Molecular biological and biochemical approaches to expand the spectrum of fungal natural products

Molekularbiologische und biochemische Ansätze zur Erweiterung des Spektrums der pilzlichen Naturprodukte

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Pharmazie

der Philipps-Universität Marburg

vorgelegt von

M. Sc. Florian Kindinger

aus Bensheim

Marburg an der Lahn, 2019

Erstgutachter: **Prof. Dr. Shu-Ming Li**

Zweitgutachter: **Prof. Dr. Peter Kolb**

Eingereicht am 19.09.2019

Tag der mündlichen Prüfung: 31.10.2019

Hochschulkennziffer: 1180

Dedicated to

my family

Table of content

List of publications

Florian Kindinger, Jonas Nies, Anke Becker, Tianjiao Zhu and Shu-Ming Li (2019) Genomic locus of a *Penicillium crustosum* pigment as integration site for secondary metabolite gene expression. ACS Chemical Biology. ACS Chem. Biol. 14, 1227-1234

Jie Fan*, Ge Liao*, **Florian Kindinger**, Lena Ludwig-Radtke, Wen-Bing Yin and Shu-Ming Li (2019) Peniphenone and penilactone formation in *Penicillium crustosum* via 1,4-Michael additions of ortho-quinone methide from hydroxyclavatol to γ-butyrolactones from crustosic acid. Journal of the American Chemical Society. 141, 4225–4229

Viola Wohlgemuth*, **Florian Kindinger*** and Shu-Ming Li (2018) Convenient synthetic approach for tri- and tetraprenylated cyclodipeptides by consecutive enzymatic prenylations. Applied Microbiology and Biotechnology 102: 2671-2681

Viola Wohlgemuth, **Florian Kindinger**, Xiulan Xie, Bin-Gui Wang, and Shu-Ming Li (2017) Two Prenyltransferases govern a consecutive prenylation cascade in the biosynthesis of echinulin and neoechinulin. Organic Letters. 19: 5928-5931

Jakub Gunera*, **Florian Kindinger***, Shu-Ming Li and Peter Kolb (2017) PrenDB: A substrate prediction database to enable biocatalytic use of prenyltransferases. Journal of Biological Chemistry 292: 4003-4021

*The authors contributed equally to the works

Share of author contributions

Publication	Authors	Estimated equity ratio [%]
Genomic locus of a *Penicillium crustosum* pigment as integration site for secondary metabolite gene expression, (2019), ACS Chem. Biol. 14, 1227-1234.	**Kindinger, F.**; Nies, J.; Becker, A.; Zhu, T.; Li, S.-M.	65
Convenient synthetic approach for tri- and tetraprenylated cyclodipeptides by consecutive enzymatic prenylations, (2018), Appl. Microbiol. Biotechnol. 102, 2671-2681.	Wohlgemuth, V.*; **Kindinger, F.***; Li, S.-M.	40
PrenDB: A substrate prediction database to enable biocatalytic use of prenyltransferases, (2017), J. Biol. Chem. 292, 4003-4021. (published)	Gunera, J.*; **Kindinger, F.***; Li, S.-M.; Kolb, P.	35
Peniphenone and penilactone formation in *Penicillium crustosum* via 1,4-Michael additions of ortho-quinone methide from hydroxyclavatol to γ-butyrolactones from crustosic acid, (2019), J. Am. Chem. Soc. 141, 4225-4229.	Fan, J.; Liao, G.; **Kindinger, F.**; Ludwig-Radtke, L.; Yin, W.-B.; Li, S.-M.	5
Two Prenyltransferases govern a consecutive prenylation cascade in the biosynthesis of echinulin and neoechinulin, (2017), Org. Lett. 19, 5928-5931.	Wohlgemuth, V.; **Kindinger, F.**; Xie, X.; Wang, B.-G.; Li, S.-M.	5

*The authors contributed equally to the works

Signature candidate

Signature supervisor

Abbreviations

The international system of units and units derived thereof have been used.

[M+H]$^+$	molecular ion plus hydrogen
[M-H]$^-$	molecular ion minus hydrogen
× g	gravitational acceleration
5-FOA	5-fluoroorotic acid
6-MSA	6-methylsalicyclic acid
A domain	adenylation domain
A. cristatus	*Aspergillus cristatus*
A. fumigatus	*Aspergillus fumigatus*
A. glaucus	*Aspergillus glaucus*
A. nidulans	*Aspergillus nidulans*
A. niger	*Aspergillus niger*
A. oryzae	*Aspergillus oryzae*
A. ruber	*Aspergillus ruber*
aa	amino acid
ACP domain	acylcarrier protein domain
AT domain	acyl transferase domain
ATP	adenosine triphosphate
BGC	biosynthetic gene cluster
br	broad (NMR signal)
C domain	condensation domain
C. purpurea	*Claviceps purpurea*
CD$_3$OD	deuterated methanol
CDB	czapek dox broth
CDCl$_3$	deuterated chloroform
cDNA	copy deoxyribonucleic acid
CLF	chain length factor
CoA	coenzyme A
COSY	correlation spectroscopy
cyclo-Trp-Ala	*cyclo*-tryptophanyl-alaninyl
cyclo-Trp-Pro	*cyclo*-tryptophanyl-prolinyl
d	doublet
Da	dalton
dd	double doublet
ddd	double double doublet
DEBS	6-deoxyerythronolide B synthase
DH domain	dehydratase domain
DHN	dihydroxynaphthalene
DKP	diketopiperazine
DMA	dimethylallyl
DMAPP	dimethylallyl diphosphate
DMATS	dimethylallyltryptophan synthase
DMOA	3,5-dimehtylorsellinic acid
DMSO	dimethyl sulfide
DMSO-d$_6$	deuterated dimethyl sulfoxid
DNA	deoxyribonucleic acid
dq	double quartet
dt	double triplet
E. coli.	*Escherichia coli*
e.g.	*exempli gratia*

ECFP	extended connectivity fingerprint
EIC	extracted ion chromatogram
ER domain	enoylreductase domain
ESI	electrospray ionization
EtOAc	ethyl acetate
FPP	farnesyl diphosphate
GA	general prenyl moiety acceptor
GB	general base
GD	general prenyl moiety donor
GGPP	geranylgeranyl diphosphate
GMM	glucose minimal medium
GP	general prenylation product
GPP	geranyl diphosphate
GTP	guanosine triphosphate
His$_6$	hexahistidine
HMBC	heteronuclear multiple bond correlation
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HR-PKS	highly reducing-polyketide synthase
HSQC	heteronuclear single quantum coherence
Hz	hertz
i.e.	*id est*
IPP	isopentenyl pyrophosphate
J	coupling constant
JGI	Joint Genome Institute
kBp	kilo base pairs
k_{cat}	turnover number
kDa	kilodaltons
K_M	Michaelis-Menten constant
KR domain	ketoreductase domain
KS domain	ketoacyltransferase domain
LB	lysogeny broth
LC-MS	liquid chromatography–mass spectrometry
m	multiplet
m/z	mass-to-charge ratio
mAU	milliabsorbance unit
Mb	mega base pairs
MCAT	malonyl-CoA:ACP transferase
MEP	methylerythritol phosphate
MHz	mega hertz
mRNA	messenger ribonucleic acid
MT domain	methyl transferase domain
multi	multiplicity
MVA	mevalonate
n × C5	number of C5 units, minimum 1
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
NMR	nuclear magnetic resonance
NP	natural product
NR-PKS	non-reducing-polyketide synthase
NRPS	nonribosomal peptide synthetase
OSMAC	one strain many compounds
P. brevicompactum	*Penicillium brevicompactum*

P. crustosum	*Penicillium crustosum*
P. rubens	*Penicillium rubens*
PCP	peptidyl carrier protein domain
PCR	polymerase chain reaction
PD	potato dextrose
PDB	potato dextrose broth
PEG	polyethylene glycol
PFAM	protein family (database)
PGB	protonated general base
PKS	polyketide synthase
PKS-NRPS	polyketide synthase-nonribosomal peptide synthetase
PPi	inorganic pyrophosphate
ppm	parts per million
PR-PKS	partially reducing-polyketide synthase
PT	prenyltransferase
PT domain	product template domain
q	quartet
R domain	release domain
RBS	ribosome binding site
RNA	ribonucleic acid
rpm	revolutions per minute
s	singlet
S. cerevisiae	*Saccharomyces cerevisiae*
SAT domain	starter unit acyl transferase domain
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SM	secondary metabolite
SMARTS	SMILES arbitrary target specification
SMILES	simplified molecular input line entry system
SMIRKS	a hybrid language of SMILES and SMARTS
t	triplet
T domain	thiolation domain
TB	terrific broth
td	triple doublet
TE domain	thioesterase domain
Tris	2-amino-2-(hydroxymethyl)-propan-1,3-diol
tsept	triple septet
UTR	untranslated region
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume
WT	wild type
δ_H	chemical shift of ^{1}H

Summary

At least 3.5 billion years ago, the first life on earth arose. This was the starting point of the evolutionary development of numerous living beings. According to current estimations, there are 10^{12} different species on our planet. Most of this enormous biodiversity originates from the kingdom of bacteria and archaea. Based on these estimations, only 0.001 % of all species are known to this day. The omnipresent competition between living beings led to the development of secondary metabolism. The metabolites derived from this metabolism are not essential for survival, yet their production offers the organism various selection advantages. Plants, bacteria, and fungi are the main producers of secondary metabolites. The more than 2,140,000 million known secondary metabolites can be divided into five large groups: (1) non-ribosomal polypeptides, (2) polyketides, (3) alkaloids, (4) terpenoids and steroids and (5) enzyme cofactors. Many of these natural compounds show a biological or pharmaceutical activity and were used for the development of drugs. The large number of not yet identified microorganisms harbors an enormous, mostly unused genetic potential to produce further new natural compounds. Such compounds may be suitable for the development of urgently needed new drugs. Various approaches, such as heterologous expression in suitable host organisms, are being investigated to make this potential accessible. Additionally, through synthetic biology approaches, the diversity of natural substances can be further extended, and new natural substances can be discovered or produced.

In the context of research on secondary metabolites, this work focuses on three main topics: 1. The extension of the spectrum of possible substrates for prenyltransferases, by using a database to predict new substrates. 2. The identification and characterization of previously unknown biosynthetic gene clusters, as well as the investigation of a possible application of the enzymes involved to produce new natural substances. 3. The generation of a host for the heterologous expression of secondary metabolite genes and investigation of their unknown products.

Prenyltransferases catalyze the transfer of prenyl units (n × C5) to their target substrates. This is of importance, as an increase in the biological activity of prenylated compounds compared to their unprenylated counterparts has been observed for many compounds. A special property of prenyltransferases is their promiscuity with respect to the substrates. This makes them suitable candidates to produce pharmaceutically active substances. However, in practice, it is difficult to identify new substrates for prenyltransferases. In order to address this problem, a database, PrenDB, was developed for the prediction of such substrates. The predictive power of this database was experimentally tested with 38 predicted substrates by their acceptance with the prenyltransferases FtmPT1, FgaPT2, and CdpNPT. For 27 of the 38 substrates, prenylation by at least one of the three tested enzymes was observed, 17 with conversion yields of more than 50 %. This proved the predictive power of the developed database and enabled the targeted selection of new potential substrates and the identification of new substrate classes.

The identification of biosynthetic gene clusters and the subsequent biochemical characterization of the enzymes involved in the biosynthetic pathways form the basis for synthetic biology approaches to produce natural products. Based on the cyclic dipeptide echinulin, a possible procedure for the identification of the responsible gene cluster and the use of the involved enzymes for the biosynthesis of new substances was described. The enzymatic prerequisites for the biosynthesis of echinulin were determined based on the structural peculiarities of echinulin. Potential candidate gene clusters must encode one non-ribosomal peptide synthetase and several prenyltransferases. In the genome of the echinulin producer *Aspergillus ruber,* a gene cluster with these prerequisites was identified. Enzyme assays with the echinulin precursor *cyclo*-L-tryptophanyl-L-alaninyl and the heterologously produced prenyltransferases EchPT1 and EchPT2 led to a well-founded biosynthetic hypothesis and confirmed the involvement of this cluster in the biosynthesis of echinulin. The combination of EchPT1 and EchPT2 with *cyclo*-L-tryptophanyl-L-alaninyl as a substrate led to the formation of 7 products with different degrees of prenylation. This special property was subsequently used to prenylate further cyclic dipeptides. The stereoisomers of *cyclo*-tryptophanyl-alaninyl and *cyclo*-tryptophanyl-prolinyl were used for this purpose. Analogous to the biosynthesis of echinulin, this led to the formation of tri-prenylated main products prenylated at position C2, C5 and C7, as well as further di-, tri- and tetra-prenylated side products.

Another possibility to investigate and produce secondary metabolites is the heterologous expression in a suitable host. A potential new host for heterologous expression, *Penicillium crustosum*, was examined in this thesis. The genome of the fungus was sequenced and the involvement of the polyketide synthase Pcr4401 in the biosynthesis of the melanin precursor YWA1 was confirmed by deletion and expression experiments. Successful integration of foreign genes in the *pcr4401* gene locus can easily be recognized by the occurrence of an albino phenotype. For better use as an expression host, a *pyrG* deficient strain and two plasmids were generated to integrate foreign genes into the *pcr4401* gene locus. The applicability as an expression host was subsequently verified by the successful expression of three PKS genes and the structural elucidation of the formed products.

Zusammenfassung

Vor mindestens 3,5 Milliarden Jahren entstand das erste Leben auf der Erde. Dies war der Startpunkt der evolutionären Entwicklung zahlreicher Lebewesen. Nach aktuellen Schätzungen existieren 10^{12} verschiedene Arten auf unserem Planeten. Der überwiegende Anteil dieser enormen Artenvielfalt stammt aus dem Reich der Bakterien und Archaeen. Basierend auf dieser Schätzung sind bis zum heutigen Tag gerade einmal 0,001 % aller Arten bekannt. Die allgegenwärtige Konkurrenz zwischen den Lebewesen führte unter anderem zu der Entwicklung des Sekundärstoffwechsels. Die aus diesem Stoffwechsel stammenden Sekundärmetabolite sind nicht essenziell für das Überleben, dennoch bietet deren Produktion dem Organismus zahlreiche Selektionsvorteile. Pflanzen, Bakterien und Pilze gehören zu den Hauptproduzenten von Sekundärmetaboliten. Die über 2.140.000 Millionen bekannten Sekundärmetabolite lassen sich in fünf große Gruppen einteilen: (1) nicht-ribosomale Polypeptide, (2) Polyketide, (3) Alkaloide, (4) Terpenoide und Steroide und (5) Enzymkofaktoren. Viele dieser Naturstoffe weißen eine biologische bzw. pharmazeutische Aktivität auf und dienten zur Entwicklung von Arzneimitteln. Die große Anzahl bisher nicht identifizierter Mikroorganismen beherbergt ein gewaltiges ungenutztes genetisches Potential zur Produktion weiterer neuer Naturstoffe. Solche Stoffe können sich zur Entwicklung dringend gebrauchter neuer Arzneimittel eignen. Verschiedene Ansätze, wie z.B. die heterologe Expression in geeigneten Wirtsorganismen werden verfolgt, um dieses Potential zugänglich zu machen. Des Weiteren können durch Ansätze der synthetischen Biologie die Diversität der Naturstoffe erweitert werden und so neue Naturstoffe entdeckt bzw. produziert werden.

Im Rahmen der Forschung an Sekundärmetaboliten beschäftigt sich diese Arbeit mit drei thematischen Schwerpunkten: 1. Die Erweiterung des Spektrums möglicher Substrate für Prenyltransferasen mittels einer Datenbank zur Vorhersage neuer Substrate. 2. Die Identifizierung und Charakterisierung bisher unbekannter Biosynthesegencluster, sowie die Untersuchung einer möglichen Anwendung der beteiligten Enzyme für die Produktion neuer Naturstoffe. 3. Die Generierung eines Stamms für die heterologe Expression von Sekundärmetabolitgenen und die Untersuchung ihrer unbekannten Produkte.

Prenyltransferasen katalysieren den Transfer von Prenyleinheiten (n × C5) auf ihre Zielsubstrate. Dies ist von besonderer Bedeutung, da bei zahlreichen Verbindungen eine Zunahme der biologischen Aktivität von prenylierten Verbindungen im Vergleich zu ihren nicht prenylierten Vorstufen beobachtet wurde. Eine besondere Eigenschaft von Prenyltransferasen ist ihre Promiskuität bezüglich der von ihnen akzeptierten Substrate. Dies macht sie zu geeigneten Kandidaten für die Produktion pharmazeutisch wirksamer Substanzen. Jedoch gestaltet es sich in der Praxis schwierig neue passende Substrate für Prenyltransferasen zu identifizieren. Um dieses Problem anzugehen, wurde im Rahmen dieser Studie eine Datenbank, PrenDB, für die Vorhersage solcher Substrate entwickelt. Die Vorhersagekraft dieser Datenbank wurde anschließend mit 38 vorhergesagten Substraten und deren Akzeptanz durch die Prenyltransferasen FtmPT1, FgapT2 und CdpNPT experimentell überprüft. Bei 27

der 38 Substrate konnte eine Prenylierung von mindestens einem der drei getesteten Enzymen beobachtet werden. Des Weiteren zeigten 17 Substrate mit mindestens einer Prenyltransferase Umsatzrate von mehr als 50 %. Dies belegt die Vorhersagekraft der entwickelten Datenbank und ermöglicht so die zielgerichtete Auswahl neuer potenzieller Substrate und Identifizierung neuer Substratklassen.

Die Identifizierung von Biosynthesegenclustern und die anschließende biochemische Charakterisierung der an der Biosynthese beteiligten Enzyme bildet die Grundlage für die Ansätze der synthetischen Biologie zur Produktion von Naturstoffen. Anhand von dem cyclischen Dipeptid Echinulin wurde in dieser Arbeit ein mögliches Vorgehen zur Identifizierung des verantwortlichen Genclusters und die Verwendung der beteiligten Enzyme zur Biosynthese neuer Substanzen beschrieben. Basierend auf den strukturellen Eigenschaften von Echinulin wurde die enzymatischen Voraussetzungen für dessen Biosynthese ermittelt. Ein Gencluster dafür muss demnach Gene für eine nichtribosomalen Peptidsynthetase und mehrere Prenyltransferasen enthalten. Im Genom des Echinulin-Produzenten *Aspergillus ruber* konnte ein Gencluster mit diesen Vorrausetzungen identifiziert werden. Enzymtests mit der Echinulin-Vorstufe *cyclo*-L-Tryptophanyl-L-Alaninyl und den heterolog produzierten Prenyltransferasen EchPT1 und EchPT2 führten zu einer fundierte Biosynthesehypothese und bestätigten die Beteiligung des Genclusters an der Biosynthese von Echinulin. Das Zusammenwirken von EchPT1 und EchPT2 mit *cyclo*-L-Tryptophanyl-L-Alaninyl als Substrat führte zu Bildung von insgesamt 7 Produkten mit unterschiedlicher Anzahl an Prenylierungen. Diese besondere Eigenschaft wurde anschließend genutzt um weitere zyklische Dipeptide mehrfach zu prenylieren. Hierfür wurden die Stereoisomere von *cyclo*-Tryptophanyl-Alaninyl und *cyclo*-Tryptophanyl-Prolinyl verwendet. Analog zur Biosynthese von Echinulin führte dies zur Bildung von an Position C2, C5 und C7 dreifach-prenylierten Hauptprodukten, sowie weiteren zweifach-, dreifach- und vierfach-prenylierten Nebenprodukten.

Eine weitere Möglichkeit Sekundärmetabolite zu produzieren und zu untersuchen ist die heterologe Expression in einem geeigneten Stamm. Ein potenzieller neuer Stamm für die heterologe Expression, *Penicillium crustosum*, wurde in dieser Arbeit untersucht. Dafür wurde das Genom des Pilzes sequenziert und durch Deletions- und Expressionsexperimente die Beteiligung der Polyketidsynthase Pcr4401 an der Biosynthese der Melaninvorstufe YWA1 bestätigt. Die erfolgreiche Integration von fremden Genen in den *pcr4401* Genlokus lässt sich durch das Auftreten eines Albinophänotyps leicht erkennen. Zur besseren Verwendung als Expressionsstamm wurde ein *pyrG* defizienter Mutant und zwei Plasmide zur Integration fremder Gene in den *pcr4401* Genelokus generiert. Die Eignung als Expressionswirt wurde anschließend durch die erfolgreiche Expression von drei PKS Genen und die Strukturaufklärung der gebildeten Produkte verifiziert.

1. Introduction

1.1. Ascomycota

The phylum of Ascomycota is the most abundant phylum in the fungal kingdom with more than 64.000 known species.[1] From all yet described fungal species, 65 % belong to the phylum of Ascomycota.[2] This phylum can be subdivided into three major subphyla, the Taphrinomycotina, the Saccharomycotina, and the Pezizomycotina.[3] Ascomycota can occur as unicellular organisms and reproduce themselves by budding *e.g. Saccharomyces cerevisiae* (Saccharomycotina) or by fission *e.g. Schizosaccharomyces pombe* (Taphrinomycotina). In contrast, most members of the subphylum Pezizomycotina *e.g. Aspergillus* or *Penicillium* species grow as filamentous fungi, forming a dense network of hyphae, the so-called mycelium. The filamentous fungi reproduce themselves either asexually by the formation of conidia or sexually by the formation of ascospores. In their asexual reproduction form, called anamorph, the conidia are formed at specially developed hyphae known as conidiophores. The sexual reproduction form, called teleomorph, produce fruiting bodies for ascospores production. At this fruit bodies, also known as ascocarp, the ascospores are produced in specialized tube- or sac-like cells termed asci.[4]

The evolutionary divergence of the Pezizomycotina started over 450 million years ago.[5] Within million years of diversification and adaption, Ascomycota conquered almost all possible terrestrial and aquatic habitats. Furthermore, they developed a wide variety of lifestyles, ranging from saprotrophic or necrotrophic to biotrophic, parasitic or symbiotic behavior.[4] Fungi can be a curse but also a blessing for humankind. Fungal pathogens affect humans directly ranging from simple dermatophytosis caused by dermatophytes to life-threatening aspergillosis caused by *Aspergillus fumigatus*.[6] Every year more than one billion people suffer from a fungus-related disease, with more than 1.5 million deaths.[7] Besides these direct effects on humankind, molds and fungal plant-pathogens affect human health and wealth indirectly. Since mankind became agriculturally active, it encountered fungal plant-pathogens *e.g. Blumeria graminis* responsible for the powdery mildew disease in various types of grains. It is estimated that fungi are responsible for the destruction of approximately one-third of all food crops each year.[8] Moreover, the production of mycotoxins can lead to food spoilage and human poisoning. Mycotoxins are compounds with low molecular weight, typically less than 1,000 Da, which are particular products of the secondary metabolism from filamentous fungi. So far, more than 400 mycotoxins are known, whose main producers belong to the genus of *Aspergillus*, *Penicillium* and *Fusarium*.[9] Despite their great structural diversity, the most important food-related mycotoxins belong to one of the following six subgroups: Aflatoxins, ochratoxins, fumonisins, zearalenone, trichothecenes, and ergot alkaloids.[10] All these mycotoxins can be found in daily consumed foodstuff, including grains, maize, nuts and fruits. Aflatoxins, especially aflatoxin B_1 (**1**) mainly produced by *Aspergillus flavus* and *Aspergillus parasiticus*, are highly hepatotoxic and cancerogenic. *Aspergillus*

ochraceus, *Aspergillus carbonarius,* and *Penicillium verrucosum* are the main producers of the nephrotoxic and potential cancerogenic ochratoxins, *e.g.* ochratoxin A (**2**). Fumonisins, *e.g.* fumonisin B$_1$ (**3**) are produced by members of the genus *Fusarium e.g. Fusarium verticillioides* and are cancer promoting mycotoxins. Also, the mycotoxin zearalenone (**4**) is produced by *Fusarium* species, in particular by *Fusarium graminearum*. The acute toxicity is considered low, but it shows potent estrogenic effects on mammalian livestock and humans. The subgroup of trichothecenes are sesquiterpenes, including more than 100 compounds sharing a common 12,13-epoxytrichothec-9-ene structure, *e.g.* deoxynivalenol (**5**). Due to the huge variety of the members, the effects on human health are diverse as well, ranging from immunosuppressive effects to gastrointestinal discomfort. Main producers are *Fusarium* species *e.g. F. graminearum*. *Claviceps purpurea* is the main producer of the ergot alkaloids, *e.g.* ergotamine (**6**). Intoxication with ergot alkaloids can result in ergotism, which is characterized by symptoms like diarrhea, muscle spasms, psychosis, vomiting and vasoconstriction.[11,12] In contrast to these harmful mycotoxins, there are also secondary fungal metabolites with potential benefits for human health. For example, ergometrine (**7**) an ergot alkaloid produced by *Claviceps purpurea* is used as a medication for the treatment of bleeding after childbirth because of its uterine contracting effects.[13] Chemical or chemoenzymatic modification of natural secondary metabolites is used to convert these metabolites to suitable drugs or to alter the molecular parameters *e.g.* target specificity, bioavailability, or stability for more convincing pharmaceutical use.[14] Besides *Streptomyces* species, fungi from the Ascomycota are a valuable source for identification and/or production of antibiotics. The best-known representative is penicillin G (**8**) discovered by Alexander Fleming in 1928, which is produced by *Penicillium rubens* (formerly known as *Penicillium chrysogenum*).[15]
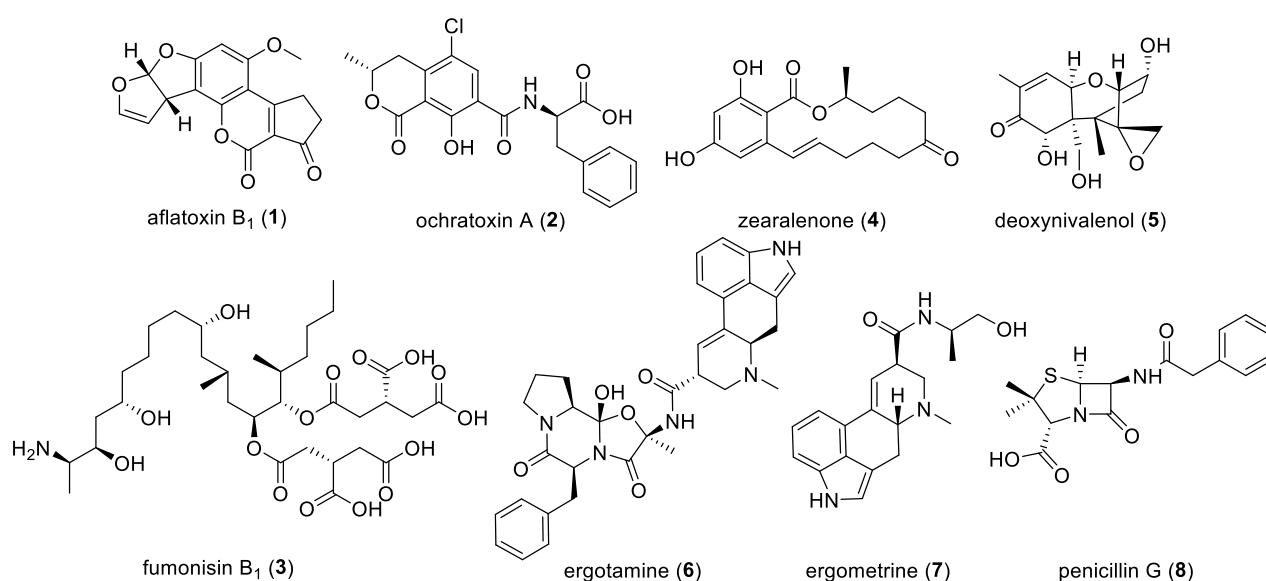

Figure 1: Secondary metabolites produced by fungi of Ascomycota.

Another important utilization for ascomycetous fungi is the biotechnological application within the food and additive (*e.g.* enzymes) production. *Penicillium roqueforti* and *Penicillium camemberti* are used for a long time in the cheese production and *Saccharomyces cerevisiae* in beer, wine, and bread production.[16] An established fungal strain for the production of soy sauce and sake is *Aspergillus oryzae*. Additional to the use in food technology, *Aspergillus oryzae* is also established in biotechnological applications *e.g.* the production of enzymes like α-amylases.[17,18] Microbial enzymes are commonly used in textile, starch processing, food, detergent, pulp, and paper industries. Roughly half of these commercially used enzymes originate from fungi. Industrially interesting enzymes often belong to amylases, amyloglucosidases, cellulases, glucose oxidases, lipases, proteases and xylanases.[19] Additional to the wide biotechnological applications, the scientific interest in Ascomycota is steadily increasing. *Saccharomyces cerevisiae, Aspergillus nidulans* and *Penicillium chrysogenum* are well-known model organisms concerning cellular, genetic, and regulatory studies. Ongoing research aims to further increase their biotechnological applicability as cell factories for the production of enzymes or natural products.[20-22]

1.2. Secondary metabolism and biosynthetic gene cluster

Metabolism can be subdivided into primary and secondary metabolism. The primary metabolism is defined as the whole set of biochemical reactions necessary for growth, development, reproduction, energy transformation, and to maintain the vitality of living organisms or cells. Catabolism and anabolism are the two major aspects of primary metabolism. Catabolism includes all reactions to break down molecules for energy generation or to biosynthesize precursors for anabolic reactions. In contrast, anabolic processes are responsible for building up macromolecules like proteins, nucleic acids, carbohydrates, and lipids.[23] Besides these essential life-sustaining processes, especially microorganisms like bacteria and fungi developed an enormous repertoire of biosynthetic pathways for the production of a diverse set of nonessential molecules in the so-called secondary metabolism. The building blocks used to produce secondary metabolites originate mainly from the central carbon metabolism with acetyl-CoA and amino acids as the most relevant precursors (**Figure 2**).[24] During the evolutionary adaptation to different ecological niches, microorganisms developed the ability to produce a variety of secondary metabolites. The biosynthesis of these substances led to diverse selection advantages based on properties such as hindering of competitor´s growth, protection against other microorganisms, acquisition of nutrition factors *e.g.* siderophores and for intra and inter-species communication purposes.[25] Most fungal secondary metabolites belong to one of the four main groups *i.e.* polyketides, non-ribosomal polypeptides, alkaloids and terpenoids and steroids (**Figure 2**).[26] Up to the year 2010 over 500,000 natural products were discovered in total and approximately 30,000 originated from fungi. Between the years 2001 to 2010, around 61 % of the identified bioactive

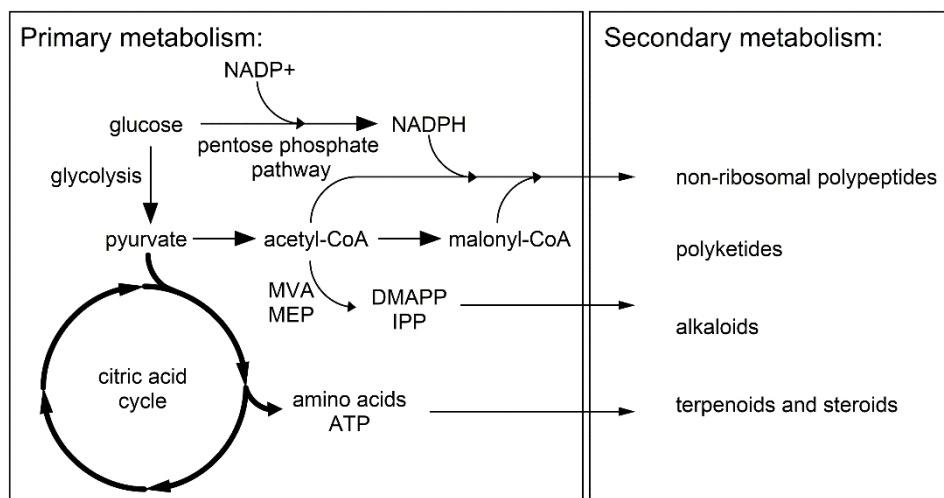secondary metabolites were from fungal origin, underlining their potential for finding and development of new drugs.[27]

Figure 2: Schematic presentation of the connection between primary and secondary metabolism.
MVA: mevalonate pathway; MEP: methylerythritol phosphate pathway; DMAPP: dimethylallyl pyrophosphate; IPP: isopentenyl pyrophosphate

In the year 1976, the first complete genome sequence with a total length of 3,569 nucleotides from the bacteriophage MS2 was elucidated and published.[28] This can be considered as the early starting point of the genomic era. The first genome of the prokaryotic organism *Haemophilus influenzae* was published 19 years later, followed by the first eukaryotic genome from *Saccharomyces cerevisiae* one year later.[29,30] Analysis of the genomic sequence revealed that genes are not randomly distributed in microorganisms. Genes can be organized in close proximity to form a gene cluster. In prokaryotes, gene clusters are coregulated and cotranscribed forming functional metabolic groups, known as operons (**Figure 3A**).[31,32] Around 1990 and in the following years it was proven that also in fungi genes encoding for enzymes of a given biosynthetic pathway are also often organized in close proximity.[26] These biosynthetic gene clusters (BGCs) share a common set-up, usually consisting of genes for a backbone enzyme, such as a non-ribosomal peptide synthetase (NRPS), a polyketide synthase (PKS) or a terpene synthase, and different kinds of tailoring enzymes for modification of the formed metabolic scaffold (**Figure 3B**). Optionally, further genes for regulation of the whole gene cluster or transporters for secretion of the products or uptake of special substrates are included in the BGCs.[24]

The size of a fungal genome is in the range from 9 to 178 Mb with an average size for ascomycetous genomes of 37 Mb.[33] The search and identification of biosynthetic gene clusters in this plethora of genomic data has led to the term genome mining. In particular, the rapid development in the area of next-generation sequencing and the associated increasing amount of available data required computationally high-throughput methods for genome analysis. Nowadays, various user-friendly computational tools for genome analysis and BGC identification exist, *e.g.* antiSMASH 5.0[34], SMURF[35], PRISM3,[36] and others[37]. The fundamental concept of these genome mining tools for gene cluster

detection is based on the common set-up of BGCs (see above), a database with already characterized BGCs and protein families (PFAM) and the presence of conserved domain sequences within the backbone and tailoring genes. In short, for BGCs detection, a comparative BLAST analysis, PFAM analysis, and backbone genes domain analysis is performed. Additionally, the simultaneous comparison with characterized BGCs assists in the detection of putatively unknown BGCs.[38] The above mentioned ongoing development in next-generation sequencing techniques resulted in a tremendous cost and time reduction for sequencing of a genome.[39] As a result, the scientific community today has access to numerous genomic sequences. For example, the database of the "1,000 fungal genomes" project from the Joint Genome Institute's (JGI) consists of 1,322 sequenced genomes (as of July 2019).
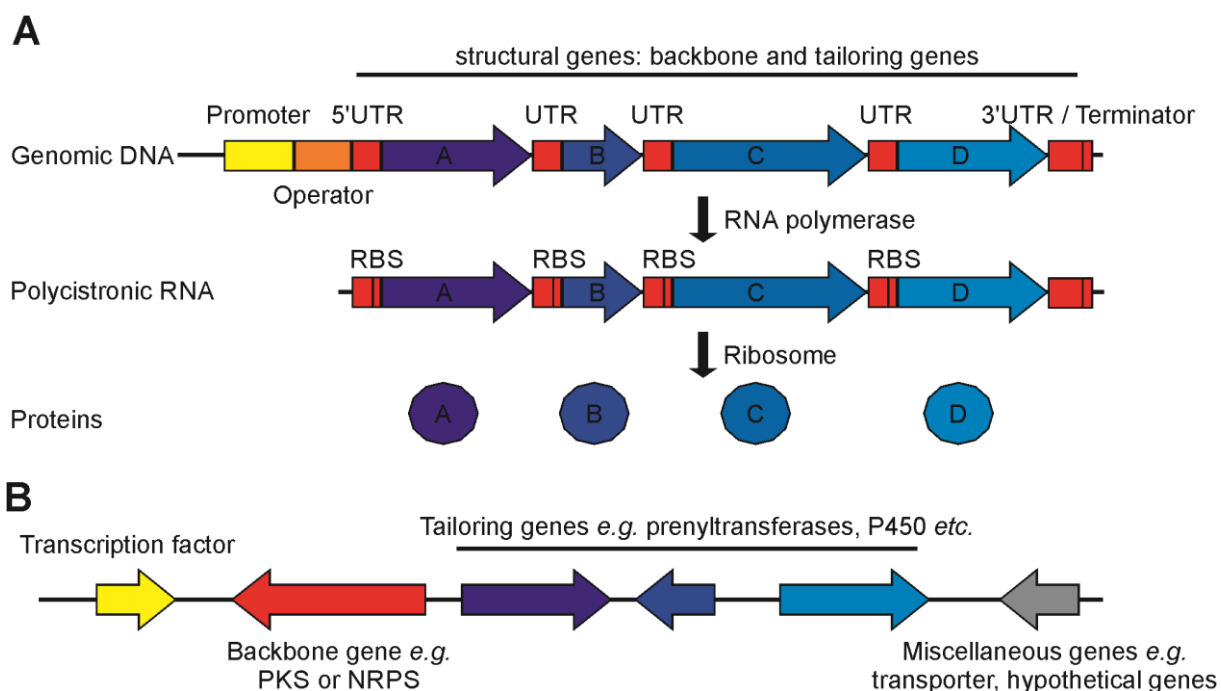
Figure 3: Schematic presentation of typical prokaryotic operon gene cluster (**A**) and eukaryotic biosynthetic gene cluster (**B**). UTR: Untranslated region; RBS: Ribosome binding site.

Microbial genome mining revealed the presence of a plethora of BGCs in these organisms. On average, Actinomycetes and Ascomycetes harbor around 20-50 and 40-80 BGCs per species, respectively. A comparison of the number of BGCs with known secondary metabolites isolated from these species demonstrated that the capability for secondary metabolite production is far from being exhausted. The reason for this is the silence of most of the gene clusters, *i.e.* they are transcriptionally inactive under standard laboratory or naturally observed cultivation conditions.[40-42] To exploit this hidden potential for secondary metabolite production and identification of new natural products, several strategies for activation of these silent gene clusters were developed. The early observation that microorganisms produce various secondary metabolites depending on the cultivation conditions led to the OSMAC (one strain many compounds) strategy. This easy and fast to realize approach is based on the adjustment of media composition, cultivation temperature, pH value, aeration *etc.* without the requirement of the genetic accessibility of the producer.[43,44] Another strategy is to mimic a

competitional microbial environment through co-cultivation of different microorganisms. It is suggested that this strategy is based on signaling molecules and metabolic precursors produced by either one or both co-cultivated species.[45] Both strategies result in a change of the environmental factors that the microorganism experiences. In order to adapt to or react to this changes a complex framework consisting of sensory systems and genetic regulatory circuits and pathways has been developed by the organisms. The model organism *Aspergillus nidulans* is frequently used to study these genetic regulation systems. For this, the heterotrimeric velvet complex consisting of the three proteins VelB, VeA, and LaeA was identified, and its crucial role in secondary metabolite regulation and reproduction was investigated. In the end, the velvet complex acts as a transcriptional activator mediated by the secondary metabolism main regulator LaeA.[46,47] Beside these global transcriptional regulators, approximately 50 % of the BGCs harbor a biosynthetic pathway specific transcription factor, responsible for BGC transcriptionally regulation.[24,48] Epigenetic modification is another important transcriptional regulation mechanism, which is based on the condensation state of the genomic DNA. Chromatin can be divided into heterochromatin consisting of condensed DNA and proteins and euchromatin being much less condensed. Genes and BGCs in less condensed genomic regions are more accessible for transcription factors and the RNA-polymerase and are therefore more often transcriptionally active. The DNA condensation state is regulated *inter alia* by the methylation and acetylation status of the histones. This is adjusted by the action of histone acetyl- and methyltransferases as well as by histone deacetylases and demethylases.[49,50] For example, an acetyl group from acetyl-CoA can be transferred to a lysine residue of the N terminal histone tail. This leads to a reduced positive charge of the histone accompanied by a reduction of the DNA-histone interaction, resulting in a less tight chromatin condensation state and activation of otherwise inactive genes.[51] Therefore, inactivation of histone deacetylases by gene deletion or *via* the use of chemical inhibitors resulted in the increased expression of formerly inactive BGCs and thus to the identification of newly produced secondary metabolites.[52,53]

A prerequisite for the strategies described above is the cultivability of the targeted microorganisms under laboratory conditions. Even so, the often low yields of the newly produced substances, hamper the identification and characterization of these compounds. With progress in DNA manipulation and cloning methods, heterologous expression in appropriate hosts became a promising method to overcome these limitations.[54] The first requirement for an effective heterologous expression of BGCs or synthetic biology approaches is a fast and reliable cloning and manipulating method of large DNA fragments. Therefore, methods for capturing and cloning of whole BGCs were developed.[55] Sometimes, even further modifications steps *e.g.* promoter exchanges are necessary for BGCs activation. This is realized for example, *via* ligation-independent DNA manipulating methods like homologous recombination cloning in *E. coli*.[56] The second requirement is the availability of

appropriate heterologous expression host. The ideal host must meet the following criteria: good genetic tractability, easy and inexpensive cultivability combined with rapid growth and sufficient supply with metabolic precursors with a clear native secondary metabolite background profile. In addition, the host should have a similar DNA splicing and glycosylation pattern compared to the BGC donor organism.[57] Fungal systems for heterologous expression were proven to be a promising choice for secondary metabolite production, isolation and characterization with yields up to 950 mg per liter medium for certain compounds.[54]

1.2.1. Polyketide synthases

The members of the large and structurally diverse polyketide family are mainly produced in bacteria, fungi, and plants by enzymes called polyketide synthases (PKS). Based on their structural composition, they can be divided into three types. Type I PKSs are large multidomain megaenzymes organized in distinct modules. Type II PKSs consist of monofunctional proteins forming multienzyme complexes. Type III PKSs are small homodimeric enzyme complexes without a so-called acylcarrier protein domain. The fundamental chemical aspect in the biosynthesis of polyketides is the Claisen condensation reaction, where the polyketide backbone is formed by the condensation of starter units *e.g.* acetyl-CoA with extender units *e.g.* malonyl-CoA. On the one hand, structural variety is achieved by PKS inherent factors *e.g.* the use of various starter and extender units (**Figure 5B**), the difference in polyketide length, degree of reduction and methylation, as well as by different release and cyclization mechanisms. On the other hand, further diversification can be achieved by polyketide modifying tailoring enzymes for oxidation, reduction, rearrangement, and transfer reactions.[58] Prominent representatives of the clinically used polyketides and their derivatives are lovastatin (**9**), a cholesterol-lowering agent,[59] griseofulvin (**10**) used due to its antimycotic effect,[60] doxorubicin (**11**) as a chemotherapeutic agent in the treatment of cancer,[61] and the macrolide antibiotic erythromycin A (**12**).[62] In addition to the potential within the development of new drugs, PKSs can also be used for production of bulky compounds and biofuel in the chemical industry, as well as pigments in the textile industry.[63,64]

The domain-based architecture is typical for type I PKS. Enzymes of this family consist of multiple domains with distinct functions and can be divided into modular and iterative type I PKS. The biosynthesis of polyketides proceeds in three phases, (1) starter unit loading, (2) chain elongation and reduction, (3) polyketide cyclization and release (**Figure 5A**). All these steps are catalyzed by the concerted action of the different domains. Except for the modular type I PKS, the chain elongation is carried out by iterative cycles. In contrast, each module of a type I modular PKS, consisting of different domains, is used only once during polyketide biosynthesis. Depending on the domain architecture, the type I iterative PKS can be further subdivided into non-reducing- (NR-), partially reducing- (PR-) and highly reducing- (HR-) PKS (**Figure 6**).
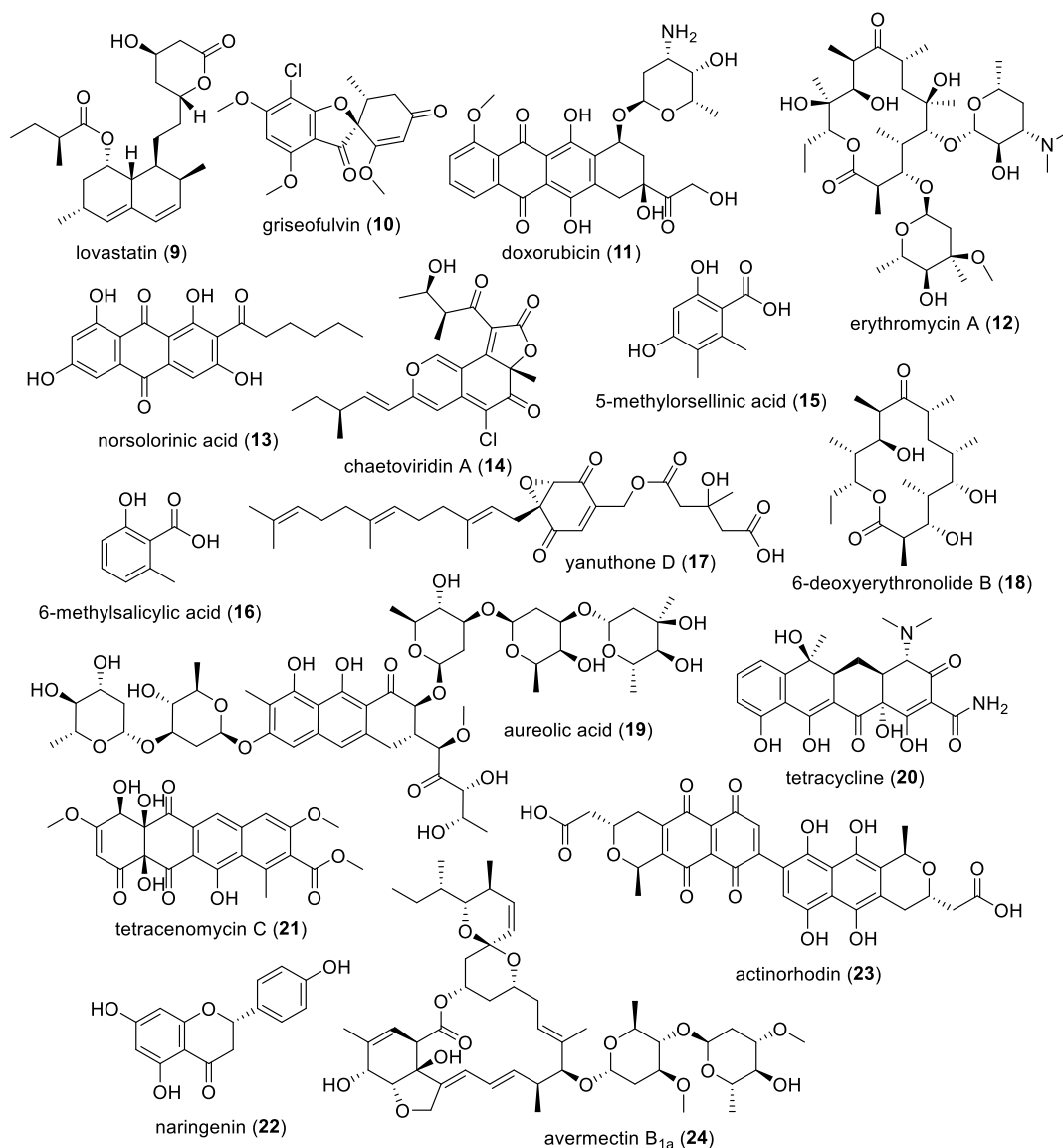
Figure 4: Polyketides and derivatives thereof produced by various microorganisms and plants.

The minimal domain architecture of a NR-PKS consists of the starter unit acyl transferase (SAT), ketoacyltransferase (KS), acyltransferase (AT), product template (PT), acyl carrier protein (ACP), and the thioesterase (TE) or the release (R) domains (**Figure 6**).[65] Biosynthesis of the polyketide begins with the selection and loading of the starter unit to the SAT domain. Beside acetyl-CoA, SAT domains are known to use other compounds as starter units, *e.g.* a C6 fatty acid in the case of norsolorinic acid (**13**) biosynthesis. Moreover, in the biosynthesis of chaetoviridin A (**14**) a highly reduced triketide, produced by another PKS, is used as starter unit.[66-68] The AT domain selects the extender unit and catalyzes its transfer to the ACP domain. In the next step, the KS domain catalyzes the decarboxylative Claisen-like condensation of the starter and extender unit, resulting in an elongated polyketide chain bounded to the ACP domain. The elongation steps are repeated iteratively until the polyketide chain reaches a certain length. Afterward, cyclization occurs at the PT domain eventually followed by the release of the polyketide at the TE or R domain. In addition to the already mentioned domains, some PKS harbor

a methyltransferase (MT) domain, catalyzing the methylation of the elongated polyketide chain as in the case of 5-methylorsellinic acid (**15**) biosynthesis.[69]
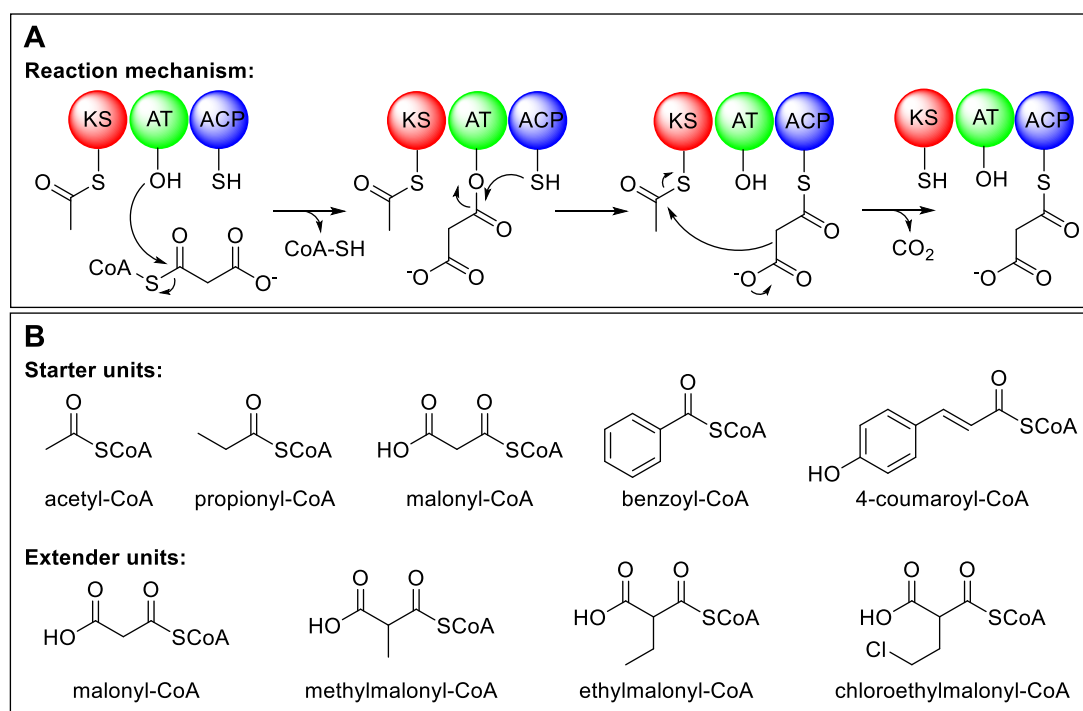
Figure 5: Reaction mechanism of polyketide chain elongation (**A**) and commonly used starter and extender units (**B**).

PR- and HR-PKSs harbor additional domains to reduce the growing polyketide chain. The ketoreductase (KR) domain is NADPH-dependent and catalyzes the reduction of the keto group to a hydroxyl group. Further reduction is achieved by water elimination catalyzed by the dehydratase (DH) domain. Action of the NADPH-dependent enoyl reductase (ER) domain finally leads to a fully saturated extender unit covalently bound to the growing polyketide chain (**Figure 6**).[70] It is worth noting that not necessarily every extender unit becomes fully or partially reduced. A well-studied PR-PKS is the 6-methylsalicylic acid synthase, which is involved in the biosynthesis of the 6-methylsalicylic acid (**16**) related meroterpenoids *e.g.* yanuthone D (**17**).[71,72] Most prominent representative of the HR-PKS is LovB, participating in the biosynthesis of lovastatin (**9**).[73]
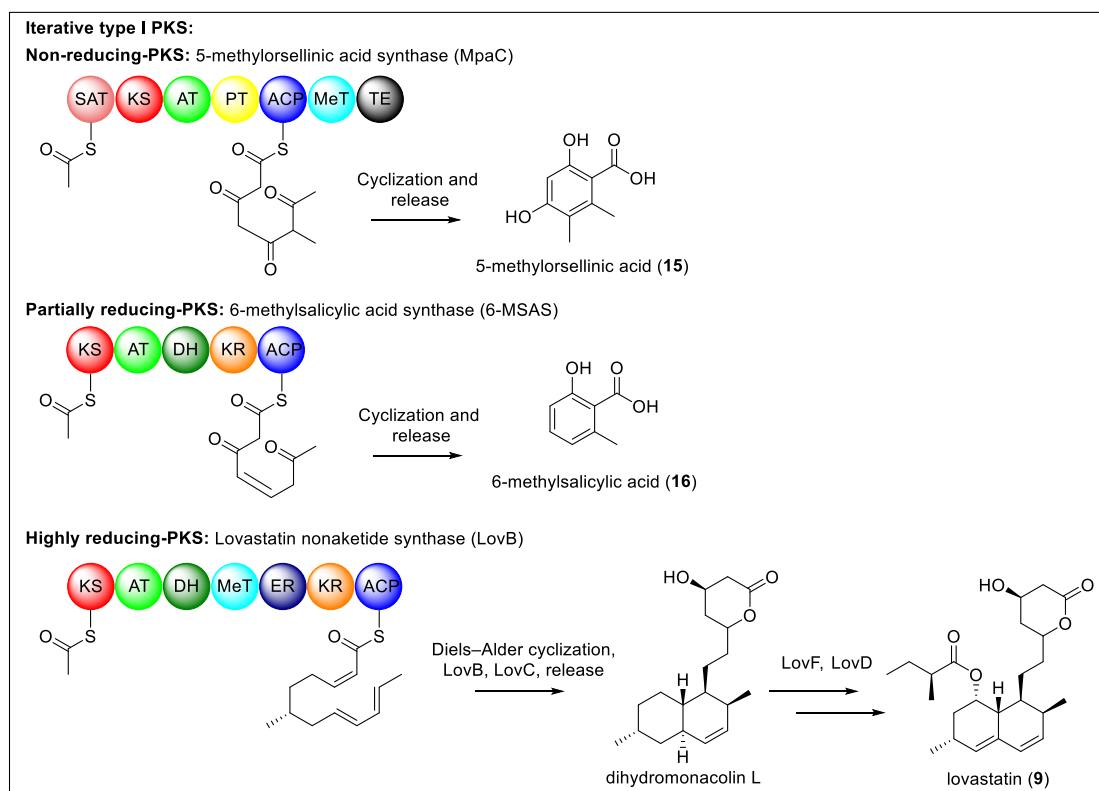
Figure 6: Representatives of the NR-, PR-, and HR-PKS, their domain architecture and their reaction products.

As mentioned above, modular type I PKSs don't work iteratively. Instead, the growing polyketide is handed to each module once where the extender unit is added and modified according to the domain set-up. Since each module is used only once, it is easier to predict the formed polyketide compared to iterative type I PKS. The modular structural architecture results in a much larger enzyme (**Figure 7**).
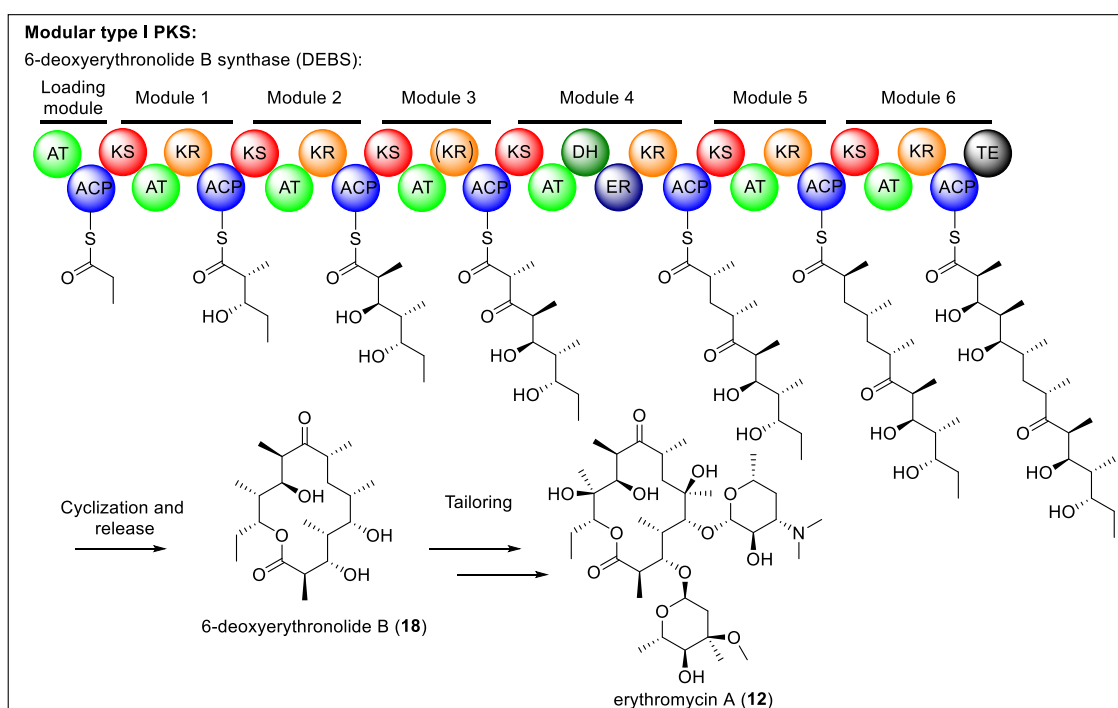
Figure 7: Representative of the modular type I PKS, its domain architecture and the reaction product.

20

The 6-deoxyerythronolide B synthase (DEBS) responsible for building the macrolide scaffold 6-deoxyerythronolide B (**18**) in the biosynthesis of the antibiotic erythromycin (**12**), is one of the best-studied PKS.[74] As with DEBS, modular type I PKSs do not necessarily consist of multiple polypeptides. Contrarily, DEBS comprises loading domain, six extending modules and one TE domain organized in three polypeptides.[75]

Unlike type I PKS, type II PKSs are not large megaenzymes, but consist of monofunctional proteins that form a functional unit (**Figure 8A**). Fungal type II PKSs are not known up to date, therefore type II PKSs are currently limited to the bacterial kingdom, especially to *Streptomyces* species.[76,77] Nevertheless, type II PKSs contribute significantly to polyketide diversity. For example, members of, bacterial anthracyclines, aureolic acid (**19**), tetracycline (**20**) and tetracenomycin C (**21**) are type II PKS based.[76] The minimal set-up of a type II PKS displays the KS domain, catalyzing the Claisen-like condensation of acetyl- and malonyl-CoA units in an iterative fashion, the ACP domain which anchors the polyketide to the enzyme complex and the chain length factor (CLF), which shows very high similarity to the KS and determines the length of the growing polyketide (**Figure 8A**). Additionally, in some cases, the CLF domain can catalyze the decarboxylation of malonyl-CoA to acetyl-CoA to generate the starter unit.[78] Not part of the enzyme complex is the malonyl-CoA:ACP transferase (MCAT) responsible for loading malonyl-CoA to the ACP domain. A well-studied example for type II PKS is TcmKLM involved in catalyzing the formation of the antibiotic tetracenomycin C (**21**).[79]

Type III PKS are small homodimeric complexes with a common size of 80 – 90 kDa. Therefore, they are the smallest types of PKS known up to this point (**Figure 8B**). These homodimers are able to catalyze all necessary reactions for polyketide biosynthesis.[80] Type III PKS were not only firstly identified in plants but are also widely distributed over the plant kingdom and were initially thought to be exclusive to plants.[81] However, type III PKSs were later also identified in bacteria.[82] Fungal type III PKSs are rare, but they are more frequently identified. Currently, five examples are characterized.[83] Type III PKSs catalyze the biosynthesis of small aromatic polyketides. Typically, ring-form and chain-form acyl-CoAs are used as starter units and are extended by the condensation of malonyl-CoA. For cyclization and product release, various types of cyclization reactions are used like Claisen condensation, aldol or lactone formation.[80] The best-studied type III PKS belongs to the family of chalcone synthases catalyzing the formation of naringenin (**22**).[84]
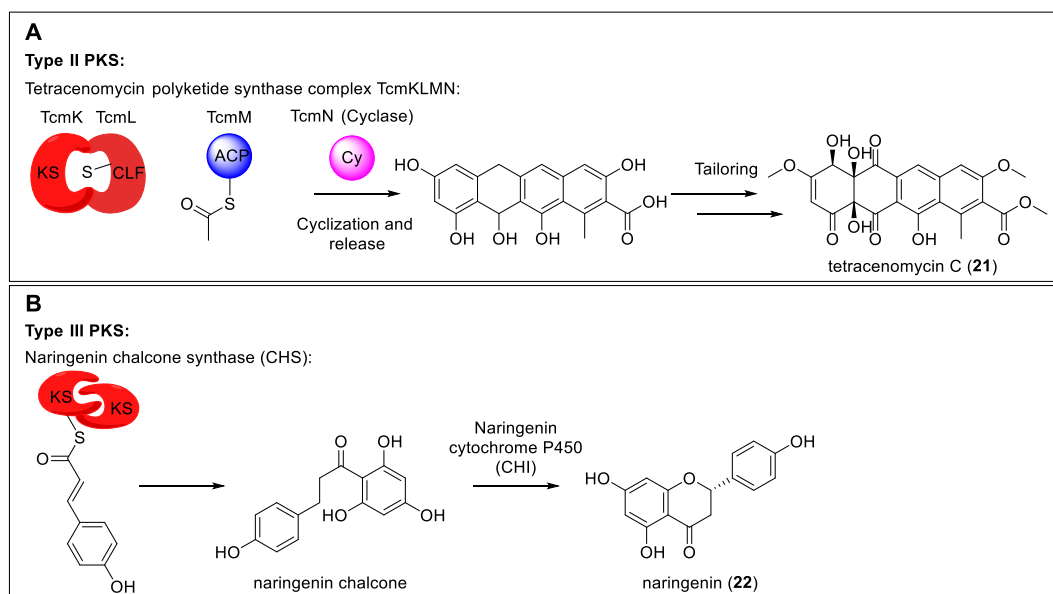
Figure 8: Representatives of type II and III PKS and their reaction products

As described in this chapter, PKS of all types contribute significantly to natural product diversity. Based on the growing knowledge about structural functionality in the fields of starter/extender unit selection, cyclization and release reactions, approaches for PKS engineering emerged to increase polyketide product diversification further. In the context of PKS engineering, modular type I PKS are of particular interest because of their linear order of catalytic domains and the associated predictability of the formed product. Truncation of several modules from the DEBS combined with the fusion to the native TE domain leads to the production of new polyketides. Generation of hybrid-PKS by adding domains from other PKS *e.g.* DH or KR domains or replacement of the AT domain responsible for starter unit selection furthermore increased product diversity.[85-87] Interfering with the cyclization pattern during polyketide biosynthesis is another approach to produce new products. This is exemplified by the combination of the DEBS with the aromatases and cyclases from the bacterial actinorhodin (**23**) gene cluster.[88] Studies on the structural functionality of the AT domain of the modular avermectin B_{1a} (**24**) PKS, which shows high promiscuity towards starter units, revealed deeper insights into starter unit selection. In these studies, the authors were able to modify and alter the accepted substrate space.[89] Although there are already successful PKS engineering approaches, due to the complex interactions between the growing polyketide chain and the domains, module-module interactions, and domain structure-function relationships, many questions and uncertainties remain in order to fully exploit the PKS engineering potential.[90]

1.2.2. Nonribosomal peptide synthetases

Nonribosomal peptides, biosynthesized by nonribosomal peptide synthetase (NRPS), are another rich and diverse family of secondary metabolites. Structural diversity is achieved by using the 20 proteinogenic amino acids or many other nonproteinogenic amino acids as condensation substrate for

building a linear, cyclic or branched-cyclic product.[91] Complexity of cyclic nonribosomal peptides can range from simple cyclic dipeptides with a 2,5-diketopiperazine backbone *e.g.* brevianamide F (**25**) to large cyclic peptides *e.g.* the immunosuppressive drug cyclosporin A (**26**) or to branched-cyclic peptides *e.g.* the antibiotic bacitracin (**27**).[92-94] Linear peptides can vary in length ranging from 3 to 20 amino acids *e.g.* the virostatic and antibiotic agent feglymycin (**28**).[95] As with the PKS, various tailoring enzymes carry out numerous modifications on the nonribosomal peptide scaffold.[96] Main producers of nonribosomal peptides are bacteria, especially species belonging to the phylum Firmicutes, Proteobacteria and Actinobacteria, and filamentous fungi of the phylum Ascomycota.[97]
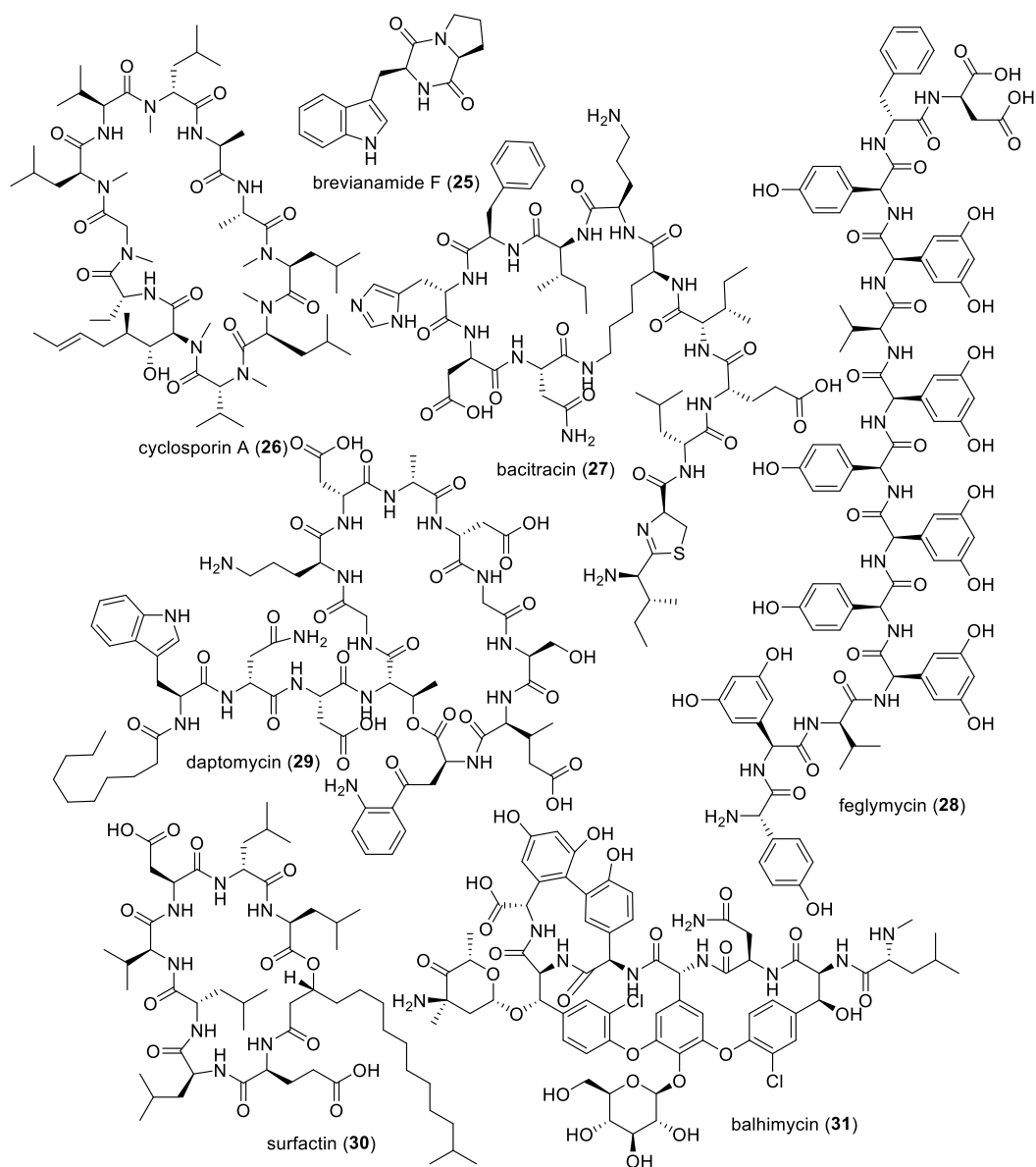
Figure 9: Nonribosomal peptides and derivatives thereof produced by various microorganisms.

Like PKS, NRPS are large modular mega proteins. Depending on the sequential assembly of the condensation substrates or whether individual modules are used several times, three different NRPS types are defined. Linear type NRPSs use every module only once and the sequence of the modules

determines the order of the condensed substrates. In iterative NRPSs, modules are used several times until the product is released. For non-linear type NRPSs, the order of the condensation product does not match the order of the modules (**Figure 10**).[98] The minimal domain set up of a module consists of an adenylation (A), a thiolation (T), also known as peptidyl carrier protein (PCP), and a condensation (C) domain (**Figure 10A**). To incorporate an amino acid into a growing peptide chain, the amino acid is first selected and activated *via* adenylation by the A domain. Consequently, the A domain catalyzes the transfer of the aminoacyl residue to the thiol group of the T domain. In the second step, the C domain catalyzes the assembly of a downstream amino acid bound to a T domain with an upstream amino acid *via* peptide bond formation. The elongated peptide chain is finally bound to the downstream T domain and ready for further elongation at the next module. Elongation goes on in a linear fashion depending on the module order until the thioester domain (TE) is reached. This domain catalyzes the release of the formed peptide either by hydrolysis resulting in a linear peptide or by an intramolecular cyclization reaction resulting in a cyclic peptide.[99] In addition to the minimal domains, a module can harbor additional domains that catalyze different modifications at the activated amino acid. Such modifications include epimerization, heterocyclization, formylation, halogenation, and oxidation or reduction reactions.[98]

As with PKSs, NRPSs are interesting candidates for protein engineering due to their modular nature and their diversity in accepted substrates. Various approaches were developed to redesign an NRPS so that completely new substances were generated, or the products were modified so that they meet the desired pharmaceutical properties.[100] Domain or module exchange approaches were successfully used to alter the polypeptide sequence. For example, the antibiotic daptomycin (**29**) is a NRPS product from *Streptomyces roseosporus*. The exchange of one or multiple modules leads to the predicted change of the amino acid sequence.[101] Instead of whole module exchanges, also single domain swapping is possible as well to generate new products. This was done in a combinatorial study to get new pyoverdine derivatives by swapping A or C domains.[102] Additionally, deletion or insertion of modules is another way to alter the produced products. This is exemplified by the reduction of the ring size of surfactin (**30**) through NRPS module deletion or by elongation of balhimycin (**31**) with additional amino acids *via* NPRS module insertion.[103,104]
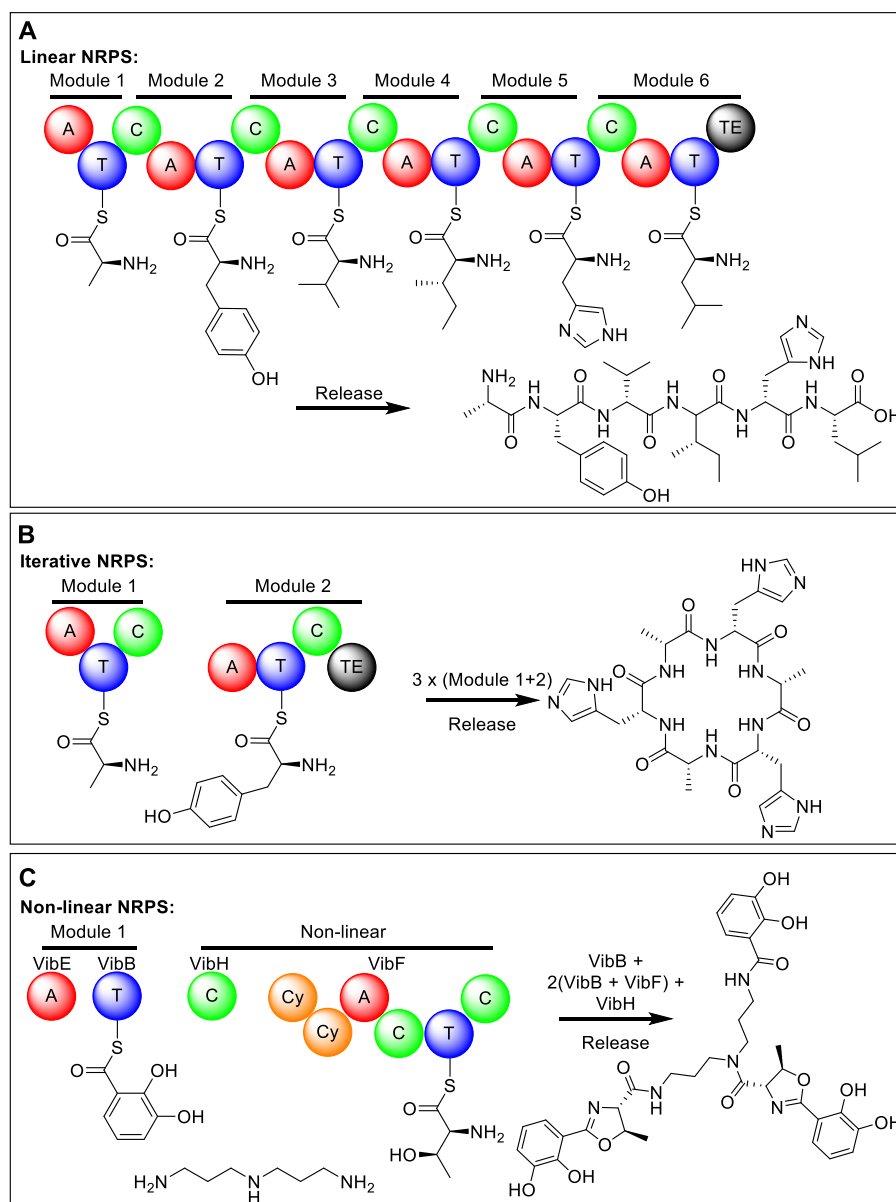
Figure 10: Schematic presentation of the linear (**A**), iterative (**B**) and non-linear (**C**) NRPS types.

1.2.3. Prenyltransferases

Prenyltransferases fulfill an important function in biosynthesis within primary and secondary metabolism. By catalyzing the attachment of a prenyl moiety from a donor to a wide range of prenyl acceptors, they contribute substantially to natural product diversity. The prenyl donors consisting of n × C5 units *e.g.* dimethylallyl (DMAPP; n = 1), geranyl (GPP; n = 2), farnesyl (FPP; n = 3) or geranylgeranyl (GGPP; n = 4) diphosphates (**Figure 11A**). These prenyl donors are derived from the terpenoid biosynthetic pathway.[105,106] In the primary metabolism they are involved, for example in the mena- and ubiquinone or heme biosynthesis.[107,108] Furthermore, prenyltransferases are also able to catalyze the farnesylation and geranylgeranylation of proteins *e.g.* members of the Ras superfamily GTPases. The correct and stable localization in membranes of these members is crucial for their proper function, for example in growth or intracellular protein transport regulation.[109] Most of the known and characterized prenyltransferases participate in the biosynthesis of various secondary metabolites. For

this, they transfer the prenyl moiety in a regular or reverse fashion onto a plethora of substance classes such as flavonoids, coumarins, xanthones, quinones, and especially indoles (**Figure 11B**). Furthermore, they are the key enzymes involved in the biosynthesis of isoprenoid natural products.[106]
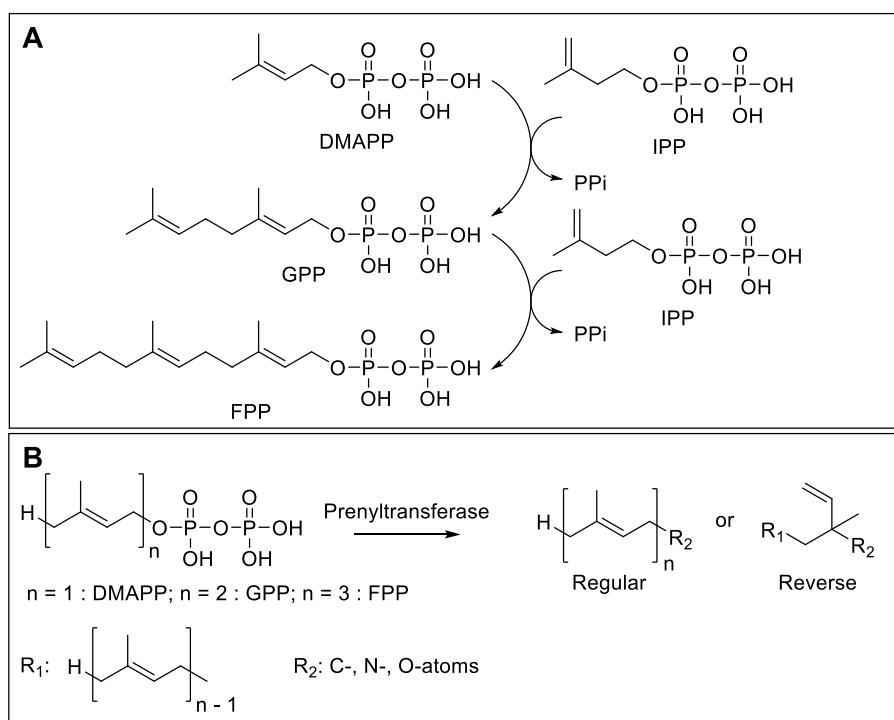
Figure 11: Schematic presentation of IPP, DMAPP, GPP and FPP structures (**A**) and the regular or reverse prenylation (**B**).

Terpenes, terpenoids and steroids are belonging to the large family of the isoprenoids with more than 80,000 known members.[110] The biosynthetic precursors, DMAPP and isopentenyl diphosphate (IPP) are either derived from the mevalonate (MVA) or methylerythritol phosphate (MEP) pathway and provide the basis for the isoprenoid biosynthesis.[111] With the utilization of these precursors, *cis-* and *trans-*prenyltransferases are the key enzymes in the biosynthetic pathway of isoprenoid production. They catalyze the head-to-tail consecutive condensation of DMAPP and IPP to form a carbon backbone with different sizes in length. The configuration of the formed double bond can be either *cis* (*Z*) or *trans* (*E*). Therefore, the prenyltransferases can be divided into *cis-* and *trans-*prenyltransferases.[111,112] *Trans-*prenyltransferases catalyze the formation of linear allylic diphosphates with a size between C_{10} to C_{50}.[113] In contrast, *cis-*prenyltransferases catalyze the formation of polyprenyl diphosphates ranging from C_{10} to $C_{10.000}$.[112] Although both groups of prenyltransferases catalyze similar reactions and are depending on Mg^{2+} ions for their reaction, they differ clearly in their three-dimensional structure.[106]

Due to their ability to catalyze the prenylation of aromatic substrates, a further group of prenyltransferases is classified as aromatic prenyltransferases. As tailoring enzymes, they contribute significantly to structural diversification in various biosynthetic pathways. In general, they catalyze the transfer of a prenyl moiety from a prenyl donor to an acceptor in regular or reverse fashion. Prenylation typically occurs at C, N or O atoms of a broad spectrum of substrate groups *e.g.* phenols, flavonoids,

coumarins, naphthalenes, indoles and derivatives thereof.[105,106,114] Depending on enzyme inherent factors such as soluble or membrane-bound state, metal ion dependency, structural organization, and motifs further classification into subgroups is possible. The assignment to the subgroups comes from the archetypical representative of the family. These are the UbiA family, the CloQ / NphB group, the LtxC group and the superfamily of dimethylallyltryptophan synthase (DMATS).[105,106]

Members of the UbiA family are membrane-bound proteins and show one or more NDxxDxxxD aspartate-rich motifs being responsible for coordination of the prenyl donor during the prenylation reaction.[115] In the majority of this family, the coordination step is assisted by divalent ions. They have an essential role in primary metabolism.[116] The bacterial UbiA, as well as the eukaryotic homolog COQ2, is involved in the biosynthesis of ubiquinones and menaquinones. Thereby, they catalyze the transfer of a polyprenyl chain to 4-hydroxybenzoate.[117,118] Another representative is the COX10 enzyme, responsible for the farnesylation of protoheme during heme A biosynthesis.[119] In addition to participation in primary metabolism, members of the UbiA family also play an important role in the biosynthesis of secondary metabolites. For example, a prenyltransferase from a *Streptomyces* species transfers a farnesyl chain to a quinoline chromophore during the biosynthesis of aurachin SS (**32**).[120] More recently, the UbiA prenyltransferase AscA from *Acremonium egyptiacum* was reported to catalyze the farnesylation of orsellinic acid in the biosynthesis of ascofuranone (**33**) and ascochlorin (**34**).[121]

The reactions catalyzed by soluble members of the CloQ / NphB group are mostly independent of the presence of divalent ions. Representatives of this group are found in bacteria and fungi. They share the same structural feature of an *αββα*-fold, also known as PT barrel, with prenyltransferases of the DMATS superfamily. This fold contains ten anti-parallel β-strands surrounded by *α*-helixes.[106,122] The eponymous enzymes CloQ und NphB of this group were identified in *Streptomyces roseochromogenes* and *Steptomyces* sp. CL190, respectively. NovQ is responsible for the prenylation of 4-Hydroxyphenylpyruvate in the biosynthesis of the antibiotic novobiocin (**35**), whereas, CloQ catalyzes the same reaction in the biosynthesis of clorobiocin (**36**). NphB catalyzes the prenylation of the naphthalene flaviolin in the biosynthetic pathway of naphterpin (**37**).[105,122]

The LtxC group consists of soluble, mostly bacterial prenyltransferases, which show no aspartate-rich motifs and are independent of divalent ions. The natural substrates of these prenyltransferases are usually indole-containing substances.*[123]* They share a relatively high sequence similarity with the members of fungal DMATS (see below), so that these two groups can theoretically be merged. LtxC was identified in the cyanobacterium *Lyngbya majuscula*. It is involved in the biosynthesis of lyngbyatoxin A (**38**), where it catalyzes the reverse prenylation of (-)-indolactam V.[123] Resolving the crystal structure of the LtxC homolog TleC showed that this group of prenyltransferases also share the

structural *αββα*-fold motif.[124] The bacterial prenyltransferase PriB from *Streptomyces* sp. RM-5-8 also owns this motif and catalyzes the prenylation of ʟ-tryptophan. Furthermore, it shows a wide promiscuity towards the accepted prenyl donors and acceptors.[125]
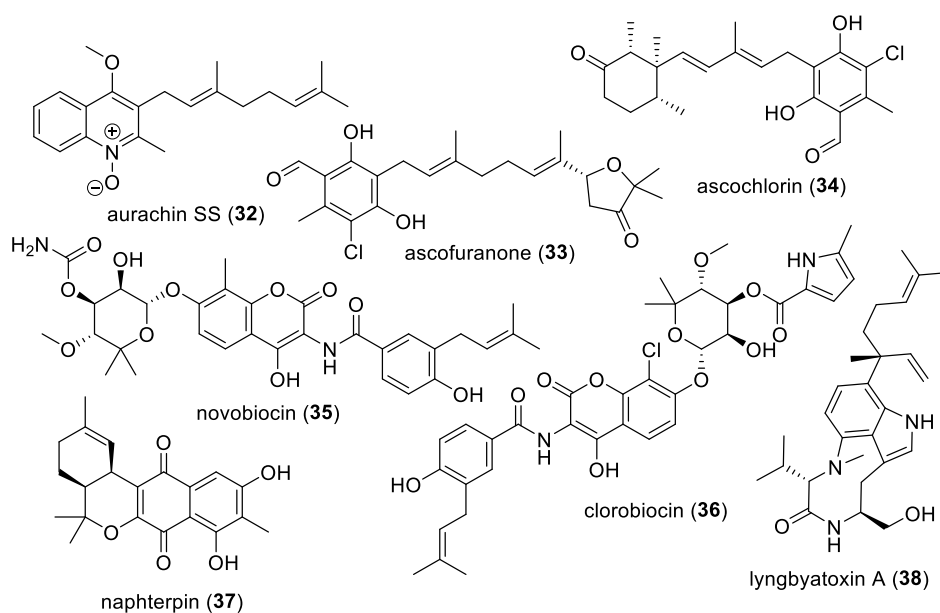

Figure 12: Prenylated compounds produced by various microorganisms

The fourth group of aromatic prenyltransferases belongs to the DMATS superfamily. To date, more than 60 representatives of this family have been identified and characterized. For this reason, they are the best-studied group of prenyltransferases.[106] In addition to biochemical characterization, the crystal structure of some prenyltransferases such as FtmPT1, AtaPT, FgaPT2, and CdpNPT has also been elucidated.[126-129] This information allowed the identification of characteristic structural features of the DMATS superfamily. Members of this family are soluble enzymes with no aspartate-rich motifs. Consequently, their reactions are independent of divalent metal ions. However, numerous examples have shown that the presence of such ions, *e.g.* Mg^{2+} and Ca^{2+} have a positive influence on their activities.[106,130] The members of the DMATS superfamily share low or no sequence similarity with the CloQ / NphB group on the amino acid level. Nevertheless, both groups show the characteristic tertiary structure of the PT barrel.[126] The elucidation of the crystal structure also enabled the clarification of the reaction mechanism. Formation of a dimethylallylic cation by cleavage of the diphosphate group from the prenyl donor initiates the Friedel-Crafts alkylation. The dimethylallylic cation is then nucleophilically attacked by the substrate to be prenylated, for example by the indole nucleus. Rearomatization finally leads to the release of the prenylated product.[126] One important feature of the DMATS prenyltransferases is their high flexibility toward aromatic substrates. Common substrates are indole-containing substances such as tryptophan and tryptophan-containing cyclic dipeptides. In contrast, most DMATS are specific for their prenyl donors and use DMAPP as a prenyl donor. However, there are also examples using FPP and GPP for their reactions.[106]
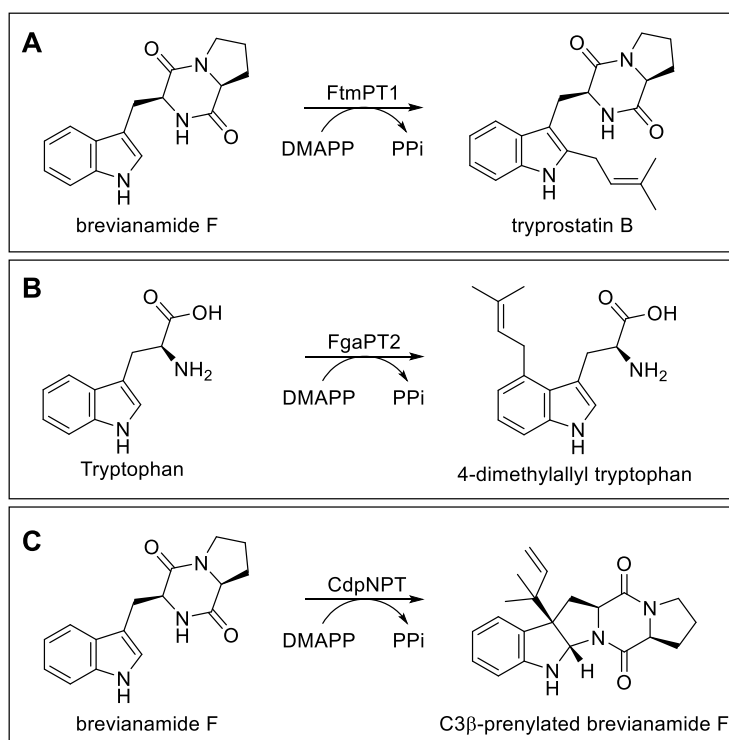
Figure 13: Prenylation reactions catalyzed by the prenyltransferases FtmPT1 (**A**), FgaPT2 (**B**) and CdpNPT (**C**).

In 1992, the first representative of the DMATS superfamily, DmaW, was identified in *Claviceps fusiformis (reported as C. purpurea)*. DmaW catalyzes the attachment of the prenyl moiety at position C4 of tryptophan. Therefore, it is also known as 4-DMATS.[131,132] In the following two decades, the research on DMATS was intensified. As a result, a wide variety of members are now identified, so that for every possible prenylation position of tryptophan or tryptophan-containing cyclic dipeptides at least one DMATS is described.[106,133] FtmPT1 from *Aspergillus fumigatus* is a cyclic dipeptide prenyltransferase involved in the biosynthesis of fumitremorgin-type alkaloids. In this pathway, it catalyzes the regular prenylation of brevianamide F (**24**) at position C2 of the indole nucleus (**Figure 13A**). Later it was shown that it could prenylate further cyclic dipeptides. In addition to the C2-prenylation, prenylation at position C3 may also occur, accompanying the formation of a fused ring system between the indole nucleus and the diketopiperazine backbone.[134,135] FgaPT2 from *Aspergillus fumigatus* is an orthologue of DmaW (4-DMATS) (**Figure 13B**), catalyzing the initial step in the biosynthesis of ergot alkaloids.[131,136] A recent study showed that FgaPT2 could also transfer other unnatural prenyl donors to tryptophan. It was shown that the prenyl donor could have an influence on the prenylation position. In addition to position C4, the unnatural prenyl donors were also transferred to positions N1, C3 and C5.[137] Furthermore, the cyclic dipeptide prenyltransferase CdpNPT identified in *Aspergillus fumigatus* catalyzes the reverse prenylation of tryptophan-containing cyclic peptides at position C3 (**Figure 13C**).[128,138,139]

2. Aim of this thesis

The following issues have been addressed in this thesis:

<u>Expanding the prenyltransferase substrate diversity by *in silico* methods and experimental proof of the prediction strength:</u>

Members of the DMATS superfamily contribute immensely to the structural diversity of natural products. They are also known for their promiscuity toward accepted substrates. Previous strategies to identify further substrates were based on comparisons to known substrates by trial and error approach. Prof. Peter Kolb and Dr. Jakub Gunera developed a database based on reactions of known DMATS enzymes and structural information to predict new possible substrates. The predictive strength was then tested experimentally. Therefore, the following experiments were carried out by the author of this thesis:

➢ Overproduction and purification of the DMATS enzymes, FtmPT1, FgaPT2, and CdpNPT.

➢ Assessing the acceptance of the 38 selected substrates by *in vitro* enzyme assays and LC-MS analysis.

➢ Determination of the conversion rates of the accepted substrates with the tested enzymes.

➢ Isolation and structure elucidation of the enzyme products of two substrates

<u>Application of EchPT1 and EchPT2 as biocatalysts for production of tri- and tetraprenylated cyclodipeptides:</u>

The two DMATS prenyltransferases EchPT1 and EchPT2 govern the consecutive prenylation of the NRPS product *cyclo*-L-tryptophanyl-L-alaninyl during the biosynthesis of echinulin and derivatives thereof in *Aspergillus ruber*. In this biosynthetic pathway, up to four prenylation reactions are involved. Because of the known promiscuity of the DMATS enzymes, the potential use of EchPT1 and EchPT2 in the production of tri- and tetraprenylated cyclodipeptides was investigated. For this, the following experiments were carried out:

➢ Overproduction and purification of the DMATS enzymes EchPT1 and EchPT2

➢ Consecutive enzymatic assays with EchPT1 and EchPT2 with stereoisomers of cyclo-tryptophanyl-alaninyl and *cyclo*-tryptophanyl-prolinyl.

➢ Isolation and structure elucidation of the obtained products

➢ Determination of the kinetic parameters of the enzymes

<u>Manipulation of *Penicillium crustosum* for use as a potential heterologous expression host:</u>

Heterologous expression of silent biosynthetic gene clusters is an excellent method for the discovery of new secondary metabolites. A suitable host for heterologous expression must fulfill various requirements. Firstly, it should grow rapidly, provide enough precursors for the biosynthesis of natural products and show good genetic tractability. Furthermore, the successful integration of foreign genes should be easy to verify, and the own metabolic background should be clear for easy identification of newly produced products. In this study, the Ascomycota *Penicillium crustosum* was investigated as a new putative host for heterologous expression. The following experiments were carried out:

- ➢ Sequencing and analysis of the *Penicillium crustosum* genome.
- ➢ Identification of the putative conidial pigment synthase PKS Pcr4401 by sequence comparison.
- ➢ Establishment of the deletion protocol for *Penicillium crustosum*
- ➢ Verification of the involvement of PKS Pcr4401 in melanin biosynthesis by gene deletion and heterologous expression experiments.
- ➢ Isolation and structure elucidation of the heterologously produced PKS Pcr4401 products.
- ➢ Creation of a *pyrG* deficient *Penicillium crustosum* strain
- ➢ Creation of two plasmids for the heterologous expression of foreign genes at the *pcr4401* locus in *Penicillium crustosum*.
- ➢ Testing the potential of *Penicillium crustosum* for heterologous expression by the integration of three different PKS genes.
- ➢ Extraction, isolation and structure elucidation of the heterologously produced PKS products from the generated *Penicillium crustosum* strains.

3. Results and discussion

3.1. The substrate prediction database PrenBD for further expanding the prenyltransferase substrate diversity by *in silico* methods – experimental proof of the prediction strength

Enzymes belonging to the diverse family of prenyltransferases participating in primary and especially in secondary metabolism. As tailoring enzymes, they contribute to structural diversification in various biosynthetic pathways.[106] At the beginning of this study, more than 40 fungal and bacterial prenyltransferases belonging to the dimethylallyltryptophan synthase superfamily have been identified and biochemically characterized.[106] They show a broad spectrum of aromatic substrates belonging to different substance classes *e.g.* xanthones, naphthalenes or indole derivatives including tryptophan-containing dipeptides. Whereas prenyl acceptor promiscuity for a given prenyltransferase is commonly observed, they are mostly strict with prenyl donors.[140] However, there are also prenyltransferases, *e.g.* AtaPT, with both prenyl acceptor and donor promiscuity.[129] Resolving the structures of prenyltransferases and multiple sequence analyses opened new opportunities to expand their catalytic applications. By identifying and changing amino acid residues involved in substrate binding, the prenylation position could be altered[141] and/or the number of accepted substrates could be expanded.[142-144] Additionally, by identifying and changing amino acids involved in prenyl donor binding, the promiscuity prenyl donors could be increased or altered.[129,145,146]

Beside protein-engineering, the acceptor promiscuity itself offers unused potential to further increase product diversity by using prenyltransferases as biocatalysts. Up to this study, the selection of substrates of interest was done by similarity comparison with known substrates/products and by trial and error. To address and use the promiscuity of prenyltransferases in a more logical and determined way the PrenDB database was created. This database is based on the data of 32 characterized prenyltransferases, 167 substrates and their reaction information. Referring to this information, an algorithm was created to automatically identify and extract the reactive atom, where the prenylation occurs and its surrounding, the so-called repitope. The next step was the multistep virtual screening of putative substrates (**Figure 14**). The collected repitope information was used to search the ZINC database, with over five million commercially available compounds, for substances that share at least one of the identified repitopes. The identified substrates were further screened by their physico-chemical properties and their one- and three-dimensional shape congruency to known accepted substrates. In the last step, the remaining putative substrates were docked into the resolved crystal structure of FtmPT1, FgaPT2, and CdpNPT to ensure the best possible fitting into the enzyme. The best fitting results were analyzed manually, and 38 predicted substrates were selected for experimental validation.
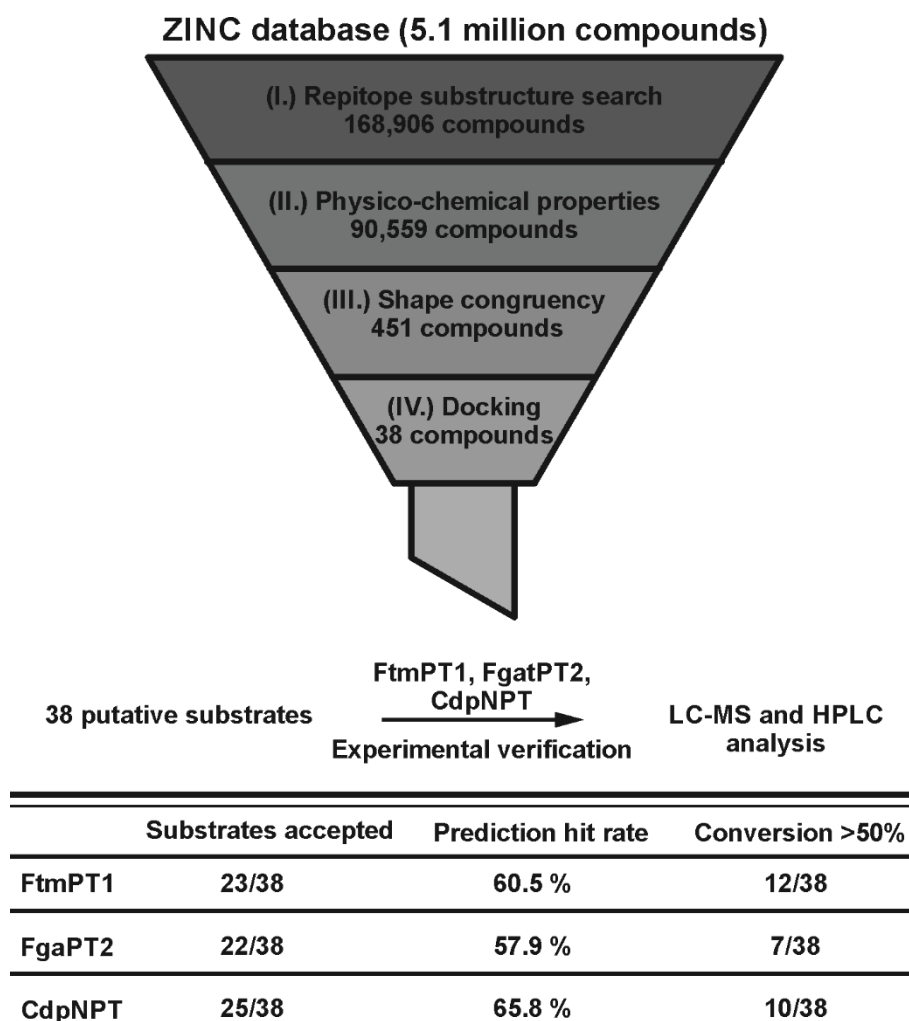
ZINC database (5.1 million compounds)

(I.) Repitope substructure search
168,906 compounds

(II.) Physico-chemical properties
90,559 compounds

(III.) Shape congruency
451 compounds

(IV.) Docking
38 compounds

38 putative substrates → **FtmPT1, FgatPT2, CdpNPT** — **Experimental verification** → **LC-MS and HPLC analysis**

	Substrates accepted	Prediction hit rate	Conversion >50%
FtmPT1	23/38	60.5 %	12/38
FgaPT2	22/38	57.9 %	7/38
CdpNPT	25/38	65.8 %	10/38

Figure 14: Schematic presentation of the four-step screening process and information about substrate acceptance and prediction hit rate.

To test the predictive strength of PrenDB, the 38 predicted substrates were assayed with the two cyclic dipeptide prenyltransferases FtmPT1 and CdpNPT as well as the tryptophan prenyltransferase FgaPT2 for acceptance and conversion yield (**Figure 14**). To analyze the acceptance, the reaction mixtures of the predicted substrates, the corresponding purified enzymes, and DMAPP as prenyl donor were incubated and afterward analyzed *via* LC-MS. Acceptance was evaluated by detection of product peaks with a longer retention time compared to the substrate and 68 Da larger [M+H]$^+$ ions for each prenyl moiety than that of the substrate. Multiple prenylation attachments have not been observed in this study. The acceptance of 23 substrates by FtmPT1, 22 by FgaPT2, and 25 by CdpNPT corresponded to prediction hit rates of 60.5 %, 57.9 %, and 65.8 %, respectively. Remarkably, product yields greater than 50 % were detected in FtmPT1, FgaPT2, and CdpNPT assays with 12, 7, and 10 substrates, respectively (**Figure 14**). The fact that the native substrates for 78 % of the prenyltransferases present in PrenDB are tryptophan or indole derivatives leads to a clearly indole-biased database regarding the identified epitopes. This is also manifested in the predicted substrates. Nevertheless, substrates with a good conversion could be identified, even though no comparable natural substrate counterparts are known for these enzymes. This is exemplified by substances with an indole-4-imidazolin-2-one motif

or with a benzylated hydroxyl-indole structure, respectively. After conversion with FtmPT1, the products of indole 3-ethyl ethylsulfonamide and 6-benzyloxyindole were investigated and their structure elucidated. 3-ethyl ethylsulfonamide shows a conversion rate of 98.6% with FtmPT1. Two products in a ratio of 1 to 1 could be identified. In one product, the prenyl group was attached to position N1. Whereas, the other product was prenylated reversely at position C3, which simultaneously led to the formation of a 6/5/5 fused ring system. 6-benzyloxyindole shows a conversion rate of 60.1% with FtmPT1. Two products with regular prenylation at position C2 and C5 respectively and a ratio of 1 to 0.85 were identified. This study shows the strength of systematic analysis of known reactions of an enzyme class to predict new possible substrates. Experimental verification of the prediction showed that 71 % of the predicted substrates were accepted by at least one of the three tested prenyltransferases. This provides an excellent opportunity to identify new possible substrate classes and thus significantly expand the repertoire of prenylated substances.

This project was a cooperation with Dr. Jakub Gunera and Prof. Peter Kolb. Dr. Jakub Gunera was responsible for PrenDB design and implementation of the cataloged prenyltransferase reaction, substrate space analysis, the automatic repitope extraction and virtual substrate screening, docking, and selection. Subsequently, I was responsible for the design of the enzyme assays, their execution using heterologously produced enzymes and the analysis using LC-MS. Furthermore, I isolated the enzymatic products and elucidated their structure.

For further details about this publication see section 4.1:

Jakub Gunera*, **Florian Kindinger***, Shu-Ming Li and Peter Kolb (2017) PrenDB: A substrate prediction database to enable biocatalytic use of prenyltransferases. Journal of Biological Chemistry 292: 4003-4021

*The authors contributed equally to this work.

3.2. Biochemical investigations on the both prenyltransferases EchPT1 and EchPT2 from the echinulin biosynthetic cluster and their application as biocatalysts for production of tri- and tetraprenylated cyclodipeptides

Nonribosomal peptides, especially those with an indole nucleus, are interesting starting candidates for the development of new drugs. This indole nucleus is also found in the essential amino acid tryptophan, which is used for the biosynthesis of serotonin and almost all proteins. Numerous receptors such as G protein-coupled receptors, possess a binding domain for the recognition of the indole motif. This explains why many indole motif-containing substances have a pharmaceutical effect *e.g.* anti-inflammatory effects or why they act as 5-hydroxytryptamine receptor agonists or antagonists.[147] An interesting group of nonribosomal peptides are cyclic dipeptides with a 2,5-diketopiperazine nucleus consisting of tryptophan and another amino acid which are mainly produced by microorganisms.[148,149]

Echinulin (**39**) and its congeners also belong to this group. In two studies, we examined the biosynthesis of these substances and the enzymes involved. Echinulin consists of L-tryptophan and L-alanine with a reverse attached prenyl moiety at position C2 and two regular prenyl residues, one at positions C5 and another at position C7 of the indole nucleus. Well-known producers of echinulin and its congeners *e.g.* preechinulin (**40**), variecolorin L (**41**), neoechinulin A (**42**), dehydroechinulin (**43**), and neoechinulin B (**44**) are *Aspergillus* species such as *Aspergillus amstelodami* and *Aspergillus ruber* (**Figure 15A**).[150] In the genome sequence of *Aspergillus ruber*, genes coding for three prenyltransferases of the DMATS superfamily were identified. Two of them, *echPT1* and *echPT2*, are closely located to the NRPS gene *echPS*. In addition, genes for a putative transporter and a cytochrome P450 enzyme, *echP450*, are also found in close proximity (**Figure 15B**). This putative gene cluster encodes the enzymatic requirements for the biosynthetic steps of echinulin, postulated from its structure. Since deletion experiments were not successful, due to the insufficient genetic tractability of *Aspergillus ruber*, we decided to characterize the biosynthetic pathway of echinulin by biochemical characterization of the involved enzymes. Therefore, the prenyltransferases EchPT1 and EchPT2 were heterologously produced in *E. coli* and subsequently isolated and purified. Afterward, the acceptance of the putative EchPS product *cyclo*-L-Trp-L-Ala (**45**) with EchPT1 and EchPT2 was investigated by in *vitro* enzymatic assays. Through extensive investigations using *in vitro* enzyme assays, isolation, and structure elucidation of intermediates and end products, we were able to provide a well-founded hypothesis for the biosynthesis of echinulin (**Figure 15C**). According to this hypothesis, EchPS catalyzes the formation of *cyclo*-L-Trp-L-Ala, which is subsequently prenylated reversely at position C2 by EchPT1 leading to the formation of preechinulin (**40**). Only this product is accepted by EchPT2. As a result, regular prenylation occurs at position C5 leading to the formation of tardioxopiperazine A (**46**). This

product is then further prenylated at position C7 resulting in the formation of echinulin (**39**). Both prenylations are catalyzed by EchPT2. The fact that the prenyltransferase EchPT2 catalyzes two prenylation reactions during the biosynthesis of echinulin is already remarkable. There are only a few other known prenyltransferases that also catalyze several prenylation reactions within one biosynthetic pathway. For example, the prenyltransferase HsPTpat from *Hypericum calycinum* catalyzes the *gem*-diprenylation during the biosynthesis of patulone.[151] However, EchPT2 has another unique feature and is able to transfer up to three additional prenyl moieties to preechinulin (**40**). Meaning, the resulted di- and triprenylated products simultaneously serve as substrates for EchPT2. The enzyme EchP450 is expected to catalyze the single or double dehydrogenation reaction at the diketopiperazine core from preechinulin (**40**). Thereby, neoechinulin A (**42**) and neoechinulin B (**44**), the precursors of the neoechinulin A and B series, are created, which can also be prenylated several times by EchPT2 to form isoechinulin A (**47**), dehydroechinulin (**43**), isoechinulin B (**48**) and tetrahydroechinulin (**49**).
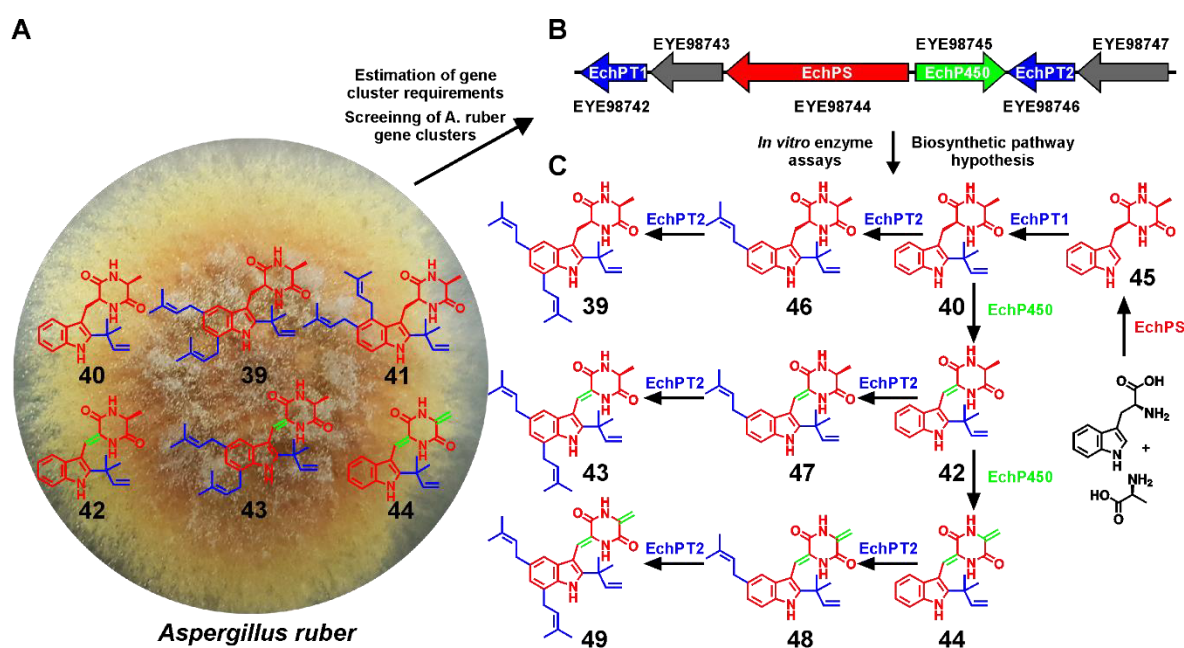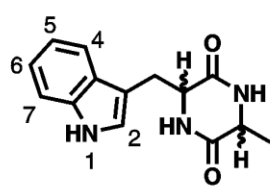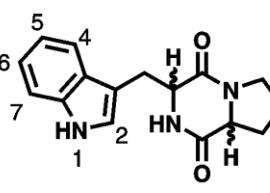
Figure 15: Schematic presentation of some echinulin derivatives isolated from *A. ruber* (**A**), the putative echinulin gene cluster (**B**) and the echinulin biosynthetic pathway hypothesis (**C**).

In a further study, the possible application of the prenyltransferases EchPT1 and EchPT2 for the biosynthetic production of tri- and tetraprenylated cyclic dipeptides and the influence of the substrate stereochemistry was investigated. Prenyltransferases show high promiscuity toward the accepted substrates and in some cases also toward the prenyl donor.[106,129] This provides an excellent basis for the chemoenzymatic synthesis of new prenylated compounds. Previous studies have already shown that stereochemistry can have a significant influence on the acceptance of the substrate.[139] Therefore, we decided to study a total of eight substrates consisting of all the four stereoisomers of *cyclo*-Trp-Ala and *cyclo*-Trp-Pro, respectively. With the help of EchPT1, we successfully prenylated all eight

substrates reversely at position C2. A total of six of the eight substrates examined showed good conversion yields from 52.6 ± 2.4 to 87.3 ± 1.8 % (**Table 1**). Only the two exclusively D-configured cyclodipeptides *cyclo*-D-Trp-D-Ala and *cyclo*-D-Trp-D-Pro were poorly accepted, with conversion rates of 16.3 ± 3.9 and 42.5 ± 1.9 %, respectively (**Table 1**). A similar effect has already been observed for the prenyltransferases AnaPT and CdpC3PT.[139] The C2 reversely prenylated substrates were isolated and used as substrates for further investigations with EchPT2, which also accepted all eight reversely monoprenylated *cyclo*-Trp-Ala and *cyclo*-Trp-Pro stereoisomers. The conversion yields were calculated between 21.2 ± 2.4 and 98.3 ± 1.2 % (**Table 1**). Again, we were able to observe that the exclusive D-configured isomers with conversions rate of 21.2 ± 2.4 % for *cyclo*-D-2-*tert*-DMA-Trp-D-Pro and 47.0 ± 0.9 % for *cyclo*-D-2-*tert*-DMA-Trp-D-Ala were clearly less converted compared to the other isomers. Similar to the biosynthesis of echinulin, EchPT2 also catalyzes the formation of di-, tri- and partly tetra-prenylated products with the substrates mentioned above. Thereby, the main products are the C2, C5, and C7 triprenylated derivatives. Except for *cyclo*-L-2-*tert*-DMA-Trp-L-Pro and *cyclo*-D-2-*tert*-DMA-Trp-D-Pro, the formation of one or two tetra-prenylated products was observed after the conversion of the substrates with EchPT2. In the formation of tetra-prenylated products, a clear influence of the substrate configuration on the formation of these products was observed. The ratio of tetra-prenylated products in (S, S) and (R, R) configured substrates was only 4% of total yields. In comparison, the ratio of C2, C4, C5, and C7-tetraprenylated products of the (S, R) and (R, S) configured substrates was determined at about 30 % (**Table 1**).

Table 1: Conversion rate resp. proportion of total turnover in the biosynthetic production of mono- to tetraprenylated cyclic dipeptides catalyzed by EchPT1 and EchPT2.

Substrate and configuration			Conversion rate	Multi-prenylated product ratio of total yields		
			Monoprenylated	Diprenylated	Triprenylated	Tetraprenylated
	(S, S)		67.4 %	4.1 %	90.3 %	3.6 %
	(S, R)	EchPT1 → EchPT2 →	57.7 %	2.0 %	65.2 %	28.6 %
	(R, R)		16.3 %	3.0 %	40.2 %	3.7 %
	(R, S)		52.6 %	1.9 %	56.7 %	25.2 %
	(S, S)		87.3 %	3.3 %	83.1 %	-
	(S, R)	EchPT1 → EchPT2 →	82.1 %	1.3 %	53.7 %	36.4 %
	(R, R)		42.5 %	1.6 %	19.6 %	-
	(R, S)		74.7 %	1.5 %	49.6 %	36.2 %

With their prenyl acceptor promiscuity, the prenyltransferases EchPT1 and EchPT2 join a steadily growing group of prenyltransferases.[106,140] In this study, we were able to show that EchPT1 and EchPT2 are well suited for use in the chemoenzymatic production of multiple prenylated cyclic dipeptides. This

is of particular importance since an increase in the biological activity of prenylated compounds compared to their unprenylated counterparts was observed for numerous compounds.[152,153] For example, it could be shown that the presence of a prenyl group in various polyphenols is crucial for the inhibitory effect on melanin biosynthesis in B16 melanoma cells.[154] The increase of the biological activity is presumably based on a higher lipophilicity of the prenylated substances, which can lead to a better interaction with biological membranes.[155] In another study, it was demonstrated by *in silico* modeling that the prenyl group of a flavonoid fits very well into the hydrophobic binding pocket of the human estrogen receptor resulting in a higher agonistic activity.[156] The structural elucidation and analysis of some enzymes such as EchPT1 and EchPT2 is of fundamental importance for the targeted structure-based enzyme design. By using structure-based engineering of some prenyltransferases, the spectrum of prenyl acceptors and donors has already been increased considerably. In a recent study, the accepted prenyl donor could be changed from DMAPP to FPP or GPP by identification of the amino acid Met328 in FgaPT2 and its mutation to glycine or alanine.[146] Similar prenyl donor gatekeeper amino acids have also been identified in other prenyltransferases such as CdpNPT, CdpC2PT, CdpC3PT, FtmPT1, and BrePT. Their mutation also led to a clear increase in the acceptance of GPP.[145] Such structure-based changes would further increase the applicability of EchPT1 and EchPT2 as biocatalysts for the synthesis of new natural products.

The studies were realized in cooperation with Viola Wohlgemuth. The identification of the putative echinulin gene cluster by bioinformatic methods, as well as the culturing of *Aspergillus ruber* and the isolation of a few natural substances were carried out by me. Furthermore, I was involved in the purification of the prenyltransferases EchPT1 and EchPT2, as well as in the enzyme assays and isolation of some of the formed products.

For further details about these publications see section 4.2 and 4.3:

Viola Wohlgemuth, **Florian Kindinger**, Xiulan Xie, Bin-Gui Wang, and Shu-Ming Li (2017) Two Prenyltransferases govern a consecutive prenylation cascade in the biosynthesis of echinulin and neoechinulin. Organic Letters. 19: 5928-5931

Viola Wohlgemuth*, **Florian Kindinger*** and Shu-Ming Li (2018) Convenient synthetic approach for tri- and tetraprenylated cyclodipeptides by consecutive enzymatic prenylations. Applied Microbiology and Biotechnology 102: 2671-2681

*The authors contributed equally to this work.

3.3. Expanding the available filamentous fungal expression hosts: Manipulation on *Penicillium crustosum* for use as a potential heterologous expression host

The ongoing progress in the development of sequencing technologies and the associated reduction in costs for the sequencing of a genome is leading to a flood of genomic information from various organisms.[157] At the same time, analysis of this information led to the conclusion that the genetic potential for the production of secondary metabolites is much greater than that of the natural products actually produced. The explanation for this is that most of the biosynthetic gene clusters are silent, *i.e.* not active, under the standard laboratory cultivation conditions.[40] Therefore, different methods have been developed to exploit the unused potential of these silent gene clusters. A particularly suitable approach to realize this is the heterologous expression of the gene clusters in a suitable expression host.[158] Hence, the potential of *Penicillium crustosum* for heterologous gene expression was investigated in order to expand the availability of appropriate expression hosts.

First, the genome of *Penicillium crustosum* was sequenced and analyzed, leading to the identification of 56 putative BGCs. We were able to show that this fungus produces almost no native secondary metabolites when cultivated in minimal medium. This is an ideal prerequisite to identify and isolate heterologously produced new substances easily. The disruption of the PKS *pcr4401* gene by deletions experiments was accompanied by the occurrence of albino phenotype. Subsequently, heterologous expression of the PKS *pcr4401* gene in *Aspergillus nidulans* confirmed the involvement of this PKS in the biosynthesis of the melanin precursor YWA1 (**50**) which is responsible for the green pigmentation of the conidia. Furthermore, the dehydrated YWA1 derivative nor-rubrofusarin (**51**) was observed as a non-enzymatic side product (**Figure 16**). [53,159]
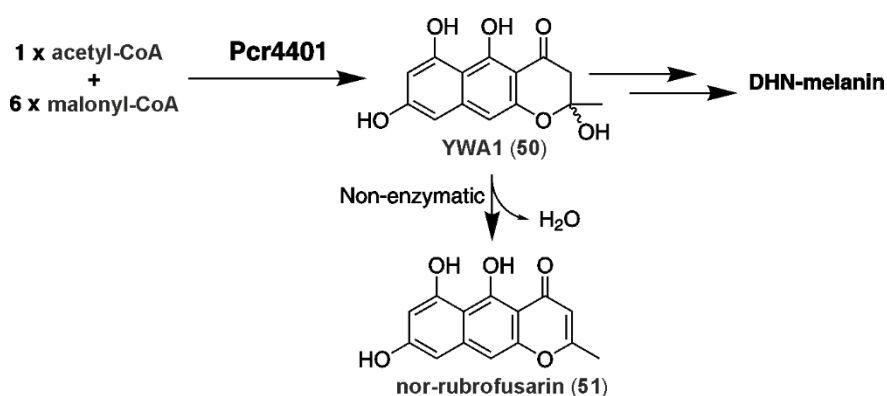
Figure 16: Formation of YWA1 and nor-rubrofusarin by *Penicillium crustosum* PKS Pcr4401.

Therefore, the deletion of the PKS Pcr4401 leads to the observed albino phenotype. The successful introduction of a foreign gene into this locus can therefore be easily determined by the phenotype. This, together with the clean metabolic profile, rapid growth, good genetic manipulability as well as simple and inexpensive cultivation conditions, makes *Penicillium crustosum* a potentially usable expression host. To verify this, two unknown PKS genes from *Penicillium camemberti* and *Penicillium*

brevicompactum were successfully introduced into *Penicillium crustosum* based on the hygromycin B resistance selection marker. For this purpose, the plasmid pFK20 was created, allowing the simple cloning of a foreign gene and targeting the *pcr4401* gene locus. Subsequently, the heterologously produced products were isolated and their structure elucidated as 3,5-dimethylorsellinic acid (**52**) and nor-toralactone (**53**) (**Figure 17**) by NMR and MS analyses. In order to ensure a better applicability of *Penicillium crustosum* as a heterologous expression host, *pyrG* was deleted afterward. The resulting uracil and uridine auxotrophic phenotype allows the use of *pyrG* as a further selection marker. Similar to pFK20, an additional plasmid pFK23 was constructed for foreign gene integration at the *pcr4401* locus. Instead of hygromycin B resistance as in the case of pFK20, the *pyrG* gene from *Aspergillus fumigatus* was used for selection. With this plasmid, we were also able to successfully introduce an additional PKS gene from *Aspergillus ruber* into the *Penicillium crustosum pcr4401* gene locus. From the strain generated in this way, 6-methylsalicylic acid (**54**) was isolated as the accumulated product and its structure was elucidated by NMR and MS analyses (**Figure 17**).
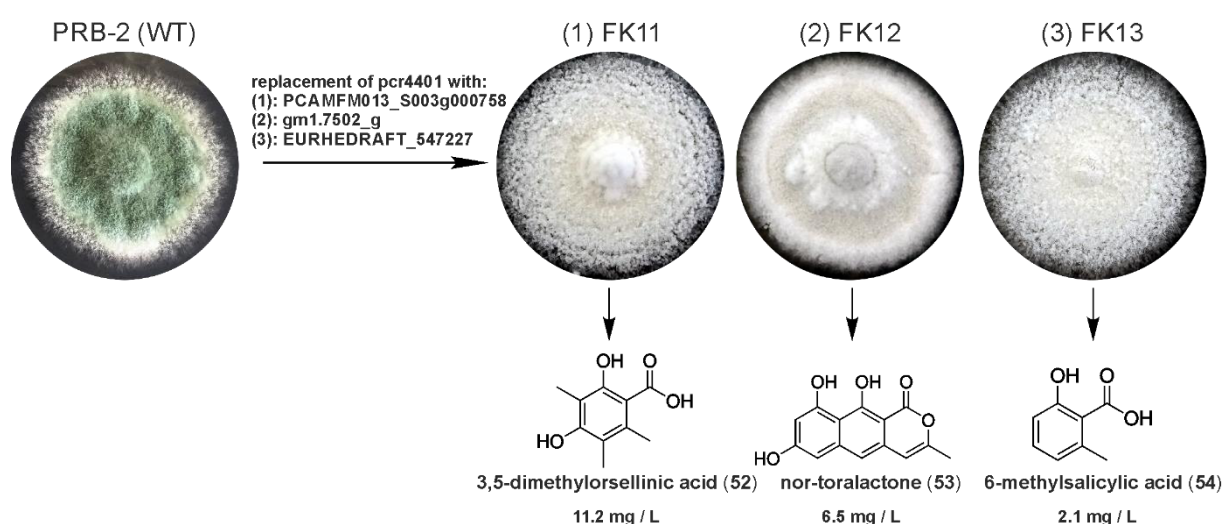
Figure 17: Heterologous production and product yield of the compounds **52**, **53** and **54** in *Penicillium crustosum*.

Although various fungi are already established and used as heterologous expression hosts, there is still a demand for alternative expression systems. The most frequently used fungi for heterologous expression are strains of the genus *Aspergillus, e.g. Aspergillus nidulans, Aspergillus niger or Aspergillus oryzae*.[160-162] However, since the splicing pattern can already differ between species and especially between different genera, this can lead to problems during heterologous gene expression. This was the case in the heterologous expression of a PKS-NRPS hybrid gene of *Magnaporthe oryzae* in the heterologous host *Aspergillus oryzae*.[163] Another potential problem is the possibility of host-dependent cross-chemistry events. Depending on the chosen expression host, different products may be formed. This was observed in the heterologous expression of a NRPS-like enzyme. Although the same gene was expressed heterologously in the hosts *Aspergillus niger* and *Aspergillus oryzae*, different products were formed.[164] An important step in the transformation of filamentous fungi by

polyethylene glycol-mediated protoplasts is the formation of the protoplasts. By using lysing enzymes from *Trichoderma harzianum* and yatalase from Corynebacterium sp. OZ-21, a protoplast concentration between 1.2×10^7 and 2.1×10^8 protoplasts per milliliter could be reached. These concentrations are comparable to the reported values for other fungi.[165] The gene targeting rate is between 5 - 10 %. In combination with the simple visual identifiability upon successful integration into the *pcr4401* locus *Penicillium crustosum* is well genetically tractable. Therefore, *Penicillium crustosum* can be a useful alternative for heterologous expression. Already with cultivation in minimal medium, heterologous production of PKS products with yields of 2.1 to 11.2 mg per liter medium was possible (**Figure 17**). When cultivated in complex media such as potato dextrose broth or rice, *Penicillium crustosum* produces large amounts of its own native secondary metabolites. This hinders the identification and isolation of new products under the given cultivation conditions. However, recently, we identified the natively produced secondary metabolites and the gene clusters responsible for the biosynthesis of these products.[166] A double deletion of the PKS backbone genes *claF* and *traA* would block the formation of all *Penicillium crustosum*'s own secondary metabolites. Thus, the clean metabolic profile favored for heterologous expression would also be available under these conditions. At the same time, a better supply of nutrients due to this medium would probably lead to higher yields.[167] In conclusion, we proved *Penicillium crustosum* as a suitable host for the expression of fungal secondary metabolite genes and have thus expanded the spectrum of applicable expression systems.

For further details about these publications see section 4.4:

Florian Kindinger, Jonas Nies, Anke Becker, Tianjiao Zhu and Shu-Ming Li (2019) Genomic locus of a *Penicillium crustosum* pigment as integration site for secondary metabolite gene expression. ACS Chemical Biology 14: 1227-1234

4. Publications

4.1. PrenDB, a substrate prediction database to enable biocatalytic use of prenyltransferases.

Gunera, J.*, **Kindinger, F.***, Li, S. M., and Kolb, P. (2017)

J. Biol. Chem. 292, 4003-4021

DOI: 10.1074/jbc.M116.759118

*The authors contributed equally to this work.

PrenDB, a Substrate Prediction Database to Enable Biocatalytic Use of Prenyltransferases*[S]

Jakub Gunera[‡§1,2], Florian Kindinger[¶1], Shu-Ming Li[§¶3], and ⬤ Peter Kolb[‡§4]

From the ‡Department of Pharmaceutical Chemistry, Philipps-University, Marburg, Hesse 35032, Germany, the ¶Institute of Pharmaceutical Biology and Biotechnology, Philipps-University, Marburg, Hesse 35032, Germany, and §Synmikro, LOEWE Centre for Synthetic Microbiology, Philipps-University, Marburg, Hesse 35043, Germany

Edited by Ruma Banerjee

Prenyltransferases of the dimethylallyltryptophan synthase (DMATS) superfamily catalyze the attachment of prenyl or prenyl-like moieties to diverse acceptor compounds. These acceptor molecules are generally aromatic in nature and mostly indole or indole-like. Their catalytic transformation represents a major skeletal diversification step in the biosynthesis of secondary metabolites, including the indole alkaloids. DMATS enzymes thus contribute significantly to the biological and pharmacological diversity of small molecule metabolites. Understanding the substrate specificity of these enzymes could create opportunities for their biocatalytic use in preparing complex synthetic scaffolds. However, there has been no framework to achieve this in a rational way. Here, we report a chemoinformatic pipeline to enable prenyltransferase substrate prediction. We systematically catalogued 32 unique prenyltransferases and 167 unique substrates to create possible reaction matrices and compiled these data into a browsable database named PrenDB. We then used a newly developed algorithm based on molecular fragmentation to automatically extract reactive chemical epitopes. The analysis of the collected data sheds light on the thus far explored substrate space of DMATS enzymes. To assess the predictive performance of our virtual reaction extraction tool, 38 potential substrates were tested as prenyl acceptors in assays with three prenyltransferases, and we were able to detect turnover in >55% of the cases. The database, PrenDB (www.kolblab.org/prendb.php), enables the prediction of potential substrates for chemoenzymatic synthesis through substructure similarity and virtual chemical transformation techniques. It aims at making prenyltransferases and their highly regio- and stereoselective reactions accessible to the research community for integration in synthetic work flows.

Prenylated primary and secondary metabolites, including indole alkaloids, flavonoids, coumarins, xanthones, quinones, and naphthalenes, are widely distributed in terrestrial and marine organisms. They exhibit a wide range of biological activities, including cytotoxic, antioxidant, and antimicrobial activities (1–3). Compared with their non-prenylated precursors, these compounds usually demonstrate distinct and often improved biological and pharmacological activities, which makes them promising candidates for drug discovery and development (1, 2, 4–6). These compounds could be considered hybrid molecules of prenyl moieties of different chain lengths ($n \cdot C_5$, where n is an integer number) and aromatic skeletons originating from various biosynthetic pathways (7, 8). Prenyl transfer reactions (i.e. the connections of prenyl moieties to the aromatic nucleus) are catalyzed by a diverse family of prenyltransferases. Interestingly, this step usually represents the key transformation in the biosynthesis of such compounds. A prenyl moiety can be attached by prenyltransferases via its C1 (regular prenylation) or C3 (reverse prenylation) to carbon, oxygen, or nitrogen atoms of an acceptor (Fig. 1, A and B) (7, 8). Together with the observed regiospecific prenylations at different positions of an acceptor molecule, prenyltransferases contribute significantly to the structural and biological diversity of natural products (7).

Based on their amino acid sequences and biochemical and structural characteristics, prenyltransferases are categorized into different subgroups (7). In the last decade, significant progress has been achieved with the members of the dimethylallyltryptophan synthase (DMATS)[5] superfamily, and >40 enzymes of this group were identified and characterized by mining of fungal and bacterial genomes (7). These enzymes catalyze transfer reactions of a prenyl moiety from prenyl diphosphate (e.g. dimethylallyl diphosphate (DMAPP)) to diverse acceptors, such as tryptophan, tyrosine, tryptophan-containing cyclic dipeptides, xanthones, tricyclic or tetracyclic aromatic moieties, or even non-aromatic compounds. Among the acceptors, indole derivatives, including tryptophan and tryptophan-containing cyclic dipeptides, are substrates of most of the DMATS enzymes investigated so far (7, 9).

The DMATS enzymes have already been demonstrated to display high substrate and catalytic promiscuity. They not only cata-

[5] The abbreviations used are: DMATS, dimethylallyltryptophan synthase; DMAPP, dimethylallyl diphosphate; SMILES, simplified molecular input line entry system; SMARTS, SMILES arbitrary target specification; SMIRKS, a hybrid language of SMILES and SMARTS; ECFP, extended connectivity fingerprint; ESI, electrospray ionization; dma, dimethylallyl; HRMS, high resolution MS.
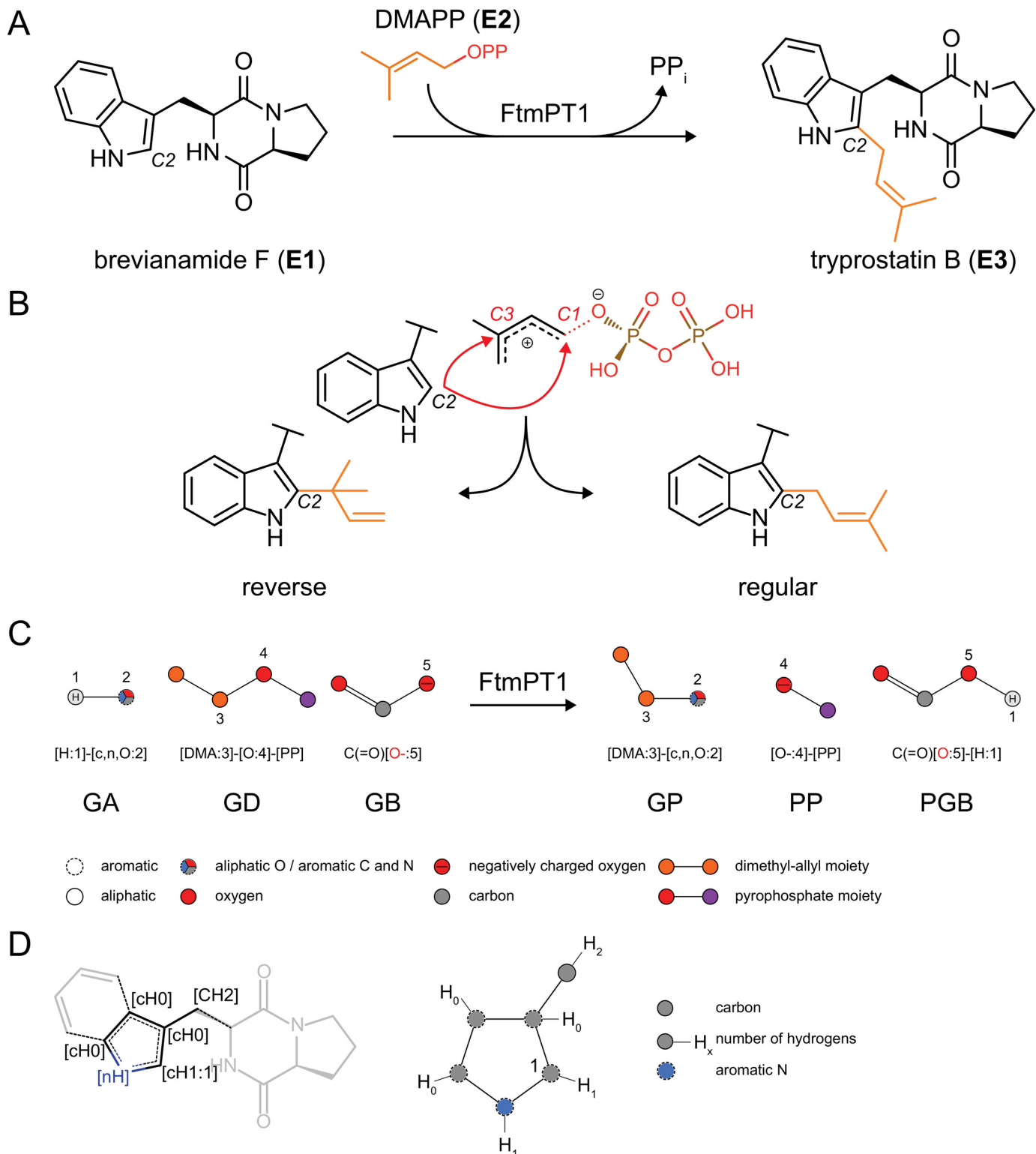
FIGURE 1. *A*, exemplary transformation of brevianamide F (E1) to tryprostatin B (E3). *B*, regiochemistry of the nucleophilic attack on the prenyl moiety. For regular prenylation, bond formation occurs between C2 and the carbon adjacent to the pyrophosphate group. An attack of C2 on the tertiary carbon of DMAPP leads to the reversely prenylated product. *C*, illustration of the SMIRKS-like notation derived from *A* (generated by SMARTSviewer) (22). *GA*, general prenyl moiety acceptor; *GD*, general prenyl moiety donor; *GB*, general base; *GP*, general prenylation product; *PP*, pyrophosphate; *PGB*, protonated general base. *D* (*left*), a reactive epitope indicated around the reactive atom ([cH1:1]) with the atomic properties given in SMARTS nomenclature (*brackets*). *D* (*right*), reactive epitope as generated by SMARTSviewer.

A

B
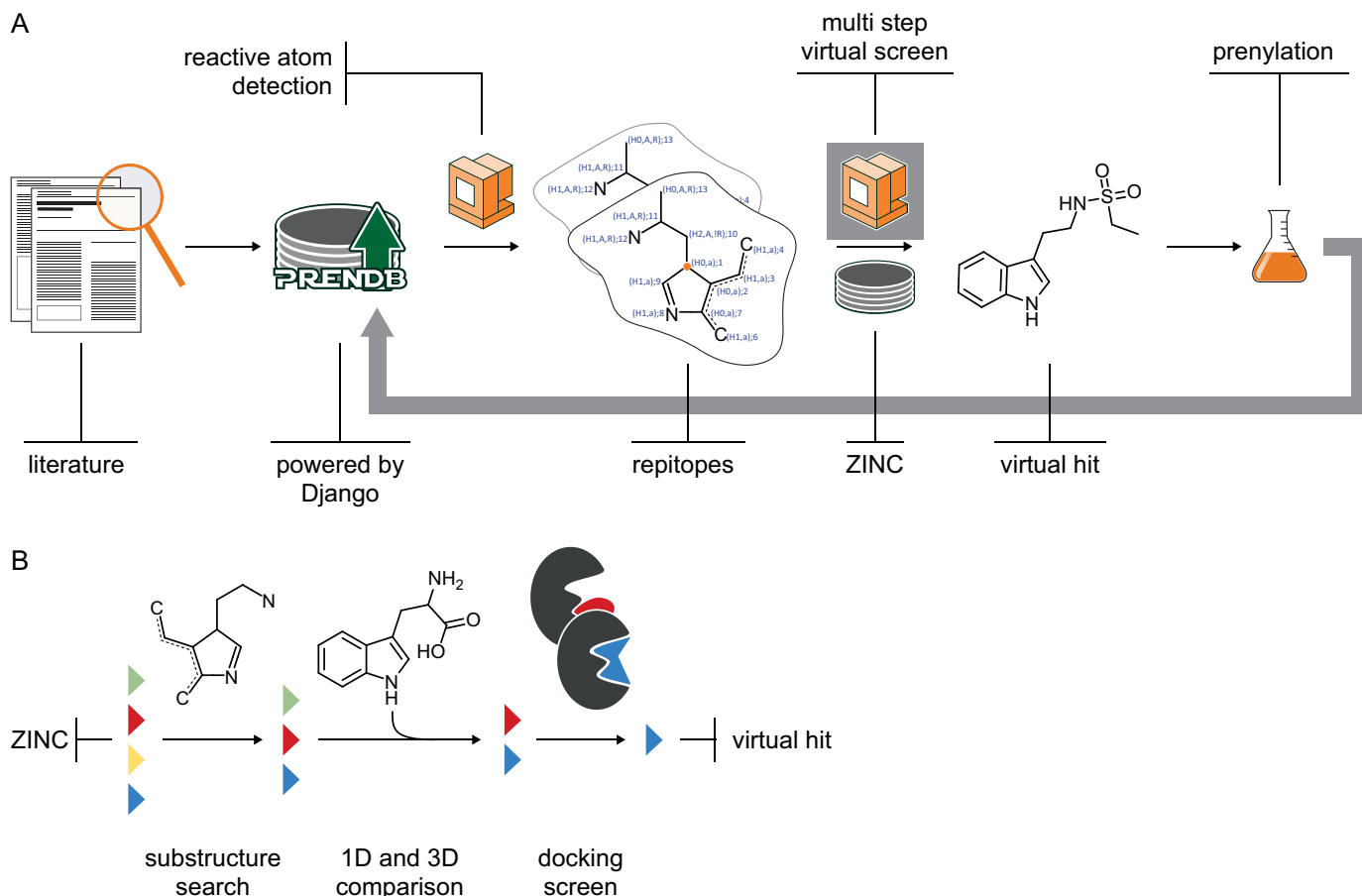
FIGURE 2. **Schematic of the underlying work flow.** *A*, reactions of prenyltransferases were digitalized and stored in PrenDB. The reactive atoms were detected algorithmically and reconstituted to reactive epitopes (repitopes). Prenylation candidates were selected by a multistep virtual screening approach, and their transformation potentials were evaluated experimentally on selected prenyltransferases. *B*, multistep virtual screening of database compounds (*colored triangles*): (i) repitope-based substructure search based on the substrate space covered in PrenDB; (ii) one-dimensional property and three-dimensional shape comparison with known substrates; and (iii) docking screen on three promiscuous prenyltransferases with available crystal structure.

lyze prenylation of their substrates and closely related compounds, but they also use structurally quite different compounds as prenyl acceptors (10). Therefore, these enzymes were successfully used for production of a large number of prenylated derivatives, including prenylated tryptophan and tyrosine analogs, tryptophan-containing peptides and derivatives thereof, hydroxyxanthones, hydroxynaphthalenes, flavonoids, indolocarbazoles, and acylphloroglucinols (10). For example, N1-, C4-, C5-, C6- and C7-prenylated tryptophan and N1-, C2-, C3-, C4- and C7-prenylated tryptophan-containing peptides and derivatives were obtained by using DMATS enzymes as biocatalysts (9, 10).

One of the challenges in discovery and use of DMATS enzymes as biocatalysts in a rational and targeted manner is the prediction of the acceptance of a putative substrate. On the one hand, the enzymes share similar structures, albeit often at low sequence identities, and catalyze, in many cases, similar reactions. On the other hand, different enzymes with similar natural substrates accept further non-native aromatic substances with clearly different activities (7). Therefore, bioinformatic and chemoinformatic approaches for the prediction of the catalytic activity of these enzymes are welcome and necessary to harness the full biosynthetic potential of this enzyme class.

We describe in this work the creation and evaluation of a database that catalogs and stores prenyltransferase reaction

information. Because storage of the reactions is automated, the database is not static but will grow with each new reaction described in the literature. Furthermore, we present an application of PrenDB, where we predict and validate putative substrates for prenyltransferases (Fig. 2).

Results

PrenDB Statistics—Digitalization and chemoinformatic encoding of enzymatic reactions of the DMATS superfamily allow for a deep analysis of their substrate space and reactivity toward distinct chemical epitopes (Fig. 3). In total, 32 unique enzymes were found throughout the literature examined. The three most prominent prenyltransferases in terms of the number of annotated transformations are 7-DMATS, FgaPT2, and SirD, accounting for 15, 14, and 13% of the reactions in the database, respectively. At the other end of the spectrum, there are seven enzymes for which only a single reaction has been published. With respect to promiscuity, the number of unique reactive epitopes–molecular substructures centered around the reactive atom and henceforth called repitopes in this work– was used as a descriptor (*cf.* Fig. 1*D* for an exemplary repitope). The enzymes 7-DMATS, FgaPT2, and AnaPT transfer prenyl moieties onto the broadest range of chemical epitopes. Together, these three enzymes contribute >65% of the repitope
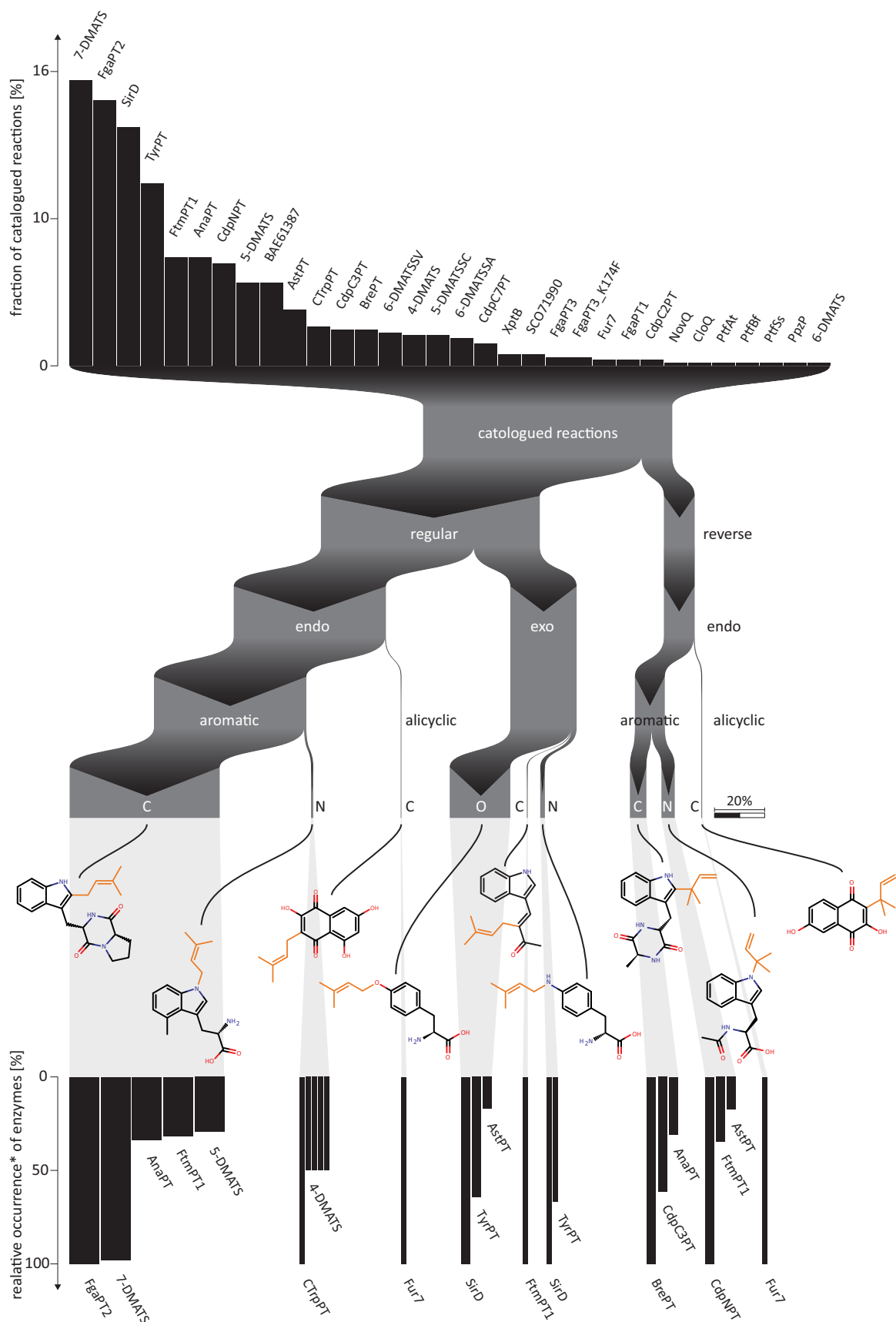
FIGURE 3. **Illustrative interpretation of the analysis of PrenDB.** Thirty-two enzymes contribute to the entirety of digitalized reactions (*top histogram*). The reactions can be further subdivided into categories based on the chemistry of the transformation (*middle*). Each subgroup corresponds to one or more enzymes (*bottom histogram*). *, the contribution of the enzyme with the highest contribution to a particular group was set to 100.
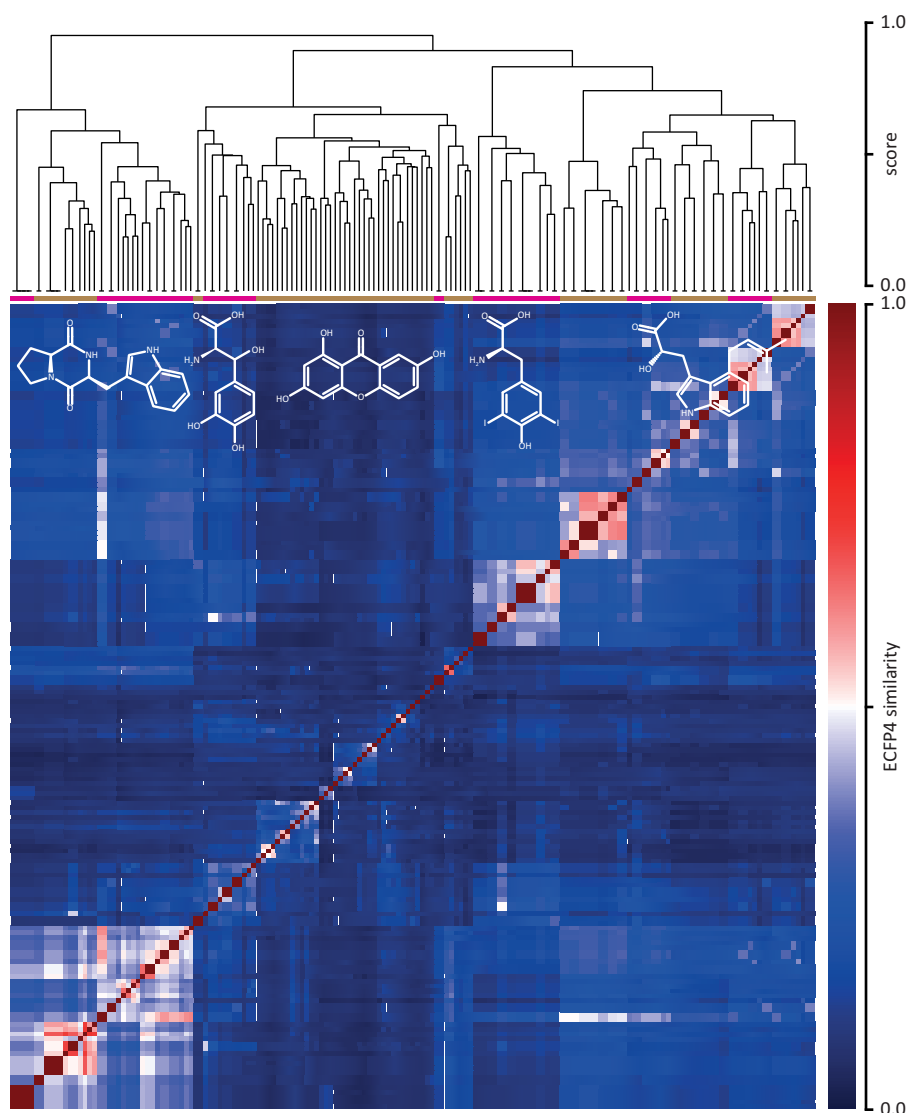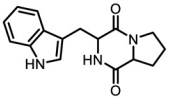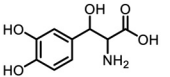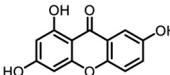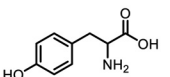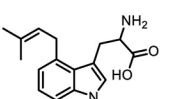
FIGURE 4. **Hierarchical clustering of the substrates extracted from PrenDB.** Individual clusters derived from the dendrograms are annotated with an image of the corresponding cluster representative. *Magenta* and *brown bars* indicate 14 detected clusters. *Black horizontal lines* on the leaves of the dendrogram indicate the number of molecules grouped together.

space. Knowledge of the reactive atom and its surroundings makes a further distinction of enzymatic transformations possible; the Sankey diagram in Fig. 3 shows how the cataloged reactions not only can be linked to their prenyltransferases, but also can be subdivided into types per the reactive atom. The clear majority of reactions (87%) corresponds to the regular type of prenyl moiety transfer (Fig. 1*B*, *right*), where the thermodynamically more stable regioisomer is formed. In 73% of all reactions and in all reverse attachments, the reactive atom is a member of a ring system (endo). Only a small part of regular prenylations occur at exocyclic atoms (exo, 26%). There, prenyl moieties are transferred onto oxygen and nitrogen atoms of tyrosine and aniline-like moieties by SirD, FtmPT2, or 4-DMATS. Reverse prenylation can be observed at carbon and nitrogen atoms only. They are incorporated in aromatic ring systems and less frequently also in alicyclic moieties, such as benzoquinones. More than 60% of all reactions occur at aromatic carbon atoms; derivatives of indole, including tryptophan, are the most frequent repitopes in this largest reaction

subclass. Much rarer are prenylations at nitrogen (8%) or at exocyclic oxygen atoms (23%). By comparison, 5% (33 entries) of reactions of tryptophan-like moieties (*e.g.* at atom position C2 in brevianamide F (E1) (Fig. 1*A*) lead to formation of compounds with a fused ring system, where the atomic environment of the reactive atom becomes dearomatized during the reaction (FtmPT1, Fur7, and 4-DMATS).

Substrate Space—Throughout the analyzed literature (44 articles from 17 journals), 167 unique substrates were found. To analyze the substrate space diversity of prenyltransferases, a similarity matrix based on the pairwise ECFP4 (11) fingerprint molecular similarity was calculated (Fig. 4, Table 1, and supplemental Table S1), followed by a hierarchical clustering. This allows the grouping of substrates based on their chemical structure. From the corresponding dendrograms and supported by the reorganized distance matrix, five substrate classes distributed over 14 clusters can be deduced: (i) unsubstituted indoles, derivatives of tryptophan and proline-tryptophan diketopiperazines; (ii) derivatives of tyrosine with modifica-

47

TABLE 1
Cluster size and substrate space coverage derived from hierarchical clustering

cluster representative	clusters	size	substrate space coverage [%]	description
	1, 2, 3	38	22	unsubstituted indole moieties and tryptophan-containing diketopiperazines
	4, 5	13	8	derivatives of tyrosine: modifications on benzene ring, homologues
	6, 7, 8	45	27	derivatives of xanthone, naphthalene, chinone and flavonoid
	9	18	11	tyrosine derivatives with modifications on benzene atoms
	10, 11, 12, 13, 14	53	32	side-chain-modified and multiply prenylated tryptophans

tions on benzene and aliphatic atoms; (iii) naphthalene, quinone, and flavonoid derivatives; (iv) side chain-modified tyrosines; (v) side chain-modified and multiply prenylated tryptophans. More than 50% of the substrate space is covered by indole-containing compounds. Molecules with tyrosine and flavonoid or xanthone motifs contribute 18 and 26% to the substrate space, respectively. Furthermore, the space spanned by the fragments obtained via bond cleavage during the fragment-based subgraph isomorphism perception process (*cf.* "Materials and Methods") is covered to an extent of 64% by tryptophan and diketopiperazine epitopes. This predominance of indoles can be explained by tryptophan and indole derivatives being the native substrates for 78% of the enzymes in PrenDB. This, combined with the DMATS bias in the literature, eventually leads to strongly indole-biased data.

Fig. 5 shows the knowledge about prenyltransferase reactions, as digitalized and stored within PrenDB, in terms of catalyzed transformations (combinations of a particular substrate and an enzyme) and the corresponding yield achieved. In the *top right corner* of the matrix, transformations of the most abundant substrates (tryptophan, tyrosine, diketopiperazines, and their derivatives, respectively) together with the most promiscuous enzymes (7-DMATS, FtmPT1, CdpNPT, SirD, and FgaPT2) can be found. At the same time, the matrix is sparse (*i.e.* contains a lot of blanks). This sparsity is the result of the availability of data and thus represents the research focus of the prenyltransferase field in the past. It presents a challenging starting situation for model building.

Repitopes—For each of the 665 cataloged reactions (each defined as a unique triplet of a substrate, product, and enzyme), a repitope (*i.e.* reactive epitope; see "Materials and Methods" for a complete definition) was extracted using the reactive atom detection and repitope reconstruction routines of the algorithm developed in this work (see "Materials and Methods"). Each

repitope comprises four reconstitution depths (from 2 to 5 bond distances). Over all repitope depths, 276 unique repitopes, defined by their unique SMARTS string, were extracted. A SMARTS string is a one-dimensional encoding of chemical substructures and an efficient way to store a complete definition of each repitope (see "Materials and Methods"). A reconstitution depth of 5 delivered the largest contribution to the repitope space with 135 (49%) members. This is consistent with expectations, because larger depths will lead to more diverse descriptions. Depths 2, 3, and 4 account for 25 (9%), 69 (25%), and 94 (34%) repitopes, respectively. Interestingly, the sum over all amounts of each reconstitution level is greater than the total number of unique repitopes. This means that distinct combinations of the substrate molecule, its reactive atom, and the depth level are not mutually exclusive, thus resulting in duplicate entries. Fig. 6 shows the distribution of repitopes for the various depth levels of reconstitution, underlining the relationship between repitope size and diversity. Although the largest class of repitopes contributes the most to repitope space, also smaller, more general and ambiguous repitopes are among the top 10 most frequent repitopes; substructures of tyrosine, benzene, and indole are at ranks 3, 5, 6, 7, and 9, respectively, whereas complete and extended indole and tyrosine structures are at ranks 1, 2, and 4 and, at 8, 10, and 11.

Prediction of Novel Substrates via a Multistep Screening Procedure—In a sequential application of virtual screening tools (Fig. 2), beginning with prenylation prediction through repitopes stored in PrenDB and concluding with docking into three prenyltransferases with a known crystal structure (FgaPT2, FtmPT1, and CdpNPT), 38 virtual hits were selected through the following procedure. (i) A compound was considered as a virtual hit if any PrenDB repitope could be found within its molecular framework at least once. Supplemental Table S2 shows the number of repitopes matching a particular hit, with a high repitope hit rate indicating promiscuous compounds (*i.e.* molecules that are classified as substrates of multiple enzymes). Using repitopes based on a reconstitution depth of 3, 168,906 compounds were selected in this first step. (ii) Comparison of molecular properties with those of known substrates and removal of molecules outside the respective ranges (Table 2) reduced the number of virtual hits to 90,559. Going beyond one- and two-dimensional molecular descriptors and ensuring that the (iii) three-dimensional shape (judged by a high score in the OEChem shape congruency tool; see "Materials and Methods") matched between putative and known substrates led to a selection of 451 compounds. This repitope-, property-, and shape-based determination of prenylation potential of the selected compounds was further refined by the (iv) docking results; for each compound, an optimal enzyme structure for docking was selected based on a compound's structural overlap with the co-crystallized substrate. The amount of this overlap was quantified by the same shape congruency methodology mentioned above but was automatically invoked from within the docking application HYBRID (see "Materials and Methods"). The generated poses, from which 38 molecules were selected for experimental validation, show a distinct geometrical consensus of the key interactions with the enzymes (Fig. 7): first, polar interactions
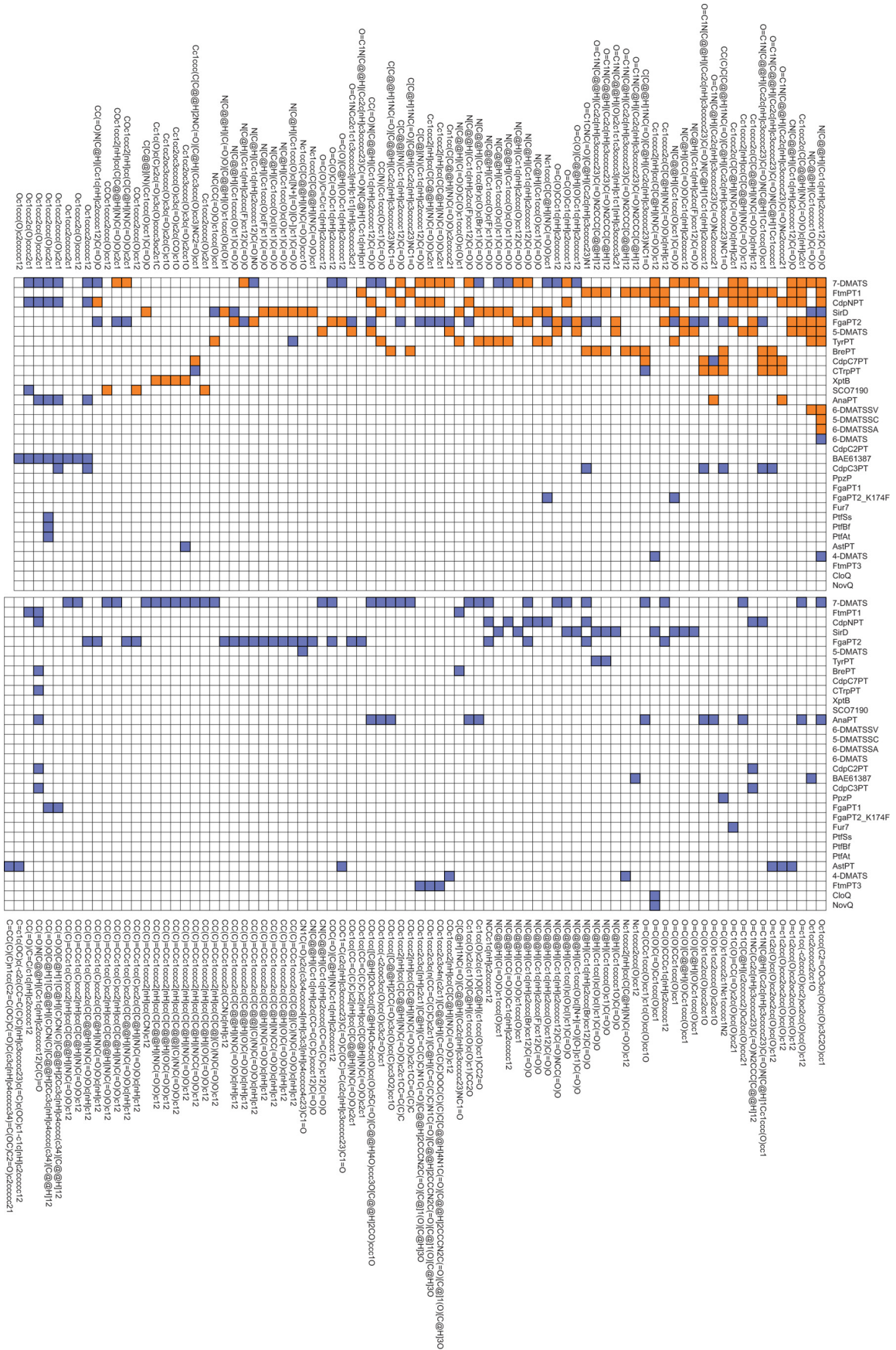
48

FIGURE 5. Coverage of enzyme-substrate space. *Orange squares* indicate transformation yields >10%, and *cornflower squares* show a yield <10% for the corresponding enzyme-substrate combination, respectively. *White squares*, absence of data.
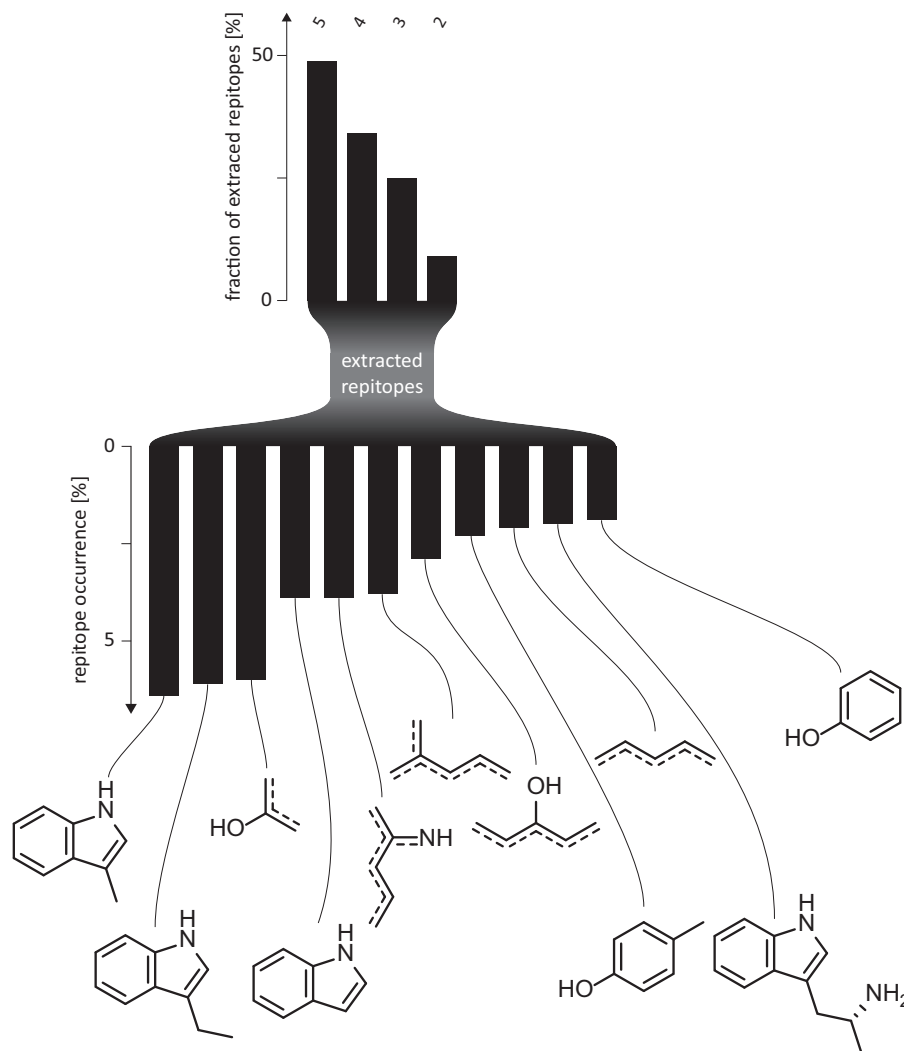
49

FIGURE 6. **Distribution of repitopes of different depth in repitope space.** *Dashed lines*, conjugated bonds.

TABLE 2

Range limits of physico-chemical properties derived from the substrate space stored in PrenDB

Physico-chemical property	Minimum	Maximum
Molecular mass (Da)	140	515
No. of heavy atoms	11	37
No. of carbon atoms	8	27
No. of heteroatoms	1	10
No. of chiral centers	0	5
Hydrogen bond acceptors	0	6
Hydrogen bond donors	1	6
No. of atoms in a ring system	6	25
No. of rotatable bonds	0	6
No. of rigid bonds	9	40
XLogP	−3.94	3.76
Minimal solubility attribute	Poorly[a]	
2D polar surface area (Å²)	20.0	1125
Removal of known aggregators	True[b]	

[a] Solubility categories (insoluble, poorly, moderately, soluble, very, highly) are derived from reparametrized atom types from the XLogP algorithm within the OEChem toolkit.

[b] An aggregator is considered an exact match with one of approximately 400 published aggregators compiled in the OEChem toolkit.

FIGURE 7. **Molecular features extracted from poses of the selected virtual hits.** *Left*, *red spheres*, hydrogen bond acceptors; *yellow discs*, aromatic moieties. The majority of acceptor functionalities can be found around the basic residues Arg-255 and His-279. *Right*, *blue spheres* indicate hydrogen bond donors. They are located around the highly conserved glutamate (Glu-89/102/116) and in the vicinity of backbone carbonyls (omitted for clarity).

with the general base Glu-89/102/116; second, occupation of the apolar indole-subpocket and hydrogen bond interactions in the vicinity of the opening of the active site, residues His-279 and Arg-244.

Novel Substrates for Prenyltransferases FgaPT2, FtmPT1, and CdpNPT—To assess the predictive performance of our virtual screening, the 38 potential substrates were tested as prenyl acceptors in enzyme assays with the tryptophan prenyltransferase FgaPT2 and the two tryptophan-containing cyclic dipeptide prenyltransferases FtmPT1 and CdpNPT. The selected substances clearly differ structurally from the substrates for the DMATS prenyltransferases reported previously (7, 10). The

50
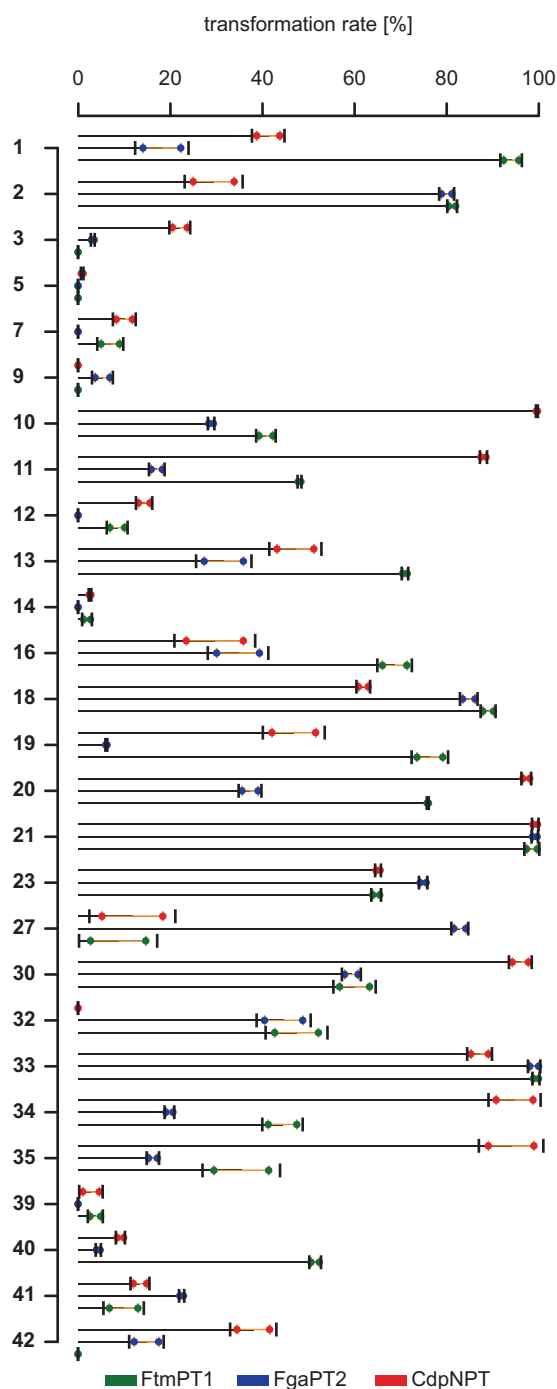
FIGURE 8. Transformation rates of virtual hits relative to L-tryptophan obtained for the three examined enzymes FtmPT1, CdpNPT, and FgaPT2. *Horizontal bars*, mean; *vertical bars*, S.D.; *orange interval*, S.E.; *colored circles*, data points.

reaction mixtures were analyzed with LC-MS to detect the formation of prenylated products. As shown in supplemental Table S2 and Fig. 8, 23 of these substances were accepted by FtmPT1, 22 by FgaPT2, and 25 by CdpNPT. In relation to the number of hits selected from our virtual screen, this corresponds to a hit rate of 60.5% for FtmPT1, 57.9% for FgaPT2, and 65.8% for CdpNPT. Product yields of >50% were observed for 12 substrates with FtmPT1, 7 with FgaPT2, and 10 with CdpNPT, respectively. The prenylated products can be

detected in a straightforward manner as signals in their corresponding mass spectra; their $[M + H]^+$ ions are shifted by 68 Da relative to their educts. Overall, we thus obtained high hit rates and yields >50% in the case of 29 reactions (25% of all attempted reactions).

Similarity Analysis of Known Substrates and Selected Compounds—To assess the novelty of the 38 selected compounds, the similarity with the substrate space cataloged within PrenDB (167 substrates) was calculated and visualized by generating a similarity matrix based on the ECFP4 fingerprint-based Tanimoto similarity (Fig. 9, *left*). The matrix shows an overall low similarity score between our selection and the known substrate space. This points toward the potential to access truly novel substrate space by employing repitopes. Of note, the similarity is higher in columns corresponding to compounds that were successfully prenylated in our assays by at least two of our test enzymes. The *right panel* of Fig. 9 shows similarity scores as calculated by our in-house fragment-based method RedFrag (12); in contrast to ECFP4, RedFrag compares the fragmental composition of molecules and the two-dimensional arrangement of fragments. RedFrag accentuates the commonalities and differences between known substrates and our selections. Compound **1**, a tryptophan-homoproline-diketopiperazine (94.1% yield on FtmPT1), shows high similarity scores with tryptophan, its indole core derivatives, and, expectedly, tryptophan-tryptophan-, tryptophan-alanine-, tryptophan-glycine-, and tryptophan-proline-diketopiperazines from clusters 1, 2, and 3, respectively. Of note, RedFrag emphasizes the similarity of compound **1** and cluster 3 based on the presence of the indole scaffold. In contrast, ECFP4 emphasizes the dissimilarity of this compound originating from the absence of the diketopiperazine motif in the same cluster. Compounds **12**, **26**, and **32** show high similarity with substrates from cluster 3 (also 9, 10, and 11). Compounds **12** and **32** are regioisomers of a brominated tryptophan derivative. They show distinctly different yields: 14.4 and 8.5% for **12** in CdpNPT and FtmPT1; 47.5 and 44.7% for **32** in FtmPT1 and FgaPT2, respectively. It is evident that the position of the bromine atom has a major impact on the role of such compounds as substrates. The influence of regiochemistry of indole core substitutions or single-atom replacements at this core is further exemplified by compound **26**. Its benzothiophene moiety (replacing the nitrogen atom in an indole by a sulfur atom) is not accepted as a substrate by any of the three test enzymes.

Selected compounds with low similarity but remarkable yields indicate novel substrate classes or motifs; compound **16** shows a good yield in FtmPT1 and moderate yields in FgaPT2 and CdpNPT (68.8, 34.8, and 29.7%, respectively). Its conjugated indole-4-imidazolin-2-one motif has no similar counterparts within the known substrate space. This is also true for compound **30** (yields of 60.1, 59.4, and 96.1%, respectively) and its benzylated hydroxyl-indole structure. Compound **27**, a pyrimidine-indole, shows excellent yield in FgaPT2 (82.9%). Further examples with high yields but RedFrag similarity scores <0.6 are compounds **11**, **13**, **20**, **23**, and **33**.

Structure Elucidation of the Products of Compounds 21 and 30—To investigate at which position within a given substrate the prenylation occurred, we carried out exemplary FtmPT1
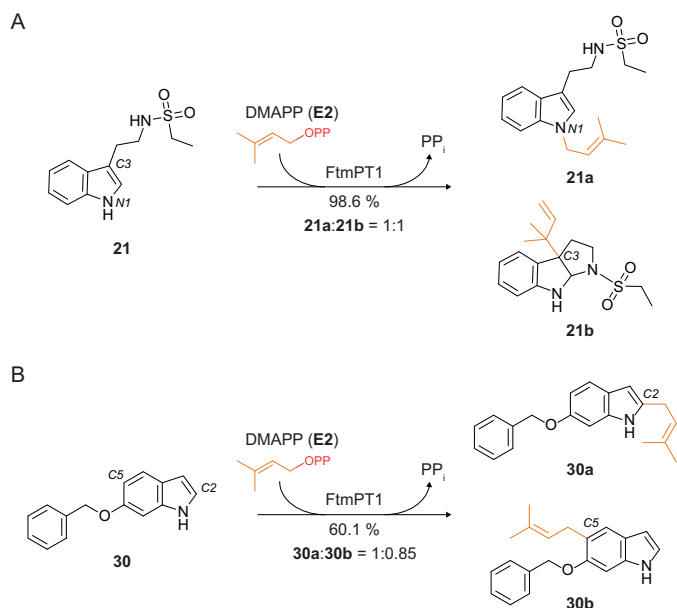
A

B
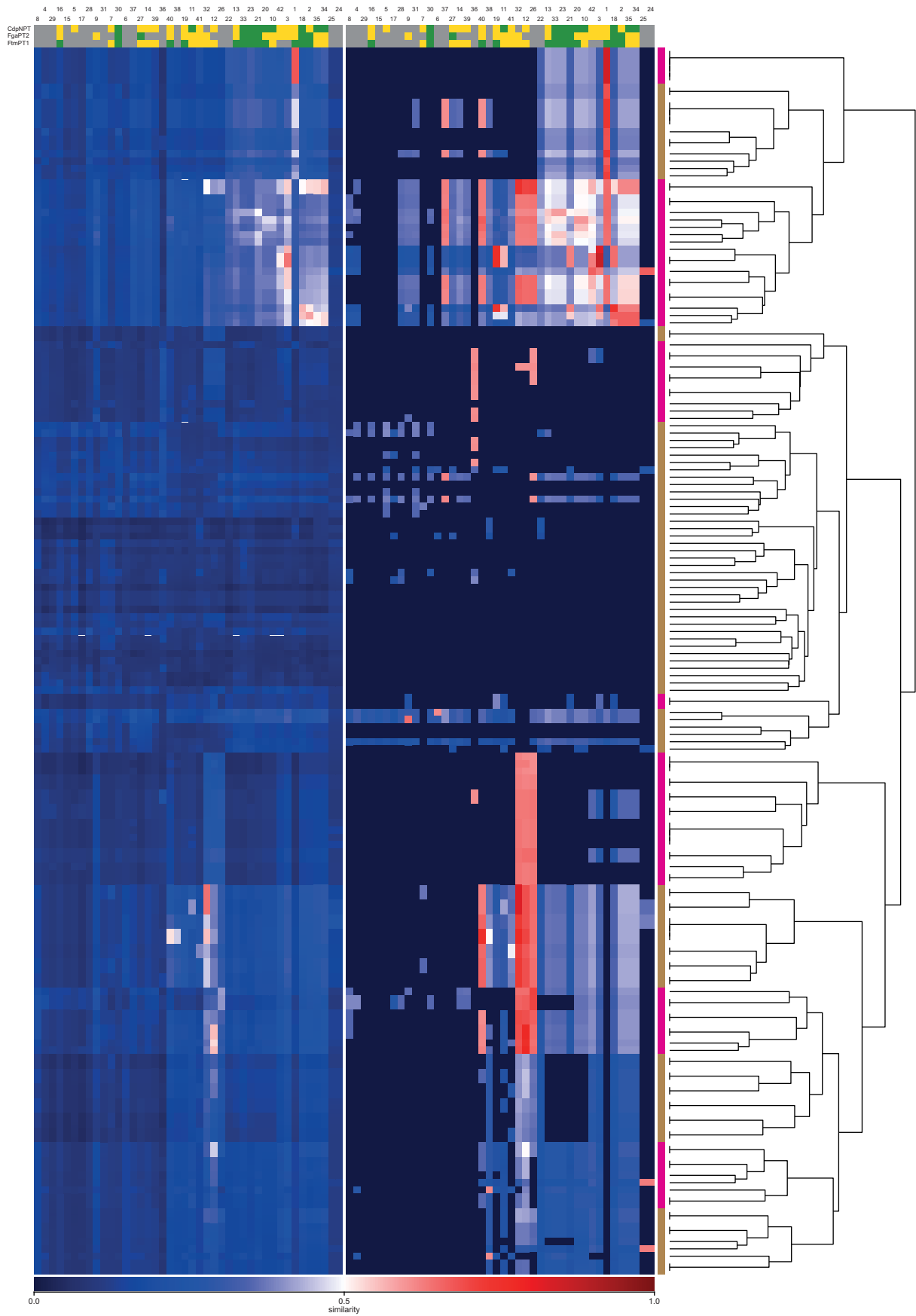
FIGURE 10. **Transformations of virtual hits 21 and 30 by FtmPT1.** *A*, indole 3-ethyl ethylsulfonamide (**21**) was regularly prenylated at position N1, leading to **21a**, and reversely prenylated at position C3 with a simultaneous formation of a 6/5/5 fused ring system (**21b**). Typically for C3-prenylation at indole substructures, a dearomatization and intramolecular cyclization accompany the prenylation reaction. *B*, 6-benzyloxyindole (**30**) was regularly prenylated at positions C2 and C5.

incubations with two indole derivatives, indole 3-ethyl ethylsulfonamide (**21**) and 6-benzyloxyindole (**30**), which were very well (98.6% yield) and moderately (60.1% yield) accepted by this enzyme, respectively (Fig. 10 and supplemental Table S2). As shown in Fig. 11*A*, a single dominant peak was observed in the LC-MS chromatograms of the incubation mixtures, which were isolated on a Multospher 120 RP-18 column for structure elucidation. ^{1}H NMR data revealed, surprisingly, the presence of two compounds in each reaction mixture. **21a** and **21b** originated from **21** in a ratio of 1:1 from the reaction mixture, and **30a** and **30b** originated in a ratio of 1:0.85 from **30**. Further purification resulted in four pure products. Through NMR and MS analyses, the structure of **21a** was subsequently elucidated as a regularly N^1-prenylated derivative. The second product, **21b,** was identified as a reversely C^3-prenylated derivative with a simultaneous cyclization of C2 of the indole with the nitrogen atom of the side chain located at C^3, resulting in the formation of a 6/5/5 fused ring system. Compounds **30a** and **30b** were proven to be regularly C^2- and C^5-prenylated derivatives, respectively. These results unequivocally proved specific prenylations at the indole ring without or with additional modifications, such as cyclization. Detailed studies of the relationships of enzymes, substrates listed in supplemental Table S2, and their products are under further investigation.

A comparison of the elucidated structures of the products **21**, **21b**, **30a**, and **30b** with the PrenDB-predicted prenylation sites of their corresponding educts reveals that the prenylation site was correctly predicted in two of four cases. However, the responsible enzyme, FtmPT1, was only proposed for the prenylation of **21** to yield **21a**. In the case of **30**, the product **30b** was predicted to originate from the enzyme CdpNPT or FgaPT2.

Discussion

This study demonstrates the power of systematically organizing and analyzing diverse and disparate experimental enzymatic data by means of chemoinformatic methods. Besides a comprehensive repository of the existing knowledge about prenyltransferase reactions, the determination of repitopes allowed us to predict novel substrates that are distinctly different from the ones that have been identified previously, both natural and synthetic. Moreover, we achieved an overall high hit rate of 71% in terms of molecules that were accepted by at least one prenyltransferase. However, it has to be noted that the repitopes stored in PrenDB are not yet accurate enough in all cases to precisely predict the correct enzyme and/or the correct reactive atom. This shortcoming is presumably correlated with the comparatively small number of instances in the database. Although the existing body of literature clearly represents a considerable experimental effort, chemistry and the biochemical reactivity of enzymes are so diverse that even higher numbers of substrate-enzyme-product triplets would be necessary to obtain more complete repitopes that also account for the different reactivity of certain substructures. The chemoinformatic strategy that we employed in this work is certainly flexible enough to accurately model more fine-grained patterns.

At the same time, a database such as PrenDB can provide excellent help in determining which reactions and substrates would be worthwhile to test next. On a basic level, one could simply be guided by the number of reactions already described for each enzyme and focus on the underrepresented ones. But also more sophisticated approaches can be envisioned; enzyme phylogenetic trees could be based not on amino acid sequence, but on substrate similarity. Further exploration would thus focus on filling in the missing links. Ultimately, such strategies might merge with machine learning approaches, where the algorithm itself would suggest which enzyme-substrate pairs to test next based on the maximum information gain of each investigation.

Nevertheless, despite some shortcomings, our database and resulting prediction algorithm are already useful for correctly predicting a large number of substrates and thereby aiding in the creation of novel chemical matter. The imprecisions could also be taken as a strength in this context, in that they allow for serendipitous discoveries (*e.g.* the reverse prenylation of **21** to **21b** by FtmPT1).

Last, it has to be emphasized that the concept of repitopes and their fragment-based determination can easily be extended to other enzymatic reactions. The automatic processing of potentially large numbers of reactions and the concomitant conversion into the reaction principles (*i.e.* repitopes) will lead

FIGURE 9. **Similarity matrix between the selected compounds and known substrates for prenyltransferases extracted from PrenDB.** *Left*, ECFP4 fingerprint similarity. *Middle*, RedFrag scores calculated with ECFP4 fingerprints. *Color coding (top)*, *green*, yield >50%; *yellow*, yield between 1 and 50%; *gray*, no transformation. *Right*, *magenta* and *brown bars* indicate 14 detected clusters. *Black vertical lines* on the leaves of the dendrogram indicate the number of molecules grouped together.
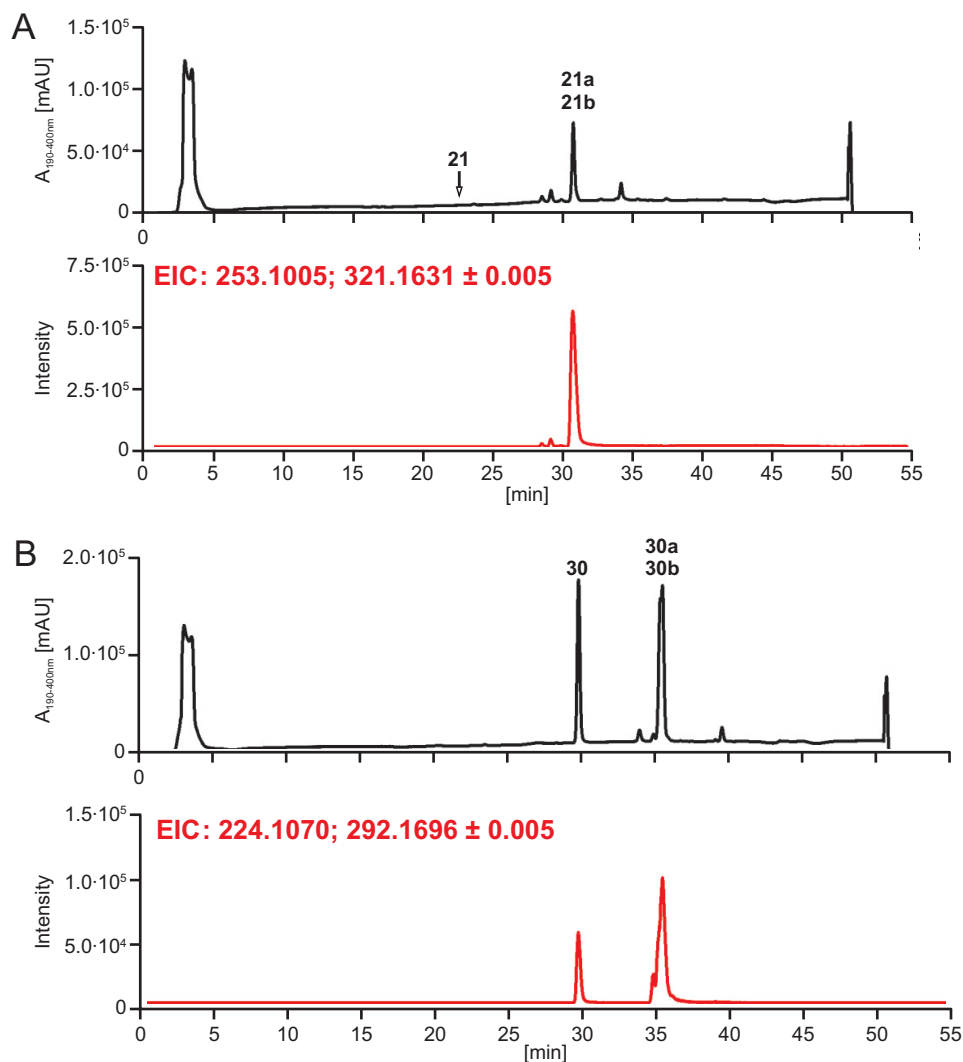
FIGURE 11. **LC-MS analysis of the reaction mixtures and prenyl transfer reactions of FtmPT1 with 21 (*A*) and 30 (*B*).** *EIC*, extracted ion chromatogram; *mAU*, milliabsorbance unit.

to facile systematizations and gain of knowledge from the analyses of the emerging data.

The high hit rates (58–66%) for each enzyme and the fact that one-fourth of the reactions had a yield of ≥50% demonstrate the excellent performance of our knowledge-based repitope approach. The combination of PrenDB and its ligand-based approach with protein structure-based tools, such as docking, therefore seems to constitute a powerful combination of strategies. Furthermore, these results prove the potential usefulness of the tested enzymes for the production of prenylated derivatives.

Materials and Methods

The prenylation reaction as conducted by the enzymes of the DMATS superfamily formally corresponds to a substitution reaction occurring on carbon, oxygen, and nitrogen atoms of small metabolites through the transfer of small apolar moieties (denoted as dma (short for <u>di</u><u>m</u><u>ethyl</u><u>a</u><u>llyl</u>) in the following example). The leaving group is always a pyrophosphate (PP_i) and a formal proton accepted by a general base. The reaction can be written in a symbolic way (SMIRKS notation, Daylight

Chemical Information System, Inc. website, accessed March 16, 2016; see below); the atoms taking part in the chemical transformation are arranged in a one-line notation showing the bond cleavages and formations.

$$[C,O,N:1][H:2] + dma[C:3][O:4]PP_i + [gB:5]$$

$$\rightarrow$$

$$[C,O,N:1][C:3]dma + [O:4]PP_i + [gB:5][H:2]$$
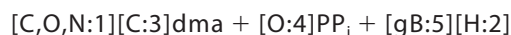
REACTION 1

Square brackets enclose individual atoms, and adjacent atoms are taken to be linked by covalent bonds. Letters denote elements, gB is the general base, and PP_i is pyrophosphate; commas represent a logical OR; and numbers are arbitrary labels to allow for unambiguous tracking of each atom. In the above example, it can be seen that the hydrogen atom with label 2 ([H:2]) is substituted by the dma group and moves from its adjacent carbon, oxygen, or nitrogen atom (labeled 1) to the general base (label 5). Fig. 1*A* illustrates this general transfor-

54

mation in a two-dimensional way exemplarily for the reaction between brevianamide F (**E1**) and DMAPP (**E2**) catalyzed by FtmPT1. With this symbolic notation and common chemoinformatic tools in hand, it is possible to virtually transform, for example, any carbon atom [C:1] bearing a hydrogen atom [H:2] (*GA* in Fig. 1*C*) into the prenylated product (*GP*), at the same time generating a protonated general base (*PGB*) and pyrophosphate (*PP*) as by-products. Although feasible *in silico*, a chemical transformation based solely on the reactive atom is unreasonable and ambiguous; in reality, only carbon, nitrogen, and oxygen atoms located within the correct atomic surroundings can undergo prenylation. Thus, the entire molecule, or at least a crucial motif within it, is necessary to completely characterize a reactive environment. We call such a set of atoms consisting of the reactive atom, to which the transferred moiety will be attached, and its neighboring atoms a "reactive epitope" (repitope for short). In the case of the transformation of **E1**, the corresponding repitope is shown in Fig. 1*D*. The specification of the carbon atom can now be extended to its full repitope notation,

$$[cH0][nH][\textbf{cH1:1}][cH0]([CH2])[cH0]$$

SCHEME 1

where lowercase letters denote membership of an atom in an aromatic system, HX indicates the presence of X adjacent hydrogen atoms, and parentheses indicate branching of the molecular framework. From this notation, it can be concluded that the reactive atom is aromatic and is bound to one hydrogen atom; its direct neighbors are an aromatic nitrogen and another aromatic carbon without any attached hydrogen atoms. The second neighbor shell consists of two aromatic carbon atoms and one aliphatic carbon atom. This convenient one-line notation of chemical environments is called SMARTS (one-line molecular patterns; Daylight Chemical Information System, Inc. website; accessed March 16, 2016) and is widely used in the field of chemoinformatics, especially for substructure searches. Atoms, their properties, and binding characteristics are encoded with alphanumeric characters. Multiple molecule SMARTS together with the information about bond breakage and formation yield the SMIRKS of a reaction. With a repitope, such as the one described above, and the enzymatic transformation encoded in SMIRKS notation, it is possible to virtually transform any substrate molecule into its corresponding prenylation product or to easily search for putative substrates by invoking substructure-based virtual screens in publicly available vendor databases.

The corresponding SMARTS notation for each repitope could in principle be deduced by hand, given the chemical structures of the substrate and product molecules. To achieve an efficient handling of several hundred enzymatic transformations, with only the two-dimensional structures of substrate, product, and the transferred moiety as input, an automated procedure for the extraction of transformation SMARTS, and thus repitopes, appears to be as indispensable as it is difficult to accomplish. A fully specified repitope requires knowledge of the reactive atom as well as its surroundings. Repitope deduction can be accomplished by applying subgraph isomorphism-based algorithms followed by the reconstitution of the chemical environment. Both steps (reactive atom perception and repitope reconstitution) will be described in detail below. All coding was done in python. For chemoinformatic calculations, the python wrappers of the RDKit (open source toolkit for chemoinformatics; accessed March 16, 2016) library were utilized. Fingerprint-based similarity calculation was carried out with the OEChem toolkit (OpenEye Scientific Software, Inc., Santa Fe, NM).

Perception of the Reactive Atom

In the case of a simple linear substitution reaction, as depicted in Fig. 1*A*, the reactive atom can be found by mapping the molecular structure of the substrate molecule onto the molecular structure of the product. For non-symmetric molecules, this leads to a unique match with an atom-to-atom correspondence between substrate and product. Because the number of atoms in the product is always greater than the number in the substrate, the substrate is a substructure of the product (*i.e.* its complete molecular skeleton can be found within that of the product). With the same approach, the atom-to-atom correspondence between the transferred moiety (*i.e.* the prenyl group) and the product molecule can be obtained. The intersection of the atom-to-atom matched sets of the substrate and the transferred moiety consists of only one atom, the reactive atom (Fig. 12*A*). If, however, the enzymatic transfer of a moiety is accompanied by a subsequent (or concerted) rearrangement of the molecular skeleton of the product (*e.g.* a cyclization), the substrate cannot be considered to be a direct substructure of the product anymore. Thus, the reactive atom can no longer be determined through the atom-wise substrate-to-product mapping as described above. In such a case, a possible strategy for establishing a substructure correspondence (*i.e.* a subgraph isomorphism) would be to weaken atom or bond type matching criteria. The resulting atom-to-atom correspondences are allowed to be more general in that way but are often ambiguous at best. To circumvent this problem, we assumed that, although the entire substrate may undergo dramatic changes in its molecular skeleton, smaller structural motifs (molecular fragments) remain unaffected by such transformations and can therefore still be unambiguously mapped onto the substrate structure *before* and *after* the reaction. Fig. 12*B* illustrates the consecutive steps in this *fragment-based* substructure isomorphism approach. In contrast to the aforementioned subgraph isomorphism based on the entire substrate structure, an additional fragmentation step has to be performed. Breakable bonds (bonds connecting ring systems with other ring systems or with acyclic motifs (see Fig. 13 (*A* and *B*) for exemplary fragmentations and the breakable bond definition)) are cleaved, leading to a set of fragments. An intermediate filter step ensures that very small fragments (single atoms, linker moieties, terminal groups) are not considered further. The remaining fragments are mapped onto the product structure, and their atom-to-atom correspondence is investigated for an intersection with the atom mappings of the transferred moiety and the product molecule. By preserving the atom-to-atom correspondences between the substrate molecule and its fragments, the reactive atom can be identified by finding the intersecting atom of one of the matching fragments in the substrate.
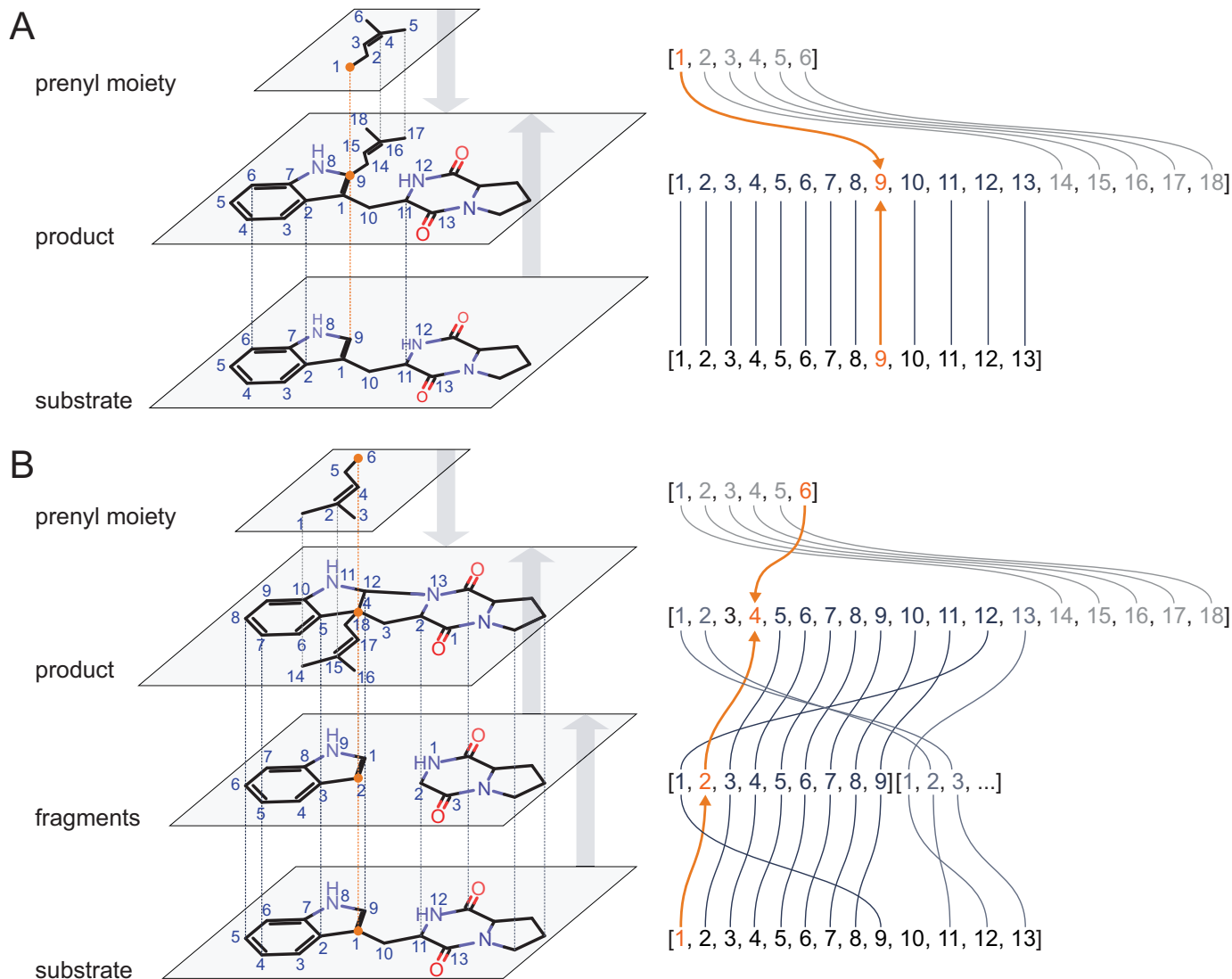
PrenDB, a Substrate Prediction Database

FIGURE 12. *A*, substrate-based subgraph isomorphism. The substrate structure matches the product as a whole. The intersection of atom overlaps between substrate and prenyl moiety delivers the reactive atom (*orange arrows*, index 9). *B*, fragment-based subgraph isomorphism. Substrate structure is fragmented into smaller epitopes preserving the substrate-fragment atom matchings. By matching fragments onto the product and analyzing the intersection with the prenyl moiety, the reactive atom can be found within the structure of the substrate (*orange arrows*, index 1).

Reconstitution of the Reactive Epitope

As already mentioned, knowledge of the reactive atom alone is only of limited use for substructure searches or virtual transformations, because both methods yield ambiguous results when only a single atom is given as input. It is therefore necessary to rebuild the chemical environment of the reactive atom to obtain a description of a particular transformation that has discriminative power. To obtain such a description, the reactive atom is augmented with additional atoms from its first, second, third, (etc.) neighbor shells (Fig. 14) (*i.e.* by traversing the atomic neighborhood of the reactive atom up to a fixed distance (*i.e.* number of bonds)). The traversed atoms are then extracted as a molecular subset and converted into a regular molecular object and, eventually, a SMARTS string. Different depths of reconstitution lead to either small and nonspecific repitopes ($d = 1$) or larger and more stringent ones for large depths ($d > 3$). *In extremis*, at the largest possible depth, the repitope becomes identical to the molecule itself. Thus, a balance has to be found, and the

most useful repitopes are able to represent reasonable chemical environments for a particular reaction but still allow for certain flexibility and diversity in retrieving putative substrates.

Database of Prenylation Reactions

With the algorithmic tools to deconstruct a given transformation catalyzed by a prenyltransferase into a reaction SMARTS and the corresponding repitopes in hand, investigation of as many transformations as possible can readily be conducted. Hence, we decided to create a database (PrenDB) storing the known transformations in an efficiently browsable and queryable manner. For this purpose, a literature search was performed to extract substrates, products, enzymes, and available metadata (such as kinetics and yields) from 44 publications (full articles, reviews, and communications) across 17 journals. Each enzymatic reaction is represented by the SMILES strings (13) of the product and substrate molecules, combined with the preferred name of the involved enzyme. The advantage of using

56

SMARTS is that each reaction can be visualized and processed with common chemoinformatic software. Furthermore, each reaction entry contains multiple repitopes, generated with the aforementioned algorithm and different environmental depths (2–5 bonds around the detected reactive atom). This reaction table (a table is a collection of database entries that are semantically equal) is supported by and connected with further tables

holding metadata extracted from the literature and/or calculated with chemoinformatic tools (Fig. 15). The molecule dictionary table comprises all small molecules involved in the reaction: substrates, products, transferred moieties (such as DMAPP or benzylpyrophosphate), and fragments. Additionally, each entry comes with a molecular properties table, where basic physico-chemical properties can be looked up. The reference table contains the literature used for data extraction together with hyperlinks to articles and entries on PubMed and UniProtKB. PrenDB can be browsed and extended with python scripts bundled with the algorithms for repitope generation described in this work (or more conveniently via a web interface) in a straightforward manner. Because of access speed and portability considerations, we decided to use the sqlite3 backend as the underlying database architecture and the Django python package for middleware and frontend.
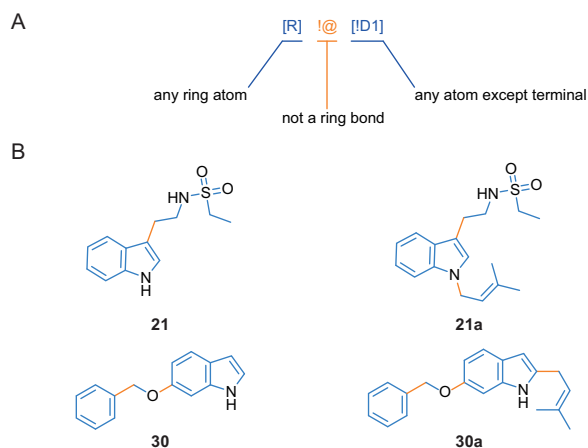
FIGURE 13. *A*, fragmentation rule expressed as SMARTS string. The rule consists of two atom definitions and a bond definition. Enclosed in *square brackets* (*blue*) are atoms connected by any bond type except for a ring bond (*orange*). The atom on the *left* can be of any type but must be a member of a ring system (hydrogen atoms are excluded indirectly because they are not allowed to form ring systems). On the *right*, the atom must not be a terminal atom (hydrogen atoms are excluded indirectly because they are always terminal). *B*, breakable bonds (*orange*) as defined by the fragmentation rule and the resulting fragments (*blue*) based on the substrates **21** and **30** and their corresponding prenylation products **21a** and **30a**.

Virtual Screen for Putative Substrates of Prenyltransferases

To predict novel substrates for transformation by a prenyltransferase, a multistep screening process was carried out with a subset of the ZINC database (14), which stores commercially available small molecules in ready-to-be-processed formats. First, the ZINC clean leads database, with a total of 5.1 million compounds, was filtered for the presence of any of the extracted repitopes from PrenDB. The repitope depth was 3. The screening was carried out as substructure searches using the python wrappers of the OEChem toolkit. Second, compounds were filtered utilizing the MolProp toolkit (OpenEye Scientific Software). Only compounds with physico-chemical properties within the range spanned by known substrates cataloged in
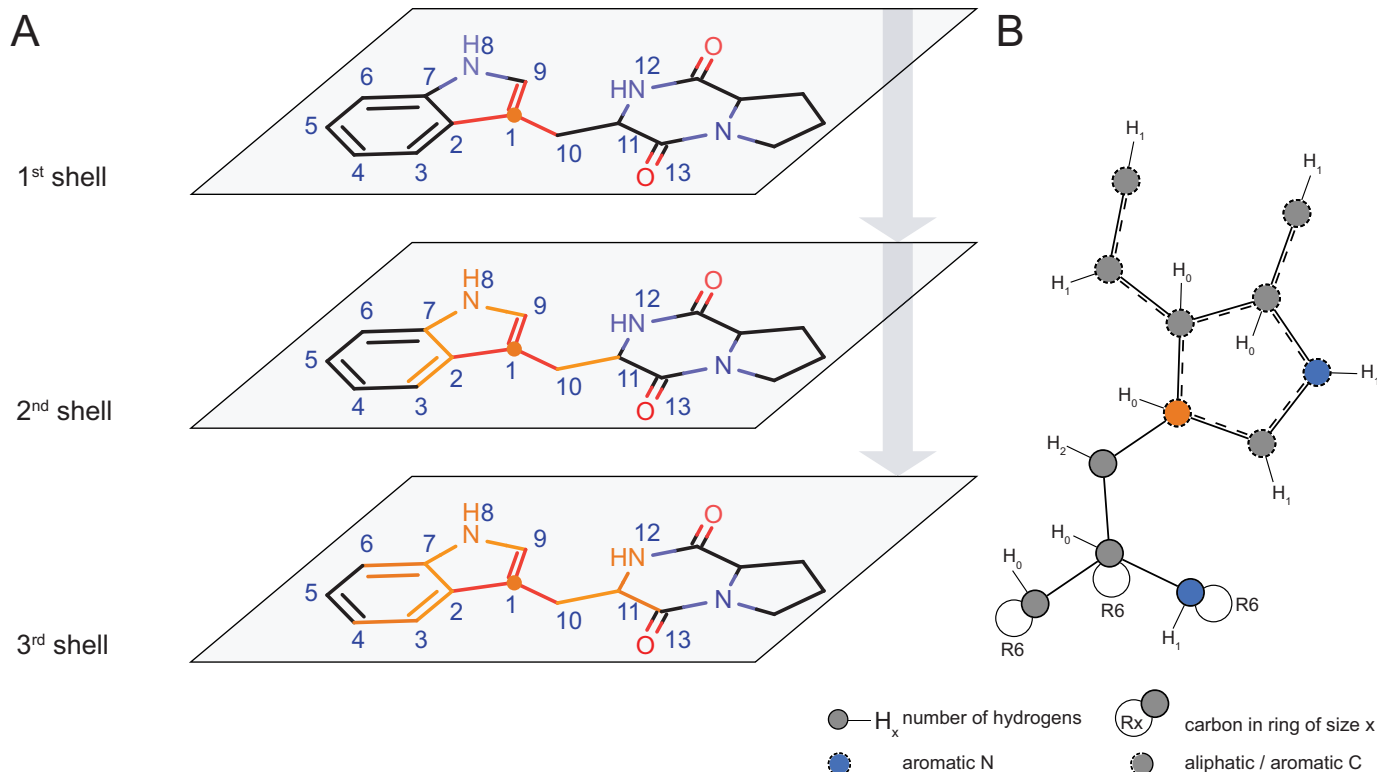
FIGURE 14. *A*, a repitope is generated by sequentially rebuilding the substrate molecule shell by shell with the reactive atom (index 1) as anchor point. Each iteration adds another neighbor shell to the repitope, resulting in a fully defined depth-3 repitope as depicted by SMARTSviewer (*B*).

PrenDB, a Substrate Prediction Database
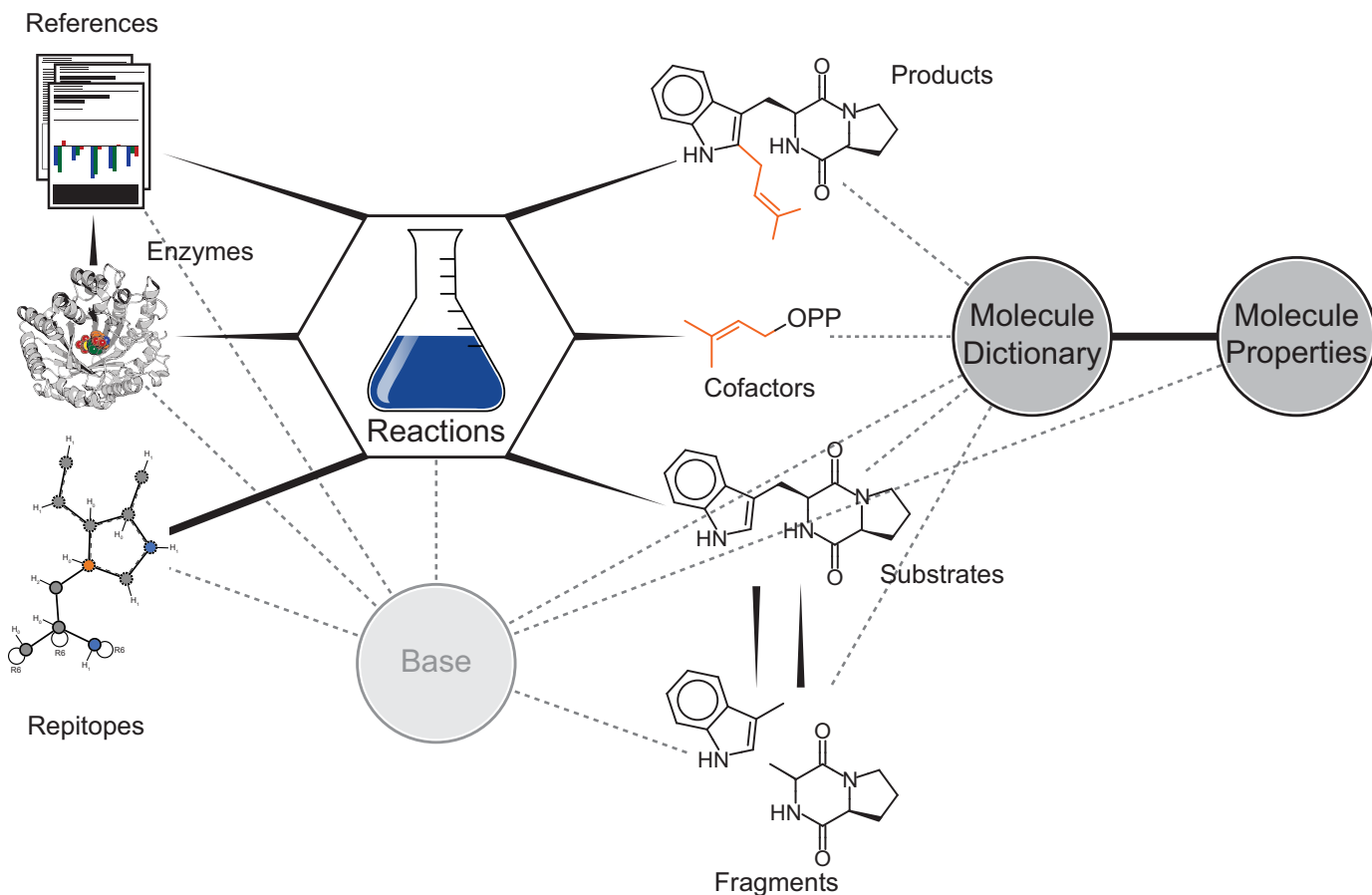
FIGURE 15. **Design of PrenDB.** The database tables are related to each other in a one-to-one (reactions and repitopes), one-to-many (substrates and reactions), or many-to-many (substrates and fragments) relationship, reflecting their real world correspondence. The central reaction dictionary holds the necessary data to encode a reaction based on substrate, product, and cofactor molecules, the enzyme, and the resulting repitope. A reference table is added to supplement the database with metadata and enhance its usability. *Dashed lines*, abstract inheritance; *solid wedged lines*, one-to-many relationships (*e.g.* a molecule can act as substrate in as many reactions as an enzyme). A repitope belonging solely to one particular reaction is indicated by a *straight solid line* (one-to-one relationship).

PrenDB were allowed (Table 2). The remaining compounds were subsequently submitted to the shape congruency analysis based on the OEChem API. In short, for each compound, a low energy conformer was generated, and its 3-dimensional overlay with each known substrate was optimized. Compounds with an overlay score >0.9, reflecting excellent three-dimensional shape matching, were allowed for the next step. Fourth, the remaining compounds were docked into the three most promiscuous prenyltransferases for which a crystal structure had been determined (FgaPT2, FtmPT1, and CdpNPT; Protein Data Bank codes 3I4X, 3O2K, and 4E0U, respectively), employing the multi-target HYBRID (15) engine; for each compound, up to 200 conformers were generated with OMEGA (16). The ensemble of conformations of each molecule was then overlaid with the co-crystallized ligand in each of the three selected crystal structures to determine the best suited enzyme for the following exhaustive docking. The method for overlaying conformers is built directly into the HYBRID engine and is based on the same methodology as implemented in the OEChem API and the ROCS application (17). For the actual docking step (translational and rotational optimization of a compound conformer within the binding site of the protein), HYBRID scores for a given protein-ligand complex were calculated based on the shape and electrostatic complementa-

rity of the ligand and protein's binding site (Fig. 16). Shape and electrostatic features are represented by Gaussian potentials. During optimization, the overlap between ligand and protein features is maximized. After docking, calculated poses were visually inspected to remove those that form improbable interactions that are not sufficiently penalized by present-day scoring functions, and the selected compounds were acquired from their respective vendors and experimentally tested.

Experimental Validation

Chemicals, Bacterial Strains, and Culture Conditions—DMAPP was synthesized according to the method described for geranyl diphosphate reported previously (18). The 38 tested substrates were purchased from Enamine Ltd. (Kiev, Ukraine), Chem-Bridge Corp. (San Diego, CA), MolPort (Riga, Latvia), Vitas-M Ltd. (Apeldoorn, Netherlands), and Mcule, Inc. (Budapest, Hungary).

Escherichia coli strains XL1 Blue MRF′ (Stratagene, Heidelberg, Germany) and *E. coli* BL21 (DE3) (Invitrogen, Karlsruhe, Germany) were used for protein overproduction. The strains with expression plasmids were cultivated in lysogeny broth or Terrific broth medium at 37 °C with 50 μg·ml^{-1} carbenicillin or 25 μg·ml^{-1} kanamycin as selection marker. Overproduction of
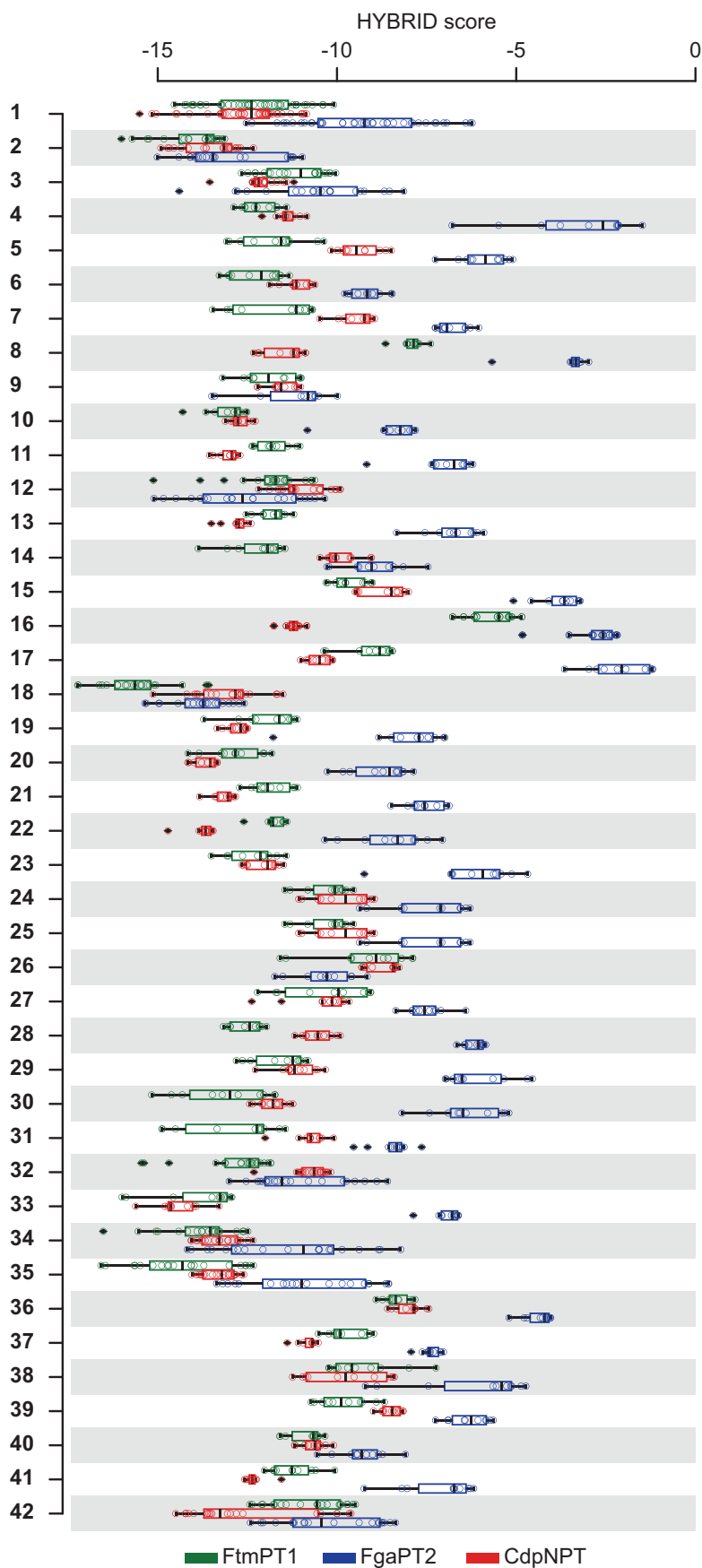
FIGURE 16. **Distribution of docking scores for each of the 42 initially selected poses.** Each distribution (represented as a *box plot*) shows 10 scores (*circles*)/compound. *Boxes* embrace 50% of the scores, and *horizontal lines* (*whiskers*) cover 99.3% of the scores. *Circles* with *black diamonds*, outliers.

59

FtmPT1 with pAG012, FgaPT2 with pIU18, and CdpNPT with pHL5 were carried out as reported previously (19–21).

Enzyme Assays with Recombinant Proteins—In the assays to determine the acceptance of the different substrates, the enzyme reaction mixtures contained 50 mM Tris-HCl, pH 7.5, 10 mM CaCl$_2$, 2 mM DMAPP, 2–7.5% (v/v) glycerol, 1–2% (v/v) DMSO, 1 mM aromatic substrate, and 0.4 mg·ml^{-1} purified recombinant protein in a volume of 100 μl. The reaction mixtures were incubated at 37 °C for 16 h and terminated by the addition of an equal volume of methanol. The reaction mixtures were brought to dryness by vacuum evaporation and subsequently resuspended in 100 μl of methanol and centrifuged at 13,000 rpm for 15 min. Five μl of the supernatants were analyzed on LC-MS.

For isolation of the enzyme products, the reaction mixtures were scaled up to 10 ml, containing 50 mM Tris-HCl, pH 7.5, 10 mM CaCl$_2$, 2 mM DMAPP, 2–7.5% (v/v) glycerol, 1–2% (v/v) DMSO, 1 mM aromatic substrate, and 0.4 mg·ml^{-1} purified recombinant protein, and incubated at 37 °C for 16 h. The reactions were terminated by the addition of 10 ml of methanol and brought to dryness by using a rotary evaporator at 37 °C. The residues were resuspended in 1 ml of methanol, centrifuged at 13,000 rpm for 15 min, and purified on an HPLC device.

LC-ESI-HRMS Analysis of the Reaction Mixtures—The treated enzyme reaction mixtures (5 μl) mentioned above were analyzed on an Agilent 1260 Infinity HPLC system (Böblingen, Germany) in combination with a photodiode array detector and a Bruker micrOTOF-Q III mass spectrometer. For separation, a Multospher 120 RP-18 column (250 × 2 mm, 5 μm, CS-Chromatographie Service, Langerwehe, Germany) with a flow rate of 0.25 ml·min^{-1} was used. Water (solvent A) and MeCN (solvent B), both containing 0.1% (v/v) formic acid, were used for a linear gradient of 5–100% (v/v) solvent B in A in 40 min. Subsequently, the column was washed with 100% solvent B for 5 min and equilibrated with 5% (v/v) solvent B for 10 min. The separations were monitored with the Bruker micrOTOF-Q III mass spectrometer using the positive ion ESI. HPLC and MS data were processed by using Bruker Compass DataAnalysis version 4.2 (build 383.1) software.

Isolation of Enzymatic Products—Isolation of the enzyme products was performed on an Agilent HPLC series 1200. The separation was carried out on a MultoHigh Chiral AM-RP column (250 × 10 mm, 5 μm, CS-Chromatographie Service) with a flow rate of 1 ml·min^{-1} and different linear gradients of methanol in water.

NMR Analysis—The isolated enzyme products were brought to dryness by using a rotary evaporator at 37 °C and dissolved in 0.7 ml of CD$_3$OD. NMR spectra were recorded on a JEOL ECA 500-MHz spectrometer (JEOL Germany GmbH, Munich, Germany). The signal of CD$_3$OD at 3.31 ppm was used as an internal reference for chemical shifts. Data processing was done by using MestReNova version 6.0.2-5475 software.

Compound 21a—^{1}H NMR (methanol-d_4, 500 MHz) δ = 7.55 (dt, J = 8.0, 0.9 Hz), 7.30 (dt, J = 8.2, 0.9 Hz), 7.12 (ddd, J = 8.2, 7.0, 0.9 Hz), 7.06 (s), 7.02 (ddd, J = 8.0, 7.0, 0.9 Hz), 5.35 (m), 4.70 (d, J = 6.8 Hz) Approx. 3.33 (t, J = 7.3 Hz, signal overlapping with those of solvent), 2.9 (t, J = 7.3 Hz), 2.89 (q, J = 7.4

Hz), 1.85 (s), 1.76 (s), 1.19 (t, J = 7.4 Hz); HR-ESI-MS: m/z = 321.1647, Calcd. for C$_{17}$H$_{25}$N$_2$O$_2$S, [M + H]$^+$: 321.1631.

Compound 21b—^{1}H NMR (methanol-d_4, 500 MHz) δ = 7.14 (d, J = 7.6 Hz), 7.05 (td, J = 7.6, 1.1 Hz), 6.70 (td, J = 7.6, 0.9 Hz), 6.60 (d, J = 7.6 Hz), 6.07 (dd, J = 17.4, 10.9 Hz), 5.38 (s), 5.12 (dd, J = 10.9, 1.3 Hz), 5.08 (dd, J = 17.4, 1.3 Hz), 3.54 (dd, J = 10.0, 8.4 Hz), 3.10 (q, J = 7.4 Hz), 2.94 (ddd, J = 11.5, 9.7, 5.3 Hz), 2.40 (ddd, J = 12.1, 11.9, 7.9 Hz), 2.07 (dd, J = 12.3, 5.3 Hz), 1.29 (t, J = 7.4 Hz), 1.10 (s), 0.98 (s); HR-ESI-MS: m/z = 321.1642, Calcd. for C$_{17}$H$_{25}$N$_2$O$_2$S, [M + H]$^+$: 321.1631.

Compound 30a—^{1}H NMR (methanol-d_4, 500 MHz) δ = 7.46 (br d, J = 7.5 Hz), 7.37 (d, J = 8.6 Hz), 7.35 (br t, J = 7.5 Hz), 7.30 (br t, J = 7.5 Hz), 6.92 (d, J = 2.2 Hz), 6.81 (s), 6.73 (dd, J = 8.6, 2.2 Hz), 5.40 (m), 5.08 (s), 3.38 (d, J = 7.0 Hz), 1.76 (s), 1.74 (s); HR-ESI-MS: m/z = 292.1703, Calcd. for C$_{20}$H$_{22}$NO, [M + H]$^+$: 292.1696.

Compound 30b—^{1}H NMR (methanol-d_4, 500 MHz) δ = 7.48 (br d, J = 7.5 Hz), 7.37 (br t, J = 7.5 Hz), 7.29 (br t, J = 7.5 Hz), 7.24 (s), 7.04 (d, J = 3.2 Hz), 6.96 (s), 6.28 (dd, J = 3.2 Hz, 0.9), 5.35 (m), 5.09 (s), 3.39 (d, J = 7.5 Hz), 1.72 (s), 1.67 (s); HR-ESI-MS: m/z = 292.1704, Calcd. for C$_{20}$H$_{21}$NO, [M + H]$^+$: 292.1696.

Author Contributions—J. G. and P. K. designed the study and wrote the paper. J. G. extracted the data of prenyltransferase reactions from the literature, designed and implemented PrenDB, analyzed cataloged reactions and the substrate space in an automatic fashion, programmed automatic repitope extraction, conducted the virtual screen for novel substrates, and conducted and analyzed the similarity study. F. K. designed, performed, and analyzed the enzyme assays with recombinant proteins; performed and analyzed LC-ESI-HRMS analysis of the reaction mixtures; performed the isolation and structure elucidation of compounds **21a**, **21b**, **30a**, and **30b**; and contributed to writing of the paper. S.-M. L. contributed to database design, designed enzyme assays, analyzed LC-ESI-HRMS spectra, and contributed to writing the paper. All authors reviewed the results and approved the final version of the manuscript.

References

1. Liu, A. H., Liu, D. Q., Liang, T. J., Yu, X. Q., Feng, M. T., Yao, L. G., Fang, Y., Wang, B., Feng, L. H., Zhang, M. X., and Mao, S. C. (2013) Caulerprenylols A and B, two rare antifungal prenylated *para*-xylenes from the green alga *Caulerpa racemosa. Bioorg. Med. Chem. Lett.* **23**, 2491–2494
2. Oya, A., Tanaka, N., Kusama, T., Kim, S. Y., Hayashi, S., Kojoma, M., Hishida, A., Kawahara, N., Sakai, K., Gonoi, T., and Kobayashi, J. (2015) Prenylated benzophenones from *Triadenum japonicum. J. Nat. Prod.* **78**, 258–264
3. Sunassee, S. N., and Davies-Coleman, M. T. (2012) Cytotoxic and antioxidant marine prenylated quinones and hydroquinones. *Nat. Prod. Rep.* **29**, 513–535
4. Li, S.-M. (2010) Prenylated indole derivatives from fungi: structure diversity, biological activities, biosynthesis and chemoenzymatic synthesis. *Nat. Prod. Rep.* **27**, 57–78

5. Wollinsky, B., Ludwig, L., Hamacher, A., Yu, X., Kassack, M. U., and Li, S.-M. (2012) Prenylation at the indole ring leads to a significant increase of cytotoxicity of tryptophan-containing cyclic dipeptides. *Bioorg. Med. Chem. Lett.* **22,** 3866 –3869

6. Botta, B., Vitali, A., Menendez, P., Misiti, D., and Delle Monache, G. (2005) Prenylated flavonoids: pharmacology and biotechnology. *Curr. Med. Chem.* **12,** 717–739

7. Winkelblech, J., Fan, A., and Li, S.-M. (2015) Prenyltransferases as key enzymes in primary and secondary metabolism. *Appl. Microbiol. Biotechnol.* **99,** 7379 –7397

8. Heide, L. (2009) Prenyl transfer to aromatic substrates: genetics and enzymology. *Curr. Opin. Chem. Biol.* **13,** 171–179

9. Mai, P., Zocher, G., Ludwig, L., Stehle, T., and Li, S.-M. (2016) Actions of tryptophan prenyltransferases toward fumiquinazolines and their potential application for the generation of prenylated derivatives by combining chemical and chemoenzymatic syntheses. *Adv. Synth. Catal.* **358,** 1639 –1653

10. Fan, A., Winkelblech, J., and Li, S.-M. (2015) Impacts and perspectives of prenyltransferases of the DMATS superfamily for use in biotechnology. *Appl. Microbiol. Biotechnol.* **99,** 7399 –7415

11. Rogers, D., and Hahn, M. (2010) Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50,** 742–754

12. Gunera, J., and Kolb, P. (2015) Fragment-based similarity searching with infinite color space. *J. Comput. Chem.* **36,** 1597–1608

13. Weininger, D. (1988) SMILES, a chemical language and information system 1: introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28,** 31–36

14. Irwin, J. J., and Shoichet, B. K. (2005) ZINC: a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45,** 177–182

15. McGann, M. (2012) FRED and HYBRID docking performance on standardized datasets. *J. Comput. Aided Mol. Des.* **26,** 897–906

16. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **50,** 572–584

17. Hawkins, P. C. D., Skillman, A. G., and Nicholls, A. (2007) Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **50,** 74 –82

18. Woodside, A. B., Huang, Z., and Poulter, C. D. (1988) Trisammonium geranyl diphosphate. *Org. Synth.* **66,** 211–215

19. Grundmann, A., and Li, S.-M. (2005) Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from *Aspergillus fumigatus*. *Microbiology* **151,** 2199 –2207

20. Unsöld, I. A., and Li, S.-M. (2005) Overproduction, purification and characterization of FgaPT2, a dimethylallyltryptophan synthase from *Aspergillus fumigatus*. *Microbiology* **151,** 1499 –1505

21. Yin, W.-B., Ruan, H.-L., Westrich, L., Grundmann, A., and Li, S.-M. (2007) CdpNPT, an *N*-prenyltransferase from *Aspergillus fumigatus*: overproduction, purification and biochemical characterisation. *Chem. Bio Chem.* **8,** 1154 –1161

22. Schomburg, K., Ehrlich H.-C., Stierand K., and Rarey, M. (2010) From structure diagrams to visual chemical patterns. *J. Chem. Inf. Model.* **50,** 1529 –1535

61

SUPPORTING INFORMATION

PrenDB: A substrate prediction database to enable biocatalytic use of prenyltransferases

Jakub Gunera [‡†1], Florian Kindinger [§1], Shu-Ming Li [§†2] and Peter Kolb[‡†3]

‡ Department of Pharmaceutical Chemistry, Philipps-University, Marburg, Hesse, 35032, Germany
§ Institute of Pharmaceutical Biology and Biotechnology, Philipps-University, Marburg, Hesse, 35032, Germany
† Synmikro, LOEWE Centre for Synthetic Microbiology, Philipps-University, Marburg, Hesse, 35043, Germany
[1]These authors contributed equally to this study.
[2]To whom correspondence should be addressed: shuming.li@staff.uni-marburg.de
[3]To whom correspondence should be addressed: peter.kolb@uni-marburg.de

Table S1: List of substrates with their PrenDB ID, SMILES string and cluster membership (C)
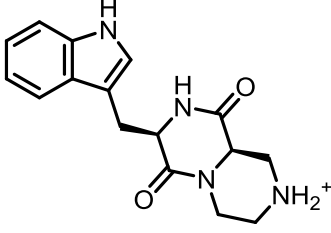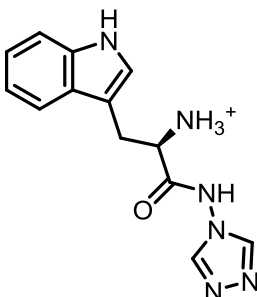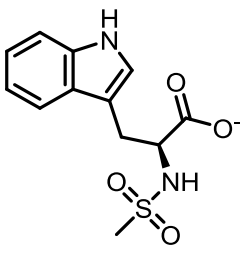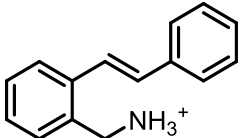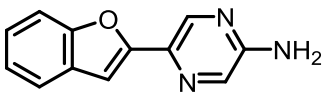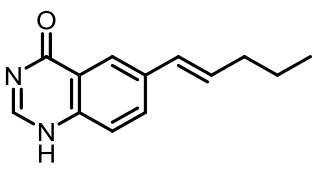
ID	PrenDB ID	SMILES	C
S1	PTDBSUB00695	O=C1NC(Cc2c[nH]c3ccccc23)C(=O)N2CCC[C@@H]12	1
S2	PTDBSUB00027	O=C1N[C@H](Cc2c[nH]c3ccccc23)C(=O)N2CCC[C@H]12	1
S3	PTDBSUB00022	O=C1N[C@H](Cc2c[nH]c3ccccc23)C(=O)N2CCC[C@@H]12	1
S4	PTDBSUB00001	O=C1N[C@@H](Cc2c[nH]c3ccccc23)C(=O)N2CCC[C@@H]12	1
S5	PTDBSUB00017	O=C1N[C@@H](Cc2c[nH]c3ccccc23)C(=O)N2CCC[C@H]12	1
S6	PTDBSUB00286	O=C1N[C@H](Cc2c[nH]c3ccccc23)C(=O)Nc2ccccc21	2
S7	PTDBSUB00392	O=C1N[C@@H](Cc2c[nH]c3ccccc23)C(=O)Nc2ccccc21	2
S8	PTDBSUB00049	C[C@@H]1NC(=O)[C@@H](Cc2c[nH]c3ccccc23)NC1=O	2
S9	PTDBSUB00044	C[C@H]1NC(=O)[C@@H](Cc2c[nH]c3ccccc23)NC1=O	2
S10	PTDBSUB00032	C[C@@H]1NC(=O)[C@H](Cc2c[nH]c3ccccc23)NC1=O	2
S11	PTDBSUB00039	C[C@H]1NC(=O)[C@H](Cc2c[nH]c3ccccc23)NC1=O	2
S12	PTDBSUB00012	O=C1N[C@@H](Cc2c[nH]c3ccccc23)C(=O)N[C@H]1Cc1ccccc1	2
S13	PTDBSUB00005	O=C1N[C@@H](Cc2c[nH]c3ccccc23)C(=O)N[C@H]1Cc1ccc(O)cc1	2
S14	PTDBSUB00412	O=C1N[C@H](Cc2c[nH]c3ccccc23)C(=O)N[C@H]1Cc1ccc(O)cc1	2
S15	PTDBSUB00059	O=C1CNC(=O)[C@H](Cc2c[nH]c3ccccc23)N1	2
S16	PTDBSUB00072	O=C1N[C@@H](Cc2c[nH]c3ccccc23)C(=O)N[C@H]1Cc1c[nH]c2ccccc12	2
S17	PTDBSUB00054	O=C1N[C@@H](Cc2c[nH]c3ccccc23)C(=O)N[C@H]1Cc1cnc[nH]1	2
S18	PTDBSUB00064	CC(C)C[C@@H]1NC(=O)[C@H](Cc2c[nH]c3ccccc23)NC1=O	2
S19	PTDBSUB00103	[NH3+][C@@H](Cc1c[nH]c2ccccc12)C(=O)[O-]	3
S20	PTDBSUB00173	[NH3+][C@H](Cc1c[nH]c2ccccc12)C(=O)[O-]	3
S21	PTDBSUB00186	C[C@@]([NH3+])(Cc1c[nH]c2ccccc12)C(=O)[O-]	3
S22	PTDBSUB00189	C[C@]([NH3+])(Cc1c[nH]c2ccccc12)C(=O)[O-]	3
S23	PTDBSUB00179	[NH3+]CCc1c[nH]c2ccccc12	3
S24	PTDBSUB00415	O=C([O-])Cc1c[nH]c2ccccc12	3
S25	PTDBSUB00491	O=C([O-])C(=O)Cc1c[nH]c2ccccc12	3
S26	PTDBSUB00297	O=C([O-])CCc1c[nH]c2ccccc12	3
S27	PTDBSUB00496	O=C([O-])CCCc1c[nH]c2ccccc12	3
S28	PTDBSUB00676	CC(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(C)=O	3
S29	PTDBSUB00317	CC(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)[O-]	3
S30	PTDBSUB00423	CC(=O)N[C@H](Cc1c[nH]c2ccccc12)C(=O)[O-]	3
S31	PTDBSUB00183	C[NH2+][C@@H](Cc1c[nH]c2ccccc12)C(=O)[O-]	3
S32	PTDBSUB00192	O=C([O-])[C@@H](O)Cc1c[nH]c2ccccc12	3
S33	PTDBSUB00195	O=C([O-])[C@H](O)Cc1c[nH]c2ccccc12	3
S34	PTDBSUB00176	[NH3+][C@H](CC(=O)[O-])Cc1c[nH]c2ccccc12	3
S35	PTDBSUB00418	[NH3+][C@@H](CC(=O)[O-])Cc1c[nH]c2ccccc12	3
S36	PTDBSUB00139	[NH3+][C@@H](Cc1c[nH]c2ccccc12)C(=O)NCC(=O)[O-]	3
S37	PTDBSUB00488	[NH3+][C@H](Cc1c[nH]c2ccccc12)C(=O)NO	3
S38	PTDBSUB00499	COC(=O)[C@H]([NH3+])Cc1c[nH]c2ccccc12	3
S39	PTDBSUB00593	[NH3+][C@@H](C(=O)[O-])[C@@H](O)c1ccc(O)c(O)c1	4
S40	PTDBSUB00722	[NH3+][C@@H](C(=O)[O-])C(O)c1ccc(O)c(O)c1	4

2
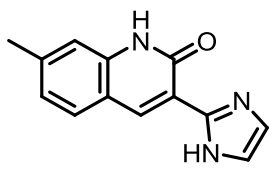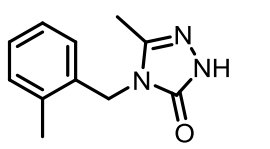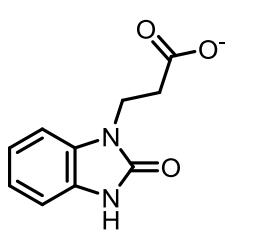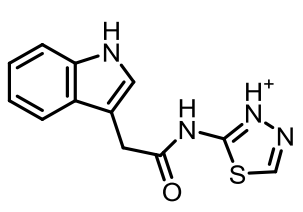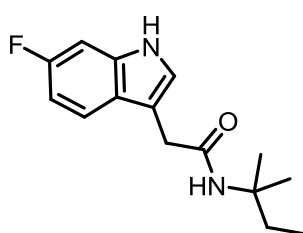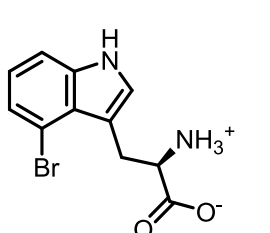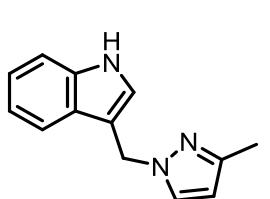
S41	PTDBSUB00689	Cc1ccc(C[C@@H]2NC(=O)[C@H](Cc3ccc(O)cc3)NC2=O)cc1	5
S42	PTDBSUB00571	O=C([O-])[C@@H](O)Cc1ccc(O)cc1	5
S43	PTDBSUB00574	O=C([O-])[C@H](O)Cc1ccc(O)cc1	5
S44	PTDBSUB00554	[NH3+][C@@H](C(=O)[O-])c1ccc(O)cc1	5
S45	PTDBSUB00551	[NH3+][C@@H](CC(=O)[O-])c1ccc(O)cc1	5
S46	PTDBSUB00718	[NH3+]C(CC(=O)[O-])c1ccc(O)cc1	5
S47	PTDBSUB00376	CC([NH3+])(Cc1ccc(O)cc1)C(=O)[O-]	5
S48	PTDBSUB00578	C[C@@]([NH3+])(Cc1ccc(O)cc1)C(=O)[O-]	5
S49	PTDBSUB00467	O=C(CCc1ccc(O)cc1)c1c(O)cc(O)cc1O	5
S50	PTDBSUB00079	O=C([O-])C(=O)Cc1ccc(O)cc1	5
S51	PTDBSUB00568	O=C([O-])CCc1ccc(O)cc1	5
S52	PTDBSUB00666	Oc1cc2ccccc2cc1O	6
S53	PTDBSUB00652	Oc1ccc2ccccc2c1	6
S54	PTDBSUB00289	Oc1ccc2ccc(O)cc2c1	6
S55	PTDBSUB00613	Oc1ccc2cc(O)ccc2c1	6
S56	PTDBSUB00647	Nc1cccc2ccc(O)cc12	6
S57	PTDBSUB00448	COc1cccc2ccc(O)cc12	6
S58	PTDBSUB00451	CCOc1cccc2ccc(O)cc12	6
S59	PTDBSUB00656	Oc1ccc(O)c2ccccc12	6
S60	PTDBSUB00661	Oc1cccc2c(O)cccc12	6
S61	PTDBSUB00399	Oc1ccc2c(O)cccc2c1	6
S62	PTDBSUB00596	Oc1cccc2ccccc12	6
S63	PTDBSUB00445	Cc1ccc2cccc(O)c2c1	6
S64	PTDBSUB00601	Oc1ccc2cccc(O)c2c1	6
S65	PTDBSUB00088	COc1ccc2c3c4n(c2c1)[C@@H](C=C(C)C)OOC(C)(C)C[C@@H]4N1C(=O)[C@@H]2CCCN2C(=O)[C@]1(O)[C@H]3O	6
S66	PTDBSUB00082	COc1ccc2c3c([nH]c2c1)[C@H](C=C(C)C)N1C(=O)[C@@H]2CCCN2C(=O)[C@]1(O)[C@H]3O	6
S67	PTDBSUB00085	COc1ccc2c3c(n(CC=C(C)C)c2c1)[C@H](C=C(C)C)N1C(=O)[C@@H]2CCCN2C(=O)[C@]1(O)[C@H]3O	6
S68	PTDBSUB00094	Cc1cc2oc3cccc(=O)c-3c([O-])c2cc1O	6
S69	PTDBSUB00100	Cc1c([O-])c(C)c2c(O)c3c(=O)cccc-3oc2c1C	6
S70	PTDBSUB00091	Cc1cc2oc3cccc(=O)c-3c(O)c2c(CO)c1[O-]	6
S71	PTDBSUB00097	Cc1cc2oc3cccc(=O)c-3c(O)c2c(C)c1[O-]	6
S72	PTDBSUB00479	COc1ccc(-c2coc3cc(=O)cc(O)c-3c2[O-])cc1	6
S73	PTDBSUB00472	O=c1cc2oc(-c3ccc(O)cc3)cc([O-])c-2c(O)c1	6
S74	PTDBSUB00343	O=c1cc2oc3cc(O)cc([O-])c3c([O-])c-2c(O)c1	6
S75	PTDBSUB00339	O=c1cc2oc3ccc(O)cc3c([O-])c-2c(O)c1	6
S76	PTDBSUB00351	O=c1cc2oc3cc(O)ccc3c([O-])c-2c(O)c1	6
S77	PTDBSUB00292	O=C1C=C([O-])c2c([O-])cc(O)cc2C1=O	6
S78	PTDBSUB00606	O=C(O)c1cc2cccc([O-])c2cc1O	6
S79	PTDBSUB00610	O=C(O)c1cc2cc(O)ccc2cc1[O-]	6
S80	PTDBSUB00483	O=C1C[C@H](c2ccccc2)Oc2cc(O)ccc21	6
S81	PTDBSUB00454	Cc1cc([O-])c2c(c1)O[C@H](c1ccc(O)cc1)CC2=O	6
S82	PTDBSUB00460	COc1ccc([C@@H]2CC(=O)c3c([O-])cc(C)cc3O2)cc1O	6

3
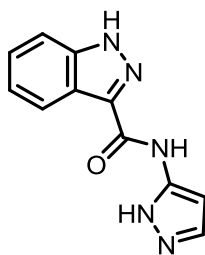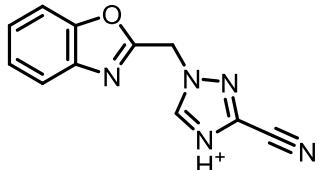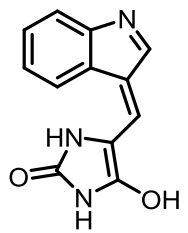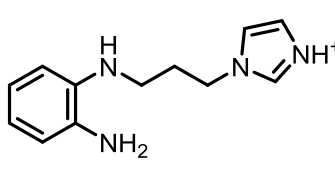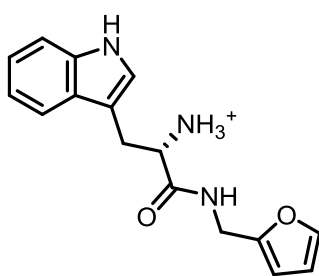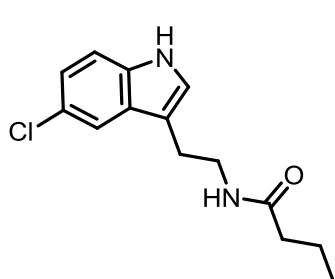
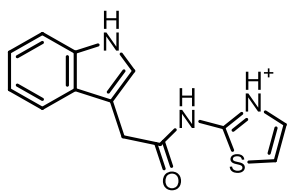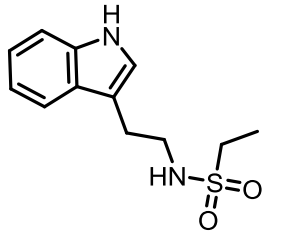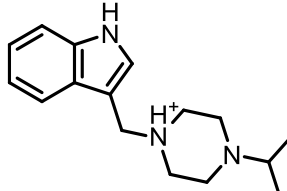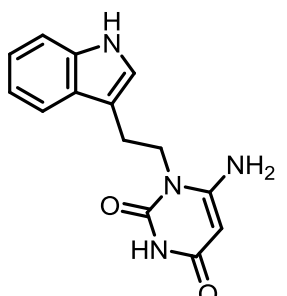| S83	PTDBSUB00464	COc1cc([C@H]2Oc3cc([C@H]4Oc5cc(O)cc([O-])c5C(=O)[C@@H]4O)ccc3O[C@@H]2CO)ccc1O	6
S84	PTDBSUB00457	Cc1cc(O)c2c(c1)O[C@H](c1ccc(O)c(O)c1)CC2O	6
S85	PTDBSUB00475	Oc1ccc(C2=COc3cc(O)cc(O)c3C2O)cc1	6
S86	PTDBSUB00367	Cn1c(O)c2c3c(c4c(c2c1O)=c1ccccc1=N4)N=c1ccccc1=3	6
S87	PTDBSUB00144	COc1c(O)c(=C2C=Nc3ccccc32)c(OC)c(O)c1=C1C=Nc2ccccc21	6
S88	PTDBSUB00396	O=C([O-])c1cccc2c1Nc1ccccc1N2	6
S89	PTDBSUB00385	CC(=O)O[C@H]1[C@@H](C)C[NH+](C)[C@@H]2Cc3c[nH]c4cccc(c34)[C@@H]12	7
S90	PTDBSUB00388	CC(=O)O[C@H]1[C@H](C)C[NH+](C)[C@@H]2Cc3c[nH]c4cccc(c34)[C@@H]12	7
S91	PTDBSUB00136	CC(=O)/C=C/c1c[nH]c2ccccc12	8
S92	PTDBSUB00300	O=C([O-])/C=C/c1c[nH]c2ccccc12	8
S93	PTDBSUB00359	O=C1NCc2c1c1c3ccccc3[nH]c1c1[nH]c3ccccc3c21	8
S94	PTDBSUB00364	O=C1N[C@@H](O)c2c1c1c3ccccc3[nH]c1c1[nH]c3ccccc3c21	8
S95	PTDBSUB00148	C=c1c(OC)c(-c2c(CC=C(C)C)[nH]c3ccccc23)c(=C)c(OC)c1-c1c[nH]c2ccccc12	8
S96	PTDBSUB00151	C=CC(C)(C)n1cc(C2=C(OC)C(=O)C(c3c[nH]c4ccccc34)=C(OC)C2=O)c2ccccc21	8
S97	PTDBSUB00703	[NH3+][C@H](Cc1cc(I)c([O-])c(I)c1)C(=O)[O-]	9
S98	PTDBSUB00370	[NH3+][C@@H](Cc1cc(Br)c([O-])c(Br)c1)C(=O)[O-]	9
S99	PTDBSUB00373	[NH3+][C@@H](Cc1cc(I)c([O-])c(I)c1)C(=O)[O-]	9
S100	PTDBSUB00548	[NH3+][C@@H](Cc1ccc([O-])c([N+](=O)[O-])c1)C(=O)[O-]	9
S101	PTDBSUB00706	[NH3+][C@H](Cc1ccc([O-])c([N+](=O)[O-])c1)C(=O)[O-]	9
S102	PTDBSUB00106	[NH3+][C@@H](Cc1ccc(O)cc1)C(=O)[O-]	9
S103	PTDBSUB00542	[NH3+][C@H](Cc1ccc(O)cc1)C(=O)[O-]	9
S104	PTDBSUB00321	Nc1ccc(C[C@H]([NH3+])C(=O)[O-])cc1	9
S105	PTDBSUB00545	Nc1ccc(C[C@@H]([NH3+])C(=O)[O-])cc1	9
S106	PTDBSUB00587	[NH3+][C@H](Cc1ccc(O)c(I)c1)C(=O)[O-]	9
S107	PTDBSUB00584	[NH3+][C@@H](Cc1ccc(O)c(F)c1)C(=O)[O-]	9
S108	PTDBSUB00382	[NH3+][C@@H](Cc1ccc(O)c(I)c1)C(=O)[O-]	9
S109	PTDBSUB00581	[NH3+][C@H](Cc1ccc(O)c(F)c1)C(=O)[O-]	9
S110	PTDBSUB00112	[NH3+][C@@H](Cc1cccc(O)c1)C(=O)[O-]	9
S111	PTDBSUB00565	[NH3+][C@H](Cc1cccc(O)c1)C(=O)[O-]	9
S112	PTDBSUB00562	Nc1cc(C[C@@H]([NH3+])C(=O)[O-])ccc1O	9
S113	PTDBSUB00379	[NH3+][C@@H](Cc1ccc(O)c(O)c1)C(=O)[O-]	9
S114	PTDBSUB00590	[NH3+][C@H](Cc1ccc(O)c(O)c1)C(=O)[O-]	9
S115	PTDBSUB00161	Cc1ccc2c(C[C@H]([NH3+])C(=O)[O-])c[nH]c2c1	10
S116	PTDBSUB00164	Cc1ccc2c(C[C@@H]([NH3+])C(=O)[O-])c[nH]c2c1	10
S117	PTDBSUB00310	[NH3+][C@H](Cc1c[nH]c2cc(F)ccc12)C(=O)[O-]	10
S118	PTDBSUB00441	[NH3+][C@@H](Cc1c[nH]c2cc(F)ccc12)C(=O)[O-]	10
S119	PTDBSUB00516	[NH3+][C@H](Cc1c[nH]c2ccc(F)cc12)C(=O)[O-]	10
S120	PTDBSUB00438	[NH3+][C@@H](Cc1c[nH]c2ccc(F)cc12)C(=O)[O-]	10
S121	PTDBSUB00432	[NH3+][C@@H](Cc1c[nH]c2ccc(Br)cc12)C(=O)[O-]	10
S122	PTDBSUB00435	[NH3+][C@H](Cc1c[nH]c2ccc(Br)cc12)C(=O)[O-]	10
S123	PTDBSUB00167	COc1ccc2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c2c1	10
S124	PTDBSUB00170	COc1ccc2[nH]cc(C[C@@H]([NH3+])C(=O)[O-])c2c1	10

4
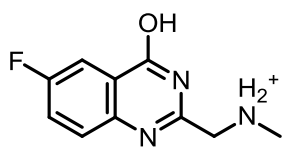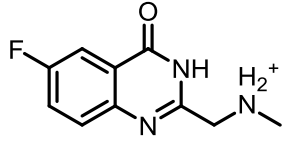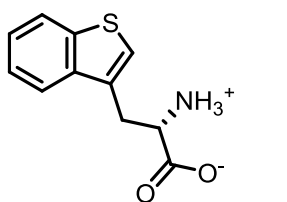
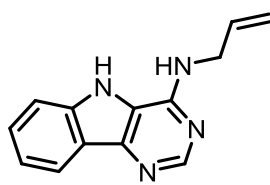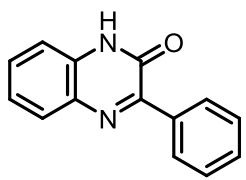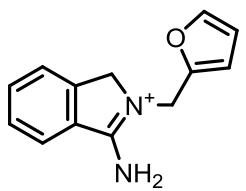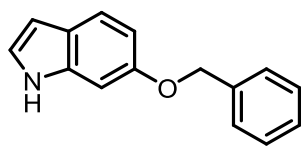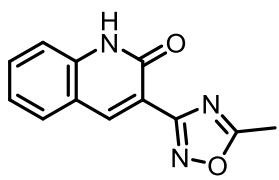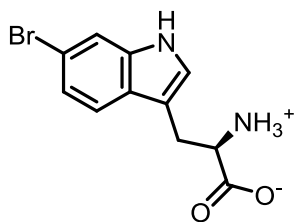| S125	PTDBSUB00155	Cc1ccc2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c2c1	10
S126	PTDBSUB00158	Cc1ccc2[nH]cc(C[C@@H]([NH3+])C(=O)[O-])c2c1	10
S127	PTDBSUB00429	[NH3+][C@@H](Cc1c[nH]c2ccc(O)cc12)C(=O)[O-]	10
S128	PTDBSUB00511	[NH3+][C@H](Cc1c[nH]c2ccc(O)cc12)C(=O)[O-]	10
S129	PTDBSUB00303	Cn1cc(C[C@@H]([NH3+])C(=O)[O-])c2ccccc21	11
S130	PTDBSUB00109	[NH3+][C@@H](Cc1ccccc1O)C(=O)[O-]	11
S131	PTDBSUB00558	[NH3+][C@H](Cc1ccccc1O)C(=O)[O-]	11
S132	PTDBSUB00313	Cc1cccc2c(C[C@@H]([NH3+])C(=O)[O-])c[nH]c12	11
S133	PTDBSUB00407	Cc1cccc2c(C[C@H]([NH3+])C(=O)[O-])c[nH]c12	11
S134	PTDBSUB00120	Cc1cccc2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c12	11
S135	PTDBSUB00306	Cc1cccc2[nH]cc(C[C@@H]([NH3+])C(=O)[O-])c12	11
S136	PTDBSUB00123	COc1cccc2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c12	11
S137	PTDBSUB00126	Nc1cccc2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c12	11
S138	PTDBSUB00274	CC(C)=CCc1cc(C)cc2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c12	12
S139	PTDBSUB00276	CC(C)=CCc1cc(C)cc2[nH]cc(C[C@@H]([NH3+])C(=O)[O-])c12	12
S140	PTDBSUB00270	CC(C)=CCc1c(C)ccc2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c12	12
S141	PTDBSUB00272	CC(C)=CCc1c(C)ccc2[nH]cc(C[C@@H]([NH3+])C(=O)[O-])c12	12
S142	PTDBSUB00278	COc1ccc2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c2c1CC=C(C)C	12
S143	PTDBSUB00281	COc1ccc2[nH]cc(C[C@@H]([NH3+])C(=O)[O-])c2c1CC=C(C)C	12
S144	PTDBSUB00239	CC(C)=CCc1c(C)ccc2c(C[C@H]([NH3+])C(=O)[O-])c[nH]c12	12
S145	PTDBSUB00242	CC(C)=CCc1c(C)ccc2c(C[C@@H]([NH3+])C(=O)[O-])c[nH]c12	12
S146	PTDBSUB00233	CC(C)=CCc1cc(C)cc2c(C[C@H]([NH3+])C(=O)[O-])c[nH]c12	12
S147	PTDBSUB00236	CC(C)=CCc1cc(C)cc2c(C[C@@H]([NH3+])C(=O)[O-])c[nH]c12	12
S148	PTDBSUB00245	COc1cc(CC=C(C)C)c2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c2c1	12
S149	PTDBSUB00248	COc1cc(CC=C(C)C)c2[nH]cc(C[C@@H]([NH3+])C(=O)[O-])c2c1	12
S150	PTDBSUB00251	CC(C)=CCc1cccc2[nH]cc(C[C@H]([NH3+])C(=O)[O-])c12	13
S151	PTDBSUB00253	CC(C)=CCc1cccc2[nH]cc(C[C@@H]([NH3+])C(=O)[O-])c12	13
S152	PTDBSUB00266	CC(C)=CCc1cccc2[nH]cc(C[C@H](O)C(=O)[O-])c12	13
S153	PTDBSUB00268	CC(C)=CCc1cccc2[nH]cc(C[C@@H](O)C(=O)[O-])c12	13
S154	PTDBSUB00255	CC(C)=CCc1cccc2[nH]cc(C[C@H]([NH3+])CC(=O)[O-])c12	13
S155	PTDBSUB00260	C[NH2+][C@@H](Cc1c[nH]c2cccc(CC=C(C)C)c12)C(=O)[O-]	13
S156	PTDBSUB00257	CC(C)=CCc1cccc2[nH]cc(CC[NH3+])c12	13
S157	PTDBSUB00262	CC(C)=CCc1cccc2[nH]cc(C[C@@](C)([NH3+])C(=O)[O-])c12	13
S158	PTDBSUB00264	CC(C)=CCc1cccc2[nH]cc(C[C@](C)([NH3+])C(=O)[O-])c12	13
S159	PTDBSUB00205	CC(C)=CCc1cccc2c(C[C@H]([NH3+])C(=O)[O-])c[nH]c12	14
S160	PTDBSUB00208	CC(C)=CCc1cccc2c(C[C@@H]([NH3+])C(=O)[O-])c[nH]c12	14
S161	PTDBSUB00226	CC(C)=CCc1cccc2c(C[C@H](O)C(=O)[O-])c[nH]c12	14
S162	PTDBSUB00229	CC(C)=CCc1cccc2c(C[C@@H](O)C(=O)[O-])c[nH]c12	14
S163	PTDBSUB00211	CC(C)=CCc1cccc2c(C[C@H]([NH3+])CC(=O)[O-])c[nH]c12	14
S164	PTDBSUB00217	C[NH2+][C@@H](Cc1c[nH]c2c(CC=C(C)C)cccc12)C(=O)[O-]	14
S165	PTDBSUB00214	CC(C)=CCc1cccc2c(CC[NH3+])c[nH]c12	14
S166	PTDBSUB00220	CC(C)=CCc1cccc2c(C[C@@](C)([NH3+])C(=O)[O-])c[nH]c12	14
S167	PTDBSUB00223	CC(C)=CCc1cccc2c(C[C@](C)([NH3+])C(=O)[O-])c[nH]c12	14

5

6

Table S2: Structures, IDs, enzyme-related yields and number of matched PrenDB reactions.
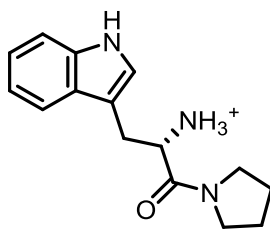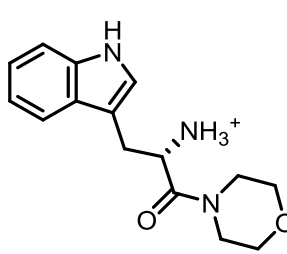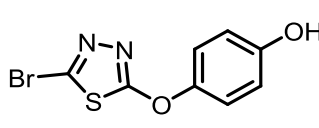
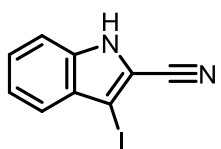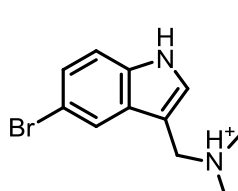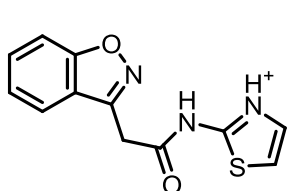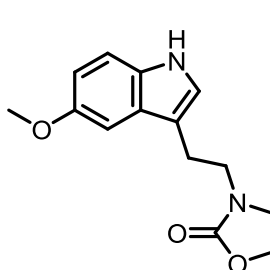| Substrate	ID	Product yield [%]			Matched reactions
		FtmPT1	FgaPT2	CdpNPT	
	1	94.1	18.2	41.3	299
	2	81.3	80.1	29.5	299
	3	-	3.2	22.1	299
	4	-	-	-	6
	5	-	-	0.9	33
	6	-	-	-	57

7
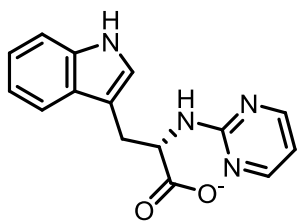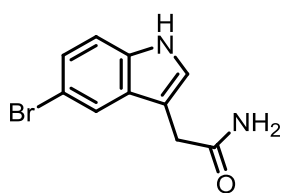
7	7.0	-	10.1	57
8	-	-	-	6
9	-	5.3	-	17
10	40.8	28.9	99.7	233
11	48.1	17.1	88.1	238
12	8.5	-	14.4	299
13	71.0	31.7	47.2	231

8

	14	2.0	-	2.6	50
	15	-	-	-	58
	16	68.8	34.8	29.7	20
	17	-	-	-	14
	18	89.1	84.9	62.0	299
	19	76.4	6.2	46.9	233

20	76.0	37.4	97.4	233
21	98.6	99.2	99.3	233
22	-	-	-	231
23	64.8	75.0	65.2	233
24[a]	-	-	-	50
25	-	-	-	50
26	-	-	-	33

10

27	8.7	82.9	11.8	137
28	-	-	-	23
29[b]	-	-	-	6
30	60.1	59.4	96.1	229
31	-	-	-	50
32	47.5	44.7	-	299
33	99.5	99.1	87.2	233

	34	44.4	19.9	94.8	299
	35	35.5	16.3	94.1	299
	36[b]	-	-	-	77
	37	-	-	-	137
	38	-	-	-	235
	39[b]	3.8	-	2.8	33
	40	51.5	4.4	9.2	237

| | 41 | 9.9 | 22.5 | 13.5 | 325 |

| | 42 | - | 14.9 | 38.1 | 299 |

---

[a] Compound **24** is the imidic acid tautomer of **25** and was thus excluded from further consideration in this work. Total and relative numbers throughout the manuscript reflect the number of unique compounds, i.e. 38.

[b] These compounds could not be obtained as ordered and were excluded from further consideration in this work

13

4.2. Two Prenyltransferases govern a consecutive prenylation cascade in the biosynthesis of echinulin and neoechinulin.

Viola Wohlgemuth, **Florian Kindinger**, Xiulan Xie, Bin-Gui Wang, and Shu-Ming Li (2017)

Two Prenyltransferases Govern a Consecutive Prenylation Cascade in the Biosynthesis of Echinulin and Neoechinulin

Viola Wohlgemuth,[†] Florian Kindinger,[†] Xiulan Xie,[‡] Bin-Gui Wang,[§] and Shu-Ming Li*[†]

[†]Institut für Pharmazeutische Biologie und Biotechnologie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany

[‡]Fachbereich Chemie, Philipps-Universität Marburg, Hans Meerwein-Straße, 35032 Marburg, Germany

[§]Key Laboratory of Experimental Marine Biology, Institute of Oceanology of the CAS, 266071 Qingdao, China

S *Supporting Information*

ABSTRACT: Two prenyltransferases from *Aspergillus ruber* control the echinulin biosynthesis via exceptional sequential prenylations. EchPT1 catalyzes the first prenylation step, leading to preechinulin. The unique EchPT2 attaches, in a consecutive prenylation cascade, up to three dimethylallyl moieties to preechinulin and its dehydro forms neo-echinulins A and B, resulting in the formation of at least 23 2- to 4-fold prenylated derivatives. Confirming these products in fungal extracts unravels the unprecedented catalytic relevance of EchPT2 for structural diversity.
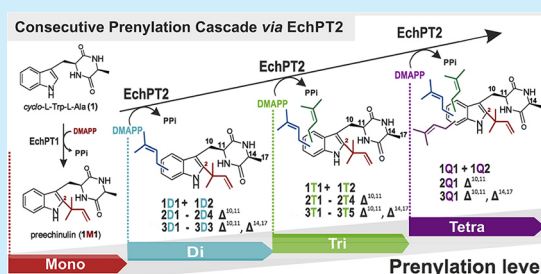
Indole diketopiperazine (DKP) alkaloids have been established as a steadily growing and reliable source for compounds of significant biological activity.[1,2] Their indole nucleus backbone, as a privileged structure, has become a focus of fragment-based drug discovery.[1,3] This core ring system is commonly assembled via one L-tryptophan and a second amino acid, usually catalyzed by a nonribosomal peptide synthetase (NRPS),[4] providing an initial biosynthetic starting point in the form of an indole DKP skeleton.[5] Subsequent modification reactions including prenylations increase not only structural complexity but also biological and pharmacological activities.[4−8]

Echinulins and neoechinulins derived from L-tryptophan and L-alanine exemplify such DKPs and are highly decorated with dimethylallyl (DMA) moieties.[1,2,5] Their eponymous group member echinulin (Figure 1) was first isolated from *Aspergillus*
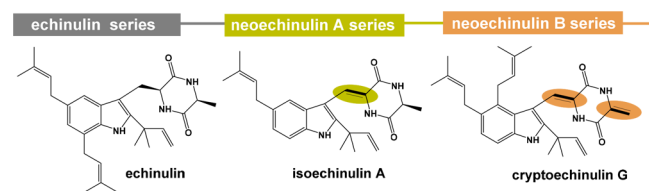

Figure 1. Representatives of echinulin and neoechinulins.

amstelodami[5] and later, together with congeners, from different terrestrial and marine-derived fungi,[5] e.g. *Aspergillus cristatus*[9] and *Aspergillus glaucus*.[10] These prenylated *cyclo*-L-Trp-L-Ala derivatives can be classified into the echinulin, neoechinulin A, and neoechinulin B series (Figures 1 and S1 in Supporting Information (SI)).[11] They share a reverse *C2*-prenyl moiety and differ from each other in *exo* double bounds at the DKP ring. In turn, the members of each series differ from each other in the

number and position of additional DMA moieties. C-5 of the indole ring seems to be the second preferred prenylation position in echinulins and neoechinulins. A third DMA moiety is mainly found at C-7 or at C-4.[11,12] Various intriguing pharmacological activities have been identified for echinulin and congeners,[5] such as protection against neuronal cell death[13] and antiviral[14] and antitumor activities.[15] In spite of over 40 years of studies on structures, as well as biological and pharmacological activities of this intriguing substance group, the enzymes involved in the biosynthesis, especially those for the transfer of the different prenyl moieties, have not been reported prior to this study.[2,5]

In nature, prenyl transfer reactions are catalyzed by prenyltransferases (PTs). These enzymes employ isoprenic precursors of various number of C_5-units such as DMAPP (C_5).[8] Known indole PTs from fungi belong to the dimethylallyl-tryptophan synthase (DMATS) superfamily and usually catalyze regio- and stereoselective regular or reverse prenylations. In most cases, one PT only catalyzes one specific transfer reaction.[8] For example, fumitremorgin A from *Neosartorya fischeri* contains three prenyl moieties, which are transferred from DMAPP by three different PTs.[16] Contrary examples of one PT involved in more than one prenylation step are rare. Two membrane-bound PTs, from *Humulus lupulus*, have been reported to be responsible for three sequential prenylation steps in the biosynthesis of β-bitter acid,[17] and in the biosynthesis of the fungal metabolite shearinine D, JanD from the DMATS superfamily catalyzes a tandem diprenylation.[18]

A large number of echinulin congeners with different numbers of prenyl moieties have been identified in and isolated from
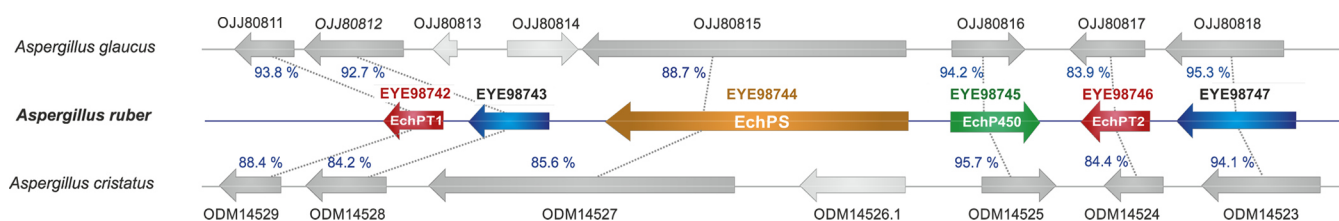
Figure 2. Putative echinulin gene clusters in *Aspergillus* strains. Genes with high sequence identities at the amino acid level are indicated by dotted lines. Details of alignment and the putative gene functions can be found in Table S1.

Aspergillus ruber.[12,14] In analogy to most biosynthetic pathways, it could be speculated that three or even four PTs are necessary for the attachment of the prenyl moieties in echinulin and its congeners. Unexpectedly, mining the *A. ruber* CBS 135680 genome[19] reveals just three DMATS PT genes. Two of them, coding for EYE98742 (termed EchPT1 in this study) and EYE98746 (EchPT2), build a cluster with an NRPS (EYE98744) gene. The third one (EYE95342) is located in a separate cluster with a polyketide synthase gene. Homologues of the putative NRPS-containing cluster harboring only two PT genes are also identified in the genomes of *Aspergillus cristatus* YKY807 and *Aspergillus glaucus* CBS 516.65 (Figure 2, Table S1). This unusual genetic organization prompted us to investigate the functions of the three PTs from *A. ruber*.

Sequence analysis revealed the incorrect annotation of the putative NRPS EYE98744 in the database by lacking 789 amino acids at the N-terminus, which was corrected in this study (SI). The revised sequence of EYE98744 comprises 2113 amino acid residues (Table S1) and shares a sequence identity of 23.5% with the known *cyclo*-L-Trp-L-Pro synthetase FtmPS.[20] It can be speculated that this enzyme, tentatively named EchPS, might be responsible for the assembly of *cyclo*-L-Trp-L-Ala. The two putative echinulin prenyltransferases EchPT1 and EchPT2 are polypeptides of 417 and 408 amino acids, respectively. They share a clear sequence similarity with the members of the DMATS superfamily (Figure S2).[8] It seems unbelievable that these two PTs should catalyze three or four prenylations deduced from the number and positions of DMA moieties in echinulins and congeners.

To investigate the possible involvement of the three PTs in the biosynthesis of echinulin, the coding sequences of EchPT1, EchPT2, and EYE95342 were PCR amplified from cDNA of the endophytic fungus *A. ruber* QEN-0407-G2[12] and cloned into pQE9 and pQE70 for heterologous expression in *E. coli* (SI, Figures S3−S5). All three purified proteins were incubated separately with the putative product of EchPS, *cyclo*-L-Trp-L-Ala (1), in the presence of DMAPP. LC-MS analysis revealed product formation in the reaction mixture with EchPT1, but with neither EchPT2 nor EYE95342 (Figure 3i−iii). The $[M + H]^+$ ion of the single EchPT1 product **1M1** indicates a monoprenylation of 1. For a better understanding, we use **M**, **D**, **T**, and **Q** after the substrate number for mono-, di-, tri-, and tetraprenylation, respectively. The number after these letters refers to the order of the identified products. Formation of **1M1** is strictly dependent on the presence of 1, DMAPP, and active EchPT1. NMR and MS analyses of the isolated product confirmed **1M1** to be preechinulin. This proved unequivocally that EchPT1 catalyzes the first prenylation in the biosynthesis of echinulin in *A. ruber* (for structure elucidation and kinetic parameters see SI, Tables S2 and S3, Figures S9−11, S16, and S18).

To verify the roles of the two other PTs in the biosynthesis of echinulins, we carried out incubations containing either two or all
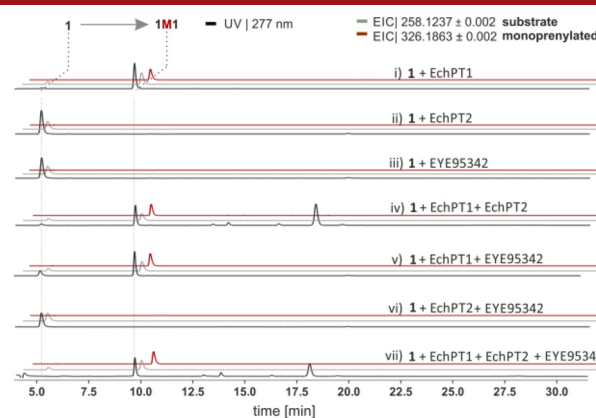

Figure 3. LC-MS analysis of the incubation mixtures of 1 with EchPT1, EchPT2, and EYE95342 alone or in combinations.

three PTs (Figure 3iv−vii). Assaying 1 with EchPT1 and EchPT2 resulted in the formation of **1M1** and at least four additional products (iv). The same peaks were also detected in the assay with all three enzymes (vii). In contrast, the combination of EchPT1 and EYE95342 (v) yields just the EchPT1 product **1M1**, while EchPT2 and EYE95342 catalyze no conversion of 1 at all (vi). These results prove that EYE95342 is likely not involved in the biosynthesis of echinulins and EchPT2 catalyzes the further metabolism of **1M1**.

To gain detailed insights into the reaction mechanism of 1 with EchPT1 and EchPT2 (Figure 3iv), we assayed **1M1** with EchPT2 in the presence of DMAPP. LC-MS analysis reveals six product peaks on three consecutive prenylation levels (Figures 4a and S6a), i.e. two di-, tri-, and tetraprenylated products each, which was confirmed by detection of their exact $[M + H]^+$ ions (Table S3). The $[M + H]^+$ ions of the first two products in ascending order of retention times (Figure S6a) are 68 Da larger than that of **1M1**, proving the presence of an additional prenyl moiety in their structures. These putative diprenylated products are termed **1D1** and **1D2**. The $[M + H]^+$ ions of the following two triprenylated products, **1T1** and **1T2**, indicate the attachment of two prenyl residues to **1M1**, and the $[M + H]^+$ ions of the last two products **1Q1** and **1Q2** indicate even three additional prenyl units attached to the already monoprenylated substrate **1M1**. Notably, while a large number of echinulin-related structures were described in the literature,[5,12] no derivative with four prenyl moieties has been reported to date. From Figure 4a, it is obvious that **1T2** is the main product of the EchPT2 reaction. To the best of our knowledge, six products across three consecutive prenylation levels from an incubation mixture with only one enzyme have not been reported prior to this study. These fascinating results prompted us to investigate the relationships between these products and their dependence on reaction time, DMAPP and protein concentrations (Table S2, Figures S12−15 and S17). HPLC analysis of the reaction mixtures revealed **1T2** as the
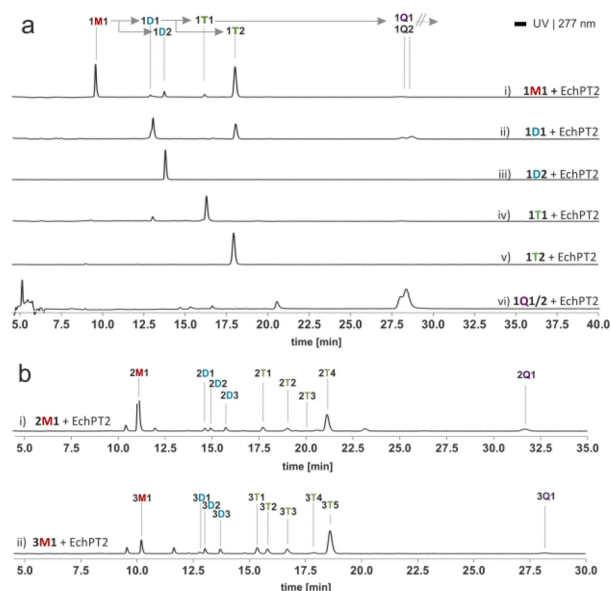
Figure 4. LC-MS analysis of EchPT2 assays with **1M1**, **1D1**, **1D2**, **1T1**, **1T2**, and **1Q1/1Q2** (a), neoechinulins A (**2M1**) and B (**3M1**) (b).

predominant product in all enzyme assays. Most interestingly, there are two divergent fate patterns for the enzyme products. In all three dependency assays, the formation of **1D1** and **1T1** showed an initially rapid, yet short increase, followed by a continuous decrease. In contrast, the yields of **1D2**, **1T2**, **1Q1**, and **1Q2** increased steadily in all assays (Figure S13). This indicates that **1D1** and **1T1** could serve as intermediates in a consecutive prenylation cascade, whereas **1D2** and **1T2** as well as **1Q2** and **1Q2** may represent end products of the (branch) pathways.

To confirm our hypothesis, we isolated EchPT2 products for structure elucidation. Interpretation of the spectra and literature search confirmed **1D1** and **1D2** to be tardioxopiperazines A and

B with an additional regular prenyl moiety at C-5 and C-7, respectively, compared to that of **1M1**. The two triprenylated products **1T1** and **1T2** bear prenyl moieties at C-2 and C-5 and differ from each other in the position of the third prenyl residue, which is located at C-4 in **1T1** and C-7 in the case of **1T2**. **1T1** and **1T2** can therefore be unequivocally identified as variecolorin L and echinulin. Due to the low amount, the structure of **1Q1** cannot be determined in this study. **1Q2** was isolated from fungal extracts (Figure S8) and identified as a reversely $C2$-, regularly $C4$-, $C5$-, and $C7$-tetraprenylated derivative. Judging by incubation results described below, **1Q1** must be a reversely $C2$-, regularly $C4$-, $C5$-, and $C6$-tetraprenylated derivative (Scheme 1; Tables S3−S7; Figures S19−29).

The single EchPT2 products were then incubated independently with EchPT2 (Figures 4a and S6a). LC-MS analysis showed further conversion of **1D1** by EchPT2 to **1T1**, **1T2**, **1Q1**, and **1Q2** (ii). The last two products were also detected in the assay of **1T1** with EchPT2 (iv). Conversely, **1D2**, **1T2**, **1Q1**, and **1Q2** underwent no further conversion by EchPT2 (iii, v, and vi). These confirmed the hypothesis that **1D2** and **1T2** are end products of different branches of the EchPT2 reaction and $C7$-prenylation serves as a termination step in a consecutive prenylation cascade as shown in Scheme 1. To reach the tetraprenylated products, the putative NRPS product **1** will be reversely prenylated at C-2 by EchPT1. The resulting product **1M1** undergoes regular prenylation at C-5, then at C-4, and finally at C-6 or C-7. No pentaprenylated derivative was detected in the reaction mixtures of **1M1**, **1D1**, **1T1**, **1Q1**, or **1Q2** with EchPT2 (Figure S6a), indicating that the tetraprenylated derivatives **1Q1** and **1Q2** are the final products of the cascade. **1M1**, **1D1**, **1D2**, **1T1**, **1T2**, and a mixture of **1Q1** and **1Q2** were also incubated with EYE95342. LC-MS analysis revealed no conversion of these substrates (Figure S7).

Having explored the outstanding catalytic skill of EchPT2, we wondered if the products of such a prenylation cascade, especially the previously undescribed tetraprenylated derivatives **1Q1** and **1Q2**, coexist in the fungal cultures. Based on previous results

Scheme 1. Proposed Biosynthetic Pathway of Echinulin and Neoechinulin Series via Consecutive Multiprenylations by EchPT1 and EchPT2 in *A. ruber*
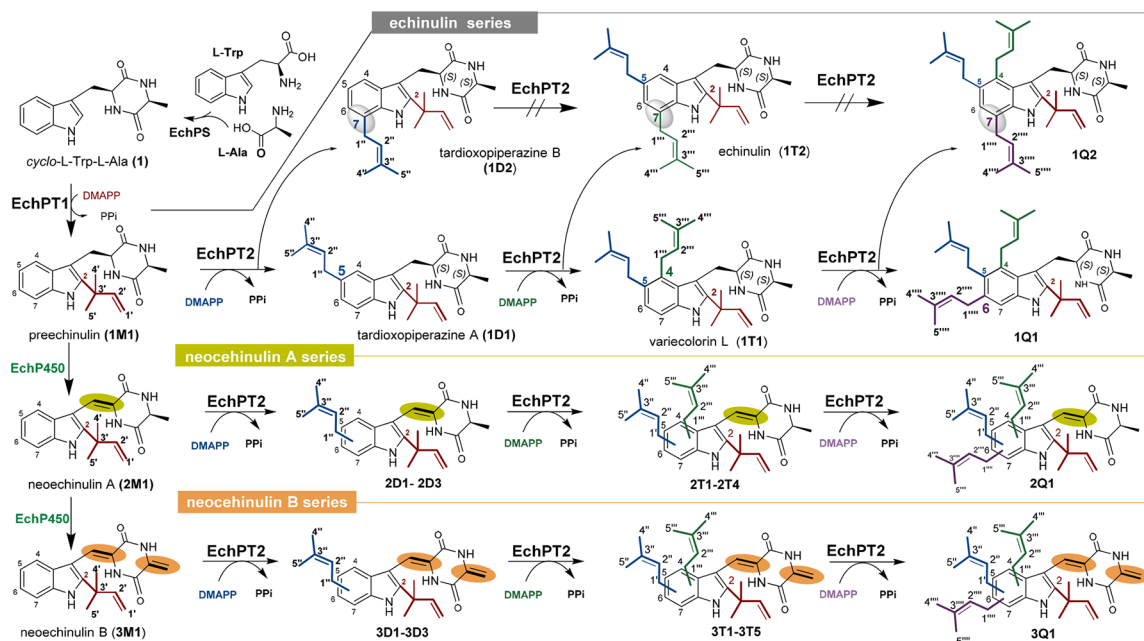
reporting a strong influence of salt concentration on metabolism,[19] we cultivated *A. ruber* QEN-0407-G2 under different conditions (SI). The fungal cultures were extracted and analyzed by LC-MS (Figure S1). The enzyme products described above, i.e. **1M1** of the EchPT1 reaction, as well as **1D1**, **1D2**, **1T1**, **1T2**, and **1Q2** of the EchPT2 reaction, were detected by LC-MS analysis, corresponding to the enzyme reactions with EchPT1 and EchPT2 (Figures S1, S6a, and S8).

As previously mentioned, neoechinulins with one (A series) or two *exo* double bonds (B series) at the DKP ring are also frequently identified in echinulin producers.[14] LC-MS analysis of the obtained fungal extract indeed revealed the presence of members of both series with up to three prenylation levels (Figure S1). In total, one mono- (neoechinulin A, **2M1**), two di-, and two triprenylated derivatives from the A series were detected. In the case of the B series, one mono- (neoechinulin B, **3M1**), two di-, and four triprenylated derivatives were observed. In contrast to the triprenylated **1T2** as the main metabolite of the echinulin series (Figure S1), the monoprenylated derivatives **2M1** and **3M1** are found to be the major products of the neoechinulin series A and B. Furthermore, *cyclo*-L-Trp-L-Ala (**1**), but not its di- or tetradehydrogenated derivatives, i.e. the unprenylated precursors of **2M1** and **3M1**, was detected in the fungal cultures. This could indicate that the first prenylation catalyzed by EchPT1 takes place before dehydrogenation, probably catalyzed by the cytochrome P450 enzyme EchP450 (Scheme 1). We speculated that both series of neoechinulins are also EchPT2 products of sequential prenylations and, thus, isolated the monoprenylated **2M1** and **3M1** from the extracts (SI for structure elucidation, Tables S3 and S8, Figures S30 and S31). LC-MS analysis of the reaction mixture of **2M1** and **3M1** with EchPT2 indeed demonstrated the clear acceptance of both substrates (Figures 4b, S6b) and the formation of eight and nine products with two to four prenyl residues, respectively. In comparison to the fungal extract, tetraprenylated products **2Q1** and **3Q1** were also clearly detected in the EchPT2 assays with **2M1** and **3M1**. This proves that EchPT2 also catalyzes a prenylation cascade with **1M1** analogs bearing exo double bonds and that the biosynthetic pathway illustrated in Scheme 1 can be expanded by neoechinulins. That is, conversion of **1M1** to neoechinulins A (**2M1**) and B (**3M1**) by a putative cytochrome P450 enzyme (EchP450)[21] marks the starting point of the neoechinulin formation. In analogy to **1M1**, **2M1** and **3M1** undergo a prenylation cascade catalyzed by EchPT2, resulting in the formation of products with different prenylation grades.

In conclusion, this study provides the first example of a prenyltransferase catalyzing an exceptional consecutive prenylation cascade. The unique feature of EchPT2 is its ability to accept its own mono-, di-, and triprenylated derivatives as substrates and to catalyze prenylations at different positions, leading to the formation of echinulin and congeners. It is of the utmost interest to solve such intriguing enzymatic structures and to comprehend the ability of EchPT2 to bind different substrates and catalyze diverse prenylations. This knowledge would also provide a basis for controlling the prenylation cascade by site-directed mutagenesis.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.orglett.7b02926.

Detailed experimental procedures including structural elucidation, kinetic parameters, MS and NMR data, and NMR spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: shuming.li@staff.uni-marburg.de.

ORCID Ⓞ

Bin-Gui Wang: 0000-0003-0116-6195
Shu-Ming Li: 0000-0003-4583-2655

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Netz, N.; Opatz, T. *Mar. Drugs* **2015**, *13*, 4814.
(2) Zhang, P.; Li, X.; Wang, B.-G. *Planta Med.* **2016**, *82*, 832.
(3) de Sá Alves, F. R.; Barreiro, E. J.; Fraga, C. A. M. *Mini-Rev. Med. Chem.* **2009**, *9*, 782.
(4) Xu, W.; Gavia, D. J.; Tang, Y. *Nat. Prod. Rep.* **2014**, *31*, 1474.
(5) Ma, Y. M.; Liang, X. A.; Kong, Y.; Jia, B. *J. Agric. Food Chem.* **2016**, *64*, 6659.
(6) Giessen, T. W.; Marahiel, M. A. *Front. Microbiol.* **2015**, *6*, 1.
(7) Terao, J.; Mukai, R. *Arch. Biochem. Biophys.* **2014**, *559*, 12.
(8) Winkelblech, J.; Fan, A.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 7379.
(9) Du, F.; Li, X.; Li, C.; Shang, Z.; Wang, B. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 4650.
(10) Cardani, C.; Casnati, G.; Piozzi, F.; Quilico, A. *Tetrahedron Lett.* **1959**, *1*, 1.
(11) Wang, W.-L.; Lu, Z.-Y.; Tao, H.-W.; Zhu, T.-J.; Fang, Y.-C.; Gu, Q.-Q.; Zhu, W.-M. *J. Nat. Prod.* **2007**, *70*, 1558.
(12) Li, D.-L.; Li, X.-M.; Li, T.-G.; Dang, H.-Y.; Wang, B.-G. *Helv. Chim. Acta* **2008**, *91*, 1888.
(13) Dewapriya, P.; Li, Y.-X.; Himaya, S. W. A.; Pangestuti, R.; Kim, S.-K. *NeuroToxicology* **2013**, *35*, 30.
(14) Chen, X.; Si, L.; Liu, D.; Proksch, P.; Zhang, L.; Zhou, D.; Lin, W. *Eur. J. Med. Chem.* **2015**, *93*, 182.
(15) Wijesekara, I.; Li, Y.-X.; Vo, T.-S.; Van Ta, Q.; Ngo, D.-H.; Kim, S.-K. *Process Biochem.* **2013**, *48*, 68.
(16) Mundt, K.; Wollinsky, B.; Ruan, H. L.; Zhu, T.; Li, S.-M. *ChemBioChem* **2012**, *13*, 2583.
(17) Li, H.; Ban, Z.; Qin, H.; Ma, L.; King, A. J.; Wang, G. *Plant Physiol.* **2015**, *167*, 650.
(18) Liu, C.; Minami, A.; Dairi, T.; Gomi, K.; Scott, B.; Oikawa, H. *Org. Lett.* **2016**, *18*, 5026.
(19) Kis-Papo, T.; Weig, A. R.; Riley, R.; Peršoh, D.; Salamov, A.; Sun, H.; Lipzen, A.; Wasser, S. P.; Rambold, G.; Grigoriev, I. V.; Nevo, E. *Nat. Commun.* **2014**, *5*, 3745.
(20) Maiya, S.; Grundmann, A.; Li, S.-M.; Turner, G. *ChemBioChem* **2006**, *7*, 1062.
(21) Ali, H.; Ries, M. I.; Nijland, J. G.; Lankhorst, P. P.; Hankemeier, T.; Bovenberg, R. A.; Vreeken, R. J.; Driessen, A. J. *PLoS One* **2013**, *8*, e65328.

SUPPLEMENTARY MATERIALS

Two Prenyltransferases Govern a Consecutive Prenylation Cascade in the Biosynthesis of Echinulin and Neoechinulin

Viola Wohlgemuth,[†] Florian Kindinger,[†] Xiulan Xie,[‡] Bin-Gui Wang,[§] Shu-Ming Li[†*]

[†]Institut für Pharmazeutische Biologie und Biotechnologie,
Philipps-Universität Marburg, Robert-Koch-Straße, 4, 35037 Marburg, Germany.

[‡]Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Straße, 35032 Marburg, Germany

[§]Key Laboratory of Experimental Marine Biology, Institute of Oceanology of the CAS, 266071 Qingdao, China.

[*]Corresponding author: *shuming.li@staff.uni-marburg.de*

S1

CONTENTS

S2

81

S3

1 EXPERIMENTAL SECTION

1.1 Computer-Assisted Sequence Analysis and Genome Mining for Gene Clusters

1.1.1 Sequence Analysis of Prenyltransferases

Sequence identities for the prenyltransferases were obtained by alignments of amino acid sequences using BLAST (www.ncbi.nlm.nih.gov) and antiSMASH.[1] The programs 2ndfind (biosyn.nih.go.jp/2.find) and FGENESH-M (Softberry, Mount Kisco, NY; www.softberry.com) were used for further sequence analysis and intron prediction.

1.1.2 Gene Cluster Prediction

Based on the NRPS protein sequence (GenBank: EYE98744) from the putative echinulin gene cluster of *Aspergillus ruber* CBS 135680,[2] a BLAST (NCBI BLAST), analysis was performed to identify further homologous proteins. Two homologous NRPSs from *A. glaucus* (GenBank: OJJ80815)[3] and *A. cristatus* (GenBank: ODM14527)[4] with 2141 aa and 2131 aa, respectively, clearly differ from the annotated NRPS in *A. ruber*, which has a length of 1285 aa. For further analysis, a multiple protein alignment with the three mentioned NRPSs was performed with "T-Coffee" (http://tcoffee.crg.cat/apps/tcoffee/do:regular). The alignments revealed a high identity between the two NRPS from *A. glaucus* and *A. cristatus*. Additionally, the annotated protein sequence from *A. ruber* also shows a high identity with the beginning at the amino acid position 790. These results indicate that the NRPS of *A. ruber* is perhaps not correctly annotated and lacks 789 aa at the N-terminus. Therefore, the DNA sequence of the putative echinulin gene cluster was used to predict the protein sequence of the NRPS by employing the web tool "2ndFind"(http://biosyn.nih.go.jp/2ndfind/), which uses the AUGUSTUS protein prediction algorithm for predicting eukaryotic proteins (http://bioinf.uni-greifswald.de/augustus/). Based on this prediction, the re-annotated putative NRPS has a polypeptide length of 2113 aa, encoded by 5 exons (bp 1209589-1210198, 1206494-1209533, 1206164-1206374, 1206019-1206108 and 1203548-1205935 of KK088413.1). The re-annotated protein was used for sequence analysis and comparison (Figure S2, Table S1).

S4

1.2 Chemicals

DMAPP was synthesized according to the previously described method for geranyl diphosphate (GPP)[5].

1.3 Bacterial Strains, Plasmids, and Culture Conditions

The vectors pGEM-T Easy, pQE-9 and pQE-70 were purchased from Promega (Mannheim, Germany) and QIAGEN (Hilden, Germany), and used for cloning and expression experiments, respectively. *Escherichia coli XL1 Blue* MRF' (Agilent Technologies, Amsterdam, the Netherlands) and M15 (pREP4) cells (QIAGEN, Hilden, Germany) were used as cloning and expression hosts. For cultivation, they were grown in liquid Lysogeny broth (LB) or Terrific broth (TB) medium and on solid LB medium with 1.5 % (w/v) agar at 30 or 37 °C. For selection of recombinant *E. coli* strains, 50 µg mL^{-1} of carbenicillin or 25 µg mL^{-1} of kanamycin were used.

1.4 Cultivation of *A. ruber*

1.4.1 Cultivation of *A. ruber* for DNA and RNA Isolation.

100 mL liquid potato dextrose medium (23.93g L^{-1}) in a 300 mL cylindrical flask were inoculated with 100 µL spore suspension of *A. ruber* QEN-0407-G2,[6] and incubated at 25 °C on a rotary shaker (150 rpm) in darkness. For messenger RNA and genomic DNA isolation, fungal mycelia were harvested after 8 and 14 days.

1.4.2 Cultivation of *A. ruber* for Secondary Metabolite Production

Metabolite production by *A. ruber* is strongly dependent on the culture conditions, particularly on salt conditions.[2,7] We therefore tested cultivation with three different media PD, PD-ASW and ASW, and at four different cultivation times (1, 2, 3 and 4 weeks). 500 mL media in Fernbach flasks were inoculated with spore suspension of *A. ruber* and statically incubated in darkness at 25 °C.

S5

PD	**potato dextrose liquid media** consisting of 23.93 g L^{-1} potato dextrose broth
PD-ASW	**potato dextrose liquid media/artificial seawater (1:1)**
ASW	**artificial sea water**, consisting of (g L^{-1} solution) sodium chloride (23.93), sodium sulfate (4.01), potassium chloride (0.70), sodium bicarbonate (0.20), potassium bromide (0.10), boric acid (0.03), sodium fluoride (0.003), magnesium chloride (0.05), calcium chloride (0.01)

LC-MS analysis of the extracts for the presence of echinulin and congeners confirmed that cultivation in PD for four weeks was the best one and was therefore used for all following cultivations.

1.4.3 Upscaled Cultivation of *A. ruber* and Extraction for Structure Elucidation

For isolation and structure elucidation of echinulin analogs, the fungal strain was statically cultivated in liquid PD media for four weeks. The cultural broth was extracted three times with one volume of ethyl acetate each and the combined organic phases were concentrated on a rotating vacuum evaporator at 35 °C and dissolved in methanol. The mycelium was frozen in liquid nitrogen, homogenised exhaustively by pestling, and extracted with methanol/acetone (1:1) for 2 h. The extract was concentrated on a rotating vacuum evaporator at 35 °C and dissolved in methanol. The combined extracts of culture broth and mycelium were subjected to a silica gel column chromatography and eluted with mixtures of dichloromethan and acetone in different ratios of increasing polarity to yield 20 fractions on the basis of TLC analysis. Echinulin (**1T2**), neoechinulin A (**2M1**), neoechinulin B (**3M1**), and the tetraprenylated derivative **1Q2** were then purified from the respective fractions on HPLC.

1.5   DNA and RNA Isolation, cDNA Synthesis, PCR Amplification and Gene Cloning

Standard procedures for DNA manipulation and propagation in *E. coli* were performed as described previously.[8] Genomic DNA was isolated from *A. ruber* mycelia harvested after 14 days using phenol/chloroform extraction. RNA was isolated from 8 day-old fungal mycelia by using E.Z.N.A. Fungal RNA Kit (Omega Bio-Tek) according

S6

to the manufacturer's protocol, and complementary DNA (cDNA) was synthesized by utilizing the ProtoScript M-MuLV First Strand cDNA Synthesis Kit (NEB) with oligo(dT) primers.

To get the genomic sequences of the homologous genes of EYE98742 (EchPT1), EYE98746 (EchPT2) and EYE95342 in *A. ruber* QEN-0407-G2, PCR amplification was carried out using genomic DNA as template. The primer pairs vwRbPT1-f (5'-CCGCA TGCCC TCTGA AGTAT TGAC-3') and vwRbPT1-rL (5'-GTGCT CACCC AGCTC CAGC-3') were used for *echPT1,* vwRbPT2-f (5'-CGCAT GCAGC CTTAC CACAC A-3') and vwRbPT2-rL (5'-GATGAA ACCAA GGTCA GAAGA CCTTC-3') for *echPT2*, and vwRBPT3-f (5'-CGCAT GCCTT TACAA ACGAC CAA-3') and vwRBTP3-rL (5'-CACAT TTCCC ATAAT CCATA TTTAT ACCGC-3') for the EYE95342 homologue. The PCR amplifications were carried out at 59 °C as annealing temperature on an iCycler from BioRad with Expand High Fidelity PCR Kit (Roche Diagnostic, Mannheim, Germany). The fragments were cloned into pGEM-T Easy, resulting in the plasmids pVW75, pVW76, and pVW77, which were subseqeuently sequenced to verify the putative gene sequences (Seqlab Sequence Laboratories, Göttingen, Germany).

The entire coding sequences of *echPT1*, *echPT2* and the gene for EYE95342 homologue were obtained by using cDNA as template with an annealing temperature at 60 °C. The primer pairs used were vwRbPT1-f2 (5'-C<u>GTCG AC</u>CCC TCTGA AGTAT TGAC-3') and vwRbPT1-r2 (5'- CC<u>AAG CTT</u>TT ATTTG ACCTG ATCAC TTCC-3') for *echPT1*, vwRbPT2-f2 (5'- <u>CTGCA G</u>CAGC CTTAC CACAC-3') and vwRbPT2-r2 (5'- CC<u>AAG CTT</u>CT AAAAC AATGA <u>CATGT AGATG</u> -3') for *echPT2*, and vwRBPT3-f (5'-C<u>GCAT G</u>CCTT TACAA ACGAC CAA-3') and vwRBTP3-r (5'- C<u>GGAT CC</u>TAC CGCAG TTTTC TGC -3') for EYE95342 homologue. Underlined letters represent the introduced restriction sites of SalI/HindIII, PstI/HindIII and SphI/BamHI. The resulting PCR products for *echPT1* (1266bp), *echPT2* (1238bp) and EYE95342 homologue (1306bp) were cloned into pGEM-T Easy giving the plasmids pVW88, pVW82, and pVW81, which were subsequently sequenced to confirm sequence integrity. Cloning of the coding seqeuences at the SalI/HindIII (*echPT1*) and PstI/HindIII (*echPT2*) sites in pQE-9 or alternatively at SphI/BamHI (EYE95342 homologue) in pQE-70 gave the expression plasmids pVW90 (*echPT1*), pVW83 (*echPT2*), and pVW84 (EYE95342 homologue) as illustrated in Figures S3-5.

S7

1.6 Overproduction and Purification of His-tagged Recombinant Proteins, Protein Analysis and Enzyme Assays

1.6.1 Protein Overproduction and Purification

To improve gene expression, the initial pH of the media, induction temperature and time were optimized in shaking cultures seperately for each construct.

For *echPT1* expression, *E. coli* M15 cells harboring pVW90 were cultivated in 500 mL TB media supplemented with carbenicillin (50 μg mL^{-1}) in a 2 L flask. After growing at 37 °C and 230 rpm, to A$_{600}$=0.60, isopropyl thiogalactosid was added to a final concentration of 0.5 mM for induction of gene expression. The cells were cultivated for another 16 h at 37 °C. The cell pellets were resuspended in lysis buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 10 mM imidazole, pH 8.0) at 2–5 mL per gram wet weight. After addition of lysozyme (1 mg mL^{-1}), the cell suspension was incubated on ice for 30 min, and then sonicated eight times for 10 s each at 200 W. Cellular debris were separated from the soluble proteins by centrifugation (20 000 *xg*, 30 min) at 4 °C. The recombinant His$_6$-tagged fusion protein was purified by affinity chromatography with Ni-NTA agarose resin (Quiagen) according to the manufacturer's protocol. The obtained protein fraction was passed through a Sephadex G25 column (PD-10 , GE Healthcare), that had been equilibrated with 50 mM Tris-HCl, 15 % glycerol (v/v), pH 7.5, previously, in order to change the buffer. Purified His$_6$-EchPT1 was eluted with the same buffer and stored at -80 °C. A protein yield was calculated to be 25 mg of purified protein per liter of culture.

For overproduction of His$_6$-EchPT2, *E. coli* M15 cells harboring pVW83 were cultivated in 2 L flasks, containing 250 mL liquid TB medium, supplemented with carbenicillin (50 μg mL^{-1}) whithout any induction, to minimize insoluble protein amount. The cultures were grown at 30 °C for 24 h and harvested by centrifugation. Protein purification was carried out by affinity chromatography with Ni-NTA agarose resin (Quiagen) as aforementioned. Protein yield was calculated to be 1.8 mg of purified protein per liter of culture.

S8

For overproduction of the EYE95342 homologue, *E. coli* M15 cells harboring pVW84 were cultivated in 2 L flasks, containing 500 mL liquid TB medium, supplemented with carbenicillin (50 µg mL^{-1}). After growing at 37 °C and 230 rpm to A_{600}=0.60, the gene expression was induced by addition of isopropyl thiogalactoside to a final concentration of 0.6 mM. The cultures were grown at 30 °C and 230 rpm for another 16 h and then centrifuged to harvest the cells. Protein purification was carried out by affinity chromatography with Ni-NTA agarose resin (Quiagen) as aforementioned. Protein yield was calculated to be 4 mg of purified protein per liters of culture.

1.6.2. Protein Analysis

Purity of the obtained recombinant proteins was analysed on SDS-PAGE (15 %) according to the method described by Laemmli,[9] and stained with Coomassie Brilliant Blue G-250, while protein quantification was determined on a NanoDrop 2000c Spectrophotometer (Thermo Fisher Scientific, Waltham, USA). Major protein bands with migrations near the 45 kDa size marker were observed for the three purified proteins, as shown in Figures S3-5, and correspond well to the calculated masses of 49.4 kDa, 47.7 kDa, and 50.9 kDa for EchPT1-His$_6$, EchPT2-His$_6$, and His$_6$-EYE95342, respectively. The molecular mass of the native recombinant proteins for EchPT1 and EchPT2 were determined by size-exclusion chromatography on an ÄKTA Purifier FPLC system (GE Healthcare, Freiburg, Germany) by using a HiLoad 16/60 Superdex 200 column (GE Healthcare) and Tris-HCl buffer (50 mM, pH 7.5) containing 150 mM NaCl as elution buffer. The column was calibrated with blue dextran 2000 (2000 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), carbonic anhydrase (29 kDa) and ribonuclease A (13.7 kDa (GE Healthcare). Separation was performed with a flow rate of 1 mL min^{-1}. The multimeric state of the native recombinant EchPT1-His$_6$ was determined as 276.5 kDa, and that of EchPT2-His$_6$ as 149.0 kDa, indicating presumably pentameric (EchPT1) and trimeric (EchPT2) species. Previous results on DMATS enzymes proved that each subunit contains a reaction chamber and thereby forms one catalytic unit[10].

S9

1.7 Enzyme Assays of EchPT1, EchPT2 and EYE95342.

1.7.1 Assays for EchPT1, EchPT2 and EYE95342 Activity Toward 1

To determine the enzyme activity toward **1**, standard assays (100 µL) contained Tris-HCl (50 mM, pH 7.50), **1** (0.5 mM), CaCl$_2$ (5 mM), DMAPP (3 mM), glycerol (0.5 - 5 %, v/v), DMSO (2.5 %, v/v) and the purified recombinant EchPT1 (1 µg, 0.20 µM), EchPT2 (10 µg, 2.13 µM), EYE95342 (20 µg, 4 µM), or their combinations. After incubation at 37 °C for 2 h, the reaction mixtures were extracted three times with 200 µL ethyl acetate each and the combined organic phases were evaporated to dryness. After desolving the residues in 100 µL methanol, 90 µL thereof were analysed on an LC-MS system described below. Additional assays containing EYE95342 were incubated for 4, 8 and 16 h.

1.7.2 Ion Dependency Assays

For determination of ion dependency of EchPT1 and EchPT2 reactions, enzyme assays contained Tris-HCl (50 mM, pH 7.50), DMAPP (1 mM for EchPT1 or 3 mM for EchPT2), 5 mM of metal ions or EDTA, and 1 mM **1** (for EchPT1 reaction) or 1 mM **1M1** (for EchPT2 reaction). The assays were incubated at 37 °C for 20 min with EchPT1 (1 µg, 0.20 µM), or 15 and 20 min with EchPT2 (5 µg 1.07 µM). The respective enzyme activity without additive was defined as 100 %.

1.7.3 Enzyme Assays for determination of Kinetic Parameters

The assays for determination of the kinetic parameters of EchPT1 toward **1** contained 1 mM DMAPP, 2 µg EchPT1 (Table S2) and the aromatic substrates at final concentrations from 0.002 to 5.0 mM, and were carried out at 37 °C for 1 min (Table S2). For kinetic parameters toward DMAPP, reaction mixtures with 2 µg EchPT1, **1** (1 mM), CaCl$_2$ (5 mM) and DMAPP at final concentrations of 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 1.0, 2.0 and 5.0 mM were carried out for 1 min at 37 °C. The assays were performed as duplicates and subsequently terminated by addition of 100 µL methanol. The precipitated protein was removed by centrifugation at 20000 *xg* for 20 min, and 100 µL

S10

were analyzed on HPLC as described below. The conversion yields were calculated from the peak areas of substrates and products in HPLC chromatogram, and their signal intensities in the ^{1}H NMR spectra of the incubation mixtures.

The assays for determination of the kinetic parameters of EchPT2 toward **1M1** were carried out with 3 mM DMAPP, 2 µg EchPT2 (Table S2) and the aromatic substrates at final concentrations of 0.002 to 5.0 mM, at 37 °C for 3 min. For kinetic parameters of EchPT2 toward DMAPP, reaction mixtures with 2 µg EchPT2, **1M1** (1 mM), CaCl$_2$ (5 mM) and DMAPP at final concentrations of 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 1.0, 2.0 and 5.0 mM were carried out for 3 min at 37 °C.

1.7.4 Preparation and Isolation of Enzyme Products for Structure Elucidation.

Enzyme assays for product isolation were scaled up to a volume of 12 mL reaction mixture, containing Tris-HCl (50 mM, pH 7.5), DMAPP (2 - 3 mM), the respective aromatic substrate (1 mM), and 1 - 10 mg purified recombinant proteins. The enzyme assays were incubated at 37 °C for 16 h and extracted three times with two volumes of ethyl acetate each. The resulting organic phases were combined and concentrated on a rotating vacuum evaporator at 35 °C to dryness and dissolved in 200 µL methanol for isolation.

1.8   HPLC and LC-MS Conditions for Analysis and Isolation of Enzyme Products and Fungal Metabololites.

1.8.1 HPLC Conditions

The enzyme assays and fungal extracts were analyzed on an Agilent HPLC series 1200 (Agilent Technologies) by using an Agilent Eclipse XDB-C18 column (5 µm, 4.6 x 150 mm). H$_2$O (solvent A) and MeCN (B), both containing 0.1 % (v/v) formic acid, were adjusted to a flow rate of 0.5 mL/min. The separation began with a linear gradient from 20 - 80 % MeCN over 10 min, followed by an isocratic elution at 80 % MeCN for 25 min, and then with a linear gradient from 80 - 100 % MeCN in 5 min. After each run, the column was washed with 100 % B for

S11

5 min and equilibrated with 20 % MeCN for 5 min. Detection was carried out with a photodiode array detector, and absorptions at 277 nm were illustrated in this study.

To isolate the enzyme products and fungal metabolites, the upscaled reaction mixtures and the fractions from a silica gel column were separated with the same HPLC equipment and a semipreparative Multospher 120 RP-18 column (5 µm, 10 x 250 mm). H_2O (solvent A) and MeCN (B) were used as solvents with a flow rate of 2.5 mL/min. A linear gradient of 30 – 80 % MeCN in 15 min was followed by an isocratic elution at 80 % MeCN for 18 min, and an additional linear gradient from 80 - 100 % MeCN in 7 min. The column was washed with 100 % B for 10 min and subsequently equilibrated with 20 % MeCN for another 10 min. The obtained fractions were further purified on an Agilent Eclipse XDB-C18 column (5 µm, 4.6 x 150 mm). The same method was used as for the analysis of the enzyme assays, modified by using H_2O and MeCN without formic acid as solvents.

1.8.2 LC-MS Condition for Analysis of the Enzyme Products

The enzyme products were analyzed on an Agilent HPLC 1260 series system equipped with a photo diode array detector and a Bruker microTOF QIII mass spectrometer by using the Agilent Eclipse XDB C18 column (5 µm, 4.6 x 150 mm) according to the standard condition for this column described above.

1.9   NMR and Mass Spectrometric Analysis

The MS data were extracted from the LC-MS analysis and are listed in Table S3.

For structural elucidation, the isolated enzyme products were dissolved in DMSO-d_6 or CD_3OD and subjected to NMR analysis including ^1H-^1H-DFC COSY, HSQC and HMBC spectra. The spectra were recorded at room temperature on a JEOL ECA-500 (JEOL, Akishima, Tokyo, Japan), or a Bruker Avance III 500 MHz spectrometer installed with a cyro probe 5 mm Prodigy for Broad Band Observation. All spectra were processed with MestReNova 6.0.2 (Metrelab Research, Santiago de Compostella, Spain) and the chemical shifts were referenced to those of the solvent. NMR data are given in Tables S4-8.

S12

1.10 Structure Elucidation of the Identified Products and NMR Assignments

1.10.1 Enzyme Product 1M1 of the EchPT1 Reaction

1.5 mg of **1M1** were isolated as white powder and subjected to NMR analysis. In the ^{1}H NMR spectrum (Figure S18), signals of a *tert*-DMA moiety have clearly been detected as three double doublets of olefinic protons around 5.15, 5.10 and 6.20 ppm and two methyl singlets at 1.55 ppm. In the downfield region for aromatic protons, signals of four coupling protons were clearly detected in the spectrum (Figure S18). In comparison with the respective substrate, the signals of H-2 of the indole ring disappear. The data listed in Table S4 are congruent with those reported previously.[11] This proves unequivocallly the reverse *C2*-prenylation in the structures of **1M1.**

1.10.2 1D1 and 1D2 as well as 1T1 and 1T2 of the EchPT2 Reactions

0.2 mg of **1D1** and 0.8 mg of **1D2** were isolated as pale yellow powders. In the ^{1}H NMR spectra of **1D1** and **1D2** (Figures S19 - 20), signals of a regular prenyl moiety each were observed with a typical triplet at about 5.30 ppm and two methyl singlets around 1.7 ppm. The three coupling aromatic protons in the spectrum of **1D2** prove the prenylation at the benzene ring. In the case of **1D1**, two doublets and one singlet for aromatic protons also indicate the prenylation at the benzene ring. Comparison of the NMR data listed in Tables S4 and S5 with known compounds in literature proves **1D1** to be tardioxopiperazine A [12,13] and **1D2** to be tardioxopiperazine B.[13]

1T1 (0.7 mg, white powder) and **1T2** (1.5 mg, white powder) were isolated from reaction mixtures of EchPT2 with **1M1**. 1.9 mg of **1T2** were also isolated as white powder from fungal extracts. In comparison to that of **1D1**, the ^{1}H NMR spectra of **1T1** and **1T2** (Figures S21 and 22) show signals for an additional regular prenyl moiety. Their identity as triprenylated *cyclo*-L-Trp-L-Ala were confirmed by HRMS analysis. The two doublets with coupling constant of 8.2 Hz in the spectrum of **1T1** indicate the C2, C4, C5-triprenylation. Comparing the NMR data listed in Table S5 with known compounds in literature proves **1T1** to be variecolorin L.[6,14]

S13

The two singlets for one aromatic proton each were detected in the ^{1}H NMR spectrum of **1T2** at 7.03 and 7.62 ppm. These could represent H-4 and H-7 and indicate a *C2*, *C5*, *C6*-triprenylation. However, this would be somewhat surprising, because **1T2** was detected as the main product of *cyclo*-L-Trp-L-Ala with EchPT1 and EchPT2 as well as preechinulin (**1M1**) with EchPT2. We expected echinulin with a C2, C5, C7-triprenylation as the main product, due to the fact that this compound was identified as the major metabolite in the producers including *A. ruber*.[6,15] Literature search reveals that the H-4 and H-6 of echinulin indeed appear as two singlets.[6,15] To confirm the structure, we isolated echinulin from the fungal culture and subjected it to NMR and MS analyses. The ^{1}H NMR spectra of the isolated echinulin and the enzyme product **1T2** are perfectly in consistence with each other. Extensive spectroscopic analysis in different solvents including ^{1}H-^{1}H-DQF COSY, HSQC, and HMBC (Table S6, Figures S23 - 28) unequivocally confirmed the structure of echinulin to be reversely *C2*-, regularly *C5*- and *C7*-triprenylated *cyclo*-L-Trp-L-Ala.

1.10.3 The tetraprenylated 1Q2 isolated from fungal extract

1Q2 (approximate 0.05 mg) was isolated as white powder. Characteristic tripletts for olefinic protons of three regular prenyl moieties were observed at 5.37, 5.13, and 4.86 ppm, respectively (Figures S29, Table S7). A close comparison of these signals with those of **1T1** and **1T2** (Figures S21, S22, and S29) revealed that the three regular residues of **1Q2** are located at C-4, C-5 and C-7 of the indole ring. This was further verified by the observation of only one singlet for one aromatic proton at 6.62 ppm (Figure S29), which corresponds perfectly to that of H-6 in **1T2**.

1.10.4 Neoechinulin A (2M1) and B (3M1) Isolated from Fungal Culture

2M1 (1.8 mg, white powder) and **3M1** (1.2 mg, yellow powder) were purified from fungal extract and subjected to NMR analysis. In comparison to those of preechinulin (**1M1**), the signals for the three coupling protons (H10 and H11) disappear in the ^{1}H NMR spectra of **2M1** and **3M1** (Figures S30 and S31). Instead, a singlet of one

S14

olefinic proton was observed at 6.88 and 6.96 ppm, respectively. In the spectrum of **3M1**, the doublet of the methyl group at C-14 in **1M1** was changed to two singlets of olefinic protons at 5.23 and 4.92 ppm. Comparing the NMR data listed in Table S8 with known compounds in literature proves **2M1** and **3M1** to be neoechnulin A[16,17] and B,[18,19] respectively (Table S8).

1.11 Accession Numbers.

The nucleotide sequences of the genes from *A.ruber* QEN-0407-G2 reported in this study have been deposited in the GenBank database under accession numbers MF538774 (*echPT1*), MF538775 (*echPT2*) and MF538776 (EYE95342 homologue).

1.12 References of supplementary materials

(1)   Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015**, *43*, W237.

(2)   Kis-Papo, T.; Weig, A. R.; Riley, R.; Peršoh, D.; Salamov, A.; Sun, H.; Lipzen, A.; Wasser, S. P.; Rambold, G.; Grigoriev, I. V; Nevo, E. *Nat. Commun.* **2014**, *5*, 3745.

(3)   Vries, R. P. De; Riley, R.; Wiebenga, A.; Aguilar-Osorio, G.; Amillis, S.; Uchima, C. A.; Anderluh, G.; Asadollahi, M.; Askin, M.; Barry, K.; Battaglia, E.; Bayram, Ö.; Benocci, T.; Braus-Stromeyer, S. A.; Caldana, C.; Cánovas, D.; Ricardo, A.; Damásio, D. L.; Diallinas, G.; Emri, T.; Fekete, E.; Flipphi, M. *Genome Biol.* **2017**, 18:28.

(4)   Ge, Y.; Wang, Y.; Liu, Y.; Tan, Y.; Ren, X.; Zhang, X.; Hyde, K. D. *BMC Genomics* **2016**, *17*, 1.

(5)   Woodside, A. B.; Huang, Z.; Poulter, C. D. *Org.Synth.* **1988**, *66*, 211.

(6)   Li, D.-L.; Li, X.-M.; Li, T.-G.; Dang, H.-Y.; Wang, B.-G. *Helv.Chim.Acta* **2008**, *91*, 1888.

(7)   Kamauchi, H.; Kinoshita, K.; Sugita, T.; Koyama, K. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 4911.

(8)   Sambrook, J.; Russell, D. *Molecular Cloning: A Laboratory Manual*; New York, 2012.

(9)   Laemmli, U. K. *Nature* **1970**, *227*, 680.

(10)   Winkelblech, J.; Xie, X.; Li, S. *Org. Biomol. Chem.* **2016**, *14*, 9883.

(11)   Yin, S.; Yu, X.; Wang, Q.; Liu, X.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 1649.

(12)   Dai, Q.; Xie, X.; Xu, S.; Ma, D.; Tang, S.; She, X. *Org.Lett.* **2011**, *13*, 2302.

(13)   Fujimoto, H.; Fujimaki, T.; Okuyama, E.; Yamazaki, M. *Chem.Pharm.Bull.* **1999**, *47*, 1426.

(14)   Wang, W.-L.; Lu, Z.-Y.; Tao, H.-W.; Zhu, T.-J.; Fang, Y.-C.; Gu, Q.-Q.; Zhu, W.-M. *J. Nat. Prod.* **2007**, *70*, 1558.

(15)   Meng, L.-H. H.; Du, F.-Y. Y.; Li, X.-M. M.; Pedpradab, P.; Xu, G.-M. M.; Wang, B.-G. *J. Nat. Prod.* **2015**, *78*, 909.

(16)   Aoki, T. *Synlett* **2006**, *5*, 677.

(17)   Kuramochi, K.; Aoki, T.; Nakazaki, A.; Kamisuki, S.; Takeno, M.; Ohnishi, K.; Kimoto, K.; Watanabe, N.; Kamakura, T.; Arai, T.; Sugawara, F.; Kobayashi, S. *Synthesis.* **2008**, No. 23, 3810.

(18)   Marchelli, R.; Dossena, A.; Pochini, A.; Dradi, E. *J.Chem.Soc., Perkin Trans.1* **1977**, No. 7, 713.

(19)   Gao, H.; Zhu, T.; Li, D.; Gu, Q.; Liu, W. *Arch.Pharm.Res.* **2013**, *36*, 952.

S16

2 TABLES OF GENETIC INFORMATION

Table S1 Proteins of the putative echinuline gene clusters and their proposed function

Protein (length in aa) of	Identity to proteins (length in aa) of (%)				Putative function
Aspergillus ruber	*Aspergillus glaucus*		*Aspergillus cristatus*		
EYE98742 (**EchPT1**) (417)	OJJ80811 (417)	93.8	ODM14529 (431)	88.4	**prenyltransferase**
EYE98743 (480)	OJJ80812 (461)	92.7	ODM14528 (463)	84.2	putative transporter
-	OJJ80813 (155)	-	-		hypothetical protein
-	OJJ80814 (392)	-	-		hypothetical protein
EYE98744 (2113)	OJJ80815 (2141)	88.7	ODM14527 (2131)	85.6	NRPS
-	-		ODM14526 (672)	-	hypothetical protein
EYE98745 (534)	OJJ80816 (519)	94.2	ODM14525 (532)	95.7	Cytochrome P450 enzyme
EYE98746 (**EchPT2**) (408)	OJJ80817 (467)	83.9	ODM14524 (408)	84.4	**prenyltransferase**
EYE98747 (767)	OJJ80818 (767)	95.3	ODM14523 (767)	94.1	hypothetical protein

3 TABLES OF KINETIC PARAMETERS

Table S2 Kinetic parameters of EchPT1 toward **1** and EchPT2 toward **1M1** in the presence of DMAPP as well as those toward the prenyl donor DMAPP

Enzyme	Substrates	Protein amount and incubation time		K_M [mM]	k_{cat} [s⁻¹]	k_{cat}/K_M [s⁻¹M⁻¹]
		Kinetic parameters for acceptors				
EchPT1	*cyclo*-L-Trp-L-Ala (**1**)	2 µg	1 min	0.09 ± 0.0340	6.63 ± 0.0378	73666.7
EchPT2	*cyclo*-L-2-*tert*-DMA-Trp-L-Ala (**1M1**)	2 µg	3 min	0.03 ± 0.0021	0.36 ± 0.0015	12000.0
		Kinetic parameters for the donor				
EchPT1	DMAPP with **1**	2 µg,	1 min	0.18 ± 0.0056	2.96 ± 0.0360	16444.4
EchPT2	DMAPP with **1M1**	2 µg,	3 min	0.10 ± 0.004	0.47 ± 0.0061	4700.0

S18

4 TABLES OF HR-ESI-MS DATA

Table S3 HR-ESI-MS data of EchPT1 and EchPT2 reaction products

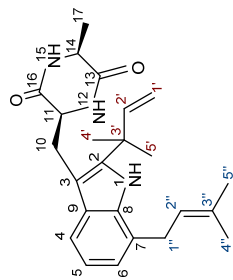| compound	chemical formula	HR-ESI-MS [M+H]$^+$		Deviation (ppm)
		calculated	measured	
preechinulin (**1M1**)	$C_{19}H_{23}N_3O_2$	326.1863	326.1860	-0.9
tardioxopiperazine A (**1D1**)	$C_{24}H_{31}N_3O_2$	394.2489	394.2466	-5.8
tardioxopiperazine B (**1D2**)	$C_{24}H_{31}N_3O_2$	394.2489	394.2485	-1.0
variecolorin L (**1T1**)	$C_{29}H_{39}N_3O_2$	462.3115	462.3133	3.8
echinulin (**1T2**)	$C_{29}H_{39}N_3O_2$	462.3115	462.3120	3.9
1Q1/2	$C_{34}H_{47}N_3O_2$	530.3741	530.3749	1.5
neoechinulin A (**2M1**)	$C_{19}H_{22}N_3O_2$	324.1707	324.1702	-1.5
2D1	$C_{24}H_{30}N_3O_2$	392.2330	392.2313	-4.3
2D2	$C_{24}H_{30}N_3O_2$	392.2330	392.2311	-4.8
2D3	$C_{24}H_{30}N_3O_2$	392.2330	392.2332	0.5
2T1	$C_{29}H_{38}N_3O_2$	460.2959	460.2949	-2.1
2T2	$C_{29}H_{38}N_3O_2$	460.2959	460.2968	1.9
2T3	$C_{29}H_{38}N_3O_2$	460.2959	460.2952	-1.5
2T4	$C_{29}H_{38}N_3O_2$	460.2959	460.2959	0
2Q1	$C_{34}H_{46}N_3O_2$	528.3585	528.3593	1.5
neoechinulin B (**3M1**)	$C_{19}H_{20}N_3O_2$	322.1550	322.1552	0.6
3D1	$C_{24}H_{28}N_3O_2$	390.2176	390.2175	-0.3
3D2	$C_{24}H_{28}N_3O_2$	390.2176	390.2168	-2.0
3D3	$C_{24}H_{28}N_3O_2$	390.2176	390.2177	-0,2
3T1	$C_{29}H_{36}N_3O_2$	458.2802	458.2783	-4.1
3T2	$C_{29}H_{36}N_3O_2$	458.2802	458.2789	-2.8
3T3	$C_{29}H_{36}N_3O_2$	458.2802	458.2781	-4.5
3T4	$C_{29}H_{36}N_3O_2$	458.2802	458.2791	-2.4
3T5	$C_{29}H_{36}N_3O_2$	458.2802	458.2798	-0.8
3Q1	$C_{34}H_{44}N_3O_2$	526.3248	526.3258	1.9

S19

5 TABLES OF NMR DATA

Table S4 ^{1}H NMR data of the enzyme product of EchPT1 reaction (**1M1**) in CD$_3$OD and the enzyme products of EchPT2 tardioxopiperazine A (**1D1**) in DMSO-d_6

cyclo-L-2-tert DMA-Trp-L-Ala **1M1**

cyclo-L-2-tert-DMA-5-DMA-Trp-L-Ala **tardioxopiperazine A (1D1)**

Pos.	1M1 δ$_H$, multi, J in Hz		1D1 δ$_H$, multi, J in Hz	
NH-1	1H	-	1H	10.43, s
4	1H	7.51, dd 7.9, 1.1	1H	7.20, s
5		6.99, ddd, 7.9, 7.0, 1.1		-
6	1H	7.06, ddd, 7.9, 7.0, 1.1	1H	6.83, d, 8.2
7	1H	7.33, dd, 7.9, 1.1	1H	7.22, d, 8.2
10a	1H	3.54, dd, 14.8, 3.8	1H	3.33#
10b	1H	3.30#	1H	3.02, m
11	1H	4.28, ddd, 9.6, 3.8, 1.2	1H	3.93, m
NH-12	1H	-	1H	8.19, d, 2.4
14	1H	3.95, dq, 7.0, 0.7	1H	3.81, m
NH-15	1H	-	1H	7.46, d, 2.8
17	3H	1.3, d, 7.1	3H	1.30, d, 6.9
1'	1H	5.15, dd, 17.5, 1.2	1H	5.04, d, 17.1
	1H	5.11, dd, 10.6, 1.2	1H	5.01, d, 10.4
2'	1H	6.21, dd, 17.5, 10.6	1H	6.15, dd, 17.2, 10.4
4'	3H	1.565, s	3H	1.47, s
5'	3H	1.560, s	3H	1.46, s
1"			2H	3.33#
2"			1H	5.32, t, 7.3
4"			3H	1.70, s
5"			3H	1.68, s

#Signals overlapping with those of water; Spectral data are consistent with those previously reported in the literature, for **1M1** (Yin, S.; Yu, X.; Wang, Q.; Liu, X.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 1649) and **1D1** (Dai, Q. et al. *Org.Lett.* **2011**, *13*, 2302) (Fujimoto, H.; Fujimaki, T.; Okuyama, E.; Yamazaki, M. *Chem.Pharm.Bull.* **1999**, *47*, 1426).

Table S5 ¹H NMR data of the enzyme products of EchPT2 tardioxopiperazine B (**1D2**) and variecolorin L (**1T1**) in DMSO-d_6
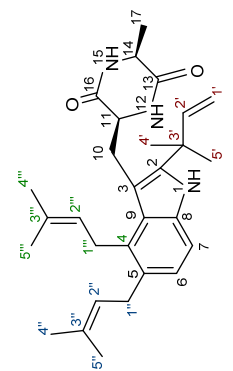

cyclo-L-2-*tert*-DMA-7-DMA-Trp-L-Ala
tardioxopiperazine B (1D2)

cyclo-L-2-*tert*-DMA-4,5-di-DMA-Trp-L-Ala
variecolorin L (1T1)

Pos.	1D2 δ_H, multi, J in Hz		1T1 Pos.	1T1 δ_H, multi, J in Hz	
NH-1	1H	9.83, s	NH-1	1H	10.52, s
4	1H	7.26, d, 7.9	4		-
5	1H	6.88, dd, 7.9, 7.0	5		-
6	1H	6.80, d, 7.0	6	1H	6.82, d, 8.2
7		-	7	1H	7.15, d, 8.2
10a	1H	3.33#	10a	1H	3.45, dd, 14.6, 3.7
10b	1H	3.07, dd, 14.4, 9.3	10b	1H	3.14, dd, 14.6, 10.9
11	1H	3.93, m	11	1H	3.99, dd, 10.9, 3.7
NH-12	1H	7.53, d, 2.7	NH-12	1H	8.18, d, 1.7
14	1H	3.79, dq, 7.1, 2.6	14	1H	3.86, qd, 7.0, 1.5
NH-15	1H	8.18, d, 2.6	NH-15	1H	6.86, br s
17	3H	1.27, d, 7.1	17	3H	1.30, d, 7.0
1'	1H	5.09, dd, 17.5, 1.1	1'	1H	5.06, d, 17.5
	1H	5.05, dd, 10.6, 1.1		1H	5.02, d, 10.6
2'	1H	6.24, dd, 17.5, 10.6	2'	1H	6.16, dd, 17.5, 10.5
4'	3H	1.52, s	4'	3H	1.49, s
5'	3H	1.51, s	5'	3H	1.47, s
1''	2H	3.60, d, 7.3	1''	2H	3.33#
2''	1H	5.40, br. t, 7.3	2''	1H	5.16, t, 6.5
4''	3H	1.735, s	4''	3H	1.67, s
5''	3H	1.730, s	5''	3H	1.66, s
			1'''	2H	3.33#
			2'''	1H	4.89, t, 5.5
			4'''	3H	1.71, s
			5'''	3H	1.70, s

#Signals overlapping with those of water; Spectral data are consistent with those previously reported in the literature, for, **1D2** (Fujimoto, H.; Fujimaki, T.; Okuyama, E.; Yamazaki, M. *Chem.Pharm.Bull.* **1999**, *47*, 1426). and **1T1** (Fujimoto, H.; Fujimaki, T.; Okuyama, E.; Yamazaki, M. *Chem.Pharm.Bull.* **1999**, *47*, 1426) (Li, D.-L.; Li, X.-M.; Li, T.-G.; Dang, H.-Y.; Wang, B.-G. *Helv.Chim.Acta* **2008**, *91*, 1888.)

100

Table S6 ¹H NMR data of the isolated echinulin (**1T2**) from the fungal culture in DMSO-d_6

*cyclo*-L-2-*tert*-DMA-5,7-di-DMA-Trp-L-Ala   echinulin (**1T2**)
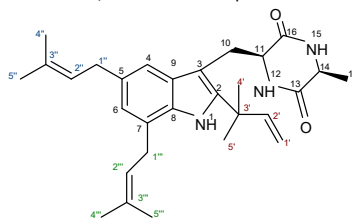
Pos.	δH, multi J in Hz		COSY Corr	δC derived from HMBC data .	,	HMBC Corr.
NH-1	1H	9.71, s	-	-		2, 3, 4, 8, 9, 3'
2				141.77		
3				105.26		
4	1H	7.03, s	H-6	114.88		3, 6, 8, 9, 1''
5		-	-	131.78		-
6	1H	6.62, s	H 4	121.25		4, 8, 1'', 1'''
7		-	-	123.83		-
8				132.45		
9				129.50		
10a	1H	3,02, dd, 14.4, 9.6	H-11, H-10b	31.34		2, 3, 9, 11, 16
10b	1H	3,33#	H-11, H-10a			2, 3, 9, 11, 16
11	1H	3.91, m	NH-12, H-10a, H-10b	55.74		3, 10, 13
NH-12	1H	7.45, d, 2.95	H-11	-		11, 14, 16
13				168.27		
14	1H	3.80, dq, 7.1, 2.5	H-17, NH-15	50.42		17, 13,
NH-15	1H	8.18, d, 2.5	H-14	-		11, 13, 14
16				167.72		
17	3H	1.32, d, 7.1	H-14	20.96		13, 14
1'	1H	5.07, dd, 17.5, 1.1	H-2', H-1b'	112.13		2, 2', 3', 4', 5'
	1H	5.03, dd, 10.6, 1.1	H-2', H-1a'			2, 2', 3' 4' 5'
2'	1H	6.21, dd, 17.5, 10.6	H-1'	147.09		2, 3' 4' 5'
3'				39.10		
4'	3H	1.51, s	-	28.01		2, 2', 3', 5
5'	3H	1.49, s	-	28.01		2, 2', 3', 4
1''	2H	3,27#	H-2'', H-4''/ H-5''	34.23		4, 5, 6, 2'', 3'', 4'', 5''
2''	1H	5.29, t, 7.3	H-1'', H-4''/ H-5''	125.14		5, 4'', 5'', 1''
3''				130.64		
4''	3H	1,683, m	H-1'', H-2''	17.69		1''', 2''', 3''', 5'''
5''	3H	1,677, m	H-1'', H-2''	25.55		1''', 2''', 3''' 4'''
1'''	2H	3.56, d, 7.2	H-2''', H-4'''/ H-5'''	29.02		6, 2'', 3'', 4'', 5''
2'''	1H	5.37, t, 7.2	H-1''', H-4'''/ H-5'''	123.00		7, 1''', 4'', 5''
3'''				132.02		
4'''	3H	1,732, s	H-1'''	17.84		1'', 2'', 3'', 5''
5'''	3H	1,727, s	H-1'''	25.66		1'', 2'', 3'', 4''

#Signals overlapping with those of water. Spectral data of **1T2** correspond well to those reported previously (Wang, W. L. *et al. J. Nat. Prod.* **2007**, *70*, 1558) (Li, D.-L.; Li, X.-M.; Li, T.-G.; Dang, H.-Y.; Wang, B.-G. *Helv.Chim.Acta* **2008**, *91*, 1888)
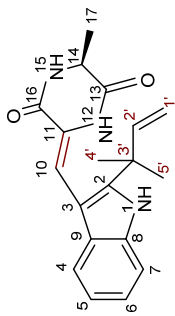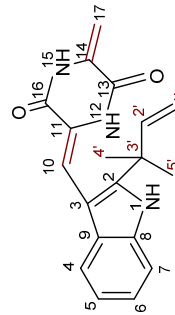
cyclo-L-2-*tert*-DMA-4,5,7-tri-DMA-Trp-L-Ala (**1Q2**)
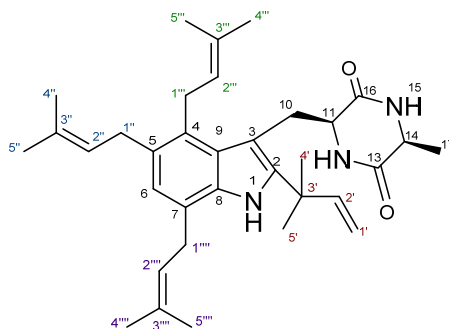
Pos.	δ_H, multi	*J* in Hz
NH-1	1H	9.56, s
4	-	-
5	-	-
6	1H	6.62, s
7	-	-
10a	1H	3.43#
10b	1H	3.18#
11	1H	3.95, br. d, 10.3
NH-12	1H	6.83, d, 1.7
14	1H	3.86, q, 6.7
NH-15	1H	8.19, d, 2.5
17	3H	1.31, d, 7.0
1'	1H	5.09, dd, 17.5, 1.0
	1H	5.04, dd, 10.6, 1.0
2'	1H	6.21, dd, 17.5, 10.6
3'		
4'	3H	1.51, s
5'	3H	1.54, s
1"	2H	3.33#
2"	1H	5.13, t, 6.7
4"	3H	1,67, s
5"	3H	1,66, s
1'''	2H	3.33#
2'''	1H	4.86, t, 5.2
4'''	3H	1,70, s
5'''	3H	1,61, s
1''''	1H	3.54, d, 7.4
2''''	1H	5.37, t, 7.3
4''''	3H	1.74, s
5''''	3H	1.73, s

#Signals overlapping with those of water

Table S8 ¹H NMR data of neoechinulin A (**2M1**) and neoechinulin B (**3M1**) isolated from the fungal culture in DMSO-d_6

	cyclo-L-2-*tert*-DMA-Trp-L-Ala $\Delta^{10,11}$ neoechinulin **A** (**2M1**)		*cyclo*-L-2-*tert*-DMA-Trp-L-Ala $\Delta^{10,11}$ $\Delta^{14,15}$ neoechinulin **B** (**3M1**)	
Pos.		δ_H, multi $\;J$ in Hz		δ_H, multi $\;J$ in Hz
NH-1	1H	11.02, s	1H	11.11, s
4	1H	7.18, dd, 8.1, 1.1	1H	7.21, d, 7.8
5	1H	7.00, ddd, 8.1, 7.3, 1.1	1H	7.02, dd, 7.8, 7.2
6	1H	7.07, ddd, 8.5, 7.3, 1.1	1H	7.10, dd, 7.8, 7.2
7	1H	7.41, dd, 8.5, 1.1	1H	7.43, d, 7.8
10	1H	6.88, s	1H	6.96, s
NH-12	1H	8.29, s	1H	8.86, s
14	1H	4.15, q, 7.2		
NH-15	1H	8.59, s	1H	10.82, s
17	3H	1.37, d, 7.2	1H	5.23, s
			1H	4.92, s
1'	1H	5.04, dd, 10.4, 1.3	1H	5.06, dd, 10.5, 1.2
	1H	5.01, dd, 17.2, 1.3	1H	5.03, dd, 17.4, 1.2
2'	1H	6.07, dd, 17.2, 10.4	1H	6.08, dd, 17.4, 10.5
4'/5'	6H	1.47, br. s	6H	1.48, br. s

aSignals overlapping with those of solvent, spectral data are consistent with those previously reported in the literature, for **2M1** (Aoki, T. *et al. Synlett* **2006**, No. 5, 677) (Kuramochi, K.; *et al. Synthesis* **2008**, 3810) and **3M1** (Marchelli, R.; Dossena, A.; Pochini, A.; Dradi, E. *J.Chem.Soc., Perkin Trans.1* **1977**, 713)(Gao, H.; Zhu, T.; Li, D.; Gu, Q.; Liu, W. *Arch.Pharm.Res.* **2013**, 36, 952)
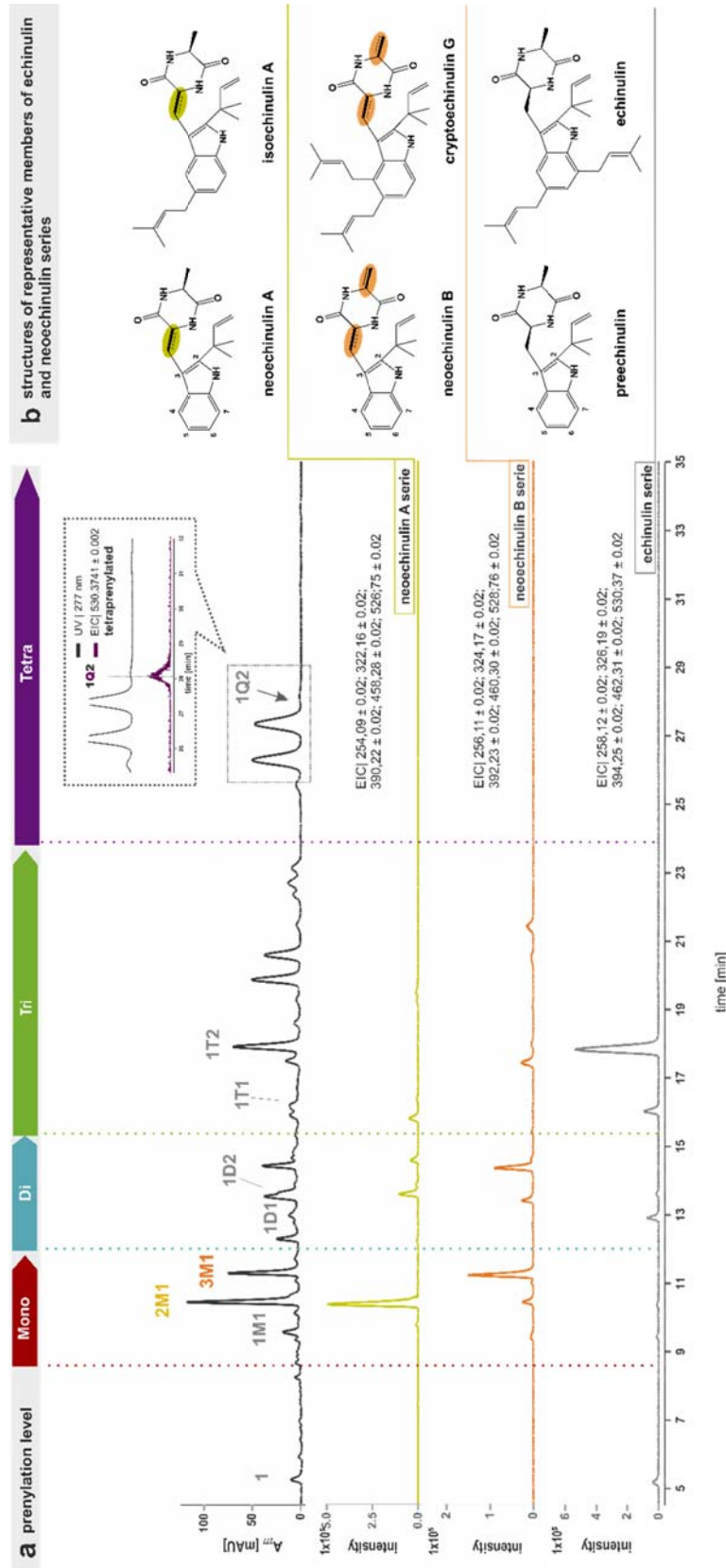
6 FIGURES OF EXPERIMENTAL SECTION

Figure S1 Mapping of the echinulin prenylation profile (A) The echinulin prenylation profile in the culture extract of *A. ruber* via LC-MS analysis confirms also the natural occurrence of tetraprenylated echinulins in *A. ruber* extract. (B) Structures of representative members of echinulin and neoechinulin series
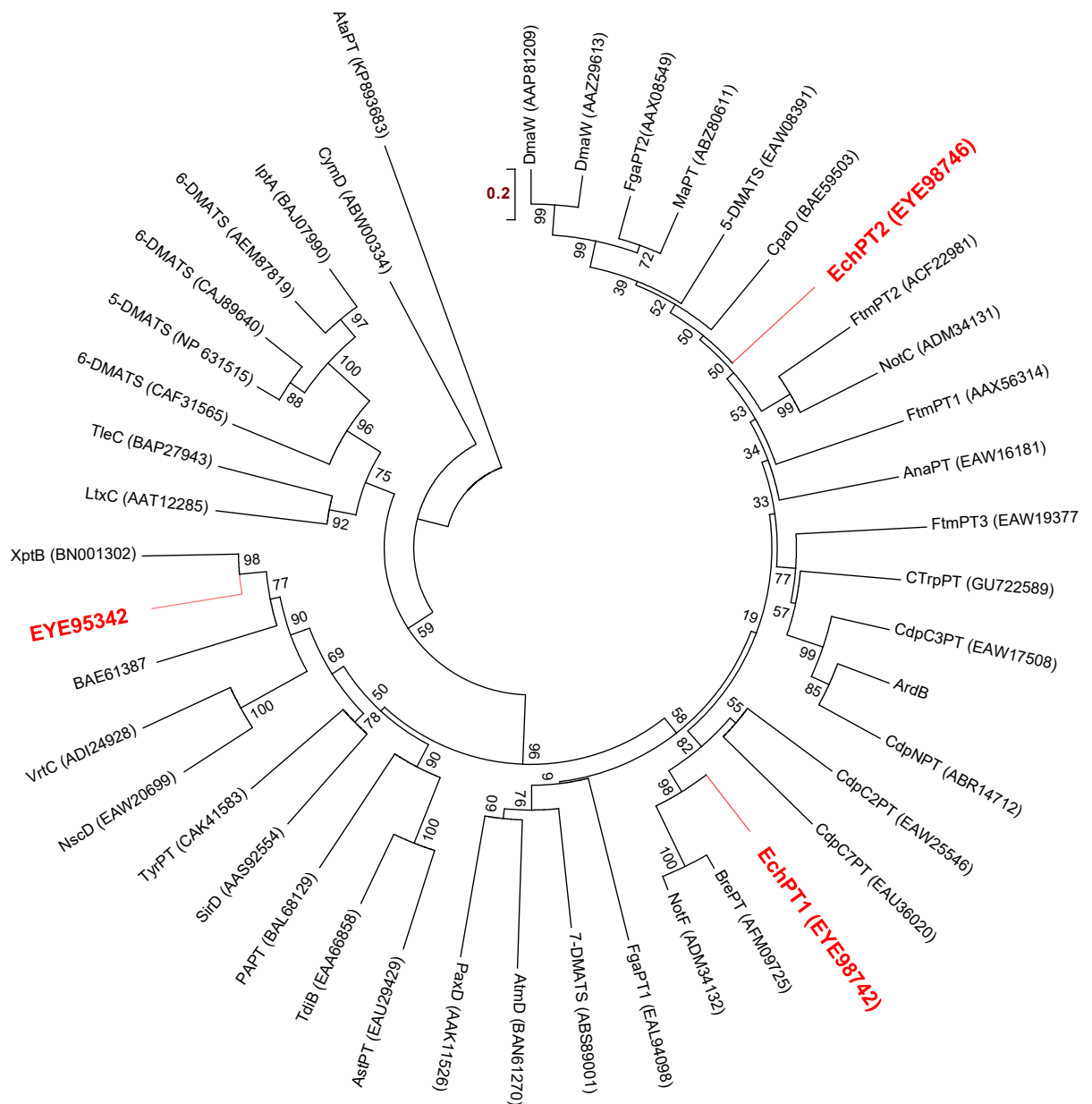
Figure S2 Phylogenetic relationship of known prenyltransferases of the DMATS superfamily.

EchPT1, EchPT2 and EYE95342 are highlitghted in red. The protein accession numbers are given in parenthesis. The sequence alignments were created using MUSCLE (Edgar, R. C. *Nucleic Acids Res.* **2004**, *32*, 1792) and the phylogenetic tree was generated by MEGA6.0 (Tamura, K.; Stecher, G.; Peterson, D.; Filipski. A., Kumar, S. *Mol. Biol. Evol.* **2013**, *30*, 2725) across the Minimum Evolution methode (Rzhetsky, A.; Nei, M. *Mol. Biol. Evol.* **1992**, *9*, 945) for interfering the evolutionary relationship, and the Bootstrap consensus tree interfered from 1000 replicates.

A

B
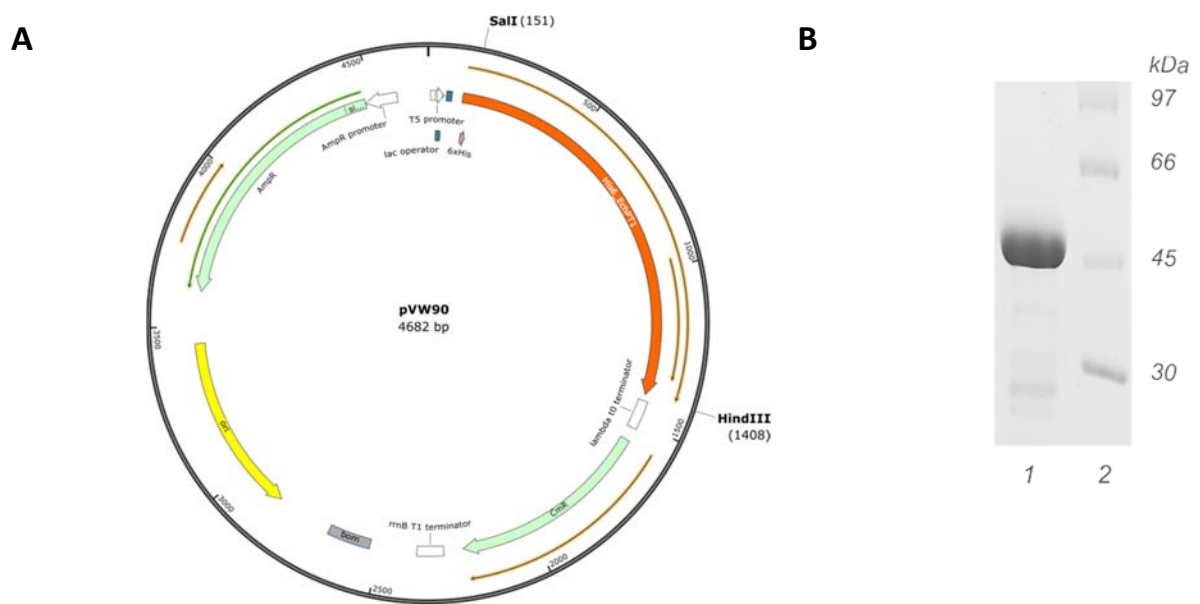
Figure S3 Plasmid map of the expression construct and SDS gel of the purified EchPT1

(A) Plasmid map of the EchPT1 expression construct and (B) SDS-PAGE of the purified His$_6$-EchPT1 (47.9kDa) (lane 1) and molecular weight marker (lane 2).
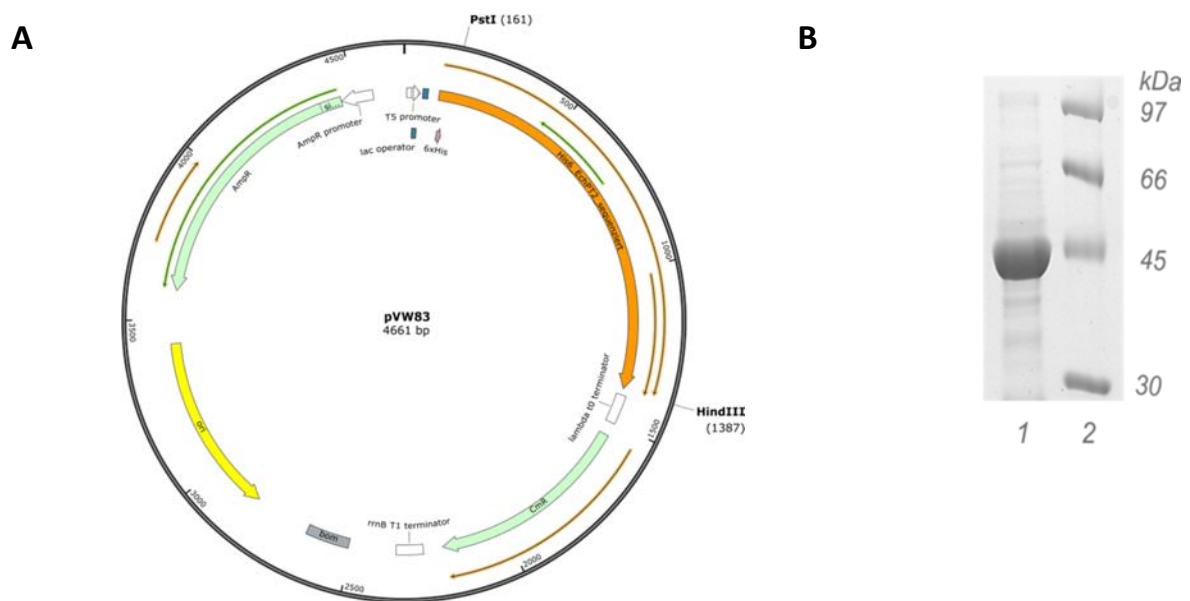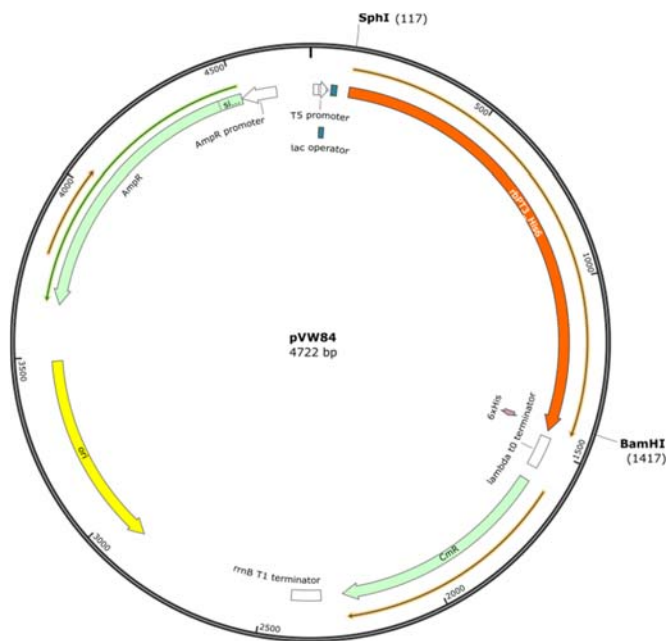
A

B

Figure S4 Plasmid map of the expression construct and SDS gel of the puriefied EchPT2

(A) Plasmid map of the EchPT2 expression construct and (B) SDS-PAGE of the purified His$_6$-EchPT2 (42.9kDa) (lane 1) and molecular weight marker (lane 2).
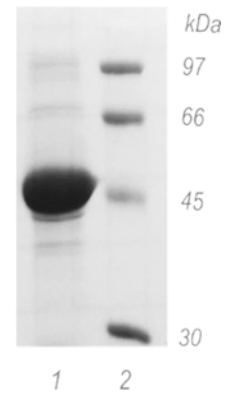
S27

Figure S5 Plasmid map of the expression construct and the SDS gel of the purified EYE95342

(A) Plasmid map of the EYE95342 expression construct and (B) SDS-PAGE of the purified protein (50.9kDa) (lane 1) and molecular weight marker (lane 2).
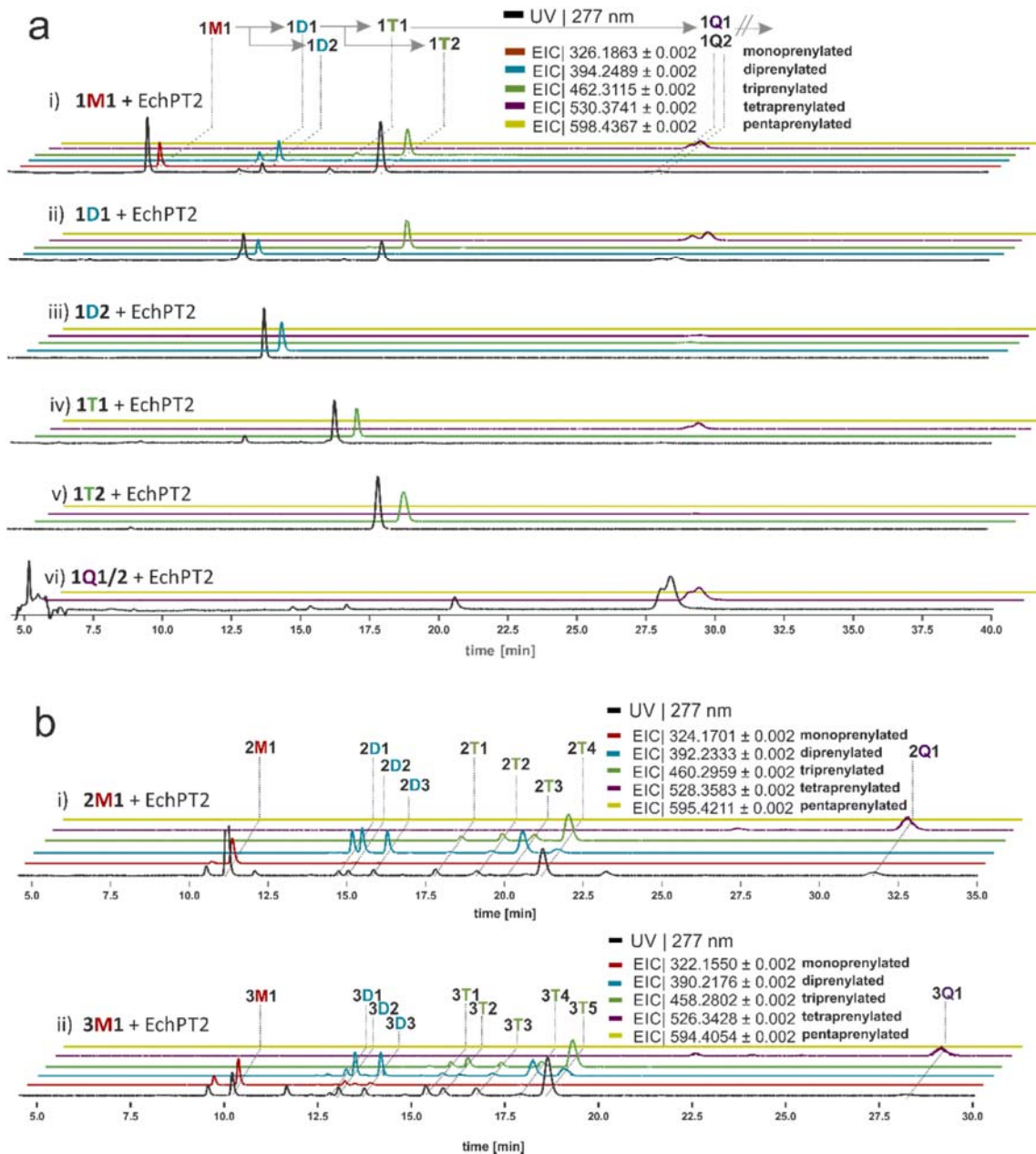
S28

Figure S6 LC-MS analysis of EchPT2 assays with **1M1**, **1D1, 1D2, 1T1, 1T2,** and a mixture of **1Q1** and **1Q2** (a) and neoechinulins A (**2M1**) and B (**3M1**) (b).
Absorptions at 277 nm and extracted ion chromatograms (EICs) of the substrates and prenylated products are illustrated.
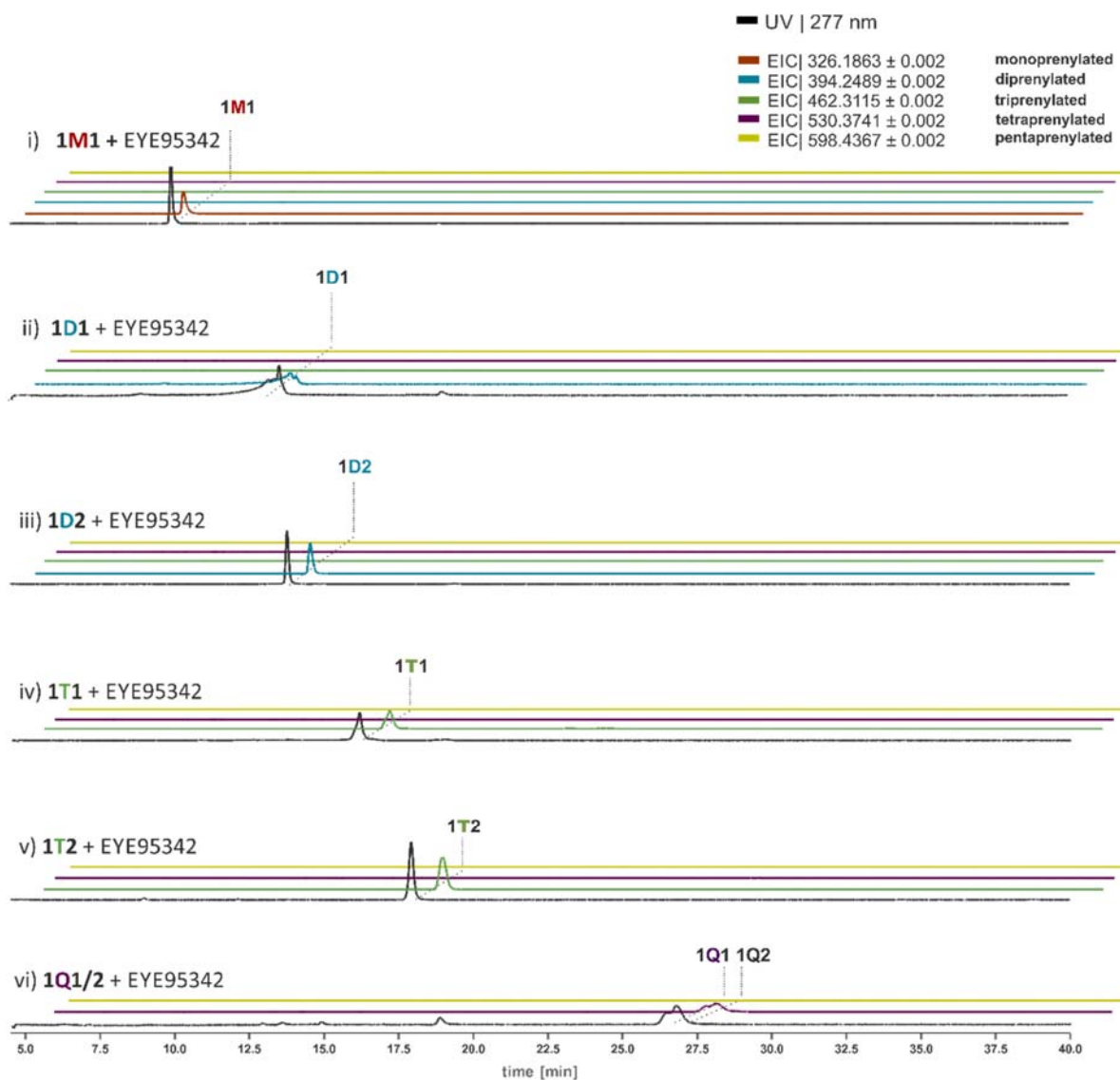
Figure S7 LC-MS analysis of EYE95342 assays with **1M1**, **1D1, 1D2, 1T1,** and **1T2** as well as a mixture of **1Q1** and **1Q2**.

Absorptions at 277 nm and extracted ion chromatograms (EICs) of the respective substrates and proposed prenylated products are illustrated. Due to a new column, the retention times of the substrates differ slightly from those in Figure 4 as well as in Figures S1 and S6.
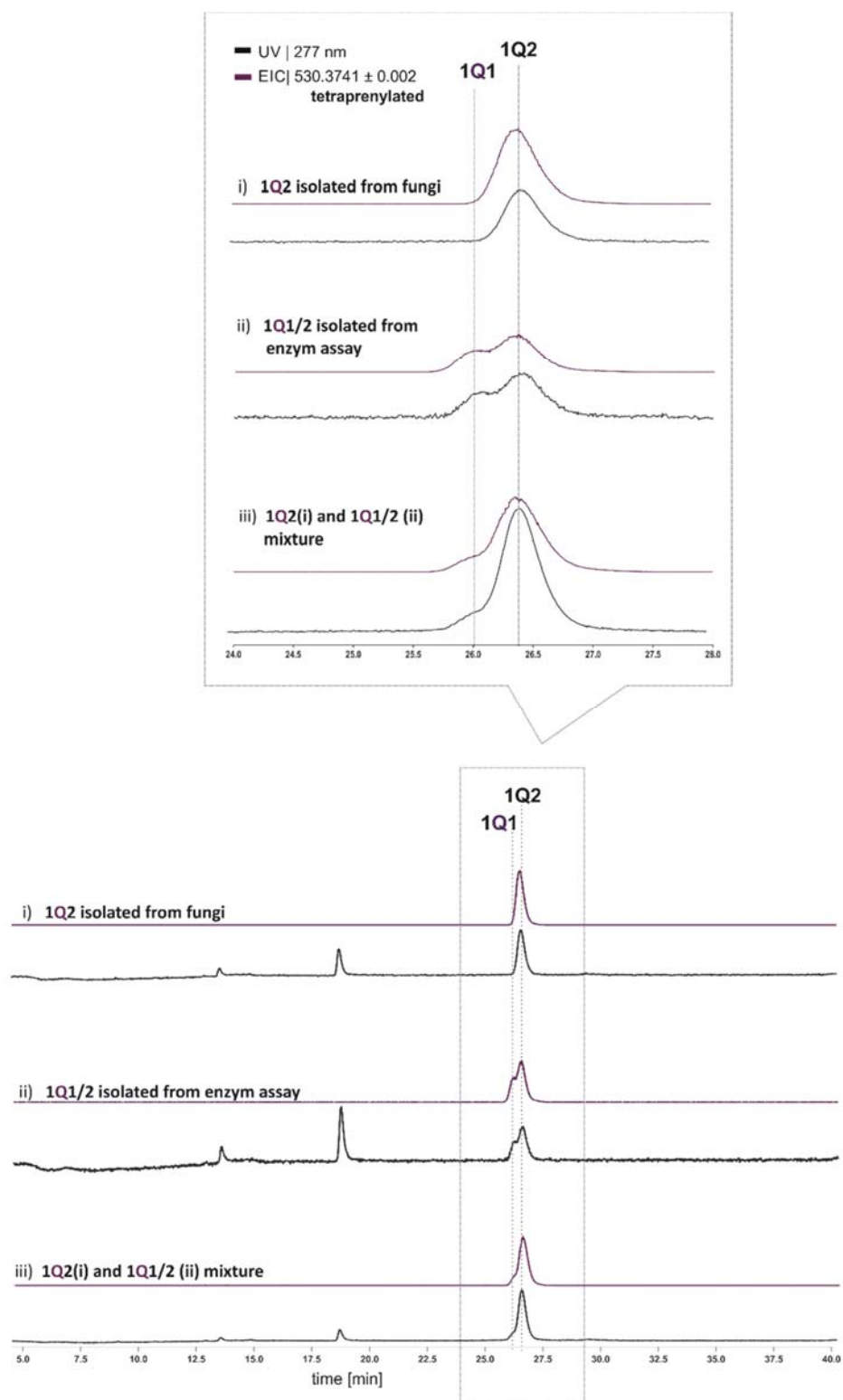
Figure S8 Mapping of the tetraprenylated EchPT2 products **1Q1** and **1Q2** by LC-MS.

Due to a new column, the retention times of the substrates differ slightly from those in Figure 4 as well as in Figures S1 and S6.
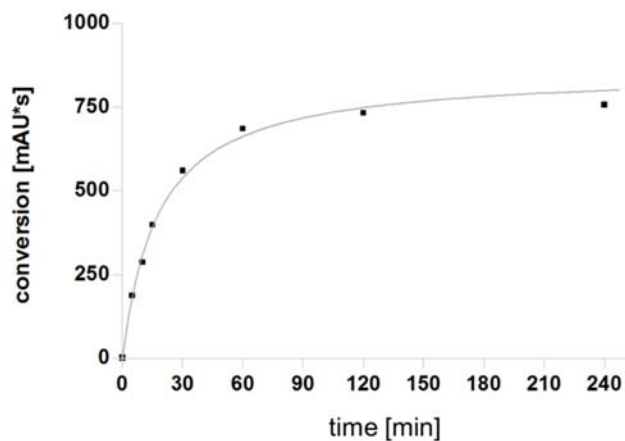
7 FIGURES OF KINETIC PARAMETERS

Figure S9 Dependence of the product formation of the EchPT1 reaction with **1** and DMAPP on time.

The reaction mixture contained 1 mM of **1,** 5 mM of CaCl$_2$, 2 mM of DMAPP and 0.5 µg of the purified recombinant protein and was incubated at 37 °C and pH 7.5. Data points represent the mean of two measurements.
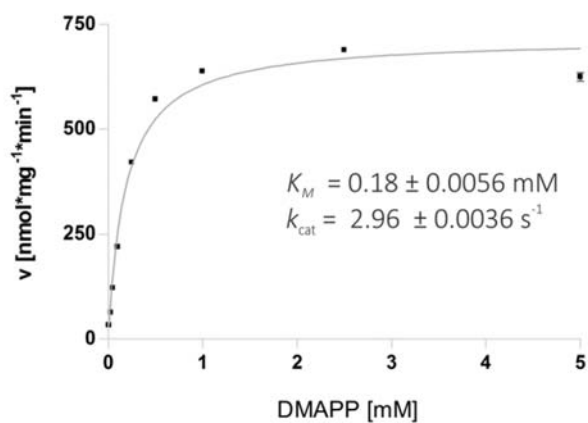

$$K_M = 0.18 \pm 0.0056 \text{ mM}$$
$$k_{cat} = 2.96 \pm 0.0036 \text{ s}^{-1}$$

Figure S10 Dependence of the product formation of the EchPT1 reaction with **1** on DMAPP concentration.

The reaction mixture contained 1 mM of **1,** 5 mM of CaCl$_2$, 2 µg of the purified recombinant protein and was incubated at 37 °C and pH 7.5 for 1 min. Data points represent the mean of two measurements.
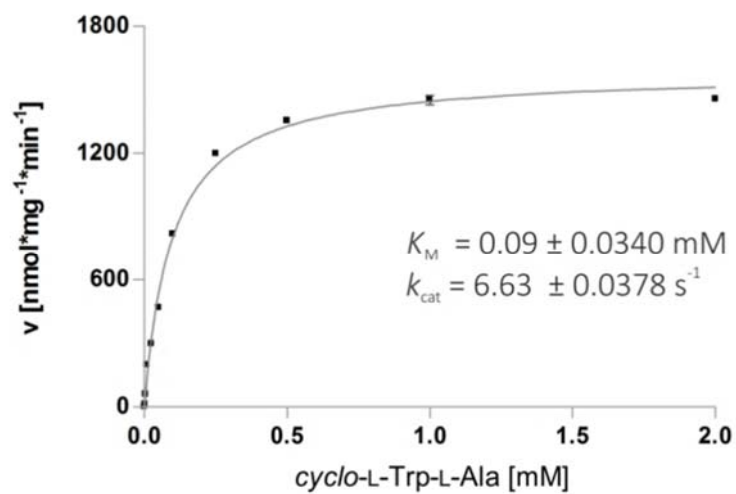
S32

Figure S11 Determination of the kinetic parameters of EchPT1 toward *cyclo*-L-Trp-L-Ala (**1**)

The reaction mixtures contained 5 mM of $CaCl_2$, 1 mM of DMAPP and 2 µg of the purified recombinant protein. The mixtures were incubated at 37 °C and pH 7.5 for 1 min.
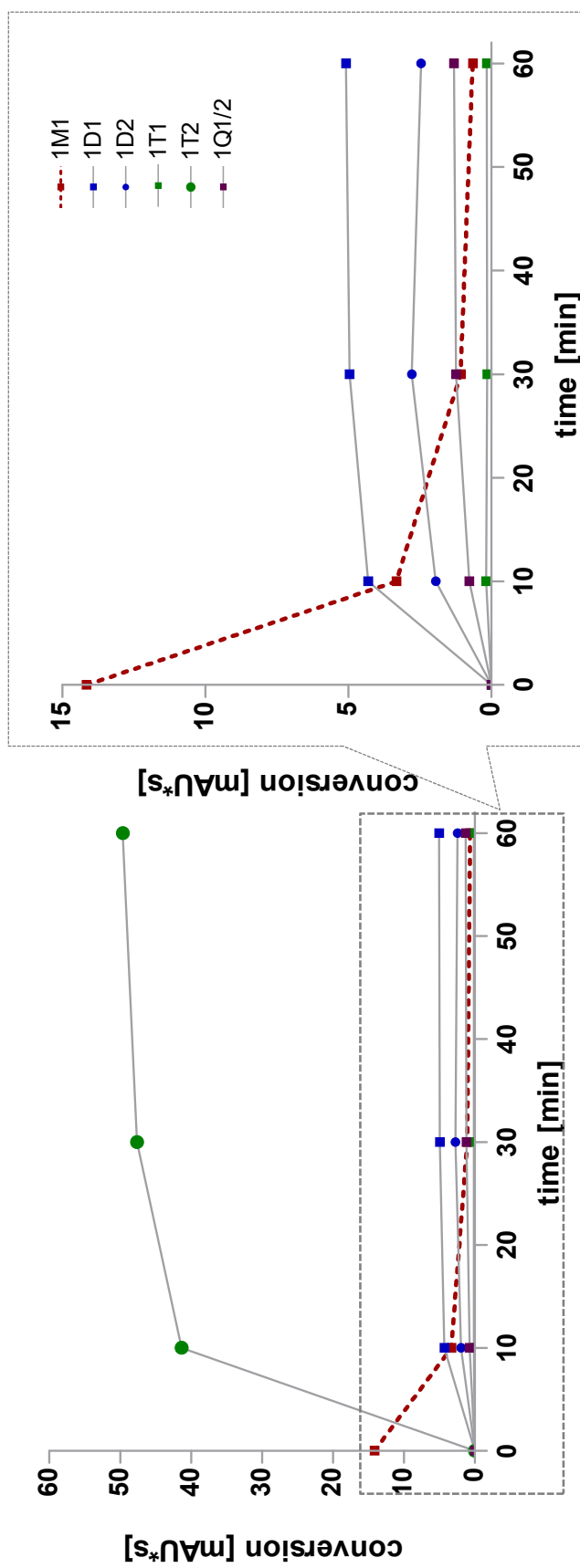
S33

Figure S12 Time dependence of the product formation of EchPT2 reaction with *cyclo*-L-2-*tert*-DMA-Trp-L-Ala (**1M1**).

The reaction mixture contained 1 mM of **1M1**, 5 mM of $CaCl_2$, 3 mM of DMAPP and 0.5 µg of the purified recombinant protein and was incubated at 37 °C and pH 7.5.
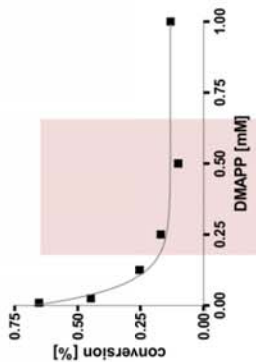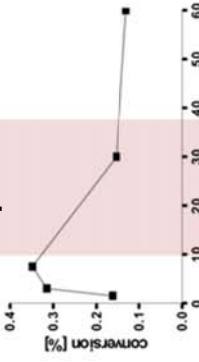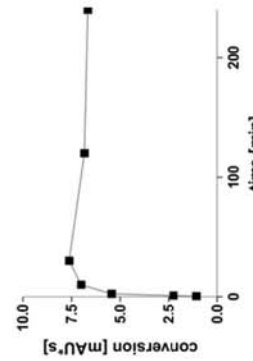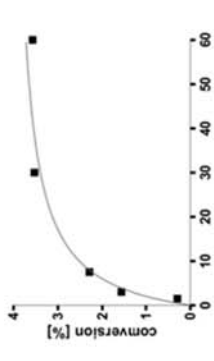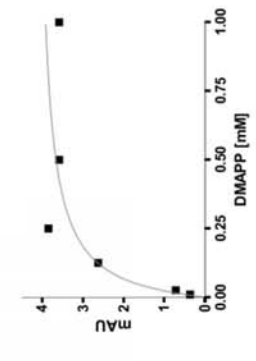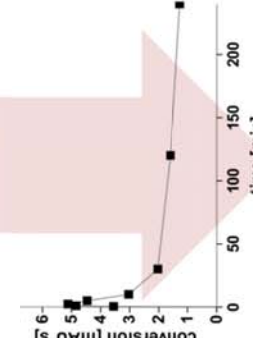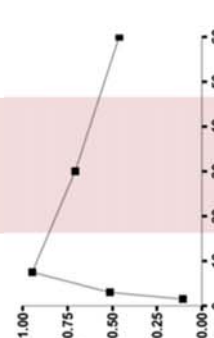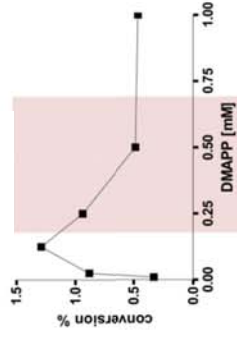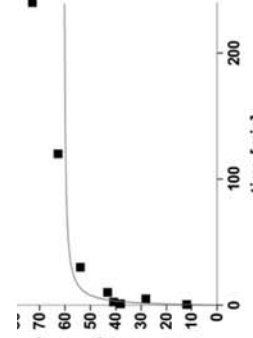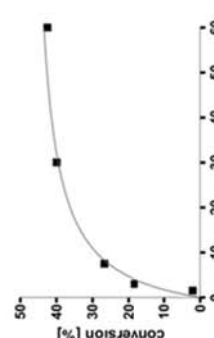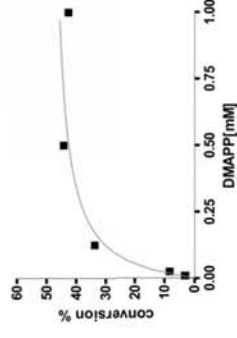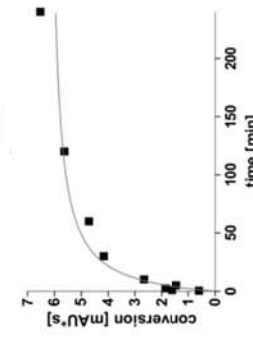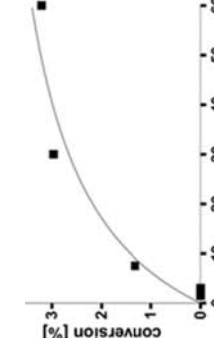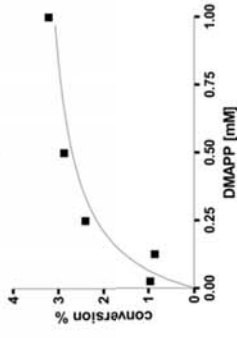
Figure S13 Dependence of the product formation of the EchPT2 reaction with *cyclo*-L-2-*tert*-DMA-Trp-L-Ala (**1M1**) on DMAPP, protein concentration, and incubation time.
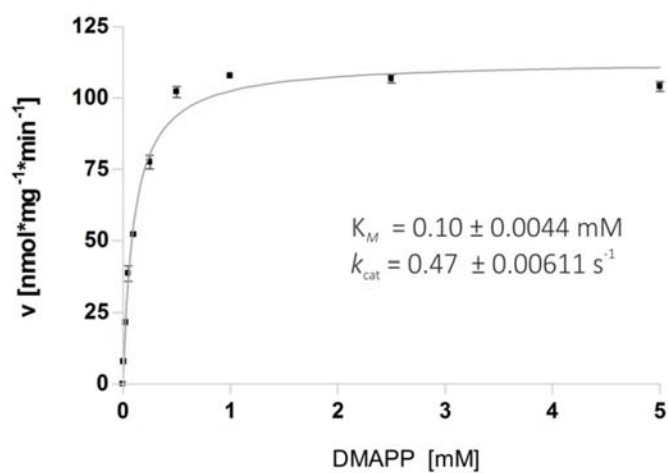
Figure S14 Dependence of the product formation of EchPT2 reaction in presence of *cyclo*-L-2-*tert*-DMA-Trp-L-Ala (**1M1**) on DMAPP concentration.

The reaction mixture contained 1 mM of **1M1**, 5 mM of $CaCl_2$, 2 µg of the purified recombinant protein and was incubated at 37 °C and pH 7.5 for 3 min.
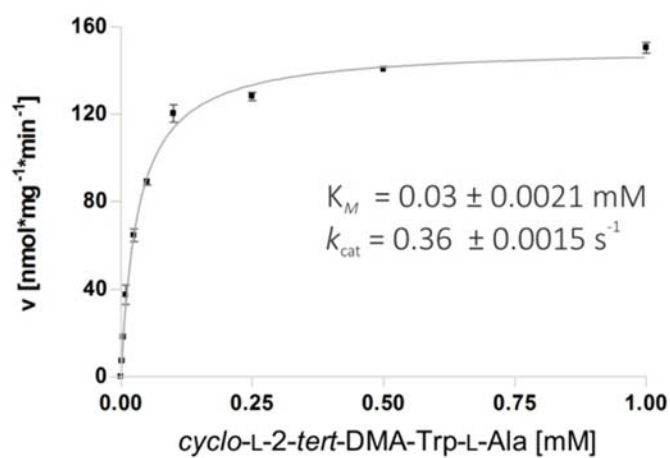
Figure S15 Determination of the kinetic parameters of the EchPT2 reaction toward *cyclo*-L-2-*tert*-DMA-Trp-L-Ala (**1M1**).

The reaction mixtures contained 5 mM of $CaCl_2$, 3 mM of DMAPP and 2 µg of the purified recombinant protein. The mixtures were incubated at 37 °C and pH 7.5 for 3 min.

S36

8 ION DEPENDENCE OF THE ECHPT1 AND ECHPT2 REACTIONS
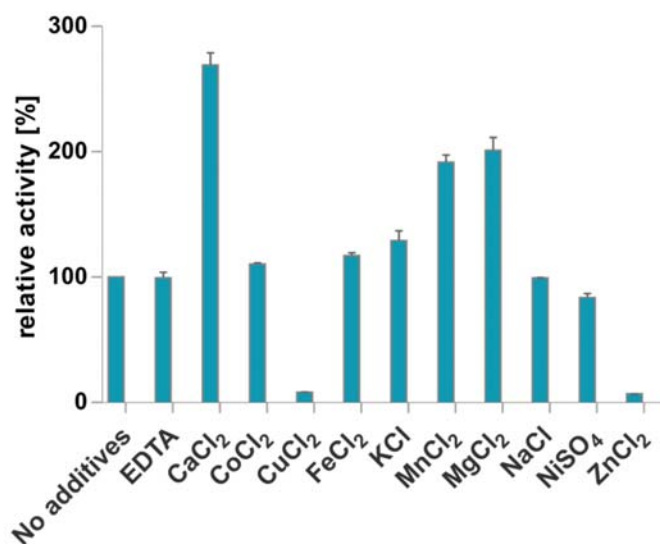
Figure S16 Dependence of the EchPT1 reaction on the presence of additives

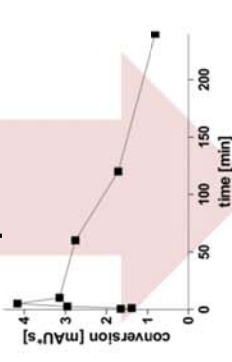The reaction mixtures contained 1 mM of **1**, 5 mM of additives, 1 mM of DMAPP and 1 µg of the purified recombinant protein and were incubated at 37 °C and pH 7.5 for 20 min. The enyme activity without additives was defined as 100 %. Error bars indicate ranges of the two measurements.
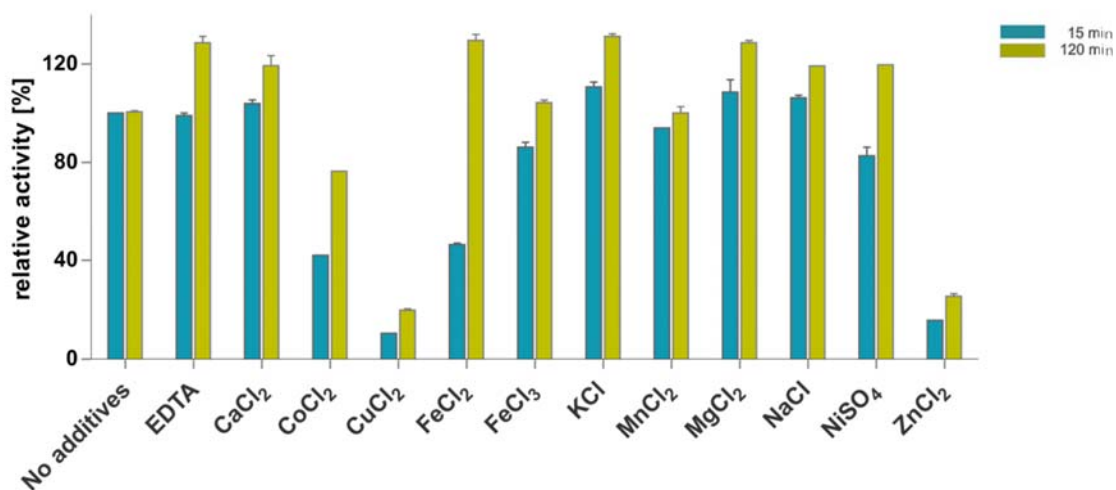
Figure S17 Dependence of the EchPT2 reaction on the presence of ions.

The reaction mixtures contained 1 mM of **1M1**, 5 mM of additives, 3 mM of DMAPP and 5 µg of the purified recombinant protein and were incubated at 37 °C and pH 7.5 for 15 or 120 min. The enzyme activities without additives were defined as 100 %. Error bars indicate ranges of the two measurements.

9 FIGURES OF NMR Spectra

Figure S18 ^{1}H NMR spectrum of *cyclo*-L-2-*tert*-DMA-Trp-L-Ala (**1M1**) in CD$_3$OD (500MHz)

117

Figure S19 ^{1}H NMR spectrum of tardioxopiperazine A (**1D1**) in DMSO-d_6 (500MHz)
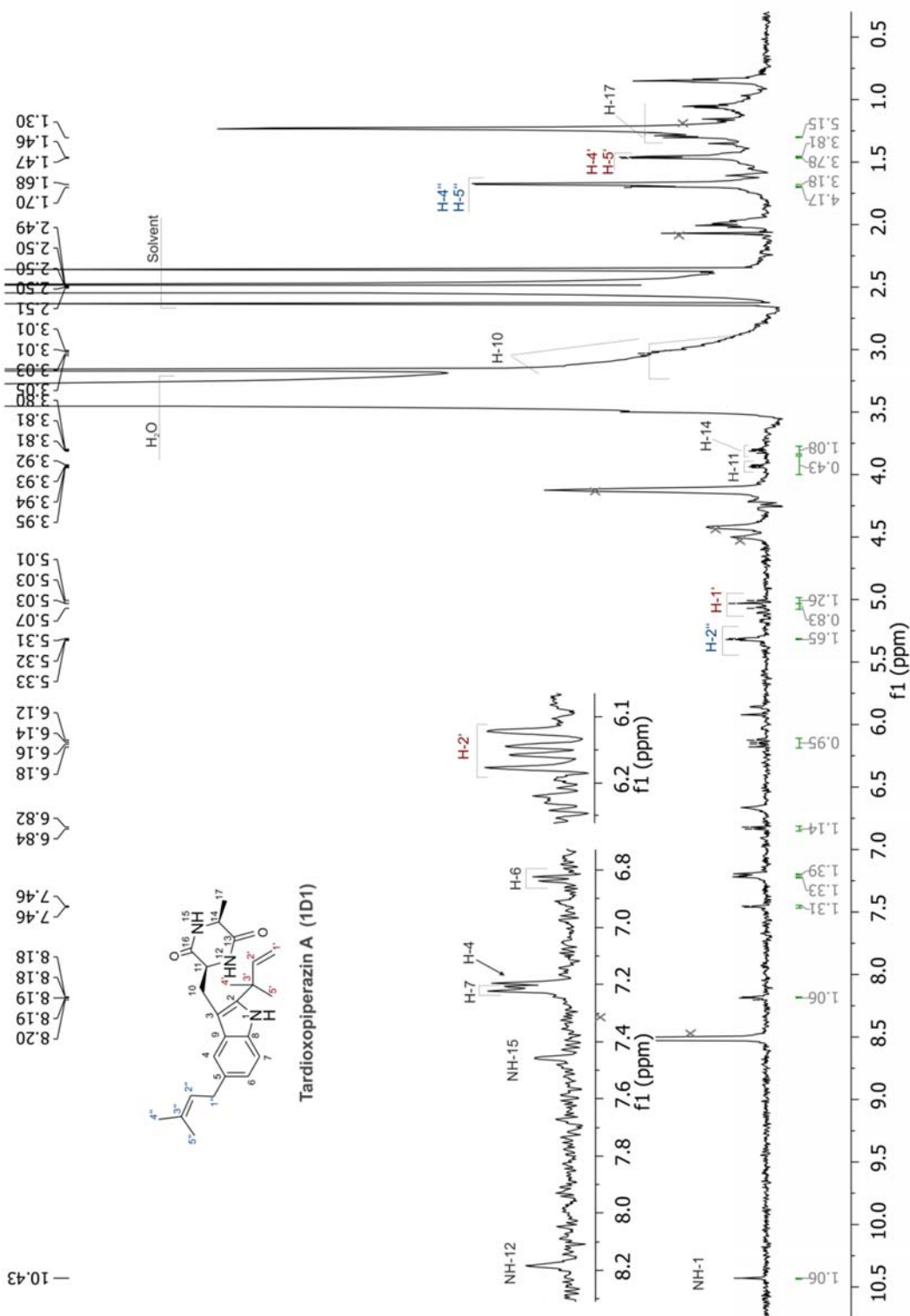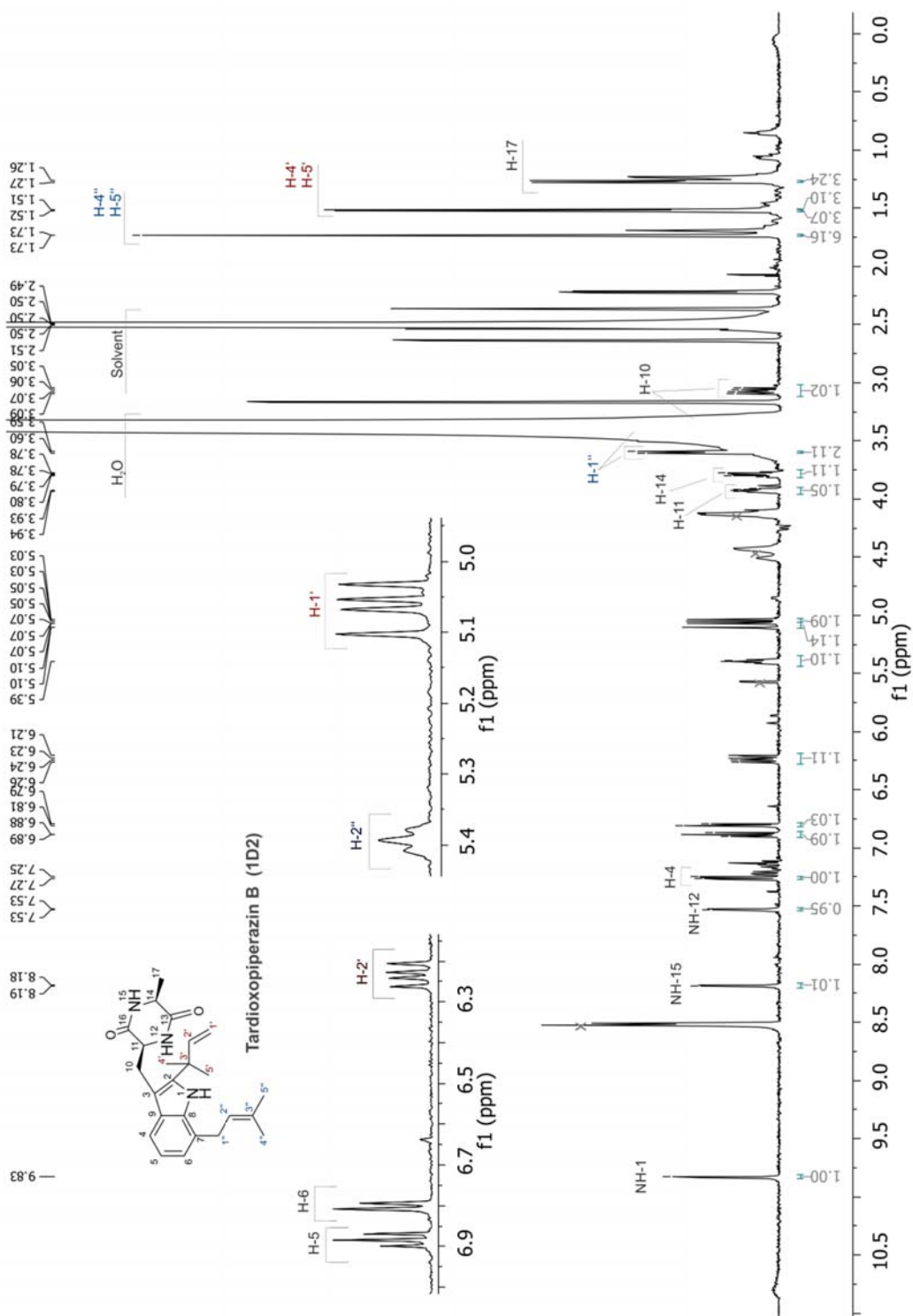
Figure S20 ¹H NMR spectrum of tardioxopiperazine B (**1D2**) in DMSO-d_6 (500MHz)
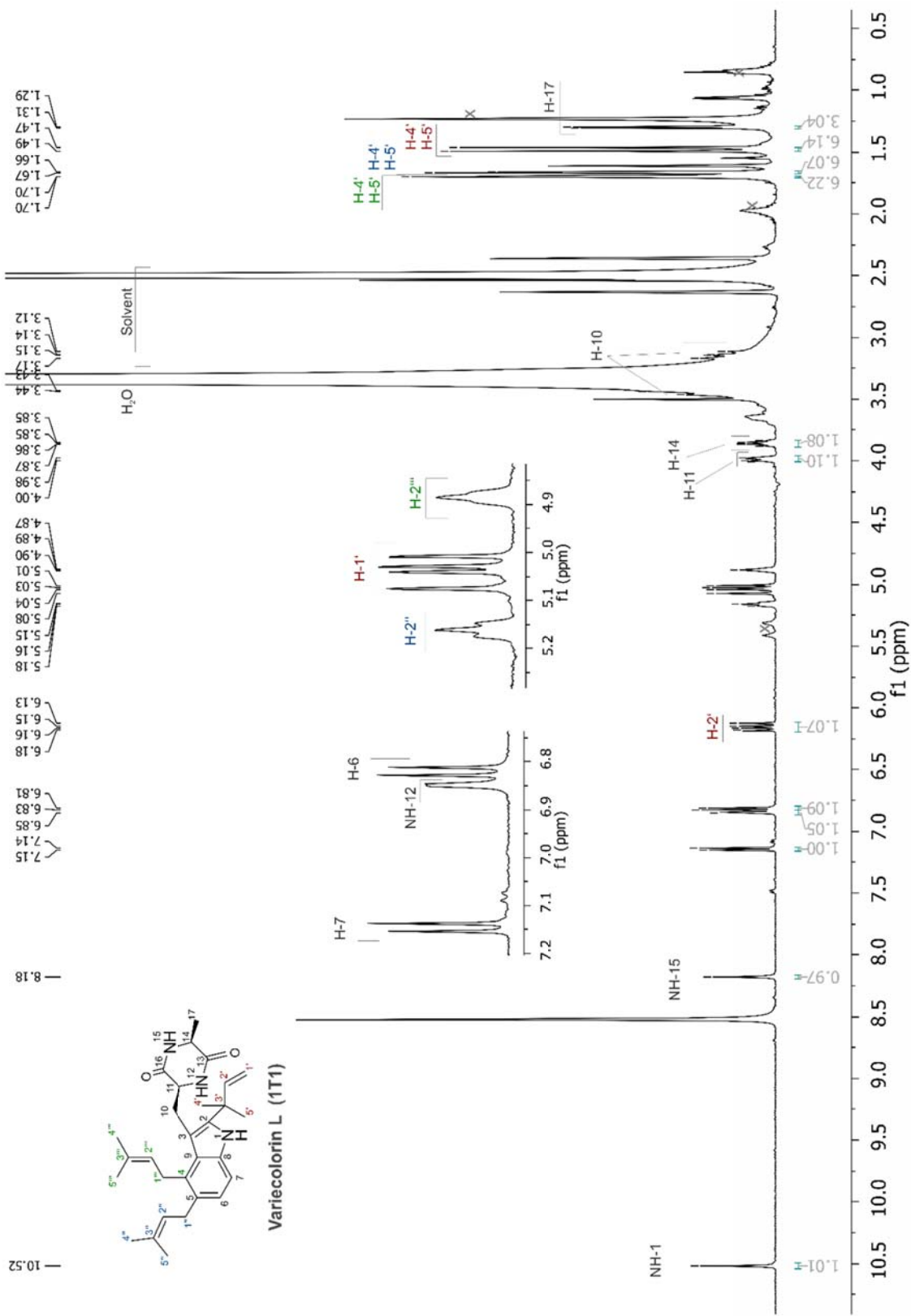
Figure S21 ^{1}H NMR spectrum of variecolorin L (**1T1**) in DMSO-d_6 (500MHz)

S41

120
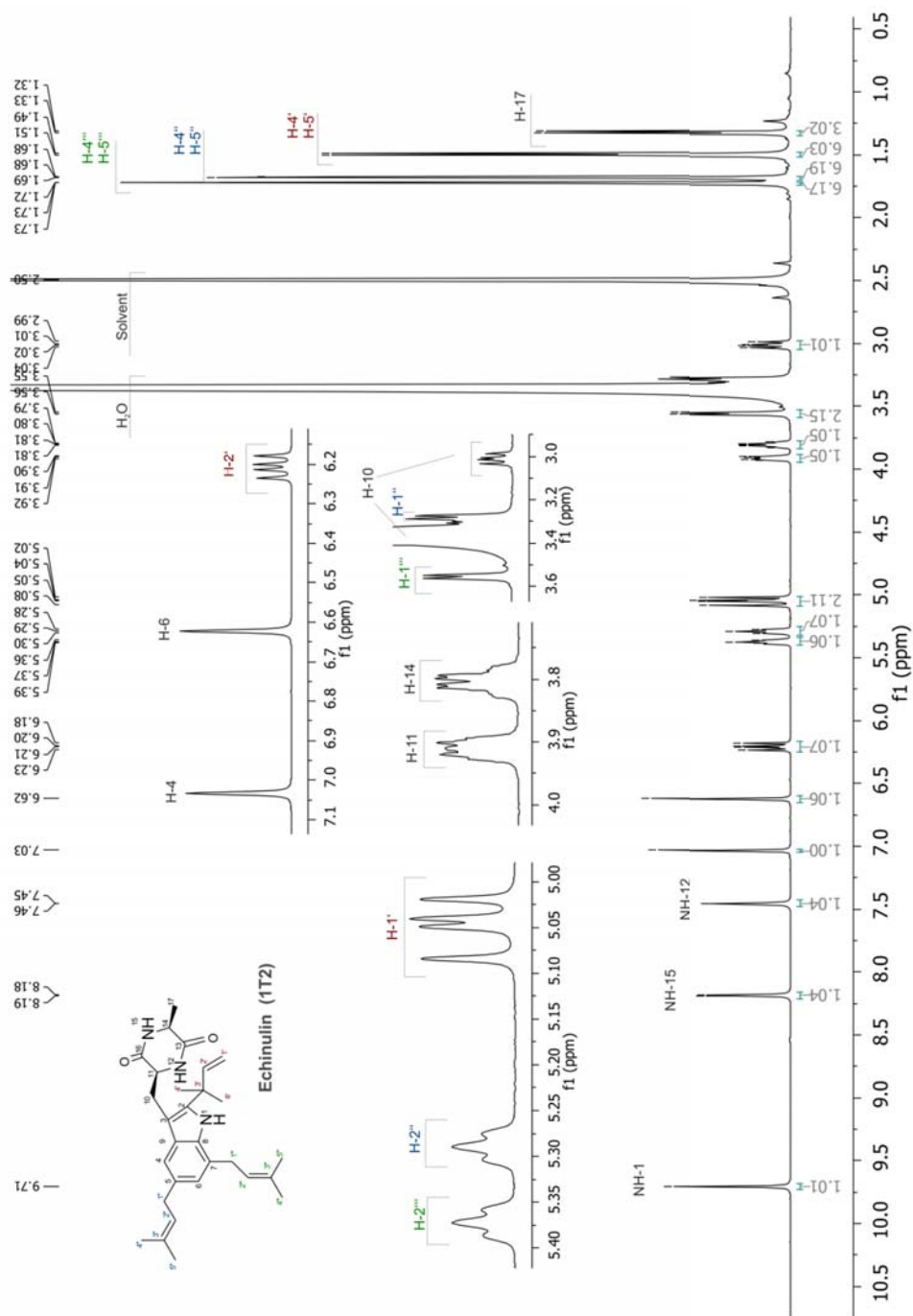
Figure S22 ^{1}H NMR spectrum of echinulin (**1T2**) in DMSO-d_6 (500MHz)
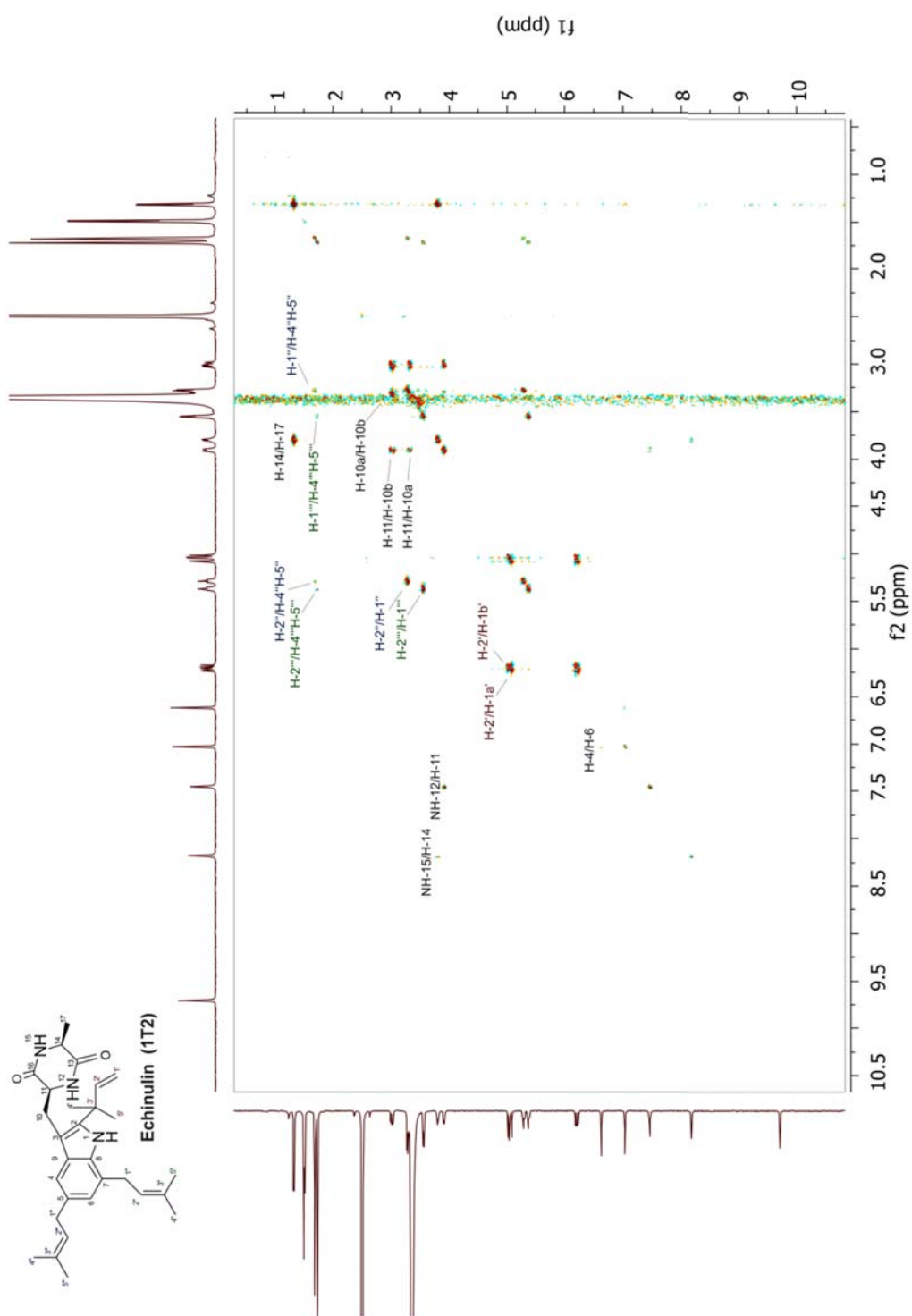
Figure S23 ¹H-¹H COSY spectrum of echinulin (**1T2**) in DMSO-d_6 (500MHz)

Figure S24 selected region (4.5. – 7.6 ppm) of ^{1}H–^{1}H COSY spectrum of echinulin (**1T2**) in DMSO-d_6 (500MHz)

123
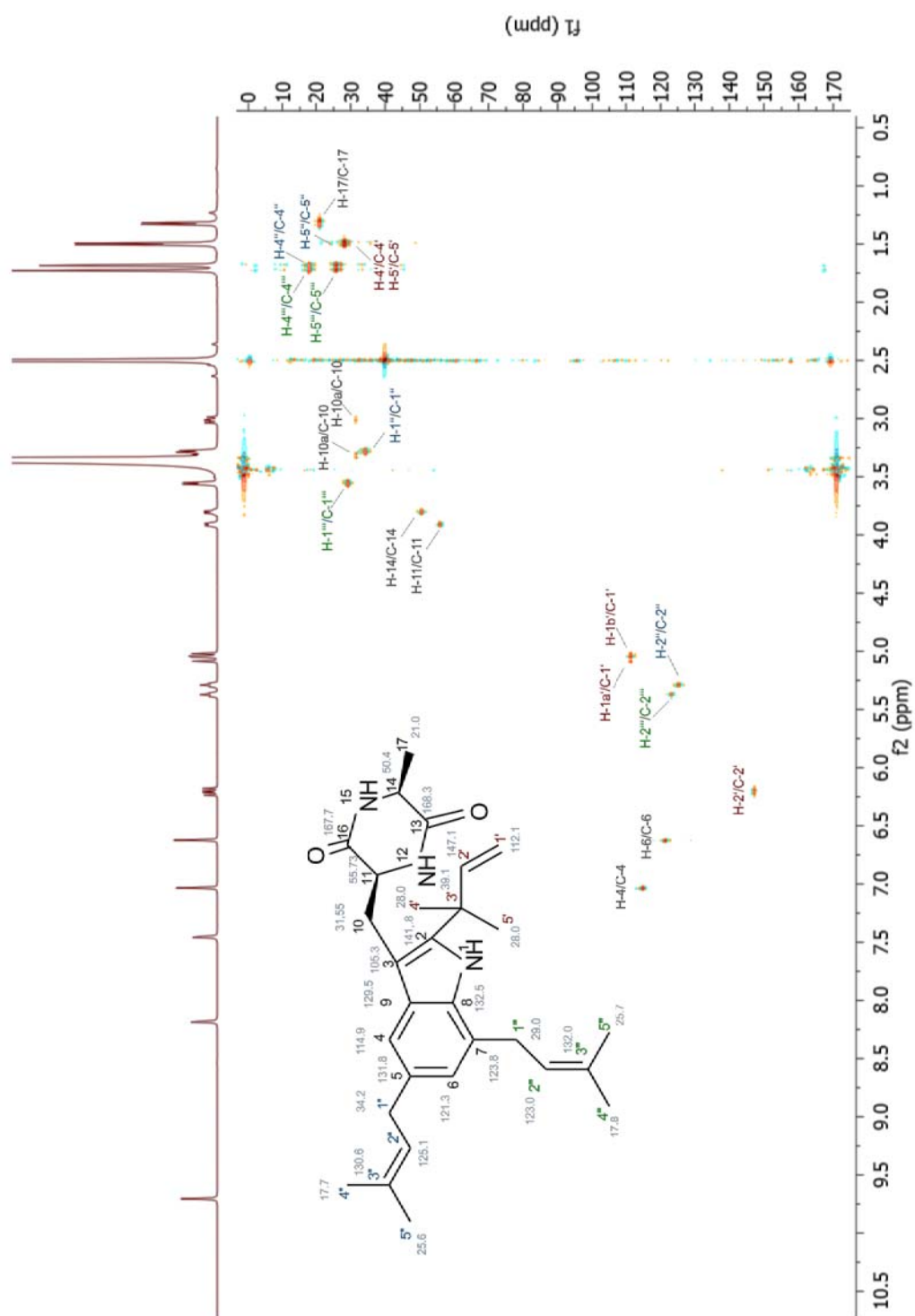
Figure S25 HSQC spectrum of echinulin (**1T2**) in DMSO-d_6 (500MHz)

Figure S26 HMBC spectrum of echinulin (**1T2**) in DMSO-d_6 (500MHz)

125
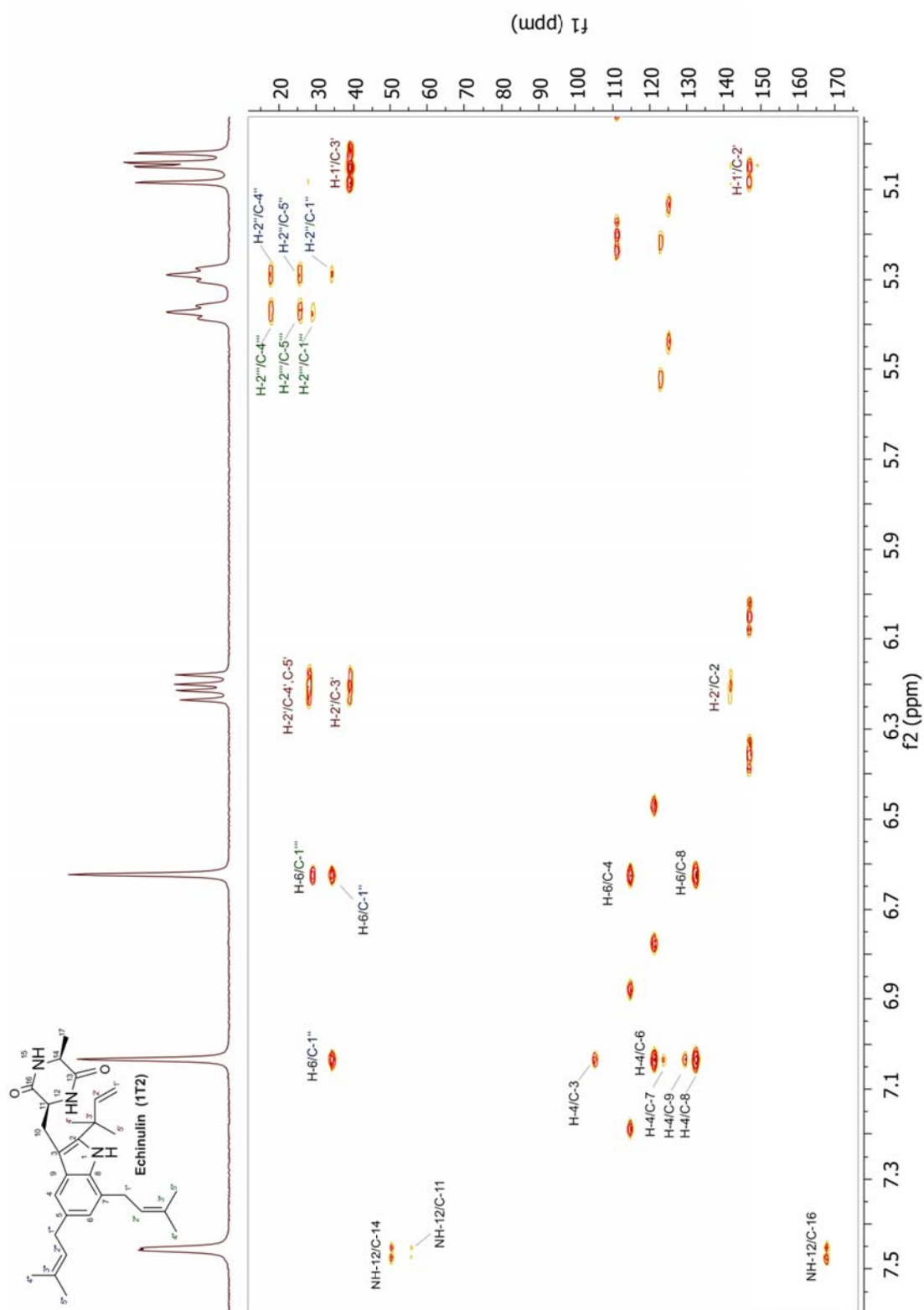
Figure S27 selected region (5.1 – 7.6 ppm) of HMBC spectrum of echinulin (**1T2**) in DMSO-d_6 (500MHz)

126

HMBC connections
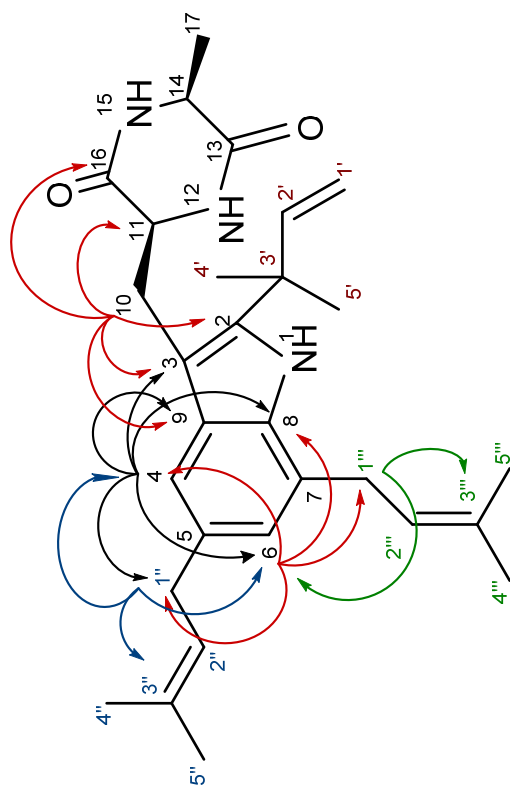
Figure S28 Selected HMBC correlations of echinulin (**1T2**) in DMSO-d_6 (500MHz)

Figure S29 ^{1}H NMR spectrum of **1Q2** in DMSO-d_6 (500MHz)
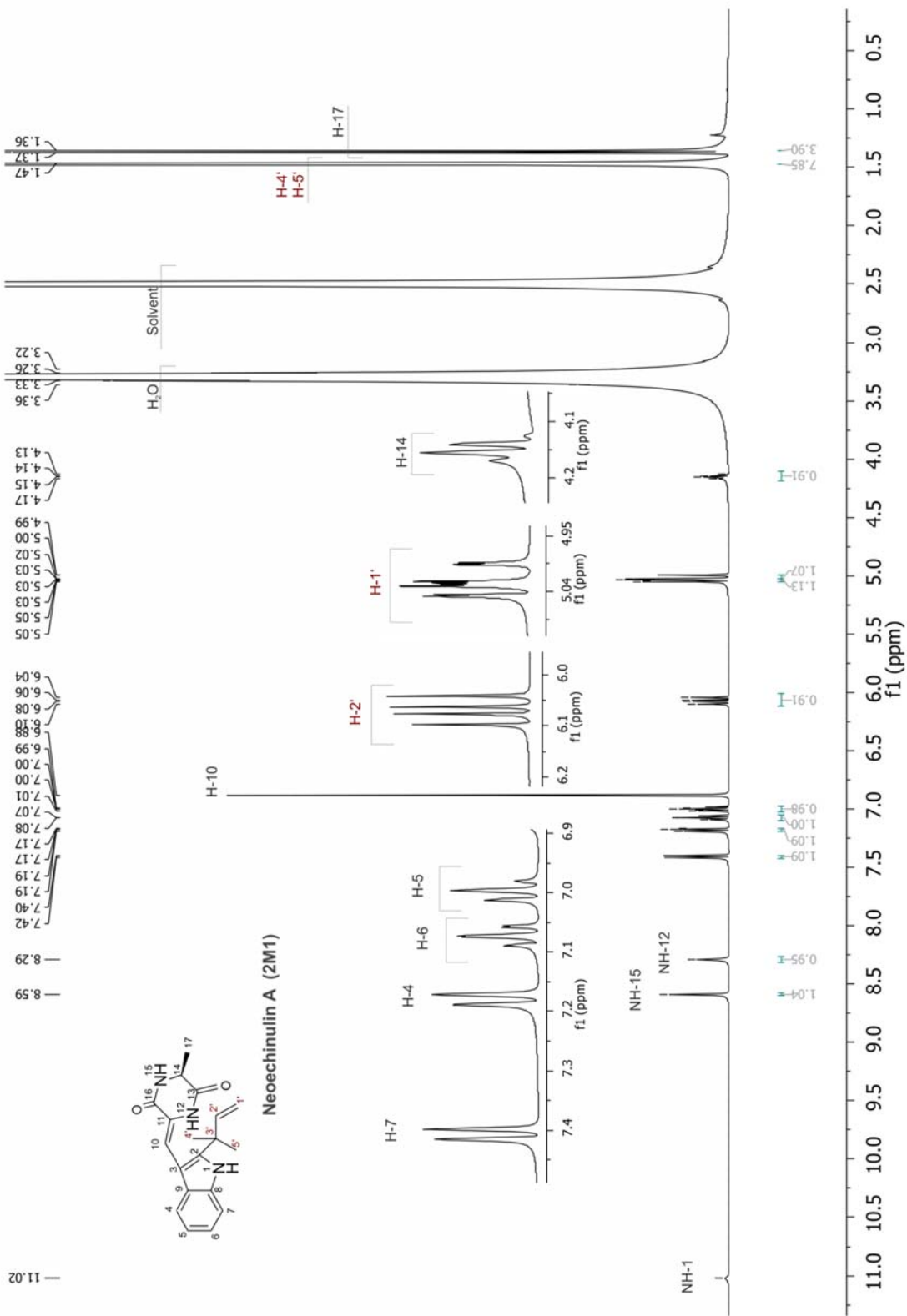
Figure S30 ^{1}H NMR spectrum of neoechinulin A (**2M1**) in DMSO-d_6 (500MHz)

Neoechinulin A (2M1)

Figure S31 ^{1}H NMR spectrum of neoechinulin B (**3M1**) in DMSO-d_6 (500MHz)
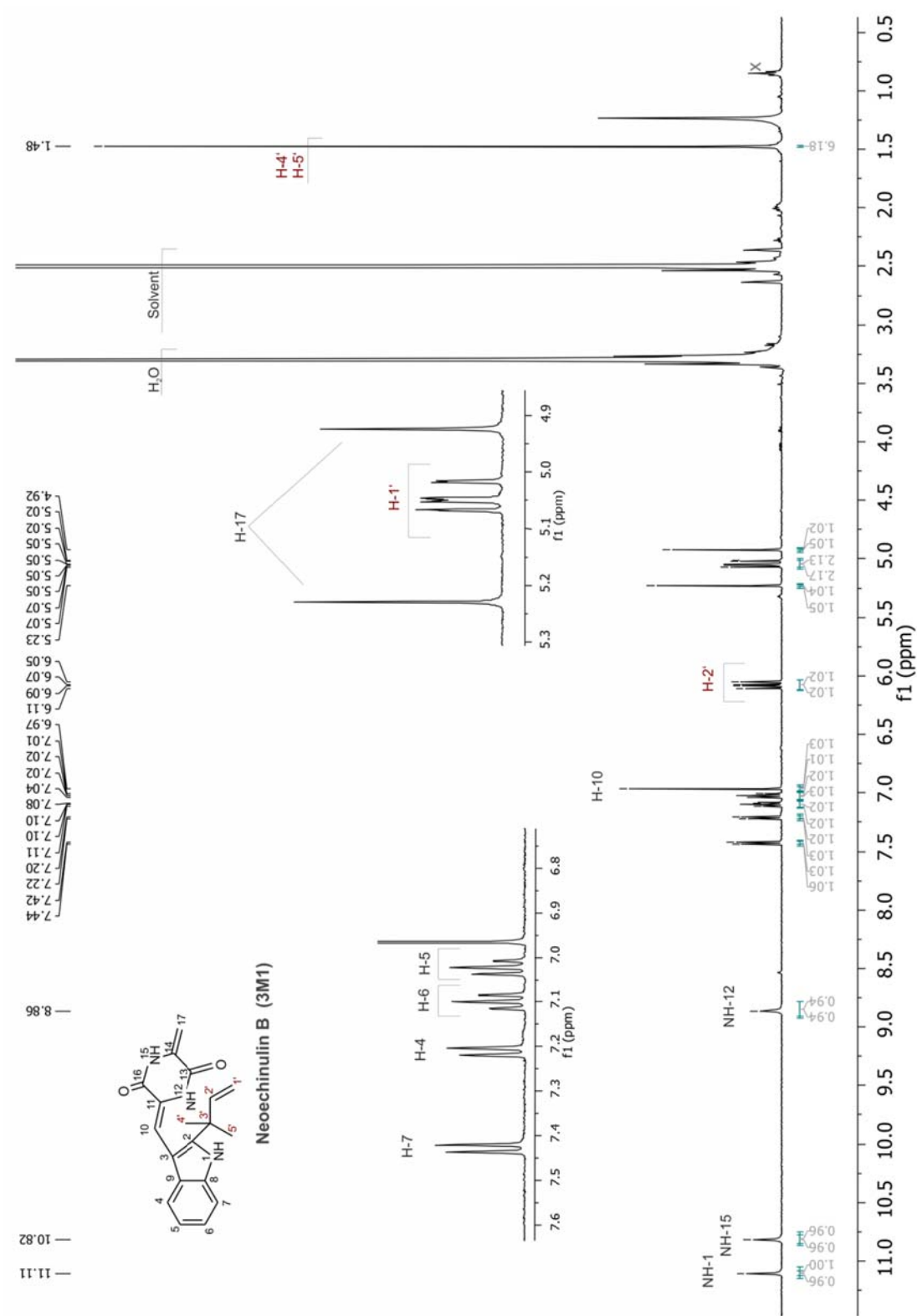
4.3. Convenient synthetic approach for tri- and tetraprenylated cyclodipeptides by consecutive enzymatic prenylations.

Wohlgemuth, V.*, **Kindinger, F.***, and Li, S. M. (2018)

*The authors contributed equally to this work.

BIOTECHNOLOGICALLY RELEVANT ENZYMES AND PROTEINS

CrossMark

Convenient synthetic approach for tri- and tetraprenylated cyclodipeptides by consecutive enzymatic prenylations

Viola Wohlgemuth [1] · Florian Kindinger [1] · Shu-Ming Li [1]

Abstract

The prenyltransferases EchPT1 and EchPT2 from *Aspergillus ruber* are responsible for the consecutive prenylations of *cyclo*-L-Trp-L-Ala, leading to the formation of the triprenylated echinulin as the predominant product. In this study, we demonstrate that EchPT1 also accepts all stereoisomers of *cyclo*-Trp-Ala and *cyclo*-Trp-Pro and catalyses regiospecific reverse *C2*-prenylation at the indole nucleus. EchPT1 products were well accepted by EchPT2 for multiple consecutive prenylations, with conversion yields of 84 to 98% for six of the eight substrates. *C2*-, *C5*- and *C7*-triprenylated derivatives are identified as major enzyme products, with product yields of 40 to 86% in seven cases. High product yields of 25–36%, i.e. approximate 30% of the total enzyme products, were observed for tetraprenylated derivatives in the four reaction mixtures with one D- and one L-configured amino acid residues. To the best of our knowledge, enzymatic preparation of tetraprenylated cyclodipeptides with such high efficacy has not been reported prior to this study.

Keywords *Aspergillus ruber* · Chemoenzymatic synthesis · Cyclodipeptides · Multiprenylated derivatives · Prenyltransferases

Introduction

Indole diketopiperazine (DKP) alkaloids exemplify the structural richness of natural products and are predominantly produced by filamentous fungi belonging to the genera *Aspergillus* and *Penicillium* of Ascomycota (Ma et al. 2016). Most DKP alkaloids are derived biogenetically from cyclodipeptides of two tryptophan molecules or a tryptophan and an additional amino acid. In fungi, the formation of such dipeptides is catalysed by non-ribosomal peptide synthetases (Xu et al. 2014). Structural diversification of the DKP backbone is subsequently achieved by different tailoring enzymes including oxidoreductases, methyltransferases,

prenyltransferases and cyclases (Giessen and Marahiel 2015; Ma et al. 2016; Winkelblech et al. 2015; Xu et al. 2014). This leads not only to more structural complexity of natural products but also improvement of their biological and pharmaceutical activities (de Sá Alves et al. 2009; Ma et al. 2016; Netz and Opatz 2015). Attachment of nxC$_5$-units to the DKP skeleton by prenyltransferases increases the lipophilicity and improves the bioavailability by better interactions with their targets such as biological membranes or proteins (Botta et al. 2005). The majority of the known fungal prenyltransferases belong to the dimethylallyltryptophan synthase (DMATS) superfamily and catalyse mostly the transfer of a dimethylallyl moiety from dimethylallyl diphosphate (DMAPP) to the indole ring in a regular or reverse manner (Winkelblech et al. 2015). For example, the biologically active indole alkaloids echinulin and congeners are multiprenylated *cyclo*-L-Trp-L-Ala derivatives. In a previous study, we demonstrated that two prenyltransferases from the DMATS superfamily, EchPT1 and EchPT2, are involved in the biosynthesis of echinulin and neoechinulins in *Aspergillus ruber* (Wohlgemuth et al. 2017).

Knowledge gained in the last years showed that the prenyltransferases of the DMATS superfamily are in general relatively strict in their acceptance of the prenyl donor, but often use a wide variety of aromatic substrates as prenyl

---

Viola Wohlgemuth and Florian Kindinger contributed equally to this work.

✉ Shu-Ming Li
shuming.li@staff.uni-marburg.de

[1] Institut für Pharmazeutische Biologie und Biotechnologie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany

Springer

acceptors (Fan et al. 2015; Winkelblech et al. 2015; Yu and Li 2012). Together with the high regio- and stereoselectivity of the reactions, this substrate promiscuity makes these enzymes interesting tools for creation of new drug candidates and for protein engineering (Chen et al. 2017; Fan et al. 2015; Yu and Li 2012). We demonstrated in previous studies that the stereochemistry of the amino acid residues in DKPs can influence the acceptance and stereoselectivity of the enzymes (Mundt and Li 2013; Yu et al. 2013). The unique feature of the aforementioned EchPT2 is its ability to catalyse a prenylation cascade of the reversely *C2*-prenylated *cyclo*-L-Trp-L-Ala, preechinulin (**1M1**), leading to the formation of di-, tri- and tetraprenylated products including tardioxopiperazine A (**1D1**), tardioxopiperazine B (**1D2**), variecolorin L (**1T1**), echinulin (**1T2**) and two tetraprenylated compounds (**1Q1** and **1Q2**) (Fig. 1) (Wohlgemuth et al. 2017). For better understanding, we adopted the nomenclature from the previous publication (Wohlgemuth et al. 2017), where "**M**" after the substrate number refers for monoprenylated, "**D**" for diprenylated, "**T**" for triprenylated and "**Q**" for tetraprenylated. The number after these letters indicates the order of the identified products. Both in in vitro assays and in the producer, **1T2** was identified as the major product. The yields of the tetraprenylated **1Q1**/**1Q2** were very low, contributing only to 4% of total conversion in enzyme assay, while in the producer just **1Q2** appears as a single tetraprenylated product. The outstanding ability of EchPT2 to catalyse a consecutive prenylation cascade prompted us to test its behaviour towards other stereoisomers, especially the degree of the prenylation cascade and the ratio of different products. For this purpose, we prepared reversely *C2*-prenylated derivatives of *cyclo*-Trp-Ala and *cyclo*-Trp-Pro isomers by EchPT1 and used them as substrates for EchPT2 reactions.

Materials and methods

Chemicals

DMAPP was synthesised according to the previously described method for geranyl diphosphate (Woodside et al. 1988). *Cyclo*-Trp-Ala and *cyclo*-Trp-Pro stereoisomers were synthesised as reported elsewhere (Yin et al. 2013).

Bacterial strains and culture conditions

Escherichia coli strain M15 (pREP4) (QIAGEN, Hilden, Germany) was used as expression host and cultivated in Terrific broth (TB) medium at 30 or 37 °C. For cultivation on solid LB medium, 1.5% (*w/v*) agar was added. Selection of recombinant *E. coli* strains was achieved with 50 μg mL^{-1} carbenicillin and 25 μg mL^{-1} of kanamycin in the precultures and 50 μg mL^{-1} carbenicillin in the main cultures.

Protein overproduction and purification

Induction of *echPT1* and *echPT2* expression was carried out in *E. coli* M15 cells with the plasmids pVW90 and pVW83, respectively (Wohlgemuth et al. 2017). Subsequent protein purification was achieved by affinity chromatography with Ni-NTA agarose resin (Qiagen) according to the manufacturer's protocol. Twenty-five milligrams of EchPT1 and 1.8 mg of EchPT2 were purified from 1 L of culture.

Enzyme assays of EchPT1 and EchPT2 reactions

EchPT1 reactions were carried out with *cyclo*-L-Trp-L-Ala (**1**) and its stereoisomers *cyclo*-L-Trp-D-Ala (**2**), *cyclo*-D-Trp-D-Ala (**3**) and *cyclo*-D-Trp-L-Ala (**4**) as well as the four stereoisomers *cyclo*-L-Trp-L-Pro (**5**), *cyclo*-L-Trp-D-Pro (**6**), *cyclo*-D-Trp-D-Pro (**7**) and *cyclo*-D-Trp-L-Pro (**8**). Standard assays (100 μL) contained Tris-HCl (50 mM, pH 7.5), one of the eight cyclodipeptides (0.5 mM), CaCl$_2$ (5 mM), DMAPP (1 mM), glycerol (0.5–5%, *v/v*), DMSO (2.5%, *v/v*) and the purified recombinant EchPT1 (1 μg, 0.20 μM).

Standard assays (100 μL) of the EchPT2 reactions contained Tris-HCl (50 mM, pH 7.5), one of the eight reversely *C2*-prenylated cyclodipeptides (**1M1**–**8M1**) (0.5 mM), CaCl$_2$ (5 mM), DMAPP (3 mM), glycerol (0.5–5%, *v/v*), DMSO (2.5%, *v/v*) and the purified recombinant EchPT2 (10 μg, 2.13 μM). Unless stated otherwise, the reaction mixtures were incubated at 37 °C for 2 h and subsequently extracted with 200 μL ethyl acetate each for three times. The organic phases were brought to dryness and the residues were dissolved in 100 μL methanol. Ninety microlitres thereof were analysed on a LC-MS system as described below.

Enzyme assays for determination of kinetic parameters

The assays for determination of the kinetic parameters of EchPT1 towards **2**–**8** contained 1 mM DMAPP, 0.05–2 μg EchPT1 and the aromatic substrates at final concentrations from 0.002 to 5.0 mM and were carried out at 37 °C for 1–30 min (Figs. S1–S7, Table 1).

The assays for determination of the kinetic parameters of EchPT2 towards **2M1**–**8M1** were carried out with 3 mM DMAPP, 2 μg EchPT2 and the aromatic substrates at final concentrations of 0.002 to 5.0 mM at 37 °C for 3–30 min (Figs. S8–S14, Table 1). The assays were performed as duplicates and terminated by addition of 100 μL methanol. The precipitated proteins were removed by centrifugation at 20000×*g* for 20 min, and 100 μL were analysed on HPLC as described below. The conversion yields were calculated from the peak areas of the respective substrate and products in HPLC chromatograms and verified by signal intensities in the ^{1}H NMR spectra of the incubation mixtures.
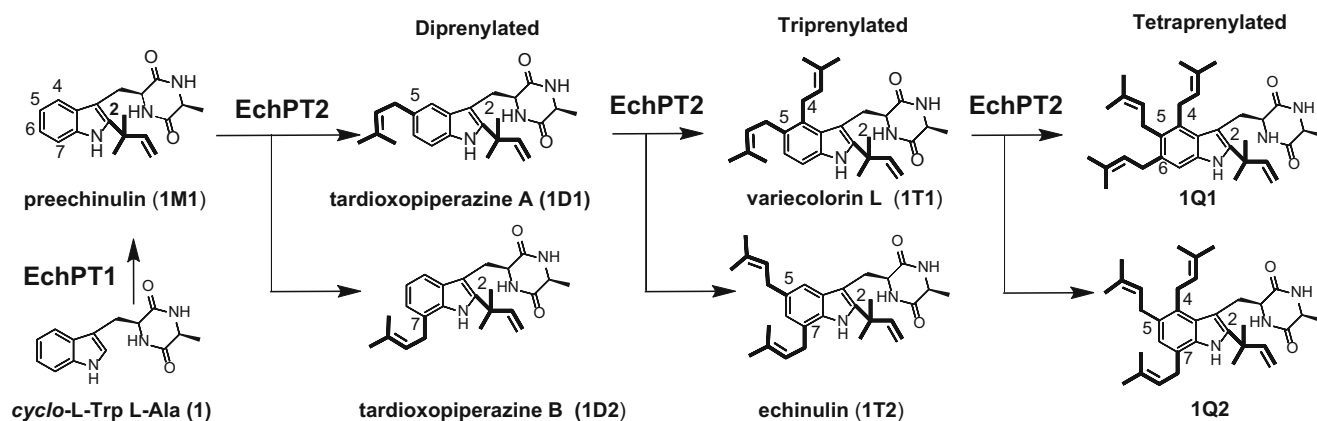
Fig. 1 Prenylation cascade catalysed by EchPT1 and EchPT2 in the biosynthesis of echinulin and derivatives

Assays for isolation of enzyme products and subsequent structure elucidation

The volumes of the enzyme assays were scaled up to 12 mL containing Tris-HCl (50 mM, pH 7.5), DMAPP (3 mM), CaCl$_2$ (10 mM), the respective aromatic substrate (1 mM) and 1–10 mg purified recombinant protein. These enzyme assays were incubated at 37 °C for 16 h and extracted three times with two volumes of ethyl acetate each. The organic phases were combined and brought to dryness on a rotating vacuum evaporator at 35 °C. The residues were dissolved afterwards in 200 μL methanol for isolation on an HPLC system with the conditions given below.

HPLC conditions for analysis and isolation

Analysis of the enzyme assays was carried out on an Agilent HPLC series 1200 (Agilent Technologies) equipped with an Agilent Eclipse XDB-C18 column (5 μm, 4.6 × 150 mm). Separation was achieved with H$_2$O and MeCN, both containing 0.1% (v/v) formic acid and a flow rate of 0.5 mL/min. A linear gradient from 20 to 80% MeCN over 10 min was followed by 25 min isocratic elution at 80% MeCN and a further linear gradient from 80 to 100% MeCN in 5 min. The column was afterwards washed with 100% MeCN for 5 min and equilibrated with 20% MeCN for 5 min. Detection was carried out with a photodiode array detector and absorptions at 277 nm were illustrated in this study.

The same HPLC system equipped with a semipreparative Multospher 120 RP-18 column (5 μm, 10 × 250 mm) was used for isolation of the enzyme products from the upscaled reaction mixtures. H$_2$O and MeCN without acid were used as solvents at a flow rate of 2.5 mL/min. Separation was done with a linear gradient of 30–80% MeCN in 15 min, followed by an isocratic elution at 80% MeCN for 18 min and an

additional linear gradient from 80 to 100% MeCN in 7 min. The column was then washed with 100% B for 10 min and subsequently equilibrated with 20% MeCN for another 10 min. The obtained fractions were further purified on an Agilent Eclipse XDB-C18 column (5 μm, 4.6 × 150 mm). The same elution profile was applied for the analysis of the enzyme assays by using H$_2$O and MeCN without acid as solvents.

LC-MS analysis

Enzyme products were analysed on an Agilent HPLC 1260 series system equipped with a photo diode array detector and a Bruker microTOF QIII mass spectrometer by using the Agilent Eclipse XDB C18 column (5 μm, 4.6 × 150 mm) according to the standard condition for this column described above. The parameters of the spectrometer were set as the following: electrospray positive ion mode for ionisation, capillary voltage with 4.5 kV, collision energy with 8.0 eV. HR-ESI-MS data of the enzyme products are given in Table S1.

NMR analysis

NMR spectra of the isolated enzyme products were recorded at room temperature on a JEOL ECA-500 (JEOL, Akishima, Tokyo, Japan). The samples were dissolved in CD$_3$OD or CDCl$_3$. All spectra were processed with MestReNova 6.0.2 (Metrelab Research, Santiago de Compostella, Spain) and the chemical shifts were referenced to those of the solvents. The NMR spectra of the monoprenylated products are supplied as Figs. S15–S24, triprenylated as Figs. S25–S39 and tetraprenylated as Figs. S40–S52. The NMR data are given in Tables S2–S6.

Table 1 Kinetic parameters of EchPT1 towards **1–8** and EchPT2 towards **1M1–8M1** in the presence of DMAPP

Enzyme	Substrates	Protein amount and incubation time		K_M [mM]	k_{cat} [s^{-1}]	k_{cat}/K_M [s^{-1} M^{-1}]
EchPT1[a]	cyclo-L-Trp-L-Ala (**1**)	2 µg	1 min	0.09 ± 0.0340	6.63 ± 0.0378	73,666.7
EchPT1	cyclo-L-Trp-D-Ala (**2**)	2 µg	5 min	0.18 ± 0.0096	0.34 ± 0.0036	1888.9
EchPT1	cyclo-D-Trp-D-Ala (**3**)	0.25 µg	30 min	1.82 ± 0.1983	0.08 ± 0.0421	95.3
EchPT1	cyclo-D-Trp-L-Ala (**4**)	2 µg	5 min	0.24 ± 0.0010	0.69 ± 0.0068	2875.0
EchPT1	cyclo-L-Trp-L-Pro (**5**)	2 µg	1 min	0.06 ± 0.0031	2.81 ± 0.0852	46,833.3
EchPT1	cyclo-L-Trp-D-Pro (**6**)	2 µg	5 min	0.17 ± 0.0097	0.34 ± 0.0055	2000.0
EchPT1	cyclo-D-Trp-D-Pro (**7**)	0.25 µg	30 min	0.39 ± 0.0015	0.06 ± 0.0102	153.8
EchPT1	cyclo-D-Trp-L-Pro (**8**)	0.05 µg	30 min	0.18 ± 0.0302	0.29 ± 0.0212	1945.0
EchPT2[a]	cyclo-L-2-tert-DMA-Trp-L-Ala (**1M1**)	2 µg	3 min	0.03 ± 0.0021	0.36 ± 0.0015	12,000.0
EchPT2	cyclo-L-2-tert-DMA-Trp-D-Ala (**2M1**)	2 µg	8 min	0.04 ± 0.0031	0.29 ± 0.0060	7250.0
EchPT2	cyclo-D-2-tert-DMA-Trp-D-Ala (**3M1**)	2 µg	10 min	0.14 ± 0.0223	0.15 ± 0.0089	1071.4
EchPT2	cyclo-D-2-tert-DMA-Trp-L-Ala (**4M1**)	2 µg	8 min	0.05 ± 0.0010	0.36 ± 0.0080	6468.2
EchPT2	cyclo-L-2-tert-DMA-Trp-L-Pro (**5M1**)	2 µg	5 min	0.07 ± 0.0014	0.42 ± 0.0047	7200.5
EchPT2	cyclo-L-2-tert-DMA-Trp-D-Pro (**6M1**)	2 µg	10 min	0.07 ± 0.0006	0.36 ± 0.0099	5142.9
EchPT2	cyclo-D-2-tert-DMA-Trp-D-Pro (**7M1**)	2 µg	30 min	0.19 ± 0.0026	0.11 ± 0.0015	578.9
EchPT2	cyclo-D-2-tert-DMA Trp-L-Pro (**8M1**)	2 µg	10 min	0.09 ± 0.0029	0.65 ± 0.0031	7222.2

[a] Data of the kinetic parameters are from a previous study and are included for comparison (Wohlgemuth et al. 2017)

Results

EchPT1 accepts very well cyclo-Trp-Ala and cyclo-Trp-Pro isomers and catalyses regiospecific reverse C2-prenylations

In a previous study, we demonstrated the regiospecific reverse C2-prenylation of cyclo-L-Trp-L-Ala (**1**) by EchPT1. In this study, **1** and its stereoisomers **2–4** as well as the four stereoisomers cyclo-Trp-Pro **5–8** were selected as prenyl acceptors. After incubation of **1–8** with 0.5 µg EchPT1 and 1 mM DMAPP at 37 °C for 10 min, the reaction mixtures were analysed on HPLC. As shown in Fig. 2, **1**, **2**, **4**, **5**, **6** and **8** were very well accepted by EchPT1, with conversion yields of 52.6 ± 2.4 to 87.3 ± 1.8%. Among the eight tested substrates, the two cyclodipeptides comprising exclusively D-configured amino acids, i.e. **3** and **7**, were poorly accepted, with conversion yields of 16.3 ± 3.9 and 42.5 ± 1.9%, respectively. Interestingly, better acceptance was observed for a given cyclo-Trp-Pro isomer than the respective counter partner of cyclo-Trp-Ala, i.e. **1** versus **5**, **2** versus **6**, **3** versus **7** as well as **4** versus **8**.

To get insights into the catalytic efficiency, kinetic parameters were determined for EchPT1 towards **2–8** under the respective best tested conditions and compared with those for **1** determined previously (Wohlgemuth et al. 2017). As shown in Figs. S1–S7, the reactions of EchPT1 with all the tested substrates followed Michaelis-Menten kinetics. Low K_M values of less than 0.1 mM and high turnover numbers of more than 2.8 s^{-1} were determined for the two substrates **1** and **5**

comprising exclusively L-configured amino acids. In contrast, high K_M values and low turnover numbers were determined for the two substrates **3** and **7** consisting of only D-configured amino acids. The other four substrates display moderate K_M values between 0.17 ± 0.0097 and 0.24 ± 0.0010 mM and turnover numbers between 0.29 ± 0.0212 and 0.69 ± 0.0068 s^{-1} (Table 1). These data correspond well to the observed conversion yields (Fig. 2). It is obvious that the stereochemistry of the amino acids plays an essential importance for their acceptance by EchPT1.

Detailed inspection of the HPLC chromatograms revealed that, as in the reaction mixture of its natural substrate **1**, only one product peak each was detected for substrates **2–8**, indicating also regiospecific prenylations for these substances. To elucidate their structures, the products **2M1–8M1** were isolated on HPLC from reaction mixtures of the respective substrates **2–8** with EchPT1 and subjected to NMR and MS analyses. HR-ESI-MS data (Table S1) confirmed the monoprenylation in **2M1–8M1** by detection of [M + H]$^+$ ions, which are 68 Da larger than those of the respective substrates. Inspection of the ^{1}H NMR spectra of **2M1–8M1** (Figs. S15–S22) revealed clear presence of signals for a reverse prenyl moiety each (Table S2) at 6.20–6.22 (dd, J = 17.5, 10.6 Hz), 5.13–5.15 (dd, J = 17.5, 1.0–1.2 Hz) and 5.10–5.12 ppm (dd, J = 10.6, 1.0–1.2 Hz). Signals of four coupling aromatic protons are present in all spectra. Comparison of these spectra with that of **1M1** reported in a previous study (Figs. S22–S24) revealed nearly the same chemical shifts and coupling constants for the aforementioned aromatic and olefinic protons in **1M1–4M1**. The same conclusion can also be drawn
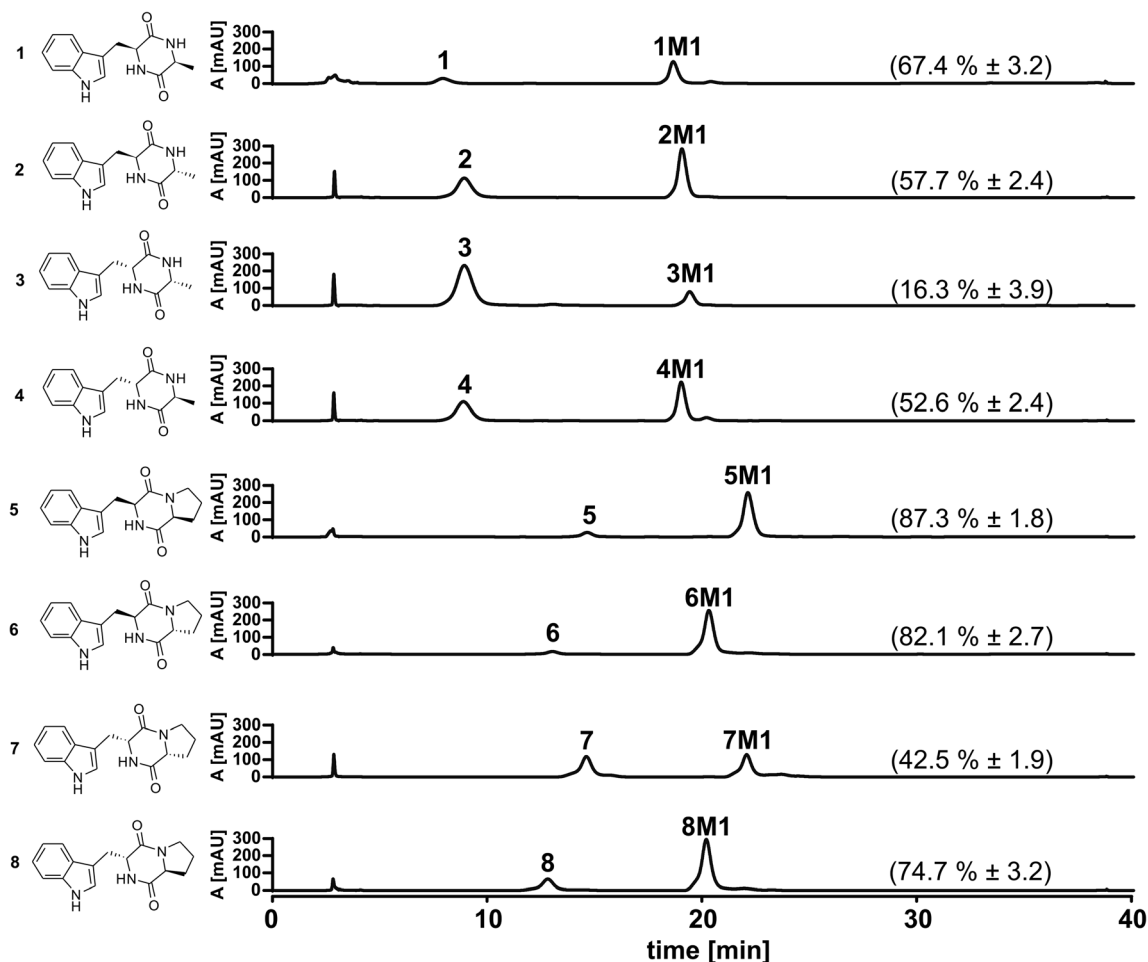
Fig. 2 HPLC analysis of EchPT1 reactions with **1–8**. The enzyme assays contained 0.5 mM aromatic substrate, 5 mM of CaCl₂, 1 mM of DMAPP and 0.5 µg of purified recombinant EchPT1 and were incubated at 37 °C and pH 7.5 for 10 min. Conversion rates are given in parentheses

for **5M1–8M1.** As expected, the spectra of the enantiomeric pairs, i.e. **1M1** and **3M1**, **2M1** and **4M1**, **5M1** and **7M1** as well as **6M1** and **8M1** are nearly identical. All these data proved unequivocally the reverse *C2*-prenylation of **2–8** by EchPT1.

EchPT2 accepts very well reverse *C2*-prenylated cyclodipeptides and catalyses similar di- and triprenylations with *C2*-, *C5*- and *C7*-prenylated derivatives as main products

Preliminary tests indicated that EchPT2 also shows broad substrate specificity for DKPs with a *C2*-prenyl moiety. **1M1–8M1** were subsequently incubated with 2.13 µM of EchPT2 in the presence of 3 mM DMAPP at 37 °C for 2 h. LC-MS analysis of the reaction mixtures showed good acceptance by EchPT2, with total product yields in the range of 21.2 ± 2.4 to 98.3 ± 1.2% (Figs. 3, 4 and 5). Again, the isomers with a (*R*, *R*)-configuration, i.e. **3M1** and **7M1**, were poorly consumed by EchPT2, which was also confirmed by determination of the

kinetic parameters (Figs. S8–S14, Table 1) (see below for details). The characteristic multiprenylation profiles with di- and triprenylated products are similar to that of **1M1**, proving the consecutive prenylations of **2M1**–**8M1** by EchPT2 (Figs. 3 and 4).

Detailed inspection of the LC-MS chromatograms of the EchPT2 reactions with **2M1**–**8M1** (Figs. 3 and 4) revealed, as in the case of **1M1**, the presence of two diprenylated derivatives **2D1**–**8D1** and **2D2**–**8D2** as minor peaks each with 68 Da larger [M + H]⁺ ions than those of **2M1**–**8M1**. A triprenylated derivative each (**2T2**–**8T2**) with 136 Da larger [M + H]⁺ ions than those of **2M1**–**8M1** was detected as the predominant product. These peaks are found in the relative position like echinulin (**1T2**) in the reaction of EchPT2 with **1M1** (Figs. 3 and 4). To elucidate their structures, **2T2**–**8T2** were isolated from incubation mixtures of **2M1**–**8M1** with EchPT2 and subjected to NMR analysis including ¹H-¹H COSY. For better comparison, the previously isolated and identified echinulin (**1T2**) (Wohlgemuth et al. 2017) was redissolved in CD₃OD and also subjected to ¹H NMR analysis (Figs. S25 and S26).

cyclo-Trp-Ala derivatives
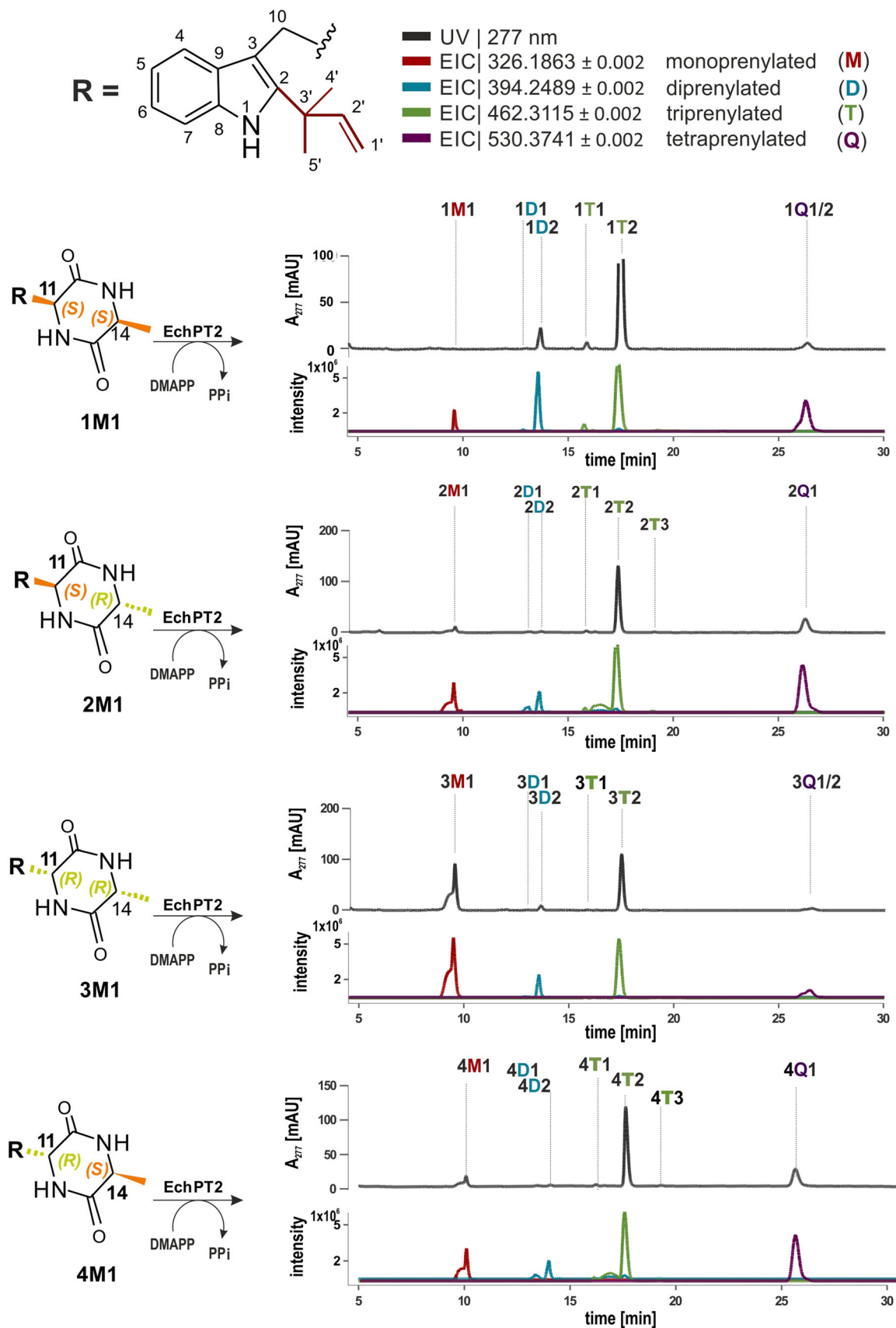
As shown in Figs. S25–S39, in addition to the above-described reverse prenyl moiety at position C-2 of the indole ring, the characteristic signals for two additional regularly attached prenyl moieties were detected (Tables S3 and S4). These include signals of two triplets of septets around 5.35 ppm, two doublets at about 3.35 and 3.55 ppm and four methyl singlets between 1.74 and 1.80 ppm. Inspection of the downfield range between 6.6–7.2 ppm for aromatic protons revealed the presence of two singlets each, with nearly the same chemical shifts in **1T2**–**4T2**. The two signals of the enantiomers **5T2** and **7T2** are slightly closer than those of another pair **6T2** and **8T2**. Comparing these results with those from previous studies (Li et al. 2008; Meng et al. 2015; Wohlgemuth et al. 2017) and the high-resolution mass data (Table S1) proves **2T2**–**8T2** to be reversely *C2*-, regularly *C5*- and *C7*-triprenylated cyclodipeptides (Fig. 6). This indicates that EchPT2 also prefers two sequential prenylations at C-5 and C-7 for all stereoisomers of *cyclo*-Trp-Ala and *cyclo*-Trp-Pro. **2T2** was identified as a racemisation product of **1T2** by treatment with triethylamine for 2 to 3 days (Westley et al. 1968). **4T2** was obtained as epiechinulin by chemical synthesis (Inoue et al. 1977; Takamatsu et al. 1971). To the best of our knowledge, the triprenylated substances **3T2** and **5T2**–**8T2** have not been described prior to this study. Differing from that of **1M1**, all the assays of **2M1**, **4M1**, **6M1** and **8M1** showed an additional small triprenylated product, namely **2T3**, **4T3**, **6T3** and **8T3**, with a slightly larger retention time than that of **2T2**, **4T2**, **6T2** and **8T2**, respectively (Figs. 3 and 4).

Product yields of *C2*-, *C4*-, *C5*- and *C7*-tetraprenylated derivatives are strongly dependent on the stereochemistry of the substrates

Most interestingly, just one tetraprenylated product each, **2Q1**, **4Q1**, **6Q1** or **8Q1**, was found in the assays of cyclodipeptide derivatives with one L- and one D-configured amino acid (Figs. 3 and 4). These products represent approximately 30% of the total conversion yields in the respective reactions (Fig. 5). In comparison, the **1M1** reaction has two tetraprenylated derivatives, **1Q1** and **1Q2**, as minor products with a total conversion yield of 4% (Figs. 3 and 5). Two minor products, **3Q1** and **3Q2**, with a total conversion yield of 4% were also observed in the assay with **3M1**. No tetraprenylated product was detected in the assays with **5M1** and **7M1** (Fig. 4).

The enzyme products **2Q1**, **4Q1**, **6Q1** and **8Q1** were isolated from the enzyme assays of the respective monoprenylated substrate with EchPT2 and subjected to NMR analysis. As shown in Figs. S40–S51, signals of one reverse and three regular prenyl moieties are clearly observed in their spectra. Comparison of the spectra of the tetraprenylated derivatives with those of the respective triprenylated counterpart (Fig. S52) revealed the presence of signals for an additional regular prenyl moiety and the disappearance of the singlet for H-4. The singlet for H-6 is still present in the range of 6.7–6.8 ppm. Together with the high-resolution mass data (Table S1), these results prove unequivocally **2Q1**, **4Q1**, **6Q1** and **8Q1** to be reversely *C2*-, regularly *C4*-, *C5*- and *C7*-tetraprenylated cyclodipeptides (Fig. 6, Tables S5 and S6). To the best of our knowledge, these compounds have not been described prior to this study.

To get further insights into the reaction efficacy of the EchPT2 reactions, kinetic parameters were determined for **2M1**–**8M1** in the presence of 3 mM DMAPP under their respective best conditions. As shown in Figs. S8–S14, the EchPT2 reactions also followed the Michaelis-Menten kinetics. High affinities with K_M values between 0.03 ± 0.0021 and 0.19 ± 0.0026 mM were determined for **2M1**–**8M1**. Turnover numbers were found from the highest value at 0.36 ± 0.0015 s^{-1} for **1M1** to the lowest value for **7M1** at 0.11 ± 0.0015 s^{-1}. Again, the two substrates, **3M1** and **7M1** derived from two D-configured amino acids, are poorly accepted by EchPT2.

Discussion

Cyclodipeptides with a DKP backbone are widely distributed in nature and mainly produced by bacteria and fungi. The DKP scaffolds are chemically condensation products of two amino acids. In fungi, DKPs are usually assembled by bimodular non-ribosomal peptide synthetases (Maiya et al. 2006; Walsh 2016; Xu et al. 2014), whereas in bacteria mainly by tRNA-dependent cyclodipeptide synthases (Giessen and Marahiel 2014; Gondry et al. 2009; James et al. 2015; Moutiez et al. 2017). Investigation in the last decades proved DKPs to be interesting candidates for drug discovery and development, due to their structural richness and their diverse biological activities (Borthwick 2012; Ma et al. 2016; Mishra et al. 2017). Their structural complexity as well as biological and pharmacological activities can be further increased by different tailoring enzymes like methyltransferase, hydroxylases, cyclases and prenyltransferases (Borthwick 2012; Giessen et al. 2013; Giessen and Marahiel 2015; Li 2010; Xu et al. 2014). Among the DKPs, tryptophan-containing members have received a remarkable increasing interest in recent years, due to the presence of the indole moiety for

cyclo-Trp-Pro derivatives
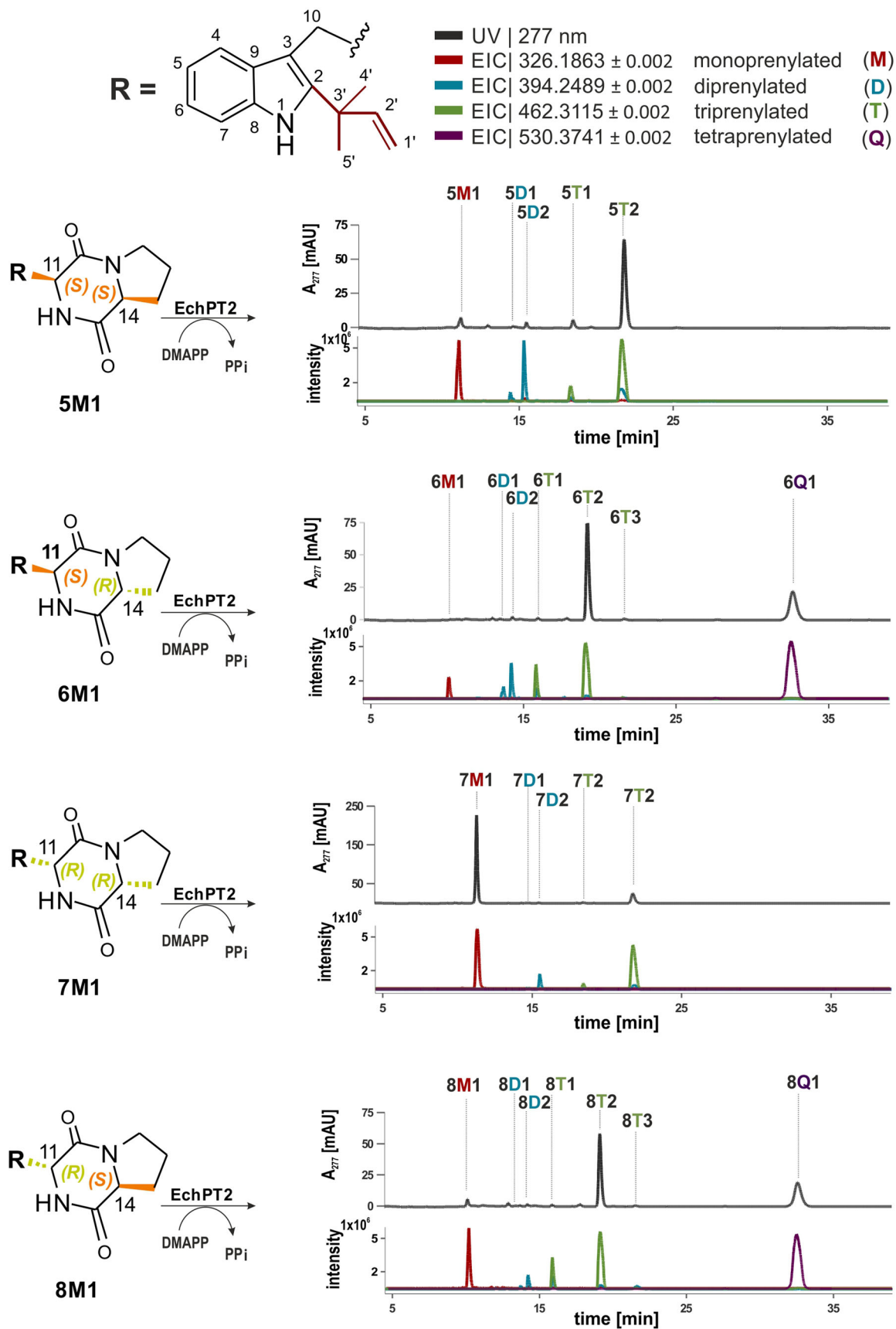
◀ **Fig. 4** LC-MS analysis of the prenylation cascade catalysed by EchPT2 with *C2*-prenylated *cyclo*-Trp-Pro isomers. The stereochemistry is visualised with orange for *S* and green for *R*. Extracted ion chromatogram (EIC) traces in the respective colours indicate the prenylation level

diversification by a large number of prenyltransferases (Winkelblech et al. 2015).

Prenylated derivatives including prenylated DKPs are products of prenyltransferases alone or together with other modification enzymes. They are important secondary metabolites in plants and microorganisms (Li 2010; Xu et al. 2014; Yazaki et al. 2009). By prenylation, the small molecules become more lipophilic and can better interact with biological membranes or proteins (Botta et al. 2005, 2009; Kim et al. 2012). Among the prenylated natural products, there are also examples of polyprenylated derivatives such as polycyclic polyprenylated xanthones and acylphloroglucinols from plants (Cottet et al. 2015; Guttroff et al. 2017) and polyprenylated indole alkaloids from fungi (Liu et al. 2014; Mundt et al. 2012). Usually, one prenyltransferase catalyses one specific prenyltransfer reaction and several prenyltransferases are involved in the formation of a given polyprenylated derivative (Mundt et al. 2012; Zou et al. 2015).

In a previous study (Wohlgemuth et al. 2017), we demonstrated the involvement of the two prenyltransferases EchPT1 and EchPT2 in the biosynthesis of echinulin (**1T2**). EchPT1 catalyses the reverse *C2*-prenylation of *cyclo*-L-Trp-L-Ala and EchPT2 afterwards a reaction cascade of up to three prenylation steps (Fig. 1). By incubation of EchPT2 with the isolated mono-, di- and triprenylated intermediates, we have shown that prenylations can take place at C-4, C-5, C-6 and C-7 of the indole ring and *C7*-prenylation leads to termination of the reaction cascade (Fig. 1). The first prenylation catalysed by EchPT2 is either at C-5 or C-7. *C2*, *C5*-diprenylated **1D1** is then converted to *C2*, *C4*, *C5*-triprenylated **1T1** and *C2*, *C5*, *C7*-triprenylated **1T2** (echinulin). **1T1** finally undergoes prenylation at C-6 or C-7, leading to the formation of the tetraprenylated **1Q1** and **1Q2** (Fig. 1). With the exception for **1Q1**, other products were also detected in the cultures of the producer (Wohlgemuth et al. 2017). However, the yields of most products are too low to be used biotechnologically. In this study, we successfully used EchPT1 and EchPT2 as biocatalysts for enzymatic preparation of eight triprenylated cyclodipeptides with high efficiency. As for its natural substrate **1M1**, *C2*, *C5*, *C7*-triprenylated derivatives were also detected as main products of EchPT2 with seven other unnatural substrates **2M1**–**8M1** (Figs. 3, 4 and 5). 17 and 42% of **7M1** and **3M1** were converted to **7T2** and **3T2**, respectively.

Fig. 5 Product yields of EchPT2 reactions

Product yields of enantiomeric pairs with (*S*, *S*) or (*R*, *R*)-configuration.
A maximum of 4 % of total yields was calculated for tetraprenylated derivatives

- 1M1 — 98.3 % ± 1.2
- 3M1 — 47 % ± 0.9
- 5M1 — 86.4 % ± 1.8
- 7M1 — 21.2 % ± 2.4

Product yields of enantiomeric pairs with (*S*, *R*) or (*R*, *S*)-configuration.
Approximate 30 % of total yields were calculated tetraprenylated derivatives

- 2M1 — 95.8 % ± 1.2
- 4M1 — 83.8 % ± 0.6
- 6M1 — 91.4 % ± 1.8
- 8M1 — 87.3 % ± 1.3

Springer

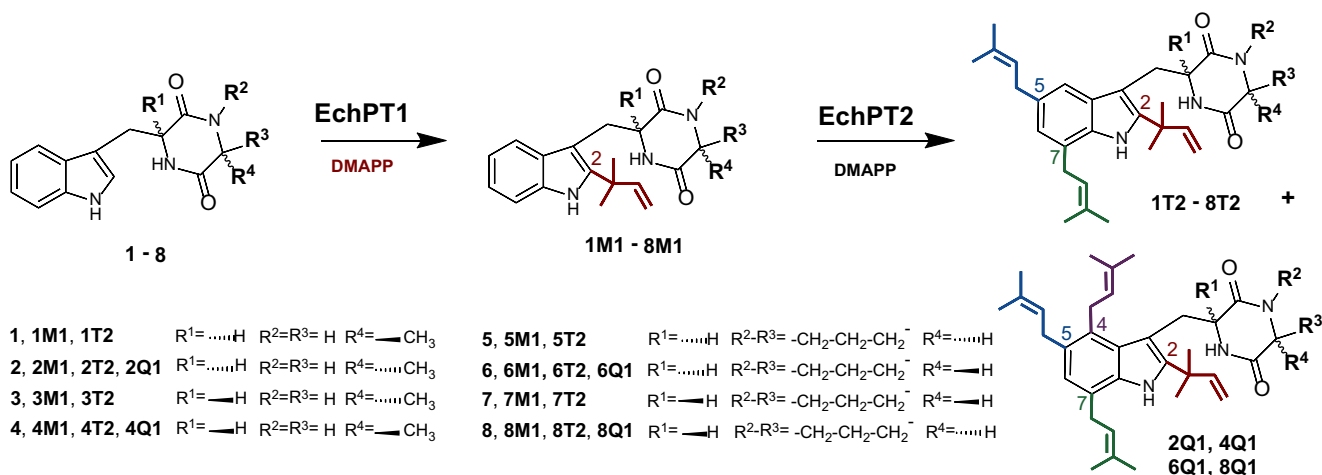**Fig. 6** Predominant products of EchPT1 and EchPT2 reactions

Higher conversion yields of 56 to 90% were found for other substrates to their respective triprenylated derivatives. In four cases, high product yields, approximate 30% of the total conversion, were also achieved for tetraprenylated derivatives from reversely *C2*-prenylated cyclodipeptides comprising one L- and one D-amino acids, i.e. **2M1**, **4M1**, **6M1** and **8M1**. It is obvious that the stereochemistry had a strong influence on the prenylation cascade catalysed by EchPT2. The stereoisomers with solely L,L-configuration, as in the cases of **1M1** and **5M1**, or D,D-configuration as in the cases of **3M1** and **7M1**, showed only low (about 4% from **1M1** and **3M1**) or no conversion (**5M1** and **7M1**) to tetraprenylated derivatives (Figs. 3, 4 and 5). The spatially larger proline compared to alanine residue seems to hinder the formation of tetraprenylated products with an L,L- or D,D-configuration. It can be expected that EchPT2 could also be successfully used for production of other tri- and tetraprenylated cyclodipeptides. Structure analysis for EchPT2, determination of the substrate binding sites and structure-based enzyme design will increase the accessibility for biotechnological application, which is now under investigation.

Compliance with ethical standards

Conflict of interest   The authors declare that they have no conflict of interest.

Human and animal rights   This article does not contain any studies with human participants or animals performed by any of the authors.

References

Borthwick AD (2012) 2,5-Diketopiperazines: synthesis, reactions, medicinal chemistry, and bioactive natural products. Chem Rev 112(7):3641–3716. https://doi.org/10.1021/cr200398y

Botta B, Vitali A, Menendez P, Misiti D, Delle MG (2005) Prenylated flavonoids: pharmacology and biotechnology. Curr Med Chem 12: 717–739

Botta B, Menendez P, Zappia G, de Lima RA, Torge R, Monachea GD (2009) Prenylated isoflavonoids: botanical distribution, structures, biological activities and biotechnological studies. An update (1995-2006). Curr Med Chem 16(26):3414–3468. https://doi.org/10.2174/092986709789057662

Chen R, Gao B, Liu X, Ruan F, Zhang Y, Lou J, Feng K, Wunsch C, Li S-M, Dai J, Sun F (2017) Molecular insights into the enzyme promiscuity of an aromatic prenyltransferase. Nat Chem Biol 13:226–234

Cottet K, Neudorffer A, Kritsanida M, Michel S, Lallemand MC, Largeron M (2015) Polycyclic polyprenylated xanthones from *Symphonia globulifera*: isolation and biomimetic electrosynthesis. J Nat Prod 78(8):2136–2140. https://doi.org/10.1021/acs.jnatprod.5b00239

de Sá Alves FR, Barreiro EJ, Fraga CAM (2009) From nature to drug discovery: the indole scaffold as a "privileged structure". Mini Rev Med Chem 9(7):782–793. https://doi.org/10.2174/138955709788452649

Fan A, Winkelblech J, Li S-M (2015) Impacts and perspectives of prenyltransferases of the DMATS superfamily for use in biotechnology. Appl Microbiol Biotechnol 99:7399–7415

Giessen TW, Marahiel MA (2014) The tRNA-dependent biosynthesis of modified cyclic dipeptides. Int J Mol Sci 15(8):14610–14631. https://doi.org/10.3390/ijms150814610

Giessen TW, Marahiel MA (2015) Rational and combinatorial tailoring of bioactive cyclic dipeptides. Front Microbiol 6:785. https://doi.org/10.3389/fmicb.2015.00785

Giessen TW, von Tesmar AM, Marahiel MA (2013) A tRNA-dependent two-enzyme pathway for the generation of singly and doubly methylated ditryptophan 2,5-diketopiperazines. Biochemistry 52(24): 4274–4283. https://doi.org/10.1021/bi4004827

Gondry M, Sauguet L, Belin P, Thai R, Amouroux R, Tellier C, Tuphile K, Jacquet M, Braud S, Courcon M, Masson C, Dubois S, Lautru S, Lecoq A, Hashimoto S, Genet R, Pernodet JL (2009)

Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. Nat Chem Biol 5:414–420

Guttroff C, Baykal A, Wang H, Popella P, Kraus F, Biber N, Krauss S, Gotz F, Plietker B (2017) Polycyclic polyprenylated acylphloroglucinols: an emerging class of non-peptide-based MRSA- and VRE-active antibiotics. Angew Chem Int Ed Engl 56: 15852–15856. https://doi.org/10.1002/anie.201707069

Inoue S, Takamatsu N, Kishi Y (1977) Synthetic studies on echinulin and related natural products. II. A total synthesis of echinulin. Yakugaku Zasshi 97:558–563

James ED, Knuckley B, Alqahtani N, Porwal S, Ban J, Karty JA, Viswanathan R, Lane AL (2015) Two distinct cyclodipeptide synthases from a marine actinomycete catalyze biosynthesis of the same diketopiperazine natural product. ACS Synth Biol 5:547–553

Kim AY, Lee CG, Lee DY, Li H, Jeon R, Ryu JH, Kim SG (2012) Enhanced antioxidant effect of prenylated polyphenols as Fyn inhibitor. Free Radic Biol Med 53(5):1198–1208. https://doi.org/10.1016/j.freeradbiomed.2012.06.039

Li S-M (2010) Prenylated indole derivatives from fungi: structure diversity, biological activities, biosynthesis and chemoenzymatic synthesis. Nat Prod Rep 27(1):57–78. https://doi.org/10.1039/B909987P

Li D-L, Li X-M, Li T-G, Dang H-Y, Wang B-G (2008) Dioxopiperazine alkaloids produced by the marine mangrove derived endophytic fungus Eurotium rubrum. Helv Chim Acta 91(10):1888–1892. https://doi.org/10.1002/hlca.200890202

Liu C, Noike M, Minami A, Oikawa H, Dairi T (2014) A fungal prenyltransferase catalyzes the regular di-prenylation at positions 20 and 21 of paxilline. Biosci Biotechnol Biochem 78:448–454

Ma YM, Liang XA, Kong Y, Jia B (2016) Structural diversity and biological activities of indole diketopiperazine alkaloids from fungi. J Agric Food Chem 64(35):6659–6671. https://doi.org/10.1021/acs.jafc.6b01772

Maiya S, Grundmann A, Li S-M, Turner G (2006) The fumitremorgin gene cluster of Aspergillus fumigatus: identification of a gene encoding brevianamide F synthetase. Chembiochem 7(7):1062–1069. https://doi.org/10.1002/cbic.200600003

Meng LH, Du FY, Li XM, Pedpradab P, Xu GM, Wang BG (2015) Rubrumazines A-C, indolediketopiperazines of the isoechinulin class from Eurotium rubrum MA-150, a fungus obtained from marine mangrove-derived rhizospheric soil. J Nat Prod 78(4):909–913. https://doi.org/10.1021/np5007839

Mishra AK, Choi J, Choi SJ, Baek KH (2017) Cyclodipeptides: an overview of their biosynthesis and biological activity. Molecules 22(10): 1796. https://doi.org/10.3390/molecules22101796

Moutiez M, Belin P, Gondry M (2017) Aminoacyl-tRNA-utilizing enzymes in natural product biosynthesis. Chem Rev 117:5578–5618

Mundt K, Li S-M (2013) CdpC2PT, a reverse prenyltransferase from Neosartorya fischeri with distinct substrate preference from known C2-prenyltransferases. Microbiology 159(Pt_10):2169–2179. https://doi.org/10.1099/mic.0.069542-0

Mundt K, Wollinsky B, Ruan HL, Zhu T, Li S-M (2012) Identification of the verruculogen prenyltransferase FtmPT3 by a combination of chemical, bioinformatic and biochemical approaches. Chembiochem 13:2583–2592

Netz N, Opatz T (2015) Marine indole alkaloids. Mar Drugs 13(8):4814–4914. https://doi.org/10.3390/md13084814

Takamatsu N, Inoue S, Kishi Y (1971) Synthetic study on echinulin and related compounds. II. Stereoselective total synthesis of optically active echinulin. Tetrahedron Lett 12(48):4665–4668

Walsh CT (2016) Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. Nat Prod Rep 33(2):127–135. https://doi.org/10.1039/C5NP00035A

Westley JW, Close VA, Nitecki DN, Halpern B (1968) Determination of steric purity and configuration of diketopiperazines by gas-liquid chromatography, thin-layer chromatography and nuclear magnetic resonance spectrometry. Anal Chem 40(12):1888–1890. https://doi.org/10.1021/ac60268a004

Winkelblech J, Fan A, Li S-M (2015) Prenyltransferases as key enzymes in primary and secondary metabolism. Appl Microbiol Biotechnol 99:7379–7397

Wohlgemuth V, Kindinger F, Xie X, Wang BG, Li S-M (2017) Two prenyltransferases govern a consecutive prenylation cascade in the biosynthesis of echinulin and neoechinulin. Org Lett 19(21):5928–5931. https://doi.org/10.1021/acs.orglett.7b02926

Woodside AB, Huang Z, Poulter CD (1988) Trisammonium geranyl diphosphate. Org Synth 66:211–215

Xu W, Gavia DJ, Tang Y (2014) Biosynthesis of fungal indole alkaloids. Nat Prod Rep 31(10):1474–1487. https://doi.org/10.1039/C4NP00073K

Yazaki K, Sasaki K, Tsurumaru Y (2009) Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. Phytochemistry 70:1739–1745

Yin S, Yu X, Wang Q, Liu XQ, Li S-M (2013) Identification of a brevianamide F reverse prenyltransferase BrePT from Aspergillus versicolor with a broad substrate specificity towards tryptophan-containing cyclic dipeptides. Appl Microbiol Biotechnol 97(4): 1649–1660. https://doi.org/10.1007/s00253-012-4130-0

Yu X, Li S-M (2012) Prenyltransferases of the dimethylallyltryptophan synthase superfamily. Methods Enzymol 516:259–278. https://doi.org/10.1016/B978-0-12-394291-3.00005-8

Yu X, Zocher G, Xie X, Liebhold M, Schütz S, Stehle T, Li S-M (2013) Catalytic mechanism of stereospecific formation of cis-configured prenylated pyrroloindoline diketopiperazines by indole prenyltransferases. Chem Biol 20(12):1492–1501. https://doi.org/10.1016/j.chembiol.2013.10.007

Zou Y, Zhan Z, Li D, Tang M, Cacho RA, Watanabe K, Tang Y (2015) Tandem prenyltransferases catalyze isoprenoid elongation and complexity generation in biosynthesis of quinolone alkaloids. J Am Chem Soc 137:4980–4983

Convenient synthetic approach for tri- and tetraprenylated cyclodipeptides by consecutive enzymatic prenylations

Viola Wohlgemuth[1], Florian Kindinger[1], Shu-Ming Li†*

Institut für Pharmazeutische Biologie und Biotechnologie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany.

[1]These authors contributed equally to this study.

Corresponding Author

*Tel +49 6421 28 22461. shuming.li@staff.uni-marburg.de

ORCID Shu-Ming Li: 0000-0003-4583-2655

CONTENTS

1 TABLE OF LC-MS DATA

Table S1 HR-ESI-MS data of EchPT1 and EchPT2 reaction products

compound	chemical formula	HR-ESI-MS [M + H]⁺		Deviation (ppm)
		calculated	measured	
cyclo-L-2-*tert*-DMA-Trp-D-Ala (**2M1**)	$C_{19}H_{23}N_3O_2$	326.1863	326.1864	0.3
2D1	$C_{24}H_{31}N_3O_2$	394.2489	394.2491	0.5
2D2	$C_{24}H_{31}N_3O_2$	394.2489	394.2486	-0.7
2T1	$C_{29}H_{39}N_3O_2$	462.3115	462.3125	2.1
cyclo-L-2-*tert*-DMA-5,7-di-DMA-Trp-D-Ala (**2T2**)	$C_{29}H_{39}N_3O_2$	462.3115	462.3126	2.4
2T3	$C_{29}H_{39}N_3O_2$	462.3115	462.3116	0.2
cyclo-L-2-*tert*-DMA-4,5,7-tri-DMA-Trp-D-Ala (**2Q1**)	$C_{34}H_{47}N_3O_2$	530.3741	350.3740	-0.2
cyclo-D-2-*tert*-DMA-Trp-D-Ala (**3M1**)	$C_{19}H_{23}N_3O_2$	326.1863	326.1874	3.4
3D1	$C_{24}H_{31}N_3O_2$	394.2489	394.2468	-5.3
3D2	$C_{24}H_{31}N_3O_2$	394.2489	394.2474	-3.8
3T1	$C_{29}H_{39}N_3O_2$	462.3115	462.3094	-4.5
cyclo-D-2-*tert*-DMA-5,7-di-DMA-Trp-D-Ala (**3T2**)	$C_{29}H_{39}N_3O_2$	462.3115	462.3122	1.5
3Q1/2	$C_{34}H_{47}N_3O_2$	530.3741	530.3732	-1.6
cyclo-D-2-DMA-Trp-L-Ala (**4M1**)	$C_{19}H_{23}N_3O_2$	326.1863	326.1859	-1.2
4D1	$C_{24}H_{31}N_3O_2$	394.2489	394.2473	-4.1
4D2	$C_{24}H_{31}N_3O_2$	394.2489	394.2483	-1.5
4T1	$C_{29}H_{39}N_3O_2$	462.3115	462.3109	-1.3
cyclo-D-2-DMA-5,7-di-DMA-Trp-L-Ala (**4T2**)	$C_{29}H_{39}N_3O_2$	462.3115	462.3123	1.7
4T3	$C_{29}H_{39}N_3O_2$	462.3115	462.3113	-0.4
cyclo-D-2-DMA-Trp-L-Ala (**4Q1**)	$C_{34}H_{47}N_3O_2$	530.3741	530.3732	-1.7

compound	chemical formula	HR-ESI-MS [M + H]$^+$		Deviation (ppm)
		calculated	measured	
cyclo-L-2-DMA-Trp-L-Pro (**5M1**)	$C_{21}H_{25}N_3O_2$	352.2020	352.2031	3.1
5D1	$C_{26}H_{33}N_3O_2$	420.2646	420.2628	-4.1
5D2	$C_{26}H_{33}N_3O_2$	420.2646	420.2633	-3.1
5T1	$C_{31}H_{41}N_3O_2$	488.3272	488.3264	-1.6
cyclo-L-2-DMA-5,7-di-DMA- Trp-L-Pro (**5T2**)	$C_{31}H_{41}N_3O_2$	488.3272	488.3274	-0.4
cyclo-L-2-DMA-Trp-D-Pro (**6M1**)	$C_{21}H_{25}N_3O_2$	352.2020	352.2015	-1.4
6D1	$C_{26}H_{33}N_3O_2$	420.2646	420.2633	-2.8
6D2	$C_{26}H_{33}N_3O_2$	420.2646	420.2632	-3.1
6T1	$C_{31}H_{41}N_3O_2$	488.3272	488.3265	-1.4
cyclo-L-2-DMA-5,7-di-DMA-Trp-D-Pro (**6T2**)	$C_{31}H_{41}N_3O_2$	488.3272	488.3261	-2.2
6T3	$C_{31}H_{41}N_3O_2$	488.3272	488.3259	-2.6
cyclo-L-2-DMA-4,5,7-tri-DMA-Trp-D-Pro (**6Q1**)	$C_{36}H_{49}N_3O_2$	556.3898	556.3886	-2.2
cyclo-D-2-DMA-Trp-D-Pro (**7M1**)	$C_{21}H_{25}N_3O_2$	352.2020	352.2032	3.4
7D1	$C_{26}H_{33}N_3O_2$	420.2646	420.2634	-2.8
7D2	$C_{26}H_{33}N_3O_2$	420.2646	420.2638	-1.9
7T1	$C_{31}H_{41}N_3O_2$	488.3272	488.3261	-2.3
cyclo-D-2-DMA-5,7-di-DMA-Trp-D-Pro (**7T2**)	$C_{31}H_{41}N_3O_2$	488.3272	488.3265	-1.4
cyclo-D-2-DMA-Trp-L-Pro (**8M1**)	$C_{21}H_{25}N_3O_2$	352.2020	352.2025	1.4
8D1	$C_{26}H_{33}N_3O_2$	420.2646	420.2643	-0.7
8D2	$C_{26}H_{33}N_3O_2$	420.2646	420.2635	-2.6
8T1	$C_{31}H_{41}N_3O_2$	488.3272	488.3273	0.2
cyclo-D-2-DMA-5,7-di-DMA-Trp-L-Pro (**8T2**)	$C_{31}H_{41}N_3O_2$	488.3272	488.3277	1.9
8T2	$C_{31}H_{41}N_3O_2$	488.3272	488.3255	-3.5
cyclo-D-2-DMA-4,5,7-tri-DMA-Trp-L-Pro (**8Q1**)	$C_{36}H_{49}N_3O_2$	556.3898	556.3908	1.8

2 TABLES OF NMR DATA

Table S2 ^{1}H NMR data of the enzyme products of EchPT1 reactions (**1M1 - 8M1**) in CD₃OD

Pos.	**1M1** cyclo-L-2-tert DMA-Trp-L-Ala[a] δH, multi	J in Hz	**2M1** cyclo-L-2-tert DMA-Trp-D-Ala δH, multi	J in Hz	**3M1** cyclo-D-2-tert DMA-Trp-D-Ala δH, multi	J in Hz	**4M1** cyclo-D-2-tert DMA-Trp-L-Ala δH, multi	J in Hz	**5M1** cyclo-L-2-tert DMA-Trp-L-Pro δH, multi	J in Hz	**6M1** cyclo-L-2-tert DMA-Trp-D-Pro δH, multi	J in Hz	**7M1** cyclo-D-2-tert DMA-Trp-D-Pro δH, multi	J in Hz	**8M1** cyclo-D-2-tert DMA-Trp-L-Pro δH, multi	J in Hz
NH-1	-		-		-		-		-		-		-		-	
4	7.51, dd	7.9, 1.1	7.52, dd	8.0, 1.0	7.51, dd	8.0, 1.1	7.52, dd	8.0, 1.0	7.47, dd	8.0, 1.1	7.46, dd	7.9, 1.1	7.47, br d	8.0	7.45, dd	8.0, 1.1
5	6.99, ddd	7.9, 7.0, 1.1	6.99, ddd	8.0, 7.1, 1.1	6.99, ddd	8.0, 7.1, 1.1	6.99, ddd	8.0, 7.1, 1.1	7.02, ddd	8.0, 7.1, 1.1	6.98, ddd	8.0, 7.1, 1.1	7.02, ddd	8.0, 7.1, 1.0	6.97, ddd	8.0, 7.1, 1.1
6	7.06, ddd	7.9, 7.0, 1.1	7.05, ddd	8.0, 7.1, 1.1	7.06, ddd	8.0, 7.1, 1.1	7.05, ddd	8.0, 7.1, 1.1	7.09, ddd	8.1, 7.1, 1.1	7.04, ddd	8.1, 7.1, 1.1	7.09, ddd	8.1, 7.1, 1.0	7.04, ddd	8.0, 7.1, 1.1
7	7.33, dd	7.9, 1.1	7.32, dd	8.0, 1.0	7.33, dd	8.0, 1.1	7.32, dd	8.0, 1.0	7.37, dd	8.1, 1.1	7.30, dd	8.1, 1.1	7.37, br d	8.1	7.30, dd	8.0, 1.1
10a	3.54, dd	14.8, 3.8	3.44, dd	14.6, 4.4	3.54, dd	14.8, 3.8	3.44, dd	14.6, 4.4	3.65, dd	15.1, 4.3	3.45, dd	14.8, 5.1	3.65, dd	15.1, 4.3	3.44, dd	14.8, 5.1
10b	3.30#		3.30#		3.30#		3.30#		3.14, dd	15.1, 11.2	3.38, dd	14.8, 6.6	3.14, dd	15.1, 11.2	3.37, dd	14.8, 6.6
11	4.28, ddd	9.6, 3.8, 1.2	4.21, dd	9.0, 4.4	4.27, dd	9.6, 3.8	4.21, dd	9.1, 4.4	4.52, ddd	11.1, 4.2, 1.8	4.18, dd	6.4, 5.2	4.53, ddd	11.2, 4.2, 1.8	4.18, dd	6.5, 5.2
NH-12	-		-		-		-		-		-		-		-	
14	3.95, dq	7.0, 0.7	3.80, q	6.9	3.94, q	7.1	3.80, q	7.0	4.23, ddd	8.5, 6.5, 1.7	3.30#		4.23, ddd	8.4, 6.4, 1.7	3.30#	
NH-15	-		-		-		-									
17	1.3, d	7.1	1.32, d	6.9	1.3, d	7.1	1.32, d	7.0	3.61, m; 3.54, m		3.47, m; 3.30#		3.61, m; 3.54, m		3.46, m; 3.30#	
18									2.03, m; 1.93, m		1.88, m; 1.51, m		2.03, m; 1.94, m		1.88, m; 1.51, m	
19									2.28, m; 1.93, m		2.09, m; 1.74, m		2.28, m; 1.94, m		2.08, m; 1.73, m	
1'	5.15, dd; 5.11, dd	17.5, 1.2; 10.6, 1.2	5.15, dd; 5.10, dd	17.5, 1.2; 10.6, 1.2	5.15, dd; 5.11, dd	17.5, 1.2; 10.6, 1.2	5.15, dd; 5.10, dd	17.5, 1.2; 10.6, 1.2	5.15, dd; 5.10, dd	17.5, 1.1; 10.6, 1.1	5.16, dd; 5.12, dd	17.5, 1.1; 10.6, 1.1	5.13, dd; 5.10, dd	17.5, 1.0; 10.6, 1.0	5.15, dd; 5.11, dd	17.5, 1.2; 10.6, 1.2
2'	6.21, dd	17.5, 10.6	6.22, dd	17.7, 10.6	6.21, dd	17.5, 10.6	6.22, dd	17.5, 10.6	6.22, dd	17.5, 10.6,	6.21, dd	17.5, 10.6	6.22, dd	17.5, 10.6	6.20, dd	17.5, 10.6
4'	1.565, s		1.55, s		1.57, s		1.565, s		1.57, s		1.54, s		1.57, s		1.53, s	
5'	1.560, s		1.54, s		1.56, s		1.560, s		1.56, s		1.53, s		1.56, s		1.52, s	

#Signals overlapping with those of solvent; All spectral data are consistent with those reported previously(Yin et al. 2013); [a]: ^{1}H NMR data are from a previous study and are included for comparison(Wohlgemuth et al. 2017)

148

Table S3 ^{1}H NMR data of the enzyme products of EchPT2 reactions (**1T2 – 4T2**) in CD$_3$OD

Pos.		**1T2** cyclo-L-2-tert-DMA-5,7-di-DMA-Trp-L-Ala δ$_H$, multi, J in Hz	COSY	**2T2** cyclo-L-2-tert-DMA-5,7-di-DMA-Trp-D-Ala δ$_H$, multi, J in Hz	COSY	**3T2** cyclo-D-2-tert-DMA-5,7-di-DMA-Trp-D-Ala δ$_H$, multi, J in Hz	**4T2** cyclo-D-2-tert-DMA-5,7-di-DMA-Trp-L-Ala δ$_H$, multi, J in Hz
NH-1		-		-		-	-
4	1H	7.14, s	H-6	7.15, s	H-6; H-1''	7.14, s	7.15, s
5		-		-		-	-
6	1H	6.70, s	H-4	6.70, s	H-4; H-1'', H-1'''	6.70, s	6.70, s
7		-		-			
10a	1H	3.52, dd, 14.7, 3.8	H-11, H-10b	3.41, dd, 14.6, 4.2		3.52, dd, 14.7, 3.8	3.41, dd, 14.6, 4.2
10b	1H	3.23, dd, 14.7, 9.9	H-11, H-10a	3.30#		3.23, dd, 14.7, 9.9	3.30#
11	1H	4.26, ddd, 9.9, 3.8, 0.7	H-10a, H-10b	4.19, dd, 9.6, 4.2	H-10a, H-10b	4.25, ddd, 9.9, 3.9, 0.7	4.19, dd, 9.7, 4.2
NH-12		-		-		-	
14	1H	3.96, dq, 7.3, 1.0	H-17	3.90, q, 6.9	H-17	3.96, dq, 7.3, 0.9	3.90, q, 6.9
NH-15		-		-		-	
17	3H	1.4, d, 7.1	H-14	1.36, d, 7.0	H-14	1.4, d, 7.1	1.36, d, 7.0
18							
19							
1'	1H	5.16, dd, 17.5, 1.2	**H-2', H-1b'**	5.16, dd, 17.5, 1.1	**H-2', H-1b'**	5.16, dd, 17.5, 1.1	5.16, dd, 17.5, 1.2
	1H	5.12, dd, 10.6, 1.2	**H-2', H-1a'**	5.12, dd, 10.6, 1.1	**H-2', H-1a'**	5.12, dd, 10.6, 1.1	5.12, dd, 10.6, 1.2
2'	1H	6.21, dd, 17.5, 10.6	**H-1'**	6.22, dd, 17.5, 10.6	**H-1'**	6.21, dd, 17.5, 10.6	6.22, dd, 17.5, 10.6
4'	3H	1.57, s		1.55, s		1.57, s	1.55, s
5'	3H	1.56, s		1.54, s		1.56, s	1.54, s
1''	2H	3.37, d, 6.9	**H-2'', H-4''/H-5''**	3.37, d, 7.3	**H-2'', H-4''/H-5''**	3.38, d, 7.0	3.37, d, 7.3
2''	1H	5.36 tsept, 7.3, 1.4a	**H-1'', H-4''/H-5''**	5.36, tsept, 7.3, 1.4b	**H-1'', H-4''/H-5''**	5.36, tsept, 7.3, 1.4c	5.38, tsept, 7.3, 1.4 d
4''	3H	1.745, s	**H-1'', H-2''**	1.75, s	**H-1'', H-2''**	1.745, s	1.75, s
5''	3H	1.740, s	**H-1'', H-2''**	1.74, s	**H-1'', H-2''**	1.740, s	1.74, s
1'''	2H	3.56, d, 7.2	**H-2''', H-4'''/H-5'''**	3.55, d, 7.3	**H-2''', H-4'''/H-5'''**	3.56, d, 7.3	3.55, d, 7.3
2'''	1H	5.38, tsept, 7.3, 1.4a	**H-1''', H-4'''/H-5'''**	5.38, tsept, 7.3, 1.4b	**H-1''', H-4'''/H-5'''**	5.38, tsept, 7.3, 1.4c	5.36, tsept, 7.3, 1.4 d
4'''	3H	1.80, s	**H-1''', H-2'''**	1.80, s	**H-1''', H-2'''**	1.80, s	1.80, s
5'''	3H	1.78, s	**H-1''', H-2'''**	1.78, s	**H-1''', H-2'''**	1.78, s	1.78, s

#Signals overlapping with those of solvent.; a–d signals with same letters overlap with each other;

149

Table S4 ¹H NMR data of the enzyme products of EchPT2 reactions (**5T2 – 8T2**) in CD₃OD

Pos.		5T2 cyclo-L-2-tert-DMA-5,7-di-DMA-Trp-L-Pro δH, multi, J in Hz	5T2 COSY	6T2 cyclo-L-2-tert-DMA-5,7-di-DMA-Trp-D-Pro δH, multi, J in Hz	6T2 COSY	7T2 cyclo-D-2-tert-DMA-5,7-di-DMA-Trp-D-Pro δH, multi, J in Hz	8T2 cyclo-D-2-tert-DMA-5,7-di-DMA-Trp-L-Pro δH, multi, J in Hz
NH-1		-		-		-	-
4	1H	7.08, s	H-6; H-1''	7.11, s	H-6;	7.08, s	7.11, s
5		-		-		-	-
6	1H	6.73, s	H-4; H-1'''	6.69, s	H-4; H-1'''	6.74, s	6.69, s
7		-		-		-	-
10a	1H	3.64, dd, 15.1, 4.3	H-10b, H-11	3.43, dd, 14.5, 5.2	H-10b, H-11	3.64, dd^f, 15.1, 4.3	3.43, dd^k, 14.6, 5.2
10b	1H	3.09, dd, 15.1, 11.2	H-10a, H-11	3.36, dd, 14.5, 6.5	H-10a, H-11	3.10, dd, 15.1, 11.2	3.36^h
11	1H	4.46, ddd, 11.2, 4.3, 1.9	H-10a, H-10b	4.15, t, 5.2	H-10a, H-10b	4.47, ddd, 11.2, 4.3, 1.7	4.15, t, 5.2
NH-12							
14	1H	4.22, ddd, 10.3, 6.9, 1.6	-	3.30$^\#$	-	4.23, ddd, 10.4, 7.0, 1.8	3.30$^\#$
17	1H	3.61, m^a	H-18, H-19	3.45, m	H-18	3.61, m^f	3.45, m^k
	1H	3.53, m	H-18, H-19	3.30$^\#$	H-18	3.53, m	3.30$^\#$
18	1H	2.01, m	H-19, H-17,	1.88, m	H-19, H-17	2.01, m	1.88, m
	1H	1.92, m^b	H-19, H-17,	1.51, m^c	H-19, H-17	1.93, m^g	1.51, m^l
19	1H	2.27, m	H18, H-17, H-14	2.08, m	H18, H-14	2.28, m	2.08, m
	1H	1.92, m^b	H18, H-17, H-14	1.73, m^d	H18, H-14	1.93, m^g	1.73, m^j
1'	1H	5.14, dd, 17.5, 1.1	H-2', H-1b'	5.16, dd, 17.5, 1.1	H-2', H-1b'	5.15, dd, 17.6, 1.0	5.16, dd, 17.5, 1.1
	1H	5.11, dd, 10.1, 1.1	H-2', H-1a'	5.12, dd, 10.6, 1.1	H-2', H-1a'	5.12, dd, 10.7, 1.0	5.12, dd, 10.6, 1.1
2'	1H	6.23, dd, 17.5, 10.6	H-1'	6.21, dd, 17.5, 10.6	H-1'	6.23, dd, 17.5, 10.6	6.21, dd, 17.5, 10.6
4'	3H	1.57, s		1.54, br. s^c		1.58, s	1.54, br. s^j
5'	3H	1.56, s		1.54, br. s^c		1.56, s	1.54, br. s^j
1''	2H	3.37, d, 7.3	H-2'', H-4''/H-5'', H-4	3.36, d, 7.3	H-2'', H-4''/H-5'', H-4	3.36, d, 7.3	3.36, d^h, 7.3
2''	1H	5.34, tsept, 7.3, 1.4	H-1'', H-4''/H-5''	5.35, tsept, 7.3, 1.4^e	H-1'', H-4''/H-5''	5.34, tsept, 7.3, 1.4	5.35, t^l 7.3
4''	3H	1.74, s	H-1'', H-2''	1.76, s^d	H-1'', H-2''	1.74, s	1.76, s^j
5''	3H	1.73, s	H-1'', H-2''	1.74, s^d	H-1'', H-2''	1.73, s	1.74, s^j
1'''	2H	3.58, d^b, 7.3	H-2''', H-4'''/H-5'''; H-6	3.54, d, 7.3	H-2''', H-4'''/H-5'''; H-6	3.58, d^f, 7.3	3.54, d, 7.3
2'''	1H	5.40, tsept, 7.3, 1.4	H-1''', H-4''/H-5'''	5.37, tsept, 7.3, 1.4^e	H-1''', H-4''/H-5'''	5.41, tsept, 7.3, 1.4	5.35, t^l 7.3
4'''	3H	1.80, s	H-1''', H-2'''	1.78, s	H-1''', H-2'''	1.80, s	1.78, s
5'''	3H	1.79, s	H-1''', H-2'''	1.77, s	H-1''', H-2'''	1.79, s	1.77, s

#Signals overlapping with those of solvent; *Signals overlapping with those of water; $^{a-l}$ signals with same letters overlap with each other;

150

Table S5 ^{1}H NMR and COSY data of the enzyme products of EchPT2 reactions (**2Q1** and **4Q1**)

Pos.	#H	*cyclo*-L-2-*tert*-DMA-4,5,7-tri-DMA-Trp-D-Ala **2Q1** δ_H, multi, *J* in Hz in CD$_3$OD	δ_H, multi, *J* in Hz in CDCl$_3$	COSY	*cyclo*-D-2-*tert*-DMA-4,5,7-tri-DMA-Trp-L-Ala **4Q1** δ_H, multi, *J* in Hz in CD$_3$OD	δ_H, multi, *J* in Hz in CDCl$_3$	COSY
NH-1		-	8.18, s	-	-	8.17, s	-
4		-	-	-	-	-	-
5		-	-	-	-	-	-
6	1H	6.71, s	6.81, s	-	6.71, s	6.81, s	-
7		-	-	-	-	-	H-1'''
10a	1H	3.40, dd, 15.3, 4.0	3.71, dd^a, 14.6, 4.4	H-10b, H-11	§	3.71, dd^d, 14.6, 4.3	H-10b, H-11
10b	1H	3.30#	3.36, dd^b, 14.6, 9.9	H-10a, H-11	3.30#	3.36, dd^e, 14.6, 9.9	H-10a, H-11
11	1H	4.23, dd, 9.7, 4.0	4.42, d, 9.9	H-10a, H-10b	4.23, dd, 9.7, 3.9	4.42, d, 9.9	H-10a, H-10b
NH-12		-	5.57, s	-	-	5.56, s	-
14	1H	4.01, q, 7.0	4.03, qd, 7.1, 2.0	NH-15, H-17	4.02, dq, 7.4, 1.1	4.03, qd, 7.0, 2.0	NH-15, H-17
NH-15		-	5.90, s	H-14	-	5.90, s	H-14
17	3H	1.40, d, 7.0	1.46, d, 7.1	-	1.40, d, 7.0	1.46, d, 7.0	-
18			3.45, dd, 14.8, 5.1 3.38, dd, 14.8, 6.6				
19			4.18, dd, 6.4, 5.2				
1'	1H	5.17, dd, 17.5, 1.1	5.16, dd, 17.5, 1.0	H-2', H-1b'	5.17, dd^c, 17.5, 1.0	5.15, dd, 17.5, 1.0	H-2', H-1b'
	1H	5.10, dd, 10.6, 1.1	5.12, dd, 10.6, 1.0	H-2', H-1a'	5.10, dd, 10.6, 1.0	5.11, dd, 10.6, 1.0	H-2', H-1a'
2'	1H	6.23, dd, 17.5, 10.6	6.10, dd, 17.5, 10.6	H-1'	6.23, dd, 17.5, 10.6	6.10, dd, 17.5, 10.6	H-1'
4'	3H	1.56, s	1.50, s	-	1.57, s	1.49, s	-
5'	3H	1.53, s	1.49, s	-	1.53, s	1.48, s	-
1''	2H	3.30#	3.32, d^b, 7.3	H-2'', H-4''/H-5''	3.30#	3.33, t^e, 7.3	H-2'', H-4''/H-5''
2''	1H	5.19, br. t, 7.3	5.23, tsept, 7.3, 1.4	H-1'', H-4''/H-5''	5.19, br. T^c, 7.3	5.23, tsept, 7.3, 1.4	H-1'', H-4''/H-5''
4''	3H	1.70, s	1.72, s	H-1'', H-2''	1.70, s	1.71, s	H-1'', H-2''
5''	3H	1.71, s	1.71, s	H-1'', H-2''	1.71, s	1.70, s	H-1'', H-2''
1'''	2H	3.54, d, 7.7	3.52, d, 7.4	H-2''', H-4'''/H-5'''	3.54, d, 7.5	3.51, d, 7.4	H-2''', H-4'''/H-5'''
2'''	1H	5.39, tsept, 7.3, 1.4	5.42, tsept, 7.3, 1.4	H-1''', H-4'''/H-5'''	5.39, tsept, 7.3, 1.4	5.42, tsept, 7.3, 1.4	H-1''', H-4'''/H-5'''
4'''	3H	1.80, s	1.88, s	H-1''', H-2'''	1.82, s	1.81, s	H-1''', H-2'''
5'''	3H	1.78, s	1.81, s	H-1''', H-2'''	1.79, s	1.78, s	H-1''', H-2'''
1''''	1H	3.70, d, 5.0	3.71^a	H-2'''', H-4''''/H-5''''	3.70, d, 4.8	3.71^d	H-2'''', H-4''''/H-5''''
	1H	3.58, m,	3.60, dd, 17.5, 5.2	H-2'''', H-4''''/H-5''''	3.60, m,	3.59, dd, 17.4, 5.1	H-2'''', H-4''''/H-5''''
2''''	1H	4.87*	5.03, t, 7.3,	H-1'''', H-4''''/H-5''''	4.87*	5.02, br s	H-1'''', H-4''''/H-5''''
4''''	3H	1.76, s	1.75, s	H-1'''', H-2''''	1.76, s	1.75, s	H-1'''', H-2''''
5''''	3H	1.66, s	1.66, s	H-1'''', H-2''''	1.66, s	1.66, s	H-1'''', H-2''''

#Signals overlapping with those of solvent; *Signals overlapping with those of water; $^{a-e}$ signals with same letters overlap with each other; §Due to low quality, signals were not detected

Table S6 ¹H NMR and COSY data of the enzyme products of EchPT2 reactions (**6Q1** and **8Q1**)

Position	int.	6Q1 δ_H, multi *J* in Hz	6Q1 δ_H, multi *J* in Hz in CDCl₃	6Q1 COSY	8Q1 δ_H, multi *J* in Hz in CD₃OD	8Q1 δ_H, multi *J* in Hz in CDCl₃	8Q1 COSY
NH-1		-	8.11, s	-	-	8.11, s	-
4		-	-	-	-	-	-
5		-	-	-	-	-	-
6	1	6.70, s	6.80, s	H-1'''	6.70, s	6.80, s	H-1'''
7		-	-	-	-	-	-
10a	1	3.44, mᵃ	3.55, dd^c, 14.8, 5.1	H-10b, H-11	3.40, m^g	3.57, dd^k, 14.8, 5.1	H-10b, H-11
10b	1	3.30#	3.32, dd, 14.8, 6.7	H-10a, H-11	3.30#	3.32, dd^h, 14.8, 6.5	H-10a, H-11
11	1	4.15, m	4.27, dt, 11.3, 3.3	NH-12, H-10a, H-	4.11, m	4.27, dt, 11.6, 5.1	NH-12, H-10a, H-10b
NH-12		-	5.58, s	H-11	-	5.58, s	H-11
14	1	3.30#	4.03, dd, 10.1, 6.7	H-19	3.30#	4.03, dd, 9.9, 6.6	H-19
17	1	3.44, mᵃ	3.70, m^d	H-18	3.40, m^g	3.71, m^i	H-18
	1	3.30#	3.54, m^c	H-18	3.30#	3.57, m^k	H-18
18	1	2.00, m	2.04, m^e	H-17	2.00, m	2.04, m^j	H-19, H-17
	1	§	1.91, m^e	H-19, H-17	§	1.92, m^j	H-19, H-17
19	1	2.30, m	2.40, m	H-14	2.31, m	2.41, m	H-18, H14
	1	§	1.98, m^e	H-18, H14	§	1.96, m^j	H-18, H14
1'	1	5.18, dd^b, 17.5, 1.2	5.16, dd, 17.5, 1.1	H-2', H-1b'	5.19, dd^f, 17.5, 1.1	5.16, dd, 17.5, 1.1	H-2', H-1b'
	1	5.12, dd, 10.6, 1.2	5.10, dd, 10.5, 1.1	H-2',	5.12, dd, 10.6, 1.1	5.10, dd, 10.5, 1.1	H-2', H-1a'
2'	1	6.25, dd, 17.5, 10.6	6.13, dd, 17.5, 10.5	H-1'	6.25, dd, 17.5, 10.6	6.13, dd, 17.5, 10.5	H-1'
4'	3	1.56, s	1.49, s				
5'	3	1.52, s	1.47, s				
1''	2	3.30#	3.32, d, 6.9	H-2'', H-4''/H-5''	3.30#	3.34, d^h, 7.5	H-2'', H-4''/H-5''
2''	1	5.20, tsept^b, 7.3, 1.4	5.23, tsept, 7.3, 1.4	H-1'', H-4''/H-5''	5.20, tsept^f, 7.3, 1.4	5.24, tsept, 7.3, 1.4	H-1'', H-4'''/H-5'''
4''	3	1.72, s	1.72, s	H-1'', H-2''	1.72, s	1.72, s	H-1'', H-2''
5''	3	1.70, s	1.70, s	H-1'', H-2''	1.70, s	1.69, s	H-1'', H-2''
1'''	2	3.52, d, 7.3	3.50, d, 7.3	H-2''', H-4'''/H-5'''	3.52, d, 7.3	3.5, d, 7.5	H-2''', H-4'''/H-5'''
2'''	1	5.37, tsept, 7.3, 1.4	5.41, tsept, 7.3, 1.4	H-1''', H-4'''/H-5'''	5.37, tsept, 7.3, 1.4	5.42, tsept, 7.3, 1.4	H-1''', H-4'''/H-5'''
4'''	3	1.80, s	1.87, s	H-1''', H-2'''	1.80, s	1.86, s	H-1''', H-2'''
5'''	3	1.79, s	1.79, s	H-1''', H-2'''	1.79, s	1.79, s	H-1''', H-2'''
1''''	1	3.70, dd, 17.2, 4.9	3.78, dd, 17.5, 5.1	H-2'''', H-4''''/H-5''''	3.70, dd, 17.4, 4.9	3.79, dd, 17.6, 5.1	H-2'''', H-4''''/H-5''''
	1	§	3.65^d	H-2'''', H-4''''/H-5''''	§	3.65^l	H-2'''', H-4''''/H-5''''
2''''	1	4.87*	4.97, br s,	H-1'''', H-4''''/H-5''''	4.87*	4.97, br s	H-1'''', H-4''''/H-5''''
4''''	3	1.78, s	1.81, s	H-1'''', H-2''''	1.78, s	1.81, s	H-1'''', H-2''''
5''''	3	1.65, s	1.64, s	H-1'''', H-2''''	1.65, s	1.64, s	H-1'''', H-2''''

cyclo-L-2-*tert*-DMA-4,5,7-tri-DMA-Trp-D-Pro **6Q1**

cyclo-D-2-*tert*-DMA-4,5,7-tri-DMA-Trp-L-Pro **8Q1**

#Signals overlapping with those of solvent; *Signals overlapping with those of water; §Due to low quality, signal were not detected; a-k signals with same letters overlap with each other; *Signals overlapping with each other

3 KINTETIC PARAMETERS

Figure S1 Determination of the kinetic parameters of EchPT1 toward *cyclo*-ʟ-Trp-ᴅ-Ala (**2**).

The reaction mixtures contained 5 mM of CaCl$_2$, 1 mM of DMAPP and 2 µg of the purified recombinant. The mixtures were incubated at 37 °C and pH 7.5 for 5 min.

Figure S2 Determination of the kinetic parameters of EchPT1 toward *cyclo*-ᴅ-Trp-ᴅ-Ala (**3**).

The reaction mixtures contained 5 mM of CaCl$_2$, 1 mM of DMAPP and 0.25 µg of the purified recombinant protein. The mixtures were incubated at 37 °C and pH 7.5 for 30 min.

Figure S3 Determination of the kinetic parameters of EchPT1 toward *cyclo*-D-Trp-L-Ala (**4**).

The reaction mixtures contained 5 mM of $CaCl_2$, 1 mM of DMAPP and 2 µg of the purified recombinant protein and were incubated at 37 °C and pH 7.5 for 5 min.

Figure S4 Determination of the kinetic parameters of EchPT1 toward cyclo-L-Trp-L-Pro (**5**).

The reaction mixtures contained 5 mM of $CaCl_2$, 1 mM of DMAPP and 2 µg of the purified recombinant protein-The mixtures were incubated at 37 °C and pH 7.5 for 1 min.

S12

154

Figure S5 Determination of the kinetic parameters of EchPT1 toward *cyclo*-L-Trp-D-Pro (**6**).

The reaction mixtures contained 5 mM of $CaCl_2$, 1 mM of DMAPP and 2 µg of the purified recombinant protein. The mixtures were incubated at 37 °C and pH 7.5 for 5 min.

Figure S6 Determination of the kinetic parameters of EchPT1 toward cyclo-D-Trp-D-Pro (**7**).

The reaction mixtures contained 5 mM of $CaCl_2$, 1 mM of DMAPP and 0.25 µg of the purified recombinant protein. The mixtures were incubated at 37 °C and pH 7.5 for 30 min.
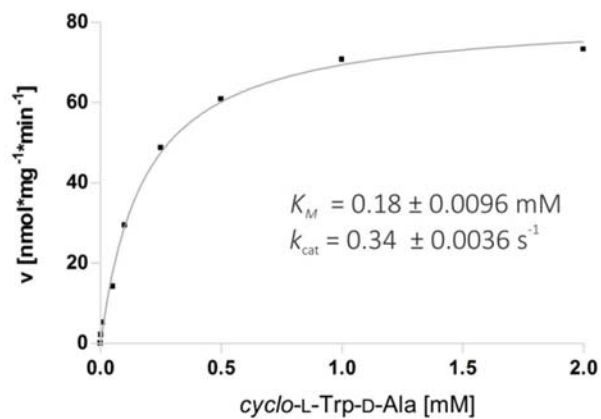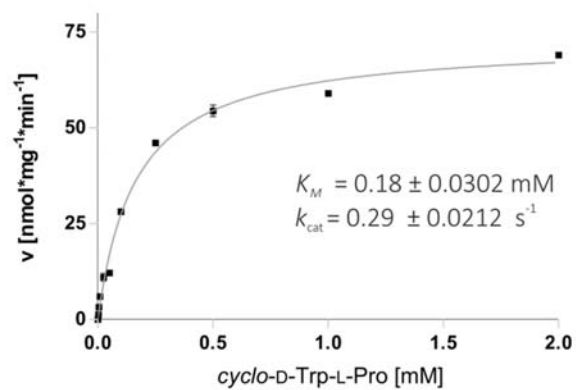
Figure S7 Determination of the kinetic parameters of EchPT1 toward cyclo-D-Trp-L-Pro (**8**)

The reaction mixtures contained 5 mM of $CaCl_2$, 1 mM of DMAPP and 0.05 µg of the purified recombinant protein. The mixtures were incubated at 37 °C and pH 7.5 for 30 min.
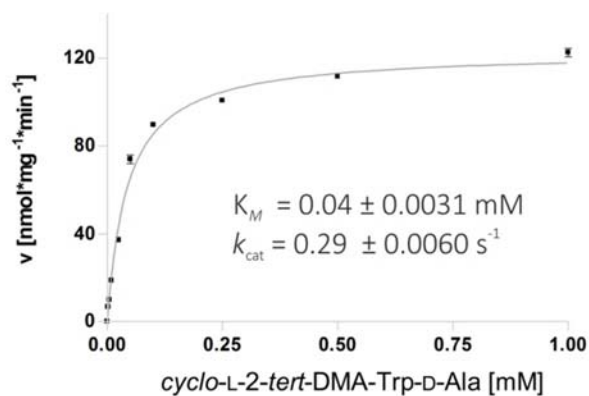
Figure S8 Determination of the kinetic parameters of the EchPT2 reaction toward *cyclo*-L-2-*tert*-DMA-Trp-D-Ala (**2M1**).

The reaction mixtures contained 5 mM of $CaCl_2$, 3 mM of DMAPP and 2µg of the purified recombinant protein and were incubated at 37 °C and pH 7.5 for 8 min.

S14
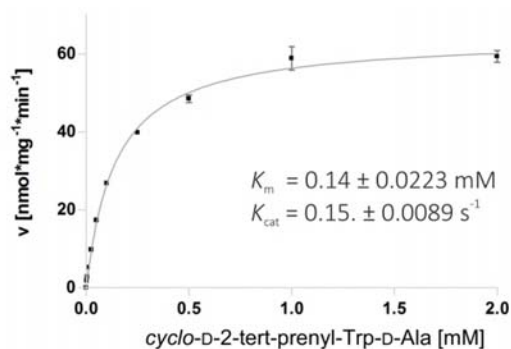
Figure S9 Determination of the kinetic parameters of the EchPT2 reaction toward *cyclo*-D-2-*tert*-DMA-Trp-D-Ala (**3M1**).

The reaction mixtures contained 5 mM of $CaCl_2$, 3 mM of DMAPP and 2 µg of the purified recombinant protein. The mixtures were incubated at 37 °C and pH 7.5 for 10 min (**3M1**).
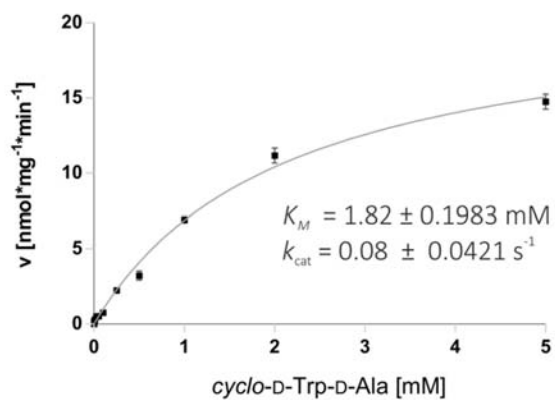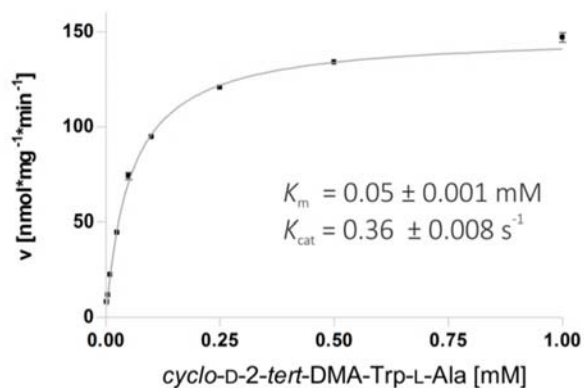

Figure S10 Determination of the kinetic parameters of the EchPT2 reaction toward *cyclo*-D-2-*tert*-DMA-Trp-L-Ala (**4M1**).

The reaction mixtures contained 5 mM of $CaCl_2$, 3 mM of DMAPP and 2µg of the purified recombinant protein and were incubated at 37 °C and pH 7.5 for 8 min.
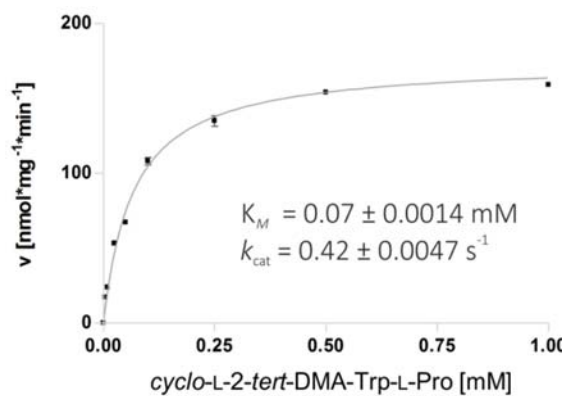
Figure S11 Determination of the kinetic parameters of the EchPT2 reaction toward *cyclo*-L-2-*tert*-DMA-Trp-L-Pro (**5M1**).

The reaction mixtures contained 5mM of $CaCl_2$, 3 mM of DMAPP and 2 µg of the purified recombinant protein. The mixtures were incubated at 37 °C and pH 7.5 for 5 min.
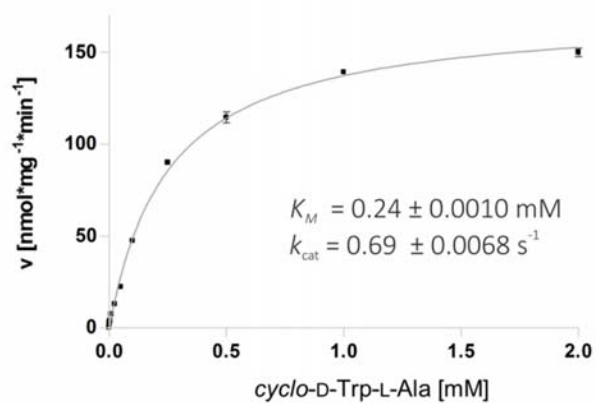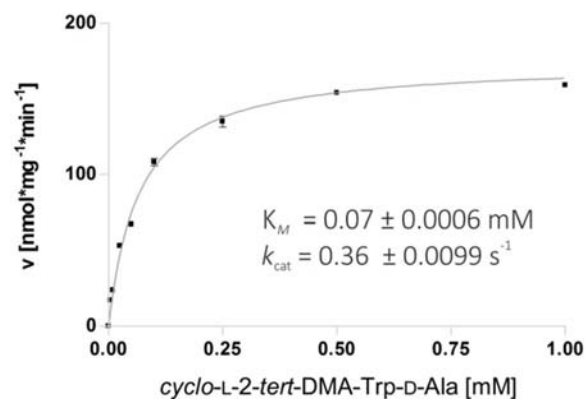

Figure S12 Determination of the kinetic parameters of the EchPT2 reaction toward *cyclo*-L-2-*tert*-DMA-Trp-D-Pro (**6M1**).

The reaction mixtures contained 5 mM of $CaCl_2$, 3 mM of DMAPP and 2 µg of the purified recombinant protein and were incubated at 37 °C and pH 7.5 for 10 min.
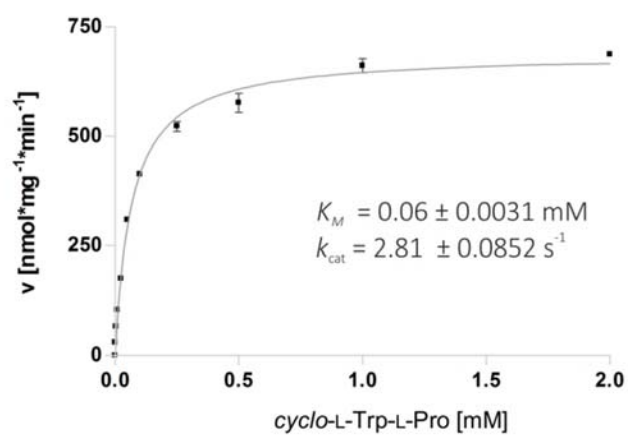
S16
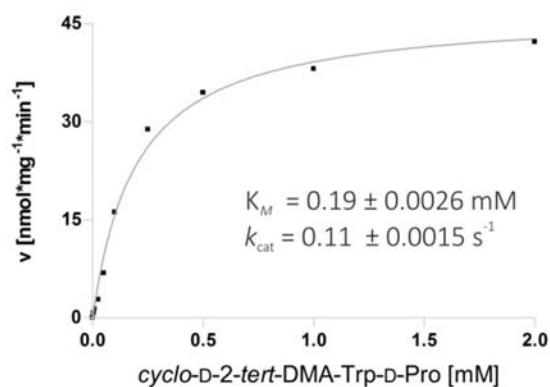
Figure S13 Determination of the kinetic parameters of the EchPT2 reaction toward *cyclo*-D-2-*tert*-DMA-Trp-D-Pro (**7M1**)

The reaction mixtures contained 5mM of $CaCl_2$, 3 mM of DMAPP and 0.25 µg of the purified recombinant protein. The mixtures were incubated at 37 °C and pH 7.5 for 30 min (**7M1**).
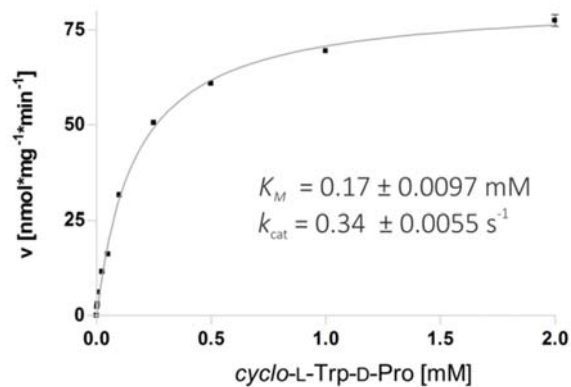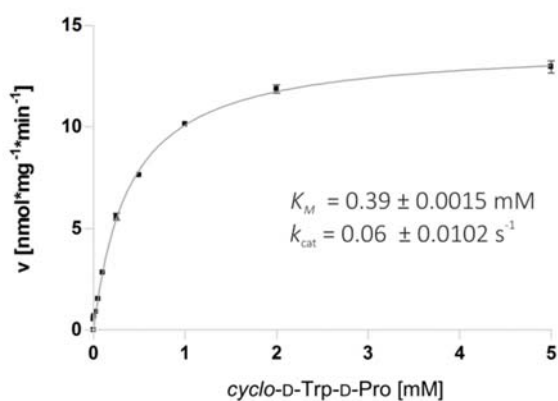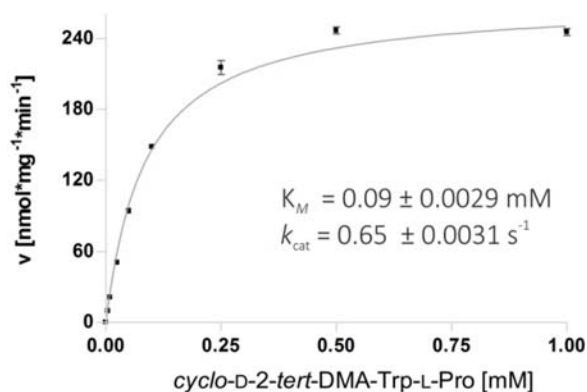

Figure S14 Determination of the kinetic parameters of the EchPT2 reaction toward cyclo-D-2-*tert*-DMA-Trp-L-Pro (**8M1**).

The reaction mixtures contained 5 mM of $CaCl_2$, 3 mM of DMAPP and 2 µg of the purified recombinant protein and were incubated at 37 °C and pH 7.5 for 10 min

## 4   FIGURES OF NMR SPECTRA

Figure S15 ^{1}H NMR spectrum of *cyclo*-L-2-*tert*-DMA-Trp-D-Ala (**2M1**) in CD$_3$OD (500MHz)
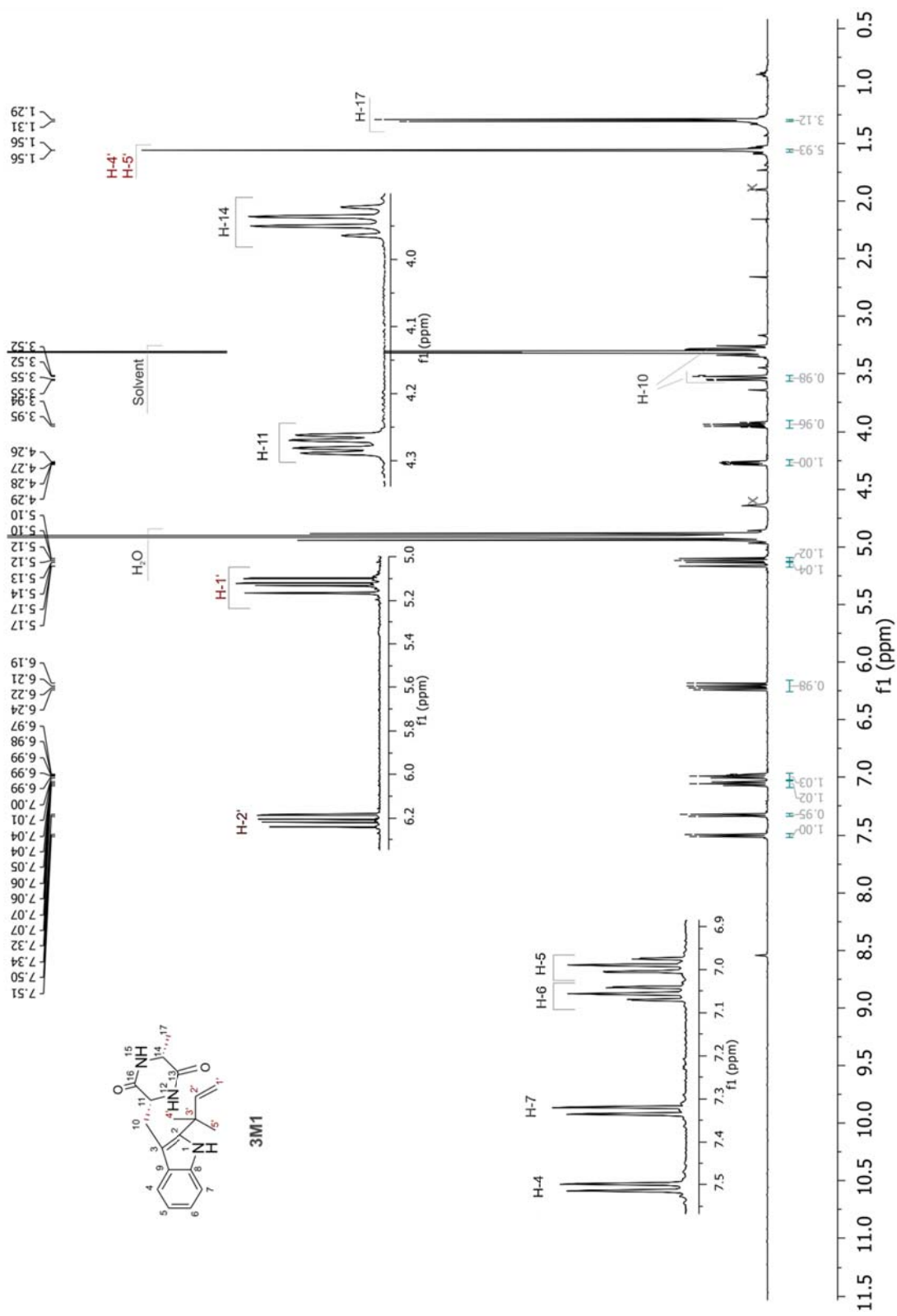
Figure S16 ¹H NMR spectrum of *cyclo*-D-2-*tert*-DMA-Trp-D-Ala (**3M1**) in CD₃OD (500MHz)
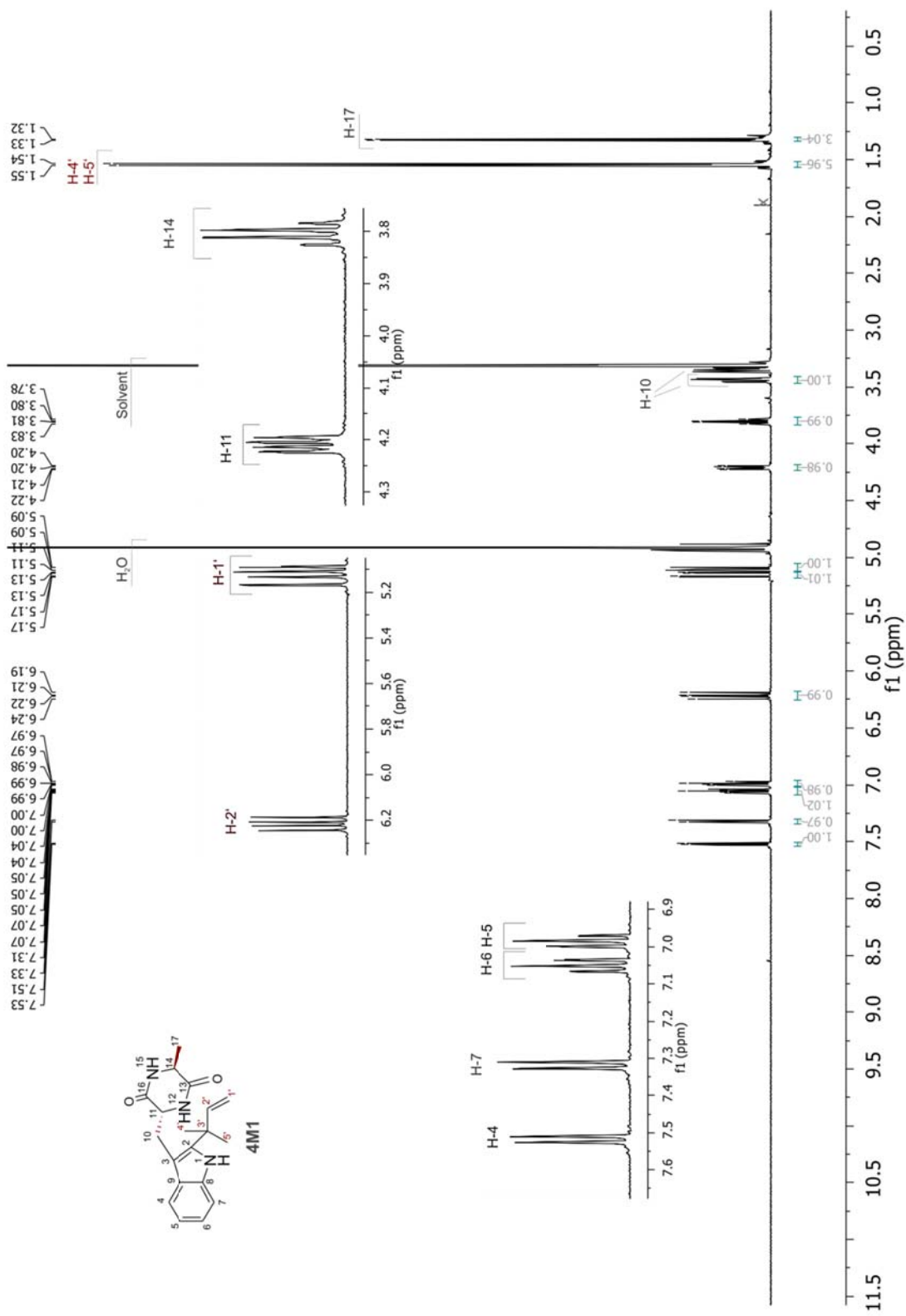
Figure S17 ^{1}H NMR spectrum of *cyclo*-ᴅ-2-*tert*-DMA-Trp-ʟ-Ala (**4M1**) in CD$_3$OD (500MHz)
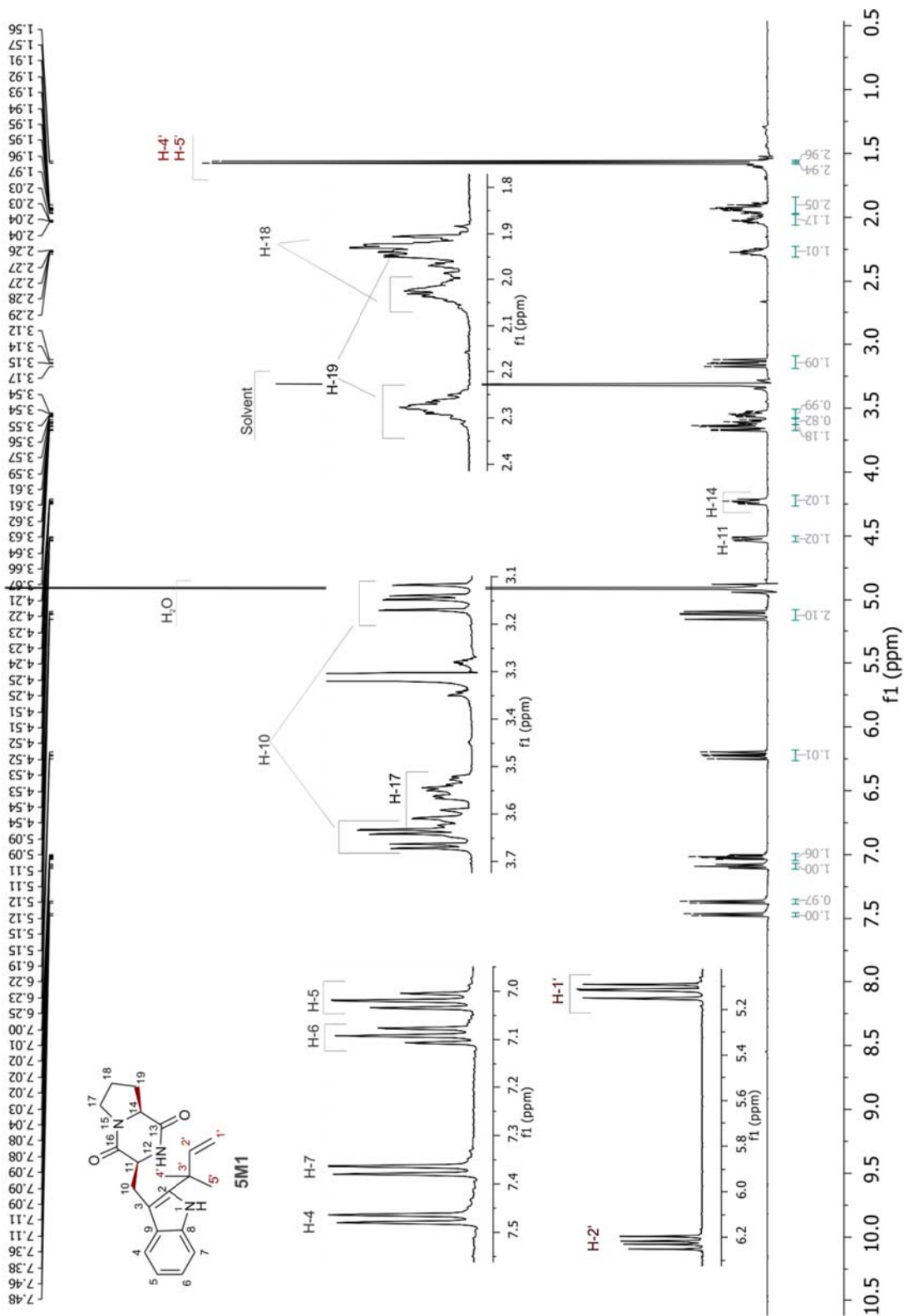
Figure S18 ^{1}H NMR spectrum of *cyclo*-L-2-*tert*-DMA-Trp-L-Pro (**5M1**) in CD$_3$OD (500MHz)

Figure S19 ^{1}H NMR spectrum of *cyclo*-L-2-*tert*-DMA-Trp-D-Pro (**6M1**) in CD$_3$OD (500MHz)
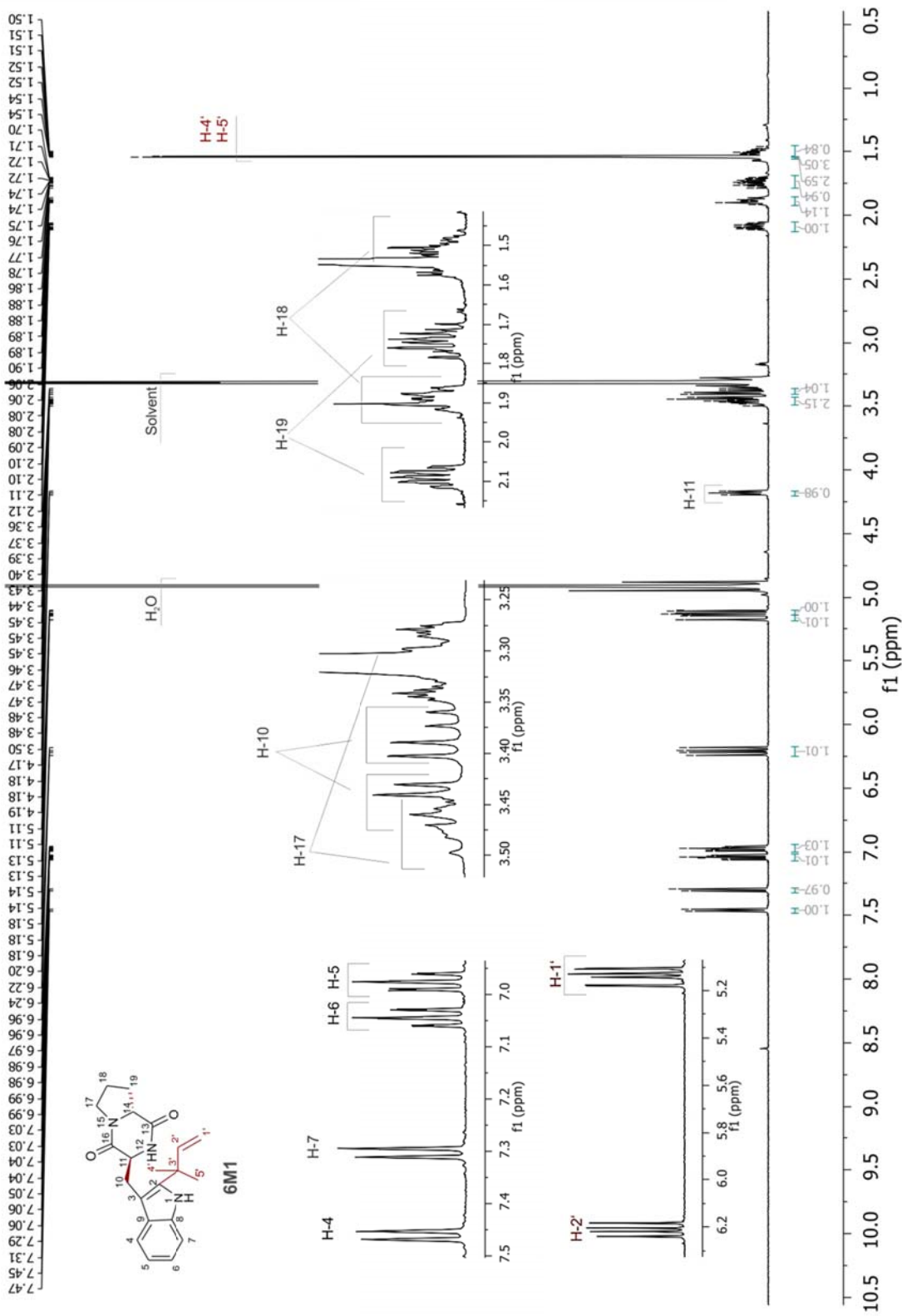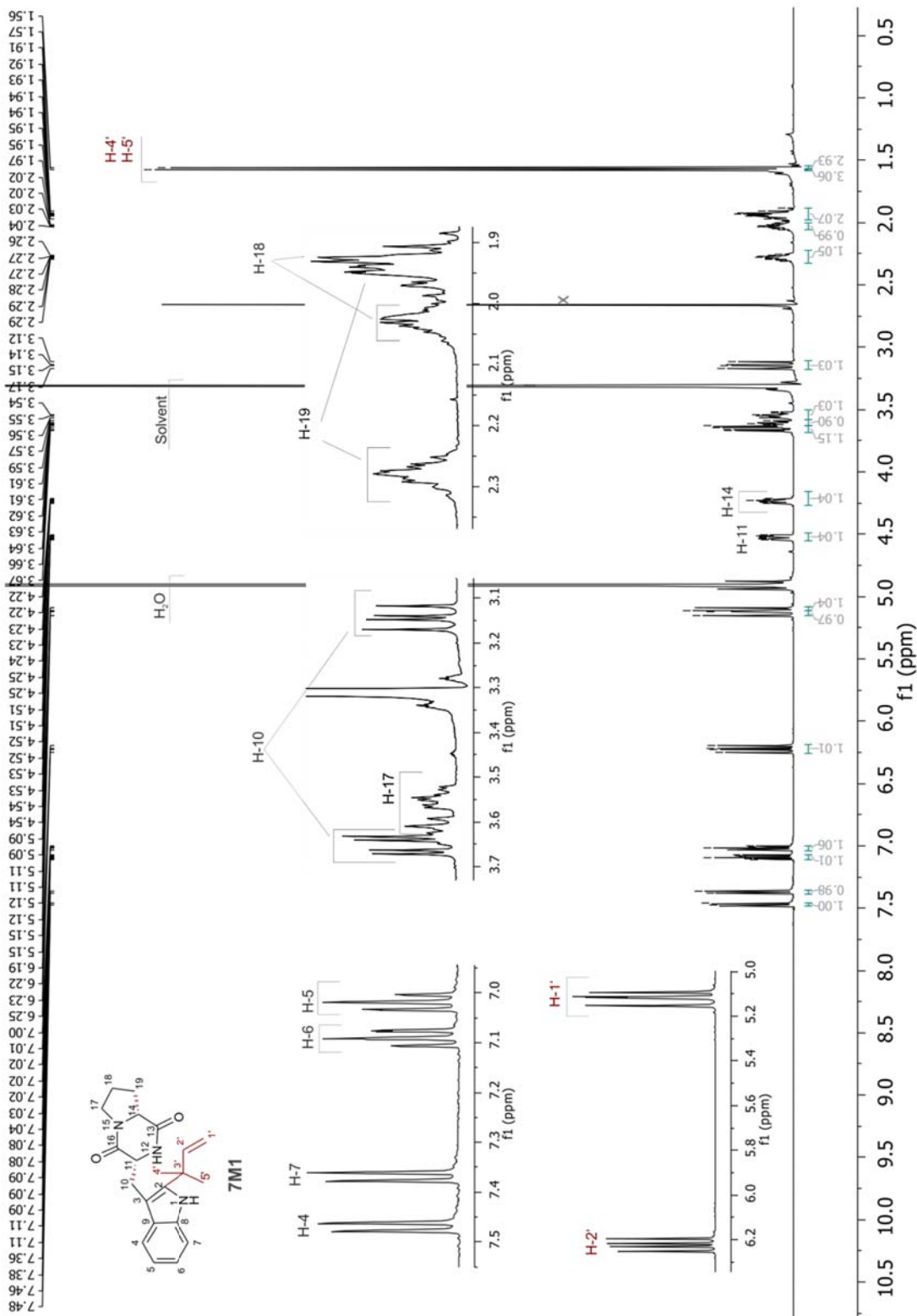
Figure S20 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-Trp-D-Pro (**7M1**) in CD$_3$OD (500MHz)

165

S23
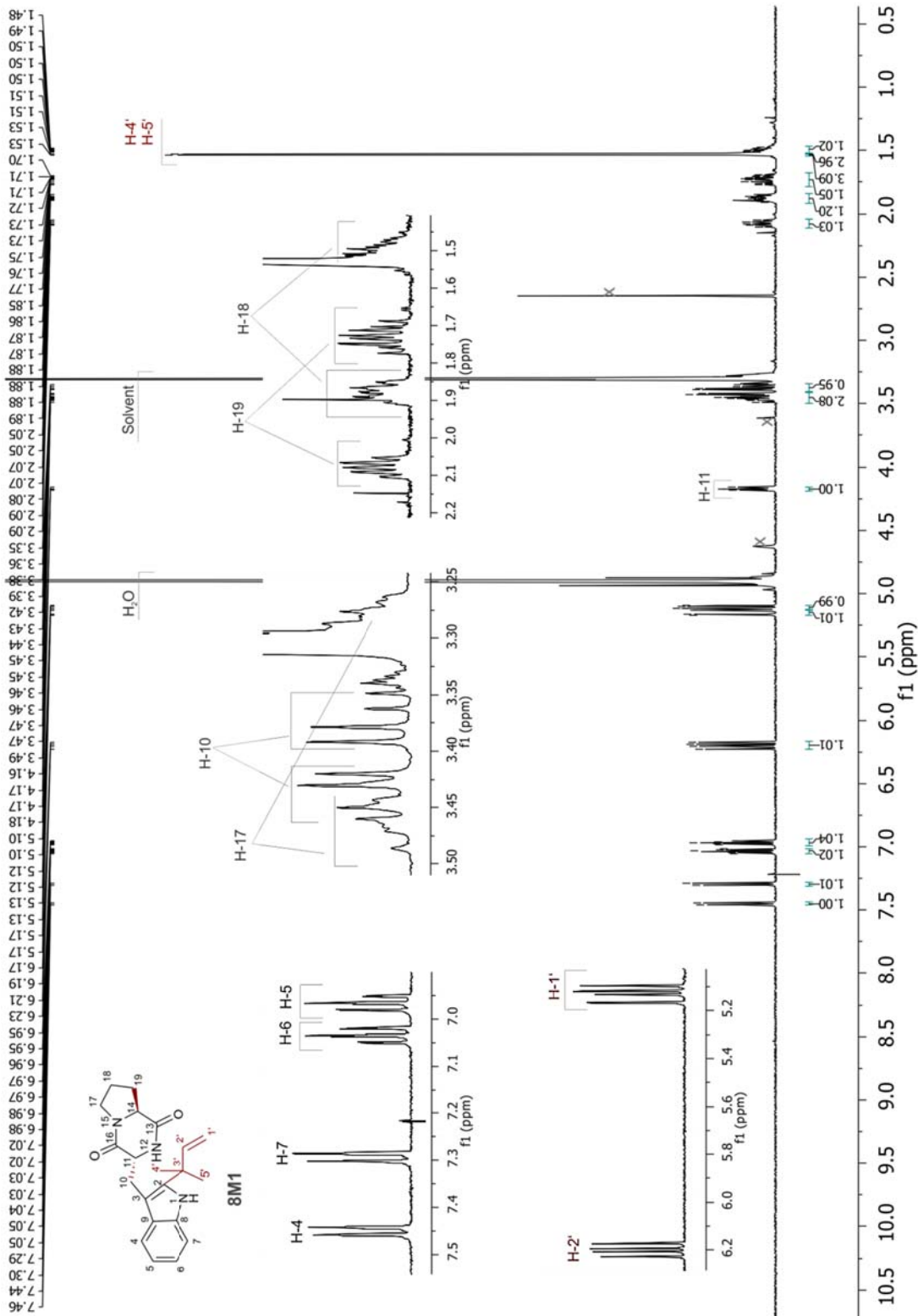
Figure S21 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-Trp-L-Pro (**8M1**) in CD$_3$OD (500MHz)
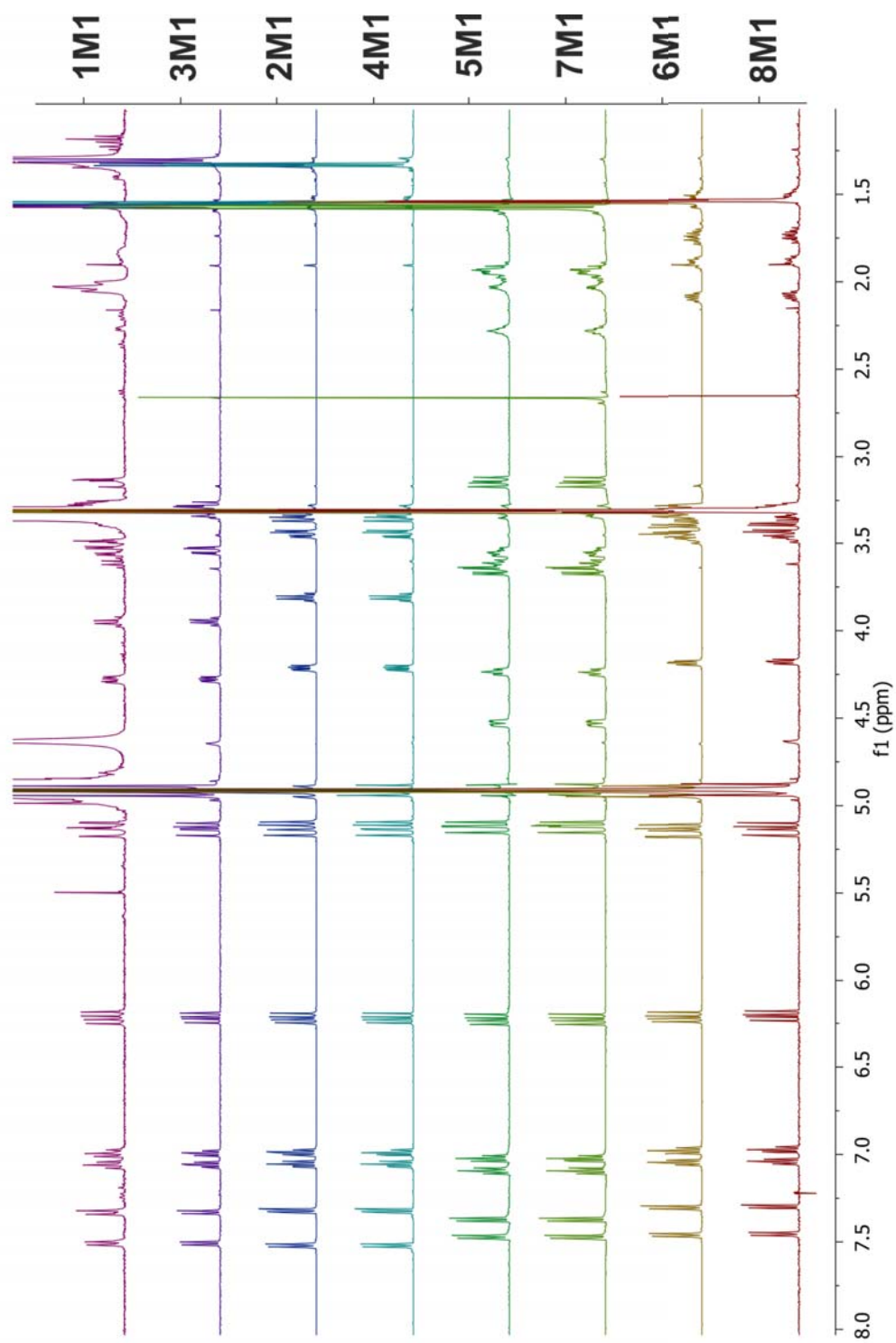
Figure S22 Comparison of the ^{1}H NMR spectra of *cyclo*-L-2-tert-DMA-Trp-L-Ala (**1M1**) (Wohlgemuth et al. 2017), *cyclo*-L-2-tert-DMA-Trp-D-Ala (**2M1**), *cyclo*-D-2-tert-DMA-Trp-D-Ala (**3M1**), *cyclo*-D-2-tert-DMA-Trp-L-Ala (**4M1**), *cyclo*-L-2-tert-DMA-Trp-L-Pro (**5M1**), *cyclo*-D-2-tert-DMA-Trp-L-Pro (**6M1**), *cyclo*-D-2-tert-DMA-Trp-D-Pro (**7M1**) and *cyclo*-D-2-tert-DMA-Trp-L-Pro (**8M1**) in CD₃OD

167
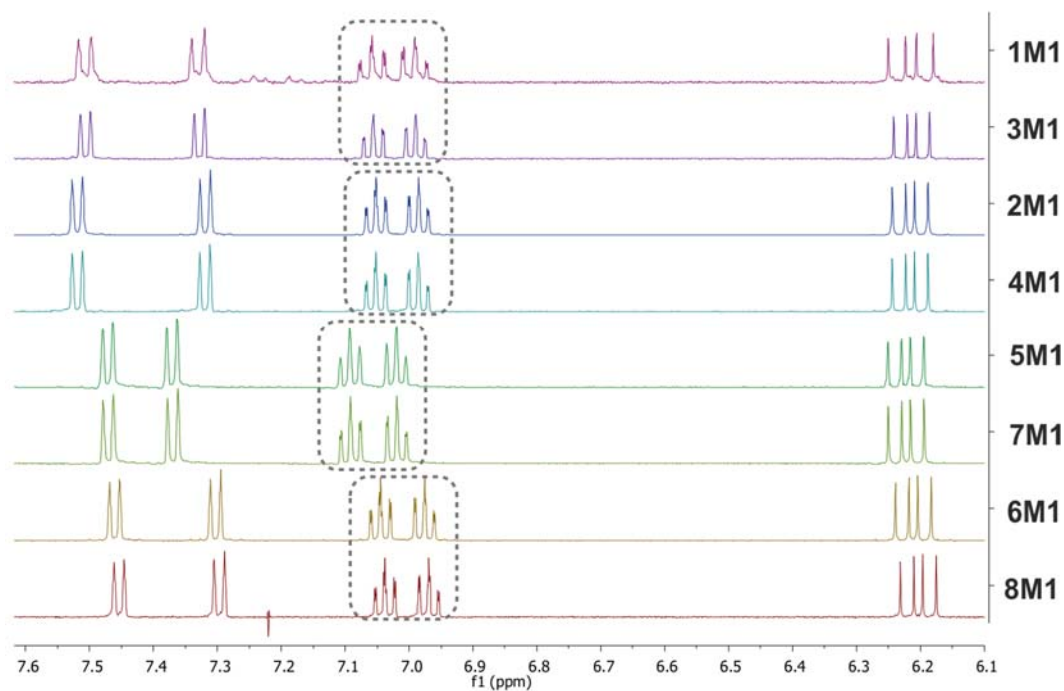
Figure S23 selected region (6.1 – 7.6 ppm) of the ^{1}H NMR spectra of **1M1** (Wohlgemuth et al. 2017) and **2M1 – 8M1** in CD$_3$OD
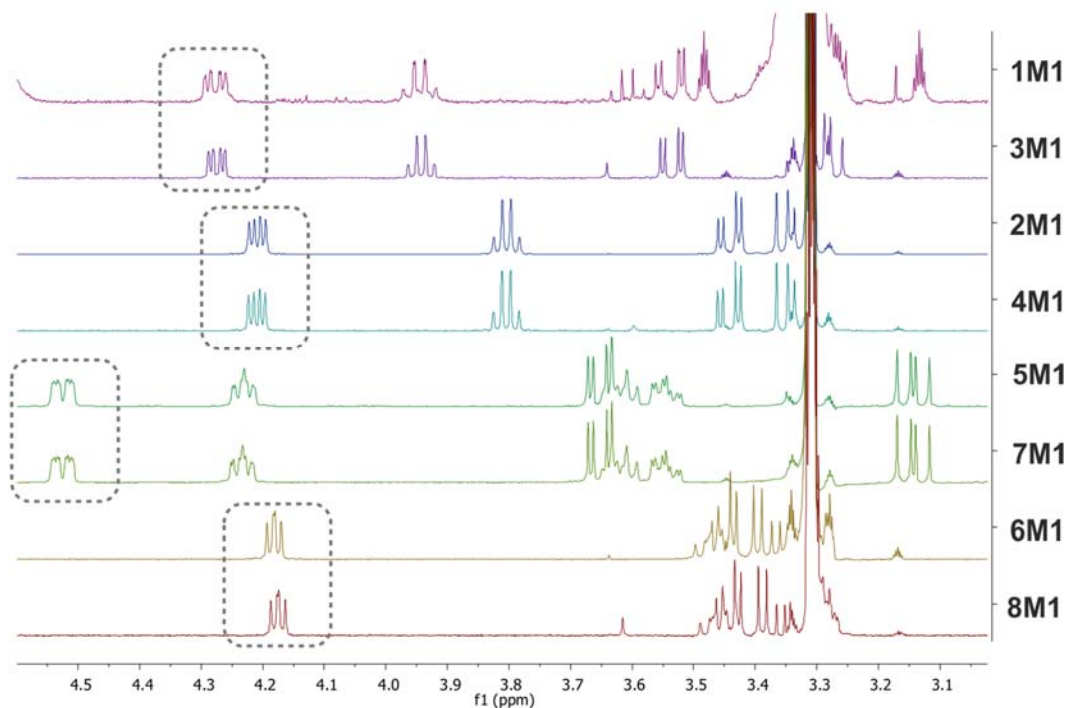
Figure S24 selected region (3.1 – 4.5 ppm) of the ^{1}H NMR spectra of **1M1** (Wohlgemuth et al. 2017) and **2M1 – 8M1** in CD$_3$OD

Figure S25 ^{1}H NMR spectrum of echinulin (**1T2**) in CD$_3$OD (500MHz)
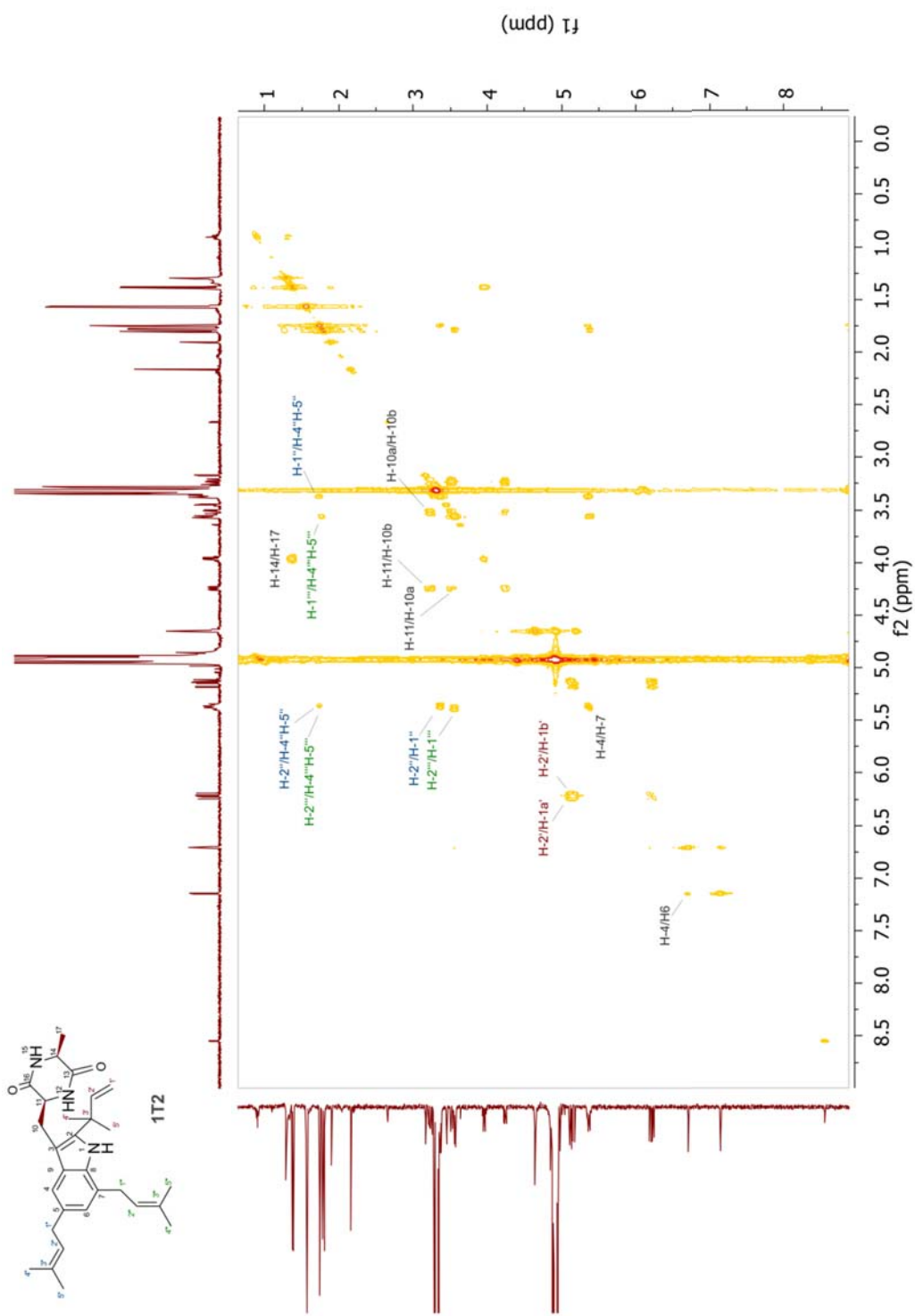
Figure S26 ^{1}H-^{1}H COSY spectrum of echinulin (**1T2**) in CD$_3$OD (500MHz)
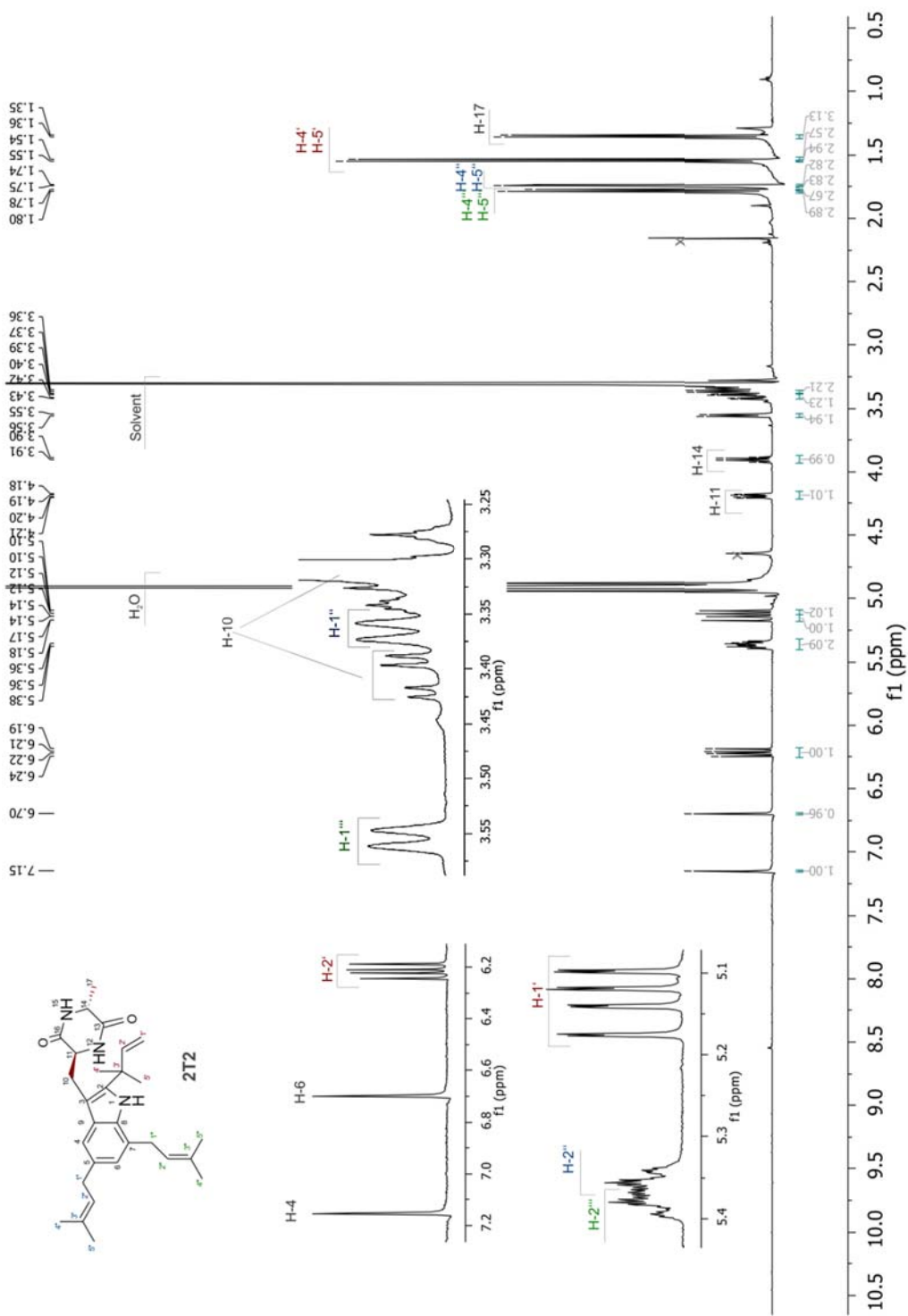
Figure S27 ¹H NMR spectrum of *cyclo*-L-2-*tert*-DMA-5,7-di-DMA-Trp-D-Ala (**2T2**) in CD₃OD (500MHz)

171
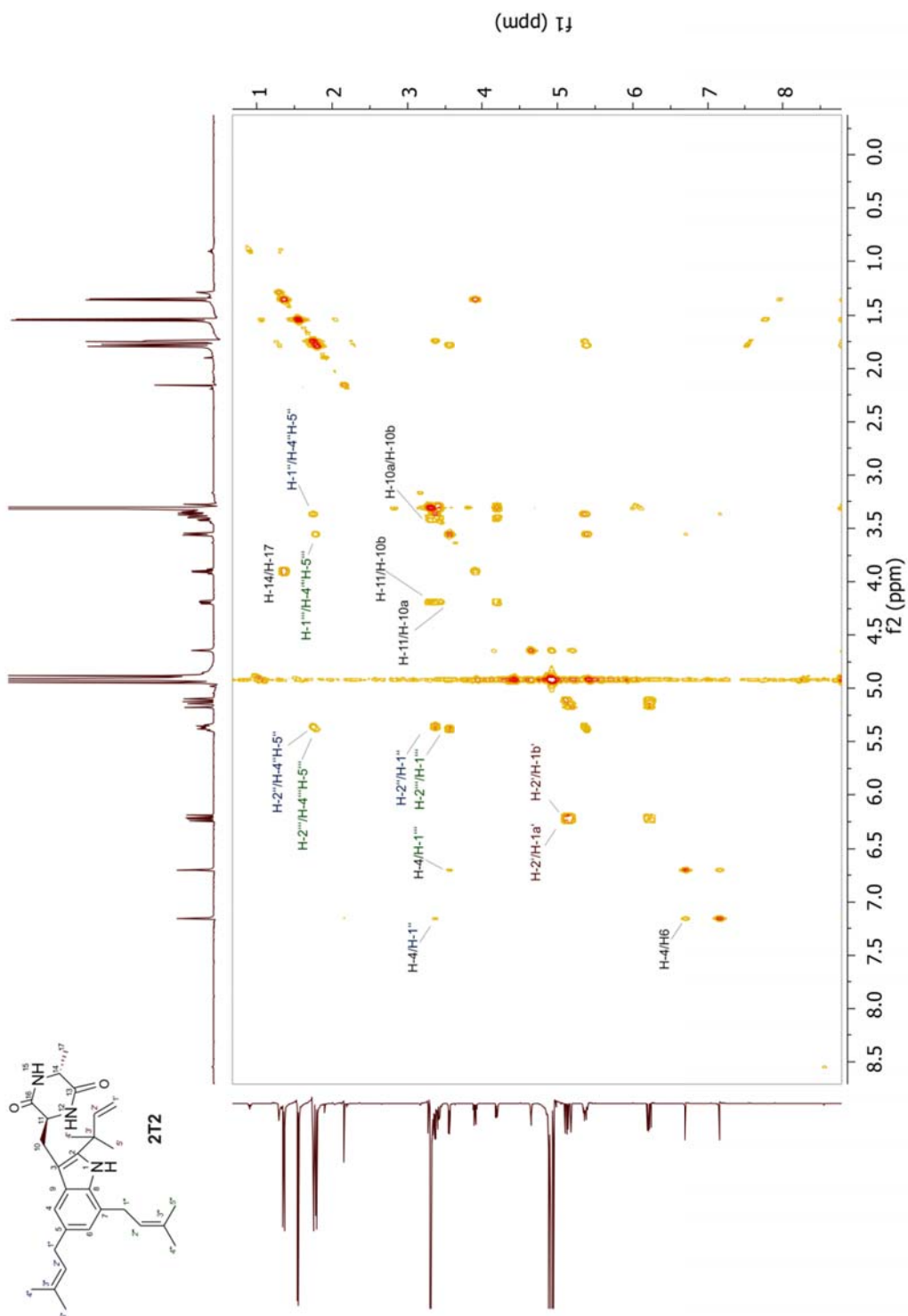
Figure S28 ^{1}H-^{1}H COSY spectrum of (**2T2**) in CD$_3$OD (500MHz)
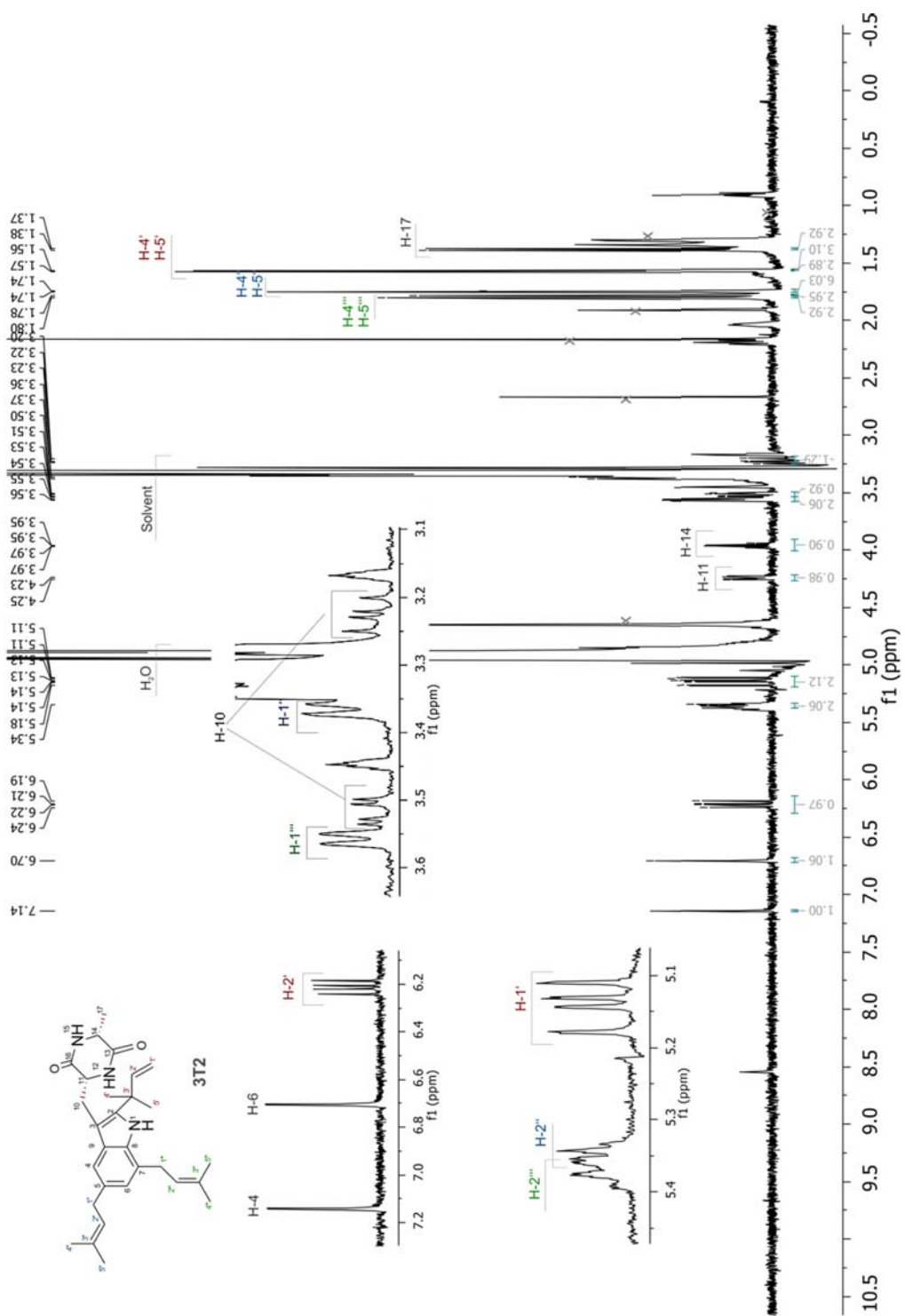
Figure S29 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-5,7-di-DMA-Trp-D-Ala (**3T2**) in CD_3OD (500MHz)

173

Figure S30 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-5,7-di-DMA-Trp-L-Ala (**4T2**) in CD$_3$OD (500MHz)
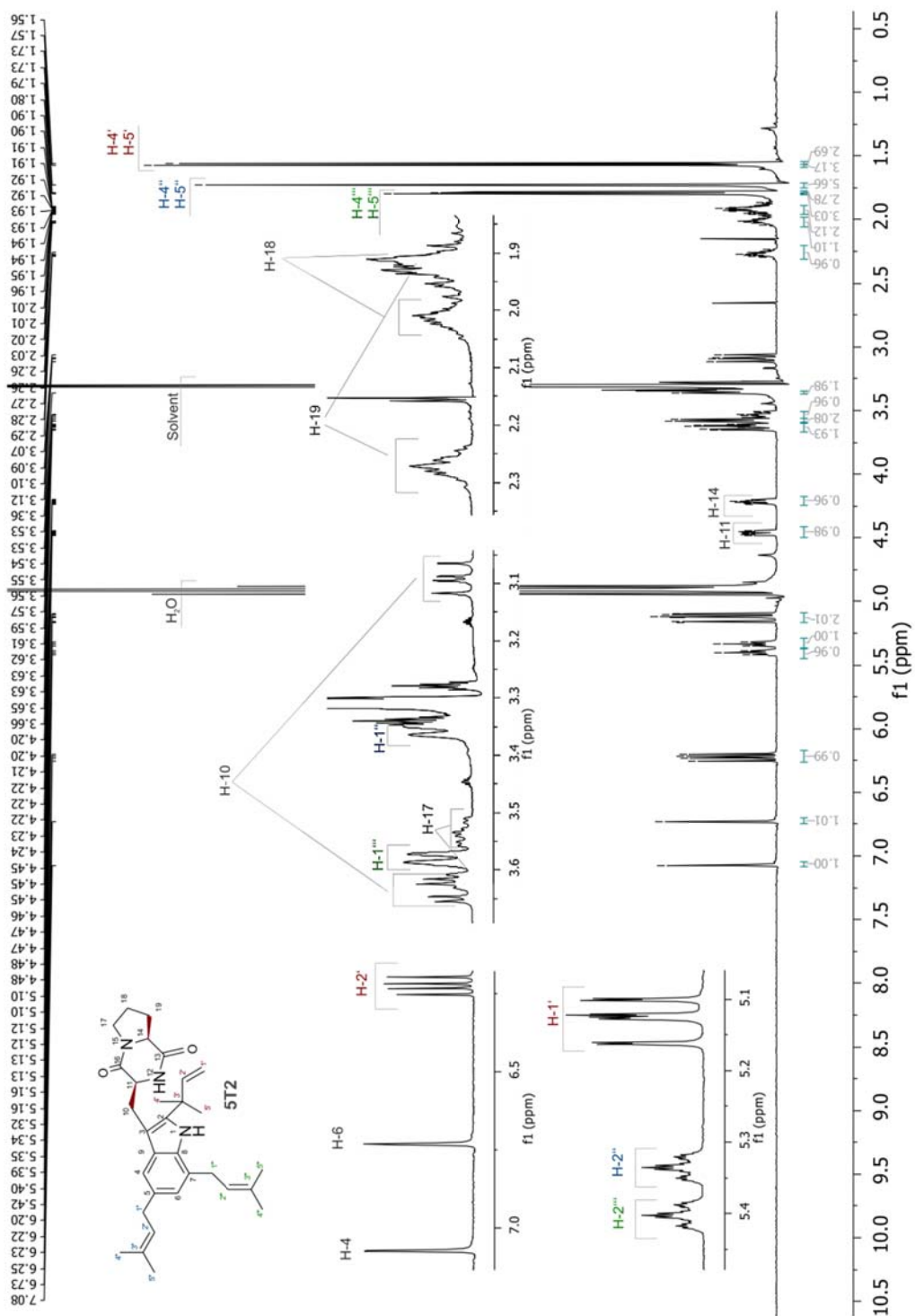
Figure S31 ^{1}H NMR spectrum of *cyclo*-L-2-*tert*-DMA-5,7-di-DMA-Trp-L-Pro (**5T2**) in CD$_3$OD (500MHz)
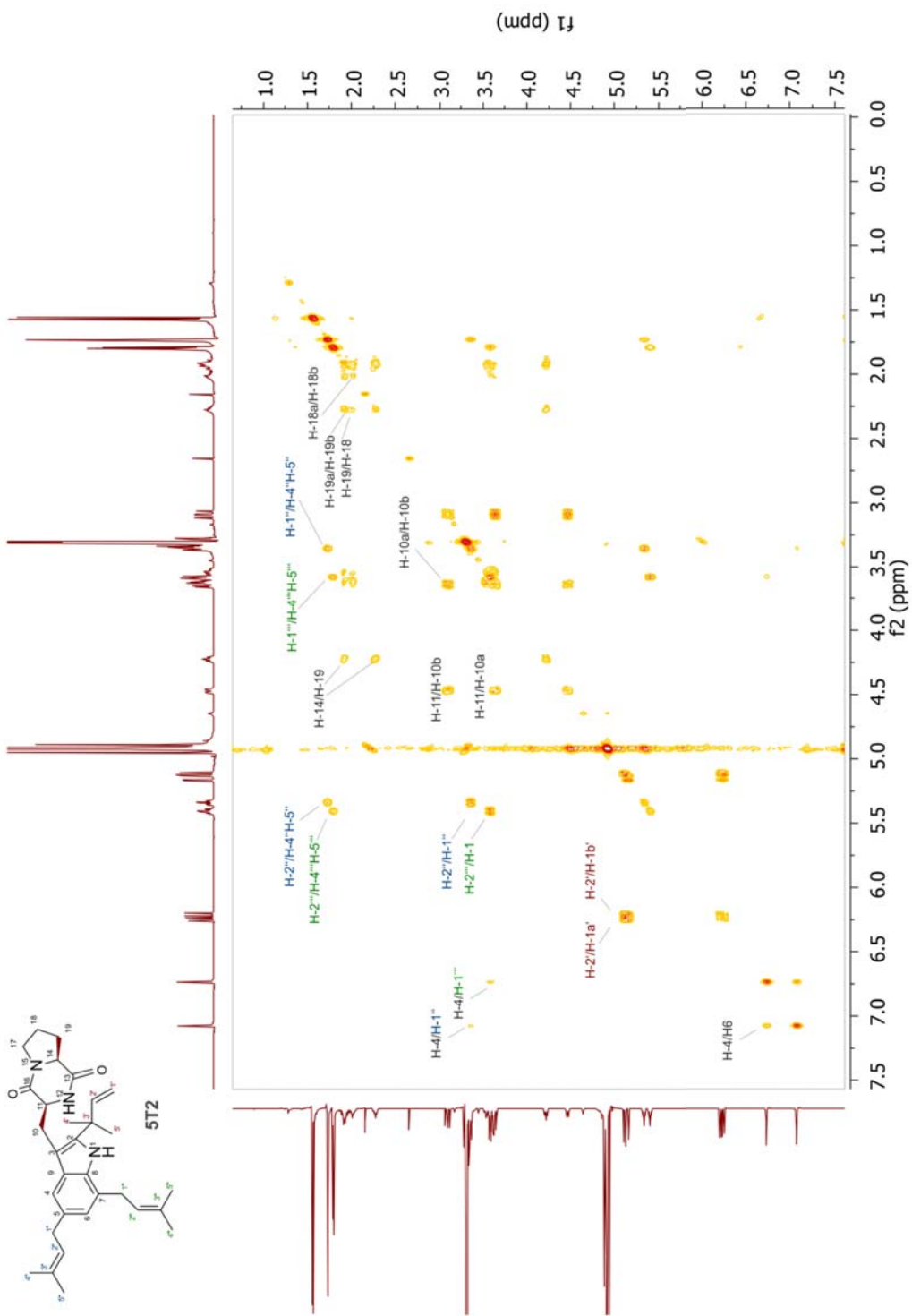
Figure S32 ^{1}H-^{1}H COSY spectrum of (**5T2**) in CD$_3$OD (500MHz)
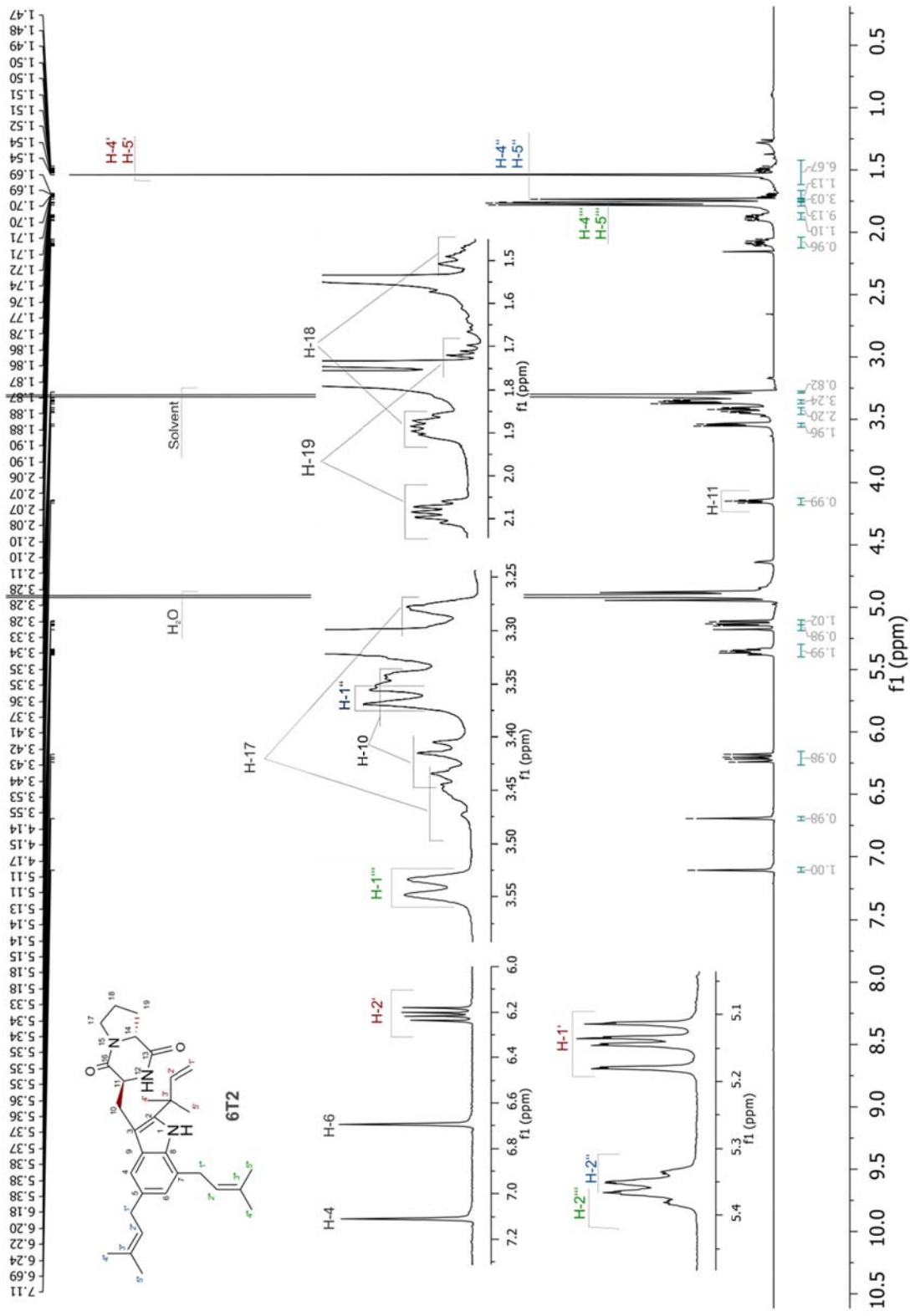
Figure S33 ^{1}H NMR spectrum of *cyclo*-L-2-*tert*-DMA-5,7-di-DMA-Trp-D-Pro (**6T2**) in CD$_3$OD (500MHz)
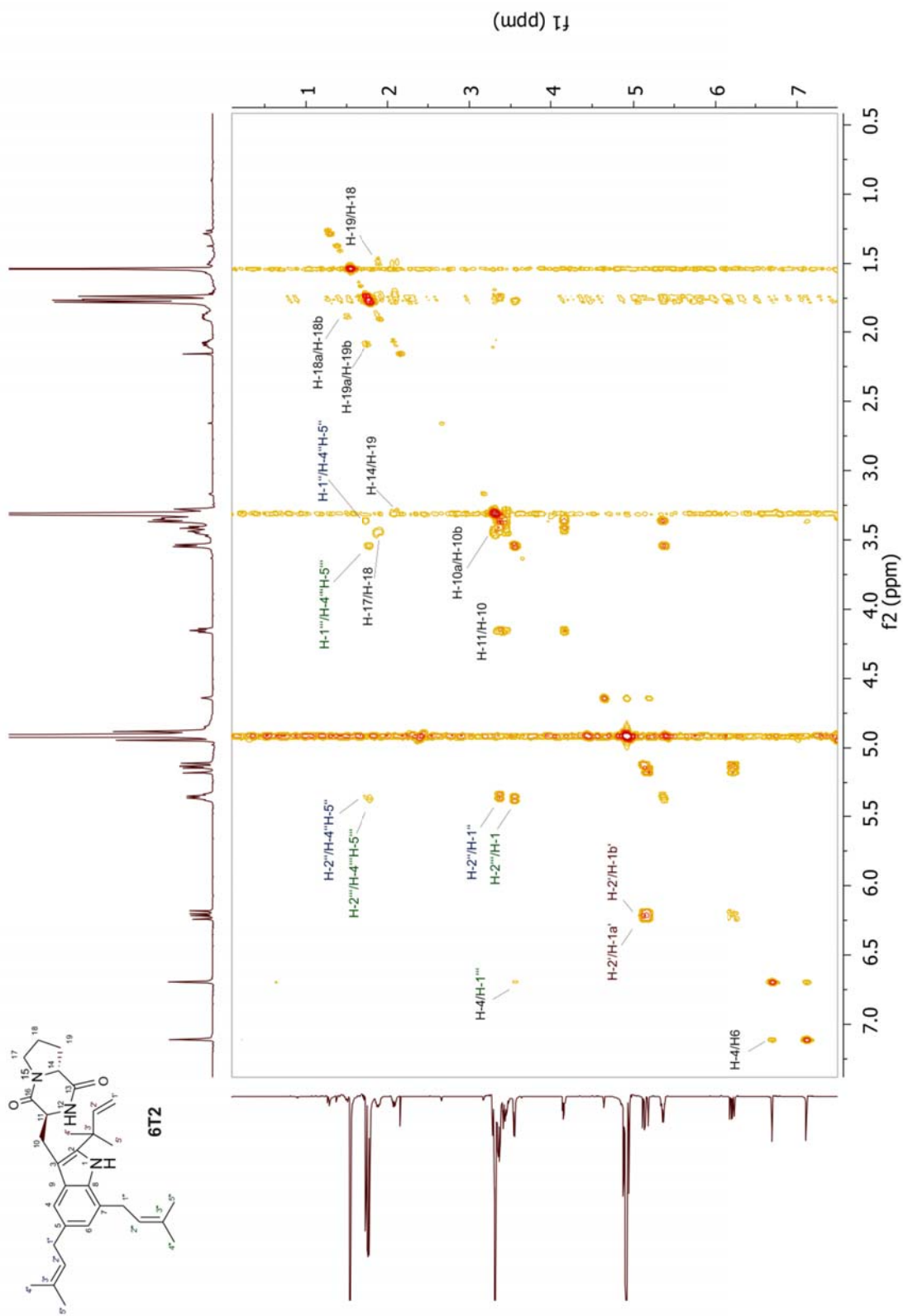
Figure S34 ^{1}H-^{1}H COSY spectrum of (**6T2**) in CD$_3$OD (500MHz)
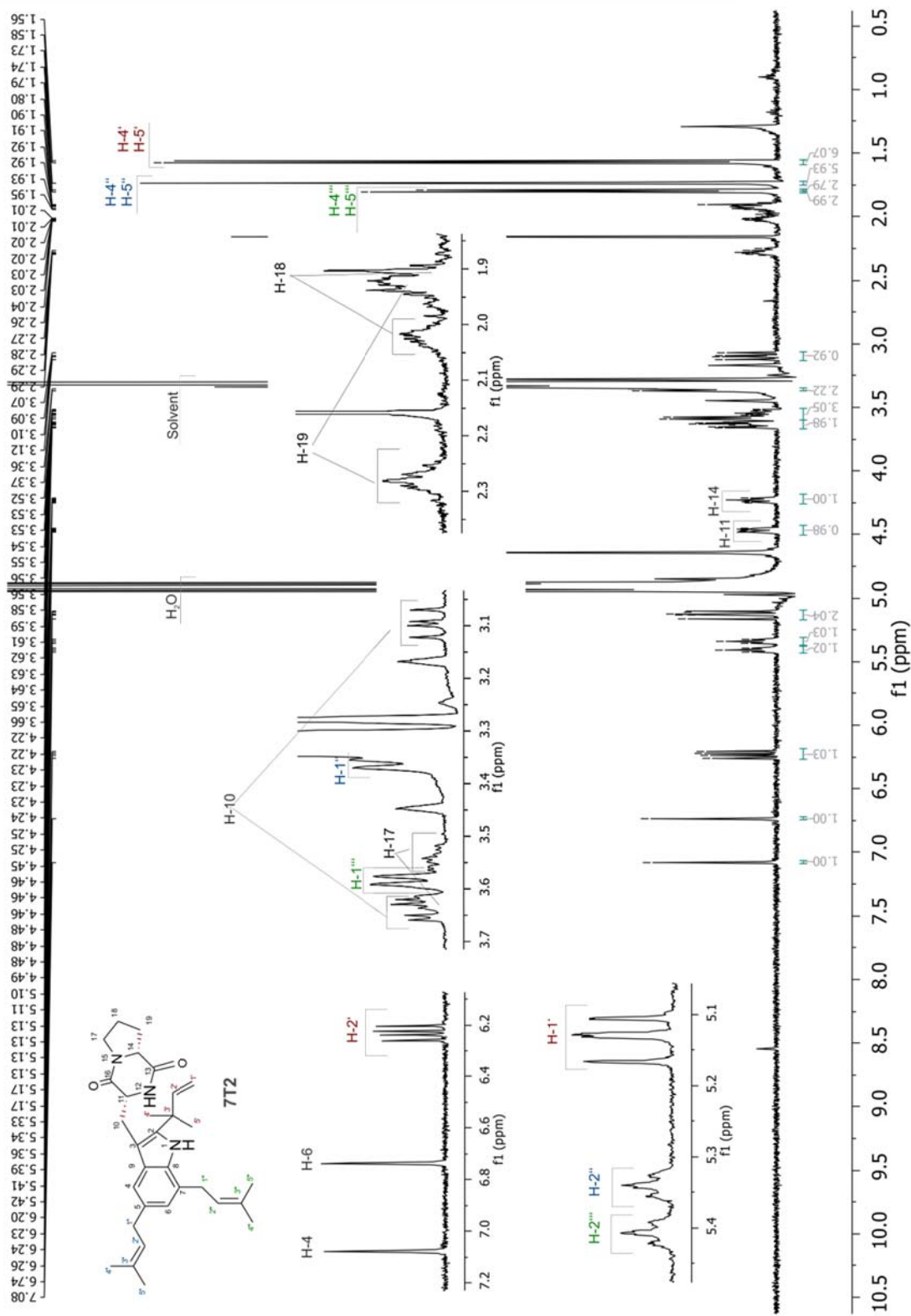
Figure S35 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-5,7-di-DMA-Trp-D-Pro (**7T2**) in CD$_3$OD (500MHz)

Figure S36 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-5,7-di-DMA-Trp-L-Pro (**8T2**) in CD$_3$OD (500MHz)
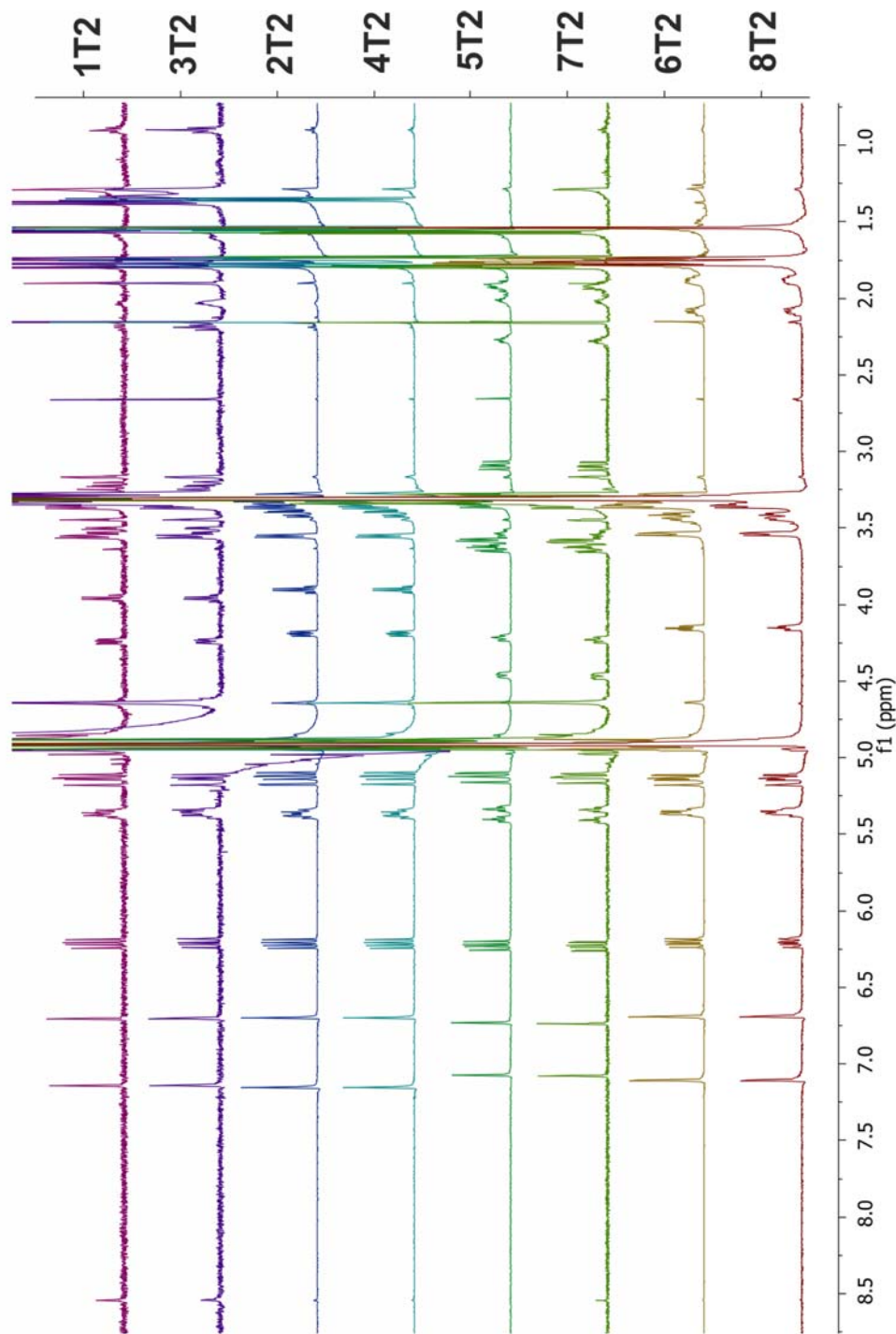
Figure S37 Comparison of ^{1}H NMR spectra of *cyclo*-L-2-tert-DMA-4,7-di-DMA-Trp-L-Ala (**1T2**), *cyclo*-L-2-tert-DMA-4,7-di-DMA-Trp-D-Ala (**2T2**), *cyclo*-D-2-tert-DMA-4,7-di-DMA-Trp-D-Ala (**3T2**), *cyclo*-D-2-tert-DMA-4,7-di-DMA-Trp-L-Ala (**4T2**), *cyclo*-L-2-tert-DMA-4,7-di-DMA-Trp-L-Pro (**5T2**), *cyclo*-D-2-tert-DMA-4,7-di-DMA-Trp-L-Pro (**6T2**), *cyclo*-D-2-tert-DMA-4,7-di-DMA-Trp-D-Pro (**7T2**) and *cyclo*-D-2-tert-DMA-4,7-di-DMA-Trp-L-Pro (**8T2**) in CD$_3$OD
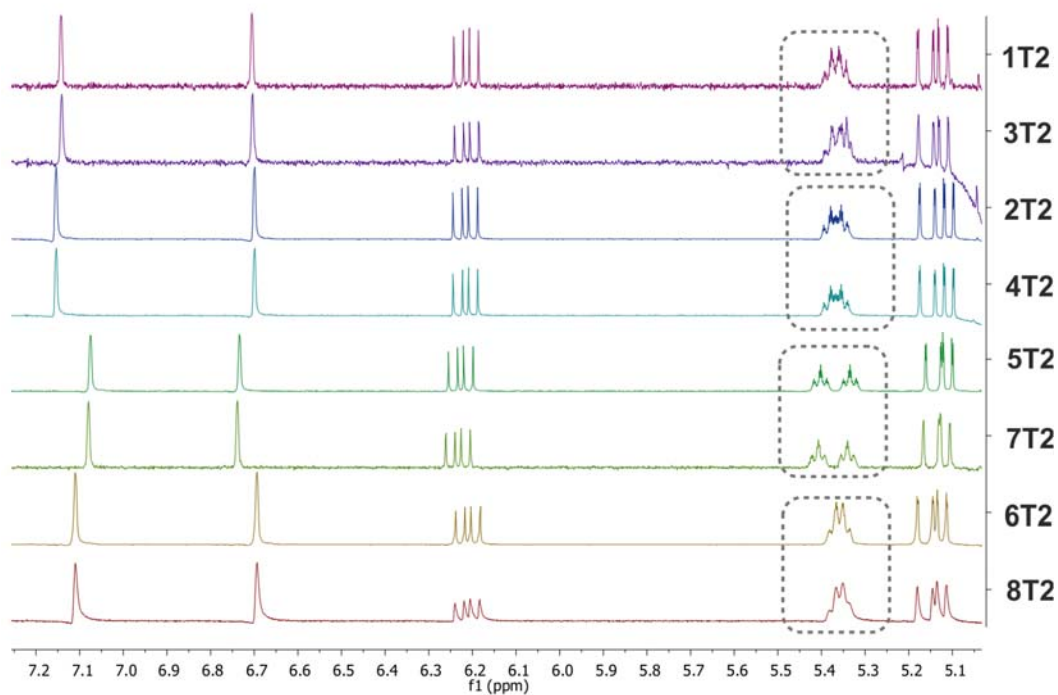
Figure S38 selected region (5.1 – 7.2 ppm) of the ^{1}H NMR of spectra **1T2 – 8T2** in CD$_3$OD
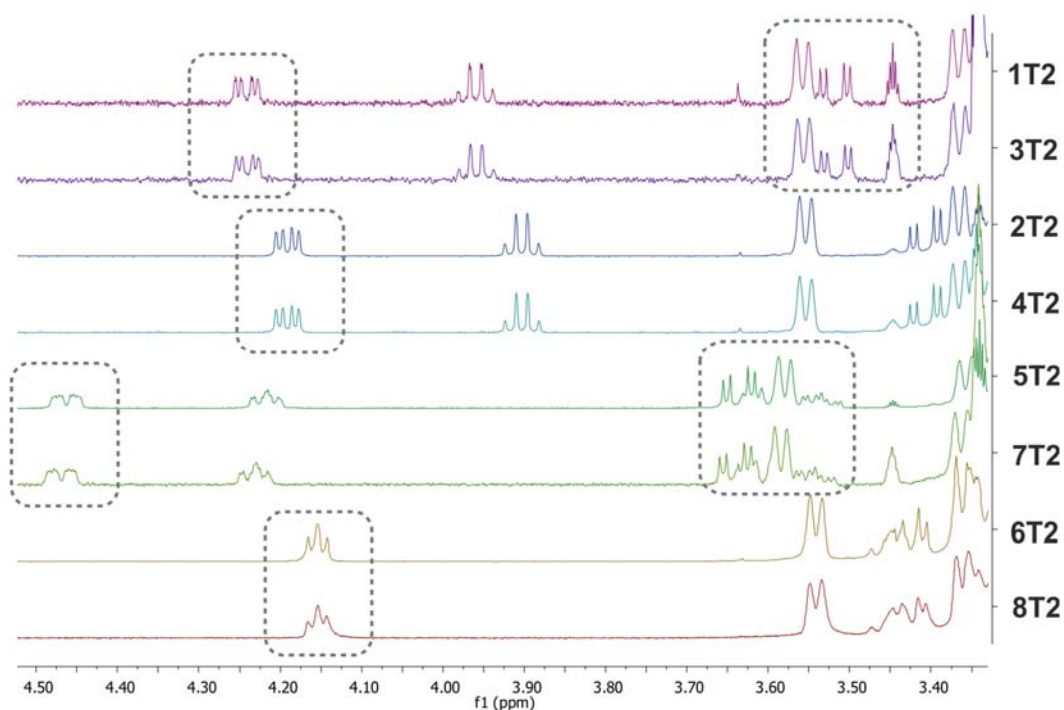
Figure S39 selected region (3.4 – 4.5 ppm) of the ^{1}H NMR of spectra **1T2 – 8T2** in CD$_3$OD
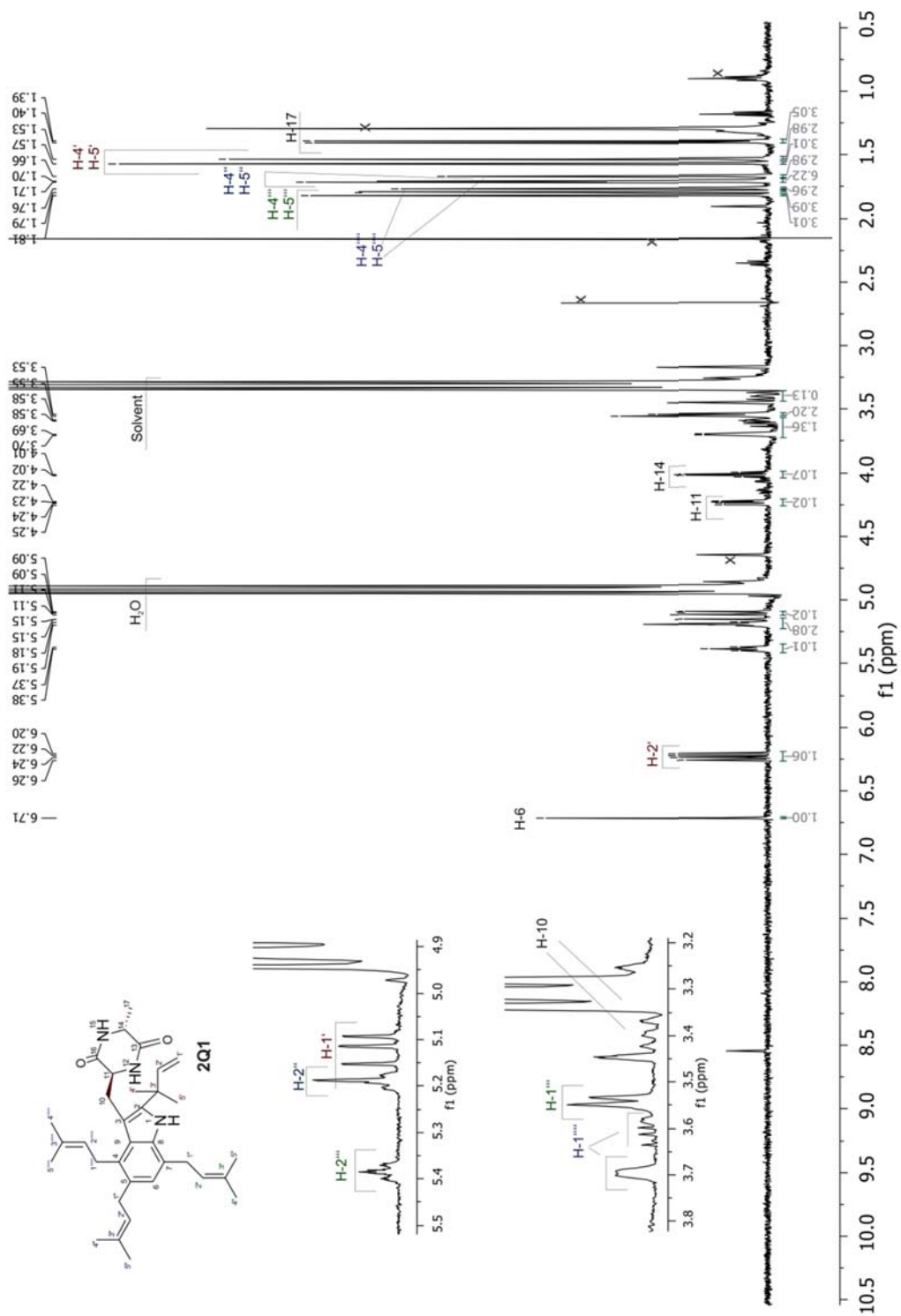
Figure S40 ^{1}H NMR spectrum of *cyclo*-L-2-*tert*-DMA-4,5,7-tri-DMA-Trp-D-Ala (**2Q1**) in CD$_3$OD (500MHz)

183
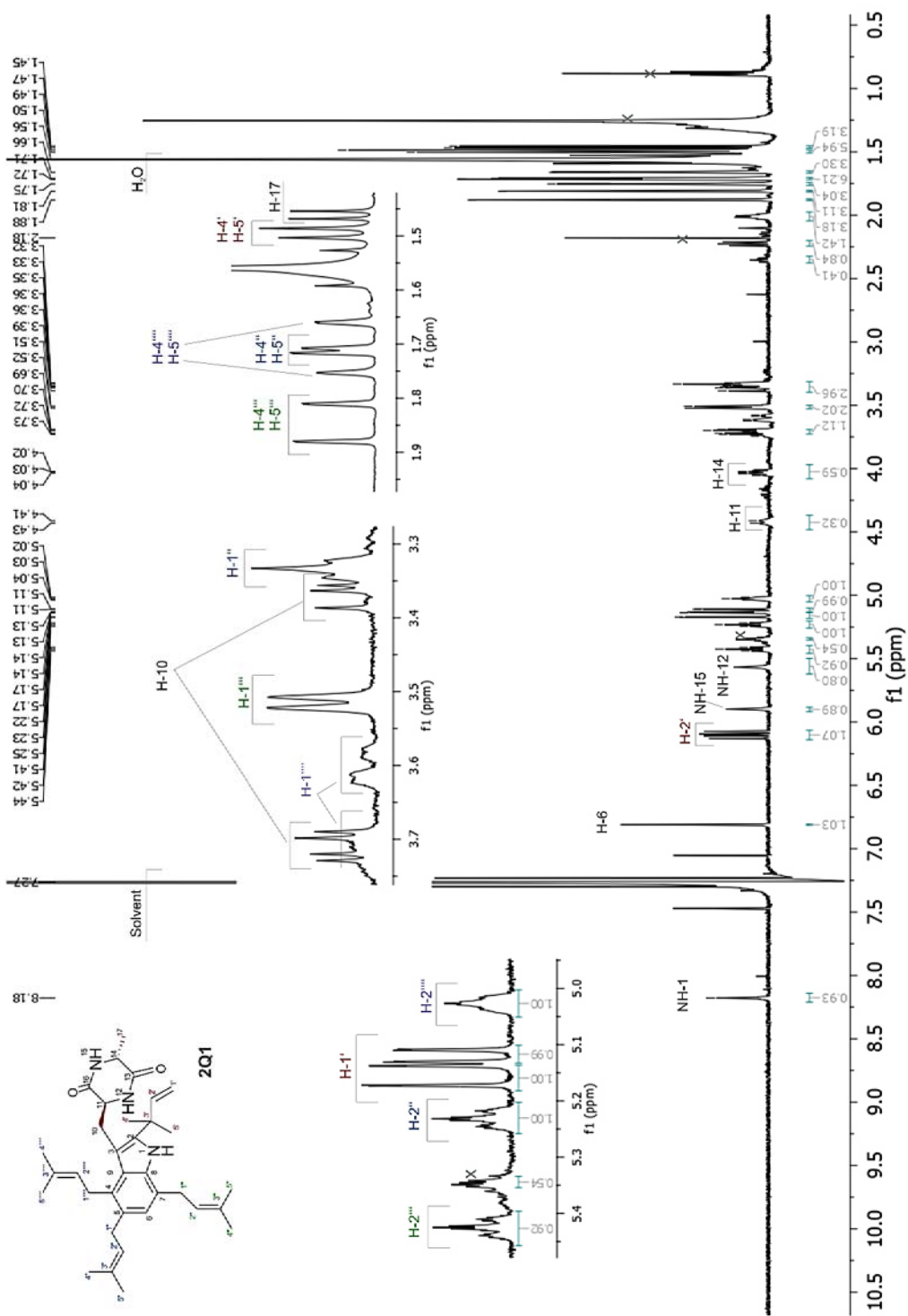
Figure S41 [^1]H NMR spectrum of *cyclo*-L-*2-tert*-DMA-4,5,7-tri-DMA-Trp-D-Ala (**2Q1**) in CDCl$_3$ (500MHz)
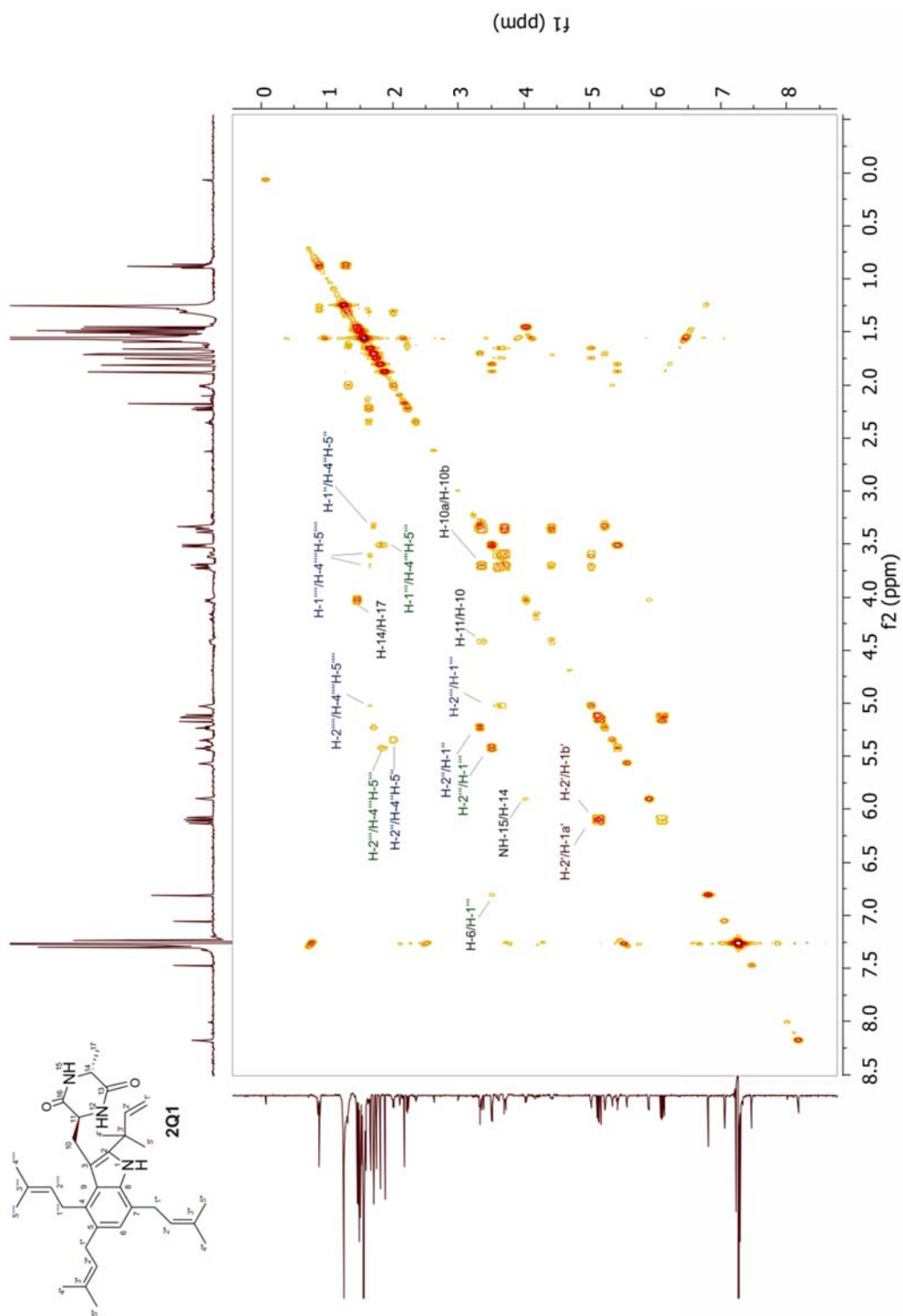
Figure S42 ^{1}H–^{1}H COSY spectrum of **2Q1** in CDCl$_3$ (500MHz)
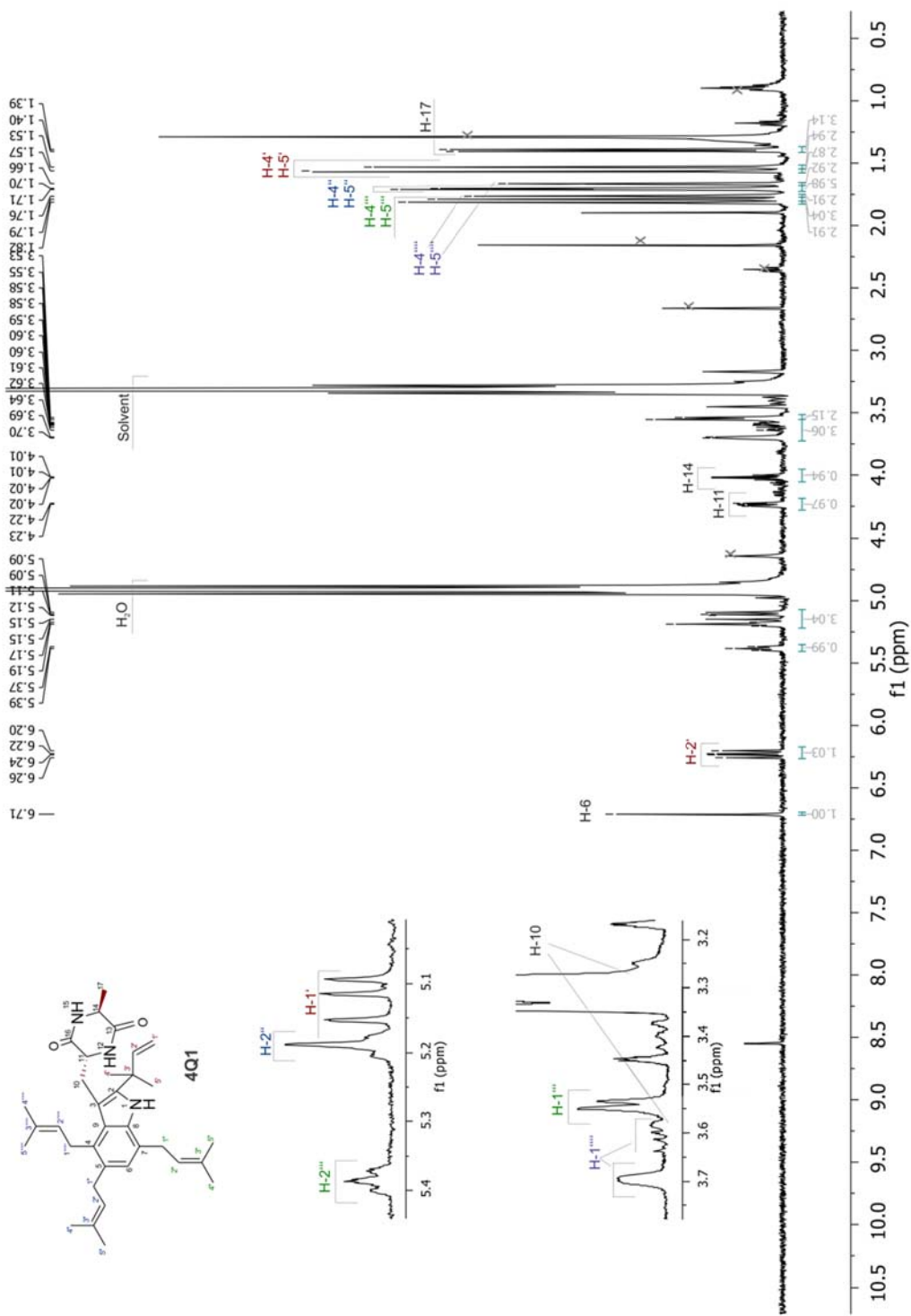
Figure S43 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-4,5,7-tri-DMA-Trp-L-Ala (**4Q1**) in CD$_3$OD (500MHz)

186

Figure S44 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-4,5,7-tri-DMA-Trp-L-Ala (**4Q1**) in CDCl$_3$ (500MHz)
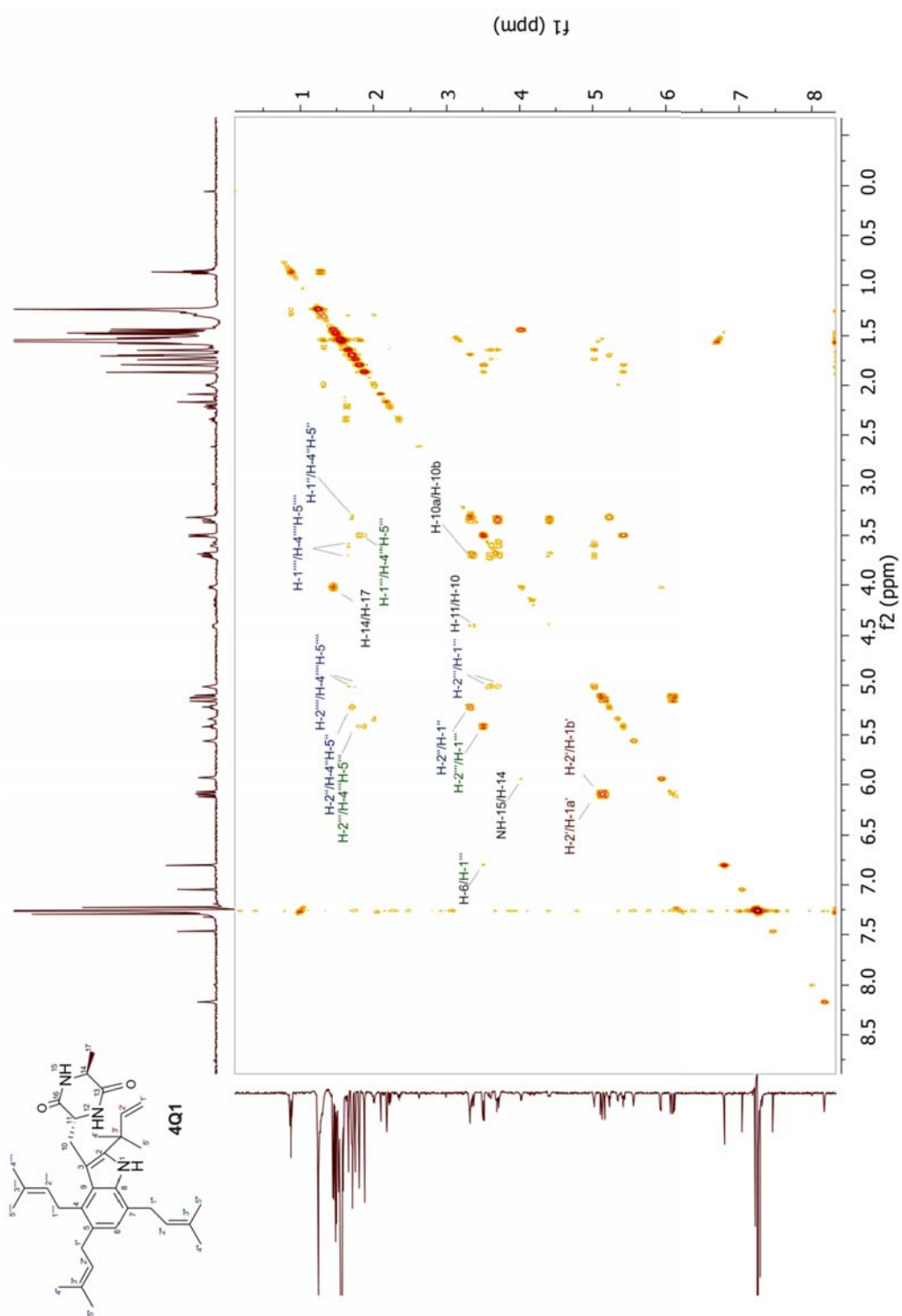
Figure S45 ^{1}H-^{1}H COSY spectrum of **4Q1** in CDCl$_3$ (500MHz)
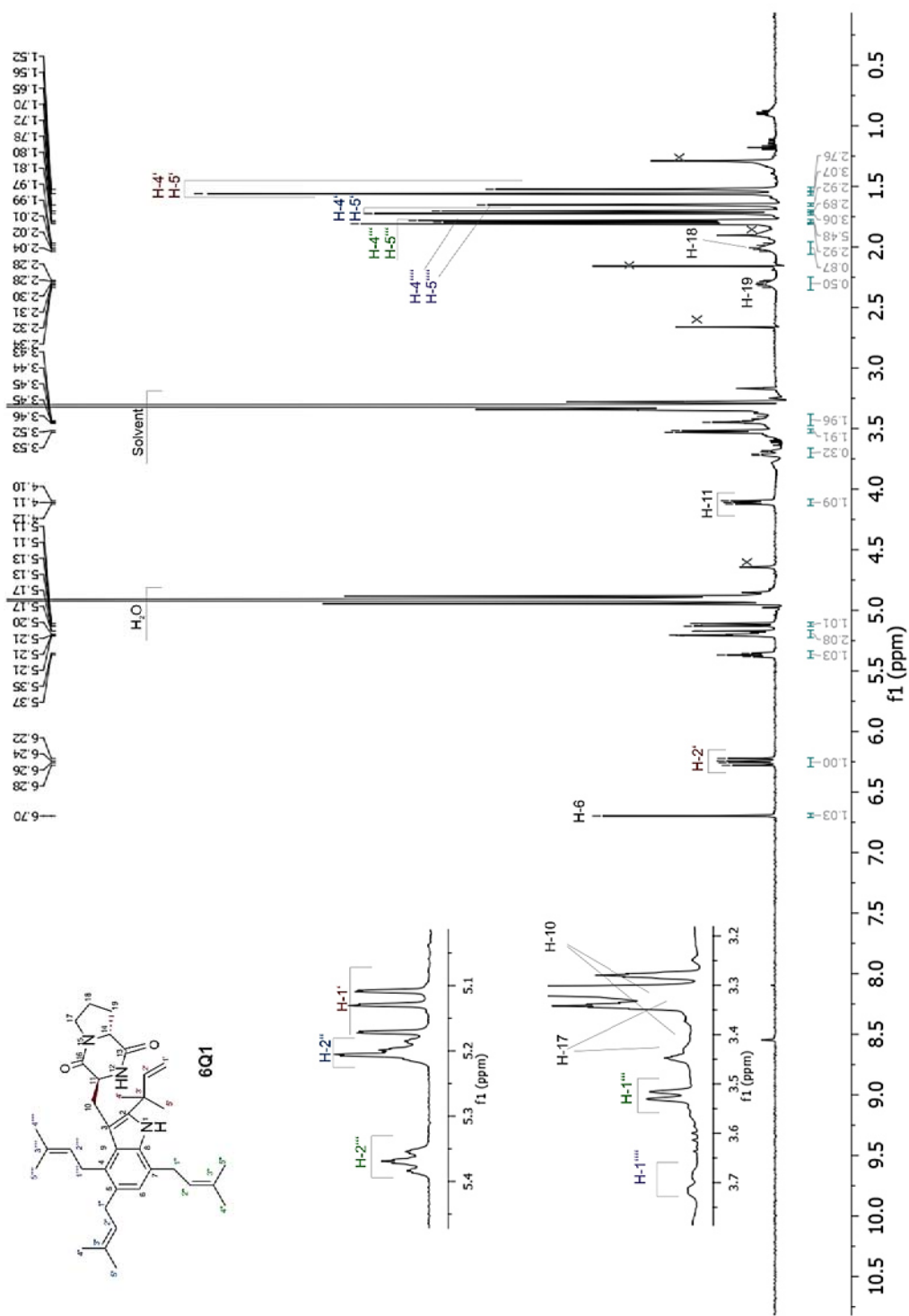
Figure S46 ^{1}H NMR spectrum of *cyclo*-L-2-*tert*-DMA-4,5,7-tri-DMA-Trp-D-Pro (**6Q1**) in CD$_3$OD (500MHz)
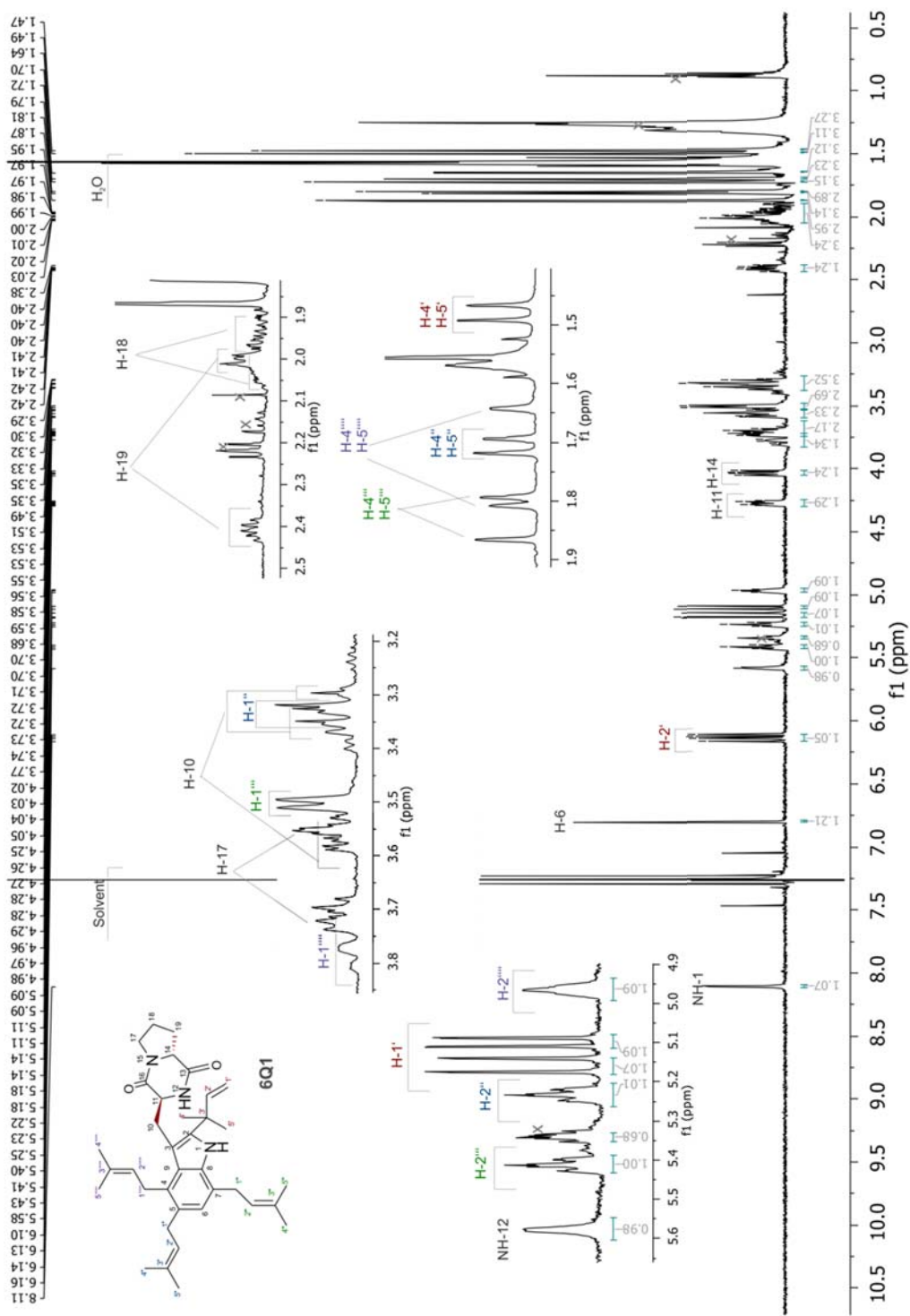
Figure S47 ^{1}H NMR spectrum of *cyclo*-L-2-*tert*-DMA-4,5,7-tri-DMA-Trp-D-Pro (**6Q1**) in CDCl$_3$ (500MHz)
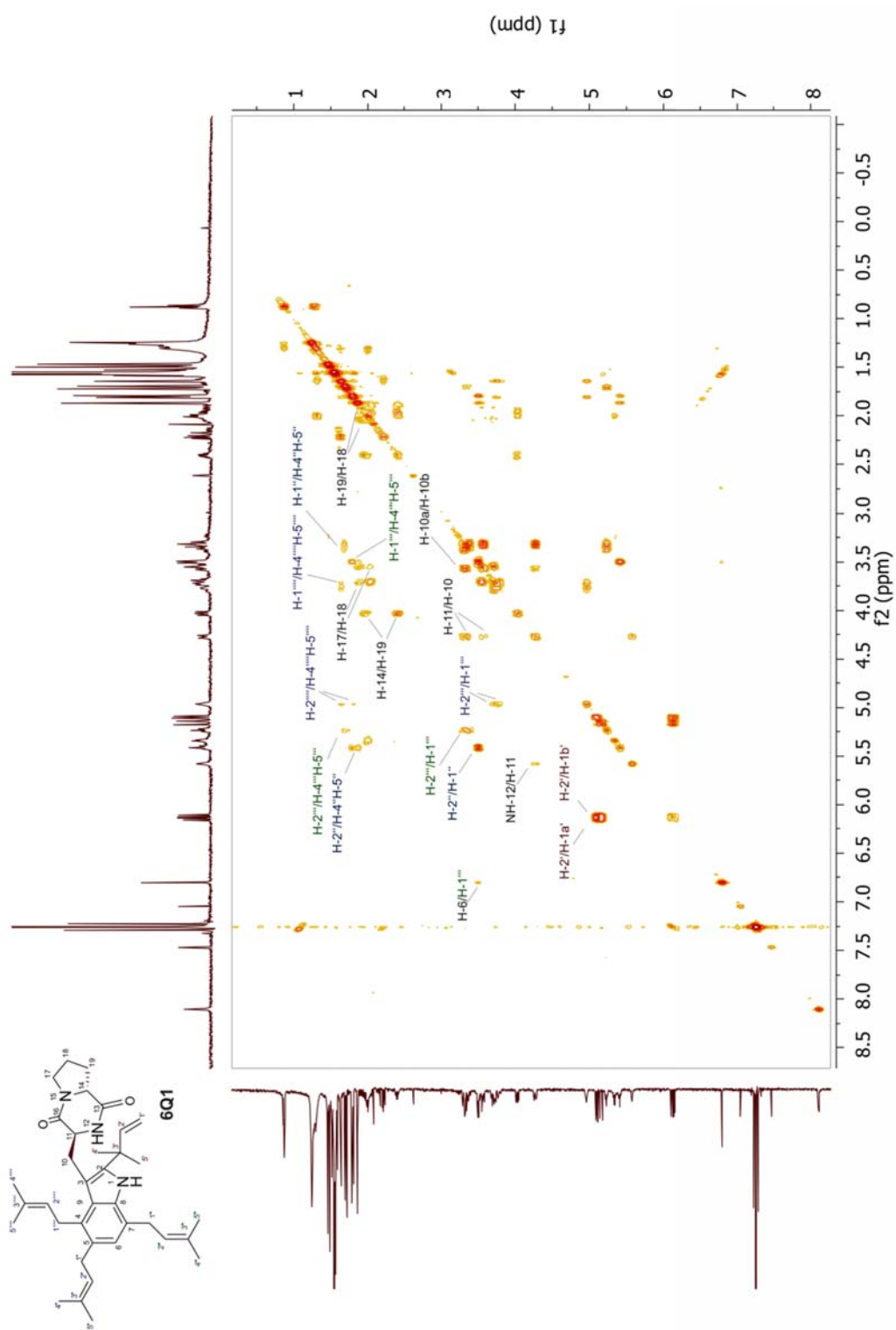
Figure S48 ^{1}H-^{1}H COSY spectrum of **6Q1** in CDCl$_3$ (500MHz)
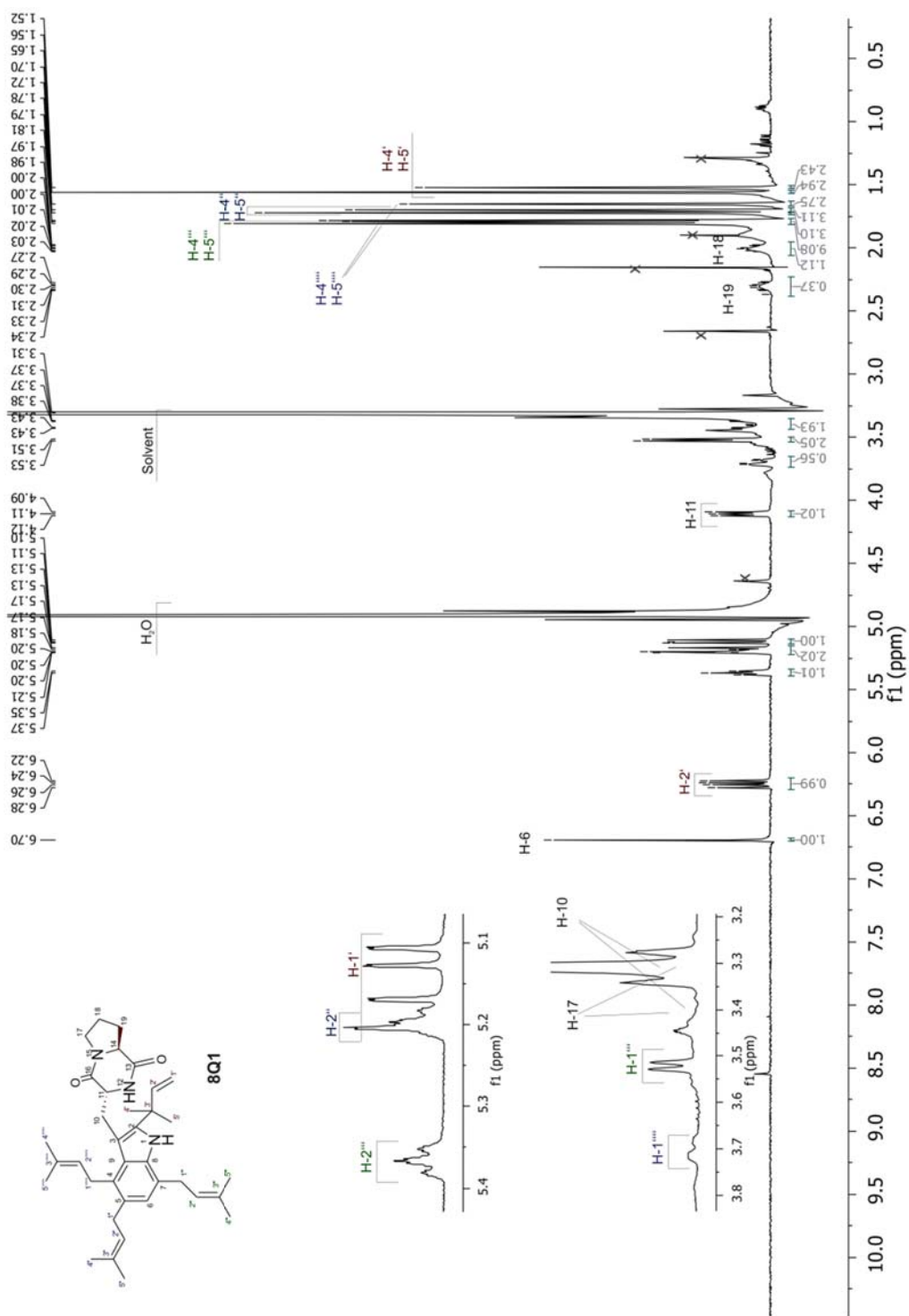
Figure S49 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-4,5,7-tri-DMA-Trp-L-Pro (**8Q1**) in CD$_3$OD (500MHz)
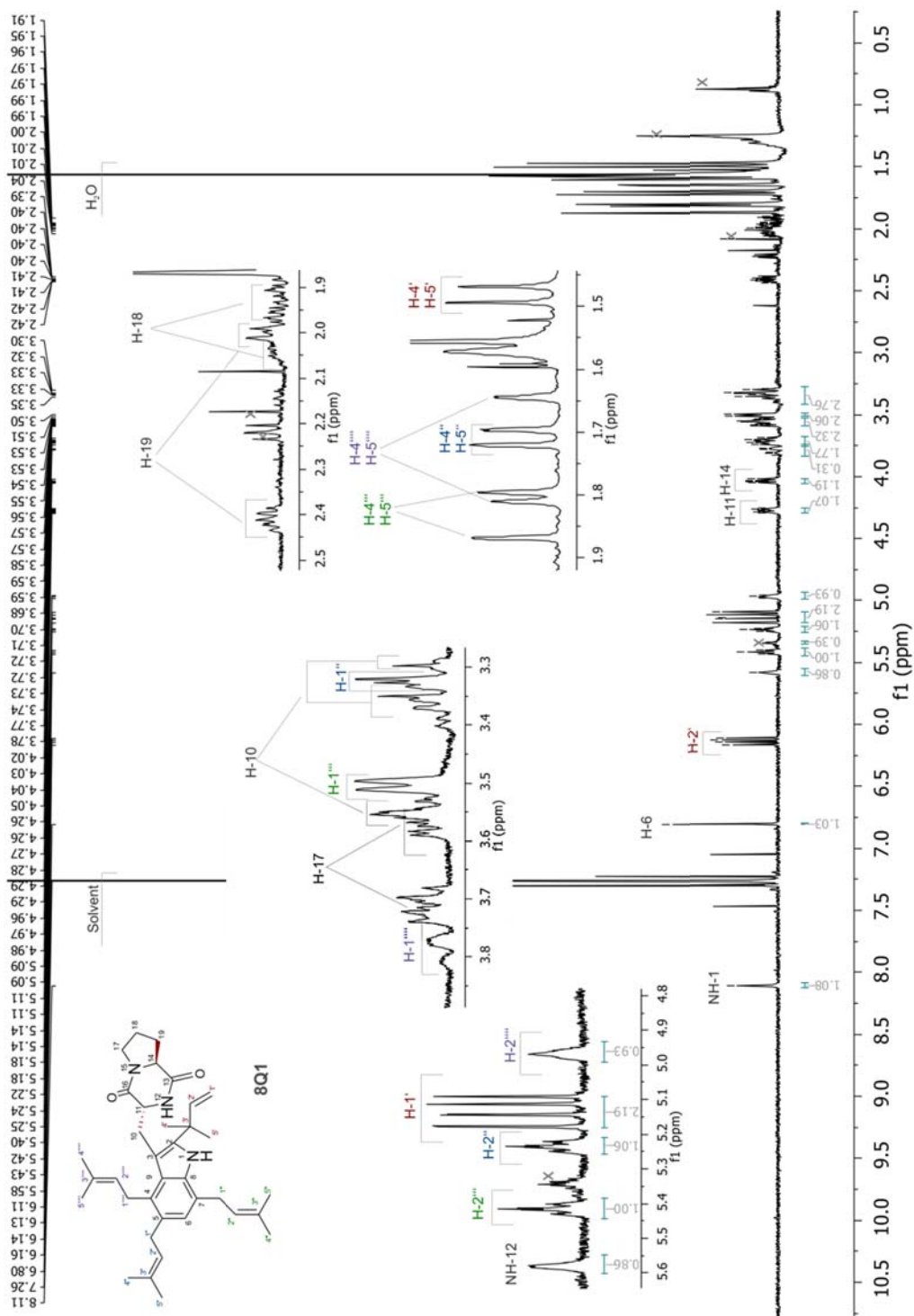
Figure S50 ^{1}H NMR spectrum of *cyclo*-D-2-*tert*-DMA-4,5,7-tri-DMA-Trp-L-Pro (**8Q1**) in CDCl$_3$ (500MHz)
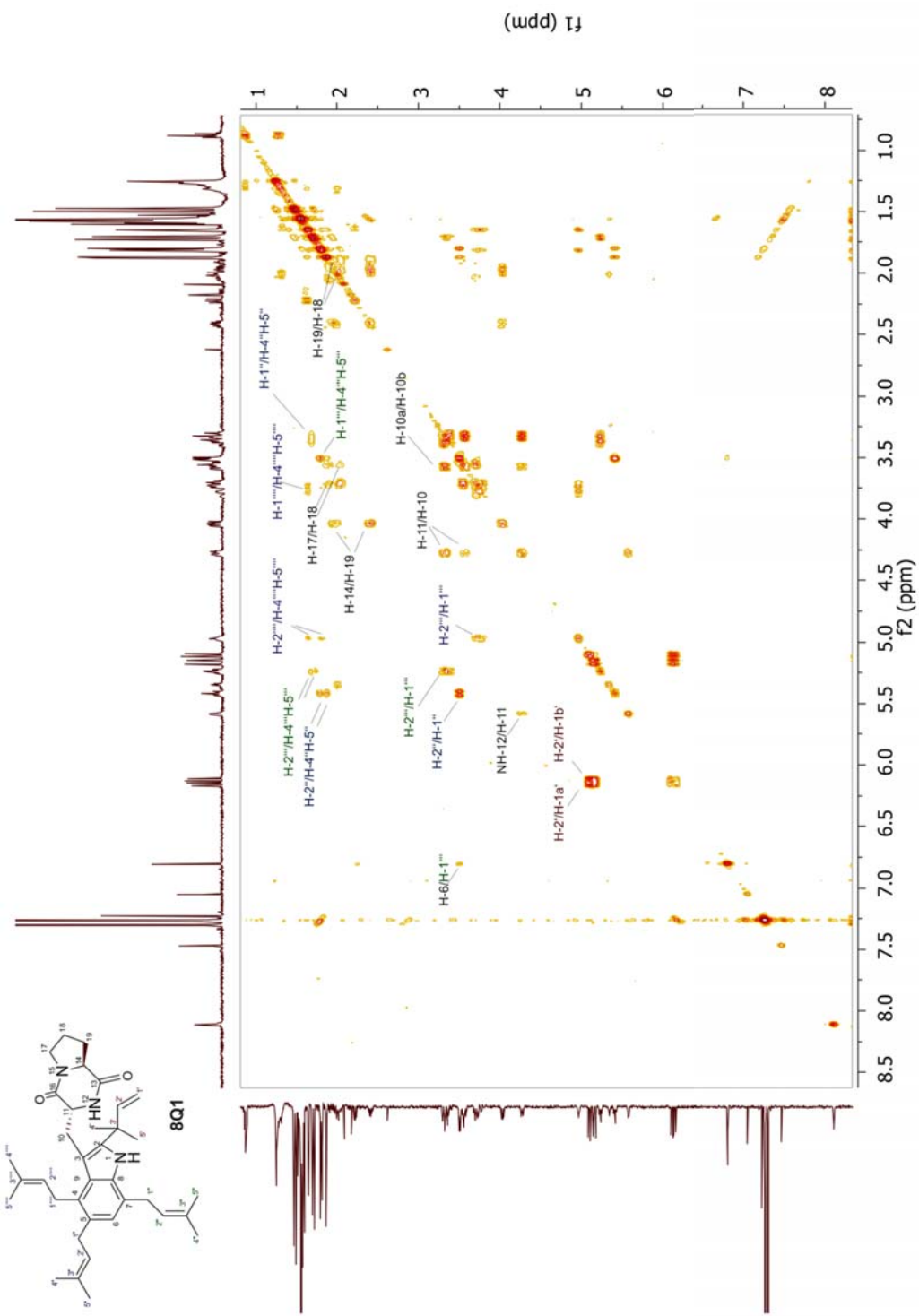
Figure S51 ^{1}H-^{1}H COSY spectrum of **8Q1** in CDCl$_3$ (500MHz)
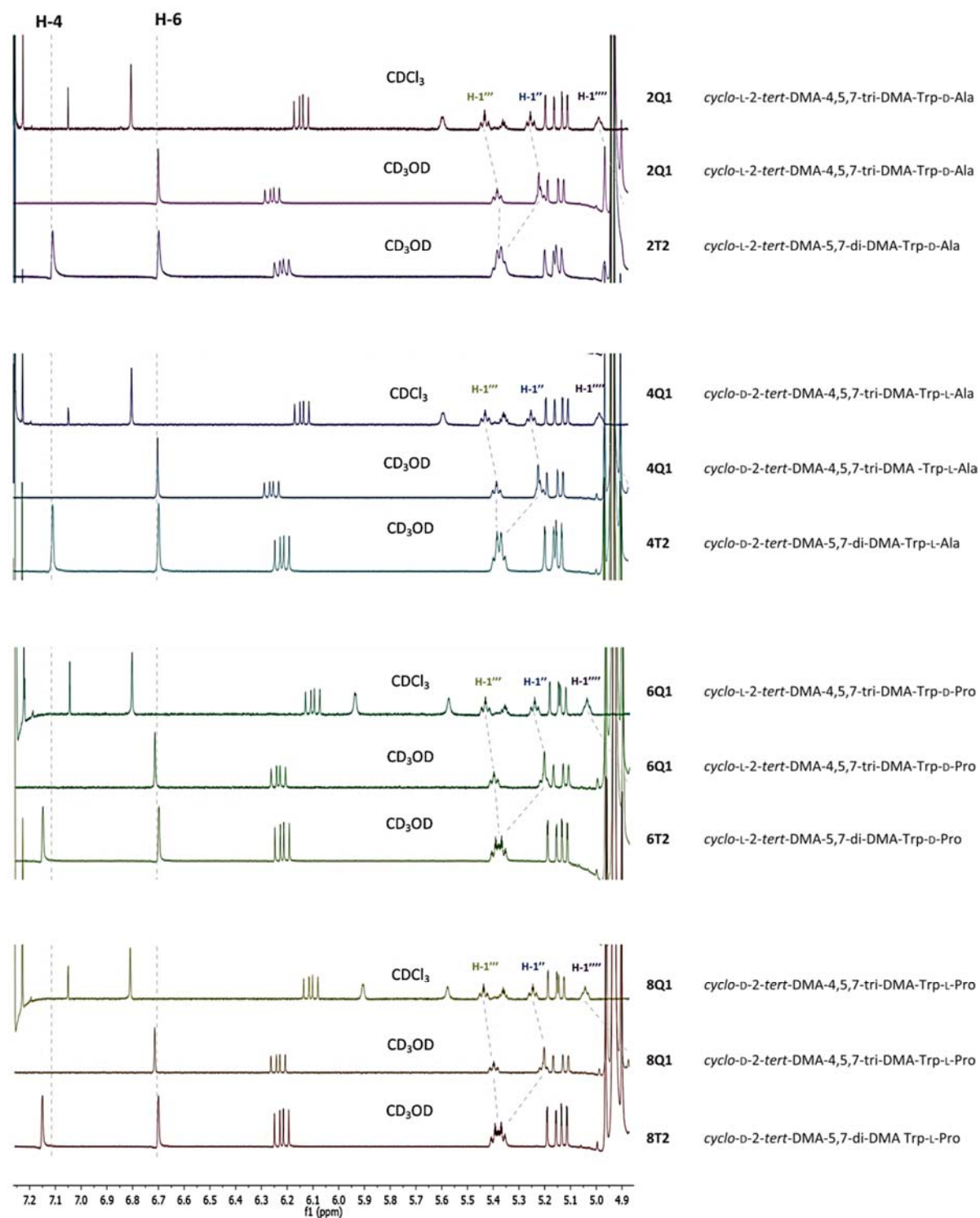
Figure S52 Visualization of the prenyl position at C-4 in the case of the fourth attached prenyl moiety in **2Q1, 4Q1, 6Q1** and **8Q1** by comparison of their ^{1}H NMR spectra with those of their precursors in CD$_3$OD and CDCl$_3$.

5 REFERENCES

References

Wohlgemuth V, Kindinger F, Xie X, Wang BG, Li SM (2017) Two Prenyltransferases Govern a Consecutive Prenylation Cascade in the Biosynthesis of Echinulin and Neoechinulin. Org Lett

Yin S, Yu X, Wang Q, Liu XQ, Li S-M (2013) Identification of a brevianamide F reverse prenyltransferase BrePT from *Aspergillus versicolor* with a broad substrate specificity towards tryptophan-containing cyclic dipeptides. Appl Microbiol Biotechnol 97:1649-1660

4.4. Genomic locus of a *Penicillium crustosum* pigment as an integration site for secondary metabolite gene expression.

Kindinger, F., Nies, J., Becker, A., Zhu, T., and Li, S. M. (2019)

Genomic Locus of a *Penicillium crustosum* Pigment as an Integration Site for Secondary Metabolite Gene Expression

Florian Kindinger,[†] Jonas Nies,[†] Anke Becker,[§] Tianjiao Zhu,[‡] and Shu-Ming Li*,[†]

[†]Institut für Pharmazeutische Biologie und Biotechnologie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany

[§]LOEWE Zentrum für Synthetische Mikrobiologie, Philipps-Universität Marburg, Hans-Meerwein-Straße, 35032 Marburg, Germany

[‡]Ocean University of China, School of Medicine and Pharmacy, Key Laboratory of Marine Drugs, Chinese Ministry of Education, Yushan Road 5, 266003 Qingdao, China

Ⓢ *Supporting Information*
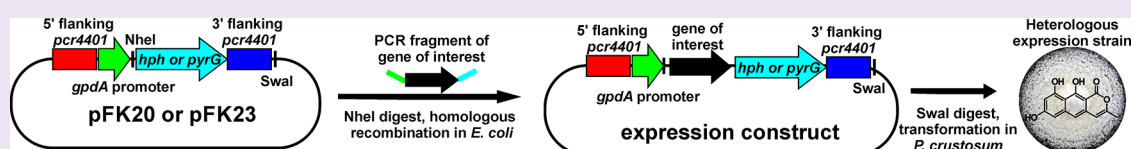
ABSTRACT: Heterologous expression of secondary metabolite genes and gene clusters has been proven to be a successful strategy for identification of new natural products of cryptic or silent genes hidden in the genome sequences. It is also a useful tool to produce designed compounds by synthetic biology approaches. In this study, we demonstrate the potential usage of the gene locus *pcr4401* in the fast-growing filamentous fungus *Penicillium crustosum* as an integration site for heterologous gene expression. The deduced polyketide synthase (PKS) Pcr4401 is involved in the dihydroxynaphthalene (DHN)-melanin pigment formation, and its deletion in *P. crustosum* PRB-2 led to an albino phenotype. Heterologous expression of *pcr4401* in *Aspergillus nidulans* proved its function as the melanin precursor YWA1 synthase. To ensure gene expression after genomic integration and to easily identify the potential transformants by visualization, the gene locus of *pcr4401* was chosen as an integration site. For heterologous expression in *P. crustosum*, the expression constructs were created by ligation-independent homologous recombination in *Escherichia coli* or *Saccharomyces cerevisiae*. A *pyrG* deficient strain was also created, so that both the *pyrG* and *hph* resistance gene can be used as selection markers. Successful expression in *P. crustosum* was demonstrated by using one uncharacterized PKS gene from *Aspergillus* and two from *Penicillium* strains. All three genes were successfully introduced, heterologously expressed, and their biosynthetic products elucidated. The results presented in this study demonstrated that *P. crustosum* can be used as a suitable host for heterologous expression of secondary metabolite genes.

Natural products (NPs) produced by microorganisms are a valuable source for finding and development of new potential drugs. In total, 46% of the 1562 approved drugs between the years 1981 and 2014 originated from natural sources including NPs (4%) and their derivatives (21%) as well as NP mimics or pharmacophores (21%).[1] Bacteria, especially the actinobacteria, are often producers of promising NPs.[2] Eukaryotic fungi of the phylum Ascomycota, *e.g., Penicillium* and *Aspergillus* belong to another important group of NP producers.[3] In recent years, bioactive fungal secondary metabolites (SMs) have been increasingly identified. From 2001 to 2010, 61% of the total identified bioactive microbial metabolites originated from fungi.[4] The ongoing process in sequencing technology significantly decreased the cost for genome sequencing. Therefore, the scientific community easily gets access to a rapidly increasing number of microbial genomes.[5] Mining the bacterial and fungal genomes revealed the presence of a huge genetic potential for SM production, which is far from being exhausted. Most biosynthetic gene clusters (BGCs) for SM production remain silent under laboratory culture conditions. Such silent BGCs are estimated to be approximately 90%.[6] Several approaches including

alterations in the cultivation conditions, use of biotic and abiotic elicitors, or interference with the condensation state of the genomic DNA, which regulates the accessibility of a DNA region for transcription factors, were developed to activate the silent BGCs.[7] In addition, heterologous expression of single genes or whole BGCs in an appropriate host is very promising to further access microbial SM production capability or to increase product diversity by synthetic biology approaches.[8]

Next to *Escherichia coli* and *Saccharomyces cerevisiae*, filamentous fungi belonging to the genus *Aspergillus* are the most used heterologous expression hosts for SM production.[8−10] In particular, the model organism *Aspergillus nidulans* is an already established and frequently used host for heterologous gene expression. *Aspergillus oryzae*, known from food technology, also proved to be a reliable host. Furthermore, *Aspergillus niger* can also be a suitable choice for heterologous expression. In comparison, there are only a

few reports on heterologous expression in *Penicillium* species. For example, the penicillin producer *Penicillium rubens* (formerly known as *Penicillium chrysogenum*) was successfully applied as an expression host.[11]

Heterologous expression may be hindered by host-dependent factors, *e.g.*, insufficient or missing supply of biosynthetic precursors, different intron splicing behavior compared to the native host or host-dependent cross-chemistry. Song *et al.*[12] reported the difficulties for the heterologous expression of a polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) hybrid gene. The reason was different splicing in the native host *Magnaporthe oryzae* and in the expression host *A. oryzae*. In a recent study, Geib *et al.*[13] investigated NRPS-like enzymes by domain swapping and heterologous expression in *A. niger* and *A. oryzae* and observed host-dependent product formation. The authors suggested to use different heterologous expression hosts for characterization of yet unknown NRPS-like enzymes.

Penicillium crustosum is a ubiquitously occurring filamentous fungus and has been identified and isolated from a wide range of ecological niches including goods for human consumption like corn, bread, meat, cheese, fruits, and nuts.[14] It is a fast-growing fungus on various media within a temperature range of 5 to 35 °C with an optimal growth temperature of 25 °C.[15] *P. crustosum* strains are also known producers of SMs including different mycotoxins like the neurotoxic penitrems, andrastins, terrestric acid, roquefortines, viridicatols, peniphenones, and penilactones.[16,17] Initially interested in the biosynthetic pathway of polyketides produced by *P. crustosum* PRB-2,[17] we sequenced its genome and identified a nonreducing polyketide synthase (NR-PKS), named Pcr4401, being responsible for the biosynthesis of the DHN-melanin precursor YWA1. In this study, we report the potential use of the *pcr4401* gene locus in *P. crustosum* as an integration site for heterologous gene expression.

■ RESULTS AND DISCUSSION

Genome Sequencing for *P. crustosum* PRB-2. Sequencing the genome of *P. crustosum* PRB-2 resulted in a draft genome sequence with an estimated genome size of 32.4 Mbp organized in 145 contigs. This sequence was used for the identification of SM backbone genes by antiSMASH,[18] combined with a BLAST search using PKS and NRPS domains as queries. Open reading frame and protein sequence prediction were achieved with FGENESH[19] and secondFind (http://biosyn.nih.go.jp/2ndfind/). In total, 52 backbone genes including 23 PKSs, 19 NRPSs, 6 NRPS-like enzymes, as well as 4 NRPS-PKSs have been identified.

Identification of Pcr4401 as a Putative Conidial Pigment Polyketide Synthase. Six of the identified PKSs are NR-PKSs. To identify possible biosynthetic products of the NR-PKSs, a BLAST search and sequence comparison with known PKSs were performed. Thereby, PKS Pcr4401 showed sequence identities of 66.6% on the amino acid level to the polyketide synthase for naphthopyrone YWA1 encoded by the *wA* gene in *A. nidulans* and 65.8% to the conidial pigment polyketide synthase Alb1 from *A. fumigatus*. All three PKS share the same SAT-KS-AT-PT-ACP-ACP-TE domain architecture (see ref 20 for the PKS domain abbreviations). We proved Pcr4401 function by gene deletion in *P. crustosum*. A deletion cassette was prepared by inserting the sequence of the *hph* gene under the control of the constitutive *gpdA* promoter and the *trpC* terminator from *A. nidulans* between the up- and

downstream sequences of *pcr4401* (Tables S1 and S2 in the Supporting Information). Introducing this DNA fragment into wildtype strain PRB-2 and selection on hygromycin B-containing plates led to identification of 215 resistant colonies. A total of 21 of them showed an albino phenotype (*e.g.*, strain FK02 in Figure 1). PCR amplification verified the replacement
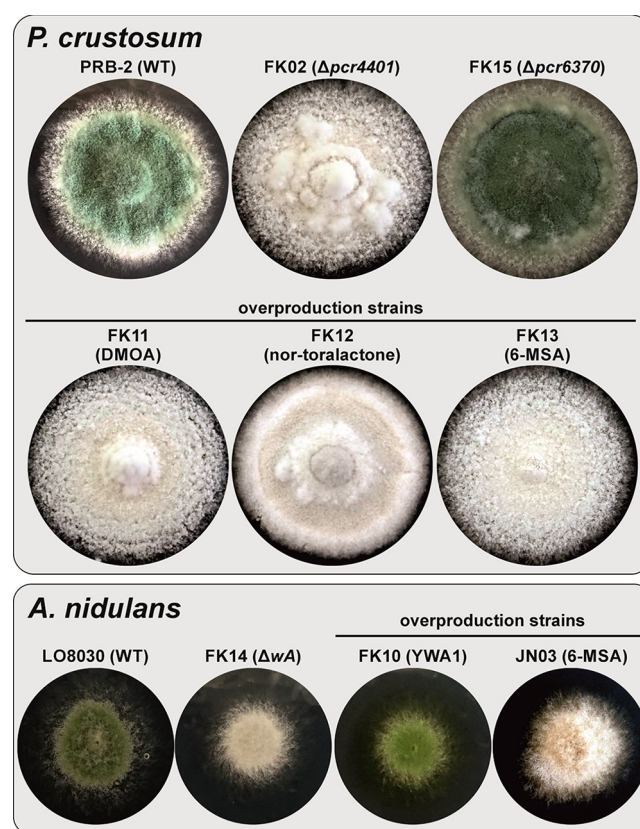
Figure 1. Phenotypes of fungal strains grown on GMM for 7 days at 25 °C. Deletion and replacement of *pcr4401* in *P. crustosum* led to the albino phenotype as detected in strains FK02 and FK11−FK13. Similar phenomenon was also observed in *A. nidulans* strain FK14 after vector integration into the *wA* locus of LO8030. Heterologous expression of *pcr4401* in *A. nidulans* LO8030 restored the green conidial color of the wild type (FK10).

of *pcr4401* by the *hph* gene (Figure 2A,B and Table 1), indicating the crucial role of Pcr4401 in the pigmentation. No changes on other SMs or growing behavior (Table S3) could be detected in the Δ*pcr4401* strain FK02, compared to the wildtype. Moreover, cultivation of *P. crustosum* in Czapek dox broth (CDB) and metabolite analysis showed a relatively clear SM background (Figure S1), providing a solid basis for detection and isolation of heterologously produced compounds.

The gene targeting efficiency was estimated to be 9.8% (Table 2). In total, 194 or 90.2% of the transformants showed hygromycin B resistance without deletion of the gene *pcr4401*, most likely due to ectopic integration of the gene deletion cassette instead of homologous recombination at the targeted gene. Gene targeting efficiency could be increased by using a nonhomologous end joining deficient (Δ*ku70/80*) strain or using split-marker gene deletion method.[21,22] For deletion or disruption of genes related to pigment formation, *e.g.*, the *pcr4401* gene locus, the albino phenotype upon successful targeted integration can be easily detected by visualization, so
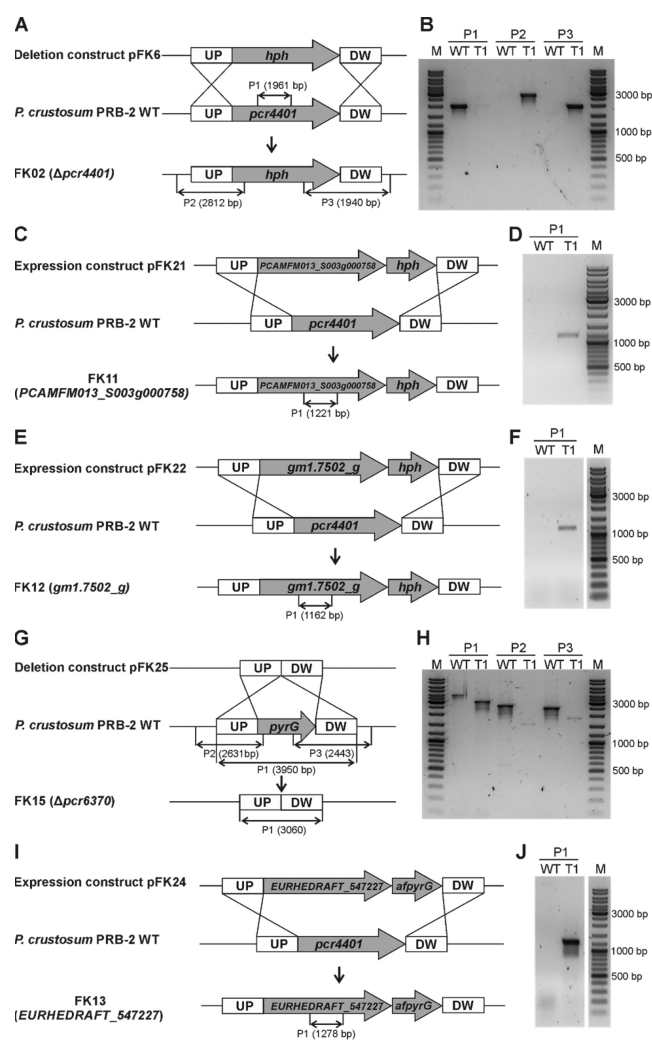
Figure 2. Schematic presentation of the gene deletions/replacements at the *pcr4401* and *pcr6370* (*pyrG*) locus in *P. crustosum* (A, C, E, G, and I) and PCR verification (B, D, F, H, and J). UP, upstream; DW, downstream; hph, hygromycin B resistance; *afpyrG*, *A. fumigatus pyrG*.

that these strategies are not essential for heterologous expression at such loci.

Heterologous Expression Proved Pcr4401 as an YWA1 Synthase. To clarify whether Pcr4401 is indeed

Table 2. Overview on Transformation Results

				transformants		
construct	DNA [μg]	marker	protoplasts	total	proven	positive [%]
pFK06	10	*hph*	2.1×10^8	134	15	11.2
	10	*hph*	2.0×10^7	38	2	5.2
	10	*hph*	1.2×10^7	43	4	9.3
pFK21	2	*hph*	5.4×10^7	51	4	7.8
	2	*hph*	2.0×10^7	62	4	6.5
	2	*hph*	1.2×10^7	47	3	6.4
pFK22	2	*hph*	3.1×10^7	36	3	8.3
	2	*hph*	2.0×10^7	68	5	7.4
	2	*hph*	1.2×10^7	40	3	7.5
pFK24	5	*pyrG*	8.7×10^7	21	2	9.5
	5	*pyrG*	1.7×10^7	19	2	9.5
	5	*pyrG*	1.5×10^7	32	1	6.3
pFK25	10	5-FOA[a]	1.0×10^8	2	2	100
	10	5-FOA[a]	1.7×10^7	2	1	50
	10	5-FOA[a]	1.5×10^7	4	1	25

[a]5-Fluoroorotic acid was used for selection of *pyrG* deficient *P. crustosum* strains.

involved in the biosynthesis of the green conidial pigment, the whole *pcr4401* sequence was amplified from PRB-2, cloned into the expression vector pYH-wA-pyrG,[23] and introduced site-specifically into the *wA* locus of *A. nidulans* LO8030,[24] resulting in the strain FK10. In parallel, the *wA* locus was also replaced by the empty vector and the resulted strain FK14 was used as a control. Cultivation of FK10 on GMM showed secretion of yellow compounds to the medium and restored green conidial pigmentation compared to strain FK14 (Figure 1). This proved that *pcr4401* can complement the function of *wA*. Liquid chromatography−mass spectrometry (LC−MS) analysis of the extract of a 3 day old rice culture led to the detection of two additional product peaks at 7.7 min with a [M + H]$^+$ ion at *m/z* 275.0568 ($C_{14}H_{12}O_6$) and 9.0 min with a [M + H]$^+$ ion at *m/z* 259.0599 ($C_{14}H_{10}O_5$), which are absent in the extract of FK14 (Figure 3). The mass data and UV spectra of the two compounds correspond very well to those of the known heptaketide YWA1 (**1**) and its dehydrated derivative YWA2, also known as nor-rubrofusarin (**2**) (Figure S2A,B). To confirm their structures, both compounds were isolated by high-performance liquid chromatography (HPLC) and subjected to ^{1}H NMR analysis, which unequivocally proved their integrity (Figures S5 and S6, Table S5).[25,26] The yields of

Table 1. Strains Used and Created in This Study

strains	genotype	source
P. crustosum PRB-2	wild type	27
P. crustosum FK02	Δ*pcr4401*::P*gpdA*-*hph* in *P. crustosum* PRB-2	this study
P. crustosum FK11	Δ*pcr4401*::P*gpdA*-PCAMFM013_S003g000758:*hph* in *P. crustosum* PRB-2	this study
P. crustosum FK12	Δ*pcr4401*::P*gpdA*-gm1.7502_g:*hph* in *P. crustosum* PRB-2	this study
P. crustosum FK13	Δ*pcr4401*::P*gpdA*-EURHEDRAFT_547227:*afpyrG*, Δ*pcr6370* in *P. crustosum* PRB-2	this study
P. crustosum FK15	Δ*pcr6370*	this study
A. nidulans LO8030	*pyroA4, riboB2, pyrG89, nkuA::argB*, deletion of eight SM biosynthetic gene clusters	24
A. nidulans FK10	Δ*wA*::P*gpdA*-*pcr4401*:*afpyrG* in *A. nidulans* LO8030	this study
A. nidulans FK14	Δ*wA*::*afpyrG* in *A. nidulans* LO8030	this study
A. nidulans JN03	Δ*wA*::P*gpdA*-EURHEDRAFT_547227:*afribo* in *A. nidulans* LO8030	this study
P. camemberti FM013	wild type	28
P. brevicompactum NRRL864	wild type	NRRL
A. ruber QEN-0407-G2	wild type	29
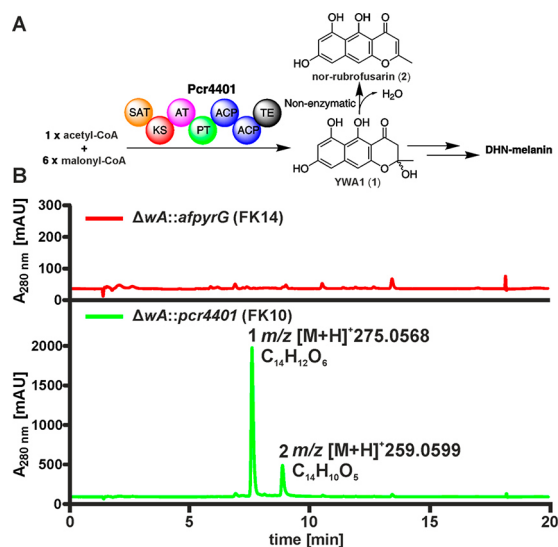
Figure 3. Heterologous expression of *pcr4401* from *P. crustosum* in *A. nidulans* LO8030: (A) domain architecture and the expected products and (B) LC−MS analysis of the control strain FK14 and the *pcr4401* expression strain FK10 after cultivation on rice for 3 days.

YWA1 and nor-rubrofusarin were calculated to be 122 mg and 60 mg per kg of 7 day-old rice culture, respectively. Elongated cultivation of FK10 increased the green conidial color intensity, indicating the conversion of YWA1 to the conidial pigments.

Guzman-Chavez et al.[30] demonstrated by gene deletion and heterologous overexpression experiments that PKS17 (CAP96497.1) from *P. rubens* is an orthologue of *A. fumigatus* naphthopyrone synthase Alb1[31] and also catalyzes the formation of YWA1. Thereby they proved the crucial role of PKS17 in the DHN-melanin biosynthetic pathway and in conidial pigmentation in *P. rubens*. Pcr4401 from *P. crustosum* showed a high sequence identity of 93.9% at the amino acid level to PKS17 from *P. rubens*. From the genome sequence of *P. crustosum*, we were able to identify genes coding for the same orthologous enzymes in the DHN-melanin pathway in *A. fumigatus* (Figure S3, Table S4), with sequence identities of 48.6−69.9%. We assume that the DHN-melanin formation in *P. crustosum* is catalyzed in the same manner as in *A. fumigatus* and *P. rubens*.

Heterologous Gene Expression in *P. crustosum* PRB-2. Successful deletion of *pcr4401* with the hygromycin B resistance marker under the control of the *A. nidulans* gpdA promoter proved that this system is also working in *P. crustosum*. The next step was to investigate the possibility of heterologous gene expression in *P. crustosum*. For proof of principal, we selected two genes coding for the PKS CRL19966 from *Penicillium camemberti* and Gm1.7502_g from *Penicillium brevicompactum* for heterologous expression. For this purpose, an expression vector pFK20 (Figure S4) with the hygromycin B resistance marker was constructed. The gene to be expressed will be cloned after the gpdA promoter sequence. The up- and downstream flanking regions of *pcr4401* were inserted before the gpdA promoter responsible for expression of the gene of interest and behind the hygromycin B resistance gene. This ensures the site-specific integration at the *pcr4401* locus for active gene expression and easy identification of the transformants by visual observation of the albino phenotype.

The PKS CRL19966.1 from *P. camemberti* was reannotated with secondFind leading to a revised protein sequence with 2636 amino acids and a SAT-KS-AT-PT-ACP-ACP-MT-TE domain architecture (Figure 4B). To the best of our
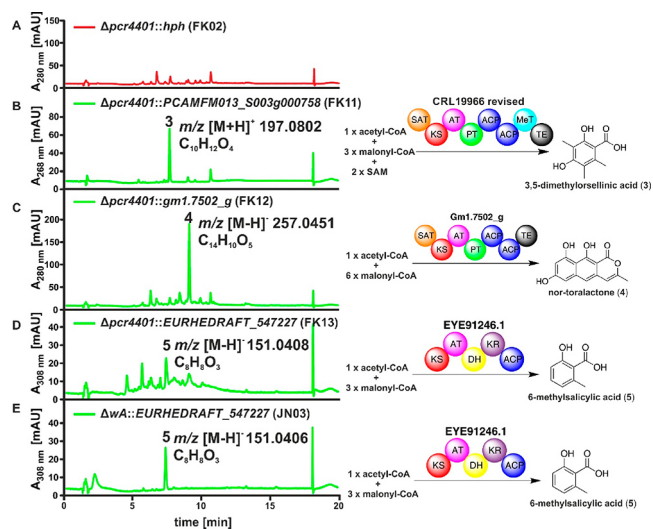
Figure 4. Heterologous expression of PKS genes in *P. crustosum* (B−D) and *A. nidulans* (E). LC−MS analysis of the extracts from different transformants and schematic representation of the PKS domain architectures with products.

knowledge, this PKS has not been investigated experimentally. A BLAST analysis against the UniProt database[32] revealed that the top three hits of known PKSs are AndM,[33] AusA,[34] and MpaC[35] with 42.4, 41.3, and 40.4% sequence identities at the amino acid level, respectively. The revised protein sequence is supported by a multiple sequence alignment with AndM, AusA, and MpaC (Figure S10). These PKSs are involved in the biosynthesis of meroterpenoids anditomin, austinol, and mycophenolic acid. Thereby AndM and AusA catalyze the formation of 3,5-dimethylorsellinic acid (DMOA) and MpaC that of 5-methylorsellinic acid.[33−35]

Cloning the genomic sequence of CRL19966 into pFK20 resulted in the expression construct pFK21. Eleven of the 160 resistant pFK21 transformants showed the albino phenotype (e.g., strain FK11 in Figure 1, Tables 1 and 2), indicating successful integration into the *pcr4401* locus, which was subsequently confirmed by PCR (Figure 2C,D). The strain FK11 was cultivated in CDB for 3 days and extracted with EtOAc for LC−MS analysis. In comparison to that of the Δpcr4401 strain FK02, one predominant peak with a [M + H]+ ion at m/z 197.0802 was detected at 7.6 min (Figure 4A,B). Isolation and structural elucidation by UV and NMR analyses (Figures S2C and S7, Table S5) proved its structure as DMOA.[33] The product yield was calculated to be 11.2 mg per liter of culture. This small aromatic polyketide is a common intermediate in the biosynthesis of diverse meroterpenoids. In contrast to the BGCs of the PKS AndM, AusA, and MpaC with a farnesyltransferase gene, no homologous genes could be identified in the proximity of CRL19966. Until now, there has been no report on DMOA related meroterpenoids or other DMOA based SMs in *P. camemberti*.

The PKS Gm1.7502_g from *P. brevicompactum* shows a SAT-KS-AT-PT-ACP-ACP-TE domain architecture (Figure 4C). UniProt BLAST analysis revealed sequence identities of 43.2, 42.3, and 42.1% to CTB1 from *Cercospora nicotianae*,

Fsr1 from *Fusarium fujikuroi*, and wA from *A. nidulans* in the aforementioned DHN-melanin biosynthesis.[25] CTB1 catalyzes the formation of nor-toralactone, which is further metabolized to cercosporin.[36] Fsr1 is responsible for the formation of 6-*O*-demethylfusarubinaldehyde, a precursor of fusarubins.[37] In analogy to CRL19966, the sequence of Gm1.7502_g was amplified from the genomic DNA of *P. brevicompactum* and integrated into the *pcr4401* locus of *P. crustosum* by transformation of the construct pFK22 resulting in the strain FK12 (Figure 2E,F). A total of 11 of the 144 resistant transformants showed the albino phenotype (Figure 1, Table 1) and were subsequently cultivated in CDB for 4 days. The SMs were then extracted and analyzed *via* LC−MS (Figure 4C). Again, one dominant product peak at 9.2 min with a [M − H]⁻ ion at m/z 257.0451 was detected, which is absent in the control strain (Figure 4A). Isolation and structural elucidation with the help of UV and ¹H NMR spectrum (Figures S2D and S8, Table S5) led to the identification of nor-toralactone (Figure 4C).[38] The product yield was assumed to be 6.5 mg per liter of culture. To the best of our knowledge, there are no reports on the production of nor-toralactone or related SMs in *P. brevicompactum*.

Heterologous Expression in a *pyrG* Deficient *P. crustosum* Strain. Heterologous overexpression of the genes coding for CRL19966 and Gm1.7502_g and the subsequent isolation of the accumulated compounds proved that *P. crustosum* PRB-2 is a suitable fast-growing host for heterologous overexpression with reliable genetic tractability (Tables 1 and 2). Furthermore, heterologous expression of both PKS genes did not affect the growing behavior of *P. crustosum* (Table S3). We manipulated therefore *P. crustosum* for more convenience as a heterologous expression host by identification and deletion of the *pyrG* gene (*pcr6370*) encoding the orotidine 5′-phosphate decarboxylase. This deletion provides the possibility to use *pyrG* as a second selection marker besides the hygromycin B resistance gene. Furthermore, using *pyrG* as selection marker enables multiple gene deletion or gene insertions by selection marker recycling based on the bidirectional selection of *pyrG* deficient or *pyrG* intact transformants. For gene deletion, the flanking regions of the *pyrG* gene were fused and cloned into the vector pGEM-T Easy to give the plasmid pFK25. Afterward, the fused flanking regions were amplified from pFK25 *via* PCR and transformed into *P. crustosum* (Figure 2G). Selection of *pyrG* deficient mutants was achieved by supplementing 5-fluoroorotic acid (5-FOA), uracil, and uridine. Success of the deletion was additionally proven by PCR (Figure 2H) and by the uracil and uridine auxotrophic selection of the transformants.

To prove the functionality of the *pyrG* selection marker, we constructed a second plasmid pFK23 (Figure S4) for heterologous expression based on the *pyrG* selection marker from *A. fumigatus* (*afpyrG*), and the *hph* resistance cassette in pFK20 was replaced by the *afpyrG* selection marker. Subsequently, a putative 6-methylsalicylic acid synthase (*EURHEDRAFT_547227*) from *Aspergillus ruber* was cloned into pFK23 and used for integration into the *pcr4401* locus of *P. crustosum*. Successful integration was subsequently confirmed *via* PCR (Figure 2I,J). A total of 5 of the 72 transformants showed the albino phenotype (Tables 1 and 2, FK13 in Figure 1). Transformant FK13 was selected for cultivation in CDB and SM screening. Growth rate of the *pyrG* deficient strain FK15 was reduced compared to the wildtype or FK02 strain. However, complementation of the *pyrG*

deficiency through transformation of the heterologous expression construct restored the growth rate in strain FK13 back to the wildtype level (Table S3). LC−MS analysis of an EtOAc extract of a 3 day old culture showed a major product peak at 7.4 min with a [M − H]⁻ ion at m/z 151.0408 (Figure 4D). Isolation and structural elucidation with the help of ultraviolet (UV) spectroscopy and ¹H NMR spectroscopy (Figures S2E and S9, Table S5) confirmed its integrity as the expected 6-methylsalicylic acid.[39] The product yield was estimated to be 2.1 mg per liter of culture. Heterologous expression of the same 6-methylsalicylic acid synthase gene in *A. nidulans* (Figure 1 and Table 1) under the identical conditions resulted in comparable product yields (Figure 4E). This proved the functionality of the *afpyrG* as selection marker for heterologous gene expression in *P. crustosum* and the convenient identification of desired transformants by insertion into the *pcr4401* locus.

In conclusion, we demonstrated that the *pcr4401* gene locus of *P. crustosum* is a suitable integration site for heterologous expression of fungal genes and shows a good and reliable genetic tractability. Successful transformation of *P. crustosum* resulted, on average, in more than 18 transformants per microgram of transformed DNA. By using the *pcr4401* locus for integration, selection of transformants with correct integration is easily achieved by the albino phenotype. Three independent experiments proved the gene targeting efficiency for deletion or heterologous expression in the range of 5−11% (Table 2). Together with both of the plasmids pFK20 and pFK23 for site-specific expression in *P. crustosum* PRB-2 and the created *pyrG* deficient strain, we expand herewith the spectrum of fungal expression hosts.

■ METHODS

Strains, Media, and Culture Conditions. *Escherichia coli* strains XL1 Blue MRF′ (Stratagene) and DH5α (Invitrogen) were used for cloning and plasmid propagation. They were cultivated in lysogeny broth (LB) at 37 °C. For cultivation on solid LB medium, 1.5% (w/v) agar was added. For selection of the recombinant *E. coli* strains, 50 μg mL⁻¹ carbenicillin was used.

Fungal strains used and created in this study are listed in Table 1. *P. crustosum* PRB-2[27] was cultivated under static conditions at 25 °C in glucose minimal medium (GMM),[40] potato dextrose broth (PDB, Sigma-Aldrich), or Czapek dox broth (CDB, BD). For cultivation on solid medium, 1.5% (w/v) agar was added to the respective medium.

For selection of *P. crustosum* PRB-2 Δ*pyrG* mutants, GMM supplemented with 2.05 mM uridine, 4.46 mM uracil, and 7.5 mM 5-FOA at pH 4 was used. *P. crustosum* PRB-2 Δ*pyrG* was cultivated in GMM supplemented with uridine and uracil.

A. nidulans LO8030 was cultivated under static conditions at 37 °C in GMM supplemented with riboflavin, pyridoxine, uridine, and uracil as reported previously.[24] Cultivation of *A. nidulans* strains FK10 and FK14 was done on rice supplemented with 6.6 μM riboflavin and 3 μM pyridoxine at 25 °C. Strain JN03 was cultivated in CDB medium supplemented with uridine, uracil, and pyridoxine.

Isolation of Genomic DNA. For genomic DNA isolation, *P. crustosum*, *Penicillium camemberti* (kindly provided by Joelle Dupont, the National Museum of Natural History), *Penicillium brevicompactum* (provided by ARS Culture Collection), *A. ruber* (provided by Bingui Wang, Institute of Oceanology, Chinese Academy of Sciences), and *A. nidulans* were cultivated in PDB for 2 days at 25 °C. Genomic DNA was isolated according to the method described previously.[21]

Genome Sequencing and Sequence Analysis. The draft genome sequence was obtained from BerryGenomics by using Nova-seq6000/X-ten (Illumina). For prediction of open reading frames and protein sequences, secondFind (http://biosyn.nih.go.jp/2ndfind/) and FGENESH[19] were used. Nucleotide sequences and predicted

proteins reported in this study are available at GenBank under accession numbers MK575554 (*pcr4401*), MK575555 (*pcr4448–pcr4453*), MK575556 (*pcr8116*), and MK575557 (*pcr6370*).

The nucleotide and protein sequences of the investigated polyketide synthases from *P. camemberti* and *A. ruber* originate from NCBI database. By using secondFind, the coding sequence of CRL19966 from *P. camemberti* was revised to have five exons (bp 1803195−1803854, 1803909−1805009, 1805071−1806264, 1806326−1810358, and 1810441−1811363 of HG793136.1) with a deduced product of 2636 amino acids. The data of the *P. brevicompactum* PKS Gm1.7502_g were obtained from the Joint Genome Institute (JGI) (https://genome.jgi.doe.gov/Penbr2/Penbr2.home.html). The coding gene is located between bp 2335606−2342342 bp at scaffold 4 with the protein ID 60339 and the transcript ID 60527.

Gene Cloning, Plasmid Construction, and Genetic Manipulation. The plasmids generated and used in this study are listed in Table S1. Oligonucleotide primers used for construction of deletion and heterologous expression vectors and transformant verification are listed in Table S2. Phusion high-fidelity DNA polymerase (New England Biolabs) was used for amplifications on a T100 thermal cycler from Bio-Rad.

Construction of the deletion and heterologous expression vectors was done by ligation-independent homologous recombination in *E. coli* or *S. cerevisiae*.[41,42] Briefly, vector backbones were linearized *via* PCR with subsequent DpnI digestion or *via* restriction digestion with subsequent purification *via* a HiYield PCR Clean-up and Gel-Extraction Kit (SLG Südlaborbedarf). Primers with overlapping homologous parts (20−25 bp) were used for DNA fragment amplification. Subsequently the fragments with homologous overlaps were mixed and transformed into *E. coli* DH5α or *S. cerevisiae* HOD114-2B. The recombinant plasmids were verified by restriction analysis.

For the *pcr4401* deletion construct pFK6, two 1.5 kbp fragments of its up- and downstream regions were amplified from *P. crustosum* genomic DNA by using primers given in Table S2 with overlapping to pGEM-T Easy vector and the *hph* cassette. The pGEM-T Easy vector was linearized *via* PCR, and the remaining template was destroyed by DpnI digestion. The *hph* cassette was amplified from plasmid pUChph *via* PCR, followed by DpnI treatment. A *pcr4401* deletion fragment of 5916 bp was PCR amplified from pFK6 and subsequently transformed into *P. crustosum*.

Construct pFK19 for heterologous expression of *pcr4401* in *A. nidulans* LO8030 is a derivative of the vector pYH-wA-pyrG.[23] The vector was linearized by restriction with NheI followed by SAP treatment. *Pcr4401* was amplified from genomic DNA of *P. crustosum* PRB-2 (Table S2) as a full-length fragment with an additional 549 bp at the 3′ end to include the native terminator sequence and homologous overlaps to pYH-wA-pyrG. Assembled plasmid pFK19 was linearized with SwaI prior to transformation.

For the *P. crustosum* PRB-2 *pyrG* (Δ*pcr6370*) deletion construct pFK25, two 1.5 kbp fragments of up- and downstream regions were amplified, mixed with PCR linearized pGEM-T Easy vector, and transformed into *E. coli* DH5α. Using the obtained plasmid as a template, the *pyrG* deletion cassette was amplified by PCR.

The vectors pFK20 and pFK23 for heterologous gene expression in *P. crustosum* with *hph* or *afpyrG* as a selection marker are based on the plasmid pYH-wA-pyrG. For targeting the *pcr4401* locus as an integration site, the 5′ and 3′ *wA* flanking regions from pYH-wA-pyrG were replaced with 1 kbp flanking regions of the *pcr4401* gene. For creation of pFK23, the up- and downstream regions of *pcr4401*, the *pyrG* selection marker with the *gpdA* promoter, and the *E. coli* ori/ampR together with the ScURA-CEN/ARS were amplified by PCR with the primers given in Table S2. For creation of pFK20, the same up- and downstream flanking regions of *pcr4401* were amplified by PCR. Additionally, the *gpdA* promoter without the *pyrG* selection marker and the *E. coli* ori/ampR together with the ScURA-CEN/ARS were amplified from the vector pYH-wA-pyrG. The hygromycin B resistance cassette originated from pUChph.

The vectors pFK20 and pFK23 were linearized *via* NheI digestion, and the respective PKS genes were amplified by PCR from genomic DNA with an additional 0.5 kbp at the 3′ ends. Corresponding linearized vector and full-length PCR fragments of *PCAMFM013_S003g000758*, *gm1.7502_g*, and *EURHE-DRAFT_547227* were transformed into *E. coli* DH5α to give the plasmids pFK21, pFK22, and pFK24, which were verified by restriction analysis. The correct plasmids were linearized *via* SwaI digestion prior to transformation.

The vector pYWB2 was cut with BamHI and NheI and used for cloning the amplified *afribo* fragment coding for the riboflavin biosynthetic gene and *gpdA* promoter *via* homologous recombination in *S. cerevisiae*. The resulting vector pJN017 was then cut with SfoI and transformed together with the PCR product of *EURHE-DRAFT_547227* into *S. cerevisiae* to create pJN015.

Transformation of *P. crustosum* PRB-2 and *A. nidulans* LO8030 was achieved by polyethylene glycol (PEG) mediated protoplast transformation according to the protocol of Yin *et al.*[23] with modifications for *P. crustosum* PRB-2. Germination time and temperature were adjusted to 12 h at 25 °C. For protoplastation, the mycelia were incubated with 50 mg of lysing enzymes from *Trichoderma harzianum* (Sigma-Aldrich) and 20 mg of yatalase from *Corynebacterium sp.* OZ-21 (OZEKI Co., Ltd.) per 10 mL of osmotic medium at 30 °C for 2 h. Resistant transformants were selected by using 200 μg of hygromycin B per mL bottom GMM and 100 μg per mL top GMM.

Cultivation and Product Extraction and Isolation. For heterologous expression of *pcr4401* in *A. nidulans* LO8030, the strain FK10 was cultivated on rice medium supplemented with riboflavin and pyridoxine at 25 °C for 7 days. SMs of the cultures were extracted with EtOAc three times. Organic phases were combined and dried by using a rotary evaporator and dissolved in methanol for HPLC and LC−MS analyses. Heterologous expression in *P. crustosum* PRB-2 was done in CDB as static cultures for 3−7 days at 25 °C. SMs were extracted and analyzed as mentioned above.

YWA1 (**1**) and nor-rubrofusarin (**2**) were isolated from the crude extract of FK10 on an Agilent HPLC series 1200 (Agilent Technologies) with a semipreparative ZORBAX Eclipse XDB-C18 column (9.4 mm × 250 mm, 5 μm) (Agilent). Water (A) and acetonitrile (B), both with 0.05% (v/v) trifluoroacetic acid, were used as solvents. Elution of the substances was achieved isocratically with 45% (v/v) B in A at a flow rate of 4.0 mL/min. This column was also used for 6-methylsalicylic acid (**5**) isolation from the FK13 extract with a linear gradient from 10% to 55% B in 20 min at a flow of 3.0 mL/min.

3,5-Dimethylorsellinic acid (**3**) was isolated from strain FK11. Crude extract was subjected to silica gel column chromatography (silica gel 60, 0.04−0.063 mm) with a stepwise elution with CH_2Cl_2/MeOH (100:0 to 70:30, v/v). A total of 13 fractions were achieved and analyzed on LC−MS. Two fractions were further purified on the HPLC mentioned above by using a Multospher 120 RP18 HP column (10 mm × 250 mm, 5 μm) (CS-Chromatographie Service GmbH) and with 40% solvent B in solvent A at a flow of 4.0 mL/min. This column was also used for nor-toralactone (**4**) isolation from strain FK12, and 50% solvent B in solvent A with a flow rate of 5.0 mL/min was applied.

6-Methylsalicylic acid (**5**) was isolated from the extract of strain FK13 by semipreparative HPLC. Water (A) and acetonitrile (B), both with 0.02% (v/v) trifluoroacetic acid were used as solvents at a flow of 3.0 mL/min. Separation was done with a linear gradient from 10% to 55% B in 20 min. Afterward, the column was washed with 100% B for 5 min and subsequently equilibrated with 10% B for 5 min.

LC−MS and NMR Analyses. For LC−MS analysis, 5 μL samples were injected into an Agilent 1260 series HPLC system (Agilent) combined with a Bruker microTOF QIII mass spectrometer under conditions described previously.[43] ^{1}H NMR spectra were taken on a JEOL ECA-500 MHz spectrometer (JEOL) and processed with the software MestReNova 6.1.0 (Metrelab). Signals of the solvents were used as reference of the chemical shifts.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: shuming.li@staff.uni-marburg.de.

ORCID ⓘ
Anke Becker: 0000-0003-4561-9184
Shu-Ming Li: 0000-0003-4583-2655

Author Contributions
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes
The authors declare no competing financial interest.

■ ABBREVIATIONS

DHN , 1,8-dihydroxynaphthalene
PKS , polyketide synthase
NP , natural product
SM , secondary metabolite
BGC , biosynthetic gene cluster
PKS-NRPS , polyketide synthase-nonribosomal peptide synthetase
NRPS , nonribosomal peptide synthetase
NR-PKS , nonreducing-polyketide synthase
DMOA , 3,5-dimethylorsellinic acid
6-MSA , 6-methylsalicylic acid
afpyrG , *A. fumigatus pyrG*
CBD , Czapek dox broth
PDB , potato dextrose broth
5-FOA , 5-fluoroorotic acid

■ REFERENCES

(1) Newman, D. J., and Cragg, G. M. (2016) Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod. 79*, 629−661.
(2) Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H. P., Clément, C., Ouhdouch, Y., and Van Wezel, G. P. (2016) Taxonomy, physiology, and natural products of actinobacteria. *Microbiol. Mol. Biol. Rev. 80*, 1−43.
(3) Bills, G. F., and Gloer, J. B. (2016) Biologically active secondary metabolites from the fungi. *Microbiol Spectr. 4*, 1−32.
(4) Bérdy, J. (2012) Thoughts and facts about antibiotics: Where we are now and where we are heading. *J. Antibiot. 65*, 441.
(5) Land, M., Hauser, L., Jun, S. R., Nookaew, I., Leuze, M. R., Ahn, T. H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S.,

and Ussery, D. W. (2015) Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics 15*, 141−161.
(6) Baltz, R. H. (2017) Gifted microbes for genome mining and natural product discovery. *J. Ind. Microbiol. Biotechnol. 44*, 573−588.
(7) Gacek, A., and Strauss, J. (2012) The chromatin code of fungal secondary metabolite gene clusters. *Appl. Microbiol. Biotechnol. 95*, 1389−1404.
(8) He, Y., Wang, B., Chen, W., Cox, R. J., He, J., and Chen, F. (2018) Recent advances in reconstructing microbial secondary metabolites biosynthesis in *Aspergillus* spp. *Biotechnol. Adv. 36*, 739−783.
(9) Anyaogu, D. C., and Mortensen, U. H. (2015) Heterologous production of fungal secondary metabolites in Aspergilli. *Front. Microbiol. 6*, 77.
(10) Lubertozzi, D., and Keasling, J. D. (2009) Developing *Aspergillus* as a host for heterologous expression. *Biotechnol. Adv. 27*, 53−75.
(11) Guzman-Chavez, F., Zwahlen, R. D., Bovenberg, R. A. L., and Driessen, A. J. M. (2018) Engineering of the filamentous fungus *Penicillium chrysogenum* as cell factory for natural products. *Front. Microbiol. 9*, 2768.
(12) Song, Z., Bakeer, W., Marshall, J. W., Yakasai, A. A., Khalid, R. M., Collemare, J., Skellam, E., Tharreau, D., Lebrun, M. H., Lazarus, C. M., Bailey, A. M., Simpson, T. J., and Cox, R. J. (2015) Heterologous expression of the avirulence gene *ACE1* from the fungal rice pathogen *Magnaporthe oryzae*. *Chem. Sci. 6*, 4837−4845.
(13) Geib, E., Baldeweg, F., Doerfer, M., Nett, M., and Brock, M. (2019) Cross-chemistry leads to product diversity from atromentin synthetases in Aspergilli from section *Nigri*. *Cell Chem. Biol. 26*, 223−234.
(14) Yadav, A. N., Verma, P., Kumar, V., Sangwan, P., Mishra, S., Panjiar, N., Gupta, V. K., and Saxena, A. K. (2018) in *New and Future Developments in Microbial Biotechnology and Bioengineering* (Gupta, V. K., and Rodriguez-Couto, S., Eds.) pp 3−18, Elsevier, Amsterdam, The Netherlands.
(15) Camardo Leggieri, M., Decontardi, S., Bertuzzi, T., Pietri, A., and Battilani, P. (2017) Modeling growth and toxin production of toxigenic fungi signaled in cheese under different temperature and water activity regimes. *Toxins 9*, 4.
(16) Sonjak, S., Frisvad, J. C., and Gunde-Cimerman, N. (2005) Comparison of secondary metabolite production by *Penicillium crustosum* strains, isolated from Arctic and other various ecological niches. *FEMS Microbiol. Ecol. 53*, 51−60.
(17) Fan, J., Liao, G., Kindinger, F., Ludwig-Radtke, L., Yin, W.-B., and Li, S.-M. (2019) Peniphenone and penilactone formation in *Penicillium crustosum via* 1,4-Michael additions of *ortho*-quinone methide from hydroxyclavatol to γ-butyrolactones from crustosic acid. *J. Am. Chem. Soc. 141*, 4225−4229.
(18) Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015) antiSMASH 3.0 - a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res. 43*, W237−W243.
(19) Salamov, A. A., and Solovyev, V. V. (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res. 10*, 516−522.
(20) Chooi, Y. H., and Tang, Y. (2012) Navigating the fungal polyketide chemical space: from genes to molecules. *J. Org. Chem. 77*, 9933−9953.
(21) Goswami, R. S. (2012) Targeted gene replacement in fungi using a split-marker approach. *Methods Mol. Biol. 835*, 255−269.
(22) Haarmann, T., Lorenz, N., and Tudzynski, P. (2008) Use of a nonhomologous end joining deficient strain (Deltaku70) of the ergot fungus *Claviceps purpurea* for identification of a nonribosomal peptide synthetase gene involved in ergotamine biosynthesis. *Fungal Genet. Biol. 45*, 35−44.
(23) Yin, W. B., Chooi, Y. H., Smith, A. R., Cacho, R. A., Hu, Y., White, T. C., and Tang, Y. (2013) Discovery of cryptic polyketide metabolites from dermatophytes using heterologous expression in *Aspergillus nidulans*. *ACS Synth. Biol. 2*, 629−634.

(24) Chiang, Y. M., Ahuja, M., Oakley, C. E., Entwistle, R., Asokan, A., Zutz, C., Wang, C. C., and Oakley, B. R. (2016) Development of genetic dereplication strains in *Aspergillus nidulans* results in the discovery of aspercryptin. *Angew. Chem., Int. Ed.* 55, 1662−1665.

(25) Watanabe, A., Fujii, I., Sankawa, U., Mayorga, M. E., Timberlake, W. E., and Ebizuka, Y. (1999) Re-identification of *Aspergillus nidulans* wA gene to code for a polyketide synthase of naphthopyrone. *Tetrahedron Lett.* 40, 91−94.

(26) Frandsen, R. J., Nielsen, N. J., Maolanon, N., Sorensen, J. C., Olsson, S., Nielsen, J., and Giese, H. (2006) The biosynthetic pathway for aurofusarin in *Fusarium graminearum* reveals a close link between the naphthoquinones and naphthopyrones. *Mol. Microbiol.* 61, 1069−1080.

(27) Wu, G., Ma, H., Zhu, T., Li, J., Gu, Q., and Li, D. (2012) Penilactones A and B, two novel polyketides from Antarctic deep-sea derived fungus *Penicillium crustosum* PRB-2. *Tetrahedron* 68, 9745−9749.

(28) Cheeseman, K., Ropars, J., Renault, P., Dupont, J., Gouzy, J., Branca, A., Abraham, A. L., Ceppi, M., Conseiller, E., Debuchy, R., Malagnac, F., Goarin, A., Silar, P., Lacoste, S., Sallet, E., Bensimon, A., Giraud, T., and Brygoo, Y. (2014) Multiple recent horizontal transfers of a large genomic region in cheese making fungi. *Nat. Commun.* 5, 2876.

(29) Li, D.-L., Li, X.-M., Li, T.-G., Dang, H.-Y., and Wang, B.-G. (2008) Dioxopiperazine alkaloids produced by the marine mangrove derived endophytic fungus *Eurotium rubrum*. *Helv. Chim. Acta* 91, 1888−1892.

(30) Guzman-Chavez, F., Salo, O., Samol, M., Ries, M., Kuipers, J., Bovenberg, R. A. L., Vreeken, R. J., and Driessen, A. J. M. (2018) Deregulation of secondary metabolism in a histone deacetylase mutant of *Penicillium chrysogenum*. *MicrobiologyOpen* 7, No. e00598.

(31) Watanabe, A., Fujii, I., Tsai, H., Chang, Y. C., Kwon-Chung, K. J., and Ebizuka, Y. (2000) *Aspergillus fumigatus alb1* encodes naphthopyrone synthase when expressed in *Aspergillus oryzae*. *FEMS Microbiol. Lett.* 192, 39−44.

(32) The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45, D158−D169.

(33) Matsuda, Y., Wakimoto, T., Mori, T., Awakawa, T., and Abe, I. (2014) Complete biosynthetic pathway of anditomin: nature's sophisticated synthetic route to a complex fungal meroterpenoid. *J. Am. Chem. Soc.* 136, 15326−15336.

(34) Lo, H. C., Entwistle, R., Guo, C. J., Ahuja, M., Szewczyk, E., Hung, J. H., Chiang, Y. M., Oakley, B. R., and Wang, C. C. (2012) Two separate gene clusters encode the biosynthetic pathway for the meroterpenoids austinol and dehydroaustinol in *Aspergillus nidulans*. *J. Am. Chem. Soc.* 134, 4709−4720.

(35) Regueira, T. B., Kildegaard, K. R., Hansen, B. G., Mortensen, U. H., Hertweck, C., and Nielsen, J. (2011) Molecular basis for mycophenolic acid biosynthesis in *Penicillium brevicompactum*. *Appl. Environ. Microbiol.* 77, 3035−3043.

(36) Newman, A. G., Vagstad, A. L., Belecki, K., Scheerer, J. R., and Townsend, C. A. (2012) Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem. Commun. (Cambridge, U. K.)* 48, 11772−11774.

(37) Studt, L., Wiemann, P., Kleigrewe, K., Humpf, H. U., and Tudzynski, B. (2012) Biosynthesis of fusarubins accounts for pigmentation of *Fusarium fujikuroi* perithecia. *Appl. Environ. Microbiol.* 78, 4468−4480.

(38) Newman, A. G., and Townsend, C. A. (2016) Molecular characterization of the cercosporin biosynthetic pathway in the fungal plant pathogen *Cercospora nicotianae*. *J. Am. Chem. Soc.* 138, 4219−4228.

(39) Bedford, D. J., Schweizer, E., Hopwood, D. A., and Khosla, C. (1995) Expression of a functional fungal polyketide synthase in the bacterium *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 177, 4544−4548.

(40) Barratt, R. W., Johnson, G. B., and Ogata, W. N. (1965) Wild-type and mutant stocks of *Aspergillus nidulans*. *Genetics* 52, 233−246.

(41) Jacobus, A. P., and Gross, J. (2015) Optimal cloning of PCR fragments by homologous recombination in *Escherichia coli*. *PLoS One* 10, No. e0119221.

(42) Oldenburg, K. R., Vo, K. T., Michaelis, S., and Paddon, C. (1997) Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast. *Nucleic Acids Res.* 25, 451−452.

(43) Ran, H., Wohlgemuth, V., Xie, X., and Li, S.-M. (2018) A non-heme FeII/2-oxoglutarate-dependent oxygenase catalyzes a double bond migration within a dimethylallyl moiety accompanied by hydroxylation. *ACS Chem. Biol.* 13, 2949−2955.

SUPPORTING INFORMATION

Genomic locus of a *Penicillium crustosum* pigment as integration site for secondary metabolite gene expression

Florian Kindinger,[†] Jonas Nies,[†] Anke Becker,[§] Tianjiao Zhu,[‡] and Shu-Ming Li*,[†]

[†] Institut für Pharmazeutische Biologie und Biotechnologie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany

[§] LOEWE Zentrum für Synthetische Mikrobiologie, Philipps-Universität Marburg, Hans-Meerwein-Straße, 35032 Marburg, Germany

[‡] Ocean University of China, School of Medicine and Pharmacy, Key Laboratory of Marine Drugs, Chinese Ministry of Education Yushan Road 5, 266003, Qingdao, China

[*]To whom correspondence may be addressed. E-mail: shuming.li@staff.uni-marburg.de

Table of content

Supplementary Tables

Table S1: Plasmids created and used in this study

Plasmid	Description	Source
pUChph	hygromycin B resistance vector, *A. nidulans gpdA* promoter, *hph* gene, *A. nidulans trpC* terminator	[1]
pGEM-T easy	*ampR*, f1 ori, ColE1 ori	Promega
pYH-wA-pyrG	*URA3*, *wA* flanking, *A. nidulans gpdA* promoter, *A. fumigatus pyrG* (*afpyrG*), *ampR*	[2]
pESC-URA	*Saccharomyces cerevisiae* and *E. coli* shuttle vector	Agilent
pFK6	a 1500 bps PCR fragment of the upstream region of *pcr4401* from genomic DNA of *P. crustosum* PRB-2 was fused to the 2875 bps *hph* resistance cassette from pUChph, which is connected to the 1500 bps PCR fragment of the downstream region of *pcr4401* from *P. crustosum* PRB-2. The fused fragment was inserted into pGEM-T easy	This study
pFK19	a PCR fragment of 6660 bps containing *pcr4401* and its native terminator region (549 bps) was amplified from genomic DNA of *P. crustosum* PRB-2 and inserted into pYH-wA-pyrG	This study
pFK20	a 1014 bps PCR fragment of the upstream region of *pcr4401* from genomic DNA of *P. crustosum* PRB-2 was fused to the 3481 bps PCR fragment including the *E. coli* ori, *ampR,* and the *ScURA*-CEN/ARS from pYH-wA-pyrG, the 720 bps *gpdA* promoter originated from pYH-*wA-pyrG*, the 2875 bps *hph* resistance cassette from pUChph, and the 983 bps downstream region of *pcr4401* from genomic DNA of *P. crustosum* PRB-2.	This study
pFK21	a 8169 bps fragment of *PCAMFM013_S003g000758* coding for the PKS CRL19966 with native terminator region 502 (bps) was amplified from *P. camemberti* genomic DNA and inserted into pFK20	This study
pFK22	a 6737 bps fragment of *gm1.7502_g* coding for the PKS Gm1.7502_g with its native terminator region (526 bps) was amplified from *P. brevicompactum* genomic DNA and inserted into pFK20	This study
pFK23	a 1014 bps PCR fragment of the upstream region of *pcr4401* from genomic DNA *of P. crustosum* PRB-2 was fused to a 2401 bps PCR fragment including the *gpdA* promoter and the *afpyrG* marker from pYH-*wA-pyrG*, the 983 bps PCR fragment of the downstream region of *pcr4401* from genomic DNA of *P. crustosum* PRB-2, and the 3481 bps PCR fragment including the *E. coli* ori, *ampR* and the *ScURA*-CEN/ARS from pYH-*wA-pyrG*.	This study
pFK24	a 5596 bps fragment of *EURHEDRAFT_547227* coding for the PKS EYE91246 with its native terminator region (499 bps) was amplified from genomic DNA from *A. ruber* and inserted into pFK23	This study
pFK25	a 1500 bps PCR fragment of the upstream region of *pcr6370* was fused to the 1500 bps PCR fragment of the downstream region of *pcr6370* and inserted into pGEM-T easy	This study
pYWB2	*URA3*, *wA* flanking, *A. fumigatus riboB* (*afribo*), ampR	[3]
pJN017	a 1445 bps fragment of the *gpdA* promoter amplified from *A. nidulans* inserted into pYWB2.	This study
pJN015	a 6096 bps fragment containing *EURHEDRAFT_547227* coding for the PKS EYE91246 with its native terminator region (698 bps) was amplified from genomic DNA from *A. ruber* and inserted into pJN017	This study

Table S2: Primers used in this study

Primer	Oligonucleotide sequence (5'-3')	Uses
pGEM_linear1	ACTAGTGAATTCGCGGCCG	Linearization of pGEM-t easy vector via PCR
pGEM_linear2	AATCGAATTCCCGCGGCC	
HygB_pUChph_f	AGCTCTCCAAAGGGCG	Amplification of the *hph* gene under the control of the *A. nidulans gpdA* promoter and *trpC* terminator
HygB_pUChph_r	TCGCGTGGAGCCAAG	
PKS_pcr4401_U_f	CCATGGCGGCCGCGGGAATTCGATTCTCCGTGGATTTGGGATTT	Amplification of the upstream region of *pcr4401* from *P. crustosum*
PKS_pcr4401_U_r	TTGCATTGTGCAACGCCCTTTGGAGAGCTTATGGTAACCCAGCTGGAAC	
PKS_pcr4401_D_f	GGAATCCGCTCTTGGCTCCACGCGATTAAACAGCTAGCTGAAATCTAAC	Amplification of the downstream region of *pcr4401* from *P. crustosum*
PKS_pcr4401_D_r	AGGCGGCCGCGAATTCACTAGTGATCGGCTCAGAAACGCAC	
pcr4401_con_f	CCAGTGACACGATCTACCTC	Screening of Δ*pcr4401* transformants
pcr4401_con_r	GCTGCTGCAATCCATGTG	
Δpcr4401_con_f	GTCCTTCTACACGAGATGGC	Screening of Δ*pcr4401* transformants
Δpcr4401_con_r	GAACATATCGAGCGTTGTTTG	
hph_con_f	AGATACAGCTCATCTGCAATG	Screening of Δ*pcr4401* transformants
hph_con_r	ACAAATGTTGGACTGACGC	
NewPYRG_UP_F	CGGCCGCCATGGCGGCCGCGGGAATTCGATTGAATGATCGGCACGAAAG	Amplification of the upstream region of *pyrG* (*pcr6370*) from *P. crustosum*
NewPYRG_UP_R	ATAACAACCACTTCTTCACCCCGCCAACCCCACCAGCGTGCAGTCCACG	
NewPYRG_DOWN_F	TCATGTAACTACAGTGGAACGTGGACTGCACGCTGGTGGGGTTGGCGGG	Amplification of the downstream region of *pyrG* (*pcr6370*) from *P. crustosum*
NewPYRG_DOWN_R	TGCAGGCGGCCGCGAATTCACTAGTGATGGAAACTGCATCTTCTGGGTC	
PyrGcd_k_r	AGGGCCTTCAGAGCCTCATC	Screening of Δ*pcr6370* transformants
PyrG_k_gen_f	TCCAGAAACAGTACCACAACGGC	
PyrG_k_r_gen	TGGTACTGTTTCTGGACTGTGTTGC	Screening of Δ*pcr6370* transformants
PyrGcd_k_f	CTGGCAGTGTTCATTTGGCTTCTG	
Pcr4401-5F	AATCATGGTCATAGCTGTTTCCTGTGTGCTTGAGAACATTTGGGTCGAT	Amplification of the upstream region of *pcr4401* from *P. crustosum* for construction of the heterologous expression vectors pFK20 and pFK23 for *P. crustosum*
Pcr4401-5F-5R	CTCTGCGACCGTCCGTCTCTCCGCATGTATGGTAACCCAGCTGGAACCA	
Pcr4401-3F	GCCTCCTCTCAGACAGAATGCGGCCGCTAACATTGAACACCTCCCAGC	Amplification of the downstream region of *pcr4401* from *P. crustosum* for construction of the heterologous expression vector pFK23 for *P. crustosum*
Pcr4401-3R	CATGTTCTTTCCTGCGTTATCCCCTGGTGGTAGTTCTGCTGAACTGG	
Pcr4401-gpdA-F2	CTGTCTCCAGTTCAGCAGAACTACCACCAGGGGATAACGCAGGAAAG	Amplification of the *E. coli* ori/*ampR* and *ScURA*-CEN/ARS from pYH-*wA-pyrG* for vectors pFK20 and pFK23
Pcr4401-gpdA-R1	GACGATCGACCCAAATGTTCTCAAGCACACAGGAAACAGCTATGACC	

Pcr4401-gpdA-F1	CCCAATGGTTCCAGCTGGGTTACCATACATGCGGAGAGACGGACG	Amplification of af*pyrG* gene and *gpdA* promoter from pYH-*wA-pyrG* for vector pFK23
Pcr4401-gpdA-R2	CCAGGGCTGGGAGGTGTTCAATGTTAGCGGCCGCATTCTGTCTGAG	
Pcr4401-3R-HygB-F	AATCCGCTCTTGGCTCCACGCGACTAACATTGAACACCTCCCAGC	Amplification of the downstream region of *pcr4401* from *P. crustosum* for construction of heterologous expression vector pFK20 for *P. crustosum*
Pcr4401-3R	CATGTTCTTTCCTGCGTTATCCCCTGGTGGTAGTTCTGCTGAACTGG	
gpdA-HygB-R	CATCATAACATTCCACCTGGAACTGCTAGCGATTAAAGGTTCTTGGATG	Amplification of the *gpdA* promoter from pYH-*wA-pyrG* for vector pFK20
Pcr4401-gpdA-F1	CCCAATGGTTCCAGCTGGGTTACCATACATGCGGAGAGACGGACG	
HygB-gpdA-F	CCATCCAAGAACCTTTAATCGCTAGCAGTTCCAGGTGGAATGTTATG	Amplification of *hph* resistance cassette from pUChph for vector pFK20
Hyg-Pcr4401-R	GGCTTCCAGGGCTGGGAGGTGTTCAATGTTAGTCGCGTGGAGCCAAGAG	
CRL19966_F	TTCCATCCAAGAACCTTTAATCGCTAGATGCCAGCCCTTGTTAATATG	Amplification of the full length *PCAMFM013_S003g000758* including the native terminator region from *P. camemberti*
CRL19966_R	TCATAACATTCCACCTGGAACTGCTAGGAGAACGACTACTTATTCCCGG	
CRL19966_con_F	CGAAGAAAACAAGGTCAATGG	Screening of *PCAMFM013_S003g000758* transformants
CRL19966_con_R	TTCGTGGCATGGATACAATTC	
Gm17502_F	TTCCATCCAAGAACCTTTAATCGCTAGATGGCACGCGTTCAACAAATC	Amplification of the full length *gm1.7502_g* including its native terminator region from *P. brevicompactum* NRRL864
Gm17502_R	CATAACATTCCACCTGGAACTGCTAGCGGATTTGTCCAGCCACAGC	
Gm17502_con_F	AAAGACGTCGAATCTGTGG	Screening of *gm1.7502_g* transformants
Gm17502_con_R	GGTTATGACGAGAGAAAGGAG	
6-MSA-F	CCCATCCAAGAACCTTTAATCGCTAGATGTGTACCTCTACTCCTTCCAC	Amplification of the full length *EURHEDRAFT_547227* including its native terminator region from *A. ruber*
6-MSA-R	TCAGACACAGAATAACTCTCGCTAGGGCACGTGAGAAGTTTTTTTATTG	
6-MSAS_con_f	TGCTATTCTCGAAGAGTTCCC	Screening of *EURHEDRAFT_547227* transformants
6-MSAS_con_r	ACTGCATGAAGAATGTTCTTTATTGC	
Pcr4401_full_f	TATATTCATCTTCCCATCCAAGAACCTTTAATCATGGAAGGCCCCAGTC	Amplification of the full length *pcr4401* including its native terminator region from *P. crustosum*
Pcr4401_full_r	CAACATATTTCGTCAGACACAGAATAACTCTCCTGGCGTTCCATCCATC	
6-MSA_2-F	ATATTTTAATCCCATGTGGGCGCCCGGCACGTGAGAAGTTTTTTTATTG	Amplification of the full length *EURHEDRAFT_547227* including its native terminator region from *A. ruber*
6-MSA_2-R	GCTACCCCGCTTGAGCAGACATCACCGGCATGTGTACCTCTACTCCTTC	
Afumigatus_Ribo_f	CACATGGGATTAAAATATGG	Amplification of the *A. fumigatus afribo* selection marker
Afumigatus_Ribo_r	GACCGGGGTCTCTAG	
ANidulans_gpdA_f	GGACTTGACTCTCCTTCTCCTGATCGGATCCCCGATAGCTCTGCAAAGG	Amplification of the *A. nidulans gpdA* promoter
ANidulans_gpdA_r	GCCTCAACACCATATTTTAATCCCATGTGGGCGCCGGTGATGTCTGCTC	

Table S3: Colony size of *P. crustosum* strains grown on GMM plates. Values are means of three replicates.

Strain	Colony diameter in mm						
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
WT	0.8 ± 0.2	3.8 ± 0.6	9.5 ± 0.7	13.8 ± 0.6	18.0 ± 0.4	23.0 ± 0.8	26.8 ± 0.6
FK02	1.0 ± 0.0	4.3 ± 0.2	10.8 ± 0.2	14.7 ± 0.5	18.7 ± 0.5	23.8 ± 0.6	27.2 ± 0.2
FK11	1.0 ± 0.0	4.8 ± 0.2	11.5 ± 0.4	15.7 ± 0.5	19.5 ± 0.4	24.8 ± 0.2	28.7 ± 0.2
FK12	0.7 ± 0.2	4.2 ± 0.2	11.8 ± 0.2	16.7 ± 0.5	19.3 ± 0.5	24.3 ± 0.5	27.8 ± 0.2
FK13	1.0 ± 0.0	4.7 ± 0.5	12.2 ± 0.6	16.3 ± 0.9	19.0 ± 0.8	24.2 ± 1.0	28.2 ± 0.8
FK15	0.7 ± 0.2	4.3 ± 0.2	8.5 ± 0.4	11.2 ± 0.2	11.8 ± 0.2	15.5 ± 0.4	17.2 ± 0.2

Table S4: DHN-melanin biosynthetic candidates in *P. crustosum* and homologues from *A. fumigatus* and *P. rubens*.

Protein name in genome	Size (AA)	*A. fumigatus* homologue	Identity (%)	*P. rubens* homologue	Identity (%)	Putative function
Pcr4401	2139	EAL94057.1 (Alb1)	65.8	XP_002568608.1	93.9	polyketide synthase
Pcr4448	686	EAL94051.2 (Abr1)	48.6	XP_002568645.1	76.3	oxidase
Pcr4449	354	-	-	XP_002568646.1	58.8	hypothetical protein
Pcr4450	191	-	-	XP_002568647.1	12.5	hypothetical protein
-	764	-	-	XP_002568648.1	-	hypothetical protein
Pcr4451	171	EAL94055.1 (Arp1)	68.4	XP_002568649.1	81.9	scytalone dehydratase
Pcr4452	268	EAL94053.1 (Arp2)	69.9	XP_002568650.1	95.1	1,3,6,8-tetrahydroxynaphthalene reductase
Pcr4453	409	EAL94052.1 (Ayg1)	67.3	XP_002568651.1	91.4	polyketide shortening
Pcr8116	586	EAL94050.2 (Abr2)	64.4	XP_002564855.1	89.1	oxidase

Table S5: NMR data of the isolated compounds

YWA1 (1, acetone-D6)

Position	δ_H, multi., J in Hz
2-OH	6.06, br s, 1H
3	2.87, d, 17.1, 1H 3.18, d, 17.1, 1H
5-OH	15.37, br s, 1H
6-OH	9.44, br s, 1H
7	6.28, d, 2.2, 1H
8-OH	9.18, br s, 1H
9	6.52, d, 2.2, 1H
10	6.45, s, 1H
11	1.72, s, 3H

nor-rubrofusarin (2, acetone-D6)

Position	δ_H, multi., J in Hz
3	6.16, s
5-OH	-
6-OH	9.77, br s, 1H
7	6.40, d, 2.0, 1H
8-OH	9.29, br s, 1H
9	6.70, d, 2.0, 1H
10	7,01, s, 1H
11	2.44, s, 3H

DMOA (3, CD3OD)

Position	δ_H, multi., J in Hz
8	2.06, s, 3H
9	2.11, s, 3H
10	2.46, s, 3H

nor-toralactone (4, CDCl3)

Position	δ_H, multi., J in Hz
3-OH	13.60, s, 1H
4-OH	9.50, s, 1H
5	6.48, d, 2.4, 1H
6-OH	-
7	6.61, d, 2.4, 1H
8	6.92, s, 1H
9	6.23, s, 1H
11	2.27, s, 3H

6-methylsalicylic acid (5, CDCl3)

Position	δ_H, multi., J in Hz
3	6.79, d, 8.0, 1H
4	7.32, t, 7.9, 1H
5	6.77, d, 7.5, 1H
7	2.58, s, 3H
8-OH	11.81, br s, 1H

The ^{1}H NMR data of the isolated products correspond well to those reported previously, see [4] for **1**, [5] for **2**, [6] for **3**, [7] for **4**, and [8] for **5**

213
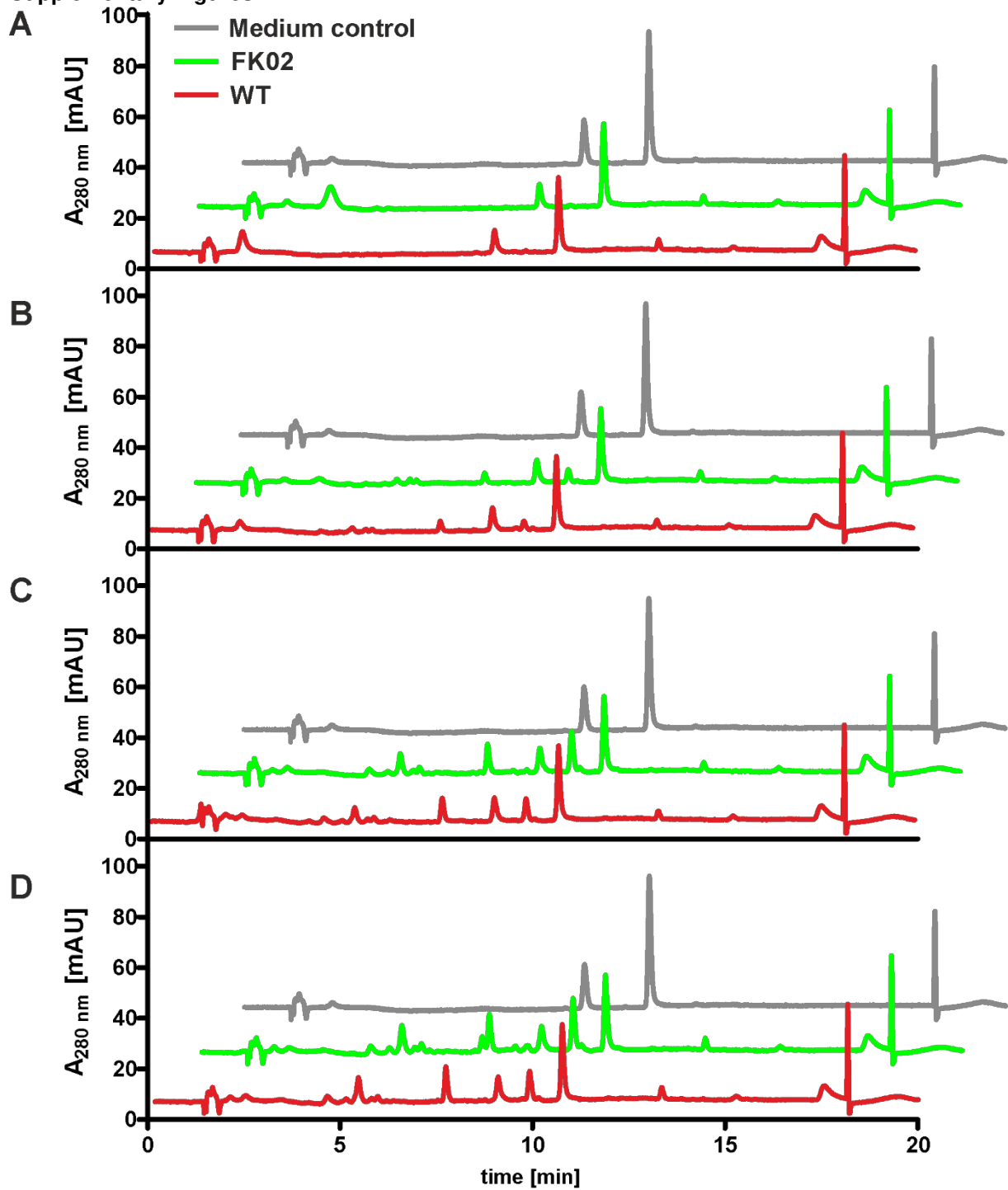
Figure S1: HPLC analysis of ethyl acetate extracts of CDB medium (negative control), *P. crustosum* WT, and FK02 mutant cultivated in CDB medium after two (A), four (B), six (C), and eight days (D).
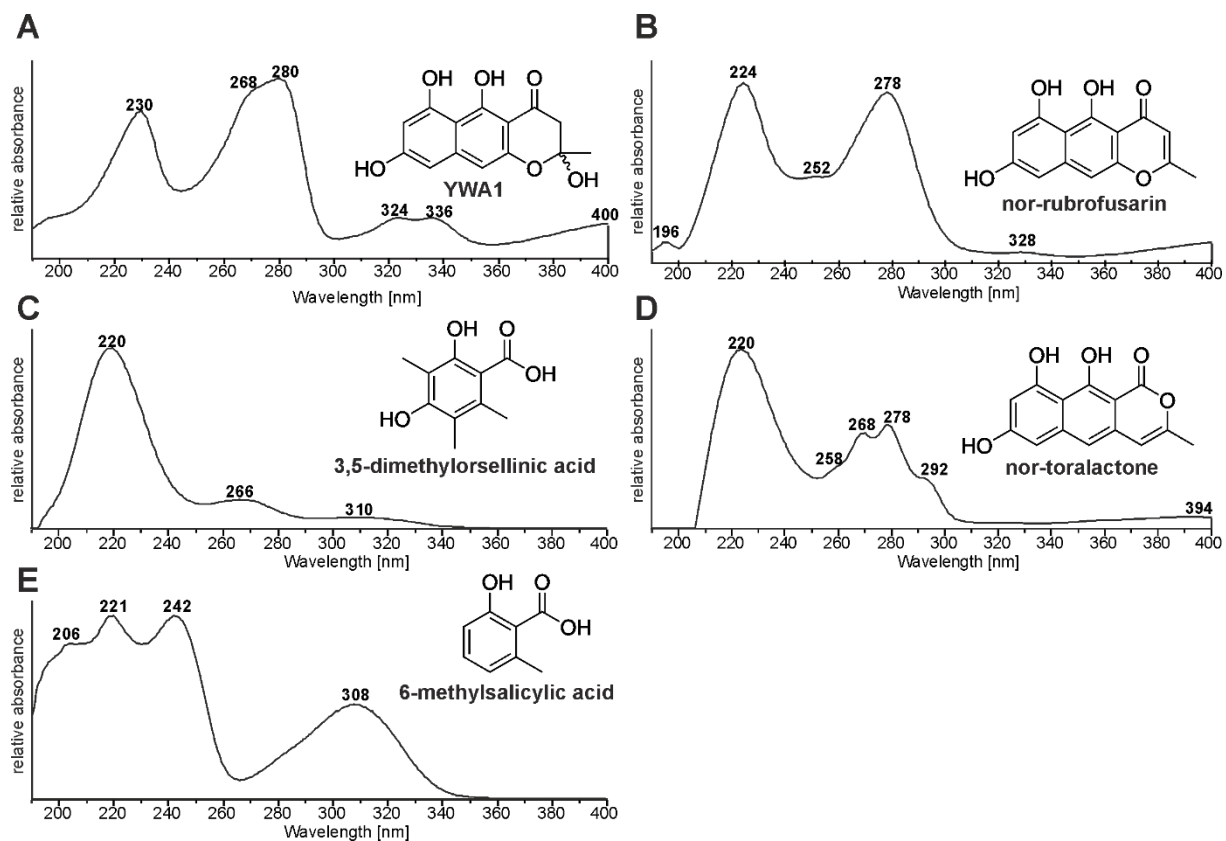
Figure S2: UV spectra from the isolated compounds. See reference [9] for YWA1, [9] for nor-rubrofusarin, [10] for 3,5-dimethylorsellinic acid (DMOA), [7] for nor-toralactone, and [11] for 6-methylsalicylic acid (6-MSA)
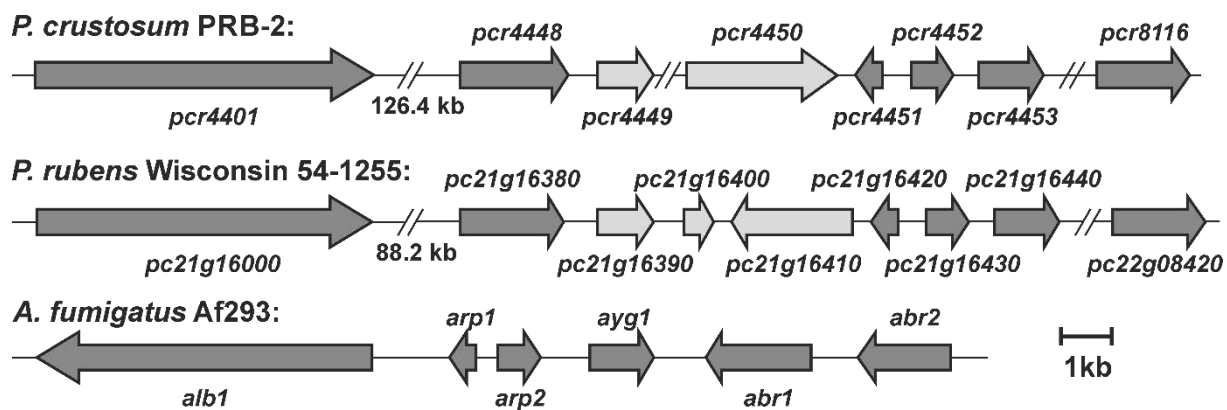
Figure S3: Schematic presentation of the DHN-melanin BGCs in *P. crustosum*, *P. rubens*, and *A. fumigatus*. In comparison to that in *A. fumigatus*, the BGCs in *P. crustosum* and *P. rubens* are partially divided
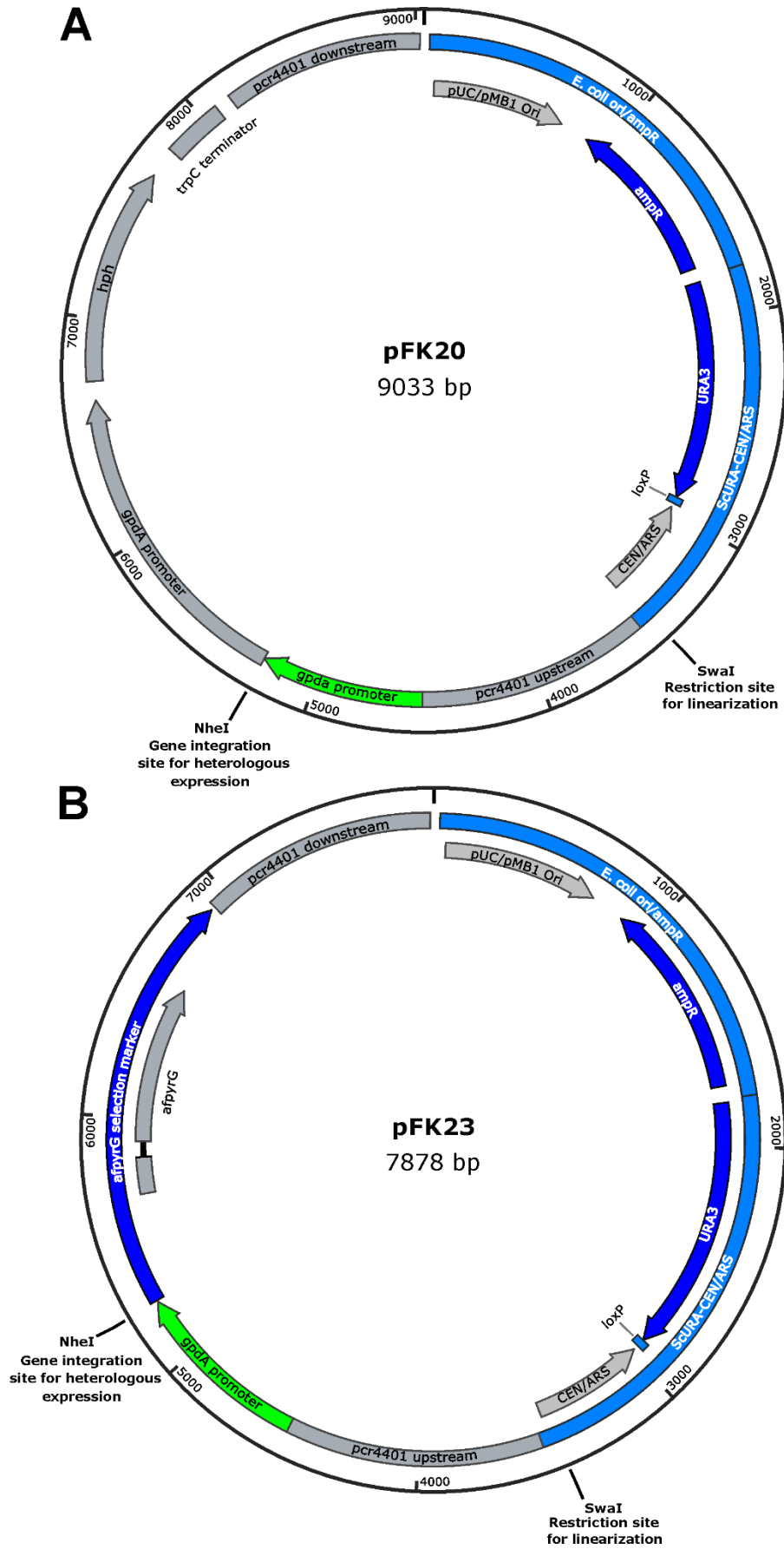
Figure S4: Schematic representation of the two plasmids pFK20 (**A**) and pFK23 (**B**) for site-specific integration of genes to be expressed at the *pcr4401* gene locus**.**

S12

Figure S5: ^{1}H NMR spectrum of YWA1 in aceton-d_6 (500 MHz)

S13

218
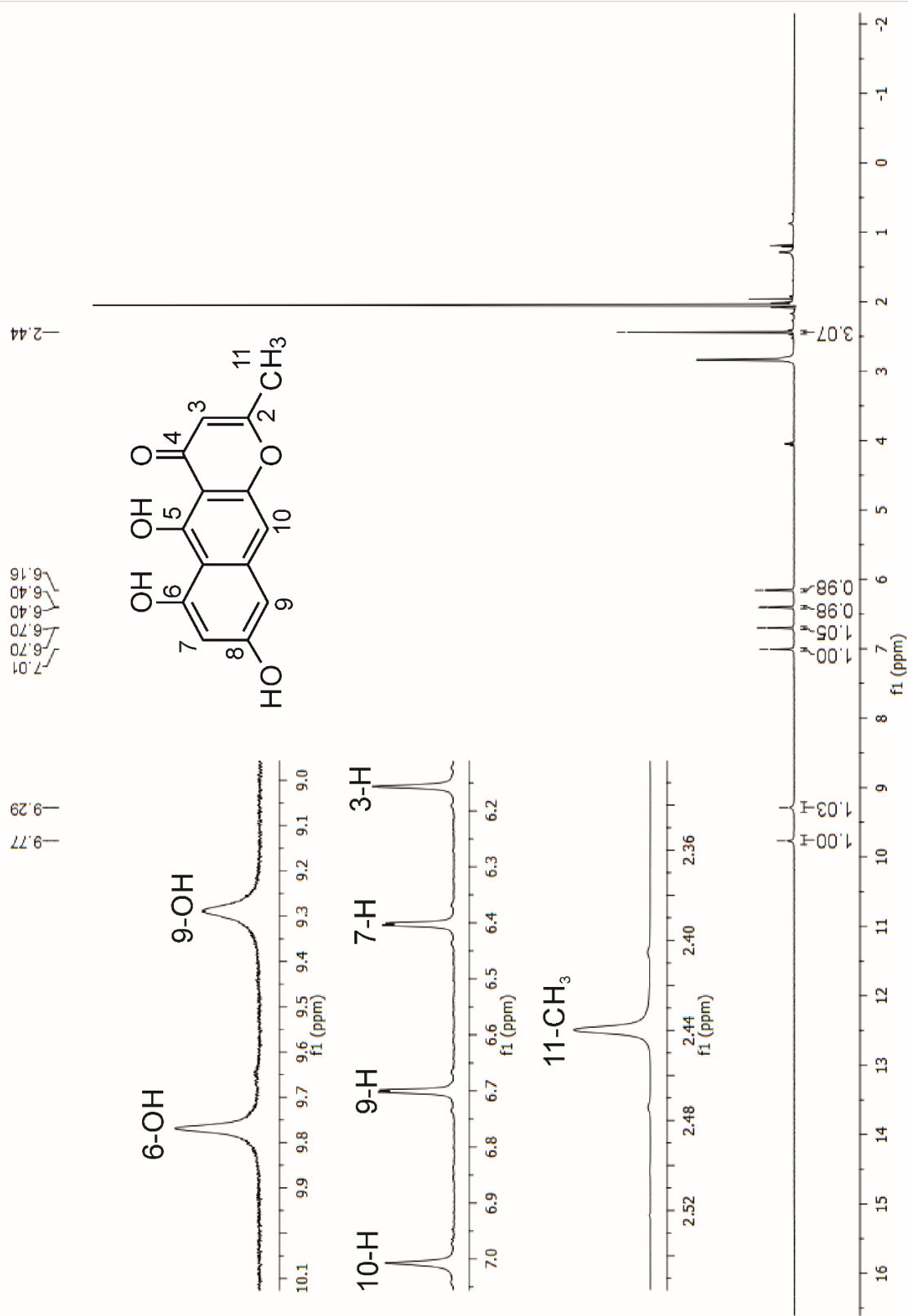
Figure S6: ^{1}H NMR spectrum of nor-rubrofusarin in aceton-d$_6$ (500 MHz)

S14
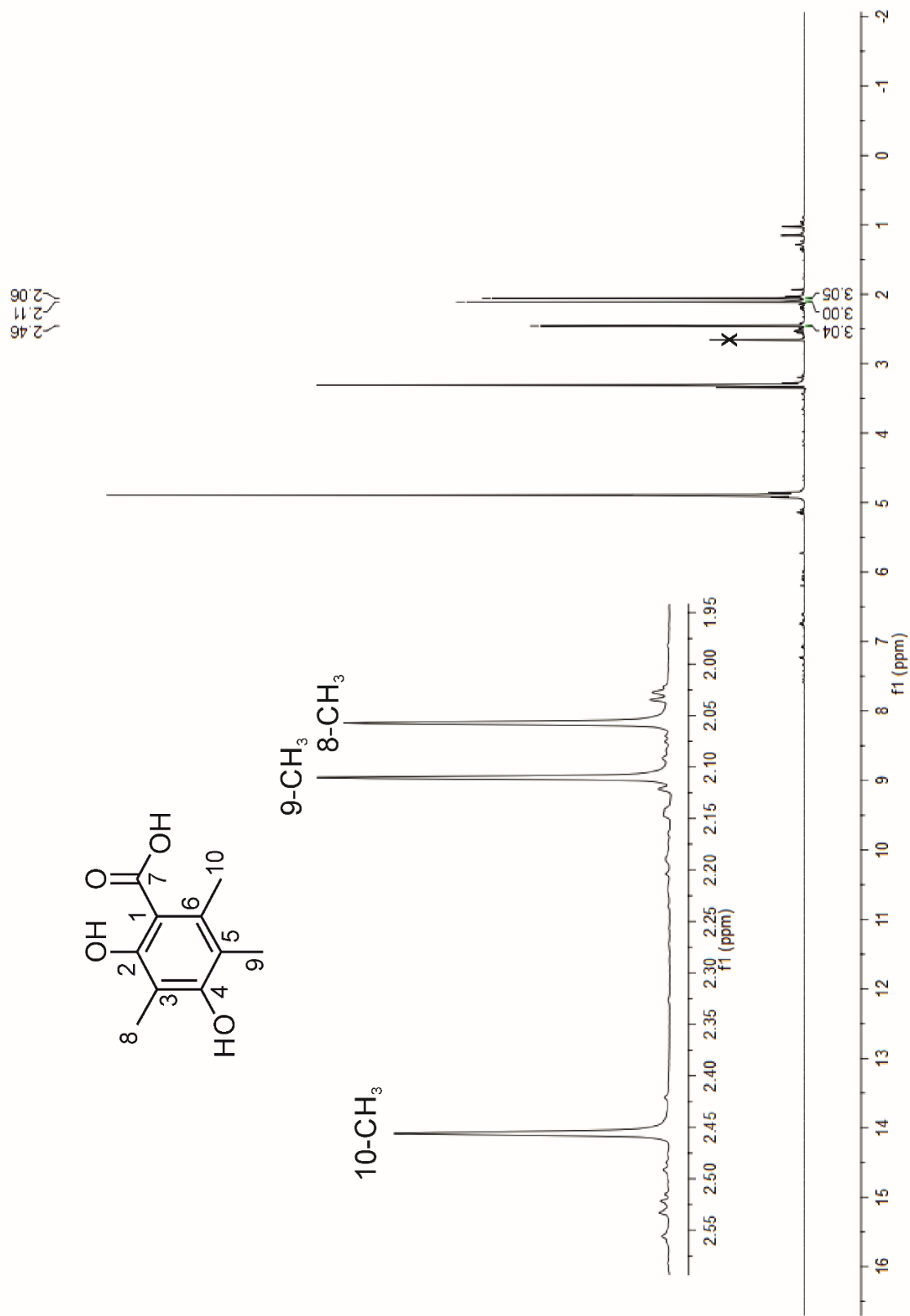
Figure S7: ^{1}H NMR spectrum of 3,5-dimethylorsellinic acid (DMOA) in CD$_3$OD (500 MHz)

S15

220

Figure S8: ^1H NMR spectrum of nor-toralactone in CDCl$_3$ (500 MHz)

S16

Figure S9: ^{1}H NMR spectrum of 6-methylsalicylic acid (6-MSA) in CDCl$_3$ (500 MHz)

S17

Figure S10: Multiple protein alignments of fungal PKSs AndM, AusA, MpaC, and the revised CRL19966

S18

223

Figure S10 (continued)

```
                1920      1930      1940      1950      1960      1970      1980      1990      2000
AndM            KILEIGAGTGGTTRYLVSQLAA.RGL..RFEYTFTDISSSLVMLAKKGFKEYDFMRYMTLNIEQDPPQELLGQYDIISTNCIHATRNIA
AusA            KVLEIGAGTGGTTNYLVQELAAVPGL..RFQYTFTDISSSLVTLARKRFKAYDFMRYTTLDIENDPSPELQGQYDIISTNCIHATRNLI
MpaC            EILEIGAGTGGTTKNLLEKLSTIPGLSTRLEYTFTDISPSLVAAGRKTFANYNFMRYETLNVENDPPSALSGQYDIVISTNCVHATRNLR
CRL19966-revised RVLELGAGTGGTTKRLVEALITT.GK..EFEYTFTDISASLVAAARRKFAQHSFFHYAVIDIEQDPPTQYQNQYDIVISTNCIHATKDLT
consensus>70    k!LEiGAGTGGTT..Lve.L....Gl..rf#YTFTDIS.SLV..ark.F.eydFmrY.tl#!E#DPp.el.gQYDI!iSTNC!HATr#1.


                2010      2020      2030      2040      2050      2060      2070      2080      2090
AndM            YSCSNIRRMLRPDGILCLIELTRNLFWFDLVFGLLEGWWLFDDGRKHALASEHLWNETLDKACYSWIDWSRNDLKESEFLRLIVASPSKP
AusA            TSCTNIRRLLRPEGILCLIELTRNLFWFDLVFGLLEGWWLFNDGRSHALAHERLWDHNLRQAGFNWVDWTDNDSAESDILRLIVASSTQP
MpaC            ESCTNIRKLLRPDGILCLIVELTRDIFWLDLVFGLLEGWWRPEDGREHALATEMMWDQTLRQSGFEWVDWTNNETVESNALRVIVASPTGN
CRL19966-revised VSCTNINKLLHSNGLLCIVELTRNLFWFDLVFGLLDGWWRPEDGRQHALASETMWRKCLTRSGFQWVDWTAGVMEESSILRVIVASPSNV
consensus>70    .SCtNIr.$Lrp#GiLCL!ELTR#1FWfDLVFGLL#GWW.F#DGR.HALA.E.$Wd..L....G%nW!DWt.nd..ESe.LR.IVASp...


                2100      2110      2120      2130      2140      2150      2160
AndM            E.............MPLVTQETVKFDEKDGVSLLADIYYPSVADDANQRRPIALLLHGGGHIMLSRKDIRADQTMLLLDAGFLPVSIDYRL
AusA            FYALEGDDECEADCNTVQEQTVLYNTRDGLELFADIYYPEKTDRSGAKRPIALLIHGGGHIMLSRKEIHHEQVRMLFDMGFLPVSIDYRL
MpaC            SSSTAT.....MSPSKLTKMETVVWGERDNLQLRADIYYPETVDTTRKQRPIALMIHGGGHVMLSRKDIRPAQTQTLLDAGFLPVSIDYRL
CRL19966-revised I.............HTTTETVEFKRVGDLSLKADIYYPRNVVSRDKRLPVALMIHGGGHIMLSRKDIRPSQVETLLDHGFLPVSVDYRL
consensus>70    .................#TV.f...d.l.L.ADIYYP...d.....rP!AL$iHGGGH!MLSRK#Ir..Q...L1D.GFLPVS!DYRL


                2170      2180      2190      2200      2210      2220      2230      2240      2250
AndM            CPETSLIEGPMRDARDAIRWARQTLPTLQIQREDIHPNGDQVVAIGWSTGGHLAMSIAWTAAELAVAAPEAVLSFYCPTDYIDPFWSQPN
AusA            CPEVSLLDGPMQDACDAIAWARNKLPQLQIQRRDILPDGNNVVAVGWSTGGHLAMTLAWTAPARGVSAPEAILSFYSPTDYTDPFWSKPN
MpaC            CPEVSLAEGPMADARDAISWVRRVIPNIPILRADIRPDGNQVVAIGWSTGGHLAMTLPFTAPAAGISAPNAVLAFYCPTNYEDPFWSNPN
CRL19966-revised CPEITLSEGPMTDVYDALAWARTVLPRLPLRRRDISIDGECVVGIGWSTGGHLALTLGFTSPLRTIRPPDATLVFYCPLDYEDPFWSEPN
consensus>70    CPEvsL.#GPM.Da.DAL.WaR..LP.l.L.R.DI.p#G#qVVa!GWSTGGHLA$tL..Tap...!.aP#AvL.FYcPt#Y.DPFWSqPN


                2260      2270      2280      2290      2300      2310      2320      2330
AndM            FPFGKQIAPPG..EIDIRAGMSPNPITAYNPPRSKRALGGWMSTTDPRSRIALHMNWKGQTIPVLLNAAK.I.....TAP..GEMLLDPT
AusA            FPYRVDVSTSDIQTGNPLDALQDAPISGYNPPPSKRALGGWMAPSDPRSRIALYMNWTGQTIPVLFYGCNYRARAAESGQDYEVVLPEPI
MpaC            FPFGQTVASNE.MEYDVWEGLQSMPIAGYNPALKERPLGGWMSTRDPRSRIALHMNWTGQTIPVLLKACTIKGNTEKCS...PDDLSRPT
CRL19966-revised LPFGDSASPSTLMQHDLLHGVHDNVITAYNPPKSRRALGGWMAPSDDRSLIALHMNWKGHTLRILLNGLSKHS..GVRS...SDELPEPT
consensus>70    fP%g..v......e.#...gl...pI..YNPp.s.RaLGGWM...DpRSrIALhMNW.GqTLp!Ll.................d.L.ePt


                2340      2350      2360      2370      2380      2390      2400      2410      2420
AndM            KEDIVRVCPTSQASAGNYRSPTFTIHGSLDDLIPMSQVRHTAKVMKANGVDVHLRELEGSTHLFDIASTYGDKPDEVQAVADGIQFLKDR
AusA            LSEVQKVCPFSQISAGSVRAPTELIHGTLDDLIPVQQAQRTHDKMQACGVDSDLRIVRDGLHLFDLEANFAGNQHAFQAVVDGYEFLRRH
MpaC            EEEIQAVSPNYQIRVGRLNTPTFLIHGTSDDLVPCAQTESTHGALTASGVEAELRVVQEAAHLFDLYPASHAGQEAKAAVAEGYEFLRRH
CRL19966-revised QSQIQAVSPLAQIRRGAYTTPTFHIHGTEDDLIPWEQAVRTFEALRQRNVIADLRILDGTEHLFDLYPSYKDDLQAVKAVQDGYEFLRRH
consensus>70    ..#!q.V.P..Qi..G.Y..PTF.IHGt.DDL!P..Q...T...$.a.gVd.dLRi.e...HLFD1...y....da..AV.#Gy#FLrrh


                2430
AndM            VQC
AusA            VGL
MpaC            VQL
CRL19966-revised VGL
consensus>70    V.1
```

Figure S10 (continued)

S20

225

Supplementary References

(1) Xu, J. and Gong, Z. Z. (2003) Intron requirement for AFP gene expression in *Trichoderma viride*, *Microbiology 149*, 3093-3097.

(2) Yin, W. B., Chooi, Y. H., Smith, A. R., Cacho, R. A., Hu, Y., White, T. C., and Tang, Y. (2013) Discovery of cryptic polyketide metabolites from dermatophytes using heterologous expression in *Aspergillus nidulans*, *ACS Synth. Biol. 2*, 629-634.

(3) Zhang, P., Wang, X., Fan, A., Zheng, Y., Liu, X., Wang, S., Zou, H., Oakley, B. R., Keller, N. P., and Yin, W. B. (2017) A cryptic pigment biosynthetic pathway uncovered by heterologous expression is essential for conidial development in *Pestalotiopsis fici*, *Mol. Microbiol. 105*, 469-483.

(4) Watanabe, A., Fujii, I., Sankawa, U., Mayorga, M. E., Timberlake, W. E., and Ebizuka, Y. (1999) Re-identification of *Aspergillus nidulans* wA gene to code for a polyketide synthase of naphthopyrone, *Tetrahedron Lett. 40*, 91-94.

(5) Frandsen, R. J., Nielsen, N. J., Maolanon, N., Sorensen, J. C., Olsson, S., Nielsen, J., and Giese, H. (2006) The biosynthetic pathway for aurofusarin in *Fusarium graminearum* reveals a close link between the naphthoquinones and naphthopyrones, *Mol. Microbiol. 61*, 1069-1080.

(6) Matsuda, Y., Wakimoto, T., Mori, T., Awakawa, T., and Abe, I. (2014) Complete biosynthetic pathway of anditomin: nature's sophisticated synthetic route to a complex fungal meroterpenoid, *J. Am. Chem Soc. 136*, 15326-15336.

(7) Newman, A. G. and Townsend, C. A. (2016) Molecular characterization of the cercosporin biosynthetic pathway in the fungal plant pathogen *Cercospora nicotianae*, *J. Am. Chem. Soc. 138*, 4219-4228.

(8) Bedford, D. J., Schweizer, E., Hopwood, D. A., and Khosla, C. (1995) Expression of a functional fungal polyketide synthase in the bacterium *Streptomyces coelicolor* A3(2), *J. Bacteriol. 177*, 4544-4548.

(9) Frandsen, R. J., Schutt, C., Lund, B. W., Staerk, D., Nielsen, J., Olsson, S., and Giese, H. (2011) Two novel classes of enzymes are required for the biosynthesis of aurofusarin in *Fusarium graminearum*, *J. Biol. Chem. 286*, 10419-10428.

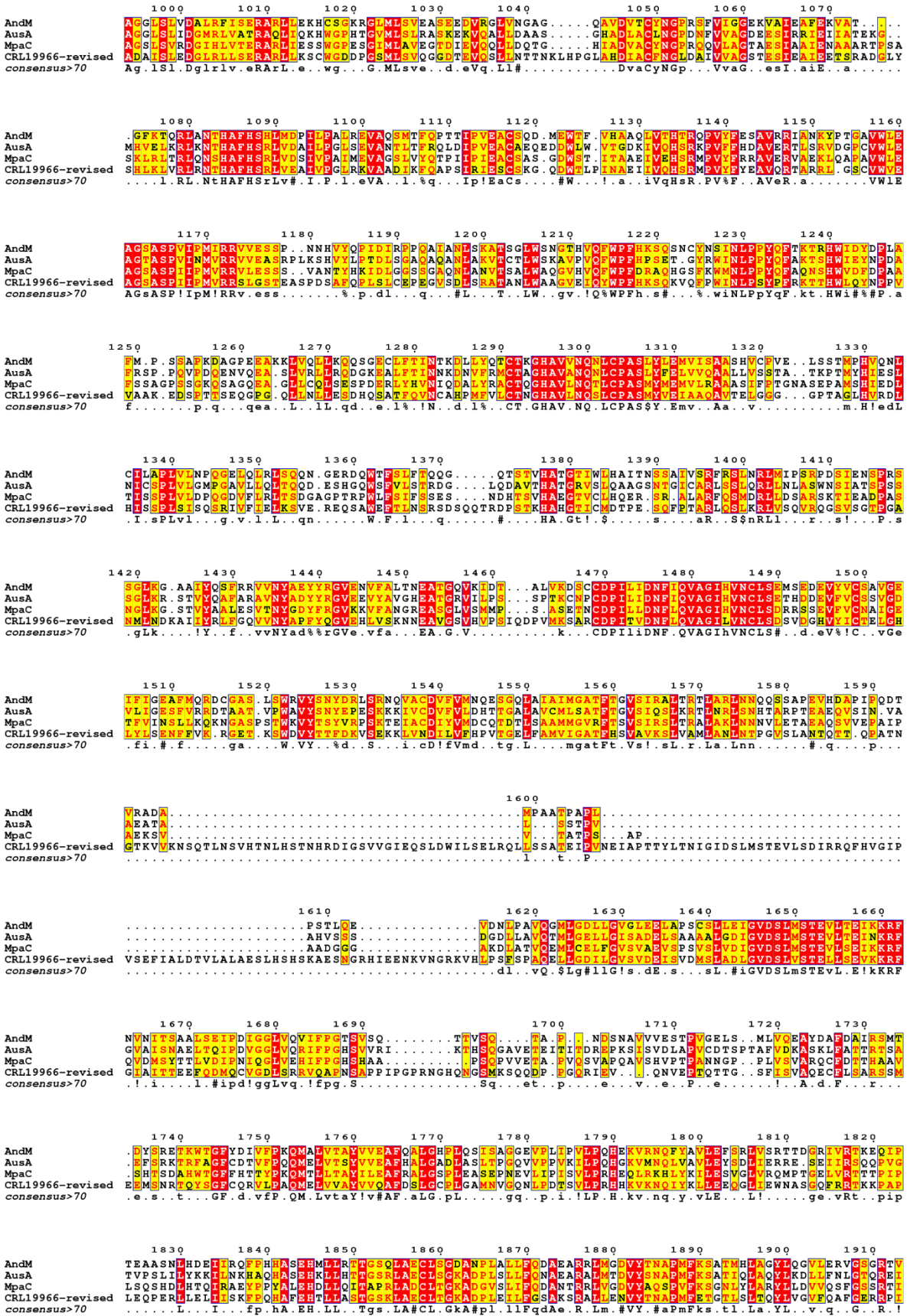(10) Yeh, H. H., Chang, S. L., Chiang, Y. M., Bruno, K. S., Oakley, B. R., Wu, T. K., and Wang, C. C. (2013) Engineering fungal nonreducing polyketide synthase by heterologous expression and domain swapping, *Org. Lett. 15*, 756-759.

(11) Yalpani, N., Altier, D. J., Barbour, E., Cigan, A. L., and Scelonge, C. J. (2001) Production of 6-methylsalicylic acid by expression of a fungal polyketide synthase activates disease resistance in tobacco, *Plant Cell 13*, 1401-1410.

5. Conclusion and outlook

In this thesis, different approaches have been used for expanding of the diversity of natural products. The biochemical characterization of enzymes with respect to their specific reaction conditions, substrates and cofactors involved as well as the elucidation of their crystal structures, are increasingly becoming standard procedures. This forms the basis for their further usage as biocatalysts in synthetic biology approaches to produce new compounds. The DMATS prenyltransferase family with more than 60 well-characterized members provides numerous biochemical information on their reactions. The systematic analysis of these data combined with chemoinformatic methods led to the development of the PrenDB to predict new potential substrates for prenyltransferases. The experimental verification of these predictions confirmed the proof of principle and the strength of this database. The possibility to extend this database by implementing newly characterized prenyltransferases will further improve the accuracy of the predictions.

The prenyltransferases EchPT1 and EchPT2 are such two examples of newly characterized DMATS members. Their characterization was the prerequisite for establishing the biosynthetic hypothesis of echinulin. Interaction of both prenyltransferases leads to reverse prenylation at position C2 of *cyclo*-L-tryptophanyl-L-alaninyl with subsequent multiple prenyl transfers to the indole nucleus. The promiscuity known from prenyltransferases with respect to their accepted substrates was also confirmed for these two prenyltransferases. This allowed their application in the production of multiple prenylated tryptophan-containing cyclic dipeptides. It was observed that the configuration of the substrates has a strong influence on the formation of tetra-prenylated products. An even better usability of EchPT1 and EchPT2 for the biosynthesis of new substances can certainly be achieved by elucidating their crystal structure.

In contrast to *in vitro* methods for the extension and investigation of natural products, heterologous expression in host organisms is an alternative *in vivo* method. Various hosts are already successfully used for the heterologous expression of secondary metabolites. However, obstacles such as different splicing patterns compared to the donor organism are encountered, making the use of these hosts sometimes difficult. To overcome this, alternative hosts for heterologous expression are necessary and welcome. Sequencing the genome of *Penicillium crustosum* and the generation of a *pyrG* deficient strain built the basis for using this fungus as a heterologous expression host. With the two generated plasmids pFK20 and pFK23, the easy detection upon successful integration into the *pcr4401* locus is ensured by the occurrence of an albino phenotypes. The suitability as an expression host was then proven by the successful expression of three PKS genes and the subsequent isolation and structural elucidation of the formed product. Additional genetic modification can further improve the potential of *Penicillium crustosum* as an expression host.

The following subjects are planned for further investigations:

➢ Ongoing maintenance and improvement of the PrenDB database by implementing newly characterized DMATS prenyltransferases and their reactions.

➢ Further investigation of the newly identified prenyltransferase substrate classes by *in vitro* enzyme assays and elucidation of the formed products.

➢ Elucidation and analysis of the crystal structures of the prenyltransferases EchPT1 and EchPT2.

➢ Targeted protein engineering of the prenyltransferases EchPT1 and EchPT2 to expand the prenyl acceptor and prenyl donor space.

➢ Improving the applicability of *Penicillium crustosum* as heterologous expression host through deletion of the backbone genes *claF* and *traA* involved in the biosynthesis of peniphenones and penilactones.

➢ Generation of a *Penicillium crustosum* KU70 deletion strain for the improvement of the gene targeting rate.

➢ Further use of *Penicillium crustosum* for the heterologous expression of whole gene clusters with associated structural elucidation of the possible products formed.

6. References

(1) Kirk, P. M., Cannon P.F., Minter D.W., and Stalpers J.A. (2008) *Dictionary of the fungi, 10th edition* CABI, Wallingford, UK.

(2) Kavanagh K. (2018) *Fungi: Biology and applications, 3rd edition* Wiley.

(3) Bennett, R. J. and Turgeon, B. G. (2016) Fungal sex: The Ascomycota, *Microbiol. Spectr. 4*.

(4) Webster, J. and Weber, R. (2007) *Introduction to fungi* Cambridge University Press., Cambridge, UK.

(5) Beimforde, C., Feldberg, K., Nylinder, S., Rikkinen, J., Tuovila, H., Dorfelt, H., Gube, M., Jackson, D. J., Reitner, J., Seyfullah, L. J., and Schmidt, A. R. (2014) Estimating the Phanerozoic history of the Ascomycota lineages: Combining fossil and molecular data, *Mol. Phylogenet. Evol. 78*, 386-398.

(6) Cadena, J., Thompson, G. R., and Patterson, T. F. (2016) Invasive aspergillosis: current strategies for diagnosis and management, *Infect. Dis. Clin. North Am. 30*, 125-142.

(7) Bongomin, F., Gago, S., Oladele, R. O., and Denning, D. W. (2017) Global and multi-national prevalence of fungal diseases-estimate precision, *J. Fungi. (Basel) 3*, 57.

(8) (2017) Stop neglecting fungi, *Nat. Microbiol. 2*, 17120.

(9) Liew, W. P. and Mohd-Redzwan, S. (2018) Mycotoxin: Its impact on gut health and microbiota, *Front Cell Infect. Microbiol. 8*, 60.

(10) Vidal, A., Mengelers, M., Yang, S., De Saeger, S., and De Boevre, M. (2018) Mycotoxin biomarkers of exposure: A comprehensive review, *Comprehensive Reviews in Food Science and Food Safety 17*, 1127-1155.

(11) Zain, M. E. (2011) Impact of mycotoxins on humans and animals, *Journal of Saudi Chemical Society 15*, 129-144.

(12) Hussein, H. S. and Brasel, J. M. (2001) Toxicity, metabolism, and impact of mycotoxins on humans and animals, *Toxicology 167*, 101-134.

(13) Vogel, J. P., Williams, M., Gallos, I., Althabe, F., and Oladapo, O. T. (2019) WHO recommendations on uterotonics for postpartum haemorrhage prevention: what works, and which one?, *BMJ Glob. Health 4*, e001466.

(14) Guo, Z. (2017) The modification of natural products for medical use, *Acta Pharm. Sin. B 7*, 119-136.

(15) Houbraken, J., Frisvad, J. C., and Samson, R. A. (2011) Fleming's penicillin producing strain is not *Penicillium chrysogenum* but *P. rubens*, *IMA Fungus. 2*, 87-95.

(16) Jakobsen, M., Cantor, M. D., and Jespersen, L. (2002) Production of bread, cheese and meat in *Industrial Applications* (Osiewacz, H. D., Ed.) pp 3-22, Springer Berlin Heidelberg, Berlin, Heidelberg.

(17) Kitamoto, K. (2015) Cell biology of the Koji mold *Aspergillus oryzae*, *Biosci. Biotechnol. Biochem. 79*, 863-869.

(18) Porfirif, M. C., Milatich, E. J., Farruggia, B. M., and Romanini, D. (2016) Production of alpha-amylase from *Aspergillus oryzae* for several industrial applications in a single step, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 1022*, 87-92.

(19) McKelvey, S. M. and Murphy, R. A. (2017) Biotechnological use of fungal enzymes in *Fungi* pp 201-225, Wiley Online Books.

(20) Karathia, H., Vilaprinyo, E., Sorribas, A., and Alves, R. (2011) *Saccharomyces cerevisiae* as a model organism: a comparative study, *PLoS One. 6*, e16015.

(21) Guzman-Chavez, F., Zwahlen, R. D., Bovenberg, R. A. L., and Driessen, A. J. M. (2018) Engineering of the filamentous fungus *Penicillium chrysogenum* as cell factory for natural products, *Front Microbiol. 9*, 2768.

(22) MacCabe, A. P., Orejas, M., and Ramon, D. (2001) *Aspergillus nidulans* as a model organism for the study of the expression of genes encoding enzymes of relevance in the food industry in *Applied Mycology and Biotechnology*

Agriculture and Food Production (Khachatourians, G. G. and Arora, D. K., Eds.) pp 239-265, Elsevier.

(23) Nielsen, J. (2017) Systems biology of metabolism, *Annu. Rev. Biochem. 86*, 245-275.

(24) Keller, N. P. (2019) Fungal secondary metabolism: regulation, function and drug discovery, *Nat. Rev. Microbiol. 17*, 167-180.

(25) Zeilinger, S., Garcia E.C., and Martin, J. F. (2015) Fungal secondary metabolites in the "OMICS" era in *Biosynthesis and Molecular Genetics of Fungal Secondary Metabolites* pp 1-12.

(26) Keller, N. P., Turner, G., and Bennett, J. W. (2005) Fungal secondary metabolism - from biochemistry to genomics, *Nat. Rev. Microbiol. 3*, 937-947.

(27) Berdy, J. (2012) Thoughts and facts about antibiotics: where we are now and where we are heading, *J. Antibiot. (Tokyo) 65*, 385-395.

(28) Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min, J. W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., and Ysebaert, M. (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene, *Nature 260*, 500-507.

(29) Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., and . (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science 269*, 496-512.

(30) Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996) Life with 6000 genes, *Science 274*, 546, 563-546, 567.

(31) Koonin, E. V. (2009) Evolution of genome architecture, *Int. J. Biochem. Cell Biol. 41*, 298-306.

(32) Salgado, H., Moreno-Hagelsieb, G., Smith, T. F., and Collado-Vides, J. (2000) Operons in *Escherichia coli*: genomic analyses and predictions, *Proc. Natl. Acad. Sci. U. S. A 97*, 6652-6657.

(33) Mohanta, T. K. and Bae, H. (2015) The diversity of fungal genome, *Biol. Proced. Online. 17*, 8.

(34) Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., and Weber, T. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline, *Nucleic Acids Res.*

(35) Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., and Fedorova, N. D. (2010) SMURF: Genomic mapping of fungal secondary metabolite clusters, *Fungal Genet. Biol. 47*, 736-741.

(36) Skinnider, M. A., Merwin, N. J., Johnston, C. W., and Magarvey, N. A. (2017) PRISM 3: expanded prediction of natural product chemical structures from microbial genomes, *Nucleic Acids Res. 45*, W49-W54.

(37) Blin, K., Kim, H. U., Medema, M. H., and Weber, T. (2017) Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters, *Briefings in Bioinformatics*.

(38) Fedorova, N. D., Moktali, V., and Medema, M. H. (2012) Bioinformatics approaches and software for detection of secondary metabolic gene clusters in *Fungal Secondary Metabolism* pp 23-45, Humana Press, Totowa, NJ.

(39) Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., and Waterston, R. H. (2017) DNA sequencing at 40: past, present and future, *Nature 550*, 345-353.

(40) Baltz, R. H. (2017) Gifted microbes for genome mining and natural product discovery, *J. Ind. Microbiol. Biotechnol. 44*, 573-588.

(41) Lind, A. L., Wisecaver, J. H., Smith, T. D., Feng, X., Calvo, A. M., and Rokas, A. (2015) Examining the evolution of the regulatory circuit controlling secondary metabolism and development in the fungal genus *Aspergillus*, *PLoS Genet. 11*, e1005096.

(42) Hoogendoorn, K., Barra, L., Waalwijk, C., Dickschat, J. S., van der Lee, T. A. J., and Medema, M. H. (2018) Evolution and diversity of biosynthetic gene clusters in *Fusarium*, *Front Microbiol. 9*, 1158.

(43) Bode, H. B., Bethe, B., Hofs, R., and Zeeck, A. (2002) Big effects from small changes: possible ways to explore nature's chemical diversity, *Chembiochem. 3*, 619-627.

(44) Pan, R., Bai, X., Chen, J., Zhang, H., and Wang, H. (2019) Exploring structural diversity of microbe secondary metabolites using OSMAC strategy: A literature review, *Front Microbiol. 10*, 294.

(45) Wakefield, J., Hassan, H. M., Jaspars, M., Ebel, R., and Rateb, M. E. (2017) Dual induction of new microbial secondary metabolites by fungal bacterial co-cultivation, *Front Microbiol. 8*, 1284.

(46) Bok, J. W. and Keller, N. P. (2004) LaeA, a regulator of secondary metabolism in *Aspergillus* spp, *Eukaryot. Cell 3*, 527-535.

(47) Bayram, O., Krappmann, S., Ni, M., Bok, J. W., Helmstaedt, K., Valerius, O., Braus-Stromeyer, S., Kwon, N. J., Keller, N. P., Yu, J. H., and Braus, G. H. (2008) VelB/VeA/LaeA complex coordinates light signal with fungal development and secondary metabolism, *Science 320*, 1504-1506.

(48) Fernandes, M., Keller, N. P., and Adams, T. H. (1998) Sequence-specific binding by *Aspergillus nidulans* AflR, a C6 zinc cluster protein regulating mycotoxin biosynthesis, *Mol. Microbiol. 28*, 1355-1365.

(49) Pfannenstiel, B. T. and Keller, N. P. (2019) On top of biosynthetic gene clusters: How epigenetic machinery influences secondary metabolism in fungi, *Biotechnol. Adv.*

(50) Williams, R. B., Henrikson, J. C., Hoover, A. R., Lee, A. E., and Cichewicz, R. H. (2008) Epigenetic remodeling of the fungal secondary metabolome, *Org. Biomol. Chem. 6*, 1895-1897.

(51) Trojer, P., Brandtner, E. M., Brosch, G., Loidl, P., Galehr, J., Linzmaier, R., Haas, H., Mair, K., Tribus, M., and Graessle, S. (2003) Histone deacetylases in fungi: novel members, new facts, *Nucleic Acids Res. 31*, 3971-3981.

(52) Shwab, E. K., Bok, J. W., Tribus, M., Galehr, J., Graessle, S., and Keller, N. P. (2007) Histone deacetylase activity regulates chemical diversity in *Aspergillus*, *Eukaryot. Cell 6*, 1656-1664.

(53) Guzman-Chavez, F., Salo, O., Samol, M., Ries, M., Kuipers, J., Bovenberg, R. A. L., Vreeken, R. J., and Driessen, A. J. M. (2018) Deregulation of secondary metabolism in a histone deacetylase mutant of *Penicillium chrysogenum*, *Microbiologyopen. 7*, e00598.

(54) He, Y., Wang, B., Chen, W., Cox, R. J., He, J., and Chen, F. (2018) Recent advances in reconstructing microbial secondary metabolites biosynthesis in Aspergillus spp, *Biotechnol. Adv. 36*, 739-783.

(55) Greunke, C., Duell, E. R., D'Agostino, P. M., Glockle, A., Lamm, K., and Gulder, T. A. M. (2018) Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential, *Metab Eng 47*, 334-345.

(56) Jacobus, A. P. and Gross, J. (2015) Optimal cloning of PCR fragments by homologous recombination in *Escherichia coli*, *PLoS One. 10*, e0119221.

(57) Zhang, J. J., Tang, X., and Moore, B. S. (2019) Genetic platforms for heterologous expression of microbial natural products, *Nat. Prod. Rep.*

(58) Cox, R. J., Skellam, E., and Williams, K. (2018) Biosynthesis of fungal polyketides in *Physiology and Genetics. The Mycota (A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research)* pp 385-412, Springer International Publishing, Cham.

(59) Alberts, A. W., Chen, J., Kuron, G., Hunt, V., Huff, J., Hoffman, C., Rothrock, J., Lopez, M., Joshua, H., Harris, E., Patchett, A., Monaghan, R., Currie, S., Stapley, E., Albers-Schonberg, G., Hensens, O., Hirshfield, J., Hoogsteen, K., Liesch, J., and Springer, J. (1980) Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent, *Proc. Natl. Acad. Sci. U. S. A 77*, 3957-3961.

(60) Harris, C. M., Roberson, J. S., and Harris, T. M. (1976) Biosynthesis of griseofulvin, *J. Am. Chem. Soc. 98*, 5380-5386.

(61) Lomovskaya, N., Otten, S. L., Doi-Katayama, Y., Fonstein, L., Liu, X. C., Takatsu, T., Inventi-Solari, A., Filippini, S., Torti, F., Colombo, A. L., and Hutchinson, C. R. (1999) Doxorubicin overproduction in *Streptomyces peucetius*: cloning and characterization of the *dnrU* ketoreductase and *dnrV* genes and the *doxA* cytochrome P-450 hydroxylase gene, *J. Bacteriol. 181*, 305-318.

(62) Staunton, J. and Wilkinson, B. (1997) Biosynthesis of erythromycin and rapamycin, *Chem. Rev. 97*, 2611-2630.

(63) Lebeau, J., Venkatachalam, M., Fouillaud, M., Petit, T., Vinale, F., Dufosse, L., and Caro, Y. (2017) Production and new extraction method of polyketide red pigments produced by ascomycetous fungi from terrestrial and marine habitats, *J. Fungi. (Basel) 3.*

(64) Cai, W. and Zhang, W. (2018) Engineering modular polyketide synthases for production of biofuels and industrial chemicals, *Curr. Opin. Biotechnol. 50*, 32-38.

(65) Chooi, Y. H. and Tang, Y. (2012) Navigating the fungal polyketide chemical space: from genes to molecules, *J. Org. Chem. 77*, 9933-9953.

(66) Winter, J. M., Cascio, D., Dietrich, D., Sato, M., Watanabe, K., Sawaya, M. R., Vederas, J. C., and Tang, Y. (2015) Biochemical and structural basis for controlling chemical modularity in fungal polyketide biosynthesis, *J. Am. Chem. Soc. 137*, 9885-9893.

(67) McKeown, D. S. J., McNicholas, C., Simpson, T. J., and Willett, N. J. (1996) Biosynthesis of norsolorinic acid and averufin: substrate specificity of norsolorinic acid synthase, *Chem. Commun.* 301-302.

(68) Crawford, J. M., Dancy, B. C., Hill, E. A., Udwary, D. W., and Townsend, C. A. (2006) Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase, *Proc. Natl. Acad. Sci. U. S. A 103*, 16728-16733.

(69) Regueira, T. B., Kildegaard, K. R., Hansen, B. G., Mortensen, U. H., Hertweck, C., and Nielsen, J. (2011) Molecular basis for mycophenolic acid biosynthesis in *Penicillium brevicompactum*, *Appl. Environ. Microbiol. 77*, 3035-3043.

(70) Herbst, D. A., Townsend, C. A., and Maier, T. (2018) The architectures of iterative type I PKS and FAS, *Nat. Prod. Rep. 35*, 1046-1069.

(71) Beck, J., Ripka, S., Siegner, A., Schiltz, E., and Schweizer, E. (1990) The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide synthases, *Eur. J. Biochem. 192*, 487-498.

(72) Holm, D. K., Petersen, L. M., Klitgaard, A., Knudsen, P. B., Jarczynska, Z. D., Nielsen, K. F., Gotfredsen, C. H., Larsen, T. O., and Mortensen, U. H. (2014) Molecular and chemical characterization

of the biosynthesis of the 6-MSA-derived meroterpenoid yanuthone D in *Aspergillus niger*, *Chem. Biol. 21*, 519-529.

(73) Campbell, C. D. and Vederas, J. C. (2010) Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes, *Biopolymers 93*, 755-763.

(74) Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A., and Cane, D. E. (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase, *Annu. Rev. Biochem. 76*, 195-221.

(75) Edwards, A. L., Matsui, T., Weiss, T. M., and Khosla, C. (2014) Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase, *J. Mol. Biol. 426*, 2229-2245.

(76) Hertweck, C., Luzhetskyy, A., Rebets, Y., and Bechthold, A. (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork, *Nat. Prod. Rep. 24*, 162-190.

(77) Kim, J. and Yi, G. S. (2012) PKMiner: a database for exploring type II polyketide synthases, *BMC. Microbiol. 12*, 169.

(78) Bisang, C., Long, P. F., Cortes, J., Westcott, J., Crosby, J., Matharu, A. L., Cox, R. J., Simpson, T. J., Staunton, J., and Leadlay, P. F. (1999) A chain initiation factor common to both modular and aromatic polyketide synthases, *Nature 401*, 502-505.

(79) Decker, H., Summers, R. G., and Hutchinson, C. R. (1994) Overproduction of the acyl carrier protein component of a type II polyketide synthase stimulates production of tetracenomycin biosynthetic intermediates in *Streptomyces glaucescens*, *The Journal of antibiotics 47*, 54-63.

(80) Shimizu, Y., Ogata, H., and Goto, S. (2017) Type III polyketide synthases: Functional classification and phylogenomics, *Chembiochem. 18*, 50-65.

(81) Kreuzaler, F. and Hahlbrock, K. (1972) Enzymatic synthesis of aromatic compounds in higher plants: Formation of naringenin (5,7,4-trihydroxyflavanone) from p-coumaroyl coenzyme A and malonyl coenzyme A, *FEBS Letters 28*, 69-72.

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

(83) Hashimoto, M., Nonaka, T., and Fujii, I. (2014) Fungal type III polyketide synthases, *Nat. Prod. Rep. 31*, 1306-1317.

(84) Austin, M. B. and Noel, J. P. (2003) The chalcone synthase superfamily of type III polyketide synthases, *Natural Product Reports 20*, 79-110.

(85) Leadlay, P. F. (1997) Combinatorial approaches to polyketide biosynthesis, *Current Opinion in Chemical Biology 1*, 162-168.

(86) Yuzawa, S., Eng, C. H., Katz, L., and Keasling, J. D. (2013) Broad substrate specificity of the loading didomain of the lipomycin polyketide synthase, *Biochemistry 52*, 3791-3793.

(87) Musiol-Kroll, E. M. and Wohlleben, W. (2018) Acyltransferases as tools for polyketide synthase engineering, *Antibiotics (Basel) 7*.

(88) Khosla, C. and Zawada, R. J. (1996) Generation of polyketide libraries via combinatorial biosynthesis, *Trends Biotechnol. 14*, 335-341.

(89) Wang, F., Wang, Y., Ji, J., Zhou, Z., Yu, J., Zhu, H., Su, Z., Zhang, L., and Zheng, J. (2015) Structural and functional analysis of the loading acyltransferase from avermectin modular polyketide synthase, *ACS Chem. Biol. 10*, 1017-1025.

(90) Klaus, M. and Grininger, M. (2018) Engineering strategies for rational polyketide synthase design, *Nat. Prod. Rep. 35*, 1070-1081.

(91) Sussmuth, R. D. and Mainz, A. (2017) Nonribosomal peptide synthesis-principles and prospects, *Angew. Chem. Int. Ed Engl. 56*, 3770-3821.

(92) Maiya, S., Grundmann, A., Li, S. M., and Turner, G. (2006) The fumitremorgin gene cluster of *Aspergillus fumigatus*: identification of a gene encoding brevianamide F synthetase, *Chembiochem. 7*, 1062-1069.

(93) Dittmann, J., Wenger, R. M., Kleinkauf, H., and Lawen, A. (1994) Mechanism of cyclosporin A biosynthesis. Evidence for synthesis via a single linear undecapeptide precursor, *The Journal of biological chemistry 269*, 2841-2846.

(94) Konz, D., Klens, A., Schoergendorfer, K., and Marahiel, M. A. (1997) The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases, *Chemistry & biology 4*, 927-937.

(95) Gonsior, M., Muhlenweg, A., Tietzmann, M., Rausch, S., Poch, A., and Sussmuth, R. D. (2015) Biosynthesis of the peptide antibiotic feglymycin by a linear nonribosomal peptide synthetase mechanism, *Chembiochem. 16*, 2610-2614.

(96) Walsh, C. T., Chen, H., Keating, T. A., Hubbard, B. K., Losey, H. C., Luo, L., Marshall, C. G., Miller, D. A., and Patel, H. M. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines, *Current Opinion in Chemical Biology 5*, 525-534.

(97) Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014) Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes, *Proc. Natl. Acad. Sci. U. S. A 111*, 9259-9264.

(98) Hur, G. H., Vickery, C. R., and Burkart, M. D. (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology, *Nat. Prod. Rep. 29*, 1074-1098.

(99) Weber, T. and Marahiel, M. A. (2001) Exploring the domain structure of modular nonribosomal peptide synthetases, *Structure 9*, R3-R9.

(100) Brown, A. S., Calcott, M. J., Owen, J. G., and Ackerley, D. F. (2018) Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines, *Nat. Prod. Rep. 35*, 1210-1228.

(101) Nguyen, K. T., Ritz, D., Gu, J. Q., Alexander, D., Chu, M., Miao, V., Brian, P., and Baltz, R. H. (2006) Combinatorial biosynthesis of novel antibiotics related to daptomycin, *Proc. Natl. Acad. Sci. U. S. A 103*, 17462-17467.

(102) Calcott, M. J., Owen, J. G., Lamont, I. L., and Ackerley, D. F. (2014) Biosynthesis of novel Pyoverdines by domain substitution in a nonribosomal peptide synthetase of Pseudomonas aeruginosa, *Appl. Environ. Microbiol. 80*, 5723-5731.

(103) Mootz, H. D., Kessler, N., Linne, U., Eppelmann, K., Schwarzer, D., and Marahiel, M. A. (2002) Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes, *J. Am. Chem. Soc. 124*, 10980-10981.

(104) Butz, D., Schmiederer, T., Hadatsch, B., Wohlleben, W., Weber, T., and Sussmuth, R. D. (2008) Module extension of a non-ribosomal peptide synthetase of the glycopeptide antibiotic balhimycin produced by *Amycolatopsis balhimycina*, *Chembiochem. 9*, 1195-1200.

(105) Heide, L. (2009) Prenyl transfer to aromatic substrates: genetics and enzymology, *Curr. Opin. Chem. Biol. 13*, 171-179.

(106) Winkelblech, J., Fan, A., and Li, S. M. (2015) Prenyltransferases as key enzymes in primary and secondary metabolism, *Appl. Microbiol. Biotechnol. 99*, 7379-7397.

(107) Meganathan, R. and Kwon, O. (2009) Biosynthesis of Menaquinone (Vitamin K2) and Ubiquinone (Coenzyme Q), *EcoSal. Plus. 3*.

(108) Boronat, A. and Rodriguez-Concepcion, M. (2015) Terpenoid biosynthesis in prokaryotes, *Adv. Biochem. Eng Biotechnol. 148*, 3-18.

(109) Palsuledesai, C. C. and Distefano, M. D. (2015) Protein prenylation: enzymes, therapeutics, and biotechnology applications, *ACS Chem. Biol. 10*, 51-62.

(110) Pemberton, T. A., Chen, M., Harris, G. G., Chou, W. K., Duan, L., Koksal, M., Genshaft, A. S., Cane, D. E., and Christianson, D. W. (2017) Exploring the influence of domain architecture on the catalytic function of diterpene synthases, *Biochemistry 56*, 2010-2023.

(111) Bian, G., Ma, T., and Liu, T. (2018) *In vivo* platforms for terpenoid overproduction and the generation of chemical diversity, *Methods Enzymol. 608*, 97-129.

(112) Grabinska, K. A., Park, E. J., and Sessa, W. C. (2016) Cis-prenyltransferase: New insights into protein glycosylation, rubber synthesis, and human diseases, *J. Biol. Chem. 291*, 18582-18590.

(113) Wallrapp, F. H., Pan, J. J., Ramamoorthy, G., Almonacid, D. E., Hillerich, B. S., Seidel, R., Patskovsky, Y., Babbitt, P. C., Almo, S. C., Jacobson, M. P., and Poulter, C. D. (2013) Prediction of function for the polyprenyl transferase subgroup in the isoprenoid synthase superfamily, *Proc. Natl. Acad. Sci. U. S. A 110*, E1196-E1202.

(114) Li, S. M. (2009) Evolution of aromatic prenyltransferases in the biosynthesis of indole derivatives, *Phytochemistry 70*, 1746-1757.

(115) Stec, E. and Li, S. M. (2012) Mutagenesis and biochemical studies on AuaA confirmed the importance of the two conserved aspartate-rich motifs and suggested difference in the amino acids for substrate binding in membrane-bound prenyltransferases, *Arch. Microbiol. 194*, 589-595.

(116) Li, W. (2016) Bringing bioactive compounds into membranes: The UbiA superfamily of intramembrane aromatic prenyltransferases, *Trends Biochem. Sci. 41*, 356-370.

(117) Ashby, M. N., Kutsunai, S. Y., Ackerman, S., Tzagoloff, A., and Edwards, P. A. (1992) COQ2 is a candidate for the structural gene encoding para-hydroxybenzoate:polyprenyltransferase, *The Journal of biological chemistry 267*, 4128-4136.

(118) Melzer, M. and Heide, L. (1994) Characterization of polyprenyldiphosphate: 4-hydroxybenzoate polyprenyltransferase from *Escherichia coli*, *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism 1212*, 93-102.

(119) Nobrega, M. P., Nobrega, F. G. d., and Tzagoloff, A. (1990) COX10 codes for a protein homologous to the ORF1 product of *Paracoccus denitrificans* and is required for the synthesis of yeast cytochrome oxidase, *undefined*.

(120) Zhang, M., Yang, C. L., Xiao, Y. S., Zhang, B., Deng, X. Z., Yang, L., Shi, J., Wang, Y. S., Li, W., Jiao, R. H., Tan, R. X., and Ge, H. M. (2017) Aurachin SS, a new antibiotic from *Streptomyces* sp. NA04227, *J. Antibiot. (Tokyo) 70*, 853-855.

(121) Araki, Y., Awakawa, T., Matsuzaki, M., Cho, R., Matsuda, Y., Hoshino, S., Shinohara, Y., Yamamoto, M., Kido, Y., Inaoka, D. K., Nagamune, K., Ito, K., Abe, I., and Kita, K. (2019) Complete biosynthetic pathways of ascofuranone and ascochlorin in *Acremonium egyptiacum*, *Proc. Natl. Acad. Sci. U. S. A 116*, 8269-8274.

(122) Kuzuyama, T., Noel, J. P., and Richard, S. B. (2005) Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products, *Nature 435*, 983-987.

(123) Edwards, D. J. and Gerwick, W. H. (2004) Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase, *J. Am. Chem. Soc. 126*, 11432-11433.

(124) Mori, T., Zhang, L., Awakawa, T., Hoshino, S., Okada, M., Morita, H., and Abe, I. (2016) Manipulation of prenylation reactions by structure-based engineering of bacterial indolactam prenyltransferases, *Nat. Commun. 7*, 10849.

(125) Elshahawi, S. I., Cao, H., Shaaban, K. A., Ponomareva, L. V., Subramanian, T., Farman, M. L., Spielmann, H. P., Phillips, G. N., Jr., Thorson, J. S., and Singh, S. (2017) Structure and specificity of a permissive bacterial C-prenyltransferase, *Nat. Chem. Biol. 13*, 366-368.

(126) Jost, M., Zocher, G., Tarcz, S., Matuschek, M., Xie, X., Li, S. M., and Stehle, T. (2010) Structure-function analysis of an enzymatic prenyl transfer reaction identifies a reaction chamber with modifiable specificity, *J. Am. Chem. Soc. 132*, 17849-17858.

(127) Metzger, U., Schall, C., Zocher, G., Unsold, I., Stec, E., Li, S. M., Heide, L., and Stehle, T. (2009) The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria, *Proc. Natl. Acad. Sci. U. S. A 106*, 14309-14314.

(128) Schuller, J. M., Zocher, G., Liebhold, M., Xie, X., Stahl, M., Li, S. M., and Stehle, T. (2012) Structure and catalytic mechanism of a cyclic dipeptide prenyltransferase with broad substrate promiscuity, *J. Mol. Biol. 422*, 87-99.

(129) Chen, R., Gao, B., Liu, X., Ruan, F., Zhang, Y., Lou, J., Feng, K., Wunsch, C., Li, S.-M., Dai, J., and Sun, F. (2017) Molecular insights into the enzyme promiscuity of an aromatic prenyltransferase, *Nat. Chem. Biol. 13*, 226-234.

(130) Yu, X. and Li, S. M. (2012) Prenyltransferases of the dimethylallyltryptophan synthase superfamily, *Methods Enzymol. 516*, 259-278.

(131) Gebler, J. C. and Poulter, C. D. (1992) Purification and characterization of dimethylallyl tryptophan synthase from *Claviceps purpurea*, *Archives of Biochemistry and Biophysics 296*, 308-313.

(132) Tsai, H. F., Wang, H., Gebler, J. C., Poulter, C. D., and Schardl, C. L. (1995) The *Claviceps purpurea* gene encoding dimethylallyltryptophan synthase, the committed step for ergot alkaloid biosynthesis, *Biochem. Biophys. Res. Commun. 216*, 119-125.

(133) Li, S. M. (2010) Prenylated indole derivatives from fungi: structure diversity, biological activities, biosynthesis and chemoenzymatic synthesis, *Nat. Prod. Rep. 27*, 57-78.

(134) Grundmann, A. and Li, S. M. (2005) Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from *Aspergillus fumigatus*, *Microbiology 151*, 2199-2207.

(135) Wollinsky, B., Ludwig, L., Xie, X., and Li, S. M. (2012) Breaking the regioselectivity of indole prenyltransferases: identification of regular C3-prenylated hexahydropyrrolo[2,3-b]indoles as side products of the regular C2-prenyltransferase FtmPT1, *Org. Biomol. Chem. 10*, 9262-9270.

(136) Unsold, I. A. and Li, S. M. (2005) Overproduction, purification and characterization of FgaPT2, a dimethylallyltryptophan synthase from *Aspergillus fumigatus*, *Microbiology 151*, 1499-1505.

(137) Bandari, C., Scull, E. M., Bavineni, T., Nimmo, S. L., Gardner, E. D., Bensen, R. C., Burgett, A. W., and Singh, S. (2019) FgaPT2, a biocatalytic tool for alkyl-diversification of indole natural products, *MedChemComm*.

(138) Yin, W. B., Ruan, H. L., Westrich, L., Grundmann, A., and Li, S. M. (2007) CdpNPT, an N-prenyltransferase from *Aspergillus fumigatus*: overproduction, purification and biochemical characterisation, *Chembiochem. 8*, 1154-1161.

(139) Yu, X., Zocher, G., Xie, X., Liebhold, M., Schutz, S., Stehle, T., and Li, S. M. (2013) Catalytic mechanism of stereospecific formation of cis-configured prenylated pyrroloindoline diketopiperazines by indole prenyltransferases, *Chem. Biol. 20*, 1492-1501.

(140) Fan, A., Winkelblech, J., and Li, S. M. (2015) Impacts and perspectives of prenyltransferases of the DMATS superfamily for use in biotechnology, *Appl. Microbiol. Biotechnol. 99*, 7399-7415.

(141) Zhou, K., Zhao, W., Liu, X.-Q., and Li, S.-M. (2016) Saturation mutagenesis on Tyr205 of the cyclic dipeptide C2-prenyltransferase FtmPT1 results in mutants with strongly increased C3-prenylating activity, *Appl Microbiol Biotechnol.* DOI: 10.1007/s00253-016-7663-9.

(142) Fan, A. and Li, S.-M. (2016) Saturation mutagenesis on Arg244 of the tryptophan C4-prenyltransferase FgaPT2 leads to enhanced catalytic ability and different preferences for tryptophan-containing cyclic dipeptides, *Appl. Microbiol. Biotechnol. 100*, 5389-5399.

(143) Zhao, W., Fan, A., Tarcz, S., Zhou, K., Yin, W. B., Liu, X. Q., and Li, S.-M. (2017) Mutation on Gly115 and Tyr205 of the cyclic dipeptide *C2*-prenyltransferase FtmPT1 increases its catalytic activity toward hydroxynaphthalenes, *Appl. Microbiol. Biotechnol.*

(144) Zheng, L., Mai, P., Fan, A., and Li, S.-M. (2018) Switching a regular tryptophan C4-prenyltransferase to a reverse tryptophan-containing cyclic dipeptide C3-prenyltransferase by sequential site-directed mutagenesis, *Org Biomol. Chem 16*, 6688-6694.

(145) Liao, G., Mai, P., Fan, J., Zocher, G., Stehle, T., and Li, S.-M. (2018) Complete decoration of the indolyl residue in *cyclo*-L-Trp-L-Trp with geranyl moieties by using engineered dimethylallyl transferases, *Org. Lett. 20*, 7201-7205.

(146) Mai, P., Zocher, G., Stehle, T., and Li, S.-M. (2018) Structure-based protein engineering enables prenyl donor switching of a fungal aromatic prenyltransferase, *Org. Biomol. Chem. 16*, 7461-7469.

(147) de Sa Alves, F. R., Barreiro, E. J. B., and Fraga, C. A. M. (2009) From nature to drug discovery: the indole scaffold as a 'privileged structure', *Mini-rev. Med. Chem. 9*, 782-793.

(148) Wang, Y., Wang, P., Ma, H., and Zhu, W. (2013) Developments around the bioactive diketopiperazines: a patent review, *Expert. Opin. Ther. Pat 23*, 1415-1433.

(149) Liu, J., Yu, H., and Li, S. M. (2018) Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from *Streptomyces* strains, *Appl. Microbiol. Biotechnol. 102*, 4435-4444.

(150) Slack, G. J., Puniani, E., Frisvad, J. C., Samson, R. A., and Miller, J. D. (2009) Secondary metabolites from Eurotium species, *Aspergillus calidoustus* and *A. insuetus* common in Canadian homes with a review of their chemistry and biological activities, *Mycol. Res. 113*, 480-490.

(151) Nagia, M., Gaid, M., Biedermann, E., Fiesel, T., El-Awaad, I., Hansch, R., Wittstock, U., and Beerhues, L. (2019) Sequential regiospecific gem-diprenylation of tetrahydroxyxanthone by prenyltransferases from *Hypericum* sp, *New Phytol. 222*, 318-334.

(152) Chen, X., Mukwaya, E., Wong, M. S., and Zhang, Y. (2014) A systematic review on biological activities of prenylated flavonoids, *Pharm. Biol. 52*, 655-660.

(153) Mukai, R. (2018) Prenylation enhances the biological activity of dietary flavonoids by altering their bioavailability, *Biosci. Biotechnol. Biochem. 82*, 207-215.

(154) Arung, E. T., Shimizu, K., and Kondo, R. (2007) Structure-activity relationship of prenyl-substituted polyphenols from *Artocarpus heterophyllus* as inhibitors of melanin biosynthesis in cultured melanoma cells, *Chem. Biodivers. 4*, 2166-2171.

(155) Botta, B., Vitali, A., Menendez, P., Misiti, D., and Delle, M. G. (2005) Prenylated flavonoids: pharmacology and biotechnology, *Curr. Med. Chem. 12*, 717-739.

(156) van de Schans, M. G., Ritschel, T., Bovee, T. F., Sanders, M. G., de, W. P., Gruppen, H., and Vincken, J. P. (2015) Involvement of a hydrophobic pocket and helix 11 in determining the modes of action of prenylated flavonoids and isoflavonoids in the human estrogen receptor, *Chembiochem. 16*, 2668-2677.

(157) Land, M., Hauser, L., Jun, S. R., Nookaew, I., Leuze, M. R., Ahn, T. H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., and Ussery, D. W. (2015) Insights from 20 years of bacterial genome sequencing, *Funct. Integr. Genomics 15*, 141-161.

(158) Alberti, F., Foster, G. D., and Bailey, A. M. (2017) Natural products from filamentous fungi and production by heterologous expression, *Appl. Microbiol. Biotechnol. 101*, 493-500.

(159) Tsai, H. F., Fujii, I., Watanabe, A., Wheeler, M. H., Chang, Y. C., Yasuoka, Y., Ebizuka, Y., and Kwon-Chung, K. J. (2001) Pentaketide melanin biosynthesis in *Aspergillus fumigatus* requires chain-length shortening of a heptaketide precursor, *J. Biol. Chem. 276*, 29292-29298.

(160) van Dijk, J. W. and Wang, C. C. (2016) Heterologous expression of fungal secondary metabolite pathways in the *Aspergillus nidulans* host system, *Methods Enzymol. 575*, 127-142.

(161) Gressler, M., Hortschansky, P., Geib, E., and Brock, M. (2015) A new high-performance heterologous fungal expression system based on regulatory elements from the *Aspergillus terreus* terrein gene cluster, *Front Microbiol. 6*, 184.

(162) Sakai, K., Kinoshita, H., and Nihira, T. (2012) Heterologous expression system in *Aspergillus oryzae* for fungal biosynthetic gene clusters of secondary metabolites, *Appl. Microbiol. Biotechnol. 93*, 2011-2022.

(163) Song, Z., Bakeer, W., Marshall, J. W., Yakasai, A. A., Khalid, R. M., Collemare, J., Skellam, E., Tharreau, D., Lebrun, M. H., Lazarus, C. M., Bailey, A. M., Simpson, T. J., and Cox, R. J. (2015) Heterologous expression of the avirulence gene *ACE1* from the fungal rice pathogen *Magnaporthe oryzae*, *Chem. Sci. 6*, 4837-4845.

(164) Geib, E., Baldeweg, F., Doerfer, M., Nett, M., and Brock, M. (2019) Cross-chemistry leads to product diversity from atromentin synthetases in *Aspergilli* from section *Nigri*, *Cell Chem. Biol. 26*, 223-234.

(165) Li, D., Tang, Y., Lin, J., and Cai, W. (2017) Methods for genetic transformation of filamentous fungi, *Microb. Cell Fact. 16*, 168.

(166) Fan, J., Liao, G., Kindinger, F., Ludwig-Radtke, L., Yin, W. B., and Li, S. M. (2019) Peniphenone and penilactone formation in *Penicillium crustosum* via 1,4-Michael additions of ortho-quinone methide from hydroxyclavatol to gamma-butyrolactones from crustosic acid, *J. Am. Chem. Soc. 141*, 4225-4229.

(167) Vandermolen, K. M., Raja, H. A., El-Elimat, T., and Oberlies, N. H. (2013) Evaluation of culture media for the production of secondary metabolites in a natural products screening program, *AMB. Express 3*, 71.

Statutory declaration

Ich, Florian Kindinger, versichere, dass ich meine Dissertation

"Molecular biological and biochemical approaches to expand the spectrum of fungal natural products"

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den...........................

....

Florian Kindinger

Acknowledgment

First of all, I would like to express my gratitude to Prof. Dr. Shu-Ming Li for his excellent mentoring and guidance during my Ph.D. study. At all times, he was interested and ready to discuss ideas, problems, and scientific questions together. Even during tough times, he could motivate me again and again to develop new solutions for problems. At the same time, he gave me much freedom to develop and realize my own ideas. I would also like to thank him for the opportunity to get to know China, its culture and its people.

I am very grateful to Prof. Dr. Peter Kolb for the good collaboration during our cooperation and for his work as second referee and examiner. I would also like to take this opportunity to thank Viola Wohlgemuth and Dr. Jakub Gunera for the good work during our cooperation and the interesting discussions. Furthermore, I would like to thank Stefan Newel for taking all NMR spectra.

A heartfelt thank you goes to Basti and Jonas with whom I had incredible fun at work and in my free time. My thanks go to Johanna, Alex and Kirsten for the wonderful time and the supply with freshly cooked coffee and "Stückchen". I would also like to thank Rixa, especially for her open ear for private and job-related problems and her friendship. I'm thankful to Mike for the time we spent together in sports and I am very happy that we stayed in contact after he finished his doctoral thesis. I would like to thank Peter-Mai and Liu for an exciting time together in China.

To all my former and present colleagues, Katja, Kathrin, Carsten, Julia, Lena, Lena (Mimi), Linus, Kristin, Lennart, Julia, Lindsay, Nina, Elisabeth, Huili, Kang, Aili, Jie, Huomiao, Ge, Jing and Pan I would like to say thank you for the wonderful time together and the wonderful individual moments.

Last but not least I would like to thank my family and Denise for their support during my university years.

"Die Seite 244 (Lebenslauf) enthält persönliche Daten. Sie ist deshalb nicht Bestandteil der Online-Veröffentlichung."