

**Genome mining-directed discovery of novel 2,5-
diketopiperazines from actinobacteria**

**Genome mining-gestützte Entdeckung neuartiger
2,5-Diketopiperazine aus Actinobakterien**

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem Fachbereich Pharmazie
der Philipps-Universität Marburg

vorgelegt von

Jing Liu
aus Jining, China

Marburg an der Lahn, 2021

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

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

Eingereicht am 06. Januar 2021

Tag der mündlichen Prüfung: 03. März 2021

Hochschulkennziffer: 1180

Dedicated to my family

Table of contents

List of publications	III
Abbreviations	VII
Summary	1
Zusammenfassung	3
1 Introduction	5
1.1 Actinobacteria: the most prolific source of natural products	5
1.2 Genome mining for novel secondary metabolites from actinobacteria	6
1.2.1 Genome mining	7
1.2.2 Manipulation of pathway-specific regulatory genes	9
1.2.3 Heterologous expression	9
1.3 2,5-Diketopiperazines	11
1.3.1 DKP scaffolds from NRPS pathways	13
1.3.2 DKP scaffolds from CDPS pathways	14
1.4 Tailoring enzymes in CDPS-dependent pathways	16
1.4.1 Cytochrome P450s	16
1.4.1.1 The P450 catalytic cycle	16
1.4.1.2 P450-mediated diverse reactions in natural product biosynthesis	18
1.4.1.3 P450s involved in CDPS-dependent pathways	21
1.4.2 Other characterized enzymes in CDPS-dependent pathways	22
2 Aims of this thesis	25
3 Results and discussion	29
3.1 Identification of nine CDPSs for the biosynthesis of tryptophan-containing cyclodipeptides from <i>Streptomyces</i>	29
3.2 Identification and biosynthetic study of novel C3-guaninyl indole alkaloids guanitrypmycins	33
3.3 Expanding the spectrum of cytochrome P450s by identification of two distinct dimerases in CDPS-dependent pathways	39
4 Publications	43

TABLE OF CONTENTS

4.1 Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from <i>Streptomyces</i> strains.....	43
4.2 Guanitrypmycin biosynthetic pathways imply cytochrome P450-mediated regio- and stereospecific guaninyl transfer reactions	73
4.3 Increasing cytochrome P450 enzyme diversity by identification of two distinct cyclodipeptide dimerases	175
5 Conclusions and future prospects	231
6 References.....	233
Statutory Declaration.....	243
Acknowledgements	245
Curriculum Vitae	247

List of publications

1. **Jing Liu**, Xiulan Xie and Shu-Ming Li (2020). Increasing cytochrome P450 enzyme diversity by identification of two distinct cyclodipeptide dimerases. *Chemical Communications*, 56, 11042–11045. DOI: 10.1039/D0CC04772D.
2. **Jing Liu**, Xiulan Xie, and Shu-Ming Li (2019). Guanitrypmycin biosynthetic pathways imply cytochrome P450-mediated regio- and stereospecific guaninyl transfer reactions. *Angewandte Chemie International Edition*, 58, 11534–11540. DOI: 10.1002/anie.201906891.
3. **Jing Liu**,* Huili Yu* and Shu-Ming Li (2018). Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from *Streptomyces* strains. *Applied Microbiology and Biotechnology*, 102, 4435–4444. DOI: 10.1007/s00253-018-8908-6. (*equal contribution)
Jing Liu,* Huili Yu* and Shu-Ming Li (2018). Correction to: Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from *Streptomyces* strains. *Applied Microbiology and Biotechnology*, 102, 5787–5789. (*equal contribution)

Erklärung zum Eigenanteil

Titel der Publikation und Journal incl. Jahr, Heft, Seitzahl + doi	Autoren	geschätzter Eigenanteil in %	Bitte angeben: angenommen/ eingereicht
<p>O: Originalarbeit Ü: Übersichtartikel/Review</p> <p>Increasing cytochrome P450 enzyme diversity by identification of two distinct cyclodipeptide dimerses. <i>Chemical Communications</i>, 2020, 56, 11042–11045. DOI: 10.1039/D0CC04772D Originalarbeit</p>	Jing Liu , Xiulan Xie, and Shu-Ming Li	65	angenommen
<p>Guanitrypmycin biosynthetic pathways imply cytochrome P450-mediated regio- and stereospecific guaninyl transfer reactions. <i>Angewandte Chemie International Edition</i>, 2019, 58, 11534-11540. DOI: 10.1002/anie.201906891 Originalarbeit</p>	Jing Liu , Xiulan Xie, and Shu-Ming Li	65	angenommen
<p>Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from <i>Streptomyces</i> strains. <i>Applied Microbiology and Biotechnology</i>, 2018, 102: 4435–4444. DOI: 10.1007/s00253-018-8908-6 Originalarbeit</p>	Jing Liu *, Huili Yu*, and Shu-Ming Li	40	angenommen

*: These authors contributed equally to this work.

Kandidat(in)

Unterschrift Betreuer(in)

Abbreviations

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

NP	natural product
SM	secondary metabolite
BGC	biosynthetic gene cluster
DKP	diketopiperazine
NRPS	nonribosomal peptide synthetase
CDP	cyclodipeptide
CDPS	cyclodipeptide synthase
A domain	adenylation domain
PCP domain	peptide carrier protein domain
C domain	condensation domain
<i>E. coli</i>	<i>Escherichia coli</i>
HPLC	high performance liquid chromatography
LC-MS	liquid chromatography–mass spectrometry
bp	base pair
HR-EIMS	high-resolution electron ionization mass spectrometry
NMR	nuclear magnetic resonance
[M+H] ⁺	molecular ion plus proton
GC content	guanine-cytosine content
aa	amino acid
aa-tRNA	aminoacyl-tRNA
P450	cytochrome P450
gDNA	genomic DNA
LB	Luria-Bertani
NRRL	ARS Culture Collection
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH
PCR	polymerase chain reaction
nm	nanometer
OD ₆₀₀	optical density at 600 nm
IPTG	Isopropyl β-D-thiogalactopyranoside

ABBREVIATIONS

rpm	revolutions per minute
EIC	extracted ion chromatography
DMSO- <i>d</i> ₆	deuterated dimethyl sulfoxide
CDCl ₃	deuterated chloroform
CD ₃ OD	deuterated methanol
cWX	cyclo-L-Trp-Xaa
cWW	<i>cyclo</i> -L-Trp-L-Trp
cWA	<i>cyclo</i> -L-Trp-L-Ala
cWL	<i>cyclo</i> -L-Trp-L-Leu
cWP	<i>cyclo</i> -L-Trp-L-Pro
cWY	<i>cyclo</i> -L-Trp-L-Tyr
cWF	<i>cyclo</i> -L-Trp-L-Phe
cWM	<i>cyclo</i> -L-Trp-L-Met
cWV	<i>cyclo</i> -L-Trp-L-Val
cWG	<i>cyclo</i> -L-Trp-Gly
cFL	<i>cyclo</i> -L-Phe-L-Leu
cLL	<i>cyclo</i> -L-Leu-L-Leu
cIL	<i>cyclo</i> -L-Ile-L-Leu
cYY	<i>cyclo</i> -L-Tyr-L-Tyr
cFF	<i>cyclo</i> -L-Phe-L-Phe
w/v	weight per volume
δ_H	chemical shift of ¹ H
s	singlet
d	doublet
t	triplet
dd	double doublet
ddd	double double doublet
m	multiplet
dt	double triplet
dq	double quartet
<i>J</i>	coupling constant
multi	multiplicity

ABBREVIATIONS

Hz	hertz
ppm	parts per million
MHz	mega hertz
<i>m/z</i>	mass-to-charge ration
PT	prenyltransferase
MT	methyltransferase
CDO	cyclodipeptide oxidase
D ₂ O	deuterium oxide
<i>S. coelicolor</i>	<i>Streptomyces coelicolor</i>
¹³ C APT	¹³ C attached proton test
COSY	correlation spectroscopy
HSQC	heteronuclear single quantum coherence
HMBC	heteronuclear multiple bond correlation
NOESY	nuclear overhauser enhancement spectroscopy
NOE	nuclear overhauser effect
CD	circular dichroism
UV	ultraviolet
UV-Vis	ultraviolet-visible
NADPH	reduced nicotinamide adenine dinucleotide phosphate
NADH	nicotinamide adenine dinucleotide
Cpd 0	compound 0
Cpd I	compound I
Cpd II	compound II
Ni-NTA	nickel-nitrilotriacetic acid
His	histidine
kDa	kilo Dalton
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
CO	carbon monoxide
Tris	tris(hydroxymethyl)aminomethane
kb	kilo base pairs
<i>S. albus</i>	<i>Streptomyces albus</i>
<i>k</i> _{cat}	turnover number

ABBREVIATIONS

K_M	Michaelis-Menten constant
mAU	milli absorbance unit
v/v	volume per volume
WT	wild-type
δ_C	chemical shift of ^{13}C
<i>i.e.</i>	that is
<i>e.g.</i>	for example

Summary

Natural products derived from cyclodipeptides (CDPs) with a 2,5-diketopiperazine (DKP) skeleton comprise an important class of secondary metabolites, especially indole alkaloids derived from tryptophan-containing CDPs, which are widespread in fungi, bacteria, and plants. They play vital roles in drug discovery and development owing to their significant biological and pharmacological activities. In nature, the DKP cores can be generated by two distinct enzyme groups, that is, the nonribosomal peptide synthetases (NRPSs) and the aminoacyl tRNA-dependent cyclodipeptide synthases (CDPSs). Afterwards, different types of tailoring enzymes, such as cytochrome P450s, FAD-dependent oxidoreductases, cyclodipeptide oxidases (CDOs), prenyltransferases (PTs), and methyltransferases (MTs) are involved in installing a number of functional groups to the DKP scaffolds, thus generating various chemical structures. Although CDPSs belong to a newly defined family of enzymes, a large set of CDPSs have been identified. Among them, only several CDPS-associated biosynthetic pathways have been functionally characterized. In recent years, huge amounts of microbial genome sequences have been released in public databases and revealed numerous silent or cryptic gene clusters hiding in their genomes, including those for 2,5-DKPs, indicating great potential for discovery of novel metabolites. Therefore, full exploration of these untapped gene clusters could be a promising way to expand the chemical range of 2,5-DKPs accessible to the medical industry in the future.

In the first project, in cooperation with Dr. Huili Yu, eleven CDPSs from *Streptomyces* strains were selected for investigation based on phylogenetic analysis. Their functions were characterized *via* heterologous expression in *Escherichia coli*. The coding sequences of these CDPSs were individually cloned into pET28a (+) vector and overexpressed in solubl21 host. The fermentation cultures of generated transformants were then analyzed by LC-MS. Combined with structural elucidation of accumulated products by NMR analysis, nine CDPSs for the assembly of tryptophan-containing cyclodipeptides (cWXs) were identified. Therefore, these nine CDP synthases represented new members of the CDPS family that are responsible for cWX biosynthesis. Among them, there is one *cyclo*-L-Trp-L-Leu synthase, two *cyclo*-L-Trp-L-Pro synthases, and three *cyclo*-L-Trp-L-Trp synthases, as well as three unspecific CDPSs producing up to seven products with *cyclo*-L-Trp-L-Ala or *cyclo*-L-Trp-L-Tyr as the major product. Under optimized cultivation conditions, total product yields of generated CDPs in the *E. coli* supernatants reached 46 to 211 mg/L. In recent years, tryptophan-containing DKPs have received increasing attention due to their promising scaffolds for structural modification. Therefore, our study provides a valid experimental basis for further combination of these CDPSs with other tailoring enzymes to generate more interesting chemical entities in the field of synthetic biology.

Afterwards, sequence analysis revealed that eight of nine cWX synthase genes identified in the first project are surrounded by a putative cytochrome P450 gene. Among them, two CDPS genes, *gutA*₂₄₃₀₉ from *Streptomyces monomycini* NRRL B-24309 and *gutA*₃₅₈₉ from *Streptomyces varsoviensis* NRRL

SUMMARY

B-3589, are located in the similar gene loci containing four additional genes coding for three modification enzymes, *i.e.*, CDO, cytochrome P450, and MT. Heterologous expression of these two *p450*-associated *cdps*-containing gene clusters in *Streptomyces coelicolor* led to the identification of eight rare and novel C3-guaninyl indole alkaloids, named guanitrypmycins. Expression of different gene combinations and precursor feeding experiments proved the biosynthetic steps of guanitrypmycins. The CDP skeletons, *cyclo*-L-Trp-L-Phe and *cyclo*-L-Trp-L-Tyr assembled by the CDPS GutA, will be dehydrogenated merely at the phenylalanyl/tyrosyl side by the CDO Gut(BC) and subsequently connected with a guanine moiety by the P450 GutD. Furthermore, the MT GutE governs the last modification step to transfer a methyl group to N9' of the guaninyl residue. Moreover, the non-enzymatic epimerization of the enzymatic pathway products *via* keto–enol tautomerism increases the structural diversity of guanitrypmycins. In addition, biochemical characterization further confirmed that the P450 enzyme GutD functions as the key biocatalyst and catalyzes the unprecedented regio- and stereospecific 3 α -guaninylation at the indole ring of the tryptophanyl moiety. Therefore, this study highlights the promise of CDPS-containing pathways as sources of novel biosynthetic transformations and natural products.

In analogy, two *cdps-p450*-containing operons were identified in *Saccharopolyspora antimicrobica* *via* genome mining. Heterologous expression, biochemical characterization, together with structural elucidation proved that the two P450 enzymes TtpB1 and TtpB2 catalyze distinct regio- and stereospecific dimerizations of *cyclo*-L-Trp-L-Trp, which are differing from those previously reported in bacteria. TtpB1 represents the first bacterial P450 that catalyzes the stereospecific C3 (sp³)–C3' (sp³) bond formation between two monomers, both from the opposite side of H-11/H-11', while TtpB2 is characterized as the first P450 to mainly catalyze the unusual linkage between C3 (sp³) of a hexahydropyrroloindole unit and N1' of the tryptophanyl moiety of the second monomer from the H-11 side. Thus, our finding significantly increases the repertoire of DKP-tailoring enzymes. Additionally, in comparison with chemical synthesis, this study provides a simple, direct, and efficient approach for enzymatic one-step preparation of structurally complex DKP dimers.

Zusammenfassung

Naturstoffe aus Cyclodipeptiden (CDPs) mit einem 2,5-Diketopiperazin (DKP) als Grundgerüst sind wichtige Vertreter von Sekundärmetaboliten. Insbesondere Indolalkaloide, die sich von tryptophanhaltigen CDPs ableiten, sind weit verbreitet in Bakterien, Pilzen und Pflanzen. Sie spielen aufgrund ihrer vielseitigen biologischen und pharmakologischen Aktivitäten eine entscheidende Rolle bei der Entwicklung von Arzneistoffen. In der Natur wird das DKP von zwei verschiedenen Enzymfamilien gebildet, entweder von nichtribosomalen Peptidsynthetasen oder von aminoacyl-tRNA-abhängigen Cyclodipeptidsynthetasen (CDPSs). Anschließend können diverse Enzyme, z.B. FAD-abhängige Oxidoreduktasen, Cytochrom P450, Cyclodipeptidoxidasen (CDOs), Methyltransferasen oder Prenyltransferasen das DKP-Grundgerüst durch Einführung verschiedener funktioneller Gruppen modifizieren. Hierdurch entsteht eine Vielzahl an chemischen Verbindungen. Obwohl CDPSs eine erst kürzlich entdeckte Enzymfamilie darstellen, sind bereits viele unterschiedliche Vertreter gefunden worden. Jedoch wurden nur wenige Biosynthesewege inklusive modifizierender Enzyme vollständig aufgeklärt. Seit einigen Jahren wächst die Anzahl sequenzierter Genome von Mikroorganismen in öffentlichen Datenbanken stetig, sodass etliche nicht aktive oder kryptische Gencluster identifiziert werden konnten. Diese Gencluster weisen großes Potential zur Entdeckung neuer Sekundärmetabolite einschließlich 2,5-DKPs auf. Daher stellt die vollständige Aufklärung solch unerforschter Gencluster einen vielversprechenden Ansatz für die pharmazeutische Industrie dar, um das Spektrum an 2,5-DKPs zu erweitern.

In meinem ersten Projekt, in Zusammenarbeit mit Dr. Huili Yu, haben wir auf Basis phylogenetischer Analysen elf CDPSs aus verschiedenen *Streptomyces*-Stämmen zur genaueren Untersuchung ausgewählt. Ihre Funktion konnte durch heterologe Expression in *Escherichia coli* aufgeklärt werden. Hierzu wurden die für die CDPSs kodierenden Genabschnitte einzeln in den pET28a (+)-Vektor kloniert und anschließend im *E. coli* solubl21 Host überexprimiert. Danach wurden die Transformanten in Flüssigmedium kultiviert, extrahiert und deren Produkte per LC-MS und NMR-Analyse strukturell aufgeklärt. Dabei haben wir neun CDPSs für die Synthese tryptophanhaltiger Cyclodipeptide (cWXs) identifiziert. Dementsprechend handelt es sich bei diesen neun CDPSs um neue Mitglieder der CDPS-Familie, die die Biosynthese tryptophanhaltiger CDPs katalysieren. Genauer gesagt handelt es sich um eine *cyclo*-L-Trp-L-Leu-Synthase, zwei *cyclo*-L-Trp-L-Pro-Synthasen, drei *cyclo*-L-Trp-L-Trp-Synthasen sowie drei unspezifische CDPSs, die als Hauptprodukt *cyclo*-L-Trp-L-Ala oder *cyclo*-L-Trp-L-Phe bilden. Unter optimierten Kultivierungsbedingungen konnten wir Ausbeuten von 46 - 211 mg/L erzielen. In letzter Zeit ist den tryptophanhaltigen DKPs vermehrte Aufmerksamkeit zuteil geworden, da ihr Grundgerüst ein vielversprechender Vorläufer für strukturelle Modifizierungen ist. Unsere Arbeit bietet daher eine valide, experimentelle Basis für die synthetische Biologie, um durch Kombinationen mit anderen nachgeschalteten Enzymen weitere interessante Verbindungen zu generieren.

Nach erfolgreichem Abschluss des ersten Projekts konnte ich durch weitergehende Sequenzanalysen der für cWX Synthasen kodierenden Gene bei acht von neun Genen jeweils ein Gen in unmittelbarer Nachbarschaft finden, welches mutmaßlich für Cytochrom P450-Enzyme kodiert. Hierbei habe ich mich auf zwei Gencluster aus *Streptomyces monomycini* NRRL B-24309 bzw. *Streptomyces varsoviensis* NRRL B-3589 konzentriert. Die beiden CDPS-Gene *gutA*₂₄₃₀₉ bzw. *gutA*₃₅₈₉ sind von vier weiteren Genen, die für drei Enzyme kodieren, umgeben – nämlich einer Cyclodipeptidoxidase, einem Cytochrom P450 und einer Methyltransferase. Nach heterologer Expression dieser biosynthetischen Gencluster in *Streptomyces coelicolor* konnte ich acht außergewöhnliche und neue Guanitrypmycine strukturell aufklären, welche an Position C3 des Tryptophanrestes guanyliert sind. Durch unterschiedliche Kombinationen der einzelnen Gene sowie Zufütterungsversuche konnte ich die einzelnen Schritte der Guanitrypmycin-Biosynthese entschlüsseln. Das von der CDPS GutA gebildete CDP wird im ersten Schritt durch die CDO Gut(BC) am Phenylalanin- bzw. am Tyrosinrest oxidiert. Danach überträgt das P450 GutD ein Guanin auf den Tryptophanrest und anschließend wird der Stickstoff an Position 9' des Guaninrestes durch die Methyltransferase GutE methyliert. Darüber hinaus wird das Spektrum der Guanitrypmycine durch nicht-enzymatische Epimerisierung in Form einer Keto-Enol-Tautomerie erweitert. Anhand weiterführender biochemischer Untersuchungen konnte die Funktion des Cytochrom P450 GutD als Schlüsselkatalysator dieser bisher beispiellosen, regio- und stereoselektiven 3 α -Guanylierung des Tryptophanrestes bestätigt werden. Diese Arbeit unterstreicht, wie aussichtsreich CDPSs beinhaltende Biosynthesewege als Quelle neuer biochemischer Reaktionen und der damit verbundenen Entdeckung unbekannter Naturstoffe sind.

In gleicher Herangehensweise habe ich zwei weitere CDPSs in *Saccharopolyspora antimicrobica* mittels Genomanalyse entdeckt, die mit P450s zusammen lokalisiert sind. Durch heterologe Expression, biochemische Untersuchung und LC-MS und NMR-gestützter Strukturaufklärung konnte ich beweisen, dass die zwei P450s TtpB1 und TtpB2 die regio- und stereospezifische Dimerisierung zweier *cyclo*-L-Trp-L-Trp katalysieren. Diese Art der Dimerisierung war zuvor nicht in Bakterien bekannt. TtpB1 ist das erste identifizierte P450, welches eine stereospezifische C3 (sp³)-C3' (sp³) Bindung verknüpft, jeweils gegenüber zu H-11 bzw. H-11'. TtpB2 ist hingegen das erste P450, welches eine ungewöhnliche Bindung zwischen dem C3 (sp³) der Hexahydro-Pyrroloindoleinheit und dem N1' des Tryptophanrestes einfügt. Hiermit konnte ich das Repertoire an DKP-modifizierenden Enzymen signifikant erweitern und darüber hinaus einen einfachen, direkten und effizienten Weg zur enzymatisch kontrollierten Synthese für synthetisch schwer zugängliche DKP-Dimere aufzeigen.

1 Introduction

1.1 Actinobacteria: the most prolific source of natural products

Actinobacteria, also known as actinomycetes, are filamentous gram-positive bacteria and constitute one of the largest bacterial phyla in nature. Similar to the filamentous fungi, many actinobacteria can produce a mycelium, and many of these mycelial actinobacteria reproduce by sporulation.¹ Therefore, they were originally recognized as a transitional form between fungi and bacteria. In general, actinobacteria have a high guanine-plus-cytosine (G+C) content in their genomes (commonly ranging from 51% to over 70%).² As free-living microbes, actinobacteria are widely distributed in diverse natural habitats, including soil environments, marine and freshwater systems.³ Actinobacteria play critical roles in helping to sustain ecosystems and are implied as important contributors to the global carbon cycle through the breakdown of plant biomass into simple sugars.⁴ They are also capable of mediating community dynamics by producing a range of small molecules.⁴

On the other hand, actinobacteria have made the most significant contribution in the field of biotechnology as the versatile producers of chemically diverse and biologically active natural products (NPs) with broad applications in medicine, agriculture, and environment.⁵ They are known as the most important source of antibiotics.⁶ Notably, approximately two-thirds of clinically used antibiotics of natural origin are produced by actinobacteria, predominantly by *Streptomyces*.⁷ Different classes of actinobacteria-derived antibiotics are used as human drugs ranging from macrolides (erythromycin⁸), tetracyclines⁹, ansamycins (rifamycins¹⁰), amphenicols (chloramphenicol¹¹), aminoglycosides (kanamycin¹², streptomycin¹²), glycopeptides (vancomycin¹³, teichoplanin¹⁴), lipopeptides (daptomycin¹⁵), to oxazolidinones (cycloserine¹⁶) (**Figure 1**). In addition to these useful antibiotics, actinobacteria are also efficient producers of numerous bioactive substances with important antifungal, anticancer, insecticidal, immunosuppressive and enzyme inhibition activities, including various polyketides, peptides, alkaloids, terpenoids, and saccharides.¹⁷ As shown in **Figure 2**, nikkomycin, a uridine-based nucleoside-peptide originally isolated from *Streptomyces tendae*, is a potential antifungal agent which competitively inhibits the chitin synthesis in fungi.¹⁸ Salinosporamide A, a proteasome inhibitor from the marine actinobacterial strain *Salinispora tropica*, displays remarkable *in vitro* cytotoxicity toward a variety of cancer cell lines. Currently, it undergoes Phase II human clinical trials for the treatment of multiple myeloma.¹⁹ Avermectins are a novel class of macrocyclic lactones produced by *Streptomyces avermitilis*, which are generally used as pesticides for the treatment of insect pests and parasitic worms.²⁰ Rapamycin has immunosuppressive properties by targeting mTORC1, a ubiquitous kinase-containing complex promoting cell growth and proliferation.²¹

Nowadays, there is urgent need for new antibiotics and bioactive metabolites due to the emergence of multidrug-resistant pathogens.²² It is believed that chemical diversity from actinobacterial natural products will continue to play important roles in the future drug discovery and development.²³

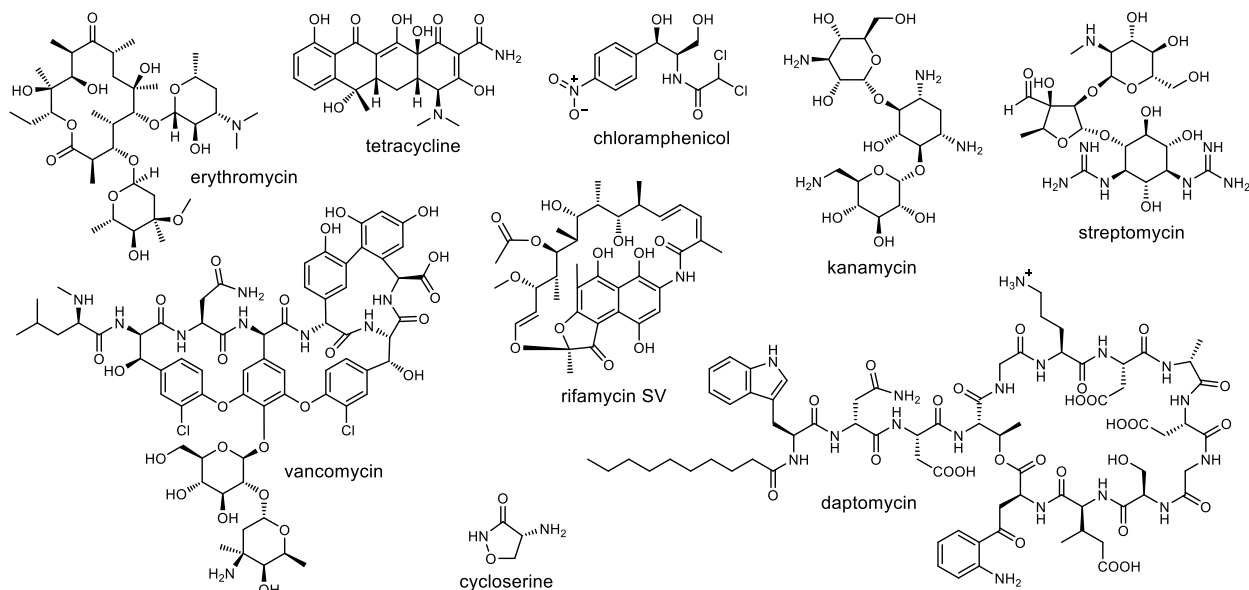


Figure 1 Examples of clinically used antibiotics originated from actinobacteria.

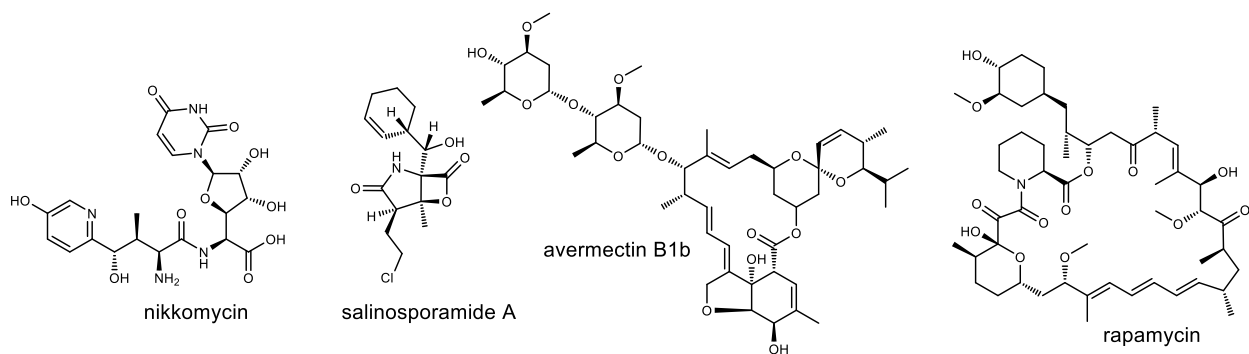


Figure 2 Examples of the bioactive substances isolated from actinobacteria.

1.2 Genome mining for novel secondary metabolites from actinobacteria

The discovery of streptomycin from *Streptomyces griseus* has triggered the search for new antibiotics from actinomycetes.²⁴ Since then, multiple new classes of antibiotics were discovered and introduced over the following two decades, so called the “golden era” of antibiotics.²⁵ However, in the 1980s, owing to the repeated isolation of actinobacteria and the frequent rediscovery of known compounds thereof, the discovery of new compounds through the traditional bioactivity-guided screening strategy became more and more difficult.²⁶ Whereas, at the beginning of the 21st century, the rapid developments in sequencing technologies and bioinformatic approaches revealed that the ability of actinobacteria to synthesize novel bioactive natural products had been far underestimated.¹⁷ Typically, the encoding genes for the enzymes that synthesize a specific NP are located in a constitute region on the microbial chromosome, so-called biosynthetic gene cluster (BGC). Bioinformatic analysis shows that each actinobacterial genome has 20 – 40 BGCs, much more than previously isolated NPs thereof.²⁷ For instance, the model actinomycete *Streptomyces coelicolor*, which has been extensively

studied for more than 70 years, was well known to produce five compounds. However, the analysis of its complete genome sequence revealed the presence of 18 additional putative BGCs related to secondary metabolite (SM) biosynthesis. Most of them have been characterized by now.²⁸ In general, the majority of BGCs appear “silent” or express very poorly under normal laboratory conditions in their natural host.²⁹ Therefore, a range of strategies, e.g., genome mining, have been developed to identify these cryptic BGCs, thereby gaining access to novel lead compounds.³⁰ This has given birth to the new, so-called field of genomics-driven natural product discovery, which complements the traditional approaches.³¹

1.2.1 Genome mining

The term “genome mining” refers to the utilization of genomic information for the discovery of novel processes, targets, and products. It involves bioinformatic analysis and identification of unknown BGCs in the target genomes,³² sequence analysis of the encoding genes, and the experimental identification of the products synthesized by these BGCs (**Figure 3A**).³³ Therefore, the prediction and selection of uncharacterized BGCs by bioinformatic methods is always the high priority for genome mining of novel NPs.³⁴ Since the rapid development of next generation sequencing (NGS) technologies has allowed to acquire the genomic data in a faster and cheaper way, a huge number of genome sequences are deposited and accessible in the public databases.³² Currently, the basic strategy to find and identify novel BGCs is to target signature genes responsible for a specific NP biosynthesis by comparative analysis.³⁵ Several computational tools have been developed to analyze and functionally annotate the draft genome sequences. One representative tool is the widely used web software “antibiotics and Secondary Metabolite Analysis SHell” (antiSMASH).³⁶ This platform provides a rapid identification and annotation of BGCs in microbial genomes and prediction of their putative corresponding metabolic products by comparison to the known BGCs, in turn, identifying new biosynthetic pathways. “Genome neighborhood networks” (GNNs) as a high-throughput approach, facilitates the discovery of uncharacterized metabolic pathways by large-scale visualization and analysis of the genome context.³⁷ This analysis tool is the extended variants of sequence similarity networks (SNNs), thereby enabling the prediction of both the *in vitro* enzymatic activities and *in vivo* metabolic functions. For targeted genome mining, these computational tools are powerful to analyze the cryptic BGCs and give important insights into the structural features of their potential product(s).³⁸

In order to activate these intriguing and cryptic gene clusters and trigger the corresponding SM overproduction, multiple new strategies have been developed to reveal the chemical potential of actinomycetes in recent years. Generally, these strategies fall into two broad categories: pleiotropic and pathway-specific approaches.³¹ The pleiotropic approaches may influence and activate the expression of more than one gene or gene cluster. Such approaches include growth condition optimization, co-cultivation with other microorganisms sharing the common ecological niches, addition of chemical elicitors, ribosome engineering, as well as manipulation of global regulators thereby initiating global changes in the regulation (**Figure 3B**).^{29, 39-42} In contrast to pleiotropic approaches, pathway-specific strategies enable a more targeted approach to mine novel NPs with a higher degree of control and predictability.³¹ Currently, two main approaches for the induction of a candidate BGC

expression have been proven effective, *i.e.*, (i) genetic manipulation in the native host, such as overexpression of the pathway-specific activator genes, the replacement of the native promoters with strong and inducible promoters, and inactivation of the negative regulatory genes; and (ii) heterologous expression of the interested BGCs in well-studied hosts (**Figure 3C**).^{31, 43-45}

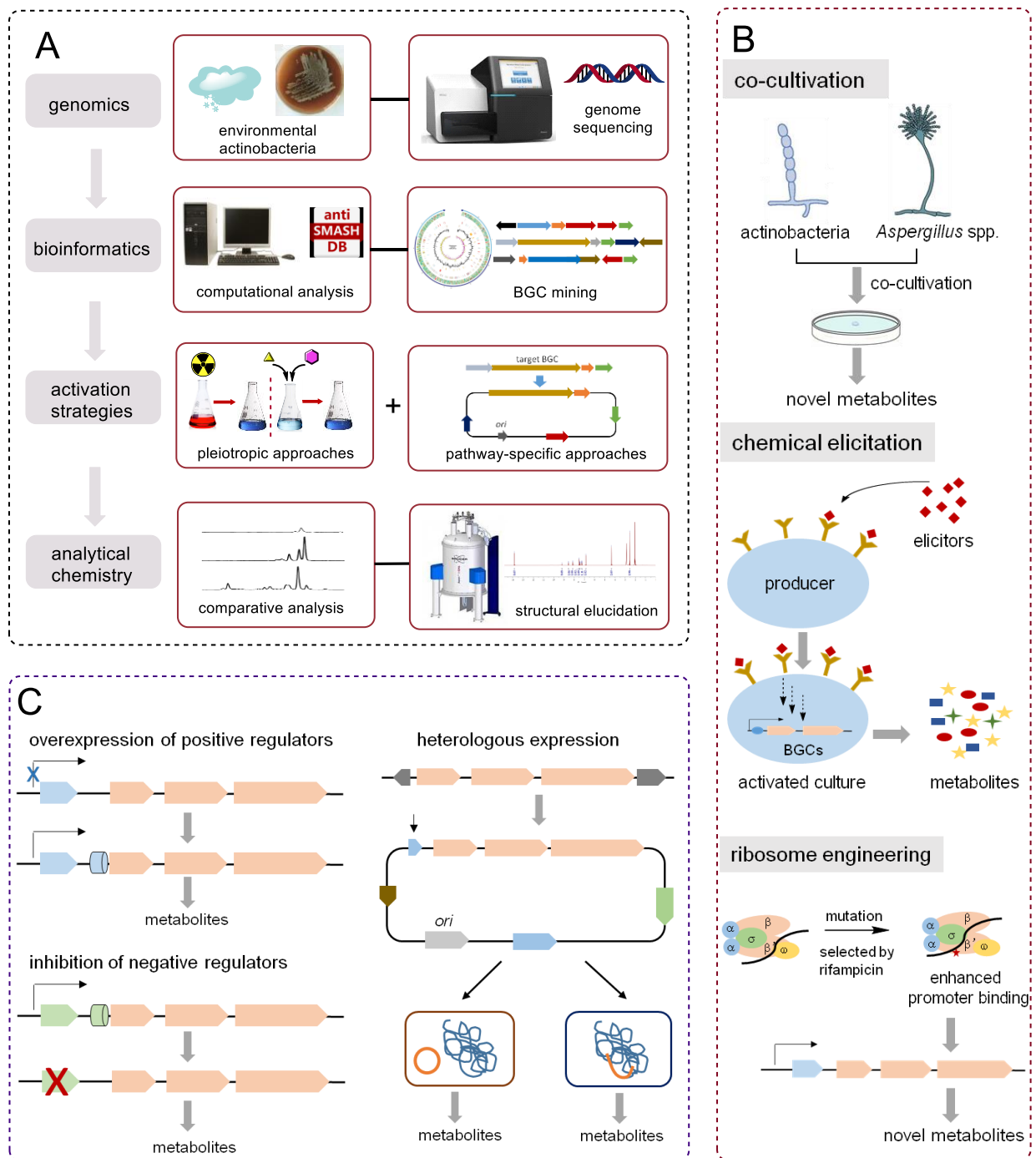


Figure 3 Genome mining approach for natural product discovery in actinobacteria.

1.2.2 Manipulation of pathway-specific regulatory genes

The biosynthesis of SMs in actinobacteria is stringently controlled by complex regulatory networks with multiple level regulators, including global regulators and pathway-specific regulators.⁴⁶⁻⁴⁸ The global regulators always regulate more than one metabolic pathways and may not directly affect any specific BGC, while pathway-specific regulators are mainly involved in the biosynthesis of a specific metabolite.⁴⁹ Comprehensive analysis revealed that there are one or more putative regulatory genes located inside each BGC, encoding so-called pathway-specific regulators (PSRs), together with the backbone and resistance genes.^{26, 32, 50} Therefore, these encoded transcription regulators, also named “cluster-situated regulators” (CSRs), act either as an activator or a repressor to affect the production of the respective metabolite by directly controlling the transcriptional initiation of a neighboring gene within the gene cluster.^{49, 51} In general, the repressors bind to the promoter region, thereby preventing the RNA polymerase from starting the gene transcription.^{52, 53} In contrast, most activators bind to the upstream region of the transcriptional promoter, helping recruit the RNA polymerase to start or accelerate the transcription.^{52, 54} Thus, overexpression of activators or deletion of repressors are the straightforward and promising approaches to activate the cryptic BGCs of interest in the genomic-driven era.

The SARP (*Streptomyces* antibiotic regulatory proteins) family and the LAL (large ATP-binding regulators of the LuxR family) family are two widely recognized families of pathway-specific activators in actinobacteria.^{55, 56} Recently, several clusters have been reported to wake up by inducing the expression of these positive regulators. A common strategy is to replace the natural promoter with a strong inducible or constitutive promoter. For example, a giant type I PKS gene cluster, spanning nearly 150 kb, was activated by insertion of the strong constitutive *ermE** promoter in front of a LAL regulator in the genome of *Streptomyces ambofaciens*, leading to the discovery of the unusual glycosylated macrolide stambomycins.⁵⁷ In contrast, the TetR (tetracycline repressor protein) family usually belong to the transcriptional repressors, which can repress the expression of their own genes as well as that of other genes.^{29, 58, 59} Knocking out *calR3*, encoding for a TetR family repressor in *Streptomyces chartreusis* NRRL 3882, led to the significantly increased production of calcimycin and its intermediate cezomycin.⁶⁰

Very recently, novel approaches, e.g., the CRISPR-Cas (Clustered regularly interspaced short palindromic repeat-CRISPR-associated protein) dependent genome editing system, have been developed to efficiently manipulate the regulatory genes.⁶¹ Nevertheless, in most cases, the complex mechanisms of regulatory networks in actinobacteria are still poorly explored.⁴⁶ Hence, a better understanding of their regulatory networks in the future would greatly promote the discovery of new actinobacterial metabolites.

1.2.3 Heterologous expression

Heterologous expression in well-characterized hosts has been proven to be a powerful approach to activate BGCs and discover the corresponding products.^{62, 63} Principally, this approach relies on efficient cloning of intact BGCs into suitable expression vectors.⁶⁴ The typical workflow for

heterologous expression of BGCs derived from actinobacteria is shown in **Figure 4**. After DNA isolation from the native producers, BGCs of interest are usually captured from genomic DNA libraries, which are constructed by using cosmid, fosmid, bacterial artificial chromosome (BAC), or phage artificial chromosome (PAC) vectors.^{62, 65-69} However, this method always requires highly efficient screening assays afterwards. Alternatively, the desired BGCs can be directly cloned from the chromosomal DNA of the actinomycetes. Direct Pathway Cloning (DiPaC) is an efficient strategy for cloning small- and mid-size BGCs by using the long-amplification polymerase chain reaction (PCR) to obtain the linear DNA fragments, followed by *in vitro* DNA assembly with the vector *via* sequence- and ligation-independent cloning (SLIC) or Gibson assembly.^{70, 71} In addition, several methods based on *in vivo* homologous recombination have been developed for direct cloning of large-size BGCs, such as linear-linear homologous recombination (LLHR), linear-circular homologous recombination (LCHR) in *E. coli*, exonuclease combined with RecET recombination (ExoCET), Cas9-assisted targeting of chromosome segments (CATCH), as well as transformation-associated recombination (TAR) cloning in *Saccharomyces cerevisiae*.⁷²⁻⁷⁷ These powerful cloning methods have highly accelerated the process of NP discovery and greatly contributed to the research on NP biosynthesis.

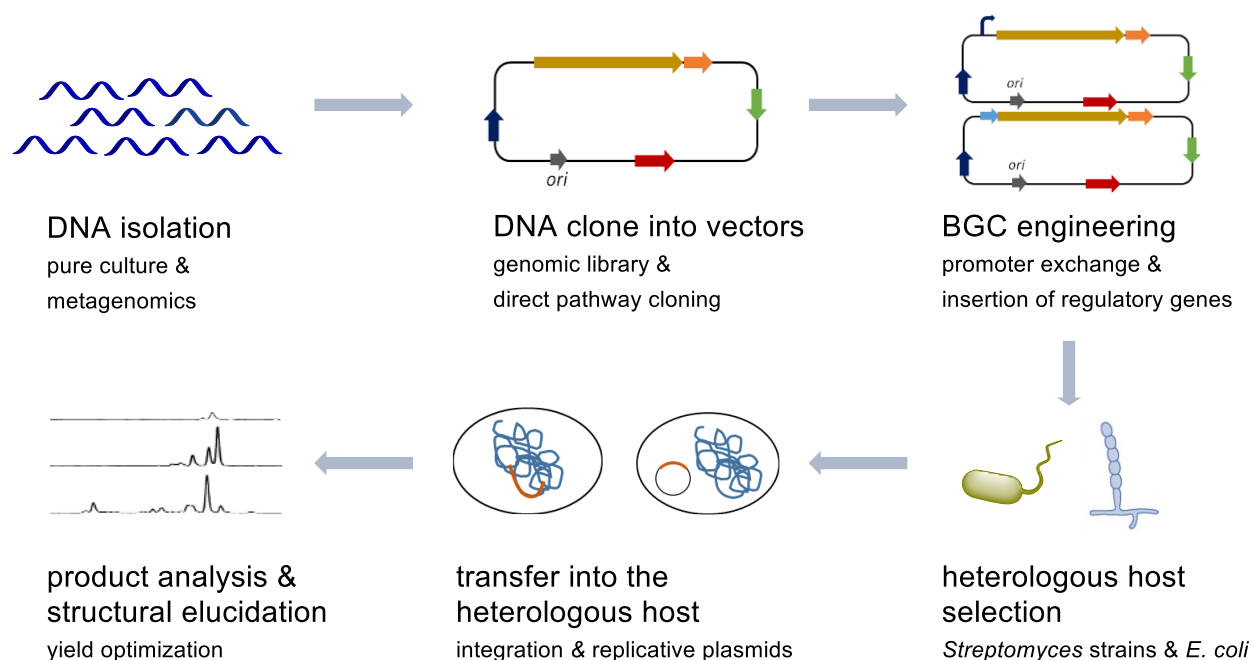


Figure 4 Typical workflow for heterologous expression of gene clusters from actinobacteria.

Once the candidate BGC is determined and cloned, it needs to be transferred into the heterologous host. Selection of a reliable host is an essential component to ensure the robust expression of BGCs. Generally, the optimized host should have an abundance of natural product precursors and a promiscuous transcriptional machinery for the efficient transcription of various BGCs, *e.g.*, regulatory elements, promoters, and ribosome binding sites.^{78, 79} Moreover, a suitable host organism should possess the following features: simplicity of use, excellent growth characteristics, and a plethora of readily accessible genetic tools.⁸⁰ For efficient expression of BGCs derived from actinobacteria,

several optimized *Streptomyces* strains have been developed as hosts. *Streptomyces coelicolor*, *Streptomyces lividans* and *Streptomyces albus* are the most widely used laboratory strains.⁸¹⁻⁸⁴ They are closely related species and exhibit a similar codon usage with the BGC's native holders, therefore they could provide a higher translational efficiency.⁸⁵ Additionally, these strains have a relative clear background by removing endogenous genes, thereby providing a better detection of new produced compounds.⁸⁶ *E. coli* could also serve as a viable option for heterologous expression with a rapid growth rate and easier genetic manipulation.⁸⁷ The utility of *E. coli* as hosts for overproduction of many biological products is also established.^{88, 89} Nevertheless, due to lack of some necessary biosynthetic components or substrates, some important modifications and improvements are usually required in *E. coli* strains before heterologous expression.⁹⁰

Once transferred, the stability of the exogenous BGCs inside the heterologous host is crucial for the sustainable accumulation of the product(s).⁷⁸ Basically, the foreign BGCs can be either kept on a stable host-compatible and replicative plasmid or integrated into the chromosome. For plasmid-based overexpression, the intact BGCs can be maintained in a single plasmid or multiple plasmids with individually coding regions *in cis*.⁷⁸ In comparison, the genomic integration of a target BGC within the heterologous host is a much more stable option for successful expression. In *Streptomyces* hosts, site-specific recombination is the widespread strategy to achieve chromosomal integration.⁹¹ Afterwards, the new metabolites can be identified by comparing the changes of the metabolite profiles, usually *via* high-performance liquid chromatography (HPLC). Furthermore, large-scale fermentation and chemical extraction of cultures lead to isolation and subsequent structural elucidation of the new compounds by 1D and 2D NMR spectroscopy. Sometimes, even though the expression of BGCs is successful, the yields may remain stuck at low level. Given the requirement for high titers of production, some additional approaches, such as refactoring of essential clusters with promoter exchange and insertion of regulatory regions, are performed to modify the expression constructs.^{31, 78}

Very recently, dozens of cryptic BGCs from actinobacteria have been successfully overexpressed in different systems, thereby diverse novel compounds with potential bioactivities were obtained.⁶² Metagenomics reveals that uncultivated organisms including actinobacteria could be a rich source of cryptic BGCs for special metabolites.⁹² As a result, a big challenge for the future is how to exploit these untapped potentials to gain more bioactive NPs. It is expected that innovative heterologous expression systems and novel genome mining strategies, combined with synthetic biotechnology can promote the discovery of novel scaffolds from diverse natural resources and broaden our knowledge about new biosynthetic mechanisms.^{31, 62}

1.3 2,5-Diketopiperazines

2,5-diketopiperazines (2,5-DKPs), the smallest class of cyclic peptides, are achieved by the condensation of two α -amino acids. They are heterocyclic compounds and characterized by a central diketopiperazine (DKP) ring. The general core of 2,5-DKPs is shown in **Figure 5A**. Substitution of side chain groups R_1 and R_2 with different types of amino acids will generate the simplest cyclodipeptides (CDPs). Their central scaffold, the six membered ring, can be then modified by various substitutions and different stereochemistry.

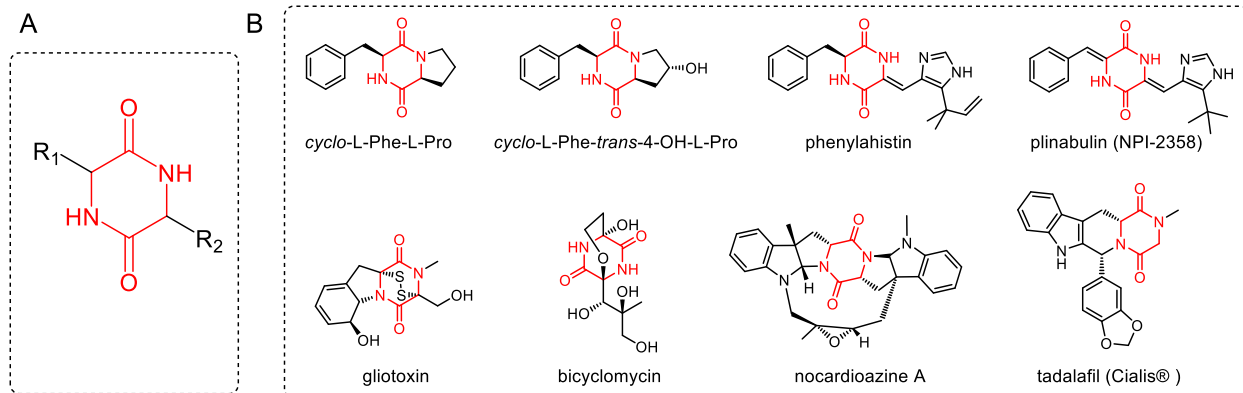


Figure 5 General structure of 2,5-DKPs (A) and examples of bioactive DKPs and derivatives (B)

2,5-DKPs are ubiquitous in nature and often found as side products of polypeptides, especially during the production process of food and beverages.⁹³ They also constitute a large class of SMs obtained from fungi, bacteria, plants, and mammals.⁹⁴ Despite the increasing number of isolated and characterized DKPs in recent years, their physiological roles in the producing organisms have yet to be investigated. In some cases, they are reported to be capable of activating or inhibiting LuxR-type proteins in N-acylated L-homoserine lactone (AHL) biosensor strains, thus suggesting that they are potentially involved in the biochemical communication as a new class of Quorum sensing (QS) molecules for interspecies or even interkingdom signaling.⁹⁵

In recent years, CDPs and their derivatives DKPs have attracted an increasing interest due to their important and diverse biological and potential pharmacological properties, including antibacterial, antifungal, antiviral, antitumor, and immunosuppressive effects.⁹⁶ Representatives of bioactive DKPs are shown in **Figure 5B**. *cyclo*-L-Phe-L-Pro and *cyclo*-L-Phe-*trans*-4-OH-L-Pro show antifungal activities.⁹⁷ Phenylahistin, produced by *Aspergillus ustus*, shows an inhibitory effect on the cell cycle progression.⁹⁸ Plinabulin (BPI-2358), the derivative of phenylahistin, is in a world-wide Phase III clinical trial for non-small cell lung cancer treatment. Gliotoxin is an immunosuppressive cytotoxin, which has been isolated from several fungal species, such as *Aspergillus fumigatus*.⁹⁹ Bicyclomycin exhibits activities against a broad spectrum of gram-negative bacteria and is currently used as a commercial antibiotic to treat diarrhea in Japan.¹⁰⁰ Nocardioazine A, originally isolated from a marine-derived actinobacterial strain, shows inhibitory activity against the membrane protein efflux pump P-glycoprotein, which is overexpressed in many multidrug resistant cancer cells.¹⁰¹ Tadalafil, achieved *via* chemical synthesis, is used to treat male erectile dysfunction.¹⁰²

The DKP scaffolds can be easily obtained from α -amino acids by conventional methodology.¹⁰³ In recent years, the synthesis of 2,5-DKPs *via* solid-phase intramolecular cyclization has been the most utilized method, which is useful for the construction of chemical libraries for drug lead discovery.¹⁰⁴ In nature, the 2,5-DKP scaffolds are synthesized by two different types of enzymes, the nonribosomal peptide synthetases (NRPSs) and the cyclodipeptide synthases (CDPSs). Furthermore, the tailoring enzymes introduce specific modifications to the DKP cores and (or) the side chains to generate more complex DKP-containing NPs.

1.3.1 DKP scaffolds from NRPS pathways

So far, most of the DKP-scaffold containing NPs are biosynthesized by the dedicated NRPSs.¹⁰⁵ A large number of DKP derivatives from different families, including indole diketopiperazine alkaloids (indole DKPs, such as acetylaszonalenin¹⁰⁶ and fumitremorgin B¹⁰⁷), ergot alkaloids (such as ergotamine¹⁰⁸), epidithiodioxopiperazines (ETPs, such as gliotoxin¹⁰⁹ and sirodesmin¹¹⁰), dimerumic acid (such as erythrochelin¹¹¹ and coprogens¹¹²) and thaxtomin¹¹³, are products of NRPS-dependent pathways (**Figure 6**).

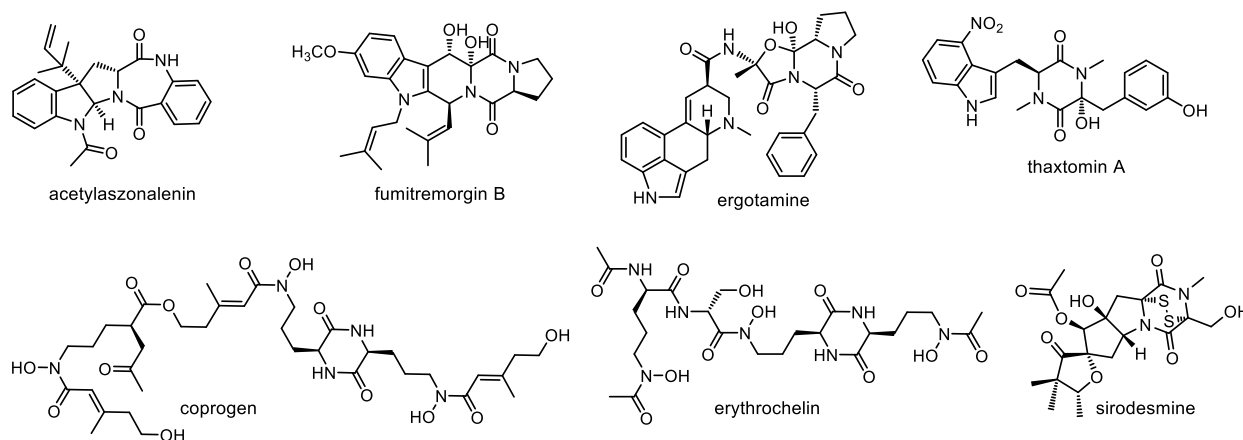


Figure 6 Examples of DKP derivatives synthesized by NRPS pathways.

NRPSs are long-studied, modular multidomain enzyme complexes and widely spread in fungi and bacteria. A typical NRPS module comprises three essential domains: the adenylation (A) domain, the peptidyl carrier protein (PCP) domain (also called thiolation (T) domain) and the condensation (C) domain, for the incorporation of one amino acid to the peptide product (**Figure 7A**).¹¹⁴ The initial step of the biosynthesis of nonribosomal peptides is performed by the A domain, which is responsible for the substrate recognition and activation *via* adenylation, thus resulting in an aminoacyl-AMP intermediate. The activated intermediate is then transferred onto the 4'-phosphopantetheine (PP) arm of the PCP domain *via* a thioester linkage. The C domain catalyzes the peptide bond formation between the current amino acyl unit and the peptidyl chain that is tethered to the adjacent upstream PCP domain, hereby elongating the growing peptide chain. The chain elongation is terminated by the action of the fourth essential NRPS catalytic unit, the thioesterase (TE) domain, which catalyzes the product release by either hydrolysis or macrocyclization.¹¹⁵ In this assemble line, the number and the order of modules in the NRPSs matches those of amino acids in the final peptide product.¹¹⁶ In addition to the proteinogenic amino acids, NRPSs can also catalyze the incorporation of a broad range of non-proteinogenic amino acid building blocks, e.g., D-amino acids and α -hydroxy acids, to generate complex scaffold architectures.¹¹⁷

For the synthesis of DKP scaffolds, the dedicated NRPSs lack the C-terminal TE domain, which is replaced by the condensation-like (C_T) domain (**Figure 7B**). It is suggested that the terminal C_T domain could be responsible for the cyclization of the dipeptide to form the diketopiperazine structure, thereby

releasing the peptide from NRPS assemble line.¹¹⁸ CDPs synthesized by NRPSs can be further modified by the tailoring enzymes, usually encoded by genes that are clustered with the NRPS genes, to generate the final diketopiperazine metabolites.¹¹⁹ Various chemical transformations ranging from prenylation to dehydrogenation, hydroxylation or methylation have been found in NRPS pathways to modify the DKP skeletons.⁹⁶

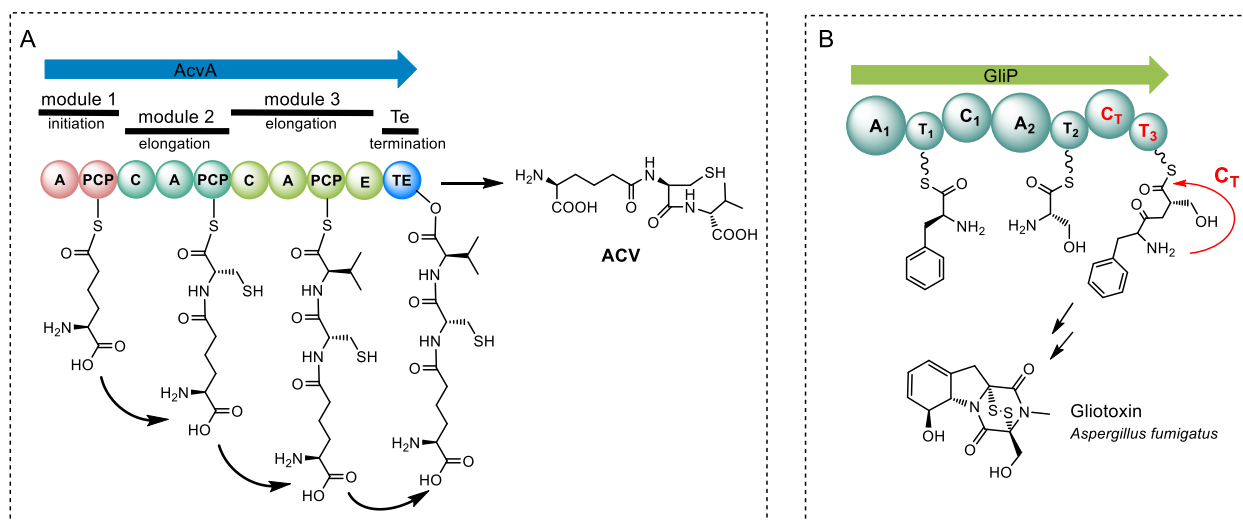


Figure 7 Examples of NRPS assembly lines.

In a few cases, the DKP scaffolds were generated through the premature release of NRPS-tethered dipeptidyl-thioester intermediates during the chain elongation process. For instance, *cyclo*-(D-Phe-L-Pro) was observed as a dipeptide intermediate during the biosynthesis of gramicidin S and tyrocidine A in *Bacillus brevis*.¹²⁰ Cyclomarazines A and B were characterized during the biosynthesis of cyclomarin in *Salinispora arenicola* CNS-205.¹²¹ It was proposed that the formation of the truncated cyclomarazine DKPs is catalyzed by the type II TE, which hydrolyze the incompletely processed dipeptide, or *via* a nonenzymatic process because of the ineffective catalysis by the third module of the NRPS CymA.¹²¹

1.3.2 DKP scaffolds from CDPs pathways

In contrast to NRPSs, CDPs are a newly defined family of enzymes that directly hijack the aminoacyl-tRNAs (aa-tRNAs) from the primary metabolism as substrates to form the DKP scaffolds. The first member of this family, AlbC, which catalyzes the formation of *cyclo*-L-Phe-L-Leu, was characterized in *Streptomyces noursei* in 2002.¹²² To date, more than 120 CDPs have been characterized, including more than 40 members in the past two years.¹²³ Over 75 different cyclodipeptides have been assembled by CDPs, consisting of 18 of the 20 proteinogenic amino acids. Very recently, CDPs have been also demonstrated to incorporate non-canonical amino acids (ncAAs) to produce non-canonical 2,5-DKPs. Compared with NRPSs, CDPs are small enzymes (~30 kDa) typically with 200 - 300 amino acid residues.¹²⁴ The majority of CDPs have been identified from bacteria, especially from actinobacteria.¹²³ Only a few cases were distributed in eukaryotes with one enzyme from the sea

anemone *Nematostella vectensis*.¹²⁵

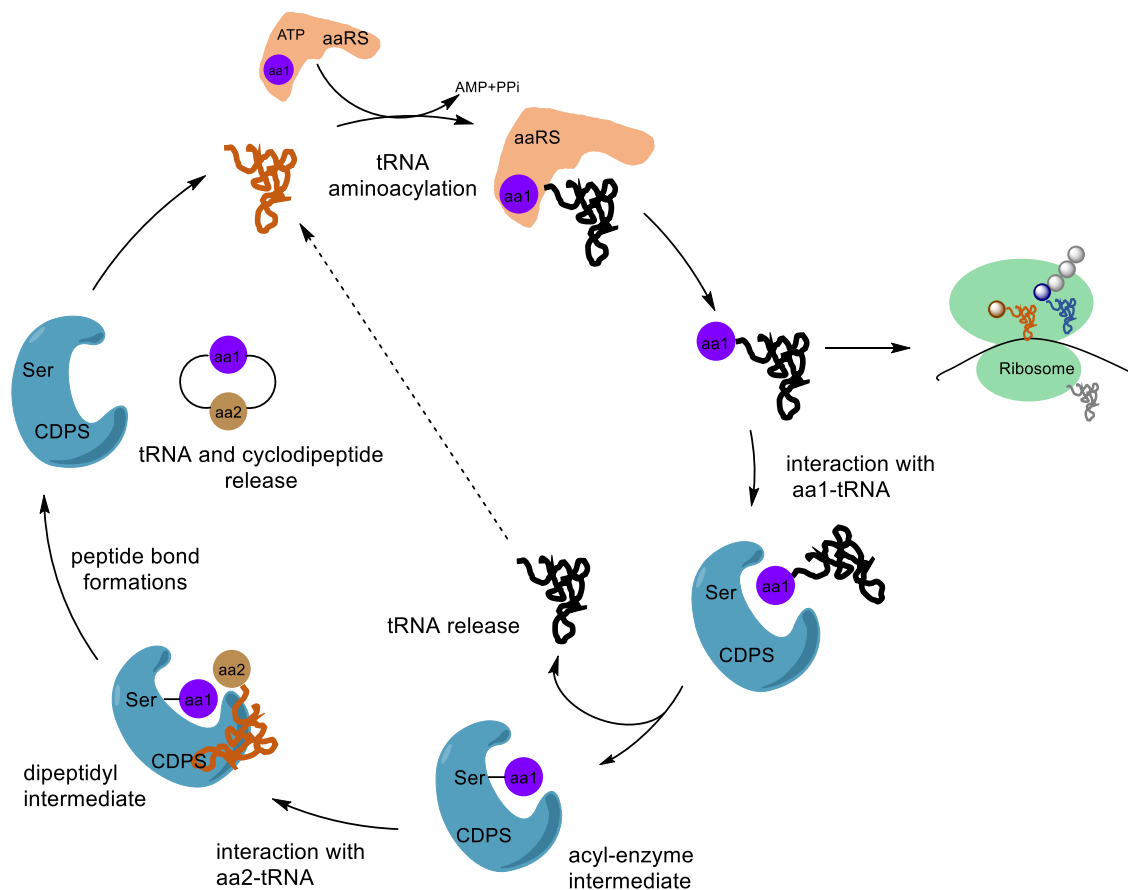


Figure 8 The proposed catalytic mechanism of CDPSs for cyclodipeptide biosynthesis.⁹⁶

The available crystallographic structures of three CDPSs, AlbC (PDB 3OQV), Rv2275 (PDB 2X9Q) and YmC (PDB 3OQH), provide insights into the catalytic mechanism of the CDP formation.¹²⁶⁻¹²⁸ Their monomeric protein possesses a common compact α/β fold and a conserved Rossmann-fold domain.¹²⁸ Although displaying only about 15% sequence similarities, the mentioned three CDPSs share high degree of structural similarity to the catalytic domains of class-1c aminoacyl-tRNA synthetases (aaRSs), *i.e.*, the Rossmann-fold subdomain and a helical connective polypeptide 1 (CP1) subdomain.⁹⁶ In addition, all CDPSs possess two surface-accessible pockets for the substrate selection and catalysis: pocket 1 (P1), which is corresponding to the aminoacyl binding pocket in class-1c aaRSs, and pocket 2 (P2), which is missing in the aaRSs.¹²³ It was proposed that CDPSs use a sequential ping-pong mechanism to achieve the synthesis of cyclodipeptides (**Figure 8**).¹²⁹ Following the recognition of the first substrate, the catalytic step initial with the binding of the first aa-tRNA to the CDPSs and the subsequent transfer of the aminoacyl group to the conserved serine residue of P1 to form an acyl-enzyme intermediate. Then, the resulted intermediate reacts with the aminoacyl moiety of the second aa-tRNA to form a dipeptidyl intermediate, which will further undergo intramolecular

cyclization, leading to the formation of the second peptide bond and the yield of final CDP product.⁹⁶ It is noteworthy that most of the CDPSs exhibit some promiscuity with respect to the recognition of the aa-tRNA substrates, thus producing a series of CDP products. Several CDPS-derived NPs have been characterized to date, including albonoursin from *S. noursei*,¹²² pulcherrimin from *Bacillus subtilis*,¹³⁰ mycocyclosin from *Mycobacterium tuberculosis*,¹³¹ nocazines from *Nocardiopsis dassonvillei*,¹³² and bicyclomycin from *Streptomyces sapporonensis* (**Figure 9**).^{133, 134}

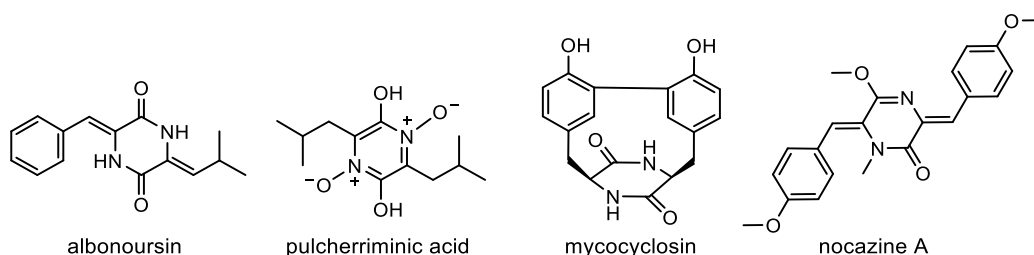


Figure 9 Examples of DKPs generated by CDPS pathways.

1.4 Tailoring enzymes in CDPS-dependent pathways

Similar to most NPs from microorganisms, the enzymes modifying the assembled DKP-scaffolds can be found in nearly all of the putative NRPS- or CDPS-associated gene clusters. Genes coding for multiple types of tailoring enzymes from different families have been found close to the respective CDPS gene in the available genome databases. These putative modification enzymes include cytochrome P450s (P450s), α -ketoglutarate/ Fe^{2+} -dependent dioxygenases (α -KGDs), flavin-containing oxidoreductases, α/β -hydrolases, prenyltransferases (PTs) and methyltransferases (MTs), cyclases, and ligases. Once the DKP scaffolds are synthesized, various types of chemical transformations can be introduced on the CDP ring core and amino acid residues by associated tailoring enzymes, generating diverse DKP-containing compounds.

1.4.1 Cytochrome P450s

Cytochrome P450s (P450s) are a superfamily of heme-dependent monooxygenases. They are widely called as monooxygenases because they are able to catalyze the scission of dioxygen (O_2), leading to the insertion of one oxygen atom into the substrate, whereas the other one is reduced to water.¹³⁵ The P450 enzymes were named for the maximum absorbance of the heme (the Soret peak) in the ferrous-CO-bound complex at 450 nm.^{136, 137} P450s are the most versatile biocatalysts in nature and widely distributed throughout all life kingdoms, from archaea, prokaryotes, to eukaryotes.¹³⁸ In addition to their well-known roles in human health, such as xenobiotic detoxification, steroid biosynthesis and drug metabolism, P450s also play significant roles in the biosynthesis of NPs.¹³⁹

1.4.1.1 The P450 catalytic cycle

The catalytic core of P450s is composed of a conserved four helix (D, E, I, and L helices) bundle, forming a trigonal prism-shaped structure.¹⁴⁰ P450s involved in natural product assembly and tailoring

reactions exhibit diverse sequences and structures and lack conserved sequence motifs.¹³⁹ To date, most of the structurally defined P450 proteins associated with NP biosynthesis originate from actinomycetes, especially from *Streptomyces* species, with sequence identities less than 43%.¹⁴¹ The cysteine (Cys) residue is the only absolutely conserved residue among all members of P450 family, serving as the heme iron ligand located at the N-terminal end of a helix. A conserved G³⁴⁷XXXC³⁵¹ motif (residue numbering according to the EryF sequence) in the heme-binding loop harboring the Cys residue is found in bacterial P450s.¹⁴¹

The currently accepted catalytic mechanism for P450s (in this case for the normal hydroxylation reaction) employs a sophisticated, multi-step catalytic cycle involving a range of transient intermediates (**Figure 10**).¹⁴² (i) The catalytic cycle starts from the substrate-free resting state. In addition to the conserved Cys residue and four nitrogen atoms of the porphyrin (Por) ring, one water molecule is coordinated to the ferric heme-iron (Fe^{III}) as the sixth ligand. (ii) The substrate (R-H) binds to the active site and displaces the water ligand, resulting in a substrate-bound ferric complex with the shift of the low-spin ferric state to high-spin one. (iii) A single electron is transferred from a redox partner and reduces the ferric (Fe^{III}) state to the ferrous (Fe^{II}) state. (iv) Then, one molecular oxygen binds to the ferrous heme iron (Fe^{II}) to form the ferrous dioxy [Fe^{II}-O₂] complex and (v) followed by the second electron reduction event from the redox partner to generate a peroxo-ferric [Fe^{II}-OO²⁻] intermediate. (vi) This intermediate is protonated to form the ferric hydroperoxy [Fe^{III}-OOH] complex, which is also known as Compound 0 (Cpd 0). (vii) The second protonation and further heterolytic cleavage of the O-O bond with concurrent release of a water molecule gives rise to the transient and highly reactive ferryl-oxo intermediate [Fe^{IV}=O], referred as compound I (Cpd I). (viii) Cpd I then abstracts a hydrogen atom from the substrate to form the ferryl-hydroxo compound II (Cpd II) with a substrate radical. (ix) The radical rebounds the hydroxyl group of Cpd II to form the hydroxylated product (R-OH). Dissociation of the monooxygenated product (R-OH) from the active site and the rebound of a water molecule as the sixth heme ligand lead to the regeneration of the resting state of the P450 enzyme, thus completing the catalytic cycle. It should be noted that some substrate-P450 complex can directly convert into Cpd 0 by utilizing H₂O₂ as the sole electron and proton donor, known as the peroxide shunt pathway.

The two electrons required for the heme-Fe^{III} reduction events in this cycle are derived from the cofactors NADPH or NADH and transported through the redox partner systems. There are two major types of P450 redox systems: (i) Class I is a two-component system, comprised of a flavin adenine dinucleotide (FAD)-containing reductase (FdR) along with an iron-sulfur containing ferredoxin (Fdx). In this case, the P450 and its redox partners FdR and Fdx are soluble cytoplasmic enzymes, which are usually present in most bacterial and mitochondrial P450s. (ii) The Class II P450 system has a single FAD- and flavin mononucleotide (FMN)-containing reductase, referred to cytochrome P450 reductase (CPR), as redox partner. In this class, both the P450 and its partner are membrane-bound enzymes, which are mainly found in eukaryotic organisms.¹⁴³

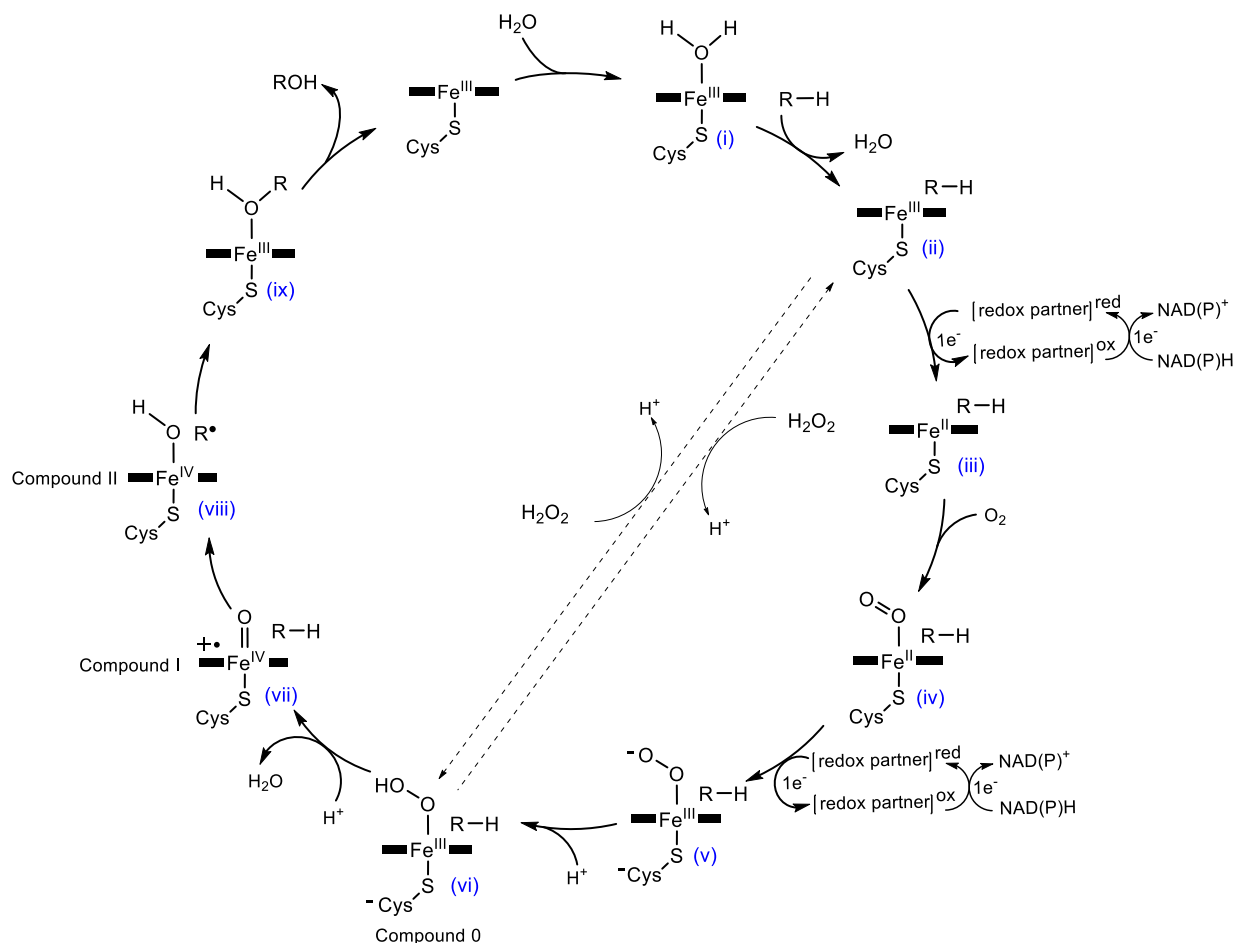


Figure 10 The P450 catalytic cycle with hydroxylation as an example.

1.4.1.2 P450-mediated diverse reactions in natural product biosynthesis

In addition to the initial assembled building blocks, the structural complexity of natural products mainly arises from post-decorating enzymes to highly functionalize the skeletons with a set of chemical transformations. Among them, P450s are one of the most utilized and talented enzymes that catalyze a great variety of reactions to modify diverse NP scaffolds (**Figure 11**).¹³⁹

The well-known reactions catalyzed by P450s in NP biosynthesis are C-H or N-H bond hydroxylation and C=C double bond epoxidation. Notably, over two-thirds of the characterized P450s from *Streptomyces* catalyze hydroxylations.¹³⁹ P450s for these oxidative reactions exhibit high regio- and stereoselectivity, resulting in the rigid order of the biosynthetic steps. Recently, several P450s have been characterized as multifunctional enzymes that can catalyze a sequential hydroxylation and epoxidation reactions.^{50, 144} For instance, MycG is capable of catalyzing two consecutive oxidation reactions on the 16-membered ring macrolide to generate mycinamicin V and mycinamicin II, respectively.¹⁴⁴ TamI, together with the FAD-dependent oxidase TamL as the partner, catalyzes multiple oxidative steps in a defined order, including C10 hydroxylation, C11/C12 epoxidation, and C18 hydroxylation in tirandamycin biosynthetic pathway.¹⁴⁵

INTRODUCTION

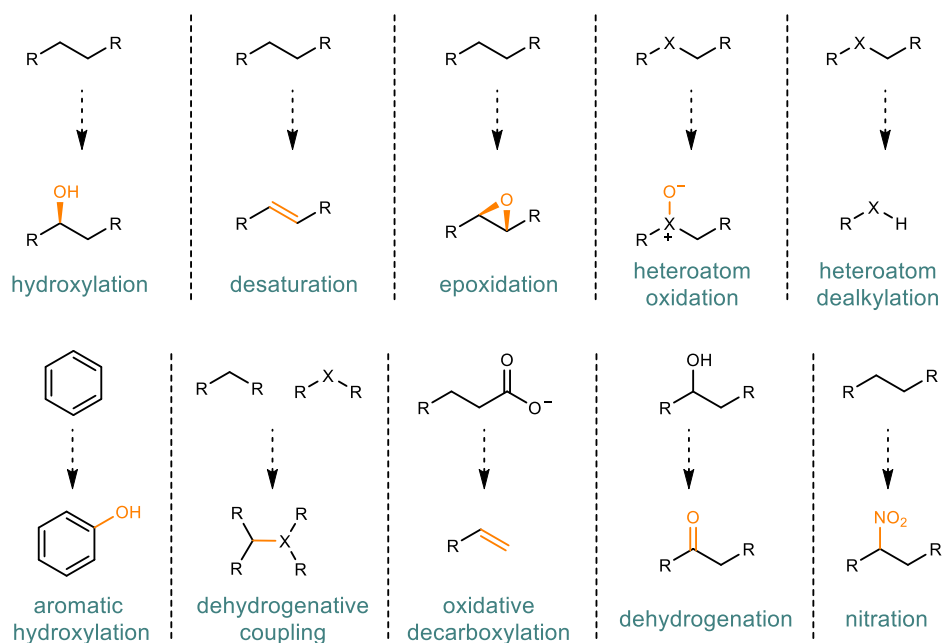
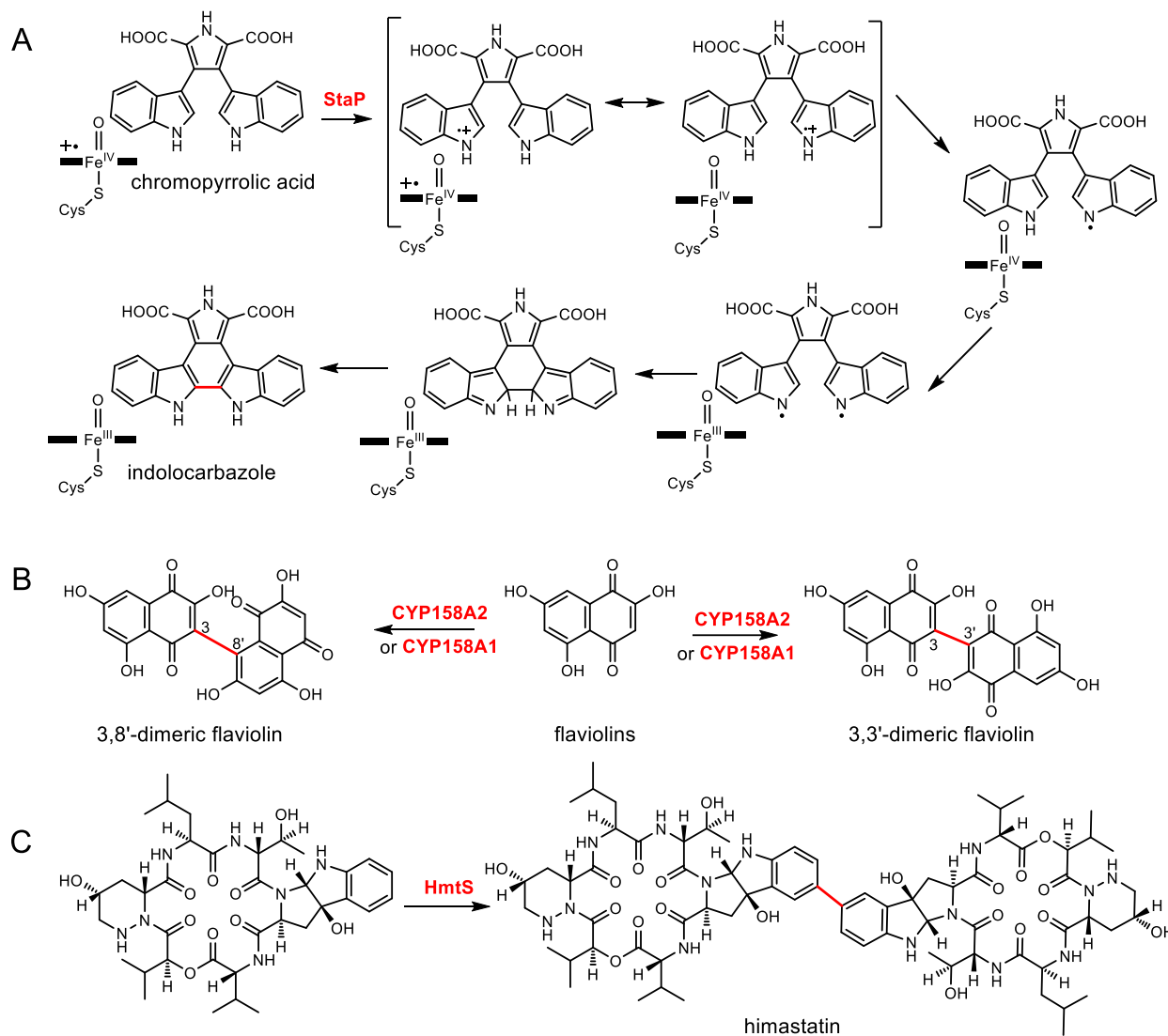


Figure 11 Diverse transformations catalyzed by P450s in natural product biosynthesis.

The P450-mediated common transformations like hydroxylation and epoxidation, do not significantly alter or affect the substrate skeletons. Whereas, some other unusual or intriguing chemical reactions performed by P450s may lead to dramatic structural modifications.¹³⁸ A broad range of unique P450 reactions leading to skeleton construction or structure re-shaping, such as bond coupling, cleavage, and migration, have attracted much more attention recently.¹³⁸

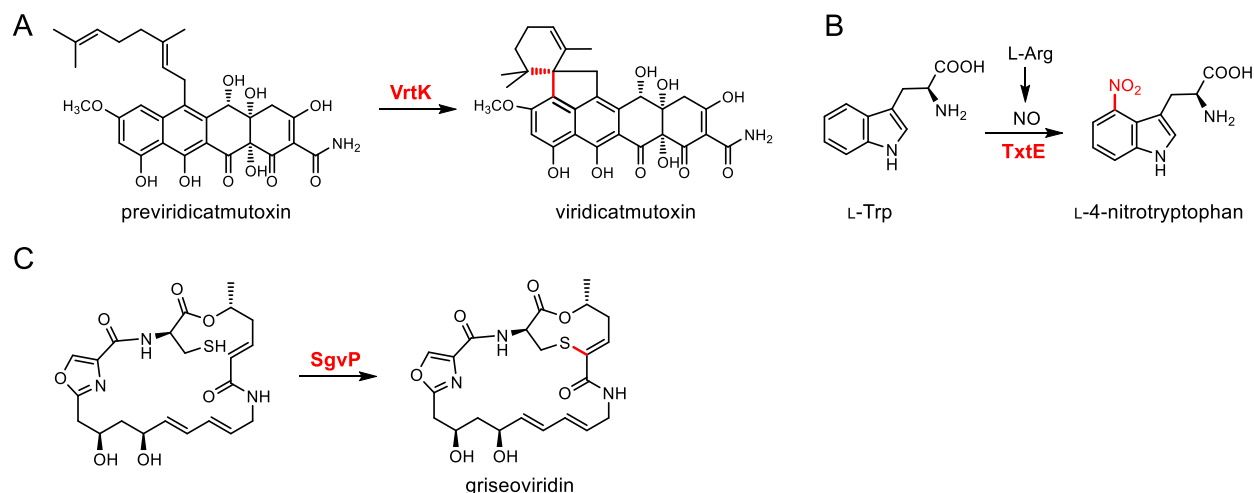
The biaryl axis is the central structural element in a large number of structurally attractive and biologically important chemical entities.¹⁴⁶ Nevertheless, phenol coupling usually requires complex steps and is difficult to achieve *via* chemical synthesis due to the high challenges for controlling the linkage of two monomers and the configuration of the biaryl axis at the same time.^{146, 147} It has been reported that many P450 enzymes are capable of catalyzing aromatic coupling reactions. For instance, the well-known and studied P450 enzyme StaP (CYP245A1) catalyzes an intramolecular C–C bond formation, leading to the generation of the indolocarbazole core in staurosporine biosynthesis.¹⁴⁸⁻¹⁵⁰ An indole cation radical mechanism was proposed for the StaP catalyzed aryl–aryl coupling reaction (**Scheme 1A**). In addition to intramolecular aromatic coupling, P450s have been identified to possess the capability of mediating intermolecular aryl ring coupling as well. During the pigment biosynthesis in *S. coelicolor* A3(2), two P450 homologues CYP158A1 and CYP158A2, sharing 61% sequence identity, catalyze the polymerization of flaviolin but with different product profiles (**Scheme 1B**).^{148, 151} An even more challenging aromatic coupling reaction is performed by the P450 enzyme HmtS in the biosynthesis of himastatin, which takes two bulky monomers as the substrate (**Scheme 1C**).¹⁵² In recent years, an increasing number of P450s have been reported to catalyze C–C linkage reaction, leading to ring formation. As a result, the structural “phenotype”, such as the shape, configuration and rigidity of the given substrate, can be significantly changed, which play important roles in constructing the skeletons of NPs.¹³⁸ A striking example is known as VrtK that represents the first P450 capable of

catalyze terpene cyclization (**Scheme 2A**).¹⁵³



Scheme 1 Examples of P450-mediated intra- and intermolecular aryl–aryl coupling reactions.

Besides the C–C crosslinking reactions, some P450s are also characterized to catalyze C–N bond and C–S bond formation. It is widely acknowledged that the C–N bond formation is achieved by the oxidative formation of an electrophilic functional group, which can subsequently be attacked by the nitrogen atom.¹³⁹ Notably, in the biosynthesis of thaxtomin, the unique P450 TxtE nitrates the free L-tryptophan at the C4 position to afford a C–N linkage in the presence of O₂ and NO from L-arginine (**Scheme 2B**).¹³⁵ During griseoviridin biosynthesis, the cytochrome P450 monooxygenase SgvP is involved in a special C–S bond formation through the direct coupling of the free SH group, leading to the formation of a nine-membered thioene-containing lactone ring (**Scheme 2C**).¹⁵⁴ Interestingly, some P450-mediated uncommon reactions including decarboxylation, C–C bond cleavage or migration, can result in a range of skeleton construction reactions, like ring opening and contraction, contributing significantly to the structural diversity of NPs.^{138, 155-157}



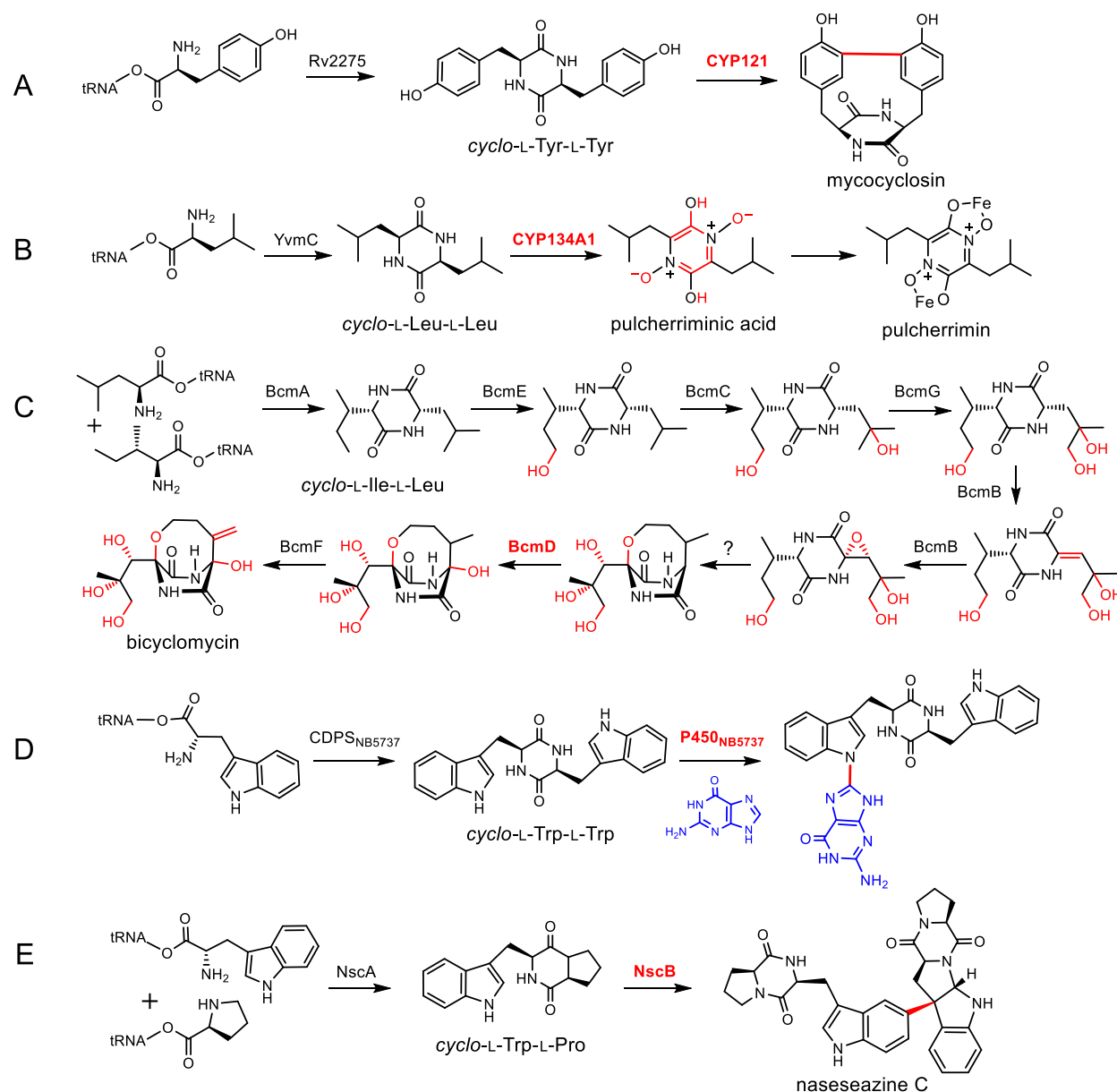
Scheme 2 Examples of P450-mediated bond formation reactions.

1.4.1.3 P450s involved in CDPS-dependent pathways

In the past years, a large number of CDPSs have been functionally characterized from bacteria, most of which are clustered with predicted DKP tailoring genes. However, CDPS-dependent biosynthetic pathways remain largely unexplored. Only nine pathways have been fully studied prior to this thesis, including five P450-associated ones. The first example is the biosynthetic pathway of mycrocyclosin in *M. tuberculosis*. Following the synthesis of *cyclo*-L-Tyr-L-Tyr by the CDPS Rv2275, the cytochrome P450 enzyme CYP121 (Rv2276) catalyzes the intramolecular diaryl C–C coupling reaction to generate mycrocyclosin.¹³¹ The P450 CYP121 showed high substrate specificity, as the oxidative C–C coupling was not observed for other aromatic cyclodipeptide substrates. A bi-carbon radical mechanism has been postulated for the C–C bond formation based on its crystal structure and biochemical studies (**Scheme 3A**). Another example is the biosynthesis of pulcherrimin acid in *B. subtilis*. After *cyclo*-L-Leu-L-Leu assembled by the CDPS YvmC, the P450 CYP134A1 (CypX) performs two-step N-hydroxylations and the subsequent dehydrogenation on the DKP ring to yield pulcherriminic acid, which is transformed to the red pigment pulcherrimin *via* non-enzymatic chelation with iron (**Scheme 3B**).¹³⁰

Three other examples of P450s were identified in 2018. The P450 enzyme BcmD responsible for the hydroxylation on the DKP ring was characterized in the biosynthesis of bicyclomycin. The initial *cyclo*-L-Ile-L-Leu undergoes a cascade of dehydrogenation and epoxidation reactions by five α -ketoglutarate/ Fe^{2+} -dependent dioxygenases and the hydroxylation by BcmD to afford the final bicyclic structure (**Scheme 3C**).¹³³ Yu *et al.* characterized the novel P450_{NB5737} responsible for the coupling of a guanine moiety with *cyclo*-L-Trp-L-Trp *via* C–N linkage *via in vivo* heterologous expression and *in vitro* biochemical investigation (**Scheme 3D**).¹⁵⁸ This case represents the first example of nucleobase modification of a peptide natural product from biosynthetic origin and thereby extends the spectrum of transformations mediated by P450s. The third case is NascB, which was identified as a dimerase for intermolecular C–C bond formation between two *cyclo*-L-Trp-L-Pro precursors during the biosynthetic

pathway of nasezeazine C (**Scheme 3E**).¹⁵⁹ In addition, a whole-cell biocatalysis system was developed by incorporating NascB into an engineered *E. coli* strain, which thereby efficiently created 30 dimeric nasezeazine analogs by using different CDPs as substrates. Very recently, the structures of two dimerases, NzeB and Nas_{F5053} involved in the CDPS pathway, were solved *via* X-ray analysis, providing us new insights into the molecular basis for DKP dimerization.^{160, 161}



Scheme 3 Examples of P450s involved in CDPS-dependent pathways.

1.4.2 Other characterized enzymes in CDPS-dependent pathways

Cyclodipeptide oxidases (CDOs) represent the first recognized tailoring enzymes in CDPS-containing biosynthetic pathways.¹²² They are composed of two subunits A and B, of which subunit A displays a

fairly high degree to nitroreductases and catalyzes the dehydrogenation reaction in the presence of flavin as the cofactor.¹²² In the biosynthesis of albonoursin, AlbA together with AlbB firstly catalyzes the α,β -dehydrogenation at the phenylalanine residue of *cyclo*-L-Phe-L-Leu, then installs the second C–C double bond at the leucine hemisphere to yield the di-dehydrogenated product (**Scheme 4A**).¹²² The second characterized CDO, consist of the two subunits Ndas_1146 and Ndas_1147, was proven to be responsible for the α,β -dehydrogenation in the biosynthesis of the nocazine family in *N. dassonvillei* (**Scheme 4B**).¹³²

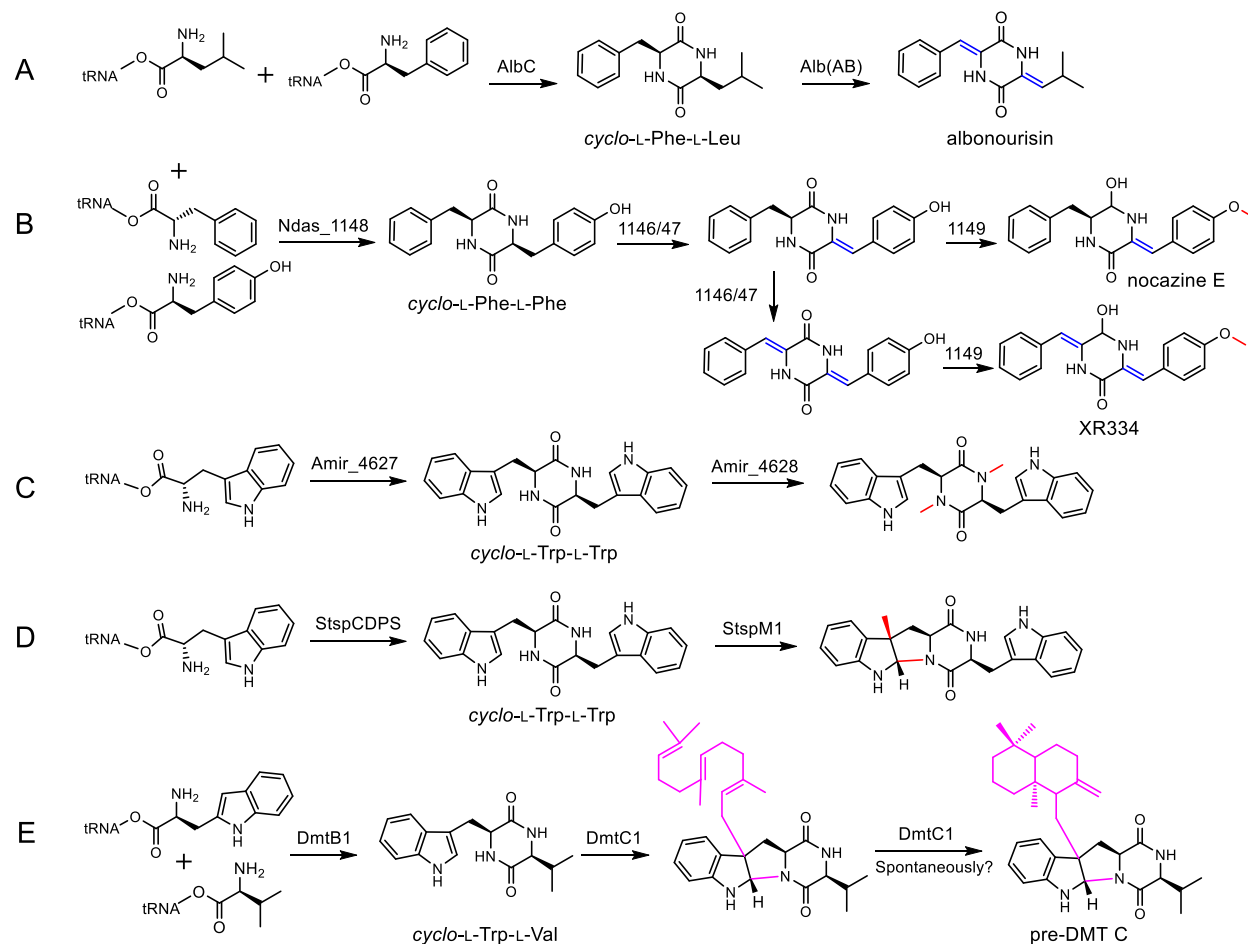
α -Ketoglutarate/ Fe^{2+} -dependent dioxygenases (α -KGDs) are another class of oxidoreductases discovered in CDPS pathways. They catalyze a wide range of oxidative transformations and play critical roles in biochemical processes.¹⁶² Five α -KGDs were verified in the biosynthesis of bicyclomycin, a commercial antibiotic inhibiting the transcription termination factor *rho*.^{133, 134} As described in the **section 1.4.1.3**, following the CDP assembled by BcmA, three α -KGDs BcmE, BcmC, and BcmG successively perform the hydroxylations on three unactivated sp^3 carbons. Subsequently, another α -KGD BcmB catalyzes the desaturation and epoxidation at the leucine hemisphere. Then, the next heterocyclization leads to the formation of the O-bridged bicyclo-[4.2.2]piperazinedione ring. Finally, a tertiary hydroxyl group installed by the P450 monooxygenase BcmD, and an exo-methylene moiety introduced by the remaining α -KGD BcmF-catalyzed dehydrogenation, complete the biosynthesis of bicyclomycin (**Scheme 3C**).

S-Adenosylmethionine (SAM) dependent methyltransferases (MTs) are common tailoring enzymes in the biosynthesis of a large member of NPs.¹⁶³ The introduction of methyl substituents could influence the biological activities and physicochemical properties of the generated products.¹⁶⁴ The common MTs associated with NP biosynthesis catalyze the transmethylation *via* $\text{S}_{\text{N}}2$ -like nucleophilic substitution, in which a nucleophilic atom from the substrate attacks the electrophilic carbon atom of SAM, resulting in the simultaneous breaking of the sulphur-carbon (S–C) bond and the formation of a new carbon-nucleophile bond.^{163, 165} To date, four MTs from *cdps* gene clusters have been characterized. Giessen and co-workers verified that Ndas_1149 catalyzes O-methylation of DKP tyrosine residues, and Ndas_1145 is responsible for N- and/or O-methylation of the DKP ring in the biosynthetic pathway of nocazine derivatives (**Scheme 4B**).¹³² They also reported another MT homologue Amir_4628 catalyzing the methylation on one or both DKP nitrogens of *cyclo*-L-Trp-L-Trp (**Scheme 4C**).¹⁶⁶ In 2019, a novel indole C3 methyltransferase from *Streptomyces* sp. HPH0547 was described (**Scheme 4D**).¹⁶⁷

Prenyltransferases (PTs) are widely spread in nature, leading to a wide range of prenylated NPs with important biological activities. So far, various PTs that carry out normal or reverse prenylations on different positions of the CDP scaffolds assembled by NRPSs, have been intensively studied.¹⁶⁸ The phytoene-synthase-like family prenyltransferase DmtC1 capable of the prenylation of an indole alkaloid DKP in a CDPS-associated pathway, has been reported very recently.¹⁶⁹ DmtC1 catalyzes the condensation of a farnesyl diphosphate (FPP) onto the C3 position of the indole ring, thus generating the pre-drimentines, which can be subsequently cyclized by the membrane-associated terpene cyclase DmtA to give rise to the final drimentines (**Scheme 4E**).

INTRODUCTION

In general, all these mentioned studies definitely expand the spectrum of DKP biosynthetic pathways. Yet genomics has revealed much DKP biosynthetic potential remains unexplored. Plenty of predicted tailoring enzymes from different families, including glycosyltransferases, acetyltransferases, and sulfotransferases are closely associated with CDPSs.¹⁷⁰ In the future, exploration of these cryptic *cdps*-containing gene clusters will significantly increase the chemical diversity of 2,5 DKPs and give rise to novel biologically active molecules for drug leads.



Scheme 4 Other tailoring enzymes involved in CDPS-dependent pathways.

2 Aims of this thesis

In this thesis, the following issues have been addressed:

Identification of nine CDPSs for the biosynthesis of tryptophan-containing cyclodipeptides from *Streptomyces*

CDPSs, mainly found in bacteria, can directly hijack the activated aminoacyl-tRNAs (aa-tRNAs) from the ribosomal machinery to produce cyclodipeptides, the simplest representative of 2,5-DKPs. Among them, tryptophan-containing DKPs have received an increasing attention due to their promising scaffolds for structural modification. In recent years, the number of putative gene coding for CDPSs in genomic databases has increased continuously. However, except some *cyclo*-L-Trp-L-Trp (cWW) synthases identified from eight actinomycetes strains, only one CDPS originating from the sea anemone *Nematostella vectensis* was verified to produce several tryptophan-containing CDPSs including *cyclo*-L-Trp-L-Leu, *cyclo*-L-Trp-L-Ala, and *cyclo*-L-Trp-L-Gly, but with very low product yields in *E. coli* culture. No other types of tryptophan-containing cyclodipeptides have been reported in microorganisms prior this project. Hence, the aim of this project is to identify new CDPSs that catalyze the formation of cyclodipeptides comprising tryptophan and another amino acid residues from actinobacteria. The following experiments were carried out in cooperation with Dr. Huili Yu:

- Searching CDPS homologues in public databases *via* blast by using the known actinobacterial cWW synthases as probes
- Phylogenetic analysis and sequence alignments of the candidate CDPSs with the characterized ones
- Cloning and overexpression of the candidate CDPSs in *E. coli* SoluBL21
- LC-MS analysis of the cultures of *E. coli* transformants harboring the CDPS expression constructs
- Isolation and structural elucidation of accumulated CDPs by HR-EIMS and ¹H NMR analysis
- Determination of the product yields of generated CDPs in each *E. coli* transformant *via* HPLC analysis

Identification and biosynthetic study of novel C3-guaninyl indole alkaloids guanitrypmycins

CDPS genes are often closely associated with those coding for modification enzymes, such as oxidoreductases, cytochrome P450s, cyclodipeptide oxidases (CDOs), prenyltransferases (PTs), and methyltransferases (MTs), which significantly expand the diversity of DKP-based natural products. Bioinformatic analysis of nine CDPSs that assemble tryptophan-containing CDPs in the first project, revealed that two CDPS genes, *gut*₂₄₃₀₉ from *Streptomyces monomycini* NRRL B-24309 and *gut*₃₅₈₉ from *Streptomyces varsoviensis* NRRL B-3589, are located in the similar gene loci containing four additional genes coding for three modification enzymes, *i.e.*, CDO, cytochrome P450, and MT. This

unique *cdps*-associated genetic organization, which is distinct from those studied before and has not been investigated yet, could be involved in the biosynthesis of novel DKP derivatives. Therefore, it is interesting to explore and investigate the possible products of these two gene clusters. The following experiments were carried out:

- Bioinformatic analysis of the two P450s present in *gut* gene clusters
- Heterologous expression of CDPS gene *gutA*₂₄₃₀₉ in *S. coelicolor* M1146 with the replicative vector pPWW50A to establish the genetic protocol
- Heterologous expression of the whole gene cluster *gut*₂₄₃₀₉ from *S. monomycini* NRRL B-24309
- Large-scale fermentation of the expression transformants for isolation and structural elucidation of the generated 2,5-DKPs guanitrypmycins
- Expression of different gene combinations of *gut*₂₄₃₀₉ and precursors feeding for determination of the biosynthetic steps of guanitrypmycins
- Incubation of guanitrypmycins in the deuterium-enriched condition to prove the non-enzymatic process of guanitrypmycins
- Different gene combination expressions of the second gene cluster *gut*₃₅₈₉ from *S. varsoviensis* NRRL B-3589 as well as isolation and structural elucidation of other guanitrypmycin derivatives
- Cultivation of the native strains in different media to validate the productivity of guanitrypmycins thereof
- *In vitro* biochemical investigation of the P450 GutD₃₅₈₉ as the key C3-guaninyl transferase

Expanding the spectrum of cytochrome P450s by identification of two distinct dimerases in CDPS-dependent pathways

Dimeric CDPs possess enormous chemical complexity due to the densely functionalized core and multiple stereogenic centers in their structures. Taking their biological activities together, dimeric CDPs hold significant promise for medicinal chemistry. Since a wide variety of actinobacterial genome sequences have been released in the past years, it provides a solid basis for the discovery of novel compounds and intriguing enzymes by genome mining. Comprehensive bioinformatic analysis revealed the presence of two *cdps-p450* operons in *Saccharopolyspora antimicrobica*. The two putative P450s TtpB1 and TtpB2 with sequence identity of 40%, are located in the phylogenetic clade near to the known dimerases, which indicate they could be involved in different dimerization substrates or/and new patterns. To verify the functions of the candidate genes and gene clusters, the following experiments were carried out:

- Bioinformatic and phylogenetic analysis to identify two *cdps-p450* operons
- Functional proof the two CDPS gene *ttpA1* and *ttpA2* by heterologously expressed in *S. albus* J1074

AIMS OF THIS THESIS

- Heterologous expression of the two *cdps-p450* gene clusters in *S. albus* J1074 and subsequent analysis of the accumulated products by LC-MS
- Large scale fermentation and isolation of the dimeric DKPs
- Structural elucidation of the purified products by 1D and 2D NMR
- Further confirmation of the functions of the two P450s *via* precursor feeding experiments
- *In vitro* biochemical investigation of the two P450s
- Evaluation of the antibacterial activity of the isolated dimeric DKPs
- Proposed mechanism for P450-mediated intermolecular coupling reactions

3 Results and discussion

3.1 Identification of nine CDPSs for the biosynthesis of tryptophan-containing cyclodipeptides from *Streptomyces*

CDPSs belong to a newly characterized enzyme family that catalyze the formation of cyclodipeptides by using aminoacyl-tRNA as substrates.⁹⁶ More than 75 different cyclodipeptides, spanning 18 of the 20 canonical amino acids, have been identified from the CDPS machinery in the past years.¹²⁴ Among them, tryptophan-containing CDPs have recently gained an increasing attention due to their promising scaffolds for structural modification.^{171, 172} Various enzymatic modifications including prenylation, oxidation, methylation, and dehydrogenation, as well as spontaneous rearrangements can occur on the electron-rich indole ring of the tryptophanyl moiety to generate diverse chemical complexity.^{173, 174} Prior to this thesis, over ten CDPSs have been verified to assemble cWW as the sole product.^{124, 166, 175} However, CDPSs synthesizing other types of tryptophan-containing CDPs have not been reported in microorganisms. In order to get CDPSs producing cyclodipeptides consisting of tryptophanyl moiety and other amino acid residues instead of cWW, we searched in the public databases for the putative homologs of the actinobacterial cWW synthases. Subsequently, based on the phylogenetic analysis with the known CDPSs, eleven candidates with moderate sequence identities were chosen for further investigation (**Figure 12**).

For functional proof, the coding region of the candidate CDPSs were amplified from the genomic DNA by PCR and cloned into pGEM T easy vector. After confirming sequence integrity, the fragments were released and ligated into pET28a (+) vector to generate the expression constructs, which were then transformed into *E. coli* SoluBL21 for gene expression. For CDPS overproduction, the transformants were induced with Isopropyl β -D-thiogalactopyranoside (IPTG). Afterwards, the cultures were extracted with ethyl acetate and then analyzed on LC-MS. In comparison to the culture of *E. coli* harboring the empty vector, the additional peak(s) could be easily detected in those of *E. coli* transformants. Subsequent isolation and structural elucidation of the generated products *via* HR-EIMS and NMR analysis, led to the undoubted identification of the generated cyclodipeptides.

Three new CDPSs, WP_078950527 from *Streptomyces lavendulae* NRRL B-2774, WP_019889609 from *Streptomyces purpureus* NRRL B-5737, and WP_063768158 from *Streptomyces xanthophaeus* NRRL B-5414, produced cWW as the sole product. WP_078873129 from *Streptomyces* sp. NRRL S-1868 and WP_051847149 from *Streptomyces* sp. NRRL F-5053 could produce *cyclo*-L-Trp-L-Pro (cWP) and *cyclo*-L-Trp-L-Leu (cWL) as a sole and predominant product, respectively. Two product peaks were detected in the transformant harboring WP_052397358 from *Streptomyces* sp. NRRL F-5123 with cWP and cWL as the major and minor product. cWP and *cyclo*-L-Trp-L-Ala (cWA) were the major and minor product of WP_078872750 from *Streptomyces* sp. NRRL S-1868, respectively. The transformant with KOG90878 from *S. varsoviensis* NRRL B-3589 had a similar and complex product spectrum as WP_078624487 from *S. monomykini* NRRL B-24309 with *cyclo*-L-Trp-L-Phe (cWF) and

RESULTS AND DISCUSSION

cyclo-L-Trp-L-Tyr (cWY) as the predominant and five other tryptophan-containing cyclodipeptides as minor products (**Figure 13**). In contrast, the two CDPs, WP_031028810 from *Streptomyces* sp. NRRL F-5639 and BAU83478 from *Streptomyces laurentii* DSM41684 were proven to catalyze the formation of *cyclo*-L-Phe-L-Leu (cFL). In this project, six expression constructs were prepared by Dr. Huili Yu. This PhD candidate was responsible for other five constructs and performed isolation and structural elucidation of the all generated CDPs.

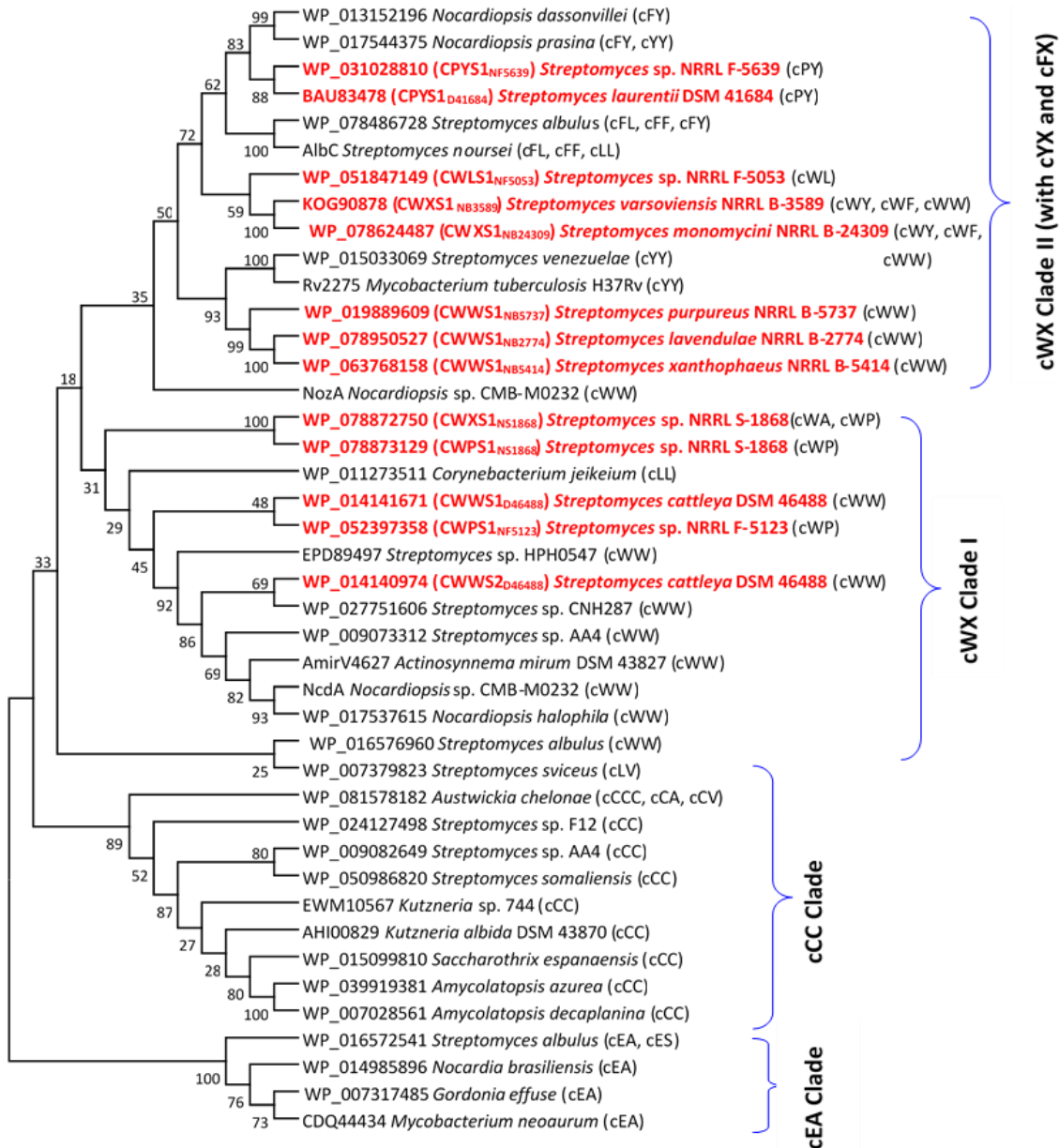


Figure 12 Phylogenetic analysis of CDPs from actinobacteria.

RESULTS AND DISCUSSION

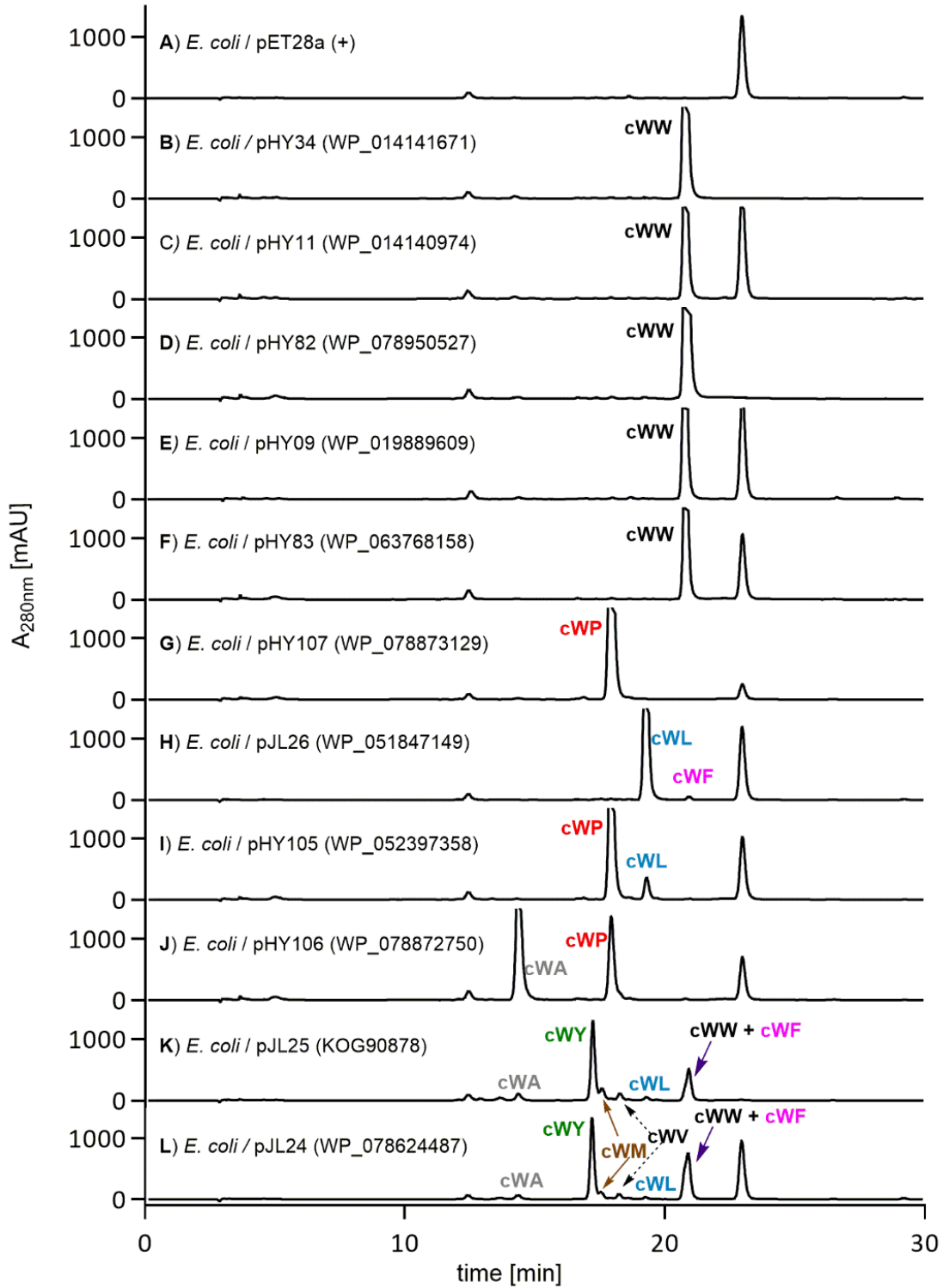


Figure 13 LC-MS analysis of recombinant *E. coli* strains with empty vector (A) and CDPs (B-L).

To determine the total product yields of these CDPs in *E. coli* culture, different growth parameters were optimized under three induction temperatures (20 °C, 30 °C or 37 °C) plus two IPTG concentrations (0.2 or 0.5mM). Product yields were calculated to be 46 to 211 mg/L via HPLC analysis after cultivation for 36 h under the optimized conditions, some of which are much higher than those previously reported.

In summary, three new cWW, one cWL, and two cWP synthases were identified from *Streptomyces* strains. In addition, three CDPSs which can produce cWA or cWY as the major metabolite were also characterized. Thus, this study represents rare examples of CDPS family derived from actinobacteria that can form various tryptophan-containing cyclodipeptides. Furthermore, this study highlights the potential of the microbial machinery for tryptophan-containing cyclodipeptide biosynthesis and provides a solid basis for further combination of *cdps* with other modification genes in synthetic biology.

For details on this work, please see the publication (section 4.1)

Jing Liu,* Huili Yu* and Shu-Ming Li (2018). Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from *Streptomyces* strains. *Applied Microbiology and Biotechnology*, 102, 4435–4444, DOI: 10.1007/s00253-018-8908-6. (*equal contribution)

Jing Liu,* Huili Yu* and Shu-Ming Li (2018). Correction to: Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from *Streptomyces* strains. *Applied Microbiology and Biotechnology*, 102, 5787-5789. (*equal contribution)

3.2 Identification and biosynthetic study of novel C3-guaninyl indole alkaloids guanitrypmycins

The biosynthetic genes that are responsible for a specific natural product tend to cluster in a certain region of microbial chromosomes, which means that locating one gene from a biosynthetic pathway typically facilitates the discovery of others from the same pathway.¹⁷⁶ In analogy, the utilization of CDPS-encoding genes as “biosynthetic hooks” is an efficient strategy to identify the genes coding for the tailoring enzymes in CDPS-dependent pathways.¹²³ Different types of putative tailoring enzymes, such as FAD-dependent oxidoreductases, cytochrome P450s, cyclodipeptide oxidases (CDOs), prenyltransferases (PTs), methyltransferases (MTs) and cyclases, are found close to the CDPSs, which may significantly expand the diversity of DKP-based natural products. Although hundreds of CDPSs have been identified in recent years, only nine CDPS-containing biosynthetic pathways have been functionally characterized prior to this thesis, including five P450-associated pathways.^{124, 176-178} These include CYP121 for an intramolecular C–C bond formation within one *cyclo*-L-Tyr-L-Tyr molecule,¹³¹ CYP134A1 for the aromatization and N-oxide formation of *cyclo*-L-Leu-L-Leu,¹³⁰ BcmD for installing a tertiary hydroxyl group,¹³³ P450_{NB5737} for the coupling of guanine with *cyclo*-L-Trp-L-Trp via C–N bond formation,¹⁵⁸ as well as NascB for dimerization of *cyclo*-L-Trp-L-Pro.¹⁵⁹ As the most versatile biocatalysts in nature, cytochrome P450s play important roles in natural product biosynthesis, which encouraged us to investigate more *p450*-associated *cdps*-containing gene clusters for generation of various DKP derivatives.

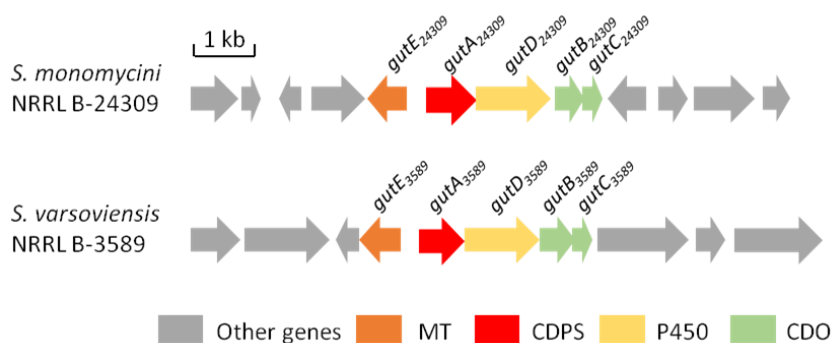


Figure 14 Comparative illustration of the *gut* clusters.

In the first project, we identified nine new CDPSs that assemble tryptophan-containing cyclodipeptides from different *Streptomyces* strains. Further sequence analysis revealed that two CDPS-coding genes, *gutA*₂₄₃₀₉ from *S. monomycini* NRRL B-24309 and *gutA*₃₅₈₉ from *S. varsoviensis* NRRL B-3589, are located in the similar gene loci containing four more genes coding for three modification enzymes, *i.e.*, CDO, cytochrome P450, and MT (**Figure 14**). Multiple sequence alignments showed that the two P450 enzymes encompass the conserved motifs as other bacterial P450s, *i.e.*, the G³⁴⁷XXXC³⁵¹ (referring to EryF¹⁴¹) motif in the heme-binding loop and the highly conserved A/Gⁿ⁻¹-Gⁿ-XX-Tⁿ⁺³ motif in the long

I-helix running over the distal surface of the heme in P450 scaffold. Phylogenetic analysis showed that the two P450 enzymes are closely located to P450_{NB5737} from *S. purpureus*¹⁵⁸ and CYP121 from *M. tuberculosis*¹³¹ (**Figure 15**), which catalyze inter- and intramolecular coupling reactions, respectively. However, the functions of the P450 candidates from the mentioned clusters cannot be predicted by sequence analysis and comparison with known enzymes. In addition, the unique *cdps*-associated genetic organization presented in **Figure 14** is distinct from those studied before and has not been investigated yet, which could be involved in the biosynthesis of novel DKP derivatives. In this study, we identified and characterized the two similar five-gene operons for the formation of unprecedented C3-guaninyl DKPs, named guanitrypmycins, whereby P450 catalyzes the unprecedented C–C linkage between the indole ring of the CDPs with the nucleobase guanine.

The corresponding proteins of the two aforementioned *gut* clusters, share sequence identities between 79 – 90% on the amino acid level, indicating that they may produce identical or very similar metabolites. Thus, the cluster *gut*₂₄₃₀₉ was investigated as a representative. Firstly, in order to establish a genetic protocol, *gutA*₂₄₃₀₉ was amplified from the genomic DNA by PCR and cloned into the replicative vector pPWW50A, then transformed into *S. coelicolor* M1146 for heterologous expression. LC-MS analysis revealed that *gutA*₂₄₃₀₉ expression strain can produce cWF (**1**) as the predominant and cWY (**2**) as the trace product, differing clearly from the result obtained from *E. coli*, which may be caused by different expression hosts.

For functional proof of the BGC, the whole *gut*₂₄₃₀₉ cluster was introduced into *S. coelicolor*. LC-MS analysis of the generated transformant harboring *gut(ABCDE)*₂₄₃₀₉ revealed the presence of four new peaks in two pairs regarding their molecular weights, **3a** and **3b**, **4a** and **4b**. **3a** and **3b** with [M+H]⁺ ions at m/z 495.189 ± 0.005 are 14 Daltons larger than **4a** and **4b**. While only **4a** and **4b** were detected as the predominant metabolites of the expression transformant with *gut(ABCD)*₂₄₃₀₉, indicating the methylation performed by the MT enzyme GutE₂₄₃₀₉. Large-scale fermentation and isolation afforded analytically pure compounds. Their structures were subsequently elucidated *via* detailed interpretation of NMR analysis including ¹H, ¹³C APT, COSY, HSQC, HMBC, and NOESY. Both **3a** and **3b** bear an exocyclic C–C double bond at the phenylalanyl side and a methylguaninyl moiety is attached to C-3 position of cWF (**1**), thereby forming a hexahydropyrrolo[2,3-b]indole framework. **3a** and **3b** differ from each other merely in the stereochemistry at C-11. H-11 in **3a** has an α -configuration while **3b** has a β -orientation. Comprehensive NMR analysis confirmed that **4a** and **4b** are demethylated counterparts of **3a** and **3b**, respectively. Due to the C3-guaninyl pyrroloindoline feature, **3a**, **3b**, **4a**, and **4b** were named herewith guanitrypmycins A1-1, A1-2, B1-1, and B1-2, respectively. Afterwards, the conversion of **3a** to **3b** as well as **4a** to **4b** was detected during the cultivation process. Incubation of these compounds separately in GYM media also confirmed the non-enzymatic transformation. Furthermore, to prove the spontaneous conversion formed *via* keto–enol tautomerism, **3a**, **3b**, **4a**, and **4b** were incubated in 50% CD₃OD in D₂O at different pH values. Subsequent LC-MS analysis of the incubation

RESULTS AND DISCUSSION

mixtures proved the incorporation of one deuterium atom into **3b** and **4b**, thereby confirming the existence of the keto–enol tautomerism equilibration.

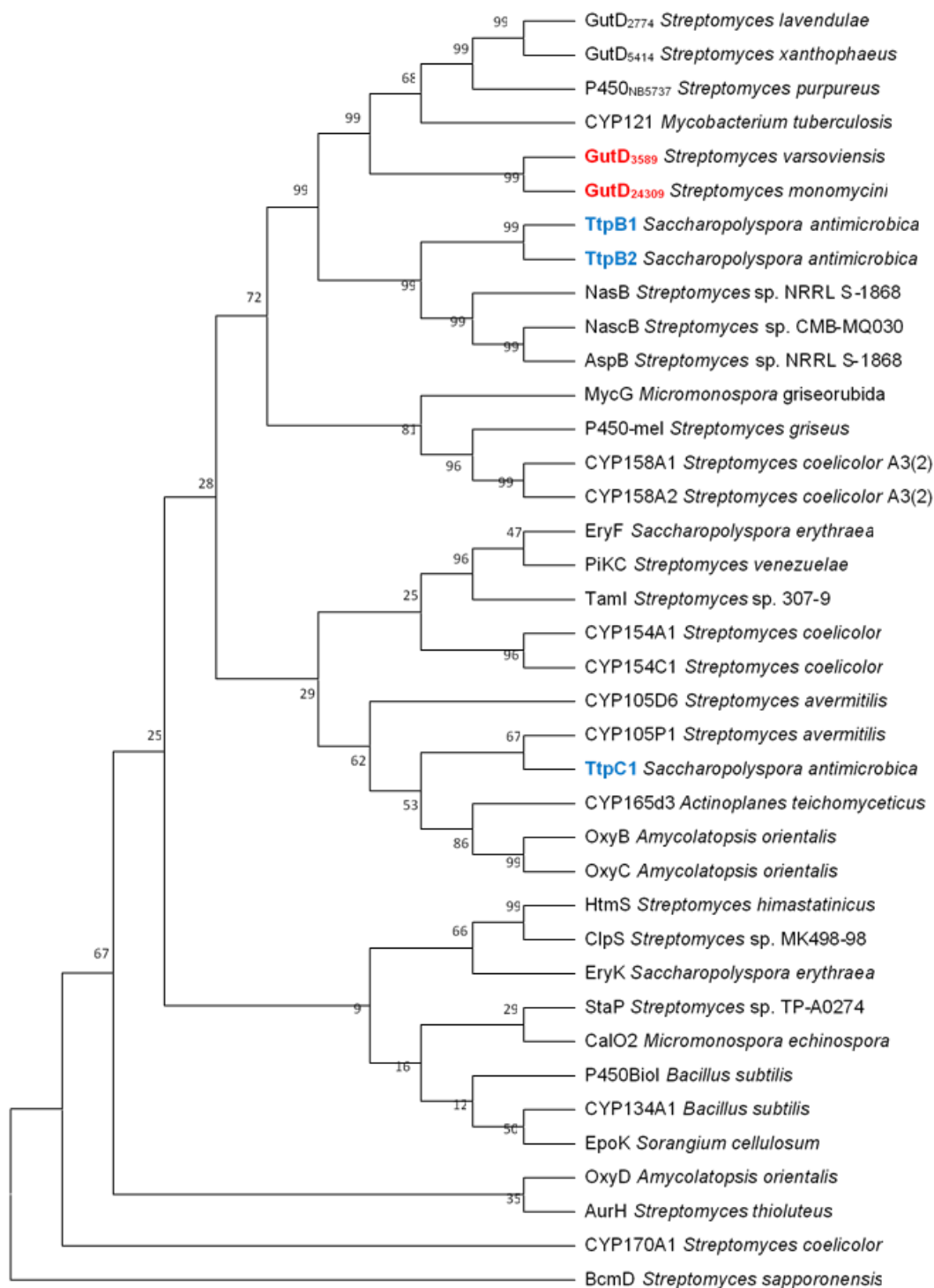


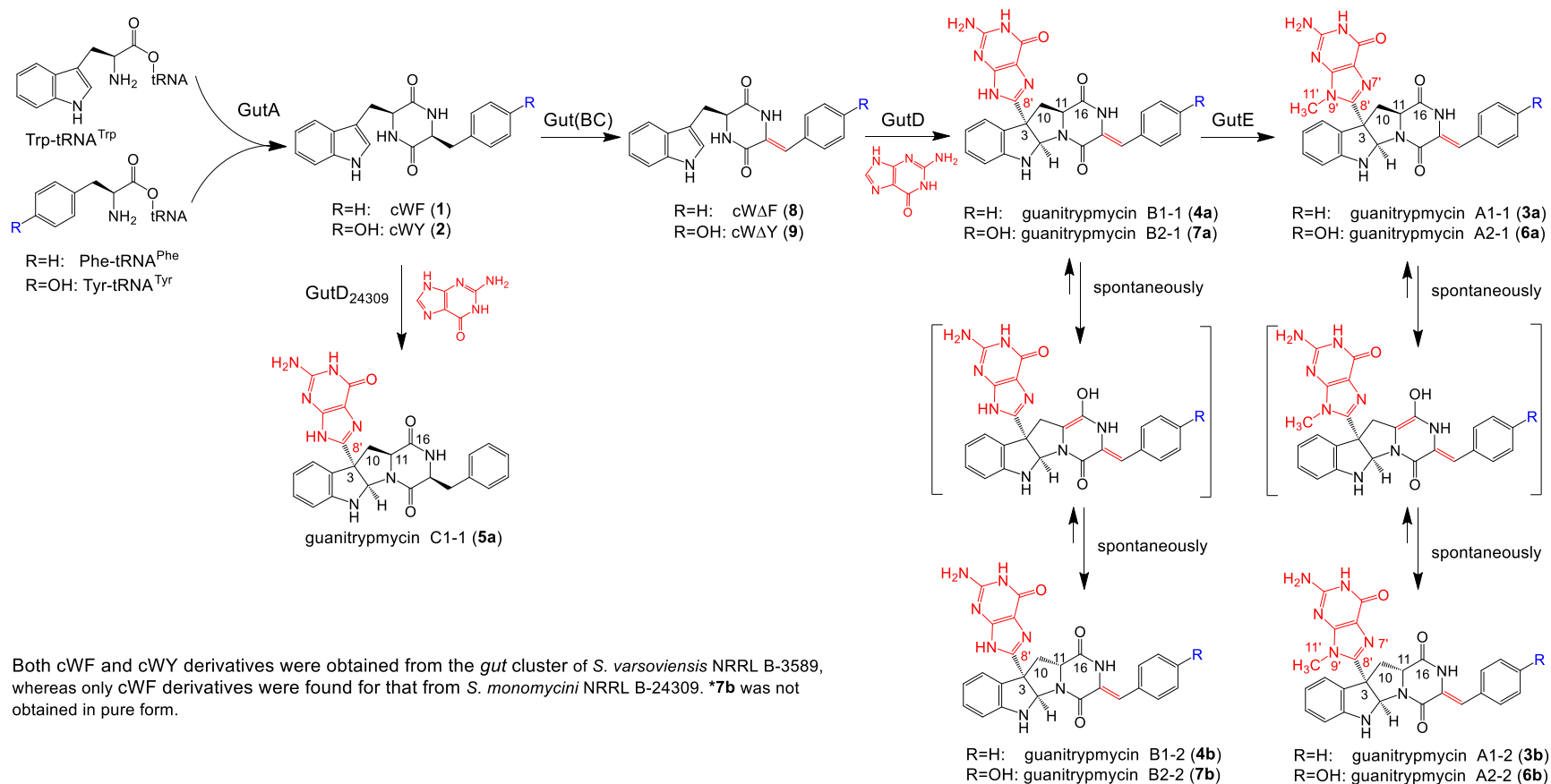
Figure 15 Phylogenetic analysis of P450s investigated in this study (in bold red) and functionally characterized P450s from bacteria.

The above data proved the enzyme functions in the biosynthesis of guanitrypmycins. GutA₂₄₃₀₉ assembles cWF (**1**), which will be dehydrogenated by Gut(BC)₂₄₃₀₉ and connected with a guanine moiety by GutD₂₄₃₀₉, finally methylated by GutE₂₄₃₀₉. The accurate reaction order of Gut(BC)₂₄₃₀₉ and GutD₂₄₃₀₉ was then determined by expression of different gene combinations together with precursor feeding experiments. When *gut(AD)*₂₄₃₀₉, *gut(ADE)*₂₄₃₀₉, and *gut(AE)*₂₄₃₀₉ were expressed in *S. coelicolor*, only slight or nearly no consumption of cWF (**1**) was observed in the transformants. In the transformant with *gut(AD)*₂₄₃₀₉, an additional compound **5a**, the guaninylated cWF (**1**), as a minor product was detected, along with cWF (**1**) as the predominant product (ca. 85%). Feeding cWF (**1**) to the *gutD*₂₄₃₀₉ transformant led to the formation of **5a** in very low yield (ca. 9%). In contrast, the fed cWΔF (**8**) was completely converted to **4a** and **4b**. The biosynthetic pathway of guanitrypmycins in *S. monomycini* NRRL B-24309 and their non-enzymatic epimerization is depicted in **Scheme 5**. Moreover, this proposed pathway was also proven by heterologous expression of different gene combinations of the *gut*₃₅₈₉ cluster from *S. varsoviensis* NRRL B-3589. Since GutA₃₅₈₉ can produce both cWF (**1**) and cWY (**2**) with a ratio of 2:1 in *S. coelicolor* M1146, a more complex product spectrum, four chemical pairs **3a/3b**, **4a/4b**, **6a/6b**, **7a/7b** along with cWΔF (**8**) and cWΔY (**9**), was detected and obtained from this gene cluster (**Scheme 5**). Our results were further confirmed by a study performed by Shi and co-workers, in which a complementary strategy for the identification of purincyclamide in *Streptomyces chrestomyceticus* NA4264 was described.¹⁷⁹

To validate the productivity of guanitrypmycins in the native strains, *S. monomycini* NRRL B-24309 and *S. varsoviensis* NRRL B-3589 were cultivated in six different media. *S. monomycini* can produce **3a/3b** and **4a/4b** only in modified R5 medium. In comparison, neither any CDPs nor their derivatives could be detected in the extract of *S. varsoviensis* broth, indicating the *gut*₃₅₈₉ cluster is totally silent in its native host.

In vivo heterologous expression experiments have proven that GutD catalyzes the key C–C linkage between cWΔF (**8**)/cWΔY (**9**) and guanine during the biosynthesis of guanitrypmycins. To confirm the GutD function *in vitro*, the recombinant GutD₃₅₈₉ was successfully overproduced in *S. coelicolor* M1146, then purified as an N-(His)₁₀-fused protein to near homogeneity, and confirmed on SDS-PAGE (**Figure 16A**). The typical absorption maximum of the sodium dithionite reduced Fe^{II}-CO complex of GutD₃₅₈₉ was clearly observed at approximately 450 nm (**Figure 16B**). The GutD₃₅₈₉ activity was probed by using guanine and cWΔF (**8**) or cWΔY (**9**) as substrates, NADPH, spinach ferredoxin and ferredoxin-NADP⁺ reductase as cofactors for electron transport. As expected, **4a** and **4b** were clearly detected in the reaction mixture of cWΔF (**8**), while cWΔY (**9**) was converted to **7a** and **7b**.

RESULTS AND DISCUSSION



Both cWF and cWY derivatives were obtained from the *gut* cluster of *S. varsoviensis* NRRL B-3589, whereas only cWF derivatives were found for that from *S. monovici* NRRL B-24309. ***7b** was not obtained in pure form.

Scheme 5 Formation of guanitrypmycins in *S. monovici* NRRL B-24309 and *S. varsoviensis* NRRL B-3589.

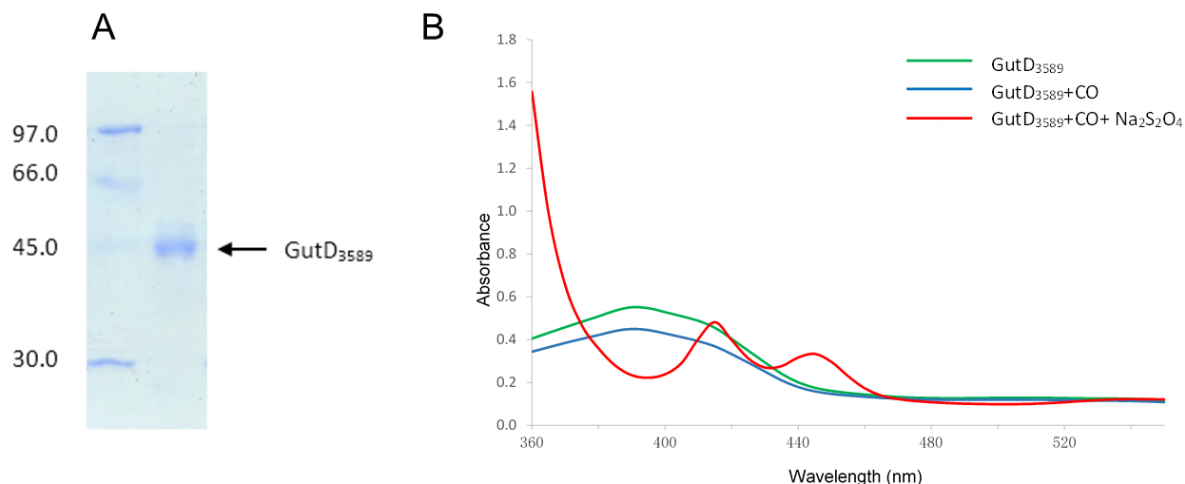


Figure 16 SDS-PAGE analysis of the purified GutD₃₅₈₉ (A) and the absorption spectra for GutD₃₅₈₉ and its ferrous-CO complex (B).

In conclusion, eight C3-guaninyl pyrroloindolines guanitrypmycins were identified from two homologous *p450*-associated *cdps*-containing gene clusters (*gut*) in *Streptomyces* (**Scheme 5**). Heterologous expression of different gene combinations in *S. coelicolor* proved that the CDP skeleton assembled by GutA initiates the biosynthesis of guanitrypmycins. Gut(BC) specifically introduces an exocyclic double bond merely to the phenylalanyl/tyrosyl hemisphere, but not the tryptophanyl side. Both *in vivo* expression and *in vitro* biochemical characterization revealed that the P450 GutD acts as the key enzyme and catalyzes the stereospecific coupling of cW Δ F (and cW Δ Y) with a guanine moiety *via* C3–C8' linkage. Furthermore, the MT GutE transfers a methyl group to N9' of the guaninyl residue after P450-catalyzed coupling reaction. Moreover, the non-enzymatic epimerization of **3a** and **4a** *via* keto–enol tautomerism increases the structural diversity of guanitrypmycins. In summary, this study provides an excellent example for the discovery of novel natural products by genome mining and exploring the proposed functions of individual biosynthetic enzymes *via* the heterologous expression approach.

[For details on this work, please see the publication \(section 4.2\)](#)

Jing Liu, Xiulan Xie, and Shu-Ming Li (2019) Guanitrypmycin biosynthetic pathways imply cytochrome P450-mediated regio- and stereospecific guaninyl transfer reactions. *Angewandte Chemie International Edition*, 58, 11534–11540, DOI: 10.1002/anie.201906891.

3.3 Expanding the spectrum of cytochrome P450s by identification of two distinct dimerases in CDPS-dependent pathways

Among the tailoring enzymes for the modification of DKPs assembled by CDPSs, P450s have been proven to catalyze diverse intriguing chemical transformations.¹⁷⁶ In addition to the aforementioned second project performed by this PhD candidate, Dr. Huili Yu has in parallel identified five P450 homologues involved in CDPS-dependent pathways. These include three P450s, P450_{NB5737} from *S. pupureus*, GutD₂₇₇₄ from *S. lavendulae*, as well as GutD₅₄₁₄ from *S. xanthophaeus*, for the formation of rare cWW adducts with the nucleobase guanine or hypoxanthine on different positions of the indole ring.^{158, 180, 181} The other two cytochrome P450 enzymes AspB and NasB from *Streptomyces* sp. NRRL S-1868, catalyze two new dimerization patterns of tryptophan-containing cyclodipeptides.¹⁸² Dimeric DKPs hold significant promise for medicinal chemistry, due to their chemical complexity and interesting biological activities. In recent years, great achievements have been obtained in the isolation and chemical synthesis of dimeric DKPs. In contrast, only four pathways for the formation of dimeric CDPSs have been characterized prior to this thesis, *i.e.*, the NRPS-dependent pathway of (+)-ditryptophenalanine in *Aspergillus flavus* and three CDPS-related ones in bacteria.^{159, 182, 183} The formation of (+)-nasesezaine C in *Streptomyces* sp. CMB-MQ03021 as well as (+)-nasesezaine A and (+)-aspergilazine A in *Streptomyces* sp. NRRL-S1868 is catalyzed by P450s *via* C–C bonds between C-5 or C-6 at the benzene ring of one unit and N1 or C3 of another one.^{159, 182}

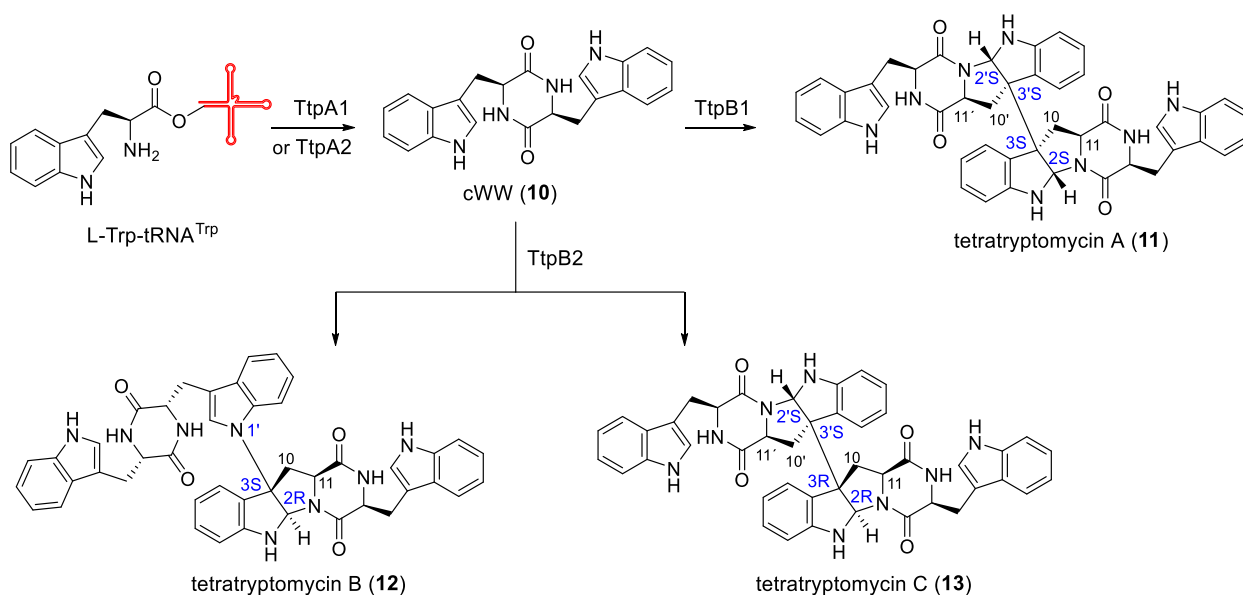
With the aim to explore more novel CDP modification enzymes, we analyzed numerous bacterial *cdps-p450*-containing clusters by using known proteins as probes. Genome mining revealed the presence of two *cdps-p450* operons comprising one *cdps* (*ttpA1* or *ttpA2*) and one (*ttpB2*) or two P450 genes (*ttpB1* and *ttpC1*) in *Saccharopolyspora antimicrobica* (**Figure 17**). Multiple sequence alignments revealed that the three putative P450s possess the conserved motifs as other bacterial P450 proteins. Phylogenetic analysis showed that TtpB1 and TtpB2, with sequence identity of 71% to each other, are located near to these characterized cyclodipeptide dimerases NasB, NascB and AspB, while TtpC1 is far away from this clade (**Figure 15**, TtpB1, TtpB2, and TtpC1 investigated in this study are labeled in bold blue). Furthermore, TtpB1/TtpB2 share a moderate sequence identities (approximately 40%) with the known dimerases. All these features made us curious about their roles in the CDP metabolism.



Figure 17 Genetic organizations of the two *ttp* gene clusters in *S. antimicrobica*

To verify their functions, the candidate genes and gene clusters were heterologously expressed in *S. albus* J1074. Firstly, the two *cdps* genes, *ttpA1* and *ttpA2* were cloned into the expression vector

pPWW50A and transformed into *S. albus* J1074. LC-MS analysis of the transformants revealed that both two CDPSs can assemble cWW (**10**) as the sole product. Afterwards, co-expression of *ttpA* with *ttpB1* and *ttpC1* led to the identification of a new peak **11** with a $[M + H]^+$ ion at $m/z 743.309 \pm 0.005$, corresponding well to that of a dimeric cWW. Detailed inspection of the NMR spectra revealed compound **11** to be a cWW homodimer with a C3–C3' linkage from the same side, thereby forming two hexahydropyrrolo[2,3-b]indole frameworks. The configurations of the new stereo centers of **11**, tetratryptomycin A, were determined as 2*S*, 3*S*, 2'*S*, and 3'*S* (**Scheme 6**). To further determine which P450 is responsible for the formation of **11**, *ttp(AB)1* and *ttp(AC)1* were separately expressed in *S. albus*. The transformant harboring *ttp(AB)1* showed the same product profile as that of *ttp(ABC)1* by comparison of molecular ions, retention time, MS² pattern and UV spectrum, while the *ttp(AC)1* transformant can only produce the CDPS product cWW (**10**). Furthermore, feeding **10** into to the *ttpB1* transformant led to the accumulation of **11**, whereas no consumption of **10** was observed in the transformant harboring *ttpC1*. This proved that TtpB1, but not TtpC1, is involved in the biosynthesis of **11** and indicated that the *ttp1* cluster comprises just *ttpA1* and *ttpB1*.



Scheme 6 The biosynthetic pathways of tetratryptomycins.

In analogy, the second gene cluster *ttp(AB)2* was also expressed in *S. albus*. LC-MS analysis revealed that the transformant harboring *ttp(AB)2* can produce two new cWW dimers **12** and **13** in a ratio of 20 : 1. Detailed interpretation of the NMR spectra revealed compound **12**, tetratryptomycin B, encompasses a C3–N1' connection between the two cWW monomers with a chirality of (2*R*, 3*S*) at the newly formed stereo centers (**Scheme 6**). Whereas, compound **13** is a stereoisomer of **11**, with a C3–C3' connection of cWW moieties. The newly formed stereo centers in **13** were proven to be (2*R*, 3*R*, 2'*S*, and 3'*S*) (**Scheme 6**). As expected, feeding **10** to the transformant with the P450 gene *ttpB2*

led to the detection of peaks **12** and **13** in a similar ratio (20 : 1) as in the *ttp(AB)2* transformant, which unequivocally proved TtpB2 as the second cWW dimerase from *S. antimicrobica*.

To further verify the functions of the two P450s *in vitro*, TtpB1 and TtpB2 were overproduced in *E. coli* BL21 (DE3) and *S. albus*, respectively, and purified to near homogeneity as confirmed on SDS-PAGE (**Figure 18A**). The purified recombinant TtpB1 has a brown color and shows an absorption shift from 420 to 450 nm after treatment with CO and Na₂S₂O₄ (**Figure 18B**). Incubation of TtpB1 with **10** in the presence of the commercial spinach ferredoxin, ferredoxin-NADP⁺ reductase, and NADPH led to the formation of the product peak **11**. In the case of TtpB2, under the same condition, **10** was consumed to give **12** and **13**. However, in the incubation mixture of TtpB1 or TtpB2 with other CDPs including cWY, cWF, cWP, cWL, cWA, cWG, cYY, and *cyclo*-L-Phe-L-Phe (cFF), no product formation was detected, proving the high substrate specificity of both P450s. A proposed mechanism involves radical mediated intramolecular cyclization and intermolecular addition reactions was postulated for the DKP coupling in the biosynthesis of tetratryptomycins (**Figure 19**). Moreover, the isolated compounds **11** - **13** were tested for their antibacterial activities against *E. coli*, *B. subtilis*, *Staphylococcus aureus* and *Pseudomonas aeruginosa*. However, no inhibitory activity was observed.

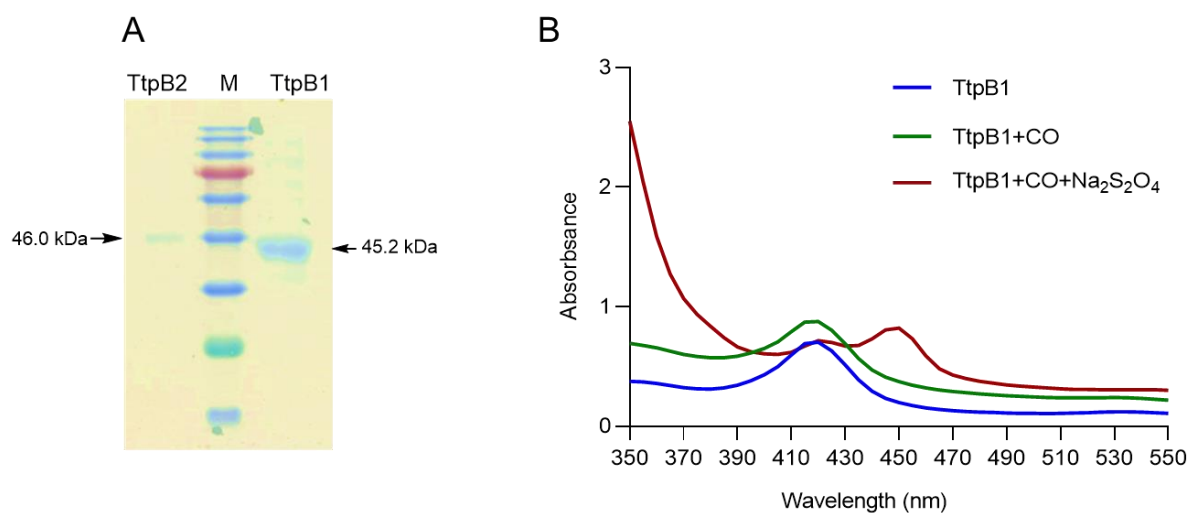


Figure 18 (A) SDS-PAGE analysis of the purified P450s and (B) UV-Vis spectroscopic analysis of TtpB1.

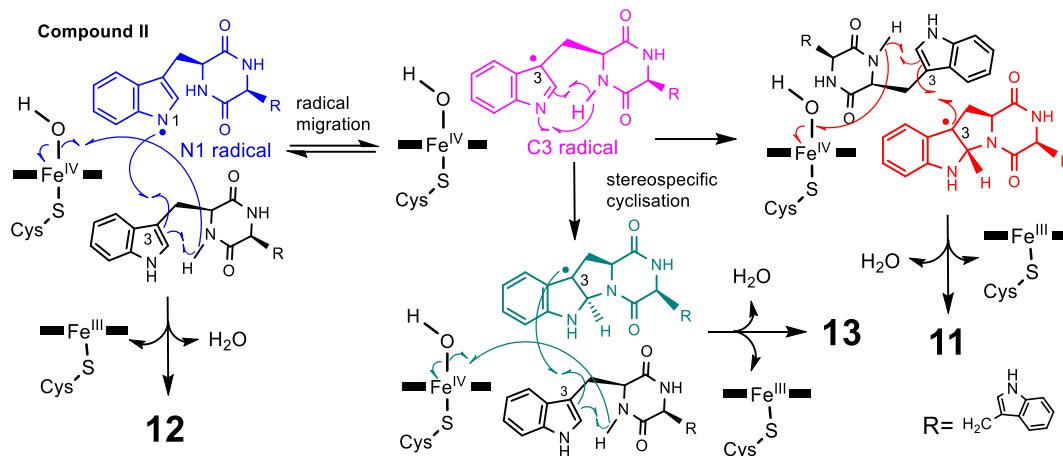


Figure 19 Proposed mechanism for P450-mediated intermolecular coupling reactions.

Taking the results together, two *cdps-p450*-containing operons *ttp1* and *ttp2* for the dimerization of the same CDP cWW were identified in *S. antimicrobica* via genome mining (**Scheme 6**). TtpB1 represents the first bacterial P450 to catalyze the stereospecific C3 (sp³)–C3' (sp³) bond formation between two CDPs, while TtpB2 is characterized as the first P450 to catalyze not only the unusual C3 (sp³)–N1' linkage, but also the intermolecular C3–C3' bond formation. It would be interesting to characterize more members of this intriguing enzyme group. Total product yields of **11** - **13** in their transformants were calculated to be 205, 200 and 9.5 mg/L, respectively. Therefore, our study provides a simple, direct, and efficient approach for enzymatic one-step preparation of structurally complex DKP dimers.

For details on this work, please see the publication (section 4.3)

Jing Liu, Xiulan Xie and Shu-Ming Li (2020). Increasing cytochrome P450 enzyme diversity by identification of two distinct cyclodipeptide dimerases. *Chemical Communications*, 56, 11042–11045, DOI: 10.1039/D0CC04772D.

4 Publications

4.1 Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from *Streptomyces* strains



Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from *Streptomyces* strains

Jing Liu¹ · Huili Yu¹ · Shu-Ming Li¹

Received: 17 January 2018 / Accepted: 28 February 2018 / Published online: 24 March 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Cyclodipeptide synthases (CDPSs) comprise normally 200–300 amino acid residues and are mainly found in bacteria. They hijack aminoacyl-tRNAs from the ribosomal machinery for cyclodipeptide formation. In this study, nine new CDPS genes from eight *Streptomyces* strains were cloned into pET28a vector and expressed in *Escherichia coli*. Structural elucidation of the isolated products led to the identification of one *cyclo*-L-Trp-L-Leu, two *cyclo*-L-Trp-L-Pro, and three *cyclo*-L-Trp-L-Trp synthases. Other three CDPSs produce *cyclo*-L-Trp-L-Ala or *cyclo*-L-Trp-L-Tyr as the major cyclodipeptide. Total product yields of 46 to 211 mg/L *E. coli* culture were obtained. Our findings represent rare examples of CDPS family derived from actinobacteria that form various tryptophan-containing cyclodipeptides. Furthermore, this study highlights the potential of the microbial machinery for tryptophan-containing cyclodipeptide biosynthesis and provides valid experimental basis for further combination of these CDPS genes with other modification genes in synthetic biology.

Keywords Aminoacyl t-RNA · Diketopiperazine · Cyclodipeptide synthase · *Streptomyces* · Tryptophan-containing cyclodipeptide

Introduction

Cyclodipeptides (CDPs) with a 2,5-diketopiperazine (DKP) skeleton are chemically condensation products of two amino acids and represent a class of secondary metabolites with simple scaffold, but different modification possibilities (Borthwick 2012). A large number of DKP derivatives show a broad range of biological and pharmacological activities such as antibacterial, antifungal, antiviral, and immunosuppressive activities (Borthwick 2012; Giessen et al. 2013), which make them attractive molecules for drug discovery and development. So far, DKP derivatives are mainly isolated from microorganism including bacteria and fungi. The DKP scaffold in fungi is usually assembled by bimodular

nonribosomal peptide synthetases (NRPSs) with a typical polypeptide chain length of 2300 amino acids (Maiya et al. 2006; Walsh 2016; Xu et al. 2014), whereas in bacteria mainly by tRNA-dependent cyclodipeptide synthases (CDPSs) comprising normally 200–300 amino acid residues (Giessen and Marahiel 2014; Gondry et al. 2009; James et al. 2015; Moutiez et al. 2017). NRPSs can use both proteinogenic and non-proteinogenic free amino acids as substrates, while CDPSs directly hijack the activated aminoacyl-tRNAs (aa-tRNAs) from the ribosomal machinery for cyclodipeptide formation (Huang et al. 2010).

The DKP scaffold is frequently modified by methylation, hydroxylation, prenylation, dimerization, and further cyclization (Borthwick 2012; Giessen et al. 2013; Li 2010; Xu et al. 2014). Among the DKPs, tryptophan-containing CDPs have received an increasing interest in recent years due to their promising scaffolds for structural modification (Li 2010; Xu et al. 2014). The electron-rich indole ring in tryptophanyl moiety can undergo different enzymatic and spontaneous modifications and rearrangement like prenylation, oxidation, methylation, and dehydrogenation to generate increased chemical complexity (Alkhalaf and Ryan 2015; Walsh 2014).

Since 2005, we have intensively studied prenyltransferases (PTs) involved in the biosynthesis of indole alkaloids in microorganisms, especially those for the prenylation of

Jing Liu and Huili Yu contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00253-018-8908-6>) contains supplementary material, which is available to authorized users.

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

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

tryptophan and tryptophan-containing CDPs (Winkelblech et al. 2015). The tryptophan-containing CDP PTs catalyzed regio- and stereospecific prenylations at the indole ring. PTs for specific prenylation at nearly all the positions of the indole ring have been characterized (Winkelblech et al. 2015; Wohlgenuth et al. 2017). Most of these PTs are from ascomycetous fungi and are involved in the modification of CDPs assembled by NRPSs. In the view of synthetic biology, it would be of essential importance to combine such PT genes with the small CDPS genes forming tryptophan-containing CDPs, especially those for *cyclo*-L-Trp-L-Pro (cWP), *cyclo*-L-Trp-L-Ala (cWA), *cyclo*-L-Trp-L-Leu (cWL), and *cyclo*-L-Trp-L-Trp (cWW). All these CDPs serve as precursors in the biosynthesis of diverse pharmacologically active compounds and are very well accepted by microbial PTs (Fan et al. 2015; Li 2010; Winkelblech et al. 2015).

By the end of July 2016, about 450 putative CDPS genes were identified by sequence homology search and their number is increasing steadily (Moutiez et al. 2017). Sixty-five of these genes have been proven to be responsible for the formation of different DKPs and about 30 are from actinomycetes (Brockmeyer and Li 2017; Jacques et al. 2015; James et al. 2015; Moutiez et al. 2017). At least ten of the identified CDPSs from eight actinomycetes catalyze the formation of cWW as the sole product (Fig. S1). These include Amir_4627 from *Actinosynnema mirum*, NozA, and NcdA from *Nocardioopsis* sp. CMB-M0232, whose products had been isolated and identified (Giessen et al. 2013; James et al. 2015). The products of other known cWW synthases were identified by LC-MS analysis (Jacques et al. 2015). CDPSs synthesizing other tryptophan-containing cyclodipeptides have not been reported in microorganisms. Until now, only one CDPS originating from the sea anemone *Nematostella vectensis* was demonstrated to produce a number of tryptophan-containing cyclodipeptides like cWL, cWA, *cyclo*-L-Trp-L-Phe (cWF), *cyclo*-L-Trp-L-Met (cWM), and *cyclo*-L-Trp-L-Gly (cWG) with very low product yields of 0.8–1.0 mg/L in *Escherichia coli* (Seguin et al. 2011). It is therefore absolutely essential to find new CDPSs for the formation of DKPs comprising tryptophan and another amino acid.

Materials and methods

Computer-assisted sequence analysis

Protein sequences listed in Figs. 1 and S1 were taken from the NCBI database (<http://www.ncbi.nlm.nih.gov/protein>) and compared with each other by using BLAST programs (<http://blast.ncbi.nlm.nih.gov/>). Multiple sequence alignments were carried out by using the program ClustalW and visualized with ESPript 3.0 (<http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>) to identify strictly conserved amino acid residues.

The phylogenetic trees in Figs. 1 and S1 were created by MEGA version 5.0 (<http://www.megasoftware.net>).

Bacterial strains, plasmids, and culture conditions

pGEM-T Easy used for cloning and pET28a (+) for expression experiments were obtained from Promega (Mannheim, Germany) and Novagen (Darmstadt, Germany), respectively. *E. coli* DH5 α (Invitrogen) was used as host for cloning and SoluBL21 (Genlantis) for expression experiments. They were grown at 37 °C in liquid Luria-Bertani (LB) medium (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) or on solid LB medium with 1.5% agar (*w/v*). For selection of recombinant *E. coli* strains, 100 μ g/mL of ampicillin or 50 μ g/mL of kanamycin were supplied in the media.

Streptomyces strains listed in Table S1 were kindly provided by ARS Culture Collection (NRRL) or purchased from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ). They were maintained at 28 °C in GYM (4.0 g/L glucose, 4.0 g/L yeast extract, 10.0 g/L malt extract, 15.0 g/L agar, and pH 7.2) or ISP4 medium (10.0 g/L soluble starch, 1.0 g/L K₂HPO₄, 1.0 g/L MgSO₄ 7H₂O, 1.0 g/L NaCl, 2.0 g/L (NH₄)₂SO₄, 1.0 g/L CaCO₃, 1.0 mL/L trace element solution, 15.0 g/L agar, and pH 7.2).

Plasmid construction and gene expression in *E. coli*

Genetic manipulation in *E. coli* was carried out according to the protocol by Sambrook and Russell (2001). Isolation of genomic DNA from *Streptomyces* was performed as described previously (Kieser et al. 2000). The genes encoding CDPSs were amplified by PCR from genomic DNA by using primers listed in Table 1 with Phusion® High-Fidelity DNA Polymerase from New England Biolabs (NEB). The generated PCR fragments were cloned into pGEM-T Easy vector and sequenced by SEQLAB (Göttingen, Germany) to confirm sequence integrity. After sequencing, the fragments were released with the appropriate restriction endonucleases from pGEM-T Easy vector and ligated into pET28a (+) vector, which was digested with the same enzymes previously. The generated plasmids (Table 1) were transformed into *E. coli* SoluBL21 for gene expression.

For CDPS overproduction, 0.5 mL of 16-h cultures of different expression constructs was used to inoculate 50 mL LB medium containing 50 μ g/mL kanamycin. The cultures were maintained at 230 rpm and 37 °C to an absorption at 600 nm of about 0.6. Isopropyl β -D-thiogalactopyranoside (IPTG) was then added to the cultures to a final concentration of 0.2 or 0.5 mM. The induction was kept at 20, 30, or 37 °C for 20 h for condition optimization (Table S2) or for 36 h for quantification. One milliliter of culture was extracted with one volume of ethyl acetate for three times. The organic phases were combined and evaporated, and the residues were afterward

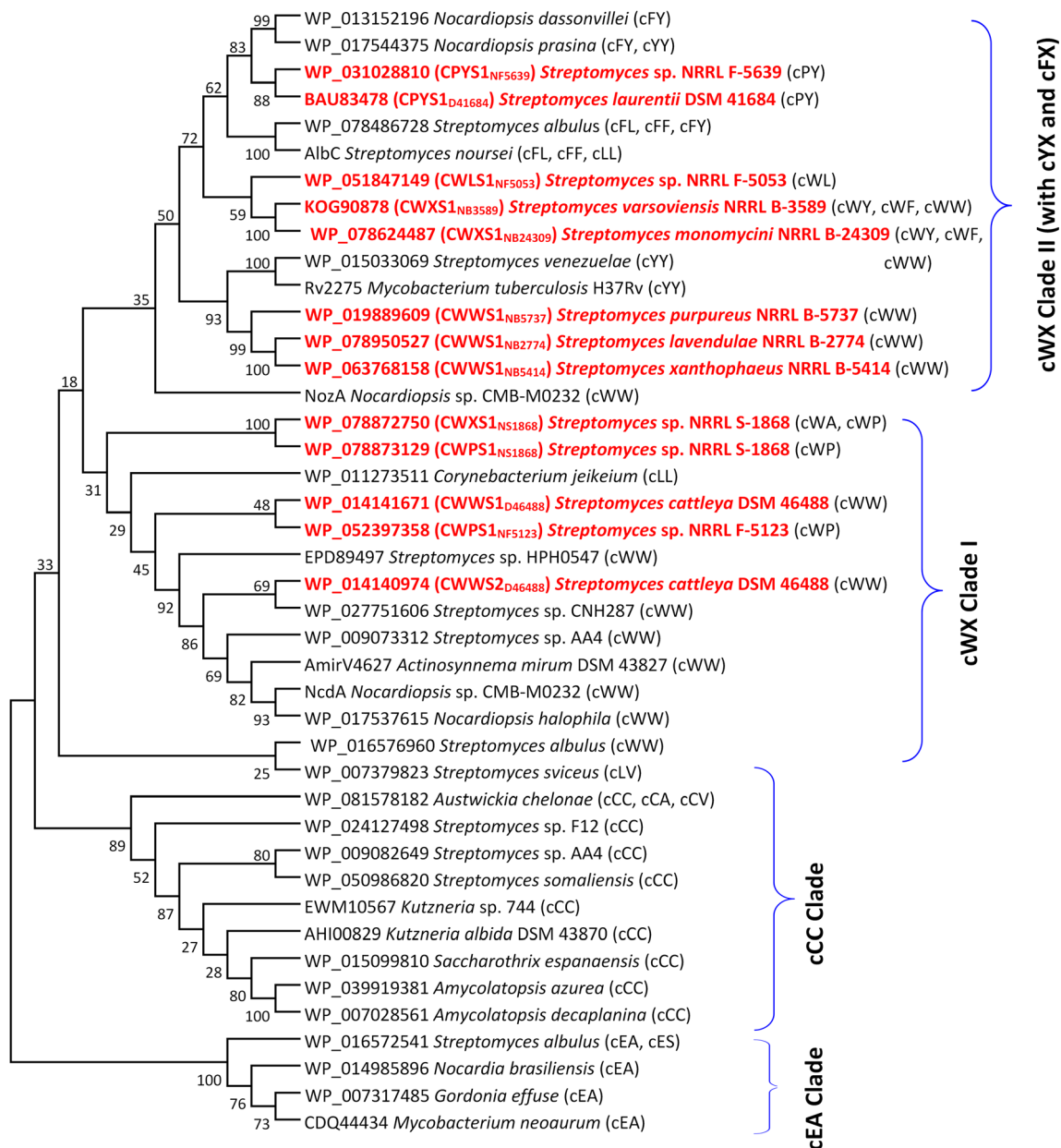


Fig. 1 Phylogenetic analysis of CDPSs from actinobacteria. The main CDPS products are given in parenthesis. Enzymes investigated in this study are highlighted in red

dissolved in 40 μ L of methanol. Five microliters of such samples were taken for LC-MS analysis.

HPLC and LC-MS analyses

The ethyl acetate extracts were analyzed on an Agilent HPLC 1260 series equipped with a photo diode array detector and a Bruker microTOF QIII mass spectrometer by using the Agilent Eclipse XDB C18 column (5 μ m, 4.6 \times 150 mm). A linear gradient of 5–100% acetonitrile in water, both containing 0.1% formic acid, in 40 min and a flow rate at 0.25 mL/min were used. The column was then washed with 100%

acetonitrile containing 0.1% formic acid for 5 min and equilibrated with 5% acetonitrile in water for 5 min. The parameters of the spectrometer were set as following: electrospray positive ion mode for ionization, capillary voltage with 4.5 kV, collision energy with 8.0 eV.

For quantification, an Agilent HPLC 1200 series equipped with a photo diode array detector and an identical column were used. A linear gradient of 10 to 100% acetonitrile in water without acid in 40 min was followed by 100% acetonitrile for 5 min and 10% acetonitrile in water for 5 min. The flow rate was 0.5 mL/min. The absorption at 280 nm was used for quantification.

Table 1 Information on CDPSs identified in this study

CDPS	Accession number	Protein length (aa)	Primer sequences (5'-3')	Cloning constructs in pGEM-T easy	Cloning sites in pET28a	Expression constructs
CWWS1 _{D46488}	WP_014141671	278	AACATATGCACGACAACGGTCATCGGCC TTGGATCCTTACGTCACCGGCT	pHY02	<u>NdeI</u> <u>BamHI</u>	pHY34
CWWS2 _{D46488}	WP_014140974	234	GCCATATGCTGCATCGAACGTCCTT TAGGATCCTCACGGCTCGGCGGCAGTT	pHY01	<u>NdeI</u> <u>BamHI</u>	pHY11
CWWS1 _{NB2774}	WP_078950527	265	CATATGACGATCACAGCTGACGCATCATTC GGATCCTCAAGCTGCTCGACGCTCAT	pHY72	<u>NdeI</u> <u>BamHI</u>	pHY82
CWWS1 _{NB5737}	WP_019889609	239	GCCATATGACTCTCATCGAAGACAC GGATCCTCAAGCGGCTCGACGGTCAT	pHY08	<u>NdeI</u> <u>BamHI</u>	pHY09
CWWS1 _{NB5414}	WP_063768158	273	CATATGAGGGCGATCACACAGGTGAC GGATCCTCAAGCTGCTCGACGCTCAT	pHY77	<u>NdeI</u> <u>BamHI</u>	pHY83
CWPS1 _{NS1868}	WP_078873129	241	CATATGAACACTTCCCTCGCTGC GAATTCTCAGGTTTCGGCCGCCGGTC	pHY103	<u>NdeI</u> <u>EcoRI</u>	pHY107
CWLS1 _{NF5053}	WP_051847149	243	CATATGTGCGAGGGCGCCGATGTGC GGATCCTCAGGATTCGTCACCGG	pJL03	<u>NdeI</u> <u>BamHI</u>	pJL26 ^a
CWPS1 _{NF5123}	WP_052397358	263	CATATGACCAGCAGGACCGAAAC GAATTCTCACGGAAGCAGCCGGGG	pHY100	<u>NdeI</u> <u>EcoRI</u>	pHY105
CWXS1 _{NS1868}	WP_078872750	267	CATATGGCCACACACGCCCTCCGC GAATTCTCACTGCTGCGTCACGTGGTC	pHY101	<u>NdeI</u> <u>EcoRI</u>	pHY106
CWXS1 _{NB3589}	KOG90878	255	CATATGGGGGCCCCGAGCCC GGATCCTCACGTCAAGTCCCTTTCTCC	pJL02	<u>NdeI</u> <u>BamHI</u>	pJL25
CWXS1 _{NB24309}	WP_078624487	282	GCGCATATGAGTGCATCGCAGGTGCTG GGATCCTCACGTACGTCACCCCTGC	pJL01	<u>NdeI</u> <u>BamHI</u>	pJL24 ^b
CPYS1 _{NF5639}	WP_031028810	278	CATATGAACCGCCGCTGTTCTTCG GGATCCTCATGCTCCTGGGGCACTG	pJL04	<u>NdeI</u> <u>BamHI</u>	pJL29
CPYS1 _{D41684}	BAU83478	224	CGCATATGAACCAGTTCGACGTGCTGCC GGATCCCCGGACCACGAGGAAGCC	pJL05	<u>NdeI</u> <u>BamHI</u>	pJL30 ^c

Restriction sites for cloning are underlined in primer sequences

^a The 12 base pairs at the 5'-end of the sequence coding for WP_051847149 were not included in the expression construct pJL26

^b The 96 base pairs at the 5'-end of the sequence coding for WP_078624487 were not included in the expression construct pJL24

^c According to the annotation in database, BAU83478 comprises 484 amino acids and consists of CDPS sequence at its N-terminus and that for a putative methyltransferase at its C-terminus. The sequence coding for amino acids 1-224 of BAU83478 was cloned into pET28a generating pJL30

Isolation and structural elucidation of generated cyclodipeptides

For structural elucidation of the accumulated CDPs, *E. coli* SoluBL21 cultures carrying the respective expression constructs were prepared under the best tested conditions (Table S2) and extracted with ethyl acetate. The CDPs were then isolated on an Agilent HPLC 1260 series equipped with a photo diode array detector by using a semi-preparative Multospher 120 RP-18 column (10 × 250 mm, 5 μm) with the same solvents and gradient as for quantification. The flow rate was set to 2.5 mL/min. The obtained products were purified on the same column with slightly improved gradients, if necessary. In this study, 5, 3, and 4 mg of cWW were isolated as white powders from 300 mL *E. coli* cultures with pHY34, pHY11, and pHY09, respectively. Thirty milligrams of cWP and 12 mg of cWL were isolated as white powders from 1000 mL *E. coli*/pHY107 and *E. coli*/pJL26, respectively. From 2000 mL *E. coli* culture with pJL25, 1.5 mg of cWA, 10 mg of *cyclo*-L-Trp-L-Tyr (cWY), 5 mg of cWM, 2.5 mg of *cyclo*-L-Trp-L-Val (cWV), and 2.5 mg of cWF were obtained as white powders.

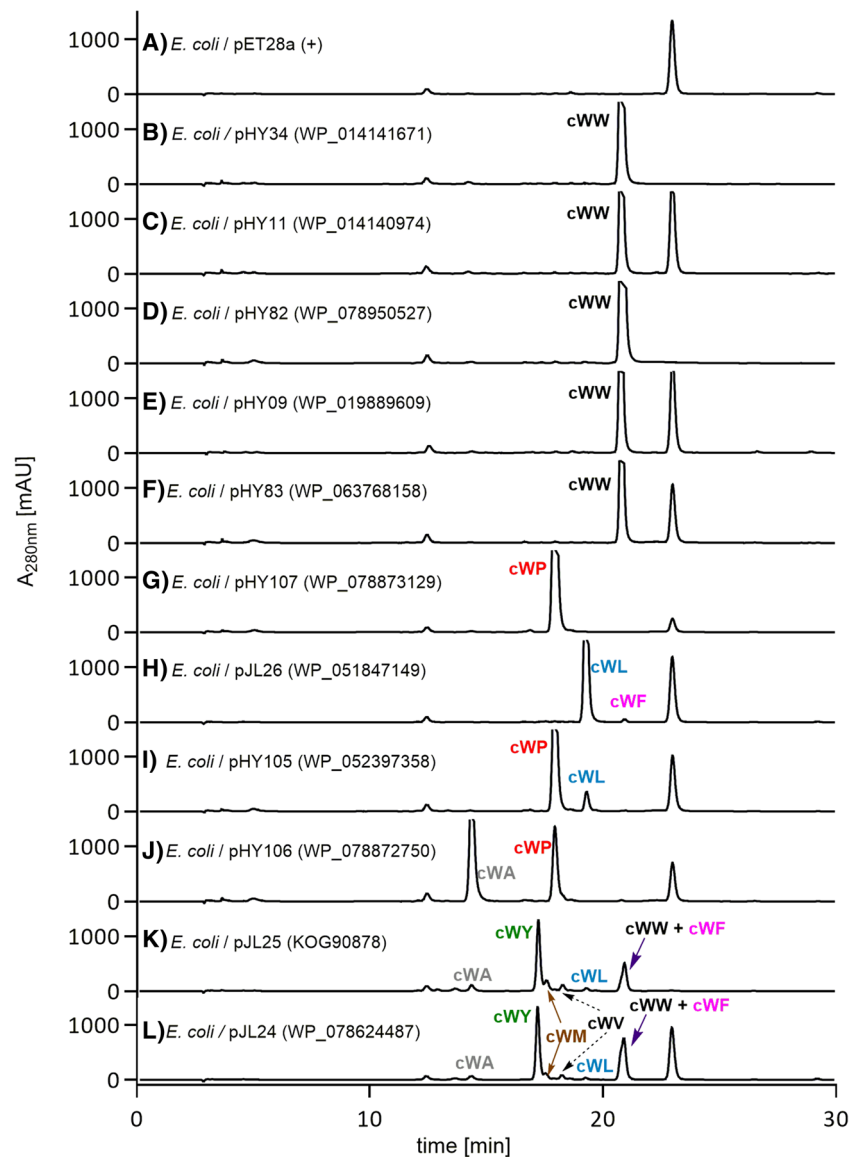
The isolated compounds were subjected to NMR and MS analyses. ¹H NMR spectra (Figs. S2–S9) were recorded at room temperature on an ECX-500 spectrometer (JEOL, Tokyo, Japan). All spectra were processed with MestReNova 5.2.2 (Metrelab Research, S5 Santiago de Compostella, Spain). MS and NMR data of the cyclodipeptides are provided as Tables S3 and S4. Interpretation of the NMR and MS data of the isolated products and comparison with those of known compounds led to the undoubted identification of the cyclodipeptides.

Results

Phylogenetic analysis of CDPSs

Phylogenetic analysis of known CDPSs from actinomycetes (Fig. S1) revealed that the known cWW synthases are located close together, with the exception for WP_016576960 from *Streptomyces albulus*. CDPSs for DKPs consisting of aliphatic amino acids like *cyclo*-L-Cys-L-Cys (cCC) and *cyclo*-L-Glu-L-Ala (cEA) and those

Fig. 2 LC-MS analysis of recombinant *E. coli* strains with empty vector (a) or with CDPSs (b–l). The ethyl acetate extracts were analyzed on an Agilent HPLC series 1260 with UV and mass detections. Absorptions at 280 nm are illustrated in this figure. The products are predicted by inspection of their exact $[M + H]^+$ ions and proven by NMR analysis after isolation



comprising aromatic amino acids such as *cyclo*-L-Tyr-L-Tyr (cYY) and *cyclo*-L-Tyr-L-X (cYX) build their own groups. To get CDPSs for other tryptophan-containing DKPs rather than cWW, we searched in databases for homologs of the actinobacterial cWW synthases and integrated them into Fig. S1 generating Fig. 1. Thirteen candidate CDPSs close to cWW and cYY synthases from 11 *Streptomyces* strains (highlighted in red in Fig. 1) were selected as investigation objectives.

Overproduction of CDPSs in *E. coli* and identification of the products

For functional proof, we established our expression system in *E. coli* SoluBL21 by using the two known CDPSs WP_014141671 and WP_014140974 from *Streptomyces cattleya* DSM 46488, termed CWWS1_{D46488} (cWW synthase

1 from strain DSM 46488) and CWWS2_{D46488} in this study, respectively (Fig. 1). After optimization of expression conditions (Table S2), the 20-h-old cultures after induction at 30 °C with 0.5 mM IPTG for CWWS1_{D46488} and 0.2 mM for CWWS2_{D46488} were extracted with ethyl acetate and then analyzed on LC-MS. In comparison to the culture of *E. coli* harboring the empty vector (Fig. 2a), one additional predominant peak each with the expected $[M + H]^+$ ion at m/z 373.1636 for cWW was detected in *E. coli* transformants (Fig. 2b, c). Isolation and structure elucidation of both peaks with the help of NMR confirmed their integrity as cWW (Tables S3 and S4, Fig. S2).

In analogy, the other nine candidate genes were individually cloned into pET28a (+) (Table 1) and expressed in *E. coli* SoluBL21 under different optimized conditions (Table S2). The ethyl acetate extracts of the 20-h cultures were analyzed on LC-MS. In comparison to that of negative control (Fig. 2a),

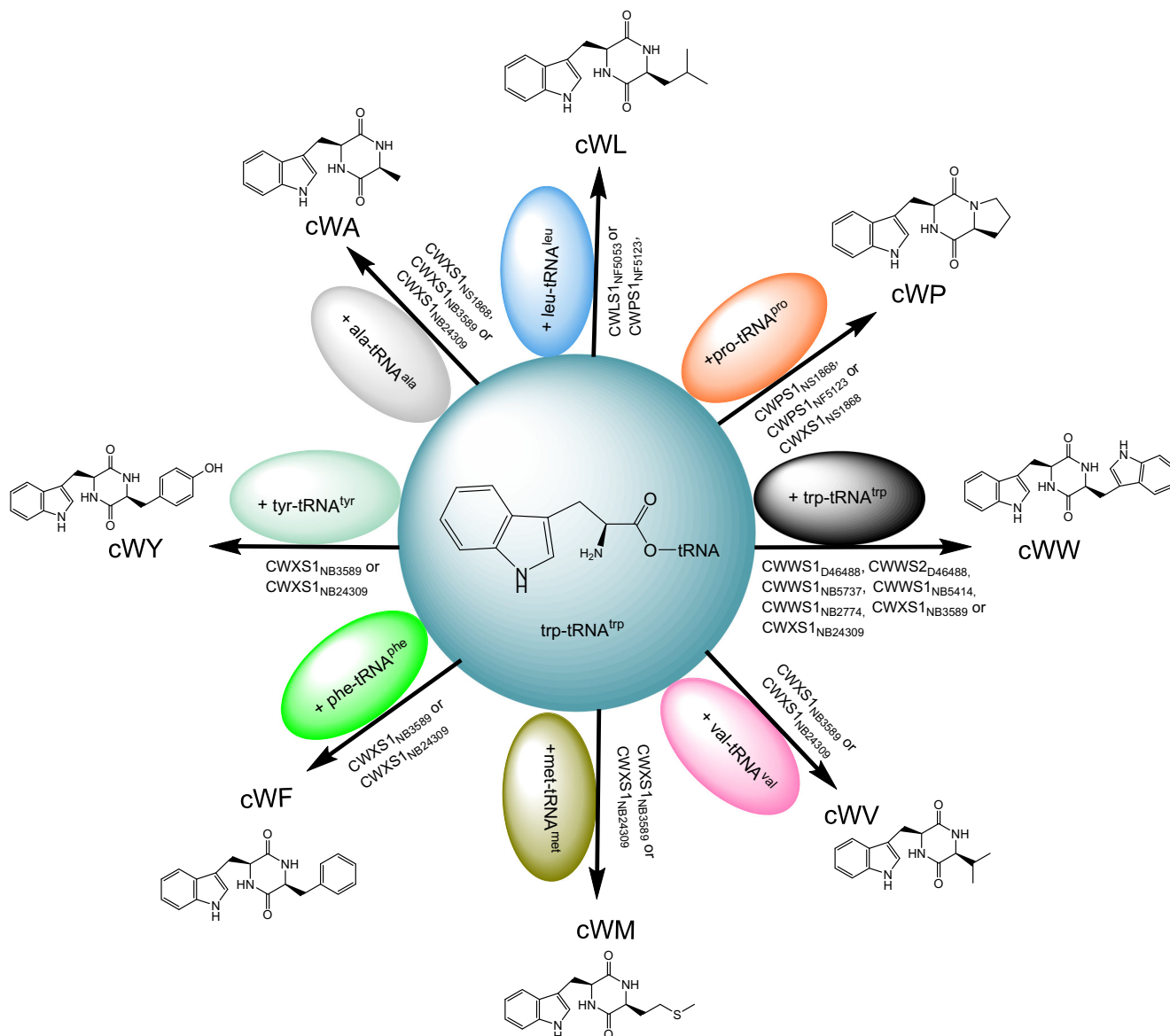


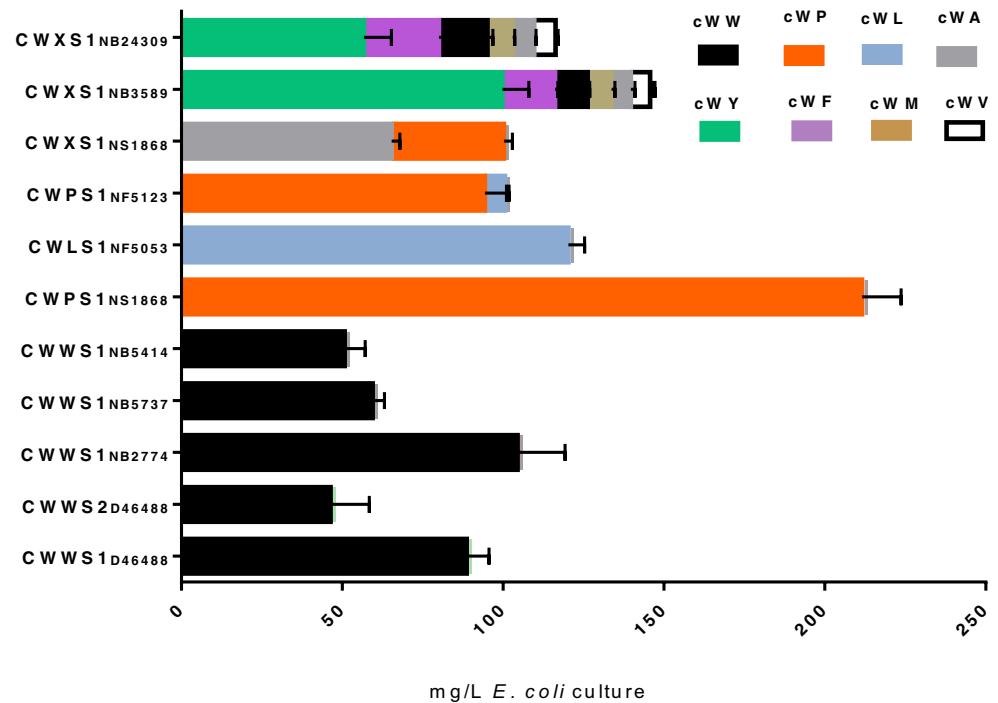
Fig. 3 Formation of tryptophan-containing cyclodipeptides by CDPSs

the chromatograms of five CDPS transformants showed one predominant product peak each (Fig. 2d–h). These include WP_078950527 from *Streptomyces lavendulae* NRRL B-2774 (Fig. 2d), WP_019889609 from *Streptomyces purpureus* NRRL B-5737 (Fig. 2e), WP_063768158 from *Streptomyces xanthophaeus* NRRL B-5414 (Fig. 2f), WP_078873129 from *Streptomyces* sp. NRRL S-1868 (Fig. 2g), and WP_051847149 from *Streptomyces* sp. NRRL F-5053 (Fig. 2h). The products of the first three transformants (Fig. 2d–f) share the same retention time and $[M + H]^+$ ion with those of the two known CWWS1_{D46488} and CWWS2_{D46488} (Fig. 2b, c), indicating the presence of cWW. Interpretation of the NMR and MS data of the isolated peak from the culture with PHY09

(Tables S3 and S4, Fig. S2) confirmed these products to be cWW (Giessen et al. 2013; Lu et al. 2017). Therefore, these three enzymes were termed CWWS1_{NB2774}, CWWS1_{NB5737}, and CWWS1_{NB5414}, respectively (Fig. 3).

For structure elucidation, the products of WP_078873129 (Fig. 2g) and WP_051847149 (Fig. 2h) were isolated from the bacterial cultures and subjected to NMR and MS analyses (Tables S3 and S4; Figs. S3 and S4), which proved unequivocally cWP (Grundmann and Li 2005; He et al. 2013; Kumar et al. 2014) as the sole product of WP_078873129, termed hereafter CWPS1_{NS1868} (Fig. 3). The two products of WP_051847149 with a ratio of 120:1 were identified as cWL (Kumar et al. 2014) and cWF (Chu et al. 2011),

Fig. 4 Product yields of CDPSs. The data were obtained from three independent 36-h old cultures



respectively. Due to the significant dominance of cWL, this enzyme was named as CWLS1_{NF5053} (Fig. 3).

Two product peaks each were detected in the transformants with WP_052397358 from *Streptomyces* sp. NRRL F-5123 (Fig. 2i) and another CDPS WP_078872750 from *Streptomyces* sp. NRRL S-1868 (Fig. 2j). One of them was found in both extracts and identified as cWP by comparison of its retention time and $[M+H]^+$ ion with those of CWPS1_{NS1868} product (Fig. 2g). cWP in the transformant with WP_052397358 contributes about 94% to the total CDPSs, and this enzyme was therefore termed CWPS1_{NF5123}. The minor product of CWPS1_{NF5123} showed a $[M+H]^+$ of cWL. Another main product in *E. coli* with pHY106 (Fig. 2j) had a $[M+H]^+$ of cWA, which was identified by comparison with the product from KOG90878 of *Streptomyces varsoviensis* NRRL B-3589 (Fig. 2k). WP_078872750 was therefore termed CWXS1_{NS1868}.

The transformant with KOG90878 had a similar and complex product spectrum as WP_078624487 from *Streptomyces monomycini* NRRL B-24309 (Fig. 2l). Two predominant and at least four minor peaks were detected for both cultures at 280 nm. They differ slightly from each other in the product yields and ratios, especially those of the two dominant peaks. Product isolation from the culture with KOG90878 and structural elucidation proved the main peak at 15.4 min to be cWY and the second major peak as a mixture of cWF and cWW. The four minor peaks were identified as cWA, cWM, cWV, and cWL, respectively (Tables S3 and S4; Figs. S5–S9).

KOG90878 and WP_078624487 are therefore responsible for the formation of at least seven tryptophan-containing cyclodipeptides and termed CWXS1_{NB3589} and CWXS1_{NB24309}, respectively (Fig. 3).

Product yields of CDPSs

Generally, high product yield is a prerequisite for potential application in the biotechnology. The product yields of cWW from *E. coli* culture at 29 and 42 mg/L were reported for the known cWW synthases Amir_4627 and NozA after codon optimization, respectively (Giessen et al. 2013; James et al. 2015). To prove the productivity of the recombinant strains, we determined the CDP contents of 36-h-old cultures by HPLC analysis using the isolated products as standards. As shown in Fig. 4, the product yields achieved in this study for several enzymes are much higher than those mentioned above. The product yields of the five cWW synthases were detected between 46 ± 10 and 104 ± 10 mg/L with the best one for CWWS1_{NB2774}. The highest product yield of 211 ± 12 mg cWP per liter culture was observed for the culture with CWPS1_{NS1868}, followed by 120 ± 5 mg cWL in the culture with CWLS1_{NF5053}. Total product yields between 100 and 140 mg/L were calculated for other four CDPSs. These data provide a solid basis for further combination of these CDPS genes with other modification genes in synthetic biology.

Table 2 Amino acid residues of selected CDPSs constituting the two binding pockets

	Amino acids constituting binding pocket P1														Amino acids constituting binding pocket P2														Reported product	Predicted product*	Identified in this study
AlbC	L ₃₃	G ₃₅	V ₆₅	V ₆₇	L ₁₁₉	L ₁₈₅	F ₁₈₆	L ₂₀₀	M ₁₅₂	A ₁₅₅	V ₁₅₆	N ₁₅₉	I ₂₀₄	T ₂₀₆	P ₂₀₇	cFL, cFF, cFY															
Amir_4627	V ₃₃	G ₃₅	V ₆₅	P ₆₇	M ₁₁₉	F ₁₈₄	F ₁₈₅	N ₁₉₉	M ₁₅₂	E ₁₅₅	V ₁₅₆	L ₁₅₉	R ₂₀₃	L ₂₀₅	P ₂₀₆	cWW (Giessen et al. 2013)															
NozA	L ₃₂	G ₃₄	V ₆₄	A ₆₆	L ₁₁₈	F ₁₈₃	F ₁₈₄	S ₁₉₈	M ₁₅₁	A ₁₅₄	V ₁₅₅	S ₁₅₈	M ₂₀₂	L ₂₀₄	P ₂₀₅	cWW (James et al. 2015)															
NcdA	V ₃₆	G ₃₈	V ₅₈	P ₆₀	M ₁₁₂	F ₁₇₇	F ₁₇₈	N ₁₉₂	M ₁₄₅	D ₁₄₈	V ₁₄₉	A ₁₅₂	R ₁₉₆	L ₁₉₈	P ₁₉₉	cWW (James et al. 2015)															
WP_016576960	L ₅₃	G ₅₅	I ₈₅	P ₈₇	L ₁₃₉	L ₂₀₄	C ₂₀₅	H ₂₁₉	A ₁₇	K ₁₇₅	A ₁₇₆	S ₁₇₉	Q ₂₂₃	I ₂₂₅	P ₂₂₆	cWW (Jacques et al. 2015)															
WP_009073312	V ₂₆	G ₂₈	I ₅₈	P ₆₀	M ₁₁₂	F ₁₇₇	F ₁₇₈	N ₁₉₂	M ₁₄₅	E ₁₄₈	V ₁₄₉	A ₁₅₂	Q ₁₉₆	L ₁₉₈	P ₁₉₇	cWW (Jacques et al. 2015)															
EPD89497	V ₂₉	G ₃₁	I ₆₁	P ₆₃	M ₁₁₅	F ₁₈₀	F ₁₈₁	C ₁₉₅	A ₁₄₈	T ₁₅₁	V ₁₅₂	A ₁₅₅	R ₁₉₉	L ₂₀₁	P ₁₉₈	cWW (Jacques et al. 2015)															
WP_027751606	V ₂₅	G ₂₇	I ₅₇	P ₅₉	M ₁₁₁	F ₁₇₆	F ₁₇₇	N ₁₉₁	M ₁₄₄	D ₁₄₇	V ₁₄₈	A ₁₅₁	Q ₁₉₅	L ₁₉₇	P ₂₀₈	cWW (Jacques et al. 2015)															
WP_017537615	V ₃₅	G ₃₇	I ₆₇	P ₆₉	M ₁₂₁	F ₁₈₆	F ₁₈₇	N ₂₀₁	M ₁₅₄	E ₁₅₇	V ₁₅₉	A ₁₇₁	R ₂₀₅	L ₂₀₇	P ₁₉₇	cWW (Jacques et al. 2015)															
CWWS1 _{D46488}	V ₇₇	G ₇₉	V ₁₀₉	P ₁₁₁	L ₁₆₄	F ₂₂₉	F ₂₃₀	C ₂₄₂	M ₁₉₇	A ₂₀₀	A ₂₀₁	S ₂₀₄	K ₂₄₆	V ₂₄₈	P ₂₄₉	cWW (Jacques et al. 2015)		cWW													
CWWS2 _{D46488}	V ₂₇	G ₂₉	V ₅₉	P ₆₁	M ₁₁₃	F ₁₇₈	F ₁₇₉	C ₁₉₃	M ₁₄₆	E ₁₄₉	A ₁₅₀	A ₁₅₃	R ₁₉₇	I ₁₉₉	P ₂₀₀	cWW (Jacques et al. 2015)		cWW													
CWWS1 _{NB2774}	L ₂₉	G ₃₁	V ₆₁	A ₆₃	L ₁₁₅	F ₁₈₁	F ₁₈₂	A ₁₉₆	M ₁₄₈	H ₁₅₁	F ₁₅₂	S ₁₅₅	V ₂₀₀	M ₂₀₂	P ₂₀₃	<i>cFX</i>	cWW														
CWWS1 _{NB5737}	L ₂₈	G ₃₀	V ₆₀	A ₆₂	L ₁₁₄	F ₁₈₀	F ₁₈₁	S ₁₉₅	M ₁₄₇	T ₁₅₀	F ₁₅₁	P ₁₅₄	Q ₁₉₉	I ₂₀₁	P ₂₀₂	cFX	cWW														
CWWS1 _{NB5414}	L ₃₅	G ₃₇	V ₆₇	A ₆₉	L ₁₂₁	F ₁₈₇	F ₁₈₈	A ₂₀₀	M ₁₅₄	H ₁₅₇	F ₁₅₈	S ₁₆₁	V ₂₀₄	M ₂₀₆	P ₂₀₇	<i>cFX</i>	cWW														
CWPS1 _{NS1868}	L ₃₂	G ₃₄	I ₆₄	P ₆₆	L ₁₁₈	F ₁₈₉	F ₁₉₀	V ₂₀₄	A ₁₅₁	R ₁₅₄	A ₁₅₅	V ₁₅₈	M ₂₀₈	V ₂₁₀	P ₂₁₁	<i>cFX</i>	cWP														
CWLS1 _{NF5053}	L ₃₆	G ₃₈	V ₆₈	A ₇₀	L ₁₂₂	F ₁₈₈	L ₁₈₉	S ₂₀₃	M ₁₅₅	S ₁₅₈	F ₁₅₉	G ₁₆₂	M ₂₀₇	A ₂₀₉	Q ₂₁₀	cFX	cWL														
CWPS1 _{NF5123}	L ₄₂	G ₄₄	V ₇₄	P ₇₆	L ₁₂₈	L ₁₉₃	L ₁₉₄	A ₂₀₈	A ₁₆₁	S ₁₆₄	V ₁₆₅	G ₁₆₈	K ₂₁₂	I ₂₁₄	D ₂₁₅	<i>cFX</i>	cWP														
CWXS1 _{NS1868}	L ₄₃	G ₄₅	I ₇₅	P ₇₇	L ₁₂₉	F ₂₀₅	F ₂₀₆	A ₂₂₀	L ₁₆₂	R ₁₆₅	A ₁₆₆	V ₁₆₉	A ₂₂₄	V ₂₂₆	G ₂₂₇	<i>cFX</i>	cWA, cWP														
CWXS1 _{NB3589}	L ₃₀	G ₃₂	I ₆₂	A ₆₄	L ₁₁₆	F ₁₉₃	F ₁₉₄	S ₂₀₈	M ₁₄₉	R ₁₅₂	H ₁₅₃	G ₁₅₆	M ₂₁₂	T ₂₁₄	P ₂₁₅	cFX	cWY, cWF, cWW														
CWXS1 _{NB24309}	–	–	I ₃₀	A ₃₂	L ₈₄	F ₁₇₇	F ₁₇₈	S ₁₉₂	M ₁₁₇	I ₁₂₀	H ₁₂₁	G ₁₂₄	L ₁₉₆	T ₁₉₈	P ₁₉₉	cFX	cWY, cWF, cWW														

*The products of the CDPSs of interest were predicted in a previous report (in bold) (Jacques et al. 2015) and in this study (in italic) according to the conserved residues suggested by Jacques et al. (Jacques et al. 2015)

Discussion

To find candidates for the formation of tryptophan-containing CDPSs, we searched the database by using known cWW synthases as probes and carried out phylogenetic analysis. Gene expression and identification of CDP products led to identification of nine new members for the biosynthesis of tryptophan-containing CDPSs. Inspection the phylogenetic tree (Fig. 1) revealed that the tryptophan-containing cyclodipeptide synthases from actinobacteria are located in two clades. With the exception for *cyclo*-L-Leu-L-Leu (cLL), the clade I comprises only cWX synthases, while the members of clade II are responsible for the formation of not only cWX, but also cYX and cFX. To prove the diversity of cWX clade II, we took two genes close to cYX synthases from this clade and expressed them in *E. coli*. As shown in Fig. S10, WP_031028810 from *Streptomyces* sp. NRRL F-5639 and BAU83478 from *Streptomyces laurentii* DSM 41684 were

proven to be responsible for the formation of *cyclo*-L-Pro-L-Tyr (cPY). Together with the CDPSs for tryptophan-containing cyclodipeptide biosynthesis (Figs. 2, 3, and 4), these results provide evidence that phylogenetic analysis could be successfully used for prediction and finding of desired CDPS candidates. However, exact CDPS product cannot be predicted by just phylogenetic analysis. For example, the 13 specific cWW synthases are distributed in two cWX clades. One of the cWP synthases, CWPS1_{NF5123} is closer to CWWS1_{D46488} than to the second one CWPS1_{NS1868}, which is again closer to CWXS1_{NS1868}. CWPS1_{NS1868} and CWXS1_{NS1868} share a sequence identity of 62% on the amino acid level, and the latter produces a mixture of cWP and cWA with a ratio of 15:7 (Figs. 2 and 4).

Based on crystal structure and mutagenesis experiments, Moutiez et al. proposed that two pockets of CDPSs are for aminoacyl-tRNA binding (Jacques et al. 2015; Moutiez et al. 2014a; Moutiez et al. 2014b). It was postulated that P1 is more

specific for binding the first aminoacyl-tRNA, and the wider pocket P2 is responsible for the less specific binding of the second and often variable aminoacyl-tRNA (Moutiez et al. 2014a; Moutiez et al. 2014b). The amino acid residues of the two pockets were used for prediction of CDPS products (Jacques et al. 2015). With the help of the residues in the two pockets, the functions of the nine tryptophan-containing CDPSs identified in this study were predicted to be responsible for the formation of cFX (Table 2) (Jacques et al. 2015). However, as shown in Figs. 2, 3, and 4, only CWLS1_{NF5053}, CWXS1_{NB3589}, and CWXS1_{NB24309} produce cWF as minor or second major product, with 1, 11, and 22% of total CDPs, respectively. No cFX was detected in other CDPS transformants. Comparing the amino acid residues in pockets 1 and 2 of tryptophan-containing cyclodipeptide synthases with each other and with those of *cyclo*-L-Phe-L-Leu (cFL) synthase AlbC (Table 2) revealed no clear indications for cWX synthases. In a previous study, we demonstrated that mutation of the related residues in the pocket P1 did not lead to the desired changes of product spectrum, but their product yields (Brockmeyer and Li 2017). This indicates that these prediction tools still need to be optimized by identification of more CDPS structures and amino acid residues in the binding pockets.

In summary, after confirmation of two known cWW synthases, we identified three new cWW, one cWL, and two cWP synthases from *Streptomyces*, which catalyze the formation of one predominant DKP. Three additional CDPSs produce one DKP like cWA or cWY as the major metabolite (Figs. 2 and 4). These CDPSs are thus ideal candidates for production of tryptophan-containing CDPs for further modification by tailoring enzymes like PTs. It should be also mentioned that cWV was not reported as CDPS product prior to this study. Therefore, our results expand significantly the product spectrum of tryptophan-containing cyclodipeptide synthases and raise hopes to find other members of this important enzyme group.

Acknowledgements We thank ARS Culture Collection (NRRL) for providing *Streptomyces* strains, S. Newel, and R. Kraut (University Marburg) for taking NMR and MS spectra.

Funding information The Bruker micrOTOF QIII mass spectrometer was financially supported in part by a grant from the Deutsche Forschungsgemeinschaft (INST 160/620-1 to S.-M. L.). J.L. and H.Y. are scholarship recipients of China Scholarship Council (201608310118 and 201306220024).

Compliance with ethical standards

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

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

References

- Alkhalaf LM, Ryan KS (2015) Biosynthetic manipulation of tryptophan in bacteria: pathways and mechanisms. *Chem Biol* 22:317–328
- Borthwick AD (2012) 2,5-diketopiperazines: synthesis, reactions, medicinal chemistry, and bioactive natural products. *Chem Rev* 112:3641–3716
- Brockmeyer K, Li S-M (2017) Mutations of residues in pocket P1 of a cyclodipeptide synthase strongly increase product formation. *J Nat Prod* 80:2917–2922
- Chu D, Peng C, Ding B, Liu F, Zhang F, Lin H, Li Z (2011) Biological active metabolite *cyclo* (L-Trp-L-Phe) produced by South China Sea sponge *Holoxea* sp. associated fungus *Aspergillus versicolor* strain TS08. *Bioprocess Biosyst Eng* 34:223–229
- Fan A, Winkelblech J, Li S-M (2015) Impacts and perspectives of prenyltransferases of the DMATS superfamily for use in biotechnology. *Appl Microbiol Biotechnol* 99:7399–7415
- Giessen TW, Marahiel MA (2014) The tRNA-dependent biosynthesis of modified cyclic dipeptides. *Int J Mol Sci* 15:14610–14631
- Giessen TW, von Tesmar AM, Marahiel MA (2013) A tRNA-dependent two-enzyme pathway for the generation of singly and doubly methylated ditryptophan 2,5-diketopiperazines. *Biochemistry* 52:4274–4283
- Gondry M, Sauguet L, Belin P, Thai R, Amouroux R, Tellier C, Tuphile K, Jacquet M, Braud S, Courcon M, Masson C, Dubois S, Lautru S, Lecoq A, Hashimoto S, Genet R, Pernodet JL (2009) Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nat Chem Biol* 5:414–420
- Grundmann A, Li S-M (2005) Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from *Aspergillus fumigatus*. *Microbiology* 151:2199–2207
- He F, Bao J, Zhang XY, Tu ZC, Shi YM, Qi SH (2013) Asperterrestide A, a cytotoxic cyclic tetrapeptide from the marine-derived fungus *Aspergillus terreus* SCSGAF0162. *J Nat Prod* 76:1182–1186
- Huang R, Zhou X, Xu T, Yang X, Liu Y (2010) Diketopiperazines from marine organisms. *Chem Biodivers* 7:2809–2829
- Jacques IB, Moutiez M, Witwinowski J, Darbon E, Martel C, Seguin J, Favry E, Thai R, Lecoq A, Dubois S, Pernodet JL, Gondry M, Belin P (2015) Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat Chem Biol* 11:721–727
- James ED, Knuckley B, Alqahtani N, Porwal S, Ban J, Karty JA, Viswanathan R, Lane AL (2015) Two distinct cyclodipeptide synthases from a marine actinomycete catalyze biosynthesis of the same diketopiperazine natural product. *ACS Synth Biol* 5:547–553
- Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA (2000) Practical *Streptomyces* genetics. John Innes Foundation, Norwich
- Kumar SN, Mohandas C, Nambisan B (2014) Purification, structural elucidation and bioactivity of tryptophan containing diketopiperazines, from *Comamonas testosteroni* associated with a rhabditid entomopathogenic nematode against major human-pathogenic bacteria. *Peptides* 53:48–58
- Li S-M (2010) Prenylated indole derivatives from fungi: structure diversity, biological activities, biosynthesis and chemoenzymatic synthesis. *Nat Prod Rep* 27:57–78
- Lu C, Xie F, Shan C, Shen Y (2017) Two novel cyclic hexapeptides from the genetically engineered *Actinosynnema pretiosum*. *Appl Microbiol Biotechnol* 101:2273–2279
- Maiya S, Grundmann A, Li S-M, Turner G (2006) The fumitremorgin gene cluster of *Aspergillus fumigatus*: identification of a gene encoding brevianamide F synthetase. *Chembiochem* 7:1062–1069
- Moutiez M, Belin P, Gondry M (2017) Aminoacyl-tRNA-utilizing enzymes in natural product biosynthesis. *Chem Rev* 117:5578–5618

- Moutiez M, Schmitt E, Seguin J, Thai R, Favry E, Belin P, Mechulam Y, Gondry M (2014a) Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat Commun* 5:5141
- Moutiez M, Seguin J, Fonvielle M, Belin P, Jacques IB, Favry E, Arthur M, Gondry M (2014b) Specificity determinants for the two tRNA substrates of the cyclodipeptide synthase AlbC from *Streptomyces noursei*. *Nucleic Acids Res* 42:7247–7258
- Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Seguin J, Moutiez M, Li Y, Belin P, Lecoq A, Fonvielle M, Charbonnier JB, Pernodet JL, Gondry M (2011) Nonribosomal peptide synthesis in animals: the cyclodipeptide synthase of *Nematostella*. *Chem Biol* 18:1362–1368
- Walsh CT (2014) Biological matching of chemical reactivity: pairing indole nucleophilicity with electrophilic isoprenoids. *ACS Chem Biol* 9:2718–2728
- Walsh CT (2016) Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat Prod Rep* 33:127–135
- Winkelblech J, Fan A, Li S-M (2015) Prenyltransferases as key enzymes in primary and secondary metabolism. *Appl Microbiol Biotechnol* 99:7379–7397
- Wohlgemuth V, Kindinger F, Xie X, Wang BG, Li S-M (2017) Two prenyltransferases govern a consecutive prenylation cascade in the biosynthesis of echinulin and neoechinulin. *Org Lett* 19:5928–5931
- Xu W, Gavia DJ, Tang Y (2014) Biosynthesis of fungal indole alkaloids. *Nat Prod Rep* 31:1474–1487

Applied Microbiology and Biotechnology

Supplementary Materials

Expanding Tryptophan-containing Cyclodipeptide Synthase Spectrum by Identification of Nine Members from *Streptomyces* Strains

Jing Liu¹, Huili Yu¹, and Shu-Ming Li¹

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

*Corresponding author

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

Tel/Fax: +49 6421 2822461/+49 6421 2825365

Content

Table S1 <i>Streptomyces</i> strains used in this study	S3
Table S2 The yields of the main products of the 11 CDPSs under six different conditions	S4
Table S3 HR-EI-MS data of the CDPS products	S5
Table S4 ¹ H NMR data of the isolated cyclodipeptides	S6-S7
Fig. S1 Phylogenetic tree of known CDPSs from actinobacteria prior to this study	S8
Fig. S2 ¹ H NMR spectrum of <i>cyclo</i> -(L-Trp-L-Trp) in DMSO-d ₆	S9
Fig. S3 ¹ H NMR spectrum of <i>cyclo</i> -(L-Trp-L-Pro) in CDCl ₃	S10
Fig. S4 ¹ H NMR spectrum of <i>cyclo</i> -(L-Trp-L-Leu) in CDCl ₃	S11
Fig. S5 ¹ H NMR spectrum of <i>cyclo</i> -(L-Trp-L-Tyr) in DMSO-d ₆	S12
Fig. S6 ¹ H NMR spectrum of <i>cyclo</i> -(L-Trp-L-Phe) in DMSO-d ₆	S13
Fig. S7 ¹ H NMR spectrum of <i>cyclo</i> -(L-Trp-L-Ala) in CD ₃ OD	S14
Fig. S8 ¹ H NMR spectrum of <i>cyclo</i> -(L-Trp-L-Met) in CD ₃ OD	S15
Fig. S9 ¹ H NMR spectrum of <i>cyclo</i> -(L-Trp-L-Val) in DMSO-d ₆	S16
Fig. S10 LC-MS analysis of recombinant <i>E. coli</i> strains with empty vector and expression constructs for cFL	S17
References	S18

Table S1 *Streptomyces* strains used in this study

Strains	Source	Cultivation Media
<i>Streptomyces cattleya</i> DSM 46488	DSMZ	GYM
<i>Streptomyces lavendulae</i> NRRL B-2774	NRRL	GYM
<i>Streptomyces purpureus</i> NRRL B-5737	NRRL	ISP4
<i>Streptomyces xanthophaeus</i> NRRL B-5414	NRRL	ISP4
<i>Streptomyces</i> sp. NRRL S-1868	NRRL	ISP4
<i>Streptomyces</i> sp. NRRL F-5053	NRRL	ISP4
<i>Streptomyces</i> sp. NRRL F-5123	NRRL	GYM
<i>Streptomyces varsoviensis</i> NRRL B-3589	NRRL	GYM
<i>Streptomyces monomycini</i> NRRL B-24309	NRRL	GYM
<i>Streptomyces</i> sp. NRRL F-5639	NRRL	ISP4
<i>Streptomyces laurentii</i> DSM 41684	DSMZ	GYM

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen

NRRL: ARS Culture Collection

Table S2 The yields (mg/L culture) of the main products of the 11 CDPSs under six different conditions

CDPSs Temperature -IPTG final conc.	CWWS1 _{D46488} (cWW)	CWWS2 _{D46488} (cWW)	CWWS1 _{NB2774} (cWW)	CWWS1 _{NB5737} (cWW)	CWWS1 _{NB5414} (cWW)	
20 °C - 0.2 mM	77.0	33.7	69.2	18.8	55.1	
20 °C - 0.5 mM	88.8	33.7	85.8	24.4	48.8	
30 °C - 0.2 mM	88.2	59.1	95.8	41.2	34.3	
30 °C - 0.5 mM	95.5	51.8	98.6	63.7	34.5	
37 °C - 0.2 mM	81.9	19.5	82.6	9.0	3.4	
37 °C - 0.5 mM	89.7	28.5	65.0	11.3	-	
CDPSs Temperature -IPTG final conc.	CWPS1 _{NS1868} (cWP)	CWPS1 _{NF5123} (cWP)	CWXS1 _{NS1868} (cWA)	CWLS1 _{NF5053} (cWL)	CWXS1 _{NB3589} (cWY)	CWXS1 _{NB24309} (cWY)
20 °C - 0.2 mM	144.9	69.2	69.3	93.9	141.4	51.7
20 °C - 0.5 mM	173.8	85.8	75.5	96.4	111.8	64.7
30 °C - 0.2 mM	131.4	95.8	68.1	104.9	25.8	55.8
30 °C - 0.5 mM	198.2	98.6	74.7	104.2	44.9	54.9
37 °C - 0.2 mM	146.3	82.6	14.7	104.9	94.6	15.5
37 °C - 0.5 mM	128.4	65.0	12.1	84.7	106.1	13.4

Table S3 HR-EI-MS data of the CDPS products

Compound	Chemical Formula	[M+H] ⁺		Deviation (ppm)
		Calculated	Measured	
cWA	C ₁₄ H ₁₅ N ₃ O ₂	258.1237	258.1224	5.0
cWY	C ₂₀ H ₁₉ N ₃ O ₃	350.1499	350.1495	1.1
cWM	C ₁₆ H ₁₉ N ₃ O ₂ S	318.1271	318.1289	-5.7
cWV	C ₁₆ H ₁₉ N ₃ O ₂	286.1550	286.1536	4.9
cWW	C ₂₂ H ₂₀ N ₄ O ₂	373.1659	373.1636	6.2
cWF	C ₂₀ H ₁₉ N ₃ O ₂	334.1550	334.1545	1.5
cWL	C ₁₇ H ₂₁ N ₃ O ₂	300.1707	300.1709	-0.7
cWP	C ₁₆ H ₁₇ N ₃ O ₂	284.1394	284.1417	-8.1
cFL	C ₁₅ H ₂₀ N ₂ O ₂	261.1598	261.1619	-8.0

Table S4 ^1H NMR data of the isolated cyclodipeptides

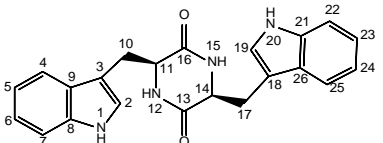
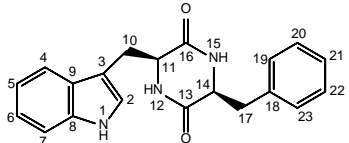
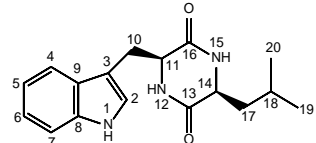
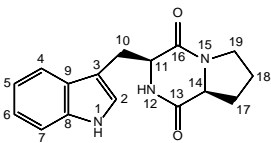
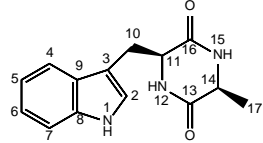
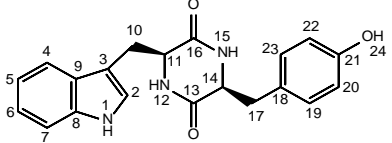
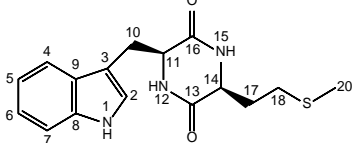
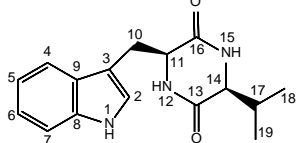
Comp.				
Pos.	δ_{H} , multi., J [Hz]	δ_{H} , multi., J [Hz]	δ_{H} , multi., J [Hz]	δ_{H} , multi., J [Hz]
1	10.81, s	10.90, s	8.15, s	8.55, s
2	6.61, d, 2.3	7.00, d, 2.3	7.11, d, 2.3	7.06, d, 2.1
4	7.35, d, 8.0	7.52, dd, 8.0, 1.1	7.63, dd, 8.0, 1.0	7.58, dd, 8.0, 1.0
5	6.95, ddd, 8.0, 7.0, 1.1	7.02, ddd, 8.0, 7.0, 1.1	7.15, ddd, 8.0, 7.1, 1.0	7.13, ddd, 8.0, 7.0, 1.0
6	7.04, ddd, 8.1, 7.0, 1.1	7.10, ddd, 8.1, 7.0, 1.1	7.23, ddd, 8.2, 7.1, 1.0	7.22, ddd, 8.1, 7.0, 1.1
7	7.28, dt, 8.1, 1.1	7.35, dt, 8.1, 1.1	7.39, dt, 8.2, 1.0	7.38, dt, 8.1, 1.1
10	2.71, dd, 14.3, 4.2 2.20, dd, 14.3, 6.7	2.84, dd, 14.5, 5.9 2.57, m	3.49, ddd, 14.7, 3.1, 0.7 3.21, dd, 14.7, 8.4	3.75, ddd, 15.1, 3.8, 1.0 2.97, dd, 15.1, 10.7
11	3.87, m	4.00, m	4.32, m	4.37, ddd, 10.7, 3.8, 1.0
12	7.66, d, 2.5	7.90, d, 2.4	5.88, s	5.83, s
14	3.87, m	3.89, m	3.89, m	4.06, bt, 7.4
15	7.66, d, 2.5	7.71, d, 2.5	5.86, s	-
17	2.71, dd, 14.3, 4.2 2.20, dd, 14.3, 6.7	2.50, m 1.92, dd, 13.6, 7.1	1.51, m 1.48, m	2.31, m 2.01, m
18	-	-	0.98, m	1.96, m 1.89, m
19	6.61, d, 2.3	6.76, dd, 8.0, 2.0	0.84, d, 6.5	3.65, m 3.58, m
20	10.81, s	7.20 ^a	0.81, d, 6.5	-
21	-	7.20 ^a	-	-
22	7.28, dt, 8.1, 1.1	7.20 ^a	-	-
23	7.04, ddd, 8.1, 7.0, 1.1	6.76, dd, 8.0, 2.0	-	-
24	6.95, ddd, 8.0, 7.0, 1.1	-	-	-
25	7.35, d, 8.0	-	-	-

Table S4 (continued)

Comp.				
Pos.	δ_H , multi., <i>J</i> [Hz]	δ_H , multi., <i>J</i> [Hz]	δ_H , multi., <i>J</i> [Hz]	δ_H , multi., <i>J</i> [Hz]
1	-	10.90, d, 1.5	-	10.85, s
2	7.05, s	7.00, d, 1.5	7.04, s	7.11, d, 2.4
4	7.58, dt, 8.0, 1.0	7.51, dd, 7.9, 1.0	7.59, dt, 8.0, 1.0	7.63, d, 7.9
5	6.97, ddd, 8.0, 7.1, 1.0	7.02, ddd, 7.9, 7.0, 1.0	6.96, ddd, 8.0, 7.0, 1.0	6.97, ddd, 7.9, 7.1, 1.0
6	7.05, ddd, 8.1, 7.1, 1.0	7.09, ddd, 8.1, 7.0, 1.0	7.05, ddd, 8.1, 7.0, 1.1	7.05, ddd, 8.1, 7.1, 1.0
7	7.30, dt, 8.1, 1.0	7.34, dt, 8.1, 1.0	7.30, dt, 8.1, 1.1	7.32, d, 8.1
10	3.44, dd, 14.6, 4.1 3.12, dd, 14.6, 4.5	2.83, dd, 14.5, 4.2 2.49 ^b	3.47, dd, 14.7, 3.5 3.09, dd, 14.7, 4.6	3.24, dd, 14.5, 5.0 3.11, dd, 14.5, 4.5
11	4.25, ddd, 4.5, 4.1, 1.3	3.97, m	4.27, ddd, 4.6, 3.5, 1.4	4.17, m
12	-	7.80, d, 2.5	-	7.96, s
14	3.67, qd, 7.1, 1.2	3.81, m	3.77, ddd, 7.4, 4.7, 1.3	3.52, m
15	-	7.62, d, 2.6	-	7.87, s
17	0.32, d, 7.1	2.45 ^b	1.17, m	1.68, m
	-	1.86, dd, 13.6, 4.5	0.60, m	-
18	-	-	1.63, m	0.64, d, 7.1
19	-	6.61, dt, 8.6, 2.0	-	0.25, d, 6.8
20	-	6.56, dt, 8.6, 2.0	1.73, s	-
22	-	6.56, dt, 8.6, 2.0	-	-
23	-	6.61, dt, 8.6, 2.0	-	-
24	-	9.18, s	-	-

^{a, b}: Signals with the same letter overlapping with each other

The data correspond well to those reported for cWA (Caballero et al. 1998), cWY (Kumar et al. 2014), cWV (He et al. 2013), cWL (Kumar et al. 2014), cWW (Giessen et al. 2013a; Lu et al. 2017), cWF (Kumar et al. 2014), and cWP (Grundmann and Li 2005), respectively.

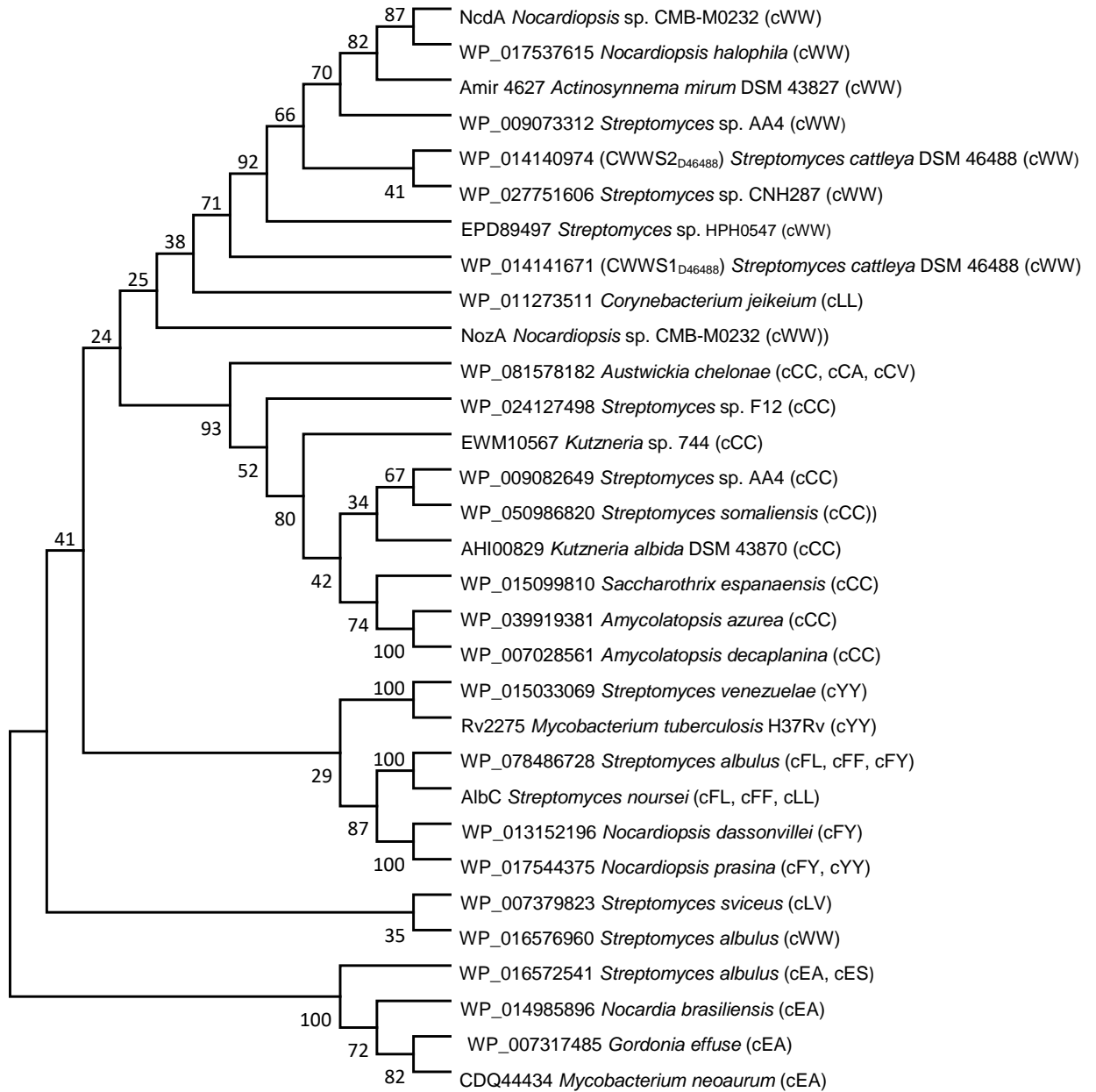


Fig. S1 Phylogenetic tree of known CDPSs from actinobacteria prior to this study (Brockmeyer and Li 2017; Giessen et al. 2013a; Giessen et al. 2013b; Jacques et al. 2015; James et al. 2015)

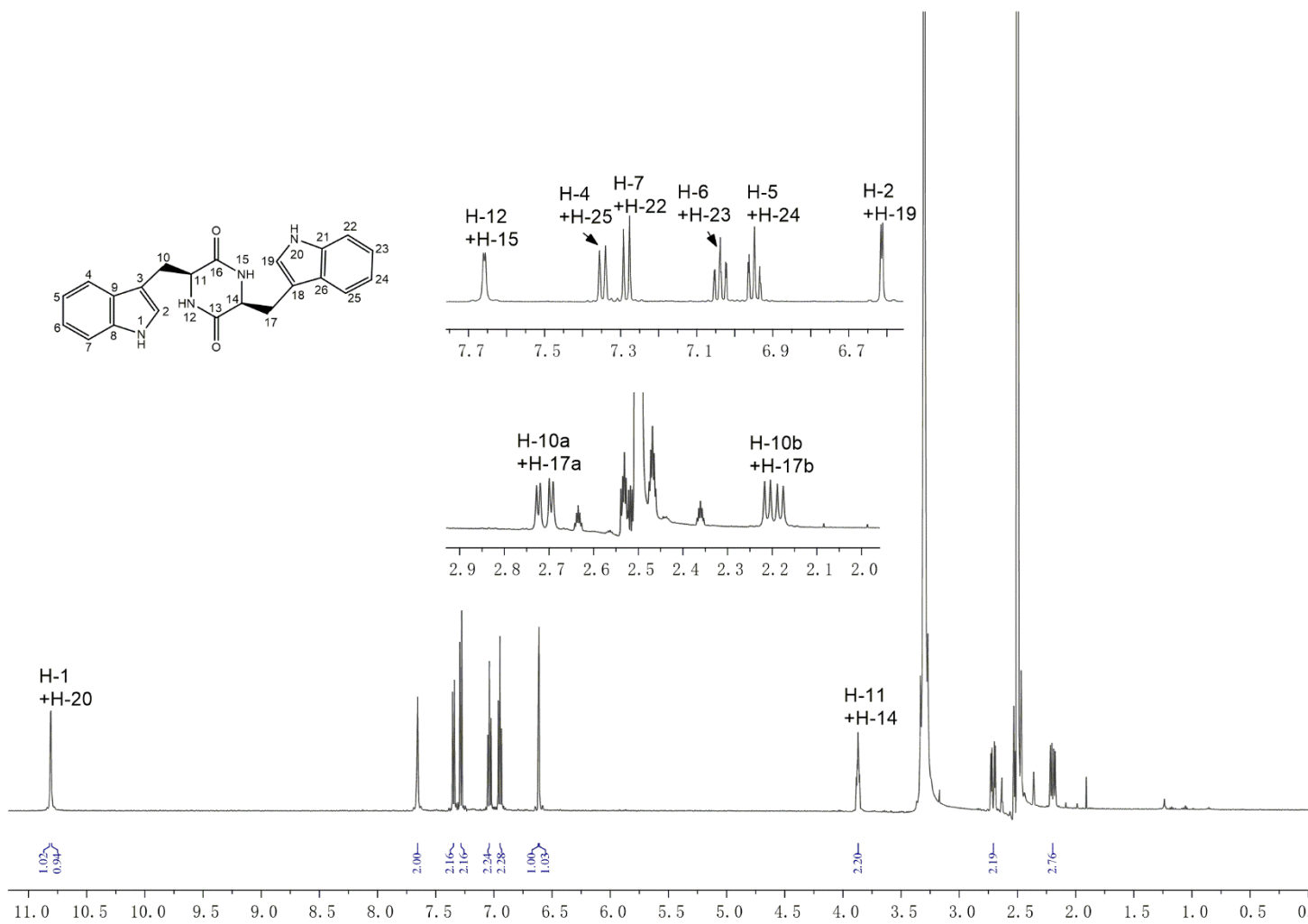


Fig. S2 ^1H NMR spectrum of *cyclo*-(L-Trp-L-Trp) in DMSO-d_6

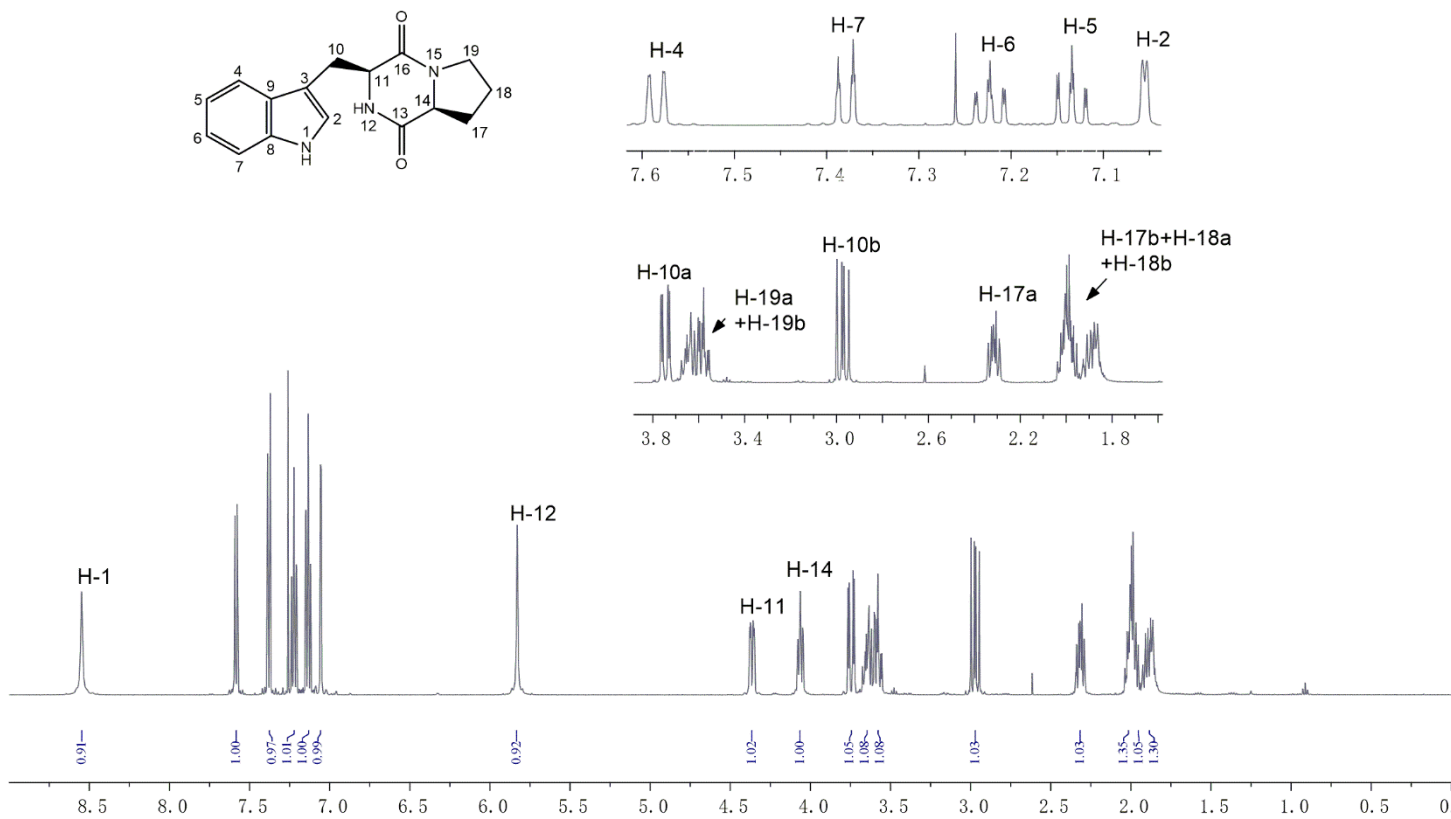


Fig. S3 ^1H NMR spectrum of *cyclo*-(L-Trp-L-Pro) in CDCl_3

S10

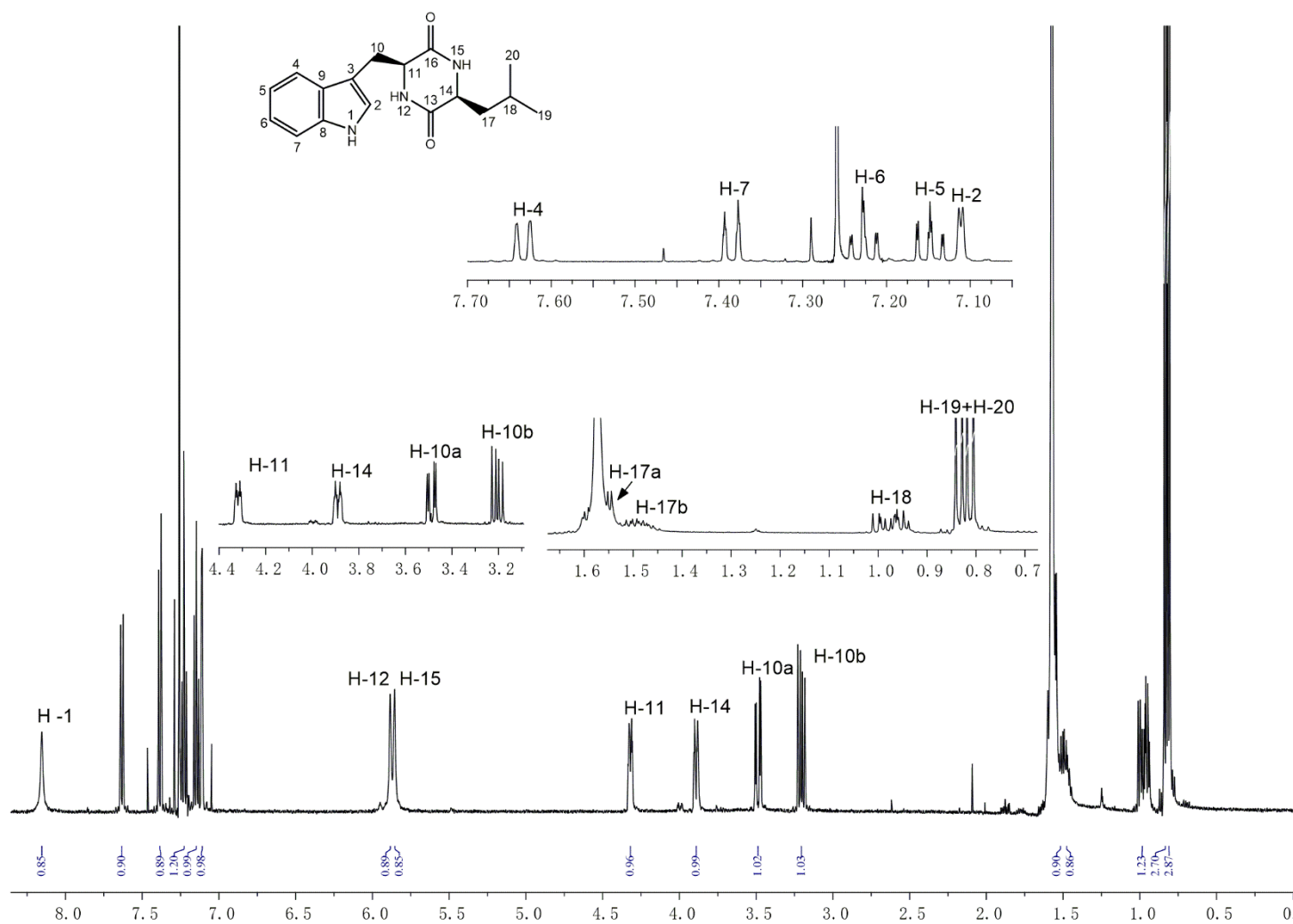


Fig. S4 ^1H NMR spectrum of *cyclo*-(L-Trp-L-Leu) in CDCl_3

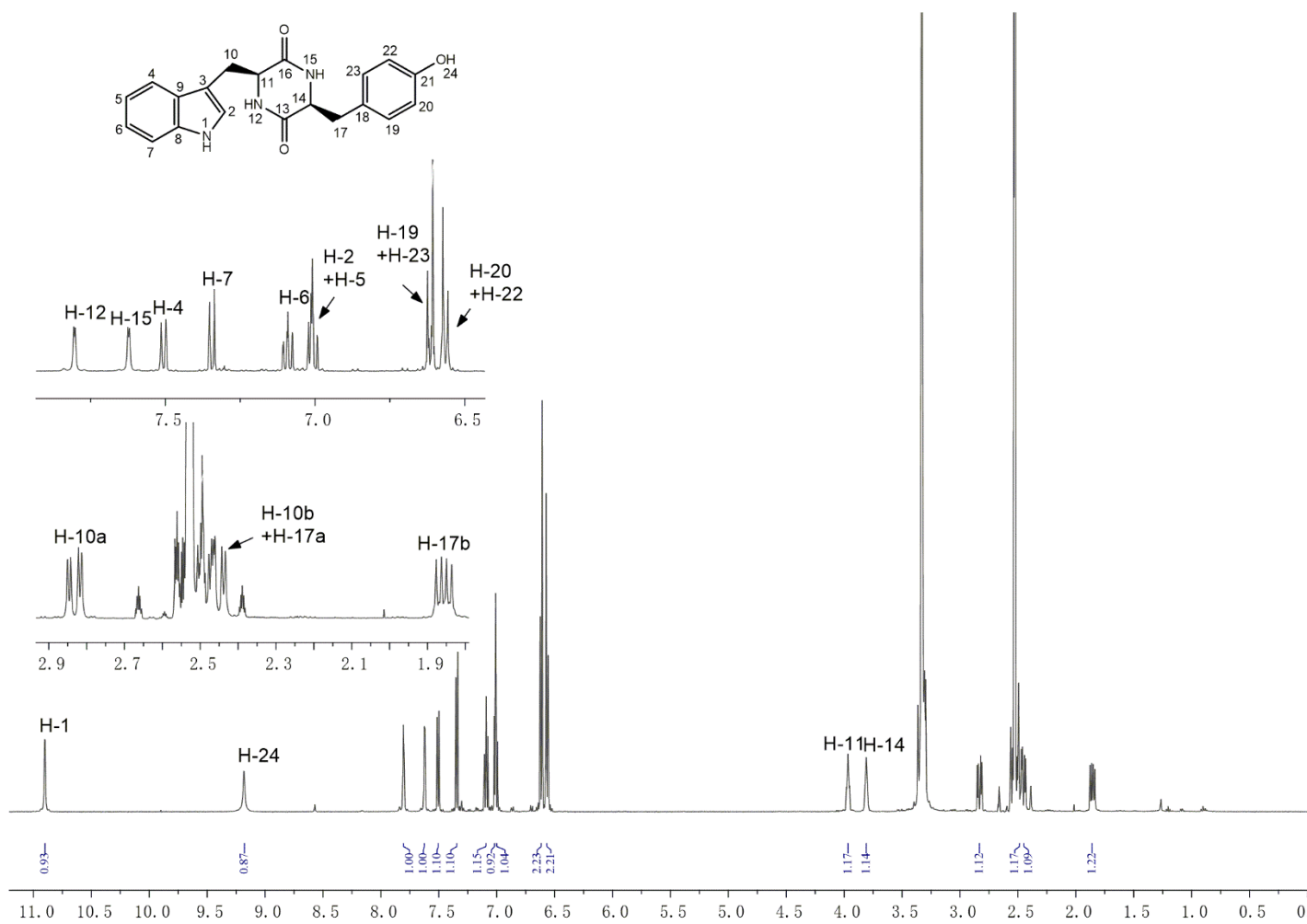


Fig. S5 ^1H NMR spectrum of in *cyclo*-(L-Trp-L-Tyr) in DMSO- d_6

S12

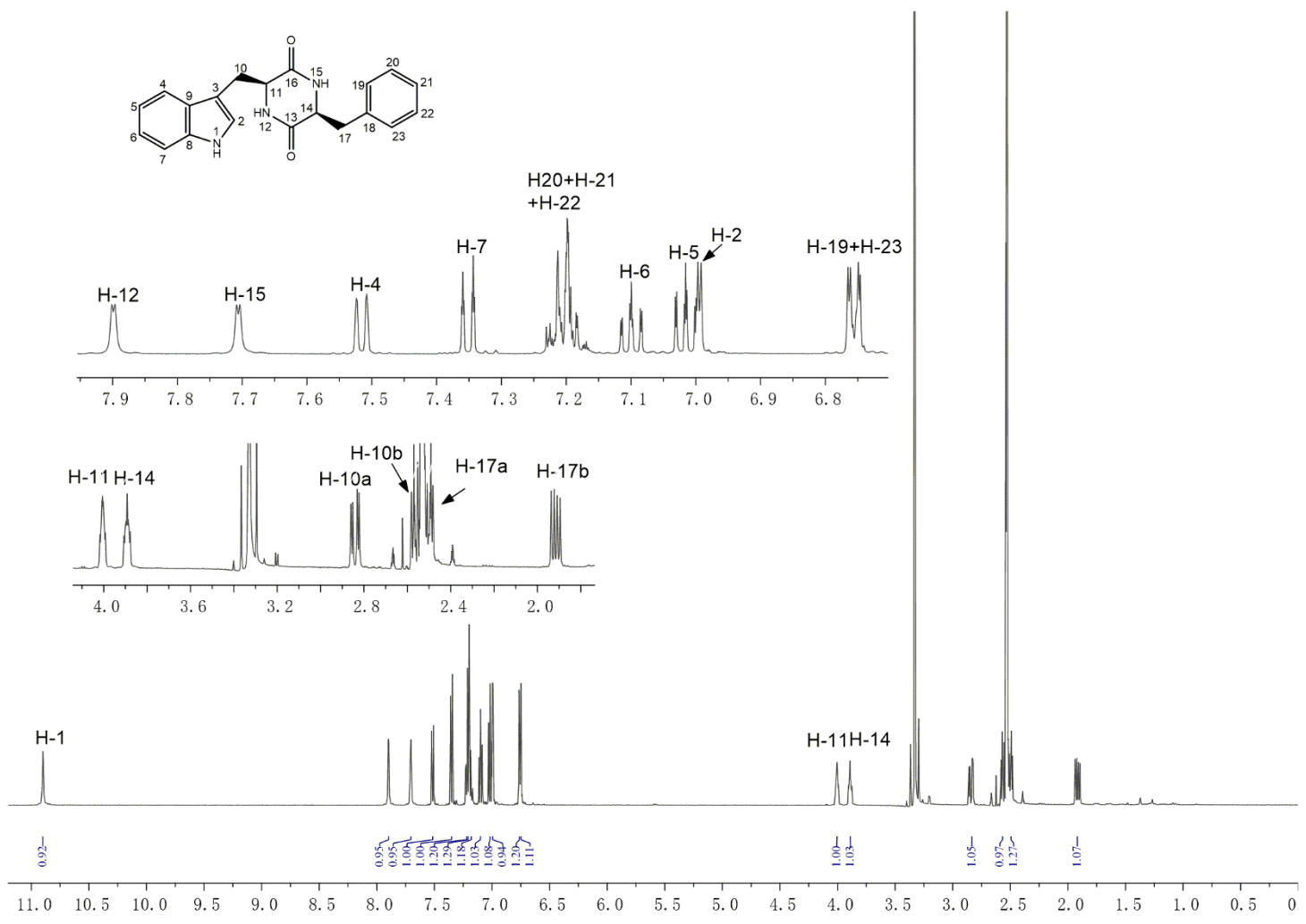


Fig. S6 ^1H NMR spectrum of *cyclo*-(L-Trp-L-Phe) in DMSO- d_6

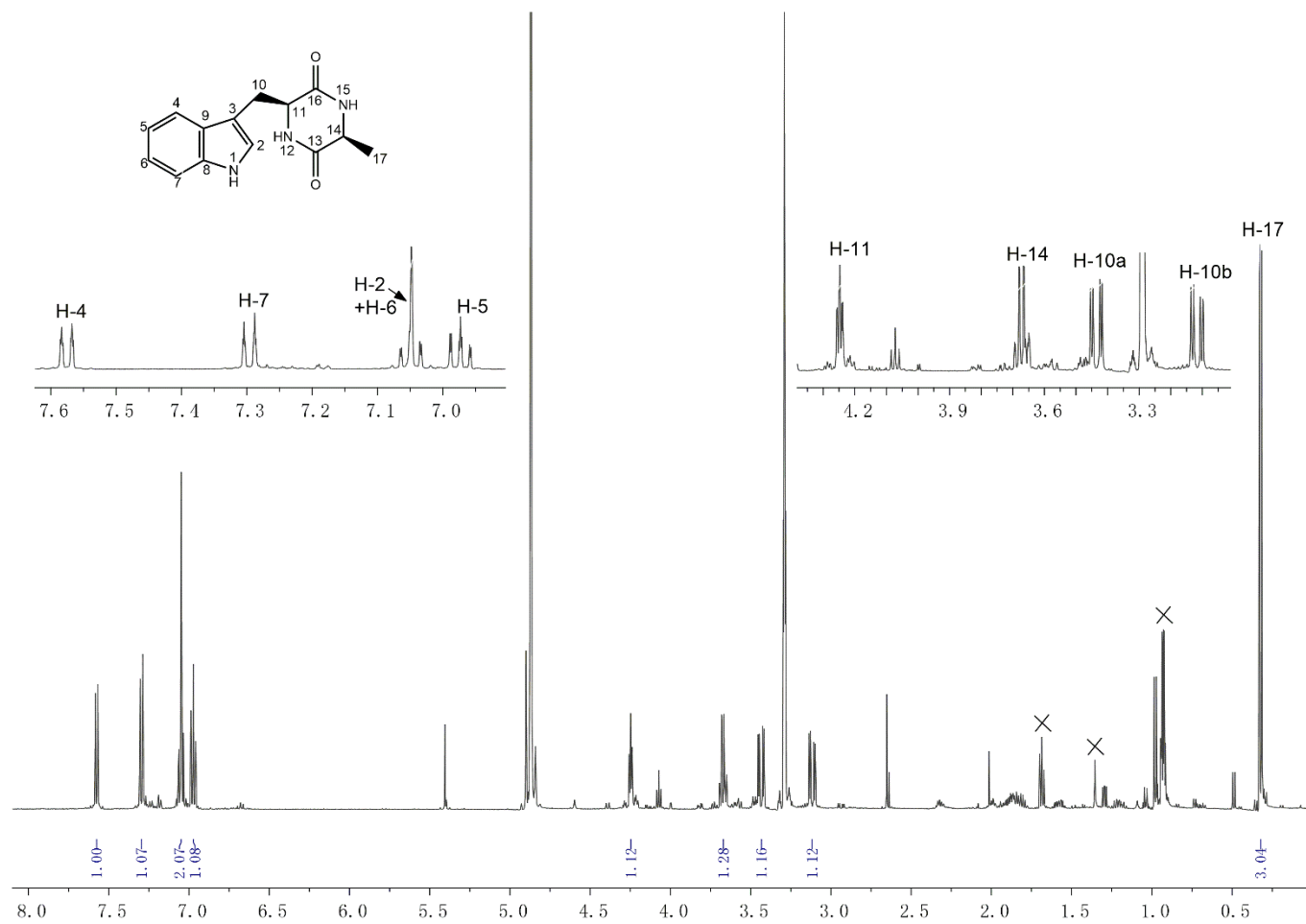


Fig. S7 ¹H NMR spectrum of *cyclo*-(L-Trp-L-Ala) in CD₃OD

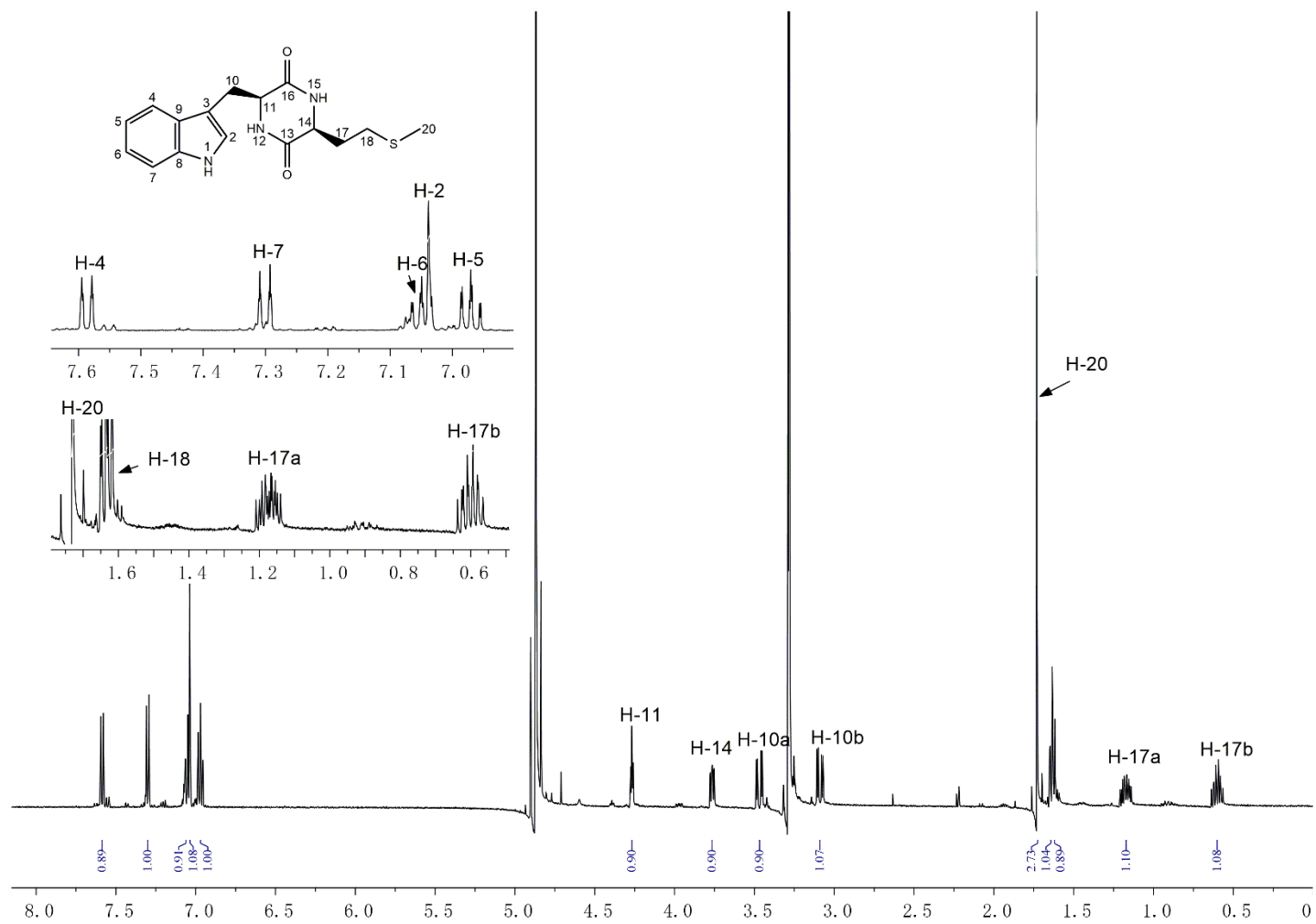


Fig. S8 ¹H NMR spectrum of *cyclo*-(L-Trp-L-Met) in CD₃OD

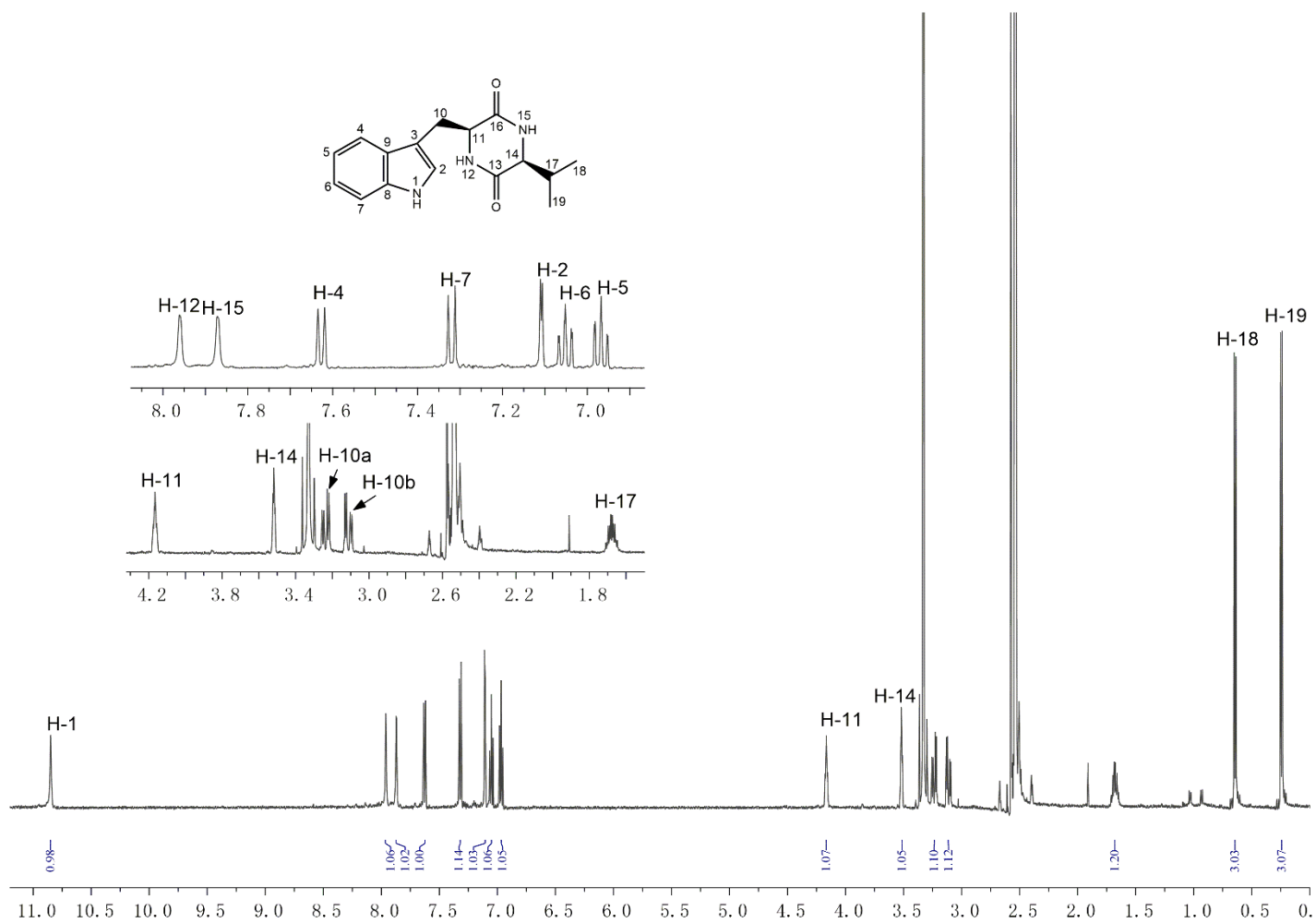


Fig. S9 ¹H NMR spectrum of *cyclo*-(L-Trp-L-Val) in DMSO-d₆

S16

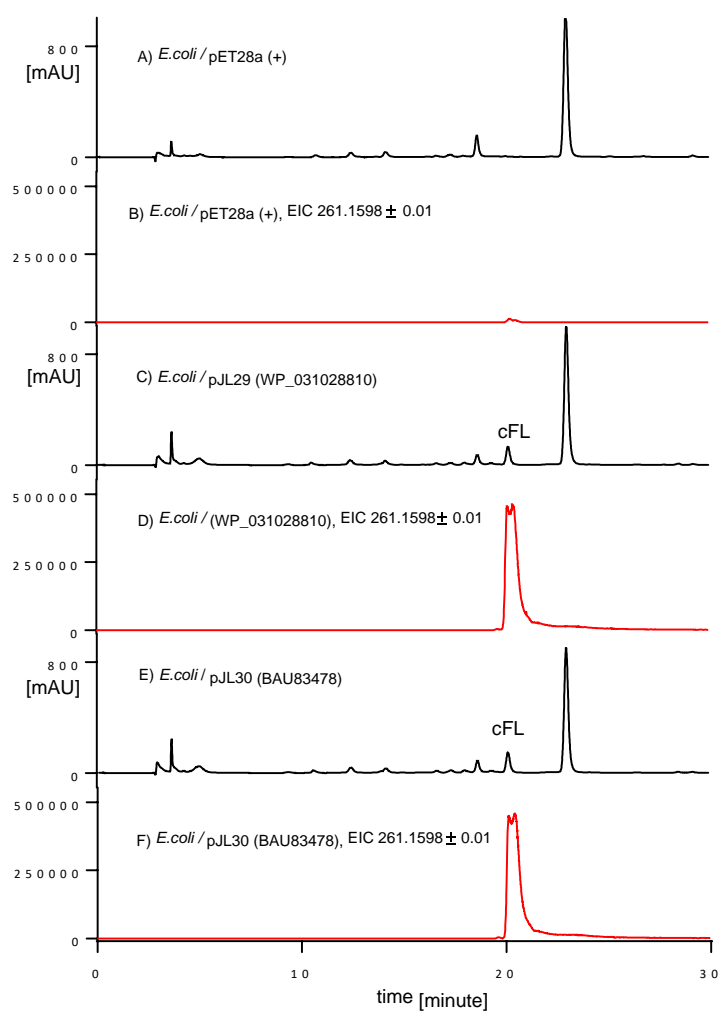


Fig. S10 LC-MS analysis of recombinant *E. coli* strains with empty vector (A and B) and expression constructs (C-F). Black lines are UV absorptions at 258 nm and the red lines refer to extracted positive ion chromatograms (EIC).

References

- Brockmeyer K, Li S-M (2017) Mutations of residues in pocket P1 of a cyclodipeptide synthase strongly increase product formation. *J Nat Prod* 80:2917-2922
- Caballero E, Avendaño C, Menéndez JC (1998) Stereochemical issues related to the synthesis and reactivity of pyrazino[2',1'-5,1]pyrrolo[2,3-b]indole-1,4-diones. *Tetrahedron: Asymmetry* 9:967-981
- Giessen TW, von Tesmar AM, Marahiel MA (2013a) A tRNA-dependent two-enzyme pathway for the generation of singly and doubly methylated ditryptophan 2,5-diketopiperazines. *Biochemistry* 52:4274-4283
- Giessen TW, von Tesmar AM, Marahiel MA (2013b) Insights into the generation of structural diversity in a tRNA-dependent pathway for highly modified bioactive cyclic dipeptides. *Chem Biol* 20:828-838
- Grundmann A, Li S-M (2005) Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from *Aspergillus fumigatus*. *Microbiology* 151:2199-2207
- He R, Wang B, Wakimoto T, Wang M, Zhu L, Abe I (2013) Cyclodipeptides from metagenomic library of a Japanese marine sponge. *J Braz Chem Soc* 24:1926-1932
- Jacques IB, Moutiez M, Witwinowski J, Darbon E, Martel C, Seguin J, Favry E, Thai R, Lecoq A, Dubois S, Pernodet JL, Gondry M, Belin P (2015) Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat Chem Biol* 11:721-727
- James ED, Knuckley B, Alqahtani N, Porwal S, Ban J, Karty JA, Viswanathan R, Lane AL (2015) Two distinct cyclodipeptide synthases from a marine actinomycete catalyze biosynthesis of the same diketopiperazine natural product. *ACS Synth Biol* 5:547-553
- Kumar SN, Mohandas C, Nambisan B (2014) Purification, structural elucidation and bioactivity of tryptophan containing diketopiperazines, from *Comamonas testosteroni* associated with a rhabditid entomopathogenic nematode against major human-pathogenic bacteria. *Peptides* 53:48-58
- Lu C, Xie F, Shan C, Shen Y (2017) Two novel cyclic hexapeptides from the genetically engineered *Actinosynnema pretiosum*. *Appl Microbiol Biotechnol* 101:2273-2279

4.2 Guanitrypmycin biosynthetic pathways imply cytochrome P450-mediated regio- and stereospecific guaninyl transfer reactions

Alkaloid Biosynthesis

International Edition: DOI: 10.1002/anie.201906891

German Edition: DOI: 10.1002/ange.201906891

Guanitrypmycin Biosynthetic Pathways Imply Cytochrome P450 Mediated Regio- and Stereospecific Guaninyl-Transfer Reactions

Jing Liu, Xiulan Xie, and Shu-Ming Li*

Abstract: Mining microbial genomes including those of *Streptomyces* reveals the presence of a large number of biosynthetic gene clusters. Unraveling this genetic potential has proved to be a useful approach for novel compound discovery. Here, we report the heterologous expression of two similar P450-associated cyclodipeptide synthase-containing gene clusters in *Streptomyces coelicolor* and identification of eight rare and novel natural products, the C3-guaninyl indole alkaloids guanitrypmycins. Expression of different gene combinations proved that the cyclodipeptide synthases assemble cyclo-L-Trp-L-Phe and cyclo-L-Trp-L-Tyr, which are consecutively and regiospecifically modified by cyclodipeptide oxidases, cytochrome P450 enzymes, and N-methyltransferases. In vivo and in vitro results proved that the P450 enzymes function as key biocatalysts and catalyze the regio- and stereospecific 3 α -guaninylation at the indole ring of the tryptophanyl moiety. Isotope-exchange experiments provided evidence for the non-enzymatic epimerization of the biosynthetic pathway products via keto–enol tautomerism. This post-pathway modification during cultivation further increases the structural diversity of guanitrypmycins.

Introduction

The genus *Streptomyces* has been well known for their ability to produce numerous compounds with diverse activities.^[1] In recent years, mining the increasing number of released genome sequences has shown that many more uncharacterized or silent clusters of secondary metabolites remain hidden in their genomes.^[2] Therefore, various strategies including genetic and cultivation-dependent activation and heterologous expression have been performed to unlock these cryptic clusters for novel metabolites.^[3] These efforts have provided deep insights into the unprecedented potential of *Streptomyces* to synthesize more metabolites than previously identified or envisaged.



Natural products derived from cyclodipeptides (CDPs) with a 2,5-diketopiperazine (DKP) skeleton comprise an important class of secondary metabolites, in particular indole alkaloids derived from tryptophan-containing CDPs,^[4,5] which are widespread in fungi, bacteria, and plants.^[5–8] In nature, the CDP core can be generated by two distinct enzyme groups, that is, the nonribosomal peptide synthetases (NRPSs)^[9] and the aminoacyl tRNA-dependent cyclodipeptide synthases (CDPSs).^[10,11] Tailoring enzymes for modifying the CDP scaffold are usually encoded by genes associated with the NRPS or CDPS in microorganisms.^[12,13] The CDPS genes are often located together with those for cytochrome P450s, oxidoreductases, cyclodipeptide oxidases (CDOs), prenyltransferases (PTs), and methyltransferases (MTs),^[12,14] which significantly expand the diversity of DKP-based natural products. Although a large number of CDPSs have been identified in the last years,^[11,15,16] only nine CDPS-containing biosynthetic pathways have been functionally characterized,^[14] indicating that much DKP tailoring potential remains unexplored. As modification enzymes, cytochrome P450s play important roles in natural product biosynthesis and constitute a large superfamily of heme-containing oxidases. They catalyze a variety of chemical reactions, such as hydroxylation, epoxidation, and demethylation.^[17,18] However, only approximately 2.4% of *Streptomyces*-originated P450s are functionally and less than 0.4% structurally characterized.^[19] Interestingly, all five functionally characterized P450s in the CDPS-dependent pathways catalyze intriguing chemical reactions. These include CYP121 for an intramolecular C–C bond formation within one cyclo-L-Tyr-L-Tyr molecule,^[20] CYP134A1 for the aromatization and N-oxide formation of cyclo-L-Leu-L-Leu,^[21] BcmD for installing a tertiary hydroxyl group^[22] (Scheme 1 i), P450_{NB5737} for a coupling of guanine with cyclo-L-Trp-L-Trp via C–N bond formation^[23] (Scheme 1 ii), and NascB for dimerization of cyclo-L-Trp-L-Pro^[24] (Scheme 1 iii). These findings encouraged us to investigate P450-associated *cdps*-containing gene clusters for the generation of various CDP derivatives.

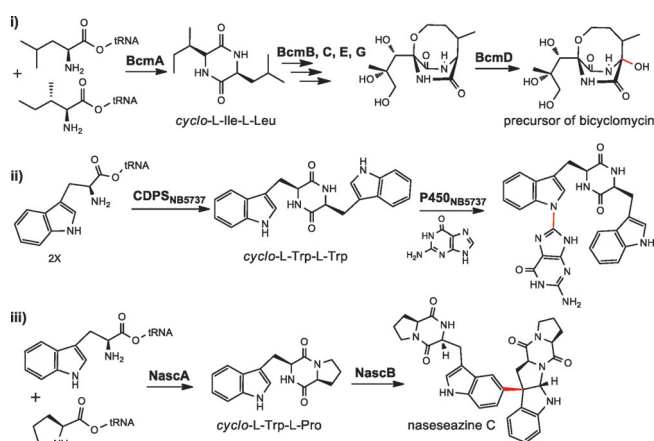
Herein, we report the characterization of two similar five-gene operons from *Streptomyces* coding for CDPS, P450, CDO, and MT proteins for the formation of unprecedented C3-guaninyl DKPs—guanitrypmycins, whereby P450 catalyzes the vital C–C linkage between indole ring of the CDPs with the nucleobase guanine. Isotope-exchange experiments proved that keto–enol tautomerism is responsible for the non-enzymatic spontaneous epimerization of the pathway products during the cultivation. Structural elucidation of eight C3-guaninyl pyrroloindoline alkaloids in *Streptomyces coelicolor* transformants confirms the proposed functions of the biosynthetic enzymes and provides an excellent example for

[*] J. Liu, Prof. S.-M. Li

Institut für Pharmazeutische Biologie und Biotechnologie
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)

 Supporting information (including experimental details) and the
 ORCID identification number(s) for the author(s) of this article can
be found under:
<https://doi.org/10.1002/anie.201906891>.



Scheme 1. Examples of biosynthetic pathways involving CDPs and P450s.

revealing the genetic potential by genome mining and heterologous expression.

Results and Discussion

Identification of the *gut* Gene Cluster

In a previous study, we identified nine CDPS genes from *Streptomyces* strains, which assemble tryptophan-containing CDPs.^[15] Sequence analysis revealed that eight of them are surrounded by a putative cytochrome P450 gene. Among them, two CDPS genes, CWXS1_{NB24309} (renamed to GutA₂₄₃₀₉ in this study) in *S. monomycini* NRRL B-24309 and CWXS1_{NB3589} (GutA₃₅₈₉) in *S. varsoviensis* NRRL B-3589, are located in the similar gene loci containing four more genes coding for three modification enzymes, that is, CDO, cytochrome P450, and MT (Figure 1). The corresponding proteins

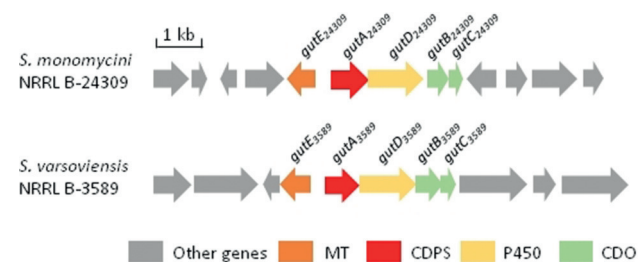


Figure 1. Comparative illustration of the *gut* clusters.

of the two clusters, termed *gut* cluster in this study, share sequence identities between 79–90% on the amino acid level (Table S1). It can be therefore speculated that these two clusters code for identical or very similar natural products.

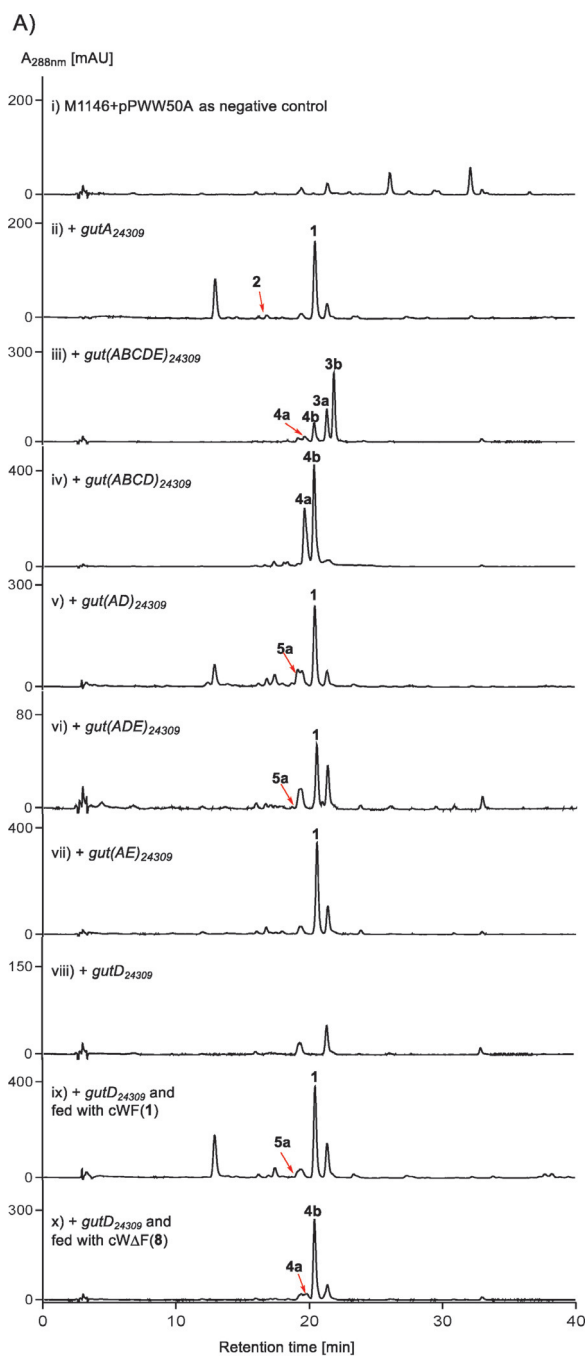
Prior to this study, two characterized MTs from CDPS-associated pathways were proven to be responsible for N- or O-methylation of cyclodipeptides.^[12] CDOs comprising two subunits A and B are flavin-dependent dehydrogenases and

usually use CDPs as substrates for installation of exocyclic C–C double bonds to the DKP rings.^[25,26] Multiple sequence alignments of the two putative P450 enzymes in the aforementioned gene clusters with other known entries revealed the presence of the conserved motifs of the bacterial P450 proteins (Figure S1), that is, the G³⁴⁷XXXC³⁵¹ (referring to EryF^[27]) motif in the heme-binding loop and the highly conserved A/Gⁿ⁻¹-Gⁿ-XX-Tⁿ⁺³ motif in the long I-helix running over the distal surface of the heme in P450 scaffold.^[27] The unique *cdps*-associated genetic organization presented in Figure 1, which is distinct from those studied before and has not been investigated yet, could be involved in the biosynthesis of novel DKP derivatives.

Heterologous Expression of *gut*₂₄₃₀₉ Cluster and Structure Elucidation of Guanitrypmycins

The two mentioned CDPs produced *cyclo*-L-Trp-L-Phe (cWF, **1**) and *cyclo*-L-Trp-L-Tyr (cWY, **2**) as main products, when heterologously expressed in *E. coli*.^[15] Initial attempts by gene inactivation experiments to find the metabolites biosynthesized by these clusters in *S. monomycini* and *S. varsoviensis* failed, due to the difficult genetic manipulation in these strains. We thereby turned to heterologous expression in *S. coelicolor* M1146 for functional proof of the gene clusters.^[28] To establish a genetic protocol, *gutA*₂₄₃₀₉ was amplified from the genomic DNA by PCR and cloned into the replicative vector pPWW50A^[29] (Strains, primers, and constructs are listed in Tables S2–4.) The M1146 cells with pPWW50A as the control strain and *gutA*₂₄₃₀₉ transformant were cultivated in GYM media at 28°C for 7 days. The fermentation broths were extracted with EtOAc and subjected to LC-MS analysis. In comparison to the extract of M1146 harboring pPWW50A (Figure 2A i), cWF (**1**) was detected as the predominant and cWY (**2**) as a trace peak in the extract of the *gutA*₂₄₃₀₉ expression strain (Figure 2A ii). These results differ clearly from those obtained from *E. coli*, in which cWY (**2**) was much higher accumulated than cWF (**1**).^[15] This is likely due to different expression hosts. Host-dependent distinct product formation has been observed in different organisms, even in different *Streptomyces* hosts. For example, the *fls* gene cluster delivers different metabolite profiles in four different hosts.^[30]

After successful confirmation of *gutA*₂₄₃₀₉ function, we cloned the complete *gut* cluster from *S. monomycini* in pPWW50A and introduced them into M1146. The transformants were cultivated and treated as previously described. LC-MS analysis of the transformant harboring *gut*-(ABCDE)₂₄₃₀₉, which codes for GutA₂₄₃₀₉ (CDPS), Gut(B/C)₂₄₃₀₉ (CDO), GutD₂₄₃₀₉ (cytochrome P450), and GutE₂₄₃₀₉ (MT), revealed the presence of four new peaks with similar UV spectra which correspond to two pairs regarding their molecular weights, **3a** and **3b** as well as **4a** and **4b** (Figure 2A iii). Peaks **3a** and **3b** with $[M+H]^+$ ions at m/z 495.189 ± 0.005 are 14 Daltons larger than **4a** and **4b** with $[M+H]^+$ ions at m/z 481.173 ± 0.005 (Figure 2A iii), indicating one additional methyl group in the structures of **3a** and **3b**. This speculation was confirmed by expression of *gut*-



B)

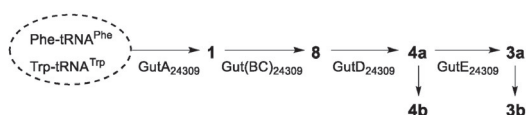
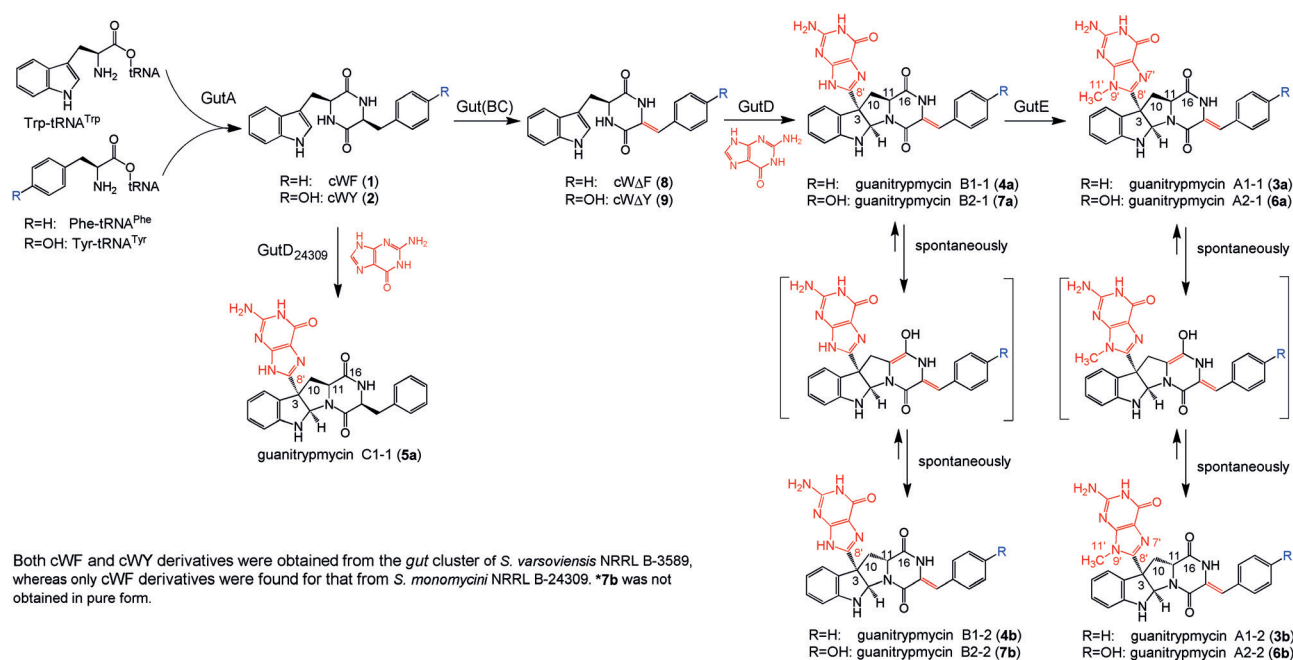


Figure 2. A) HPLC analysis of the *S. coelicolor* transformants and B) schematic presentation of the biosynthetic pathway of guanytrypmycins in *S. monomycini*. $[M+H]^+$ ions with a tolerance range of ± 0.005 were detected at m/z 334.155 (**1**), 350.150 (**2**), 495.189 (**3a/3b**), 481.1731 (**4a/4b**), and 483.1888 (**5a**), respectively.

(*ABCD*)₂₄₃₀₉ without the putative MT gene *gutE*₂₄₃₀₉. As shown in Figure 2A iv, **4a** and **4b** were detected as the

predominant metabolites of this transformant. Subsequently, fermentation of the transformant harboring *gut(ABCDE)*₂₄₃₀₉ on an 8 L scale and isolation afforded analytically pure **3a**, **3b**, **4a**, and **4b**. Their structures were characterized by detailed interpretation of NMR data of **3a** and **3b** including ¹H, ¹³C APT, COSY, HSQC, HMBC, and NOESY. (See the Supporting Information for structure elucidation. NMR data of the identified compounds are listed in Tables S5–13 and spectra provided as Figures S3–51.) Both **3a** and **3b** bear an exocyclic C–C double bond at the phenylalanyl side and a methylguaninyl moiety is attached to C-3 of cWF (**1**), thereby forming a hexahydropyrrolo[2,3-*b*]indole framework (Scheme 2). The five ¹³C signals of the guaninyl residue are found at approximately 157, 154, 154, 147, and 115 ppm, which is similar to the data reported previously.^[23,31,32] Very strong HMBC correlations from H2 to C8' and from H10 to C8' support that C-8' of guanine and C-3 of cWF (**1**) are connected via a C–C bond. The key HMBC correlations from H-11' of the methyl group to C4' and C8' suggest its attachment to N-9' of the guaninyl residue. These results imply the putative functions of the four enzymes. In this context, the P450 enzyme functions as an unusual guaninyl transferase and catalyzes the attachment of a guanine moiety to C-3 of the indole residue via a C–C bond. **3a** and **3b** differ from each other merely in the stereochemistry at C-11 and show therefore a slightly different CD spectra (Figures S52 and S53). In the NOESY spectrum of **3a** (Figure S8), strong correlations between H-2 at the indoline ring and H-11' of the guaninyl residue as well as H-2 and H-11 at the DKP ring prove that H-2, H-11', and H-11 are located at the same side. In comparison, strong NOE correlation of H-2 with H-11', but only very weak correlation between H-2 and H-11 was observed in the NOESY spectrum of **3b** (Figure S14), indicating different orientations of H-2 and H-11 in the structure. Based on the fact that **3b** is a non-enzymatic rearrangement product of **3a** (see below for details) and CDPS products from proteinogenic acids have an *S*-configuration, we assigned H-11 in **3a** an α - and in **3b** a β -orientation. Due to the C3-guaninyl pyrroloindoline feature, we named **3a** and **3b** guanytrypmycin A1-1 and A1-2 (Scheme 2), respectively.

Inspection of the ¹H NMR spectra of **4a** and **4b** indicated indeed the absence of the methyl group at the guaninyl residue of **3a** and **3b**. However, we did not have sufficient amounts of **4a** and **4b** for detailed structure elucidation. We therefore isolated them from the culture broth of M1146 harboring *gut(ABCD)*₂₄₃₀₉ (Figure 2A iv). **4a** and **4b** isolated from this strain showed identical ¹H NMR spectra to those from the strain carrying *gut(ABCDE)*₂₄₃₀₉. Comprehensive NMR analysis confirmed that **4a** and **4b** are demethylated **3a** and **3b** and named herewith guanytrypmycins B1-1 and B1-2, respectively (Scheme 2). **4a** and **4b** shared almost the same CD spectra with their counterparts **3a** and **3b**, respectively (Figures S52 and S53). Their configurations were also confirmed by interpretation of the NOE correlations of H-2, H-4, H-10, and H-11 (Figures S20 and S26).



Scheme 2. Formation of guanitrypmycins in *S. monomycini* NRRL B-24309 and *S. varsoviensis* NRRL B-3589.

Enzymatic and Non-enzymatic Formation of Guanitrypmycins

The identification of the compound pairs **3a/3b** and **4a/4b** as products of the transformant with *gut(ABCDE)*₂₄₃₀₉ as well as **4a/4b** as those with *gut(ABCD)*₂₄₃₀₉ raised the question about their relationships with each other and to the biosynthetic enzymes. We firstly proved their accumulation during the cultivation process. As shown in Figure S54, **3a** was detected as the major peak in the three-day-culture of *gut(ABCDE)*₂₄₃₀₉, whereas **4a** was the predominant product in that of *gut(ABCD)*₂₄₃₀₉. The ratios of **3b** to **3a** and **4b** to **4a** increased continuously during the fermentation process. **3b** and **4b** were even detected as the predominant products after 9 days, indicating the conversion of **3a** to **3b** and **4a** to **4b**. Incubation of **3a** and **3b** in GYM media confirmed the non-enzymatic spontaneous conversion of **3a** to **3b** (Figure S55). Almost the same results were also observed for **4a** and **4b** after incubation in GYM media (Figure S56).

Based on the fact that **3a** and **4a** have the same configuration at C11 as cWF (**1**), and **3b** and **4b** are their epimers at this position, we speculated that **3b** and **4b** are stable diastereomers of **3a** and **4a**, respectively, and formed via keto–enol tautomerism (Scheme 2). To prove this hypothesis, **3a**, **3b**, **4a**, and **4b** were incubated in 50% CD₃OD in D₂O at different pH values. The samples were subsequently subjected to LC-MS analysis (Figures S57–62). As expected for keto–enol tautomerism,^[33] the conversion of **3a** to **3b** and **4a** to **4b** are much easier under alkaline than under acidic conditions. Incorporation of one deuterium into **3b** and **4b** was determined by interpretation of the isotopic peaks of their [M+H]⁺ ions. In comparison to those in DMSO with *m/z* 495.19 and 481.17, the [M+H]⁺ ions of **3b** and **4b** are shifted to *m/z* 496.20 and 482.18 after treatment with 0.01M NaOH, respectively. At pH 8.0, the isotope peaks were partially

shifted. Furthermore, the changed isotope pattern of the [M+H]⁺ ions of the trace formed **3a** (from **3b**) and **4a** (from **4b**) confirmed the existence of the keto–enol tautomerism equilibration (Figures S57, S58, S60, and S61).

Taking the results together, GutA₂₄₃₀₉ assembles cWF (**1**), which will be dehydrogenated merely at the phenylalaninyl side by Gut(BC)₂₄₃₀₉ and connected with a guanine moiety by GutD₂₄₃₀₉. GutE₂₄₃₀₉ governs the last modification step after guanine attachment (Scheme 2, Figure 2B). However, the above data cannot give us the accurate reaction order of Gut(BC)₂₄₃₀₉ and GutD₂₄₃₀₉. Therefore, we expressed *gut(AD)*₂₄₃₀₉, *gut(ADE)*₂₄₃₀₉, and *gut(AE)*₂₄₃₀₉. Only slight (Figures 2Av, vi) or nearly no consumption of cWF (**1**) (Figure 2Avii) was observed in these transformants, indicating the importance of the exocyclic double bond for further metabolism. In the transformant with *gut(AD)*₂₄₃₀₉, cWF (**1**) was detected as the predominant product (ca. 85%) and the additional peak **5a** with a [M+H]⁺ ion at *m/z* 483.189 ± 0.005, which overlapped with a product of the host strain, as a minor product (Figure 2Av). Isolation and structure elucidation led to the identification of **5a** as guanilylated cWF (**1**), named guanitrypmycin C1-1. Expression of *gutD*₂₄₃₀₉ alone did not result in metabolite changes (Figure 2Aviii). Feeding cWF (**1**) to the *gutD*₂₄₃₀₉ transformant led to the formation of **5a** in low yield (ca. 9%) (Figure 2Aix). In contrast, the fed cWΔF (**8**) was completely converted to **4a** and **4b** (Figure 2Ax). These results proved that GutD₂₄₃₀₉ can use cWF (**1**) as substrate, but only with very low efficiency and the product cWΔF (**8**) generated by Gut(BC)₃₅₈₉ (see below for strain construction) serves as the natural substrate of GutD₂₄₃₀₉. Detection of **4a** two days after the feeding of cWΔF (**8**) to the *gutD*₂₄₃₀₉ transformant and of **4b** at the end of cultivation confirmed again the non-enzymatic conversion of **4a** to **4b** (Figure S63). The order of the four enzyme reactions in the biosynthesis of

guanitrypmycins in *S. monomycini* NRRL B-24309 and their non-enzymatic epimerization is depicted in Scheme 2.

The proposed pathway in Scheme 2 was also confirmed by different gene combination expressions of the gene cluster from *S. varsoviensis* NRRL B-3589. When *gutA*₃₅₈₉ was expressed in *S. coelicolor* M1146, both cWF (**1**) and cWY (**2**) were detected and presented in a ratio of 2:1 (Figure S64); hence a more complex product spectrum, four chemical pairs **3a/3b**, **4a/4b**, **6a/6b**, **7a/7b** along with cWΔF (**8**) and cWΔY (**9**), was detected from this cluster (Figure S64). Isolation and structure elucidation confirmed **3a/3b** and **4a/4b** are the same as those of the cluster from NRRL B-24309, which are derived from cWF (**1**). Two further product pairs, **6a/6b** and **7a/7b**, are the corresponding products derived from cWY (**2**), named guanitrypmycin A2-1/A2-2 and B2-1/B2-2, respectively.

Guanitrypmycins in *S. monomycini* NRRL B-24309

To validate the productivity for guanitrypmycins, we performed fermentation of *S. monomycini* NRRL B-24309 and *S. varsoviensis* NRRL B-3589 in six different media. LC-MS analysis revealed the presence of **3a/3b** with $[M+H]^+$ ions at m/z 495.189 ± 0.005 and **4a/4b** with $[M+H]^+$ ions at m/z 481.173 ± 0.005 only in the fermentation broth of *S. monomycini* with modified R5 medium (Figure S65). Scaled-up fermentation of this strain and HPLC-guided purification, and ¹H NMR analysis confirmed unambiguously these products as **3a/3b** and **4a/4b**. In comparison, the *gut*₃₅₈₉ cluster seemed to be totally silent in *S. varsoviensis*: neither cWF (**1**) or cWY (**2**), nor their deduced derivatives could be detected in the extract of *S. varsoviensis* broth (Figure S66).

Biochemical Characterization of GutD

Our in vivo heterologous expression experiment proved that GutD catalyzes the key C–C linkage reaction between cWΔF (**8**)/cWΔY (**9**) and guanine during the biosynthesis of guanitrypmycins. To verify the GutD function in vitro, we amplified *gutD*₂₄₃₀₉ and *gutD*₃₅₈₉ from their genomic DNA and heterologously overexpressed them in *E. coli*. Unfortunately, neither GutD₂₄₃₀₉ nor GutD₃₅₈₉ could be obtained as soluble proteins. Finally, GutD₃₅₈₉ was successfully overproduced in *S. coelicolor* M1146 by using the replicative vector pPWW50A and purified as an N-(His)₁₀-fused protein to near homogeneity (Figure S67). The purified recombinant GutD₃₅₈₉ exhibited a red color and UV absorption maxima at approximately 390 and 420 nm, indicating the presence of both high- and low-spin states of the enzyme. The first state could be due to substrate binding in the active site.^[34] The typical absorption maximum of the sodium dithionite reduced Fe^{II}-CO complex was clearly observed at approximately 450 nm (Figure S67).

The GutD₃₅₈₉ activity was probed by using guanine and cWΔF (**8**) or cWΔY (**9**) as substrates, commercial spinach ferredoxin, and ferredoxin-NADP⁺ reductase for electron transport. The reaction mixtures were analyzed by LC-MS

analysis and the products were identified by comparison of their retention times and $[M+H]^+$ ions with those from the *S. coelicolor* transformants (Figure 3). As shown in Figure 3, **4a** and **4b** were clearly detected in the reaction mixture of cWΔF (**8**), while cWΔY (**9**) was converted to **7a** and **7b**. No conversion of cWΔF (**8**) and cWΔY (**9**) was observed with heat-denatured GutD₃₅₈₉ (Figure 3). The product formation was strictly dependent on the presence of active recombinant protein (Figure 3) and cofactors (Figure S68). Negligible product formation was detected when GTP was used instead of guanine (data not shown).

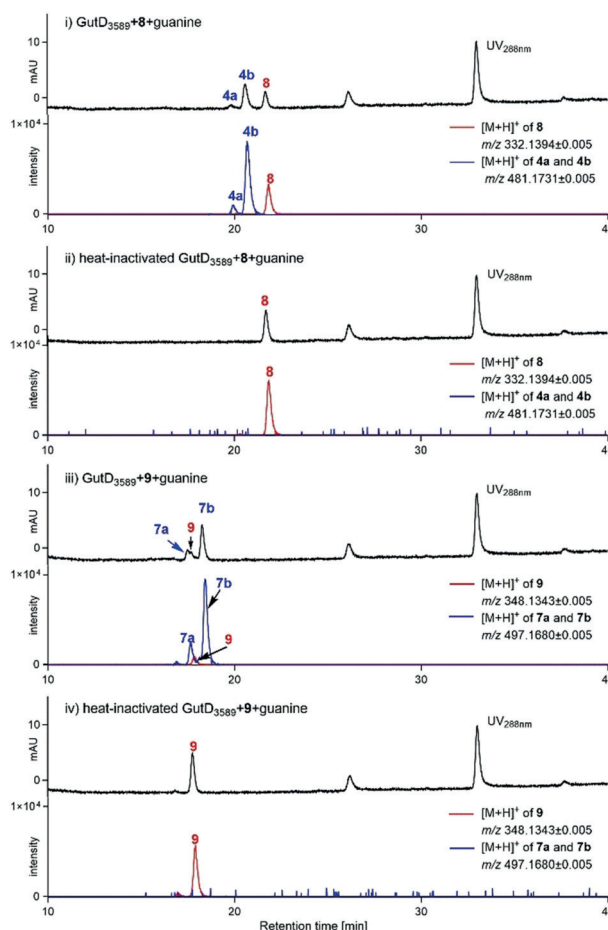


Figure 3. LC-MS monitoring of the in vitro assays of GutD₃₅₈₉.

Conclusion

In conclusion, genome mining revealed the presence of two homologous P450-associated *cdps*-containing gene clusters (*gut*) in *Streptomyces*. Eight guanitrypmycins were identified as C3-guaninyl pyrroloindolines from the transformants harboring the *gut*₂₄₃₀₉ and *gut*₃₅₈₉ gene clusters. Total product yields ranging from 37 to 58 mg per liter culture were calculated for the transformants carrying the whole clusters or lacking the MT gene (Table S14). Testing of guanitrypmycins with the human ovarian adenocarcinoma cells SK-OV-3 shows almost no cytotoxicity.^[35]

Expression of different gene combinations proved that the CDP skeleton assembled by GutA₂₄₃₀₉ and GutA₃₅₈₉ initiates the biosynthesis of guanitrypmycins. In *S. coelicolor* host, GutA₂₄₃₀₉ mainly produces cWF (1), whereas cWF (1) and cWY (2) are products of GutA₃₅₈₉, which serve as DKP precursors and are further metabolized by the three tailoring enzymes. Gut(BC) introduce an exocyclic double bond merely to the phenylalanyl/tyrosyl, but not the tryptophanyl side, which is distinctly different from the two known CDOs AlbA/B and Ndas1146/1147 inserting double bonds to both sides of the DKP rings.^[25,36] Both in vivo expression and in vitro biochemical characterizations reveal that the P450s GutD act as key enzymes and subsequently catalyze the stereospecific coupling of cWΔF (and cWΔY) with a guanine moiety via C3–C8' bond formation. This intriguing reaction is clearly distinguished from other known P450s, extending the functional spectrum of P450 tailoring enzymes (Scheme 1).^[14] As listed in the reviews,^[12,37] P450 coding sequences are predicted in a large number of *cdps*-containing clusters. It can be therefore expected that more novel P450-mediated transformations will be uncovered in the near future. Furthermore, the MTs GutE transfer a methyl group to N9' of the guaninyl residue after P450-catalyzed coupling reaction, which is different from the known MTs in *cdps*-containing clusters catalyzing methylation at the DKP skeletons.^[25,38] The three modification enzymes in the *gut* cluster, that is, CDO, cytochrome P450, and MT, catalyze strictly regioselective tailoring reactions. To the best of our knowledge, this is the first report on guanitrypmycin structures and functional proof of biosynthetic genes, especially the P450 enzymes as C3 α -guaninyl transferases for tryptophan-containing CDPs. Moreover, the non-enzymatic epimerization of 3a and 4a via keto-enol tautomerism increases the structural diversity of guanitrypmycins. In summary, this study provides an excellent example for finding novel natural products by genome mining and exploring the proposed functions of individual biosynthetic enzymes via the heterologous expression approach.

Acknowledgements

We thank Rixa Kraut (Marburg) for LC-MS analysis, Prof. Wenhan Lin (Beijing) for helping with structure elucidation, and the DFG for funding (INST 160/620-1). We also thank ARS Culture Collection (NRRL) for providing *Streptomyces* strains. J.L. (201608310118) received a scholarship from the China Scholarship Council.

Conflict of interest

The authors declare no conflict of interest.

cyclodipeptides · biosynthesis · cytochrome P450 ·
guaninyl transferase · natural products

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 11534–11540
Angew. Chem. **2019**, *131*, 11658–11664

- [1] M. Komatsu, T. Uchiyama, S. Omura, D. E. Cane, H. Ikeda, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2646–2651.
- [2] P. Cimermancic, M. H. Medema, J. Claesen, K. Kurita, L. C. Wieland Brown, K. Mavrommatis, A. Pati, P. A. Godfrey, M. Koehrsen, J. Clardy, B. W. Birren, E. Takano, A. Sali, R. G. Linington, M. A. Fischbach, *Cell* **2014**, *158*, 412–421.
- [3] C. H. R. Senges, A. Al-Dilaimi, D. H. Marchbank, D. Wibberg, A. Winkler, B. Haltli, M. Nowrousian, J. Kalinowski, R. G. Kerr, J. E. Bandow, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 2490–2495.
- [4] A. D. Borthwick, *Chem. Rev.* **2012**, *112*, 3641–3716.
- [5] S.-M. Li, *Nat. Prod. Rep.* **2010**, *27*, 57–78.
- [6] K. R. Klas, H. Kato, J. C. Frisvad, F. Yu, S. A. Newmister, A. E. Fraley, D. H. Sherman, S. Tsukamoto, R. M. Williams, *Nat. Prod. Rep.* **2018**, *35*, 532–558.
- [7] W. Xu, D. J. Gavia, Y. Tang, *Nat. Prod. Rep.* **2014**, *31*, 1474–1487.
- [8] Y. M. Ma, X. A. Liang, Y. Kong, B. Jia, *J. Agric. Food Chem.* **2016**, *64*, 6659–6671.
- [9] C. T. Walsh, *Nat. Prod. Rep.* **2016**, *33*, 127–135.
- [10] M. Gondry, L. Sauguet, P. Belin, R. Thai, R. Amouroux, C. Tellier, K. Tuphile, M. Jacquet, S. Braud, M. Courcon, C. Masson, S. Dubois, S. Lautru, A. Lecoq, S. Hashimoto, R. Genet, J. L. Pernodet, *Nat. Chem. Biol.* **2009**, *5*, 414–420.
- [11] 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.
- [12] T. W. Giessen, M. A. Marahiel, *Int. J. Mol. Sci.* **2014**, *15*, 14610–14631.
- [13] R. D. Süßmuth, A. Mainz, *Angew. Chem. Int. Ed.* **2017**, *56*, 3770–3821; *Angew. Chem.* **2017**, *129*, 3824–3878.
- [14] P. Borgman, R. D. Lopez, A. L. Lane, *Org. Biomol. Chem.* **2019**, *17*, 2305–2314.
- [15] J. Liu, H. Yu, S.-M. Li, *Appl. Microbiol. Biotechnol.* **2018**, *102*, 4435–4444.
- [16] 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.
- [17] F. P. Guengerich, A. W. Munro, *J. Biol. Chem.* **2013**, *288*, 17065–17073.
- [18] A. Greule, J. E. Stok, J. J. De Voss, M. J. Cryle, *Nat. Prod. Rep.* **2018**, *35*, 757–791.
- [19] J. D. Rudolf, C. Y. Chang, M. Ma, B. Shen, *Nat. Prod. Rep.* **2017**, *34*, 1141–1172.
- [20] 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.
- [21] M. J. Cryle, S. G. Bell, I. Schlichting, *Biochemistry* **2010**, *49*, 7282–7296.
- [22] 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.
- [23] H. Yu, X. Xie, S.-M. Li, *Org. Lett.* **2018**, *20*, 4921–4925.
- [24] W. Tian, C. Sun, M. Zheng, J. R. Harmer, M. Yu, Y. Zhang, H. Peng, D. Zhu, Z. Deng, S. L. Chen, M. Mobli, X. Jia, X. Qu, *Nat. Commun.* **2018**, *9*, 4428.
- [25] T. W. Giessen, A. M. von Tesmar, M. A. Marahiel, *Chem. Biol.* **2013**, *20*, 828–838.
- [26] T. W. Giessen, M. A. Marahiel, *Front. Microbiol.* **2015**, *6*, 785.
- [27] L. M. Podust, D. H. Sherman, *Nat. Prod. Rep.* **2012**, *29*, 1251–1266.
- [28] J. P. Gomez-Escribano, M. J. Bibb, *Microb. Biotechnol.* **2011**, *4*, 207–215.
- [29] Y. Zhu, P. Fu, Q. Lin, G. Zhang, H. Zhang, S. Li, J. Ju, W. Zhu, C. Zhang, *Org. Lett.* **2012**, *14*, 2666–2669.

- [30] C. Huang, C. Yang, W. Zhang, L. Zhang, B. C. De, Y. Zhu, X. Jiang, C. Fang, Q. Zhang, C. S. Yuan, H. W. Liu, C. Zhang, *Nat. Commun.* **2018**, *9*, 2088.
- [31] L. D. Marzilli, B. de Castro, C. Solorzano, *J. Am. Chem. Soc.* **1982**, *104*, 461–466.
- [32] M. Teijeira, L. Santana, E. Uriarte, *Magn. Reson. Chem.* **1997**, *35*, 806–807.
- [33] E. Iglesias, *J. Org. Chem.* **2003**, *68*, 2680–2688.
- [34] J. L. Grant, C. H. Hsieh, T. M. Makris, *J. Am. Chem. Soc.* **2015**, *137*, 4940–4943.
- [35] K. Backhaus, S.-M. Li, unpublished results.
- [36] S. Lautru, M. Gondry, R. Genet, J. L. Pernodet, *Chem. Biol.* **2002**, *9*, 1355–1364.
- [37] P. Belin, M. Moutiez, S. Lautru, J. Seguin, J. L. Pernodet, M. Gondry, *Nat. Prod. Rep.* **2012**, *29*, 961–979.
- [38] T. W. Giessen, A. M. von Tesmar, M. A. Marahiel, *Biochemistry* **2013**, *52*, 4274–4283.

Manuscript received: June 3, 2019

Accepted manuscript online: June 17, 2019

Version of record online: July 5, 2019

Supporting Information

Guanitrypmycin Biosynthetic Pathways Imply Cytochrome P450 Mediated Regio- and Stereospecific Guaninyl-Transfer Reactions

*Jing Liu, Xiulan Xie, and Shu-Ming Li**

anie_201906891_sm_miscellaneous_information.pdf

Table of contents	Page
1. Computer-assisted sequence analysis	S4
2. Bacterial strains, plasmids, and culture conditions	S4
3. Genetic manipulation, PCR amplification, and gene cloning	S4-S5
4. Heterologous expression in <i>S. coelicolor</i> M1146	S5
5. Overproduction and purification of GutD ₃₅₈₉ in <i>Streptomyces</i>	S5-S6
6. UV-Vis spectroscopic analysis of GutD ₃₅₈₉	S6
7. Enzyme assays of GutD ₃₅₈₉	S6
8. LC-MS analysis	S6-S7
9. Isolation of the generated metabolites	S7
10. Precursor feeding experiments	S7
11. Heterologous expression of <i>gut₃₅₈₉</i> cluster in <i>S. coelicolor</i>	S7-S8
12. Determination of product yields for 3a , 3b , 4a , and 4b in <i>Streptomyces</i> transformants	S8
13. NMR measurements	S9
14. Circular dichroism (CD) spectroscopic analysis	S9
15. Physicochemical properties of the compounds described in this study	S9-S10
16. Structural elucidation	S10
Table S1. Comparison of CDPS-containing gene cluster from <i>Streptomyces</i> strains	S11
Table S2. Bacterial strains used in this study	S12
Table S3. Gene amplification and cloning	S13
Table S4. Expression constructs in pPWW50A	S14
Table S5. NMR data of guanitrypmycin A1-1 (3a)	S15
Table S6. NMR data of guanitrypmycin A1-2 (3b)	S16
Table S7. NMR data of guanitrypmycin B1-1 (4a)	S17
Table S8. NMR data of guanitrypmycin B1-2 (4b)	S18
Table S9. NMR data of guanitrypmycin C1-1 (5a)	S19
Table S10. NMR data of guanitrypmycin A2-1 (6a)	S20
Table S11. NMR data of guanitrypmycin A2-2 (6b)	S21
Table S12. NMR data of guanitrypmycin B2-1 (7a)	S22
Table S13. NMR data of cWAF (8) and cWΔY (9)	S23
Table S14. Products yields of the transformants	S24
Figure S1. Sequence alignments of P450 enzymes	S25
Figure S2. UV spectra of the isolated products	S26
Figure S3. ¹ H NMR spectrum of guanitrypmycin A1-1 (3a)	S27
Figure S4. ¹³ C APT NMR spectrum of guanitrypmycin A1-1 (3a)	S28
Figure S5. ¹ H- ¹ H COSY spectrum of guanitrypmycin A1-1 (3a)	S29
Figure S6. HSQC spectrum of guanitrypmycin A1-1 (3a)	S30
Figure S7. HMBC spectrum of guanitrypmycin A1-1 (3a)	S31
Figure S8. NOESY spectrum of guanitrypmycin A1-1 (3a)	S32
Figure S9. ¹ H NMR spectrum of guanitrypmycin A1-2 (3b)	S33
Figure S10. ¹³ C NMR spectrum of guanitrypmycin A1-2 (3b)	S34
Figure S11. ¹ H- ¹ H COSY spectrum of guanitrypmycin A1-2 (3b)	S35
Figure S12. HSQC spectrum of guanitrypmycin A1-2 (3b)	S36

Figure S13. HMBC spectrum of guanitrypmycin A1-2 (3b)	S37
Figure S14. NOESY spectrum of guanitrypmycin A1-2 (3b)	S38
Figure S15. ¹ H NMR spectrum of guanitrypmycin B1-1 (4a)	S39
Figure S16. ¹³ C APT NMR spectrum of guanitrypmycin B1-1 (4a)	S40
Figure S17. ¹ H- ¹ H COSY spectrum of guanitrypmycin B1-1 (4a)	S41
Figure S18. HSQC spectrum of guanitrypmycin B1-1 (4a)	S42
Figure S19. HMBC spectrum of guanitrypmycin B1-1 (4a)	S43
Figure S20. NOESY spectrum of guanitrypmycin B1-1 (4a)	S44
Figure S21. ¹ H NMR spectrum of guanitrypmycin B1-2 (4b)	S45
Figure S22. ¹³ C APT NMR spectrum of guanitrypmycin B1-2 (4b)	S46
Figure S23. ¹ H- ¹ H COSY spectrum of guanitrypmycin B1-2 (4b)	S47
Figure S24. HSQC spectrum of guanitrypmycin B1-2 (4b)	S48
Figure S25. HMBC spectrum of guanitrypmycin B1-2 (4b)	S49
Figure S26. NOESY spectrum of guanitrypmycin B1-2 (4b)	S50
Figure S27. ¹ H NMR spectrum of guanitrypmycin C1-1 (5a)	S51
Figure S28. ¹³ C APT NMR spectrum of guanitrypmycin C1-1 (5a)	S52
Figure S29. ¹ H- ¹ H COSY spectrum of guanitrypmycin C1-1 (5a)	S53
Figure S30. HSQC spectrum of guanitrypmycin C1-1 (5a)	S54
Figure S31. HMBC spectrum of guanitrypmycin C1-1 (5a)	S55
Figure S32. NOESY spectrum of guanitrypmycin C1-1 (5a)	S56
Figure S33. ¹ H NMR spectrum of guanitrypmycin A2-1 (6a)	S57
Figure S34. ¹³ C APT NMR spectrum of guanitrypmycin A2-1 (6a)	S58
Figure S35. ¹ H- ¹ H COSY spectrum of guanitrypmycin A2-1 (6a)	S59
Figure S36. HSQC spectrum of guanitrypmycin A2-1 (6a)	S60
Figure S37. HMBC spectrum of guanitrypmycin A2-1 (6a)	S61
Figure S38. NOESY spectrum of guanitrypmycin A2-1 (6a)	S62
Figure S39. ¹ H NMR spectrum of guanitrypmycin A2-2 (6b)	S63
Figure S40. ¹³ C APT NMR spectrum of guanitrypmycin A2-2 (6b)	S64
Figure S41. ¹ H- ¹ H COSY spectrum of guanitrypmycin A2-2 (6b)	S65
Figure S42. HSQC spectrum of guanitrypmycin A2-2 (6b)	S66
Figure S43. HMBC spectrum of guanitrypmycin A2-2 (6b)	S67
Figure S44. NOESY spectrum of guanitrypmycin A2-2 (6b)	S68
Figure S45. ¹ H NMR spectrum of guanitrypmycin B2-1 (7a)	S69
Figure S46. ¹³ C APT NMR spectrum of guanitrypmycin B2-1 (7a)	S70
Figure S47. HSQC spectrum of guanitrypmycin B2-1 (7a)	S71
Figure S48. HMBC spectrum of guanitrypmycin B2-1 (7a)	S72
Figure S49. NOESY spectrum of guanitrypmycin B2-1 (7a)	S73
Figure S50. ¹ H NMR spectrum of cWΔF (8)	S74
Figure S51. ¹ H NMR spectrum of cWΔY (9)	S75
Figure S52. CD spectra of 3a , 4a , 6a , and 7a in MeOH	S76
Figure S53. CD spectra of 3b and 4b in MeOH	S76
Figure S54. Time-dependent product formation	S77
Figure S55. LC-MS analysis for determination of the stability of 3a and 3b in GYM	S78

Figure S56. LC-MS analysis for determination of the stability of 4a and 4b in GYM	S79
Figure S57. LC-MS analysis of 3a and 3b after incubation in CD ₃ OD/D ₂ O at pH 8.0	S80
Figure S58. LC-MS analysis of 3a and 3b after incubation in CD ₃ OD/D ₂ O with 0.01M NaOH	S81
Figure S59. LC-MS analysis of 3a and 3b after incubation in CD ₃ OD/D ₂ O with 0.01M HCl	S82
Figure S60. LC-MS analysis of 4a and 4b after incubation in CD ₃ OD/D ₂ O at pH 8.0	S83
Figure S61. LC-MS analysis of 4a and 4b after incubation in CD ₃ OD/D ₂ O with 0.01M NaOH	S84
Figure S62. LC-MS analysis of 4a and 4b after incubation in CD ₃ OD/D ₂ O with 0.01M HCl	S85
Figure S63. Time-dependent product formation in <i>S. coelicolor</i> M1146 harboring <i>gutD</i> ₂₄₃₀₉ after fed with cWΔF (8)	S86
Figure S64. HPLC analysis of <i>S. coelicolor</i> carrying genes from <i>S. varsoviensis</i> NRRL B-3589	S87
Figure S65. LC-MS analysis of metabolite profile of <i>S. monomykini</i> NRRL B-24309	S88
Figure S66. LC-MS analysis of metabolite profile of <i>S. varsoviensis</i> NRRL B-3589	S89
Figure S67. SDS-PAGE analysis of the purified GutD ₃₅₈₉ and the absorption spectra of GutD ₃₅₈₉	S90
Figure S68. LC-MS analysis of enzyme assays of GutD ₃₅₈₉	S91
References	S92

1. Computer-assisted sequence analysis

Protein sequences used in this study were taken from the NCBI database (<http://www.ncbi.nlm.nih.gov/protein>) and compared with each other by using BLASTP program (<http://blast.ncbi.nlm.nih.gov/>). Multiple sequence alignments were carried out with the program ClustalW and visualized with ESPript 3.2 (<http://esprict.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>) to identify strictly conserved amino acid residues.

2. Bacterial strains, plasmids, and culture conditions

Bacterial strains and plasmids used in this study are listed in **Tables S2** and **S3**, respectively. Liquid or solid Luria-Bertani (LB) medium with agarose was used for cultivation of *E. coli* and 100 µg/mL ampicillin, 50µg/mL kanamycin, 50 µg/mL apramycin or 25 µg/mL chloramphenicol were used for selection when necessary.

Streptomyces strains were kindly provided by ARS Culture Collection (NRRL). They were maintained in GYM (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. *Streptomyces coelicolor* M1146 and the exconjugants were cultured on MS media (mannitol 20.0 g/L, soya flour 20.0 g/L, agar 15.0 g/L). For secondary metabolite production, *S. monomycini* and *S. varsoviensis* were cultivated in modified R5 medium (sucrose 103.0 g/L, glucose 10.0 g/L, yeast extract 5.0 g/L, MgCl₂.6H₂O 10.12 g/L, K₂SO₄ 0.25 g/L, Difco casaminoacids 0.1 g/L, MOPS 21.0 g/L, trace element solution 2 mL/L, pH 7.2) at 28 °C for 7 days.¹

3. Genetic manipulation, PCR amplification, and gene cloning

Genetic manipulation in *E. coli* was carried out according to the protocol by Sambrook and Russell.² Isolation of genomic DNA from *Streptomyces* was performed as described in the literature.¹ The genes mentioned in this paper were amplified from genomic DNA by using primers listed in **Table S3** and Phusion® High-Fidelity DNA Polymerase from New England Biolabs (NEB). The generated PCR fragments were cloned into pGEM-T Easy vector and sequenced to confirm sequence integrity. After sequencing, the fragments were released with

the specific restriction endonucleases from pGEM-T Easy vector and ligated into pPWW50A vector,³ which was digested with the same enzymes, previously. The generated plasmids (**Table S4**) were transformed into *S. coelicolor* M1146 for gene expression.

4. Heterologous expression in *S. coelicolor* M1146

The constructed plasmids harboring different genes or gene clusters were firstly transformed into non-methylating *E. coli* ET12567/pUZ8002, which was subjected to conjugation with *S. coelicolor* M1146. The positive conjugants were first selected by the phenotype showing apramycin resistance and further confirmed by PCR. For gene expression, *S. coelicolor* M1146 recombinant strains were inoculated in GYM liquid media supplied with 50 µg/mL of apramycin in 250 mL baffled flask and cultured at 28 °C, 200 rpm for 7 days. 1 mL culture was extracted with one volume of ethyl acetate for three times. The organic phases were combined, the solvent evaporated and the residue afterward resolved in 50 µL of methanol. 5 µL of such samples were taken for LC-MS analysis.

5. Overproduction and purification of GutD₃₅₈₉ in *Streptomyces*

S. coelicolor M1146 harboring pJL40 (*gutD*₃₅₈₉ in pPWW50A, **Table S4**) was cultivated in 50 mL tryptic soy broth (TSB) medium containing 50 µg/mL apramycin for 48 h. 5 mL of this pre-culture were transferred to 100 mL TSB with 50 µg/mL apramycin in 500 mL conical flasks. The cultures were further incubated at 28°C and 200 rpm for 3 days. One liter of such cultures were harvested by centrifugation at 4 °C and 4500 rpm for 20 min. The pellets were resuspended in lysis buffer (50 mM Tris-HCl, 10 mM imidazole, 300 mM NaCl, pH 8.0) in a ratio of 2 – 5 mL per gram wet weight. Lysozyme from the chicken egg white was added to a final concentration of 1 mg/mL and incubated on ice for 30 min. After sonication, the crude protein extract was collected by centrifugation at 13000 rpm and 4 °C for 1 h. One-step purification of the recombinant His-tagged protein was performed by using Ni-NTA agarose (Macherey-Nagel, Düren, Germany) and eluted with 500 mM imidazole in 50 mM Tris-HCl, 300 mM NaCl buffer (pH 8.0). The purified protein was desalted through PD-10 Desalting Column (GE Healthcare, USA) according to the manufacturers' protocol and stored in 50 mM Tris-HCl buffer containing

15% (v/v) glycerol (pH 7.5) at $-80\text{ }^{\circ}\text{C}$. GutD₃₅₈₉ concentration was determined on a Nanodrop C2000 (Thermo Scientific, Braunschweig, Germany) to be 0.3 mg/mL. The purity of the obtained protein was proven on a 12% (w/v) SDS-PAGE (**Figure S67**).

6. UV-Vis spectroscopic analysis of GutD₃₅₈₉

To measure the typical absorbance of P450 ferrous-CO complex after reduction, carbon monoxide (CO) gas was bubbled into the GutD₃₅₈₉ solution (50 mM Tris-HCl, 15% (v/v) glycerol, pH 7.5) for 2 min. After addition of 0.2 mg/mL of sodium dithionite, a UV-Vis spectrum between 360 and 550 nm was recorded on a Multiskan™ GO Microplate Spectrophotometer (Thermo Scientific, Dreieich, Germany). UV-Vis spectra of GutD₃₅₈₉ before and after treatment with CO were also taken as controls. The spectra are given in **Figure S67**.

7. Enzyme assays of GutD₃₅₈₉

A standard GutD₃₅₈₉ assay contained 8 μM **8** or 13 μM **9**, 3.3 μM GutD₃₅₈₉, 0.11 mM guanine, 2 mM NADPH, 2 μM spinach ferredoxin (Sigma-Aldrich), 0.1 unit/mL spinach ferredoxin-NADP⁺ reductase (Sigma-Aldrich), 100 mM Tris-HCl buffer (pH 8.5) in a total volume of 50 μL . The reaction was performed at 30 $^{\circ}\text{C}$ for 16 h and then quenched with 50 μL MeOH. After centrifugation at 13000 rpm for 10 min, 5 μL of the supernatants were subjected to LC-MS analysis. The negative controls with heat-inactivated GutD₃₅₈₉ were performed under the same conditions.

8. LC-MS analysis

The ethyl acetate extracts were analyzed on an Agilent HPLC 1260 series system equipped with a photo diode array detector and a Bruker microTOF QIII mass spectrometer by using a Multospher 120 RP-18 column (250x4mm, 5 μm , CS-Chromatographie Service, Langerwehe, Germany). A linear gradient of 5 – 100 % acetonitrile in water, both containing 0.1 % formic acid, in 40 min and a flow rate at 0.25 mL/min were used. The column was then washed with 100 % acetonitrile containing 0.1 % formic acid for 5 min and equilibrated with 5 % acetonitrile in water 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.

9. Isolation of the generated metabolites

For structural elucidation of the accumulated compounds, different expression transformants harboring different genes or gene clusters were cultivated in GYM medium on a large scale (8 L or more) at 28 °C for 7 days. The supernatants were then collected and extracted with ethyl acetate for three times. The EtOAc phases were evaporated to dryness and dissolved in a 1:1 mixture of CH₂Cl₂:MeOH and mixed with an appropriate amount of silica gel for normal phase silica gel column chromatography, eluted with a gradient elution of CH₂Cl₂:MeOH in ratios of 100:2, 100:3, 100:5, 100:10. The products were detected by LC-MS analysis and the target compounds were found mainly in the fractions eluted with solvent ratio of 100:10. The fractions were further purified on an Agilent HPLC 1260 series equipped with a photo diode array detector by using a semi-preparative Agilent ZORBAX Eclipse XDB C18 HPLC column (250 ×9.4 mm, 5µm) to get different products. The flow rate was set to 2.0 mL/min. **3a**, **3b**, **6a**, **6b** and **8** were purified with a linear gradient of 35% ACN in water, **4a**, **4b**, **5a**, **7a** and **9** of 30% ACN in water.

10. Precursor feeding experiments

Precursor feeding was carried out by using 20 mM stock solutions in DMSO. 25 µL of these solutions were added to 50 mL of 2 day-old cultures of *Streptomyces* transformants in GYM media. After cultivation at 28°C for additional 7 days, the metabolites were extracted with EtOAc and analyzed on LC-MS.

11. Heterologous expression of *gut*₃₅₈₉ cluster in *S. coelicolor* and identification of metabolites derived from cWF (1) and cWY (2)

When *gutA*₃₅₈₉ was expressed in *S. coelicolor* M1146, cWF (1) and cWY (2) in a ratio of 2:1 were detected (**Figure S64**). Therefore, we expected a more complex product spectrum from this cluster. Indeed, expression of the whole cluster *gut(ABCDE)*₃₅₈₉ led to the accumulation of at least 12 products including four pairs **3a/3b**, **4a/4b**, **6a/6b**, and **7a/7b** (**Figure S64**). Isolation

and structure elucidation confirmed the same products **3a/3b** and **4a/4b** as those of the cluster from NRRL B-24309, which are derived from cWF (**1**). The two product pairs **6a/6b** and **7a/7b** are the corresponding products derived from cWY (**2**), named guanitrypmycin A2-1/A2-2 and B2-1/B2-2, respectively (**Scheme 2**). The configurations of **6a**, **6b** and **7a** were determined by interpretation of NOE correlations and comparison of their CD spectra with those of **3a**, **3b**, **4a**, and **4b** (**Figure S52** and **S53**). Expression of the cluster lacking the putative MT gene *gutE*₃₅₈₉ resulted in the formation of **4a/4b** and **7a/7b**, but not the methylated products **3a/3b** and **6a/6b** (**Figure S64**). Significant amount of cWY (**2**), but not cWF (**1**) was detected in both transformants with *gut(ABCDE)*₃₅₈₉ and *gut(ABCD)*₃₅₈₉ indicates that cWF (**1**) is more effectively used by the tailoring enzymes. Detection of cWΔF (**8**) and cWΔY (**9**) in both transformants provides additional evidence that CDO reaction takes places before that of P450 (**Scheme 2**). This was also confirmed by expression of *gut(AD)*₃₅₈₉, *gut(ADE)*₃₅₈₉, *gut(ABC)*_{3589'}, and *gut(AE)*₃₅₈₉. Expression of *gut(ABC)*₃₅₈₉ accumulated cWF (**1**), cWY (**2**), cWΔF (**8**), and cWΔY (**9**) in a ratio of 2.3:1.2:0.7:1.0, whereas almost no consumption of cWF (**1**) and cWY (**2**) was detected in other transformants (**Figures S64**). Precursor feeding to the *gutD*₃₅₈₉ transformant revealed that cWΔF (**8**) and cWΔY (**9**) were converted to **4a/4b** and **7a/7b**, respectively, while no changes were detected after feeding with cWF (**1**) and cWY (**2**) (**Figures 64**).

12. Determination of product yields for **3a**, **3b**, **4a**, and **4b** in *Streptomyces* transformants

An Agilent HPLC 1200 series equipped with a photo diode array detector and an Agilent Eclipse XDB C18 column (5µm, 4.6 × 150 mm) were used for quantification. A linear gradient of 10 to 100% acetonitrile in water in 40 min was followed by 100% acetonitrile for 5 min and then 10% acetonitrile in water for 5 min. The flow rate was set to 0.5 mL/min. The absorption at 288 nm was used for quantification. 1 mL culture of M1146 transformants was extracted with 1 mL ethyl acetate for 3 times. The organic phases were combined and evaporated to dryness. The residues were dissolved in 100 µL of methanol and 20 µL were analyzed on HPLC. The isolated products were used as authentic standards for quantification.

13. NMR analysis

All the purified compounds were dissolved in DMSO-*d*₆ for NMR analysis. NMR experiments were performed at 300K on a Bruker AVIII spectrometer (500 MHz) equipped with a 5 mm BBO cryo probe Prodigy with z-gradient. All spectra were processed with MestReNova 5.2.2 (Metrelab Research, S5 Santiago de Compostella, Spain). NMR spectra and data of the identified compounds are provided as **Figures S3 - S51** and **Tables S5 – S13**, respectively.

14. Circular dichroism (CD) spectroscopic analysis

CD spectra were taken on a J-815 CD spectrometer (Jasco Deutschland GmbH, Pfungstadt, Germany). The samples were dissolved in methanol and measured in the range of 200–400 nm by using a 1 mm path length quartz cuvette (Hellma Analytics, Müllheim, Germany). The CD spectra are given in **Figures S52 - S53**.

15. Physicochemical properties of the compounds described in this study

Guanitrypmycin A1-1 (**3a**): 12 mg, light yellow powder; CD (MeOH) λ max ($\Delta\epsilon$) 293 (+69.9), 250 (-10.7), 229 (+64.2), 214 (+84.7) nm; HRMS (m/z): (ESI/[M + H]⁺) calcd. for C₂₆H₂₃N₈O₃, 495.1888, found 495.1912.

Guanitrypmycin A1-2 (**3b**): 13 mg, light yellow powder; CD (MeOH) λ max ($\Delta\epsilon$) 294 (+58.2), 265 (+36.5), 246 (-72.9), 230 (+14.5); HRMS (m/z): (ESI/[M + H]⁺) calcd. for C₂₆H₂₃N₈O₃, 495.1888, found 495.1918.

Guanitrypmycin B1-1 (**4a**): 8 mg, light yellow powder; CD (MeOH) λ max ($\Delta\epsilon$) 295 (+64.8), 248 (-17.6), 227 (+48.9), 213 (+77.1) nm; HRMS (m/z): (ESI/[M + H]⁺) calcd. for C₂₅H₂₁N₈O₃, 481.1731, found 481.1755.

Guanitrypmycin B1-2 (**4b**): 10 mg, light yellow powder; CD (MeOH) λ max ($\Delta\epsilon$) 298 (+45.1), 267 (+10.8), 245 (-82.4), 230 (-38.3); HRMS (m/z): (ESI/[M + H]⁺) calcd. for C₂₅H₂₁N₈O₃, 481.1731, found 481.1761.

Guanitrypmycin C1-1 (**5a**): 8 mg, light yellow powder; HRMS (m/z): (ESI/[M + H]⁺) calcd. for C₂₅H₂₃N₈O₃, 483.1888, found 483.1901.

Guanitrypmycin A2-1 (**6a**): 3 mg, light yellow powder; CD (MeOH) λ max ($\Delta\epsilon$) 304 (+22.8),

252 (-10.9), 215 (+29.5), 206 (-1.5) nm; HRMS (m/z): (ESI/[M + H]⁺) calcd. for C₂₆H₂₃N₈O₄, 511.1837, found 511.1848.

Guanitrypmycin A2-2 (**6b**): 5 mg, light yellow powder; HRMS (m/z): (ESI/[M + H]⁺) calcd. for C₂₆H₂₃N₈O₄, 511.1837, found 511.1861.

Guanitrypmycin B2-1 (**7a**): 4 mg, light yellow powder; CD (MeOH) λ max ($\Delta\epsilon$) 304 (+31.2), 250 (-16.2), 217 (+41.6), 204 (+5.9) nm; HRMS (m/z): (ESI/[M + H]⁺) calcd. for C₂₅H₂₁N₈O₄, 497.1680, found 497.1704.

16. Structural elucidation

The structures of the isolated products were elucidated by comprehensive interpretation of their UV (**Figure S2**), NMR (**Tables S5–13** and **Figures S3–51**), and CD spectra (**Figures S52** and **S53**). The didehydro products of CDO_{24309/3589}, cW Δ F (**8**) and cW Δ F (**9**), were identified by comparing their ¹H NMR data with those of cWF (**1**) and cW Δ Y (**2**). The signals of the three coupling protons at H-14 and H-17 of **1** and **2** disappeared in the spectra of **8** and **9**. Instead, an olefinic singlet was observed at approx. 6.2 ppm. This proton has a resonance at approx. 6.7 ppm in the spectra of **3a**, **3b**, **4a**, **4b**, **6a**, **6b**, and **7a**. The characteristic ¹³C signals of the guaninyl residue were found at approx. 157, 154, 154, 147, and 115 ppm.⁴ For the typical hexahydropyrrolo [2, 3-b] indole framework in guanitrypmycins, the ring formation between H-2 and N-12 resulted in the destroying the aromatic characters of the five-membered ring, so that the signal of H-2 was upfield shifted from approx. 7.0 to 5.9–6.3 ppm. The signal of C-2 was upfield shifted from approx. 125 to 80 ppm and C-3 from 110 to 55 ppm. The connections and configurations of the structural elements were confirmed by interpretation of the key correlations in the HMBC and NOESY spectra (**Tables S5-13**). **Figures S52** and **S53** showed almost the same CD spectra for compounds with the same configuration.

Table S1. Comparison of CDPS-containing gene clusters in two *Streptomyces* strains.

<i>S. monomycini</i> NRRL B-24309				<i>S. varsoviensis</i> NRRL B-3589				Sequence identity
Accession No.	Size (aa)	Gut protein	Putative function	Accession No.	Size (aa)	Gut protein	Putative function	
WP_030022365.1	110		hypothetical protein	WP_030877678.1	276		inositol monophosphatase	
WP_078624485.1	126		divalent cation tolerance protein CutA	WP_037963123.1	474		NAD(P)/FAD-dependent oxidoreductase	
No protein_id	310		hypothetical protein	WP_030877674.1	126		hypothetical protein	
WP_078624486.1	231	GutE ₂₄₃₀₉	SAM-dependent methyltransferase	WP_078645497.1	229	GutE ₃₅₈₉	SAM-dependent methyltransferase	86%
WP_078624487.1	282	GutA ₂₄₃₀₉	cyclodipeptide synthase, CDPS	WP_078645499.1	305	GutA ₃₅₈₉	cyclodipeptide synthase, CDPS	82%
WP_050502760.1	411	GutD ₂₄₃₀₉	cytochrome P450	WP_048832742.1	415	GutD ₃₅₈₉	cytochrome P450	87%
WP_033038824.1	158	GutB ₂₄₃₀₉	cyclodipeptide oxidase, CDO subunit A	WP_048832786.1	175	GutB ₃₅₈₉	cyclodipeptide oxidase, CDO subunit A	90%
WP_050502761.1	114	GutC ₂₄₃₀₉	cyclodipeptide oxidase, CDO subunit B	WP_014908976.1	95	GutC ₃₅₈₉	cyclodipeptide oxidase, CDO subunit B	79%
WP_078624488.1	246		hypothetical protein	WP_030877658.1	490		DNA-3-methyladenine glycosylase	
WP_050502762.1	162		glutamylcyclotransferase	WP_030877655.1	167		DNA methyltransferase	
WP_050502763.1	338		adenosine deaminase	WP_048832744.1	464		hypothetical protein	

Table S2. Bacterial strains used in this study.

Strains	Source	Cultivation Media
<i>E. coli</i> DH5 α (for cloning)	Invitrogen	LB
<i>E. coli</i> ET12567/pUZ8002	5	LB
<i>Streptomyces varsoviensis</i> NRRL B-3589	NRRL	GYM
<i>Streptomyces monomycini</i> NRRL B-24309	NRRL	GYM
<i>Streptomyces coelicolor</i> M1146	5	MS

NRRL: Agricultural Research Service (ARS) Culture Collection

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

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

MS medium: mannitol 20.0 g/L, soya flour 20.0 g/L, agar 15.0 g/L, pH 7.2.

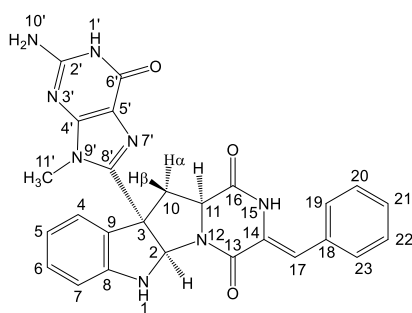
Table S3. Gene amplification and cloning.

Gene	Primer sequences (5'-3')	Cloning sites for pPWW50A	Cloning constructs
<i>gutA</i> ₂₄₃₀₉	GCG <u>CATATG</u> AGTGCATCGCAGGCTGCTG GGATCCTCACGTCACGTCCCCTGC	<i>Nde</i> I <i>Bam</i> HI	pJL31
<i>gutD</i> ₂₄₃₀₉	<u>CATATG</u> AGCGGACGGCCTCCCGGC GGATCCTCACCAGAGCACCGGCAGG	<i>Nde</i> I <i>Bam</i> HI	pJL36
<i>gutE</i> ₂₄₃₀₉	GGATCCTGTCAATTCCTGCGGATCGTTCGC <u>ACTAGTTC</u> ACCCGGAGCGATCCGGA	<i>Bam</i> HI <i>Spe</i> I	pJL34
<i>gut(AD)</i> ₂₄₃₀₉	GCG <u>CATATG</u> AGTGCATCGCAGGCTGCTG GGATCCTCACCAGAGCACCGGCAGG	<i>Nde</i> I <i>Bam</i> HI	pJL35
<i>gut(ABCD)</i> ₂₄₃₀₉	GCG <u>CATATG</u> AGCAGCAACAACAGTTACTTC GGATCCTCACCAGAGCACCGGCAGG	<i>Nde</i> I <i>Bam</i> HI	pJL33
<i>gutA</i> ₃₅₈₀₉	<u>CATATG</u> GGGGGCCCCGCAGCCC_ GGATCCTCACGTCAAGTCCCTTTCTCC	<i>Nde</i> I <i>Bam</i> HI	pJL32
<i>gut(BC)</i> ₃₅₈₀₉	GT <u>GGATCC</u> ATGAGCCGCCAGGAGCGCAC <u>ACTAGTTC</u> AGGCGGGCGGGCGGG	<i>Bam</i> HI <i>Spe</i> I	pJL41
<i>gutD</i> ₃₅₈₀₉	G <u>CATATG</u> AACGCACAGTCCGCGACGGGC CGGATCCTCACCAGAGCACCGGCAGGCG	<i>Nde</i> I <i>Bam</i> HI	pJL40
<i>gutE</i> ₃₅₈₀₉	GGATCCGTCGCGGCCTGTCACTCAC <u>ACTAGTTC</u> AGGCGCGGTTGCCGTC	<i>Bam</i> HI <i>Spe</i> I	pJL38
<i>gut(AD)</i> ₃₅₈₀₉	<u>CATATG</u> GGGGGCCCCGCAGCCC CGGATCCTCACCAGAGCACCGGCAGGCG	<i>Nde</i> I <i>Bam</i> HI	pJL39
<i>gut(ABCD)</i> ₃₅₈₀₉	<u>CATATG</u> GGGGGCCCCGCAGCCCACTGTC GGATCCTCAGGCGGGCGGGCGGGG	<i>Nde</i> I <i>Bam</i> HI	pJL37

Restriction sites for cloning are underlined in the primer sequences.

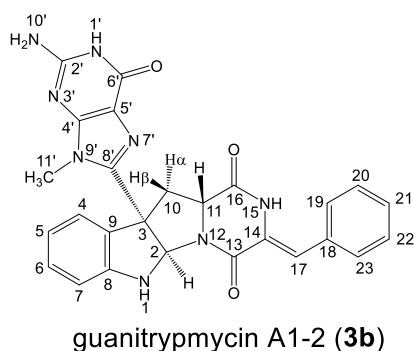
Table S4. Expression constructs in pPWW50A.

Expression constructs	Genes to be expressed	Fragments from cloning vectors
pJL51	<i>gutA</i> ₂₄₃₀₉	pJL31
pJL53	<i>gut(ABCDE)</i> ₂₄₃₀₉	pJL33+pJL34
pJL54	<i>gut(ABCD)</i> ₂₄₃₀₉	pJL33
pJL55	<i>gut(AD)</i> ₂₄₃₀₉	pJL35
pJL56	<i>gutD</i> ₂₄₃₀₉	pJL36
pJL57	<i>gutAE</i> ₂₄₃₀₉	pJL31+pJL34
pJL58	<i>gutADE</i> ₂₄₃₀₉	pJL34+pJL35
pJL52	<i>gutA</i> ₃₅₈₀₉	pJL32
pJL59	<i>gut(ABCDE)</i> ₃₅₈₀₉	pJL37+ pJL38
pJL60	<i>gut(ABCD)</i> ₃₅₈₀₉	pJL37
pJL61	<i>gut(AD)</i> ₃₅₈₀₉	pJL39
pJL62	<i>gutD</i> ₃₅₈₀₉	pJL40
pJL63	<i>gut(ABC)</i> ₃₅₈₀₉	pJL32+ pJL41
pJL64	<i>gut(AE)</i> ₃₅₈₀₉	pJL32+ pJL38
pJL65	<i>gut(ADE)</i> ₃₅₈₀₉	pJL38+pJL39

Table S5. NMR data of guanitrypmycin A1-1 (**3a**).guanitrypmycin A1-1 (**3a**)

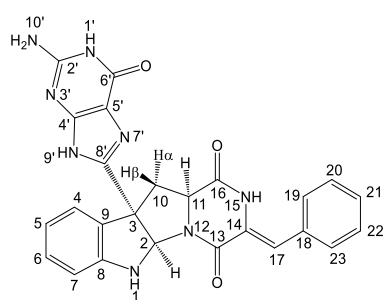
Pos.	δ_C	δ_H , multi., J in Hz	1H - 1H COSY	HMBC	NOESY(key correlations) ^a (s: strong; m: medium; w: weak)
1	-	not observed	-	-	
2	80.8	5.93, s	-	C-3, 8, 10, 11, 8'	H-10 α (w), 10 β (w), 11 (m), 11' (s)
3	55.6	-	-	-	
4	122.6	6.83, d, 7.3	H-5	C-3, 6, 8	H-10 α (w), 10 β (m), 11' (w)
5	118.4	6.64, t, 7.4	H-4, 6	C-7, 9	
6	129.3	7.10, t, 7.7	H-5, 7	C-4, 8	
7	109.6	6.87, d, 7.9	H-6	C-4, 5, 9	
8	148.9	-	-	-	
9	129.7	-	-	-	
10 α	39.9	3.82, dd, 12.3, 6.1	H-11, 10 β	C-2, 3, 8'	H-2 (w), 4 (w), 10 β (s), 11 (s)
10 β		2.01, t, 11.9	H-11, 10 α	C-3, 9, 11, 16, 8'	H-2 (w), 4 (m), 10 α (s), 11 (m)
11	57.6	4.76, dd, 11.4, 6.2	10 α , 10 β	C-10, 16	H-2 (m), 10 α (s), 10 β (m)
13	160.2	-	-	-	
14	128.3	-	-	-	
15	-	10.12, s	-	C-11, 13, 14, 16	
16	167.0	-	-	-	
17	116.1	6.75, s	-	C-13, 19, 23	
18	133.4	-	-	-	
19	129.6	7.56, d, 7.6	H-20	C-17, 21, 23	
20	128.7	7.39, t, 7.6	H-19, 21	C-18, 22	
21	128.3	7.30, t, 7.3	H-20, 22	C-19, 23	
22	128.7	7.39, t, 7.6	H-21, 23	C-18, 20	
23	129.6	7.56, d, 7.6	H-22	C-17, 19, 21	
1'	-	10.68, d, 5.4	-	C-5'	
2'	153.9	-	-	-	
4'	153.8	-	-	-	
5'	114.6	-	-	-	
6'	156.7	-	-	-	
8'	146.1	-	-	-	
10'	-	6.58, s	-	-	
11'	29.3	3.02, s	-	C-4', 8'	H-2 (s), 4 (m)

^a only related correlations for determination of configuration are listed.

Table S6. NMR data of guanitrypmycin A1-2 (**3b**).

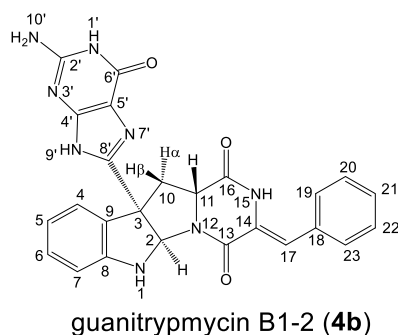
Pos.	δ_C	δ_H , multi., J in Hz	1H - 1H COSY	HMBC	NOESY(key correlations) ^a (s: strong; m: medium; w: weak)
1	-	7.16, s	-	-	
2	78.3	6.01, d, 1.2	-	C-11, 8, 9, 8'	H-11 (w), 11' (s)
3	55.0	-	-	-	
4	123.9	6.98, d, 7.3	H-5	C-6, 8	H-10 β (m), 11 (w), 11' (w)
5	118.6	6.66, t, 7.4	H-4, 6	C-7, 9	
6	129.5	7.09, t, 7.6	H-5, 7	C-4, 8	
7	108.7	6.65, d, 7.8	H-6	C-5, 9	
8	150.8	-	-	-	
9	127.8	-	-	-	
10 α	40.8	3.51, t, 12.4	H-11, 10 β	C-3, 9, 16, 8'	H-10 β (s), 11 (m)
10 β		2.71, dd, 12.9, 6.0	H-11, 10 α	C-2, 9	H-4 (m), 10 α (s), 11 (s)
11	56.3	4.34, dd, 11.9, 6.0	10 α , 10 β	C-16	H-2 (w), 4 (w), 10 α (m), 10 β (s)
13	158.1	-	-	-	
14	127.7	-	-	-	
15	-	10.07, s	-	-	
16	166.3	-	-	-	
17	115.4	6.75, s	-	C-13, 19, 23	
18	133.3	-	-	-	
19	129.4	7.52, d, 7.6	H-20	C-17, 21, 23	
20	128.6	7.39, t, 7.7	H-19, 21	C-18, 22	
21	128.1	7.30, t, 7.4	H-20, 22	C-19, 23	
22	128.6	7.39, t, 7.7	H-21, 23	C-18, 20	
23	129.4	7.52, d, 7.6	H-22	C-17, 19, 21	
1'	-	10.58, s	-	-	
2'	153.6	-	-	-	
4'	153.6	-	-	-	
5'	114.4	-	-	-	
6'	156.7	-	-	-	
8'	147.8	-	-	-	
10'	-	6.48, s	-	-	
11'	28.3	3.13, s	-	C-4', 8'	H-2 (s), 4 (w)

^a only related correlations for determination of configuration are listed.

Table S7. NMR data of guanitrypmycin B1-1 (**4a**).guanitrypmycin B1-1 (**4a**)

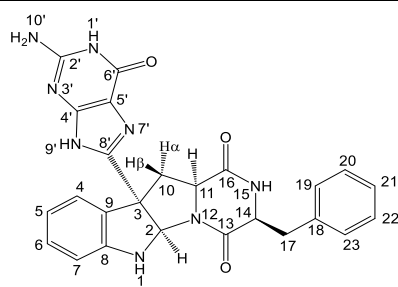
Pos.	δ_C	δ_H , multi., J in Hz	1H - 1H COSY	HMBC	NOESY (key correlations) ^a (s: strong; m: medium; w: weak)
1	-	7.05, s	H-2	C-3, 9	
2	81.3	6.32, s	H-1	-	H-11 (m), 10 α (m), 10 β (w)
3	56.0	-	-	-	
4	122.9	7.03, d, 7.4	H-5	C-3, 6, 8	H-10 α (w), 10 β (m)
5	117.9	6.60, t, 7.1	H-4, 6,	C-7, 9	
6	128.7	7.04, t, 7.2	H-5, 7	C-4, 8	
7	109.7	6.61, d, 7.6	H-6	C-5, 9	
8	148.4	-	-	-	
9	130.3	-	-	-	
10 α	39.2	3.37, dd, 13.2, 6.5	H-11, 10 β	C-2, 3, 8'	H-2 (m), 4 (w), 10 β (s), 11 (s)
10 β		2.10, t, 12.2	H-11, 10 α	C-3, 9, 16, 8'	H-2 (w), 4 (m), 10 α (s), 11 (m)
11	57.7	4.60, dd, 11.7, 6.8	10 α , 10 β	C-10, 16	H-2 (m), 10 α (s), 10 β (m)
13	160.4	-	-	-	
14	128.4	-	-	-	
15	-	10.15, s	-	-	
16	167.0	-	-	-	
17	116.3	6.77, s	-	C-13, 19, 23	
18	133.5	-	-	-	
19	129.6	7.55, d, 7.6	H-20	C-17, 21, 23	
20	128.7	7.39, t, 7.5	H-19, 21	C-18, 22	
21	128.4	7.30, t, 7.4	H-20, 22	C-19, 23	
22	128.7	7.39, t, 7.5	H-21, 23	C-18, 20	
23	129.6	7.55, d, 7.6	H-22	C-17, 19, 21	
1'	-	10.66, s	-	-	
2'	154.0	-	-	-	
4'	153.7	-	-	-	
5'	115.8	-	-	-	
6'	156.8	-	-	-	
8'	146.9	-	-	-	
9'	-	12.60, s	-	-	
10'	-	6.43, s	-	-	

^a only related correlations for determination of configuration are listed.

Table S8. NMR data of guanitrypmycin B1-2 (**4b**).

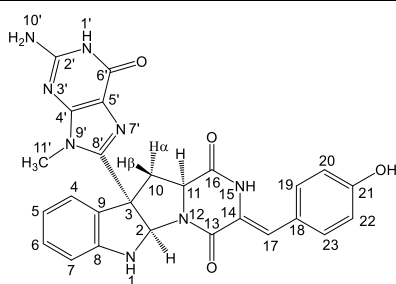
Pos.	δ_C	δ_H , multi., J in Hz	1H - 1H COSY	HMBC	NOESY (key correlations) ^a (s: strong; m: medium; w: weak)
1	-	6.87, s	-	C-3, 9	
2	79.1	6.24, s	-	-	H-10 α (w)
3	55.3	-	-	-	
4	124.3	7.34, d, 7.3	H-5	C-3, 6, 8	H-10 α (w), 10 β (m), 11 (w)
5	117.9	6.61, t, 7.2	H-4, 6,	C-7, 9	
6	129.0	6.98, t, 7.4	H-5, 7	C-4, 8	
7	108.8	6.54, d, 7.7	H-6	C-5, 9	
8	150.1	-	-	-	
9	128.3	-	-	-	
10 α	40.6	2.66, t, 12.0	H-11, 10 β	C-3, 9, 11, 16, 8'	H-2 (w), 4 (w), 10 β (s), 11 (m)
10 β		2.91, dd, 11.9, 5.9	H-11, 10 α	-	H-4 (m), 10 α (s), 11 (s)
11	57.9	4.30, dd, 11.1, 5.9	10 α , 10 β	C-10, 16	H-4 (w), 10 α (m), 10 β (s)
13	158.1	-	-	-	
14	128.0	-	-	-	
15	-	10.04, s	-	-	
16	166.1	-	-	-	
17	115.1	6.70, s	-	C-13, 16, 19, 23	
18	133.5	-	-	-	
19	129.4	7.47, d, 7.4	H-20	C-17, 21, 23	
20	128.7	7.34, t, 7.3	H-19, 21	C-18, 22	
21	128.1	7.25, t, 7.0	H-20, 22	C-19, 23	
22	128.7	7.34, t, 7.3	H-21, 23	C-18, 20	
23	129.4	7.47, d, 7.4	H-22	C-17, 19, 21	
1'	-	10.57, s	-	-	
2'	153.7	-	-	-	
4'	153.6	-	-	-	
5'	115.3	-	-	-	
6'	156.7	-	-	-	
8'	148.0	-	-	-	
9'	-	12.46, s	-	-	
10'	-	6.35, s	-	-	

^a only related correlations for determination of configuration are listed.

Table S9. NMR data of guanitrypmycin C1-1 (**5a**).guanitrypmycin C1-1 (**5a**)

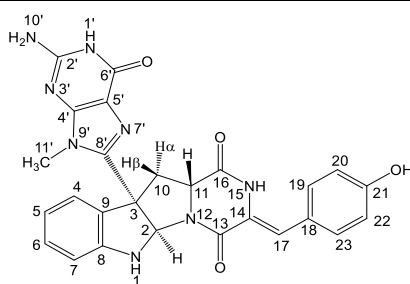
Pos.	δ_C	δ_H , multi., J in Hz	1H - 1H COSY	HMBC	NOESY (key correlations) ^a (s: strong; m: medium; w: weak)
1	-	6.98, d, 3.7	H-2	C-3, 8, 9	
2	81.2	6.20, d, 3.6	H-1	C-8, 8'	H-11 (m), 10 α (w), 10 β (w)
3	56.2	-	-	-	
4	122.6	6.92, d, 7.4	H-5	C-6, 8	H-10 α (w), 10 β (m), 11 (w)
5	117.6	6.57, t, 7.5	H-4, 6,	C-7, 9	
6	128.7	7.02, t, 7.7	H-5, 7	C-4, 8	
7	109.4	6.60, d, 7.9	H-6	C-5, 9	
8	148.4	-	-	-	
9	130.3	-	-	-	
10 α	38.8	3.27, dd, 12.7, 6.5	H-11, 10 β	C-2, 3, 8'	H-2 (w), 10 β (s), 11 (m)
10 β		1.93, t, 12.0	H-11, 10 α	C-3, 9, 11, 16, 8'	H-2 (w), 4 (m), 10 α (s), 11 (m)
11	58.2	4.40, dd, 11.0, 6.5	10 α , 10 β	C-10	H-2 (m), 10 α (m), 10 β (m)
13	167.2	-	-	-	
14	55.6	4.44, t, 5.2	17 α , 17 β	C-13, 17, 18	
15	-	8.00, s	-	C-11, 13, 14, 17	
16	169.6	-	-	-	
17 α	34.3	3.16, dd, 14.1, 5.7	H-14, 17 β	C-13, 18, 19, 23	
17 β		3.03, dd, 14.5, 5.4	H-14, 17 α	C-13, 18, 19, 23	
18	137.5	-	-	-	
19	130.3	7.34, d, 7.4	H-20	C-17, 21, 23	
20	128.0	7.21, t, 7.5	H-19, 21	C-18, 22	
21	126.3	7.16, t, 7.3	H-20, 22	C-19, 23	
22	128.0	7.21, t, 7.5	H-21, 23	C-18, 20	
23	130.3	7.34, d, 7.4	H-22	C-17, 19, 21	
1'	-	10.51, s	-	-	
2'	153.8	-	-	-	
4'	153.6	-	-	-	
5'	115.7	-	-	-	
6'	156.6	-	-	-	
8'	146.7	-	-	-	
9'	-	12.51, s	-	-	
10'	-	6.36, s	-	-	

^a only related correlations for determination of configuration are listed.

Table S10. NMR data of guanitrypmycin A2-1 (**6a**).guanitrypmycin A2-1 (**6a**)

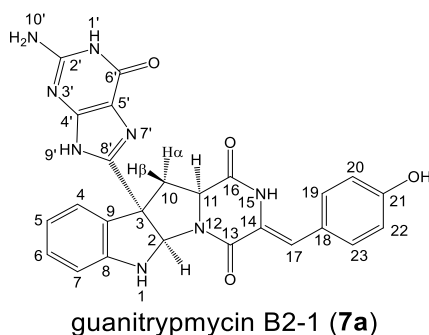
Pos.	δ_C	δ_H , multi., J in Hz	1H - 1H COSY	HMBC	NOESY (key correlations) ^a (s: strong; m: medium; w: weak)
1	-	7.24, d, 3.6	H-2	C-2, 3, 8, 9	
2	80.6	5.90, d, 3.6	H-1	C-3, 8, 10, 11, 8'	H-11 (m), 11'(s)
3	55.6	-	-	-	
4	122.5	6.83, d, 7.3	H-5	C-3, 6, 8	H-10 α (m), 10 β (m), 11'(w)
5	118.3	6.62, td, 7.5, 0.9	H-4, 6,	C-7, 9	
6	129.2	7.10, td, 7.8, 1.2	H-5, 7	C-4, 8	
7	109.5	6.66, d, 8.0	H-6	C-5, 9	
8	148.9	-	-	-	
9	129.7	-	-	-	
10 α	39.8	3.79, dd, 12.3, 6.2	H-11, 10 β	C-2, 3, 8'	H-4 (m), 10 β (s), 11 (s)
10 β	39.8	1.99, t, 11.9	H-11, 10 α	C-3, 9, 11, 16, 8'	H-4 (s), 10 α (s), 11 (m)
11	57.5	4.68, dd, 11.4, 6.2	10 α , 10 β	C-10, 16	H-2 (m), 10 α (s), 10 β (m)
13	160.8	-	-	-	
14	125.7	-	-	-	
15	-	9.99, s	-	-	
16	166.8	-	-	-	
17	117.0	6.68, s	-	C-13, 19, 23	
18	124.2	-	-	-	
19	131.4	7.43, d, 8.7	H-20	C-17, 21, 23	
20	115.6	6.79, d, 8.7	H-19	C-18, 22	
21	157.9	not observed	-	-	
22	115.6	6.79, d, 8.7	H-23	C-18, 20	
23	131.4	7.43, d, 8.7	H-22	C-17, 19, 21	
1'	-	not observed	-	-	
2'	153.8	-	-	-	
4'	153.7	-	-	-	
5'	114.6	-	-	-	
6'	156.7	-	-	-	
8'	146.1	-	-	-	
10'	-	6.56, s	-	-	
11'	29.2	3.01, s	-	C-4', 8'	H-2 (s), 4 (w)

^a only related correlations for determination of configuration are listed.

Table S11. NMR data of guanitrypmycin A2-2 (**6b**).guanitrypmycin A2-2 (**6b**)

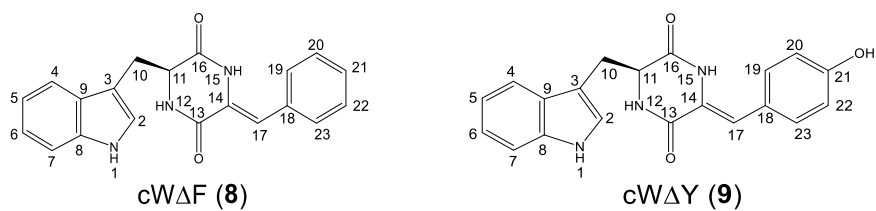
Pos.	δ_C	δ_H , multi., J in Hz	1H - 1H COSY	HMBC	NOESY (key correlations) ^a (s: strong; m: medium; w: weak)
1	-	7.14, s	H-2	C-2, 3	
2	78.3	5.99, d, 1.5	H-1	C-3, 8, 9, 11, 8'	H-11 (w), 11'(s)
3	55.0	-	-	-	
4	123.9	6.97, d, 7.6	H-5	C-3, 6, 8	H-10 α (w), 10 β (s), 11'(w)
5	108.7	6.64, t, 7.2	H-4, 6,	C-7, 9	
6	129.5	7.08, t, 7.7	H-5, 7	C-4, 8	
7	119.0	6.65, d, 7.6	H-6	C-5, 9	
8	150.8	-	-	-	
9	127.8	-	-	-	
10 α	40.9	3.50, t, 12.5	H-11, 10 β	C-3, 9, 11, 16, 8'	H-4 (w), 10 β (s), 11 (m)
10 β	40.9	2.69, dd, 12.8, 6.1	H-11, 10 α	C-2, 9	H-4 (s), 10 α (s), 11 (s)
11	56.2	4.29, dd, 11.9, 6.0	10 α , 10 β	C-16	H-2 (w), 10 α (m), 10 β (s)
13	158.7	-	-	-	
14	125.3	-	-	-	
15	-	9.92, s	-	-	
16	166.2	-	-	-	
17	114.5	6.67, s	-	C-13, 19, 23	
18	124.1	-	-	-	
19	131.2	7.39, d, 8.7	H-20	C-17, 21, 23	
20	115.6	6.78, d, 8.7	H-19	C-18, 22	
21	157.8	-	-	-	
22	115.6	6.78, d, 8.7	H-23	C-18, 20	
23	131.2	7.39, d, 8.7	H-22	C-17, 19, 21	
1'	-	not observed	-	-	
2'	153.7	-	-	-	
4'	153.6	-	-	-	
5'	114.4.	-	-	-	
6'	156.8	-	-	-	
8'	147.8	-	-	-	
10'	-	6.50, s	-	-	
11'	28.2	3.12, s		C-4', 8'	H-2 (s), 4 (w)

^a only related correlations for determination of configuration are listed.

Table S12. NMR data of guanitrypmycin B2-1 (**7a**).

Pos.	δ_C	δ_H , multi., J in Hz	HMBC	NOESY (key correlations) ^a (s: strong; m: medium; w: weak)
1	-	6.98, s	C-3, 9,	
2	81.1	6.27, s	-	H-11 (m), 10 α (w)
3	55.9	-	-	
4	122.9	7.03, d, 7.4	C-3, 6	H-10 α (w), 10 β (m)
5	117.9	6.60, t, 7.2	C-7, 9	
6	128.9	7.04, t, 7.2	C-4, 8	
7	109.6	6.60, d, 7.8	C-5, 9	
8	148.4	-	-	
9	130.3	-	-	
10 α	39.2	3.33, dd, 13.2, 6.2	-	H-2 (w), 4 (w), 10 β (s), 11 (s)
10 β		2.1, t, 12.4	C-8'	H-2 (w), 4 (m), 10 α (s), 11 (m)
11	57.5	4.53, dd, 11.4, 6.4	C-10, 16	H-2 (m), 10 α (s), 10 β (m)
13	160.9	-	-	
14	125.7	-	-	
15	-	9.96, s	-	
16	166.9	-	-	
17	117.2	6.70, s	C-13, 19, 23	
18	124.2	-	-	
19	131.4	7.41, d, 8.6	C-17, 21, 23	
20	115.6	6.78, d, 8.7	C-18, 22, 21	
21	157.9	not observed	-	
22	115.6	6.78, d, 8.7	C-18, 20, 21	
23	131.4	7.41, d, 8.6	C-17, 19, 21	
1'	-	10.54, s	-	
2'	154.0	-	-	
4'	153.5	-	-	
5'	115.8	-	-	
6'	156.7	-	-	
8'	147.0	-	-	
9'	-	12.56, s	-	
10'	-	6.35, s	-	

^a only related correlations for determination of configuration are listed.

Table S13. ^1H NMR data of cW Δ F (**8**) and cW Δ Y (**9**).

Pos.	δ_{H} , multi., J in Hz	δ_{H} , multi., J in Hz
1	10.83, s	10.82, d, 1.3
2	7.01, s	7.01, d, 1.5
4	7.54, dd, 7.0, 1.7	7.54, dd, 7.2, 1.6
5	7.00, td, 6.8, 1.5	6.99, td, 7.1, 1.5
6	7.04, td, 7.0, 1.6	7.03, td, 7.0, 1.6
7	7.16, dd, 6.9, 1.7	7.18, dd, 7.1, 1.7
10	3.38, dd, 14.4, 3.6 3.03, dd, 14.4, 4.6	3.34, dd, 14.4, 4.0 3.03, dd, 14.5, 4.7
11	4.27, dd, 7.0, 3.8	4.23, dd, 7.2, 4.2
12	8.34, d, 2.0	8.22, d, 2.0
15	9.34, s	9.18, s
17	6.20, s	6.17, s
19	6.58, dd, 7.1, 2.1	6.54, m
20	7.12, m	6.54, m
21	7.12, m	not observed
22	7.12, m	6.54, m
23	6.58, dd, 7.1, 2.1	6.54, m

Table S14. Products yields (mg/L culture) of the four transformants cultivated in GMY media at 28°C for 7 days.

M1146 harboring \ main products	3a	3b	4a	4b	total yields
<i>gut(ABCDE)</i> ₂₄₃₀₉	11.7	16.8	1.2	7.3	37.0
<i>gut(ABCD)</i> ₂₄₃₀₉	-	-	17.9	40.3	58.2
<i>gut(ABCDE)</i> ₃₅₈₉	10.7	22.3	3.5	4.9	41.4
<i>gut(ABCD)</i> ₃₅₈₉	-	-	23.2	27.0	50.2

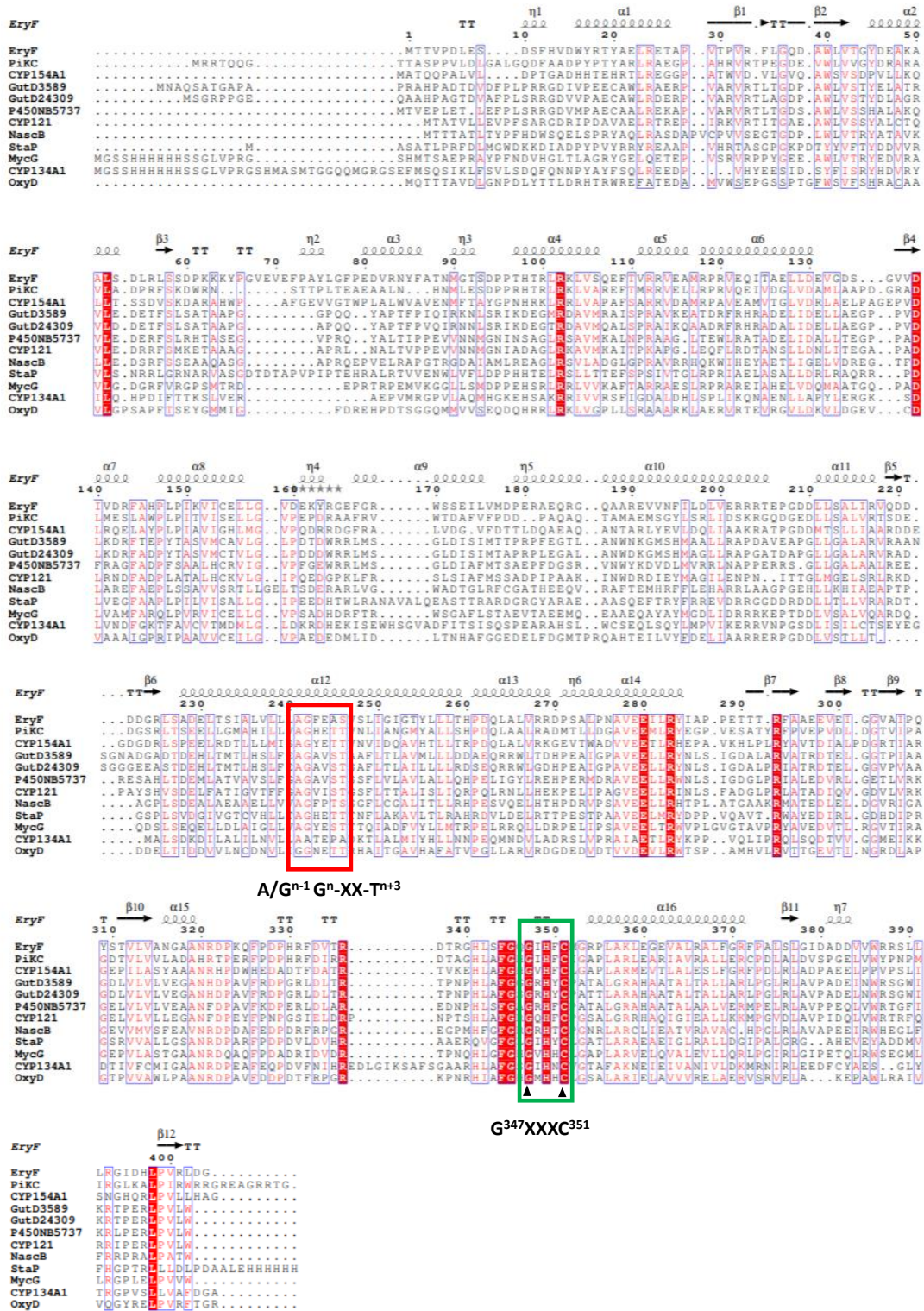


Figure S1. Sequence alignments of P450s from this study and structurally solved natural product P450s from bacteria.

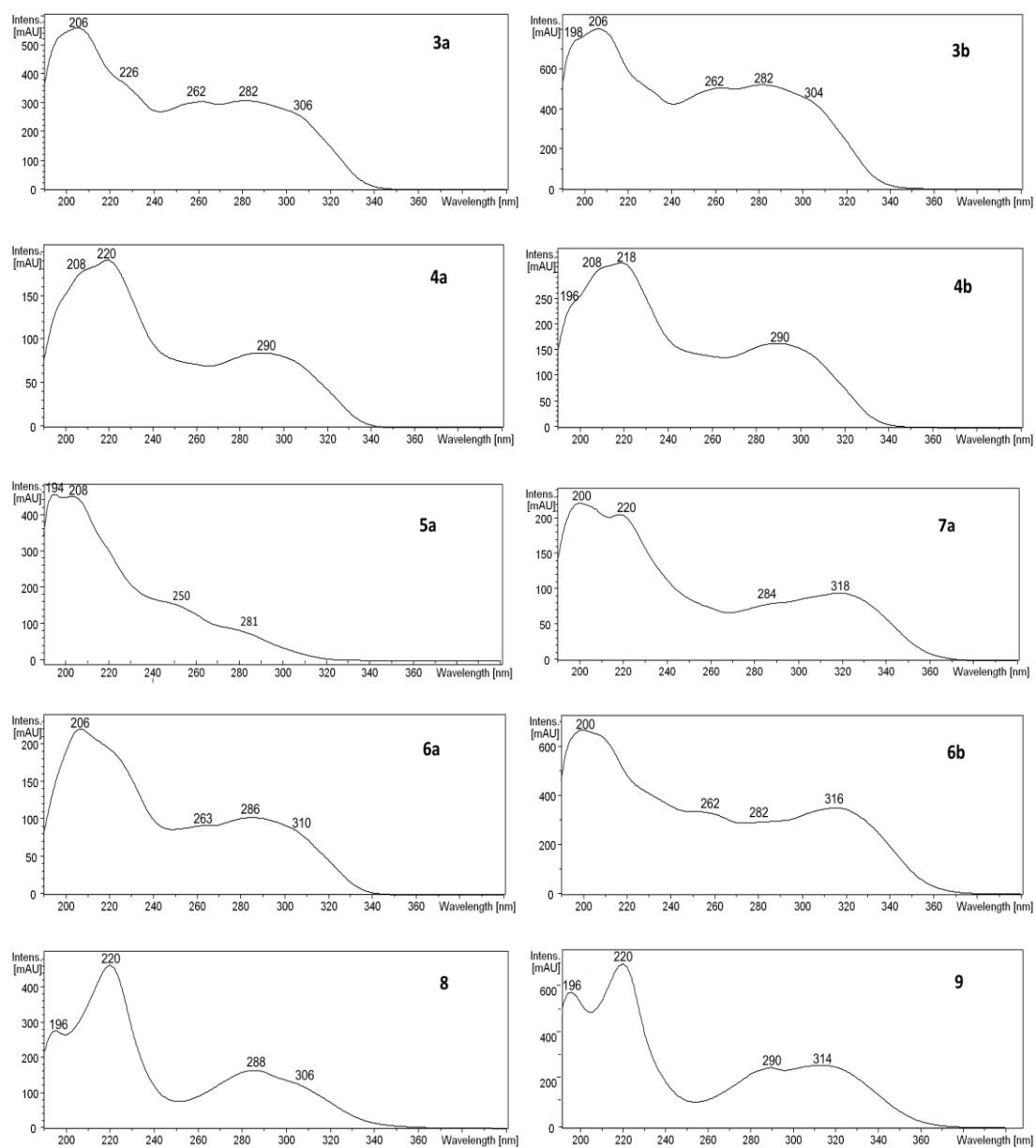


Figure S2. UV spectra of the isolated products obtained from LC-MS analysis.

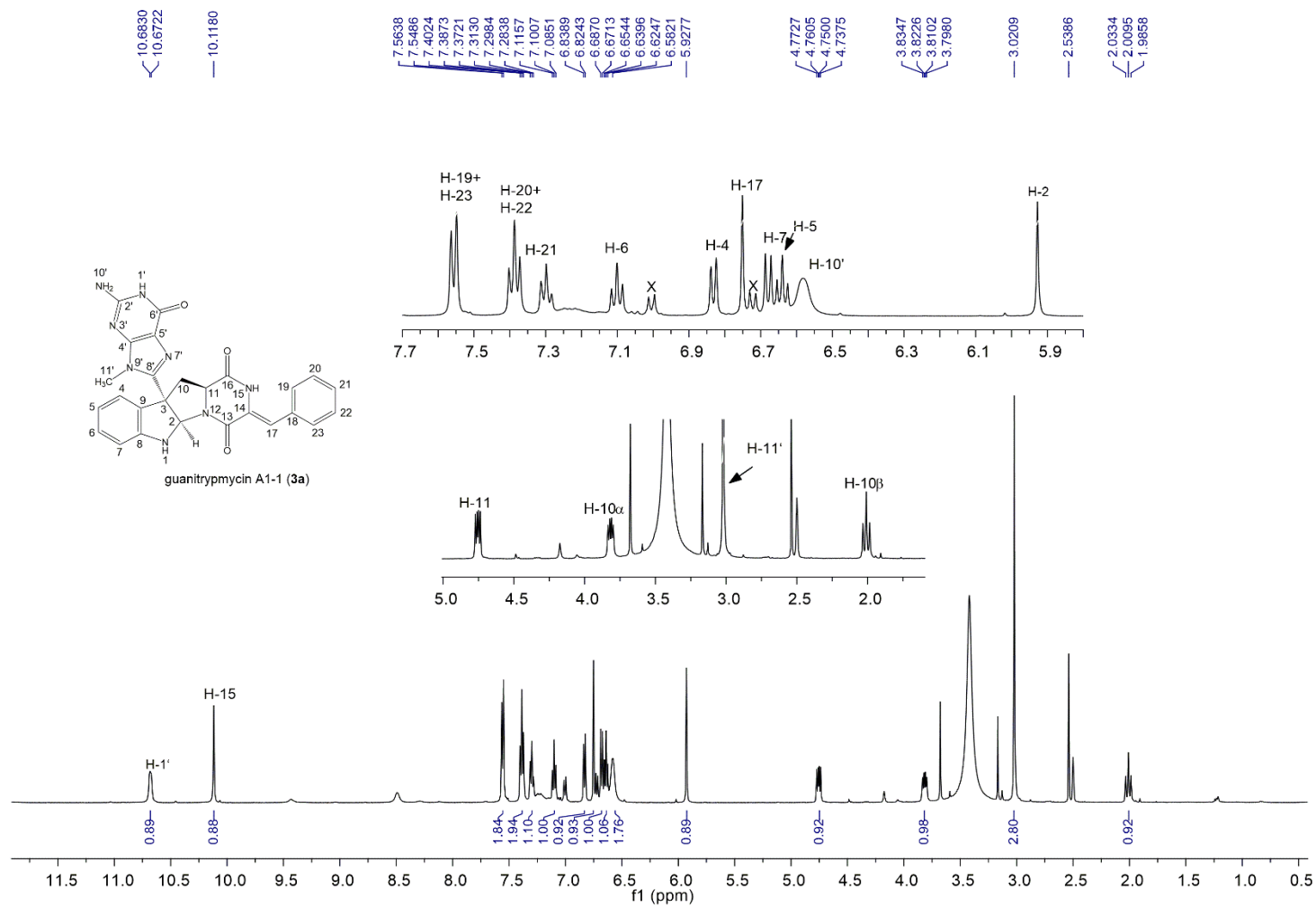
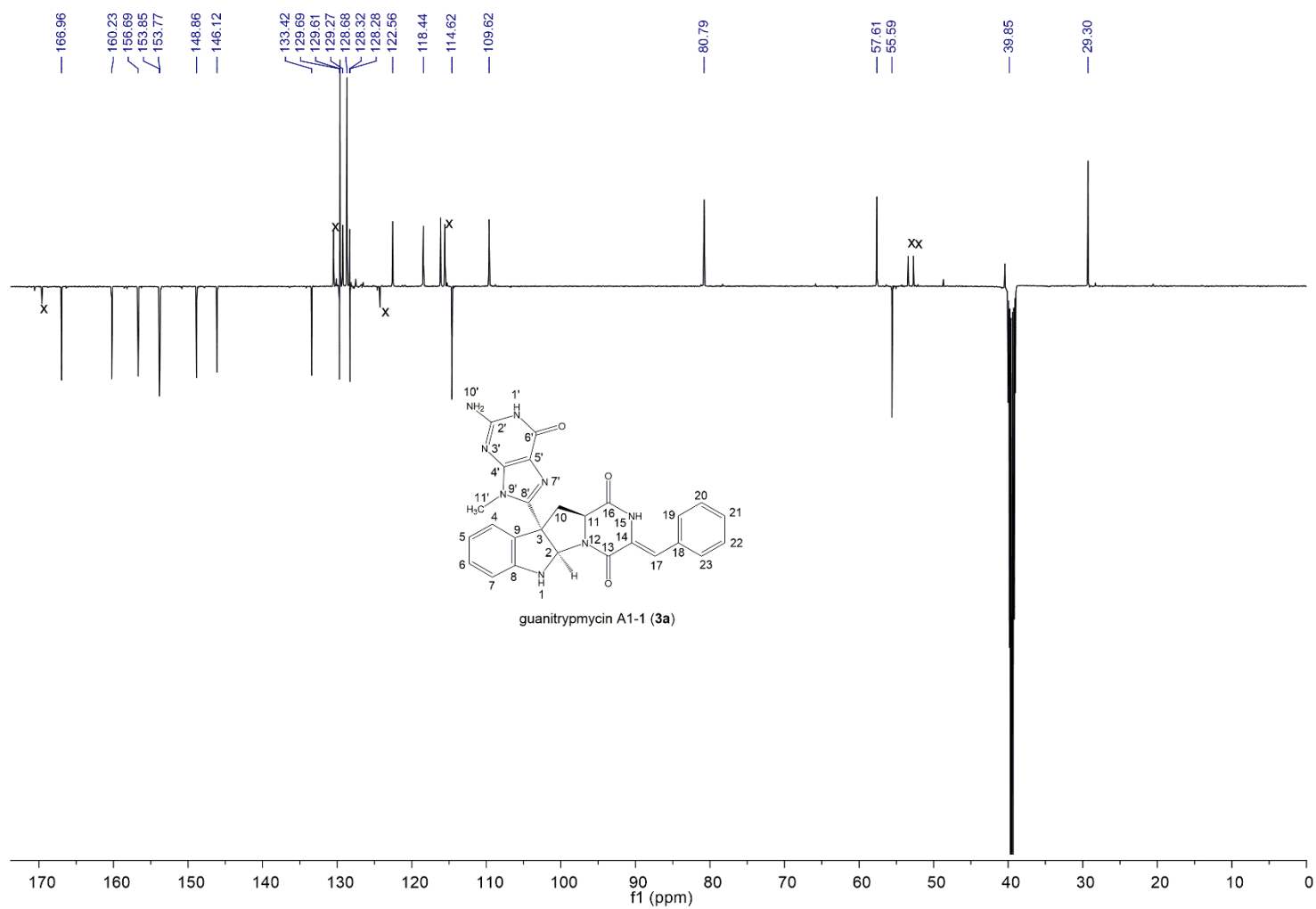


Figure S3. ¹H NMR spectrum of guanitrypmycin A1-1 (3a).



S28

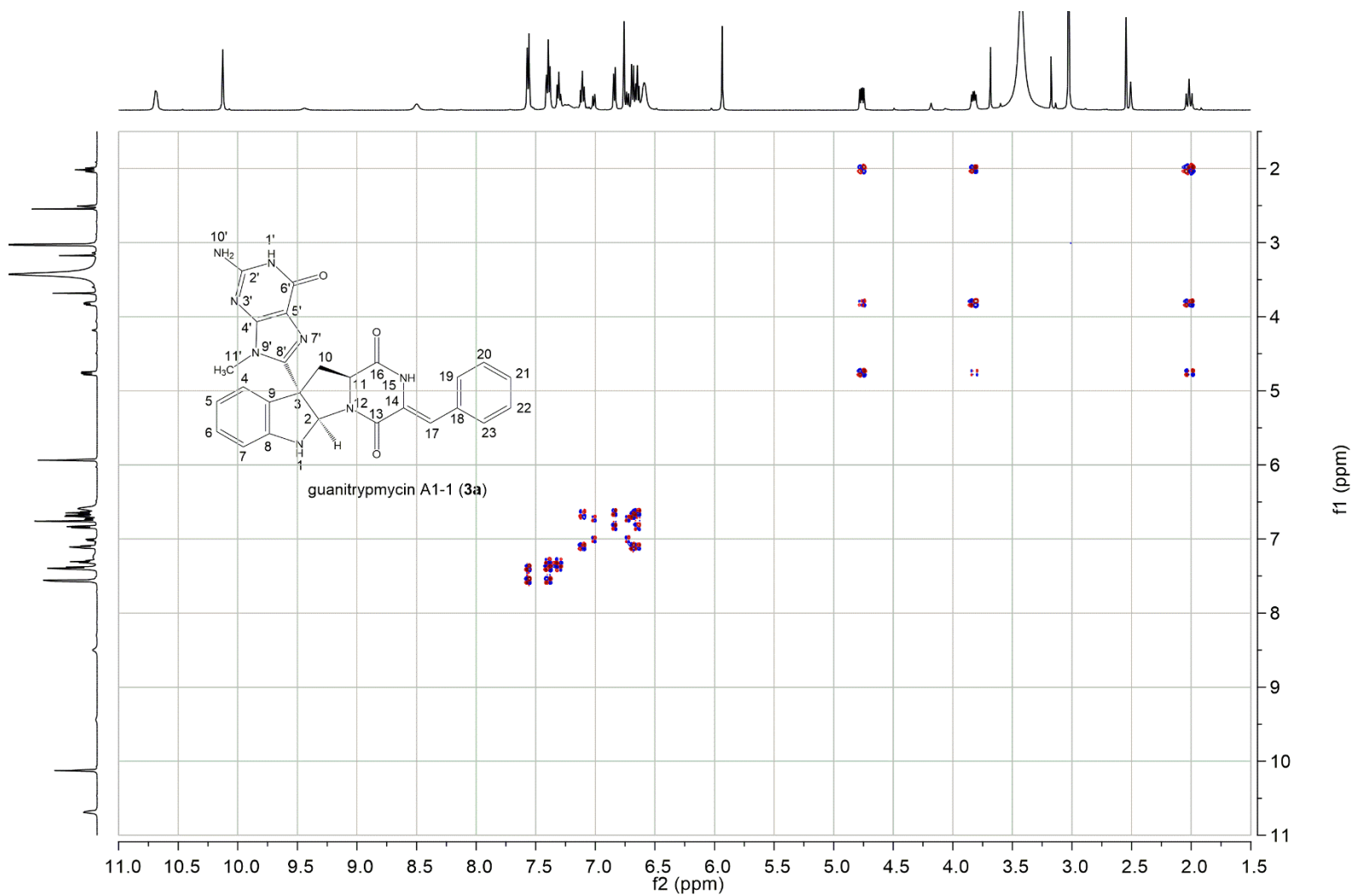


Figure S5. ^1H - ^1H COSY spectrum of guanitrypmycin A1-1 (**3a**).

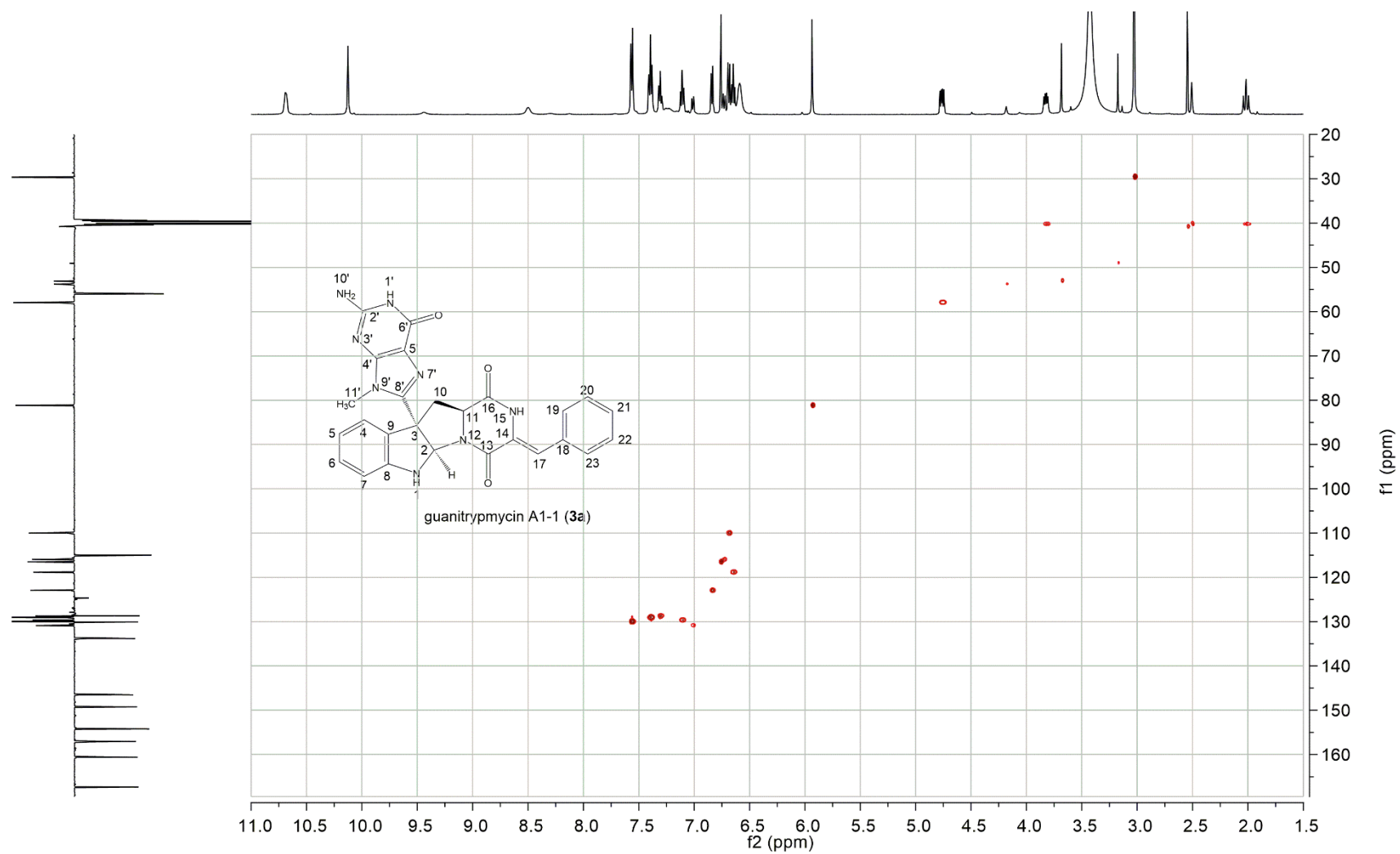


Figure S6. HSQC spectrum of guanitrypmycin A1-1 (**3a**).

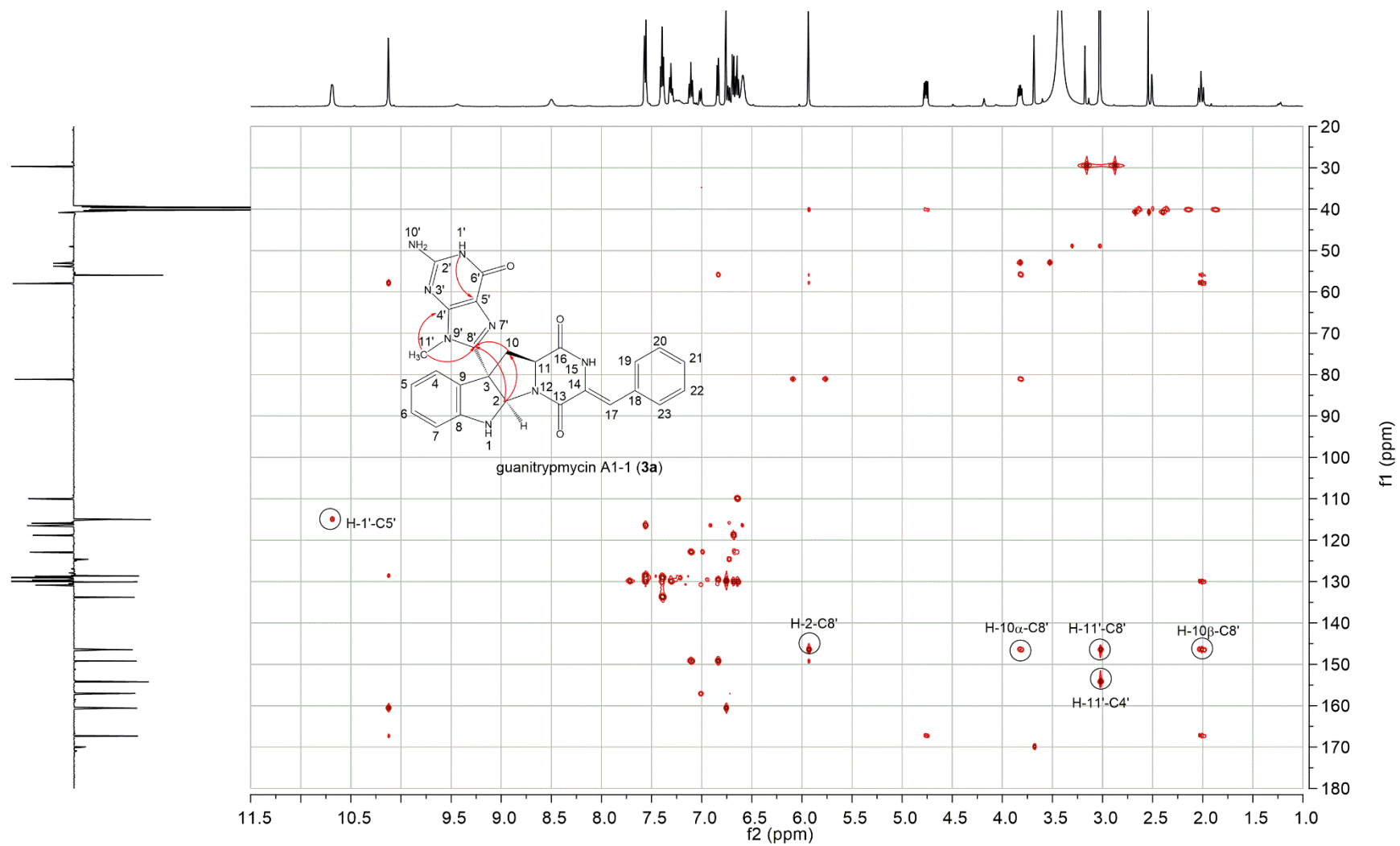


Figure S7. HMBC spectrum of guanitrypmycin A1-1 (**3a**).

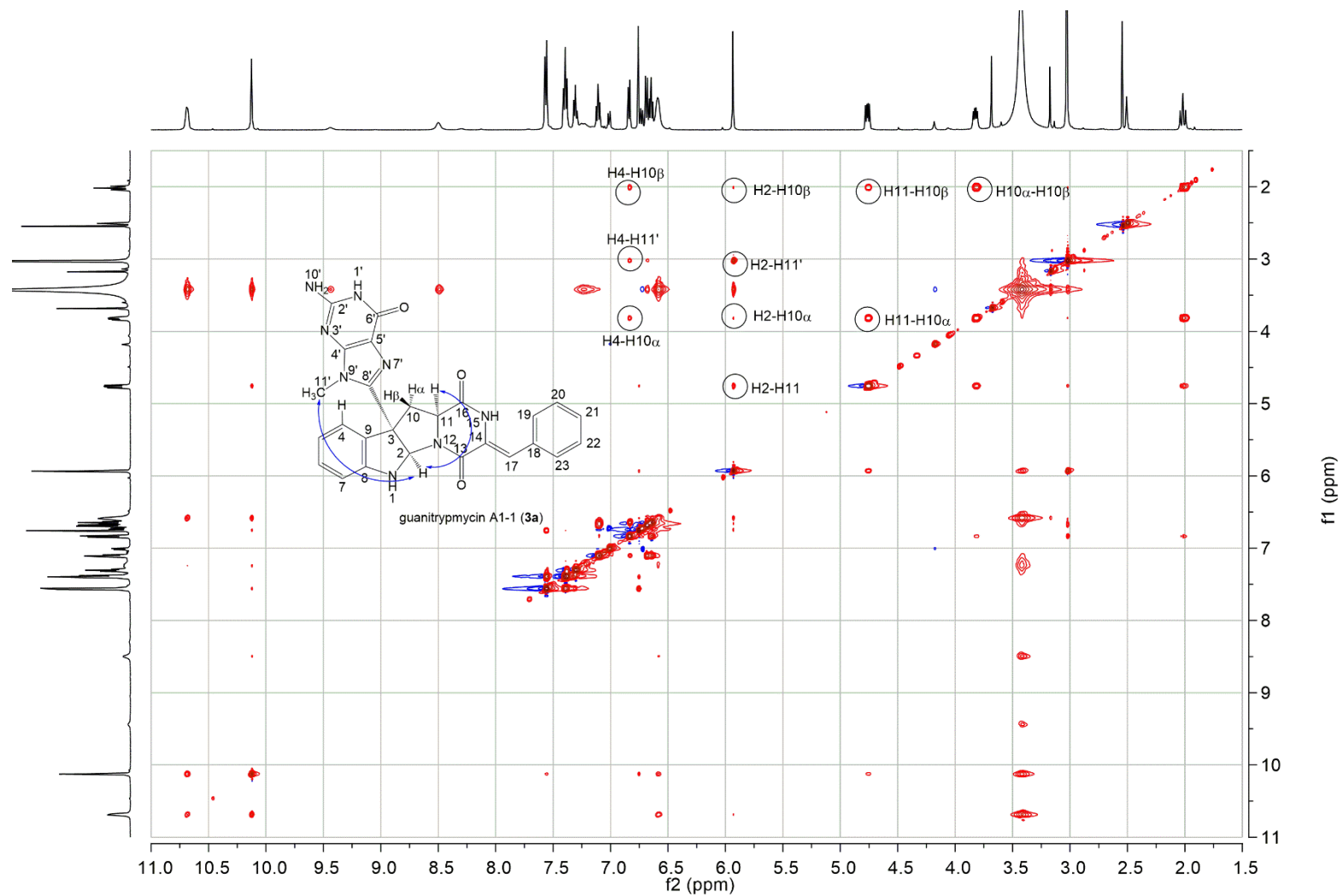


Figure S8. NOESY spectrum of guanitrypmycin A1-1 (**3a**).

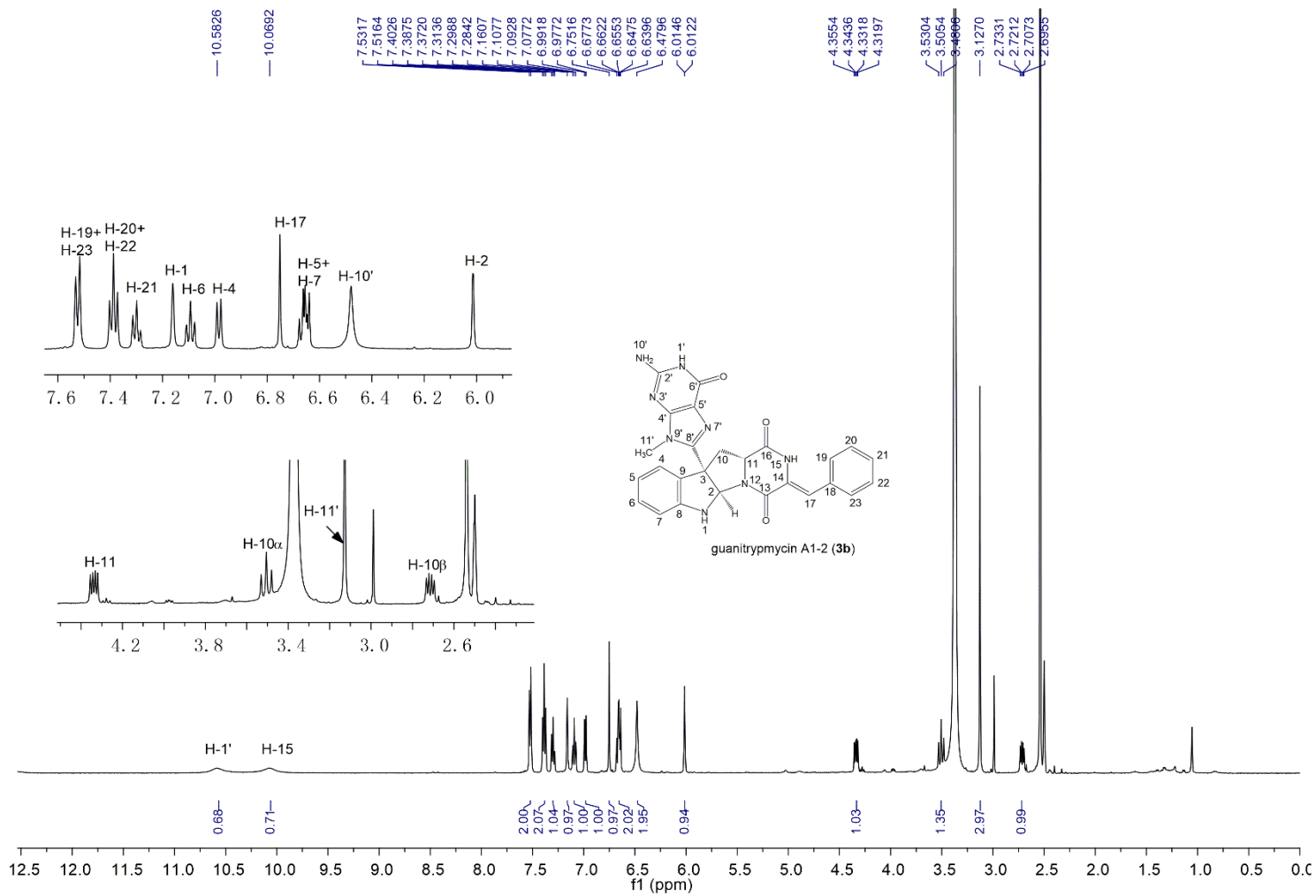


Figure S9. ¹H NMR spectrum of guanitrypmycin A1-2 (3b).

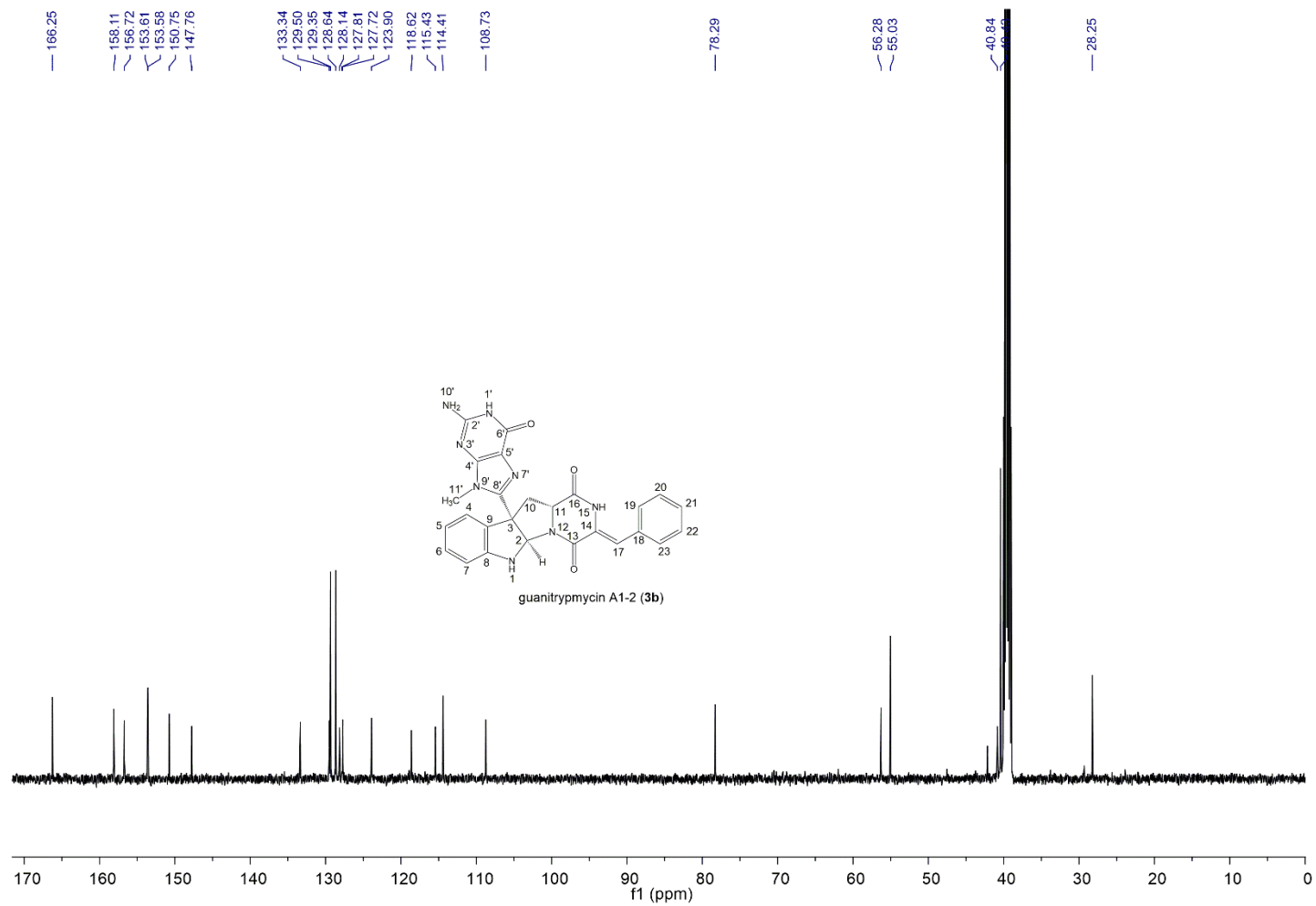


Figure S10. ^{13}C NMR spectrum of guanitrypmycin A1-2 (**3b**).

S34

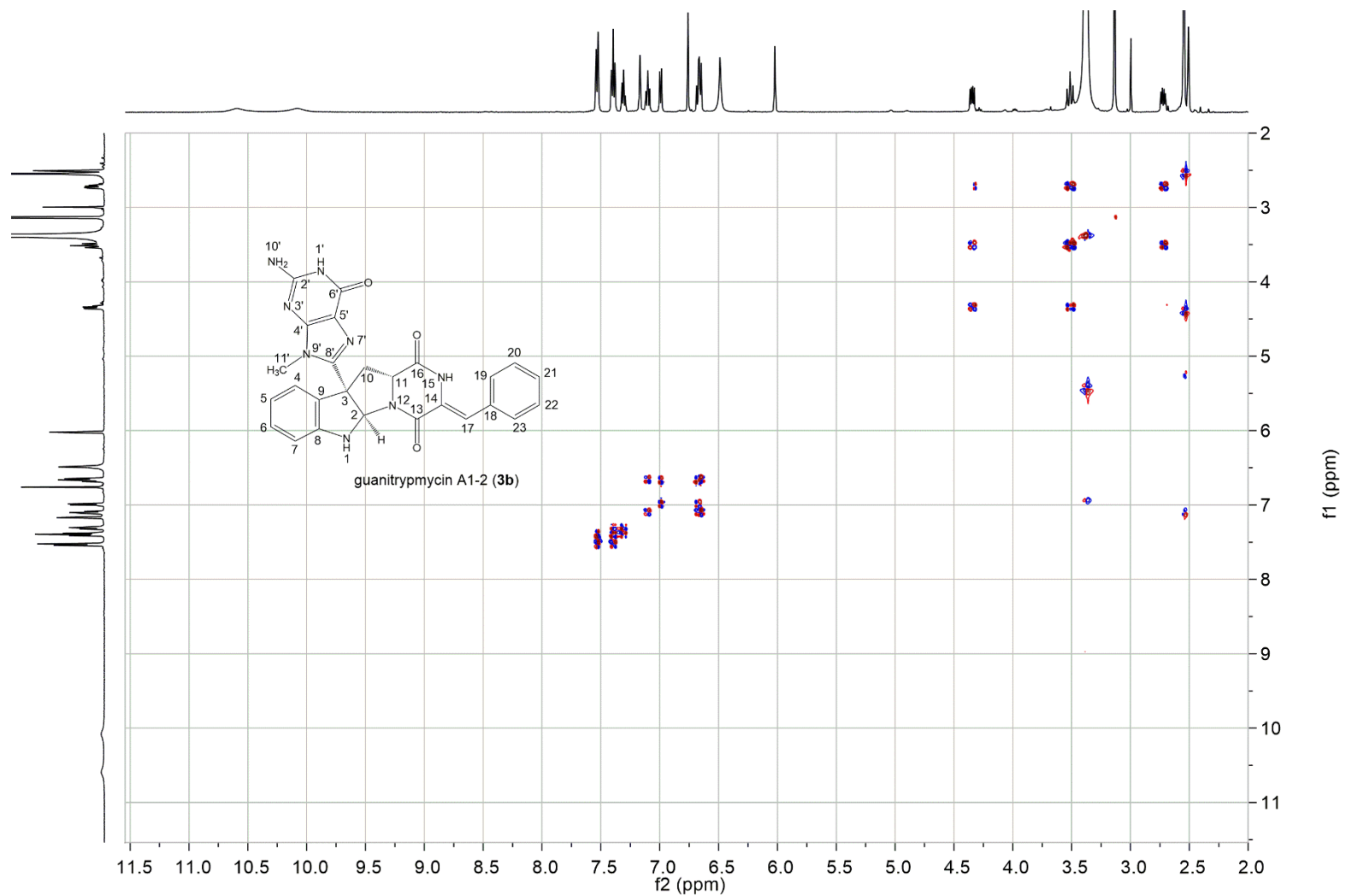


Figure S11. ^1H - ^1H COSY spectrum of guanitrypmycin A1-2 (**3b**).

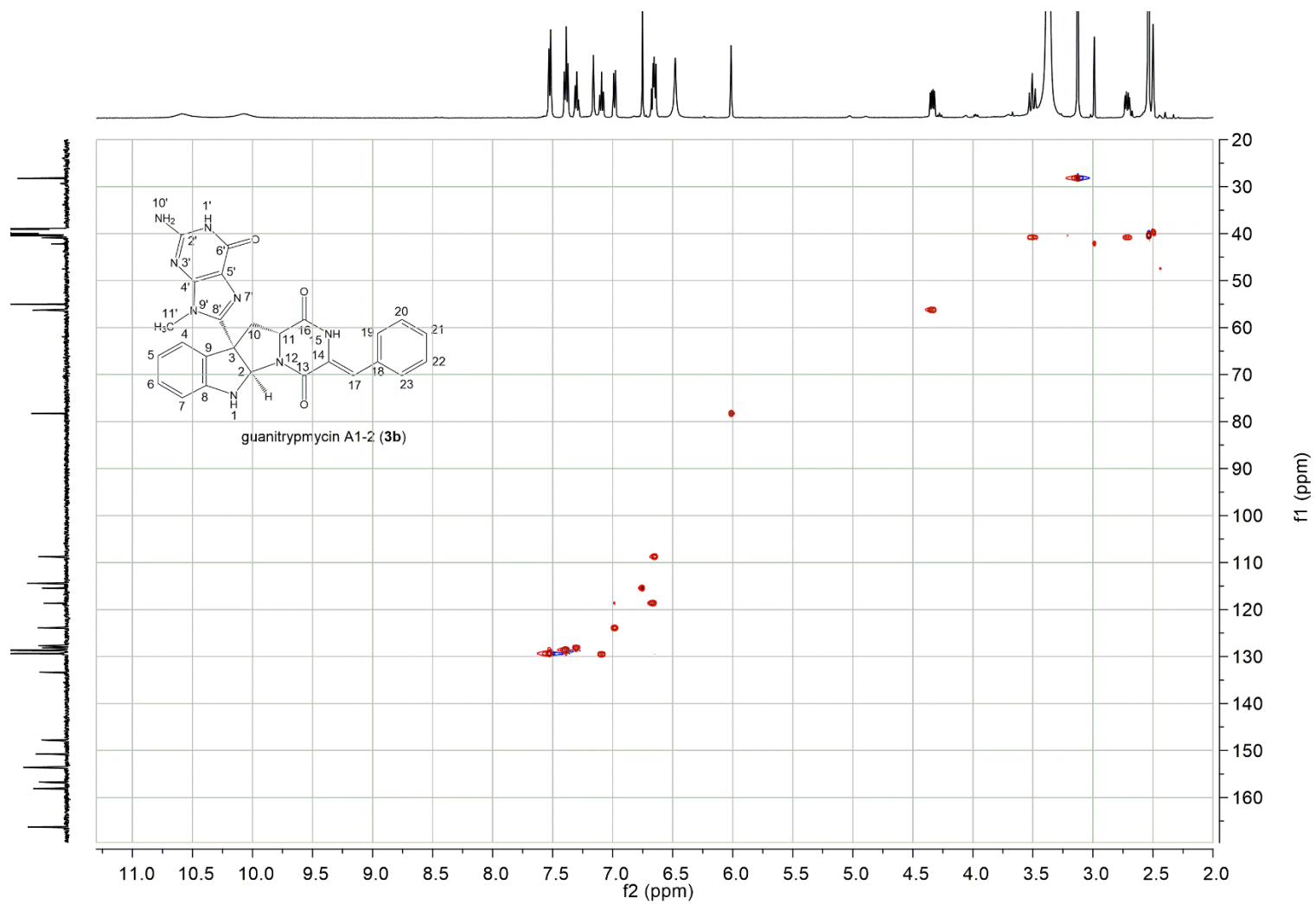


Figure S12. HSQC spectrum of guanitrypmycin A1-2 (**3b**).

S36

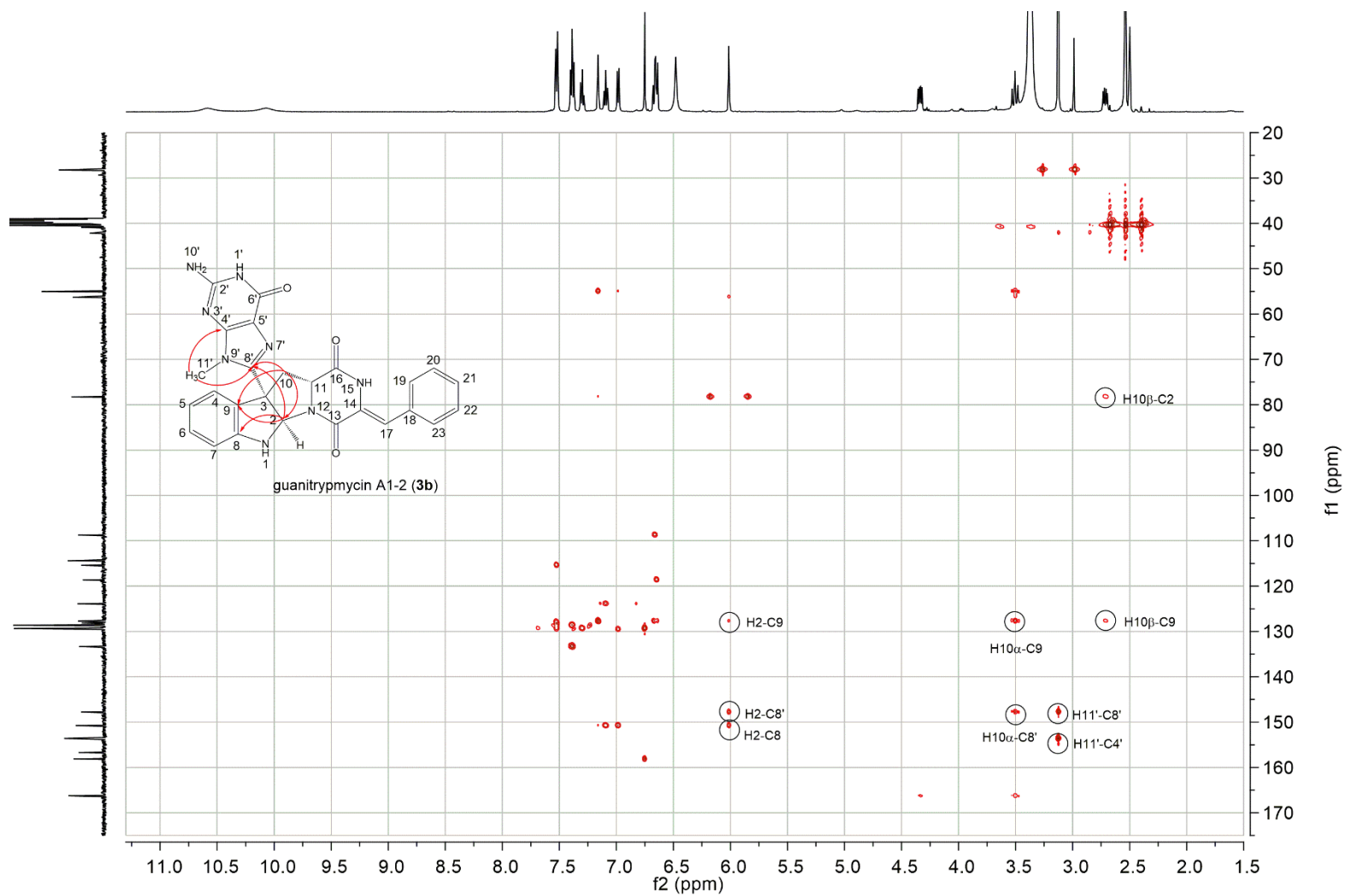


Figure S13. HMBC spectrum of guanitrypmycin A1-2 (**3b**).

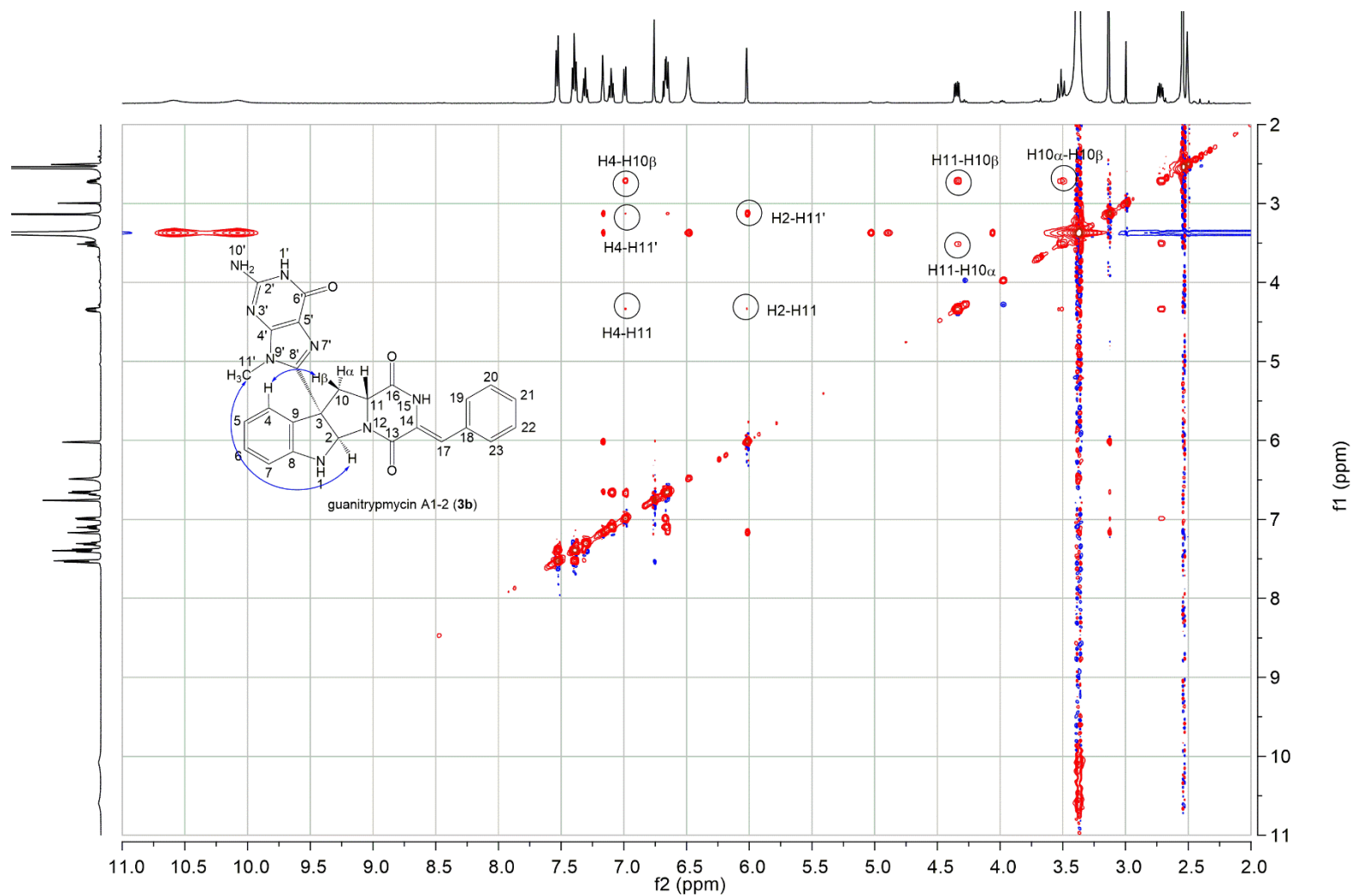


Figure S14. NOESY spectrum of guanitrypmycin A1-2 (**3b**).

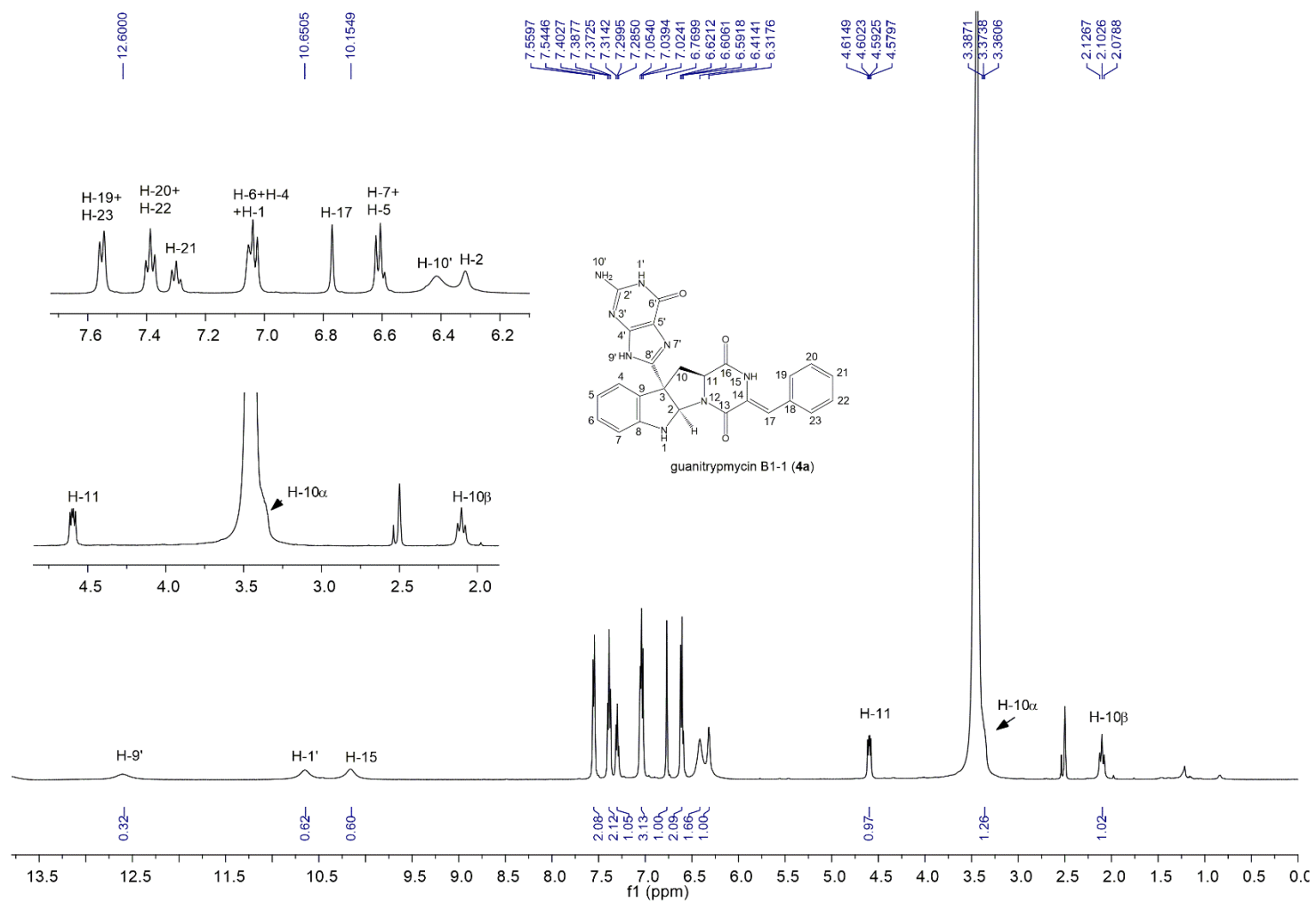


Figure S15. ¹H NMR spectrum of guanitrypmycin B1-1 (4a).

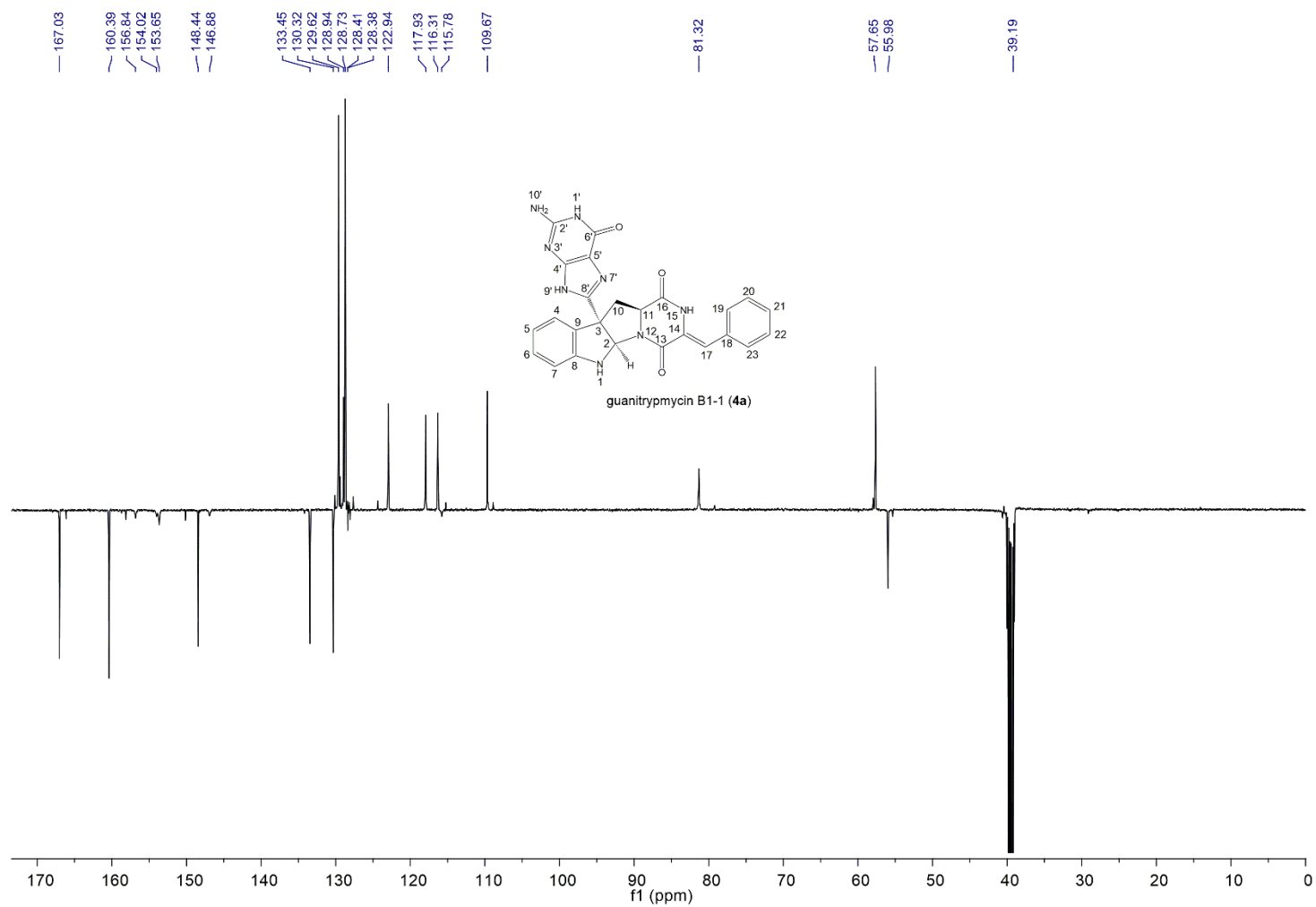


Figure S16. ^{13}C APT NMR spectrum of guanitrypmycin B1-1 (**4a**).

S40

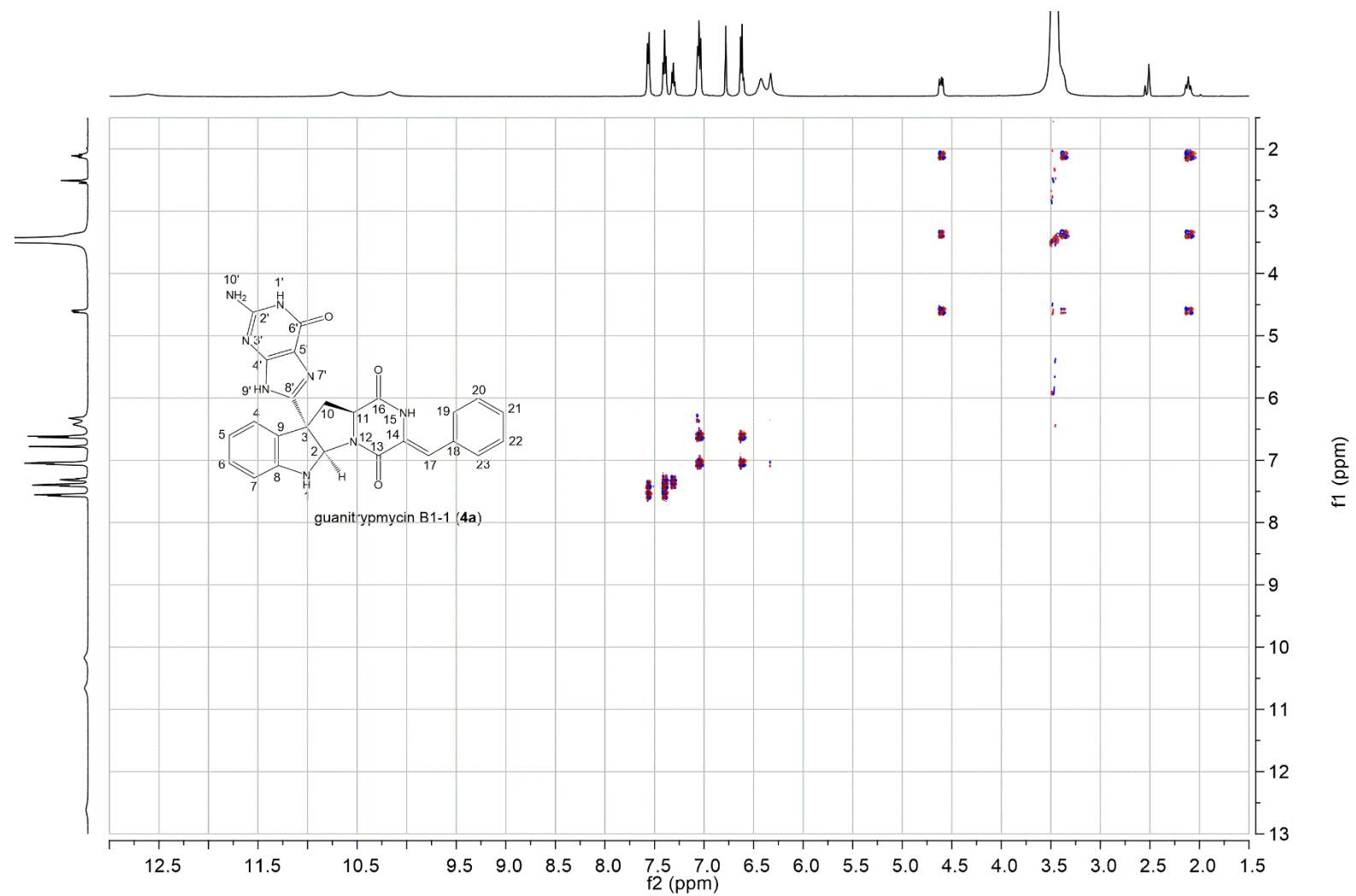


Figure S17. ^1H - ^1H COSY spectrum of guanitrypmycin B1-1 (**4a**).

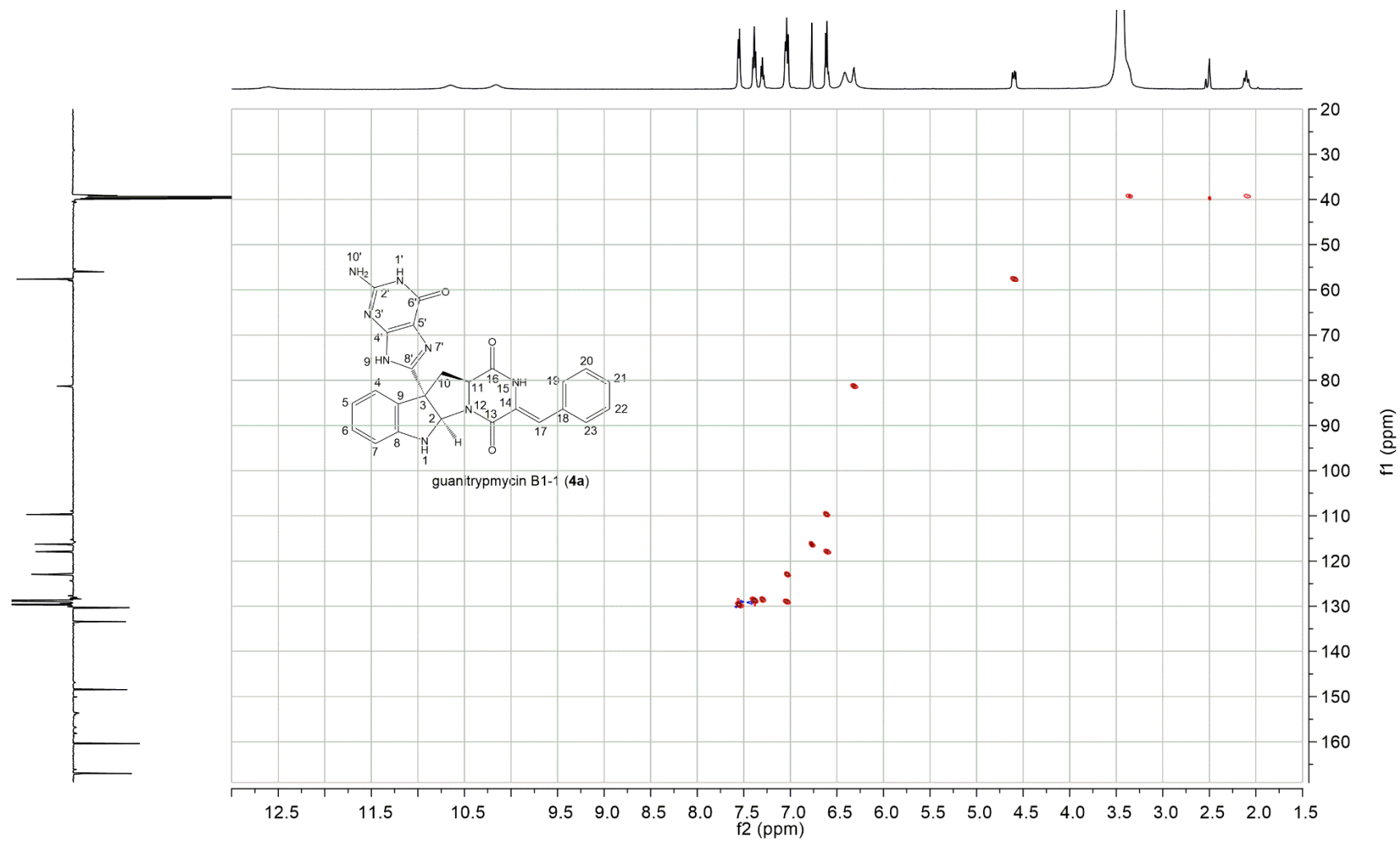


Figure S18. HSQC spectrum of guanitrypmycin B1-1 (**4a**)

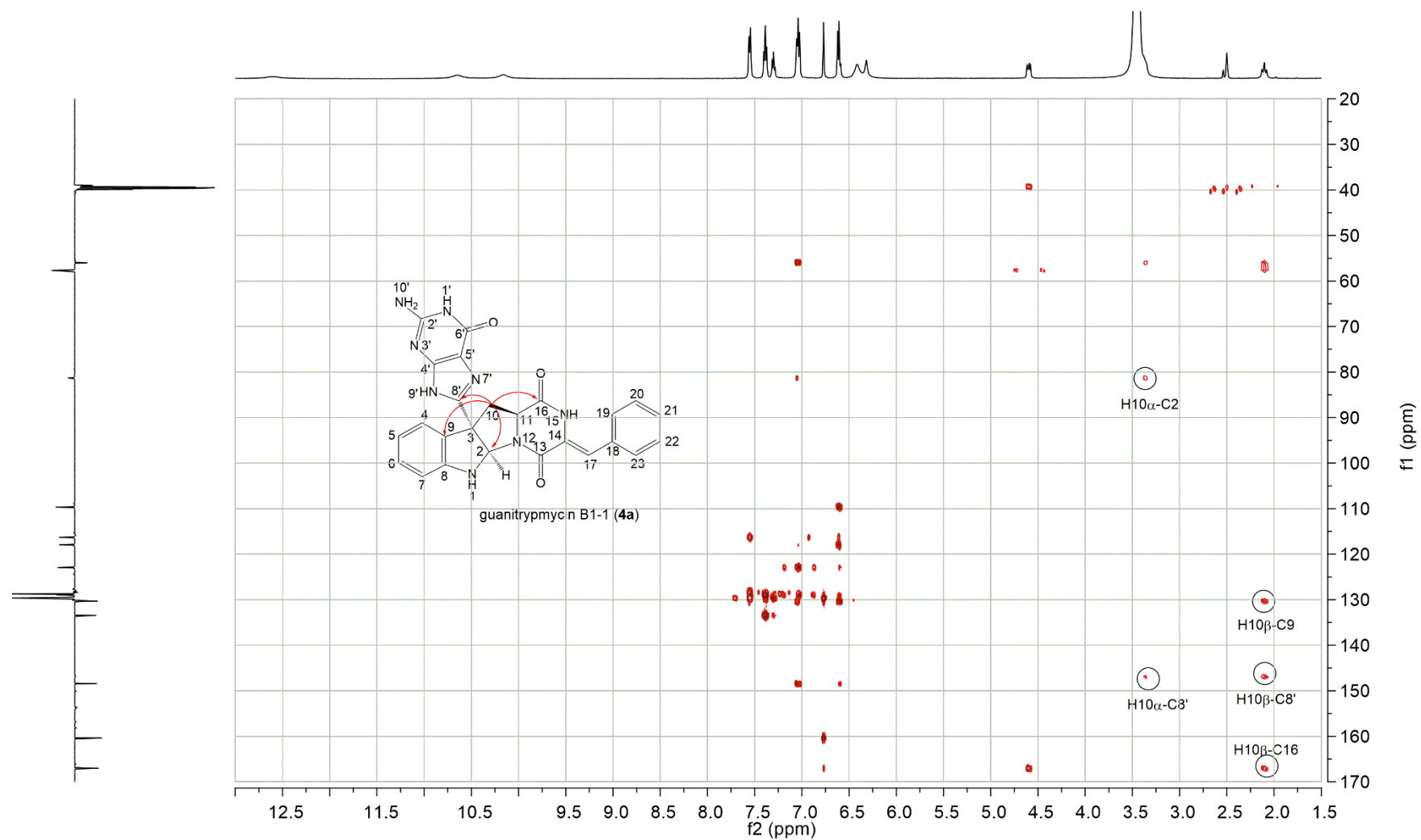


Figure S19. HMBC spectrum of guanitrymycin B1-1 (**4a**).

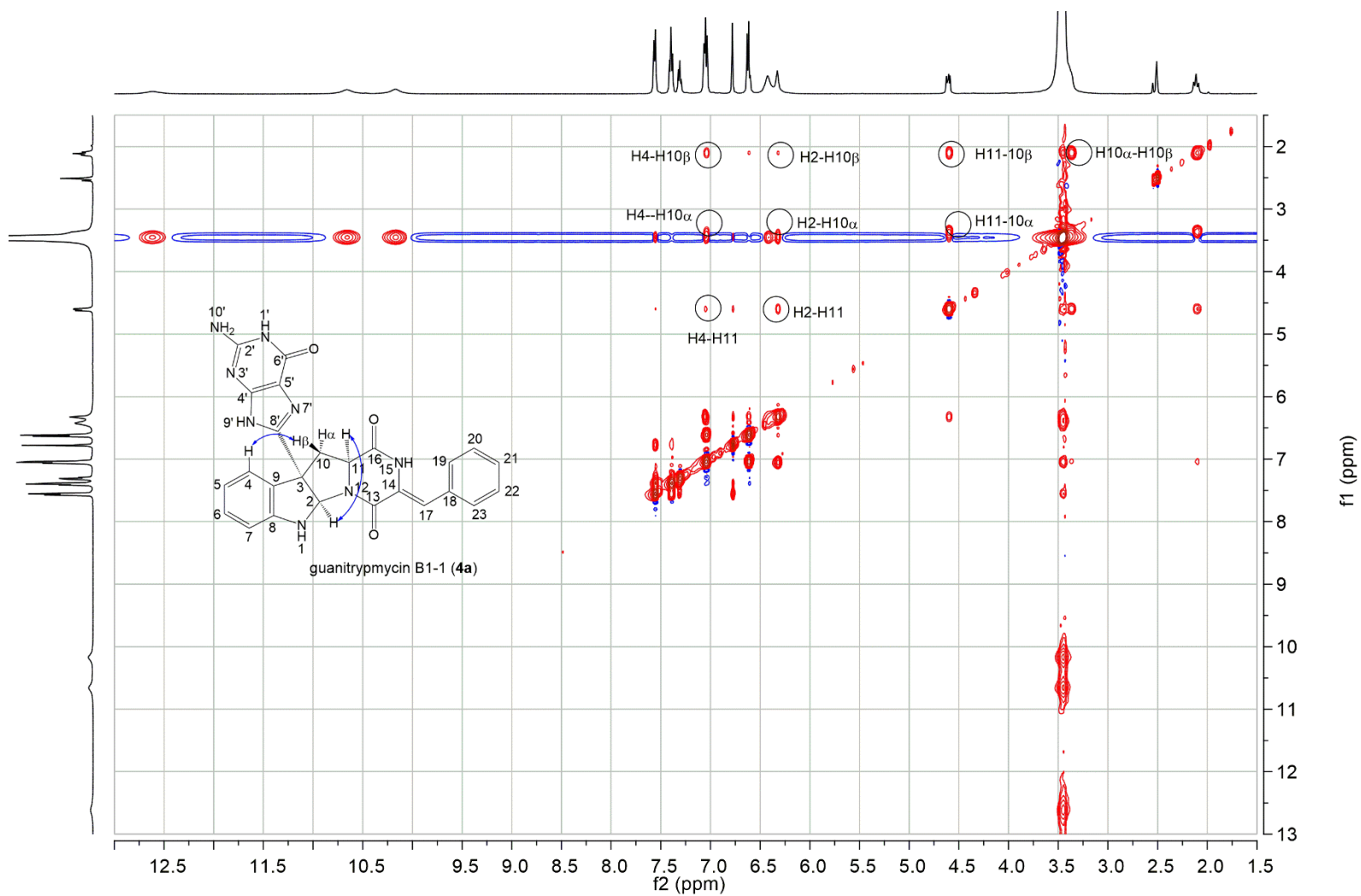


Figure S20. NOESY spectrum of guanitrypmycin B1-1 (**4a**).

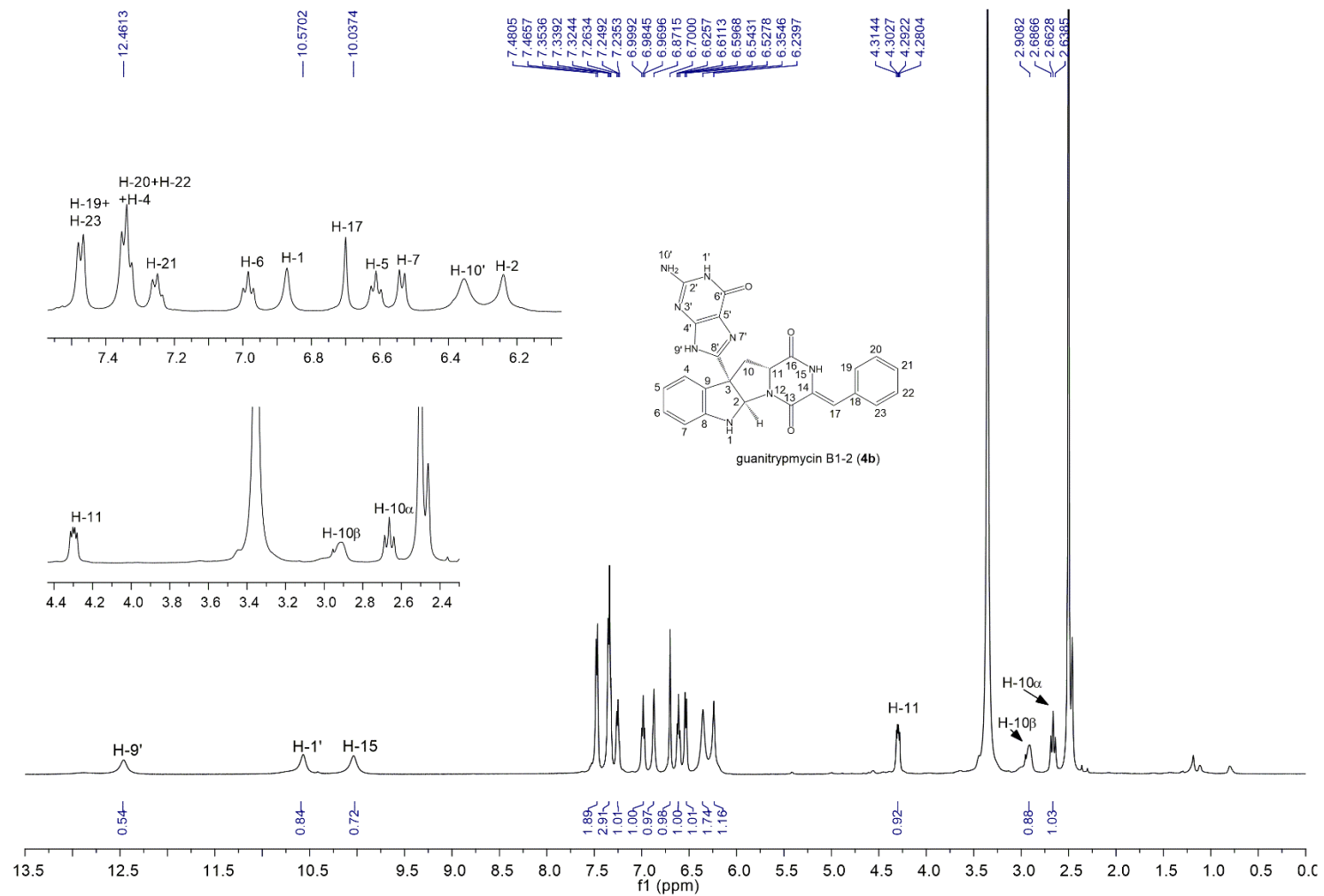


Figure S21. ¹H NMR spectrum of guanitrypmycin B1-2 (4b).

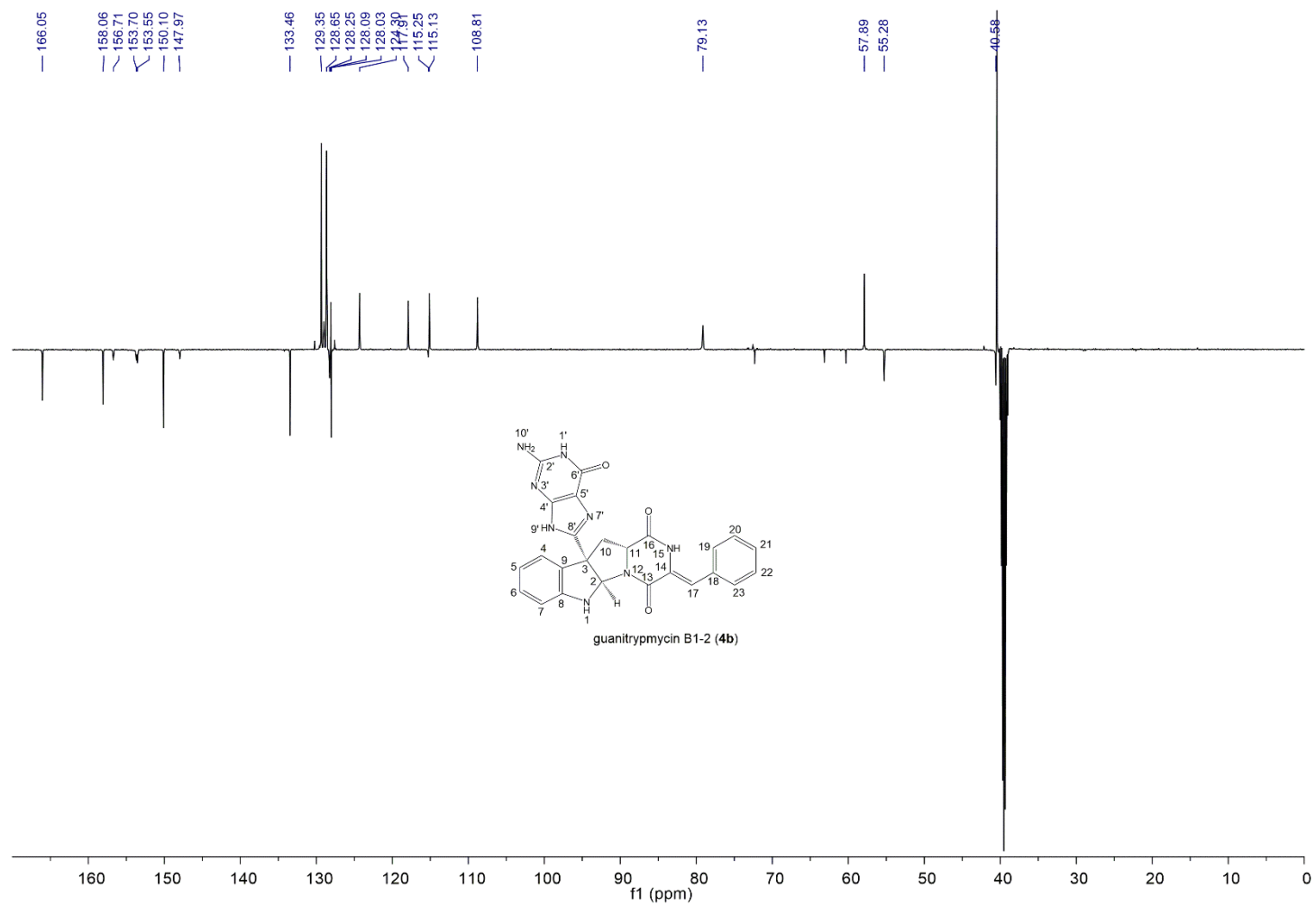


Figure S22. ^{13}C APT NMR spectrum of guanitrypmycin B1-2 (**4b**).

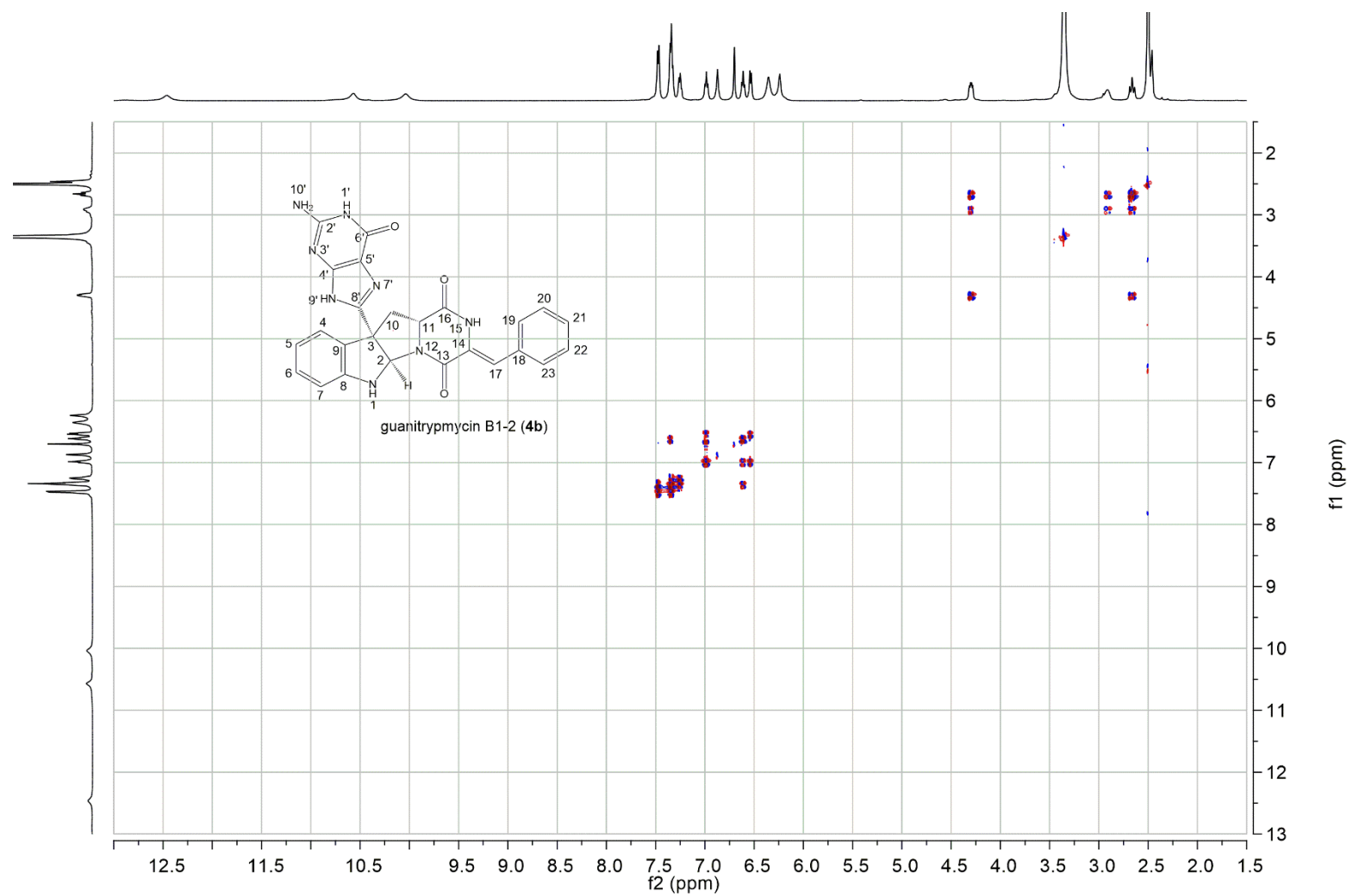


Figure S23. ^1H - ^1H COSY spectrum of guanitrypmycin B1-2 (**4b**).

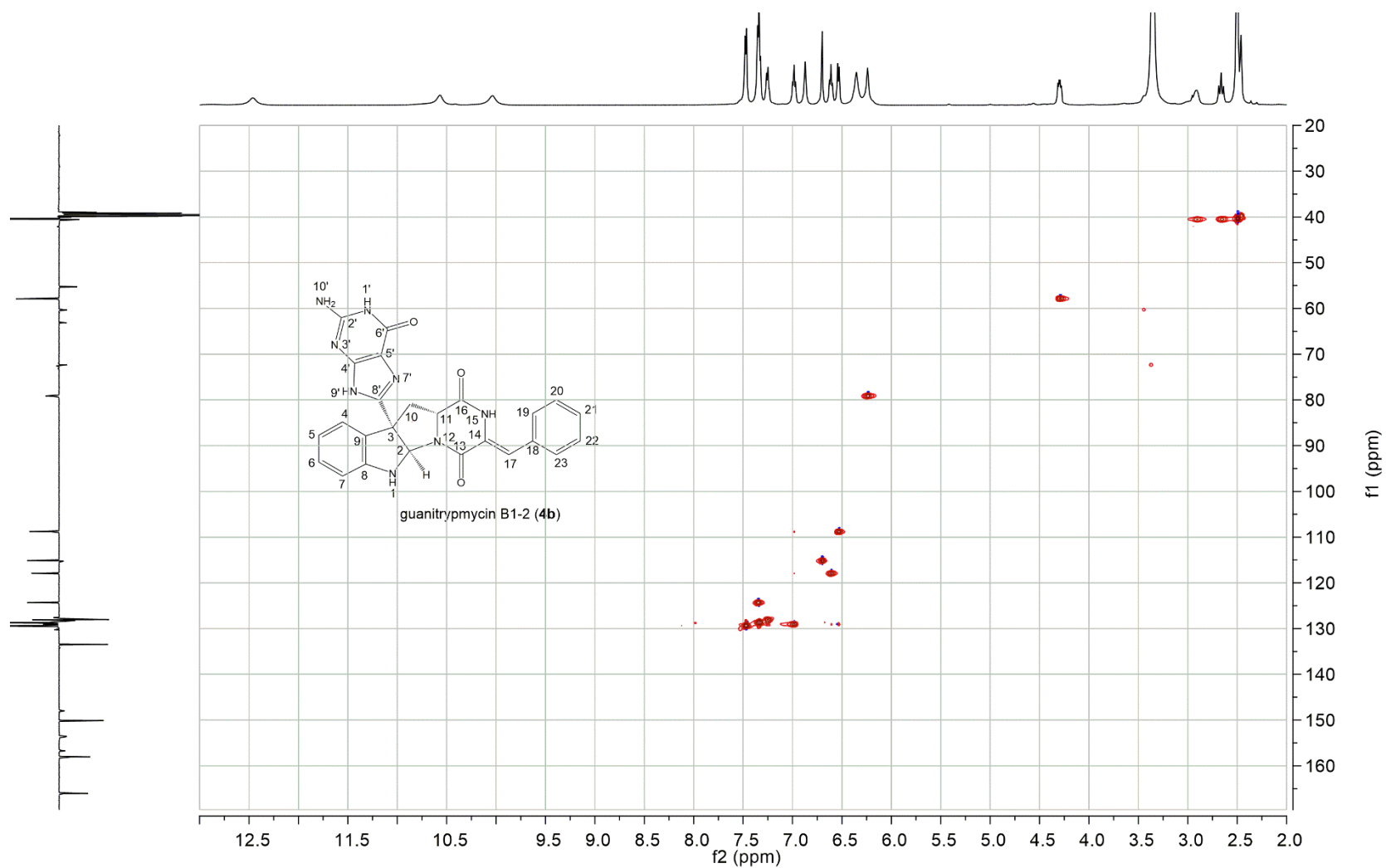


Figure S24. HSQC spectrum of guanitrypmycin B1-2 (4b).

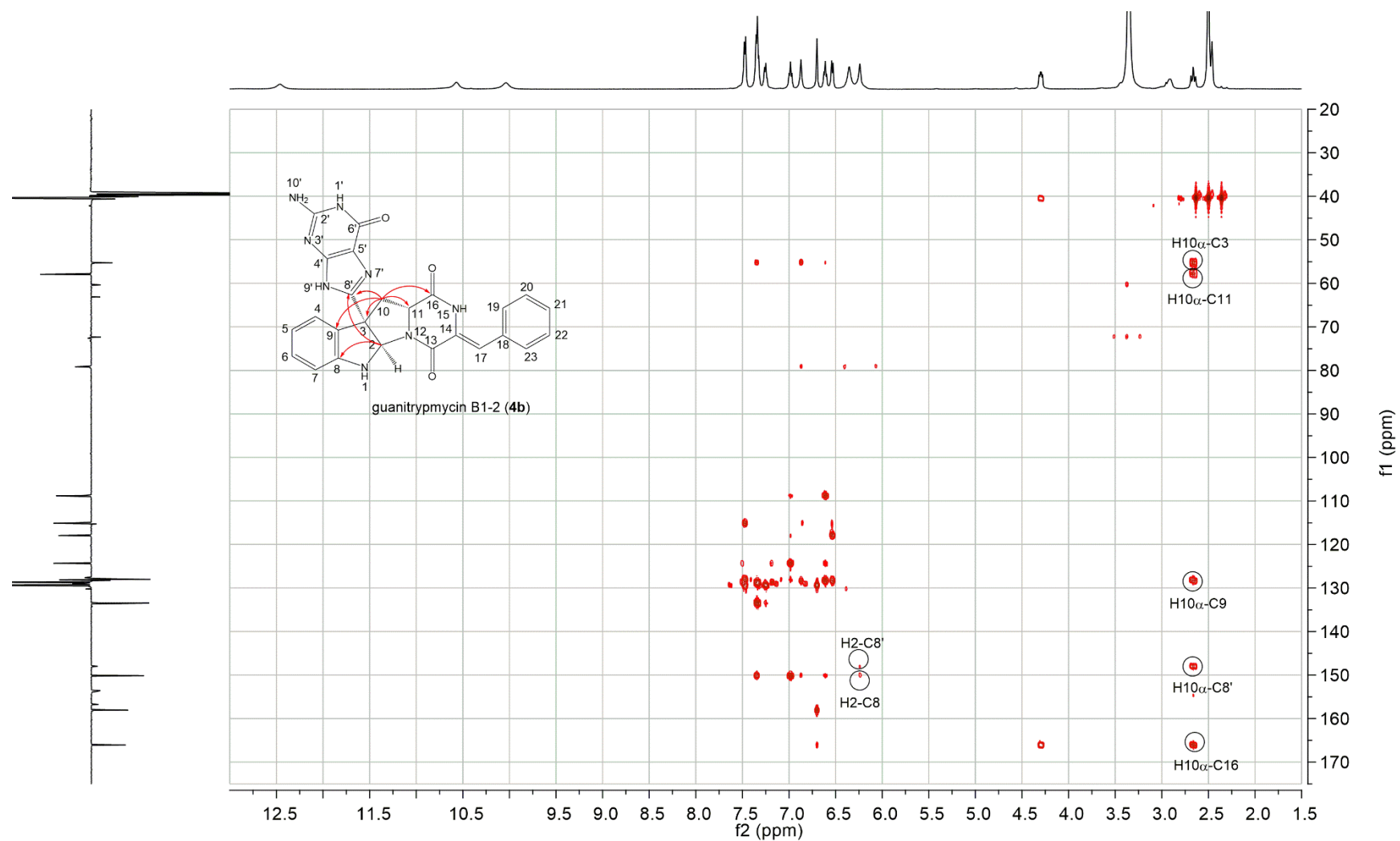


Figure S25. HMBC spectrum of guanitrypmycin B1-2 (**4b**).

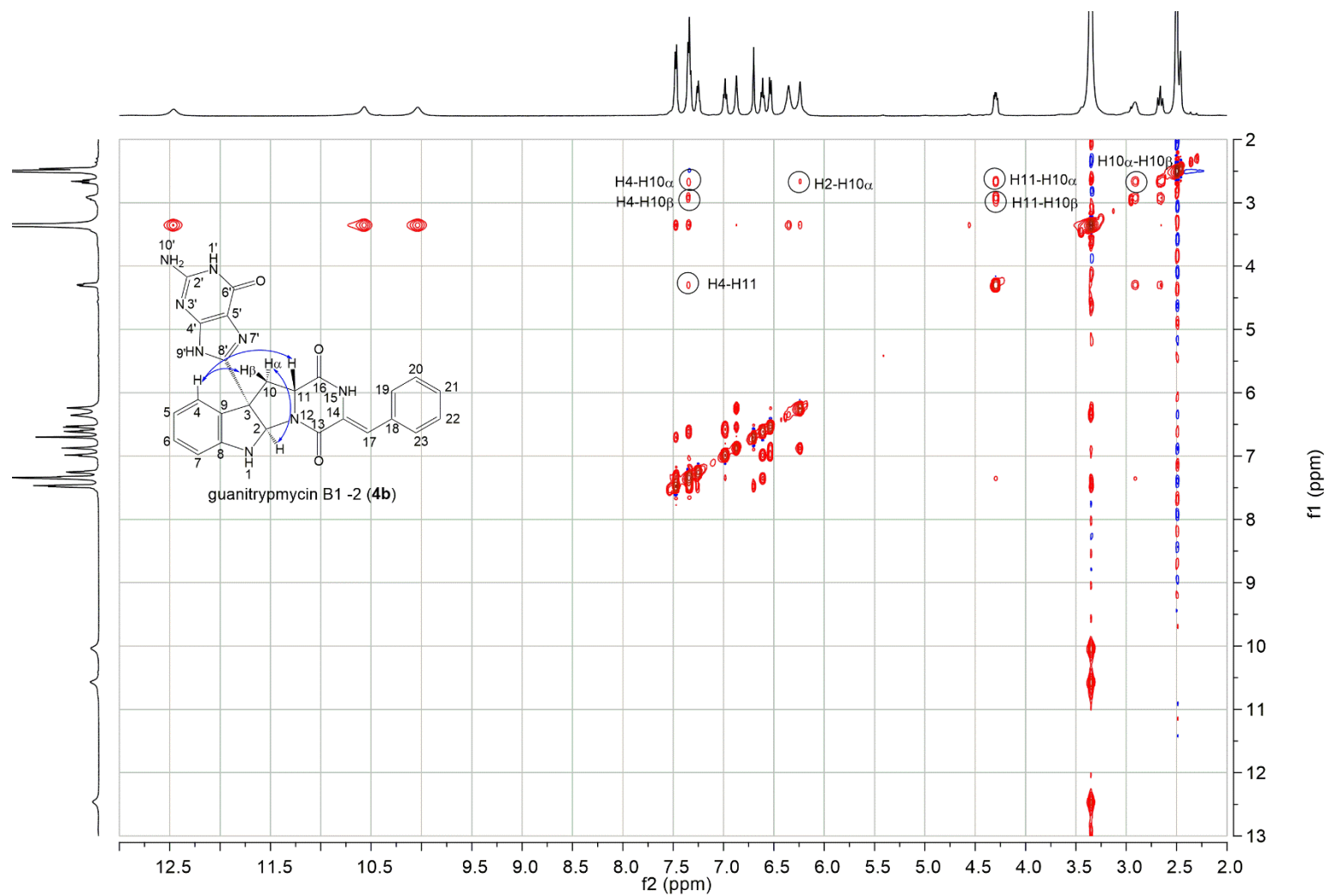


Figure S26. NOESY spectrum of guanitrypmycin B1-2 (**4b**).

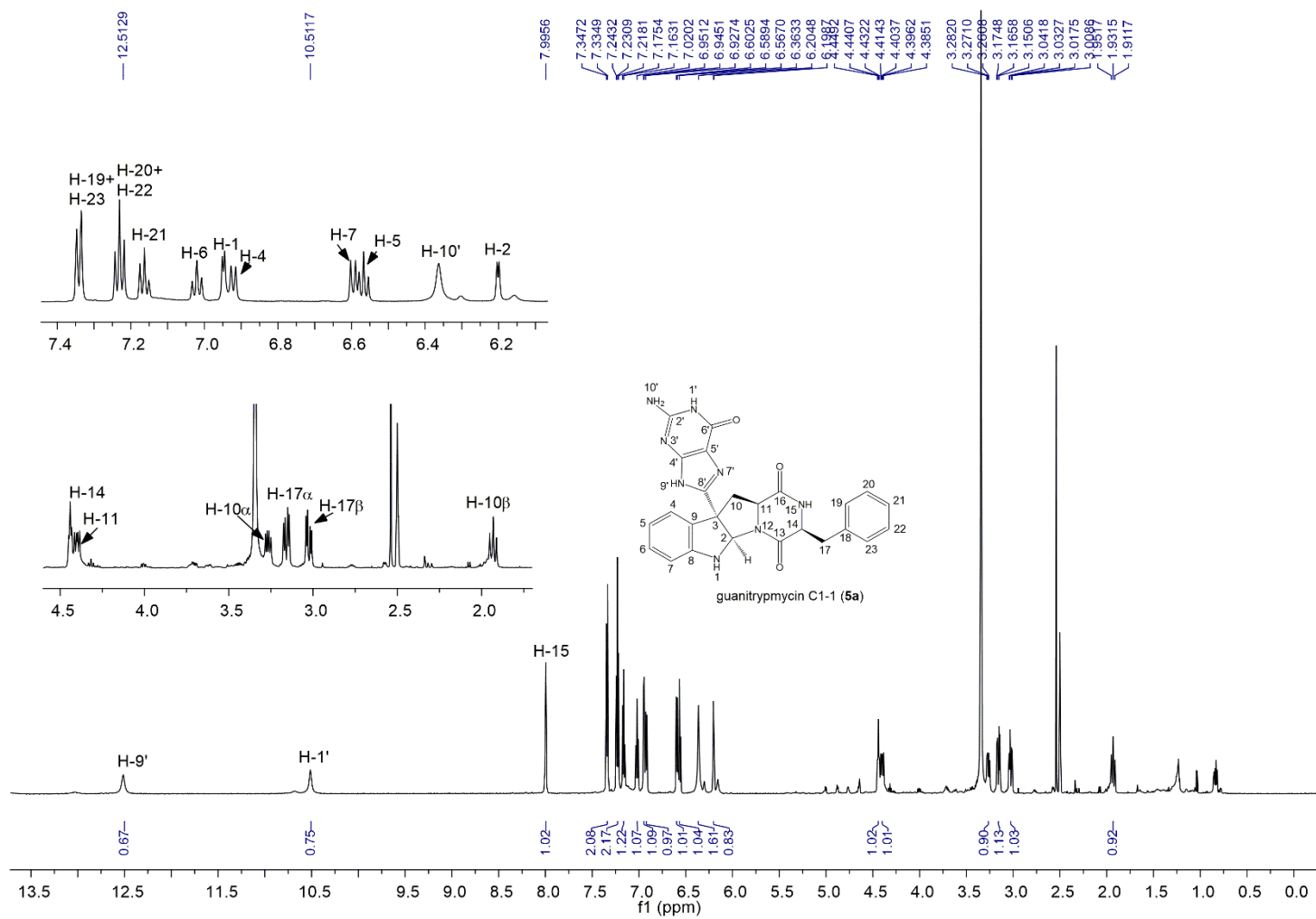


Figure S27. ¹H NMR spectrum of guanitrypmycin C1-1 (5a).

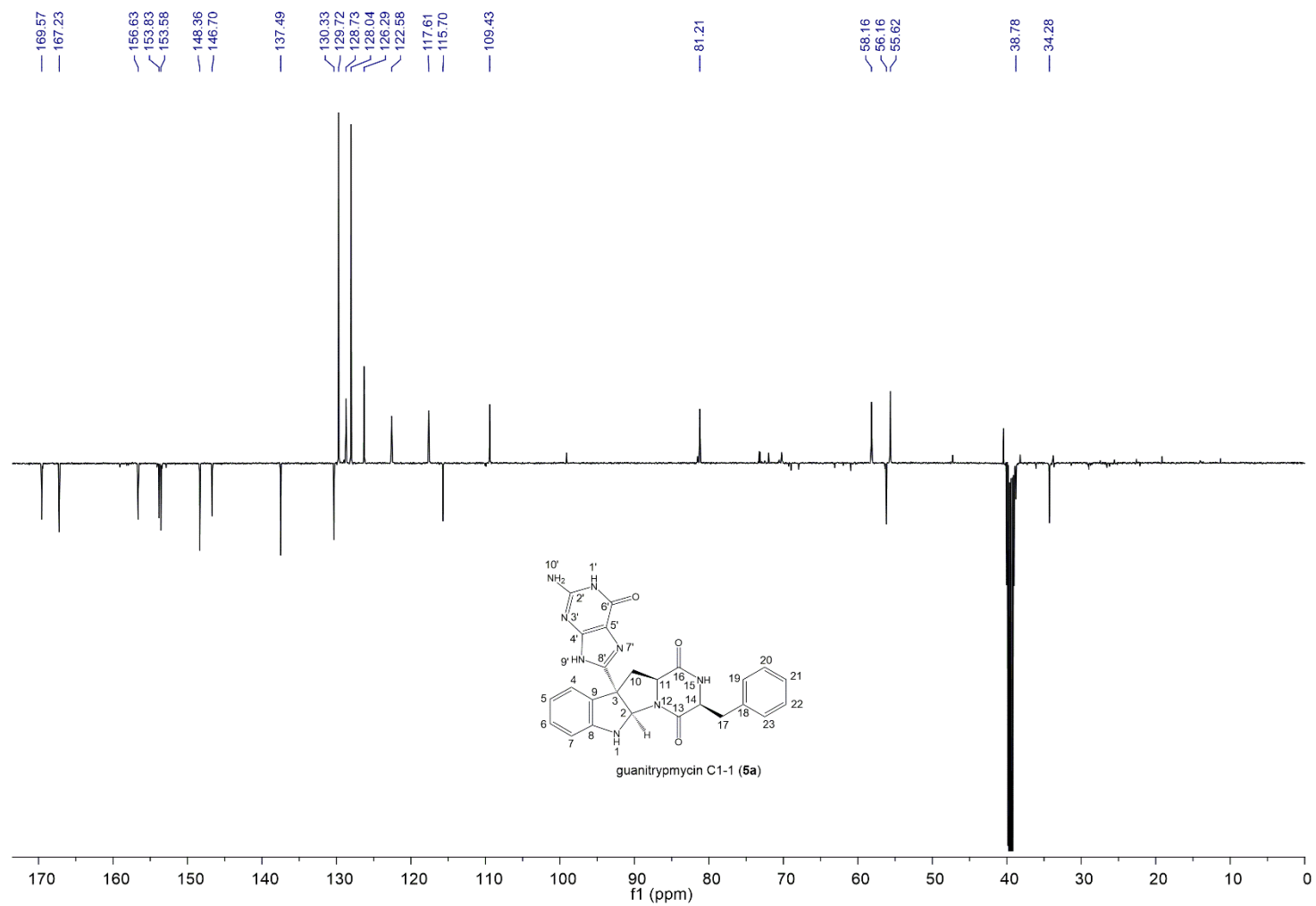


Figure S28. ^{13}C APT NMR spectrum of guanitrypmycin C1-1 (**5a**).

S52

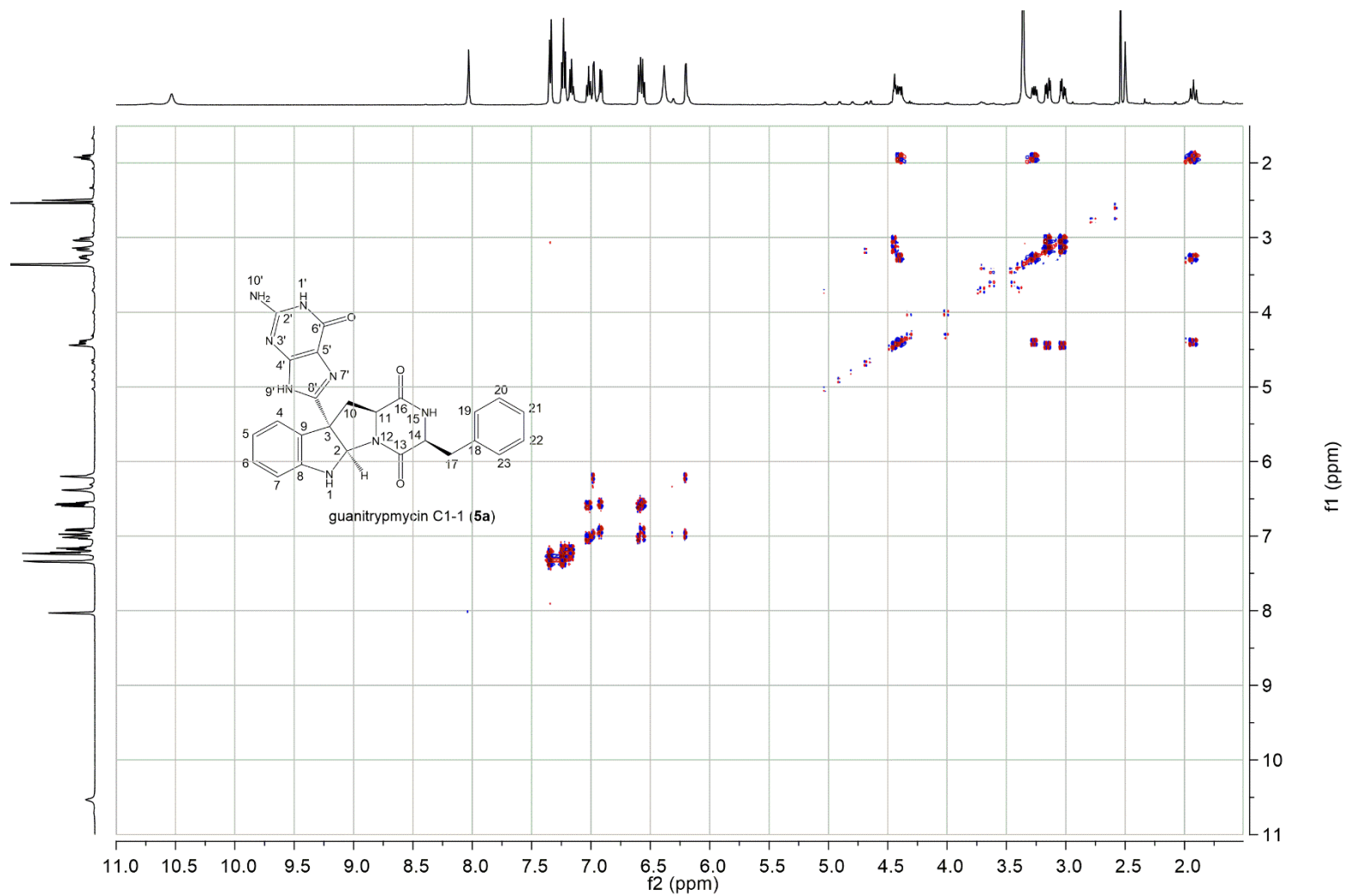


Figure S29. ^1H - ^1H COSY spectrum of guanitrypmycin C1-1 (**5a**).

S53

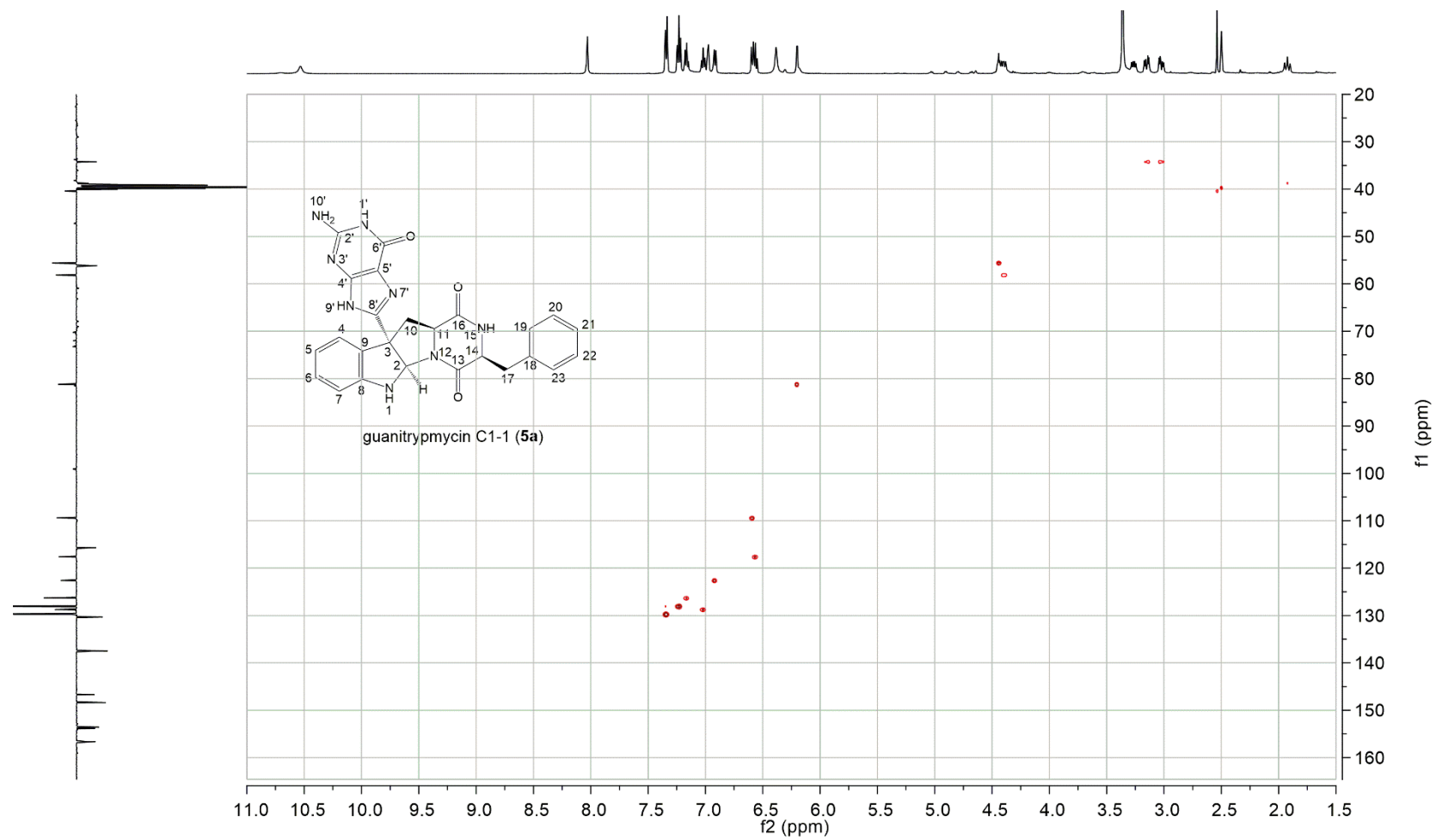


Figure S30. HSQC spectrum of guanitrypmycin C1-1 (**5a**).

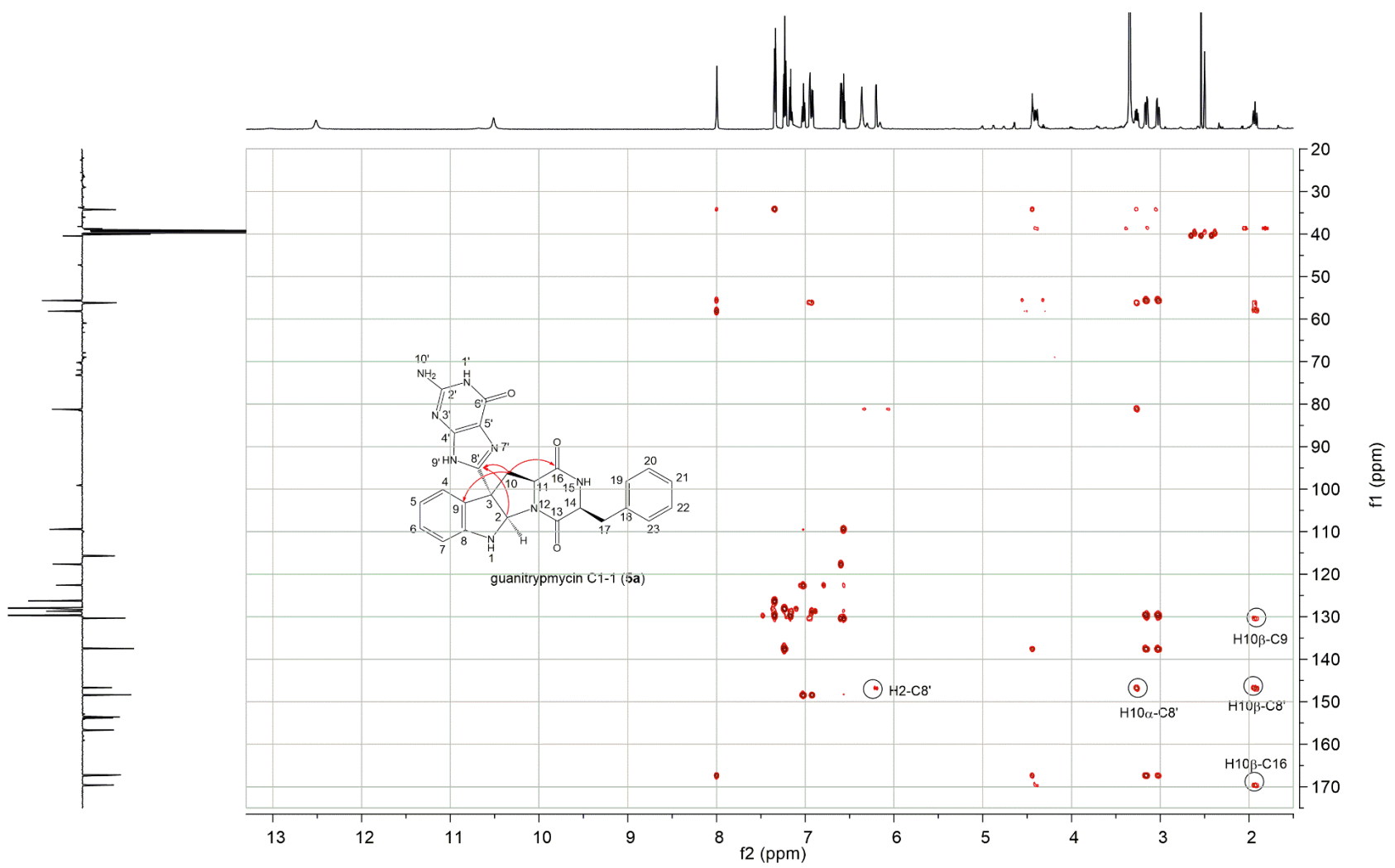


Figure S31. HMBC spectrum of guanitrypmycin C1-1 (**5a**).

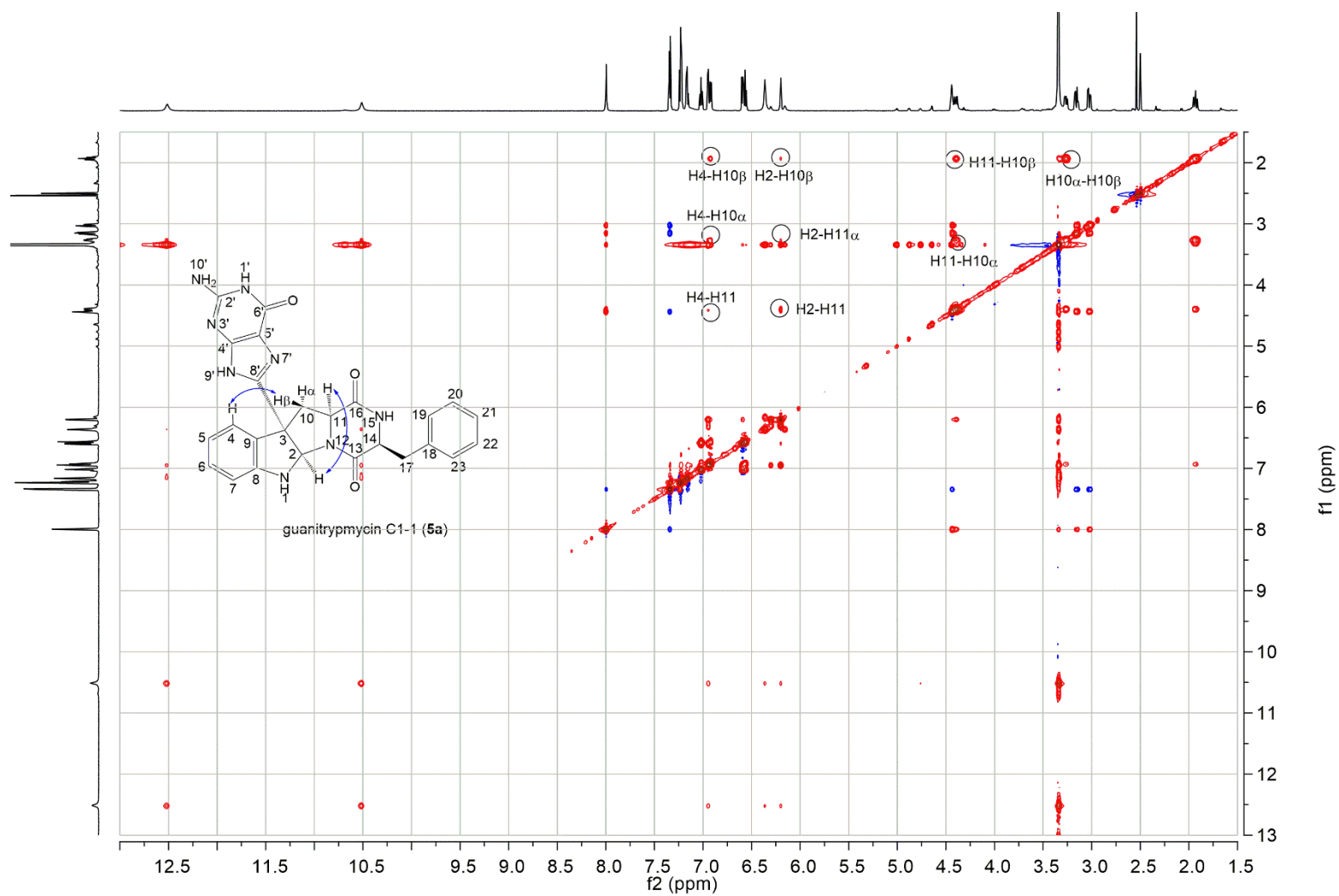


Figure S32. NOESY spectrum of guanitrypmycin C1-1 (**5a**).

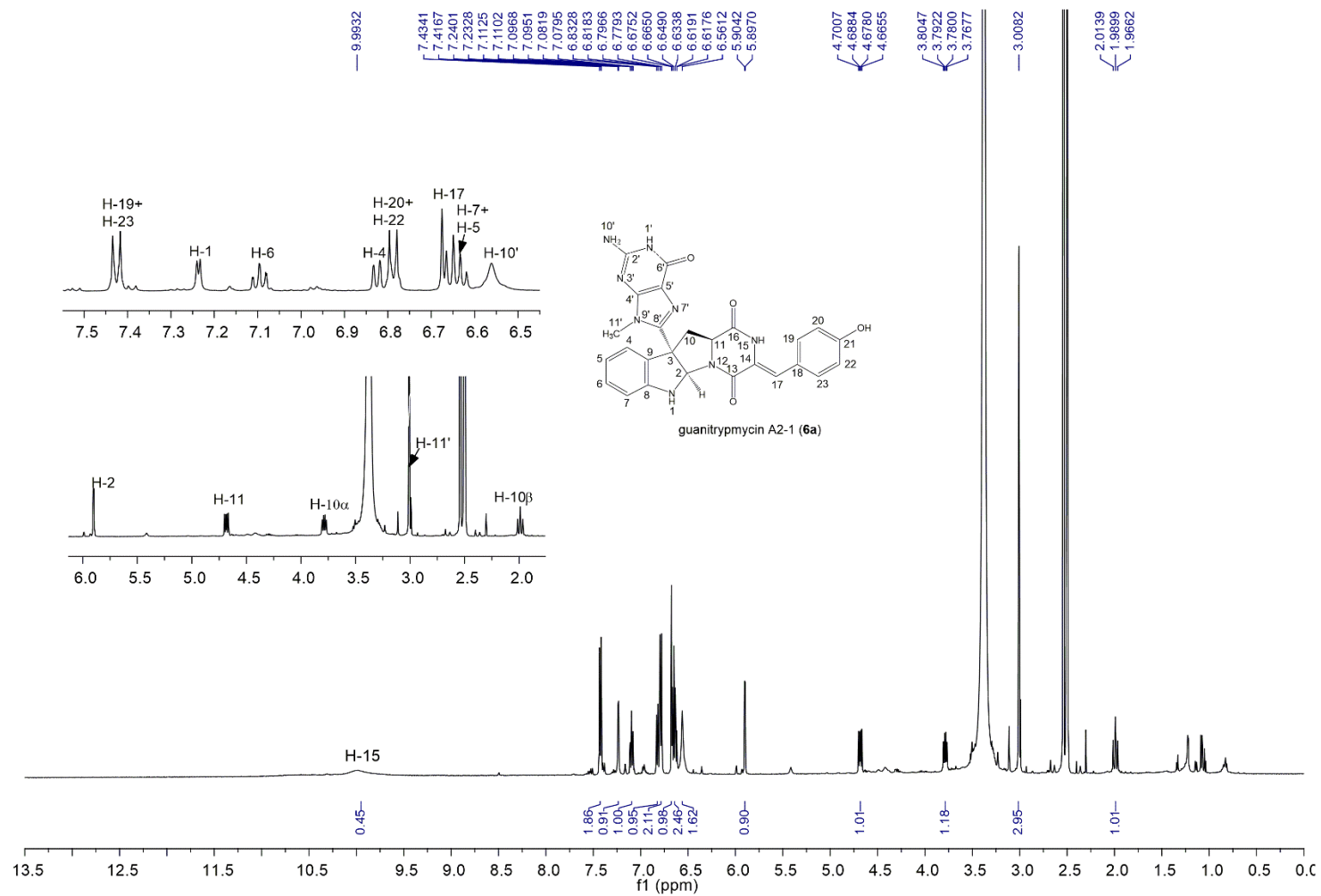


Figure S33. ¹H NMR spectrum of guanitrypmycin A2-1 (**6a**).

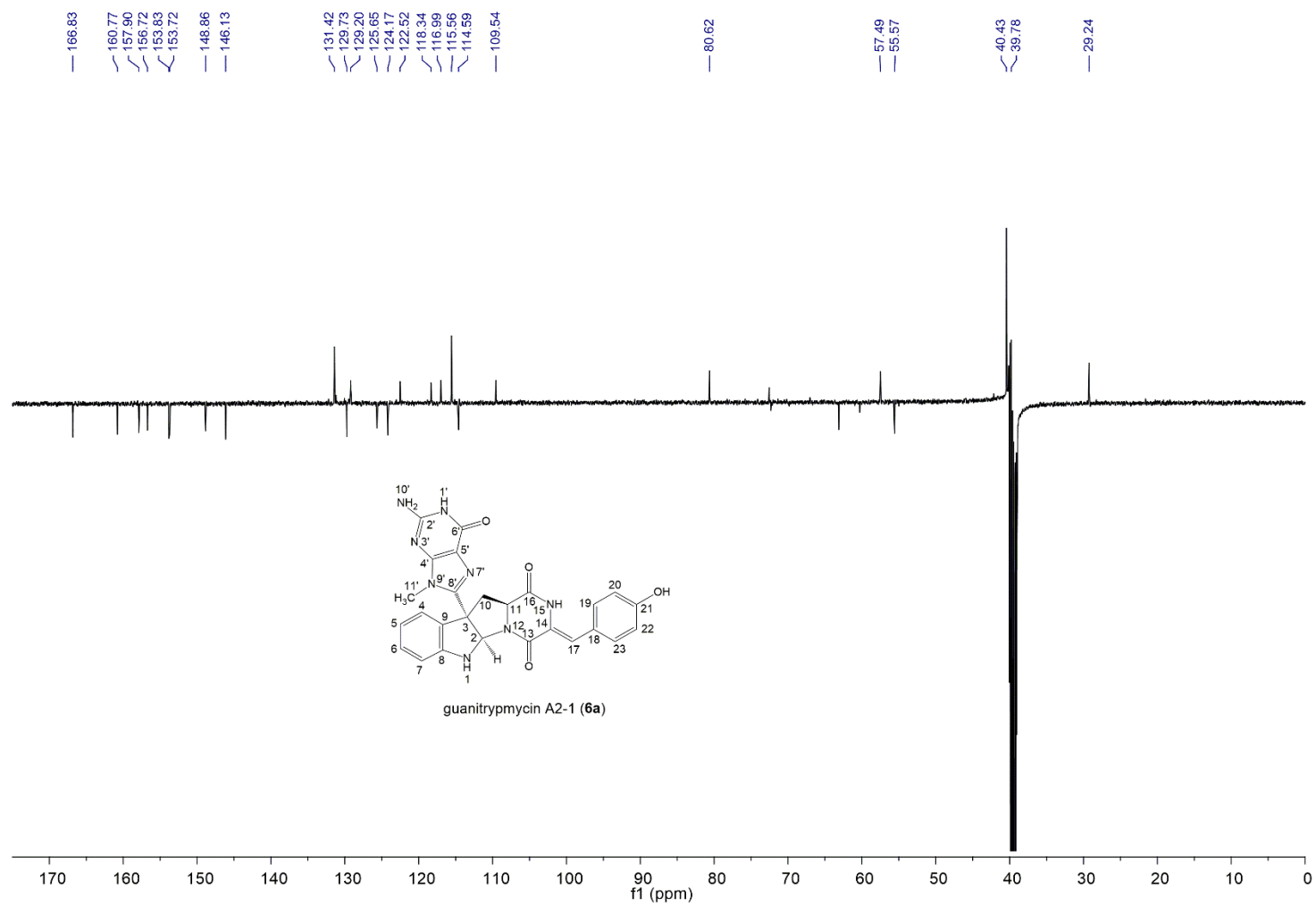


Figure S34. ¹³C APT NMR spectrum of guanitrypmycin A2-1 (6a).

S58

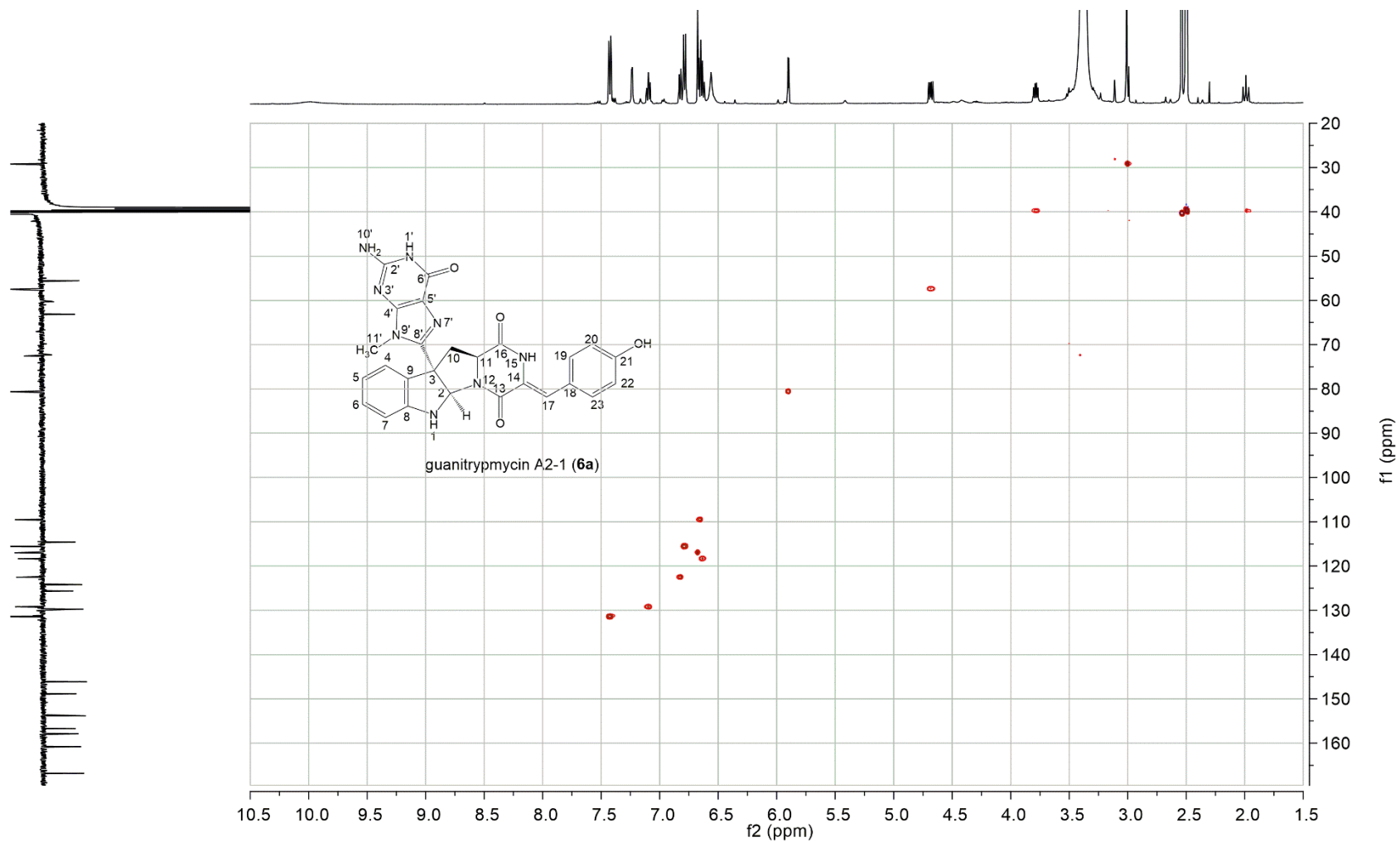


Figure S36. HSQC spectrum of guanitrypmycin A2-1 (**6a**).

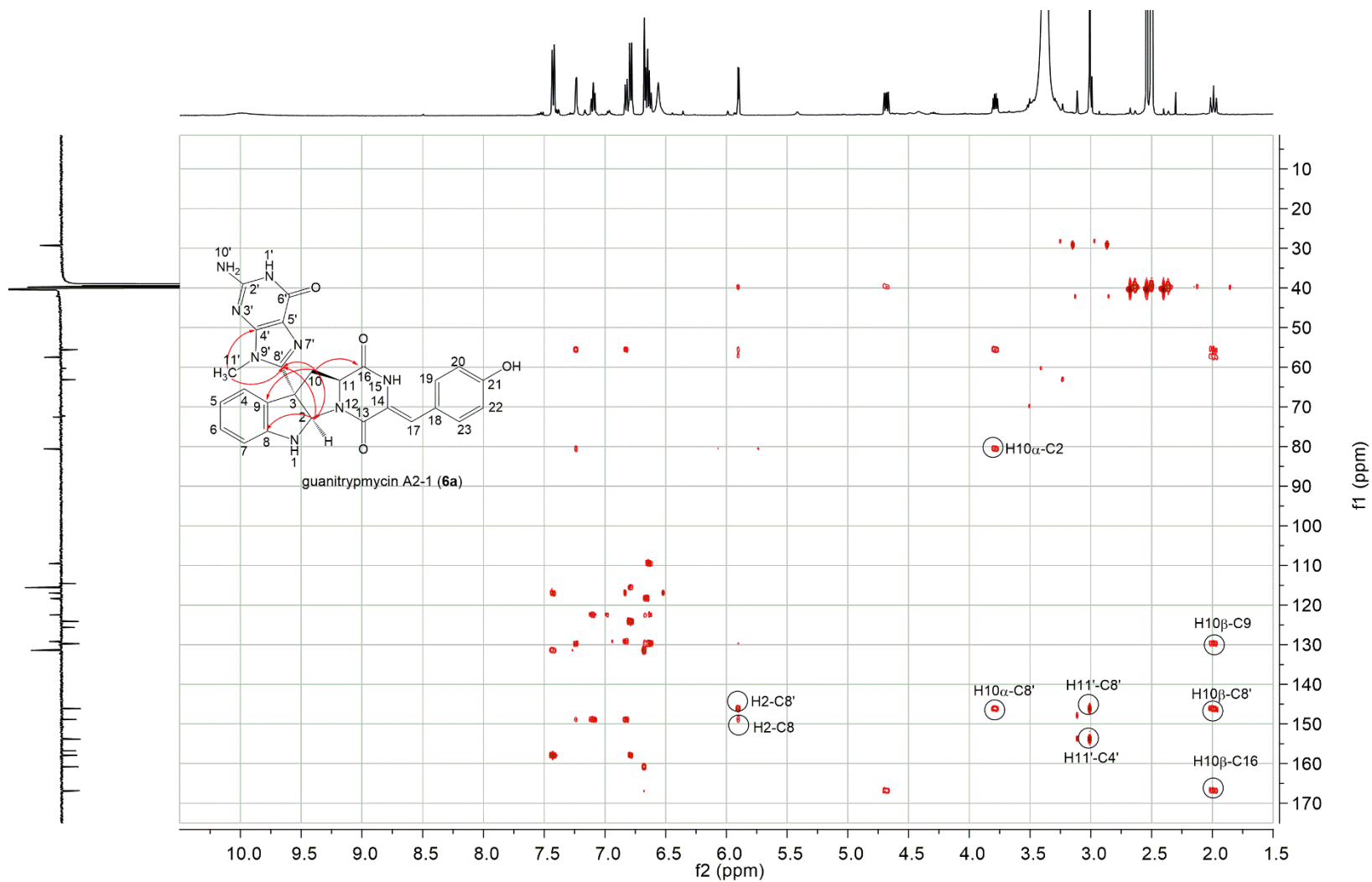


Figure S37. HMBC spectrum of guanitrypmycin A2-1 (**6a**).

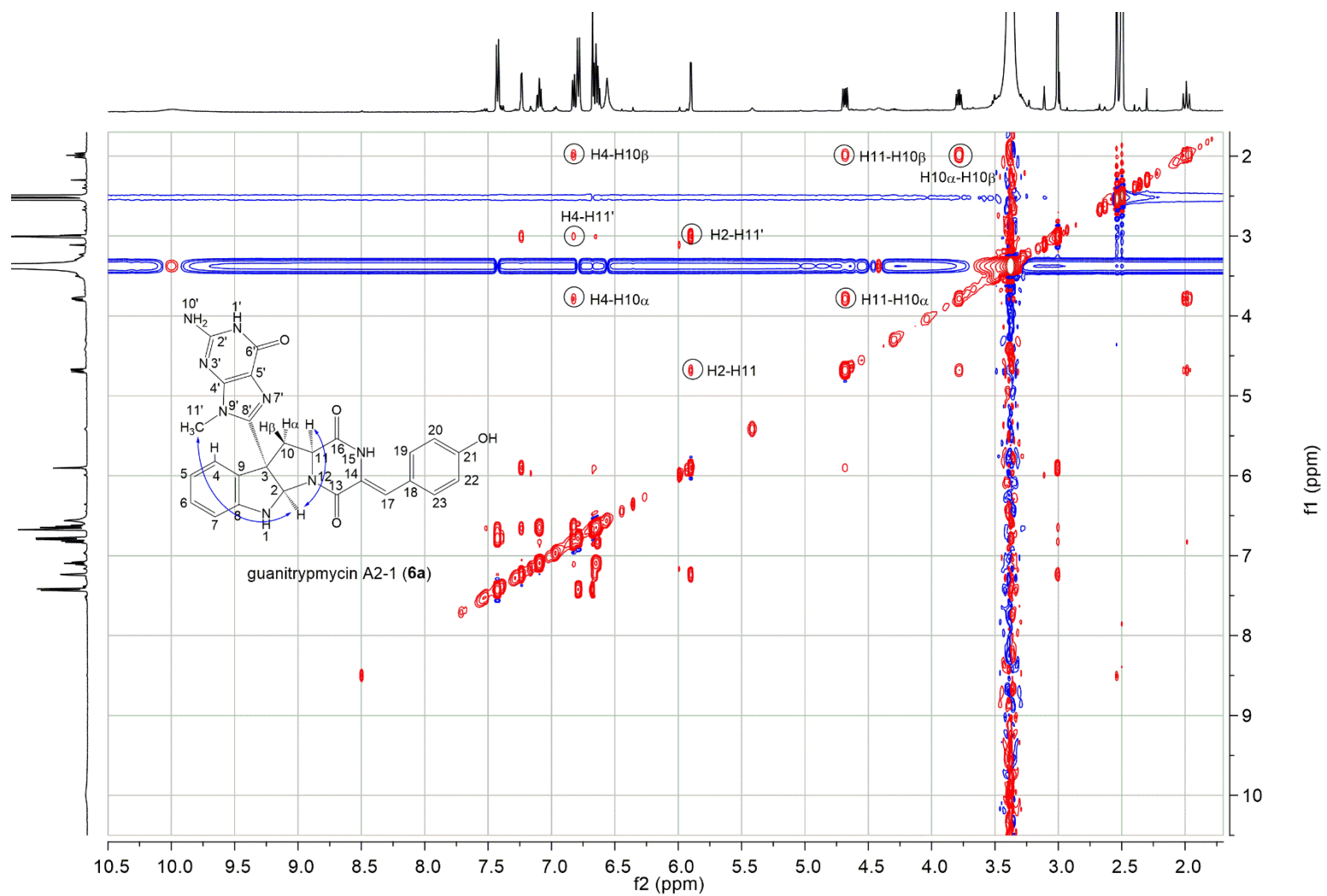


Figure S38. NOESY spectrum of guanitrypmycin A2-1 (**6a**).

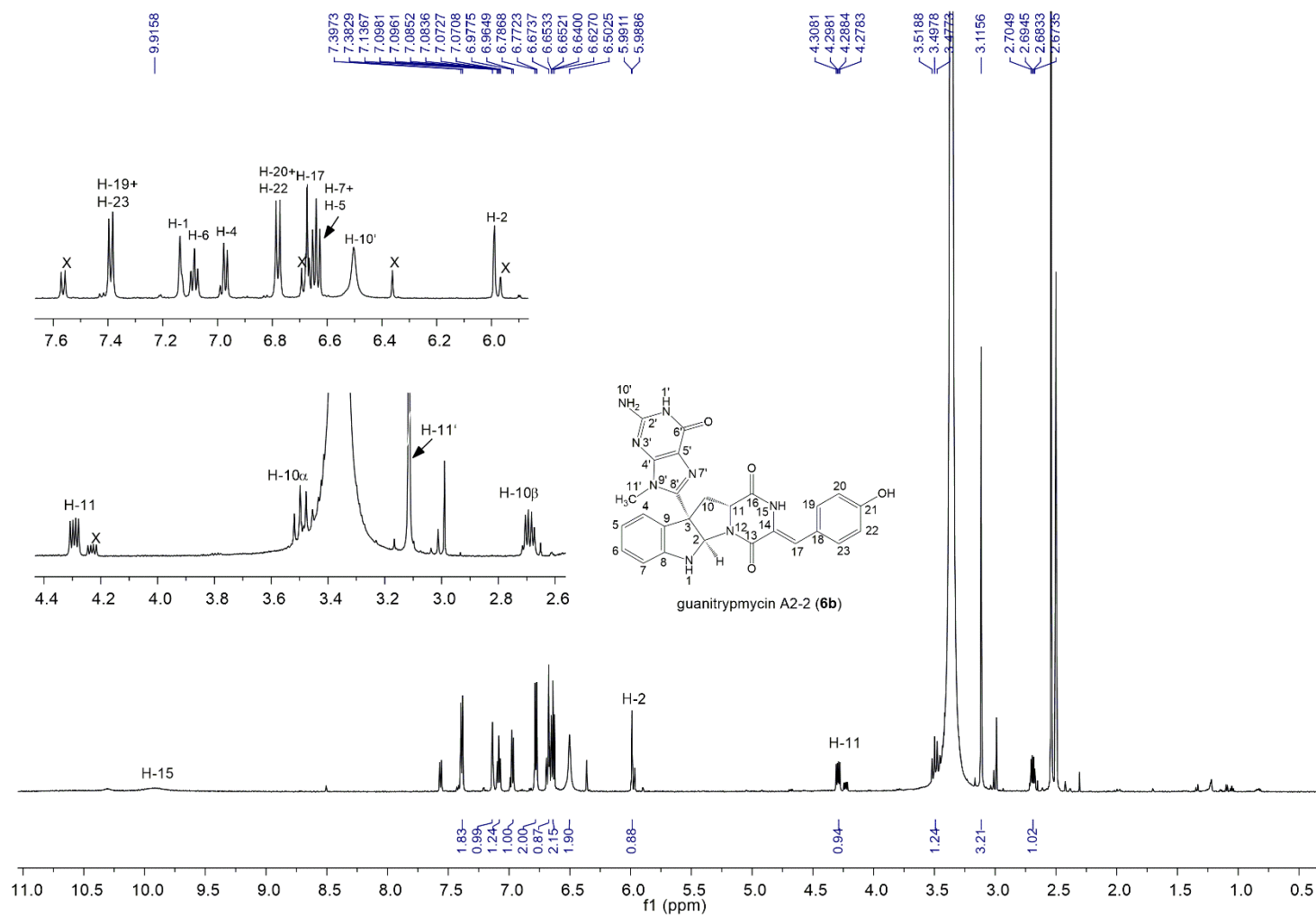


Figure S39. ¹H NMR spectrum of guanitrypmycin A2-2 (6b).

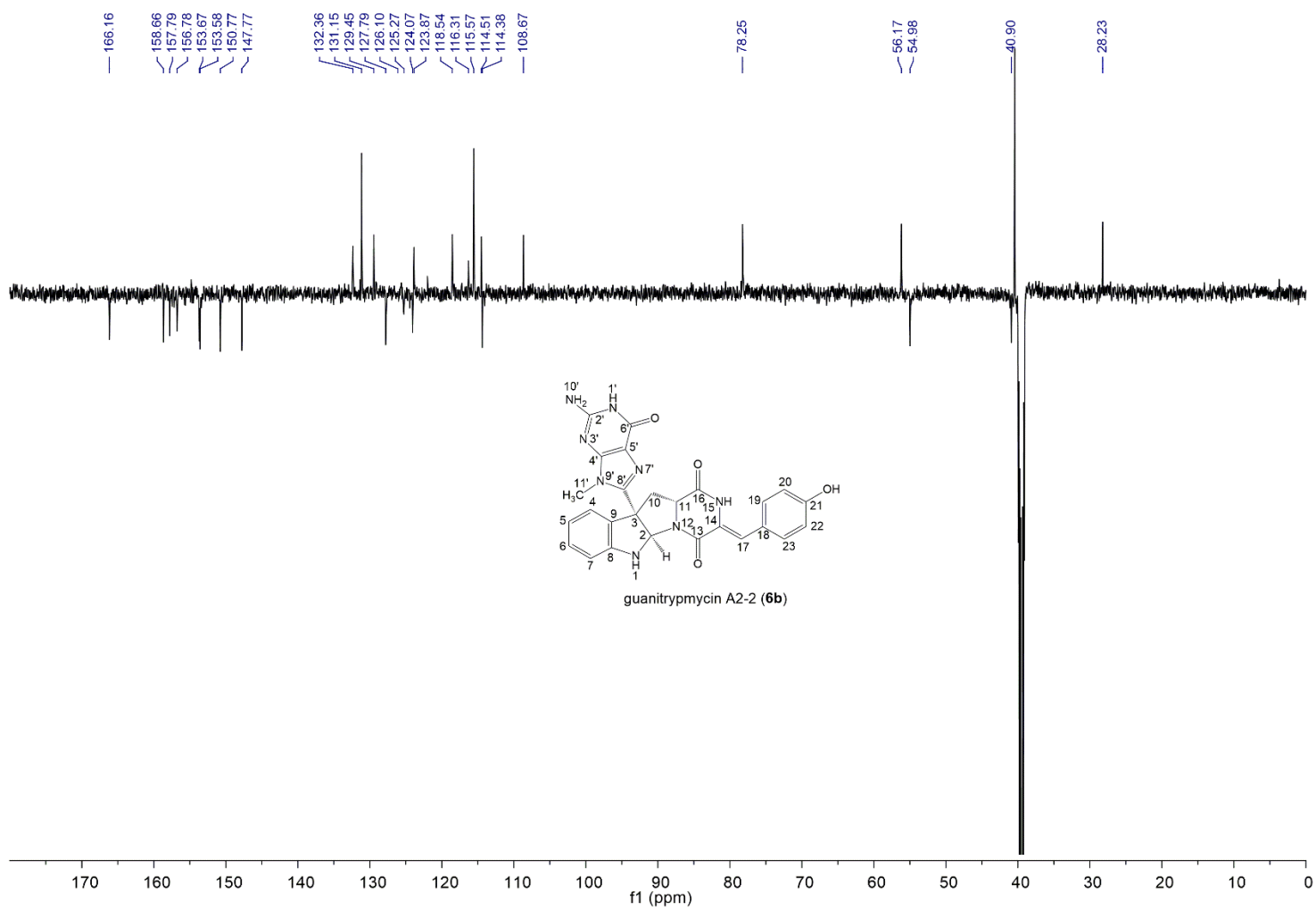
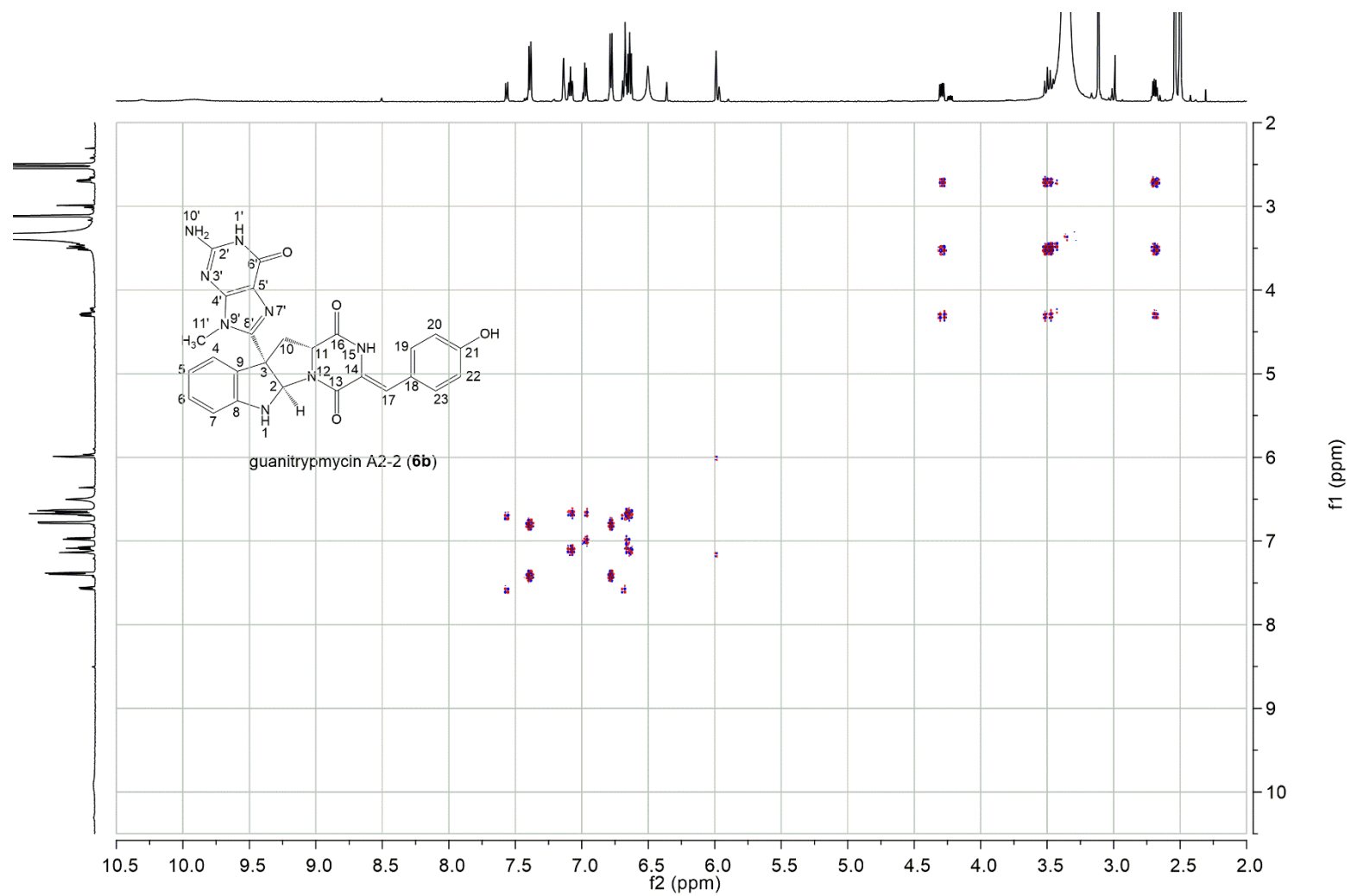
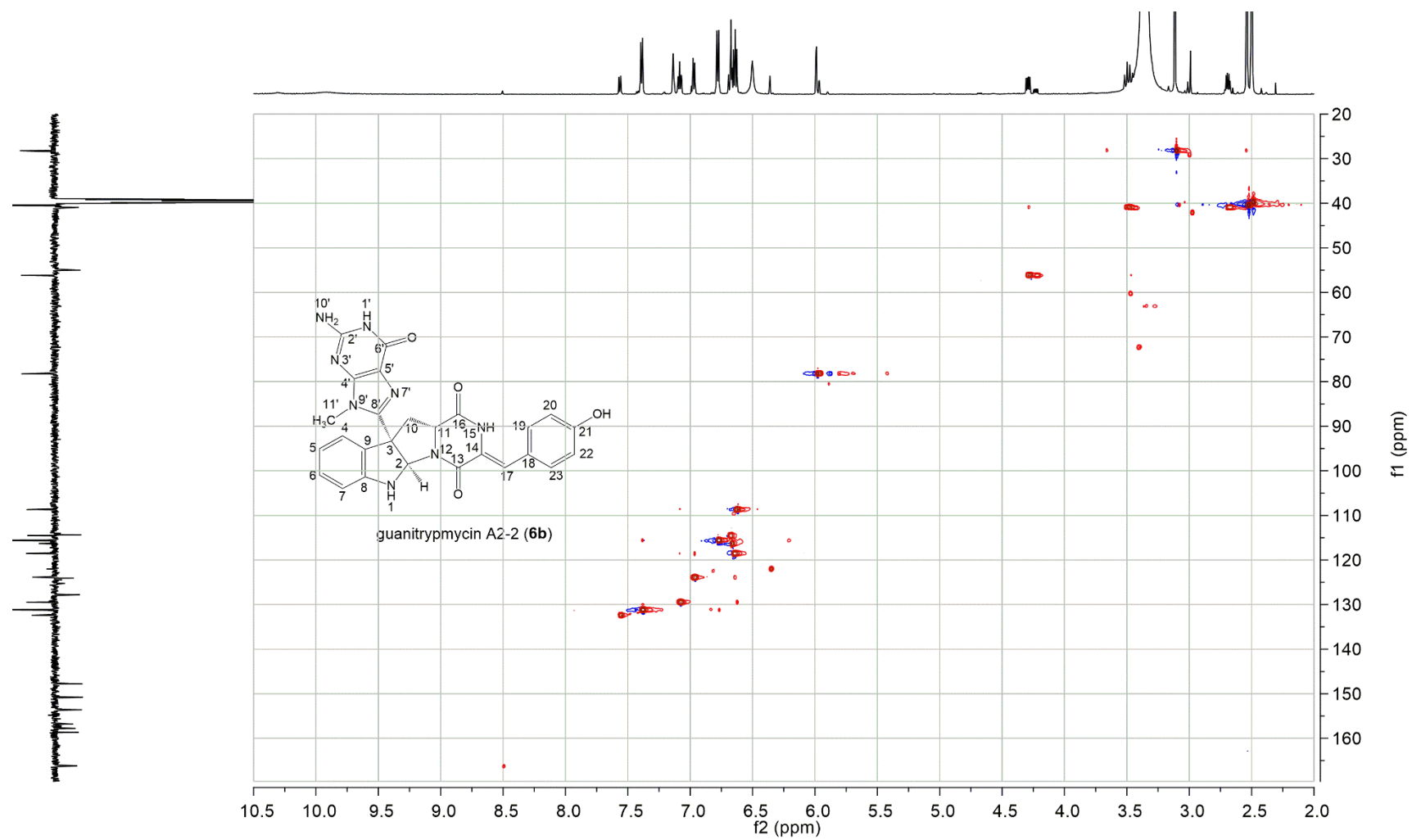


Figure S40. ^{13}C APT NMR spectrum of guanitrypmycin A2-2 (**6b**).

S64



S65



S66

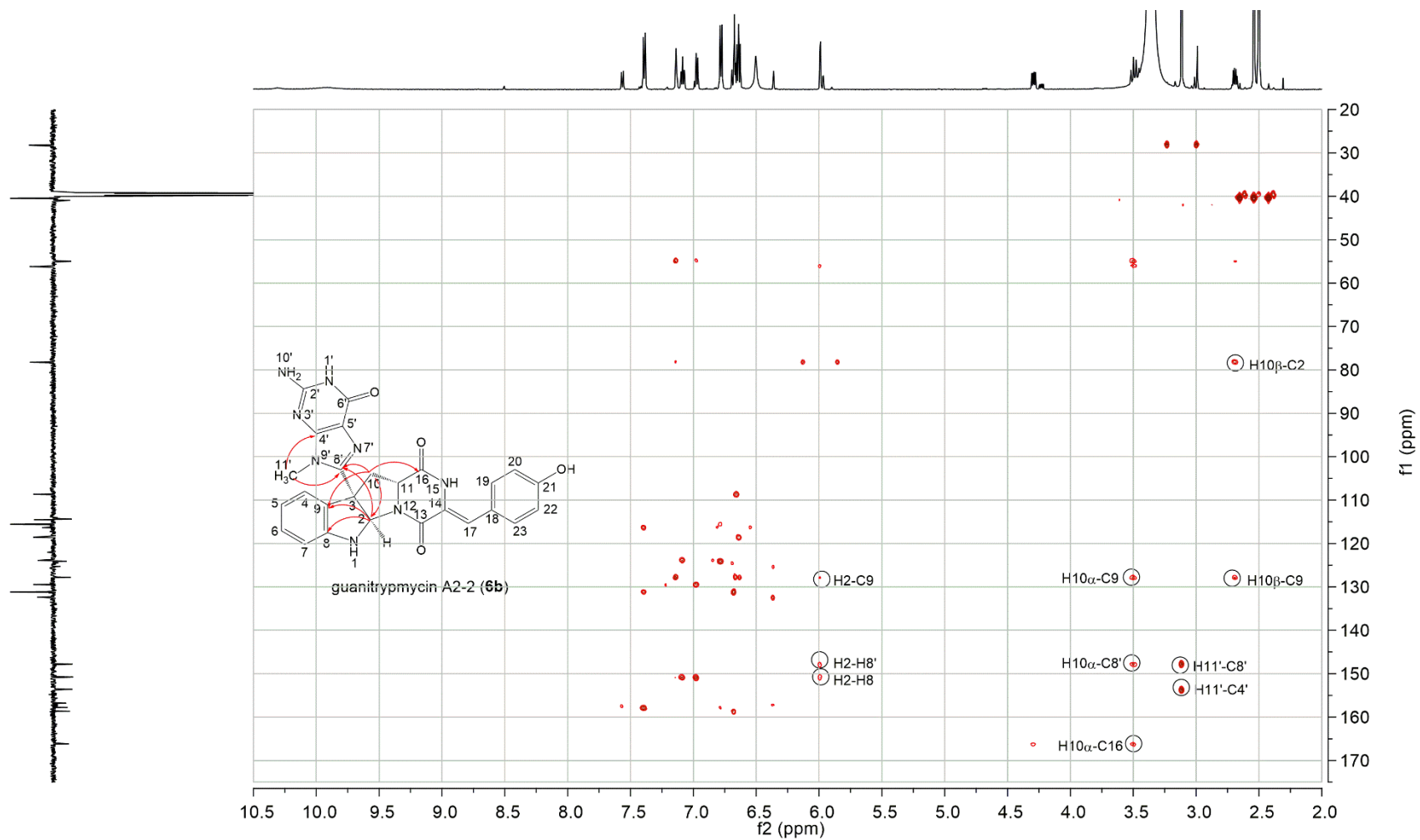


Figure S43. HMBC spectrum of guanitrypmycin A2-2 (**6b**).

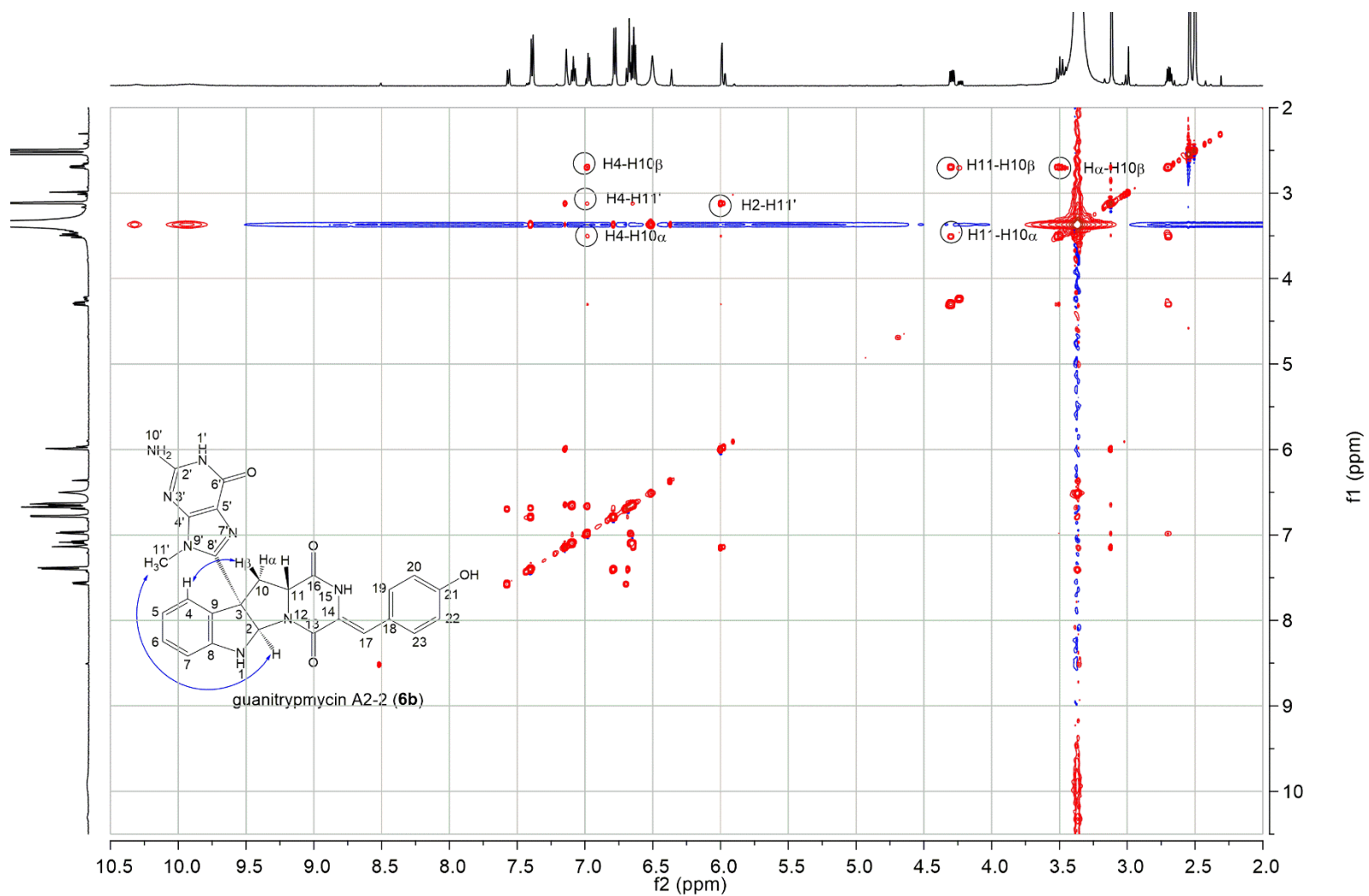


Figure S44. NOESY spectrum of guanitrypmycin A2-2 (**6b**).

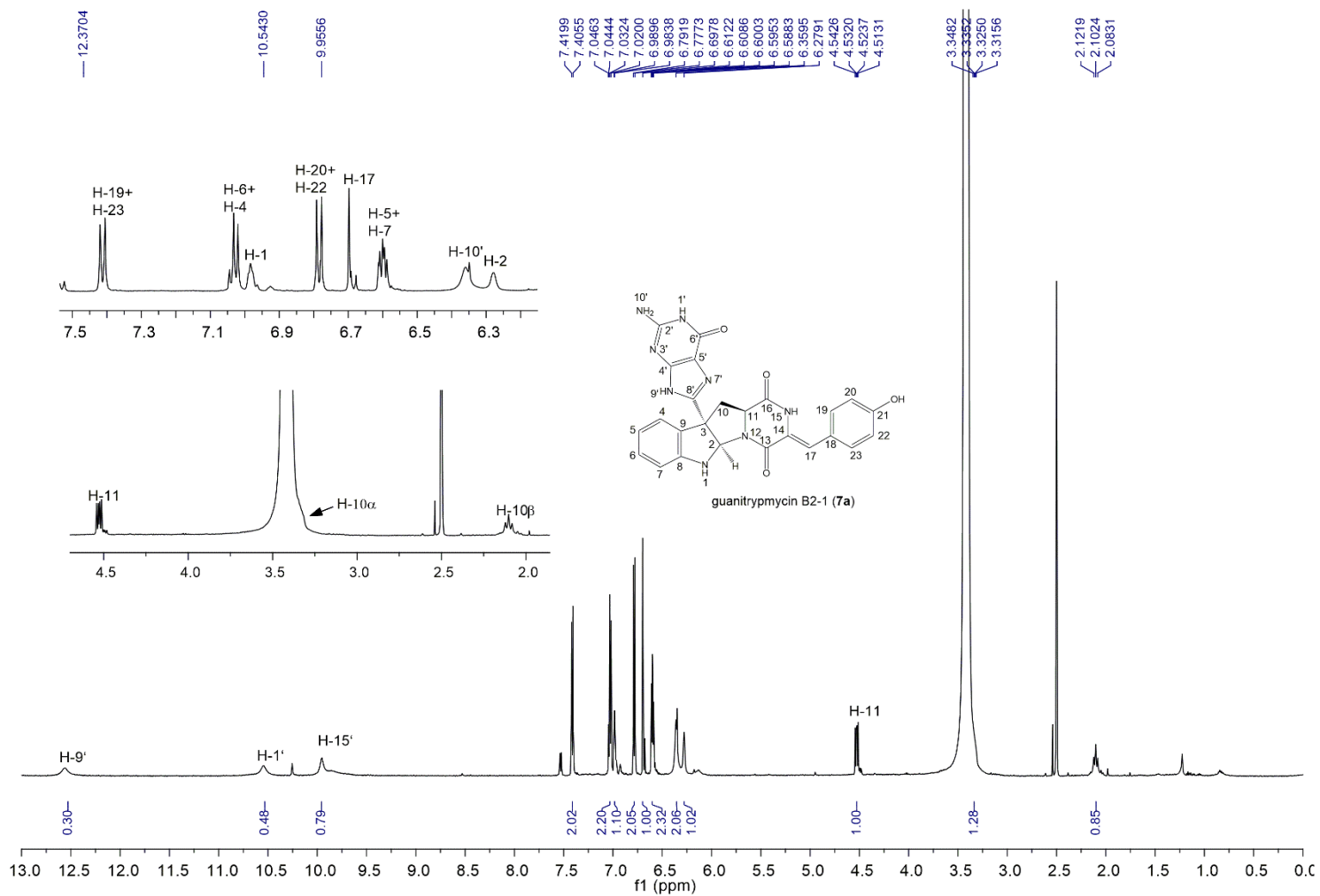


Figure S45. ¹H NMR spectrum of guanitrypmycin B2-1 (7a).

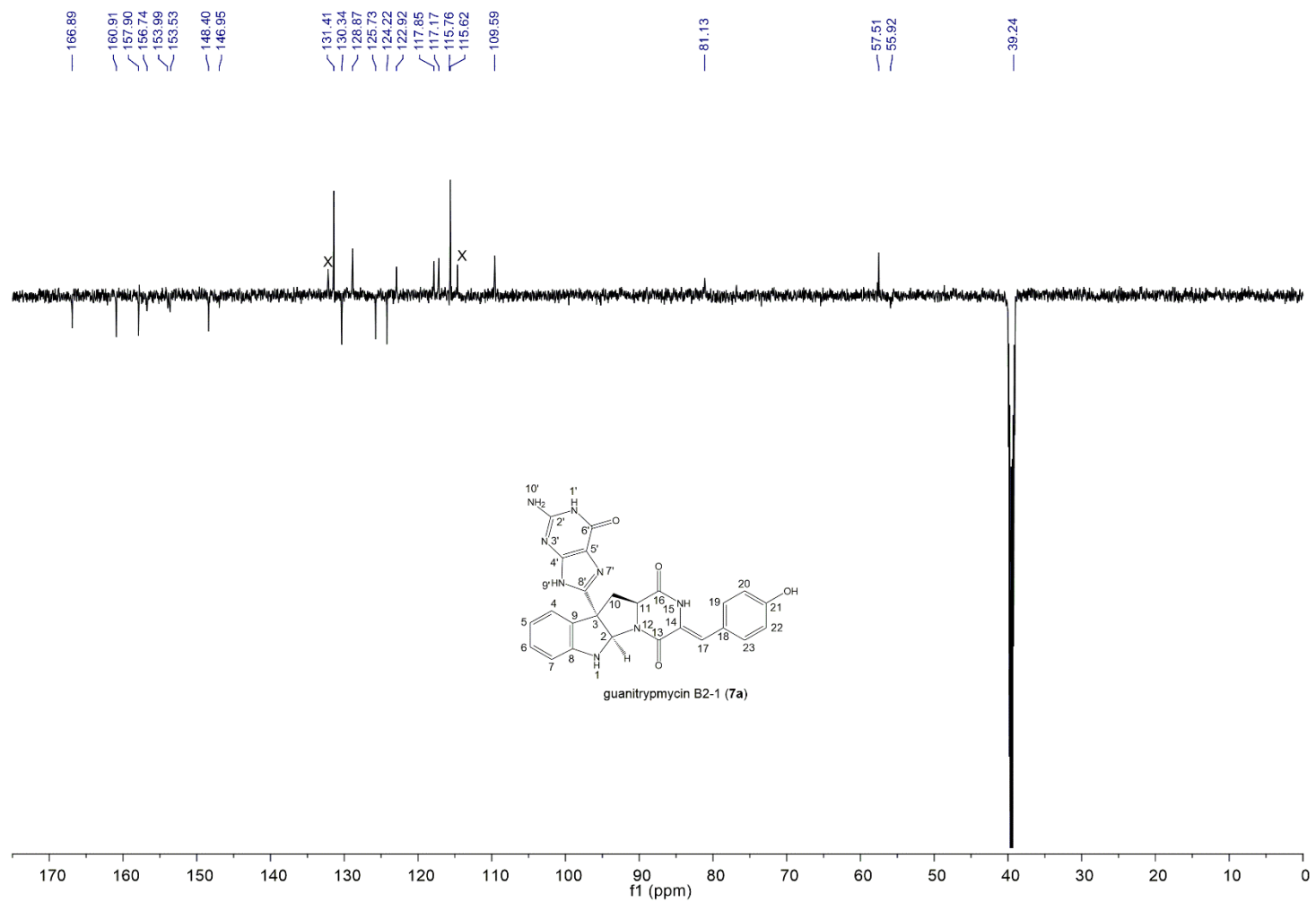


Figure S46. ¹³C APT NMR spectrum of guanitrypmycin B2-1 (7a).

S70

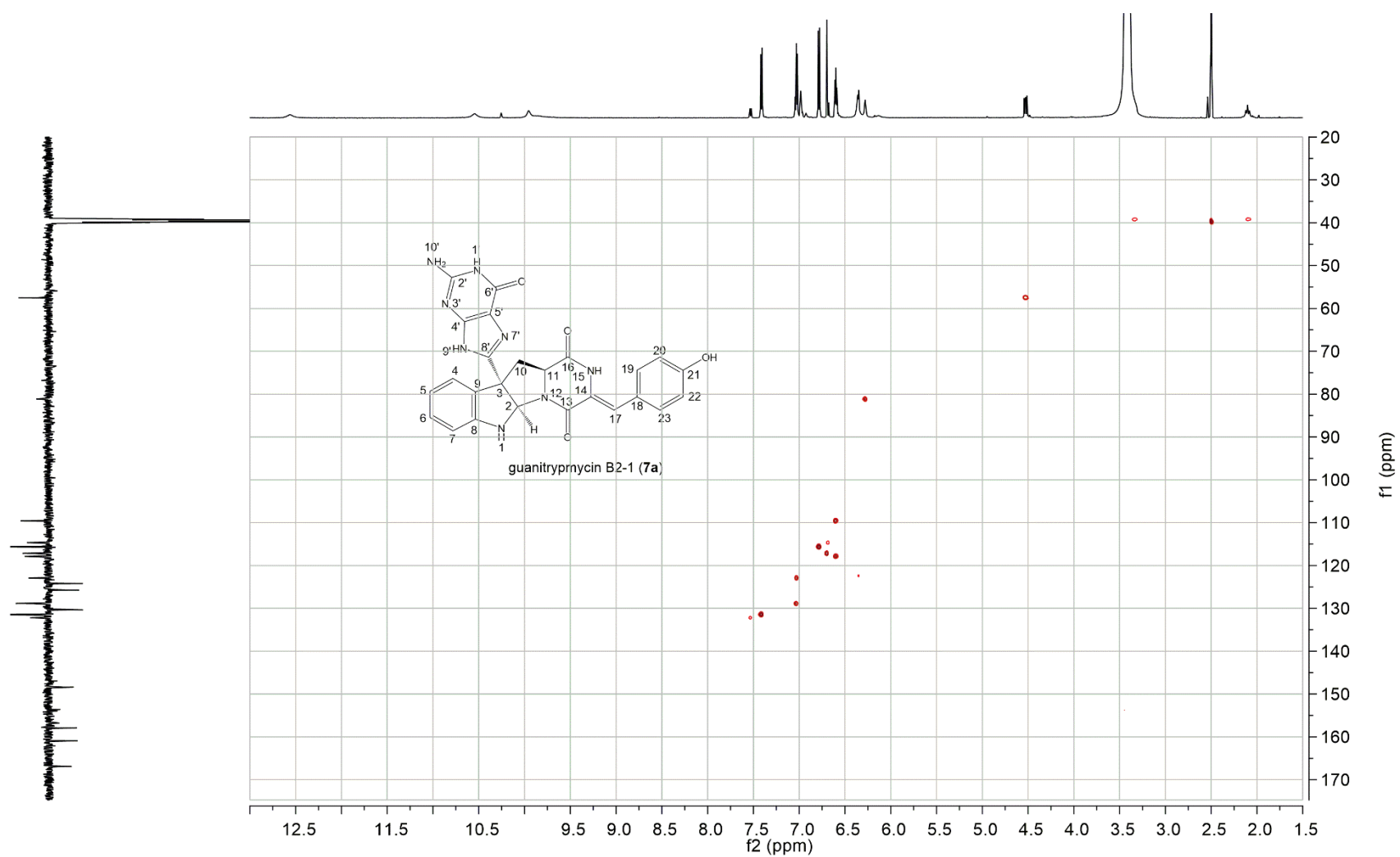


Figure S47. HSQC spectrum of guanitrypyncin B2-1 (7a).

S71

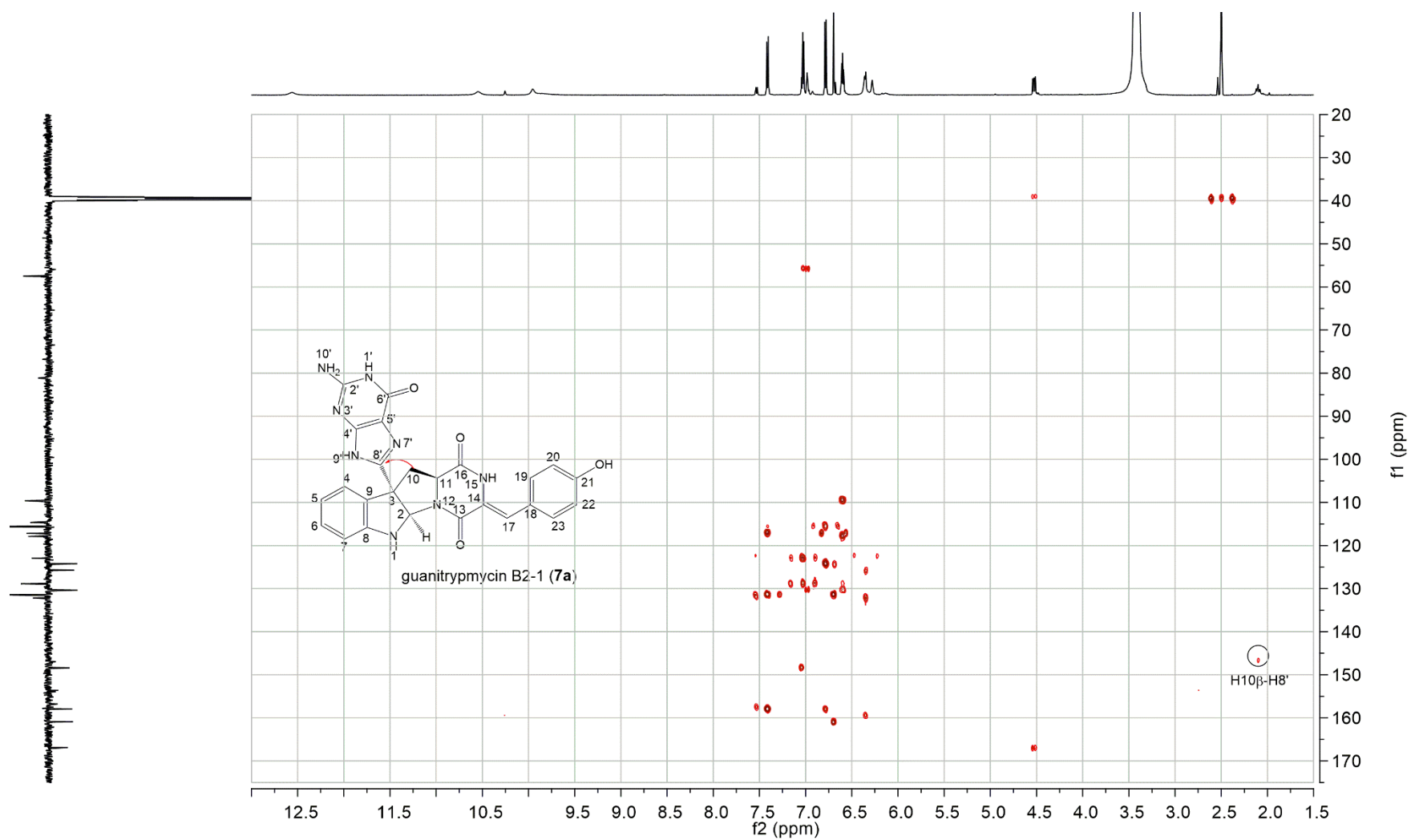


Figure S48. HMBC spectrum of guanitrypmycin B2-1 (**7a**).

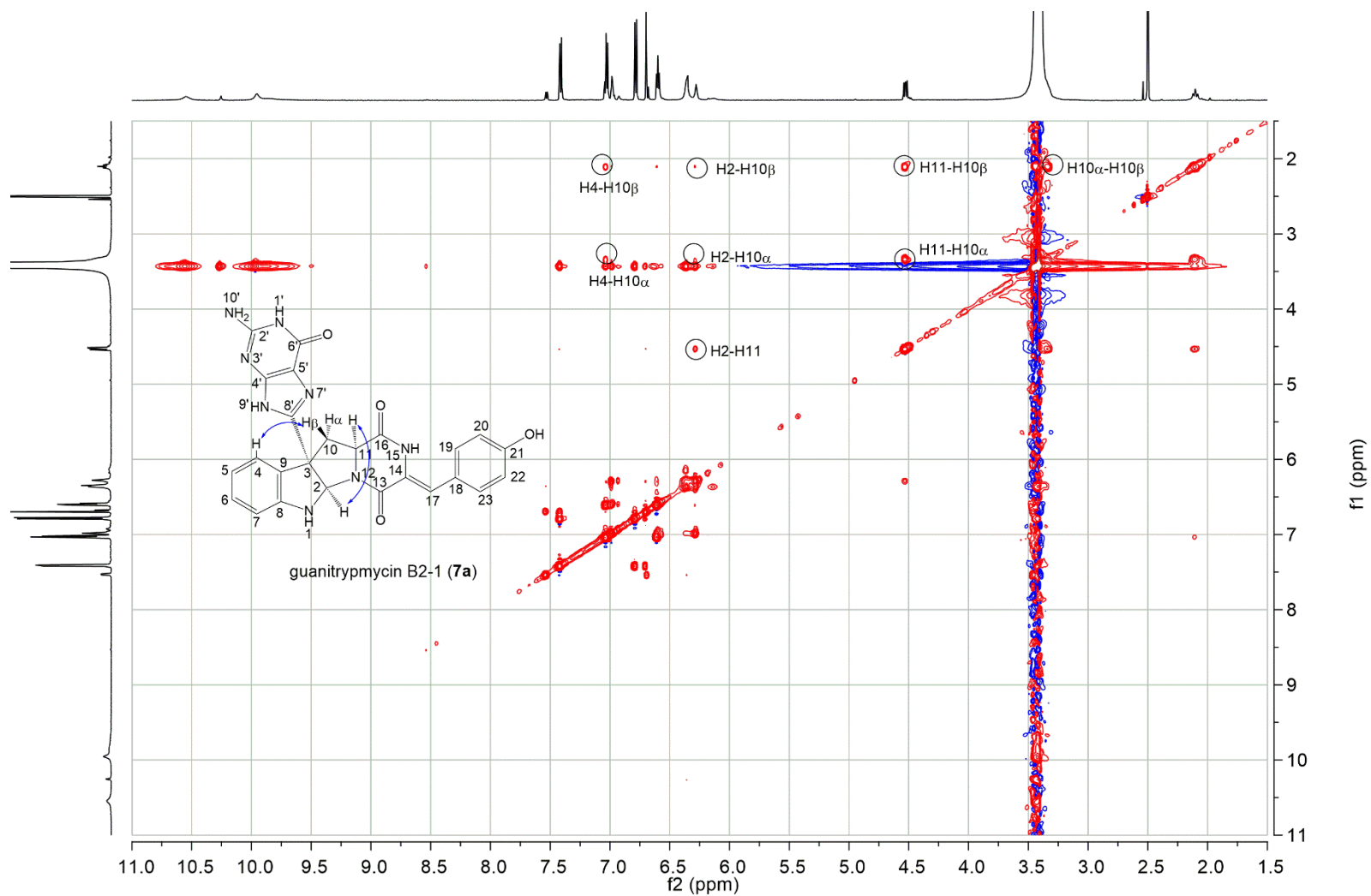


Figure S49. NOESY spectrum of guanitrypmycin B2-1 (7a).

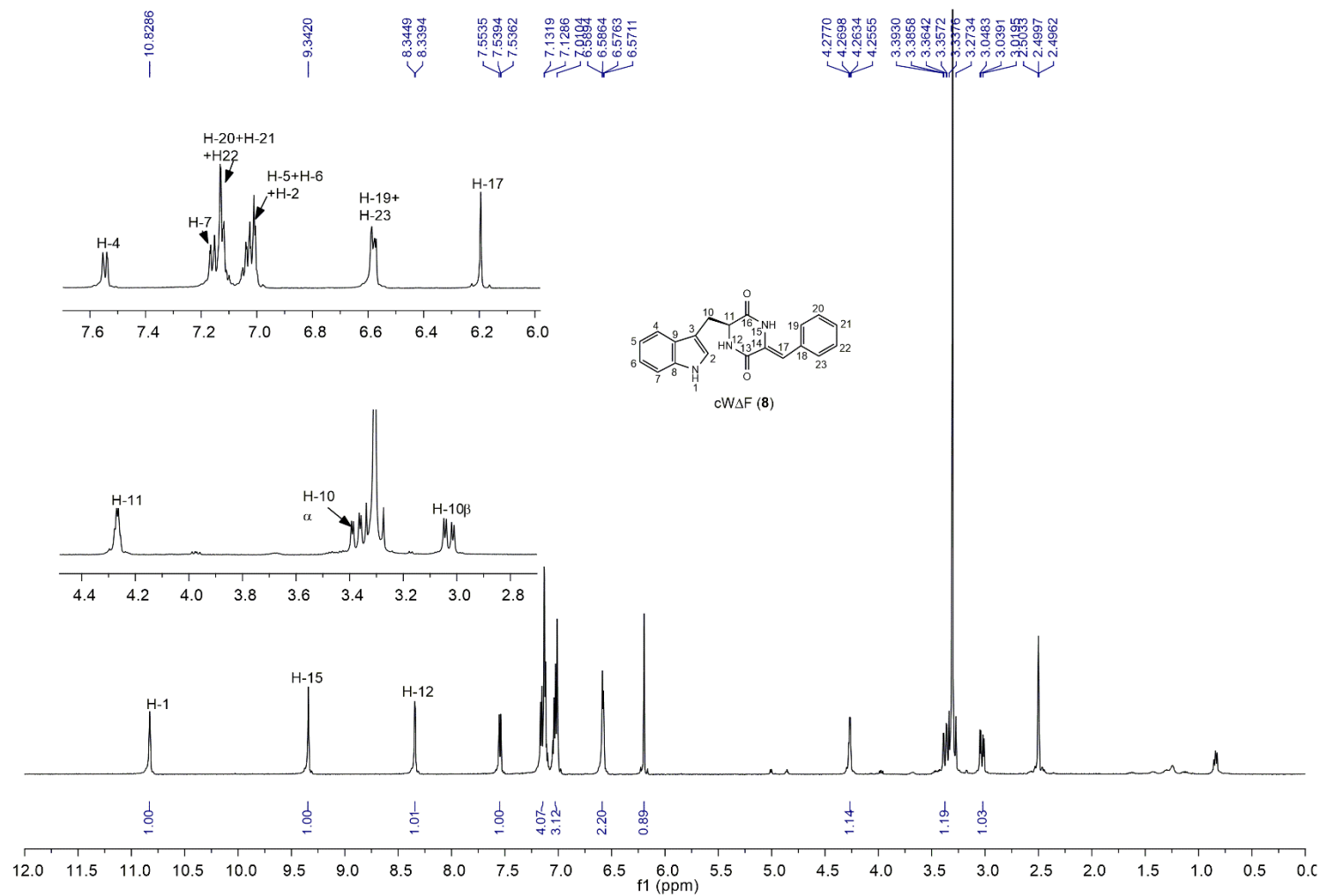


Figure S50. ^1H NMR spectrum of cW Δ F (8).

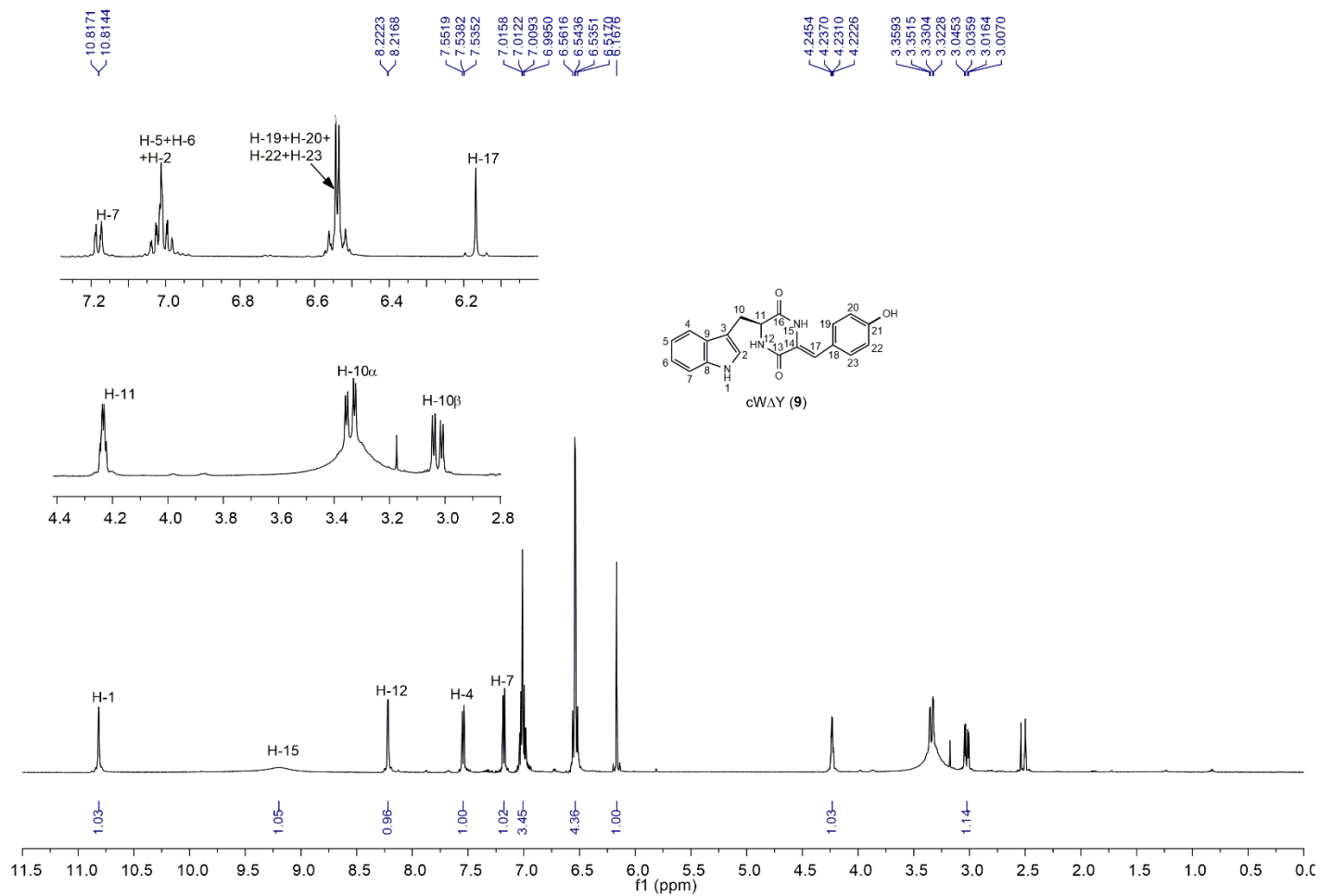


Figure S51. ¹H NMR spectrum of cWΔY (9).

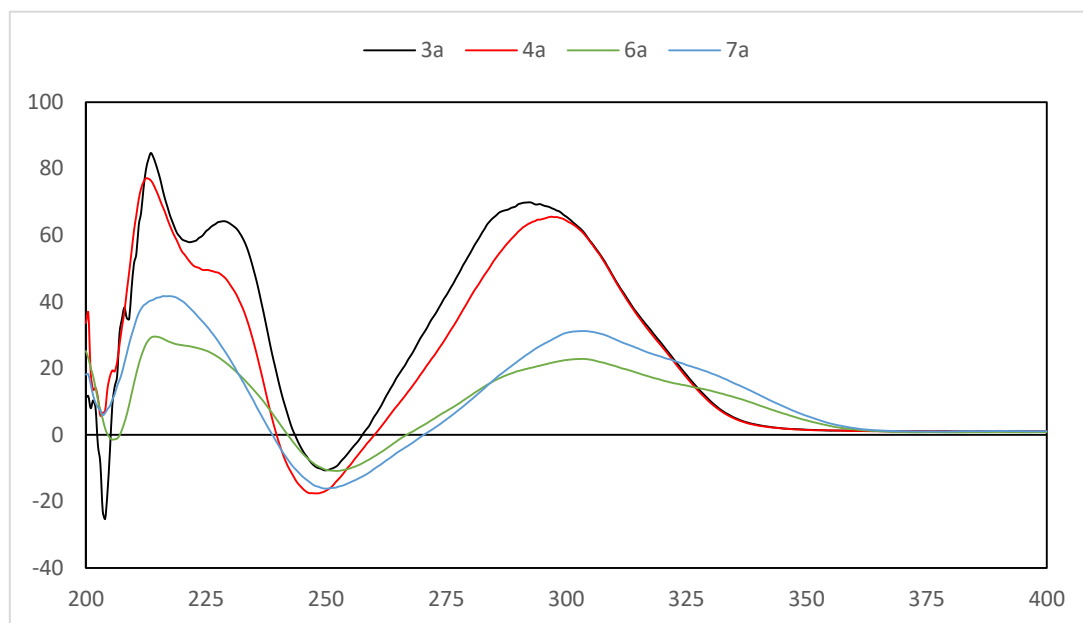


Figure S52. CD spectra of **3a**, **4a**, **6a**, and **7a** in MeOH.

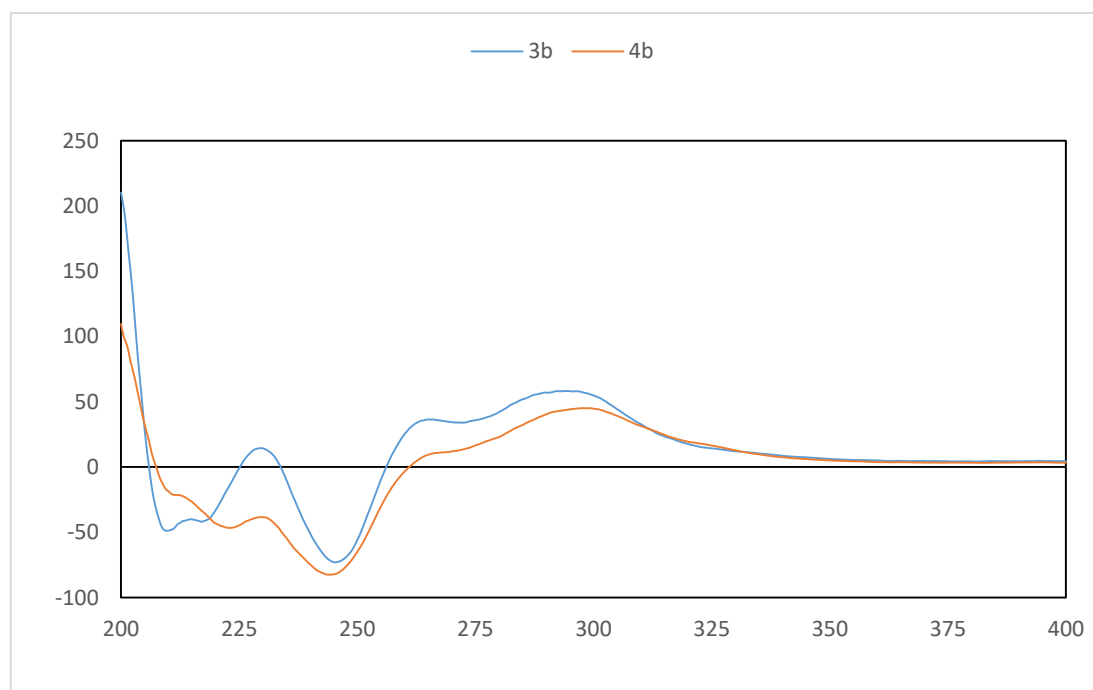


Figure S53. CD spectra of **3b** and **4b** in MeOH.

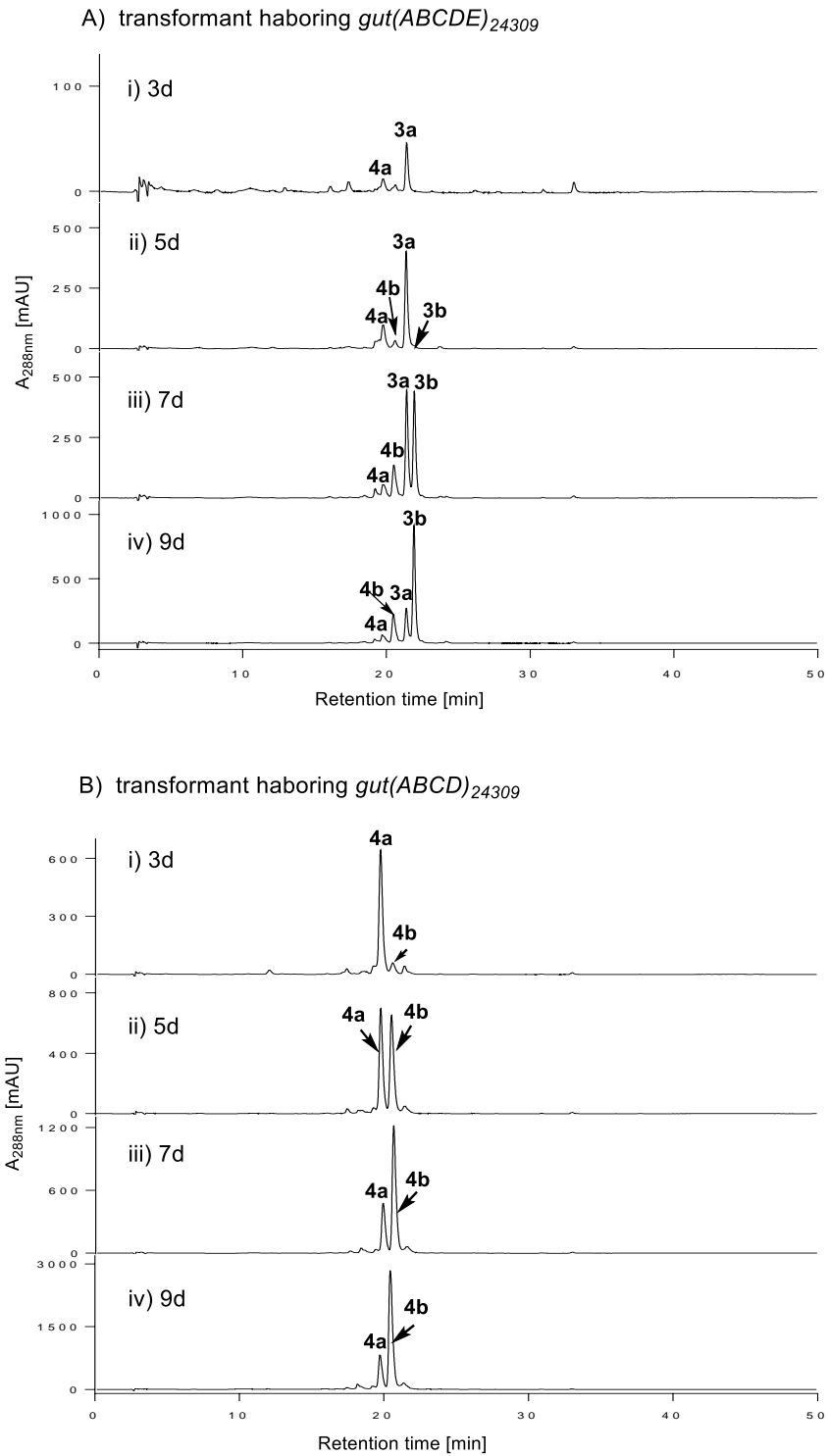


Figure S54. Time-dependent product formation of *S. coelicolor* M1146 transformants cultivated in GYM media at 28°C. UV absorptions at 288 nm are illustrated.

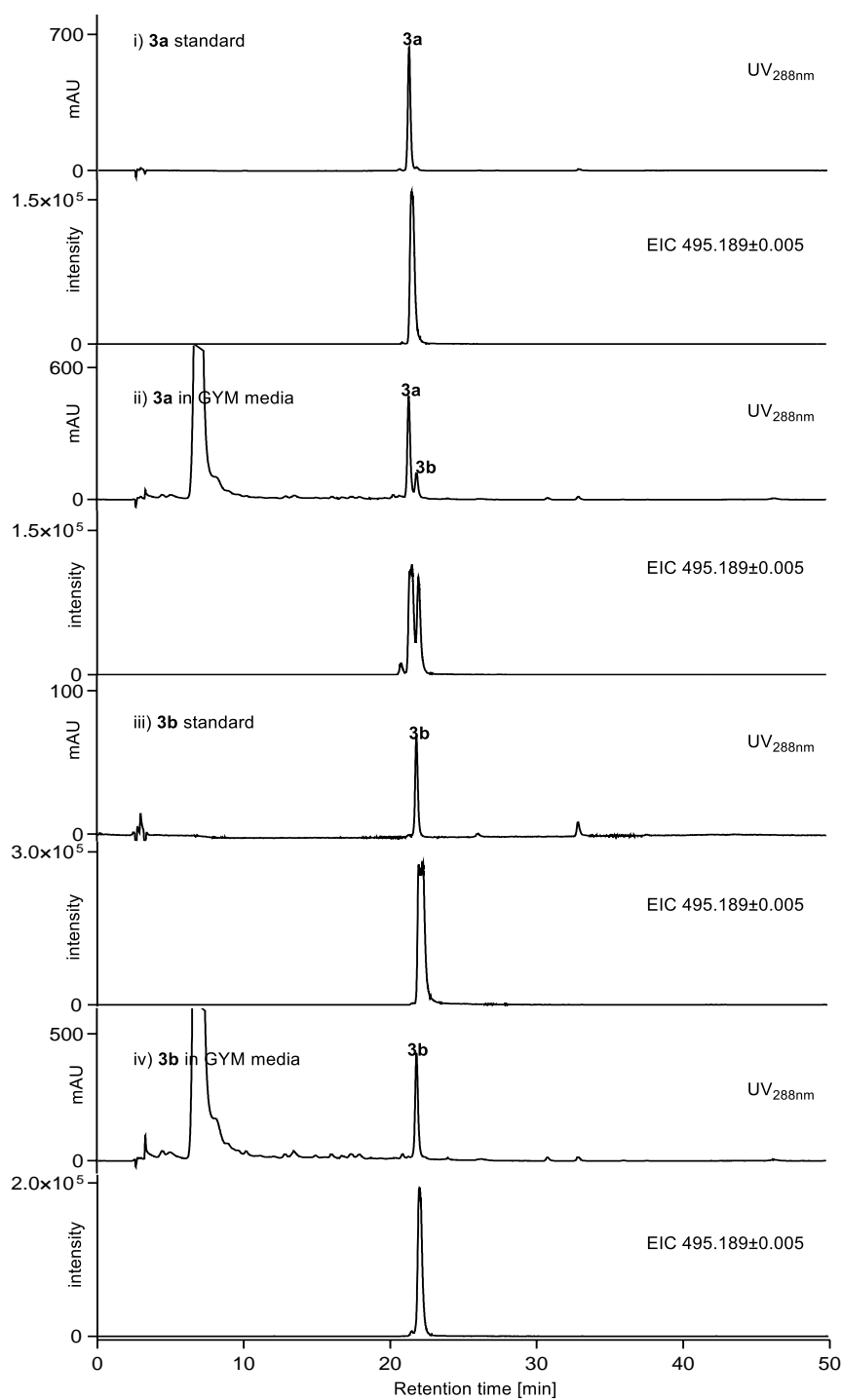


Figure S55. LC-MS analysis for determination of the stability of **3a** and **3b** in GYM media. The solutions were incubated at 28 °C for 8 days.

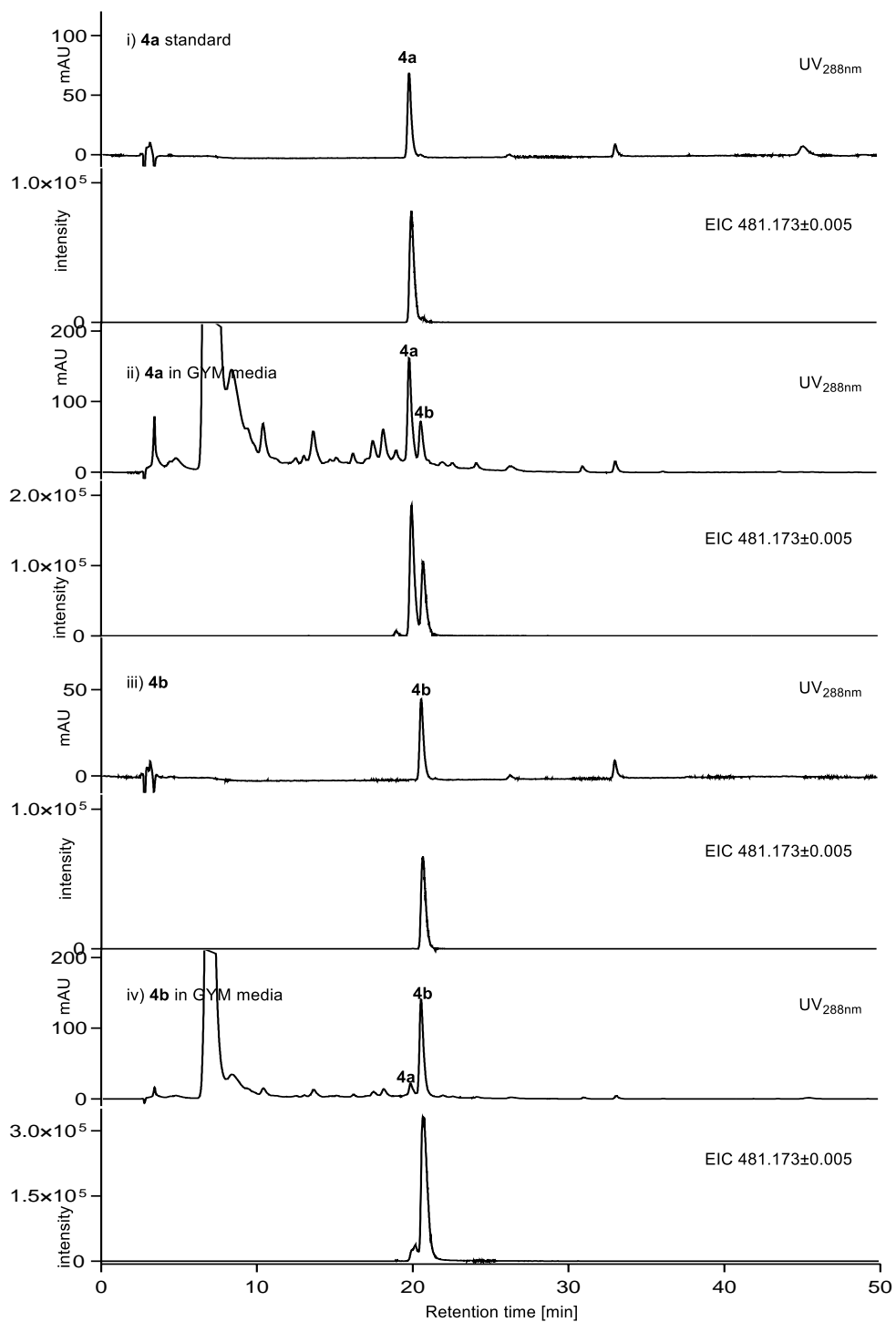


Figure S56. LC-MS analysis for determination of the stability of **4a** and **4b** in GYM media. The solutions were incubated at 28 °C for 8 days.

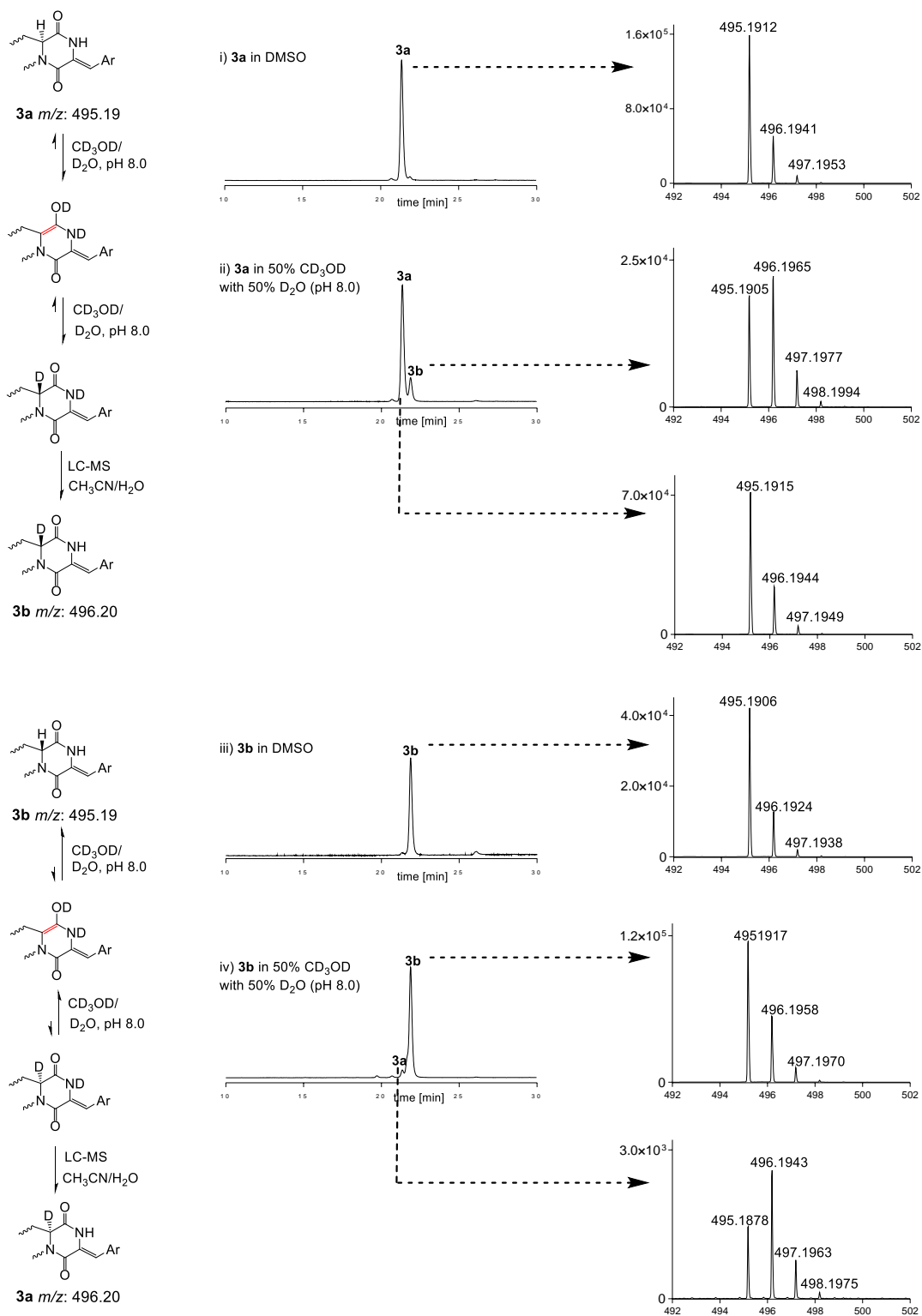


Figure S57. LC-MS analysis of **3a** and **3b** after incubation in CD₃OD/D₂O (1:1, pH 8.0) for 12 h.

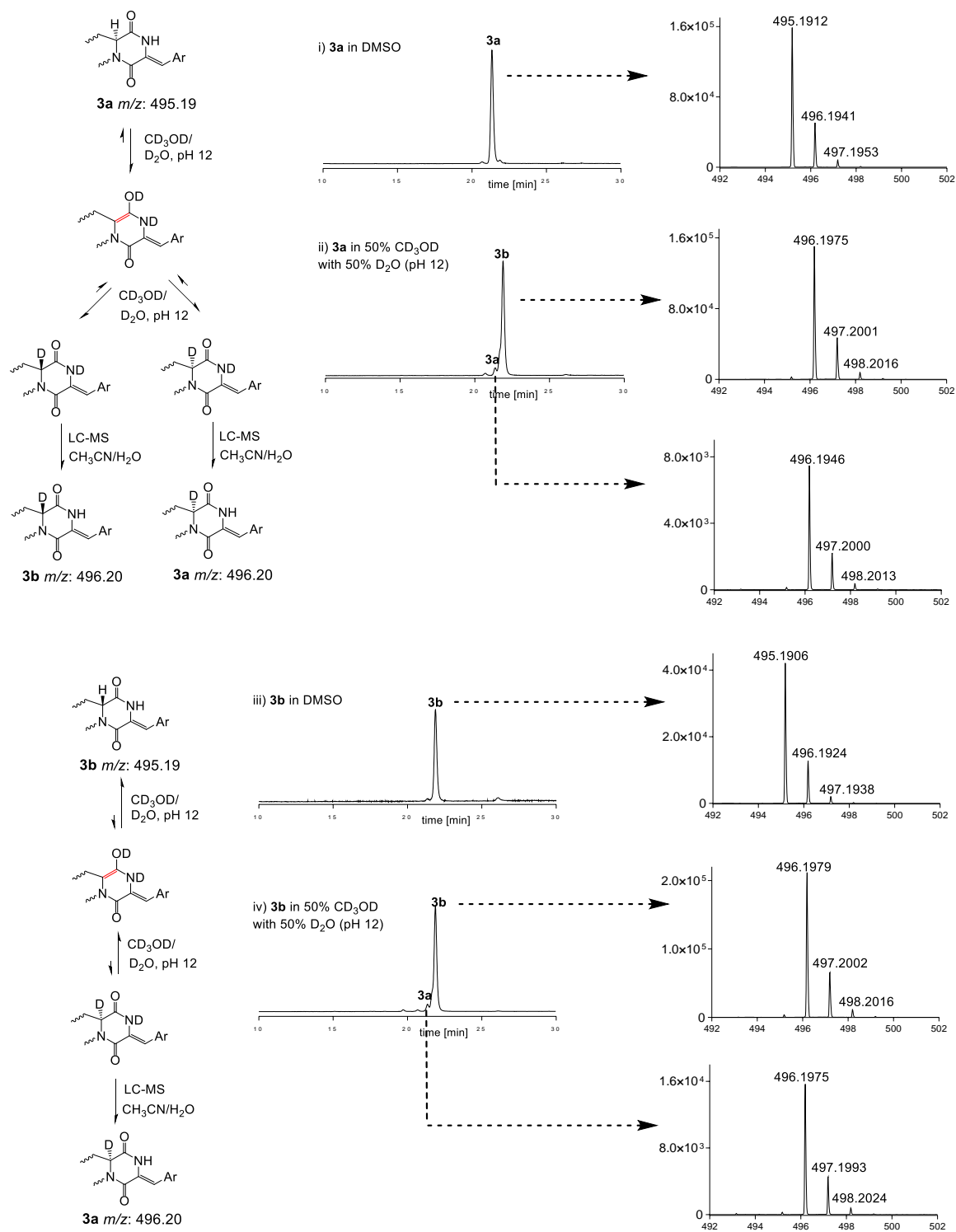


Figure S58. LC-MS analysis of **3a** and **3b** after incubation in CD₃OD/D₂O (1:1) with 0.01M NaOH for 16 h.

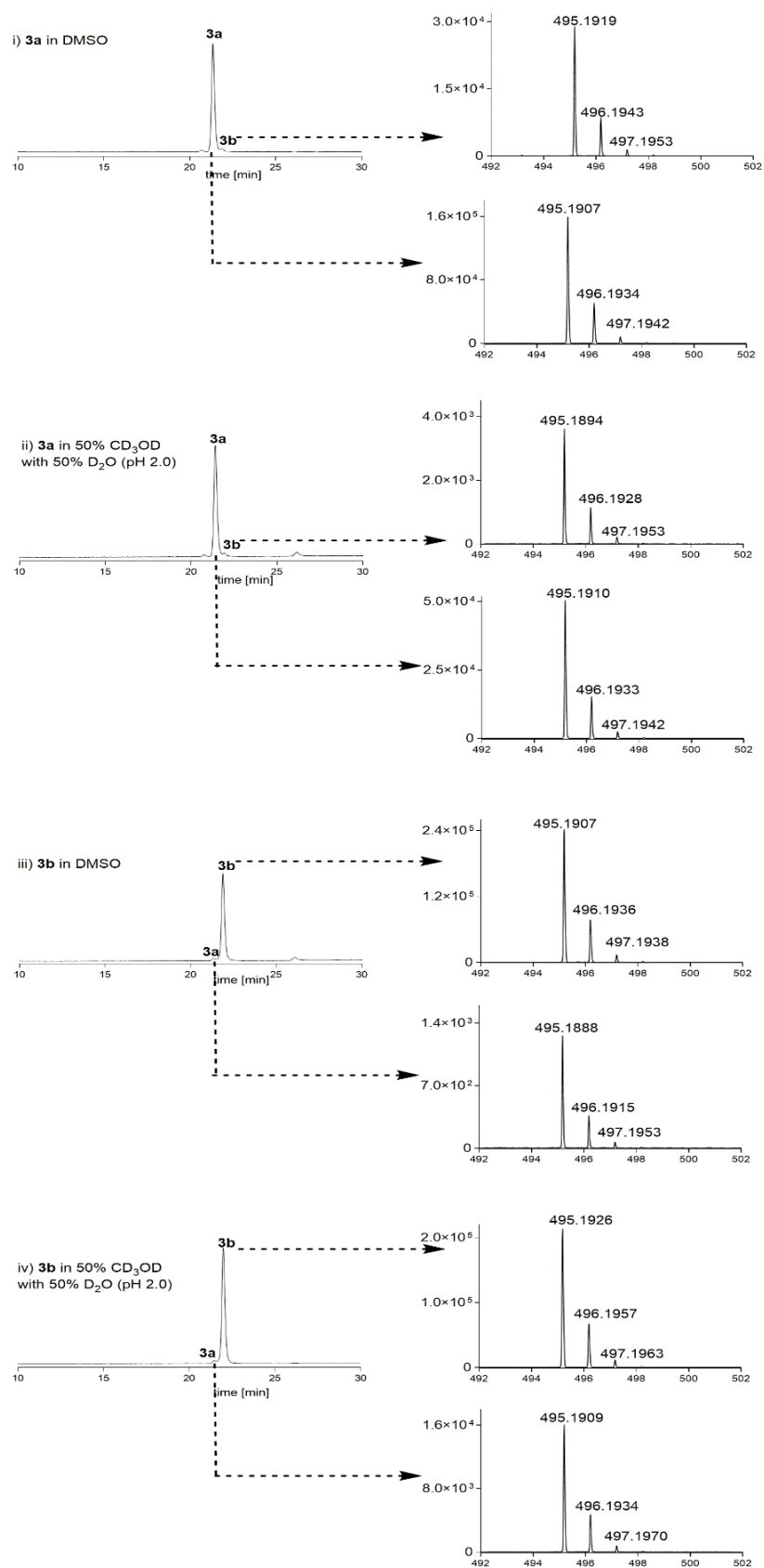


Figure S59. LC-MS analysis of **3a** and **3b** after incubation in CD₃OD/D₂O (1:1) with 0.01M HCl for 12 h.

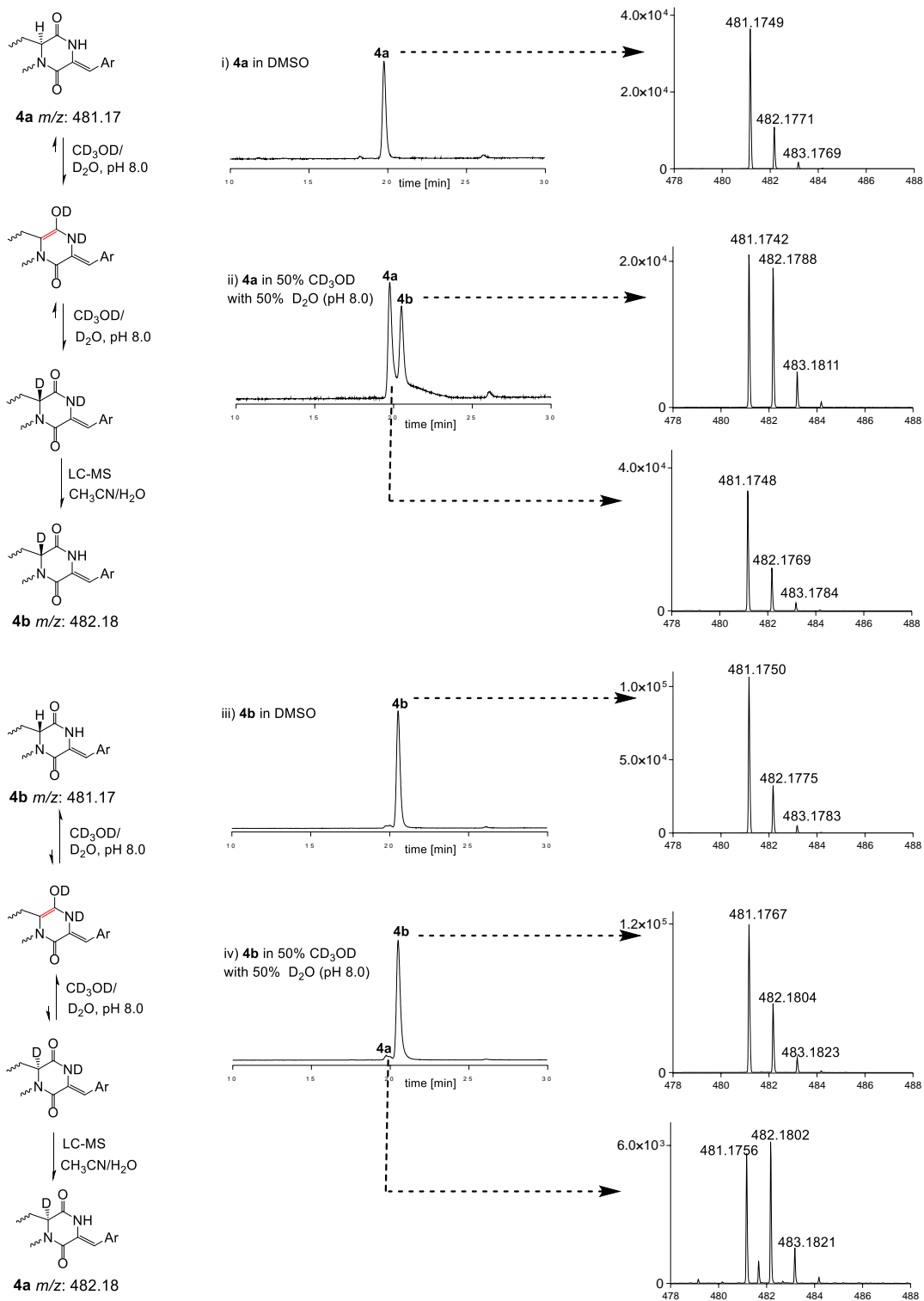


Figure S60. LC-MS analysis of **4a** and **4b** after incubation in CD₃OD/D₂O (1:1, pH 8.0) for 12 h.

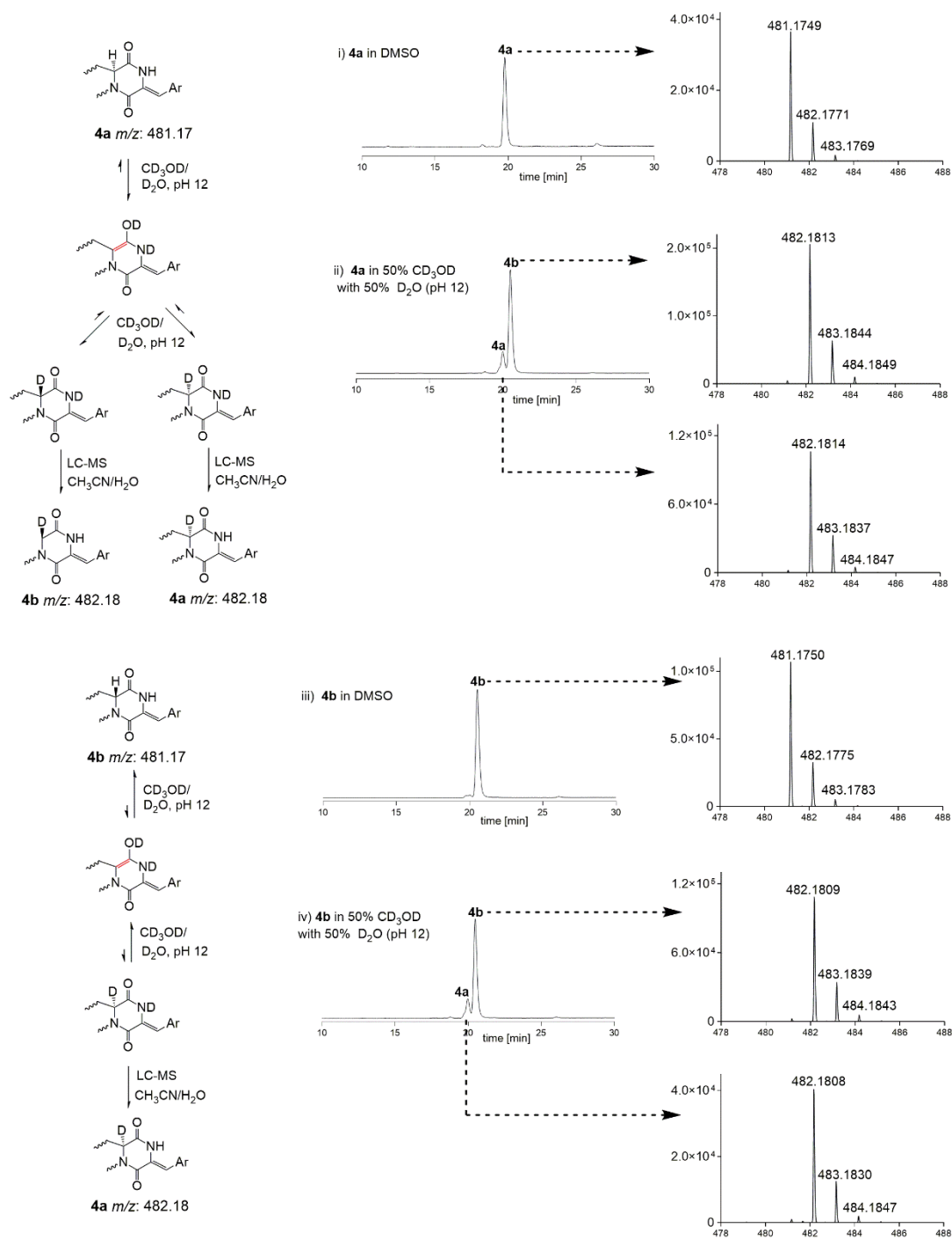


Figure S61. LC-MS analysis of **4a** and **4b** after incubation in CD_3OD/D_2O (1:1) with 0.01M NaOH for 16 h.

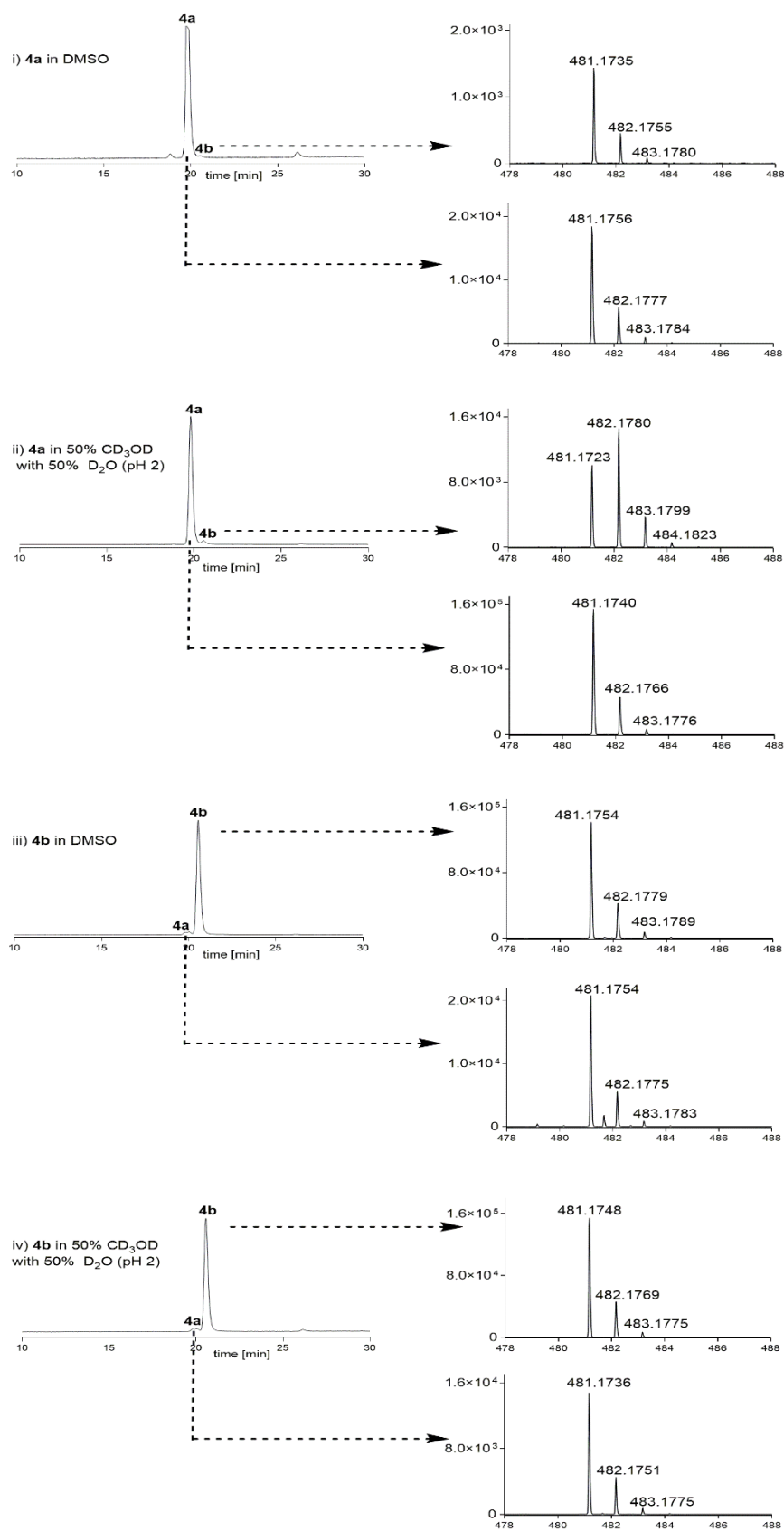


Figure S62. LC-MS analysis of **4a** and **4b** after incubation in CD_3OD/D_2O (1:1) with 0.01M HCl for 12 h.

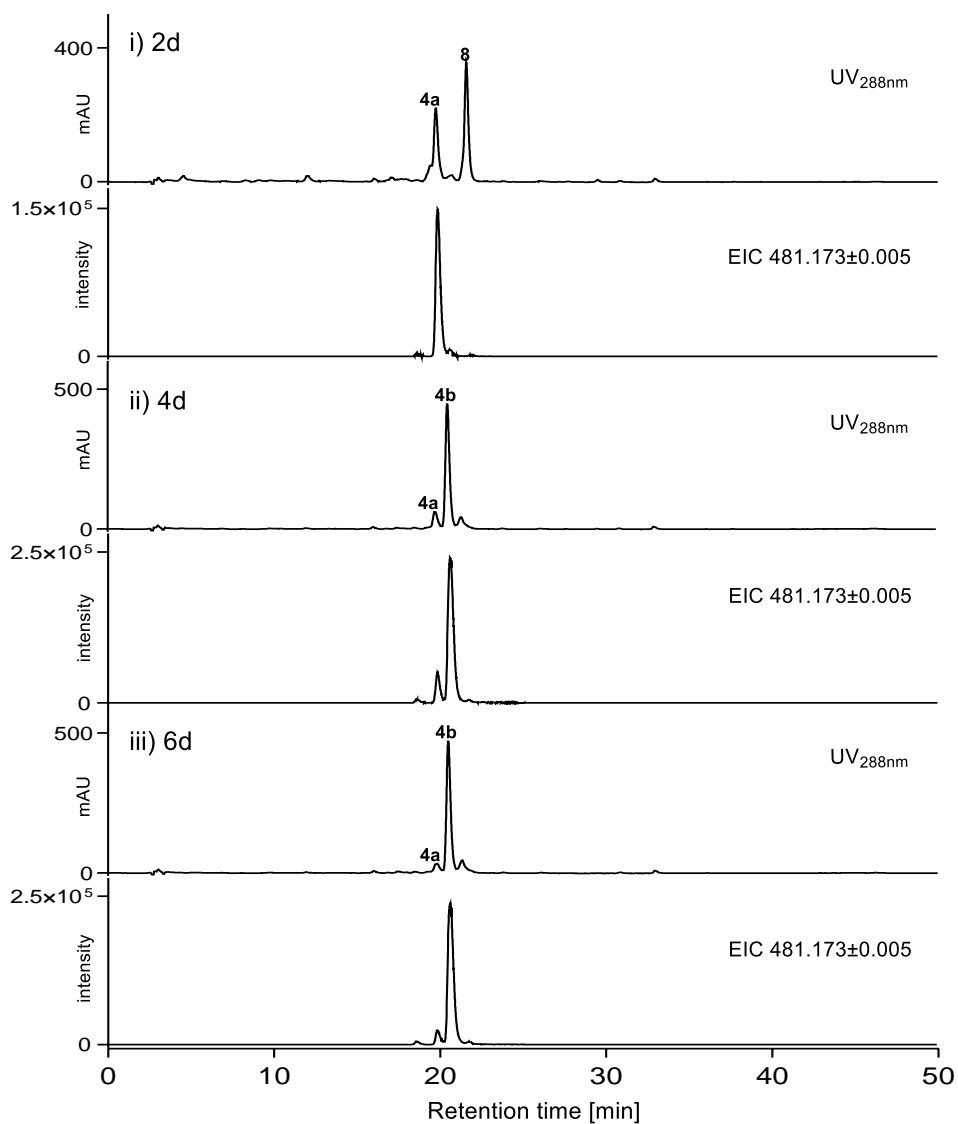


Figure S63. Time-dependent product formation in *S. coelicolor* M1146 harboring *gutD*₂₄₃₀₉ after feeding with cWΔF (8). UV absorptions at 288 nm are illustrated.

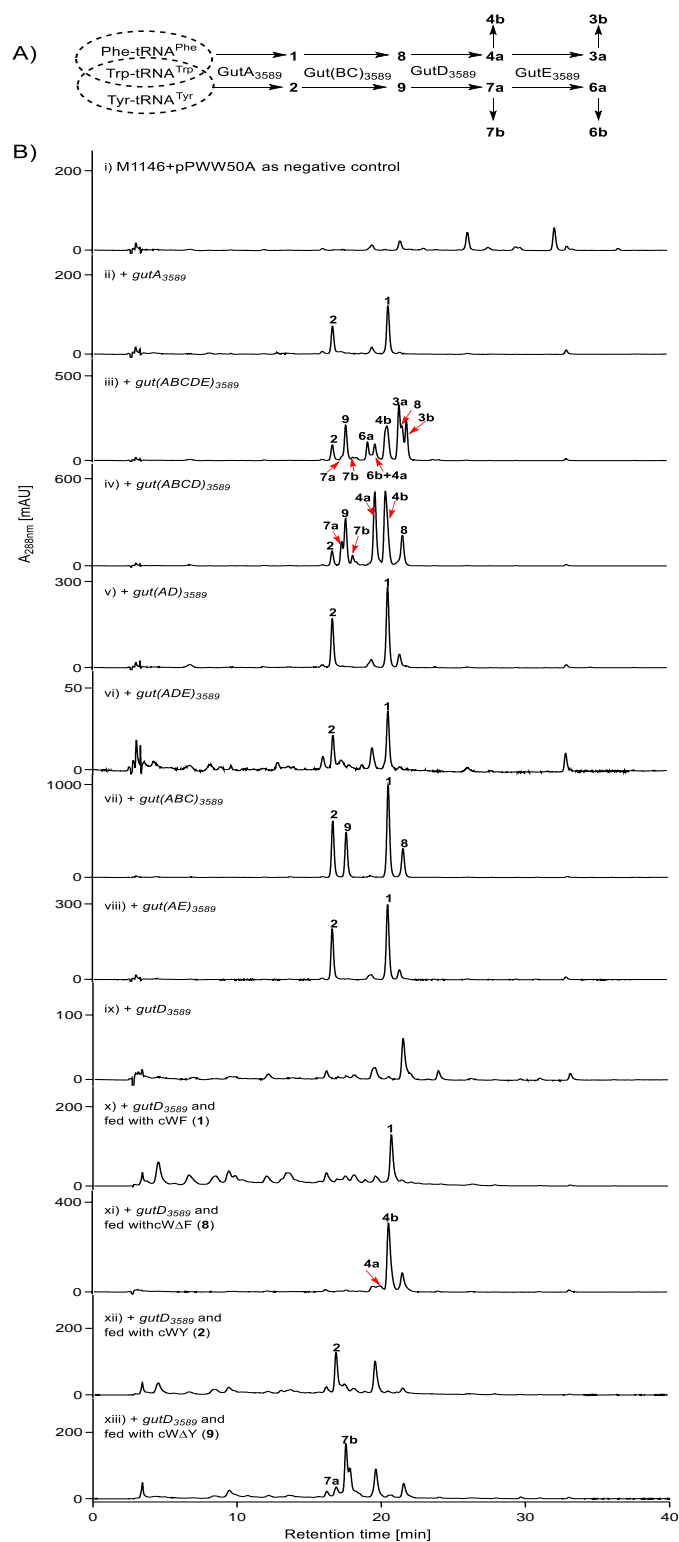


Figure S64. Schematic presentation of the biosynthetic pathway of guanitrypmycins in *S. varsoviensis* (A) and HPLC analysis of the *Str. coelicolor* transformants (B). UV absorptions at 288 nm are illustrated. $[M+H]^+$ ions, with a tolerance range of ± 0.005 , were detected at m/z 334.155 (**1**), 350.150 (**2**), 495.189 (**3a/3b**), 481.173 (**4a/4b**), 511.184 (**6a/6b**), 497.168 (**7a/7b**), 332.139 (**8**), and 348.134 (**9**), respectively.

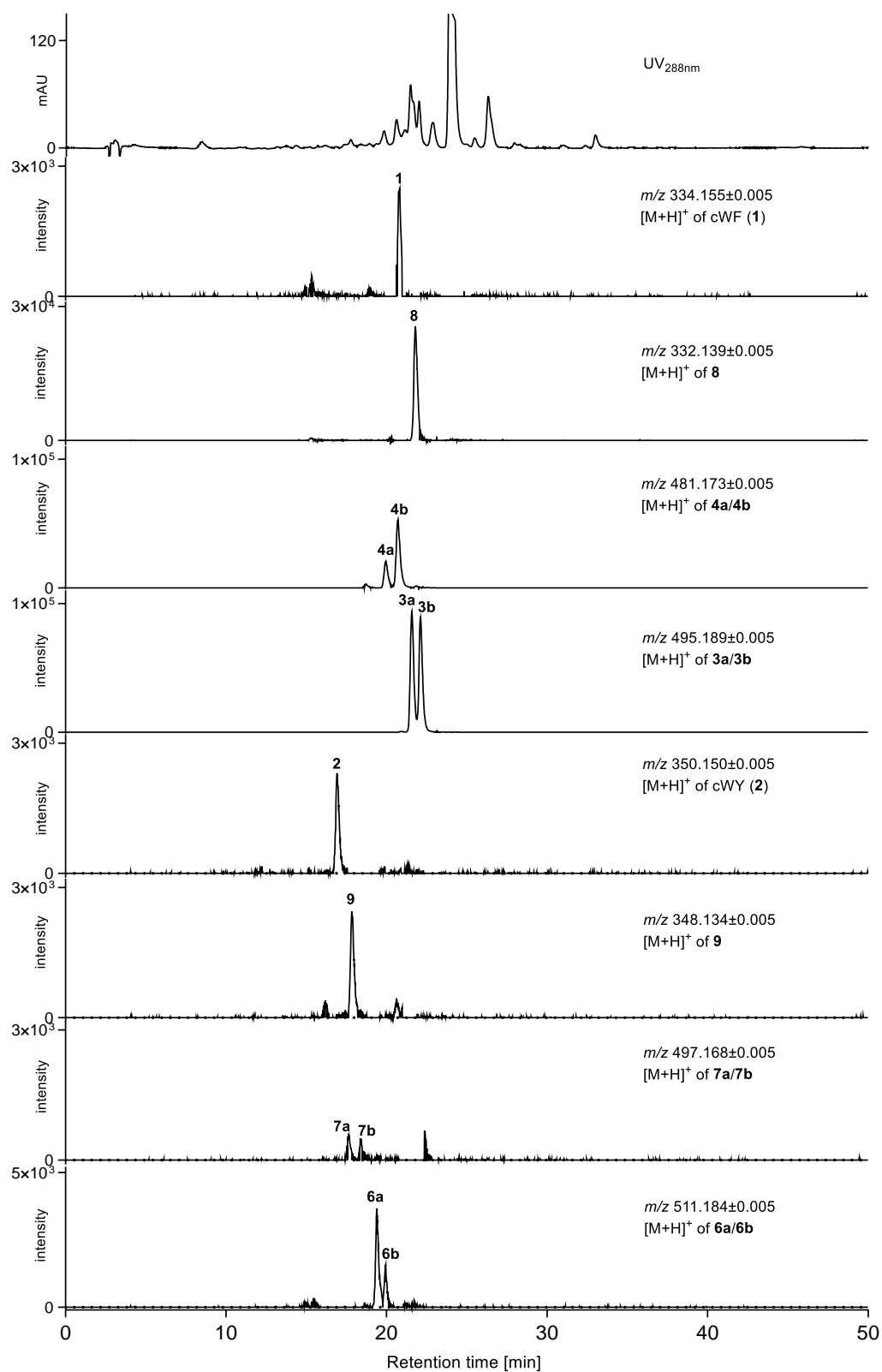


Figure S65. LC-MS analysis of metabolite profile of *S. monomycini* NRRL B-24309.

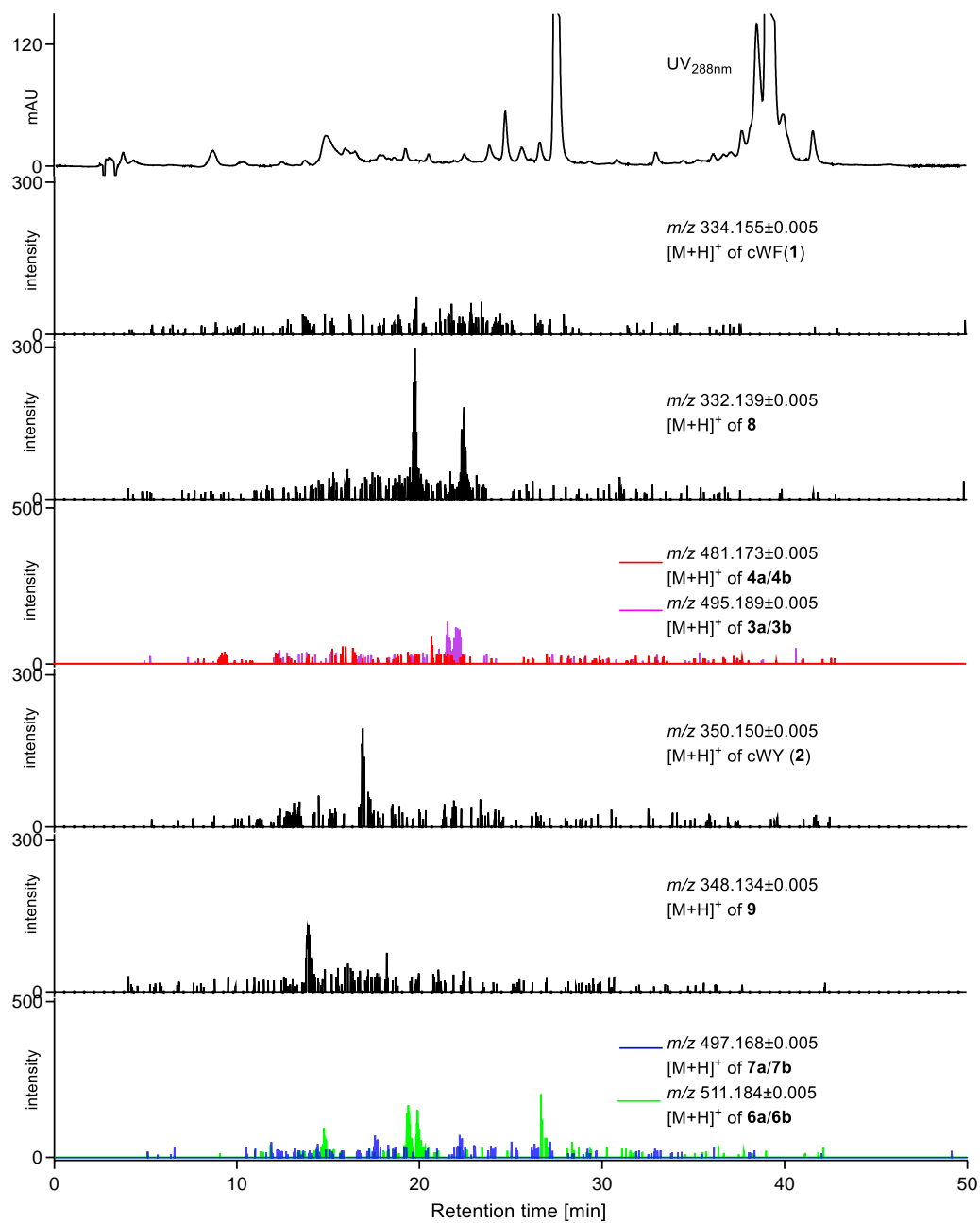


Figure S66. LC-MS analysis of metabolite profile of *S. varsoviensis* NRRL B-3589.

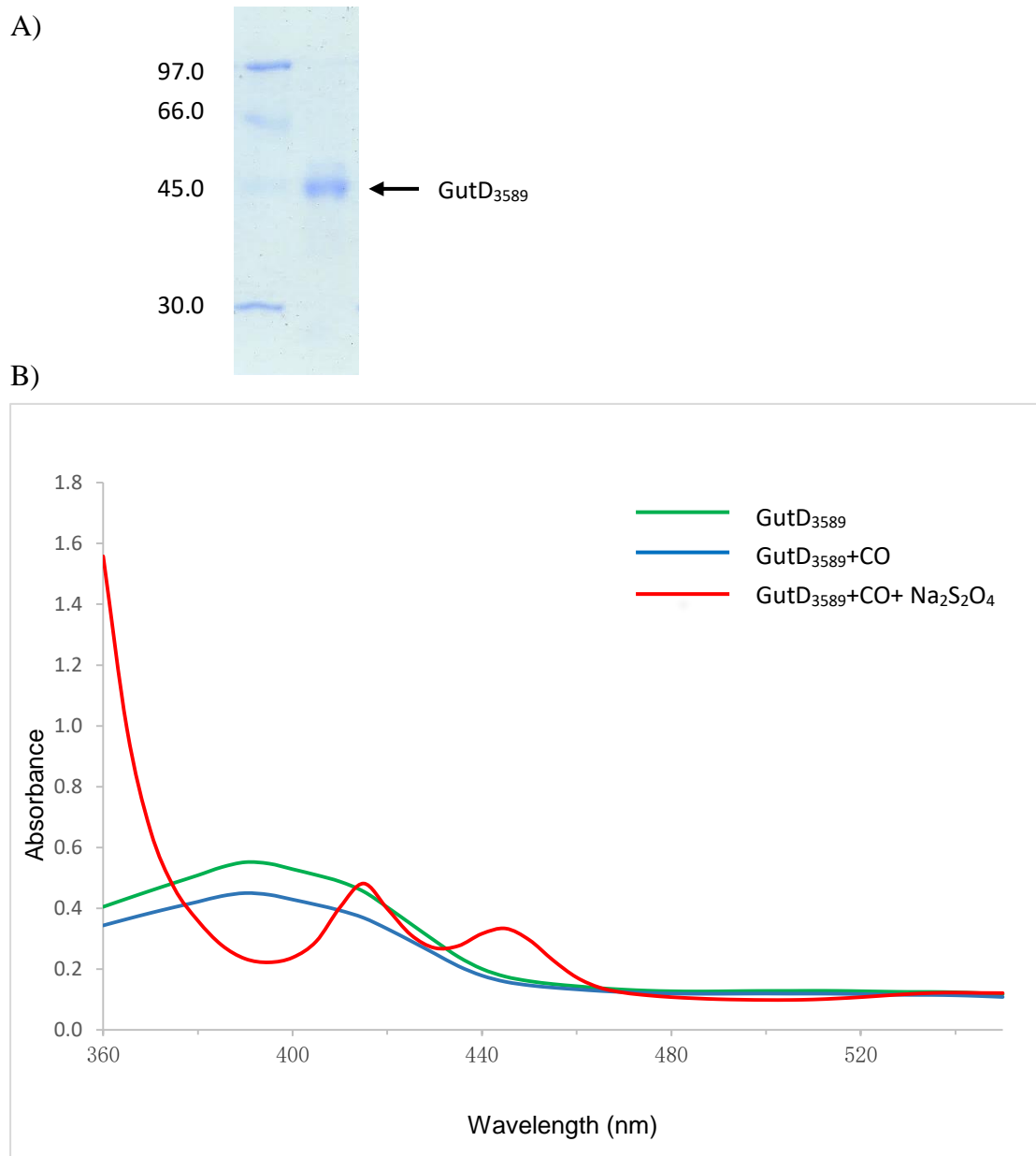


Figure S67. SDS-PAGE analysis of the purified GutD₃₅₈₉ (A) and the absorption spectra for GutD₃₅₈₉ and its ferrous-CO complex (B).

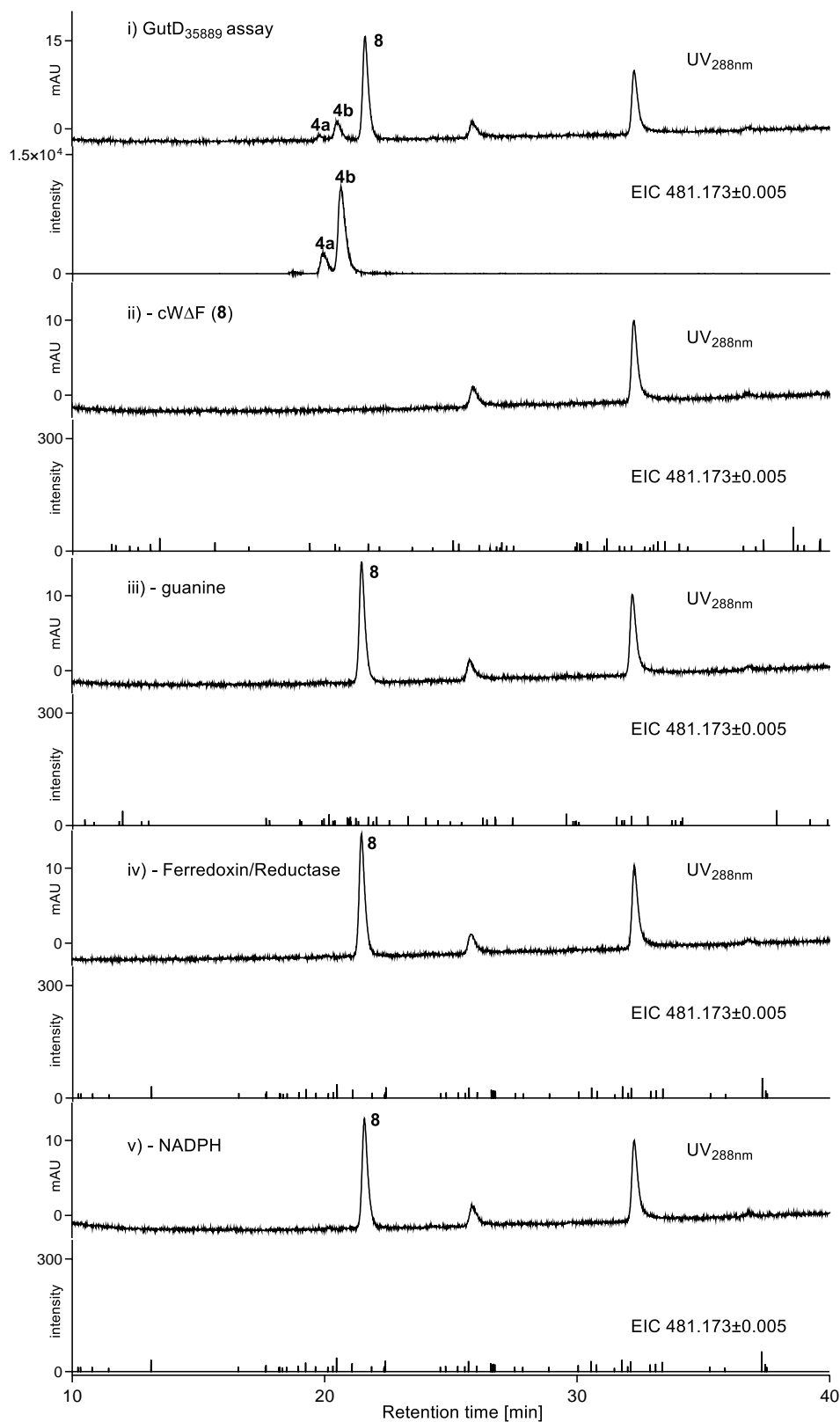


Figure S68. LC-MS analysis of enzyme assays of GutD₃₅₈₉. (i) *in vitro* GutD₃₅₈₉ full assay, (ii) without cWΔF (8), (iii) without guanine, (iv) without ferredoxin and reductase, and (v) without NADPH.

References

- (1) 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.
- (2) Sambrook, J.; Russell, D. W. *Molecular cloning: a laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor: New York, 2001.
- (3) 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.
- (4) 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.* **2018**, *20*, 4921.
- (5) 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.

4.3 Increasing cytochrome P450 enzyme diversity by identification of two distinct cyclodipeptide dimerases



Increasing cytochrome P450 enzyme diversity by identification of two distinct cyclodipeptide dimerases†

 Jing Liu,^a Xiulan Xie^b and Shu-Ming Li^{ib}*^a

 Cite this: *Chem. Commun.*, 2020, 56, 11042

 Received 10th July 2020,
Accepted 7th August 2020

DOI: 10.1039/d0cc04772d

rsc.li/chemcomm

Genome mining revealed the presence of two *cdps-p450* operons in *Saccharopolyspora antimicrobica*. Heterologous expression, biochemical characterisation and structure elucidation proved that the two P450 enzymes catalyse distinct regio- and stereospecific dimerizations of cyclo-(L-Trp-L-Trp), which significantly expands the repertoire of diketopiperazine-tailoring enzymes. TtpB1 connects the monomers via C3–C3', both from the opposite side of H-11/H-11', while TtpB2 is characterised as the first P450 to mainly catalyse the unusual linkage between N1' and C3 from the H-11 side.

Derivatives of cyclodipeptides (CDPs), the smallest cyclic peptides with a representative 2,5-diketopiperazine (DKP) heterocycle, comprise a large class of bioactive molecules.¹ Despite the structural simplicity, their privileged structural core makes them attractive scaffolds for drug discovery and development.^{2–4} In nature CDPs are assembled by either nonribosomal peptide synthetases (NRPSs), mostly in fungi⁵ or cyclodipeptide synthases (CDPSs), usually in bacteria.^{6,7} Different tailoring enzymes are afterwards involved in installing a number of functional groups, thus generating various chemical structures.^{1,8,9} Among these modification enzymes, P450s were proven to catalyse diverse intriguing chemical transformations.^{8,10} With the aim to explore novel modification enzymes, we recently identified seven P450s for the modification of DKPs assembled by CDPSs. These enzymes catalyse unprecedented coupling reactions with nucleobases^{11–13} or cyclodipeptide dimerizations.¹⁴

Dimeric CDPs possess enormous chemical complexity owing to the densely functionalized core and multiple stereogenic centres in their structures.¹⁵ Taking their biological activities

together, dimeric CDPs hold significant promise for medicinal chemistry. In recent years, a large number of dimeric DKP alkaloids have been identified from nature, mainly from fungi of the genera *Aspergillus* and *Penicillium*.³ Meanwhile, great efforts have been made for the chemical synthesis of dimeric DKPs, e.g. epidithiodiketopiperazines,¹⁶ (+)-11,11'-dideoxyverticillin A,¹⁷ (+)-WIN 64821,¹⁸ (–)-ditryptophenamine,¹⁸ and (+)-naseaezines A and B.¹⁹ These synthetic procedures imply usually multiple steps. Protected amino acids were used as reactants and metal salts like CoCl(PPh₃)₃ or AgSbF₆ as coupling reagents.

In contrast to the achievements obtained in isolation and chemical synthesis, only four pathways for the formation of dimeric CDPs are characterised to date, i.e. the NRPS-dependent pathway of (–)-ditryptophenamine in *Aspergillus flavus*²⁰ (Scheme S1, ESI†) and three CDPS-related ones in bacteria. The formation of (–)-naseaezine C in *Streptomyces* sp. CMB-MQ030²¹ as well as (+)-naseaezine A and (–)-aspergilazine A in *Streptomyces* sp. NRRL-S1868¹⁴ is catalysed by P450s via C–C bonds between C-5 or C-6 at the benzene ring of one unit and N1 or C3 of another one (Scheme S1, ESI†).

A huge number of released actinobacteria genome sequences provide a solid basis for the discovery of novel compounds and intriguing enzymes by genome mining.²² To expand the spectrum of CDP modification enzymes, we analysed different bacterial *cdps-p450*-containing clusters by using known proteins as probes. The two clusters comprising one *cdps* (*ttpA1* or *ttpA2*, Fig. S1, ESI†) and one (*ttpB2*) or two P450 genes (*ttpB1* and *ttpC1*, Fig. 1A and Table S1, ESI†) raised our interest, because the two putative P450s TtpB1 and TtpB2, with sequence identity of 71% to each other, are located in the phylogenetic tree near to the known enzymes NasB, NascB and AspB (Fig. S2, ESI†), which catalyse the dimerization of cWP or the connection of cWP with cWA.^{14,21} TtpC1 is located phylogenetically out of this clade. Multiple sequence alignments of the three aforementioned P450 enzymes with other known entries revealed the presence of the conserved motifs for bacterial P450 proteins (Fig. S3, ESI†), i.e. the G³⁴⁷XXXC³⁵¹ (referring the number of EryF²³) motif in the heme-binding

^a 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; Fax: +49-6421-28-22461; Tel: +49-6421-28-25365

^b Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Straße 4, Marburg 35032, Germany

† Electronic supplementary information (ESI) available: The experimental procedures, materials, methods and characterization of the compounds. See DOI: 10.1039/d0cc04772d

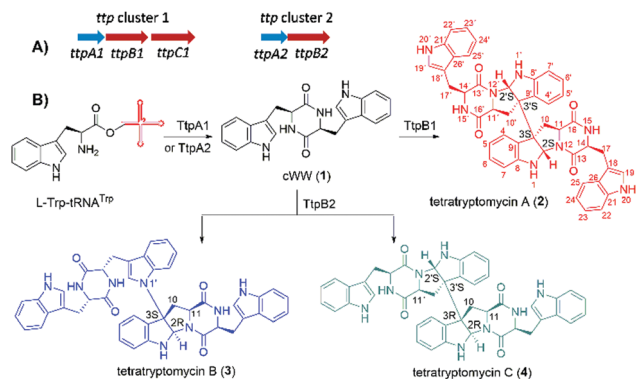


Fig. 1 (A) Genetic organizations of the two *ttp* gene clusters and (B) the biosynthetic pathways of tetratryptomycins.

loop and the highly conserved $A/G^{n-1}-G^n-XX-T^{n+3}$ motif in the long I-helix running over the distal surface of the heme in the P450 scaffold.²³ The moderate sequence identities of approximately 40% between TtpB1/TtpB2 and the known dimerization enzymes indicate different dimerization substrates or/and patterns, which made us curious about their roles in CDP metabolism.

To verify their functions, the candidate genes and gene clusters were heterologously expressed in *Streptomyces albus* J1074.²⁴ Firstly, the sequences of *ttpA1* and *ttpA2* were amplified by PCR from the genomic DNA of DSM 45119, cloned into the expression vector pPWW50A²⁵ and transformed into *S. albus* J1074 (Tables S2 and S3, ESI†). The obtained transformants were then cultivated in modified R5 media at 28 °C for 7 days, extracted with EtOAc and analysed on LC-MS. Compared to J1074 harbouring the pPWW50A vector (Fig. 2A), the sole product peak 1 with the same retention time and $[M + H]^+$ ion at m/z 373.166 ± 0.005 was detected in both transformants (Fig. 2B and J). **1** was identified as cWW by comparison with an authentic standard, confirming that both TtpA1 and TtpA2 function as cWW synthases. Two CDPS copies for the same product have also been reported previously.^{26,27}

After proof of the CDPS function, *ttpA1* was cloned together with *ttpB1* and *ttpC1* into pPWW50A and expressed in J1074. LC-MS analysis of the culture of the *ttp(ABC)1* transformant revealed the presence of one product peak 2 with a $[M + H]^+$ ion at m/z 743.309 ± 0.005 (Fig. 2C), corresponding well to that of a dimeric cWW. Large scale fermentation and isolation afforded analytically pure **2**. Interpretation of the NMR data including ¹H, ¹³C, ¹H–¹H COSY, HSQC and HMBC indicated the presence of two identical cWW units (see ESI† for details; NMR data are given in ESI† and NMR spectra as Fig. S4–S27, ESI†). Detailed inspection of the NMR spectra confirmed **2** to be a homodimer with a C3–C3' bond from the same side (Fig. 1B). This linkage results thereby in the formation of two hexahydropyrrolo[2,3-*b*]indole frameworks. Thus, the signal of H-2/H-2' is up-field shifted from approx. 7.0 to 5.1 ppm. The signal of C-2/C-2' is up-field shifted from approx. 124 to 77 ppm and C-3/C-3' from 109 to 59 ppm. Key correlations from H-2/H-2' to C-3/C-3', from H-10 to C-3' and from H-10' to C-3 were observed in the HMBC spectrum (Fig. S8, ESI†). In its NOESY spectrum, strong NOE

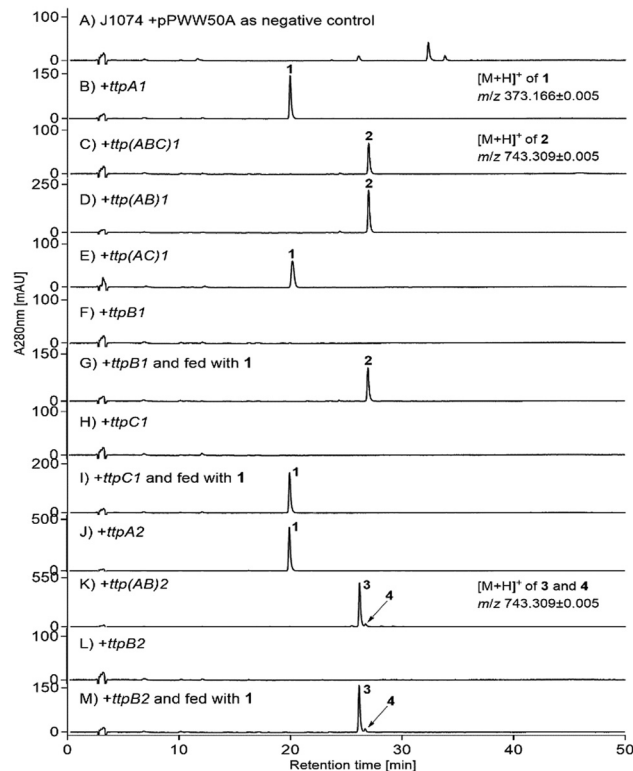


Fig. 2 HPLC analysis of the extracts of *S. albus* transformants.

correlations between H-2 and H-10β suggested that these protons are located at the same side of the five-membered rings (Fig. S9, ESI†). Due to the *cis*-fused ring system and the NMR data featuring two identical cWW units, the configurations of the new stereo centres were determined as 2*S*, 3*S*, 2'*S*, and 3'*S* (Fig. 1B). As its structure features four tryptophan residues, **2** was named tetratryptomycin A.

To verify which P450 is responsible for the formation of tetratryptomycin A (**2**), we expressed *ttp(AB)1* and *ttp(AC)1* separately. LC-MS analysis of the fermentation culture of the *ttp(AB)1* transformant showed the same product profile as that of *ttp(ABC)1* (Fig. 2D), with identical molecular ion, retention time, MS² pattern and UV spectrum. In comparison, only the CDPS product **1** was detected in the *ttp(AC)1* transformant (Fig. 2E). To further confirm that TtpB1, but not TtpC1, is responsible for the dimerization, the coding sequences of *ttpB1* and *ttpC1* were cloned in pPWW50A and expressed in J1074, respectively. Expression of *ttpB1* and *ttpC1* alone did not result in the formation of any new metabolites by comparison with J1074 harbouring pPWW50A (Fig. 2F and H). Feeding **1** into the two-day old culture of the *ttpB1* transformant at the final concentration of 100 μM led to the accumulation of **2** (Fig. 2G), whereas no consumption of **1** was observed in the *ttpC1* transformant (Fig. 2I). This proved that TtpC1 is not involved in the metabolism of **1** and indicated that the *ttp1* cluster comprises just *ttpA1* and *ttpB1*. The second duplicate P450 gene without any function has also been reported in the biosynthetic gene cluster of venezuelaene.²⁸ Taking all of the above results together, TtpB1 functions as a regio- and stereospecific C3–C3'

dimerase, generating a dimeric DKP connecting from the opposite side of H-11/H-11' of the cWW residues.

In analogy, *ttpA2* and *ttpB2* from the second gene cluster were also cloned into pPWW50A and expressed in J1074. LC-MS analysis of the transformant harbouring *ttp(AB)2* revealed the presence of two new peaks 3 and 4 in a ratio of 20:1 with the same $[M + H]^+$ ions at m/z 743.309 \pm 0.005 (Fig. 2K), also indicating the formation of cWW dimers. For structure elucidation, 3 and 4 were isolated and subjected to NMR analysis. Detailed interpretation of the NMR spectra of 3 revealed a C3–N1' connection between the two cWW moieties. Key correlation from H-2' to C-3 was clearly observed in its HMBC spectrum (Fig. S14, ESI[†]). As expected for the formed hexahydropyrrolo[2,3-*b*]indole framework, the signals of H-2, C-2 and C-3 were shifted from approx. 7.0, 124 and 109 ppm to 5.8, 81 and 73 ppm, respectively. In the NOESY spectrum of 3 (Fig. S15, ESI[†]), correlations of H-2' at the indole ring of one cWW unit with H-10 α and H-11 as well as H-2 and H-11 of another monomer were observed, suggesting that these protons are located on the same side. In comparison to that of 2, the CD spectrum of 3 showed two opposite Cotton effects at 245–250 and 300–310 nm (Fig. S28, ESI[†]), supporting the suggested configuration. Therefore, 3 carries a chirality of (2*R*, 3*S*) at the newly formed stereo centres and was named as tetra tryptomycin B (Fig. 1B).

For structure elucidation, the NMR spectra of the minor product 4 were obtained in both DMSO-*d*₆ and CD₃CN. Inspection of its ¹H and ¹³C NMR spectra in DMSO-*d*₆ at 300 K revealed the presence of two very similar cWW moieties with a hexahydropyrrolo[2,3-*b*]indole framework each, indicating a C3–C3' connection of cWW moieties and the presence of a stereoisomer of 2. The proton signals overlapped strongly with each other and the signals for aromatic protons were mostly detected with poor resolution. Nevertheless, key correlation from H-2 of one to C-3' of another cWW unit can be clearly detected in the HMBC spectrum (Fig. S20, ESI[†]). The configurations of 4 at the newly formed stereo centres were proven to be (2*R*, 3*R*, 2'*S*, and 3'*S*) by detection of the key NOE correlations between H-2 and H-11, H-2 and H-10 α , H-2' and H-10' β and H-2 and H-2' (Fig. S21, ESI[†]). 4 was named herewith tetra tryptomycin C. Obtaining the ¹H NMR spectrum of 4 in CD₃CN at 310 K led to a much better resolution, which fulfilled a full signal assignment and verified the suggested structure (Fig. 1B). For further proof of the dimerization reaction catalysed by TtpB2, its coding sequence was cloned into pPWW50A and expressed in J1074. LC-MS analysis proved that neither 1 nor other products were detected in the *ttpB2* transformant (Fig. 2L). As expected, feeding 1 to the transformant led to the detection of tetra tryptomycins B and C in a similar ratio (20:1) as in the *ttp(AB)2* transformant (Fig. 2M). This unequivocally proved TtpB2 as the second cWW dimerase from *S. antimicrobica*.

To further confirm the function of the two P450s *in vitro*, we cloned and expressed *ttpB1* and *ttpB2* in *Escherichia coli*. Unfortunately, only TtpB1 was obtained as a soluble protein and used for enzyme assays (Fig. S29, ESI[†]). With the replicative vector pPWW50A, TtpB2 can be successfully overproduced in *S. albus* J1074 and purified as an *N*-(His)₁₀-fused protein to near

homogeneity, but only with a very low yield (0.05 mg L⁻¹ culture, Fig. S29, ESI[†]). The purified recombinant TtpB1 is brown in colour and shows an absorption shift from 420 to 450 nm after treatment with CO and Na₂S₂O₄, indicating the presence of an active cytochrome P450 enzyme (Fig. S30, ESI[†]). Incubation of TtpB1 with 1 in the presence of the commercial spinach ferredoxin, ferredoxin-NADP⁺ reductase and NADPH led to the clear detection of 2. Nearly 20% conversion of 1 (500 μ M) was achieved after incubation with 7 μ M TtpB1 at 30 °C for 30 min (Fig. S31A, ESI[†]). The reaction of TtpB1 followed the Michaelis–Menten kinetics (Fig. S32, ESI[†]). The *K*_m and *k*_{cat} values were determined to be 50.8 μ M and 0.26 min⁻¹, respectively. Due to the low protein concentration, we incubated 1.1 μ M TtpB2 with 1 at 50 μ M for 12 h. 25% of 1 was converted under this condition, mainly to 3. The product 4 was only detected in the extracted ion chromatogram by LC-MS analysis (Fig. S31B, ESI[†]). In comparison, approx. 95% of 1 was consumed by 5 μ M TtpB1 under the same conditions (Fig. S31A, ESI[†]). The very low concentration of TtpB2 solution prevented us from performing the CO difference experiment and the determination of the kinetic parameters. No consumption of 1 was detected with boiled TtpB1 and TtpB2 and without redox partners or cofactors. Incubation of the recombinant TtpB1 or TtpB2 with other CDPs including cWY, cWF, cWP, cWL, cWA, cWG, cYY and cFF did not lead to any product formation (data not shown), proving the high substrate specificity of both P450s. Absorption shift from 420 to 390 nm was only detected in the UV-Visible spectrum of TtpB1 with cWW, but not in those with other CDPs mentioned above due to the absence of enzyme-substrate binding events (Fig. S33, ESI[†]).

Taking the results together, the P450 TtpB1 catalyses a regio- and stereoselective cWW dimerization, generating the C3–C3' homodimer 2. TtpB2 is responsible for the formation of the dimer 3 *via* a C–N bond as a major product and another C3–C3' homodimer 4 as a minor product. It should be mentioned that both the cWW units in 2 have the same configuration at the newly formed stereo centres, while the opposite configuration was observed for one of the cWW units at the corresponding positions in 4. Recently, Tian *et al.* proposed that the formation of the DKP dimer nasezeazine C by NascB implies a radical-mediated intramolecular cyclization and intermolecular addition reactions.²¹ Similarly, we postulated that the DKP coupling in the biosynthesis of tetra tryptomycins begins with the formation of a N1[•] radical (Fig. 3). Direct intermolecular coupling with the second cWW at C3 would result in a fused pyrroloindoline ring system in 3. In the cases of 2 and 4 with a C3–C3' connection, radical migration from N1 to C3 is plausible, before the second cWW is attacked at C3'. Solving the crystal structures of such P450 enzymes in the future would provide new insights into the reaction mechanisms.

Total product yields of 2 in the J1074 transformant harbouring *ttp(AB)1* and 3 and 4 in the transformant with *ttp(AB)2* were calculated to be 205, 200 and 9.5 mg L⁻¹, respectively. In comparison to chemical synthesis, which usually requires complex procedures, our finding provides an efficient approach to afford dimeric DKPs in a direct manner. Therefore, these

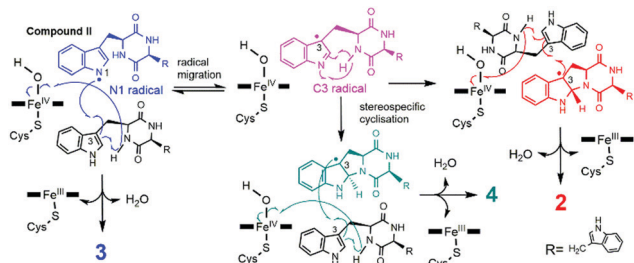


Fig. 3 Proposed mechanism for P450-mediated intermolecular coupling reactions.

transformants can be considered as excellent hosts for large-scale production of dimeric tetraptromycins. Notably, none of the tetraptromycins were detected in *S. antimicrobica* cultures (Fig. S34, ESI[†]), indicating the presence of two silent gene clusters in the native host. 2–4 were tested for their antibacterial activities against *E. coli* ATCC 25922 and K-12, *Bacillus subtilis* NCIB 3610, *Staphylococcus aureus* ATCC 29213 and *Pseudomonas aeruginosa* ATCC 27853.²⁹ Unfortunately, no inhibitory activity was observed.

In summary, we identified two *cdps-p450*-containing operons *ttp1* and *ttp2* in one bacterial strain for dimerization of the same CDP *cWW* *via* genome mining. The two P450s catalyse distinct coupling patterns differing from those previously reported in actinobacteria. Thus, our finding significantly increases the P450 repertoire for CDP modifications. As far as we know, TtpB1 represents the first bacterial P450 catalysing stereospecific C3 (sp³)–C3' (sp³) bond formation between two CDPs, while TtpB2 is characterised as the first P450 to catalyse not only the unusual linkage of C3 (sp³) of a hexahydropyrroloindole unit to N1' of the tryptophanyl moiety of the second diketopiperazine unit, but also the intermolecular C3–C3' bond formation. Therefore, our study provides a simple, direct and efficient approach for enzymatic one-step preparation of structurally complex DKP dimers and expands the classical chemical methods. Interestingly, for all the studied CDP-related dimerization gene clusters (Scheme S1, ESI[†] and Fig. 1A), there exist at least two similar *cdps-p450* operons in one native strain, which generate different dimeric CDPs. It could be speculated that the copies of these two-gene operons may come from a common ancestor and become functionalized after horizontal gene transfer in the evolutionary process, for various dimeric CDP derivatives with different linkage patterns and/or stereochemistry. It would be interesting to characterise more members of this intriguing enzyme group.

We thank R. Kraut and S. Newel for LC-MS and NMR analyses. This project was funded in part by the DFG (Li844/14-1). Jing Liu (201608310118) is a scholarship recipient from CSC.

Conflicts of interest

There are no conflicts to declare.

Notes and references

- 1 T. W. Giessen and M. A. Marahiel, *Front. Microbiol.*, 2015, **6**, 785.
- 2 A. D. Borthwick, *Chem. Rev.*, 2012, **112**, 3641–3716.
- 3 Y. M. Ma, X. A. Liang, Y. Kong and B. Jia, *J. Agric. Food Chem.*, 2016, **64**, 6659–6671.
- 4 P. Ruiz-Sanchis, S. A. Savina, F. Albericio and M. Alvarez, *Chemistry*, 2011, **17**, 1388–1408.
- 5 C. T. Walsh, *Nat. Prod. Rep.*, 2016, **33**, 127–135.
- 6 M. Gondry, L. Sauguet, P. Belin, R. Thai, R. Amouroux, C. Tellier, K. Tiphile, M. Jacquet, S. Braud, M. Courcon, C. Masson, S. Dubois, S. Lautru, A. Lecoq, S. Hashimoto, R. Genet and J. L. Pernodet, *Nat. Chem. Biol.*, 2009, **5**, 414–420.
- 7 M. Gondry, I. B. Jacques, R. Thai, M. Babin, N. Canu, J. Seguin, P. Belin, J. L. Pernodet and M. Moutiez, *Front. Microbiol.*, 2018, **9**, 46.
- 8 P. Borgman, R. D. Lopez and A. L. Lane, *Org. Biomol. Chem.*, 2019, **17**, 2305–2314.
- 9 H. Li, Y. Qiu, C. Guo, M. Han, Y. Zhou, Y. Feng, S. Luo, Y. Tong, G. Zheng and S. Zhu, *Chem. Commun.*, 2019, **55**, 8390–8393.
- 10 X. Zhang and S. Li, *Nat. Prod. Rep.*, 2017, **34**, 1061–1089.
- 11 J. Liu, X. Xie and S.-M. Li, *Angew. Chem., Int. Ed.*, 2019, **58**, 11534–11540.
- 12 H. Yu, X. Xie and S.-M. Li, *Org. Lett.*, 2018, **20**, 4921–4925.
- 13 H. Yu, X. Xie and S.-M. Li, *Org. Lett.*, 2019, **21**, 9104–9108.
- 14 H. Yu and S.-M. Li, *Org. Lett.*, 2019, **21**, 7094–7098.
- 15 N. G. M. Gomes, R. B. Pereira, P. B. Andrade and P. Valentão, *Mar. Drugs*, 2019, **17**, 551.
- 16 J. Kim and M. Movassaghi, *Acc. Chem. Res.*, 2015, **48**, 1159–1171.
- 17 J. Kim, J. A. Ashenurst and M. Movassaghi, *Science*, 2009, **324**, 238–241.
- 18 M. Movassaghi, M. A. Schmidt and J. A. Ashenurst, *Angew. Chem., Int. Ed.*, 2008, **47**, 1485–1487.
- 19 J. Kim and M. Movassaghi, *J. Am. Chem. Soc.*, 2011, **133**, 14940–14943.
- 20 T. Saruwatari, F. Yagishita, T. Mino, H. Noguchi, K. Hotta and K. Watanabe, *ChemBioChem*, 2014, **15**, 656–659.
- 21 W. Tian, C. Sun, M. Zheng, J. R. Harmer, M. Yu, Y. Zhang, H. Peng, D. Zhu, Z. Deng, S. L. Chen, M. Mobli, X. Jia and X. Qu, *Nat. Commun.*, 2018, **9**, 4428.
- 22 J. J. Hug, D. Krug and R. Müller, *Nat. Rev. Chem.*, 2020, **4**, 172–193.
- 23 L. M. Podust and D. H. Sherman, *Nat. Prod. Rep.*, 2012, **29**, 1251–1266.
- 24 N. Ziburannyi, M. Rabyk, B. Ostash, V. Fedorenko and A. Luzhetskyy, *BMC Genomics*, 2014, **15**, 97.
- 25 Y. Zhu, P. Fu, Q. Lin, G. Zhang, H. Zhang, S. Li, J. Ju, W. Zhu and C. Zhang, *Org. Lett.*, 2012, **14**, 2666–2669.
- 26 E. D. James, B. Knuckley, N. Alqahtani, S. Porwal, J. Ban, J. A. Karty, R. Viswanathan and A. L. Lane, *ACS Synth. Biol.*, 2016, **5**, 547–553.
- 27 J. Liu, H. Yu and S.-M. Li, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 4435–4444.
- 28 Z. Li, Y. Jiang, X. Zhang, Y. Chang, S. Li, X. Zhang, S. Zheng, C. Geng, P. Men, L. Ma, Y. Yang, Z. Gao, Y.-J. Tang and S. Li, *ACS Catal.*, 2020, **10**, 5846–5851.
- 29 M. Balouiri, M. Sadiqi and S. K. Ibsouda, *J. Pharm. Anal.*, 2016, **6**, 71–79.

Supporting information

Increasing cytochrome P450 enzyme diversity by identification of two distinct cyclodipeptide dimerases

Jing Liu,^a Xiulan Xie,^b and Shu-Ming Li^{*,a}

^a Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany

^b Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Straße 4, Marburg 35032, Germany

Table of Contents

Experimental Procedures	1
1. Computer-assisted sequence analysis.....	1
2. Bacterial strains, plasmids, and growth conditions	1
3. Genetic manipulation, PCR amplification, and gene cloning.....	1
4. Heterologous gene expression in <i>Streptomyces albus</i> J1074	2
5. Overproduction and purification of P450s in <i>E. coli</i> and <i>Streptomyces</i>	2
6. UV-Vis spectroscopic analysis of TtpB1	3
7. <i>In vitro</i> assays of P450s	3
8. Determination of kinetic parameters for TtpB1	3
9. LC-MS analysis	4
10. Isolation of generated metabolites from <i>S. albus</i> J1074 transformants.....	4
11. Precursor feeding experiments.....	4
12. Determination of production yields of cWW dimers in <i>Streptomyces</i> transformants.....	4
13. NMR analysis	5
14. The physicochemical properties of the identified compounds.....	5
Supplementary Tables.....	8
Table S1. Comparison of CDPSs and P450s from the two clusters in <i>Saccharopolyspora antimicrobica</i> DSM 45119.....	8
Table S2. Bacterial strains used in this study.....	9
Table S3. Cloning and expression constructs used in this study	10
Supplementary Scheme	11
Scheme S1 Known biosynthetic pathways of CDP dimers in <i>Aspergillus flavus</i> and <i>Streptomyces</i> species.....	11
Supplementary Figures	12
Fig. S1 Alignments of CDPSs from bacteria.....	12
Fig. S2 Phylogenetic analysis of P450s investigated in this study.....	13
Fig. S3 Alignments of CDP dimerization P450s from bacteria.	14
Fig. S4 ¹H NMR spectrum of tetratryptomycin A (2) in DMSO-<i>d</i>₆ at 300 K (500 MHz).	15
Fig. S5 ¹³C NMR spectrum of tetratryptomycin A (2) in DMSO-<i>d</i>₆ at 300 K (125 MHz).	16
Fig. S6 ¹H-¹H COSY spectrum of tetratryptomycin A (2) in DMSO-<i>d</i>₆ at 300 K (500 MHz).....	17

Fig. S7 HSQC spectrum of tetratryptomycin A (2) in DMSO- <i>d</i> ₆ at 300 K (500 MHz, 125 MHz).	18
Fig. S8 HMBC spectrum of tetratryptomycin A (2) in DMSO- <i>d</i> ₆ at 300 K (500 MHz, 125 MHz).	19
Fig. S9 NOESY spectrum of tetratryptomycin A (2) in DMSO- <i>d</i> ₆ at 300 K (500 MHz).	20
Fig. S10 ¹ H NMR spectrum of tetratryptomycin B (3) in DMSO- <i>d</i> ₆ at 300 K (500 MHz).	21
Fig. S11 ¹³ C NMR spectrum of tetratryptomycin B (3) in DMSO- <i>d</i> ₆ at 300 K (125 MHz).	22
Fig. S12 ¹ H- ¹ H COSY spectrum of tetratryptomycin B (3) in DMSO- <i>d</i> ₆ at 300 K (500 MHz).	23
Fig. S13 HSQC spectrum of tetratryptomycin B (3) in DMSO- <i>d</i> ₆ at 300 K (500 MHz, 125 MHz).	24
Fig. S14 HMBC spectrum of tetratryptomycin B (3) in DMSO- <i>d</i> ₆ at 300 K (500 MHz, 125 MHz).	25
Fig. S15 NOESY spectrum of tetratryptomycin B (3) in DMSO- <i>d</i> ₆ at 300 K (500 MHz).	26
Fig. S16 ¹ H NMR spectrum of tetratryptomycin C (4) in DMSO- <i>d</i> ₆ at 300 K (500 MHz).	27
Fig. S17 ¹³ C NMR spectrum of tetratryptomycin C (4) in DMSO- <i>d</i> ₆ at 300 K (125 MHz).	28
Fig. S18 ¹ H- ¹ H COSY spectrum of tetratryptomycin C (4) in DMSO- <i>d</i> ₆ at 300 K (500 MHz).	29
Fig. S19 HSQC spectrum of tetratryptomycin C (4) in DMSO- <i>d</i> ₆ at 300 K (500 MHz, 125 MHz).	30
Fig. S20 HMBC spectrum of tetratryptomycin C (4) in DMSO- <i>d</i> ₆ at 300 K (500 MHz, 125 MHz).	31
Fig. S21 NOESY spectrum of tetratryptomycin C (4) in DMSO- <i>d</i> ₆ at 300 K (500 MHz).	32
Fig. S22 ¹ H NMR spectrum of tetratryptomycin C (4) in acetonitrile- <i>d</i> ₃ at 273 K (500 MHz).	33
Fig. S23 ¹ H NMR spectrum of tetratryptomycin C (4) in acetonitrile- <i>d</i> ₃ at 300 K (500 MHz).	34
Fig. S24 ¹ H NMR spectrum of tetratryptomycin C (4) in acetonitrile- <i>d</i> ₃ at 310 K (500 MHz).	35
Fig. S25 ¹ H NMR spectrum of tetratryptomycin C (4) in acetonitrile- <i>d</i> ₃ at 320 K (500 MHz).	36
Fig. S26 ¹³ C NMR spectrum of tetratryptomycin C (4) in acetonitrile- <i>d</i> ₃ at 310 K (125 MHz).	37
Fig. S27 HSQC spectrum of tetratryptomycin C (4) in acetonitrile- <i>d</i> ₃ at 310 K (500 MHz, 125 MHz). ...	38
Fig. S28 CD spectra of tetratryptomycins.	39
Fig. S29 SDS-PAGE analysis of the purified P450s.	40
Fig. S30 UV-Vis spectroscopic analysis of TtpB1.	41
Fig. S31 HPLC analysis of the enzyme assays with TtpB1 (A) and TtpB2 (B).	42
Fig. S32 Determination of kinetic parameters of TtpB1 for cWW (1).	43
Fig. S33 UV-Vis spectroscopic analysis of TtpB1 with its substrate cWW (A) and other CDPs (B).	44
Fig. S34 LC-MS analysis for tetratryptomycin production in <i>S. antimicrobica</i> DSM 45119.	45
References	46

Experimental Procedures

1. Computer-assisted sequence analysis

The gene and protein sequences used in this study were obtained from NCBI databases (<http://www.ncbi.nlm.nih.gov>). Protein sequences were compared with each other by using BLASTP program (<http://blast.ncbi.nlm.nih.gov/>). The phylogenetic tree of P450s showing in Fig. S2 was created by MEGA version 7.0 (<http://www.megasoftware.net>). Protein sequence alignments were performed with the program ClustalW and visualized with ESPript 3.0 (<http://endscript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>) to identify strictly conserved amino acid residues (Fig. S1 and S3).

2. Bacterial strains, plasmids, and growth conditions

Strains and plasmids used in this study are listed in Tables S2 and Table S3, respectively. Recombinant *E. coli* strains were cultivated in liquid or on solid Luria-Bertani (LB) medium with 100 µg/mL ampicillin, 50 µg/mL kanamycin, 50 µg/mL apramycin or 25 µg/mL chloramphenicol, when necessary.

Saccharopolyspora antimicrobica DSM 45119 was purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). *Streptomyces albus* J1074¹ was kindly gifted by Prof. 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, MgCl₂·6H₂O 10.12 g/L, K₂SO₄ 0.25 g/L, Difco casaminoacids 0.1 g/L, MOPS 21.0 g/L, trace element solution 2 mL/L, pH 7.2) at 28 °C for 7 days.

3. Genetic manipulation, PCR amplification, and gene cloning

Genetic manipulation in *E. coli* was performed according to the protocol by Green and Sambrook.² Isolation of genomic DNA from actinomycetes was carried out as described in the literature.³ The *cdps* and *p450* genes were amplified by PCR from genomic DNA of *S. antimicrobica* DSM 45119 by using primers listed in Table S3 and Phusion® High-Fidelity DNA Polymerase from New England Biolabs (NEB). The generated PCR fragments were cloned into pGEM-T Easy vector and the sequence integrity was confirmed by sequencing. Subsequently, the fragments were released with restriction endonucleases from pGEM-T Easy and ligated into pPWW50A⁴ or pET28a (+) vector, which were digested with the same enzymes, previously. The

generated constructs (Table S3) were transformed into *S. albus* J1074 or *E. coli* BL21 (DE3) for gene expression.

4. Heterologous gene expression in *Streptomyces albus* J1074

The constructed plasmids harbouring different genes or gene clusters were firstly transformed into the non-methylating *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 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. 1 mL of the cultures was extracted with the same volume of ethyl acetate for three times. The organic phases were combined, evaporated, and the dried residues were afterwards dissolved in 400 µL of methanol. 5 µL of such samples were subjected to LC-MS for analysis.

5. Overproduction and purification of P450s in *E. coli* and *Streptomyces*

For the purification of TtpB1, pJL80 was transformed into *E. coli* BL21 (DE3). 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 grew at 37 °C and 230 rpm to an absorption of 0.6 at 600 nm. The 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 His₆-tagged protein was performed by using Ni-NTA agarose (Macherey-Nagel, Düren, Germany) according to the manufacturer's instruction. The storage buffer was changed to 50 mM Tris-HCl (pH 7.5) containing 15 % (v/v) glycerol through a PD-10 column (GE Healthcare, Freiburg, Germany), which had been equilibrated with the same buffer. The obtained protein was stored frozen at -80 °C.

S. albus J1074 harbouring pJL84 (*ttpB2* in pPWW50A, Table S3) was cultivated in 50 mL tryptic soy broth (TSB) medium containing 50 µg/mL apramycin for 48 h as preculture. 5 mL of this preculture were transferred to 100 mL TSB with 50 µg/mL apramycin in 500 mL conical flasks. The cultures were further incubated at 28°C and 200 rpm for 3 days. Two litres of such cultures were harvested by centrifugation at 4 °C and 4,500 rpm for 20 min. The protein was purified as

described for *E. coli* cells.

The concentrations of TtpB1 and TtpB2 were determined on a Nanodrop C2000 (Thermo Scientific, Braunschweig, Germany) to be 0.65 mg/L and 0.05 mg/L culture, respectively. The purity of the recombinant P450s was proven by 12 % (w/v) SDS-PAGE (Fig. S29).

6. UV-Vis spectroscopic analysis of TtpB1

To measure the typical absorbance of P450 ferrous CO complex after reduction, carbon monoxide gas was bubbled into the TtpB1 solution (14 μ M in 50 mM Tris-HCl containing 15% (v/v) glycerol, pH 7.5) for 2 min. After addition of 0.2 g/mL of sodium dithionite, a UV-Vis spectrum between 350 and 550 nm was taken on a Multiskan™ GO Microplate Spectrophotometer (Thermo Scientific, Dreieich, Germany). UV-Vis spectra of a protein sample without any treatment and another one only bubbled with CO were taken as controls. The spectra of TtpB1 is given in Fig. S30.

7. *In vitro* assays of P450s

cWW (**1**, 500 μ M) was first assayed with 7 μ M TtpB1, 5 mM NADPH, 2 μ M spinach ferredoxin (Sigma-Aldrich), 0.1 unit/mL spinach ferredoxin-NADP⁺ reductase (Sigma-Aldrich), 50 mM Tris-HCl buffer (pH 7.5) in a total volume of 50 μ L at 30 °C for 30 min. Afterwards, 50 μ M of **1** was incubated with 5 μ M TtpB1 or 1.1 μ M TtpB2 for 12 h. The reactions were quenched with 50 μ L ice-cold MeOH. After centrifugation at 13,000 rpm for 5 min, 5 μ L of the supernatants were subjected to LC-MS analysis. Incubations with heat-inactivated P450s, without ferredoxin, ferredoxin reductase, or NADPH were used as negative controls.

8. Determination of kinetic parameters for TtpB1

For determination of the kinetic parameters of TtpB1 towards cWW (**1**), the reaction mixtures (50 μ L) contained 7 μ M TtpB1, 5 mM NADPH, 2 μ M spinach ferredoxin (Sigma-Aldrich), 0.1 unit/mL spinach ferredoxin-NADP⁺ reductase (Sigma-Aldrich), 50 mM Tris-HCl buffer (pH 7.5) and **1** at final concentrations of 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, and 0.4 mM. The reactions were carried out at 30 °C for 30 min and terminated by addition of 50 μ L ice-cold MeOH. After removal of proteins by centrifugation, 50 μ L of the supernatants were subjected to HPLC analysis.

The analysis was carried out on an Agilent HPLC 1200 series equipped with a photo diode array detector and an Eclipse XDB C18 column (5 μ m, 4.6 x 150 mm). A linear gradient of 5 to 100% acetonitrile in water (0.1% formic acid) in 20 min was followed by 100% acetonitrile for 5 min and 5% acetonitrile in water for 5 min. The flow rate was set to 0.5 mL/min. Absorptions at 280 nm were illustrated in this study. The K_m and k_{cat} were proceeded with GraphPad Prism 8 (Fig. S32).

9. 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 (250 x 2 mm, 5 μ m, CS-Chromatographie Service GmbH). For secondary metabolite analysis, a linear gradient of 5 – 100 % acetonitrile in water, both containing 0.1 % formic acid, in 40 min and a flow rate at 0.25 mL/min were used. The column was then washed with 100 % acetonitrile containing 0.1 % formic acid for 5 min and equilibrated with 5 % acetonitrile in water for 5 min. For enzyme assay analysis, a linear gradient of 5 – 100 % acetonitrile in water in 10 min was used, and the column was then washed and equilibrated as described as the former method. 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.

10. Isolation of generated metabolites from *S. albus* J1074 transformants

For structural elucidation of the accumulated compounds, the *S. albus* J1074 transformants harbouring *ttp(ABC)1* and *ttp(AB)2* were fermented in modified R5 medium on a large scale (4 L) at 28 °C for 7 days. The cultures were extracted with equal volume of ethyl acetate for three times. The organic phases were combined and evaporated to dryness. The extracts were applied to a silica gel column and eluted with a gradient CH₂Cl₂: MeOH in ratios of 100:2, 100:3, 100:5, 100:10. The target compounds **2** and **4** were mainly found in the fractions eluted with CH₂Cl₂: MeOH of 100:5 and **3** in 100:10, respectively. These fractions were further purified on an Agilent HPLC 1260 series equipped with a photo diode array detector by using a semi-preparative Agilent ZORBAX Eclipse XDB C18 HPLC column (9.4 x 250 mm, 5 μ m) with 55 % ACN in water as solvent. The flow rate was set to 2.0 mL/min.

11. Precursor feeding experiments

Precursor feeding was carried out by using 20 mM cWW in DMSO. 150 μ L of this solution were added to 30 mL of 2 day-old cultures of *Streptomyces* transformants in modified R5 media. After cultivation at 28°C for additional 7 days, the metabolites were extracted with EtOAc and analysed on LC-MS.

12. Determination of production yields of cWW dimers in *Streptomyces* transformants

An Agilent HPLC 1200 series equipped with a photo diode array detector and an Agilent Eclipse XDB C18 column (5 μ m, 4.6 x 150 mm) were used for quantification. A linear gradient of 10 to 100 % acetonitrile in water in 40 min was followed by 100 % acetonitrile for 5 min and then 10 %

acetonitrile in water for 5 min. The flow rate was set to 0.5 mL/min. The absorption at 280 nm was used for quantification. To ensure complete extraction of cWW dimers from mycelia, precipitants and supernatants, 1 mL whole culture of *S. albus* J1074 transformants was extracted with 1 mL ethyl acetate for three times. The organic phases were combined and evaporated to dryness. The residues were dissolved in 200 μ L of methanol and 100 μ L were analysed on HPLC. The isolated products were used as authentic standards for quantification.

13. NMR analysis

The NMR spectra of the purified compounds **2** and **3** were recorded on a JOEL ECA-500 MHz spectrometer (JEOL, Tokyo, Japan) in DMSO- d_6 . The NMR spectra of compound **4** in DMSO- d_6 were taken at 300 K on a Bruker AVIII spectrometer (500 MHz) equipped with a 5 mm cryo BBO probe Prodigy. To obtain a better NMR signal resolution, the ^1H NMR of **4** was also recorded in acetonitrile- d_3 at 273 K, 300 K, 310 K, and 320 K on a Bruker HD AVII spectrometer (500 MHz) equipped with a cryo BBO probe Prodigy. The ^{13}C and HSQC NMR spectra of **4** were then taken at the best found temperature 310 K on the same equipment.

All spectra were processed with MestReNova 5.2.2 (Metrelab Research, S5 Santiago de Compostella, Spain). The NMR data of the identified compounds are listed as physiochemical properties and the spectra are provided in Fig. S4 – S27.

14. The physiochemical properties of the identified compounds

Tetratryptomycin A (**2**): 30 mg, light yellow powder; slightly soluble in modified R5 media (approx. 20 mg/L); CD (MeOH) λ_{max} ($\Delta\epsilon$) 306 (-18.4), 269 (-4.8), 249 (-29.4), 227 (+13) nm; HRMS (m/z): (ESI/[M+H] $^+$) calcd. for $\text{C}_{44}\text{H}_{38}\text{N}_8\text{O}_4$, 743.3089, found 743.3090. ^1H NMR (DMSO- d_6 , 500 MHz) δ 10.68 (s, 2H, H-20 and H-20'), 7.66 (s, 2H, H-15 and H-15'), 7.55 (d, $J = 7.8$ Hz, 2H, H-25 and H-25'), 7.38 (d, $J = 8.1$ Hz, 2H, H-22 and H-22'), 7.19 (d, $J = 7.4$ Hz, 2H, H-4 and H-4'), 7.07 (s, 2H, H-19 and H-19'), 7.06 (t, $J = 8.1$ Hz, 2H, H-23 and H-23'), 7.04 (t, $J = 7.8$ Hz, 2H, H-6 and H-6'), 6.96 (t, $J = 7.8$ Hz, 2H, H-24 and H-24'), 6.67 (s, 2H, H-1 and H-1'), 6.66 (t, $J = 7.4$ Hz, 2H, H-5 and H-5'), 6.57 (d, $J = 7.8$ Hz, 2H, H-7 and H-7'), 5.10 (s, 2H, H-2 and H-2'), 4.34 (t, $J = 5.0$ Hz, 2H, H-14 and H-14'), 3.82 (dd, $J = 12.1, 5.9$ Hz, 2H, H-11 and H-11'), 3.25 (dd, $J = 15.0, 5.7$ Hz, 2H, H-17 α and H-17' α), 3.04 (dd, $J = 15.0, 5.7$ Hz, 2H, H-17 β and H-17' β), 2.35 (dd, $J = 12.3, 5.9$ Hz, 2H, H-10 α and H-10' α), 2.25 (t, $J = 11.9$ Hz, 2H, H-10 β and H-10' β). ^{13}C NMR (DMSO- d_6 , 125 MHz) δ 168.1 (C-16 and C-16'), 165.4 (C-13 and C-13'), 151.0 (C-8 and C-8'), 136.0 (C-21 and C-21'), 129.1 (C-6 and C-6'), 127.5 (C-26 and C-26'), 127.0 (C-9 and C-9'), 124.6 (C-4 and C-4'), 123.7 (C-19 and C-19'), 120.9 (C-23 and C-23'), 118.4 (C-25 and C-25'), 118.2 (C-24 and C-24'), 117.4 (C-5 and C-5'), 111.4 (C-22 and C-22'), 109.3 (C-18 and C-18'), 108.7 (C-7 and C-7'), 77.2

(C-2 and C-2'), 58.7 (C-3 and C-3'), 57.8 (C-11 and C-11'), 55.0 (C-14 and C-14'), 36.2 (C-10 and C-10'), 25.4 (C-17 and C-17').

Tetratryptomycin B (**3**): 40 mg, light yellow powder; slightly soluble in modified R5 media (approx. 20 mg/L); CD (MeOH) λ_{\max} ($\Delta\epsilon$) 300 (+11.4), 268 (-5.34), 245 (+4.8), 240 (+7.0), 229 (+47.0), 214 (-66.5) nm; HRMS (m/z): (ESI/[M+H]⁺) calcd. for C₄₄H₃₈N₈O₄, 743.3089, found 743.3119. ¹H NMR (DMSO-*d*₆, 500 MHz) δ_{H} 10.95 (s, 1H, H-20'), 10.86 (s, 1H, H-20), 8.10 (s, 1H, H-15'), 7.94 (s, 1H, H-15), 7.62 (d, J = 1.9 Hz, 1H, H-12'), 7.58 (d, J = 7.5 Hz, 1H, H-25), 7.54 (d, J = 7.7 Hz, 1H, H-25'), 7.38 (d, J = 8.0 Hz, 1H, H-22'), 7.34 (d, J = 8.0 Hz, 1H, H-22), 7.28 (s, 1H, H-19), 7.24 (d, J = 7.8 Hz, 1H, H-4'), 7.19 (d, J = 3.6 Hz, 1H, H-1), 7.15 - 7.06 (m, 4H, H-6, H-23, H-23' and H-24'), 6.99 (t, J = 7.5 Hz, H-24), 6.95 (t, J = 7.8 Hz, 1H, H-5'), 6.88 (t, J = 7.8 Hz, 1H, H-6'), 6.88 (s, 1H, H-19'), 6.73 (d, J = 7.5 Hz, 1H, H-4), 6.70 (d, J = 7.8 Hz, 1H, H-7), 6.54 (d, J = 7.8 Hz, 1H, H-7'), 6.53 (t, J = 7.5 Hz, 1H, H-5), 6.47 (s, 1H, H-2'), 5.82 (d, J = 3.6 Hz, 1H, H-2), 4.67 (dd, J = 13.9, 5.9 Hz, 1H, H-11), 4.39 (t, J = 5.1 Hz, 1H, H-14), 4.04 (m, 1H, H-14'), 3.71 (m, 1H, H-11'), 3.48 (dd, J = 14.6, 5.9 Hz, 1H, H-10 α), 3.39 (dd, J = 15.4, 5.1 Hz, 1H, H-17 α), 3.08 - 2.98 (m, 2H, H-17 β and H-17' β), 2.85 (dd, J = 14.1, 4.4 Hz, 1H, H-17' α), 2.64 (dd, J = 14.0, 3.3 Hz, 1H, H-10' α), 2.23 (t, J = 14.2, Hz, 1H, H-10 β), 1.27 (dd, J = 14.0, 9.6 Hz, 1H, H-10' β). ¹³C NMR (DMSO-*d*₆, 125 MHz) δ 169.2 (C-16), 167.8 (C-13), 166.8 (C-16'), 166.6 (C-13'), 148.0 (C-8), 136.0 (C-21), 136.0 (C-21'), 135.0 (C-8'), 129.6 (C-6), 129.2 (C-9'), 128.6 (C-9), 127.7 (C-26'), 127.3 (C-26), 124.8 (C-19'), 124.7 (C-2'), 124.1 (C-19), 122.2 (C-4), 121.0 (C-6'), 120.9 (C-23'), 120.8 (C-23), 119.2 (C-5'), 119.0 (C-25'), 118.6 (C-4'), 118.5 (C-24'), 118.4 (C-25), 118.3 (C-24), 118.1 (C-5), 111.8 (C-7'), 111.5 (C-22'), 111.3 (C-22), 110.0 (C-7), 109.5 (C-18), 109.1 (C-3'), 108.6 (C-18'), 81.4 (C-2), 73.4 (C-3), 57.4 (C-11), 55.4 (C-14'), 55.2 (C-14), 54.5 (C-11'), 39.1 (C-10), 30.2 (C-10'), 29.8 (C-17'), 24.4 (C-17).

Tetratryptomycin C (**4**): 10 mg, light yellow powder; slightly soluble in modified R5 media (approx. 20 mg/L); CD (MeOH) λ_{\max} ($\Delta\epsilon$) 267 (-75.4), 216 (-2.7), 242 (+14.3), 224 (-1.4), 219 (+2.2) nm; HRMS (m/z): (ESI/[M+H]⁺) calcd. for C₄₄H₃₈N₈O₄, 743.3089, found 743.3113. ¹H NMR (DMSO-*d*₆, 500 MHz) δ 10.81 (br s, 1H, H-20), 10.73 (br s, 1H, H-20'), 7.95 (br s, 1H, H-15'), 7.63 (d, J = 7.6 Hz 1H, H-25'), 7.60 (s, 1H, H-15), 7.58 (d, J = 7.8 Hz 1H, H-25), 7.33 (d, J = 7.9 Hz, 1H, H-22), 7.28 (d, J = 7.5 Hz, 1H, H-22'), 7.19 (s, 1H, H-19), 7.15 (s, 1H, H-19'), 7.08 (t, J = 7.9 Hz, 1H, H-23), 7.02 - 6.93 (m, 5H, H-6, H-6', H-23', H-24 and H-24'), 6.45 (m, 6H, H-4, H-4', H-5, H-5', H-7 and H-7'), 5.37 (br s, 1H, H-2), 5.13 (br s, 1H, H-2'), 4.35 (m, 1H, H-14'), 4.29 (m, 1H, H-14), 4.07 (m, 1H, H-11), 3.77 (m, 1H, H-11'), 3.36 (m, 1H, H-17 α), 3.24 - 3.16 (m, 2H, H-17' α and H-17' β), 2.96 (dd, J = 14.4, 6.7 Hz, 1H, H-17 β), 2.37 (m, 1H, H-10' α), 2.27 (dd, J = 13.0, 6.3 Hz, 1H,

H-10 α), 1.76 (t, $J = 11.8$ Hz, 1H, H-10' β), 1.07 (m, 1H, H-10 β). ^{13}C NMR (DMSO- d_6 , 125 MHz) δ 169.6 (C-16), 168.6 (C-13), 167.0 (C-16'), 164.7 (C-13'), 151.2 (C-8), 148.9 (C-8'), 136.1 (C-21), 135.9 (C-21'), 130.0 (C-9), 128.9 (C-6), 128.5 (C-6'), 127.6 (C-26'), 127.5 (C-9'), 127.1 (C-26), 124.4 (C-19'), 124.2 (C-19), 123.8 (C-4), 123.7 (C-4'), 120.9 (C-23), 120.8 (C-23'), 118.9 (C-25'), 118.3 (C-25), 118.3 (C-24'), 118.3 (C-24), 117.5 (C-5), 117.4 (C-5'), 111.4 (C-22'), 111.4 (C-22), 109.5 (C-18), 108.9 (C-18'), 108.8 (C-7), 108.3 (C-7'), 78.1 (C-2'), 76.0 (C-2), 59.8 (C-3), 59.6 (C-3'), 57.3 (C-11'), 57.2 (C-11), 55.7 (C-14'), 55.4 (C-14), 38.6 (C-10'), 34.5 (C-10), 26.8 (C-17'), 24.3 (C-17). ^1H NMR (acetonitrile- d_3 , 500 MHz, 310 K) δ 9.68 (br s, 1H, H-20), 9.19 (br s, 1H, H-20'), 7.65 (d, $J = 8.0$ Hz, 1H, H-25'), 7.63 (d, $J = 8.0$ Hz, 1H, H-25), 7.41 (d, $J = 8.0$ Hz, 1H, H-22'), 7.39 (d, $J = 7.9$ Hz, 1H, H-22), 7.18 – 7.14 (m, 3H, H-19, H-23' and H-23), 7.11 – 7.04 (m, 5H, H-6, H-6', H-19', H-24' and H-24), 6.58 – 6.51 (m, 6H, H-4, H-4', H-5, H-5', H-7 and H-7'), 6.45 (s, 1H, H-15'), 6.03 (s, 1H, H-15), 5.65 (br s, 1H, H-1), 5.47 (s, 1H, H-2), 5.23 (br s, 1H, H-1'), 5.12 (s, 1H, H-2'), 4.39 (t, $J = 4.6$ Hz, 1H, H-14'), 4.36 (dd, $J = 8.0, 4.9$ Hz, 1H, H-14), 4.18 (t, $J = 8.2$ Hz, 1H, H-11), 3.78 (dd, $J = 12.0, 5.4$ Hz, 1H, H-11'), 3.47 (dd, $J = 15.0, 4.5$ Hz, 1H, H-17 α), 3.36 (dd, $J = 14.8, 5.0$ Hz, 1H, H-17' α), 3.49 (dd, $J = 15.0, 4.2$ Hz, 1H, H-17' β), 3.11 (dd, $J = 15.2, 7.6$ Hz, 1H, H-17 β), 2.42 – 2.38 (m, 2H, H-10' α and H-10 α), 1.73 (t, $J = 12.0$ Hz, 1H, H-10' β), 1.31 (m, 1H, H-10 β). ^{13}C NMR (acetonitrile- d_3 , 125 MHz, 310 K) δ 169.3 (C-16), 168.8 (C-13), 166.6 (C-16'), 164.8 (C-13'), 151.1 (C-8), 148.8 (C-8'), 136.6 (C-21), 136.4 (C-21'), 130.3 (C-9), 129.2 (C-6), 128.8 (C-6'), 127.9 (C-9'), 127.8 (C-26'), 127.3 (C-26), 124.5 (C-4), 124.2 (C-4'), 124.0 (C-19), 123.9 (C-19'), 121.7 (C-23), 121.6 (C-23'), 119.1 (C-24), 119.1 (C-24'), 118.9 (C-25'), 118.7 (C-5), 118.5 (C-5'), 118.4 (C-25), 111.7 (C-22'), 111.5 (C-22), 109.3 (C-18), 109.3 (C-18'), 109.3 (C-7), 108.8 (C-7'), 78.9 (C-2'), 76.5 (C-2), 60.1 (C-3), 60.0 (C-3'), 57.4 (C-11'), 57.3 (C-11), 56.5 (C-14'), 55.4 (C-14), 39.0 (C-10'), 35.0 (C-10), 27.9 (C-17'), 25.2 (C-17).

Supplementary Tables

Table S1. Comparison of CDPSs and P450s from the two clusters in *Saccharopolyspora antimicrobica* DSM 45119

<i>ttp</i> cluster 1			<i>ttp</i> cluster 2			Sequence identity (%)
Protein	Accession No.	Length (aa)	Protein	Accession No.	Length (aa)	
TtpA1	WP_093145978.1	257	TtpA2	WP_093145813.1	254	80
TtpB1	WP_121505431.1	400	TtpB2	WP_170210414.1	400	71
TtpC1	WP_143121525.1	412				31 to TtpB1 39 to TtpB2

Table S2. Bacterial strains used in this study

Strain	Source	Cultivation media
<i>E. coli</i> DH5 α	Invitrogen	LB
<i>E. coli</i> BL21(DE3)	Novagen	LB
<i>E. coli</i> ET12567/pUZ8002	5	LB
<i>Streptomyces albus</i> J1074	1	MS
<i>Saccharopolyspora antimicrobica</i> DSM 45119	DSMZ	modified R5

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen

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

Modified R5 medium: sucrose 103.0 g/L, glucose 10.0 g/L, yeast extract 5.0 g/L, MgCl₂·6H₂O 10.12 g/L, K₂SO₄ 0.25 g/L, Difco casaminoacids 0.1 g/L, MOPS 21.0 g/L, trace element solution 2 mL/L, pH 7.2.

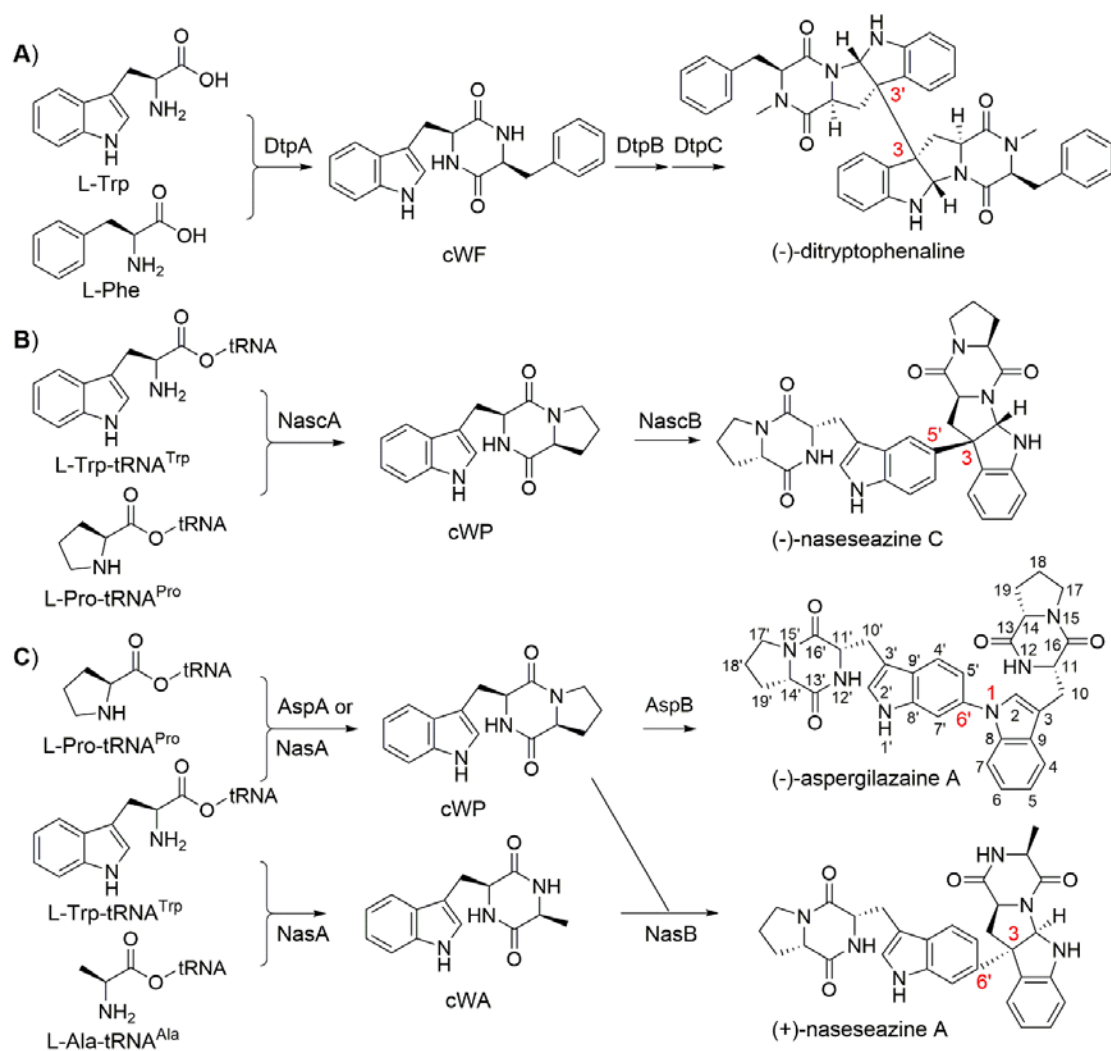
MS medium: mannitol 20.0 g/L, soya flour 20.0 g/L, agar 20.0 g/L.

Table S3. Cloning and expression constructs used in this study

Gene	Primer sequences (5'-3')	Cloning constructs	Expression vector	Cloning sites	Expression constructs
<i>ttpA1</i>	<u>CATATG</u> CCACCCACGCCTACCACTG <u>GGATCC</u> TCACGCGGTGCTGGCTCGTC	pJL66	pPWW50A	<i>NdeI/BamHI</i>	pJL75
<i>ttp(ABC)1</i>	<u>CATATG</u> CCACCCACGCCTACCACTG <u>GGATCC</u> TCACGCGGTGCTGGCTCGTC	pJL67	pPWW50A	<i>NdeI/BamHI</i>	pJL76
<i>ttp(AB)1</i>	<u>CATATG</u> CCACCCACGCCTACCACTG <u>GGATCC</u> CTACCAGCTCACGGGAAGGGC	pJL68	pPWW50A	<i>NdeI/ BamHI</i>	pJL77
<i>ttp(AC)1</i>	<u>CATATG</u> CCACCCACGCCTACCACTG <u>GGATCC</u> TCACGCGGTGCTGGCTCGTC	pJL66	pPWW50A	<i>NdeI/BamHI</i>	pJL78
	<u>AGATCT</u> GTGCCACCAGACAACGAGGGCG <u>ACTAGT</u> TCACCAGCTGACGGGGAGCTG	pJL67		<i>BglII/SpeI</i>	
<i>ttpB1</i>	<u>CATATG</u> TTCGCCATCGACGACATCCCG <u>GGATCC</u> CTACCAGCTCACGGGAAGGGC	pJL70	pPWW50A	<i>NdeI/BamHI</i>	pJL79
<i>ttpB1</i>	<u>CATATG</u> TTCGCCATCGACGACATCCCG <u>GGATCC</u> CTACCAGCTCACGGGAAGGGC	pJL70	pET28a (+)	<i>NdeI/BamHI</i>	pJL80
<i>ttpC1</i>	<u>CATATG</u> CCACCAGACAACGAGGGCG <u>GGATCC</u> TCACGTCAAGTCCCTTTCTCC	pJL71	pPWW50A	<i>NdeI/BamHI</i>	pJL81
<i>ttpA2</i>	<u>CATATG</u> CATTCCACGTGTATCGACCGAG <u>GGATCC</u> CTACTGGACAGCATCGTTCCCCC	pJL72	pPWW50A	<i>NdeI/BamHI</i>	pJL82
<i>ttp(AB)2</i>	<u>CATATG</u> CATTCCACGTGTATCGACCGAG <u>GGATCC</u> CTACCAGGTGACGGGCAGGG	pJL73	pPWW50A	<i>NdeI/BamHI</i>	pJL83
<i>ttpB2</i>	<u>CATATG</u> CTGTCCAGTGATCAGATCCCGG <u>GGATCC</u> CTACCAGGTGACGGGCAGGG	pJL74	pPWW50A	<i>NdeI/BamHI</i>	pJL84

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

Supplementary Scheme



Scheme S1 Known biosynthetic pathways of CDP dimers in *Aspergillus flavus* (A) and *Streptomyces* species (B and C)

Supplementary Figures

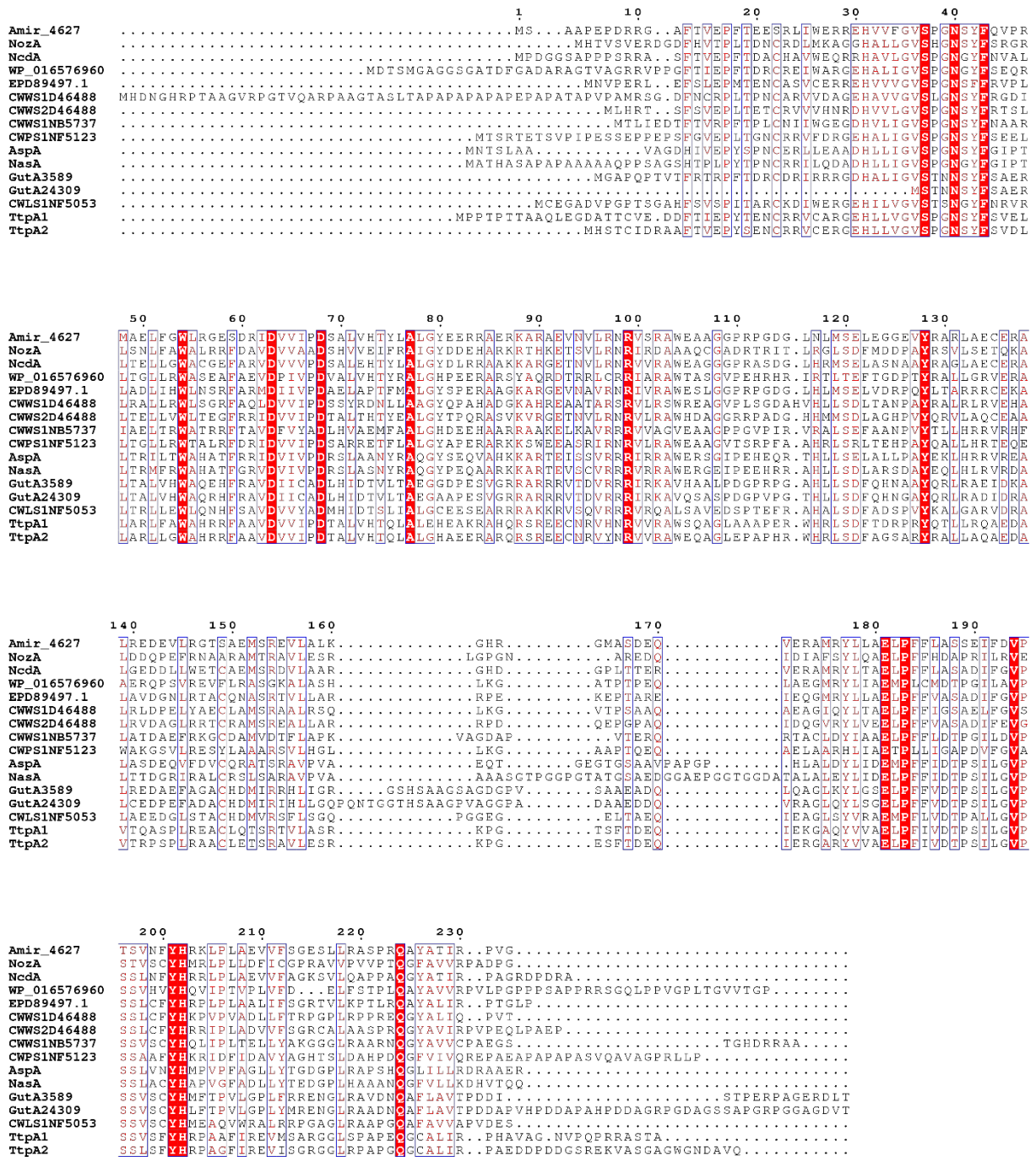


Fig. S1 Alignments of CDPs from bacteria. Amir₄₆₂₇,⁶ NozA,⁷ NcdA,⁷ WP_016576960,⁸ EPD89497.1,^{8,9} CWWS1_{D46488},¹⁰ CWWS2_{D46488},¹⁰ CWWS1_{NB5737},¹⁰ CWPS1_{NF5123},¹⁰ AspA,¹¹ NasA,¹¹ GutA₃₅₈₉,¹² GutA₂₄₃₀₉¹² and CWLS1_{NF5053}¹⁰ have been characterised as tryptophan-containing CDP synthases.

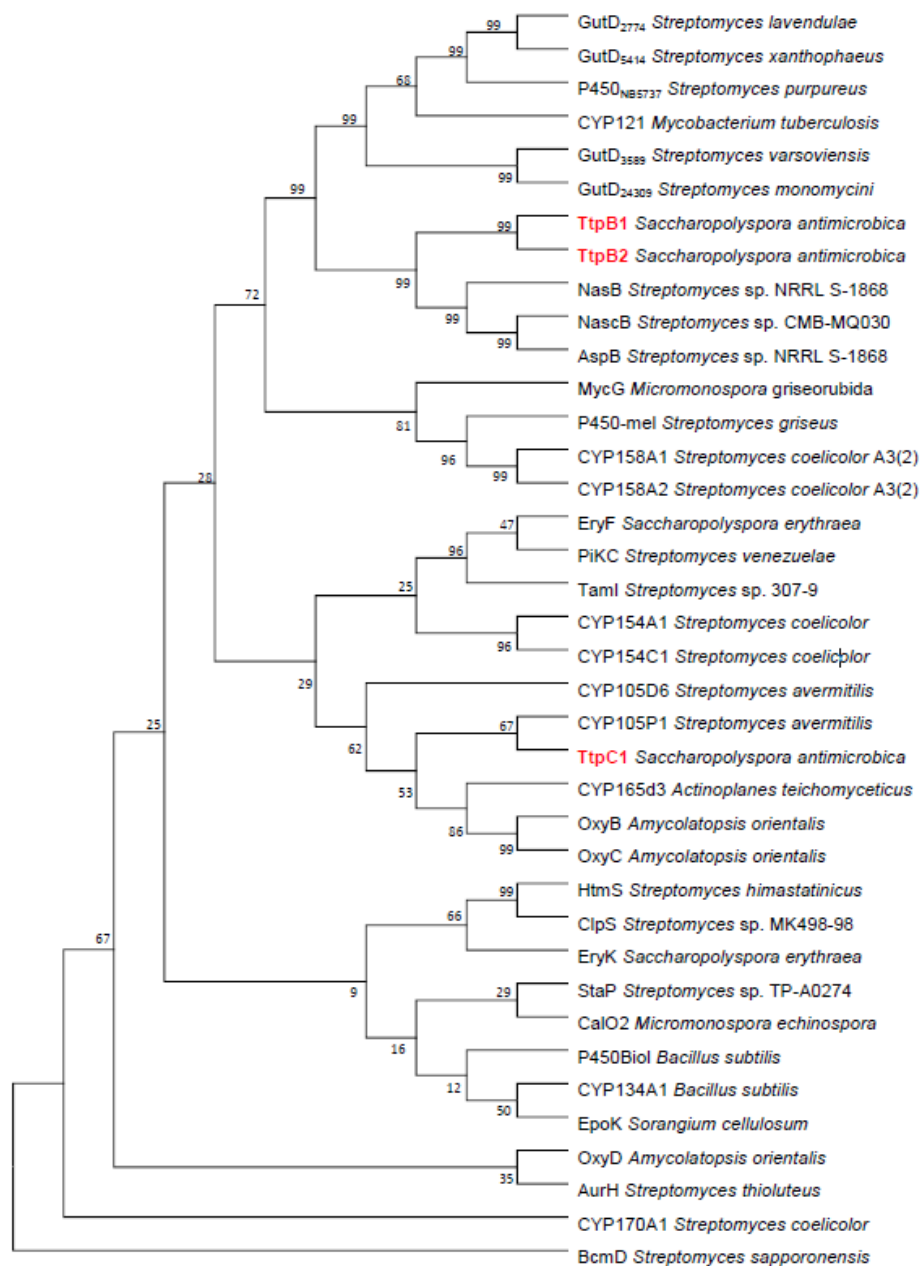


Fig. S2 Phylogenetic analysis of P450s investigated in this study (in bold red) and functionally characterised P450s from bacteria. GutD₂₇₇₄,¹³ GutD₅₄₁₄,¹³ GutD₂₄₃₀₉,¹² GutD₃₅₈₉,¹² P450₅₇₃₇,¹³ CYP121,¹⁴ NasB,¹¹ NascB,¹⁵ AspB,¹¹ CYP134A1¹⁶ and BcmD¹⁷ are members of the CDPS-related pathways. HtmS¹⁸ and ClpS¹⁹ are involved in the biosynthesis of himastatin and chloptosin, respectively. Other enzymes are structurally characterised natural product P450s mentioned in the review by Podust *et al.*²⁰ The protein sequences were downloaded from NCBI database.

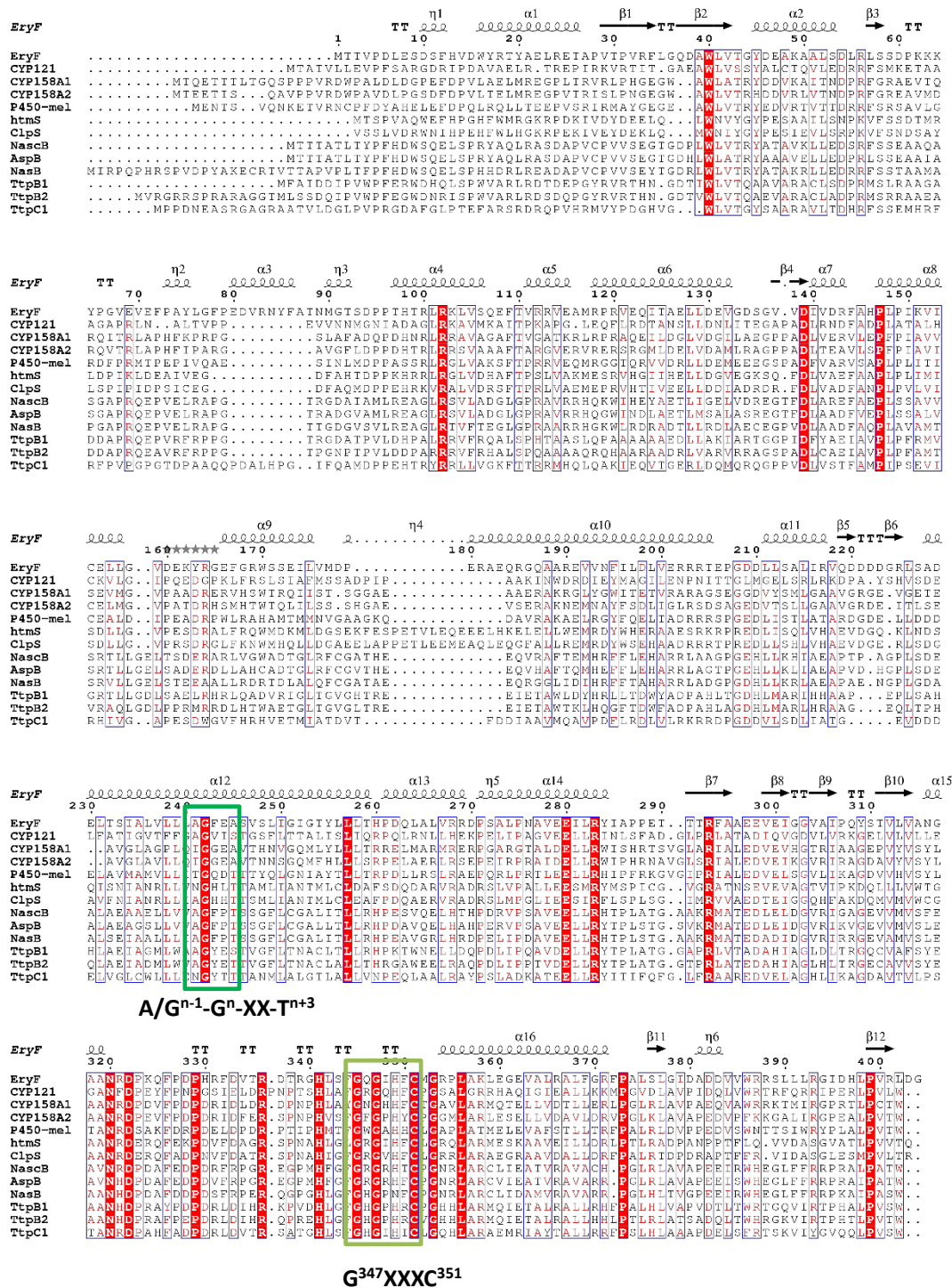


Fig. S3 Alignments of CDP dimerization P450s from bacteria. The origins of the enzymes were mentioned in the legend to Fig. S2.

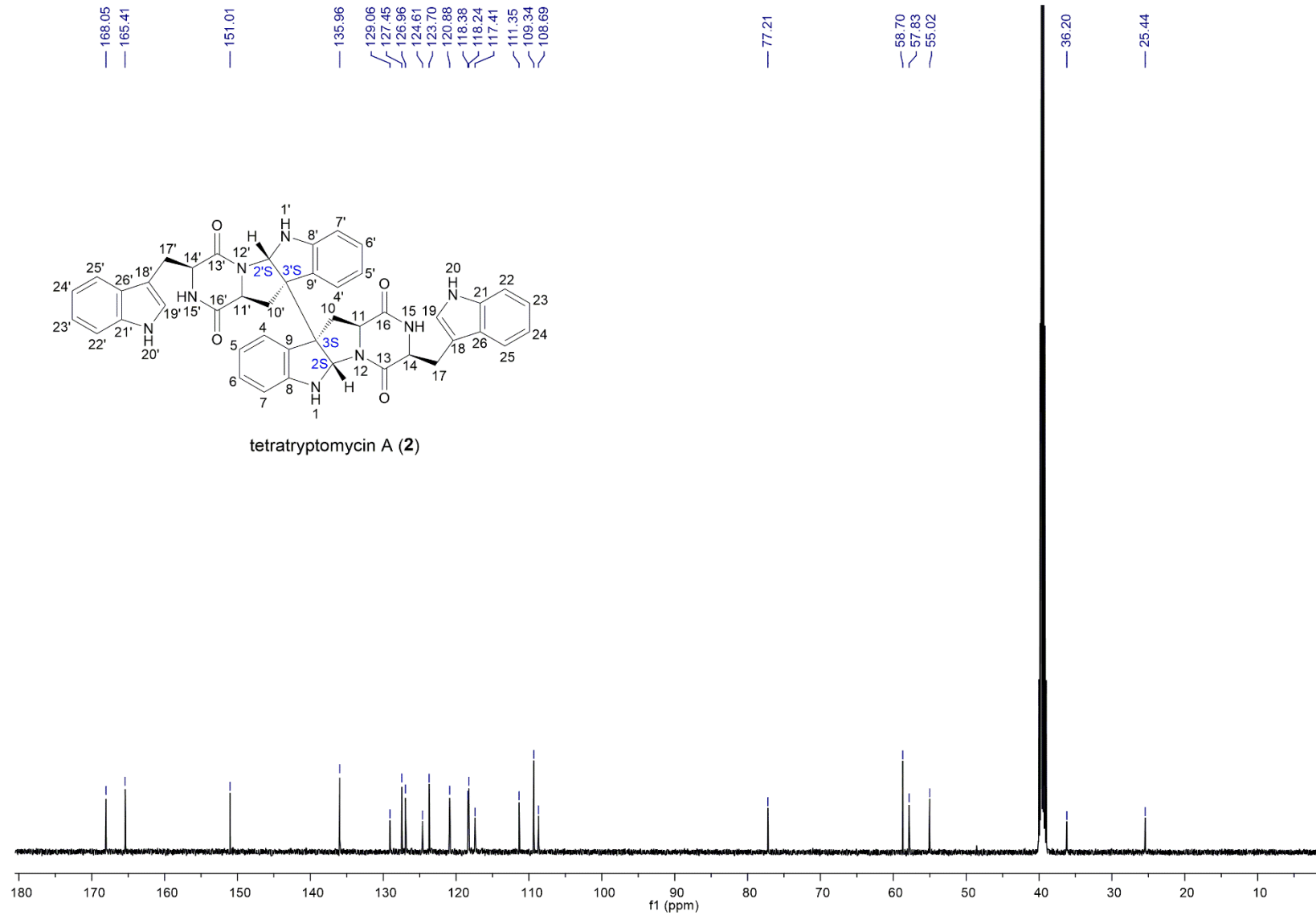


Fig. S5 ^{13}C NMR spectrum of tetratryptomycin A (2) in $\text{DMSO-}d_6$ at 300 K (125 MHz).

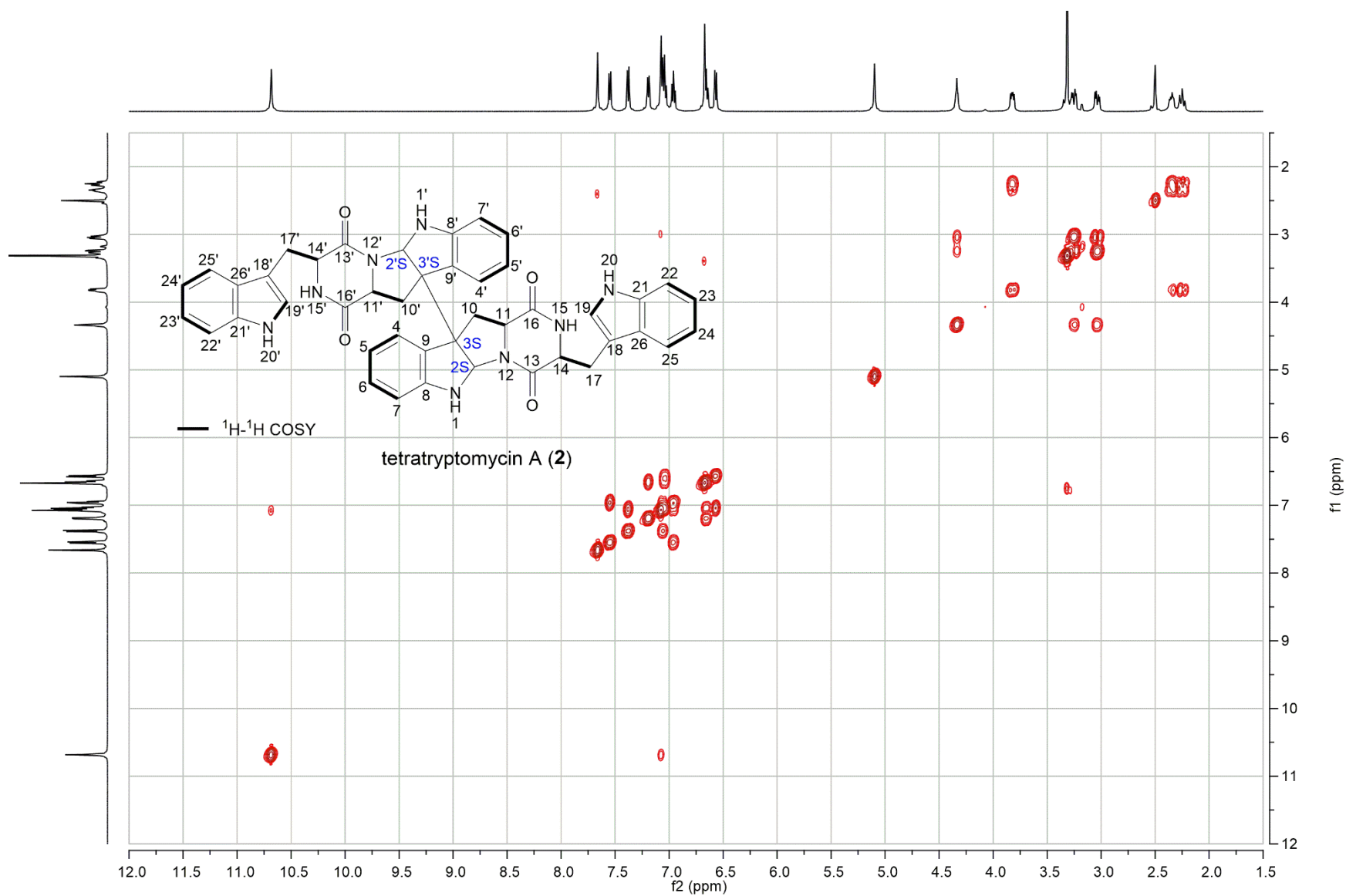


Fig. S6 ^1H - ^1H COSY spectrum of tetratryptomycin A (2) in $\text{DMSO-}d_6$ at 300 K (500 MHz).

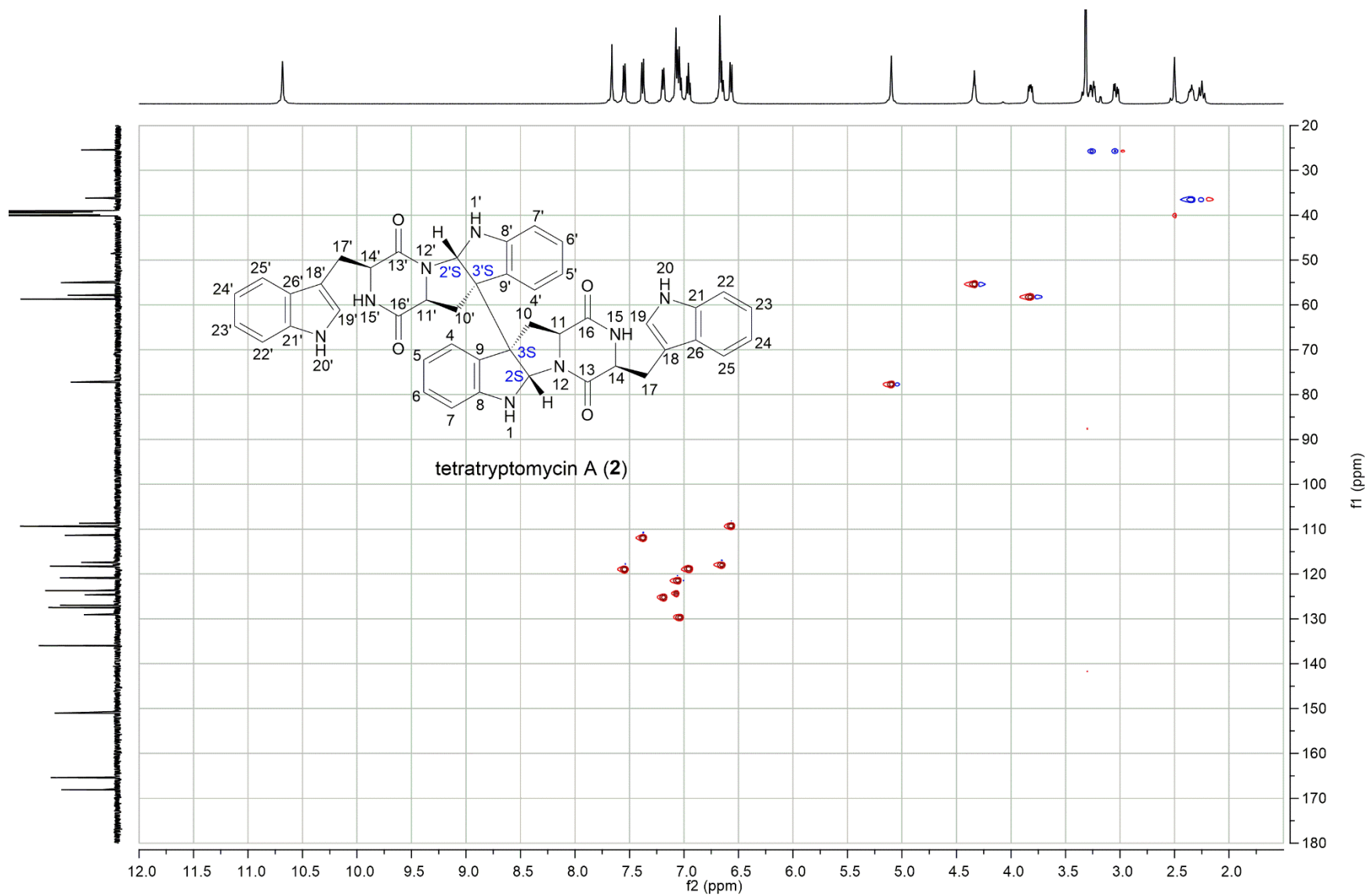


Fig. S7 HSQC spectrum of tetratryptomycin A (**2**) in DMSO- d_6 at 300 K (500 MHz, 125 MHz).

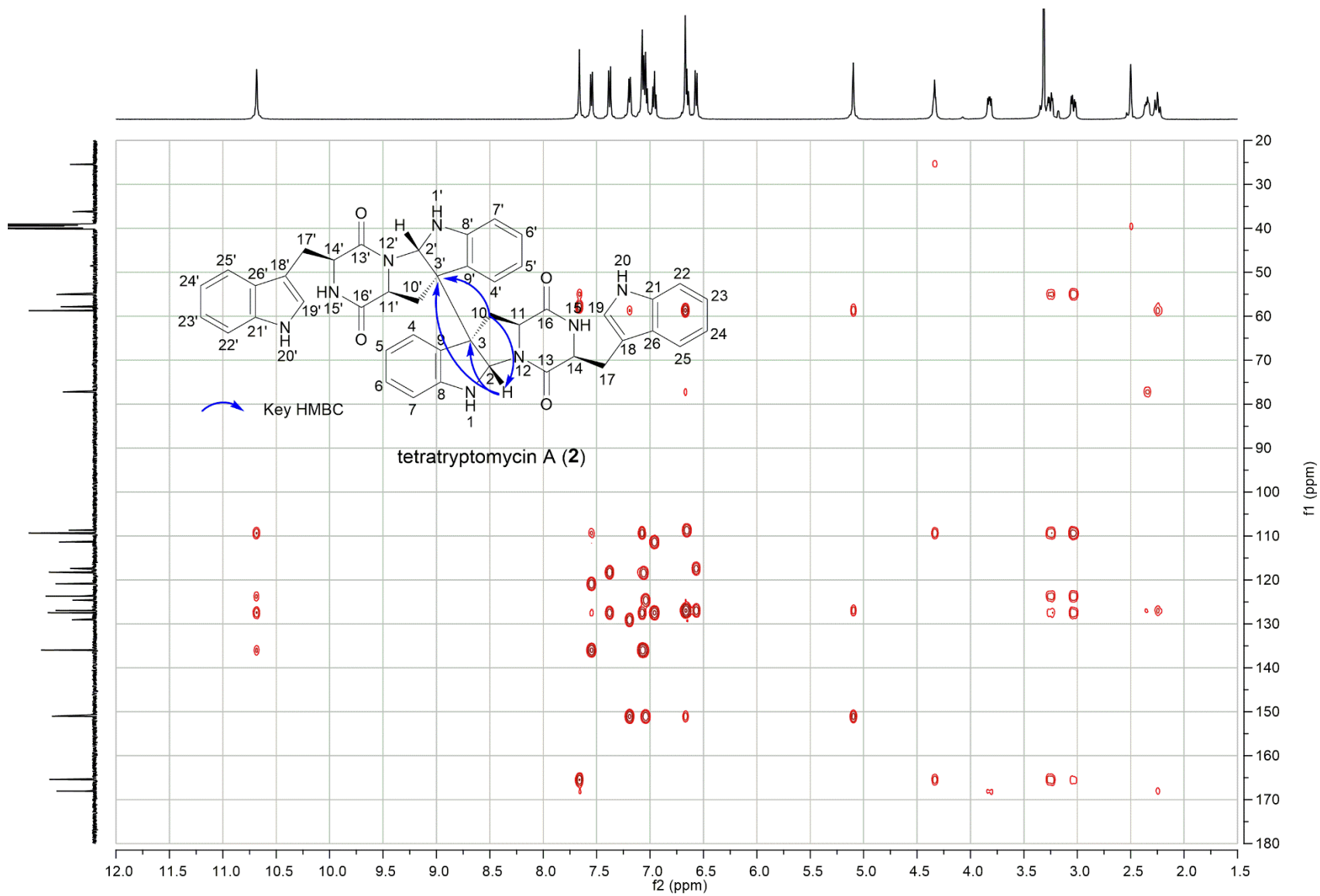


Fig. S8 HMBC spectrum of tetratryptomycin A (2) in DMSO- d_6 at 300 K (500 MHz, 125 MHz).

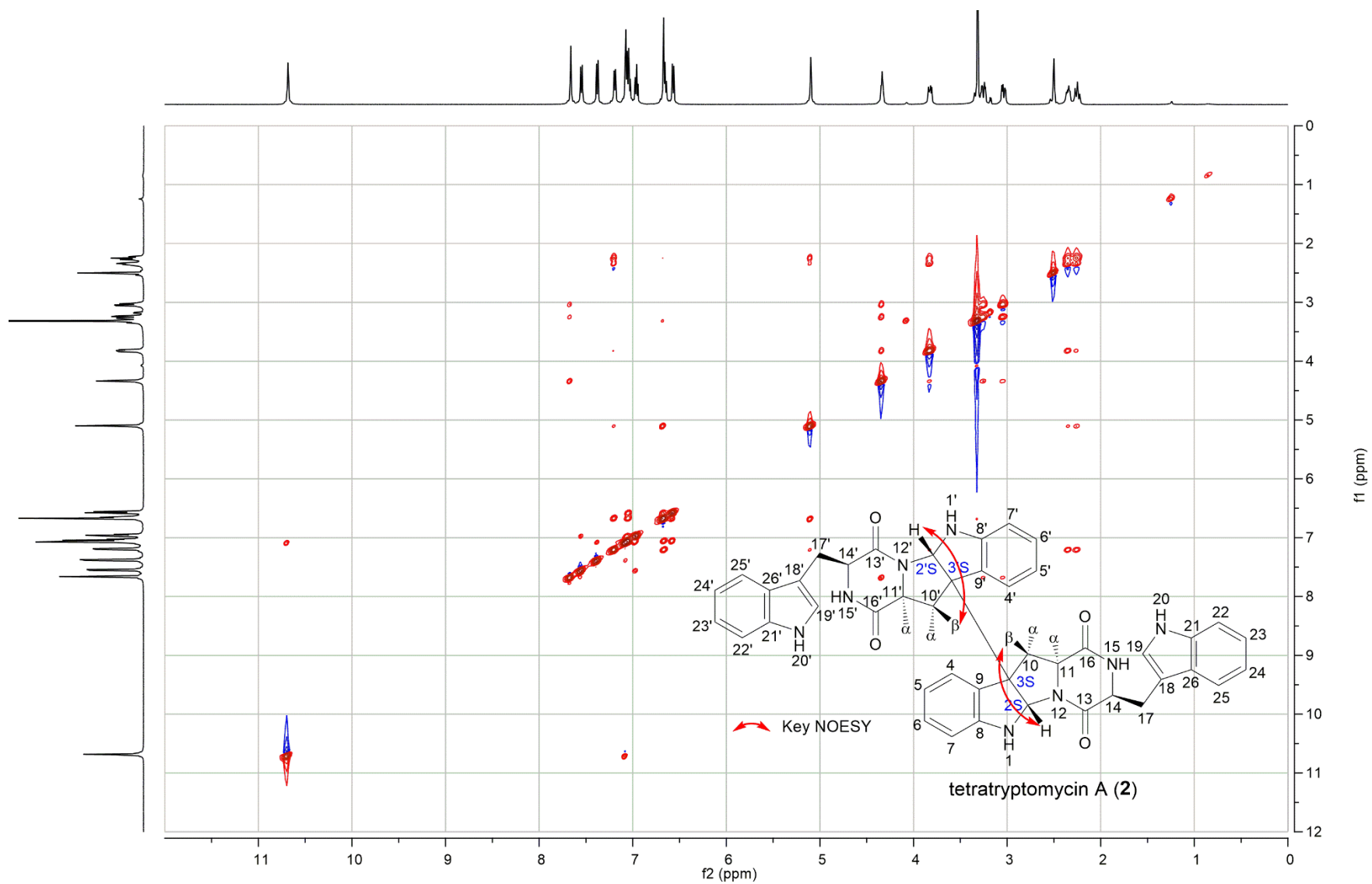


Fig. S9 NOESY spectrum of tetraptomycin A (2) in DMSO- d_6 at 300 K (500 MHz).

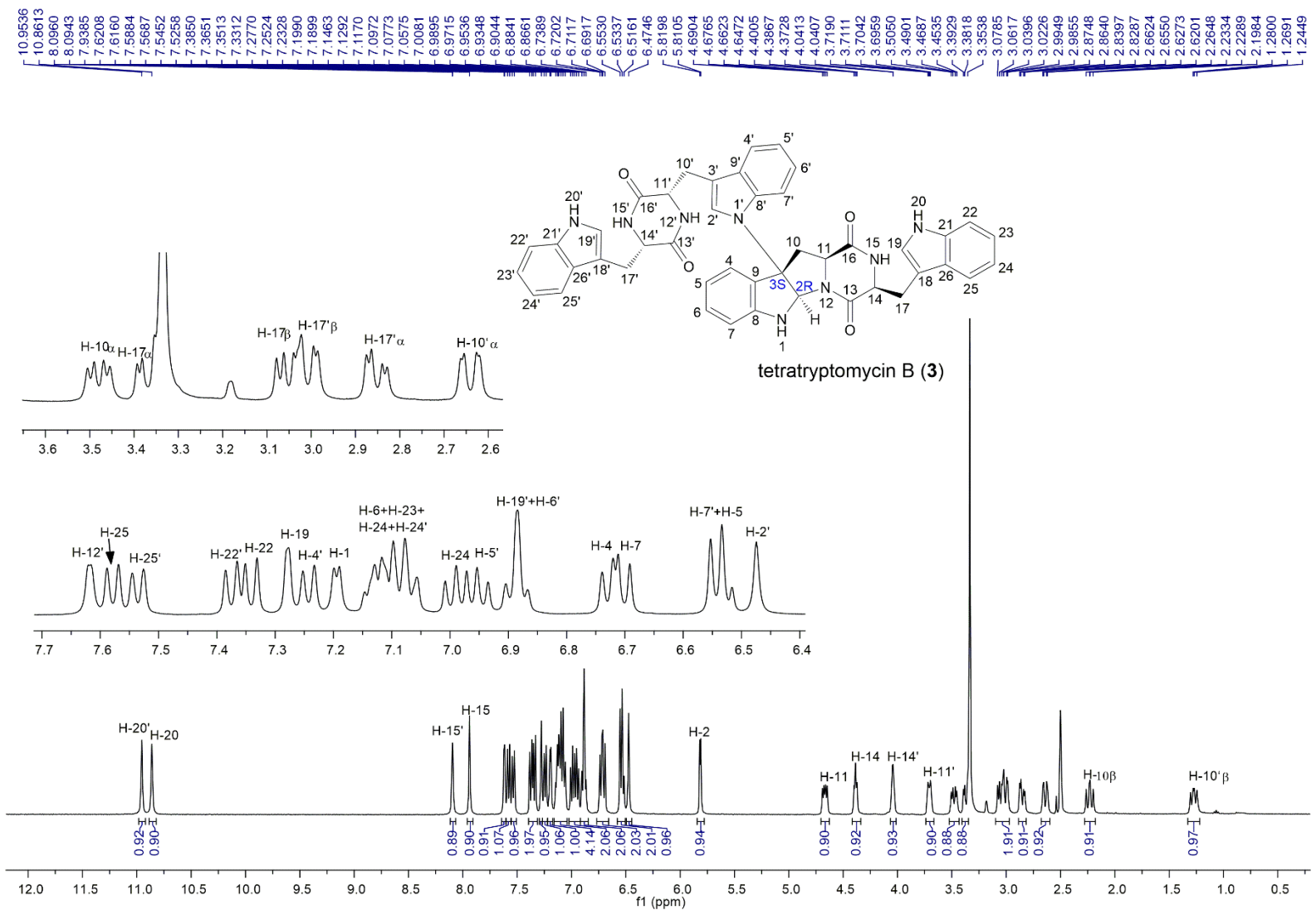


Fig. S10 ¹H NMR spectrum of tetratryptomycin B (3) in DMSO-d₆ at 300 K (500 MHz).

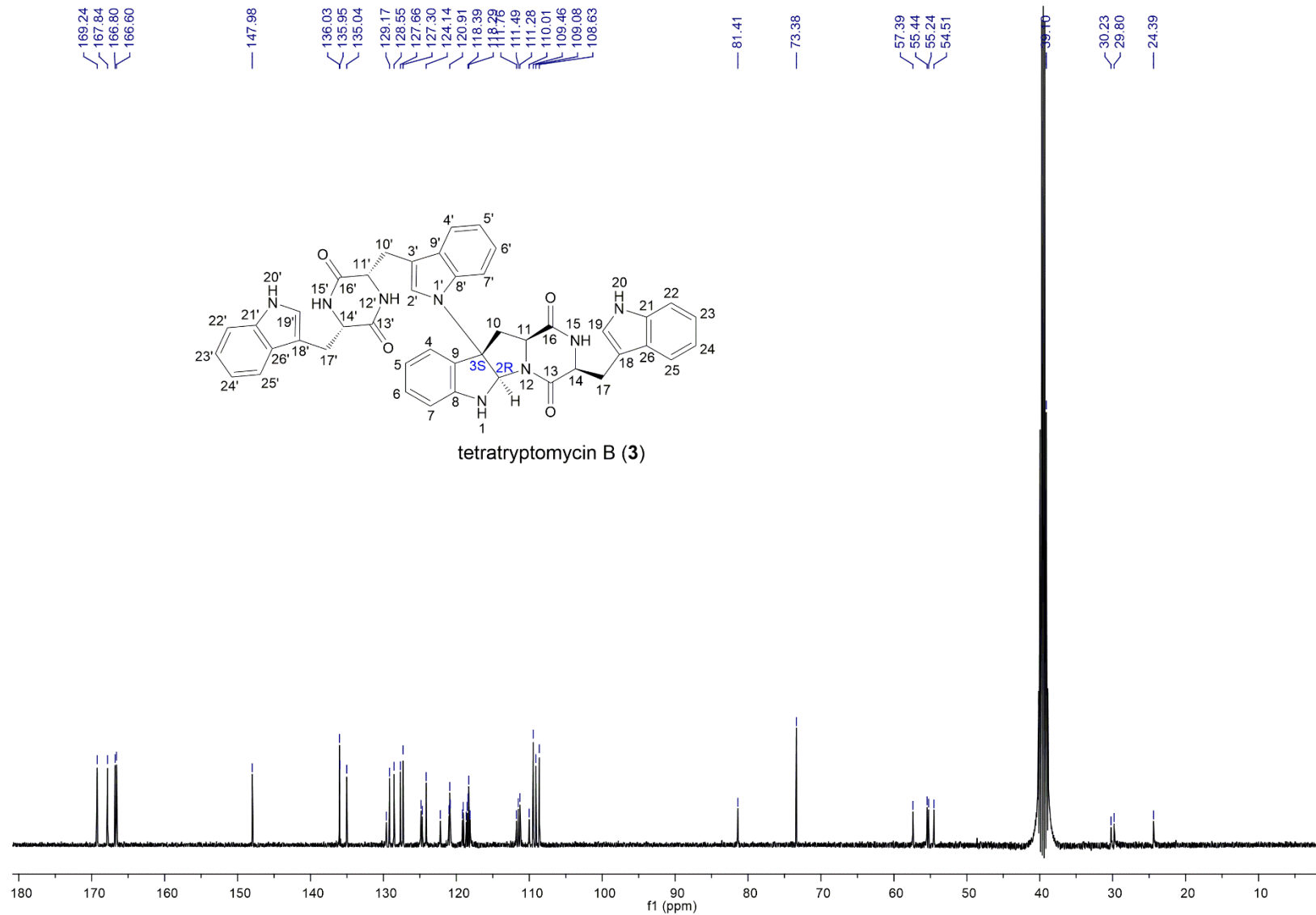


Fig. S11 ^{13}C NMR spectrum of tetratryptomycin B (**3**) in $\text{DMSO-}d_6$ at 300 K (125 MHz).

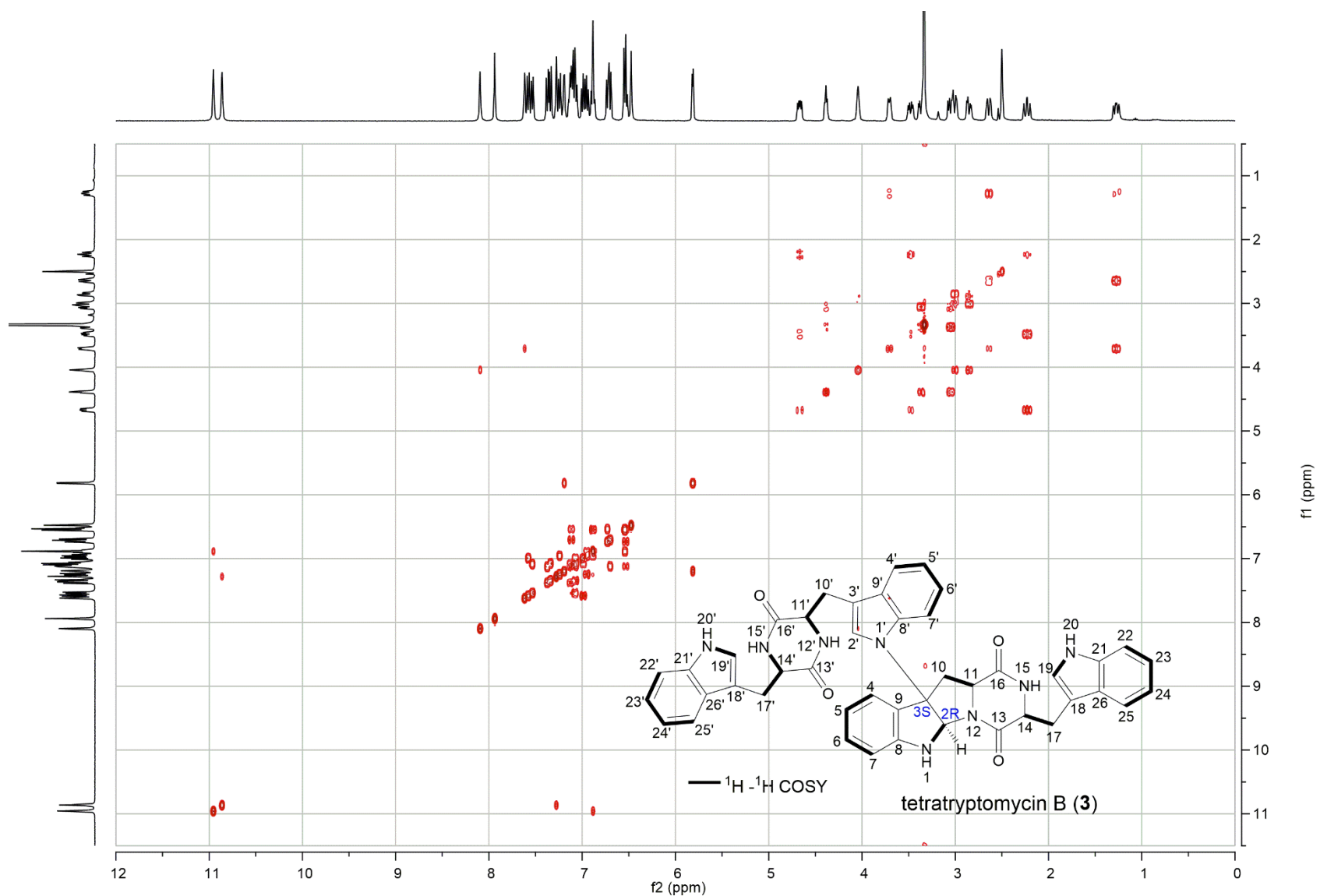


Fig. S12 ^1H - ^1H COSY spectrum of tetraptomycin B (3) in $\text{DMSO-}d_6$ at 300 K (500 MHz).

S23

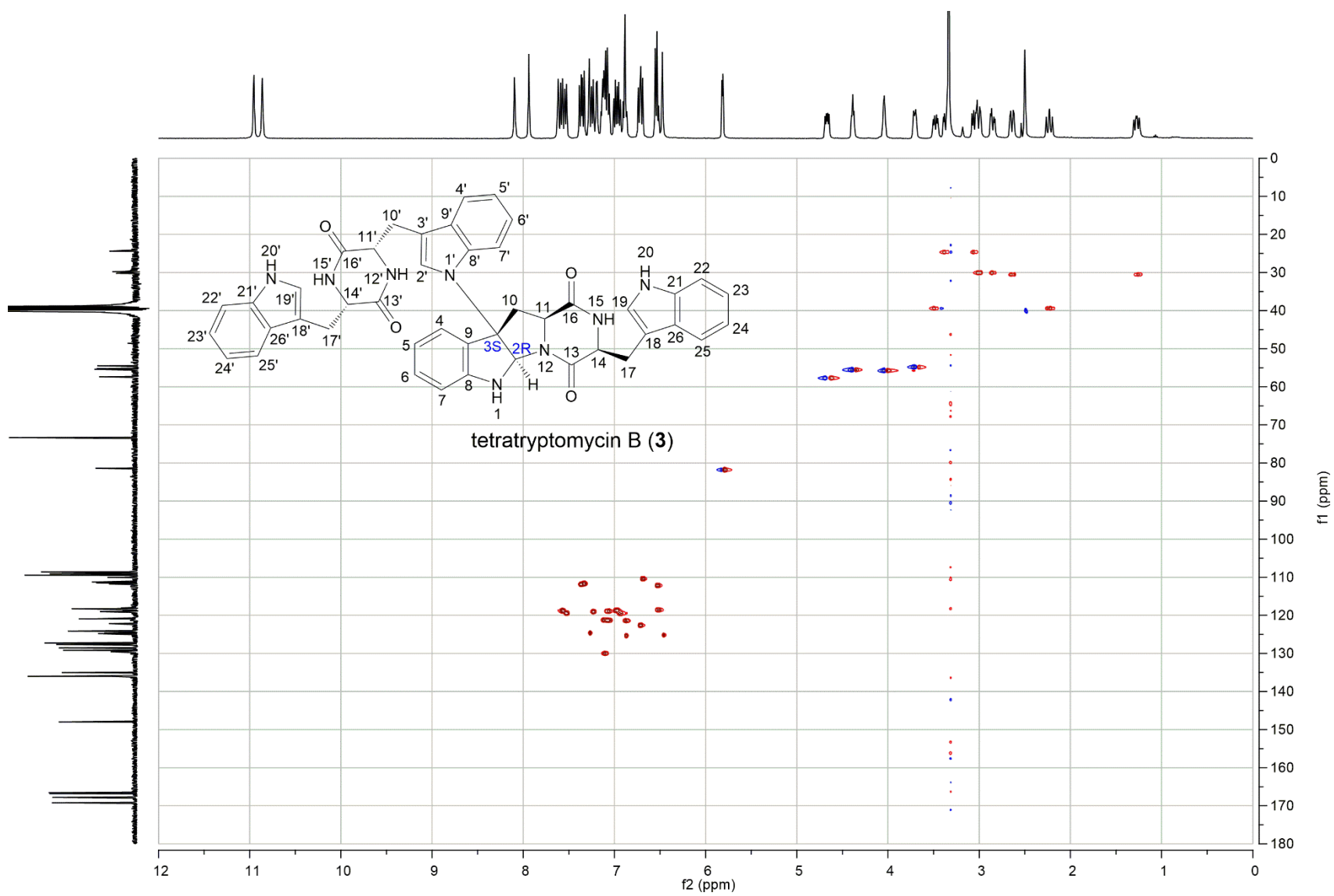


Fig. S13 HSQC spectrum of tetraptomycin B (**3**) in DMSO-*d*₆ at 300 K (500 MHz, 125 MHz).

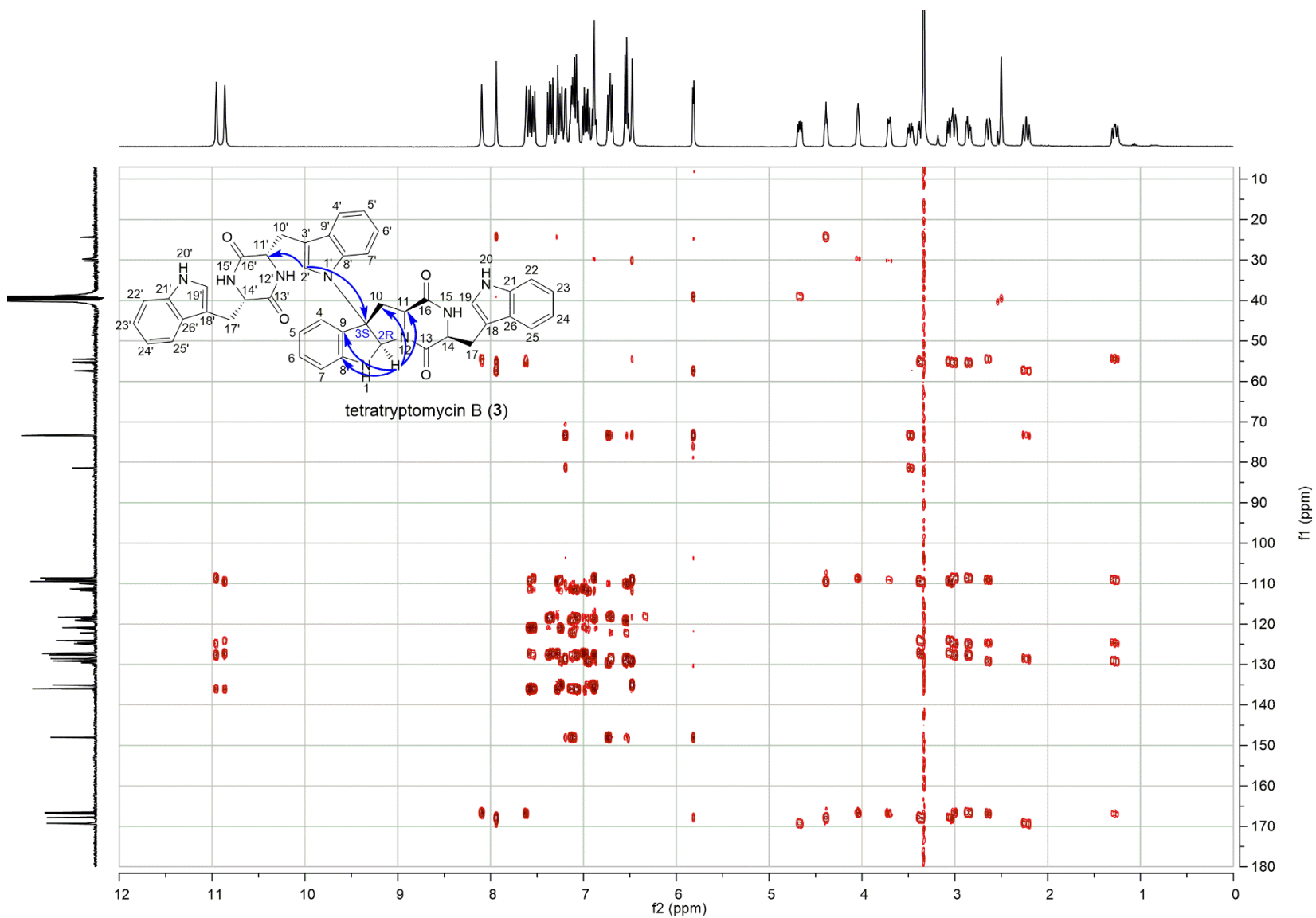


Fig. S14 HMBC spectrum of tetratryptomycin B (**3**) in DMSO- d_6 at 300 K (500 MHz, 125 MHz).

S25

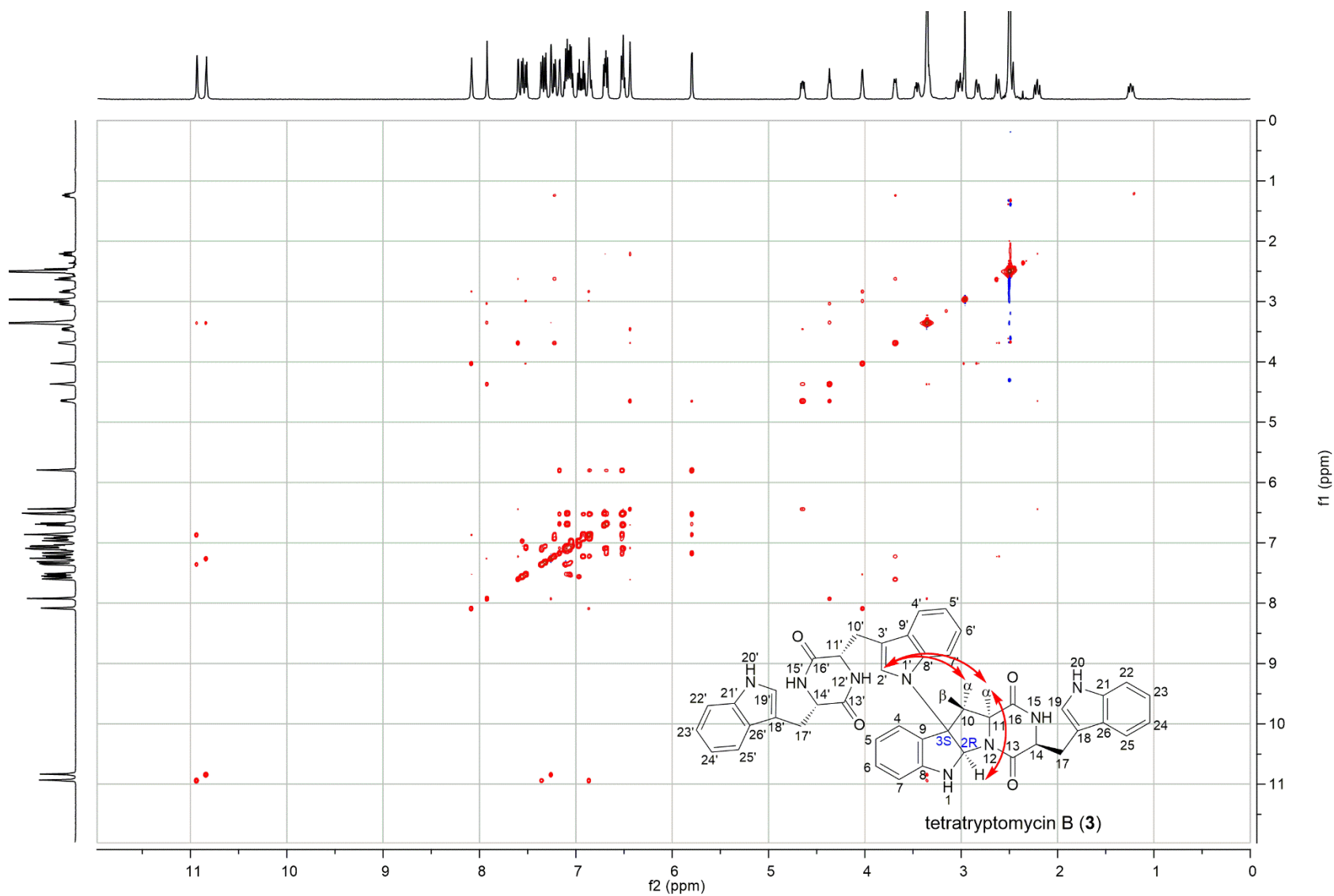


Fig. S15 NOESY spectrum of tetratryptomycin B (**3**) in DMSO-*d*₆ at 300 K (500 MHz).

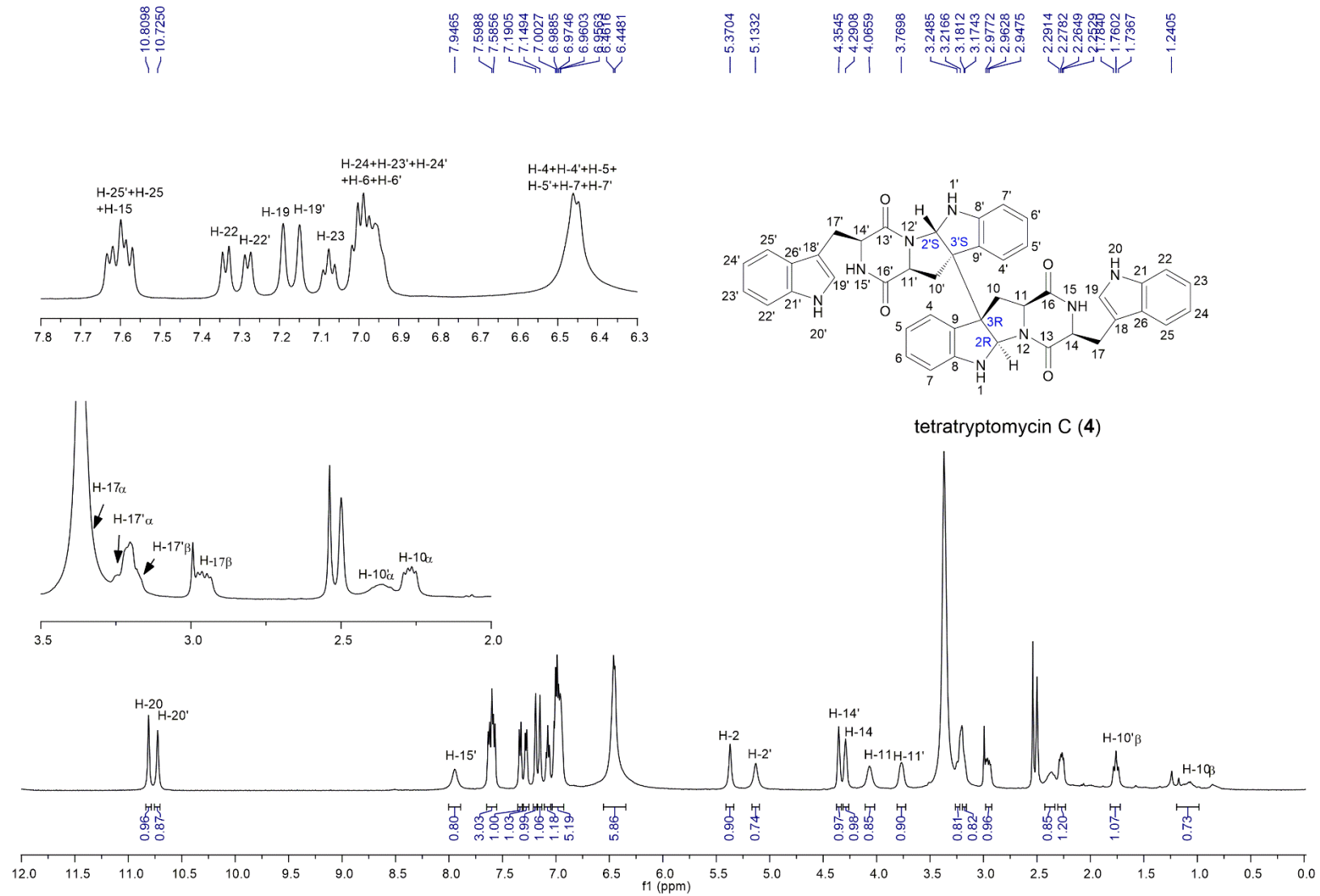


Fig. S16 ¹H NMR spectrum of tetratryptomycin C (4) in DMSO-*d*₆ at 300 K (500 MHz).

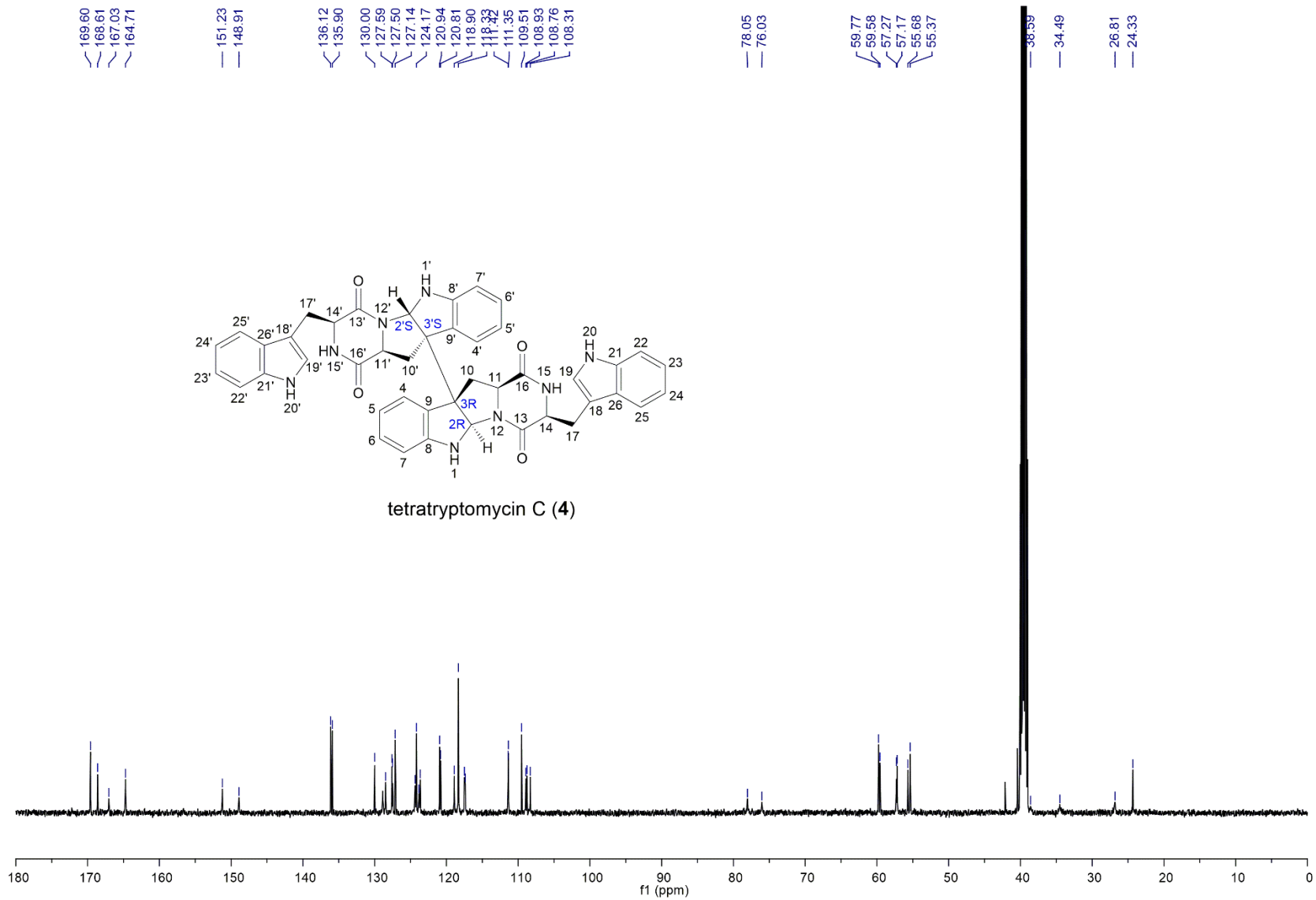


Fig. S17 ^{13}C NMR spectrum of tetratryptomycin C (**4**) in $\text{DMSO-}d_6$ at 300 K (125 MHz).

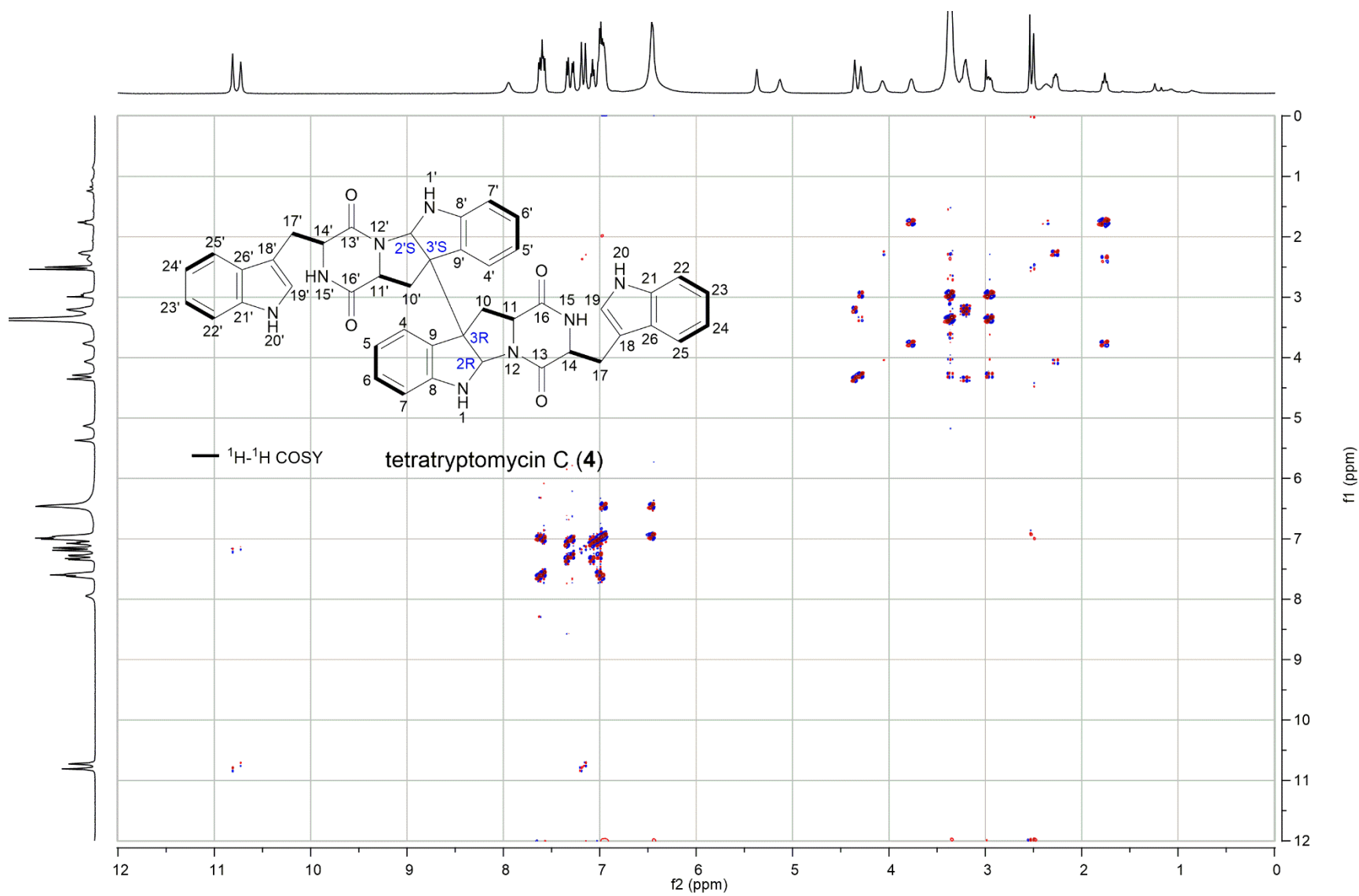


Fig. S18 ^1H - ^1H COSY spectrum of tetratryptomycin C (4) in $\text{DMSO-}d_6$ at 300 K (500 MHz).

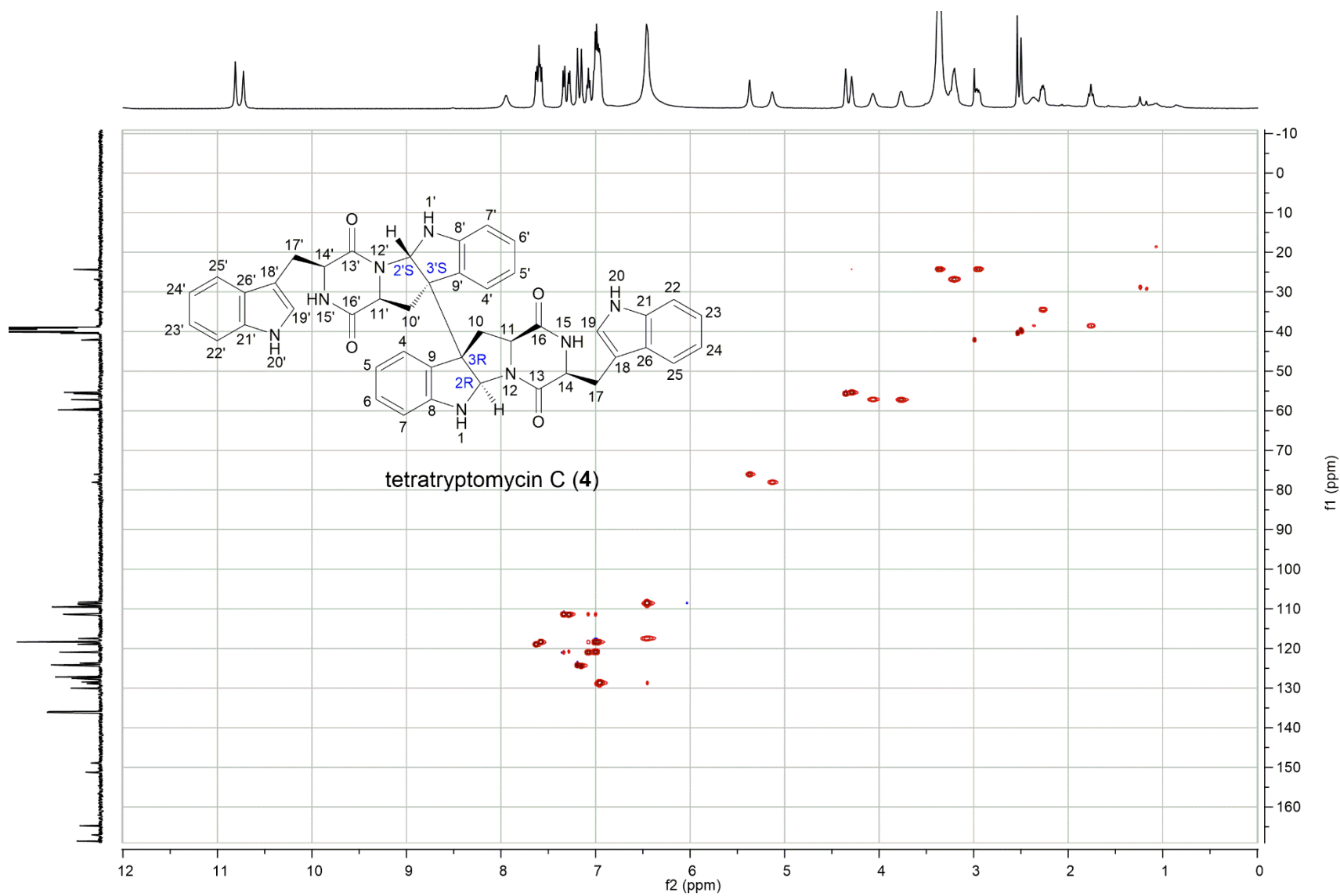


Fig. S19 HSQC spectrum of tetratryptomycin C (4) in DMSO-*d*₆ at 300 K (500 MHz, 125 MHz).

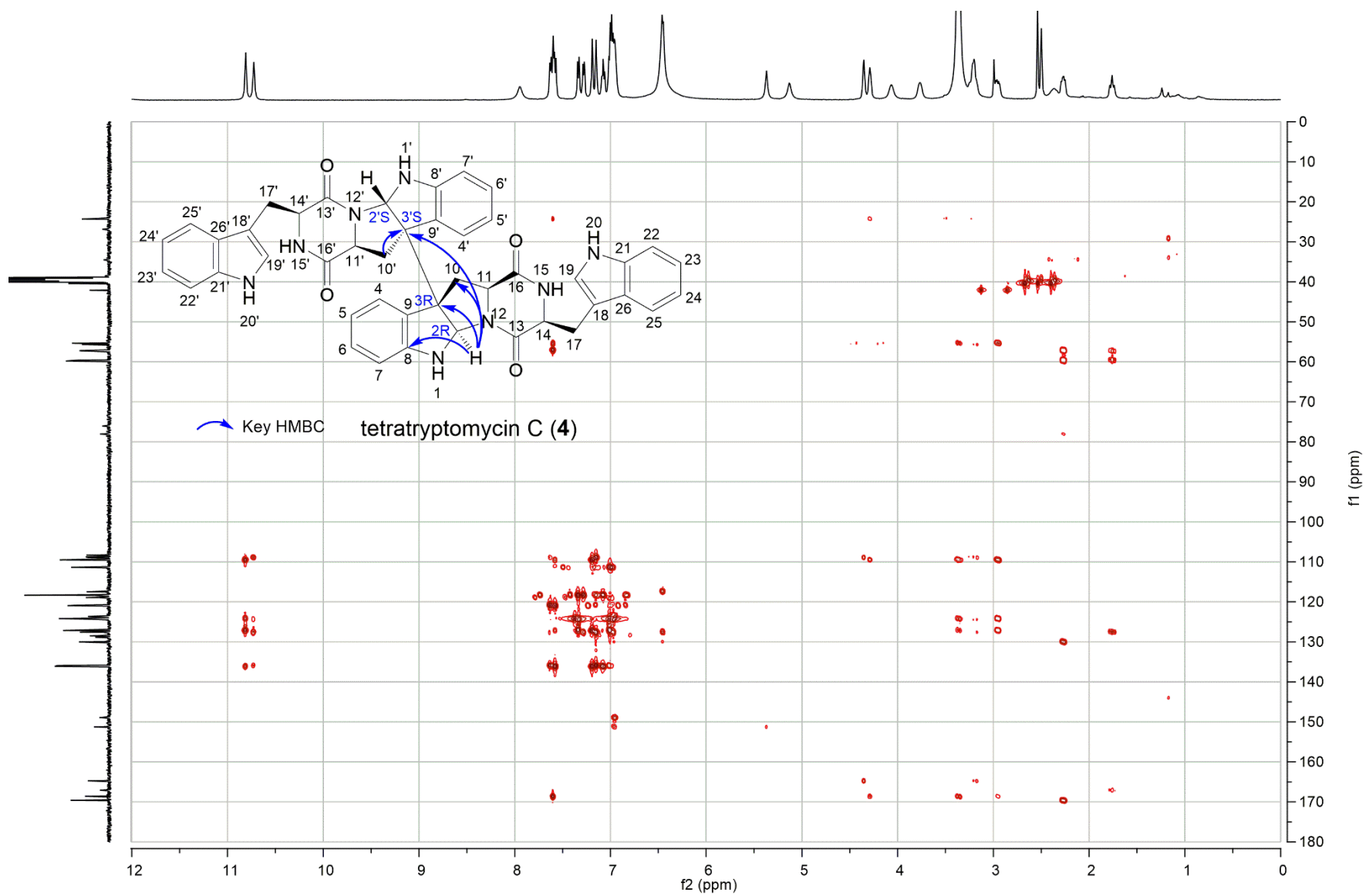


Fig. S20 HMBC spectrum of tetratryptomycin C (**4**) in $\text{DMSO-}d_6$ at 300 K (500 MHz, 125 MHz).

S31

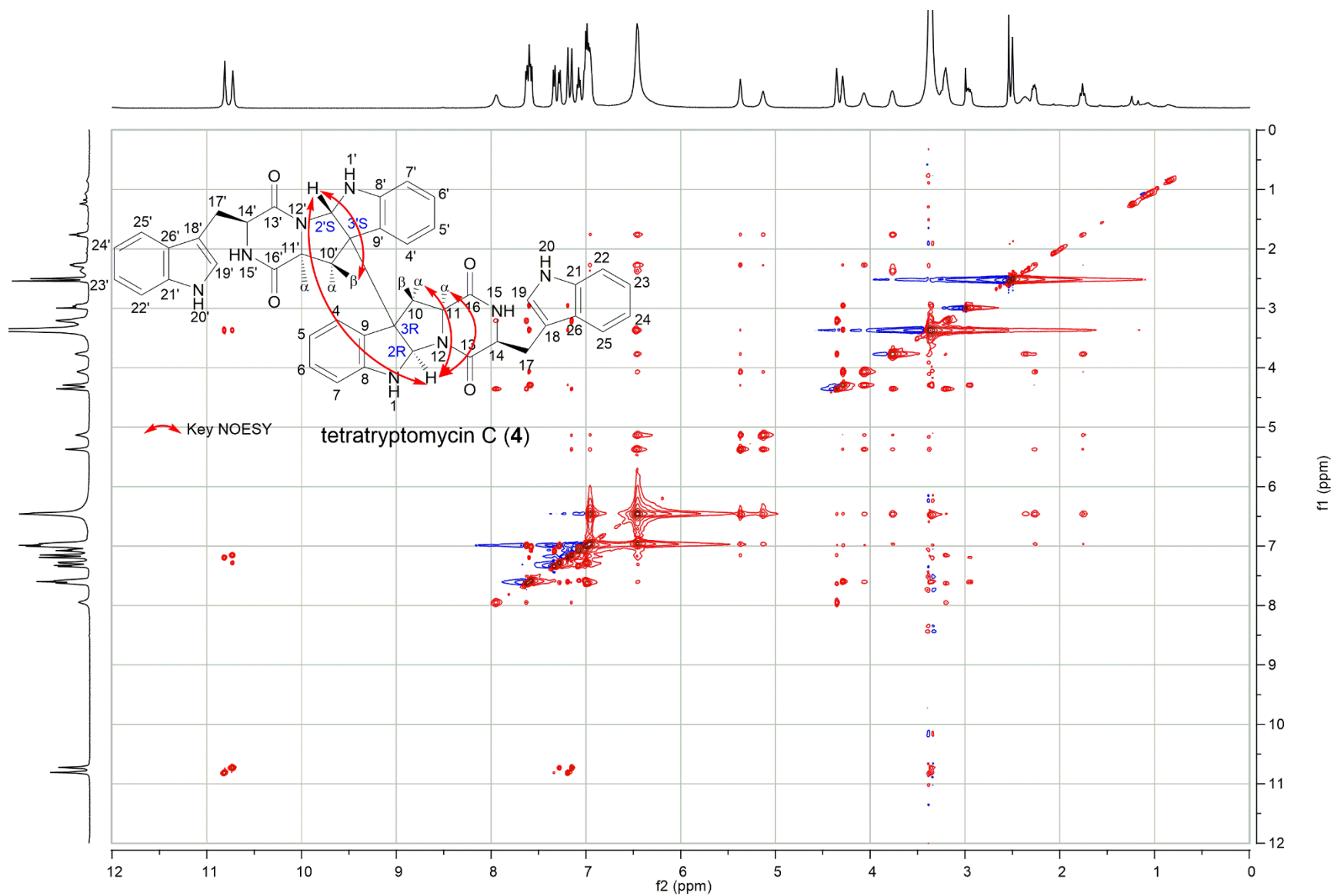


Fig. S21 NOSEY spectrum of tetraptomycin C (4) in DMSO- d_6 at 300 K (500 MHz).

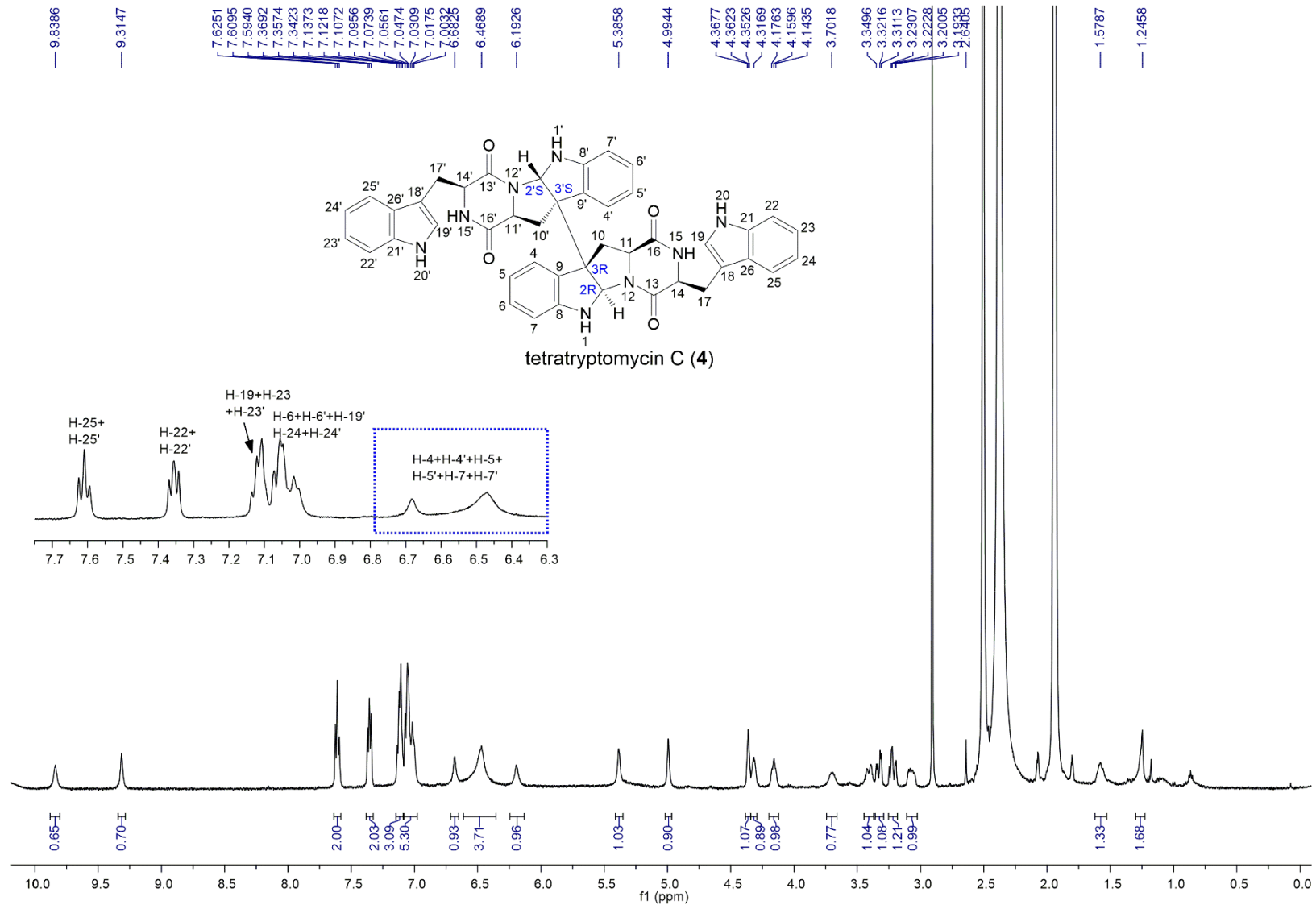


Fig. S22 ¹H NMR spectrum of tetraptomycin C (4) in acetonitrile-*d*₃ at 273 K (500 MHz).

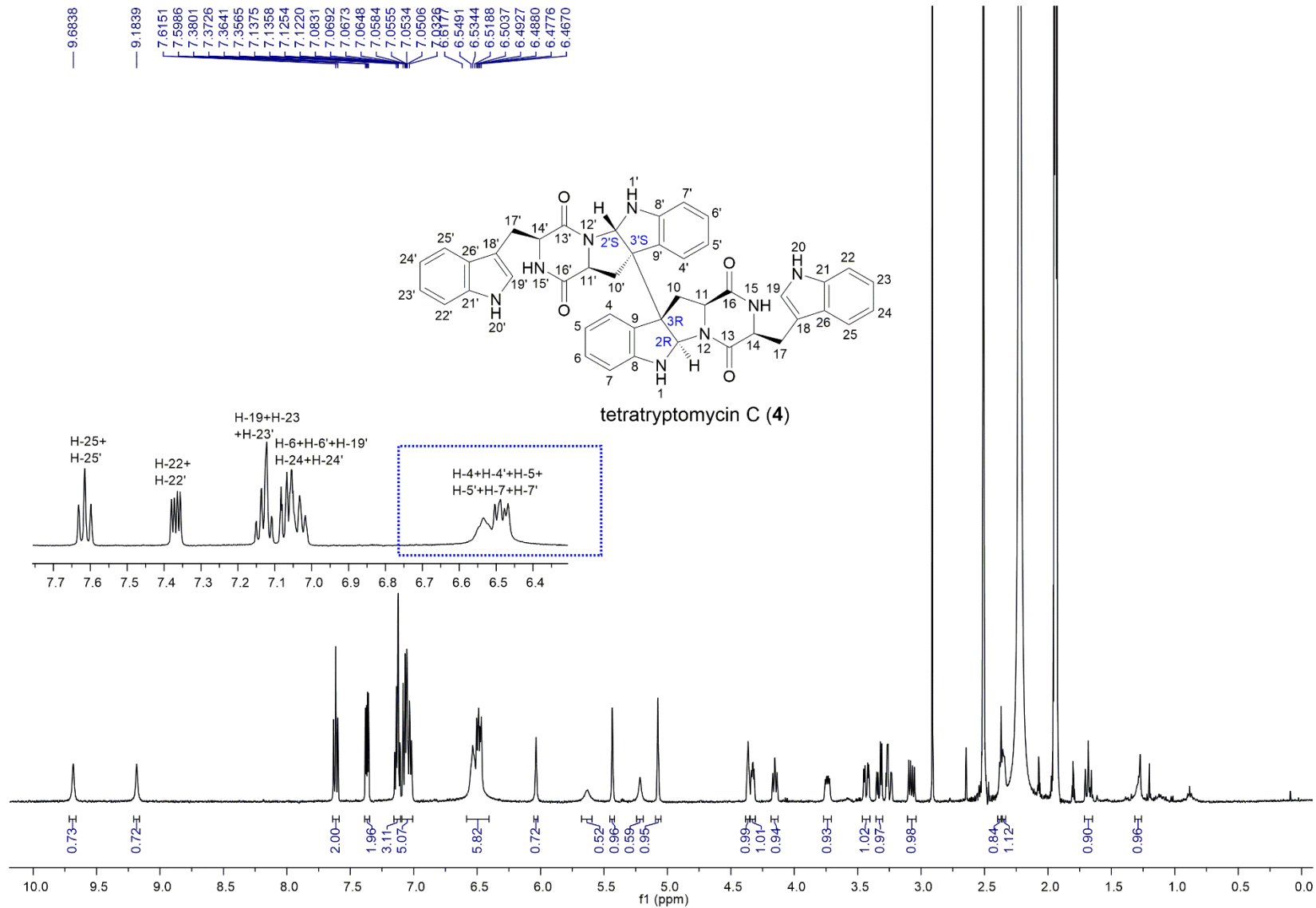


Fig. S23 ^1H NMR spectrum of tetratryptomycin C (4) in acetonitrile- d_3 at 300 K (500 MHz).

S34

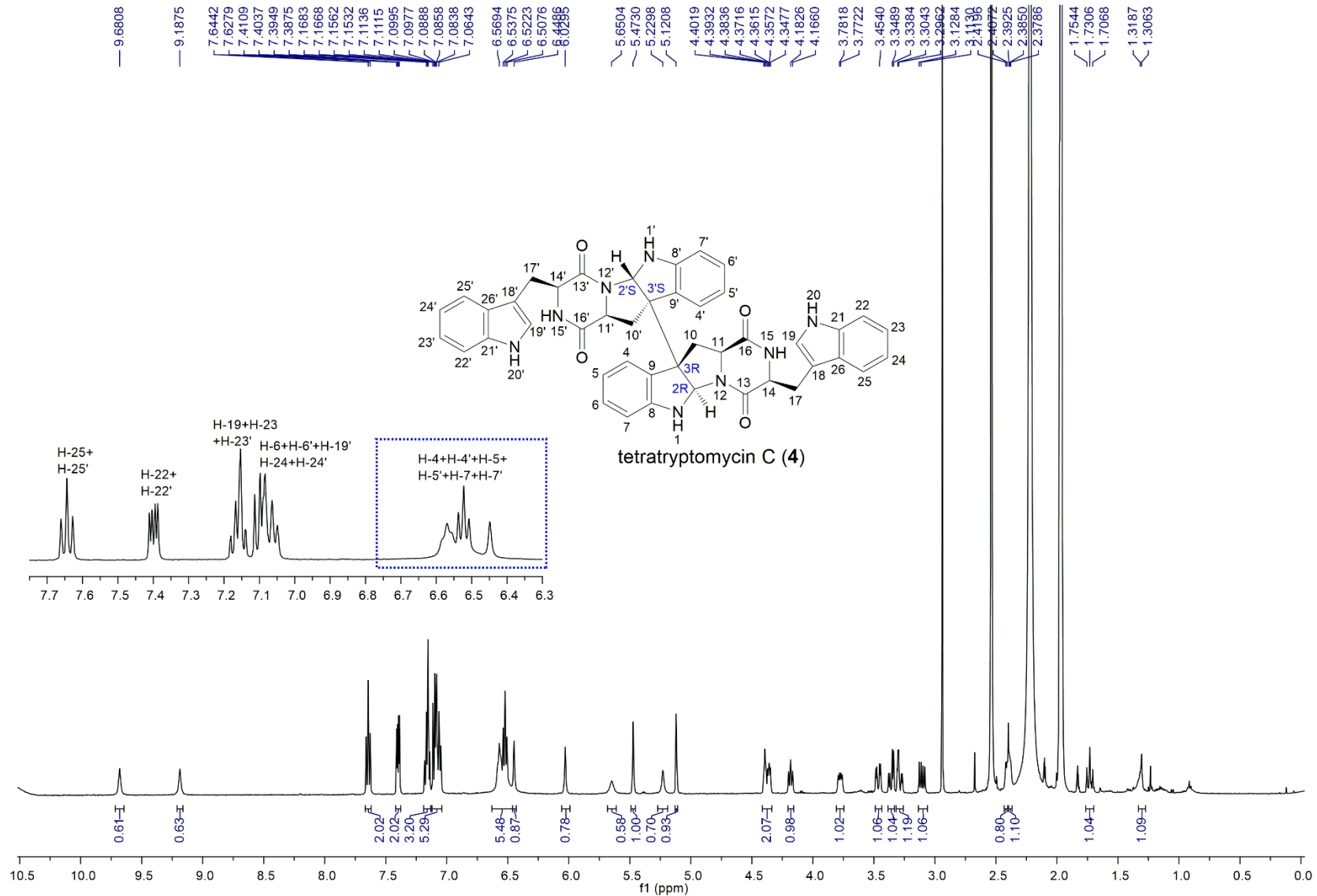


Fig. S24 ^1H NMR spectrum of tetratryptomycin C (4) in acetonitrile- d_3 at 310 K (500 MHz).

S35

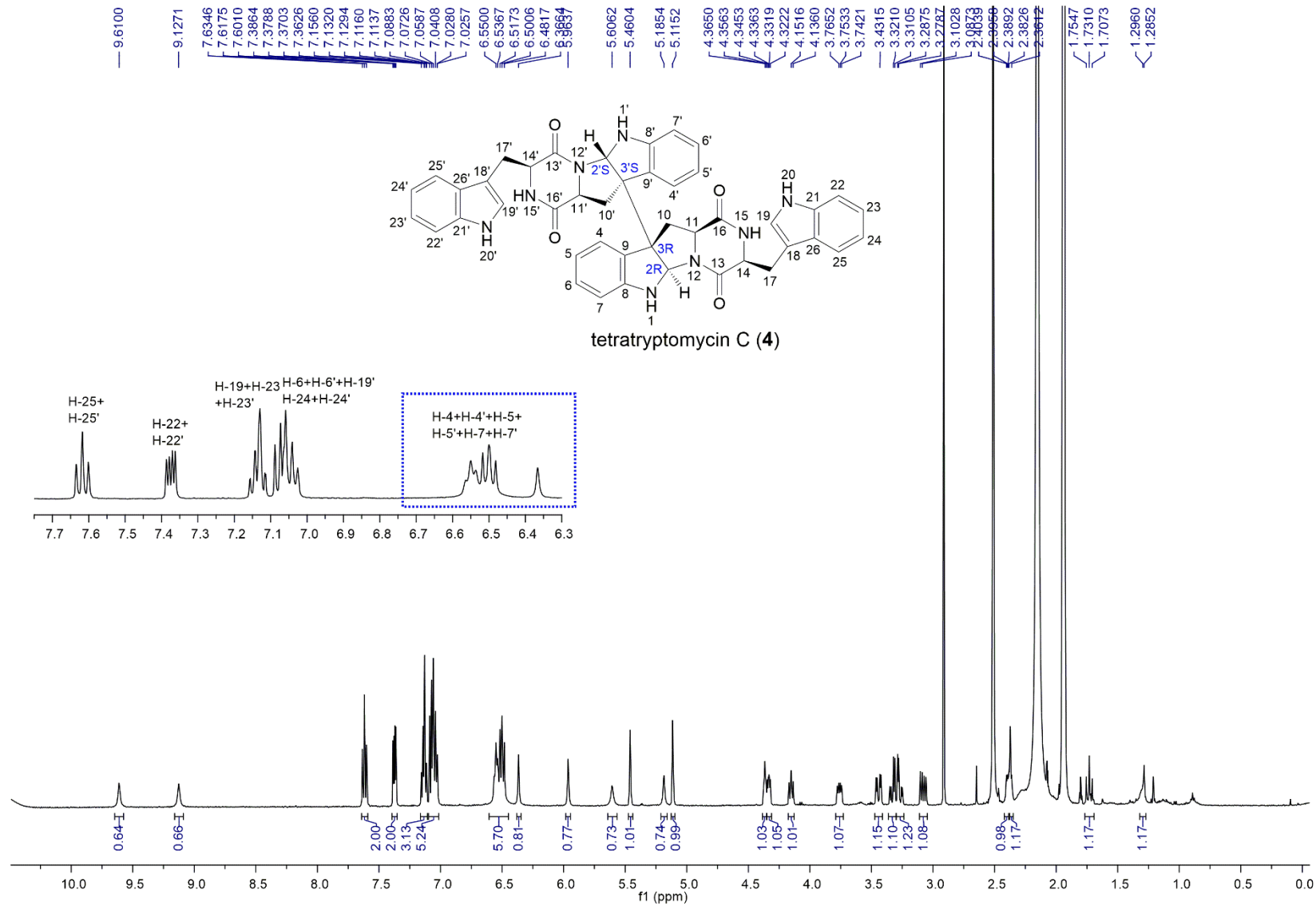


Fig. S25 ^1H NMR spectrum of tetratryptomycin C (4) in acetonitrile- d_3 at 320 K (500 MHz).

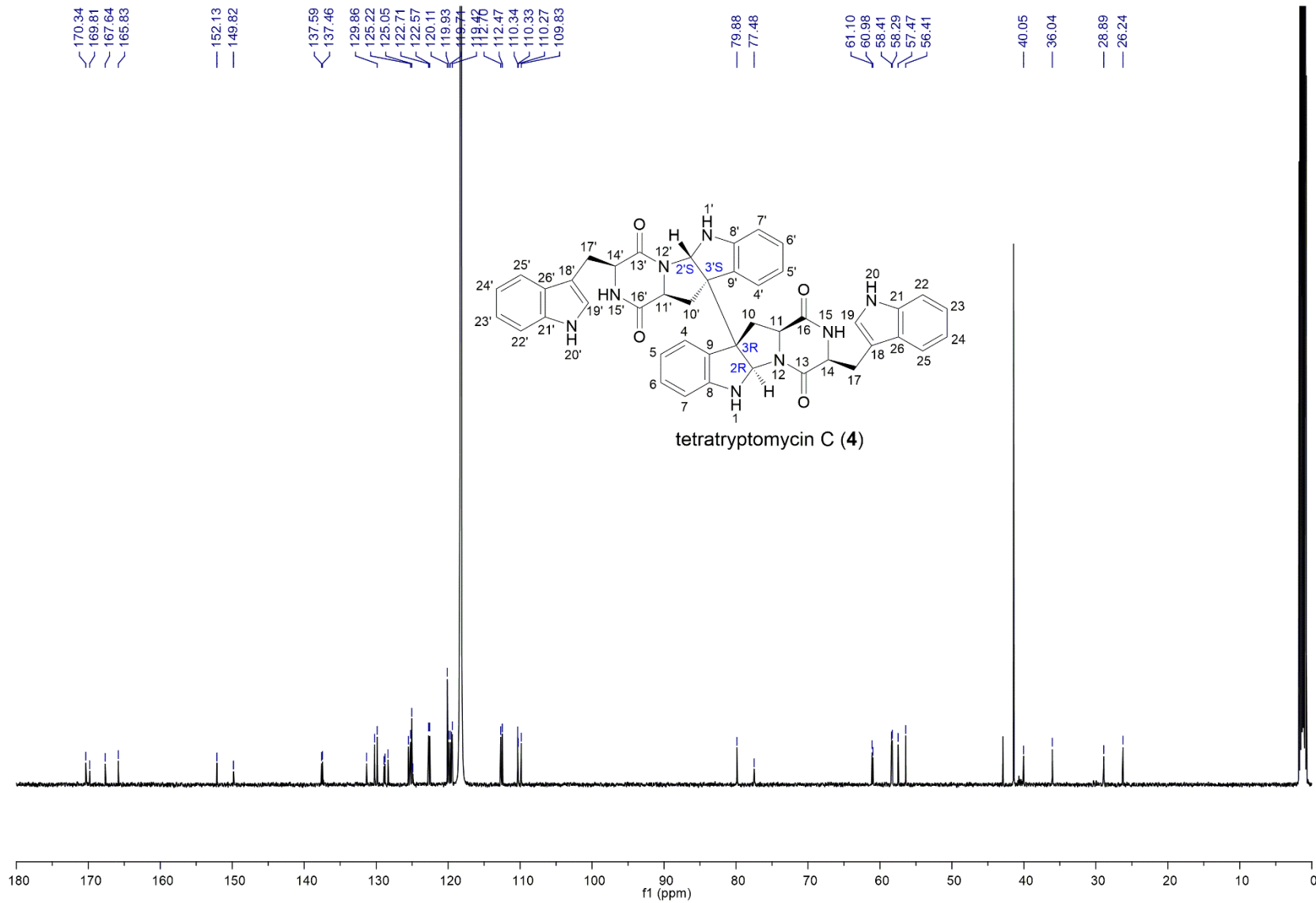


Fig. S26 ^{13}C NMR spectrum of tetratryptomycin C (**4**) in acetonitrile- d_3 at 310 K (125 MHz).

S37

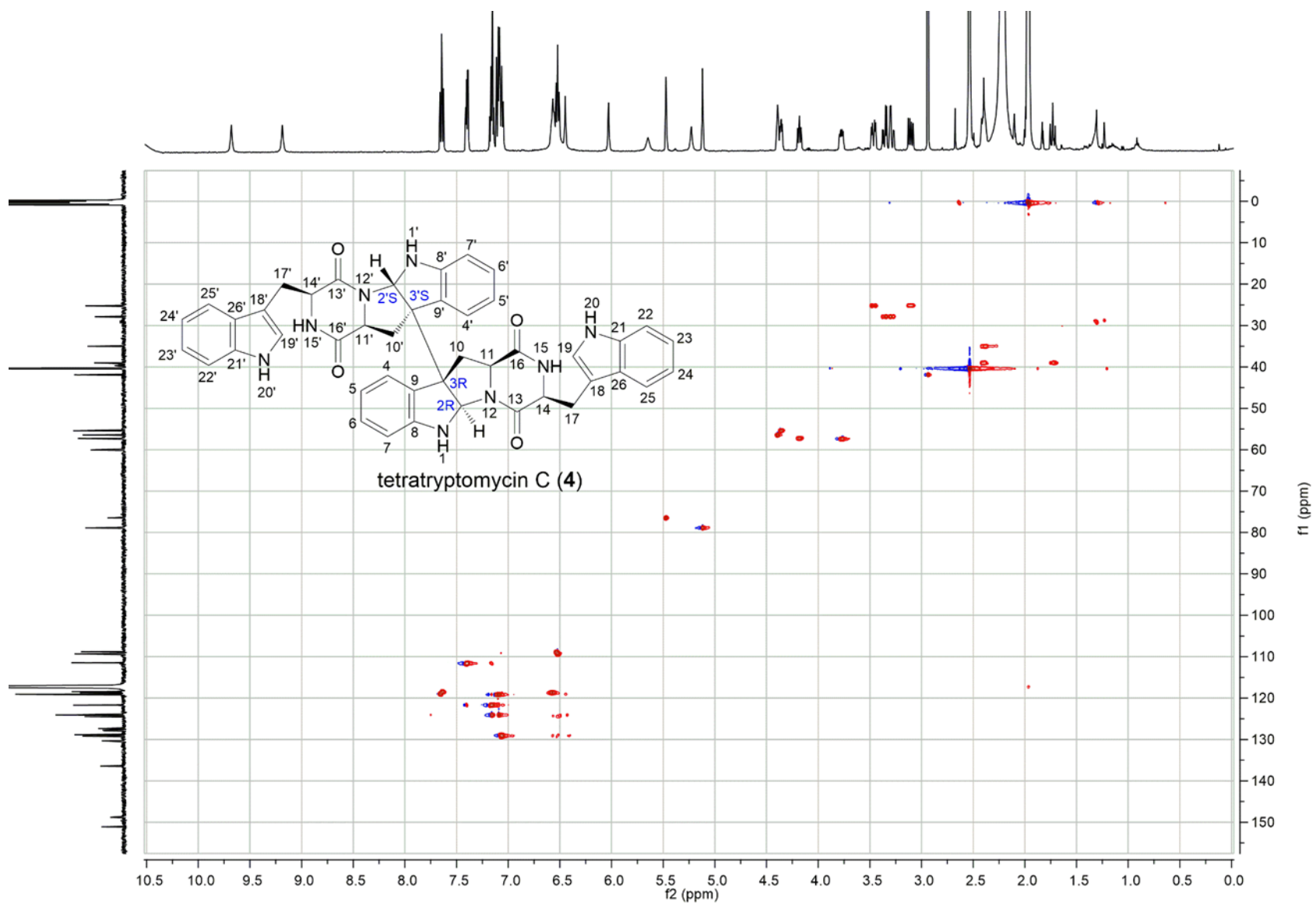


Fig. S27 HSQC spectrum of tetratryptomycin C (4) in acetonitrile-*d*₃ at 310 K (500 MHz, 125 MHz).

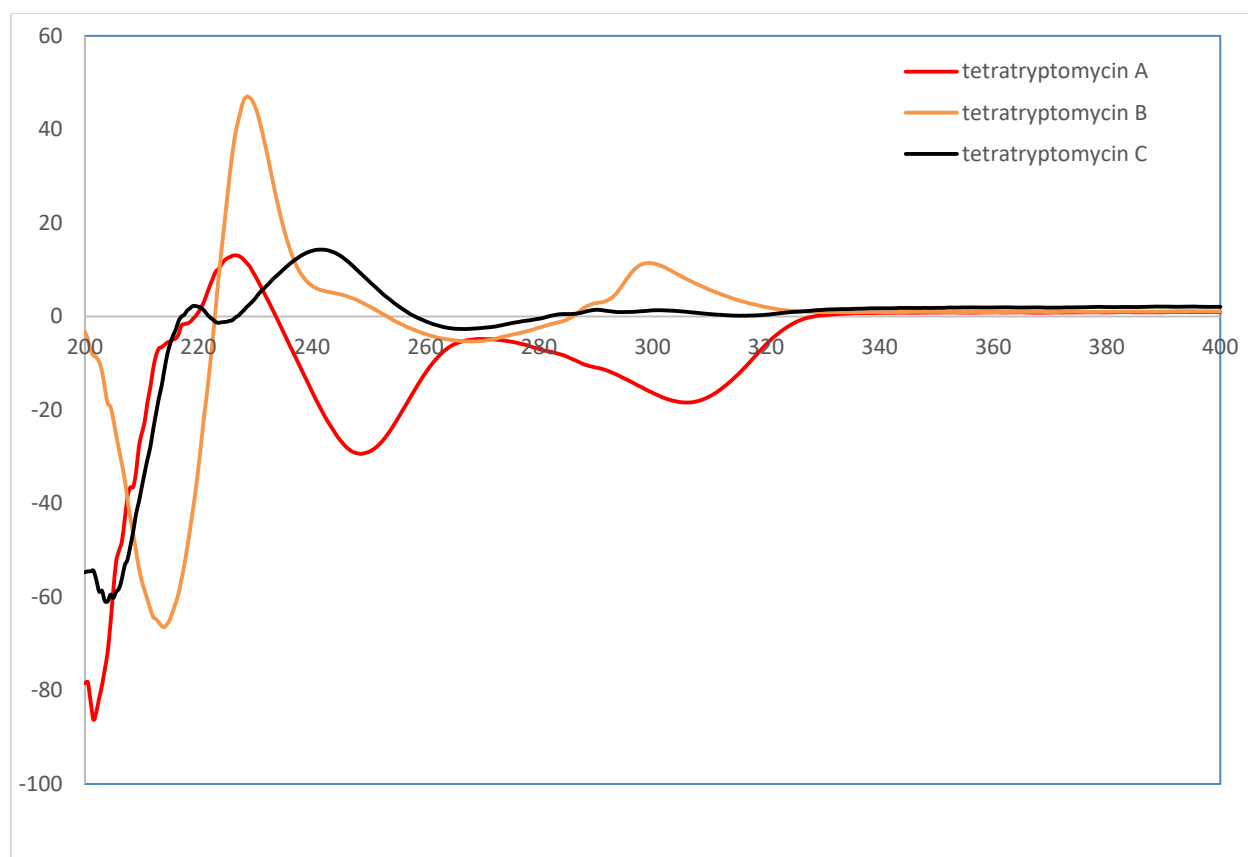


Fig. S28 CD spectra of tetratryptomycins.

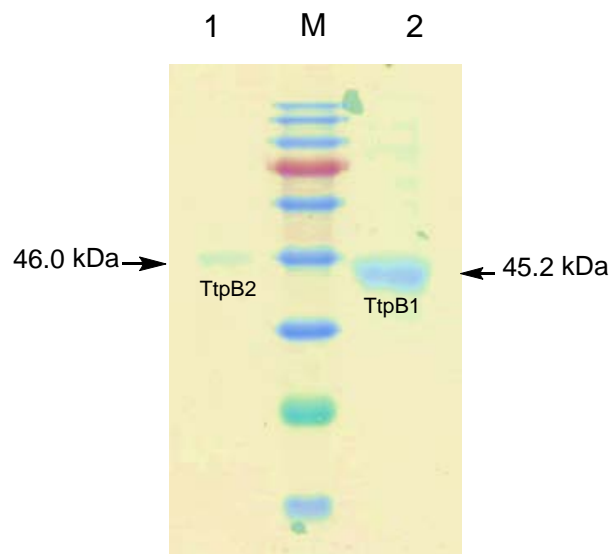


Fig. S29 SDS-PAGE analysis of the purified P450s.

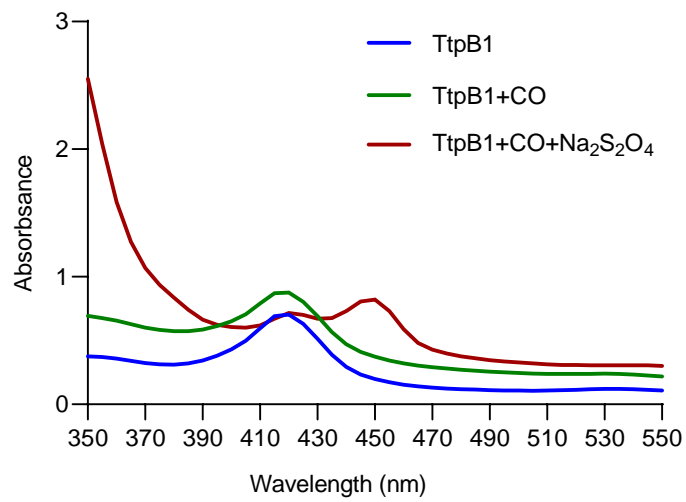


Fig. S30 UV-Vis spectroscopic analysis of TtpB1.

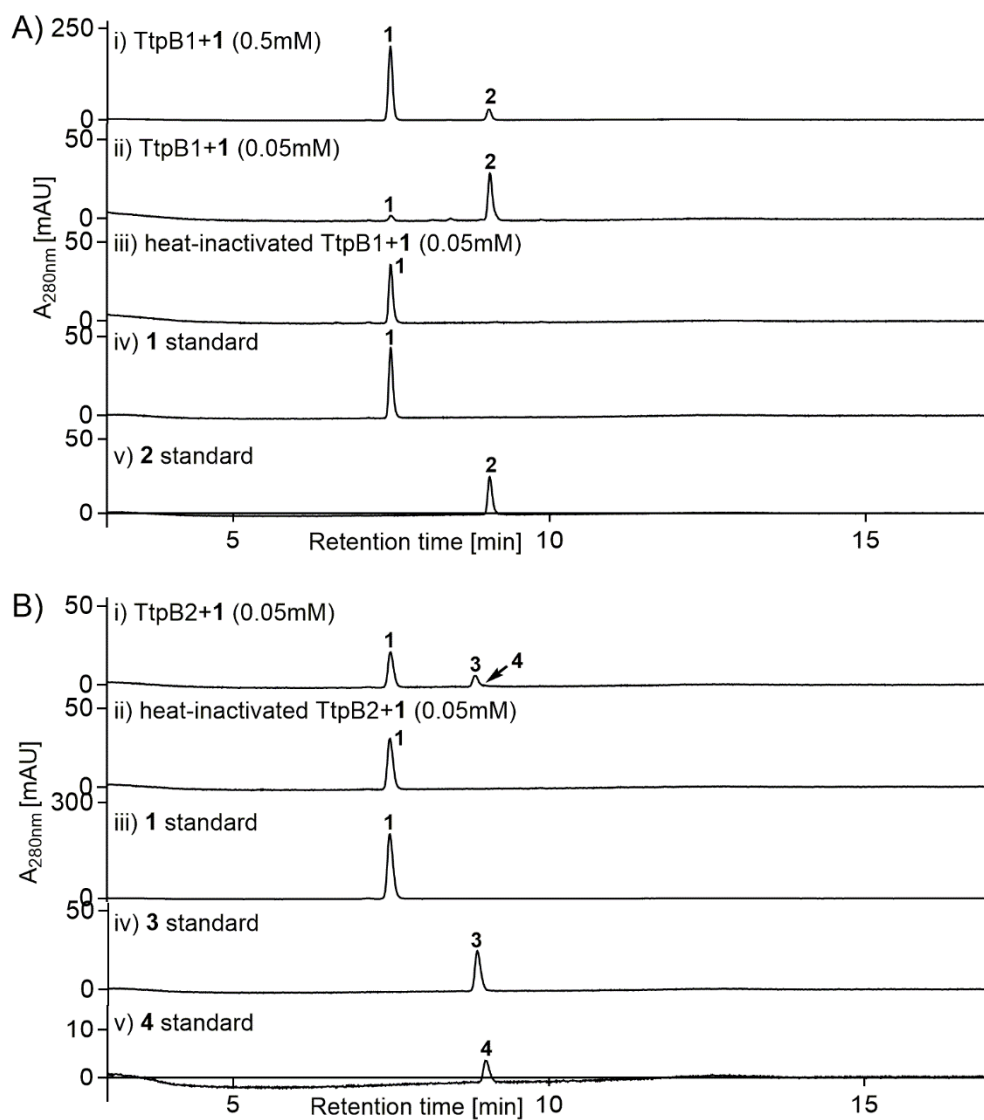


Fig. S31 HPLC analysis of the enzyme assays with TtpB1 (A) and TtpB2 (B).

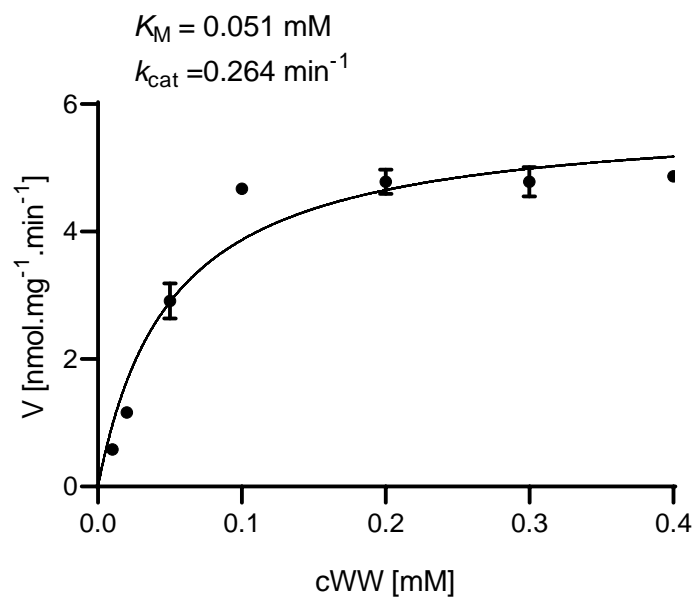


Fig. S32 Determination of kinetic parameters of TtpB1 for cWW (1).

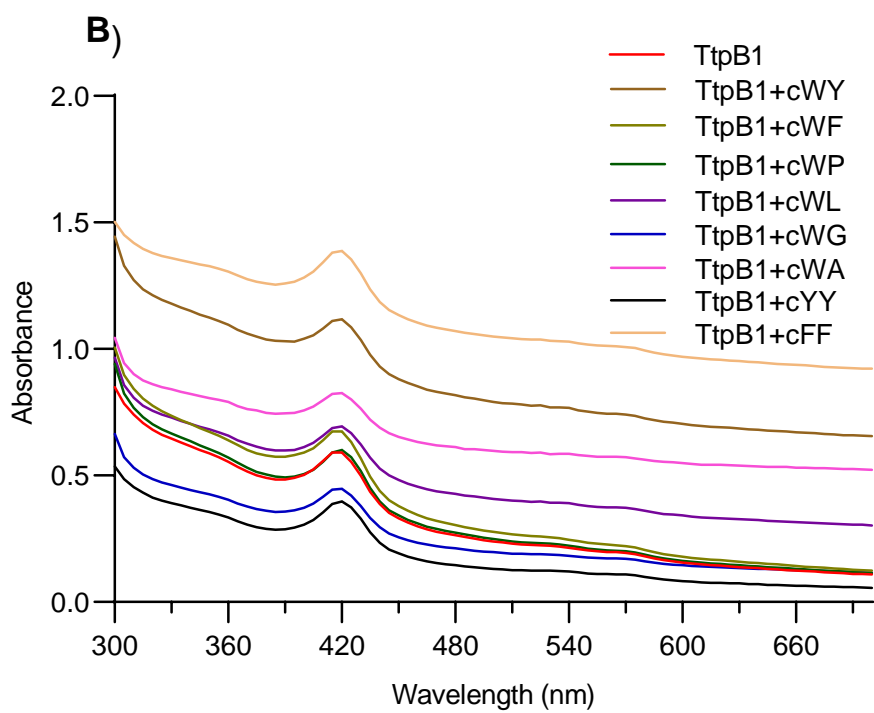
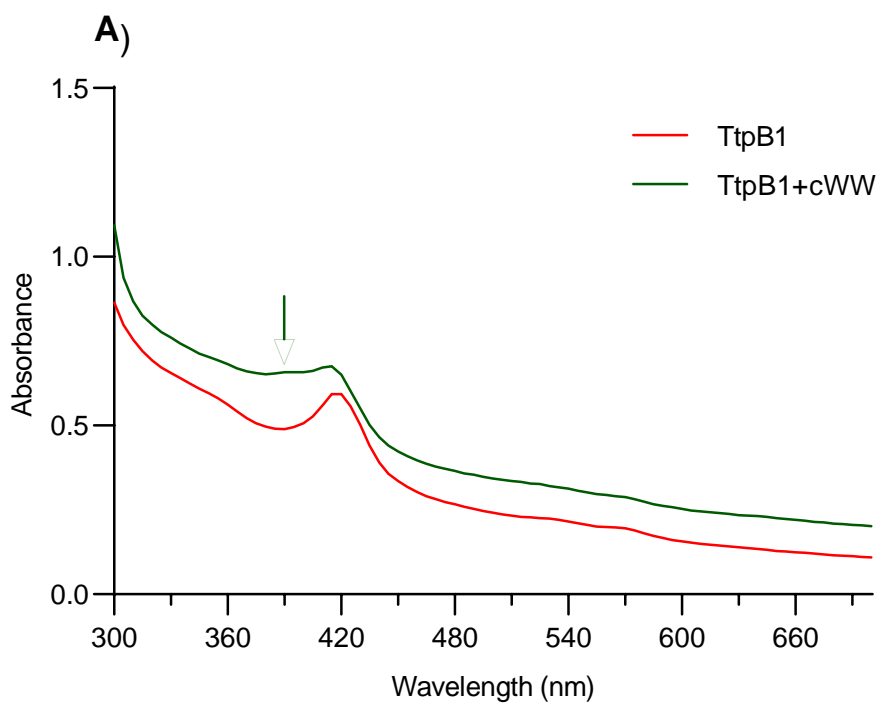


Fig. S33 UV-Vis spectroscopic analysis of TtpB1 with its substrate cWW (A) and other CDPs (B).

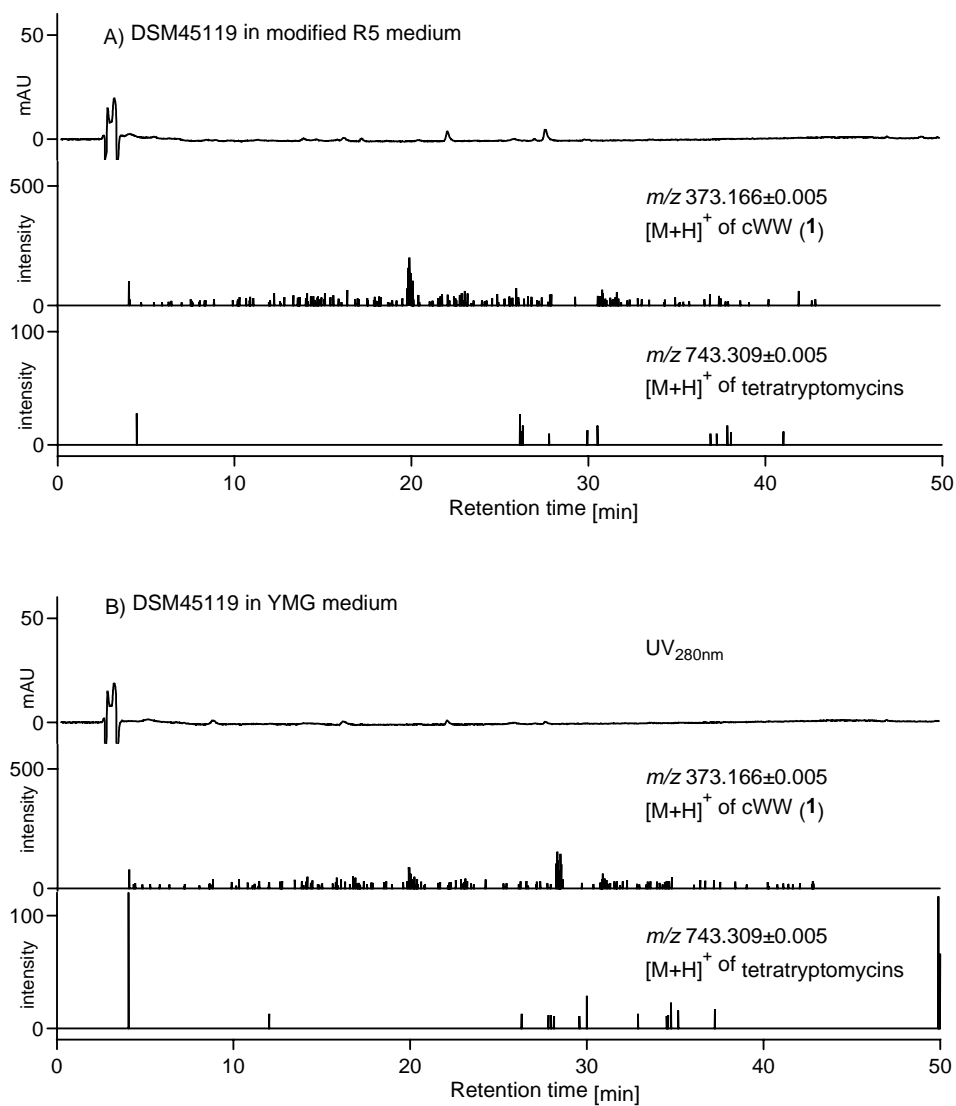


Fig. S34 LC-MS analysis for tetratryptomycin production in *S. antimicrobica* DSM 45119.

References

- 1 N. Zaburannyi, M. Rabyk, B. Ostash, V. Fedorenko, and A. Luzhetskyy, *Bmc Genomics*, 2014, **15**, 97.
- 2 M. R. Green and J. Sambrook, *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2012.
- 3 T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood, *Practical Streptomyces Genetics*, John Innes Foundation, Norwich, UK, 2000.
- 4 Y. Zhu, P. Fu, Q. Lin, G. Zhang, H. Zhang, S. Li, J. Ju, W. Zhu, and C. Zhang, *Org Lett.*, 2012, **14**, 2666-2669.
- 5 B. Gust, G. L. Challis, K. Fowler, T. Kieser, and K. F. Chater, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1541-1546.
- 6 T. W. Giessen, A. M. von Tesmar, and M. A. Marahiel, *Biochemistry*, 2013, **52**, 4274-4283.
- 7 E. D. James, B. Knuckley, N. Alqahtani, S. Porwal, J. Ban, J. A. Karty, R. Viswanathan, and A. L. Lane, *ACS Synth. Biol.*, 2016, **5**, 547-553.
- 8 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, and P. Belin, *Nat. Chem Biol.*, 2015, **11**, 721-727.
- 9 H. Li, Y. Qiu, C. Guo, M. Han, Y. Zhou, Y. Feng, S. Luo, Y. Tong, G. Zheng, and S. Zhu, *Chem. Commun.*, 2019, **55**, 8390-8393.
- 10 J. Liu, H. Yu, and S.-M. Li, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 4435-4444.
- 11 H. Yu and S.-M. Li, *Org. Lett.*, 2019, **21**, 7094-7098.
- 12 J. Liu, X. Xie, and S.-M. Li, *Angew. Chem. Int. Ed. Engl.*, 2019, **58**, 11534-11540.
- 13 H. Yu, X. Xie, and S.-M. Li, *Org Lett.*, 2018, **20**, 4921-4925.
- 14 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, and M. Gondry, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 7426-7431.
- 15 W. Tian, C. Sun, M. Zheng, J. R. Harmer, M. Yu, Y. Zhang, H. Peng, D. Zhu, Z. Deng, S. L. Chen, M. Mobli, X. Jia, and X. Qu, *Nat. Commun.*, 2018, **9**, 4428.
- 16 M. J. Cryle, S. G. Bell, and I. Schlichting, *Biochemistry*, 2010, **49**, 7282-7296.
- 17 S. Meng, W. Han, J. Zhao, X. H. Jian, H. X. Pan, and G. L. Tang, *Angew. Chem Int. Ed. Engl.*, 2018, **57**, 719-723.
- 18 J. Ma, Z. Wang, H. Huang, M. Luo, D. Zuo, B. Wang, A. Sun, Y. Q. Cheng, C. Zhang, and J. Ju, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 7797-7802.
- 19 Y. Du, Y. Wang, T. Huang, M. Tao, Z. Deng, and S. Lin, *BMC Microbiol.*, 2014, **14**, 30.
- 20 L. M. Podust and D. H. Sherman, *Nat. Prod. Rep.*, 2012, **29**, 1251-1266.

5 Conclusions and future prospects

In this thesis, several new tryptophan-containing 2,5-DKP derivatives together with some novel biosynthetic enzymes have been identified and characterized from CDPS-dependent pathways in actinobacterial strains *via* genome mining and heterologous expression. As thousands of actinobacterial genome sequences are available in the public databases, numerous gene clusters await to be explored. It is expected that advanced genome mining strategies and tools will strongly accelerate the discovery and characterization of novel and interesting biocatalysts as well as secondary metabolites from these untapped biosynthetic pathways.

In the course of searching new CDPSs for the formation of DKPs comprising tryptophan and other amino acids, functions of eleven candidate CDPS homologues from *Streptomyces* strains were verified by heterologous expression in *E. coli*. Particularly, nine of them were characterized to produce one or more tryptophan-containing cyclodipeptides, thus expanding the product spectrum of *cyclo*-L-Trp-Xaa originated from actinobacteria. The total product yields of these CDPs reached to 46 - 211 mg/L in *E. coli* culture, which makes it possible to combine these CDPSs with other modification enzymes in the field of synthetic biology in the future. This study also highlights the potential of the microbial machinery for tryptophan-containing cyclodipeptide biosynthesis.

Furthermore, two similar *p450*-associated *cdps*-containing gene clusters (*gut*) from *Streptomyces* strains were identified for the biosynthesis of rare and novel C3-guaninyl indole alkaloids guanitrypmycins by heterologous expression in *Streptomyces coelicolor* M1146. Expression of different gene combinations, precursor feeding experiments, combined with biochemical characterization proved the biosynthetic steps of guanitrypmycins. It demonstrated that *cyclo*-L-Trp-L-Phe and *cyclo*-L-Trp-L-Tyr initially assembled by the CDPS GutA serve as DKP precursors and are further dehydrogenated by cyclodipeptide oxidase Gut(BC) on the phenylalanyl/tyrosyl hemisphere. Subsequently, the cytochrome P450 GutD was proven to act as the key biocatalyst and mediate the stereospecific coupling of the generated *cW* Δ F/ *cW* Δ Y with a guanine moiety *via* C3–C8' linkage. The methyltransferase GutE governs the last modification step to transfer a methyl group to N9' of the guaninyl residue. Moreover, the non-enzymatic epimerization *via* keto-enol tautomerism further increases the structural diversity of guanitrypmycins. Therefore, this study represents an excellent example for revealing the untapped genetic potential by genome mining and heterologous expression.

Inspired by the intriguing chemical transformation performed by the P450 GutD in the biosynthesis of guanitrypmycins, we analyzed more bacterial *cdps*-containing clusters to explore novel modification enzymes. Subsequently, two *cdps-p450* operons (*tp1* and *tp2*) in *Saccharopolyspora antimicrobica* were identified and characterized to produce different *cyclo*-L-Trp-L-Trp dimers *via* genome mining. Heterologous expression and biochemical characterization revealed that the P450 TtpB1 catalyzes stereospecific C3 (sp³)–C3' (sp³) bond formation between two monomers, both from the opposite side

of H-11/H-11', while TtpB2 mainly connects the monomers *via* the unusual linkage between N1' and C3 from the H-11 side. As the two P450s catalyze distinct coupling patterns different from those previously reported in actinobacteria, this study significantly expands the P450 spectrum for CDP modifications and provides a simple and direct approach for enzymatic one-step preparation of structurally complex dimeric DKPs.

For future prospects, the following works can be performed:

- In the first project, some CDPSs exhibit promiscuity with respect to their substrates, producing a range of cyclodipeptide products. Therefore, refactoring these CDPSs with other modification enzymes, *e.g.*, different prenyltransferases, would give rise to diverse chemical entities.
- Guanitrypmycins are novel nucleobase-containing DKPs. However, testing of guanitrypmycins with the human ovarian adenocarcinoma cell SK-OV-3 showed almost no cytotoxic activity. Therefore, further bioactivity assays toward some representative screening models, such as various bacterial and fungal strains, as well as different tumor cell lines, are required to explore their potential biological and pharmaceutical activities.
- Determining the crystal structures of the mentioned P450 enzymes would provide insights into their catalytic mechanisms.
- Both of the two P450s TtpB1 and TtpB2 showed high substrate specificity, as the coupling reaction was not observed when using other tryptophan-containing or aromatic cyclodipeptides as substrates. Thus, further site-directed mutagenesis of the two P450s based on their structural information, could make them as versatile biocatalysts for preparation of diverse dimeric DKPs.

6 References

1. Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Meier-Kolthoff, J. P.; Klenk, H. P.; Clement, C.; Ouhdouch, Y.; van Wezel, G. P., Taxonomy, physiology, and natural products of actinobacteria. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, 1-43.
2. Ventura, M.; Canchaya, C.; Tauch, A.; Chandra, G.; Fitzgerald, G. F.; Chater, K. F.; van Sinderen, D., Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 495-548.
3. van der Meij, A.; Worsley, S. F.; Hutchings, M. I.; van Wezel, G. P., Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol. Rev.* **2017**, *41*, 392-416.
4. Lewin, G. R.; Carlos, C.; Chevrette, M. G.; Horn, H. A.; McDonald, B. R.; Stankey, R. J.; Fox, B. G.; Currie, C. R., Evolution and ecology of actinobacteria and their bioenergy applications. *Annu. Rev. Microbiol.* **2016**, *70*, 235-254.
5. Bérdy, J., Bioactive microbial metabolites. *J. Antibiot.* **2005**, *58*, 1-26.
6. Genilloud, O., Actinomycetes: still a source of novel antibiotics. *Nat. Prod. Rep.* **2017**, *34*, 1203-1232.
7. Papagianni, M., Recent advances in engineering the central carbon metabolism of industrially important bacteria. *Microb. Cell Fact.* **2012**, *11*, 50.
8. Gaynor, M.; Mankin, A. S., Macrolide antibiotics: binding site, mechanism of action, resistance. *Curr. Top. Med. Chem.* **2003**, *3*, 949-960.
9. Chopra, I.; Roberts, M., Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 232-260.
10. Baysarowich, J.; Koteva, K.; Hughes, D. W.; Ejim, L.; Griffiths, E.; Zhang, K.; Junop, M.; Wright, G. D., Rifamycin antibiotic resistance by ADP-ribosylation: structure and diversity of Arr. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4886-4891.
11. Omura, S.; Imamura, N.; Oiwa, R.; Kuga, H.; Rimiko, I.; Masuma, R.; Iwai, Y., Clostomicins, new antibiotics produced by *Micromonospora echinospora* subsp. *armeniaca* subsp. nov. I. Production, isolation, and physico-chemical and biological properties. *J. Antibiot.* **1986**, *39*, 1407-1412.
12. Mingeot-Leclercq, M.-P.; Glupczynski, Y.; Tulkens, P. M., Aminoglycosides: activity and resistance. *Antimicrob. Agents Chemother.* **1999**, *43*, 727-737.
13. Butler, M. S.; Hansford, K. A.; Blaskovich, M. A.; Halai, R.; Cooper, M. A., Glycopeptide antibiotics: back to the future. *J. Antibiot.* **2014**, *67*, 631-644.
14. Van Bambeke, F., Glycopeptides and glycodepsipeptides in clinical development: a comparative review of their antibacterial spectrum, pharmacokinetics and clinical efficacy. *Curr. Opin. Invest. Drugs* **2006**, *7*, 740-749.
15. Steenbergen, J. N.; Alder, J.; Thorne, G. M.; Tally, F. P., Daptomycin: a lipopeptide antibiotic for the treatment of serious Gram-positive infections. *J. Antimicrob. Chemother.* **2005**, *55*, 283-288.
16. Mulinos, M., Cycloserine: an antibiotic paradox. *Antibiotics Ann.* **1955**, *3*, 131.
17. Nett, M.; Ikeda, H.; Moore, B. S., Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **2009**, *26*, 1362-84.
18. Fiedler, H. P.; Kurth, R.; Langhärig, J.; Delzer, J.; Zähler, H., Nikkomycins: microbial inhibitors of chitin synthase. *J. Chem. Biotechnol.* **1982**, *32*, 271-280.
19. Feling, R. H.; Buchanan, G. O.; Mincer, T. J.; Kauffman, C. A.; Jensen, P. R.; Fenical, W., Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chem. Int. Ed.* **2003**, *42*, 355-357.
20. Jansson, R.; Dybas, R., Avermectins: biochemical mode of action, biological activity and agricultural importance. In *Insecticides with novel modes of action*, Springer 1998; pp 152-170.
21. Stallone, G.; Infante, B.; Di Lorenzo, A.; Rascio, F.; Zaza, G.; Grandaliano, G., mTOR inhibitors effects on regulatory T cells and on dendritic cells. *J. Transl. Med.* **2016**, *14*, 152.

REFERENCES

22. Rangseekaew, P.; Pathom-Aree, W., Cave actinobacteria as producers of bioactive metabolites. *Front. Microbiol.* **2019**, *10*, 387.
23. Subramani, R.; Aalbersberg, W., Culturable rare actinomycetes: diversity, isolation and marine natural product discovery. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 9291-9321.
24. Waksman, S. A., Streptomycin: background, isolation, properties, and utilization. *Science* **1953**, *118*, 259-266.
25. Hutchings, M. I.; Truman, A. W.; Wilkinson, B., Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **2019**, *51*, 72-80.
26. Li, Z.; Zhu, D.; Shen, Y., Discovery of novel bioactive natural products driven by genome mining. *Drug Discov. Therap.* **2018**, *12*, 318-328.
27. Wright, G. D., Unlocking the potential of natural products in drug discovery. *Microb. Biotechnol.* **2019**, *12*, 55-57.
28. Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; *al.*, *e.*, Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature* **2002**, *417*, 141-147.
29. Aigle, B.; Corre, C., Waking up *Streptomyces* secondary metabolism by constitutive expression of activators or genetic disruption of repressors. In *Methods Enzymol.*, Elsevier 2012; Vol. 517, pp 343-366.
30. Albarano, L.; Esposito, R.; Ruocco, N.; Costantini, M., Genome mining as new challenge in natural products discovery. *Mar. Drugs* **2020**, *18*, 199.
31. Rutledge, P. J.; Challis, G. L., Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015**, *13*, 509-523.
32. Challis, G. L., Genome mining for novel natural product discovery. *J. Med. Chem.* **2008**, *51*, 2618-2628.
33. Corr, C.; Challis, G. L., Exploiting genomics for new natural product discovery in prokaryotes. In *Comprehensive Natural Products II: Chemistry and Biology*, Elsevier 2010; pp 429-453.
34. Fischbach, M. A.; Walsh, C. T., Antibiotics for emerging pathogens. *Science* **2009**, *325*, 1089-1093.
35. Chu, L.; Huang, J.; Muhammad, M.; Deng, Z.; Gao, J., Genome mining as a biotechnological tool for the discovery of novel marine natural products. *Crit. Rev. Biotechnol.* **2020**, 571-589.
36. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T., antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**, *47*, W81-W87.
37. Zhao, S.; Sakai, A.; Zhang, X.; Vetting, M. W.; Kumar, R.; Hillerich, B.; San Francisco, B.; Solbiati, J.; Steves, A.; Brown, S.; Akiva, E.; Barber, A.; Seidel, R. D.; Babbitt, P. C.; Almo, S. C.; Gerlt, J. A.; Jacobson, M. P., Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks. *Elife* **2014**, *3*, e03275.
38. Ziemert, N.; Alanjary, M.; Weber, T., The evolution of genome mining in microbes—a review. *Nat. Prod. Rep.* **2016**, *33*, 988-1005.
39. Romano, S.; Jackson, S. A.; Patry, S.; Dobson, A. D. W., Extending the “one strain many compounds”(OSMAC) principle to marine microorganisms. *Mar. Drugs* **2018**, *16*, 244.
40. McAlpine, J. B.; Bachmann, B. O.; Pirae, M.; Tremblay, S.; Alarco, A.-M.; Zazopoulos, E.; Farnet, C. M., Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *J. Nat. Prod.* **2005**, *68*, 493-496.
41. Onaka, H.; Mori, Y.; Igarashi, Y.; Furumai, T., Mycolic acid-containing bacteria induce natural-product biosynthesis in *Streptomyces* species. *Appl. Environ. Microbiol.* **2011**, *77*, 400-406.
42. Onaka, H., Novel antibiotic screening methods to awaken silent or cryptic secondary metabolic pathways in actinomycetes. *J. Antibiot.* **2017**, *70*, 865-870.
43. Liu, G.; Chater, K. F.; Chandra, G.; Niu, G.; Tan, H., Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.* **2013**, *77*, 112-143.

REFERENCES

44. Zarins-Tutt, J. S.; Barberi, T. T.; Gao, H.; Mearns-Spragg, A.; Zhang, L.; Newman, D. J.; Goss, R. J. M., Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat. Prod. Rep.* **2016**, *33*, 54-72.
45. Hug, J. J.; Krug, D.; Müller, R., Bacteria as genetically programmable producers of bioactive natural products. *Nat. Rev. Chem.* **2020**, 172-193.
46. Xia, H.; Li, X.; Li, Z.; Zhan, X.; Mao, X.; Li, Y., The application of regulatory cascades in *Streptomyces*: yield enhancement and metabolite mining. *Front. Microbiol.* **2020**, *11*, 406.
47. Hoskisson, P. A.; Fernández - Martínez, L. T., Regulation of specialised metabolites in actinobacteria—expanding the paradigms. *Environ. Microbiol. Rep.* **2018**, *10*, 231-238.
48. Lu, W.; Alanzi, A. R.; Abugrain, M. E.; Ito, T.; Mahmud, T., Global and pathway-specific transcriptional regulations of pactamycin biosynthesis in *Streptomyces pactum*. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 10589-10601.
49. Romero-Rodríguez, A.; Robledo-Casados, I.; Sánchez, S., An overview on transcriptional regulators in *Streptomyces*. *Biochim. Biophys. Acta. (BBA)-Gene Regulatory Mechanisms* **2015**, *1849*, 1017-1039.
50. Sherman, D. H.; Li, S.; Yermalitskaya, L. V.; Kim, Y.; Smith, J. A.; Waterman, M. R.; Podust, L. M., The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from *Streptomyces venezuelae*. *J. Biol. Chem.* **2006**, *281*, 26289-26297.
51. Rebets, Y.; Brötz, E.; Tokovenko, B.; Luzhetskyy, A., Actinomycetes biosynthetic potential: how to bridge *in silico* and *in vivo*? *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 387-402.
52. Browning, D. F.; Busby, S. J. W., The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2004**, *2*, 57-65.
53. Ishihama, A., Prokaryotic genome regulation: a revolutionary paradigm. *Proc. Jpn. Acad. B Phys. Biol. sci.* **2012**, *88*, 485-508.
54. Lee, D. J.; Minchin, S. D.; Busby, S. J., Activating transcription in bacteria. *Annu. Rev. Microbiol.* **2012**, *66*, 125-152.
55. Wietzorrek, A.; Bibb, M., A novel family of proteins that regulates antibiotic production in *Streptomyces* appears to contain an OmpR - like DNA - binding fold. *Mol. Microbiol.* **1997**, *25*, 1181-1184.
56. De Schrijver, A.; De Mot, R., A subfamily of MalT-related ATP-dependent regulators in the LuxR family. *Microbiology* **1999**, *145*, 1287-1288.
57. Laureti, L.; Song, L.; Huang, S.; Corre, C.; Leblond, P.; Challis, G. L.; Aigle, B., Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6258-6263.
58. Ramos, J. L.; Martínez-Bueno, M.; Molina-Henares, A. J.; Terán, W.; Watanabe, K.; Zhang, X.; Gallegos, M. T.; Brennan, R.; Tobes, R., The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 326-356.
59. Cuthbertson, L.; Nodwell, J. R., The TetR family of regulators. *Microbiol. Mol. Biol. Rev.* **2013**, *77*, 440-475.
60. Gou, L.; Han, T.; Wang, X.; Ge, J.; Liu, W.; Hu, F.; Wang, Z., A novel TetR family transcriptional regulator, CalR3, negatively controls calcimycin biosynthesis in *Streptomyces chartreusis* NRRL 3882. *Front. Microbiol.* **2017**, *8*, 2371.
61. Tong, Y.; Weber, T.; Lee, S. Y., CRISPR/Cas-based genome engineering in natural product discovery. *Nat. Prod. Rep.* **2019**, *36*, 1262-1280.
62. Huo, L.; Hug, J. J.; Fu, C.; Bian, X.; Zhang, Y.; Müller, R., Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2019**, *36*, 1412-1436.
63. Luo, Y.; Enghiad, B.; Zhao, H., New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Nat. Prod. Rep.* **2016**, *33*, 174-182.

REFERENCES

64. Zhang, J. J.; Yamanaka, K.; Tang, X.; Moore, B. S., Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. In *Methods Enzymol.*, Elsevier 2019; Vol. 621, pp 87-110.
65. Zheng, X.; Cheng, Q.; Yao, F.; Wang, X.; Kong, L.; Cao, B.; Xu, M.; Lin, S.; Deng, Z.; Chooi, Y.-H.; You, D., Biosynthesis of the pyrrolidine protein synthesis inhibitor anisomycin involves novel gene ensemble and cryptic biosynthetic steps. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 4135-4140.
66. Owen, J. G.; Charlop-Powers, Z.; Smith, A. G.; Ternei, M. A.; Calle, P. Y.; Reddy, B. V. B.; Montiel, D.; Brady, S. F., Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 4221-4226.
67. Kaysser, L.; Bernhardt, P.; Nam, S.-J.; Loesgen, S.; Ruby, J. G.; Skewes-Cox, P.; Jensen, P. R.; Fenical, W.; Moore, B. S., Merochlorins A–D, cyclic meroterpenoid antibiotics biosynthesized in divergent pathways with vanadium-dependent chloroperoxidases. *J. Am. Chem. Soc.* **2012**, *134*, 11988-11991.
68. Alexander, D. C.; Rock, J.; He, X.; Brian, P.; Miao, V.; Baltz, R. H., Development of a genetic system for combinatorial biosynthesis of lipopeptides in *Streptomyces fradiae* and heterologous expression of the A54145 biosynthesis gene cluster. *Appl. Environ. Microbiol.* **2010**, *76*, 6877-6887.
69. Saugar, I.; Molloy, B.; Sanz, E.; Sánchez, M. B.; Fernández-Lobato, M.; Jiménez, A., Characterization of the biosynthetic gene cluster (*ata*) for the A201A aminonucleoside antibiotic from *Saccharothrix mutabilis* subsp. *capreolus*. *J. Antibiot.* **2017**, *70*, 404-413.
70. D'Agostino, P. M.; Gulder, T. A. M., Direct pathway cloning combined with sequence- and ligation-independent cloning for fast biosynthetic gene cluster refactoring and heterologous expression. *ACS Synth. Biol.* **2018**, *7*, 1702-1708.
71. Greunke, C.; Duell, E. R.; D'Agostino, P. M.; Glöckle, A.; Lamm, K.; Gulder, T. A. M., Direct pathway cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.* **2018**, *47*, 334-345.
72. Wang, H.; Li, Z.; Jia, R.; Hou, Y.; Yin, J.; Bian, X.; Li, A.; Müller, R.; Stewart, A. F.; Fu, J.; Zhang, Y., RecET direct cloning and Red α recombinering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* **2016**, *11*, 1175-1190.
73. Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Müller, R.; Stewart, A. F.; Zhang, Y., Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **2012**, *30*, 440-446.
74. Wang, H.; Bian, X.; Xia, L.; Ding, X.; Müller, R.; Zhang, Y.; Fu, J.; Stewart, A. F., Improved seamless mutagenesis by recombinering using *ccdB* for counterselection. *Nucleic Acids Res.* **2014**, *42*, e37.
75. Wang, H.; Li, Z.; Jia, R.; Yin, J.; Li, A.; Xia, L.; Yin, Y.; Müller, R.; Fu, J.; Stewart, A. F., ExoCET: exonuclease *in vitro* assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.* **2018**, *46*, e28-.
76. Jiang, W.; Zhu, T. F., Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat. Protoc.* **2016**, *11*, 960-975.
77. Kouprina, N.; Larionov, V., TAR cloning: perspectives for functional genomics, biomedicine, and biotechnology. *Mol. Ther. Methods Clin. Dev.* **2019**, *14*, 16-26.
78. Ongley, S. E.; Bian, X.; Neilan, B. A.; Müller, R., Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2013**, *30*, 1121-1138.
79. Gomez-Escribano, J. P.; Bibb, M. J., *Streptomyces coelicolor* as an expression host for heterologous gene clusters. In *Methods Enzymol.*, Elsevier 2012; Vol. 517, pp 279-300.
80. Wenzel, S. C.; Müller, R., Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* **2005**, *16*, 594-606.
81. Gomez - Escribano, J. P.; Bibb, M. J., Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **2011**, *4*, 207-215.

REFERENCES

82. Penn, J.; Li, X.; Whiting, A.; Latif, M.; Gibson, T.; Silva, C. J.; Brian, P.; Davies, J.; Miao, V.; Wrigley, S. K.; Baltz, R. H., Heterologous production of daptomycin in *Streptomyces lividans*. *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 121-128.
83. Myronovskiy, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A., Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316-324.
84. Zaburannyi, N.; Rabyk, M.; Ostash, B.; Fedorenko, V.; Luzhetskyy, A., Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genom.* **2014**, *15*, 97.
85. Gustafsson, C.; Govindarajan, S.; Minshull, J., Codon bias and heterologous protein expression. *Trends Biotechnol.* **2004**, *22*, 346-353.
86. Komatsu, M.; Uchiyama, T.; Ōmura, S.; Cane, D. E.; Ikeda, H., Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2646-2651.
87. Pickens, L. B.; Tang, Y.; Chooi, Y.-H., Metabolic engineering for the production of natural products. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 211-236.
88. Boghigian, B. A.; Zhang, H.; Pfeifer, B. A., Multi-factorial engineering of heterologous polyketide production in *Escherichia coli* reveals complex pathway interactions. *Biotechnol. Bioeng.* **2011**, *108*, 1360-1371.
89. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **2001**, *291*, 1790-1792.
90. Gao, X.; Wang, P.; Tang, Y., Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **2010**, *88*, 1233-1242.
91. Myronovskiy, M.; Luzhetskyy, A., Genome engineering in actinomycetes using site-specific recombinases. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 4701-4712.
92. Owen, J. G.; Reddy, B. V. B.; Ternei, M. A.; Charlop-Powers, Z.; Calle, P. Y.; Kim, J. H.; Brady, S. F., Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 11797-11802.
93. Prasad, C., Bioactive cyclic dipeptides. *Peptides* **1995**, *16*, 151-164.
94. Borthwick, A. D., 2,5-Diketopiperazines: synthesis, reactions, medicinal chemistry, and bioactive natural products. *Chem. Rev.* **2012**, *112*, 3641-3716.
95. Campbell, J.; Lin, Q.; Geske, G. D.; Blackwell, H. E., New and unexpected insights into the modulation of LuxR-type quorum sensing by cyclic dipeptides. *ACS Chem. Biol.* **2009**, *4*, 1051-1059.
96. Belin, P.; Moutiez, M.; Lautru, S.; Seguin, J.; Pernodet, J.-L.; Gondry, M., The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. *Nat. Prod. Rep.* **2012**, *29*, 961-979.
97. 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.
98. Kanoh, K.; Kohno, S.; Asari, T.; Harada, T.; Katada, J.; Muramatsu, M.; Kawashima, H.; Sekiya, H.; Uno, I., (-)-Phenylahistin: a new mammalian cell cycle inhibitor produced by *Aspergillus ustus*. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2847-2852.
99. Watanabe, A.; Kuriyama, T.; Kamei, K.; Nishimura, K.; Miyaji, M.; Sekine, T.; Waku, M., Immunosuppressive substances in *Aspergillus fumigatus* culture filtrate. *J. Antimicrob. Chemother.* **2003**, *9*, 114-121.
100. Kohn, H.; Widger, W., The molecular basis for the mode of action of bicyclomycin. *Curr. Drug Targets Infect. Disord.* **2005**, *5*, 273-295.
101. 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.* **2011**, *13*, 2770-2773.

REFERENCES

102. Gouda, M. A., Overview of the synthetic routes to tadalafil and its analogues. *Synth. Commun.* **2017**, *47*, 2269-2304.
103. Martins, M. B.; Carvalho, I., Diketopiperazines: biological activity and synthesis. *Tetrahedron* **2007**, *63*, 9923-9932.
104. Horton, D. A.; Bourne, G. T.; Smythe, M. L., Exploring privileged structures: the combinatorial synthesis of cyclic peptides. *J. Comput. Aided Mol. Des.* **2002**, *16*, 415-431.
105. Mishra, A. K.; Choi, J.; Choi, S.-J.; Baek, K.-H., Cyclodipeptides: an overview of their biosynthesis and biological activity. *Molecules* **2017**, *22*, 1796.
106. Yin, W.-B.; Grundmann, A.; Cheng, J.; Li, S.-M., Acetylaszonalenin biosynthesis in *Neosartorya fischeri* identification of the biosynthetic gene cluster by genomic mining and functional proof of the genes by biochemical investigation. *J. Biol. Chem.* **2009**, *284*, 100-109.
107. Li, S.-M., Genome mining and biosynthesis of fumitremorgin-type alkaloids in ascomycetes. *J. Antibiot.* **2011**, *64*, 45-49.
108. Haarmann, T.; Machado, C.; Lübbe, Y.; Correia, T.; Schardl, C. L.; Panaccione, D. G.; Tudzynski, P., The ergot alkaloid gene cluster in *Claviceps purpurea*: extension of the cluster sequence and intra species evolution. *Phytochemistry* **2005**, *66*, 1312-1320.
109. Balibar, C. J.; Walsh, C. T., GliP, a multimodular nonribosomal peptide synthetase in *Aspergillus fumigatus*, makes the diketopiperazine scaffold of gliotoxin. *Biochemistry* **2006**, *45*, 15029-15038.
110. Gardiner, D. M.; Cozijnsen, A. J.; Wilson, L. M.; Pedras, M. S. C.; Howlett, B. J., The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*. *Mol. Microbiol.* **2004**, *53*, 1307-1318.
111. Lazos, O.; Tosin, M.; Slusarczyk, A. L.; Boakes, S.; Cortés, J.; Sidebottom, P. J.; Leadlay, P. F., Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters. *Chem. Biol.* **2010**, *17*, 160-173.
112. Oide, S.; Moeder, W.; Krasnoff, S.; Gibson, D.; Haas, H.; Yoshioka, K.; Turgeon, B. G., NPS6, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes. *The Plant Cell* **2006**, *18*, 2836-2853.
113. Healy, F. G.; Wach, M.; Krasnoff, S. B.; Gibson, D. M.; Loria, R., The *txtAB* genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol. Microbiol.* **2000**, *38*, 794-804.
114. Schwarzer, D.; Finking, R.; Marahiel, M. A., Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* **2003**, *20*, 275-287.
115. Walsh, C. T., Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **2016**, *33*, 127-135.
116. Hur, G. H.; Vickery, C. R.; Burkart, M. D., Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **2012**, *29*, 1074-1098.
117. Walsh, C. T.; O'Brien, R. V.; Khosla, C., Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed.* **2013**, *52*, 7098-7124.
118. Baccile, J. A.; Le, H. H.; Pfannenstiel, B. T.; Bok, J. W.; Gomez, C.; Brandenburger, E.; Hoffmeister, D.; Keller, N. P.; Schroeder, F. C., Diketopiperazine formation in fungi requires dedicated cyclization and thiolation domains. *Angew. Chem. Int. Ed.* **2019**, *58*, 14589-14593.
119. Walsh, C. T.; Chen, H.; Keating, T. A.; Hubbard, B. K.; Losey, H. C.; Luo, L.; Marshall, C. G.; Miller, D. A.; Patel, H. M., Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.* **2001**, *5*, 525-534.
120. Stachelhaus, T.; Mootz, H. D.; Bergendahl, V.; Marahiel, M. A., Peptide bond formation in nonribosomal peptide biosynthesis catalytic role of the condensation domain. *J. Biol. Chem.* **1998**, *273*, 22773-22781.

REFERENCES

121. Schultz, A. W.; Oh, D.-C.; Carney, J. R.; Williamson, R. T.; Udvary, D. W.; Jensen, P. R.; Gould, S. J.; Fenical, W.; Moore, B. S., Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J. Am. Chem. Soc.* **2008**, *130*, 4507-4516.
122. 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.
123. 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.
124. Jacques, I. B.; Moutiez, M.; Witwinowski, J.; Darbon, E.; Martel, C.; Seguin, J.; Favry, E.; Thai, R.; Lecoq, A.; Dubois, S.; Pernodet, J.-L.; Gondry, M.; Belin, P., Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat. Chem. Biol.* **2015**, *11*, 721-727.
125. Seguin, J.; Moutiez, M.; Li, Y.; Belin, P.; Lecoq, A.; Fonvielle, M.; Charbonnier, J.-B.; Pernodet, J.-L.; Gondry, M., Nonribosomal peptide synthesis in animals: the cyclodipeptide synthase of *Nematostella*. *Chem. Biol.* **2011**, *18*, 1362-1368.
126. Sauguet, L.; Moutiez, M.; Li, Y.; Belin, P.; Seguin, J.; Le Du, M.-H.; Thai, R.; Masson, C.; Fonvielle, M.; Pernodet, J.-L., Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. *Nucleic Acids Res.* **2011**, *39*, 4475-4489.
127. Vetting, M. W.; Hegde, S. S.; Blanchard, J. S., The structure and mechanism of the *Mycobacterium tuberculosis* cyclodityrosine synthetase. *Nat. Chem. Biol.* **2010**, *6*, 797-799.
128. Bonnefond, L.; Arai, T.; Sakaguchi, Y.; Suzuki, T.; Ishitani, R.; Nureki, O., Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3912-3917.
129. Moutiez, M.; Schmitt, E.; Seguin, J.; Thai, R.; Favry, E.; Belin, P.; Mechulam, Y.; Gondry, M., Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat. Commun.* **2014**, *5*, 5141.
130. 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* **2010**, *49*, 7282-7296.
131. Belin, P.; Le Du, M. H.; Fielding, A.; Lequin, O.; Jacquet, M.; Charbonnier, J.-B.; Lecoq, A.; Thai, R.; Courçon, M.; Masson, C.; Dugave, C.; Genet, R.; Pernodet, J.-L.; Gondry, M., Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 7426-7431.
132. 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.
133. 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.* **2018**, *57*, 719-723.
134. 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.
135. Rittle, J.; Green, M. T., Cytochrome P450 compound I: capture, characterization, and CH bond activation kinetics. *Science* **2010**, *330*, 933-937.
136. Omura, T.; Sato, R., The carbon monoxide-binding pigment of liver microsomes I. Evidence for its hemoprotein nature. *J. Biol. Chem.* **1964**, *239*, 2370-2378.
137. Omura, T.; Sato, R., The carbon monoxide-binding pigment of liver microsomes II. Solubilization, purification, and properties. *J. Biol. Chem.* **1964**, *239*, 2379-2385.
138. Zhang, X.; Li, S., Expansion of chemical space for natural products by uncommon P450 reactions. *Nat. Prod. Rep.* **2017**, *34*, 1061-1089.

REFERENCES

139. 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.* **2017**, *34*, 1141-1172.
140. Werck-Reichhart, D.; Feyereisen, R., Cytochromes P450: a success story. *Genome Biol.* **2000**, *1*, reviews3003. 1.
141. Podust, L. M.; Sherman, D. H., Diversity of P450 enzymes in the biosynthesis of natural products. *Nat. Prod. Rep.* **2012**, *29*, 1251-1266.
142. Guengerich, F. P., Mechanisms of cytochrome P450-catalyzed oxidations. *ACS Catal.* **2018**, *8*, 10964-10976.
143. Bakkes, P. J.; Riehm, J. L.; Sagadin, T.; Rühlmann, A.; Schubert, P.; Biemann, S.; Girhard, M.; Hutter, M. C.; Bernhardt, R.; Urlacher, V. B., Engineering of versatile redox partner fusions that support monooxygenase activity of functionally diverse cytochrome P450s. *Sci. Rep.* **2017**, *7*, 1-13.
144. Kudo, F.; Motegi, A.; Mizoue, K.; Eguchi, T., Cloning and characterization of the biosynthetic gene cluster of 16 - membered macrolide antibiotic FD - 891: involvement of a dual functional cytochrome P450 monooxygenase catalyzing epoxidation and hydroxylation. *ChemBioChem* **2010**, *11*, 1574-1582.
145. Carlson, J. C.; Li, S.; Gunatilleke, S. S.; Anzai, Y.; Burr, D. A.; Podust, L. M.; Sherman, D. H., Tirandamycin biosynthesis is mediated by co-dependent oxidative enzymes. *Nat. Chem.* **2011**, *3*, 628-633.
146. Mazzaferro, L. S.; Hüttel, W.; Fries, A.; Müller, M., Cytochrome P450-catalyzed regio- and stereoselective phenol coupling of fungal natural products. *J. Am. Chem. Soc.* **2015**, *137*, 12289-12295.
147. Hugentobler, K. G.; Müller, M., Towards semisynthetic natural compounds with a biaryl axis: oxidative phenol coupling in *Aspergillus niger*. *Bioorg. Med. Chem.* **2018**, *26*, 1374-1377.
148. Zhao, B.; Lamb, D. C.; Lei, L.; Kelly, S. L.; Yuan, H.; Hachey, D. L.; Waterman, M. R., Different binding modes of two flaviolin substrate molecules in cytochrome P450 158A1 (CYP158A1) compared to CYP158A2. *Biochemistry* **2007**, *46*, 8725-8733.
149. Makino, M.; Sugimoto, H.; Shiro, Y.; Asamizu, S.; Onaka, H.; Nagano, S., Crystal structures and catalytic mechanism of cytochrome P450 StaP that produces the indolocarbazole skeleton. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11591-11596.
150. Wang, Y.; Chen, H.; Makino, M.; Shiro, Y.; Nagano, S.; Asamizu, S.; Onaka, H.; Shaik, S., Theoretical and experimental studies of the conversion of chromopyrrolic acid to an antitumor derivative by cytochrome P450 StaP: the catalytic role of water molecules. *J. Am. Chem. Soc.* **2009**, *131*, 6748-6762.
151. Zhao, B.; Guengerich, F. P.; Bellamine, A.; Lamb, D. C.; Izumikawa, M.; Lei, L.; Podust, L. M.; Sundaramoorthy, M.; Kalaitzis, J. A.; Reddy, L. M.; Kelly, S. L.; Moore, B. S.; Stec, D.; Voehler, M.; Falck, J. R.; Shimada, T.; Waterman, M. R., Binding of two flaviolin substrate molecules, oxidative coupling, and crystal structure of *Streptomyces coelicolor* A3 (2) cytochrome P450 158A2. *J. Biol. Chem.* **2005**, *280*, 11599-11607.
152. Ma, J.; Wang, Z.; Huang, H.; Luo, M.; Zuo, D.; Wang, B.; Sun, A.; Cheng, Y. Q.; Zhang, C.; Ju, J., Biosynthesis of himastatin: assembly line and characterization of three cytochrome P450 enzymes involved in the post - tailoring oxidative steps. *Angew. Chem. Int. Ed.* **2011**, *50*, 7797-7802.
153. Chooi, Y.-H.; Hong, Y. J.; Cacho, R. A.; Tantillo, D. J.; Tang, Y., A cytochrome P450 serves as an unexpected terpene cyclase during fungal meroterpenoid biosynthesis. *J. Am. Chem. Soc.* **2013**, *135*, 16805-16808.
154. Xie, Y.; Li, Q.; Song, Y.; Ma, J.; Ju, J., Involvement of SgvP in carbon-sulfur bond formation during griseoviridin biosynthesis. *ChemBioChem* **2014**, *15*, 1183-1189.
155. Lin, H.-C.; Tsunematsu, Y.; Dhingra, S.; Xu, W.; Fukutomi, M.; Chooi, Y.-H.; Cane, D. E.; Calvo, A. M.; Watanabe, K.; Tang, Y., Generation of complexity in fungal terpene biosynthesis: discovery of a multifunctional cytochrome P450 in the fumagillin pathway. *J. Am. Chem. Soc.* **2014**, *136*, 4426-4436.

REFERENCES

156. Akashi, T.; Aoki, T.; Ayabe, S.-i., Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice. *Plant Physiol.* **1999**, *121*, 821-828.
157. Tsunematsu, Y.; Ishikawa, N.; Wakana, D.; Goda, Y.; Noguchi, H.; Moriya, H.; Hotta, K.; Watanabe, K., Distinct mechanisms for spiro-carbon formation reveal biosynthetic pathway crosstalk. *Nat. Chem. Biol.* **2013**, *9*, 818-825.
158. 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.* **2018**, *20*, 4921-4925.
159. 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 C³-aryl pyrroloindoline alkaloids. *Nat. Commun.* **2018**, *9*, 4428.
160. Shende, V. V.; Khatri, Y.; Newmister, S. A.; Sanders, J. N.; Lindovska, P.; Yu, F.; Doyon, T.; Kim, J.; Movassaghi, M.; Houk, K. N.; Movassaghi, M.; Sherman, D. H., Structure and function of NzeB, aversatile C–C and C–N bond forming diketopiperazine dimerase. *J. Am. Chem. Soc.* **2020**, *142*, 17413–17424.
161. Sun, C.; Luo, Z.; Zhang, W.; Tian, W.; Peng, H.; Lin, Z.; Deng, Z.; Kobe, B.; Jia, X.; Qu, X., Molecular basis of regio- and stereo-specificity in biosynthesis of bacterial heterodimeric diketopiperazines. *Nat. Commun.* **2020**, *11*, 6251.
162. Martinez, S.; Hausinger, R. P., Catalytic mechanisms of Fe (II)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* **2015**, *290*, 20702-20711.
163. Liscombe, D. K.; Louie, G. V.; Noel, J. P., Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat. Prod. Rep.* **2012**, *29*, 1238-1250.
164. Law, B. J. C.; Struck, A.-W.; Bennett, M. R.; Wilkinson, B.; Micklefield, J., Site-specific bioalkylation of rapamycin by the RapM 16-O-methyltransferase. *Chem. Sci.* **2015**, *6*, 2885-2892.
165. Landgraf, B.; Booker, S. J., The ylide has landed. *Nature* **2013**, *498*, 45-47.
166. 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* **2013**, *52*, 4274-4283.
167. Li, H.; Qiu, Y.; Guo, C.; Han, M.; Zhou, Y.; Feng, Y.; Luo, S.; Tong, Y.; Zheng, G.; Zhu, S., Pyrroloindoline cyclization in tryptophan-containing cyclodipeptides mediated by an unprecedented indole C3 methyltransferase from *Streptomyces* sp. HPH0547. *Chem. Commun.* **2019**, *55*, 8390-8393.
168. Winkelblech, J.; Fan, A.; Li, S.-M., Prenyltransferases as key enzymes in primary and secondary metabolism. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 7379-7397.
169. Yao, T.; Liu, J.; Liu, Z.; Li, T.; Li, H.; Che, Q.; Zhu, T.; Li, D.; Gu, Q.; Li, W., Genome mining of cyclodipeptide synthases unravels unusual tRNA-dependent diketopiperazine-terpene biosynthetic machinery. *Nat. Commun.* **2018**, *9*, 4091.
170. Skinnider, M. A.; Johnston, C. W.; Merwin, N. J.; Dejong, C. A.; Magarvey, N. A., Global analysis of prokaryotic tRNA-derived cyclodipeptide biosynthesis. *BMC Genom.* **2018**, *19*, 45.
171. Li, S.-M., Prenylated indole derivatives from fungi: structure diversity, biological activities, biosynthesis and chemoenzymatic synthesis. *Nat. Prod. Rep.* **2010**, *27*, 57-78.
172. Xu, W.; Gavia, D. J.; Tang, Y., Biosynthesis of fungal indole alkaloids. *Nat. Prod. Rep.* **2014**, *31*, 1474-1487.
173. Alkhalaf, L. M.; Ryan, K. S., Biosynthetic manipulation of tryptophan in bacteria: pathways and mechanisms. *Chem. Biol.* **2015**, *22*, 317-328.
174. Walsh, C. T., Biological matching of chemical reactivity: pairing indole nucleophilicity with electrophilic isoprenoids. *ACS Chem. Biol.* **2014**, *9*, 2718-2728.

REFERENCES

175. James, E. D.; Knuckley, B.; Alqahtani, N.; Porwal, S.; Ban, J.; Karty, J. A.; Viswanathan, R.; Lane, A. L., Two distinct cyclodipeptide synthases from a marine actinomycete catalyze biosynthesis of the same diketopiperazine natural product. *ACS Synth. Biol.* **2016**, *5*, 547-553.
176. 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.
177. 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*, 46.
178. Liu, J.; Yu, H.; Li, S.-M., Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from *Streptomyces* strains. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 4435-4444.
179. 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.
180. Yu, H.; Xie, X.; Li, S.-M., Coupling of *cyclo*-L-Trp-L-Trp with hypoxanthine increases the structure diversity of guanitrypmycins. *Org. Lett.* **2019**, *21*, 9104-9108.
181. Liu, J.; Xie, X.; Li, S. M., Guanitrypmycin biosynthetic pathways imply cytochrome P450 mediated regio - and stereospecific guaninyl - transfer reactions. *Angew. Chem. Int. Ed.* **2019**, *58*, 11534-11540.
182. Yu, H.; Li, S.-M., Two cytochrome P450 enzymes from *Streptomyces* sp. NRRL S-1868 catalyze distinct dimerization of tryptophan-containing cyclodipeptides. *Org. Lett.* **2019**, *21*, 7094-7098.
183. Saruwatari, T.; Yagishita, F.; Mino, T.; Noguchi, H.; Hotta, K.; Watanabe, K., Cytochrome P450 as dimerization catalyst in diketopiperazine alkaloid biosynthesis. *ChemBioChem* **2014**, *15*, 656-659.

Statutory Declaration

Ich, Jing Liu, versichere, dass ich meine Dissertation

„Genome mining-directed discovery of novel 2,5-diketopiperazines from actinobacteria“

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

.

Marburg, den.....

.....
Jing Liu

Acknowledgements

This work was carried out during the period from October 2016 to Decemeber 2020, am Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, der Philipps-Universität Marburg.

First and foremost, I would like to express my sincere gratitude to Prof. Dr. Shu-Ming Li for giving me the opportunity to be involved in the wonderful project. I am sincerely grateful for his invaluable advice, guidance, support, and encouragement throughout my PhD study. His advices on the projects, his endeavors on the manuscripts, and his help in my daily life are of great value. All I have learned and trained in the research group of Prof. Li will benefit me greatly in my future life.

I'm very grateful to Prof. Dr. Peter Kolb for his acceptance to be my second supervisor and co-examiner, and Prof. Dr. Udo Bakowsky and Prof. Dr. Jens Kockskämper for their kindness to be co-examiners.

I would like to thank Dr. Xiulan Xie for recording NMR spectra and her specialized help in analyzing the NMR data, Dr. Regina Ortmann and Stefan Newel for taking NMR spectra, Prof. Dr. Wenhan Lin from Peking University for his help in structural elucidation, Rixa Kraut, Lena Ludwig-Radtke and Johanna Schäfer for LC-MS sample measurement, as well as Prof. Dr. Ulrich Mühlenhoff and Dr. Jens Schäfer for their assistants with CD and UV-Vis spectrum measurements, respectively.

Many thanks to Dr. Huili Yu for her help in the cooperation of the first project in this thesis. Also sincere thanks to Dr. Ge Liao and Yiling Yang for their help of NMR data analysis. Moreover, I'm really appreciated the efforts made by Lauritz Harken, Jenny Zhou and Sina Stierle for translating the summary part and proof reading of the dissertation. Also great thanks to Pan Xiang, Basti Kemmerich, Jonas Nies, Liujuan Zheng, Yu Dai, Haowen Wang, Wen Li and Zhengxi Zhang for the wonderful time we have spent together.

I also want to express my sincere gratitude to all my current and former colleagues, Dr. Kang Zhou, Dr. Shuang Zhou, Dr. Jie Fan, Dr. Huomiao Ran, Dr. Huan Liu, Dr. Jinglin Wang, Dr. Wei Li, Sonja Hiemenz, Sabine Burgers, Dr. Florian Kindinger, Dr. Katija Backhaus, Lena Mikulski, Kristin Öqvist, Dr. Elisabeth Hühner, Linus Naumann, Marlies Peter, Danniell Jonathan Janzen, Andreas Martin, Dr. Nina Gerhards, and Dr. Dieter Kreuzsch.

I would express my special thanks to Prof. Dr. Jianhua Ju, especially for his introducing me into the scientific field of biosynthesis of natural products and his encouragement during these years.

I wish to thank the constant encouragement from my friends outside of the laboratory.

I would like to acknowledge the China Scholarship Council (CSC) for financial supports.

Finally, I would extend my heartfelt thanks to my family. Deeply thanks to my parents for giving me love and supports over the past years. Special and sincerest thanks to my husband Dr. Liang Ma for his consistent support and encouragement.

