

**Elucidation of novel biosynthetic pathways for the discovery  
of cyclodipeptide derivatives from *Streptomyces* species**

**Aufklärung neuer Biosynthesewege zur Entdeckung von  
Cyclodipeptid-Derivaten aus *Streptomyces* Arten**

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Pharmazie der Philipps-Universität Marburg

vorgelegt von

**Lauritz Christian Harken**

aus Schleswig

Marburg an der Lahn, 2022

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

Zweitgutachter: **Prof. Dr. Raphael Reher**

Hochschulkennziffer: 1180

Eingereicht am: 10.10.2022

Tag der mündlichen Prüfung: 22.11.2022

Die Planung der Arbeit, Durchführung der Experimente und Auswertung der Ergebnisse dieser Dissertation wurden am Institut für Pharmazeutische Biologie und Biotechnologie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit vom August 2019 bis Juli 2022 unter der Leitung von Prof. Dr. Shu-Ming Li geleistet.

gewidmet meiner Familie





## TABLE OF CONTENTS

### Table of contents

Erklärung .....	III
Publications .....	V
Erklärung zum Eigenanteil.....	VII
Academic activities .....	IX
Abbreviations .....	XI
Summary .....	XV
Zusammenfassung .....	XVII
1. Introduction .....	1
1.1 <i>Streptomyces</i> – their biology and impact on the environment and humankind .....	1
1.1.1 Taxonomy overview .....	1
1.1.2 Occurrence and properties.....	3
1.1.3 Lifecycle .....	4
1.1.4 Impact on the environment.....	6
1.1.5 <i>Streptomyces</i> and their irreplaceable importance for modern medicine.....	6
1.2 Elucidation of biosynthetic pathways in <i>Streptomyces</i> .....	8
1.2.1 Genome mining for biosynthetic gene clusters .....	8
1.2.2 Activation of cryptic biosynthetic gene clusters.....	8
1.3 Cyclodipeptides with a 2,5-diketopiperazine scaffold .....	10
1.3.1 Chemical properties and pharmacological activities .....	10
1.3.2 Biosynthesis of cyclodipeptides – NRPS vs. CDPS.....	13
1.4 Tailoring enzymes associated to cyclodipeptide synthases .....	15
1.4.1 Cyclodipeptide oxidases .....	15
1.4.2 Cytochrome P450 enzymes .....	16
1.4.2.1 Properties .....	16
1.4.2.2 Reaction mechanism.....	16
1.4.2.3 CDPs derivatives catalyzed by P450 enzymes.....	17
1.4.3 Fe <sup>II</sup> /2-oxoglutarate dependent oxidases .....	19
1.4.4 Prenyltransferases .....	20
2. Aims of this thesis.....	21
3. Results and discussion .....	23

## TABLE OF CONTENTS

3.1 Discovery of the cWM derivative guatrypmethine C and the Fe <sup>II</sup> /2-OG oxidase GtmE as a new enzyme type for double bond formation in CDPS-associated pathways.....	23
3.2 Identification of widely distributed, bifunctional P450 oxidases from actinobacteria for intramolecular C-C bond formation of cYY and its coupling with nucleobases .....	29
3.3 Investigation of the streptoazine biosynthetic pathway from <i>Streptomyces aurantiacus</i> reveals the presence of the promiscuous prenyltransferase SasB .....	35
4. Publications .....	37
4.1 Biosynthesis of Guatrypmethine C Implies Two Different Oxidases for exo Double Bond Installation at the Diketopiperazine Ring .....	37
4.2 Widely Distributed Bifunctional Bacterial Cytochrome P450 Enzymes Catalyze both Intramolecular C-C Bond Formation in cyclo-L-Tyr-L-Tyr and Its Coupling with Nucleobases .....	119
4.3 Elucidation of the Streptoazine Biosynthetic Pathway in <i>Streptomyces aurantiacus</i> Reveals the Presence of a Promiscuous Prenyltransferase/Cyclase.....	167
5. Conclusion and future perspectives .....	221
6. References .....	223
Acknowledgements .....	231
Curriculum vitae.....	233

## Erklärung

Ich, Lauritz Christian Harken, versichere, dass ich meine Dissertation mit dem Titel:

**„Elucidation of novel biosynthetic pathways for the discovery of  
cyclodipeptide derivatives from *Streptomyces* species”**

selbstständig und ohne fremde Hilfe verfasst und dabei keine anderen als die von mir ausdrücklich bezeichneten Quellen oder Hilfsmittel benutzt habe. Des Weiteren sind alle vollständig oder sinngemäß übernommenen Zitate als solche gekennzeichnet.

Diese Dissertation wurde in vorliegender oder ähnlicher Form noch bei keiner anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht.

Marburg, 10.10.2022

.....



## Publications

1. **Lauritz Harken**, Jing Liu, Oliver Kreuz, Robert Berger, and Shu-Ming Li (2022). Biosynthesis of Guatrypmethine C Implies Two Different Oxidases for *exo* Double Bond Installation at the Diketopiperazine Ring. *ACS Catalysis* **12** (1), 648-654. DOI: 10.1021/acscatal.1c04609
2. Jing Liu,\* **Lauritz Harken**,\* Yiling Yang, Xiulan Xie, and Shu-Ming Li (2022). Widely Distributed Bifunctional Bacterial Cytochrome P450 Enzymes Catalyze both Intramolecular C-C Bond Formation in *cyclo*-L-Tyr-L-Tyr and Its Coupling with Nucleobases. *Angewandte Chemie International Edition*, DOI: 10.1002/anie.202200377
3. Jing Liu,\* Yiling Yang,\* **Lauritz Harken**, and Shu-Ming Li (2021). Elucidation of the Streptoazine Biosynthetic Pathway in *Streptomyces aurantiacus* Reveals the Presence of a Promiscuous Prenyltransferase/Cyclase. *Journal of Natural Products* **84** (12), 3100-3109 DOI: 10.1021/acs.jnatprod.1c00844
4. **Lauritz Harken** and Shu-Ming Li (2021). Modifications of Diketopiperazines Assembled by Cyclodipeptide Synthases with Cytochrome P450 Enzymes. *Applied Microbiology and Biotechnology* **105**, 2277–2285. DOI: 10.1007/s00253-021-11178-1.

\*These authors contributed equally to this work



## Erklärung zum Eigenanteil

Titel der Publikation und <b>Journal incl. Jahr, Heft, Seitenzahl, DOI</b> <b>O: Originalarbeit</b> <b>Ü: Übersichtartikel/Review</b>	Autoren	geschätzter Eigenanteil in %	<b>Bitte angeben:</b> angenommen/ eingereicht
Biosynthesis of Guatripmethine C Implies Two Different Oxidases for exo Double Bond Installation at the Diketopiperazine Ring. <i>ACS Catalysis</i> . 2022, <b>12</b> (1), 648-654 DOI: 10.1021/acscatal.1c04609 <b>Originalarbeit</b>	<b>Lauritz Harken</b> , Jing Liu, Oliver Kreuz, Robert Berger, and Shu-Ming Li	55	angenommen
Widely Distributed Bifunctional Bacterial Cytochrome P450 Enzymes Catalyze both Intramolecular C-C Bond Formation in <i>cyclo</i> -L-Tyr-L-Tyr and Its Coupling with Nucleobases <i>Angewandte Chemie Int. Ed.</i> . 2022 DOI: 10.1002/anie.202200377 <b>Originalarbeit</b>	Jing Liu*, <b>Lauritz Harken</b> *, Yiling Yang, Xiulan Xie, Shu-Ming Li	35	angenommen
Elucidation of the Streptoazine Biosynthetic Pathway in <i>Streptomyces aurantiacus</i> Reveals the Presence of a Promiscuous Prenyltransferase/Cyclase <i>Journal of Natural Products</i> . 2021, <b>84</b> (12), 3100-3109 DOI: 10.1021/acs.jnatprod.1c00844 <b>Originalarbeit</b>	Jing Liu*, Yiling Yang*, <b>Lauritz Harken</b> , and Shu-Ming Li	5	angenommen
Modifications of Diketopiperazines Assembled by Cyclodipeptide Synthases with Cytochrome P450 Enzymes. <i>Applied Microbiology and Biotechnology</i> . 2021, <b>105</b> , 2277–2285 DOI: 10.1007/s00253-021-11178-1 <b>Übersichtartikel/Review</b>	<b>Lauritz Harken</b> and Shu-Ming Li	75	angenommen

\*These authors contributed equally to this work

-----

Kandidat

-----

Unterschrift Betreuer





## Academic activities

**Lauritz Harken**, Jing Liu, and Shu-Ming Li

Discovery of Two New Guanylated Cyclodipeptides by Heterologous Expression of a Biosynthetic Gene Cluster from *Streptomyces cinnamoneus*

**Poster presentation**, annual meeting of the German pharmaceutical society (DPhG), 30th September 2021, Leipzig (online)

**Lauritz Harken**, Jing Liu, Yiling Yang, and Shu-Ming Li

Widely Distributed Bifunctional Bacterial P450 Enzymes Catalyze both Intramolecular C-C Bond Formation in *cyclo*-L-Tyr-L-Tyr and its Coupling with Nucleobases

**Poster presentation**, annual meeting of the German pharmaceutical society (DPhG), 14th September 2022, Marburg



## Abbreviations

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

The IUPAC nomenclatures for elements and amino acids have been used.

[M+H] <sup>+</sup>	molecular ion plus proton
16S-rRNA	16 svedberg ribosomal ribonucleic acid
18S-rRNA	18 svedberg ribosomal ribonucleic acid
5-HT	5-hydroxytryptamine / serotonin
Å	angstrom
aa	amino acid
aaRS	aminoacyl transporter ribonucleic acid synthetase
aa-tRNA	aminoacyl transporter ribonucleic acid
ABBA	$\alpha$ - $\beta$ - $\beta$ - $\alpha$ fold
ACN	acetonitrile
A-domain	adenylation domain
ARS	agricultural research service
AUC	area under the curve
BGC	biosynthetic gene cluster
BLAST	basic local alignment search tool
bp	base pair
br	broad
calcd	calculated
CD <sub>3</sub> OD	deuterated methanol
CDO	cyclodipeptide oxidase
C-domain	condensation domain
CDP	cyclodipeptide
CDPS	cyclodipeptide synthase
COSMO	conductor-like screening model
COSY	correlation spectroscopy
cpd	compound
cXX	<i>cyclo</i> -L-amino acid-L-amino acid
cWΔX	<i>cyclo</i> -L-tryptophan-14,17-dehydro amino acid
d	doublet
D <sub>2</sub> O	deuterated water
Da	dalton
dd	double doublet
ddd	double double doublet
DFG	Deutsche Forschungsgemeinschaft
DFT	density functional theory
DKP	2,5-diketopiperazine
DMAPP	dimethylallyl diphosphate
DMSO- <i>d</i> <sub>6</sub>	deuterated dimethyl sulfoxide

## ABBREVIATIONS

DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DOI	digital object identifier
DSM / DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
<i>E. coli</i>	<i>Escherichia coli</i>
e.g.	example given
ECD	electronic circular dichroism
EDTA	ethylenediaminetetraacetic acid
EIC	extracted ion chromatography
ESI	electrospray ionization
<i>et. al.</i>	<i>et alii</i> / and others
EtOAc	ethyl acetate
FAD	flavin adenine dinucleotide
Fe <sup>II</sup> /2-OG	iron(II) / 2-oxoglutarate dependent
fig	figure
FPP	farnesyl diphosphate
GYM	glucose yeast malt
HCOOH	formic acid
his-tag	polyhistidine tag
HMBC	heteronuclear multiple bond correlation
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum coherence
<i>i.e.</i>	<i>id est</i> / this is
IPNS	isopenicillin N synthase
IPP	isopentenyl diphosphate
J	coupling constant
$k_{cat}$	turnover number
$K_M$	Michaelis-Menten constant
LB	Luria-Bertani
LC-MS	liquid chromatography–mass spectrometry
m	multiplet
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
m/z	mass to charge ratio
mAU	milli absorbance units
mdeg	millidegrees
MHz	mega hertz
MOPS	3-(N-morpholino)propane sulfonic acid
MS	mass spectrometry
MT	methyltransferase
multi	multiplicity
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate

## ABBREVIATIONS

NCBI	national center for biotechnology information
NF- $\kappa$ B	nuclear factor kappa B
Ni-NTA	nickel nitrilotriacetic acid
nkat	nano katal
NMR	nuclear magnetic resonance
NOESY	nuclear overhauser enhancement spectroscopy
NP	natural product
NRPS	nonribosomal peptide synthetase
NRRL	northern regional research laboratory
<i>ori</i>	origin of replication
P450	cytochrome pigment 450 nm
PCP-domain	peptidyl carrier protein domain
PCR	polymerase chain reaction
pH	potential of hydrogen
pos	position
PPi	diphosphate
ppm	parts per million
PT	prenyltransferase
rel	relative
RNA	ribonucleic acid
RNAP	ribonucleic acid polymerase
RP	reverse phase
rpm	revolutions per minute
s	singlet
<i>S. xxx</i>	<i>Streptomyces xxx</i>
<i>S. albus</i>	<i>Streptomyces albus</i> J1074
<i>S. aurantiacus</i>	<i>Streptomyces aurantiacus</i> NRRL ISP-5412
<i>S. cinnamoneus</i>	<i>Streptomyces cinnamoneus</i> DSM 40646
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
std	standard
t	triplet
td	triple doublet
T-domain	thiolation domain
TE-domain	thioesterase domain
TRIS	tris(hydroxymethyl) aminomethane
TSB	tryptic soy broth
UV	ultraviolet
UV-vis	ultraviolet-visible
v/v	volume per volume
w/v	weight per volume
$\delta_{15\text{N}}$	chemical shift of $^{15}\text{N}$
$\delta_{\text{C}}$	chemical shift of $^{13}\text{C}$
$\delta_{\text{H}}$	chemical shift of $^1\text{H}$



## Summary

Cyclodipeptides (CDPs) with a 2,5-diketopiperazine (DKP) as central core occur ubiquitously in living organisms, from simple bacteria and fungi to more complex ones like plants and animals. They display various biological and pharmacological effects, including antibiotic, antifungal, and antiproliferative activities. In microorganisms, CDPs are usually synthesized by one of the two distinct enzyme families, nonribosomal peptide synthetases (NRPSs), mainly occurring in fungi, or cyclodipeptide synthases (CDPSs), commonly found in bacteria. NRPS for CDP formation are comparably large (~2500 amino acids), bi-modular enzymes, using free amino acids as substrates. CDPSs on the other hand are smaller (200 – 300 amino acids) and require activated amino acyl tRNAs for peptide bond formation. In general, the formation of the DKP ring increases the stability of CDPs against proteolysis compared to their acyclic counterparts. This enables a variety of intriguing modifications carried out by tailoring enzymes. Their genetic information often lies in direct neighborhood to that of backbone enzymes, like CDPSs, arranged in biosynthetic gene clusters (BGCs). In CDPS-associated pathways, tailoring enzymes comprise cyclodipeptide oxidases (CDOs), cytochrome P450 (P450) enzymes, Fe<sup>II</sup>/2-oxoglutarate dependent (Fe<sup>II</sup>/2-OG) oxidases, as well as methyl- (MTs) and prenyltransferases (PTs). In this thesis, eight of such BGCs from *Streptomyces* species were identified using genome mining and elucidated by a combination of heterologous expression and biochemical analyses.

In the first project, a BGC from *Streptomyces cinnamoneus* consisting of five genes was chosen for detailed investigation and termed *gtm* gene cluster. It codes for four enzymes, *i.e.* a CDPS (GtmA), a CDO (GtmBC), a P450 enzyme (GtmD), and a Fe<sup>II</sup>/2-OG oxidase (GtmE). The genes were cloned in different combinations into the replicative pPWW50A vector for heterologous expression in *Streptomyces albus* J1074 (*S. albus*). Investigation using LC-MS and NMR spectroscopy revealed that GtmA synthesizes *cyclo*-L-Trp-L-Met, GtmBC installs a double bond at the methionine residue of the DKP, GtmD transfers a guanine onto the tryptophan moiety, and GtmE forms a second double bond at another side of the DKP. Together, this cascade results in the formation of the novel secondary metabolite guatrypmethine C. As the second dehydrogenation by GtmE displayed a novel reaction for Fe<sup>II</sup>/2-OG oxidases in CDPS-dependent pathways, it was further characterized biochemically using the recombinant protein. It was proven that GtmE indeed catalyzes the conversion of the precursor guatrypmethine A to the pathway end product guatrypmethine C. No efficient conversion of the stable isomer guatrypmethine B was observed by GtmE. This experimental finding was further supported by quantum chemical calculations using density functional theory.

In the second project, in cooperation with Dr. Jing Liu, a widely distributed two-gene locus, *gymAB*, was identified in 47 different actinobacteria. It comprises the genes *gymA* and *gymB*, coding for a CDPS and a P450 oxidase, respectively. The latter is closely related to CYP121, an essential enzyme for the viability of *Mycobacterium tuberculosis*. Six representative *Streptomyces* species were selected for functional elucidation of these BGCs. In analogy to the first project, their genes were cloned into pPWW50A and overexpressed in *S. albus*. Analyses of the cultural extracts by LC-MS in combination with NMR spectroscopy of the purified compounds showed that all six CDPSs produce *cyclo*-L-Tyr-L-Tyr (cYY) as major product. Subsequently, the P450 oxidases catalyze two different kind of reactions –

## SUMMARY

either the formation of an intramolecular C-C bond within cYY resulting in mycocyclosin, or the intermolecular transfer of the nucleobases guanine or hypoxanthine, leading to the formation of the novel secondary metabolites guatyromycine A and B, respectively. The reactions catalyzed by GymBs were confirmed with biochemical assays using recombinant proteins of all six candidates. As the intramolecular coupling is the same reaction performed by CYP121 from *Mycobacterium tuberculosis*, the corresponding gene cluster was also expressed heterologously in the same manner. However, CYP121 merely catalyzes the formation of mycocyclosin, indicating that GymBs might have evolved from CYP121 and slightly changed during evolution.

In the third project, I contributed to the elucidation of a BGC from *Streptomyces aurantiacus*, coding for the CDPS SasA, the PT SasB, and the MT SasC. It was proven that the *sasABC* gene cluster is responsible for the formation of streptoazine C. The involved PT SasB catalyzes two regular prenylations at both tryptophan residues within *cyclo*-L-Trp-L-Trp. By incubation with other CDPs and dehydrogenated CDPs, it was shown that SasB possesses a broad substrate flexibility and can convert at least eight other CDP derivatives efficiently.



## Zusammenfassung

Cyclodipeptide (CDPs) mit einem 2,5-Diketopiperazin (DKP) Grundgerüst sind weit verbreitet in unterschiedlichen Lebensformen, von einfachen Bakterien und Pilzen, bis hin zu komplexeren Arten, wie Pflanzen und Tieren. Sie besitzen vielfältige biologische und pharmakologische Effekte, zum Beispiel antibiotische, antifungale und antiproliferative Wirkungen. In Mikroorganismen werden CDPs von zwei verschiedenen Enzymfamilien gebildet, entweder von nichtribosomalen Peptidsynthetasen (NRPSs), die vor allem in Pilzen vorkommen, oder von Cyclodipeptidsynthasen (CDPSs), die sich hauptsächlich in Bakterien finden lassen. NRPSs, welche CDPs herstellen, sind vergleichsweise große (~2500 Aminosäuren), bi-modulare Enzyme, die freie Aminosäuren als Substrate nutzen. CDPSs sind hingegen kleiner (200 – 300 Aminosäuren) und benötigen bereits aktivierte Aminoacyl-tRNAs zur Knüpfung von Peptidbindungen. Die Bildung des DKP-Rings führt zu einer erhöhten Stabilität der CDPs gegen Proteolyse im Vergleich zu ihren nicht cyclischen Verwandten. Dies ermöglicht eine Vielzahl faszinierender Reaktionen, die von modifizierenden Enzymen katalysiert werden. Deren genetische Information befindet sich meistens in direkter Nachbarschaft zu der von Rückgrat-Enzymen, wie CDPSs. Daher spricht man von biosynthetischen Genclustern (BGCs). In CDPS-abhängigen Biosynthesewegen umfassen solche modifizierenden Enzyme Cyclodipeptidoxidasen (CDOs), Cytochrom P450 (P450) Enzyme,  $\text{Fe}^{II}/2\text{-Oxoglutarat}$  abhängige ( $\text{Fe}^{II}/2\text{-OG}$ ) Oxidasen, sowie Methyl- (MTs) und Prenyltransferasen (PTs). In dieser Doktorarbeit wurden acht solcher BGCs durch genomgestützte Analysen entdeckt und deren Funktionen mittels heterologer Expression und biochemischen Untersuchungen entschlüsselt.

Im ersten Projekt wurde ein BGC aus *Streptomyces cinnamoneus* zur detaillierten Untersuchung ausgewählt und *gtm* Gencluster benannt. Es besteht aus fünf Genen, die für vier Enzyme codieren – eine CDPS (GtmA), eine CDO (GtmBC), ein P450 Enzym (GtmD) und eine  $\text{Fe}^{II}/2\text{-OG}$  Oxidase (GtmE). Die Gene wurden in verschiedenen Kombinationen in den replikativen Vektor pPWW50A kloniert und in *Streptomyces albus* J1074 (*S. albus*) heterolog exprimiert. Mittels LC-MS und NMR Spektroskopie konnte gezeigt werden, dass GtmA *cyclo*-L-Trp-L-Met synthetisiert, GtmBC eine Doppelbindung am DKP auf Seite des Methioninrests einfügt, GtmD ein Guanin auf den Tryptophanrest überträgt und GtmE eine zweite Doppelbindung auf der gegenüberliegenden Seite des DKP bildet. Zusammen führt diese Kaskade zur Entstehung des neuen Sekundärmetaboliten Guatripmethine C. Die Bildung der zweiten Doppelbindung stellt eine neuartige Reaktion für  $\text{Fe}^{II}/2\text{-OG}$  Oxidasen in CDPS abhängigen Biosynthesewegen dar. Daher wurde diese Reaktion biochemisch genauer untersucht, indem Enzymassays mit rekombinantem Protein durchgeführt wurden. So konnte bestätigt werden, dass GtmE tatsächlich die Umsetzung des Vorläufers Guatripmethine A zu dem Endprodukt Guatripmethine C katalysiert. Für das stabile Isomer Guatripmethine B wurde hingegen nur eine sehr geringe Umsetzung beobachtet. Diese experimentelle Beobachtung konnte durch quantenchemische Berechnung mittels Dichtefunktionaltheorie gestützt werden.

In Kooperation mit Dr. Jing Liu, konnte im zweiten Teil der Arbeit ein weit verbreiteter Genlokus mit den beiden Genen *gymAB* in 47 verschiedenen Aktinobakterien identifiziert werden. *GymA* kodiert für eine CDPS und *gymB* für ein P450 Enzym. Letzteres ist eng mit CYP121, einem überlebenswichtigen Enzym

## ZUSAMMENFASSUNG

aus *Mycobacterium tuberculosis*, verwandt. Sechs *Streptomyces* Arten wurden als Kandidaten zur Aufklärung dieser BGCs ausgesucht. Analog zum ersten Projekt wurden ihre Gene in pPWW50A kloniert und in *S. albus* überexprimiert. Die Untersuchung der Extrakte mittels LC-MS in Kombination mit NMR Analysen der aufgereinigten Produkte ergab, dass alle sechs CDPSs *cyclo*-L-Tyr-L-Tyr (cYY) als Hauptprodukt bilden. Die P450 Oxidasen katalysieren im Anschluss zwei verschiedenartige Reaktionen – entweder die Knüpfung einer intramolekularen Bindung innerhalb von cYY, wobei Mycocyclosin entsteht, oder den intermolekularen Transfer der Nukleinbasen Guanin bzw. Hypoxanthin, der zur Bildung der neuentdeckten Sekundärmetabolite Guatyromycine A bzw. B führt. Die GymB katalysierten Reaktionen wurden des Weiteren durch biochemische Untersuchungen mit rekombinant hergestellten Proteinen aller sechs Kandidaten bestätigt. Die gleiche intramolekulare Bindung knüpft auch CYP121 aus *Mycobacterium tuberculosis*, daher wurde das entsprechende Gencluster in gleicher Weise exprimiert. CYP121 bildet hingegen ausschließlich Mycocyclosin, was dafürspricht, dass die GymBs womöglich aus CYP121 hervorgegangen sind und sich im Laufe der Evolution leicht verändert haben.

Im dritten Projekt war ich an der Aufklärung eines BGC aus *Streptomyces aurantiacus* beteiligt. Das BGC kodiert für die CDPS SasA, die PT SasB und die MT SasC. Wir konnten beweisen, dass das *sasABC* Gencluster für die Bildung von Streptoazine C verantwortlich ist. Die daran beteiligte PT SasB prenyliert in regulärer Weise beide Tryptophanreste in *cyclo*-L-Trp-L-Trp. Durch Inkubation mit anderen CDPs und dehydrogenierten Derivaten konnte gezeigt werden, dass SasB eine breite Substratspezifität besitzt und mindestens acht zusätzliche CDP-Derivate effizient umsetzen kann.

# 1. Introduction

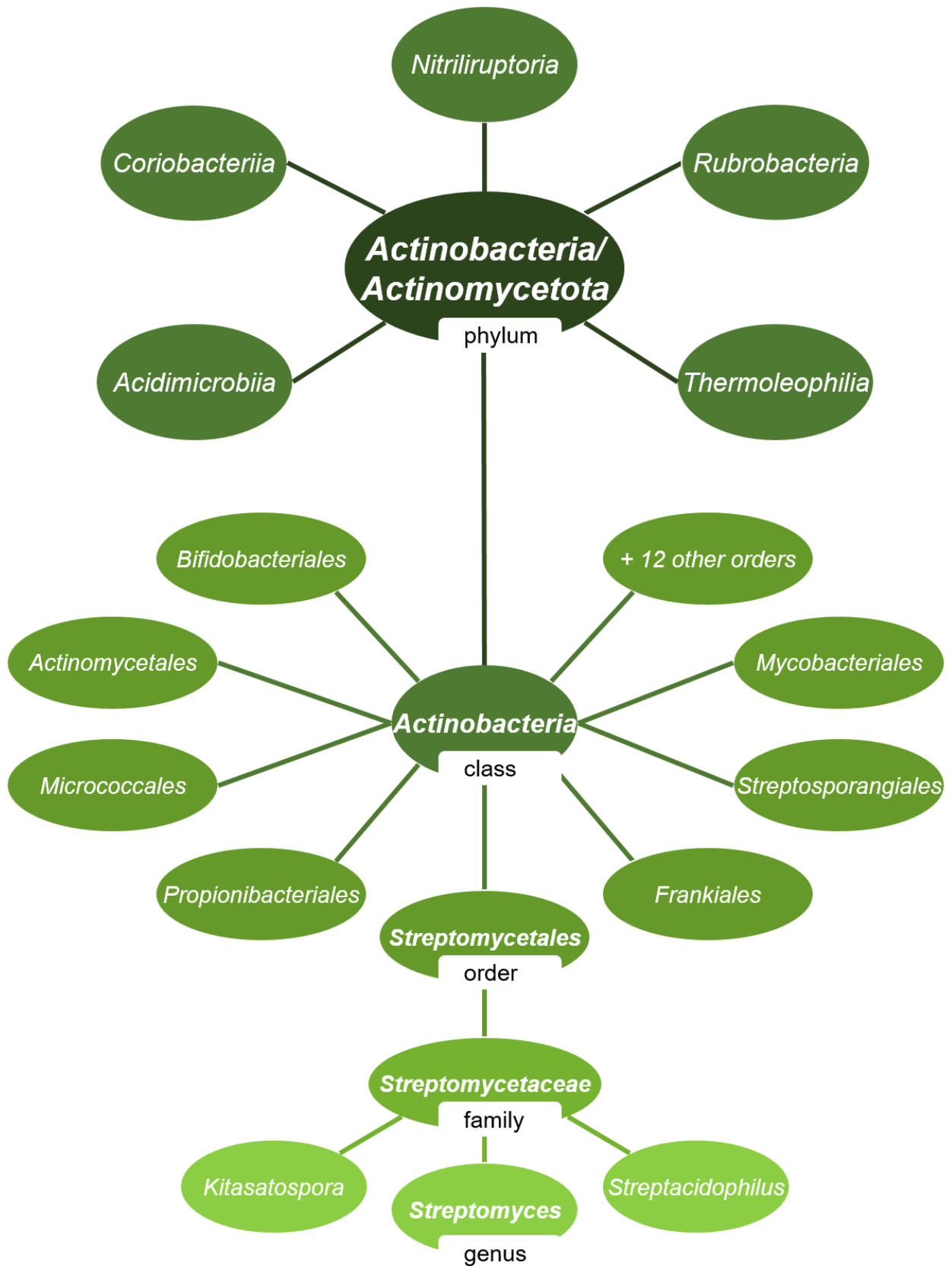
## 1.1 *Streptomyces* – their biology and impact on the environment and humankind

### 1.1.1 Taxonomy overview

On the highest taxonomic rank, life can be divided into three domains – Bacteria, Archaea and Eukarya.<sup>1, 2</sup> The first two are prokaryotic microorganisms without a cell nucleus, which distinguishes them from eukaryotes. The latter carry their genetic heritage inside a nucleus as multiple, linear chromosomes enclosed by a cell membrane, except for mitochondrial and chloroplast DNA. On the contrary, the DNA of Bacteria and Archaea usually forms a single, circular chromosome.<sup>3</sup> The classification into these three domains is based primarily on comparison of ribosomal RNA-sequences, mostly the subunits 16S-rRNA for Bacteria and Archaea, and their eukaryotic counterpart 18S-rRNA.<sup>4</sup> They also differ in phenotypic properties. For instance, the cell wall of the most bacteria consists of peptidoglycan with muramic acid as unique part. In contrast, Archaea have a different cell wall chemistry containing polysaccharides, proteins or glycoproteins. Some eukaryotes possess a cell wall, which is built of either cellulose or chitin.<sup>3</sup> Further differences are found in the linkage of membrane lipids; Bacteria and Eukarya use mostly linear fatty acids connected via ester bonds onto glycerin as backbone, Archaea on the other hand have branched hydrocarbons bound onto glycerol via ether bonds. Besides that, archaic and eukaryotic RNA-polymerases are both more complex and versatile compared to the single type of RNA-polymerases, which all bacteria share. Moreover, other differences can be found in the protein biosynthesis, the presence of introns and operons in genome sequences, occurrence of histones, and many additional factors.<sup>3</sup>

All three domains are further divided into phyla.<sup>1</sup> Within Bacteria, the phylum of *Actinobacteria* (synonym: *Actinomycetota*) constitutes one of the largest taxonomic units and its members are distributed ubiquitously in both terrestrial and aquatic habitats.<sup>5-7</sup> Various bacteria with versatile lifestyles and vital impacts on ecosystems, agriculture and medicine belong to the *Actinobacteria*, most notably soil inhabitants (*Streptomyces* and *Micromonospora* species), pathogens (*Mycobacteriaceae*, *Corynebacteriaceae* and *Propionibacteriaceae*), as well as gastrointestinal or plant symbionts (*Bifidobacteriaceae* and *Frankia* species).<sup>7</sup> On the next level, the phylum is subdivided into the six classes *Actinobacteria*, *Acidimicrobiia*, *Coriobacteriia*, *Nitrospirae*, *Rubrobacteriia*, and *Thermoleophilina*.<sup>8, 9</sup> The class of *Actinobacteria* is by far the largest one and contains 20 orders with 53 of the total 74 families belonging to the phylum of *Actinobacteria*.<sup>8</sup> Among these, the order of *Actinomycetales* harbors the family *Streptomycetaceae*, mainly containing the genera *Streptomyces*, *Kitasatospora*, and *Streptacidophilus*.<sup>8, 10</sup>

## INTRODUCTION



**Figure 1.** Taxonomic overview of the genus *Streptomyces* belonging to the phylum of *Actinobacteria* (syn.: *Actinomycetota*)

## INTRODUCTION

### 1.1.2 Occurrence and properties

Various species of *Streptomyces* occur ubiquitously in the environment. Most of them live terrestrially, especially in the soil, though some have adopted to a marine lifestyle.<sup>3, 11</sup> It is believed that the first *Streptomyces* species originated around 400 million years ago together with the colonization of the earth by vascular plants.<sup>12</sup> This indicates that they already played a key role in the spreading of plants by decomposing early plant material and thus forming the primeval soil. Up to now, *Streptomyces* are of vital importance for ecosystems due to their abilities to degrade cell walls and surface compounds of plants, fungi and insects.<sup>13, 14</sup> The distinct earthy smell of soil and forests is mainly caused by a mixture of geosmins produced and secreted by *Streptomyces* species.<sup>15</sup>

Similar to fungi, *Streptomyces* are filamentous and grow a branched mycelium into the substrate and aerial hyphae, which mature into spores.<sup>16</sup> Their development is characterized by apical tip extension instead of division by binary fission.<sup>17</sup> During the process of sporulation, cross-walls are formed by septation, leading to chains of unigenomic spore compartments resembling a ladder.<sup>18, 19</sup> This growth mechanism makes them a very rare example for multicellular bacteria.<sup>3</sup> The spores are able to persist for long periods and can survive most challenging conditions.<sup>20, 21</sup> Moreover, they allow this otherwise immobile genus to spread into new habitats.<sup>17</sup>

Like all other *Actinobacteria*, *Streptomyces* are gram positive and aerobic.<sup>3</sup> Glycolysis or the hexose monophosphate shunt way are used primarily for energy generation, although a broad spectrum of other carbon sources can be utilized.<sup>22, 23</sup> Nitrogen assimilation works almost exclusively via the glutamine synthetase in combination with the glutamine 2-oxoglutarate aminotransferase (syn.: glutamate synthase).<sup>24, 25</sup> This pathway is energetically more expensive compared to the common bacterial glutamate dehydrogenase, but enables nitrogen fixation at low ammonium concentrations.<sup>22</sup> Nevertheless, most *Streptomyces* are also able to use the latter enzyme at higher ammonium concentrations.<sup>25</sup>

The genomes of *Streptomyces* species stand out with one of the highest proportion of guanine and cytosine residues of all known bacteria (66 – 78 mol %).<sup>26</sup> Most of them have a comparably large size of around 8 Mb (6.7 – 12.3 Mb) with 5.000 – 10.000 genes of which about 2.000 are present in the central genomic region of each species.<sup>12, 27</sup> *Streptomyces* possess linear chromosomes, which is very unusual for bacterial DNA and can also be found in other members of the order *Actinomycetales* with a mycelial growth, like *Saccharopolysporae*, *Actinoplanes*, *Micromonosporae*, *Nocardiae*, and *Streptoverticilliae*.<sup>27-30</sup> In contrast, members with a non-mycelial growth, like *Corynebacteriae*, *Mycobacteriae*, or *Rhodococci*, have the common circular DNA with significantly smaller sizes.<sup>27</sup> The telomere sequences of *Streptomyces* are highly conserved and contain several tightly-packed palindromic sequences that are able to form stable secondary structures.<sup>29</sup> In addition, the chromosomal ends are capped by terminal proteins that are covalently bound to the 5' ends of the DNA.<sup>31</sup> These terminal proteins are encoded on a gene located terminally on a chromosome in one operon together with a second gene for an equally conserved telomere-associated protein. Both proteins are essential for the end patching of telomere overhangs during the replication of linear chromosomes and are as of today's knowledge unique among bacteria.<sup>32</sup> However, initiation of DNA amplification follows the same mechanism as for other bacteria, which is mediated by the centrally

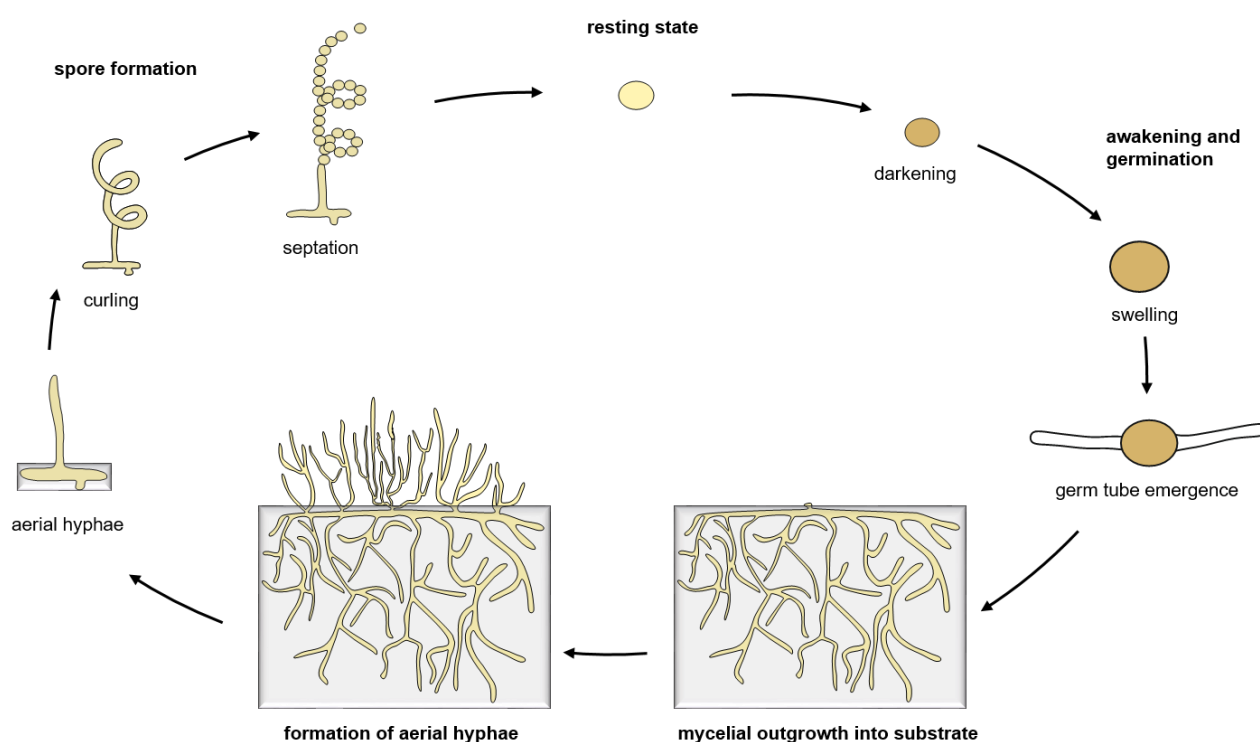
## INTRODUCTION

located origin of replication (*oriC*).<sup>33, 34</sup> Along with linear chromosomes, plasmids are also present in many *Streptomyces* species.<sup>29</sup> Accessory genes, located at the terminal regions, are highly unstable and often subject to replacement or deletion, leading to a broad genetic diversity.<sup>29</sup> The genetic exchange by conjugal DNA transfer, mainly under control of TraB proteins, is comparably easy in *Streptomyces*.<sup>35</sup> Initial plasmid transfer is usually followed by distribution of the plasmid into large parts of the mycelium.<sup>36</sup>

### 1.1.3 Lifecycle

In general, the lifecycle of *Streptomyces* species can be divided into the following stages: The resting or dormant state, the awakening and germination, the mycelial outgrowth into substrate, the formation of aerial hyphae, and the sporulation (Fig. 2).<sup>7, 37</sup>

During the resting state, spores of *Streptomyces* are able to survive most hostile environmental conditions, such as heat, freezing, desiccation, radiation, or oxidative stress.<sup>16, 37</sup> The metabolism is limited from no up to minimal activity, with the possibility to oxidize atmospheric H<sub>2</sub> to supply maintenance energy.<sup>38</sup> The genetic heritage is safeguarded in haploid state together with other macromolecules needed for future launch of germination, protected by a thick spore wall.<sup>37</sup> It is constructed of cytoskeletal actin-like proteins MreB and Mbl forming a stable rodlet layer.<sup>39-41</sup> Lack of water ensures resistance to thermal extremities. DNA and proteins are stabilized with trehalose, a non-reducing disaccharide consisting of two glucose molecules. Trehalose replaces water in the process of desiccation by formation of hydrogen bonds between membranes, nucleic acids, and proteins, leading to a gel-like phase preventing aggregation.<sup>37, 42</sup> Furthermore, it serves as a first nutrient source and eases the attachment of chaperons to proteins in the subsequent phase of germination.<sup>43</sup>



**Figure 2.** Lifecycle of *Streptomyces*, modified on the basis of Barka *et al.*<sup>7</sup> and Bobek *et al.*<sup>37</sup>

## INTRODUCTION

To trigger the awakening of spores, the environment has to be at least aqueous.<sup>44</sup> Nutrient sources, warmth, mechanical disruption, and remains of peptidoglycans accelerate this process.<sup>37, 45, 46</sup> The first step of germination is characterized by a darkening of the spores, as bivalent cations are dissolved and no longer reflect light.<sup>47, 48</sup> The spores are unfolding, which is mediated by calcium-binding protein CabC and facilitated by lysozyme-like hydrolases RpfA and SwIA–C.<sup>21, 49</sup> The spore walls lose hydrophobicity, enabling water influx, as well as absorption of extracellular nutrients. The continuous water influx causes swelling of the spores. Trehalose is gradually diluted and triggers trehalase activity to release glucose for energy supply.<sup>50</sup> The decreasing concentration of trehalose enables refolding of proteins and ribosomes into their active forms with the help of chaperons.<sup>51</sup> Re-activation of the protein biosynthesis starts with translation from a safeguarded mRNA stock, followed by DNA replication.<sup>16, 43, 52</sup> The germination is finalized by the apical emergence of germ tubes out of a spore.<sup>53</sup> This vegetative growth is mediated by various proteins, e.g. SsgA for localization of the correct sites for emergence, DivIA for new cell wall outgrowth, FilP as cell growth associated protein, and FtsZ as cell division associated protein.<sup>54-57</sup> The concentration of ATP is increased by ATP synthases serving as both energy source and precursor for the signal molecule cAMP.<sup>58-60</sup> Furthermore, expression of stress responding proteins, like sigma factors, is activated to control stress caused by the entry of water, oxygen, radicals, or salts.<sup>37</sup>

The development of the vegetative mycelium into the substrate is characterized by branching and apical tip extension of hyphae, forming a tangled network of filamentous cells. As the mycelium expands, cross-walls, so called vegetative septa, are built behind the growing tips.<sup>16</sup> The septation into compartments makes the mycelium more resilient against damage by phages or mechanical injury.<sup>61</sup> Next, the secondary metabolism is activated leading to production and secretion of various substances. These include inhibitors for further germination, e.g. germicidins or hydnosin, signaling molecules for the communication in plant-microbe symbioses, like chalcones, flavonoids and other phenolic structures, as well as antibiotics to protect against competitors, such as neomycin, streptomycin, and many more (see chapter 1.1.5).<sup>37, 62-64</sup>

With the aging of vegetative colonies and depletion of nutrients, single hyphae start to grow vertically into the air and spread parasitically on the substrate mycelial biomass. The aerial hyphae are surrounded by a specialized outer sheath with a hydrophilic part in contact with hydrated hyphal walls and a hydrophobic phase towards the air.<sup>16</sup> It is mostly built of the hydrophobic cell surface proteins chaplins and rodlines, clustered into paired rodlet structures.<sup>65, 66</sup> The hydrophobic layer decreases surface tension of water and thus enables erected apical growth.<sup>67</sup> A kind of programmed cell death of non-sporulating parts of the mycelium provides nutrients to support aerial growth.<sup>16</sup> Glycogen and polyphosphates are accumulated near the upper surface.<sup>68</sup> During this period, the secondary metabolism is most active, preventing that released nutrients and damaged parts are depleted by competitors.<sup>69, 70</sup> The subsequent process of spore formation and maturation is controlled by a complex interaction between diverse proteins operating in cascades. Of note are members of the SALP- (Ssg) and the Whi-family of which most are unique to *Actinobacteria*.<sup>71</sup> WhiA and WhiB seem to control the stop of aerial growth and initiate sporulation, since mutants without grow longer aerial hyphae without spores.<sup>72-74</sup> Conversion of glycogen to trehalose, which is needed to replace water, as well as for energy storage, is specifically dependent on WhiG.<sup>16</sup> SsgA increases the de-novo peptidoglycan synthesis to

## INTRODUCTION

build septa at the interphases, SsgG ensures the correct localization of separation, and SsgB activates FtsZ proteins.<sup>54, 75, 76</sup> FtsZ assembles the so-called Z-ring at the new site of cell division.<sup>77</sup> Contraction of the Z-ring leads to constriction into unigenomic pre-spore compartments that resemble ladder-like spirals. Spore wall thickening is directed by SsgD, DNA condensation by WhiI and pigmentation by other Whi members.<sup>16, 76</sup> Finally, the matured spores are separated by the cell wall hydrolases SsgE and SsgF and are dispersed into the environment.<sup>76</sup>

### 1.1.4 Impact on the environment

*Streptomyces* are abundantly present in soil. It is thus unsurprising that they play vital roles for ecosystems. Their ability to produce various hydrolytic enzymes like cellulase or chitinase enables them to degrade natural macromolecules from (mostly) dead plants, fungi, and insects, e.g. cellulose, chitin, xylan, and agar.<sup>13</sup> In addition to carbon recycling, *Streptomyces* facilitate the uptake of nutrients by nitrogen fixation, solubilization of minerals, as well as synthesis of siderophores. Plant growth can be stimulated by secretion of phytohormones.<sup>7</sup>

Although most *Streptomyces* species live in beneficial symbiosis with endophytes, some rare examples are plant pathogens. *S. scabies* or *S. acidiscabies* can infect tuber of potatoes, carrots, or beets, causing the common scab or acid scab, respectively.<sup>78</sup> Secretion of thaxtomins, cyclodipeptides with a 2,5-diketopiperazine backbone (see section 1.3), enables tissue penetration and suppresses defensive mechanisms of plants.<sup>79</sup> Infection of sweet potatoes by *S. ipomoea* can lead to significant yield loss and *S. parvus* can trigger early decay and death of tree branches in maple trees.<sup>7</sup>

### 1.1.5 *Streptomyces* and their irreplaceable importance for modern medicine

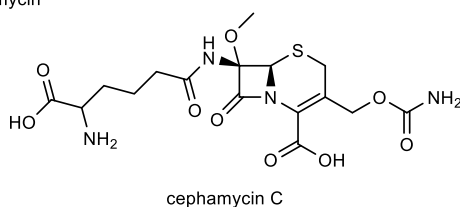
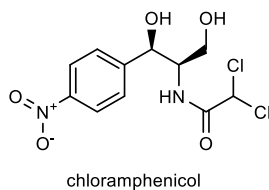
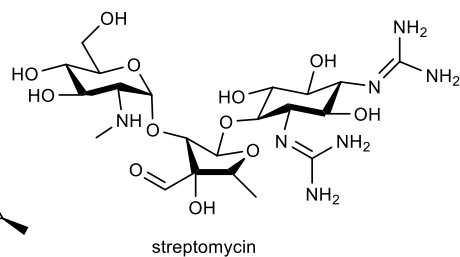
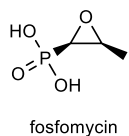
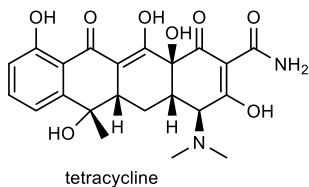
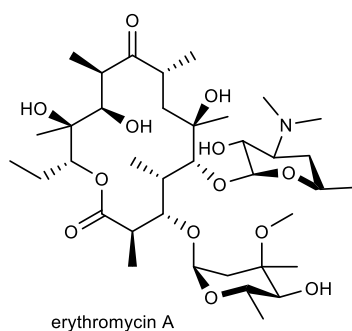
The impact of natural products (NPs) and their derivatives from *Actinobacteria*, especially *Streptomyces*, on the treatment of diseases is unequivocally immense.<sup>80</sup> Life expectancy increased significantly in the middle of the 20<sup>th</sup> century and otherwise life threatening diseases became curable.<sup>81</sup> Over 10.000 bioactive compounds have been characterized from *Actinobacteria*, out of which about three quarters are derived from *Streptomyces* species, making them the most prolific source together with fungi.<sup>82</sup> Since the discovery of actinomycin, streptothricin and streptomycin in the early 1940s, many antibiotics and whole classes thereof have been isolated from *Streptomyces* species, especially during the golden era between 1940 and 1960.<sup>83</sup> Major antibiotic classes produced by *Streptomyces* or chemically modified afterward, include: macrolides (erythromycin, clarithromycin), aminoglycosides (streptomycin, kanamycin, neomycin), tetracyclines, chloramphenicol, fosfomycin,  $\beta$ -lactams (cephamycins), as well as  $\beta$ -lactamase inhibitors (clavulanic acid).<sup>84-92</sup> Apart from antibiotics, several other secondary metabolites with pharmacological effects originating from *Streptomyces* include antifungals, such as amphotericin B, pentamycin, and natamycin, antiparasitics, e.g. avermectin and prodiginine, antitumor agents, like bleomycin, daunorubicin, and doxorubicin, immunosuppressive agents, e.g. rapamycin and tacrolimus, as well as insecticides (see Fig. 3 for structures).<sup>93-102</sup>

Besides the beneficial effects of *Streptomyces* on medicine, very few species are pathogenic to humans. *S. somaliensis* and *S. sudanensis* can cause actinomycetoma and the inhalation of spores is linked to respiratory diseases like farmer's lung.<sup>103-105</sup>

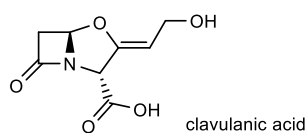


## INTRODUCTION

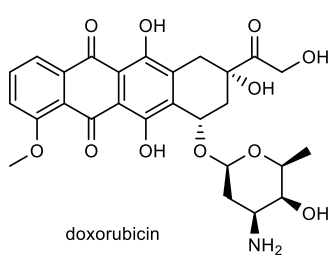
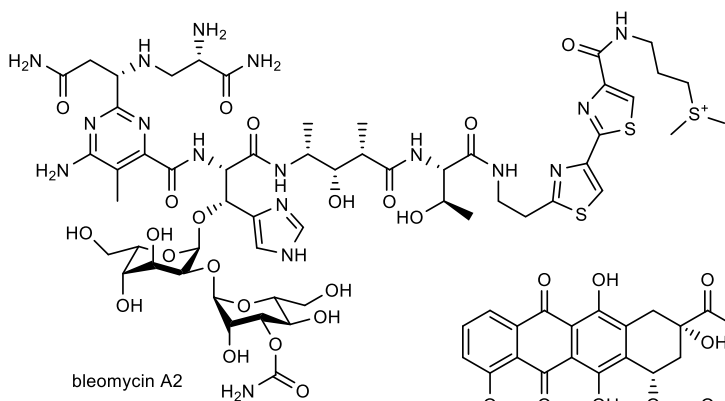
### antibiotics



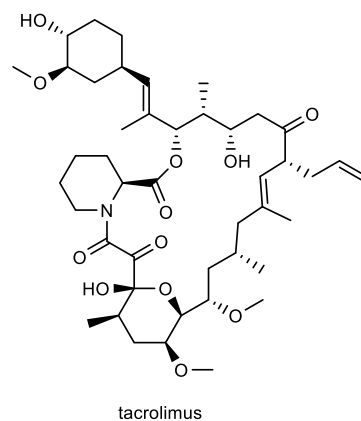
### $\beta$ -lactamase inhibitor



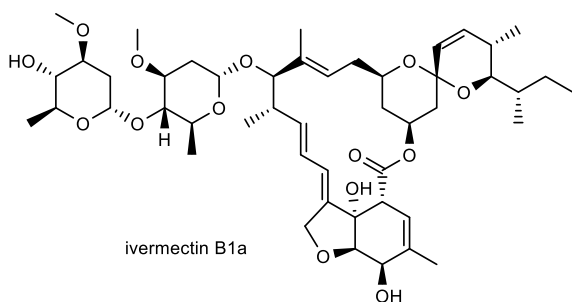
### antitumor agents



### immunosuppressant



### antiparasitic agent



### antifungal agent

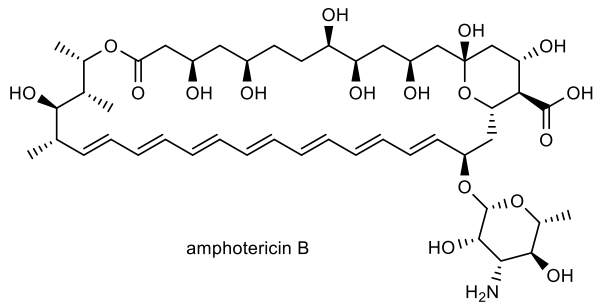


Figure 3. Bioactive natural products isolated from *Streptomyces* species

## 1.2 Elucidation of biosynthetic pathways in *Streptomyces*

### 1.2.1 Genome mining for biosynthetic gene clusters

As stated previously, secondary metabolites from *Streptomyces* are of indispensable value for medicine. With growing resistance against widely used antibiotics, it is of highest importance to find compounds with new action modes against multi-resistant pathogens.<sup>106</sup> It was estimated that up to 90 % of the biosynthetic ability has not been elucidated so far.<sup>107</sup> This means that roughly 150.000 bioactive secondary metabolites alone from *Streptomyces* might yet be undiscovered.<sup>108</sup> The information for the biosynthesis of novel NPs lies hidden in the genome of the bacteria, precisely clustered in biosynthetic gene clusters (BGCs), a defined region on the microbial chromosome.<sup>109</sup>

Identification of BGCs can be achieved by genome mining, which describes the exploitation of genomic information for the discovery of new compounds or targets. Progress in DNA sequencing in combination with isolation of new strains from different habitats, especially marine sources, led to an exponentially increasing number of microbial genomic sequences deposited in public databases.<sup>110</sup> The available sequence data can be analyzed using bioinformatical software, like BLAST or antiSMASH, to arrange a genome and predict its containing BGC by comparison with already characterized clusters. Similar to other algorithms, these tools get more precise with each elucidated cluster and are continuously updated.<sup>111</sup> Regarding *Streptomyces*, each sequenced genome contains about 20 to 40 BGCs.<sup>112</sup> Most of the BGCs are silent under laboratory conditions in their natural host, as their environmental signal for activation is absent and unknown.<sup>113</sup>

### 1.2.2 Activation of cryptic biosynthetic gene clusters

To activate cryptic BGCs in order to discover novel NPs, several methods have been developed during the past decades. In general, these can be divided into two different groups – pleiotropic and pathway specific approaches.<sup>114</sup> As the name describes, pleiotropic approaches have multiple effects that influence the expression of more than one gene. Their aim is to trigger a global change in regulation of metabolic pathways in order to find new secondary metabolites. Thus, a higher throughput can be achieved, but simultaneously a possible mixture of new compounds can impede the identification and isolation. These methods can succeed even without the knowledge of any BGC and are therefore rather empirical approaches. However, the increasing knowledge about certain factors influencing the gene expression provides the opportunity for a more specific activation. In contrary, pathway specific approaches require a more detailed expertise about specific BGCs in a species. They are usually more precise and enable single step elucidations. As each BGC needs a particular strategy for activation, the throughput is lower.<sup>114</sup>

Pleiotropic methods comprise the following:

- **variation of growth conditions**

Cultivation on/in different media, especially with different carbon or nitrogen sources, can change the secondary metabolite profile of the same strain significantly. Simple other changes, e.g. pH-value, temperature, oxygen supply (standing or shaking cultures), or cultivation time can have similar effects. Some studies also showed that the addition of rare elements or sub-lethal concentrations of antibiotics can lead to the production of new NPs.<sup>113, 114</sup>

## INTRODUCTION

- **co-cultivation**

Interspecies crosstalk between organisms occupying identical habitats can stimulate either defense mechanisms resulting in a (higher) production of antibiotics, antifungals, and herbicides, or the secretion of signaling molecules.<sup>115, 116</sup>

- **manipulation of global regulators**

The activation or suppression of a global regulator or regulatory system can increase overall transcription rates and protein levels, leading to production of several secondary metabolites. However, only few pleiotropic transcriptional regulators have been characterized until now.<sup>117, 118</sup>

- **engineering the transcription and translation machinery**

Upregulation of BGCs can be achieved by mutations or other influences on RNA polymerases (RNAP) or ribosomal proteins, e.g. by the RNAP-targeting antibiotic rifampicin or ribosome-targeting antibiotics, like streptomycin and gentamicin. This method is limited due to a small number of triggers in bacteria.<sup>119-121</sup>

- **epigenetic perturbation**

This approach can be used in eukaryotes to activate secondary metabolism as an answer to alteration of the chromatin structure by mutagenesis or use of small-molecule inhibitors, e.g. inhibition of DNA methyltransferase and histone deacetylase.<sup>122, 123</sup>

Pathway specific methods include:

- **manipulation of pathway specific regulators**

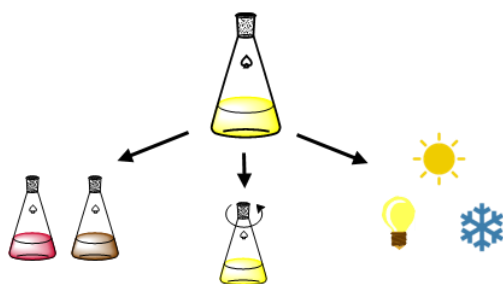
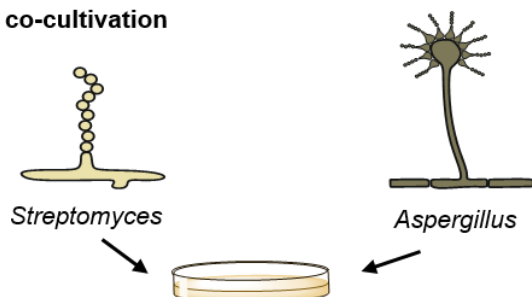
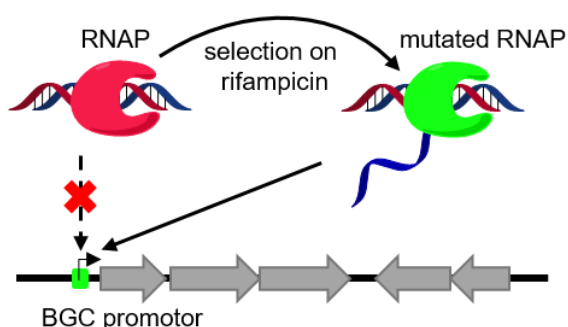
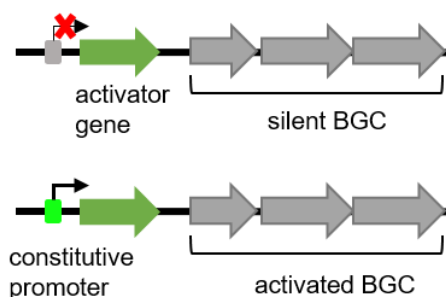
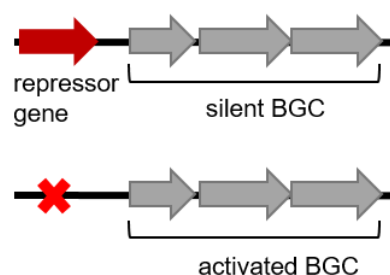
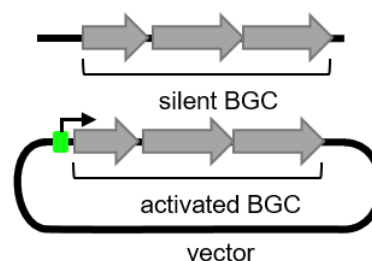
BGCs are often controlled by specific transcriptional activators or repressors. Such regulators can be identified by comparison with known ones using predictive bioinformatical software. Overexpression of an activator with a strong inducible promotor or deletion of a repressor gene can result in targeted activation of one specific BGC.<sup>124, 125</sup>

- **replacement of natural promoters**

Exchange of a natural promotor with a stronger constitutive or inducible promotor can increase or initiate the expression of a specific BGC. However, the natural host must be genetically manipulable.<sup>114</sup>

- **heterologous expression**

Complete cryptic BGCs or parts are cloned from the genome of a distinct species into an expression vector, where the genes are under control of a strong constitutive or inducible promotor. Afterward, the vector is transferred into a suitable host for expression.<sup>109</sup> Certain strains of *S. albus*, *S. coelicolor*, and *S. lividans* have been optimized for the heterologous expression of BGCs derived from other *Streptomyces* species.<sup>126-128</sup> The original host does not have to be amenable to genetic manipulation and a limited metabolic profile of the expression strains simplifies the detection of novel NPs.<sup>129, 130</sup> Complete biosynthetic pathways can be elucidated by different combinations of genes.<sup>131</sup> However, this approach is mainly suitable for smaller BGCs (<40 kbp) and problems may occur due to a lack of precursors, co-factors, or chaperons for protein folding, or else by not recognizing of promoters and ribosome-binding sites.<sup>109, 114</sup>

**pleiotropic methods****change of growth condition****co-cultivation****engineering the ribosomal machinery****pathway specific methods****expression of activator genes****deletion of repressor genes****heterologous expression**

**Figure 4.** Methods to activate BGCs, parts are modified after Rudledge and Challis (2015)<sup>114</sup>

### 1.3 Cyclodipeptides with a 2,5-diketopiperazine scaffold

#### 1.3.1 Chemical properties and pharmacological activities

Cyclodipeptides (CDPs) represent the smallest form of cyclic peptides. Condensation of two  $\alpha$ -amino acids via two *cis* amide bonds results in a 2,5-diketopiperazine (DKP) as central core.<sup>132</sup> The formation of a DKP ring provides a higher stability against proteolysis than acyclic dipeptides possess.<sup>133</sup> Furthermore, the central DKP core possesses 2H-bond acceptor and donor sites, which are often essential for binding to receptors and enzymes. Modifications can be introduced at all six positions of the DKP ring and stereochemistry controlled at four positions.<sup>133</sup> Thus, various NPs from bacteria, fungi, plants, and animals bear a DKP subunit – either alone or embedded in larger, more complex structures. Further diversity is achieved by chemical derivatization.<sup>133</sup> All these lead to NPs with a broad spectrum

## INTRODUCTION

of biological and pharmacological activities, ranging from interspecies communication as signaling molecules to antibiotic or antitumor effects exploited in medicine.<sup>134-138</sup> Some bioactive CDPs and derivatives have been chosen here as representatives (Fig. 5) and are grouped into NPs from *Streptomyces* species and fungi as well as synthetic derivatives.

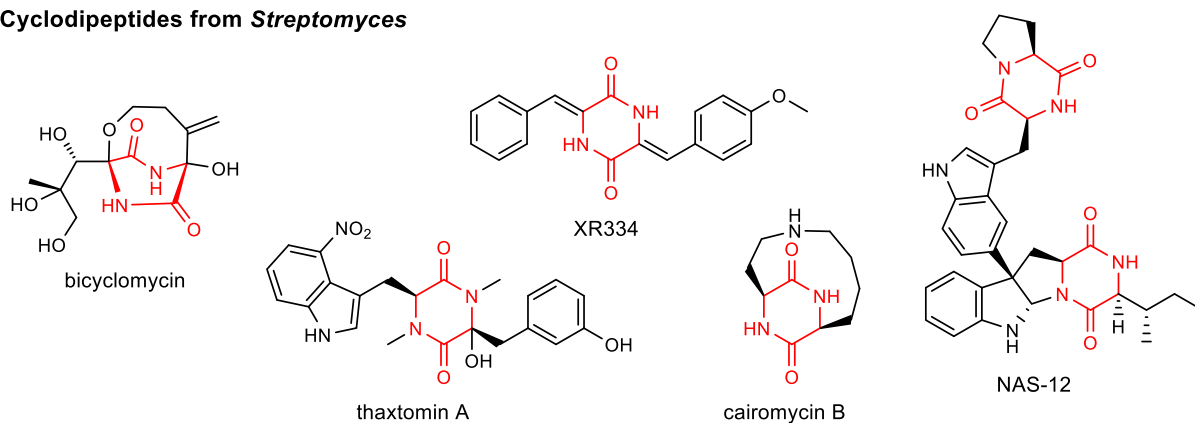
The *cyclo*-L-Leu-L-Ile backbone in bicyclomycin from *Streptomyces cinnamoneus* (formerly known as *Streptomyces sapporensis*) is altered by multistep hydroxylations carried out by a cytochrome P450 (P450) enzyme and several Fe<sup>II</sup>/2-oxoglutarate dependent oxidases.<sup>139, 140</sup> It is used as an antibiotic against Gram-negative bacteria and exhibits a unique mechanism of action inhibiting the transcription termination factor Rho.<sup>141</sup> Another CDP product from *Streptomyces* with antibiotic activity is the *cyclo*-L-Asp-L-Lys derivative caiomycin B. It mainly shows effects against Gram-positive staphylococci.<sup>142</sup> XR334, a *cyclo*-L-Tyr-L-Phe descendant produced by *Streptomyces* species, inhibits the plasminogen activator inhibitor-1, a physiological regulator of fibrinolysis.<sup>143</sup> The plant pathogenic *S. scabies* and *S. acidiscabies* use thaxtomins to penetrate tissues and to suppress defensive mechanisms of plants.<sup>79</sup> Last but not least, some dimers of CDPs like NAS-12, formed by the P450 enzyme NascB from a marine *Streptomyces* species, show neuroprotective effects against glutamate induced damage.<sup>144</sup>

Diverse fungal secondary metabolites derived from CDPs with a DKP backbone have been elucidated, most of them built of an aromatic amino acid.<sup>145</sup> Roquefortine C is a common fungal metabolite produced by numerous *Penicillium* species. It is found in the veins of blue cheeses and originates from *cyclo*-L-Trp-L-His. The amounts of roquefortine C in cheese are considered as non-hazardous for human consumption. However, high doses of the purified compound are classified as neurotoxic due to an inhibition of mammalian P450 oxidases.<sup>146</sup> Gliotoxin, a CDP derivative of *cyclo*-L-Phe-L-Ser, is not only a potent mycotoxin, but also an important virulence factor for invasive aspergillosis by pathogenic *Aspergillus* species, e.g. *Aspergillus fumigatus*.<sup>147</sup> The disulfide bridge can interact with thiol residues of cell membranes, leading to suppression of immune cells, especially neutrophilic and eosinophilic granulocytes, as well as macrophages. In addition, several other enzymes are inhibited by gliotoxin, including NF- $\kappa$ B, farnesyl transferase I, and NADPH oxidases. All effects together result in a reduced immune response.<sup>148, 149</sup> FR106969, produced by *Penicillium citrinum*, is a similar metabolite with both of its sulfur atoms methylated. It exhibits high inhibitory potency against platelet activating factor (PAF) induced platelet aggregation and therefore might be used against anaphylaxis or to protect amyloid-beta damaged neurons.<sup>133</sup> Numerous fungal CDPs, especially from *Aspergillus* strains, bear a *cyclo*-L-Trp-L-Pro backbone, also known as brevianamide F.<sup>150, 151</sup> This includes tryprostatin B, a mammalian cell cycle inhibitor, and fumitremorgin C, an inhibitor of a multidrug-resistance protein in breast cancer, as well as the neurotoxic verruculogen.<sup>133, 152</sup> Avrainvillamide is an even more complex descendant with antibiotic activity against multidrug-resistant bacteria.<sup>133, 153</sup>

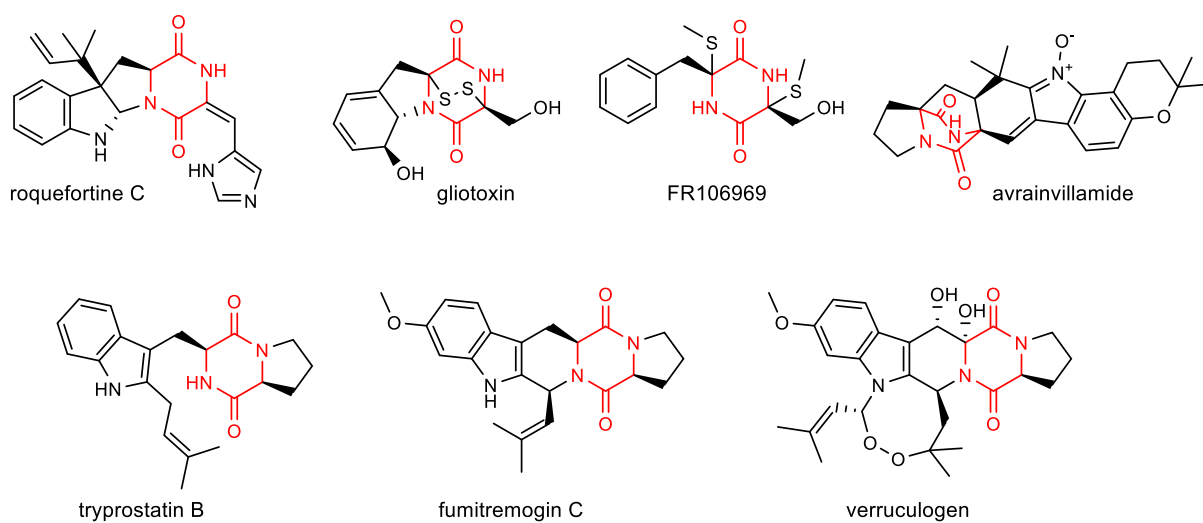
The variety of bioactive natural CDPs inspired chemical derivatization thereof or purely synthetic approaches to generate even more drug candidates with higher potency and less toxicity. Plinabulin, similar to the fungal metabolite phenylahistin, is a tubulin binding agent blocking the polymerization of tubulin. It is under investigation as a clinical candidate to prevent chemotherapy-induced neutropenia.<sup>154,</sup>

## INTRODUCTION

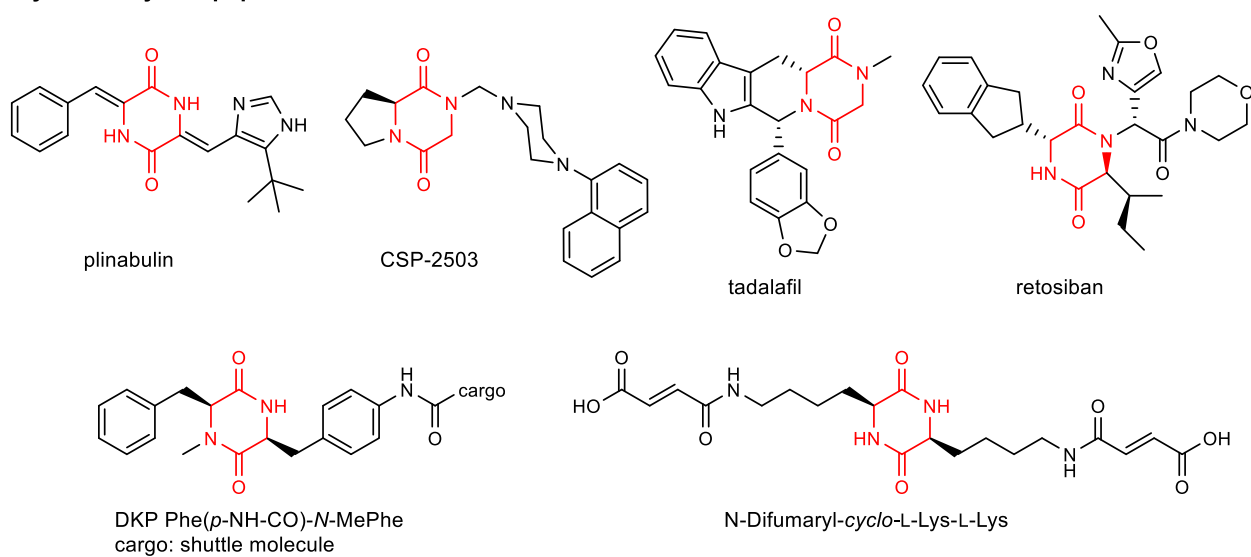
### Cyclodipeptides from *Streptomyces*



### Cyclodipeptides from fungi



### Synthetic cyclodipeptide derivatives



**Figure 5.** Bioactive CDP derivatives with a 2,5-DKP core

## INTRODUCTION

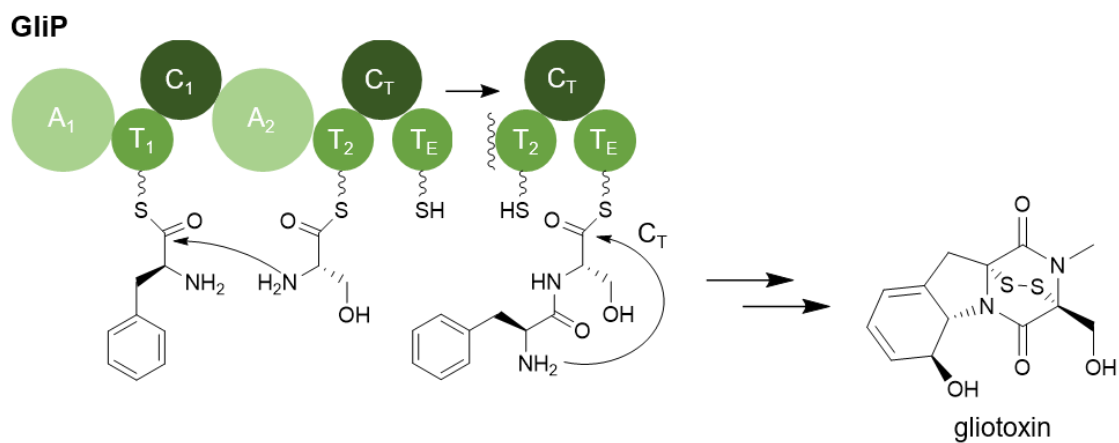
CSP-2503 shows anxiolytic effects as selective 5-HT<sub>1A</sub> agonist and 5-HT<sub>2A</sub>/5-HT<sub>3</sub> antagonist. Tadalafil is an approved drug to treat erectile dysfunction, benign prostatic hyperplasia, and pulmonary arterial hypertension. Its effect on vasodilation is caused by inhibition of phosphodiesterase type 5 and thus partly blocking the degradation of guanosine monophosphate.<sup>156, 157</sup> Retosiban is in development as a competitive oxytocin receptor antagonist to prevent oxytocin-mediated contraction of the uterus for the treatment of preterm labor.<sup>158</sup> Last but not least, some CDP derivatives are investigated as future drug carrier molecules. The linkage of dopamine or baicalin, both unable to cross the blood brain barrier, onto DKP Phe(*p*-NH-CO)-*N*-MePhe lets them surmount the barrier *in vitro*.<sup>159</sup> The inhalable insulin Afrezza® uses N-Difumaryl-*cyclo*-L-Lys-L-Lys as biomaterial to encapsulate the hormone.<sup>159</sup>

### 1.3.2 Biosynthesis of cyclodipeptides – NRPS vs. CDPS

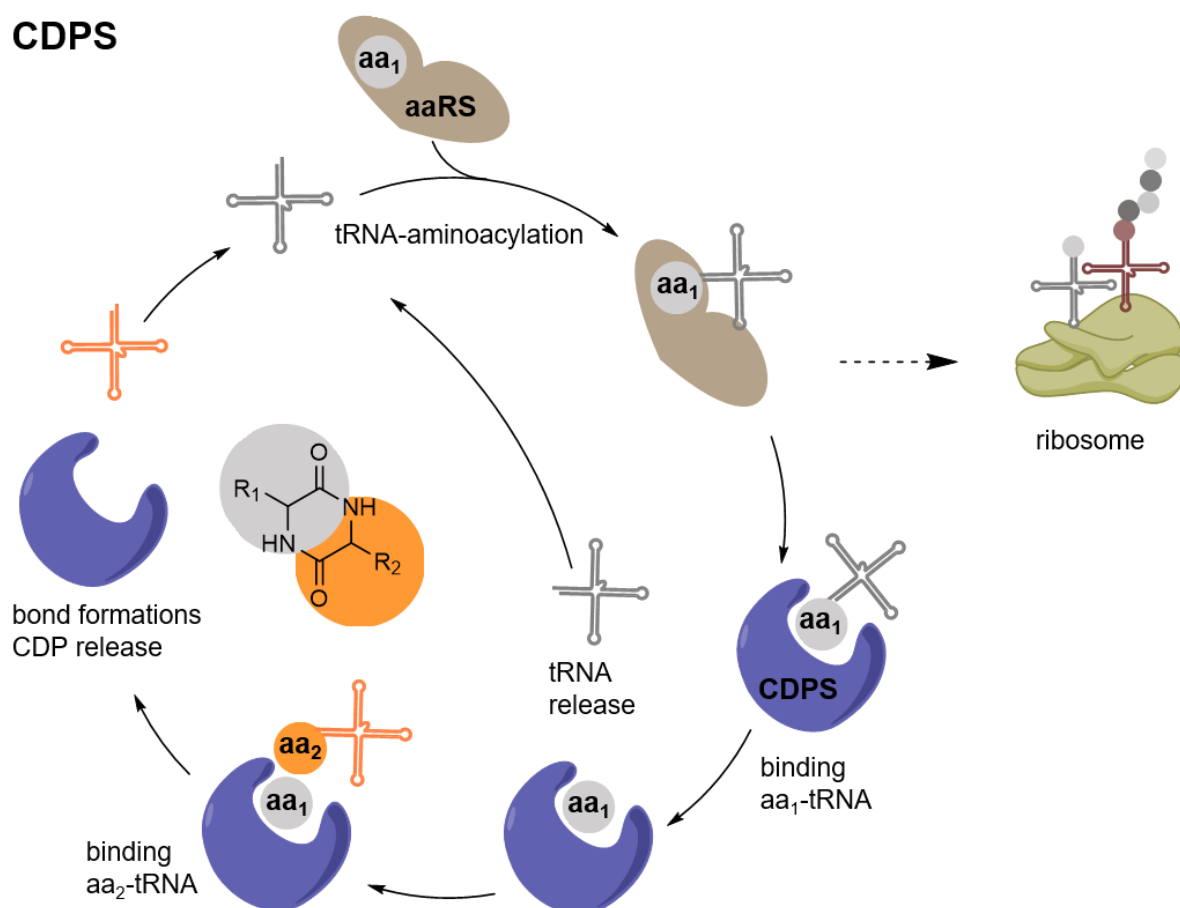
In nature, CDPs are formed by two different enzyme families – either by non-ribosomal peptide synthetases (NRPSs), commonly occurring in fungi, or by cyclodipeptide synthases (CDPSs), mainly found in bacteria.<sup>132, 160, 161</sup> NRPSs for CDP formation are comparably large bi-modular enzymes with a typical length of 2300 – 2500 amino acids, that are able to use free amino acids as substrates. They consist of three core domains, the adenylation (A)-, thiolation (T)-, also known as peptidyl carrier protein (PCP)-, and the condensation (C)-domain.<sup>160</sup> In the initial step of the CDP biosynthesis, the first amino acid is activated by the A<sub>1</sub>-domain via adenylation. Next, the resulting amino acyl intermediate is linked via thioester bond to a thiol group of the T<sub>1</sub>-domain. In analogy, another amino acid is loaded onto the second set of A<sub>2</sub>-, T<sub>2</sub>-domains. The first amino bond formation is accomplished by the C<sub>1</sub>-domain and the intermediate transferred to the T<sub>C</sub>-domain (synonym: T<sub>3</sub>-domain) and, again, tethered via thioester bond. Usually, a terminal thioesterase (TE)-domain cleaves the product from the NRPS. This domain is missing in NRPS forming the DKP ring of CDPs. Instead, the second C-domain is a conserved terminal condensation-like (C<sub>T</sub>)-domain with a distinct catalytic histidine residue. The C<sub>T</sub>-domain catalyzes both, the cyclization to the DKP core by formation of the second amide bond and the release of the CDP from the enzyme (Fig. 6).<sup>162</sup>

In contrast to NRPSs, CDPSs are smaller, consisting of only 200-300 amino acids, and require already activated amino acids to construct CDPs. Thus, they hijack aminoacyl tRNAs (aa-tRNAs), which have been originally activated by aminoacyl-tRNA synthetases (aaRSs) for protein biosynthesis on ribosomes.<sup>132, 163</sup> Each CDPS contains two surface-accessible pockets, P1 and P2, for binding of the amino acid residues. These pockets possess consensual motifs that allow precise prediction of the first incorporated amino acid and less accurate prediction of the second one.<sup>164, 165</sup> A sequential ping-pong mechanism has been suggested for the reaction, starting with the attachment of the first aa-tRNA and subsequent shuttle of its aminoacyl moiety to a conserved serine residue of the catalytic pocket. Thereby, the tRNA is released, enabling binding of the second aa-tRNA. Next, two peptide bonds are formed and the constructed CDP is released together with the second tRNA (Fig. 6).<sup>163</sup>

## NRPS



## CDPS



**Figure 6.** Mechanisms of CDP formation catalyzed by NRPS (upper part) and CDPS (lower part) modified after Belin *et al.* (2012) and Baccile *et al.* (2019).<sup>132, 162</sup>



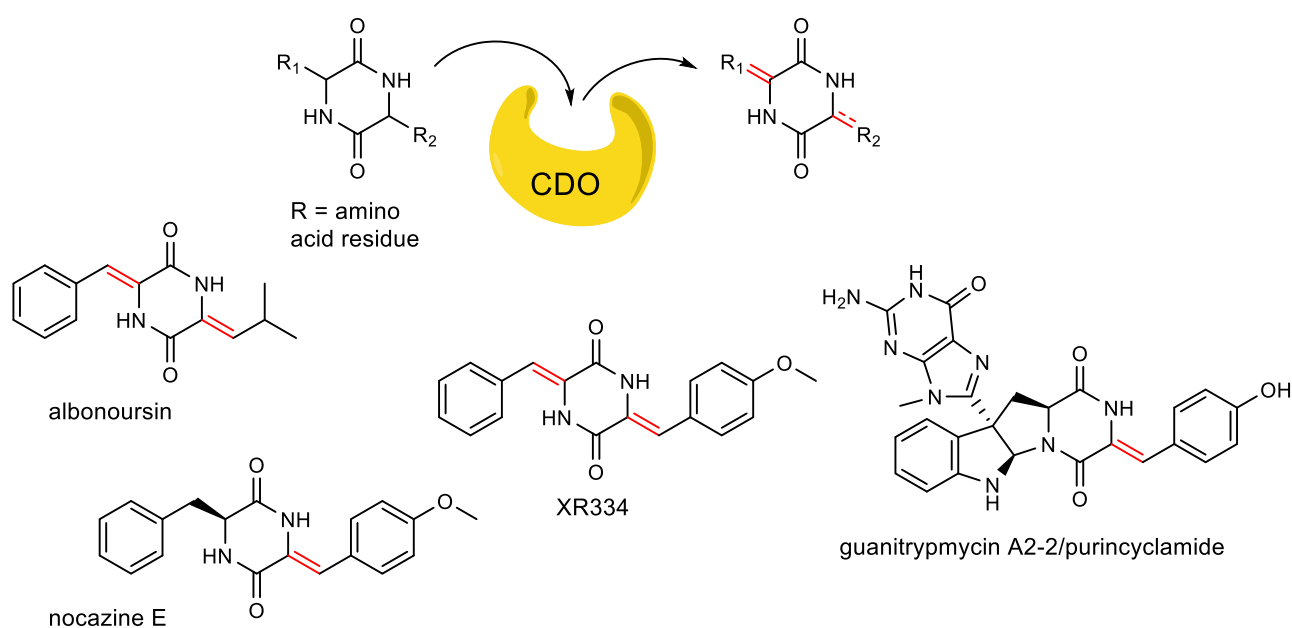
## 1.4 Tailoring enzymes associated to cyclodipeptide synthases

As mentioned above, the formation of the DKP ring as central core of the CDP results in a stable backbone.<sup>133</sup> This enables modifications by so-called tailoring enzymes that are encoded by genes in direct neighborhood to the backbone CDPS gene.<sup>161</sup> These include cyclodipeptide oxidases (CDOs), cytochrome P450 (P450s) enzymes, Fe<sup>II</sup>/2-oxoglutarate dependent (Fe<sup>II</sup>/2-OG) oxidases, as well as methyl- (MTs) and prenyltransferases (PTs). The enzyme families that are part of BGCs analyzed in this dissertation, are explained briefly in this section.

### 1.4.1 Cyclodipeptide oxidases

CDOs were the first discovered tailoring enzymes in CDPS-associated pathways.<sup>166, 167</sup> They consist of two subunits with each being approximately 21 and 11 kDa in size. These two subunits are encoded by individual genes, which often overlap by 20-30 nucleotides. The active enzyme is a heteropolymer with high molecular weight, in which both subunits together with a flavinic co-factor are embedded. It catalyzes C- $\alpha$ - $\beta$  dehydrogenation of CDPs using molecular oxygen.<sup>167-169</sup> A broad substrate specificity usually allows conversion of several different CDP scaffolds.<sup>170, 171</sup>

Albonoursin was the first described CDP (*cyclo*-L-Phe-L-Leu) formed by a CDPS, *i.e.* AlbC from *Streptomyces noursei* and *Streptomyces albus*.<sup>166, 168</sup> The CDO consisting of AlbA and AlbB installs both double bonds. The CDO Ndas 1146/1147, derived from a gene cluster of *Nocardioopsis dassonvillei*, catalyzes either one or two dehydrogenations in  $\alpha$ - $\beta$ -position, resulting in nocazine E and XR334, respectively.<sup>169</sup> Finally, the double bond at the tyrosine residue in guanitrypmycin A2-2 (syn: purincyclamide) is created by GutBC (Fig. 7).<sup>131, 172</sup>



**Figure 7.** Cyclodipeptide derivatives originating from CDPS and CDO containing pathways. CDOs catalyze the dehydrogenations depicted in red.

### 1.4.2 Cytochrome P450 enzymes

#### 1.4.2.1 Properties

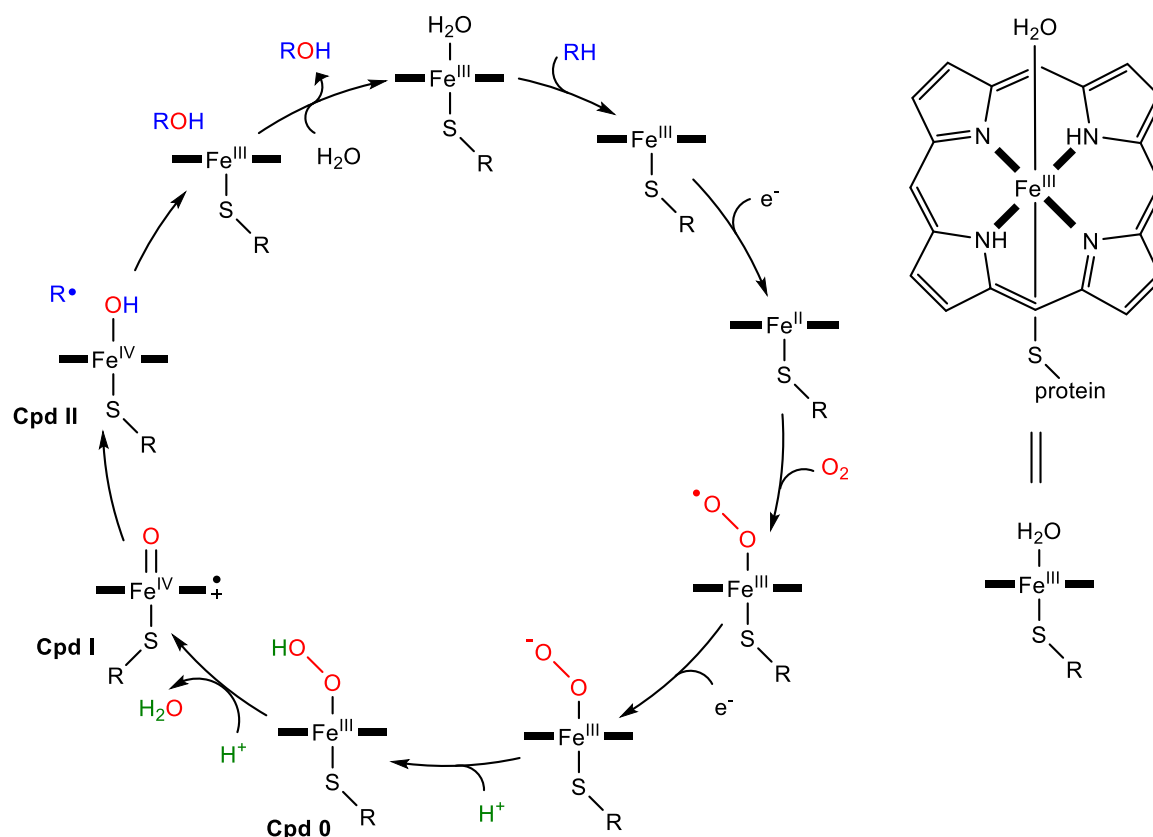
P450 oxidases constitute one of the largest and versatile enzyme superfamily, whose members occur ubiquitously in all living organisms.<sup>173</sup> They play vital roles in the metabolism, especially in biotransformation of xenobiotics, as well as the biosynthesis of NPs.<sup>174, 175</sup> The name P450 is derived from P for pigment, as the first described P450 enzyme was a pigment from rat liver microsomes, and the characteristic absorption maximum at 450 nm, when the reduced state of the protein is complexed with CO.<sup>176</sup> By reductive activation of molecular oxygen, P450 enzymes are able to functionalize inert C-H bonds, for which reason they are often referred to as mono-oxygenases. Although insertion of oxygen, like hydroxylation or epoxidation, are common reactions of P450 enzymes, this term does not reflect the extensive repertoire they are able to catalyze.<sup>174, 177</sup> In addition, it comprises aromatization, desaturation, dealkylation, nitration, rearrangements, C-C-, C-N and C-S-bond formation, as well as dimerization and nucleobase transfer.<sup>131, 177-180</sup> This variety of reactions and resulting NPs is based on versatile sizes of P450 proteins, distinct shapes of binding pockets and the usage of different co-factors.<sup>175</sup> Essential for the oxygen binding is the prosthetic group heme with its central iron atom (Fe<sup>III</sup>). Fe<sup>III</sup> is coordinated planarly to four nitrogen atoms belonging to porphyrin and bound axially to the protein via a cysteine residue and one molecule of water in the resting state.<sup>175, 181, 182</sup> Only this cysteine (C<sub>351</sub>; aligned to EryF) is an absolutely conserved residue in bacterial P450 oxidases, while a neighboring glycine and a threonine are highly conserved.<sup>177, 183, 184</sup> The lack of other conserved motifs in their sequences reflects their heterogeneity. Bacterial P450 redox systems belong to class I and are located dissolved in the cytosol. They consist of a three component system – a NAD(P)H-dependent and FAD-containing reductase (ferredoxin reductase) shuttles electrons to ferredoxin, which transfers them to the P450 protein. In contrast, eukaryotic P450 enzymes are usually membrane bound class II enzymes with a two component system.<sup>175</sup> In theory, *Streptomyces* P450 oxidases are able to accept a broad spectrum of heterologous redox partners, including eukaryotic spinach ferredoxin and ferredoxin reductase, flavodoxin and flavodoxin reductase from *E. coli*, as well as putidaredoxin and putidaredoxin reductase from *Pseudomonas putida*.<sup>177</sup>

#### 1.4.2.2 Reaction mechanism

The reaction mechanism of P450 oxidases is depicted for a hydroxylation in Fig. 8. In the resting state, one H<sub>2</sub>O is coordinated as sixth ligand to the central Fe<sup>III</sup>, which is in low spin. Binding of the substrate to the active site of the enzyme displaces the H<sub>2</sub>O and changes the Fe<sup>III</sup> to high spin. A first electron is shifted from the redox partner to Fe<sup>III</sup>, thereby reducing it to Fe<sup>II</sup>. Subsequently, O<sub>2</sub> from the environment binds to Fe<sup>II</sup>, oxidizing it again to Fe<sup>III</sup>. Next, a second electron is transferred to the distal oxygen atom and thus forming a free electron pair with a negative charge. This position is protonated resulting in the hydroperoxo-ferric intermediate compound 0 (Fe<sup>III</sup>-OOH<sup>-</sup>). Via a second protonation and release of water, the highly reactive key intermediate compound 1, an oxo-ferryl cationic porphyrin radical (Fe<sup>IV</sup>=O<sup>+</sup>), is formed. Compound 1 is able to abstract one hydrogen from less reactive C-H bonds, leading to compound 2, a hydroxo-ferryl intermediate (Fe<sup>IV</sup>-OH), as well as a radical at the substrate. The hydroxyl group is attached to the substrate and Fe<sup>IV</sup> is reduced to Fe<sup>III</sup>. In the last step, the product is released from the enzyme and one H<sub>2</sub>O is re-bound to restore the resting state.<sup>175, 177</sup> More complex

## INTRODUCTION

reactions follow the same steps, but might involve a second hydrogen abstraction from the substrate, which is carried out by compound 2, and subsequent release of H<sub>2</sub>O.<sup>185</sup> Because of the peroxide shunt pathway, some bacterial P450 oxidases are capable to use H<sub>2</sub>O<sub>2</sub> to directly generate compound 0 and substitute the two electrons from the redox partner, the atmospheric O<sub>2</sub>, and the proton.<sup>177</sup>



**Figure 8.** Mechanism of hydroxylation catalyzed by P450 enzymes. The axially coordinated porphyrin ring is abbreviated. Cpd: compound

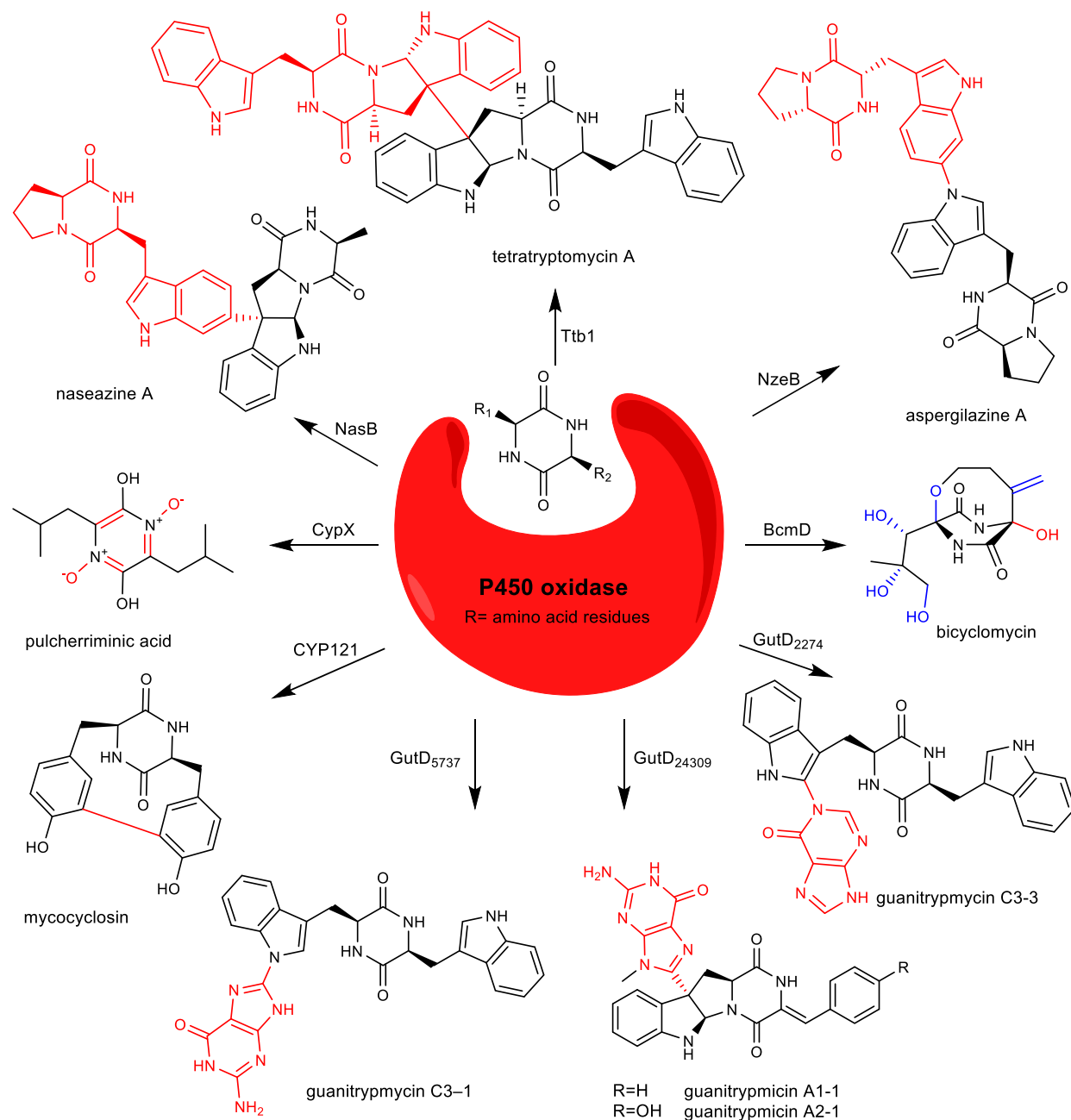
### 1.4.2.3 CDPs derivatives catalyzed by P450 enzymes

P450 enzyme-mediated reactions in CDPS containing pathways comprise hydroxylation, aromatization, coupling of intramolecular C-C bonds, dimerization, and nucleobase transfer (see Fig. 9).<sup>131, 139, 140, 144, 172, 179, 180, 185-190</sup> During the biosynthesis of the antibiotic bicyclomycin in *Streptomyces cinnamoneus*, the P450 oxidase BcmD installs the last hydroxyl group at a tertiary carbon, while four previous hydroxylations at primary and secondary carbons are catalyzed by Fe<sup>II</sup>/2-OG oxidases (see next section).<sup>139</sup> The formation of pulcherriminic acid, precursor of the red pigment pulcherrimin, involves CypX (CYP134) responsible for aromatization of the DKP in *cyclo*-L-Leu-L-Leu.<sup>179</sup> CYP121 from *Mycobacterium tuberculosis* (*M. tuberculosis*) attracted recent attention, as it seems to be essential for survival of this pathogen. Inhibition of this P450 enzyme is investigated as a new target for the treatment of tuberculosis. CYP121 forms the intramolecular C-C bond in the biogenesis of mycrocyclosin.<sup>186</sup>

## INTRODUCTION

Several P450 oxidases have been characterized as CDP-dimerases during the last years. They are able to catalyze the regio- and stereospecific connection of two identical or different CDPs – homo- or hetero-dimerases. Such P450 enzymes are members of the biosynthetic pathways of naseazine A (NasB),<sup>187</sup> aspergilazine A (NzeB),<sup>185</sup> and tetratryptomycin A (TtpB1).<sup>180</sup>

Last but not least, some intriguing P450 oxidases were elucidated in BGCs responsible for the formation of guanitryptomycins, CDP adducts with nucleobases. The nucleobases guanine and hypoxanthine are transferred by GutDs onto N-1, C-2, or C-3 of tryptophan residues.<sup>131, 189, 190</sup>

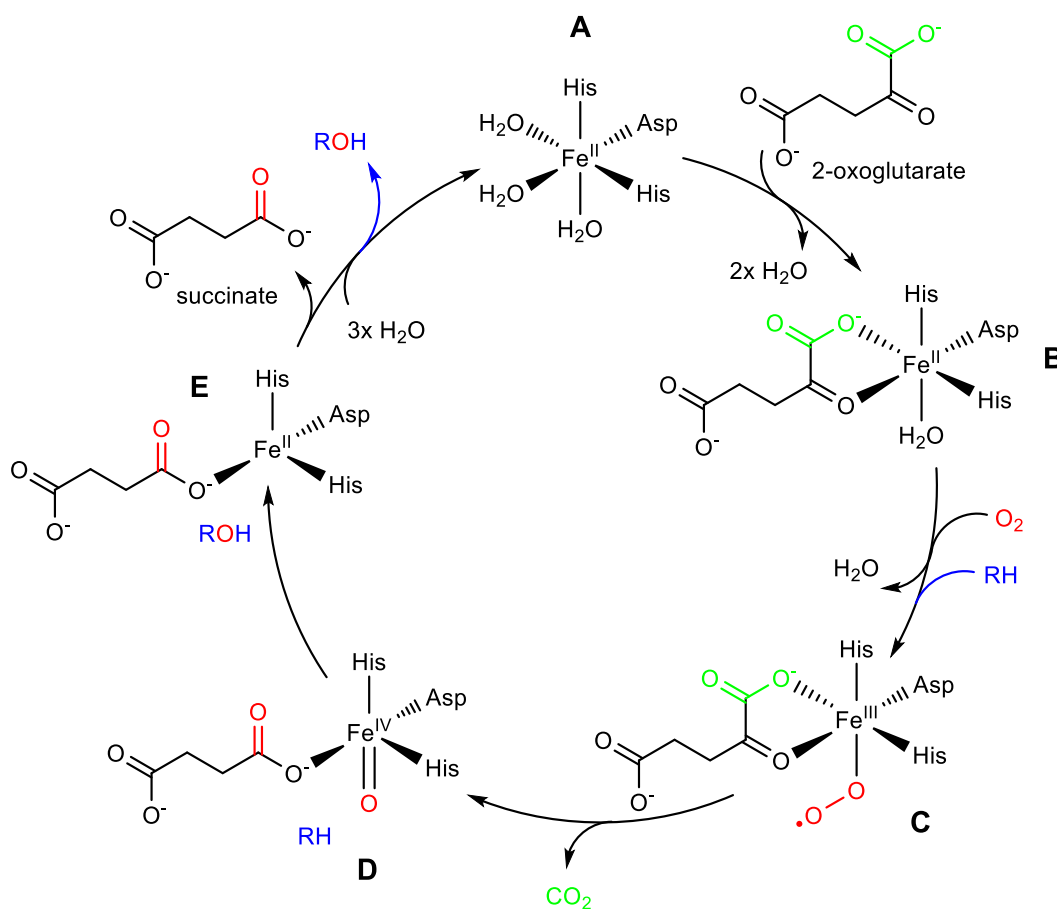


**Figure 9.** Reactions catalyzed by bacterial P450 oxidases associated to CDPs are depicted in red. The hydroxyl groups and the exomethylene group in bicyclomycin are installed by Fe<sup>II</sup>/2-OG oxidases depicted in blue.

### 1.4.3 Fe<sup>II</sup>/2-oxoglutarate dependent oxidases

Fe<sup>II</sup>/2-OG oxidases belong to a large enzyme family and are involved in primary and secondary metabolism throughout almost all forms of life.<sup>191</sup> A variety of oxidative biotransformations is accomplished by these enzymes, such as hydroxylation, epoxidation, halogenation, desaturation, and rearrangements.<sup>192-198</sup> Their protein sequences typically share a distinct catalytic triad consisting of two histidine and one aspartic acid located within a double stranded  $\beta$ -helix fold. A central iron atom (Fe<sup>II</sup>) is coordinated to each of these residues and to three H<sub>2</sub>O in the resting state (Fig. 10 **A**). When activated, two H<sub>2</sub>O are replaced by the co-substrate 2-oxoglutarate (**B**), and the third by molecular oxygen after the substrate was bound to the active site of the enzyme (**C**). 2-oxoglutarate is converted to succinate by elimination of CO<sub>2</sub> and transfer of one oxygen atom resulting in the key reactive ferryl-oxo species (Fe<sup>IV</sup>=O) (**D**), similar to P450 enzymes. This reactive intermediate is able to abstract less reactive hydrogens from substrates to form different types of products (**E**). Finally, the product and succinate are released and three H<sub>2</sub>O re-coordinate (**A**).<sup>191, 199</sup>

So far, Fe<sup>II</sup>/2-OG oxidases associated to CDPS have only been found in the biosynthesis of bicyclomycin. In this case, they are responsible for the formation of an exomethylene group and four hydroxylations, of which one induces the cyclization (see blue reactions in Fig. 9).<sup>139</sup>

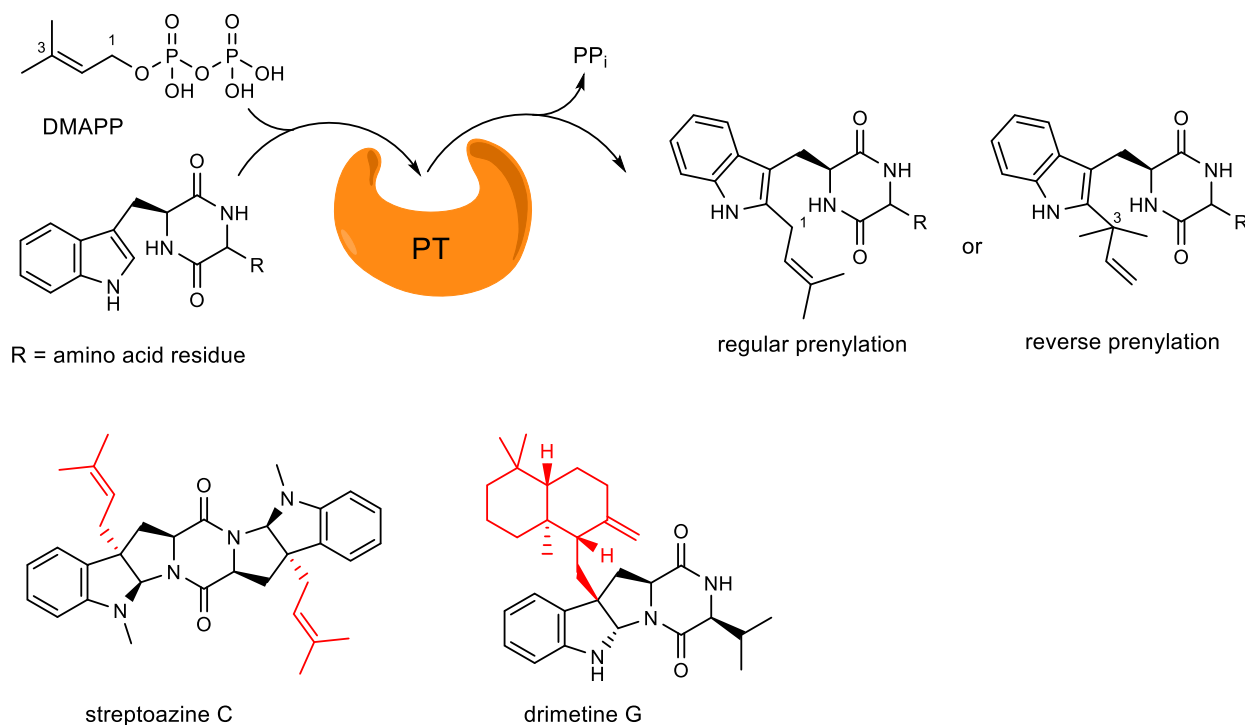


**Figure 10.** Catalytic cycle of a hydroxylation performed by a Fe<sup>II</sup>/2-oxoglutarate-dependent oxidase

### 1.4.4 Prenyltransferases

PTs catalyze the regio- and stereoselective attachment of isoprene units onto diverse substrates and are crucial in primary and secondary metabolism, e.g. post-translational modification of proteins and peptides, or prenylation steps in the biosynthesis of ubiquinones.<sup>200</sup> Dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP), derived from the acetate-mevalonate or methylerythritol phosphate pathway, are precursors for multiples of the C<sub>5</sub>-core. These multiplied units, including DMAPP (C<sub>5</sub>), IPP (C<sub>5</sub>), geranyl diphosphate (C<sub>10</sub>), farnesyl diphosphate (FPP, C<sub>15</sub>), and geranylgeranyl diphosphate (C<sub>20</sub>), serve as prenyl donors to be transferred onto various, often aromatic prenyl acceptors.<sup>200</sup> Thereby, PTs contribute not only to structural diversity of secondary metabolites, but are also essential for their biological and pharmacological effects.<sup>200, 201</sup> Prenyl moieties typically increase the lipophilic character of compounds and their binding ability to target proteins.<sup>202</sup> The attachment can be distinguished into regular prenylation via C-1 or reverse prenylation via C-3 of the prenyl donor.<sup>203, 204</sup> Afterward, the prenyl residues can be further modified by hydroxylation, epoxidation and cyclization.<sup>205-208</sup>

Members of the ABBA family from bacteria and fungi are soluble aromatic PTs, that possess a distinct three-dimensional structure, referred to as  $\alpha$ - $\beta$ - $\beta$ - $\alpha$ -fold.<sup>209</sup> Most of their reactions are metal-independent. However, addition of Ca<sup>2+</sup> or Mg<sup>2+</sup> might increase their catalytic efficiency.<sup>210-212</sup> Only two PTs (SazB and DmtC1) from CDPS containing BGCs have been characterized so far. Both transfer a prenyl donor, *i.e.* DMAPP and FPP, in a regular manner onto a tryptophan residue of a CDP, in the biosynthesis of streptoazine C and drimetine G, respectively (Fig. 11).<sup>207, 213</sup>



**Figure 11.** Regular and reverse prenylation catalyzed by a PT (above). Prenylated compounds from CDPS-associated pathways (below).

## 2. Aims of this thesis

The overall aim of this thesis was the investigation of novel CDPS-associated biosynthetic pathways from the genus *Streptomyces*, which were identified by genome mining.

The following topics were investigated in the course of this work:

### **Discovery of the cWM derivative guatrypmethine C and the Fe<sup>II</sup>/2-OG oxidase GtmE as a new enzyme type for double bond formation in CDPS-associated pathways**

Tailoring enzymes are able to catalyze intriguing reactions and thereby increase the diversity of secondary metabolites by a concerted interaction of combined modifications. A common reaction is the formation of double bonds by dehydrogenation at the DKP ring of CDPs. This reaction is typically carried out by CDOs in CDPS-dependent pathways. Prior to this work, Fe<sup>II</sup>/2-OG oxidases associated to CDPSs were solely discovered in the biosynthesis of bicyclomycin, where they catalyze four hydroxylation steps and the formation of an *exo*-methylene group. By genome mining, we identified a five-gene cluster from *Streptomyces cinnamoneus* coding for four putative enzymes – a CDPS for backbone formation, a CDO, a P450 oxidase, and a Fe<sup>II</sup>/2-OG oxidase, termed GtmA, GtmBC, GtmD, and GtmE, respectively.

The following experiments were carried out to elucidate this cryptic BGC:

- Phylogenetic analysis of the *gtm* cluster and its encoded enzymes
- Cloning of different gene combinations into the pPWW50A vector
- Heterologous expression of the generated plasmids in the host *Streptomyces albus* J1047
- LC-MS analysis of the cultural extracts to detect new metabolites
- Large scale fermentation and isolation for structure elucidation by a set of NMR measurements, including <sup>15</sup>N NMR analysis of isotopically labelled compounds
- Deuterium exchange experiments monitored by LC-MS to prove an epimerization via keto–enol tautomerism
- In vitro biochemical investigations of the dehydrogenation catalyzed by GtmE
- Quantum chemical calculations to determine energy levels of the intermediates guatrypmethine A and B (in cooperation with Oliver Kreuz and Prof. Dr. Robert Berger from the Department of Chemistry, Philipps-Universität Marburg)
- Postulation of a reaction mechanism for the dehydrogenation by GtmE

### **Identification of widely distributed, bifunctional P450 oxidases from actinobacteria for intramolecular C-C bond formation of cYY and its coupling with nucleobases**

P450 oxidases in CDPs-dependent pathways are known to catalyze hydroxylation, aromatization, formation of an intramolecular C-C bond, dimerization, and transfer of the nucleobases guanine and hypoxanthine onto tryptophan-containing CDPs. By genome mining and phylogenetic analysis, my colleague Dr. Jing Liu and I found a widely distributed, similar two-gene locus in actinobacteria. These clusters comprise genes for a CDPS as backbone enzyme and a P450 oxidase for modification. We termed them GymAx and GymBx, respectively. Interestingly, the GymBs are phylogenetically clustered

in a clade next to CYP121 from *M. tuberculosis*, which is responsible for the formation of the intramolecular C-C bond of mycocyclosin. Six representative *Streptomyces* species were chosen for detailed investigation.

This project was done in cooperation with Dr. Jing Liu. Both of us simultaneously worked on different representatives until we discovered that all six clusters produce identical secondary metabolites. I finalized this project after Dr. Jing Liu left our lab in April 2021. The following experiments were performed:

- Phylogenetic analysis of selected putative and known CDPSs and P450 oxidases
- Generation of plasmids harboring *gymAs* for expression of the CDPSs in *E. coli* and those with *gymAs* and *gymBs* for heterologous expression in *S. albus* J1074
- Analysis of the cultural extracts of the obtained transformants using LC-MS
- Large scale fermentation for isolation and structural elucidation by a combined set of NMR spectra
- Expression of the mycocyclosin BGC from *M. tuberculosis* to compare its metabolite profile
- Overexpression of the P450 oxidases GymB<sub>1</sub>-GymB<sub>6</sub> in *E. coli*, followed by enzymatic assays with the recombinant proteins for biochemical investigation in vitro (together with Yiling Yang)
- Description of the biosynthetic pathway and proposal of a mechanism for the reactions catalyzed by GymBx

### **Investigation of the streptoazine biosynthetic pathway from *Streptomyces aurantiacus* reveals the presence of the promiscuous prenyltransferase SasB**

PTs rarely occur in CDPS-associated biosynthetic pathways. Prior to this work only two such PTs have been investigated, *i.e.* DmtC1 and SazB. For SazB, a strict substrate specificity toward *cyclo*-L-Trp-L-Trp (cWW) and DMAPP was reported by Zhang *et al.* (2021).<sup>213</sup> We identified a BGC from *Streptomyces aurantiacus*, comprising the three genes *sasA*–*C* coding for a CDPS, a PT, and a MT, respectively.

Elucidation of the *sas* cluster was mainly done by Dr. Jing Liu and Yiling Yang. I took part in the structure elucidation of the generated natural products and the investigation of the substrate flexibility of the PT SasB toward different tryptophan-containing CDPs. The project was divided into the following experiments:

- Phylogenetic comparison of the *sas* cluster with the previously reported *saz* cluster
- Construction of plasmids with different gene combinations of *sasA*–*C* into pPWW50A vector for heterologous expression in *S. albus* J1074 to investigate the enzyme functions
- Cultivation of the generated transformants and investigation of their extracts using LC-MS
- Large scale fermentation and structural elucidation of the purified compounds by NMR analysis
- Feeding of pathway intermediates to prove each step of the biosynthesis
- Cultivation of the *sasB* transformant with different tryptophan-containing CDPs, including dehydrogenated forms, to investigate the substrate specificity of SasB
- Deuterium exchange experiments monitored by LC-MS to prove an observed epimerization



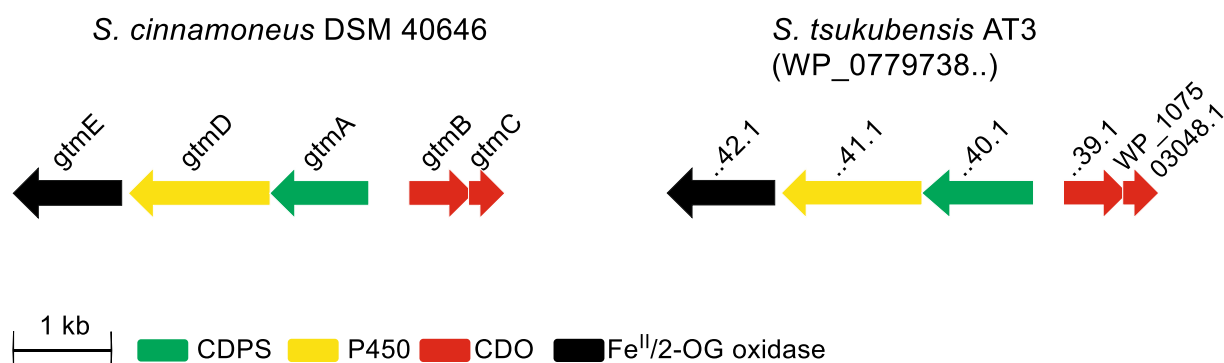
### 3. Results and discussion

#### 3.1 Discovery of the cWM derivative guatrypmethine C and the Fe<sup>II</sup>/2-OG oxidase GtmE as a new enzyme type for double bond formation in CDPS-associated pathways

So far, double bond formation at the DKP ring of CDPs was reported to be catalyzed by P450 enzymes in NRPS-dependent pathways, mostly in fungi, or by CDOs in bacterial CDPS-associated pathways. In this part of the thesis, GtmE, a member of the widely distributed Fe<sup>II</sup>/2-OG oxidase family, was identified as a novel enzyme type to catalyze the formation of a double bond at the DKP core of *cyclo*-L-Trp-L-Met (cWM) in the biosynthesis of the novel secondary metabolite guatrypmethine C.

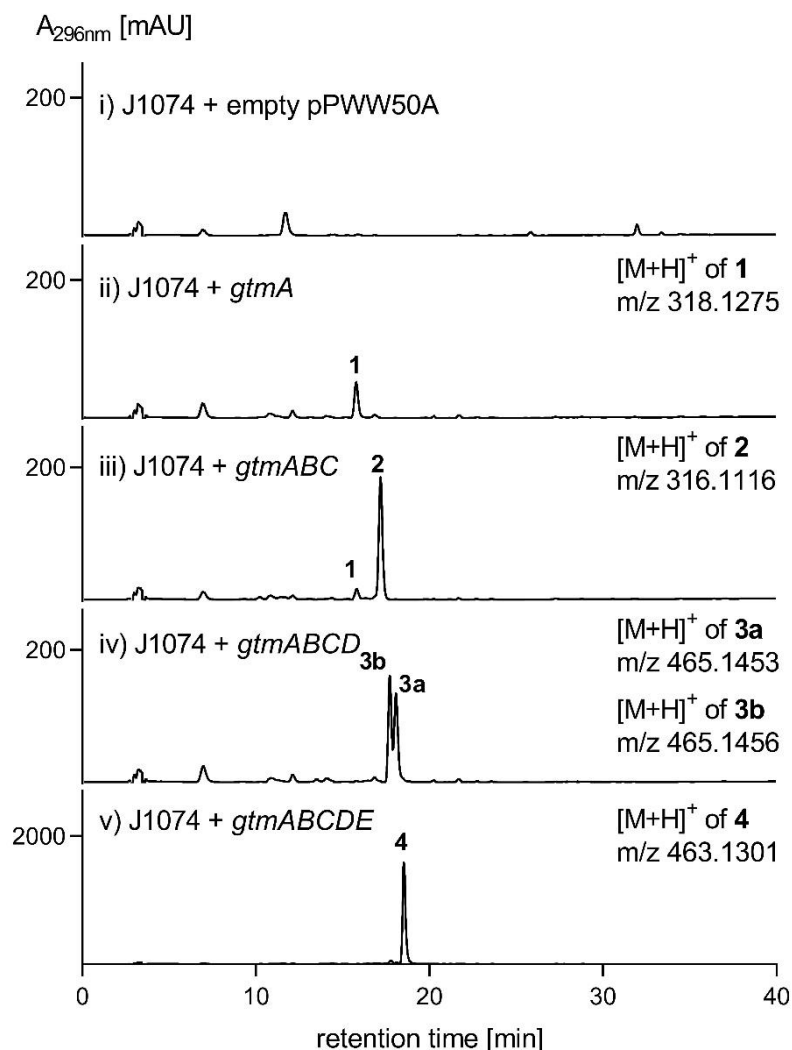
A five gene locus, termed *gtm* cluster, was identified in *Streptomyces cinnamoneus* DSM 40646 (*S. cinnamoneus*), using BLAST search with characterized P450 enzymes from CDPS-dependent biosynthetic pathways. A second homologous cluster was also found in *Streptomyces tsukubensis* AT3. The BGC codes for four enzymes, *i.e.* the CDPS GtmA, the CDO GtmBC, the P450 oxidase GtmD, and the Fe<sup>II</sup>/2-OG oxidase GtmE. The latter was annotated in the database as a hypothetical protein, but detailed sequence comparison showed an approximate identity of 60 % on the amino acid level to isopenicillin-N-synthases, which belong to the superfamily of Fe<sup>II</sup>/2-OG oxidases.<sup>214</sup> Thus, the *gtm* cluster differs from the previously reported *gut* and *pcm* cluster by replacement of the methyltransferase genes *gutE* and *pcmE* with the Fe<sup>II</sup>/2-OG oxidase gene *gtmE*, respectively.<sup>131, 172</sup>

The sequences of *gtmA*–*gtmE* were amplified by performing PCR of the genomic DNA of *S. cinnamoneus* and cloned in different combinations into the replicative vector pPWW50A under control of the constitutive promoter *ermEp*\*.<sup>215</sup> These plasmids were transferred into the model expression host *S. albus* J1074 via conjugation.<sup>127</sup> Positive transformants were cultivated in modified R5 medium and their extracts analyzed using LC-MS (Fig. 12).



**Figure 12.** Genetic organization of the *gtm* cluster from *Streptomyces cinnamoneus* and its homologous cluster from *Streptomyces tsukubensis*

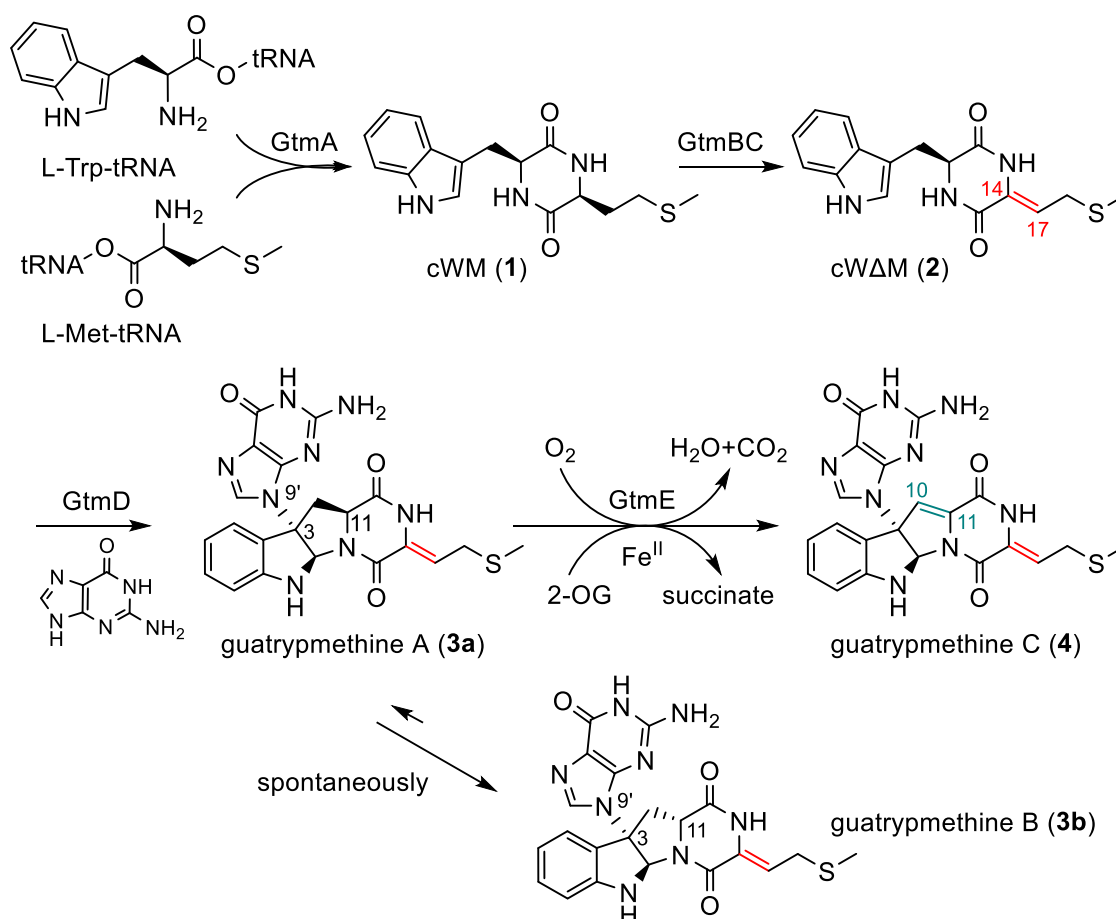
## RESULTS AND DISCUSSION



**Figure 13.** LC-MS chromatograms of extracts from *S. albus* transformants harboring *gtm* genes or the empty vector as negative control

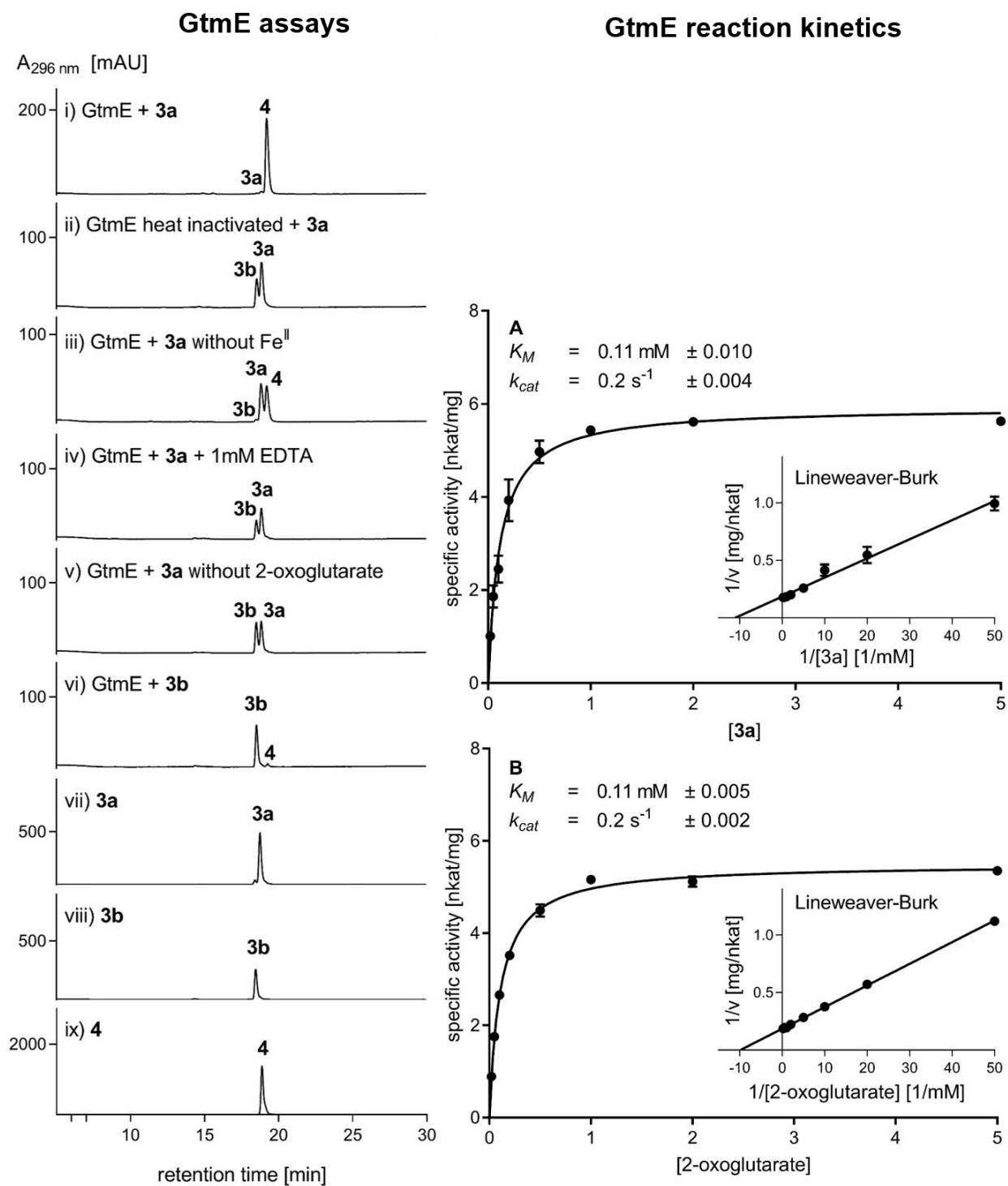
In contrast to the negative control (Fig. 13i), the transformant harboring the gene for GtmA showed a predominant peak **1** with a  $[M+H]^+$  ion corresponding to cWM (Fig. 13ii). Interpretation of the  $^1\text{H}$ -NMR spectrum of the isolated compound confirmed its structure. Combination of the genes for the tailoring enzymes together with GtmA and detailed NMR analyses ( $^1\text{H}$ ,  $^{13}\text{C}$ , COSY, HSQC, and HMBC) of the purified metabolites proved that the CDO GtmBC first installs a double bond at the methionine residue of cWM (**2**, Fig. 13iii). The P450 oxidase GtmD then transfers a guanine to C-3 of the tryptophan moiety (**3a**, Fig 13iv), leading to cyclization between C-2 and N-12 of the CDP. Finally, the  $\text{Fe}^{\text{II}}/2\text{-OG}$  oxidase GtmE forms a second double bond at the tryptophan side of the DKP ring (**4**, Fig. 13v). The connection of the guanine via its N-9' to C-3 of cWM was proven by interpretation of the  $^{15}\text{N}$  NMR spectra ( $^{15}\text{N}$  inverse gated,  $^1\text{H}$ - $^{15}\text{N}$  HSQC, and  $^1\text{H}$ - $^{15}\text{N}$  HMBC) with  $^{15}\text{N}$ -labelled **3b** and **4**. In pH neutral to alkaline aqueous environments, **3a** is converted to its stable isomer **3b** via keto-enol-tautomerism. This was demonstrated by incubation of **3a** and **3b** in a mixture of  $\text{CD}_3\text{OD}$  and  $\text{D}_2\text{O}$  (1:1) and subsequent LC-MS analysis. Due to the involvement of L-tryptophan, L-methionine, and guanine, **3a**, **3b**, and **4** were termed guatrypmethines A, B, and C, respectively.

## RESULTS AND DISCUSSION



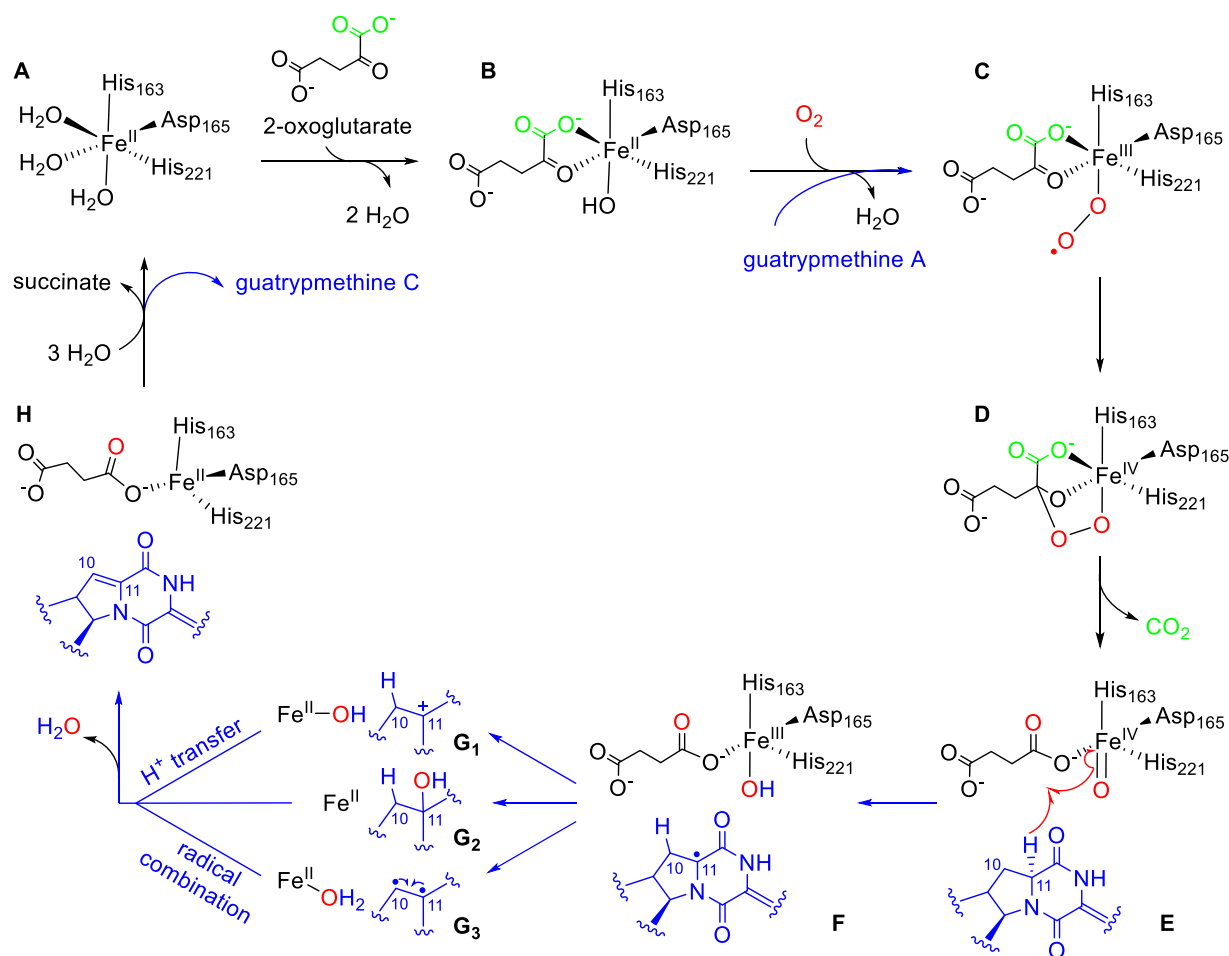
**Figure 14.** Biosynthetic pathway of guatrypmethine C (4) in *Streptomyces cinnamoneus*

Prior to this study, the formation of a double bond at the DKP ring was only known to be carried out by CDOs in CDPS-dependent pathways and has not been reported for Fe<sup>II</sup>/2-OG oxidases.<sup>161</sup> Thus, we decided to investigate the GtmE-mediated dehydrogenation biochemically. Its coding sequence was cloned into pET28a for overproduction in *E. coli* and the recombinant His<sub>6</sub>-tagged GtmE was purified on Ni-NTA agarose to near homogeneity. Enzyme assays of GtmE (1.5 μM) with 0.5 mM **3a** in the presence of (NH<sub>4</sub>)<sub>2</sub>Fe<sup>II</sup>(SO<sub>4</sub>)<sub>2</sub>, 2-oxoglutarate, and ascorbate showed that **3a** was almost completely converted to **4** (Fig. 15i). No conversion was observed in assays with heat-inactivated GtmE, additional EDTA to complex the central iron of GtmE, or without 2-oxoglutarate (Fig. 15ii, iv, v). These experiments demonstrated unequivocally that GtmE belongs to the family of Fe<sup>II</sup>/2-OG oxidases and that it is indeed responsible for the formation of the second double bond exo of the DKP ring. Furthermore, the GtmE-catalyzed dehydrogenation follows Michaelis-Menten kinetics and the kinetic parameters  $K_M$  and  $k_{cat}$  were calculated for both **3a** and 2-oxoglutarate as 0.11 mM and 0.2 s<sup>-1</sup>, respectively (Fig. 15 right). Only low product formation of 2 % was observed after incubation of the stable isomer **3b** instead of **3a** (Fig. 15vi). Quantum chemical calculations on the DFT level supported the experimental observation that **3b** is energetically more stable than **3a**. In this context, the low production yield of **4** for the incubation with **3b** is rationalized by a higher activation energy for the reaction.



**Figure 15.** Left: LC-MS chromatograms of GtmE assays; right: determination of the kinetic parameters of the GtmE reaction for **3a** (A) and 2-oxoglutarate (B)

## RESULTS AND DISCUSSION



**Figure 16.** Proposed mechanism of GtmE-mediated double bond installation in guatrypmethine A

The typical catalytic triad of Fe<sup>II</sup>/2-OG oxidases was identified as His<sub>163</sub>, Asp<sub>165</sub>, and His<sub>221</sub> by alignments with characterized protein sequences. A proposed reaction mechanism is depicted in Fig. 16. At the key step (E), the first hydrogen abstraction most likely takes place at C-11, as the radical intermediate can be mesomerically stabilized (F). Subsequently, three possible intermediates for the generation of the desaturated product 4 can be taken into account. Either via a cationic intermediate (G<sub>1</sub>), via hydroxyl rebound (G<sub>2</sub>), or through a biradical species (G<sub>3</sub>).<sup>197, 199, 216</sup>

In conclusion, we elucidated in this study the biosynthetic pathway of the three newly discovered guatrypmethines A, B, and C in *S. cinnamoneus* applying a combined approach of genome mining, heterologous expression, and biochemical investigation. Regarding GtmE, a new function of the well-known Fe<sup>II</sup>/2-OG oxidases as desaturase in CDPS-associated pathways was identified and its kinetics characterized.

For details on this work, please refer to publication 4.1.

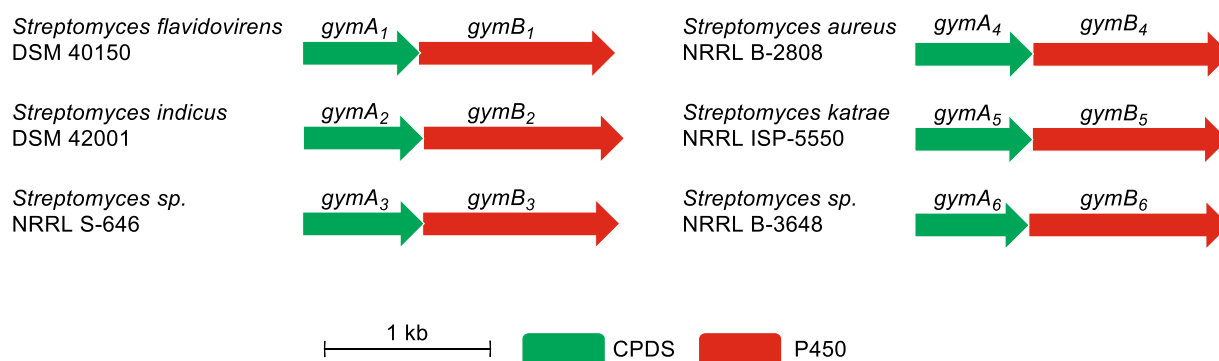
## RESULTS AND DISCUSSION

### 3.2 Identification of widely distributed, bifunctional P450 oxidases from actinobacteria for intramolecular C-C bond formation of cYY and its coupling with nucleobases

P450 enzymes are known as versatile catalysts in CDPS-dependent pathways.<sup>161</sup> A fairly new known reaction of P450 oxidases is the transfer of nucleobases, e.g. guanine and hypoxanthine. CYP121 from *M. tuberculosis*, which is responsible for the intramolecular C-C bond within mycrocyclosin (see 1.4.2.3), attracted special attention, as it is suggested to be essential for the viability of *M. tuberculosis* and is thus further investigated as a potential target for the treatment of tuberculosis.<sup>186</sup>

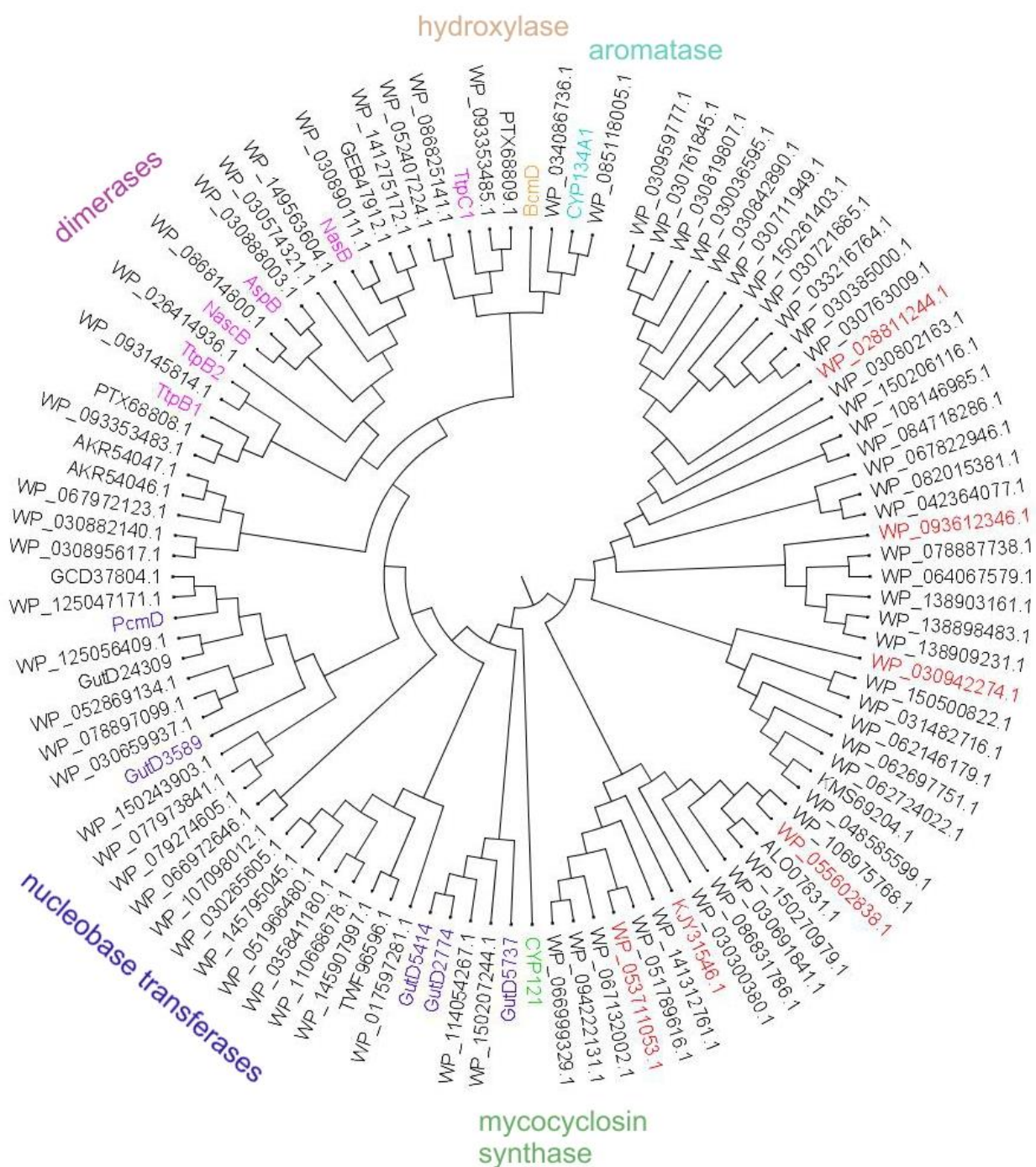
In analogy to the first project, my colleague Dr. Jing Liu and I conducted a systematic search in the NCBI database for unknown P450 homologues associated to CDPS by genome mining with characterized proteins. Phylogenetic analysis of known and putative P450 enzymes enclosed to CDPSs, revealed that the reported P450 oxidases for dimerization and nucleobase transfer are clustered in two distinct clades (Fig 18).<sup>131, 139, 179, 180, 186-190</sup> The largest clade of the constructed phylogenetic tree contains 47 unknown P450 homologues, which are closely related to CYP121. A representative of each subclade was chosen for detailed investigation in this study; in total six different *cdps* *p450* containing BGCs from *Streptomyces flavidovirens* DSM 40150 (*gymA<sub>1</sub>B<sub>1</sub>*), *Streptomyces indicus* DSM 42001 (*gymA<sub>2</sub>B<sub>2</sub>*), *Streptomyces* sp. NRRL S-646 (*gymA<sub>3</sub>B<sub>3</sub>*), *Streptomyces aureus* NRRL B-2808 (*gymA<sub>4</sub>B<sub>4</sub>*), *Streptomyces katrae* NRRL ISP-5550 (*gymA<sub>5</sub>B<sub>5</sub>*), and *Streptomyces* sp. NRRL B-3648 (*gymA<sub>6</sub>B<sub>6</sub>*) (Fig. 17).

The functions of the CDPSs GymA<sub>1</sub>–GymA<sub>6</sub> were elucidated by heterologous expression in *E. coli* BL21 DE3 with constructs in the pET28a vector. All CDPSs produced *cyclo*-L-Tyr-L-Tyr (cYY, **5**) as main and *cyclo*-L-Tyr-L-Phe (cYF, **6**) as minor product. The production of two CDPs differing in the second amino acid residue is in line with the substrate promiscuity of other described CDPSs.<sup>131, 217</sup> Afterward, *gymA<sub>6</sub>* and the complete cluster *gymA<sub>6</sub>B<sub>6</sub>* were chosen exemplarily and cloned into the replicative pPWW50A vector for heterologous expression in *S. albus* J1074.<sup>127</sup> LC-MS analysis of the extract of the *gymA<sub>6</sub>* transformant confirmed the results of the expression in *E. coli* (Fig. 19ii). The extract of the transformant harboring the complete BGC showed the presence of three additional peaks **7**, **8**, and **9** (Fig. 19viii).



**Figure 17.** Genetic organization of the *gym* clusters from six different *Streptomyces* species

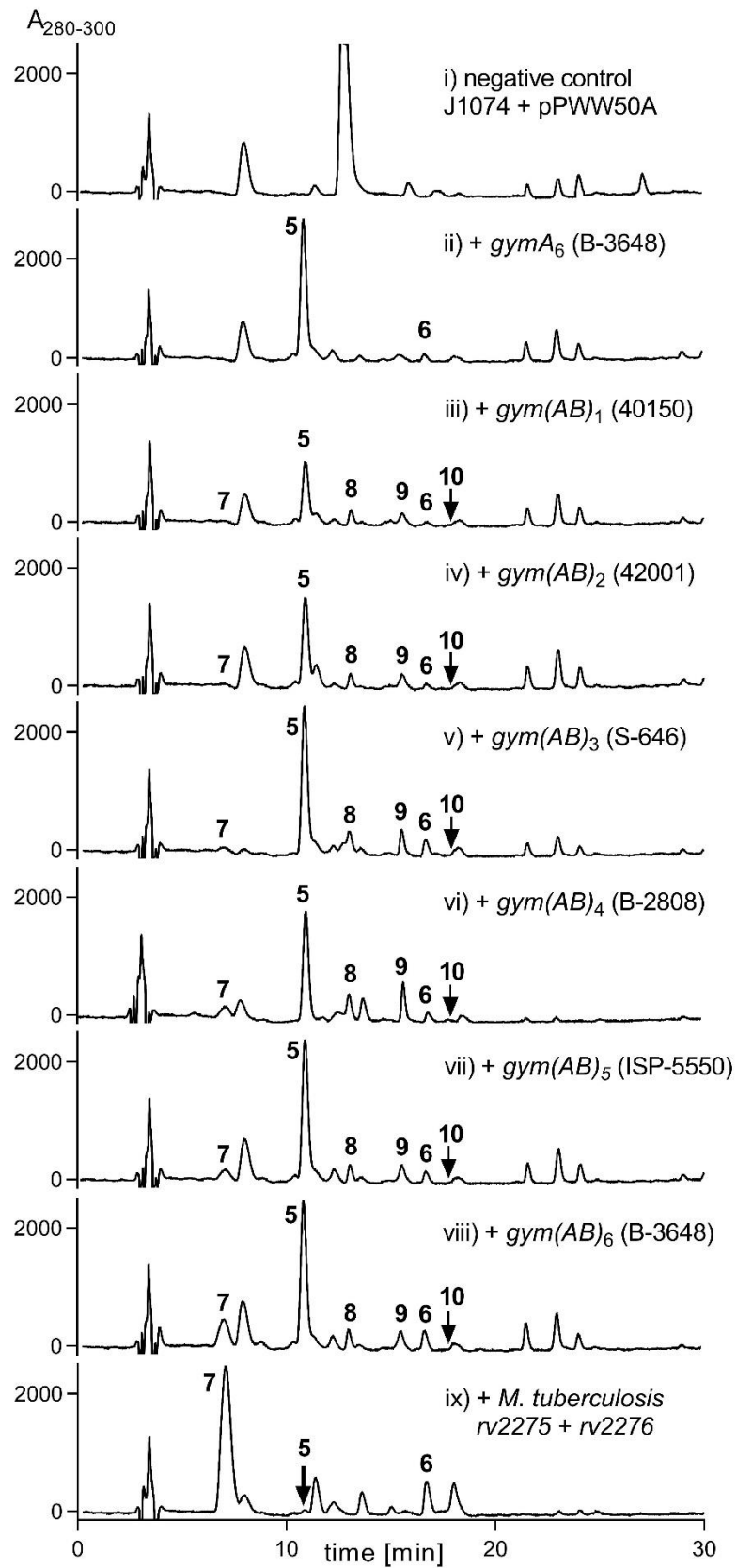
## RESULTS AND DISCUSSION



**Figure 18.** Phylogenetic tree of P450 oxidases associated to CDPSPs with accession numbers from the NCBI database (<https://www.ncbi.nlm.nih.gov>). Enzymes investigated in this study are depicted in red, CYP121 in green, other known enzymes in different colors.



## RESULTS AND DISCUSSION



**Figure 19.** LC-MS chromatograms of the extracts of *S. albus* transformants harboring the empty vector (i), *gym* genes (ii – viii), and *rv2275 + rv2276* (ix) coding for the mycocyclusin biosynthesis in *M. tuberculosis*

## RESULTS AND DISCUSSION

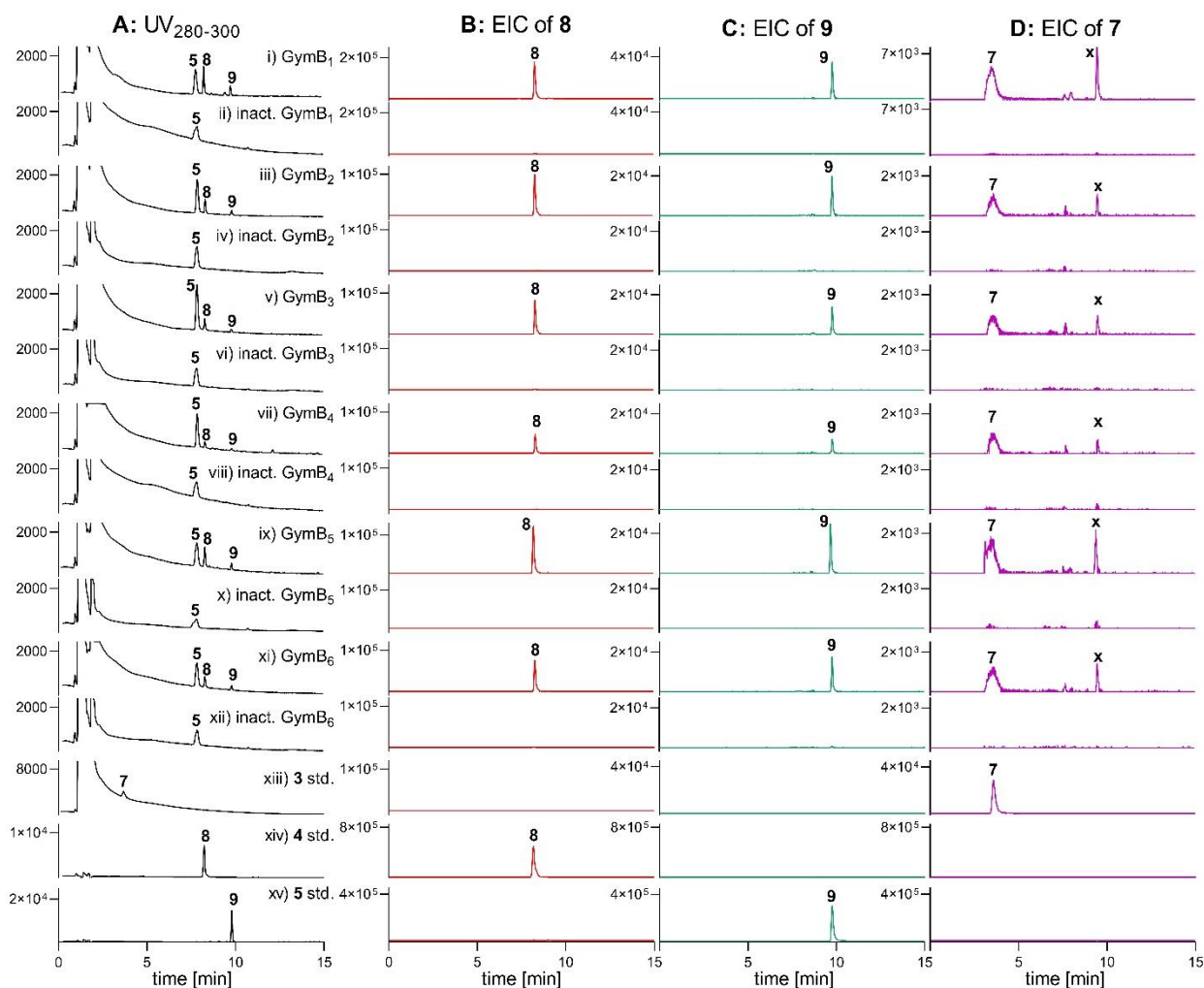
The mass of **7** is 2 Da smaller than that of cYY (**5**), corresponding well to that of mycocyclusin, which was confirmed by comparing its  $^1\text{H-NMR}$  spectrum with the previously published one.<sup>186</sup> The  $[\text{M}+\text{H}]^+$  ion of **8** is 149 Da larger compared to that of **5**, suggesting the presence of a guanine residue. The  $[\text{M}+\text{H}]^+$  ion of **9** is 134 Da larger than that of **5**, indicating the attachment of a hypoxanthine. Both compounds were isolated and their structures were elucidated with detailed NMR analysis. The NMR data confirmed the assumption that **8** is a cYY derivative with a guanine residue attached to a hydroxyl group via C-8' and **9** a descendent with a hypoxanthine connected via C-2' to C-10 of cYY. Cultivation of the *S. albus* transformant harboring *gymA<sub>6</sub>B<sub>6</sub>* in medium with  $^{15}\text{NH}_4\text{Cl}$  and subsequent LC-MS analysis of the cultural extract demonstrated that at least two  $^{15}\text{N}$  atoms are incorporated into **7**, seven in **8**, and six in **9**, which further verified their structures. As L-tyrosine and guanine/hypoxanthine are part of the structures of **8** and **9**, they were termed guatyromycine A and B, respectively.

The other five candidate BGCs *gymA<sub>1</sub>B<sub>1</sub>*–*gymA<sub>5</sub>B<sub>5</sub>* were expressed in an identical manner. The extracts of the corresponding transformants were analyzed by LC-MS, which revealed similar product profiles as that of *gymA<sub>6</sub>B<sub>6</sub>* (Fig. 19iii – vii). All five compounds, the precursors **5** and **6**, as well as the products **7**, **8**, and **9**, were observed using UV detection and verified by comparison of their retention times, UV-spectra, and MS data including MS<sup>2</sup> fragmentation pattern. Since mycocyclusin (**7**) is also the product of CYP121 from *M. tuberculosis*, the respective BGC containing the genes *rv2275* and *rv2276* was cloned and expressed as well. This cluster predominantly produced **7** with a product yield of 53.9 mg/L, together with **5** and **6** confirming that CYP121 merely forms the intramolecular C-C bond within **7** (Fig. 19ix).<sup>186</sup> Quantification of the product formation in the transformants, revealed that those located phylogenetically closer to CYP121 (*gymA<sub>4</sub>B<sub>4</sub>* – *gymA<sub>6</sub>B<sub>6</sub>*) produced **3** with 9.4 – 24.0 mg/L significantly higher than those in larger distance with 1.6 – 2.3 mg/L (*gymA<sub>1</sub>B<sub>1</sub>* – *gymA<sub>2</sub>B<sub>2</sub>*). The amounts of **8** and **9** were comparable in all respective transformants. A clear peak **10** with the mass of a coupling product of cYF (**6**) with guanine was detected using extracted ion chromatography in the extracts of all transformants carrying *gymABs*. However, due to very low quantities, its structure could not be determined in this study.

The functions of GymBs were further investigated in vitro after overproduction in *E. coli* BL21 DE3 and purification on Ni-NTA agarose. Each protein was incubated with cYY (**5**), guanine, and hypoxanthine in the presence of ferredoxin, ferredoxin reductase, and NADPH. Heat-inactivated enzymes served as negative controls. The corresponding LC-MS chromatograms showed that **8** was produced by all active proteins as the major product, followed by **9** and **7** (Fig. 20). These assays unambiguously proved that GymBs act as bifunctional enzymes catalyzing both the nucleobase transfer and the intramolecular C-C bond coupling. The conversion of **5** to **7** was much higher in the absence of guanine and hypoxanthine.

Based on the results mentioned above, it can be concluded that GymAs use L-Tyr-tRNA and L-Phe-tRNA to assemble cYY (**5**) as main and cYF (**6**) as side product. In the second step, the GymBs either introduce an intramolecular C-C bond within **5**, resulting in mycocyclusin (**7**), or transfer the nucleobases guanine or hypoxanthine onto **5**, generating guatyromycine A (**8**) and B (**9**), respectively.

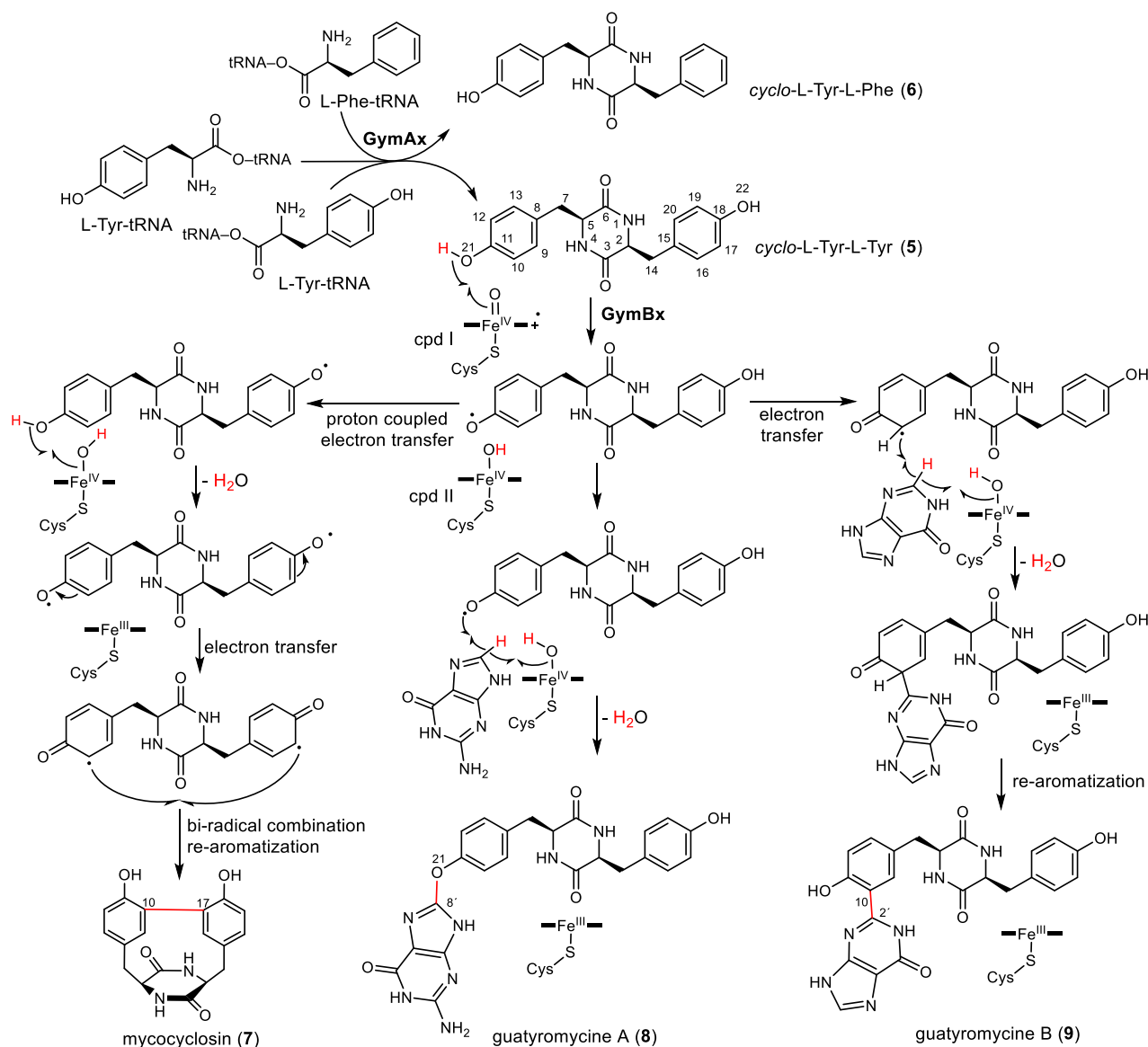
## RESULTS AND DISCUSSION



**Figure 20.** LC-MS chromatograms of enzyme assays with cYY and GymB<sub>1</sub>–GymB<sub>6</sub> in the presence of guanine and hypoxanthine. EICs of **7**, **8**, and **9** refer to  $[M + H]^+$  ions at  $m/z$  325.118, 476.184, and 461.172, respectively, with a tolerance range of  $\pm 0.005$ . Peak x: unknown mycocyclusin isomer. inact.: heat-inactivated; std.: standard

Regarding the mechanisms of the GymB catalyzed reactions (Fig 21), it can be postulated that a hydrogen at O-21 is firstly abstracted by compound I ( $\text{Fe}^{\text{IV}}=\text{O}^+$ ), in analogy to the mechanism for the biosynthesis of **7**.<sup>218</sup> The resulting radical at O-21 is the central intermediate for the formation of all three products. This radical can be transferred to different positions (C-10 and C-17) and a second hydrogen is cleaved by compound II ( $\text{Fe}^{\text{IV}}-\text{OH}$ ) either again at cYY (**5**) or at the respective nucleobase. In all three reactions one molecule  $\text{H}_2\text{O}$  is released, the respective C-O or C–C bond is formed, and the constructed compound is released from the active side of the P450 oxidase. It seems that C-10 of **5** preferably reacts with C2' of hypoxanthine in analogy to the position for the formation of mycocyclusin (**7**). However, the connection of guanine to **5** is likely caused by the amino substitution at C2' of the purine and a resulting steric hindrance in the binding pocket of the P450 oxidases.

## RESULTS AND DISCUSSION



**Figure 21.** Biosynthetic pathway of **7**, **8**, and **9**, and postulated mechanisms for the reactions catalyzed by GymBx

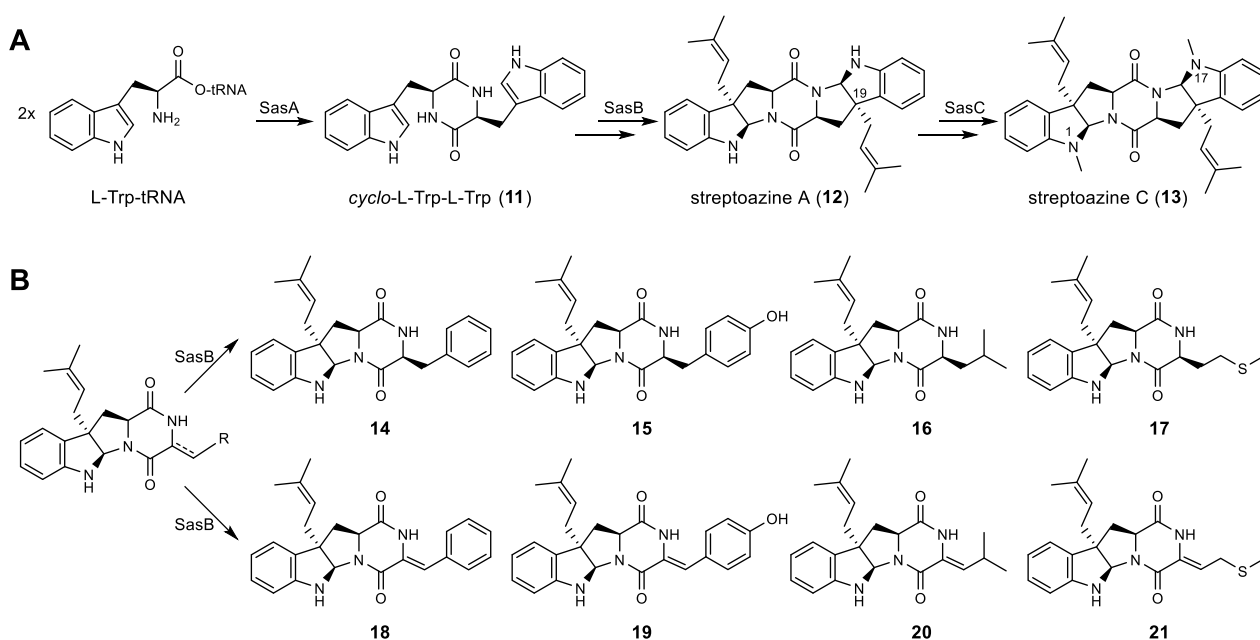
In summary, this work revealed the presence of similar, widely distributed BGCs among actinobacteria, each coding for a CDPS (GymAx) and a P450 oxidase (GymBx). They are responsible for the formation of mycocyclusin (**7**) by intramolecular C-C coupling of cYY (**5**), as well as the biosynthesis of guatyromycine A (**8**) and B (**9**) by nucleobase transfer. Their biosynthetic pathways were elucidated by combination of heterologous expression in *E. coli* and *S. albus* together with biochemical investigation using recombinant P450 enzymes. GymBs are the first described P450 oxidases in CDPS-dependent BGCs that function as unique bifunctional enzymes for both intramolecular coupling and nucleobase transfer onto tyrosyl residues.

For details on this work, please refer to publication 4.2.

### 3.3 Investigation of the streptoazine biosynthetic pathway from *Streptomyces aurantiacus* reveals the presence of the promiscuous prenyltransferase SasB

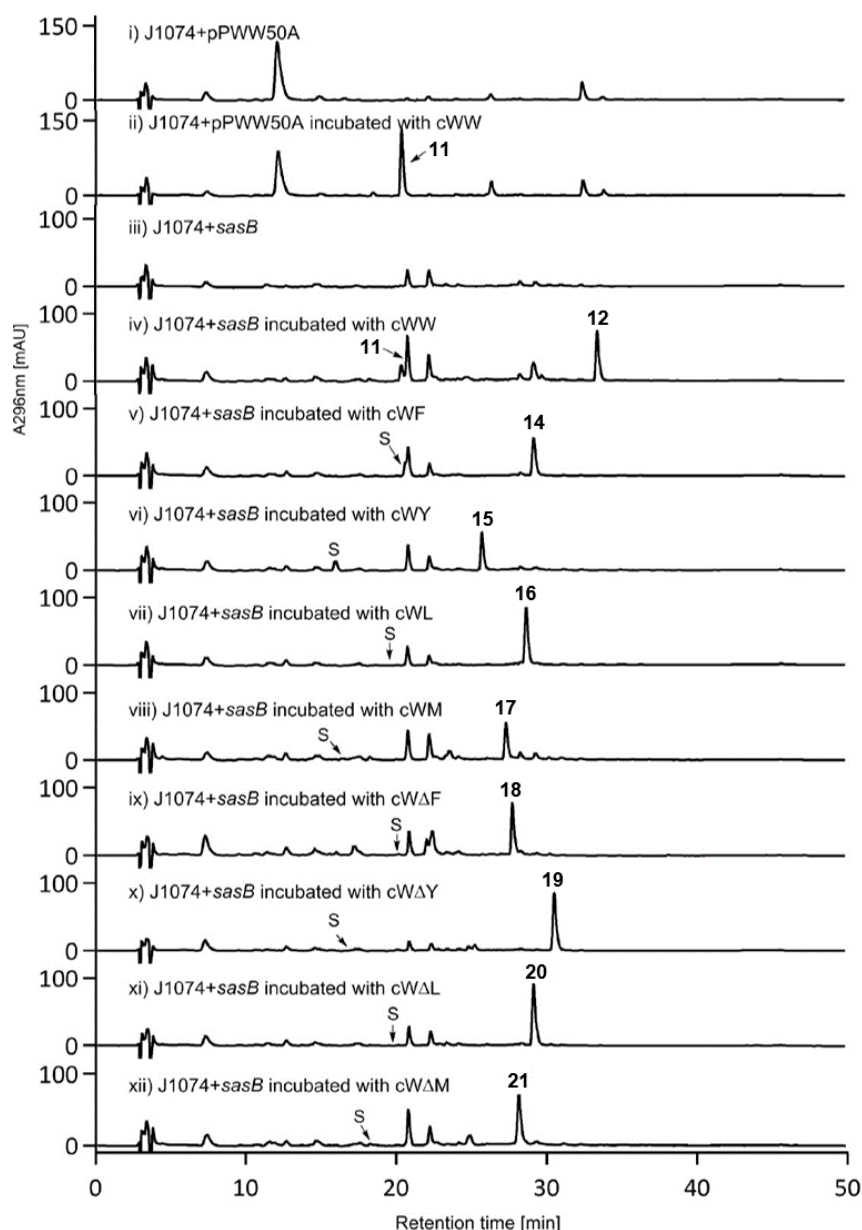
Most of the characterized prenylated CDP derivatives origin from NRPS-dependent pathways, mainly in fungi and seldom in bacteria.<sup>200, 219</sup> The corresponding PTs are usually highly permissive and are able to use structurally diverse substrates.<sup>220</sup> PTs as tailoring enzymes in CDPS-dependent BGCs are rare, as only DmtC from the drimentine G pathway and SazB, involved in the biosynthesis of streptoazine C (**13**), have been identified so far (see section 1.4.4).<sup>207, 213</sup> The latter is a bifunctional enzyme from *Streptomyces leeuwenhoekii* (*S. leeuwenhoekii*) containing both PT and MT domains for prenylation and methylation of the precursor *cyclo*-L-Trp-L-Trp (cWW, **11**). It was reported that SazB has a strict substrate specificity and only accepts DMAPP and cWW as substrates.<sup>213</sup>

Using genome mining, my colleagues identified a three-gene cluster from *Streptomyces aurantiacus* NRRL ISP-5412 (*S. aurantiacus*) coding for the CDPS SasA, the PT SasB, and the MT SasC. SasB shares 85 % sequence identity with the PT domain of SazB and SasC 82 % with the MT domain of SazB on the amino acid level. To elucidate the BGC function, the genes were cloned into pPWW50A and heterologously overexpressed in *S. albus* J1074.<sup>127</sup> LC-MS analysis revealed that SasA produces cWW (**11**) as in the case of SazA. The LC-MS chromatogram of the transformant with the complete cluster *sasABC* showed a predominant peak **13**, which is 164 Da larger than **11** and corresponds to that of streptoazine C (**13**). Three minor peaks were also detected. The structures were proven by NMR and ECD spectroscopy. The order of the tailoring reactions was elucidated by expression of the genes in different combinations. The double prenylation at C-3 of cWW (**11**) by SasB and the related cyclization to the pyrrolidine-indoline structure of streptoazine A (**12**) have to take place before SasC is able to methylate the nitrogen atoms of the former indole units (N-1/N-17) and thereby generates the final product streptoazine C (**13**, Fig. 22).



**Figure 22.** A: Biosynthetic pathway of streptoazine C in *Streptomyces aurantiacus*, B: prenylated products of SasB formed out of CDPs and dehydrogenated CDPs

## RESULTS AND DISCUSSION



**Figure 23.** LC-MS chromatograms of *S. albus* transformants harboring empty pPWW50A (i – ii) and *sasB* (iii – xii) with and without feeding of different CDPs

The function of SasB as PT for regular prenylation at C-3 was proven undoubtedly since the *sasB* transformant converted the fed precursor **11** to **12** (Fig. 23 iv). For a possible use of the PT in the field of synthetic biology, its potential substrate flexibility was investigated in vivo by feeding 12 CDPs to the *sasB* transformant. Four CDPs, *i.e.* cWF, cWY, cWL, and cWM, were almost completely converted by SasB to C-3 regularly mono-prenylated products **14** – **17** (Fig. 22B, 23 v – viii). In addition, the dehydrogenated forms of the converted CDPs, cWΔF, cWΔY, cWΔL, and cWΔM, which are common metabolites of CDOs in CDPS-associated pathways, were also C-3 prenylated by SasB to **18** – **21** (Fig. 22B, 23 ix – xii). The structures were elucidated by interpretation of their NMR spectra. The results of the in vivo feeding experiments demonstrated that SasB possesses a highly flexible substrate specificity, especially in comparison with the strictly conserved one of SazB from *S. leeuwenhoekii*.<sup>213</sup>

For details on this work, please refer to publication 4.3.

## **4. Publications**

### **4.1 Biosynthesis of Guatrypmethine C Implies Two Different Oxidases for *exo* Double Bond Installation at the Diketopiperazine Ring**

## PUBLICATIONS



Biosynthesis of Guatrypmethine C Implies Two Different Oxidases for *exo* Double Bond Installation at the Diketopiperazine Ring

Lauritz Harken, Jing Liu, Oliver Kreuz, Robert Berger, and Shu-Ming Li\*

Cite This: *ACS Catal.* 2022, 12, 648–654

Read Online

ACCESS |



Metrics &amp; More

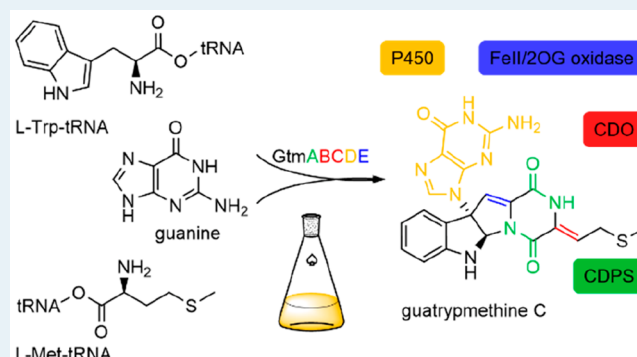


Article Recommendations



Supporting Information

**ABSTRACT:** Heterologous expression of a silent gene cluster from *Streptomyces cinnamoneus* led to identification of guatrypmethine C, a guaninylated cyclo-L-Trp-L-Met derivative with two *exo* double bonds at the diketopiperazine ring. Structural elucidation was achieved using MS,  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  NMR analyses including  $^1\text{H}$ – $^{13}\text{C}$  and  $^1\text{H}$ – $^{15}\text{N}$  HMBC. Gene combination and biochemical investigation proved that the formation of the two  $\text{C}=\text{C}$  bonds are catalyzed by two distinct enzyme families (i.e., cyclodipeptide oxidase and  $\text{Fe}^{\text{II}}/2\text{OG}$ -dependent oxygenase).



**KEYWORDS:** biosynthesis, cyclodipeptide oxidase, cyclodipeptide synthase, cytochrome P450 enzyme as guaninyl transferase,  $\text{Fe}^{\text{II}}/2\text{-oxoglutarate-dependent oxygenase}$ , natural products

Modified cyclodipeptides (CDPs) originating from either bacteria or fungi, belong to a group of natural products, exhibiting diverse structures and activities including antibacterial, antifungal, antitumor, and immuno-suppressive effects.<sup>1–8</sup> CDPs consist of two amino acids, which usually form a diketopiperazine (DKP) core via two amide bonds, resulting in increased stability compared to acyclic dipeptides.<sup>9</sup> CDPs can be assembled either by nonribosomal peptide synthetases (NRPSs), commonly occurring in fungi, or by cyclodipeptide synthases (CDPSs), mainly found in bacteria. NRPSs enabling CDP formation are two-modular proteins with a typical length of 2300–2500 amino acids and accept free amino acids as substrates. In comparison, CDPSs comprise merely 200–300 amino acids and use aminoacyl-tRNA for condensation.<sup>10–15</sup> The stable DKP backbone enables different modifications by tailoring enzymes, such as cytochrome P<sub>450</sub> enzymes (P450s), methyltransferases (MTs), prenyltransferases, cyclodipeptide oxidases (CDOs), and  $\text{Fe}^{\text{II}}/2\text{-oxoglutarate-dependent}$  ( $\text{Fe}^{\text{II}}/2\text{OG}$ ) oxidases.<sup>15–17</sup> One modification is the installation of  $\text{C}=\text{C}$  bonds *exo* of the DKP scaffold, catalyzed either by P450s, mainly in NRPS-related fungal pathways, or by CDOs in CDPS-associated bacterial pathways.<sup>18,19</sup> Both metabolites with only one *exo* double bond at the DKP ring (e.g., roquefortine C, phenylahistin, and guanitrypmycin A1-1) and those with two double bonds (e.g., neoechinulin B, albonoursin, and nocazine A) have been described in the literature (Figure 1).<sup>7,20–25</sup> The formation of the double bonds in the three bacterial products is catalyzed by one CDO for each compound,<sup>7,22,26</sup> while P450s are

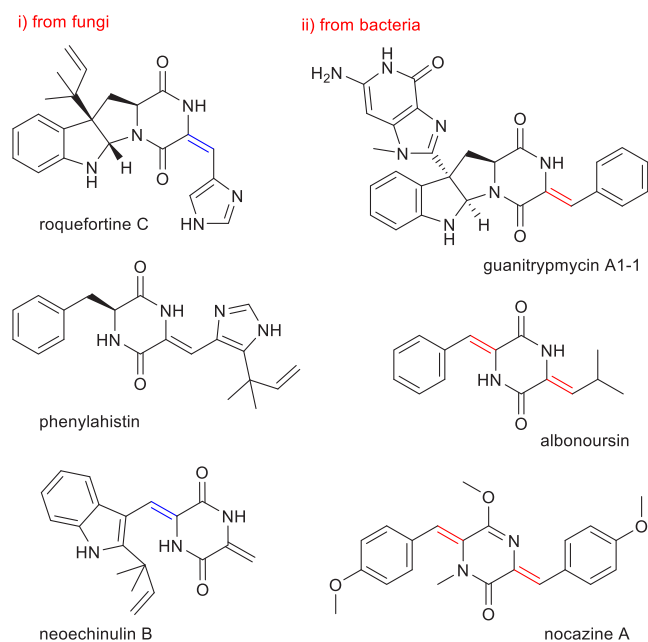
responsible for the installation of the  $\text{C}=\text{C}$  bond in roquefortine C and at the tryptophanyl side of neoechinulin B.<sup>24,27</sup>

$\text{Fe}^{\text{II}}/2\text{OG}$  oxidases belong to a large enzyme superfamily of oxidases involved in both primary and secondary metabolism.<sup>28,29</sup> They catalyze various oxidative biotransformation steps during biosynthesis of natural products, including hydroxylation, desaturation, halogenation, rearrangements, and epoxidation.<sup>30–39</sup> In CDPS-associated pathways,  $\text{Fe}^{\text{II}}/2\text{OG}$  oxidases are able to catalyze desaturation and hydroxylations.<sup>40,41</sup>

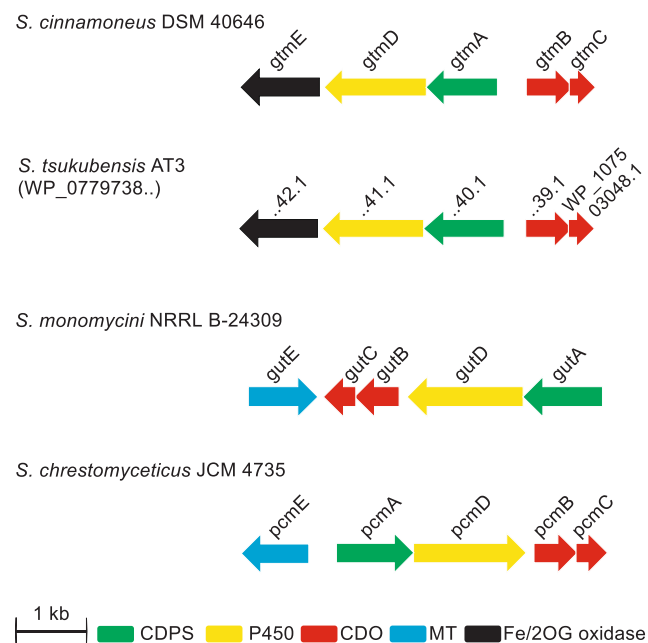
In a previous study, we identified the guanitrypmycin biosynthetic gene cluster comprising the five genes *gutABCDE* in *Streptomyces monomycini* (*S. monomycini*).<sup>22</sup> A Blastp search with the P450 enzyme GutD<sub>24309</sub> from this cluster revealed the presence of two putative P450s, that is, WP\_079274605.1 from *Streptomyces cinnamoneus* DSM 40646 (*S. cinnamoneus*) and WP\_077973841.1 from *Streptomyces tsukubensis* AT3, with sequence identities of 66% and 63% on the amino acid level, respectively. Analysis of the genetic loci in both strains led to identification of two very similar five-gene clusters with sequence identities between 60% and 77% (Figure 2, Table

Received: October 6, 2021

Revised: December 21, 2021



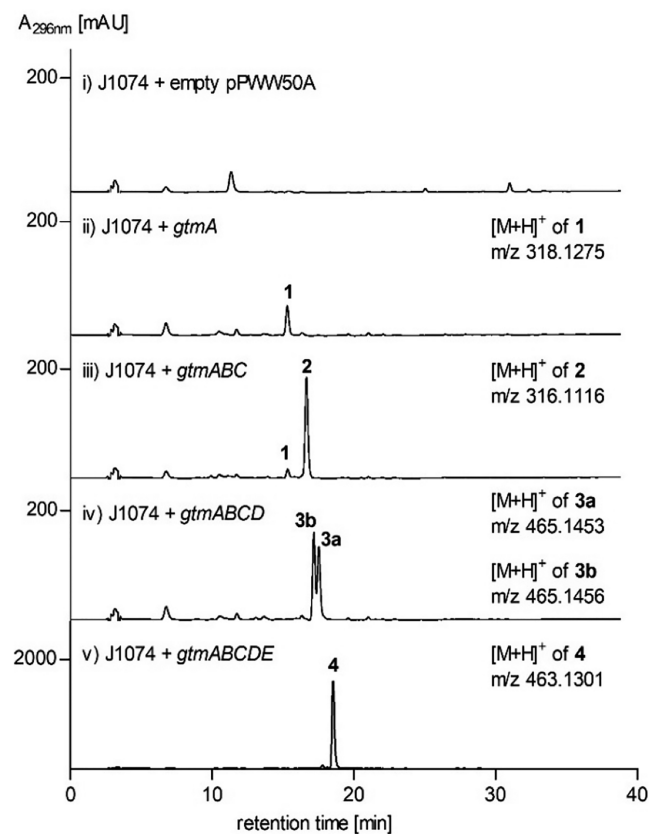
**Figure 1.** Examples of fungal and bacterial CDP derivatives with exo double bonds. The C=C bonds in red are installed by CDOs and those in blue by P450s.



**Figure 2.** Genetic organization of the *gtm* cluster from *Streptomyces cinnamoneus* and its homologous cluster from *Streptomyces tsukubensis* as well as the known *gut* cluster from *Streptomyces monomycini* and *pcm* cluster from *Streptomyces chrestomyceticus*.

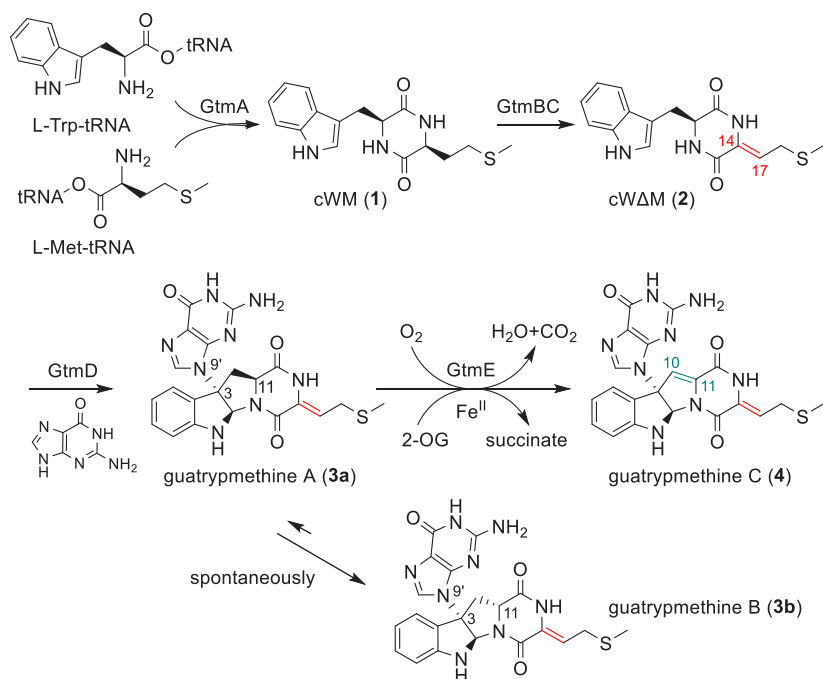
S2). In this study, we focus on the cluster from *S. cinnamoneus* and term it hereafter the *gtm* cluster. Similar to the *gut* cluster from *S. monomycini* and the orthologous *pcm* cluster from *Streptomyces chrestomyceticus*,<sup>42</sup> the *gtm* cluster contains a gene coding for a CDPS (GtmA), two genes for a CDO (GtmBC), and one gene for a putative P450 (GtmD). In contrast to *gut* and *pcm* clusters, which have the methyltransferase genes *gutE* or *pcmE*, the *gtm* cluster bears the putative Fe<sup>II</sup>/2OG-dependent oxygenase gene *gtmE*.

Targeted gene activation by heterologous expression has been proven to be a successful strategy for the discovery of novel natural products.<sup>43,44</sup> In analogy to our previous work on guanitrypmycins,<sup>22</sup> we amplified the sequences of *gtmA*–*gtmE* by performing PCR of the genomic DNA of *S. cinnamoneus* DSM 40646 using the primers listed in Table S3. The PCR fragments were cloned into the replicative vector pPWW50A under the control of the constitutive promoter *ermEp*.\*<sup>45</sup> These plasmids (Table S3) and the empty pPWW50A vector were transferred into the model expression host *Streptomyces albus* (*S. albus*) J1074 via conjugation.<sup>46</sup> The obtained transformants were cultivated in modified R5 medium (Table S1), extracted with EtOAc and analyzed using LC-MS (Figure 3).



**Figure 3.** LC-MS chromatograms of extracts from *S. albus* transformants.

In contrast to the negative control (Figure 3i), the culture extract of the *gtmA* transformant showed a predominant peak 1 with a  $[M+H]^+$  ion at  $m/z$  318.1275 (Figure 3ii), corresponding well to that of *cyclo*-L-Trp-L-Met (cWM). Isolation and comparison of the <sup>1</sup>H NMR data confirmed 1 to be cWM.<sup>47</sup> Having determined the function of GtmA as a cWM synthase, we expressed it together with the genes *gtmBC* for a CDO as *gtmABC* in *S. albus*. LC-MS analysis revealed 2 with a  $[M+H]^+$  ion at  $m/z$  316.1116 as the predominant peak in the transformant (Figure 3iii), indicating that GtmBC abstracts two hydrogen atoms from 1, as expected for a CDO. Combination of CDPS and CDO with the putative P450 GtmD, as in the transformant harboring *gtmADBC*, led to the identification of the two new peaks 3a at 18.1 min and 3b at 17.7 min with similar UV-spectra (Figure S2) and nearly the

Scheme 1. Proposed Biosynthetic Pathway of Guatrypmethine C in *Streptomyces cinnamoneus*

same  $[M+H]^+$  ions at  $m/z$  465.145 (Figure 3iv). Expression of all the five genes *gtmADEBC* in the same orientation resulted in the formation of the mere product 4 with a  $[M+H]^+$  ion at  $m/z$  463.1301, two daltons less than those of 3a and 3b.

Compounds 2, 3a, 3b, and 4 were subsequently isolated from large-scale fermentations of the respective *S. albus* transformants and subjected to NMR analysis. Comprehensive interpretation of their NMR data including  $^1\text{H}$ ,  $^{13}\text{C}$ , COSY, HSQC and HMBC (Tables S4–8, Figures S3–32) confirmed that 2 is a dehydrogenated derivative of cWM with an installed double bond *exo* of the DKP ring at the methionyl residue. Five signals for a guaninyl moiety were detected at approximately  $\delta_{\text{C}}$  118, 135, 151, 153, and 155–157 ppm in the  $^{13}\text{C}$  NMR spectra of 3a, 3b, and 4. HMBC analysis proved that this guaninyl residue is attached to the C3 of the CDP derivatives, leading to the formation of a hexahydropyrrolo-[2,3-*b*]indole framework. Interpretation of the NOESY spectra of 3a and 3b revealed that they are isomers. Correlations were observed for 3a from H-11 to H-2 and H-10b, whereas H-11 in 3b correlated with H-4 and H-10a, confirming that 3a is *s*- and 3b *r*-configured at C-11. Thus, the configuration of 3a at this position corresponds to that of the CDPS product cWM. Comparative analysis of the NOESY spectra of 3a, 3b, and 4 also revealed the configuration of C-2 as *R* and C-3 as *S* for all three compounds. H-2 correlates with H-8' in 3a, 3b, and 4, as well as with H-11 in 3a, and 4. No significant NOE was observed between H-2 and H-11 in the spectrum of 3b. Comparison of the  $^1\text{H}$  NMR spectra revealed that the signals of H-10a, H-10b, and H-11 of 3a and 3b in the upfield range (2.68–5.37 ppm) were absent in the spectrum of 4. Instead, an additional singlet at 6.91 ppm was observed. Accordingly, the  $^{13}\text{C}$  signals for C-10 and C-11 were deshielded from approximately  $\delta_{\text{C}}$  39–40 and 56–57 ppm in 3a and 3b to  $\delta_{\text{C}}$  114.6 and 135.6 ppm in the  $^{13}\text{C}$  spectrum of 4, respectively. All these data unequivocally proved the installation of a second double bond *exo* of the DKP ring between C-10 and C-11.

Three singlets at  $\delta_{\text{H}}$  6.3–6.5, 7.1–7.8, and 10.7–10.8 ppm, corresponding to those of H-10', H-8', and H-1' of the guaninyl residue, were detected in the  $^1\text{H}$  NMR spectra of 3a, 3b, and 4, indicating its linkage to C-3 of the CDP derivatives via either N-7' or N-9'. To prove the connecting position of the guaninyl residue, spores of the *gtmADBC* and *gtmADEBC* transformants were inoculated in modified RS medium containing  $^{15}\text{NH}_4\text{Cl}$ . Satisfactory signals in the  $^{15}\text{N}$  inverse-gated NMR as well as  $^1\text{H}$ – $^{15}\text{N}$  HSQC and  $^1\text{H}$ – $^{15}\text{N}$  HMBC spectra were achieved for  $^{15}\text{N}$ -labeled 3b and 4. Key correlations from H-2 to N-9' in the  $^1\text{H}$ – $^{15}\text{N}$  HMBC spectra of both compounds were clearly observed. Furthermore, correlation from H-10b to N9' also exists in the spectrum of 3b. This unambiguously proved that the guaninyl residue is attached from its N-9' to C-3 of the tryptophanyl moiety. Considering the involvement of guanine, L-Trp, and L-Met in the structures, 3a, 3b, and 4 were termed guatrypmethine A, B, and C, respectively.

In analogy to the guanitryptmycin biosynthesis in *S. monomycinii*, it can be assumed that the CDPS GtmA assembles cWM (1), the CDO GtmBC dehydrogenates 1 to cWΔM (2), and the P450 GtmD transfers a guanine onto 2 to form 3a and 3b. As the last step, the putative  $\text{Fe}^{\text{II}}$ /2OG oxidase GtmE catalyzes the second dehydrogenation, resulting in the formation of 4 (Scheme 1). Cultivation of *S. cinnamoneus* under different conditions as well as monitoring the accumulation of 1, 2, 3a, 3b, and 4 indicated that the *gtm* cluster is silent in this strain (see SI for details).

The presence of the isomeric pair 3a and 3b raised questions regarding their relationship. Since 3a has the same configuration as L-tryptophan, we hypothesized that 3a is the actual product and can be converted to 3b nonenzymatically via a keto–enol–tautomerism. To confirm this, 3a and 3b were incubated separately in a mixture of  $\text{CD}_3\text{OD}$  and  $\text{D}_2\text{O}$  (1:1) at different pH values for 16 h at 25 °C, and their changes were analyzed using LC-MS. As expected, no conversion took place at a pH of 2 (Figure S34), while at a pH of 8, 3a was slightly

converted to **3b** with the predominant  $[M+H]^+$  ion at  $m/z$  466.1515. Compared with the main  $[M+H]^+$  ion of **3a** at  $m/z$  465.1453, this value proves that one hydrogen atom is exchanged to deuterium via keto–enol–tautomerism (Figure S35). At a pH of 10, both isomers in the two samples showed exchange of one hydrogen atom to deuterium, and the **3a** peak is almost completely converted to **3b** (Figure S36). This clearly confirmed **3a** as the pathway intermediate and **3b** as the stable isomer (Scheme 1).

$\text{Fe}^{\text{II}}/2\text{OG}$  oxidases in CDPS-related pathways have only been reported in bicyclomycin biosynthesis, where BcmC, BcmE, and BcmG catalyze hydroxylations, and BcmF the conversion of a methyl to methylene group.<sup>40,41</sup> As aforementioned, the dehydrogenation at the DKP ring is usually carried out by CDOs in CDPS-related pathways and has not been reported for a  $\text{Fe}^{\text{II}}/2\text{OG}$  oxidase, which awoke our interest to characterize GtmE biochemically. The coding sequence of GtmE was cloned into pET28a for overproduction in *E. coli*. The recombinant His<sub>6</sub>-tagged GtmE was purified on Ni-NTA agarose to near homogeneity (Figure S38). GtmE (1.5  $\mu\text{M}$ ) was incubated with 0.5 mM **3a** in the presence of  $(\text{NH}_4)_2\text{Fe}^{\text{II}}(\text{SO}_4)_2$ , 2-oxoglutarate, and ascorbate at pH 7.5 and 37 °C for 30 min. As a negative control, GtmE was heat-inactivated. The reactions were terminated with MeOH and analyzed using LC-MS (Figure 4). As shown in Figure 4i, **3a** was almost completely converted to **4** by GtmE, which was identified by comparing its retention time, UV-spectrum, and MS-MS fragmentation pattern to those of an authentic standard isolated in this study. A product yield of 23% was achieved in the assay without  $(\text{NH}_4)_2\text{Fe}^{\text{II}}(\text{SO}_4)_2$  (Figure 4iii), indicating that the purified protein is partially a holoenzyme. No product formation was detected in either the negative control, the assay with EDTA, or the assay without 2-oxoglutarate (Figure 4ii, iv, v). This unequivocally proved that GtmE acts as a  $\text{Fe}^{\text{II}}/2\text{OG}$  oxidase. In these assays, **3a** was partially converted nonenzymatically to **3b**. Using **3b** instead of **3a** as a substrate, a low product formation of approximately 2% was observed (Figure 4vi). It could thus be speculated that the desaturation step catalyzed by GtmE is a critical step to avoid the spontaneous tautomerization of **3a** to **3b**. Determination of the kinetic parameters revealed that the GtmE reaction followed Michaelis–Menten kinetics.  $K_{\text{M}}$  and  $k_{\text{cat}}$  for both **3a** and 2-oxoglutarate were calculated to be 0.11 mM and 0.2  $\text{s}^{-1}$ , respectively (Figures S39–S41).

The experimental finding that **3b** is energetically more stable than **3a** in an aqueous environment is supported by quantum chemical calculations on the density functional theory (DFT) level. The energies obtained for the optimized structures are compared (see SI for details and images of the molecular structures). When applying Boltzmann statistics and taking thermal corrections into account, the mole fractions of the isomers under standard conditions are estimated to be around 4% **3a** and 96% **3b** (Table 1 and S9–S22, Figures S42 and S43, see SI for details). This result is in line with the experimental observation of 8% **3a** and 92% **3b** in equilibrium in a basic environment (Figure S36). In this framework, the low product formation shown in Figure 4vi would be rationalized by the higher activation energy for the reaction from **3b** to **4** via the common radical intermediate and thus be less favorable compared to the pathway from **3a** to **4**. However, we are aware of the possibility that the specific structure of the  $\text{Fe}^{\text{II}}/2\text{OG}$  oxidase might exert a significant influence on this selectivity.

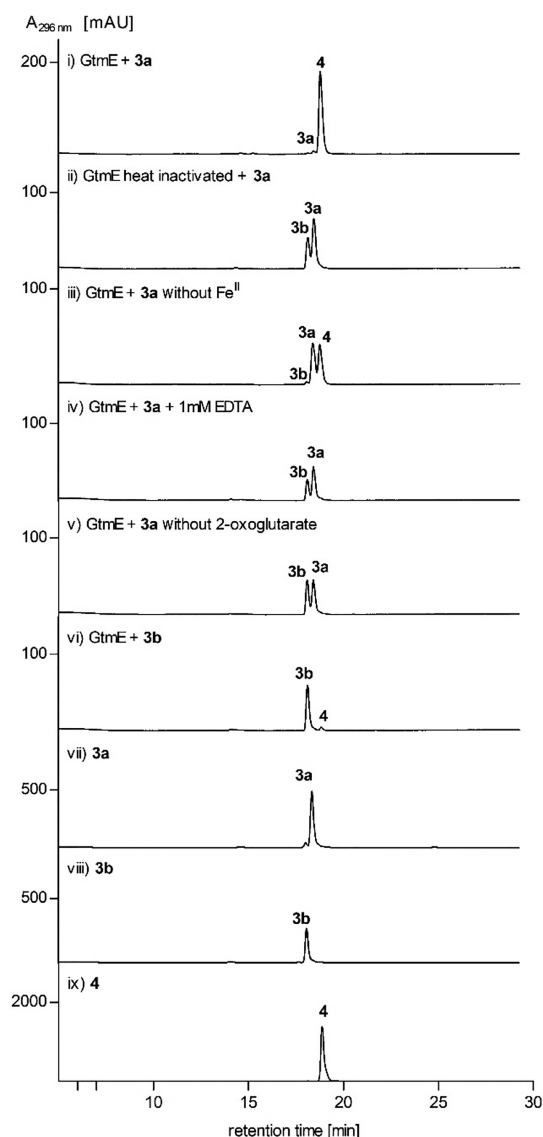


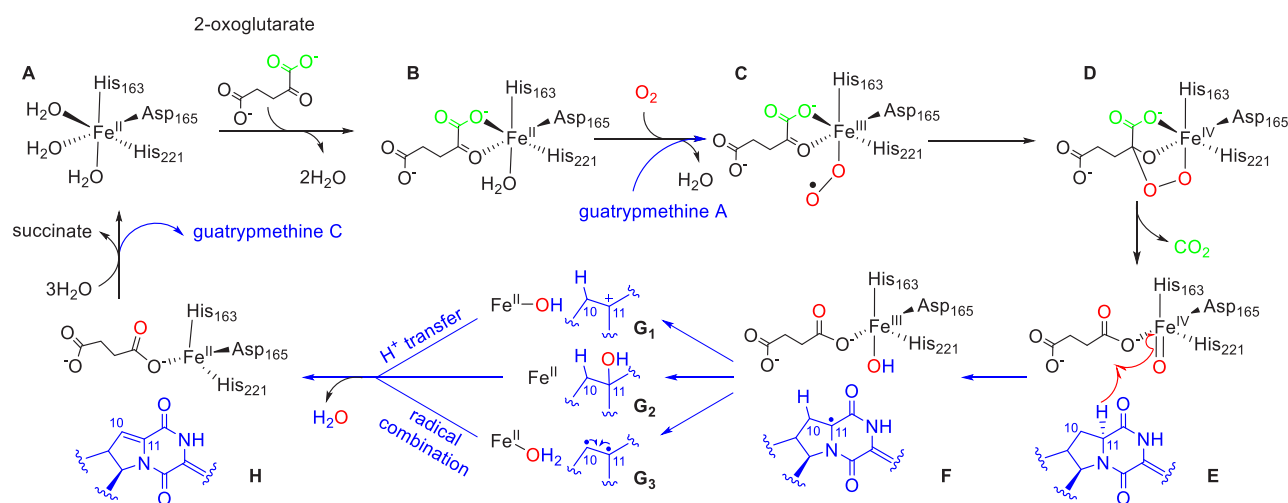
Figure 4. LC-MS chromatograms of GtmE assays.

Alignments of the GtmE sequence with those of known  $\text{Fe}^{\text{II}}/2\text{OG}$  oxidases (Figure S1) led to identification of His<sub>163</sub>, Asp<sub>165</sub>, and His<sub>221</sub> as the catalytic triad.<sup>40,48,49</sup> To understand the unusual function of a  $\text{Fe}^{\text{II}}/2\text{OG}$  oxidase for double bond installation, we proposed a reaction mechanism as depicted in Scheme 2. In the resting state, the central  $\text{Fe}^{\text{II}}$  is coordinated to His<sub>163</sub>, Asp<sub>165</sub>, His<sub>221</sub>, and three  $\text{H}_2\text{O}$  molecules (A). In the first step, 2-oxoglutarate ligates to  $\text{Fe}^{\text{II}}$  with its keto and carboxyl groups and replaces two  $\text{H}_2\text{O}$  molecules (B). When activated, **3a** binds to the active site of GtmE and one  $\text{O}_2$  molecule replaces one  $\text{H}_2\text{O}$ , forming the superoxo (C) and peroxohemiketal intermediate (D). Next, one  $\text{CO}_2$  molecule is released, and one hydrogen is abstracted from **3a** (E), very likely at C-11, since the radical intermediate can be mesomerically stabilized.  $\text{Fe}^{\text{IV}}$  is reduced to  $\text{Fe}^{\text{III}}$  (F). Thereafter, three mechanisms for the formation of the desaturated product **4** are possible, as described previously for other homologues enzymes.<sup>50–52</sup> One possibility is the formation of a carbocation intermediate (G1) through transfer of one electron, located at C-11, to  $\text{Fe}^{\text{III}}$ . Moreover, transfer of the hydroxyl radical from  $\text{Fe}^{\text{III}}$  to C-11 could result in G2. The



**Table 1. Relative Energies and Chemical Potentials of the Isomers and Their Conformers Together with the Resulting Boltzmann Weights**

structure	rel. electronic energy (in $\text{kJ}^*\text{mol}^{-1}$ )	rel. thermal correction (in $\text{kJ}^*\text{mol}^{-1}$ )	rel. chemical potential (in $\text{kJ}^*\text{mol}^{-1}$ )	Boltzmann weights	sum Boltzmann weights
3a I	12.33	−0.11	12.22	0.0042	0.0361
3a I'	8.02	0.59	8.61	0.0182	
3a II	17.08	−0.99	16.09	0.0009	
3a II'	11.52	−1.46	10.06	0.0102	
3a III	20.07	−2.24	17.83	0.0004	
3a III'	16.96	−3.07	13.89	0.0022	0.9639
3b I	0.00	0.00	0.00	0.5864	
3b I'	4.58	−0.72	3.86	0.1234	
3b II	4.39	0.59	4.98	0.0786	
3b II'	8.77	−1.17	7.60	0.0273	
3b III	5.32	−0.01	5.31	0.0687	
3b III'	8.49	−3.05	5.44	0.0654	
3b IV	4.04	5.42	9.46	0.0129	
3b IV'	13.59	1.79	15.38	0.0012	

**Scheme 2. Proposed Mechanism of GtmE-Mediated Double Bond Installation in Guatrypmethine A**

third possibility is the abstraction of one hydrogen at C-10 by  $\text{Fe}^{\text{III}}$  and formation of the biradical species ( $\text{G}_3$ ). The product **4** is formed (**H**) after  $\text{H}_2\text{O}$  elimination, released from the enzyme, and three  $\text{H}_2\text{O}$  molecules re-coordinate to  $\text{Fe}^{\text{II}}$ , associated with the loss of succinate (**A**).

In conclusion, we identified the gene cluster and elucidated the biosynthetic pathway of guatrypmethine C, a guaninylated *cyclo*-L-Trp-L-Met derivative with two double bonds *exo* of the diketopiperazine ring. We proved that these double bonds are formed by two distinct enzymes (i.e., cyclodipeptide oxidase and  $\text{Fe}^{\text{II}}$ /2OG-dependent oxygenase). To the best of our knowledge, a  $\text{Fe}^{\text{II}}$ /2OG oxidase-catalyzed double bond installation at the diketopiperazine ring has not been reported prior to this study. Furthermore, we demonstrated that feeding with inexpensive chemicals like  $^{15}\text{NH}_4\text{Cl}$  and subsequent  $^{15}\text{N}$  NMR analysis including  $^1\text{H}$ – $^{15}\text{N}$  HMBC are useful tools for elucidation of complex structures.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.1c04609>.

Additional experimental details, materials, and methods, including strains, plasmids, primers, LC-MS, NMR data, and DFT calculations are reported (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

Shu-Ming Li – Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, 35037 Marburg, Germany; [orcid.org/0000-0003-4583-2655](https://orcid.org/0000-0003-4583-2655); Phone: + 49- 6421-28-22461/25365; Email: [shuming.li@staff.uni-marburg.de](mailto:shuming.li@staff.uni-marburg.de)

### Authors

Lauritz Harken – Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, 35037 Marburg, Germany

Jing Liu – Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, 35037 Marburg, Germany

Oliver Kreuz – Fachbereich Chemie, Philipps-Universität Marburg, 35032 Marburg, Germany

Robert Berger – Fachbereich Chemie, Philipps-Universität Marburg, 35032 Marburg, Germany; [orcid.org/0000-0002-9107-2725](https://orcid.org/0000-0002-9107-2725)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acscatal.1c04609>

### Author Contributions

The manuscript was written through contributions of all authors.

### Funding

We thank DFG for funding (INST 160/620-1 and Li844/14-1). J.L. (201608310118) received a scholarship from the China Scholarship Council.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Rixa Kraut, Lena Ludwig-Radtke, Dr. Regina Ortmann, and Stefan Newel (Marburg) for MS and NMR measurement. Computer time provided by the Center for Scientific Computing (CSC) Frankfurt is gratefully acknowledged.

## ■ REFERENCES

- (1) Cain, C. C.; Lee, D.; Waldo, R. H., III; Henry, A. T.; Casida, E. J., Jr.; Wani, M. C.; Wall, M. E.; Oberlies, N. H.; Falkinham, J. O., III Synergistic antimicrobial activity of metabolites produced by a nonobligate bacterial predator. *Antimicrob. Agents Chemother.* **2003**, 47, 2113–2117.
- (2) Kohn, H.; Widger, W. The molecular basis for the mode of action of bicyclomycin. *Curr. Drug Targets. Infect. Disord.* **2005**, 5, 273–295.
- (3) Musetti, R.; Polizzotto, R.; Vecchione, A.; Borselli, S.; Zulini, L.; D'Ambrosio, M.; di Toppi, L. S.; Pertot, I. Antifungal activity of diketopiperazines extracted from *Alternaria alternata* against *Plasmodium falciparum*: an ultrastructural study. *Micron*. **2007**, 38, 643–650.
- (4) Ström, K.; Sjögren, J.; Broberg, A.; Schnürer, J. *Lactobacillus plantarum* MiLAB 393 produces the antifungal cyclic dipeptides cyclo(L-Phe-L-Pro) and cyclo(L-Phe-trans-4-OH-L-Pro) and 3-phenyllactic acid. *Appl. Environ. Microbiol.* **2002**, 68, 4322–4327.
- (5) Yamazaki, Y.; Tanaka, K.; Nicholson, B.; Deyanat-Yazdi, G.; Potts, B.; Yoshida, T.; Oda, A.; Kitagawa, T.; Orikasa, S.; Kiso, Y.; Yasui, H.; Akamatsu, M.; Chinen, T.; Usui, T.; Shinozaki, Y.; Yakushiji, F.; Miller, B. R.; Neuteboom, S.; Palladino, M.; Kanoh, K.; Lloyd, G. K.; Hayashi, Y. Synthesis and structure-activity relationship study of antimicrotubule agents phenylhistin derivatives with a dihydropiperazine-2,5-dione structure. *J. Med. Chem.* **2012**, 55, 1056–1071.
- (6) Waring, P.; Beaver, J. Gliotoxin and related epipolythiodioxopiperazines. *Gen. Pharmacol.* **1996**, 27, 1311–1316.
- (7) Giessen, T. W.; von Tesmar, A. M.; Marahiel, M. A. Insights into the generation of structural diversity in a tRNA-dependent pathway for highly modified bioactive cyclic dipeptides. *Chem. Biol.* **2013**, 20, 828–838.
- (8) Zhang, Q.; Li, S.; Chen, Y.; Tian, X.; Zhang, H.; Zhang, G.; Zhu, Y.; Zhang, S.; Zhang, W.; Zhang, C. New diketopiperazine derivatives from a deep-sea-derived *Nocardia alba* SCSIO 03039. *J. Antibiot.* **2013**, 66, 31–36.
- (9) Borthwick, A. D. 2,5-diketopiperazines: synthesis, reactions, medicinal chemistry, and bioactive natural products. *Chem. Rev.* **2012**, 112, 3641–3716.
- (10) Xu, W.; Gavia, D. J.; Tang, Y. Biosynthesis of fungal indole alkaloids. *Nat. Prod. Rep.* **2014**, 31, 1474–1487.
- (11) Izoré, T.; Cryle, M. J. The many faces and important roles of protein-protein interactions during non-ribosomal peptide synthesis. *Nat. Prod. Rep.* **2018**, 35, 1120–1139.
- (12) Payne, J. A.; Schoppet, M.; Hansen, M. H.; Cryle, M. J. Diversity of nature's assembly lines - recent discoveries in non-ribosomal peptide synthesis. *Mol. Biosyst.* **2017**, 13, 9–22.
- (13) Gondry, M.; Jacques, I. B.; Thai, R.; Babin, M.; Canu, N.; Seguin, J.; Belin, P.; Pernodet, J. L.; Moutiez, M. A comprehensive overview of the cyclodipeptide synthase family enriched with the characterization of 32 new enzymes. *Front Microbiol.* **2018**, 9, article 46.
- (14) Moutiez, M.; Belin, P.; Gondry, M. Aminoacyl-tRNA-utilizing enzymes in natural product biosynthesis. *Chem. Rev.* **2017**, 117, 5578–5618.
- (15) Harken, L.; Li, S.-M. Modifications of diketopiperazines assembled by cyclodipeptide synthases with cytochrome P450 enzymes. *Appl. Microbiol. Biotechnol.* **2021**, 105, 2277–2285.
- (16) Borgman, P.; Lopez, R. D.; Lane, A. L. The expanding spectrum of diketopiperazine natural product biosynthetic pathways containing cyclodipeptide synthases. *Org. Biomol. Chem.* **2019**, 17, 2305–2314.
- (17) Canu, N.; Moutiez, M.; Belin, P.; Gondry, M. Cyclodipeptide synthases: a promising biotechnological tool for the synthesis of diverse 2,5-diketopiperazines. *Nat. Prod. Rep.* **2020**, 37, 312–321.
- (18) Mishra, A. K.; Choi, J.; Choi, S. J.; Baek, K. H. Cyclodipeptides: An overview of their biosynthesis and biological activity. *Molecules* **2017**, 22, 1796.
- (19) Mikulski, L.; Schäfer, J.; Brockmeyer, K.; Kraut, R.; Li, S.-M. Comparative studies on similarities and differences of cyclodipeptide oxidases for installation of C-C double bonds at the diketopiperazine ring. *Appl. Microbiol. Biotechnol.* **2020**, 104, 2523–2536.
- (20) Kosalkova, K.; Dominguez-Santos, R.; Cotton, M.; Cotton, E.; Garcia-Estrada, C.; Liras, P.; Martin, J. F. A natural short pathway synthesizes roquefortine C but not meleagrin in three different *Penicillium roqueforti* strains. *Appl. Microbiol. Biotechnol.* **2015**, 99, 7601–7612.
- (21) Kanoh, K.; Kohno, S.; Asari, T.; Harada, T.; Katada, J.; Muramatsu, M.; Kawashima, H.; Sekiya, H.; Uno, I. (–)-Phenylhistin: a new mammalian cell cycle inhibitor produced by *Aspergillus ustus*. *Bioorg. Med. Chem.* **1997**, 7, 2847–2852.
- (22) Liu, J.; Xie, X.; Li, S.-M. Guanitryptamycin biosynthetic pathways imply cytochrome P450-mediated regio- and stereospecific guaninyl transfer reactions. *Angew. Chem., Int. Ed. Engl.* **2019**, 58, 11534–11540.
- (23) García-Estrada, C.; Ullán, R. V.; Álbillos, S. M.; Fernández-Bodega, M. A.; Durek, P.; von Dohren, D. H.; Martin, J. F. A single cluster of coregulated genes encodes the biosynthesis of the mycotoxins roquefortine C and meleagrin in *Penicillium chrysogenum*. *Chem. Biol.* **2011**, 18, 1499–1512.
- (24) Nies, J.; Li, S.-M. Prenylation and dehydrogenation of a C2-reversely prenylated diketopiperazine as a branching point in the biosynthesis of echinulin family alkaloids in *Aspergillus ruber*. *ACS Chem. Biol.* **2021**, 16, 185–192.
- (25) Gondry, M.; Lautru, S.; Fusai, G.; Meunier, G.; Menez, A.; Genet, R. Cyclic dipeptide oxidase from *Streptomyces noursei*. Isolation, purification and partial characterization of a novel, amino acyl -dehydrogenase. *Eur. J. Biochem.* **2001**, 268, 1712–1721.
- (26) Lautru, S.; Gondry, M.; Genet, R.; Pernodet, J. L. The albonoursin gene cluster of *S. noursei*: Biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. *Chem. Biol.* **2002**, 9, 1355–1364.
- (27) Martin, J. F.; Liras, P. Evolutionary formation of gene clusters by reorganization: the meleagrin/roquefortine paradigm in different fungi. *Appl. Microbiol. Biotechnol.* **2016**, 100, 1579–1587.
- (28) Islam, M. S.; Leissing, T. M.; Chowdhury, R.; Hopkinson, R. J.; Schofield, C. J. 2-Oxoglutarate-dependent oxygenases. *Annu. Rev. Biochem.* **2018**, 87, 585–620.
- (29) Herr, C. Q.; Hausinger, R. P. Amazing diversity in biochemical roles of Fe(II)/2-oxoglutarate oxygenases. *Trends Biochem. Sci.* **2018**, 43, 517–532.
- (30) Ran, H.; Wohlgemuth, V.; Xie, X.; Li, S.-M. A non-heme Fe(II)/2-oxoglutarate-dependent oxygenase catalyzes a double bond migration within a dimethylallyl moiety accompanied by hydroxylation. *ACS Chem. Biol.* **2018**, 13, 2949–2955.
- (31) Siitonen, V.; Selvaraj, B.; Niiranen, L.; Lindqvist, Y.; Schneider, G.; Metsä-Ketelä, M. Divergent non-heme iron enzymes in the

nogalamycin biosynthetic pathway. *Proc. Natl. Acad. Sci. U. S. A* **2016**, *113*, 5251–5256.

(32) Havemann, J.; Vogel, D.; Loll, B.; Keller, U. Cyclolization of D-lysergic acid alkaloid peptides. *Chem. Biol.* **2014**, *21*, 146–155.

(33) Jakubczyk, D.; Caputi, L.; Stevenson, C. E.; Lawson, D. M.; O'Connor, S. E. Structural characterization of EasH (*Aspergillus japonicus*) - an oxidase involved in cycloclavine biosynthesis. *Chem. Commun. (Camb.)* **2016**, *52*, 14306–14309.

(34) Mitchell, A. J.; Zhu, Q.; Maggiolo, A. O.; Ananth, N. R.; Hillwig, M. L.; Liu, X.; Boal, A. K. Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **2016**, *12*, 636–640.

(35) Zhu, Q.; Liu, X. Characterization of non-heme iron aliphatic halogenase WelO5\* from *Hapalosiphon welwitschii* C-52-3: Identification of a minimal protein sequence motif that confers enzymatic chlorination specificity in the biosynthesis of welwitindolelinones. *Beilstein J. Org. Chem.* **2017**, *13*, 1168–1173.

(36) Nakashima, Y.; Mori, T.; Nakamura, H.; Awakawa, T.; Hoshino, S.; Senda, M.; Senda, T.; Abe, I. Structure function and engineering of multifunctional non-heme iron dependent oxygenases in fungal meroterpenoid biosynthesis. *Nat. Commun.* **2018**, *9*, 104.

(37) Nakashima, Y.; Mitsunashi, T.; Matsuda, Y.; Senda, M.; Sato, H.; Yamazaki, M.; Uchiyama, M.; Senda, T.; Abe, I. Structural and computational bases for dramatic skeletal rearrangement in anditomin biosynthesis. *J. Am. Chem. Soc.* **2018**, *140*, 9743–9750.

(38) Mori, T.; Zhai, R.; Ushimaru, R.; Matsuda, Y.; Abe, I. Molecular insights into the endoperoxide formation by Fe(II)/-KG-dependent oxygenase Nvfl. *Nat. Commun.* **2021**, *12*, 4417.

(39) Gao, S.-S.; Naowarajna, N.; Cheng, R.; Liu, X.; Liu, P. Recent examples of (-ketoglutarate-dependent mononuclear non-haem iron enzymes in natural product biosyntheses. *Nat. Prod. Rep.* **2018**, *35*, 792–837.

(40) Meng, S.; Han, W.; Zhao, J.; Jian, X. H.; Pan, H. X.; Tang, G. L. A six-oxidase cascade for tandem C-H bond activation revealed by reconstitution of bicyclomycin biosynthesis. *Angew. Chem., Int. Ed. Engl.* **2018**, *57*, 719–723.

(41) Patteson, J. B.; Cai, W.; Johnson, R. A.; Santa Maria, K. C.; Li, B. Identification of the biosynthetic pathway for the antibiotic bicyclomycin. *Biochemistry* **2018**, *57*, 61–65.

(42) Shi, J.; Xu, X.; Zhao, E. J.; Zhang, B.; Li, W.; Zhao, Y.; Jiao, R. H.; Tan, R. X.; Ge, H. M. Genome mining and enzymatic total biosynthesis of purincyclamide. *Org. Lett.* **2019**, *21*, 6825–6829.

(43) Medema, M. H.; de Rond, R. T.; Moore, B. S. Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.* **2021**, *22*, 553–571.

(44) Scherlach, K.; Hertweck, C. Mining and unearthing hidden biosynthetic potential. *Nat. Commun.* **2021**, *12*, 3864.

(45) Zhu, Y.; Fu, P.; Lin, Q.; Zhang, G.; Zhang, H.; Li, S.; Ju, J.; Zhu, W.; Zhang, C. Identification of caerulomycin A gene cluster implicates a tailoring amidohydrolase. *Org. Lett.* **2012**, *14*, 2666–2669.

(46) Zaburannyi, N.; Rabyk, M.; Ostash, B.; Fedorenko, V.; Luzhetskyy, A. Insights into naturally minimised *Streptomyces albus* J1074 genome. *Bmc Genomics* **2014**, *15*, 97.

(47) Tian, W.; Sun, C.; Zheng, M.; Harmer, J. R.; Yu, M.; Zhang, Y.; Peng, H.; Zhu, D.; Deng, Z.; Chen, S. L.; Mobli, M.; Jia, X.; Qu, X. Efficient biosynthesis of heterodimeric C<sup>3</sup>-aryl pyrroloindoline alkaloids. *Nat. Commun.* **2018**, *9*, 4428.

(48) Roach, P. L.; Clifton, I. J.; Hensgens, C. M. H.; Shibata, N.; Schofield, C. J.; Hajdu, J.; Baldwin, J. E. Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **1997**, *387*, 827–830.

(49) Wang, Y.; Li, R. Cloning and sequencing the isopenicillin N synthetase (IPNS) gene from *Streptomyces cattleya*. *Wei Sheng Wu Xue. Bao.* **1996**, *36*, 87–92.

(50) Liao, H. J.; Li, J.; Huang, J. L.; Davidson, M.; Kurnikov, I.; Lin, T. S.; Lee, J. L.; Kurnikova, M.; Guo, Y.; Chan, N. L.; Chang, W. C. Insights into the desaturation of cyclopeptin and its C3 epimer catalyzed by a non-heme iron enzyme: Structural characterization and

mechanism elucidation. *Angew. Chem., Int. Ed. Engl.* **2018**, *57*, 1831–1835.

(51) Dunham, N. P.; Chang, W. C.; Mitchell, A. J.; Martinie, R. J.; Zhang, B.; Bergman, J. A.; Rajakovich, L. J.; Wang, B.; Silakov, A.; Krebs, C.; Boal, A. K.; Bollinger, J. M., Jr. Two distinct mechanisms for C-C desaturation by Iron(II)- and 2-(oxo)glutarate-dependent oxygenases: Importance of heteroatom assistance. *J. Am. Chem. Soc.* **2018**, *140*, 7116–7126.

(52) Martinez, S.; Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* **2015**, *290*, 20702–20711.

## Supporting information

### **Biosynthesis of Guatrypmethine C Implies Two Different Oxidases for *exo* Double Bond Installation at the Diketopiperazine Ring**

Lauritz Harken,<sup>1</sup> Jing Liu,<sup>1</sup> Oliver Kreuz,<sup>2</sup> Robert Berger,<sup>2</sup> and Shu-Ming Li<sup>1\*</sup>

<sup>1</sup> Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany

<sup>2</sup> Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Straße 4, 35032 Marburg, Germany

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

Tel/Fax: + 49-6421-28-22461/25365.

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



## Table of contents

<b>Experimental procedures .....</b>	<b>49</b>
1. Bioinformatic tools.....	49
2. Bacterial strains and cultivation conditions.....	49
3. Comparison of biosynthetic gene clusters and alignments of Fe <sup>II</sup> /2-oxoglutarate-dependent oxidases ....	49
4. PCR amplification, gene cloning and plasmid construction.....	49
5. Heterologous expression in <i>Streptomyces albus</i> J1074 .....	49
6. LC-MS analysis.....	50
7. Isolation of the accumulated metabolites .....	50
8. <sup>1</sup> H and <sup>13</sup> C NMR analyses.....	50
9. <sup>15</sup> N feeding and <sup>15</sup> N NMR analysis.....	50
10. Structure elucidation .....	50
11. Product yield in <i>Streptomyces albus</i> transformants .....	50
12. Non-enzymatic conversion of guatrypmethine A to B via a keto-enol-tautomerism .....	50
13. Silent gene cluster in <i>Streptomyces cinnamoneus</i> DSM 40646.....	50
14. Overproduction and purification of GtmE in <i>E. coli</i> BL21 DE3 .....	51
15. Enzyme assay with GtmE .....	51
16. DFT calculations .....	51
17. Physiochemical properties of the compounds.....	52
<b>Supplementary tables.....</b>	<b>53</b>
Table S1. Bacterial strains used in this study .....	53
Table S2. Comparison of the biosynthetic gene clusters .....	54
Table S4. NMR data of <i>cyclo</i> -L-Trp-L-Met ( <b>1</b> ) (DMSO- <i>d</i> <sub>6</sub> ).....	56
Table S5. NMR data of <i>cyclo</i> -L-Trp-Δ-Met ( <b>2</b> ) (DMSO- <i>d</i> <sub>6</sub> ).....	57
Table S6. NMR data of guatrypmethine A ( <b>3a</b> ) (DMSO- <i>d</i> <sub>6</sub> ).....	58
Table S7. NMR data of guatrypmethine B ( <b>3b</b> ) (DMSO- <i>d</i> <sub>6</sub> ).....	59
Table S8. NMR data of guatrypmethine C ( <b>4</b> ) (DMSO- <i>d</i> <sub>6</sub> ).....	60
Table S9. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3a I</b> in units of Å .....	61
Table S10. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3a I'</b> in units of Å .....	62
Table S11. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3a II</b> in units of Å .....	63
Table S12. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3a II'</b> in units of Å .....	64
Table S13. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3a III</b> in units of Å .....	65
Table S14. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3a III'</b> in units of Å .....	66
Table S15. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3b I</b> in units of Å .....	67
Table S16. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3b I'</b> in units of Å .....	68
Table S17. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3b II</b> in units of Å .....	69
Table S18. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3b II'</b> in units of Å .....	70
Table S19. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3b III</b> in units of Å .....	71
Table S20. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3b III'</b> in units of Å .....	72
Table S21. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3b IV</b> in units of Å.....	73
Table S22. Cartesian coordinates (xyz coordinates) of the equilibrium structure <b>3b IV'</b> in units of Å.....	74
<b>Supplementary figures .....</b>	<b>75</b>
Figure S1. Sequence alignments of GtmE from <i>S. cinnamoneus</i> with known Fe <sup>II</sup> /2OG oxygenases .....	75
Figure S2. UV spectra of the identified products .....	76

## PUBLICATIONS

Figure S3. <sup>1</sup> H-NMR spectrum of cyclo-L-Trp-L-Met ( <b>1</b> ) in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	77
Figure S4. <sup>1</sup> H-NMR spectrum of <i>cyclo</i> -L-Trp-Δ-Met ( <b>2</b> ) in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	78
Figure S5. <sup>13</sup> C{ <sup>1</sup> H}-NMR spectrum of <i>cyclo</i> -L-Trp-Δ-Met ( <b>2</b> ) in DMSO- <i>d</i> <sub>6</sub> (125 MHz) .....	79
Figure S6. <sup>1</sup> H- <sup>1</sup> H COSY spectrum of <i>cyclo</i> -L-Trp-Δ-Met ( <b>2</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	80
Figure S7. <sup>1</sup> H- <sup>13</sup> C HSQC spectrum of <i>cyclo</i> -L-Trp-Δ-Met ( <b>2</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	81
Figure S8. <sup>1</sup> H- <sup>13</sup> C HMBC spectrum of <i>cyclo</i> -L-Trp-Δ-Met ( <b>2</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	82
Figure S9. <sup>1</sup> H-NMR spectrum of guatrypmethine A ( <b>3a</b> ) in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	83
Figure S10. <sup>13</sup> C{ <sup>1</sup> H}-NMR spectrum of guatrypmethine A ( <b>3a</b> ) in DMSO- <i>d</i> <sub>6</sub> (125 MHz) .....	84
Figure S11. <sup>1</sup> H- <sup>1</sup> H COSY spectrum of guatrypmethine A ( <b>3a</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	85
Figure S12. <sup>1</sup> H- <sup>13</sup> C HSQC spectrum of guatrypmethine A ( <b>3a</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	86
Figure S13. <sup>1</sup> H- <sup>13</sup> C HMBC spectrum of guatrypmethine A ( <b>3a</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	87
Figure S14. NOESY spectrum of guatrypmethine A ( <b>3a</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	88
Figure S15. <sup>1</sup> H-NMR spectrum of guatrypmethine B ( <b>3b</b> ) in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	89
Figure S16. <sup>13</sup> C{ <sup>1</sup> H}-NMR spectrum of guatrypmethine B ( <b>3b</b> ) in DMSO- <i>d</i> <sub>6</sub> (125 MHz) .....	90
Figure S17. <sup>15</sup> N{ <sup>1</sup> H, <sup>13</sup> C} inverse gated spectrum of guatrypmethine B ( <b>3b</b> ) in DMSO- <i>d</i> <sub>6</sub> (40 MHz).....	91
Figure S18. <sup>1</sup> H- <sup>1</sup> H COSY spectrum of guatrypmethine B ( <b>3b</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	92
Figure S19. <sup>1</sup> H- <sup>13</sup> C HSQC spectrum of guatrypmethine B ( <b>3b</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	93
Figure S20. <sup>1</sup> H- <sup>13</sup> C HMBC spectrum of guatrypmethine B ( <b>3b</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	94
Figure S21. <sup>1</sup> H- <sup>15</sup> N HSQC spectrum of guatrypmethine B ( <b>3b</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	95
Figure S22. <sup>1</sup> H- <sup>15</sup> N HMBC spectrum of guatrypmethine B ( <b>3b</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	96
Figure S23. NOESY spectrum of guatrypmethine B ( <b>3b</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	97
Figure S24. <sup>1</sup> H-NMR spectrum of guatrypmethine C ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	98
Figure S25. <sup>13</sup> C{ <sup>1</sup> H}-NMR spectrum of guatrypmethine C ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> (125 MHz) .....	99
Figure S26. <sup>15</sup> N{ <sup>1</sup> H, <sup>13</sup> C} inverse gated spectrum of guatrypmethine C ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> (40 MHz).....	100
Figure S27. <sup>1</sup> H- <sup>1</sup> H COSY spectrum of guatrypmethine C ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	101
Figure S28. <sup>1</sup> H- <sup>13</sup> C HSQC spectrum of guatrypmethine C ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	102
Figure S29. <sup>1</sup> H- <sup>13</sup> C HMBC spectrum of guatrypmethine C ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	103
Figure S30. <sup>1</sup> H- <sup>15</sup> N HSQC spectrum of guatrypmethine C ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	104
Figure S31. <sup>1</sup> H- <sup>15</sup> N HMBC spectrum of guatrypmethine C ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	105
Figure S32. NOESY spectrum of guatrypmethine C ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .....	106
Figure S33. MS-MS of [M+H] <sup>+</sup> ions of the identified products.....	107
Figure S34. LC-MS data of <b>3a</b> and <b>3b</b> after incubation in CD <sub>3</sub> OD / D <sub>2</sub> O 1:1 at pH 2.0 for 16 h.....	108
Figure S35. LC-MS data of <b>3a</b> and <b>3b</b> after incubation in CD <sub>3</sub> OD / D <sub>2</sub> O 1:1 at pH 8.0 for 16 h.....	109
Figure S36. LC-MS data of <b>3a</b> and <b>3b</b> after incubation in CD <sub>3</sub> OD / D <sub>2</sub> O 1:1 at pH 10.0 for 16 h.....	110
Figure S37. LC-MS chromatograms of the crude extract from <i>Streptomyces cinnamoneus</i> DSM 40646. ....	111
Figure S38. 12% SDS PAGE of the purified GtmE stained with coomassie brilliant blue R-250® .....	112
Figure S39. Determination of the kinetic parameters of the GtmE reaction for <b>3a</b> (A) and 2-oxoglutarate (B).....	113
Figure S40. Calibration curve of different concentrations of <b>4</b> . .....	114
Figure S41. Linearity test of the GtmE catalyzed desaturation.....	114
Figure S42. Energy minimized structures of <b>3a</b> .....	115
Figure S43. Energy minimized structures of <b>3b</b> .....	116
<b>References</b> .....	<b>117</b>

## Experimental procedures

### Bioinformatic tools

Gene and protein sequences used in this study were obtained from the NCBI database (<https://www.ncbi.nlm.nih.gov/>), compared with each other by the Basic Local Alignment Search Tool (BLAST) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and analyzed by antiSMASH bacterial version.<sup>1</sup> Sequence alignments were carried out with Clustal omega and MEGA X and displayed with ESPript 3.0.<sup>2-4</sup> Snapgene® (GSL Biotech LLC, San Diego, United States) was used to design primers and for the simulation of enzymatic digestion.

### Bacterial strains and cultivation conditions

All bacterial strains used in this study are listed in **Table S1**.

*Streptomyces cinnamoneus* DSM 40646 was purchased from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH (Braunschweig, Germany). It was maintained on GYM medium (glucose 4.0 g/L, yeast extract 4.0 g/L, malt extract 10.0 g/L agar 15.0 g/L, pH 7.2) at 28 °C and cultivated for DNA extraction in liquid TSB medium (tryptic soy broth) purchased from Merck KGaA (Darmstadt, Germany) at 28 °C and 190 rpm for 4 days.

*Streptomyces albus* J1074<sup>5</sup> and its conjugants were cultivated on solid MS medium (mannitol 20.0 g/L, soy flour 20.0 g/L, agar 15.0 g/L) or in liquid modified R5 medium (sucrose 103.0 g/L, glucose 10.0 g/L, yeast extract 5.0 g/L, MgCl<sub>2</sub>•6 H<sub>2</sub>O 12 g/L, K<sub>2</sub>SO<sub>4</sub> 0.25 g/L, tryptone / peptone ex casein 0.1g/L, MOPS 21.0 g/L, trace element solution 2 mL/L (0.1 g/L each of ZnSO<sub>4</sub>•7 H<sub>2</sub>O, FeSO<sub>4</sub>•7 H<sub>2</sub>O, MnCl<sub>2</sub>•4 H<sub>2</sub>O, CaCl<sub>2</sub>•6 H<sub>2</sub>O and NaCl), pH 7.2) at 28 °C and 190 rpm for 7 days.<sup>5,6</sup> The transformants were selected with apramycin (50 µg/mL).

*E. coli* strains were cultivated in liquid or solid Luria-Bertani (LB) medium (tryptone / peptone ex casein 10.0 g/L, yeast extract 5.0 g/L, NaCl 10.0 g/L) and selected with ampicillin (100 µg/mL), kanamycin (50µg/mL), apramycin (50 µg/mL) or chloramphenicol (25 µg/mL), if necessary.

### Comparison of biosynthetic gene clusters and alignments of Fe<sup>II</sup>/2-oxoglutarate-dependent oxidases

The biosynthetic gene cluster (BGC) of interest was identified by BLASTP comparison of the cytochrome P450 enzyme GutD<sub>24309</sub> from *Streptomyces monomycini* NRRL B-24309, which is an orthologue of PcmD from *Streptomyces chrestomyceticus* JCM 4735.<sup>7,8</sup> A similar BGC was found in *Streptomyces tsukubensis* AT3. A detailed comparison and prediction of gene functions are given in **Table S2**. The catalytic triad of the Fe<sup>II</sup>/2-oxoglutarate-dependent oxidase GtmE, His<sub>163</sub>, Asp<sub>165</sub>, and His<sub>221</sub> was identified by alignments of the amino acid sequence with those of the known Fe<sup>II</sup>/2-oxoglutarate-dependent oxidases BcmF from *Streptomyces sapporensis* and the isopenicillin N synthases (IPNS) from *Aspergillus nidulans* and *Streptomyces cattleya* (**Fig. S1**).<sup>9-11</sup> GtmE shows, on the amino acid level, low sequence identities of 32, 21 and 18 % with BcmF, the IPNSs from *Aspergillus nidulans* and *Streptomyces cattleya*, respectively.

### PCR amplification, gene cloning and plasmid construction

Genetic manipulation was accomplished according to the protocols of Green and Sambrook for *E. coli* and Kieser *et al.* for *Streptomyces*.<sup>6,12</sup> Genes were amplified from the genomic DNA of *Streptomyces cinnamoneus* DSM 40646 by using Phusion® High-Fidelity DNA Polymerase from New England Biolabs (NEB), 5x Phusion® GC-Reaction buffer NEB and primers listed in **Table S3**, which were ordered from Microsynth AG (Balgach, Switzerland). The PCR fragments were either cloned into pGEM®-T Easy Vector (Promega GmbH, Walldorf, Germany) or directly inserted into pPWW50A *via* homologous recombination in *E. coli* DH5α.<sup>13</sup> Afterward, the sequence integrity of the plasmids was confirmed by sequencing (Seqlab - Microsynth AG, Balgach, Switzerland). Inserts on pGEM®-T Easy Vector were cut out by restriction endonucleases (see **Table S3**) and ligated with T4 ligase (Jena Bioscience GmbH, Jena, Germany) either into pET28a for protein overproduction in *E. coli* BL21 (DE3) or into pPWW50A for heterologous expression in *Streptomyces albus* J1074. All constructed plasmids are listed in **Table S3**.

### Heterologous expression in *Streptomyces albus* J1074

The plasmids with the pPWW50A vector as backbone were transferred into the non-methylating strain *E. coli* ET12567/pUZ8002.<sup>14</sup> After conjugation of the respective *E. coli* ET12567 transformants with *Streptomyces albus* J1074, positive conjugants were selected with apramycin and confirmed by PCR after DNA extraction. The respective conjugants of *Streptomyces albus* were cultivated for gene expression in 50 mL modified R5 medium with 50 µg/mL apramycin in 250 mL baffled conical flasks at 28 °C, 190 rpm for seven days. 1 mL culture was extracted twice with 800 µL ethyl acetate. The organic phases were combined and evaporated. The residue was dissolved in 300 µL methanol. 5 µL were analyzed by LC-MS.

### LC-MS analysis

The samples were measured on an Agilent HPLC 1260 series system with a photo diode array detector and a Bruker micro TOF QIII mass spectrometer by using a Multospher 120 RP-18 column (250 x 4 mm, 5  $\mu$ m, CS-Chromatographie Service GmbH, Langerwehe, Germany) for separation. A linear gradient from 5 – 100 % acetonitrile in water, containing 0.1 % formic acid, in 40 min at a flow rate of 0.25 mL/min achieved a sufficient separation. The parameters of the MS device were set as follows: electrospray positive ion mode for ionization, capillary voltage of 4.5 kV and collision energy of 8.0 eV.

### Isolation of the accumulated metabolites

*Streptomyces albus* transformants were cultivated in 1 – 5 L modified R5 medium with 50  $\mu$ g/mL apramycin (performed in 2 L conical flasks with 0.5 L medium) at 28 °C, 190 rpm for seven days. Afterward, the bacterial cultures were centrifuged at 4500 rpm for 30 min. The supernatants were then extracted twice with the same volume of ethyl acetate and the combined organic phases were evaporated to dryness. The extracts were separated on a silica gel column (0.04 – 0.063 mm) with a gradient of 2 – 20 % (v/v) methanol in dichloromethane. The resulting fractions were further purified on an Agilent HPLC 1260 series, with a semi-preparative Agilent ZORBAX Eclipse XDB C18 HPLC column (9.4 x 250 mm, 5  $\mu$ m), using 28 % acetonitrile in water as mobile phase and a flow rate of 2.0 mL/min.

### <sup>1</sup>H and <sup>13</sup>C NMR analyses

<sup>1</sup>H and <sup>13</sup>C NMR spectra of the purified compounds were recorded on a JEOL ECA-500 spectrometer (JEOL, Akishima, Tokyo, Japan). The samples were dissolved in DMSO-*d*<sub>6</sub>. The spectra were processed using MestReNov. 9.0.0 (Mestrelab Research, Santiago de Compostella, Spain). The NMR data are summarized in **Tables S4 – 8** and the spectra are given in **Figs. S3 – 32**.

### <sup>15</sup>N feeding and <sup>15</sup>N NMR analysis

To achieve a sufficient signal strength for <sup>15</sup>N NMR spectral analysis, 100 mg of <sup>15</sup>NH<sub>4</sub>Cl (Sigma-Aldrich, Darmstadt, Germany) were added to 100 mL modified R5 medium. *Streptomyces albus* J1074 harboring the gene combinations of *gtmABCD* or *gtmABCDE* were cultivated in this media and the products were isolated under the same conditions as mentioned above. The samples were measured on a JEOL ECZ400 spectrometer (JEOL, Akishima, Tokyo, Japan) as inverse gated <sup>15</sup>N{<sup>1</sup>H, <sup>13</sup>C} spectrum, and <sup>1</sup>H-<sup>15</sup>N HSQC and <sup>1</sup>H-<sup>15</sup>N HMBC according to the manufacture's parameters. The spectra were referenced as described by Wishart *et al.*<sup>15</sup>

### Structure elucidation

Structure elucidation of the compounds was accomplished by comprehensive interpretation of their UV (**Fig. S2**), NMR (**Figs. S3 – 32**) and MS-MS data (**Fig. S33**). The connections and configurations were confirmed by analysis of the <sup>1</sup>H-<sup>13</sup>C HMBC, <sup>1</sup>H-<sup>15</sup>N HMBC and NOESY spectra. The structure of cWM (**1**) was proven by comparison of the <sup>1</sup>H-NMR data (**Table S4**) with those reported previously.<sup>16</sup> Comparison of the NMR spectra of guatrypmethine A (**3a**) and B (**3b**) with those of guatrypmethine C (**4**) reveals that the proton signals of H-10a, 10b and 11 in the upfield (2.68 – 5.37 ppm) are disappeared in the <sup>1</sup>H-NMR spectrum of **4**. An additional singlet at 6.91 ppm was observed instead. This proves the formation of a double bond between C-10 and C-11, which is also confirmed by a downfield shift of the respective <sup>13</sup>C signals to 114.3 and 135.9 ppm, respectively (**Tables S6 – 8**).

### Product yield in *Streptomyces albus* transformants

The isolated compound **4** was used as authentic standard for the quantification of the product yield by using the LC-MS method as described above.

### Non-enzymatic conversion of guatrypmethine A to B via a keto-enol-tautomerism

To prove the non-enzymatic conversion of guatrypmethine A (**3a**) to its stable isomer guatrypmethine B (**3b**) via a keto-enol-tautomerism, we incubated 1 mM of the purified compound **3a** or **3b** in CD<sub>3</sub>OD and D<sub>2</sub>O (1:1) at pH 2, pH 8 and pH 10 for 16 h at room temperature. LC-MS analyses of these samples confirmed that the conversion is favored in the alkaline milieu whereas under acidic conditions no conversion took place. At pH 8, **3a** is slightly converted to **3b** and the hydrogen atom at position 11 is exchanged to deuterium. At pH 10 even both isomers in both samples show the exchange of one atom hydrogen to deuterium and the peak of **3a** is almost completely converted to **3b**.

### Silent gene cluster in *Streptomyces cinnamoneus* DSM 40646

*Streptomyces cinnamoneus* DSM 40646 was cultivated in six different media (**Table S1**) for 7 to 14 days, at 28 °C and 190 rpm. 1 mL of each culture was extracted twice with 800  $\mu$ L ethyl acetate. The organic phases were

combined and evaporated. The residue was dissolved in 300  $\mu\text{L}$  methanol. 5  $\mu\text{L}$  were analyzed by LC-MS. Detection with the respective  $[\text{M}+\text{H}]^+$  ions of **1**, **2**, **3a**, **3b** and **4** revealed only very weak signal for an ion at  $m/z$   $318.127 \pm 0.005$  in the best obtained result from modified R5 medium after 7 days (**Fig. S37**), corresponding to that of cWM (**1**). However, this peak was also visible in the extract of the pure medium. Thus, the *gtm* gene cluster seems to be silent in *Streptomyces cinnamoneus* DSM 40646.

### Overproduction and purification of GtmE in *E. coli* BL21 DE3

To obtain recombinant GtmE, *E. coli* BL21 (DE3) cells harboring the plasmid pLH18 were grown in 5 mL LB media with 50  $\mu\text{g/mL}$  kanamycin at 37 °C and 230 rpm for 16 h. 5 mL of this pre-culture were transferred into 500 mL LB media with 50  $\mu\text{g/mL}$  kanamycin in 2 L conical flasks and cultivated to an absorption of 0.6 at 600 nm. After addition of isopropyl  $\beta$ -D-1-thiogalactopyranoside at a final concentration of 0.5 mM, the bacteria were further cultivated at 30 °C and 200 rpm for 12 h. After centrifugation at 4500 rpm and 4 °C for 30 min, the cell pellets were stored at -80 °C overnight and then re-suspended in 40 mL lysis buffer (50 mM  $\text{NaH}_2\text{PO}_4$ , 300 mM NaCl, 10 mM imidazole, pH 8.0 at 4 °C). Lysozyme from the chicken egg (Sigma-Aldrich, Darmstadt, Germany) and DNase (PanReac AppliChem GmbH, Darmstadt, Germany) were added to a final concentration of 1 mg/mL and 1  $\mu\text{g/mL}$ , respectively. After incubation on ice for 30 min., the cell walls were destroyed by sonication on ice. The proteins were separated from the cell debris by centrifugation at 13000 rpm and 4 °C for 1 h. The recombinant His<sub>6</sub>-tagged GtmE was purified with Ni-NTA agarose (Cube Biotech GmbH, Monheim, Germany) according to the protocol of the manufacturer. Imidazole was removed by using a PD MiniTrap™ G-25 column (GE Healthcare, Freiburg, Germany) equilibrated with storage buffer (50 mM TRIS, 20 % (v/v) glycerol, pH 8.0 at 4 °C). The concentration of GtmE was measured on a Nanodrop C2000 (Thermo Scientific, Braunschweig, Germany) as 0.65 mg/mL. The approximate yield was 0.3 mg protein per 1 L culture of *E. coli* BL21. The purity of GtmE was proven on a 12 % (w/v) SDS-PAGE (**Fig. S38**). The purified protein was stored at -80 °C.

### Enzyme assay with GtmE

The function of GtmE, as a  $\text{Fe}^{\text{II}}$ /2-oxoglutarate-dependent oxidase that catalyzes the conversion of **3a** to **4** by installing a double bond between C-10 and C-11 exo of the diketopiperazine core, was proven by an enzyme assay. 1.5  $\mu\text{M}$  GtmE was incubated with 0.5 mM **3a**, 1 mM 2-oxoglutarate, 1 mM  $(\text{NH}_4)_2\text{Fe}^{\text{II}}(\text{SO}_4)_2$ , and 1 mM ascorbic acid in  $\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-}$  buffer at pH 7.5 at a final concentration of 50 mM at 37 °C for 30 min. As negative control, GtmE was inactivated by heating at 100 °C for 20 min. The assays were terminated with 20  $\mu\text{L}$  methanol and centrifuged at 13.000 rpm and 4 °C. 5  $\mu\text{L}$  of the mixtures were measured by LC-MS.

The isolated compound **4** was used as authentic standard for the quantification of the product yield by using the LC-MS method as described above. The calibration curve is shown in **Fig. S40**.

For determination of the kinetic parameters of **3a** and the co-substrate 2-oxoglutarate, 0.2  $\mu\text{M}$  recombinant GtmE was incubated with 1 mM 2-oxoglutarate and **3a** in the range of 0.02 – 5 mM or 1.0 mM **3a** and 2-oxoglutarate from 0.02 to 5 mM under the same conditions as mentioned above (**Fig. S39**). The linearity of the GtmE-catalyzed reaction was determined up to 30 min with 0.5  $\mu\text{M}$  protein (**Fig. S41**).

### DFT calculations

All quantum chemical calculations were performed using the Turbomole software package (Version 7.3) (Turbomole GmbH, Karlsruhe, Germany). Structures were energy-minimized at the BP86/def2-TZVP level of theory<sup>17-20</sup> with the resolution of the identity method for the Coulomb part (RI-J)<sup>21</sup> enabled on the m3 numerical integration grid using the COSMO solvation model<sup>22</sup> for water (relative permittivity set to 80) and the D3 dispersion correction<sup>23</sup> with damping (BJ).<sup>24</sup>

The self-consistent field convergence criterion was chosen to be  $10^{-9} E_h$ . The structures **3a I-III**, **3a I'-III'**, **3b I-III** and **3b I'-III'** were energy-minimized to a gradient norm below the convergence threshold of  $10^{-4} E_h/a_0$ . For **3b IV** and **3b IV'** a slightly weaker convergence criterion of  $0.5 \times 10^{-3} E_h/a_0$  was used. Numerical harmonic vibrational frequency calculations were employed to confirm the presence of local minima (absence of imaginary harmonic vibrational frequencies) on the potential energy hypersurface.

To determine energy differences between structures, single point calculations were performed at the B3LYP/def2-TZVP<sup>25-27</sup> level of theory on the m4 numerical integration grid using again the COSMO solvation model for water and the D3(BJ) dispersion correction.

Thermal corrections to the electronic energies and zero-point vibrational energy contributions were calculated *via* the freeh module of Turbomole using data from numerical harmonic vibrational frequency calculations at the BP86/def2-TZVP level of theory using the COSMO solvation model for water. An overall scaling factor of 0.9953

## PUBLICATIONS

was employed for the harmonic vibrational frequencies.<sup>28</sup> The temperature was set to 298.15 K, the pressure to 0.1 MPa. Boltzmann weights were calculated accordingly.

Examining the arrangement of the guaninyl residue of the different conformers of **3a** and **3b**, it is plausible that rotation of this unit is facile, giving rise to numerous conformations. Conformers shown here were chosen by virtue of relative energy and of arrangement for relevant correlations in NOESY. Considering more conformers likely will not change results qualitatively. Fewer conformers for **3a** than for **3b** were found. Conformations like **IV** and **IV'** for **3b** are not energetically optimal for **3a** due to steric reasons.

All energy minimized structures are shown in **Figs. S42 – 43** and all Cartesian coordinates in **Tables S9 – 22**.

### Physiochemical properties of the compounds

- cWM (1):** white powder, chemical formula:  $C_{16}H_{19}N_3O_2S$ , HRMS (m/z): (ESI/[M+H]<sup>+</sup>) calculated for  $C_{16}H_{20}N_3O_2S$ : 318.1271, found 318.1275
- cWΔM (2):** white powder, chemical formula:  $C_{16}H_{17}N_3O_2S$ , HRMS (m/z): (ESI/[M+H]<sup>+</sup>) calculated for  $C_{16}H_{18}N_3O_2S$ : 316.1114, found 316.1116
- guatrypmethine A (3a):** white powder, chemical formula:  $C_{21}H_{20}N_8O_3S$ , HRMS (m/z): (ESI/[M+H]<sup>+</sup>) calculated for  $C_{21}H_{21}N_8O_3S$ : 465.1452, found 465.1453
- guatrypmethine B (3b):** white powder, chemical formula:  $C_{21}H_{20}N_8O_3S$ , HRMS (m/z): (ESI/[M+H]<sup>+</sup>) calculated for  $C_{21}H_{21}N_8O_3S$ : 465.1452, found 465.1456
- guatrypmethine C (4):** light yellow powder, chemical formula:  $C_{21}H_{18}N_8O_3S$ , HRMS (m/z): (ESI/[M+H]<sup>+</sup>) calculated for  $C_{21}H_{19}N_8O_3S$ : 463.1295, found 463.1301

## Supplementary tables

Table S1. Bacterial strains used in this study

Strain	Source	Growth Medium
<i>Streptomyces cinnamoneus</i> DSM 40646	DSMZ	GYM (solid/liquid) TSB (liquid) ISP4 (liquid) RA2 (liquid) modified R5 (liquid) MS (liquid)
<i>Streptomyces albus</i> J1074	5	MS (solid) modified R5 (liquid)
<i>E. coli</i> DH5α	Thermo Fisher Scientific Inc. (Waltham, MA, USA)	LB
<i>E. coli</i> BL21 (DE3)	Merck KGaA (Darmstadt, Germany)	LB
<i>E. coli</i> ET12567/pUZ8002	14	LB

DSMZ:	Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany)
GYM medium:	glucose 4.0 g/L, yeast extract 4.0 g/L, malt extract 10.0 g/L agar 15.0 g/L (for solid medium), pH 7.2
TSB medium:	tryptic soy broth purchased from Sigma-Aldrich
ISP4 medium	soluble starch 20.0 g/L, K <sub>2</sub> HPO <sub>4</sub> 2 g/L, NaCl 2.0 g/L, (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> 4.0 g/L, yeast extract 0.5 g/L, tryptone / peptone ex casein 1.0 g/L, trace element solution 2 mL/L (0.1 g/L each of ZnSO <sub>4</sub> •7 H <sub>2</sub> O, FeSO <sub>4</sub> •7 H <sub>2</sub> O, MnCl <sub>2</sub> •4 H <sub>2</sub> O, CaCl <sub>2</sub> •6 H <sub>2</sub> O and NaCl), pH 7.2
RA2 medium	glucose 10.0 g/L, malt extract 10.0 g/L, corn flour 5 g/L, soluble starch 6.0 g/L, maltose 10 g/L, trace element solution 2 mL/L (0.1 g/L each of ZnSO <sub>4</sub> •7 H <sub>2</sub> O, FeSO <sub>4</sub> •7 H <sub>2</sub> O, MnCl <sub>2</sub> •4 H <sub>2</sub> O, CaCl <sub>2</sub> •6 H <sub>2</sub> O and NaCl), pH 7.2
MS medium:	mannitol 20.0 g/L, soy flour 20.0 g/L, agar 15.0 g/L (for solid medium)
modified R5 medium:	sucrose 103.0 g/L, glucose 10.0 g/L, yeast extract 5.0 g/L, MgCl <sub>2</sub> •6 H <sub>2</sub> O 12 g/L, K <sub>2</sub> SO <sub>4</sub> 0.25 g/L, tryptone / peptone ex casein 0.1g/L, MOPS 21.0 g/L, trace element solution 2 mL/L (0.1 g/L each of ZnSO <sub>4</sub> •7 H <sub>2</sub> O, FeSO <sub>4</sub> •7 H <sub>2</sub> O, MnCl <sub>2</sub> •4 H <sub>2</sub> O, CaCl <sub>2</sub> •6 H <sub>2</sub> O and NaCl), pH 7.2













**Table S2.** Comparison of the biosynthetic gene clusters

<i>Streptomyces cinnamoneus</i> DSM 40646		<i>Streptomyces tsukubensis</i> AT3		<i>Streptomyces monomycini</i> NRRL B-24309 <sup>7</sup>		<i>Streptomyces chrestomyceticus</i> JCM 4735 <sup>8</sup>		Sequence identity [%]			Function
Protein (accession No.)	Size (aa)	Accession No.	Size (aa)	Protein (accession No.)	Size (aa)	Protein (accession No.)	Size (aa)	Gtm/AT3	Gtm/Gut	Gtm/Pcm	
GtmC (WP_071967256.1)	104	WP_107503048.1	105	GutC (WP_050502761.1)	114	PcmC (QE122733.1)	114	77	55	55	cyclodipeptide oxidase, subunit B
GtmB (WP_071967255.1)	186	WP_077973839.1	208	GutB (WP_208870198.1)	158	PcmB (QE122734.1)	199	66	65	58	cyclodipeptide oxidase, subunit A
GtmA (WP_071967254.1)	245	WP_077973840.1	309	GutA (WP_078624487.1)	282	PcmA (QE122736.1)	256	60	52	51	cyclodipeptide synthase
GtmD (WP_079274605.1)	404	WP_077973841.1	404	GutD (WP_050502760.1)	411	PcmD (QE122735.1)	411	76	66	65	cytochrome P <sub>450</sub>
GtmE (WP_177338671.1)	289	WP_077973842.1	310	---	---	---	---	68	---	---	Fe <sup>II</sup> /2-oxoglutarate- dependent oxidase

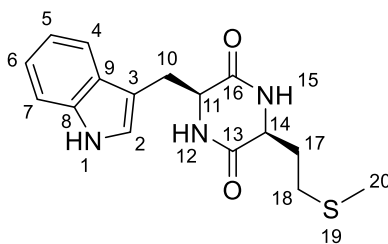


# PUBLICATIONS

**Table S3.** Primers, cloning sites and plasmids used in this study

Gene(s)	Primer sequences (5'-3')	Cloning site	Expression vector	Gene position and orientation in the construct	Plasmid
<i>gtmE</i>	catatgacctcgacacctcgacg ggatccatgacacgagtgcatcatgc	NdeI BamHI	pET28a (+)		pLH18
<i>gtmA</i>	catatgcatgtccggaaggttcagc ggatcctcacggccggttcccgc	NdeI BamHI	pPWW50A		pLH20
<i>gtmA</i>	catatgcatgtccggaaggttcagc ggatcctcacggccggttcccgc	NdeI	pPWW50A		pLH26
+ <i>gtmBC</i>	ccgcgcgatgatgcactcgtgtcatggatccgagagattcagcttggcc gattcattaatgcagagcttctagaactagttcaaggatccttgagaacctcg	BamHI		 	
<i>gtmAD</i>	catatgcatgtccggaaggttcagc ggatcctcaccacatgacgggcagcc	NdeI BamHI	pPWW50A		pLH25
+ <i>gtmBC</i>	actagtgcagagattcagcttggcc actagtcaaggatccttgagaacctcg	SpeI		 	
<i>gtmADE</i>	Catatgcatgtccggaaggttcagc ggatccatgacacgagtgcatcatgc	NdeI	pPWW50A		pLH23
+ <i>gtmBC</i>	ccgcgcgatgatgcactcgtgtcatggatccgagagattcagcttggcc gattcattaatgcagagcttctagaactagttcaaggatccttgagaacctcg	BamHI		  	

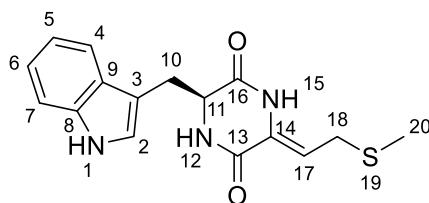
The plasmids pLH23 and pLH26 were constructed *via* homologous recombination for *gtmBC* directly in pPWW50A

**Table S4.** NMR data of *cyclo*-L-Trp-L-Met (**1**) (DMSO-*d*<sub>6</sub>)

Pos.	$\delta_{\text{H}}$	multi, J[Hz]
1	10.88	br s
2	7.04	d, 2.1
4	7.59	d, 8.0
5	6.94	td, 8.0, 1.0
6	7.02	td, 8.1, 1.1
7	7.29	d, 8.1
10	3.00	dd, 14.5, 4.6
	3.27	dd, 14.5, 4.0
11	4.13	m
12	8.11	d, 1.6
14	3.68	m
15	7.95	d, 1.8
17	0.94	m
	1.22	m
18	1.65	m
20	1.75	s

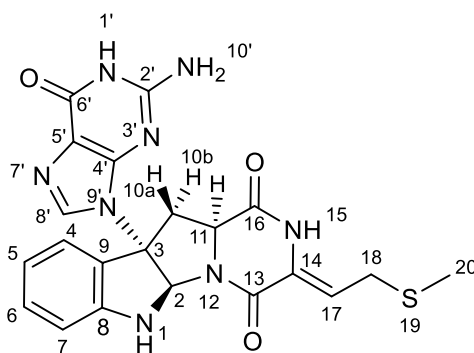
The NMR data of **1** correspond very well to those published previously.<sup>16</sup>

## PUBLICATIONS

**Table S5.** NMR data of *cyclo*-L-Trp- $\Delta$ -Met (**2**) (DMSO- $d_6$ )

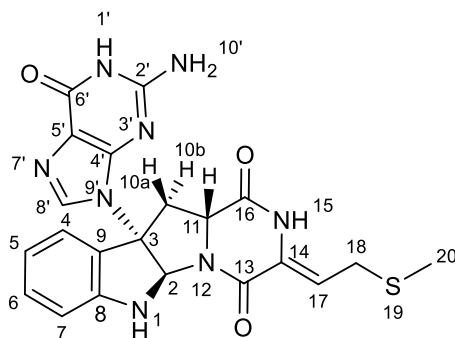
Pos.	$\delta_c$	$\delta_H$	multi, J[Hz]	$^1H$ - $^1H$ COSY	HMBC
1	-	10.84	br s	H-2	C-2, 3, 8, 9
2	124.6	6.99	d, 2.3	H-1	C-3, 9, 10
3	107.6	-	-	-	-
4	118.7	7.52	d, 8.0	H-5	C-3, 6, 7, 8
5	118.3	6.92	td, 8.0, 0.9	H-4, 6	C-7
6	120.7	7.01	td, 8.0, 0.9	H-5, 7	C-4
7	111.1	7.27	d, 8.0	H-6	C-5, 9
8	135.9	-	-	-	-
9	127.6	-	-	-	-
10	29.7	3.04 <sup>b</sup>	dd, 14.2, 8.8	H-10, 11	C-3, 9, 11
		3.29 <sup>a</sup>	dd, 14.2, 4.3	H-10, 11	C-3, 9, 11
11	55.7	4.30	m	H-10, 10, 12	C-3, 13, 16
12	-	8.31	d, 2.0	H-11	C-11, 13, 14, 16
13	158.6	-	-	-	-
14	128.2	-	-	-	-
15	-	9.81	s	-	C-11, 13, 14, 16
16	166.4	-	-	-	-
17	110.9	5.36	br t, 8.6	H-18	C-14, 16, 18
18	26.4	2.99 <sup>b</sup>	dd, 14.0, 8.5	H-17	C-14, 17, 20
		3.01 <sup>b</sup>	dd, 14.0, 8.7	H-17	
20	13.2	1.58	s		C-18

<sup>[a]</sup> overlapped with water signal<sup>[b]</sup> signals overlapped with each other

**Table S6.** NMR data of guatrypmethine A (**3a**) (DMSO-*d*<sub>6</sub>)

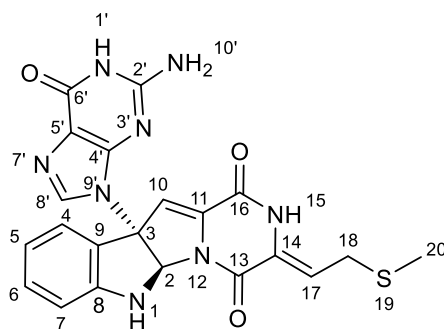
Pos.	$\delta_c$	$\delta_H$	multi, J[Hz]	$^1H$ - $^1H$ COSY	HMBC	NOESY (key correlation) s: strong, m:medium, w: weak <sup>a</sup>
1	-	6.97	d, 1.4	H-2, 6	C-2, 3, 8, 9	-
2	80.7	5.73	d, 1.4	H-1	C-3, 8, 9	H-10b (m), 11 (w), 8' (m)
3	71.3	-	-	-	-	-
4	125.2	7.46	d, 7.3	H-5	C-3, 6, 8	H-8' (w)
5	118.6	6.79	t, 7.3	H-6	C-7, 9	-
6	130.5	7.19	t, 7.7	H-1, 5, 7	C-8,9	-
7	110.2	6.74	d, 7.7	H-4, 6	C-5, 9	-
8	148.8	-	-	-	-	-
9	126.5	-	-	-	-	-
10a	39.8	2.68	dd, 14.0, 9.7	H-10b, 11	C-3, 9, 11, 16	H-11 (w)
10b		3.34 <sup>b</sup>	dd, 14.0, 8.4	H-10a, 11	C-3, 11	H-11 (s)
11	56.9	5.37	dd, 9.7, 8.4	H-10a, 10b	C-10, 16	H-2 (w), 10a (w), 10b (s)
13	159.7	-	-	-	-	-
14	129.6	-	-	-	-	-
15	-	10.24	s	-	C-11, 13	H-18 (m)
16	166.9	-	-	-	-	-
17	114.4	5.90	t, 8.6	H-18	C-2, 13	H-18 (s)
18	27.3	3.39	d, 8.6	H-17	C-14, 17, 20	H-17 (s)
20	14.1	2.02	s	-	-	-
1'	-	10.76	s	-	-	-
2'	153.2	-	-	-	-	-
4'	150.8	-	-	-	-	-
5'	118.1	-	-	-	-	-
6'	156.8	-	-	-	-	-
8'	134.8	7.12	s	-	C-4', 5'	H-2 (s), 4 (w)
10'	-	6.53	br s	-	-	-

<sup>[a]</sup> only correlations for the determination of stereochemistry are listed<sup>[b]</sup> overlapped with water signal

**Table S7.** NMR data of guatrypmethine B (**3b**) (DMSO-*d*<sub>6</sub>)

Pos.	$\delta_C$	$\delta_{15N}$	$\delta_H$	multi, J[Hz]	$^1H$ - $^1H$ COSY	HMBC	NOESY (key correlation) s: strong, m:medium, w: weak <sup>a</sup>
1	-	78.5	6.93	d, 2.0	H-2, 7	C-2, 3, 7, 8, 9	-
2	78.4	-	6.20	d, 1.9	H-1	C-3, 8, 9, 11, 13, 14, N-1, 9'	H-8' (s)
3	71.2	-	-	-	-	-	-
4	124.8	-	7.52	dt, 7.5, 1.2	H-5, 8'	C-3, 6, 7, 8	H-10a (s), 11 (w)
5	118.8	-	6.76	td, 7.5, 1.3	H-4, 6	C-4, 6, 7, 8, 9	-
6	130.7	-	7.17	ddd, 7.9, 7.5, 1.2	H-5, 7	C-4, 5, 7, 8	-
7	110.6	-	6.72	d, 7.9	H-6	C-5, 9	-
8	151.0	-	-	-	-	-	-
9	125.7	-	-	-	-	-	-
10a	38.8	-	3.33 <sup>b</sup>	m	H-10b, 11-	C-2, 3, 9, 11, 16, N-12	H-11 (s)
10b	-	-	3.21	br t, 11.9	H-10a, 11	C-3, 9, 11, N-9'	H-11 (w), 8' (w)
11	57.9	-	4.23	dd, 11.9, 5.6	H-10a, 10b	C-10, 16, N-15	H-4 (w), 10a (s), 10b (m)
12	-	133.5	-	-	-	-	-
13	157.1	-	-	-	-	-	-
14	129.4	-	-	-	-	-	-
15	-	126.6	10.24	s	-	-	H-18 (s)
16	165.1	-	-	-	-	-	-
17	113.3	-	5.82	br t, 8.8	H-18a, 18b	C-14, 13, 16	H-18 (s)
18	27.2	-	3.43	dd, 14.0, 8.9	H-17, 18	C-14, 17	H-17 (w)
			3.31 <sup>b</sup>	m	H-17, 18	-	-
20	14.1	-	2.01	s	-	C-18	-
1'	-	142.8	10.72	s	-	-	-
2'	153.2	-	-	-	-	-	-
3'	-	163.5	-	-	-	-	-
4'	151.0	-	-	-	-	-	-
5'	117.7	-	-	-	-	-	-
6'	156.9	-	-	-	-	-	-
7'	-	242.4	-	-	-	-	-
8'	135.0	-	7.43	s	H-4	C-4', 5', 6', N-7', 9'	H-2 (s), 10b (m)
9'	-	166.7	-	-	-	-	-
10'	-	69.6	6.46	br s	-	-	-

<sup>[a]</sup> only correlations for the determination of stereochemistry are listed<sup>[b]</sup> overlapped with water signal

**Table S8.** NMR data of guatrypmethine C (**4**) (DMSO-*d*<sub>6</sub>)

Pos.	$\delta_C$	$\delta_{15N}$	$\delta_H$	multi, J[Hz]	$^1H$ - $^1H$ COSY	HMBC	NOESY (key correlation) s: strong, m:medium, w: weak <sup>a</sup>
1	-	75.4	7.16	d, 3.6	-	-	H-10 w
2	81.5	-	6.74	d, 3.6	H-10	C-3, 8, 11, N-1, 9'	H-10 m, H-8' s
3	75.0	-	-	-	-	-	-
4	124.1	-	7.26	d, 7.4	H-5, 6	C-3, 6	H-5s, H-10 s, H-8' m
5	118.7	-	6.70	td, 7.4, 0.8	H-4	C-7, 9	H-4 s
6	130.1	-	7.12	td, 7.9, 1.2	H-4,5	C-4, 8	H-7 s
7	110.7	-	6.73	d, 7.9	H-6	C-9	H-6 s
8	150.2	-	-	-	-	-	-
9	126.6	-	-	-	-	-	-
10	114.6	-	6.91	s	H-2	C-2, 3, 11, 13, N-12	H-2 s, H-4 m, H-8'
11	135.6	-	-	-	-	-	-
12	-	155.6	-	-	-	-	-
13	154.6	-	-	-	-	-	-
14	128.9	-	-	-	-	-	-
15	-	130.7	10.65 <sup>b</sup>	br s	-	C-13	-
16	156.7	-	-	-	-	-	-
17	114.2	-	5.92	br t, 8.8	H-18	N-15	-
18	27.0	-	3.42	dd, 14.0, 8.7	H-17	C-14, 17, 20	-
			3.47	dd, 14.0, 8.9			-
20	13.9	-	2.01	s	-	C-18	-
1'	-	143.2	10.66 <sup>b</sup>	br s	-	-	-
2'	153.0	-	-	-	-	-	-
3'	-	163.6	-	-	-	-	-
4'	150.9	-	-	-	-	-	-
5'	117.7	-	-	-	-	-	-
6'	154.5	-	-	-	-	-	-
7'	-	242.9	-	-	-	-	-
8'	1354.8	-	7.75	s	-	C-4', 5', 6', N-7', 9'	H-2 s, H-4 w, H-10 s
9'	-	166.7	-	-	-	-	-
10'	-	69.7	6.30	br s	-	-	-

<sup>[a]</sup> only correlations for the determination of stereochemistry are listed<sup>[b]</sup> signals overlapped with each other

## PUBLICATIONS

**Table S9.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3a I** in units of Å

C	1.021602730590	4.645507524930	2.629219114321
O	1.672007152940	5.564818404126	2.125330171820
C	1.367556744004	3.196597928223	2.379517701072
C	-0.781370917754	3.893660802926	4.142080432458
C	-1.701669342241	4.214087194720	5.077096144090
C	-2.028493973818	5.604989127532	5.507659995631
C	-0.633189995392	2.471663383067	3.702004947956
C	1.715607320635	2.847570225801	0.927720230655
N	0.287136533698	2.269436185944	2.727294534313
N	-0.008032857722	4.867163258371	3.495707134222
O	-1.352735288317	1.561651571622	4.148538405769
H	2.246376604680	2.972134523177	3.012053702177
H	-2.253284378653	3.384218219282	5.515893803121
H	-2.861684256174	5.606389780334	6.219718621131
H	-2.312465424305	6.231125798749	4.645489390026
H	1.183766574371	3.509610081082	0.235967781940
H	-0.121691896809	5.835544283482	3.811356981663
S	-0.570645769025	6.419147396667	6.299928577980
C	-1.231990539407	8.101836087629	6.474194731904
C	0.506817811571	0.967930086719	2.074675177087
C	1.223247023098	1.384701510068	0.733579078860
N	-0.700268683395	0.286306783772	1.663686549797
H	1.136776846476	0.333161079334	2.712984898861
C	0.029292077441	1.674883575253	-1.591977222191
C	-1.158384365368	1.468684400159	-2.307967936283
C	0.093680677024	1.258234301888	-0.267956763605
H	-1.228480337377	1.784582278268	-3.348597679370
C	-2.257593585411	0.865374900850	-1.681043822861
C	-1.010114673522	0.652547299160	0.356058197890
H	-3.181234404529	0.718639188013	-2.243013021791
C	-2.204127929220	0.452613974285	-0.344778446154
H	-3.067637728981	-0.007282071625	0.136605365836
H	0.885880662127	2.155331092770	-2.067988837333
H	-1.434398048385	0.216496900859	2.366228239980
C	4.197857349660	-0.711846309005	0.562288816697
C	5.447381690369	-1.110718661163	1.133050269161
C	3.465034618550	0.301173002562	1.195187754018
N	2.324942301981	0.460522903987	0.431956524350
N	5.724647337964	-0.359913746284	2.304989864311
N	3.767932469064	0.985190716142	2.319573994585
H	6.605564021156	-0.591916006215	2.764142765443
C	4.929953165859	0.621334800440	2.851605990799
N	5.394797675226	1.258268185108	3.953060710996
O	6.249888656546	-1.973258013350	0.748464459159
H	6.136531719459	0.842252088197	4.505801788969
H	4.734948052828	1.845877408197	4.452246130312
N	3.524806532935	-1.172934125580	-0.560184072399
C	2.415545506379	-0.459961341090	-0.604372632277
H	1.632566630636	-0.546885588835	-1.350238815268
H	-2.111632064181	8.106493588269	7.129950242221
H	-0.437302934391	8.705214268853	6.929714132707
H	-1.488709668493	8.520332687020	5.492895024324
H	2.789034718183	2.941009047655	0.742543378501

**Table S10.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3a I'** in units of Å

C	-0.044645350971	4.661752155203	2.525291216209
O	0.287835386340	5.674763671988	1.905549784341
C	0.665752621070	3.333544438756	2.355804644116
C	-1.530905876029	3.532830431686	4.135157206509
C	-2.347070478369	3.654543596129	5.203964823173
C	-2.722622486591	4.962410935065	5.815526153749
C	-1.161393801675	2.182245227575	3.612471104942
C	1.203510992464	3.062322053554	0.947690221135
N	-0.218846447845	2.195594427486	2.637248304844
N	-1.044373058278	4.657578121481	3.451254684233
O	-1.701599316215	1.138178975682	4.015081988530
H	1.509875430218	3.320992301492	3.069113450867
H	-2.745997806752	2.733858390700	5.627232034839
H	-3.272896439685	4.819548418683	6.752102377337
H	-1.826680947853	5.570314645414	6.022062896303
H	0.632757403574	3.629821070069	0.205003985944
H	-1.568565600470	5.533110256828	3.581977678727
S	-3.768115294511	5.951030635112	4.653139525595
C	-3.749559403228	7.550430351402	5.514904040578
C	0.278235560699	0.972765099866	1.988638380756
C	1.003564607494	1.537975276609	0.711575068622
N	-0.755228704387	0.095658138319	1.482904324191
H	0.961397684627	0.444457299670	2.667486196658
C	-0.035484366731	1.642117036414	-1.702535541896
C	-1.104340413696	1.233393891012	-2.512696215588
C	-0.001476108185	1.224261593470	-0.377759852124
H	-1.148198370088	1.547016418078	-3.555445440560
C	-2.120516690895	0.429573729158	-1.977195594801
C	-1.022224408023	0.417711998867	0.154937298725
H	-2.953468092058	0.124593084618	-2.612515978421
C	-2.099371868714	0.013255884796	-0.641145894867
H	-2.899498151116	-0.604096301210	-0.232007345324
H	0.753324227218	2.278751462395	-2.106981411096
H	-1.512554554086	-0.125900903855	2.125869710360
C	4.316237697579	0.026706284565	0.789870192304
C	5.561164250010	-0.151642238998	1.472142451770
C	3.354345787860	0.870123367058	1.361600515393
N	2.277629003365	0.837582797488	0.497889585943
N	5.587385241995	0.605218966619	2.672075740355
N	3.421166692459	1.566377473713	2.517352319199
H	6.450416432684	0.526766486790	3.209812288480
C	4.577767889251	1.407290007161	3.152194114915
N	4.814000751951	2.087253371038	4.300305710518
O	6.540909790469	-0.841913758722	1.156501166417
H	5.564461135399	1.795920514673	4.917719556256
H	4.010903557346	2.521972174169	4.742932533166
N	3.843382137884	-0.520023752618	-0.394195002081
C	2.629207435379	-0.022337365446	-0.534657350381
H	1.946162952845	-0.232241114666	-1.350780909734
H	-2.722448694199	7.923392687878	5.614115440079
H	-4.331018528329	8.245454036285	4.896950852096
H	-4.219554654057	7.463414561286	6.502405462096
H	2.255853788202	3.347030531530	0.864258146351



**Table S11.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3a II** in units of Å

C	-2.383912052449	-1.950612047608	0.356731882863
O	-2.476568522783	-3.098506579995	0.797984395152
C	-1.046863990841	-1.344903719037	-0.010911888879
C	-3.469000352231	0.134513930346	-0.413132313742
C	-4.609266731073	0.781024883459	-0.736689942570
C	-5.983748113928	0.222182873357	-0.576338033441
C	-2.159138258488	0.855256846637	-0.481636054910
C	0.079033569365	-1.551229854337	1.000758883261
N	-1.077618973495	0.110541465410	-0.150393613528
N	-3.472609288151	-1.173229342984	0.094539679869
O	-2.075327615592	2.064188640880	-0.747341126986
H	-0.773044313741	-1.805779134448	-0.980037561632
H	-4.496434647856	1.797714616051	-1.110451611865
H	-6.165727269678	-0.087310044418	0.466354257351
H	-6.740296723960	0.966988696268	-0.848267584659
H	-0.286648251267	-1.309806085623	2.006713101751
H	-4.380442651490	-1.645253977979	0.155195627182
S	-6.225490172174	-1.277943175146	-1.627445761231
C	-7.856819776906	-1.799515898364	-1.022624741195
C	0.243461122290	0.712407766066	0.062544332202
C	1.120771753577	-0.480276677934	0.603920746657
N	0.261499917441	1.704096122646	1.127725285507
H	0.626973713205	1.121874003328	-0.880175464619
C	2.841158481414	-0.382161019304	2.581587150457
C	3.326737885807	0.396303778710	3.642741795162
C	1.828592510705	0.134185946287	1.787254586333
H	4.128420134533	0.016211202145	4.276080611071
C	2.785110667795	1.666501132457	3.882206163401
C	1.283742936628	1.404856278659	2.029676229877
H	3.172586817472	2.268227471023	4.706018123129
C	1.757685770358	2.187952240160	3.086504924656
H	1.348307260100	3.180469626985	3.278051292854
H	3.264913912011	-1.366290814438	2.373302000981
H	0.132374572563	2.671909880330	0.845676462968
C	3.716605784356	-1.033320524670	-1.883626754097
C	4.786572611409	-0.493736502350	-2.665709150663
C	2.994416838314	-0.170171898965	-1.048810003132
N	2.050293017833	-0.965222800038	-0.423968662569
N	4.920652833373	0.899162418398	-2.436107863163
N	3.154798471329	1.156331231727	-0.867722616882
H	5.666912729529	1.351475909965	-2.964108798764
C	4.143524043869	1.662817079826	-1.594347428256
N	4.388799000556	2.995675639862	-1.555483539818
O	5.550517939317	-1.066352387689	-3.456715063178
H	5.272077253640	3.360582459246	-1.895576057931
H	3.947436054266	3.511403478655	-0.800979548965
N	3.234906693852	-2.329588568187	-1.779175536729
C	2.250482041155	-2.255263877280	-0.904227644816
H	1.648218667058	-3.094855103516	-0.575941176963
H	-8.609662860294	-1.026893846901	-1.223597857030
H	-8.118828339519	-2.711724340444	-1.572407231503
H	-7.818294477238	-2.021376931410	0.051345094784
H	0.475401346772	-2.569166735994	1.024370946740

## PUBLICATIONS

**Table S12.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3a II'** in units of Å

C	-2.535235864329	-1.805933913502	0.470148976820
O	-2.661690578634	-2.877271336089	1.067136501370
C	-1.182565754774	-1.273722589702	0.036206968848
C	-3.524281208096	0.235013791779	-0.500452052687
C	-4.612588080325	0.821326811072	-1.042309372843
C	-5.945976632543	0.160584085647	-1.147572956665
C	-2.213857149282	0.954855619759	-0.454839508582
C	-0.035077641098	-1.538786134908	1.007254227070
N	-1.163268771920	0.186758240154	-0.080895326055
N	-3.594307042953	-1.033272004408	0.099432357461
O	-2.102438593750	2.166250093792	-0.698613078287
H	-0.969149670941	-1.733124807403	-0.948301177436
H	-4.491879611807	1.839819043085	-1.408954756452
H	-5.855356102882	-0.840975336956	-1.598054992232
H	-6.637880912795	0.756444155734	-1.753070709734
H	-0.367608782542	-1.324721078680	2.030730345666
H	-4.528686662364	-1.349526394521	0.391013122784
S	-6.680174547154	-0.082303207417	0.532625447331
C	-8.084774767525	-1.149308472830	0.099274239299
C	0.169678351220	0.749406098323	0.148147129103
C	1.026543780275	-0.482988515539	0.622564206192
N	0.222102464874	1.691496158342	1.258618743868
H	0.550923661045	1.196576267494	-0.778383918563
C	2.798057560703	-0.500552786868	2.557691146342
C	3.323027257895	0.222862564491	3.639144527600
C	1.777025949950	0.067770375586	1.811015316868
H	4.131830089225	-0.198657923037	4.236122495209
C	2.811684843259	1.491244935526	3.945616060245
C	1.261512898911	1.336279325819	2.121328835095
H	3.230395336093	2.049988074151	4.784227673557
C	1.775577136727	2.064966954778	3.198192905494
H	1.390252076692	3.055746147801	3.441681746088
H	3.197614139045	-1.482318059300	2.297384428345
H	0.111991897915	2.673320809753	1.018768361993
C	3.536786153333	-1.014749160243	-1.956221379365
C	4.597471470736	-0.476608214865	-2.751321837887
C	2.866822826559	-0.163696446929	-1.066725150727
N	1.916438558758	-0.954388737256	-0.445358530859
N	4.782974608291	0.901087993863	-2.472292564232
N	3.071563287525	1.149535389326	-0.841542276294
H	5.526364770933	1.351100887341	-3.006076729818
C	4.054471889738	1.654913283143	-1.578648215683
N	4.350274447028	2.970617403755	-1.486039752309
O	5.317648188538	-1.039908860796	-3.589107310262
H	5.186141768294	3.354391937408	-1.910945667163
H	3.919119345747	3.494622648951	-0.732850941549
N	3.017670802257	-2.299166162851	-1.887730047404
C	2.061700407121	-2.229836248994	-0.981600507804
H	1.442418015349	-3.062707035295	-0.668165383569
H	-7.734533698714	-2.078311095755	-0.367641258150
H	-8.599897857266	-1.385731161992	1.038360885049
H	-8.776943271269	-0.624384004417	-0.570962767365
H	0.335319603916	-2.566563628438	0.987537908289

**Table S13.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3a III** in units of Å

C	0.119086757212	0.924839183333	2.381300837898
O	0.540916908683	2.083336099410	2.381411806367
C	-0.015713923701	0.118317248824	1.108884310564
C	-0.690474723789	-1.051368770926	3.631830054452
C	-0.806321561543	-1.688351935298	4.816010974609
C	-0.388860403276	-1.115146826967	6.127393807346
C	-1.205759281383	-1.693343637829	2.382933529631
C	-0.585810883001	0.850963213566	-0.100331911158
N	-0.919033500357	-1.028260896947	1.238241816084
N	-0.180136454733	0.251104272719	3.529399754414
O	-1.895564437199	-2.724021922285	2.394294832873
H	1.006431896209	-0.239641573824	0.877636244632
H	-1.233565011059	-2.689466568548	4.788334655229
H	-0.758645760636	-0.085854971028	6.264262581742
H	-0.765432527648	-1.725085790058	6.956765567841
H	-1.489989801176	1.403349546986	0.186004718210
H	0.043778331091	0.742401844754	4.396246487697
S	1.454919176969	-1.061642302707	6.229352107737
C	1.640821869805	-0.288967587523	7.862849397110
C	-1.533414303192	-1.391498622495	-0.037235445999
C	-0.993702288771	-0.315393043014	-1.042307605391
N	-2.982626748039	-1.209191868985	-0.064182146371
H	-1.272424496743	-2.426143259768	-0.290924109274
C	-2.336600521285	0.704703231076	-3.072945958883
C	-3.612949010208	0.856392085298	-3.635035958396
C	-2.207804626287	-0.017463129681	-1.892566568706
H	-3.733209186219	1.409541857443	-4.566533218125
C	-4.730944514725	0.295734997231	-3.003459191599
C	-3.333857568810	-0.571940284746	-1.257278941158
H	-5.718617479631	0.415295029408	-3.451397075247
C	-4.610614534942	-0.421615416635	-1.807103961442
H	-5.484788743258	-0.858034275763	-1.322690842538
H	-1.458655702816	1.141103563709	-3.549203821367
H	-3.531778003973	-2.015611291277	0.224445942197
C	1.949954290871	-0.876633892962	-3.133085958379
C	2.954045853070	-0.352984641176	-4.008541622321
C	1.003160755293	0.000041995506	-2.586428238559
N	0.181430689990	-0.790461198622	-1.796706697437
N	2.793552558363	1.043937804699	-4.177721012313
N	0.894340904562	1.332136936897	-2.780419405647
H	3.473602490346	1.484275956912	-4.797377897532
C	1.818022713750	1.822785795834	-3.598618477547
N	1.792679944328	3.136147460435	-3.928220697003
O	3.874611221019	-0.945832433275	-4.588984330221
H	2.605951127181	3.573099873108	-4.348085256624
H	1.191358595694	3.734511586681	-3.371374669399
N	1.722815540283	-2.173977907289	-2.705057241576
C	0.667299855969	-2.089797607626	-1.920820733767
H	0.207350003317	-2.930216689333	-1.413857382577
H	1.186549152391	0.709946536222	7.874537890644
H	1.193181235601	-0.917003417677	8.643692809815
H	2.718691646858	-0.198892570813	8.045810859295
H	0.117599361676	1.537620791354	-0.576401217205

## PUBLICATIONS

**Table S14.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3a III'** in units of Å

C	0.090746394470	1.052487557466	2.386434566310
O	0.374767502950	2.251954921359	2.360424737215
C	-0.000931103245	0.210897655220	1.128907245227
C	-0.525028832750	-0.973760614566	3.654522053323
C	-0.461086890515	-1.641883992045	4.825534842148
C	0.149276530939	-1.082956673279	6.067092795089
C	-1.119172400010	-1.614331309638	2.441500671465
C	-0.577242556113	0.905980507411	-0.098490490751
N	-0.892409788611	-0.944156280214	1.286114402135
N	-0.104899245286	0.362266388957	3.545601240865
O	-1.796682360575	-2.652335619181	2.487610515181
H	1.027727824953	-0.138710914484	0.916368071597
H	-0.888843202746	-2.643166358608	4.845967169498
H	1.157796094896	-0.688022892132	5.864689450736
H	0.221471526840	-1.843864534360	6.852056593526
H	-1.478689342734	1.468362284675	0.176943356879
H	-0.064851275127	0.897896536458	4.423723095583
S	-0.849339475511	0.339553402916	6.701488669012
C	0.314237251109	1.003978963864	7.928550536190
C	-1.526610814179	-1.330870155356	0.029584643667
C	-0.993005314708	-0.283540874835	-1.007245199306
N	-2.975614577978	-1.138503938139	0.015385140584
H	-1.277423957013	-2.373896976129	-0.201785133032
C	-2.350045983830	0.679198665710	-3.056754495049
C	-3.630847954396	0.818621217758	-3.611971588078
C	-2.213404988490	-0.006062974246	-1.855538470229
H	-3.757233103059	1.342556997858	-4.559423028563
C	-4.745312269713	0.282913127740	-2.953164841743
C	-3.335901828148	-0.535983272772	-1.193311815379
H	-5.736470814032	0.392348935497	-3.395973664323
C	-4.616942092521	-0.397184414321	-1.736015205288
H	-5.488379046848	-0.814867916993	-1.230599210708
H	-1.474873911268	1.096786530178	-3.554487883744
H	-3.526024905341	-1.933846511127	0.331565826891
C	1.932046615259	-0.905160785393	-3.106643833805
C	2.930362620543	-0.406709599511	-4.003054784513
C	0.992977977411	-0.011891131379	-2.573525875917
N	0.175040213237	-0.780020235160	-1.758307754712
N	2.773156098647	0.986219621926	-4.205057737485
N	0.886965195023	1.315446130827	-2.799272350579
H	3.449272587346	1.408691988198	-4.841261608981
C	1.805200791341	1.782526170650	-3.637110867681
N	1.781399012526	3.087448294230	-3.998914602261
O	3.843921766384	-1.016835473298	-4.576754805867
H	2.592397267968	3.510839021158	-4.436702987984
H	1.187106222573	3.701633026310	-3.451835032922
N	1.703628663339	-2.190992559251	-2.645584730387
C	0.654822777206	-2.083571726164	-1.855138126473
H	0.195336817206	-2.909389013660	-1.324464205460
H	1.255964290243	1.298044846091	7.448112649699
H	-0.163491671209	1.889719240558	8.364589938188
H	0.504617741283	0.267983883939	8.719429832465
H	0.123092856534	1.579013697242	-0.597838277522

## PUBLICATIONS

**Table S15.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3b I** in units of Å

C	-2.909753410406	-1.268663120733	0.196538663407
O	-2.990953817728	-1.900970099216	1.252259118148
C	-1.576971427609	-0.977502723233	-0.466350045100
C	-4.000157967709	-0.082119635830	-1.672795203699
C	-5.077322594910	0.605362585324	-2.108090079766
C	-6.308390355030	0.825381991583	-1.294696937433
C	-2.789750081156	-0.226761923469	-2.536095094720
C	-0.569076176387	-2.124021775781	-0.405615641967
N	-1.695147343774	-0.717956374734	-1.907682045856
N	-3.995956337394	-0.735898097610	-0.429838852251
O	-2.786859921401	0.055997912205	-3.745130009733
H	-1.165037298162	-0.084376635006	0.038807647481
H	-5.026331159865	1.012234965048	-3.117015612634
H	-6.991643058228	1.524063047980	-1.790224424306
H	-6.054890295817	1.226071623083	-0.299657438921
H	-0.051830683703	-2.181568383414	0.557106097055
H	-4.916923573130	-0.932526017343	-0.015210866641
S	-7.193814802866	-0.772735702861	-1.007295476436
C	-8.403716219166	-0.227744939002	0.233329304177
C	-0.492923574968	-1.101506695987	-2.660816727075
C	0.395780507079	-1.779202143761	-1.556777711573
N	0.308753747118	0.008503047583	-3.149958087033
H	-0.786678750774	-1.771133656534	-3.475791172740
C	2.242504699418	-0.539406266390	-0.149077438582
C	3.023124645863	0.620799865743	-0.044817836296
C	1.361913973536	-0.660565646842	-1.216558841808
H	3.721234598297	0.735125710498	0.784117278884
C	2.897246808844	1.635808587426	-1.002720368695
C	1.232581588060	0.364783812133	-2.171338582842
H	3.502888974384	2.538640119781	-0.909823096254
C	2.002705711237	1.528082466692	-2.074046588600
H	1.903147388424	2.330412640042	-2.805579301995
H	2.327315778362	-1.330980335351	0.597604047796
H	-0.171243885069	0.725904098420	-3.687512545266
C	1.511148561886	-4.929444229639	-3.009921054363
C	1.182864676418	-6.179726918389	-3.623168502407
C	0.467598527692	-4.078309197513	-2.621460301620
N	1.084805221704	-2.964685014579	-2.082625715996
N	-0.223504768709	-6.333053115812	-3.726769189059
N	-0.863803329089	-4.263622402769	-2.752092166500
H	-0.529798407838	-7.208107483479	-4.151887049985
C	-1.173452015478	-5.426786090671	-3.314166084317
N	-2.474581292877	-5.777387021179	-3.457498774684
O	1.933899301850	-7.072252967241	-4.042696187792
H	-2.733923230788	-6.522032487508	-4.095632116390
H	-3.159113049236	-5.040837331746	-3.319960559318
N	2.741700321270	-4.356354481948	-2.724198392314
C	2.452243281834	-3.190600106280	-2.178463282444
H	3.170441919204	-2.456162611962	-1.830049473258
H	-9.071477792868	0.534568041137	-0.187201664230
H	-8.991560439288	-1.112202675257	0.507554802017
H	-7.894828811250	0.162408027330	1.123490698033
H	-1.072018034489	-3.076935447330	-0.610570806018

## PUBLICATIONS

**Table S16.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3b I'** in units of Å

C	-2.590303107215	-1.169991386774	0.527224579659
O	-2.508825469223	-1.570412661907	1.691232559279
C	-1.354156758951	-0.919476442882	-0.307131294964
C	-3.982397313411	-0.387547154681	-1.361791310250
C	-5.202308678304	-0.023643110308	-1.812045148712
C	-6.462150207916	-0.104950273849	-1.016228041317
C	-2.813629490272	-0.387124828902	-2.292147759082
C	-0.334246331911	-2.059648310667	-0.316220141762
N	-1.620094679601	-0.705827542226	-1.733825967258
N	-3.782167416438	-0.862419588792	-0.058056125948
O	-2.928689796007	-0.166243985582	-3.508063756368
H	-0.890996541391	-0.009438595917	0.117794643957
H	-5.251830095994	0.328537123240	-2.841113855897
H	-7.327576904635	0.174540151320	-1.627492529916
H	-6.626004852384	-1.126487090121	-0.634882378319
H	0.265078319834	-2.098216934076	0.598593392289
H	-4.594135780780	-0.857247590281	0.567410183529
S	-6.393409046084	1.009847169131	0.456275960224
C	-7.933014342081	0.507072472830	1.279326941901
C	-0.466285252637	-1.048005484377	-2.576562670687
C	0.522287525263	-1.718944047778	-1.552612229356
N	0.271668095323	0.083737214261	-3.113373756597
H	-0.807517904301	-1.715308086776	-3.374680984726
C	2.492338920280	-0.483236943091	-0.322440244608
C	3.272356371968	0.681384315918	-0.281346075152
C	1.512967994979	-0.599573936564	-1.300824143213
H	4.047606735700	0.791847976396	0.476415933019
C	3.046630646345	1.706061322903	-1.210531216603
C	1.283586411289	0.435518310577	-2.226022680232
H	3.652494912869	2.612431410605	-1.166752652673
C	2.052423194389	1.603573792346	-2.190667453916
H	1.876790657183	2.413258788609	-2.899389119435
H	2.653986631031	-1.282286069327	0.403233053154
H	-0.255369531306	0.801008711924	-3.604315266428
C	1.494213254760	-4.875022911088	-3.091374503662
C	1.107711627321	-6.125342838040	-3.669783874442
C	0.493055578233	-4.020487555295	-2.610318670438
N	1.159560033171	-2.907181450222	-2.133550019615
N	-0.302836709374	-6.274428097862	-3.643842981266
N	-0.845283443025	-4.202116469791	-2.617589978598
H	-0.649310724103	-7.149225803980	-4.037485686344
C	-1.208252880999	-5.365105767331	-3.146904602012
N	-2.518045541395	-5.712493480665	-3.168962879377
O	1.814478633793	-7.020719019680	-4.154837287232
H	-2.837503057273	-6.454458367889	-3.782603061609
H	-3.184622266070	-4.973884191045	-2.968904091867
N	2.747024241278	-4.304276741263	-2.921472765558
C	2.511820707705	-3.136329634995	-2.354521156449
H	3.260731009279	-2.402627749970	-2.076195804954
H	-7.903694318789	-0.558731560227	1.538959869090
H	-8.005286640195	1.102394486074	2.197665748950
H	-8.800295053626	0.718145846648	0.641120499650
H	-0.845569230490	-3.018467012929	-0.466161388119

## PUBLICATIONS

**Table S17.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3b II** in units of Å

C	-0.655841194697	2.405742344380	-1.114556048194
O	-1.441081689334	2.284552467443	-2.057446382967
C	-0.454550630722	1.309482266189	-0.086506097174
C	1.024115253585	3.735987774243	0.111003849906
C	1.506758650306	4.961340970418	0.407082976394
C	0.997508830059	6.223971608098	-0.202724332689
C	1.564373299551	2.520474745432	0.790109994266
C	-0.451387346713	-0.111260760169	-0.645865864880
N	0.840614104691	1.391996938127	0.596548466187
N	0.049944939409	3.548026809454	-0.882953967974
O	2.609501604891	2.528555409027	1.460384334385
H	-1.272087130372	1.419333229373	0.650560651348
H	2.318720638062	5.009251905944	1.131282579020
H	1.428845963023	7.103566819280	0.287850203303
H	-0.100522462594	6.281394806305	-0.128324916663
H	-1.454281920559	-0.492410818089	-0.850329836025
H	-0.049520571006	4.304842357657	-1.572859170040
S	1.406232610474	6.285298737375	-2.005686286217
C	0.408917856171	7.723392422081	-2.492678361026
C	1.344565061473	0.079448132114	1.031235788563
C	0.255563093760	-0.898617773298	0.471444746690
N	1.335634078832	-0.135119150365	2.467773147104
H	2.349357704150	-0.057553709192	0.614047680464
C	-1.838240814274	-1.744883289921	1.831844939913
C	-2.408745785109	-1.857975606633	3.108554380637
C	-0.596078863712	-1.136224808770	1.709784127184
H	-3.385131399294	-2.327268671003	3.228657877848
C	-1.726913187268	-1.361929928230	4.228041709903
C	0.088635294753	-0.645107109373	2.835964262750
H	-2.182602459260	-1.448495261505	5.215643315257
C	-0.471921429556	-0.750656729374	4.112130397192
H	0.048791041959	-0.365698008164	4.989449389480
H	-2.354167226829	-2.126442592358	0.949612907292
H	1.780609114939	0.581184751989	3.037049251948
C	1.094416885040	-4.197566630263	-0.833029565934
C	0.708921668060	-5.367658184998	-1.559845870679
C	0.192351786554	-3.126458579555	-0.749362425324
N	0.844568741524	-2.158206202724	-0.008772033671
N	-0.600395590079	-5.223495734121	-2.082536385584
N	-1.052595583710	-3.038942410015	-1.264966346681
H	-0.945488023554	-6.027718086560	-2.606267887390
C	-1.419802530216	-4.128090383022	-1.929164387871
N	-2.669862012002	-4.209021063409	-2.449379120557
O	1.344469005218	-6.412736874679	-1.763246581853
H	-2.876329374127	-4.899372626485	-3.163586935953
H	-3.192564846154	-3.340543395261	-2.500654141690
N	2.262181769122	-3.911144766315	-0.142662699052
C	2.075850881931	-2.700027612912	0.343626406116
H	2.780840757740	-2.156511920051	0.964291373995
H	0.751385051897	8.625840216262	-1.971092736467
H	0.552282672889	7.852619744707	-3.572402719125
H	-0.653721786897	7.544775505110	-2.285443474928
H	0.142248051635	-0.156499239460	-1.568660175028

## PUBLICATIONS

**Table S18.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3b II'** in units of Å

C	-0.944724119153	2.430812047759	-0.828070428971
O	-1.884712202470	2.363839671464	-1.623852827630
C	-0.620570610283	1.291271864147	0.112014938292
C	0.893590869896	3.736171007157	0.192226853097
C	1.517463710258	4.926273796446	0.325594193930
C	1.171291728736	6.161037870234	-0.437583779340
C	1.418295396369	2.535274499560	0.909631608174
C	-0.560275408392	-0.096337723617	-0.525219643482
N	0.685508093676	1.404506582457	0.764327600260
N	-0.186039389903	3.554493170225	-0.684239200717
O	2.483268562935	2.546921071673	1.547418555745
H	-1.417236634404	1.308416566710	0.880427040775
H	2.350125513879	4.958645563766	1.026599952440
H	1.868797260208	6.972590994480	-0.201088592895
H	1.210483251977	5.979521628158	-1.524658494096
H	-1.547330098476	-0.510380317479	-0.742126826046
H	-0.525658638344	4.384289383791	-1.180760239043
S	-0.545655939661	6.724822205338	-0.051536175397
C	-0.698901188840	8.059275377513	-1.273969530265
C	1.235535852346	0.096440283460	1.154035840741
C	0.190570575934	-0.904901086366	0.548291808581
N	1.228868451847	-0.176295278794	2.580614893097
H	2.246902950831	0.017832246687	0.738205474461
C	-1.851483300302	-1.926797605582	1.863675570578
C	-2.415706730812	-2.128233767386	3.132227378876
C	-0.644710898492	-1.245731515695	1.772727384947
H	-3.364642830986	-2.655700228674	3.228496600498
C	-1.762844872819	-1.645527282153	4.274649879835
C	0.011568629596	-0.768572977658	2.921570176506
H	-2.213494392986	-1.801070636997	5.256064180749
C	-0.543560979850	-0.961264728844	4.190095637069
H	-0.045575422321	-0.587878717187	5.085411529545
H	-2.345081608013	-2.295535414244	0.963321026965
H	1.638945269694	0.531322877533	3.185410914800
C	1.180061904994	-4.085456916326	-0.931700929469
C	0.848220255046	-5.230619438871	-1.721993722723
C	0.229976849885	-3.062831735913	-0.792262295724
N	0.837216627999	-2.107594828071	0.001797207648
N	-0.466319061772	-5.118360052440	-2.239111953571
N	-1.017842857443	-3.004589768207	-1.304973977658
H	-0.774458471146	-5.908120244669	-2.806067002332
C	-1.335238788719	-4.071943470379	-2.027761433752
N	-2.580621502521	-4.182162423423	-2.553171411707
O	1.530992714015	-6.232777551447	-1.980532665910
H	-2.757246342322	-4.843243245484	-3.302160581859
H	-3.144064581801	-3.338065184119	-2.556834016784
N	2.333133367825	-3.784491991550	-0.223083966960
C	2.091602213172	-2.611504350274	0.328080326855
H	2.771008792623	-2.071398998404	0.979548109719
H	-0.585486439138	7.666125653960	-2.292129354297
H	-1.705789685638	8.478275975899	-1.156814730396
H	0.044072717687	8.844641366534	-1.085722158874
H	0.028781728319	-0.065867094358	-1.451582068818



**Table S19.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3b III** in units of Å

C	1.424709299799	-0.305254885357	2.919602684910
O	2.604106468208	-0.543177262843	3.191803222506
C	1.023387269735	0.343159391319	1.610680468171
C	-0.966834113010	-0.397759439486	3.523641533346
C	-1.885173025894	-0.489678956416	4.508723598815
C	-1.545418144033	-0.673324179109	5.950038962239
C	-1.390978044376	-0.149600731380	2.111991147987
C	1.754259210084	-0.211328230872	0.386652194242
N	-0.378728045402	0.129935785127	1.255377030500
N	0.402663658435	-0.565878580526	3.780025963114
O	-2.575070036296	-0.211980531792	1.748779913052
H	1.223725944384	1.426257286130	1.716885839327
H	-2.930942106673	-0.419082509732	4.213344780207
H	-0.809894762124	0.079104034608	6.277907914534
H	-2.437133038247	-0.588751626809	6.580845665012
H	2.775993171341	0.169644716714	0.295517388963
H	0.645639567521	-1.044945426974	4.657960076328
S	-0.765622862121	-2.325588136355	6.238031487763
C	-0.172188847466	-2.086619628365	7.938452786629
C	-0.601887721751	0.347507185681	-0.173817443368
C	0.851267770948	0.225703225006	-0.784590241633
N	-1.058063656892	1.705151344414	-0.457665024044
H	-1.309108979265	-0.398082756202	-0.548532822359
C	2.243677493788	2.154319443456	-1.886432102614
C	2.212801112581	3.473074297367	-2.358747557550
C	1.104106115178	1.633083276968	-1.285847580382
H	3.093861546772	3.903340634875	-2.834911099481
C	1.042320998867	4.235542765657	-2.227680814807
C	-0.057772920868	2.404207582188	-1.135419602084
H	1.023878372681	5.258264319830	-2.607454392743
C	-0.103861317030	3.718134921393	-1.615063833829
H	-1.006757213214	4.320965806339	-1.513144073417
H	3.139795994689	1.542622812841	-2.007890960347
H	-2.008589029807	1.787025431417	-0.808662006680
C	0.629954044764	-1.703548667614	-3.865710724280
C	-0.038672548529	-1.965964410953	-5.102916966947
C	0.067752171888	-0.772063356708	-2.981447820505
N	0.922420684552	-0.735062161602	-1.895289069233
N	-1.218105023645	-1.186500595087	-5.209317265834
N	-1.063006468737	-0.051267464482	-3.128600075213
H	-1.748560278035	-1.324458067004	-6.069388344828
C	-1.688820331248	-0.293748127165	-4.273480488075
N	-2.869787626765	0.319249914351	-4.536338004457
O	0.278944912218	-2.735847959570	-6.021654736599
H	-3.226796137034	0.348170302984	-5.485500793581
H	-3.111385071452	1.104900473730	-3.941148968164
N	1.811500773493	-2.215790662406	-3.350989184819
C	1.959663740871	-1.613858346888	-2.186609622227
H	2.790328387220	-1.763805620574	-1.505116507467
H	-1.013615816037	-1.912463961359	8.620430744169
H	0.340619287606	-3.013022758015	8.224358242375
H	0.535201164673	-1.249157102282	7.986568153910
H	1.781003108568	-1.306371614149	0.459147053723

## PUBLICATIONS

**Table S20.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3b III'** in units of Å

C	1.475836257138	-0.116893640313	2.928892645522
O	2.662317225040	-0.159133362095	3.265058626113
C	1.061524148011	0.449340086940	1.590653696974
C	-0.906819084014	-0.495462615035	3.476053702865
C	-1.822967970394	-0.885295429132	4.388214317905
C	-1.499806674158	-1.420897397945	5.743326908554
C	-1.346234208227	-0.095302144908	2.104034375370
C	1.750687898666	-0.187621571669	0.379728232737
N	-0.351247288767	0.281109093705	1.263981748908
N	0.467704479560	-0.522304654731	3.749981538503
O	-2.531484839025	-0.152866289476	1.743814568840
H	1.301429670155	1.528395527953	1.632116470134
H	-2.866257347751	-0.818728624865	4.083574902731
H	-2.410073279156	-1.738498036147	6.264363628138
H	-0.821039457243	-2.287035737520	5.675837589589
H	2.790825342867	0.133092766634	0.266559476371
H	0.741800535354	-0.744872568038	4.713601074354
S	-0.639650598714	-0.150831740539	6.774245709362
C	-0.196033293842	-1.179635610386	8.203916411002
C	-0.576442096147	0.500808527649	-0.163899894414
C	0.857092446095	0.265866725142	-0.794303237121
N	-0.934925513370	1.883929263563	-0.462477245528
H	-1.340878214786	-0.196990951201	-0.517025222654
C	2.346745579413	2.058220410529	-1.993548505322
C	2.395251992805	3.362868425585	-2.502234534836
C	1.191910897989	1.635958598593	-1.346196382637
H	3.289562490843	3.716002792172	-3.015582393660
C	1.286422531463	4.210700259239	-2.359327826863
C	0.092338334264	2.491976095191	-1.185663240883
H	1.328880840491	5.221845145613	-2.767223693596
C	0.125371768940	3.793022566933	-1.700286961929
H	-0.729313788523	4.461202148328	-1.590649617017
H	3.193047704518	1.380770453373	-2.122828015408
H	-1.887949428416	2.034240605904	-0.783236246390
C	0.436142215846	-1.761122363832	-3.790050686336
C	-0.277477789462	-2.022804625169	-5.002135128124
C	-0.031139181212	-0.748867178921	-2.940043928897
N	0.843140365127	-0.736107217876	-1.869209341115
N	-1.389611047248	-1.152125071655	-5.125991386341
N	-1.101346434657	0.056424730699	-3.101399824854
H	-1.946440871444	-1.283495144089	-5.970184523146
C	-1.766706268622	-0.182591239198	-4.224716275402
N	-2.896973917429	0.515621728171	-4.497148629676
O	-0.045110974033	-2.858790439805	-5.887984833382
H	-3.269779379309	0.531275547920	-5.440552208170
H	-3.060232782399	1.344571342807	-3.935103626104
N	1.580972108156	-2.346607337149	-3.270351414624
C	1.799298946276	-1.709615854111	-2.135979451198
H	2.627653104212	-1.898220790569	-1.461300752961
H	0.452476091861	-2.009937828529	7.897099235653
H	0.349968145645	-0.530798949722	8.899532441022
H	-1.096746021659	-1.564465233514	8.698544005505
H	1.714319567799	-1.279368704570	0.488459022937

**Table S21.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3b IV** in units of Å

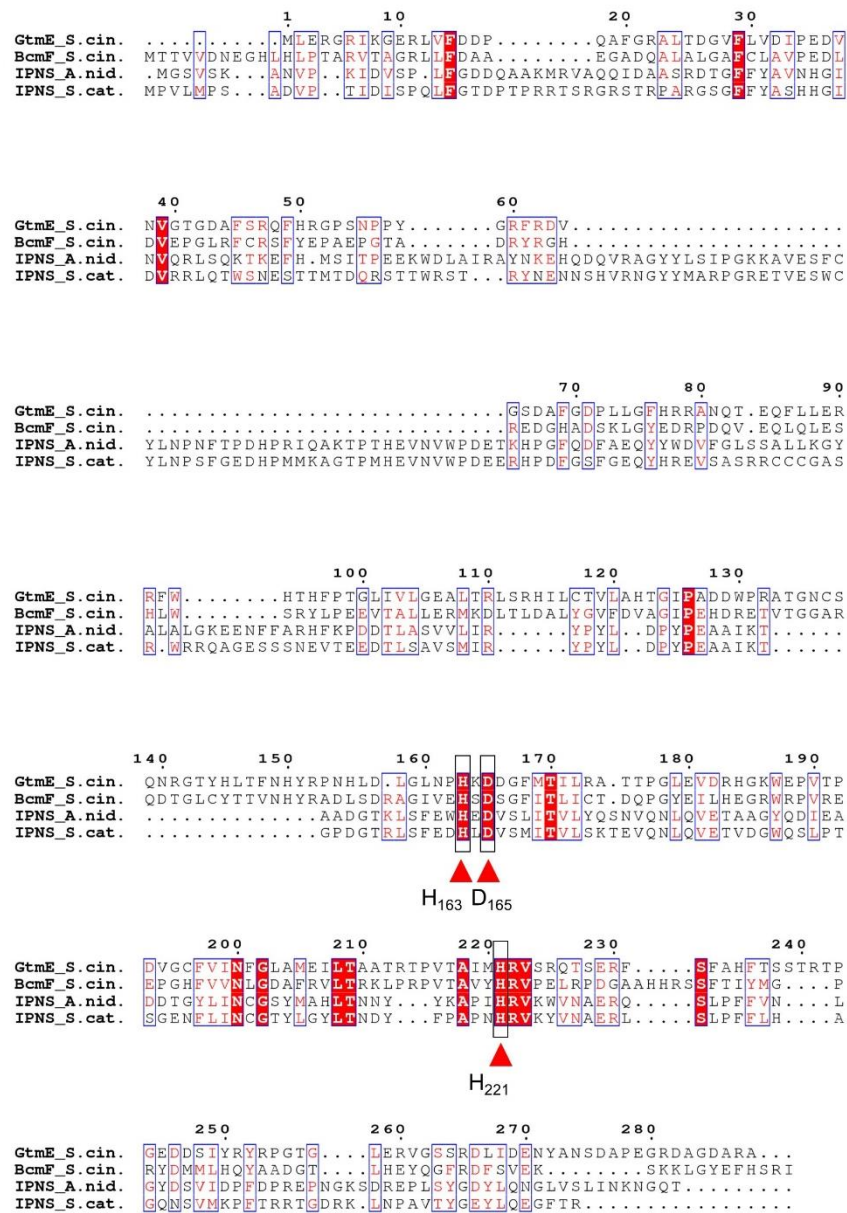
C	1.222751871946	3.995024514468	2.922410012551
O	1.984500929058	4.753044311337	2.320188141825
C	0.974656670402	2.577229317602	2.452573409061
C	-0.300148822331	3.590464581112	4.819759227024
C	-1.053057116231	4.110615389187	5.813044063738
C	-1.159728195117	5.572404249951	6.086697097630
C	-0.245644142487	2.113931101198	4.607740467467
C	2.256176533406	1.849425001890	1.990670897652
N	0.380340448480	1.708110175501	3.471920024201
N	0.492874869494	4.402820636582	3.999851584665
O	-0.727853529196	1.315860143232	5.424690887439
H	0.273514357703	2.663976402095	1.603913682893
H	-1.588474465686	3.406110763625	6.447149710399
H	-1.945022470038	5.786955736489	6.819319154144
H	-1.375287457392	6.137453138760	5.165448964027
H	2.293715002173	1.784290405264	0.899942378037
H	0.645313012241	5.368718201113	4.309247816465
S	0.453533636718	6.211015745635	6.731321174011
C	0.208719297075	7.997723914124	6.511138664740
C	0.776041362009	0.328470473366	3.280020520150
C	2.191490726713	0.419944565930	2.606028371184
N	-0.041190686094	-0.407006039314	2.312315016256
H	0.797335438633	-0.166352339916	4.258878058851
C	3.118275811210	-1.143214658134	0.700122000754
C	2.777731089970	-2.123034315159	-0.242905269216
C	2.132105968988	-0.661707823016	1.553870115428
H	3.541225728198	-2.526333954686	-0.907834913828
C	1.457498217544	-2.586924691487	-0.325201396242
C	0.801370785581	-1.103664638474	1.444304320571
H	1.202774928078	-3.355375947795	-1.056590094049
C	0.452249355285	-2.082578619221	0.507368669212
H	-0.574942704163	-2.439689228948	0.427085717653
H	4.144193061497	-0.778472656296	0.771287307688
H	-0.847274139043	-0.900874391467	2.689707455324
C	4.547111211854	0.388536096999	5.390892304794
C	5.104572281008	1.023681778042	6.545164375551
C	3.593105911141	1.080789332604	4.630760958408
N	3.246022469434	0.213650440561	3.607458986753
N	4.554532605265	2.317199004758	6.718660296114
N	3.091102104381	2.318174865814	4.847606835787
H	4.909422718463	2.830715094320	7.525422318341
C	3.611045062216	2.913963650434	5.913737133077
N	3.251621402858	4.182680572618	6.229058086274
O	5.945914523010	0.595525016220	7.348340496943
H	3.458129744469	4.546861630697	7.152941787166
H	2.428679680408	4.570442398633	5.775402225487
N	4.771302588044	-0.875120313736	4.867940333994
C	3.976506057183	-0.949893678157	3.818716355500
H	3.871405888587	-1.806829789417	3.162459814872
H	-0.622878489852	8.351719643687	7.132566198201
H	1.137453637718	8.481101131520	6.837776454414
H	0.024971144116	8.235858989114	5.456120089534
H	3.145722097239	2.385498658136	2.332885696997

## PUBLICATIONS

**Table S22.** Cartesian coordinates (xyz coordinates) of the equilibrium structure **3b IV'** in units of Å

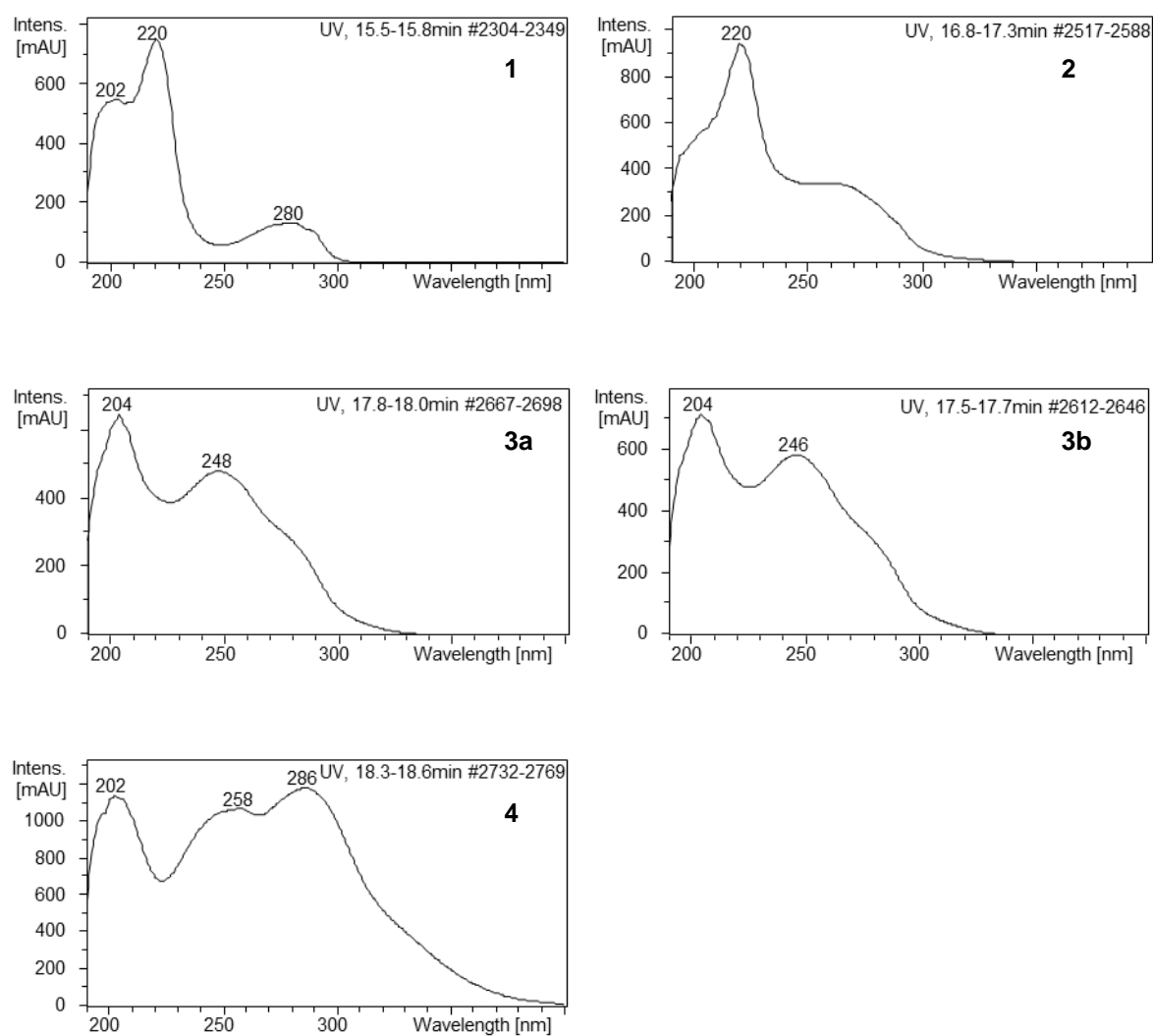
C	0.939145823641	3.938341391810	2.885858967363
O	1.524522887134	4.769846572718	2.190131900082
C	0.807387948952	2.499807944935	2.439144784936
C	-0.361966075576	3.411802742285	4.917077330871
C	-0.951335749662	3.852922642756	6.049206074896
C	-0.931607418557	5.267633699363	6.525097625012
C	-0.318096350282	1.947225948227	4.628293770028
C	2.129965690437	1.848195527345	1.968783840877
N	0.262157600093	1.591338193037	3.450079598526
N	0.296689638280	4.285340133407	4.041401037059
O	-0.748207485395	1.109899600617	5.434524461254
H	0.104104246430	2.536865883717	1.588610575366
H	-1.441536702392	3.100715164745	6.664780680437
H	-1.458798982848	5.366316027450	7.480508933801
H	0.104222887969	5.617622858636	6.675172726754
H	2.160697198788	1.783881144872	0.877832230718
H	0.121123973202	5.291478532038	4.181805186445
S	-1.698318639526	6.415788555206	5.298006658377
C	-1.189392855675	8.003865280688	6.017464425385
C	0.718749272756	0.237173231838	3.216363201191
C	2.144417674568	0.414686830120	2.581647037446
N	-0.036162253371	-0.493495477988	2.195530762172
H	0.738066835313	-0.294309769438	4.175875332029
C	3.215422752676	-1.106957168734	0.717747402638
C	2.950909123176	-2.090842658607	-0.245650619617
C	2.175342856501	-0.663872856479	1.526508249045
H	3.756866928278	-2.464229702160	-0.877107347847
C	1.652009047273	-2.596876504667	-0.391528493798
C	0.866517135234	-1.151533610244	1.356403269835
H	1.456356597006	-3.367530858157	-1.138557326831
C	0.593373873661	-2.133764606281	0.398417021595
H	-0.415903466825	-2.526320603544	0.270395244119
H	4.224196157668	-0.708929145335	0.838718427740
H	-0.842054683916	-1.020172687285	2.526831233131
C	4.398605543998	0.532224928947	5.444629772879
C	4.886705536047	1.218694749355	6.601688268775
C	3.466762630982	1.177157081738	4.619513639946
N	3.179781029798	0.271670735918	3.612271472825
N	4.303279459111	2.507460366549	6.705624882436
N	2.934637663617	2.409347475516	4.768541325283
H	4.614593376343	3.058777597882	7.505350842249
C	3.393025300362	3.055251054859	5.833002221814
N	2.986930546605	4.325691544385	6.063231632125
O	5.697421369590	0.836750565720	7.457562972184
H	3.147185838552	4.778080258498	6.955650178619
H	2.236969764690	4.673826620039	5.475691334372
N	4.670228759799	-0.743389464129	4.973333920791
C	3.923057546335	-0.869485602638	3.894204167417
H	3.863913611883	-1.749265322051	3.262367558715
H	-1.580675419639	8.786309089439	5.356147092845
H	-1.618382923495	8.127417912874	7.019645461920
H	-0.095514514322	8.077177957148	6.060286834819
H	2.992367424789	2.430025325234	2.304528771007

## Supplementary figures

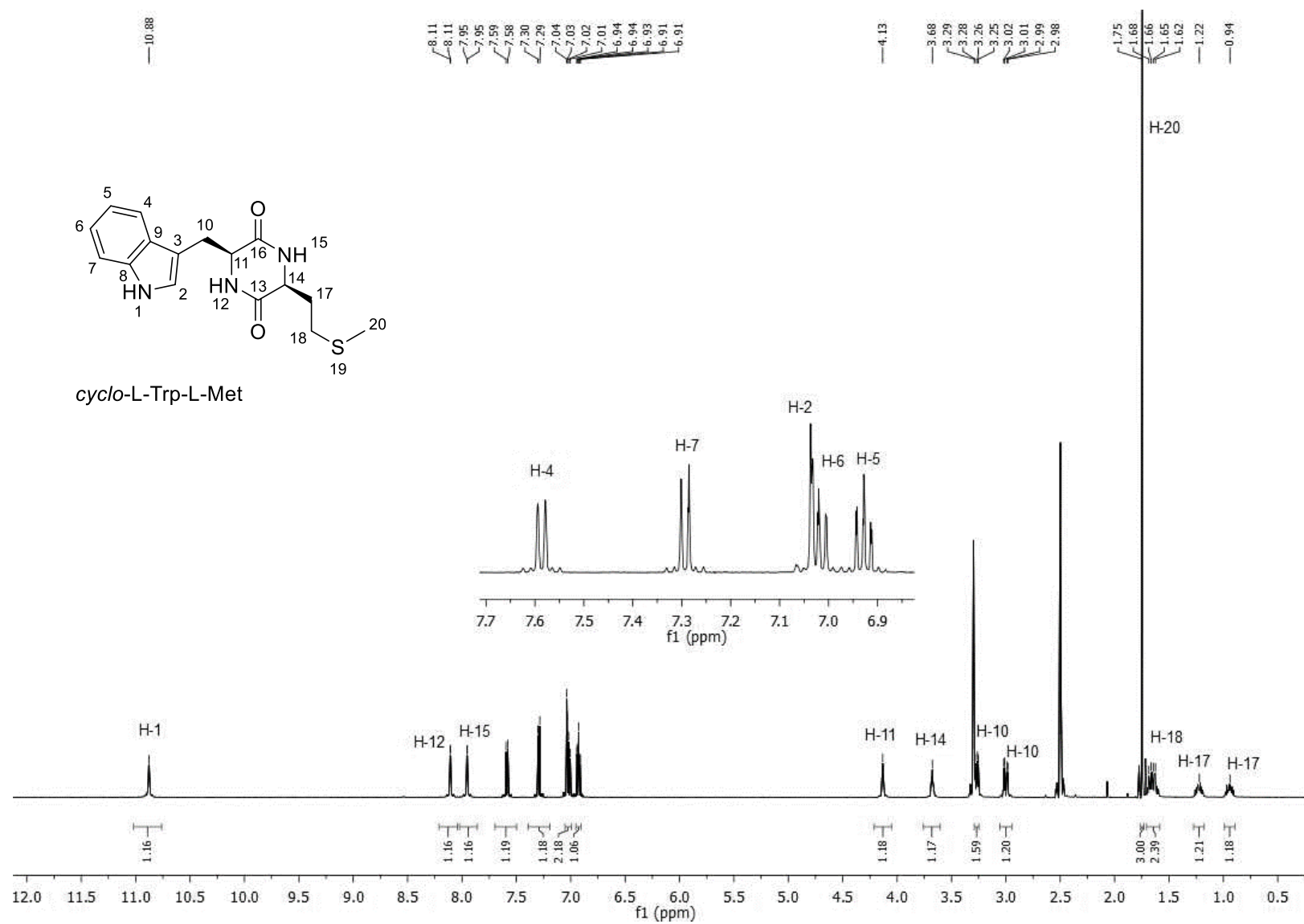


**Figure S24.** Sequence alignments of GtmE from *Streptomyces cinamoneus* (this study) with BcmF from *Streptomyces sapporonensis*, isopenicillin N synthase from *Aspergillus nidulans*, and isopenicillin N synthase from *Streptomyces cattleya*.<sup>9-11</sup> The red triangles mark the three postulated key residues for the catalytic triad of GtmE.

## PUBLICATIONS

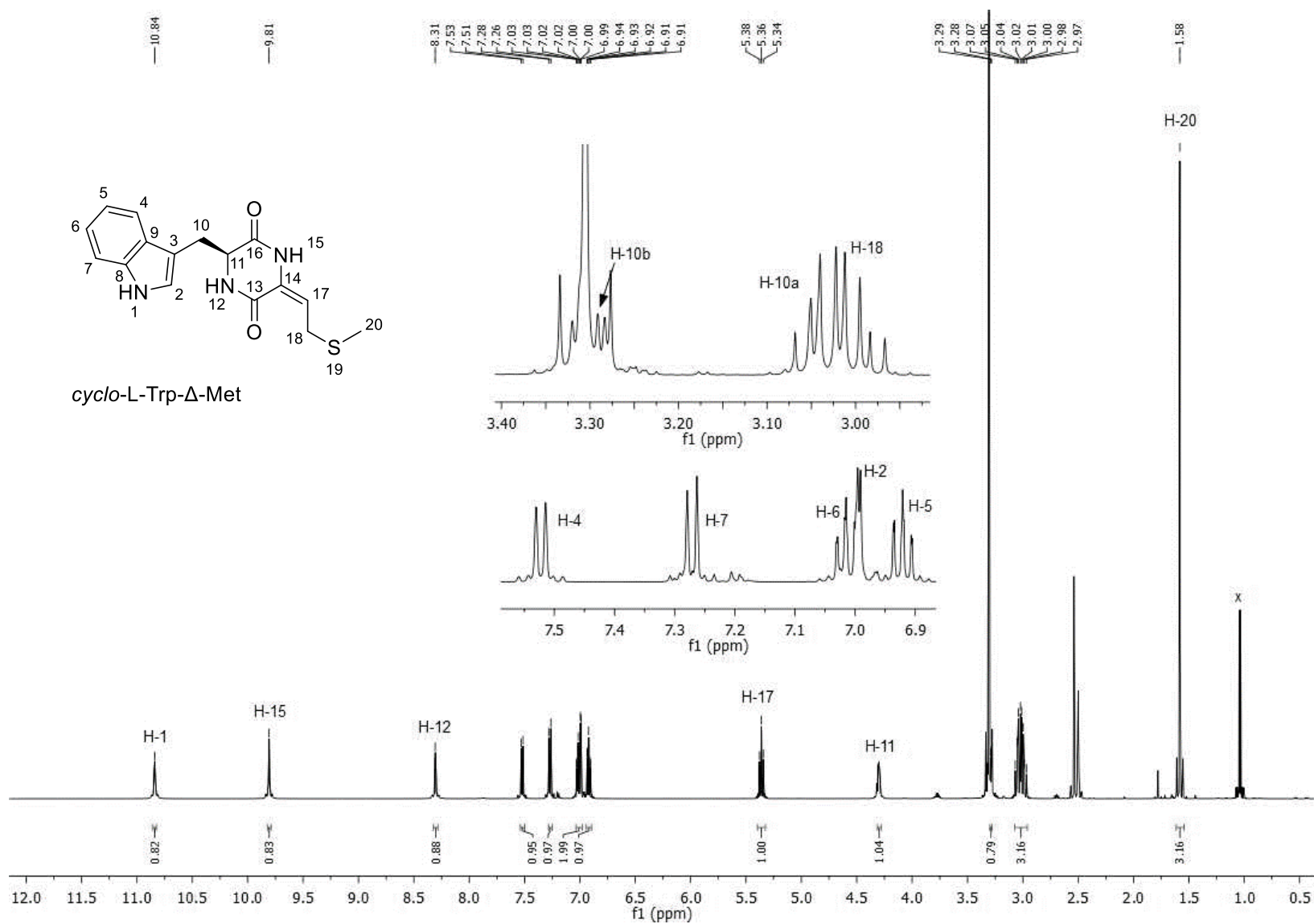


**Figure S25.** UV spectra of the identified products



**Figure S26.** <sup>1</sup>H-NMR spectrum of *cyclo*-L-Trp-L-Met (1) in DMSO-*d*<sub>6</sub> (500 MHz)

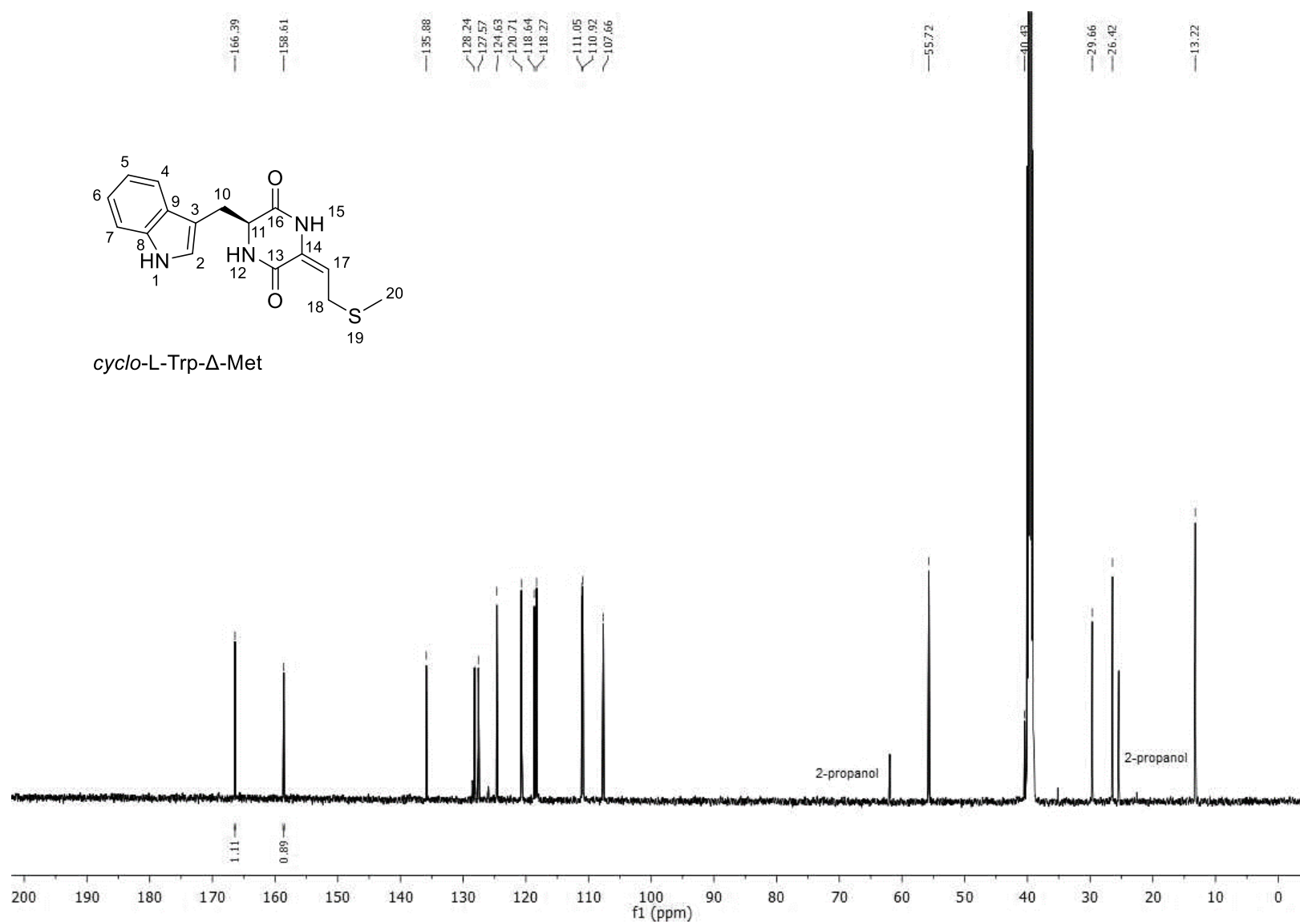
PUBLICATIONS



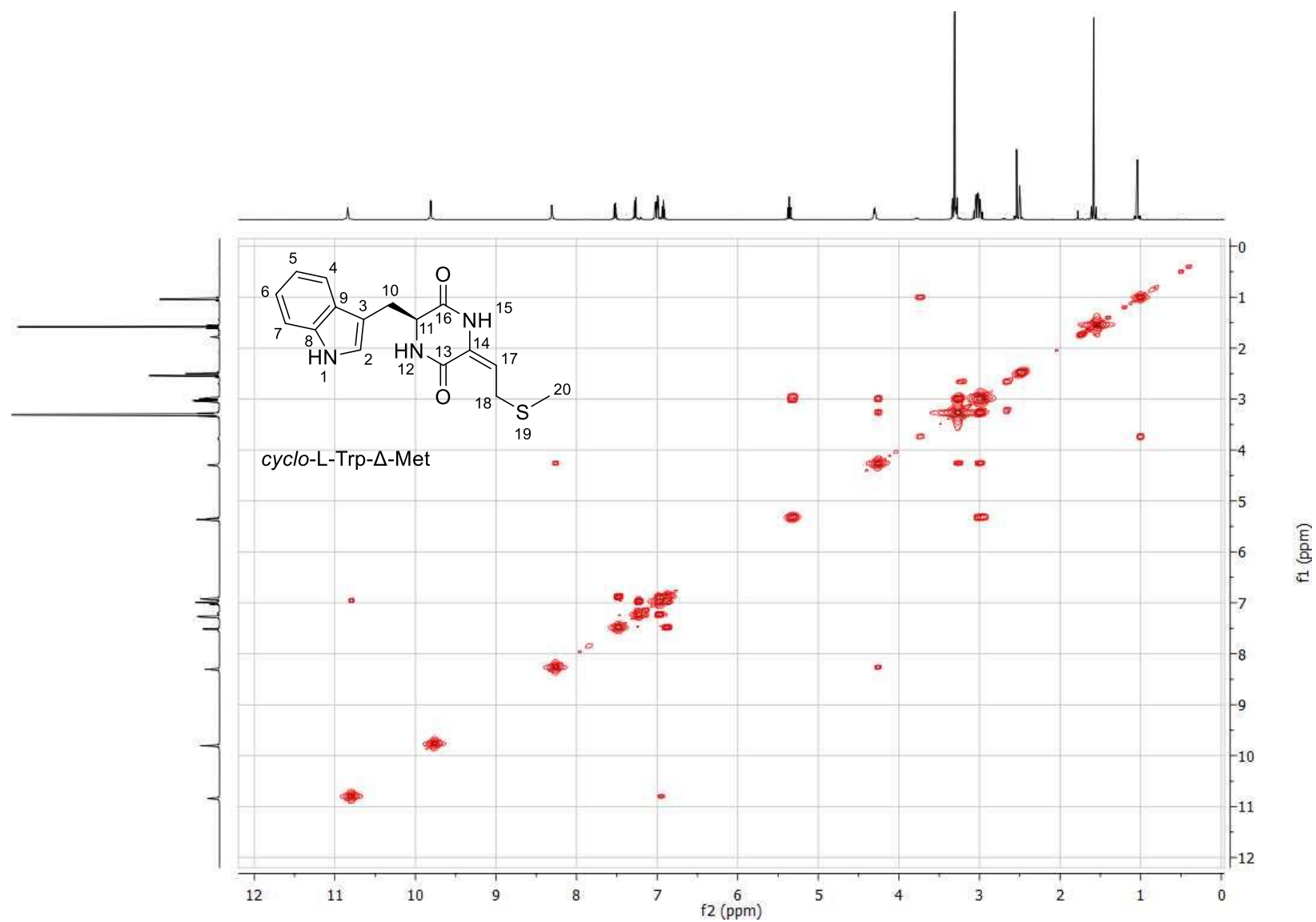
**Figure S27.**  $^1\text{H}$ -NMR spectrum of *cyclo*-L-Trp- $\Delta$ -Met (2) in  $\text{DMSO}-d_6$  (500 MHz)



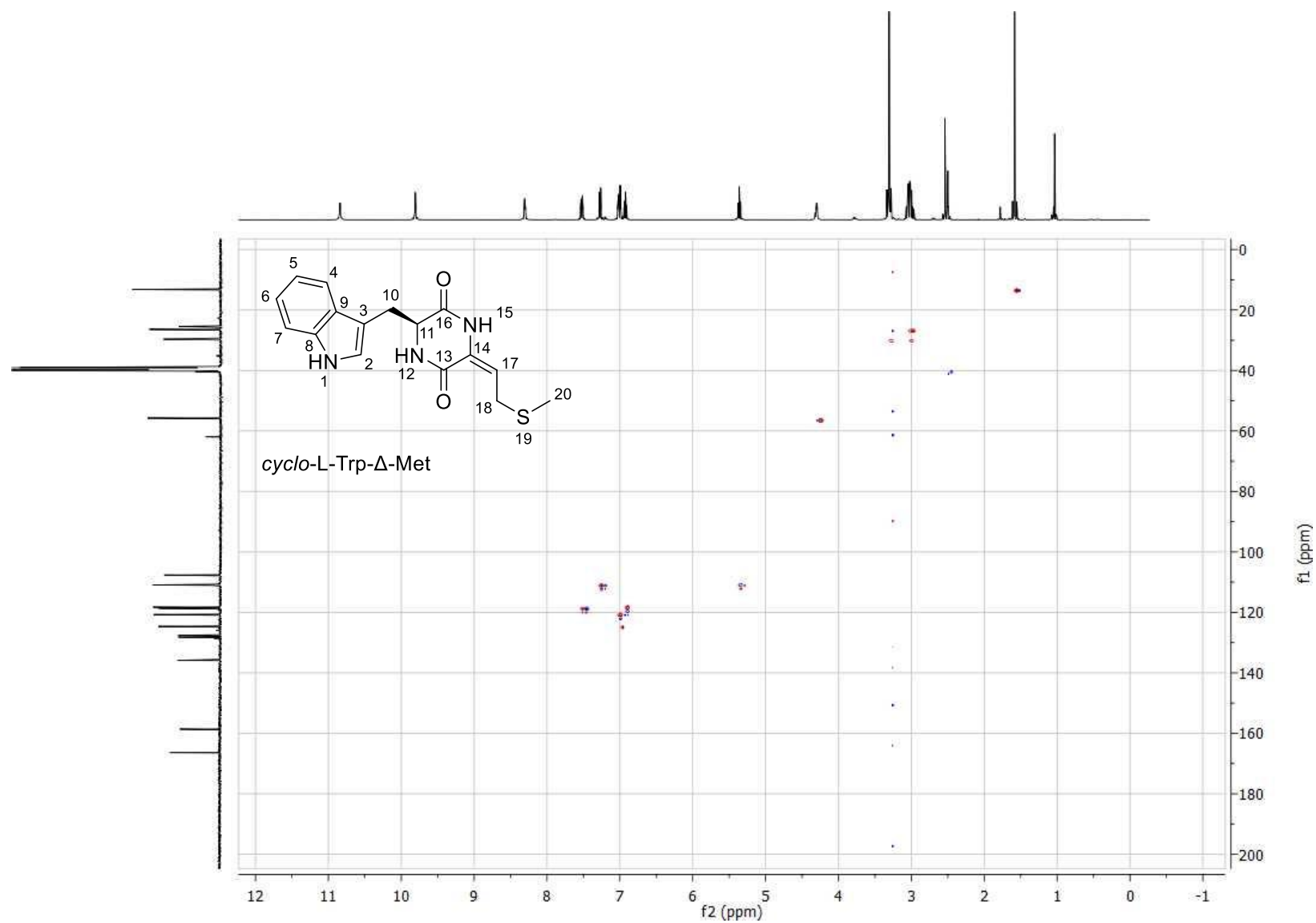
# PUBLICATIONS



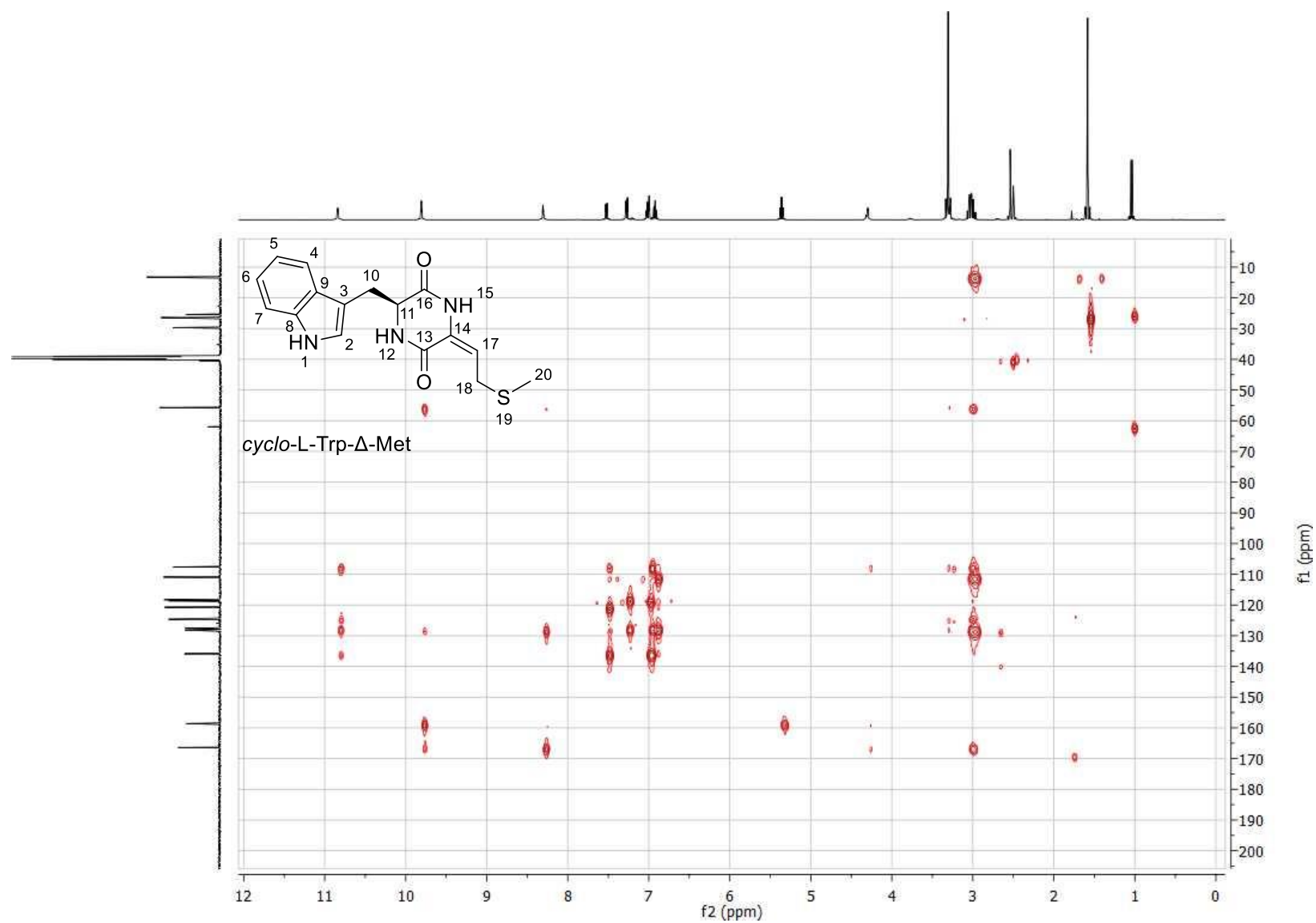
**Figure S28.**  $^{13}\text{C}\{^1\text{H}\}$ -NMR spectrum of *cyclo*-L-Trp-Δ-Met (**2**) in  $\text{DMSO-}d_6$  (125 MHz)



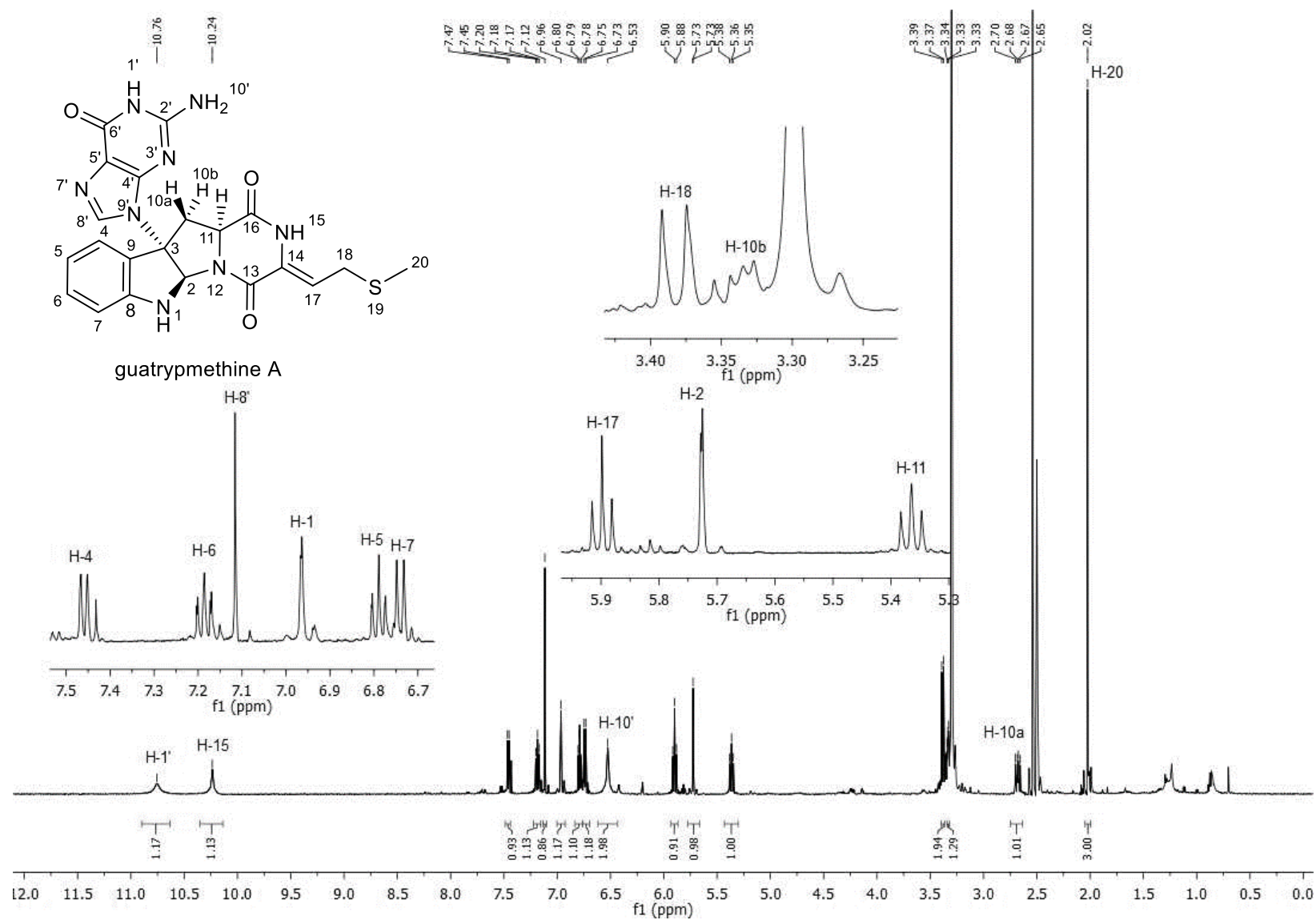
**Figure S29.**  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of *cyclo*-L-Trp- $\Delta$ -Met (2) in  $\text{DMSO}-d_6$



**Figure S30.**  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of *cyclo*-L-Trp- $\Delta$ -Met (2) in  $\text{DMSO}-d_6$

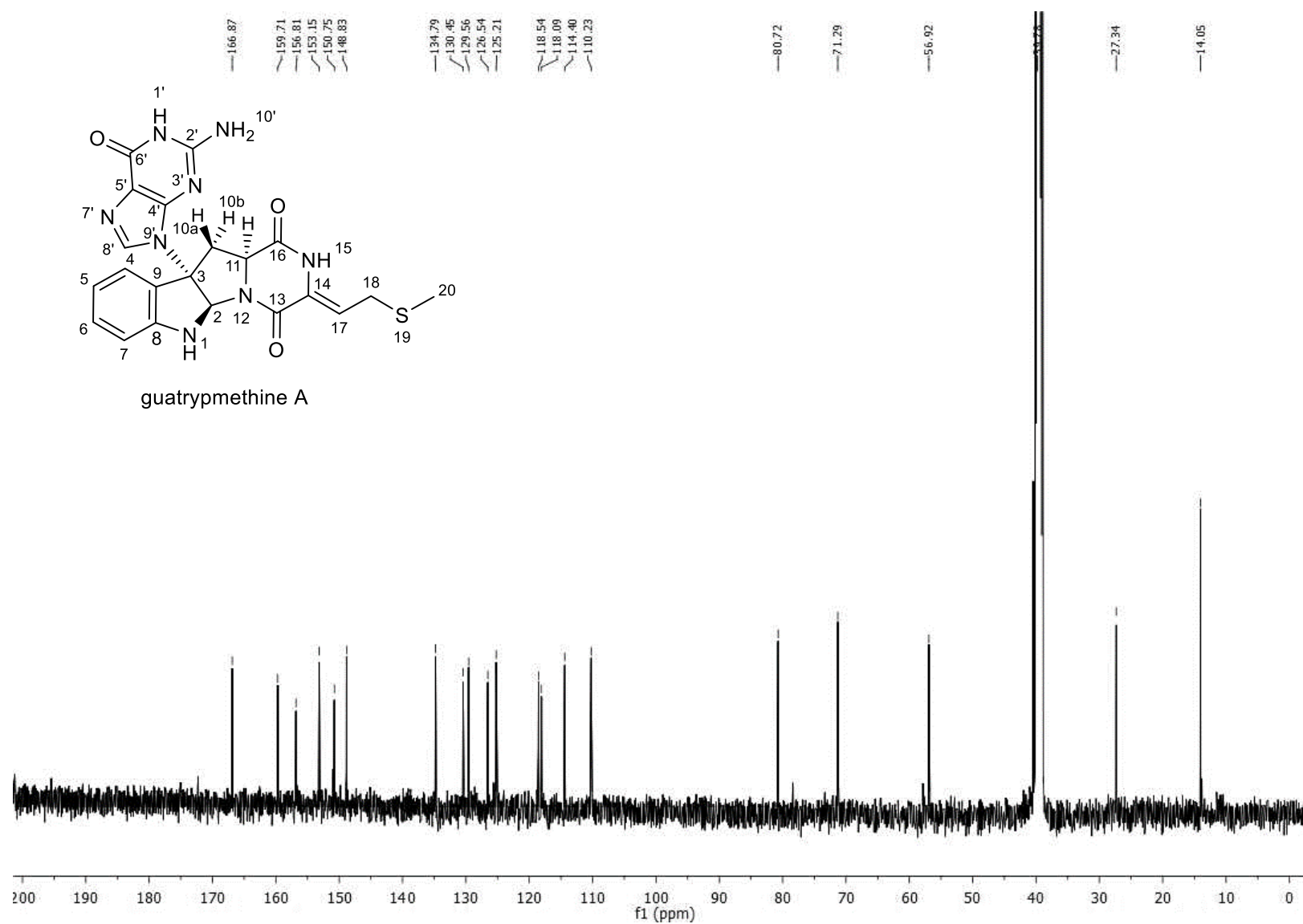


**Figure S31.**  $^1\text{H}$ - $^{13}\text{C}$  HMBC spectrum of *cyclo*-L-Trp- $\Delta$ -Met (2) in  $\text{DMSO}-d_6$

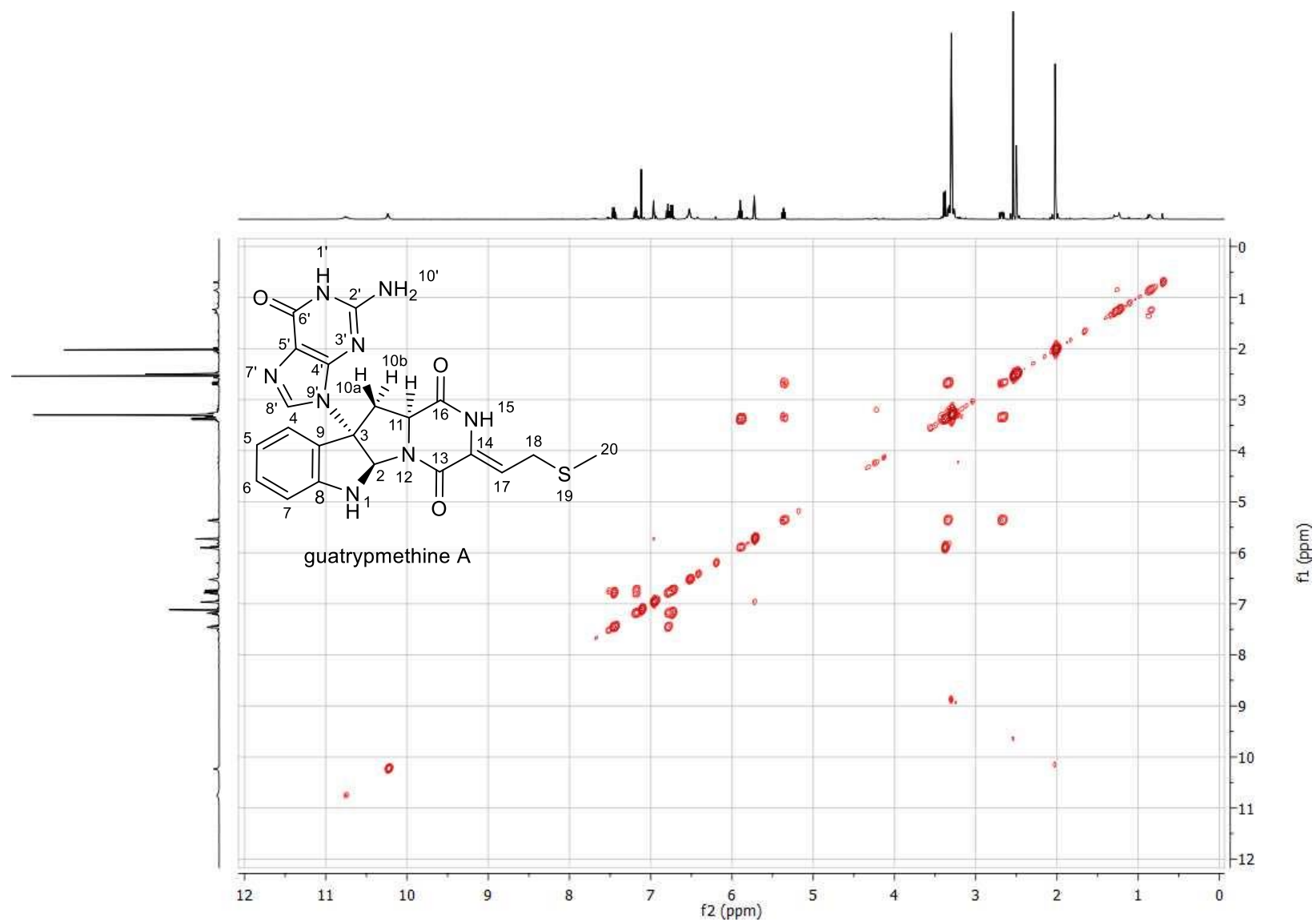


**Figure S32.** <sup>1</sup>H-NMR spectrum of guatrypmethine A (**3a**) in DMSO-*d*<sub>6</sub> (500 MHz)

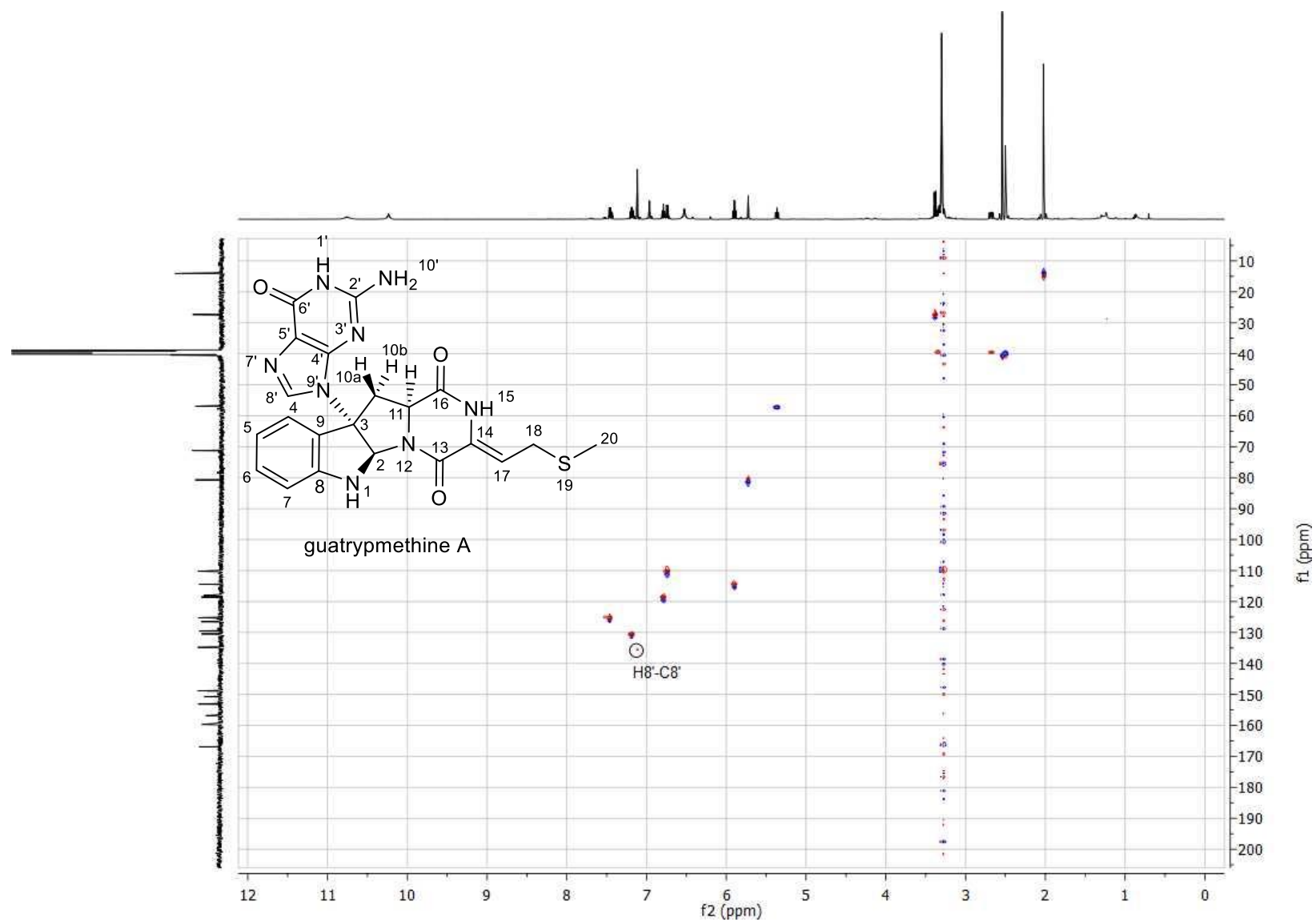
# PUBLICATIONS



**Figure S33.**  $^{13}\text{C}\{^1\text{H}\}$ -NMR spectrum of guatrypmethine A (**3a**) in  $\text{DMSO-}d_6$  (125 MHz)



**Figure S34.**  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of guatrypmethine A (3a) in  $\text{DMSO}-d_6$



**Figure S35.**  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of guatrypmethine A (**3a**) in  $\text{DMSO}-d_6$



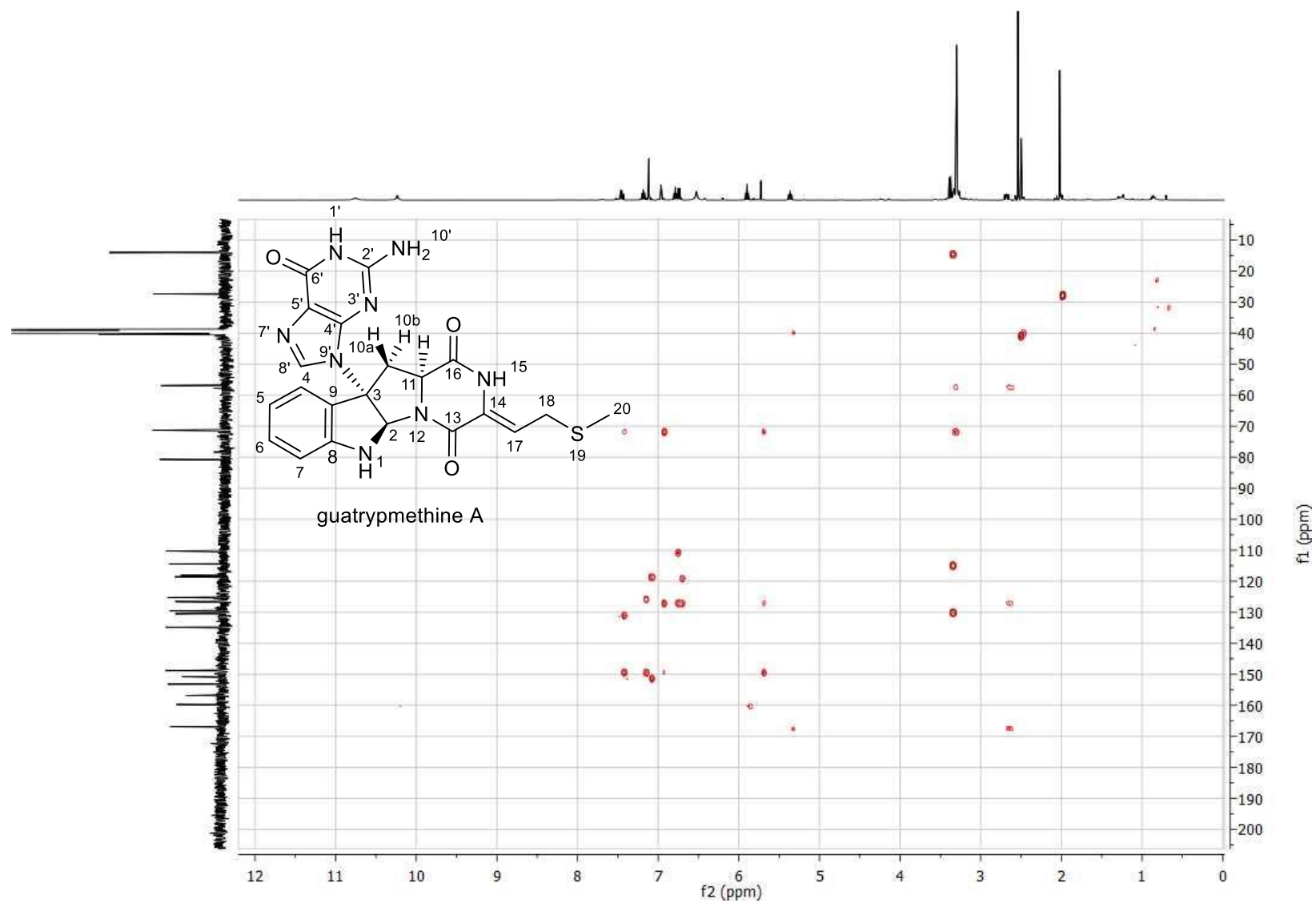


Figure S36.  $^1\text{H}$ - $^{13}\text{C}$  HMBC spectrum of guatrypmethine A (**3a**) in  $\text{DMSO}-d_6$

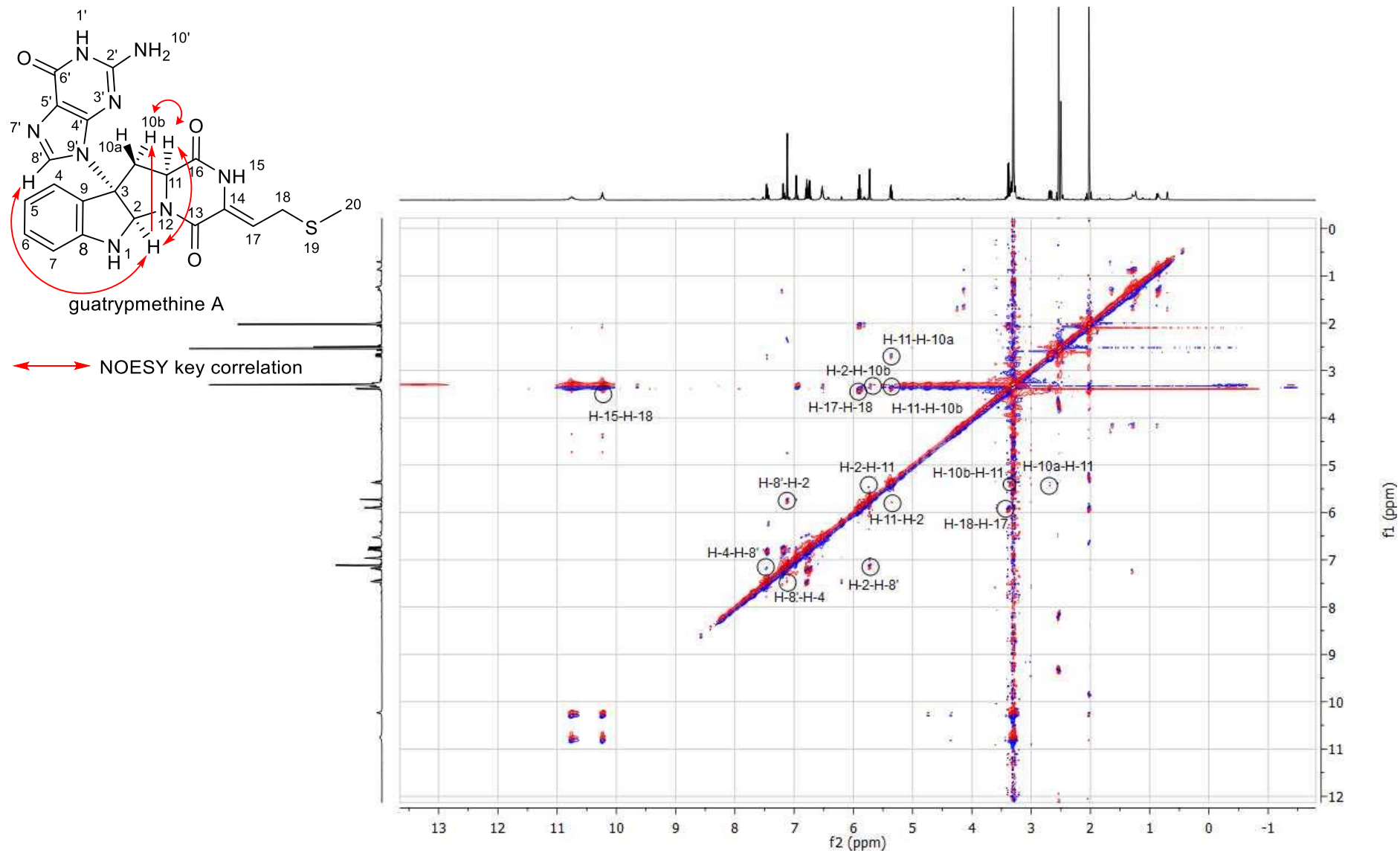
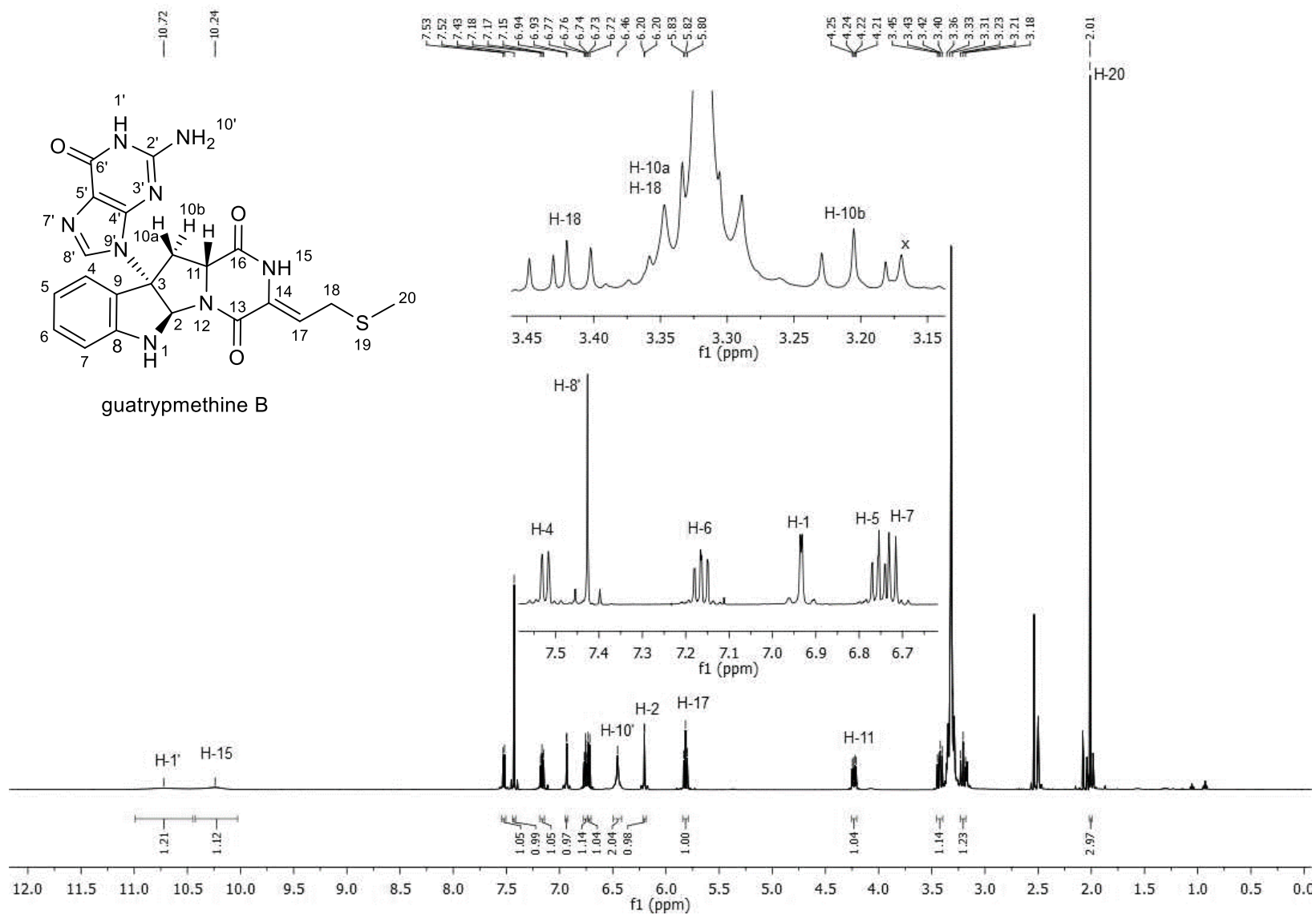


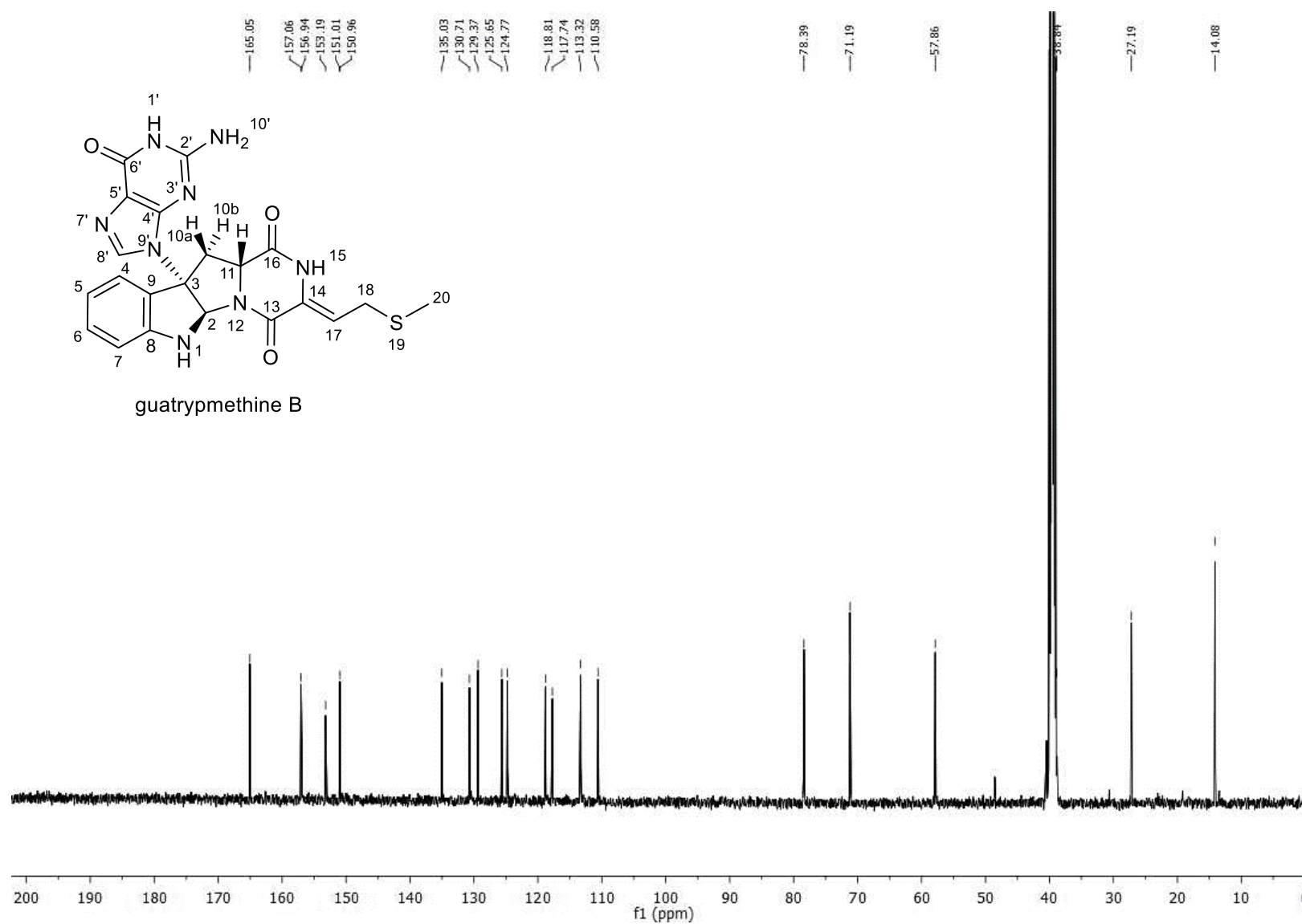
Figure S37. NOESY spectrum of guatrypmethine A (3a) in DMSO- $d_6$

PUBLICATIONS



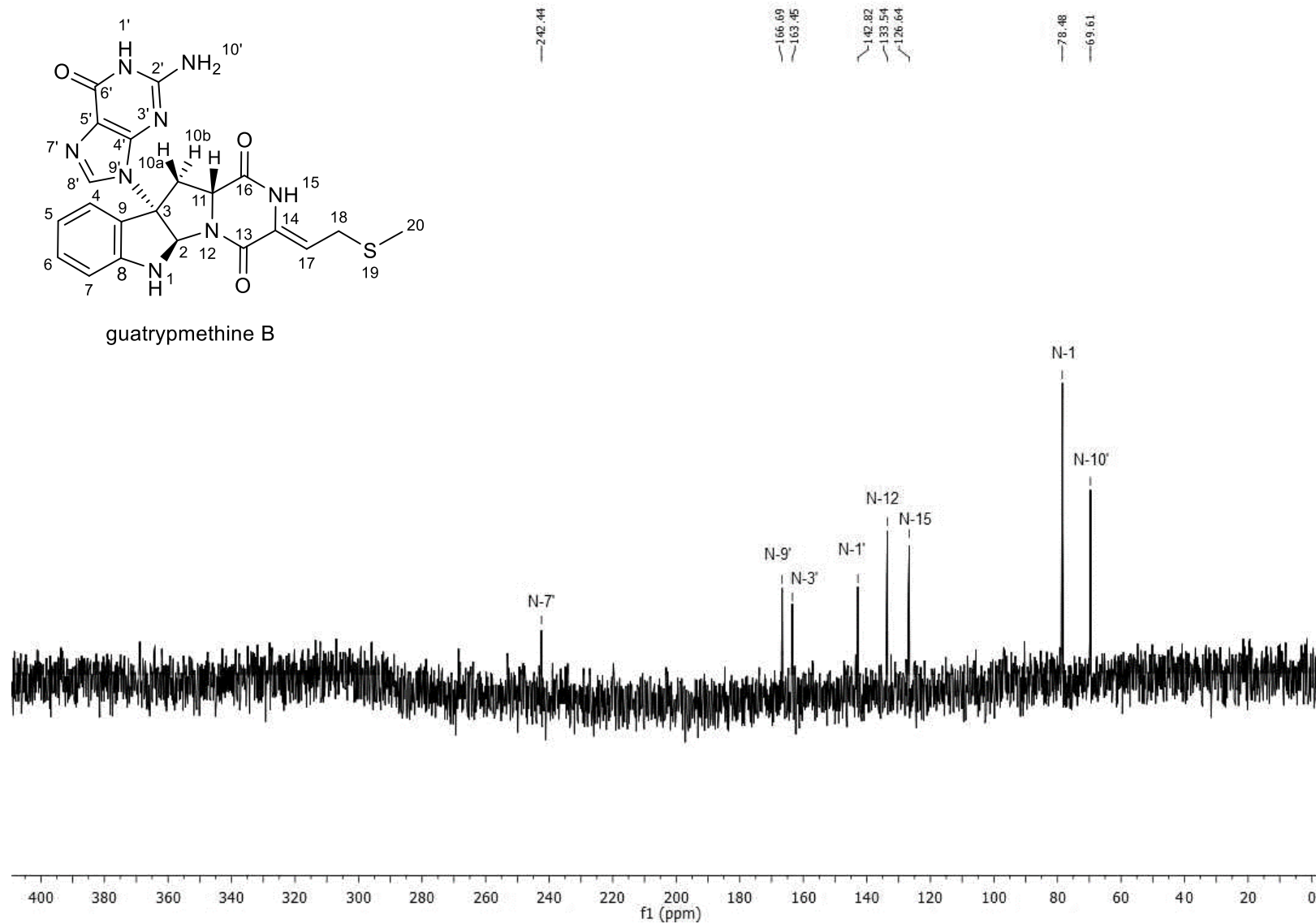
**Figure S38.**  $^1\text{H}$ -NMR spectrum of guatrypmethine B (**3b**) in  $\text{DMSO}-d_6$  (500 MHz)

# PUBLICATIONS

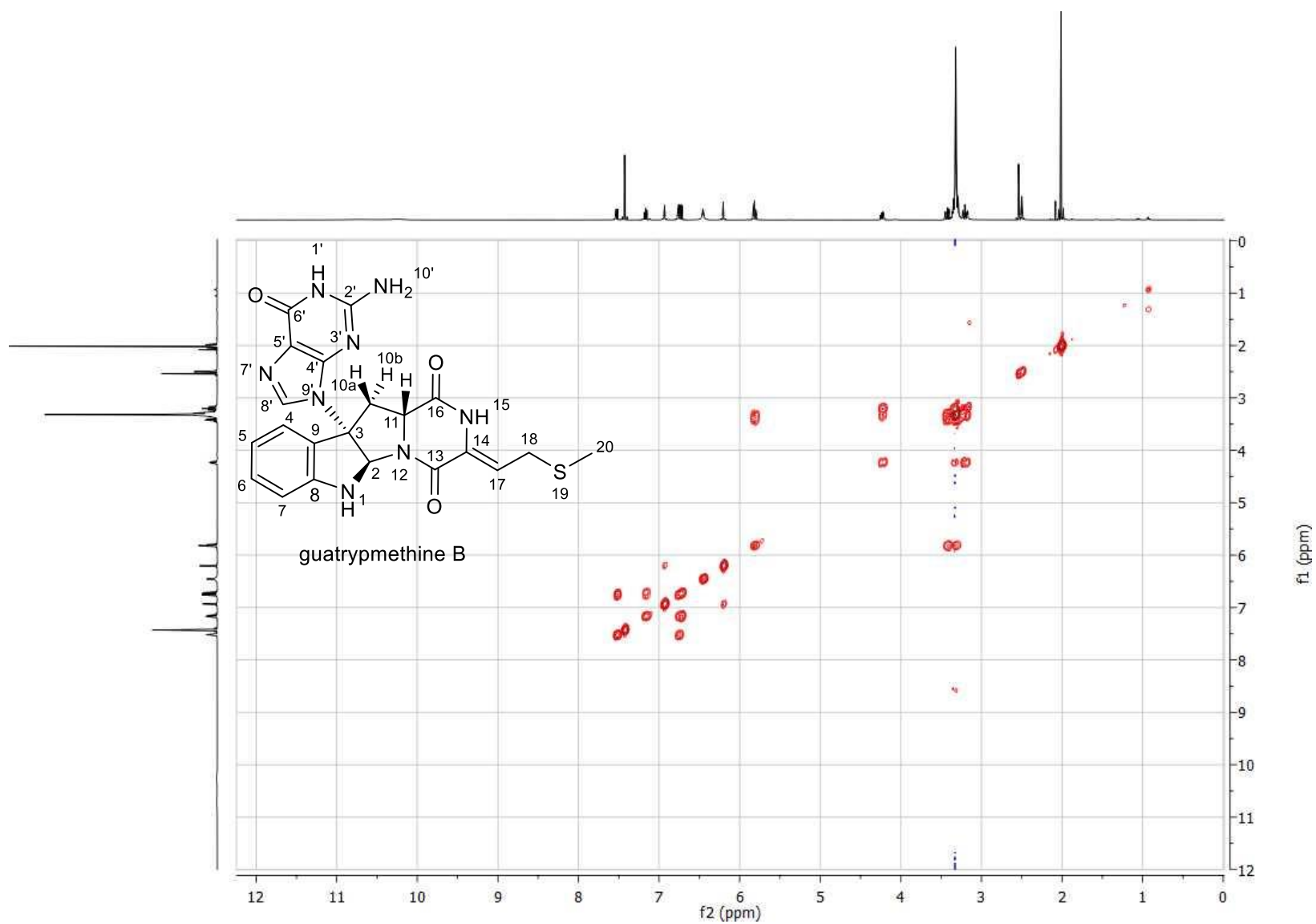


**Figure S39.**  $^{13}\text{C}\{^1\text{H}\}$ -NMR spectrum of guatrypmethine B (**3b**) in  $\text{DMSO}-d_6$  (125 MHz)

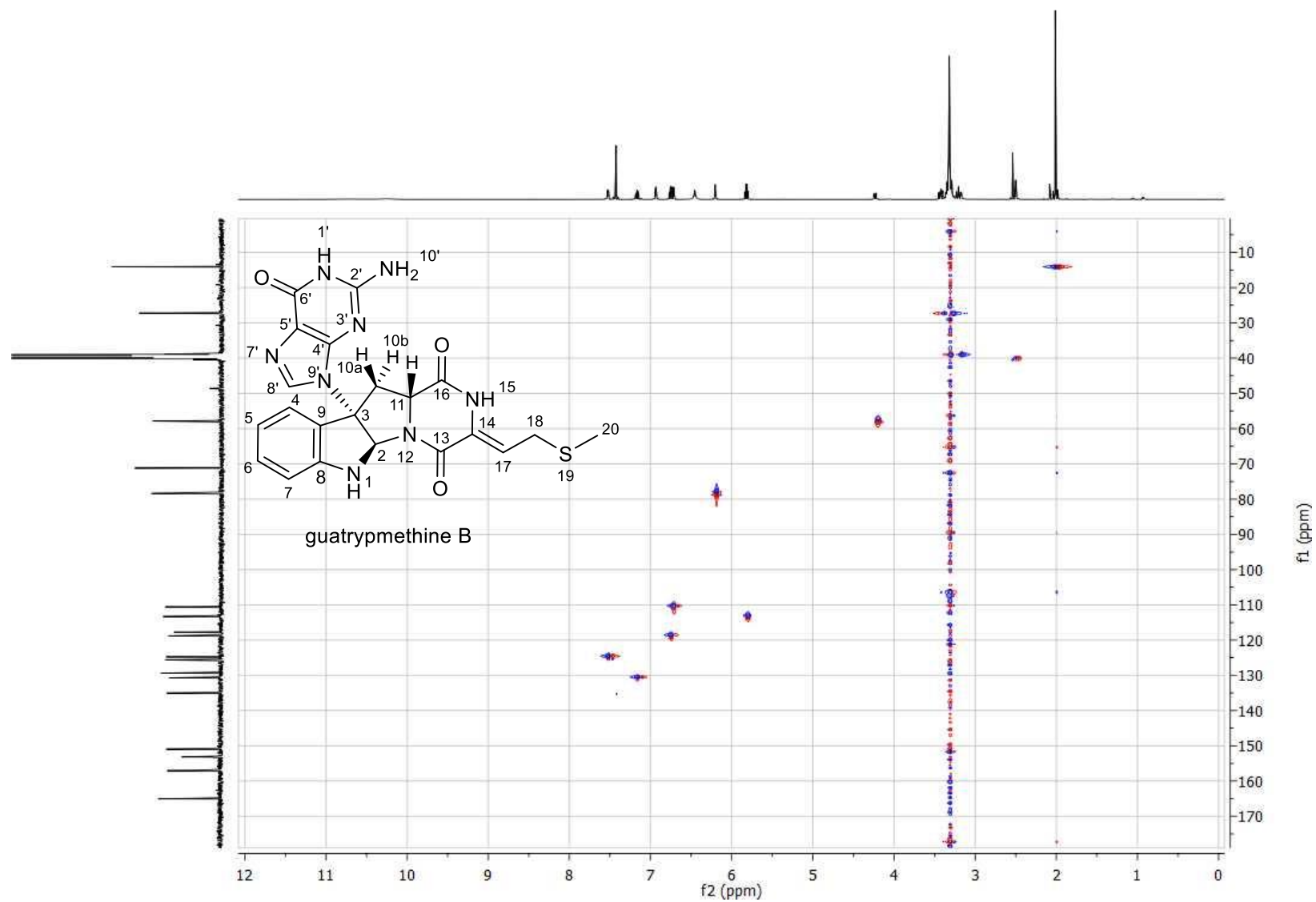
# PUBLICATIONS



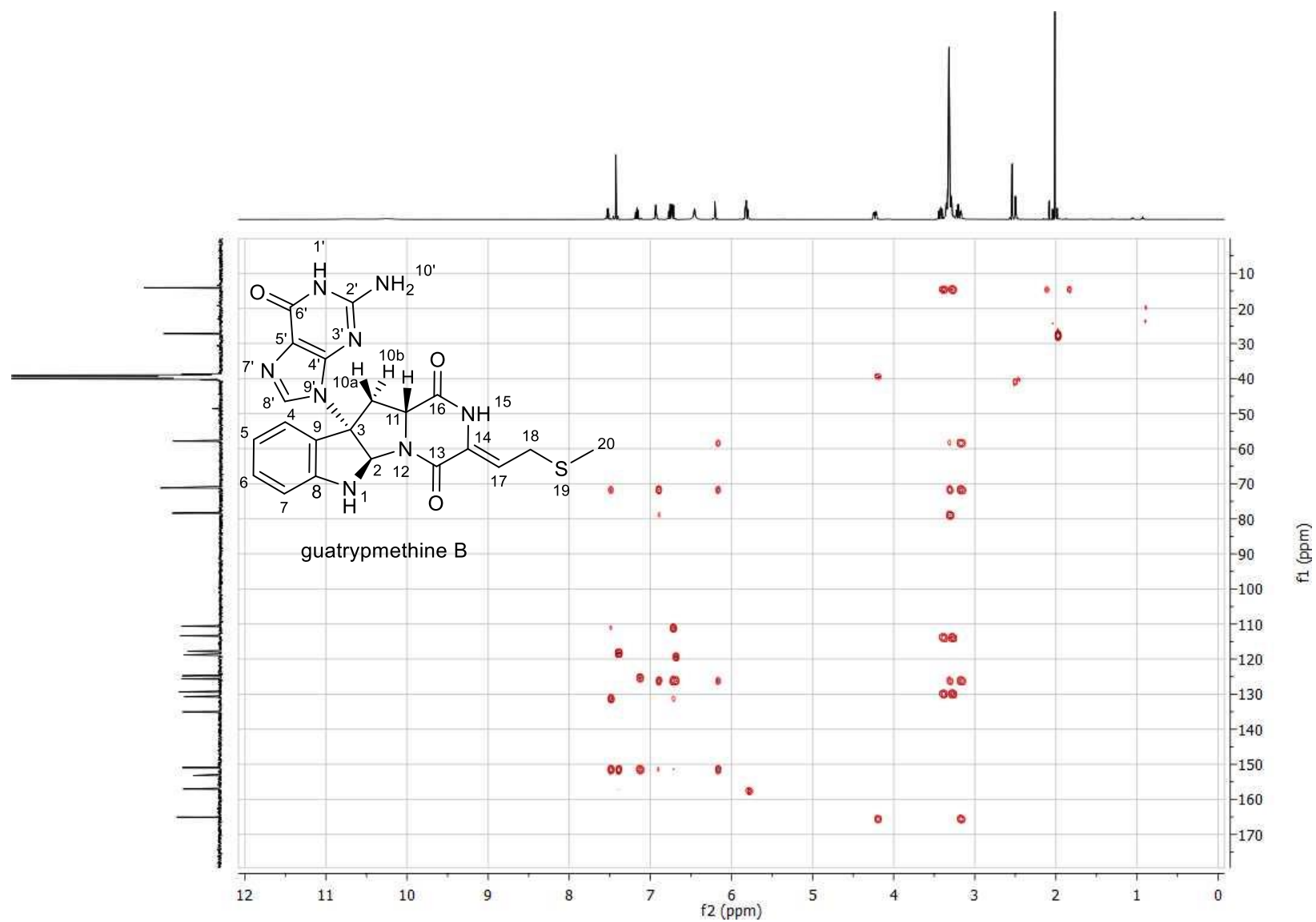
**Figure S40.**  $^{15}\text{N}\{^1\text{H},^{13}\text{C}\}$  inverse gated spectrum of guatrypmethine B (**3b**) in  $\text{DMSO}-d_6$  (40 MHz)



**Figure S41.**  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of guatrypmethine B (**3b**) in  $\text{DMSO}-d_6$

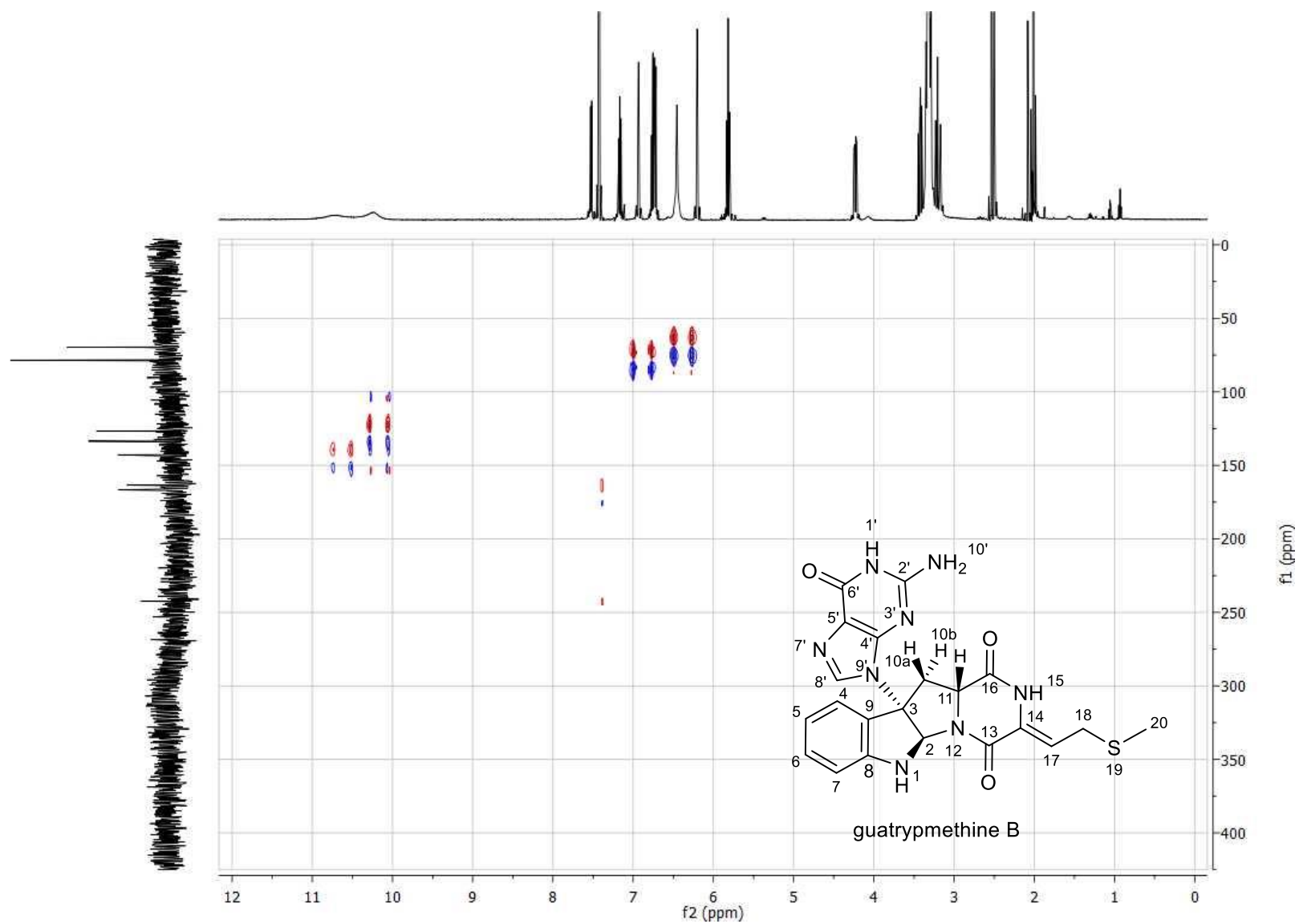


**Figure S42.**  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of guatrypmethine B (**3b**) in  $\text{DMSO}-d_6$

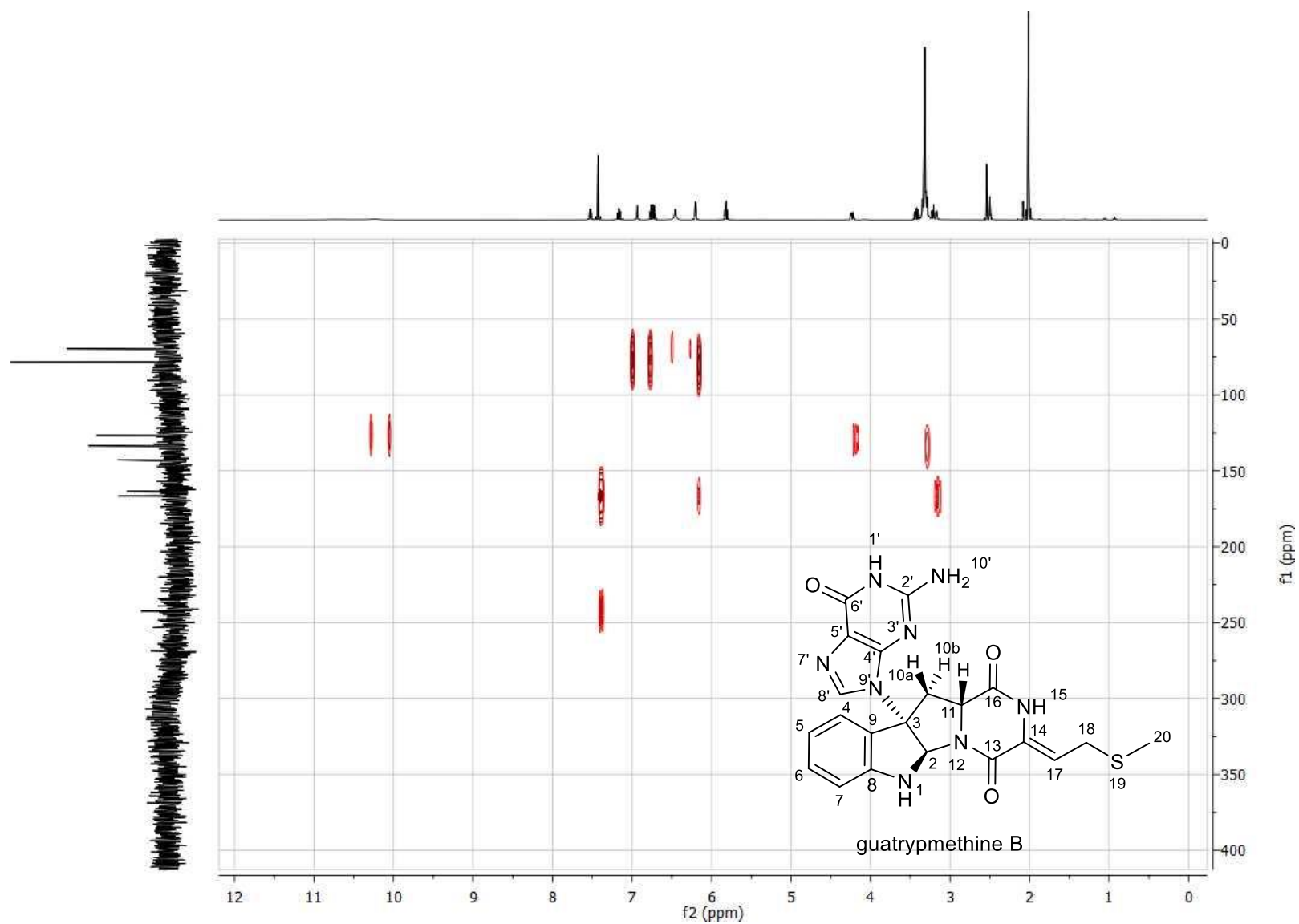


**Figure S43.**  $^1\text{H}$ - $^{13}\text{C}$  HMBC spectrum of guatrypmethine B (**3b**) in  $\text{DMSO}-d_6$

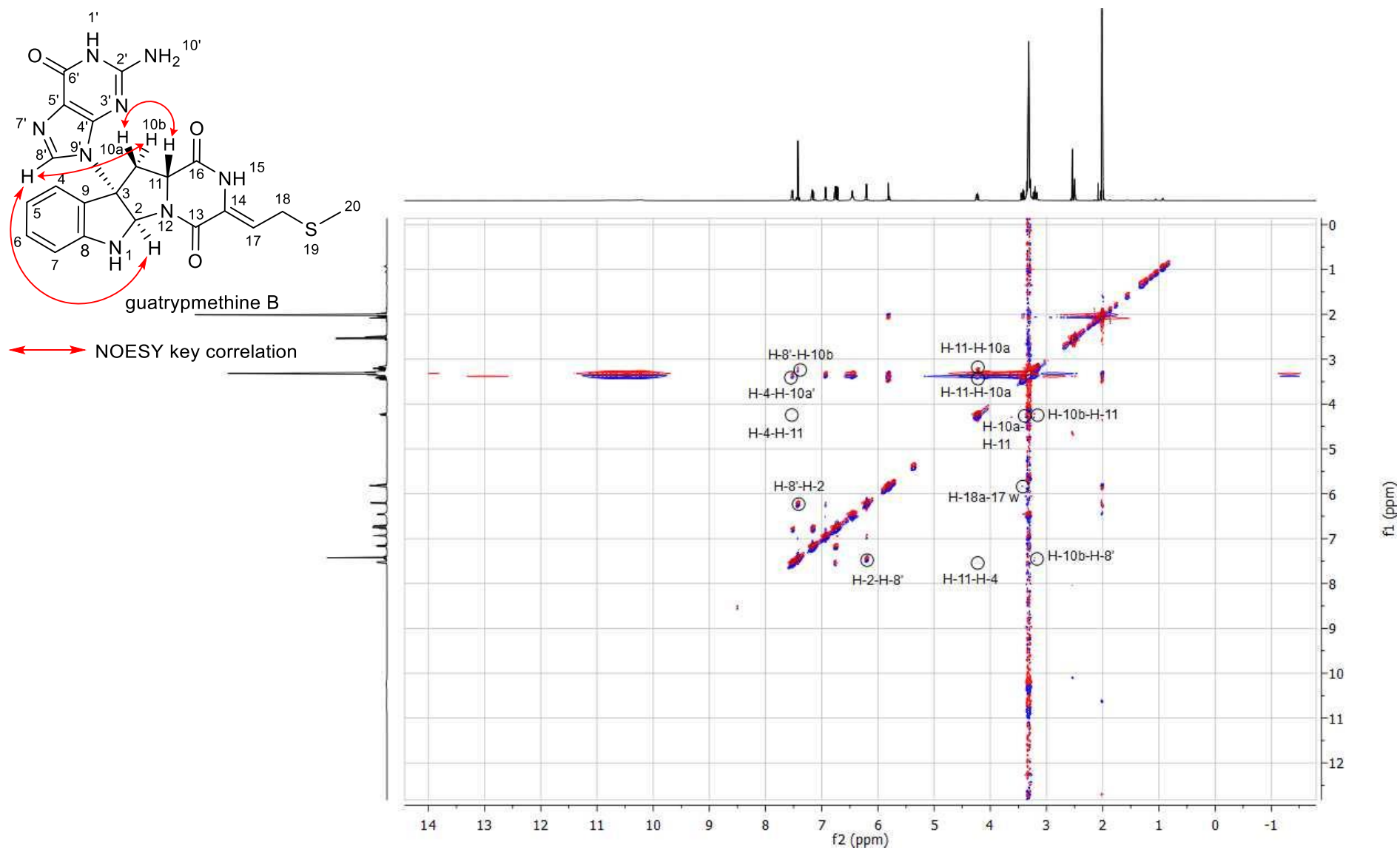




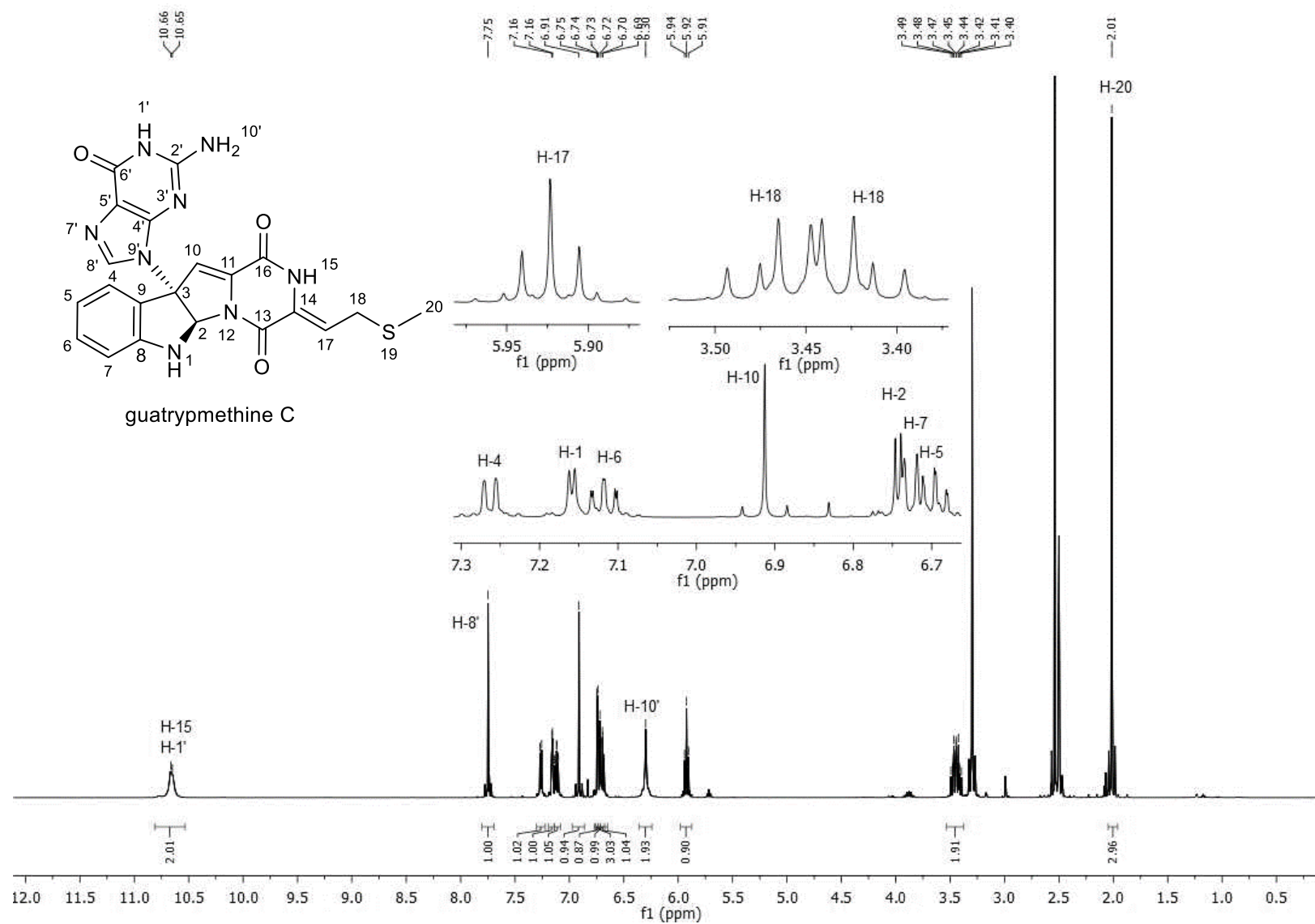
**Figure S44.**  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of guatrypmethine B (**3b**) in  $\text{DMSO}-d_6$



**Figure S45.**  $^1\text{H}$ - $^{15}\text{N}$  HMBC spectrum of guatrypmethine B (**3b**) in  $\text{DMSO}-d_6$

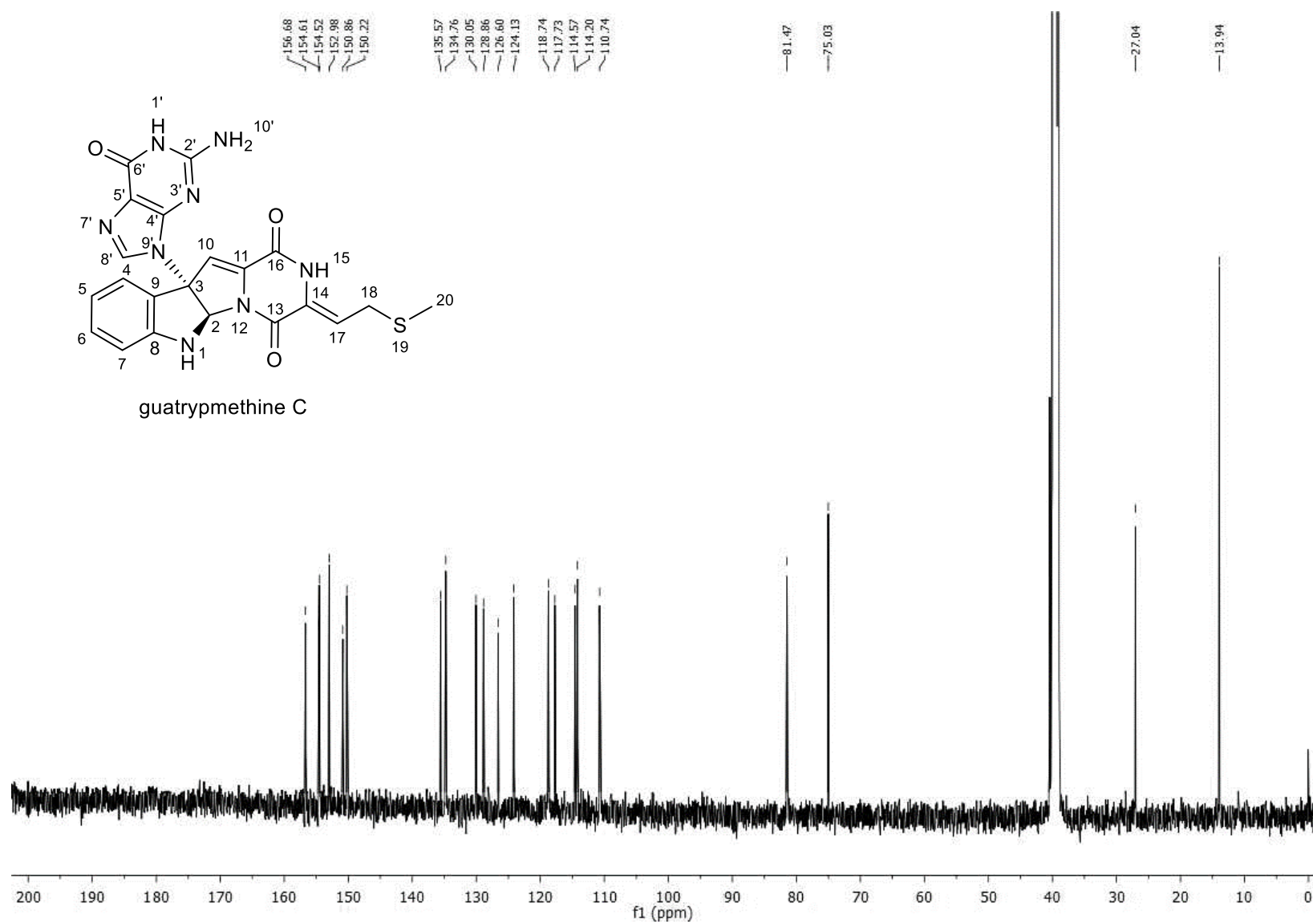


**Figure S46.** NOESY spectrum of quatrypmethine B (**3b**) in DMSO- $d_6$



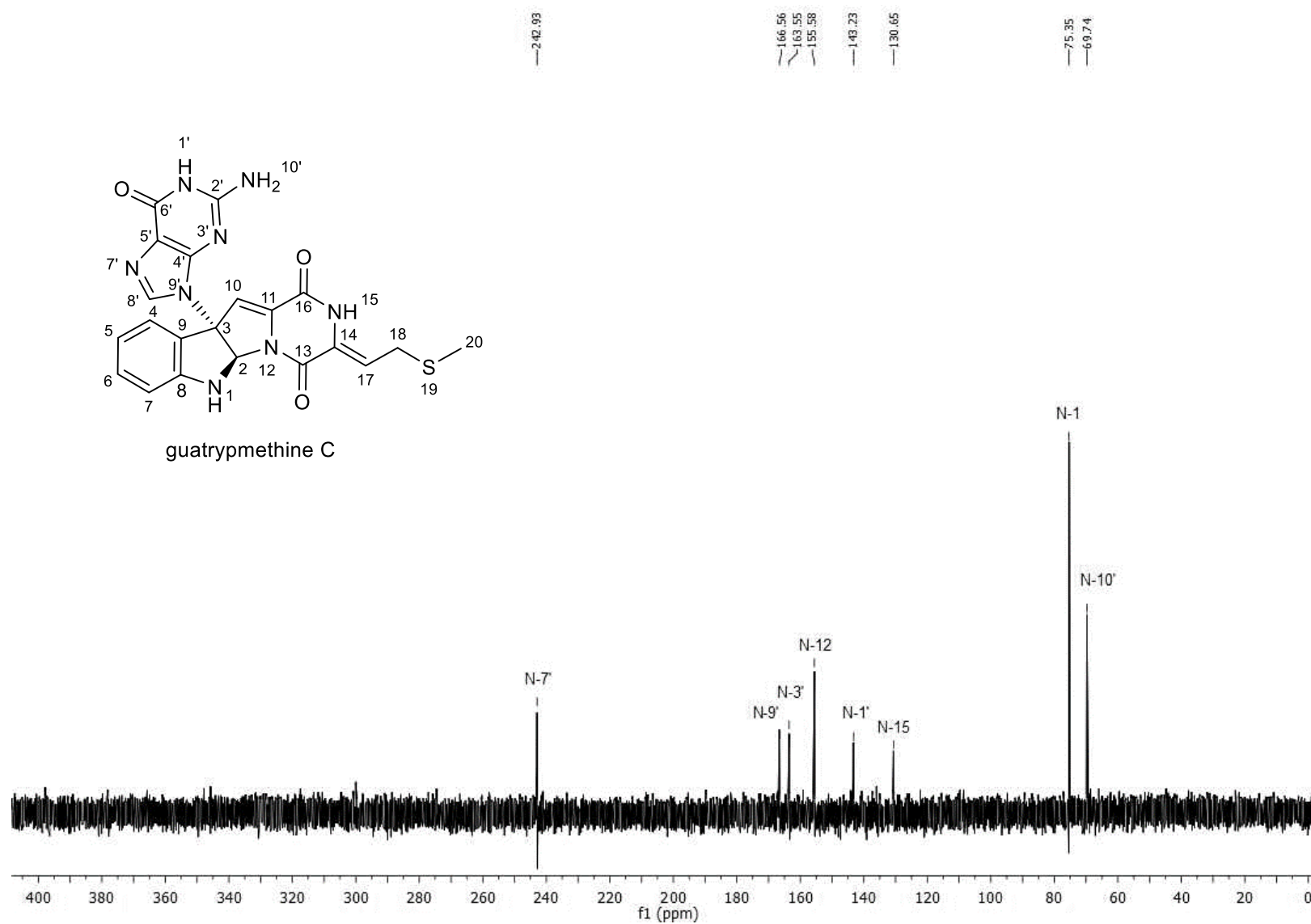
**Figure S47.**  $^1\text{H}$ -NMR spectrum of guatrypmethine C (4) in  $\text{DMSO}-d_6$  (500 MHz)

# PUBLICATIONS

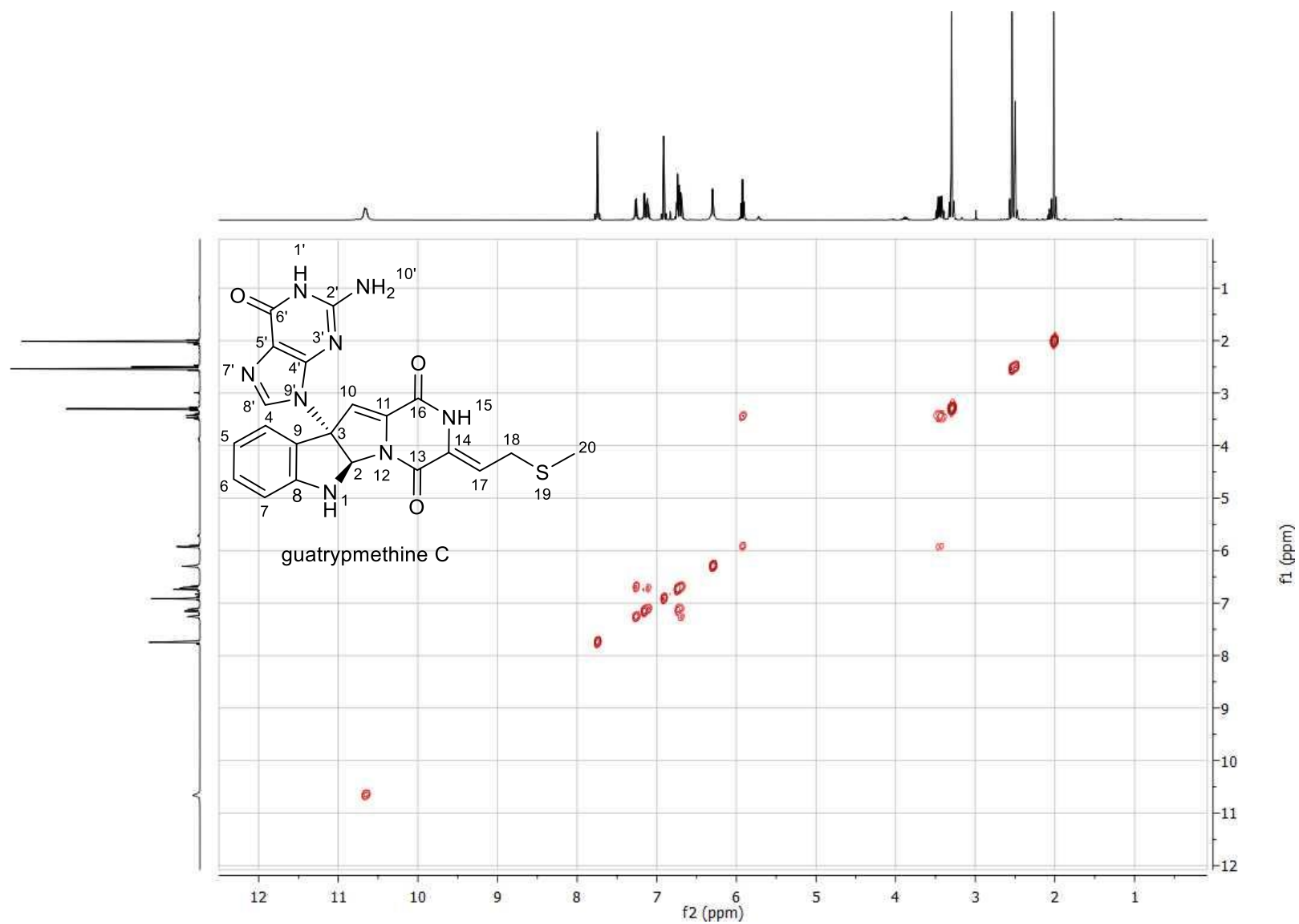


**Figure S48.**  $^{13}\text{C}\{^1\text{H}\}$ -NMR spectrum of guatrypmethine C (4) in  $\text{DMSO}-d_6$  (125 MHz)

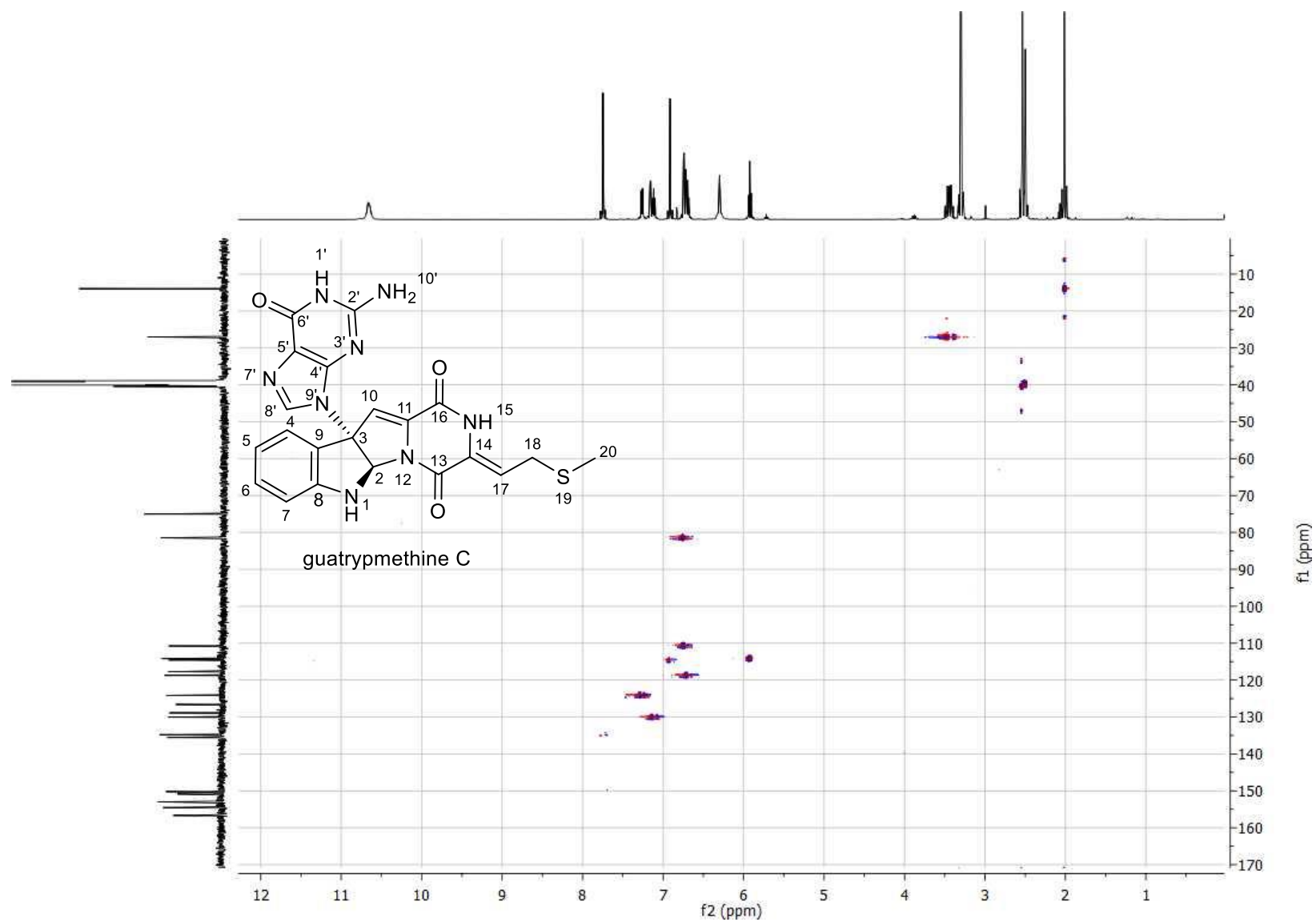
# PUBLICATIONS



**Figure S49.**  $^{15}\text{N}\{^1\text{H},^{13}\text{C}\}$  inverse gated spectrum of guatrypmethine C (4) in  $\text{DMSO-}d_6$  (40 MHz)

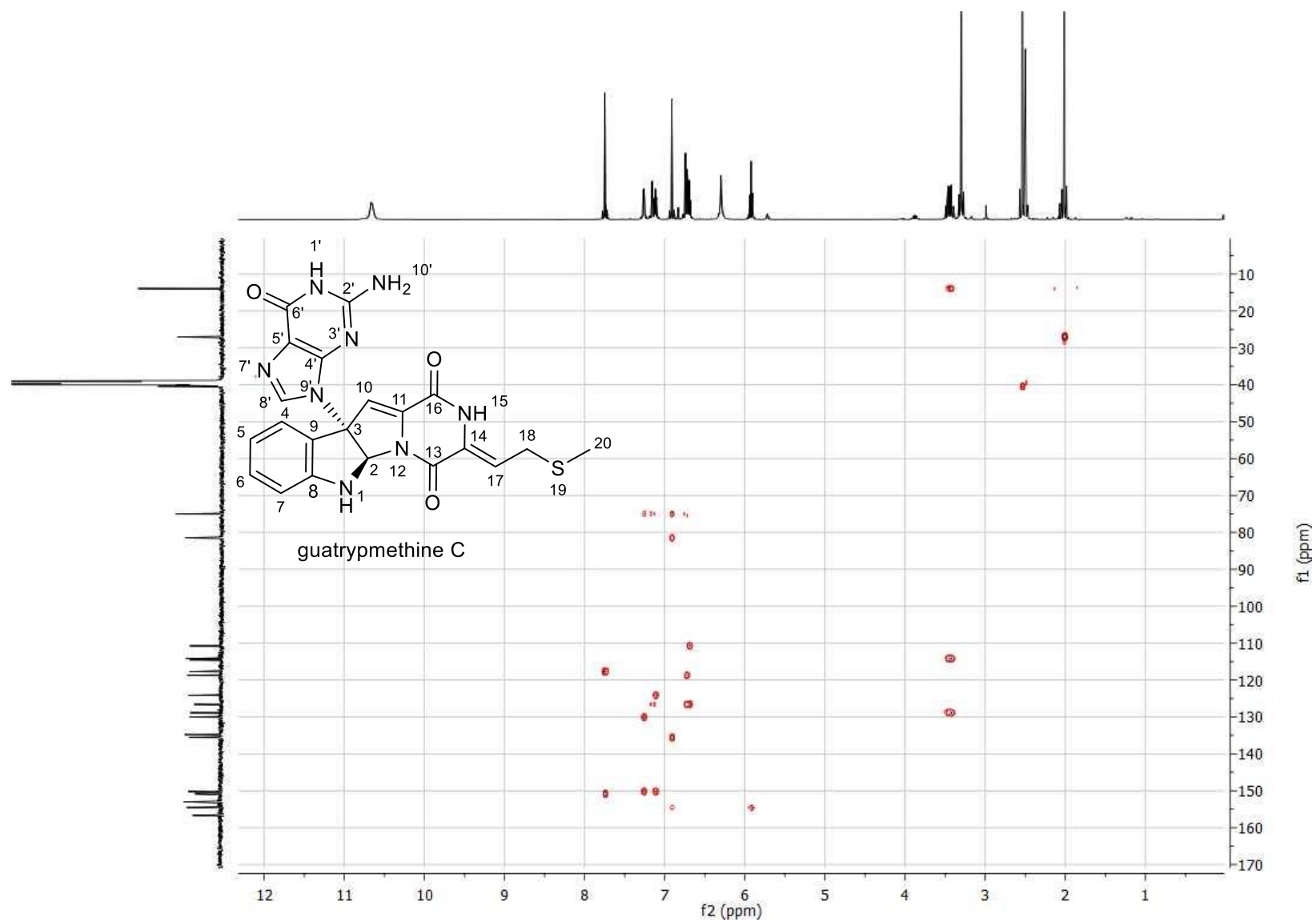


**Figure S50.**  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of guatrypmethine C (4) in  $\text{DMSO}-d_6$

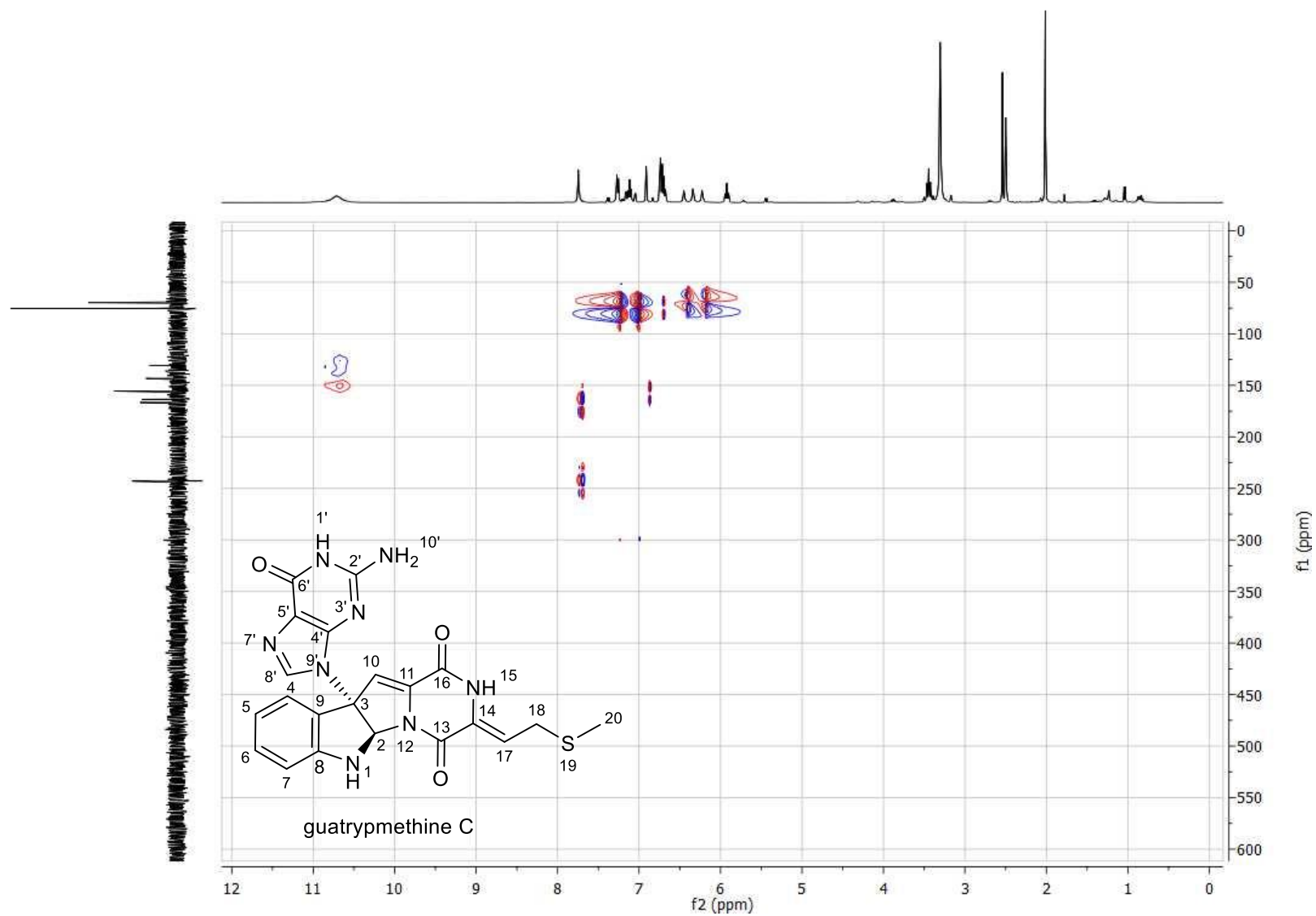


**Figure S51.**  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of guatrypmethine C (4) in  $\text{DMSO}-d_6$

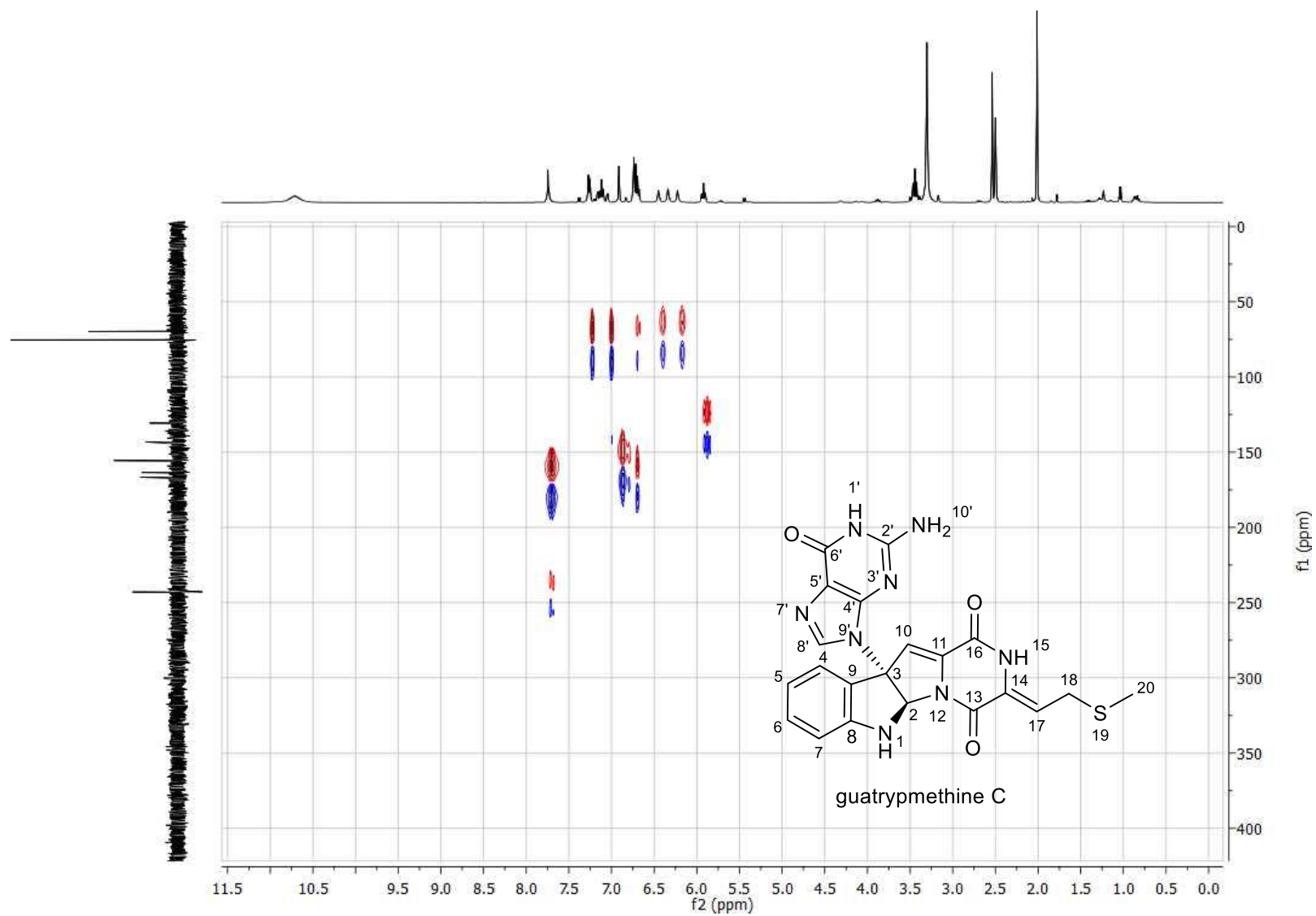




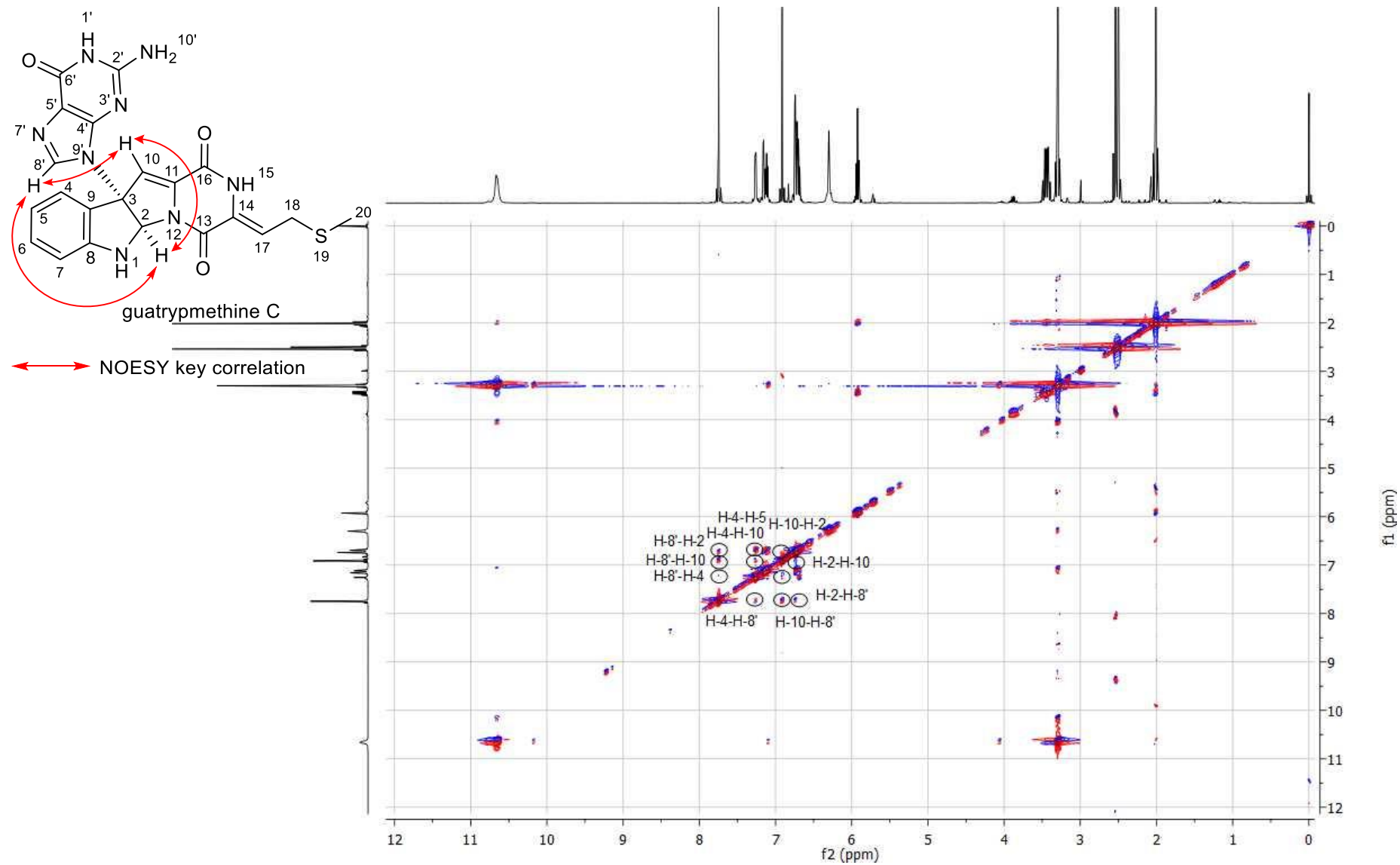
**Figure S52.**  $^1\text{H}$ - $^{13}\text{C}$  HMBC spectrum of guatrypmethine C (4) in  $\text{DMSO}-d_6$



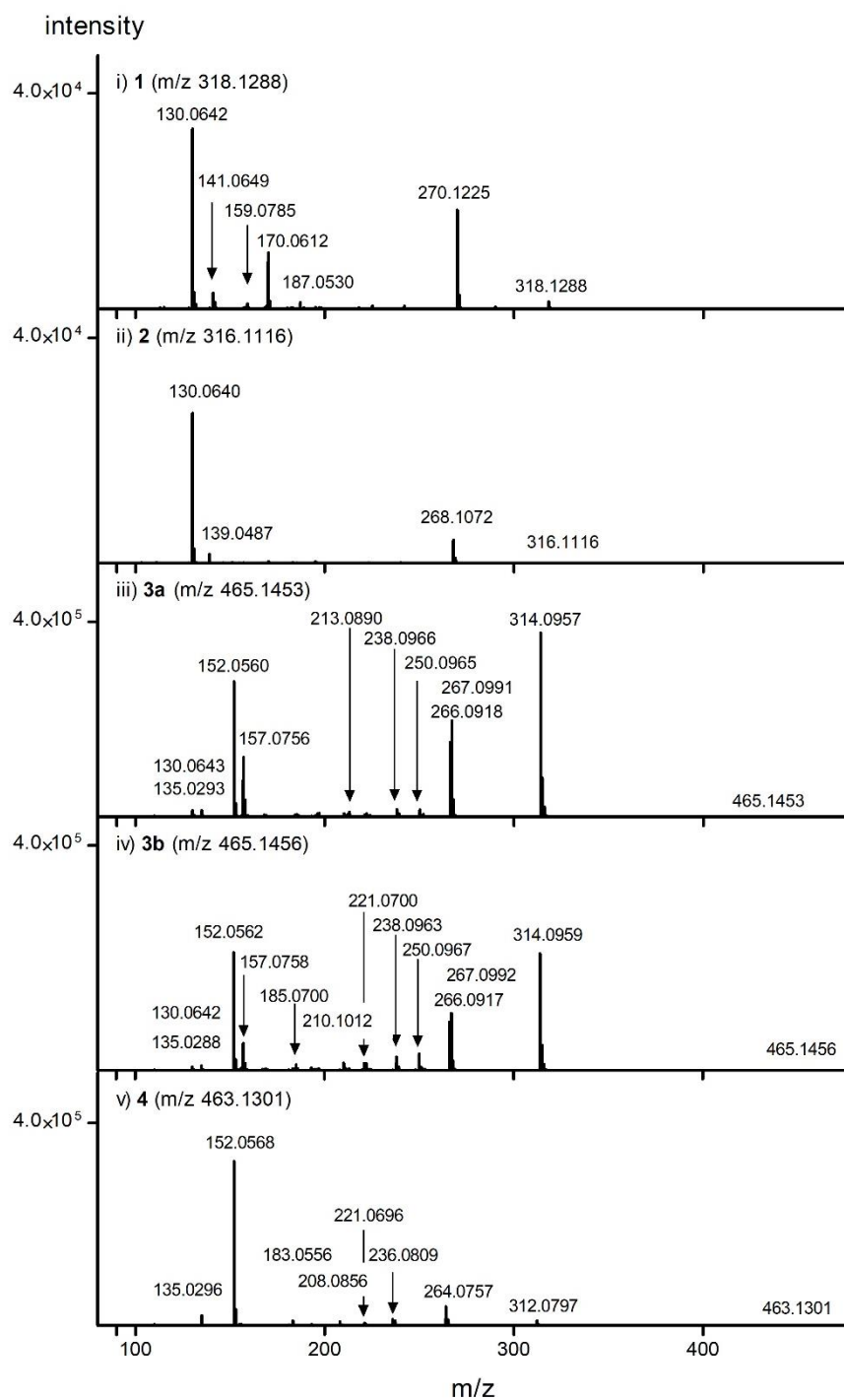
**Figure S53.**  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of guatrypmethine C (4) in  $\text{DMSO}-d_6$



**Figure S54.**  $^1\text{H}$ - $^{15}\text{N}$  HMBC spectrum of guatrypmethine C (4) in  $\text{DMSO}-d_6$

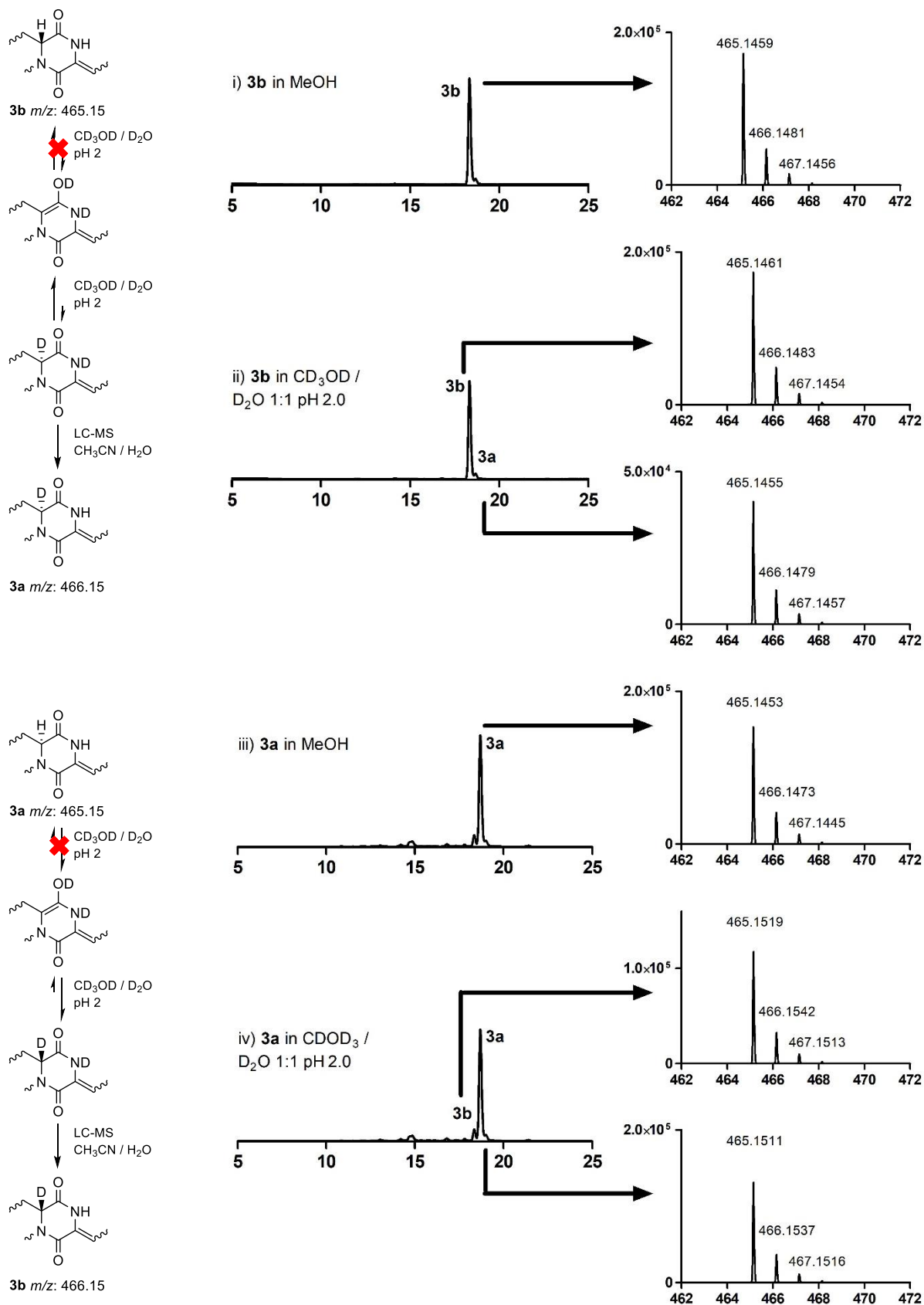


**Figure S55.** NOESY spectrum of guatrypmethine C (**4**) in DMSO- $d_6$

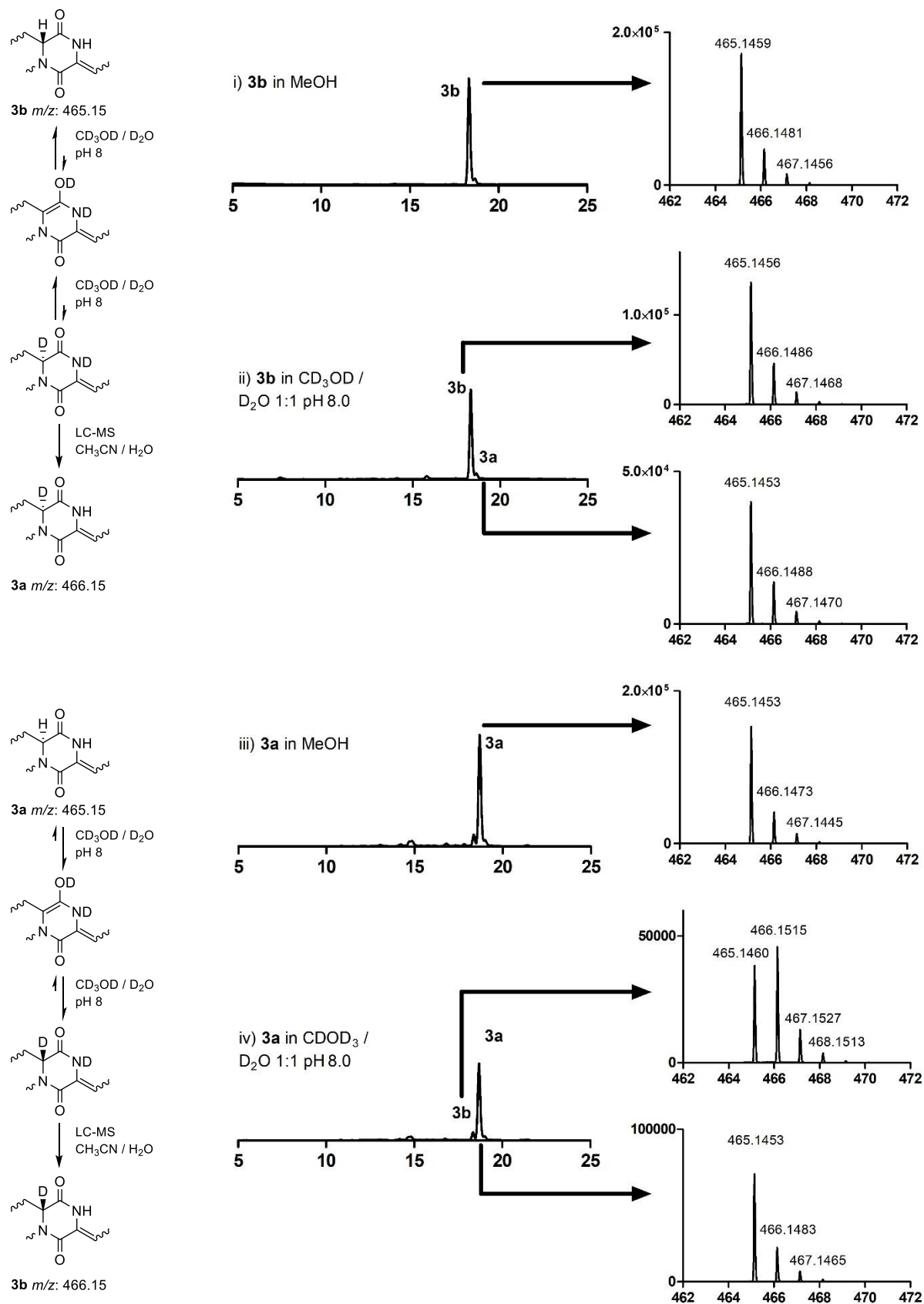


**Figure S56.** MS-MS of  $[M+H]^+$  ions of the identified products

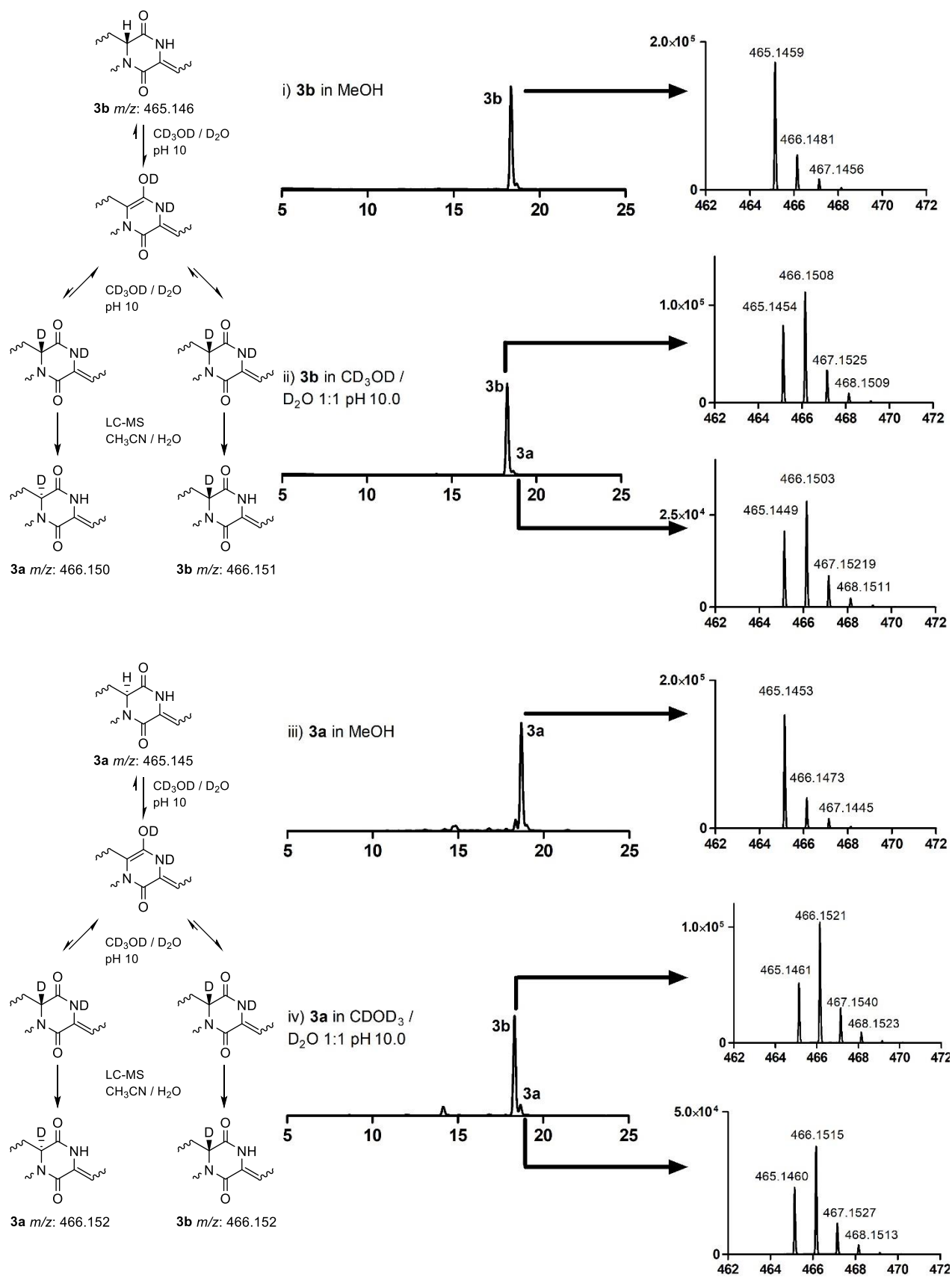
# PUBLICATIONS



**Figure S57.** LC-MS data of **3a** and **3b** after incubation in  $CD_3OD / D_2O$  1:1 at pH 2.0 for 16 h

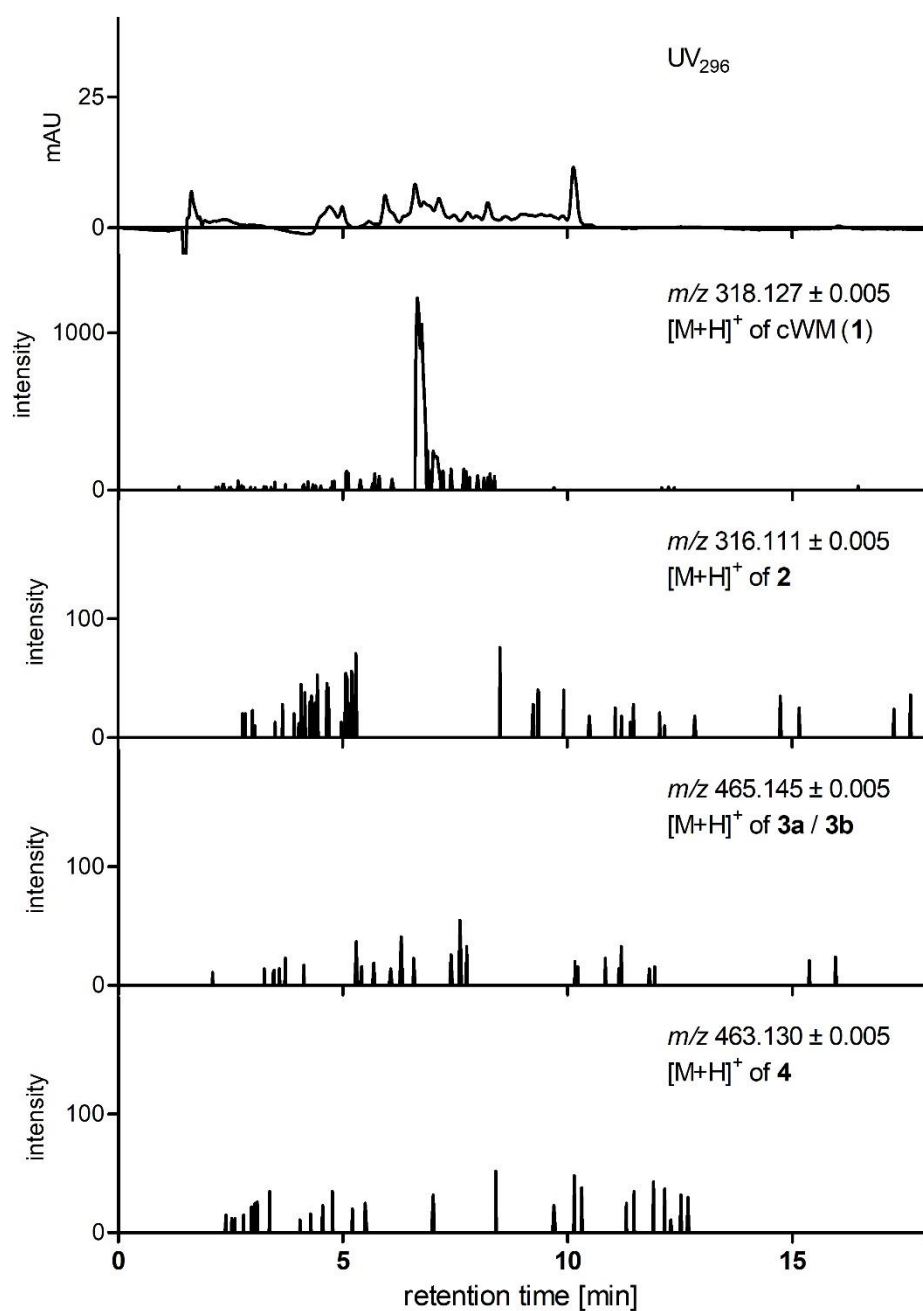


**Figure S58.** LC-MS data of **3a** and **3b** after incubation in  $\text{CD}_3\text{OD} / \text{D}_2\text{O}$  1:1 at pH 8.0 for 16 h

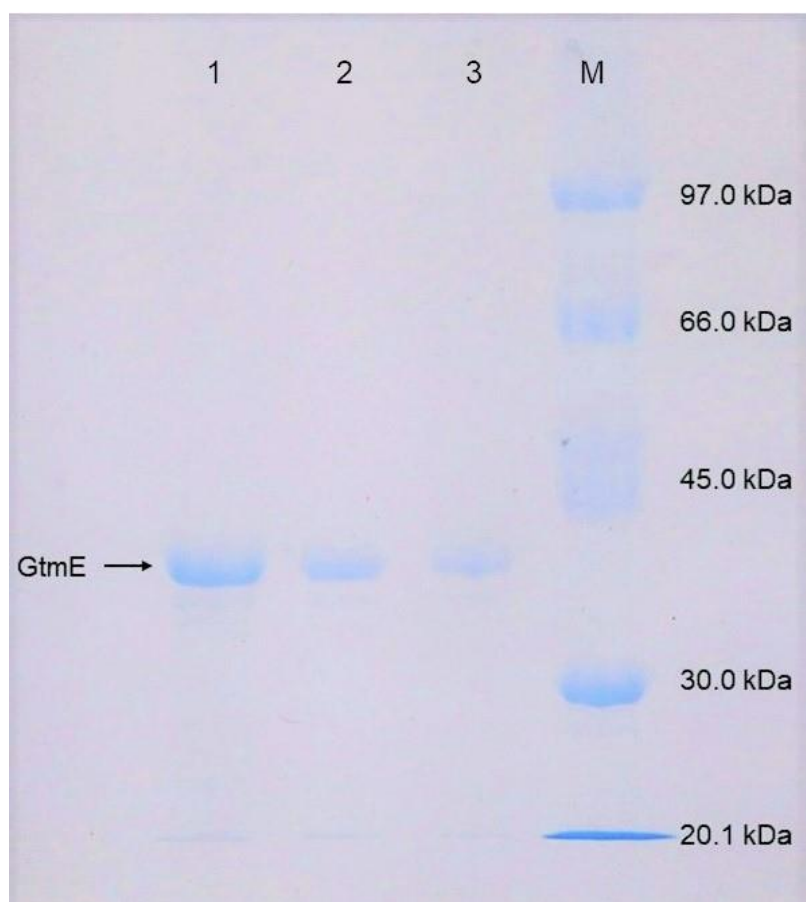


**Figure S59.** LC-MS data of **3a** and **3b** after incubation in  $\text{CD}_3\text{OD} / \text{D}_2\text{O}$  1:1 at pH 10.0 for 16 h

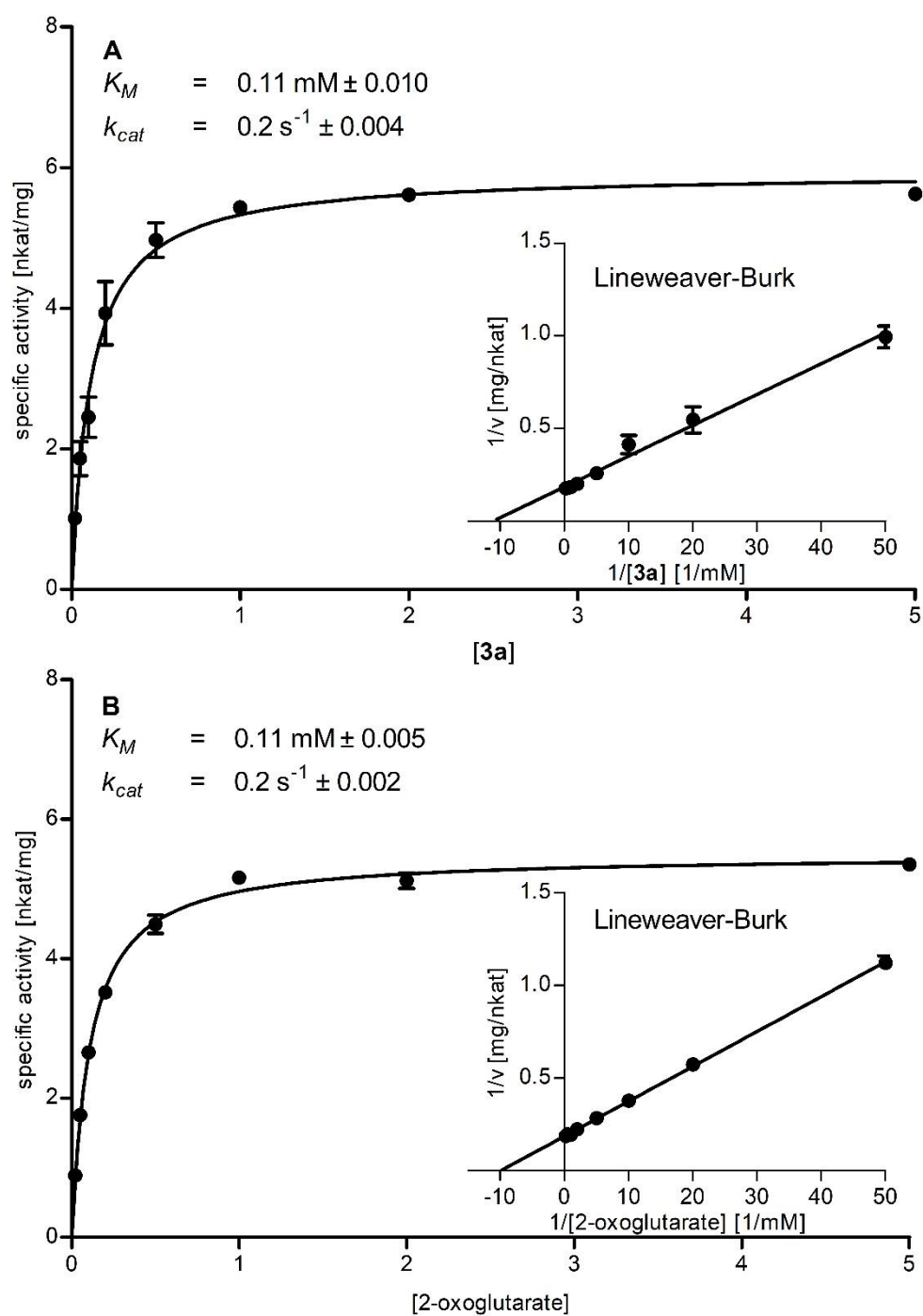




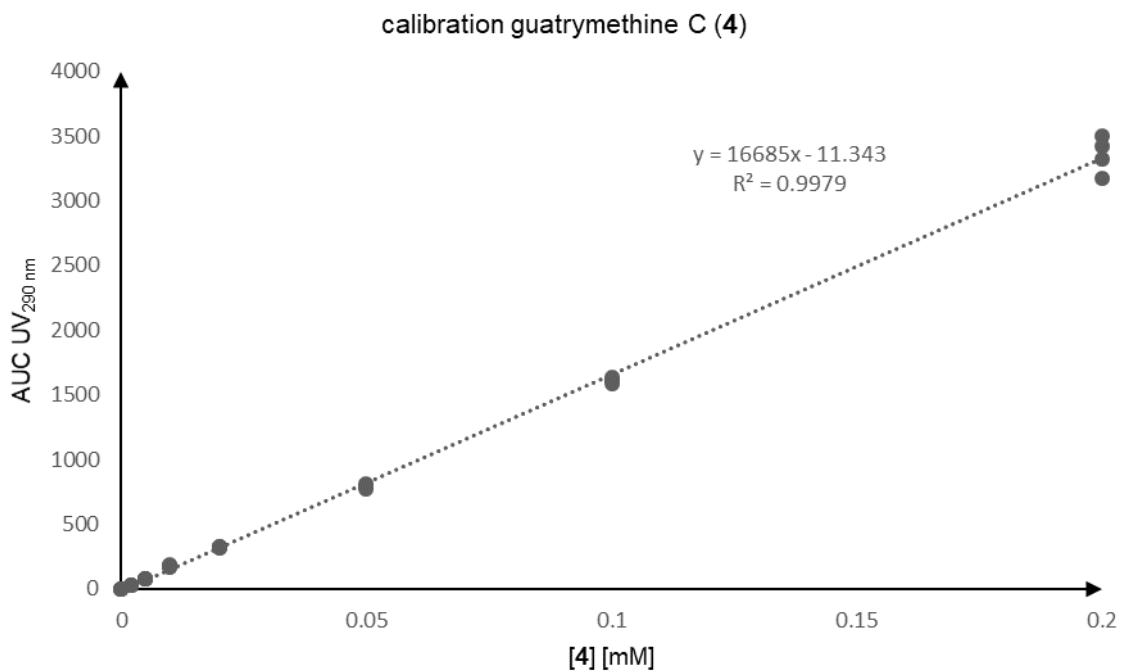
**Figure S60.** LC-MS chromatograms of the crude extract from *Streptomyces cinnamoneus* DSM 40646 with detection at UV<sub>296nm</sub> and extracted ions for compounds **1**, **2**, **3a**, **3b** and **4**.



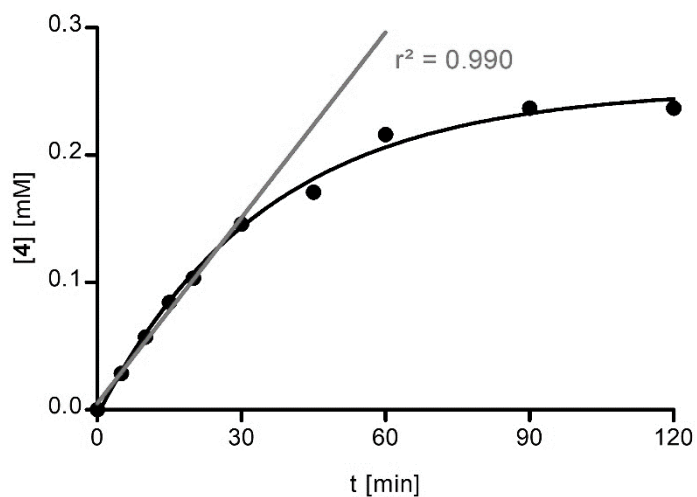
**Figure S61.** 12% SDS PAGE of the purified GtmE stained with coomassie brilliant blue R-250® lanes 1-3: GtmE in different concentrations, M: Amersham low molecular weight calibration kit



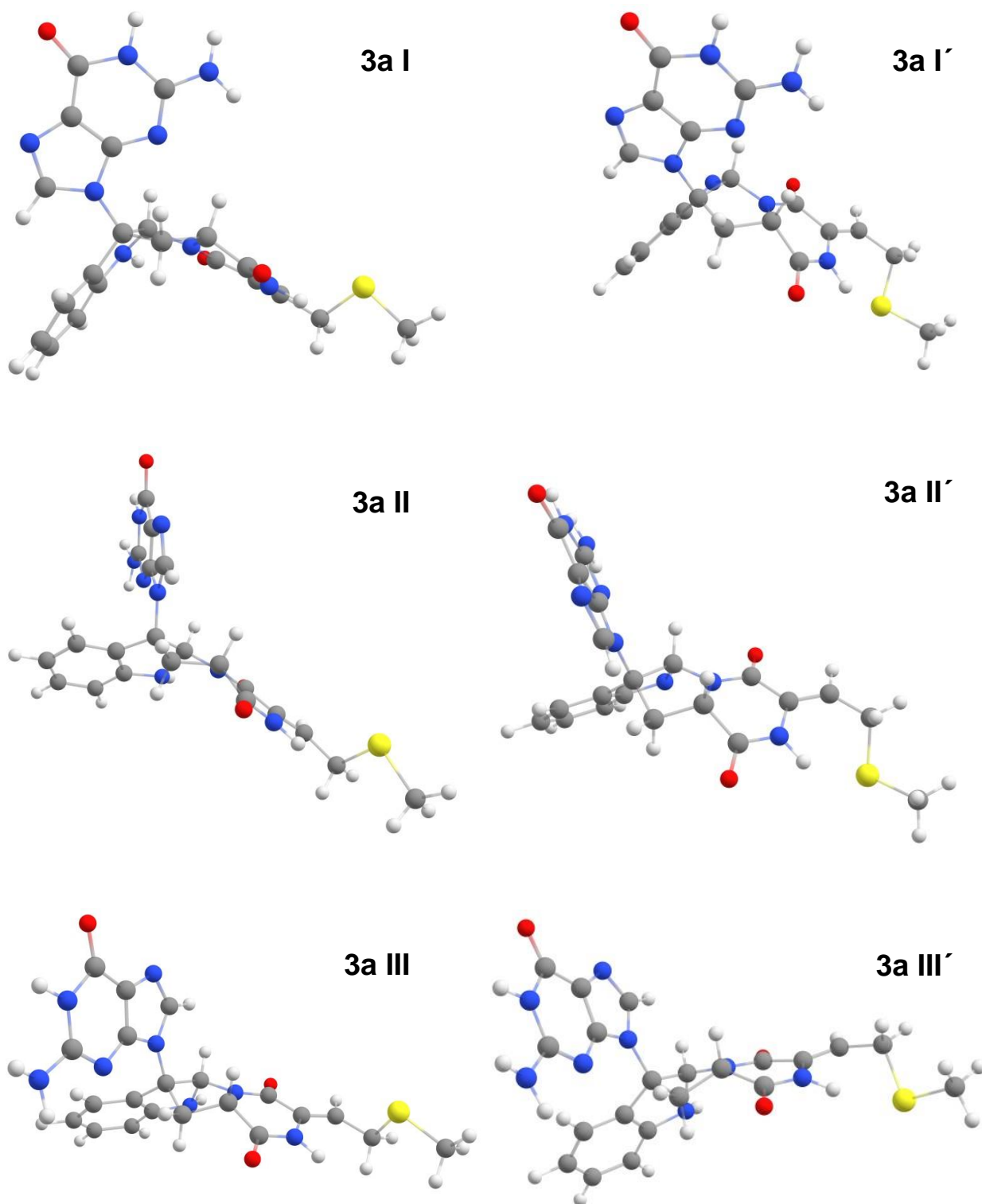
**Figure S62.** Determination of the kinetic parameters of the GtmE reaction for **3a** (A) and 2-oxoglutarate (B)



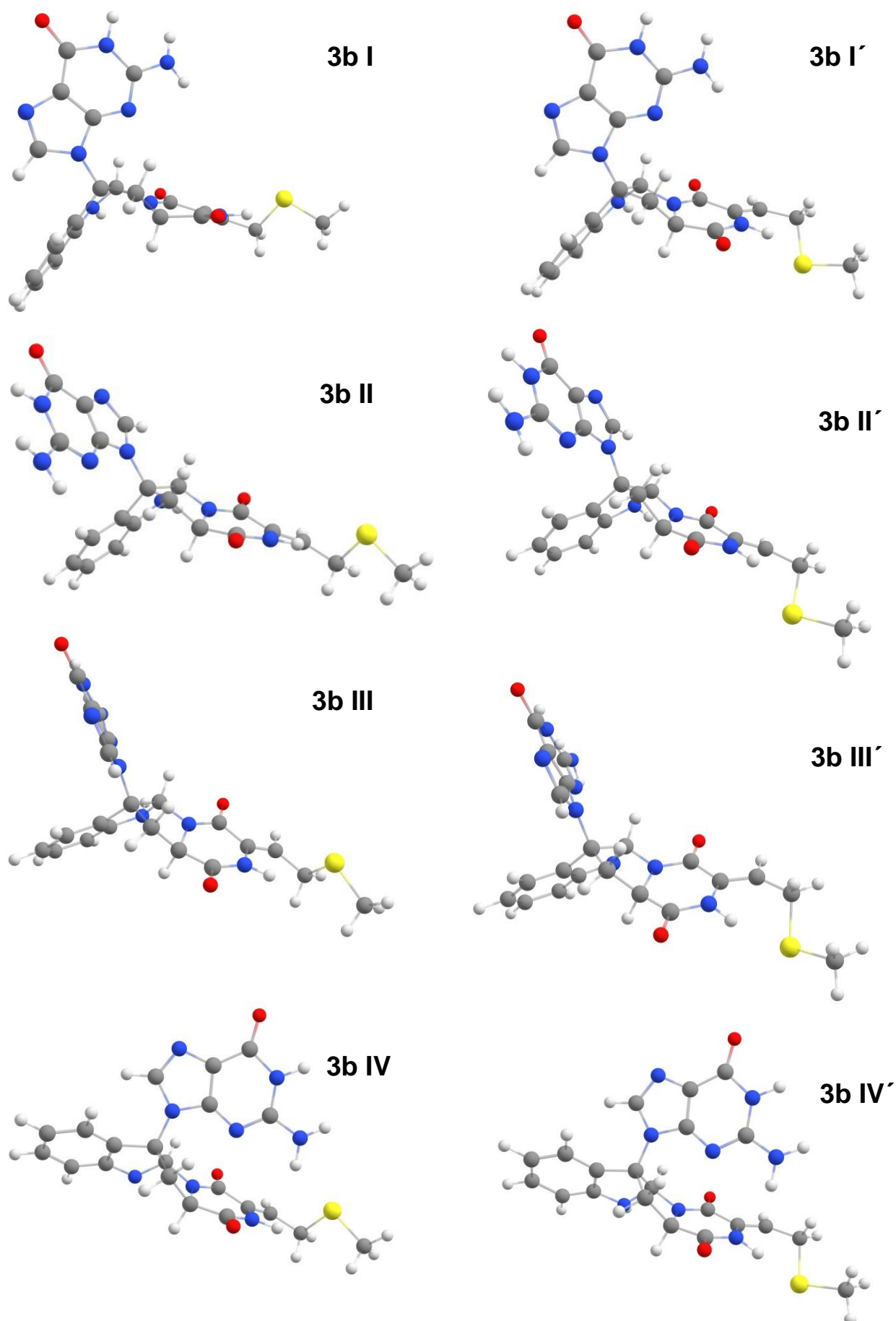
**Figure S63.** Calibration curve of different concentrations of **4**. The data represent four independent measurements.



**Figure S64.** Linearity test of the GtmE catalyzed desaturation from 0 – 120 min. In read is a linear regression curve for the values from 0 – 30 min with  $r^2 = 0.990$ , confirming that the reaction is linear until 30 min.



**Figure S65.** Energy minimized structures of **3a**



**Figure S66.** Energy minimized structures of **3b**

## References

- (1) Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; Van Wezel, G. P.; Medema, M. H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35.
- (2) Madeira, F.; Park, Y. M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641.
- (3) Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549.
- (4) Robert, X. and Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, *42*, W320–W324.
- (5) Zaburannyi, N.; Rabyk, M.; Ostash, B.; Fedorenko, V.; Luzhetskyy, A. Insights into naturally minimised *Streptomyces albus* J1074 genome. *Bmc Genomics* **2014**, *15*, 97
- (6) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces Genetics*; 2nd ed.; John Innes Foundation: Norwich, UK, 2000.
- (7) Liu, J.; Xie, X.; Li, S.-M. Guanitrypmycin biosynthetic pathways imply cytochrome P450-mediated regio- and stereospecific guaninyl transfer reactions. *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, 11534–11540.
- (8) Shi, J.; Xu, X.; Zhao, E. J.; Zhang, B.; Li, W.; Zhao, Y.; Jiao, R. H.; Tan, R. X.; Ge, H. M. Genome mining and enzymatic total biosynthesis of purincyclamide. *Org. Lett.* **2019**, *21*, 6825–6829.
- (9) Meng, S.; Han, W.; Zhao, J.; Jian, X. H.; Pan, H. X.; Tang, G. L. A six-oxidase cascade for tandem C-H bond activation revealed by reconstitution of bicyclomycin biosynthesis. *Angew. Chem Int. Ed. Engl.* **2018**, *57*, 719–723.
- (10) Roach, P. L.; Clifton, I. J.; Hensgens, C. M. H.; Shibata, N.; Schofield, C. J.; Hajdu, J.; Baldwin, J. E. Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **1997**, *387*, 827–830.
- (11) Wang, Y. and Li, R. Cloning and sequencing the isopenicillin N synthetase(IPNS) gene from *Streptomyces cattleya*. *Wei Sheng Wu Xue. Bao.* **1996**, *36*, 87–92.
- (12) Green, M. R.; Sambrook, J. *Molecular cloning: a laboratory manual*; 4th ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor: New York, 2012.
- (13) Zhu, Y.; Fu, P.; Lin, Q.; Zhang, G.; Zhang, H.; Li, S.; Ju, J.; Zhu, W.; Zhang, C. Identification of caerulomycin A gene cluster implicates a tailoring amidohydrolase. *Org Lett.* **2012**, *14*, 2666–2669.
- (14) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A* **2003**, *100*, 1541–1546.
- (15) Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **1995**, *6*, 135–140.
- (16) Tian, W.; Sun, C.; Zheng, M.; Harmer, J. R.; Yu, M.; Zhang, Y.; Peng, H.; Zhu, D.; Deng, Z.; Chen, S. L.; Mobli, M.; Jia, X.; Qu, X. Efficient biosynthesis of heterodimeric C<sup>3</sup>-aryl pyrroloindoline alkaloids. *Nat. Commun.* **2018**, *9*, 4428.
- (17) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (18) Weigend, F. and Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (19) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.
- (20) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822 (R)
- (21) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283–290.

## PUBLICATIONS

- (22) Klamt, A. and Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2* **1993**, 799–805.
- (23) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, 132, 154104
- (24) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, 32, 1456–1465.
- (25) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, 37, 785.
- (26) Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **1993**, 98, 1372–1377.
- (27) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, 98, 11623–11627.
- (28) Kesharwani, M. K.; Brauer, B.; Martin, J. M. L. Frequency and zero-point vibrational energy scale factors for double-hybrid density functionals (and other selected methods): Can anharmonic force fields be avoided? *J. Phys. Chem. A* **2015**, 119, 1701–1714.



**4.2 Widely Distributed Bifunctional Bacterial Cytochrome P450 Enzymes Catalyze both Intramolecular C-C Bond Formation in cyclo-L-Tyr-L-Tyr and Its Coupling with Nucleobases**

## PUBLICATIONS

## Biosynthesis

How to cite:

International Edition: doi.org/10.1002/anie.202200377

German Edition: doi.org/10.1002/ange.202200377

# Widely Distributed Bifunctional Bacterial Cytochrome P450 Enzymes Catalyze both Intramolecular C–C Bond Formation in *cyclo*-L-Tyr-L-Tyr and Its Coupling with Nucleobases

Jing Liu<sup>+</sup>, Lauritz Harken<sup>+</sup>, Yiling Yang, Xiulan Xie, and Shu-Ming Li\*

**Abstract:** Tailoring enzymes are important modification biocatalysts in natural product biosynthesis. We report herein six orthologous two-gene clusters for mycocyclusin and guatyromycine biosynthesis. Expression of the cyclodipeptide synthase genes *gymA<sub>1</sub>*–*gymA<sub>6</sub>* in *Escherichia coli* resulted in the formation of *cyclo*-L-Tyr-L-Tyr as the major product. Reconstruction of the biosynthetic pathways in *Streptomyces albus* and biochemical investigation proved that the cytochrome P450 enzymes GymB<sub>1</sub>–GymB<sub>6</sub> act as both intramolecular oxidases and intermolecular nucleobase transferases. They catalyze not only the oxidative C–C coupling within *cyclo*-L-Tyr-L-Tyr, leading to mycocyclusin, but also its connection with guanine and hypoxanthine, and are thus responsible for the formation of tyrosine-containing guatyromycines, instead of the reported tryptophan-nucleobase adducts. Phylogenetic data suggest the presence of at least 47 GymB orthologues, indicating the occurrence of a widely distributed enzyme class.

## Introduction

2,5-Diketopiperazines (DKPs) and their derivatives are widespread in nature and represent an important class of secondary metabolites.<sup>[1,2]</sup> They display a broad spectrum of biological and pharmacological activities, such as antibacterial, antifungal, antitumor, anti-plasmodial, and anti-Alzheimer properties.<sup>[3–8]</sup> Apart from these effects, they are also involved in quorum sensing and might be used in future as

drug delivery systems due to a high cell penetration.<sup>[9–11]</sup> All these features make them attractive molecules for drug discovery and development. In nature, the heterocyclic DKP core of cyclodipeptides (CDPs) is usually formed by condensation of two  $\alpha$ -amino acids in two manners. This reaction is either catalyzed by nonribosomal peptide synthetases (NRPSs), mainly found in fungi, or by cyclodipeptide synthases (CDPSs), commonly occurring in bacteria.<sup>[12–14]</sup> In comparison to NRPSs, CDPSs are small proteins of approximate 30 kDa and use activated aminoacyl-tRNAs from the ribosomal machinery as substrates.<sup>[15]</sup> The formation of a DKP as backbone exhibits an increased stability against proteolysis compared to acyclic dipeptides, which enables tailoring enzymes to modify the heterocycle or the side chains.<sup>[16]</sup> In the CDPS-dependent pathways, such enzymes include cytochrome P450 enzymes, cyclodipeptide oxidases, Fe<sup>II</sup>/2-oxoglutarate-dependent oxygenases, prenyltransferases, methyltransferases, and terpene cyclases.<sup>[17–22]</sup>

Members of the P450 enzyme superfamily occur ubiquitously in living organisms and play vital roles in the metabolism of xenobiotics as well as the biosynthesis of natural products.<sup>[23–25]</sup> Due to different lengths of the polypeptide chains, shapes of the binding pockets, and electron transfer partners, they can catalyze a large variety of transformations.<sup>[25]</sup> The catalytic activities of P450 enzymes in CDPS-associated pathways include hydroxylation, e.g. BcmD in the formation of bicyclomycin,<sup>[18,26]</sup> multistep oxidation as CypX in the biosynthesis of pulcherri-minic acid,<sup>[27]</sup> dimerization in the production of naseazine C and tetratryptomycins,<sup>[28–30]</sup> the formation of an intramolecular C–C bond by CYP121 in the mycocyclusin biosynthesis,<sup>[31]</sup> and transfer of nucleobases, e.g. guanine and hypoxanthine in the biosynthesis of guanitryptomycins, purin-cyclamide, and guatryptmethines (Figure 1A).<sup>[32–34]</sup>

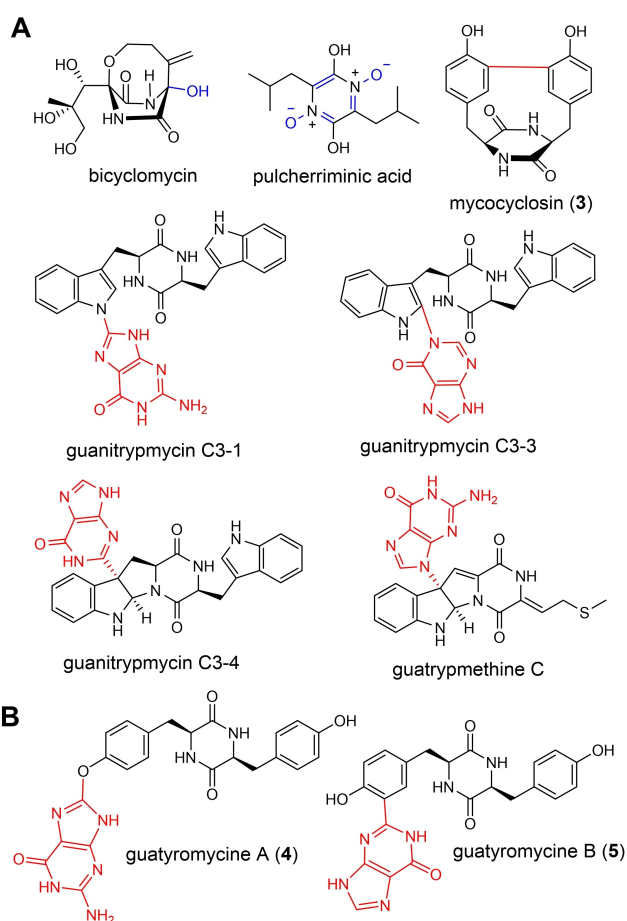
Bioinformatic analysis revealed the presence of a widely distributed family of two-gene clusters from *Streptomyces* species, termed *gym* clusters, each coding for a CDPS GymA and a P450 oxidase GymB. Herein, we report the identification of six of these clusters by heterologous expression in *E. coli* and *Streptomyces albus* (*S. albus*) as well as by biochemical characterization. Our results demonstrated that GymA<sub>1</sub>–GymA<sub>6</sub> assemble *cyclo*-L-Tyr-L-Tyr (cYY, **1**) as main and *cyclo*-L-Tyr-L-Phe (cYF, **2**) as side product. The P450 enzymes GymB<sub>1</sub>–GymB<sub>6</sub> catalyze the formation of mycocyclusin (**3**) by oxidative intramolecular C–C coupling and the transfer of a nucleobase, guanine or hypoxanthine, onto different positions of **1** for the formation

[\*] Dr. J. Liu,<sup>+</sup> L. Harken,<sup>+</sup> Y. Yang, Prof. S.-M. Li  
 Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg  
 Robert-Koch-Straße 4, 35037 Marburg (Germany)  
 E-mail: shuming.li@staff.uni-marburg.de

Dr. X. Xie  
 Fachbereich Chemie, Philipps-Universität Marburg  
 Hans-Meerwein-Straße 4, 35032 Marburg (Germany)

[†] These authors contributed equally to this work.

© 2022 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

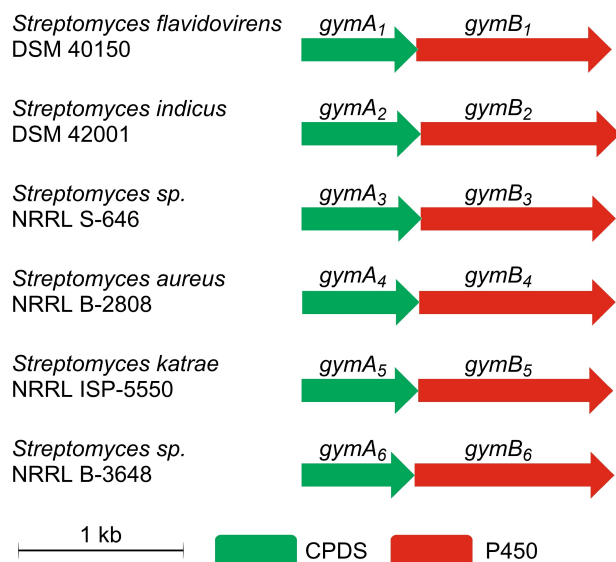


**Figure 1.** Known (A) and new (B) secondary metabolites from biosynthetic pathways containing CDPS and P450 enzymes. P450 enzymes catalyze novel CDP modifications including hydroxylation/oxidation (depicted in blue), formation of the intramolecular C–C bond in mycrocyclosin and the nucleobase transfer reactions in guanitrypmycins and guatrypmethines (depicted in red).

of guatyromycines A (4) and B (5, Figure 1B), respectively. Thus, GymBs represent the first P450 members in CDPS-related pathways that can catalyze both intra- and intermolecular coupling reactions.

## Results and Discussion

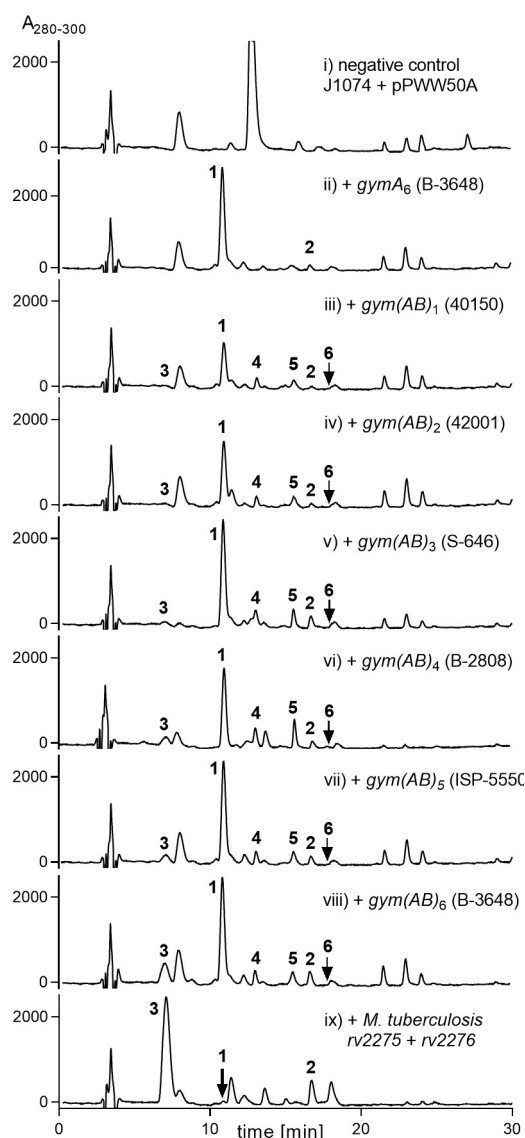
In recent years, the number of microbial genomic sequences in the public databases, especially those of bacteria and fungi, has increased exponentially.<sup>[35]</sup> The available sequence data made us aware that the number and diversity of metabolites from a candidate organism is far greater than those reported. Among the tremendous number of putative gene clusters containing *cdps* genes, only a dozen CDPS-related biosynthetic pathways have been characterized yet, implying that a great potential still remains unexplored.<sup>[15,17,19]</sup> In order to identify more *cdps*-P450-containing gene clusters, we used the sequences of the functionally characterized CDPSs as probes to search for



**Figure 2.** Genetic organization of the *gym* clusters from six *Streptomyces* species.

their homologs by using BLAST® and analyzed their neighboring genes coding for tailoring enzymes, especially for P450 enzymes. Phylogenetic analysis of known and putative CDPS and associated P450 sequences indicated the presence of a large group of unknown *cdps*-P450 gene clusters (Figures S1 and S2). Phylogenetic analysis also revealed that the reported P450 enzymes responsible for nucleobase transfer and CDP dimerization are clustered in two distinct clades (Figure S2). It is noteworthy that CYP121, which catalyzes the intramolecular oxidative C–C coupling of *cyclo*-L-Tyr-L-Tyr,<sup>[31]</sup> is located closest to the largest clade containing 47 unknown P450 homologs. Thus, we selected one representative from each of the subclades, in total six different *cdps*-P450 gene clusters, for detailed investigation. These clusters were identified in *Streptomyces flavidovirens* DSM 40150 (*gymA*<sub>1</sub>*B*<sub>1</sub>), *Streptomyces indicus* DSM 42001 (*gymA*<sub>2</sub>*B*<sub>2</sub>), *Streptomyces sp.* NRRL S-646 (*gymA*<sub>3</sub>*B*<sub>3</sub>), *Streptomyces aureus* NRRL B-2808 (*gymA*<sub>4</sub>*B*<sub>4</sub>), *Streptomyces katrae* NRRL ISP-5550 (*gymA*<sub>5</sub>*B*<sub>5</sub>), and *Streptomyces sp.* NRRL B-3648 (*gymA*<sub>6</sub>*B*<sub>6</sub>) (Figure 2, Table S2). These candidate P450s share approximately 60 % sequence identities with CYP121 encoded by Rv2276 on the amino acid sequence level (Table S1).

Alignments of the six CDPSs from the selected clusters with AlbC from *Streptomyces noursei* (Figure S3)<sup>[36]</sup> and prediction of their substrate specificities<sup>[14,37]</sup> revealed that both binding pockets are specific for L-Tyr-tRNA, so that *cyclo*-L-Tyr-L-Tyr can be expected as their product (Table S3). To investigate the *cdps* function, the coding sequences of *gymA*<sub>1</sub>-*gymA*<sub>6</sub> were cloned into pET28a (+) for heterologous expression in *E. coli* BL21 (DE3) (Tables S2 and S4). LC-MS analysis of the bacterial cultures bearing the *gymA* genes revealed similar metabolite profiles and the presence of one predominant product with a [*M* + *H*]<sup>+</sup> ion at *m/z* 327.134 ± 0.005 and a minor peak with a [*M* + *H*]<sup>+</sup> ion at *m/z* 311.139 ± 0.005 (Figure S4). Comparison of



**Figure 3.** HPLC analysis of the extracts from *S. albus* transformants. The dominant peak at 13 min in negative control was strongly reduced in the transformants, as observed in other expression studies.<sup>[22,34]</sup> Therefore, it is unlikely related to the introduced *gym* genes.

the retention times, MS<sup>2</sup> data, UV-spectra (Figure S5), and NMR data (Table S6, Figures S8 and S9) with authentic standards led to the identification of predicted *cyclo*-L-Tyr-L-Tyr (cYY, **1**) as the main product (Table S3) and *cyclo*-L-Tyr-L-Phe (cYF, **2**) as the minor one. Production of two CDPs by one CDPS is in agreement with the published substrate promiscuity of other characterized enzymes.<sup>[38,39]</sup>

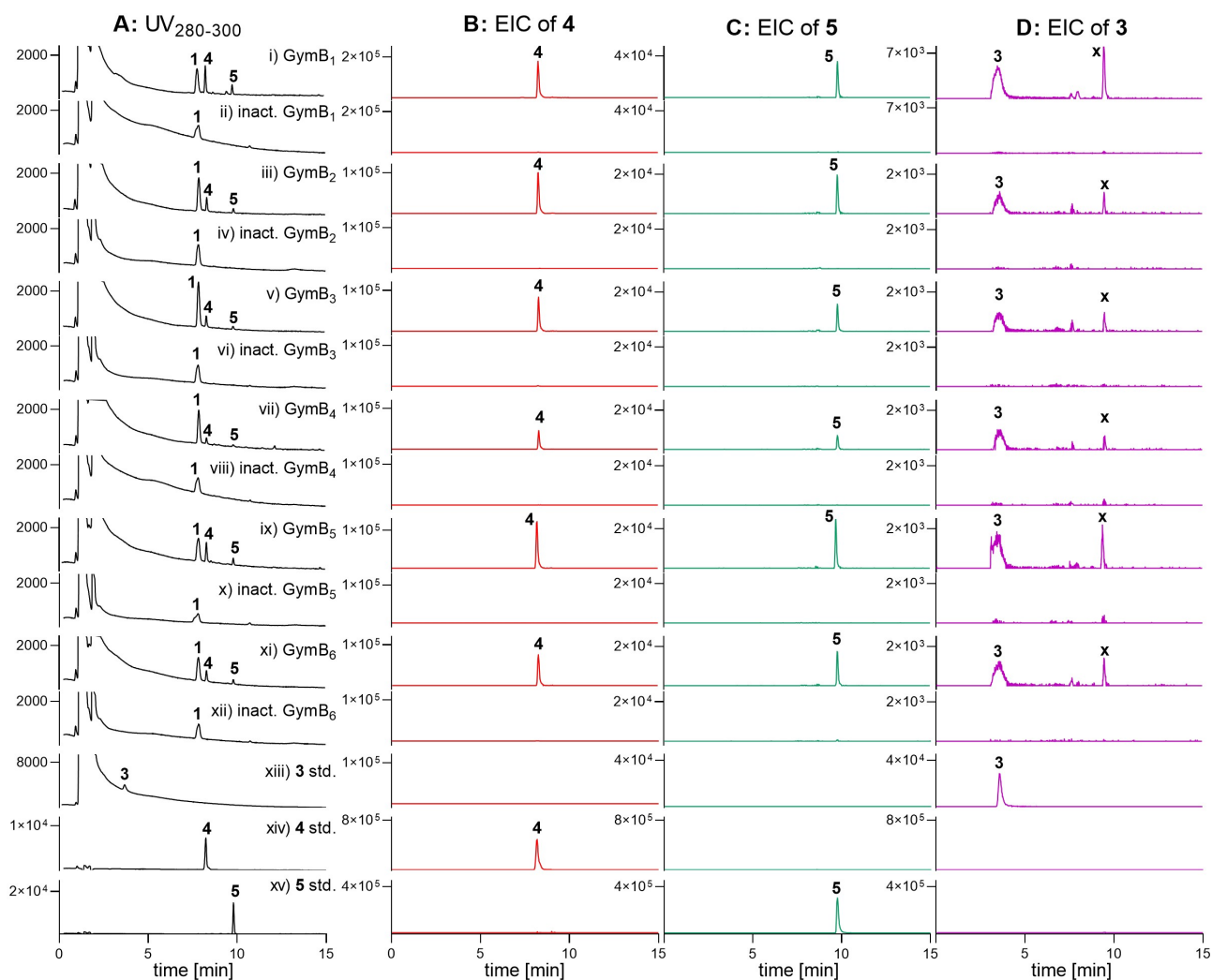
Heterologous expression of microbial biosynthetic gene clusters in well-investigated host strains has proven to be an effective approach to discover new natural products.<sup>[40]</sup> Having determined the function of the backbone enzymes GymA<sub>1</sub>–GymA<sub>6</sub> as cYY and cYF synthases, we decided to heterologously express the two-gene clusters in *S. albus* J1074 to identify the cluster products.<sup>[41]</sup> Since GymB<sub>6</sub> was located closest to CYP121 in the phylogenetic tree (Fig-

ure S2), the *gymA<sub>6</sub>B<sub>6</sub>* cluster from *Streptomyces* sp. NRRL B-3648 was taken as the first candidate. We first reproduced the expression of *gymA<sub>6</sub>* in J1074 under control of the constitutive *ermEp\** promotor by using the replicative vector pPWW50 A.<sup>[42]</sup> The transformant was cultivated in MR5 media (Table S2) and extracted with EtOAc. LC-MS analysis revealed that, in comparison to the negative control with pPWW50 A (Figure 3i), the *gymA<sub>6</sub>* transformant produced both **1** and **2** as CDPS products (Figure 3ii), corresponding very well to those identified by expression in *E. coli* (Figure S4).

The *gymA<sub>6</sub>B<sub>6</sub>* cluster was then cloned into pPWW50 A and expressed in *S. albus*. LC-MS analysis of the cultural extract displayed three new peaks (Figure 3viii) with distinct UV absorptions (Figure S5). Peak **3** has a  $[M+H]^+$  ion at  $m/z$  325.1180, corresponding well to that of mycrocyclosin.<sup>[31]</sup> The  $[M+H]^+$  ion of **4** ( $m/z$  476.1840) is 149 Da larger than that of **1**, suggesting an additional guaninyl moiety on it. The  $[M+H]^+$  ion of **5** at  $m/z$  461.1719 is 134 Da larger than that of **1**, indicating the attachment of a hypoxanthinyl residue.

In order to elucidate the structures of the newly accumulated products, analytically pure compounds **3–5** were obtained by semi-preparative HPLC and subjected to NMR analysis. Comparison of the <sup>1</sup>H NMR spectrum with that published previously supported **3** to be mycrocyclosin (see Figure 1A for structure, Table S6, Figure S10).<sup>[31]</sup> Detailed inspection of the NMR data of **4**, including <sup>1</sup>H, <sup>13</sup>C, <sup>1</sup>H–<sup>1</sup>H COSY, HSQC, and HMBC (Table S7, Figures S11–S15, see Supporting Information for detailed structure elucidation), suggested a cYY derivative with a guanine residue attached to a hydroxyl group (Figure 1B). Structure elucidation of **5** confirmed our assumption that a hypoxanthine is transferred via its C-2' to C-10 of the cYY residue. (Table S8, Figures S16–S20). Cultivation of *S. albus* transformant harboring *gymA<sub>6</sub>B<sub>6</sub>* in medium with <sup>15</sup>NH<sub>4</sub>Cl and subsequent LC-MS analysis of the cultural extract demonstrated that at least two <sup>15</sup>N atoms are incorporated into in **3**, seven in **4**, and six in **5** (Figure S21), confirming the structures of **3**, **4**, and **5** (Figure 1). Based on the fact that **4** and **5** are derived from L-tyrosine and guanine/hypoxanthine, they are therefore termed as guatyromycines A and B, respectively.

In analogy to *gymA<sub>6</sub>B<sub>6</sub>* cluster from *Streptomyces* sp. NRRL B-3648, we performed co-expression of the genes for CDPS and P450 enzyme from the other five candidates (Figure 2) in *S. albus*. The plasmids (Table S4) were constructed as described in the Supporting Information and transferred into *S. albus* by conjugation. The extracts of the corresponding transformants were analyzed by LC-MS (Figure 3), which revealed the presence of very similar product profiles as described for NRRL B-3648. Together with their precursor **1** and the side product **2**, all three products **3**, **4**, and **5** were clearly observed by UV and EIC detections (Figure 3iii–3vii, Figure S6). Their structures were further confirmed by comparison of their retention times, UV-spectra, and MS data including MS<sup>2</sup> fragmentation pattern with those of **3**, **4**, and **5** isolated in this study. As compound **3** is also the product of CYP121 from *Mycobacterium tuberculosis*, we decided to express the responsible cluster



**Figure 4.** LC-MS chromatograms of cYY assays with GymB<sub>1</sub>–GymB<sub>6</sub> in the presence of guanine and hypoxanthine. EICs of **3**, **4**, and **5** refer  $[M+H]^+$  ions at  $m/z$  325.118, 476.184, 461.172 with a tolerance range of  $\pm 0.005$ . Peak **x**: unknown mycocyclusin isomer. inact.: heat-inactivated; std.: standard.

with two genes *rv2275* and *rv2276* as well.<sup>[31]</sup> In comparison, the cluster from *M. tuberculosis* mainly produced **3** with a product yield of  $53.9 \pm 2.8 \text{ mg L}^{-1}$  (Table S5), accompanied by **1** and **2** as minor peaks (Figure 3ix, Figure S6), confirming that CYP121 merely catalyzes the formation of **3**.<sup>[31]</sup> Quantification of the productivity of the different transformants (Table S5) showed that **4** and **5** were accumulated in comparable product yields of  $2.3\text{--}5.5 \text{ mg L}^{-1}$ , while the contents of **3** varied significantly from  $1.6 \pm 0.2 \text{ mg L}^{-1}$  for *gymA<sub>2</sub>B<sub>2</sub>* to  $24.0 \pm 1.3 \text{ mg L}^{-1}$  for *gymA<sub>6</sub>B<sub>6</sub>*.

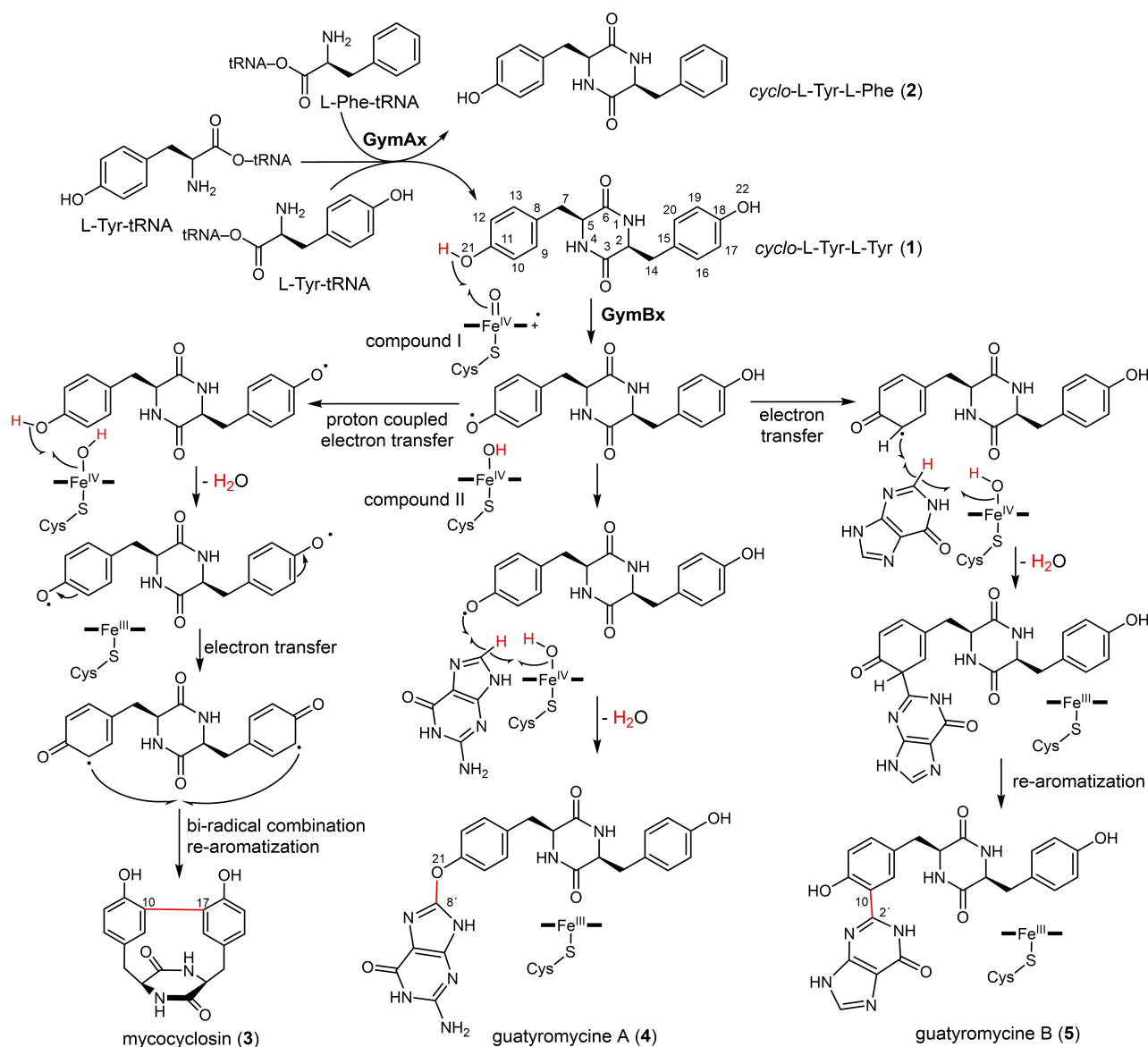
An intramolecular coupling product of cYF (**2**) or its adduct with hypoxanthine were detected neither by UV nor by  $[M+H]^+$  ion detection in all of the transformants described in this study (Figure S7). In contrast, a clear peak **6** at 17.8 min was detected with a  $m/z$  of  $460.173 \pm 0.005$ , corresponding to the  $[M+H]^+$  ion of **2** with guanine, in the transformants carrying *gymA<sub>1</sub>B<sub>1</sub>*–*gymA<sub>6</sub>B<sub>6</sub>*. However, this peak is so weak that it's almost not visible by UV detection

(Figure 3). Due to the low quality, its structure could not be elucidated in this study.

From these results, it can be postulated that the CDPSS GymA<sub>1</sub>–GymA<sub>6</sub> use tyrosyl-tRNA and phenylalanyl-tRNA as substrates to assemble cYY (**1**) as the major and cYF (**2**) as the minor product. cYY serves then as the substrate of the cytochrome P450 enzymes GymB<sub>1</sub>–GymB<sub>6</sub> for intramolecular oxidative C–C bond formation and coupling with guanine and hypoxanthine via different atoms and positions of cYY, resulting in mycocyclusin (**3**), guatyromycines A (**4**) and B (**5**), respectively (Figure 1). It is likely that **2** was also used by GymB<sub>1</sub>–GymB<sub>6</sub> as the substrate for coupling with guanine.

To prove the activities of GymB<sub>1</sub>–GymB<sub>6</sub> in vitro, their coding sequences were cloned into pET28a (+) for overproduction in *E. coli* BL21 (DE3) (Table S4). The recombinant His<sub>6</sub>-tagged GymB<sub>1</sub>–GymB<sub>6</sub> were purified on Ni-NTA agarose (Figure S22) and used for in vitro assays with **1**, guanine, and hypoxanthine in the presence of ferredoxin,





**Figure 5.** Biosynthetic pathway of **3**, **4**, and **5** and mechanism of the GymB-catalyzed reactions.

ferredoxin-reductase, and NADPH. Heat-inactivated proteins served as negative controls. After incubation at 30 °C for 16 h, the reactions were quenched with MeOH and analyzed by LC-MS. As shown in Figure 4, **4** was detected in all the assays with active enzymes as the major product, followed by **5** and **3**. These results confirmed unequivocally that GymB<sub>1</sub>–GymB<sub>6</sub> catalyze both the nucleobase transfer reactions and the intramolecular C–C bond coupling. They act therefore as bi-functional enzymes. In the absence of guanine and hypoxanthine, GymB<sub>1</sub>–GymB<sub>6</sub> converted **1** to **3** as the sole product with much higher yields than with nucleobases (Figure S23). A peak “x” at 9.5 min was detected in EIC for **3** of some assays with and without nucleobases (Figures 4 and S23), indicating the presence of a mycrocyclosin isomer. Due to the trace product amount, its structure could not be determined in this study.

It can be postulated that the abstraction of one hydrogen from OH at C-11 by the Fe<sup>IV</sup>=O species (compound I) leads to the formation of a radical at O-21, which acts as a central intermediate for the formation of mycrocyclosin (**3**)<sup>[43]</sup> and guatyromycines A (**4**) and B (**5**) (Figure 5). Transfer of a hydrogen from the second hydroxyl group to Fe<sup>IV</sup>–OH (compound II) and release of one molecule H<sub>2</sub>O will result in the formation of a diradical intermediate. Electron migration to C-11 and C-17, radical combination and subsequent rearomatization will release product **3** from the GymBx template. Direct reaction of the radical at O-21 with C-8 of guanine, transfer of one hydrogen to Fe<sup>IV</sup>–OH and water elimination lead to the formation of guatyromycine A (**4**). In the case of the formation of guatyromycine B (**5**), the radical at O-21 would first undergo electron migration, resulting in a radical at C-10, which couples with hypoxanthine at its C-2'. Water elimination and rearomatization

complete the formation of **5** (Figure 5). The Fe<sup>III</sup> species will be reduced to Fe<sup>II</sup> by reduction partner ferredoxin/ferredoxin reductase/NADPH and is ready for binding of O<sub>2</sub> for the next reaction cycle.<sup>[25,44]</sup> It seems that C-2' in hypoxanthine is preferred to react with C-10 of cYY in analogy to the formation of mycocyclusin, which has been investigated in detail.<sup>[43]</sup> The connection of guanine with cYY is likely caused by the amino substitution at C-2' of the purine skeleton and a resulting steric hindrance in the binding pocket of the enzymes.

## Conclusion

In this study, we demonstrated by heterologous expression the presence of a widely distributed family of two-gene clusters coding for a CDPS and a cytochrome P450 enzyme in *actinobacteria* (Figures S1 and S2), which produce both mycocyclusin (**3**) by intramolecular oxidative C–C coupling of cYY (**1**) and guatyromycines A (**4**) and B (**5**) by transfer of a nucleobase onto different positions of **1**. Biochemical investigations by using recombinant P450 enzymes GymB<sub>1</sub>–GymB<sub>6</sub> proved the conversion of **1** to **3**, **4**, and **5**. To the best of our knowledge, the GymBs are the first described CDPS-associated P450s that function as unique bifunctional enzymes for both intramolecular coupling and nucleobase transfer onto a tyrosyl moiety, which is different from those reported previously.<sup>[31,32,34,45,46]</sup>

Cultivation of the six *gymAB*-bearing *Streptomyces* species given in Figure 2 in two different media and LC-MS analysis of the extracts did not lead to detection of any products mentioned above, indicating the presence of silent genes in the native strains (data not shown).

Mycocyclusin (**3**) has been proven to be the product of a gene cluster with a CDPS gene *rv2275* and the gene *rv2276* for the cytochrome P450 enzyme CYP121 from *M. tuberculosis*.<sup>[31]</sup> No adduct of **1** with nucleobases was reported. We confirmed in this study these results by heterologous expression of *rv2275/rv2276* in *S. albus*. A product yield of 53.9 mg mL<sup>−1</sup> was achieved for **3**. This is two- to twenty-fold of the transformants carrying *gym* clusters (Table S5). It could be speculated that *gymBs* and potential orthologues evolved from *rv2276*. Phylogenetic analysis revealed a coherence of the product yields for **3** in the respective transformants and their sequence proximity. For example, GymB<sub>4–6</sub> are located closely to CYP121 and the *S. albus* transformants with their clusters produced **3** at 9.4–24.0 mg L<sup>−1</sup>, significantly higher than 1.6 mg mL<sup>−1</sup> in the *gymA<sub>2</sub>B<sub>2</sub>* and 2.3 mg mL<sup>−1</sup> in the *gymA<sub>1</sub>B<sub>1</sub>* transformant.

In summary, this study demonstrates again the power of the combinational approaches of genome mining in the available databases, heterologous expression in well-established hosts and biochemical characterization with recombinant proteins for the discovery of new secondary metabolites and intriguing enzymes. It can be expected that more novel natural product structures and enzymes will be discovered in the future by unravelling the biosynthetic potential hidden behind the silent/cryptic genes in the tremendous available genome sequences.

## Acknowledgements

We thank the ARS Culture Collection (NRRL) for providing *Streptomyces* strains, Rixa Kraut, Lena Ludwig-Radtke, Dr. Regina Ortmann, and Stefan Newel (Marburg) for MS and NMR measurement. This project was funded in part by the Deutsche Forschungsgemeinschaft (DFG, INST 160/620-1 and Li844/14-1). Jing Liu (201608310118) and Yiling Yang (201808530447) are scholarship recipients from the China Scholarship Council. Open Access funding enabled and organized by Projekt DEAL.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Keywords:** Biosynthesis · Cyclodipeptide Synthase · Cytochrome P450 Enzymes · Natural Products · Nucleobase Transferase

- [1] V. Apostolopoulos, J. Bojarska, T. T. Chai, S. Elnagdy, K. Kaczmarek, J. Matsoukas, R. New, K. Parang, O. P. Lopez, H. Parhiz, C. O. Perera, M. Pickholz, M. Remko, M. Saviano, M. Skwarczynski, Y. Tang, W. M. Wolf, T. Yoshiya, J. Zabrocki, P. Zielenkiewicz, M. AlKhazindar, V. Barriga, K. Kelaionis, E. M. Sarasia, I. Toth, *Molecules* **2021**, 26, 430.
- [2] Z. Song, Y. Hou, Q. Yang, X. Li, S. Wu, *Mar. Drugs* **2021**, 19, 403.
- [3] R. Musetti, R. Polizzotto, A. Vecchione, S. Borselli, L. Zulini, M. D'Ambrosio, L. S. di Toppi, I. Pertot, *Micron* **2007**, 38, 643–650.
- [4] L. Buedenbender, T. Grkovic, S. Duffy, D. I. Kurtböke, V. M. Avery, A. R. Carroll, *Tetrahedron Lett.* **2016**, 57, 5893–5895.
- [5] H. Kohn, W. Widger, *Curr. Drug Targets Infect. Disord.* **2005**, 5, 273–295.
- [6] H. Turkez, I. Cacciatore, M. E. Arslan, E. Fornasari, L. Marinelli, S. A. Di, A. Mardinoglu, *Biomolecules* **2020**, 10, 737.
- [7] K. Ström, J. Sjögren, A. Broberg, J. Schnürer, *Appl. Environ. Microbiol.* **2002**, 68, 4322–4327.
- [8] Z. Ding, F. Li, C. Zhong, F. Li, Y. Liu, S. Wang, J. Zhao, W. Li, *Bioorg. Med. Chem.* **2020**, 28, 115435.
- [9] S. J. Sun, Y. C. Liu, C. H. Weng, S. W. Sun, F. Li, H. Li, H. Zhu, *Biomolecules* **2020**, 10, 298.
- [10] J. Zhu, Y. Zhang, J. Deng, H. Jiang, L. Zhuang, W. Ye, J. Ma, J. Jiang, L. Feng, *J. Agric. Food Chem.* **2019**, 67, 12013–12025.
- [11] L. Feni, L. Jutten, S. Parente, U. Piarulli, I. Neundorff, D. Diaz, *Chem. Commun.* **2020**, 56, 5685–5688.
- [12] T. Izoré, M. J. Cryle, *Nat. Prod. Rep.* **2018**, 35, 1120–1139.
- [13] J. A. Payne, M. Schoppet, M. H. Hansen, M. J. Cryle, *Mol. Biosyst.* **2017**, 13, 9–22.
- [14] M. Gondry, I. B. Jacques, R. Thai, M. Babin, N. Canu, J. Seguin, P. Belin, J. L. Pernodet, M. Moutiez, *Front. Microbiol.* **2018**, 9, 46.
- [15] N. Canu, M. Moutiez, P. Belin, M. Gondry, *Nat. Prod. Rep.* **2020**, 37, 312–321.



- [16] A. D. Borthwick, *Chem. Rev.* **2012**, *112*, 3641–3716.
- [17] L. Harken, S.-M. Li, *Appl. Microbiol. Biotechnol.* **2021**, *105*, 2277–2285.
- [18] S. Meng, W. Han, J. Zhao, X. H. Jian, H. X. Pan, G. L. Tang, *Angew. Chem. Int. Ed.* **2018**, *57*, 719–723; *Angew. Chem.* **2018**, *130*, 727–731.
- [19] P. Borgman, R. D. Lopez, A. L. Lane, *Org. Biomol. Chem.* **2019**, *17*, 2305–2314.
- [20] Y. Zhang, T. Yao, Y. Jiang, H. Li, W. Yuan, W. Li, *Appl. Environ. Microbiol.* **2021**, *87*, e02525–20.
- [21] T. Yao, J. Liu, Z. Liu, T. Li, H. Li, Q. Che, T. Zhu, D. Li, Q. Gu, W. Li, *Nat. Commun.* **2018**, *9*, 4091.
- [22] J. Liu, Y. Yang, L. Harken, S.-M. Li, *J. Nat. Prod.* **2021**, *84*, 3100–3109.
- [23] D. R. Nelson, *Biochim. Biophys. Acta Proteins Proteomics* **2018**, *1866*, 141–154.
- [24] N. P. Dunham, F. H. Arnold, *ACS Catal.* **2020**, *10*, 12239–12255.
- [25] C.-C. Chen, J. Min, L. Zhang, Y. Yang, X. Yu, R.-T. Guo, *ChemBioChem* **2021**, *22*, 1317–1328.
- [26] J. B. Patteson, W. Cai, R. A. Johnson, K. C. Santa Maria, B. Li, *Biochemistry* **2018**, *57*, 61–65.
- [27] M. J. Cryle, S. G. Bell, I. Schlichting, *Biochemistry* **2010**, *49*, 7282–7296.
- [28] V. V. Shende, Y. Khatrri, S. A. Newmister, J. N. Sanders, P. Lindovska, F. Yu, T. J. Doyon, J. Kim, K. N. Houk, M. Movassaghi, D. H. Sherman, *J. Am. Chem. Soc.* **2020**, *142*, 17413–17424.
- [29] C. Sun, H. Peng, W. Zhang, M. Zheng, W. Tian, Y. Zhang, H. Liu, Z. Lin, Z. Deng, X. Qu, *J. Org. Chem.* **2021**, *86*, 11189–11197.
- [30] J. Liu, X. Xie, S.-M. Li, *Chem. Commun.* **2020**, *56*, 11042–11045.
- [31] P. Belin, M. H. Le Du, A. Fielding, O. Lequin, M. Jacquet, J. B. Charbonnier, A. Lecoq, R. Thai, M. Courcon, C. Masson, C. Dugave, R. Genet, J. L. Pernodet, M. Gondry, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7426–7431.
- [32] H. Yu, X. Xie, S.-M. Li, *Org. Lett.* **2019**, *21*, 9104–9108.
- [33] J. Shi, X. Xu, E. J. Zhao, B. Zhang, W. Li, Y. Zhao, R. H. Jiao, R. X. Tan, H. M. Ge, *Org. Lett.* **2019**, *21*, 6825–6829.
- [34] L. Harken, J. Liu, O. Kreuz, R. Berger, S.-M. Li, *ACS Catal.* **2022**, *12*, 648–654.
- [35] K. Scherlach, C. Hertweck, *Nat. Commun.* **2021**, *12*, 3864.
- [36] S. Lautru, M. Gondry, R. Genet, J. L. Pernodet, *Chem. Biol.* **2002**, *9*, 1355–1364.
- [37] I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, J. L. Pernodet, M. Gondry, P. Belin, *Nat. Chem. Biol.* **2015**, *11*, 721–727.
- [38] J. Liu, H. Yu, S.-M. Li, *Appl. Microbiol. Biotechnol.* **2018**, *102*, 4435–4444.
- [39] T. W. Giessen, A. M. von Tesmar, M. A. Marahiel, *Biochemistry* **2013**, *52*, 4274–4283.
- [40] M. H. Medema, R. T. de, B. S. Moore, *Nat. Rev. Genet.* **2021**, *22*, 553–571.
- [41] N. Zaburanyi, M. Rabyk, B. Ostash, V. Fedorenko, A. Luzhetskyy, *BMC Genomics* **2014**, *15*, 97.
- [42] Y. Zhu, P. Fu, Q. Lin, G. Zhang, H. Zhang, S. Li, J. Ju, W. Zhu, C. Zhang, *Org. Lett.* **2012**, *14*, 2666–2669.
- [43] V. G. Dumas, L. A. Defelipe, A. A. Petruk, A. G. Turjanski, M. A. Marti, *Proteins Struct. Funct. Bioinf.* **2014**, *82*, 1004–1021.
- [44] A. Greule, J. E. Stok, J. J. De Voss, M. J. Cryle, *Nat. Prod. Rep.* **2018**, *35*, 757–791.
- [45] J. Liu, X. Xie, S.-M. Li, *Angew. Chem. Int. Ed.* **2019**, *58*, 11534–11540; *Angew. Chem.* **2019**, *131*, 11658–11664.
- [46] H. Yu, X. Xie, S.-M. Li, *Org. Lett.* **2018**, *20*, 4921–4925.

Manuscript received: January 9, 2022

Accepted manuscript online: February 24, 2022

Version of record online: ■■, ■■



Supporting Information  
©Wiley-VCH 2021  
69451 Weinheim, Germany

## **Widely Distributed Bifunctional Bacterial Cytochrome P450 Enzymes Catalyze both Intramolecular C-C Bond Formation in *cyclo*-L-Tyr-L-Tyr and Its Coupling with Nucleobases**

Jing Liu,<sup>[a, c]</sup> Lauritz Harken,<sup>[a, c]</sup> Yiling Yang,<sup>[a]</sup> Xiulan Xie,<sup>[b]</sup> and Shu-Ming Li<sup>\*[a]</sup>

<sup>a</sup> Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany;

<sup>b</sup> Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Straße 4, 35032 Marburg, Germany;

<sup>c</sup> These authors contributed equally: Jing Liu, Lauritz Harken.

\* Email: shuming.li@staff.uni-marburg.de

## Table of Contents

<b>Methods .....</b>	<b>131</b>
Computer-assisted sequence analysis.....	131
Bacterial strains, plasmids, and growth conditions .....	131
Genetic manipulation, PCR amplification, and gene cloning .....	131
Heterologous gene expression in <i>S. albus</i> J1074 and cultivation for secondary metabolite production.....	131
LC-MS analysis.....	132
Overproduction in <i>E. coli</i> and purification of the recombinant GymB <sub>1</sub> – GymB <sub>6</sub> .....	132
<i>In vitro</i> assays with GymB <sub>1</sub> – GymB <sub>6</sub> .....	132
Large-scale fermentation, extraction, and isolation of secondary metabolites .....	133
Determination of production yields .....	133
NMR analysis and structural elucidation .....	133
Cultivation of the <i>S. albus</i> transformant in media containing <sup>15</sup> N-labeled salt .....	134
<b>Supplementary Tables .....</b>	<b>135</b>
Table S1. Comparison of <i>gym</i> genes with known entries .....	135
Table S2. Bacterial strains used in this study .....	136
Table S3. Key residues for the substrate binding pockets of CDPSs .....	137
Table S4. Cloning and expression constructs used in this study .....	138
Table S5. Product yields of the <i>S. albus</i> transformants harboring guatyromycine and mycrocyclosin clusters .....	139
Table S6. <sup>1</sup> H NMR data of cYY (1), cYF (2), and mycrocyclosin (3) in DMSO- <i>d</i> <sub>6</sub> .....	140
Table S7. NMR data of guatyromycine A (4) in DMSO- <i>d</i> <sub>6</sub> .....	141
Table S8. NMR data of guatyromycine B (5) in DMSO- <i>d</i> <sub>6</sub> .....	142
<b>Supplementary Figures.....</b>	<b>143</b>
Figure S1. Phylogenetic tree of CDPSs associated with cytochrome P450 enzymes.....	143
Figure S2. Phylogenetic tree of CDPS-associated cytochrome P450 enzymes.....	144
Figure S3. Alignments of GymA <sub>1</sub> – GymA <sub>6</sub> with AlbC .....	145
Figure S4. HPLC analysis of <i>E. coli</i> extracts harboring <i>gymA</i> <sub>1</sub> – <i>gymA</i> <sub>6</sub> .....	146
Figure S5. UV spectra of 1 – 5.....	147
Figure S6. LC-MS analysis of extracts of <i>S. albus</i> J1074 harboring <i>gym</i> (AB) <sub>1</sub> – <i>gym</i> (AB) <sub>6</sub> and <i>rv2275/rv2276</i> by UV detection and EICs for 3, 4, and 5.....	148
Figure S7. LC-MS analysis of extracts of <i>S. albus</i> J1074 harboring <i>gym</i> (AB) <sub>1</sub> – <i>gym</i> (AB) <sub>6</sub> and <i>rv2275/rv2276</i> by UV detection and EICs for potential cYF derivatives .....	149
Figure S8. <sup>1</sup> H NMR spectrum of cYY (1) in DMSO- <i>d</i> <sub>6</sub> .....	150
Figure S9. <sup>1</sup> H NMR spectrum of cYF (2) in DMSO- <i>d</i> <sub>6</sub> .....	151

## PUBLICATIONS

Figure S10. <sup>1</sup> H NMR spectrum of mycrocyclosin ( <b>3</b> ) in DMSO- <i>d</i> <sub>6</sub> .	152
Figure S11. <sup>1</sup> H NMR spectrum of guatyromycine A ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .	153
Figure S12. <sup>13</sup> C NMR spectrum of guatyromycine A ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .	154
Figure S13. <sup>1</sup> H- <sup>1</sup> H COSY spectrum of guatyromycine A ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .	155
Figure S14. HSQC spectrum of guatyromycine A ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .	156
Figure S15. HMBC spectrum of guatyromycine A ( <b>4</b> ) in DMSO- <i>d</i> <sub>6</sub> .	157
Figure S16. <sup>1</sup> H NMR spectrum of guatyromycine B ( <b>5</b> ) in DMSO- <i>d</i> <sub>6</sub> .	158
Figure S17. <sup>13</sup> C NMR spectrum of guatyromycine B ( <b>5</b> ) in DMSO- <i>d</i> <sub>6</sub> .	159
Figure S18. <sup>1</sup> H- <sup>1</sup> H COSY spectrum of guatyromycine B ( <b>5</b> ) in DMSO- <i>d</i> <sub>6</sub> .	160
Figure S19. HSQC spectrum of guatyromycine B ( <b>5</b> ) in DMSO- <i>d</i> <sub>6</sub> .	161
Figure S20. HMBC spectrum of guatyromycine B ( <b>5</b> ) in DMSO- <i>d</i> <sub>6</sub> .	162
Figure S21. MS analysis of extracts of <i>S. albus</i> J1074 harboring <i>gymA<sub>6</sub>B<sub>6</sub></i> cultivated with and without <sup>15</sup> NH <sub>4</sub> Cl.	163
Figure S22. SDS-PAGE analysis of the purified P450s.	164
Figure S23. LC-MS analysis of cYY assays with GymB <sub>1</sub> – GymB <sub>6</sub> without nucleobases.	164
<b>References</b>	<b>166</b>

## Methods

### Computer-assisted sequence analysis

Nucleotide and amino acid sequences used in this study (Table S1) were obtained from NCBI databases (<http://www.ncbi.nlm.nih.gov>). Comparison of protein sequences was carried out by using BLASTP program (<http://blast.ncbi.nlm.nih.gov/>). The phylogenetic tree of CDPs and P450s (Figures S1 and S2) were created by MEGA version X (<http://www.megasoftware.net>) using the maximum-likelihood method. Multiple sequence alignments were carried out with the program ClustalW and visualized with ESPript 3.0 (<http://esprict.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>) to identify strictly conserved amino acid residues (Figure S3).

### Bacterial strains, plasmids, and growth conditions

Strains and plasmids used and generated in this study are listed in Tables S2 and S4, respectively. Recombinant *E. coli* strains were cultivated in liquid or on solid Luria-Bertani (LB) with 100 µg/mL ampicillin, 50 µg/mL kanamycin, 50 µg/mL apramycin or 25 µg/mL chloramphenicol when necessary.

*S. flavidovirens* DSM 40150 and *S. indicus* DSM 42001 were purchased from German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). *Streptomyces* sp. NRRL S-646, *Streptomyces* sp. NRRL B-3648, *S. katrae* NRRL ISP-5550 as well as *S. aureus* NRRL B-2808 were obtained from the ARS Culture Collection (NRRL). *S. albus* J1074<sup>[1]</sup> was kindly gifted by Prof. Dr. Andriy Luzhetskyy (Saarland University). *S. albus* J1074 and the generated ex-conjugants were maintained on MS plates (mannitol 20.0 g/L, soy flour 20.0 g/L, agar 20.0 g/L) at 28 °C for sporulation. For secondary metabolite production, *S. albus* J1074 transformants were cultivated in liquid modified R5 medium (Table S2) at 28 °C for 7 days.

### Genetic manipulation, PCR amplification, and gene cloning

Genetic manipulation in *E. coli* was performed according to the protocol by Green and Sambrook.<sup>[2]</sup> Genomic DNA isolation from *Streptomyces* was performed as described in the literature.<sup>[3]</sup>

Genes were amplified by PCR from the genomic DNA of the native strains by using primers listed in Table S4 with Phusion® High-Fidelity DNA Polymerase from New England Biolabs. The generated PCR fragments were cloned into pGEM-T Easy vector (Promega) and the sequence integrity was confirmed by sequencing. Subsequently, the fragments were released with restriction endonucleases from pGEM-T Easy and ligated into pPWW50A<sup>[4]</sup> for overexpression in *S. albus* J1074, or pET28a (+) for overexpression in *E. coli* BL21 (DE3), which were both linearized with the same endonucleases. The plasmid pLH11 harboring *rv2275* and *rv2276* from *M. tuberculosis* was cloned directly into pPWW50A, linearized with restriction endonucleases, via homologous recombination in *E. coli* DH5α. The generated constructs are listed in Table S4.

### Heterologous gene expression in *S. albus* J1074 and cultivation for secondary metabolite production

The constructed plasmids harboring different genes/gene clusters were firstly transferred into non-methylated *E. coli* ET12567/pUZ8002, then conjugated with *S. albus* J1074. The positive conjugants were firstly selected by the phenotype showing apramycin resistance and further

confirmed by PCR. The spores of the *S. albus* J1074 transformants were inoculated into 50 mL modified R5 liquid media supplied with 50 µg/mL of apramycin in 250 mL baffled flasks and cultured at 28 °C and 200 rpm for 7 days. 500 µL of such cultures were extracted with the same volume of EtOAc for three times. After that, the organic phases were combined, evaporated, and the dried residues dissolved in 250 µL of methanol with 30 µL DMSO. 5 µL of such samples were subjected to LC-MS for analysis.

### LC-MS analysis

LC-MS analysis was performed on an Agilent HPLC 1260 series system equipped with a photo diode array detector and a microTOF QIII mass spectrometer (Bruker, Bremen, Germany) by using a Multospher 120 RP-18 column (250x2 mm, 5 µm, CS-Chromatographie Service GmbH). For secondary metabolite analysis, a linear gradient of 5 – 100% ACN in H<sub>2</sub>O, both containing 0.1% HCOOH, in 40 min and a flow rate at 0.25 mL/min were used. The column was then washed with 100% ACN containing 0.1% HCOOH for 5 min and equilibrated with 5% ACN in H<sub>2</sub>O for 5 min. The parameters of the mass spectrometer were set as following: electrospray positive ion mode for ionization, capillary voltage with 4.5 kV, collision energy with 8.0 eV.

For analysis of the enzyme assays, samples were measured on the same LC-MS system as described above, but with a VDSpher PUR 100 C18-M-SE column (150x2 mm, 3 µm, VDS optilab Chromatographie Technik GmbH). A linear gradient of 5 – 100% ACN in H<sub>2</sub>O, both containing 0.1% HCOOH, in 30 min and a flow rate at 0.3 mL/min were used. The column was then washed and equilibrated as the described above.

### Overproduction in *E. coli* and purification of the recombinant GymB<sub>1</sub> – GymB<sub>6</sub>

The recombinant *E. coli* cells were cultivated for 16 h in 50 mL LB media supplied with 50 µg/mL kanamycin as preculture. 5 mL of the preculture were transferred into 500 mL LB media (with 50 µg/mL kanamycin) in 2 L-Erlenmeyer flasks and grown at 37 °C and 230 rpm to an absorption of 0.6 at 600 nm. Gene expression was induced with 0.1 mM IPTG at 16 °C for 20 h. The bacterial cultures were harvested by centrifugation (4,500 rpm, 20 min, 4 °C) and the cells were resuspended in lysis buffer (50 mM Tris-HCl, 10 mM imidazole, 300 mM NaCl, pH 8.0) with 2 – 5 mL/g wet weight. Lysozyme from the chicken egg white was added to the mixture at a final concentration of 1 mg/mL, which was incubated on ice for 30 min. The cells were then lysed by sonication on ice. Cell debris was removed by centrifugation at 13,000 rpm and 4 °C for 30 min. One-step purification of the recombinant His<sub>6</sub>-tagged protein was performed by using Ni-NTA agarose (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. The storage buffer was changed to 50 mM Tris-HCl (pH 8.0) containing 15% (v/v) glycerol through a PD MiniTrap G-25 column (Cytiva, Freiburg, Germany), which had been equilibrated with the same buffer. After proof of their purity on 12 % (w/v) SDS-PAGE (Figure S22), the recombinant GymB<sub>1</sub> – GymB<sub>6</sub> were stored frozen at -80 °C.

### *In vitro* assays with GymB<sub>1</sub> – GymB<sub>6</sub>

The enzyme assays of GymB<sub>1</sub> – GymB<sub>6</sub> were carried out with 100 µM cYY (1), 10 µM each of the six P450s, 8 mM NADPH, 25 µM spinach ferredoxin (Sigma-Aldrich), 0.4 unit/mL spinach ferredoxin-NADP<sup>+</sup> reductase (Sigma-Aldrich) with or without 100 µM guanine and 100 µM hypoxanthine, and 50 mM Tris-HCl buffer (pH 8.0) in a total volume of 50 µL. The reactions were performed at 30 °C for 16 h and then quenched with 50 µL ice-cold MeOH. After

centrifugation at 13,000 rpm for 5 min, 20  $\mu$ L of the supernatants were subjected to LC-MS analysis. Assays with heat-inactivated P450s were used as negative controls.

### Large-scale fermentation, extraction, and isolation of secondary metabolites

For structural elucidation of the accumulated products, *S. albus* J1074 transformants harboring *gymA<sub>6</sub>B<sub>6</sub>* were cultivated in modified R5 medium on large scales (12 L) at 28 °C for 7 days. The culture supernatants were collected and extracted with equal volume of EtOAc for three times. The EtOAc phases were combined, evaporated to dryness, dissolved in MeOH, and applied to chromatography on a silica gel column and eluted with a gradient of CH<sub>2</sub>Cl<sub>2</sub>: MeOH in ratios of 100:2, 100:3, 100:5, 100:10, 100:20 and 100:100. The fractions containing the target products were further purified on an Agilent HPLC 1260 series by using a semi-preparative Agilent ZORBAX Eclipse XDB C18 HPLC column (250  $\times$  9.4 mm, 5  $\mu$ m). The flow rate was set to 2.0 mL/min.

### Determination of production yields

For quantification, UV absorptions at 280 nm were used for **4**, and **5** and at 294 nm for **3**. 0.5 mL culture of *S. albus* J1074 transformants harboring (*gymAB*)<sub>1</sub> — (*gymAB*)<sub>6</sub> were extracted with same volume of EtOAc for three times. The organic phases were combined and evaporated to dryness. The residues were dissolved in 250  $\mu$ L MeOH and 30  $\mu$ L DMSO. 5  $\mu$ L were analyzed on LC-MS described above. The isolated products were used as authentic standards.

### NMR analysis and structural elucidation

All the purified compounds were dissolved in DMSO-*d*<sub>6</sub> for NMR analysis. The NMR spectra of the purified compounds **1** — **3** were recorded on a JOEL ECA-500 MHz spectrometer (JEOL, Tokyo, Japan). For compound **4** and **5**, NMR experiments were performed on a Bruker AVIII 500 spectrometer equipped with a 5 mm BBO cryo probe Prodigy with z-gradient. All spectra were processed with MestReNova 5.2.2 (Mestrelab Research, Santiago de Compostella, Spain). NMR data and spectra of the identified compounds are provided in Tables S6– S8 and Figures S8 – S20, respectively.

By comparison of their <sup>1</sup>H NMR data with those of authentic standards, compounds **1** and **2** were identified to be *cyclo*-L-Tyr-L-Tyr (cYY) and *cyclo*-L-Tyr-L-Phe (cYF), respectively. Compound **3** was identified as mycocyclusin by comparison of its <sup>1</sup>H NMR data with those published previously.<sup>[5]</sup>

Inspection of the NMR data of guatyromycine A (**4**), including <sup>1</sup>H, <sup>13</sup>C, <sup>1</sup>H–<sup>1</sup>H COSY, HSQC, and HMBC (Table S7, Figures S11 – S15), suggested a cYY derivative with a guaninyl residue. Five signals at  $\delta_c$  156.2 (C-2'), 155.2 (C-4'), 115.1 (C-5'), 159.7 (C-6'), and 152.8 ppm (C-8') can be assigned to the guaninyl unit. Two sets of AA'BB' systems at  $\delta_H$  6.71 and 6.90, and at  $\delta_H$  7.02 and 7.17 ppm in the <sup>1</sup>H NMR spectrum and at  $\delta_c$  115.2 and 131.1, and at  $\delta_c$  130.7 and 119.1 ppm in the <sup>13</sup>C NMR spectrum indicated the presence of two different 1,4-disubstituted benzene rings. All these data supported that the guanine residue in **4** is attached to a hydroxyl group of cYY (Figure 1B). The loss of the characteristic singlet of H-8' at approximate 8 ppm supported the linkage between C-8' and O-21.

Structure elucidation of guatyromycine B (**5**) confirmed our assumption that a hypoxanthine is transferred onto cYY. The five <sup>13</sup>C signals of the hypoxanthinyl residue were found at  $\delta_c$  153.2

(C-2'), 155.8 (C-4'), 114.9 (C-5'), 156.1 (C-6'), and 143.4 ppm (C-8'). The signal of C-2' is de-shielded compared to that of a hypoxanthin-cWW conjugate.<sup>[6]</sup> Furthermore, the signal of H-9 in the tyrosyl moiety appears as a singlet and that of H-10 is disappeared. Thus, hypoxanthine is attached *via* its C-2' to C-10 of the cYY residue (Figure 1B), which is unequivocally confirmed by the correlation from H-9 to C-2' in the HMBC spectrum (Table S8, Figures S16 – S20).

#### **Cultivation of the *S. albus* transformant in media containing <sup>15</sup>N-labeled salt**

Cultivation of *S. albus* transformant harboring *gymA<sub>6</sub>B<sub>6</sub>* in 50 mL medium containing 100 mg <sup>15</sup>NH<sub>4</sub>Cl and subsequent LC-MS analysis of the cultural extract demonstrated that at least two <sup>15</sup>N atoms are incorporated into in **3**, seven in **4**, and six in **5** (Figure S21). These data provide additional evidence for the structures of **3**, **4**, and **5** depicted in Figure 1.



## Supplementary Tables

**Table S1.** Comparison of *gym* genes with known entries

<i>Streptomyces</i> strains	Protein	Accession No.	Length (aa)	Sequence identity (%) to
<i>S. flavidovirens</i> DSM 40150	GymA <sub>1</sub>	WP_078606819.1	236	Rv2275 (63)
	GymB <sub>1</sub>	WP_028811244.1	395	Rv2276 (60)
<i>S. indicus</i> DSM 42001	GymA <sub>2</sub>	WP_093612344.1	241	Rv2275 (65)
	GymB <sub>2</sub>	WP_093612346.1	402	Rv2276 (63)
<i>Streptomyces</i> sp. NRRL S-646	GymA <sub>3</sub>	WP_199931956.1	243	Rv2275 (63)
	GymB <sub>3</sub>	WP_030942274.1	395	Rv2276 (62)
<i>S. aureus</i> NRRL B-2808	GymA <sub>4</sub>	WP_079042211.1	237	Rv2275 (60)
	GymB <sub>4</sub>	WP_055602838.1	397	Rv2276 (60)
<i>S. katrae</i> NRRL ISP-5550	GymA <sub>5</sub>	KJY31545.1	236	Rv2275 (62)
	GymB <sub>5</sub>	KJY31546.1	393	Rv2276 (57)
<i>Streptomyces</i> sp. NRRL B-3648	GymA <sub>6</sub>	WP_078946746.1	229	Rv2275 (64)
	GymB <sub>6</sub>	WP_053711053.1	403 <sup>a</sup>	Rv2276 (59)

<sup>a</sup> Due to insolubility of the recombinant protein with its original coding region, a 186 bp genomic sequence upstream *gymB*<sub>6</sub> was added for expression in *E.coli* BL21 (DE3).

**Table S2.** Bacterial strains used in this study

Strain	Source	Cultivation media
<i>E. coli</i> DH5α	Invitrogen	LB
<i>E. coli</i> ET12567/pUZ8002	[7]	LB
<i>S. albus</i> J1074	[1]	MS (solid), modified R5 (liquid)
<i>S. flavidovirens</i> DSM 40150	DSMZ	modified R5, GYM
<i>S. indicus</i> DSM 42001	DSMZ	modified R5, GYM
<i>Streptomyces</i> sp. NRRL S-646	NRRL	modified R5, GYM
<i>S. katrae</i> NRRL ISP-5550	NRRL	modified R5, GYM
<i>Streptomyces</i> sp. NRRL B-3648	NRRL	modified R5, GYM
<i>S. aureus</i> NRRL B-2808	NRRL	modified R5, GYM
<i>M. tuberculosis</i> DSM 43990 <sup>a</sup>	DSMZ	---

<sup>a</sup> DNA was ordered instead of the bacterium

NRRL: ARS Culture Collection

DSMZ: German Collection of Microorganisms and Cell Cultures GmbH

LB medium: tryptone 10.0 g/L, yeast extract 5.0 g/L, NaCl 10.0 g/L.

MS medium: mannitol 20.0 g/L, soy flour 20.0 g/L, agar 20.0 g/L.

Modified R5 medium: sucrose 103.0 g/L, glucose 10.0 g/L, yeast extract 5.0 g/L, MgCl<sub>2</sub>·6H<sub>2</sub>O 10.12 g/L, K<sub>2</sub>SO<sub>4</sub> 0.25 g/L, Difco casaminoacids 0.1 g/L, MOPS 21.0 g/L, trace element solution 2 mL/L, pH 7.2

GYM medium: glucose 4.0 g/L, yeast extract 4.0 g/L, malt extract 10.0 g/L, pH 7.2

**Table S3.** Key residues for the substrate binding pockets of the investigated CDPSs compared to AlbC<sup>[8]</sup>

Protein	amino acid residues of pocket P1								amino acid residues of pocket P2							Predicted product
	33	35	65	67	119	185	186	200	152	155	156	159	204	206	207	
GymA <sub>1</sub>	V	G	V	T	M	L	F	N	L	T	F	T	Q	L	P	cYY
GymA <sub>2</sub>	V	G	V	T	M	L	F	N	L	T	F	D	Q	L	P	cYY
GymA <sub>3</sub>	V	G	V	T	M	L	F	N	L	S	F	T	Q	L	P	cYY
GymA <sub>4</sub>	V	G	V	T	M	L	F	N	L	I	F	G	Q	L	P	cYY
GymA <sub>5</sub>	V	G	I	T	M	L	F	N	L	I	F	S	Q	L	P	cYY
GymA <sub>6</sub>	V	G	V	T	M	L	F	N	L	R	F	A	Q	L	P	cYY

Prediction of CDPS products were carried out based on the previous reported methods.<sup>[9;10]</sup>

**Table S4.** Cloning and expression constructs used in this study

Gene	Primer sequences (5'-3')	Cloning construct	Expression vector	Cloning sites	Expression constructs
<i>gymA</i> <sub>1</sub> -D40150	<u>CATATG</u> ACCACAGCAGTAGAACTC <u>GGATCCT</u> CAGTGCTGGTCATGGCG	pLH01	pET28a(+)	NdeI/BamHI	pLH06
<i>gym(AB)</i> <sub>1</sub> -D40150	<u>CATATG</u> ACCACAGCAGTAGAACTC <u>GGATCCG</u> ATCACTACCAGGCGATG	pLH02	pPWW50	NdeI/BamHI	pLH07
<i>gymA</i> <sub>2</sub> -D42001	<u>CATATG</u> ACCGCTGTGACCGTTCTCGAATC <u>GGATCCG</u> GTCATCGGGCTTCTCCTTCG	pJL85	pET28a(+)	NdeI/BamHI	pJL95
<i>gym(AB)</i> <sub>2</sub> -D42001	<u>CATATG</u> ACCGCTGTGACCGTTCTCGAATC <u>GGATCCT</u> CACCACATCACCGGCAGCCG	pJL86	pPWW50	NdeI/BamHI	pJL96
<i>gymA</i> <sub>3</sub> -S-646	<u>CATATG</u> ACGACTGCGACCGTTTCC <u>GGATCCT</u> CAGCGGACATCGGTGTGGTC	pJL87	pET28a(+)	NdeI/BamHI	pJL97
<i>gym(AB)</i> <sub>3</sub> -S-646	<u>CATATG</u> ACGACTGCGACCGTTTCC <u>GGATCC</u> ATGACCGGCCGCCGAG	pJL88	pPWW50	NdeI/BamHI	pJL98
<i>gymA</i> <sub>4</sub> -B2808	<u>CATATG</u> ACTCAGGCGGCTTTTCGC <u>GGATCCT</u> GACCACGGTCATCGGAATCT	pJL93	pET28a(+)	NdeI/BamHI	pJL103
<i>gym(AB)</i> <sub>4</sub> -B2808	<u>CATATG</u> ACTCAGGCGGCTTTTCGC <u>GGATCCT</u> TACCAGGCCACCGGAAG	pJL94	pPWW50	NdeI/BamHI	pJL104
<i>gymA</i> <sub>5</sub> -ISP5550	<u>CATATG</u> ACAACAGCAACCCGGACC <u>GGATCCT</u> CACAGCTGGTCTCCTTCGGT	pJL89	pET28a(+)	NdeI/BamHI	pJL99
<i>gym(AB)</i> <sub>5</sub> -ISP5550	<u>CATATG</u> ACAACAGCAACCCGGACC <u>GGATCCT</u> CACCACATCACCGGCAGGC	pJL90	pPWW50	NdeI/BamHI	pJL100
<i>gymA</i> <sub>6</sub> -B3648	<u>CATATG</u> ACCACAGCATCCGTGCTGAT <u>GGATCCT</u> CATTCCGCGCTCCTTCC	pJL91	pET28a(+) & pPWW50	NdeI/BamHI	pJL101 & pJL117
<i>gym(AB)</i> <sub>6</sub> -B3648	<u>CATATG</u> ACCACAGCATCCGTGCTGAT <u>ACTAGT</u> CACCAGGCCACCGGGAG	pJL92	pPWW50	NdeI/Spel	pJL102
<i>rv2275-rv2276 M.tuberc..</i>	CACAGCAGCGGCCATATCGAAGGTCGTC ATATGGCTGCCGAACCAAGGC TTTAGGCTAGAGATCTCGATCCCGCGAAA TCTACCAGAGCACCGGAAGGC	---	pPWW50	NdeI/BamHI <sup>a</sup>	pLH11
<i>gymB</i> <sub>1</sub> -D40150	<u>CATATG</u> ACCAGCACTGACACGCTCCTC <u>GGATCCCT</u> TACCAGGCGATGGGCAGGC	pJL105	pET28a(+)	NdeI/BamHI	pJL111
<i>gymB</i> <sub>2</sub> -D42001	<u>CATATG</u> ACCGCGCCCCGCCCTC <u>GGATCCT</u> CACCACATCACCGGCAGCCG	pJL106	pET28a(+)	NdeI/BamHI	pJL112
<i>gymB</i> <sub>3</sub> -S-646	<u>CATATG</u> TCCGCTGACCTGCTCATCG <u>GGATCCT</u> CACCACATGGCAGGCAGCC	pJL107	pET28a(+)	NdeI/BamHI	pJL113
<i>gymB</i> <sub>4</sub> -B2808	<u>CATATG</u> ACCGTGGTGATGACGAACG <u>GGATCCCT</u> TACCAGGCCACCGGAAG	pJL110	pET28a(+)	NdeI/BamHI	pJL116
<i>gymB</i> <sub>5</sub> -ISP5550	<u>CATATG</u> ACCACCCACGCGGAGCC <u>GGATCCT</u> CACCACATCACCGGCAGGCG	pJL108	pET28a(+)	NdeI/BamHI	pJL114
<i>gymB</i> <sub>6</sub> -B3648	<u>CATATG</u> CGCCGAGGCCCGCT <u>AAGCTT</u> CACCAGGCCACCGGGAGC	pJL109	pET28a(+)	NdeI/HindIII	pJL115

Restriction sites for cloning are underlined in the primer sequences. Cloning constructs are based on pGEM T<sup>®</sup> EASY vector.

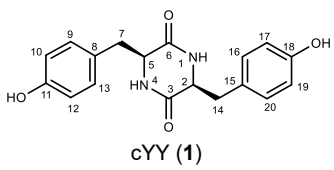
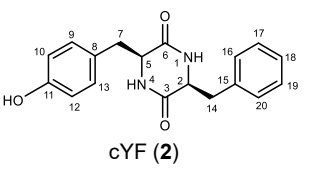
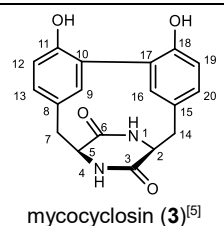
<sup>a</sup> pLH11 was cloned *via* homologous recombination directly into pPWW50A

**Table S5.** Product yields of the *S. albus* transformants harboring guatyromycine and mycocyclosin clusters

<i>S. albus</i> with	Yields [mg/L]		
	mycocyclosin ( <b>3</b> )	guatyromycine A ( <b>4</b> )	guatyromycine B ( <b>5</b> )
<i>gymA<sub>1</sub>B<sub>1</sub></i>	2.3 ± 0.1	3.4 ± 0.2	2.3 ± 0.2
<i>gymA<sub>2</sub>B<sub>2</sub></i>	1.6 ± 0.2	2.6 ± 0.2	2.4 ± 0.1
<i>gymA<sub>3</sub>B<sub>3</sub></i>	5.6 ± 0.2	5.0 ± 0.7	5.4 ± 0.5
<i>gymA<sub>4</sub>B<sub>4</sub></i>	9.4 ± 0.3	5.5 ± 0.3	5.4 ± 0.1
<i>gymA<sub>5</sub>B<sub>5</sub></i>	10.1 ± 0.1	4.1 ± 0.0	3.3 ± 0.1
<i>gymA<sub>6</sub>B<sub>6</sub></i>	24.0 ± 1.3	5.5 ± 0.1	3.6 ± 0.1
<i>rv2275/ rv2276</i>	53.9 ± 2.8	Not detected	Not detected

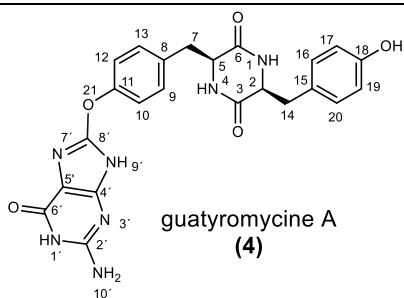
The data are mean values and the deviation are given as ± values (n = two or three independent experiments)

**Table S6.**  $^1\text{H}$  NMR data of cYY (**1**), cYF (**2**), and mycocyclosin (**3**) in  $\text{DMSO-}d_6$ 

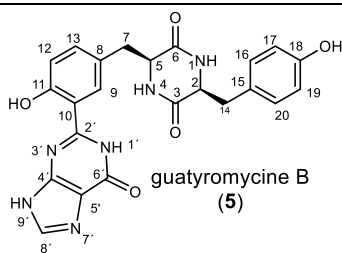
compounds			
	cYY (1)	cYF (2)	mycocyclosin (3) <sup>[5]</sup>
Position	$\delta_{\text{H}}$ , multi. ( <i>J</i> in Hz)	$\delta_{\text{H}}$ , multi. ( <i>J</i> in Hz)	$\delta_{\text{H}}$ , multi. ( <i>J</i> in Hz)
1	7.73, d (1.5)	7.82, s	7.97, s
2	3.85, t (6.1)	3.88, t (6.0)	4.33, d (5.5)
4	7.73, d (1.5)	7.82, s	7.97, s
5	3.85, t (6.1)	3.95, t (6.0)	4.33, d (5.5)
7a	2.51, dd (13.7, 6.1)	2.59 <sup>b</sup> , dd (13.7, 6.0)	3.46, d (15.6)
7b	2.12, dd (13.7, 6.1)	2.18 <sup>a</sup> , m	2.65, dd (15.6, 5.5)
9	6.84, d (8.4)	6.84, d, (8.3)	6.58, d (2.4)
10	6.67, d (8.4)	6.67, d, (8.3)	-
11	9.19 (OH), s	9.20 (OH), s	9.04 (OH), br s
12	6.67, d (8.4)	6.67, d (8.3)	6.62, d (8.1)
13	6.84, d (8.4)	6.84, d (8.3)	6.84, dd, (8.3, 2.3)
14a	2.51, dd (13.7, 6.1)	2.54 <sup>b</sup> , m	3.46, d (15.6)
14b	2.12, dd (13.7, 6.1)	2.18 <sup>a</sup> , m	2.65, dd (15.6, 5.5)
16	6.84, d (8.4)	7.04, d (7.2)	6.58, d (2.4)
17	6.67, d (8.4)	7.27, t (7.2)	-
18	9.19 (OH), s	7.20, t (7.2)	9.04 (OH), br s
19	6.67, d (8.4)	7.27, t (7.2)	6.62, d (8.1)
20	6.84, d (8.4)	7.04, d (7.2)	6.84, dd, (8.3, 2.3)

<sup>a</sup> signal overlapping with solvent peak, <sup>b</sup> signals overlapping with each other.

The spectra of **1** and **2** correspond well to those of the authentic standards and **3** to those published previously.<sup>[5]</sup>

**Table S7.** NMR data of guatyromycine A (**4**) in DMSO-*d*<sub>6</sub>

Position	$\delta_c$	$\delta_H$ , multi. ( <i>J</i> in Hz)	COSY	HMBC
1	-	7.87, d (2.1)	H-2	C-2, 3,
2	55.5	3.87, m	H-1, H-14	C-3, 6, 15
3	166.3	-	-	-
4	-	8.00, d (1.7)	H-5	C-5, 6
5	55.7	3.99, m	H-4, H-7	C-3, 7, 8
6	166.4	-	-	-
7a	38.3	2.63, dd, (13.7, 5.0)	-	C-5, 6, 8, 9, 13
7b		2.51, m	-	C-5, 6, 8, 9, 13
8	126.3	-	-	-
9	131.1	6.90, d (8.4)	H-10	C-5, 7, 10, 11, 13
10	115.2	6.71, d (8.4)	H-9	C-9, 11
11	156.3	-	-	-
12	115.2	6.71, d (8.4)	H-13	C-11, 13
13	131.1	6.90, d (8.4)	H-12	C-5, 7, 9, 11, 12
14a	39.8	2.51, m	-	C-2, 3, 15, 16, 20
14b		1.93, dd (13.6, 7.3)	-	C-2, 3, 15, 16, 20
15	133.1	-	-	-
16	130.7	7.02, d (8.5)	H-17	C-14, 17, 18, 20
17	119.1	7.17, d (8.5)	H-16	C-15, 16, 18, 19
18	153.2	-	-	-
19	119.1	7.17, d (8.5)	H-20	C-15, 17, 18, 20
20	130.7	7.02, d (8.5)	H-19	C-14, 16, 18, 19
1'	-	10.80, s	-	-
2'	156.2	-	-	-
4'	155.2	-	-	-
5'	115.1	-	-	-
6'	159.7	-	-	-
8'	152.8	-	-	-
9'	-	not observed	-	-
10'	-	6.35, br s	-	-

**Table S8.** NMR data of guatyromycine B (**5**) in DMSO-*d*<sub>6</sub>

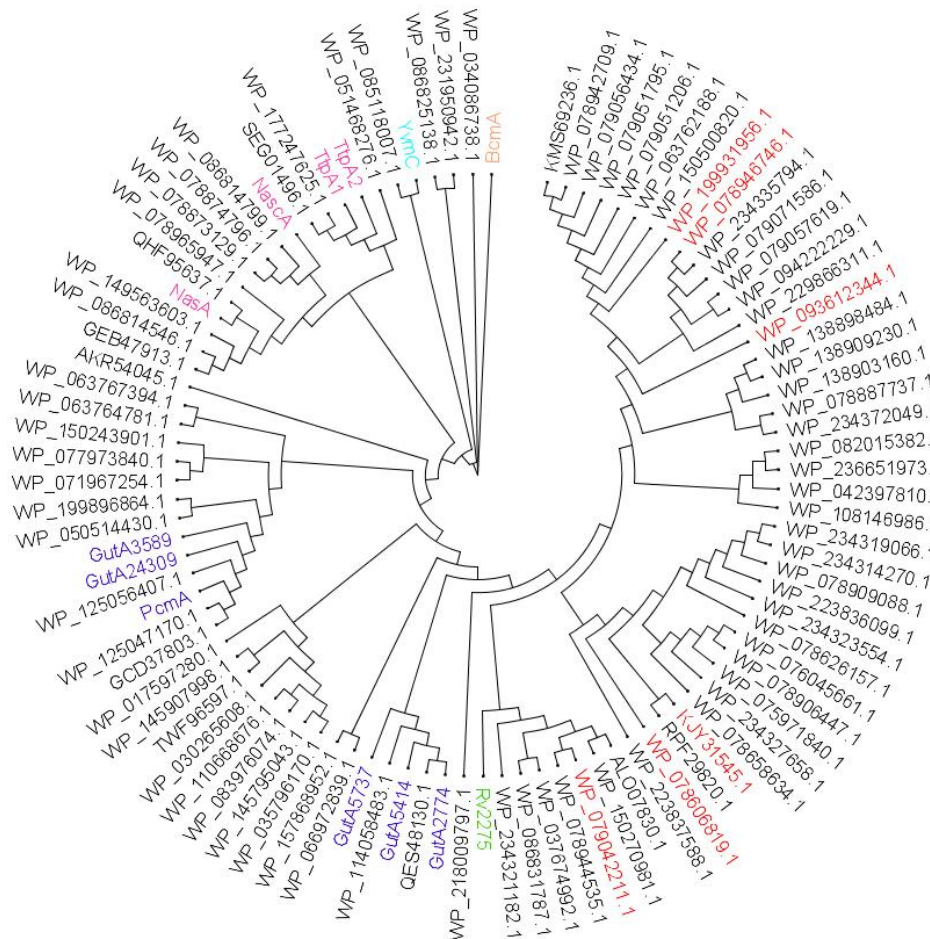
Position	$\delta_c$	$\delta_H$ , multi., (J in Hz)	COSY	HMBC
1	-	7.84, s	H-2	C-2, 6
2	55.5	3.93 <sup>b</sup> , m	H-14	C-3, 14, 15
3	166.5	-	-	-
4	-	7.76, s	H-5	C-3, 5, 7
5	55.9	3.94 <sup>b</sup> , m	H-7	C-6, 7, 8
6	166.3	-	-	-
7a	39.6 <sup>a</sup>	2.63 <sup>c</sup> , dd (13.7, 4.3)	H-5, 7	C-5, 6, 8, 13
7b		1.98, dd (13.7, 7.4)	H-5, 7	C-5, 6, 8, 13
8	126.5	-	-	-
9	127.4	7.80, s	-	C-7, 13, 2'
10	115.1	-	-	-
11	156.4	-	-	-
12	116.4	6.85 <sup>d</sup> , br d (9.3)	H-13	C-8, 10, 11
13	131.3	6.87, br d (9.3)	H-9, 12	C-9, 11
14a	38.7	2.61 <sup>c</sup> , dd (13.0, 4.2)	H-2, 14	C-2, 3, 15, 16, 20
14b		2.39, dd (13.0, 5.2)	H-2, 14	C-2, 3, 15, 16, 20
15	126.3	-	-	-
16	130.9	6.84 <sup>d</sup> , d (8.4)	H-17, 19	C-17, 18, 19
17	115.3	6.64, d (8.4)	H-16, 20	C-15, 16, 18, 20
18	156.3	-	-	-
19	115.1	6.64, d (8.4)	H-16, 20	C-15, 16, 18, 20
20	130.8	6.84 <sup>d</sup> , d (8.4)	H-17, 19	C-17, 18, 19
1'	-	not observed	-	-
2'	153.2	-	-	-
4'	155.8	-	-	-
5'	114.9	-	-	-
6'	156.1	-	-	-
8'	143.4	7.88, s	-	C-4'
9'	-	not observed	-	-

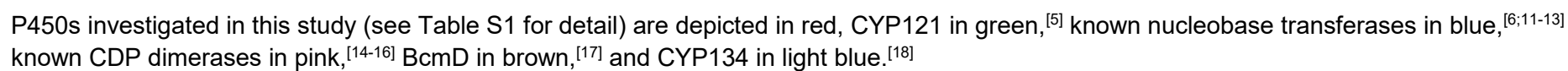
<sup>a</sup> signal overlapping with solvent peak<sup>b, c, d</sup> signals with same letter overlapping with each other



**Figure S1.** Phylogenetic tree of CDPs associated with cytochrome P450 enzymes.

CDPSs investigated in this study (see Table S1 for detail) are depicted in red, Rv2275 in green,<sup>[5]</sup> CDPs associated to nucleobase transferases in blue,<sup>[6;11-13]</sup> those associated to dimerases in pink,<sup>[14-16]</sup> BcmA in brown,<sup>[17]</sup> and YvmC in light blue.<sup>[18]</sup>





# PUBLICATIONS

	1	10	20	30	40	50		
Albc	MLAG.....	LVPAPDHGM	REEILGDR	SRLIRQ	GEHALIGIS	AGNSYFSQKNTVM	LLQWA	
WP_078606819.1	.....	MTTAVEL	FTVQPYTS	HCQVIAD	EGEHAVIGISP	AGNSYFSAQRVTD	LAHWG	
WP_093612344.1	MTA..	VTVLESQPF	TQKFHVA	PYTAHCR	VICE	DGDHAVIGVSP	AGNSYFSAQRIHD	LALWG
WP_199931956.1	MTTATV	SEIQHVTPADL	FLVRPFTP	HCEVI	RTAADHAVIGISP	AGNSYFSAQRIHD	LARWG	
WP_079042211.1	.....	MTQAAFAPAET	LRLQPYTD	HCRVIAD	EGAHVVIGVSP	AGNSYFTAQRLREL	TRWG	
KJY31545.1	.....	MTTATRTSA	FRVDPYTP	HCQVIAGE	GDHAVIGISP	AGNSYFSAQRVHAL	AEWA	
WP_078946746.1	.....	MADV	FEVQPYTS	HCEVILR	EGAHAVIGVSP	AGNSYFSARRLHD	LARWG	

	60	70	80	90	100	110						
Albc	GQR	FERTD	VYVDTHI	D	EMLIAD	GRSAQ	EAERSVKRT	LKDLRRRL	RRSL	ESVGD	HAE	RFR
WP_078606819.1	LRN	FRQVD	LVYTDLHI	G	MYEAL	GYPP	EAARRKAVKN	LRGVRAKV	TN	AALS	ADPT	GTHVR
WP_093612344.1	LAL	FRQVD	LVYTDLHV	A	MYEAF	GYAP	EAARRKFVN	LRGVRAKV	NN	AV	AALDP	EGERLR
WP_199931956.1	LAL	FERVD	LVYTDLHV	A	MYEAS	GYTAD	EAARRKAVKN	LRGVRAKV	LA	AVEA	ADPE	GNRLR
WP_079042211.1	LDH	FEQVD	LVYTDLHV	S	MYEAL	GYGA	EAARRKAVKN	LRGVRAKV	GT	AV	AEADPT	GVRVR
KJY31545.1	LDN	FRQAD	LIYTDLHV	A	MYVAM	GYPE	EAARRKTVKN	LRGVRAKV	TN	AV	ADIDP	SGTRLR
WP_078946746.1	LDH	FD	VDLVYTDLYV	A	MYEAS	GYPP	EAARRKAVKN	LRGVRAKV	RD	AV	SAADP	DGVRLLD

	120	130	140	150	160	170										
Albc	VRS	LS	SELQE	TPEYRA	VRERT	DRAFE	DAE	FATACE	DMVRA	VVMN	RPGD	GVG	ISAE	HLR	AG	
WP_078606819.1	GR	PM	SDLTS	SPAYRE	LHAHL	GDLL	LAT	DPEFRVTCE	TLVD	TLST	KVLD	GKD	STAQ	QREVC		
WP_093612344.1	AQ	PM	SSFTEN	PAYRE	IHRHL	QDR	LAH	DDFRTTCD	ALVD	TLSD	KVLD	GKG	ATAE	QREVC		
WP_199931956.1	AH	MS	SALRD	NPAYRQ	IHED	LRAR	LSR	DDEFRTTCE	KLVD	SLST	K...	AQE	VTAR	QRAVC		
WP_079042211.1	ARG	MS	SEFRAL	PAYQE	LHRQ	VVA	VDG	DPVVRQTC	ALTG	IFLAG	KLAP	GQE	ASGL	QKEVC		
KJY31545.1	GR	PM	SSLQGI	PAYDR	MHAD	LLER	LGT	DPAFAQTANE	LVDI	FLSS	KVLG	GQQ	ATAH	QRKVC		
WP_078946746.1	W	HP	MS	SEFR	TNPAYQE	IHRQL	QER	LVS	DGA	FRSVCE	TLVNR	FLMAR	...	GET	PTER	QRAVC

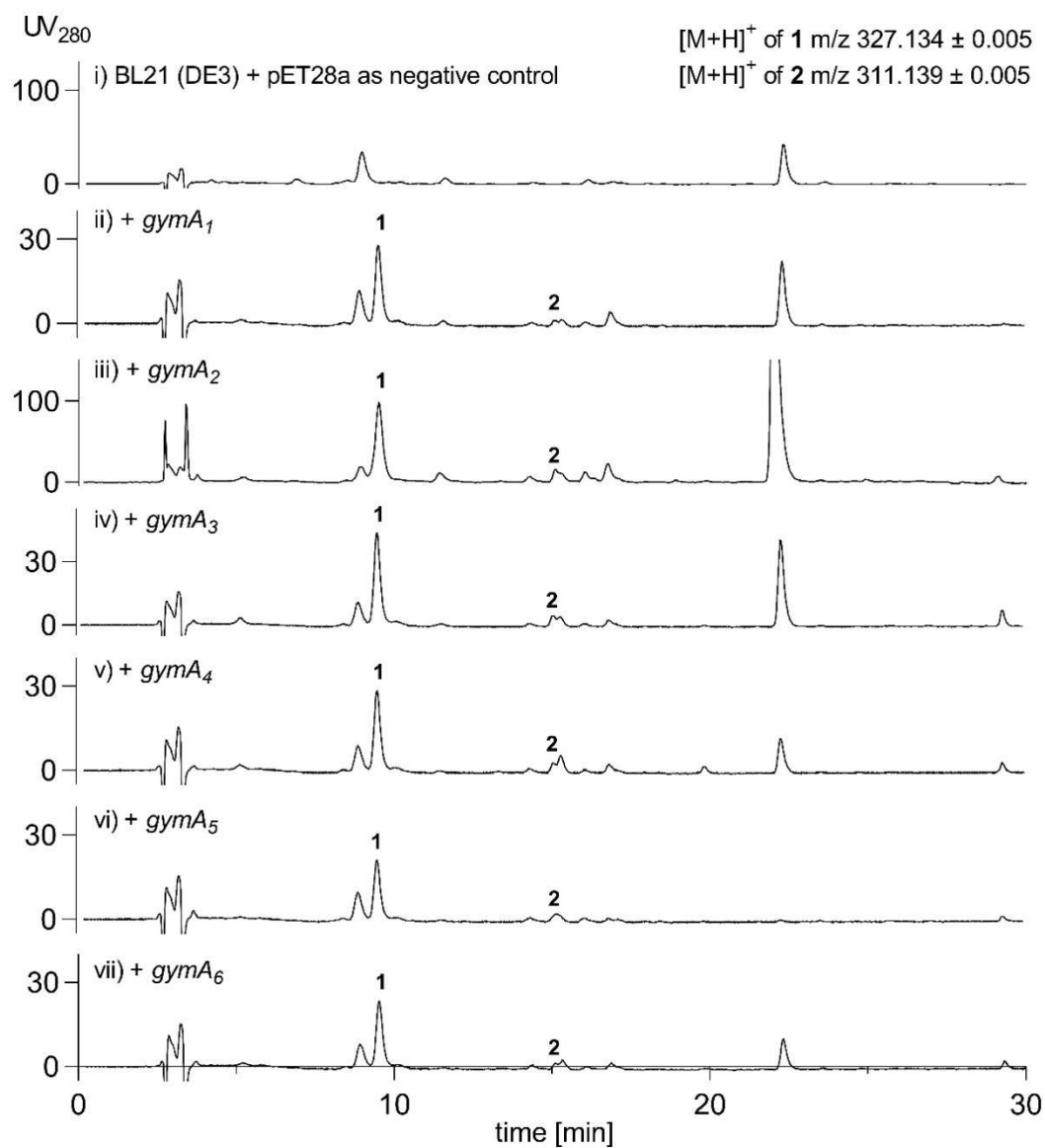
  

	180	190	200	210	220	230									
Albc	LN	YVLA	EAAPLF	ADSP	GVFS	VPSSVL	CYHIDT	PITAF	LSR	RET	G	FRA	AEQ	AYVVVR	PQEL
WP_078606819.1	LQ	YICAE	VPLFLD	TPAIL	GVSS	SSLN	CYHQAL	PLADL	LY	ARGP	G	LRA	SRN	QGHAI	VT
WP_093612344.1	LR	YVCA	EAAPLF	LDTP	AILGV	PSSLN	CYHQALL	PMAEL	LY	SRGH	G	LRA	SRN	QGHAI	VT
WP_199931956.1	ME	YVTA	EAAPLF	LDTP	AILGV	PSSLN	CYHQALL	PMAEL	LY	SPG	A	G	LRA	SRN	QGHAI
WP_079042211.1	RA	YICAE	VPLFLD	TPAIL	GVSS	SSLN	CYHQAL	PLADL	LY	ARGT	G	LRA	SRN	QGHAI	VT
KJY31545.1	LD	YVCA	EAAPLF	LDTP	AILGV	PSSLN	CYHQAL	PMADL	LY	SRGA	G	LRA	SRN	QGHAV	VT
WP_078946746.1	LE	YVCA	EAAPLF	LDTP	AILRV	PSSLN	CYHQALL	PMAEL	LY	SRGA	G	LRA	SRN	QGHAI	VT

Albc	ADAA..
WP_078606819.1	HRHDQH
WP_093612344.1	EAR...
WP_199931956.1	DHTDVR
WP_079042211.1	DSR...
KJY31545.1	GDQL..
WP_078946746.1	EGAAE..

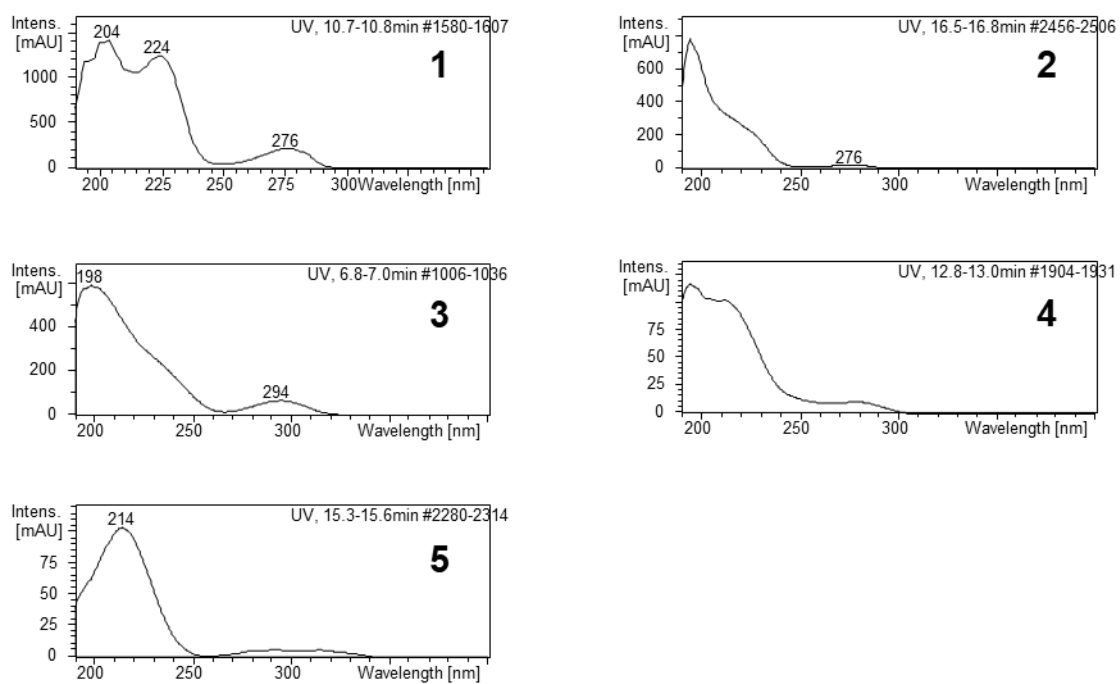
**Figure S3.** Alignments of GymA<sub>1</sub> — GymA<sub>6</sub> with AlbC<sup>[8]</sup> to determine the key residues for the substrate binding pockets P1 and P2 (Table S3) as reported previously.<sup>[9;10]</sup>



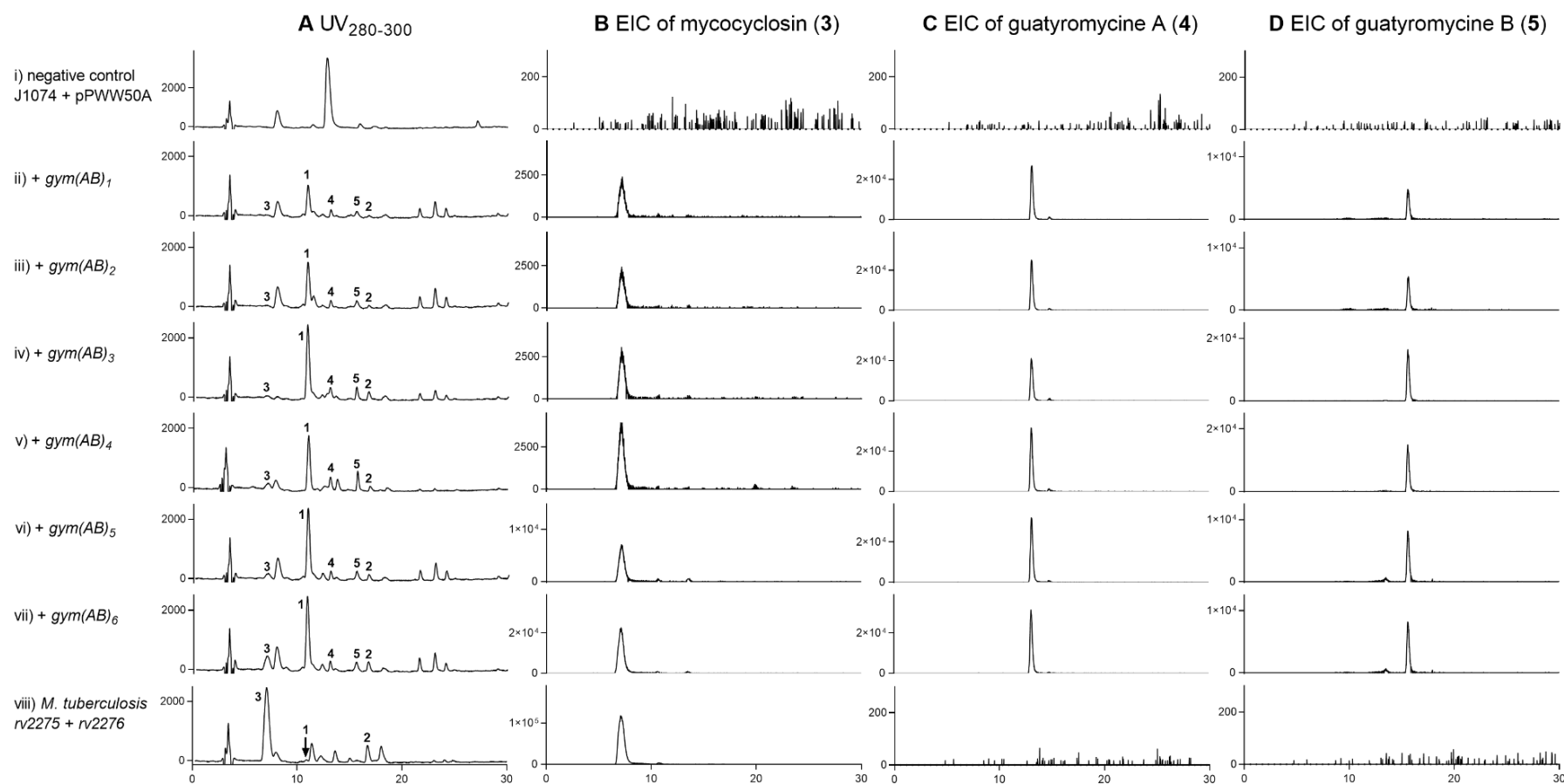
**Figure S4.** HPLC analysis of *E. coli* extracts harboring *gymA*<sub>1</sub> — *gymA*<sub>6</sub>.

The integrity of **1** (cYY) and **2** (cYF) was confirmed by <sup>1</sup>H NMR analysis after isolation (Table S6).

## PUBLICATIONS

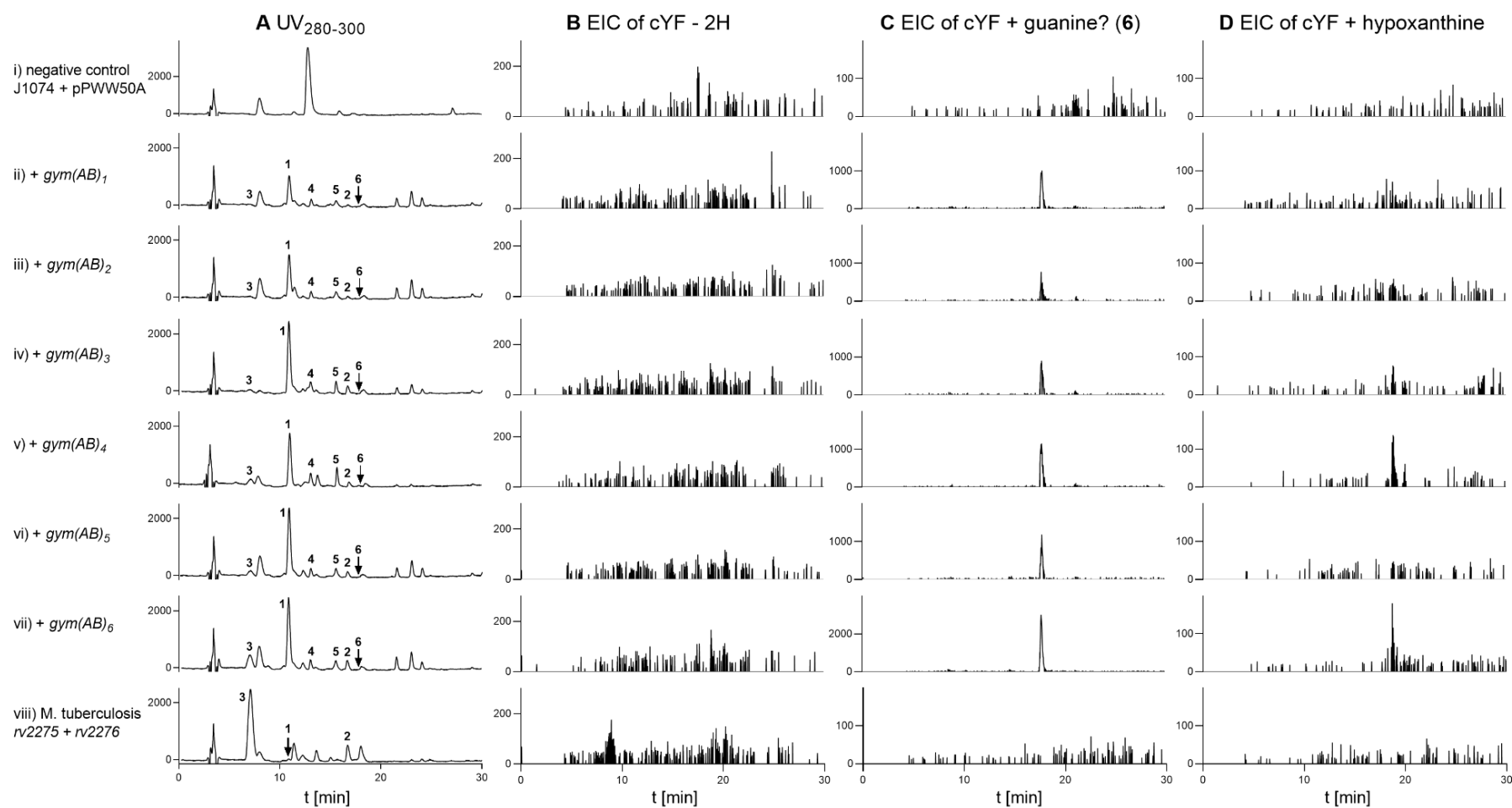


**Figure S5.** UV spectra of 1 – 5



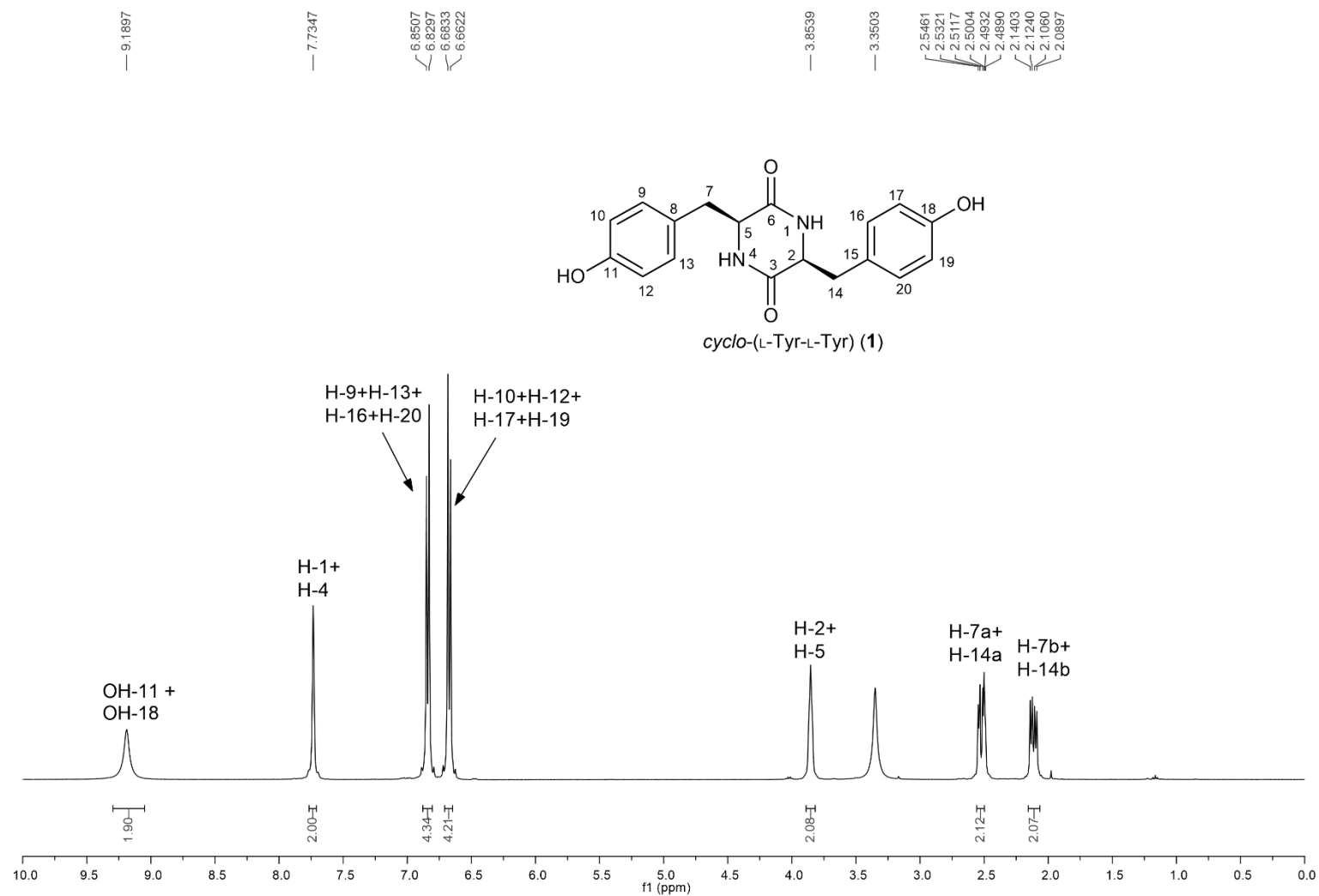
**Figure S6.** LC-MS analysis of extracts of *S. albus* J1074 harboring *gym(AB)*<sub>1</sub> – *gym(AB)*<sub>6</sub> and *rv2275/rv2276*, detected by UV absorptions at 280 – 300 nm and extracted ion chromatography (EICs) for **3**, **4**, and **5**.

# PUBLICATIONS



**Figure S7.** LC-MS analysis of extracts of *S. albus* J1074 harboring *gym*(AB)<sub>1</sub> – *gym*(AB)<sub>6</sub> and *rv2275/rv2276*, detected by UV absorptions at 280 – 300 nm and extracted ion chromatograms (EICs) for potential cYF derivatives, e.g. cYF - 2H, cYF + guanine (6), and cYF + hypoxanthine

# PUBLICATIONS



**Figure S8.** <sup>1</sup>H NMR spectrum of cYY (1) in DMSO-*d*<sub>6</sub>.



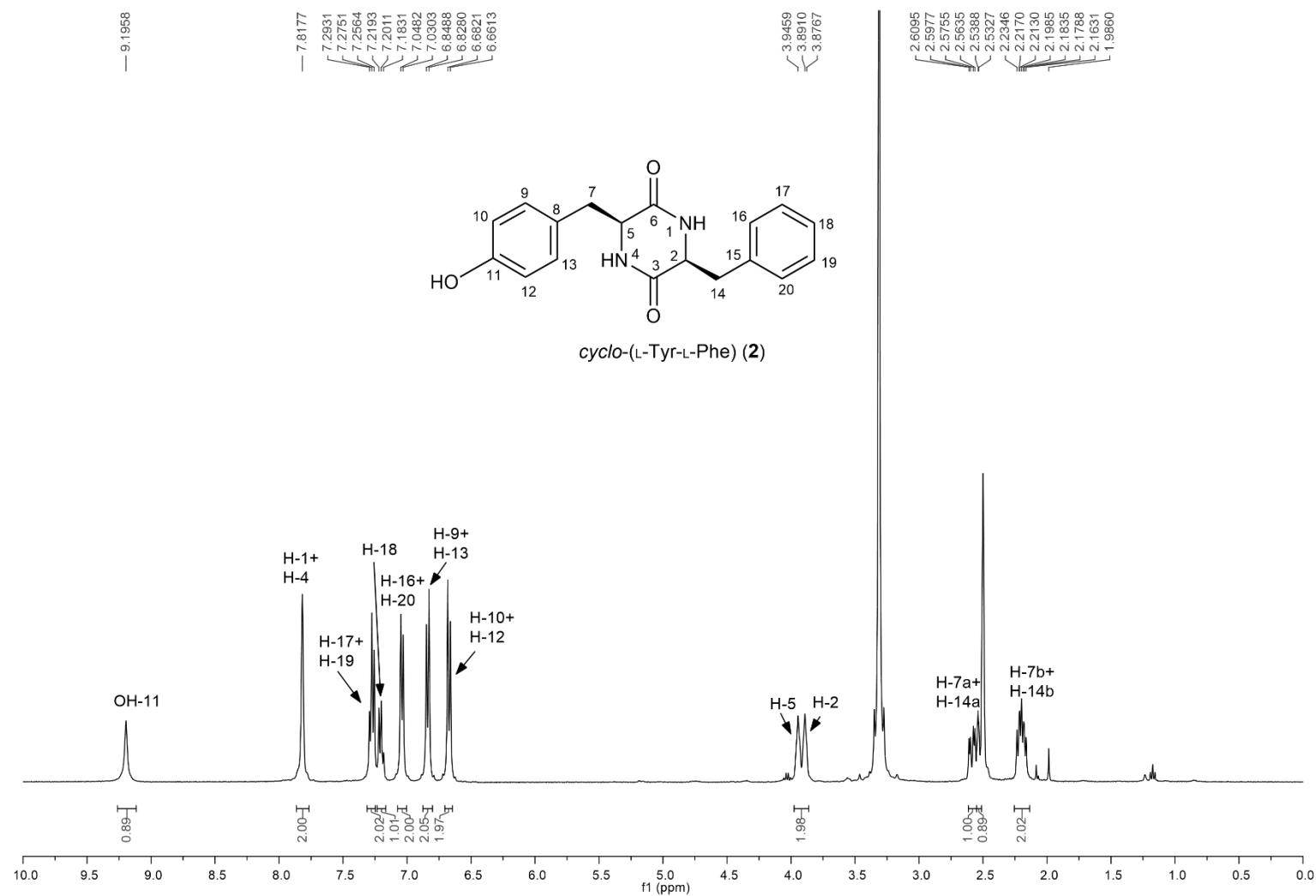
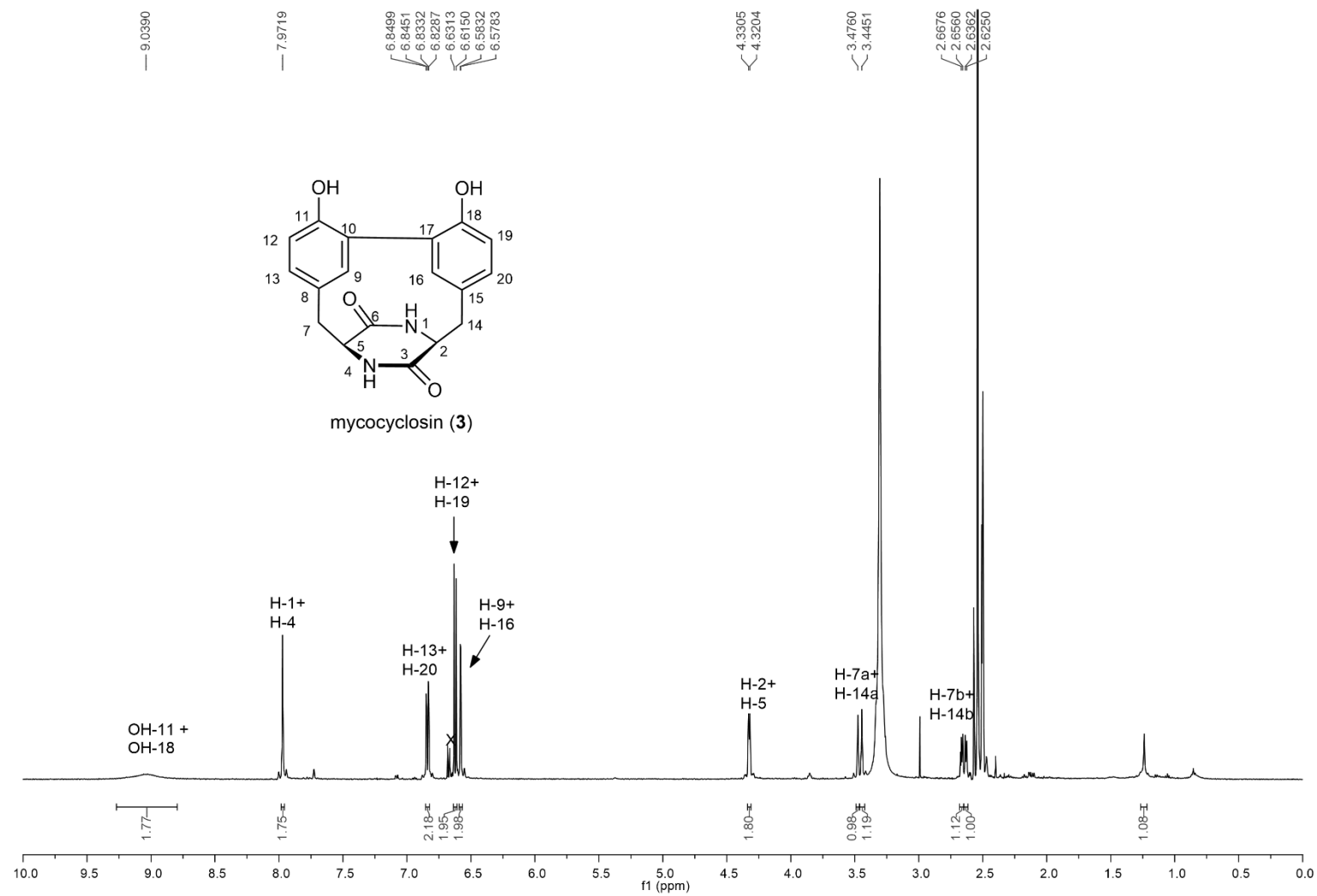
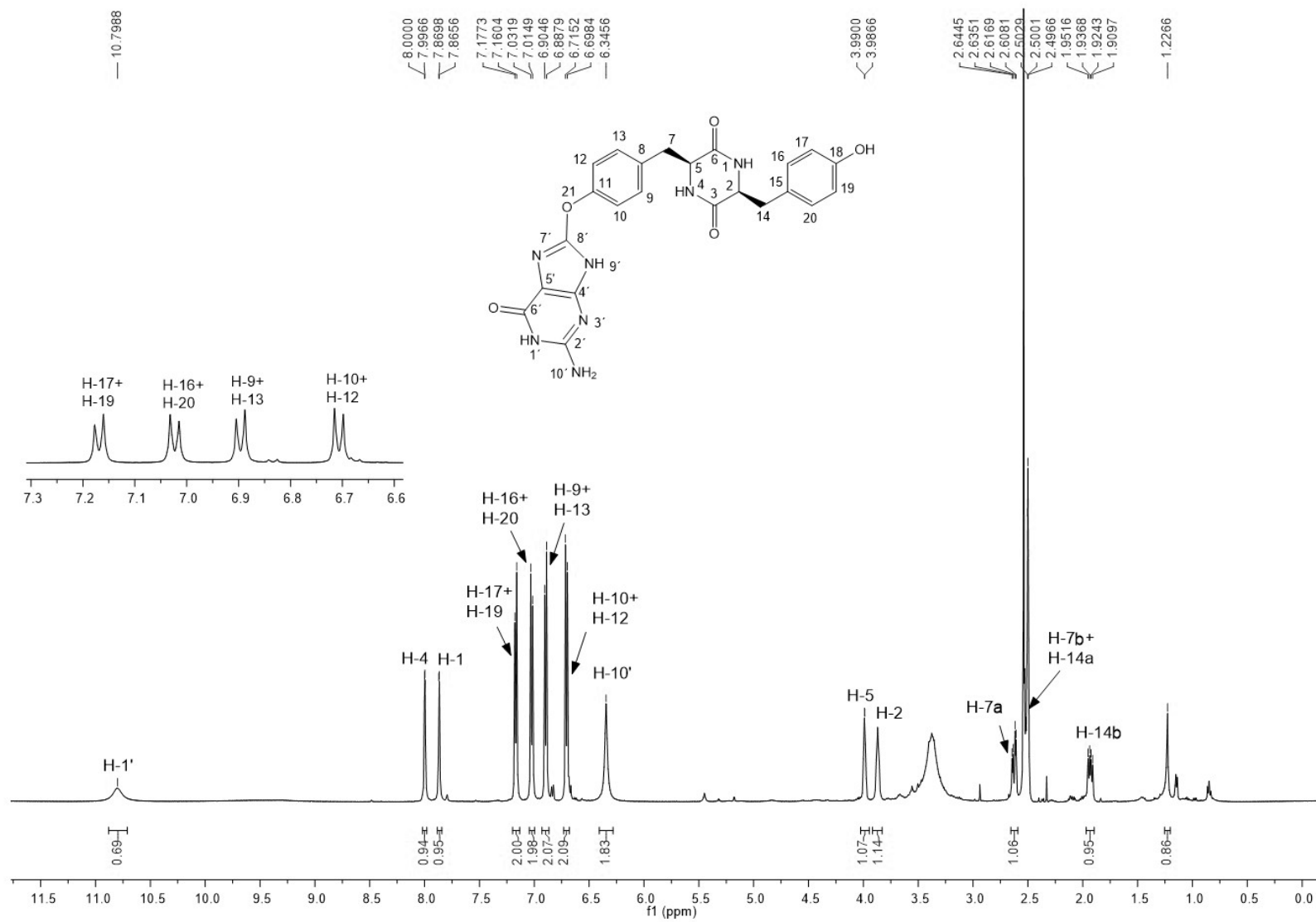


Figure S9.  $^1\text{H}$  NMR spectrum of cYF (**2**) in  $\text{DMSO}-d_6$ .

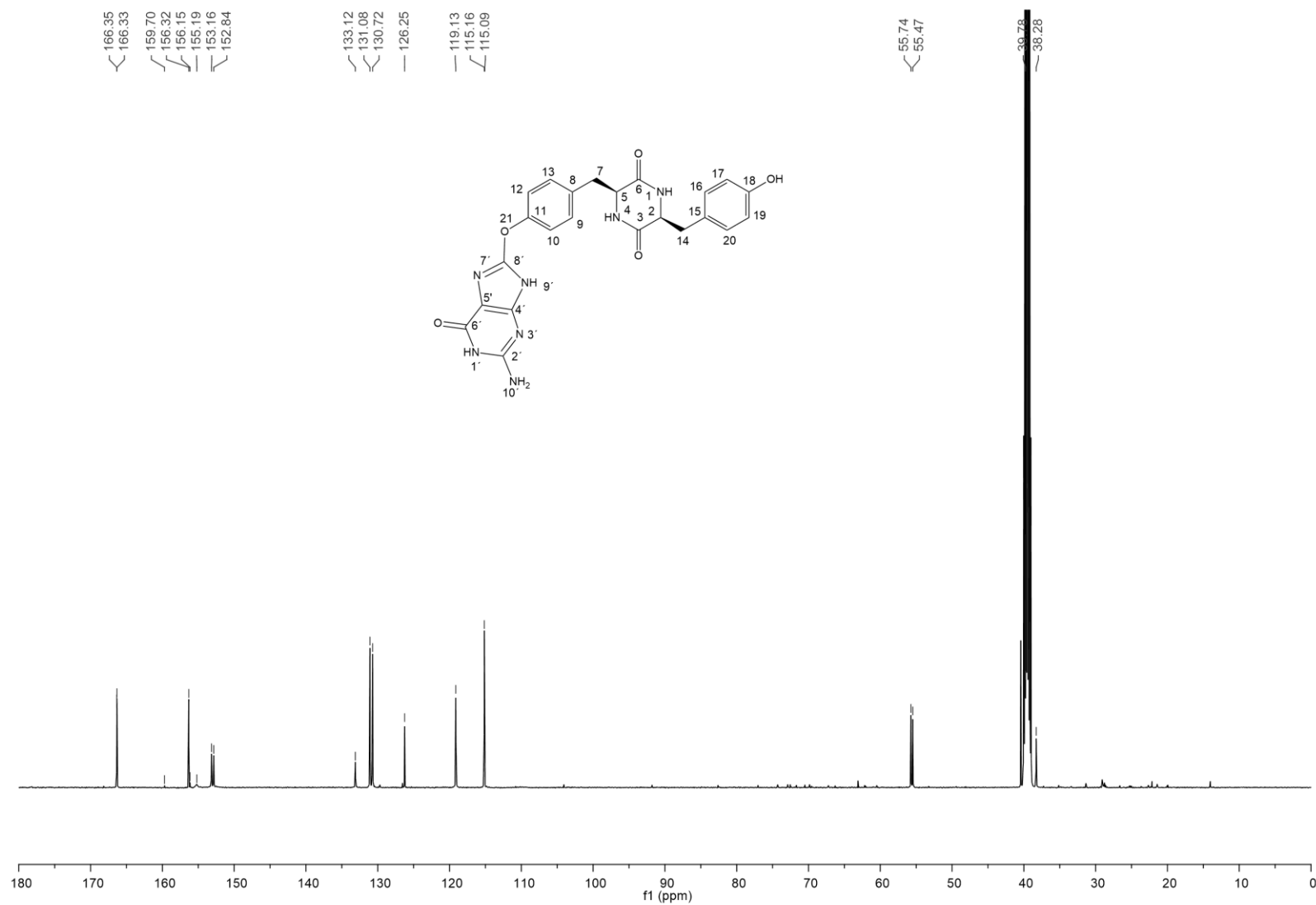


**Figure S10.**  $^1\text{H}$  NMR spectrum of mycocyclusin (3) in  $\text{DMSO}-d_6$ .

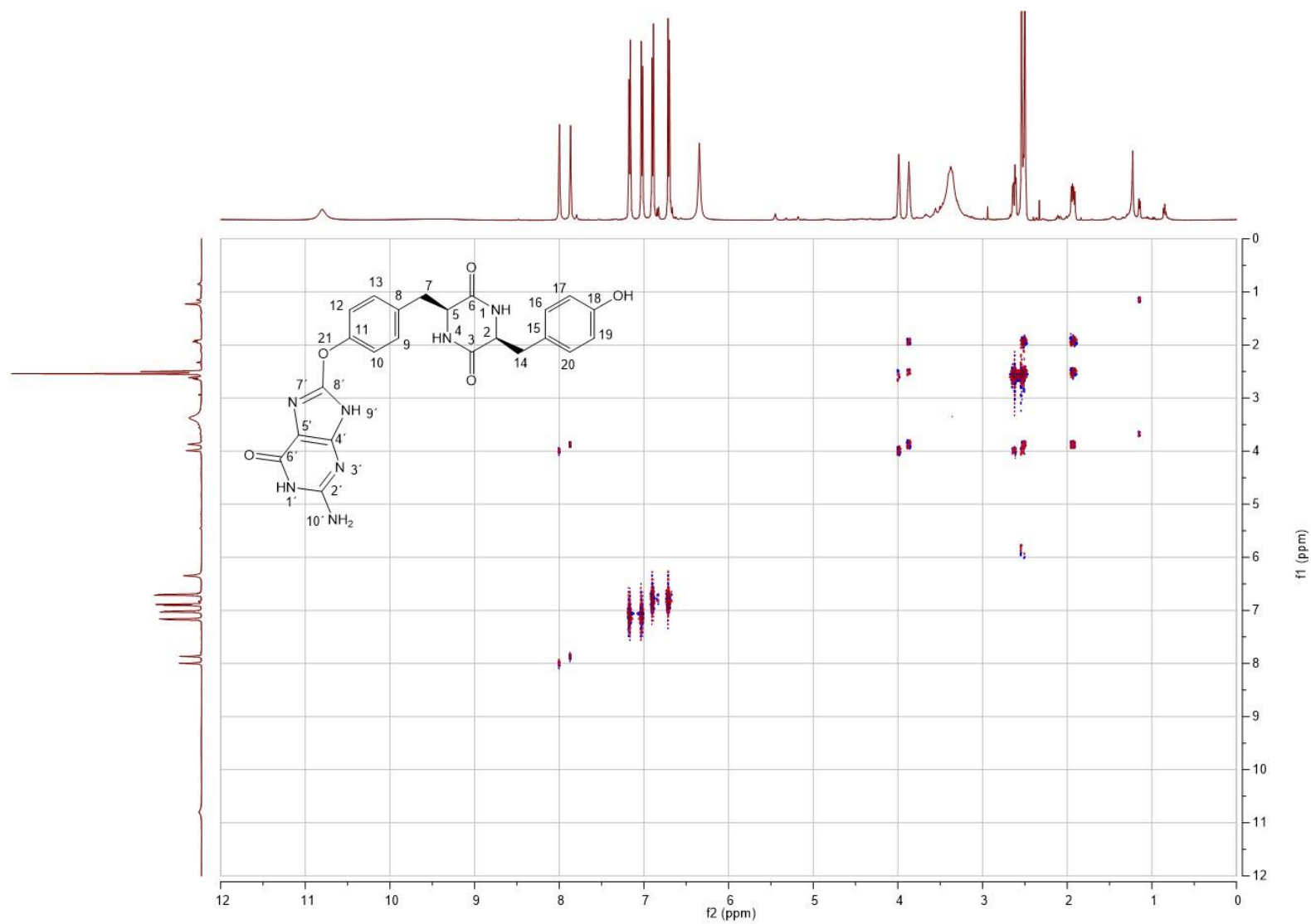


**Figure S11.** <sup>1</sup>H NMR spectrum of guatyromycine A (4) in DMSO-*d*<sub>6</sub>.

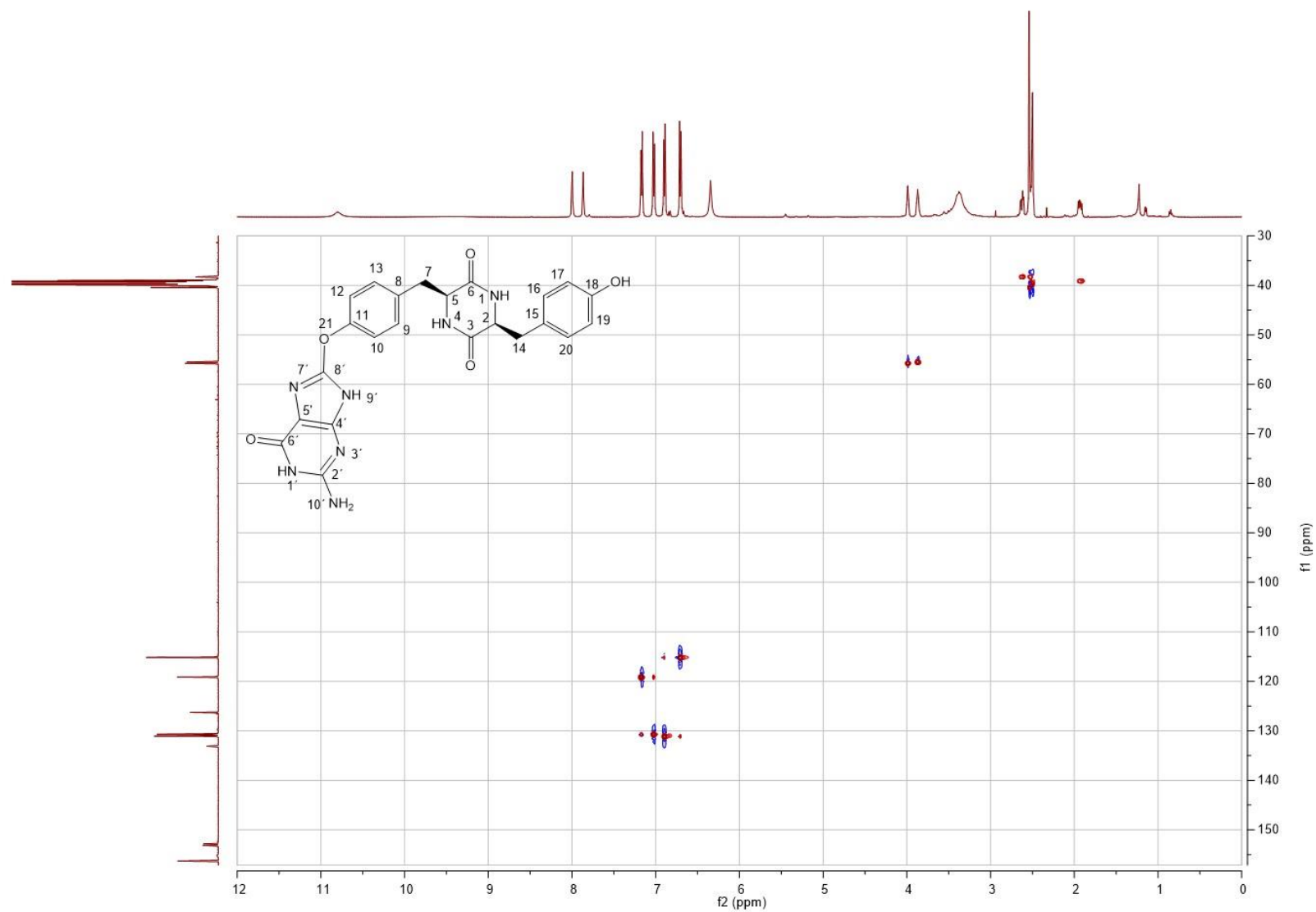
# PUBLICATIONS



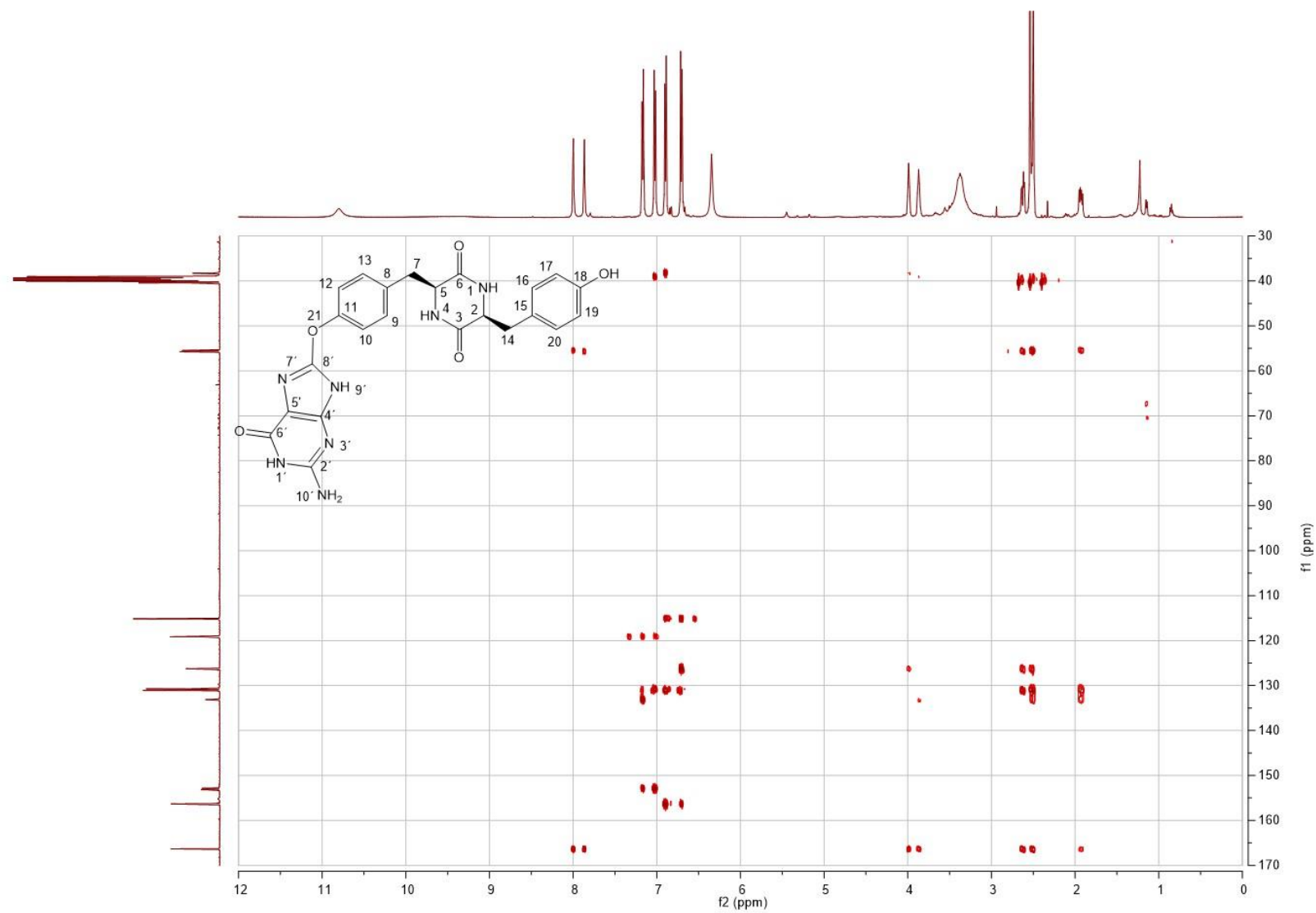
**Figure S12.**  $^{13}\text{C}$  NMR spectrum of guatyromycine A (4) in  $\text{DMSO-}d_6$ .



**Figure S13.**  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of guatyromycine A (4) in  $\text{DMSO}-d_6$ .

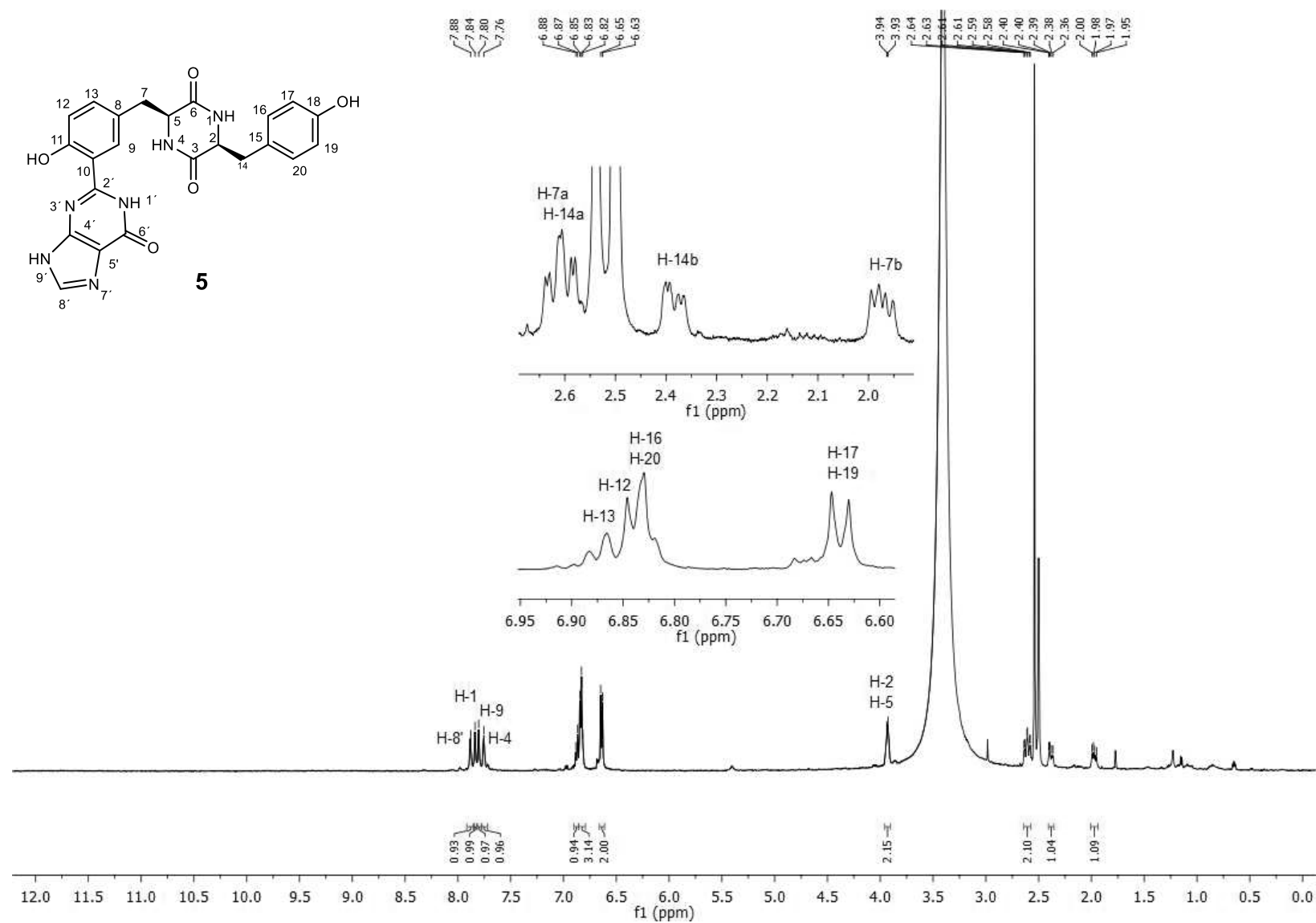


**Figure S14.** HSQC spectrum of guatyromycine A (4) in  $\text{DMSO}-d_6$ .



**Figure S15.** HMBC spectrum of guatyromycine A (4) in DMSO- $d_6$ .

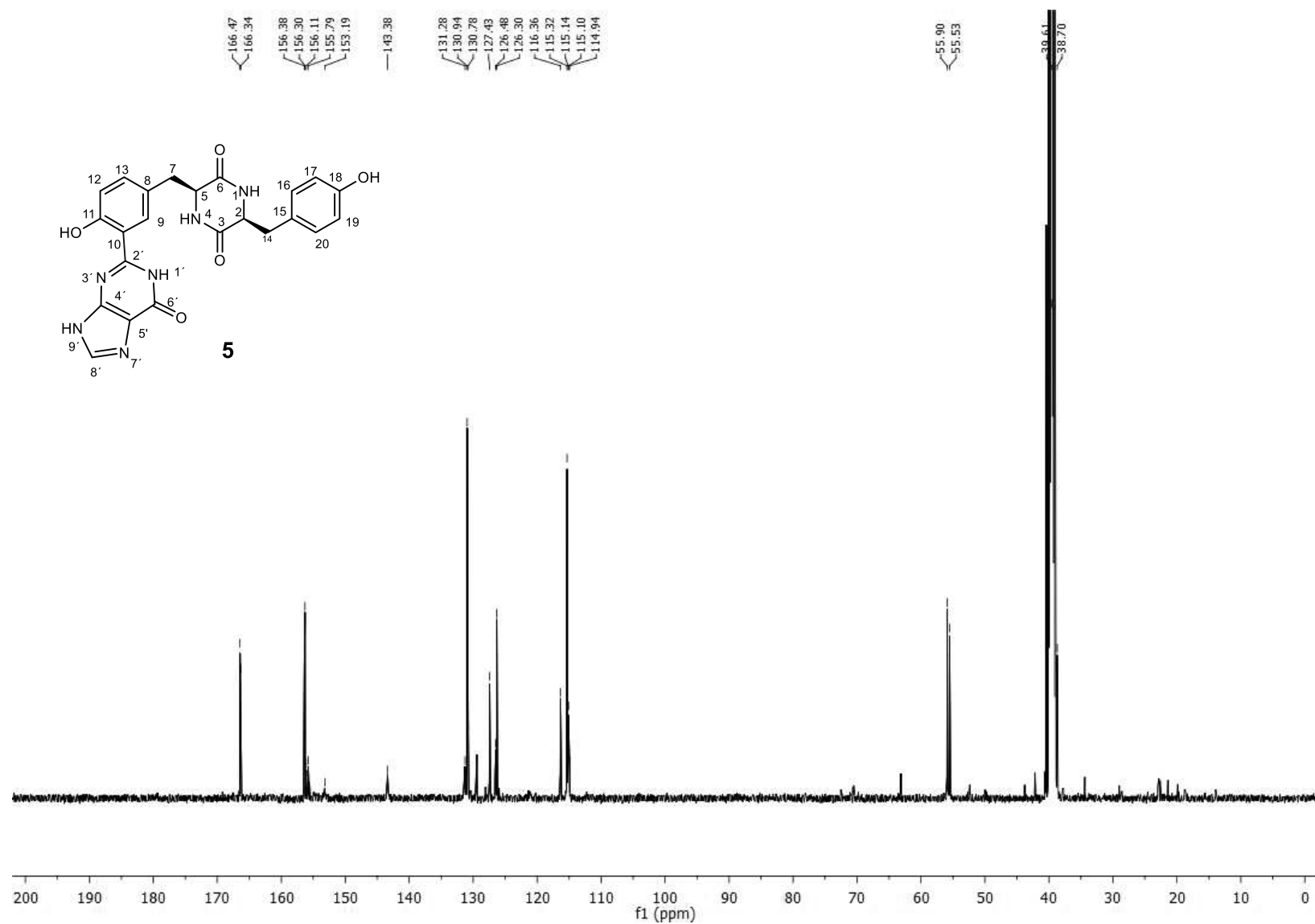
# PUBLICATIONS



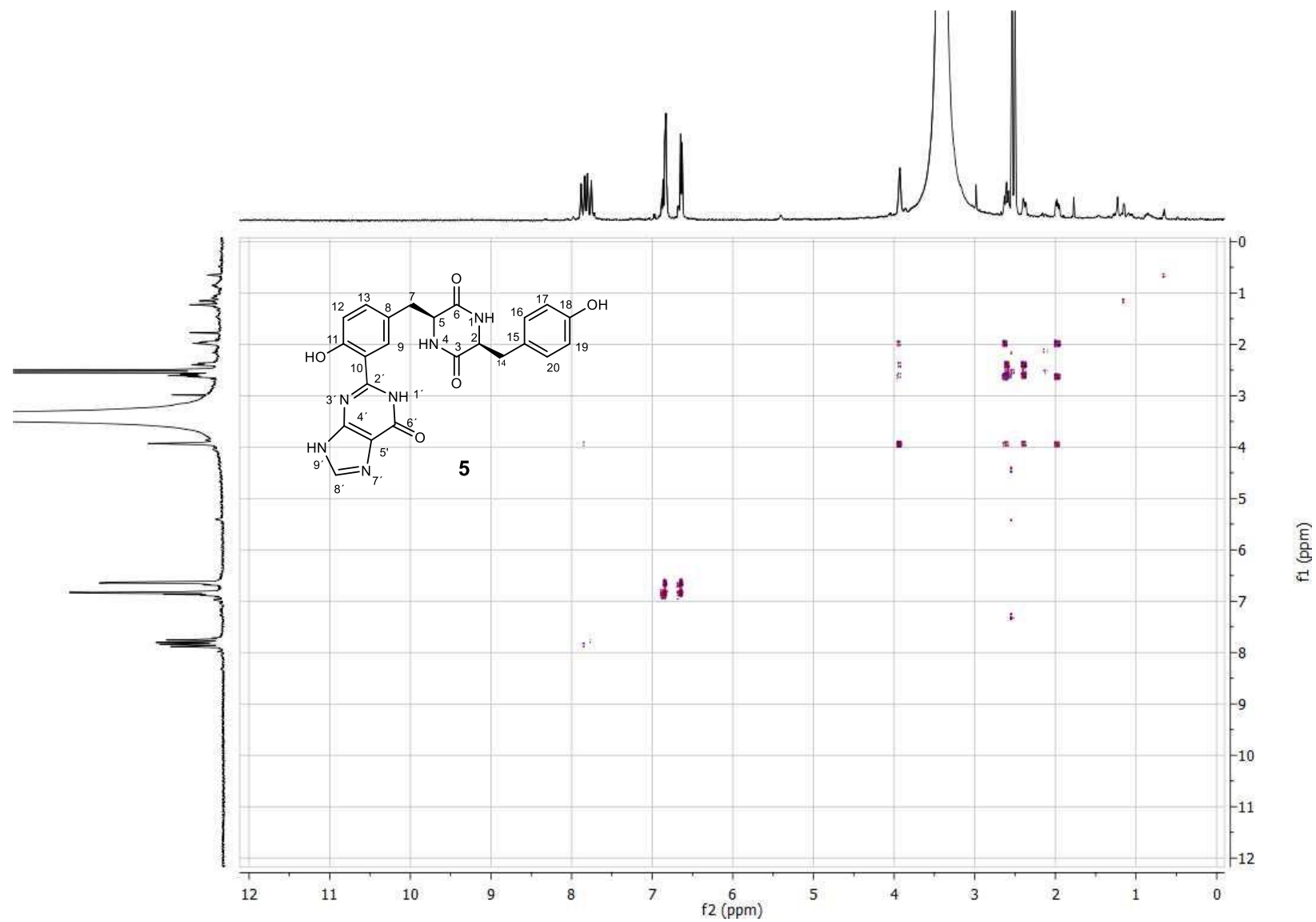
**Figure S16.** <sup>1</sup>H NMR spectrum of guatyromycine B (**5**) in DMSO-*d*<sub>6</sub>.



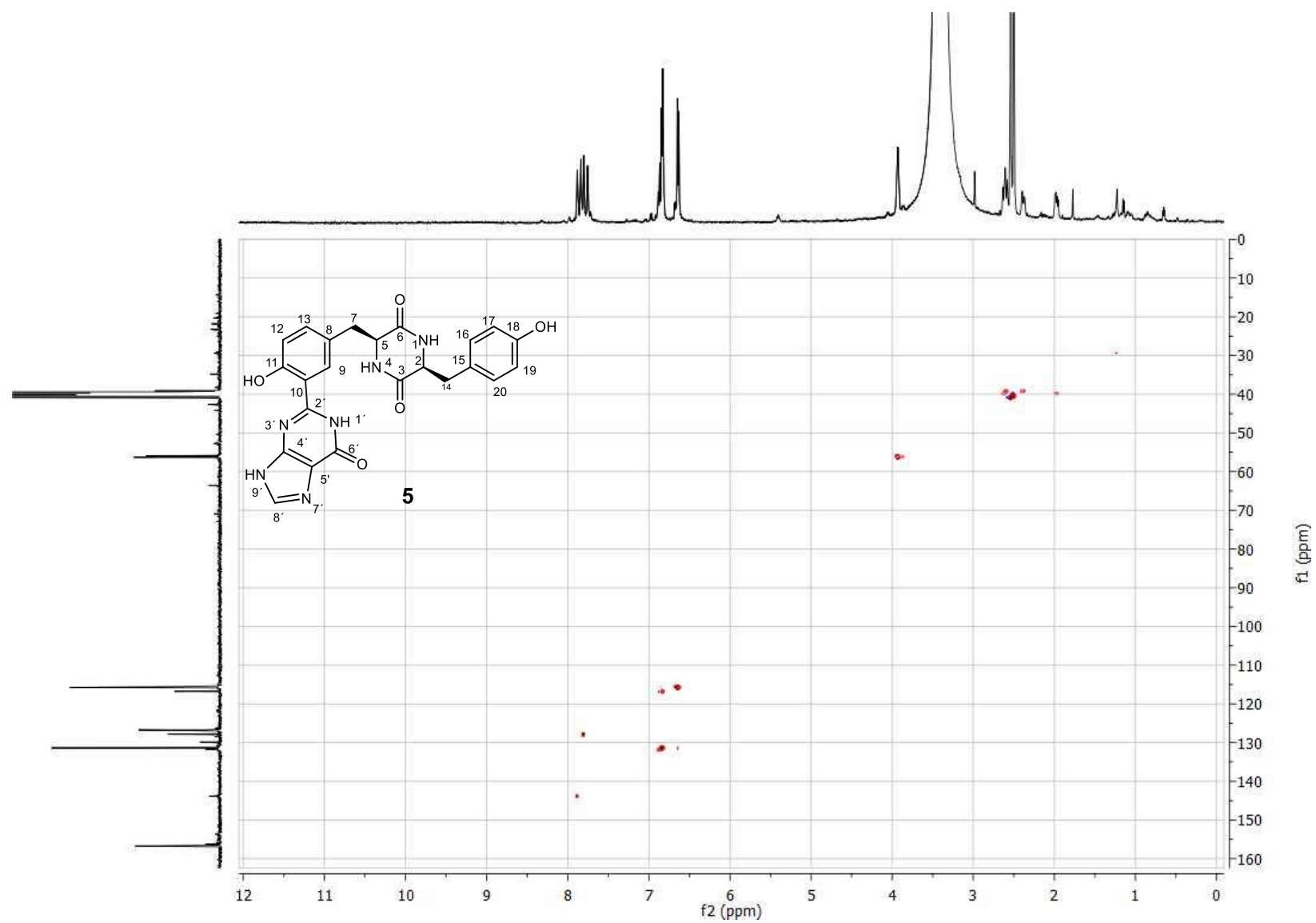
# PUBLICATIONS



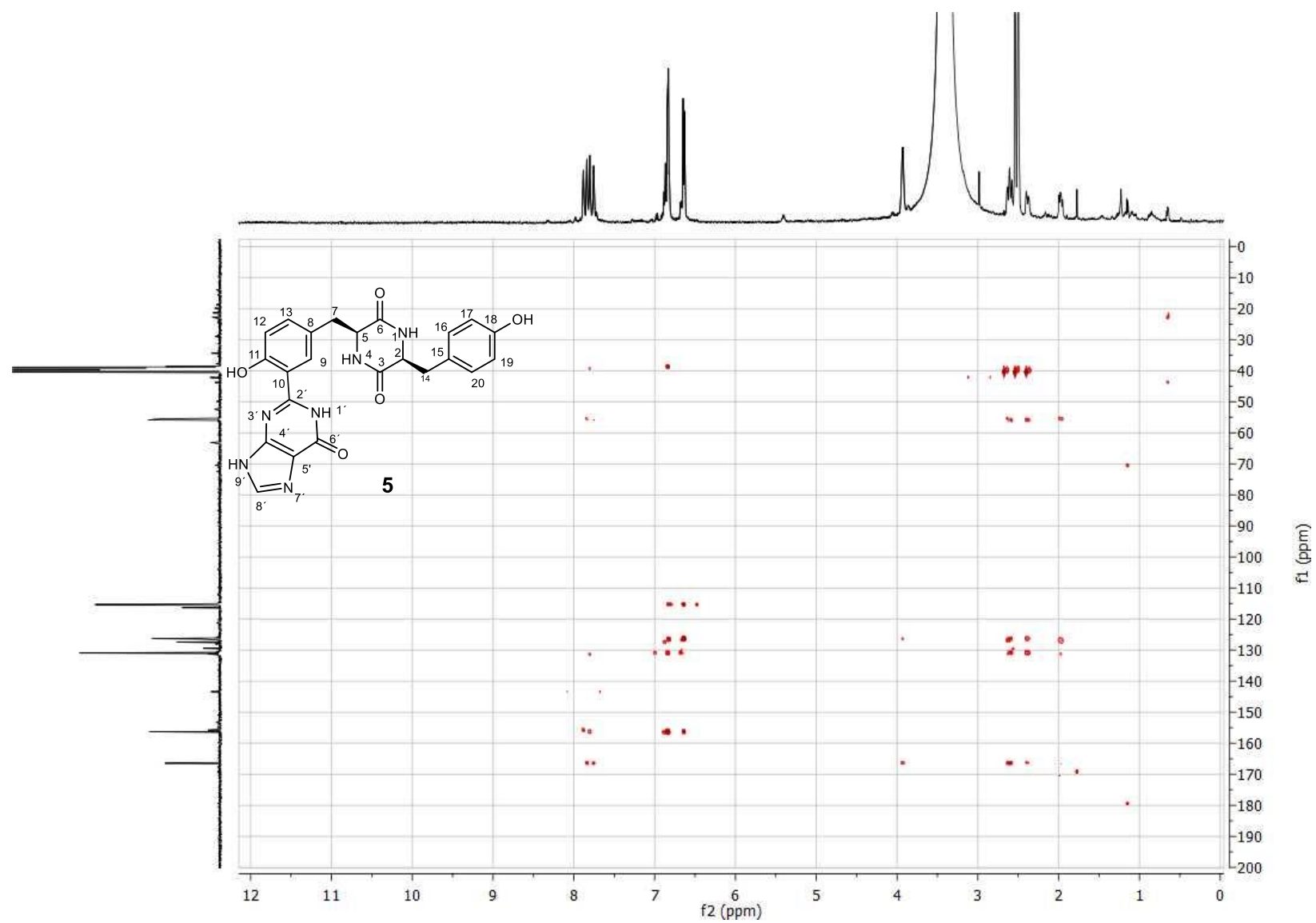
**Figure S17.**  $^{13}\text{C}$  NMR spectrum of guatyromycine B (5) in  $\text{DMSO}-d_6$ .



**Figure S18.**  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of guatyromycine B (5) in  $\text{DMSO}-d_6$ .

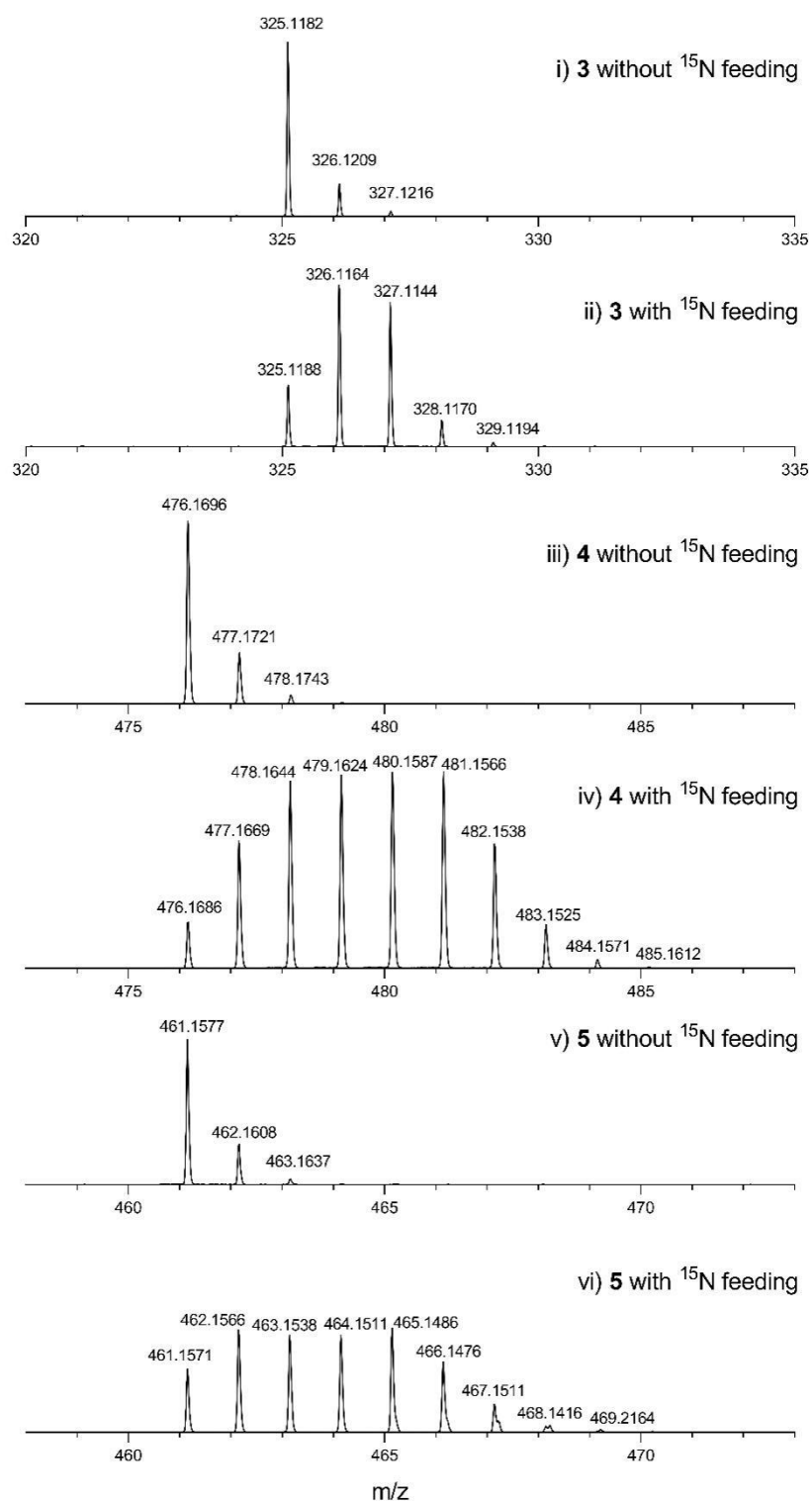


**Figure S19.** HSQC spectrum of guatyromycine B (**5**) in DMSO- $d_6$ .

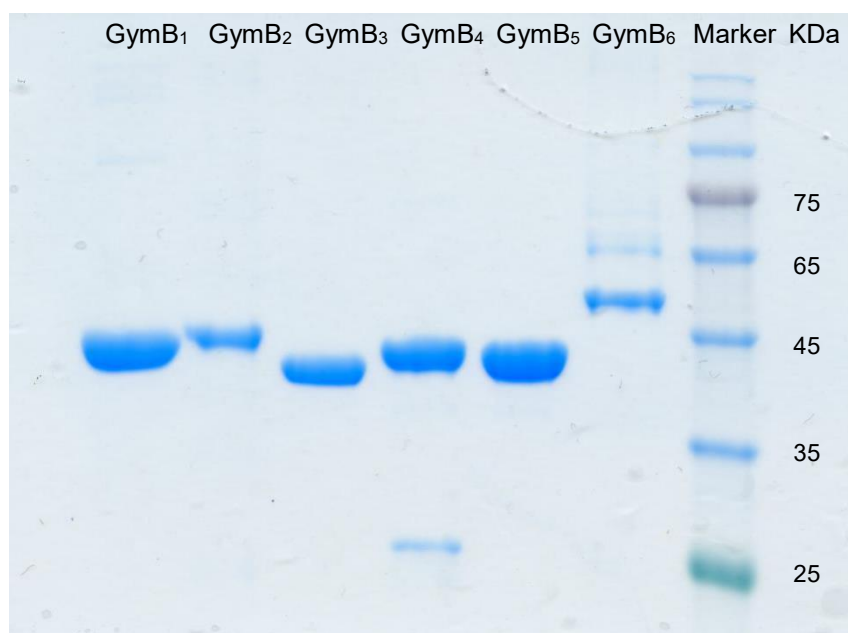


**Figure S20.** HMBC spectrum of guatyromycine B (5) in DMSO- $d_6$ .

# PUBLICATIONS

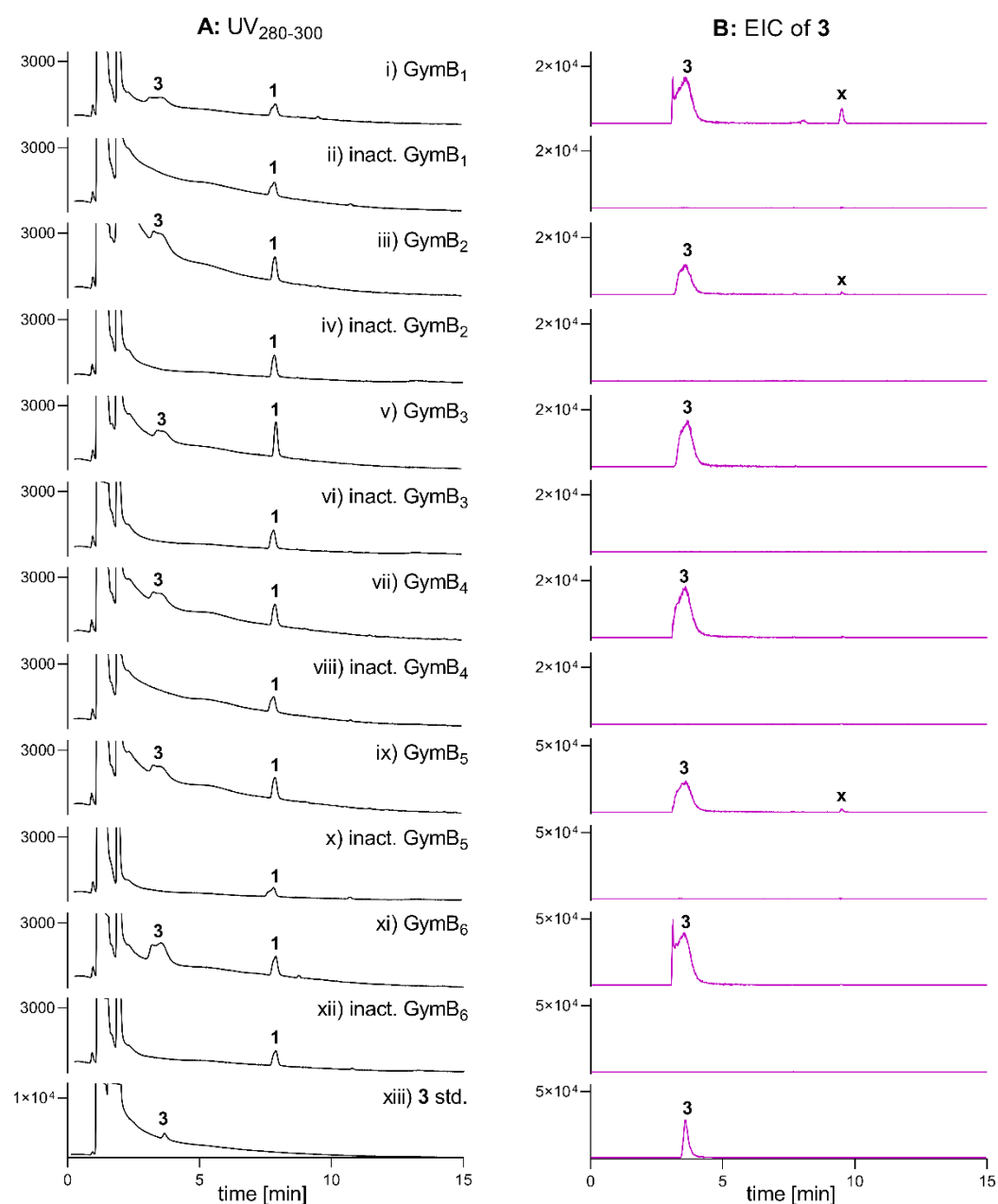


**Figure S21.** MS analysis of extracts of *S. albus* J1074 harboring *gymA<sub>6</sub>B<sub>6</sub>* cultivated in MR5 medium with and without  $^{15}\text{NH}_4\text{Cl}$ .



**Figure S22.** SDS-PAGE analysis of the purified P450s

Due to insolubility of the actual GymB<sub>6</sub>, its sequence was enlarged to 465 amino acids instead of 402 by addition of upstream sequences. Therefore, its protein mass is significant larger than those of other GymBs.



**Figure 23.** LC-MS analysis of cYY assays with GymB<sub>1</sub> – GymB<sub>6</sub> without nucleobases.

EIC of **3** refers the  $[M + H]^+$  ion at  $m/z$  325.118 with a tolerance range of  $\pm 0.005$ .

x: unknown mycocyclusin isomer.

## References

- [1] N. Zaburanyi, M. Rabyk, B. Ostash, V. Fedorenko, A. Luzhetskyy, *Bmc Genomics* **2014**, 15, 97.
- [2] M. R. Green, J. Sambrook, *Molecular cloning: a laboratory manual*, 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York **2012**.
- [3] T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, D. A. Hopwood, *Practical Streptomyces Genetics*, 2nd ed. John Innes Foundation, Norwich, UK **2000**.
- [4] Y. Zhu, P. Fu, Q. Lin, G. Zhang, H. Zhang, S. Li, J. Ju, W. Zhu, C. Zhang, *Org Lett.* **2012**, 14, 2666-2669.
- [5] P. Belin, M. H. Le Du, A. Fielding, O. Lequin, M. Jacquet, J. B. Charbonnier, A. Lecoq, R. Thai, M. Courcon, C. Masson, C. Dugave, R. Genet, J. L. Pernodet, M. Gondry, *Proc. Natl. Acad. Sci. U. S. A* **2009**, 106, 7426-7431.
- [6] H. Yu, X. Xie, S.-M. Li, *Org. Lett.* **2019**, 21, 9104-9108.
- [7] B. Gust, G. L. Challis, K. Fowler, T. Kieser, K. F. Chater, *Proc. Natl. Acad. Sci. U. S. A* **2003**, 100, 1541-1546.
- [8] S. Lautru, M. Gondry, R. Genet, J. L. Pernodet, *Chem Biol.* **2002**, 9, 1355-1364.
- [9] I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, J. L. Pernodet, M. Gondry, P. Belin, *Nat. Chem Biol.* **2015**, 11, 721-727.
- [10] M. Gondry, I. B. Jacques, R. Thai, M. Babin, N. Canu, J. Seguin, P. Belin, J. L. Pernodet, M. Moutiez, *Front Microbiol.* **2018**, 9, 46.
- [11] H. Yu, X. Xie, S.-M. Li, *Org Lett.* **2018**, 20, 4921-4925.
- [12] J. Liu, X. Xie, S.-M. Li, *Angew. Chem. Int. Ed. Engl.* **2019**, 58, 11534-11540.
- [13] J. Shi, X. Xu, E. J. Zhao, B. Zhang, W. Li, Y. Zhao, R. H. Jiao, R. X. Tan, H. M. Ge, *Org. Lett.* **2019**, 21, 6825-6829.
- [14] V. V. Shende, Y. Khatri, S. A. Newmister, J. N. Sanders, P. Lindovska, F. Yu, T. J. Doyon, J. Kim, K. N. Houk, M. Movassaghi, D. H. Sherman, *J. Am. Chem. Soc.* **2020**, 142, 17413-17424.
- [15] H. Yu, S.-M. Li, *Org. Lett.* **2019**, 21, 7094-7098.
- [16] J. Liu, X. Xie, S.-M. Li, *Chem. Commun.* **2020**, 56, 11042-11045.
- [17] S. Meng, W. Han, J. Zhao, X. H. Jian, H. X. Pan, G. L. Tang, *Angew. Chem Int. Ed. Engl.* **2018**, 57, 719-723.
- [18] M. J. Cryle, S. G. Bell, I. Schlichting, *Biochemistry* **2010**, 49, 7282-7296.



**4.3 Elucidation of the Streptoazine Biosynthetic Pathway in *Streptomyces aurantiacus* Reveals the Presence of a Promiscuous Prenyltransferase/Cyclase**

## PUBLICATIONS

# Elucidation of the Streptoazine Biosynthetic Pathway in *Streptomyces aurantiacus* Reveals the Presence of a Promiscuous Prenyltransferase/Cyclase

Jing Liu,<sup>§</sup> Yiling Yang,<sup>§</sup> Lauritz Harken, and Shu-Ming Li\*



Cite This: *J. Nat. Prod.* 2021, 84, 3100–3109



Read Online

ACCESS |



Metrics & More

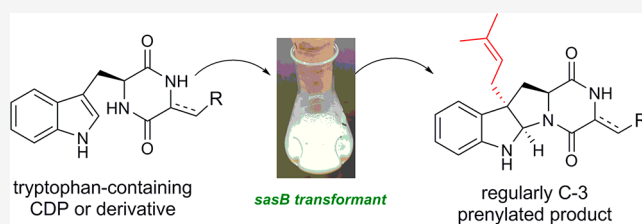


Article Recommendations



Supporting Information

**ABSTRACT:** Heterologous expression of a three-gene cluster from *Streptomyces aurantiacus* coding for a cyclodipeptide synthase, a prenyltransferase, and a methyltransferase led to the elucidation of the biosynthetic steps of streptoazine C (2). *In vivo* biotransformation experiments proved the high flexibility of the prenyltransferase SasB toward tryptophan-containing cyclodipeptides for regular C-3-prenylation. Furthermore, their corresponding dehydrogenated derivatives prepared by using cyclodipeptide oxidases were also used for prenylation. This study provides an enzyme with high substrate promiscuity from a less explored group of prenyltransferases for potential use to generate prenylated derivatives.



2,5-Diketopiperazine (DKP) alkaloids with an indole or indoline ring and isoprenoid moieties are derived from tryptophan-containing cyclodipeptides (CDPs).<sup>1–3</sup> They represent an important class of hybrid natural products and display diverse biological and pharmacological activities, including antibacterial, antitumor, anti-inflammatory, and insecticidal effects.<sup>2–4</sup> Representatives of tryptophan-containing CDP derivatives with various amino acids and one or more dimethylallyl (C<sub>5</sub>) moieties at different positions of the indole or indoline ring are shown in Chart 1. Okaramin C,<sup>5,6</sup> fellutanine D,<sup>7,8</sup> fructigenine A,<sup>9</sup> fumitremorgin B,<sup>10</sup> roquefortine E,<sup>11</sup> and echinulin<sup>12</sup> are examples of a large number of fungal products. In comparison, only a limited number of prenylated DKP derivatives, such as nocardiozine A,<sup>13</sup> drimentine G,<sup>14</sup> and streptoazine C (2),<sup>15</sup> are bacterial metabolites.

Significant progress has been achieved recently regarding the understanding of the biosynthesis of prenylated CDPs and derivatives thereof, especially of those from fungi of the genera *Penicillium* and *Aspergillus*.<sup>16–20</sup> In nature, the 2,5-DKP scaffolds are usually biosynthesized by two distinct enzyme types, either by the nonribosomal peptide synthetases (NRPSs)<sup>18,21</sup> or by the cyclodipeptide synthases (CDPSs).<sup>22,23</sup> NRPSs are modular multidomain enzyme complexes and incorporate free amino acids to form the final peptide products.<sup>24</sup> Bimodular NRPSs are responsible for the formation of CDPs.<sup>16,24</sup> In contrast, CDPs, mostly of bacterial origin, directly hijack aminoacyl-tRNAs from the protein biosynthesis as substrates to form the DKP scaffolds.<sup>25</sup> The DKP scaffolds can be further modified by diverse tailoring enzymes including prenyltransferases (PTs), methyltransferases (MTs), and cytochrome P450 enzymes.<sup>16,17,25,26</sup>

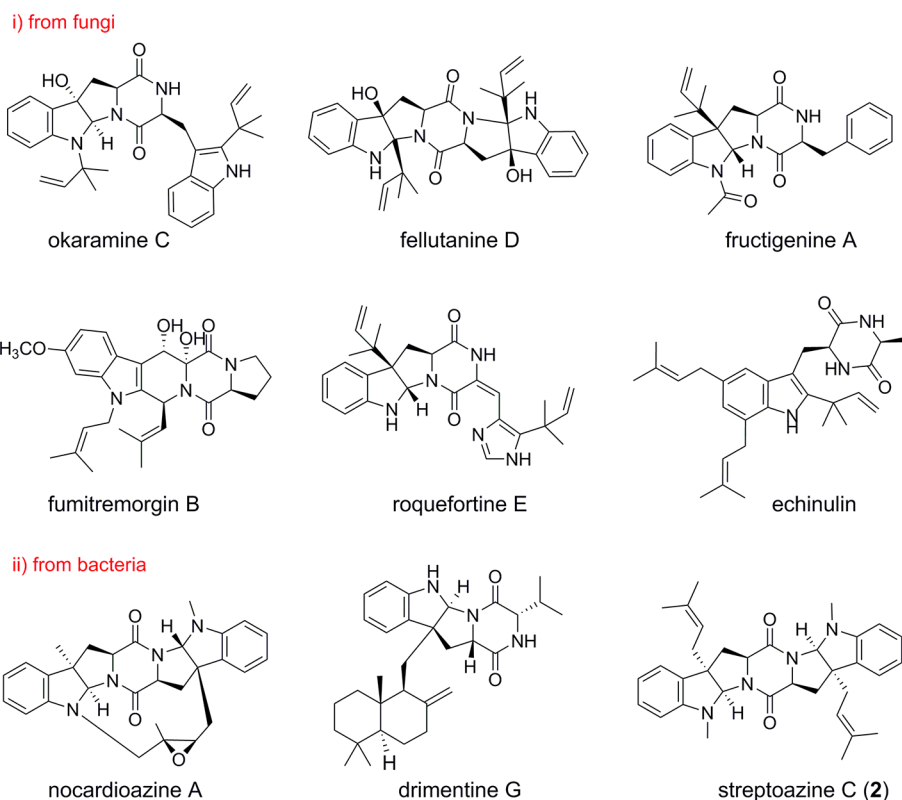
Prenylation by PTs at different positions of the indole ring of tryptophan-containing CDPs plays a key role for structural diversification of indole alkaloids and is involved in the biosynthesis of a large number of CDP derivatives.<sup>1,16</sup> So far, most of the prenylated CDP alkaloids have been identified from NRPS-dependent pathways in fungi, and at least 13 fungal CDP PTs for regular (normal) or reverse prenylation were characterized biochemically.<sup>17,27</sup> PTs from bacteria and fungi are usually highly permissive and can use structurally distinct compounds for prenylation.<sup>19,28</sup> For example, a fungal PT for a given CDP shows high flexibility toward not only CDPs but also hydroxynaphthalenes and flavonoids.<sup>28</sup> Substrate and catalytic promiscuity were frequently reported for bacterial PTs as well. NphB in the biosynthesis of the naphterpin catalyzes a C-prenylation of hydroxynaphthalenes and can accept some simple phenols, phenylpropanoids, flavonoids, and stilbenes for O- and C-prenylation as well.<sup>29,30</sup> Until now, only two PTs from CDPS-dependent pathways have been described. Zhang et al.<sup>15</sup> recently reported the identification of a two-gene cluster from *Streptomyces leeuwenhoekii*, being responsible for the biosynthesis of streptoazine C (2) (Chart 1). In this pathway, *cyclo*-(L-Trp-L-Trp) (cWW), the product of the CDPS SazA, was further prenylated and methylated by the bifunctional enzyme SazB containing both PT and MT domains. It was reported that

Received: September 5, 2021

Published: November 30, 2021



Chart 1. Examples of Prenylated Diketopiperazine Derivatives



SazB-PT showed strict substrate specificity and accepted only dimethylallyl diphosphate and cWW as substrates.<sup>15</sup> The prenyltransferase DmtC1 from *Streptomyces youssoufiensis* is involved in the biosynthesis of drimenterines and catalyzes the C-3-farnesylation of *cyclo*-(L-Trp-L-Pro), *cyclo*-(L-Trp-L-Val), *cyclo*-(L-Trp-L-Leu), and *cyclo*-(L-Trp-L-Ile).<sup>14</sup> However, all of these compounds are CDPS products of the drimenterine pathway, i.e., the natural substrates of DmtC1. The sharp contrast between the high flexibility of fungal CDP PTs and the strict substrate specificity of the bacterial SazB-PT encouraged us to investigate more PTs from CDPS-dependent pathways.

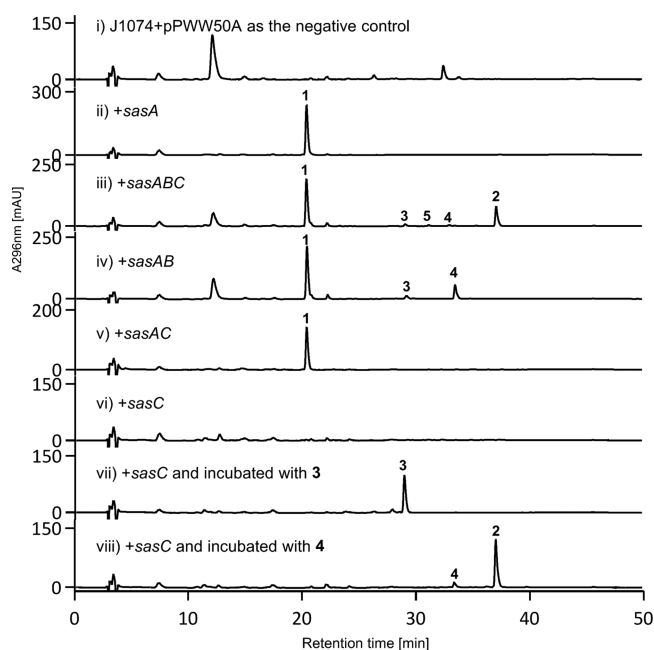
In this study, we identified by genome mining a putative *cdps*-containing gene cluster with one gene for a PT and an additional gene for an MT. Heterologous expression provided evidence for their roles in the biosynthesis of streptoazine C (2). Furthermore, we proved via biotransformation that the prenyltransferase SasB was able to prenylate diverse tryptophan-containing cyclodipeptides and their dehydrogenated derivatives, which highlights its potential as a useful biocatalyst to generate diverse prenylated DKPs.

## RESULTS AND DISCUSSION

**Identification and Analysis of the Putative *sas* Gene Cluster.** Genome mining and heterologous expression in a well-characterized host have been proven to be an efficient strategy to explore the silent/cryptic biosynthetic potential for natural product production.<sup>31–33</sup> Using this strategy, we have successfully identified several new metabolites from different CDPS-associated biosynthetic pathways and characterized intriguing chemical reactions thereof, including the novel nucleobase-containing alkaloid guanitrypmycins and several dimeric DKPs with distinct linkage patterns.<sup>26</sup> In analogy, we

analyzed a wide range of *cdps*-containing clusters by using characterized proteins as probes and identified a candidate from *S. aurantiacus* NRRL ISP-5412. The cluster of interest, termed the *sas* cluster, consists of three open reading frames coding for a putative CDPS (SasA, WP\_079103588.1) and two tailoring enzymes, SasB (WP\_121505431.1) and SasC (WP\_054413754.1) (Table S1). SasA with a polypeptide chain length of 252 amino acids shares a sequence identity of 82% on the amino acid level with SazA mentioned above (Table S1). SasB comprising 347 amino acids displays sequence identities of 85% and 38% with the known SazB-PT and DmtC1, respectively, indicating its role as a prenyltransferase. Phylogenetic analysis with functionally characterized PTs showed that SasB and SazB-PT are closely located to each other (Figure S1). The 290 amino acid bearing SasC has a high sequence identity of 82% with the MT domain of SazB. All of these data indicate that the two clusters probably evolved from the same ancestor and underwent diversification during the evolutionary process.

**Functional Proof of the Gene Cluster in the Biosynthesis of Streptoazine.** To verify their functions, we first cloned the *cdps* gene *sasA* from *S. aurantiacus* into pPWW50A<sup>34</sup> and expressed it in *Streptomyces albus* J1074 (Tables S2 and S3).<sup>35</sup> The obtained transformant harboring *sasA* was cultivated in modified RS media at 28 °C for 7 days. The bacterial culture was subsequently extracted with EtOAc and analyzed by LC-MS. In comparison to the host strain J1074 harboring pPWW50A (Figure 1-i), one predominant product (1) bearing a  $[M + H]^+$  ion at  $m/z$  373.1659 was detected (Figure 1-ii). Compound 1 was identified as cWW by comparison with an authentic standard, proving SasA to be a cWW synthase. Afterward, the whole gene cluster comprising *sasABC* was cloned into pPWW50A and overexpressed in



**Figure 1.** LC-MS analysis of *S. albus* J1074 transformants with and without precursors. Absorptions at UV 296 nm are illustrated.  $[M + H]^+$  ions with a tolerance range of  $\pm 0.005$  were detected at  $m/z$  373.166 for 1, 537.322 for 2, 441.229 for 3, 509.291 for 4, and 523.307 for 5, respectively. The  $[M + H]^+$  ion of the peak at 32.5 min in the negative control differs from that of 4.

J1074 as described above. In addition to the predominant 1, four new additional compounds were observed (Figure 1-iii). The second dominant product (2) was detected with a  $[M + H]^+$  ion at  $m/z$  537.3224, 164 Da larger than that of cWW, indicating the attachment of two prenyl and two methyl groups to 1. The three minor compounds 3, 4, and 5 with  $[M + H]^+$  ions at 441.2285, 509.2911, and 523.3068 are 68, 136, and 150 Da larger than 1, implying one prenyl, two prenyl, and two prenyl moieties plus one methyl group in their structures, respectively. Compound 2 was then isolated by semi-preparative HPLC after large-scale fermentation. Comprehensive interpretation and comparison of the  $^1\text{H}$  NMR data as well as the ECD spectrum (Table S4, Figures S2 and S59) with those reported in the literature<sup>15</sup> confirmed 2 to be

streptoazine C (Scheme 1). These data strongly support the function of SasB as a regular C-3-prenyltransferase and SasC as an indoline *N*-methyltransferase. Due to the low product yields, 3–5 could not be isolated from the *sasABC* transformant for structure elucidation by NMR analysis.

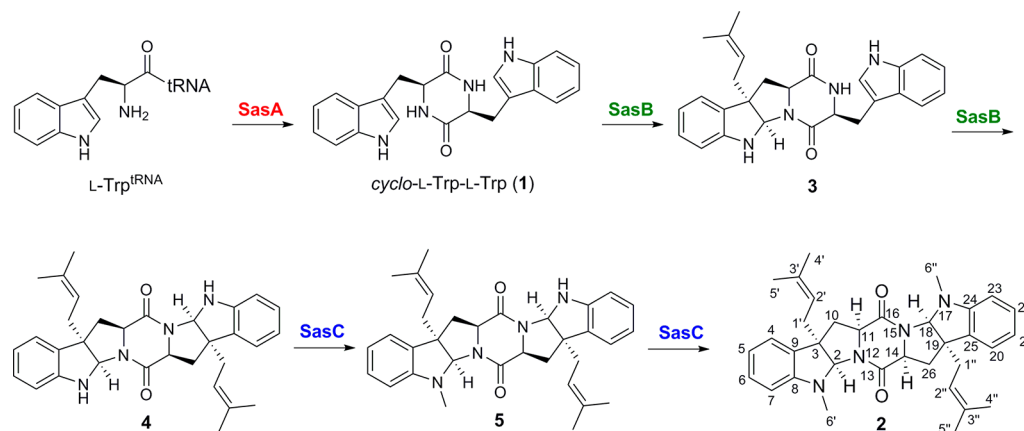
To confirm the SasB and SasC functions and figure out the reaction order, we performed the coexpression of *sasA* with *sasB* and *sasC* separately, that is, *sasAB* and *sasAC*. In addition to the CDPS product 1, two additional compounds, 3 and 4, were detected in the fermentation culture of the *sasAB* transformant (Figure 1-iv). Isolation and structure elucidation by MS and  $^1\text{H}$  NMR analyses as well as comparison with the data of known compounds<sup>15</sup> confirmed 3 and 4 to be regularly C-3 monoprenylated cWW and streptoazine A, respectively (Scheme 1, Table S4, and Figures S3 and S4). In contrast, only 1 was observed in the culture of the *sasAC* transformant (Figure 1-v). Neither mono- nor dimethylated 1 was detected in the *sasABC* transformant, even in the sensitive EIC chromatogram (data not shown). These results supported that 1 cannot be methylated by the methyltransferase SasC and prenylation and cyclization take place before methylation.

Incubating 4 with the *sasC* transformant led to the clear detection of 2 (Figure 1-vi), whereas no new peaks were observed in the culture after incubation with 3 (Figure 1-vii). This demonstrated that methylation proceeds only after the attachment of two prenyl moieties (Scheme 1). This is also the reason for the absence of methylated monoprenylated 1 in the *sasABC* transformant (Figure 1-iii). This order of reactions is the same as that recently reported for the two-gene cluster responsible for streptoazine C biosynthesis.<sup>15</sup>

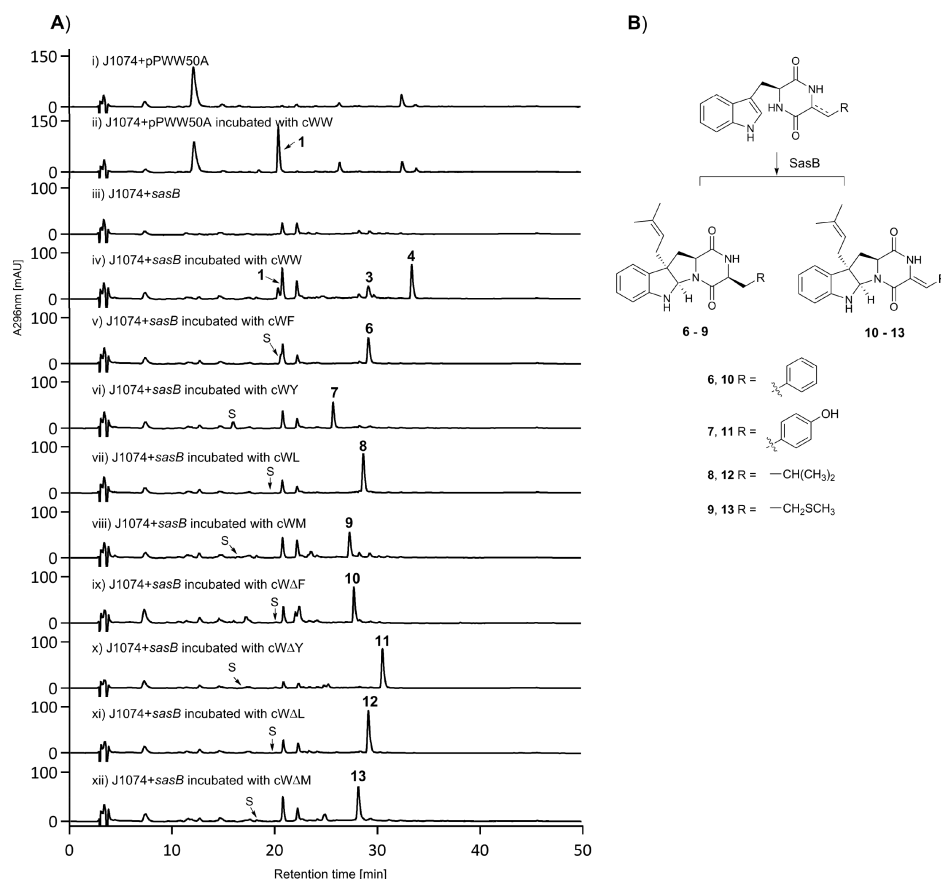
To further verify that the formation of 3 and 4 is catalyzed by SasB, its coding sequence was cloned into pPWW50A and expressed in J1074. In comparison to J1074 harboring the empty vector, neither 1 nor other additional metabolites were observed in the *sasB* transformant (Figure 2A-iii). Incubating the *sasB* transformant with 1 (100  $\mu\text{M}$ ) and cultivation for 5 days led to the identification of 3 and 4 (Figure 2A-iv), whereas no consumption of 1 was found in the control culture (Figure 2A-ii). These results demonstrated that SasB is able to catalyze the regular C-3 prenylation of 1.

**Substrate Promiscuity of SasB and Generation of Diverse Prenylated Tryptophan-Containing DKP Derivatives.** SasB acts as a C-3-prenyltransferase and complements

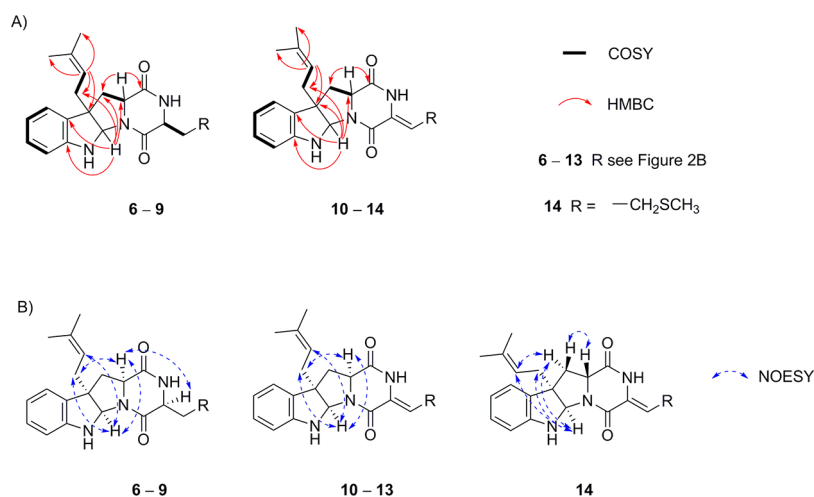
#### Scheme 1. Biosynthetic Pathway of Streptoazine in *S. aurantiacus*<sup>a</sup>



<sup>a</sup>The structures of streptoazine C (2), compound 3, and streptoazine A (4) were confirmed by NMR analysis.



**Figure 2.** (A) HPLC analysis of the *sasB* transformant with and without precursors and (B) prenylated products of SasB. The  $[M + H]^+$  ions of the peaks at 26.3 and 32.5 min in the negative control differ clearly from those of 7 and 4, respectively. S: substrate.



**Figure 3.** (A) Key COSY, and HMBC and (B) NOESY correlations in 6–14.

the reverse C-3-prenylation of fungal CDP PTs such as AnaPT, CdpNPT, and CdpC3PT<sup>17</sup> and the regular C-3-prenylation feature of FtmPT1.<sup>36</sup> Such an enzyme with flexible substrate specificity is welcome for its potential use in the production of prenylated DKPs.<sup>37</sup> We therefore investigated the substrate specificity of SasB. Because no recombinant protein was obtained after heterologous expression in *E. coli* and *Streptomyces* (data not shown), *in vitro* testing for the acceptance of tryptophan-containing CDPs by SasB cannot be achieved. Thus, we supplied 12 CDPs to the J1074

transformant with *sasB* and monitored their consumption by LC-MS analysis. As shown in Figure 2A–v–viii, cWF, cWY, cWL, and cWM were efficiently converted by SasB. The  $[M + H]^+$  ions of the products 6–9 are 68 Da larger than those of the precursors, indicating the attachment of one dimethylallyl moiety to the substrates.

Large-scale fermentation and subsequent isolation via preparative HPLC resulted in the products 6–9 of high purity for NMR analysis including <sup>1</sup>H, <sup>13</sup>C, COSY, HSQC, HMBC, and NOESY (Figures 3 and S5–S28). The typical signals of a

Table 1. NMR Data of Compounds 6–9 in DMSO- $d_6$ <sup>a</sup>

position	6		7		8		9	
	$\delta_C$ , type	$\delta_H$ , multi. (J in Hz)	$\delta_C$ , type	$\delta_H$ , multi. (J in Hz)	$\delta_C$ , type	$\delta_H$ , multi. (J in Hz)	$\delta_C$ , type	$\delta_H$ , multi. (J in Hz)
1		6.53, d (2.5)		6.53, d (3.0)		6.50, d (3.0)		6.49, d (3.0)
2	79.9, CH	5.24, d (2.5)	79.8, CH	5.24, d (3.0)	79.8, CH	5.26, d (3.0)	79.8, CH	5.26, d (3.0)
3	55.0, C		55.0, C		55.1, C		55.0, C	
4	122.4, CH	7.04, d (7.3)	122.4, CH	7.05, d (7.3)	122.4, CH	7.07, d (7.4)	122.3, CH	7.07, d (7.3)
5	117.4, CH	6.59, t (7.4)	117.4, CH	6.58, td (7.3, 1.2)	117.5, CH	6.59, dd (7.4, 0.9)	117.5, CH	6.59, td (7.3, 1.0)
6	127.7, CH	6.96, t (7.6)	127.7, CH	6.96, td (7.7, 1.2)	127.7, CH	6.95, dd (7.7, 0.9)	127.7, CH	6.95, td (7.8, 1.0)
7	108.7, CH	6.53, t (7.8)	108.7, CH	6.51, d (7.7)	108.7, CH	6.49, d (7.7)	108.7, CH	6.50, d (7.8)
8	148.5, C		148.7, C		148.5, C		148.5, C	
9	132.6, C		132.6, C		132.6, C		132.6, C	
10 $\alpha$	39.1, CH <sub>2</sub>	2.32, dd (12.8, 7.6)	39.8, CH <sub>2</sub>	2.31, dd (12.9, 8.0)	38.9, CH <sub>2</sub>	2.34, dd (13.0, 7.8)	39.3, CH <sub>2</sub>	2.35, dd (13.0, 7.6)
10 $\beta$		1.90, dd (12.8, 10.3)		1.91, dd (12.9, 10.1)		2.05, dd (13.0, 9.8)		2.01, dd (13.0, 10.2)
11	57.2, CH	4.45, t (8.7)	57.2, CH	4.44, t (8.8)	57.4, CH	4.49, t (8.8)	57.3, CH	4.48, br t (8.7)
13	167.3, C		167.5, C		168.4, C		167.5, C	
14	55.8, CH	4.37, t (5.2)	56.0, CH	4.27, t (5.3)	52.9, CH	4.04, dd (7.8, 5.1)	53.6, CH	4.18, br t (5.1)
15		7.87, s		7.71, s		7.96, s		8.14, s
16	169.6, C		169.6, C		170.3, C		170.2, C	
17 $\alpha$	34.5, CH <sub>2</sub>	3.16, dd (14.5, 5.1)	33.6, CH <sub>2</sub>	3.03, dd (14.5, 5.2)	37.6, CH <sub>2</sub>	1.79, ddd (13.6, 8.5, 4.9)	28.7, CH <sub>2</sub>	2.10, m
17 $\beta$		3.01, dd (14.5, 5.4)		2.89, dd (14.5, 5.6)		1.39, ddd (13.6, 7.8, 5.9)		1.89, m
18	137.3, C		127.3, C		24.0, CH	1.90, m	28.9, CH <sub>2</sub>	2.58, t (8.5)
19	129.5, CH	7.31, d (7.5)	130.5, CH	7.10, d (8.5)	22.8, CH <sub>3</sub>	0.87, d (6.3)		
20	127.9, CH	7.20, t (7.4)	114.8, CH	6.59, dd (8.5, 1.9)	21.8, CH <sub>3</sub>	0.88, d (6.3)	14.4, CH <sub>3</sub>	2.05, s
21	126.2, CH	7.14, t (6.9)	155.7, C					
OH-21				9.11, s				
22	127.9, CH	7.20, t (7.4)	114.8, CH	6.59, dd (8.5, 1.9)				
23	129.5, CH	7.31, d (7.5)	130.5, CH	7.10, d (8.5)				
1'	34.4, CH <sub>2</sub>	2.44, d (6.9)	34.3, CH <sub>2</sub>	2.44, d (7.2)	34.2, CH <sub>2</sub>	2.46, d (7.2)	34.2, CH <sub>2</sub>	2.46, d (7.2)
2'	119.6, CH	5.02, t (6.9)	119.6, CH	5.01, br t (7.2)	119.7, CH	5.03, t (7.2)	119.7, CH	5.02, t (7.2)
3'	133.9, C		133.9, C		133.9, C		133.9, C	
4'	17.9, CH <sub>3</sub>	1.57, s	17.9, CH <sub>3</sub>	1.57, s	17.9, CH <sub>3</sub>	1.58, s	17.9, CH <sub>3</sub>	1.58, s
5'	25.6, CH <sub>3</sub>	1.63, s	25.7, CH <sub>3</sub>	1.63, s	25.6, CH <sub>3</sub>	1.64, s	25.6, CH <sub>3</sub>	1.64, s

<sup>a</sup>See Supporting Information for structure numbering.



Table 2. NMR Data of Compounds 10 and 11 in DMSO-*d*<sub>6</sub>

position	10		11	
	$\delta_C$ , type	$\delta_H$ , multi. (J in Hz)	$\delta_C$ , type	$\delta_H$ , multi. (J in Hz)
1		6.64, d (2.8)		6.60, d (2.8)
2	79.9, CH	5.32, d (2.8)	79.8, CH	5.29, d (2.8)
3	54.7, C		54.7, C	
4	122.5, CH	7.11, d (7.1)	122.5, CH	7.10, d (7.5)
5	117.6, CH	6.62, t (7.7)	117.6, CH	6.61, td (7.5, 1.3)
6	127.8, CH	6.97, br t (7.6)	127.8, CH	6.96, td (7.9, 1.3)
7	108.8, CH	6.52, d (7.8)	108.7, CH	6.51, d (7.9)
8	148.5, C		148.5, C	
9	132.7, C		132.8, C	
10 $\alpha$	39.8, CH <sub>2</sub>	2.46 <sup>a</sup>	39.5, CH <sub>2</sub>	2.44, dd (12.8, 7.4)
10 $\beta$		2.04, dd (12.8, 10.5)		2.04, dd (12.8, 10.5)
11	56.6, CH	4.72, dd (10.5, 7.4)	56.5, CH	4.65, dd (10.5, 7.4)
13	160.4, C		161.0, C	
14	128.3, C		125.7, C	
15		9.97, br s		9.86, br s
16	167.3, C		167.2, C	
17	115.9, CH	6.77, s	116.8, CH	6.70, s
18	133.3, C		124.1, C	
19/23	129.4, CH	7.56, d (7.6)	131.2, CH	7.42, d (8.6)
20/22	128.6, CH	7.41, t (7.4)	115.5, CH	6.80, d (8.6)
21	128.2, CH	7.32, t (7.4)	157.8, C	
OH-21				9.78, br s
1'	34.6, CH <sub>2</sub>	2.46 <sup>a</sup>	34.6, CH <sub>2</sub>	2.48, d (7.2)
2'	119.6, CH	5.02, t (7.3)	119.6, CH	5.02, t (7.2)
3'	134.1, C		134.0, C	
4'	18.0, CH <sub>3</sub>	1.59, s	18.0, CH <sub>3</sub>	1.58, s
5'	25.7, CH <sub>3</sub>	1.64, s	25.7, CH <sub>3</sub>	1.64, s

<sup>a</sup>Signals overlapping with each other. See Supporting Information for structure numbering.

regular C-3-prenyl residue in the <sup>1</sup>H NMR spectra are found in the ranges of  $\delta_H$  2.44–2.46 (d, 6.9–7.2 Hz, H-1'), 5.01–5.03 (t, 6.9–7.2 Hz, H-2'), 1.57–1.58 (s, H-4'), and 1.63–1.64 (s, H-5') (Table 1). The signals of the five carbons are detected in the <sup>13</sup>C spectra at about  $\delta_C$  34 (C-1'), 120 (C-2'), 134 (C-3'), 18 (C-4'), and 26 (C-5') (Table 1). Prenylation at C-3 destroys the aromatic character of the indole system and causes a shielded shift of the H-2 signal to  $\delta_H$  5.24–5.26 as well as those of C-2 and C-3 to  $\delta_C$  80 and 55, respectively. The configuration of the products was determined based on the correlations between H-1' and H-11, H-1', and H-2 as well as H-2 and H-11 in the NOESY spectra. Comparison of their ECD spectra provided additional evidence for their configurations (Figure S60). All the obtained data confirmed that 6–9 are C-3-prenylated derivatives of the corresponding CDPs (Figure 2B). More fascinatingly, SasB performed also very effective conversions of these four substrates, at least in our experiments.

Low conversions to prenylated derivatives were also detected by LC-MS analysis for *cyclo*-(L-Trp-L-Ala), *cyclo*-(L-Trp-D-Ala), *cyclo*-(D-Trp-L-Ala), *cyclo*-(D-Trp-D-Ala), *cyclo*-(L-Trp-L-Pro), *cyclo*-(L-Trp-D-Pro), *cyclo*-(D-Trp-L-Pro), and *cyclo*-(D-Trp-D-Pro) (Figures S63 and S64). Due to the low product yields, the structures of these products could not be elucidated in this study. These results suggest a more flexible substrate specificity of SasB from *S. aurantiacus* than that of SazB from *S. leeuwenhoekii*.<sup>15</sup> It was reported that cWF, cWY, cWA, and cWP were not accepted by SazB.<sup>15</sup> In our case, all of these four CDPs were prenylated by SasB with high conversion for cWF and cWY (Figures 2, S63, and S64).

Cyclodipeptide oxidases (CDOs) are frequently found in the CDPS-related pathways and install *exo* double bonds at the DKP ring.<sup>26</sup> For combinatorial application of SasB with these oxidases, we tested its acceptance of the dehydrogenated forms of the four efficiently converted CDPs, i.e., cWΔF, cWΔY, cWΔL, and cWΔM, by incubation experiments in the *sasB* transformant. LC-MS analysis showed that all of these compounds were good substrates for SasB and were completely converted to their prenylated products (Figure 2A–ix–xii). The products 10–13 were subsequently isolated, and their structures confirmed to be regularly C-3-prenylated derivatives at the indoline ring (Figure 2B) by detailed interpretation of their NMR data and the comparison with the data of 6–9 (Tables 2 and 3, and Figures 3, S29–S52, and S61). Observation of the interaction between NH-15 and H-19/H-23 in the NOESY spectrum of 11 as well as NH-15 and H-18 in that of 12 supported the *Z*-configuration of the *exo* double bonds in their structures (Figures S40 and S46).

During the isolation procedure, we observed the conversion of the cWΔM product 13 to the new compound 14. Isolation by using a chiral-phase HPLC column (Figure S65) and structure elucidation by NMR analysis including interpretation of the NOESY data and comparison of its ECD spectrum with that of 13 (Table 3 and Figures 3, S53–S58, and S62) confirmed the epimerization at the C-11 position. As the nonenzymatic epimerization via keto–enol tautomerism was already observed for the guanitryptmycins,<sup>38</sup> we speculated a similar mechanism may explain the conversion of 13 to 14. Incubation of 13 in CD<sub>3</sub>OD/D<sub>2</sub>O (1:1) at pH 9 and 12 for 14 h and LC-MS analysis confirmed indeed the conversion of 13



Table 3. NMR Data of Compounds 12–14 in DMSO-*d*<sub>6</sub>

position	12		13		14	
	$\delta_C$ , type	$\delta_H$ , multi. ( <i>J</i> in Hz)	$\delta_C$	$\delta_H$ , multi. ( <i>J</i> in Hz)	$\delta_C$ , type	$\delta_H$ , multi. ( <i>J</i> in Hz)
1		6.55, d (2.8)		6.60, d (2.9)		6.51, d (1.0)
2	79.8, CH	5.23, d (2.8)	79.9, CH	5.26, d (2.9)	79.0, CH	5.33, br s
3	54.4, C		54.4, C		54.4, C	
4	122.5, CH	7.08, d (7.2)	122.5, CH	7.08, dd (7.4, 1.1)	123.1, CH	7.12, d (7.2)
5	117.5, CH	6.60, td (7.2, 1.0)	117.5, CH	6.60, td (7.4, 1.1)	117.6, CH	6.61, td (7.2, 1.0)
6	127.8, CH	6.95, td (7.7, 1.0)	127.8, CH	6.96, td (7.8, 1.3)	128.1, CH	6.97, td (7.4, 1.0)
7	108.6, CH	6.49, d (7.7)	108.6, CH	6.50, d (7.8)	108.2, CH	6.51, d (7.4)
8	148.5, C		148.5, C		150.2, C	
9	132.8, C		132.7, C		130.9, C	
10 $\alpha$	40.1, CH <sub>2</sub>	2.42, dd (12.8, 7.1)	40.1, CH <sub>2</sub>	2.44, dd (12.7, 7.1)	39.8, CH <sub>2</sub>	2.20, t (11.9)
10 $\beta$		1.94, dd (12.8, 10.8)		1.95, dd (12.7, 11.0)		2.46, dd (11.9, 5.9)
11	56.3, CH	4.57, dd (10.8, 7.1)	56.4, CH	4.62, dd (11.0, 7.1)	57.8, CH	4.02, dd (11.9, 5.9)
13	159.7, C		159.1, C		157.0, C	
14	127.0, C		130.1, C		129.7, C	
15		9.92, s		10.12, s		10.10, s
16	166.9, C		166.9, C		165.9, C	
17	125.5, CH	5.68, d (10.5)	113.6, CH	5.84, t (8.7)	112.8, CH	5.77, t (8.7)
18	24.4, CH	2.82, m	27.3, CH <sub>2</sub>	3.41, dd (14.0, 8.7) 3.29 <sup>a</sup>	27.2, CH <sub>2</sub>	3.39, dd (14.0, 8.7)
19	22.3, CH <sub>3</sub>	0.96, d (6.5)				
20	22.1, CH <sub>3</sub>	0.96, d (6.5)	14.0, CH <sub>3</sub>	2.01, s	14.0, CH <sub>3</sub>	1.99, s
1'	34.6, CH <sub>2</sub>	2.46, d (7.3)	34.6, CH <sub>2</sub>	2.47, d (7.2)	36.0, CH <sub>2</sub>	2.40, d (7.3)
2'	119.6, CH	4.99, t (7.3)	119.6, CH	4.99, t (7.2)	119.3, CH	5.08, t (7.3)
3'	134.0, C		134.0, C		133.8, C	
4'	18.0, CH <sub>3</sub>	1.58, s	18.0, CH <sub>3</sub>	1.58, s	18.0, CH <sub>3</sub>	1.51, s
5'	25.7, CH <sub>3</sub>	1.63, s	25.7, CH <sub>3</sub>	1.64, s	25.7, CH <sub>3</sub>	1.64, s

<sup>a</sup>Signals overlapping with that of water. See Supporting Information for structure numbering.

to **14** and incorporation of one deuterium in both molecules (Figure S66). This supported the epimerization at C-11 via keto–enol tautomerism.

Taking the above results together, a putative *cdps*-containing gene cluster with a prenyltransferase and a methyltransferase gene was identified from *S. aurantiacus* and successfully expressed in *S. albus* J1074. The diprenylated and dimethylated diketopiperazine indole alkaloid streptoazine C (**2**) was identified as the cluster end product. Heterologous expression of different gene combinations and precursor incubation experiments enabled us to evaluate the order of the biosynthetic steps to **2**. The cWW formation catalyzed by SasA is followed by two regular prenylation steps with SasB and final *N*-methylations with SasC. It is noteworthy that the same order of reaction steps occurs in the biosynthesis of the same product in *S. leuvenhoekeii*.<sup>15</sup> However, differing from *sazB* from *S. leuvenhoekeii* coding for both PT and MT activities, the two independent genes *sasB* and *sasC* are located in the streptoazine cluster in *S. aurantiacus* described in this study. More importantly, SasB displays a remarkably high substrate tolerance and can accept not only a number of tryptophan-containing CDPs but also their dehydrogenated derivatives for prenyl decoration. The successful production of dehydrogenated and prenylated CDPs by combination of cyclodipeptide oxidases and the prenyltransferases SasB will provide an excellent example for accessing diversified natural products in the combinatorial biology field, which also inspires us to explore the combination of SasB with other modification genes in order to access more diverse DKPs. This is the first report on the substrate flexibility of a CDPS-related PT toward non-natural substrates. Identification of additional PTs from

this less explored enzyme group will provide more details on their biochemical properties.

## EXPERIMENTAL SECTION

**General Experimental Procedures.** The optical rotation was measured with an A KRÜSS P3000 polarimeter at 20 °C using the D-line of the sodium lamp at  $\lambda = 589.3$  nm. Prior to the measurement, the polarimeter was calibrated with chloroform as solvent. Circular dichroism spectra were taken on a J-1500 CD spectrometer (Jasco Deutschland GmbH). The samples were dissolved in MeOH and measured in the range of 200–400 nm by using a 1 mm path length quartz cuvette. The NMR spectra of the purified compounds were recorded on a JEOL ECA-500 MHz spectrometer in DMSO-*d*<sub>6</sub>, and all spectra were processed with MestReNova 9.0.0 (Metrelab). Chemical shifts are referred to those of DMSO-*d*<sub>6</sub> ( $\delta_H$  2.50 and  $\delta_C$  39.5). High-resolution mass spectrometric analysis was performed on an Agilent 1260 HPLC system equipped with a microTOF-Q III spectrometer (Bruker) using a Multospher 120 RP-18 column (250 × 2 mm, 5  $\mu$ m) (CS-Chromatographie Service GmbH) or a Multohigh Chiral AM-RP HPLC column (150 × 4.6 mm, CS-Chromatographie Service GmbH). Electrospray positive ionization mode was selected for determination of the exact masses. The capillary voltage was set to 4.5 kV and the collision energy to 8.0 eV. Sodium formate was used in each run for mass calibration. The masses were scanned in the range of *m/z* 100–1500. Data were evaluated with the Compass DataAnalysis 4.2 software (Bruker Daltonik). Semipreparative HPLC was performed on the same HPLC equipment with an Agilent ZORBAX Eclipse XDB C18 HPLC column (250 × 9.4 mm, 5  $\mu$ m) and a Multohigh Chiral AM-RP HPLC column (250 × 10 mm, CS-Chromatographie Service GmbH). Sephadex LH-20 (Merck) was used for column chromatography, and HPLC for detection of the desired substance in the fractions.

**Chemicals.** *cyclo*-(L-Trp-L-Trp), *cyclo*-(L-Trp-L-Phe), *cyclo*-(L-Trp-L-Tyr), and *cyclo*-(L-Trp-L-Leu) were purchased from Bachem. *cyclo*-(L-Trp-L-Trp), *cyclo*-(L-Trp-L-Phe), *cyclo*-(L-Trp-L-Tyr), and *cyclo*-(L-Trp-L-Leu) were purchased from Bachem.

(L-Trp-L-Met) was isolated from the strain described previously.<sup>22</sup> The stereoisomers of *cyclo*-Trp-Ala and *cyclo*-Trp-Pro were synthesized according to the method published previously.<sup>39</sup> The dehydrogenated CDP substrates used in this study were prepared by using cyclodipeptide oxidases according to the method described in a previous work.<sup>40</sup>

#### Bacterial Strains, Plasmids, and Growth Conditions.

*Streptomyces aurantiacus* NRRL ISP-5412 was obtained from the ARS Culture Collection (NRRL). *Streptomyces albus* J1074<sup>35</sup> was kindly gifted by Prof. Dr. Andriy Luzhetskyy (Saarland University). *S. albus* J1074 and the generated exconjugants were maintained on MS plates (mannitol 20.0 g/L, soya flour 20.0 g/L, agar 20.0 g/L) at 28 °C for sporulation. For secondary metabolite production, *S. albus* J1074 transformants were cultivated in liquid modified R5 medium (sucrose 103.0 g/L, glucose 10.0 g/L, yeast extract 5.0 g/L, MgCl<sub>2</sub>·6H<sub>2</sub>O 10.12 g/L, K<sub>2</sub>SO<sub>4</sub> 0.25 g/L, Difco casamino acids 0.1 g/L, MOPS 21.0 g/L, trace element solution 2 mL/L, pH 7.2) at 28 °C for 7 days.

**Computer-Assisted Sequence Analysis.** Nucleotide and amino acid sequences used in this study were obtained from NCBI databases (<http://www.ncbi.nlm.nih.gov>). Comparison of protein sequences was carried out by using the BLASTP program (<http://blast.ncbi.nlm.nih.gov/>). The phylogenetic tree of PTs (Figure S1) was created by MEGA version 7.0 (<http://www.megasoftware.net>).

**Genetic Manipulation, PCR Amplification, and Gene Cloning.** Strains and plasmids used in this study are listed in Table S2 and Table S3, respectively. Recombinant *E. coli* strains were cultivated in liquid or on solid lysogeny-broth (LB) with 100 µg/mL ampicillin, 50 µg/mL kanamycin, 50 µg/mL apramycin, or 25 µg/mL chloramphenicol when necessary.

Genetic manipulation in *E. coli* was performed according to the protocol by Green and Sambrook.<sup>41</sup> Genomic DNA isolation from *Streptomyces* was performed as described in the literature.<sup>42</sup>

Gene regions were amplified by PCR from genomic DNA of *S. aurantiacus* by using primers listed in Table S3 and Phusion high-fidelity DNA polymerase from New England Biolabs. The generated PCR fragments were cloned into pGEM-T Easy vector (Promega), and the sequence integrity was confirmed by sequencing. Subsequently, the fragments were released with restriction endonucleases from pGEM-T Easy and ligated into pPWV50A,<sup>34</sup> which was digested with the same enzymes, previously. The generated constructs (Table S3) were used for heterologous expression in *S. albus* J1074.

**Heterologous Gene Expression in *Streptomyces albus* J1074 and Cultivation for Secondary Metabolite Production.** The constructed plasmids harboring different genes or the gene cluster were transformed into nonmethylated *E. coli* ET12567/pUZ8002, then conjugated with *S. albus* J1074. The positive conjugants were selected by the phenotype showing apramycin resistance and further confirmed by PCR. The spores of the *S. albus* J1074 transformants were inoculated into 50 mL of modified R5 liquid media supplied with 50 µg/mL of apramycin in 250 mL baffled flasks and cultured at 28 °C and 200 rpm for 7 days. This culture (1 mL) was extracted with the same volume of EtOAc three times. After that, the organic phases were combined and evaporated, and the dried residues were dissolved in 200 µL of MeOH. Samples (5 µL) were subjected to LC-MS for analysis.

**Biotransformation for the Generation of Various Prenylated DKPs.** The *S. albus* J1074 transformant harboring *sasB* was incubated in modified R5 medium at 28 °C, 200 rpm for 2 days. Tryptophan-containing CDPs or dehydrogenated derivatives were separately added to 10 mL of these precultures to final concentrations of 100 µM. After cultivation at 28 °C for an additional 5 days, the metabolites were extracted with EtOAc and analyzed by LC-MS.

**LC-MS Analysis.** For secondary metabolite analysis, a linear gradient of 5–100% MeCN in H<sub>2</sub>O, both containing 0.1% HCOOH, in 40 min and a flow rate at 0.25 mL/min were used. The column was then washed with 100% MeCN containing 0.1% HCOOH for 5 min and equilibrated with 5% MeCN in H<sub>2</sub>O for 5 min. For analysis of the samples after incubation of the prenylated dehydrogenated DKPs in CD<sub>3</sub>OD/D<sub>2</sub>O, a Multohigh Chiral AM-RP HPLC column (150 × 4.6

mm, CS-Chromatographie Service GmbH) was used. Separation was carried out with a linear gradient of 50–100% MeCN in H<sub>2</sub>O in 30 min and a flow rate of 0.5 mL/min.

**Extraction and Isolation of Secondary Metabolites.** For structure elucidation of the accumulated products, *S. albus* J1074 transformants harboring *sasABC* and *sasAB* were fermented in modified R5 medium on large scales (8 L) at 28 °C for 7 days. The culture supernatants were collected and extracted with an equal volume of EtOAc three times. The EtOAc phases were evaporated to dryness, dissolved in MeOH, and applied to chromatography on a Sephadex LH-20 column with MeOH as eluent. The fractions containing the target products were further purified on an Agilent HPLC 1260 series by using a semipreparative Agilent ZORBAX Eclipse XDB C18 HPLC column (250 × 9.4 mm, 5 µm). The flow rate was set to 2.0 mL/min. Compounds 2, 3, and 4 were purified with 95%, 70%, and 85% MeCN in H<sub>2</sub>O, respectively.

For the prenylated DKP derivatives generated by biotransformation, the extracts were obtained by extraction with EtOAc as described above. Compounds 6–12 were further purified on an Agilent HPLC 1260 series with 60–65% MeCN in H<sub>2</sub>O. Compounds 13 and 14 were separated on an Agilent HPLC 1260 series by using a semipreparative Multohigh Chiral AM-RP HPLC column (250 × 10 mm, CS-Chromatographie Service GmbH) with 80% MeCN in H<sub>2</sub>O. The flow rate was set to 2.0 mL/min.

**Streptoazine C (2):** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +180 (c 0.33, CHCl<sub>3</sub>); ECD (0.29 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 292 (+3.71), 277 (+2.64), 255 (+9.86), 229 (−3.08), 211 (+10.51) nm; <sup>1</sup>H NMR data, Table S4; HRESIMS *m/z*: 537.3245 [M + H]<sup>+</sup> (calcd for C<sub>34</sub>H<sub>41</sub>N<sub>4</sub>O<sub>2</sub>, 537.3224).

**Compound 3:** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +45 (c 0.40, CHCl<sub>3</sub>); ECD (0.27 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 290 (+3.09), 265 (+1.59), 244 (+10.82), 222 (−4.78), 214 (−4.60) nm; <sup>1</sup>H NMR data, Table S4; HRESIMS *m/z* 441.2288 [M + H]<sup>+</sup> (calcd for C<sub>27</sub>H<sub>29</sub>N<sub>4</sub>O<sub>2</sub>, 441.2285).

**Streptoazine A (4):** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +340 (c 0.26, CHCl<sub>3</sub>); ECD (0.30 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 291 (+2.69), 265 (+1.08), 245 (+11.14), 219 (−1.46), 208 (+25.47) nm; <sup>1</sup>H NMR data, Table S4; HRESIMS *m/z* 509.2911 [M + H]<sup>+</sup> (calcd for C<sub>32</sub>H<sub>37</sub>N<sub>4</sub>O<sub>2</sub>, 509.2911).

**Compound 6:** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +40.1 (c 1.87, CHCl<sub>3</sub>); ECD (0.70 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 294 (+3.70), 270 (+0.43), 243 (+9.02), 218 (−1.24) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, Table 1; HRESIMS *m/z* 402.2182 [M + H]<sup>+</sup> (calcd for C<sub>25</sub>H<sub>28</sub>N<sub>3</sub>O<sub>2</sub>, 402.2176).

**Compound 7:** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +25 (c 0.71, CHCl<sub>3</sub>); ECD (0.51 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 287 (+2.23), 267 (+0.55), 243 (+12.14), 225 (−1.80), 213 (+1.14) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, Table 1; HRESIMS *m/z* 418.2127 [M + H]<sup>+</sup> (calcd for C<sub>25</sub>H<sub>28</sub>N<sub>3</sub>O<sub>3</sub>, 418.2125).

**Compound 8:** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +113 (c 1.34, CHCl<sub>3</sub>); ECD (0.55 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 293 (+3.10), 264 (+0.83), 242 (+10.11), 223 (−0.97), 210 (+8.14) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, Table 1; HRESIMS *m/z* 368.2330 [M + H]<sup>+</sup> (calcd for C<sub>22</sub>H<sub>30</sub>N<sub>3</sub>O<sub>2</sub>, 368.2333).

**Compound 9:** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +14.2 (c 2.19, CHCl<sub>3</sub>); ECD (0.68 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 291 (+2.28), 268 (+0.57), 243 (+8.54), 224 (+1.23), 210 (+7.29) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, Table 1; HRESIMS *m/z* 368.1906 [M + H]<sup>+</sup> (calcd for C<sub>21</sub>H<sub>28</sub>N<sub>3</sub>O<sub>2</sub>S, 368.1897).

**Compound 10:** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +150 (c 0.89, CHCl<sub>3</sub>); ECD (0.34 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 293 (+8.48), 254 (+1.11), 229 (+12.71), 215 (+12.83) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, Table 2; HRESIMS *m/z* 400.2020 [M + H]<sup>+</sup> (calcd for C<sub>25</sub>H<sub>26</sub>N<sub>3</sub>O<sub>2</sub>, 400.2020).

**Compound 11:** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +25 (c 0.85, CHCl<sub>3</sub>); ECD (0.31 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 328 (+7.90), 315 (+7.98), 301 (+10.74), 258 (−2.19), 221 (+20.25) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, Table 2; HRESIMS *m/z* 416.1968 [M + H]<sup>+</sup> (calcd for C<sub>25</sub>H<sub>26</sub>N<sub>3</sub>O<sub>3</sub>, 416.1969).

**Compound 12:** yellow oil; [ $\alpha$ ]<sub>D</sub><sup>20</sup> +24 (c 0.67, CHCl<sub>3</sub>); ECD (0.28 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 306 (+2.94), 276 (+0.26), 248 (+13.56), 237

(+9.72), 219 (+26.54) nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR, Table 3; HRESIMS  $m/z$  366.2178  $[\text{M} + \text{H}]^+$  (calcd for  $\text{C}_{22}\text{H}_{28}\text{N}_3\text{O}_2$ , 366.2176).

**Compound 13:** yellow oil;  $[\alpha]_{\text{D}}^{20} +227$  ( $c$  1.29,  $\text{CHCl}_3$ ); ECD (0.40 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 306 (+2.48), 287 (+1.71), 247 (+10.40), 246 (+10.25), 220 (+29.36) nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, Table 3; HRESIMS  $m/z$  384.1746  $[\text{M} + \text{H}]^+$  (calcd for  $\text{C}_{21}\text{H}_{26}\text{N}_3\text{O}_2\text{S}$ , 384.1740).

**Compound 14:** yellow oil;  $[\alpha]_{\text{D}}^{20} +170$  ( $c$  0.61,  $\text{CHCl}_3$ ); ECD (0.48 mM, MeOH)  $\lambda_{\text{max}}$  ( $\Delta\epsilon$ ) 302 (+3.88), 287 (+3.11), 249 (+12.83), 220 (−2.94), 207 (−13.33) nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, Table 3; HRESIMS  $m/z$  384.1737  $[\text{M} + \text{H}]^+$  (calcd for  $\text{C}_{21}\text{H}_{26}\text{N}_3\text{O}_2\text{S}$ , 384.1740).

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jnatprod.1c00844>.

Strains, plasmids, primers, cluster information, NMR data of compounds 2–4, NMR spectra of compounds 2–14, experimental ECD spectra, HPLC, and LC-MS chromatograms (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

Shu-Ming Li – Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, 35037 Marburg, Germany; [orcid.org/0000-0003-4583-2655](https://orcid.org/0000-0003-4583-2655); Email: [shuming.li@staff.uni-marburg.de](mailto:shuming.li@staff.uni-marburg.de)

### Authors

Jing Liu – Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, 35037 Marburg, Germany

Yiling Yang – Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, 35037 Marburg, Germany

Lauritz Harken – Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, 35037 Marburg, Germany

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jnatprod.1c00844>

### Author Contributions

§J.L. and Y.Y. contributed equally to this work.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank the ARS Culture Collection (NRRL) for providing *Streptomyces aurantiacus* strain and R. Kraut, L. Ludwig-Radtke, and S. Newel (Fachbereich Pharmazie, University of Marburg) for LC-MS and NMR analyses. This project was funded in part by the Deutsche Forschungsgemeinschaft (DFG, Li844/14-1 and INST 160/620-1). J.L. (201608310118) and Y.Y. (201808530447) are the scholarship recipients from the China Scholarship Council.

## ■ REFERENCES

- (1) Li, S.-M. *Nat. Prod. Rep.* **2010**, *27*, 57–78.
- (2) Wang, X.; Li, Y.; Zhang, X.; Lai, D.; Zhou, L. *Molecules* **2017**, *22*, 2026.
- (3) Borthwick, A. D. *Chem. Rev.* **2012**, *112*, 3641–3716.

- (4) Mishra, A. K.; Choi, J.; Choi, S. J.; Baek, K. H. *Molecules* **2017**, *22*, E1796.

- (5) Hayashi, H.; Takiuchi, K.; Murao, S.; Arai, M. *Agric. Biol. Chem.* **1988**, *52*, 2131–2133.

- (6) Murao, S.; Hayashi, H.; Akiuchi, K.; Arai, M. *Agric. Biol. Chem.* **1988**, *52*, 885–886.

- (7) Kozlovsky, A. G.; Vinokurova, N. G.; Adanin, V. M.; Burkhardt, G.; Dahse, H. M.; Gräfe, U. *J. Nat. Prod.* **2000**, *63*, 698–700.

- (8) Kozlovsky, A. G.; Vinokurova, N. G.; Adanin, V. M.; Burkhardt, G.; Dahse, H.-M.; Gräfe, U. *J. Nat. Prod.* **2001**, *64*, 553–554.

- (9) Arai, K.; Kimura, K.; Mushiroda, T.; Yamamoto, Y. *Chem. Pharm. Bull.* **1989**, *37*, 2937–2939.

- (10) Yamazaki, M.; Sasago, K.; Miyaki, K. *J. Chem. Soc., Chem. Commun.* **1974**, 408–409.

- (11) Clark, B.; Capon, R. J.; Lacey, E.; Tennant, S.; Gill, J. H. *J. Nat. Prod.* **2005**, *68*, 1661–1664.

- (12) Cardani, C.; Casnati, G.; Piozzi, F.; Quilico, A. *Tetrahedron Lett.* **1959**, *1*, 1–8.

- (13) Raju, R.; Piggott, A. M.; Huang, X. C.; Capon, R. J. *Org. Lett.* **2011**, *13*, 2770–2773.

- (14) Yao, T.; Liu, J.; Liu, Z.; Li, T.; Li, H.; Che, Q.; Zhu, T.; Li, D.; Gu, Q.; Li, W. *Nat. Commun.* **2018**, *9*, 4091.

- (15) Zhang, Y.; Yao, T.; Jiang, Y.; Li, H.; Yuan, W.; Li, W. *Appl. Environ. Microbiol.* **2021**, *87*, e02525–20.

- (16) Xu, W.; Gavia, D. J.; Tang, Y. *Nat. Prod. Rep.* **2014**, *31*, 1474–1487.

- (17) Winkelblech, J.; Fan, A.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 7379–7397.

- (18) Li, S.-M. *J. Antibiot.* **2011**, *64*, 45–49.

- (19) Mori, T. *J. Nat. Med.* **2020**, *74*, 501–502.

- (20) Fraley, A. E.; Sherman, D. H. *FEBS J.* **2020**, *287*, 1381–1402.

- (21) Nies, J.; Li, S.-M. *ACS Chem. Biol.* **2021**, *16*, 185–192.

- (22) Liu, J.; Yu, H.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 4435–4444.

- (23) Jacques, I. B.; Moutiez, M.; Witwinowski, J.; Darbon, E.; Martel, C.; Seguin, J.; Favry, E.; Thai, R.; Lecoq, A.; Dubois, S.; et al. *Nat. Chem. Biol.* **2015**, *11*, 721–727.

- (24) Süßmuth, R. D.; Mainz, A. *Angew. Chem., Int. Ed.* **2017**, *56*, 3770–3821.

- (25) Canu, N.; Moutiez, M.; Belin, P.; Gondry, M. *Nat. Prod. Rep.* **2020**, *37*, 312–321.

- (26) Harken, L.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2021**, *105*, 2277–2285.

- (27) Wohlgemuth, V.; Kindinger, F.; Xie, X.; Wang, B. G.; Li, S.-M. *Org. Lett.* **2017**, *19*, 5928–5931.

- (28) Fan, A.; Winkelblech, J.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 7399–7415.

- (29) Kumano, T.; Richard, S. B.; Noel, J. P.; Nishiyama, M.; Kuzuyama, T. *Bioorg. Med. Chem.* **2008**, *16*, 8117–8126.

- (30) Kuzuyama, T.; Noel, J. P.; Richard, S. B. *Nature* **2005**, *435*, 983–987.

- (31) Medema, M. H.; de Rond, T.; Moore, B. S. *Nat. Rev. Genet.* **2021**, *22*, 553.

- (32) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R. *Annu. Rev. Biochem.* **2021**, *90*, 763–788.

- (33) Skinnider, M. A.; Johnston, C. W.; Gunabalasingam, M.; Merwin, N. J.; Kieliszek, A. M.; MacLellan, R. J.; Li, H.; Ranieri, M. R. M.; Webster, A. L. H.; Cao, M. P. T.; et al. *Nat. Commun.* **2020**, *11*, 6058.

- (34) Zhu, Y.; Fu, P.; Lin, Q.; Zhang, G.; Zhang, H.; Li, S.; Ju, J.; Zhu, W.; Zhang, C. *Org. Lett.* **2012**, *14*, 2666–2669.

- (35) Zaburannyi, N.; Rabyk, M.; Ostash, B.; Fedorenko, V.; Luzhetskyy, A. *BMC Genomics* **2014**, *15*, 97.

- (36) Wollinsky, B.; Ludwig, L.; Xie, X.; Li, S.-M. *Org. Biomol. Chem.* **2012**, *10*, 9262–9270.

- (37) Dubois, P.; Correia, I.; Le Chevalier, F.; Dubois, S.; Jacques, I.; Canu, N.; Moutiez, M.; Thai, R.; Gondry, M.; Lequin, O.; et al. *Sci. Rep.* **2019**, *9*, 9208.

- (38) Liu, J.; Xie, X.; Li, S.-M. *Angew. Chem., Int. Ed.* **2019**, *58*, 11534–11540.
- (39) Yin, S.; Yu, X.; Wang, Q.; Liu, X. Q.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 1649–1660.
- (40) Mikulski, L.; Schäfer, J.; Brockmeyer, K.; Kraut, R.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 2523–2536.
- (41) Green, M. R.; Sambrook, J. *Molecular Cloning: A Laboratory Manual*, 4th ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2012.
- (42) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces Genetics*, 2nd ed.; John Innes Foundation: Norwich, UK, 2000.

## Supporting Information

### **Elucidation of the Streptoazine Biosynthetic Pathway in *Streptomyces aurantiacus* Reveals the Presence of a Promiscuous Prenyltransferase/Cyclase**

Jing Liu<sup>† §</sup>, Yiling Yang<sup>† §</sup>, Lauritz Harken<sup>†</sup>, and Shu-Ming Li<sup>†\*</sup>

<sup>†</sup> Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany



## Table of Contents

<b>Supplementary Tables .....</b>	<b>183</b>
<b>Table S1.</b> Comparison of sas genes with known entries.....	183
<b>Table S2.</b> Bacterial strains used in this study.....	183
<b>Table S3.</b> Cloning and expression constructs used in this study.....	183
<b>Table S4.</b> $^1\text{H}$ NMR data of compounds <b>2</b> — <b>4</b> in $\text{DMSO}-d_6$ .....	184
<b>Supplementary Figures.....</b>	<b>185</b>
<b>Figure S1.</b> Phylogenetic analysis of PTs .....	185
<b>Figure S2.</b> $^1\text{H}$ NMR spectrum of compound <b>2</b> in $\text{DMSO}-d_6$ (500 MHz).....	186
<b>Figure S3.</b> $^1\text{H}$ NMR spectrum of compound <b>3</b> in $\text{DMSO}-d_6$ (500 MHz).....	186
<b>Figure S4.</b> $^1\text{H}$ NMR spectrum of compound <b>4</b> in $\text{DMSO}-d_6$ (500 MHz).....	187
<b>Figure S5.</b> $^1\text{H}$ NMR spectrum of compound <b>6</b> in $\text{DMSO}-d_6$ (500 MHz).....	187
<b>Figure S6.</b> $^{13}\text{C}$ NMR spectrum of compound <b>6</b> in $\text{DMSO}-d_6$ (125 MHz) .....	188
<b>Figure S7.</b> HSQC spectrum of compound <b>6</b> in $\text{DMSO}-d_6$ .....	188
<b>Figure S8.</b> COSY spectrum of compound <b>6</b> in $\text{DMSO}-d_6$ .....	189
<b>Figure S9.</b> HMBC spectrum of compound <b>6</b> in $\text{DMSO}-d_6$ .....	189
<b>Figure S10.</b> NOESY spectrum of compound <b>6</b> in $\text{DMSO}-d_6$ .....	190
<b>Figure S11.</b> $^1\text{H}$ NMR spectrum of compound <b>7</b> in $\text{DMSO}-d_6$ (500 MHz).....	190
<b>Figure S12.</b> $^{13}\text{C}$ NMR spectrum of compound <b>7</b> in $\text{DMSO}-d_6$ (125 MHz) .....	191
<b>Figure S13.</b> HSQC spectrum of compound <b>7</b> in $\text{DMSO}-d_6$ .....	191
<b>Figure S14.</b> COSY spectrum of compound <b>7</b> in $\text{DMSO}-d_6$ .....	192
<b>Figure S15.</b> HMBC spectrum of compound <b>7</b> in $\text{DMSO}-d_6$ .....	192
<b>Figure S16.</b> NOESY spectrum of compound <b>7</b> in $\text{DMSO}-d_6$ .....	193
<b>Figure S17.</b> $^1\text{H}$ NMR spectrum of compound <b>8</b> in $\text{DMSO}-d_6$ (500 MHz).....	193
<b>Figure S18.</b> $^{13}\text{C}$ NMR spectrum of compound <b>8</b> in $\text{DMSO}-d_6$ (125 MHz) .....	194
<b>Figure S19.</b> HSQC spectrum of compound <b>8</b> in $\text{DMSO}-d_6$ .....	194
<b>Figure S20.</b> COSY spectrum of compound <b>8</b> in $\text{DMSO}-d_6$ .....	195
<b>Figure S21.</b> HMBC spectrum of compound <b>8</b> in $\text{DMSO}-d_6$ .....	195
<b>Figure S22.</b> NOESY spectrum of compound <b>8</b> in $\text{DMSO}-d_6$ .....	196
<b>Figure S23.</b> $^1\text{H}$ NMR spectrum of compound <b>9</b> in $\text{DMSO}-d_6$ (500 MHz).....	196
<b>Figure S24.</b> $^{13}\text{C}$ NMR spectrum of compound <b>9</b> in $\text{DMSO}-d_6$ (125 MHz) .....	197
<b>Figure S25.</b> HSQC spectrum of compound <b>9</b> in $\text{DMSO}-d_6$ .....	197

<b>Figure S26.</b> COSY spectrum of compound <b>9</b> in DMSO- <i>d</i> <sub>6</sub> .....	198
<b>Figure S27.</b> HMBC spectrum of compound <b>9</b> in DMSO- <i>d</i> <sub>6</sub> .....	198
<b>Figure S28.</b> NOESY spectrum of compound <b>9</b> in DMSO- <i>d</i> <sub>6</sub> .....	199
<b>Figure S29.</b> <sup>1</sup> H NMR spectrum of compound <b>10</b> in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	199
<b>Figure S30.</b> <sup>13</sup> C NMR spectrum of compound <b>10</b> in DMSO- <i>d</i> <sub>6</sub> (125 MHz) .....	200
<b>Figure S31.</b> HSQC spectrum of compound <b>10</b> in DMSO- <i>d</i> <sub>6</sub> .....	200
<b>Figure S32.</b> COSY spectrum of compound <b>10</b> in DMSO- <i>d</i> <sub>6</sub> .....	201
<b>Figure S33.</b> HMBC spectrum of compound <b>10</b> in DMSO- <i>d</i> <sub>6</sub> .....	201
<b>Figure S34.</b> NOESY spectrum of compound <b>10</b> in DMSO- <i>d</i> <sub>6</sub> .....	202
<b>Figure S35.</b> <sup>1</sup> H NMR spectrum of compound <b>11</b> in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	202
<b>Figure S36.</b> <sup>13</sup> C NMR spectrum of compound <b>11</b> in DMSO- <i>d</i> <sub>6</sub> (125 MHz) .....	203
<b>Figure S37.</b> HSQC spectrum of compound <b>11</b> in DMSO- <i>d</i> <sub>6</sub> .....	203
<b>Figure S38.</b> COSY spectrum of compound <b>11</b> in DMSO- <i>d</i> <sub>6</sub> .....	204
<b>Figure S39.</b> HMBC spectrum of compound <b>11</b> in DMSO- <i>d</i> <sub>6</sub> .....	204
<b>Figure S40.</b> NOESY spectrum of compound <b>11</b> in DMSO- <i>d</i> <sub>6</sub> .....	205
<b>Figure S41.</b> <sup>1</sup> H NMR spectrum of compound <b>12</b> in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	205
<b>Figure S42.</b> <sup>13</sup> C NMR spectrum of compound <b>12</b> in DMSO- <i>d</i> <sub>6</sub> (125 MHz). .....	206
<b>Figure S43.</b> HSQC spectrum of compound <b>12</b> in DMSO- <i>d</i> <sub>6</sub> .....	206
<b>Figure S44.</b> COSY spectrum of compound <b>12</b> in DMSO- <i>d</i> <sub>6</sub> .....	207
<b>Figure S45.</b> HMBC spectrum of compound <b>12</b> in DMSO- <i>d</i> <sub>6</sub> .....	207
<b>Figure S46.</b> NOESY spectrum of compound <b>12</b> in DMSO- <i>d</i> <sub>6</sub> .....	208
<b>Figure S47.</b> <sup>1</sup> H NMR spectrum of compound <b>13</b> in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	208
<b>Figure S48.</b> <sup>13</sup> C NMR spectrum of compound <b>13</b> in DMSO- <i>d</i> <sub>6</sub> (125 MHz) .....	209
<b>Figure S49.</b> HSQC spectrum of compound <b>13</b> in DMSO- <i>d</i> <sub>6</sub> .....	209
<b>Figure S50.</b> COSY spectrum of compound <b>13</b> in DMSO- <i>d</i> <sub>6</sub> .....	210
<b>Figure S51.</b> HMBC spectrum of compound <b>13</b> in DMSO- <i>d</i> <sub>6</sub> .....	210
<b>Figure S52.</b> NOESY spectrum of compound <b>13</b> in DMSO- <i>d</i> <sub>6</sub> .....	211
<b>Figure S53.</b> <sup>1</sup> H NMR spectrum of compound <b>14</b> in DMSO- <i>d</i> <sub>6</sub> (500 MHz).....	211
<b>Figure S54.</b> <sup>13</sup> C NMR spectrum of compound <b>14</b> in DMSO- <i>d</i> <sub>6</sub> (125 MHz) .....	212
<b>Figure S55.</b> HSQC spectrum of compound <b>14</b> in DMSO- <i>d</i> <sub>6</sub> .....	212
<b>Figure S56.</b> COSY spectrum of compound <b>14</b> in DMSO- <i>d</i> <sub>6</sub> .....	213
<b>Figure S57.</b> HMBC spectrum of compound <b>14</b> in DMSO- <i>d</i> <sub>6</sub> .....	213
<b>Figure S58.</b> NOESY spectrum of compound <b>14</b> in DMSO- <i>d</i> <sub>6</sub> .....	214

<b>Figure S59.</b> Experimental ECD spectra of <b>2</b> — <b>4</b> in MeOH.....	215
<b>Figure S60.</b> Experimental ECD spectra of <b>6</b> — <b>9</b> in MeOH.....	215
<b>Figure S61.</b> Experimental ECD spectra of <b>10</b> — <b>13</b> in MeOH.....	215
<b>Figure S62.</b> Experimental ECD spectra of <b>13</b> — <b>14</b> in MeOH.....	215
<b>Figure S63.</b> LC-MS analysis of <i>sasB</i> transformant after incubating with <i>cyclo</i> -Trp-Ala isomers.....	216
<b>Figure S64.</b> LC-MS analysis of <i>sasB</i> transformant after incubating with <i>cyclo</i> -Trp-Pro isomers.....	217
<b>Figure S65.</b> Purification of <b>13</b> and <b>14</b> on the Multohigh Chiral AM-RP column. ....	218
<b>Figure S66.</b> LC-MS analysis of <b>13</b> after incubation in CD <sub>3</sub> OD/D <sub>2</sub> O (1:1) for 14 h.....	219
<b>Reference</b> .....	220



## Supplementary Tables

**Table S1.** Comparison of sas genes with known entries

Protein	Accession No.	Length (aa)	Sequence identity (length in aa) in %	
			<i>S. leeuwenhoekii</i> NRRL B-24963 <sup>1</sup>	<i>S. youssoufiensis</i> OUC6819 <sup>2</sup>
SasA	WP_079103588.1	252	SazA (252) 82	DmtA1 (233), 41
SasB	WP_121505431.1	347	SazB-PT domain (314) 85	DmtC1 (311), 38
SasC	WP_055513754.1	290	SazB-MT domain (276) 82	

**Table S2.** Bacterial strains used in this study

Strain	Source	Cultivation media
<i>E. coli</i> DH5α	Invitrogen	LB
<i>E. coli</i> ET12567/pUZ8002	3	LB
<i>Streptomyces albus</i> J1074	4	MS
<i>Streptomyces aurantiacus</i> NRRL ISP-5412	NRRL	modified R5

NRRL: ARS Culture Collection

LB medium: tryptone 10.0 g/L, yeast extract 5.0 g/L, NaCl 10.0 g/L.

MS medium: mannitol 20.0 g/L, soya flour 20.0 g/L, agar 20.0 g/L.

Modified R5 medium: sucrose 103.0 g/L, glucose 10.0 g/L, yeast extract 5.0 g/L, MgCl<sub>2</sub>·6H<sub>2</sub>O 10.12 g/L, K<sub>2</sub>SO<sub>4</sub> 0.25 g/L, Difco casaminoacids 0.1 g/L, MOPS 21.0 g/L, trace element solution 2 mL/L, pH 7.2.

**Table S3.** Cloning and expression constructs used in this study

Gene	Primer sequences (5'-3')	Cloning construct	Expression vector	Cloning sites	Expression constructs
<i>sasA</i>	CATATGTCCAGCAAGGACGTCGAC TCTAGACTATGTGCGTTGACTTCCTTC	pJL81	pPWW50	<i>NdeI/XbaI</i>	pJL87
<i>sasABC</i>	CATATGTCCAGCAAGGACGTCGAC ACTAGTCTGCGTTACCGGGTCG	pJL82	pPWW50	<i>NdeI/SpeI</i>	pJL88
<i>sasAB</i>	CATATGTCCAGCAAGGACGTCGAC ACTAGTTCACCGGTCCGTCTCCGC	pJL83	pPWW50	<i>NdeI/SpeI</i>	pJL89
<i>sasAC</i>	CATATGTCCAGCAAGGACGTCGAC AGATCTTGTCCAGCAAGGACGTCGAC	pJL84	pPWW50	<i>NdeI/BglII</i>	pJL90
	ACTAGTATGTATCAGTCCGGGACCCGTT TC	pJL86		<i>SpeI/XbaI</i>	
<i>sasB</i>	TCTAGATCACC GGTCGGACCGCTG CATATGAGCCAGCGAGAACTCACC ACTAGTTCACCGGTCCGTCTCCGC	pJL85	pPWW50	<i>NdeI/SpeI</i>	pJL91
<i>sasC</i>	ACTAGTATGTATCAGTCCGGGACCCGTT TC TCTAGATCACC GGTCGGACCGCTG	pJL86	pPWW50	<i>SpeI/XbaI</i>	pJL92

Restriction sites for cloning are underlined in the primer sequences. Cloning constructs are based on pGEM T EASY vector.

**Table S4.**  $^1\text{H}$  NMR data of compounds **2** — **4** in  $\text{DMSO-}d_6$ 

	<b>2</b>	<b>3</b>	<b>4</b>
Position	$\delta_{\text{H}}$ , multi. ( <i>J</i> in Hz)	$\delta_{\text{H}}$ , multi. ( <i>J</i> in Hz)	$\delta_{\text{H}}$ , multi. ( <i>J</i> in Hz)
1		6.53, d (3.0)	6.49, d (3.4)
2	5.10, s	5.28, d (3.0)	5.28, d (3.4)
4	7.11, dd (7.4, 0.9)	7.05, d (7.4)	7.08, d (7.3)
5	6.70, td (7.4, 0.9)	6.59, td (7.4, 1.0)	6.59, t (7.3)
6	7.08, td (7.8, 1.3)	6.98, td (7.8, 1.0)	6.94, td (7.8, 0.9)
7	6.47, d (7.8)	6.49, d (7.8)	6.44, d (7.8)
10 $\alpha$	2.38, dd (12.9, 7.2)	2.33, dd (13.0, 7.7)	2.41, dd (13.1, 7.7)
10 $\beta$	1.96, dd (12.9, 11.1)	1.98, dd (13.0, 10.0)	1.99, dd (13.1, 10.7)
11	4.76, dd (11.1, 7.2)	4.50, t (9.0)	4.66, dd (10.7, 7.7)
14	4.76, dd (11.1, 7.2)	4.37, t (5.5)	4.66, dd (10.7, 7.7)
15		7.59, s	
17		10.81, s	
18	5.10, s	7.22, d (2.3)	5.28, d (3.4)
20	7.11, dd (7.4, 0.9)	7.57, d (7.8)	7.08, d (7.3)
21	6.70, td (7.4, 0.9)	6.95, td (7.8, 1.0)	6.59, t (7.3)
22	7.08, td (7.8, 1.3)	7.07, td (8.1, 1.0)	6.94, td (7.8, 0.9)
23	6.47, d (7.8)	7.32, d (8.1)	6.44, d (7.8)
26 $\alpha$	2.38, dd (12.9, 7.2)	3.37, dd (15.2, 5.5)	2.41, dd (13.1, 7.7)
26 $\beta$	1.96, dd (12.9, 11.2)	3.00, dd (15.2, 6.7)	1.99, dd (13.1, 10.7)
1' / 1''	2.54, d (7.4)	2.45, d (7.1)	2.51, d (7.3)
2' / 2''	4.99, br t (7.3)	5.02, br t (7.3)	5.03, t (7.1)
4' / 4''	1.61, s	1.58, s	1.60, s
5' / 5''	1.64, s	1.64, s	1.64, s
6' / 6''	2.87, s		

The NMR data of **2** — **4** correspond well to those reported previously.<sup>1</sup>

## Supplementary Figures

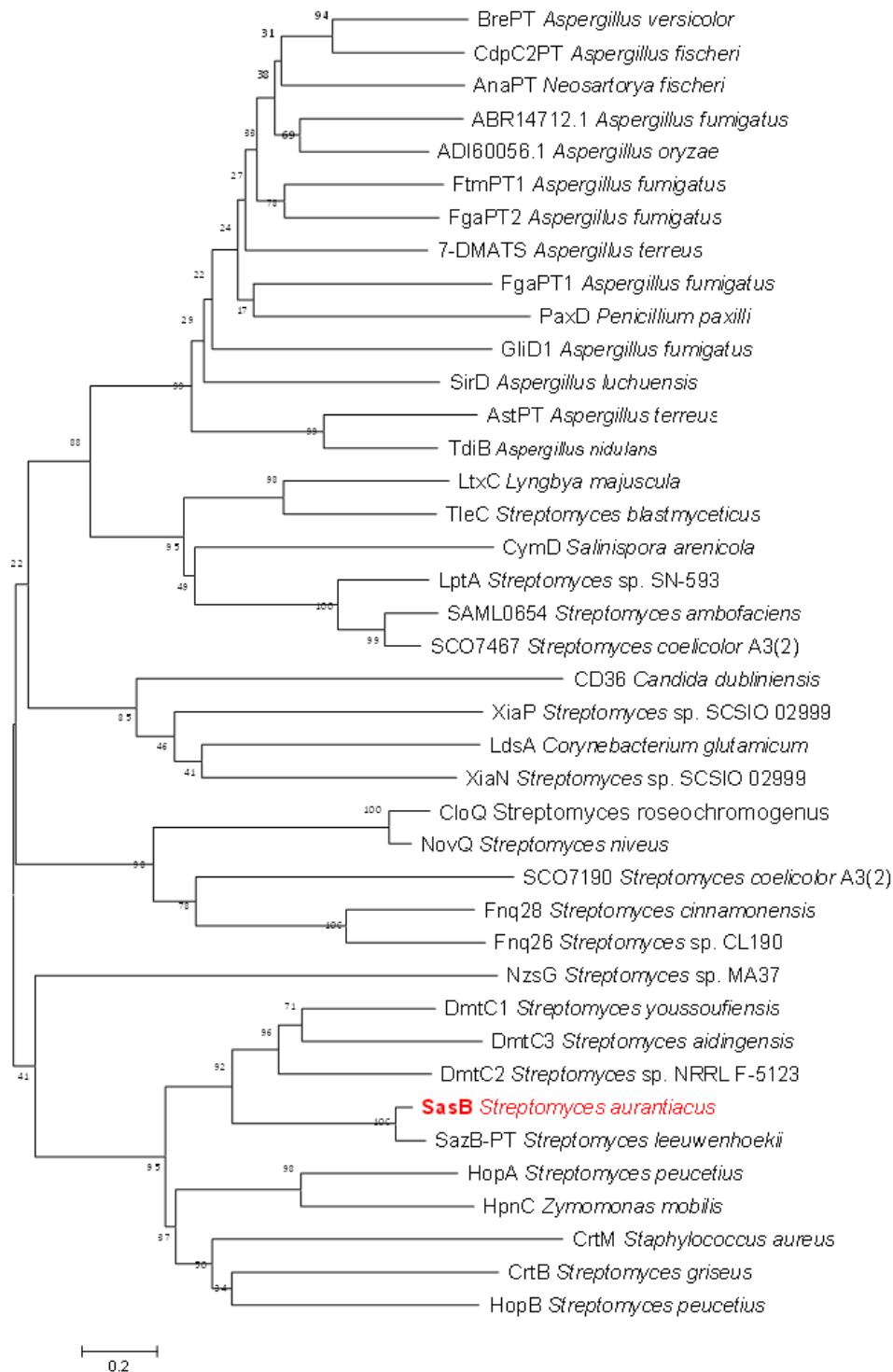
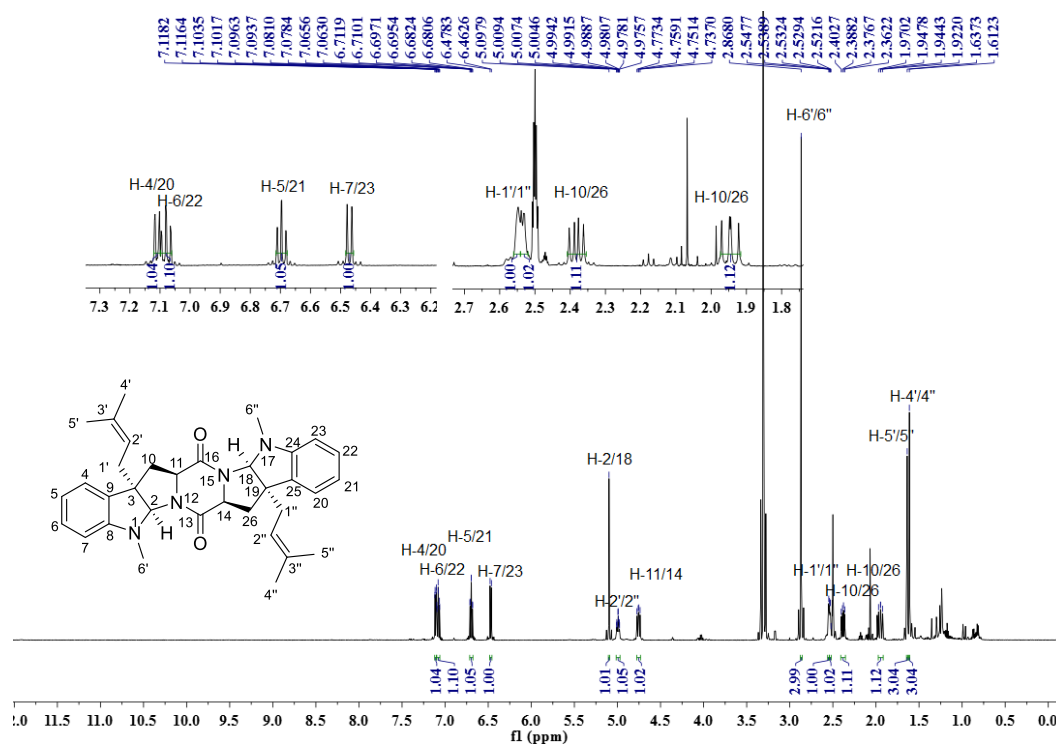
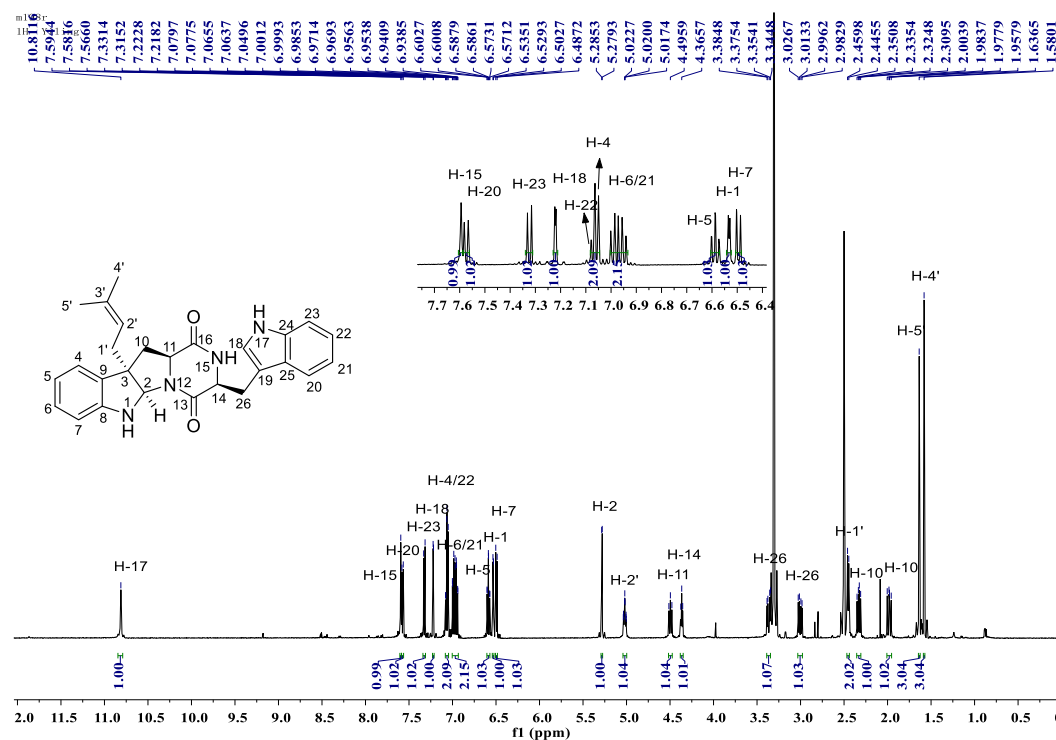


Figure S1. Phylogenetic analysis of PTs

Investigated in this study (in bold red) and functionally characterized PTs from bacteria and fungi. The protein sequences were downloaded from NCBI database.

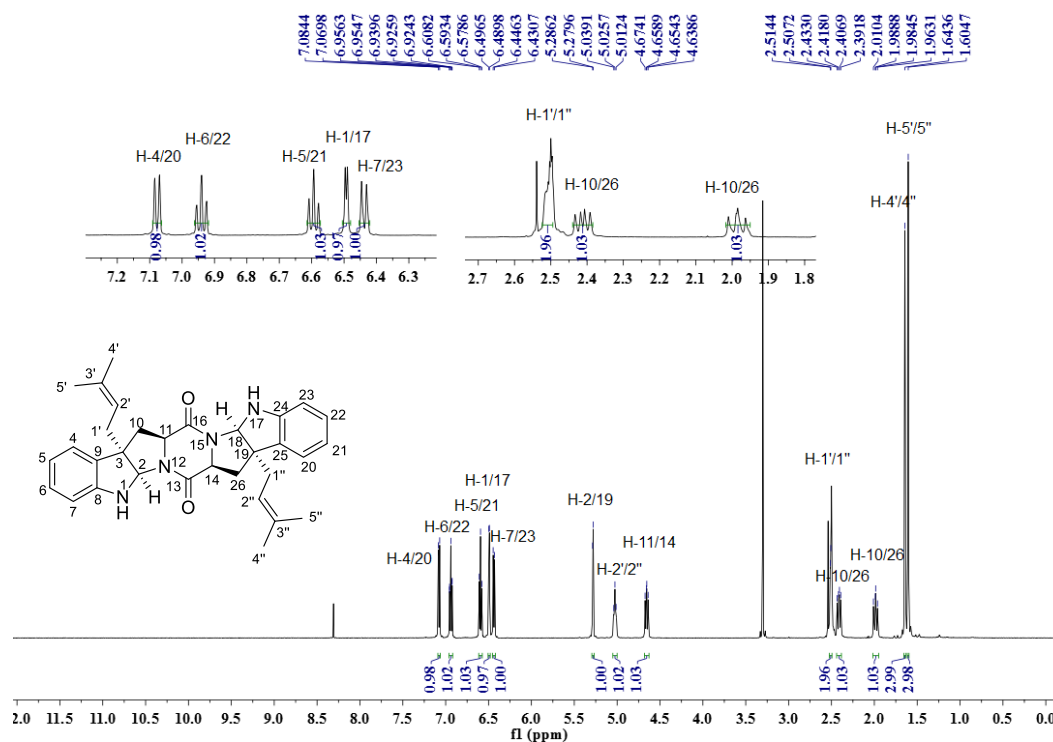


**Figure S2.**  $^1\text{H}$  NMR spectrum of compound **2** in  $\text{DMSO}-d_6$  (500 MHz)

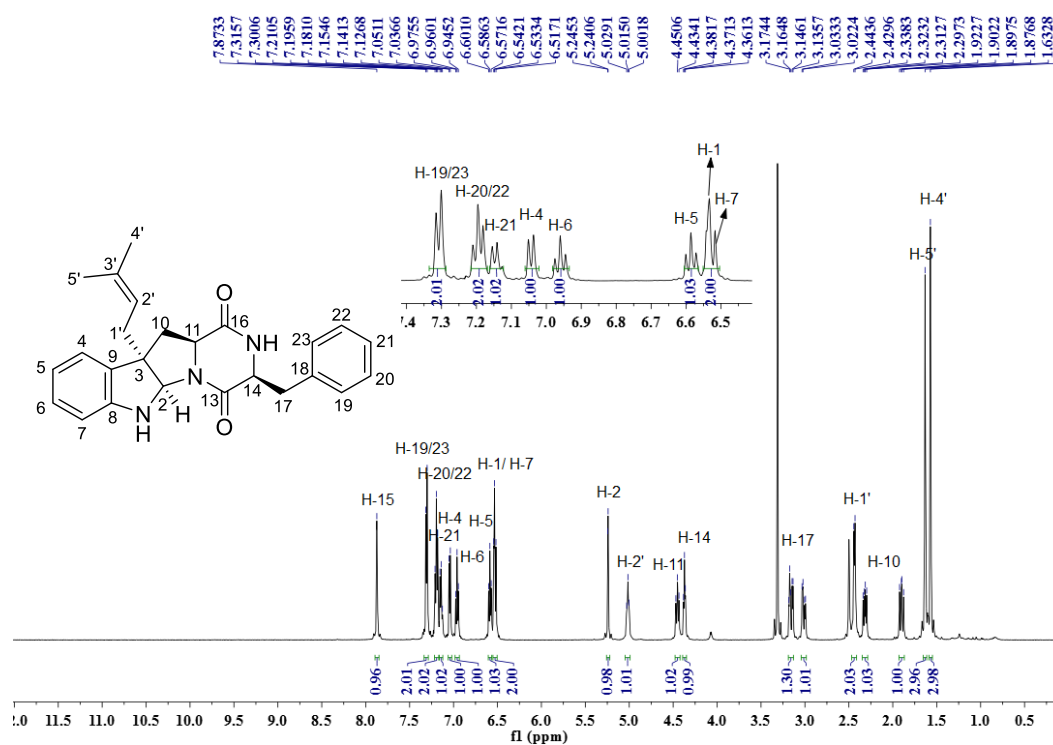


**Figure S3.**  $^1\text{H}$  NMR spectrum of compound **3** in  $\text{DMSO}-d_6$  (500 MHz)

# PUBLICATIONS

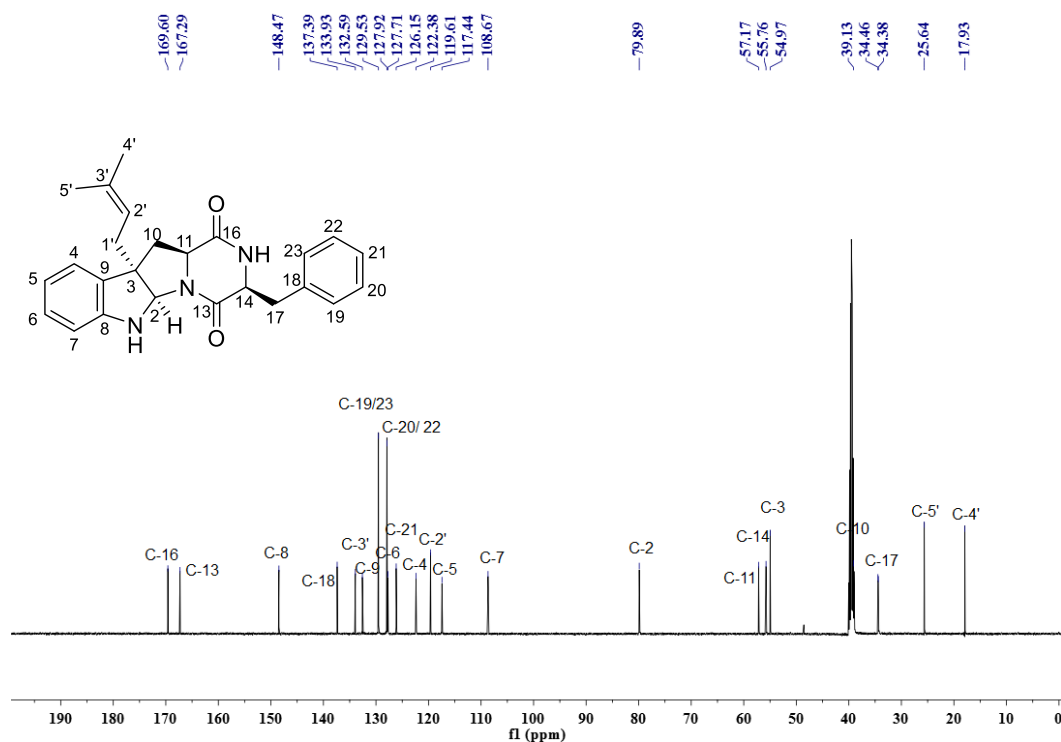


**Figure S4.**  $^1\text{H}$  NMR spectrum of compound **4** in  $\text{DMSO}-d_6$  (500 MHz)

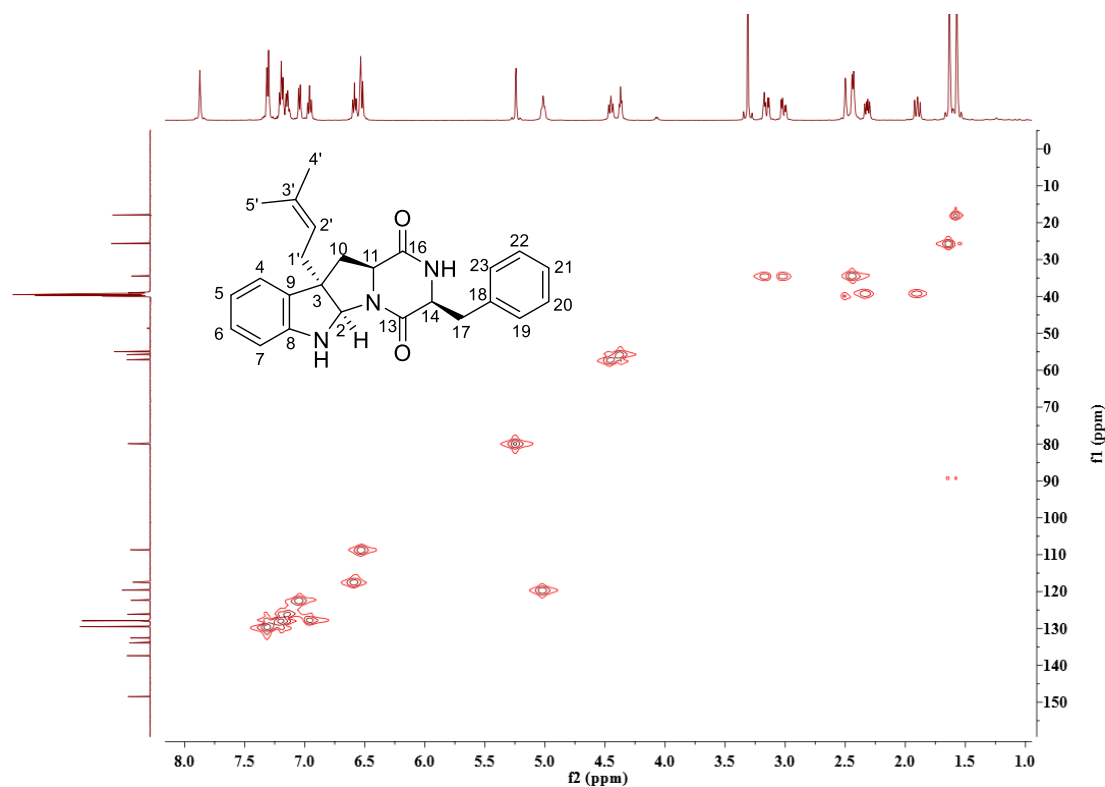


**Figure S5.**  $^1\text{H}$  NMR spectrum of compound **6** in  $\text{DMSO}-d_6$  (500 MHz)

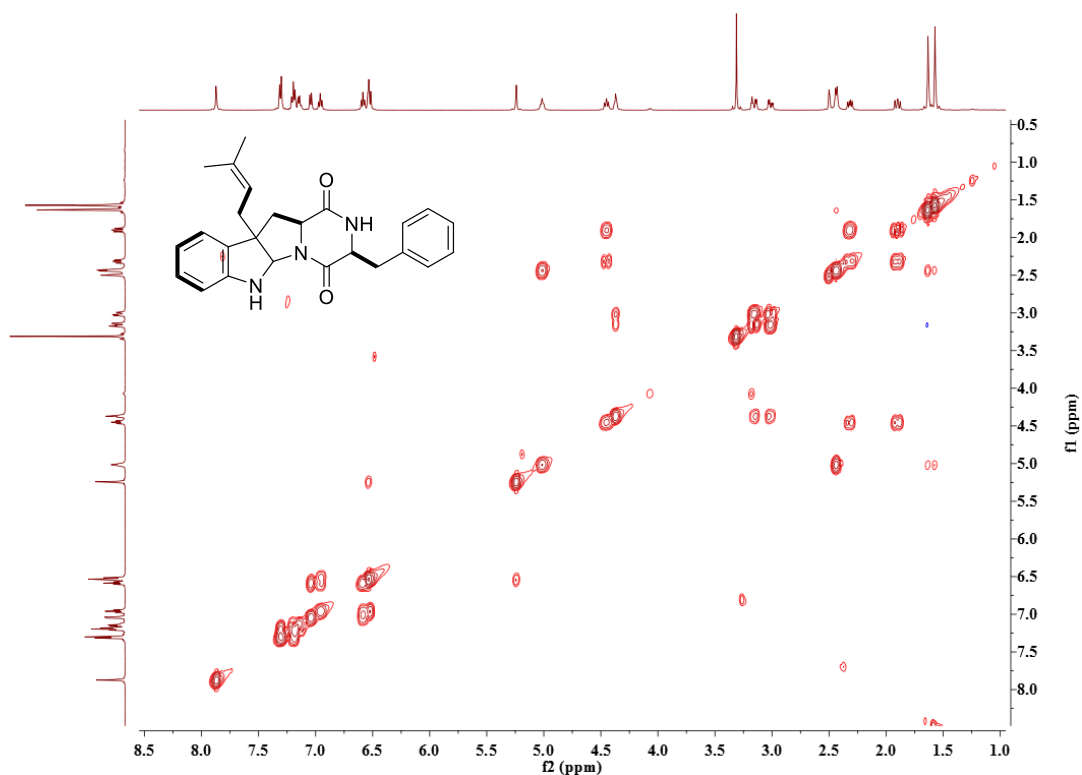
# PUBLICATIONS



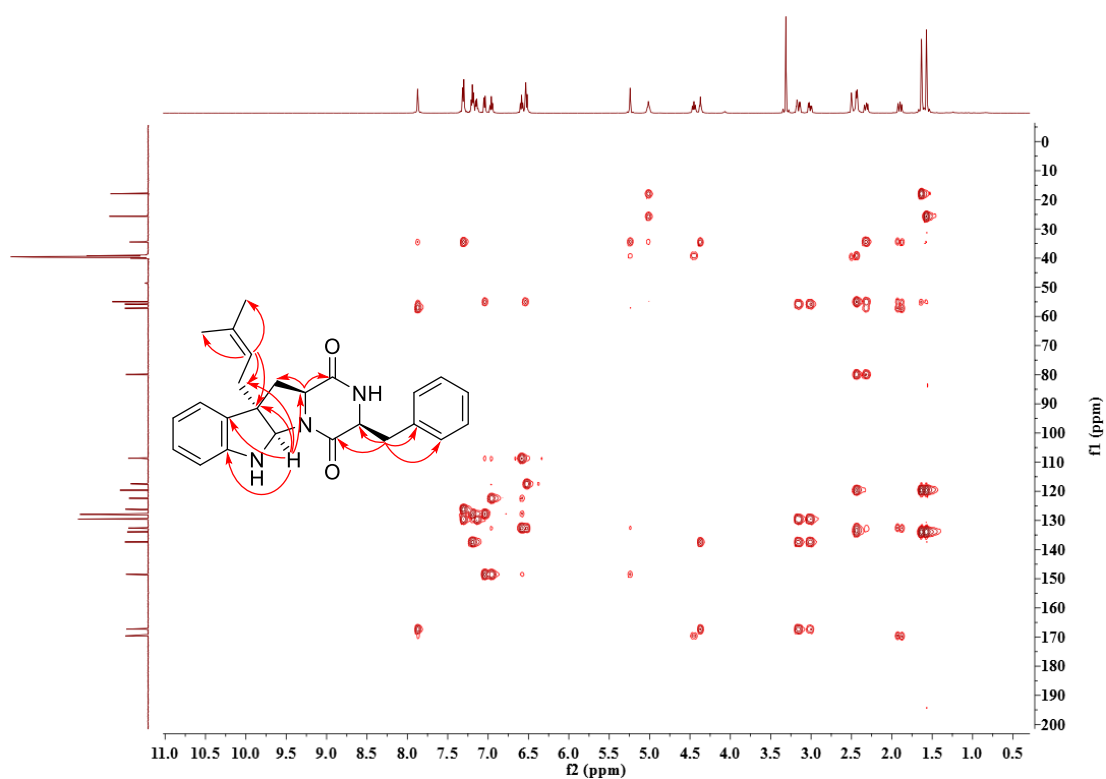
**Figure S6.**  $^{13}\text{C}$  NMR spectrum of compound **6** in  $\text{DMSO}-d_6$  (125 MHz)



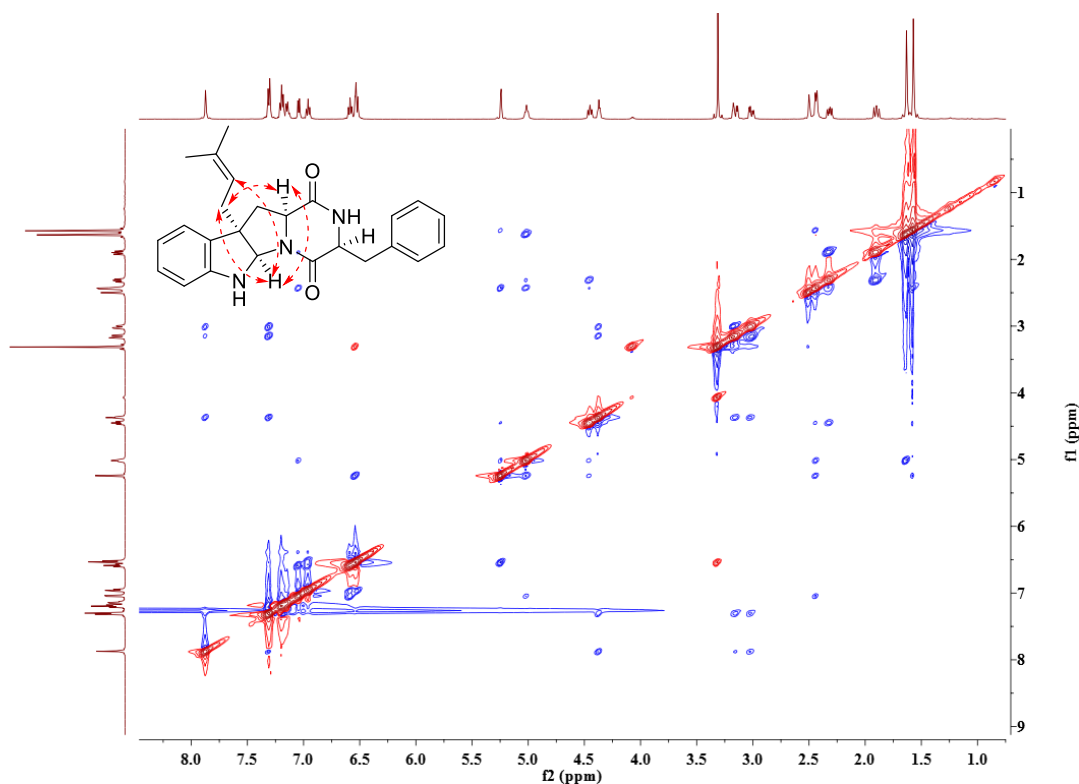
**Figure S7.** HSQC spectrum of compound **6** in  $\text{DMSO}-d_6$



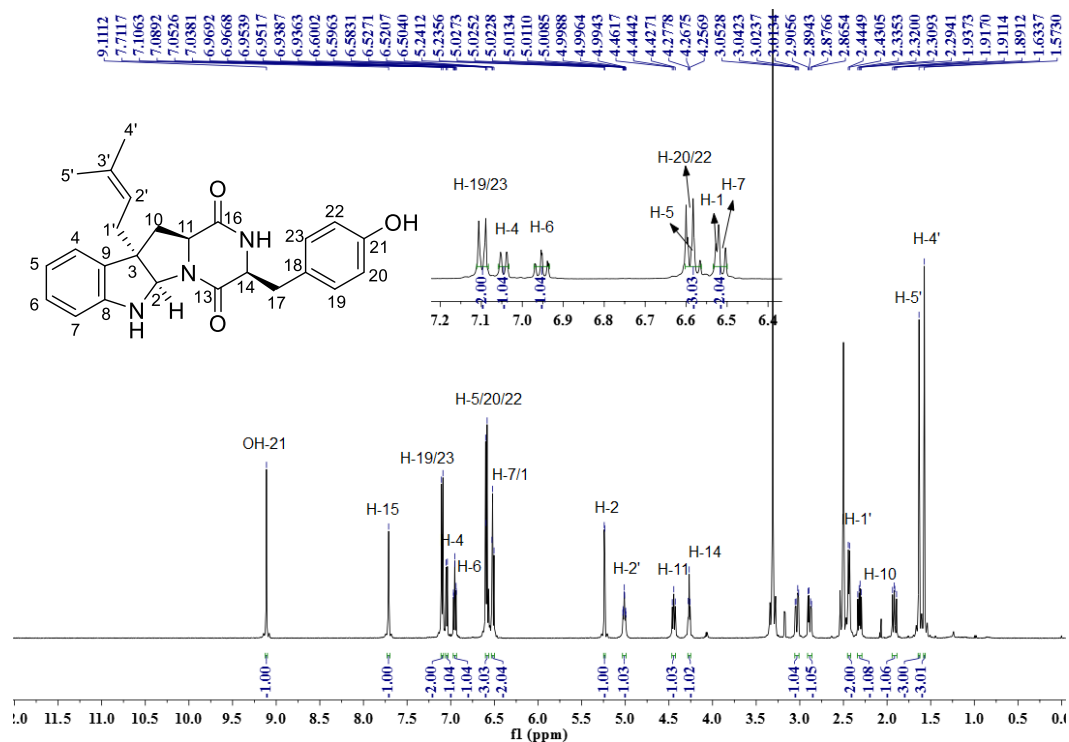
**Figure S8.** COSY spectrum of compound **6** in DMSO-*d*<sub>6</sub>



**Figure S9.** HMBC spectrum of compound **6** in DMSO-*d*<sub>6</sub>



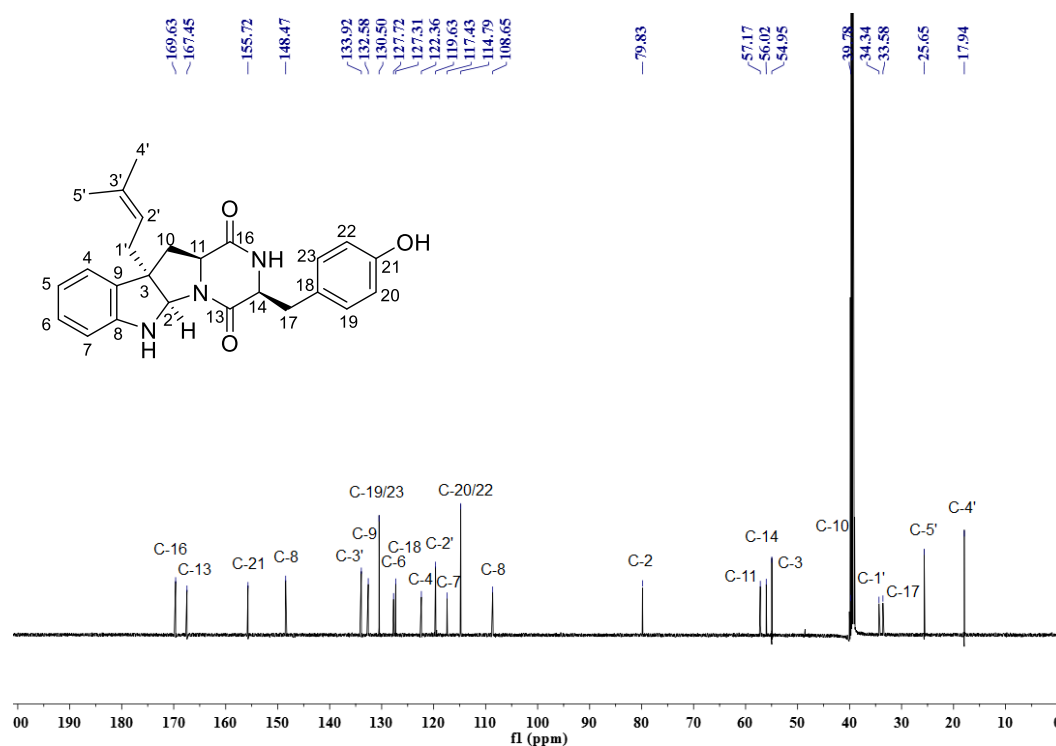
**Figure S10.** NOESY spectrum of compound **6** in DMSO- $d_6$



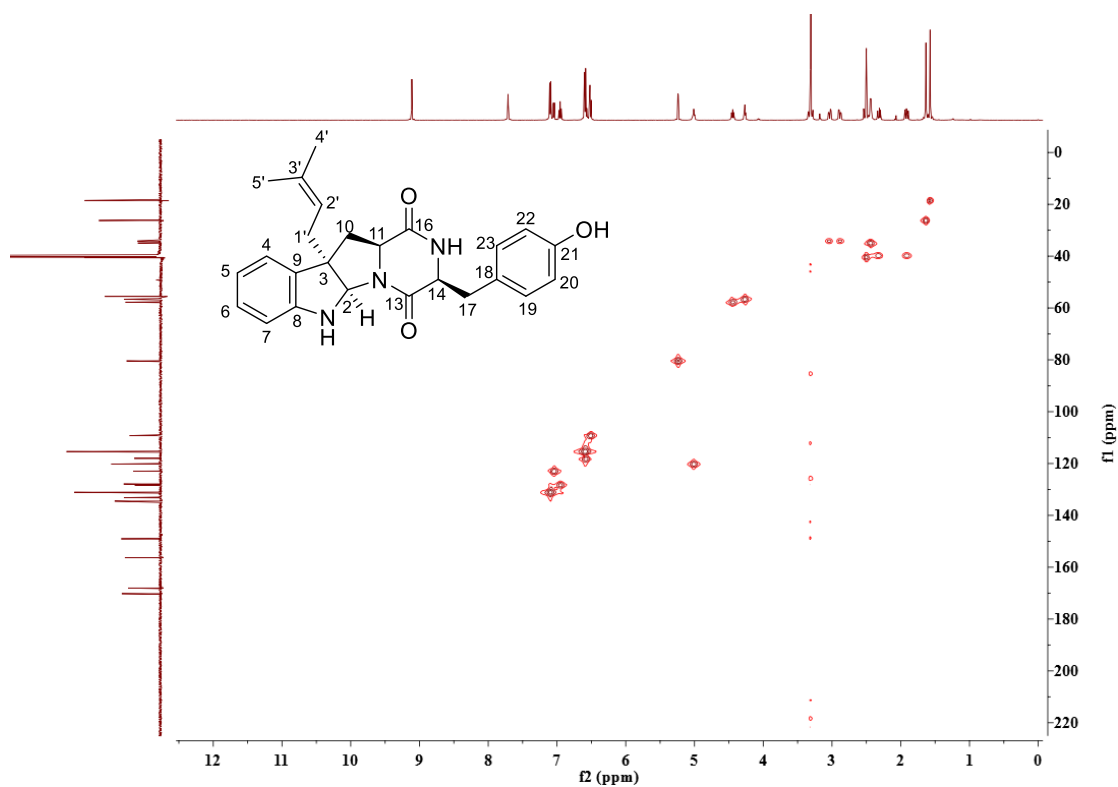
**Figure S11.**  $^1\text{H}$  NMR spectrum of compound **7** in DMSO- $d_6$  (500 MHz)



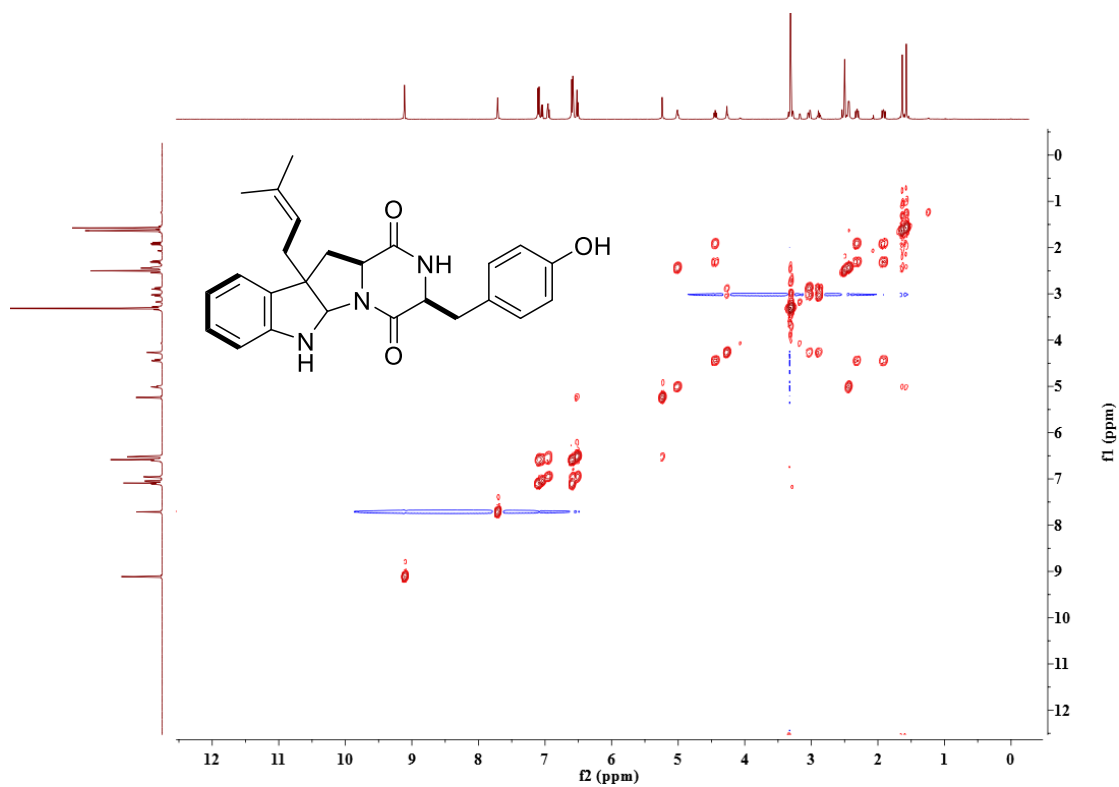
# PUBLICATIONS



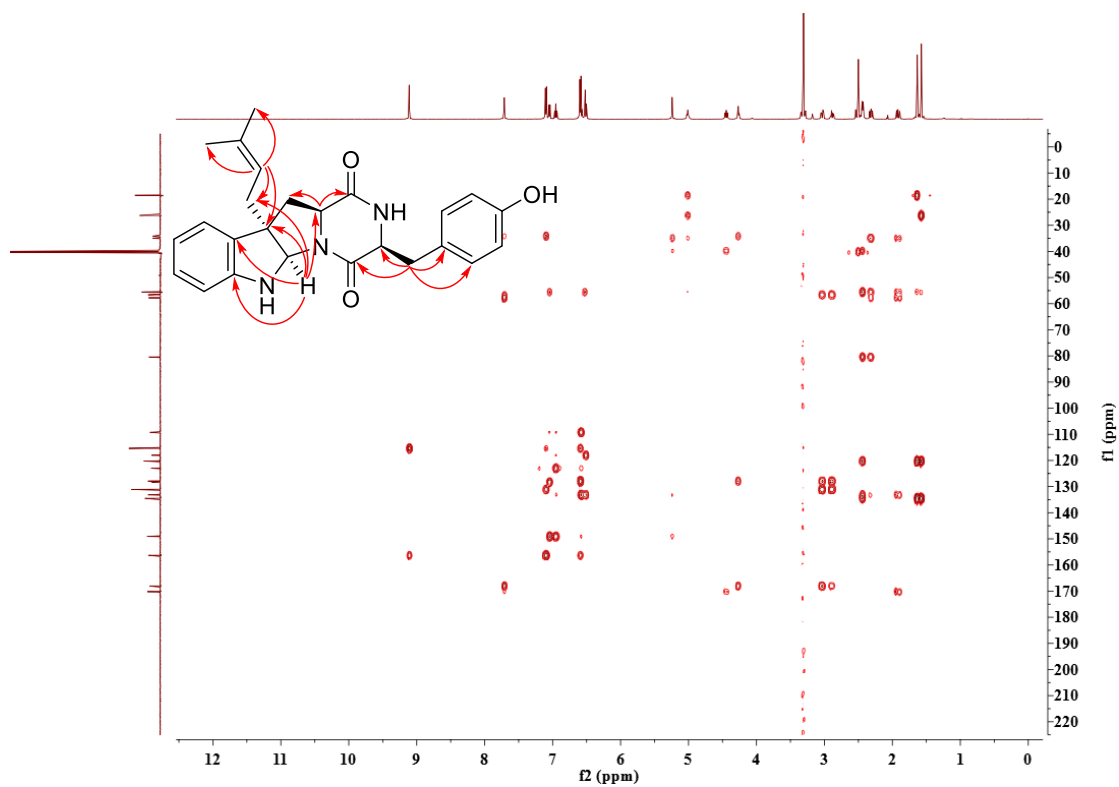
**Figure S12.**  $^{13}\text{C}$  NMR spectrum of compound **7** in  $\text{DMSO}-d_6$  (125 MHz)



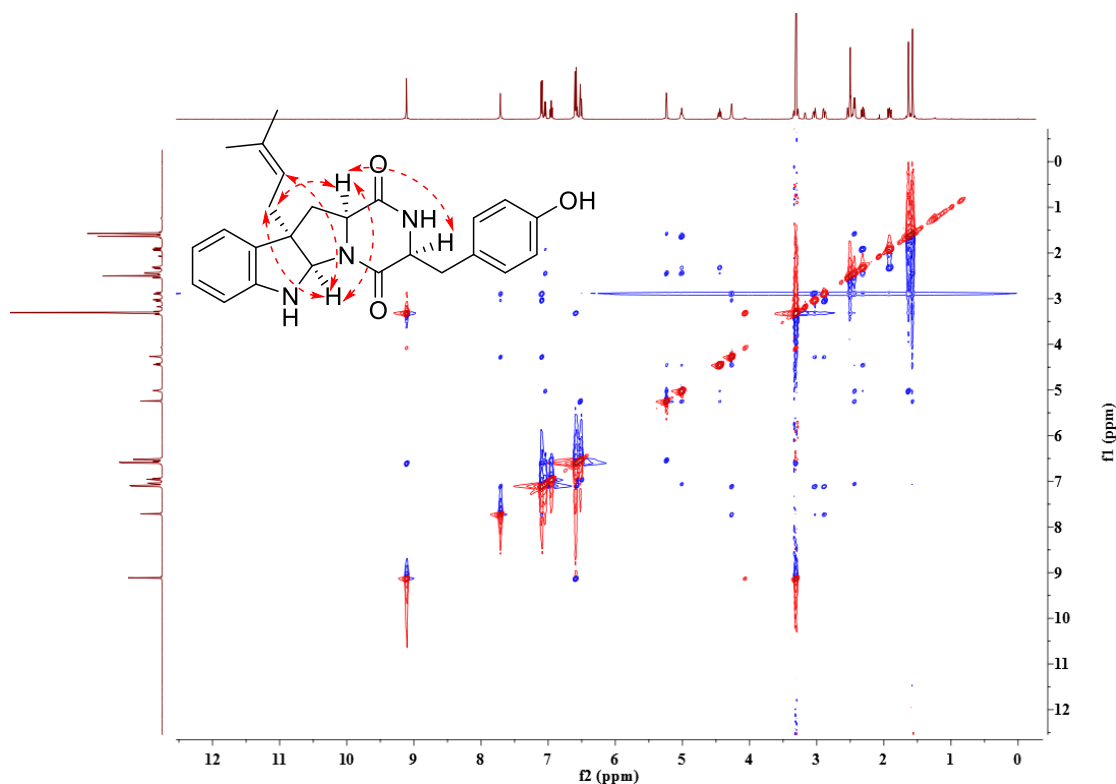
**Figure S13.** HSQC spectrum of compound **7** in  $\text{DMSO}-d_6$



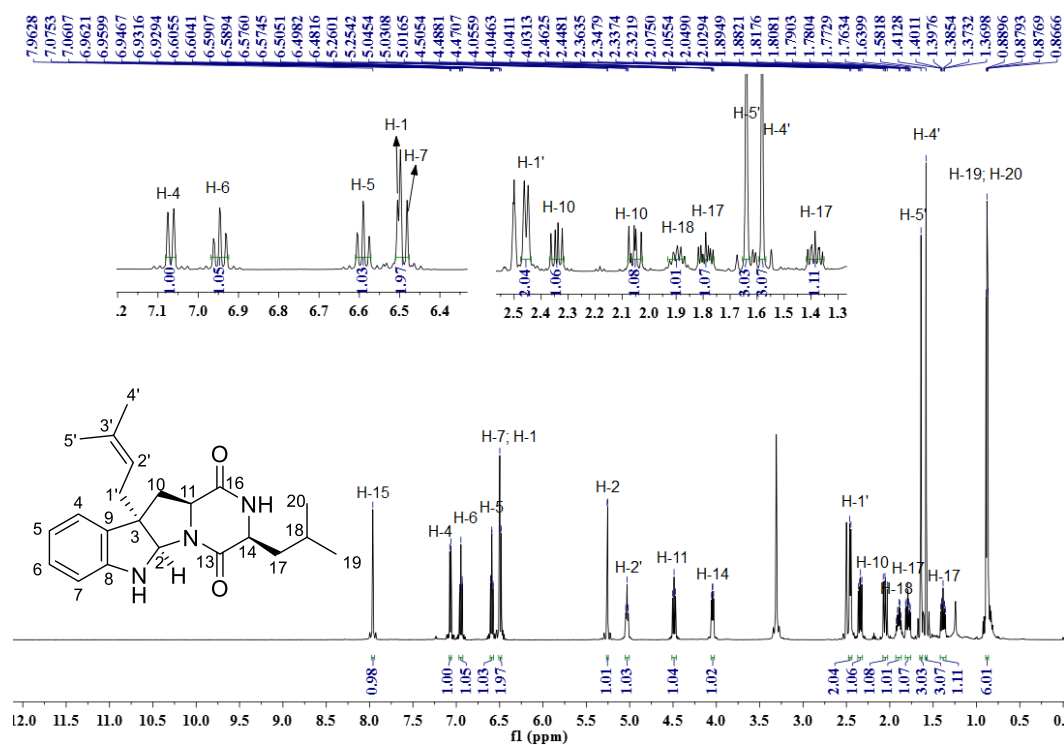
**Figure S14.** COSY spectrum of compound **7** in DMSO-*d*<sub>6</sub>



**Figure S15.** HMBC spectrum of compound **7** in DMSO-*d*<sub>6</sub>

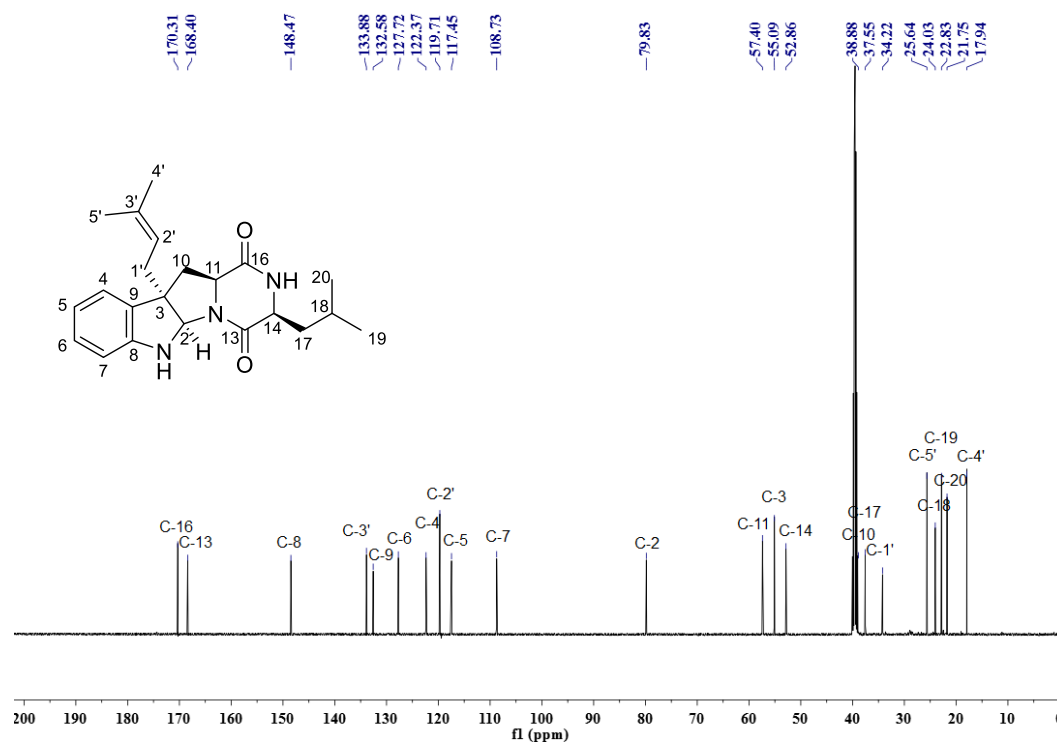


**Figure S16.** NOESY spectrum of compound **7** in DMSO- $d_6$

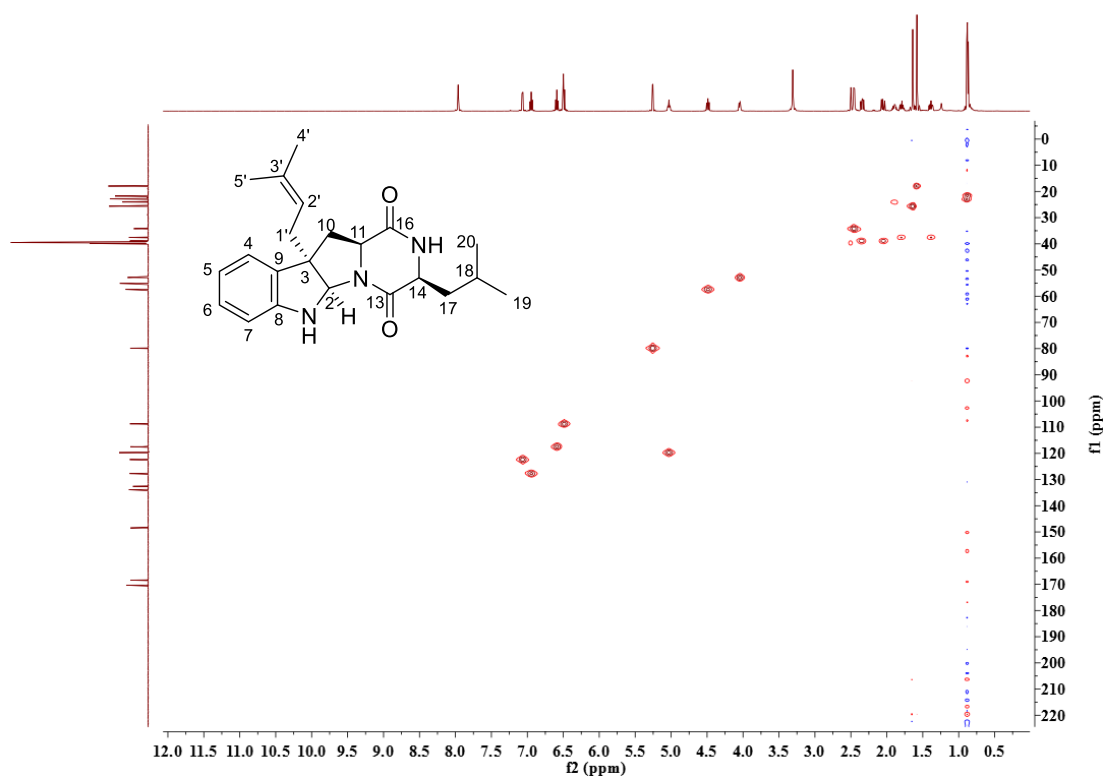


**Figure S17.**  $^1\text{H}$  NMR spectrum of compound **8** in DMSO- $d_6$  (500 MHz)

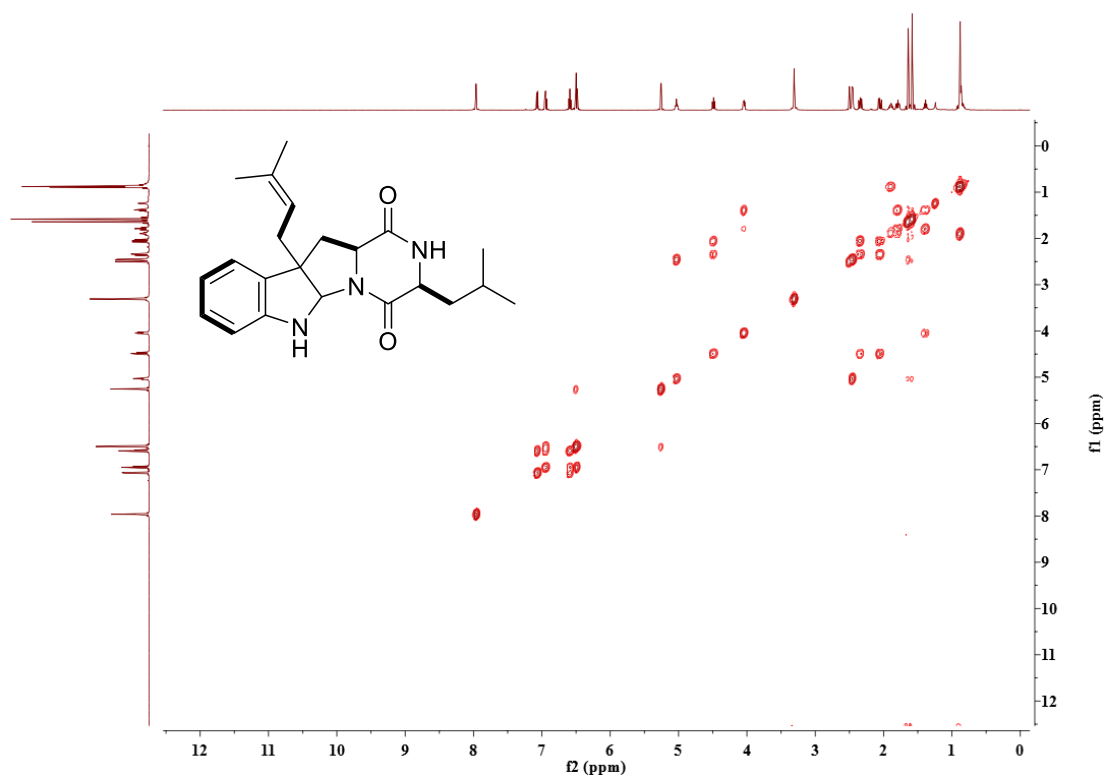
# PUBLICATIONS



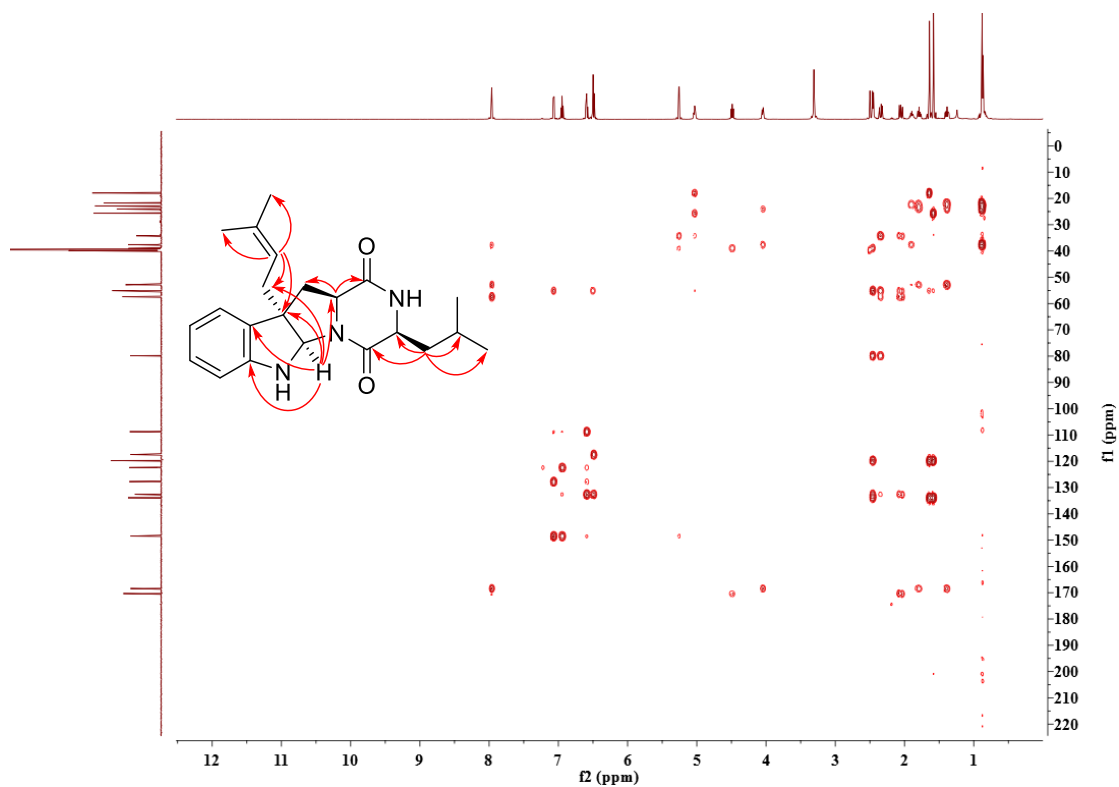
**Figure S18.**  $^{13}\text{C}$  NMR spectrum of compound **8** in  $\text{DMSO}-d_6$  (125 MHz)



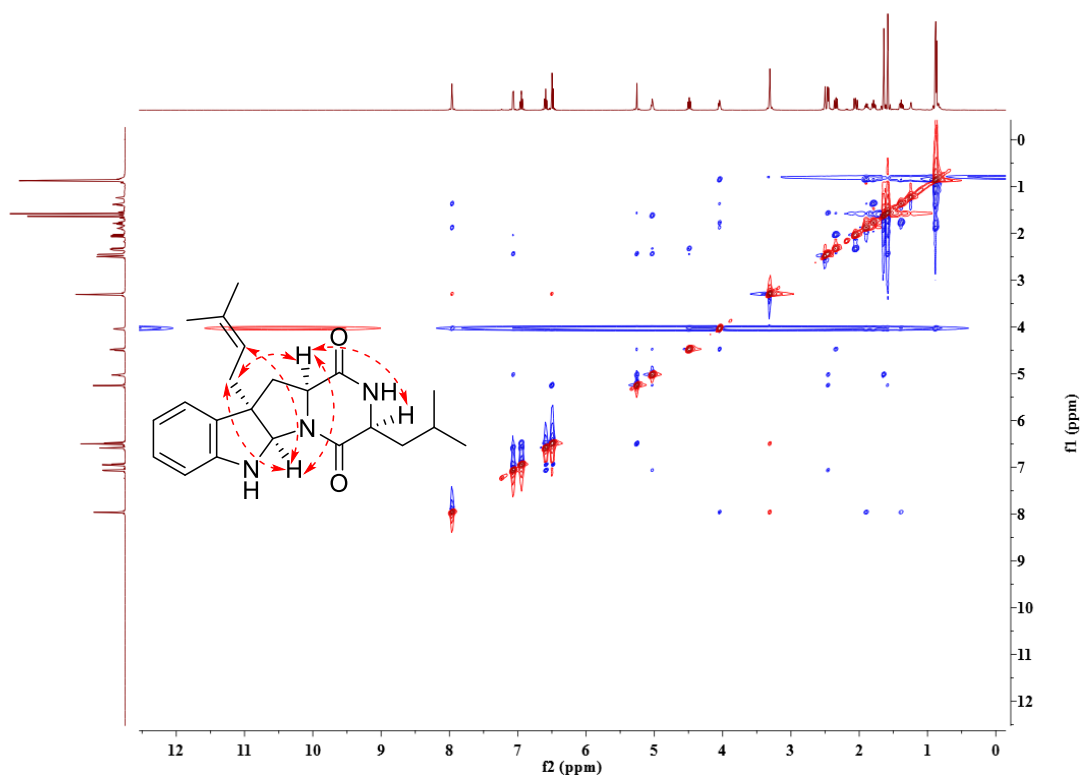
**Figure S19.** HSQC spectrum of compound **8** in  $\text{DMSO}-d_6$



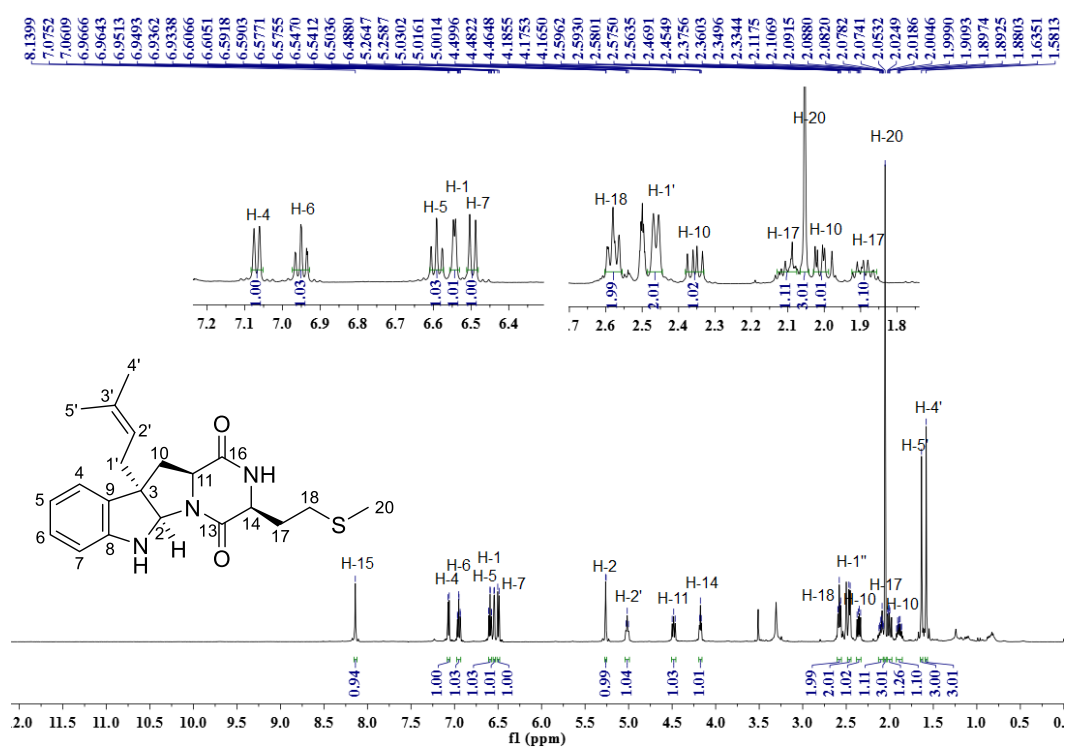
**Figure S20.** COSY spectrum of compound **8** in DMSO- $d_6$



**Figure S21.** HMBC spectrum of compound **8** in DMSO- $d_6$

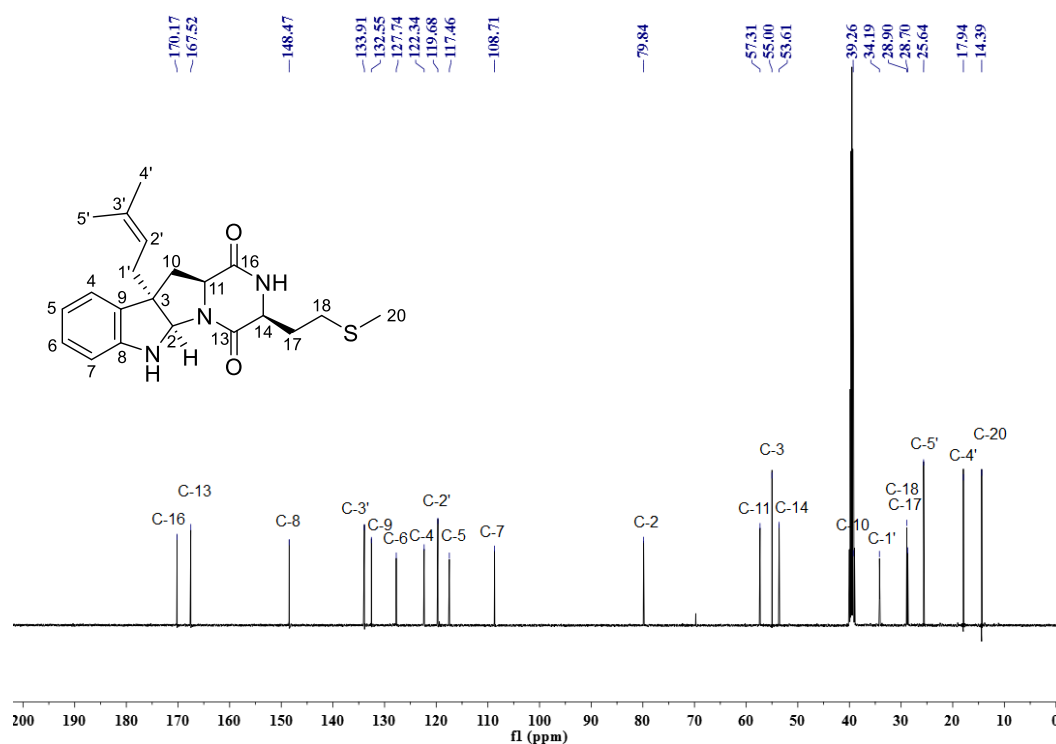


**Figure S22.** NOESY spectrum of compound **8** in DMSO- $d_6$

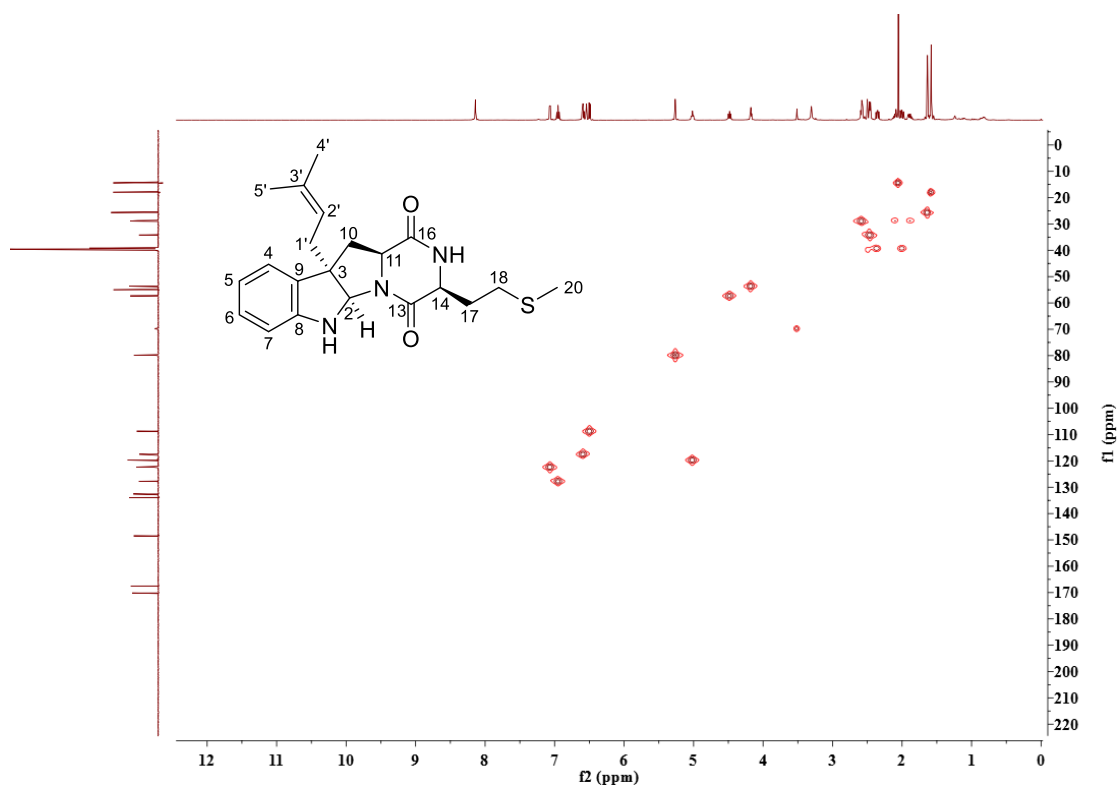


**Figure S23.**  $^1\text{H}$  NMR spectrum of compound **9** in DMSO- $d_6$  (500 MHz)

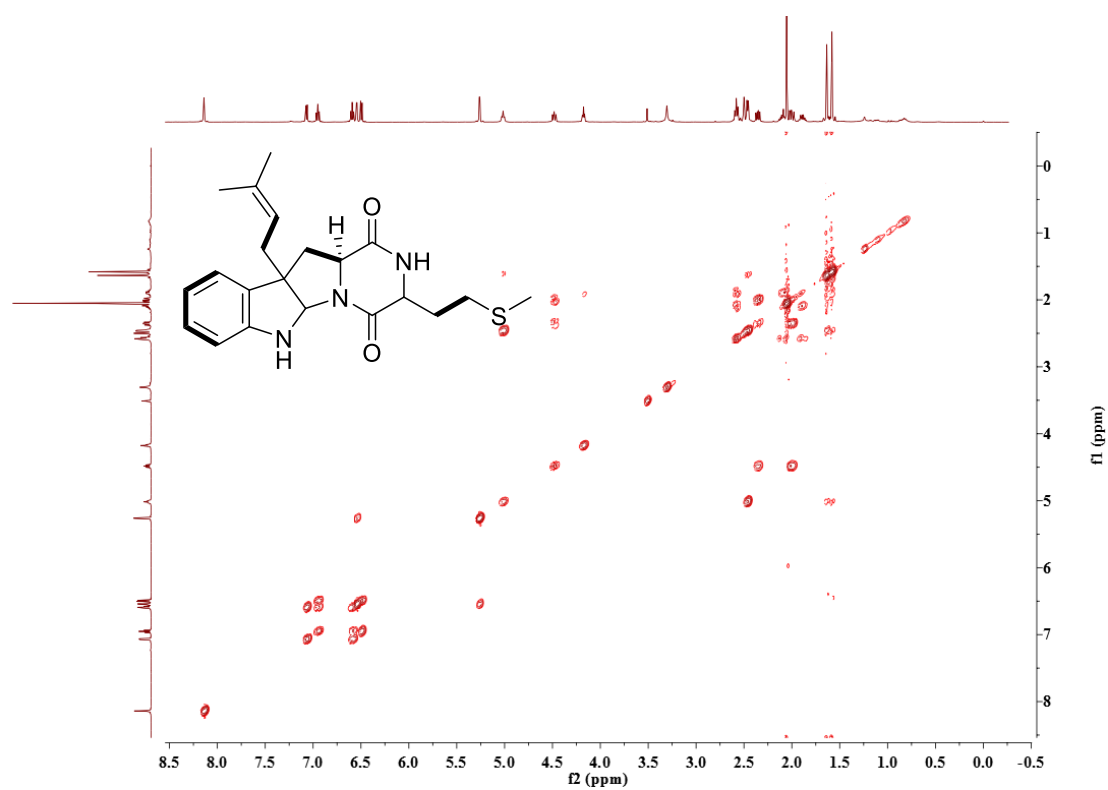
# PUBLICATIONS



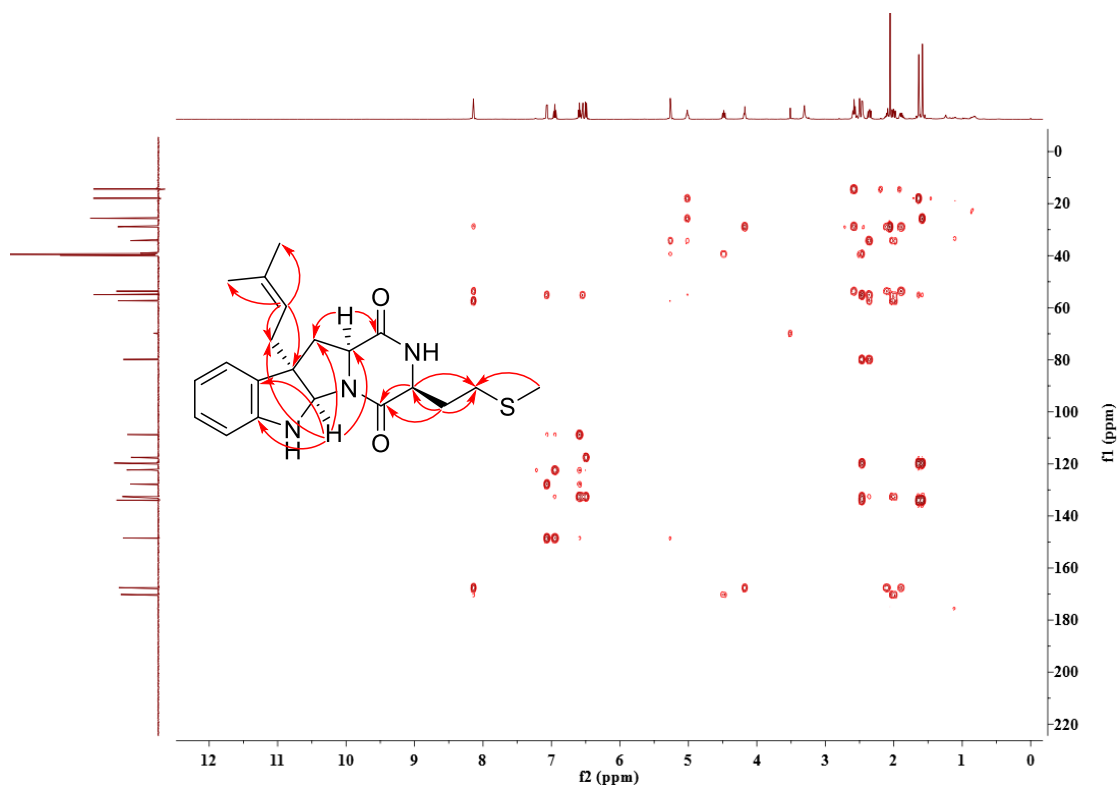
**Figure S24.**  $^{13}\text{C}$  NMR spectrum of compound **9** in  $\text{DMSO}-d_6$  (125 MHz)



**Figure S25.** HSQC spectrum of compound **9** in  $\text{DMSO}-d_6$

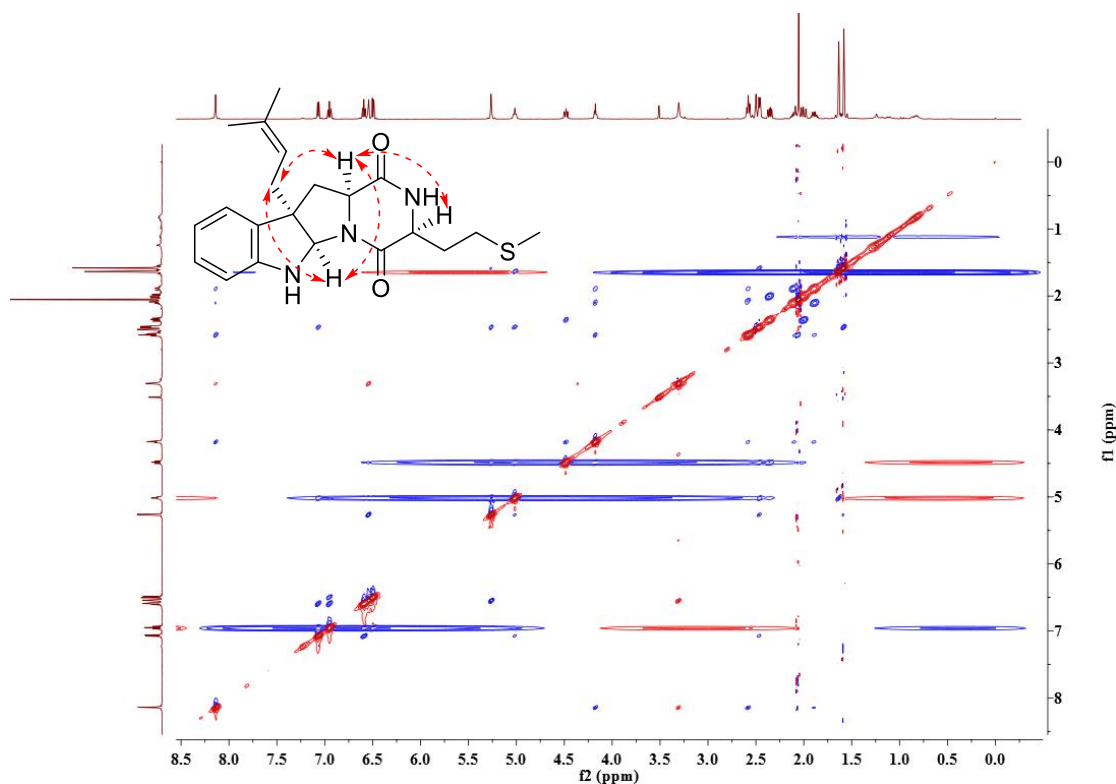


**Figure S26.** COSY spectrum of compound **9** in  $\text{DMSO}-d_6$ .

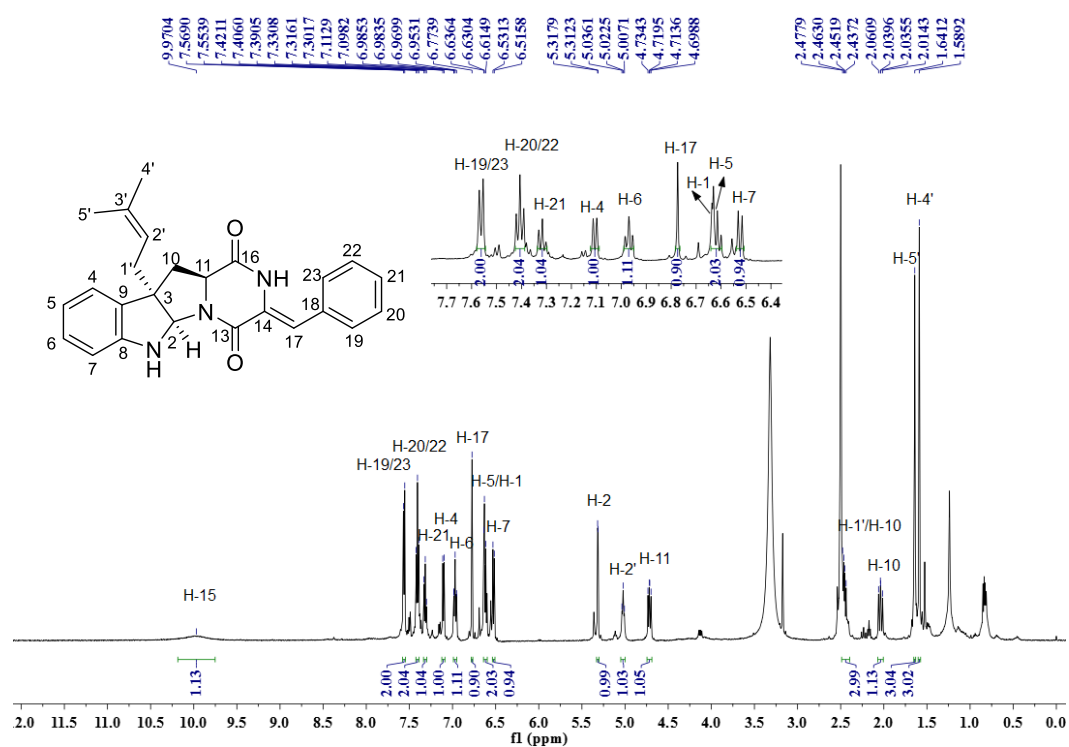


**Figure S27.** HMBC spectrum of compound **9** in  $\text{DMSO}-d_6$ .



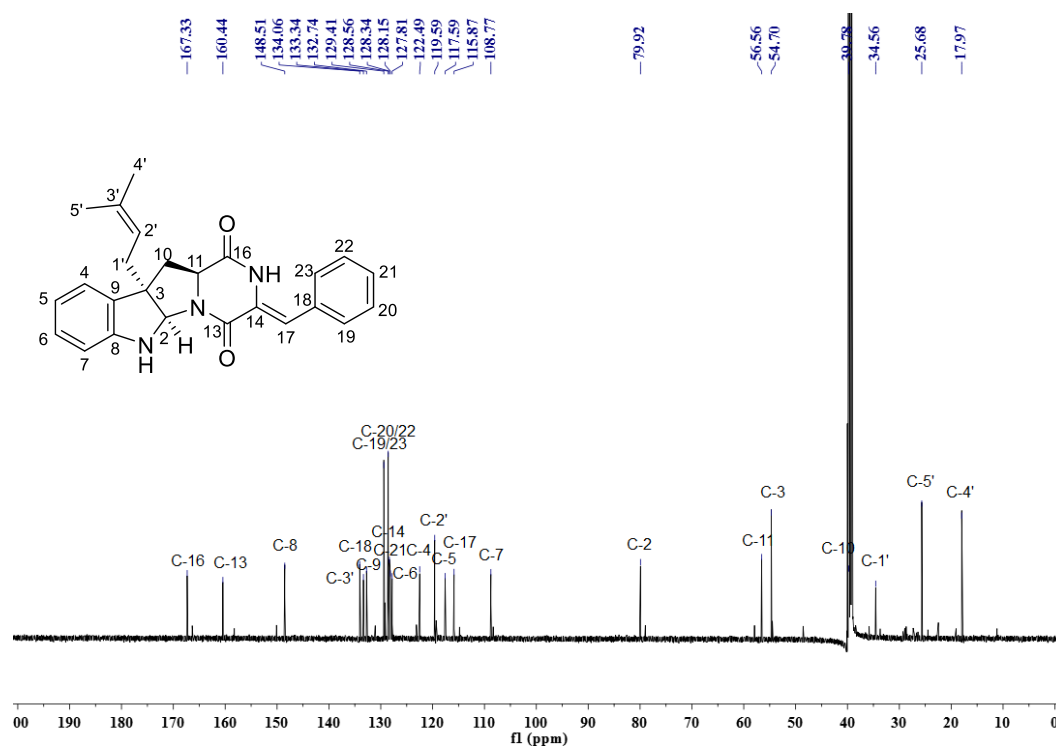


**Figure S28.** NOESY spectrum of compound **9** in DMSO- $d_6$

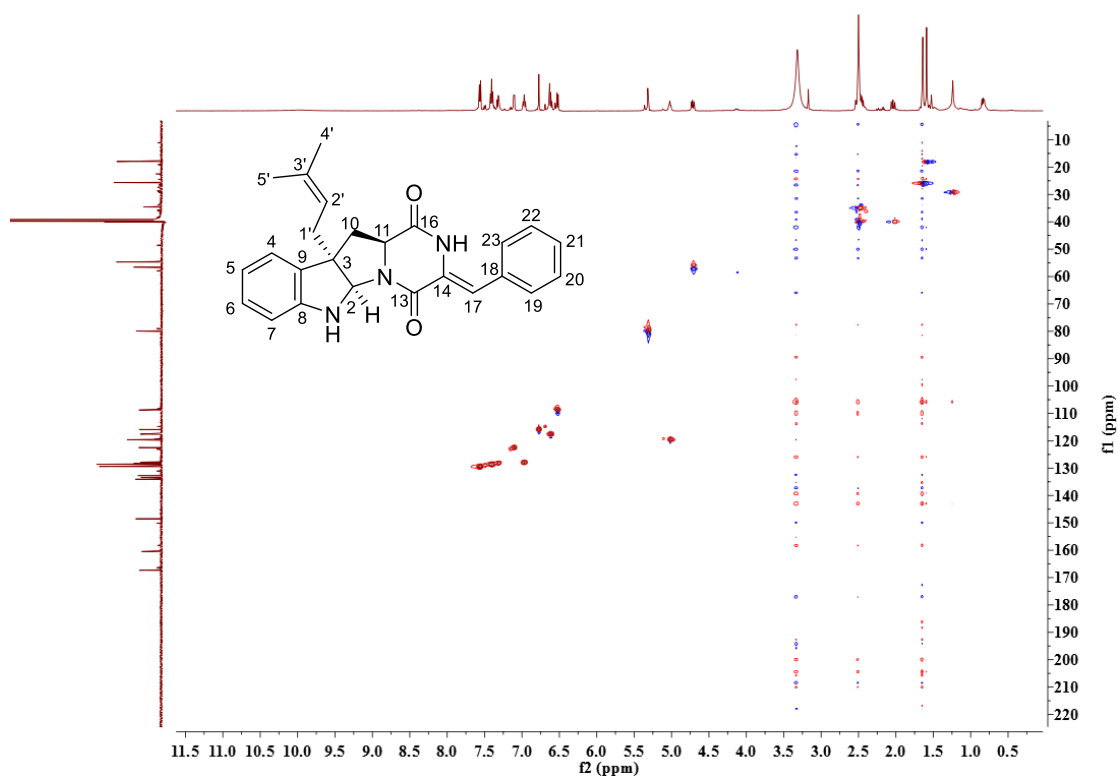


**Figure S29.**  $^1\text{H}$  NMR spectrum of compound **10** in DMSO- $d_6$  (500 MHz)

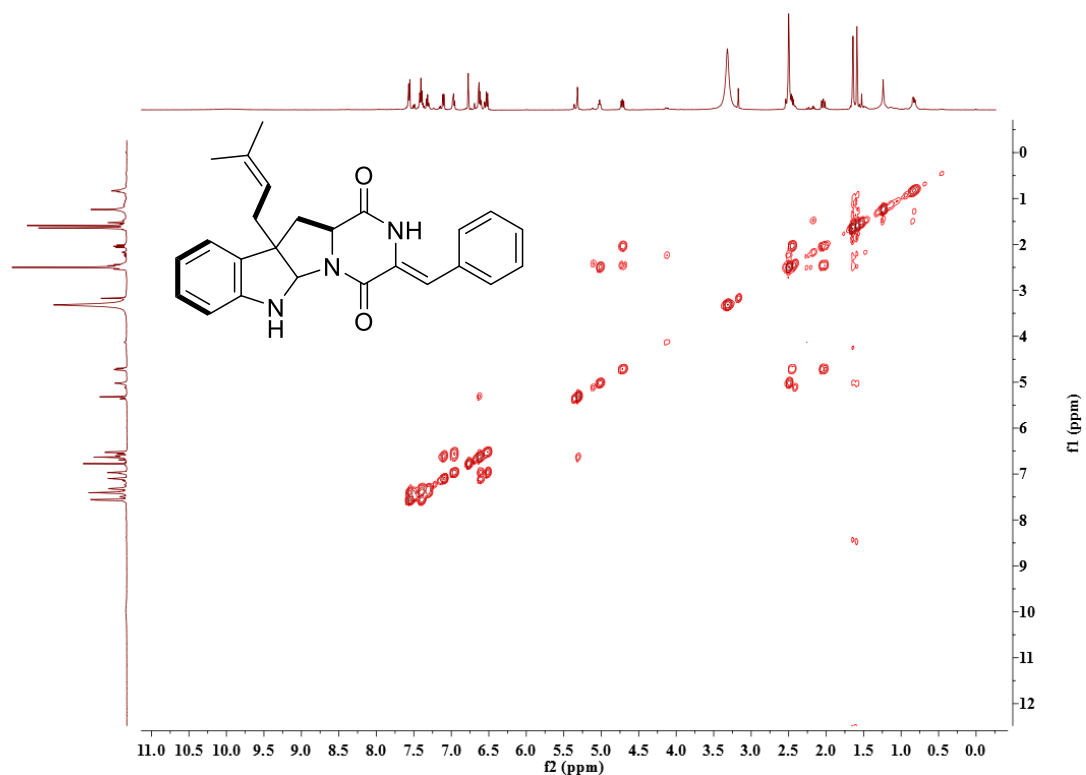
# PUBLICATIONS



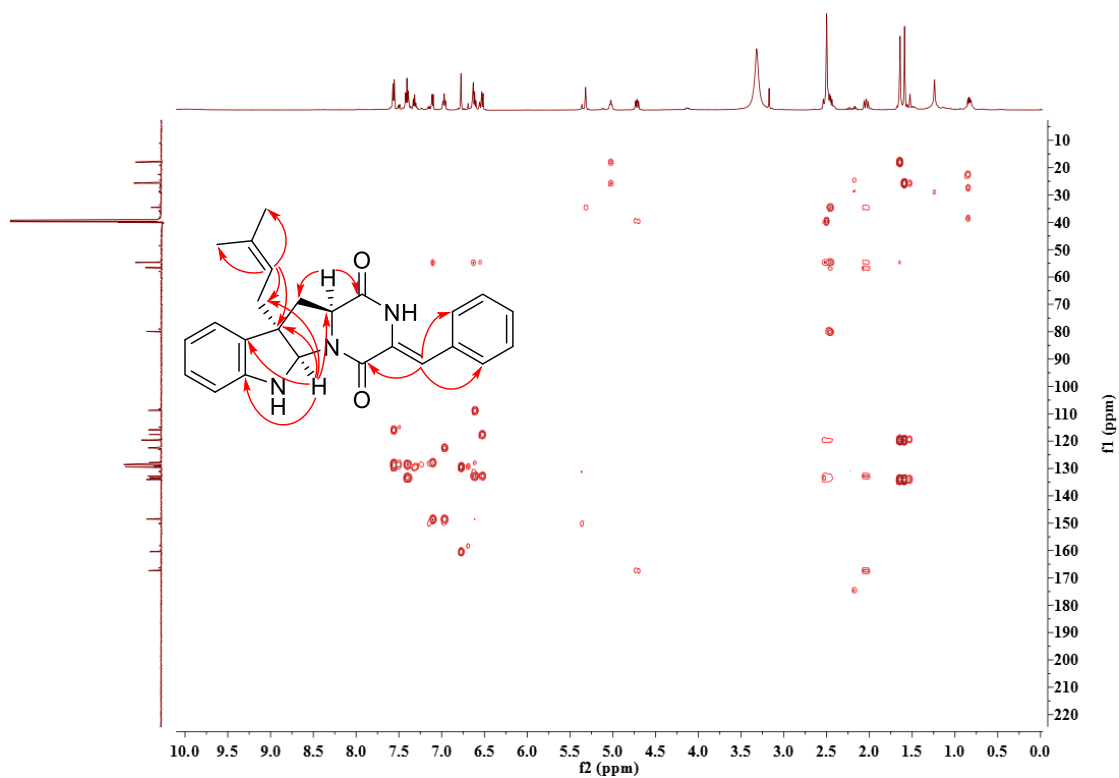
**Figure S30.**  $^{13}\text{C}$  NMR spectrum of compound **10** in  $\text{DMSO}-d_6$  (125 MHz)



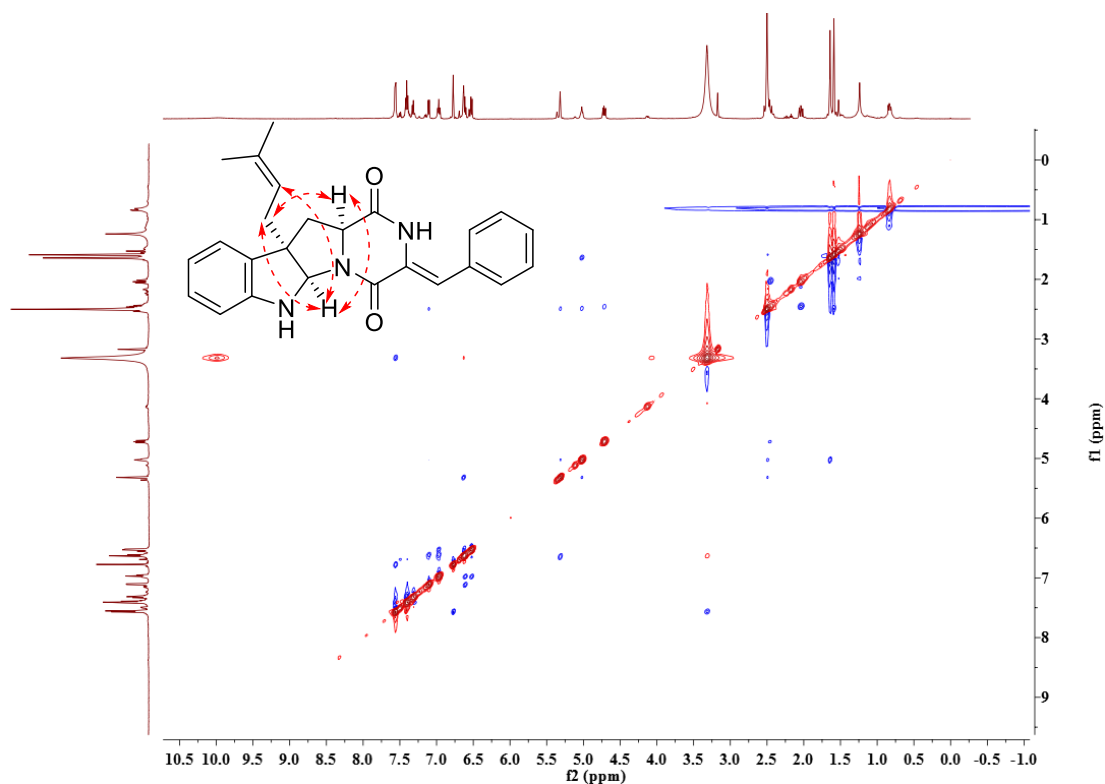
**Figure S31.** HSQC spectrum of compound **10** in  $\text{DMSO}-d_6$



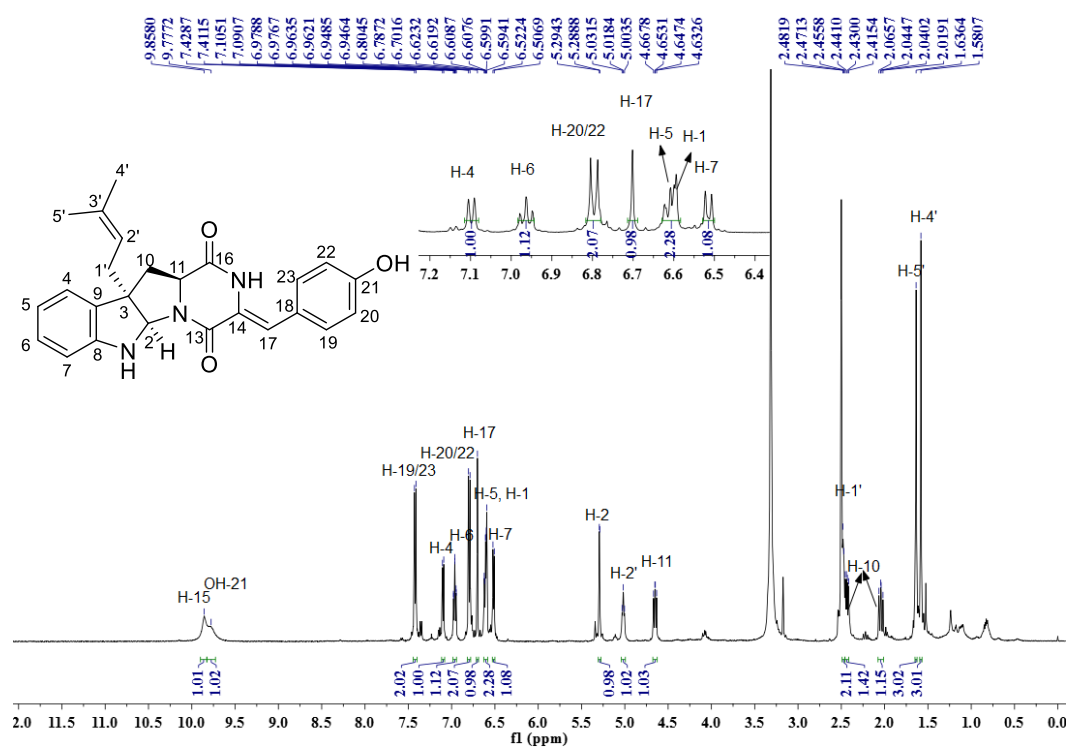
**Figure S32.** COSY spectrum of compound **10** in DMSO- $d_6$



**Figure S33.** HMBC spectrum of compound **10** in DMSO- $d_6$

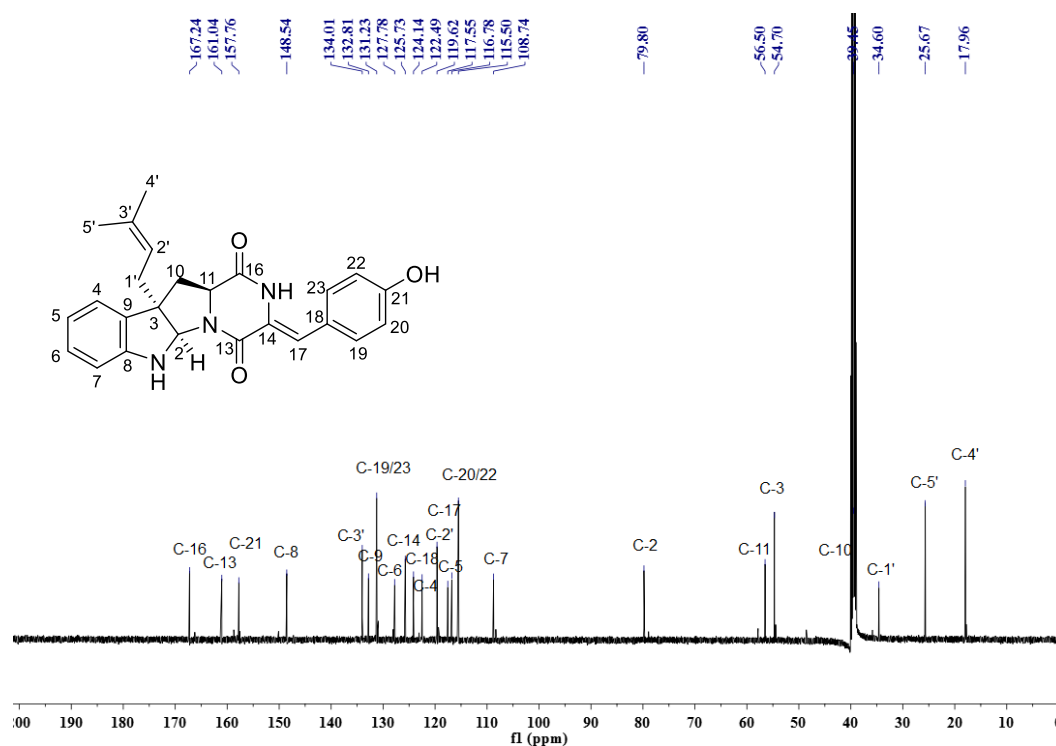


**Figure S34.** NOESY spectrum of compound **10** in DMSO- $d_6$

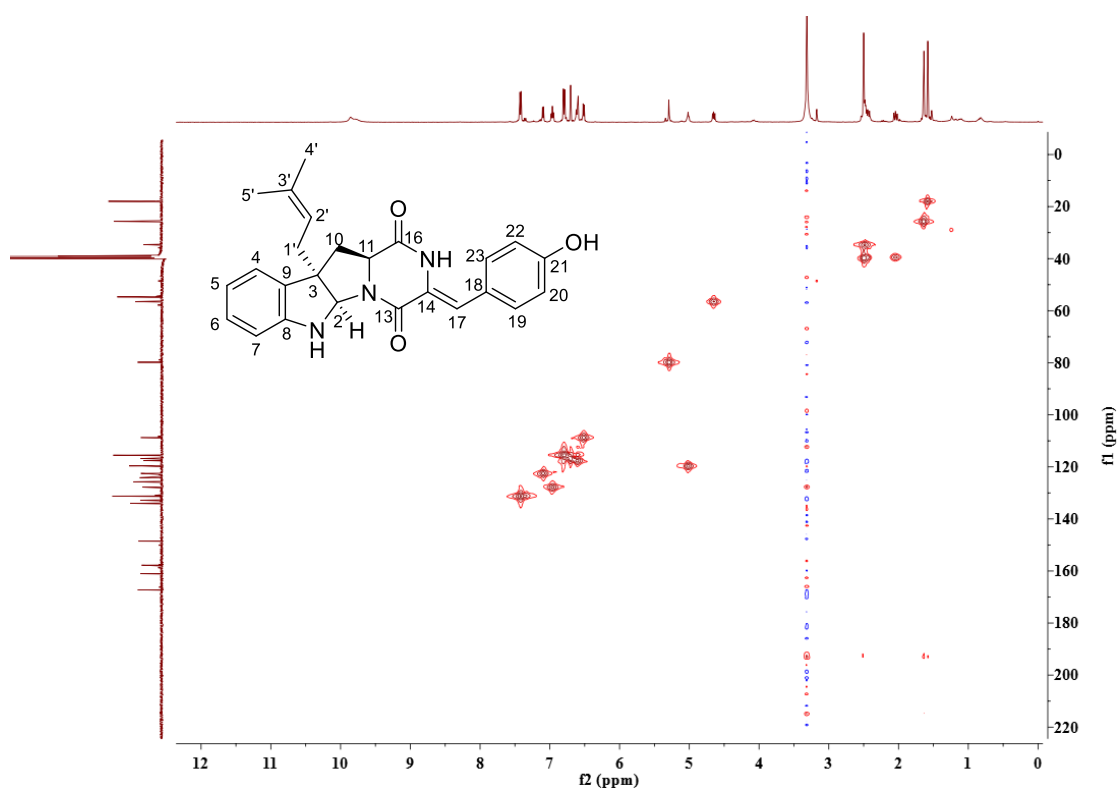


**Figure S35.**  $^1\text{H}$  NMR spectrum of compound **11** in DMSO- $d_6$  (500 MHz)

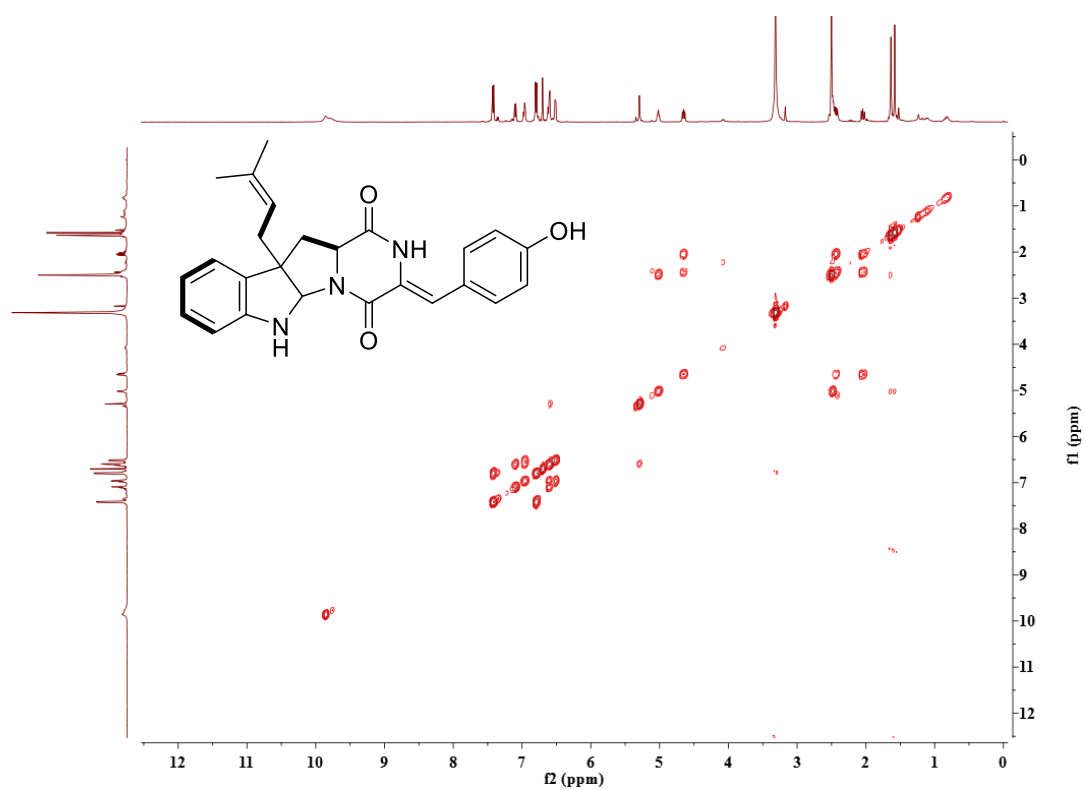
# PUBLICATIONS



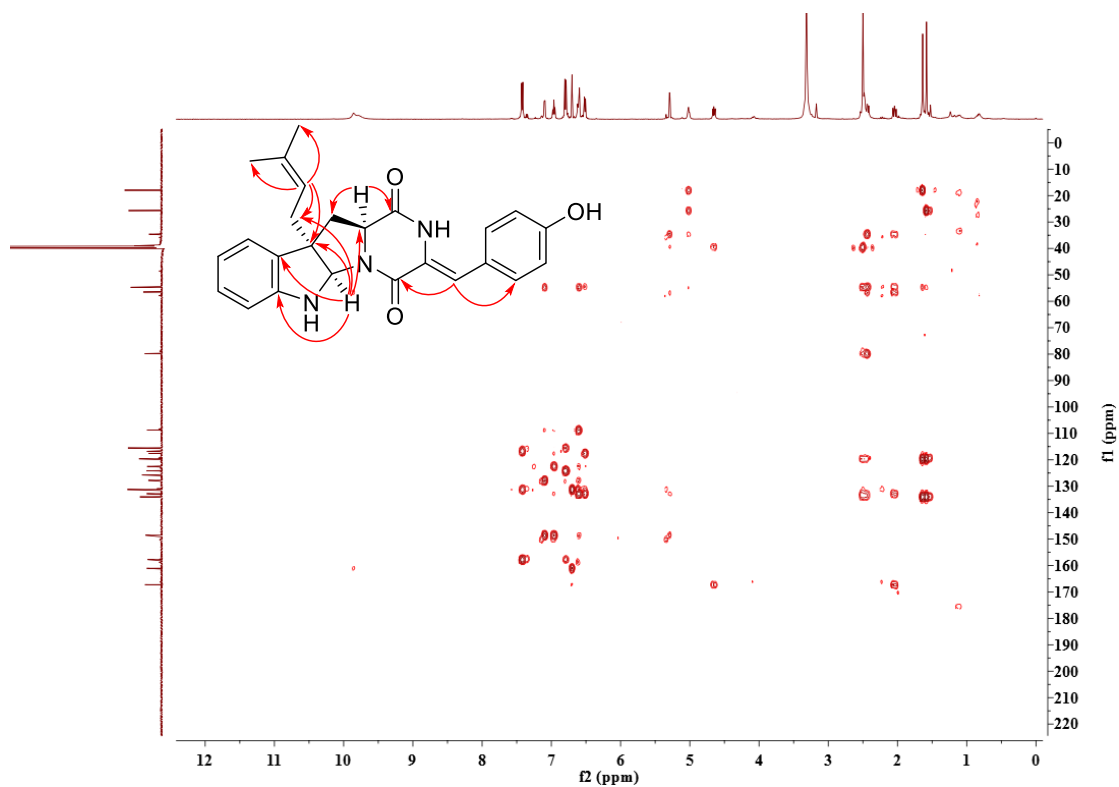
**Figure S36.**  $^{13}\text{C}$  NMR spectrum of compound 11 in  $\text{DMSO}-d_6$  (125 MHz)



**Figure S37.** HSQC spectrum of compound 11 in  $\text{DMSO}-d_6$



**Figure S38.** COSY spectrum of compound **11** in  $\text{DMSO}-d_6$



**Figure S39.** HMBC spectrum of compound **11** in  $\text{DMSO}-d_6$

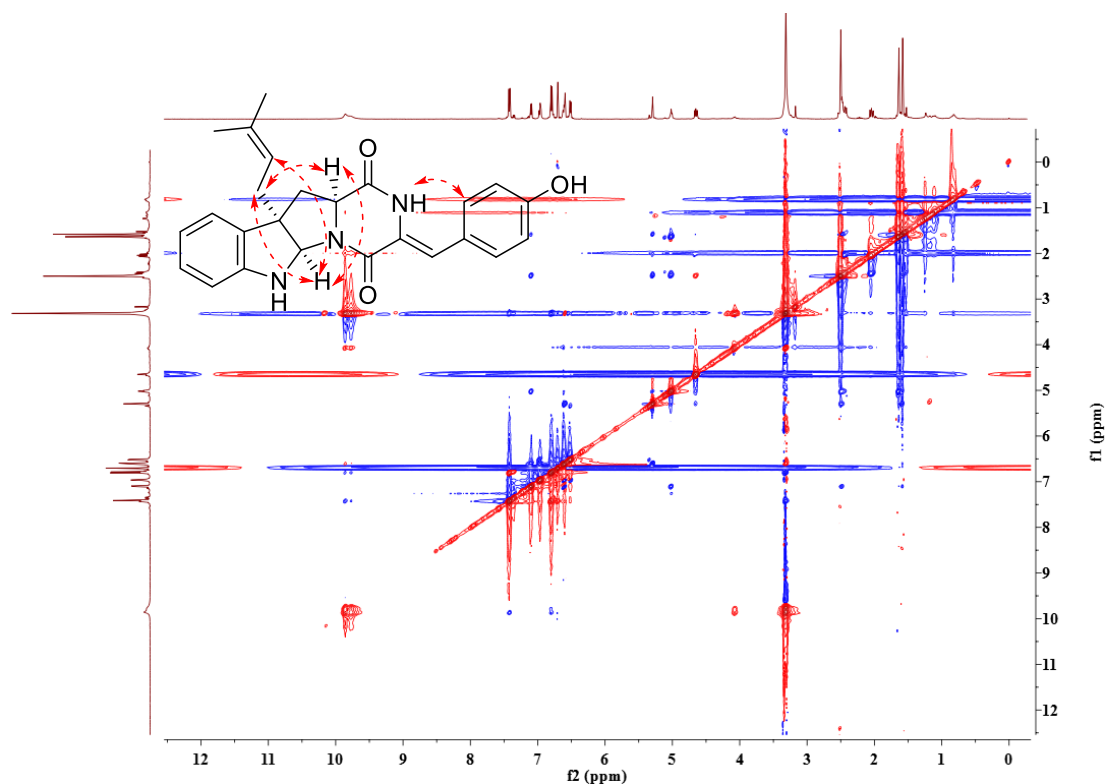


Figure S40. NOESY spectrum of compound **11** in DMSO- $d_6$

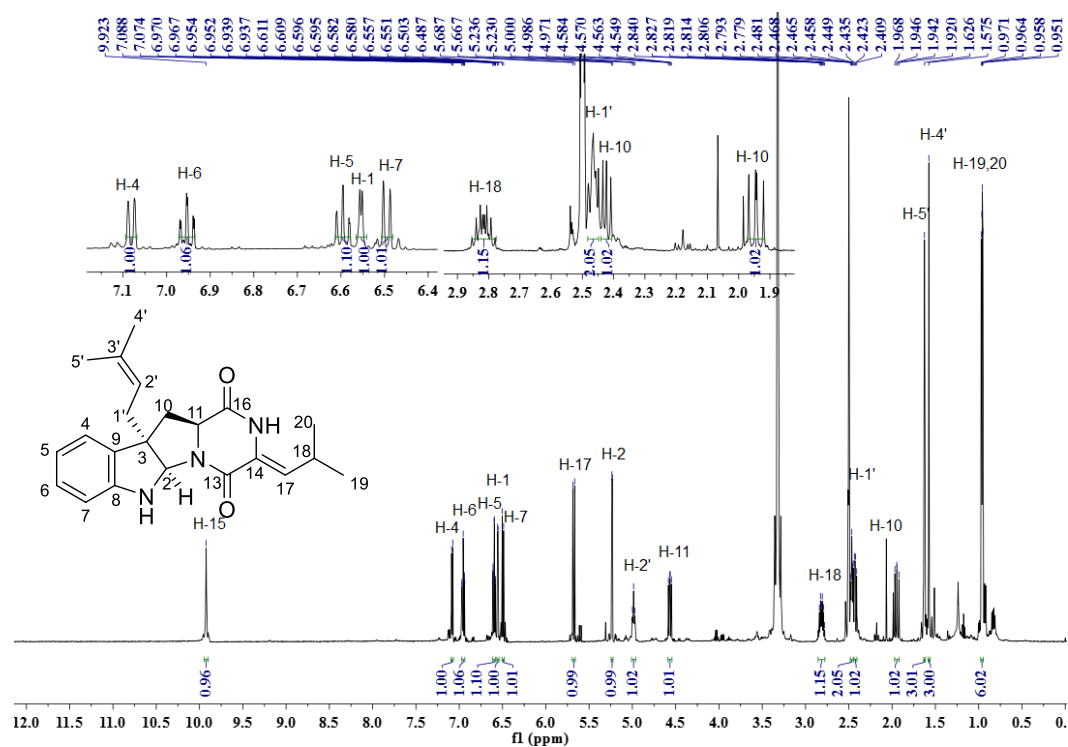
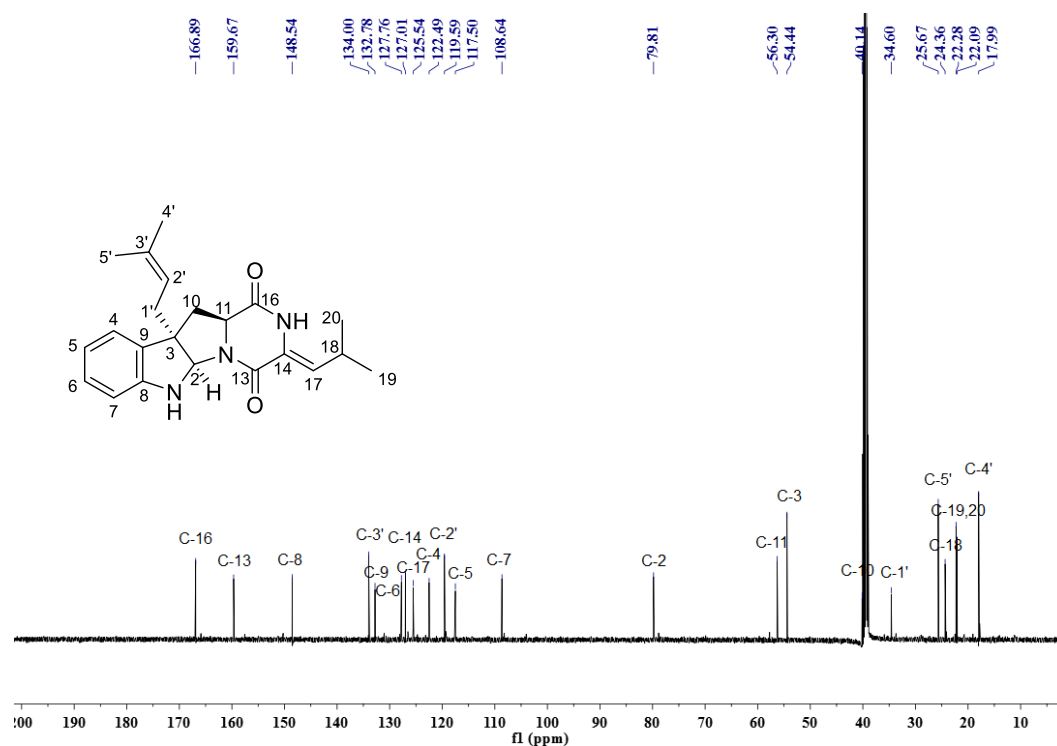
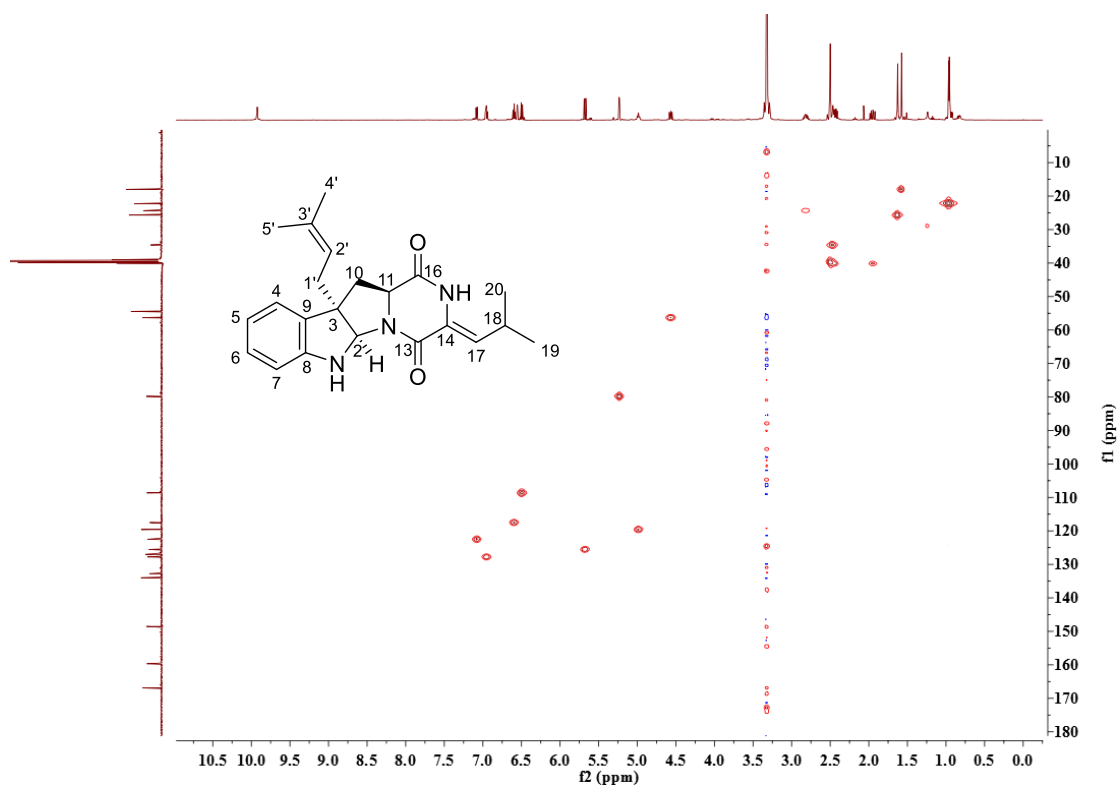


Figure S41.  $^1\text{H}$  NMR spectrum of compound **12** in DMSO- $d_6$  (500 MHz)

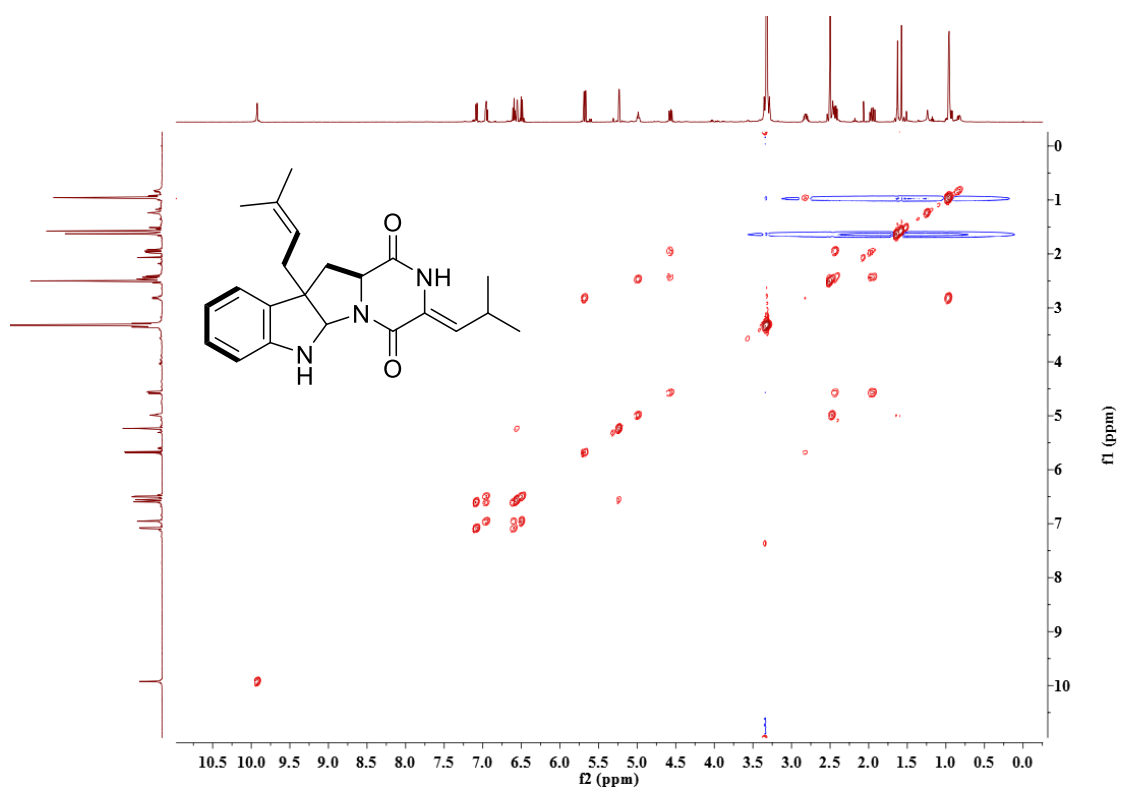


**Figure S42.**  $^{13}\text{C}$  NMR spectrum of compound **12** in  $\text{DMSO}-d_6$  (125 MHz).

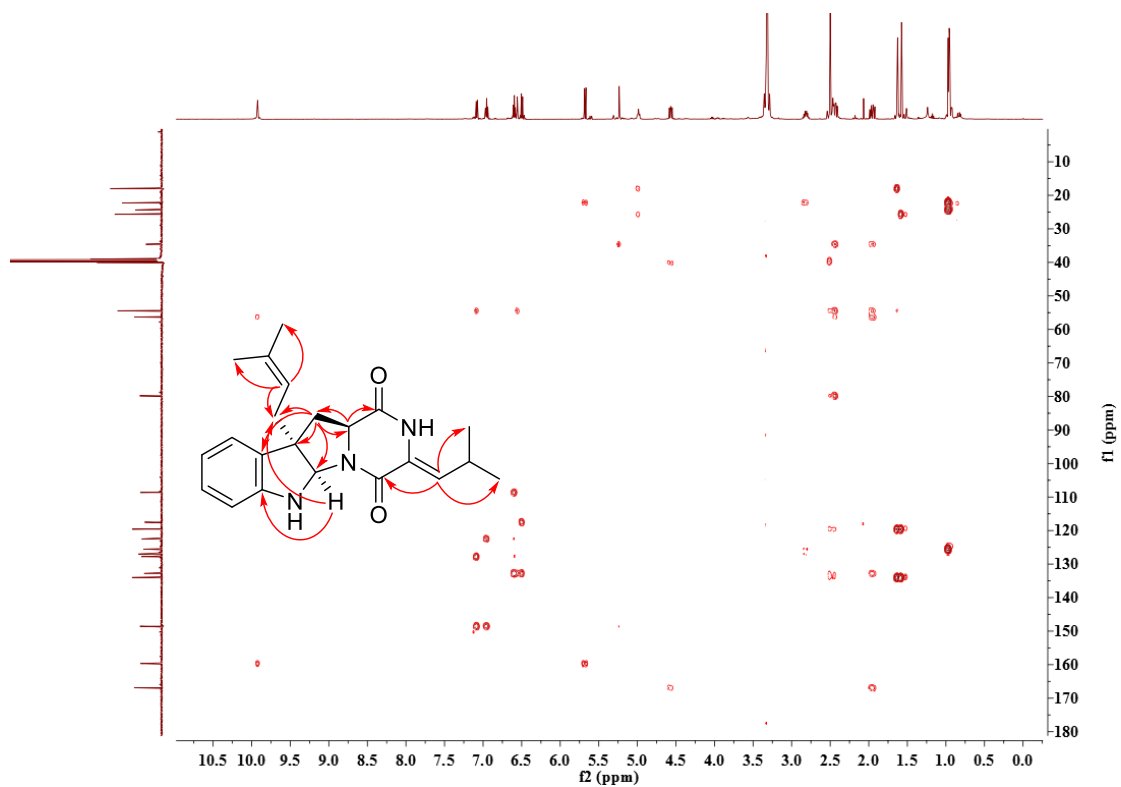


**Figure S43.** HSQC spectrum of compound **12** in  $\text{DMSO}-d_6$





**Figure S44.** COSY spectrum of compound **12** in DMSO-*d*<sub>6</sub>



**Figure S45.** HMBC spectrum of compound **12** in DMSO-*d*<sub>6</sub>

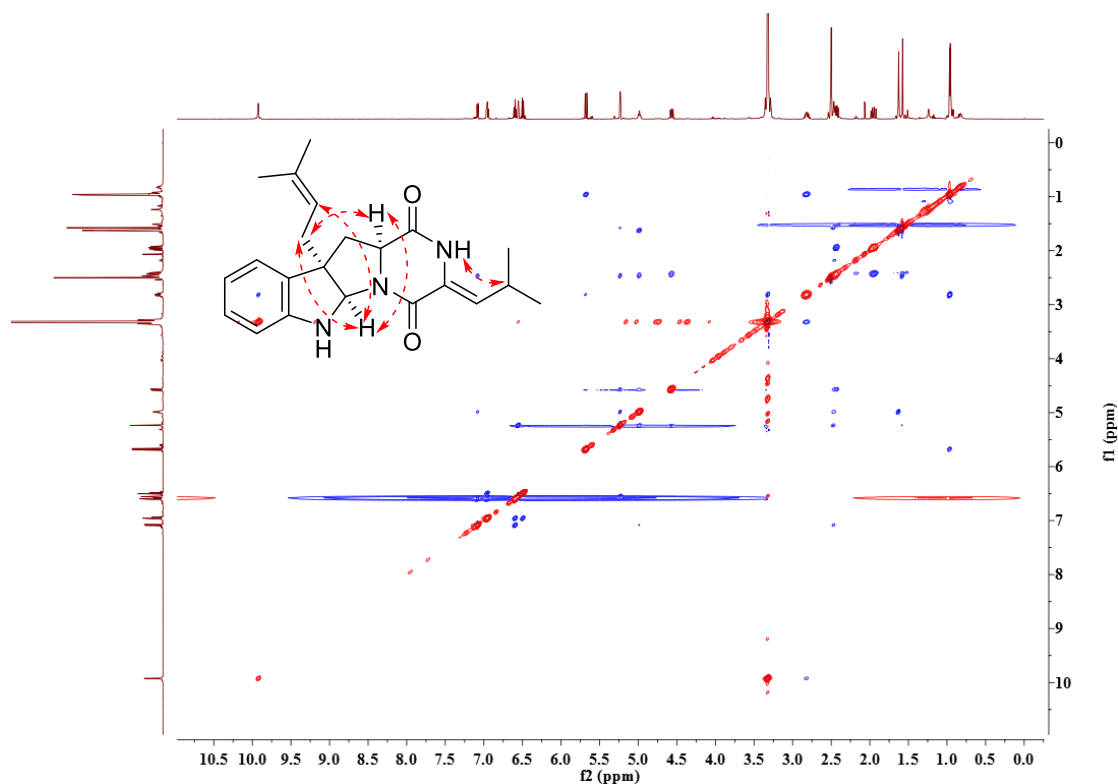


Figure S46. NOESY spectrum of compound **12** in DMSO- $d_6$

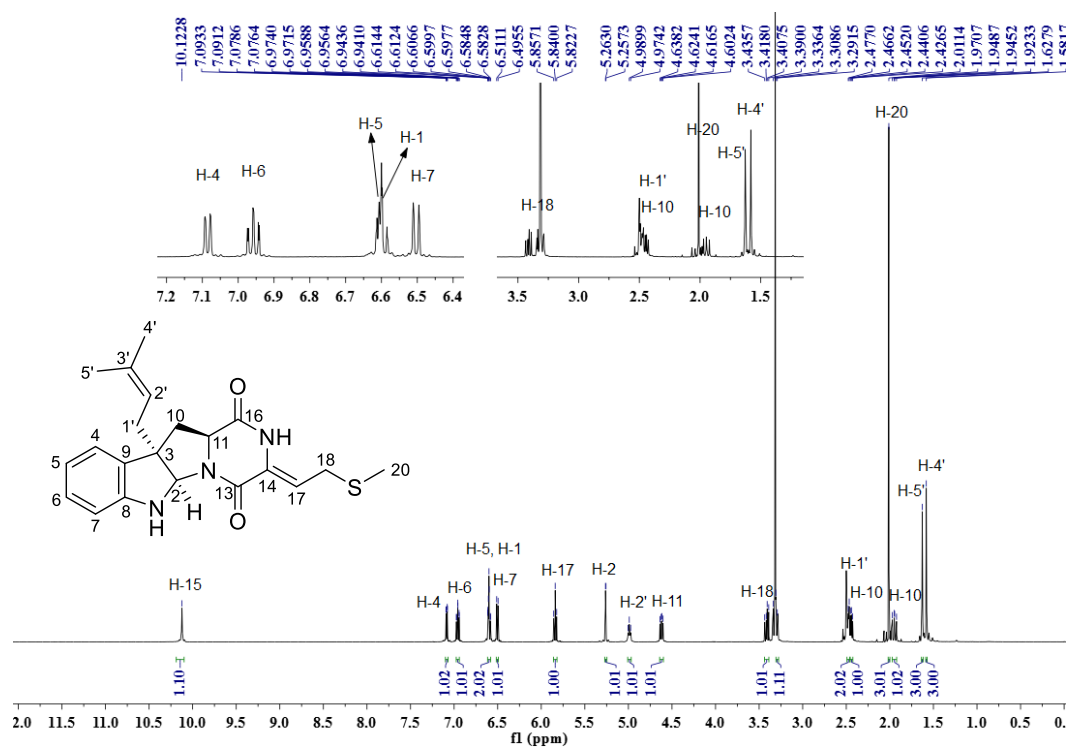
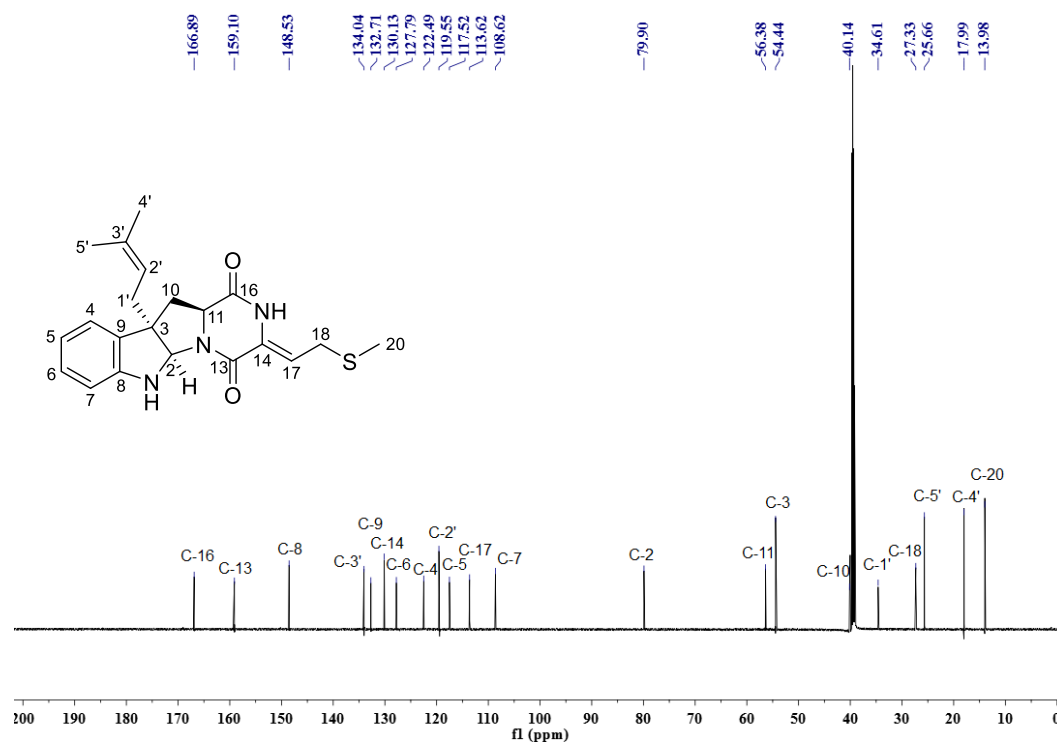
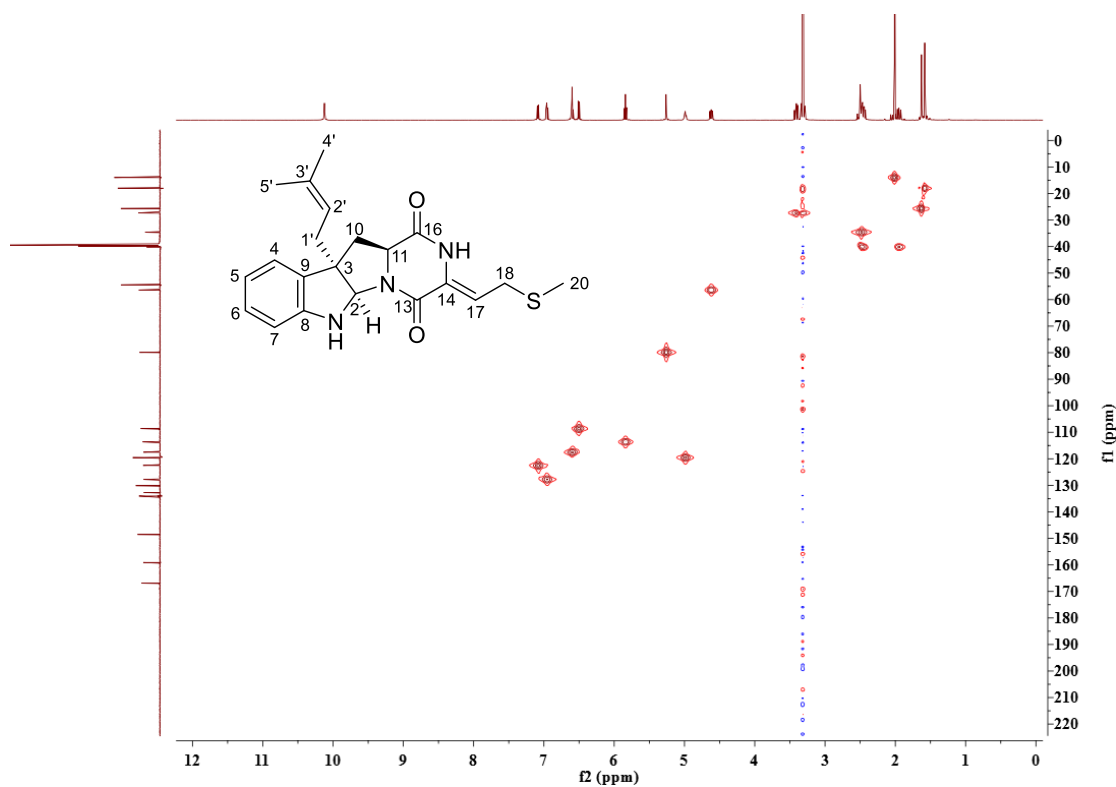


Figure S47.  $^1\text{H}$  NMR spectrum of compound **13** in DMSO- $d_6$  (500 MHz)

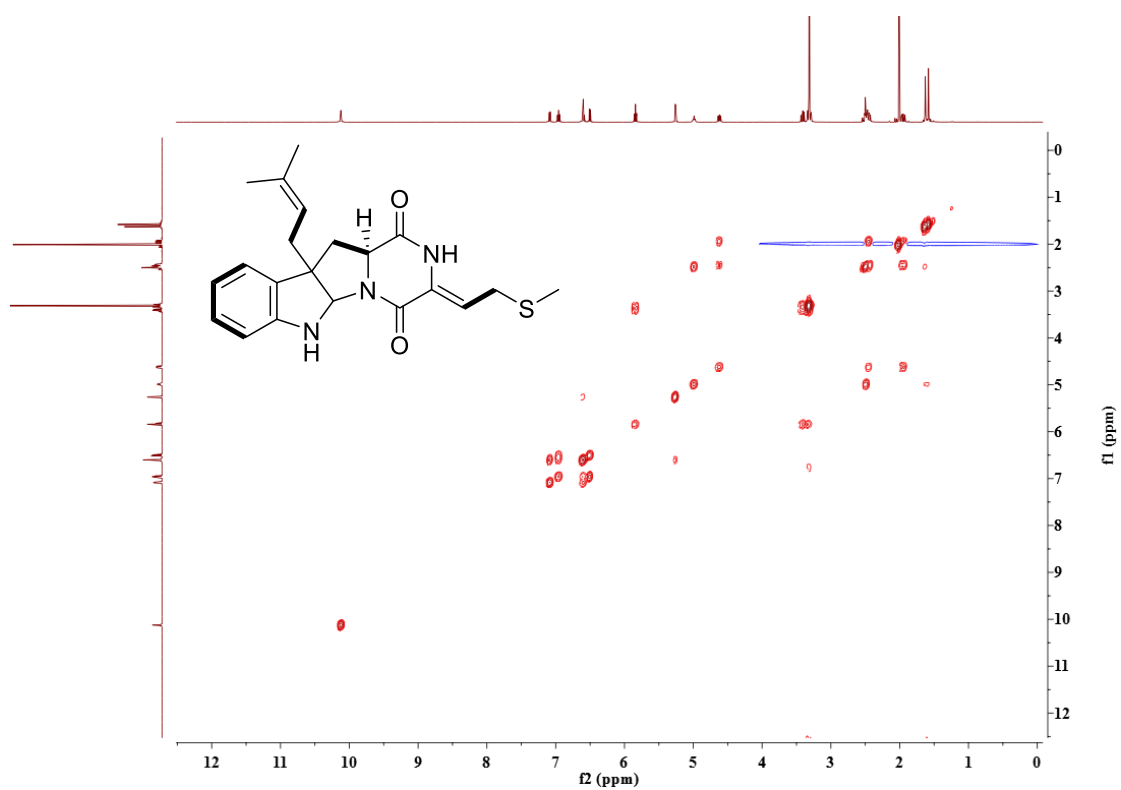
# PUBLICATIONS



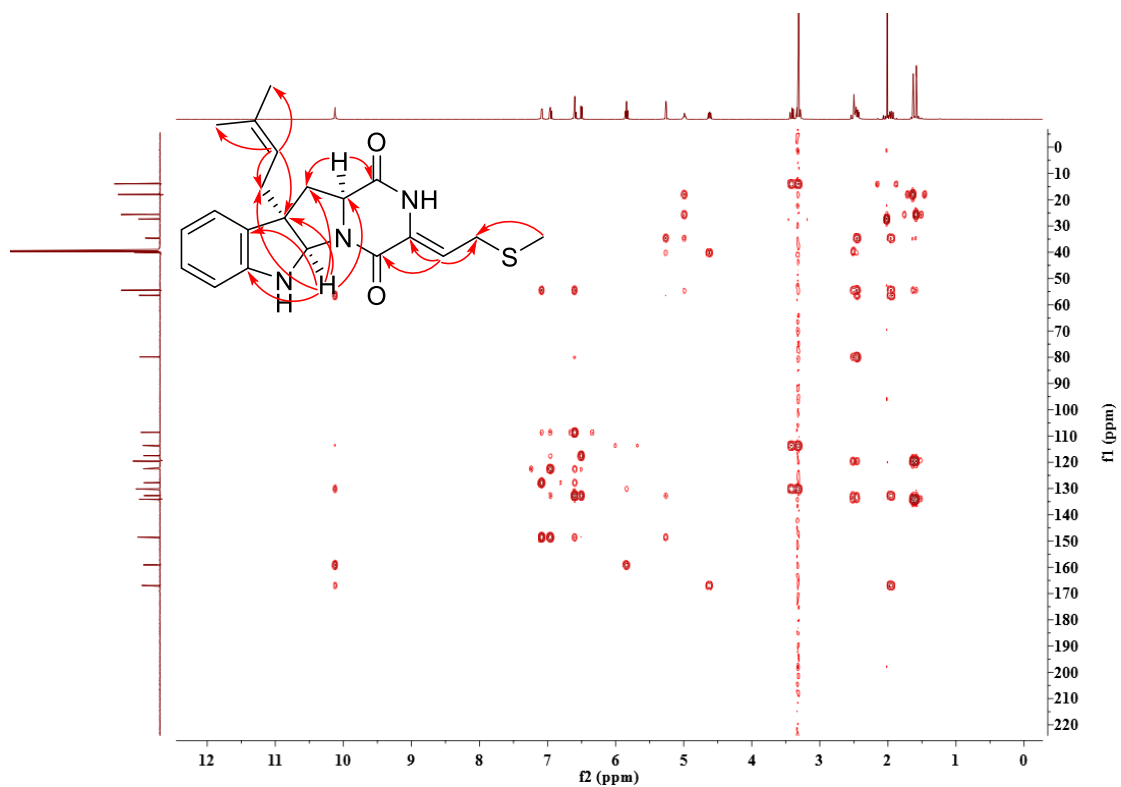
**Figure S48.**  $^{13}\text{C}$  NMR spectrum of compound **13** in DMSO- $d_6$  (125 MHz)



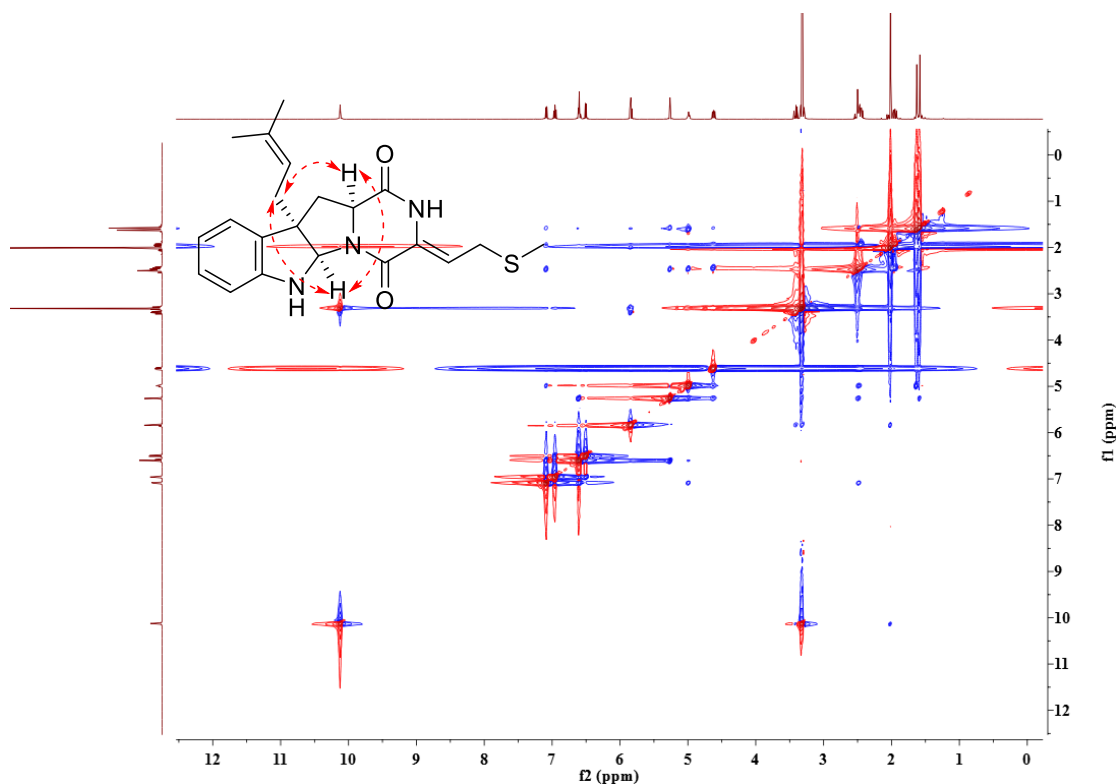
**Figure S49.** HSQC spectrum of compound **13** in DMSO- $d_6$



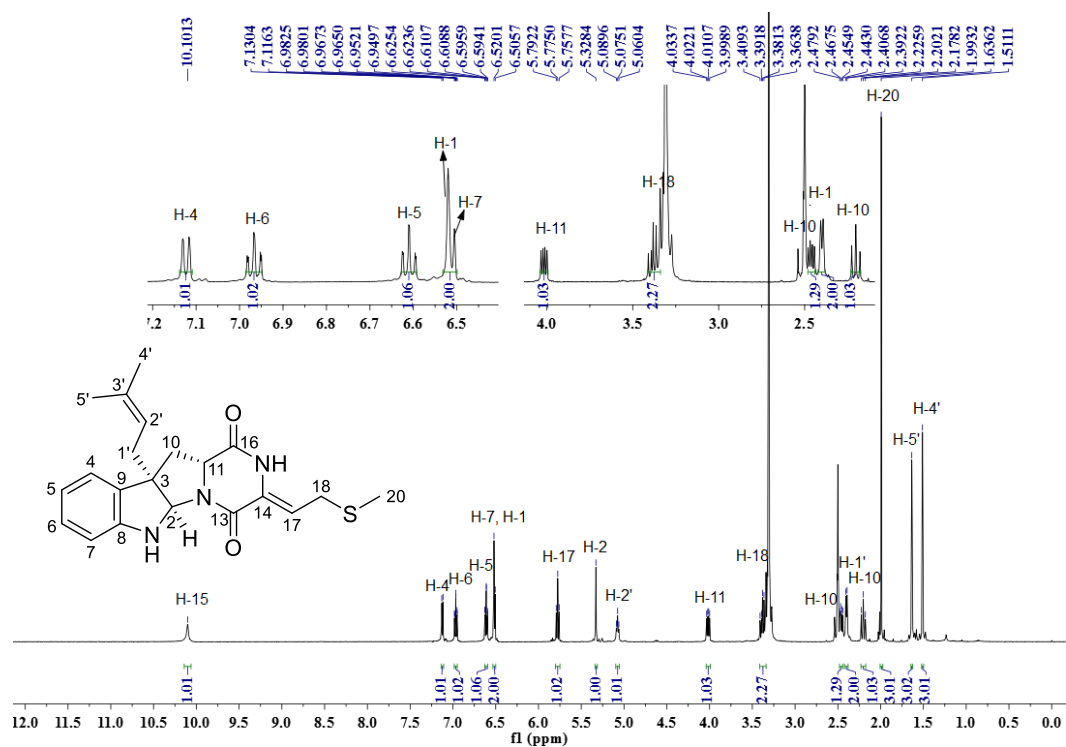
**Figure S50.** COSY spectrum of compound **13** in  $\text{DMSO}-d_6$



**Figure S51.** HMBC spectrum of compound **13** in  $\text{DMSO}-d_6$

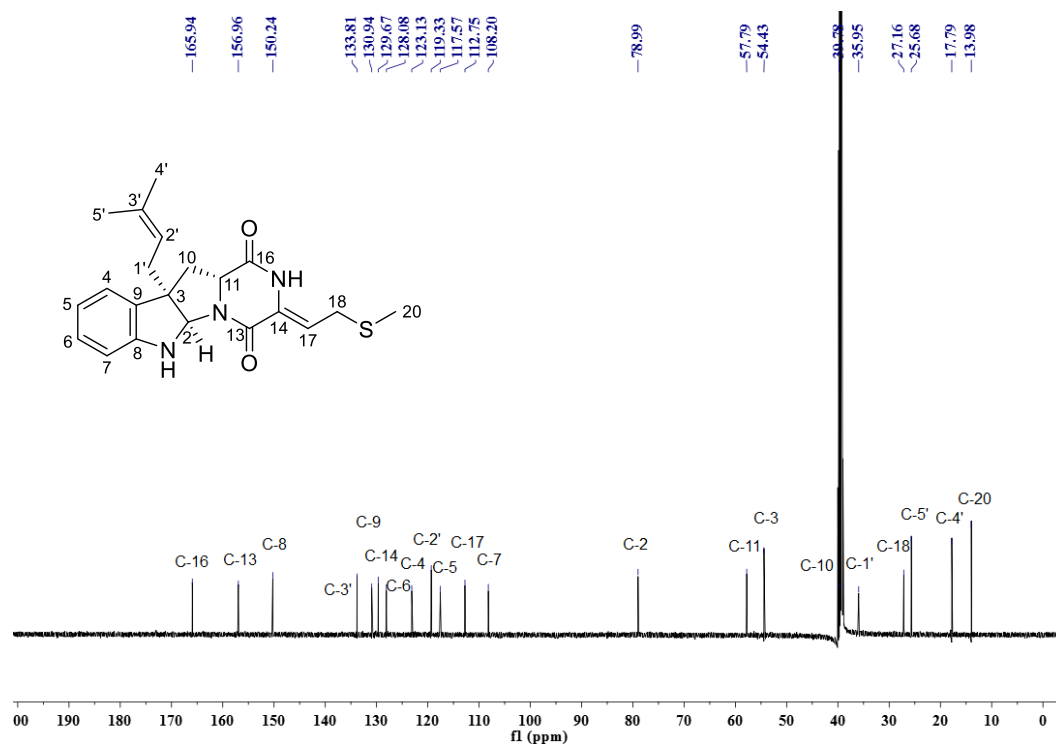


**Figure S52.** NOESY spectrum of compound **13** in DMSO- $d_6$ .

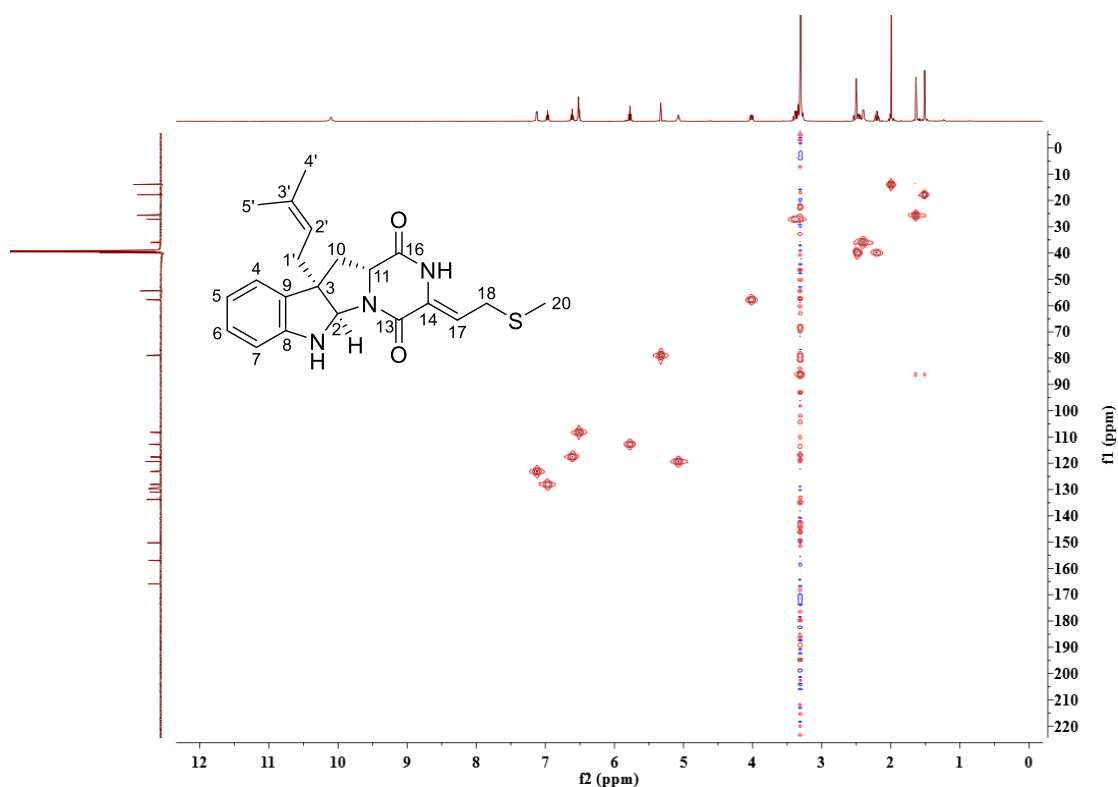


**Figure S53.**  $^1\text{H}$  NMR spectrum of compound **14** in DMSO- $d_6$  (500 MHz)

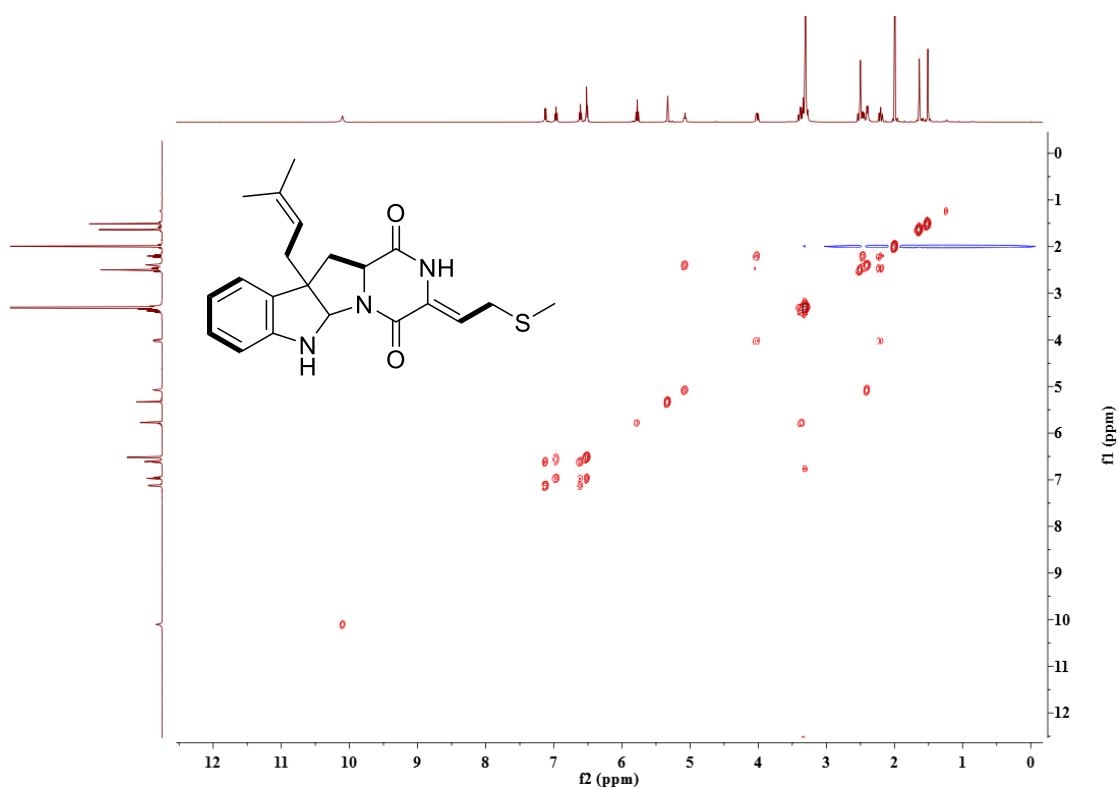
# PUBLICATIONS



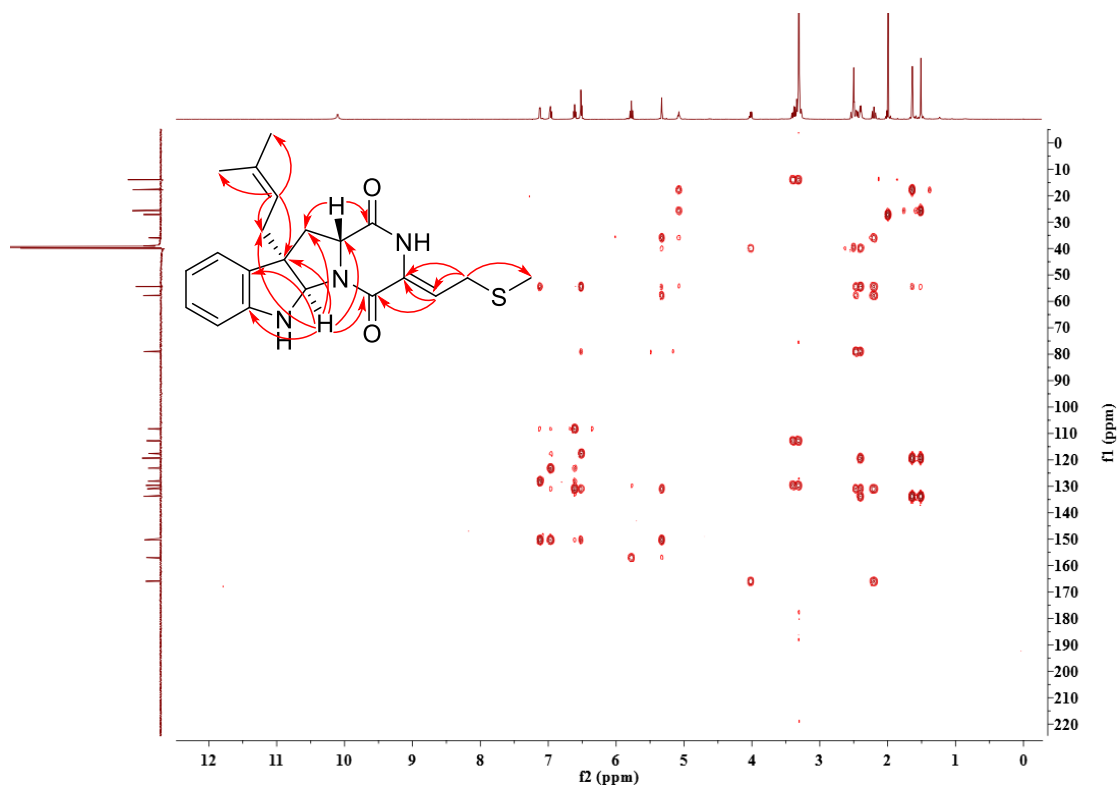
**Figure S54.**  $^{13}\text{C}$  NMR spectrum of compound **14** in  $\text{DMSO}-d_6$  (125 MHz)



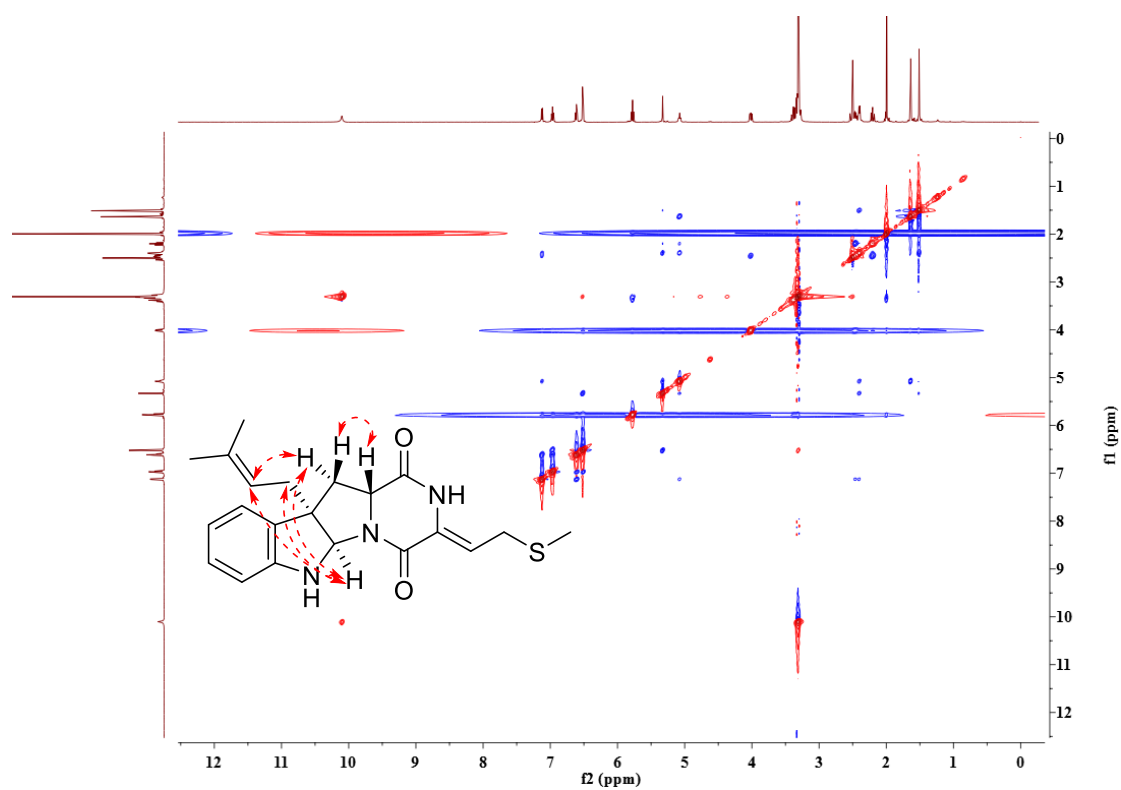
**Figure S55.** HSQC spectrum of compound **14** in  $\text{DMSO}-d_6$



**Figure S56.** COSY spectrum of compound **14** in  $\text{DMSO}-d_6$

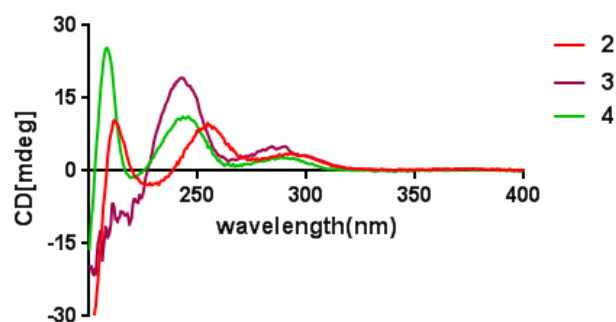


**Figure S57.** HMBC spectrum of compound **14** in  $\text{DMSO}-d_6$

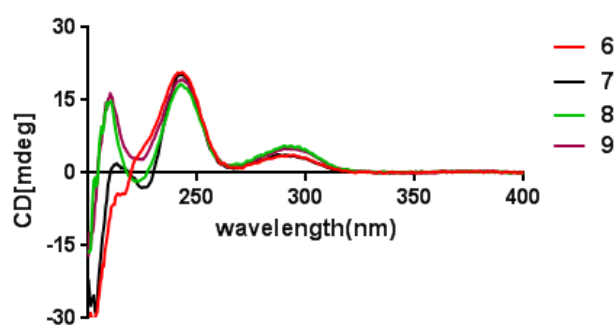


**Figure S58.** NOESY spectrum of compound **14** in DMSO- $d_6$

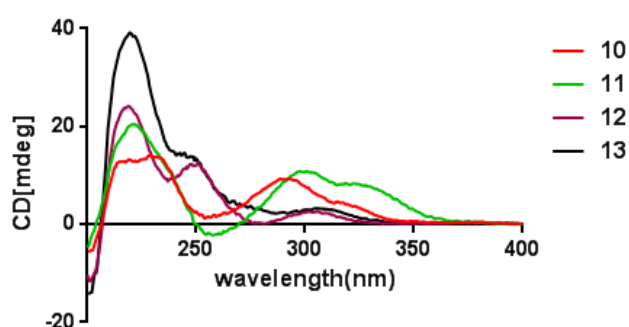




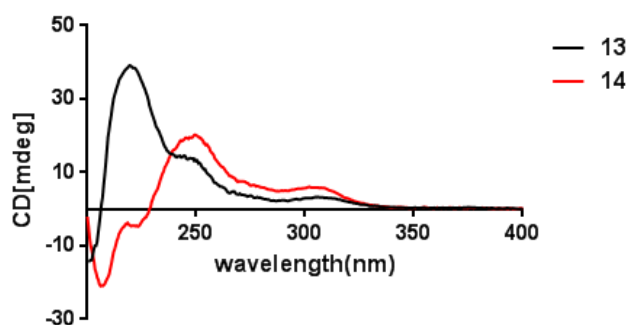
**Figure S59.** Experimental ECD spectra of **2** — **4** in MeOH



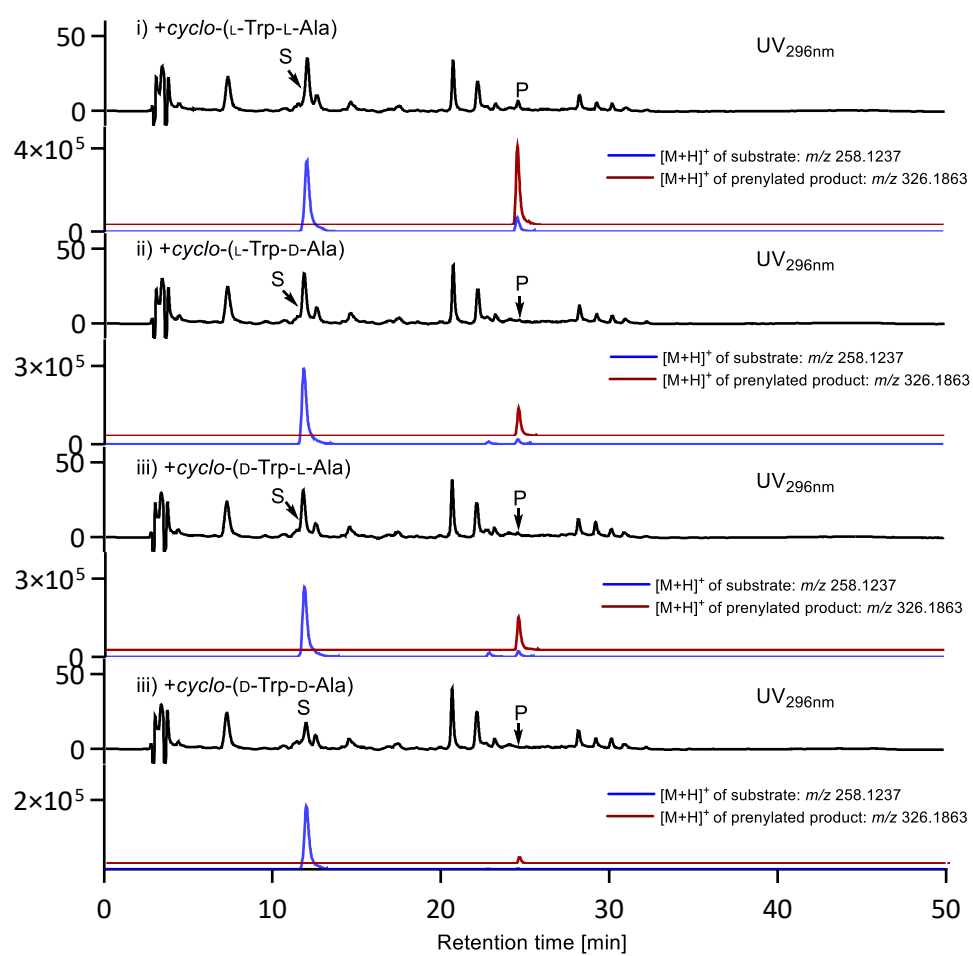
**Figure S60.** Experimental ECD spectra of **6** — **9** in MeOH



**Figure S61.** Experimental ECD spectra of **10** — **13** in MeOH

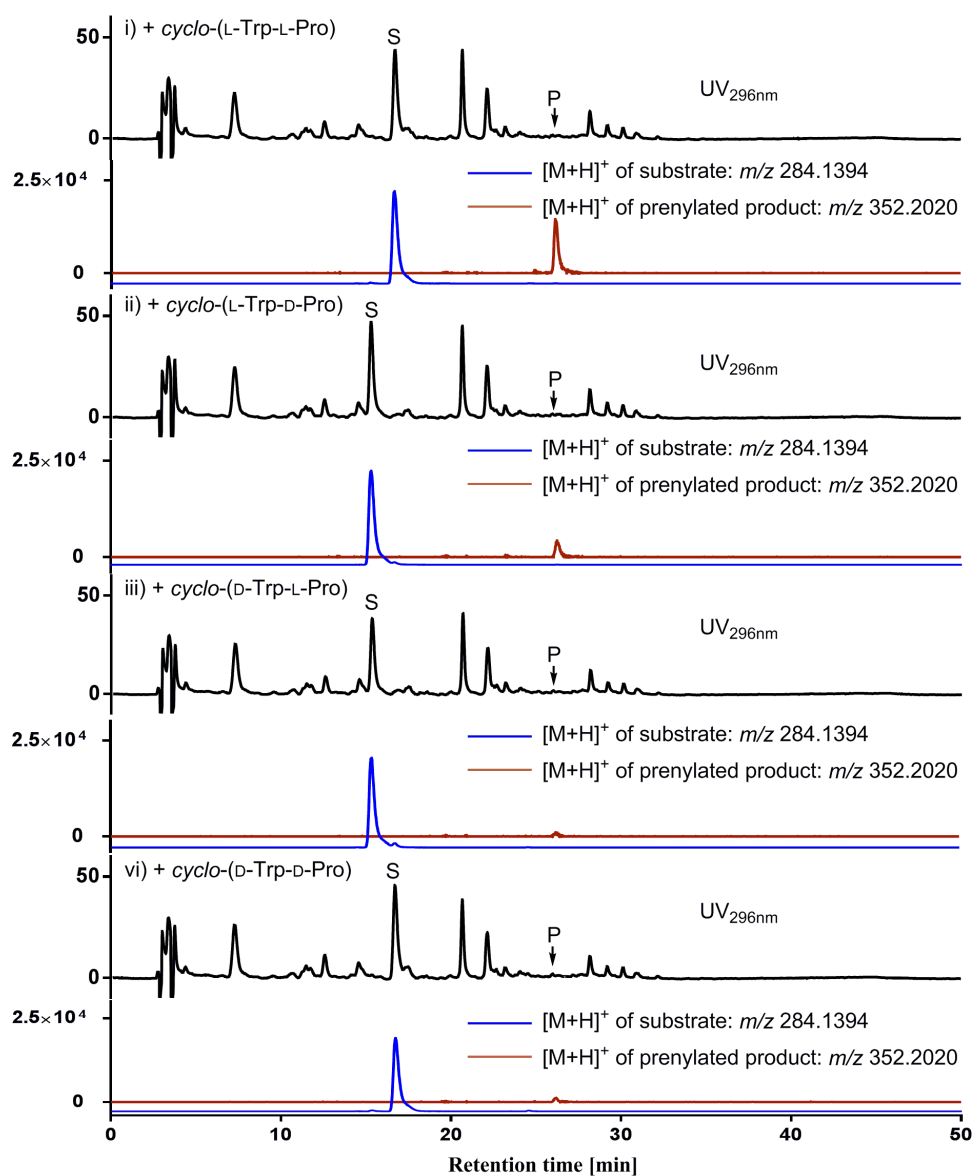


**Figure S62.** Experimental ECD spectra of **13** — **14** in MeOH.



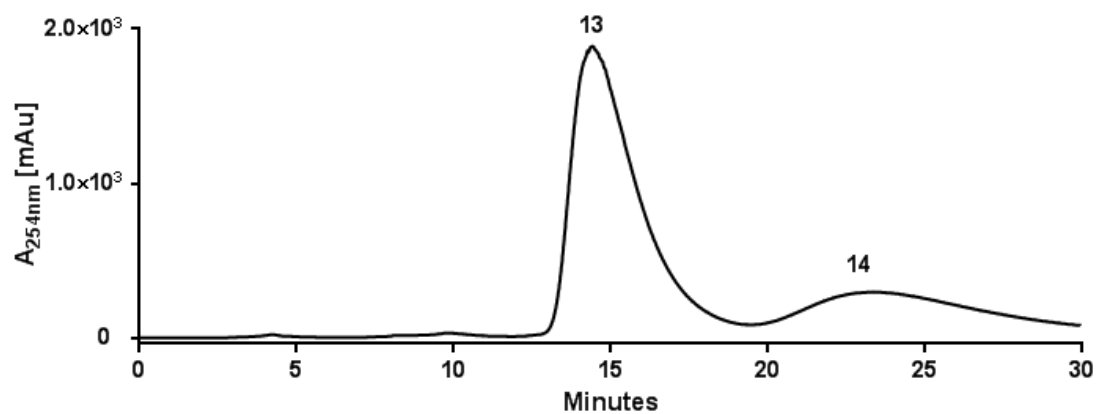
**Figure S63.** LC-MS analysis of *sasB* transformant after incubating with *cyclo*-Trp-Ala isomers

S: substrate, P: product



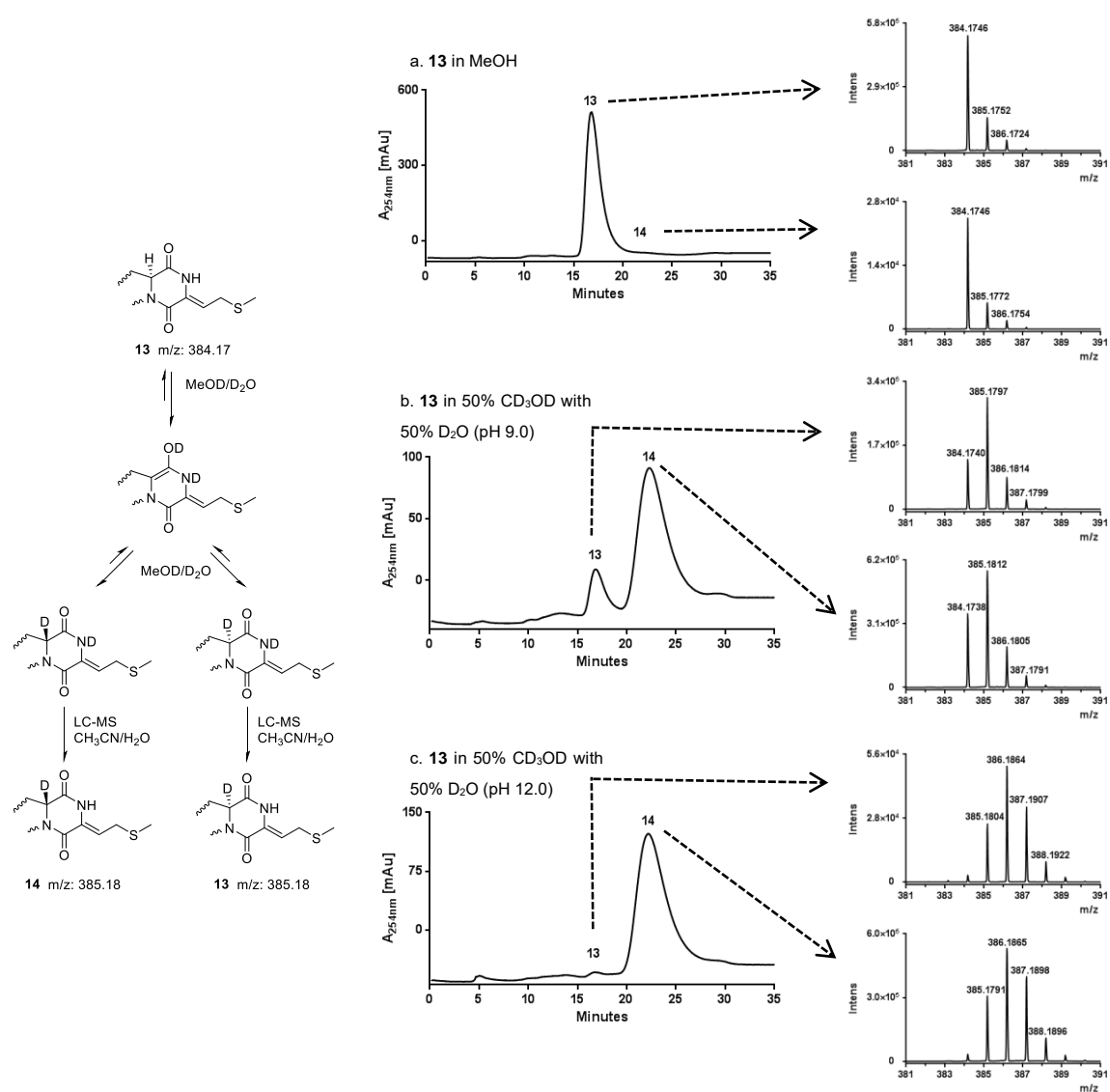
**Figure S64.** LC-MS analysis of *sasB* transformant after incubating with *cyclo*-Trp-Pro isomers

S: substrate, P: product



**Figure S65.** Purification of **13** and **14** on the Multohigh Chiral AM-RP column.

# PUBLICATIONS



**Figure S66.** LC-MS analysis of **13** after incubation in CD<sub>3</sub>OD/D<sub>2</sub>O (1:1) for 14 h.

**Reference**

- (1) Zhang, Y.; Yao, T.; Jiang, Y.; Li, H.; Yuan, W.; Li, W. *Appl. Environ. Microbiol.* **2021**, *87*, e02525-20.
- (2) Yao, T.; Liu, J.; Liu, Z.; Li, T.; Li, H.; Che, Q.; Zhu, T.; Li, D.; Gu, Q.; Li, W. *Nat. Commun.* **2018**, *9*, 4091.
- (3) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. *Proc. Natl. Acad. Sci. U. S. A* **2003**, *100*, 1541-1546.
- (4) Zaburannyi, N.; Rabyk, M.; Ostash, B.; Fedorenko, V.; Luzhetskyy, A. *Bmc Genomics* **2014**, *15*, 97.

## 5. Conclusion and future perspectives

In the course of this thesis, several novel CDP derivatives from different *Streptomyces* species were discovered and their biosynthetic pathways as well as that of previously characterized streptoazine C were elucidated in detail, using a combinatorial approach of genome mining, heterologous expression in *S. albus*, and biochemical investigation. All three projects are based on the knowledge of preliminary work, carried out by my colleagues Dr. Huili Yu and Dr. Jing Liu. This knowledge, especially the gene sequences, served as starting point to discover the BGCs investigated in this thesis. Apart from the obvious characterization of the biosynthesis of the secondary metabolites guatrypmethine C, guatyromycine A and B, and streptoazine C, each project had its own highlights.

The Fe<sup>II</sup>/2-OG oxidase GtmE extends the set of tailoring enzymes in CDPS-dependent pathways for dehydrogenation at the DKP ring, recently limited to CDOs, and enriches the repertoire of this ubiquitously occurring enzyme family. It was shown that the GtmE-catalyzed transformation from guatrypmethine A to C, by installation of a double bond exo of the DKP ring, follows Michaelis-Menten kinetics and the parameters  $K_M$  and  $k_{cat}$  were determined for substrate and co-substrate as 0.11 mM and 0.2 s<sup>-1</sup>, respectively. The experimental observation that only guatrypmethine A, but not its stable isomer guatrypmethine B, is efficiently converted by GtmE, was substantiated by quantum chemical calculations on the density functional theory level.

The versatility of P450 enzymes is reflected by the investigated GymBs that act as bifunctional oxidases, catalyzing the intramolecular C-C coupling within mycrocyclosin, as well as the transfer of the nucleobases guanine and hypoxanthine onto cYY, resulting in guatyromycine A and B. Their functions were proven for six different representative *Streptomyces* species by in vivo and in vitro experiments. However, the presence of the *gymAB* gene locus in at least 47 sequenced and deposited actinobacteria species and their relation to CYP121, an essential P450 enzyme in *M. tuberculosis*, raises new questions about the physiological or biological function of the GymBs.

In the third project, the substrate flexible PT SasB has been identified during the elucidation of the biosynthesis of streptoazine C. Efficient regular prenylation at C-3 of tryptophan containing CDP derivatives catalyzed by SasB was achieved for at least eight non-native precursors. Thus, SasB might serve as productive tool for prenylation in the field of synthetic biology.

Especially the steadily increasing discovery of novel guanylated CDP derivatives raises questions about their physiological, biological, or potentially pharmacological effects. It is noteworthy, that under laboratory conditions, none of these secondary metabolites were produced in their native strains. Discovery of the original trigger for the activation of the elucidated BGCs might provide insights into whether their products are used against competitors, as signaling molecules for symbionts, or for any other unknown task.

As it was stated in the introduction, predictions suggest that 90 % of the biosynthetic potential of *Streptomyces* species is yet unknown. Regarding P450 enzymes for nucleobase transfer for example, it would be interesting to know whether other nucleobases, like adenine, cytosine, or thymine, are substrates of yet unknown P450 oxidases. In analogy, there might be other PTs associated to CDPSs

## CONCLUSION AND FUTURE PERSPECTIVES

to prenylate either other positions of tryptophan containing CDPs or other CDPs in general. Furthermore, members of other enzyme families could be involved in the modification of CDPS products as tailoring enzymes, like flavin-dependent oxidases or halogenases.

Another promising aspect for future studies is the utilization of the characterized enzymes in the field of synthetic biology. In this way, rationally engineered compounds could be created in a stereo- and regio-specific manner, to e.g. improve pharmacokinetic or pharmacodynamic parameters of a potential drug candidate.



## 6. References

- 1 Woese, C.R., Kandler, O., & Wheelis, M.L. Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4576-4579 (1990).
- 2 Hug, L.A. *et al.* A New View of the Tree of Life. *Nat. Microbiol.* **1**, 16048 (2016).
- 3 Madigan, M. & Martinko, J. *Brock Biology of Microorganisms* (Pearson Education, Inc., London, UK, 2006).
- 4 Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635-645 (2014).
- 5 Whitman, W.B. *Bergey's Manual of Systematic Bacteriology Volume Five The Actinobacteria, Part A and B* (Springer New York, US, 2012).
- 6 Stach, J.E. & Bull, A.T. Estimating and comparing the diversity of marine actinobacteria. *Antonie Van Leeuwenhoek* **87**, 3-9 (2005).
- 7 Barka, E.A. *et al.* Taxonomy, physiology, and natural products of actinobacteria. *Microbiol. Mol. Biol. Rev.* **80**, 1-43 (2016).
- 8 Parte, A.C., Sarda, C.J., Meier-Kolthoff, J.P., Reimer, L.C., & Goker, M. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **70**, 5607-5612 (2020).
- 9 Gao, B. & Gupta, R.S. Phylogenetic framework and molecular signatures for the main clades of the phylum *Actinobacteria*. *Microbiol. Mol. Biol. Rev.* **76**, 66-112 (2012).
- 10 Girard, G. *et al.* Analysis of novel *Kitasatosporae* reveals significant evolutionary changes in conserved developmental genes between *Kitasatospora* and *Streptomyces*. *Antonie Van Leeuwenhoek* **106**, 365-380 (2014).
- 11 Ian, E. *et al.* Genomics of sponge-associated *Streptomyces* spp. closely related to *Streptomyces albus* J1074: insights into marine adaptation and secondary metabolite biosynthesis potential. *PLoS. One.* **9**, e96719 (2014).
- 12 Chater, K.F. Recent advances in understanding *Streptomyces*. *F1000Res.* **5**, 2795 (2016).
- 13 Chater, K.F., Biro, S., Lee, K.J., Palmer, T., & Schrempf, H. The complex extracellular biology of *Streptomyces*. *FEMS Microbiol. Rev.* **34**, 171-198 (2010).
- 14 Wibberg, D. *et al.* Complete genome sequence of *Streptomyces reticuli*, an efficient degrader of crystalline cellulose. *J. Biotechnol.* **222**, 13-14 (2016).
- 15 Citron, C.A., Gleitzmann, J., Laurenzano, G., Pukall, R., & Dickschat, J.S. Terpenoids are widespread in actinomycetes: a correlation of secondary metabolism and genome data. *Chembiochem* **13**, 202-214 (2012).
- 16 Chater, K.F. Differentiation in *Streptomyces*: the properties and programming of diverse cell types in *Streptomyces Molecular Biology and Biotechnology* (ed. Dyson, P.) 43-86 (Caister Academic Press, Norfolk, UK, 2011).
- 17 Elliot, M.A., Buttner, M.J., & Nodwell, J.R. Multicellular development in *Streptomyces* in *Myxobacteria: multicellularity and differentiation* (ed. Whitworth, D.E.) 419-439 (ASM Press, Washington, D.C., US, 2008).
- 18 Schwedock, J., McCormick, J.R., Angert, E.R., Nodwell, J.R., & Losick, R. Assembly of the cell division protein FtsZ into ladder-like structures in the aerial hyphae of *Streptomyces coelicolor*. *Mol. Microbiol.* **25**, 847-858 (1997).
- 19 Flärdh, K. & Buttner, M.J. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* **7**, 36-49 (2009).
- 20 Ensign, J.C. Formation, properties, and germination of actinomycete spores. *Annu. Rev. Microbiol.* **32**, 185-219 (1978).
- 21 Haiser, H.J., Yousef, M.R., & Elliot, M.A. Cell wall hydrolases affect germination, vegetative growth, and sporulation in *Streptomyces coelicolor*. *J. Bacteriol.* **191**, 6501-6512 (2009).
- 22 Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., & Hopwood, D.A. *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, UK, 2000).
- 23 Keulen van, G., Siebring, J., & Dijkhuizen, L. Central carbon metabolic pathways in *Streptomyces* in *Streptomyces molecular biology and biotechnology* (ed. Dyson, P.) 105-124 (Caister Academic Press, Norfolk, UK, 2011).
- 24 Fisher, S.H. Glutamine synthesis in *Streptomyces*--a review. *Gene* **115**, 13-17 (1992).
- 25 Wohlleben, W., Mast, Y., & Reuther, J. Regulation of nitrogen assimilation in *Streptomyces* and other *Actinobacteria* in *Streptomyces molecular biology and biotechnology* (ed. Dyson, P.) 125-136 (Caister Academic Press, Norfolk, UK, 2011).
- 26 Kämpfer, P. *Streptomyces* in *Bergey's Manual of Systematics of Archaea and Bacteria* (Wiley, Hoboken, US, 2015).
- 27 Redenbach, M., Scheel, J., & Schmidt, U. Chromosome topology and genome size of selected actinomycetes species. *Antonie Van Leeuwenhoek* **78**, 227-235 (2000).
- 28 Reeves, A.R., Post, D.A., & Vanden Boom, T.J. Physical-genetic map of the erythromycin-producing organism *Saccharopolyspora erythraea*. *Microbiology* **144**, 2151-2159 (1998).
- 29 Kirby, R. & Chen, C.W. Genome architecture in *Streptomyces molecular biology and biotechnology* (ed. Dyson, P.) 5-26 (Caister Academic Press, Norfolk, UK, 2011).
- 30 McLeod, M.P. *et al.* The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15582-15587 (2006).
- 31 Yang, C.C. *et al.* The terminal proteins of linear *Streptomyces* chromosomes and plasmids: a novel class of replication priming proteins. *Mol. Microbiol.* **43**, 297-305 (2002).

## REFERENCES

- 32 Yang, C.C., Tseng, S.-M., Pan, H.-Y., Huang, C.-H., & Chen, C.W. Telomere associated primase tap repairs truncated telomeres of *Streptomyces*. *Nucleic Acids Research* **45**, 5838-5849 (2017).
- 33 Chen, C.W. *Streptomyces* linear plasmids: replication and telomeres in *Microbial linear plasmids* (eds. Friedhelm, M. & Roland, K.) 33-61 (Springer, Heidelberg, Germany, 2007).
- 34 Zawilak-Pawlik, A. *et al.* Architecture of bacterial replication initiation complexes: orisomes from four unrelated bacteria. *Biochem. J.* **389**, 471-481 (2005).
- 35 Thoma, L. & Muth, G. The conjugative DNA-transfer apparatus of *Streptomyces*. *Int. J. Med. Microbiol.* **305**, 224-229 (2015).
- 36 Vogelmann, J., Wohlleben, W., & Muth, G. *Streptomyces* conjugative genetic elements in *Streptomyces molecular biology and biotechnology* (ed. Dyson, P.) 27-42 (Caister Academic Press, Norfolk, UK, 2011).
- 37 Bobek, J., Šmídová, K., & ěihák, M. A waking review: old and novel insights into the spore germination in *Streptomyces*. *Front. Microbiol.* **8**, 2205 (2017).
- 38 Liot, Q. & Constant, P. Breathing air to save energy--new insights into the ecophysiological role of high-affinity [NiFe]-hydrogenase in *Streptomyces avermitilis*. *MicrobiologyOpen* **5**, 47-59 (2016).
- 39 Wildermuth, H., Wehrli, E., & Horne, R.W. The surface structure of spores and aerial mycelium in *Streptomyces coelicolor*. *J. Ultrastruct. Res.* **35**, 168-180 (1971).
- 40 Mazza, P. *et al.* MreB of *Streptomyces coelicolor* is not essential for vegetative growth but is required for the integrity of aerial hyphae and spores. *Mol. Microbiol.* **60**, 838-852 (2006).
- 41 Heichlinger, A. *et al.* The MreB-like protein Mbl of *Streptomyces coelicolor* A3(2) depends on MreB for proper localization and contributes to spore wall synthesis. *J. Bacteriol.* **193**, 1533-1542 (2011).
- 42 Sola-Penna, M. & Meyer-Fernandes, J.R. Stabilization against thermal inactivation promoted by sugars on enzyme structure and function: why is trehalose more effective than other sugars? *Arch. Biochem. Biophys.* **360**, 10-14 (1998).
- 43 Bobek, J., Halada, P., Angelis, J., Vohradsky, J., & Mikulik, K. Activation and expression of proteins during synchronous germination of aerial spores of *Streptomyces granaticolor*. *Proteomics* **4**, 3864-3880 (2004).
- 44 Hirsch, C.F. & Ensign, J.C. Nutritionally defined conditions for germination of *Streptomyces viridochromogenes* spores. *J. Bacteriol.* **126**, 13-23 (1976).
- 45 Hirsch, C.F. & Ensign, J.C. Heat activation of *Streptomyces viridochromogenes* spores. *J. Bacteriol.* **126**, 24-30 (1976).
- 46 Shah, I.M., Laaberki, M.H., Popham, D.L., & Dworkin, J. A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. *Cell* **135**, 486-496 (2008).
- 47 Hardisson, C., Manzanal, M.B., Salas, J.A., & Suarez, J.E. Fine structure, physiology and biochemistry of arthrospore germination in *Streptomyces antibioticus*. *J. Gen. Microbiol.* **105**, 203-214 (1978).
- 48 Salas, J.A., Guijarro, J.A., & Hardisson, C. High calcium content in *Streptomyces* spores and its release as an early event during spore germination. *J. Bacteriol.* **155**, 1316-1323 (1983).
- 49 Wang, S.L. *et al.* CabC, an EF-hand calcium-binding protein, is involved in Ca<sup>2+</sup>-mediated regulation of spore germination and aerial hypha formation in *Streptomyces coelicolor*. *J. Bacteriol.* **190**, 4061-4068 (2008).
- 50 McBride, M.J. & Ensign, J.C. Regulation of trehalose metabolism by *Streptomyces griseus* spores. *J. Bacteriol.* **172**, 3637-3643 (1990).
- 51 Singer, M.A. & Lindquist, S. Multiple effects of trehalose on protein folding in vitro and in vivo. *Mol. Cell* **1**, 639-648 (1998).
- 52 Mikulik, K., Janda, I., Weiser, J., Stastna, J., & Jiranova, A. RNA and ribosomal protein patterns during aerial spore germination in *Streptomyces granaticolor*. *Eur. J. Biochem.* **145**, 381-388 (1984).
- 53 Strakova, E. *et al.* Systems insight into the spore germination of *Streptomyces coelicolor*. *J. Proteome. Res.* **12**, 525-536 (2013).
- 54 Noens, E.E. *et al.* Loss of the controlled localization of growth stage-specific cell-wall synthesis pleiotropically affects developmental gene expression in an ssgA mutant of *Streptomyces coelicolor*. *Mol. Microbiol.* **64**, 1244-1259 (2007).
- 55 Flårdh, K. Essential role of DivIVA in polar growth and morphogenesis in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **49**, 1523-1536 (2003).
- 56 Flårdh, K. Growth polarity and cell division in *Streptomyces*. *Curr. Opin. Microbiol.* **6**, 564-571 (2003).
- 57 Grantcharova, N., Lustig, U., & Flårdh, K. Dynamics of FtsZ assembly during sporulation in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **187**, 3227-3237 (2005).
- 58 Eaton, D. & Ensign, J.C. *Streptomyces viridochromogenes* spore germination initiated by calcium ions. *J. Bacteriol.* **143**, 377-382 (1980).
- 59 Süssstrunk, U., Pidoux, J., Taubert, S., Ullmann, A., & Thompson, C.J. Pleiotropic effects of cAMP on germination, antibiotic biosynthesis and morphological development in *Streptomyces coelicolor*. *Mol. Microbiol.* **30**, 33-46 (1998).
- 60 Piette, A. *et al.* From dormant to germinating spores of *Streptomyces coelicolor* A3(2): new perspectives from the crp null mutant. *J. Proteome. Res.* **4**, 1699-1708 (2005).
- 61 McCormick, J.R. Cell division is dispensable but not irrelevant in *Streptomyces*. *Curr. Opin. Microbiol.* **12**, 689-698 (2009).
- 62 Grund, A.D. & Ensign, J.C. Properties of the germination inhibitor of *Streptomyces viridochromogenes* spores. *J. Gen. Microbiol.* **131**, 833-847 (1985).
- 63 Aoki, Y., Matsumoto, D., Kawaide, H., & Natsume, M. Physiological role of germicidins in spore germination and hyphal elongation in *Streptomyces coelicolor* A3(2). *J. Antibiot. (Tokyo)* **64**, 607-611 (2011).
- 64 Xu, Y. & Vetsigian, K. Phenotypic variability and community interactions of germinating *Streptomyces* spores. *Sci. Rep.* **7**, 699 (2017).

## REFERENCES

- 65 Elliot, M.A. *et al.* The chaplins: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in *Streptomyces coelicolor*. *Genes Dev.* **17**, 1727-1740 (2003).
- 66 Claessen, D. *et al.* The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlines and chaplins. *Mol. Microbiol.* **53**, 433-443 (2004).
- 67 Claessen, D., de Jong, W., Dijkhuizen, L., & Wosten, H.A. Regulation of *Streptomyces* development: reach for the sky! *Trends Microbiol.* **14**, 313-319 (2006).
- 68 Chandra, G., Chater, K.F., & Bornemann, S. Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. *Microbiology* **157**, 1565-1572 (2011).
- 69 Bibb, M.J. Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **8**, 208-215 (2005).
- 70 van Wezel, G.P. & McDowall, K.J. The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat. Prod. Rep.* **28**, 1311-1333 (2011).
- 71 Sigle, S., Ladwig, N., Wohleben, W., & Muth, G. Synthesis of the spore envelope in the developmental life cycle of *Streptomyces coelicolor*. *Int. J. Med. Microbiol.* **305**, 183-189 (2015).
- 72 Flårdh, K., Findlay, K.C., & Chater, K.F. Association of early sporulation genes with suggested developmental decision points in *Streptomyces coelicolor* A3(2). *Microbiology* **145**, 2229-2243 (1999).
- 73 Flårdh, K., Leibovitz, E., Buttner, M.J., & Chater, K.F. Generation of a non-sporulating strain of *Streptomyces coelicolor* A3(2) by the manipulation of a developmentally controlled *ftsZ* promoter. *Mol. Microbiol.* **38**, 737-749 (2000).
- 74 Kormanec, J. *et al.* The *Streptomyces aureofaciens* homologue of the *whiB* gene is essential for sporulation; its expression correlates with the developmental stage. *Folia Microbiol.* **43**, 605-612 (1998).
- 75 Noens, E.E. *et al.* SsgA-like proteins determine the fate of peptidoglycan during sporulation of *Streptomyces coelicolor*. *Mol. Microbiol.* **58**, 929-944 (2005).
- 76 Traag, B.A. & van Wezel, G.P. The SsgA-like proteins in actinomycetes: small proteins up to a big task. *Antonie Van Leeuwenhoek* **94**, 85-97 (2008).
- 77 Willemse, J., Borst, J.W., de Waal, E., Bisseling, T., & van Wezel, G.P. Positive control of cell division: FtsZ is recruited by SsgB during sporulation of *Streptomyces*. *Genes Dev.* **25**, 89-99 (2011).
- 78 Loria, R., Bukhalid, R.A., Fry, B.A., & King, R.R. Plant pathogenicity in the genus *Streptomyces*. *Plant Dis.* **81**, 836-846 (1997).
- 79 King, R.R. & Calhoun, L.A. The thaxtomin phytotoxins: sources, synthesis, biosynthesis, biotransformation and biological activity. *Phytochemistry* **70**, 833-841 (2009).
- 80 Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770-803 (2020).
- 81 Riley, J.C. Estimates of regional and global life expectancy, 1800 – 2001. *Population and Development Review* **31**, 537-543 (2005).
- 82 Olano, C., Méndez, C., & Salas, J.A. Gene clusters for bioactive natural products in *Actinomycetes* and their use in combinatorial biosynthesis in *Streptomyces molecular biology and biotechnology* (ed. Dyson, P.) 195-232 (Caister Academic Press, Norfolk, UK, 2011).
- 83 Katz, L. & Baltz, R.H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155-176 (2016).
- 84 Thompson, R.M. & Strong, F.M. Identification of erythromycin A in cultures of *Streptomyces griseoplanus*. *Biochem. Biophys. Res. Commun.* **43**, 213-216 (1971).
- 85 Gaynor, M. & Mankin, A.S. Macrolide antibiotics: binding site, mechanism of action, resistance. *Curr. Top. Med. Chem.* **3**, 949-961 (2003).
- 86 Yanai, K. & Murakami, T. The kanamycin biosynthetic gene cluster from *Streptomyces kanamyceticus*. *J. Antibiot. (Tokyo)* **57**, 351-354 (2004).
- 87 Piepersberg, W. Streptomycin and related aminoglycosides. *Biotechnology* **28**, 531-570 (1995).
- 88 Behal, V. & Hunter, I.S. Tetracyclines. *Biotechnology* **28**, 359-384 (1995).
- 89 Doull, J.L. *et al.* Conjugational fertility and location of chloramphenicol biosynthesis genes on the chromosomal linkage map of *Streptomyces venezuelae*. *J. Gen. Microbiol.* **132**, 1327-1338 (1986).
- 90 Rogers, T.O. & Birnbaum, J. Biosynthesis of fosfomycin by *Streptomyces fradiae*. *Antimicrob. Agents Chemother.* **5**, 121-132 (1974).
- 91 Ward, J.M. & Hodgson, J.E. The biosynthetic genes for clavulanic acid and cephamycin production occur as a 'super-cluster' in three *Streptomyces*. *FEMS Microbiol. Lett.* **110**, 239-242 (1993).
- 92 Paradkar, A.S. & Jensen, S.E. Functional analysis of the gene encoding the clavamate synthase 2 isoenzyme involved in clavulanic acid biosynthesis in *Streptomyces clavuligerus*. *J. Bacteriol.* **177**, 1307-1314 (1995).
- 93 Caffrey, P., Lynch, S., Flood, E., Finnan, S., & Oliynyk, M. Amphotericin biosynthesis in *Streptomyces nodosus*: deductions from analysis of polyketide synthase and late genes. *Chem. Biol.* **8**, 713-723 (2001).
- 94 Zhou, S. *et al.* Pentamycin Biosynthesis in Philippine *Streptomyces* sp. S816: Cytochrome P450-Catalyzed Installation of the C-14 Hydroxyl Group. *ACS Chem. Biol.* **14**, 1305-1309 (2019).
- 95 Aparicio, J.F. *et al.* Biotechnological production and application of the antibiotic pimaricin: biosynthesis and its regulation. *Appl. Microbiol. Biotechnol.* **100**, 61-78 (2016).
- 96 Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., & Omura, S. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proc. Natl. Acad. Sci. U. S. A* **96**, 9509-9514 (1999).
- 97 Williamson, N.R., Fineran, P.C., Leeper, F.J., & Salmond, G.P. The biosynthesis and regulation of bacterial prodiginines. *Nat. Rev. Microbiol.* **4**, 887-899 (2006).

## REFERENCES

- 98 Shen,B. *et al.* Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003(1). *J. Nat. Prod.* **65**, 422-431 (2002).
- 99 Otten,S.L., Stutzman-Engwall,K.J., & Hutchinson,C.R. Cloning and expression of daunorubicin biosynthesis genes from *Streptomyces peucetius* and *S. peucetius* subsp. *caesius*. *J. Bacteriol.* **172**, 3427-3434 (1990).
- 100 Hutchinson,C. & Colombo,A. Genetic engineering of doxorubicin production in *Streptomyces peucetius*: a review. *J. Ind. Microbiol. Biotechnol.* **23**, 647-652 (1999).
- 101 Bolourian,A. & Mojtahedi,Z. Immunosuppressants produced by *Streptomyces*: evolution, hygiene hypothesis, tumour rapalog resistance and probiotics. *Environ. Microbiol. Rep.* **10**, 123-126 (2018).
- 102 Jizba,J. *et al.* Macrotetrolide antibiotics produced by *Streptomyces globisporus*. *Folia Microbiol.* **36**, 437-443 (1991).
- 103 Kirby,R. *et al.* Draft genome sequence of the human pathogen *Streptomyces somaliensis*, a significant cause of actinomycetoma. *J. Bacteriol.* **194**, 3544-3545 (2012).
- 104 Quintana,E.T. *et al.* *Streptomyces sudanensis* sp. nov., a new pathogen isolated from patients with actinomycetoma. *Antonie Van Leeuwenhoek* **93**, 305-313 (2008).
- 105 Herbrink,A. *et al.* A human lung-associated *Streptomyces* sp. TR1341 produces various secondary metabolites responsible for virulence, cytotoxicity and modulation of immune response. *Front Microbiol.* **10**, 3028 (2020).
- 106 Czaplewski,L. *et al.* Alternatives to antibiotics-a pipeline portfolio review. *Lancet Infect. Dis.* **16**, 239-251 (2016).
- 107 Nett,M., Ikeda,H., & Moore,B.S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26**, 1362-1384 (2009).
- 108 Watve,M.G., Tickoo,R., Jog,M.M., & Bhole,B.D. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **176**, 386-390 (2001).
- 109 Medema,M.H., de Rond,T., & Moore,B.S. Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.* **22**, 553-571 (2021).
- 110 Scherlach,K. & Hertweck,C. Mining and unearthing hidden biosynthetic potential. *Nat. Commun.* **12**, 3864 (2021).
- 111 Blin,K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29-W35 (2021).
- 112 Wright,G.D. Unlocking the potential of natural products in drug discovery. *Microb. Biotechnol.* **12**, 55-57 (2019).
- 113 Scherlach,K. & Hertweck,C. Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **7**, 1753-1760 (2009).
- 114 Rutledge,P.J. & Challis,G.L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13**, 509-523 (2015).
- 115 Onaka,H., Mori,Y., Igarashi,Y., & Furumai,T. Mycolic acid-containing bacteria induce natural-product biosynthesis in *Streptomyces* species. *Appl. Environ. Microbiol.* **77**, 400-406 (2011).
- 116 König,C.C. *et al.* Bacterium induces cryptic meroterpenoid pathway in the pathogenic fungus *Aspergillus fumigatus*. *Chembiochem* **14**, 938-942 (2013).
- 117 Hoskisson,P.A. & Fernandez-Martinez,L.T. Regulation of specialised metabolites in *Actinobacteria* - expanding the paradigms. *Environ. Microbiol. Rep.* **10**, 231-238 (2018).
- 118 Romero-Rodriguez,A., Robledo-Casados,I., & Sanchez,S. An overview on transcriptional regulators in *Streptomyces*. *Biochim. Biophys. Acta* **1849**, 1017-1039 (2015).
- 119 Ochi,K. & Hosaka,T. New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl. Microbiol. Biotechnol.* **97**, 87-98 (2013).
- 120 Xu,J., Tozawa,Y., Lai,C., Hayashi,H., & Ochi,K. A rifampicin resistance mutation in the *rpoB* gene confers ppGpp-independent antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Genet. Genomics* **268**, 179-189 (2002).
- 121 Hosaka,T. *et al.* Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol.* **27**, 462-464 (2009).
- 122 Brosch,G., Loidl,P., & Graessle,S. Histone modifications and chromatin dynamics: a focus on filamentous fungi. *FEMS Microbiol. Rev.* **32**, 409-439 (2008).
- 123 Cichewicz,R.H. Epigenome manipulation as a pathway to new natural product scaffolds and their congeners. *Nat. Prod. Rep.* **27**, 11-22 (2010).
- 124 Gomez-Escribano,J.P. *et al.* Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the *cpk* gene cluster of *Streptomyces coelicolor* M145. *Chem. Sci.* **3**, 2716-2720 (2012).
- 125 Chiang,Y.M. *et al.* A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in *Aspergillus nidulans*. *J. Am. Chem. Soc.* **131**, 2965-2970 (2009).
- 126 Gomez-Escribano,J.P. & Bibb,M.J. *Streptomyces coelicolor* as an expression host for heterologous gene clusters. *Methods Enzymol.* **517**, 279-300 (2012).
- 127 Zaburanyi,N., Rabyk,M., Ostash,B., Fedorenko,V., & Luzhetskyy,A. Insights into naturally minimised *Streptomyces albus* J1074 genome. *Bmc Genomics* **15**, 97 (2014).
- 128 Anne,J. & Van Mellaert,L. *Streptomyces lividans* as host for heterologous protein production. *FEMS Microbiol. Lett.* **114**, 121-128 (1993).
- 129 Zhang,J.J., Tang,X., & Moore,B.S. Genetic platforms for heterologous expression of microbial natural products. *Nat. Prod. Rep.* **36**, 1313-1332 (2019).

## REFERENCES

- 130 Huo,L. *et al.* Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **36**, 1412-1436 (2019).
- 131 Liu,J., Xie,X., & Li,S.-M. Guanitrypmycin biosynthetic pathways imply cytochrome P450-mediated regio- and stereospecific guaninyl transfer reactions. *Angew. Chem. Int. Ed. Engl.* **58**, 11534-11540 (2019).
- 132 Belin,P. *et al.* The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. *Nat. Prod. Rep.* **29**, 961-979 (2012).
- 133 Borthwick,A.D. 2,5-diketopiperazines: synthesis, reactions, medicinal chemistry, and bioactive natural products. *Chem. Rev.* **112**, 3641-3716 (2012).
- 134 Sun,S.J. *et al.* Cyclic dipeptides mediating quorum sensing and their biological effects in *Hypsizygus marmoreus*. *Biomolecules* **10**, 298 (2020).
- 135 Zhu,J. *et al.* Diketopiperazines synthesis gene in *Shewanella baltica* and roles of diketopiperazines and resveratrol in quorum sensing. *J. Agric. Food Chem.* **67**, 12013-12025 (2019).
- 136 Apostolopoulos,V. *et al.* A global review on short peptides: Frontiers and perspectives. *Molecules* **26**, 430 (2021).
- 137 Song,Z., Hou,Y., Yang,Q., Li,X., & Wu,S. Structures and biological activities of diketopiperazines from marine organisms: A review. *Mar. Drugs* **19**, 403 (2021).
- 138 Ding,Z. *et al.* Structure-based design and synthesis of novel furan-diketopiperazine-type derivatives as potent microtubule inhibitors for treating cancer. *Bioorg. Med. Chem.* **28**, 115435 (2020).
- 139 Meng,S. *et al.* A six-oxidase cascade for tandem C-H bond activation revealed by reconstitution of bicyclomycin biosynthesis. *Angew. Chem Int. Ed. Engl.* **57**, 719-723 (2018).
- 140 Patteson,J.B., Cai,W., Johnson,R.A., Santa Maria,K.C., & Li,B. Identification of the biosynthetic pathway for the antibiotic bicyclomycin. *Biochemistry* **57**, 61-65 (2017).
- 141 Kohn,H. & Widger,W. The molecular basis for the mode of action of bicyclomycin. *Curr. Drug Targets. Infect. Disord.* **5**, 273-295 (2005).
- 142 Shimi,I.R., Abedallah,N., & Fathy,S. Cairomycin B, a new antibiotic. *Antimicrob. Agents Chemother.* **11**, 373-375 (1977).
- 143 Bryans,J. *et al.* Inhibition of plasminogen activator inhibitor-1 activity by two diketopiperazines, XR330 and XR334 produced by *Streptomyces* sp. *J. Antibiot. (Tokyo)* **49**, 1014-1021 (1996).
- 144 Tian,W. *et al.* Efficient biosynthesis of heterodimeric C<sup>3</sup>-aryl pyrroloindoline alkaloids. *Nat. Commun.* **9**, 4428 (2018).
- 145 Wang,X., Li,Y., Zhang,X., Lai,D., & Zhou,L. Structural diversity and biological activities of the cyclodipeptides from fungi. *Molecules* **22**, 2026 (2017).
- 146 Aninat,C., Hayashi,Y., Andre,F., & Delaforge,M. Molecular requirements for inhibition of cytochrome p450 activities by roquefortine. *Chem. Res. Toxicol.* **14**, 1259-1265 (2001).
- 147 Hof,H. & Kupfahl,C. Gliotoxin in *Aspergillus fumigatus*: an example that mycotoxins are potential virulence factors. *Mycotoxin. Res.* **25**, 123-131 (2009).
- 148 Vigushin,D.M. *et al.* Gliotoxin is a dual inhibitor of farnesyltransferase and geranylgeranyltransferase I with antitumor activity against breast cancer in vivo. *Med. Oncol.* **21**, 21-30 (2004).
- 149 Dolan,S.K., O'Keeffe,G., Jones,G.W., & Doyle,S. Resistance is not futile: gliotoxin biosynthesis, functionality and utility. *Trends Microbiol.* **23**, 419-428 (2015).
- 150 Wauters,I. *et al.* Beyond the diketopiperazine family with alternatively bridged brevianamide F analogues. *J. Org. Chem.* **80**, 8046-8054 (2015).
- 151 Maiya,S., Grundmann,A., Li,S.-M., & Turner,G. The fumitremorgin gene cluster of *Aspergillus fumigatus*: identification of a gene encoding brevianamide F synthetase. *Chembiochem* **7**, 1062-1069 (2006).
- 152 Zhao,S. *et al.* Biological activity of the tryprostatins and their diastereomers on human carcinoma cell lines. *J. Med. Chem.* **45**, 1559-1562 (2002).
- 153 Sugie,Y. *et al.* A new antibiotic CJ-17,665 from *Aspergillus ochraceus*. *J. Antibiot. (Tokyo)* **54**, 911-916 (2001).
- 154 Nicholson,B. *et al.* NPI-2358 is a tubulin-depolymerizing agent: in-vitro evidence for activity as a tumor vascular-disrupting agent. *Anticancer Drugs* **17**, 25-31 (2006).
- 155 Blayney,D.W. *et al.* Efficacy of plinabulin vs pegfilgrastim for prevention of docetaxel-induced neutropenia in patients with solid tumors: a randomized clinical trial. *JAMA Netw. Open.* **5**, e2145446 (2022).
- 156 Coward,R.M. & Carson,C.C. Tadalafil in the treatment of erectile dysfunction. *Ther. Clin. Risk Manag.* **4**, 1315-1330 (2008).
- 157 Falk,J.A., Philip,K.J., & Schwarz,E.R. The emergence of oral tadalafil as a once-daily treatment for pulmonary arterial hypertension. *Vasc. Health Risk Manag.* **6**, 273-280 (2010).
- 158 Borthwick,A.D. & Liddle,J. The design of orally bioavailable 2, 5-diketopiperazine oxytocin antagonists: from concept to clinical candidate for premature labor. *Med. Res. Rev.* **31**, 576-604 (2011).
- 159 Balachandra,C., Padhi,D., & Govindaraju,T. Cyclic dipeptide: a privileged molecular scaffold to derive structural diversity and functional utility. *ChemMedChem* **16**, 2558-2587 (2021).
- 160 Payne,J.A., Schoppet,M., Hansen,M.H., & Cryle,M.J. Diversity of nature's assembly lines - recent discoveries in non-ribosomal peptide synthesis. *Mol. Biosyst.* **13**, 9-22 (2017).
- 161 Canu,N., Moutiez,M., Belin,P., & Gondry,M. Cyclodipeptide synthases: a promising biotechnological tool for the synthesis of diverse 2,5-diketopiperazines. *Nat. Prod. Rep.* **37**, 312-321 (2020).
- 162 Baccile,J.A. *et al.* Diketopiperazine formation in fungi requires dedicated cyclization and thiolation domains. *Angew. Chem.* **131**, 14731-14735 (2019).
- 163 Moutiez,M. *et al.* Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat. Commun.* **5**, 5141 (2014).

## REFERENCES

- 164 Jacques, I.B. *et al.* Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat. Chem. Biol.* **11**, 721-727 (2015).
- 165 Gondry, M. *et al.* A comprehensive overview of the cyclodipeptide synthase family enriched with the characterization of 32 new enzymes. *Front. Microbiol.* **9**, 46 (2018).
- 166 Kanzaki, H., Imura, D., Nitoda, T., & Kawazu, K. Enzymatic dehydrogenation of cyclo(L-Phe-L-Leu) to a bioactive derivative, albonoursin. *J. Mol. Cat. B: Enzym.* **6**, 265-270 (1999).
- 167 Gondry, M. *et al.* Cyclic dipeptide oxidase from *Streptomyces noursei*. Isolation, purification and partial characterization of a novel, amino acyl  $\alpha,\beta$ -dehydrogenase. *Eur. J. Biochem.* **268**, 1712-1721 (2001).
- 168 Lautru, S., Gondry, M., Genet, R., & Pernodet, J.L. The albonoursin gene cluster of *S. noursei*: Biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. *Chem. Biol.* **9**, 1355-1364 (2002).
- 169 Giessen, T.W., von Tesmar, A.M., & Marahiel, M.A. Insights into the generation of structural diversity in a tRNA-dependent pathway for highly modified bioactive cyclic dipeptides. *Chem. Biol.* **20**, 828-838 (2013).
- 170 Le Chevalier, F. *et al.* In vivo characterization of the activities of novel cyclodipeptide oxidases: new tools for increasing chemical diversity of bioproduced 2,5-diketopiperazines in *Escherichia coli*. *Microb. Cell Fact.* **19**, 178 (2020).
- 171 Mikulski, L., Schäfer, J., Brockmeyer, K., Kraut, R., & Li, S.-M. Comparative studies on similarities and differences of cyclodipeptide oxidases for installation of C-C double bonds at the diketopiperazine ring. *Appl. Microbiol. Biotechnol.* **104**, 2523-2536 (2020).
- 172 Shi, J. *et al.* Genome mining and enzymatic total biosynthesis of purincyclamide. *Org. Lett.* **21**, 6825-6829 (2019).
- 173 Nelson, D.R. Cytochrome P450 diversity in the tree of life. *Biochim. Biophys. Acta Proteins Proteom.* **1866**, 141-154 (2018).
- 174 Dunham, N.P. & Arnold, F.H. Nature's machinery, repurposed: Expanding the repertoire of iron-dependent oxygenases. *ACS Catal.* **10**, 12239-12255 (2020).
- 175 Chen, C.-C. *et al.* Advanced understanding of electron transfer pathway of cytochrome P450s. *ChemBiochem* **22**, 1317-1328 (2021).
- 176 Klingenberg, M. Pigments of rat liver microsomes. *Arch. Biochem. Biophys.* **409**, 2-6 (1957).
- 177 Rudolf, J.D., Chang, C.Y., Ma, M., & Shen, B. Cytochromes P450 for natural product biosynthesis in *Streptomyces*: sequence, structure, and function. *Nat. Prod. Rep.* **34**, 1141-1172 (2017).
- 178 Greule, A., Stok, J.E., De Voss, J.J., & Cryle, M.J. Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism. *Nat. Prod. Rep.* **35**, 757-791 (2018).
- 179 Cryle, M.J., Bell, S.G., & Schlichting, I. Structural and biochemical characterization of the cytochrome P450 CypX (CYP134A1) from *Bacillus subtilis*: a cyclo-L-leucyl-L-leucyl dipeptide oxidase. *Biochemistry* **49**, 7282-7296 (2010).
- 180 Liu, J., Xie, X., & Li, S.-M. Increasing cytochrome P450 enzyme diversity by identification of two distinct cyclodipeptide dimerases. *Chem. Commun.* **56**, 11042-11045 (2020).
- 181 Kalb, V.F. & Loper, J.C. Proteins from eight eukaryotic cytochrome P-450 families share a segmented region of sequence similarity. *Proc. Natl. Acad. Sci. U. S. A* **85**, 7221-7225 (1988).
- 182 McNaught, A.D. & Wilkinson, A. *Compendium of Chemical Terminology Gold Book* (International Union of Pure and Applied Chemistry, 2014).
- 183 Podust, L.M. & Sherman, D.H. Diversity of P450 enzymes in the biosynthesis of natural products. *Nat. Prod. Rep.* **29**, 1251-1266 (2012).
- 184 Rupasinghe, S. *et al.* The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine. *FEBS Lett.* **580**, 6338-6342 (2006).
- 185 Shende, V.V. *et al.* Structure and function of NzeB, a versatile C-C and C-N bond-forming diketopiperazine dimerase. *J. Am. Chem. Soc.* **142**, 17413-17424 (2020).
- 186 Belin, P. *et al.* Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A* **106**, 7426-7431 (2009).
- 187 Yu, H. & Li, S.-M. Two cytochrome P450 enzymes from *Streptomyces* sp. NRRL S-1868 catalyze distinct dimerization of tryptophan-containing cyclodipeptides. *Org. Lett.* **21**, 7094-7098 (2019).
- 188 Sun, C. *et al.* Molecular basis of regio- and stereo-specificity in biosynthesis of bacterial heterodimeric diketopiperazines. *Nat. Commun.* **11**, 6251 (2020).
- 189 Yu, H., Xie, X., & Li, S.-M. Coupling of guanine with cyclo-L-Trp-L-Trp mediated by a cytochrome P450 homologue from *Streptomyces purpureus*. *Org. Lett.* **20**, 4921-4925 (2018).
- 190 Yu, H., Xie, X., & Li, S.-M. Coupling of cyclo-L-Trp-L-Trp with hypoxanthine increases the structure diversity of guanitryptmycins. *Org. Lett.* **21**, 9104-9108 (2019).
- 191 Herr, C.Q. & Hausinger, R.P. Amazing diversity in biochemical roles of Fe(II)/2-oxoglutarate oxygenases. *Trends Biochem. Sci.* **43**, 517-532 (2018).
- 192 Hibi, M. *et al.* Novel enzyme family found in filamentous fungi catalyzing trans-4-hydroxylation of L-pipecolic acid. *Appl. Environ. Microbiol.* **82**, 2070-2077 (2016).
- 193 Li, C., Junaid, M., Almuqri, E.A., Hao, S., & Zhang, H. Structural analysis of a phosphonate hydroxylase with an access tunnel at the back of the active site. *Acta Crystallogr. F. Struct. Biol. Commun.* **72**, 362-368 (2016).
- 194 Tao, H. *et al.* Molecular insights into the unusually promiscuous and catalytically versatile Fe(II)/ $\alpha$ -ketoglutarate-dependent oxygenase SptF. *Nat. Commun.* **13**, 95 (2022).

## REFERENCES

- 195 Chang, W.C., Li, J., Lee, J.L., Cronican, A.A., & Guo, Y. Mechanistic investigation of a non-heme iron enzyme catalyzed epoxidation in (-)-4'-methoxycyclophenin biosynthesis. *J. Am. Chem. Soc.* **138**, 10390-10393 (2016).
- 196 Hillwig, M.L., Zhu, Q., Ittiamornkul, K., & Liu, X. Discovery of a promiscuous non-heme iron halogenase in ambiguine alkaloid biogenesis: implication for an evolvable enzyme family for late-stage halogenation of aliphatic carbons in small molecules. *Angew. Chem. Int. Ed Engl.* **55**, 5780-5784 (2016).
- 197 Dunham, N.P. *et al.* Two distinct mechanisms for C-C desaturation by Iron(II)- and 2-(oxo)glutarate-dependent oxygenases: Importance of  $\alpha$ -heteroatom assistance. *J. Am. Chem. Soc.* **140**, 7116-7126 (2018).
- 198 Li, J. *et al.* Epoxidation catalyzed by the nonheme iron(II)- and 2-oxoglutarate-dependent oxygenase, AsqJ: Mechanistic elucidation of oxygen atom transfer by a ferryl intermediate. *J. Am. Chem. Soc.* **142**, 6268-6284 (2020).
- 199 Martinez, S. & Hausinger, R.P. Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* **290**, 20702-20711 (2015).
- 200 Winkelblech, J., Fan, A., & Li, S.-M. Prenyltransferases as key enzymes in primary and secondary metabolism. *Appl. Microbiol. Biotechnol.* **99**, 7379-7397 (2015).
- 201 Sunassee, S.N. & Davies-Coleman, M.T. Cytotoxic and antioxidant marine prenylated quinones and hydroquinones. *Nat. Prod. Rep.* **29**, 513-535 (2012).
- 202 Qiu, C., Liu, Y., Wu, Y., Zhao, L., & Pei, J. Biochemical Characterization of a Novel Prenyltransferase from *Streptomyces* sp. NT11 and Development of a Recombinant Strain for the Production of 6-Prenylnarigenin. *J. Agric. Food Chem.* **69**, 14231-14240 (2021).
- 203 Heide, L. Prenyl transfer to aromatic substrates: genetics and enzymology. *Curr. Opin. Chem. Biol.* **13**, 171-179 (2009).
- 204 Yu, X., Liu, Y., Xie, X., Zheng, X.-D., & Li, S.-M. Biochemical characterization of indole prenyltransferases: Filling the last gap of prenylation positions by a 5-dimethylallyltryptophan synthase from *Aspergillus clavatus*. *J. Biol. Chem.* **287**, 1371-1380 (2012).
- 205 Raju, R., Piggott, A.M., Huang, X.C., & Capon, R.J. Nocardioazines: a novel bridged diketopiperazine scaffold from a marine-derived bacterium inhibits p-glycoprotein. *Org. Lett.* **13**, 2770-2773 (2011).
- 206 Tagami, K. *et al.* Reconstitution of biosynthetic machinery for indole-diterpene paxilline in *Aspergillus oryzae*. *J. Am. Chem. Soc.* **135**, 1260-1263 (2013).
- 207 Yao, T. *et al.* Genome mining of cyclodipeptide synthases unravels unusual tRNA-dependent diketopiperazine-terpene biosynthetic machinery. *Nat. Commun.* **9**, 4091 (2018).
- 208 Ran, H. & Li, S.-M. Fungal benzene carbaldehydes: occurrence, structural diversity, activities and biosynthesis. *Nat. Prod. Rep.* **38**, 240-263 (2021).
- 209 Tello, M., Kuzuyama, T., Heide, L., Noel, J.P., & Richard, S.B. The ABBA family of aromatic prenyltransferases: broadening natural product diversity. *Cell Mol. Life Sci.* **65**, 1459-1463 (2008).
- 210 Li, S.-M. Applications of dimethylallyltryptophan synthases and other indole prenyltransferases for structural modification of natural products. *Appl. Microbiol. Biotechnol.* **84**, 631-639 (2009).
- 211 Pockrandt, D., Ludwig, L., Fan, A., König, G.M., & Li, S.-M. New insights into the biosynthesis of prenylated xanthenes: XptB from *Aspergillus nidulans* catalyses an O-prenylation of xanthenes. *Chembiochem* **13**, 2764-2771 (2012).
- 212 Yu, X. & Li, S.-M. Prenyltransferases of the dimethylallyltryptophan synthase superfamily. *Methods Enzymol.* **516**, 259-278 (2012).
- 213 Zhang, Y. *et al.* Deciphering a cyclodipeptide synthase pathway encoding prenylated indole alkaloids in *Streptomyces leeuwenhoekii*. *Appl. Environ. Microbiol.* **87**, e02525-20 (2021).
- 214 Roach, P.L. *et al.* Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827-830 (1997).
- 215 Zhu, Y. *et al.* Identification of caerulomycin A gene cluster implicates a tailoring amidohydrolase. *Org. Lett.* **14**, 2666-2669 (2012).
- 216 Liao, H.J. *et al.* Insights into the desaturation of cyclopeptin and its C3 epimer catalyzed by a non-heme iron enzyme: structural characterization and mechanism elucidation. *Angew. Chem. Int. Ed. Engl.* **57**, 1831-1835 (2018).
- 217 Giessen, T.W., von Tesmar, A.M., & Marahiel, M.A. A tRNA-dependent two-enzyme pathway for the generation of singly and doubly methylated ditryptophan 2,5-diketopiperazines. *Biochemistry* **52**, 4274-4283 (2013).
- 218 Dumas, V.G., Defelipe, L.A., Petruk, A.A., Turjanski, A.G., & Marti, M.A. QM/MM study of the C-C coupling reaction mechanism of CYP121, an essential cytochrome p450 of *Mycobacterium tuberculosis*. *Proteins* **82**, 1004-1021 (2014).
- 219 Wohlgemuth, V., Kindinger, F., Xie, X., Wang, B.G., & Li, S.-M. Two prenyltransferases govern a consecutive prenylation cascade in the biosynthesis of echinulin and neoechinulin. *Org. Lett.* **19**, 5928-5931 (2017).
- 220 Fan, A., Winkelblech, J., & Li, S.-M. Impacts and perspectives of prenyltransferases of the DMATS superfamily for use in biotechnology. *Appl. Microbiol. Biotechnol.* **99**, 7399-7415 (2015).





## Acknowledgements

In the first place, I sincerely want to thank my doctoral supervisor Prof. Dr. Shu-Ming Li for the opportunity to do research on this marvelous field of microbiological biotechnology. I am truthfully grateful for his indispensable knowledge, suggestions, and guidance throughout my PhD studies. I could always consult him straightforward and got immediate advice. On the other hand he offered me a beneficial liberty and trusted my work. Thank you for this great time in Marburg and for the invaluable experience I got in your lab.

I honestly thank Prof. Dr. Raphael Reher for being my second supervisor and his encouraging and productive remarks. Furthermore, my special thanks to Prof. Dr. Wibke Diederich and Prof. Dr. Jens Kockskämper for being members of the examination board.

My greatest thanks to my lab mother and coach Dr. Jing Liu for her patience and detailed explanations. Her strict and accurate nature always reminded me to work carefully and precisely. My projects might not have been successful without her supervision and extensive introduction into the laboratory work. Besides her yelling at me for doing stupid mistakes, I am very thankful for the times we laughed and had fun. I also thank Yiling Yang for cooperating in the second (very spontaneously) and third project, and her diligent work attitude.

I also deeply thank Lina Keutzer, Sina Stierle, and Shaun Österlind for their critical remarks on my thesis and their help to correct as many mistakes as possible.

I thank:

Chroissie, Tim, and Hauke for the energy I could reload during every holiday we spent together.

Lena Ludwig-Radtke, Rixa Kraut, Stefan Newel, Dr. Regina Ortmann, and Dr. Xiulan Xie for LC-MS and NMR measurement as well as their constructive advice thereof.

Sina Stierle and Johanna Schäfer for the organization of important parts in our lab especially taking care of primer, enzyme, and sequencing ordering. Our work would be much more complicated without your precise overview.

Jenny, Daniel, Sandra, Philipp, and David for somehow managing to control a mass of students every semester without any severe incidents. You were always reliable and uncomplicated. Special thanks to Dr. Lindsay Coby for handing over your former student lab to me and for all your useful information.

Andi for suffering together with me, when our beloved autoclave decided to break again, again, and again...

Basti, Dr. Jonas, coach Jing, Andi, Hendrik, and David for all the times we laughed and made fun in the coolest lab that I could imagine, but also for enabling a trustful, peaceful, and cooperative atmosphere.

All my other colleagues, also from Prof. Reher's and Prof. Petersen's group, for the nice time we spent together in and outside the lab, as well as for your invaluable scientific advices. Thanks to Wen, Danielo,

## ACKNOWLEDGEMENTS

Marlies, Haowen, Pan, Xuping, Yiling, Yu, Zhengxi, Zhanghai, Dr. Jing Zhou, Sonja, Tim, Amira, Kristin, Dr. Katja Backhaus, Dr. Liujuan Zheng, Dr. Ge Liao, Dr. Huomiao Ran, and Dr. Jie Fan.

Above all, my biggest thanks go to my beloved girlfriend Sina and my dearest family, especially my parents and my sister, for your infinite love and support throughout the time in Marburg and my entire life. I am truthfully glad to have all of you at my side and that I can count on you.