 4.7.10-1.

Table 4.7.10-1: BLAST search results for the deduced peptide sequence of *ab7*, found on fosmid E8. It was also confirmed by contig 00522, obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
Halogenase McnD [*Microcystis aeruginosa*]	367/556 (66%)	438/556 (78%)	CAO82181 (Cadel-Six *et al.*, 2008)
Halogenase AerJ [*Microcystis aeruginosa* NIES-98]	359/554 (64%)	432/554 (77%)	ACM68683 (Ishida *et al.*, 2009)
Halogenase AdpC [*Anabaena circinalis* 90]	355/556 (63%)	433/556 (77%)	CAC01605 (Rouhiainen *et al.*, 2000)

Ab7 contains the two conserved signature motifs for $FADH_2$-dependent halogenases described in 4.7.4 and a conserved lysine at position 78 [figure 4.7.10-1].

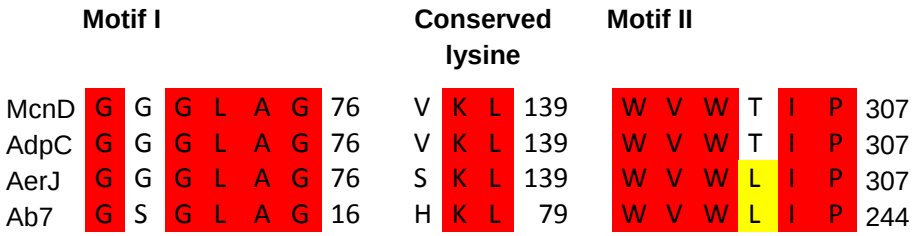


Figure 4.7.10-1: Multiple sequence alignment of Ab7 from *F. ambigua* with amino acid sequences of published $FADH_2$-dependent phenolic halogenases of *Microcystis* sp. and *Anabaena circinalis*. For GenBank accession numbers, see table 4.7.10-1. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

The peptide sequence of Ab7 is similar to halogenases from the biosynthesis of chlorinated cyclic peptides, i.e. McnD known from the biosynthesis of cyanopeptolin 984 in several *Microcystis* strains (Rounge *et al.*, 2007), AerJ from *Microcystis aeruginosa* NIES-98 (Cadel-Six *et al.*, 2008) from aeruginosin production and ApdC, which is involved in the anabaenopeptilide 90B assembly line in *Anabaena circinalis* 90 (Rouhiainen *et al.*, 2000). It is remarkable that the two halogenases Ab1 and Ab7 share only 22 % amino acid sequence identity.

4.7.11 Orf23, a putative DAHP synthetase

The gene *orf23* possibly encodes a 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP) synthetase. It shares 65 % identity with DAHP synthetases from *Nodularia spumigena CCY9414*, *Nostoc punctiforme* PCC 73102 and *Anabaena variabilis* ATCC 29413 [table 4.7.11-1]. These belong to the DAHP synthetase I family, i.e. class I DAHP synthetases according to the classification by Birck and Woodard (Wu & Woodard, 2006; Birck & Woodard, 2001). DAHP synthetases catalyse the first, crucial step in the shikimate pathway, in which chorismate and prephenate are formed as important intermediates for the production of the aromatic amino acids, i.e. phenylalanine, tyrosine and tryptophan. As well, chorismate is a precursor for the formation of aromatic aryl acids, e.g. 4-hydroxybenzoic acid [figure 5.5-1] (Wu *et al.*, 2003). DAHP synthetases exist as many isoforms and can be specifically inhibited,

either by allostery feedback inhibition or by transcriptional regulation (Grove & Gunsalus, 1987).

Table 4.7.11-1: BLAST search results for the deduced amino acid sequence of *orf23*, found on fosmid E8. It was also confirmed by contig 00522 obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number
3-Deoxy-7-phosphoheptulonate synthase [*Nodularia spumigena CCY9414*]	233/363 (65%)	295/363 (82%)	ZP_01628888
Phospho-2-dehydro-3-deoxyheptonate aldolase [*Nostoc punctiforme* PCC 73102]	233/363 (65%)	294/363 (81%)	YP_001864699
Phospho-2-dehydro-3-deoxyheptonate aldolase [*Anabaena variabilis* ATCC 29413]	235/364 (65%)	290/364 (80%)	YP_320855

4.8 Elucidation of the putative tjipanazole D biosynthetic gene cluster

Tjipanazoles have been isolated from *Tolypothrix tjipanasensis* (Bonjouklian *et al.*, 1991) as well as from *F. ambigua* (Falch *et al.*, 1995). Tjipanazole D consists of an indolocarbazole matrix with two tryptophan residues that are chlorinated at position 5 of the indole ring system. PCR studies on genes encoding for the respective tryptophan halogenase in *F. ambigua* were accomplished previously and led to the identification of a 730 bp long PCR fragment that showed highest identity with the tryptophan 5-halogenase of *Streptomyces rugosporus* LL-42D005 (Wagner, 2008). The fosmid library of *F. ambigua* was screened using the specific primers T-neu-fw and T-neu-rev (3.4.14.8), in order to identify clones containing the 730 bp long partial halogenase sequence. No positive clone could be identified in the genomic library. Therefore, the genome assembly from the recent 454 sequencing of *F. ambigua* (4.6) was searched for tryptophan halogenases using the software CLUSEAN (3.1.13) (Weber *et al.*, 2009).

4.8.1 Overall sequence analysis of contig 15287

Contig 15287 with a size of 59 kb was found to carry the complete putative tjipanazole D biosynthetic gene cluster, boardered by a transposase gene *orf1* and a probable regulator gene *orf4*. Detailed BLAST search results are listed in table 4.8.1-1.

Table 4.8.1-1: BLAST search results for the transposase-boardered assembly of genes putatively involved in tjipanazole D production, found on contig 15287. The sequences were obtained from 454 sequencing of *F. ambigua*. The deduced amino acid sequences of *orf*s found on contig 15287 (first column) were aligned with homologous proteins by BLAST (3.1.13). In column four, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). In the first column, the orientation of *orf*s identified on contig 15287 is given: > (5',3'-direction); < (3',5'-direction).

Gene	Size (kb)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
***orf1* >**	0.3	Transposase [*Methanosarcina acetivorans* C2A]	60/101 (59%)	NP_615790
***tj1* >**	1.54	Putative L-tryptophan oxidase [*Lechevalieria aerocolonigenes*]	208/492 (42%)	CAC93714 (Nishizawa *et al.*, 2005)
***tj2* >**	3.45	Chromopyrrolic acid synthase *StaD* [*Streptomyces* sp. TP-A0274]	558/1211 (46%)	BAC15759 (Asamizu *et al*., 2006)
***orf2* >**	1.1	Hypothetical protein (putative O-methyltransferase) [*Nodularia spumigena* CCY9414]	158/342 (46%)	ZP_01629838
***orf3* >**	0.63	NAD(P)H-dehydrogenase [*Anabaena variabilis* ATCC 29413]	135/212 (63%)	YP_321581
***tj3* >**	1.59	Monooxygenase FAD-binding [*Methylobacterium radiotolerans* JCM 2831]	79/203 (38%)	YP_001754189

Table 4.8.1-1 (continued):

Gene	Size (kb)	Highest homology (protein level)	Identity of aligned amino acids	GenBank accession number
tj4 **>**	1.29	Putative cytochrome P450 enzyme [*Actinomadura melliaura*]	190/424 (44%)	ABC02792 (Gao *et al.*, 2006)
tj5 **<**	1.72	Tryptophan 5-Halogenase PyrH [*Streptomyces rugosporus*]	257/517 (49%)	AAU95674 (Zhu *et al.*, 2009)
orf4 **>**	0.49	CMP/dCMP deaminase, zinc-binding [*Nostoc punctiforme* PCC 73102]	123/142 (86%)	YP_001869809
orf5 **>**	0.72	Hypothetical protein Ava_3794 [*Anabaena variabilis* ATCC 29413]	177/238 (74%)	YP_324294

An overview of genes possibly involved in tjipanazole D production and their arrangement is given by figure 4.8.1-1.

Figure 4.8.1-1: Organisation of genes putatively involved in the biosynthesis of tjipanazole D. The genes *orf1* and *orf4* are likely the boarders of the biosynthetic gene cluster and presumably not part of the tjipanazole D biosynthetic pathway. Putative proteins, encoded by *orf2* and *orf3* could not be ascribed to the production of tjipanazole D. The *orf*s 1-4, which are of unknown function are coloured in black. Genes *tj1*-5, which are probably related to the tjipanazole D biosynthesis are shown in blue.

4.8.2 Tj1, a putative L-tryptophan oxidase

A BLAST search on the deduced peptide sequence of *tj1* revealed homology with the L-tryptophan oxidase RebO from the rebeccamycin biosynthetic gene cluster identified in the actinomycete *L. aerocolonigenes*. Furthermore, Tj1 is similar to AtmO, an L-tryptophan oxidase from the recently characterised biosynthetic gene cluster of AT2433-A1 [table 4.8.2-1] (Gao *et al.*, 2006).

Table 4.8.2-1: BLAST search results for the deduced amino acid sequence of *tj1*, found on contig 15287, which was obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
Putative L-tryptophan oxidase RebO [*Lechevalieria aerocolonigenes*]	208/492 (42%)	284/492 (57%)	CAC93714 (Nishizawa *et al.*, 2005)
Amine oxidase [*Salinispora arenicola* CNS-205]	204/496 (41%)	287/496 (57%)	YP_001537178 (Penn *et al.*, 2009)
Putative L-tryptophan oxidase AtmO [*Actinomadura melliaura*]	213/493 (43%)	284/493 (57%)	ABC02789 (Gao *et al.*, 2006)

Based on an amino acid alignment among putative amine oxidases listed in table 4.8.2-1 and the deduced peptide sequence of *tj1*, two conserved motifs were recognised [figure 4.8.2-1], i.e. G-X-G-X-X-G-X-X-X-[G/A] flagging the FAD binding site of amine oxidases (Nishizawa *et al.*, 2005), and secondly a G-G motif. The dinucleotide binding motif and the GG doublet occur in miscellaneous families of flavoproteins (Vallon, 2000). The FAD-binding motif is also a typical conserved region found in halogenases (4.7.4, 4.7.10, 4.8.8).

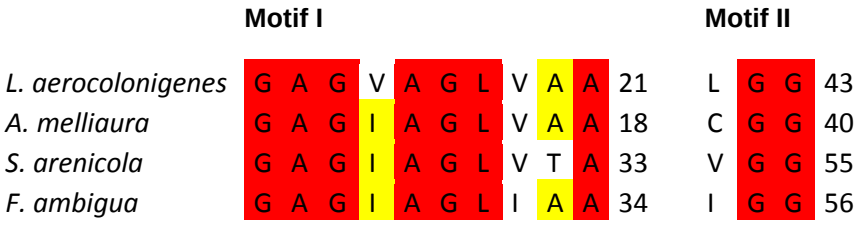


Figure 4.8.2-1: Multiple sequence alignment of the deduced amino acid sequence of *tj1* from *F. ambigua* with amino acid sequences of putative amine oxidases from *Lechevalieria aerocolonigenes*, *Salinispora arenicola* CNS-205 and *Actinomadura melliaura*. For GenBank accession numbers, see table 4.8.2-1. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

4.8.3 Tj2, a possible chromopyrrolic acid synthase-like protein

The deduced amino acid sequence of *tj2* was analysed by BLAST search for homologous proteins implicating a high degree of identity to the chromopyrrolic acid (CPA) synthase StaD from *Streptomyces* sp. TP-A0274] and the homologous protein RebD from *Lechevalieria aerocolonigenes* (Howard-Jones & Walsh, 2005). It also bears high similarity to VioB, from the violacein biosynthetic pathway (Balibar & Walsh, 2006) [table 4.8.3-1].

Table 4.8.3-1: BLAST search results for the deduced peptide sequence of *tj2*, found on contig 15287, which was obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino aid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
Chromopyrrolic acid synthase StaD [*Streptomyces* sp. TP-A0274]	558/1211 (46%)	713/1211 (58%)	BAC15759 (Asamizu *et al.*, 2006; Onaka *et al.*, 2002)
VioB [*Salinispora arenicola* CNS-205]	535/1204 (44%)	707/1204 (58%)	YP_001537179 (Penn *et al.*, 2009)
Putative CCA synthetase [*Actinomadura melliaura*]	484/1205 (40%)	648/1205 (53%)	ABC02790 (Gao *et al.*, 2006)
RebD [*Lechevalieria aerocolonigenes*]	480/1209 (39%)	634/1209 (52%)	CAC93715 (Sánchez *et al.*, 2002; Onaka *et al.*, 2003)
Chromopyrrolic acid synthase InkD [*Nonomuraea longicatena*]	480/1218 (39%)	623/1218 (51%)	ABD59213 (Kim *et al.*, 2007)

StaD is a tetrameric hemoprotein that was demonstrated to catalyse the coupling reaction of two molecules of indole-3-pyruvic acid imine to yield chromopyrrolic acid (Ryan & Drennan, 2009; Asamizu *et al.*, 2006; Howard-Jones & Walsh, 2005). RebD and VioB, which participate in the rebeccamycin and violacein biosynthetic pathway, respectively, are like StaD supposed to mediate the formation of a carbon-carbon bond between the β-carbons of two indole-pyruvate-imine units, in order to give the corresponding iminophenylpyruvate dimer (5.7.3.2). The latter is thought to be spontaneously transformed to CPA (Ryan & Drennan, 2009). For the assembly of the

rebeccamycin aglycone, a tandem mechanism of RebD together with RebO was postulated (Howard-Jones & Walsh, 2005). Proteins of the StaD family comprise about 1000 amino acids. As reported by Asamizu *et al*. (2006), a search of the amino acid sequence database did not give significant conserved domains among the StaD family and other proteins including heme containing proteins.

4.8.4 Orf2, a probable O-methyltransferase

The gene *orf2* was translated to the corresponding peptide sequence and searched for homologous primary structures using BLAST (3.1.13). Search results are presented in detail in table 4.8.4-1 revealing identity of Orf2 with putative O-methyltransferases. Furthermore, a classification of Orf2 to the Methyltransferase 2-superfamily (GenBank accession: cl14604) and to class I of the family of S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase; GenBank accession: cl12011) was suggested by BLAST.

Table 4.8.4-1: BLAST search results for the deduced peptide sequence of *orf2*, found on contig 15287, which was obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number
Hypothetical protein (putative O-methyltransferase) [*Nodularia spumigena* CCY9414]	158/342 (46%)	231/342 (67%)	ZP_01629838
O-methyltransferase family protein [*Nostoc punctiforme* PCC 73102]	153/337 (45%)	225/337 (66%)	YP_001869557
O-demethylpuromycin-O-methyltransferase [*Microcystis aeruginosa* NIES-843]	153/336 (45%)	221/336 (65%)	YP_001659523
O-methyltransferase family 2 [*Cyanothece* sp. PCC 7425]	147/335 (43%)	221/335 (65%)	YP_002481441

The SAM-binding site can be recognised by a glycine-rich consensus sequence G-X-G-X-G (Kozbial & Mushegian, 2005), which was found in the primary structure of Orf2

shown in figure 4.8.4-1. The peptide sequence of the probable O-methyltransferase of *F. ambigua* was aligned with similar proteins, listed in table 4.8.4-1.

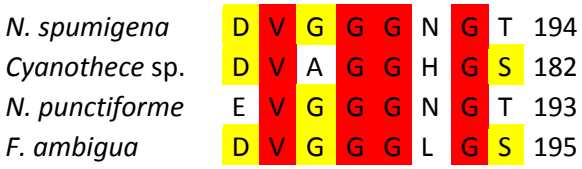


Figure 4.8.4-1: Multiple sequence alignment of the deduced amino acid sequence of *orf2* from *F. ambigua* with amino acid sequences of related O-methyltransferases. For GenBank accession numbers, see table 4.8.4-1. The typical consensus motif G-X-G-G, which is the hallmark of the SAM-binding site, was recognised. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow

A nucleotide BLAST aligning the sequence of gene *orf2* with that of *rebM*, which codifies for the indolocarbazole sugar O-methyltransferase RebM in rebeccamycin biosynthesis (Singh *et al.*, 2008), did not show significant homology. However, it has to be considered that SAM-dependent methyltransferases share little sequence identity, especially their substrate-binding site is highly variable. At the same time, they exhibit a highly conserved structural fold (Martin & McMillan, 2002). Therefore, the sole comparison of primary structures is probably not sufficient to classify Orf2 and to hypothesise on its possible function. RebM was characterised *in vitro* through its crystal structure (Singh *et al.*, 2008). It catalyses the final tailoring step in the assembly of rebeccamycin, the methylation of the C-4 glucosyl atom. The distinct sugar O-alkylations performed by RebM, AtM (AT2433-A1-biosynthesis) and StaMB (staurosporine biosynthesis) have a wide influence on the biological activity of the corresponding indolocarbazole metabolites (Singh *et al.*, 2008).

4.8.5 Orf3, a possible NAD(P)H- dehydrogenase

The deduced amino acid sequence of the gene *orf3* reveals significant similarity to putative NAD(P)H-dehydrogenases, e.g. from *Anabaena variabilis* ATCC 29413 and a probable acyl carrier protein phosphodiesterase from *Nostoc* sp. PCC 7120 [table 4.8.5-1].

Table 4.8.5-1: BLAST search results for the deduced peptide sequence of *orf3*, found on contig 15287, which was obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number
Putative NAD(P)H dehydrogenase (quinone) [*Anabaena variabilis* ATCC 29413]	135/212 (63%)	165/212 (77%)	YP_321581
Putative acyl carrier protein phosphodiesterase [*Nostoc* sp. PCC 7120]	136/212 (64%)	162/212 (76%)	NP_486145
NAD(P)H dehydrogenase (quinone) [*Nostoc punctiforme* PCC 73102]	133/211 (63%)	152/211 (72%)	YP_001869506
FMN-dependent NADH-azoreductase 2 [*Oscillatoria* sp. PCC 6506]	122/216 (56%)	154/216 (71%)	CBN55427

Orf3 keeps a canonical pyridine nucleotide binding domain that is stamped by the conserved motif G-X-X-P (Ojha *et al.*, 2007; Roma *et al.*, 2005) [figure 4.8.5-1]. The related gene *orf3* is located adjacently to gene *tj3* [figure 4.8.1-1], which probably encodes an FAD-binding monooxygenase (4.8.6). Therefore, a possible protein-protein interaction between Orf3 and Tj3 for cofactor regeneration and transfer might be assumed. Similar considerations could be made for an interaction between Orf3 and the putative $FADH_2$-dependent halogenase Tj5 (discussed in 5.7.4).

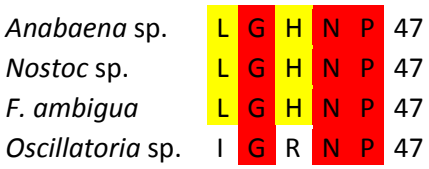


Figure 4.8.5-1: Multiple sequence alignment of the deduced amino acid sequence of *orf3* from *F. ambigua* with amino acid sequences of related putative NAD(P)H-dehydrogenases. For GenBank accession numbers, see table 4.8.5-1. The alignment shows the consensus motif G-X-X-P, which is a signature motif of the pyridine nucleotide binding site. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

4.8.6 Tj3, a putative FAD-binding monooxygenase

A BLAST search performed on the amino acid sequence of Tj3 revealed similarity to different FAD-binding monooxygenases [table 4.8.6-1]. Near the N-terminus of the protein, the typical dinucleotide binding motif G-X-G-X-X-G was recognised [figure 4.8.6-1].

Table 4.8.6-1: BLAST search results for the deduced peptide sequence of *tj3*, found on contig 15287, which was obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
Monooxygenase FAD-binding [*Methylobacterium radiotolerans* JCM 2831]	211/548 (38%)	293/548 (53%)	YP_001754189
Monooxygenase FAD-binding [*Roseomonas cervicalis* ATCC 49957]	207/544 (38%)	294/544 (54%)	ZP_06896947
Putative monooxygenase, FAD binding (hydroxylase) [*Bradyrhizobium* sp. ORS278]	181/565 (32%)	274/565 (48%)	YP_001208727 (Giraud *et al.*, 2007)
Monooxygenase FAD-binding [*Cyanothece* sp. PCC 7822]	183/547 (33%)	287/547 (52%)	ZP_03153217
Putative 2-polyprenyl-6-methoxyphenol hydroxylase [*Sorangium cellulosum* 'So ce 56']	191/538 (35%)	191/538 (35%)	YP_001618529
Pentachlorophenol 4-monooxygenase [*Stigmatella aurantiaca* DW4/3-1]	179/540 (33%)	253/540 (46%)	ZP_01461750

M. radiotolerans	V	V	G	A	G	P	I	G	L	24
R. cervicalis	V	I	G	A	G	P	V	G	L	24
F. ambigua	I	V	G	A	G	P	V	G	L	12
S. aurantiaca	I	A	G	A	G	P	T	G	L	19

Figure 4.8.6-1: Multiple sequence alignment of the deduced amino acid sequence of *tj3* from *F. ambigua* with amino acid sequences of related putative FAD-binding monooxygenases. For GenBank accession numbers, see table 4.8.6-1. The alignment shows the consensus motif G-X-G-X-X-G, which is a signature motif of the FAD-binding site. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

Aligning Tj3 with RebC from the rebeccamycin biosynthetic pathway resulted in 25% identity. Recent studies on RebC and StaC support the idea that both enzymes stabilise reactive intermediates generated by the CYP 450-dependent enzymes RebP and StaP (4.8.7), respectively, in order to mediate the formation of a single product, i.e. the indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazole core structure of rebeccamycin and staurosporine, respectively (Howard-Jones & Walsh, 2006). An alignment comparing the amino acid sequence of Tj3 with StaC revealed 24% identity.

4.8.7 Tj4, a putative CYP 450 enzyme

BLAST search results with Tj4 are presented in table 4.8.7-1. An identity of 45 % was found to the CYP 450 enzyme StaP from *Streptomyces* sp. TP-A0274. Further, Tj4 shares 44 % sequence identity with RebP from *L. aerocolonigenes* on the peptide level.

Table 4.8.7-1: BLAST search results for the deduced peptide sequence of *tj4*, found on contig 15287, which was obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
Putative cytochrome P450 enzyme [*Actinomadura melliaura*]	190/424 (44%)	256/424 (60%)	ABC02792 (Gao *et al.*, 2006)
Cytochrome P450 [*Roseiflexus* sp. RS-1]	180/412 (43%)	245/412 (59%)	YP_001275131
Cytochrome P450 [*Roseiflexus castenholzii* DSM 13941]	183/415 (44%)	242/415 (58%)	YP_001434483
StaP [*Streptomyces* sp. TP-A0274]	184/407 (45%)	245/407 (60%)	BAC55212 (Onaka *et al.*, 2002)
StaP [*Streptomyces longisporoflavus*]	191/424 (45%)	251/424 (59%)	ABI94389 (Howard-Jones & Walsh, 2006)
RebP [*Lechevalieria aerocolonigenes*]	176/415 (42%)	244/415 (58%)	BAC15753 (Onaka *et al.*, 2003)

StaP and RebP are well-known CYP 450 enzymes from the staurosporine and rebeccamycin biosynthetic pathways, respectively. StaP has been demonstrated to act on chromopyrrolic acid, in order to catalyse an intramolecular aryl-aryl coupling between two indole moieties to give the indolocarbazole skeleton (Howard-Jones & Walsh, 2007; Makino *et al.*, 2007; Howard-Jones & Walsh, 2006). The latter occurs with auxiliary modifications, i.e. glycosylation (Sánchez *et al.*, 2006b) and in case of rebeccamycin also halogenation (Yeh *et al.*, 2005). Although to date only StaP has been studied in detail, all results are supposed to be likewise valid for its homologue, RebP (Ryan & Drennan, 2009).

The amino acid sequence of Tj4 carries three conserved signature motifs as described in 4.7.5 suggesting a classification to the P450-superfamily (GenBank accession: cl12078). Figure 4.8.7-1 shows an alignment of the conserved regions of Tj4 with related CYP 450 enzymes. These are: motif A-(A,G)-X- (E,D)-T of helix I with a generally conserved threonine residue, the generally conserved E-X-X-R motif that is located in helix K and the haem binding motif F-X-X-G-X-X-X-C-X-G at the start of the L-helix, which is located in the active center. The latter region contains the absolutely conserved cysteine ligand.

The compared sequences were taken from BLAST search results obtained for the deduced amino acid sequence of gene *tj4* [table 4.8.7-1].

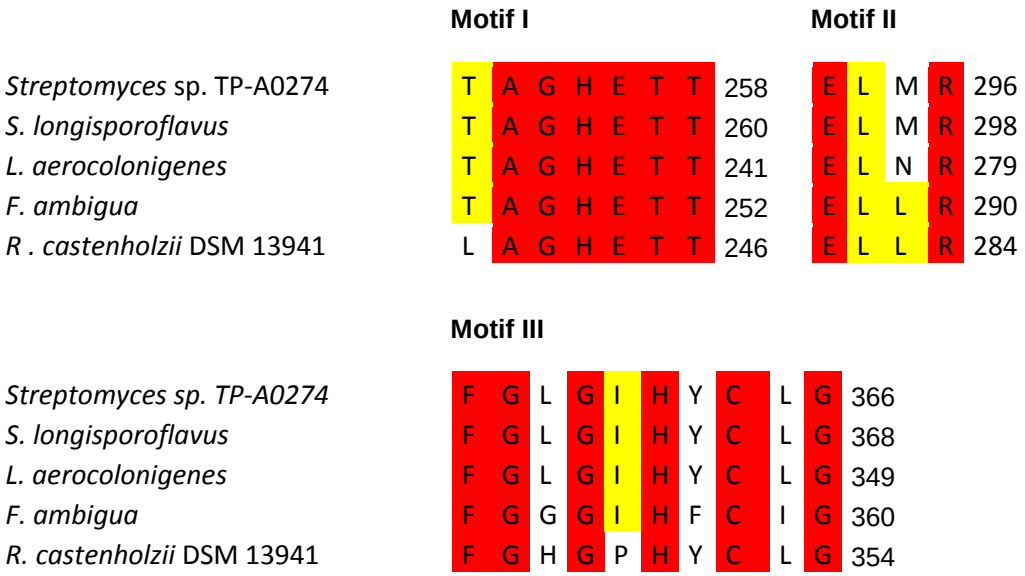


Figure 4.8.7-1: Multiple sequence alignment of the deduced amino acid sequence of *tj4* from *F. ambigua* with amino acid sequences of related putative CYP 450-dependent monooxygenases. For GenBank accession numbers, see table 4.8.7-1. Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to four compared sequences are coloured in yellow

4.8.8 Tj5, a putative tryptophan halogenase

The gene *tj5* was translated to the corresponding amino acid sequence (3.1.13). The protein sequence was searched for homologous proteins using BLAST. It bears significant identity to $FADH_2$-dependent halogenases, which are involved in the chlorination of tryptophan moieties. Results and references are presented in table 4.8.8-1 for most related tryptophan halogenases.

Table 4.8.8-1: BLAST search results for the deduced amino acid sequence of *tj5*, found on contig 15287, which was obtained from 454 sequencing of *F. ambigua*. In column two, the ratio of identical amino acids (first value) to all compared amino acids (second value) is given (percentage in brackets). Positives are amino acid residues that are similar to each other concerning their chemical properties (e.g. polarity, aromaticity, side chains, and acidity).

Homologous amino acid sequence	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number/Reference
Tryptophan 5-halogenase PyrH [*Streptomyces rugosporus*]	257/517 (49%)	361/517 (69%)	AAU95674 (Zhu *et al.*, 2009)
KtzR [*Kutzneria* sp. 744]	242/503 (48%)	337/503 (66%)	ABV56598 (Fujimori *et al.*, 2007)
Putative tryptophan halogenase [*Burkholderia ambifaria* MC40-6]	232/539 (43%)	329/539 (61%)	YP_001811924
Tryptophan halogenase PrnA [*Myxococcus fulvus*]	231/536 (43%)	321/536 (59%)	AF161185_4 (Hammer *et al.*, 1999)
Putative tryptophan halogenase *[Cyanothece* sp. PCC 8802]	232/535 (43%)	316/535 (59%)	YP_003137148
Tryptophan halogenase PrnA [*Pseudomonas fluorescens*]	223/539 (41%)	329/539 (61%)	AF161184_1 (Hammer *et al.*, 1999)

The protein is characterised by the typical conserved regions as described in 4.7.4. Figure 4.8.8-1 presents a multiple sequence alignment of three consensus motifs, which represent a hallmark of $FADH_2$-dependent halogenases (Wagner *et al.*, 2009). The conserved lysine moiety that is involved in the mechanistic action of the halogenation reaction was recognised at position 74 (Yeh *et al.*, 2007; Buedenbender *et al.*, 2009). Interestingly, this halogenase shows an identity of about 26 % to the putative phenolic halogenase Ab1, which is supposed to be involved in ambigol formation (4.7.4). This is a higher degree of identity than it was found between Ab1 and the second putative phenolic halogenase Ab7 (4.7.10).

	Motif I											Conserved lysine				
S. rugosporus	V	G	G	G	T	A	G	W	M	T	17	Y	K	L	G	76
F. ambigua	V	G	G	G	T	A	G	W	M	S	20	Y	K	T	A	79
P. fluorescens	V	G	G	G	T	A	G	W	M	A	20	F	K	A	A	80
Cyanothece sp. PCC 8801	V	G	G	G	T	A	G	W	M	T	27	F	K	T	A	87

	Motif II								
S. rugosporus	G	W	M	W	T	I	P	L	285
F. ambigua	G	W	V	W	N	I	P	L	295
P. fluorescens	G	W	T	W	K	I	P	M	278
Cyanothece sp. PCC 8801	G	W	I	W	K	I	P	M	283

Figure 4.8.8-1: Multiple sequence alignment of the deduced amino acid sequence of *tj4* from *F. ambigua* with amino acid sequences of related tryptophan halogenases. For GenBank accession numbers, see table 4.8.8-1. The alignment shows the consensus motifs of flavin-dependent halogenases (4.7.4). Amino acids identical in all compared sequences are shaded in red and those, which are only identical among up to three compared sequences are coloured in yellow.

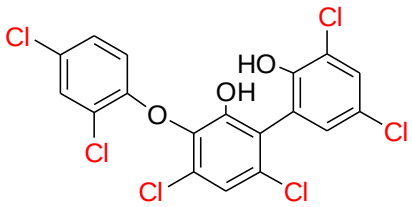
5. Discussion

Cyanobacteria produce a great variety of secondary metabolites. These include cyanobacterial toxins (Wiegand & Pflugmacher, 2005; Smith *et al.*, 2008) but also compounds interesting for therapeutical use (Singh *et al.*, 2005). The most prevalent cyanobacterial natural products are NRPS and PKS derived (Liu & Rein, 2010; Sielaff *et al.*, 2006; Dittmann *et al.*, 2001). Also, many typical cyanobacterial bioactive secondary metabolites are produced in combinatorial pathways like mixed PKS/NRPS systems, for instance curacin A and the cryptophycins (Jones *et al.*, 2010; Wagner *et al.*, 2009; Jones *et al.*, 2009). Mixed NRPS/PKS sytems are of particular interest, because they combine principals of the peptide biosynthesis with those of fatty acid biosynthesis. Both are modular systems that apply analogous acyl chain elongation routes with differences in substrate recognition and catalysed bond formation (Walsh & Fischbach, 2010; Cane & Walsh, 1999).

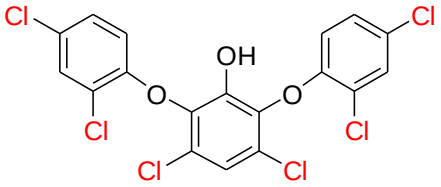
Besides their variety in constructing complex core structures with high diversity (Welker & von Doehren, 2006), cyanobacteria possess impressing skills in elaborating structural modifications, performed by a considerable repertoire of specific enzymes. These reactions include halogenation of aromatic and aliphatic moieties, cryptic halogenation, SAM-dependent methylations, reduction and HMG-CoA-dependent transformations (Jones *et al.*, 2010), to mention some of them. Usually these reactions are carried out in the course of peptide or mixed NRPS/PKS biosynthesis, and therefore are often dependent on certain carrier protein domains (Walsh, 2008; Rausch *et al.*, 2007; Walsh *et al.*, 2001).

5.1 Bioactive secondary metabolites from *F. ambigua*

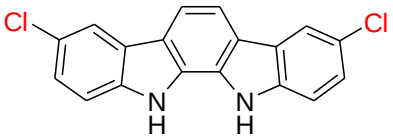
F. ambigua is a gliding, filamentously growing cyanobacterium producing highly interesting chlorinated metabolites, which exhibit a spectrum of biological activities. These include antibacterial, antifungal, cytotoxic and molluscicidal properties for the polychlorinated phenolic ethers ambigol A and B [figure 5.1-1]. A moderate antibacterial activity was found for the indolo[2,3-*a*]carbazole tjipanazole D (Falch, S. *et al*, 1995).



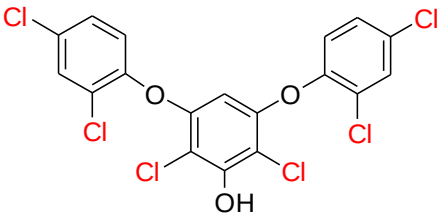
ambigol A



ambigol B



tjipanazole D



ambigol C

Figure 5.1-1: Halogenated aromatic metabolites, isolated from the cyanobacterium *F. ambigua*

With the ambigols, *F. ambigua* produces unprecedented compounds concerning their presence in cyanobacteria. Similar structures, however with a dimeric core scaffold, are known from sponges, e.g. from *Lamellodysidea herbacea* (formerly *Dysidea herbacea)* (Unson *et al.*, 1994). The biosynthesis of the latter, however has been ascribed to cyanobacteria. The sponge-derived structures contain only a dimeric backbone connected by an ether bridge and are brominated. By contrast, the ambigols are trimeric natural products endowed with both, C-O and C-C linkages. Further, the aromatic framework of the ambigols is ornated with a high density of chlorine atoms preserving strict regioselectivity in the halogenation pattern [figure 5.1-1]. Compareable structures from other organisms, which also have ether and biphenyl linkages, do not possess the high number of chlorine substituents, e.g. phloroglucinol compounds from brown algae [figure 1.3.1.2-1].

The therapeutic potential of polyhalogenated, hydroxylated aromatic compounds was shown for numerous related natural products, especially polybrominated biphenyls from *Lamellodysidea* (formerly *Dysidea)* spp. (Gribble, 2010; Shridhar *et al.*, 2009).

As outlined in 1.3.1.4, the synthetic antimicrobial agent triclosan is structurally related to the overall ambigol structure. It is medically used, e.g. as a disinfectant against clinically important bacteria, i.e. methicillin-resistant *Staphylococcus aureus* (MRSA). The use of triclosan as part of daily applied products, e.g. soaps, toothpastes or

textiles, however is regarded as critical due to resistance development in microorganisms. Triclosan possesses an inhibitory effect on fatty acid synthetatase systems of microbes (Freundlich *et al.*, 2009; Freundlich *et al.*, 2007). The latter may also be the mode of action for ambigol C having antiplasmodial and trypanocidal activity (Wright *et al.*, 2005).

To date, no data are available, explaining the biosynthesis of compounds like ambigol B, which was the focus of the current work.

The unique core structure of tjipanazole D, i.e. its indolo[2,3-*a*]carbazole scaffold was to date only found in cyanobacteria (Falch *et al.*, 1995; Bonjouklian *et al.*, 1991). Similar compounds, which possess an indolo[2,3-*a*]pyrrolo[3,4-*c*] were isolated from actinomycetes, e.g. rebeccamycin and staurosporine (Bush *et al.*, 1987; Tamaoki *et al.*, 1986). Due to their promising therapeutic effects, e.g. as anticancer agents (Nakano & Omura, 2009), they have been studied extensively towards structure-activity relationships (Marminon *et al.*, 2008; Voldoire *et al.*, 2004; Anizon *et al.*, 2003; Prudhomme, 2003; Bailly *et al.*, 1999; Pereira *et al.*, 1996) and their biosynthetic pathways (Chiu *et al.*, 2009a; Sánchez *et al.*, 2006b; Onaka *et al.*, 2002; Onaka *et al.*, 2003).

5.2 Problems in DNA extraction from filamentous cyanobacteria

Aiming the identification of the biosynthetic gene clusters of the ambigols and tjipanazole D, genomic DNA had to be isolated from *F. ambigua*. DNA was then used to establish a fosmid library (4.3.3), which was screened with specific primers for the putative $FADH_2$-dependent halogenases Ab7 (4.7.10) and Tj5 (4.8.8). In addition, sequence data were completed by whole genome sequencing and analysis (4.6).

Extraction of good quality genomic DNA from cyanobacteria, especially filamentously growing strains, often includes difficulties with sheath-associated bacteria, effective filament breakage and cell lysis as well as removal of solved trace contaminants, in particular polysaccharides of the mucilaginous envelope (Hoiczyk & Baumeister, 1995). The production of these polysaccharides seems to be related directly to the extent of illumination of cyanobacterial cultures and the release of carotenoids (Ehling-Schulz *et al.*, 1997). In accordance with studies on the terrestrial

cyanobacterium *Nostoc commune*, an excessive light exposure of *F. ambigua* cultures resulted in slower growth with enhanced release of carotenoids and accumulation of polysaccharides, changing the translucent culture medium into a viscous, orange coloured liquid [figure 4.3.2-2] (Ehling-Schulz *et al.*, 1997). When high concentrations of carotenoids were present, they could not be removed completely from the final DNA solution, which revealed an orange color.

The biological role of the enhanced carotenoid production was suggested to be the protection of cyanobacterial cells from intensive light exposure to maintain optimal conditions for photosynthesis (Liang *et al.*, 2006; Paerl, 1984). The coincident increase of polysaccharide production is necessary to provide a matrix for other UV light absorbing substances, i.e. the aromatic pigment scytonemin or oligosaccharide mycosporine amino acids (OS-MAAs), which are closely attached to the glycan by non-covalent interactions (Ehling-Schulz *et al.*, 1997; Böhm *et al.*, 1995; Hill *et al.*, 1994). When high amounts of polysaccharides were present in the cyanobacterial cultures, traces of these contaminants were retrieved in the final DNA solution increasing its viscosity. Due to the presence of anionic groups in the polysaccharide chains, their separation from DNA was ineffective when using typical DNA purification columns (3.4.2; 4.3.1). For the genus *Fischerella*, it was in particular ascertained that released polysaccharides consist of deoxysugars (Nicolaus *et al.*, 1999; De Philippis *et al.*, 2001), and as a consequence, they might interfere with downstream applications, in particular PCR amplifications and ligation reactions (Sambrook & Russell, 2001; Porter, 1988).

In conclusion, the well-dosed application of light together with a well-adjusted culture time (21 days) was a quality-determining factor for the obtained DNA solution. This was in particular important, because *F. ambigua* showed in general a low growth rate, with an obviously long generation time (population doubling time), whereas the induction of carotenoid production started after a relatively short time depending on light intensity (Fernández-González *et al.*, 1998). Furthermore, it has to be considered that for the cyanobacterium *Synechococcus elongatus* PCC7942, a correlation between its division rate and environmental factors, i.e. light intensity, temperature, and nutrients was reported, whereby LD (light/dark) 12:12 cycles administer a fitness advantage to cyanobacterial colonies (Hellweger, 2010; Ouyang *et al.*, 1998). The circadian clock (Yang *et al.*, 2010) in cyanobacteria affects gene

expression, and as a consequence cellular processes (Johnson *et al.*, 2008b; Johnson *et al.*, 2008a). For instance, nitrogen fixation, which is an oxygen sensitive process reaches its maximum at the night phase, and thus is uncoupled from the oxygen producing photosynthesis (Grobbelaar *et al.*, 1986). Very recently, an important link between three so called “clock proteins” (KaiA-KaiC) and the cell division control of *S. elongates* has been discovered (Dong *et al.*, 2010a; Dong *et al.*, 2010b).

Cyanobacteria, e.g. *F. ambigua* are well-known to harbor associated heterotrophic bacteria in their polysaccharide- and peptide-containing sheath providing a suitable biotope to them (Hube *et al.*, 2009). Detectable amounts of sheath material are usually released to the culture medium by polysaccharide-producing cyanobacteria (De Philippis *et al.*, 2001) allowing heterotrophic bacteria to settle even in BG-11 medium, constituting a pure salt medium. A removal of the identified *P. stutzeri* strain (4.2) by antibiotic treatment revealed that the presence of these heterotrophs seems to play a role in growing conditions, because they significantly accelerate the growth of *F. ambigua*. Rippka *et al* (1988) postulated that cyanobacteria require aerobic heterotrophic bacteria for growth (Rippka, 1988). This is likely one reason, why axenic cultures of cyanobacteria are known to be difficult to obtain (Choi *et al.*, 2008). Antibiotically pretreated cultures were not useful for DNA-extraction due to highly decreased amount of harvested cells. Interestingly, treatment with lysozyme as described in 3.3.3 resulted in a compareable slow growth rate (4.2.2). An approach to remove the *P. stutzeri* strain by excessive UV-light exposure was also not successful, because as mentioned above, light absorbing pigments are always present in effective amount within the polysaccharide sheath (Hoiczyk & Hansel, 2000). Thus, the associated bacteria were probably sheltered during excessive UV-light exposure.

Although cyanobacterial cell walls are affiliated to the Gram-negative type, the peptidoglycan layer present in cyanobacterial cell walls is significantly thicker than that of most other Gram-negative bacteria [figure 4.3.1.1-1b] (Hoiczyk & Hansel, 2000). Hence, a yielding DNA extraction requires effective cell breakage measurements by application of strong tensides, whereby the most effective turned out to be SDS (4.3.1). The robustness of cyanobacteria offers an important advantage allowing the application of milder cell lysis conditions to eliminate the

accompanying bacteria without or slightly affecting the intactness of the cyanobacterial cells (4.3.1.3).

5.3 Construction and screening of a genomic library

A genomic library aims to represent the whole genome of an organism in assessable sequence fragments of about 40 kb. This requires, as a first step, a ligation of isolated genomic DNA into an appropriate vector. The ligation of extracted *F. ambigua* DNA into the fosmid vector pCC1FOS turned out to be rather complicated (4.3.3). This may have several reasons. Without doubt, the discussed trace contaminants (5.2) in the DNA solution may affect the T4-DNA ligase, perhaps by a kind of competitive inhibition caused by anionic polysaccharides. Another possibility is that acidic or alkaline groups of these contaminants changed the optimal buffer conditions for the T4-DNA ligase in an unfavorable way. With respect to the sensitivity of blunt end ligations towards the vector:template-ratio, the inhibitory effect of possible contaminants is even more significant.

Another problem did arise with the screening of the genomic library, when some sequences could not be retrieved using sequence-specific primers. In particular, a partial sequence of the tryptophan 5-halogenase gene *tj5* (4.8.8), obtained in a previous work (Wagner, 2008), could not be detected in the library by PCR (4.8). When searching for fosmids overlapping with fosmid E8 using specific primers designed from boarder sequences, i.e. gene *ab2* (4.7.5) and a gene encoding a putative acetate kinase (4.4.2), no positive clones could be detected. This lack of representatives in the library may be reasoned by possible traces of *P. stutzeri* DNA in the ligation mixture, which were presumably incorporated in a number of fosmids. In this case, the stored 2,800 clones were obviously not enough to represent all of the *Fischerella* genome.

Calculating the cost of fosmid sequencing in comparison to the diminishing costs for whole genome sequencing (4.6), a fosmid library might be no longer the best way, if only nucleotide sequences shall be analysed, i.e. elucidation of new biosynthetic gene clusters. However, the great advantage of having the actual DNA fragments present in a vector is their use in possible downstream experiments, in particular for heterologous expression studies.

With regard to the time consuming procedure of library construction and to the relative cost of fosmid sequencing, a 454 sequencing of the *F. ambigua* genome was favoured over the construction of a further, bigger fosmid library. The size of *orf*s related to the ambigols and tjipanazole D biosynthesis was estimated to be suitable for PCR amplifications. Therefore, sequence data, obtained from fosmid E8 and 454 sequencing, enable the amplification of genes probably related to ambigol and tjipanazole D biosynthesis for heterologous expression studies in future projects.

5.4 Whole genome analysis

In the last five years, rapid advances have emerged in the development of alternatives to the automated Sanger method for DNA sequencing, which was the prevalent method for almost two decades (Sanger *et al.*, 1977). These new strategies, referred to as next-generation sequencing (NGS) include different improved technologies, i.e. Roche/454, Illumina/Solexa, Life/APG and Helicos BioSciences (for a comprehensive review, see Metzker, 2010). The 454 sequencing applied for this project is based on massive parallelisation of sequencing reactions by performing millions of the latter individually at the same time in picotiter plates (3.4.15). By this method, 500 MB of DNA per run and 1 GB DNA per day can be sequenced (Grody *et al.*, 2010).

The overwhelming yield of sequence information per read along with significantly lower cost is a result of the high speed, in which sequencing reactions maybe carried out simultaneously. This development has changed the way of planning scientific projects, since the sole elucidation of sequence data has been facilitated to an extreme extent. As a consequence, this technical advance in methods of molecular biology, in particular the automation of DNA sequencing, has implied the involvement of bioinformatics and the establishment of databases to enable fast gene identification along with prediction of the putative function of corresponding proteins. Current and future projects are highly dependent on the use of systematised collections of proteomics data (Swiss Institute of Bioinformatics http://www.isb-sib.ch; http://www.ncbi.nlm.nih.gov/protein; Marchler-Bauer *et al*., 2011) with special regard to the knowledge of protein-protein interactions, so that it might be possible to dissect even proteins from complex enzymatic machineries, i.e. from ambigol biosynthesis for biochemical experiments.

In order to assure a sufficient coverage of the *Fischerella* genome by 454 sequencing, it was important to assure that the utilised DNA solution did not contain any foreign DNA, i.e. from the associated *P. stutzeri* strain (4.2). Therefore, the applied DNA was obtained by method B described in 3.4.2.2, which was the preferred method for DNA extraction in this work (4.3.1.3). The method includes the lysis of *P. stutzeri* cells and removal of foreign DNA prior to lysis of cyanobacterial cells. Furthermore, the resulting DNA solution did undergo a 16S rDNA analysis for *Pseudomonas*, to approve successfully its purity (4.1). As a result of the 454 sequencing, a 26.5-fold coverage of the *F. ambigua* genome was estimated.

5.5 Overview of the biosynthesis of ambigols and tjipanazole D

At a first glance, ambigols and tjipanazole D are structurally completely different compounds [figure 5.1-1]. However, with the knowledge of genes encoding proteins that probably participate in the biosynthesis of these secondary metabolites, a high degree of similarity can be constituted concerning their biosynthetic assembly. Both, the ambigols and tjipanazole D contain an aromatic unit that most likely arises from the shikimate pathway, and in particular from chorismate. Chorismic acid is the key intermediate in the production of several aromatic acids, i.e. phenylalanine, tyrosine and tryptophan but also of some aryl acids like 4-hydroxybenzoic acid (4-HBA) [figure 5.5-1] (Dewick, 2009; Van Lanen *et al.*, 2008). As a consequence, many of the known aromatic secondary metabolites contain chorismate-derived structural units. These include the large group of NRPS- and mixed PKS/NRPS-compounds. As well, more simple bioactive molecules, such as the small trimeric nonribosomal peptide enterobactin and the antibiotic chloramphenicol, contain aromatic building blocks, which arise from chorismate [figure 5.5-1]. In ambigols and tjipanazole D biosynthesis, the supposed monomers are 4-hydroxybenzoic acid (4-HBA) and L-tryptophan, respectively.

The indole-containing aromatic amino acid L-tryptophan is produced from anthranilic acid and a phosphoribosyl unit in a few reaction steps (Dewick, 2009). Tryptophan does not only serve as a building block for the biosynthesis of nonribosomal peptides but also functions as a precursor of a variety of alkaloid structures. The structural diversity of these indolic secondary metabolites produced by bacteria has been reviewed recently (Ryan & Drennan, 2009).

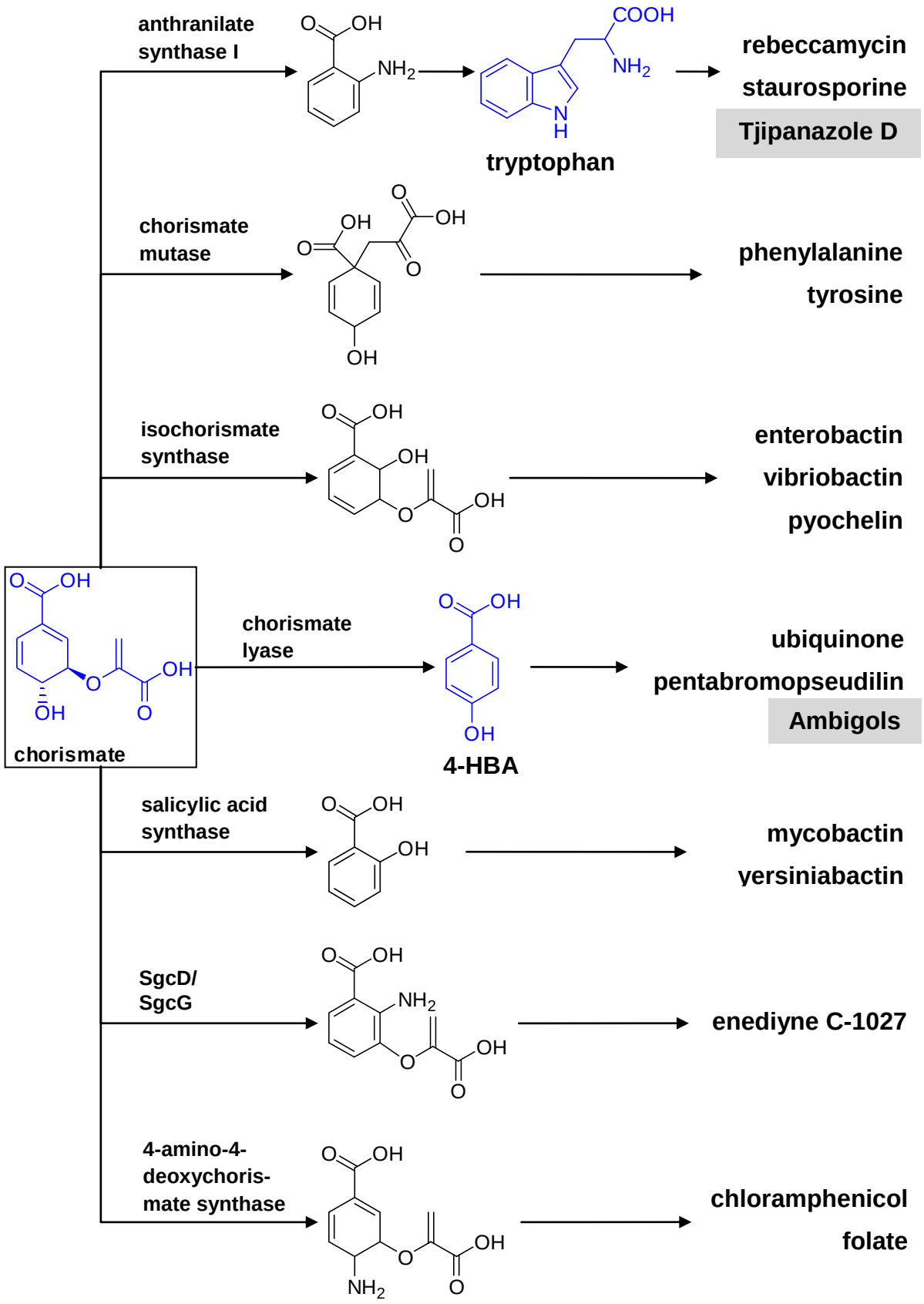


Figure 5.5-1: Overview of some chorismate derived metabolites. The ambigols and tjipanazole D are also supposed to arise from chorismate. 4-HBA = 4-hydroxybenzoic acid. SgcD and SgcG are two enzymes from enediyne C-1027 biosynthesis (Van Lanen *et al.*, 2008).

In contrast to plants, bacteria are capable to generate 4-hydroxybenzoic acid (4-HBA) directly from chorismic acid by a single elimination reaction, catalysed by a 4-HBA synthetase via removal of the enolpyruvic acid side chain of chorismate (Dewick, 2009). In the putative ambigol biosynthetic gene cluster, a sequence that probably encodes a chorismate lyase (4.7.7; 1.4.2) was identified. This might allow anticipating that the origin of the phenolic structure of these compounds is 4-HBA.

Apart from their common origin, both, tryptophan and 4-HBA are suggested to be substrates for oxidative coupling reactions giving rise to the structural matrix of tjipanazole D and the ambigols, respectively.

A third common feature of these compounds is their tailoring modification, i.e. chlorination (4.7.4; 4.7.10; 4.8.8). However, differences in the halogenations process have to be considered (Buedenbender *et al*., 2009; Zhu *et al*., 2009; Flecks *et al*., 2008). As described previously, halogenation of phenolic residues is often performed on a carrier-attached substrate (1.4.1; 1.4.2). Furthermore, it is well-established that in the assembly of complex glycopeptide antibiotics, an NRPS-bound peptide chain is modified by intramolecular, CYP 450-catalysed crosslinking (Woithe *et al.*, 2007). By contrast, the production of indolocarbazole frameworks, for instance that of rebeccamycin, does involve free diffusible substrates for chlorination as well as for the production of the indolocarbazole scaffold (5.7.1). Therefore, it is not surprising that in the putative biosynthetic gene cluster of the ambigols (*ab*-operon), an NRPS-like module, i.e. a possible carrier protein, was identified, whereas such a protein is not encoded by the putative tjipanazole D operon *tj*.

Within this work, *ab* and *tj* genes were identified and characterised (4.7; 4.8). Proof of their involvement in the production of ambigols and tjipanazole D awaits gene inactivation experiments and/or heterologous expression of relevant enzymes. Both clusters are boardered at one side by a transposase gene. For the *ab*-operon, this transposase gene was named *orf25* [figure 5.6-1] (4.7.2) and for *tj* (4.8) *orf4* [figure 5.7-1]. Some other cyanobacterial gene clusters, i.e. those for cryptophycin (Magarvey *et al.*, 2006), curacin A (Chang *et al.*, 2004) and the jamaicamides (Edwards *et al.*, 2004), are also flanked by transposases. Thus, this feature seems to be characteristic for cyanobacterial biosynthetic gene clusters.

The *ab*-operon is possibly only responsible for the production of ambigol B. This will be discussed further in 5.6.

5.6 The putative biosynthetic gene cluster for the ambigols

Genes probably involved in the biosynthesis of ambigols A-C [figure 5.1-1] were found by three different strategies, i.e. PCR techniques, genomic library construction and whole genome analysis (3.4.9; 3.4.14; 3.4.15). First, partial sequences encoding for the halogenase Ab7 (4.7.10) were obtained by the design of degenerate primers in a previous work (Wagner, 2008) using published data on $FADH_2$-dependent halogenases.

A PCR screening performed in this study led to the identification of additional sequence information on the phenolic halogenase Ab7 (4.3.2; 4.7.10). Subsequently, a genomic library was established (4.3.3). The high specifity of halogenase genes, which is in particular due to the presence of highly conserved motifs (4.7.4; 4.7.10; 4.8.8), facilitated the localization of the putative ambigol biosynthetic gene cluster in the fosmid library.

Fosmid E8 contained most of the genetic information required to construct a hypothetical biosynthetic pathway for the ambigols. However, the sequence of the gene *ab2* encoding a putative CYP 450 (4.7.5) was not completely covered by E8. As well, the halogenase gene *ab1* (4.7.4) was not identified [figure 5.6-1]. Screening for fosmids overlapping with E8 (4.5) was not successful. As mentioned in 5.3, traces of DNA from the accompanying *Pseudomonas* strain (4.2) did probably interfere with the ligation reaction during library production. Therefore, the calculated 2,800 clones of the fosmid library were not sufficient to represent the complete *F. ambigua* genome. Instead of constructing a new fosmid library, whole genome sequencing was preferred, as this was judged a cheaper and faster alternative with regard to the cost of further fosmid sequencing.

The putative ambigol biosynthetic gene cluster has a size of approximately 13 kb and is boardered by *orf21* and *orf24* [table 4.7.2-2; figure 5.6-1]. The deduced peptide sequences of the latter two genes could only be assigned to hypothetical proteins with unknown function. In case of Orf21 (4.7.3), similarity to regulative proteins, i.e. a possible AAA-ATPase-like protein (i.e. ATPases associated with diverse cellular

activities; identity = 63%) and a putative WD (tryptophan/aspartate)-repeat containing protein (identity = 62%) was found. Immediatialy downstream of *orf24*, the gene *orf25* was deciphered as a putative transposase [table 4.7.2-2], which most likely is the boarder of the ambigol biosynthetic gene cluster.

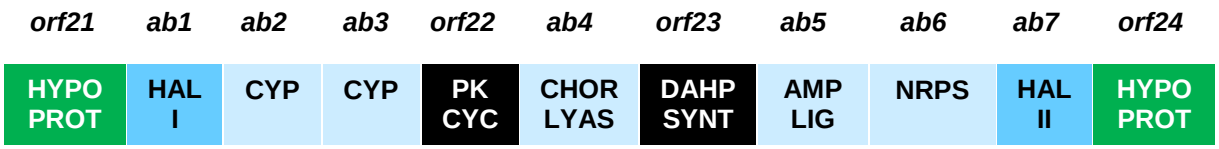


Figure 5.6-1: Schematic overview of the arrangement of genes (*ab1-ab7*), which are putatively related to ambigol-production in *F. ambigua*. The coloured areas represent the corresponding, putative proteins. Areas coloured in light blue are involved in the formation of the ambigol core structure. Proteins coloured in blue are probable halogenases (HAL), which are responsible for tailoring reactions of ambigols, i.e. regiospecific chlorination. The black shaded areas represent proteins that may fulfil regulative functions. The cluster is boardered by *orf21* and *orf24*. HYPOPROT = hypothetical protein; CYP = CYP 450 enzyme; PKCYC = polyketide cyclase; CHORLYAS = chorismate lyase; DAHPSYNT = DAHP synthetase; AMPLIG = AMP ligase. For detailed information and the putative function of the respective proteins, see table 4.7.2-2.

The sequences of *ab1* (4.7.4) and *ab7* (4.7.10) encode for putative phenolic halogenases, whereas *ab2* and *ab3* (4.7.5) show significant identity with genes related to CYP 450 enzymes. The sequence of *orf22* (4.7.6) resembles that of a gene encoding a putative cyclase/dehydrase-like protein from the cyanobacterium *Nodularia spumigena* CCY9414 (identity = 61%). Its function in the biosynthetic process could not be deduced. A chorismate lyase with the corresponding gene *ab4* (4.7.7) is likely responsible for the formation of 4-HBA (5.6.1), which presumably represents the origin of the basic monomer of the ambigols, i.e. their phenolic building unit (1.3.1). The gene *orf23* was deciphered as a possible DAHP synthetase. These proteins exist as different isoenzymes and are specifically feedback-inhibited, and thus *orf23* probably has a regulative function in the biosynthesis of antibacterial ambigols (Wu *et al.*, 2003) (1.3; 5.6.7). Adjacently, the gene *ab5* (4.7.8) encodes a putative AMP-ligase with similarity to CoA synthetases, which is probably involved in the activation of 4-HBA to its CoA-ester, or alternatively to an adenylate (5.6.2).

If 4-HBA-CoA arises during ambigol biosynthesis, it is proposed to be transferred to the PCP domain of Ab6 (4.7.9), in order to be presented for a first halogenation reaction that is assumed to be mediated by Ab1 (4.7.4; 5.6.4). The two CYP 450

enzymes, Ab2 and Ab3 are postulated to construct the trimeric ambigol backbone by phenolic oxidative coupling performed on the carrier-bound substrate (5.6.5). A second halogenation reaction is supposed to be mediated by halogenase Ab7 while the trimeric framework is still bound to the PCP domain of Ab6 (5.6.6). The following considerations on a possible ambigol biosynthetic pathway, do not necessarily take the order of involved genes of the cluster into account.

5.6.1 4-HBA, the possible key intermediate in ambigol assembly

Chorismic acid is an important branch point intermediate [figure 5.5-1] in the biosynthesis of aromatic amino acids, i.e. phenylalanine, tyrosine and tryptophan as well as numerous bioactive aromatic compounds, for instance the antitumor agent enediyne C-1027 or the antibiotic compound chloramphenicol (Podzelinska *et al.*, 2010; Van Lanen *et al.*, 2008). Ab4 shares 58% identity with a putative 4-HBA synthetase from *Microcoleus chthonoplastes* PCC 7420 and 55% to a putative chorismate lyase from *Beggiatoa* sp. PS (4.7.7). It should be noted that the information on functions of similar proteins relies exclusively on homology in their peptide sequences and not on in vitro proof for enzymatic activity of Ab4. This has to be proven by in vitro experiments with the heterologously expressed and purified protein (Ingram-Smith *et al.*, 2006b). The superfamily of chorismate lyases (GenBank accession: cl01230) includes the most extensively studied representative UbiC from ubiquinone production (Smith *et al.*, 2006; Holden *et al.*, 2002). It catalyses the conversion of chorismate to 4-HBA by elimination of pyruvate (Gallagher *et al.*, 2001). Relying on sequence identity found for Ab4, it can only be assumed that it is responsible for the formation of 4-HBA [figure 5.6.1-1].

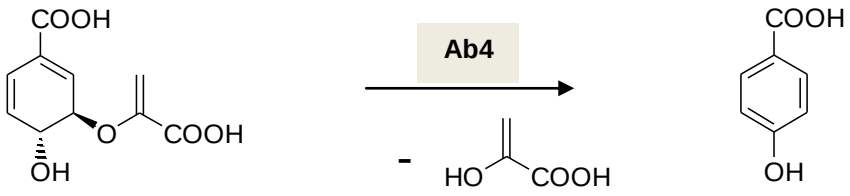


Figure 5.6.1-1: Possible reaction catalysed by the putative chorismate lyase Ab4. Removal of pyruvic acid leads to 4-HBA.

The latter, however is most likely an intermediate in ambigol biosynthesis, because it was proven to be the precursor of the structurally similar, marine antibiotic pentabromopseudilin, produced by *Alteromonas luteoviolaceus* [figure 5.6.1-2]. This compound contains a dihalogenated phenolic unit similar to that present in ambigols (Peschke *et al.*, 2005; Hanefeld *et al.*, 1994). A closely related analogue of this secondary metabolite is pentachloropseudilin produced by *Actinoplanes* sp. ATCC 33002 (Wynands & van Pee, 2004).

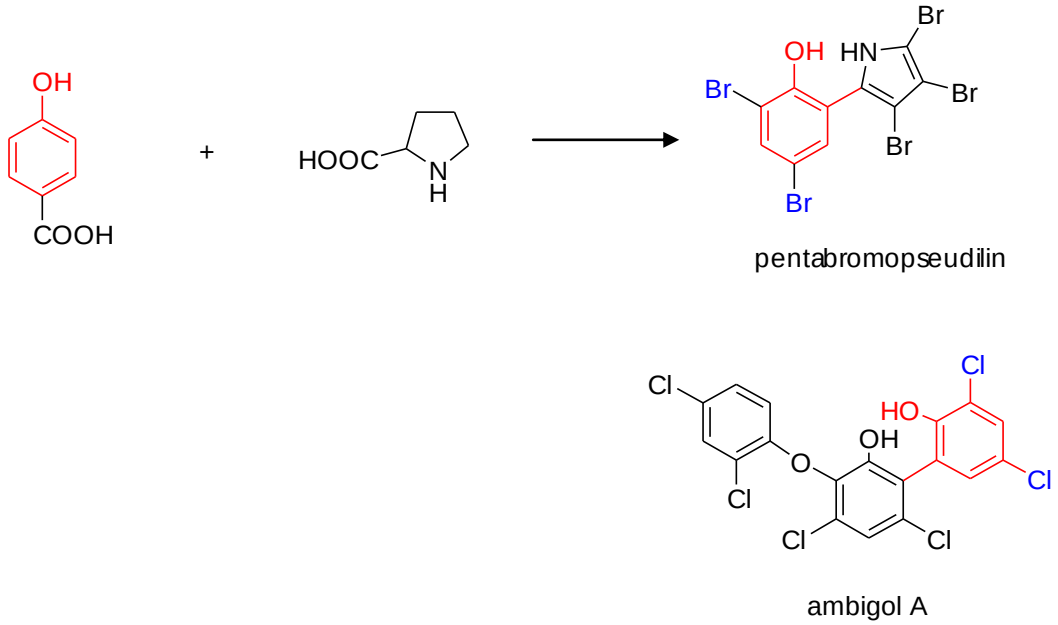


Figure 5.6.1-2: 4-HBA is a building unit of the marine antibiotic pentabromopseudilin

The green marine alga *Ulva lactuca* was analysed for the presence of phenolic compounds [figure 5.6.1-3]. Among other isolated secondary metabolites, 3,5-dibromo-4-hydroxybenzoic acid (3,5-dibromo-4-HBA) was identified. It was suggested that 4-HBA, which also was found in this organism, could be a possible precursor of 3,5-dibromo-4-HBA and 2,4,6-tribromophenol (Flodin & Whitfield, 1999b; Flodin & Whitfield, 1999a).

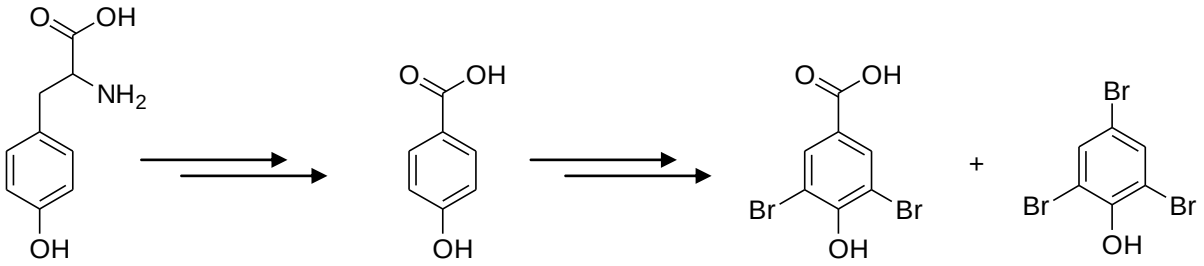


Figure 5.6.1-3: Possible precursors of brominated phenolic compounds from *U. lactuca*

5.6.2 Possible activation of 4-HBA to 4-HBA-CoA

Ab5 (4.7.8) was recognised as an AMP-binding enzyme bearing 50% amino acid identity with an AMP-dependent synthetase and ligase from *Herpetosiphon aurantiacus* ATCC 23779. Secondly, 41% identity to an AMP-binding enzyme from *Teredinibacter turnerae* T7901 was found. AMP-dependent enzymes exist as three variants (1.6), one of which is represented by CoA synthetases. Ab5 resembles several acyl CoA synthetases on the peptide level including a putative acyl CoA synthetase found in *Hahella chejuensis* KCTC 2396 (identity = 36%). Therefore, it might be anticipated that Ab5, whose gene is located downstream of *ab4* (5.6.1) [figure 5.6-1], mediates the formation of the CoA-ester of the aryl acid 4-HBA, and thus activates it for following biosynthetic reaction steps [figure 5.6.2-1].

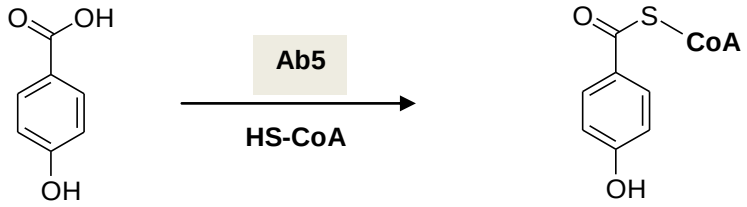


Figure 5.6.2-1: Possible reaction catalysed by the putative CoA synthetase Ab5. It is supposed to activate the aryl acid 4-HBA to its corresponding CoA-ester.

In this context, another possible activation reaction should be mentioned. As described in 1.4.3, AMP ligases may directly transfer an activated adenylate (aryl acid-AMP) onto a PCP domain. An example was described in 1.4.4 for the biosynthesis of the siderophore enterobactin, in which the freestanding A domain EntE generates 2,3-DHB-AMP and transfers it to the ArCP domain of EntB (Roche & Walsh, 2003). Similarly, in the lipidation reaction during daptomycin biosynthesis, the

acyl ACP ligase DptE fulfils a similar function in that it activates fatty acids and transfers them to the cognate ACP DptF (Wittmann *et al*., 2008). By analogy, it is possible that the AMP-ligase Ab5 is responsible for the adenylation of 4-HBA to 4-HBA-AMP along with the transfer of the latter onto the PCP domain of Ab6.

On the other hand, it was reported that EncN, a benzoate CoA-ligase from enterocin biosynthesis is capable to transfer the benzoyl unit of benzoyl-CoA to the ACP domain EncC (transthiolation reaction) (Izumikawa *et al*., 2006). Therefore, it is absolutely essential to apply recombinant Ab5 in biochemical experiments, to investigate wether this enzyme exhibits CoA-ligase- or ACP-ligase-activity, or even both (Izumikawa *et al*., 2006; Ingram-Smith *et al*., 2006b, Knobloch & Hahlbrock 1977). This is especially of interest with regard to the possible function of the starter C domain of the putative “NRPS”-module Ab6, which is proposed to display acyltransferase activity (5.6.3).

5.6.3 Putative carrier protein domain Ab6

The gene codifying for the “NRPS”-module Ab6 (4.7.9) is located adjacent to the putative aryl CoA synthetase gene *ab5* (5.6.2) [figure 5.6-1]. It is composed of a C domain, an unspecific boundary sequence, a PCP and a TE domain. The C domain reveals 51% identity with OciA, a protein domain of an NRPS from *Planktothrix rubescens* NIVA-CYA 98, likely responsible for the production of the cyanopeptolin oscillapeptin G (Rounge *et al.*, 2009). The PCP domain of Ab6 shows 56% identity with the PP-binding domain of MicD from the microginin gene cluster of *Planktothrix rubescens* NIVA-CYA 98 (Rounge *et al.*, 2009). For the TE domain, 39% identity to the TE domain of HctF from the biosynthesis of hectochlorin in *Lyngbya majuscula* was found (Ramaswamy *et al.*, 2007).

Canonical conserved motifs for these three domains, which are required for enzymatic activity, were recognised (4.7.9). Thus, it could be followed that the three domains of the NRPS module Ab6 are functional units. However, their involvement in the biosynthesis of the ambigols remains to be proven by gene disruption studies. A possible function of this NRPS-like module could be the binding of 4-hydroxybenzoyl-CoA (5.6.2) for following reaction steps. The C domain of Ab6 could be an ancient form of acyltransferases, bearing a divergent C3 core motif H-H-I-V-I-D-F, however with the presence of the essential second histidine residue (Samel *et al.*, 2007)

(4.7.9.1). The latter is supposed to be involved in the amide bond formation in nonribosomal peptide biosynthesis by activation of the amino group of a downstream amino acid for its nucleophilic attack on the carboxy group of an upstream amino acid (Roche & Walsh, 2003; Bergendahl *et al*., 2002).

An example for bacterial acyltransferases can be found in the biosynthesis of neutral lipids, i.e. triacylglycerols and wax esters. For enzymes of the wax ester synthase/diacylglycerol acyltransferase (WS/DGAT) family, it could be shown that even the first histidine residue is necessary for the catalytic acyl transfer (Stöveken *et al.*, 2009). Further, the H-H-X-X-X-D-G motif is also present in the chloramphenicol-inactivating, bacterial enzyme chloramphenicol acetyltransferase, which catalyses the transfer of an acetyl group onto a hydroxyl group (Shaw *et al.*, 1979). Moreover, dihydrolipoamid-S-acetyltransferase contains the H-H-X-X-X-D-G motif (Keating *et al*., 2002; Marahiel *et al.*, 1997). This enzyme mediates an interthiol acetyl-transfer between the cofactor S-acetylliponamid and CoA to generate acetyl-CoA.

The starter C domain of Ab6 (1.4.5; 4.7.9) may receive the CoA-bound substrate to transfer it to the PCP domain [figure 5.6.3-1]. Here, an analogy to the biosynthesis of daunorubicin is worthwhile to mention. In the biosynthesis of this anthracycline antibiotic in *S. peucetius*, the KS III domain DpsC fulfils a dual function transferring a propionyl group from a propionyl-CoA starter unit to the 4'-PP cofactor of an ACP while it is also capable to exhibit condensation activity (Bao *et al.*, 1999a). DpsC showed distinct specifity towards propionyl-CoA (Bao *et al.*, 1999b).

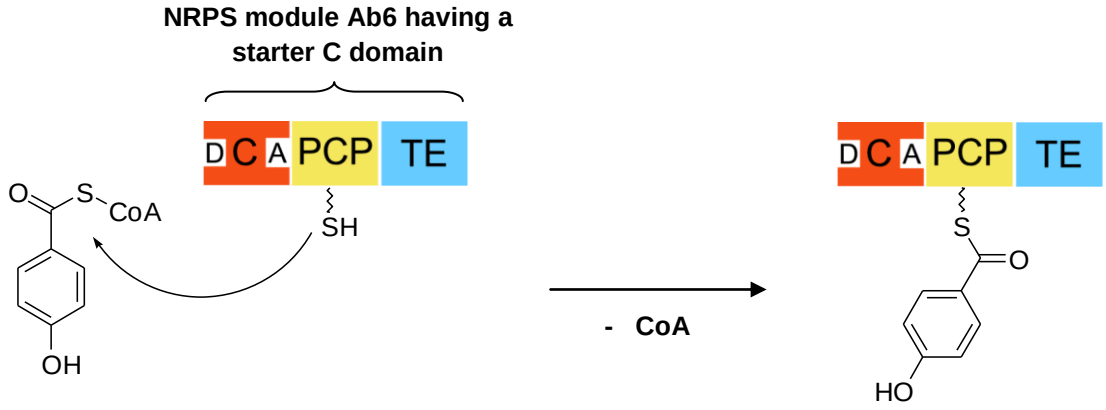


Figure 5.6.3-1: Possible interthiol transfer reaction mediated by the C domain of Ab6. D = donor site; A = acceptor site.

C domains as starter units of an NRPS (1.4.5) are a characteristic for biosynthetic pathways of mixed PKS/NRPS products, lipopeptides and siderophores (Chooi & Tang, 2010; Rausch *et al.*, 2007; Roongsawang *et al.*, 2005). These compounds are characterised by a polyketide derived carboxylic acid, a fatty acid or a hydroxylated carboxylic acid (e.g. 2,3-dihydroxybenzoic acid), respectively, which is received at the donor site by a starter C domain, in order to be linked to an amino acid bound to the acceptor site.

The biosynthesis of surfactin includes a simplified variant of this procedure. The biotenside surfactin is a cyclic lipoheptapeptide lactone isolated from *Bacillus subtilis* (Bonmatin *et al.*, 2003). It is composed of seven amino acids and a β-hydroxy fatty acid. The surfactin biosynthetic gene cluster comprises four biosynthetic genes codifying for the three-modular NRPSs SrfA and SrfB, the one-modular NRPS SrfC, and a dissected type II TE domain, i.e. SrfD. The latter fulfils repair functions during peptide assembly (Yeh *et al.*, 2004). Thus, seven amino acid activating modules were identified in the biosynthetic gene cluster. These are, SrfA1, SrfA2, SrfA3, SrfB1, SrfB2, SrfB3, and SrfC.

Surfactin-production is initiated by the NRPS-module SrfA1, which bears an N-terminal C domain, also referred to as starter C domains (Steller *et al.*, 2004). The role of this unique starter C domain has been extensively studied by Kraas *et al*. (2010). It was shown to catalyse the initial acylation step by linking 3-hydroxy myristic acid to glutamate. Thereby, the lipid chain is not tethered to an ACP domain but instead, 3-hydroxymyristoyl-CoA is directly transferred to the PCP-bound glutamate (Kraas *et al.*, 2010). Interestingly, the surfactin biosynthetic machinery does not include a dedicated CoA-ligase for the formation of the fatty acyl-CoA substrate but instead, this substrate is suggested to be produced by primary metabolism (Chooi & Tang, 2010; Kraas *et al.*, 2010). This example shows that CoA-esters may be produced separately by a lone standing CoA ligase, in order to be recognised and transferred by a modular protein. A stimulating effect of the TE domain SrfD in the uptake of the 3-hydroxymyristoyl-CoA substrate was observed [5.5.1.3-1] (Wittmann *et al.*, 2008; Steller *et al.*, 2004).

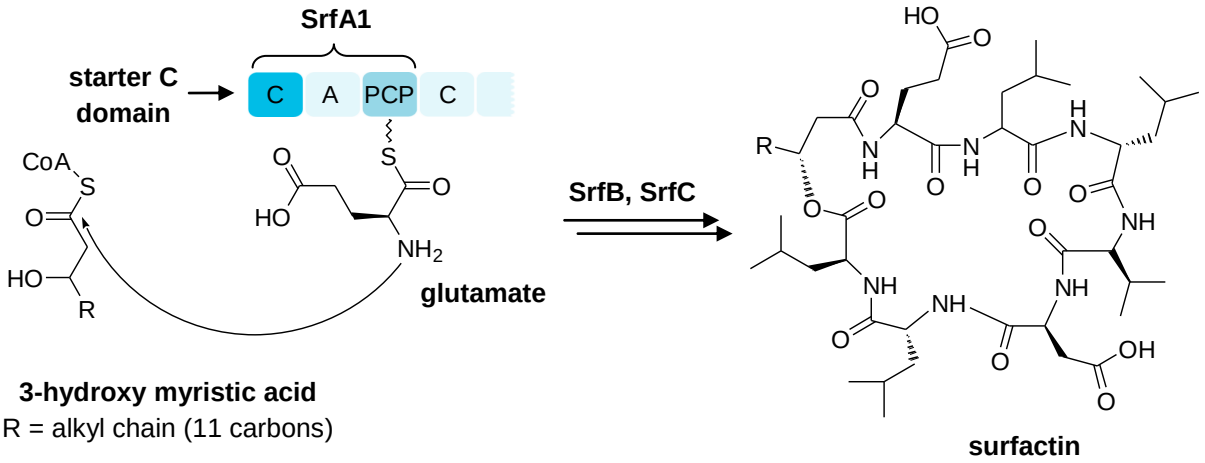


Figure 5.5.1.3-1: SrfA1 is the first module of the three-modular NRPS SrfA. It contains a dissected, N-terminal C domain, an A and a PCP domain. The C domain is involved in the direct utilisation of a fatty acid-CoA, which is linked to glutamic acid by amide bond formation.

Assuming that the C domain of Ab6 is even simpler than that of SrfA1 [figure 5.5.1.3-1], it is conceivable that the NRPS-like module Ab6 would not catalyse a repetitive assembly of amino acids via peptide bond formation but rather transfers the first 4-HBA-CoA unit to the PCP domain. This PCP-tethered 4-HBA-molecule is supposed to undergo dichlorination. Furthermore, the starter C domain of Ab6 could play an assisting role in the oxidative coupling reaction by providing 4-HBA units at its donor site for coupling with the PCP-bound 2,6-dichloro-4-HBA (5.6.4; 5.6.5).

As mentioned in 5.6.2, a direct transfer of an aryl adenylate, i.e. 4-HBA-AMP onto the PCP domain of Ab6 may also be possible. In that case, the function of the starter C domain of Ab6 would be unknown.

Downstream of the C domain, a boundary sequence of about 170 amino acids was identified (4.7.9.2) but is not clearly defined. Although BLAST classifies it partially (~40 amino acids) as TIGR01720, the typical conserved motifs of this protein family are not present in the sequence from *F. ambigua*. It is supposed that the latter is an artefact and not related to the ambigol biosynthesis. TIGR01720 has been found in e.g. linear gramicidin synthetase subunit A (Kessler *et al.*, 2004) or in surfactin synthetase subunit 1, i.e. SrfA, which carries a starter C domain, discussed above (Nakano *et al.*, 1991).

The PCP domain of Ab6 consists of approximately 80 amino acids revealing the typical conserved core region D-X-F-F-X-G-G-(H/D)-S-(L/I). It resembles thiolation

domains from NRPS assembly lines that are involved in the biosynthesis of cyanobacterial nonribosomal cyclic peptides, e.g. the PP binding domain of MicD from microginin production in *Planktothrix rubescens* NIVA-CYA 98 (Rounge *et al.*, 2009) (4.7.9.3).

5.6.4 Probable halogenation of a PCP-bound substrate

The putative flavin-dependent halogenase Ab1 reveals sequence identity with the phenolic halogenase CrpH (identity=58%), which catalyses the chlorination of tyrosine in *ortho*-position to the O-Met moiety in the cryptophycin assembly line (Magarvey *et al.*, 2006). Ab1 also resembles the tryptophan-2-halogenase, CmdE from the myxobacterium *Chondromyces crocatus* with an identity of 44 %. According to Kling *et al*. 2005, $FADH_2$-dependent halogenases share two highly conserved regions, the G-X-G-X-X-G-motif, which is located in the flavin binding site near the N-terminus and a W-X-W-X-I-P sequence found in the middle of the enzyme (Kling *et al.*, 2005). The latter one is supposed to sterically inhibit the active site from substrate binding and thus from catalysing monooxygenase reactions (Dong *et al.*, 2005). These two motifs are a characteristic combination only found in halogenases and are very strong criteria to identify halogenase peptide sequences.

The biosynthesis of chondramide B and D as well as that of cryptophycin is maintained by a modular pathway of mixed NRPS and PKS, in which the halogenation step is integrated, and thus takes place while the substrate is carrier-bound (Rachid *et al.*, 2006; Magarvey *et al.*, 2006). Within both of these biosynthetic gene clusters (*cmd* and *crp*), the halogenase genes are located immediately downstream of the associated NRPS genes. The corresponding chlorinating enzymes CmdE and CrpH, respectively, mediate a final decorating modification of the produced mixed PKS/NRPS chains. By contrast, the halogenase gene *ab1*, unlike the second halogenase gene *ab7*, is located clearly separated from the NRPS gene *ab6*, rather at the beginning of the biosynthetic gene cluster [figure 5.6-1]. It resides directly upstream of the genes for the two CYP 450 enzymes, Ab2 and Ab3. It is likely that the CYP 450's act in tandem with the halogenase. Due to the high sequence homology of Ab1 to NRPS-associated halogenases, it could be expected that this enzyme like most other phenolic halogenases requires a PCP-attached substrate, rather than a free diffusible one. As described in 1.4.1.1, carrier-bound

halogenation may occur as di- or even trichlorination, however with strict regioselectivity (Jones *et al.*, 2010). Ab1 is postulated to perform dichlorination of PCP-tethered 4-HBA at *meta*-position [figure 5.6.4-1]. It cannot be excluded, however that Ab7 (4.7.10) is involved in these halogenations.

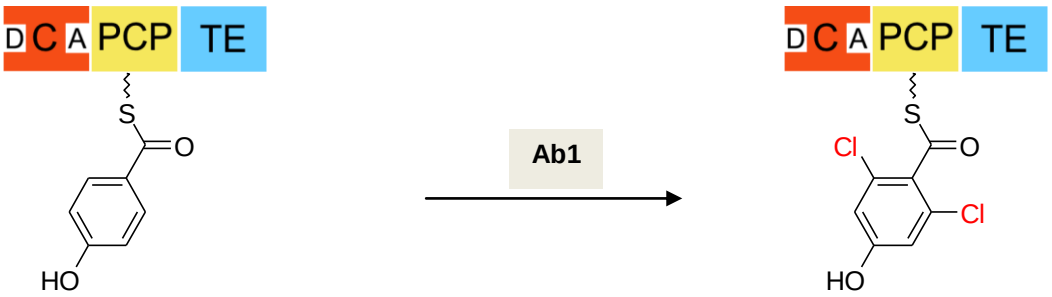


Figure 5.6.4-1: Possible reaction catalysed by the putative halogenase Ab1. It is supposed to dichlorinate the activated substrate 4-HBA-S-PCP in *meta*-position to the OH-group. D = donor site; A = acceptor site.

5.6.5 Putative oxidative coupling reaction

CYP 450 enzymes are able to mediate phenolic oxidative coupling reactions, and thereby produce phenolic ethers or biphenyl compounds (Walsh, 2004) (1.4.4). The two proteins Ab2 and Ab3 show homology to vertebrate CYP 450 enzymes, which fulfil metabolic functions. For Ab2 highest identity was determined to CYP 450 1D1 from the zebrafish *Danio rerio*. Ab3 possesses 35 % identity with CYP 1C2 from the rainbow trout *Oncorhynchus mykiss* (4.7.5).

As might be expected, Ab2 and Ab3 share 73% sequence identity. Taking the ambigol structure into account, one could expect that both enzymes act on a similar substrate, maybe 4-HBA but differ concerning the type of catalysed linkage, i.e. C-O- or C-C-coupling. Another possibility could be that both, Ab2 and Ab3, mediate C-O-coupling, while one of them acts to produce a dimer, whereas the second one is responsible for the formation of the trimeric scaffold. In this case, the identified *ab*-operon would most likely be responsible for the formation of ambigol B only, which constitutes a symmetric molecule.

Both enzymes are supposed to act in tandem with the putative halogenase Ab1 while the substrate is still tethered to the PCP domain of Ab6. As described previously

(1.4.4), an interaction of CYP 450's with PCP domains is well-established. For instance, vancomycin and balhimycin comprise a common heptapeptide aglycone backbone, in which aryl residues are linked via biphenyl and biphenyl ether connections formed via phenolic oxidative crosslinking of the PCP-attached peptide chain (Walsh, 2008; Bischoff *et al.*, 2005; Nolan & Walsh, 2009).

Remarkably, the halogenated trimeric building unit of the peptide scaffold in balhimycin and vancomycin [figure 5.6.5-1] is reminiscent of the ambigols isolated from *F. ambigua* (Falch *et al.*, 1995). The oxidative chemistry in the biosynthesis of vancomycin-type antibiotics has been recently reviewed (Widboom & Bruner, 2009).

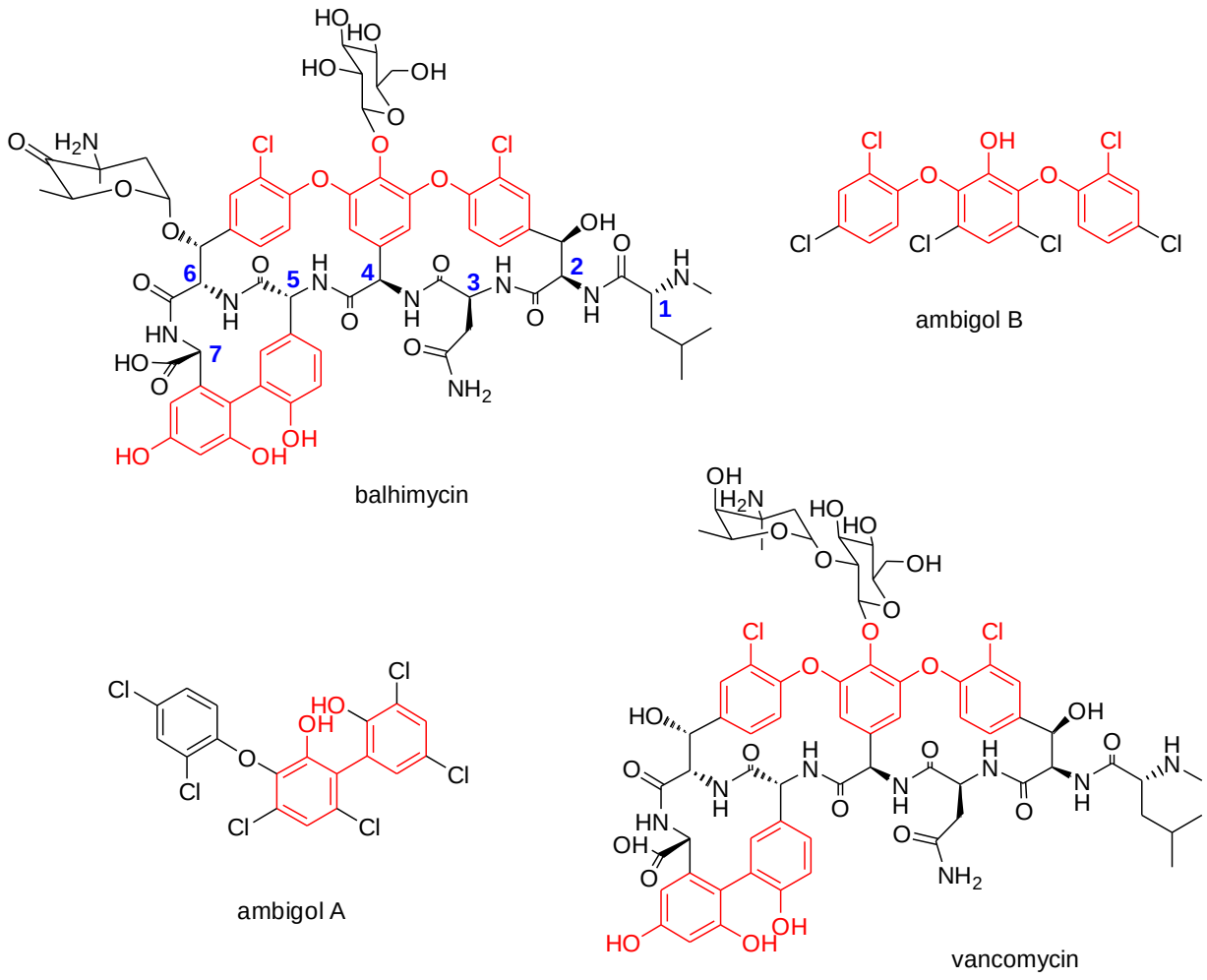


Figure 5.6.5-1: The ambigol structure is highly reminiscent of partial structures in glycopeptide antibiotics, e.g. vancomycin and balhimycin.

Analysis of the biosynthetic gene clusters of different glycopeptide antibiotics, heterologous expression of several involved genes in *E. coli* as well as gene

inactivation studies have given insights into the complex assembly of the crosslinked heptapeptide aglykone scaffold (Hubbard & Walsh, 2003). Knockout studies on the balhimycin-producing strain initially revealed that each crosslinking step within the heptapeptide is accomplished by a particular CYP 450 enzyme, i.e. OxyA, OxyB and OxyC, respectively, by sequential action. Thus, OxyB and OxyA catalyse aryl ether connections between residues 4 and 6 and residues 2 and 4, respectively, whereas OxyC by contrast is responsible for the C-C-crosslinking step between residues 5 and 7 [figure 5.6.5-1] (Nolan & Walsh, 2009; Hadatsch *et al.*, 2007; Stegmann *et al.*, 2006; Walsh, 2004; Bischoff *et al.*, 2001; Pelzer *et al.*, 1999). The crystal structures of OxyB and OxyC from the vancomycin producer *A. orientalis* have been solved recently and the typical P450-fold was reported (Pylypenko *et al.*, 2003; Zerbe *et al.*, 2002). For OxyB, a model hexapeptide bound to a PCP domain served as a substrate to display CYP 450 activity giving rise to a monocyclic product (Widboom & Bruner, 2009; Woithe *et al.*, 2008; Geib *et al.*, 2008).

As mentioned in 5.6.3, the starter C domain of Ab6 could play an assisting role in that it receives the CoA-activated aryl acid 4-HBA at its donor site to promote substrate recognition by the CYP 450 enzymes. The latter may depend on possible enzymatic interactions with the C and PCP domain, respectively. In accordance with the explanations in 1.4.4, a PCP-associated phenolic coupling reaction is reasonable to anticipate for ambigol production.

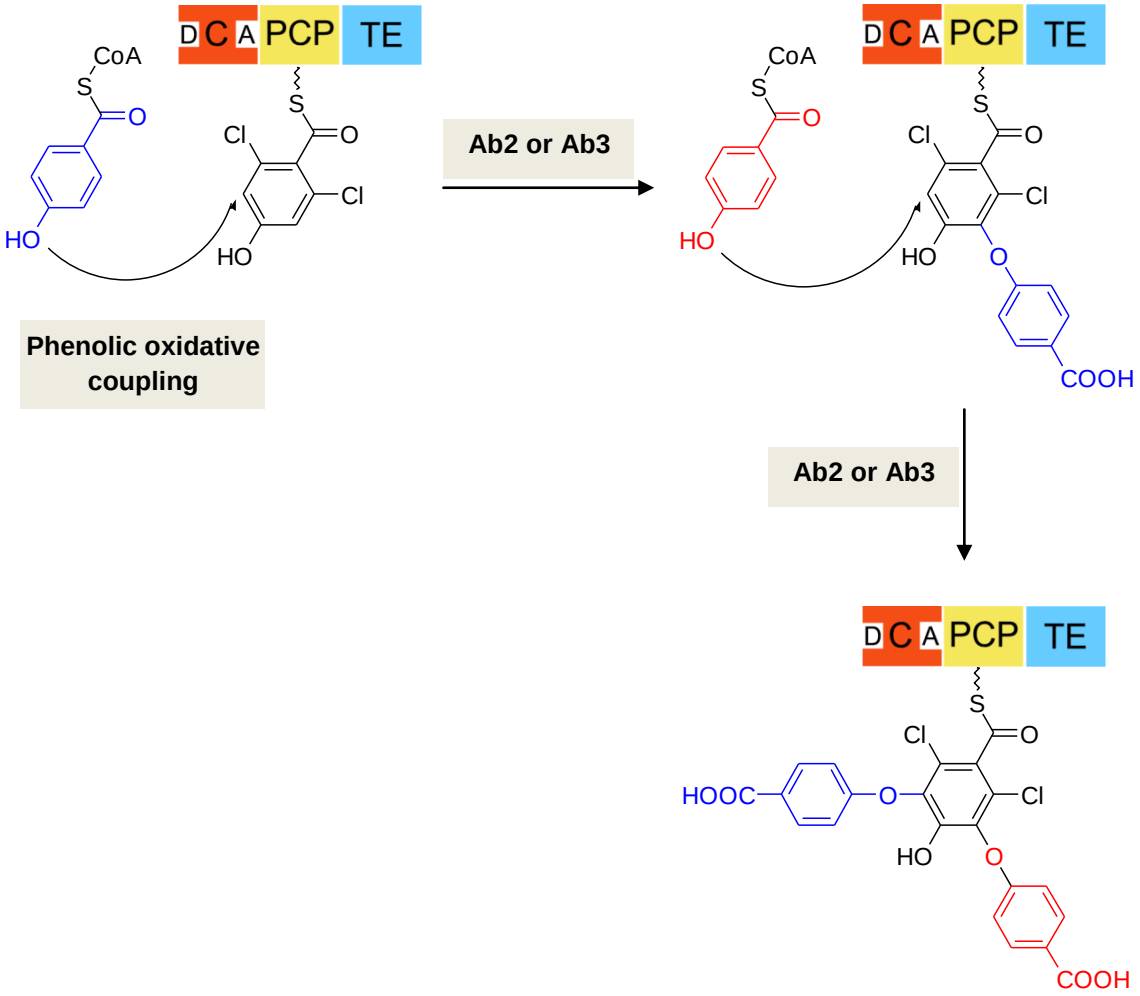


Figure 5.6.5-2: Possible phenolic oxidative coupling catalysed by Ab2 and Ab3, respectively. D = donor site; A = acceptor site.

By analogy to glycopeptides biosynthesis, the first chlorination step is suggested to be performed prior to oxidative coupling reactions (5.6.4) (Puk *et al.*, 2002). On the contrary to this, the second and final dihalogenation step is proposed to occur subsequently to the oxidative crosslinking reactions. Chlorination in *ortho*- and *para*-position to the oxygen substituted position of the outer rings of the ambigols [figure 5.1-1] is probably mediated by the $FADH_2$-dependent halogenase Ab7 (5.6.6).

This would require a previous decarboxylation step to remove the carboxyl group from the *para*-position of the oxidatively linked 4-HBA units. However, a putative decarboxylase could not be identified in the putative ambigol biosynthetic gene cluster. An involvement of Ab2 and Ab3 in the decarboxylation of crosslinked 4-HBA units in the course of coupling reactions may be conceivable. However, no specific information on this reaction step is available.

The barbamide biosynthetic pathway involves two specific enzymes BarH and BarI, which putatively mediate oxidative decarboxylation of the barbamide precursor to produce the thiazol moiety of the compound (Jones *et al.*, 2010; Flatt *et al.*, 2006; Chang *et al.*, 2002). Another type of oxidative decarboxylase, i.e. CndG was recognised only recently in the biosynthetic gene cluster of chondrochloren. This FAD-dependent enzyme catalyses the formation of the styryl moiety of the tyrosine derived natural product (Rachid *et al.*, 2010).

Concerning the signature motifs of CYP 450 enzymes described in 4.7.5, the A-(A,G)-X-(E,D)-T-motif is not completely conserved in Ab2 and Ab3, revealing a lysine residue instead of threonine. Although the threonine is conserved in many CYP 450's, there are also functionally active enzymes reported, where replacement of threonine is critical for the enzymatic activity, e.g. in OxyB from vancomycin production, Asn-240 takes the position of the conserved threonine. The amido group of the side chain of Asn-240 is located in the active site of OxyB forming a hydrogen bond with the iron-bound water molecule. It thus plays a stabilising role in the catalytic process mediated by OxyB (Woithe *et al.*, 2007; Zerbe *et al.*, 2002).

Another example is found in the biosynthetic pathway of erythromycin, which includes 6-deoxyerythronolide B synthase, a PKS, which produces the erythromycin aglycone (Khosla *et al.*, 2007). EryF, a CYP 450 enzyme performs 6(*S*)-hydroxylation of the macrocyclic polyketide aglycone 6-deoxyerythronolide B. This reaction depends on a substrate-CYP 450 interaction (Walsh & Fischbach, 2010; Isin & Guengerich, 2008). As found for Ab2 and Ab3, the threonine of motif I (4.7.5) is not absolutely conserved in EryF. It contains an alanine residue in place of threonine, which appeared to be important for dioxygen bond cleavage (Cupp-Vickery *et al.*, 1996).

It has to be mentioned that CYP 450 enzymes in general share low sequence homology. Despite that, they contain a conserved overall tertiary structure combined with regions of high flexibility for substrate positioning (Lewis & Wiseman, 2005; Graham & Peterson, 1999). As a consequence, the sole knowledge of the primary structure of Ab2 and Ab3 does not allow predictions, neither on their possible substrates nor on their catalysed reactions. Further studies are necessary including knockout experiments and heterologous expression of both CYP 450 enzymes to prove their involvement in ambigol biosynthesis and to test possible substrates by in vitro assays.

An experimental approach to inactivate a CYP 450 enzyme was managed by Wang (2008). In a knockout mutant of *Chaetomium chiversii*, the producer of the chlorinated secondary metabolite radicicol, the gene *radP* was inactivated. This gene encodes a CYP 450 epoxidase that is responsible for epoxidation of a PKS-generated double bond. As a result, a deepoxy analogue was obtained (Wang *et al.*, 2008).

5.6.6 Final halogenation steps

With regard to the successive position of *ab6* and *ab7* in the putative ambigol biosynthetic gene cluster [figure 5.1-1], a tandem action of the corresponding enzymes could be estimated. The deduced peptide sequence of Ab7 [4.7.10] is similar to prominent NRPS associated $FADH_2$-dependent halogenases, which perform tailoring halogenation in the production of cyclic and linear nonribosomal peptides, e.g. cyanopeptolins and aeruginosins [figure 1.2-1]. Highest identity (66%) was found to the halogenase McnD from *M. aeruginosa*, which produces the chlorinated heptapeptide cyanopeptolin 984 (Cadel-Six *et al.*, 2008; Rounge *et al.*, 2007; Tooming-Klunderud *et al.*, 2007). The close relationship to these NRPS-associated halogenases implies that Ab7 most likely acts on an NRPS-attached substrate. Furthermore, its catalysed reaction is expected to be the final step in ambigol biosynthesis due to the terminal location of the corresponding gene *ab7* in the assembly of putative ambigol biosynthesis genes [figure 5.1-1]. Thus, halogenase Ab7 may perform dichlorination of the aromatic moieties in the putative trimeric ambigol precursor [figure 5.6.6-1] in *ortho* and *para* to the oxygen substituted position of the outer rings. Dichlorination on phenolic or pyrrole substrates is well-established in literature. One example was given in 1.4.1.1, i.e. the production of pyoluteorin by *P. fluorescens* Pf-5 (Dorrestein *et al*., 2005). CrpH, the $FADH_2$-dependent halogenase of the cryptophycins, was proposed to generate dichlorinated cryptophycin congeners (Magarvey *et al.*, 2006). Likewise, the flavin-dependent halogenase ChlA has been proven in vivo and in vitro to be capable to mediate both halogenations in the biosynthesis of DIF-1 (Neumann *et al.*, 2010) (1.4.1).

Similar to ChlA, which utilises a free PKS product as a substrate, the $FADH_2$-dependent halogenase Rdc2 from the fungus *Pochonia chlamydosporia* acts on a resorcylic acid lactone, likely produced by two iterative PKSs (Zhou *et al.*, 2010). Furthermore, in vitro assays using recombinant Rdc2 revealed that the latter was

even capable of generating dichlorinated products, although the endogenous secondary metabolite is only monochlorinated. Rdc2 was also able to incorporate bromide instead of chlorine and to halogenate structurally divergent lactones (Zeng & Zhan, 2010).

The PCP-tethered ambigol precursor is suggested to be composed of a trimeric phenolic backbone with six chlorine substituents [figure 5.6.6-1]. The halogenase Ab7 is proposed to chlorinate the aromatic residues of the oxidatively coupled 4-HBA units. Release of the halogenated product is most likely mediated by action of the putative TE domain of Ab6 (4.7.9.4).

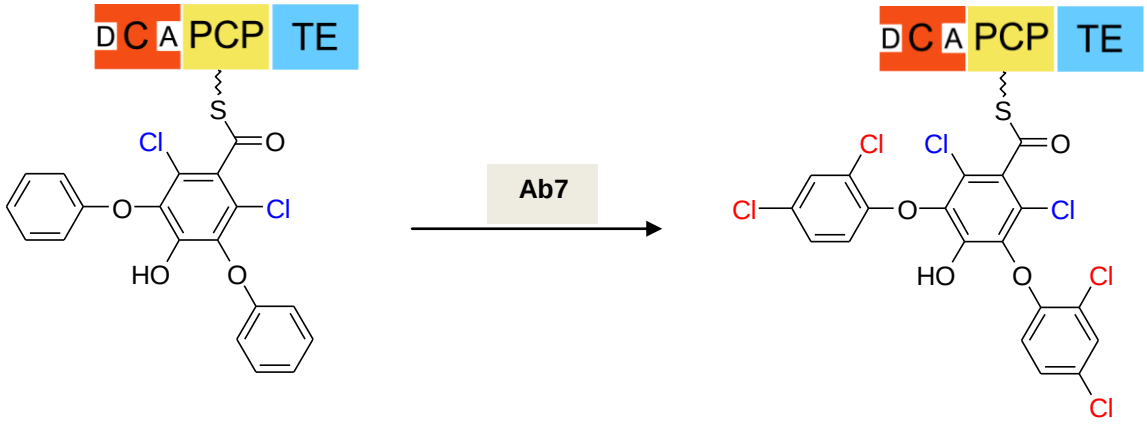


Figure 5.6.6-1: Possible reaction catalysed by the putative halogenase Ab7. It is supposed to chlorinate the outer rings of the trimeric ambigol scaffold. D = donor site; A = acceptor site.

A final decarboxylation of the middle ring would be necessary to produce ambigols. Since no decarboxylating enzyme was identified in the putative ambigol biosynthetic gene cluster, it might be possible that housekeeping enzymes are involved or that the TE domain is capable of removing the carboxyl group in the course of ester cleavage (Mortison & Sherman, 2010).

Moreover, the ambigol biosynthetic gene cluster did not contain an FAD-reductase, which is usually required for regeneration of the essential cofactor $FADH_2$ (van Pée & Patallo, 2006; van Pée, 2001; Keller *et al.*, 2000). However, within the biosynthetic gene cluster of pyrrolnitrin, which comprises four genes, also no flavin reductase was identified. In fact, the endogenous flavin reductase from *Pseudomonas fluorescens* could be even exchanged by reductases from other bacteria, e.g. Fre and SsuE from

E. coli (Unversucht *et al.*, 2005; Dorrestein *et al*., 2005; Keller *et al.*, 2000). Similarly, the halogenated indole derivative pyrroindomycin B (Ding *et al.*, 1994) was produced in vitro using SsuE (Zehner *et al.*, 2005). Therefore, a specifity between the flavin reductase and the halogenase is obviously not required (van Pee *et al.*, 2006). The chemical lability of free diffusible $FADH_2$, however suggests at least a halogenase-reductase interaction for a possible interenzymatic cofactor-transfer (Flecks *et al.*, 2008).

5.6.7 DAHP synthetase

The peptide sequence of ORF23 (4.7.11) aligns with a DAHP synthetase from *Nodularia spumigena* CCY9414 bearing 64 % identity. Same degree of identity was found to a DAHP sythetase from *Nostoc punctiforme* PCC 73102 and *Anabaena variabilis* ATCC 29413, respectively. The DAHP enzyme catalyses the stereospecific aldol-like condensation of phosphoenol pyruvate and D-erythrose 4-phosphate to give 3-deoxy-D-arabino-heptulosonate 7-phosphate, which is the initial step of the shikimate pathway. The latter represents the source of aromatic amino acids in bacteria and plants (Knaggs, 2003; Herrmann & Weaver, 1999).

DAHP synthetases are present in most bacteria, fungi and plants. Birck and Woodard separated this protein family into two classes, class I and class II, which was proposed due to their metal requirements (Birck & Woodard, 2001). An alternative classification was suggested by Jensen *et al*. (2002) dividing DAHP synthetases into two homology families ($AroA_I$ and $AroA_{II}$) (Jensen *et al.*, 2002). $AroA_{II}$ comprises herbal DAHP synthetases including higher plant proteins, but also numerous bacteria have been found to possess this class of enzyme (Gosset *et al.*, 2001). According to amino acid similarity, ORF3 presumbly belongs to class I, which is typical for DAHP synthetases from microorganisms.

It is well established that different strategies take place to control the activity of enzymes within the biosynthetic pathway for aromatic amino acids (Bentley, 1990). Regulation of DAHP synthetase activity is accomplished either transcriptionally or by feedback inhibition (Walker *et al.*, 1996). The three DAHP synthetase isoforms expressed by *E. coli* are representatives of $AroA_{I\alpha}$ (class II) and are specifically feedback-inhibited by tyrosine, phenylalanine or tryptophan, respectively. Early studies on the isoenzymes AroF, AroG, and AroH revealed that slight mutation in

AroF by replacing Leu-148 with a proline residue resulted in a tyrosine-insensitive enzyme (Weaver & Herrmann, 1990). Substitution of Lys-8 by an Asn-residue also led to a tyrosine-insensitive DAHP synthetase (Jossek *et al.*, 2001). The DAHP synthetase from *B. subtilis*, which belongs to the $AroA_{I\beta}$ family (class II), is inhibited by the shikimate pathway intermediates prephenate and chorismate (Wu *et al.*, 2003). Another example for the specifity of feedback-inhibition mechanisms is a DAHP synthetase from *Xanthomonas campestris* that has been shown to be sensitive to chorismate as an allosteric inhibitor, however less sensitive to L-tryptophan (Gosset *et al.*, 2001).

Thus, the presence of a *dahp* gene in a biosynthetic gene cluster probably encoding a specific isoform of DAHP synthetases enables a distinct control of precursor availability. This, as a consequence would be a limiting factor in the production of a natural product, i.e. the potentially toxic ambigols.

Only recently, this correlation has been shown for the balhimycin producer *A. balhimycina* (Stegmann *et al.*, 2010; Thykaer *et al.*, 2010). Balhimycin and related glycopeptide antibiotics, e.g. vancomycin and teicoplanin, contain the non-proteinogenic amino acid (*S*)-4-hydroxyphenylglycine (Hpg), which arises from the shikimate pathway (Chen *et al.*, 2001; Widboom & Bruner, 2009). This amino acid is also known from calcium-dependent antibiotic (CDA), an acidic lipopeptide from *Streptomyces coelicolor* (Strieker & Marahiel, 2009). Both biosynthetic gene clusters encode three similar proteins, a 4-hydroxymandelate synthase (HmaS), 4-hydroxymandelate oxidase (Hmo) and the 4-hydroxyphenyl-glycine transaminase (HpgT), which mediate the formation of Hpg (Strieker & Marahiel, 2009). Also both biosynthetic gene clusters revealed the presence of a *dahp* gene (Hojati *et al.*, 2002; Pelzer *et al.*, 1999). A recombinant strain of *A. balhimycina* overexpressing the *dahp* gene was able to produce significantly higher amounts of balhimycin, whereas the deletion mutant clearly revealed a reduced production of balhimycin (Thykaer *et al.*, 2010).

5.6.8 Summary of genes probably involved in ambigol biosynthesis

The putative biosynthetic gene cluster of the ambigols comprises seven genes (5.6) codifying for a chorismate lyase (Ab4), a CoA synthetase (Ab5), an NRPS-like module (Ab6), two CYP 450 enzymes (Ab2 and Ab3) and two $FADH_2$-dependent

halogenases (Ab1 and Ab7). These enzymes are most likely responsible for the production of an aromatic scaffold modified by regioselective halogenation. The central building block for ambigol A and B is proposed to be 4-HBA, which is generated by the probable chorismate lyase Ab4 (5.6.1). Activation of this aryl acid to the corresponding CoA-ester may be ascribed to the AMP-dependent CoA synthetase Ab5 (5.6.2). The thioesterified substrate is supposed to be transferred to the PCP domain of the NRPS-like module Ab6 (5.6.3) by interthiol transfer probably mediated by the starter C domain of Ab6 [figure 5.6.8-1]. Alternatively, Ab5 could mediate the adenylation of 4-HBA to 4-HBA-AMP, and subsequently catalyse the transfer of the latter onto the PCP domain of Ab6.

Genes encoding the halogenase Ab1 (5.6.4) and the CYP 450's Ab2 and Ab3 (5.6.5) are resided contiguously on the biosynthetic gene cluster. They are therefore assumed to act collectively on the PCP-bound substrate. Ab1 could be responsible for the dichlorination of PCP-tethered 4-HBA in *meta*-position to the OH-group [figure 5.6.8-1]. The trimeric ambigol backbone would emerge by action of Ab2 and Ab3 catalysing either C-O- or C-C-coupling reactions. It is also possible however that Ab2 and Ab3 both catalyse C-O-couplings. In this case, the first CYP 450 enzyme (Ab2 or Ab3) acts on a monomeric substrate only, while the second would utilise the resulting dimeric product to couple it to another 4-HBA unit. This hypothesis is in accordance with findings for the biosynthesis of vancomycin-type antibiotics, in which every single coupling step is mediated by a dedicated CYP 450 enzyme (5.6.5) (Widboom & Bruner, 2009; Walsh, 2004). Following this theory, it is suggested that the identified putative ambigol biosynthetic gene cluster is responsible only for the production of the symmetric molecule ambigol B [figure 5.6.8-1].

Decarboxylation of the two added 4-HBA units after their linkage to 2,6-dichlorobenzoyl-S-PCP is probably managed by housekeeping enzymes or in the course of the phenolic oxidative coupling reaction (5.6.5). Subsequently, Ab7 putatively performs the final halogenations in *ortho* and *para* to the oxygen substituted position of the outer rings of the ambigols (5.6.6). Release of the final product is likely mediated by the TE domain of Ab6. The latter could probably involve an integrated decarboxylation step (5.6.6). In figure 5.6.8-1, the proposed biosynthetic pathway of the ambigols A and B is summarised.

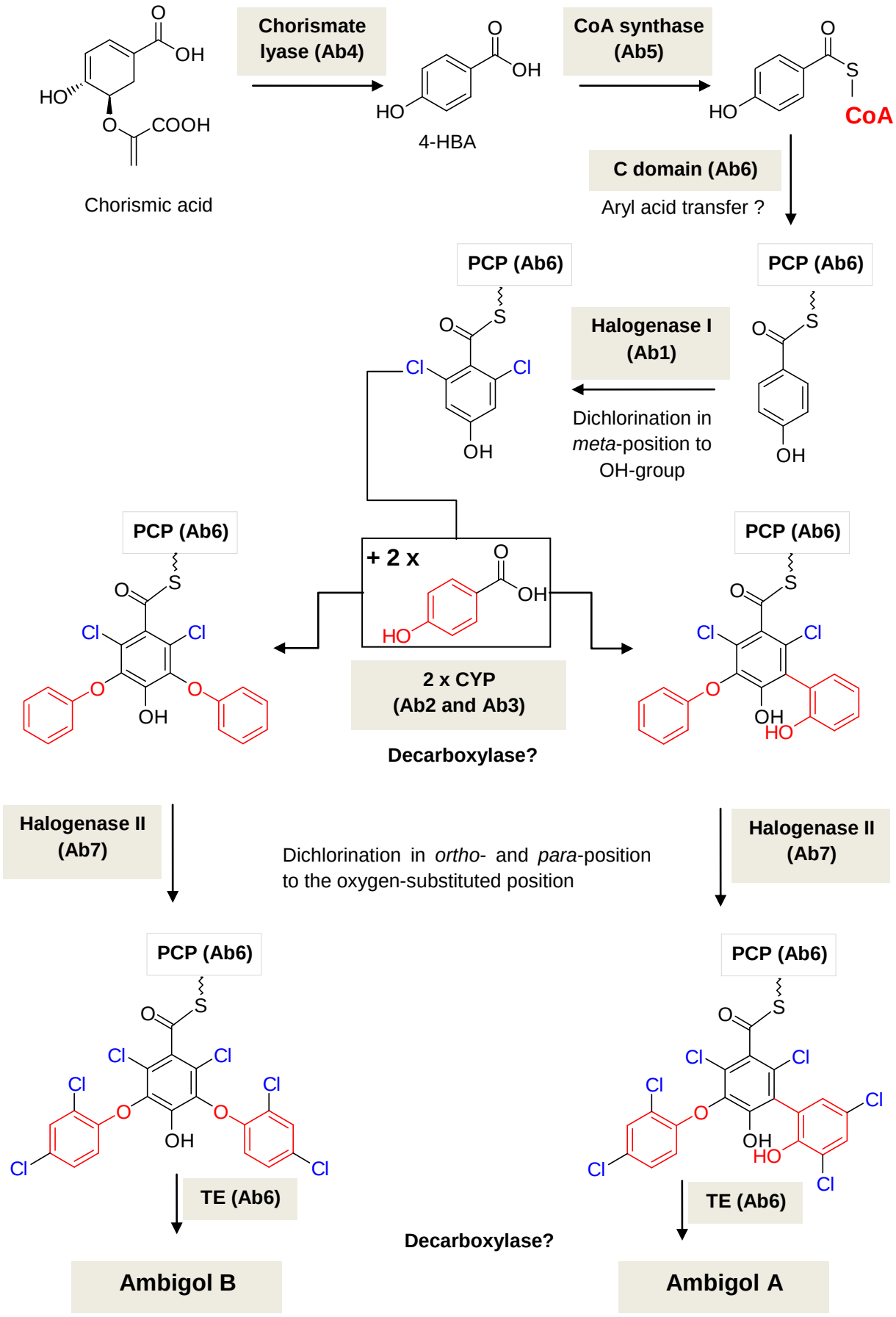


Figure 5.6.8-1: Proposed biosynthesis for ambigols A and B.

On the basis of this theory, a possible biosynthesis for ambigol A and B could be deduced, whereby the biosynthetic pathway of ambigol B is favoured to be related to the identified gene cluster. Concerning ambigol C [figure 5.1-1], the linkage of the outer phenolic units in *meta*-position to the OH-group of the central ring can not be explained by a phenolic oxidative coupling reaction. Since the delocated electron can only be present in *ortho* and *para* positions to the OH-group of the central building block, a coupling reaction in *meta*-position would be impossible (1.4.4). Thus, the formulation of mesomeric structures of the respective phenolic radicals excludes a phenolic oxidative coupling reaction in *meta*-position to an OH-group.

5.7 The putative biosynthetic gene cluster of tjipanazole D

Sequence data obtained in this project show that the arrangement of genes encoding the putative set of tjipanazole D biosynthesis enzymes is to a high extent congruent with that of rebeccamycin production (5.7.1). To date, published information is only available on the biosynthesis of the indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazole scaffold found in rebeccamycin and related compounds (Onaka, 2009; Salas & Mendez, 2009; Nakano & Omura, 2009; Sánchez *et al.*, 2006b) (1.3.2) but not on the production of the indolo[2,3-*a*]carbazole core structure as present in tjipanazole D. Therefore, the presented study may contribute to the understanding of the biosynthesis of this divergent type of indole alkaloid secondary metabolites.

The putative tjipanazole D biosynthetic gene cluster (4.8) has a size of approximately 12 kb and is boardered by the transposase gene *orf1* and the downstream located *orf4*, which was deciphered as a putative CMP/dCMP deaminase (4.8.1). The latter enzyme is likely to participate in primary metabolism (Reizer *et al.*, 1994). Five genes are probably directly ascribed to the assembly of the bisindole natural product tjipanazole D (4.8). These are: *tj1*, *tj2*, *tj3*, which codify for a putative L-tryptophan oxidase, a probable chromopyrrolic acid (CPA) synthase-like enzyme, and a putative FAD-dependent monooxygenase, respectively. Furthermore, the deduced peptide sequence of *tj4* resembles CYP 450 enzymes involved in the biosynthesis of prominent indolocarbazoles, i.e. RebP and StaP, respectively (Ryan & Drennan, 2009). The gene *tj5* most probably encodes a tryptophan 5-halogenase. All above mentioned genes are arranged in 5',3'-direction, except the halogenase gene *tj5*, which is read in 3',5'-direction. Therefore, the halogenase is predicted to be

expressed independently from the remaining enzymes of the biosynthetic gene cluster. As it was proposed for most investigated tryptophan halogenases (Zhu *et al.*, 2009; Flecks *et al.*, 2008; Yeh *et al.*, 2005; Sánchez *et al.*, 2005; Zehner *et al.*, 2005) (1.4.2), Tj5 is assumed to catalyse the first step in tjipanazole D production, the 5-chlorination of tryptophan.

Figure 5.6-1: Schematic overview of the arrangement of genes, which putatively participate in tjipanazole D-formation in *F. ambigua*. The coloured areas represent the corresponding, putative proteins. For detailed information and the putative function of the respective proteins, see table 4.8.1-1. Proteins coloured in light blue are involved in the formation of the tjipanazole D core structure. The probable halogenase Tj5 and the possible flavin reductase Orf3 are coloured in blue. The black shaded area (gene *orf2*) represents a protein of unknown function, i.e. a SAM-dependent methyltransferase. The gene cluster is boardered by *orf1* and *orf4*. TRANSPOS = transposase; TRPOXI: L-tryptophan oxidase; CPASYNTH: chromopyrrolic acid synthase-like protein; O-METTRANS = O-methyltransferase; NAD(P)HDEHYD = NAD(P)H-dehydrogenase; FADMONO = FAD-binding monooxygenase; CYP = CYP 450 enzyme; HAL = $FADH_2$-dependent halogenase; CMP/dCMP DEAMIN = CMP/dCMP deaminase, zinc-binding

The nucleotide sequences for *tj1-tj4* correspond to respective genes of the biosynthetic gene clusters of rebeccamycin (*reb*) and staurosporine (*sta)* (Onaka *et al.*, 2003; Sánchez *et al.*, 2002; Onaka *et al.*, 2002). *Orf*s related to the tjipanazole D biosynthesis and assigned genes from *reb* and *sta* biosynthetic gene clusters are listed in table 5.7.1-1.

Rebeccamycin and the related natural product staurosporine, are composed of an indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazole skeleton [figure 5.7-2], which is produced by dimerisation of two molecules of indole-3-pyruvate (IPA) imine, catalysed by StaD and RebD, to give chromopyrrolic acid (CPA) and 11,11'-dichlorochromopyrrolic acid [figure 5.7.1-1], respectively. The precursor, IPA-imine arises by action of the L-amino acid oxidases RebO and StaO, respectively. Subsequent oxidation steps with ring closure mediated by the enzyme pairs StaP/StaC and RebP/RebC, respectively, lead to the indolocarbazole core structure (Singh *et al.*, 2008; Sánchez *et al.*, 2006b; Howard-Jones & Walsh, 2006; Howard-Jones & Walsh, 2005).

By analogy, *tj1* and *tj2* are supposed to encode proteins that trigger the formation of a CPA-like intermediate [figure 5.7.3.2-1], which could be the precursor of the indolo[2,3-*a*]carbazole scaffold of tjipanazole D (Ryan *et al*., 2007).

The halogenase RebH together with the flavin reductase RebF was proven by in vitro assay to catalyse the 7-chlorination of L-tryptophan as the initiating step of rebeccamycin biosynthesis (Yeh *et al.*, 2005). Similarly, PyrH, the tryptophan 5-halogenase from *Streptomyces rugosporus* was heterologously expressed to show its halogenating activity towards free tryptophan (Zehner *et al.*, 2005). As might be expected with regard to the chlorination pattern in tjipanazole D, Tj5 reveals highest homology to PyrH. Therefore, it is reasonable to anticipate that Tj5 is the starting enzyme in tjipanazole D production (5.7.2).

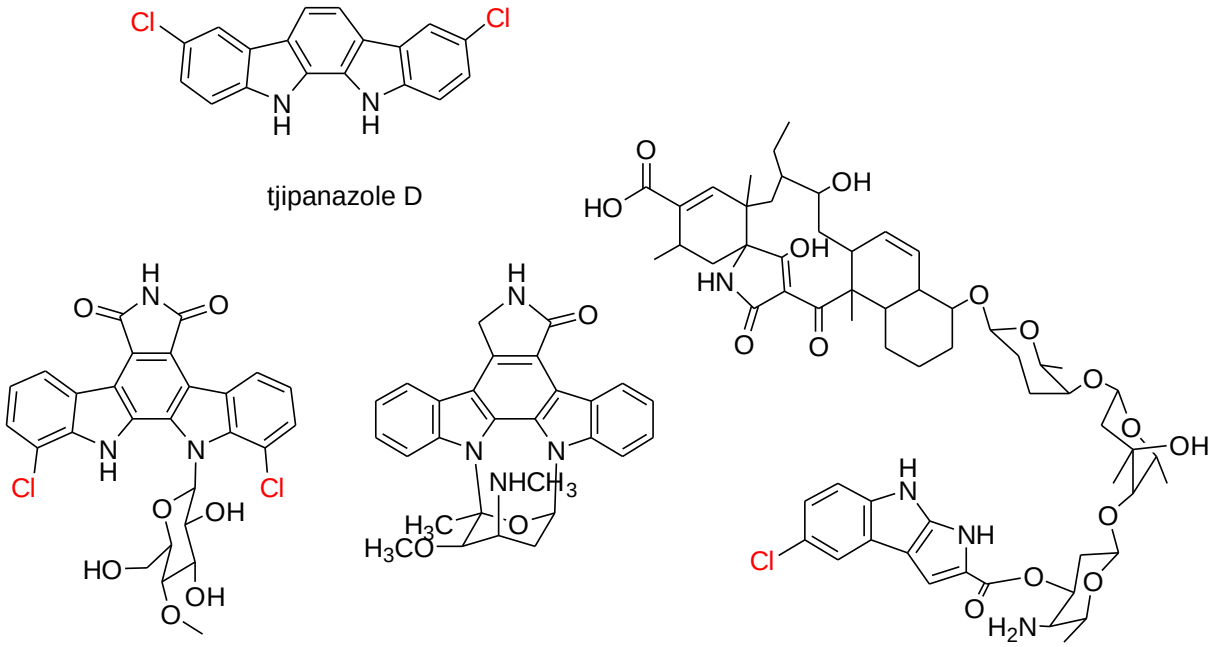


Figure 5.7-2: Indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazoles (staurosporine, rebeccamycin) and tjipanazole D. Pyrroindomycin contains, just as tjipanazole D, a tryptophan-derived moiety that is chlorinated in 5-position.

5.7.1 Analogy of tjipanazole D to indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazoles

A key step in staurosporine and rebeccamycin biosynthesis is the formation of the indolocarbazole core by intramolecular C-C bond formation and oxidative decarboxylation of chromopyrrolic acid (Ryan & Drennan, 2009).

The biosynthetic gene clusters of staurosporine, rebeccamycin as well as those for related indolocarbazoles, i.e. K252a and At2433-A1, have been cloned and characterised (Onaka *et al.*, 2003; Onaka *et al.*, 2002; Sánchez *et al.*, 2002; Chiu *et al.*, 2009b; Chiu *et al.*, 2009a; Gao *et al.*, 2006). The rebeccamycin biosynthetic gene cluster has been heterologously expressed completely and partially in *Streptomyces albus* and *E. coli*, respectively (Hyun *et al.*, 2003; Sánchez *et al.*, 2002). Furthermore, genes encoding for the biosynthesis of staurosporine were proven by heterologous expression in *Streptomyces lividans* and by gene disruption (Onaka, 2009; Makino *et al.*, 2007; Onaka *et al.*, 2002). Taking these results together with the combinatorial biosynthetic approach performed by Sánchez *et al*. (2005), four genes *staO*, *staD*, *staP*, and *staC* in *Streptomyces* sp. TP-A0274, and the homologous genes *rebO*, *rebD*, *rebP*, and *rebC* in *Lechevalieria aerocolonigenes*, encode for the biosynthetic enzymes that catalyse the formation of the indolocarbazole backbone [figure 5.7.1-1] of staurosporine and rebeccamycin, respectively (Onaka, 2009).

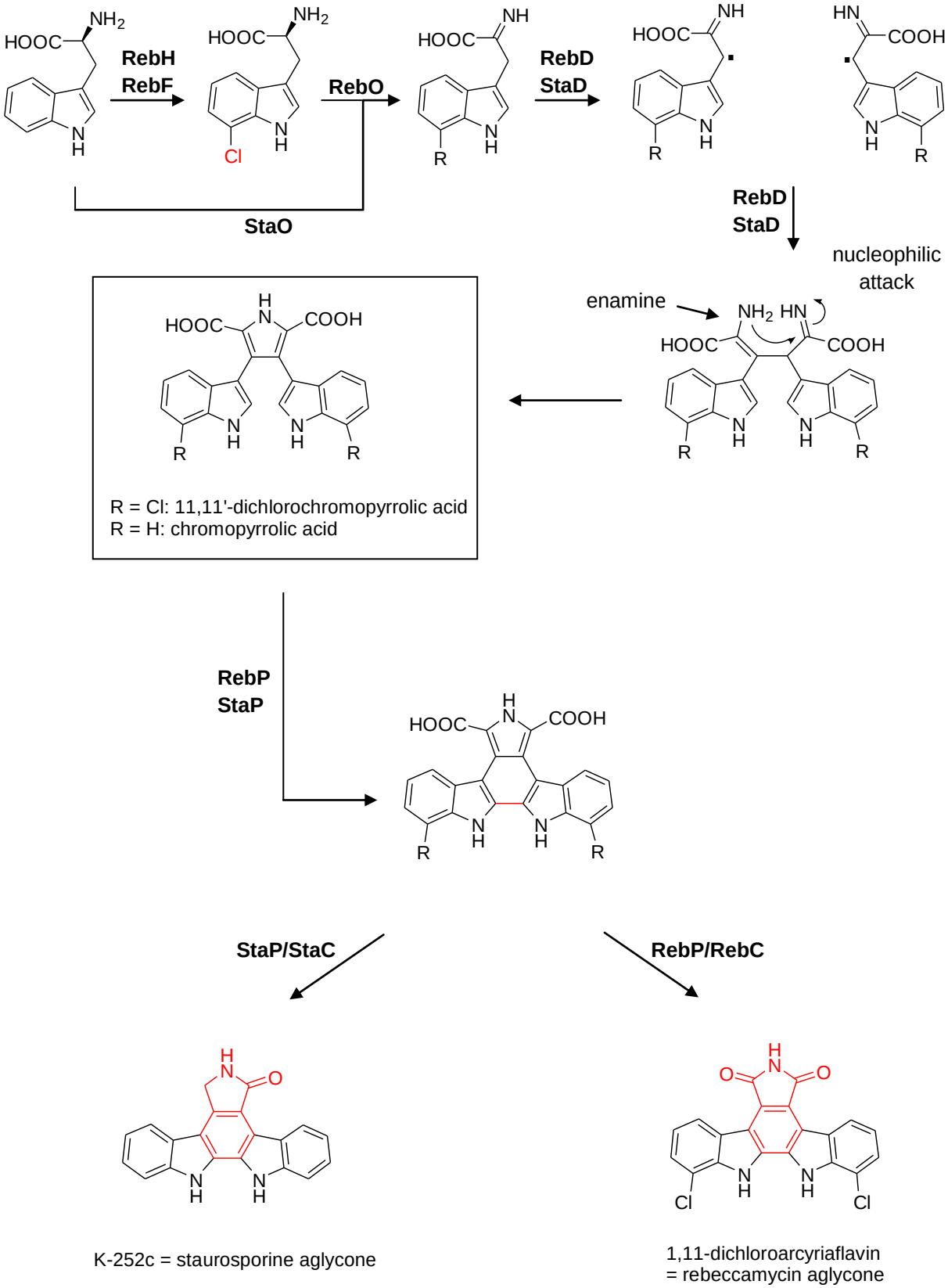


Figure 5.7.1-1: Proposed biosynthesis for the rebeccamycin- and stauroporine-aglycones, respectively, according to Ryan and Drennan (2009). For details on the involved proteins, see table 5.7.1-1.

On the basis of the structural similarity of tjipanazole D to rebeccamycin and staurosporine, respectively (5.1), an assignment of *tj* genes to possibly corresponding genes of the *reb* and *sta* biosynthetic gene clusters is suggested in table 5.7.1-1. The role of involved enzymes in staurosporine and rebeccamycin biosynthesis has been thoroughly investigated and reviewed, and therefore is well-established in literature (Onaka, 2009; Onaka *et al.*, 2003; Onaka *et al.*, 2002; Sánchez *et al.*, 2002; Sánchez *et al.*, 2005; Sánchez *et al.*, 2006b; Ryan & Drennan, 2009; Howard-Jones & Walsh, 2007; Howard-Jones & Walsh, 2006).

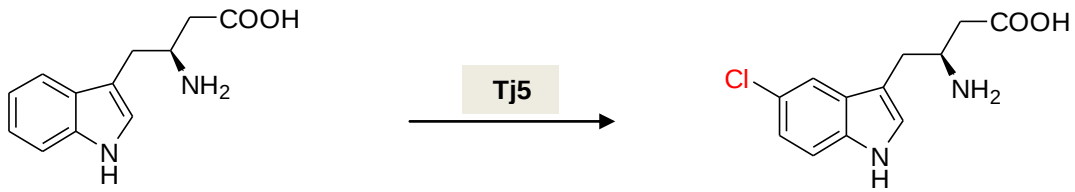
Table 5.7.1-1: Enzymes encoded by the putative tjipanazole D biosynthetic gene cluster (first column) and corresponding proteins of the biosynthetic pathways of rebeccamycin and staurosporine (second column), respectively.

Enzyme encoded by *tj* operon	Possible equivalent encoded by *reb* / *sta*	Function in rebeccamycin-/staurosporine-biosynthesis
Tj1	RebO/StaO	formation of indole-3-pyruvate imine
Tj2	RebD/StaD	formation of chromopyrrolic acid (CPA)/11,11'-CPA
Tj3	RebC/StaC	stabilising reactive intermediate formed by RebP/StaP and directing the biosynthesis towards a single product
Tj4	RebP/StaP	intramolecular aryl-aryl coupling and decarboxylation
Tj5	RebH (+RebF)	7-chlorination of tryptophan in rebeccamycin biosynthesis

5.7.2 Possible start of tjipanazole D assembly: chlorination of L-tryptophan

The putative $FADH_2$-dependent halogenase Tj5 (4.8.8) shares 49% identity with the tryptophan 5-halogenase PyrH from *Streptomyces rugosporus* (Zehner *et al.*, 2005). It also resembles KtzR from *Kutzneria* sp. 744 (48%), which performs 6-chlorination of the precursor 7-dichloro-L-tryptophan in the kutznerides biosynthetic pathway (1.4.1.2) (Fujimori *et al.*, 2007). Furthermore, an identity of 41% to the tryptophan 7-halogenase PrnA from pyrrolnitrin biosynthesis (Dong *et al.*, 2005) was found. Based on sequence homology, an involvement of Tj5 in 5-chlorination of free tryptophan

[figure 5.7.2-1] is proposed and is postulated to be the first step in the production of tjipanazole D.



L-Trp **1** 5-chloro-L-Trp **2**

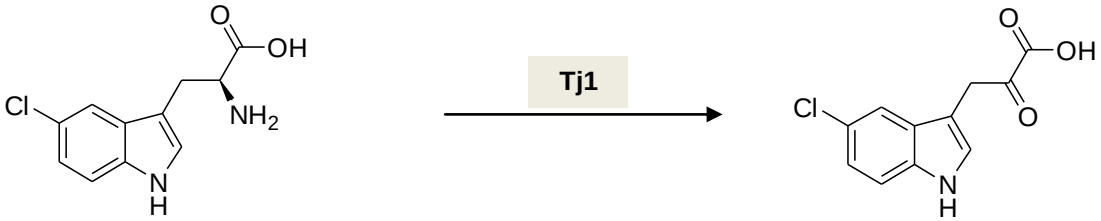
Figure 5.7.2-1: L-Trp (**1**) is proposed to be chlorinated in 5-position to give the chlorinated analogue (**2**). This reaction is likely catalysed by Tj5, which shows homology to the tryptophan 5-halogenase PyrH from *Streptomyces rugosporus*.

5.7.3 Biosynthesis of the indolo [2,3-*a*]carbazole core of tjipanazole D

5.7.3.1 Possible oxidation of 5-chloro-L-tryptophan

The translated peptide sequence of gene *tj1* aligns with several amine oxidases, among which RebO from *L. aerocolonigenes* revealed highest identity to Tj1 (4.8.2). This FAD-dependent enzyme was characterised as an L-tryptophan oxidase, which converts 7-chloro-L-tryptophan to 7-chloro-indole-3-pyruvic acid (IPA) imine (Nishizawa *et al.*, 2005). Relying on sequence homology, Tj1 is like RebO probably specific for the L-amino acid enantiomer. In addition, RebO was shown to exhibit selectivity to the chlorinated derivative 7-chloro-L-tryptophan, which is oxidised more efficiently than the non-chlorinated precursor (Sánchez *et al.*, 2006b; Nishizawa *et al.*, 2005). An identical reaction is presumed for StaO from the biosynthesis of staurosporine. StaO however acts on non-chlorinated L-tryptophan to yield indole-3-pyruvic acid imine (Ryan & Drennan, 2009).

For Tj1, a slightly divergent mechanism has to be postulated [figure 5.7.3.1-1], since tjipanazole D in contrast to rebeccamycin and staurosporine lacks the upper imide heterocycle. Therefore, it might be anticipated that in a probable Tj1-catalysed reaction, indole-3-pyruvic acid is formed, congruent with the product of TdiD from terrequinone A biosynthesis in *Aspergillus nidulans* (Ryan & Drennan, 2009; Balibar *et al.*, 2007).



5-chloro-L-Trp

2

5-chloro-IPA

3

Figure 5.7.3.1-1: Intermediate **2** (figure 5.7.2-1) is proposed to be oxidatively desaminated to give **3**.

5.7.3.2 Putative dimerisation of 5-chloro-IPA

Immediately downstream of *tj1*, the gene *tj2* (4.8.3) was identified, probably codifying for a protein, which belongs to the StaD-family (Asamizu *et al.*, 2006). It shows highest identity (46%) to the prototype of this family, the CPA-synthase StaD from *Streptomyces* sp. TP-A0274. This heme-containing enzyme from the biosynthesis of staurosporine has been heterologously expressed in *E. coli* and was shown to catalyse the coupling of two molecules of indole-3-pyruvic acid imine to give chromopyrrolic acid, which is the key intermediate in the assembly of the indolocarbazole framework (Ryan & Drennan, 2009; Howard-Jones & Walsh, 2006; Asamizu *et al.*, 2006).

In case of RebD, a tandem mechanism with RebO, whose gene is located upstream of that related to RebD, was postulated (Nishizawa *et al.*, 2005; Howard-Jones & Walsh, 2005) (5.7.3.1). Herein, the reactive species indole-3-pyruvate imine, rather than indole-3-pyruvic acid is the preferred substrate of RebD and is thought to be channeled directly to this enzyme (Howard-Jones & Walsh, 2005). StaD and RebD mediate the formation of chromopyrrolic acid and 11,11'-dichlorochromopyrrolic acid, respectively, by linkage of the β-carbons of two molecules of indole-pyruvic acid imine (Ryan & Drennan, 2009; Howard-Jones & Walsh, 2005; Nishizawa *et al.*, 2005). Very recently, the homologous genes *inkO* and *inkD* from the biosynthetic gene cluster of the staurosporine analogue K252a have been heterologously coexpressed in *Streptomyces albus* J1074 yielding chromopyrrolic acid (Chae *et al.*, 2009).

In contrast to the indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazole and violacein biosyntheses (Onaka, 2009; Balibar & Walsh, 2006), which utilise indole-3-pyruvic acid imine as a precursor, Tj2 is proposed to act on 5-chloro-IPA. In this reaction Tj2 might catalyse a radical coupling of the aliphatic side chains of two phenylpyruvate units (**4**) to produce a coupled phenylpyruvate dimer (**5**). The latter represents a completely symmetric molecule with α-keto acid side chains [figure 5.7.3.2-1].

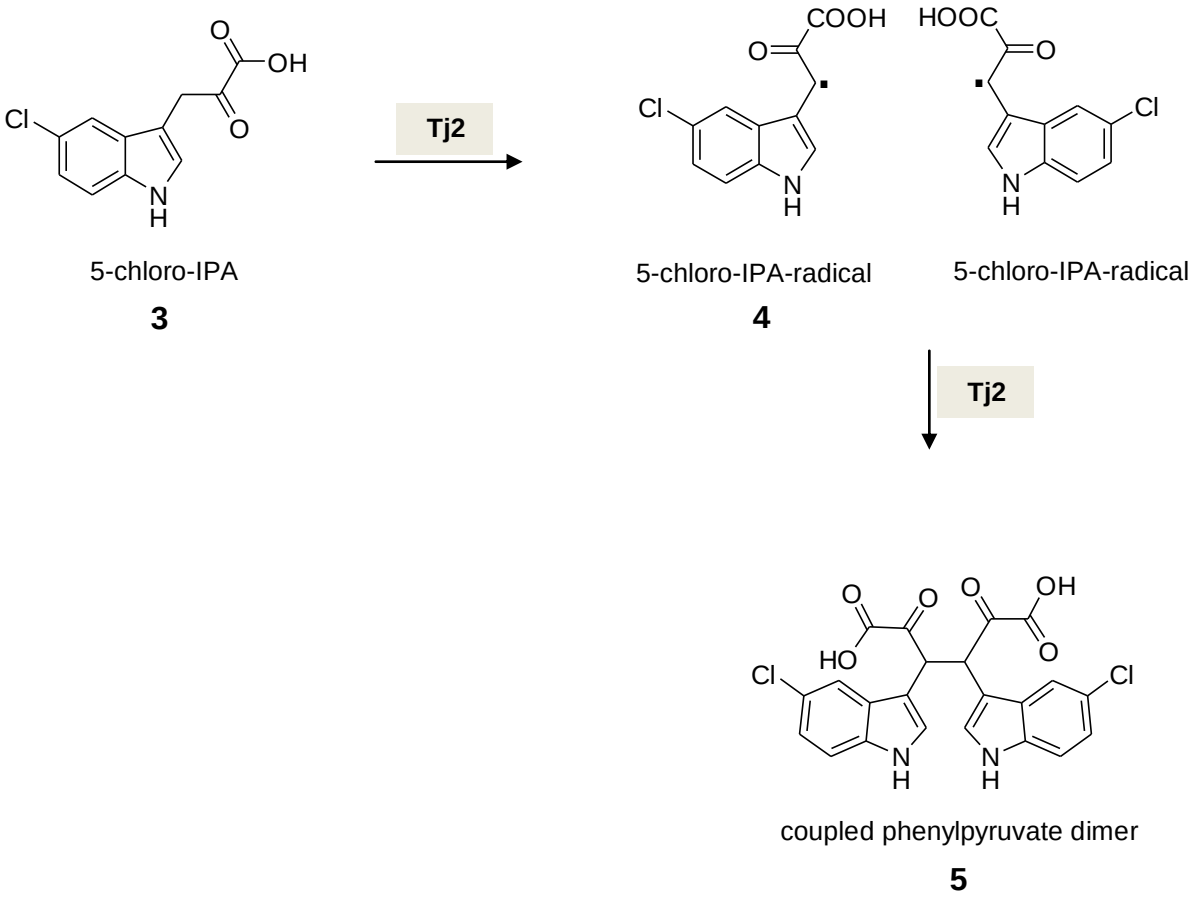
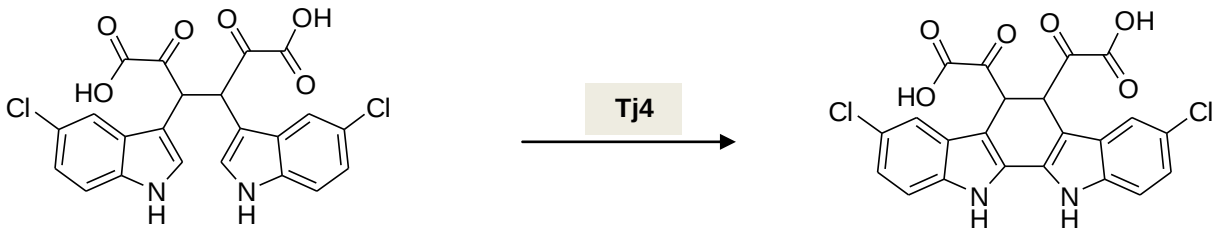


Figure 5.7.3.2-1: Possible dimerisation of 5-chloro-IPA (**4**) by a possible radical mechanism catalysed by Tj2 during tjipanazole D biosynthesis.

5.7.3.3 Possible intramolecular ary-aryl coupling

Although differences in the indolocarbazole scaffold exist between tjipanazole D (1.3.2) and indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazoles [figure 5.7.2], all these natural compounds keep the same interesting structural feature, i.e. the aryl-aryl coupling between the two indole rings of the tryptophan building blocks. Crosslinking between two aryl moieties is often catalysed by CYP 450 enzymes (1.4.4).

The primary structure of Tj4 resembles StaP *Streptomyces* sp. TP-A0274 with 45% identity and also aligns with RebP from *L. aerocolonigenes* revealing 42% identity (4.8.7). RebP from the biosynthetic gene cluster of rebeccamycin and its analogue StaP (staurosporine biosynthesis) represent functionally equal proteins (Sánchez *et al.*, 2005). They catalyse the formation of the indolocarbazole core of rebeccamycin and staurosporine, respectively, via an intramolecular C-C bond formation in the bisindole intermediate chromopyrrolic acid. An analogous coupling reaction can be formulated for tjipanazole D biosynthesis, in which Tj3 probably catalyses an intramolecular aryl-aryl coupling of the phenylpyruvate dimer **5** [figure 5.7.3.3-1].



coupled phenylpyruvate dimer **5**

chlorinated bisindole intermediate **6**

Figure 5.7.3.3-1: Possible intramolecular aryl-aryl coupling in tjipanazole D biosynthesis, probably catalysed by the putative CYP 450 enzyme Tj3.

During their coexpression experiments using rebeccamycin genes and genes from other microorganisms, Sánchez *et al*. discovered that StaP and RebP may be replaced by one another without any change in the core structure of yielded metabolites. Secondly, the initial chlorination of tryptophan is not essential for downstream biosynthetic enzymes, which all can utilise non-chlorinated or even alternately chlorinated substrates (Sánchez *et al.*, 2005). The crystal structure of StaP implies that this enzyme exhibits a peroxidase-like mechanism to mediate C-C-coupling via an indole cation radical intermediate (Makino *et al.*, 2007).

With regard to the order of genes in the *tj*-operon [table 4.8.1-1; figure 5.6-1], it should be considered that the gene *tj3*, encoding a putative FAD-binding monooxygenase, is located prior to *tj4*. Therefore, an alternative biosynthetic

pathway is possible, in which a reduction of the C_2-side chain of intermediate **5**, probably catalysed by Tj3, precedes intramolecular aryl-aryl coupling.

5.7.3.4 Putative side chain reduction of the bisindole intermediate

The deduced peptide sequence of *tj3* reveals 38% identity with a probable FAD-binding monooxygenase from *Methylobacterium radiotolerans* JCM 2831 and a putative 2-polyprenyl-6-methoxyphenol hydroxylase from *Sorangium cellulosum* 'So ce 56'. Further, Tj3 aligns with StaC and RebC only with 24% and 28% identity, respectively (4.8.6). By contrast, StaC and RebC, which act on similar substrates and thereby catalyse a similar reaction step [figure 5.7.3.4-1], share 65 % sequence identity.

As outlined in 5.7.3.3, Tj4 is proposed to be responsible for an intramolecular aryl-aryl coupling reaction in tjipanazole D biosynthesis. Herein, it is supposed to act collectively with Tj3 to generate the indolo[2,3-*a*]carbazole scaffold of tjipanazole D [figure 5.7-2]. The formation of this final product would require a reduction of the α-keto acid side chains in structure **6** [figure 5.7.3.3-1], perhaps via oxidative decarboxylation. In addition, a final decarboxylation step and aromatisation of the middle ring would be necessary.

In the biosynthesis of rebeccamycin and staurosporine, the intramolecular radical coupling between the two indole rings of CPA is mediated by the CYP 450 enzymes RebP and StaP, respectively [figure 5.7.3.4-1] (Howard-Jones & Walsh, 2007; Makino *et al.*, 2007). This step is followed by an oxidative decarboxylation, which is suggested to be also catalysed by the respective CYP450's in a concerted action with their corresponding enzymes RebC and StaC, respectively. The latter two proteins contain sequence similarity to flavin-dependent oxygenases, as does Tj3 [table 4.8.6-1]. Both, RebC and StaC play an important role in receiving and directing the outcome of the CYP 450 enzymes RebP and StaP, respectively, towards the formation of a single biosynthesis product [figure 5.7.3.4-1] (Ryan & Drennan, 2009; Ryan *et al.*, 2007; Howard-Jones & Walsh, 2006; Sánchez *et al.*, 2005).

A possible mechanism for the formation of the rebeccamycin aglycone 11,11'-arcyriaflavin A (Howard-Jones & Walsh, 2006) through the FAD-dependent enzyme

RebC was proposed by Ballou (2007). It suggests a sequential hydroxylation with spontaneous decarboxylation at C5 and C7 of the pyrrole ring (Ballou, 2007).

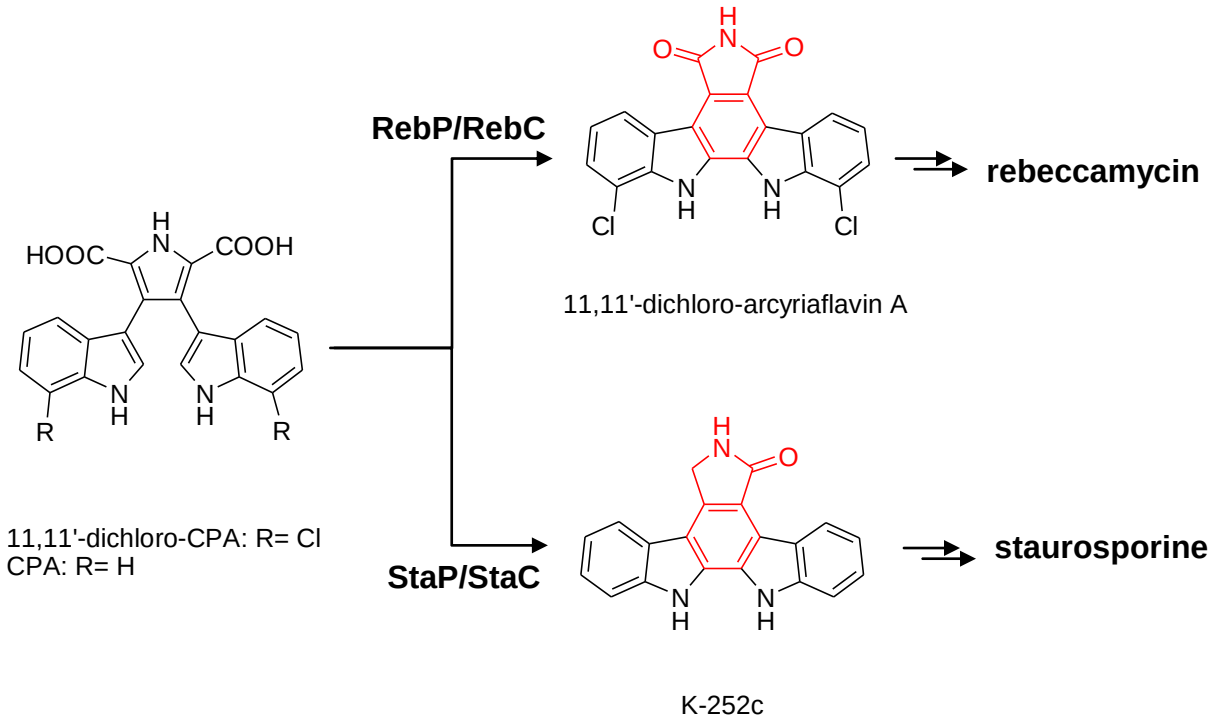


Figure 5.7.3.4-1: Collective action of the enzyme pairs RebP/RebC and StaP/StaC to produce the rebeccamycin and staurosporine aglycone, respectively (Howard-Jones & Walsh, 2006).

It has to be noted that all of the characterised four biosynthetic gene clusters of indolo[2,3-a]pyrrolo[3,4-c]carbazoles i.e. rebeccamycin (*reb*), staurosporine (*sta*), K252a (*ink*) and At2433-A1 (*atm*) contain a similar enzyme pair that is sought to produce the indolocarbazole aglycone of the respective compounds (RebC/RebP, StaC/StaP, AtmC/AtmP and InkE/InkP) (5.7.1) (Ryan & Drennan, 2009).

Wether and how decarboxylation and subsequent aromatisation of intermediate **6** [figure 5.7.3.4-2] is carried out in tjipanazole D biosynthesis, has to be investigated further.

5.7.4 Two orfs with unknown function

Orf2 and *orf3*, which are resided next to each other were deciphered as a probable O-methyltransferase (4.8.4) and a putative NAD(P)H dehydrogenase (4.8.5), respectively. Their function in the biosynthesis of tjipanazole D is not clear at this point of time.

Concerning Orf3 (4.8.5), a possible function is proposed with respect to the location of its related gene (*orf3*) adjacently to gene *tj3* that possibly codifies for an FAD-binding monooxygenase. Such two-component systems are very common, e.g. the alkane sulfonate monooxygenase system from *E. coli* (SsuD and SsuE) or the chlorophenol 4-monooxygenase system (TftD and TftC) of *Burkholderia cepacia* AC1100 (Webb *et al*., 2010). Tj3 revealed 33 % identity to a pentachlorophenol 4-monooxygenase from *Stigmatella aurantiaca* DW4/3-1 [table 4.8.6-1]. A two-component system, i.e. Orf3/Tj3 might be one possibility.

Predictions on putative functions of proteins based on amino acid similarity are limited. For instance, RebF from rebeccamycin biosynthesis revealed highest identity with a putative FMN:NADH oxidoreductase from *Streptomyces violaceoruber* (Sánchez *et al*., 2002). Despite that, the enzyme RebF was shown to be essential for catalytic activity of the $FADH_2$-dependent halogenase RebH, and thus is necessary, in order to supply reduced FAD to the halogenase (Yeh *et al*., 2005). Therefore, Orf3 could be responsible for FAD reduction either. Furthermore, the *E. coli* flavin mononucleotide (FMN) reductase SsuE was also applied for in vitro experiments with the $FADH_2$-dependent halogenase PltA (Dorrestein *et al*., 2005). In this regard, a probable two-component system consisting of Orf3 and Tj5 is conceivable.

Hence, further studies are required to clarify the role Orf3 and its possible interaction with Tj3 or Tj5, respectively.

5.7.5 Summary of the proposed biosynthetic pathway for tjipanazole D

Similar to the biosynthesis of rebeccamycin, five genes are probably involved in the production of the tjipanzole D molecule [figure 5.7.5-1]. Following this analogy, the first catalysed step is postulated to be the chlorination of L-tryptophan in 5-position, possibly catalysed by the $FADH_2$-dependent halogenase Tj5 (Yeh *et al.*, 2005; Zehner *et al.*, 2005). By contrast to rebeccamycin biosynthesis, the product of Tj1, an

amine oxidase, is proposed to be 5-chloro-IPA, similar to the product of TdiD in terrequinone A assembly (Balibar *et al.*, 2007; Nishizawa *et al.*, 2005). Tj2 shows similarity to RebD and StaD (4.8.3), which are responsible for the formation of CPA and 11,11'-dichloro-CPA as the key intermediates in staurosporine and rebeccamycin production, respectively (Nishizawa *et al.*, 2005). Tj2 is supposed to catalyse a similar reaction via radical coupling of the aliphatic side chains of the 5-chloro-IPA-molecules (Asamizu *et al.*, 2006; Howard-Jones & Walsh, 2005). Tj4 as a probable analogue of StaP and RebP is proposed to mediate an intramolecular aryl-aryl coupling reaction (Howard-Jones & Walsh, 2007). As suggested for RebC and StaC, Tj3 is suggested to act concertedly with Tj4 receiving its possible reactive intermediate, in order to initiate the formation of a stable final product, i.e. tjipanazole D.

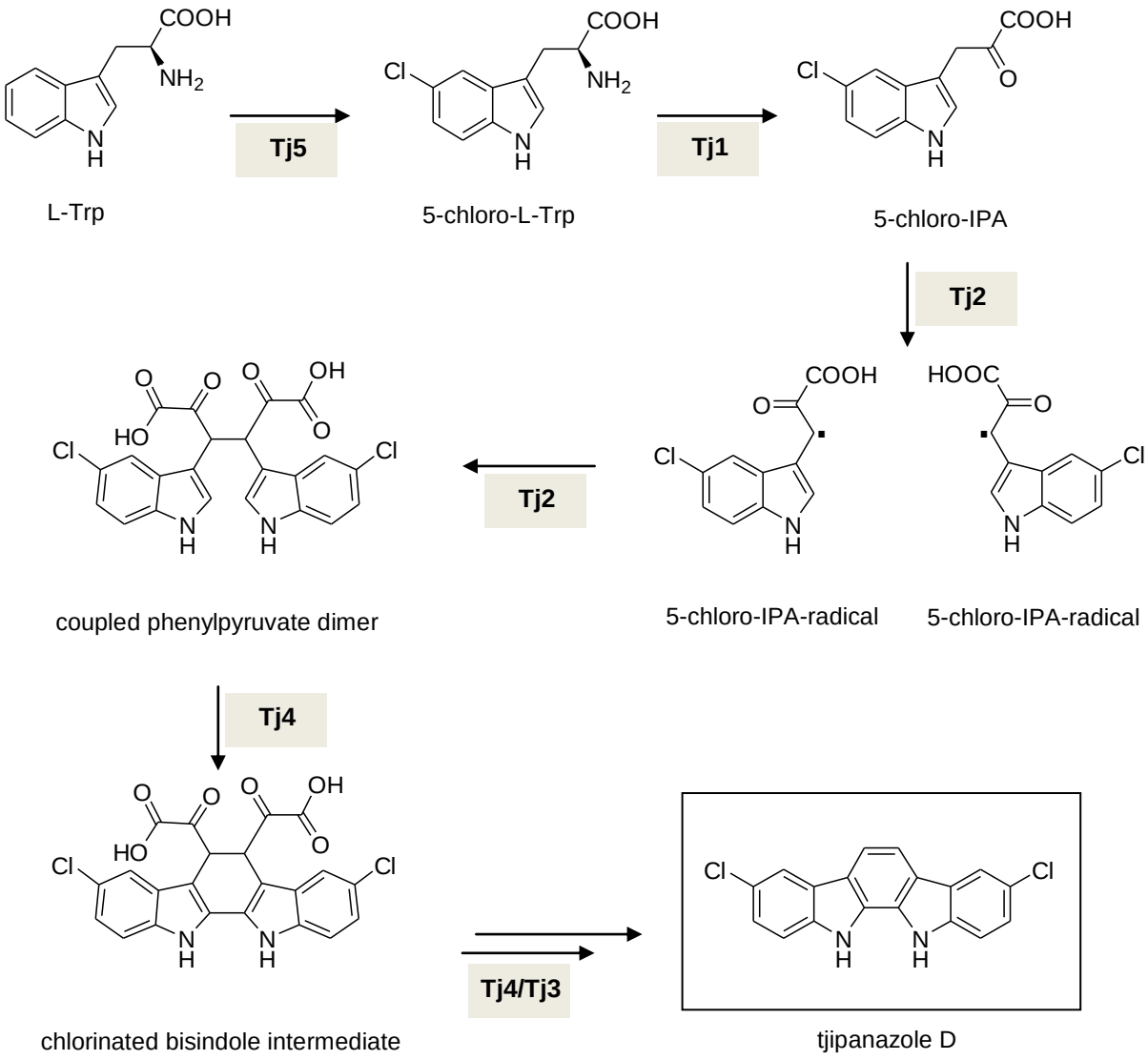


Figure 5.7.5-1: Proposed biosynthesis for tjipanazole D.

6. Summary

Cyanobacteria are Gram negative, photoautotrophic bacteria, which are distinguished by an impressive array of chemically diverse natural products. The latter are frequently products of nonribosomal peptide synthetases and polyketide synthases like the liver toxins microcystins and the antitumor compound cryptophycin 1.

The here investigated cyanobacterial strain of *Fischerella ambigua* was shown to produce the phenolic compounds ambigols A-C and the indole alkaloid tjipanazole D.

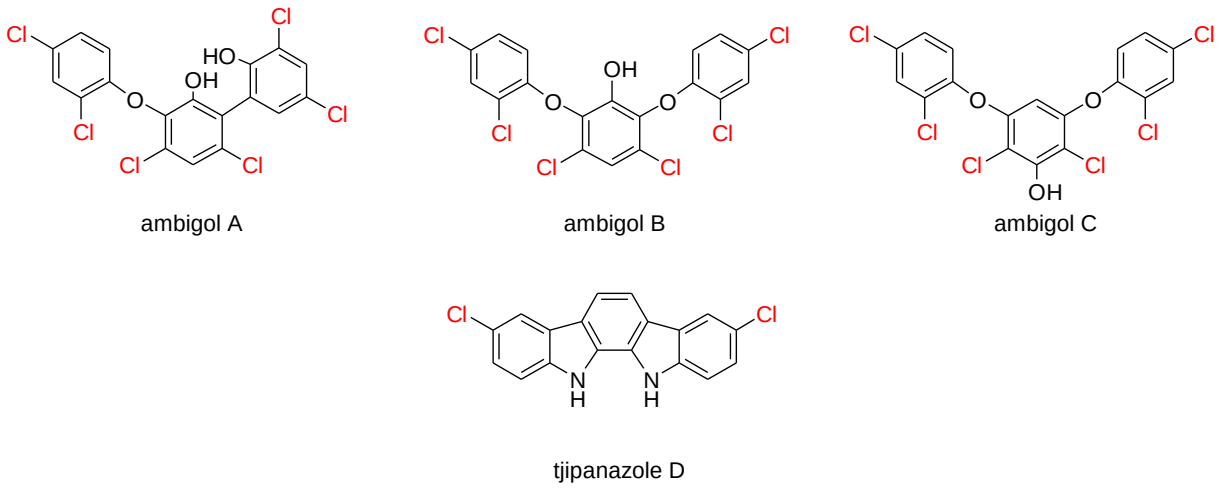


Figure 6-1: Structures of the polychlorinated phenolic ethers ambigols A-C and the indolocarbazole tjipanazole D, produced by the terrestrial cyanobacterium *Fischerella ambigua*.

The ambigol core structure is characterised by ether and aryl-aryl bridges which are strongly reminiscent of those found in glycopeptide antibiotics of the vancomycin-family. Based on this analogy, the ambigol skeleton was hypothesised to be generated by an oxidative, radical mechanism (i.e. radical coupling reaction) carried out by CYP 450 enzymes. With respect to the phenolic character of the involved ambigol monomers, each bearing a single OH-group, their biosynthesis is likely related to the shikimate and not to the polyketide pathway. In this regard, the ambigol biosynthetic pathway would represent a link between primary metabolite production and the conversion of these simple building blocks to complex natural products. This correlation is likewise established for glycopeptides, e.g. balhimycin. The chlorination

of the phenolic components are proposed to be performed by halogenases on carrier-bound substrates, as judged from reports for related compounds with aromatic moieties, e.g. chondrochloren. The indolocarbazole tjipanazole D is composed of two indole units linked via a central six-membered ring. The aromatic framework of this compound resembles that of related indolocarbazole natural products, for instance that of rebeccamycin and staurosporine, whose biosyntheses have been thoroughly investigated. Regarding the structural analogy of tjipanazole D to these latter metabolites, it was hypothesised that tjipanazole D arises from two tryptophan units. These are proposed to be linked via oxidative metabolism, catalysed by heme-containing enzymes. In this regard, the biosynthetic pathway of tjipanazole D is supposed to follow a similar logic as that described above for the ambigols.

The current work focused on sequence information related to the ambigol and tjipanazole D biosynthetic gene clusters and its annotation using bioinformatic tools. First, a fosmid library of the *F. ambigua* genome had to be established. For this purpose a method for DNA isolation from the filamentous cyanobacterial strain had to be developed, which proved to be rather difficult due to interfering polysaccharides produced by the bacterium. The procedure also had to include the removal of DNA resulting from an associated *Pseudomonas stutzeri* strain, in order to prevent this DNA to be integrated in the genomic library.

Subsequently, the genomic library was screened for halogenase genes and fosmid E8 was identified as possibly harbouring ambigol related biosynthetic genes. Fosmid E8 was then sequenced completely and comprised significant, however only partial information on a putative ambigol biosynthetic gene cluster. The missing data related to the putative CYP 450 enzyme Ab2 and the $FADH_2$-dependent halogenase Ab1, respectively. Information on these additional genes were subsequently obtained from a 454 sequencing of the *F. ambigua* genome, in that contig 00522 with a size of approximately 123 kb confirmed and completed the sequences found on fosmid E8. Using bioinformatic tools the sequence information was analysed for the putative function of the respective proteins.

Thus, the putative ambigol biosynthetic gene cluster (size = 13 kb) comprises seven genes encoding enzymes that would be expected regarding the structural features of ambigols. These enzymes were a 4-hydroxybenzoate (4-HBA) synthetase that is possibly involved in the formation of the starter molecule via the shikimate pathway,

an AMP-ligase putatively activating the starter molecule, an NRPS module that may serve as a carrier protein for the activated starter unit. Two CYP 450 enzymes and two halogenases are proposed to mediate the following biosynthesis steps, i.e. phenolic oxidative coupling and aromatic moiety chlorination. With respect to the symmetric character of ambigol B and the presence of two dedicated CYP 450 enzymes, the hypothesis is supported that the identified biosynthetic gene cluster is responsible for the production of ambigol B only.

The putative tjipanazole D biosynthetic gene cluster (size = 12 kb) was identified on contig 15287, which has a size of approximately 59 kb. It was also obtained by 454 sequencing and bears all sequence information on genes probably related to the tjipanazole D production. Based on the highly conserved biosynthetic pathways for the indolocarbazoles rebeccamycin, staurosporine, K252a and At2433-A1, an assignment of all genes of the *tj* operon was possible. Similar to the rebeccamycin biosynthetic gene cluster, a set of five genes is putatively responsible for tjipanazole D formation. These are proposed to encode an L-tryptophan halogenase, an L-tryptophan oxidase, a chromopyrrolic acid synthase-like protein, a CYP 450 enzyme and an FAD-binding monooxygenase. Whereas the halogenase is proposed to chlorinate free tryptophan, the other four enzymes most likely mediate the assembly of the indolocarbazole skeleton.

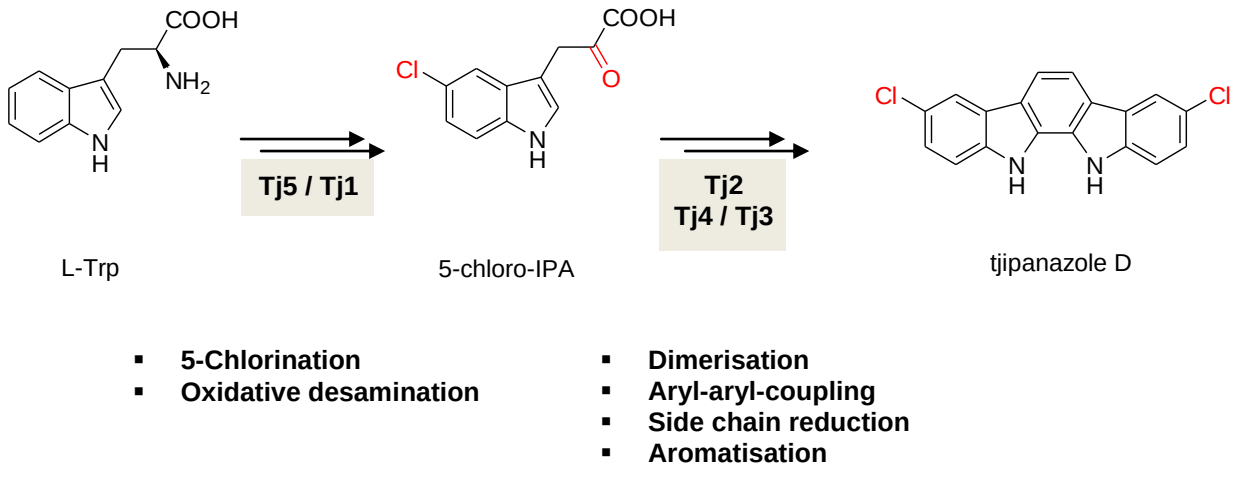


Figure 6-2: Scheme showing possible biosynthesis intermediates of tjipanazole D production. Tj1= tryptophan amine oxidase; Tj2 = Chromopyrrolic acid synthase-like protein; Tj3 = FAD-binding monooxygenase; Tj4 = CYP 450 enzyme; Tj5= $FADH_2$-dependent halogenase.

The current study thus sheds light on sequence information which is putatively related to the biosynthesis of unusual cyanobacterial metabolites, i.e. ambigols and tjipanazole D, respectively. Further studies will have to proof the relationship of the here presented information with secondary metabolite production in *F. ambigua*.

7. References

Adams DG. **2001.** How do cyanobacteria glide? *Microbiology Today* **28:** 131-133.

Adams DG, Duggan PS. **2008.** Cyanobacteria-bryophyte symbioses. *J.Exp.Bot.* **59:** 1047-1058.

Anderson JLR, Chapman SK. **2006.** Molecular mechanisms of enzyme-catalysed halogenation. *Mol. BioSyst.* **2:** *350–357*.

Anizon F, Golsteyn RM, Leonce S, Pfeiffer B, Prudhomme M. **2009.** A three-step synthesis from rebeccamycin of an efficient checkpoint kinase 1 inhibitor. *Eur.J.Med.Chem.* **44:** 2234-2238.

Anizon F, Moreau P, Sancelme M, Laine W, Bailly C, Prudhomme M. **2003.** Rebeccamycin analogues bearing amine substituents or other groups on the sugar moiety. *Bioorg.Med.Chem.* **11:** 3709-3722.

Arcuri HA, Zafalon GF, Marucci EA, Bonalumi CE, da Silveira NJ, Machado JM, de AW, Jr., Palma MS. **2010.** SKPDB: a structural database of shikimate pathway enzymes. *BMC Bioinformatics.* **11:** 12.

Asamizu S, Kato Y, Igarashi Y, Furumai T, Onaka H. **2006.** Direct formation of chromopyrrolic acid from indole-3-pyruvic acid by StaD, a novel hemoprotein in indolocarbazole biosynthesis. *Tetrahedron Lett.* **47:** 473-475.

Austin MB, Saito T, Bowman ME, Haydock S, Kato A, Moore BS, Kay RR, Noel JP. **2006.** Biosynthesis of *Dictyostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nat Chem.Biol.* **2:** 494-502.

Bai R, Schwartz RE, Kepler JA, Pettit GR, Hamel E. **1996.** Characterization of the interaction of cryptophycin 1 with tubulin: binding in the *Vinca* domain, competitive inhibition of dolastatin 10 binding, and an unusual aggregation reaction. *Cancer Res.* **56:** 4398-4406.

Bailly C, Qu X, Anizon F, Prudhomme M, Riou JF, Chaires JB. **1999.** Enhanced binding to DNA and topoisomerase I inhibition by an analog of the antitumor antibiotic rebeccamycin containing an amino sugar residue. *Mol.Pharmacol.* **55:** 377-385.

Balibar CJ, Howard-Jones AR, Walsh CT. **2007.** Terrequinone A biosynthesis through L-tryptophan oxidation, dimerization and bisprenylation. *Nat.Chem.Biol.* **3:** 584-592.

Balibar CJ, Walsh CT. **2006.** In vitro biosynthesis of violacein from L-tryptophan by the enzymes VioA-E from *Chromobacterium violaceum*. *Biochemistry* **45:** 15444-15457.

Ballou DP. **2007.** Crystallography gets the jump on the enzymologists. *Proc.Natl.Acad.Sci U.S.A* **104:** 15587-15588.

Balunas MJ, Linington RG, Tidgewell K, Fenner AM, Urena LD, Togna GD, Kyle DE, Gerwick WH. **2010.** Dragonamide E, a modified linear lipopeptide from *Lyngbya majuscula* with antileishmanial activity. *J.Nat Prod* **73:** 60-66.

Bao W, Sheldon PJ, Hutchinson CR. **1999a.** Purification and properties of the *Streptomyces peucetius* DpsC beta-ketoacyl:acyl carrier protein synthase III that specifies the propionate-starter unit for type II polyketide biosynthesis. *Biochemistry* **38:** 9752-9757.

Bao W, Sheldon PJ, Wendt-Pienkowski E, Hutchinson CR. **1999b.** The *Streptomyces peucetius* dpsC gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide. *J.Bacteriol.* **181:** 4690-4695.

Beck ZQ, Aldrich CC, Magarvey NA, Georg GI, Sherman DH. **2005.** Chemoenzymatic synthesis of cryptophycin/arenastatin natural products. *Biochemistry* **44:** 13457-13466.

Bentley R. **1990.** The shikimate pathway--a metabolic tree with many branches. *Crit Rev.Biochem.Mol.Biol.* **25:** 307-384.

Bergendahl V, Linne U, Marahiel MA. **2002.** Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur.J.Biochem.* **269:** 620-629.

Billi D, Grilli CM, Paolozzi L, Ghelardini P. **1998.** A method for DNA extraction from the desert cyanobacterium *Chroococcidiopsis* and its application to identification of *ftsZ*. *Appl.Environ.Microbiol.* **64:** 4053-4056.

Birck MR, Woodard RW. **2001.** *Aquifex aeolicus* 3-deoxy-D-manno-2-octulosonic acid 8-phosphate synthase: a new class of KDO 8-P synthase? *J.Mol.Evol.* **52:** 205-214.

Bischoff D, Bister B, Bertazzo M, Pfeifer V, Stegmann E, Nicholson GJ, Keller S, Pelzer S, Wohlleben W, Sussmuth RD. **2005.** The biosynthesis of vancomycin-type glycopeptide antibiotics--a model for oxidative side-chain cross-linking by oxygenases coupled to the action of peptide synthetases. *Chembiochem.* **6:** 267-272.

Bischoff D, Pelzer S, Bister B, Nicholson GJ, Stockert S, Schirle M, Wohlleben W, Jung G, Sussmuth RD. **2001.** The biosynthesis of vancomycin-type glycopeptide antibiotics-the order of the cyclization steps . *Angew.Chem.Int.Ed Engl.* **40:** 4688-4691.

Böhm GA, Pfleiderer W, Boger P, Scherer S. **1995.** Structure of a novel oligosaccharide-mycosporine-amino acid ultraviolet A/B sunscreen pigment from the terrestrial cyanobacterium *Nostoc commune*. *J.Biol.Chem.* **270:** 8536-8539.

Bonjouklian R, Smitka TA, Doolin LE, Molloy RM, Debono M, Shaffer SA, Moore RE, Stewart JB, Patterson GML. **1991.** Tjipanazoles, new antifungal agents from the blue-green-alga *Tolypothrix tjipanasensis*. *Tetrahedron* **47:** 7739-7750.

Bonmatin JM, Laprevote O, Peypoux F. **2003.** Diversity among microbial cyclic lipopeptides: iturins and surfactins. Activity-structure relationships to design new bioactive agents. *Comb.Chem.High Throughput.Screen.* **6:** 541-556.

Buedenbender S, Rachid S, Muller R, Schulz GE. **2009.** Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases. *J.Mol.Biol.* **385:** 520-530.

Bush JA, Long BH, Catino JJ, Bradner WT. **1987.** Production and biological activity of rebeccamycin, a novel antitumor agent. *J.Antibiot.* **40:** 668-678.

Butler A, Sandy M. **2009.** Mechanistic considerations of halogenating enzymes. *Nature* **460:** 848-854.

Cadel-Six S, Dauga C, Castets AM, Rippka R, Bouchier C, Tandeau dM, Welker M. **2008.** Halogenase genes in nonribosomal peptide synthetase gene clusters of *Microcystis* (cyanobacteria): sporadic distribution and evolution. *Mol.Biol.Evol.* **25:** 2031-2041.

Campos-Bermudez VA, Bologna FP, Andreo CS, Drincovich MF. **2010.** Functional dissection of *Escherichia coli* phosphotransacetylase structural domains and analysis of key compounds involved in activity regulation. *FEBS J.* **277:** 1957-1966.

Cane DE, Walsh CT. **1999.** The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem.Biol.* **6:** R319-R325.

Cardellina JH, Marner FJ, Moore RE. **1979.** Seaweed dermatitis: structure of lyngbyatoxin A. *Science* **204:** 193-195.

Carté B, Faulkner DJ. **1981.** Polybrominated diphenyl ethers from *Dysidea herbacea*, *Dysidea chlorea* and *Phyllospongia foliascens*. *Tetrahedron* **37:** 2335-2339.

Chae CS, Park JS, Chung SC, Kim TI, Lee SH, Yoon KM, Shin J, Oh KB. **2009.** Production of chromopyrrolic acid by coexpression of *inkOD* in a heterologous host *Streptomyces albus*. *Bioorg.Med.Chem.Lett.* **19:** 1581-1583.

Chang KH, Xiang H, Dunaway-Mariano D. **1997.** Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate:coenzyme A ligase. *Biochemistry* **36:** 15650-15659.

Chang Z, Flatt P, Gerwick WH, Nguyen VA, Willis CL, Sherman DH. **2002.** The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* **296:** 235-247.

Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH. **2004.** Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J.Nat Prod* **67:** 1356-1367.

Chen H, Tseng CC, Hubbard BK, Walsh CT. **2001.** Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc.Natl.Acad.Sci U.S A* **98:** 14901-14906.

Chen H, Walsh CT. **2001.** Coumarin formation in novobiocin biosynthesis: beta-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI. *Chem.Biol.* **8:** 301-312.

Chiu HT, Chen YL, Chen CY, Jin C, Lee MN, Lin YC. **2009a.** Molecular cloning, sequence analysis and functional characterization of the gene cluster for biosynthesis of K-252a and its analogs. *Mol.Biosyst.* **5:** 1180-1191.

Chiu HT, Lin YC, Lee MN, Chen YL, Wang MS, Lai CC. **2009b.** Biochemical characterization and substrate specificity of the gene cluster for biosyntheses of K-252a and its analogs by in vitro heterologous expression system of *Escherichia coli*. *Mol.Biosyst.* **5:** 1192-1203.

Choi GG, Bae MS, Ahn CY, Oh HM. **2008.** Induction of axenic culture of *Arthrospira* (*Spirulina*) *platensis* based on antibiotic sensitivity of contaminating bacteria. *Biotechnol.Lett.* **30:** 87-92.

Chooi YH, Tang Y. **2010.** Adding the lipo to lipopeptides: do more with less. *Chem.Biol.* **17:** 791-793.

Clarke L, Carbon J. **1976.** A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. *Cell* **9:** 91-99.

Cryle MJ, De Voss JJ. **2004.** Carbon-carbon bond cleavage by cytochrome P450(BioI) (CYP107H1). *Chem.Commun.(Camb.)* 86-87.

Cryle MJ, Meinhart A, Schlichting I. **2010.** Structural characterization of OxyD, a cytochrome P450 involved in beta-hydroxytyrosine formation in vancomycin biosynthesis. *J.Biol.Chem.* **285:** 24562-24574.

Cryle MJ, Schlichting I. **2008.** Structural insights from a P450 carrier protein complex reveal how specificity is achieved in the P450(BioI) ACP complex. *Proc.Natl.Acad.Sci U.S A* **105:** 15696-15701.

Cupp-Vickery JR, Han O, Hutchinson CR, Poulos TL. **1996.** Substrate-assisted catalysis in cytochrome P450eryF. *Nat.Struct.Biol.* **3:** 632-637.

De Philippis R, Sili C, Paperi R, Vincenzini M. **2001.** Exopolysaccharide-producing cyanobacteria and their possible exploitation: A review. *J.Appl.Phycol.* **13:** 293-299.

Dewick PM. **2009.** *Medicinal Natural Products: A Biosynthetic Approach*.

Ding W, Williams DR, Northcote P, Siegel MM, Tsao R, Ashcroft J, Morton GO, Alluri M, Abbanat D, Maiese WM, . 1994. Pyrroindomycins, novel antibiotics produced by *Streptomyces rugosporus* sp. LL-42D005. I. Isolation and structure determination. *J.Antibiot.(Tokyo)* **47:** 1250-1257.

Ding Y, Seufert WH, Beck ZQ, Sherman DH. **2008.** Analysis of the cryptophycin P450 epoxidase reveals substrate tolerance and cooperativity. *J.Am.Chem.Soc.* **130:** 5492-5498.

Dittmann E, Neilan BA, Borner T. **2001.** Molecular biology of peptide and polyketide biosynthesis in cyanobacteria. *Appl.Microbiol.Biotechnol.* **57:** 467-473.

Dittmann E, Wiegand C. **2006.** Cyanobacterial toxins--occurrence, biosynthesis and impact on human affairs. *Mol.Nutr.Food Res.* **50:** 7-17.

Dong C, Flecks S, Unversucht S, Haupt C, van Pee KH, Naismith JH. **2005.** Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* **309:** 2216-2219.

Dong C, Kotzsch A, Dorward M, van Pee KH, Naismith JH. **2004.** Crystallization and X-ray diffraction of a halogenating enzyme, tryptophan 7-halogenase, from *Pseudomonas fluorescens*. *Acta Crystallogr.D.Biol.Crystallogr.* **60:** 1438-1440.

Dong G, Kim YI, Golden SS. **2010a.** Simplicity and complexity in the cyanobacterial circadian clock mechanism. *Curr.Opin.Genet.Dev.* **20:** 619-625.

Dong G, Yang Q, Wang Q, Kim YI, Wood TL, Osteryoung KW, van Oudenaarden A, Golden SS. **2010b.** Elevated ATPase activity of KaiC applies a circadian checkpoint on cell division in Synechococcus elongatus. *Cell* **140:** 529-539.

Donia MS, Ravel J, Schmidt EW. **2008.** A global assembly line for cyanobactins. *Nat Chem.Biol.* **4:** 341-343.

Dorrestein PC, Yeh E, Garneau-Tsodikova S, Kelleher NL, Walsh CT. **2005.** Dichlorination of a pyrrolyl-S-carrier protein by $FADH_2$-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc.Natl.Acad.Sci.U.S.A* **102:** 13843-13848.

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

Du L, Shen B. **2001.** Biosynthesis of hybrid peptide-polyketide natural products. *Curr.Opin.Drug Discov.Devel.* **4:** 215-228.

Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH. **2004.** Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem.Biol.* **11:** 817-833.

Ehling-Schulz M, Bilger W, Scherer S. **1997.** UV-B-induced synthesis of photoprotective pigments and extracellular polysaccharides in the terrestrial cyanobacterium *Nostoc commune*. *J.Bacteriol.* **179:** 1940-1945.

Eustaquio AS, Gust B, Luft T, Li SM, Chater KF, Heide L. **2003.** Clorobiocin biosynthesis in *Streptomyces*: identification of the halogenase and generation of structural analogs. *Chem.Biol.* **10:** 279-288.

Falch BS, Konig GM, Wright AD, Sticher O, Angerhofer CK, Pezzuto JM, Bachmann H. **1995.** Biological activities of cyanobacteria: evaluation of extracts and pure compounds. *Planta Med.* **61:** 321-328.

Faulkner J, Unson MD, Bewley CA. **1994.** The chemistry of some sponges and their symbionts. *Pure Appl.Chem.* **66:** 1983-1990.

Fernández-González B, Martinez-Ferez IM, Vioque A. **1998.** Characterization of two carotenoid gene promoters in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochim.Biophys.Acta* **1443:** 343-351.

Finking R, Marahiel MA. **2004.** Biosynthesis of nonribosomal peptides. *Annu.Rev.Microbiol.* **58:** 453-488.

Fiore MF, Moon DH, Tsai SM, Lee H, Trevors JT. **2000.** Miniprep DNA isolation from unicellular and filamentous cyanobacteria. *J.Microbiol.Methods* **39:** 159-169.

Flatt PM, Gautschi JT, Thacker RW, Musafija-Girt M, Crews P, Gerwick WH. **2005.** Identification of the cellular site of polychlorinated peptide biosynthesis in the marine sponge *Dysidea* (*Lamellodysidea*) *herbacea* and symbiotic cyanobacterium *Oscillatoria spongeliae* by CARD-FISH analysis. *Marine Biology* **147:** 761-774.

Flatt PM, O'Connell SJ, McPhail KL, Zeller G, Willis CL, Sherman DH, Gerwick WH. **2006.** Characterization of the initial enzymatic steps of barbamide biosynthesis. *J.Nat.Prod.* **69:** 938-944.

Flecks S, Patallo EP, Zhu X, Ernyei AJ, Seifert G, Schneider A, Dong C, Naismith JH, van Pee KH. **2008.** New insights into the mechanism of enzymatic chlorination of tryptophan. *Angew.Chem.Int.Ed Engl.* **47:** 9533-9536.

Flodin C, Whitfield FB. **1999a.** 4-hydroxybenzoic acid: a likely precursor of 2,4,6-tribromophenol in *Ulva lactuca*. *Phytochemistry* **51:** 249-255.

Flodin C, Whitfield FB. **1999b.** Biosynthesis of bromophenols in marine algae. *Water Sci.Technol.* **40:** 53-58.

Freundlich JS, Wang F, Tsai HC, Kuo M, Shieh HM, Anderson JW, Nkrumah LJ, Valderramos JC, Yu M, Kumar TR, Valderramos SG, Jacobs WR, Jr., Schiehser GA, Jacobus DP, Fidock DA, Sacchettini JC. **2007.** X-ray structural analysis of *Plasmodium falciparum* enoyl acyl carrier protein reductase as a pathway toward the optimization of triclosan antimalarial efficacy. *J.Biol.Chem.* **282:** 25436-25444.

Freundlich JS, Wang F, Vilcheze C, Gulten G, Langley R, Schiehser GA, Jacobus DP, Jacobs WR, Jr., Sacchettini JC. **2009.** Triclosan derivatives: towards potent inhibitors of drug-sensitive and drug-resistant *Mycobacterium tuberculosis*. *ChemMedChem.* **4:** 241-248.

Fujii K, Harada KI, Suzuki M, Kondo F, Ikai Y, Carmichael WW, Ivonen K. **1996.** Occurrence of novel cyclic peptides together with microcystins from toxic cyanobacteria, *Anabaena* species. *Intergovermental Oceanographic Commission of UNESCO.* **Harmful and toxic algal blooms.Yasumoto, T., Oshima, Y., and Fukuyo, Y.(editors):** 559-562.

Fujimori DG, Hrvatin S, Neumann CS, Strieker M, Marahiel MA, Walsh CT. **2007.** Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc.Natl.Acad.Sci.U.S.A* **104:** 16498-16503.

Fujimori DG, Walsh CT. **2007.** What's new in enzymatic halogenations. *Curr.Opin.Chem.Biol.* **11:** 553-560.

Gallagher DT, Mayhew M, Holden MJ, Howard A, Kim KJ, Vilker VL. **2001.** The crystal structure of chorismate lyase shows a new fold and a tightly retained product. *Proteins* **44:** 304-311.

Gao Q, Zhang C, Blanchard S, Thorson JS. **2006.** Deciphering indolocarbazole and enediyne aminodideoxypentose biosynthesis through comparative genomics: insights from the AT2433 biosynthetic locus. *Chem.Biol.* **13:** 733-743.

Garrity G, Boone DR, Castenholz RW. **2001.** *Bergeys Manual of Systematic Bacteriology 1. The Archea and the Deeply Branching and Phototrophic Bacteria*.

Gehring AM, Bradley KA, Walsh CT. **1997.** Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* **36:** 8495-8503.

Gehring AM, Mori I, I, Perry RD, Walsh CT. **1998a.** The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* **37:** 17104.

Gehring AM, Mori I, Walsh CT. **1998b.** Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37:** 2648-2659.

Geib N, Woithe K, Zerbe K, Li DB, Robinson JA. **2008.** New insights into the first oxidative phenol coupling reaction during vancomycin biosynthesis. *Bioorg.Med.Chem.Lett.* **18:** 3081-3084.

Ginard M, Lalucat J, Tummler B, Romling U. **1997.** Genome organization of *Pseudomonas stutzeri* and resulting taxonomic and evolutionary considerations. *Int.J.Syst.Bacteriol.* **47:** 132-143.

Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y, Bena G, Hannibal L, Fardoux J, Kojadinovic M, Vuillet L, Lajus A, Cruveiller S, Rouy Z, Mangenot S, Segurens B, Dossat C, Franck WL, Chang WS, Saunders E, Bruce D, Richardson P, Normand P, Dreyfus B, Pignol D, Stacey G, Emerich D, Vermeglio A, Medigue C, Sadowsky M. **2007.** Legumes symbioses: absence of *nod* genes in photosynthetic bradyrhizobia. *Science* **316:** 1307-1312.

Gkelis S, Harjunpaa V, Lanaras T, Sivonen K. **2005.** Diversity of hepatotoxic microcystins and bioactive anabaenopeptins in cyanobacterial blooms from Greek freshwaters. *Environ.Toxicol.* **20:** 249-256.

Glombitza KW, Koch M, Eckhardt G. **1977.** Chlorophlorethols from *Laminaria ochroleuca*. *Phytochemistry* **16:** 796-798.

Goldstone JV, Jonsson ME, Behrendt L, Woodin BR, Jenny MJ, Nelson DR, Stegeman JJ. **2009.** Cytochrome P450 1D1: a novel CYP1A-related gene that is not transcriptionally activated by PCB126 or TCDD. *Arch.Biochem.Biophys.* **482:** 7-16.

Gosset G, Bonner CA, Jensen RA. **2001.** Microbial origin of plant-type 2-keto-3-deoxy-D-arabino-heptulosonate 7-phosphate synthases, exemplified by the chorismate- and tryptophan-regulated enzyme from *Xanthomonas campestris*. *J.Bacteriol.* **183:** 4061-4070.

Graham SE, Peterson JA. **1999.** How similar are P450s and what can their differences teach us? *Arch.Biochem.Biophys.* **369:** 24-29.

Gribble GW. **2010.** Diphenyl ethers. *Fortschr.Chem.Org.Naturst.* **91:** 273-281.

Grobbelaar N, Huang TC, Lin HY, Chow TJ. **1986.** Dinitrogen-fixing endogenous rhythm in *Synechococcus* Rf-1. *FEMS Microbiol.Lett.* **37:** 173-177.

Grody WW, Nakamura RM, Kiechle FL, Strom C. **2010.** *Molecular Diagnostics: Techniques and Applications for the Clinical Laboratory*.

Grove CL, Gunsalus RP. **1987.** Regulation of the aroH operon of *Escherichia coli* by the tryptophan repressor. *J.Bacteriol.* **169:** 2158-2164.

Guengerich FP. **2001.** Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem.Res.Toxicol.* **14:** 611-650.

Gugger MF, Hoffmann L. **2004.** Polyphyly of true branching cyanobacteria (Stigonematales). *Int.J.Syst.Evol.Microbiol.* **54:** 349-357.

Hadatsch B, Butz D, Schmiederer T, Steudle J, Wohlleben W, Sussmuth R, Stegmann E. **2007.** The biosynthesis of teicoplanin-type glycopeptide antibiotics: assignment of p450 mono-oxygenases to side chain cyclizations of glycopeptide a47934. *Chem.Biol.* **14:** 1078-1089.

Hammer PE, Burd W, Hill DS, Ligon JM, van Pee K. **1999.** Conservation of the pyrrolnitrin biosynthetic gene cluster among six pyrrolnitrin-producing strains. *FEMS Microbiol.Lett.* **180:** 39-44.

Hammer PE, Hill DS, Lam ST, van Pee KH, Ligon JM. **1997.** Four genes from *Pseudomonas fluorescens* that encode the biosynthesis of pyrrolnitrin. *Appl.Environ.Microbiol.* **63:** 2147-2154.

Handayani D, Edrada RA, Proksch P, Wray V, Witte L, Van Soest RW, Kunzmann A, Soedarsono. **1997.** Four new bioactive polybrominated diphenyl ethers of the sponge *Dysidea herbacea* from West Sumatra, Indonesia. *J.Nat Prod* **60:** 1313-1316.

Hanefeld U, Floss HG, Laatsch H. **1994.** Biosynthesis of the marine antibiotic pentabromopseudilin.1. the Benzene-Ring. *J.Org.Chem.* **59:** 3604-3608.

Harrigan GG, Goetz GH, Luesch H, Yang S, Likos J. **2001.** Dysideaprolines A-F and barbaleucamides A-B, novel polychlorinated compounds from a *Dysidea* species. *J.Nat Prod* **64:** 1133-1138.

Heath RJ, White SW, Rock CO. **2001.** Lipid biosynthesis as a target for antibacterial agents. *Prog.Lipid Res.* **40:** 467-497.

Heemstra JR, Jr., Walsh CT. **2008.** Tandem action of the O_2- and $FADH_2$-dependent halogenases KtzQ and KtzR produce 6,7-dichlorotryptophan for kutzneride assembly. *J.Am.Chem.Soc.* **130:** 14024-14025.

Hellweger FL. **2010.** Resonating circadian clocks enhance fitness in cyanobacteria *in silico*. *Ecol.Model.* **221:** 1620-1629.

Hentschel U, Bringmann G. **2010.** Search for new drugs from natural sources. Marine sponges: cornucopia of the marine. *Pharm.Unserer Zeit* **39:** 62-66.

Hentschel U, Usher KM, Taylor MW. **2006.** Marine sponges as microbial fermenters. *FEMS Microbiol.Ecol.* **55:** 167-177.

Herdman M, Janvier M, Rippka R, Stanier RY. **1979.** Genome size of cyanobacteria. *Journal of General Microbiology* **111:** 73-85.

Herrmann KM, Weaver LM. **1999.** The shikimate pathway. *Annu.Rev.Plant Physiol Plant Mol.Biol.* **50:** 473-503.

Hill DR, Hladun SL, Scherer S, Potts M. **1994.** Water stress proteins of *Nostoc commune* (Cyanobacteria) are secreted with UV-A/B-absorbing pigments and associate with 1,4-beta-*D*-xylanxylanohydrolase activity. *J.Biol.Chem.* **269:** 7726-7734.

Hoffmann D, Hevel JM, Moore RE, Moore BS. **2003.** Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224. *Gene* **311:** 171-180.

Hoiczyk E, Baumeister W. **1995.** Envelope structure of four gliding filamentous cyanobacteria. *J.Bacteriol.* **177:** 2387-2395.

Hoiczyk E, Hansel A. **2000.** Cyanobacterial cell walls: news from an unusual prokaryotic envelope. *J.Bacteriol.* **182:** 1191-1199.

Hojati Z, Milne C, Harvey B, Gordon L, Borg M, Flett F, Wilkinson B, Sidebottom PJ, Rudd BA, Hayes MA, Smith CP, Micklefield J. **2002.** Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem.Biol.* **9:** 1175-1187.

Holden MJ, Mayhew MP, Gallagher DT, Vilker VL. **2002.** Chorismate lyase: kinetics and engineering for stability. *Biochim.Biophys.Acta* **1594:** 160-167.

Horswill AR, Escalante-Semerena JC. **2002.** Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: Residue Lys592 is required for propionyl-AMP synthesis. *Biochemistry* **41:** 2379-2387.

Howard-Jones AR, Walsh CT. **2005.** Enzymatic generation of the chromopyrrolic acid scaffold of rebeccamycin by the tandem action of RebO and RebD. *Biochemistry* **44:** 15652-15663.

Howard-Jones AR, Walsh CT. **2006.** Staurosporine and rebeccamycin aglycones are assembled by the oxidative action of StaP, StaC, and RebC on chromopyrrolic acid. *J.Am.Chem.Soc.* **128:** 12289-12298.

Howard-Jones AR, Walsh CT. **2007.** Nonenzymatic oxidative steps accompanying action of the cytochrome P450 enzymes StaP and RebP in the biosynthesis of staurosporine and rebeccamycin. *J.Am.Chem.Soc.* **129:** 11016-11017.

Hu JL, Xue YC, Xie MY, Zhang R, Otani T, Minami Y, Yamada Y, Marunaka T. **1988.** A new macromolecular antitumor antibiotic, C-1027. I. Discovery, taxonomy of producing organism, fermentation and biological activity. *J.Antibiot.(Tokyo)* **41:** 1575-1579.

Huang X, Holden HM, Raushel FM. **2001.** Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Annu.Rev.Biochem.* **70:** 149-180.

Hubbard BK, Walsh CT. **2003.** Vancomycin assembly: nature's way. *Angew.Chem.Int.Ed Engl.* **42:** 730-765.

Hube AE, Heyduck-Soller B, Fischer U. **2009.** Phylogenetic classification of heterotrophic bacteria associated with filamentous marine cyanobacteria in culture. *Syst.Appl.Microbiol.* **32:** 256-265.

Hung T, Mak K, Fong K. **1990.** A specificity enhancer for polymerase chain reaction. *Nucleic Acids Res.* **18:** 4953.

Hutchinson CR. **2003.** Polyketide and non-ribosomal peptide synthases: falling together by coming apart. *Proc.Natl.Acad.Sci.U.S.A* **100:** 3010-3012.

Hyun CG, Bililign T, Liao J, Thorson JS. **2003.** The biosynthesis of indolocarbazoles in a heterologous *E. coli* host. *Chembiochem* **4:** 114-117.

Ingram-Smith C, Martin SR, Smith KS. **2006a.** Acetate kinase: not just a bacterial enzyme. *Trends Microbiol.* **14:** 249-253.

Ingram-Smith C, Woods BI, Smith KS. **2006b.** Characterization of the acyl substrate binding pocket of acetyl-CoA synthetase. *Biochemistry* **45:** 11482-11490.

Ishida K, Christiansen G, Yoshida WY, Kurmayer R, Welker M, Valls N, Bonjoch J, Hertweck C, Borner T, Hemscheidt T, Dittmann E. **2007.** Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety. *Chem.Biol.* **14:** 565-576.

Ishida K, Okita Y, Matsuda H, Okino T, Murakami M. **1999.** Aeruginosins, protease inhibitors from the cyanobacterium *Microcystis aeruginosa*. *Tetrahedron* **55:** 10971-10988.

Ishida K, Welker M, Christiansen G, Cadel-Six S, Bouchier C, Dittmann E, Hertweck C, Tandeau dM. **2009.** Plasticity and evolution of aeruginosin biosynthesis in cyanobacteria. *Appl.Environ.Microbiol.* **75:** 2017-2026.

Isin EM, Guengerich FP. **2008.** Substrate binding to cytochromes P450. *Anal.Bioanal.Chem.* **392:** 1019-1030.

Izumikawa M, Cheng Q, Moore BS. 2006. Priming type II polyketide synthases via a type II nonribosomal peptide synthetase mechanism. *J Am Chem Soc.* **128:** 1428–1429.

Jensen RA, Xie G, Calhoun DH, Bonner CA. **2002.** The correct phylogenetic relationship of KdsA (3-deoxy-D-manno-octulosonate 8-phosphate synthase) with one of two independently evolved classes of AroA (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase). *J.Mol.Evol* **54:** 416-423.

Jiménez JI, Scheuer PJ. **2001.** New lipopeptides from the Caribbean cyanobacterium *Lyngbya majuscula*. *J.Nat Prod* **64:** 200-203.

Johnson CH, Egli M, Stewart PL. **2008a.** Structural insights into a circadian oscillator. *Science* **322:** 697-701.

Johnson CH, Mori T, Xu Y. **2008b.** A cyanobacterial circadian clockwork. *Curr.Biol.* **18:** R816-R825.

Jones AC, Gu L, Sorrels CM, Sherman DH, Gerwick WH. **2009.** New tricks from ancient algae: natural products biosynthesis in marine cyanobacteria. *Curr.Opin.Chem.Biol.* **13:** 216-223.

Jones AC, Monroe EA, Eisman EB, Gerwick L, Sherman DH, Gerwick WH. **2010.** The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat Prod Rep* **27:** 1048-1065.

Jonsson ME, Gao K, Olsson JA, Goldstone JV, Brandt I. **2010a.** Induction patterns of new CYP1 genes in environmentally exposed rainbow trout. *Aquat.Toxicol.* **98:** 311-321.

Jonsson ME, Gao K, Olsson JA, Goldstone JV, Brandt I. **2010b.** Induction patterns of new CYP1 genes in environmentally exposed rainbow trout. *Aquat.Toxicol.* **98:** 311-321.

Jossek R, Bongaerts J, Sprenger GA. **2001.** Characterization of a new feedback-resistant 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase AroF of *Escherichia coli*. *FEMS Microbiol.Lett.* **202:** 145-148.

Kazlauskas R, Lidgard RO, Wells RJ. **1977.** A novel hexachloro-metabolite from the sponge *Dysidea herbacea*. *Tetrahedron Lett.* 3183-3186.

Keating TA, Marshall CG, Walsh CT. **2000a.** Reconstitution and characterization of the *Vibrio cholerae* vibriobactin synthetase from VibB, VibE, VibF, and VibH. *Biochemistry* **39:** 15522-15530.

Keating TA, Marshall CG, Walsh CT. **2000b.** Vibriobactin biosynthesis in *Vibrio cholerae*: VibH is an amide synthase homologous to nonribosomal peptide synthetase condensation domains. *Biochemistry* **39:** 15513-15521.

Keating TA, Marshall CG, Walsh CT, Keating AE. **2002.** The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol.* **9:** 522-526.

Keller S, Wage T, Hohaus K, Holzer M, Eichhorn E, van Pee KH. **2000.** Purification and partial characterization of tryptophan 7-halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew.Chem.Int.Ed Engl.* **39:** 2300-2302.

Kessler N, Schuhmann H, Morneweg S, Linne U, Marahiel MA. **2004.** The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J.Biol.Chem.* **279:** 7413-7419.

Khare D, Wang B, Gu L, Razelun J, Sherman DH, Gerwick WH, Hakansson K, Smith JL. **2010.** Conformational switch triggered by alpha-ketoglutarate in a halogenase of curacin A biosynthesis. *Proc.Natl.Acad.Sci.U.S.A* **107:** 14099-14104.

Khosla C, Tang Y, Chen AY, Schnarr NA, Cane DE. **2007.** Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu.Rev.Biochem.* **76:** 195-221.

Kim SY, Park JS, Chae CS, Hyun CG, Choi BW, Shin J, Oh KB. **2007.** Genetic organization of the biosynthetic gene cluster for the indolocarbazole K-252a in *Nonomuraea longicatena* JCM 11136. *Appl.Microbiol.Biotechnol.* **75:** 1119-1126.

Kirner S, Hammer PE, Hill DS, Altmann A, Fischer I, Weislo LJ, Lanahan M, van Pee KH, Ligon JM. **1998.** Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. *J.Bacteriol.* **180:** 1939-1943.

Kling E, Schmid C, Unversucht S, Wage T, Zehner S, van Pee KH. **2005.** Enzymatic incorporation of halogen atoms into natural compounds. *Ernst.Schering.Res.Found.Workshop* 165-194.

Knaggs AR. **2003.** The biosynthesis of shikimate metabolites. *Nat Prod Rep* **20:** 119-136.

Knobloch KH, Hahlbrock K. **1977.** 4-Coumarate:CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm. *Arch. Biochem. Biophys* **184:** 237-248.

Kobayashi M, Kitagawa I. **1999.** Marine spongean cytotoxins. *J.Nat Toxins.* **8:** 249-258.

Kobayashi M, Wang W, Ohyabu N, Kurosu M, Kitagawa I. **1995.** Improved total synthesis and structure-activity relationship of arenastatin A, a potent cytotoxic spongean depsipeptide. *Chem.Pharm.Bull.(Tokyo)* **43:** 1598-1600.

Koglin A, Walsh CT. **2009.** Structural insights into nonribosomal peptide enzymatic assembly lines. *Nat Prod Rep* **26:** 987-1000.

Konz D, Doekel S, Marahiel MA. **1999.** Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. *J.Bacteriol.* **181:** 133-140.

Konz D, Marahiel MA. **1999.** How do peptide synthetases generate structural diversity? *Chem.Biol.* **6:** R39-R48.

Kozbial PZ, Mushegian AR. **2005.** Natural history of S-adenosylmethionine-binding proteins. *BMC.Struct.Biol.* **5:** 19.

Kraas FI, Helmetag V, Wittmann M, Strieker M, Marahiel MA. **2010.** Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation. *Chem.Biol.* **17:** 872-880.

Kurihara H, Mitani T, Kawabata J, Takahashi K. **1999a.** Inhibitory potencies of bromophenols from Rhodomelaceae algae against alpha-glucosidase activity. *Fish Sci* **65:** 300-303.

Kurihara H, Mitani T, Kawabata J, Takahashi K. **1999b.** Two new bromophenols from the red alga *Odonthalia corymbifera*. *J.Nat.Prod.* **62:** 882-884.

La Barre S, Potin P, Leblanc C, Delage L. **2010.** The halogenated metabolism of brown algae (Phaeophyta), its biological importance and its environmental significance. *Mar.Drugs* **8:** 988-1010.

Lang H, Polnick S, Nicke T, William P, Patallo EP, Naismith JH, van Pée KH. **2011.** Änderung der Regioselektivität der Tryptophan-7-halogenase PrnA durch ortsspezifische Mutagenese. *Angew. Chem.* **123**, 3007 –3010.

Laatsch H, Pudleiner H. **1989.** Marine bacteria, I. - Synthesis of pentabromopseudiline, a cytotoxic phenylpyrrole from *Alteromonas luteoviolaceus*. *Liebigs Ann.Chem.* 863-881.

Laub LB, Jones BD, Powell WH. **2010.** Responsiveness of a *Xenopus laevis* cell line to the aryl hydrocarbon receptor ligands 6-formylindolo[3,2-b]carbazole (FICZ) and 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). *Chem.Biol.Interact.* **183:** 202-211.

Lautru S, Challis GL. **2004.** Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology* **150:** 1629-1636.

Leao PN, Pereira AR, Liu WT, Ng J, Pevzner PA, Dorrestein PC, Konig GM, Vasconcelosa VM, Gerwick WH. **2010.** Synergistic allelochemicals from a freshwater cyanobacterium. *Proc.Natl.Acad.Sci.U.S.A.* **107:** 11183-11188.

Leikoski N, Fewer DP, Jokela J, Wahlsten M, Rouhiainen L, Sivonen K. **2010.** Highly diverse cyanobactins in strains of the genus *Anabaena*. *Appl.Environ.Microbiol.* **76:** 701-709.

Leikoski N, Fewer DP, Sivonen K. **2009.** Widespread occurrence and lateral transfer of the cyanobactin biosynthesis gene cluster in cyanobacteria. *Appl.Environ.Microbiol.* **75:** 853-857.

Lemloh ML, Fromont J, Brummer F, Usher KM. **2009.** Diversity and abundance of photosynthetic sponges in temperate Western Australia. *BMC.Ecol.* **9:** 4.

Lewis DFV, Wiseman A. **2005.** A selective review of bacterial forms of cytochrome P450 enzymes - Review. *Enzyme and Microbial Technology* **36:** 377-384.

Li SY, Wackett LP. **1993.** Reductive dehalogenation by cytochrome P450(Cam) - substrate binding and catalysis. *Biochemistry* **32:** 9355-9361.

Li TL, Huang F, Haydock SF, Mironenko T, Leadlay PF, Spencer JB. **2004.** Biosynthetic gene cluster of the glycopeptide antibiotic teicoplanin: characterization of two glycosyltransferases and the key acyltransferase. *Chem.Biol.* **11:** 107-119.

Liang C, Zhao F, Wei W, Wen Z, Qin S. **2006.** Carotenoid biosynthesis in cyanobacteria: structural and evolutionary scenarios based on comparative genomics. *Int.J.Biol.Sci.* **2:** 197-207.

Lin S, Van Lanen SG, Shen B. **2007.** Regiospecific chlorination of (S)-beta-tyrosyl-S-carrier protein catalyzed by SgcC3 in the biosynthesis of the enediyne antitumor antibiotic C-1027. *J.Am.Chem.Soc.* **129:** 12432-12438.

Linne U, Schafer A, Stubbs MT, Marahiel MA. **2007.** Aminoacyl-coenzyme A synthesis catalyzed by adenylation domains. *FEBS Lett.* **581:** 905-910.

Liu L, Rein KS. **2010.** New peptides isolated from *Lyngbya* species: a review. *Mar.Drugs* **8:** 1817-1837.

Liu WL, Zhang JC, Jiang FQ, Fu L. **2009.** Synthesis and cytotoxicity studies of new cryptophycin analogues. *Arch.Pharm.(Weinheim)* **342:** 577-583.

Lu S, Archer MC. **2005.** Fatty acid synthase is a potential molecular target for the chemoprevention of breast cancer. *Carcinogenesis* **26:** 153-157.

Luecking R, Lawrey JD, Sikaroodi M, Gillevet PM, Chaves JL, Sipman HJM, Bungartz F. **2009.** Do lichens domesticate photobionts like farmers domesticate crops? Evidence from a previously unrecognized lineage of filamentous cyanobacteria. *Am.J.Bot.* **96:** 1409-1418.

Magarvey NA, Beck ZQ, Golakoti T, Ding Y, Huber U, Hemscheidt TK, Abelson D, Moore RE, Sherman DH. **2006.** Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from cyanobionts. *ACS Chem.Biol.* **1:** 766-779.

Makino M, Sugimoto H, Shiro Y, Asamizu S, Onaka H, Nagano S. **2007.** Crystal structures and catalytic mechanism of cytochrome P450 StaP that produces the indolocarbazole skeleton. *Proc.Natl.Acad.Sci.U.S.A* **104:** 11591-11596.

Marahiel MA, Stachelhaus T, Mootz HD. **1997.** Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem.Rev.* **97:** 2651-2674.

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH. **2009.** CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* **37:** D205-D210.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. **2011.** CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39:** D225–D229.

Marminon C, Anizon F, Moreau P, Pfeiffer B, Pierre A, Golsteyn RM, Peixoto P, Hildebrand MP, David-Cordonnier MH, Lozach O, Meijer L, Prudhomme M. **2008.** Rebeccamycin derivatives as dual DNA-damaging agents and potent checkpoint kinase 1 inhibitors. *Mol.Pharmacol.* **74:** 1620-1629.

Martin JL, McMillan FM. **2002.** SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr.Opin.Struct.Biol.* **12:** 783-793.

McPhail KL, Correa J, Linington RG, Gonzalez J, Ortega-Barria E, Capson TL, Gerwick WH. **2007.** Antimalarial linear lipopeptides from a Panamanian strain of the marine cyanobacterium *Lyngbya majuscula*. *J.Nat Prod* **70:** 984-988.

Meeks JC, Elhai J. **2002.** Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. *Microbiol.Mol.Biol.Rev.* **66:** 94-+.

Mehner C, Muller D, Kehraus S, Hautmann S, Gutschow M, Konig GM. **2008.** New peptolides from the cyanobacterium *Nostoc insulare* as selective and potent inhibitors of human leukocyte elastase. *Chembiochem* **9:** 2692-2703.

Metzker ML. **2010.** Sequencing technologies - the next generation. *Nat Rev.Genet.* **11:** 31-46.

Miethke M, Bisseret P, Beckering CL, Vignard D, Eustache J, Marahiel MA. **2006.** Inhibition of aryl acid adenylation domains involved in bacterial siderophore synthesis. *FEBS J.* **273:** 409-419.

Moffitt MC, Neilan BA. **2004.** Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. *Appl.Environ.Microbiol.* **70:** 6353-6362.

Moon YJ, Park YM, Chung YH, Choi JS. **2004.** Calcium is involved in photomovement of cyanobacterium *Synechocystis* sp. PCC 6803. *Photochem.Photobiol.* **79:** 114-119.

Moore RE. **1996.** Cyclic peptides and depsipeptides from cyanobacteria: a review. *J.Ind.Microbiol.* **16:** 134-143.

Moreau P, Holbeck S, Prudhomme M, Sausville EA. **2005.** Cytotoxicities of three rebeccamycin derivatives in the National Cancer Institute screening of 60 human tumor cell lines. *Anticancer Drugs* **16:** 145-150.

Morris HR, Taylor GW, Masento MS, Jermyn KA, Kay RR. **1987.** Chemical structure of the morphogen differentiation inducing factor from *Dictyostelium discoideum*. *Nature* **328:** 811-814.

Mortison JD, Sherman DH. **2010.** Frontiers and opportunities in chemoenzymatic synthesis. *J.Org.Chem.* **75:** 7041-7051.

Murakami M, Ishida K, Okino T, Okita Y, Matsuda H, Yamaguchi K. **1995.** Aeruginosin-98-A and Aeruginosin-98-B, trypsin-Inhibitors from the blue-green-alga *Microcystis aeruginosa* Nies-98. *Tetrahedron Lett.* **36:** 2785-2788.

Murakami N, Tamura S, Koyama K, Sugimoto M, Maekawa R, Kobayashi M. **2004.** New analogue of arenastatin A, a potent cytotoxic spongean depsipeptide, with anti-tumor activity. *Bioorg.Med.Chem.Lett.* **14:** 2597-2601.

Nagle DG, Paul VJ. **1998.** Chemical defense of a marine cyanobacterial bloom. *J.Exp.Mar.Biol.Ecol.* **225:** 29-38.

Nakano H, Omura S. **2009.** Chemical biology of natural indolocarbazole products: 30 years since the discovery of staurosporine. *J.Antibiot.(Tokyo)* **62:** 17-26.

Nakano MM, Magnuson R, Myers A, Curry J, Grossman AD, Zuber P. **1991.** srfA is an operon required for surfactin production, competence development, and efficient sporulation in *Bacillus subtilis*. *J.Bacteriol.* **173:** 1770-1778.

Neer EJ, Schmidt CJ, Nambudripad R, Smith TF. **1994.** The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371:** 297-300.

Neres J, Wilson DJ, Celia L, Beck BJ, Aldrich CC. **2008.** Aryl acid adenylating enzymes involved in siderophore biosynthesis: fluorescence polarization assay, ligand specificity, and discovery of non-nucleoside inhibitors via high-throughput screening. *Biochemistry* **47:** 11735-11749.

Nett M, Koenig GM. **2007.** The chemistry of gliding bacteria. *Nat Prod Rep* **24:** 1245-1261.

Neumann CS, Fujimori DG, Walsh CT. **2008.** Halogenation strategies in natural product biosynthesis. *Chem.Biol.* **15:** 99-109.

Neumann CS, Walsh CT, Kay RR. **2010.** A flavin-dependent halogenase catalyzes the chlorination step in the biosynthesis of *Dictyostelium* differentiation-inducing factor 1. *Proc.Natl.Acad.Sci.U.S.A* **107:** 5798-5803.

Nicolaus B, Panico A, Lama L, Romano I, Manca MC, De Giulio A, Gambacorta A. **1999.** Chemical composition and production of exopolysaccharides from representative members of heterocystous and non-heterocystous cyanobacteria. *Phytochemistry* **52:** 639-647.

Nishiwaki-Matsushima R, Nishiwaki S, Ohta T, Yoshizawa S, Suganuma M, Harada K, Watanabe MF, Fujiki H. **1991.** Structure-function relationships of microcystins, liver tumor promoters, in interaction with protein phosphatase. *Jpn.J.Cancer Res.* **82:** 993-996.

Nishizawa T, Aldrich CC, Sherman DH. **2005.** Molecular analysis of the rebeccamycin L-amino acid oxidase from *Lechevalieria aerocolonigenes* ATCC 39243. *J.Bacteriol.* **187:** 2084-2092.

Nolan EM, Walsh CT. **2009.** How nature morphs peptide scaffolds into antibiotics. *Chembiochem.* **10:** 34-53.

Nowak-Thompson B, Chaney N, Wing JS, Gould SJ, Loper JE. **1999.** Characterization of the pyoluteorin biosynthetic gene cluster of *Pseudomonas fluorescens* Pf-5. *J.Bacteriol.* **181:** 2166-2174.

Oh KB, Lee JH, Chung SC, Shin J, Shin HJ, Kim HK, Lee HS. **2008.** Antimicrobial activities of the bromophenols from the red alga *Odonthalia corymbifera* and some synthetic derivatives. *Bioorg.Med.Chem.Lett.* **18:** 104-108.

Oh KB, Lee JH, Lee JW, Yoon KM, Chung SC, Jeon HB, Shin J, Lee HS. **2009.** Synthesis and antimicrobial activities of halogenated bis(hydroxyphenyl)methanes. *Bioorg.Med.Chem.Lett.* **19:** 945-948.

Ohta T, Takahashi A, Matsuda M, Kamo S, Agatsuma T, Endo T, Nozoe S. **1995.** Russuphelol, a novel optically active chlorohydroquinone tetramer from the mushroom *Russula subnigricans*. *Tetrahedron Lett.* **36:** 5223-5226.

Ojha S, Meng EC, Babbitt PC. **2007.** Evolution of function in the "two dinucleotide binding domains" flavoproteins. *PLoS.Comput.Biol.* **3:** e121.

Onaka H. **2009.** Biosynthesis of indolocarbazole and goadsporin, two different heterocyclic antibiotics produced by actinomycetes. *Biosci.Biotechnol.Biochem.* **73:** 2149-2155.

Onaka H, Taniguchi S, Igarashi Y, Furumai T. **2002.** Cloning of the staurosporine biosynthetic gene cluster from *Streptomyces* sp. TP-A0274 and its heterologous expression in *Streptomyces lividans*. *J.Antibiot.(Tokyo)* **55:** 1063-1071.

Onaka H, Taniguchi S, Igarashi Y, Furumai T. **2003.** Characterization of the biosynthetic gene cluster of rebeccamycin from *Lechevalieria aerocolonigenes* ATCC 39243. *Biosci.Biotechnol.Biochem.* **67:** 127-138.

Orjala J, Gerwick WH. **1996.** Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium *Lyngbya majuscula*. *J.Nat Prod* **59:** 427-430.

Ouyang Y, Andersson CR, Kondo T, Golden SS, Johnson CH. **1998.** Resonating circadian clocks enhance fitness in cyanobacteria. *Proc.Natl.Acad.Sci.U.S.A* **95:** 8660-8664.

Paerl HW. **1984.** Cyanobacterial carotenoids - Their roles in maintaining optimal photosynthetic production among aquatic bloom forming genera. *Oecologia* **61:** 143-149.

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

Penn K, Jenkins C, Nett M, Udwary DW, Gontang EA, McGlinchey RP, Foster B, Lapidus A, Podell S, Allen EE, Moore BS, Jensen PR. **2009.** Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME.J.* **3:** 1193-1203.

Pereira ER, Belin L, Sancelme M, Prudhomme M, Ollier M, Rapp M, Severe D, Riou JF, Fabbro D, Meyer T. **1996.** Structure-activity relationships in a series of substituted indolocarbazoles: topoisomerase I and protein kinase C inhibition and antitumoral and antimicrobial properties. *J.Med.Chem.* **39:** 4471-4477.

Perozzo R, Kuo M, Sidhu AS, Valiyaveettil JT, Bittman R, Jacobs WR, Jr., Fidock DA, Sacchettini JC. **2002.** Structural elucidation of the specificity of the antibacterial agent triclosan for malarial enoyl acyl carrier protein reductase. *J.Biol.Chem.* **277:** 13106-13114.

Peschke JD, Hanefeld U, Laatsch H. **2005.** Biosynthesis of the marine antibiotic pentabromopseudilin. 2. The pyrrole ring. *Biosci.Biotechnol.Biochem.* **69:** 628-630.

Podzelinska K, Latimer R, Bhattacharya A, Vining LC, Zechel DL, Jia Z. **2010.** Chloramphenicol biosynthesis: the structure of CmlS, a flavin-dependent halogenase showing a covalent flavin-aspartate bond. *J.Mol.Biol.* **397:** 316-331.

Pojer F, Kahlich R, Kammerer B, Li SM, Heide L. **2003a.** CloR, a bifunctional non-heme iron oxygenase involved in clorobiocin biosynthesis. *J.Biol.Chem.* **278:** 30661-30668.

Pojer F, Wemakor E, Kammerer B, Chen H, Walsh CT, Li SM, Heide L. **2003b.** CloQ, a prenyltransferase involved in clorobiocin biosynthesis. *Proc.Natl.Acad.Sci U.S.A* **100:** 2316-2321.

Porter RD. **1988.** DNA transformation. *Methods Enzymol.* **167:** 703-712.

Prudhomme M. **2003.** Rebeccamycin analogues as anti-cancer agents. *Eur.J.Med.Chem.* **38:** 123-140.

Prudhomme M. **2004.** Combining DNA damaging agents and checkpoint 1 inhibitors. *Curr.Med.Chem.Anticancer Agents* **4:** 435-438.

Puk O, Huber P, Bischoff D, Recktenwald J, Jung G, Sussmuth RD, van Pee KH, Wohlleben W, Pelzer S. **2002.** Glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908: function of a halogenase and a haloperoxidase/perhydrolase. *Chem.Biol.* **9:** 225-235.

Pylypenko O, Vitali F, Zerbe K, Robinson JA, Schlichting I. **2003.** Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis. *J.Biol.Chem.* **278:** 46727-46733.

Qiao C, Wilson DJ, Bennett EM, Aldrich CC. **2007.** A mechanism-based aryl carrier protein/thiolation domain affinity probe. *J.Am.Chem.Soc.* **129:** 6350-6351.

Rachid S, Krug D, Kunze B, Kochems I, Scharfe M, Zabriskie TM, Blocker H, Muller R. **2006.** Molecular and biochemical studies of chondramide formation-highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. *Chem.Biol.* **13:** 667-681.

Rachid S, Revermann O, Dauth C, Kazmaier U, Muller R. **2010.** Characterization of a novel type of oxidative decarboxylase involved in the biosynthesis of the styryl moiety of chondrochloren from an acylated tyrosine. *J.Biol.Chem.* **285:** 12482-12489.

Rachid S, Scharfe M, Blocker H, Weissman KJ, Muller R. **2009.** Unusual chemistry in the biosynthesis of the antibiotic chondrochlorens. *Chem.Biol.* **16:** 70-81.

Ramaswamy AV, Sorrels CM, Gerwick WH. **2007.** Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *J.Nat.Prod.* **70:** 1977-1986.

Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH. **2007.** Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol.Biol.* **7:** 78.

Reimmann C, Patel HM, Walsh CT, Haas D. **2004.** PchC thioesterase optimizes nonribosomal biosynthesis of the peptide siderophore pyochelin in *Pseudomonas aeruginosa*. *J.Bacteriol.* **186:** 6367-6373.

Reizer J, Buskirk S, Bairoch A, Reizer A, Saier MH. **1994.** A novel zinc-binding motif found in 2 ubiquitous deaminase families. *Protein Sci* **3:** 853-856.

Rippka R. **1988.** Isolation and purification of cyanobacteria. *Methods Enzymol.* **167:** 3-27.

Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. **1979.** Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Journal of General Microbiology* **111:** 1-61.

Roche ED, Walsh CT. **2003.** Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry* **42:** 1334-1344.

Roma GW, Crowley LJ, Davis CA, Barber Mj. **2005.** Mutagenesis of glycine 179 modulates both catalytic efficiency and reduced pyridine nucleotide specificity in cytochrome b5 reductase. *Biochemistry* **44:** 13467–13476.

Roongsawang N, Lim SP, Washio K, Takano K, Kanaya S, Morikawa M. **2005.** Phylogenetic analysis of condensation domains in the nonribosomal peptide synthetases. *FEMS Microbiol.Lett.* **252:** 143-151.

Rouhiainen L, Paulin L, Suomalainen S, Hyytiainen H, Buikema W, Haselkorn R, Sivonen K. **2000.** Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena* strain 90. *Mol.Microbiol.* **37:** 156-167.

Rouhiainen L, Vakkilainen T, Siemer BL, Buikema W, Haselkorn R, Sivonen K. **2004.** Genes coding for hepatotoxic heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90. *Appl.Environ.Microbiol.* **70:** 686-692.

Rounge TB, Rohrlack T, Nederbragt AJ, Kristensen T, Jakobsen KS. **2009.** A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. *BMC.Genomics* **10:** 396.

Rounge TB, Rohrlack T, Tooming-Klunderud A, Kristensen T, Jakobsen KS. **2007.** Comparison of cyanopeptolin genes in *Planktothrix*, *Microcystis*, and *Anabaena* strains: evidence for independent evolution within each genus. *Appl.Environ.Microbiol.* **73:** 7322-7330.

Rupasinghe S, Schuler MA, Kagawa N, Yuan H, Lei L, Zhao B, Kelly SL, Waterman MR, Lamb DC. **2006.** The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine. *FEBS Lett.* **580:** 6338-6342.

Ryan KS, Drennan CL. **2009.** Divergent pathways in the biosynthesis of bisindole natural products. *Chem.Biol.* **16:** 351-364.

Ryan KS, Howard-Jones AR, Hamill MJ, Elliott SJ, Walsh CT, Drennan CL. **2007.** Crystallographic trapping in the rebeccamycin biosynthetic enzyme RebC. *Proc.Natl.Acad.Sci.U.S.A* **104:** 15311-15316.

Salas JA, Mendez C. **2009.** Indolocarbazole antitumour compounds by combinatorial biosynthesis. *Curr.Opin.Chem.Biol.* **13:** 152-160.

Sambrook J, Russell D. **2001.** *Molecular Cloning: A Laboratory Manual (Third edition)*.

Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen LO. **2007.** Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15:** 781-792.

Sánchez C, Butovich IA, Brana AF, Rohr J, Mendez C, Salas JA. **2002.** The biosynthetic gene cluster for the antitumor rebeccamycin: characterization and generation of indolocarbazole derivatives. *Chem.Biol.* **9:** 519-531.

Sánchez C, Mendez C, Salas JA. **2006a.** Engineering biosynthetic pathways to generate antitumor indolocarbazole derivatives. *J.Ind.Microbiol.Biotechnol.* **33:** 560-568.

Sánchez C, Mendez C, Salas JA. **2006b.** Indolocarbazole natural products: occurrence, biosynthesis, and biological activity. *Nat.Prod.Rep.* **23:** 1007-1045.

Sánchez C, Zhu L, Brana AF, Salas AP, Rohr J, Mendez C, Salas JA. **2005.** Combinatorial biosynthesis of antitumor indolocarbazole compounds. *Proc.Natl.Acad.Sci.U.S.A* **102:** 461-466.

Sánchez LM, Lopez D, Vesely BA, Della TG, Gerwick WH, Kyle DE, Linington RG. **2010.** Almiramides A-C: discovery and development of a new class of leishmaniasis lead compounds. *J.Med.Chem.* **53:** 4187-4197.

Sanger F, Nicklen S, Coulson AR. **1977.** DNA sequencing with chain-terminating inhibitors. *Proc.Natl.Acad.Sci.U.S.A* **74:** 5463-5467.

Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J. **2005.** Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc.Natl.Acad.Sci.U.S.A* **102:** 7315-7320.

Schneider A, Marahiel MA. **1998.** Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch.Microbiol.* **169:** 404-410.

Sharma A, Johri BN. **2003.** Growth promoting influence of siderophore-producing *Pseudomonas* strains GRP3A and PRS9 in maize (Zea mays L.) under iron limiting conditions. *Microbiol.Res.* **158:** 243-248.

Shaw WV, Packman LC, Burleigh BD, Dell A, Morris HR, Hartley BS. **1979.** Primary structure of a chloramphenicol acetyltransferase specified by R plasmids. *Nature* **282:** 870-872.

Shridhar DM, Mahajan GB, Kamat VP, Naik CG, Parab RR, Thakur NR, Mishra PD. **2009.** Antibacterial activity of 2-(2', 4'-dibromophenoxy)-4, 6-dibromophenol from *Dysidea granulosa*. *Mar.Drugs* **7:** 464-471.

Sieber SA, Marahiel MA. **2005.** Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem.Rev.* **105:** 715-738.

Sielaff H, Christiansen G, Schwecke T. **2006.** Natural products from cyanobacteria: Exploiting a new source for drug discovery. *IDrugs.* **9:** 119-127.

Sikora AL, Wilson DJ, Aldrich CC, Blanchard JS. **2010.** Kinetic and inhibition studies of dihydroxybenzoate-AMP ligase from *Escherichia coli*. *Biochemistry* **49:** 3648-3657.

Silva-Stenico ME, Silva CS, Lorenzi AS, Shishido TK, Etchegaray A, Lira SP, Moraes LA, Fiore MF. **2010.** Non-ribosomal peptides produced by Brazilian cyanobacterial isolates with antimicrobial activity. *Microbiol.Res.*

Simmons TL, Coates RC, Clark BR, Engene N, Gonzalez D, Esquenazi E, Dorrestein PC, Gerwick WH. **2008.** Biosynthetic origin of natural products isolated from marine microorganism-invertebrate assemblages. *Proc.Natl.Acad.Sci.U.S.A* **105:** 4587-4594.

Singh IP, Sidana J, Bansal P, Foley WJ. **2009.** Phloroglucinol compounds of therapeutic interest: global patent and technology status. *Expert.Opin.Ther.Pat* **19:** 847-866.

Singh IP, Sidana J, Bharate SB, Foley WJ. **2010.** Phloroglucinol compounds of natural origin: synthetic aspects. *Nat Prod Rep* **27:** 393-416.

Singh S, Kate BN, Banerjee UC. **2005.** Bioactive compounds from cyanobacteria and microalgae: an overview. *Crit Rev.Biotechnol.* **25:** 73-95.

Singh S, McCoy JG, Zhang C, Bingman CA, Phillips GN, Jr., Thorson JS. **2008.** Structure and mechanism of the rebeccamycin sugar 4'-O-methyltransferase RebM. *J.Biol.Chem.* **283:** 22628-22636.

Sisay MT, Hautmann S, Mehner C, Konig GM, Bajorath J, Gutschow M. **2009.** Inhibition of human leukocyte elastase by brunsvicamides A-C: cyanobacterial cyclic peptides. *ChemMedChem.* **4:** 1425-1429.

Sivonen K, Leikoski N, Fewer DP, Jokela J. **2010.** Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl.Microbiol.Biotechnol.* **86:** 1213-1225.

Smith CD, Zhang X. **1996.** Mechanism of action cryptophycin. Interaction with the *Vinca* alkaloid domain of tubulin. *J.Biol.Chem.* **271:** 6192-6198.

Smith CD, Zhang X, Mooberry SL, Patterson GM, Moore RE. **1994.** Cryptophycin: a new antimicrotubule agent active against drug-resistant cells. *Cancer Res.* **54:** 3779-3784.

Smith JL, Boyer GL, Zimba PV. **2008.** A review of cyanobacterial odorous and bioactive metabolites: Impacts and management alternatives in aquaculture. *Aquaculture* **280:** 5-20.

Smith N, Roitberg AE, Rivera E, Howard A, Holden MJ, Mayhew M, Kaistha S, Gallagher DT. **2006.** Structural analysis of ligand binding and catalysis in chorismate lyase. *Arch.Biochem.Biophys.* **445:** 72-80.

Snider J, Houry WA. **2008.** AAA+ proteins: diversity in function, similarity in structure. *Biochem.Soc.Trans.* **36:** 72-77.

Snider J, Thibault G, Houry WA. **2008.** The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **9:** 216.

Stachelhaus T, Huser A, Marahiel MA. **1996a.** Biochemical characterization of peptides carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chemistry & Biology* **3:** 913-921.

Stachelhaus T, Schneider A, Marahiel MA. **1996b.** Engineered biosynthesis of peptide antibiotics. *Biochemical Pharmacology* **52:** 177-186.

Steffensky M, Li SM, Heide L. **2000a.** Cloning, overexpression, and purification of novobiocic acid synthetase from *Streptomyces spheroides* NCIMB 11891. *J.Biol.Chem.* **275:** 21754-21760.

Steffensky M, Muhlenweg A, Wang ZX, Li SM, Heide L. **2000b.** Identification of the novobiocin biosynthetic gene cluster of *Streptomyces spheroides* NCIB 11891. *Antimicrob.Agents Chemother.* **44:** 1214-1222.

Stegmann E, Frasch HJ, Wohlleben W. **2010.** Glycopeptide biosynthesis in the context of basic cellular functions. *Curr.Opin.Microbiol.* **13:** 595-602.

Stegmann E, Pelzer S, Bischoff D, Puk O, Stockert S, Butz D, Zerbe K, Robinson J, Sussmuth RD, Wohlleben W. **2006.** Genetic analysis of the balhimycin (vancomycin-type) oxygenase genes. *J.Biotechnol.* **124:** 640-653.

Steller S, Sokoll A, Wilde C, Bernhard F, Franke P, Vater J. **2004.** Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochemistry* **43:** 11331-11343.

Stöveken T, Kalscheuer R, Malkus U, Reichelt R, Steinbuchel A. **2005.** The wax ester synthase/acyl coenzyme A: diacylglycerol acyltransferase from *Acinetobacter* sp strain ADP1: Characterization of a novel type of acyltransferase. *J.Bacteriol.* **187:** 1369-1376.

Stöveken T, Kalscheuer R, Steinbuchel A. **2009.** Both histidine residues of the conserved HHXXXDG motif are essential for wax ester synthase/acyl-CoA:diacylglycerol acyltransferase catalysis. *Eur.J.Lipid Sci.Technol.* **111:** 112-119.

Strieker M, Marahiel MA. **2009.** The structural diversity of acidic lipopeptide antibiotics. *Chembiochem.* **10:** 607-616.

Strieker M, Nolan EM, Walsh CT, Marahiel MA. **2009.** Stereospecific synthesis of threo- and erythro-beta-hydroxyglutamic acid during kutzneride biosynthesis. *J.Am.Chem.Soc.* **131:** 13523-13530.

Strieker M, Tanovic A, Marahiel MA. **2010.** Nonribosomal peptide synthetases: structures and dynamics. *Curr.Opin.Struct.Biol.* **20:** 234-240.

Svenning MM, Eriksson T, Rasmussen U. **2005.** Phylogeny of symbiotic cyanobacteria within the genus *Nostoc* based on 16S rDNA sequence analyses. *Arch.Microbiol.* **183:** 19-26.

Takahashi A, Agatsuma T, Matsuda M, Ohta T, Nunozawa T, Endo T, Nozoe S. **1992.** Russuphelin A, a new cytotoxic substance from the mushroom *Russula subnigricans* Hongo. *Chem.Pharm.Bull.(Tokyo)* **40:** 3185-3188.

Tamaoki T, Nomoto H, Takahashi I, Kato Y, Morimoto M, Tomita F. **1986.** Staurosporine, a potent inhibitor of phospholipid/Ca++dependent protein kinase. *Biochem.Biophys.Res.Commun.* **135:** 397-402.

Tan LT. **2007.** Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry* **68:** 954-979.

Teaumroong N, Innok S, Chunleuchanon S, Boonkerd N. **2002.** Diversity of nitrogen-fixing cyanobacteria under various ecosystems of Thailand: I. Morphology, physiology and genetic diversity. *World J.Microbiol.Biotechnol.* **18:** 673-682.

Thacker RW, Starnes S. **2003.** Host specificity of the symbiotic cyanobacterium *Oscillatoria spongeliae* in marine sponges, *Dysidea* spp. *Mar.Biol.* **142:** 643-648.

Thomas MG, Burkart MD, Walsh CT. **2002.** Conversion of L-proline to pyrrolyl-2-carboxyl-S-PCP during undecylprodigiosin and pyoluteorin biosynthesis. *Chem.Biol.* **9:** 171-184.

Thurston EL, Ingram LO. **1971.** Morphology and fine structure of *Fischerella ambigua*. *Journal of Phycology* **7:** 203-&.

Thykaer J, Nielsen J, Wohlleben W, Weber T, Gutknecht M, Lantz AE, Stegmann E. **2010.** Increased glycopeptide production after overexpression of shikimate pathway genes being part of the balhimycin biosynthetic gene cluster. *Metab Eng* **12:** 455-461.

Tooming-Klunderud A, Rohrlack T, Shalchian-Tabrizi K, Kristensen T, Jakobsen KS. **2007.** Structural analysis of a non-ribosomal halogenated cyclic peptide and its putative operon from *Microcystis*: implications for evolution of cyanopeptolins. *Microbiology* **153:** 1382-1393.

Townsend CA. **1997.** Structural studies of natural product biosynthetic proteins. *Chem.Biol.* **4:** 721-730.

Unson MD, Faulkner DJ. **1993.** Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia* **49:** 349-353.

Unson MD, Holland ND, Faulkner DJ. **1994.** A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Marine Biology* **119:** 1-11.

Unson MD, Rose CB, Faulkner DJ, Brinen LS, Steiner JR, Clardy J. **1993.** New polychlorinated amino acid derivatives from the marine sponge *Dysidea herbacea*. *J.Org.Chem.* **58:** 6336-6343.

Unversucht S, Hollmann F, Schmid A, van Pee KH. **2005.** $FADH_2$-dependence of tryptophan 7-halogenase. *Adv.Synth.Catal.* **347:** 1163-1167.

Vaillancourt FH, Yeh E, Vosburg DA, Garneau-Tsodikova S, Walsh CT. **2006.** Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem.Rev.* **106:** 3364-3378.

Valerio E, Chambel L, Paulino S, Faria N, Pereira P, Tenreiro R. **2009.** Molecular identification, typing and traceability of cyanobacteria from freshwater reservoirs. *Microbiology* **155:** 642-656.

Vallon O. **2000.** New sequence motifs in flavoproteins: evidence for common ancestry and tools to predict structure. *Proteins* **38:** 95-114.

Van Lanen SG, Lin S, Shen B. **2008.** Biosynthesis of the enediyne antitumor antibiotic C-1027 involves a new branching point in chorismate metabolism. *Proc.Natl.Acad.Sci.U.S.A* **105:** 494-499.

van Pée KH. **2001.** Microbial biosynthesis of halometabolites. *Arch.Microbiol.* **175:** 250-258.

van Pee KH, Dong CJ, Flecks S, Naismith J, Patallo EP, Wage T. **2006.** Biological halogenation has moved far beyond haloperoxidases. *Advances in Applied Microbiology, Vol 59* **59:** 127-157.

van Pée KH, Patallo EP. **2006.** Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl.Microbiol.Biotechnol.* **70:** 631-641.

van Pée KH, Zehner S. **2003.** Enzymology and molecular genetics of biological halogenation. In: 171-199.

Van Sande J, Deneubourg F, Beauwens R, Braekman JC, Daloze D, Dumont JE. **1990.** Inhibition of iodide transport in thyroid cells by dysidenin, a marine toxin, and some of its analogs. *Mol.Pharmacol.* **37:** 583-589.

Van Wagoner RM, Drummond AK, Wright JLC. **2007.** Biogenetic diversity of cyanobacterial metabolites. *Adv.Appl.Microbiol.* **61:** 89-217.

Voldoire A, Moreau P, Sancelme M, Matulova M, Leonce S, Pierre A, Hickman J, Pfeiffer B, Renard P, Dias N, Bailly C, Prudhomme M. **2004.** Analogues of antifungal tjipanazoles from rebeccamycin. *Bioorg.Med.Chem.* **12:** 1955-1962.

von Elert E, Oberer L, Merkel P, Huhn T, Blom JF. **2005.** Cyanopeptolin 954, a chlorine-containing chymotrypsin inhibitor of *Microcystis aeruginosa* NIVA Cya 43. *J.Nat Prod* **68:** 1324-1327.

Wagner C. **2008.** Investigation towards the biohalogenation of the chlorinated natural products from *Fischerella ambigua* (Näg.) Gomont.

Wagner C, El Omari M, Konig GM. **2009.** Biohalogenation: nature's way to synthesize halogenated metabolites. *J.Nat Prod* **72:** 540-553.

Walker GE, Dunbar B, Hunter IS, Nimmo HG, Coggins JR. **1996.** Evidence for a novel class of microbial 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase in *Streptomyces coelicolor* A3(2), *Streptomyces rimosus* and *Neurospora crassa*. *Microbiology* **142 (Pt 8):** 1973-1982.

Walsh CT. **2003.** Enzymatic assembly lines for nonribosomal peptides. *FASEB J.* **17:** A148.

Walsh CT. **2004.** Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303:** 1805-1810.

Walsh CT. **2008.** The chemical versatility of natural-product assembly lines. *Acc.Chem.Res.* **41:** 4-10.

Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM. **2001.** Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr.Opin.Chem.Biol.* **5:** 525-534.

Walsh CT, Fischbach MA. **2010.** Natural products version 2.0: connecting genes to molecules. *J.Am.Chem.Soc.* **132:** 2469–2493.

Walsh CT, Garneau-Tsodikova S, Howard-Jones AR. **2006.** Biological formation of pyrroles: nature's logic and enzymatic machinery. *Nat.Prod.Rep.* **23:** 517-531.

Walsh CT, Gehring AM, Weinreb PH, Quadri LE, Flugel RS. **1997.** Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr.Opin.Chem.Biol.* **1:** 309-315.

Wang S, Xu Y, Maine EA, Wijeratne EM, Espinosa-Artiles P, Gunatilaka AA, Molnar I. **2008.** Functional characterization of the biosynthesis of radicicol, an Hsp90 inhibitor resorcylic acid lactone from *Chaetomium chiversii*. *Chem.Biol.* **15:** 1328-1338.

Wang ZX, Li SM, Heide L. **2000.** Identification of the coumermycin A(1) biosynthetic gene cluster of *Streptomyces rishiriensis* DSM 40489. *Antimicrob.Agents Chemother.* **44:** 3040-3048.

Weaver LM, Herrmann KM. **1990.** Cloning of an aroF allele encoding a tyrosine-insensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. *J.Bacteriol.* **172:** 6581-6584.

Webb BN, Ballinger JW, Kim E, Belchik SM, Lam KS, Youn B, Nissen MS, Yun L, Kang CH. **2010.** Characterization of Chlorophenol 4-Monooxygenase (TftD) and NADH:FAD Oxidoreductase (TftC) of *Burkholderia cepacia* AC1100. *J. Biol. Chem.* **285:** 2014–2027.

Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson DH, Wohlleben W. **2009.** CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *J.Biotechnol.* **140:** 13-17.

Welker M, Marsalek B, Sejnohova L, von Dohren H. **2006.** Detection and identification of oligopeptides in *Microcystis* (cyanobacteria) colonies: toward an understanding of metabolic diversity. *Peptides* **27:** 2090-2103.

Welker M, von Doehren H. **2006.** Cyanobacterial peptides - nature's own combinatorial biosynthesis. *FEMS Microbiol.Rev.* **30:** 530-563.

Welsh EA, Liberton M, Stockel J, Loh T, Elvitigala T, Wang C, Wollam A, Fulton RS, Clifton SW, Jacobs JM, Aurora R, Ghosh BK, Sherman LA, Smith RD, Wilson RK, Pakrasi HB. **2008.** The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc.Natl.Acad.Sci.U.S.A* **105:** 15094-15099.

Werck-Reichhart D, Feyereisen R. **2000.** Cytochromes P450: a success story. *Genome Biol.* **1:** REVIEWS3003.

White EH, Branchini BR. **1975.** Letter: Modification of firefly luciferase with a luciferin analog. A red light producing enzyme. *J.Am.Chem.Soc.* **97:** 1243-1245.

White SR, Lauring B. **2007.** AAA+ ATPases: achieving diversity of function with conserved machinery. *Traffic.* **8:** 1657-1667.

Widboom PF, Bruner SD. **2009.** Complex oxidation chemistry in the biosynthetic pathways to vancomycin/teicoplanin antibiotics. *Chembiochem.* **10:** 1757-1764.

Wiegand C, Pflugmacher S. **2005.** Ecotoxicological effects of selected cyanobacterial secondary metabolites: a short review. *Toxicol.Appl.Pharmacol.* **203:** 201-218.

Wilcox AJ, Laney JD. **2009.** A ubiquitin-selective AAA-ATPase mediates transcriptional switching by remodelling a repressor-promoter DNA complex. *Nat.Cell Biol.* **11:** 1481-1486.

Wittmann M, Linne U, Pohlmann V, Marahiel MA. **2008.** Role of DptE and DptF in the lipidation reaction of daptomycin. *FEBS J.* **275:** 5343-5354.

Woithe K, Geib N, Meyer O, Wortz T, Zerbe K, Robinson JA. **2008.** Exploring the substrate specificity of OxyB, a phenol coupling P450 enzyme involved in vancomycin biosynthesis. *Org.Biomol.Chem.* **6:** 2861-2867.

Woithe K, Geib N, Zerbe K, Li DB, Heck M, Fournier-Rousset S, Meyer O, Vitali F, Matoba N, Abou-Hadeed K, Robinson JA. **2007.** Oxidative phenol coupling reactions catalyzed by OxyB: a cytochrome P450 from the vancomycin producing organism. implications for vancomycin biosynthesis. *J.Am.Chem.Soc.* **129:** 6887-6895.

Wright AD, Papendorf O, Konig GM. **2005.** Ambigol C and 2,4-dichlorobenzoic acid, natural products produced by the terrestrial cyanobacterium *Fischerella ambigua*. *J.Nat Prod* **68:** 459-461.

Wu J, Howe DL, Woodard RW. **2003.** *Thermotoga maritima* 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase: the ancestral eubacterial DAHP synthase? *J.Biol.Chem.* **278:** 27525-27531.

Wu J, Woodard RW. **2006.** New insights into the evolutionary links relating to the 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase subfamilies. *J.Biol.Chem.* **281:** 4042-4048.

Wynands I, van Pee KH. **2004.** A novel halogenase gene from the pentachloropseudilin producer *Actinoplanes* sp ATCC 33002 and detection of in vitro halogenase activity. *FEMS Microbiol.Lett.* **237:** 363-367.

Yang Q, Pando BF, Dong G, Golden SS, van Oudenaarden A. **2010.** Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science* **327:** 1522-1526.

Yeh E, Blasiak LC, Koglin A, Drennan CL, Walsh CT. **2007.** Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* **46:** 1284-1292.

Yeh E, Garneau S, Walsh CT. **2005.** Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc.Natl.Acad.Sci.U.S.A* **102:** 3960-3965.

Yeh E, Kohli RM, Bruner SD, Walsh CT. **2004.** Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *Chembiochem.* **5:** 1290-1293.

Yoshizawa S, Matsushima R, Watanabe MF, Harada K, Ichihara A, Carmichael WW, Fujiki H. **1990.** Inhibition of protein phosphatases by microcystins and nodularin associated with hepatotoxicity. *J.Cancer Res.Clin.Oncol.* **116:** 609-614.

Zanette J, Jenny MJ, Goldstone JV, Woodin BR, Watka LA, Bainy AC, Stegeman JJ. **2009.** New cytochrome P450 1B1, 1C2 and 1D1 genes in the killifish *Fundulus heteroclitus*: Basal expression and response of five killifish CYP1s to the AHR agonist PCB126. *Aquat.Toxicol.* **93:** 234-243.

Zehner S, Kotzsch A, Bister B, Sussmuth RD, Mendez C, Salas JA, van Pee KH. **2005.** A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem.Biol.* **12:** 445-452.

Zeng J, Zhan J. **2010.** A novel fungal flavin-dependent halogenase for natural product biosynthesis. *Chembiochem.* **11:** 2119-2123.

Zerbe K, Pylypenko O, Vitali F, Zhang W, Rouset S, Heck M, Vrijbloed JW, Bischoff D, Bister B, Sussmuth RD, Pelzer S, Wohlleben W, Robinson JA, Schlichting I. **2002.** Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. *J.Biol.Chem.* **277:** 47476-47485.

Zhao B, Bower MJ, McDevitt PJ, Zhao H, Davis ST, Johanson KO, Green SM, Concha NO, Zhou BB. **2002.** Structural basis for Chk1 inhibition by UCN-01. *J.Biol.Chem.* **277:** 46609-46615.

Zhou H, Qiao K, Gao Z, Meehan MJ, Li JW, Zhao X, Dorrestein PC, Vederas JC, Tang Y. **2010.** Enzymatic synthesis of resorcylic acid lactones by cooperation of fungal iterative polyketide synthases involved in hypothemycin biosynthesis. *J.Am.Chem.Soc.* **132:** 4530-4531.

Zhou Z, Lai JR, Walsh CT. **2007.** Directed evolution of aryl carrier proteins in the enterobactin synthetase. *Proc.Natl.Acad.Sci.U.S.A* **104:** 11621-11626.

Zhu X, De Laurentis W, Leang K, Herrmann J, Ihlefeld K, van Pee KH, Naismith JH. **2009.** Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J.Mol.Biol.* **391:** 74-85.

8. Appendix

8.1 Translated sequences of the *ab*-operon

Sequences of genes probably related to the ambigol biosynthesis were obtained from fosmid E8 and from 454 sequencing of *F. ambigua*.

8.1.1 Amino acid sequence of the putative $FADH_2$-dependent halogenase Ab1

M S N L P K S T K V L V V G G G P A G T T A A T L L A R E G F D I T L L E R E V
F P R Y H I G E S L L P S S L K V L D L L G V R D K I D A H G F Q Y K P G G H Y
H W G D E H W D L N F S D L S G N I T H S Y Q V R R D E F D K L L L D H A K S
Q G V K V F D G I G V S S L S F E N E R P K S A I W S Q T N D K N H T G E I S F
D F L I D A T G R Y G L M A N H H L K N R E Y H D V F Q N V A I W G Y W K N A
D R L D N G R E G A I I I E S L K D G W L W G I P L H D G T I S V G L V V H K T I
Y K E K R S K S L K D I Y L E G I A E S L D L K R L L E P G E L A S E V R S E Q D
Y S Y A A D S F A G Q G Y F M I G D A A C F L D P L L S T G V H L A T F S G L L
S A A S L A S V I R N H I T E E Q A I S F F E R T Y K Q A Y L R L M A M V S A F Y
E N S K K E S Y F W Q A Q Q L T K T R Q S N E D K E K L H Q M F L N V V S G M
E D M S D A E E N S E E L F L E L S E R L R E N W S L R H K Q T A N D L D Q T
E E E K L R A S N Q F V S R L N G L F S L S K E S A V E G L Y I V T T P Q L G L
V Q V N

8.1.2 Amino acid sequences of the putative CYP 450 enzymes Ab2 and Ab3

Ab2:

M L Y Q E V A S T A F N K I A P P G P P V L P F V D M L P C L G K H L H L A L N
Q L A K K Y G N I F Q I R V G A K T L V V L N G L E T I K E A L V K Q P D S F N A
R A D F D I Y Q Q P P Q A Q F L E L K S G E S W R K H H N I L G Q A M H T F V
V G K P D M L E S W A L E E A A D L A N I F F K F S G Q P F D P D L Y M P L A T
L S F M Q R L I F D K R G A I K N P E E D H E F V A S A Y T L K H I P T V L E A V
R L E Y I P K I W Q P I F R L S R W K S L R N F L K S L V A L E S Y V S K N V A Q
H Q E S F D P E N L R D I T D A L L K A S S E L T E S D R N N L H L S E N D I V N
G S L M Q L A G A G A G L A S F M L R W G V L Y M M T Y P A I Q A E I H K E L
D E V V G R Q Q Q P C L E H R G K L P F T E A C I H E I F R H S S I T T M P P I T
Y A T T T D V T L E D Y F I P Q N T P L L I N Y Y S L T R D Q R Y W E E P E Q F
N P Y R F L D E N G K L R K N L L D K F Y P F G M G S R R C I G E Y L G R L L V
F T F F T N L M H K C K F E K V P G E K L S F E S I P G A F I I P E K Y R V V V K
P S F

Ab3:

M V S Q Q S A S T E L N T L N K I S P P G P P A L P F V G M L P F L G K H L H L
A L N Q L A K K Y G N I Y Q I R V G A R T L V V L N G L E T I K E A L V K Q P D S
F N A R A D F D L Y Q Q P P Q C H F M E Q K S G E S W R K H R T I I G Q V M H
T F V A S R S G T L E S W A L E E A A D L A N I F V N S S G Q A F D P N L Y L P
L A T L S F I Q R L I F G K R G N L D D P E K D T D F V T T A H S M G K L N N G
A Q N L T K L V L L P T I W R P I L M I S I W K S L R G F V K A A D A A E G Y L I K
N V E Q H Q K S F D P E N L R D I T D A L L K A S S E L T E S D R N N L G L S E
N D I V N G S L T Q F V G A G T E L P S L M L R W A L L Y M I T Y P A I Q A E I H
K E L D E V V G R Q Q Q P C S E H R G K L P F T E A C I H E V F R H S S A T T T
P A F I Y A T T T D V T L D G Y F I P Q N T P L L V D Y Y S L T R D E R Y W E K P
E Q F N P Y R F L D E N G K L R K N L L D K F H P F G I G S R R C I G E Y I G R
L L I F T F F T H L M H K C K F E K V P G E K L S L D P Q P A I I L P P Q N Y K V I
A K P R F

8.1.3 Amino acid sequence of the putative polyketide cyclase Orf22

M S K I Y D L N Q K S D F S G A T D S F K T E K E P E S L N S A T N P A I L Q D
V E I K I E K L E G R Q R R I F A K I Q I P Y P L E Q V W Q V L T D Y E A F A K F
M P N M T Q S R R L E H S T A S I C V E Q V R T K S F M G M K F S A R S V F D
V E E K F P H E I H Y Q L I E G D F K A C S G Y W R L E P W N S S D E K A G V
D L I Y N F L I L P K P I F P M P L V E N I L S H D I P V G I L A I R Q R V E E L F S
S K

8.1.4 Amino acid sequence of the putative chorismate lyase Ab4

M I H L Q K E S L S T V D L Q D E A S L P I S I L T N N S Q E S P S R N P I D P S
T L S T F Q R I L L T T N G T V T D I L E Y Y A F E Q I R V V K L A E Q L V S L A H
E I P M M E L K E G T E V L V R K I L L Q G K I S R K N F L Y A D S I I V P E R L D
E R F R K A L L E T K M P I G K L W F E L R V E T F K E V L D T S K E V A G N L
A D Y F Q I Q P D D N I L S R T Y R V I N N R K P V M M I T E K F P E N Y Y L K C
S

8.1.5 Amino acid sequence of the putative AMP-ligase Ab5

M L T Q L F T E V V S N Y P E K T A I V Y D Q T K I S Y Q T L Y S Q I K S F S Q G
L G S I G I D Q G D C V A L L L P N C P E F V I S F Y A I A R L N G V V L P L N H
L F K A E E V S H Y L N D S D V K A I I T D S Q R A D I C K K I I F N L G K K I E L I
V V D Q A P P P A K Y F Y D L I L P N S T E I H E S V L P Y E G N V L Y Q Y S S
G S T G R P K R V S R T Q K N L Y H E A R N F T E S V K V T P S D N I L C T V P
L Y H A H G L G N C L L A A T C N G L T L V I L E Q S I Q N G V S V E V P F V F K
C P R I L E L I K T E K I S I F P G V P Y I F N S L A E T P V N I Q A D L S T L K L C
I S A G N F L G K D V F N K F L Q R F G V P I R Q L Y G C T E A G A M C I N L D

E N P E Q T W D S V G T P L K N V G I K I I N E Q G H E L S V G Q T G E I L I K S
Q A L T N G Y D N I P D L N Q Q A F K E G T F F T G D L G K L D E A G R L Y I T
G R K K I L I D T G G R K V D P I E I E D I L N T H P Q V K E A V V V G V K G A H
A G E L V K A V I V L K E P E Q C D E Q K I F S Y C K E R L A E F K V P K I I E F
R N E I P K S P L G K I L R K A L V V

8.1.6 Amino acid sequence of the putative DAHP synthetase Orf23

M D I N K I L D T N I K S L S V L M A P M Q M K E Q L P I T P V A T E T V L R G R
Q A V K E I L D G K D S R K F I I V G P C S I H D V K A T L E Y A E K L K T L A D
K V Q D K L L I L M R V Y F E K P R T T I G W K G L I N D P D L D D S F N I Q K G
L L T A R N L L I N I A E L G L P S A T E A L D P V T P Q Y I S D L I S W A A I G A
R T I E S Q T H R E M A S G L S M P V G F K N G T D G N I Q V A L D A I Q S S R
N P H H F L G I D Q I G Q I S I F Q T K G N V Y G H I I L R G G G G Q P N F D A A
T V A W V E K K L E N L K L P K R I V I D C S H G N S Y K N H Q L Q T A V F N N
V L Q Q I T D G N Q S M I G M M L E S N L Y E G N Q K I P S D L N Q L K Y G V S
V T D K C I G W E E T E E I I L S A H E R L S A D R N V M L H T C G M V V S G T
P V R N L M A T K G

8.1.7 Amino acid sequence of the putative NRPS-module Ab6

M I S S M S L E T L R D R N Q N I H P A S G L Q P Q N Q T K S S E E A Q K D L
W E A I K T V I S L Q N Q A P P L V S V S R Q G N I P L S F S Q E R L W F L E Q
L E P S R S S A Y N M P S A F R I T G A L N V S V L Q Q S L N E I L R R H E A L
R T T F A F R E G K S V Q V I H P A L T L N L P I I E L Q N I S S E Q Q H I K T M
Q L I R E E V Q R P F D L S Q L P L L R A T L L R L S E N E H L L L L S V H H I V I
D F W S K G I L F Q E L S V L Y E A F S T G K P S P L S E L P I Q Y A D F A V W
Q R Q W L K G E F L E V L L N Y W K Q Q L D S N L S E L H L P T D R A R S M L
Q T R D G A N Q K L V L S K E L T K E L K A L S R Q E G T T L F V V L L A A F K
V L L Y R Y T E Q D D I F V C S P I A N R N R K E V K G L I G Y F V N L L I L R T
S L S D N P S F R E L L G R V R K V T S G G Y A Y Q D L P V Q Q L V K S L N L L
Q T P L S R V M F G L Q N T A I H S L N L P G L T V R S V D I E G G T A D F D L
Y L Y V L E E G S T L T A T L K Y N T D L F D D S T I V Q L L N H F Q T V L E N I
A V D S G Q S I P L L L P L S T A E Q Q Q L T D K R L E Q S S L K P E G V Y V A
P R N P L E L Q L T Q I W S Q V L G I Q S V G V K D N F F E L G G E S L L A M S
L F A K I E K I F G K T L P L T T L L Q A P T V E Q F A Q L L T Q D A N S V S W S
S L V P I Q P S G T K P P L F C I H G Q Q G N V L N F R K L S Q Y L G S D Q P F
Y G L Q A K G L D G K E L P L F R I E D I A T H Y I Q E I R T L Q P E G P Y F L A
G N S M G G T I A F E M A Q Q L H K Q G Q K V A L L V M F D T F G L D C F P R
L S L R R Q H Y W A Y L L Q L G I S K F L L N E V N E L C Q R R L K E M I S R L
Y L S L G R P L P Q N L R D E L V A E A N M Q A K I G Y Q A Q V Y P G R V T L L
R A S Q P A L F P K L Y L P T S E D W Y N R N P E H G W S E V V G G G L E I H
D V P G D H F S I F E E P H V Q V L A E K L K A C L D E A Q T K Y

8.1.8 Amino acid sequence of the putative $FADH_2$-dependent halogenase Ab7

M K N I Y D V A I C G S G L A G L T L A R Q L K L K M P D I S V V V L D R L A R P
L P E A G F K V G E S S V E V G A F Y L A H I V Q L E D Y L E K Q H L H K L G L
R Y F L G D T K G P F H K R P E I G L S K Y H F P N S Y Q L D R G K L E N D L R
S I N T E A G V E L L E G C L V K D I E L G D P Q Q L H Q I I Y T Q E N N K A T Q
A I Q A R W V V D S M G Y R R F L Q R K L G L A K P K N S Q F S A V W F R V E
G R F D V S D F V P S T E I E W H E R V P H N N R Y Y S T N H L C G E G Y W V
W L I P L S T G Y T S I G I V T N E E I H P F G T Y H T Y E K A F Q W L E K H E P
V V A F H L K S N P P V D F M K I P Q Y S Y S S N Q V F S I N R W A C V G V A G
V F A D P F Y S P G T D L I G F G N S L I T Q M V E L D R E N K L T P E I V N E A
N R F L I T Y N E S V T S N I H N A Y L C F G N E T V M V M K Y I W D V L S A W
A F S A P M M F N S L F L D S D K R A K V R K G T G Q F F L L A Q R M N Q L F
R D W A V Q S Q R R T S F E F I D Y L Q I P F V R E L R A R N L K T N K T E Q E
L I D D H L A S I K L F E E L A Q V I F L L A L E D T M P E K S A D F P S P V W L
N A W V V S L D D K R W E I D G L F R P T S K P R D L R P M M E Q L W Q N I H
F R A A D R D S S L I T A

8.2 Translated sequences of the *tj*-operon

Sequences of genes probably related to the tjipanazole D biosynthesis were obtained from 454 sequencing of *F. ambigua*.

8.2.1 Amino acid sequence of the putative L-tryptophan oxidase Tj1

M N I E T K L Q Q C K R A N H Q P S K H V T I L G A G I A G L I A A Y E L E R L G
H Q V D V M E G S P R I G G R I W T H R F G D P I D G P Y G E L G A M R I P P
Q H E Y V L H Y V Q T L G L G D K L R K F V T M F E E H N A L V N I D G T V L R
I Q D A P R S I Q M H Y R G V F V D E R Y S E K T R L F A A W L K T I V D A I G S
G E L R E C L E R D L N S H L M D E L K K L D L D P F F D V C G K T I D L H A F I
K A Y P S F R N K C S K A L N I F L S D I L I E T S H D L W Q L E G G M D Q L I H
S L A A A I N G P I R C N Q N V V A L R V Y Q D G V E I S W L E A G K L H T R L
C D Y V L C T I P F S V L R Q I E L S G L D E D K L A T I H N I I Y W P A T K V L F
L C A R P F W Q E S G I F G G A S F S D D A V R Q I Y Y P S T K S H S S L N S V
L L A S Y T I G N D A R Y L G M M S E Q E R H S Y V K E A L S K L H P E I K V P
G T L L D M T S I A W G N Y K W S A A G C S I P W D R S A F Y Q S P Q H Q D A
I P H T E S V Y H Q K A A S P Q N T L F F A G E H C S K L P A W L Q G S V E S A
L Q A V Y D I V S H N S S V N P N K K P S L V L E K G

8.2.2 Amino acid sequence of the putative chromopyrrolic acid synthase Tj2

M S I F D F P R L H F Q G F A R I H A P T G H K N G L V D L S T N T V Y M N G D
R F D H R R P V S E Y H E Y L K N L G P K F N V E G H W D E D G P F S M A M
G W D F G G N G H F A I E A S I V S V Q R Q P G E L D Q R D P V V S R N I D M
W G H Y N D Y I G T T F N R A R I F D C D P A S H W T T T I M I G Q L T F G R Q
G A S Y E V P Y M L S A P V E G V Q P A R W Q D F N H I H E L S E H C L N N E
F K R A A V H Q F T I S K E A E N F L W G E E S T L S P T V S L L R E A M N Q D
D V L G L V V Q F G L S N M S S P L K P D F P V F W E L H G T I G L W R K G E L
T N Y P Y G R L L T P H R A A H D L E Q T T S R S L S N L T V Q V T P Q G V S L
N M V T A V P C V R R A A K P G P G P I H A I G D K L N L G D L E L R T V A S Q
L L V A R I P Q Q A Y Q R Q A Y Q L T S G I V D V P L A A P F D Q L R D E I E H
Q G L C I T V T Q S D G Q H R V L V E E E E I N I Q V E Q A C L F V E F P D W K
R S E E H V L E M E V H S F I R G R P A S V E A I Y L R Q F Y N P R G L P Q L R
Y A F E H D R R K H E V Q A L S T L S V S A G I R L R E S T G G N S A Q T E H T
F H F P R S S E L D I V H F K P G R K E E T G D F A S T C V I S T D E Q G R G W
V T L R G A Q P G T A R I L L S D R S N V F P C D P N H P D E A I I S Y D N D D L
L G F W S G I G S C A V R V L T N D W H L E E V E D K A V D F D L I Y E H I L A
Y Y E L A F S F M K A E V F S L A D R C K V E N Y A R L M W Q M C D P R N K D
K T Y Y M P P T R D M S Q P K A M L L R K F L R N Q Q Q I G Y V P N S T P V P
K R P Q R I L Q T R D D L V I A L K H A A E L E A A V M L Q Y I Y A A Y S I P N Y
A T G Q E Y V R Q G L W T L N Q L L L A C G D D Q E V R N F G I R G V L L D I S
H E E M I H F L M V N N I L M A L G E P F Y P A M P D F G K L N Q R F P I E V D
F A L E P F S A T T L Q R F M R F E W P D F L A E D L T N K A A S T D E P S V N
N L H G Y S S L S E L Y R Q I R E A L K N I P D L I L A E K G R V G G E H H L F M
R Q D F N T V H P D Y Q L Q V D D I D S A L F A I D Y I V E Q G E G C D P K S P
K F E Q S H F Q K F R R L A E A L A K E H I N D E T G C Q V P W T P A Y P A L R
N P S L Y H Q D Y H T H I V T I P Q T R A V M Q I F D E C Y F L M M Q L M V Q Q
F G W M P T G S L R R S K L M N A A I D I M S G M M R P L G E L L M T M P S G
K R G K T A G P S F Q T I T L P Q Y I P T P K V A Y Q L I A R R F E D L T H Q A R
A C E A I H S T V C D L F E F Y A R F F E D L A N Q P I D A A

8.2.3 Amino acid sequence of the putative O-methyltransferase Orf2

M T Q A P V A P A N L N Q Q N A L P P S E A L M Q M L D G H Y L S Q A I F V A
A K L G I A D L L K D G T K S T D E L A K V T E V N S Q F L Y R I L R A L S S V G
I F A E V G D R N F E L T P L A K Y L Q S D V P G S M R L P A I L V G E E W H W
Q A W G N M F N A V K N G T S A F E A K F G T N I V D Y F G Q N P Q Q S K V F
F E A M T T Y S V I V N N A I L E V Y D F S A F S K L V D V G G G L G S L L T D I
L K A N P Q L T G I L L E L P P V I E R A K Q Q N H F Q T K E I S G R Y E I V G G
D F L E S V P S G G D V Y I L K Q I I H N L N N D D A I K L L Q N C H D A M S T N
G K L L V I D P V I P S S N E P S F S K L L D L Q M M V T H G A Q V H T A N E F
Q D I L T K T G F Q I T N I I P T K S P C T I I E S V K K

8.2.4 Amino acid sequence of the putative NAD(P)H-dehydrogenase Orf3

M A H I L H I D S S P R G E R S F S R L F S N E F I S K W K N A H P E D T I T Y R
D L G H N P V P P V D E L W V A G A F T P P E V R P P E L A K G I E L S D A L I
D E F L A A D R Y V F G V P M F N L I A P S T F K A Y I D Q I V R I G R T F A Q D
N Y G V I K G L V E G K K M L I I T A R G A S Y R P E T P V A H L D F Q E P Y L R
A I F G F I G I T D I S F I H V D N L N M G D D V R Q Q S L A R A R E E I A R T I T
T W

8.2.5 Amino acid sequence of the putative FAD-binding monooxygenase Tj3

M A Q T K N Q E D L I D V L I V G A G P V G L T L A A E L L R L G V Q F R I L E T
R P E P E K Y S K A A N L W P R T Q E V F A A I G V I D R L L A E S V P I R T A T
L Y A Y G K R M G H V T I D G Y S S P Y G T P V M I G Q N R I E R I L S D H L V
Q A G R S V E R G I T F T G L H Q N T D Y V E A T V E G N G K H E T V H C R F
L V G C D G N K S R V R N A I G L S I H P E R L E R R F M R Q I D A R V R W S R
S V R D D Q I W F F L F D T G Y I G V L P L P E G Y H R F W I I D D D Q G V P D
R D P T L E E M Q E V V Q R I T G D T Q V E L Y D P I W F S H G R F Q H G V A
S A L R K N R V I L A G D A G H I P I P I S G K G M N T G I Q D A F N L G W K L A
A T L H E Q V S S V V L N S Y S I E R Q K I R Q Q L D T T Q V A G F H W I M E P
S K I Q Q H L V R K L G S V W L N L I A A R F F K Q R L S Q L D I A Y P D S V L S
Q D L L G K K G S C A G D R A P D A V V V A I P G Q Y T I T L F K L I Y E G L N
W T L L L F D G A Q G S E I L E Q L Q T I A T A I A K E F S T I R V W P V L I A P V
L A D P K H I P M L L D F D S F A H K A F G L E N P A L V L I R P D G H V A F R A
A I N D Y Q A L Q M Y A R Q V F K M P P Q N N Y K S V S Q A E L T K I Q S

8.2.6 Amino acid sequence of the putative CYP 450 enzyme Tj4

M I E R L N P F L P E F I A N P Y A F Y S R Y R E E D P V H W G I T S N S R L P
G A W Y L F R Y E D V V K V F E D P R F G R E A R R V L S D N Q A P L V P P A
Y K G F L S M V S N W L V F R D P P D H T R L R S L V N K V F S Q R V V E N I R
P A I F S I A D S L L D Q V H A R G E M D L I E E F A F P L P V M V I A A L L G V
D P K D R P L F R Q W S M A L L E A S A S R L T P S P E I Y S R A E Q A T Q G F
I D Y F T E A I A Q R R A E P R E D L I T D L I K A Q D E G D K L S E Q E V L S M
C I H L L T A G H E T T V N L I S K G M L S L L R N R D T F K I L R T H P E L L P
G A V E E L L R Y D S P V Q M V T R W A Y E D V E I G G K L I R R G D S V G L
M I G A A N R D P L R F E N P D V L D I K R E D C R H C V F G G G I H F C I G S
A L A R A E G Q I A L N V L L N R L P E L R L A E Q T L E W H G T I V F H G P K
H L W V T F R P P T I P S A M P T P V V T S V N S D

8.2.7 Amino acid sequence of the putative $FADH_2$-dependent halogenase Tj5

M E C L L S W T F L M N V S D A F R N L Q P R S Q D K L K G E T L I M S Q P V
Q N L V I V G G G T A G W M S A V Y L R K A L D K N I Q I T L V E S S D V T T I G
V G E A T F S T I K L F F D Y L G L Q E H E W M P K C N A S Y K T A I K F N N W
N T Q G Q H F Y H P F Q R Y E T V N G F D I A E W W L K I K R D E E P F D Y A
C F L I P S I C D H K K S P R Y L D G T V F D D K V K D L I S R E F I P E K N V L
S D H K V Q Y P Y A Y H F D A N L L A R F L K D Y A K Q R G V K H V I D H V E N
V K L A E D G S I D S I I T R E H G N V S G D L F V D C T G F R G L L I N E A L D
E P F N Y F S D S L L C D R A I A M Q I P T D I K K D G I N P F T A A T A L S S G
W V W N I P L Y G R D G T G Y V Y C S A F L S E E E A E Q E F R Q H L G P A A
L N C K A K H I K I R V G R N R N S W V K N C V A I G L A S G F V E P L E S T G I
F F I Q H G I E E L V S H F P D K T F N E E L I K S Y N N T I A E C I D G V R D F L
T L H Y C A S D R T D T P F W K A T K Q E I K I P E Q L N E K L R L W K T R L P
N N K N V N H N Y H G F D S Y S Y S V M L L G L N Y L P E S S L P A L N H I D E
R E A I A V F N S I K Q K A N H L A A T L P S Q Y E Y L T Y V I K S Q E Q S E Y L
R E E S L V G V

8.3 Nucleotide Sequences of 16S rDNA analyses

8.3.1 Partial 16S rDNA sequences of *Fischerella ambigua*

FA16S1-T7-4.3:

ATTAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGTATGCTTAACACATGC
AAGTCGAACGGAGTGCTTCGGCACTTAGTGGCGGACGGGTGAGTAACGCGTGA
GAATCTAGCTTTGGGTTCGGGACAACCACGGGAAACGGTGGCTAATACCGGAT
GTGCCGAGAGGTAAAAGATTAATTGCCCGAAGAAGAGCTCGCGTCTGATTAGCT
AGTTGGTGTGGTAAGAGCGCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGG
ATGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAG
CAGTGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAATACCGCGTGA
GGGAGGAAGGCTCTTGGGTTGTAAACCTCTTTTCTTAGGGAAGAAGAACTGACG
GTACCTAAGGAATAAGCATCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACG
GAGGATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCCGCAGGTGGCA
ATGAAAGTCTGCTGTCAAAGAGTCTGGCTTAACCAGATAAAGGCAGTGGAAACT
ACATAGCTAGAGTGAGGTAGGGGCAGAGGGAATTCCTGGTGTAGCGGTGAAAT
GCGTAGAGATCAGGAAGAACACCGGTGGCGAAGGCGCTCTGCTGGACCGCAA
CTGACACTGAGGGACGAAAGCTAGGGGAGCGAATGGGATTAGATACCCCAGTA
GTCCTAGCTGTAAACGATGGATACTAGGCGTTGCCCGTATCGACCCG

FA16S2-T7-4.3:

GGAGGTGATCCAGCCCCACCTTCCGGTACGGCTACCTTGTTACGACTTCACCC
CAGTCACCAGTCCTACCTTCGGCATCCCCCTCCTCGAAAGGTTAGGGTAACGAC
TTCGGGCGTGACCAGCTTCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGG
GAACGAATTCACTGCAGTATGCTGACCTGCAATTACTAGCGATTCCGACTTCAC
GCAGGCGAGTTGCAGCCTGCGATCTGAACTGAGCCGTGGTTTCTGGGATTTGC
TTGCGTTCGCACGCTTGCTGCCCTTTGTCCACAGCATTGTAGTACGTGTGTAGC
CCAGGACGTAAGGGGCATGCTGACTTGACGTCATCCCCACCTTCCTCCGGTTT
GTCACCGGCAGTCTCCTTAGAGTTCCCAACTTAATGATGGCAACTAAGTACGAG
GGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGAC
AGCCATGCACCACCTGTGTTCCGGTTCCCGAAGGCACCCCTCCCTTTCAAGAG
GGTTCCGGACATGTCAAGTCCTGGTAAGGTTCTTCGCGTTGCATCGAATTAAAC
CACATACTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCACACTT
GCGTGCGTACTCCCCAGGCGGGATACTTAACGCNTTAGCTACGACACTGCCCG
GGTCGATACGGGCAACGCCTAGTATCCATCGTTTACAGCTAGGACTACTGGGGT
ATCTAATCCCATTCGCTCCCCTAGCTTTC

FA16S3-T7-4.3:

TTAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGTATGCTTAACACATGCA
AGTCGAACGGAGTACTTCGGTACTTAGTGGCGGACGGGTGAGTAACGCGTGAG
AATCTAGCTTTGGGTTCGGGACAACCACGGGAAACGGTGGCTAATACCGGATG
TGCCGAGAGGTAAAAGATTAATTGCCCGAAGAAGAGCTCGCGTCTGATTAGCTA
GTTGGTGTGGTAAGAGCGCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGA
TGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGC
AGTGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAATACCGCGTGAG
GGAGGAAGGCTCTTGGGTTGTAAACCTCTTTTCTTAGGGAAGAAGAACTGACGG
TACCTAAGGAATAAGCATCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGG
AGGATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCCGCAGGTGGCAA
TGAAAGTCTGCTGTCAAAGAGTCTGGCTTAACCAGATAAAGGCAGTGGAAACTA
CATAGCTAGAGTGAGGTAGGGGCAGAGGGAATTCCTGGTGTAGCGGTGAAATG
CGTAGAGATCAGGAAGAACACCGGTGGCGAAGGCGCTCTGCTGGACCGCAACT
GACACTGAGGGACGAAAGCTAGGGGAGCGAATGGGATTAGATACCCCAGTAGT
CCTAGCTGTAAACGATGGATACTAGGCGTTGCCCGTATCGACCCGGGCAGTGT
CGTAGCTAACGCGTTAAGTATCCCGCCTGGGGAGTACGCACGCAAGTGTGAAA
CTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGTATGTGGTTTAATTC
GATGCAACGCGAAGAACCTTACCAGGACTTGACATGTCCGGAACCCTCTTGA

FA16S2-T7-27.2:

GGAGGTGATCCAGCCCCACCTTCCGGTACGGCTACCTTGTTACGACTTACCCCA
GTCACCAGTCCTACCTTCGGCATCCCCCTCCTCGAAAGGTTAGGGTAACGACTT
CGGGCGTGACCAGCTTCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGA
ACGAATTCACTGCAGTATGCTGACCTGCAATTACTAGCGATTCCGACTTCACGC
AGGCGAGTTGCAGCCTGCGATCTGAACTGAGCCGTGGTTTCTGGGATTTGCTT
GCGTTCGCACGCTTGCTGCCCTTTGTCCACAGCATTGTAGTACGTGTGTAGCCC

AGGACGTAAGGGGCATGCTGACTTGACGTCATCCCCACCCTCCTCCGGTTTGT
CACCGGCAGTCTCCTTAGAGTTCCCAACTTAATGATGGCAACTAAGTACGAGGG
TTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAG
CCATGCACCACCTGTGTTCCGGTTCCCGAAGGCACCCCTCCCTTTCAAGAGGG
TTCCGGACATGTCAAGTCCTGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCA
CATACTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCACACTTGC
GTGCGTACTCCCCAGGCGGGATACTTAACGCGTTAGCTACGGCACTGCCCGGG
TCGATACGGGCAACGCCTAGTATCCATCGTTTACAGCTAGGACTACTGGGGTAT
CTAATCCCATTCGCTCCCCTAGCTTTCGTCCCTCAGTGTCAGTTGCGGTCCAGC
AGAGCGCCTTCGCCACCGGTGTTCTTCCTGATCTCTACGCATTTCACCGCTACA
CCAGGAATTCCCTCTGCCCCTACCTCACTCTAGCTATGTAGTTTCCACTGCCTTT
ATCTGGTTAAGCCAGACTCTTTGACAGCAGACTTTCATTGCCACCTGCGGACGC
TTTACGCCCAATCA

FA16S4-T7-27.2

GGAGGTGATCCAGCCCCACCTTCCGGTACGGCTACCTTGTTACGACTTCACCC
CAGTCACCAGTCCTACCTTCGGCATCCCCCTCCTCGAAAGGTTAGGGTAACGAC
TTCGGGCGTGACCAGCTTCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGG
GAACGAATTCACTGCAGTATGCTGACCTGCAATTACTAGCGATTCCGACTTCAC
GCAGGCGAGTTGCAGCCTGCGATCTGAACTGAGCCGTGGTTTCTGGGATTTGC
TTGCGTTCGCACGCTTGCTGCCCTTTGTCCACAGCATTGTAGTACGTGTGTAGC
CCAGGACGTAAGGGGCATGCTGACTTGACGTCATCCCCACCTTCCTCCGGTTT
GTCACCGGCAGTCTCCTTAGAGTTCCCAACTTAATGATGGCAACTAAGTACGAG
GGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGAC
AGCCATGCACCACCTGTGTTCCGGTTCCCGAAGGCACCCCTCCCTTTCAAGAG
GGTTCTGGACATGTCAAGTCCTGGTAAGGTTCTTCGCGTTGCATCGAATTAAAC
CACATACTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCACACTT
GCGTGCGTACTCCCCAGGCGGGATACTTAACGCGTTAGCTACGACACTGCCCG
GGTCGATACGGGCAACGCCTAGTATCCATCGTTTACAGCTAGGACTACTGGGGT
ATCTAATCCCATTCGCTCCCCTAGCTTTCGTCCCTCAGTGTCAGTTGCGGTCCA
GCAGAGCGCCTTCGCCACCGGTGTTCTTCCTGATCTCTACGCATTTCACCGCTA
CACCAGGAATTCCCTCTGCCCCTACCTCACTCTAGCTATGTAGTTTCCACTGCCT
TTATCTGGTTAAGCCAGACTCTTTGACAGCAGACTTTCATTGCCACCTGCGGAC
GCTTTACGCCCAATCATTCCGGATAACGCTTGCATCCTCCGT

FA16S5-T7-27.2

TTAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGTATGCTTAACACATGCA
AGTCGAACGGAGTACTTCGGTACTTAGTGGCGGACGGGTGAGTAACGCGTGAG
AATCTAGCTTTGGGTTCGGGACAACCACGGGAAACGGTGGCTAATACCGGATG
TGCCGAGAGGTAAAAGATTAATTGCCCGAAGAAGAGCTCGCGTCTGATTAGCTA
GTTGGTGTGGTAAGAGCGCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGA
TGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGC
AGTGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAATACCGCGTGAG
GGAGGAAGGCTCTTGGGTTGTAAACCTCTTTTCTTAGGGAAGAAGAACTGACGG
TACCTAAGGAATAAGCATCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGG

AGGATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCCGCAGGTGGCAA
TGAAAGTCTGCTGTCAAAGAGTCTGGCTTAACCAGATAAAGGCAGTGGAAACTA
CATAGCTAGAGTGAGGTAGGGGCAGAGGGAATTCCTGGTGTAGCGGTGAAATG
CGTAGAGATCAGGAAGAACACCGGTGGCGAAGGCGCTCTGCTGGACCGCAACT
GACACTGAGGGACGAAAGCTAGGGGAGCGAATGGGATTAGATACCCCAGTAGT
CCTAGCTGTAAACGATGGATACTAGGCGTTGCCCGTATCGACCCGGGCAGTGT
CGTAGCTAACGCGTTAAGTATCCCGCCTGGGGAGTACGCACGCAAGTGTGAAA
CTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGTATGTGGTTTAATTC
GATGCAACGCGAAGAACCTTACCAGGACTTGACATGTC

8.3.2 Partial 16S rDNA sequences of the associated *Pseudomonas* sp.

Pstu-1-T7

CTCCGGCCGCATGGCCGCGGGATTCAGATTGAACGCTGGCGGCAGGCCTAAC
ACATGCAAGTCGAGCGGATGAGTGGAGCTTGCTCCATGATTCAGCGGCGGACG
GGTGAGTAATGCCTAGGAATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGA
AACGCTAATACCGCATACGTCCTACGGGAGAAAGTGGGGGATCTTCGGACCTC
ACGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGAGGTAAAGGCTCAC
CAAGGCGACGATCCGTAACTGGTCTGAGAGGATGATCAGTCACACTGGAACTG
AGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGG
GCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTA
AAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTTTGACG
TTACCAACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGA
AGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGT

Pstu-3-T7

CACTAGTGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTG
GATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGG
TCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTC
ACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGC
CAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGG
GCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGC
GGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCA
GGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACG
AGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTA
TAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCG
ACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTAT
CCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG
CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACA
GAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC

Pstu-4-T7

CTCCGGCCGCCTGGCCGCGGGATTCAGATTGAACGCTGGCGGCAGGCCTAAC
ACATGCAAGTCGAGCGGATGAGTGGAGCTTGCTCCATGATTCAGCGGCGGACG
GGTGAGTAATGCCTAGGAATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGG
AACGCTAATACCGCATACGTCCTACGGGAGAAAGTGGGGGATCTTCGGACCTC
ACGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGAGGTAAAGGCTCAC
CAAGGCGACGATCCGTAACTGGTCTGAGAGGATGATCAGTCACACTGGAACTG
AGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGG
GCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTA
AAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTTTGACG
TTACCAACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAA

Pstu-5-T7

CTCCGGCCGCCTGGCCGCGGGATTTCAGATTGAACGCTGGCGGCAGGCCTAA
CACATGCAAGTCGAGCGGATGAGTGGAGCTTGCTCCATGATTCAGCGGCGGAC
GGGTGAGTAATGCCTAGGAATCTGCCTGGTAATGGGGGACAACGTTTCGAAAG
GAACGCTAATACCGCATACGTCCTAAGGGAGAAAGTGGGGGATCTTCGGACCT
CACGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGAGGTAAAGGCTCA
CCAAGGCGACGATCCGTAACTGGTCTGAGAGGATGATCAGTCACACTGGAACT
GAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATG
GGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGT
AAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTTTGAC
GTTACCAACAGAATA

Pstu-7-T7

CTCCGGCCGCATGGCCGCGGGATTTCAGATTGAACGCTGGCGGCAGGCCTAAC
ACATGCAAGTCGAGCGGATGAGTGGAGCTTGCTCCATGATTCAGCGGCGGACG
GGTGAGTAATGCCTAGGAATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGG
AACGCTAATACCGCATACGTCCTACGGGAGAAAGTGGGGGATCTTCGGACCTC
ACGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGAGGTAAAGGCTCAC
CAAGGCGACGATCCGTAACTGGTCTGAGAGGATGATCAGTCACACTGGAACTG
AGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGG
GCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTA
AAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTTTGACG
TTACCAACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGA
AGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTT

Pstu-7-T7

CTCCGGCCGCCTGGCCGCGGGATTTCAGATTGAACGCTGGCGGCAGGCCTAA
CACATGCAAGTCGAGCGGATGAGTGGAGCTTGCTCCATGATTCAGCGGCGGAC
GGGTGAGTAATGCCTAGGAATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAG
GAACGCTAATACCGCATACGTCCTACGGGAGAAAGTGGGGGATCTTCGGACCT
CACGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGAGGTAAAGGCTCA
CCAAGGCGACGATCCGTAACTGGTCTGAGAGGATGATCAGTCACACTGGAACT
GAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATG

GGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGT
AAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTTTGAC
GTTACCAACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATAC
GAAGGGTGCAAGCGTTAATCG