

---

# Advancing Myxobacterial Natural Product Discovery by Combining Genome and Metabolome Mining with Organic Synthesis

Dissertation  
zur Erlangung des Grades  
des Doktors der Naturwissenschaften  
der Naturwissenschaftlich-Technischen Fakultät  
der Universität des Saarlandes

von  
**Nicolas Frank**  
Saarbrücken  
2022

Tag des Kolloquiums: 15. Juli 2022

Dekan: Prof. Dr. Jörn Eric Walter

Berichterstatte:r: Prof. Dr. Rolf Müller  
Prof. Dr. Anna K.H. Hirsch

Vorsitz: Prof. Dr. Uli Kazmaier

Akad. Mitarbeiter: PD Dr. Martin Frotscher

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Rolf Müller am Institut für Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes von September 2017 bis April 2022.

**There's no sense crying over every mistake.  
You just keep on trying till you run out of cake.  
And the science gets done.**

Jonathan Coulton in the song „Still Alive” by Ellen McLain

## Danksagung

In erster Linie bedanken möchte ich mich bei meinem Doktorvater Prof. Dr. Rolf Müller für die Aufnahme in seine Arbeitsgruppe, die fortwährende Unterstützung über die gesamte Promotionszeit, sowie für das entgegengebrachte Vertrauen zur eigenständigen Bearbeitung spannender Projekte. Zudem gilt mein Dank Prof. Dr. Anna Hirsch, sowie vertretungsweise Prof. Dr. Alexander Titz, für die Unterstützung als wissenschaftliche Begleiter und für die vielen fruchtbaren Diskussionen im Rahmen der Promotionskomitees.

Ein besonderer Dank geht zudem an meine Betreuer Dr. Daniel Krug und Dr. Fabian Panter sowie Dr. Chantal Bader, die mich tatkräftig bei allen Belangen meiner Dissertation unterstützt haben, immer genug Zeit und ein offenes Ohr für mich hatten, mir stets mit höchst innovativen Lösungsansätzen bei den diversen Herausforderungen meiner Dissertation helfen konnten und viel Mühe in die Überarbeitung meiner Manuskripte und dieser Dissertation investiert haben. Darüber hinaus bin ich froh, dass sich auch privat eine Freundschaft entwickelt hat, die mit Sicherheit auch in Zukunft bestehen bleibt.

Mein Dank geht auch an die gesamte Arbeitsgruppe, die mir während meiner Promotion immer mit wissenschaftlichem oder auch weniger wissenschaftlichem Rat und Tat zur Seite stand. Durch euch konnte ich meinen Horizont konstant erweitern und fachlich sowie persönlich wachsen.

Besonderer Dank gebührt dabei den Bewohnern des Analytik-Büros, Fabian Panter, Chantal Bader und Patrick Haack, sowie den gebührenden Nachfolgern Christine Walt, Alexander Voltz und Joy Birkelbach, und den vielen Besuchern und Ehrenbüromitgliedern wie Jan Schlemmer, Sebastian Groß, Joachim Hug, Sebastian Walesch, Sebastian Adam, Laura Franz, Jake Haeckl, Hu Zeng, Markus Neuber, Selina Deckarm, Anna-Lena Huber, Carsten Seyfert, Christoph Porten, Sabine Backes, Kathrin Andres, Ronald Garcia, Jan Dastbaz, Timo Risch, Sergi Akone, u.v.m., die mit fachlicher Unterstützung glänzen konnten, aber auch in den stressigsten Phasen für die zwingend notwendige Ablenkung von allen negativen Gefühlen sorgen konnten.

Außerdem möchte ich meiner Familie danken, die mich bei allen Entscheidungen immer unterstützt hat und mir stets seelischen und moralischen Beistand geleistet hat. Abschließend bedanken möchte ich mich bei meiner Freundin Christine für ihre unglaubliche Unterstützung sowohl im Labor als auch im Privaten. Du bist stets bemüht mein Leben zu verbessern und das ist dir bisher auch in allen Punkten gelungen. Vielen Dank für deine Unterstützung und auch für Merlin, den süßesten Cavalier auf der ganzen Welt.

## Vorveröffentlichungen aus dieser Dissertation

Teile dieser Arbeit wurden vorab mit Genehmigung der Naturwissenschaftlich-Technischen Fakultäten, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen veröffentlicht oder sind derzeit in Vorbereitung zur Veröffentlichung:

### Publikationen

Nicolas A. Frank,<sup>1</sup> Márió Széles,<sup>1</sup> Sergi H. Akone, Sari Rasheed, Stephan Hüttel, Simon Frewert, Mostafa M. Hamed, Jennifer Herrmann, Sören M. M. Schuler, Anna K. H. Hirsch and Rolf Müller\*: Expanding the Myxochelin Natural Product Family by Nicotinic Acid Containing Congeners. *Molecules* 2021 Aug 14, **26**(16):4929

DOI: [10.3390/molecules26164929](https://doi.org/10.3390/molecules26164929)

Joachim J. Hug,<sup>1</sup> Nicolas A. Frank,<sup>1</sup> Christine Walt, Petra Šenica, Fabian Panter and Rolf Müller\*: Genome-guided Discovery of the First Myxobacterial Biarylittide Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis. *Molecules* 2021 Dec 10, **26**(24):7483

DOI: [10.3390/molecules26247483](https://doi.org/10.3390/molecules26247483)

### Publikationen, die nicht Teil dieser Arbeit sind

Justus Horstmann, Annabelle Laric, Annette Boese, Daniela Yildiz, Teresa Roehrig, Martin Empting, Nicolas Frank, Daniel Krug, Rolf Müller, Nicole Schneider-Daum, Cristiane de Souza Carvalho Wodarz\* and Claus-Michael Lehr\*: Transferring Microclusters of *P. Aeruginosa* Biofilms to the Air–Liquid Interface of Bronchial Epithelial Cells for Repeated Deposition of Aerosolized Tobramycin. *ACS Infect. Dis.* 2021 Dec 17, **8**(2), 137-149

DOI: [10.1021/acsinfecdis.1c00444](https://doi.org/10.1021/acsinfecdis.1c00444)

---

<sup>1</sup> These Authors contributed equally to the manuscript

Thomas Kuhn, Lina Zielke, Ahmad Aljohmani, **Nicolas Frank**, Mina Mehanny, Jessica Hoppstädter, Daniela Yildiz and Gregor Fuhrmann\*: A Cell-Free, Biomimetic Hydrogel Based on *Lactobacillus* Membrane Vesicles Reduces Scar Formation *in Vivo*. **Small** 2022 (Manuscript submitted)

**DOI:**

Philipp Lapuhs, Eilien Heinrich, Lukas Bollenbach, Veronika Stibane, **Nicolas Frank**, Adriely Goes, Thomas Kuhn, Marcus Koch, Alexandra K. Kiemer, Rolf Müller, Kathrin Fuhrmann, Ronald Garcia and Gregor Fuhrmann\*: The Inherent Antibiotic Activity of Bacteria-Derived Autofluorescent Membrane Vesicles Is Switched On and Off by Light Stimulation. **PNAS** 2022 (Manuscript submitted)

**DOI:**

## Tagungsbeiträge

**Nicolas A. Frank**, Fabian Panter, Daniel Krug and Rolf Müller (**2018**) Identification, isolation and structure elucidation of novel myxobacterial secondary metabolites (Oral Communication). **VII. Symposium of the Interdisciplinary Graduate School of Natural Product Research**. 2018 July 12, Saarbrücken, Germany.

**Nicolas A. Frank**, Chantal Bader and Rolf Müller (**2021**) Identification of non-covalent protein binders from a myxobacterial natural product library using native protein mass spectrometry (Poster). **HIPS Symposium 2021**. 2021 May 20, Saarbrücken, Germany.

## Zusammenfassung

Myxobakterien sind eine reichhaltige Quelle für neue Naturstoffe mit vielfältigen chemischen Grundgerüsten und faszinierenden biologischen Aktivitäten. Diese Arbeit behandelt verschiedene aktuelle Methoden, den Sekundärstoffwechsel von Myxobakterien zu erschließen. Das ribosomale Peptid Myxarylin wurde mithilfe eines genomgeleiteten Ansatzes entdeckt. Beschrieben wird außerdem die semisynthesegestützte Isolierung, Strukturaufklärung und heterologe Produktion. Mit statistischer Metabolomanalyse wurde eine Familie lichtinstabiler Verbindungen mit bislang unbekanntem Strukturen entdeckt. Über Inaktivierungsexperimente konnte ein Gencluster identifiziert werden, das vermutlich die Biosynthesemaschinerie dieser Naturstoffe kodiert. Weiterhin wurden neue Myxochelin-Derivate entdeckt, die sich durch den Einbau von Nikotinsäure auszeichnen. Mittels Totalsynthese konnten die Strukturen inklusive Stereochemie aufgeklärt und weitere Derivate hergestellt werden. Zuletzt wurden neue Derivate der Sandacrabine synthetisiert, eine Familie von Terpenoid-Alkaloiden mit vielversprechender antiviraler Aktivität. Das Ziel dabei ist es, die gewünschte Aktivität zu erhöhen und die Zytotoxizität zu verringern. Im Rahmen dieser Arbeit wurden erfolgreich moderne Ansätze in der Naturstoffforschung, insbesondere UHPLC-*hr*MS-basierte Methoden, mit organischer Synthese kombiniert, um die Isolierung zu erleichtern, Strukturen zu bestätigen und neue Derivate myxobakterieller Naturstoffe herzustellen.



## Abstract

Myxobacteria represent a viable source for natural products with a broad variety of chemical scaffolds and intriguing biological activities. This thesis covers different contemporary ways to approach myxobacterial secondary metabolism. The ribosomal peptide myxarylin was discovered through a genome-guided approach. This study describes the discovery, semi-synthesis-assisted isolation, structure elucidation and heterologous production. Furthermore, statistics-based metabolome mining revealed a family of light-sensitive compounds with yet elusive structures. A biosynthetic gene cluster putatively encoding the biosynthetic machinery, could be identified by cluster inactivation experiments. Metabolome mining additionally revealed new myxochelin congeners featuring a rare nicotinic acid moiety. Total synthesis was applied to confirm structures, elucidate the absolute stereochemistry and to generate additional non-natural derivatives. Finally, total synthesis was used to create a small library of sandacrabins, a family of terpenoid-alkaloids that feature promising antiviral activities, with the aim to develop improved congeners with increased target activity and reduced cytotoxicity. The combination of up-to-date approaches in natural products discovery, especially focusing on UHPLC-*hr*MS workflows, and small-scale organic synthesis was successfully applied to facilitate compound isolation, confirm structures and to create novel congeners of myxobacterial natural products.

---

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>15</b>
1.1	Myxobacteria: Prolific Producers of Natural Products	16
1.2	The Biosynthetic Potential of Myxobacteria	18
1.3	The Biosynthesis of Myxobacterial Natural Products	19
1.3.1	Multimodular Polyketide Biosynthesis	19
1.3.2	Non-ribosomal Peptide Biosynthesis	21
1.3.3	Biosynthesis of Ribosomally Produced and Post-translationally Modified Peptides	22
1.3.4	Terpene Biosynthesis	23
1.4	Discovery of Natural Products by Combined Genome and Metabolome Mining	25
1.4.1	Accessing the Myxobacterial Secondary Metabolome	26
1.4.2	Reducing the Complexity of Metabolome Data	26
1.4.3	Dereplication and Prioritization of Metabolites	27
1.4.4	Genome-guided Discovery of Natural Products	28
1.4.5	Linking Genomics to Metabolomics Data	29
1.5	Increasing Yields and Modifying Natural Products	30
1.6	Outline of This Thesis	31
1.7	References	33
<b>2</b>	<b>Genome-guided Discovery of the First Myxobacterial Biarylittide</b>	
	<b>Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis</b>	<b>42</b>
2.1	Abstract	44
2.2	Introduction	45
2.3	Results and Discussion	47
2.3.1	Discovery and Purification of the Myxobacterial Biarylittide MeYLH	47
2.3.2	Structure Elucidation of <b>1</b> and <b>2</b>	50
2.3.3	Heterologous Expression and Biosynthesis	53
2.4	Conclusion	55
2.5	Materials and Methods	56
2.5.1	Applied Software, DNA Sequence Analysis and Bioinformatics Methods	56

---

2.5.2	Myxobacterial Fermentation and Extraction Procedure for LC-MS Analysis	56
2.5.3	Standardized UHPLC-MS Conditions	57
2.5.4	Isolation of <b>1</b> by Supernatant Derivatization and Semipreparative HPLC	58
2.5.5	Structure Elucidation	59
2.5.6	Assessment of Antimicrobial Activities	61
2.5.7	Molecular Cloning, Construction of Plasmids, Maintenance of Bacterial Cultures	61
2.6	Acknowledgments	62
2.7	References	63
S 2	Supporting Information	67
S 2.1	Myxobacterial Culture Media	68
S 2.2	Molecular Cloning, Construction of Expression Plasmids and Genetic Investigation	71
S 2.2.1	Genetic Investigations	72
S 2.2.2	Heterologous Production of Myxarylin in <i>M. xanthus</i> DK1622	73
S 2.3	Compound Characterization	75
S 2.3.1	Tandem MS Spectra	75
S 2.3.2	Structure Elucidation	76
S 2.4	References	95
<b>3</b>	<b>Metabolome-guided Discovery of a Family of Light-sensitive Compounds from <i>Coralloccoccus</i> sp. MCy9049</b>	<b>96</b>
3.1	Abstract	98
3.2	Introduction	99
3.3	Results	100
3.3.1	Discovery of a Novel Family of Chlorinated Compounds	100
3.3.2	Purification of the Most Abundant Compounds <b>5</b> and <b>6</b>	105
3.3.3	Characterization of This Compound Class	107
3.3.4	Metabolome-assisted Genome Analysis	108
3.4	Discussion	111
3.5	Conclusion	115
3.6	Materials and Methods	116
3.6.1	Applied Software, DNA Sequence Analysis and Bioinformatics Methods	116

---

3.6.2	Myxobacterial Fermentation and Extraction for LC-MS Analysis	116
3.6.3	Standardized UHPLC-MS Conditions for Bacterial Crude Extracts	117
3.6.4	Statistical Metabolomics and Similarity Clustering via GNPS	117
3.6.5	Stability Assay and Catalytic Hydrogenation	118
3.6.6	Compound Isolation	119
3.7	Acknowledgments	120
3.8	References	120
S 3	Supporting Information	124
S 3.1	Myxobacterial Culture Media	125
S 3.1.1	Media Optimization	127
S 3.2	Compound Characterization	128
S 3.2.1	Tandem MS Spectra	128
S 3.2.2	UV/Vis Spectra of the Most Abundant Compounds	141
S 3.2.3	Stable Isotope Feeding Experiments	142
S 3.2.4	Proposed Peptide Sequences	153
S 3.3	Analysis of the Putative BGC	154
S 3.4	References	164
<b>4</b>	<b>Expanding the Myxochelin Natural Product Family by Nicotinic Acid Containing Congeners</b>	<b>165</b>
4.1	Abstract	167
4.2	Introduction	167
4.3	Results	169
4.3.1	Discovery of Myxochelin N1	169
4.3.2	Precursor-directed Biosynthesis and Isolation of Myxochelins N1 and N3	170
4.3.3	Structure Elucidation	172
4.3.4	Total Synthesis	175
4.3.5	Biological Activities	178
4.3.6	Biosynthetic Origin of the New Myxochelins N1-N3	179
4.4	Discussion	182
4.5	Conclusion	184
4.6	Materials and Methods	184
4.6.1	Myxobacterial Fermentation and Small-scale Precursor Feeding	184
4.6.2	Analysis of the Secondary Metabolome of Bacterial Crude Extracts	184

---

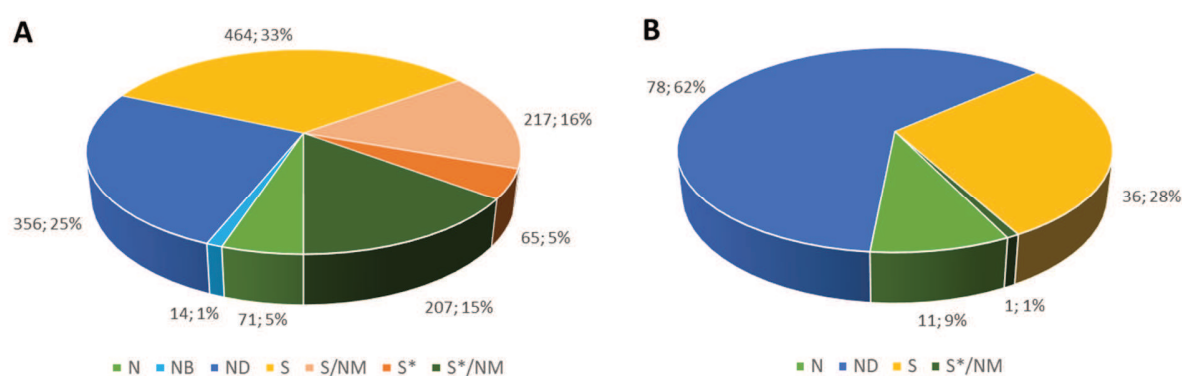
4.6.3	Compound Isolation	185
4.6.4	Structure Elucidation, Chiroptical and Circular Dichroism Measurements	186
4.6.5	Chemistry	187
4.6.6	Characterization of Synthetic Intermediates and Final Products	189
4.6.7	Reaction Monitoring and Purification of Synthetic Compounds	195
4.6.8	Bioactivity	196
4.7	Acknowledgments	197
4.8	References	197
S 4	Supporting Information	202
S 4.1	Characterization of Myxochelin Derivatives	203
S 4.1.1	UV Spectra of the Myxochelins	203
S 4.1.2	Tandem MS Spectra	207
S 4.1.3	Circular Dichroism Spectra	210
S 4.1.4	HPLC-MS Chromatograms of Crude Extracts	212
S 4.1.5	HPLC-MS Chromatograms of the Synthetic Compounds	215
S 4.1.6	Synthesis of Aryl Substituents with an Fmoc Protected Amino Group	223
S 4.1.7	NMR-based Structure Elucidation	225
S 4.2	<i>In-silico</i> Analysis of the Myxochelin N1-N3 Biosynthesis	248
S 4.2.1	Metabolome Database Search of <b>1-3</b>	248
S 4.2.2	Analysis of the Myxochelin BGC	249
S 4.3	Assessment of Biological Activities	251
S 4.4	References	252
<b>5</b>	<b>SAR Studies and Biological Characterization of Sandacrabin Derivatives as Antiviral RNA Polymerase Inhibitors</b>	<b>253</b>
5.1	Abstract	255
5.2	Introduction	256
5.3	Results	257
5.3.1	Evaluation of Biological Activities and Selectivity	257
5.3.2	<i>In Silico</i> Docking Studies with the SARS-CoV-2 RdRp Complex	259
5.3.3	Assessment of Metabolic Stability and Toxicity	260
5.3.4	Modifications of the Benzimidazole Core at Positions 5 and 6	261
5.3.5	Modifications of the Benzimidazole Core at Position 2	262
5.3.6	Bissubstitution of the Modified Benzimidazole Cores	264

---

5.4	Discussion and Outlook	265
5.5	Materials and Methods	267
5.5.1	UHPLC- <i>hr</i> MS Measurements	267
5.5.2	NMR Measurements	268
5.5.3	Computational Docking Studies	268
5.5.4	Biological Assays	268
5.5.5	Assessment of Metabolic Stability	269
5.5.6	Chemical Synthesis of Sandacrabins	270
5.6	References	328
S 5	Supporting Information	331
S 5.1	Antiviral Activities and Estimated Binding Affinities	332
S 5.2	DEREK Analysis	335
S 5.3	ADMET Profiling	337
S 5.4	NMR Spectra	338
S 5.5	References	393
<b>6</b>	<b>Discussion and Outlook</b>	<b>394</b>
6.1	Challenges and Opportunities in Natural Products Discovery	395
6.1.1	Expanding the Detectable Metabolome from Bacterial Crude Extracts	396
6.1.2	Statistical Treatment of UHPLC-MS Data	397
6.1.3	Extracting Early-stage Structural Information from Tandem MS Data	398
6.1.4	Genome Mining	399
6.1.5	Connecting Genome to Metabolome Data	400
6.1.6	Conclusions for Contemporary Natural Products Discovery	403
6.2	Organic Synthesis in Natural Products Discovery, Isolation and Characterization Workflows	404
6.2.1	Total Synthesis of Natural Products	404
6.2.2	Semi-synthetic Approaches in Natural Products Isolation and Characterization Workflows	405
6.2.3	Conclusions on the Use of Organic Synthesis in Natural Products Research	406
6.3	Assessing the Function of Natural Products	407
6.4	Concluding Remarks	409
6.5	References	409

# 1 Introduction

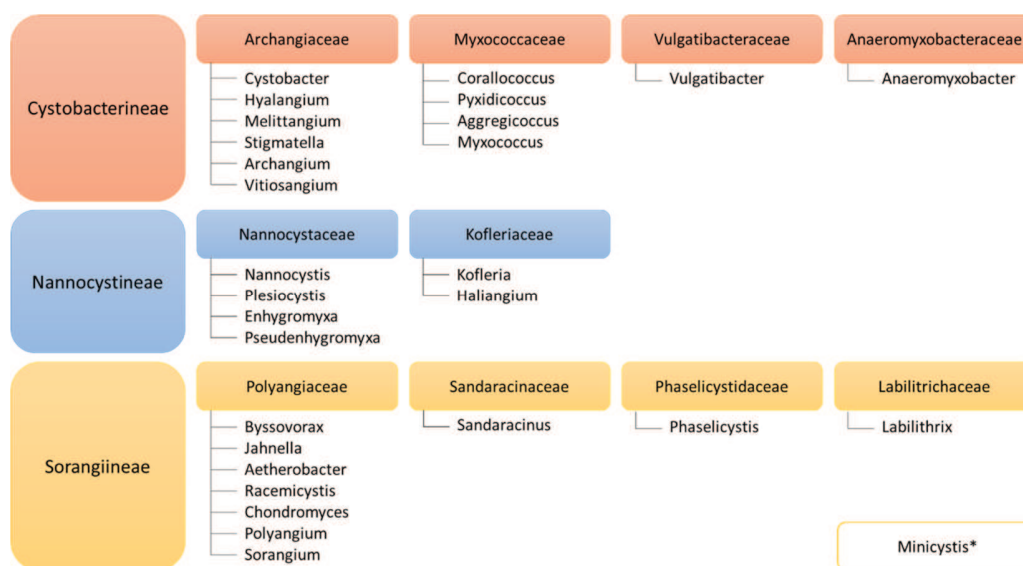
Natural products, also termed secondary metabolites, are a remarkably diverse group of small molecules that are mainly produced by plants, fungi and eubacteria [1]. Although they are usually not involved in primary metabolism and thus not imminently essential for cell survival [2], the high energy demand during their production suggests that the producing organism by some means benefits from the biosynthesis of natural products, which grant it an evolutionary advantage over competitors [3]. Many natural products are known to exhibit distinct biological functions such as cell signaling, uptake of trace metal ions or defense against competing organisms, which requires them to specifically interact with their target molecules. Albeit in various instances, their ecological purpose remains elusive, as evolution impelled the optimization of natural products towards highly specific and potent compounds, natural products have contributed many lead structures to human medicine [4]. Therapeutic application is typically closely related to the proposed biological functions such as antimicrobial, antiviral or cytotoxic compounds that are used to treat infections caused by bacteria, viruses, fungi or parasites, or find their application in cancer treatment. On the other hand, natural products are also used for other indications like analgesics, hypocholesterolemic or immunosuppressants, just to name a few. Over the last roughly 40 years, 66% of all approved small-molecule drugs and 71% of all antibacterial drugs were natural products or natural product-derived (Figure 1).



**Figure 1.** All small-molecule approved drugs (A) and all antibacterial drugs (B) from 01/1981 to 09/2019 by source. N: unaltered natural product; NB: botanical drug; ND: natural product derivative; S: synthetic drug; S/NM: synthetic drug/mimic of natural product; S\*: synthetic drug with natural product pharmacophore; S\*/NM: synthetic drug with natural product pharmacophore/mimic of natural product. Adapted from [4].

## 1.1 Myxobacteria: Prolific Producers of Natural Products

Bacteria of the order *Myxococcales*, commonly named myxobacteria, are rod-shaped Gram-negative  $\delta$ -proteobacteria that are ubiquitously distributed throughout the world. They are able to colonize diverse natural habitats such as temperate zones, tropical rain forests, arctic tundra, deserts, caves as well as marine and saline environments [5–9]. They are mostly aerobic, chemoheterotrophic and mesophilic [10], with exceptions like the facultative anaerobic *Anaeromyxobacter dehalogenans* [11] or several moderately thermophilic myxobacteria of *Cystobacterinae* and *Sorangiiineae* suborders [12]. As of May 2018, the order *Myxococcales* consists of 3 suborders, 10 families and 29 genera (Figure 2).

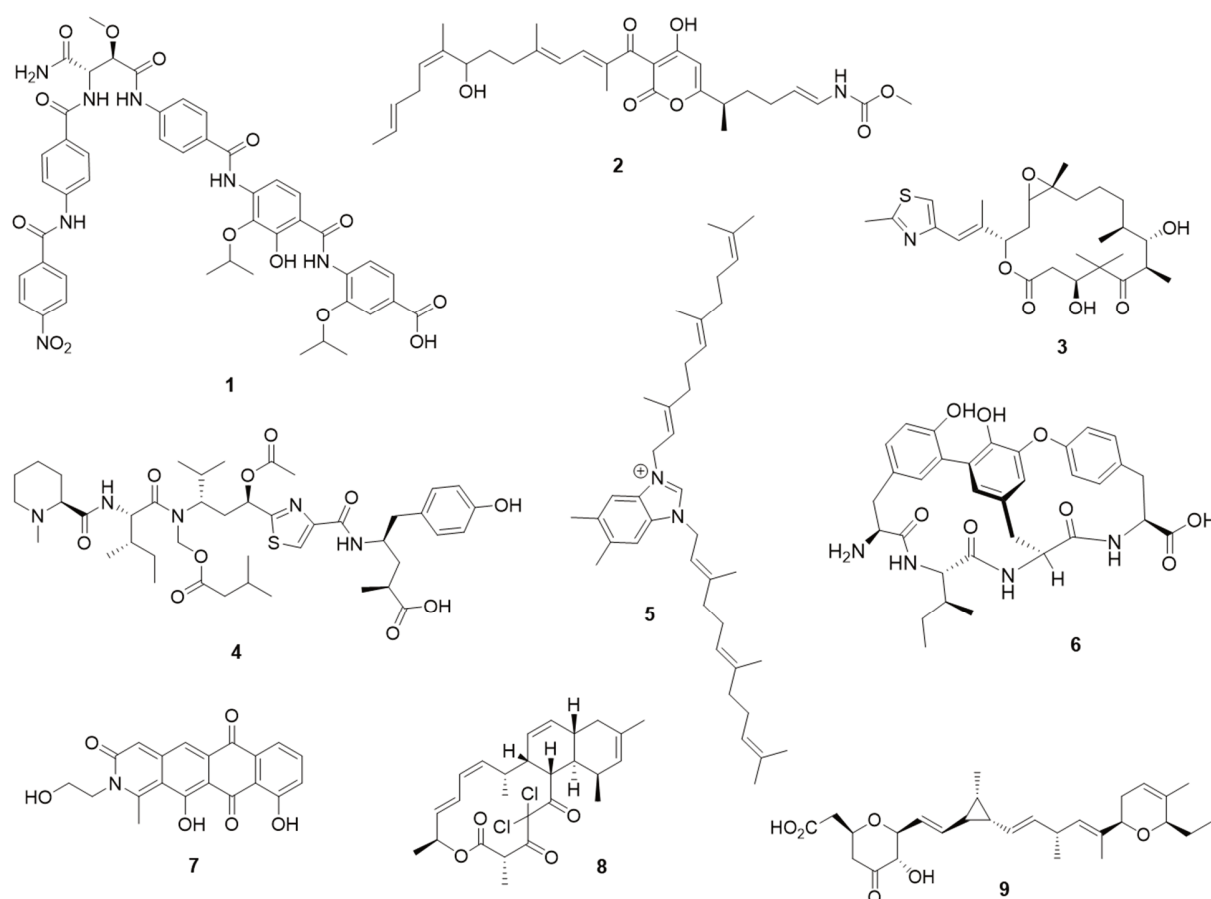


**Figure 2.** Suborders, families and genera of the order *Myxococcales*. \* The genus *Minicystis* is considered “incertae sedis” and thus not assigned to a family in the *Sorangiiineae* suborder. Adapted from [10,13].

Myxobacteria have mainly been isolated from soil, bark, rotting wood, leaves, compost and dung, which are usually nutrient-rich, but highly competitive environments [5,14]. Many myxobacteria, especially from the *Cystobacteriineae* suborder are predatory and feed on other soil-dwelling microorganisms. Other myxobacteria however, in particular from the genera *Sorangium* and *Byssovorax*, are non-predatory decomposers that are able to efficiently degrade organic macromolecules such as cellulose, chitin or xylan [7,15]. Unlike other bacteria, myxobacteria are social bacteria that move in coordinated multicellular groups, termed swarms, and feed cooperatively in predatory groups. Their social behavior enables them to share secreted hydrolytic enzymes and secondary metabolites in order to increase the concentrations in their local environment and to optimize the use of degradation products [16]. Furthermore, under starvations conditions, myxobacterial swarms have the ability to form



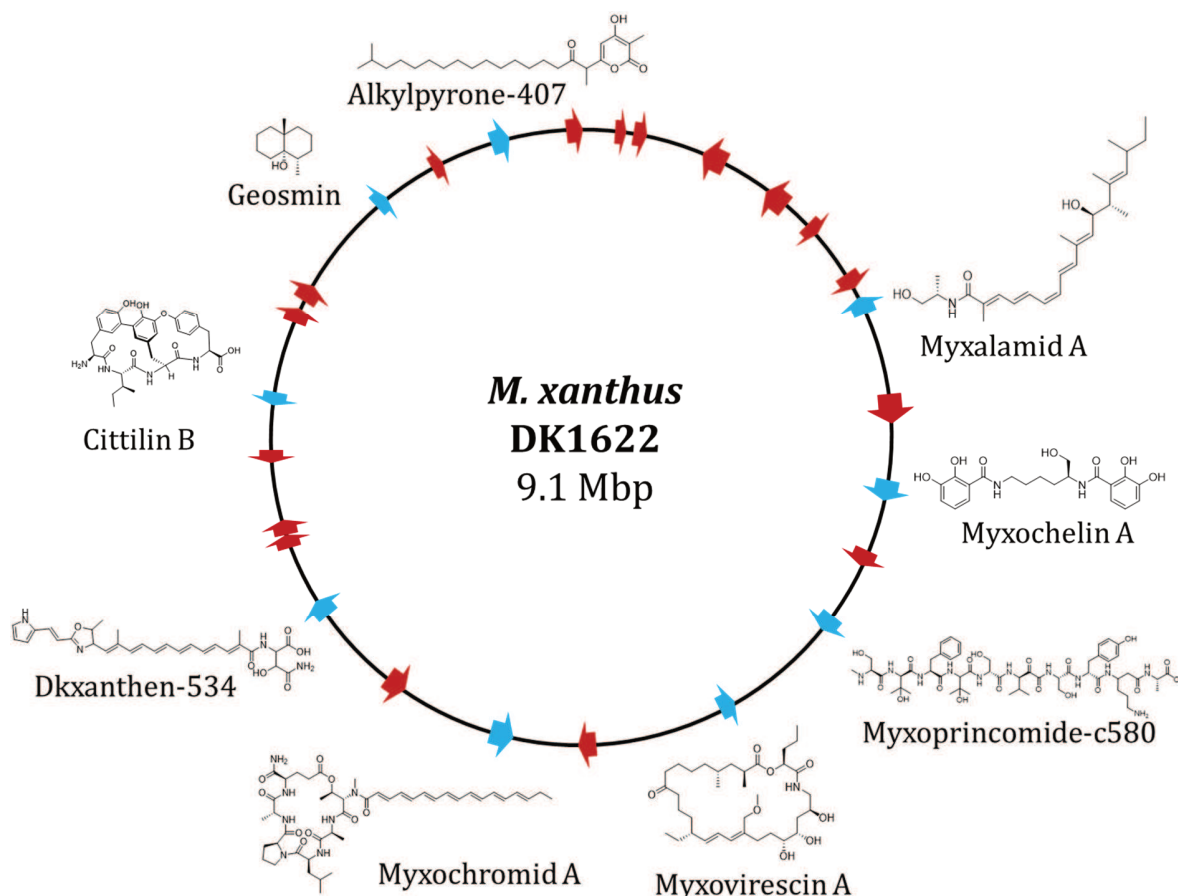
macroscopic fruiting bodies, within which a small fraction of the cells differentiates into spherical dormant spores, while most of the cells lyse or differentiate into persistor-like rods [17]. In addition to their remarkable biological characteristics, myxobacteria are an excellent source for natural products featuring a broad variety of chemical scaffolds as well as biological activities [18] as exemplified in Scheme 1. While many myxobacterial natural products are known to exhibit strong biological activities, several compounds have been isolated that have no alleged biological activity yet. However, since these scaffolds have been evolutionary developed to gain an advantage over competing microorganisms, it is likely that these compounds show activity against so far unknown target structures and might become useful therapeutics in the future.



**Scheme 1.** Examples for the structural, functional and biosynthetic diversity of myxobacterial natural products. **1:** Cystobactamid 919-2 (Antibacterial, Non-ribosomal peptide synthetase (NRPS)-derived) [19]; **2:** Corallopyronin A (Antibacterial, Type I polyketide synthase (PKS)-derived) [20]; **3:** Epothilon B (Cytotoxic, Type I PKS-NRPS hybrid-derived) [21]; **4:** Tubulysin A (Cytotoxic, NRPS-derived) [22]; **5:** Sandacrabin B (Antiviral, Terpenoid-alkaloid) [23]; **6:** Cittalil B (in vitro inhibition of CsrA, Ribosomally synthesized and post-translationally modified peptide (RiPP)) [24]; **7:** Pyxidicyclin A (Cytotoxic, Type II PKS-derived) [25]; **8:** Chlorotonil A (Antiplasmodial, Type I PKS-derived) [26]; **9:** Ambruticin (Antifungal, Type I PKS-derived) [27].

## 1.2 The Biosynthetic Potential of Myxobacteria

Myxobacteria possess the largest prokaryotic genomes known to date with chromosome sizes of up to 15 MBps [28–30]. While the absolute genome size provides only limited information of the biosynthetic potential, the number of biosynthetic gene clusters (BGCs) encoded in the genome reflects the theoretical capacity of natural product production more effectively. Nowadays, a variety of bioinformatics tools such as the “antibiotics and secondary metabolites analysis shell” (antiSMASH) [31] or “prediction informatics for secondary metabolites (PRISM) [32] are available that allow rapid and comprehensive prediction as well as annotation of BGCs in bacterial genome data. The biosynthetic proficiency of myxobacteria can be judged by the huge discrepancy between the estimated number of biosynthetic pathways encoded in the genome and the number of natural products actually detected or isolated [33]. Even the best investigated producers like the myxobacterial model strain *Myxococcus xanthus* DK1622 proved to incorporate a much higher biosynthetic potential than expected from the number of already isolated natural products [34,35].



**Figure 3.** Schematic overview of BGCs and known compounds from the myxobacterial model strain DK1622. Blue arrows represent known BGCs; red arrows represent BGCs with no associated product yet. For visibility reasons, some BGCs are depicted larger than to scale. Adapted from [34,36].

## 1.3 The Biosynthesis of Myxobacterial Natural Products

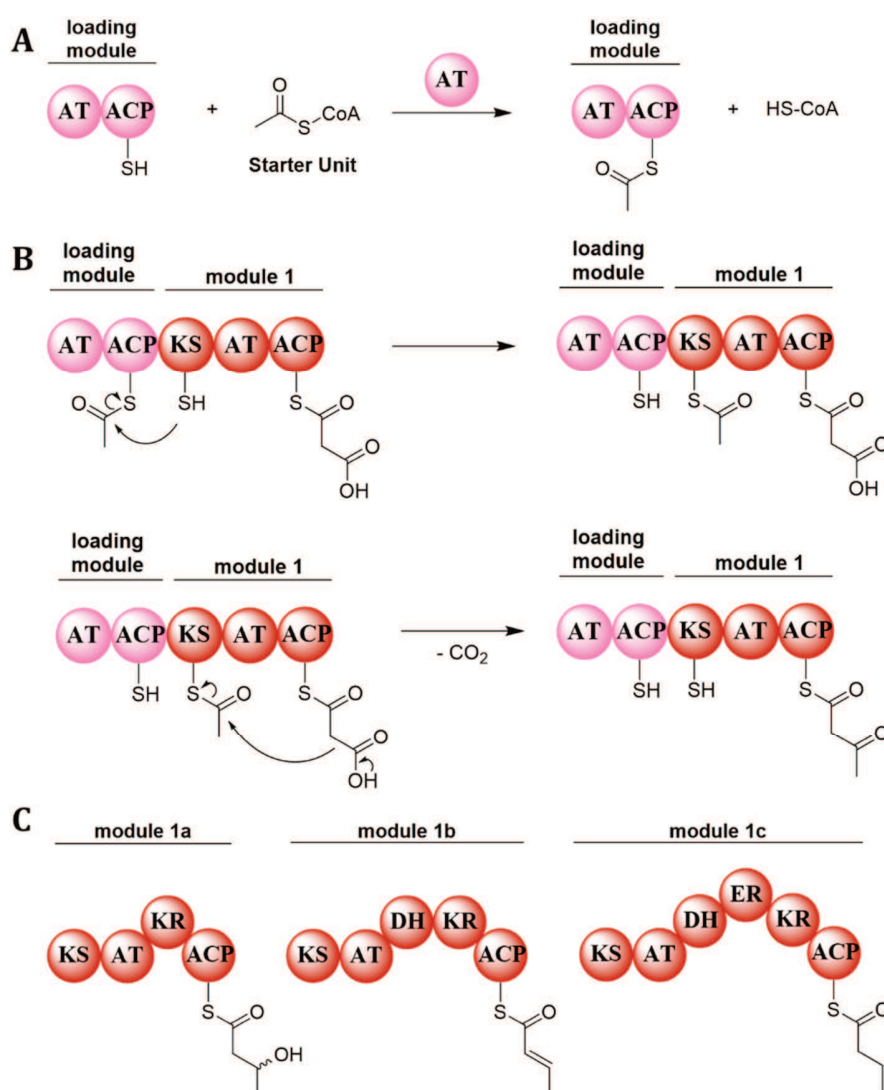
The broad functional variety of natural products is reflected by a vast structural diversity, which is generated by specialized biosynthetic machinery. Myxobacterial natural products belong to several compound classes such as polyketides, non-ribosomal peptides, ribosomally produced and post-translationally modified peptides (RiPPs) and terpenes, as well as hybrids of these classes. The majority of myxobacterial natural products known to date derive from modular nonribosomal peptide synthetases (NRPS), polyketide synthases (PKS) and hybrids thereof, while natural products from RiPP and terpene BGCs have been isolated less frequently [35]. Due to their relevance, especially considering the natural products described in this thesis, the biosynthetic machinery of PKS, NRPS, RiPPs and terpene pathways are briefly described in the following sections.

### 1.3.1 Multimodular Polyketide Biosynthesis

The core biosynthetic principle of polyketide biosynthesis involves repetitive Claisen condensation of an extender unit with an activated carboxylic acid starter unit similar to fatty acid biosynthesis [37]. This reaction forms a 1,3-diketo moiety that represents the nascent polyketide chain and is subsequently modified and further extended. Polyketides are synthesized by huge biosynthetic enzyme complexes that are divided into functional units called modules which again consist of several domains exhibiting different functions. On the basis of different architecture and biochemistry, polyketide biosynthesis pathways are classified into three distinct classes [38]. In bacterial type I PKS systems, also termed modular polyketide synthases, modules are covalently connected in multifunctional megasynthases where each module catalyzes chain elongation by one extender unit as well as subsequent modification. Contrary to that, type II and III PKS systems, which are not further discussed here, involve freestanding monofunctional and iteratively-acting enzymes [37].

In modular type I PKS systems, biosynthesis is initiated by condensation of an acyl starter unit to the phosphopantetheinyl thiol group of an acyl carrier protein (ACP) by an acyl transferase (AT) domain (Scheme 2A). Typical starter units include acetyl-CoA, propionyl-CoA, or less commonly other acyl-CoA thioesters like malonyl-CoA and benzoyl-CoA. Furthermore, also several free carboxylic acids can be activated and loaded by NRPS-like adenylation-thiolation (A-T) didomains to serve as starters in PKS systems [39]. The starter unit is next transferred to the ketosynthase (KS) domain of the adjacent elongation module. Following loading of an extender unit to the ACP domain of the elongation module by its AT domain, the ACP delivers the loaded extender unit to the KS domain for a decarboxylative Claisen condensation that appends it to the nascent polyketide chain (Scheme 2B). In addition to malonyl-CoA and methylmalonyl-

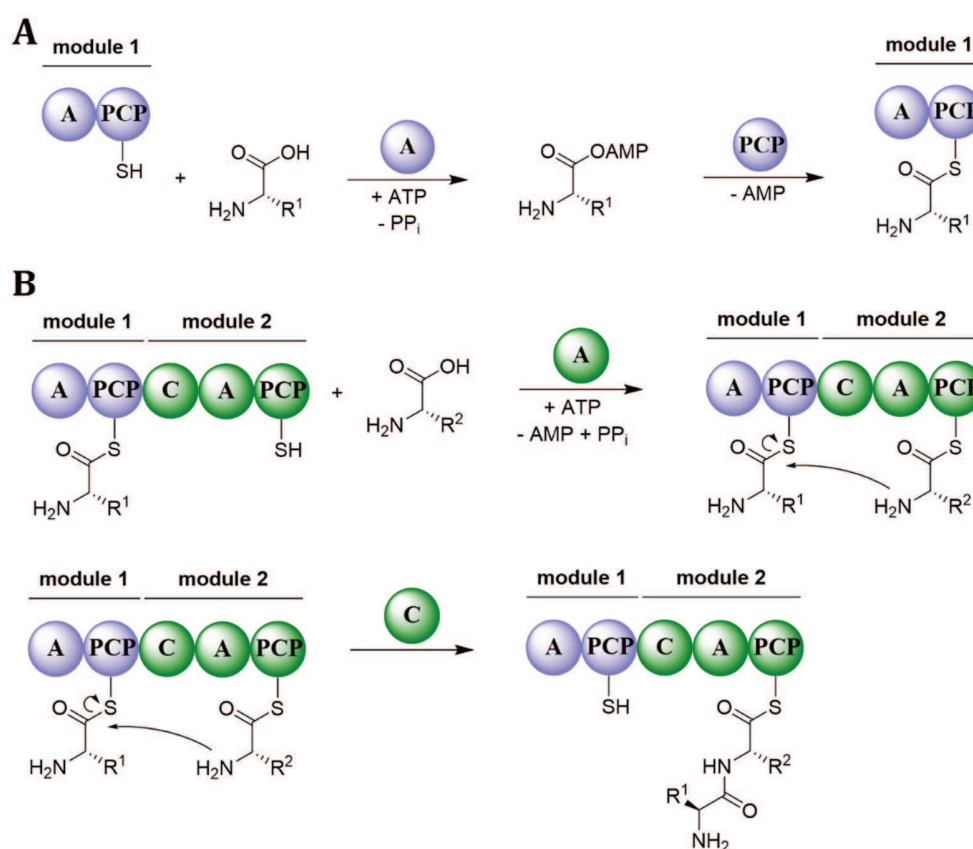
CoA, which are mainly used as an extender units, other CoA- and ACP-linked extender units like ethylmalonyl-CoA and methoxymalonyl-ACP can also be used [40,41]. Elongation modules can additionally include further domains that catalyze tailoring reactions on the resulting  $\beta$ -ketone moiety such as reduction of the ketone to a secondary alcohol by ketoreductase (KR) domains, double bond formation by dehydration of the primary alcohol by dehydratase (DH) domains and formation of a saturated  $C_2$  unit by reduction of the double bond by enoyl-reductase (ER) domains subsequently (Scheme 2C) [42]. The mature polyketide chain is usually released from the assembly line by a terminal thioesterase (TE) by hydrolysis or macrocyclization. Finally, the released polyketide is prone to further modifications like hydroxylation, methylation or glucosylation to yield the final PKS product.



**Scheme 2.** Loading of an acetyl-CoA starter unit on an ACP catalyzed by an AT domain (A), chain elongation by decarboxylative Claisen condensation (B) and products of differently organized elongation modules (C). Each color represents one module. Abbreviations are explained in the main text above.

### 1.3.2 Non-ribosomal Peptide Biosynthesis

Analogous to modular type I PKS systems as described in section 1.3.1, non-ribosomal peptide (NRP) biosynthesis relies on huge biosynthetic enzyme complexes. These so-called megasynthetase enzymes are organized into modules consisting of several domains exhibiting different functions. Each module is thereby linked to elongation of the nascent peptide chain by one specific amino acid [43]. Generally, NRP biosynthesis is initiated by activation of the carboxylic acid function of distinct amino acid monomer by an adenylation (A) domain to yield an adenosylmonophosphatidyl (AMP) ester, which is subsequently condensed to the phosphopantheinyl thiol group of a peptidyl carrier protein (PCP) domain (Scheme 3A). In the adjacent elongation module, the next amino acid is activated and loaded to its PCP domain by the successive A domain. Consequently, a condensation (C) domain catalyzes peptide bond formation by transferring the incoming peptide chain to the freshly bound amino acid (Scheme 3B). The mature peptide is usually released from the assembly line by a terminal thioesterase (TE) or reduction (Red) domain and can be further modified by tailoring enzymes.

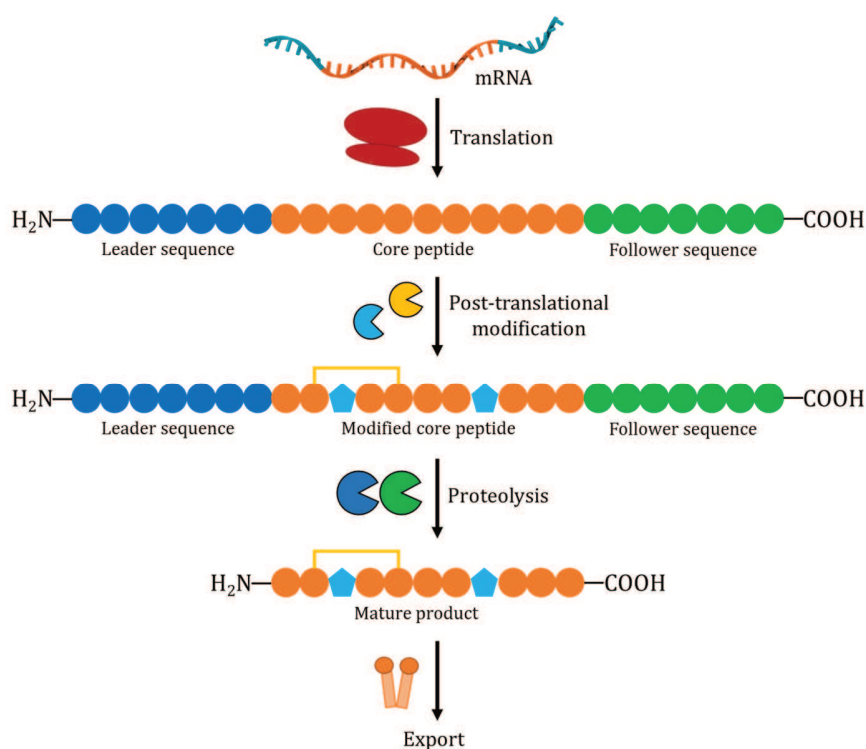


**Scheme 3.** Activation of an amino acid and loading on a PCP catalyzed by an A domain (A) and chain elongation by peptide bond formation catalyzed by a C domain (B). Abbreviations are explained in the main text above.

Similar to type I PKS, several additional domains can be included in NRPS modules that expand the chemical space available from NRP biosynthesis. Examples are epimerase (E), C-, N- and O-methyltransferase (cMT, nMT, oMT), oxidase (Ox), heterocyclization (HC) and formylation (F) domains. Furthermore, in contrast to ribosomal peptide and protein assembly, NRP biosynthesis is not limited to the 21 canonical amino acid building blocks, but allows incorporation of many uncommon amino acids as for example p-aminobenzoic acid in cystobactamid biosynthesis [19], tubuvalin in tubulysin biosynthesis [44] or 6-chloromethyl-5-methoxy-pipecolic acid in chloromyxamide biosynthesis [45]. The starter unit that is loaded onto the first PCP domain is even more variable as it does not necessarily require a primary or secondary amine moiety for chain elongation as demonstrated by incorporation of long chain fatty acids in vioprolide biosynthesis [46] or several benzoic and nicotinic acid derivatives in myxochelin biosynthesis [47–49]. So far, about 500 naturally occurring amino acids have been identified that can be incorporated in non-ribosomal peptides [50].

### **1.3.3 Biosynthesis of Ribosomally Produced and Post-translationally Modified Peptides**

Ribosomally produced and post-translationally modified peptides (RiPPs) are, as their name implies, ribosomally produced peptides that undergo some degree of post-translational modification. Their biosynthesis is referred to as post-ribosomal peptide synthesis (PRPS) analogous to non-ribosomal peptide synthesis (NRPS) discussed in the previous section [51]. Their biosynthesis is typically initiated by ribosomal translation of a precursor peptide encoded by a structural gene. Beside the core peptide, which is modified by tailoring enzymes to yield the mature product, the precursor peptide may also contain an N-terminal leader sequence or a C-terminal follower sequence that are important for recognition by post-translational modification enzymes and export from the cell [52]. Eukaryotic precursor peptides often additionally include an N-terminal signal sequence that directs the peptide to specific cellular compartments. After modification of the core peptide by tailoring enzymes, the leader and follower sequences are cleaved off by proteolysis and the mature product is exported (Scheme 4).



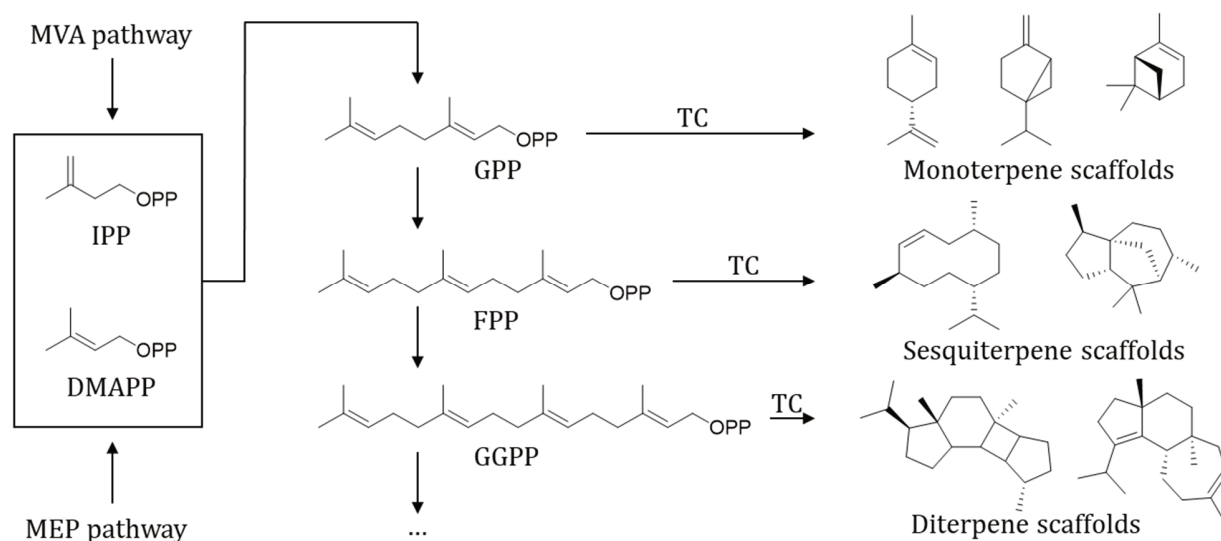
**Scheme 4.** Schematic overview of post-translational peptide synthesis (PRPS). Adapted from [51].

Unlike in NRP biosynthesis, the chemical space of ribosomal peptides is initially limited to the canonical 21 proteinogenic amino acids. Nevertheless, extensive post-translational modification leads to high structural diversity comparable to NRPS [53]. In fact, many ribosomally produced natural products contain structures very similar to non-ribosomal peptides, which highlights the similarity in chemical space obtained by both biosynthetic machineries. By the types of post-translational modification, RiPPs are subdivided into a growing number of various compound classes as for examples the lanthipeptides containing lanthionine and 3-methylanthionine moieties.

### 1.3.4 Terpene Biosynthesis

All terpenoids, or isoprenoids, are biosynthesized from the universal  $C_5$  precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are assembled through the 2-C-methyl-D-erythritol-4-phosphate (MEP) or the mevalonic acid (MVA) pathway [54,55]. While most bacteria synthesize IPP and DMAPP through the MEP pathway, myxobacteria use the MVA pathway for terpenoid biosynthesis [56]. The isoprenoid precursors are joined together by isoprenyl transferases to form hydrocarbon backbones of varying chain lengths such as geranyl diphosphate (GPP,  $C_{10}$ ), farnesyl

diphosphate (FPP, C<sub>15</sub>) and geranylgeranyl diphosphate (GGPP, C<sub>20</sub>). The linear hydrocarbon backbones can consist of more than 5000 C<sub>5</sub>-bodies and reach molecular weights of more than 2 million Dalton as for example reported for the polyterpenes gutta-percha and natural rubber [57]. Subsequently, the linear terpenoid backbones are cyclized by oligoprenyl synthetases or terpene cyclases (TC) to generate for instance diverse mono- (C<sub>10</sub>), sesqui- (C<sub>15</sub>) and diterpene (C<sub>20</sub>) scaffolds (Scheme 5). In a second phase of the biosynthesis, the terpene scaffolds are heavily modified by tailoring enzymes to generate the mature product [54,55]. Although all terpenoids originate from the same hydrocarbon backbones, the variety of different cyclization enzymes and subsequent terpene scaffold functionalization lead to an enormous structural diversity with over 80,000 compounds grouped into more than 400 structural families described in the dictionary of natural products that can largely be explained by carbocation induced chemistry [58].



**Scheme 5.** Biosynthesis of representative mono-, sesqui- and diterpene scaffolds. MVA: mevalonic acid pathway, MEP: 2-C-methyl-d-erythritol-4-phosphate pathway, IPP: isopentenyl diphosphate, DMAPP: dimethylallyl diphosphate, GPP: geranyl diphosphate, FPP: farnesyl diphosphate, GGPP: geranylgeranyl diphosphate, TC: terpene cyclase. Adapted from [55].



## 1.4 Discovery of Natural Products by Combined Genome and Metabolome Mining

Since the discoveries of the first antibiotic natural products in the 1930s, natural product discovery workflows changed significantly to adapt towards newly emerging challenges and improve the outcome of natural product research [33]. Most of the antibiotic drugs from the so-called golden era of antibiotics discovery that lasted until around the 1960s were discovered by bioactivity-guided isolation from *Actinobacteria* [59]. As antibiotic drugs that are produced by well-known strains in high yields and those that feature strong mass spectrometry (MS) and ultra-violet (UV) spectroscopy signals have predominantly been discovered and well characterized, the rediscovery of already known natural products is a commonly arising problem in natural product research [33,60]. Still, most of the already investigated microbial strains have only been used to isolate compounds from a single natural product family despite the fact that almost all strains are, according to *in-silico* predictions, able to produce many more natural products [34]. On top of that, the cultivation of less-investigated microorganisms vastly enhances the odds of discovering natural products not described yet [61]. Metagenomics findings furthermore suggest that the majority of microorganisms have not been cultivated under laboratory conditions yet, potentially leaving a huge chemical space still untouched [62]. In order to increase the chances of discovering new natural products, novel microbial sources have to be explored while using improved analytical systems and methods to detect the metabolites that are hard to access [63,64]. While advanced techniques make it possible to detect more metabolites despite low production rates or unfavorable polarity or ionization properties, methods for dereplication and prioritization of MS signals are becoming increasingly important. With the increase of publicly available genome data due to the development of fast and cheap genome sequencing technologies like next generation sequencing (NGS) and single molecule real time (SMRT) sequencing as well as reliable gene cluster prediction, the probability of identifying novel natural products can be significantly increased by selection of microbial species with a high theoretical secondary metabolite production potential [65]. Combining the fields of secondary metabolomics and microbial genomics in association with bioinformatics tools can accelerate natural product discovery and enable the correlation of cryptic biosynthetic gene clusters with their associated products [66]. The following sections will focus on general approaches for mining large quantities of metabolome and genome data in order to identify promising microbial natural products.

### 1.4.1 Accessing the Myxobacterial Secondary Metabolome

Bacterial crude extracts are highly complex mixtures of primary and secondary metabolites as well as components co-extracted from cultivation media and other contaminants that usually span the entire polarity spectrum and cover a wide mass range [67]. Secondary metabolites are often minor constituents of these extracts featuring MS and UV signals orders of magnitude less intense than media-derived compounds, which impedes reliable identification and analysis [68]. The so-called 'low-hanging fruits' of the secondary metabolome that are easy to detect due to prominent signals and high yields have mainly been isolated and characterized shifting the mission towards compounds that are more challenging to detect [69]. Current secondary metabolome analysis workflows commonly feature a combination of highly standardized liquid chromatography (LC) conditions with fast scanning high resolution MS acquisition enabling broad detection coverage of natural products including additional information, like the elemental composition of the analytes [33,70]. Despite continuous advances in LC-MS based techniques, a variety of natural product scaffolds are still difficult to detect because of challenging polarity and ionization properties. Very polar substances like pseudouridimycin or myxarylin have almost no retention in reversed phase high performance liquid chromatography (RP-HPLC) and suffer from ion suppression in the injection peak range [71,72], whereas highly nonpolar compounds such as terpenoids or steroids might not elute from RP-HPLC columns in standardized gradients at all. Furthermore, the commonly used electrospray ionization (ESI) fails to ionize natural product scaffolds with low oxygen and nitrogen content, further complicating the detection of various terpenes and steroids [73]. Last but not least, eluents for LC usually contain organic acids to enhance column separation and facilitate ionization that might also degrade acid-sensitive compounds. Summing up, although technology continues to lower the limit of detection and increase separation power, no single method can map the entire bacterial metabolome. Nevertheless, as high-throughput LC-MS-based workflows generate huge amounts of complex data, the use of highly standardized analytical methods is inevitable for natural product discovery as it allows for statistical data treatment as well as automated database dereplication of already known compounds.

### 1.4.2 Reducing the Complexity of Metabolome Data

The complexity and quantity of measurement data generated by high-throughput HPLC-MS-based workflows exceeds the capabilities for manual data curation and analysis by far [68]. To effectively reduce complexity and data size, automated background subtraction by statistical methods has to be performed in order to enable efficient mining for new natural products [74]. Assuming that LC-MS parameters are standardized and tightly controlled, the resulting chromatogram can be divided into so-called molecular features. A molecular feature is the noise-free combination of all isotopic peaks of a single putative

molecular entity together with the retention time (RT) of its intensity maximum and the mass-to-charge ratio ( $m/z$ ) of the monoisotopic peak. Subsequently, the whole RT and  $m/z$  range of the chromatogram is divided into rectangular subspaces with defined RT and  $m/z$  widths called buckets. If a bucket contains a molecular feature, this bucket gets subsequently assigned the maximum intensity value of this molecular feature, else the bucket gets assigned the value zero. The bucket table, or feature table, generated in this way can then be used for further comparative statistical analysis [68,75]. For bacterial crude extracts, for example, separating and subtracting 'blank' culture medium-derived molecular features from LC-MS features associated with microbial growth can be a key step to ensure that relevant precursor ions are adequately covered by subsequent tandem MS experiments [76]. A frequently used approach to statistically compare large data sets such as bucket tables generated from LC-MS data is principal component analysis (PCA) [77]. Metabolomics experiments can effectively use PCA to compare large sets of LC-MS data from two or more separate groups enabling the separation of features derived from culture medium from those of the bacterial metabolome. The PCA results can be used to reserve the duty cycle of subsequent tandem MS spectra acquisition to precursor ions associated with the bacterial metabolome and thus to massively improve the quality and significance of the resulting data [33,76].

### 1.4.3 Dereplication and Prioritization of Metabolites

Even after excessive data preparation and filtering of unwanted background features, the challenge of perpetual rediscovery of known metabolites remains eminent. One way to face this challenge is the annotation of feature tables with information from compound databases such as the Dictionary of Natural Products or the Natural Products Atlas [58,78]. These publicly available databases contain data on several thousand natural products that can be used to annotate feature tables using calculated  $m/z$  values, as well as simulated isotopic patterns expressed by  $m$ Sigma values. Additional information can be obtained from metadata, for example on phylogeny, biosynthesis or related compounds. However, highly standardized analytical conditions make it possible to develop own databases from known metabolites, which significantly increase the certainty of annotation by integrating additional parameters like retention time, ion mobility or even tandem MS fragmentation data. To facilitate the process of database-assisted dereplication of knowns, several annotation tools such as Bruker Metaboscape or DEREP-NP can be used [79]. Another way to dereplicate metabolome data is through the analyses of tandem MS fragmentation data. Comparison of fragmentation data allows not only to identify the parent compound, but additionally to detect related compounds by assigning it to groups of similar fragmentation patterns. Furthermore, tandem MS analysis tools make it possible to predict interesting structural elements at an early stage and to prioritize metabolites on the basis of these elements [80]. Finally, spectral networking tools like Global

Natural Products Social Molecular Networking (GNPS) [81] are highly beneficial in bacterial secondary metabolomics. Spectral networking works by linking molecular features with similar fragmentation spectra, creating a network in which structurally related metabolites are clustered together. Additionally, these networks can be automatically annotated with reference data from various community-curated repositories allowing efficient dereplication and prioritization of metabolites by looking for outlier clusters in the spectral network containing putative uncommon structural elements. Summing up, the various approaches and tools introduced in the previous sections efficiently prevent rediscovery and enable prioritization of molecular features towards the discovery of structurally interesting natural products.

#### **1.4.4 Genome-guided Discovery of Natural Products**

The development of next generation sequencing techniques significantly dropped the cost of full genome sequencing while simultaneously increasing reliability. This improvement led to a completely new approach in natural product research called genome mining, which describes the prediction and subsequent isolation of natural products solely on the basis of genetic information without any knowledge about the underlying structures [82]. The genome sequencing revolution triggered the development of a variety of bioinformatics tools such as antiSMASH or PRISM allowing efficient prediction and annotation of biosynthetic genes and BGCs and furthermore grant early insights into the putative structure of encoded metabolites [31,32]. Similar to the previously described metabolome data, dereplication of genome data using publicly available data repositories such as “Minimal Information about a Biosynthetic Gene cluster” (MIBiG) can significantly reduce the probability of rediscovery of known natural products and thus prevent repetition of work that has already been performed [83,84]. In contrast to genome sequencing, which can be conducted quickly in a high-throughput manner, the subsequent progress from obtaining BGC data to the isolation of a new natural product family is time-consuming and expensive, especially if it involves genetic engineering steps like gene cluster activation or heterologous expression [33]. This discrepancy highlights that to avoid unnecessary work, an efficient strategy to prioritize gene clusters is inevitable [85]. One way to prioritize BGCs encoding for antimicrobial natural products is to screen for self-resistance markers or ABC transporters in close proximity to predicted gene clusters as exemplified by the discovery of the pyxidicyclins [25]. While this approach vastly increases the chances of discovering a bioactive compound, it is based on known resistance markers and therefore often fails to uncover novel modes of actions reducing the clinical usability. Nevertheless, even a copy of a known target gene in proximity to a BGC can still provide a novel mode of action. Another approach towards BGC prioritization is based on the complexity and organization of gene clusters. For many gene clusters, the biosynthetic energy demand for the production of a natural product can be estimated by the size and complexity of the corresponding

BGC. Assuming that a microorganism has an optimized energy budget, it can be anticipated that large and highly complex clusters produce compounds that somehow justify the high energy consumption by granting a survival benefit [86]. Last but not least, gene cluster prioritization can also be performed based on their predicted products. In this rather classical approach, genomes are screened for homologs of conserved genes involved in natural product biosynthesis such as PKS and NRPS domains or specific tailoring enzymes. Especially modular PKS and NRPS pathways allow for comprehensive prediction of natural product scaffolds by identifying specific domains involved in the putative biosynthesis [87,88]. In contrast to PKS and NRPS clusters containing catalytic domains with high sequence homology, BGCs associated with RiPP and terpene biosynthesis are way more difficult to identify and predict as they usually contain small and poorly conserved open reading frames (ORFs) [89,90]. This dilemma also highlights the general disadvantage of genome mining as most bioinformatics tools rely on homology to already known BGCs or genes limiting the capability to identify novel classes of BGCs or genes encoding for novel enzymes involved in natural products biosynthesis. Nevertheless, genome mining is still an excellent tool to prioritize strains or interesting BGCs for further investigations leading to the discovery of novel metabolites.

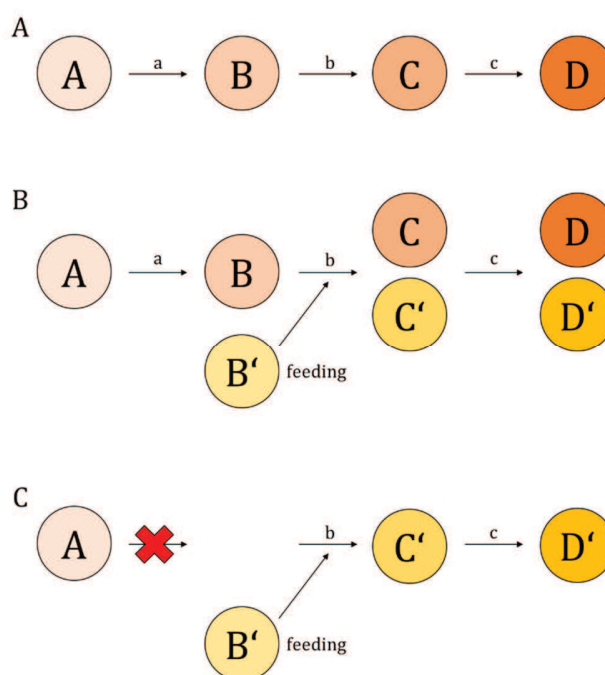
#### 1.4.5 Linking Genomics to Metabolomics Data

In many cases, genomics and metabolomics are still considered separate or consecutive approaches to discover and characterize new natural products. However, if these two methods are combined, the benefits of both can be reaped to accelerate natural product research significantly [66]. For instance, comparable statistics experiments as described in the previous sections can help to link MS signals to biosynthetic gene clusters, *e.g.* by comparison of the metabolome of different strains and subsequent correlation of the appearance of MS features and BGCs, or by comparing the metabolome data of wild type strains with that of cluster inactivation or overexpression mutants. In addition, best chances for the discovery of structurally novel metabolites arise by prioritizing strains that are taxonomically distant from the well-characterized type strains as also their secondary metabolite profiles are usually distinct from known strains [61]. For most of these strains, genetic engineering tools, for example for direct activation of a BGC, do not exist, thus only allowing for labor-intensive methods such as heterologous expression of the selected BGC in a different production strain. Preceding this laborious process by screening the strains' genome and metabolome and dereplicate using large public databases can significantly reduce rediscovery and save time and resources [33]. Furthermore, despite continuous development, many BGCs as for example observed for the citrilins or the biarylittides are still not automatically detected by bioinformatics tools like antiSMASH [24,91]. Especially for putative RiPPs detected in the secondary metabolome,

putative BGCs can be identified by mining the corresponding genome for ORFs encoding for the putative core sequences and so-called RiPP recognition elements (RREs) thus unlocking the whole panel of genetic engineering.

## 1.5 Increasing Yields and Modifying Natural Products

As mentioned earlier, natural products are often only produced in very small amounts, which conflicts with the high compound consumption of follow-up experiments. NMR, the gold standard for *de novo* structure elucidation of new natural products, has a significantly lower sensitivity than the preceding LC-MS experiments. Besides NMR, which at least is usually not a destructive method, higher compound amounts are also needed for bioactivity screenings, toxicity assessment or mode-of-action studies, all of which render the isolated compounds unusable for other experiments. The most obvious method for increasing yields is, of course, initially selecting the best producer strain available, as well as choosing the most advantageous cultivation conditions. However, many strains - especially myxobacteria - grow slow and might still not reach compound levels sufficient for all follow-up experiments. While heterologous expression enables production in faster growing microorganisms, it is a time-consuming process and does not guarantee higher yields than in the original wild type strain. In many cases, precursor supply represents a bottleneck in natural product biosynthesis like the supply of 2,3-dihydroxybenzoate in myxochelin biosynthesis [49,92]. Feeding artificial precursor molecules, a process called precursor-directed biosynthesis (PDB) can hereby dramatically increase yields as for example observed for myxochelin N1-N3 biosynthesis [47]. Furthermore, in case of promiscuous enzymes, a supply of artificial non-native precursors can generate additional products allowing for early structure-activity relationship (SAR) studies. By feeding deuterium,  $^{13}\text{C}$ - or  $^{15}\text{N}$ -labeled building blocks and subsequent MS-analysis, it is also possible to draw conclusions about the underlying biosynthesis at an early stage. The precursor molecules needed for such PDB experiments can oftentimes be anticipated via specificity of NRPS or PKS modules in the BGC or via tandem MS fragments. Mutasynthesis represents a further development of PDB, in which the intrinsic supply of the native precursor is blocked by inactivation of certain biosynthetic enzymes thus eliminating the production of the native natural product (Scheme 6). Subsequently, modified analogues of these building blocks are fed to the culture in the hope that they will be incorporated into the resulting unnatural derivatives.

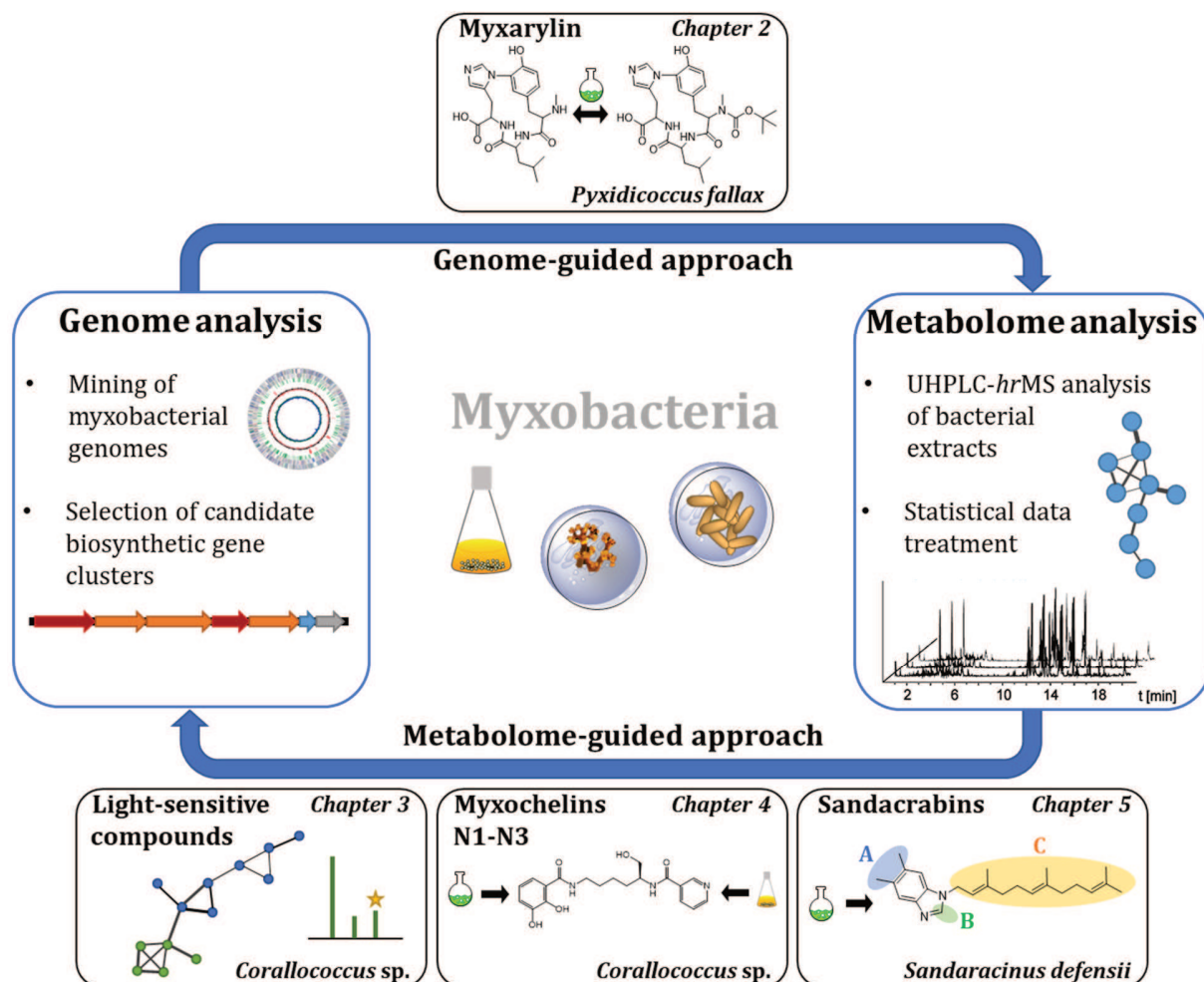


**Scheme 6.** Native biosynthetic pathway (A), precursor-directed biosynthesis (B) and mutasynthesis (C). A-D represent native biosynthesis educts, intermediates or products, B'-D' represent non-native intermediates or products, a-c represent biosynthetic enzymes. Adapted from [93].

Another highly beneficial approach is early total synthesis of natural products. While total synthesis on the one hand shows the promise of high yields as well as the possibility to make various modifications to the natural product of interest, it also loses all the negative factors of using biological systems such as long growing times, differences in between batches and yield limitations *e.g.* by self-toxicity. Furthermore, early available total synthesis can help to solve or prove elucidated structures, especially regarding the absolute stereochemistry of new natural products. Although this approach is for sure not applicable for all natural products because of their three dimensional complexity, there are examples like the sandacrabins, where early stage total synthesis proved to be highly beneficial by allowing many follow-up experiments that would not have been possible with the rather small amounts isolated from the native producer [23].

## 1.6 Outline of This Thesis

The work described in this thesis is focused on the discovery of novel natural products of myxobacterial origin. Thereby, different approaches are applied in order to approach myxobacterial secondary metabolism from both the genome and the metabolome side, ultimately leading to the discovery of new natural products originating from several different biosynthetic pathways (Figure 4).



**Figure 4.** Representation of the different parts of this thesis as part of metabolome- and genome-guided natural products discovery and development workflows.

The study on Myxarylin (Chapter 2) exemplifies a genome-guided natural products discovery approach. The amino acid sequence of the precursor and the DNA sequence of a gene encoding a modifying enzyme involved in the recently published biaryllytite biosynthesis [91] are used as a query to mine our in-house genome library for homologs and subsequent metabolome analysis is employed to facilitate compound identification. In contrast, the study on a family of light-sensitive compounds presented in chapter 3 describes a statistics-based metabolome analysis approach based on high-resolution mass spectrometry and analyses of tandem MS fragmentation data by spectral networking. A link between metabolome and genome data is subsequently achieved by conducting gene cluster inactivation experiments. Similarly, the myxochelins N1-N3 (Chapter 4) are discovered in a metabolome-guided approach using high-resolution MS and tandem MS data. Additionally, extensive database-assisted analysis of genome and metabolome data is employed to unveil the biosynthetic origin of these



compounds. Furthermore, also the sandacrabins presented in chapter 5 have previously been discovered by in-depth analysis of the metabolome of their myxobacterial producer strain [23].

Moreover, throughout the work presented in this thesis, small-scale organic synthesis is applied in several instances in order to facilitate compound isolation, confirm structures, elucidate the absolute stereochemistry and to additionally create novel non-natural congeners of myxobacterial natural products with the aim of improving their properties for human application. By integrating small-scale organic synthesis into contemporary genome and metabolome mining workflows, this work highlights the outstanding potential for the discovery and isolation of novel natural products of myxobacterial origin.

## 1.7 References

1. Bérdy, J. Bioactive Microbial Metabolites. *J. Antibiot.* **2005**, *58*, 1–26, doi:10.1038/ja.2005.1.
2. Firn, R.D.; Jones, C.G. The evolution of secondary metabolism - a unifying model. *Mol. Microbiol.* **2000**, *37*, 989–994, doi:10.1046/j.1365-2958.2000.02098.x.
3. Davies, J. Specialized microbial metabolites: functions and origins. *J. Antibiot.* **2013**, *66*, 361–364, doi:10.1038/ja.2013.61.
4. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803, doi:10.1021/acs.jnatprod.9b01285.
5. Dawid, W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol. Rev.* **2000**, *24*, 403–427, doi:10.1111/j.1574-6976.2000.tb00548.x.
6. Mohr, K.I.; Zindler, T.; Wink, J.; Wilharm, E.; Stadler, M. Myxobacteria in high moor and fen: An astonishing diversity in a neglected extreme habitat. *Microbiologyopen* **2017**, *6*, e00464, doi:10.1002/mbo3.464.
7. Reichenbach, H. The ecology of the myxobacteria. *Environ. Microbiol.* **1999**, *1*, 15–21, doi:10.1046/j.1462-2920.1999.00016.x.
8. Iizuka, T.; Jojima, Y.; Fudou, R.; Hiraishi, A.; Ahn, J.W.; Yamanaka, S. *Plesiocystis pacifica* gen. nov., sp. nov., a marine myxobacterium that contains dihydrogenated menaquinone, isolated from the pacific coasts of Japan. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 189–195, doi:10.1099/ijs.0.02418-0.
9. Albataineh, H.; Stevens, D.C. Marine Myxobacteria: A Few Good Halophiles. *Mar. Drugs* **2018**, *16*, doi:10.3390/md16060209.
10. Mohr, K.I. Diversity of Myxobacteria-We Only See the Tip of the Iceberg. *Microorganisms* **2018**, *6*, doi:10.3390/microorganisms6030084.

11. Sanford, R.A.; Cole, J.R.; Tiedje, J.M. Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halo-respiring facultative anaerobic myxobacterium. *Appl. Environ. Microbiol.* **2002**, *68*, 893–900, doi:10.1128/AEM.68.2.893–900.2002.
12. Gerth, K.; Müller, R. Moderately thermophilic myxobacteria: novel potential for the production of natural products isolation and characterization. *Environ. Microbiol.* **2005**, *7*, 874–880, doi:10.1111/j.1462-2920.2005.00761.x.
13. Landwehr, W.; Wolf, C.; Wink, J. Actinobacteria and Myxobacteria—Two of the Most Important Bacterial Resources for Novel Antibiotics. *Curr. Top. Microbiol. Immunol.* **2016**, doi:10.1007/82\_2016\_503.
14. Mohr, K.I.; Stechling, M.; Wink, J.; Wilharm, E.; Stadler, M. Comparison of myxobacterial diversity and evaluation of isolation success in two niches: Kiritimati Island and German compost. *Microbiologyopen* **2016**, *5*, 268–278, doi:10.1002/mbo3.325.
15. Garcia, R.; Müller, R. The Family Polyangiaceae. In *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; pp 247–279, ISBN 978-3-642-39044-9.
16. Muñoz-Dorado, J.; Marcos-Torres, F.J.; García-Bravo, E.; Moraleda-Muñoz, A.; Pérez, J. Myxobacteria: Moving, Killing, Feeding, and Surviving Together. *Front. Microbiol.* **2016**, *7*, doi:10.3389/fmicb.2016.00781.
17. Cao, P.; Dey, A.; Vassallo, C.N.; Wall, D. How Myxobacteria Cooperate. *J. Mol. Biol.* **2015**, doi:10.1016/j.jmb.2015.07.022.
18. Herrmann, J.; Fayad, A.A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2016**, doi:10.1039/c6np00106h.
19. Baumann, S.; Herrmann, J.; Raju, R.; Steinmetz, H.; Mohr, K.I.; Hüttel, S.; Harmrolfs, K.; Stadler, M.; Müller, R. Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **2014**, *53*, 14605–14609, doi:10.1002/anie.201409964.
20. Jansen, R.; Höfle, G.; Irschik, H.; Reichenbach, H. Antibiotika aus Gleitenden Bakterien, XXIV. Corallopyronin A, B und C – drei neue Antibiotika aus *Coralloccoccus coralloides* Cc c127 (Myxobacterales). *Liebigs Ann. Chem.* **1985**, *1985*, 822–836, doi:10.1002/jlac.198519850418.
21. Mulzer, J.; Altmann, K.H.; Höfle, G.; Müller, R.; Prantz, K. Epothilones - a fascinating family of microtubule stabilizing antitumor agents. *C. R. Chimie* **2008**, *11*, 1336–1368, doi:10.1016/j.crci.2008.02.005.

22. Sasse, F.; Steinmetz, H.; Heil, J.; Höfle, G.; Reichenbach, H. Tubulysins, new cytostatic peptides from myxobacteria acting on microtubuli. Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **2000**, *53*, 879–885, doi:10.7164/antibiotics.53.879.
23. Bader, C.D.; Panter, F.; Garcia, R.; Tchesnokov, E.P.; Haid, S.; Walt, C.; Spröer, C.; Kiefer, A.F.; Götte, M.; Overmann, J.; et al. Sandacrabins - Structurally Unique Antiviral RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry – A European Journal* **2022**, *28*, e202104484, doi:10.1002/chem.202104484.
24. Hug, J.J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. Biosynthesis of Cittilins, Unusual Ribosomally Synthesized and Post-translationally Modified Peptides from *Myxococcus xanthus*. *ACS Chem. Biol.* **2020**, *15*, 2221–2231, doi:10.1021/acscchembio.0c00430.
25. Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9*, 4898–4908, doi:10.1039/C8SC01325J.
26. Gerth, K.; Steinmetz, H.; Höfle, G.; Jansen, R. Chlorotonil A, a macrolide with a unique gem-dichloro-1,3-dione functionality from *Sorangium cellulosum*, So ce1525. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 600–602, doi:10.1002/anie.200703993.
27. Ringel, S.M.; Greenough, R.C.; Roemer, S.; Connor, D.; Gutt, A.L.; Blair, B.; Kanter, G.; von Strandtmann. Ambruticin (W7783), a new antifungal antibiotic. *J. Antibiot.* **1977**, *30*, 371–375, doi:10.7164/antibiotics.30.371.
28. Schneiker, S.; Perlova, O.; Kaiser, O.; Gerth, K.; Alici, A.; Altmeyer, M.O.; Bartels, D.; Bekel, T.; Beyer, S.; Bode, E.; et al. Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **2007**, *25*, 1281–1289, doi:10.1038/nbt1354.
29. Han, K.; Li, Z.-F.; Peng, R.; Zhu, L.-P.; Zhou, T.; Wang, L.-g.; Li, S.-g.; Zhang, X.-b.; Hu, W.; Wu, Z.-H.; et al. Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Scientific Reports* **2013**, *3*, 2101, doi:10.1038/srep02101.
30. Land, M.; Hauser, L.; Jun, S.-R.; Nookaew, I.; Leuze, M.R.; Ahn, T.-H.; Karpinets, T.; Lund, O.; Kora, G.; Wassenaar, T.; et al. Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* **2015**, *15*, 141–161, doi:10.1007/s10142-015-0433-4.
31. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35, doi:10.1093/nar/gkab335.

32. Skinnider, M.A.; Johnston, C.W.; Gunabalasingam, M.; Merwin, N.J.; Kieliszek, A.M.; MacLellan, R.J.; Li, H.; Ranieri, M.R.M.; Webster, A.L.H.; Cao, M.P.T.; et al. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Commun.* **2020**, *11*, 6058, doi:10.1038/s41467-020-19986-1.
33. Panter, F.; Bader, C.D.; Müller, R. Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Chem. Sci.* **2021**, 5994–6010, doi:10.1039/D0SC06919A.
34. Bader, C.D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **2020**, *39*, 107480, doi:10.1016/j.biotechadv.2019.107480.
35. Wenzel, S.C.; Müller, R. Myxobacteria—‘microbial factories’ for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, *5*, 567–574, doi:10.1039/b901287g.
36. Bader, C.D.; Haack, P.A.; Panter, F.; Krug, D.; Müller, R. Expanding the Scope of Detectable Microbial Natural Products by Complementary Analytical Methods and Cultivation Systems. *J. Nat. Prod.* **2021**, doi:10.1021/acs.jnatprod.0c00942.
37. Stuart Smith; Shiou-Chuan Tsai. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **2007**, *24*, 1041–1072, doi:10.1039/B603600G.
38. Hopwood, D.A. Genetic contributions to understanding polyketide synthases. *Chem Rev* **1997**, *97*, 2465–2497, doi:10.1021/cr960034i.
39. Bradley S. Moore; Christian Hertweck. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep.* **2002**, *19*, 70–99, doi:10.1039/B003939J.
40. Chan, Y.A.; Podevels, A.M.; Kevany, B.M.; Thomas, M.G. Biosynthesis of Polyketide Synthase Extender Units. *Nat. Prod. Rep.* **2009**, *26*, 90–114, doi:10.1039/b801658p.
41. Wenzel, S.C.; Williamson, R.M.; Grünanger, C.; Xu, J.; Gerth, K.; Martinez, R.A.; Moss, S.J.; Carroll, B.J.; Grond, S.; Unkefer, C.J.; et al. On the biosynthetic origin of methoxymalonyl-acyl carrier protein, the substrate for incorporation of “glycolate” units into ansamitocin and soraphen A. *J. Am. Chem. Soc.* **2006**, *128*, 14325–14336, doi:10.1021/ja064408t.
42. Dutta, S.; Whicher, J.R.; Hansen, D.A.; Hale, W.A.; Chemler, J.A.; Congdon, G.R.; Narayan, A.R.; Hakansson, K.; Sherman, D.H.; Smith, J.L.; et al. Structure of a modular polyketide synthase. *Nature* **2014**, *510*, 512–517, doi:10.1038/nature13423.
43. Miller, B.R.; Gulick, A.M. Structural Biology of Nonribosomal Peptide Synthetases. *Methods Mol. Biol.* **2016**, *1401*, 3–29, doi:10.1007/978-1-4939-3375-4\_1.

44. Sandmann, A.; Sasse, F.; Müller, R. Identification and analysis of the core biosynthetic machinery of tubulysin, a potent cytotoxin with potential anticancer activity. *Chem. Biol.* **2004**, *11*, 1071–1079, doi:10.1016/j.chembiol.2004.05.014.
45. Gorges, J.; Panter, F.; Kjaerulff, L.; Hoffmann, T.; Kazmaier, U.; Müller, R. Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxy-pipecolic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 14270–14275, doi:10.1002/anie.201808028.
46. Yan, F.; Auerbach, D.; Chai, Y.; Keller, L.; Tu, Q.; Hüttel, S.; Glemser, A.; Grab, H.A.; Bach, T.; Zhang, Y.; et al. Biosynthesis and Heterologous Production of Vioprolides: Rational Biosynthetic Engineering and Unprecedented 4-Methylazetidinecarboxylic Acid Formation. *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 8754–8759, doi:10.1002/anie.201802479.
47. Frank, N.A.; Széles, M.; Akone, S.H.; Rasheed, S.; Hüttel, S.; Frewert, S.; Hamed, M.M.; Herrmann, J.; Schuler, S.M.M.; Hirsch, A.K.H.; et al. Expanding the Myxochelin Natural Product Family by Nicotinic Acid Containing Congeners. *Molecules* **2021**, *26*, 4929, doi:10.3390/molecules26164929.
48. Korp, J.; König, S.; Schieferdecker, S.; Dahse, H.-M.; König, G.M.; Werz, O.; Nett, M. Harnessing Enzymatic Promiscuity in Myxochelin Biosynthesis for the Production of 5-Lipoxygenase Inhibitors. *ChemBioChem* **2015**, *16*, 2445–2450, doi:10.1002/cbic.201500446.
49. Gaitatzis, N.; Kunze, B.; Müller, R. Novel insights into siderophore formation in myxobacteria. *ChemBioChem* **2005**, *6*, 365–374, doi:10.1002/cbic.200400206.
50. 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. Engl.* **2013**, *52*, 7098–7124, doi:10.1002/anie.201208344.
51. Arnison, P.G.; Bibb, M.J.; Bierbaum, G.; Bowers, A.A.; Bugni, T.S.; Bulaj, G.; Camarero, J.A.; Campopiano, D.J.; Challis, G.L.; Clardy, J.; et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **2013**, *30*, 108–160, doi:10.1039/C2NP20085F.
52. Oman, T.J.; van der Donk, W. A. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat. Chem. Biol.* **2010**, *6*, 9–18, doi:10.1038/nCheMBio.286.
53. McIntosh, J.A.; Donia, M.S.; Schmidt, E.W. Ribosomal Peptide Natural Products: Bridging the Ribosomal and Nonribosomal Worlds. *Nat. Prod. Rep.* **2009**, *26*, 537–559, doi:10.1039/B714132G.

54. Reddy, G.K.; Leferink, N.G.H.; Umemura, M.; Ahmed, S.T.; Breitling, R.; Scrutton, N.S.; Takano, E. Exploring novel bacterial terpene synthases. *PLoS ONE* **2020**, *15*, e0232220, doi:10.1371/journal.pone.0232220.
55. Helfrich, E.J.N.; Lin, G.-M.; Voigt, C.A.; Clardy, J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem.* **2019**, *15*, 2889–2906, doi:10.3762/bjoc.15.283.
56. Kuzuyama, T.; Seto, H. Two distinct pathways for essential metabolic precursors for isoprenoid biosynthesis. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* **2012**, *88*, 41–52, doi:10.2183/pjab.88.41.
57. Hess, D. Terpenoids. In *Plant Physiology: Molecular, Biochemical, and Physiological Fundamentals of Metabolism and Development*; Hess, D., Ed.; Springer: S.I., 1975; pp 99–116, ISBN 978-3-540-06643-9.
58. CRC Press. Dictionary of Natural Products 30.2. Available online: <https://dnp.chemnetbase.com> (accessed on 3 January 2022).
59. Lyddiard, D.; Jones, G.L.; Greatrex, B.W. Keeping it simple: lessons from the golden era of antibiotic discovery. *FEMS Microbiol. Lett.* **2016**, *363*, fnw084, doi:10.1093/femsle/fnw084.
60. Ortholand, J.Y.; Ganesan, A. Natural products and combinatorial chemistry: back to the future. *Curr. Opin. Chem. Biol.* **2004**, *8*, 271–280, doi:10.1016/j.cbpa.2004.04.011.
61. Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **2018**, *9*, 803, doi:10.1038/s41467-018-03184-1.
62. Nichols, D.; Cahoon, N.; Trakhtenberg, E.M.; Pham, L.; Mehta, A.; Belanger, A.; Kanigan, T.; Lewis, K.; Epstein, S.S. Use of Ichip for High-Throughput *In Situ* Cultivation of “Uncultivable” Microbial Species. *Appl. Environ. Microbiol.* **2010**, *76*, 2445–2450, doi:10.1128/AEM.01754-09.
63. Hug, J.J.; Bader, C.D.; Remškar, M.; Cirnski, K.; Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **2018**, *7*, 44, doi:10.3390/antibiotics7020044.
64. Monciardini, P.; Iorio, M.; Maffioli, S.; Sosio, M.; Donadio, S. Discovering new bioactive molecules from microbial sources. *Microb. Biotechnol.* **2014**, *7*, 209–220, doi:10.1111/1751-7915.12123.
65. Harvey, A.L.; Edrada-Ebel, R.; Quinn, R.J. The re-emergence of natural products for drug discovery in the genomics era. *Nature Reviews Drug Discovery* **2015**, *14*, 111–129, doi:10.1038/nrd4510.
66. Doroghazi, J.R.; Albright, J.C.; Goering, A.W.; Ju, K.-S.; Haines, R.R.; Tchalukov, K.A.; Labeda, D.P.; Kelleher, N.L.; Metcalf, W.W. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **2014**, *10*, 963–968, doi:10.1038/nchembio.1659.
67. Sticher, O. Natural product isolation. *Nat. Prod. Rep.* **2008**, *25*, 517–554, doi:10.1039/b700306b.

68. Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **2014**, *86*, 10780–10788, doi:10.1021/ac502805w.
69. Krug, D.; Müller, R. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **2014**, *31*, 768–783, doi:10.1039/c3np70127a.
70. Chen, F.; Wang, M.; Cheng, K.-W. Liquid Chromatography-Mass Spectrometry in Natural Product Research. In *Bioactive natural products: Detection, isolation, and structural determination / edited by Steven M. Colegate, Russell J. Molyneux*, 2<sup>nd</sup> ed.; Colegate, S.M., Molyneux, R.J., Eds.; CRC Press: Boca Raton, 2008; pp 245–2651, ISBN 978-0-8493-7258-2.
71. Böhringer, N.; Patras, M.A.; Schäberle, T.F. Heterologous Expression of Pseudouridimycin and Description of the Corresponding Minimal Biosynthetic Gene Cluster. *Molecules* **2021**, *26*, doi:10.3390/molecules26020510.
72. Hug, J.J.; Frank, N.A.; Walt, C.; Šenica, P.; Panter, F.; Müller, R. Genome-Guided Discovery of the First Myxobacterial Biarylite Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis. *Molecules* **2021**, *26*, 7483, doi:10.3390/molecules26247483.
73. Gao, W.; Stalder, T.; Foley, P.; Rauh, M.; Deng, H.; Kirschbaum, C. Quantitative analysis of steroid hormones in human hair using a column-switching LC-APCI-MS/MS assay. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2013**, *928*, 1–8, doi:10.1016/j.jchromb.2013.03.008.
74. Ito, T.; Masubuchi, M. Dereplication of microbial extracts and related analytical technologies. *J. Antibiot* **2014**, *67*, 353–360, doi:10.1038/ja.2014.12.
75. Cortina, N.S.; Krug, D.; Plaza, A.; Revermann, O.; Müller, R. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem. Int. Ed.* **2012**, *51*, 811–816, doi:10.1002/anie.201106305.
76. Panter, F.; Krug, D.; Müller, R. Novel Methoxymethacrylate Natural Products Uncovered by Statistics-Based Mining of the *Myxococcus fulvus* Secondary Metabolome. *ACS Chem. Biol.* **2019**, *14*, 88–98, doi:10.1021/acscchembio.8b00948.
77. Kuhnert, N.; Jaiswal, R.; Eravuchira, P.; El-Abassy, R.M.; Kammer, B. von der; Materny, A. Scope and limitations of principal component analysis of high resolution LC-TOF-MS data: the analysis of the chlorogenic acid fraction in green coffee beans as a case study. *Anal. Methods* **2011**, *3*, 144–155, doi:10.1039/c0ay00512f.
78. Linington, R.G. npatlas - The Natural Products Atlas. Available online: <https://www.npatlas.org>.

- 
79. Zani, C.L.; Carroll, A.R. Database for Rapid Dereplication of Known Natural Products Using Data from MS and Fast NMR Experiments. *J. Nat. Prod.* **2017**, *80*, 1758–1766, doi:10.1021/acs.jnatprod.6b01093.
80. Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A.A.; Melnik, A.V.; Meusel, M.; Dorrestein, P.C.; Rousu, J.; Böcker, S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **2019**, *16*, 299–302, doi:10.1038/s41592-019-0344-8.
81. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapon, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837, doi:10.1038/nbt.3597.
82. Ziemert, N.; Alanjary, M.; Weber, T. The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **2016**, *33*, 988–1005, doi:10.1039/c6np00025h.
83. Kautsar, S.A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J.C.; Terlouw, B.R.; van der Hooft, Justin J J; van Santen, J.A.; Tracanna, V.; Suarez Duran, H.G.; Pascal Andreu, V.; et al. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* **2020**, *48*, D454-D458, doi:10.1093/nar/gkz882.
84. Medema, M.H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J.B.; Blin, K.; Bruijn, I. de; Chooi, Y.H.; Claesen, J.; Coates, R.C.; et al. Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **2015**, *11*, 625–631, doi:10.1038/nchembio.1890.
85. Phuong Nguyen Tran; Ming-Ren Yen; Chen-Yu Chiang; Hsiao-Ching Lin; Pao-Yang Chen. Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 3277–3287, doi:10.1007/s00253-019-09708-z.
86. Bode, H.B.; Müller, R. Secondary metabolism in myxobacteria. In *Myxobacteria: Multicellularity and differentiation*; Whitworth, D., Ed.; ASM Press: Chicago, 2007; pp 259–282, ISBN 978-1-55581-567-7.
87. Hopwood, D.A. Genetic Contributions to Understanding Polyketide Synthases. *Chem. Rev.* **1997**, *97*, 2465–2498, doi:10.1021/cr960034i.
88. Chevrette, M.G.; Aicheler, F.; Kohlbacher, O.; Currie, C.R.; Medema, M.H. SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. *Bioinformatics* **2017**, *33*, 3202–3210, doi:10.1093/bioinformatics/btx400.
89. de Jong, A.; van Hijum, S.A.F.T.; Bijlsma, J.J.E.; Kok, J.; Kuipers, O.P. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res* **2006**, *34*, W273-W279, doi:10.1093/nar/gkl237.



- 
90. Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-Ya, K.; Omura, S.; Cane, D.E.; Ikeda, H. Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 857–862, doi:10.1073/pnas.1422108112.
  91. Zdouc, M.M.; Alanjary, M.M.; Zarazúa, G.S.; Maffioli, S.I.; Crüsemann, M.; Medema, M.H.; Donadio, S.; Sosio, M. A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters. *Cell Chem. Biol.* **2020**, doi:10.1016/j.chembiol.2020.11.009.
  92. Silakowski, B.; Kunze, B.; Nordsiek, G.; Blöcker, H.; Höfle, G.; Müller, R. The myxochelin iron transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15. *Eur. J. Biochem.* **2000**, *267*, 6476–6485, doi:10.1046/j.1432-1327.2000.01740.x.
  93. Kirschning, A.; Taft, F.; Knobloch, T. Total synthesis approaches to natural product derivatives based on the combination of chemical synthesis and metabolic engineering. *Org. Biomol. Chem.* **2007**, *5*, 3245–3259, doi:10.1039/b709549j.

## Chapter 2

# Genome-guided Discovery of the First Myxobacterial Biarylittide Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis

Previously published in:

Joachim J. Hug<sup>†</sup>, Nicolas A. Frank<sup>†</sup>, Christine Walt, Petra Šenica, Fabian Panter,  
Rolf Müller\*

*Molecules*, 2021 December 10; **26**(24):7483

DOI: 10.3390/molecules26247483

### Affiliation

Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

---

# Contributions and Acknowledgments

## Author's effort:

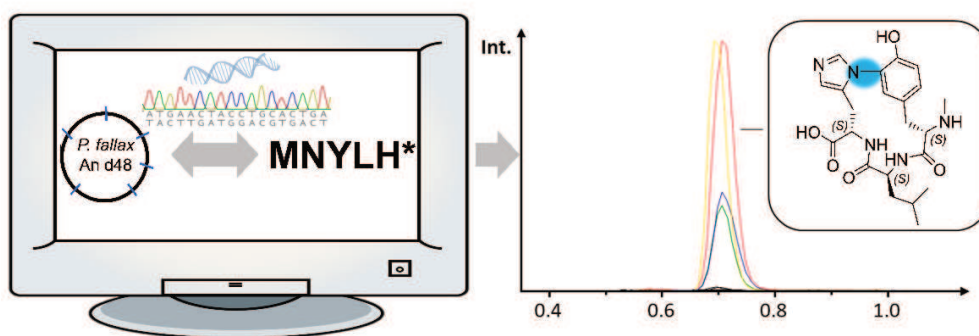
The author significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The laboratory and *in silico* work regarding cultivation, purification and elucidation of the absolute stereochemistry were performed by the author. Furthermore, the author contributed significantly to conceiving and writing this manuscript.

## Contributions by others:

Joachim Hug contributed to conception of the study, performed genome-guided compound discovery, heterologous expression of the BGC and proposal of the biosynthesis and contributed significantly to conceiving and writing the manuscript. Christine Walt contributed through NMR structure elucidation as well as conceiving and writing the manuscript. Petra Šenica contributed through heterologous expression of the BGC. Fabian Panter contributed through compound discovery, cultivation, purification and proposal of the biosynthesis as well as writing the manuscript. Rolf Müller contributed by supervision of the project and conceiving, editing and proofreading of the manuscript.

## 2.1 Abstract

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a structurally diverse group of natural products. They feature a wide range of intriguing post-translational modifications, as exemplified by the biarylittides. These are a family of cyclic tripeptides found in *Planomonospora*, carrying a biaryl linkage between two aromatic amino acids. Recent genomic analyses revealed that the minimal biosynthetic prerequisite of biarylittide biosynthesis consists of only one ribosomally synthesized pentapeptide precursor as the substrate and a modifying cytochrome-P450-dependent enzyme. In silico analyses revealed that minimal biarylittide RiPP clusters are widespread among natural product producers across phylogenetic borders, including myxobacteria. We report here the genome-guided discovery of the first myxobacterial biarylittide MeYLH, termed Myxarylin, from *Pyxidicoccus fallax* An d48. Myxarylin was found to be an *N*-methylated tripeptide that surprisingly exhibits a C–N biaryl crosslink. In contrast to Myxarylin, previously isolated biarylittides are *N*-acetylated tripeptides that feature a C–C biaryl crosslink. Furthermore, the formation of Myxarylin was confirmed by the heterologous expression of the identified biosynthetic genes in *Myxococcus xanthus* DK1622. These findings expand the structural and biosynthetic scope of biarylittide-type RiPPs and emphasize the distinct biochemistry found in the myxobacterial realm.



**Graphical Abstract.** Genome-guided screening of our in-house database revealed a genetic operon in *Pyxidicoccus fallax* An d48, which contains the gene *bytA*, encoding for the pentapeptide MNYLH (the asterisk represents a stop codon). Subsequent metabolome analysis led to the discovery of Myxarylin, the first myxobacterial biarylittide featuring a C–N biaryl crosslink.

## 2.2 Introduction

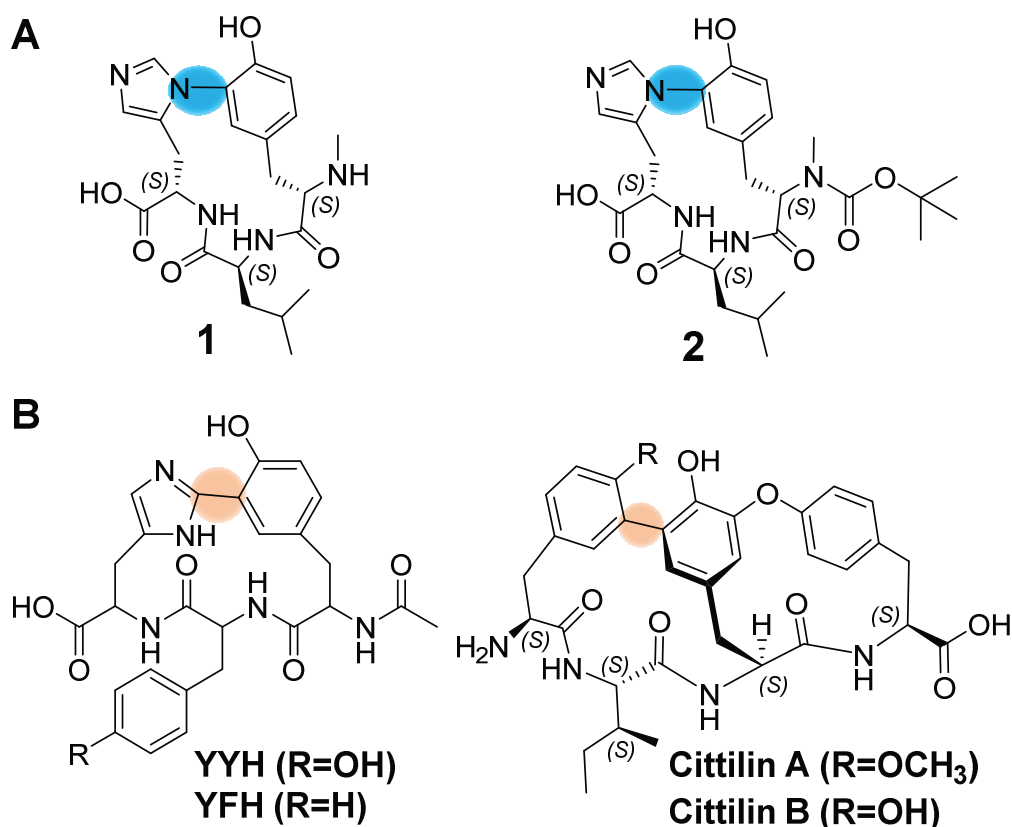
Myxobacteria are a phylum of Gram-negative bacteria that display a variety of unusual “behavioral” traits, such as coordinated swarming and the formation of macroscopic, multicellular fruiting bodies [1,2]. In addition to their unique “social behavior”, myxobacteria are also a viable source for a multitude of natural products, exhibiting diverse biological activities due to their biosynthetic gene cluster (BGC)-rich genomes [3]. The majority of myxobacterial secondary metabolites known to date derive from huge biosynthetic enzyme complexes, such as modular nonribosomal peptide synthetase (NRPS), polyketide synthase (PKS) and hybrids thereof, while natural products from other biosynthetic machineries have been isolated less frequently [4]. One of the reasons for this finding is that the discovery of natural products has shifted from a “grind and find” approach towards a more genome-guided discovery of microbial natural products [5]. This approach mostly relies on bioinformatics tools such as the “antibiotics and secondary metabolite analysis shell” (antiSMASH), allowing genome-wide identification and analysis of BGCs [6]. In contrast to NRPS and PKS gene clusters, which encode large enzyme complexes containing catalytic domains with high sequence homology, other types of BGCs, such as ribosomally synthesized and post-translationally modified peptide (RiPP) BGCs, are more difficult to identify and annotate as they are usually encoded by small, oftentimes poorly conserved open reading frames (ORFs) [7].

Although several bioinformatics tools have recently been developed [8–13] or successively updated [14] to allow for the automated detection of certain classes of RiPPs, there is likely still an abundance of yet-undiscovered RiPPs that may have been overlooked in the past [15]. Since the fulvocins [16,17], xanthacin [18], the crocagins [19,20] and the cittelins [21–23] are the only myxobacterial RiPPs that have been identified and partly characterized up to date, myxobacteria likely provide an underexploited reservoir for the discovery of new RiPPs. Among these few known myxobacterial RiPPs, the biosynthesis of the cittelins is remarkable, since it only requires a 27-amino-acid precursor peptide which is enzymatically modified by the cytochrome-P450-dependent enzyme CitB to form a bridged tetrapeptide containing a biaryl and an aryl-oxy-aryl link [21].

In a recent study, the biosynthetic genes of a natural product family closely related to the cittelins, termed biarylites, were initially found in *Planomonospora* strains, but further in silico analysis revealed that these minimal RiPP BGCs also appear in several other genera, including the myxobacterium *Pyxidicoccus* sp. CA032A. Similar to the bridging mechanism observed in cittelin biosynthesis, these biarylites also contain a biaryl linkage between two aromatic amino acids introduced by a cytochrome-P450-dependent enzyme [24]. This natural product class features an unprecedented RiPP biosynthesis, as it is produced from a precursor peptide encoded by an ORF termed *bytA* that encodes a five-amino-acid

substrate. Biosynthetically characterized members of the biaryllytide family, such as biaryllytide YYH and YFH, are exclusively *N*-acetylated tripeptides with a C–C biaryl crosslink between the aromatic side chains of amino acids one and three.

Consequently, the combined evaluation of genome and metabolome data covering myxobacteria led to the discovery and elucidation of the full structure, via spectroscopic techniques such as 2D NMR and high-resolution mass spectrometry, of the first myxobacterial biaryllytide MeYLH (**1**) and its semi-synthetic Boc derivative (**2**), which we named Myxarylin and Myxarylin-Boc, respectively (Figure 1). In contrast to previous members of the biaryllytide family, **1** features an *N*-methylated tripeptide with a C–N biaryl crosslink. Additionally, the formation of **1** was confirmed by the heterologous expression of the identified biosynthetic genes in *Myxococcus xanthus* DK1622.

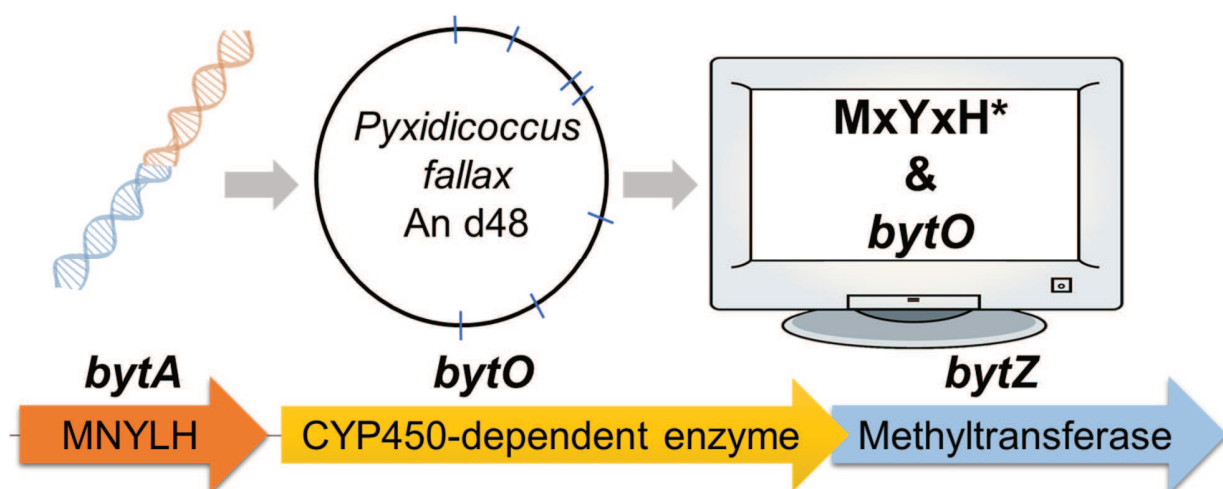


**Figure 1.** (A) Chemical structures of the myxobacterial biaryllytide MeYLH termed Myxarylin (**1**) and its semi-synthetic Boc-derivative Myxarylin-Boc (**2**). (B) Biosynthetically characterized members of the biaryllytide family such as biaryllytide YYH and YFH originating from *Planomonospora* spp., and cittilin A and B. Blue circles indicate C–N biaryl bonds and the beige-colored circle indicates the C–C biaryl bond.

## 2.3 Results and Discussion

### 2.3.1 Discovery and Purification of the Myxobacterial Biarylittide MeYLH

A previously conducted phylogenetic analysis of *bytO*-related genes revealed that the myxobacterium *Pyxidiccoccus* sp. CA032A harbors a genetic operon that putatively encodes a cytochrome-P450-dependent enzyme and a pentapeptide precursor [24], indicating the presence of an associated myxobacterial biarylittide. Similar to *Pyxidiccoccus* sp. CA032A, the myxobacterial strain *Pyxidiccoccus fallax* An d48 (formerly known as *Angiococcus disciformis* An d48)—a strain that has already been shown to be a prolific producer of bacterial secondary metabolites [25–27]—contains the biarylittide ORF *bytA* with the encoded peptide sequence MNYLH. In contrast to previously reported members of the biarylittide BGC family, the precursor peptide encoding gene *bytA* is not only clustered with a *bytO* gene homolog encoding a cytochrome-P450-dependent enzyme (91.5% amino acid identity to *Pyxidiccoccus* sp. CA032A homolog and 40.7% identical sites to *Planomonospora* sp. ID82291/ID107089, Figures S2 and S3) but also with an *S*-adenosyl-methionine-(SAM)-dependent methyltransferase gene that we named *bytZ* (Figure 2).

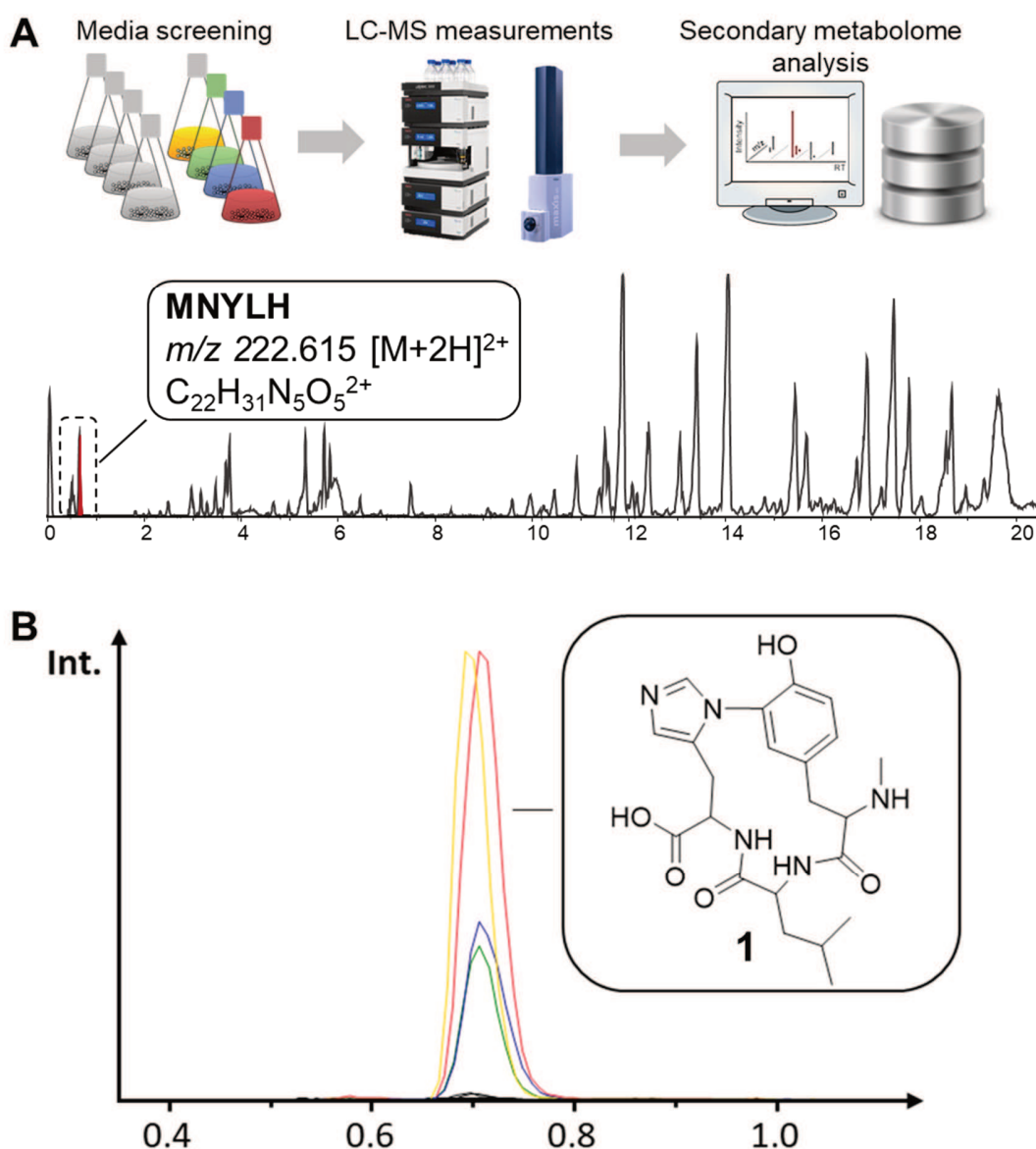


**Figure 2.** Genome-wide investigation in our in-house genome library using the amino acid sequence MxYxH\* (the asterisk represents a stop codon) and the DNA sequence of *bytO* homolog from *Pyxidiccoccus* sp. CA032A as query, revealed a genetic operon in *Pyxidiccoccus fallax* An d48, which comprises the gene *bytA* encoding the pentapeptide MNYLH, *bytO* encoding a cytochrome P450 (CYP450) dependent enzyme and *bytZ* which encodes an *S*-adenosyl-methionine-(SAM)-dependent methyltransferase.

Analysis of the secondary metabolome of the corresponding bacterium *P. fallax* An d48 revealed compound **1** as a peak in the liquid chromatography–mass spectrometry (LC-MS) chromatogram at a mass-to-charge ratio ( $m/z$ ) 222.615  $[M + 2H]^{2+}$ , supporting the deduced molecular formula  $C_{22}H_{31}N_5O_5^{2+}$  at a retention time of 0.7 min (Figure 3A). The polar nature of **1**, which is reflected by its early elution during LC-MS analysis, is consistent with the characteristics of previously reported biarylittides [24]. In addition, the deduced molecular formula of **1** ( $C_{22}H_{31}N_5O_5^{2+}$ ) matches the mass of a methylated cyclic tripeptide with the sequence YLH. In addition to that, the identified molecule shows a neutral loss fragment, which was clearly assigned in the tandem MS spectrum of **1** as leucine/isoleucine (Figure S7). These findings strongly supported our assumption that the discovery of **1** can be assigned as the first myxobacterial biarylittide MeYLH.

Interestingly, the production of **1** is dependent on the myxobacterial culture medium (Figure 3B). As shown in Figure 3B, the production of **1** is only observed when *P. fallax* An d48 is cultivated in a medium containing microorganisms that have been inactivated by autoclaving. Apart from that, the type of inactivated microorganism in the fermentation medium (in our case *E. coli* or *S. cerevisiae*) seemed to only play a minor role in the production levels. The finding that biotic or abiotic elicitors can enhance the production of natural products has been described before for numerous different microorganisms [28–30]; myxobacteria in particular are well-known for their complex connection between secondary metabolite production and cellular development stages [2,31]. Nevertheless, the molecular regulation mechanism responsible for biarylittide production in *P. fallax* An d48 remains elusive.

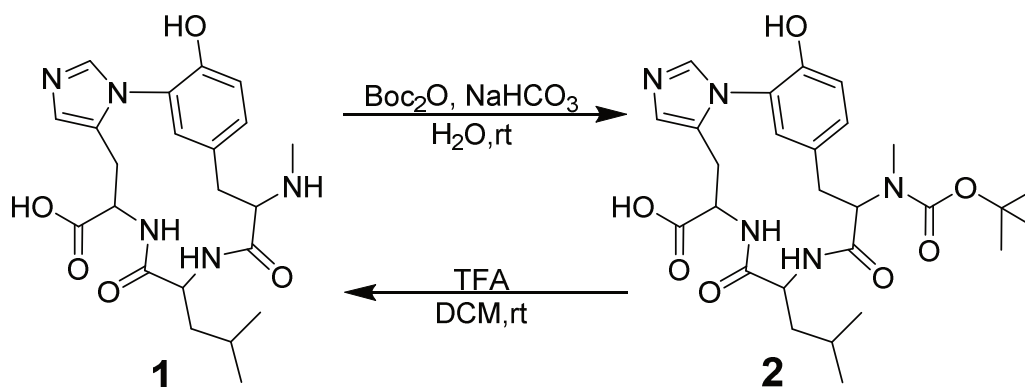




**Figure 3.** (A) Secondary metabolome analyses via HPLC-MS of *P. fallax* An d48 revealed the target natural product at  $m/z$  222.615  $[M + 2H]^{2+}$ , supporting the deduced molecular formula  $C_{22}H_{31}N_5O_5^{2+}$  at a retention time of 0.7 min. (B) Production of **1** in different media as observed as an LC-MS EIC at  $222.615 \pm 0.05$  Da  $[M + 2H]^{2+}$ . The black traces represent media without significant production of **1**, namely AMB, YM, CFL, CTT, M and P medium. The blue trace represents P medium containing autoclaved *Escherichia coli* (Pec), the yellow trace represents P medium containing autoclaved *Saccharomyces cerevisiae* (Py), the green trace represents low concentration starch based medium containing autoclaved *S. cerevisiae* (VY/2) and the red trace represents high concentration starch based medium containing autoclaved *S. cerevisiae* (VY).

To isolate secondary metabolite **1**, *P. fallax* An d48 was cultivated in a VY medium in larger-scale batch fermentation. Since the poor solubility of **1** in organic solvents and water hindered efficient compound purification, a *tert*-butoxycarbonyl (Boc) protecting group was introduced at the *N*-terminal secondary amine moiety by stirring the culture supernatant with di-*tert*-butyl dicarbonate (Boc<sub>2</sub>O) at an

ambient temperature to yield the semi-synthetic compound **2** (Scheme 1). The purification of **2** was achieved by liquid/liquid partitioning of the reaction mixture followed by semi-preparative HPLC with MS detection. The deprotection of **2** to regain the original secondary metabolite, **1**, was achieved by stirring the pure compound suspended in dichloromethane with 20% (v/v) trifluoroacetic acid (TFA) at room temperature until full conversion was observed.



**Scheme 1.** Protection of **1** with a tert-butyloxycarbonyl (Boc) protecting group and deprotection of **2** under acidic conditions.

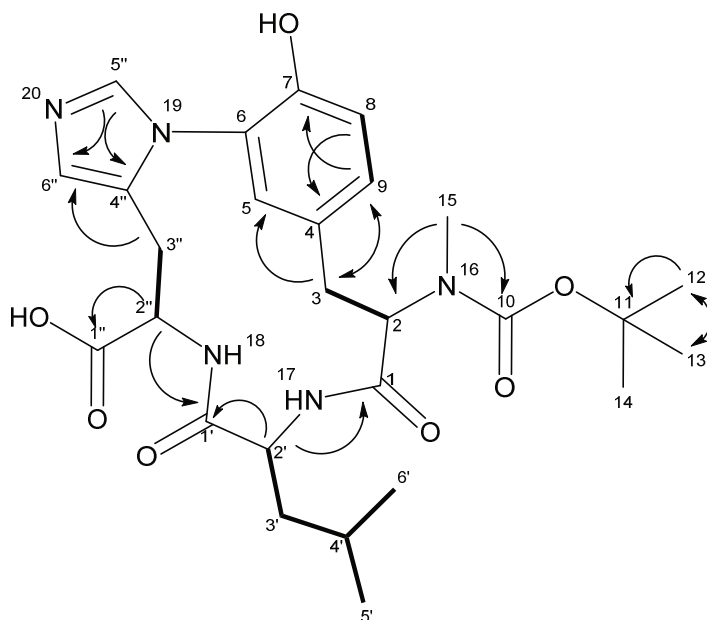
### 2.3.2 Structure Elucidation of **1** and **2**

HRESIMS indicates an  $[M + H]^+$  monoisotopic mass peak for **2** at  $m/z$  544.2765, consistent with the molecular formula  $C_{27}H_{38}N_5O_7$  ( $m/z$  calculated for  $[M + H]^+$  544.2766) with 12 double-bond equivalents (DBEs).  $^1\text{H-NMR}$  and HSQC spectra in methanol- $d_4$  reveal five aromatic double-bond signals at  $\delta(^1\text{H}) = 7.94$  (1H, s)  $\delta(^{13}\text{C}) = 138.5$ ,  $\delta(^1\text{H}) = 7.03$  (1H, d,  $J = 8.13$  Hz)  $\delta(^{13}\text{C}) = 130.2$ ,  $\delta(^1\text{H}) = 6.91$  (1H, s)  $\delta(^{13}\text{C}) = 118.9$ ,  $\delta(^1\text{H}) = 6.89$  (1H, brs)  $\delta(^{13}\text{C}) = 118.1$  and  $\delta(^1\text{H}) = 6.84$  (1H, m)  $\delta(^{13}\text{C}) = 126.1$  ppm. Three  $\alpha$ -proton signals are located at  $\delta(^1\text{H}) = 4.86$  (1H, s)  $\delta(^{13}\text{C}) = 52.8$ ,  $\delta(^1\text{H}) = 4.66$  (1H, m)  $\delta(^{13}\text{C}) = 59.7$  and  $\delta(^1\text{H}) = 4.61$  ppm (1H, dd,  $J = 12.62, 2.78$  Hz)  $\delta(^{13}\text{C}) = 55.2$ . Furthermore, two diastereotopic methylene signals could be observed at  $\delta(^1\text{H}) = 3.55, 2.56$  (2H, dd,  $J = 15.19, 11.77$  Hz)  $\delta(^{13}\text{C}) = 34.2$  as well as at  $\delta(^1\text{H}) = 3.32, 2.82$  (2H, dd, 16.47, 12.62 Hz)  $\delta(^{13}\text{C}) = 33.0$  with another methylene signal at  $\delta(^1\text{H}) = 1.58$  (2H, m)  $\delta(^{13}\text{C}) = 43.7$  ppm. Lastly, one methine signal appears at  $\delta(^1\text{H}) = 1.60$  (1H, m)  $\delta(^{13}\text{C}) = 26.1$  and three methyl signals at  $\delta(^1\text{H}) = 2.90$  (3H, s)  $\delta(^{13}\text{C}) = 30.7$ ,  $\delta(^1\text{H}) = 1.48$  (9H, s)  $\delta(^{13}\text{C}) = 28.8$  and  $\delta(^1\text{H}) = 0.94$  (6H, m)  $\delta(^{13}\text{C}) = 23.8$  ppm. Comprehensive analysis of COSY and HMBC correlations confirmed the presence of a cyclic tripeptide consisting of leucine, histidine and a Boc-protected, *N*-methylated tyrosine with a characteristic C–N-biaryl bond between histidine and tyrosine (Figure 4). The signal at  $\delta(^{13}\text{C}) = 177.8$  ppm exhibits a downfield shift characteristic for free carboxyl functions and shows HMBC correlations to the methine signal at  $\delta(^1\text{H}) = 4.61$  ppm as well as the diastereotopic methylene signal at  $\delta(^1\text{H}) = 3.32, 2.82$  ppm of the histidine part. Moreover, HMBC

correlation of the methyl signal at  $\delta(^1\text{H}) = 2.90$  ppm to the  $\alpha$ -proton signal at  $\delta(^1\text{H}) = 4.66$  ppm proves the *N*-methylation of the tyrosine part. Additionally, Boc protection of the same nitrogen could be underpinned by the HMBC correlation of the mentioned methyl signal at  $\delta(^1\text{H}) = 2.90$  ppm to the carbonyl signal at  $\delta(^{13}\text{C}) = 157.6$  ppm possessing a characteristic shift for carbamate function. The corresponding three methyl groups of the Boc moiety can be observed at  $\delta(^1\text{H}) = 1.48$  ppm exhibiting HMBC correlations to the quaternary carbon at  $\delta(^{13}\text{C}) = 81.9$  ppm. For additional evidence an  $^1\text{H}$ - $^{15}\text{N}$  HMBC experiment was carried out using standard parameters and uncovered a characteristic low field  $^{15}\text{N}$  NMR chemical shift for N20 of the imidazole moiety with  $\delta(^{15}\text{N}) = 247.3$  ppm showing correlations to the respective aromatic double-bond signals at  $\delta(^1\text{H}) = 7.94$  and  $6.91$  ppm as well as to the diastereotopic methylene signals at  $\delta(^1\text{H}) = 3.32, 2.82$  ppm. Furthermore,  $\delta(^{15}\text{N}) = 174.6$  ppm is the chemical shift for N19, which also evince correlations to the respective two aromatic double-bond signals. The peptidic NH18 of the histidine exhibits a signal at  $\delta(^{15}\text{N}) = 123.1$  ppm with correlations to the diastereotopic methylene signals at  $\delta(^1\text{H}) = 3.32, 2.82$  ppm. Unfortunately, no  $^1\text{H}$ - $^{15}\text{N}$  HMBC signal could be observed for peptidic NH17, but the presence of leucine was underpinned by Marfey's derivatization experiments. Finally, the methylcarbamate N16, with a characteristic high field chemical shift at  $\delta(^{15}\text{N}) = 84.7$  ppm, possesses a correlation to the methyl signal at  $\delta(^1\text{H}) = 2.90$  ppm, which is additional proof for the position of methylation. Selective 1D NOESY experiments were performed to propose the biaryl bond to be located between C6 and N19 or N20 using standard parameters and a mixing time of  $350 \mu\text{s}$ . The first experiment was adapted for the selective excitation of C5'' at  $\delta(^1\text{H}) = 7.94$  ppm with a distance of  $17.20$  Hz resulting in no additional signals in the  $^1\text{H}$  spectrum, which could indicate the spatial distance to the tyrosine ring. The second measurement was optimized for the excitation of C6'' at  $\delta(^1\text{H}) = 6.91$  ppm with a distance of  $20.32$  Hz. Hereby, an additional  $^1\text{H}$  signal at  $\delta(^1\text{H}) = 7.03$  ppm (C9) appeared, which is likely to be the correlation between C9 and C5 at  $\delta(^1\text{H}) = 6.84$  ppm that was also excited under the given conditions. Since no correlation between C5'' and C5 could be found, the biaryl bond is most likely located between C6 and N19.

NMR experiments for **1** — after Boc deprotection — were carried out in  $\text{DMSO-}d_6$  due to previously mentioned solubility issues. Therefore, the majority of observed signals are more upfield shifted. More prominent shift changes could be observed for the two proton signals of the histidine at  $\delta(^1\text{H}) = 9.35$  (1H, s)  $\delta(^{13}\text{C}) = 135.6$  and  $\delta(^1\text{H}) = 7.16$  (1H, s)  $\delta(^{13}\text{C}) = 120.4$  ppm. Due to the presence of both histidine protons, the characteristic C–N-biaryl bond between histidine and tyrosine is again ensured. The same is true for the *N*-methylation of the tyrosine, with a methyl signal at  $\delta(^1\text{H}) = 2.50$  (3H, m)  $\delta(^{13}\text{C}) = 31.3$ . Furthermore, COSY and HMBC spectra exhibit correlations for two amide protons at  $\delta(^1\text{H}) = 8.86$  (1H, d,  $J = 9.52$  Hz) and  $\delta(^1\text{H}) = 9.21$  (1H, d,  $J = 8.56$  Hz) ppm as well as for the N-H proton of the *N*-methylated amine at  $\delta(^1\text{H}) =$

8.79 (1H, brs) ppm. The latter seems to isomerize with a N-H signal at  $\delta(^1\text{H}) = 9.31$  (1H, brs), which could be due to the deprotection reaction at this amine.



**Figure 4.** Atom numbering and selected COSY (bold) and HMBC (plain) correlations of **2**.

The assignment of the absolute configuration was based on Marfey's derivatization method. [32] Acidic hydrolysis and the derivatization of the hydrolysis product of **1** with D-respective L-(1-fluoro-2,4-dinitrophenyl-5-leucine amide) (FDLA) revealed C2, C2' and C2'' to be S and the tyrosine, leucine and histidine to thus be L-configured as the respective products have the same retention time as the derivatized L-amino acid standards (Figures S24–S29). These results are in line with expectations, since **1** most likely originates from a RiPP biosynthesis pathway. In order to validate the ribosomal origin of **1** (and the prediction of the stereochemical configurations in **1**) we attempted to perform heterologous expression of the biosynthetic genes *bytA*, *bytO* and *bytZ* in the established myxobacterial host, *M. xanthus* DK1622 [31].

Compounds **1** and **2** showed no antimicrobial activity in the performed biological assays (Table 1). The outcome of the bioactivity profiling of **1** and **2** reflects previous findings that these bridged tri/tetrapeptide RiPPs such as the cittilins and biarylites are not active as antimicrobial compounds [21,24]. Thus, the biological function of these compounds remains unknown.

**Table 1.** Minimum inhibitory concentration (MIC) values ( $\mu\text{g/mL}$ ) of Myxarylin and Myxarylin-Boc (**1** and **2**) against common microbial pathogens.

Microorganism	MIC of 1	MIC of 2
<i>Acinetobacter baumannii</i> DSM 30008	>64	>64
<i>Mucor hiemalis</i> DSM 2656	>64	>64
<i>Cryptococcus neoformans</i> DSM 11959	>64	>64
<i>Staphylococcus aureus</i> Newman	>64	>64
<i>Pseudomonas aeruginosa</i> PA14 (DSM 19882)	>64	>64
<i>Escherichia coli</i> <i>acrB</i> JW0451-2	>64	>64
<i>E. coli</i> wild type BW25113 (DSM 27469)	>64	>64
<i>Bacillus subtilis</i> DSM 10	>64	>64
<i>Candida albicans</i> DSM 1665	>64	>64
<i>Pichia anomala</i> DSM 6766	>64	>64
<i>Citrobacter freundii</i> DSM 30039	>64	>64
<i>Mycobacterium smegmatis</i> MC <sup>2</sup> 155	>64	>64

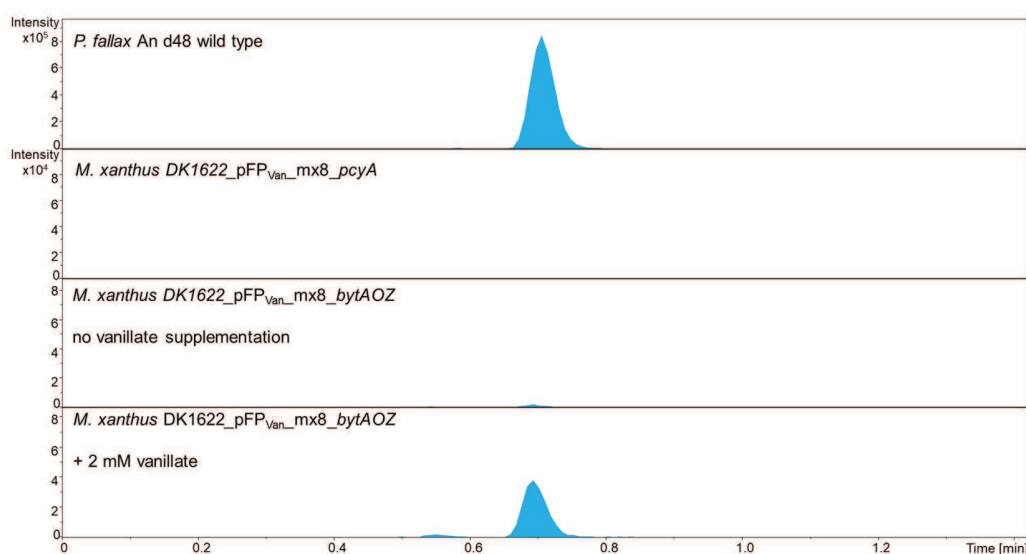
### 2.3.3 Heterologous Expression and Biosynthesis

The small genetic operon in *P. fallax* An d48 consisting of *bytA*, *bytO* and *bytZ* was amplified by PCR and subcloned into the expression vector pFP<sub>van\_mx8</sub>, a derivative based on the plasmid pFP<sub>van\_pcyA</sub> [26] (Tables S9–S13). The resulting expression plasmid, pFP<sub>van\_mx8\_bytAOZ</sub>, features a kanamycin resistance gene for clonal selection in the heterologous host *M. xanthus* DK1622, a vanillate-inducible promoter system to control the expression of *bytAOZ* and the Mx8 integrase gene, which enables the integration of the expression plasmid pFP<sub>van\_mx8\_bytAOZ</sub> into the attB1 or attB2 sites of *M. xanthus* DK1622 [31].

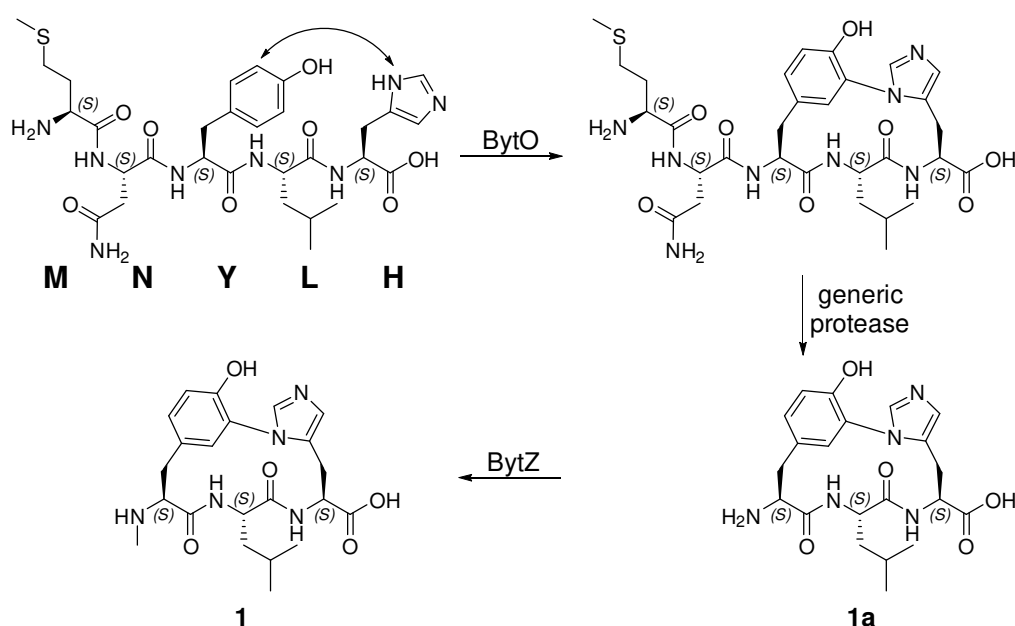
The host strain *M. xanthus* DK1622 was transformed with the generated expression construct via electroporation, and genetically verified transformants were investigated via LC-MS for the production of **1**. Subsequent heterologous expression of the myxobacterial biarylittide MeYLH BGC did yield **1** at low concentrations according to LC-MS analysis after vanillate supplementation of the transformants' fermentation cultures (Figure 5). The respective mutant strain showed, without vanillate supplementation, only trace amounts of **1**, presumably due to the basal expression of *bytAOZ*; previously developed biotechnological production platforms using the vanillate-inducible promoter system displayed even higher production titers of the associated natural product under un-induced cultivation conditions [33]. One possible explanation for the rather inefficient heterologous production of **1** might be that in the heterologous production setup the catalytic enzyme responsible for the proteolytic cleavage of the *N*-terminal residue is missing. For that reason, the *N*-terminal cleavage might be catalyzed by a different protease in *M. xanthus* DK1622 that shows lower catalytic activity.

Combining the results from the heterologous expression with the current knowledge of biarylittide biosynthesis, we conclude that the biosynthesis of **1** starts with the ribosomally synthesized pentapeptide

MNYLH, which is processed by the cytochrome-P450-dependent enzyme BytO to install the C–N biaryl linkage between the histidine and tyrosine. This modified pentapeptide is cleaved by the catalytic action of a generic protease to yield the non-methylated biosynthetic intermediate **1a**. This generic protease could also derive from another (RiPP) BGC, as previously described for the biosynthesis of argicyclamides [34]. The final tailoring step is probably catalyzed by BytZ to methylate the *N*-terminal free amine (Scheme 2). In conclusion, the successful heterologous production of **1** in *M. xanthus* DK1622 does not only prove the connection of the initially identified biarylptide BGC to its secondary metabolite, but also strongly supports the prediction of the absolute configuration of **1** as (2*S*, 2'*S*, 2''*S*).



**Figure 5.** Heterologous production of the myxobacterial biarylptide MeYLH (**1**) in *M. xanthus* DK1622. HPLC–MS EIC of crude extracts from *P. fallax* An d48 wild type as positive control and *M. xanthus* DK1622\_pFP<sub>van\_mx8</sub> mutant strain as negative control. While the HPLC–MS EIC of the crude extract from *M. xanthus* DK1622\_pFP<sub>van\_bytAOZ</sub> mutant without vanillate supplementation showed only trace amounts of **1** due to leaky expression, the HPLC–MS EIC of the crude extract from *M. xanthus* DK1622\_pFP<sub>van\_bytAOZ</sub> without 2 mM vanillate supplementation indicates low but significant production of **1**. EIC: Extracted ion chromatogram; blue, 222.615±0.05 Da [*M* + 2*H*]<sup>2+</sup>.



**Scheme 2.** Proposed biosynthesis of **1** from the ribosomally synthesized precursor peptide MNYLH.

## 2.4 Conclusion

In this study, we describe the discovery, isolation, full-structure elucidation and heterologous production of the first myxobacterial biarylittide with the sequence MeYLH called Myxarylin, which displays biaryl crosslinking distinct from previously described members of the biarylittide family. The different crosslinking and methylation decoration in the course of myxobacterial biarylittide biosynthesis raise the question of the biological function of this intriguing RiPP natural product family, since its biosynthetic gene clusters seem to be common among prokaryotes and have been identified across phylogenetic borders [24]. Although the ecological purpose of the biarylittides remains elusive to this date, the widespread conservation of biarylittide BGCs within numerous genera indicates an important role in producing microorganisms, making them potentially more multifaceted than simply being antimicrobial molecules.

The two distinctive features of Myxarylin are the *N*-methylation of the *N*-terminus and the C–N biaryl crosslinking. The methylation decoration and the underlying genetic operon consisting of genes that encode a precursor peptide, a cytochrome-P450-dependent enzyme and a methyltransferase strongly parallel the BGC of the cittelins. It appears that myxobacteria feature for both the cittelins and the biarylittides an obligatory methyltransferase, whereas all genome sequence data available for other microorganisms strictly lack a gene encoding a methyltransferase. Nevertheless, the previously isolated and biosynthetically characterized biarylittides undergo *N*-terminal acetylation by a yet-unknown acetyltransferase, a reaction that is not observed in the biosynthesis of **1**. The C–N biaryl crosslink found

in **1** points towards an essential question: whether the BytO enzyme found in *P. fallax* An d48 is capable of forming C–N biaryl crosslinks with other pentapeptide substrates, especially those found in *Planomonospora*. In reverse, the heterologous expression platform based on the BytO enzyme found in *Planomonospora* could be analyzed for the formation of **1** after the genetic alteration of the myxobacterial motif sequence MNYLH found in this study. These experiments would exclude the possibility that the amino acid sequence of BytA, rather than the architecture of the oxidative cyclization enzyme BytO, is responsible for the different biaryl crosslinkings.

In summary, the findings from this study set the foundation for further biochemical investigation of this small but diverse RiPP pathway as well as the underlying mechanism of biaryl crosslinking. The relatively low amino acid sequence identity of ~40% between the BytO homologs originating from *Pyxidicoccus* spp. and *Planomonospora* spp. implies inherent differences in the biochemistry of the C–N and C–C linkage observed in **1** and the other biarylites. The herein presented C–N biaryl crosslink of **1** might help to pinpoint the required molecular differences to form natural products with distinct biaryl crosslinks. Understanding biarylite catalysis might result in a versatile biotechnological tool to form different biaryl tripeptides using a minimalistic biosynthesis platform.

## 2.5 Materials and Methods

### 2.5.1 Applied Software, DNA Sequence Analysis, and Bioinformatic Methods

The genome of the terrestrial myxobacterial strain *P. fallax* An d48 (producer strain of **1**) was screened for the presence of *bytO* and *bytA* homologs with the software Geneious Prime<sup>®</sup> (Biomatters Ltd., Auckland, New Zealand, 2020.0.5) [35]. In order to find homologous genes or proteins, either the nucleotide or amino acid sequence of interest was aligned with the basic local alignment search tool (BLAST) against an in-house genome database or a publicly available nucleotide database. Raw data from alignments for the in silico evaluation of biarylite BGCs are stored on the in-house server. The functional prediction of ORFs was performed by using protein blast and/or blastx programs as well as Pfam [36]. The nucleotide sequence of the BGC of **1** (originating from *P. fallax* An d48) has been deposited in GenBank under the accession number OL539738.

### 2.5.2 Myxobacterial Fermentation and Extraction Procedure for LC-MS Analysis

Cultures for UHPLC-*hr*MS analysis are grown in 300 mL shake flasks containing 50 mL of fermentation medium inoculated with 1 mL of preculture. After inoculation the medium is supplemented with 2% of



sterile XAD-16 adsorber resin (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) suspension in water to bind secondary metabolites in the culture medium. Small-scale cultures were grown for 10–12 days. After fermentation the culture is pelleted in a 50 mL falcon at 6000 rcf for 10 min using an Eppendorf falcon table centrifuge. The supernatant is filtered through glass wool and dried using a rotary evaporator at 20 mbar and 40 °C water bath temperature. The dry supernatant is taken up in 1 mL water and centrifuged for 5 min at 15,000 rpm in a Hitachi table centrifuge to remove residual cells and cell debris. The supernatant is transferred into a novel 2 mL Eppendorf vial and mixed with 1 mL of cold methanol. The mixture is centrifuged again and the supernatant is transferred into an HPLC vial for UHPLC-*hr*MS analysis. The pellet containing cells and XAD-16 is transferred into a 100 mL Erlenmeyer flask and a magnetic stirrer is added. Fifty milliliters of acetone are added onto the pellet and the mixture is stirred for 60 min on a magnetic stirrer. The acetone extract is left to settle in order to sediment cell debris and XAD-16 resin for a second extraction step. The extract is filtered with a 125 micron folded filter, keeping cells and XAD-16 resin in the Erlenmeyer flask for a second extraction step. The residual pellet and XAD-16 resin are extracted again with 30 mL of distilled acetone for 60 min on a magnetic stirrer and filtered through the same folded filter. The combined extracts are transferred into a 100 mL round-bottom flask. The acetone is evaporated using a rotary evaporator at 260 mbar and 40 °C water bath temperature. The residual water is evaporated at 20 mbar until the residue in the flask is completely dry. The residue is taken up in 550  $\mu$ L of methanol and transferred into a 1.5 mL Eppendorf tube. This tube is centrifuged with a table centrifuge (Hitachi Koki Co., Tokyo, Japan) at 15,000 rpm for 2 min to remove residual insolubilities such as salts, cell debris and XAD-16 fragments. The residual extract is diluted 1:10 for UHPLC-*hr*MS analysis.

### 2.5.3 Standardized UHPLC-MS conditions

UHPLC-*hr*MS analysis is performed on a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Bruker maXis 4G ultra-high-resolution quadrupole time-of-flight (UHR-qTOF) MS equipped with a high-resolution electrospray ionization (HRESI) source (Bruker Daltonics, Billerica, MA, USA). The separation of a 1  $\mu$ L sample is achieved with a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH<sub>2</sub>O with 0.1% formic acid on an ACQUITY BEH C18 column (100  $\times$  2.1 mm, 1.7  $\mu$ m d<sub>p</sub>) (Waters, Eschborn, Germany) equipped with a Waters VanGuard BEH C18 1.7  $\mu$ m guard column at a flow rate of 0.6 mL/min and 45 °C for 18 min with detection by a diode array detector at 200–600 nm. The LC flow is split into 75  $\mu$ L/min before entering the mass spectrometer. Mass spectrograms are acquired in centroid mode ranging from 150–2500 *m/z* at an acquisition rate of 2 Hz in positive MS mode. Source parameters are set to 500 V end-

plate offset; 4000 V capillary voltage; 1 bar nebulizer gas pressure; 5 L/min dry gas flow; and 200 °C dry gas temperature. Ion transfer and quadrupole parameters are set to 350 V<sub>pp</sub> funnel RF; 400 V<sub>pp</sub> multipole RF; 5 eV ion energy; and 120 *m/z* low-mass cut-off. Collision cell is set to 5.0 eV and pre-pulse storage is set to 5 μs. Calibration is conducted automatically before every HPLC-MS run by the injection of sodium formate and calibration on the respective clusters formed in the ESI source. All MS analyses are acquired in the presence of the lock masses C<sub>12</sub>H<sub>19</sub>F<sub>12</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, C<sub>18</sub>H<sub>19</sub>F<sub>24</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub> and C<sub>24</sub>H<sub>19</sub>F<sub>36</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, which generate the [M + H]<sup>+</sup> ions of 622.0289, 922.0098 and 1221.9906. The HPLC-MS system is operated by HyStar 5.1 (Bruker Daltonics, Billerica, MA, USA), and LC chromatograms as well as UV spectra and mass spectrograms are analyzed with DataAnalysis 4.4 (Bruker Daltonics, Billerica, MA, USA). LC and MS conditions for scheduled precursor list (SPL)-guided tandem MS data acquisitions are kept constant according to standard UHPLC-MS conditions. Tandem MS data acquisition parameters are set to exclusively fragment SPL entries within a retention time tolerance of 0.2 min and a mass tolerance of 0.05 *m/z* for precursor ion selection. The method picks up to two precursors per cycle, applies smart exclusion after five spectra and performs CID and MS/MS spectra acquisition time ramping. CID energy is ramped from 35 eV for 500 *m/z* to 45 eV for 1000 *m/z* and 60 eV for 2000 *m/z*. MS full-scan acquisition rate is set to 2 Hz and MS/MS spectra acquisition rates are ramped from 1 to 4 Hz for precursor ion intensities of 10 kcts to 100 kcts.

#### 2.5.4 Isolation of 1 by Supernatant Derivatization and Semipreparative HPLC

For compound isolation, *P. fallax* An d48 was fermented in 50 mL of a VY/2 medium as a seed culture in 300 mL shake flasks on an Orbitron shaker at 180 rpm and 30 °C. The opaque culture medium becomes translucent and the supernatant turns green after 7 to 14 days of fermentation. This pre-culture is used to inoculate 6 × 2 L VY medium supplemented with 2% XAD-16 resin suspension in sterilized water in 6 × 5 L shake flasks on an Orbitron shaker at 160 rpm and 30 °C. Fermentation is complete after 14 days. Cells and XAD-16 resin are separated from the supernatant by centrifugation on a Beckmann Avanti J-26 XP equipped with a JLA 8.1 rotor at 8000 rcf. The supernatant is filtered through glass wool and dried using a rotary evaporator at 25 mbar and 40 °C water bath temperature. The dry supernatant is dissolved in 500 mL of a 1:1 mixture of milliQ water and acetone, transferred into 50 mL falcons and centrifuged at 8000 rcf and 4 °C for 15 min using an Eppendorf falcon table centrifuge to remove proteins and bigger peptides. The remaining acetone in the supernatant is evaporated using a rotary evaporator at 200 mbar and 40 °C water bath temperature to yield 250 mL of a concentrated aqueous supernatant.

Supernatant derivatization is carried out by the addition of 100 mL of Di-*tert*-butyl dicarbonate (Boc<sub>2</sub>O) and 100 mL of 1 M NaHCO<sub>3</sub> solution with a pH of 8.8 to the concentrated supernatant and stirring the mixture at room temperature on a magnetic stirrer. The reaction was controlled by HPLC-MS and

stopped after the conversion yield of the biaryllytite (**1**) to its Boc-protected congener (**2**) was stagnating and no longer improved. Afterwards, the aqueous reaction broth is transferred to a separating funnel and extracted twice with equal volumes of ethyl acetate, whereas the Boc-protected biaryllytite (**2**) stays in the water layer. Excess ethyl acetate is removed using a rotary evaporator at 100 mbar and 40 °C water bath temperature. The water layer is extracted again twice with equal volumes of n-butanol. The Boc-protected biaryllytite (**2**) separates into the n-butanol layer. The remaining solvent is evaporated using a rotary evaporator at 25 mbar and 60 °C water bath temperature. The dry n-butanol layer is redissolved with 20 mL of methanol. The purification of **2** was performed using a Dionex UltiMate 3000 semi-preparative system equipped with an automated fraction collector (Thermo Fisher Scientific, Waltham, MA, USA). Compound separation was achieved with a gradient of acetonitrile (B) in ddH<sub>2</sub>O (A) on an ACQUITY BEH C18 column (250 × 10 mm, 5 μm dp) (Waters, Eschborn, Germany) at a flow rate of 5.0 mL/min and 45 °C. The initial gradient was held at 5% B for 2 min and then elevated to 25% B within 1 min. After that, the B level was increased to 58% within 25 min and subsequently raised to 95% within 0.5 min and held there for 2 min. Finally, the gradient was ramped back to 5% B in 0.5 min and re-equilibrated for the next injection for 1.5 min. Detection was performed using the 3D plot of a DAD detector by absorption at 222 and 287 nm and a Thermo Scientific ISQ EC single quadrupole MS system with a HESI ion source in alternating polarity mode. The measuring parameters were 296 °C vaporizer temperature; 300 °C ion transfer tube temperature; 51.6 psig sheath gas pressure; 5.9 psig aux gas pressure; 0.5 psig sweep gas pressure; +3000 V source voltage (positive mode); and -2000 V source voltage (negative mode). Fraction collection was performed by time according to the single-ion monitoring (SIM) chromatogram peaks of 544.3 *m/z* ([M + H]<sup>+</sup>) and 542.3 *m/z* ([M - H]<sup>-</sup>). After solvent evaporation by lyophilization, **2** was obtained as a pale yellow solid. LC-hrMS analysis shows a single peak with an exact mass of 544.2765 *m/z* [M + H]<sup>+</sup> (calculated as 544.2766 *m/z*).

## 2.5.5 Structure Elucidation

### 2.5.5.1 NMR Conditions and Spectroscopic Data

One-dimensional and two-dimensional NMR data used for the elucidation of the structure of **2** are acquired on a Bruker Ascend 700 spectrometer equipped with a 5 mm TXI cryoprobe (<sup>1</sup>H at 700 MHz, <sup>13</sup>C at 175 MHz). All observed chemical shift values ( $\delta$ ) are given in ppm and coupling constant values (*J*) in Hz. Standard pulse programs are used for HMBC, HSQC and gCOSY experiments. HMBC experiments are optimized for <sup>2,3</sup>J<sub>C-H</sub> = 6 Hz. The spectra are recorded in methanol-*d*<sub>4</sub>, and chemical shifts of the solvent signals at  $\delta^H$  3.31 ppm and  $\delta^C$  49.2 ppm are used as reference signals for spectra calibration. Respective measurements for **1** are carried out in DMSO-*d*<sub>6</sub> with reference signals at  $\delta^H$  2.50 ppm and  $\delta^C$  39.5 ppm for

calibration; temperature is set to 313 K for resolution improvement. To increase sensitivity, all measurements are conducted in a 5 mm Shigemi tube (Shigemi Inc., Allison Park, PA, USA). The NMR signals are grouped in tables and correspond to the numbering in the schemes corresponding to every table. All structure formulae devised by NMR will be made publicly available under their corresponding name in NPatlas [37,38].

#### 2.5.5.2 Elucidation of the Absolute Stereochemistry

To determine the absolute configurations of amino acids Marfey's derivatization method is employed. Approximately 50  $\mu\text{g}$  of the peptide to analyze is dried in a glass vial at 110 °C. One hundred microliters of 6 N HCl is added, the vial is filled with  $\text{N}_2$  gas and incubated at 110 °C for 45 min (for leucine) to 8 h (for histidine and tyrosine) for peptide hydrolysis. Notably, even after 8 h of incubation, the C–N biaryl bond as well as the *N*-terminal *N*-methyl group are only very poorly cleaved to release free histidine and tyrosine, while **1** is mostly hydrolyzed to free leucine and an intact biaryl backbone. The vial is subsequently opened and the contained fluid is dried at 110 °C. The residue is taken up in 100  $\mu\text{L}$  of  $\text{ddH}_2\text{O}$  and split into two 2 mL Eppendorf tubes. To each tube 20  $\mu\text{L}$  of 1 N  $\text{NaHCO}_3$  is added as well as 20  $\mu\text{L}$  of D- or L-(1-fluoro-2,4-dinitrophenyl-5-leucine amide) (FDLA) as a 1% solution in acetone. The mixture is incubated at 40 °C and centrifuged at 700 rpm. Ten microliters of 2 N HCl solution is added to quench the reaction and 300  $\mu\text{L}$  of ACN is added to obtain a total volume of 400  $\mu\text{L}$ . The Eppendorf tube is centrifuged at 15,000 rpm in a table centrifuge and transferred into a conical HPLC vial.

UHPLC-*hr*MS analysis is performed on a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Bruker maXis 4G ultra-high-resolution quadrupole time-of-flight (UHR-qTOF) MS equipped with a high-resolution electrospray ionization (HRESI) source (Bruker Daltonics, Billerica, MA, USA), as described in Section 3.5.3. The separation of a 1  $\mu\text{L}$  sample is achieved with a gradient of acetonitrile with 0.1% formic acid (B) in  $\text{ddH}_2\text{O}$  with 0.1% formic acid (A) on an ACQUITY BEH C18 column (100  $\times$  2.1 mm, 1.7  $\mu\text{m}$   $d_p$ ) (Waters, Eschborn, Germany) equipped with a Waters VanGuard BEH C18 1.7  $\mu\text{m}$  guard column at a flow rate of 0.6 mL/min and 45 °C. The separation method starts with a linear gradient from 5 to 10% B over 1 min, followed by a linear gradient to 35% B over 14 min. After that, the B level is raised to 55% over 7 min, subsequently increased to 80% within 3 min and held there for one minute. Finally, the gradient is ramped back to 5% B in 0.5 min and the column is re-equilibrated for the next injection for 4.5 min. The detection of Marfey's derivatives is done by mass spectrometry and UV detection at 340 nm. Identification of the correct stereochemistry of the amino acid is done via comparison of retention times to FDLA derivatized standards [32].

### 2.5.6 Assessment of Antimicrobial Activities

All microorganisms used in this study were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ), the Coli Genetic Stock Center (CGSC) or were part of our internal collection, and were handled according to standard sterile microbiological procedures and techniques.

Microbroth dilution assays were used to test **1** and **2** on the following panel of bacteria and fungi: *Escherichia coli* wild-type BW25113 (DSM 27469), *E. coli* JW0451-2 (*acrB*-efflux pump deletion mutant of *E. coli* BW25113), *Pseudomonas aeruginosa* PA14 (DSM 19882), *Bacillus subtilis* DSM 10, *Staphylococcus aureus* Newman, *Candida albicans* DSM 1665, *Citrobacter freundii* DSM 30039, *Pichia anomala* DSM 6766, *Mycobacterium smegmatis* Mc<sup>2</sup> 155, *Cryptococcus neoformans* DSM 11959, *Mucor hiemalis* DSM 2656 and *Acinetobacter baumannii* DSM 30008. For microbroth dilution assays, the respective overnight cultures were prepared from cryogenically preserved cultures and were diluted to achieve a final inoculum of 10<sup>4</sup>–10<sup>5</sup> colony-forming units (cfu)/mL. The tested derivatives were prepared as DMSO stocks (5 mg/mL).

Serial dilutions of **1** and **2** in the respective growth medium (0.06 to 64 µg/mL) were prepared in sterile 96-well plates and the suspension of bacteria or fungi were added. The cell suspension was added and microorganisms were grown for 24 h at either 30 °C or 37 °C. Minimum inhibitory concentrations (MIC) are defined as the lowest compound concentration where no visible growth is observed.

### 2.5.7 Molecular Cloning, Construction of Plasmids, Maintenance of Bacterial Cultures

Routine handling of nucleic acids, such as isolation of plasmid DNA, restriction endonuclease digestions, DNA ligations and other DNA manipulations, was performed according to standard protocols [39]. *E. coli* HS996 (Invitrogen, Waltham, MA, USA) was used as the host for standard cloning experiments. *E. coli* strains were cultured in an LB liquid medium or on LB agar (1% tryptone, 0.5% yeast extract and 0.5% NaCl (1.5% agar) at 30–37 °C and 200 rpm) overnight. The antibiotic kanamycin was used at the following final concentration: 50 µg/mL. The transformation of *E. coli* strains was achieved via electroporation in 0.1-cm-wide cuvettes at 1250 V with a resistance of 200 Ω and a capacitance of 25 µF. Plasmids were purified either by standard alkaline lysis [39] or by using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, MA, USA) or the NucleoBond PC100 kit (Macherey-Nagel, Duren, Germany). Restriction endonucleases, alkaline phosphatase (FastAP) and T4 DNA ligase were purchased from Thermo Fisher Scientific. Oligonucleotides used for PCR and sequencing were acquired from Sigma-Aldrich and are listed in Tables S9 and S10.

PCRs were carried out in a Mastercycler® pro (Eppendorf) using Phusion™ High-Fidelity according to the manufacturer's protocol. The temperature and duration settings for each thermocycling step in PCR

with Phusion™ High-Fidelity polymerase were performed as follows: initial denaturation (30 s, 98 °C); 33 cycles of denaturation (15 s, 98 °C), annealing (15 s, 53–72 °C, depending on the melting temperature of primers) and elongation (based on PCR product length 30 s/1 kb, 72 °C); and final extension (10 min, 72 °C). PCR products or DNA fragments from restriction digestions were purified by agarose gel electrophoresis and isolated using the PCR clean-up gel extraction kit using Nucleo Spin® (Macherey-Nagel). After selection with a suitable antibiotic, clones harboring correct recombination products were identified by plasmid isolation and restriction analysis with a set of different restriction endonucleases. In addition to restriction analysis, the integrity of the constructs for induced gene expression was verified by sequencing.

According to previously established electroporation procedures for *M. xanthus* DK1622 [40,41], the host strain *M. xanthus* DK1622 was transformed with the generated expression constructs (Table S13). *M. xanthus* DK1622 transformants were routinely cultivated at 30 °C in a CTT medium or on CTT agar (1% casitone, 10 mM Tris buffer pH 7.6, 1 mM KH<sub>2</sub>PO<sub>4</sub> pH 7.6 and 8 mM MgSO<sub>4</sub> (1.5% agar) pH adjusted to 7.6). Liquid cultures were grown in Erlenmeyer flasks on an orbital shaker at 180 rpm for 3–6 days. *M. xanthus* transformants were selected by adding 50 µg/mL kanamycin to the fermentation culture. The correct chromosomal integration of the expression constructs into the *mx8 attb1* site was confirmed by PCR. Genomic DNA of the transformants was isolated using the Gentra® Puregene® Yeast/Bacteria Genomic DNA Purification Kit (Qiagen) according to the manufacturer's instructions.

For each expression construct, the correct chromosomal integration was confirmed according a previous study [42], using two different primer combinations to reveal PCR products of the expected sizes: *mx8-attB-up2/Mx8-attP-down* (427 bp) and *mx8-attP-up2/Mx8-attB-down* (403 bp) (Table S9). The genomic DNA of *M. xanthus* DK1622 was used as the negative control. A complementary experiment using the following primer combination revealed a specific PCR product for *M. xanthus* DK1622 wild type, but not for any of the *M. xanthus* DK1622 transformants harboring one of the generated constructs: *mx8-attB-up2/Mx8-attB-down* (449 bp).

## 2.6 Acknowledgments

The authors thank Viktoria George and Alexandra Amann for performing bioactivity assays.

## 2.7 References

1. Cao, P.; Dey, A.; Vassallo, C.N.; Wall, D. How Myxobacteria Cooperate. *J. Mol. Biol.* **2015**, doi:10.1016/j.jmb.2015.07.022.
2. Muñoz-Dorado, J.; Marcos-Torres, F.J.; García-Bravo, E.; Moraleda-Muñoz, A.; Pérez, J. Myxobacteria: Moving, killing, feeding, and surviving together. *Front. Microbiol.* **2016**, *7*, 781, doi:10.3389/fmicb.2016.00781.
3. Herrmann, J.; Fayad, A.A.; Müller, R. Natural products from myxobacteria: Novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34*, 135–160, doi:10.1039/C6NP00106H.
4. Wenzel, S.C.; Müller, R. Myxobacteria—‘Microbial factories’ for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, *5*, 567–574, doi:10.1039/b901287g.
5. Panter, F.; Bader, C.D.; Müller, R. Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Chem. Sci.* **2021**, 5994–6010, doi:10.1039/D0SC06919A.
6. Medema, M.H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M.A.; Weber, T.; Takano, E.; Breitling, R. antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **2011**, *39*, W339–W346, doi:10.1093/nar/gkr466.
7. De Jong, A.; van Hijum, S.A.F.T.; Bijlsma, J.J.E.; Kok, J.; Kuipers, O.P. BAGEL: A web-based bacteriocin genome mining tool. *Nucleic Acids Res.* **2006**, *34*, W273–W279, doi:10.1093/nar/gkl237.
8. Agrawal, P.; Khater, S.; Gupta, M.; Sain, N.; Mohanty, D. RiPPMiner: A bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res.* **2017**, *45*, W80–W88, doi:10.1093/nar/gkx408.
9. Santos-Aberturas, J.; Chandra, G.; Frattaruolo, L.; Lacret, R.; Pham, T.H.; Vior, N.M.; Eyles, T.H.; Truman, A.W. Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. *Nucleic Acids Res.* **2019**, *47*, 4624–4637, doi:10.1093/nar/gkz192.
10. Kloosterman, A.M.; Shelton, K.E.; van Wezel, G.P.; Medema, M.H.; Mitchell, D.A. RRE-Finder: A Genome-Mining Tool for Class-Independent RiPP Discovery. *mSystems* **2020**, *5*, doi:10.1128/mSystems.00267-20.
11. Merwin, N.J.; Mousa, W.K.; Dejong, C.A.; Skinnider, M.A.; Cannon, M.J.; Li, H.; Dial, K.; Gunabalasingam, M.; Johnston, C.; Magarvey, N.A. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 371–380, doi:10.1073/pnas.1901493116.

12. De Los Santos, E.L.C. NeuRiPP: Neural network identification of RiPP precursor peptides. *Sci. Rep.* **2019**, *9*, 13406, doi:10.1038/s41598-019-49764-z.
13. Tietz, J.I.; Schwalen, C.J.; Patel, P.S.; Maxson, T.; Blair, P.M.; Tai, H.-C.; Zakai, U.I.; Mitchell, D.A. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat. Chem. Biol.* **2017**, *13*, 470–478, doi:10.1038/nchembio.2319.
14. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35, doi:10.1093/nar/gkab335.
15. Russell, A.H.; Truman, A.W. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comput. Struct. Biotechnol. J.* **2020**, doi:10.1016/j.csbj.2020.06.032.
16. Hirsch, H.J. Bacteriocins from *Myxococcus fulvus* (Myxobacterales). *Arch. Microbiol.* **1977**, *115*, 45–49, doi:10.1007/bf00427843.
17. Tsai, H.; Hirsch, H.-J. The primary structure of fulvocin C from *Myxococcus fulvus*. *Biochim. Biophys. Acta Protein Struct.* **1981**, *667*, 213–217, doi:10.1016/0005-2795(81)90082-9.
18. Mccurdy, H.D.; MacRae, T.H. Xanthacin. A bacteriocin of *Myxococcus xanthus* fb. *Can. J. Microbiol.* **1974**, *20*, 131–135, doi:10.1139/m74-021.
19. Viehrig, K.; Surup, F.; Volz, C.; Herrmann, J.; Abou Fayad, A.; Adam, S.; Kohnke, J.; Trauner, D.; Müller, R. Structure and biosynthesis of crocagins: Polycyclic postranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem.* **2017**, 1–5, doi:10.1002/anie.201612640.
20. Adam, S.; Klein, A.; Surup, F.; Koehnke, J. The structure of CgnJ, a domain of unknown function protein from the crocagin gene cluster. *Acta Crystallogr. F Struct. Biol. Commun.* **2019**, *75*, 205–211, doi:10.1107/S2053230X19000712.
21. Hug, J.J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. Biosynthesis of Cittilins, Unusual Ribosomally Synthesized and Post-translationally Modified Peptides from *Myxococcus xanthus*. *ACS Chem. Biol.* **2020**, *15*, 2221–2231, doi:10.1021/acscchembio.0c00430.
22. *Cittilins: Bicyclic Isotriptyrosines from Myxococcus xanthus: 24th General meeting German Chemists' Society*; Trowitzsch-Kienast, W., Ed.; Hamburg, Germany; 5–11 September 1993.
23. Reichenbach, H.; Höfle, G. Myxobacteria as producers of secondary metabolites. In *Drug Discovery from Nature*; Grabley, S., Thiericke, R., Eds.; Springer: Berlin, Germany, 1999; pp 149–179.
24. Zdouc, M.M.; Alanjary, M.M.; Zarazúa, G.S.; Maffioli, S.I.; Crüsemann, M.; Medema, M.H.; Donadio, S.; Sosio, M. A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters. *Cell Chem. Biol.* **2020**, doi:10.1016/j.chembiol.2020.11.009.



25. Chai, Y.; Pistorius, D.; Ullrich, A.; Weissman, K.J.; Kazmaier, U.; Müller, R. Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* **2010**, *17*, 296–309, doi:10.1016/j.chembiol.2010.01.016.
26. Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9*, 4898–4908, doi:10.1039/C8SC01325J.
27. Sandmann, A.; Frank, B.; Müller, R. A transposon-based strategy to scale up myxothiazol production in myxobacterial cell factories. *J. Biotechnol.* **2008**, *135*, 255–261, doi:10.1016/j.jbiotec.2008.05.001.
28. Moore, J.M.; Bradshaw, E.; Seipke, R.F.; Hutchings, M.I.; McArthur, M. Use and discovery of chemical elicitors that stimulate biosynthetic gene clusters in *Streptomyces* bacteria. *Methods Enzymol.* **2012**, *517*, 367–385, doi:10.1016/B978-0-12-404634-4.00018-8.
29. Seyedsayamdost, M.R. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7266–7271, doi:10.1073/pnas.1400019111.
30. De Felício, R.; Ballone, P.; Bazzano, C.F.; Alves, L.F.G.; Sigrist, R.; Infante, G.P.; Niero, H.; Rodrigues-Costa, F.; Fernandes, A.Z.N.; Tonon, L.A.C.; et al. Chemical Elicitors Induce Rare Bioactive Secondary Metabolites in Deep-Sea Bacteria under Laboratory Conditions. *Metabolites* **2021**, *11*, 107, doi:10.3390/metabo11020107.
31. Hug, J.J.; Müller, R. Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products: 6.09. In *Comprehensive Natural Products III*; Liu, H.-W.B., Begley, T.P., Eds.; Elsevier: Oxford, UK, 2020; pp 149–216, ISBN 978-0-08-102691-5.
32. B'Hymer, C.; Montes-Bayon, M.; Caruso, J.A. Marfey's reagent: Past, present, and future uses of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide. *J. Sep. Sci.* **2003**, *26*, 7–19.
33. Pogorevc, D.; Müller, R. Biotechnological production optimization of argyrins—A potent immunomodulatory natural product class. *Microb. Biotechnol.* **2021**, doi:10.1111/1751-7915.13959.
34. Phan, C.-S.; Matsuda, K.; Balloo, N.; Fujita, K.; Wakimoto, T.; Okino, T. Argicyclamides A-C Unveil Enzymatic Basis for Guanidine Bis-prenylation. *J. Am. Chem. Soc.* **2021**, *143*, 10083–10087, doi:10.1021/jacs.1c05732.
35. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649, doi:10.1093/bioinformatics/bts199.

36. Finn, R.D.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–D285, doi:10.1093/nar/gkv1344.
37. Linington, R.G. npatlas—The Natural Products Atlas. Available online: <https://www.npatlas.org> (accessed on 22 November 2021).
38. Van Santen, J.A.; Jacob, G.; Singh, A.L.; Aniebok, V.; Balunas, M.J.; Bunsko, D.; Neto, F.C.; Castaño-Espriu, L.; Chang, C.; Clark, T.N.; et al. The natural products atlas: An open access knowledge base for microbial natural products discovery. *ACS Cent. Sci.* **2019**, *5*, 1824–1833, doi:10.1021/acscentsci.9b00806.
39. Sambrook, J.; Russell, D.W. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2001.
40. Kashafi, K.; Hartzell, P.L. Genetic suppression and phenotypic masking of a *Myxococcus xanthus* *frzF*-defect. *Mol. Microbiol.* **1995**, *15*, 483–494.
41. Hug, J.J.; Panter, F.; Krug, D.; Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319–334, doi:10.1007/s10295-018-2105-6.
42. Pogorevc, D.; Panter, F.; Schillinger, C.; Jansen, R.; Wenzel, S.C.; Müller, R. Production optimization and biosynthesis revision of coralopyronin A, a potent anti-filarial antibiotic. *Metab. Eng.* **2019**, *55*, 201–211, doi:10.1016/j.ymben.2019.07.010.

## Supporting Information

# Genome-guided Discovery of the First Myxobacterial Biarylittide Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis

Previously published in:

Joachim J. Hug<sup>†</sup>, Nicolas A. Frank<sup>†</sup>, Christine Walt, Petra Šenica, Fabian Panter,  
Rolf Müller\*

*Molecules*, 2021 December 10; **26**(24):7483

DOI: 10.3390/molecules26247483

### Affiliation

Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

## S 2.1 Myxobacterial Culture Media

**Table S1.** Recipe for VY/2 medium.

VY/2 – Medium			
Amount	Ingredient	Concentration	Supplier
5 g/L	Yeast (entire cells)	-	Deutsche Hefewerke GmbH, Nürnberg, Germany
5 g/L	Soluble Starch	-	Carl Roth GmbH, Karlsruhe, Germany
1 g/L	CaCl <sub>2</sub> • 2H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 g/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	-	Grüssing GmbH, Filsum, Germany
10 mL/L	TRIS • HCl pH 8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, pH adjusted to 7.2 with 1 N HCl			

**Table S2.** Recipe for VY medium.

VY – Medium			
Amount	Ingredient	Concentration	Supplier
10 g/L	Yeast (entire cells)	-	Deutsche Hefewerke GmbH, Nürnberg, Germany
5 g/L	Soluble Starch	-	Carl Roth GmbH, Karlsruhe, Germany
1 g/L	CaCl <sub>2</sub> • 2H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 g/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	-	Grüssing GmbH, Filsum, Germany
10 mL/L	TRIS • HCl pH 8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, pH adjusted to 7.2 with 1 N HCl			

**Table S3.** Recipe for CFL medium.

CFL – Medium			
Amount	Ingredient	Concentration	Supplier
3 g/L	Tryptone	-	Becton, Dickinson and Company, Sparks, MD, USA
1 g/L	Soytone	-	Becton, Dickinson and Company, Sparks, MD, USA
3.5 g/L	Soluble Starch	-	Carl Roth GmbH, Karlsruhe, Germany
4 g/L	Maltose Monohydrate	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
2 g/L	Glucose	-	Carl Roth GmbH, Karlsruhe, Germany
0.5 g/L	CaCl <sub>2</sub> • 2H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 g/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	-	Grüssing GmbH, Filsum, Germany
10 mL/L	TRIS • HCl pH 8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, pH adjusted to 7.5 with 1 N KOH			

**Table S4.** Recipe for P medium.

<b>P – Medium</b>			
Amount	Ingredient	Concentration	Supplier
2 g/L	Peptone (Phytone)	-	Becton, Dickinson and Company, Sparks, MD, USA
4 g/L	Procion FM582	-	Hoechst GmbH, Frankfurt, Germany
8 g/L	Soluble Starch	-	Carl Roth GmbH, Karlsruhe, Germany
2 g/L	Glucose	-	Carl Roth GmbH, Karlsruhe, Germany
1 g/L	CaCl <sub>2</sub> • 2H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 g/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	-	Grüssing GmbH, Filsum, Germany
10 mL/L	TRIS • HCl pH 8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, pH adjusted to 7.5 with 1 N KOH			

**Table S5.** Recipe for M medium.

<b>M – Medium</b>			
Amount	Ingredient	Concentration	Supplier
10 g/L	Soy Peptone	-	Becton, Dickinson and Company, Sparks, MD, USA
10 g/L	Maltose monohydrate	-	Becton, Dickinson and Company, Sparks, MD, USA
1 g/L	CaCl <sub>2</sub> • 2H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 g/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	-	Grüssing GmbH, Filsum, Germany
10 mL/L	TRIS • HCl pH 8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, pH adjusted to 7.5 with 1 N KOH			

**Table S6.** Recipe for YM medium.

<b>YM – Medium</b>			
Amount	Ingredient	Concentration	Supplier
3 g/L	Yeast extract	-	Becton, Dickinson and Company, Sparks, MD, USA
3 g/L	Malt extract	-	Becton, Dickinson and Company, Sparks, MD, USA
5 g/L	Peptone (Phytone)	-	Becton, Dickinson and Company, Sparks, MD, USA
10 g/L	Glucose	-	Carl Roth GmbH, Karlsruhe, Germany
1 mL/L	K <sub>x</sub> H <sub>y</sub> PO <sub>4</sub> buffer pH 8.0	0.8 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 g/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	-	Grüssing GmbH, Filsum, Germany
10 mL/L	TRIS • HCl pH 8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, pH adjusted to 7.5 with 1 N KOH			

**Table S7.** Recipe for CTT medium.

<b>CTT – Medium</b>			
Amount	Ingredient	Concentration	Supplier
10 g/L	Casitone	-	Becton, Dickinson and Company, Sparks, MD, USA
10 mL/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	0.8 M	Grüssing GmbH, Filsum, Germany
1 mL/L	K <sub>x</sub> H <sub>y</sub> PO <sub>4</sub> buffer pH 8.0	0.1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
10 mL/L	TRIS • HCl pH8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, pH adjusted to 7.6 with 1 N KOH			

**Table S8.** Recipe for AMB medium.

<b>AMB – Medium</b>			
Amount	Ingredient	Concentration	Supplier
2.5 g/L	Casitone	-	Becton, Dickinson and Company, Sparks, MD, USA
5 g/L	Soluble starch	-	Carl Roth GmbH, Karlsruhe, Germany
10 mL/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	0.8 M	Grüssing GmbH, Filsum, Germany
1 mL/L	K <sub>x</sub> H <sub>y</sub> PO <sub>4</sub> buffer pH 8.0	0.1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
10 mL/L	TRIS • HCl pH 8.0	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, pH adjusted to 7.6 with 1 N KOH			

## S 2.2 Molecular Cloning, Construction of Expression Plasmids and Genetic Investigation

**Table S9.** List of oligonucleotides used in this study.

No.	Primer name	Primer sequence 5'–3'
1	Fw_mx8_PagI	ATATTCATGAGGCCATCGTCGAAAGAGTCG
2	Rv_mx8_ScaI	ATATAGTACTAGGACTCCTCTGGCTGGGTG
3	Fw_An_d48_NdeI	ATATCATATGAACTACCTGCACTGA <u>GAGGAGCCTCC</u> GTGCCCA
4	Rv_An_d48_EcoRI	TATATGAATTCGAGGCGGCGTAGCCCTCA
5	Mx8-attP-up2	CGACGGTGCCGACAAATAC
6	Mx8-attB-up2	GCGCACTGGACCATCACGTC
7	Mx8-attP-down	GGCTTGTCAGTCAACTGCG
8	Mx8-attB-down	CGGATAGCTCAGCGGTAGAG

Restriction sites in bold, *bytA* and *bytO* start are underlined, RBS of *bytO* in yellow.

**Table S10.** List of oligonucleotides for sequencing used in this study.

No.	Primer name	Primer sequence 5'–3'
1	P <sub>van</sub> _Seq1	TGTCAAGCTGCTGTTTTCGC
2	M13-29R	CAGGAAACAGCTATGACC
3	BiarylSeq1	GAGGCGGCGTAGCCCTCA
4	BiarylSeq2	GAGGAGCCTCCGTGCCCA

**Table S11.** List of PCR-amplified constructs.

No.	PCR product name/characteristics	Size [bp]	template	Primers used
1	mx8_integrase gene_PagI_ScaI	1945	pBen39	primer No.1 primer No.2
2	An_d48_bytAOZ	2034	gDNA from <i>P. fallax</i> An d48	primer No.1 primer No.2

**Table S12.** List of plasmids used in this study.

No.	Plasmid name/ characteristic	Size [kb]	Function	Reference
1	pFP <sub>van</sub> _pcyA	6.181	pCR 2.1 TOPO derivative; used as backbone for heterologous expression of <i>bytAO</i>	[1]
2	pBen39	7.616	mx8_integrase fragment mobilized on pUC-based plasmid pBen38/ <i>kanR</i> , <i>ampR</i>	unpublished

**Table S13.** List of genetic constructs generated in this study.

No.	Plasmid name	Construction details/ characteristics
1	pFP <sub>van</sub> _pcyA_mx8	Construct obtained by conventional restriction ligation of plasmid pFP <sub>van</sub> _pcyA and PCR product No. 1
2	pFP <sub>van</sub> _bytAOZ_mx8	Construct obtained by conventional restriction ligation of construct No. 1 and PCR product No. 2

S 2.2.1 Genetic Investigations

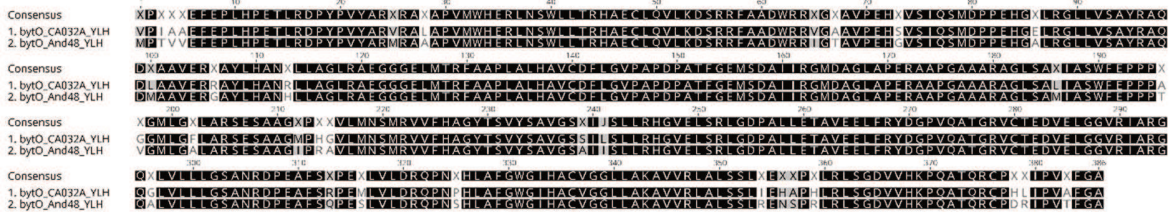


Figure S1. Amino acid alignment of myxobacterial BytO homologs from *Pyxidicoccus* sp. CA032A (*bytO\_CA032A\_YLH*) and *Pyxidicoccus fallax* An d48 (*bytO\_And48\_YLH*). Pairwise identity and identical sites: 91.5%.



Figure S2. Amino acid alignment of BytO homologs from *Planomonospora* sp. ID107089 (*bytO\_ID107089\_YFH*), *Planomonospora* sp. ID82291 (*bytO\_ID82291\_YYH*) and *Pyxidicoccus fallax* An d48 (*bytO\_And48\_YLH*). Pairwise identity: 59.5%, Identical sites: 40.7%.

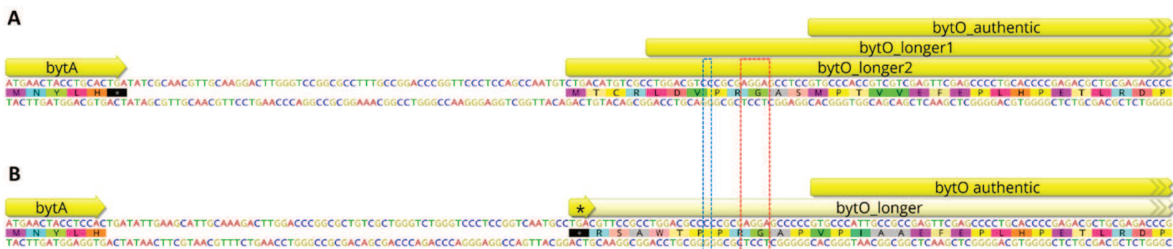
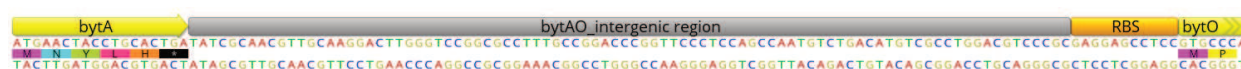


Figure S3. Nucleotide sequence and different orfs of *bytO* in *Pyxidicoccus fallax* An d48 (A) and *Pyxidicoccus* sp. CA032A (B) that are varying in nucleotide length. The length of a hypothetical *bytO* gene in *Pyxidicoccus* sp. CA032A is limited by the presence of a stop codon in the same translational frame. *Pyxidicoccus fallax* An d48 features in this genetic locus a single nucleotide deletion, which causes a frame shift (blue dashed box). Therefore two hypothetical longer orfs of *bytO* could be possibly expressed. Nevertheless, the most likely start of the myxobacterial *bytO* is in both nucleotide sequences labeled as *bytO authentic* for the following two reasons; firstly, a putative RBS (red dashed box) is only present in front of *bytO authentic*, and secondly in *Pyxidicoccus* sp. CA032A, only *bytO authentic* can be expressed, since any other longer version does not feature an appropriate start codon.



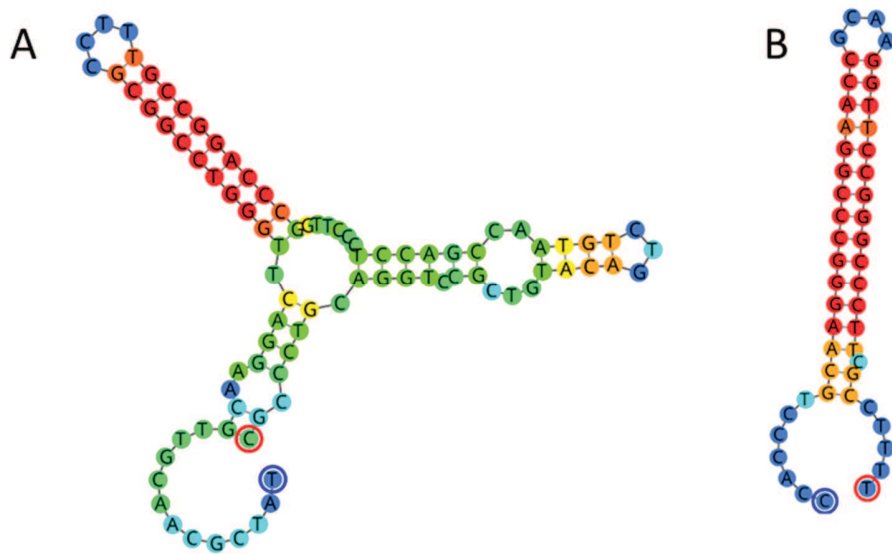
### S 2.2.2 Heterologous Production of Myxarylin in *M. xanthus* DK1622

To express the identified Myxarylin-BGC in *M. xanthus* DK1622, an expression vector was constructed based on the plasmid pFP<sub>van\_</sub>*pcyA* [1]. To enable efficient genomic integration of the expression plasmid, the mx8 integrase gene originating from the bacteriophage Mx8 was PCR-amplified (Phusion™ High-Fidelity polymerase, Thermo Fisher Scientific) from pBen39 and the resulting PCR product subcloned into pFP<sub>van\_</sub>*pcyA* via conventional restriction ligation (PacI, Scal). The resulting plasmid pFP<sub>van\_</sub>*pcyA\_mx8* (genetic construct 1) was subsequently used to incorporate the PCR-amplified Myxarylin-BGC (An\_d48\_*bytAOZ*, PCR product 2) from *P. fallax* An d48, which led to the final expression vector pFP<sub>van\_</sub>*bytAOZ\_mx8* (genetic construct 2). The PCR-amplified operon (An\_d48\_*bytAOZ*, PCR product 2) differs from the natural sequence in that sense, it lacks the intergenic region between *bytA* and *bytO* (termed *bytAO*\_intergenic region, Figure S3).

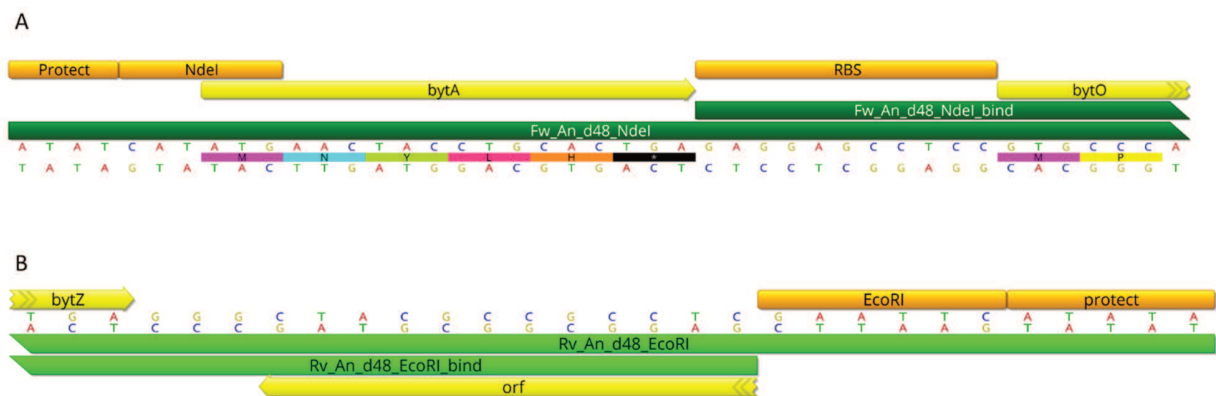


**Figure S4.** Nucleotide sequence encoding the precursor peptide *BytA*, the intergenic region termed *bytAO*, the putative ribosome binding site of *bytO*, and the first seven nucleotides encoding the cytochrome P450 dependent enzyme *BytO*.

Since the intergenic region *bytAO* features *in silico* the formation of secondary structures (Figure S5) – resembling those of known transcriptional terminators such as the *TD1* terminator from *M. xanthus* bacteriophage Mx8 [2] – the heterologous expression construct was designed in a way to omit this 90 bp intergenic region and only includes 11 bp in which a putative ribosome binding site is incorporated (..AGGA..) (Figure S6).



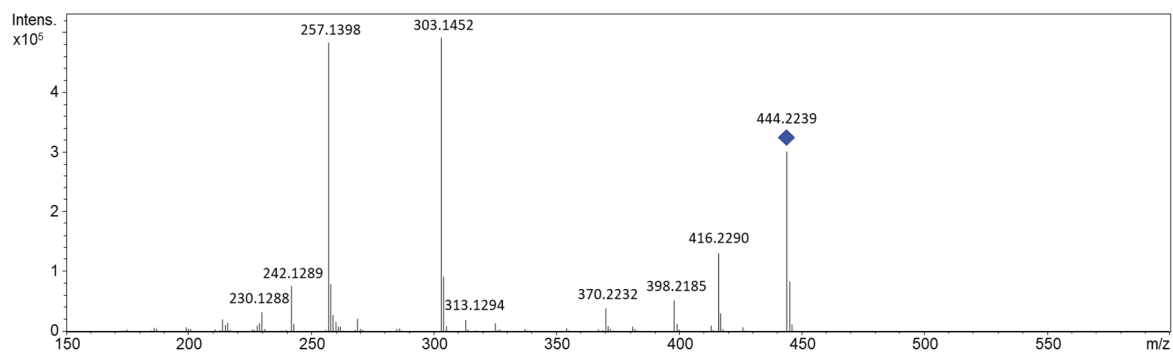
**Figure S5.** Predicted secondary structure of the 90 bp intergenic region bytAO (A), and the hair pin structure of the previously described 49 bp transcriptional terminator tD1 (B).



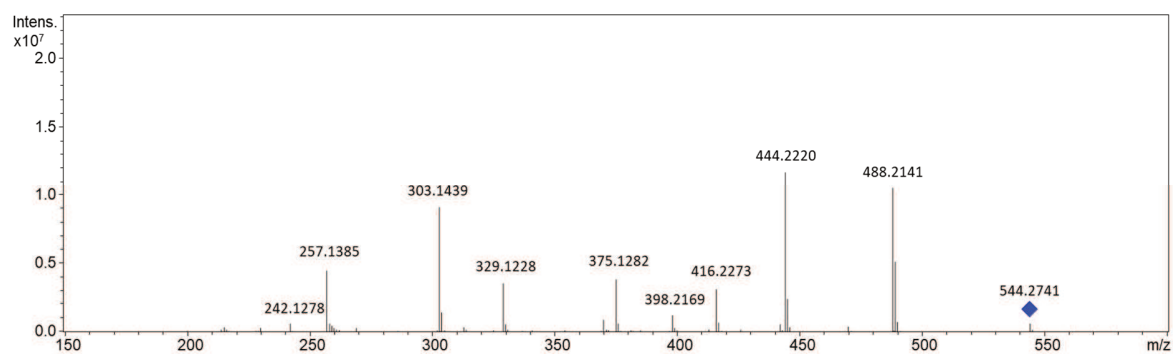
**Figure S6.** (A) Nucleotide sequence and annotation of the primer Fw\_An\_d48\_Ndel (Primer 3, 43 bp). Primer 3 comprises at the 5' end extensions which include a protection site for the NdeI restriction site, the authentic bytA nucleotide sequence, the identified RBS site of bytO and the first seven nucleotides of the gene encoding BytO. Within Primer 3, 18 bp at the 3' end is responsible for the site-specific annealing (termed Fw\_An\_d48\_Ndel\_bind). (B) The associated reverse primer Rv\_An\_d48\_EcoRI (Primer 4, 29 bp), was designed in a way that the genetic region downstream of bytZ remained unmodified in the respective PCR product. Hence, Primer 4 features at the 5' end an extension for the EcoRI restriction site including a 5 bp protection site and 18 bp at the 3' end for site-specific annealing (termed Rv\_An\_d48\_EcoRI\_bind).

## S 2.3 Compound Characterization

### S 2.3.1 Tandem MS Spectra



**Figure S7.** Tandem MS spectrum of **1** (collision energy: 35.0 eV). Leucine loss ( $\Delta m/z = 113.0841$ ) is visible e.g. between 416.2290 m/z and 303.1452 m/z (err. 0.0003 m/z) or 370.2232 m/z and 257.1398 m/z (err. 0.0007 m/z).



**Figure S8.** Tandem MS spectrum of **2** (collision energy: 35.9 eV).

## S 2.3.2 Structure Elucidation

## S 2.3.2.1 NMR-based Structure Elucidation

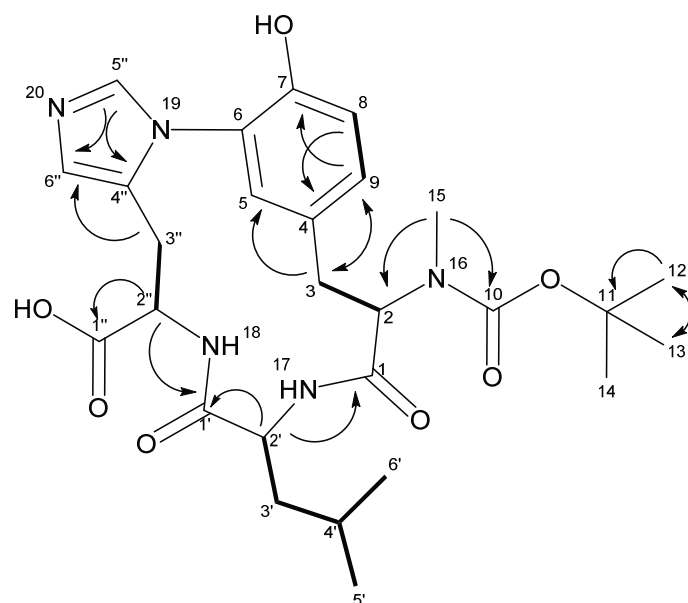
Table S14. NMR spectroscopic data of **2** measured in methanol-*d*<sub>4</sub> at 700/175 MHz.

#	$\Delta^{13}\text{C}$ [PPM]	$\Delta^1\text{H}$ [PPM], MULT (J [HZ])	$\Delta^{15}\text{N}$ [PPM]	COSY	HMBC	SEL NOESY	$^1\text{H}$ - $^{15}\text{N}$ HMBC
1	171.8	-	-	-	-	n.m.	-
2	59.7	4.66, m	-	3	-	n.m.	-
3	34.2	3.55,2.56, dd (15.19, 11.77)	-	2,3,5	1,2,5,9	n.m.	-
4	131.1	-	-	-	-	n.m.	-
5	126.1	6.84, m	-	-	-	n.m.	-
6	-	-	-	-	-	n.m.	-
7	149.7	-	-	-	-	n.m.	-
8	118.1	6.91, brs	-	9	4,5,7,9	9	-
9	130.2	7.03, d (8.13)	-	8	3,5,7,8	n.m.	-
10	157.6	-	-	-	-	n.m.	-
11	81.9	-	-	-	-	n.m.	-
12,13,14	28.8	1.48, s	-	-	11,12,13,14-	n.m.	-
15	30.7	2.90, s	-	-	2,10	n.m.	16
1'	173.7	-	-	-	-	n.m.	-
2'	52.8	4.86, s	-	3'	1,1',3',4'	n.m.	-
3'	43.7	1.58, m	-	2',5',6'	4'	n.m.	-
4'	26.1	1.60, m	-	5',6'	-	n.m.	-
5',6'	23.8	0.94, m	-	3'	-	n.m.	-
1''	177.8	-	-	-	-	n.m.	-
2''	55.2	4.61, dd (12.62, 2.78)	-	3''	1',1'',3'',4''	n.m.	-
3''	33.0	3.32,2.82, dd (16.47, 12.62)	-	2'',3'',6''	1'',2'',4'',6''	n.m.	18,20
4''	139.3	-	-	-	-	n.m.	-
5''	138.5	7.94, s	-	6''	4'',6''	-	19,20
6''	118.9	6.89, s	-	3'',5''	4'',5''	-	19,20
16	-	-	84.7	-	-	n.m.	-
17	-	-	-	-	-	n.m.	-
18	-	-	123.1	-	-	n.m.	-
19	-	-	174.6	-	-	n.m.	-
20	-	-	247.3	-	-	n.m.	-

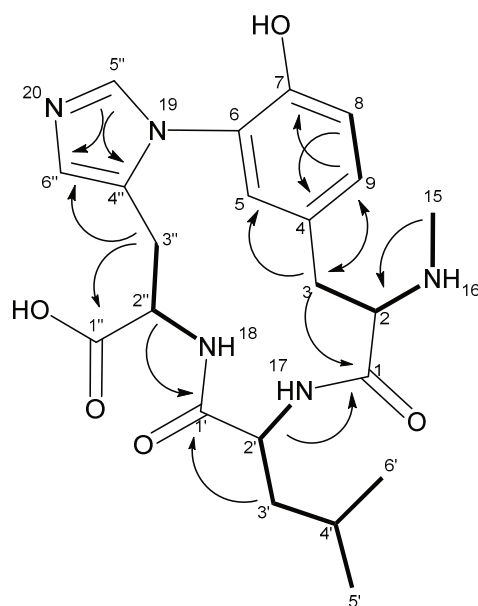
n.m.: not measured

Table S15. NMR spectroscopic data of **1** measured in DMSO-*d*<sub>6</sub> at 700/175 MHz.

#	$\Delta^{13}\text{C}$ [PPM]	$\Delta^1\text{H}$ [PPM], MULT (J [HZ])	COSY	HMBC
1	166.6	-	-	-
2	61.1	4.08, m	3,16	-
3	33.1	3.16, m	2,5,9	1,2,5,9
4	-	-	-	-
5	127.8	6.81, d (2.25)	3,9	3,7,9
6	125.3	-	-	-
7	149.3	-	-	-
8	117.1	7.08, d (8.34)	9	6,7
9	131.8	7.23, dd (8.24,2.25)	3,5,8	3,7,8
15	31.3	2.50, m	16	2
1'	171.9	-	-	-
2'	51.5	4.71, td (9.06,5.4)	3',17	-
3'	42.1	1.54,1.48, m	2',4'	1',2',4',5',6'
4'	24.1	1.67, m	5',6'	3',5',6'
5'	22.7	0.93, d (6.63)	4'	3',4',6'
6'	21.6	0.90 d (6.63)	4'	4',5'
1''	171.1	-	-	-
2''	50.6	4.58, ddd (12.25, 9.52, 2.78)	3'',18	-
3''	27.0	3.35,2.87, dd (15.19, 12.41)	2'',6''	1'',2'',4'',6''
4''	130.6	-	-	-
5''	135.6	9.35, s	6''	4'',6''
6''	120.4	7.16, s	3'',5''	4'',5''
16	-	8.79, brs	2	-
17	-	9.21, d (8.56)	2'	1
18	-	8.86, d (9.52)	2''	1'



**Figure S9.** Chemical structure and atom numbering of **2**. COSY correlations are represented as bold lines, HMBC correlations are marked with arrows.



**Figure S10.** Chemical structure and atom numbering of **1**. COSY correlations are represented as bold lines, HMBC correlations are marked with arrows.

## S 2.3.2.2 NMR Spectra Employed in Myxarylin Structure Elucidation

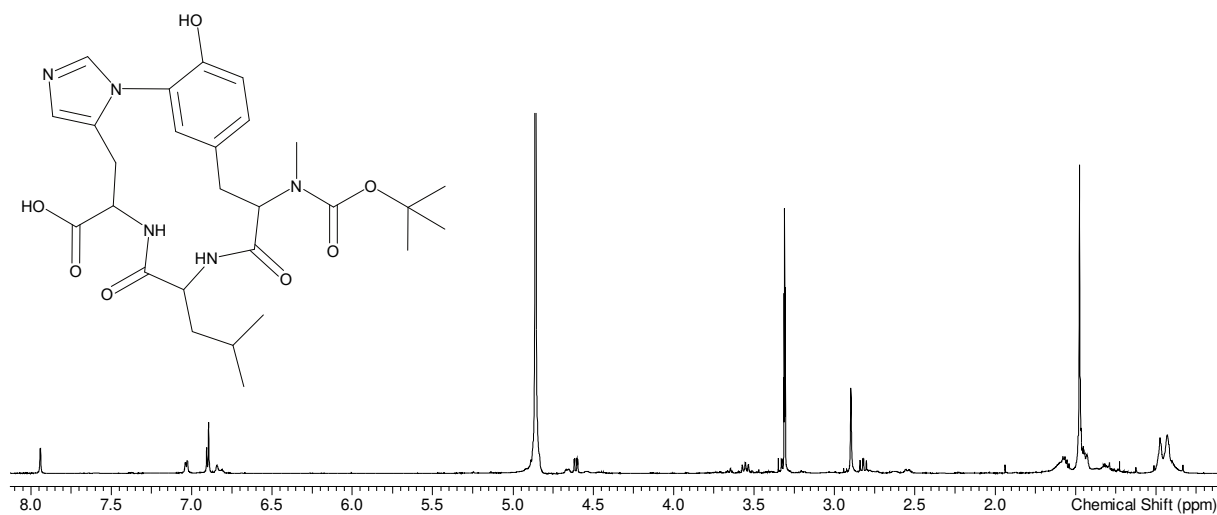


Figure S11.  $^1\text{H}$  spectrum of **2** measured in methanol- $d_4$  at 700 MHz.

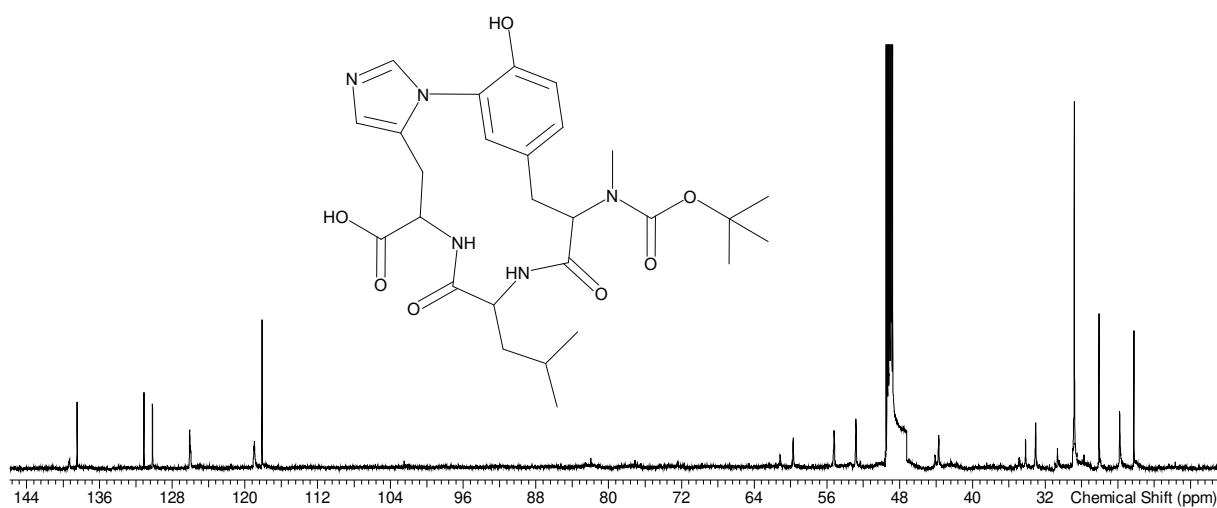


Figure S12.  $^{13}\text{C}$  spectrum of **2** measured in methanol- $d_4$  at 175 MHz.

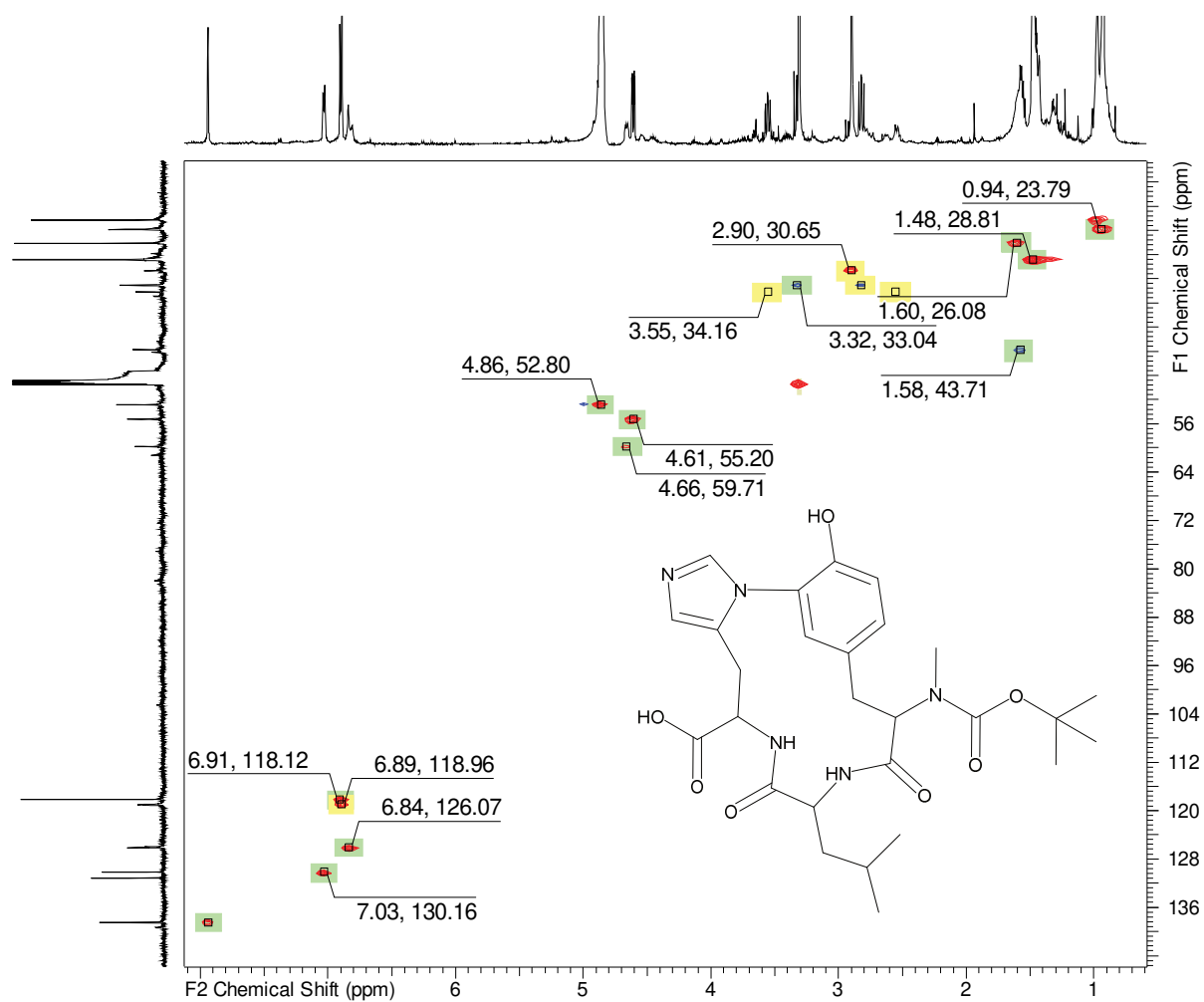


Figure S13. HSQC spectrum of **2** measured in methanol- $d_4$  at 700 ( $^1H$ ) and 175 ( $^{13}C$ ) MHz.



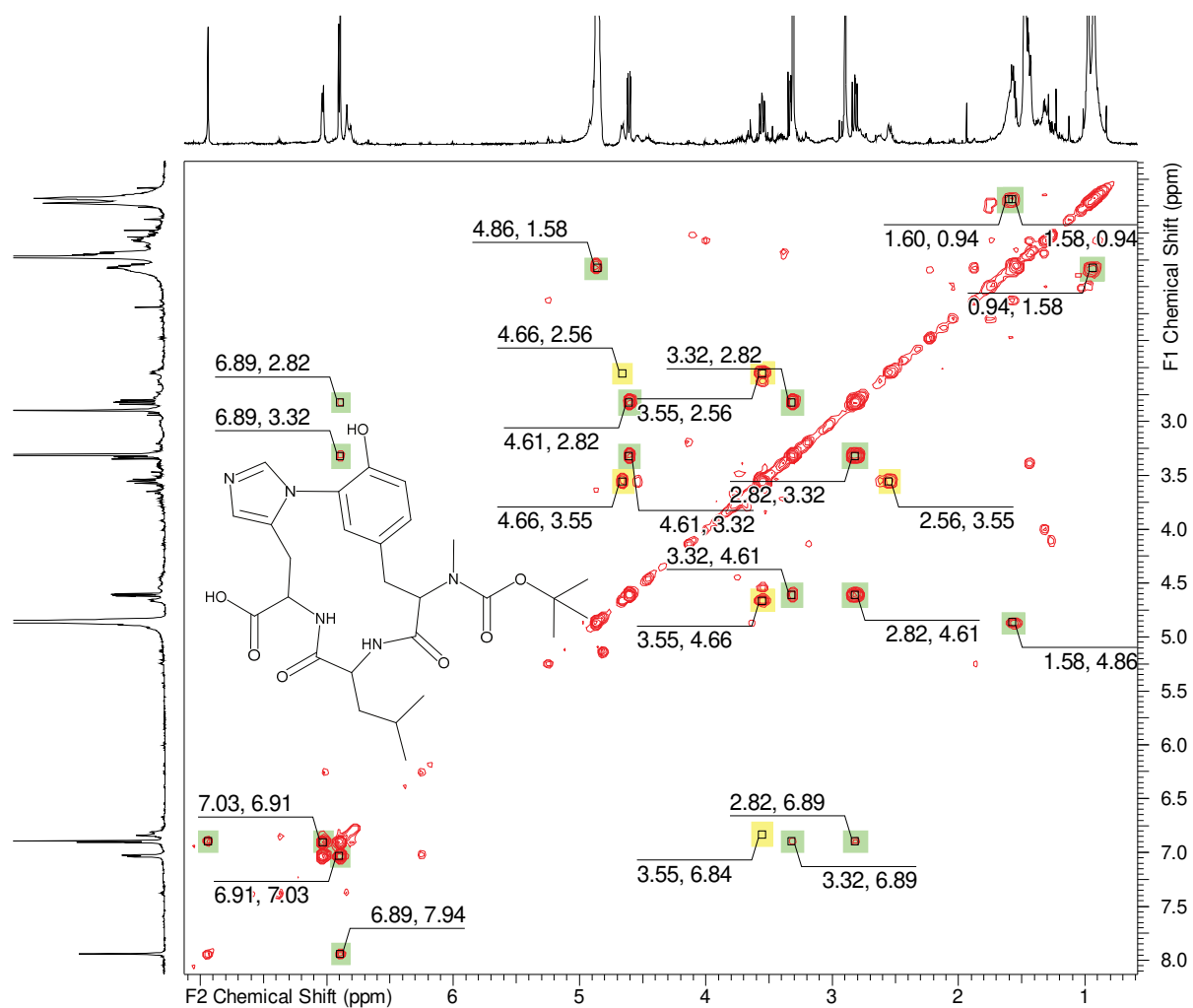


Figure S14. COSY spectrum of **2** measured in methanol- $d_4$  at 700 MHz.

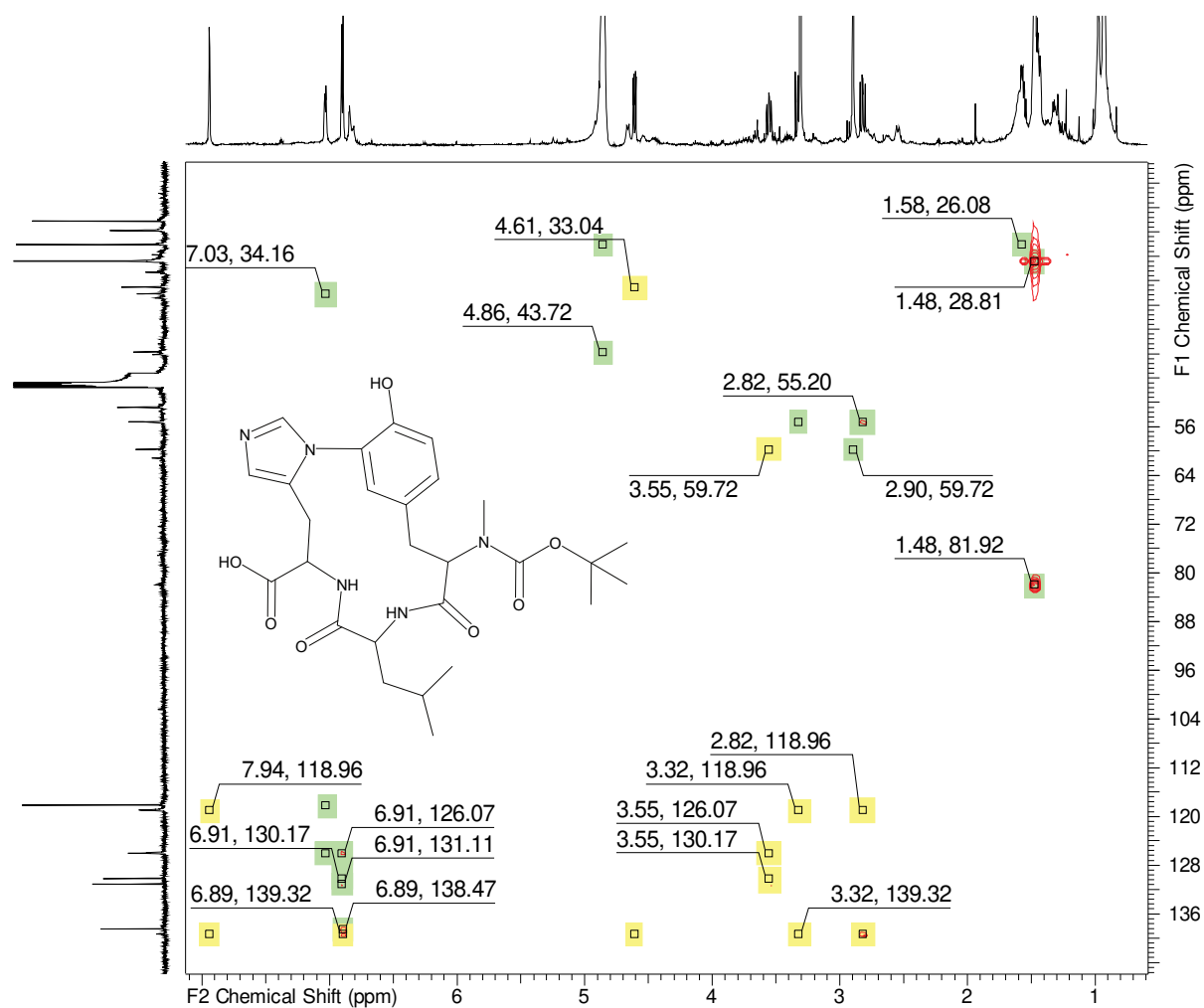


Figure S15. HMBC spectrum of **2** measured in methanol- $d_4$  at 700 ( $^1H$ ) and 175 ( $^{13}C$ ) MHz.

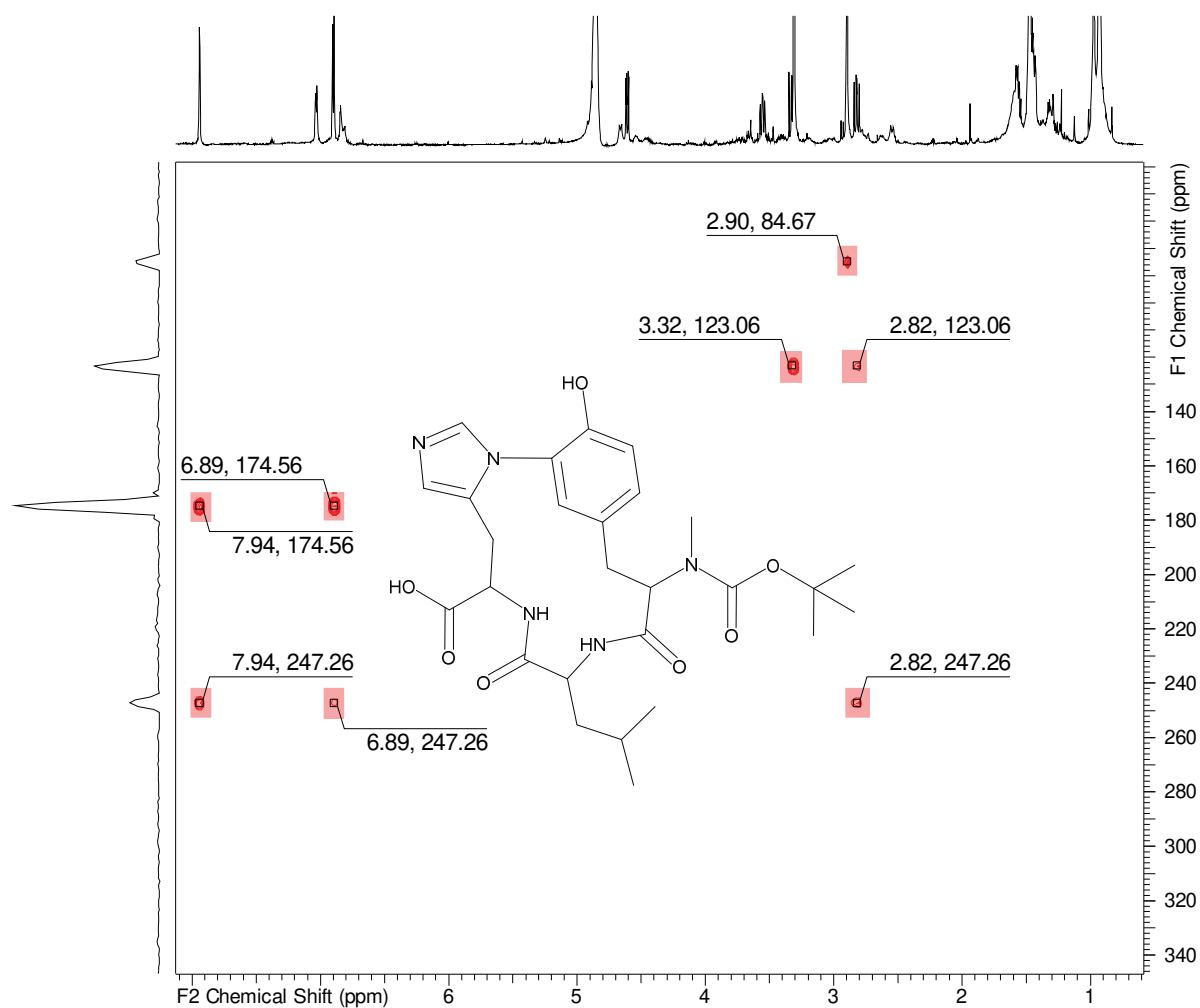
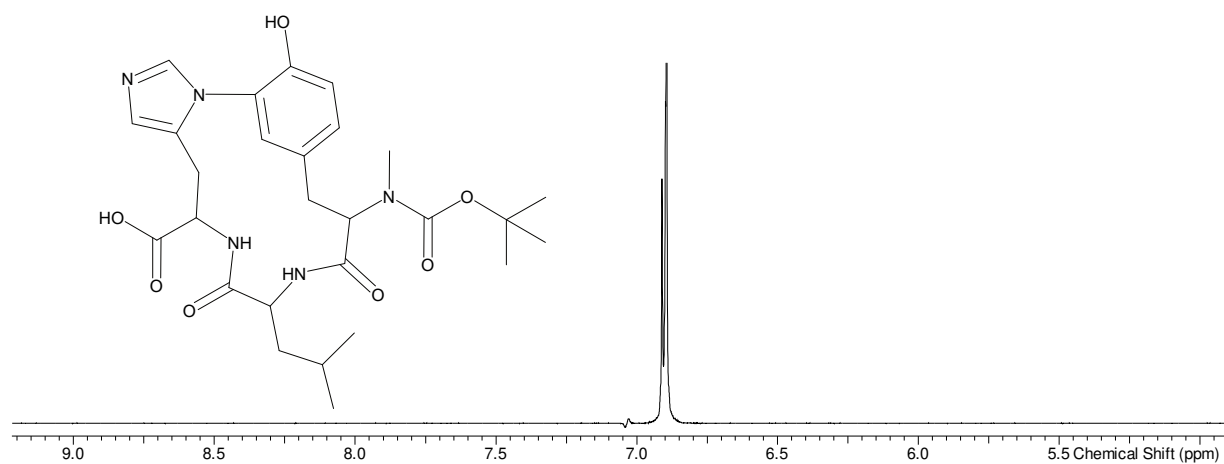
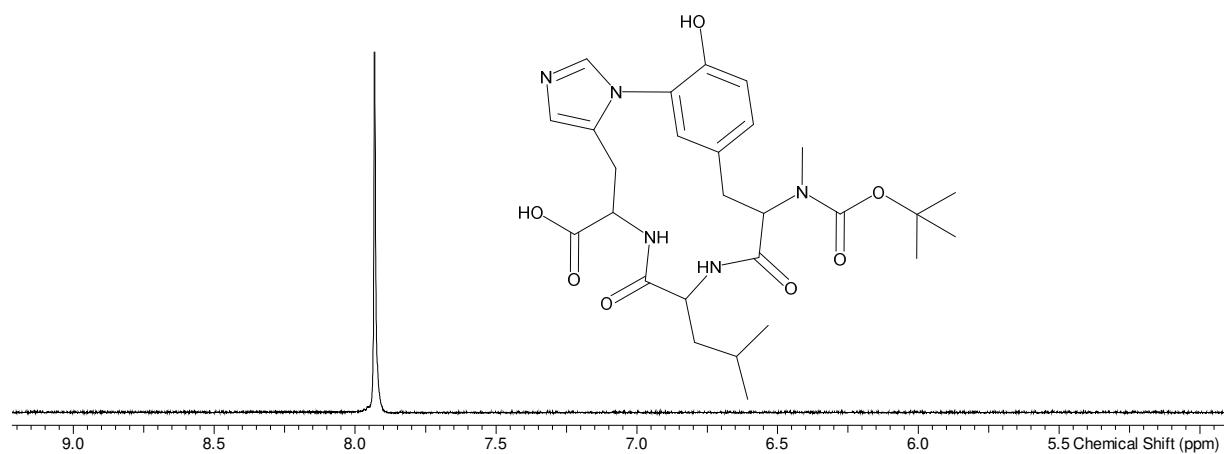


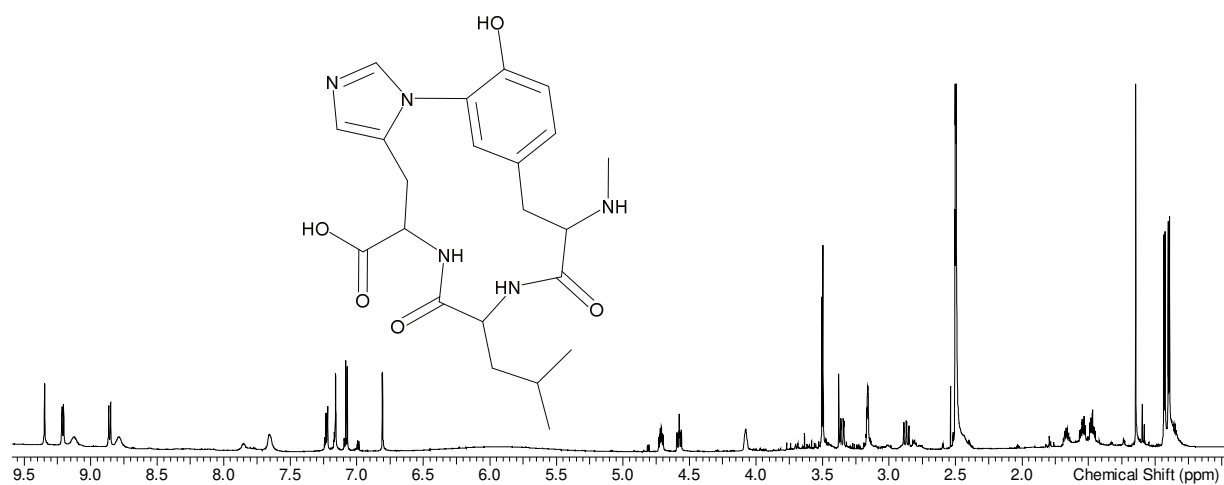
Figure S16.  $^1\text{H}$ - $^{15}\text{N}$  HMBC spectrum of **2** measured in methanol- $d_4$  at 700 ( $^1\text{H}$ ) and 175 ( $^{15}\text{N}$ ) MHz.



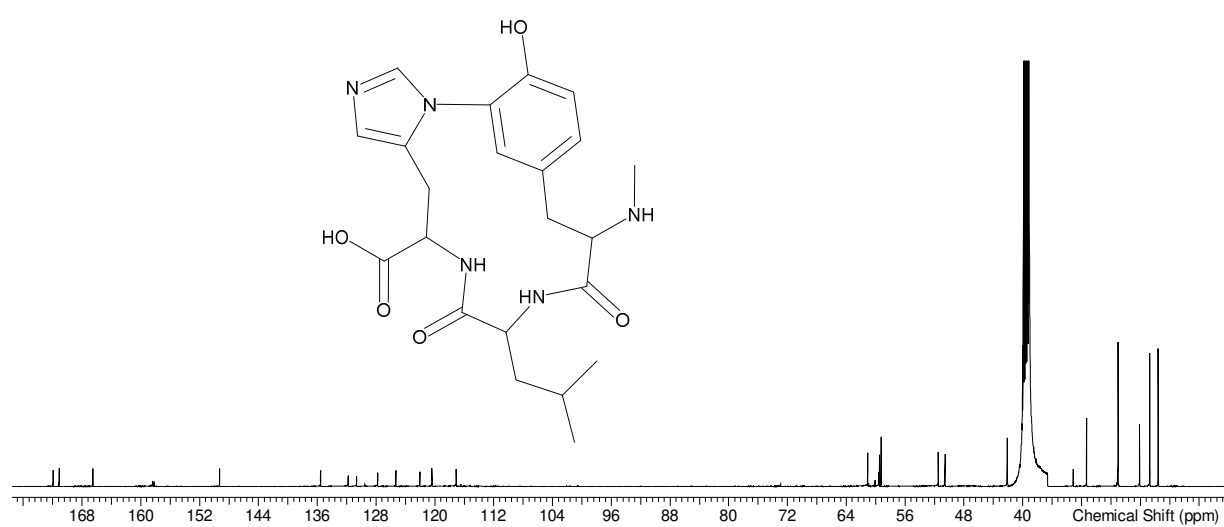
**Figure S17.** Spectrum of **2** exciting at  $\delta(^1\text{H}) = 6.91$  ppm with a distance of 20.32 Hz in methanol- $d_4$  at 700 MHz.



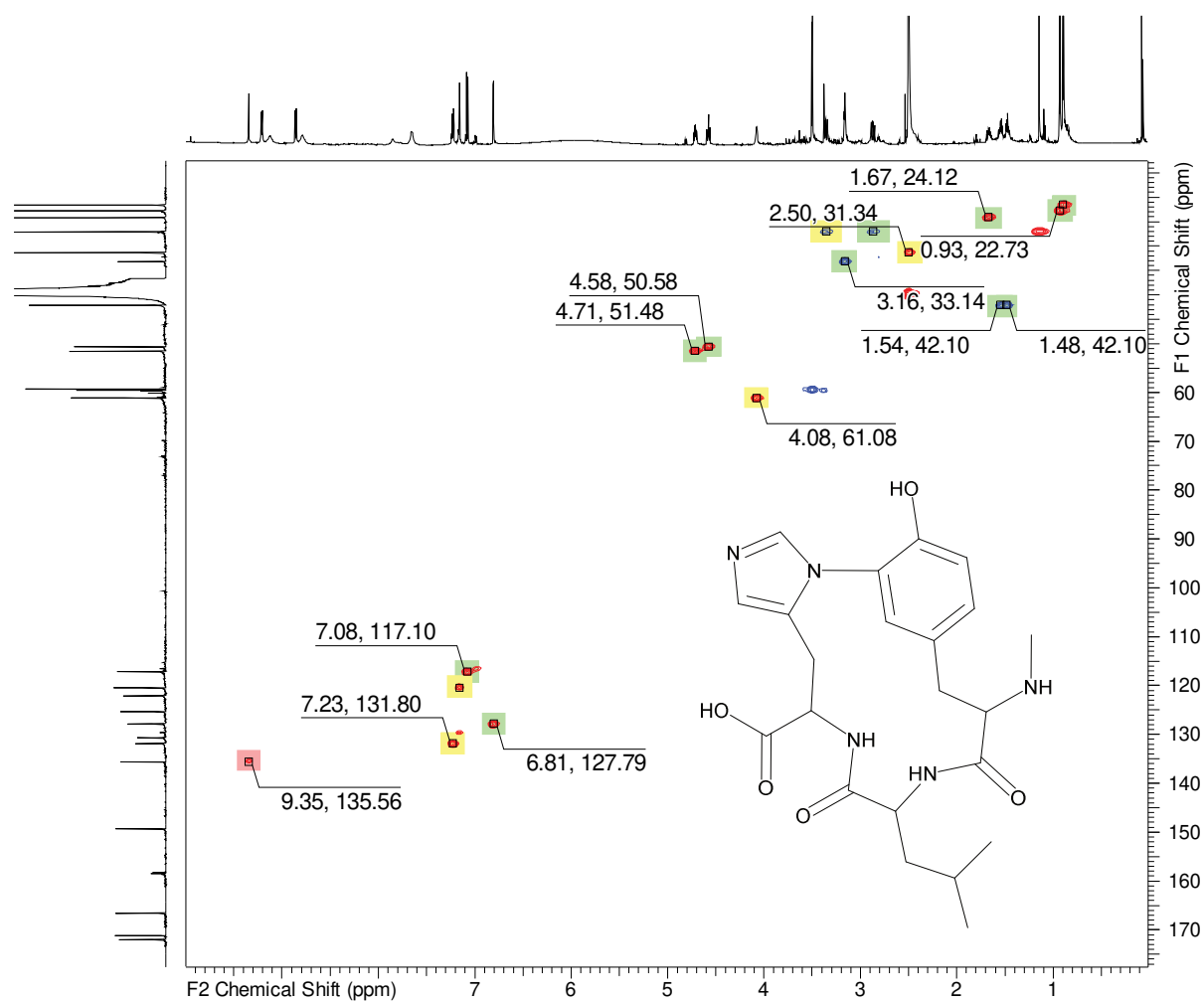
**Figure S18.** Selective 1D NOESY spectrum of **2** exciting at  $\delta(^1\text{H}) = 7.94$  ppm with a distance of 17.20 Hz in methanol- $d_4$  at 700 MHz.



**Figure S19.** <sup>1</sup>H spectrum of **1** measured in DMSO-d<sub>6</sub> at 700 MHz.



**Figure S20.** <sup>13</sup>C spectrum of **1** measured in DMSO-d<sub>6</sub> at 175 MHz.



**Figure S21.** HSQC spectrum of **1** measured in DMSO-*d*<sub>6</sub> at 700 (<sup>1</sup>H) and 175 (<sup>13</sup>C) MHz.

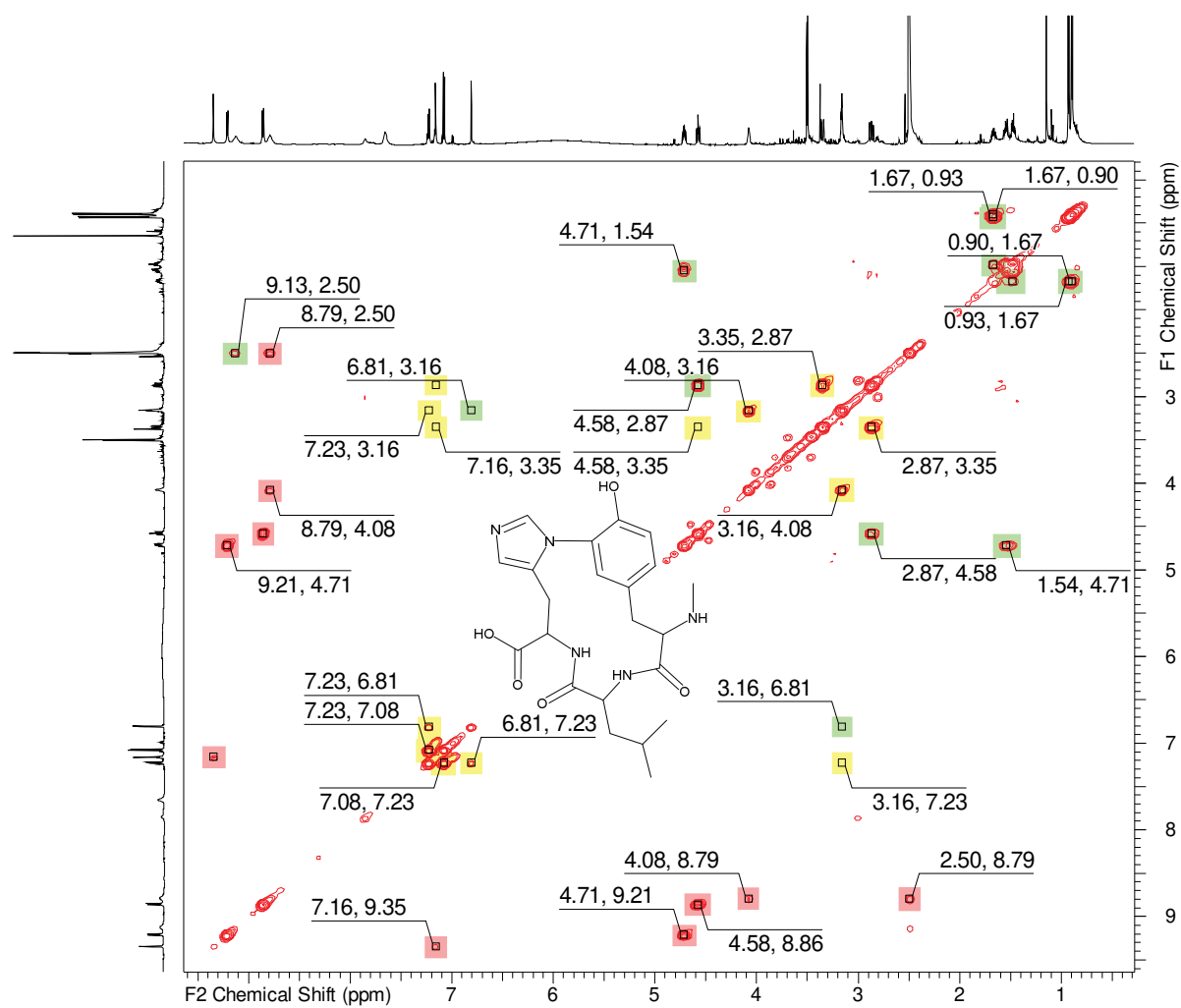


Figure S22. COSY spectrum of **1** measured in DMSO- $d_6$  at 700 MHz.

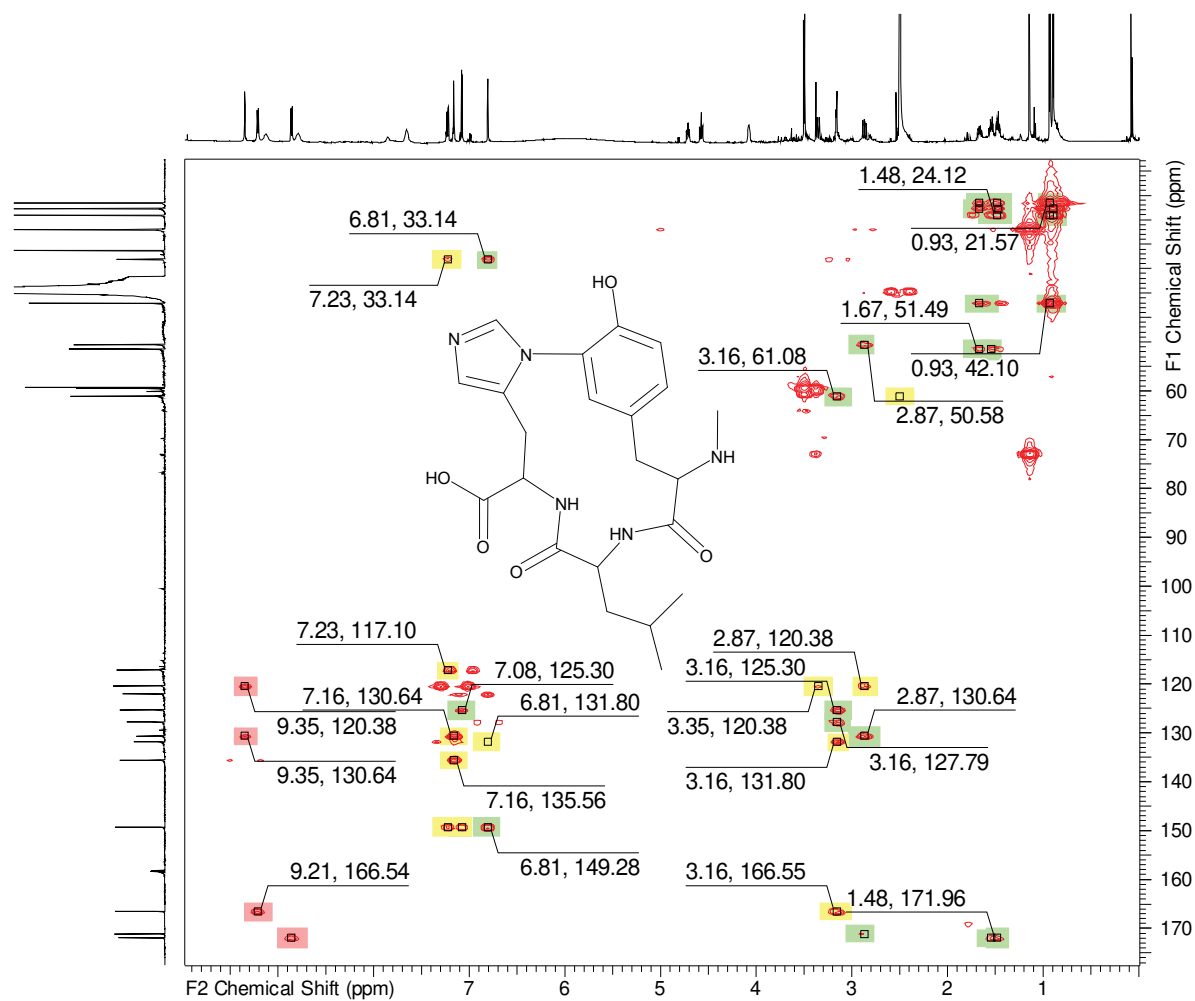
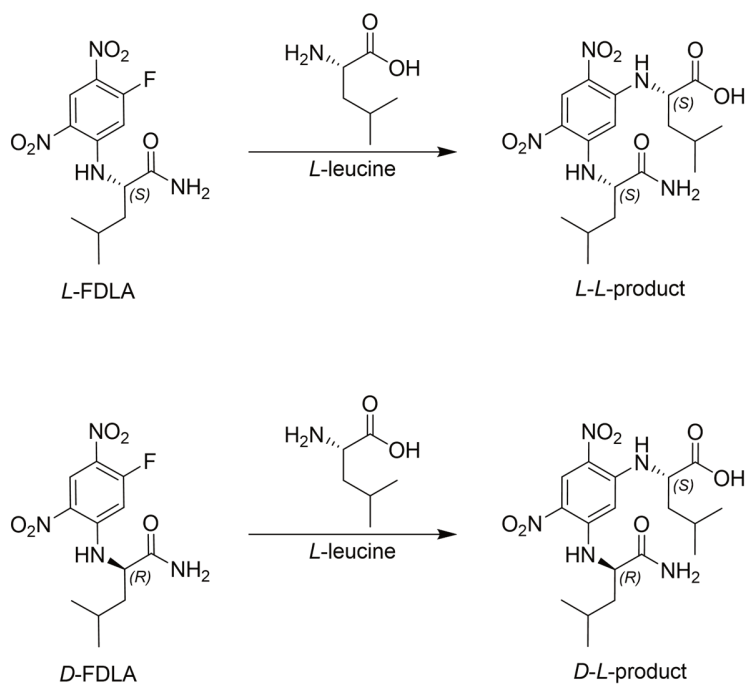


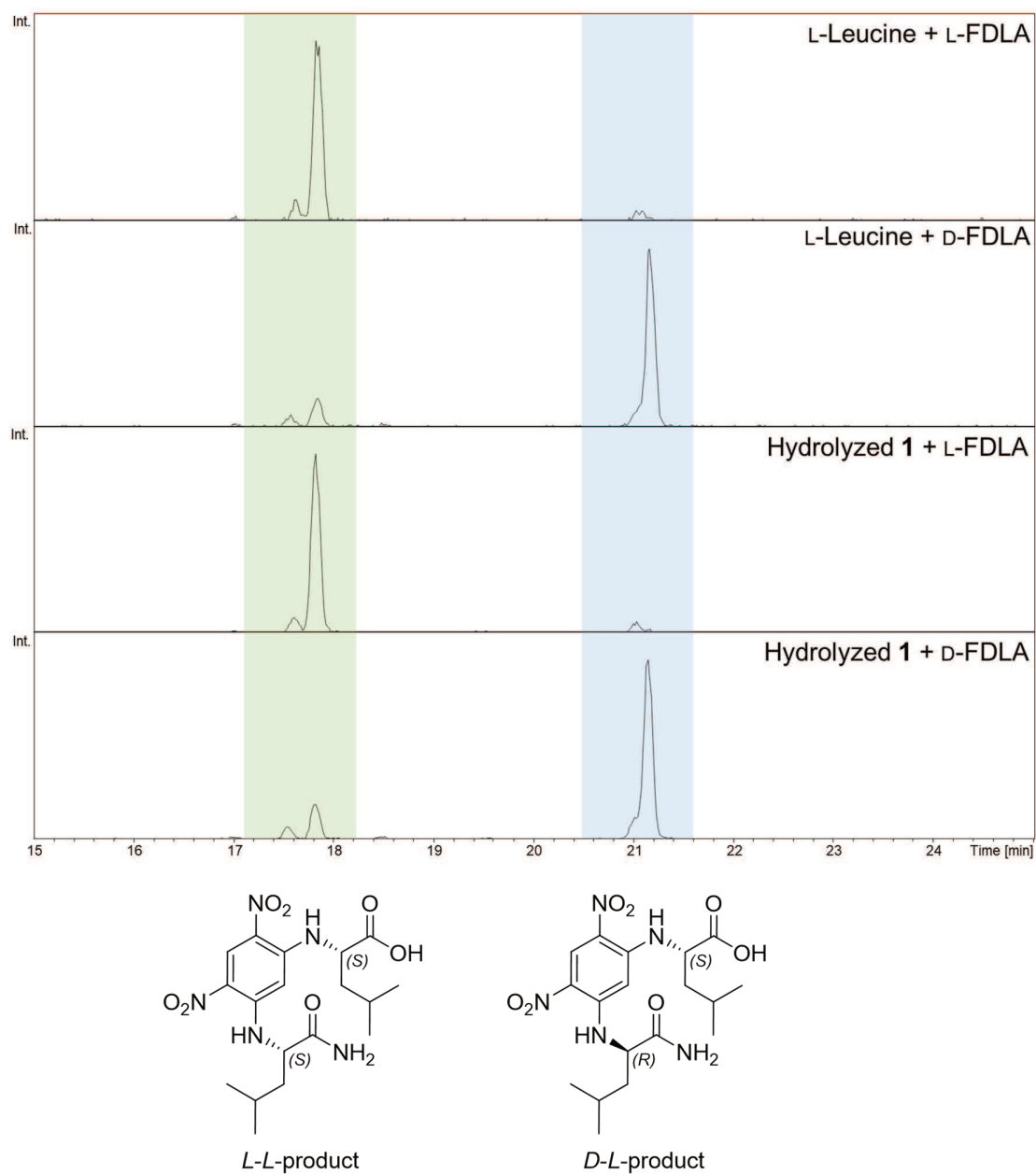
Figure S23. HMBC spectrum of **1** measured in DMSO- $d_6$  at 700 ( $^1H$ ) and 175 ( $^{13}C$ ) MHz.



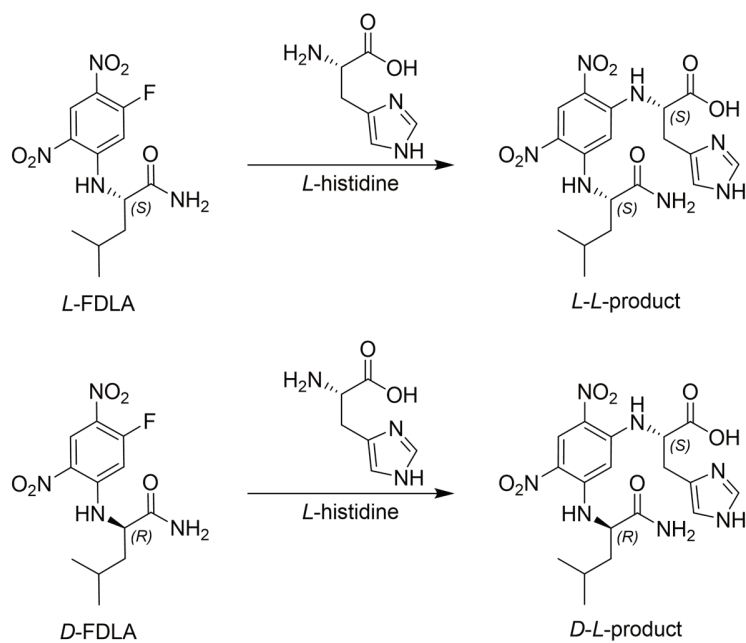
## S 2.3.2.3 Elucidation of the Absolute Stereochemistry



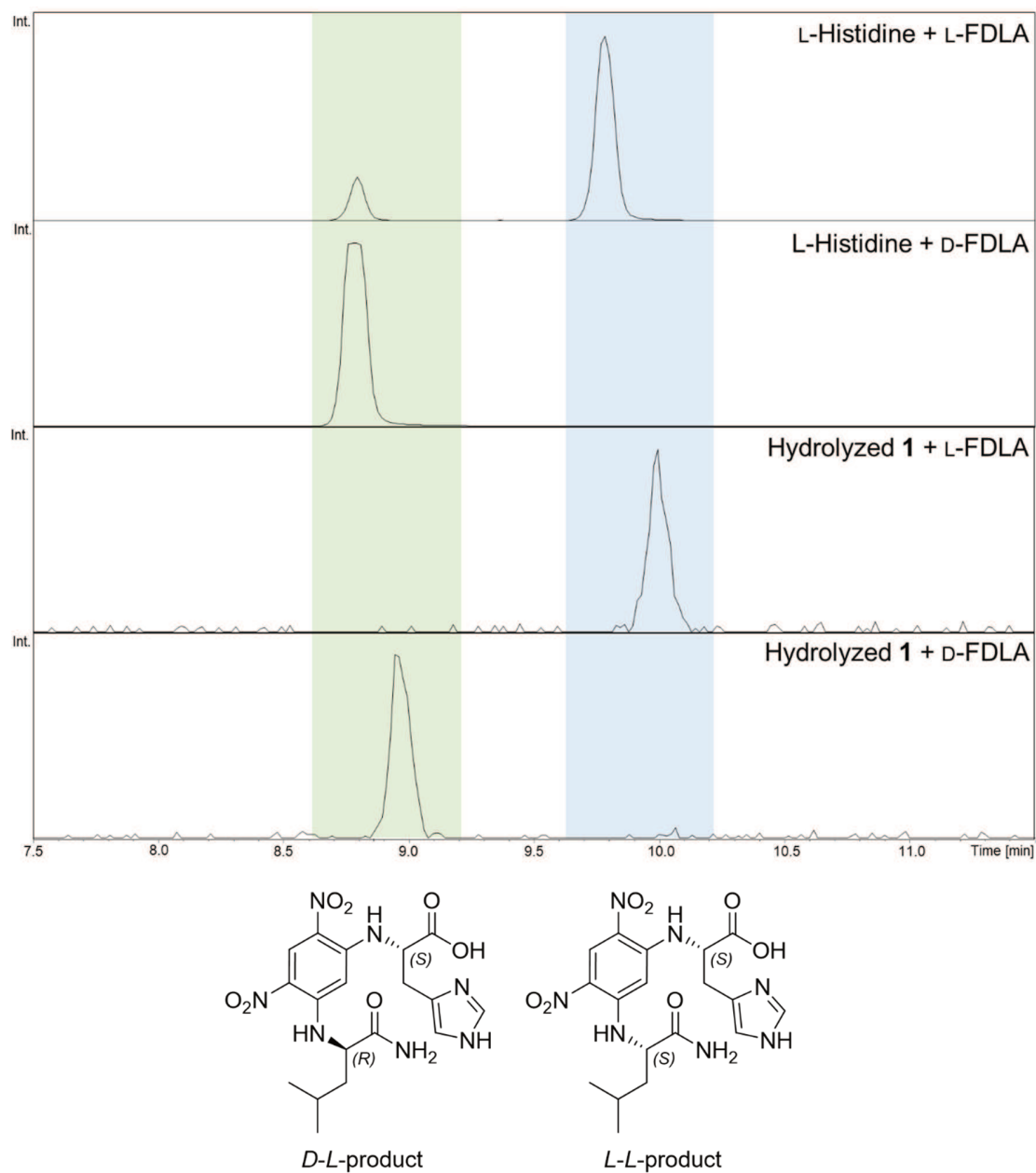
**Figure S24.** Marfey's derivatization reaction of L-leucine standard with L- respective D-(1-fluoro-2,4-dinitrophenyl-5-leucine amide) (FDLA).



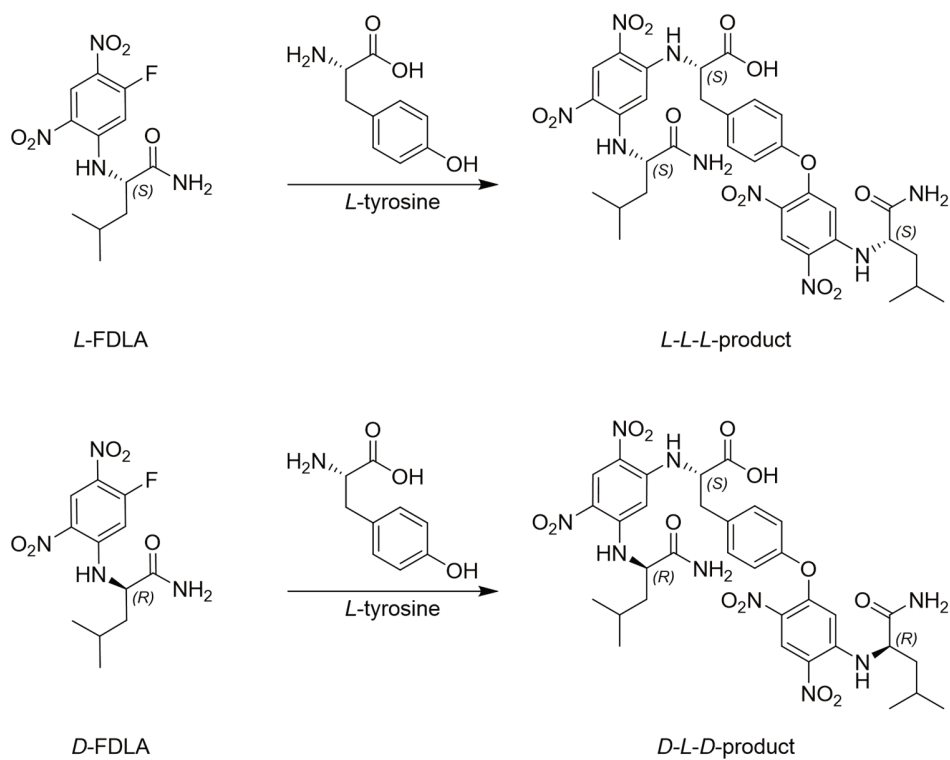
**Figure S25.** HPLC-MS extracted ion chromatograms (EICs) of 426.19833  $m/z$  with a width of 0.005  $m/z$  from Marfey's derivatization reaction with commercial L-leucine standard and hydrolyzed **1**. Highlighted in green: L-L-product peaks; highlighted in blue: D-L-product peaks.



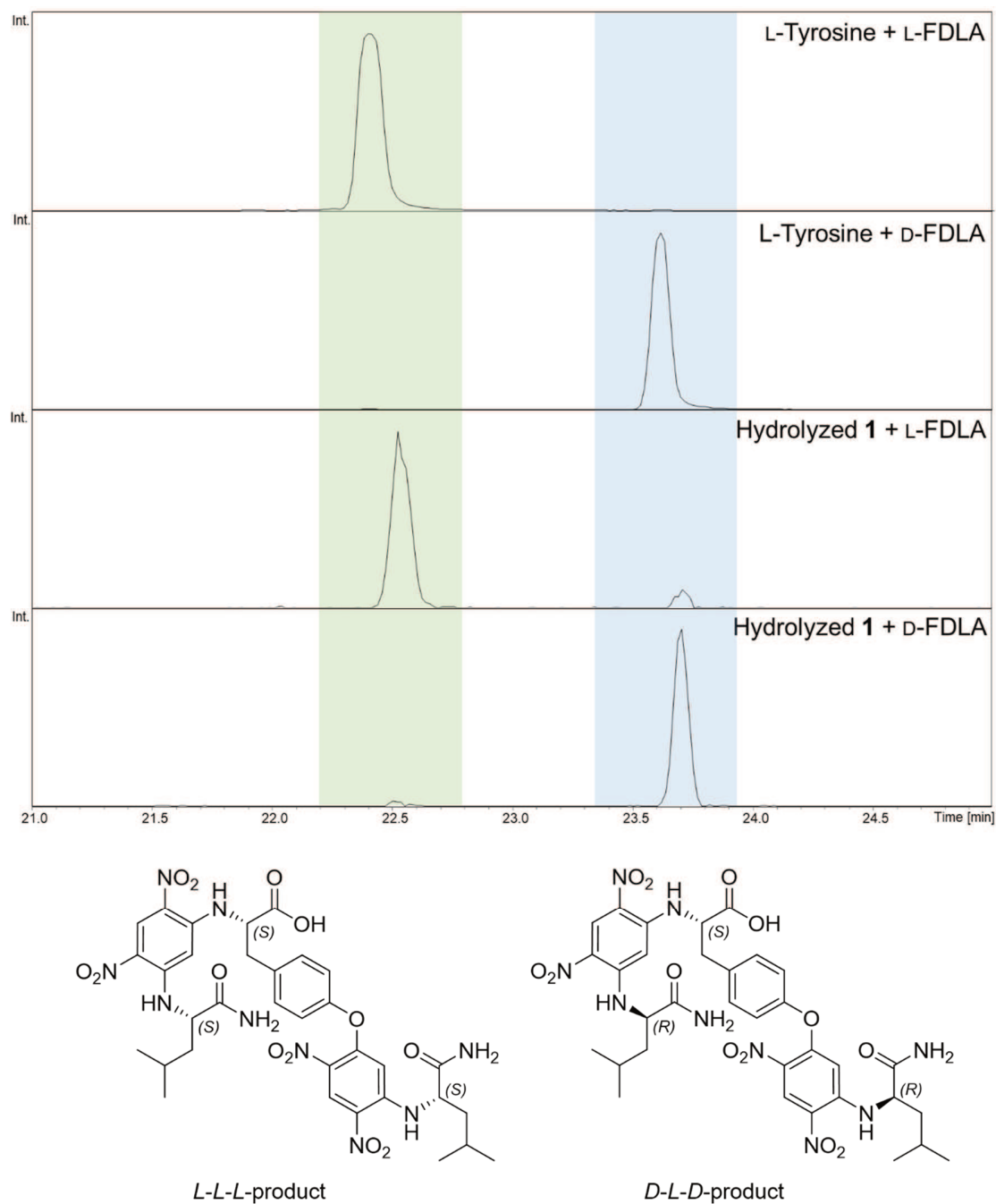
**Figure S26.** Marfey's derivatization reaction of L-histidine standard with L- respective D-(1-fluoro-2,4-dinitrophenyl-5-leucine amide) (FDLA).



**Figure S27.** HPLC-MS extracted ion chromatograms (EICs) of 450.17317  $m/z$  with a width of 0.005  $m/z$  from Marfey's derivatization reaction with commercial L-histidine standard and hydrolyzed **1**. Highlighted in green: D-L-product peaks; highlighted in blue: L-L-product peaks.



**Figure S28.** Marfey's derivatization reaction of L-tyrosine standard with L- respective D-(1-fluoro-2,4-dinitrophenyl)-5-leucine amide) (FDLA). In this case, we mainly observed the double derivatized product as shown above.



**Figure S29.** HPLC-MS extracted ion chromatograms (EICs) of 770.27401  $m/z$  with a width of 0.005  $m/z$  from Marfey's derivatization reaction with commercial L-tyrosine standard and hydrolyzed **1**. Highlighted in green: L-L-L-product peaks; highlighted in blue: D-L-D-product peaks.

## **S 2.4** References

1. Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9*, 4898–4908, doi:10.1039/C8SC01325J.
2. Magrini, V.; Creighton, C.; Youderian, P. Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: genetic elements required for integration. *J. Bacteriol.* **1999**, *181*, 4050–4061, doi:10.1128/jb.181.13.4050-4061.1999.

## Chapter 3

Metabolome-guided Discovery of a  
Family of Light-sensitive Compounds  
from *Corallococcus* sp. MCy9049



# Contributions and Acknowledgments

## Author's effort:

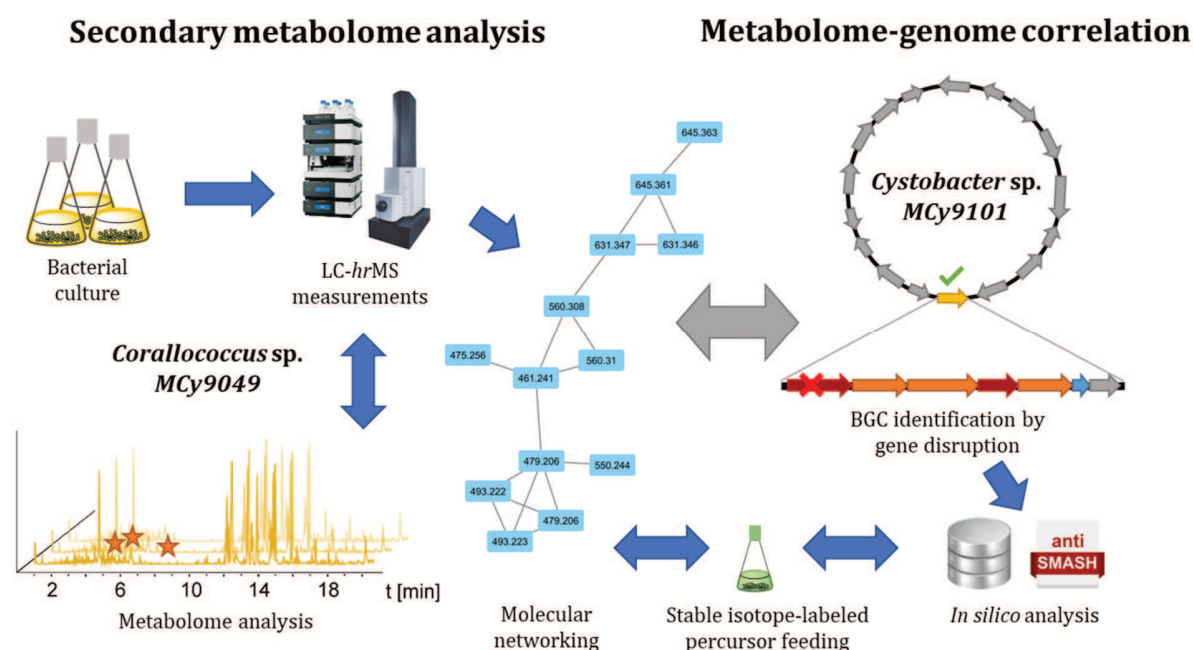
The author significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The laboratory and *in silico* work regarding discovery, isolation, gene cluster analysis and feeding experiments were performed by the author. Furthermore, the author contributed significantly to conceiving and writing this chapter.

## Contributions by others:

Fabian Panter contributed to conception of the study, performed initial cultivation and screening experiments and helped with isolation. Hu Zeng performed genetic engineering experiments and contributed through feeding experiments with labeled methionine in MCy9101. Tatiana Malygina performed de-novo reassembly and improvement of the genome data of MCy9101. Daniel Krug contributed to conception and supervision of the study and editing of this chapter. Rolf Müller was responsible for conception of the project, supervision and proofreading of this chapter.

### 3.1 Abstract

Myxobacteria represent a viable source of natural products featuring unique chemical structures and biosynthetic pathways. Nevertheless, their biosynthetic potential is still vastly underexplored considering the gap between biosynthetic gene clusters identified *in-silico* and actually isolated compounds. In this chapter, we report the metabolome-guided discovery of a family of light-sensitive compounds from the terrestrial myxobacterium *Corallocooccus* sp. MCy9049. Gene cluster inactivation by gene disruption through plasmid integration in the alternative producer strain *Cystobacter* sp. MCy9101 led to the discovery of a cryptic type I PKS-NRPS hybrid biosynthetic gene cluster putatively encoding the biosynthetic machinery of the newly discovered compounds. As isolation of the most abundant derivatives could not provide sufficient amounts for NMR structure elucidation and follow-up experiments, compound characterization was based on tandem MS fragments, feeding of stable-isotope labeled building blocks and *in-silico* analysis of the putative BGC. Despite these efforts, only fragments of the chemical structure of these natural products could be determined and thus the full structure remains elusive so far. Metabolome-mining against our in-house database revealed that the compound class is frequently encountered in the genus *Corallocooccus*, but also appears in other myxobacteria belonging to *Cystobacter* and *Myxococcus* species.



**Graphical abstract.** Analysis of the metabolome of *Corallocooccus* sp. MCy9049 led to the discovery of a family of light-sensitive compounds. The corresponding biosynthetic gene cluster was identified in the alternative producer *Cystobacter* sp. MCy9101. Characterization of the compounds was performed by *in-silico* analysis of the BGC, stable isotope-labeled precursor feeding and LC-hrMS data.

## 3.2 Introduction

Myxobacteria are a phylum of soil dwelling ubiquitously occurring Gram-negative  $\delta$ -proteobacteria that display remarkable characteristics [1]. They show unique social behavioral traits such as coordinated swarming, preying on other microorganisms and formation of macroscopic fruiting bodies [2,3]. Furthermore, myxobacteria are known to have large BGC-rich genomes and have emerged as a fruitful reservoir of secondary metabolites with various chemical scaffolds and promising biological activities [4–6]. However, due to the production of already known bioactive compounds or the lack of bioactivity in initial screenings, many myxobacterial strains are not considered for further follow-up work. Nevertheless, many strains harbor numerous BGCs without assigned compounds.

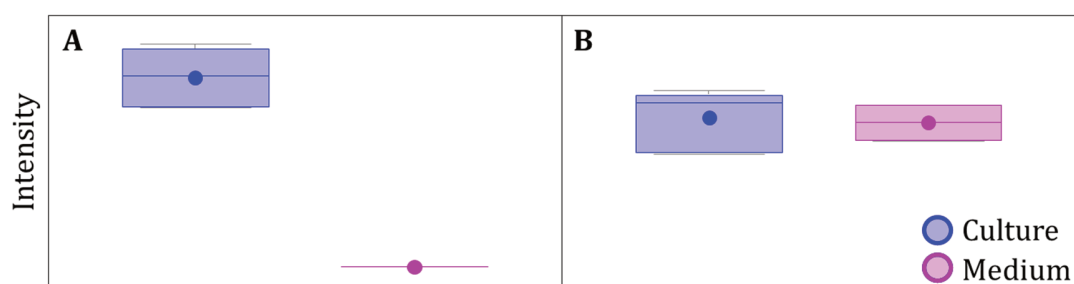
*Corallococcus* sp. MCy9049 is a myxobacterial strain whose extracts showed no bioactivity in an initial extract screening. Still, the strain contains a large unexplored potential as it harbors 40 putative biosynthetic gene clusters in its genome. Apart from the BGCs encoding for the biosynthetic machineries of myxochelins [7], alkylnones [8] and geosmin [9], no natural product could be assigned to the other 37 clusters.

In the following chapter we describe the metabolome-guided discovery, purification and subsequent characterization of the members of a newly identified family consisting of 28 light-sensitive compounds. Furthermore, we report the discovery of the putative biosynthetic gene cluster by gene disruption experiments and present *in-silico* data for each coding DNA sequence in the cluster. Although the final structure of the compounds remains elusive as compound isolation could not provide sufficient amounts for NMR structure elucidation, the data presented in this chapter show that these are promising compounds worthy of further experiments to provide a complete characterization including elucidated structure, biosynthesis hypothesis and antimicrobial activities.

## 3.3 Results

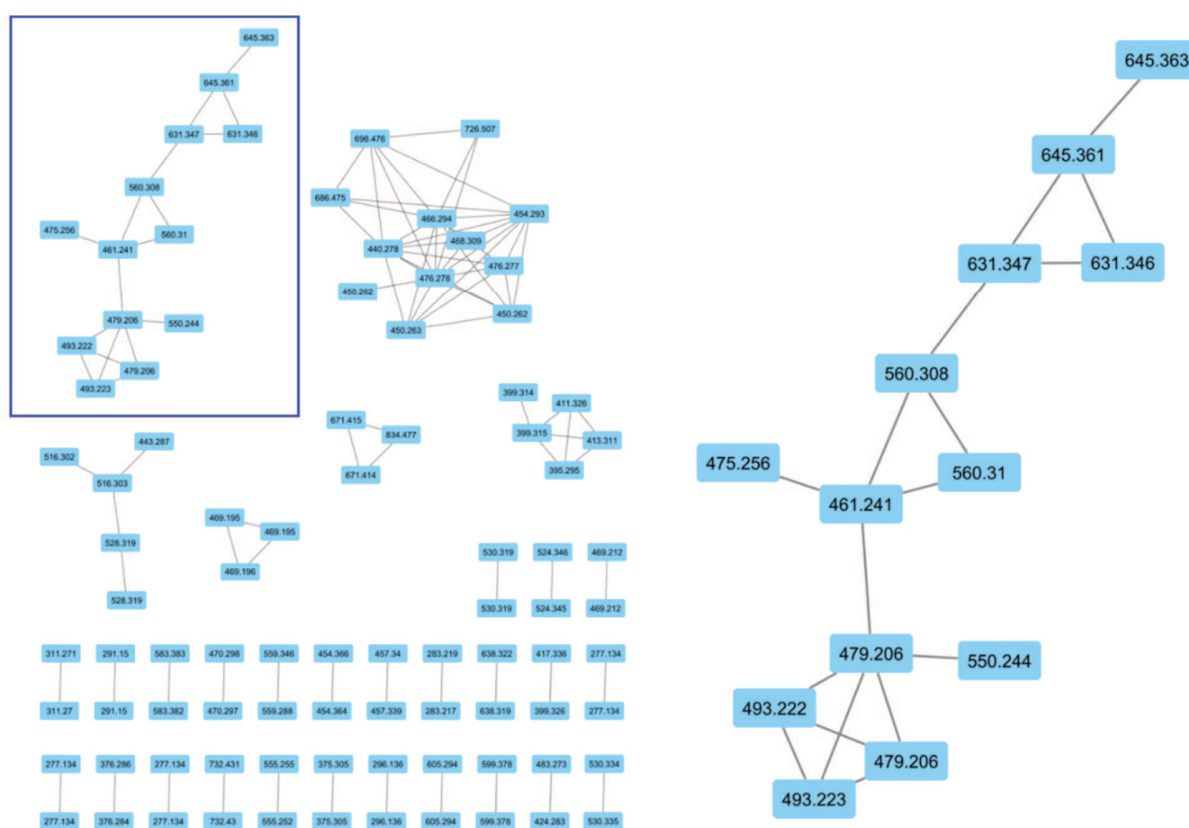
### 3.3.1 Discovery of a Novel Family of Chlorinated Compounds

In metabolomics analyses, structural information about microbial natural products can be obtained at an early stage via MS/MS experiments. Various methods such as spectral networking, fingerprinting or *in-silico* fragmentation tools have been developed for this purpose, with spectral networking being the most prominent method for disclosure of novel natural product scaffolds [10,11]. However, as complex biological samples such as myxobacterial crude extracts contain a large number of abundant medium-derived MS signals automatic precursor selection based on signal intensity in a tandem MS experiment might be hindered. Therefore, an unbiased preselection of relevant features is useful to obtain a set of tandem MS data only consisting of signals that are related to bacterial metabolism. Filtering of precursor ion signals was performed based on comparison of myxobacterial cultures with the respective media blank samples. Both sample types were prepared in triplicates and handled the same way throughout the experiment. In this study, liquid cultures of the myxobacterium *Corallococcus sp.* MCy9049 in VY/2S medium were prepared alongside the respective media blanks and incubated for 10 days. After extraction and dilution, each sample was measured twice by LC-MS resulting in a total of six datasets for both sample types. The resulting datasets were further processed and transformed to a feature table as described in section 3.6.4. The resulting bucket table was filtered to specifically select features that never appear in media blank measurements while being present in all six measurements of bacterial culture extracts. Subsequently, the filtered bucket table was used to create a scheduled precursor list (SPL) for targeted LC-MS/MS analysis of a mixture of the three biological replicates of myxobacterial cultures obtained previously. While this approach distinctly reduced the size of the resulting dataset, it also enabled acquisition of tandem MS data for bacterial secondary metabolites which overlap with abundant medium signals.



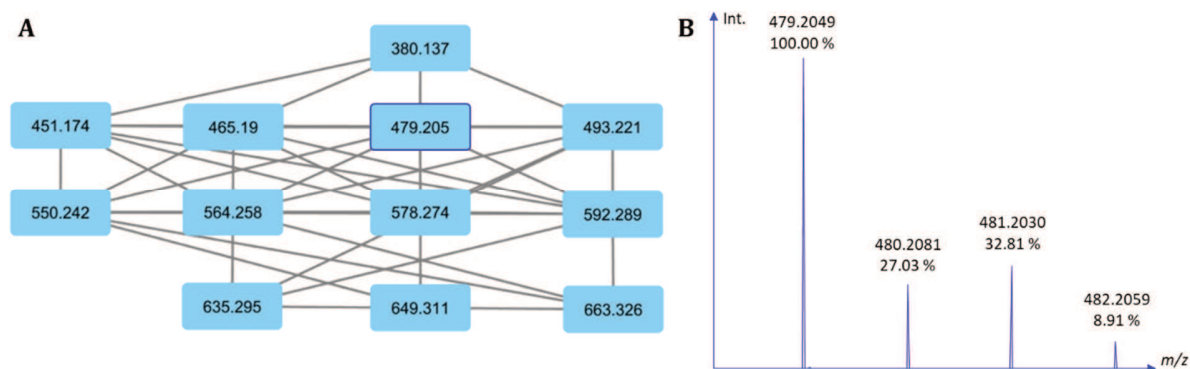
**Figure 1.** Intensity box-and-whisker diagrams for a feature specific for the myxobacterial culture (A) and a media-derived feature (B). The circles represent the mean intensities, the lines represent the median intensities, the boxes range from the 75% quartile (upper) to the 25% quartile (lower) with whiskers indicating upper and lower extremes.

Most mycobacterial secondary metabolites are not single compounds, but rather whole compound families of multiple derivatives sharing the same backbone structure. Molecular networking is a bioinformatics tool to visualize and annotate mass spectrometry data and organize related molecules into ‘molecular families’ based on their fragmentation spectra [12]. With this in hand, it is possible to discover putative novel secondary metabolites by similarity clustering of MS/MS spectra obtained from crude bacterial extracts. To perform such similarity clustering, the MS/MS chromatograms obtained by UHPLC-SPL-MS/MS as described before were exported, submitted to the GNPS online workflow [11] and the resulting spectral networks were visualized by Cytoscape 3.6.1 [13].



**Figure 2.** Spectral network of MCy9049 in VY/2S medium with highlighted molecular target family.

The spectral network of the MS/MS spectra acquired from the crude extract of MCy9049 in VY/2S medium shown above revealed a molecular family of novel compounds which were chosen for further analyses. Manual inspection of the mass spectra with special focus on isotope patterns and main fragments raised the assumption that the compound family consists of peptides, which are partly chlorinated. Chlorine only has two stable isotopes,  $^{35}\text{Cl}$  and  $^{37}\text{Cl}$  with a natural abundance of 75.77 % and 24.23 %, respectively [14], which results in a characteristic intensity ratio of  $\frac{0.2423}{0.7577} = 0.32$  between the first and third isotope peak of monochlorinated compounds (Figure 3B). In order to identify more members of the target compound family, isotope cluster analysis was performed in DataAnalysis 4.4 (Bruker Daltonics, Billerica, MA, USA) with intensity ratio of 0.32 (tolerance 10 %) and exact mass difference of 2.00 Da (tolerance 0.1 Da). The resulting features were used to create a second SPL for targeted LC-MS/MS analysis and the resulting chromatogram was once more exported and submitted to the GNPS workflow resulting in the cluster shown below in Figure 3A.



**Figure 3.** (A) Spectral network of monochlorinated compounds from extracts of MCy9049. (B) Characteristic isotope pattern of monochlorinated compounds as exemplified by the MS<sup>1</sup> spectrogram of 5. Percentages represent the relative abundance of the respective signal.

The high similarity of fragmentation spectra as well as retention times and predicted sum formulas reinforced the hypothesis that the clustered features belong to a family of related secondary metabolites. As every horizontal line in the spectral network presented above shows shifts of 14 Da and all compounds are located in a narrow retention time window, manual inspection led to the discovery of two additional features with  $m/z$  of 394.153 and 621.280, respectively. From their fragmentation patterns the different features were further divided into two subgroups depending on whether the features show a neutral loss fragment specific for valine or isoleucine (Table 1). All library searches for the same or similar natural products based on exact mass, calculated sum formulae and fragmentation patterns showed no results, so it is assumed that this molecular family consisting of monochlorinated peptides indeed is a novel family

of myxobacterial secondary metabolites. Interestingly, the upper part of the spectral family displayed in Figure 2 was found to represent a second homologous set of compounds that feature putative hydroxylation instead of chlorination. Again, manual inspection of the LC-MS data led to the discovery of total 14 representatives that are summarized in Table 2.

**Table 1.** Family of newly discovered monochlorinated compounds from MCy9049.

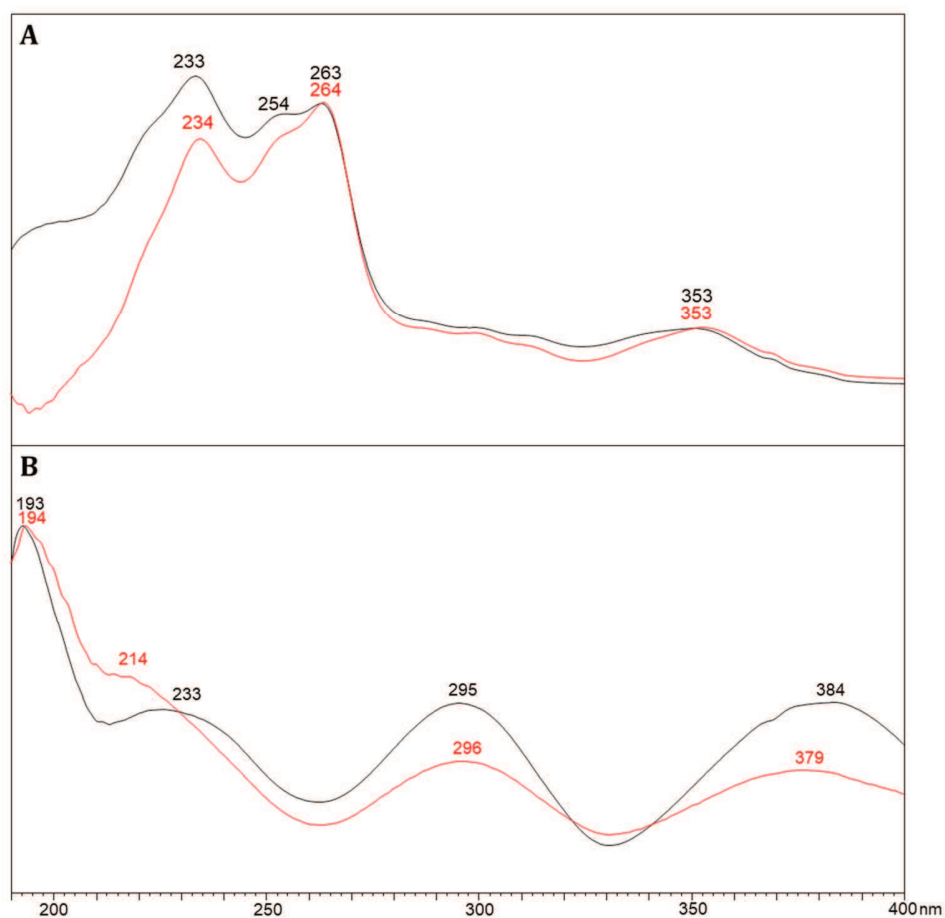
#	$m/z$ [M + H] <sup>+</sup>	Retention times [min]	Predicted sum formula [M + H] <sup>+</sup>	Subgroup
1	380.137	6.78 / 7.37	C <sub>18</sub> H <sub>23</sub> ClN <sub>3</sub> O <sub>4</sub>	Valine
2	394.153	7.47	C <sub>19</sub> H <sub>25</sub> ClN <sub>3</sub> O <sub>4</sub>	Isoleucine
3	451.174	6.32 / 6.94	C <sub>21</sub> H <sub>28</sub> ClN <sub>4</sub> O <sub>5</sub>	Valine
4	465.190	7.00 / 7.50	C <sub>22</sub> H <sub>30</sub> ClN <sub>4</sub> O <sub>5</sub>	Isoleucine
5	479.206	7.23 / 7.89	C <sub>23</sub> H <sub>32</sub> ClN <sub>4</sub> O <sub>5</sub>	Valine
6	493.221	7.77 / 8.43	C <sub>24</sub> H <sub>34</sub> ClN <sub>4</sub> O <sub>5</sub>	Isoleucine
7	550.242	6.94 / 7.53	C <sub>26</sub> H <sub>37</sub> ClN <sub>5</sub> O <sub>6</sub>	Valine
8	564.258	7.49 / 8.09	C <sub>27</sub> H <sub>39</sub> ClN <sub>5</sub> O <sub>6</sub>	Isoleucine
9	578.274	7.84 / 8.42	C <sub>28</sub> H <sub>41</sub> ClN <sub>5</sub> O <sub>6</sub>	Valine
10	592.289	8.36 / 8.43 / 8.95	C <sub>29</sub> H <sub>43</sub> ClN <sub>5</sub> O <sub>6</sub>	Isoleucine
11	621.280	6.88 / 7.43	C <sub>29</sub> H <sub>42</sub> ClN <sub>6</sub> O <sub>7</sub>	Valine
12	635.295	7.39 / 7.44 / 8.00	C <sub>30</sub> H <sub>44</sub> ClN <sub>6</sub> O <sub>7</sub>	Isoleucine
13	649.311	7.49 / 7.57 / 8.08	C <sub>31</sub> H <sub>46</sub> ClN <sub>6</sub> O <sub>7</sub>	Valine
14	663.326	8.00 / 8.11 / 8.59	C <sub>32</sub> H <sub>48</sub> ClN <sub>6</sub> O <sub>7</sub>	Isoleucine

**Table 2.** Family of newly discovered putatively hydroxylated compounds from MCy9049.

#	$m/z$ [M + H] <sup>+</sup>	Retention times [min]	Predicted sum formula [M + H] <sup>+</sup>	Subgroup
15	362.171	4.34 / 4.90	C <sub>18</sub> H <sub>24</sub> N <sub>3</sub> O <sub>5</sub>	Valine
16	376.187	5.00 / 5.16	C <sub>19</sub> H <sub>26</sub> N <sub>3</sub> O <sub>5</sub>	Isoleucine
17	433.208	4.06 / 4.63	C <sub>21</sub> H <sub>29</sub> N <sub>4</sub> O <sub>6</sub>	Valine
18	447.224	4.74 / 5.26	C <sub>22</sub> H <sub>31</sub> N <sub>4</sub> O <sub>6</sub>	Isoleucine
19	461.239	4.95 / 5.59	C <sub>23</sub> H <sub>33</sub> N <sub>4</sub> O <sub>6</sub>	Valine
20	475.255	5.55 / 6.18	C <sub>24</sub> H <sub>35</sub> N <sub>4</sub> O <sub>6</sub>	Isoleucine
21	532.277	4.78 / 5.37	C <sub>26</sub> H <sub>38</sub> N <sub>5</sub> O <sub>7</sub>	Valine
22	546.292	5.41 / 5.96	C <sub>27</sub> H <sub>40</sub> N <sub>5</sub> O <sub>7</sub>	Isoleucine
23	560.308	5.76 / 6.32	C <sub>28</sub> H <sub>42</sub> N <sub>5</sub> O <sub>7</sub>	Valine
24	574.324	6.36 / 6.89	C <sub>29</sub> H <sub>44</sub> N <sub>5</sub> O <sub>7</sub>	Isoleucine
25	603.314	4.83 / 5.41	C <sub>29</sub> H <sub>43</sub> N <sub>6</sub> O <sub>8</sub>	Valine
26	617.329	5.35 / 5.61 / 6.14	C <sub>30</sub> H <sub>45</sub> N <sub>6</sub> O <sub>8</sub>	Isoleucine
27	631.345	5.47 / 6.03 / 6.71	C <sub>31</sub> H <sub>47</sub> N <sub>6</sub> O <sub>8</sub>	Valine
28	645.361	6.00 / 6.37 / 6.91	C <sub>32</sub> H <sub>49</sub> N <sub>6</sub> O <sub>8</sub>	Isoleucine

In the first experiments, it was already noticeable that the compounds of interest tend to isomerize and fragment. A short stability assay revealed that isomerization is mainly acid-induced, while compound degradation is induced by light (Section 3.6.5). From the shift in retention time as well as the different UV absorption maxima for different isomers, it was suspected that the isomerization might be an acid catalyzed cis-trans-isomerization of a double bond (Figure 4). In order to suppress both processes that effectively hinder compound purification and follow-up experiments, catalytic hydrogenation using hydrogen and palladium over activated carbon was performed on the most abundant compound **5**. Although the unknown core scaffold, which comprises the chromophore, could successfully be hydrogenated to yield a dihydro derivative of  $m/z = 481.222$  and a tetrahydro derivative of  $m/z = 483.237$ , neither product appeared to be protected from isomerization or fragmentation.



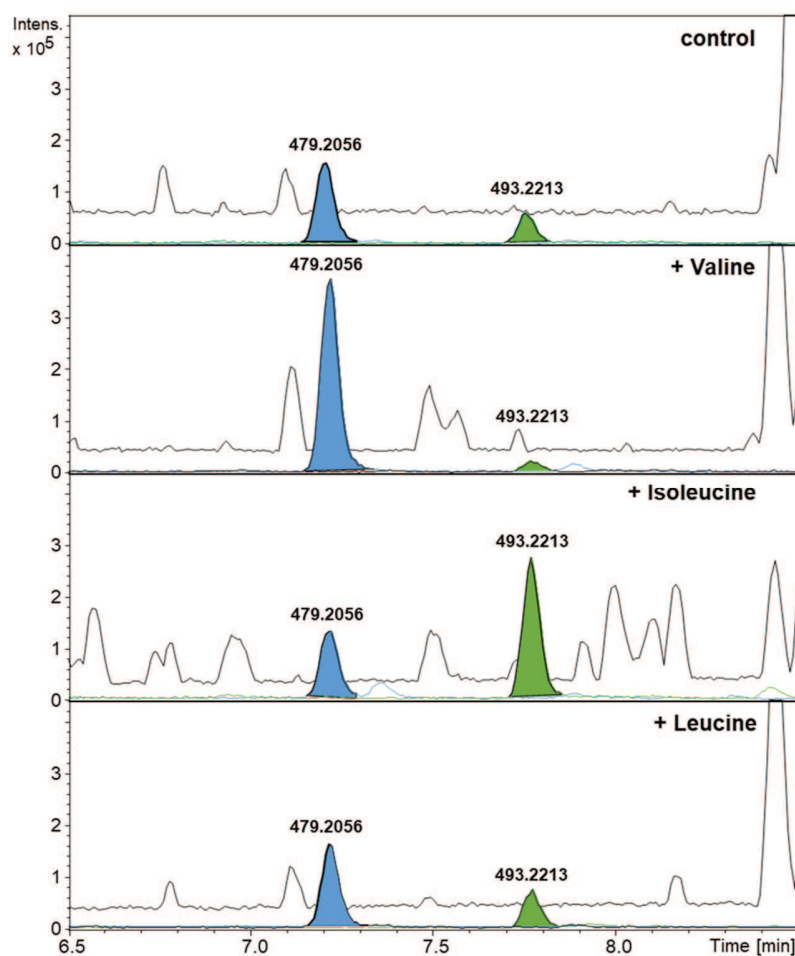


**Figure 4.** UV/Vis spectra of **5** (black) and **6** (red) in water/acetonitrile mixture with 0.1% formic acid. (A) Most abundant “native” isomer (RT(**5**) = 7.23 min, RT(**6**) = 7.77 min). (B) Most abundant isomer after acid-induced isomerization (RT(**5**) = 7.89 min, RT(**6**) = 8.43 min).

### 3.3.2 Purification of the Most Abundant Compounds **5** and **6**

Previous analyses performed by Fabian Panter already demonstrated that growing the strain *Corallococcus* sp. MCy9049 in VY/2S medium yields the highest amounts of the novel chlorinated compounds among the basic media tested. Based on that knowledge, further media optimization was achieved by the addition of amino acids and vitamins. In total eight different additives were tested and the peak area of the two most abundant peaks with the masses 479.206 (**5**) and 493.221 (**6**) was compared to the respective peak area in the control (Table S5). All additives that improved the production of either of the compounds were mixed to create the production media entitled 479 and 493 (Tables S2-S3). Notable is that particularly the incorporation of valine shifted the balance between both derivatives towards compound **5** while the addition of isoleucine lead to an improved production of compound **6**. Leucine incorporation however had no effect at all indicating leucine is not present in the metabolites under investigation. This was needed as an additional experiment as leucine and isoleucine cannot be

distinguished by LC-MS/MS since they are isomeric, while also showing isomeric MS/MS fragments. Overall, the media optimization led to a 2.5-fold increase of peak area for compound **5** and a 5-fold increase for compound **6**. These results suggest that valine is the preferred amino acid to be incorporated in the molecule as even in the isoleucine enriched medium the production was comparable to the production in the control whereas the production of compound **6** drops to almost zero in a valine enriched medium.



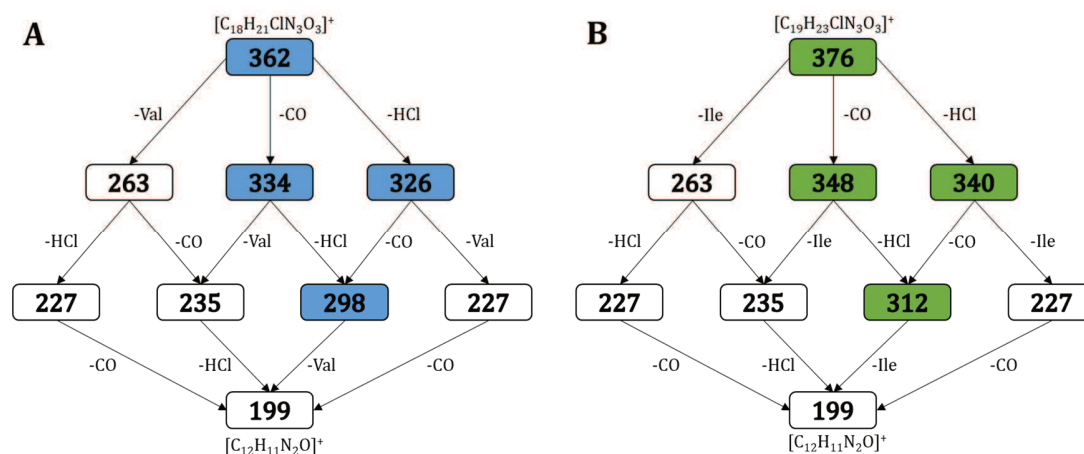
**Figure 5.** Changes in the production of compounds **5** and **6** by the addition of valine, isoleucine or leucine to the production medium. Black: BPC 150 – 2500 m/z, blue: EIC 479.2 m/z, green: EIC 493.2 m/z.

To isolate the secondary metabolites **5** and **6**, *Corallocooccus* sp. MCy9049 was cultivated in 479 and 493 medium in larger-scale batch fermentation. Purification of **5** and **6** was achieved by liquid/liquid partitioning of the methanol extract of the culture followed by semi-preparative HPLC with MS detection. All steps, including cultivation, were carried out in absence of light and if that was not possible, e.g. for

liquid/liquid partitioning, indirect red light of 660 nm was used. Unfortunately, although pure compound was eventually obtained, the yields were negligible and thus insufficient for NMR structure elucidation.

### 3.3.3 Characterization of this Compound Class

Since structure elucidation by NMR spectroscopy was impeded by insignificant compound yields, insights on the structure could only be obtained by different approaches. First of all, the sum formula for compound **5** was predicted and confirmed correct by comparison of the isotopic fine structure from a direct infusion experiment on a Bruker solarix XR FTICR-MS system. Furthermore, an MS<sup>n</sup> multistep fragmentation experiment was conducted with five consecutive steps of precursor isolation and in-cell fragmentation with increasing collision energy to build a fragmentation tree (Figure 6A). The product ion  $m/z$  199.087 (C<sub>12</sub>H<sub>11</sub>N<sub>2</sub>O<sup>+</sup>) that is the unknown core structure of the molecule further generated several smaller fragments for which no unambiguous molecular formula could be determined. As most of the fragments of **5** were already observed in the previously recorded SPL-MS/MS spectra (Section 3.3.1) and largely comparable with the spectra of **6**, a further tree was constructed analogously from the SPL-MS/MS spectra of **6** (Figure 6B). Each of the fragments depicted could be observed for all monochlorinated derivatives of the corresponding subfamily.



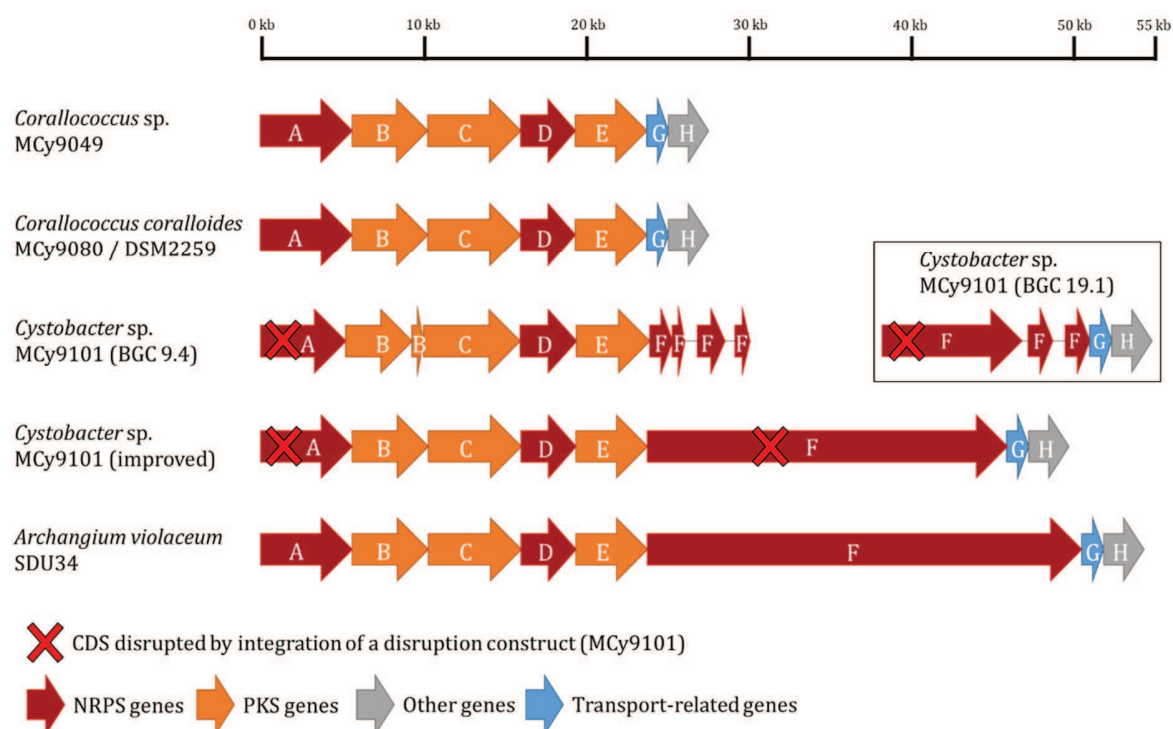
**Figure 6.** Fragmentation trees of **5** (A) generated from FT-ICR in-cell MS<sup>n</sup> experiments and **6** (B) generated from SPL-MS/MS data. Larger fragments as shown were not observed. The difference of the largest fragment to the respective parent mass fits H<sub>2</sub>O + C<sub>5</sub>H<sub>9</sub>NO (Valine) = 117.08 Da in both cases. Highlighted in blue: Fragments specific for compounds of the valine subfamily; highlighted in green: Fragments specific for compounds of the isoleucine subfamily. The number in the boxes represents the observed  $m/z$  of each fragment.

From the fragmentation spectra, it can be concluded that the structures share a common core structure of  $m/z$  = 263.06 (C<sub>13</sub>H<sub>12</sub>ClN<sub>2</sub>O<sub>2</sub><sup>+</sup>), featuring nine double bond equivalents, that is substituted by either two valine (**5**) or one valine and one isoleucine (**6**) moiety. In order to gather more information

about the structure and biosynthesis of the unknown core scaffold as well as to prove the incorporation of valine and isoleucine, a feeding study using 23 different stable-isotope labeled precursors was performed (Table S6). From this broad set of building blocks, for compound **5**, incorporation was only observed for  $^{13}\text{C}_5\text{-}^{15}\text{N}_1$ -methionine (mass shift +5) and d8-valine (mass shift +16), for compound **6**,  $^{13}\text{C}_5\text{-}^{15}\text{N}_1$ -methionine (mass shift +5), d8-valine (mass shift +8) and d10-isoleucine (mass shift +10) were incorporated. Interestingly, the larger compounds, *e.g.* compounds **13** and **14** incorporated multiple valine molecules (**13**: 3, **14**: 2), but always only one or no isoleucine (**13**: 0, **14**: 1). In addition, all compounds showed a uniform mass shift of +5 Da when fed with labeled methionine, suggesting that methionine is somehow used as a building block for the unknown core structure.

### 3.3.4 Metabolome-assisted Genome Analysis

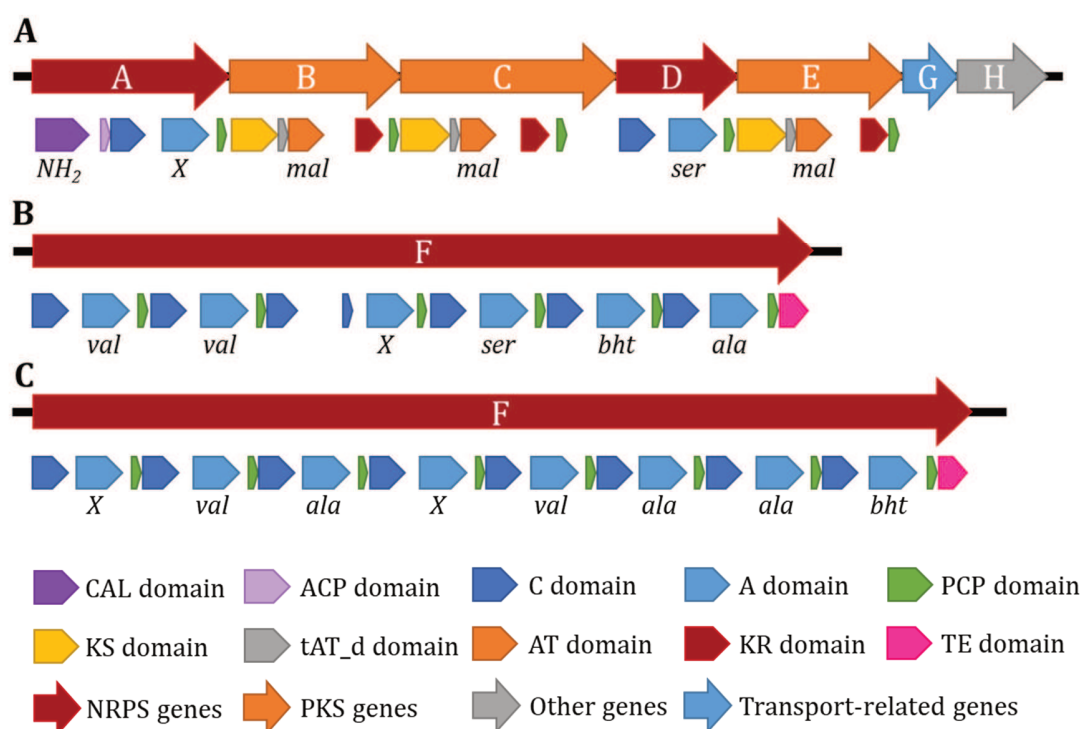
In order to identify alternative producers of the compound family a metabolome mining in our in-house metabolome database was performed. After manual control of each hit, 50 alternative producers of compounds **5** and **6** were identified of which 44 strains are classified in the genus *Coralloccoccus*, 5 in *Cystobacter* and 1 in *Myxococcus*. Besides the original producer strain MCy9049 five other strains had a full genome sequence available in our in-house genome database. One of the strains, *Cystobacter* sp. MCy9101, had previously been extensively characterized by Hu Zeng, also giving access to cluster inactivation mutants for 38 of the 50 BGCs of the strain [15]. The mutants were created by integrating a disruption construct with a kanamycin resistance marker into key biosynthesis genes by single crossover to inactivate the biosynthesis cluster. Cultivation of these KO mutants in triplicates and metabolome analysis of each mutant for the three most abundant compounds in the wild type, namely **5**, **6** and **19**, revealed two target gene clusters named BGC 9.4 and BGC 19.1, where the production of all three compounds was completely abolished, while in all other cluster inactivation mutants, at least two out of three replicates produced the target compounds at levels comparable to the wild type. De novo reassembly of the original Illumina paired-end reads using shovill [16] and subsequent improvement of the assembly with the genome sequence of the closely related strain *Archangium violaceum* SDU34 (GenBank: CP069338) using ragout [17] suggested that the two biosynthetic gene clusters, both located on contig edges, are probably parts of the same gene cluster (Figure 7).



**Figure 7.** Organization of representative BGCs putatively encoding for the biosynthetic machinery of compounds 1-28. *Corallocooccus* sp. MCy9049, *Corallocooccus coralloides* MCy9080 and *Cystobacter* sp. MCy9101 have confirmed production of the target compounds. *Archangium violaceum* SDU34 was included as closest homolog of both target clusters from MCy9101 and has no metabolome data available.

The putative BGC responsible for the biosynthesis of the newly discovered compounds is a Type I PKS-NRPS hybrid cluster that features several peculiar characteristics (Figure 8, Figure S48). First of all, CDS region A contains a CoA-ligase domain with unknown specificity (Table S17) and the corresponding ACP domain as well as a NRPS module consisting of a C, A and PCP domains again with unknown specificity (Table S14). CDS regions B, C and E each consist of a ketosynthase (KS) domain, a putative transacyltransferase docking (tAT\_d) domain (annotated with SeMPI 2.0 [18]) or dehydratase (Dht) domain (annotated with antiSMASH 6.0.1 [19]), an acyltransferase (AT) domain specific for malonyl-CoA (Table S16), a ketoreductase (KR) domain annotated as inactive as well as a carrier protein annotated as peptidyl carrier protein (PCP) domain instead of the expected acyl carrier protein (ACP) domain. The KS domain in CDS region B is presumably condensation-incompetent as the active site cysteine is replaced by serine (KS<sup>S</sup>), a modification typically associated with decarboxylation of ACP-bound extender units [20,21]. CDS region D contains a single NRPS module consisting of a C, A and PCP domain with an A domain specific for serine (Table S14). While the BGCs found in *Corallocooccus* sp. MCy9049 and *Corallocooccus coralloides* MCy9080 do not contain any terminal thioesterase (TE) or reduction domains for releasing the mature product from the assembly line, the BGCs in *Cystobacter* sp. MCy9101 and *Archangium violaceum* SDU34

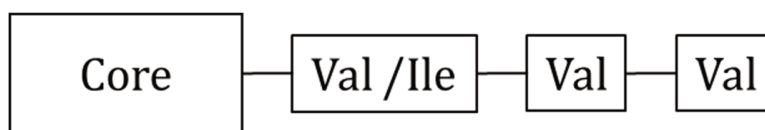
feature an additional NRPS gene that consists of six (MCy9101) or eight (SDU34) modules specific for valine, serine,  $\beta$ -hydroxytyrosine and alanine (Table S15) as well as a terminal thioesterase. Interestingly, although both NRPS genes contain similar modules, the module order differs between both strains. The predicted core cluster regions of each strain end with CDS regions G and H encoding for a putative major facilitator superfamily transporter and a putative penicillin amidase respectively. Needless to say, without the corresponding structures of the target substances, a biosynthesis hypothesis or structure prediction based on these peculiar features is very difficult to impossible.



**Figure 8.** (A) Predicted domains in the biosynthetic gene cluster from *Corallocooccus* sp. MCy9049. (B) Predicted domains in CDS region F from *Cystobacter* sp. MCy9101. (C) Predicted domains in CDS region F from *Archangium violaceum* SDU34. Domain predictions were assigned by SeMPI 2.0 [18] Predicted specificities for A and AT domains are written below the corresponding domain. X: unknown/inconclusive; CAL: CoA ligase; ACP: acyl carrier protein; C: condensation; A: adenylation; PCP: peptidyl carrier protein; KS: ketosynthase; tAT\_d: trans-acyltransferase docking domain (annotated as DHT (dehydratase variant more commonly found in trans-AT PKS clusters) by antiSMASH 6.0.1 [19]); AT: acyltransferase; KR: ketoreductase; TE: thioesterase.

### 3.4 Discussion

This chapter describes the discovery, isolation and characterization of a putatively novel family of monochlorinated and hydroxylated compounds. Due to low yields, full structure elucidation by NMR spectroscopy was not successful. These results highlight a potential constraint of purely metabolome-guided natural product discovery. Although peak intensities and areas in both UV/Vis spectroscopy and MS spectrometry usually correlate with the successive yield of a natural product, it is almost impossible to determine the actual amount of a still unknown compound at first. Although the first reasonable estimates can be already made after some pre-purification steps, a reliable determination of the yield can only be made after successful isolation. Depending on presence or absence of chromophores or easily ionizable functional groups, relative estimates of yield may even be misleading and thus result in the prioritization of minor derivatives of natural product families or compounds produced in amounts too small to resolve their structure. Furthermore, light-sensitive compounds as described in this chapter can face rapid degradation under UV light, resulting in a constant decrease in amount of the compounds during purification processes. Nevertheless, the available analytical data allowed a preliminary insight into the structure of the natural products presented in this chapter. To begin, tandem MS spectra as well as stable isotope feeding experiments demonstrated that the amino acids valine and isoleucine are incorporated directly into the molecules. The tandem MS spectra furthermore suggested that the amino acids likely substitute a yet unknown core structure as a linear peptide chain with the amino acid directly attached to the core assigning the resulting compounds to either the valine or isoleucine subfamily (Scheme 1). In contrast to that, some members of the compound family seem to also incorporate alanine or aminobutyric acid as they show fragments specific for  $C_3H_7NO_2$  (Alanine),  $C_3H_5NO$  (Alanine- $H_2O$ ) or  $C_4H_7NO$  (Aminobutyric acid- $H_2O$ ), but feeding experiments employing d2-alanine did not reveal any incorporation (Table S6). This inconsistency could possibly be explained by an unsuccessful feeding experiment in general as the alanine metabolism is very quick and often disassembled by the primary metabolism before it is used as a building block or the presence of a constitutional isomer of alanine such as *N*-methylglycine or  $\beta$ -alanine, which are indistinguishable from  $\alpha$ -alanine in tandem MS. However, in bacteria  $\beta$ -alanine is mostly formed by decarboxylation of aspartic acid [22] or by uracil degradation [23] and at least feeding of  $^{13}C_4$ - $^{15}N$ -aspartic acid also showed no labeling that would hint on amino acid incorporation.

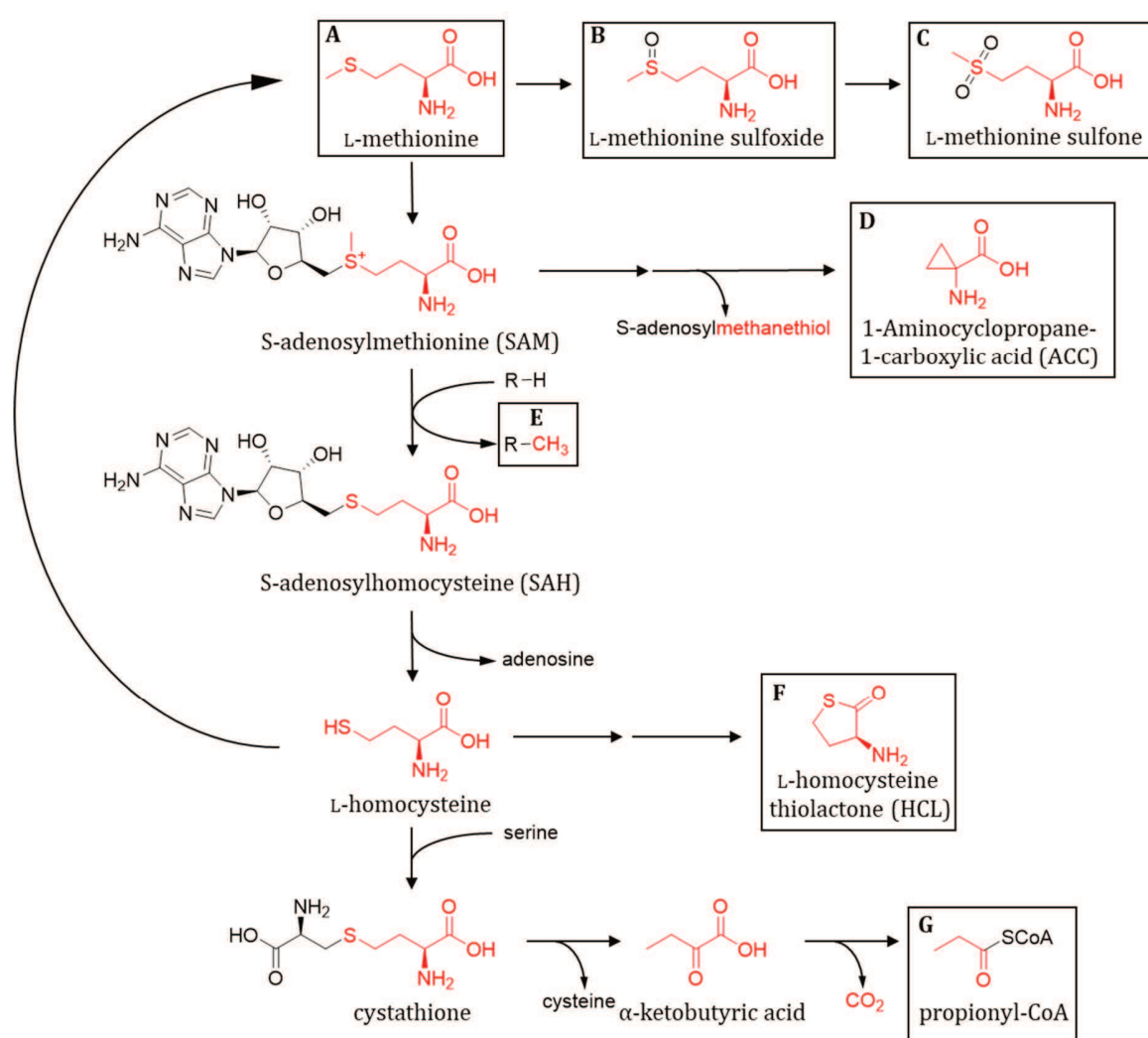


**Scheme 1.** Proposed substructure of compounds **9/23** (Val) and **10/24** (Ile). The ambiguous valine or isoleucine moiety assigns the compounds into the respective subfamily.

Even more peculiar seems the incorporation of  $^{13}\text{C}_5\text{-}^{15}\text{N}_1$ -methionine resulting in a mass shift of +5.01 Da for all members of the compound family. Although there is no tandem MS data available from these feeding experiments, the uniform mass shift of +5.01 Da suggests that incorporation of methionine occurs at the core structure. Direct incorporation of methionine by NRPS adenylation domains as described in the biosynthesis of oscillamide B [24] is only rarely observed (Figure 8A), and even rarer is incorporation of the oxidized forms methionine sulfoxide as in tiglicamide [25] biosynthesis (Figure 8B) and methionine sulfone as seen in nodulapeptin A [26] biosynthesis (Figure 8C). One argument against such direct incorporation is that the compounds most likely do not contain sulfur and furthermore incorporation of methionine should correspond to a mass shift of 6 Da, since all stable-isotope labeled atoms are retained. Another methionine-derived building block in nonribosomal peptide biosynthesis is 1-aminocyclopropane-1-carboxylic acid (ACC, Figure 8D), which is for example found in colibactin [27] and guanganmycin [28] biosynthesis. In both biosynthetic pathways, ACC is produced by intramolecular cyclization of *S*-adenosylmethionine, either bound to a PCP domain or prior to loading to the NRPS assembly line. Even though incorporation of methionine-derived ACC would result in the observed mass shift of +5 Da and does not add any sulfur atoms to the natural product, no homologs for *gnmY* or other aminotransferase domains are found within the putative BGC regions of any of the producer strains. The most common way of incorporating of methionine-derived atoms into natural products is by using SAM-dependent methyltransferases [29]. In general, these enzymes catalyze methyl transfer from SAM to C, O, N or S atoms of various secondary and primary metabolites as well as proteins and DNA [30]. A transfer of five methyl groups to the core structure would also result in a mass shift of +5 Da in feeding experiments and does not introduce sulfur into the resulting structures (Figure 8E). Nevertheless, the putative BGC regions do not contain any methyltransferase domains and it seems unlikely that the putative core fragment ( $\text{C}_{13}\text{H}_{12}\text{N}_2\text{O}_2^+$ ) carries five methyl groups, which add up to  $\text{C}_5\text{H}_{10}$  leaving only a fictitious desmethylated core fragment of  $\text{C}_8\text{H}_2\text{ClN}_2\text{O}_2^+$ . Incorporation of homocysteine or homocysteine thiolactone (HCL, Figure 8F) also appears unlikely, as both building blocks would introduce sulfur into the resulting natural products. Finally, methionine, among other molecules like valine, isoleucine, threonine and odd chain fatty acids, is catabolized to propionyl-CoA (Figure 8G), which is further carboxylated to form methylmalonyl-CoA [31]. Methylmalonyl-CoA is the second most commonly used extender unit in polyketide and fatty acid



biosynthesis, but might also be converted to succinate and re-enter primary metabolism [32]. Again, it seems unlikely that propionyl-CoA is incorporated into the compounds described in this chapter as each propionyl-CoA unit would result in a +3 Da shift in labeling experiments which is inconsistent with the observed shift of +5 Da. Furthermore, prediction of the substrate specificity of the PKS AT domains in the putative BGC suggest the incorporation of malonyl-CoA instead of methylmalonyl-CoA. Last but not least, incorporation of other primary metabolites from succinate also seems peculiar as methionine shows strong and specific incorporation and the succinate building block can as well be produced by various other precursors where no incorporation to the core structure was observed.



**Figure 9.** Integration of methionine in natural product biosynthesis pathways. Highlighted in red are the atoms or subunits that originate from methionine. Labeled with A-G are methionine-derived building blocks found in bacterial natural products.

Structural predictions based on the putative biosynthetic gene cluster should in this case only be taken with great caution because of many uncertainties in the predicted biosynthesis. Modular NRPS and PKS biosynthesis pathway usually follow a so-called collinearity rule, which states that the number and order of catalytic domains in a BGC corresponds to the amino acid and acyl building block number, order and structure in the final product [33]. While the prediction of substrate specificity of A and AT domains is not always unambiguous, the rule still allows at least a prediction of the backbone. However, there are many examples where the final secondary metabolite deviates from the expected product or the collinearity rule does not apply at all [34]. For instance, *Corallocooccus* sp. MCy9049 and *Corallocooccus coralloides* MCy9080 completely lack CDS\_F encoding for the NRPS that is thought to be responsible for the incorporation of valine and/or isoleucine, which are clearly evident in all products. Nevertheless, all producer strains seemed to produce the same metabolites with the only difference being the relative abundance of the products. Even in the two cluster inactivation mutants of MCy9101, only complete abolishment of production was detected, but no newly appearing or increased shunt products were observed. Taken together, it can be stated that a prediction of the biosynthesis and consequently of the underlying BGC is probably only possible with reliably elucidated structures of the resulting products.

Despite the issues discussed above, the compounds described in this chapter also face other difficulties such as their sensitivity to light as they seem to degrade quickly under UV radiation. While the purification of light-unstable compounds may require more effort and time, it also lowers the chances of compound rediscovery as mainly the “low-hanging” fruits of compound discovery, meaning abundant and stable secondary metabolites, have been purified in the past. Furthermore, although crude extracts of MCy9049 did not show any bioactivity in the initial screenings, as activity screenings are usually not conducted in absence of light, it might be possible that the compounds nevertheless possess antimicrobial activity even against the standard panel of pathogens. Moreover, compound purification in absence of light was comparatively easy to accomplish. For most processing steps, it was sufficient to use amber glass vessels wrapped in aluminum foil in order to avoid any light-induced degradation. Only few exceptions like *e.g.* liquid/liquid partitioning, where visual inspection was necessary, were performed in the dark using only a source of indirect red light of 660 nm. For the general benefit of an antimicrobial compound, light sensitivity should also not play too big of a role, since most compounds are not used as they are found in nature anyway and the stability can still be improved afterwards while retaining the bioactivity. Another possible way could be to encapsulate the candidate compound in a light-shielding formulation to enable therapeutic use.

In the last chapter it was already shown that semi-synthesis can be quite advantageous in a compound purification workflow. In the project presented in this chapter, it was also possible to generate

a semisynthetic derivative from the original substance, but the hoped-for advantage in the purification process, namely improved compound stability, unfortunately failed to materialize. In contrast to Myxarylin and Myxarylin-Boc described in the previous chapter, the semisynthetic product generated in this chapter would also not be as easy to revert to the original secondary metabolite. Nevertheless, both chapters demonstrate that organic synthesis can be a useful tool in purification of new secondary metabolites.

### 3.5 Conclusion

In this chapter, we describe the metabolome-guided discovery of a family of light-sensitive compounds from the myxobacterium *Coralloccoccus* sp. MCy9049 that is, according to metabolome mining results, frequently encountered in the genus *Coralloccoccus*. Isolation of the most abundant derivatives could not provide yields sufficient for NMR-based structure elucidation. Tandem MS fragmentation revealed that all 28 members of this compound family share the same core scaffold, which seems to be either chlorinated or hydroxylated. A cryptic type I PKS-NRPS hybrid BGC could be identified by gene disruption experiments in the alternative host *Cystobacter* sp. MCy9101, which putatively encodes for the biosynthetic machinery of these compounds. Feeding experiments with stable-isotope labeled building blocks showed strong incorporation of methionine, valine and isoleucine, which cannot be fully explained by the predicted substrate specificity of the adenylation domains present in the anticipated BGC. Despite the efforts, predictions of the structure of the common core scaffold were unsuccessful so far. Because of many uncertainties in the predicted biosynthesis, structural predictions based on the BGC are very difficult to impossible.

In summary, the discovery of the compound class described in this chapter and the putative assignment of a corresponding biosynthetic gene cluster set the foundation for further investigations ultimately leading to full structure elucidation and a proposed biosynthesis. Increased yields sufficient for full compound characterization could for example be achieved by switching to more efficient production hosts, improvement of the purification procedure or by homologous overexpression of the operon in *Cystobacter* sp. MCy9101. Assessment of antimicrobial and cytotoxic activities in absence of light might furthermore uncover a yet unknown biological function of these molecules.

## 3.6 Materials and Methods

### 3.6.1 Applied Software, DNA Sequence Analysis and Bioinformatics Methods

The genomes of *Corallococcus* sp. MCy9049, *Corallococcus coralloides* MCy9080 (DSM2259), *Cyterbacter* sp. MCy9101 and *Archangium violaceum* SDU34 are screened with the software Geneious Prime® (Biomatters Ltd., Auckland, New Zealand, 2021.2.1) [35]. In order to find homologous genes or proteins, the nucleotide or amino acid sequence is aligned with the basic local alignment search tool (BLAST) against an in-house genome database or publicly available nucleotide or protein databases. Functional prediction of ORFs is performed using protein blast and the Pfam [36] database to search for domains, repeats or motifs and classify hypothetical proteins into putative protein families. Predictions of NRPS and PKS domain specificity and activity are conducted with antiSMASH 6.0.1 [19], NRPSpredictor2 [37] and SeMPI 2.0 [18].

### 3.6.2 Myxobacterial Fermentation and Extraction for LC-MS Analysis

*Cystobacterineae* species *Corallococcus* sp. MCy9049 is routinely cultivated at 30 °C in VY/2 medium (Table S1). Liquid cultures are grown in 300 mL shake flasks without baffles on an orbital shaker at 180 rpm for 10-14 days. Standard fermentation and media optimization is performed by cultivating the strain in 50 mL VY/2 medium using 1 mL inoculum (2% inoculum volume). The cultures are supplemented with 1 mL (v/v) sterile amberlite resin XAD-16 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) and with aqueous solutions of the additives listed in Table S5. After seven days, the combined cells and resin are separated from the supernatant by centrifugation at 6,000 rcf for 10min using an Eppendorf falcon table centrifuge. While the supernatant is discarded, the pellet is extracted in a 100 mL Erlenmeyer flask by adding 50 mL of methanol and stirring for 60 min on a magnetic stirrer. The methanol extract is left to settle in order to sediment the cell debris and XAD-16 resin for a second extraction step. The extract is filtered through a 125-micron folded filter paper, keeping cells and XAD-16 resin in the Erlenmeyer flask. Subsequently, the filter residue is extracted again with 50 mL acetone by stirring for 1 h and filtered through the same folded filter paper. The combined extract is transferred into a 100 mL round bottom flask and dried using a rotary evaporator at 260 mbar and 40 °C water bath temperature (acetone), 180 mbar and 40 °C (methanol) and 20 mbar and 40 °C (residual water) until the residue in the flask is completely dry. The residue is redissolved in 1.5 mL of methanol, further diluted 1:5 with methanol (v/v) and centrifuged for 5 min at 15,000 rpm with a table centrifuge (Hitachi Koki Co., Tokyo, Japan) to remove residual insolubilities such as salts, cell debris and XAD-16 fragments. The supernatant is analyzed by high-performance liquid chromatography (HPLC) coupled to a mass spectrometer (MS).

### 3.6.3 Standardized UHPLC-MS Conditions for Bacterial Crude Extracts

UHPLC-hrMS analysis is performed on a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Bruker maXis 4G ultra-high-resolution quadrupole time-of-flight (UHR-qTOF) MS equipped with a high-resolution electrospray ionization (HRESI) source (Bruker Daltonics, Billerica, MA, USA). The separation of 1  $\mu\text{L}$  sample is achieved with a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH<sub>2</sub>O with 0.1% formic acid on an ACQUITY BEH C18 column (100 mm  $\times$  2.1 mm, 1.7  $\mu\text{m}$  d<sub>p</sub>) (Waters, Eschborn, Germany) equipped with a Waters VanGuard BEH C18 1.7  $\mu\text{m}$  guard column at a flow rate of 0.6 mL/min and 45 °C for 18 min with detection by a diode array detector at 200–600 nm. The LC flow is split into 75  $\mu\text{L}/\text{min}$  before entering the mass spectrometer. Mass spectrograms are acquired in centroid mode ranging from 150–2500 m/z at an acquisition rate of 2 Hz in positive MS mode. Source parameters are set to 500 V end-plate offset; 4000 V capillary voltage; 1 bar nebulizer gas pressure; 5 L/min dry gas flow; and 200 °C dry gas temperature. Ion transfer and quadrupole parameters are set to 350 V<sub>pp</sub> funnel RF; 400 V<sub>pp</sub> multipole RF; 5 eV ion energy; and 120 m/z low-mass cut-off. Collision cell is set to 5.0 eV and pre-pulse storage is set to 5  $\mu\text{s}$ . Calibration is conducted automatically before every HPLC-MS run by the injection of sodium formate and calibration on the respective clusters formed in the ESI source. All MS analyses are acquired in the presence of the lock masses C<sub>12</sub>H<sub>19</sub>F<sub>12</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, C<sub>18</sub>H<sub>19</sub>F<sub>24</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub> and C<sub>24</sub>H<sub>19</sub>F<sub>36</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, which generate the [M + H]<sup>+</sup> ions of 622.0289, 922.0098 and 1221.9906. The HPLC-MS system is operated by HyStar 5.1 (Bruker Daltonics, Billerica, MA, USA), and LC chromatograms as well as UV spectra and mass spectrograms are analyzed with DataAnalysis 5.3 (Bruker Daltonics, Billerica, MA, USA). LC and MS conditions for scheduled precursor list (SPL)-guided tandem MS data acquisitions are kept constant according to standard UHPLC-MS conditions. Tandem MS data acquisition parameters are set to exclusively fragment SPL entries within a retention time tolerance of 0.2 min and a mass tolerance of 0.05 m/z for precursor ion selection. The method picks up to two precursors per cycle, applies smart exclusion after five spectra and performs CID and MS/MS spectra acquisition time ramping. CID energy is ramped from 35 eV for 500 m/z to 45 eV for 1000 m/z and 60 eV for 2000 m/z. MS full-scan acquisition rate is set to 2 Hz and MS/MS spectra acquisition rates are ramped from 1 to 4 Hz for precursor ion intensities of 10 kcts to 100 kcts.

### 3.6.4 Statistical Metabolomics and Similarity Clustering via GNPS

For statistical metabolomics analysis, both the myxobacterial culture and medium blanks are incubated and extracted in triplicates as described before. Each extract is measured as technical duplicate, giving a total number of six replicates each for the bacterial culture and medium blank extracts. T-ReX-3D molecular feature finder of MetaboScape 2021b (Bruker Daltonics, Billerica, MA, USA) is used to obtain

molecular features. Detection parameters are set to intensity threshold  $5 \times 10^3$  and minimum peak length of five spectra to ensure a precursor selection that has a sufficient intensity to generate tandem MS data. Identification of bacterial features is performed with the built-in ANOVA/t-test routine and filtered to appearance in all six bacterial extracts and in none of the medium blank extracts. The t-test table is used to create a scheduled precursor list (SPL) for tandem MS fragmentation. The resulting tandem MS chromatogram is exported as MZML file using DataAnalysis 5.3 (Bruker Daltonics, Billerica, MA, USA) and submitted to the GNPS online workflow [11]. In the GNPS networking pipeline MS/MS spectra are window filtered by choosing only the top six peaks in the  $\pm 50$  Da window throughout the spectrum. The data is clustered with MS-Cluster with a parent mass tolerance of 0.05 Da and a MS/MS fragment ion tolerance of 0.1 Da to create consensus spectra. A spectral network is created where edges are filtered to have a cosine score above 0.8 [38] and more than 7 matched peaks. Further edges between two nodes are kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the resulting spectral network is visualized by Cytoscape 3.6.1 [13].

### 3.6.5 Stability Assay and Catalytic Hydrogenation

In order to determine the rate of compound degradation and isomerization crude standard extracts obtained as described before are diluted 1:5 with methanol and analyzed using the standard UHPLC-*hr*MS setup to determine reference peak areas of the compounds of interest. Subsequently, 100  $\mu$ L aliquots of the diluted extracts are transferred into 300  $\mu$ L conical vials which are then exposed to different conditions, namely direct and indirect white light, temperatures of 4 °C, 20 °C and 45 °C, addition of 0.1 % and 1 % formic acid (v/v) as well as potassium phosphate buffer pH 6.8, addition of dithiothreitol (DTT) and flushing the vial with inert gas (nitrogen) and compressed air. After 24 h of exposure, the samples are filled up again to the 100  $\mu$ L mark, centrifuged and analyzed again by UHPLC-*hr*MS. Determination of degradation and isomerization is conducted by comparison of the respective peak areas to the reference from extracted ion chromatograms (EICs) generated in DataAnalysis 5.3 (Bruker Daltonics, Billerica, MA, USA).

For catalytic hydrogenation, the crude extract is first split into 20 fractions using a Dionex UltiMate 3000 semi-preparative system equipped with an automated fraction collector (Thermo Fisher Scientific, Waltham, MA, USA). The separation of 100  $\mu$ L sample is achieved with a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH<sub>2</sub>O with 0.1% formic acid on an ACQUITY BEH C18 column (250  $\times$  10 mm, 5  $\mu$ m d<sub>p</sub>) (Waters, Eschborn, Germany) at a flow rate of 5.0 mL/min and 45 °C for 20 min. Because of the light-sensitive compounds, the lamps of the diode array detector are turned off during fractionation. Every minute of the gradient is collected in one tube, analyzed by UHPLC-*hr*MS as described before and

dried by freezing in liquid nitrogen and subsequent lyophilization. About 25 mg of the fraction containing compound **5** is dissolved in 1.5 mL methanol + 0.1% formic acid and transferred to a pear-shaped flask. Afterwards, 10% (*m/v*) palladium on activated charcoal (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) are added to the reaction flask. The flask is first evacuated and refilled with nitrogen gas three times and further evacuated and refilled with hydrogen gas. The mixture is stirred for 3 days in total. After 2, 4, 12, 24, 48 and 72 hours, the flask is evacuated and refilled with nitrogen gas. 5  $\mu\text{L}$  of the mixture are removed from the reaction and the flask is evacuated again with nitrogen and hydrogen respectively as described above. The aliquot take from the reaction is diluted 1:4 with methanol, centrifuged for 5 min at 15,000 rpm with a table centrifuge (Hitachi Koki Co., Tokyo, Japan) and the supernatant is analyzed by UHPLC-MS. Afterwards, to determine differences in compound stability for the hydrogenated semisynthetic congeners, the vial is again exposed to indirect light for 24 h and analyzed again by UHPLC-MS.

### 3.6.6 Compound Isolation

The myxobacterial strain *Corallococcus* sp. MCy9049 is cultivated in six 5 L shake flasks each containing 2 L 479 medium for isolation of compounds of the valine-subfamily or 493 medium for isolation of compounds of the isoleucine subfamily, all supplemented with 2% XAD-16 resin (*v/v*) as described before. After 14 days of fermentation, combined cells and resin are harvested by centrifugation and extracted with 1 L methanol twice. After drying under vacuum, the extract is redissolved in 500 mL methanol and partitioned using 500 mL *n*-hexane. The methanol layer is dried, redissolved in 500 mL ddH<sub>2</sub>O, and further partitioned using 500 mL chloroform. The aqueous layer is once more partitioned using 500 mL ethyl acetate. The aqueous layer containing the compounds of interest is dried by lyophilization and redissolved in 10 mL ddH<sub>2</sub>O. Purification is performed using a Dionex UltiMate 3000 semi-preparative system equipped with an automated fraction collector (Thermo Fisher Scientific, Waltham, MA, USA) that was entirely covered in aluminum foil to block light. Compound separation is achieved with a gradient of acetonitrile with 0.1% formic acid (B) in ddH<sub>2</sub>O with 0.1% formic acid (A) on a Phenomenex Luna C18(2) column (250  $\times$  10 mm, 5  $\mu\text{m}$  d<sub>p</sub>) (Phenomenex Ltd., Aschaffenburg, Germany) at a flow rate of 5.0 mL/min and 45 °C. The initial gradient is held at 10% B for 2 min and then elevated to 20% B within 3 min. After that, the B level is increased to 40% within 28 min and subsequently elevated to 95% within 2.5 min and held there for 1.5 minutes. Finally, the gradient is ramped back to 10% B in 0.5 min and re-equilibrated for the next injection for 3 min. Detection is performed using an HCT ultra ESI-MS ion trap mass spectrometer (Bruker Daltonics, Billerica, MA, USA) while the built-in Dionex UltiMate 3000 DAD compartment is turned off to assure compound stability. The LC flow is split into 500  $\mu\text{L}/\text{min}$  before entering the mass

spectrometer. Mass spectrograms are acquired in “standard enhanced” scan mode ranging from 200–1000  $m/z$  in positive MS mode. Source parameters are set to 500 V end-plate offset; 4500 V capillary voltage; 30 psi nebulizer gas pressure; 12 L/min dry gas flow; and 300 °C dry gas temperature. Ion transfer and quadrupole parameters are automatically set by Smart Parameter Setting (SPS) with a target mass of 500  $m/z$ , 100% compound stability and 100% trap drive level. The ion trap is set to ICC smart target of 100000 and 100 ms accumulation time. After drying the HPLC fractions containing the target compounds by lyophilization and redissolving in 2 mL ddH<sub>2</sub>O, the fractions undergo a second step of semi-preparative HPLC using the same analytical setup as described above. Compound separation is achieved with a gradient of acetonitrile with 0.1% formic acid (B) in ddH<sub>2</sub>O with 0.1% formic acid (A) on a Phenomenex Kinetex Biphenyl column (250 × 10 mm, 5 μm d<sub>p</sub>) (Phenomenex Ltd., Aschaffenburg, Germany) at a flow rate of 5.0 mL/min and 45 °C. The initial gradient is held at 20% B for 2 min and then elevated to 40% B within 24.5 min. After that, the B level was increased to 95% within 0.5 min and held there for 1 minute. Finally, the gradient was ramped back to 20% B in 0.5 min and re-equilibrated for the next injection for 3 minutes. The isolated pure compounds were dried by lyophilization.

### 3.7 Acknowledgments

The authors thank Nestor Zaburanyi for initial genome assembly of MCy9049 and MCy9101 and Joachim Hug for assistance with the interpretation of genomic data and fruitful discussions about structures and biosynthesis of the members of the newly discovered compound family.

### 3.8 References

1. Dawid, W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol. Rev.* **2000**, *24*, 403–427, doi:10.1111/j.1574-6976.2000.tb00548.x.
2. Muñoz-Dorado, J.; Marcos-Torres, F.J.; García-Bravo, E.; Moraleda-Muñoz, A.; Pérez, J. Myxobacteria: moving, killing, feeding, and surviving together. *Front. Microbiol.* **2016**, *7*, 781, doi:10.3389/fmicb.2016.00781.
3. Cao, P.; Dey, A.; Vassallo, C.N.; Wall, D. How Myxobacteria Cooperate. *J. Mol. Biol.* **2015**, doi:10.1016/j.jmb.2015.07.022.
4. Herrmann, J.; Fayad, A.A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34*, 135–160, doi:10.1039/C6NP00106H.



5. Weissman, K.J.; Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **2010**, *27*, 1276–1295, doi:10.1039/c001260m.
6. Wenzel, S.C.; Müller, R. Myxobacteria—“microbial factories’ for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, *5*, 567–574, doi:10.1039/b901287g.
7. Silakowski, B.; Kunze, B.; Nordsiek, G.; Blöcker, H.; Höfle, G.; Müller, R. The myxochelin iron transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15. *Eur. J. Biochem.* **2000**, *267*, 6476–6485, doi:10.1046/j.1432-1327.2000.01740.x.
8. Hug, J.J.; Panter, F.; Krug, D.; Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319–334, doi:10.1007/s10295-018-2105-6.
9. Dickschat, J.S.; Bode, H.B.; Mahmud, T.; Müller, R.; Schulz, S. A novel type of geosmin biosynthesis in myxobacteria. *J. Org. Chem.* **2005**, *70*, 5174–5182, doi:10.1021/jo050449g.
10. Panter, F.; Krug, D.; Müller, R. Novel Methoxymethacrylate Natural Products Uncovered by Statistics-Based Mining of the *Myxococcus fulvus* Secondary Metabolome. *ACS Chem. Biol.* **2019**, *14*, 88–98, doi:10.1021/acscchembio.8b00948.
11. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837, doi:10.1038/nbt.3597.
12. Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; et al. Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* **2020**, *17*, 905–908, doi:10.1038/s41592-020-0933-6.
13. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504, doi:10.1101/gr.1239303.
14. Meija, J.; Coplen, T.B.; Berglund, M.; Brand, W.A.; Bièvre, P. de; Gröning, M.; Holden, N.E.; Irrgeher, J.; Loss, R.D.; Walczyk, T.; et al. Atomic weights of the elements 2013 (IUPAC Technical Report). *Pure and Applied Chemistry* **2016**, *88*, 265–291, doi:10.1515/pac-2015-0305.
15. Zeng, H. Exploring the biosynthetic potential of *Cystobacter* sp. SBCb004. Dissertation; Saarland University, Saarbrücken, 2021.
16. Seemann, T.; Kwong, J.; Gladman, S.; Gonçalves da Silva, A. *shovill*, available online: <https://github.com/tseemann/shovill> (accessed on 18 January 2022).

17. Kolmogorov, M.; Armstrong, J.; Raney, B.J.; Streeter, I.; Dunn, M.; Yang, F.; Odom, D.; Flicek, P.; Keane, T.M.; Thybert, D.; et al. Chromosome assembly of large and complex genomes using multiple references. *Genome Res.* **2018**, *28*, 1720–1732, doi:10.1101/gr.236273.118.
18. Zierep, P.F.; Ceci, A.T.; Dobrusin, I.; Rockwell-Kollmann, S.C.; Günther, S. SeMPI 2.0—A Web Server for PKS and NRPS Predictions Combined with Metabolite Screening in Natural Product Databases. *Metabolites* **2021**, *11*, doi:10.3390/metabo11010013.
19. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* **2021**, *49*, W29–W35, doi:10.1093/nar/gkab335.
20. Keatinge-Clay, A.T. The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. *Chem. Rev.* **2017**, *117*, 5334–5366, doi:10.1021/acs.chemrev.6b00683.
21. Witkowski, A.; Joshi, A.K.; Lindqvist, Y.; Smith, S. Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. *Biochemistry* **1999**, *38*, 11643–11650, doi:10.1021/bi990993h.
22. Cronan, J. E., Jr. Beta-alanine synthesis in Escherichia coli. *J. Bacteriol.* **1980**, *141*, 1291–1297, doi:10.1128/jb.141.3.1291-1297.1980.
23. López-Sámamo, M.; Beltrán, L.F.L.-A.; Sánchez-Thomas, R.; Dávalos, A.; Villaseñor, T.; García-García, J.D.; García-de Los Santos, A. A novel way to synthesize pantothenate in bacteria involves  $\beta$ -alanine synthase present in uracil degradation pathway. *Microbiologyopen* **2020**, *9*, e1006, doi:10.1002/mbo3.1006.
24. Sano, T.; Usui, T.; Ueda, K.; Osada, H.; Kaya, K. Isolation of new protein phosphatase inhibitors from two cyanobacteria species, *Planktothrix* spp. *J. Nat. Prod.* **2001**, *64*, 1052–1055, doi:10.1021/np0005356.
25. Matthew, S.; Paul, V.J.; Luesch, H. Tiglicamides A–C, cyclodepsipeptides from the marine cyanobacterium *Lyngbya confervoides*. *Phytochemistry* **2009**, *70*, 2058–2063, doi:10.1016/j.phytochem.2009.09.010.
26. Fujii, K.; Sivonen, K.; Adachi, K.; Noguchi, K.; Sano, H.; Hirayama, K.; Suzuki, M.; Harada, K.-i. Comparative study of toxic and non-toxic cyanobacterial products: Novel peptides from toxic *Nodularia spumigena* AV1. *Tetrahedron Lett.* **1997**, *38*, 5525–5528, doi:10.1016/S0040-4039(97)01192-1.
27. Zha, L.; Jiang, Y.; Henke, M.T.; Wilson, M.R.; Wang, J.X.; Kelleher, N.L.; Balskus, E.P. Colibactin assembly line enzymes use S-adenosylmethionine to build a cyclopropane ring. *Nat. Chem. Biol.* **2017**, *13*, 1063 EP -, doi:10.1038/nchembio.2448.

28. Xu, Z.; Pan, G.; Zhou, H.; Shen, B. Discovery and Characterization of 1-Aminocyclopropane-1-carboxylic Acid Synthase of Bacterial Origin. *J. Am. Chem. Soc.* **2018**, doi:10.1021/jacs.8b11463.
29. Kozbial, P.Z.; Mushegian, A.R. Natural history of S-adenosylmethionine-binding proteins. *BMC Struct. Biol.* **2005**, *5*, 1–25, doi:10.1186/1472-6807-5-19.
30. Sun, Q.; Huang, M.; Wei, Y. Diversity of the reaction mechanisms of SAM-dependent enzymes. *Acta Pharm. Sin. B.* **2021**, *11*, 632–650, doi:10.1016/j.apsb.2020.08.011.
31. Wongkittichote, P.; Ah Mew, N.; Chapman, K.A. Propionyl-CoA carboxylase – A review. *Molecular Genetics and Metabolism* **2017**, *122*, 145–152, doi:10.1016/j.ymgme.2017.10.002.
32. Chan, Y.A.; Podevels, A.M.; Kevany, B.M.; Thomas, M.G. Biosynthesis of Polyketide Synthase Extender Units. *Nat. Prod. Rep.* **2009**, *26*, 90–114, doi:10.1039/b801658p.
33. Callahan, B.; Thattai, M.; Shraiman, B.I. Emergent gene order in a model of modular polyketide synthases. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19410–19415.
34. Aleti, G.; Sessitsch, A.; Brader, G. Genome mining: Prediction of lipopeptides and polyketides from *Bacillus* and related Firmicutes. *Computational and Structural Biotechnology Journal* **2015**, *13*, 192–203, doi:10.1016/j.csbj.2015.03.003.
35. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649, doi:10.1093/bioinformatics/bts199.
36. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–285, doi:10.1093/nar/gkv1344.
37. Rottig, M.; Medema, M.H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **2011**, *39*, W362–7, doi:10.1093/nar/gkr323.
38. Ramakrishnan, S.R.; Mao, R.; Nakorchevskiy, A.A.; Prince, J.T.; Willard, W.S.; Xu, W.; Marcotte, E.M.; Miranker, D.P. A fast coarse filtering method for peptide identification by mass spectrometry. *Bioinformatics* **2006**, *22*, 1524–1531, doi:10.1093/bioinformatics/btl118.

## Supporting Information

Metabolome-guided Discovery of a  
Family of Light-sensitive Compounds  
from *Corallococcus* sp. MCy9049

### S 3.1 Myxobacterial Culture Media

**Table S1.** Recipe for VY/2 medium.

<b>VY/2 – Medium</b>			
Amount	Ingredient	Concentration	Supplier
5 g/L	Yeast (entire cells)	-	Deutsche Hefewerke GmbH, Nürnberg, Germany
5 g/L	Soluble Starch	-	Carl Roth GmbH, Karlsruhe, Germany
1 g/L	CaCl <sub>2</sub> • 2H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 g/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	-	Grüssing GmbH, Filsum, Germany
10 mL/L	TRIS • HCl pH 8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, pH adjusted to 7.2 with 1 N HCl			

**Table S2.** Recipe for 479 medium.

<b>479 – Medium</b>			
Amount	Ingredient	Concentration	Supplier
10 g/L	Yeast (entire cells)	-	Deutsche Hefewerke GmbH, Nürnberg, Germany
5 g/L	Soluble Starch	-	Carl Roth GmbH, Karlsruhe, Germany
1 g/L	CaCl <sub>2</sub> • 2H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 g/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	-	Grüssing GmbH, Filsum, Germany
10 mL/L	TRIS • HCl pH 8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
0.5 g/L	Valine	-	Fagron GmbH & Co. KG, Glinde, Germany
0.5 g/L	Proline	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Riboflavin solution (added after autoclaving)	1 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Nicotinic acid solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
100 µL/L	Sterile Pyridoxine solution (added after autoclaving)	1 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 mL/L	Sterile TE solution	-	self-prepared (see Table S4)
Dissolved in milli-Q water, pH adjusted to 7.2 with 1 N HCl			

**Table S3.** Recipe for 493 medium.

493 – Medium			
Amount	Ingredient	Concentration	Supplier
10 g/L	Yeast (entire cells)	-	Deutsche Hefewerke GmbH, Nürnberg, Germany
5 g/L	Soluble Starch	-	Carl Roth GmbH, Karlsruhe, Germany
1 g/L	CaCl <sub>2</sub> • 2H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 g/L	MgSO <sub>4</sub> • 7H <sub>2</sub> O	-	Grüssing GmbH, Filsum, Germany
10 mL/L	TRIS • HCl pH 8	1 M	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
0.5 g/L	Isoleucine	-	Fagron GmbH & Co. KG, Glinde, Germany
0.5 g/L	Proline	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Riboflavin solution (added after autoclaving)	1 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
100 µL/L	Sterile Nicotinic acid solution (added after autoclaving)	1 mg/mL	Carl Roth GmbH, Karlsruhe, Germany
100 µL/L	Sterile Pyridoxine solution (added after autoclaving)	1 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
200 µL/L	Sterile FeNaEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
1 mL/L	Sterile TE solution	-	self-prepared (see Table S4)
Dissolved in milli-Q water, pH adjusted to 7.2 with 1 N HCl			

**Table S4.** Recipe for Trace Elements (TE) solution.

Trace Elements Solution 1000x			
Amount	Ingredient	Concentration	Supplier
0.1 g/L	MnCl <sub>2</sub> • 4H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
0.02 g/L	CoCl <sub>2</sub> • 6H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
0.01 g/L	CuSO <sub>4</sub>	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
0.01 g/L	Na <sub>2</sub> MoO <sub>4</sub> • 2H <sub>2</sub> O	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
0.02 g/L	ZnCl <sub>2</sub>	-	Carl Roth GmbH, Karlsruhe, Germany
0.005 g/L	LiCl	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
0.005 g/L	SnCl <sub>2</sub>	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
0.01 g/L	H <sub>3</sub> BO <sub>3</sub>	-	Carl Roth GmbH, Karlsruhe, Germany
0.02 g/L	KBr	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
0.02 g/L	KI	-	Sigma Aldrich Chemie GmbH, Taufkirchen, Germany
Dissolved in milli-Q water, sterilized by filtration (0.22 µm pore size)			

## S 3.1.1 Media Optimization

**Table S5.** Production optimization by addition of amino acids and vitamins to the culture medium.

Additive to VY/2 medium	Amount added	$\Delta$ (Peak Area) Compound 5	$\Delta$ (Peak Area) Compound 6
none (control)	-	-	-
Sterile proline solution 1 mg/mL	1 mL/L	slight increase	slight increase
Sterile valine solution 0.1 mg/mL	10 mL/L	strong increase	slight decrease
Glucose	1 g/L	slight increase	slight increase
Maltose • H <sub>2</sub> O	1 g/L	similar	similar
TE solution 1000x	1 mL/L	strong increase	similar
Sterile riboflavin solution 1 mg/mL	100 $\mu$ L/L	slight increase	similar
Sterile nicotinic acid solution 1 mg/mL	100 $\mu$ L/L	slight increase	slight increase
Sterile pyridoxin solution 1 mg/mL	100 $\mu$ L/L	slight increase	slight increase
Sterile cobalamin solution 1 mg/mL	100 $\mu$ L/L	slight decrease	similar
Sterile FeNaEDTA solution 8 mg/mL	200 $\mu$ L/L	similar	similar

All additive solutions were sterilized by filtration (0.22  $\mu$ m pore size) and added to the medium after autoclaving, all solid additives were added during media preparation before autoclaving. Changes in peak area are compared to the extracts of the control (cultivation in standard VY/2 medium (Table S1)) and rated into five categories:

- More than 50% increase in peak area: strong increase
- 20 – 50% increase in peak area: slight increase
- Less than 20% increase or decrease: similar
- 20 – 50% decrease in peak area: slight decrease

All additives that showed to be beneficial for production of either compound were added into the modified VY/2 media named “479 medium” and “493 medium” (Tables S2-3). FeNaEDTA was added despite having no apparent effect on production levels because its addition suppresses the production of siderophores such as the myxochelins and thus facilitates compound purification. While only changes in peak area for the most abundant compounds **5** and **6** are presented here, similar behavior was observed for every compound of the respective subfamily of both the chlorinated and hydroxylated members.

## S 3.2 Compound Characterization

### S 3.2.1 Tandem MS Spectra

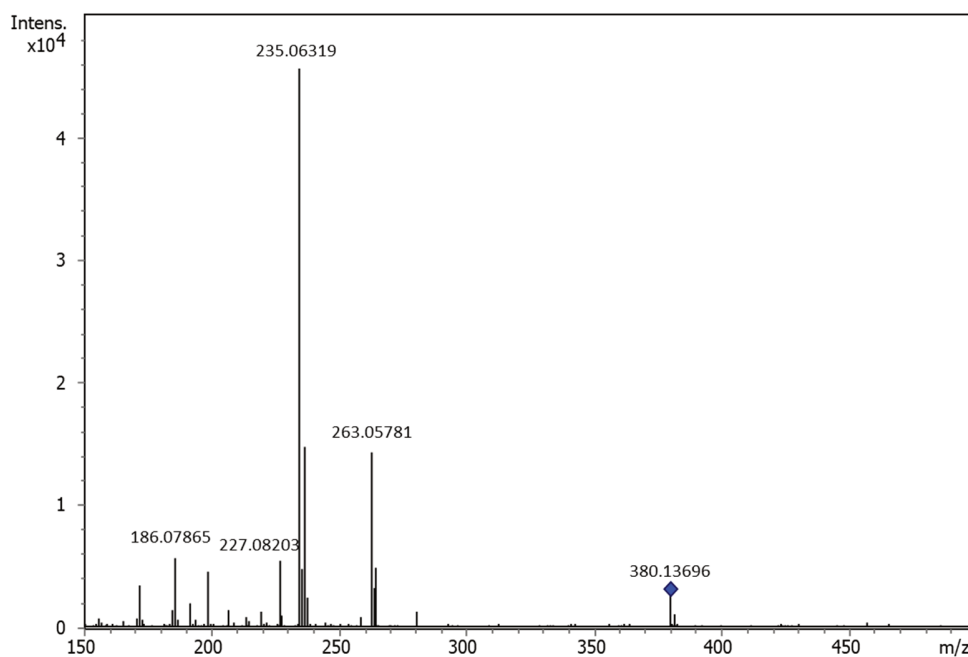


Figure S1. Tandem MS spectrum of 1.

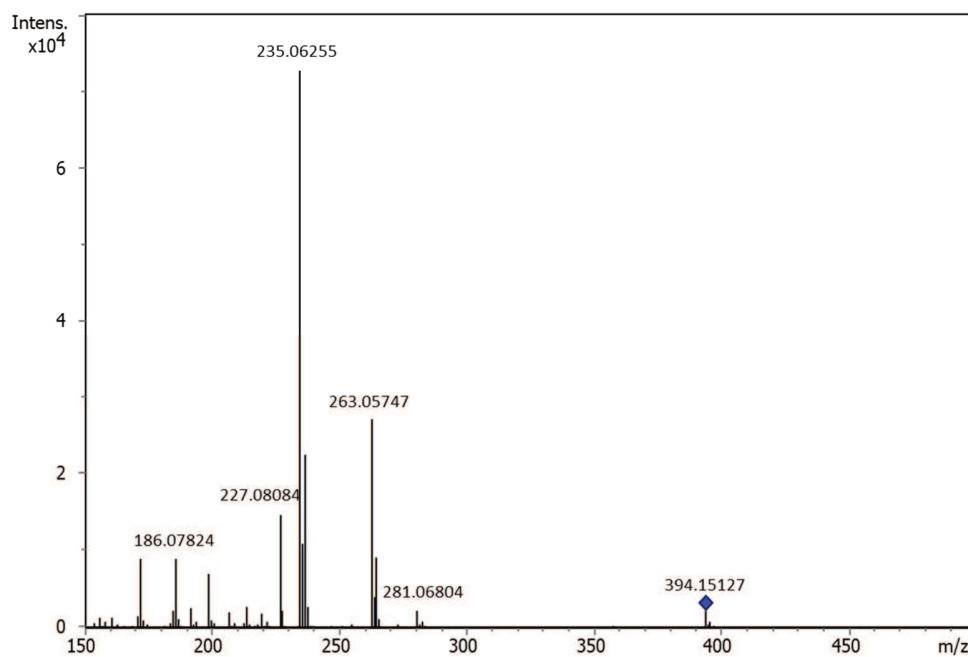


Figure S2. Tandem MS spectrum of 2.



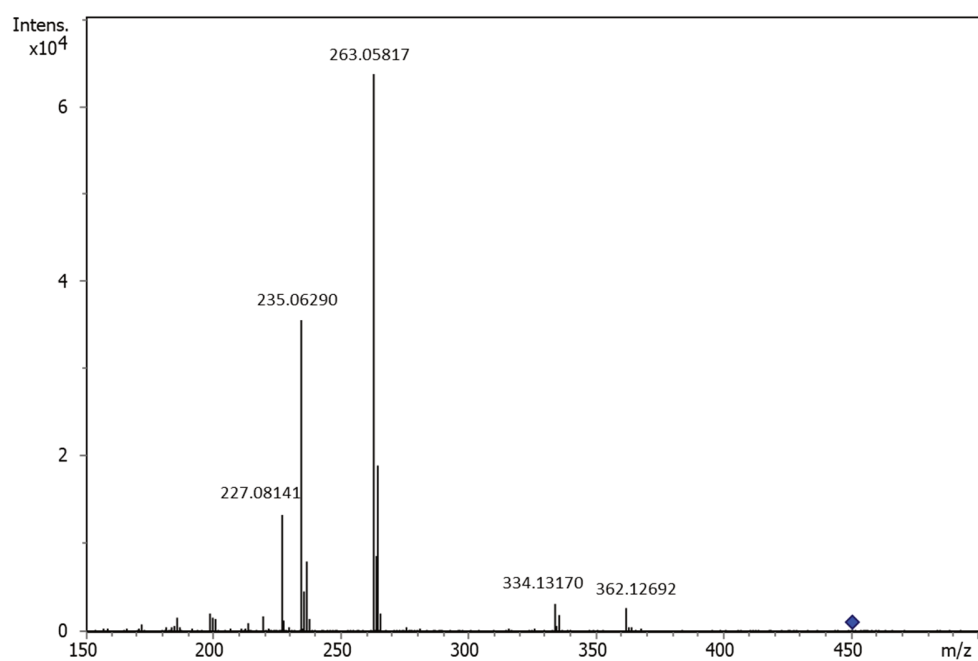


Figure S3. Tandem MS spectrum of 3.

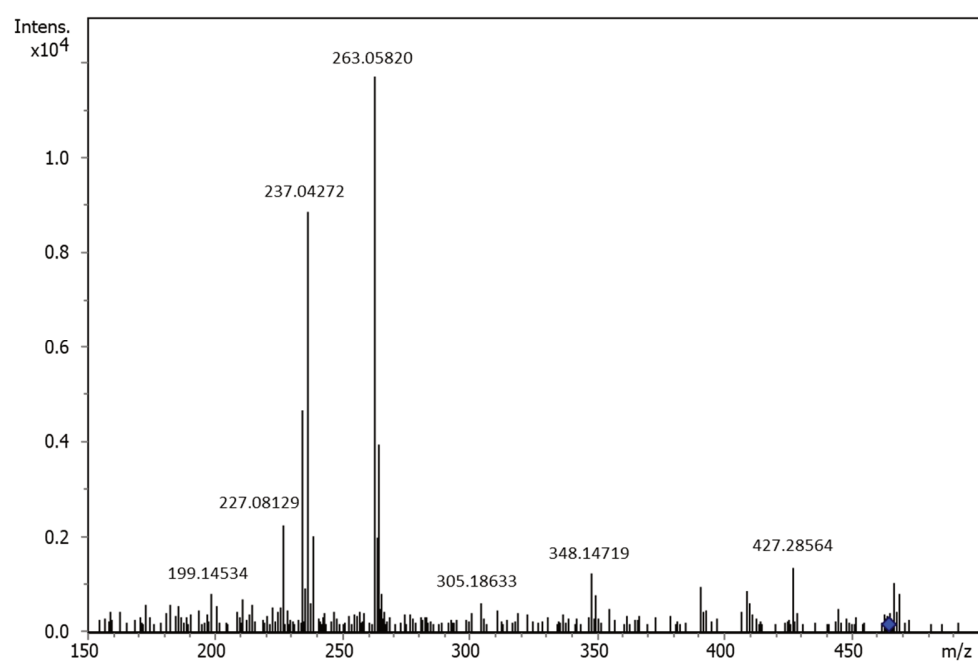


Figure S4. Tandem MS spectrum of 4.

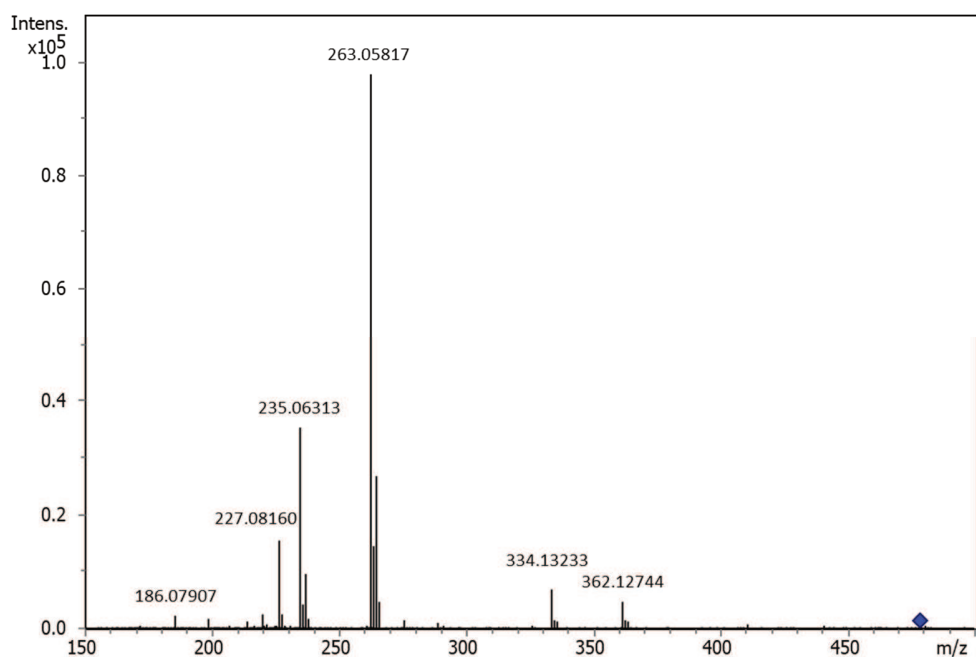


Figure S5. Tandem MS spectrum of 5.

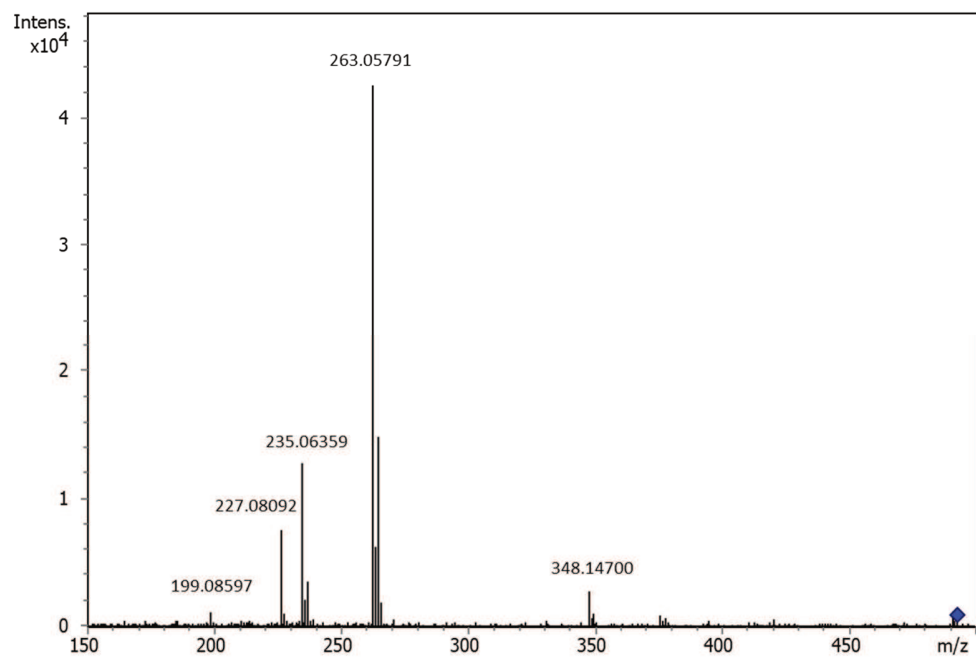


Figure S6. Tandem MS spectrum of 6.

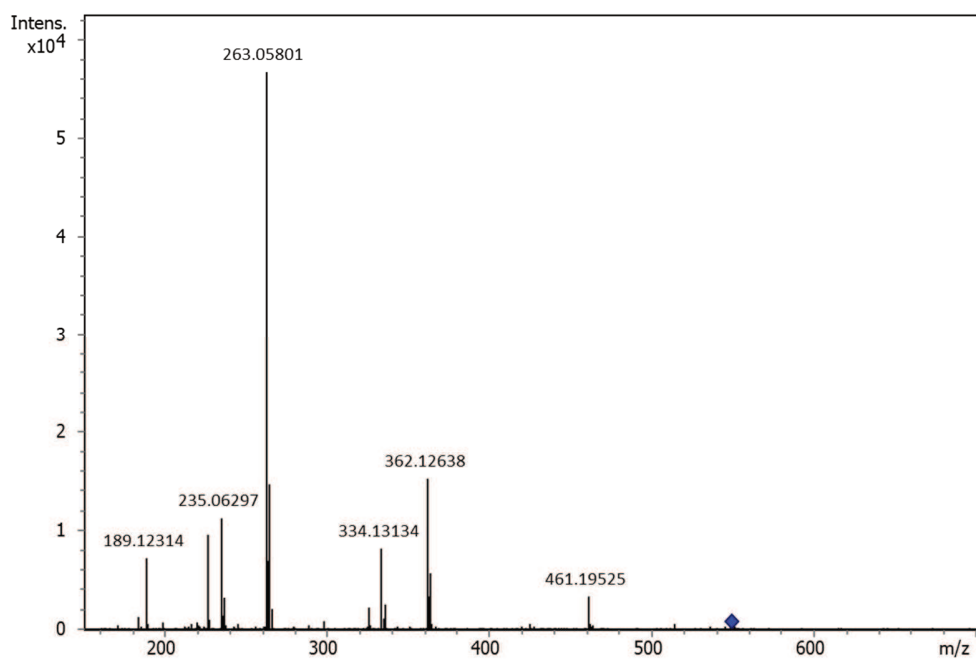


Figure S7. Tandem MS spectrum of **7**.

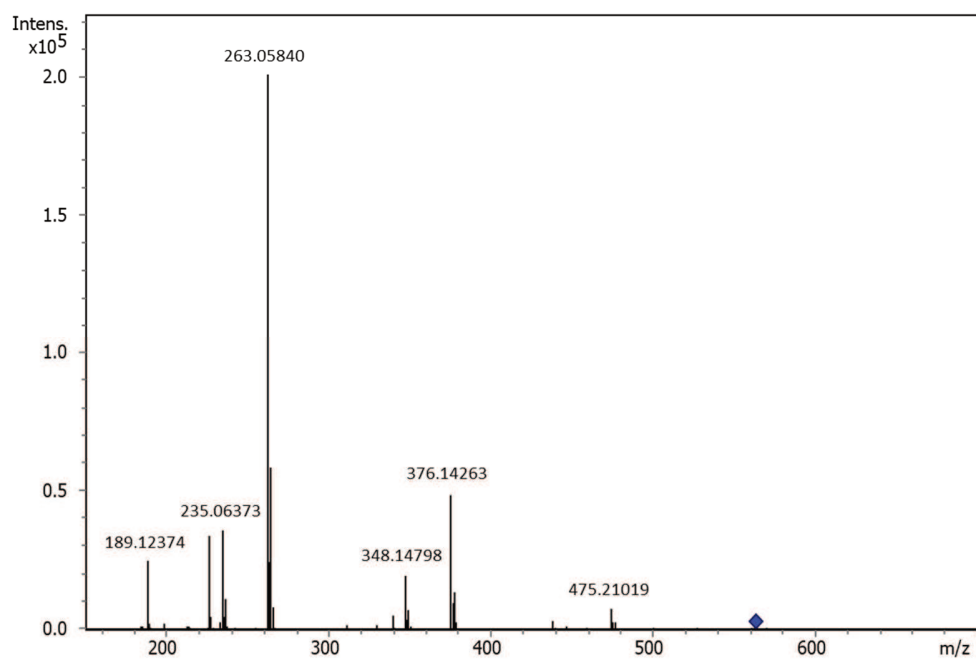


Figure S8. Tandem MS spectrum of **8**.

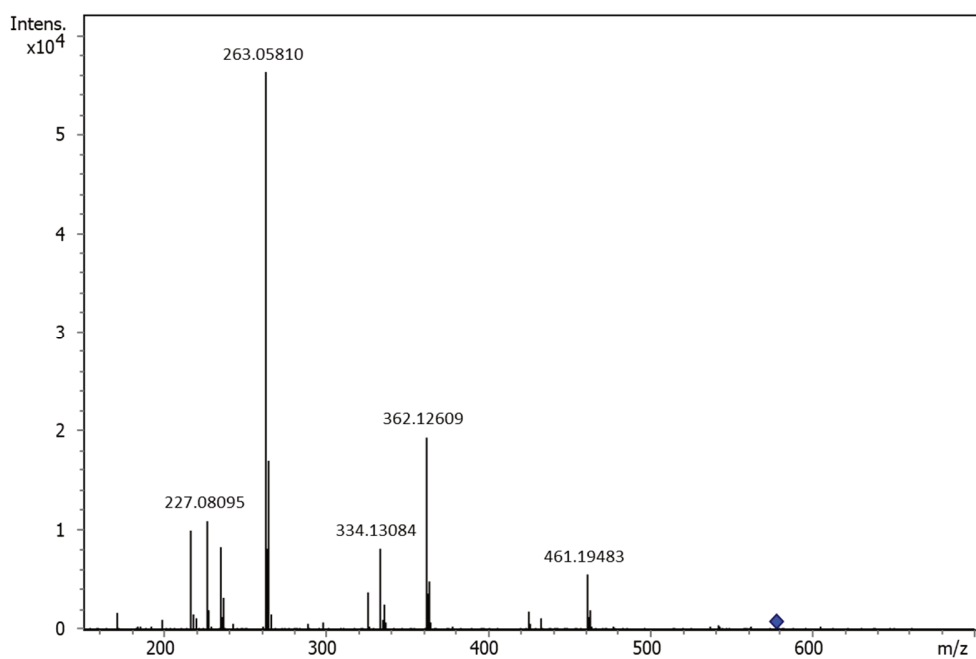


Figure S9. Tandem MS spectrum of 9.

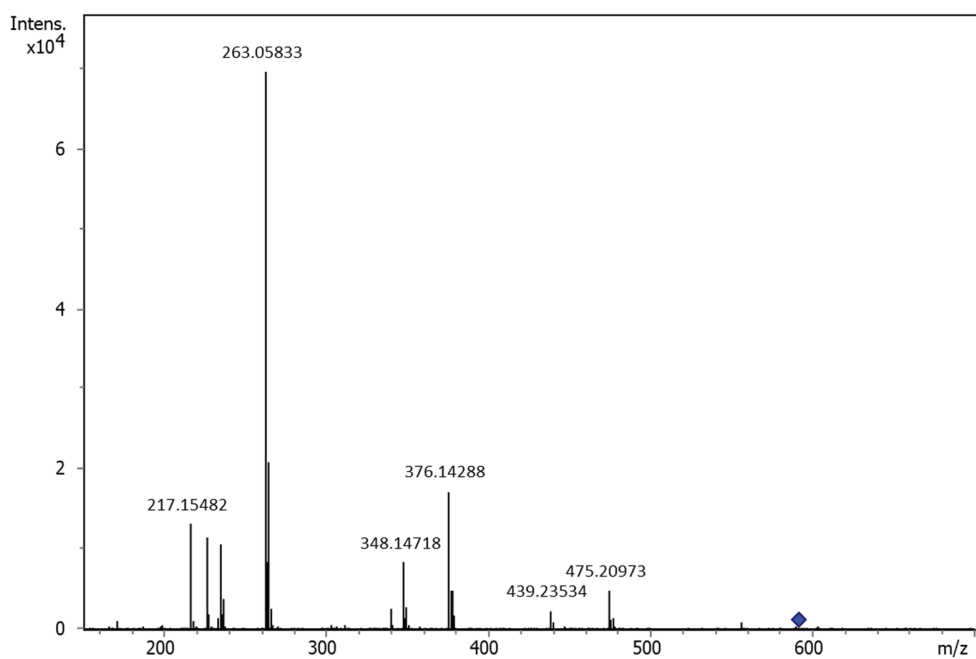


Figure S10. Tandem MS spectrum of 10.

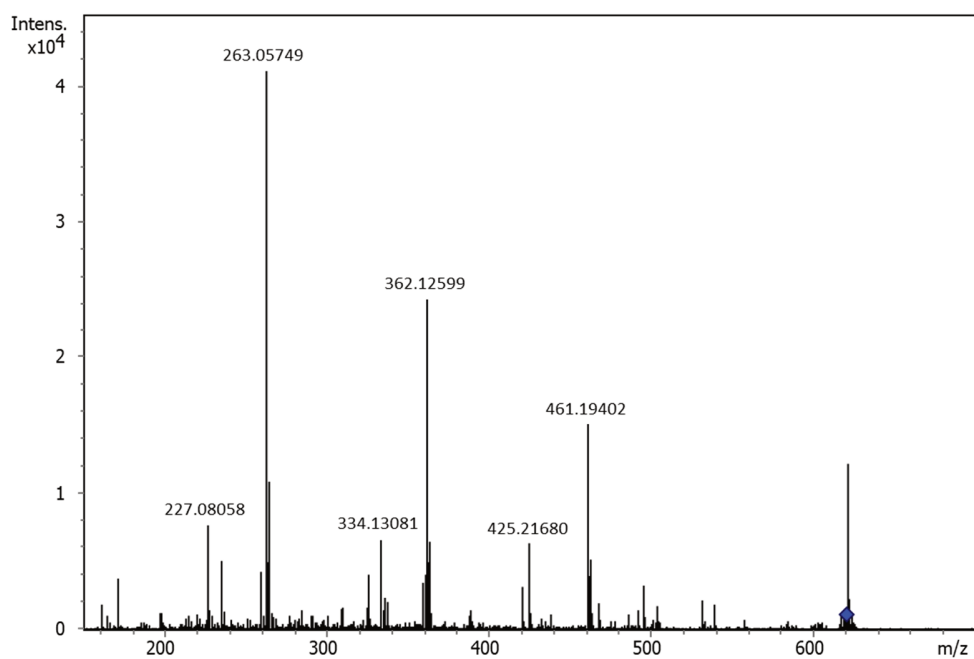


Figure S11. Tandem MS spectrum of 11.

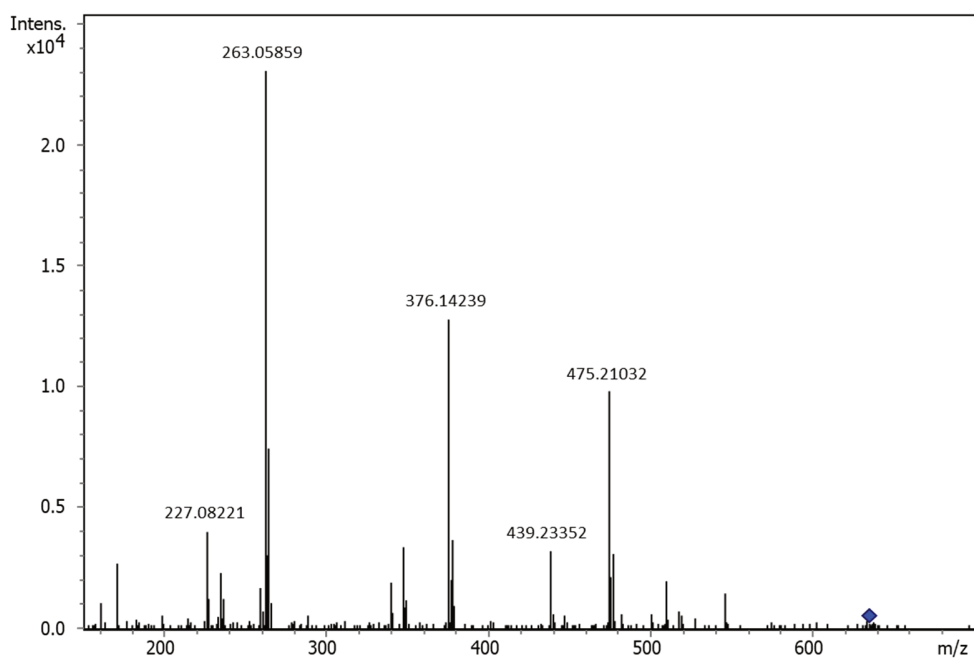


Figure S12. Tandem MS spectrum of 12.

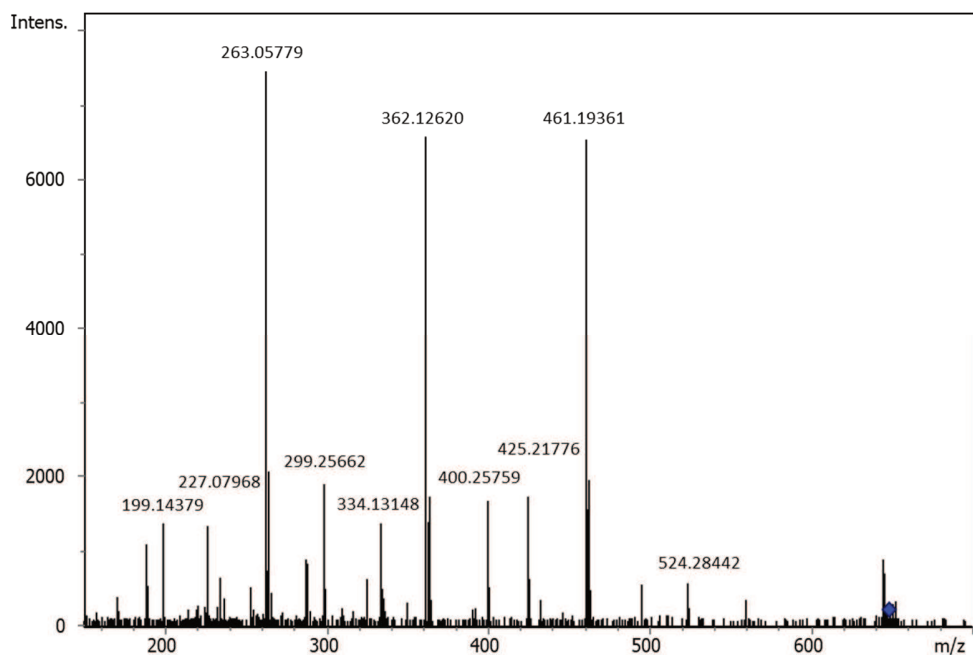


Figure S13. Tandem MS spectrum of 13.

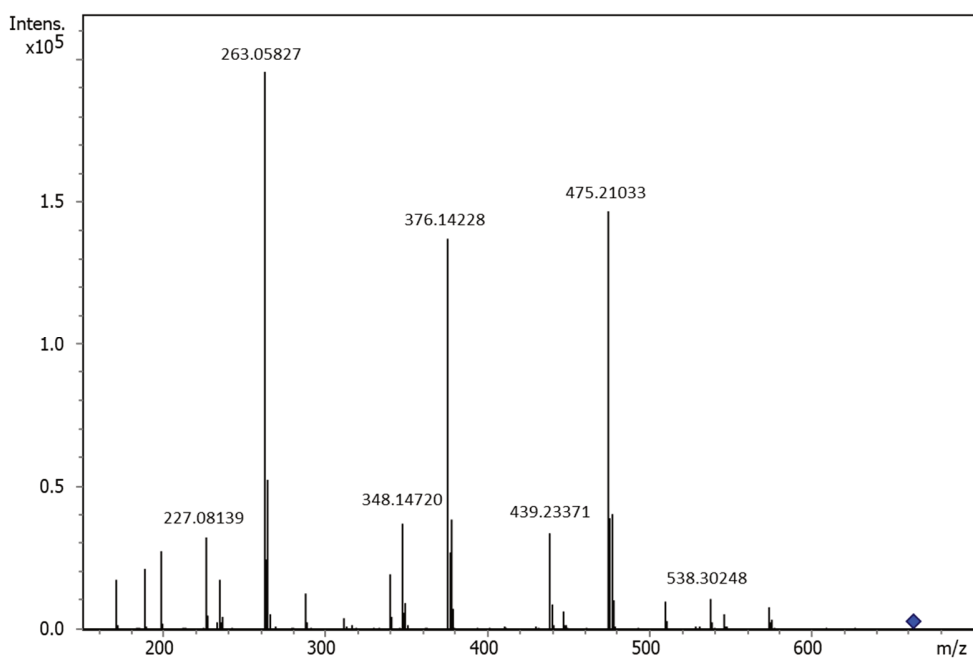


Figure S14. Tandem MS spectrum of 14.

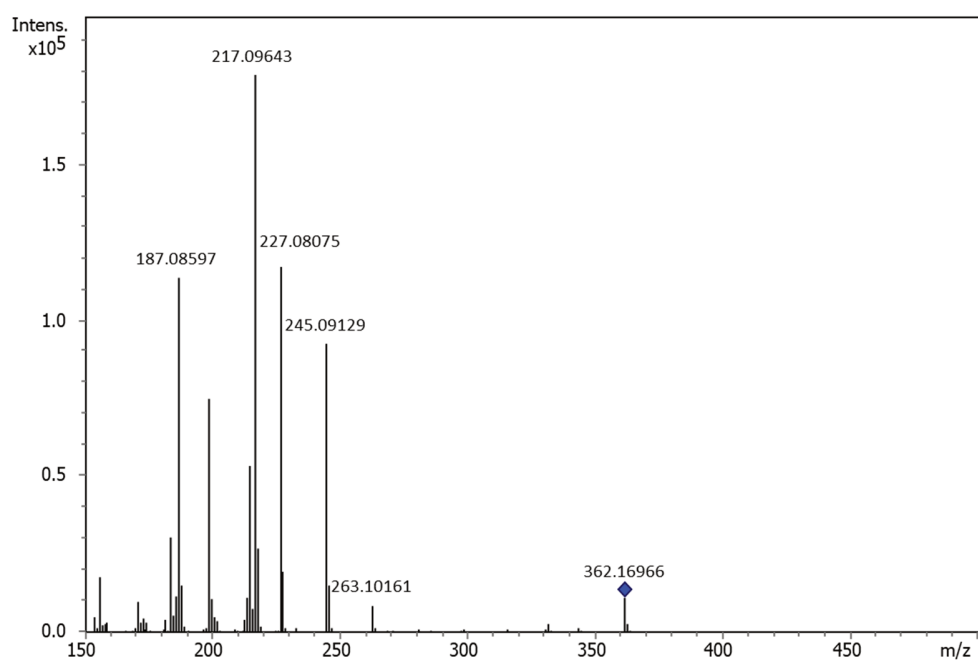


Figure S15. Tandem MS spectrum of 15.

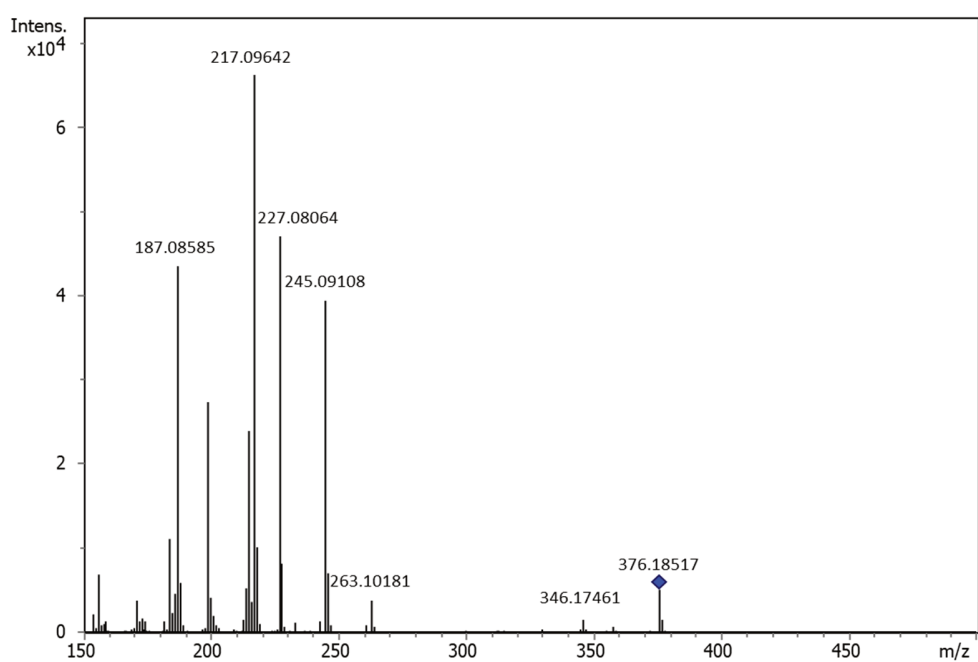


Figure S16. Tandem MS spectrum of 16.

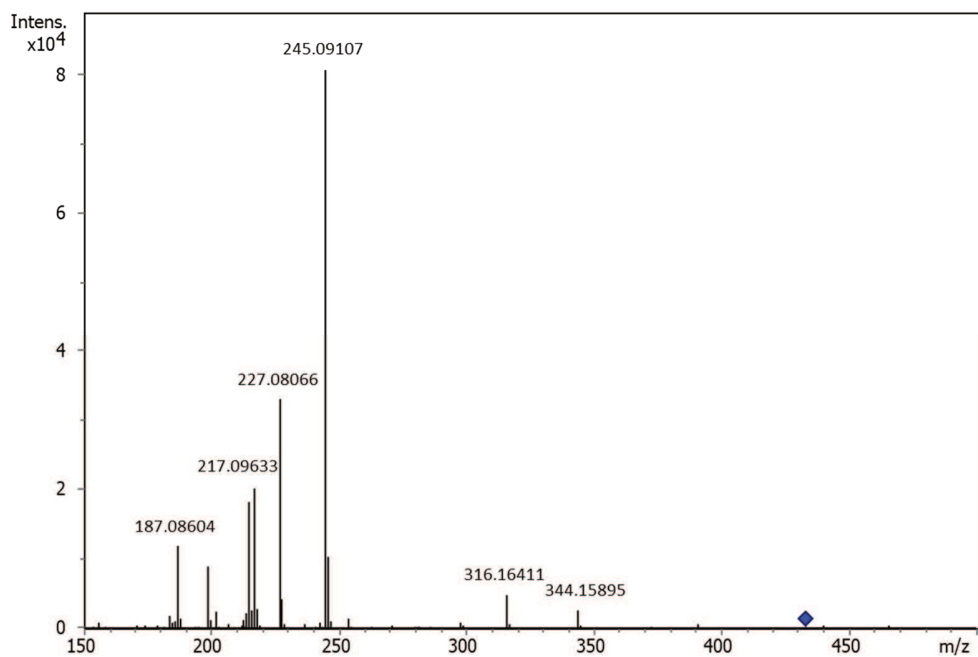


Figure S17. Tandem MS spectrum of 17.

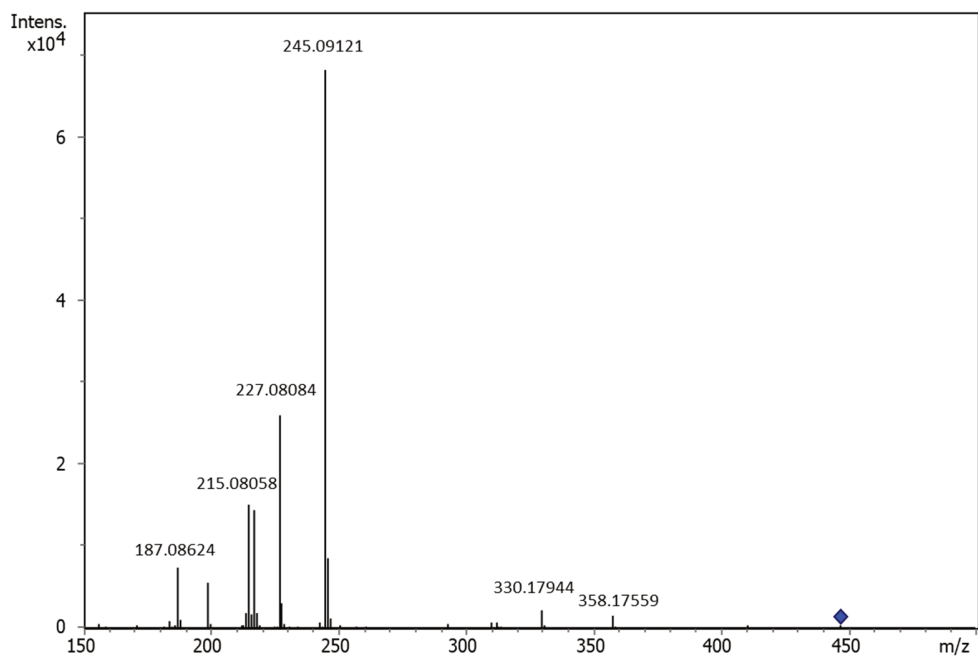


Figure S18. Tandem MS spectrum of 18.



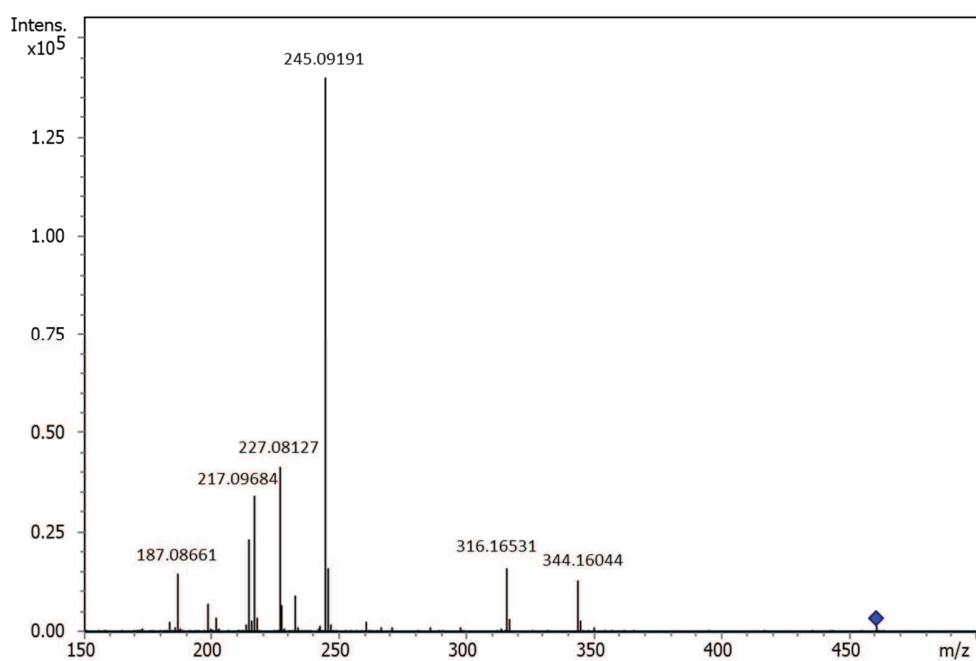


Figure S19. Tandem MS spectrum of **19**.

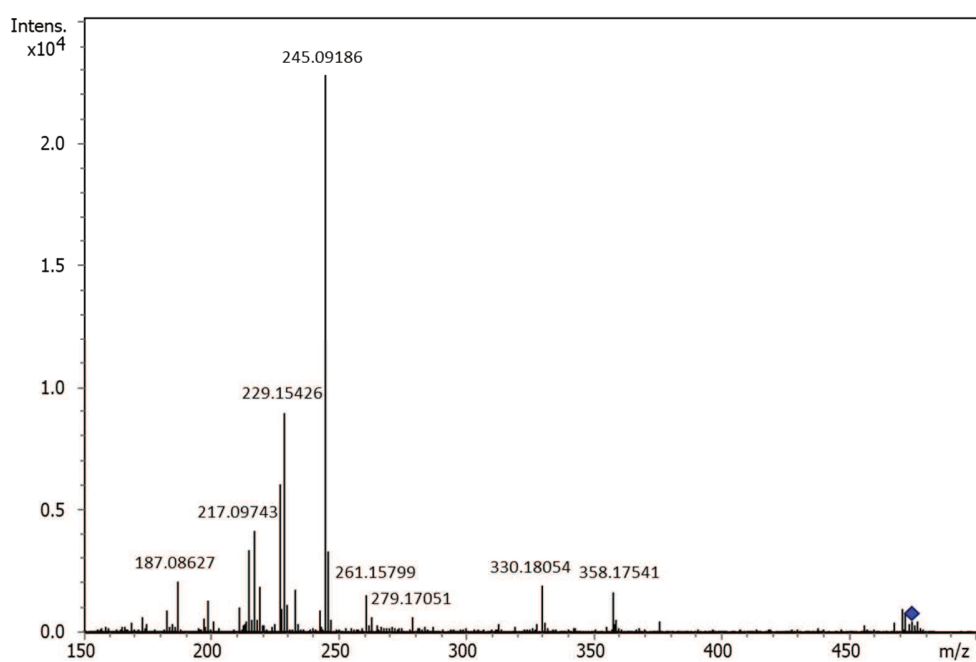


Figure S20. Tandem MS spectrum of **20**.

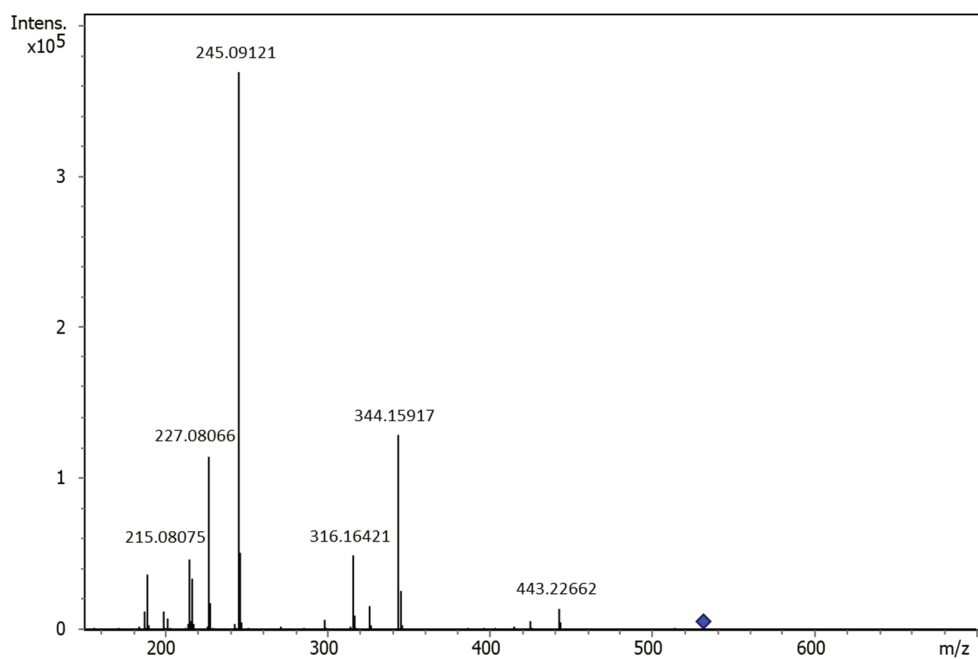


Figure S21. Tandem MS spectrum of 21.

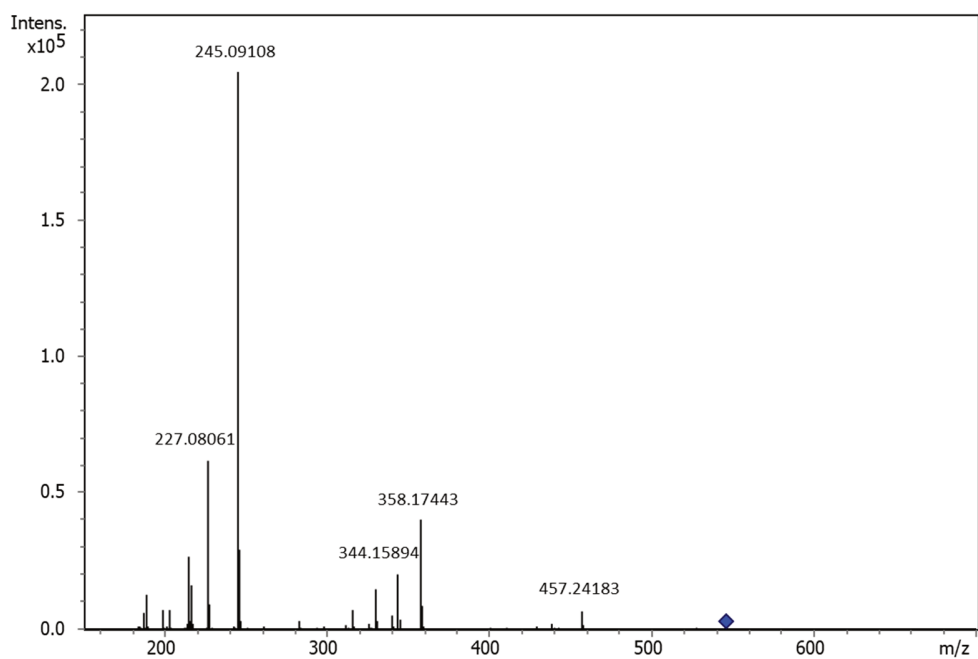


Figure S22. Tandem MS spectrum of 22.

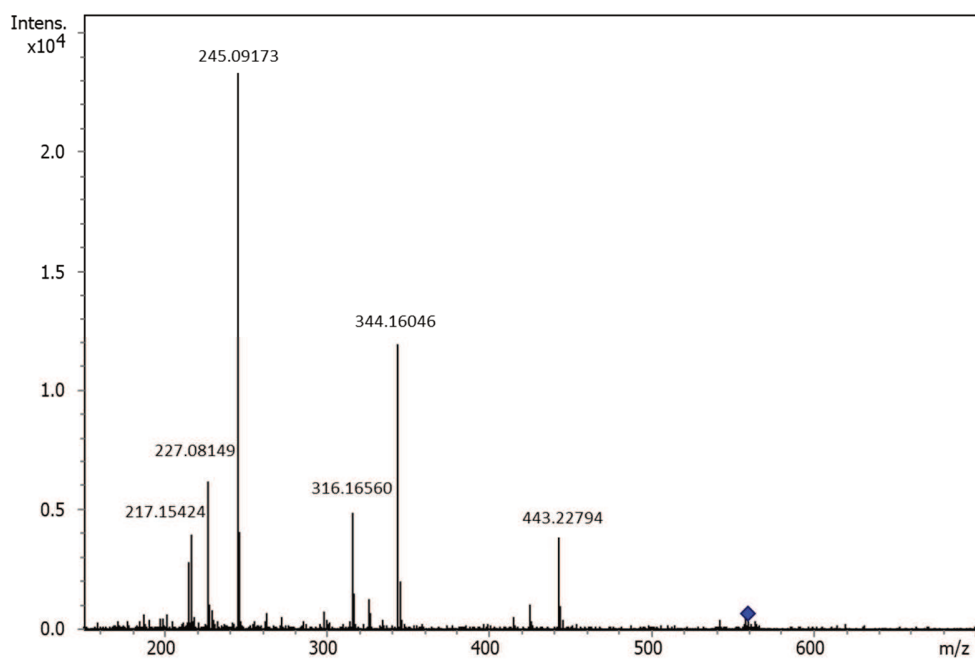


Figure S23. Tandem MS spectrum of **23**.

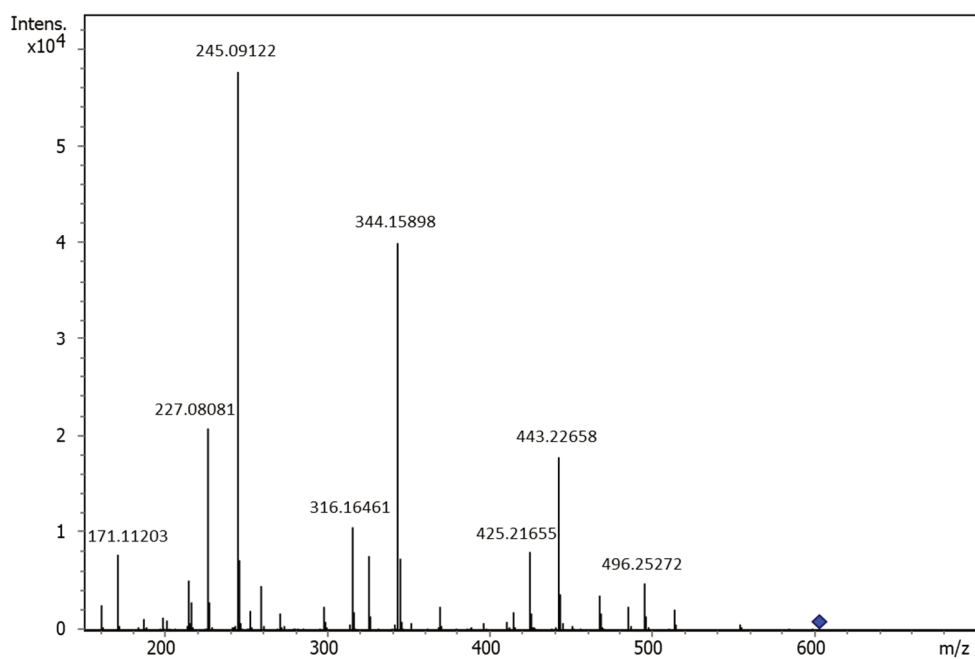


Figure S24. Tandem MS spectrum of **25**.

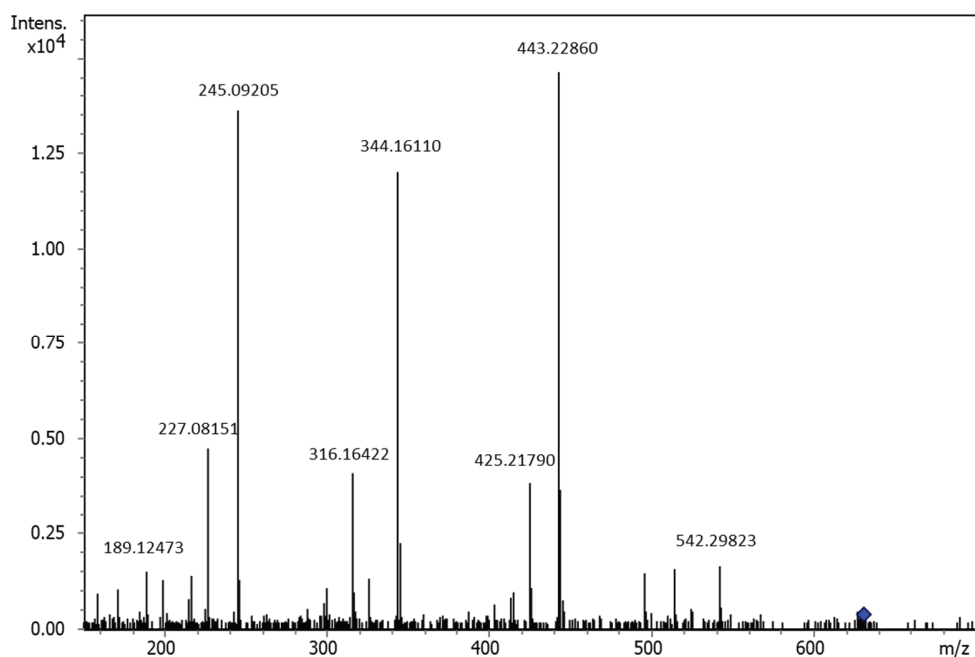


Figure S25. Tandem MS spectrum of **27**.

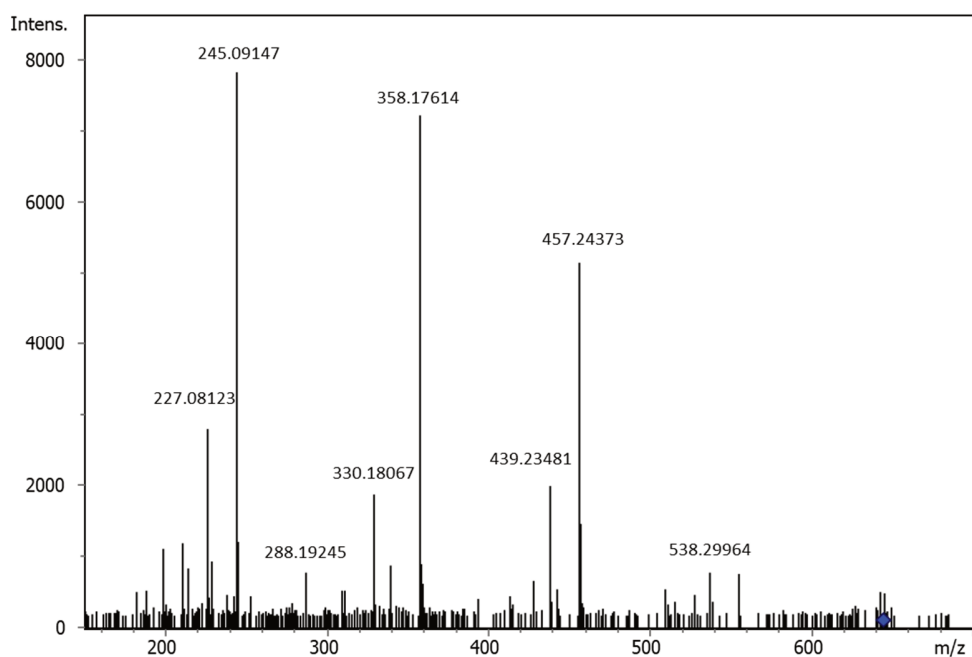
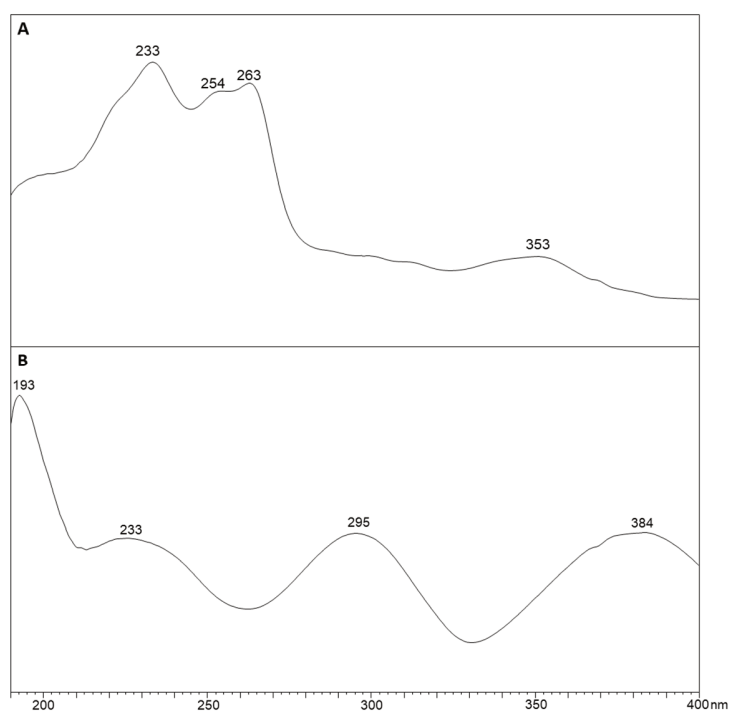


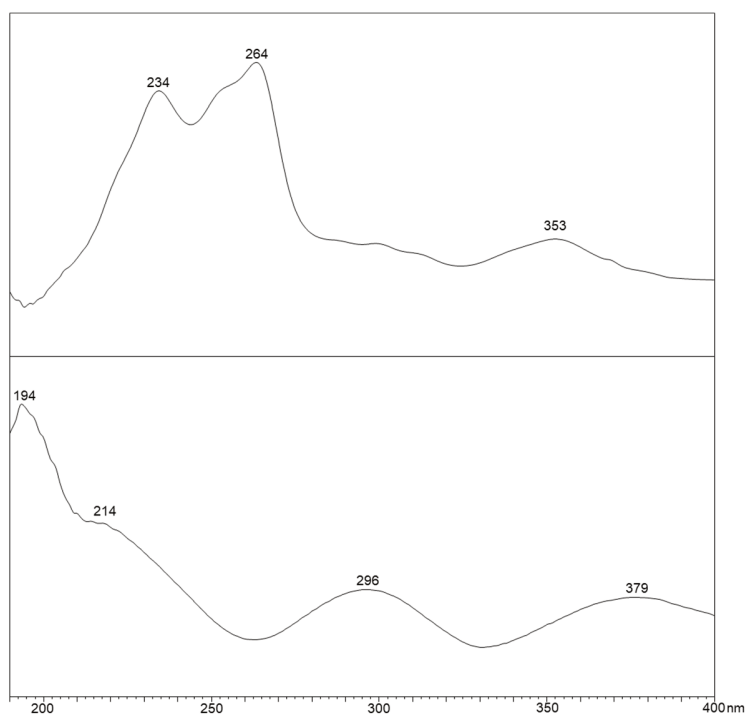
Figure S26. Tandem MS spectrum of **28**.

Compounds **24** ( $m/z = 574.324$ ,  $C_{29}H_{44}N_5O_7$ ) and **26** ( $m/z = 617.329$ ,  $C_{30}H_{45}N_6O_8$ ) were too low in abundance to be selected as precursors for tandem MS spectrum acquisition.

## S 3.2.2 UV/Vis Spectra of the Most Abundant Compounds



**Figure S27.** UV/Vis spectrum of **5** in water/acetonitrile mixture with 0.1% formic acid. A: Isomer at RT = 7.23 min, B: Isomer at RT = 7.89 min (Table 1).

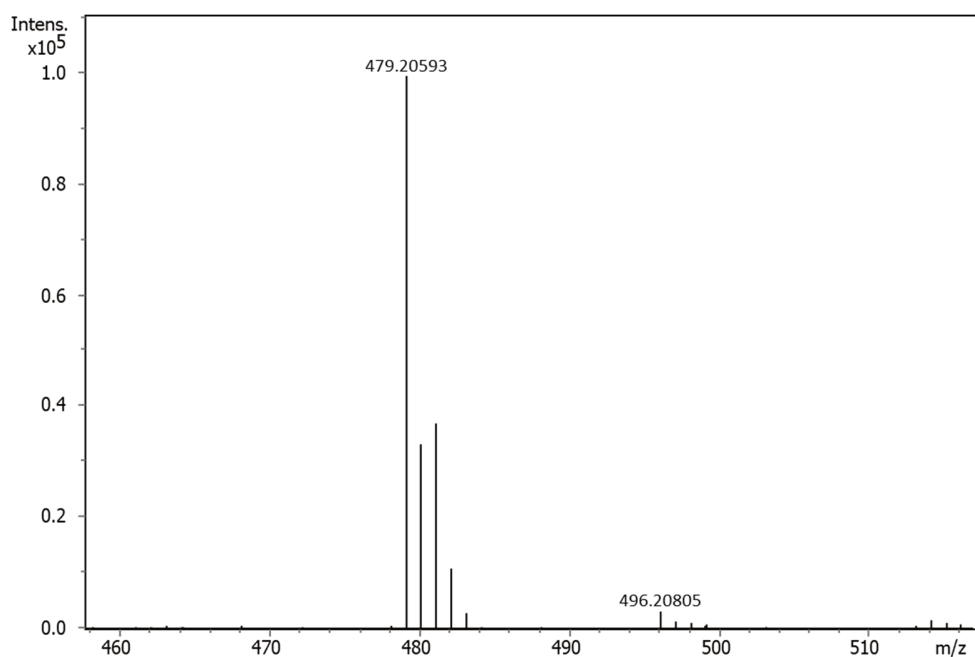


**Figure S28.** UV/Vis spectrum of **6** in water/acetonitrile mixture with 0.1% formic acid. A: Isomer at RT = 7.77 min, B: Isomer at RT = 8.43 min (Table 1).

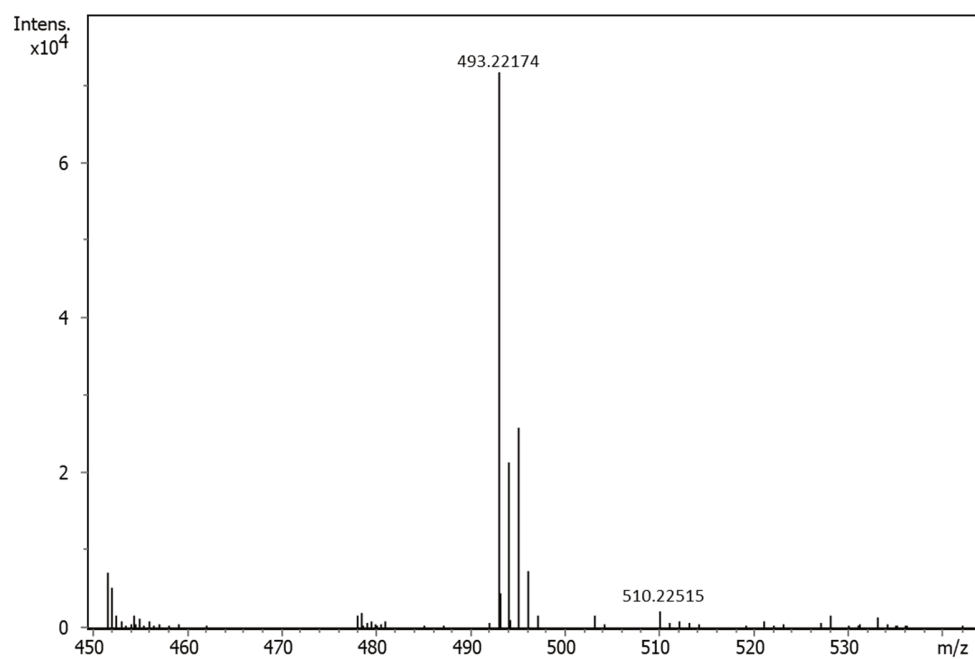
## S 3.2.3 Stable Isotope Feeding Experiments

**Table S6.** Stable-isotope labeled building blocks used for feeding experiments and expected mass shifts from incorporation of one unit. Highlighted in red are building blocks that showed incorporation into at least one of the novel compounds.

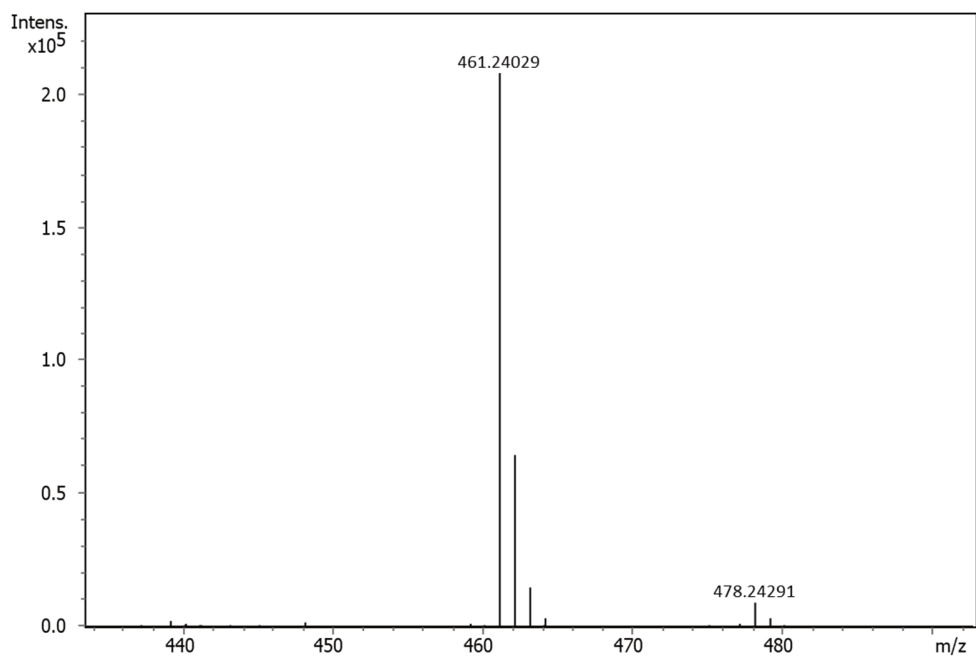
Building Block	Label	Expected Mass Shift per Incorporated Unit
L-Alanine	d2	+ 2 Da
L-Arginine	$^{13}\text{C}_6, ^{15}\text{N}_4$	+ 10 Da
L-Asparagine	$^{13}\text{C}_4, ^{15}\text{N}_2$	+ 6 Da
L-Aspartate	$^{13}\text{C}_4, ^{15}\text{N}_1$	+ 5 Da
L-Glutamine	$^{13}\text{C}_5, ^{15}\text{N}_2$	+ 7 Da
L-Glutamate	$^{15}\text{N}_1$ or d5	+ 1 Da or + 5 Da
L-Glycine	d2	+ 2 Da
L-Histidine	$^{15}\text{N}_3$	+ 3 Da
L-Isoleucine	d10	+ 10 Da
L-Leucine	d3	+ 3 Da
L-Lysine	$^{13}\text{C}_6, ^{15}\text{N}_2$	+ 8 Da
L-Methionine	$^{13}\text{C}_5, ^{15}\text{N}_1$	+ 6 Da
L-Proline	$^{15}\text{N}_1$	+ 1 Da
L-Serine	$^{13}\text{C}_1$	+ 1 Da
L-Threonine	$^{13}\text{C}_4, ^{15}\text{N}_1$	+ 5 Da
L-Tryptophane	$^{15}\text{N}_2$ or d8	+ 2 Da or + 8 Da
L-Tyrosine	d4	+ 4 Da
L-Valine	d8	+ 8 Da
Acetate	$^{13}\text{C}_1$ or $^{13}\text{C}_2$	+ 1 Da or + 2 Da
Propionate	$^{13}\text{C}_1$	+ 1 Da
Succinate	d4	+ 4 Da
Glucose	$^{13}\text{C}_1$	+ 1 Da
Glycerol	$^{13}\text{C}_3$	+ 3 Da



**Figure S29.** Isotope pattern of 5 in VY/2 medium (control).

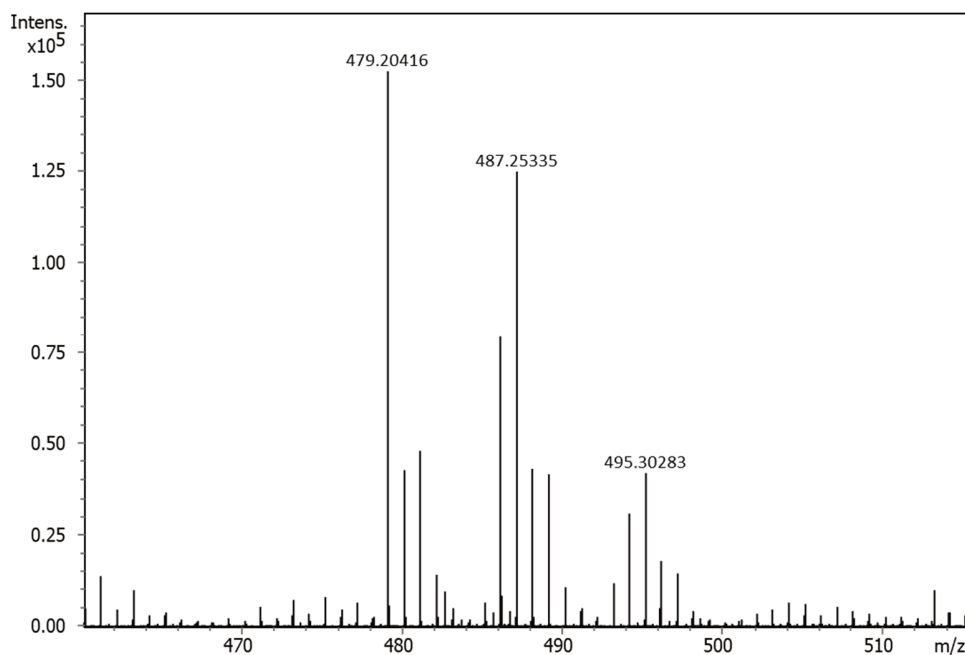


**Figure S30.** Isotope pattern of 6 in VY/2 medium (control).

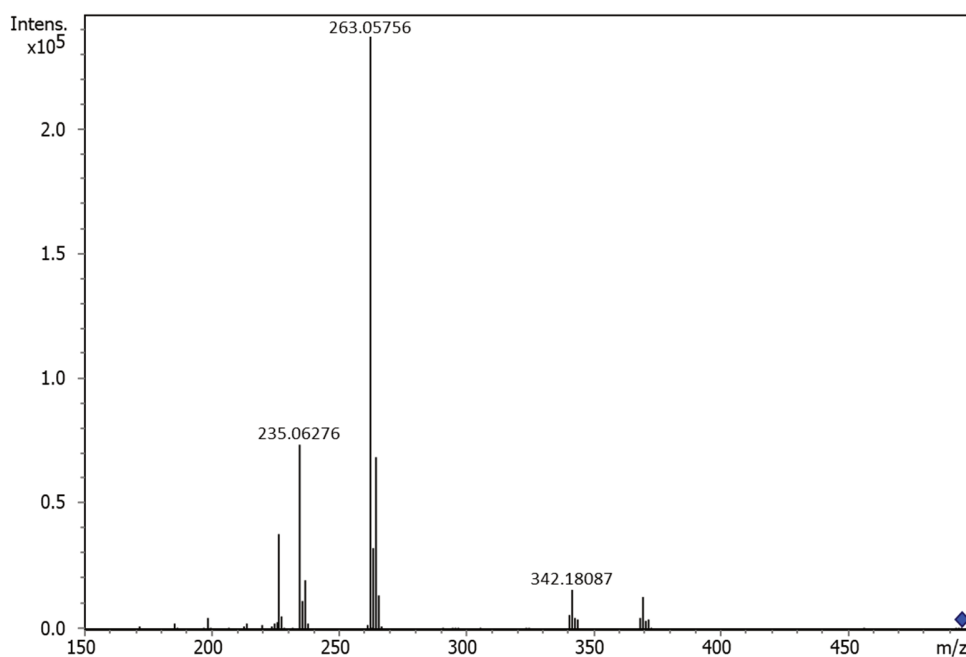


**Figure S31.** Isotope pattern of **19** in VY/2 medium (control).

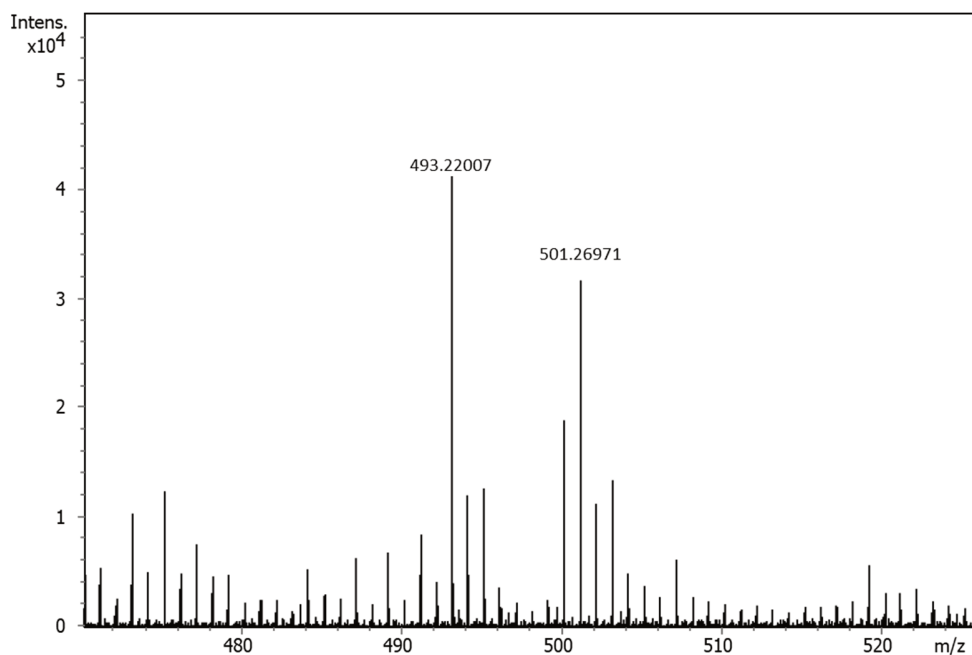




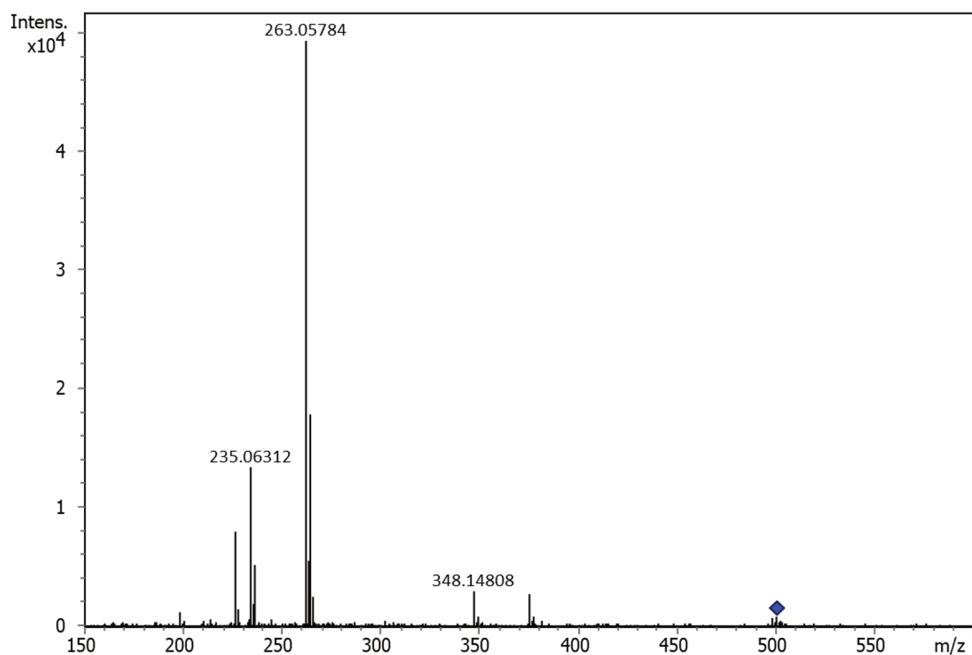
**Figure S32.** Isotope pattern of **5** in VY/2 medium substituted with L-valine-d8. Mass shifts of +8.05 and +16.10 Da suggest incorporation of two labeled valine units.



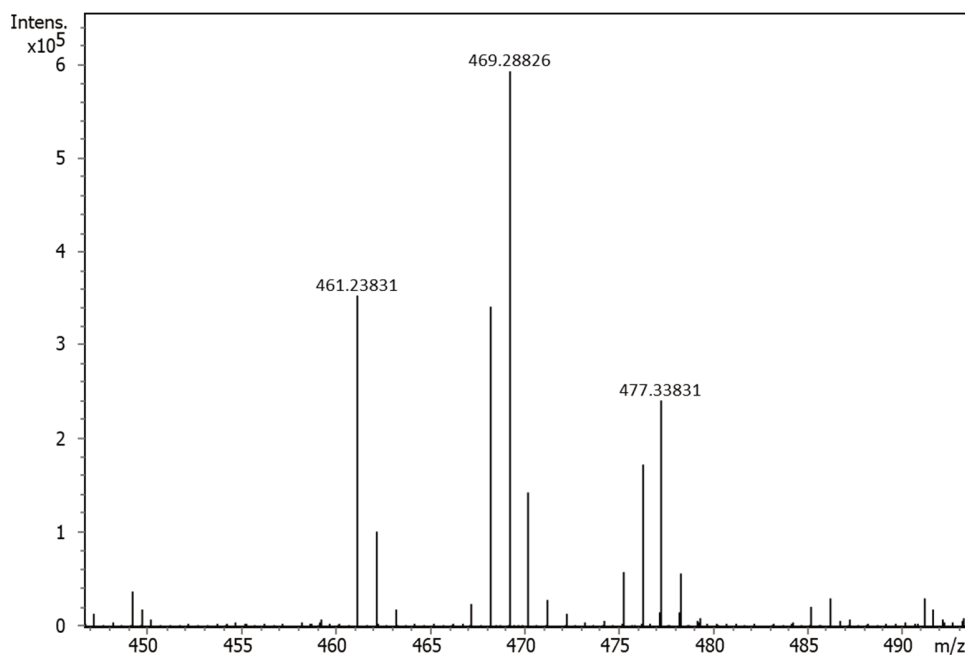
**Figure S33.** Tandem MS spectrum of m/z = 495.30 (see Figure S32). Mass shifts in comparison to Figure S5 are observed for the fragments 334.13 (+8.05) and 362.13 (+8.05). All other fragment masses remain unchanged.



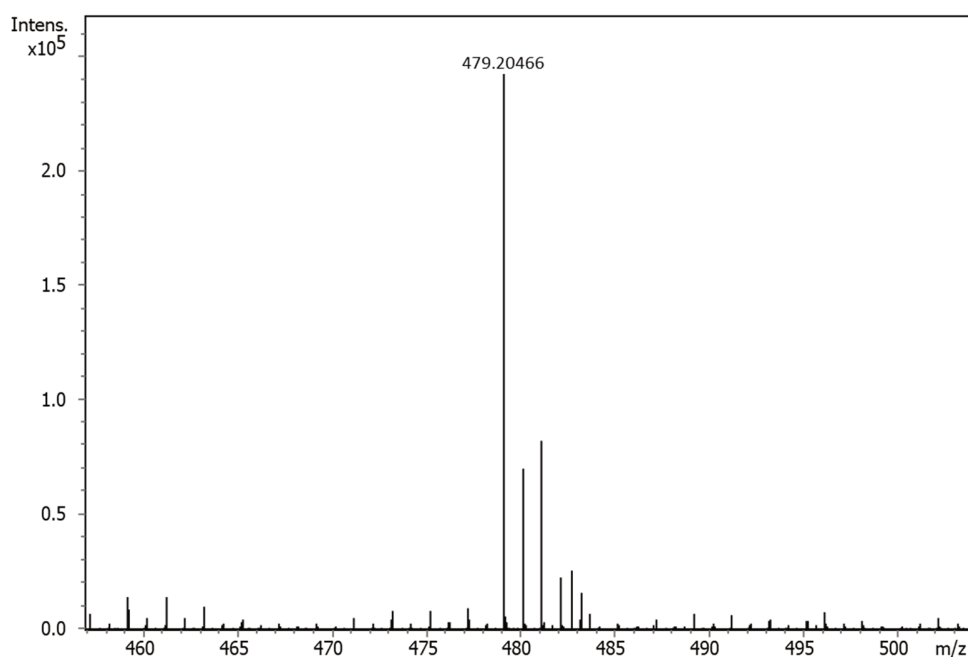
**Figure S34.** Isotope pattern of **6** in VY/2 medium substituted with *L*-valine- $d_8$ . A mass shift of +8.05 Da suggests incorporation of one labeled valine unit.



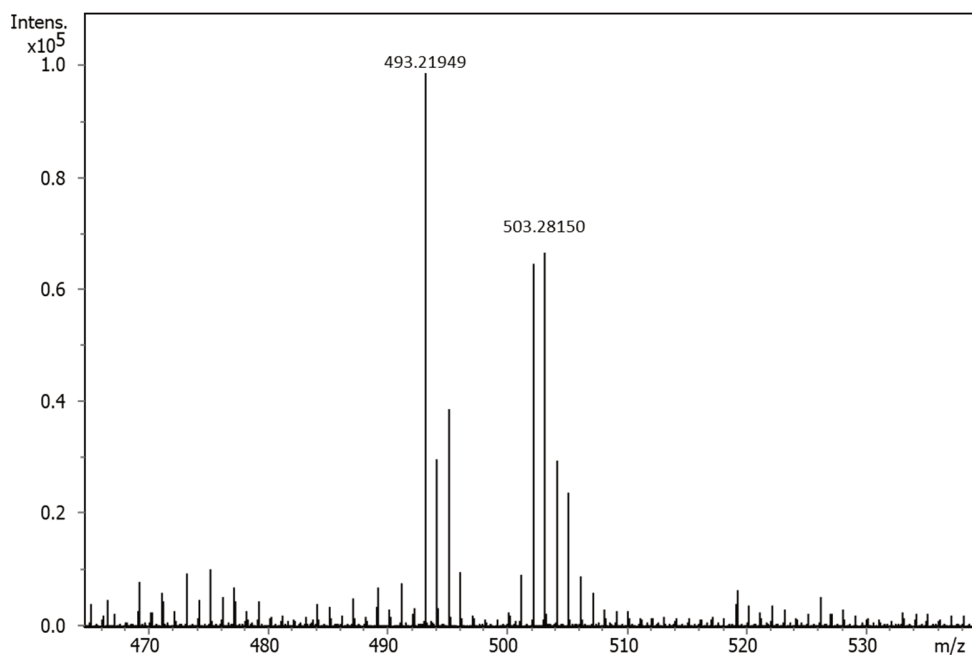
**Figure S35.** Tandem MS spectrum of  $m/z = 501.27$  (see Figure S34). No mass shifts in comparison to Figure S6 are observed suggesting that the labeled valine moiety is the first fragment to dissociate in the collision cell.



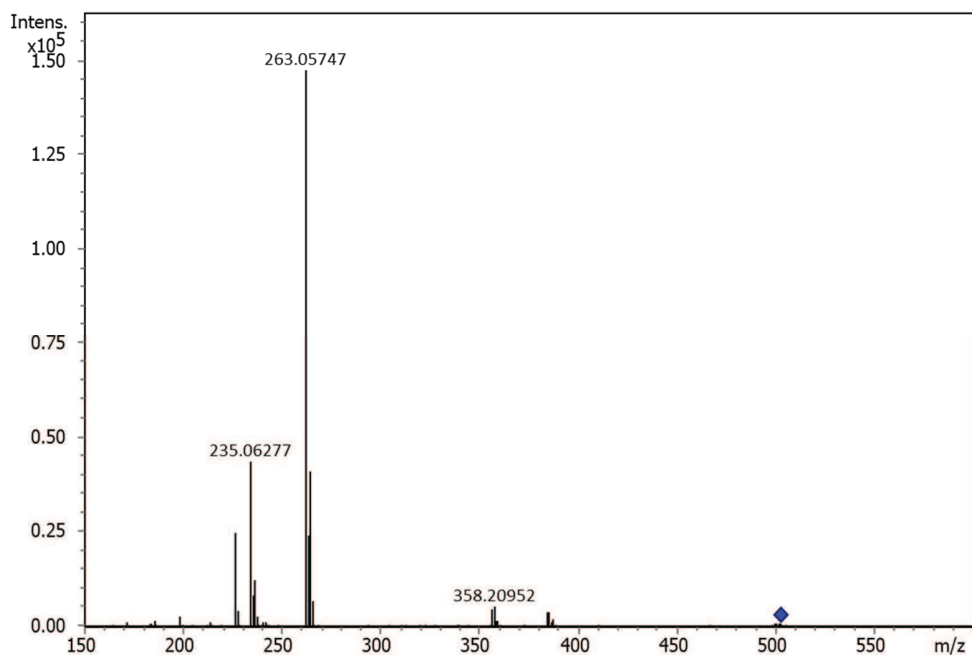
**Figure S36.** Isotope pattern of **19** in VY/2 medium substituted with *L*-valine-*d*8. Mass shifts of +8.05 and +16.10 Da suggest incorporation of two labeled valine units. The labeled precursor has not been picked for tandem MS fragmentation.



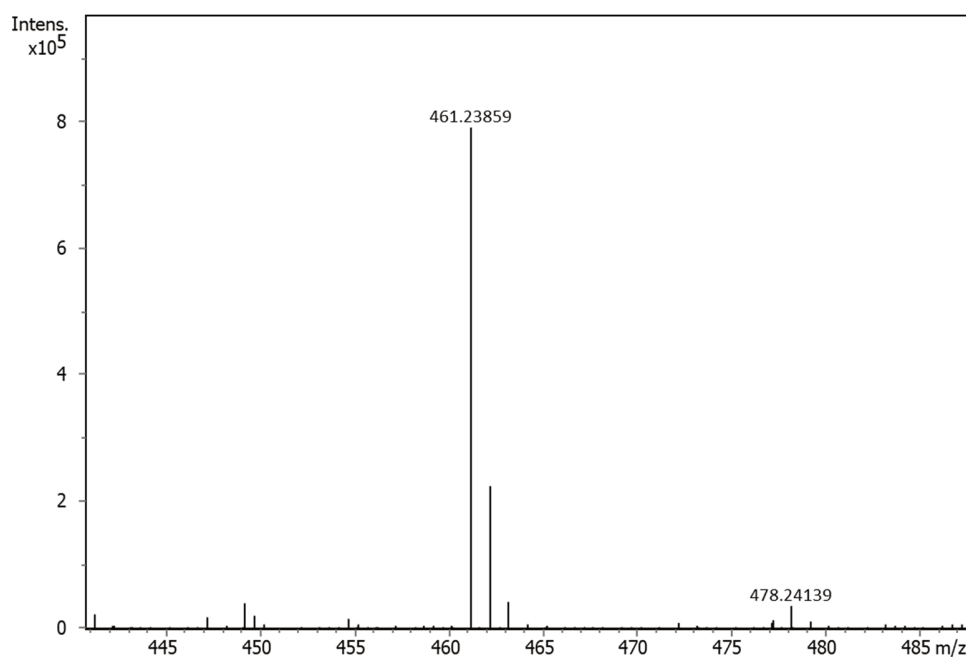
**Figure S37.** Isotope pattern of **5** in VY/2 medium substituted with *L*-isoleucine-*d*10. No mass shift and thus no incorporation was observed.



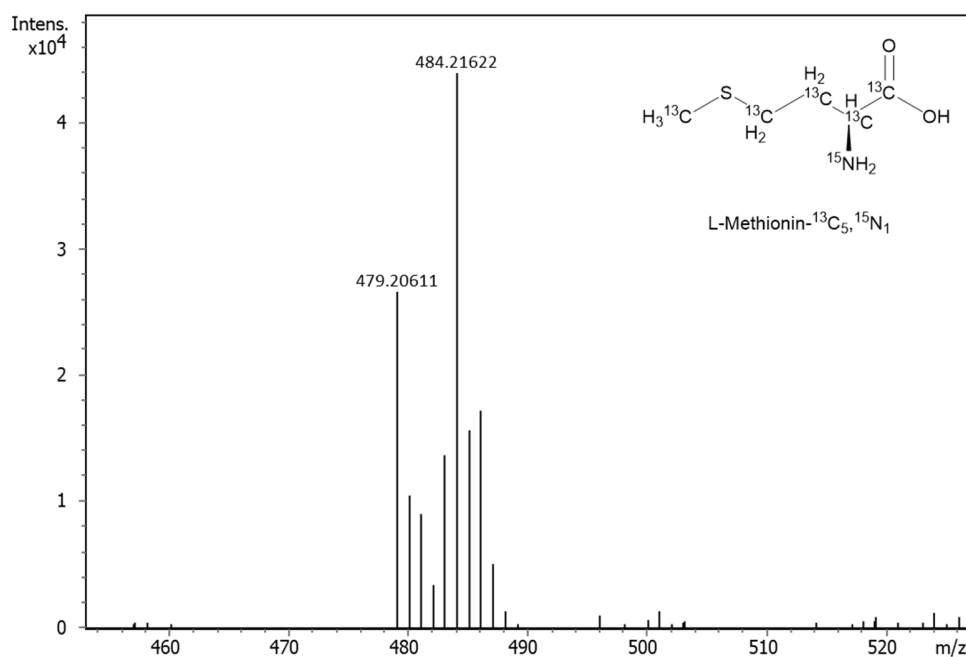
**Figure S38.** Isotope pattern of **6** in VY/2 medium substituted with L-isoleucine-d10. A mass shift of +10.06 Da suggests incorporation of one labeled isoleucine unit.



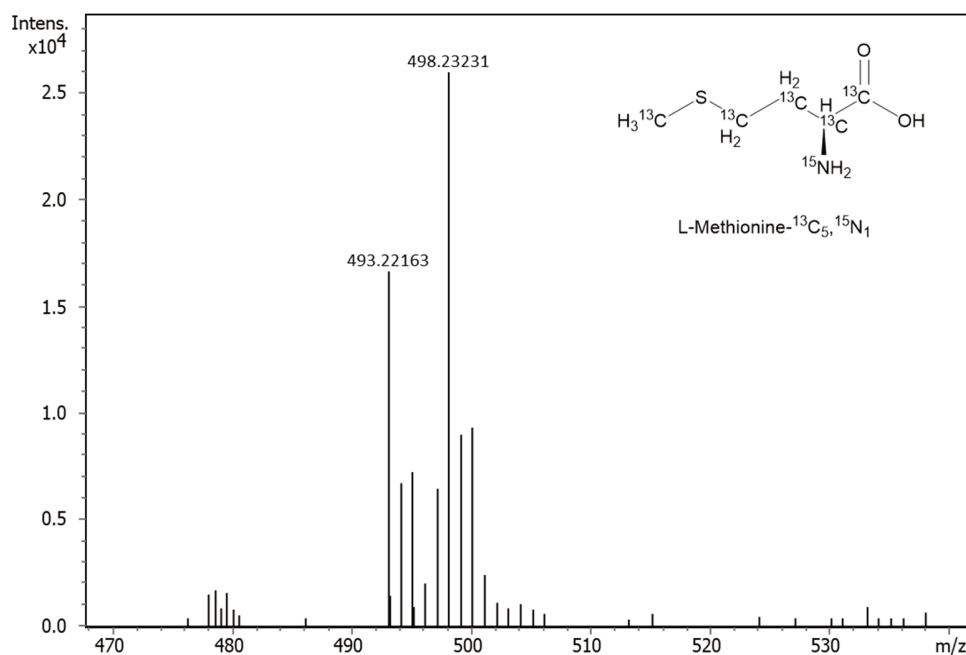
**Figure S39.** Tandem MS spectrum of m/z = 503.28 (see Figure S38). Mass shifts in comparison to Figure S6 are observed for the fragments 348.15 (+10.06) and 376.14 (+10.06). All other fragment masses remain unchanged.



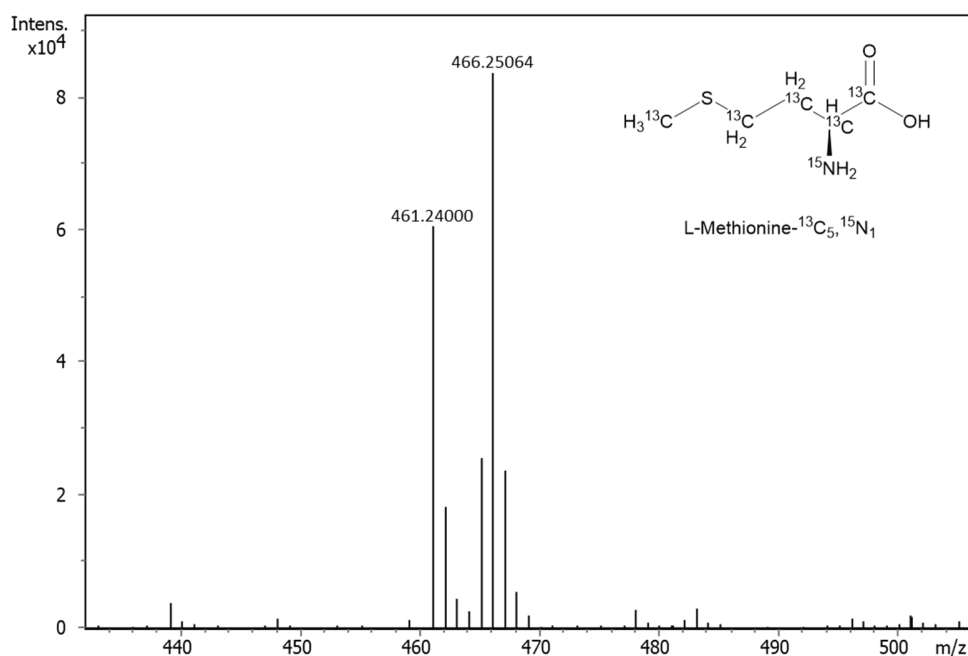
**Figure S40.** Isotope pattern of **19** in VY/2 medium substituted with L-isoleucine-d10. No mass shift and thus no incorporation was observed.



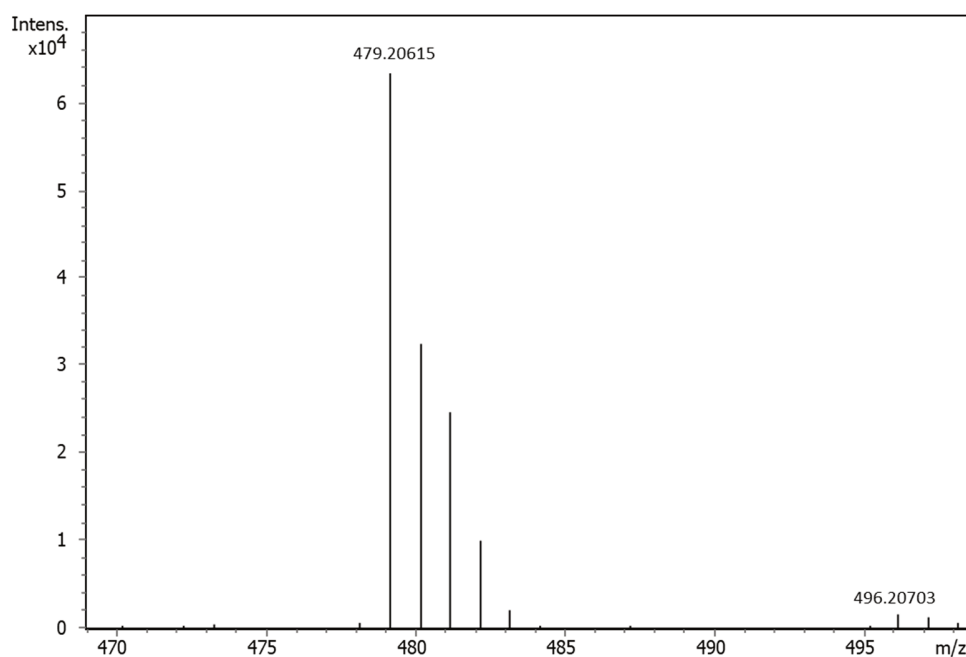
**Figure S41.** Isotope pattern of **5** in VY/2 medium substituted with L-methionine- $^{13}\text{C}_5, ^{15}\text{N}_1$ . A mass shift of +5.01 Da is observed. The labeled precursor has not been picked for tandem MS fragmentation.



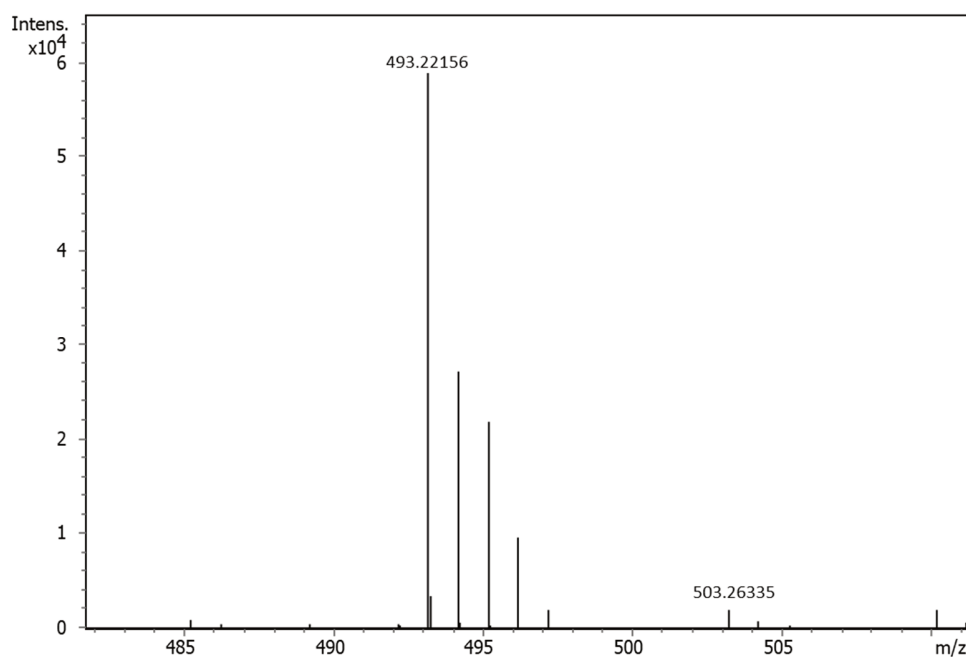
**Figure S42.** Isotope pattern of **6** in VY/2 medium substituted with L-methionine-<sup>13</sup>C<sub>5</sub>, <sup>15</sup>N<sub>1</sub>. A mass shift of +5.01 Da is observed. The labeled precursor has not been picked for tandem MS fragmentation.



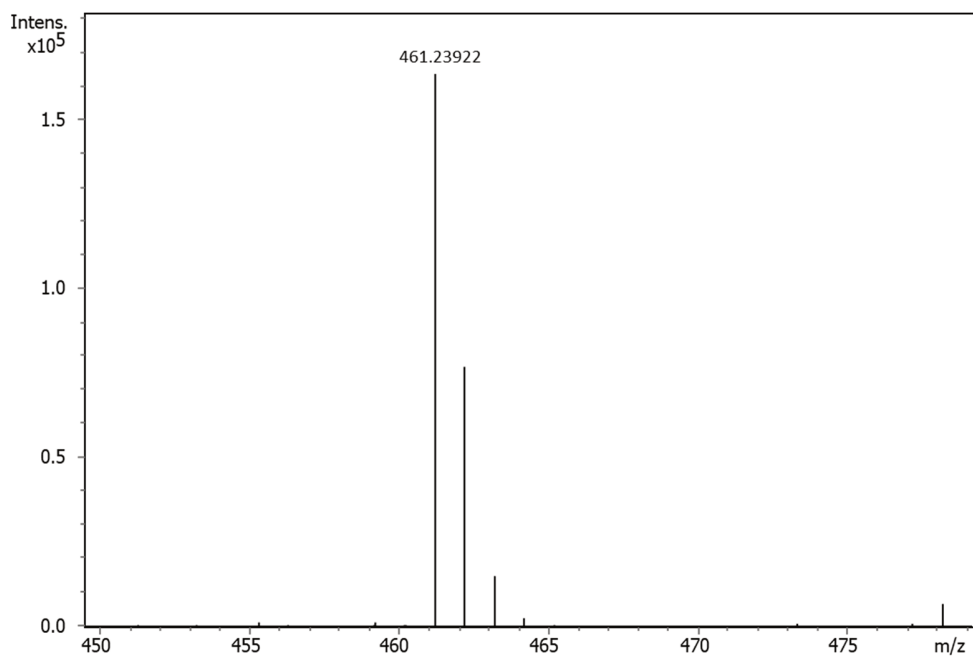
**Figure S43.** Isotope pattern of **19** in VY/2 medium substituted with L-methionine-<sup>13</sup>C<sub>5</sub>, <sup>15</sup>N<sub>1</sub>. A mass shift of +5.01 Da is observed. The labeled precursor has not been picked for tandem MS fragmentation.



**Figure S44.** Isotope pattern of **5** in VY/2 medium substituted with *L*-serine- $^{13}\text{C}_1$ . The relative intensity of the second isotope peak is slightly increased. The labeled precursor has not been picked for tandem MS fragmentation.



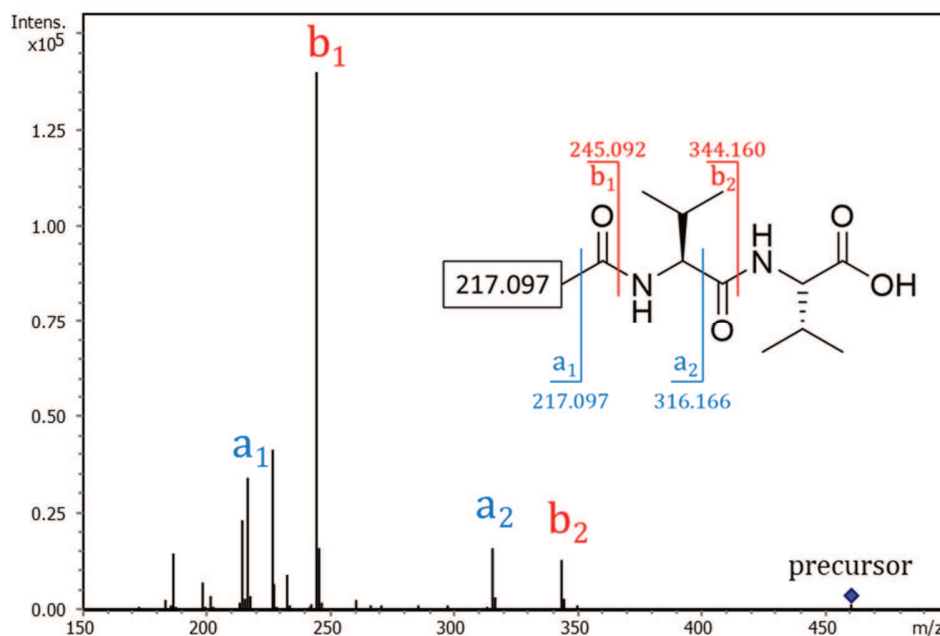
**Figure S45.** Isotope pattern of **6** in VY/2 medium substituted with *L*-serine- $^{13}\text{C}_1$ . The relative intensity of the second isotope peak is slightly increased. The labeled precursor has not been picked for tandem MS fragmentation.



**Figure S46.** Isotope pattern of **19** in VY/2 medium substituted with L-serine-<sup>13</sup>C<sub>1</sub>. The relative intensity of the second isotope peak is slightly increased. The labeled precursor has not been picked for tandem MS fragmentation.



## S 3.2.4 Proposed Peptide Sequences



**Figure S47.** Tandem MS spectrum of **19** annotated with putative *a* and *b* ions that were used for generation of the proposed peptide sequences shown below; the respective *y* ions were not detected. Ion nomenclature from [1].

**Table S7.** Proposed peptide chains based on tandem MS fragmentation. [Core] represents the proposed *b*<sub>1</sub> fragment of 263.06 (chlorinated core, compounds **1-14**) or 245.09 (hydroxylated core, compounds **15-28**) as presented above. Amino acids in brackets had matching fragments in the tandem MS spectra, but did not have their incorporation confirmed by stable-isotope feeding.

Compound #	Proposed peptide sequence
<b>1, 15</b>	[Core]-Val
<b>2, 16</b>	[Core]-Ile
<b>3, 17</b>	[Core]-Val-(Ala)
<b>4, 18</b>	[Core]-Ile-(Ala)
<b>5, 19</b>	[Core]-Val-Val
<b>6, 20</b>	[Core]-Ile-Val
<b>7, 21</b>	[Core]-Val-Val-(Ala)
<b>8, 22</b>	[Core]-Ile-Val-(Ala)
<b>9, 23</b>	[Core]-Val-Val-Val
<b>10, 24</b>	[Core]-Ile-Val-Val
<b>11, 25</b>	[Core]-Val-Val-(Ala)-(Abu)
<b>12, 26</b>	[Core]-Ile-Val-(Ala)-(Abu)
<b>13, 27</b>	[Core]-Val-Val-Val-(Ala)
<b>14, 28</b>	[Core]-Ile-Val-Val-(Ala)

Abu: α-aminobutyric acid

### S 3.3 Analysis of the Putative BGCs

Every coding DNA sequence (CDS) in the putative biosynthetic gene cluster of MCy9080 / *Corallocooccus coralloides* DSM2259 was extracted translated and searched with the blastp algorithm against the NCBI RefSeq non-redundant protein sequence database.[2] Locus tags COCOR\_05735 – COCOR\_05741 represent the CDS within the cluster borders predicted by the ClusterFinder algorithm [3], while COCOR\_05722 – COCOR\_05734 are located upstream and COCOR\_05742 – COCOR\_05750 are located downstream of the predicted borders.

**Table S8.** Tabulated blastP results for the CDS regions present in the MCy9080 putative biosynthetic gene cluster.

Locus Tag	Length [AA]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Accession Nr.
<b>COCOR_05735</b> <i>cdsA</i>	1815	Amino acid adenylation domain-containing protein [Corallocooccus sp. CA054B]	98.9 / 1815	WP_120588875
<b>COCOR_05736</b> <i>cdsB</i>	1552	Type I polyketide synthase [Corallocooccus sp. CA054B]	99.1 / 1552	WP_120588876
<b>COCOR_05737</b> <i>cdsC</i>	1990	Acyltransferase domain-containing protein [Corallocooccus sp. CA054B]	98.6 / 1989	WP_120588877
<b>COCOR_05738</b> <i>cdsD</i>	1107	Amino acid adenylation domain-containing protein [Corallocooccus coralloides CA044C]	99.1 / 1107	WP_171425671
<b>COCOR_05739</b> <i>cdsE</i>	1511	Type I polyketide synthase [Corallocooccus sp. CA054B]	98.0 / 1511	WP_158616635
<b>COCOR_05740</b> <i>cdsG</i>	467	MFS transporter [Corallocooccus sp. CA054B]	97.4 / 467	WP_120588879
<b>COCOR_05741</b> <i>cdsH</i>	812	Penicillin acylase family protein [Corallocooccus sp. CA054B]	97.5 / 812	WP_120588880

Table S9. Tabulated blastP results for the CDS regions upstream of the MCy9080 putative biosynthetic gene cluster.

Locus Tag	Length [AA]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Accession Nr.
<b>COCOR_05722</b>	629	M4 family metallopeptidase [Coralloccoccus coralloides CA044C]	92.4 / 629	WP_171425278
<b>COCOR_05723</b>	185	DUF2378 family protein [Coralloccoccus sp. CA054B]	99.5 / 185	WP_120591693
<b>COCOR_05724</b>	364	Hypothetical protein [Coralloccoccus sp. CA054B]	99.5 / 364	WP_120591692
<b>COCOR_05725</b>	325	Hypothetical protein [Coralloccoccus sp. AB011P]	96.8 / 279	WP_120563490
<b>COCOR_05726</b>	145	DUF2383 domain-containing protein [Coralloccoccus sp. CA054B]	97.9 / 145	WP_120591690
<b>COCOR_05727</b>	272	Xanthine dehydrogenase accessory protein XdhC [Coralloccoccus sp. CA054B]	99.3 / 272	WP_120591689
<b>COCOR_05728</b>	1277	Xanthine dehydrogenase molybdopterin binding subunit [Coralloccoccus coralloides B035]	98.3 / 1277	WP_128798441
<b>COCOR_05729</b>	285	Metallo-mystery pair system four-Cys motif protein [unclassified Coralloccoccus]	94.4 / 285	WP_120563494
<b>COCOR_05730</b>	382	di-heme enzyme [Coralloccoccus sp. CA054B]	98.7 / 382	WP_120588869
<b>COCOR_05731</b>	568	Hypothetical protein [Coralloccoccus sp. CA054B]	91.0 / 568	WP_120588886
<b>COCOR_05732</b>	447	Hypothetical protein [Coralloccoccus sp. CA054B]	92.5 / 438	WP_233585373
<b>COCOR_05733</b>	239	DUF1775 domain-containing protein [Coralloccoccus sp. CA054B]	98.3 / 239	WP_120588873
<b>COCOR_05734</b>	189	Hypothetical protein [Coralloccoccus sp. CA054B]	94.7 / 189	WP_147447112

**Table S10.** Tabulated blastP results for the CDS regions downstream of the MCy9080 putative biosynthetic gene cluster.

Locus Tag	Length [AA]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Accession Nr.
<b>COCOR_05742</b>	226	Crp/Fnr family transcriptional regulator [Coralloccoccus sp. CA054B]	99.6 / 226	WP_120588881
<b>COCOR_05743</b>	513	GAF domain-containing protein [Coralloccoccus sp. CA054B]	99.4 / 513	WP_120588882
<b>COCOR_05744</b>	672	Protein kinase [Coralloccoccus sp. CA054B]	99.0 / 672	WP_120588883
<b>COCOR_05745</b>	288	NAD-dependent protein deacetylase [Coralloccoccus sp. CA054B]	97.6 / 288	WP_120588884
<b>COCOR_05746</b>	1041	Tetratricopeptide repeat-containing serine/threonine protein kinase [Coralloccoccus sp. CA054B]	99.5 / 1041	WP120588885
<b>COCOR_05747</b>	489	Hypothetical protein [Coralloccoccus sp. CA054B]	91.9 / 489	WP_233585371
<b>COCOR_05748</b>	499	Amidase [Coralloccoccus coralloides]	96.6 / 499	WP_128798457
<b>COCOR_05749</b>	245	Hypothetical protein [Coralloccoccus sp. CA054B]	96.3 / 245	WP_120589699
<b>COCOR_05750</b>	798	Hypothetical protein [Coralloccoccus sp. CA054B]	86.2 / 798	WP_120589700

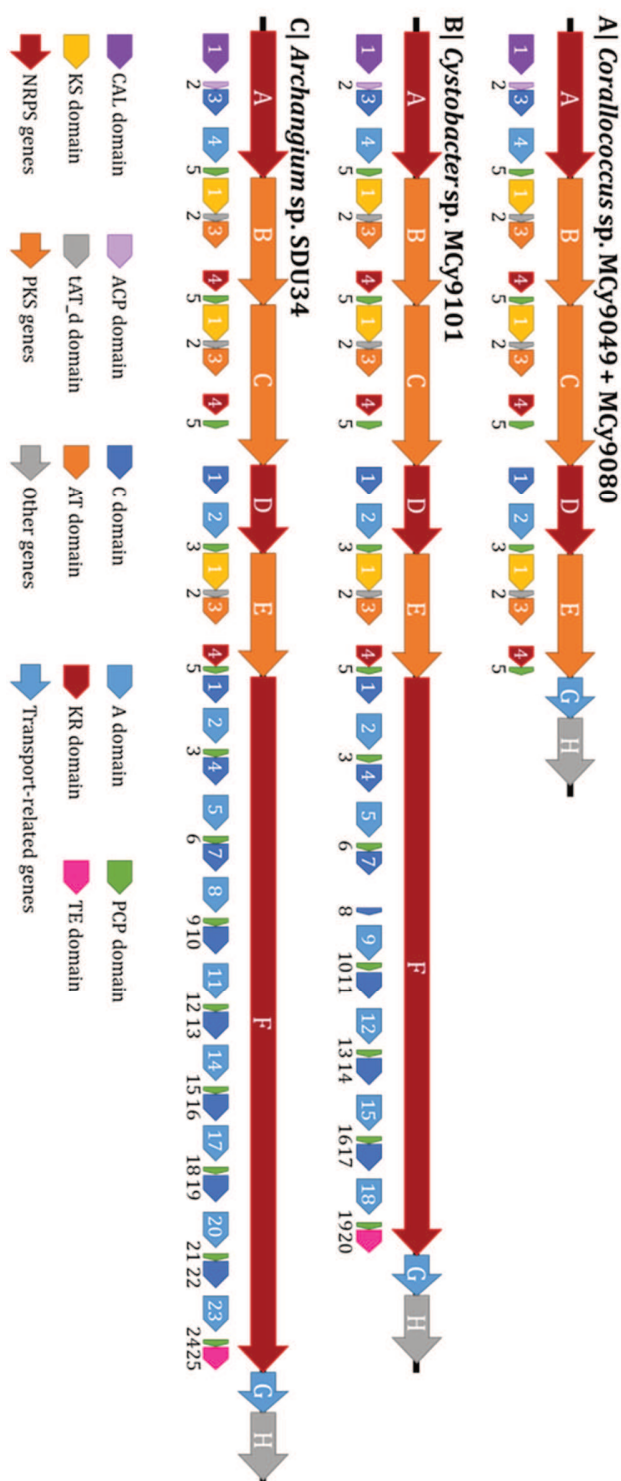
Additionally, every CDS in the two putative biosynthetic gene clusters of MCy9101 was extracted translated and searched again with the blastp algorithm against the NCBI RefSeq non-redundant protein sequence database.[2] The CDS that have been disrupted by single-crossover in previous experiments by Hu Zeng are highlighted in red.

**Table S11.** Tabulated blastP results for the CDS regions present in MCy9101 cluster 9.4.

CDS	Length [AA]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Accession Nr.
<b>653707-659140</b>	1811	Amino acid adenylation domain-containing protein [Archangium violaceum SDU34]	96.5 / 1810	WP_203404059
<b>659191-663175</b>	1328	Type I polyketide synthase [Archangium violaceum SDU34]	95.8 / 1321	WP_203404058
<b>663171-663849</b>	226	Type I polyketide synthase [Archangium violaceum SDU34]	95.6 / 226	WP_203404058
<b>663882-669786</b>	1968	SDR family NAD(P)-dependent oxidoreductase [Archangium violaceum SDU34]	96.3 / 1967	WP_203404057
<b>669789-673119</b>	1110	Amino acid adenylation domain-containing protein [Archangium violaceum SDU34]	95.7 / 1109	WP_203404056
<b>673112-677566</b>	1485	Type I polyketide synthase [Archangium violaceum SDU34]	97.0 / 1484	WP_203404055

**Table S12.** Tabulated blastP results for the CDS regions present in MCy9101 cluster 19.1.

CDS	Length [AA]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Accession Nr.
<b>0-9081</b>	3027	Non-ribosomal peptide synthetase [Archangium violaceum SDU34]	74.3 / 3027	WP_203404054
<b>9338-11021</b>	561	Non-ribosomal peptide synthetase [Archangium violaceum SDU34]	86.1 / 560	WP_203404054
<b>11845-13452</b>	536	Non-ribosomal peptide synthetase [Archangium violaceum SDU34]	90.8 / 534	WP_203404054
<b>13448-14867</b>	473	MFS transporter [Archangium violaceum SDU34]	93.6 / 472	WP_203404053
<b>14863-17317</b>	818	Penicillin acylase family protein [Archangium violaceum SDU34]	98.0 / 817	WP_203404052



**Figure S48.** Organization of biosynthetic gene clusters from MCy9049, MCy9080, MCy9101 and *A. violaceum* SDU34. Letters represent CDS regions, numbers represent domains as predicted by SeMPI 2.0 [4] The genome sequence of MCy9101 contains a gap between *cdsF\_7* and *cdsF\_8*. CAL: CoA ligase; ACP: acyl carrier protein; C: condensation; A: adenylation; PCP: peptidyl carrier protein; KS: ketosynthase; tAT\_d: trans-acyltransferase docking domain (annotated as DHT (dehydratase variant more commonly found in trans-AT PKS clusters) by antiSMASH 6.0.1 [5]); AT: acyltransferase; KR: ketoreductase; TE: thioesterase.

**Table S13.** Tabulated blastP results for the CDS regions present in MCy9101 after reassembly with SDU34 reference.

CDS	Length [AA]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Accession Nr.
<i>cdsA</i>	1811	Amino acid adenylation domain-containing protein [Archangium violaceum SDU34]	96.5 / 1810	WP_203404059
<i>cdsB</i>	1328	Type I polyketide synthase [Archangium violaceum SDU34]	95.8 / 1321	WP_203404058
<i>cdsC</i>	226	Type I polyketide synthase [Archangium violaceum SDU34]	95.6 / 226	WP_203404058
<i>cdsD</i>	1968	SDR family NAD(P)-dependent oxidoreductase [Archangium violaceum SDU34]	96.3 / 1967	WP_203404057
<i>cdsE</i>	1110	Amino acid adenylation domain-containing protein [Archangium violaceum SDU34]	95.7 / 1109	WP_203404056
<i>cdsF</i>	7200	Non-ribosomal peptide synthetase [Archangium violaceum SDU34]	76.8 / 4030	WP_203404054
<i>cdsG</i>	473	MFS transporter [Archangium violaceum SDU34]	93.6 / 472	WP_203404053
<i>cdsH</i>	818	Penicillin acylase family protein [Archangium violaceum SDU34]	98.0 / 817	WP_203404052

**Table S14.** Predicted specificity of A domains from *cdsA* and *cdsD* of the putative biosynthetic gene clusters.

A domain	Stachelhaus sequence	Nearest Stachelhaus code (match)	NRSPredictor2	Consensus
<i>MCy9049</i> <i>cdsA_4</i>	asGsLGGVCK	val (70%)	hydrophobic-aromatic	-
<i>MCy9080</i> <i>cdsA_4</i>	asGsLGGVCK	val (70%)	hydrophobic-aromatic	-
<i>MCy9101</i> <i>cdsA_4</i>	asGsLGGVCK	val (70%)	hydrophobic-aromatic	-
<i>SDU34</i> <i>cdsA_4</i>	asGsLGGiCK	val (60%)	hydrophobic-aromatic	-
<i>MCy9049</i> <i>cdsD_2</i>	DVWHVSLVDK	ser (100%)	ser	ser
<i>MCy9080</i> <i>cdsD_2</i>	DVWHVSLVDK	ser (100%)	ser	ser
<i>MCy9101</i> <i>cdsD_2</i>	DVWHVSLVDK	ser (100%)	ser	ser
<i>SDU34</i> <i>cdsD_2</i>	DVWHVSLVDK	ser (100%)	ser	ser

**Table S15.** Predicted specificity of A domains from *cdsF* of the putative biosynthetic gene clusters. Modules with identical Stachelhaus sequence and specificity prediction are highlighted in the same color.

A domain	Stachelhaus sequence	Nearest Stachelhaus code (match %)	NRPSPredictor2	Consensus
<b>MCy9101</b> <i>cdsF_2</i>	DALWLGaTFK	val (90%)	val	val
<b>MCy9101</b> <i>cdsF_5</i>	DALWLGaTFK	val (90%)	val	val
<b>MCy9101</b> <i>cdsF_9</i>	DiWHFgqIDK	ser (70%)	hydrophilic	-
<b>MCy9101</b> <i>cdsF_12</i>	DVWHVSLVDK	ser (100%)	ser	ser
<b>MCy9101</b> <i>cdsF_15</i>	DASTVAGVCK	bht (100%)	bht	bht
<b>MCy9101</b> <i>cdsF_18</i>	DLFNNALTYK	ala (100%)	ala	ala
<b>SDU34</b> <i>cdsF_2</i>	DAFFfGaTFK	ile (80%)	val	-
<b>SDU34</b> <i>cdsF_5</i>	DAfWLGgTFK	val (90%)	val	val
<b>SDU34</b> <i>cdsF_8</i>	DLFNNALTYK	ala (100%)	ala	ala
<b>SDU34</b> <i>cdsF_11</i>	DiWHFgqIDK	ser (70%)	hydrophilic	-
<b>SDU34</b> <i>cdsF_14</i>	DALWLGgTFK	val (100%)	val	val
<b>SDU34</b> <i>cdsF_17</i>	DLFNNALTYK	ala (100%)	ala	ala
<b>SDU34</b> <i>cdsF_20</i>	DLFNNALTYK	ala (100%)	ala	ala
<b>SDU34</b> <i>cdsF_23</i>	DASTVAGVCK	bht (100%)	bht	bht

bht:  $\beta$ -hydroxytyrosine; hydrophobic-aromatic: phe, trp, phg (2-phenylglycine), tyr, bht; hydrophobic-aliphatic: gly, ala, val, leu, ile, abu (2-aminobutyric acid), iva (isovaline); hydrophilic: asp, asn, glu, gln, aad (2-aminoadipic acid)



Table S16. Predicted specificity of AT domains from the putative biosynthetic gene clusters.

AT domain	ATSignature (Top 3 matches, score)	Minowa (Top 3 prediction, score)	Consensus
<b>MCy9049</b> <i>cdsB_3</i>	Malonyl-CoA: 91.7% inactive: 87.5% 2-Hydroxymalonyl-CoA: 54.2%	Malonyl-CoA: 188.3 inactive: 131.5 Methoxymalonyl-CoA: 71.0	Malonyl-CoA
<b>MCy9080</b> <i>cdsB_3</i>	Malonyl-CoA: 91.7% inactive: 87.5% 2-Hydroxymalonyl-CoA: 54.2%	Malonyl-CoA: 186.1 inactive: 132.3 Methoxymalonyl-CoA: 70.8	Malonyl-CoA
<b>MCy9101</b> <i>cdsB_3</i>	Malonyl-CoA: 91.7% inactive: 87.5% 2-Hydroxymalonyl-CoA: 54.2%	Malonyl-CoA: 191.0 inactive: 130.5 Methoxymalonyl-CoA: 72.6	Malonyl-CoA
<b>SDU34</b> <i>cdsB_3</i>	Malonyl-CoA: 91.7% inactive: 87.5% 2-Hydroxymalonyl-CoA: 54.2%	Malonyl-CoA: 192.4 inactive: 134.1 Methoxymalonyl-CoA: 72.8	Malonyl-CoA
<b>MCy9049</b> <i>cdsC_3</i>	Malonyl-CoA: 91.7% inactive: 83.3% 2-Hydroxymalonyl-CoA: 62.5%	Malonyl-CoA: 194.5 inactive: 117.3 Methoxymalonyl-CoA: 80.8	Malonyl-CoA
<b>MCy9080</b> <i>cdsC_3</i>	Malonyl-CoA: 91.7% inactive: 83.3% 2-Hydroxymalonyl-CoA: 62.5%	Malonyl-CoA: 194.5 inactive: 117.3 Methoxymalonyl-CoA: 80.8	Malonyl-CoA
<b>MCy9101</b> <i>cdsC_3</i>	Malonyl-CoA: 91.7% inactive: 79.2% 2-Hydroxymalonyl-CoA: 58.3%	Malonyl-CoA: 164.1 inactive: 114.0 Methoxymalonyl-CoA: 72.0	Malonyl-CoA
<b>SDU34</b> <i>cdsC_3</i>	Malonyl-CoA: 91.7% inactive: 79.2% 2-Hydroxymalonyl-CoA: 58.3%	Malonyl-CoA: 165.3 inactive: 114.5 Methoxymalonyl-CoA: 71.3	Malonyl-CoA
<b>MCy9049</b> <i>cdsE_3</i>	Malonyl-CoA: 91.7% inactive: 87.5% 2-Hydroxymalonyl-CoA: 54.2%	Malonyl-CoA: 184.5 inactive: 118.6 Methoxymalonyl-CoA: 82.0	Malonyl-CoA
<b>MCy9080</b> <i>cdsE_3</i>	Malonyl-CoA: 91.7% inactive: 87.5% 2-Hydroxymalonyl-CoA: 54.2%	Malonyl-CoA: 182.2 inactive: 115.3 Methoxymalonyl-CoA: 80.5	Malonyl-CoA
<b>MCy9101</b> <i>cdsE_3</i>	Malonyl-CoA: 91.7% inactive: 75.0% 2-Hydroxymalonyl-CoA: 58.3%	Malonyl-CoA: 174.1 inactive: 108.2 Methoxymalonyl-CoA: 75.6	Malonyl-CoA
<b>SDU34</b> <i>cdsE_3</i>	Malonyl-CoA: 91.7% inactive: 75.0% 2-Hydroxymalonyl-CoA: 58.3%	Malonyl-CoA: 175.7 inactive: 108.2 Methoxymalonyl-CoA: 75.3	Malonyl-CoA

**Table S17.** Predicted specificity of the Acyl-CoA ligase (CAL) domains from the putative biosynthetic gene clusters.

CAL domain	Minowa (Top 3 prediction, score)
<b>MCy9049</b> <i>cdsA_1</i>	NH2: 152.6 Acetyl-CoA: 17.8 fatty_acid: 3.9
<b>MCy9080</b> <i>cdsA_1</i>	NH2: 159.1 Acetyl-CoA: 18.3 fatty_acid: 3.2
<b>MCy9101</b> <i>cdsA_1</i>	NH2: 28.7 fatty_acid: 9.1 shikimic_acid: 5.4
<b>SDU34</b> <i>cdsA_1</i>	NH2: 197.6 fatty_acid: 10.1 AHBA: 3.1

AHBA: Aminohydroxybenzoate

**Figure S49.** Consensus sequence of conserved active site residues of all 26 KS domains in Table S18.**Figure S50.** Consensus sequence of conserved C-terminal regions of all 26 KS domains in Table S18.

**Table S18.** Alignment of the KS domain amino acid sequences encompassing the active site and a conserved C-terminal region. [6] The active site residue and a conserved serine in KS<sup>S</sup> and extender KS domains or glycine in KS<sup>Q</sup> domains are highlighted in red. Reference KS domains: Nys: nystatin (GenBank accession number: AF263912); Amph: amphotericin (AF357202); Pim: pimarin (AJ278573); Myc: mycinamicin (AB017641); Nid: niddamycin (AF016585); Ole: oleandomycin (AF220951); Pik: pikromycin (AF079138).

KS domain	Conserved Active Site	Conserved C-terminal Region
MCy9049 cdsB_1	GPAVTVQSS <b>S</b> STSLTC	GVTS <b>LGM</b> GGTNAHAILEE
MCy9080 cdsB_1	GPAVTVQSS <b>S</b> STSLTC	GVTS <b>LGM</b> GGTNAHAILEE
MCy9101 cdsB_1	GPAVTIQSS <b>S</b> STSLTA	GVTS <b>LGM</b> GGTNAHAILEE
SDU34 cdsB_1	GPAVTIQSS <b>S</b> STSLTA	GVTS <b>LGM</b> GGTNAHAILEE
MCy9049 cdsC_1	GPSYSITSAC <b>S</b> TSLVA	GVSSFGV <b>G</b> GTNAHVILEE
MCy9080 cdsC_1	GPSYSITSAC <b>S</b> TSLVA	GVSSFGV <b>G</b> GTNAHVILEE
MCy9101 cdsC_1	GPSYSILSAC <b>S</b> TSLVA	GVSSFGV <b>G</b> GTNAHVVLQE
SDU34 cdsC_1	GPSYSILSAC <b>S</b> TSLVA	GVSSFGV <b>G</b> GTNAHVVLQE
MCy9049 cdsE_1	GPAATVQTAC <b>S</b> SSLVA	GVSAFGI <b>G</b> GTNTHVILEE
MCy9080 cdsE_1	GPAATVQTAC <b>S</b> SSLVA	GVSAFGI <b>G</b> GTNTHVILEE
MCy9101 cdsE_1	GPSLTVQTAC <b>S</b> SSLVA	GIS <b>S</b> FGI <b>G</b> GTNAHLVILEE
SDU34 cdsE_1	GPSLTVQTAC <b>S</b> SSLVA	GIS <b>S</b> FGI <b>G</b> GTNAHLVILEE
NysA (KS <sup>S</sup> )	GPAVTVDT <b>T</b> SSSSLVA	GVSSFGI <b>S</b> GTNAHVILEQ
AmphA (KS <sup>S</sup> )	GPAVTLDT <b>A</b> SSSSLVA	AVSSFGI <b>S</b> GTNAHVILEE
PimS0 (KS <sup>S</sup> )	GPALTVDT <b>A</b> SSSSLVA	GVSSFGL <b>T</b> GTNAHVILEE
NysKS1 (Extender KS)	GPAVTVDT <b>A</b> CSSSLVA	AVSSFGI <b>S</b> GTNAHALLEQ
AmphKS1 (Extender KS)	GPAVTVDT <b>A</b> CSSSLVA	AVSSFGI <b>S</b> GTNAHAVLEQ
PimKS1 (Extender KS)	GPALTVDT <b>A</b> CSSSLVA	AVSSFGI <b>S</b> GTNAHVILEQ
MycKS1 (Extender KS)	GPAVTVDT <b>A</b> CSSSLVA	AVSSFGV <b>S</b> GTNAHVILEQ
NidKS1 (Extender KS)	GPALTVDT <b>A</b> CSSSLVA	GISA <b>F</b> GV <b>S</b> GTNAHVILEE
OleKS1 (Extender KS)	GPAVTVDT <b>A</b> CSSSLVA	GVSSFGI <b>S</b> GTNAHVIVLEE
PikKS1 (Extender KS)	GPAVTVDT <b>A</b> CSSSLVA	AVSSFGV <b>S</b> GTNAHVILEQ
MycKSQ (KS <sup>Q</sup> )	GPSFVVDSA <b>Q</b> SSSLVA	AVSSFG <b>M</b> GGTNAHLVLAQ
NidKSQ (KS <sup>Q</sup> )	GPSLVVDT <b>G</b> QSSSLVA	GVSAFG <b>M</b> GGTNAHVILEE
OleKSQ (KS <sup>Q</sup> )	GPSLTVDT <b>G</b> QSSSLAA	GVSSFG <b>M</b> GGTNAHVVLSE
PikKSQ (KS <sup>Q</sup> )	GPSMVVDS <b>G</b> QSSSLVA	GVSSFG <b>M</b> GGTNAHVILEE

## S 3.4 References

1. Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry* **1984**, *11*, doi:10.1002/bms.1200111109.
2. O’Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciuffo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733-45, doi:10.1093/nar/gkv1189.
3. Cimermancic, P.; Medema, M.H.; Claesen, J.; Kurita, K.; Wieland Brown, L.C.; Mavrommatis, K.; Pati, A.; Godfrey, P.A.; Koehrsen, M.; Clardy, J.; et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014**, *158*, 412–421, doi:10.1016/j.cell.2014.06.034.
4. Zierep, P.F.; Ceci, A.T.; Dobrusin, I.; Rockwell-Kollmann, S.C.; Günther, S. SeMPI 2.0—A Web Server for PKS and NRPS Predictions Combined with Metabolite Screening in Natural Product Databases. *Metabolites* **2021**, *11*, doi:10.3390/metabo11010013.
5. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29-W35, doi:10.1093/nar/gkab335.
6. Brautaset, T.; Borgos, S.E.F.; Sletta, H.; Ellingsen, T.E.; Zotchev, S.B. Site-specific mutagenesis and domain substitutions in the loading module of the nystatin polyketide synthase, and their effects on nystatin biosynthesis in *Streptomyces noursei*. *J. Biol. Chem.* **2003**, *278*, 14913–14919, doi:10.1074/jbc.M212611200.

## Chapter 4

# Expanding the Myxochelin Natural Product Family by Nicotinic Acid Containing Congeners

Previously published in:

Nicolas A. Frank<sup>†</sup>, Márió Széles<sup>†</sup>, Sergi H. Akone, Sari Rasheed, Stephan Hüttel, Simon Frewert, Mostafa M. Hamed, Jennifer Herrmann, Sören M. M. Schuler, Anna K. H. Hirsch, Rolf Müller\*

*Molecules*, 2021 August 14; **26**(16):4929

DOI: 10.3390/molecules26164929

### Affiliation

Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

# Contributions and Acknowledgments

## Author's effort:

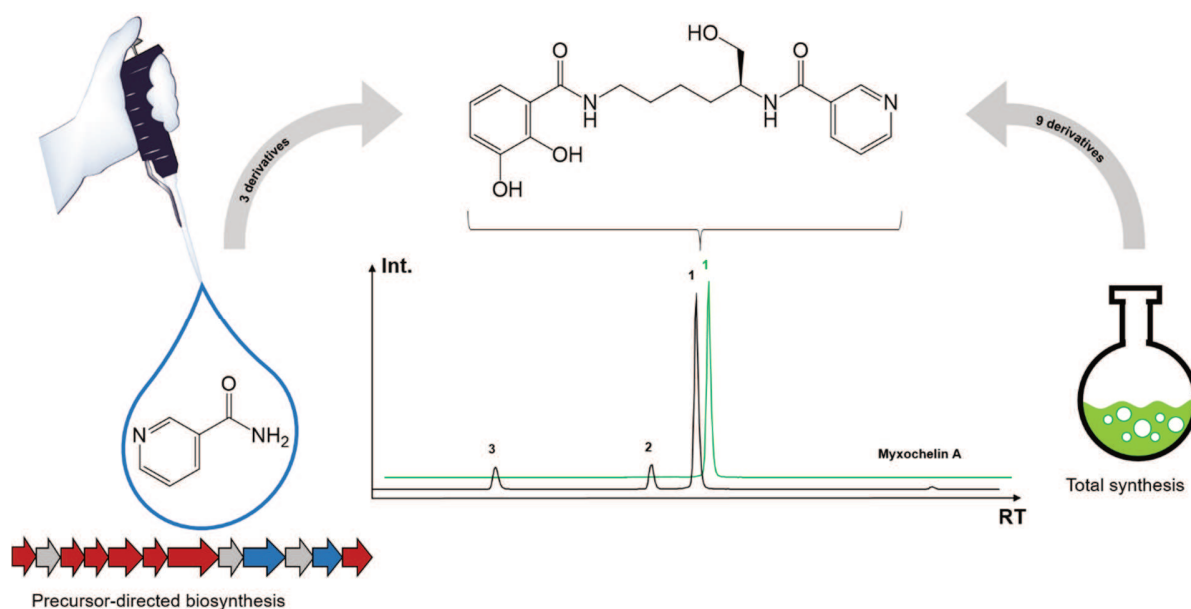
The author significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The laboratory and *in silico* work regarding metabolome-guided compound discovery, precursor-directed biosynthesis and proposal of the biosynthesis were performed by the author. Furthermore, the author contributed significantly to conceiving and writing this manuscript.

## Contributions by others:

Márió Széles contributed to conception of this study, performed total synthesis and contributed through writing of the manuscript. Sergi Akone contributed to conception, performed cultivation, precursor-directed biosynthesis and NMR structure elucidation and contributed through writing of the manuscript. Sari Rasheed assessed potential synergistic activities. Stephan Hüttel and Simon Frewert contributed preliminary data of PDB experiments and elucidated structures. Mostafa Hamed contributed through assistance in total synthesis and structure confirmation of synthetic derivatives. Jennifer Herrmann contributed through conception and organization of biological activity assays and writing of the manuscript. Sören Schuler contributed through assistance in total synthesis and writing of the manuscript. Anna Hirsch and Rolf Müller contributed by supervision of the project and conceiving, editing and proofreading of the manuscript.

## 4.1 Abstract

Myxobacteria represent a viable source of chemically diverse and biologically active secondary metabolites. The myxochelins are a well-studied family of catechol-type siderophores produced by various myxobacterial strains. Here, we report the discovery, isolation, and structure elucidation of three new myxochelins N1–N3 from the terrestrial myxobacterium *Coralloccoccus* sp. MCy9049, featuring an unusual nicotinic acid moiety. Precursor-directed biosynthesis (PDB) experiments and total synthesis were performed in order to confirm structures, improve access to pure compounds for bioactivity testing, and to devise a biosynthesis proposal. The combined evaluation of metabolome and genome data covering myxobacteria supports the notion that the new myxochelin congeners reported here are in fact frequent side products of the known myxochelin A biosynthetic pathway in myxobacteria.



**Graphical Abstract.** Metabolome-guided screening of *Coralloccoccus* sp. MCy9049 led to the discovery of myxochelin N1 (1). Precursor-directed biosynthesis (PDB) by feeding of nicotinamide led to increased yield and discovery of myxochelins N2 (2) and N3 (3). Total synthesis was performed to confirm the structures and to generate additional non-natural derivatives.

## 4.2 Introduction

Myxobacteria are Gram-negative bacteria with a variety of remarkable characteristics. They are able to glide on surfaces and show coordinated swarming [1]. Unlike many other bacteria, they are able to form macroscopic fruiting bodies and exhibit unique multicellular behavior mediated through chemical communication [2]. Myxobacteria are an excellent source for secondary metabolites with a broad variety of chemical scaffolds as well as biological activities [3], such as the antibacterial cystobactamids [4], the antifungal soraphens [5], the antiplasmodial chlorotonils [6], and the cytotoxic epothilones [7]. In addition, myxobacteria are also known to produce secondary metabolites with other functions such as pigments, signaling molecules, and iron-chelating siderophores [8].

The myxochelins represent a class of catecholate-type siderophores produced by various myxobacterial strains that have been reported in literature for more than 30 years [9]. The corresponding biosynthetic gene cluster (BGC) encodes the enzymes MxcCD assembling 2,3-dihydroxybenzoic acid (DHBA), MxcE activating DHBA by adenylation and loading it to the ArCP domain of MxcF and the non-ribosomal peptide synthetase (NRPS) subunit MxcG condensing DHBA with the two primary amino groups of L-lysine. The NADPH-dependent reductase domain of MxcG releases an aldehyde intermediate from the assembly line by reduction of the thioester, which is further reduced by MxcG to form myxochelin A or transaminated by MxcL to form myxochelin B [10,11]. Although it has been shown that the promiscuity of MxcE allows the activation and loading of other benzoic acid derivatives to the assembly line to form precursor-derived non-natural myxochelin derivatives [12–14], the incorporation of heteroaromatic carboxylic acid precursors has so far not been described.

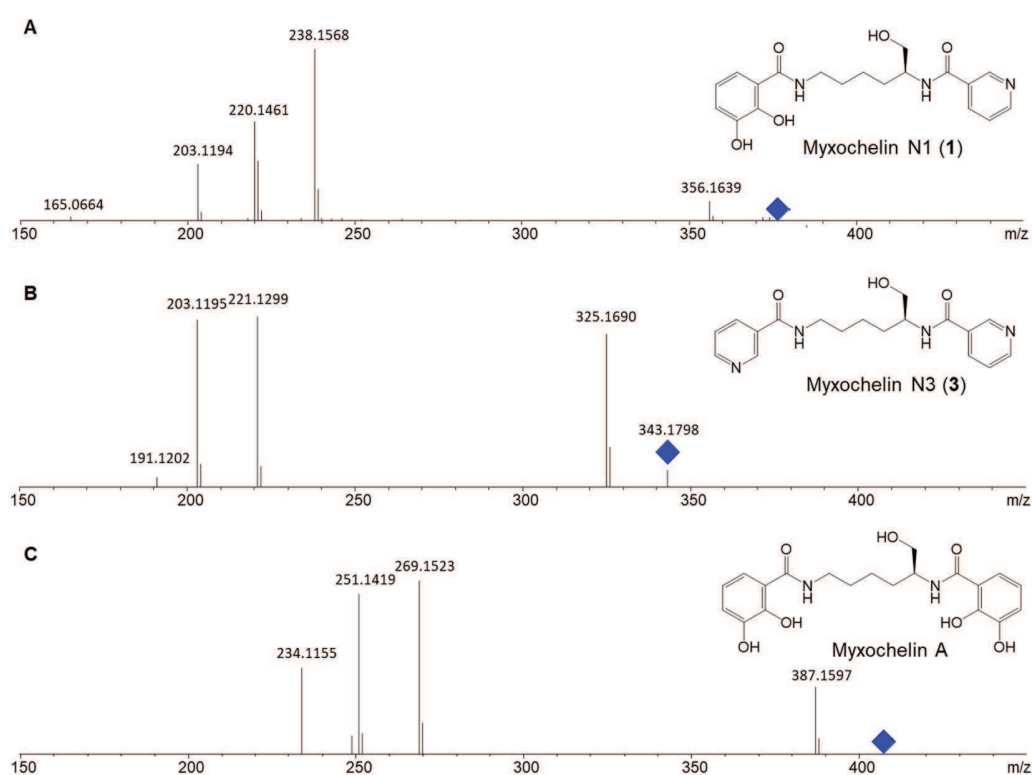
We hereby report the discovery, isolation, and full structure elucidation of myxochelins N1–N3 from crude organic extracts of the myxobacterial producer *Coralloccoccus* sp. MCy9049, featuring an unusual nicotinic acid moiety. In order to probe the biosynthetic origin of the newly discovered myxochelins and to achieve higher yields in the natural producer strain, precursor-directed biosynthesis (PDB) was carried out utilizing nicotinic acid as well as nicotinamide as substrates. In addition, we utilized a total synthesis entry to provide higher quantities of the compounds, to prove the chemical structure of the isolated natural products, and to facilitate the generation of further derivatives.



## 4.3 Results

### 4.3.1 Discovery of Myxochelin N1

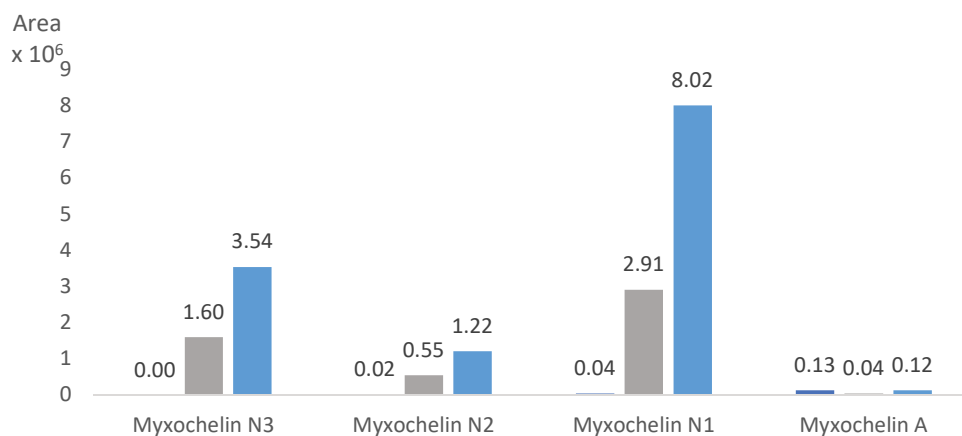
Analysis of the secondary metabolome of Cystobacterineae species *Coralloccoccus* sp. MCy9049 [15] revealed the secondary metabolite **1** featuring an ion peak in the liquid chromatography-mass spectrometry (LC-MS) chromatogram at mass-to-charge ratio ( $m/z$ ) 374.1710  $[M + H]^+$ , supporting the deduced molecular formula  $C_{19}H_{24}N_3O_5^+$  at a retention time of 3.5 min. The peculiar UV absorption at 219, 249, and 309 nm (Figure S1 in Supplementary Materials) sparked further interest in isolation of the compound. The  $MS^2$  fragment pattern of **1** was characterized by  $m/z$  356.1638 ( $C_{19}H_{22}N_3O_4^+$ ),  $m/z$  238.1563 ( $C_{12}H_{20}N_3O_2^+$ ),  $m/z$  220.1457 ( $C_{12}H_{18}N_3O^+$ ), and  $m/z$  203.1189 ( $C_{12}H_{15}N_2O^+$ ), which according to neutral losses is similar to the known natural product myxochelin A (Figure 1). The quantity of **1** in the crude organic extracts of *Coralloccoccus* sp. MCy9049, however, was initially insufficient for full structure elucidation and evaluation of biological activities. Therefore, we employed a two-pronged approach to access this alleged new secondary metabolite, using both precursor-directed biosynthesis and total synthesis.



**Figure 1.**  $MS^2$  fragmentation spectra of myxochelin N1 (A), myxochelin N3 (B), and myxochelin A (C). The blue diamond represents the  $m/z$  of the precursor ion.

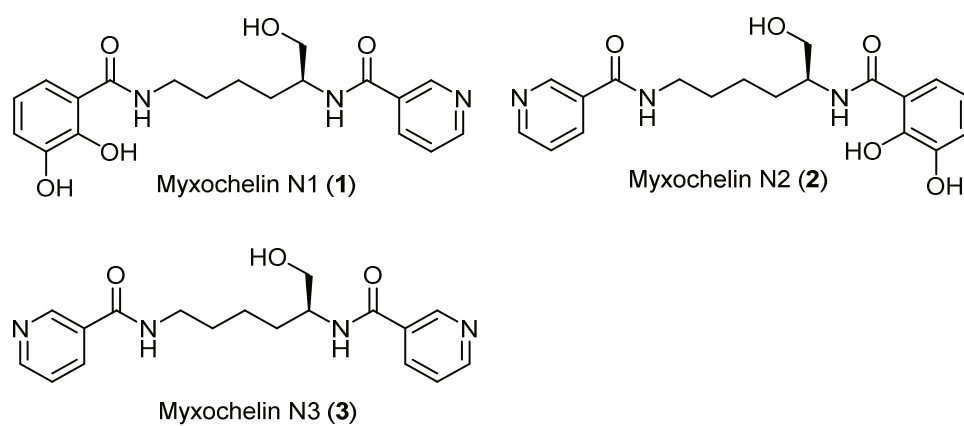
### 4.3.2 Precursor-Directed Biosynthesis and Isolation of Myxochelin N1 and N3

In order to access sufficient quantities of **1** for structure elucidation by NMR, a PDB approach was conducted aiming at increasing the yield available from the natural producer. Comparison of the deduced molecular formulae of **1** ( $C_{19}H_{23}N_3O_5$ ) and myxochelin A ( $C_{20}H_{24}N_2O_7$ ) as well as their  $MS^2$  fragments suggested the formal presence of a pyridinecarboxylic acid instead of one of the two DHBA moieties present in myxochelin A. As different hypothetical placements of the carboxyl group give rise to three constitutional isomers that cannot be distinguished from each other with the data available, PDB was conducted by feeding of 2-, 3-, and 4-pyridinecarboxylic acid, respectively, as well as nicotinamide to the culture. While 2- and 4-pyridinecarboxylic acid were only incorporated to a minor degree and showed ion peaks at different retention times (Figure S34), supplementing the producing culture with 3-pyridinecarboxylic acid led to a 70-fold increase in production of compound **1**. Surprisingly, feeding of nicotinamide even resulted in a 193-fold increase of production in comparison to the non-supplemented culture (Figure 2). In addition, the PDB approach revealed the two additional metabolites **2** and **3** with  $m/z$  374.1712 ( $C_{19}H_{24}N_3O_5^+$ , retention time 3.2 min) and  $m/z$  343.1766 ( $C_{18}H_{23}N_4O_3^+$ , retention time 2.0 min), respectively, possessing similar  $MS^2$  fragmentations compared to **1** and myxochelin A (Figures S14–S17 and S32–S34).



**Figure 2.** LC-MS chromatogram peak areas of myxochelin N1–N3 and myxochelin A in control (dark blue) and with nicotinic acid (gray) and nicotinamide (light blue) supplementation. Numbers above bars indicate EIC peak area  $\times 10^6$ .

In order to isolate **1–3** (Figure 3), *Coralloccoccus* sp. MCy9049 was cultivated in larger-scale batch fermentation in shake flasks with feeding of nicotinamide. Purification of the secondary metabolites was performed by liquid/liquid partitioning of the crude extract followed by semi-preparative HPLC with UV detection. However, only compounds **1** and **3** were obtained in sufficient amounts allowing their full structure elucidation by NMR.



**Figure 3.** Chemical structure of the natural products **1–3**. Structure elucidation of **1** and **3** was performed with the isolated compounds (Section 4.3.3), structure elucidation of **2** was performed with its synthetic homolog **9b** (Sections 4.3.4 and 4.6.6).

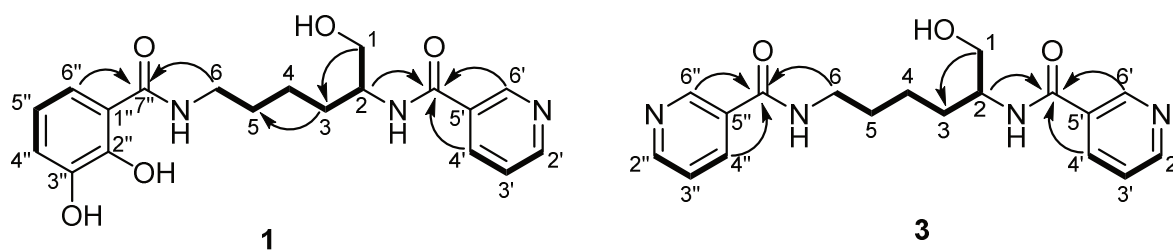
### 4.3.3 Structure Elucidation

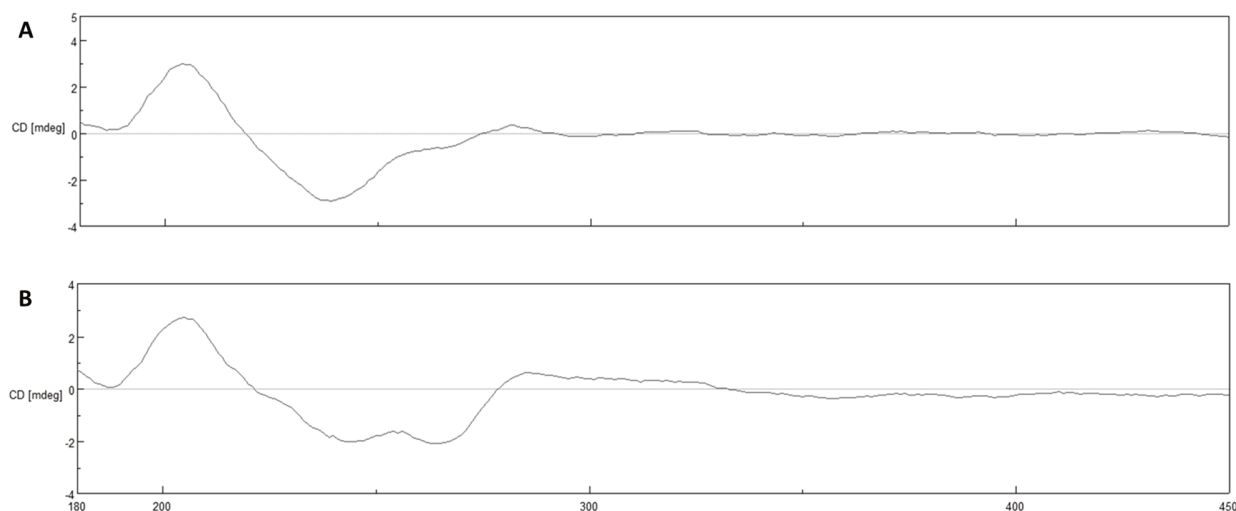
Compound **1** was obtained as a brownish oil. Its molecular formula was assigned as  $C_{19}H_{23}N_3O_5$  based on a prominent ion peak at  $m/z$  374.1710 ( $[M + H]^+$ , calcd for  $C_{19}H_{24}N_3O_5^+$ , 374.1711) in HRESIMS, which requires ten degrees of unsaturation. Inspection of the  $^1H$  NMR spectrum (Table 1) displayed two separate aromatic systems. The first consisted of the signals at  $\delta_H$  7.16 (d,  $J = 8.1$  Hz, H-6''), 6.90 (d,  $J = 8.0$  Hz, H-4''), and 6.67 (t,  $J = 8.0$  Hz, H-5''), which is typical of a 1,2,3-trisubstituted benzene system (Table 1). This observation was confirmed by the HMBC correlations from H-5'' to C-1'' ( $\delta_C$  115.3) and C-3' ( $\delta_C$  145.9), as well as from H-4'' and H-6'' to C-2'' ( $\delta_C$  148.6) (Figure 4). Based on the HMBC correlations from H-6'' to C-7'' ( $\delta_C$  170.1), it was proven that this aromatic group was attached to an amide group at C-1'', confirming the presence of a benzamide moiety in **1** (Figure 4). The second spin system consisted of the signals at  $\delta_H$  8.96 (brd,  $J = 2.2$  Hz, H-6'), 8.65 (dd,  $J = 4.8, 1.5$  Hz, H-2'), 8.20 (brd,  $J = 7.7$  Hz, H-4'), and 7.49 (dd,  $J = 8.1, 5.0$  Hz, H-3') ascribed to a pyridine moiety as underpinned by the HMBC correlations from H-6' to C-4' ( $\delta_C$  135.7), C-2' ( $\delta_C$  151.0), as well as from H-3' to C-5' ( $\delta_C$  130.9) (Figure 4). The HMBC correlation from H-6' and H-4' to C-7' (166.7) confirmed the presence of a nicotinic acid moiety in **1**. The remaining signals of three aliphatic methylene groups at  $\delta_H$  1.49 (m, H<sub>2</sub>-4), 1.63 (m, H<sub>2</sub>-3), 1.72 (m, H<sub>2</sub>-5), an aminomethylene group at  $\delta_H$  3.39 (t,  $J = 7$  Hz, H<sub>2</sub>-6), an oxygenated methylene group at  $\delta_H$  3.63 (brd,  $J = 3.4$  Hz, H<sub>2</sub>-1a) and  $\delta_H$  3.61 (brd,  $J = 3.8$  Hz, H<sub>2</sub>-1b), and an aminomethine proton at  $\delta_H$  4.15 (m, H<sub>2</sub>-2) were attributed to a 2,6-diaminohexan-1-ol moiety. This assumption was supported by the long spin system, observed in the  $^1H$ - $^1H$  COSY spectrum starting from the oxymethylene proton H-1a and H-1b via the aminomethine proton H-2, terminating at the aminomethylene proton H<sub>2</sub>-6 (Figure 4); as well as by the HMBC correlations from H-1a and H-1b to C-3 ( $\delta_C$  30.2), H<sub>2</sub>-3 to C-1 ( $\delta_C$  64.4), and C-5 ( $\delta_C$  28.9). The linkage between the hexanol moiety and the nicotinic acid residue was elucidated based on the HMBC correlations from H-2 to C-7' ( $\delta_C$  166.7). Moreover, the 1-hexanol moiety was also found to be attached to the benzamide moiety as an HMBC correlation was observed from H<sub>2</sub>-6 to C-7'' ( $\delta_C$  170.1). These data were similar to those reported for myxochelin A [16], with the only difference being the replacement of one of the two DHBA moieties by a nicotinic acid moiety, which is consistent with the MS<sup>2</sup> spectrum of **1** showing the loss of the dihydroxybenzoyl group (Figure 1).

Despite the fact that from a biogenetic point of view, compound **1** should possess the same stereochemistry as the co-occurring myxochelin A, we confirmed the absolute configuration of **1** through its total synthesis (Section 4.3.4). The total synthesis doubtlessly revealed the absolute configuration of **1** to be *S* as reported for myxochelin A [16]. This finding was further confirmed by comparison of the CD spectra (Figure 5) and optical rotation of **1** and its synthetic analogue **9a**. Thus, compound **1** was confirmed as a new myxochelin derivative featuring an unusual nicotinic acid moiety.

**Table 1.** NMR data of **1** measured in MeOH-*d*<sub>4</sub> at 500 (<sup>1</sup>H) and 125 (<sup>13</sup>C) MHz.

<b>1 (in CD<sub>3</sub>OD)</b>			
No.	$\delta_{\text{H}}$ , Mult. ( <i>J</i> in Hz)	$\delta_{\text{C}}$ , Type	HMBC
1	3.63, brd (3.4) 3.61, brd (3.8)	64.4, CH <sub>2</sub>	2, 3
2	4.15, m	53.2, CH	4, 7'
3	1.63, m	30.2, CH <sub>2</sub>	1, 2, 4, 5
4	1.49, m	22.6, CH <sub>2</sub>	2, 5, 6
5	1.72, m	28.9, CH <sub>2</sub>	3, 6
6	3.39, t (7.0)	39.4, CH <sub>2</sub>	4, 5, 7''
2'	8.65, brd (8.2)	151.0, CH	3', 4', 6'
3'	7.49, dd (8.1, 5.0)	123.6, CH	2', 5'
4'	8.20, brd (7.7)	135.7, CH	2', 3', 6'
5'	-	130.9, C	-
6'	8.96, brd (2.2)	147.7, CH	2', 4', 7'
7'	-	166.7, C	-
1''	-	115.3, C	-
2''	-	148.6, C	-
3''	-	145.9, C	-
4''	6.90, d (8.0)	118.1, CH	2'', 3'', 6''
5''	6.67, t (8.0)	118.1, CH	1'', 3''
6''	7.16, d (8.1)	117.1, CH	1'', 2'', 4'', 7''
7''	-	170.1, C	-

**Figure 4.** Atom numbering and selected COSY (bold) and HMBC (plain) correlations of **1** and **3**.



**Figure 5.** CD spectra of **1** (A) and its synthetic analogue **9a** (B).

Compound **3** was isolated as colorless oil. Its molecular formula was determined as  $C_{18}H_{22}N_4O_3$  as a prominent protonated ion peak was observed at  $m/z$  343.1766  $[M + H]^+$  in HRESIMS spectrum bearing ten degrees of unsaturation. The  $^1H$  NMR spectrum of **3** displayed resonances of two aromatic spin systems consisting of overlapping signals due to protons with a similar chemical environment (Table 2). These signals included resonances at  $\delta_H$  8.95 (H-6'/6''), 8.66 (H-2'/2''), 8.20 (H-4'/4''), and 7.51 (H-3'/3'') suggesting the presence of two pyridine moieties in **3**. This observation was confirmed by the HMBC correlations from H-6' to C-4' ( $\delta_C$  136.4) and C-2' ( $\delta_C$  151.7), as well as H-3' to C-5' ( $\delta_C$  131.2) (Figure 4). Furthermore, the  $^1H$  NMR also exhibited the signals of three aliphatic methylene groups at  $\delta_H$  1.49 (m, H<sub>2</sub>-4), 1.64 (m, H<sub>2</sub>-3), 1.74 (m, H<sub>2</sub>-5), an aminomethylene at  $\delta_H$  3.42 (m, H<sub>2</sub>-6), an oxygenated methylene group at  $\delta_H$  3.62 (dd,  $J = 5.4, 10.7$  Hz, H<sub>2</sub>-1a) and  $\delta_H$  3.64 (dd,  $J = 5.2, 10.6$  Hz, H<sub>2</sub>-1b), and an aminomethine proton at  $\delta_H$  4.15 (m, H<sub>2</sub>-2) assigned to a 2,6-diaminohexan-1-ol moiety. Comparison of the NMR and UV data of **1** and **3** revealed a high degree of similarity except for the replacement of the DHBA moiety by an additional nicotinic acid moiety in **3**. This assumption was further underpinned by the absence of the signals of the DHBA moiety in the proton spectrum of **3** as well as by the HMBC correlations from H-4'/4'' and H-6'/6'' to C-7'/7'' ( $\delta_C$  167.2) connecting the hexanol moiety on its both sides with the nicotinic acid moieties (Figure 4). Moreover, this assumption is consistent with the MS<sup>2</sup> spectrum of **3** showing the loss of the nicotinic acid group (Figure 1).

From a biogenetic point of view, compound **3** should also retain the same stereochemistry as that of the co-occurring **1**, which is in congruence with the negative  $[\alpha]_D$  values measured for both compounds. Thus, compound **3** is a double nicotinic acid derivative of **1**.

**Table 2.** NMR data of **3** measured in MeOH-*d*<sub>4</sub> at 500 (<sup>1</sup>H) and 125 (<sup>13</sup>C) MHz.

<b>3 (in CD<sub>3</sub>OD)</b>			
No.	δ <sub>H</sub> , Mult. (J in Hz)	δ <sub>C</sub> , Type	HMBC
1	3.62, dd (10.7, 5.4) 3.64, dd (10.6, 5.2)	64.8, CH <sub>2</sub>	2, 3
2	4.15, m	53.7, CH	4, 7'
3	1.64, m	30.2, CH <sub>2</sub>	1, 2, 4, 5
4	1.49, m	23.4, CH <sub>2</sub>	2, 5, 6
5	1.74, m	29.2, CH <sub>2</sub>	3, 6
6	3.42, m	40.4, CH <sub>2</sub>	4, 5, 7''
2'/2''	8.66, o	151.7, CH *	4', 6'
3'/3''	7.51, dd (7.9, 2.3)	123.6, CH	5'
4'/4''	8.20, o	136.4, CH	2', 6', 7'
5'/5''	-	131.2, C *	-
6'/6''	8.95, o	148.2, CH	4', 7'
7'/7''	-	167.2, C	-

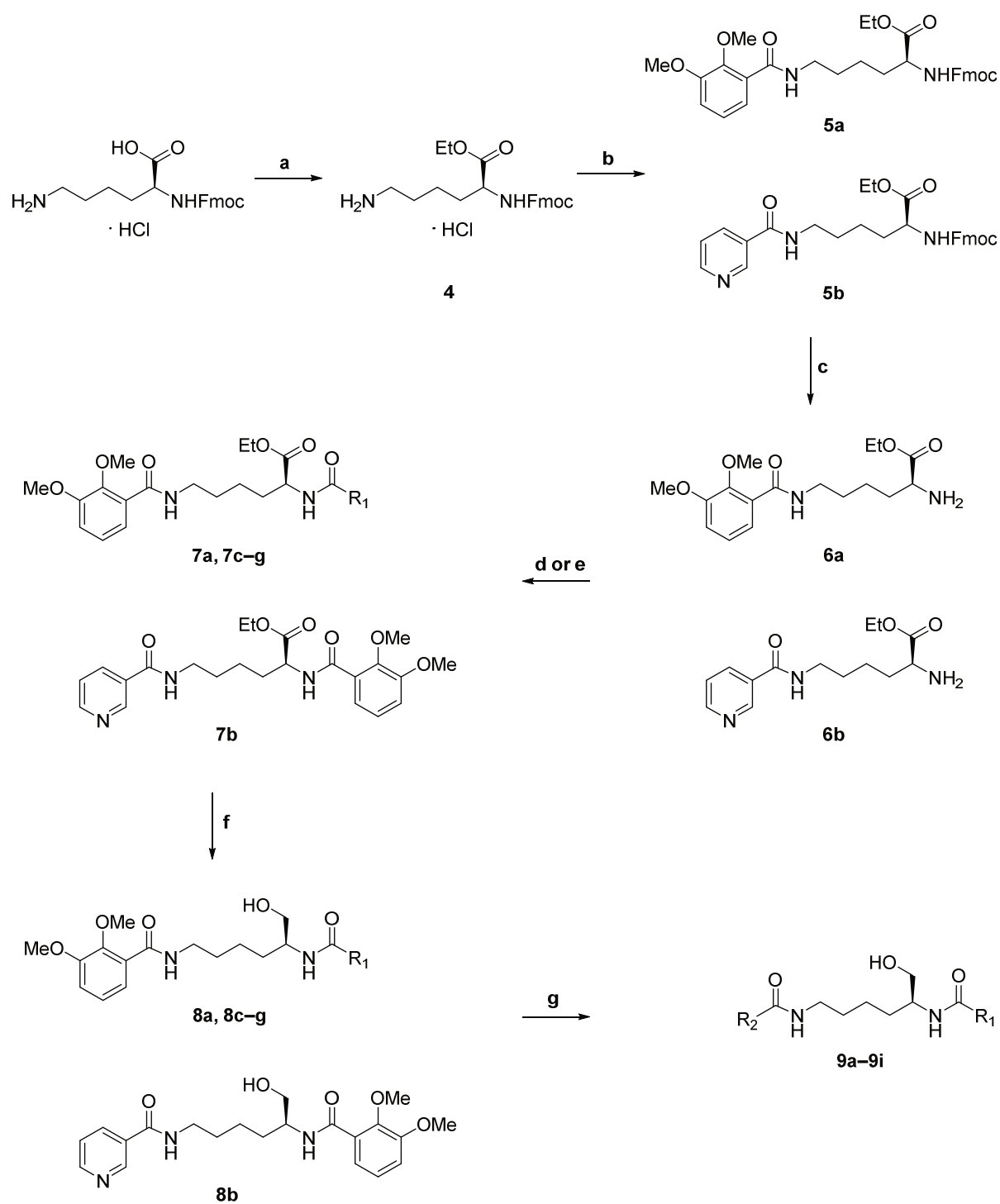
#### 4.3.4 Total Synthesis

In parallel to PDB, we sought to establish a total synthesis route for the new myxochelin congeners. This decision was motivated by the finding that none of the published PDB approaches with myxochelin had previously shown incorporation of heteroaromatic carboxylic acids [12] and thus success with the strategy described in Section 2.2 was all but guaranteed beforehand. Indeed, despite our success with **1** and **3**, we were not able to obtain **2** through PDB with yields sufficient for structure elucidation by NMR.

In 2009, Miyanaga et al. reported the total synthesis of several myxochelin analogues starting from either methyl L-lysinate or methyl α-*N*-Boc-L-lysinate in case of target compounds with different aryl moieties on the α- and ε-amino groups. They then used benzotriazol-1-yloxytris(dimethylamino)phosphonium hexafluorophosphate (BOP) to perform an amide coupling first on the ε-position, followed by Boc deprotection and subsequent coupling to the α-position. Finally, to obtain the target compounds, they removed the benzyl protecting groups of the 2,3-bis(benzyloxy)benzamide moiety under reductive conditions [17]. Recently, Schieferdecker et al. reported a fast and efficient synthetic route to synthesize myxochelin A and several derivatives. They used L-lysine ethyl ester hydrochloride as a starting point and subjected it to amide coupling conditions to employ mainly identical aryl groups on the α- and ε-amino groups. Finally, myxochelin A and derivatives were obtained after reduction of the ethyl ester to the corresponding primary hydroxyl group with lithium borohydride and demethylation of the 2,3-dimethoxybenzamide moiety with boron tribromide [18]. Later, the same group also used this strategy to synthesize a large set of myxochelin analogues to explore the structure–activity relationship with 5-lipoxygenase, an enzyme associated with various types of cancer [19].


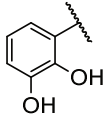
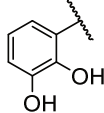


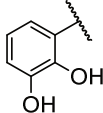
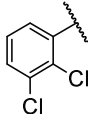
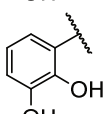
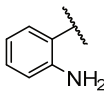
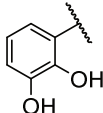
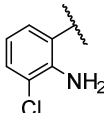
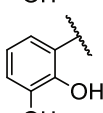
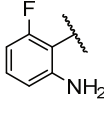
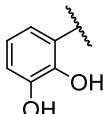

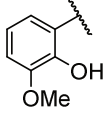
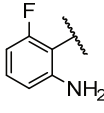
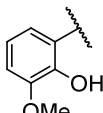
Based on the methodologies reported in the literature, in the present study we synthesized the new myxochelins **1** and **2**. Moreover, we used the established synthesis platform to generate six novel and one literature reported analogues, which are not readily accessible by PDB (Scheme 1). After converting the free carboxylic acid of Fmoc- L-lysine hydrochloride to the corresponding ethyl ester **4** [20], the side-chain  $\epsilon$ -amino group was subjected to amide coupling using either 2,3-dimethoxybenzoic acid or nicotinic acid. We achieved the fastest conversion with hexafluorophosphate azabenzotriazole tetramethyl uronium (HATU) as activating agent, triethylamine (TEA) as base, and *N,N*-dimethylformamide (DMF) as solvent. Afterwards, the  $\alpha$ -amino group of **5a** and **5b** was deprotected by removing the fluorenylmethoxycarbonyl (Fmoc) group under basic conditions [21]. By applying the same amide coupling conditions, the second aryl building block was attached providing intermediates **7a–g**. In case of aryl substituents with an Fmoc-protected amino group, the Fmoc group was removed one-pot under basic conditions after completion of the amide coupling. The subsequent reduction of the ethyl ester moiety with lithium borohydride yielded the primary hydroxyl group of **8a–g** in a pure reaction profile after optimization of the conditions. Lastly, boron tribromide was used to achieve partial or complete demethylation of the 2,3-dimethoxybenzamide system (Table 3). HRMS data of the synthetic products **9a/9b** and the isolated natural products **1/2**, and NMR data of **9a** and **1** were identical. As aforementioned, to prove the absolute configuration of myxochelin **1**, optical rotation was measured and circular dichroism spectrophotometry was utilized to compare the spectrum with that of synthetic compound **9a** (Figure 5). Finally, the seven target compounds **9a–g** of the total synthetic approach were obtained over six steps in sufficient amounts enabling further investigation.





**Scheme 1.** a: TMSCl, EtOH, r.t., 16 h, quant.; b: HATU, TEA, DMF, 2,3-dimethoxybenzoic acid (for **5a**) or nicotinic acid (for **5b**), r.t., 16 h, 51–83%; c: DMA, DCM (1:1), r.t., 2 h; d: HATU, TEA, DMF,  $R^1$ COOH, r.t., 16 h, 17–53% over two steps for **7a–7d**; e: HATU, TEA, DMF,  $R^2$ COOH, r.t., 16 h, then DMA, r.t., 2 h, 13–14% over two steps for **7e–7g**; f:  $\text{LiBH}_4$ , THF, EtOH, 4 °C, 15 min, then r.t., 2 h, 49–71%; g:  $\text{BBr}_3$ , DCM, r.t., 16 h, 18–27%.

Table 3. Synthetic analogues provided for biological activity testing.

	R <sup>1</sup>	R <sup>2</sup>
9a = 1		
9b = 2		
9c [12]		
9d		
9e		
9f		
9g		
9h		
9i		

#### 4.3.5 Biological Activities

Myxochelin A has been described to possess weak antimicrobial activity [9] which might be contributed to its iron-chelating properties. Although the new myxochelin derivatives **9a–9i** lack the symmetrical 2,3-dihydroxybenzoyl moieties, we found some marginal antibacterial activity in fractions of the extract of the natural producer strain containing myxochelin N1 (**1**) (data not shown). Thus, all synthetic derivatives were tested against a small panel of rather sensitive indicator strains and we also assessed potential synergistic activity of **9a** (= **1**) with selected reference antibiotics in *Staphylococcus aureus* str. Newman (Table S3). Unfortunately, **9a–9i** did not show any activity (minimum inhibitory

concentration, MIC > 128 µg/mL) against *Escherichia coli*, *S. aureus*, *Candida albicans*, and *Mucor hiemalis*. Furthermore, **9a** was also not able to potentiate the activity of the reference drugs ciprofloxacin, linezolid, gentamicin, and daptomycin against *S. aureus*.

In 2015, it was found that myxochelins target human 5-lipoxygenase and inhibit the growth of human K-562 leukemic cells at low micromolar concentrations [22]. Here, we assessed the antiproliferative activity of **9a–9i** using human HCT-116 colorectal carcinoma and human HepG2 hepatocellular carcinoma cells (Table 4). Intriguingly, derivatives **9c**, **9f**, and **9g** efficiently inhibited the growth of both cell lines at low to mid micromolar concentrations being in line with previously reported activities for myxochelin A [22,23]. Moreover, **9b**, **9d**, and **9e** also showed some cytotoxic activity, while **9a**, **9h**, and **9i** were inactive against both cell lines ( $IC_{50} > 37$  µg/mL). In summary, it can be stated that compounds with a nicotinic acid moiety or fewer free phenolic OH groups appeared to be less cytotoxic.

**Table 4.** Antiproliferative activity of synthetic myxochelin derivatives against HCT-116 and HepG2 cell lines. Half-inhibitory concentrations ( $IC_{50}$ ) were determined as duplicates and are given as average values from two independent experiments  $\pm$  SD.

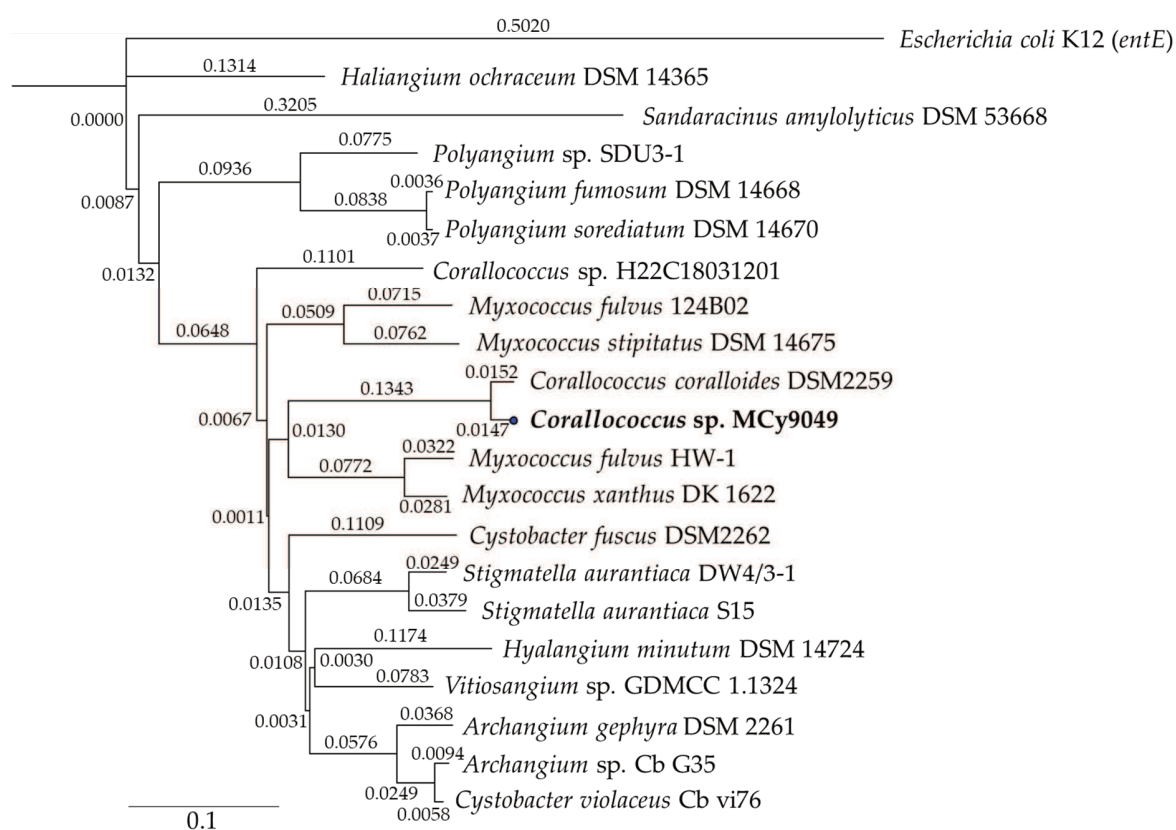
Compound	$IC_{50}$ [µg/mL]	
	HCT-116	HepG2
<b>9a</b>	> 37	> 37
<b>9b</b>	29.1 $\pm$ 3.0	> 37
<b>9c</b>	6.1 $\pm$ 0.1	6.7 $\pm$ 0.9
<b>9d</b>	29.5 $\pm$ 9.2	35.9 $\pm$ 0.9
<b>9e</b>	17.4 $\pm$ 8.0	11.3 $\pm$ 0.7
<b>9f</b>	6.7 $\pm$ 2.0	10.6 $\pm$ 2.6
<b>9g</b>	9.2 $\pm$ 2.3	11.8 $\pm$ 1.7
<b>9h</b>	> 37	> 37
<b>9i</b>	> 37	> 37
Doxorubicin	0.2	0.2

#### 4.3.6 Biosynthetic Origin of the New Myxochelins N1-N3

The production of myxochelin A by myxobacteria is widespread among the *Cystobacterineae* suborder, but also *Nannocystineae* and *Sorangineae* species are able to biosynthesize myxochelins (Figure S71). We performed an extended survey covering the occurrence of **1–3** across myxobacterial taxa, using a previously established collection of high-resolution LC-MS datasets from ca. 2600 myxobacterial strains [24]. This data evaluation was based on exact mass, isotope pattern, and retention time matching, whereas search parameters were: exact mass deviation below 5 ppm and retention time deviation below 0.3 min. Our findings indicate that similar to myxochelin A, the new congeners **1–3** occur predominantly in

*Cystobacterineae* species and exclusively co-occur with iron-chelating myxochelin A (Figure S72). Furthermore, supplementing the fermentation media with ferric ethylenediaminetetraacetic acid (EDTA) led to complete abolishment of myxochelin production, including the nicotinic acid congeners (Figure S33), indicating that negative regulation of biosynthetic genes via iron availability took place.

Analysis of the genome of MCy9049 further showed one myxochelin-like biosynthetic gene cluster (BGC) that could be responsible for the biosynthesis of **1–3**. In order to investigate the loading module MxcE in silico, the sequence of *mxcE* was extracted, translated, and aligned with *mxcE* homologs from published myxobacterial genomes in a phylogenetic tree based on primary amino acid sequence similarity (Figure 6). Additionally, the web server tool NRPSpredictor2 [25] was used to extract the Stachelhaus specificity-conferring codes as well as the 8 Å signature sequences from translated *mxcE* homologs, revealing highly conserved A domain specificity (Table S2).

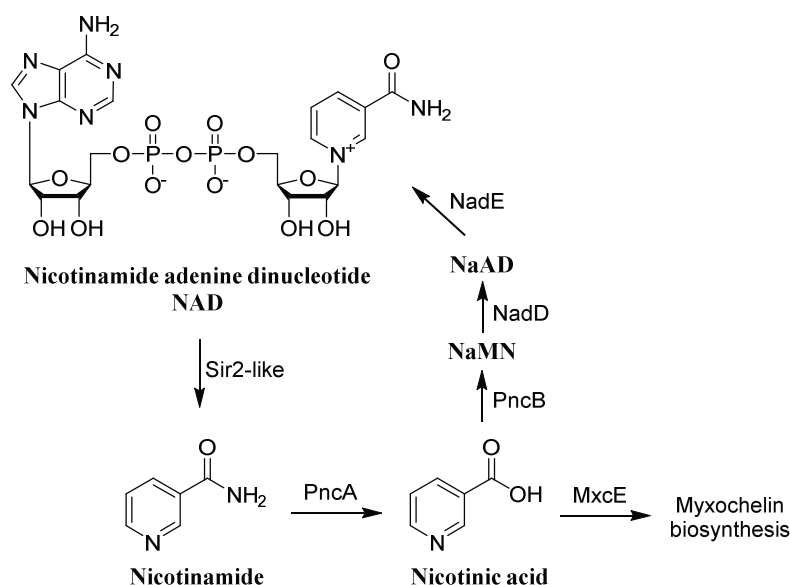


**Figure 6.** Phylogenetic tree based on translation of *mxcE* genes in myxobacterial genomes using *Escherichia coli* K12 enterobactin synthase *entE* (UniProt P10378) as an outgroup.

We conclude from these findings that **1–3** are likely produced via the canonical myxochelin biosynthesis pathway, subject to the availability of nicotinic acid as an alternative precursor. Under

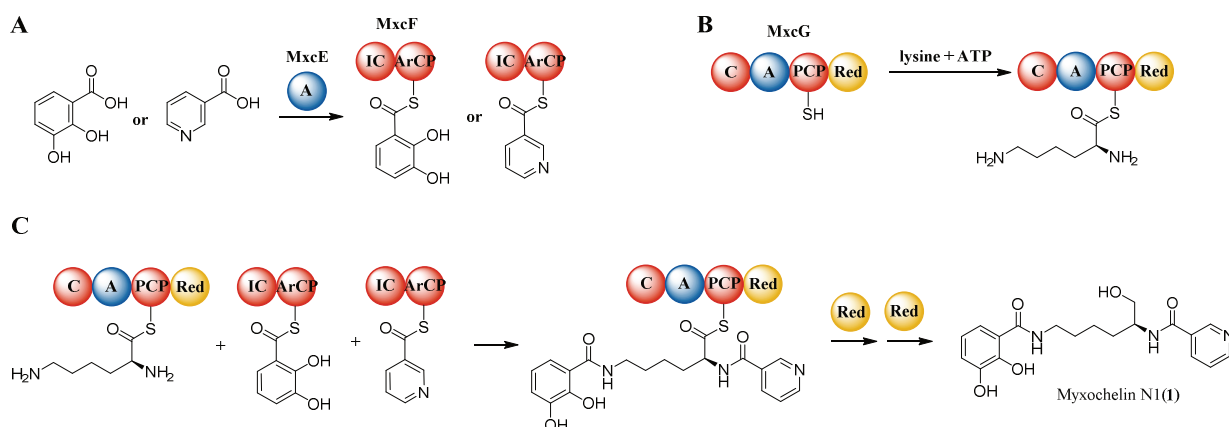
standard laboratory growth conditions, nicotinamide as well as nicotinic acid appear in the nicotinamide adenine dinucleotide (NAD) salvage pathway (Figure 7). NAD-consuming enzymes such as silent information regulator 2 (Sir2)-like deacetylases [26] or ADP-ribosyltransferases [27] produce nicotinamide as enzymatic product. Nicotinamide deaminase (PncA) converts nicotinamide to nicotinate, which can then be used for the biosynthesis of **1–3**. Besides that, the presence of a *pncA* homolog in the genome of MCy9049 could explain why supplementation of nicotinamide leads to an increase in production of **1–3**, albeit NRPS adenylation domains specifically select carboxylic acid moieties for incorporation [28].

Further salvation involves nicotinate phosphoribosyltransferase (PncB) to synthesize nicotinamide mononucleotide (NaMN), and nicotinate mononucleotide adenylyltransferase (NadD) and NAD synthetase (NadE) to re-synthesize NAD via nicotinate adenine dinucleotide (NaAD).



**Figure 7.** Nicotinamide adenine dinucleotide (NAD) salvage pathway similar to [29–31].

Based on the elucidated chemical structure of **1**, we propose the natural product to be biosynthesized as shown in Figure 8. MxcE activates nicotinic acid and DHBA by adenylation and loads it to the ArCP domain of MxcF (Figure 8A). The non-ribosomal peptide synthetase (NRPS) subunit MxcG activates and loads lysine (Figure 8B) and further condenses nicotinic acid and DHBA with the two primary amino groups of L-lysine (Figure 8C). The NADPH-dependent reductase domain of MxcG finally releases an aldehyde intermediate from the assembly line by reduction of the thioester and further reduces it to form **1**. The biosynthesis of **2** and **3** occurs likewise with the other (**2**) or both (**3**) primary amino groups of L-lysine substituted with nicotinic acid instead of DHBA.



**Figure 8.** Proposed biosynthesis of **1** similar to [11]. (A) Activation and loading of DHBA or nicotinic acid to MxcF; (B) Activation and loading of lysine to MxcG; (C) Condensation and reductive release of the product **1**.

## 4.4 Discussion

In this study, we report the isolation and structure elucidation of **1** and **3**, novel secondary metabolites featuring an unusual nicotinic acid moiety, from crude organic extracts of the myxobacterium *Coralloccoccus* sp. MCy9049 as well as the total synthesis of **1**, **2**, and the artificial derivatives **9c–9i**.

Whenever new variants of known natural products—such as nicotinic acid containing congeners of myxochelins in this study—are found, the question emerges how their appearance could be explained by the underlying biosynthetic pathway. From an evolutionary perspective, plausible explanations include the mutational change of substrate specificity in a building block-activating domain to recruit new moieties, possibly following a duplication of the domain or the entire biosynthetic gene cluster. However, the combined evidence from metabolome- and genome-based analysis spanning 105 and 20 myxobacteria, respectively, led us to dismiss these hypothetical explanations in this case. Firstly, we found no indication of a second myxochelin biosynthetic gene cluster in any publicly accessible myxobacterial genome; secondly, the loading module MxcE seems highly conserved across the myxobacteria clade and specificity-conferring residues show little variability; and lastly, production of new myxochelins **1–3** appears stringently coupled to formation of myxochelin A as judged on the basis of LC-MS analysis. We therefore reason that the nicotinic acid containing myxochelin congeners **1–3** reported here are common byproducts of myxochelin A synthesis, wherein their formation is enabled by substrate promiscuity of the A domain MxcE and biogenic precursor availability. A similar production mode depending on availability of nicotinic acid as precursor has also been assumed for asperphenamate biosynthesis in *Penicillium astrolabium* [32].

Nicotinic acid is a rarely occurring precursor for biosynthesis of natural products. Among the few known exceptions are a fungal polyketide synthase (PKS) that incorporates nicotinyI-CoA for

meroterpenoid biosynthesis [33], a PKS from a marine mollusk that uses nicotinic acid as a starter unit to synthesize 3-alkylpyridine alkaloids [34], an NRPS-unit from *Micromonospora* sp. that incorporates nicotinic acid to produce kosinostatin [35], and a fungal NRPS that uses either benzoic acid or nicotinic acid for the biosynthesis of asperphenamates [32]. Furthermore, also in the biosynthesis of bacillibactins, the incorporation of nicotinic or benzoic acid instead of DHBA has been reported and assumed to happen due to a promiscuous DhbE [36], which has a pairwise identity of 67.7% (coverage 99.1%) to MxcE as well as the identical A domain Stachelhaus specificity-conferring code of PLPAQGVVVK (Table S2).

In myxochelin biosynthesis, it has been demonstrated previously that the enzymatic promiscuity of MxcE allows the activation and loading of several benzoic acid derivatives to the assembly line to form precursor-derived non-natural myxochelin derivatives [12–14]. However, incorporation of heteroaromatic carboxylic acid precursors to form natural or artificial myxochelins has so far not been described. Unlike in bacillibactin biosynthesis, where also picolinic acid can be incorporated [37], feeding of pyridinecarboxylic acids only gave relevant yields for nicotinic acid and nicotinamide. This might be an effect of either differences in enzyme promiscuity between DhbE and MxcE, or due to precursor availability as nicotinic acid and nicotinamide, can be actively transported into bacterial cells by niacin transport proteins NiaX, PnuC, NiaP, or NiaY [38]. Surprisingly, supplementation with nicotinamide in our study resulted in about 3-fold higher yield of **1–3** compared to nicotinic acid, which is likely due to decreased cell density in the nicotinic acid-supplemented culture. This is consistent with the finding that also the peak area of myxochelin A is about 3-fold lower with nicotinic acid than with nicotinamide and in the control (Figure 2).

The nicotinic acid containing myxochelins **9a** and **9b** as well as the synthetic analogues **9c–9i** displayed no antimicrobial activity at concentrations up to 128 µg/mL against a small test panel of bacteria and fungi. The reported antibacterial activity of myxochelin A [16] and other catecholate-type iron chelators [39] is attributed to iron competition, and thus, dependent on the presence of both DHBA moieties. As **1–3/9a–9i** are lacking one catecholate moiety needed for iron chelation, these compounds are presumably not functioning as siderophores anymore. This is supported by the finding that the peak area of myxochelin A of cultures supplemented with nicotinamide versus the respective control barely differs (Figure 2). Natural myxochelin A inhibits the growth of various cancer cell lines [9,23]. Encouragingly, three of the new synthetic derivatives (**9c**, **9f**, **9g**) also displayed prominent antiproliferative activity against human cancer cell lines at concentrations in the one-digit µg/mL range.

## 4.5 Conclusion

Taken together, the combined evaluation of metabolome and genome data in this study supports the notion that **1–3** are previously overlooked byproducts of myxochelin A biosynthesis, which are commonly produced by myxobacteria upon precursor availability. Although the biological function of these new myxochelin congeners remains elusive at present, the presented discovery of **1–3**, featuring an unusual nicotinic acid moiety, as well as the total synthesis of **9a–9i**, set the stage for further biosynthetic studies, the generation of new myxochelin derivatives, and identification of the biological function of **1–3**.

## 4.6 Materials and Methods

### 4.6.1 Myxobacterial Fermentation and Small-Scale Precursor Feeding

*Cystobacterineae* species *Corallococcus* sp. MCy9049 was routinely cultivated at 30 °C in VY/2S medium (1.0% baker's yeast (F. X. Wieninger GmbH, Passau, Germany), 0.5% soluble starch (Carl Roth GmbH, Karlsruhe, Germany), 0.1% CaCl<sub>2</sub> × 2 H<sub>2</sub>O (VWR International GmbH, Darmstadt, Germany), 0.1% MgSO<sub>4</sub> × 7 H<sub>2</sub>O (Grüssing GmbH, Filsum, Germany), 10 mM TRIS (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) adjusted to pH 7.2 with 1 M HCl before autoclaving). Liquid cultures were grown in Erlenmeyer shake flasks without baffles on an orbital shaker at 180 rpm for seven days. Feeding of precursors was performed by cultivating the strain in 50 mL VY/2S medium using 1 mL inoculum (2% inoculum volume). The cultures were supplemented with 1 mL (v/v) sterile amberlite resin XAD-16 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) and with either 500 µL of a 0.1 M aqueous solution of nicotinic acid, nicotinamide, or picolinic acid, or with 1.25 mL of a 0.04 M solution of isonicotinic acid (abcr GmbH, Karlsruhe, Germany), resulting in a final concentration of 1 mM each. After seven days, the combined cells and resin were separated from the supernatant by centrifugation. While the supernatant was discarded, the pellet was extracted with 50 mL methanol, stirred for 1 h, and filtered through filter paper. After the solvent was evaporated under vacuum, the extracts were redissolved in 1.5 mL of methanol, further diluted (1:5 with methanol, v/v) and centrifuged, and the supernatant was analyzed by high-performance liquid chromatography (HPLC) coupled to a mass spectrometer (MS) (Bruker Daltonics, Billerica, MA, USA).

### 4.6.2 Analysis of the Secondary Metabolome of Bacterial Crude Extracts

The secondary metabolome of bacterial crude extracts was analyzed on a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to



a Bruker maXis 4G ultra-high-resolution quadrupole time-of-flight (UHR-qTOF) MS equipped with a high-resolution electrospray ionization (HRESI) source (Bruker Daltonics, Billerica, MA, USA). Separation of 1  $\mu\text{L}$  sample was achieved with a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH<sub>2</sub>O with 0.1% formic acid on an ACQUITY BEH C18 column (100  $\times$  2.1 mm, 1.7  $\mu\text{m}$  d<sub>p</sub>) (Waters, Eschborn, Germany) at a flow rate of 0.6 mL/min and 45 °C over 18 min with detection by a diode array detector at 200–600 nm. The LC flow was split to 75  $\mu\text{L}$  before entering the mass spectrometer. Mass spectrograms were acquired in centroid mode ranging from 150–2500  $m/z$  at an acquisition rate of 2 Hz in positive MS mode. Source parameters were set to 500 V end plate offset; 4000 V capillary voltage; 1 bar nebulizer gas pressure; 5 L/min dry gas flow; and 200 °C dry gas temperature. Calibration was conducted automatically before every HPLC-MS run by injection of sodium formate and calibration on the respective clusters formed in the ESI source. All MS analyses were acquired in the presence of the lock masses C<sub>12</sub>H<sub>19</sub>F<sub>12</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, C<sub>18</sub>H<sub>19</sub>F<sub>24</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, and C<sub>24</sub>H<sub>19</sub>F<sub>36</sub>N<sub>3</sub>O<sub>6</sub>P<sub>3</sub>, which generate the [M + H]<sup>+</sup> ions of 622.0289, 922.0098, and 1221.9906. The HPLC-MS system was operated by HyStar 5.1 (Bruker Daltonics, Billerica, MA, USA) and LC chromatograms as well as UV spectra and mass spectrograms were analyzed with DataAnalysis 4.4 (Bruker Daltonics, Billerica, MA, USA).

For statistical metabolomics analysis, both the strain and medium blanks were cultivated and extracted in triplicates as described before. Each extract was measured as technical duplicates giving a total number of six replicates each for the bacterial and medium blank extracts. T-ReX-3D molecular feature finder of MetaboScape 6.0.2 (Bruker Daltonics, Billerica, MA, USA) was used to obtain molecular features. Detection parameters were set to intensity threshold  $5 \times 10^3$  and minimum peak length of five spectra. Identification of bacterial features was performed with the built in t-test routine and filtered to appearance in all six bacterial extracts and in none of medium blank extracts.

### 4.6.3 Compound Isolation

For compound isolation, *Coralloccoccus* sp. MCy9049 was cultivated in 10 L VY/2S medium containing 1 mm Nicotinamide and 2% XAD-16 resin ( $v/v$ ) as described before. After seven days of fermentation, combined cells and resin were harvested by centrifugation and extracted with 1 L methanol twice. After drying under vacuum, the extract was redissolved in 500 mL methanol and partitioned using 500 mL *n*-hexane. The methanol layer was dried, redissolved in 500 mL ddH<sub>2</sub>O, and further partitioned using 500 mL chloroform. The aqueous layer was once more partitioned using 500 mL ethyl acetate. The ethyl acetate layer was dried to yield 55.2 mg of extract and redissolved in 2 mL methanol. Purification was performed using a Dionex UltiMate 3000 semi-preparative system equipped with an automated fraction collector (Thermo Fisher Scientific, Waltham, MA, USA). Compound separation was achieved with a gradient of

acetonitrile with 0.1% formic acid (B) in ddH<sub>2</sub>O with 0.1% formic acid (A) on an ACQUITY BEH C18 column (250 × 10 mm, 5 μm d<sub>p</sub>) (Waters, Eschborn, Germany) at a flow rate of 5.0 mL/min and 45 °C. The initial gradient was held at 5% B for 1 min and then elevated to 35% B within 15 min. After that, the B level was increased to 95% and held there for one minute. Finally, the gradient was ramped back to 5% B in 1 min and re-equilibrated for the next injection for 3 min. Detection was performed using the 3D plot of a DAD detector by absorption at 220, 245, 260, and 310 nm. The isolated pure compounds were dried by lyophilization.

#### 4.6.4 Structure Elucidation, Chiroptical and Circular Dichroism Measurements

The chemical structures of all the compounds were determined via multidimensional NMR analysis. <sup>1</sup>H-NMR, <sup>13</sup>C-NMR, and 2D spectra were recorded at 700 and 500 MHz (<sup>1</sup>H)/175 and 126 MHz (<sup>13</sup>C), conducted in the Ascend 700 spectrometer using a cryogenically cooled triple resonance probe (Bruker BioSpin, Rheinstetten, Germany) or in the Bruker Avance Neo 500 MHz, equipped with a Prodigy Cryo-probe. Samples were dissolved in methanol-d<sub>4</sub>, acetone-d<sub>6</sub>, or dimethyl sulfoxide-d<sub>6</sub>. Chemical shifts are reported in ppm relative to tetramethylsilane; the solvent was used as the internal standard. Coupling constants are reported in Hertz (Hz). Multiplicity is reported with the usual abbreviations (s: singlet, br s: broad singlet, d: doublet, dd: doublet of doublets, ddd: doublet of doublet of doublets, t: triplet, dt: doublet of triplets, q: quartet, p: pentet, dp: doublet of pentets, m: multiplet). Chiroptical measurements of all the compounds in MeOH ([α]<sub>D</sub>) were obtained on a model 341 polarimeter (PerkinElmer Inc. Waltham, MA, USA) in a 100 × 2 mm cell at 25 °C. Circular dichroism measurements were performed for **1** and its synthetic analogue **9a** at 1.0 mg/mL in MeOH (190–400 nm) with the J-1500 CD spectrophotometer (JASCO, Easton, MD, USA).

*Myxochelin N1* (**1**), yellow oil (0.3 mg), [α]<sub>D</sub><sup>20</sup> –18.6 (c 0.025, MeOH); <sup>1</sup>H NMR (700 MHz, methanol-d<sub>4</sub>): δ 8.96 (d, *J* = 2.2 Hz, 1H), 8.66 (brd, *J* = 8.2 Hz, 1H), 8.23 (brd, *J* = 7.7 Hz, 1H), 7.49 (dd, *J* = 8.0, 5.0 Hz, 1H), 7.16 (d, *J* = 8.1 Hz, 1H), 6.90 (d, *J* = 8.0 Hz, 1H), 6.67 (t, *J* = 8.0 Hz, 1H), 4.15 (m, 1H), 3.63 (brd, *J* = 3.4 Hz, 1H), 3.61 (brd, *J* = 3.6 Hz, 1H), 3.39 (t, *J* = 7.0 Hz, 2H), 1.72 (m, 2H), 1.63 (m, 2H), 1.49 (m, 2H); <sup>13</sup>C NMR (126 MHz, methanol-d<sub>4</sub>): δ 170.1, 166.7, 151.0, 148.6, 147.7, 145.9, 135.7, 130.9, 123.6, 118.1, 118.1, 117.1, 115.3, 64.4, 53.2, 39.4, 30.2, 28.9, 22.6; HRMS (ESI) (*m/z*): calcd for C<sub>19</sub>H<sub>24</sub>N<sub>3</sub>O<sub>5</sub> 374.1710, found 374.1710 [M + H]<sup>+</sup>.

*Myxochelin N3* (**3**), yellow oil (0.1 mg), [α]<sub>D</sub><sup>20</sup> –26.4 (c 0.025, MeOH); <sup>1</sup>H NMR (700 MHz, methanol-d<sub>4</sub>): δ 8.95 (dd, *J* = 2.2 Hz, 1H), 8.65 (dt, *J* = 8.4, 1.6 Hz, 1H), 8.20 (ddt, *J* = 8.4, 2.2, 2.0 Hz, 1H), 7.51 (dd, *J* = 7.9, 2.3 Hz, 1H), 4.15 (m, 1H), 4.15 (m, 1H), 3.62 (dd, *J* = 10.7, 5.4 Hz, 1H), 3.61 (dd, *J* = 10.6, 5.2 Hz, 1H), 3.42 (m, 2H), 1.74 (m, 2H), 1.64 (m, 2H), 1.49 (m, 2H); <sup>13</sup>C NMR (126 MHz, methanol-d<sub>4</sub>): δ 167.2, 151.7,

148.2, 136.4, 131.2, 123.6, 64.8, 53.7, 40.4, 30.2, 29.2, 23.4.; HRMS (ESI) ( $m/z$ ): calcd for  $C_{18}H_{23}N_4O_3$  343.1765, found 343.1766  $[M + H]^+$ .

#### 4.6.5 Chemistry

##### 4.6.5.1 General Procedure for Synthetic Step a

Fmoc-L-lysine hydrochloride (1 eq) was suspended in ethanol and was cooled to 0 °C. Then, chlorotrimethylsilane (2 eq) was added and the resulting suspension was allowed to warm up to r. t. and it was stirred for 16 h. The remaining ethanol and chlorotrimethylsilane were evaporated under reduced pressure. The crude product of **4** was obtained as a white powder and was used without further purification.

##### 4.6.5.2 General Procedure for Synthetic Step b

Fmoc-L-lysine ethyl ester hydrochloride (**4**) (1 eq) was dissolved in dry DMF (5.0 mL). Subsequently, 2,3-dimethoxybenzoic acid (2 eq) or nicotinic acid (2 eq) and HATU (2 eq) were dissolved in dry DMF (10.0 mL) and were added to the reaction flask. The resulting solution was cooled to 0 °C and TEA (2.5 eq) was added dropwise over 5 min. Then, the solution was allowed to warm to r. t. and it was stirred for 16 h. After completion of the reaction (monitored via LC-MS), the reaction solution was poured into ice water (100 mL). The mixture was exhaustively extracted with ethyl acetate (3 × 50 mL). The organic layers were combined, dried over sodium sulfate, and filtered. After evaporation of the solvent, the crude product was purified by automated flash chromatography using petroleum ether/ethyl acetate (100%:0% → 0%:100%) as eluent. The product was a yellow oil.

##### 4.6.5.3 General Procedure for Synthetic Step c

To the ethyl  $N^2$ -Fmoc- $N^6$ -aryl-L-lysinate compound **5a** or **5b** (1 eq) a mixture of DMA and DCM (1:1 v/v) was added and the resulting solution was stirred at r. t. for 2 h. Then, the solvent and residual dimethylamine was removed under reduced pressure. Subsequently, the residue was dissolved in a mixture of saturated ammonium chloride solution and ethyl acetate (1:1 v/v, 50 mL), the aqueous phase was extracted with ethyl acetate (3 × 50 mL). Afterwards, saturated sodium bicarbonate solution was added to the aqueous phase until the pH was set to ~8. Then, extraction of the aqueous phase was repeated with ethyl acetate (3 × 50 mL). The organic phases were combined, dried over sodium sulfate, and filtered. After the evaporation of the solvent, the crude product was obtained as a viscous yellow oil and was used without further purification.

#### 4.6.5.4 General Procedure for Synthetic Step **d**

The ethyl  $N^6$ -aryl-L-lysinate compound **6a** or **6b** (1 eq) was dissolved in dry DMF (5.0 mL). Subsequently, HATU (2 eq) and nicotinic acid (2 eq) or a benzoic acid derivative (2 eq) were dissolved in dry DMF (10.0 mL) and were added to the reaction flask. Lastly, TEA (2.5 eq) was added dropwise and the solution was stirred at r. t. for 16 h. After completion, the reaction mixture was poured into ice water (100 mL). The mixture was exhaustively extracted with ethyl acetate (3 × 50 mL). The organic layers were combined, dried over sodium sulfate and filtered. After evaporation of the solvent, the crude product was taken forward either without further purification or it was purified by automated reverse-phase flash chromatography, using water/acetonitrile (95%:5% → 5%:95%) as eluent. The product was a viscous yellow oil.

#### 4.6.5.5 General Procedure for Synthetic Step **e**

Ethyl  $N^6$ -(2,3-dimethoxybenzoyl)-L-lysinate **6a** (1 eq) was dissolved in dry DMF (5.0 mL). Subsequently, a Fmoc-protected aminobenzoic acid derivative (2 eq) and HATU (2 eq) were dissolved in dry DMF (10.0 mL) and were added to the reaction flask. The resulting solution was cooled to 0 °C and TEA (2.5 eq) was added dropwise over 5 min and the solution was stirred at r. t. for 16 h. After completion, DMA (5.0 mL) was added and the solution was stirred at r. t. for 2 h. Then, the reaction mixture was poured into ice water (100 mL). The mixture was exhaustively extracted with ethyl acetate (3 × 50 mL). The organic layers were combined, dried over sodium sulfate, and filtered. After evaporation of the solvent, the crude product was taken forward without further purification or it was purified by automated reverse-phase flash chromatography, using water/acetonitrile (95%:5% → 5%:95%) as eluent. The product was a viscous yellow oil.

#### 4.6.5.6 General Procedure for Synthetic Step **f**

Ethyl  $N^2$ -aryl- $N^6$ -aryl-L-lysinate **7a–7g** (1 eq) was dissolved in dry THF (1.0 mL) under a  $N_2$  atmosphere. Then, lithium borohydride (2 M in THF, 1 eq) was added at r. t. and the resulting mixture was stirred for 2 h. After completion, the reaction was quenched by addition of saturated aqueous ammonium chloride solution (3 mL). The resulting mixture was stirred at r. t. for 30 min. Then, the solution was exhaustively extracted with DCM (3 × 20 mL). The organic layers were combined, dried over sodium sulfate, and filtered. After evaporation of the solvent, the crude product was purified by automated flash chromatography, using DCM/methanol (100%:0% → 90%:10%) as eluent. The product was a viscous yellow oil.

#### 5.5.7. General Procedure for Synthetic Step **g**

The dimethoxybenzamide containing compound **8a–8g** (1 eq) was dissolved in dry DCM (1.0 mL) under a N<sub>2</sub> atmosphere. Boron tribromide (1 M in DCM, 4 eq) was added and the reaction mixture was stirred at r. t. for 16 h. After completion, the reaction was quenched by addition of water (2 mL) and the solvent mixture was evaporated to dryness. Then, the crude product was purified by preparative HPLC, using water/acetonitrile (95%:5% → 80%:20%) as eluent. The fractions with the product were collected and freeze dried. The product was a light-yellow powder.

#### 4.6.6 Characterization of Synthetic Intermediates and Final Products

Ethyl Fmoc-L-lysinate (**4**), General procedure **a** was followed with Fmoc-L-lysine hydrochloride (2.43 g, 6.0 mmol) and chlorotrimethylsilane (1.50 mL, 12.0 mmol) to afford intermediate **4** as a white solid (yield: 2.60 g, quant.). <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ 7.90 (d, *J* = 7.5 Hz, 2H), 7.82 (br s, 2H), 7.71 (t, *J* = 7.5 Hz, 2H), 7.42 (t, *J* = 7.4 Hz, 2H), 7.33 (t, *J* = 7.4 Hz, 2H), 4.39–4.17 (m, 3H), 4.09 (qd, *J* = 7.2, 2.3 Hz, 2H), 3.97 (ddd, *J* = 9.6, 7.8, 4.9 Hz, 1H), 2.75 (h, *J* = 6.5 Hz, 2H), 1.74–1.27 (m, 6H), 1.17 (t, *J* = 7.1 Hz, 3H); ESI-MS (*m/z*): calcd for C<sub>23</sub>H<sub>29</sub>N<sub>2</sub>O<sub>4</sub> 397.21, found 397.14 [M + H]<sup>+</sup>.

Ethyl N<sup>2</sup>-Fmoc-N<sup>6</sup>-(2,3-dimethoxybenzoyl)-L-lysinate (**5a**). General procedure **b** was followed with **4** (1.95 g, 4.5 mmol), 2,3-dimethoxybenzoic acid (1.64 g, 9.0 mmol), HATU (3.42 g, 9.0 mmol), and TEA (2.50 mL, 18 mmol) to afford intermediate **5a** as a yellow oil (2.28 g, 90%, purity >83%). <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ 8.19 (t, *J* = 5.8 Hz, 1H), 7.89 (d, *J* = 7.5 Hz, 2H), 7.76 (d, *J* = 7.8 Hz, 1H), 7.71 (dd, *J* = 7.5, 2.9 Hz, 2H), 7.41 (t, *J* = 7.5 Hz, 2H), 7.32 (t, *J* = 7.4 Hz, 2H), 7.23–7.16 (m, 1H), 7.16–7.05 (m, 1H), 4.35–4.18 (m, 3H), 4.18–3.94 (m, 3H), 3.81 (s, 3H), 3.74 (s, 3H), 3.26–3.21 (m, 2H), 1.88–1.61 (m, 2H), 1.52 (dq, *J* = 14.3, 6.9 Hz, 2H), 1.40 (td, *J* = 14.2, 12.5, 8.1 Hz, 2H), 1.16 (t, *J* = 7.0 Hz, 3H); <sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>): δ 172.5, 165.4, 156.1, 152.5, 146.2, 143.8, 140.8, 130.3, 127.7, 127.1, 125.2, 124.0, 120.6, 120.1, 114.5, 65.6, 60.9, 60.4, 55.9, 54.0, 46.6, 38.6, 30.4, 28.6, 23.0, 14.1; HRMS (ESI) (*m/z*): calcd for C<sub>32</sub>H<sub>37</sub>N<sub>2</sub>O<sub>7</sub> 561.2601, found 561.2584 [M + H]<sup>+</sup>.

Ethyl N<sup>2</sup>-Fmoc-N<sup>6</sup>-nicotinoyl-L-lysinate (**5b**). General procedure **b** was followed with **4** (650 mg, 1.5 mmol), nicotinic acid (369 mg, 3.0 mmol), HATU (1.14 g, 3.0 mmol), and TEA (0.850 mL, 6.0 mmol) to afford intermediate **5b** as a yellow oil (701 mg, 93%, purity >87%). <sup>1</sup>H NMR (500 MHz, methanol-*d*<sub>4</sub>): δ 8.83 (d, *J* = 2.3 Hz, 1H), 8.51 (dd, *J* = 4.9, 1.6 Hz, 1H), 8.12 (dt, *J* = 8.0, 2.0 Hz, 1H), 7.63 (d, *J* = 7.5 Hz, 2H), 7.50 (t, *J* = 7.7 Hz, 2H), 7.40 (dd, *J* = 8.0, 5.0 Hz, 1H), 7.22 (t, *J* = 7.5 Hz, 2H), 7.14 (t, *J* = 7.5 Hz, 2H), 4.21–4.10 (m, 2H), 4.07–3.98 (m, 2H), 3.94 (q, *J* = 7.1 Hz, 2H), 3.29 (t, *J* = 7.0 Hz, 2H), 1.68–1.29 (m, 6H), 1.15 (t, *J* = 7.3 Hz, 3H); <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>): δ 172.5, 166.0, 165.9, 156.4, 152.2, 148.1, 144.0, 143.7, 141.4, 135.2, 130.4, 130.3, 127.9, 127.2, 125.2, 123.5, 120.1, 67.2, 61.8, 53.5, 47.2, 39.7, 32.7, 28.6, 22.6, 14.3; HRMS (ESI) (*m/z*): calcd for C<sub>29</sub>H<sub>32</sub>N<sub>3</sub>O<sub>5</sub> 502.2342, found 502.2323 [M + H]<sup>+</sup>.

Ethyl N<sup>6</sup>-(2,3-dimethoxybenzoyl)-L-lysinate (**6a**). General procedure **c** was followed with **5a** (2.28 g, 4.1 mmol) and DMA/DCM (1:1 v/v, 9 mL) to afford intermediate **6a** as a yellow oil (2.34 g, 70%, purity >56%). HRMS (ESI) (*m/z*): calcd for C<sub>17</sub>H<sub>27</sub>N<sub>2</sub>O<sub>5</sub> 339.1920, found 339.1904 [M + H]<sup>+</sup>.

Ethyl N<sup>6</sup>-nicotinoyl-L-lysinate (**6b**). General procedure **c** was followed with **5b** (701 mg, 1.4 mmol) and DMA/DCM (1:1 v/v, 3 mL) to afford crude intermediate **6b** as a yellow oil (not isolated, purity >53%). HRMS (ESI) (*m/z*): calcd for C<sub>14</sub>H<sub>22</sub>N<sub>3</sub>O<sub>3</sub> 280.1661, found 280.1646 [M + H]<sup>+</sup>.

Ethyl N<sup>6</sup>-(2,3-dimethoxybenzoyl)-N<sup>2</sup>-nicotinoyl-L-lysinate (**7a**). General procedure **d** was followed with **6a** (1.53 g, 4.5 mmol), nicotinic acid (1.11 g, 9.0 mmol), HATU (3.42 g, 9.0 mmol), and TEA (2.50 mL, 18.0 mmol) and purified by automated reverse-phase flash chromatography to afford intermediate **7a** as a yellow oil (850 mg, 43%). <sup>1</sup>H NMR (500 MHz, methanol-*d*<sub>4</sub>): δ 8.99 (d, *J* = 2.3 Hz, 1H), 8.69 (dt, *J* = 4.9, 1.2 Hz, 1H), 8.26 (dt, *J* = 8.0, 2.2 Hz, 1H), 7.53 (dd, *J* = 8.0, 4.9 Hz, 1H), 7.26 (dd, *J* = 7.5, 1.9 Hz, 1H), 7.18–7.08 (m, 2H), 4.67–4.56 (m, 1H), 4.21 (q, *J* = 7.1 Hz, 2H), 3.87 (s, 3H), 3.83 (s, 3H), 3.45–3.39 (m, 2H), 2.08–1.86 (m, 2H), 1.70 (m, 2H), 1.67–1.51 (m, 2H), 1.28 (t, *J* = 7.1, 3H); <sup>13</sup>C NMR (126 MHz, methanol-*d*<sub>4</sub>): δ 173.7, 168.7, 168.2, 154.2, 152.8, 149.3, 148.4, 137.2, 131.6, 129.5, 125.4, 125.1, 122.2, 116.4, 62.5, 61.8, 56.6, 54.6, 40.2, 31.8, 30.0, 24.5, 14.5; HRMS (ESI) (*m/z*): calcd for C<sub>23</sub>H<sub>30</sub>N<sub>3</sub>O<sub>6</sub> 444.2135, found 444.2116 [M + H]<sup>+</sup>.

Ethyl N<sup>2</sup>-(2,3-dimethoxybenzoyl)-N<sup>6</sup>-nicotinoyl-L-lysinate (**7b**). General procedure **d** was followed with **6b** (508 mg, 1.5 mmol), 2,3-dimethoxybenzoic acid (546 mg, 3.0 mmol), HATU (1.14 g, 3.0 mmol), and TEA (0.850 mL, 6.0 mmol) and purified by automated reverse-phase flash chromatography to afford intermediate **7b** as a yellow oil (701 mg, 93%, purity >65%). <sup>1</sup>H NMR (500 MHz, methanol-*d*<sub>4</sub>): δ 8.93 (d, *J* = 2.3 Hz, 1H), 8.66 (dd, *J* = 4.9, 1.7 Hz, 1H), 8.20 (dt, *J* = 8.1, 2.0 Hz, 1H), 7.51 (dd, *J* = 8.0, 5.0 Hz, 1H), 7.37 (dd, *J* = 7.7, 1.7 Hz, 1H), 7.20 (dd, *J* = 8.2, 1.8 Hz, 1H), 7.14 (t, *J* = 8.0 Hz, 1H), 4.70–4.63 (m, 1H), 4.22 (qd, *J* = 7.1, 1.9 Hz, 2H), 3.89 (br s, 6H), 3.46–3.37 (m, 2H), 2.07–1.96 (m, 1H), 1.94–1.83 (m, 1H), 1.77–1.63 (m, 2H), 1.57–1.43 (m, 2H), 1.28 (t, *J* = 7.1 Hz, 3H); <sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>): δ 172.0, 165.4, 164.7, 152.6, 151.7, 148.3, 146.5, 134.9, 130.1, 128.7, 124.1, 123.4, 120.8, 115.2, 61.0, 60.6, 56.0, 52.3, 38.9, 30.7, 28.6, 22.8, 14.1; HRMS (ESI) (*m/z*): calcd for C<sub>23</sub>H<sub>30</sub>N<sub>3</sub>O<sub>6</sub> 444.2135, found 444.2119 [M + H]<sup>+</sup>.

Ethyl N<sup>2</sup>-(2-chlorobenzoyl)-N<sup>6</sup>-(2,3-dimethoxybenzoyl)-L-lysinate (**7c**). General procedure **d** was followed with **6a** (676 mg, 2.0 mmol), 2-chlorobenzoic acid (312 mg, 2.0 mmol), HATU (760 mg, 2.0 mmol), and TEA (0.600 mL, 4.0 mmol) to afford intermediate **7c** as a yellow oil (161 mg, 17%). <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ 8.82 (d, *J* = 7.5 Hz, 1H), 8.20 (t, *J* = 5.6 Hz, 1H), 7.51–7.48 (m, 1H), 7.48–7.43 (m, 1H), 7.41–7.36 (m, 2H), 7.14 (dd, *J* = 6.7, 3.2 Hz, 1H), 7.11–7.06 (m, 2H), 4.35 (ddd, *J* = 9.6, 7.4, 4.8 Hz, 1H), 4.21–4.02 (m, 2H), 3.78 (s, 3H), 3.71 (s, 3H), 3.24 (m, 2H), 1.84–1.41 (m, 6H), 1.21 (t, *J* = 7.1 Hz, 3H); <sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>): δ 172.3, 167.0, 165.9, 153.0, 146.6, 136.9, 131.4, 130.8, 130.5, 130.1, 129.4, 127.5, 124.5,

121.0, 115.0, 61.4, 61.0, 56.4, 53.0, 39.1, 30.7, 29.1, 23.4, 14.6; HRMS (ESI) ( $m/z$ ): calcd for  $C_{24}H_{30}ClN_2O_6$  477.1792, found 477.1776 [M + H]<sup>+</sup>.

Ethyl N<sup>2</sup>-(2,3-dichlorobenzoyl)-N<sup>6</sup>-(2,3-dimethoxybenzoyl)-L-lysinate (**7d**). General procedure **d** was followed with **6a** (676 mg, 2.0 mmol), 2,3-dichlorobenzoic acid (380 mg, 2.0 mmol), HATU (760 mg, 2.0 mmol), and TEA (0.600 mL, 4.0 mmol) to afford intermediate **7d** as a yellow oil (51.0 mg, 5%). <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>): 8.91 (d,  $J$  = 7.3 Hz, 1H), 8.72 (dd,  $J$  = 4.8, 1.7 Hz, 1H), 8.30–8.02 (m, 2H), 7.51 (ddd,  $J$  = 8.0, 4.9, 0.9 Hz, 1H), 7.13 (dd,  $J$  = 6.9, 3.0 Hz, 1H), 7.10–6.94 (m, 2H), 4.40 (ddd,  $J$  = 9.1, 7.2, 5.6 Hz, 1H), 4.11 (q,  $J$  = 7.1, 2H), 3.81 (s, 3H), 3.72 (s, 3H), 3.25 (q,  $J$  = 6.6 Hz, 2H), 1.91–1.37 (m, 6H), 1.19 (t,  $J$  = 7.1 Hz, 3H); HRMS (ESI) ( $m/z$ ): calcd for  $C_{24}H_{29}Cl_2N_2O_6$  511.1403, found 511.1388 [M + H]<sup>+</sup>.

Ethyl N<sup>2</sup>-(2-aminobenzoyl)-N<sup>6</sup>-(2,3-dimethoxybenzoyl)-L-lysinate (**7e**). General procedure **e** was followed with **6a** (676 mg, 2.0 mmol), *N*-Fmoc-2-aminobenzoic acid (**SI-1**) (718 mg, 2.0 mmol), HATU (760 mg, 2.0 mmol), and TEA (0.600 mL, 4.0 mmol) to afford intermediate **7e** as a yellow oil (128 mg, 14%, purity >74%). <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ 8.38 (d,  $J$  = 7.2 Hz, 1H), 8.20 (t,  $J$  = 5.8 Hz, 1H), 7.57 (dd,  $J$  = 8.0, 1.6 Hz, 1H), 7.18–7.01 (m, 4H), 6.68 (dd,  $J$  = 8.4, 1.2 Hz, 1H), 6.49 (ddd,  $J$  = 8.1, 7.1, 1.2 Hz, 1H), 6.35 (s, 2H), 4.31 (q,  $J$  = 7.2 Hz, 1H), 4.17–4.01 (m, 2H), 3.82 (s, 3H), 3.73 (s, 3H), 3.25 (q,  $J$  = 6.5 Hz, 2H), 1.80 (t,  $J$  = 7.6 Hz, 2H), 1.59–1.24 (m, 4H), 1.18 (t,  $J$  = 7.1 Hz, 3H); HRMS (ESI) ( $m/z$ ): calcd for  $C_{24}H_{32}N_3O_6$  458.2291, found 458.2273 [M + H]<sup>+</sup>.

Ethyl N<sup>2</sup>-(2-amino-3-chlorobenzoyl)-N<sup>6</sup>-(2,3-dimethoxybenzoyl)-L-lysinate (**7f**). General procedure **e** was followed with **6a** (676 mg, 2.0 mmol), *N*-Fmoc-2-amino-3-chlorobenzoic acid (**SI-2**) (786 mg, 2.0 mmol), HATU (760 mg, 2.0 mmol), and TEA (0.600 mL, 4.0 mmol) to afford intermediate **7f** as a yellow oil (125 mg, 13%). <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ 8.92 (d,  $J$  = 7.5 Hz, 1H), 8.20 (t,  $J$  = 5.8 Hz, 1H), 7.71 (dd,  $J$  = 8.0, 1.6 Hz, 1H), 7.42 (t,  $J$  = 7.8 Hz, 1H), 7.35 (dd,  $J$  = 7.6, 1.6 Hz, 1H), 7.18–7.05 (m, 3H), 4.36 (ddd,  $J$  = 9.6, 7.5, 4.8 Hz, 1H), 4.20–4.06 (m, 2H), 3.82 (s, 3H), 3.74 (s, 3H), 3.24 (qd,  $J$  = 6.7, 1.8 Hz, 2H), 1.89–1.65 (m, 2H), 1.63–1.38 (m, 4H), 1.21 (t,  $J$  = 7.1 Hz, 3H); <sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>): δ 171.7, 165.9, 165.5, 152.5, 146.2, 138.8, 132.1, 131.2, 130.3, 128.5, 128.2, 127.4, 124.0, 120.6, 114.5, 60.9, 60.6, 56.0, 52.5, 38.6, 30.2, 28.6, 22.9, 14.1; HRMS (ESI) ( $m/z$ ): calcd for  $C_{24}H_{31}ClN_3O_6$  492.1901, found 492.1881 [M + H]<sup>+</sup>.

Ethyl N<sup>2</sup>-(2-amino-6-fluorobenzoyl)-N<sup>6</sup>-(2,3-dimethoxybenzoyl)-L-lysinate (**7g**). General procedure **e** was followed with **6a** (676 mg, 2.0 mmol), *N*-Fmoc-2-amino-6-fluorobenzoic acid (**SI-3**) (754 mg, 2.0 mmol), HATU (760 mg, 2.0 mmol), and TEA (0.600 mL, 4.0 mmol) to afford intermediate **7g** as a yellow oil (130 mg, 14%, purity >76%). <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ 8.57 (dd,  $J$  = 7.1, 2.0 Hz, 1H), 8.19 (dd,  $J$  = 7.0, 4.5 Hz, 1H), 7.91–7.78 (m, 1H), 7.76–7.57 (m, 1H), 7.44–7.23 (m, 1H), 7.14–7.02 (m, 1H), 6.50 (dd,  $J$  = 8.3, 0.9 Hz, 1H), 6.35–6.25 (m, 1H), 5.78 (s, 2H), 4.34 (ddd,  $J$  = 9.2, 7.0, 5.1 Hz, 1H), 4.17–4.07 (m, 2H), 3.82 (s, 3H),

3.74 (s, 3H), 3.23 (qd,  $J = 6.7, 2.6$  Hz, 2H), 1.83–1.71 (m, 2H), 1.57–1.37 (m, 4H), 1.20 (t,  $J = 7.1$  Hz, 3H); HRMS (ESI) ( $m/z$ ): calcd for  $C_{24}H_{31}FN_3O_6$  476.2197, found 476.2180  $[M + H]^+$ .

(*S*)-*N*-(6-(2,3-Dimethoxybenzamido)-1-hydroxyhexan-2-yl)nicotinamide (**8a**). General procedure **f** was followed with **7a** (850 mg, 1.92 mmol) and lithium borohydride (2 M in THF, 0.950 mL) to afford intermediate **8a** as a yellow oil (330 mg, 43%).  $^1H$  NMR (500 MHz, methanol- $d_4$ ):  $\delta$  8.97 (d,  $J = 2.3$  Hz, 1H), 8.66 (dd,  $J = 4.9, 1.6$  Hz, 1H), 8.23 (dt,  $J = 8.1, 2.0$  Hz, 1H), 7.50 (dd,  $J = 8.0, 4.9$  Hz, 1H), 7.26 (dd,  $J = 7.6, 1.9$  Hz, 1H), 7.17–7.07 (m, 2H), 4.61 (d,  $J = 19.5$  Hz, 1H), 4.16 (dq,  $J = 10.3, 5.3$  Hz, 1H), 3.87 (s, 3H), 3.81 (s, 3H), 3.80 (d,  $J = 11.2$  Hz, 1H), 3.54–3.35 (m, 4H), 1.86–1.44 (m, 6H); ESI-MS ( $m/z$ ): calcd for  $C_{21}H_{28}N_3O_5$  402.20, found 402.06  $[M + H]^+$ .

(*S*)-*N*-(5-(2,3-Dimethoxybenzamido)-6-hydroxyhexyl)nicotinamide (**8b**). General procedure **f** was followed with **7b** (280 mg, 0.63 mmol) and lithium borohydride (2 M in THF, 0.31 mL) to afford intermediate **8b** as a yellow oil (91 mg, 36%). No  $^1H$  NMR spectrum was recorded; ESI-MS ( $m/z$ ): calcd for  $C_{21}H_{28}N_3O_5$  402.20, found 402.08  $[M + H]^+$ .

(*S*)-*N*-(5-(2-Chlorobenzamido)-6-hydroxyhexyl)-2,3-dimethoxybenzamide (**8c**). General procedure **f** was followed with **7c** (161 mg, 0.34 mmol) and lithium borohydride (2 M in THF, 0.200 mL) to afford intermediate **8c** as a yellow oil (90.0 mg, 61%).  $^1H$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  8.18 (t,  $J = 5.7$  Hz, 1H), 8.10 (d,  $J = 8.5$  Hz, 1H), 7.47 (dd,  $J = 8.1, 1.3$  Hz, 1H), 7.47–7.35 (m, 3H), 7.18–7.05 (m, 3H), 5.75 (s, 1H), 4.69 (t,  $J = 5.7$  Hz, 1H), 3.82 (s, 3H), 3.75 (s, 3H), 3.46 (dt,  $J = 10.7, 5.4$  Hz, 2H), 3.27–3.20 (m, 2H), 1.59–1.32 (m, 6H); HRMS (ESI) ( $m/z$ ): calcd for  $C_{22}H_{28}ClN_2O_5$  435.1687, found 435.1668  $[M + H]^+$ .

(*S*)-2,3-Dichloro-*N*-(6-(2,3-dimethoxybenzamido)-1-hydroxyhexan-2-yl)benzamide (**8d**). General procedure **f** was followed with **7d** (51.0 mg, 0.10 mmol) and lithium borohydride (2 M in THF, 0.050 mL) to afford intermediate **8d** as a yellow oil (23.0 mg, 49%). Intermediate **8d** was taken to the next step without purification. No  $^1H$  NMR spectrum was recorded; ESI-MS ( $m/z$ ): calcd for  $C_{22}H_{27}Cl_2N_2O_5$  469.13, found 469.15  $[M + H]^+$ .

(*S*)-*N*-(5-(2-Aminobenzamido)-6-hydroxyhexyl)-2,3-dimethoxybenzamide (**8e**). General procedure **f** was followed with **7e** (128 mg, 0.28 mmol) and lithium borohydride (2 M in THF, 0.280 mL) to afford crude intermediate **7e** as a yellow oil (90.0 mg, 61%). Intermediate **8e** was taken to the next step without purification. No  $^1H$  NMR spectrum was recorded; ESI-MS ( $m/z$ ): calcd for  $C_{22}H_{30}N_3O_5$  416.21, found 416.16  $[M + H]^+$ .

(*S*)-2-Amino-3-chloro-*N*-(6-(2,3-dimethoxybenzamido)-1-hydroxyhexan-2-yl)benzamide (**8f**). General procedure **f** was followed with **7f** (125 mg, 0.25 mmol) and lithium borohydride (2 M in THF, 0.160 mL) to afford intermediate **8f** as a yellow oil (57.0 mg, 51%, purity >70%).  $^1H$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  8.17 (t,  $J = 5.6$  Hz, 1H), 7.97 (d,  $J = 8.5$  Hz, 1H), 7.50 (dd,  $J = 7.9, 1.5$  Hz, 1H), 7.37 (td,  $J = 7.6, 1.4$  Hz, 1H), 7.16–7.04



(m, 3H), 6.56 (t,  $J = 7.8$  Hz, 1H), 6.41 (d,  $J = 9.8$  Hz, 2H), 4.69 (t,  $J = 5.8$  Hz, 1H), 3.92 (td,  $J = 8.5, 8.0, 4.2$  Hz, 1H), 3.81 (s, 3H), 3.72 (s, 3H), 3.47–3.40 (m, 2H), 3.23 (q,  $J = 6.7$  Hz, 2H), 1.68–1.26 (m, 6H); ESI-MS ( $m/z$ ): calcd for  $C_{22}H_{29}ClN_3O_5$  450.18, found 450.13  $[M + H]^+$ .

(*S*)-*N*-(5-(2-Amino-6-fluorobenzamido)-6-hydroxyhexyl)-2,3-dimethoxybenzamide (**8g**). General procedure **f** was followed with **7g** (130 mg, 0.27 mmol) and lithium borohydride (2 M in THF, 0.140 mL) to afford **8g** as a yellow oil (84.0 mg, 72%).  $^1H$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  8.17 (t,  $J = 3.85$  Hz, 1H), 7.83 (dd,  $J = 8.7, 2.5$  Hz, 1H), 7.13 (dd,  $J = 8.1, 2.0$  Hz, 1H), 7.10 (d,  $J = 8.0$  Hz, 1H), 7.09 (d,  $J = 8.0$  Hz, 1H), 7.04 (dd,  $J = 8.3, 1.9$  Hz, 1H), 6.49 (dd,  $J = 8.2, 1.0$  Hz, 1H), 6.29 (ddt,  $J = 11.1, 8.1, 1.0$  Hz, 1H), 5.80 (s, 1H), 5.76 (s, 1H), 4.73 (t,  $J = 5.5$  Hz, 1H), 3.95–3.90 (m, 1H), 3.82 (s, 3H), 3.74 (s, 3H), 3.46–3.33 (m, 2H), 3.23–3.20 (m, 2H), 1.65–1.23 (m, 6H); HRMS (ESI) ( $m/z$ ): calcd for  $C_{22}H_{29}FN_3O_5$  434.2086, found 434.2086  $[M + H]^+$ .

(*S*)-*N*-(6-(2,3-Dihydroxybenzamido)-1-hydroxyhexan-2-yl)nicotinamide (**9a**). General procedure **g** was followed with **8a** (330 mg, 0.83 mmol) and boron tribromide (1 M in DCM, 3.32 mL) to afford intermediate **9a** as a light-yellow powder (85.0 mg, 27%).  $[\alpha]_D^{25} = -15.9$  ( $c = 0.1$ , MeOH);  $^1H$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  12.40 (s, 1H), 9.17 (s, 1H), 8.99 (d,  $J = 2.2$  Hz, 1H), 8.80 (t,  $J = 5.6$  Hz, 1H), 8.68 (dd,  $J = 4.8, 1.7$  Hz, 1H), 8.25 (d,  $J = 8.4$  Hz, 1H), 8.16 (dt,  $J = 7.9, 2.0$  Hz, 1H), 7.47 (dd,  $J = 8.0, 4.8$  Hz, 1H), 7.24 (dd,  $J = 8.1, 1.5$  Hz, 1H), 6.88 (dd,  $J = 7.8, 1.4$  Hz, 1H), 6.63 (t,  $J = 7.9$  Hz, 1H), 4.72 (s, 1H), 3.96 (dq,  $J = 9.3, 4.6$  Hz, 1H), 3.47–3.44 (m, 1H), 3.39 (dd,  $J = 10.8, 6.0$  Hz, 2H), 3.26 (s, 1H), 1.68–1.48 (m, 4H), 1.42–1.29 (m, 2H);  $^{13}C$  NMR (126 MHz, DMSO- $d_6$ ):  $\delta$  169.7, 164.8, 151.6, 149.9, 148.5, 146.3, 135.0, 130.3, 123.3, 118.6, 117.7, 117.1, 114.9, 63.3, 51.5, 39.2, 30.3, 28.9, 23.2; HRMS (ESI) ( $m/z$ ): calcd for  $C_{19}H_{24}N_3O_5$  374.1710, found 374.1711  $[M + H]^+$ .

(*S*)-*N*-(5-(2,3-Dihydroxybenzamido)-6-hydroxyhexyl)nicotinamide (**9b**). General procedure **g** was followed with **8b** (91.0 mg, 0.23 mmol) and boron tribromide (1 M in DCM, 0.920 mL) to afford intermediate **9b** as a light-yellow powder (12.0 mg, 14%).  $[\alpha]_D^{25} = -19.6$  ( $c = 0.1$ , MeOH);  $^1H$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  12.69 (s, 1H), 9.30 (s, 1H), 8.96 (d,  $J = 2.2$  Hz, 1H), 8.67 (dd,  $J = 4.9, 1.7$  Hz, 1H), 8.64 (t,  $J = 5.7$  Hz, 1H), 8.37 (d,  $J = 8.4$  Hz, 1H), 8.13 (dt,  $J = 8.0, 2.0$  Hz, 1H), 7.46 (dd,  $J = 7.9, 4.8$  Hz, 1H), 7.34 (dd,  $J = 8.1, 1.5$  Hz, 1H), 6.89 (dd,  $J = 7.8, 1.4$  Hz, 1H), 6.65 (t,  $J = 7.9$  Hz, 1H), 4.77 (s, 1H), 3.99 (td,  $J = 9.1, 4.6$  Hz, 1H), 3.51–3.46 (m, 1H), 3.44 (d,  $J = 7.0$  Hz, 2H), 3.25 (s, 1H), 1.66–1.48 (m, 4H), 1.41–1.29 (m, 2H);  $^{13}C$  NMR (126 MHz, DMSO- $d_6$ ):  $\delta$  169.5, 164.7, 151.7, 149.7, 148.3, 146.2, 134.9, 130.1, 123.4, 118.6, 117.6, 117.5, 115.2, 63.2, 51.0, 39.2, 30.2, 29.0, 23.2; HRMS (ESI) ( $m/z$ ): calcd for  $C_{19}H_{24}N_3O_5$  374.1710, found 374.1710  $[M + H]^+$ .

(*S*)-*N*-(5-(2-Chlorobenzamido)-6-hydroxyhexyl)-2,3-dihydroxybenzamide (**9c**). General procedure **g** was followed with **8c** (90.0 mg, 0.21 mmol) and boron tribromide (1 M in DCM, 0.840 mL) to afford **9c** as a light-yellow powder (4.0 mg, 5%) and partially demethylated **9h** as a light-yellow powder (16.0 mg, 18%).  $[\alpha]_D^{25} = -14.8$  ( $c = 0.1$ , MeOH);  $^1H$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  12.91 (s, 1H), 9.10 (s, 1H), 8.82 (s, 1H), 8.10

(d,  $J = 8.6$  Hz, 1H), 7.46 (dd,  $J = 8.0, 1.2$  Hz, 1H), 7.41 (dd,  $J = 7.0, 2.0$  Hz, 1H), 7.39–7.35 (m, 1H), 7.32 (td,  $J = 7.3, 1.3$  Hz, 1H), 7.27 (dd,  $J = 8.2, 1.5$  Hz, 1H), 6.89 (dd,  $J = 7.8, 1.5$  Hz, 1H), 6.65 (t,  $J = 8.0$  Hz, 1H), 4.69 (s, 1H), 3.87 (s, 1H), 3.45 (dd,  $J = 10.6, 5.3$  Hz, 1H), 3.28 (d,  $J = 6.1$  Hz, 3H), 1.68–1.46 (m, 4H), 1.37 (tt,  $J = 11.3, 6.1$  Hz, 2H);  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ ):  $\delta$  169.7, 166.2, 150.0, 146.3, 137.5, 130.5, 129.8, 129.5, 128.8, 127.0, 118.6, 117.7, 117.1, 114.9, 63.5, 51.2, 39.0, 30.3, 28.9, 23.1; HRMS (ESI) ( $m/z$ ): calcd for  $\text{C}_{20}\text{H}_{24}\text{ClN}_2\text{O}_5$ , 407.1368, found 407.1367 [M + H] $^+$ .

(*S*)-2,3-Dichloro-*N*-(6-(2,3-dihydroxybenzamido)-1-hydroxyhexan-2-yl)benzamide (**9d**). General procedure **g** was followed with **8d** (23.0 mg, 0.05 mmol) and boron tribromide (1 M in DCM, 0.200 mL) to afford **9d** as a light-yellow powder (4.0 mg, 18%).  $[\alpha]_D^{25} = -13.6$  ( $c = 0.1$ , MeOH);  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  12.92 (s, 1H), 9.15 (s, 1H), 8.81 (t,  $J = 5.7$  Hz, 1H), 8.25 (d,  $J = 8.6$  Hz, 1H), 7.66 (dd,  $J = 7.6, 1.9$  Hz, 1H), 7.38–7.30 (m, 2H), 7.28 (d,  $J = 8.0$  Hz, 1H), 6.90 (d,  $J = 7.7$  Hz, 1H), 6.67 (t,  $J = 7.9$  Hz, 1H), 4.76 (s, 1H), 3.90–3.83 (m, 1H), 3.45 (d,  $J = 5.3$  Hz, 1H), 3.29 (dq,  $J = 12.7, 6.7$  Hz, 3H), 1.69–1.43 (m, 4H), 1.36 (dq,  $J = 15.9, 8.5$  Hz, 2H);  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ ):  $\delta$  169.8, 165.6, 149.9, 146.3, 139.9, 132.0, 130.8, 128.4, 128.0, 127.3, 118.8, 117.9, 117.1, 114.9, 63.4, 51.3, 39.4, 30.3, 28.9, 23.2; HRMS (ESI) ( $m/z$ ): calcd for  $\text{C}_{20}\text{H}_{23}\text{Cl}_2\text{N}_2\text{O}_5$  441.0979, found 441.0981 [M + H] $^+$ .

(*S*)-*N*-(5-(2-Aminobenzamido)-6-hydroxyhexyl)-2,3-dihydroxybenzamide (**9e**). General procedure **g** was followed with **8e** (57.0 mg, 0.14 mmol) and boron tribromide (1 M in DCM, 0.560 mL) to afford **9e** as a light-yellow powder (14.0 mg, 26%).  $[\alpha]_D^{25} = -11.2$  ( $c = 0.1$ , MeOH);  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  12.84 (s, 1H), 9.09 (s, 1H), 8.81 (t,  $J = 5.8$  Hz, 1H), 7.77 (d,  $J = 8.4$  Hz, 1H), 7.45 (dd,  $J = 8.0, 1.5$  Hz, 1H), 7.25 (dd,  $J = 8.1, 1.5$  Hz, 1H), 7.11 (ddd,  $J = 8.4, 7.0, 1.5$  Hz, 1H), 6.88 (dd,  $J = 7.8, 1.4$  Hz, 1H), 6.65 (t,  $J = 8.2$  Hz, 2H), 6.51–6.45 (m, 1H), 6.27 (s, 2H), 4.75 (s, 1H), 3.93–3.88 (m, 1H), 3.41 (d,  $J = 5.5$  Hz, 1H), 3.35–3.32 (m, 1H), 3.25 (q,  $J = 6.7$  Hz, 2H), 1.65–1.43 (m, 4H), 1.40–1.25 (m, 2H);  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ ):  $\delta$  169.9, 169.0, 150.0, 149.5, 146.4, 128.6, 128.4, 118.9, 118.0, 117.2, 116.4, 115.6, 115.1, 114.8, 63.7, 50.9, 39.3, 30.5, 29.1, 23.4; HRMS (ESI) ( $m/z$ ): calcd for  $\text{C}_{20}\text{H}_{26}\text{N}_3\text{O}_5$  388.1867, found 388.1864 [M + H] $^+$ .

(*S*)-2-Amino-3-chloro-*N*-(6-(2,3-dihydroxybenzamido)-1-hydroxyhexan-2-yl)benzamide (**9f**). General procedure **g** was followed with **8f** (57.0 mg, 0.13 mmol) and boron tribromide (1 M in DCM, 0.520 mL) to afford **9f** as a light-yellow powder (11.0 mg, 21%).  $[\alpha]_D^{25} = -15.4$  ( $c = 0.1$ , MeOH);  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  12.91 (s, 1H), 9.13 (s, 1H), 8.79 (t,  $J = 5.7$  Hz, 1H), 8.00 (d,  $J = 8.5$  Hz, 1H), 7.49 (d,  $J = 8.1$  Hz, 1H), 7.36 (d,  $J = 7.9$  Hz, 1H), 7.26 (d,  $J = 7.1$  Hz, 1H), 6.89 (d,  $J = 7.9$  Hz, 1H), 6.66 (t,  $J = 7.9$  Hz, 1H), 6.56 (t,  $J = 7.8$  Hz, 1H), 6.39 (s, 2H), 4.73 (s, 1H), 3.92 (tq,  $J = 10.3, 5.5$  Hz, 1H), 3.43 (t,  $J = 7.9$  Hz, 2H), 3.26 (q,  $J = 6.7$  Hz, 2H), 1.66–1.44 (m, 4H), 1.41–1.27 (m, 2H);  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ ):  $\delta$  169.8, 168.1, 149.8, 146.3, 144.9, 131.5, 127.4, 118.8, 118.8, 117.9, 117.7, 117.1, 115.2, 114.9, 63.4, 51.1, 39.0, 30.8, 28.9, 23.3; HRMS (ESI) ( $m/z$ ): calcd for  $\text{C}_{20}\text{H}_{25}\text{N}_3\text{O}_5\text{Cl}$  422.1477, found 422.1477 [M + H] $^+$ .

(*S*)-*N*-(5-(2-Amino-6-fluorobenzamido)-6-hydroxyhexyl)-2,3-dihydroxybenzamide (**9g**). General procedure **g** was followed with **8g** (84.0 mg, 0.19 mmol) and boron tribromide (1 M in DCM, 0.760 mL) to afford **9g** as a light-yellow powder (15.0 mg, 19%) and partially demethylated **9i** as a light-yellow powder (37.0 mg, 46%, purity >92%).  $[\alpha]_D^{25} = -12.8$  ( $c = 0.1$ , MeOH);  $^1\text{H NMR}$  (500 MHz, DMSO- $d_6$ ):  $\delta$  12.85 (s, 1H), 9.05 (s, 1H), 8.80 (s, 1H), 7.85 (dd,  $J = 8.6, 2.4$  Hz, 1H), 7.26 (dd,  $J = 8.2, 1.5$  Hz, 1H), 7.07–7.03 (m, 1H), 6.89 (dd,  $J = 7.8, 1.5$  Hz, 1H), 6.65 (t,  $J = 8.0$  Hz, 1H), 6.49 (dd,  $J = 8.3, 1.0$  Hz, 1H), 6.29 (ddd,  $J = 10.4, 8.1, 1.0$  Hz, 1H), 5.78 (s, 2H), 4.73 (s, 1H), 3.92 (s, 1H), 3.42 (dd,  $J = 10.5, 5.7$  Hz, 2H), 3.26 (d,  $J = 7.3$  Hz, 2H), 1.66–1.48 (m, 4H), 1.41 (ddt,  $J = 12.9, 7.8, 4.0$  Hz, 2H);  $^{13}\text{C NMR}$  (126 MHz, DMSO- $d_6$ ):  $\delta$  169.7, 164.1, 160.2 (d,  $^1J_{CF} = 247.5$  Hz), 150.0, 149.2, 146.3, 130.7, 118.6, 117.7, 117.1, 114.9, 111.2, 108.3 (d,  $^2J_{CF} = 19.3$  Hz), 101.8 (d,  $^2J_{CF} = 23.2$  Hz), 63.3, 50.9, 40.4, 30.3, 28.9, 23.0; HRMS (ESI) ( $m/z$ ): calcd for  $\text{C}_{20}\text{H}_{25}\text{FN}_3\text{O}_5$  406.1773, found 406.1773  $[\text{M} + \text{H}]^+$ .

(*S*)-*N*-(5-(2-Chlorobenzamido)-6-hydroxyhexyl)-2-hydroxy-3-methoxybenzamide (**9h**). Synthesis see **9c**.  $[\alpha]_D^{25} = -19.6$  ( $c = 0.1$ , MeOH);  $^1\text{H NMR}$  (500 MHz, DMSO- $d_6$ ):  $\delta$  12.94 (s, 1H), 8.86 (t,  $J = 5.7$  Hz, 1H), 8.15 (d,  $J = 8.6$  Hz, 1H), 7.46–7.42 (m, 2H), 7.42–7.36 (m, 2H), 7.32 (td,  $J = 7.4, 1.3$  Hz, 1H), 7.08 (dd,  $J = 8.0, 1.4$  Hz, 1H), 6.80 (t,  $J = 8.1$  Hz, 1H), 4.75 (s, 1H), 3.93–3.87 (m, 1H), 3.78 (s, 3H), 3.48 (dd,  $J = 10.6, 5.3$  Hz, 1H), 3.32 (dq,  $J = 13.4, 6.7$  Hz, 3H), 1.70–1.46 (m, 4H), 1.43–1.33 (m, 2H);  $^{13}\text{C NMR}$  (126 MHz, DMSO- $d_6$ ):  $\delta$  169.6, 166.4, 151.2, 148.6, 137.6, 130.6, 129.9, 129.6, 128.9, 127.0, 118.6, 117.8, 115.3, 114.8, 63.6, 55.8, 48.7, 39.1, 30.4, 28.9, 23.2; HRMS (ESI) ( $m/z$ ): calcd for  $\text{C}_{21}\text{H}_{26}\text{ClN}_2\text{O}_5$  421.1524, found 421.1525  $[\text{M} + \text{H}]^+$ .

(*S*)-*N*-(5-(2-Amino-6-fluorobenzamido)-6-hydroxyhexyl)-2-hydroxy-3-methoxybenzamide (**9i**). Synthesis see **9g**.  $[\alpha]_D^{25} = -17.3$  ( $c = 0.1$ , MeOH);  $^1\text{H NMR}$  (500 MHz, DMSO- $d_6$ ):  $\delta$  12.92 (s, 1H), 8.83 (t,  $J = 5.7$  Hz, 1H), 7.88 (d,  $J = 8.4$  Hz, 1H), 7.41 (d,  $J = 8.2$  Hz, 1H), 7.07 (ddd,  $J = 14.3, 8.1, 1.5$  Hz, 2H), 6.79 (t,  $J = 8.1$  Hz, 1H), 6.49 (d,  $J = 8.3$  Hz, 1H), 6.32–6.25 (m, 1H), 5.79 (s, 2H), 4.77 (s, 1H), 3.93 (td,  $J = 9.0, 4.7$  Hz, 1H), 3.77 (s, 3H), 3.42 (t,  $J = 8.4$  Hz, 2H), 3.27 (q,  $J = 7.1$  Hz, 2H), 1.65–1.50 (m, 3H), 1.39 (qt,  $J = 16.3, 10.3$  Hz, 3H);  $^{13}\text{C NMR}$  (126 MHz, DMSO- $d_6$ ):  $\delta$  169.6, 164.1, 160.3 (d,  $^1J_{CF} = 240.2$  Hz), 151.1, 149.5, 148.4, 130.9 (d,  $^2J_{CF} = 11.2$  Hz), 118.5, 117.7, 115.3, 114.8, 111.2 (d,  $^3J_{CF} = 2.3$  Hz), 108.5 (d,  $^2J_{CF} = 19.3$  Hz), 101.9, 63.4, 55.8, 51.0, 39.1, 30.3, 28.9, 23.1; HRMS (ESI) ( $m/z$ ): calcd for  $\text{C}_{21}\text{H}_{27}\text{FN}_3\text{O}_5$  420.1929, found 420.1928  $[\text{M} + \text{H}]^+$ .

#### 4.6.7 Reaction Monitoring and Purification of Synthetic Compounds

Reactions were monitored by a liquid chromatography-mass spectrometry (LC-MS) system equipped with Dionex UltiMate 3000 pump, autosampler, column compartment, detector, and ESI quadrupole MS (MSQ Plus or ISQ EC) from Thermo Fisher Scientific, Dreieich, Germany. Purification of the final products, when necessary, was performed using preparative HPLC (Dionex UltiMate 3000 UHPLC+ focused, Thermo

Scientific) on a reversed-phase column (C18 column, 5  $\mu\text{m}$ , MACHEREY-NAGEL, Düren, Germany). The solvents used for the chromatography were water with 0.05% formic acid and acetonitrile with 0.05% formic acid. Unless indicated otherwise, reagents and substrates were purchased from commercial sources and used as received. Solvents not required to be dry were purchased as technical grade and used as received. Dry solvents were purchased from commercial sources in Sure/Seal™ bottles and used as received and stored under a dry inert gas ( $\text{N}_2$  or Ar). Inert atmosphere experiments were performed with standard Schlenk techniques with dried ( $\text{P}_2\text{O}_5$ ) nitrogen gas. All reported compounds were characterized by  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR and compared with literature data. All final products were fully characterized by  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR and HRMS techniques. The purity of the final products and intermediates was determined by LC-MS and found to be >95%, unless otherwise stated.

#### 4.6.8 Bioactivity

##### 4.6.8.1 Assessment of Antimicrobial Activities (MIC and Synergy)

All microorganisms used in this study were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ), the Coli Genetic Stock Center (CGSC), or were part of our internal collection, and were handled according to standard procedures. Single colonies of *S. aureus* and *E. coli* were inoculated from CASO agar plates into cation-adjusted Mueller Hinton Broth to obtain a final inoculum of  $10^5$  colony-forming units (CFU)/mL. Single colonies of *C. albicans* on Sabouraud agar plates and spores of *M. hiemalis* were used to prepare the inoculums in Myc2 medium (10 g/L Bacto-peptone, 10 g/L yeast extract, 0.2% glycerol, pH 6.3). The tested derivatives were prepared as DMSO stocks (10 mg/mL). Serial dilutions of derivatives in the respective growth medium (0.06 to 128  $\mu\text{g}/\text{mL}$ ) were prepared in sterile 96-well plates and the bacterial or fungal suspensions were added. Growth inhibition was assessed after incubation at 37 °C (*S. aureus*, *E. coli*) and 30 °C (*C. albicans*, *M. hiemalis*), respectively. Minimum inhibitory concentrations (MIC) are defined as the lowest compound concentration where no visible growth is observed. The checkerboard assay was used to evaluate synergism, antagonism, or indifference between the tested derivative **9a** and a number of antibiotics (daptomycin, gentamicin, linezolid, and ciprofloxacin). Briefly, 96-well microtiter plates were prepared: panel A was prepared for a serial dilution of the antibiotics and panel B was used for the serial dilution of the tested derivative. Then, 50  $\mu\text{L}$  was transferred from wells of panel A and dispensed in the corresponding wells of panel B. Bacterial suspension of *S. aureus* str. Newman were added to the plate to achieve approximately  $10^5$  CFU/mL. The plates were incubated at 37 °C for 24 h. The MICs of panel A and panel B alone as well as the combination wells were determined.

#### 4.6.8.2 Cytotoxic Activity (IC<sub>50</sub>)

HepG2 cells (human hepatoblastoma cell line; ACC 180, DSMZ) and HCT-116 cells (human colon carcinoma cell line; ACC 581, DSMZ) were cultured under conditions recommended by the depositor and cells were propagated in Dulbecco's modified Eagle medium (DMEM) and McCoy's 5A medium, respectively, supplemented with 10% fetal bovine serum (FBS). For determining the antiproliferative activity of test compounds, cells were seeded at  $6 \times 10^3$  cells per well of 96-well plates in 120  $\mu$ L complete medium. After 2 h of equilibration, compounds were added in serial dilution in 60  $\mu$ L complete medium. Compounds, as well as the solvent control and doxorubicin as reference, were tested as duplicates in two independent experiments. After 5 d incubation, 20  $\mu$ L of 5 mg/mL MTT (thiazolyl blue tetrazolium bromide) in PBS was added per well and cells were further incubated for 2 h at 37 °C. The medium was then discarded and cells were washed with 100  $\mu$ L PBS before adding 100  $\mu$ L 2-propanol/10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader (Tecan Infinite M200Pro), and cell viability was expressed as percentage relative to the respective solvent control. IC<sub>50</sub> were determined by sigmoidal curve fitting using GraphPad PRISM 8 (GraphPad Software, San Diego, CA, USA).

## 4.7 Acknowledgments

The authors thank Alexandra Amann, Tanja Küssau, and Stefanie Schmidt for performing bioactivity assays, and Eric Bacqué and Laurent Goullieux for fruitful discussions.

## 4.8 References

1. Muñoz-Dorado, J.; Marcos-Torres, F.J.; García-Bravo, E.; Moraleda-Muñoz, A.; Pérez, J. Myxobacteria: moving, killing, feeding, and surviving together. *Front. Microbiol.* **2016**, *7*, 781, doi:10.3389/fmicb.2016.00781.
2. Cao, P.; Dey, A.; Vassallo, C.N.; Wall, D. How Myxobacteria Cooperate. *J. Mol. Biol.* **2015**, doi:10.1016/j.jmb.2015.07.022.
3. Herrmann, J.; Fayad, A.A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34*, 135–160, doi:10.1039/C6NP00106H.
4. Baumann, S.; Herrmann, J.; Raju, R.; Steinmetz, H.; Mohr, K.I.; Hüttel, S.; Harmrolfs, K.; Stadler, M.; Müller, R. Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent

- antibacterial activity. *Angew. Chem. Int. Ed.* **2014**, *53*, 14605–14609, doi:10.1002/anie.201409964.
5. Gerth, K.; Bedorf, N.; Irschik, H.; Höfle, G.; Reichenbach, H. The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen A<sub>1α</sub>: fermentation, isolation, biological properties. *J. Antibiot.* **1994**, *47*, 23–31, doi:10.7164/antibiotics.47.23.
  6. Held, J.; Gebru, T.; Kalesse, M.; Jansen, R.; Gerth, K.; Müller, R.; Mordmüller, B. Antimalarial activity of the myxobacterial macrolide chlorotonil A. *Antimicrob. Agents Chemother.* **2014**, *58*, 6378–6384, doi:10.1128/AAC.03326-14.
  7. Höfle, G.; Bedorf, N.; Steinmetz, H.; Schomburg, D.; Gerth, K.; Reichenbach, H. Epothilone A and B—Novel 16-Membered macrolides with cytotoxic activity: Isolation, crystal structure, and conformation in solution. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1567–1569, doi:10.1002/anie.199615671.
  8. Weissman, K.J.; Müller, R. A brief tour of myxobacterial secondary metabolism. *Bioorg. Med. Chem.* **2009**, *17*, 2121–2136, doi:10.1016/j.bmc.2008.11.025.
  9. Kunze, B.; Bedorf, N.; Kohl, W.; Höfle, G.; Reichenbach, H. Myxochelin A, a new iron-chelating compound from *Angiococcus disciformis* (Myxobacteriales). Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **1989**, *42*, 14–17, doi:10.7164/antibiotics.42.14.
  10. Gaitatzis, N.; Kunze, B.; Müller, R. In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11136–11141, doi:10.1073/pnas.201167098.
  11. Li, Y.; Weissman, K.J.; Müller, R. Myxochelin biosynthesis: Direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. *J. Am. Chem. Soc.* **2008**, *130*, 7554–7555, doi:10.1021/ja8025278.
  12. Korp, J.; König, S.; Schieferdecker, S.; Dahse, H.-M.; König, G.M.; Werz, O.; Nett, M. Harnessing Enzymatic Promiscuity in Myxochelin Biosynthesis for the Production of 5-Lipoxygenase Inhibitors. *ChemBioChem* **2015**, *16*, 2445–2450, doi:10.1002/cbic.201500446.
  13. Sester, A.; Winand, L.; Pace, S.; Hiller, W.; Werz, O.; Nett, M. Myxochelin- and Pseudochelin-Derived Lipoxygenase Inhibitors from a Genetically Engineered *Myxococcus xanthus* Strain. *J. Nat. Prod.* **2019**, *82*, 2544–2549, doi:10.1021/acs.jnatprod.9b00403.
  14. Gaitatzis, N.; Kunze, B.; Müller, R. Novel insights into siderophore formation in myxobacteria. *ChemBioChem* **2005**, *6*, 365–374, doi:10.1002/cbic.200400206.

15. Panter, F.; Popoff, A.; Garcia, R.; Krug, D.; Müller, R. Myxobacteria of the Cystobacterineae Suborder Are Producers of New Vitamin K2 Derived Myxoquinones. *Microorganisms* **2022**, *10*, 534, doi: 10.3390/microorganisms10030534.
16. Miyanaga, S.; Obata, T.; Onaka, H.; Fujita, T.; Saito, N.; Sakurai, H.; Saiki, I.; Furumai, T.; Igarashi, Y. Absolute configuration and antitumor activity of myxochelin A produced by *Nonomurea pusilla* TP-A0861. *J. Antibiot.* **2006**, *59*, 698–703, doi:10.1038/ja.2006.93.
17. Miyanaga, S.; Sakurai, H.; Saiki, I.; Onaka, H.; Igarashi, Y. Synthesis and evaluation of myxochelin analogues as antimetastatic agents. *Bioorg. Med. Chem.* **2009**, *17*, 2724–2732, doi:10.1016/j.bmc.2009.02.040.
18. Schieferdecker, S.; Nett, M. A fast and efficient method for the preparation of the 5-lipoxygenase inhibitor myxochelin A. *Tetrahedron Lett.* **2016**, *57*, 1359–1360, doi:10.1016/j.tetlet.2016.02.047.
19. Schieferdecker, S.; König, S.; Pace, S.; Werz, O.; Nett, M. Myxochelin-Inspired 5-Lipoxygenase Inhibitors: Synthesis and Biological Evaluation. *ChemMedChem* **2017**, *12*, 23–27, doi:10.1002/cmdc.201600536.
20. Li, J.; Sha, Y. A convenient synthesis of amino acid methyl esters. *Molecules* **2008**, *13*, 1111–1119, doi:10.3390/molecules13051111.
21. Fields, G.B.; Noble, R.L. Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int. J. Pept. Protein Res.* **1990**, *35*, 161-214, doi:10.1111/j.1399-3011.1990.tb00939.x.
22. Schieferdecker, S.; König, S.; Koeberle, A.; Dahse, H.-M.; Werz, O.; Nett, M. Myxochelins target human 5-lipoxygenase. *J. Nat. Prod.* **2015**, *78*, 335–338, doi:10.1021/np500909b.
23. Ahn, J.-W.; Lee, C.-O.; Baek, S.-H. Myxochelin A, a cytotoxic antibiotic from the myxobacterium *Angiococcus Disciformis*. *Orient. Pharm. Exp. Med.* **2002**, *2*, 64–67, doi:10.3742/OPEM.2002.2.1.064.
24. Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **2018**, *9*, 803, doi:10.1038/s41467-018-03184-1.
25. Röttig, M.; Medema, M.H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **2011**, *39*, W362–W367, doi:10.1093/nar/gkr323.
26. Schmidt, M.T.; Smith, B.C.; Jackson, M.D.; Denu, J.M. Coenzyme specificity of Sir2 protein deacetylases: implications for physiological regulation. *J. Biol. Chem.* **2004**, *279*, 40122–40129, doi:10.1074/jbc.M407484200.

27. Okazaki, I.J.; Moss, J. Mono-ADP-ribosylation: A Reversible Posttranslational Modification of Proteins. *Adv Pharmacol.* **1996**, *35*, 247–280, doi:10.1016/s1054-3589(08)60277-x.
28. McErlean, M.; Overbay, J.; van Lanen, S. Refining and expanding nonribosomal peptide synthetase function and mechanism. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 493–513, doi:10.1007/s10295-018-02130-w.
29. Gholson, R.K. The Pyridine Nucleotide Cycle. *Nature* **1966**, *212*, 933–935, doi:10.1038/212933a0.
30. Rodionov, D.A.; de Ingeniis, J.; Mancini, C.; Cimadamore, F.; Zhang, H.; Osterman, A.L.; Raffaelli, N. Transcriptional regulation of NAD metabolism in bacteria: NrtR family of Nudix-related regulators. *Nucleic Acids Res.* **2008**, *36*, 2047–2059, doi:10.1093/nar/gkn047.
31. Denu, J.M. Linking chromatin function with metabolic networks: Sir2 family of NAD<sup>+</sup>-dependent deacetylases. *Trends Biochem. Sci.* **2003**, *28*, 41–48, doi:10.1016/S0968-0004(02)00005-1.
32. Subko, K.; Wang, X.; Nielsen, F.H.; Isbrandt, T.; Gotfredsen, C.H.; Ramos, C.; Mackenzie, T.; Vicente, F.; Genilloud, O.; Frisvad, J.C.; et al. Mass Spectrometry guided discovery and design of novel Asperphenamate analogues from *Penicillium astrolabium* reveals an extraordinary NRPS flexibility. *Front. Microbiol.* **2021**, *11*, 3510, doi:10.3389/fmicb.2020.618730.
33. Itoh, T.; Tokunaga, K.; Matsuda, Y.; Fujii, I.; Abe, I.; Ebizuka, Y.; Kushiro, T. Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases. *Nat. Chem.* **2010**, *2*, 858–864, doi:10.1038/NCHEM.764.
34. Cutignano, A.; Tramice, A.; de Caro, S.; Villani, G.; Cimino, G.; Fontana, A. Biogenesis of 3-alkylpyridine alkaloids in the marine mollusc *Haminoea orbignyana*. *Angew. Chem. Int. Ed.* **2003**, *42*, 2633–2636, doi:10.1002/anie.200250642.
35. Ma, H.M.; Zhou, Q.; Tang, Y.M.; Zhang, Z.; Chen, Y.S.; He, H.Y.; Pan, H.X.; Tang, M.C.; Gao, J.F.; Zhao, S.Y.; et al. Unconventional origin and hybrid system for construction of pyrrolopyrrole moiety in kosinostatin biosynthesis. *Chem. Biol.* **2013**, *20*, 796–805, doi:10.1016/j.chembiol.2013.04.013.
36. Wu, Q.; Throckmorton, K.; Maity, M.; Chevrette, M.G.; Braun, D.R.; Rajski, S.R.; Currie, C.R.; Thomas, M.G.; Bugni, T.S. Bacillibactins E and F from a Marine Sponge-Associated *Bacillus* sp. *J. Nat. Prod.* **2020**, doi:10.1021/acs.jnatprod.0c01170.
37. Miethke, M.; Klotz, O.; Linne, U.; May, J.J.; Beckering, C.L.; Marahiel, M.A. Ferri-bacillibactin uptake and hydrolysis in *Bacillus subtilis*. *Mol. Microbiol.* **2006**, *61*, 1413–1427, doi:10.1111/j.1365-2958.2006.05321.x.
38. Jaehme, M.; Slotboom, D.J. Diversity of membrane transport proteins for vitamins in bacteria and archaea. *Biochim. Biophys. Acta* **2015**, *1850*, 565–576, doi:10.1016/j.bbagen.2014.05.006.



39. Bergeron, R.J.; Gavanaugh, P.F.; Kline, S.J.; Hughes, R.G.; Elliott, G.T.; Porter, C.W. Antineoplastic and antiherpetic activity of spermidine catecholamide iron chelators. *Biochem. Biophys. Res. Commun.* **1984**, *121*, 848–854, doi:10.1016/0006-291X(84)90755-1.

## Supporting Information

# Expanding the Myxochelin Natural Product Family by Nicotinic Acid Containing Congeners

Previously published in:

Nicolas A. Frank, Márió Széles, Sergi H. Akone, Sari Rasheed, Stephan Hüttel, Simon Frewert, Mostafa M. Hamed, Jennifer Herrmann, Sören M. M. Schuler, Anna K. H. Hirsch, Rolf Müller\*

*Molecules*, 2021 August 14; **26**(16):4929

DOI: 10.3390/molecules26164929

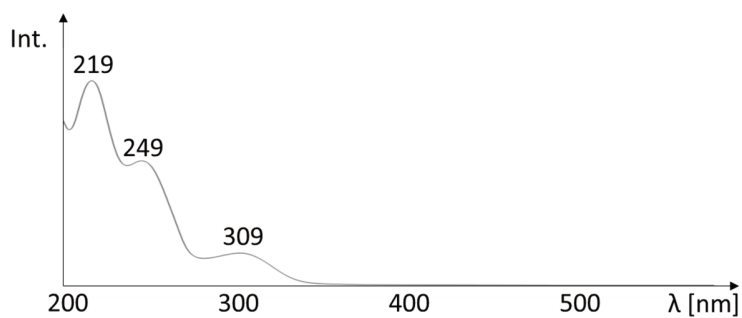
### Affiliation

Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

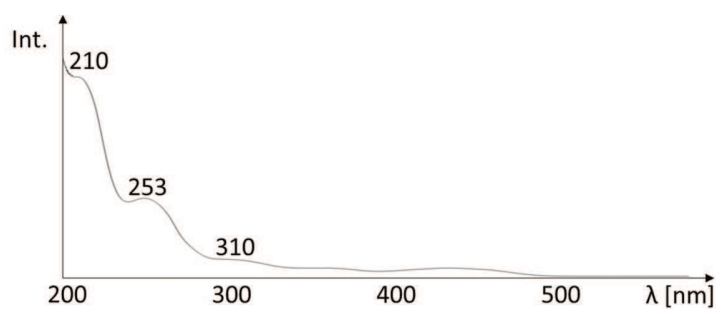
## S 4.1 Characterization of Myxochelin Derivatives

### S 4.1.1 UV Spectra of the Myxochelins

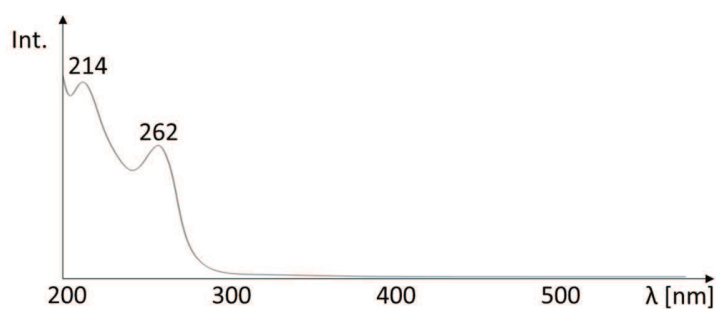
#### S 4.1.1.1 UV Spectra of the Naturally Occurring Myxochelins



**Figure S1.** UV spectrum of **1** (natural product) in water/acetonitrile mixture with 0.1% formic acid.



**Figure S2.** UV spectrum of **2** (natural product) in water/acetonitrile mixture with 0.1% formic acid.



**Figure S3.** UV spectrum of **3** (natural product) in water/acetonitrile mixture with 0.1% formic acid.

S 4.1.1.2 UV Spectra of Synthetic Myxochelins

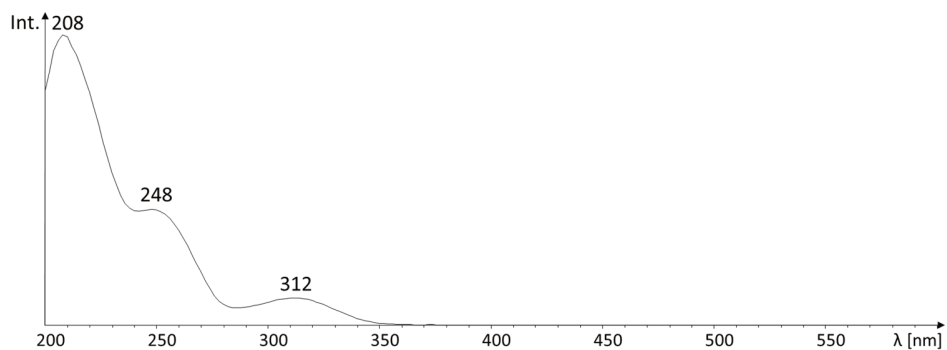


Figure S4. UV spectrum of **9a = 1** (synthetic compound) in water/acetonitrile mixture with 0.1% formic acid.

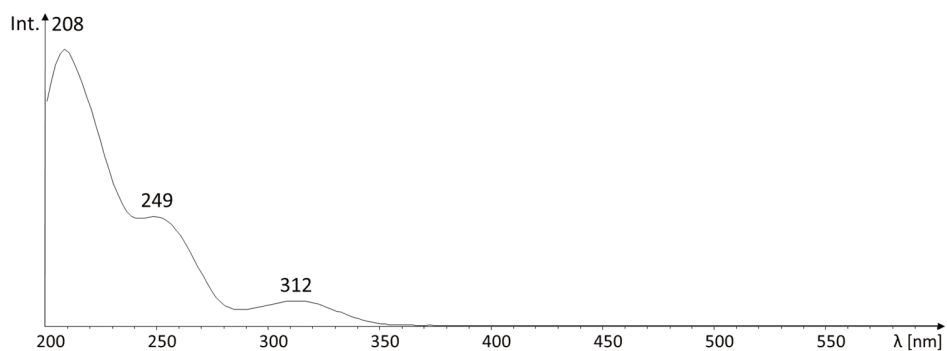


Figure S5. UV spectrum of **9b = 2** (synthetic compound) in water/acetonitrile mixture with 0.1% formic acid.

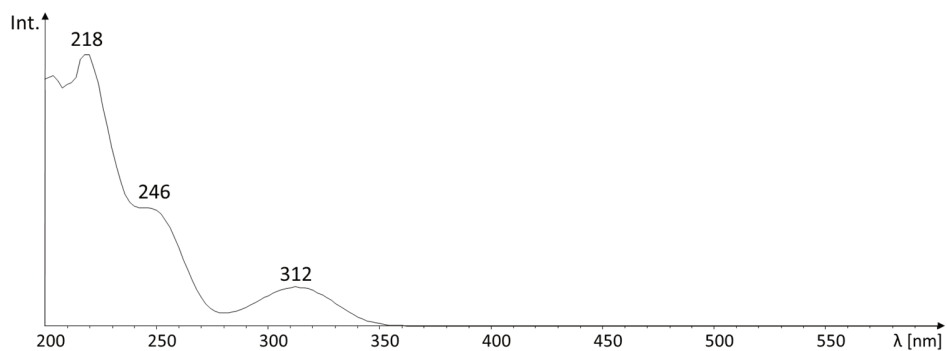
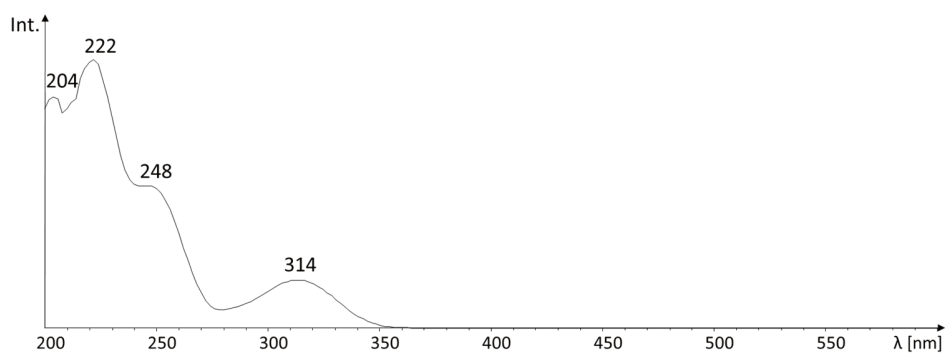
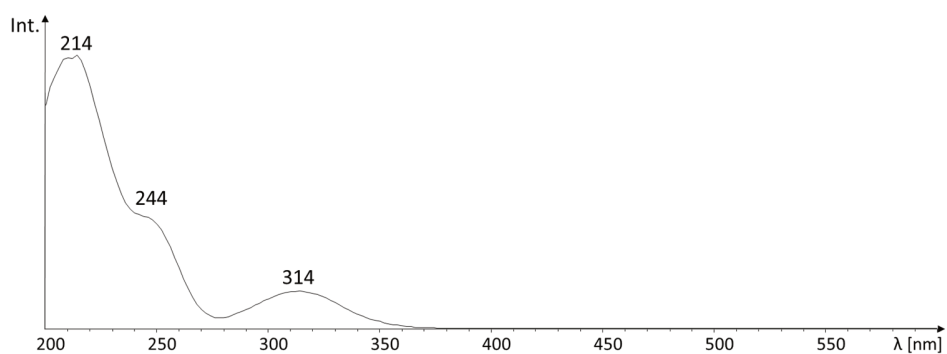


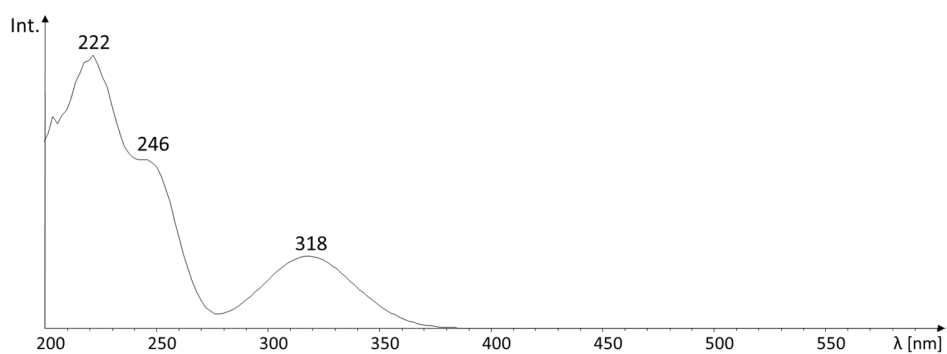
Figure S6. UV spectrum of **9c** in water/acetonitrile mixture with 0.1% formic acid.



**Figure S7.** UV spectrum of **9d** in water/acetonitrile mixture with 0.1% formic acid.



**Figure S8.** UV spectrum of **9e** in water/acetonitrile mixture with 0.1% formic acid.



**Figure S9.** UV spectrum of **9f** in water/acetonitrile mixture with 0.1% formic acid.

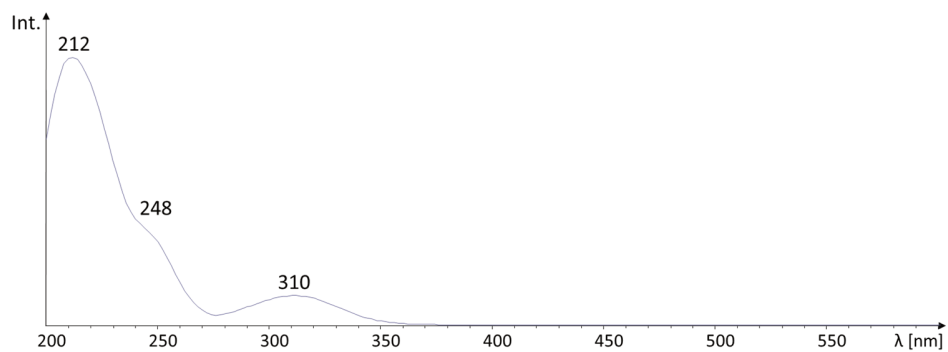


Figure S10. UV spectrum of **9g** in water/acetonitrile mixture with 0.1% formic acid.

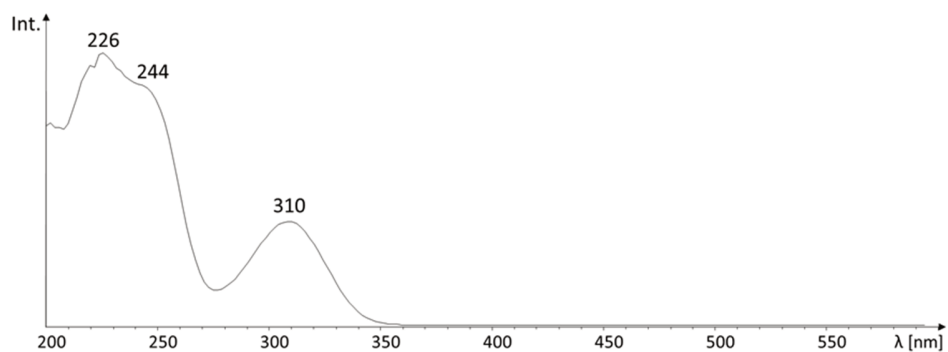


Figure S11. UV spectrum of **9h** in water/acetonitrile mixture with 0.1% formic acid.

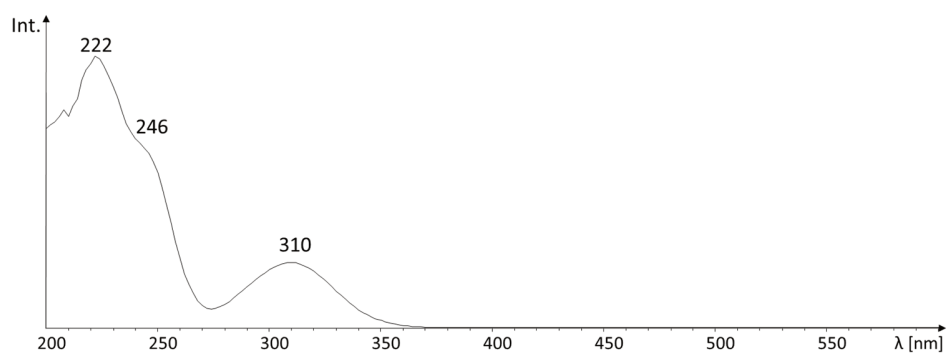


Figure S12. UV spectrum of **9i** in water/acetonitrile mixture with 0.1% formic acid.

## S 4.1.2 Tandem MS Spectra

## S 4.1.2.1 Tandem MS spectra of naturally occurring myxochelins

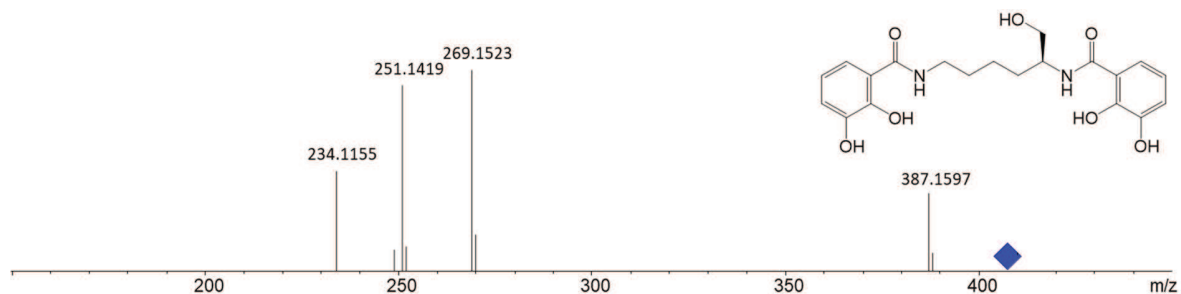


Figure S13. Tandem MS spectrum of myxochelin A.

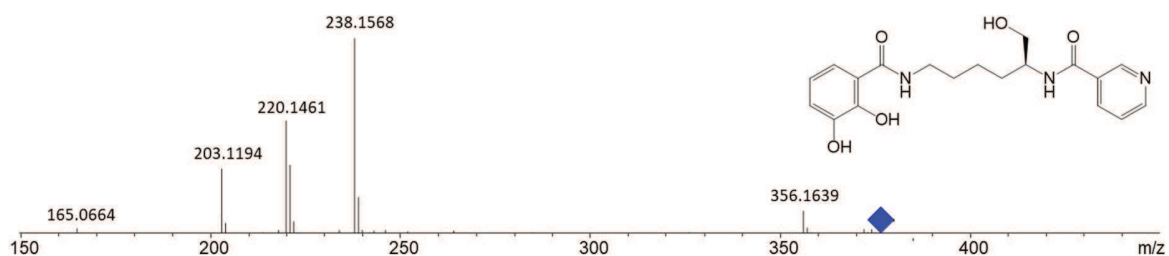


Figure S14. Tandem MS spectrum of 1.

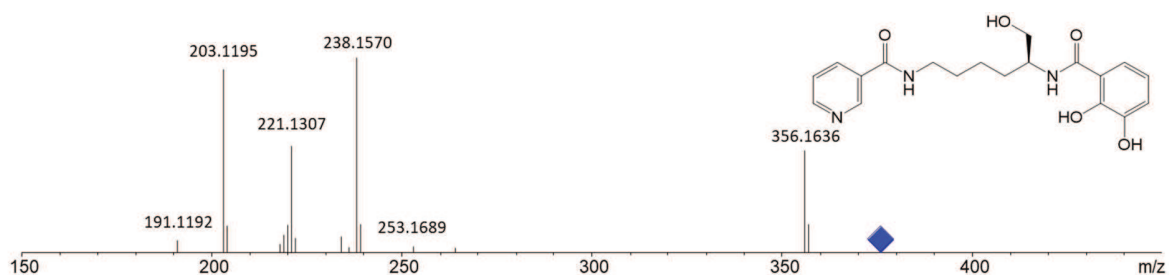


Figure S15. Tandem MS spectrum of 2.

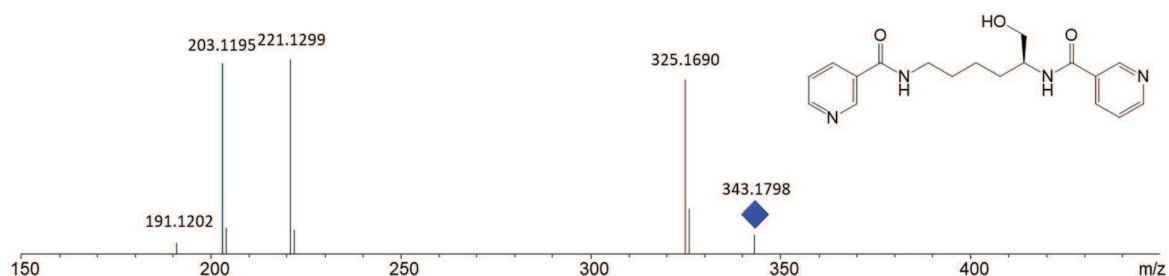


Figure S16. Tandem MS spectrum of 3.

S 4.1.2.2 Tandem MS spectra of synthetic myxochelins

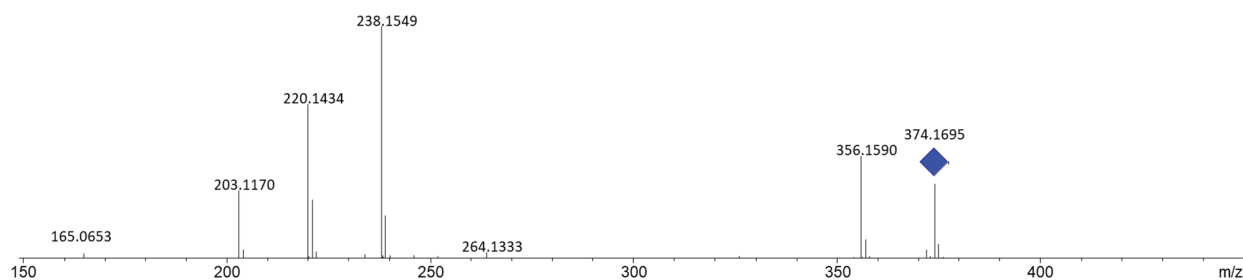


Figure S17. Tandem MS spectrum of **9a = 1**.

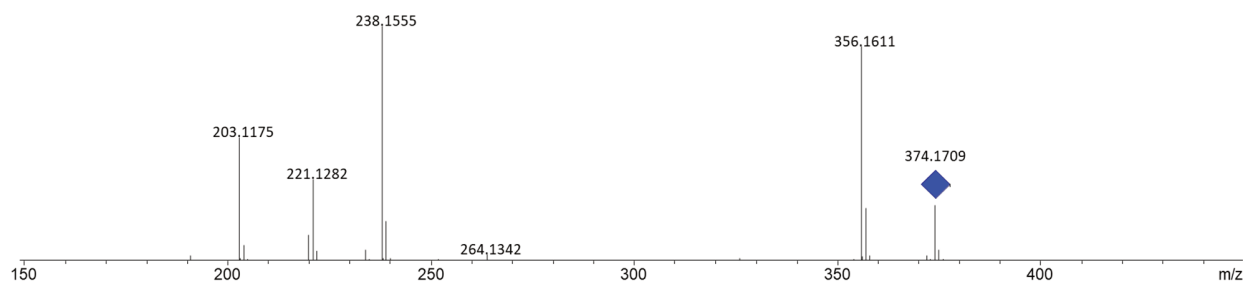


Figure S18. Tandem MS spectrum of **9b = 2**.

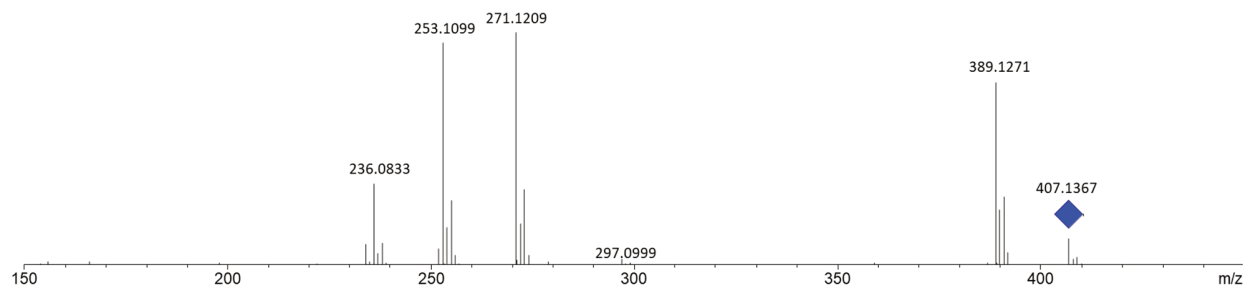


Figure S19. Tandem MS spectrum of **9c**.

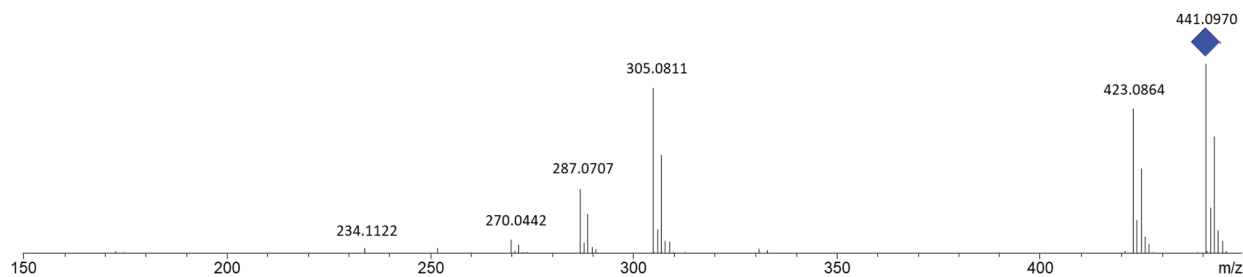


Figure S20. Tandem MS spectrum of **9d**.



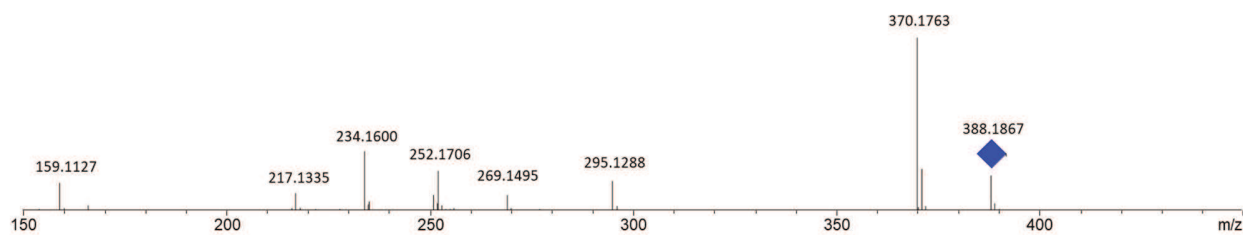


Figure S21. Tandem MS spectrum of 9e.

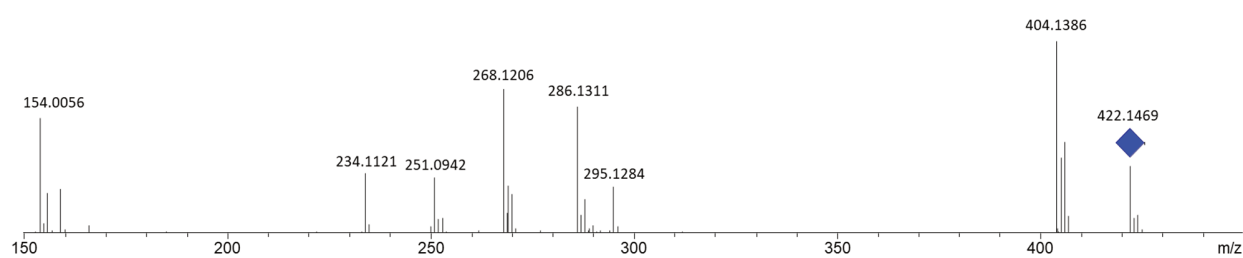


Figure S22. Tandem MS spectrum of 9f.

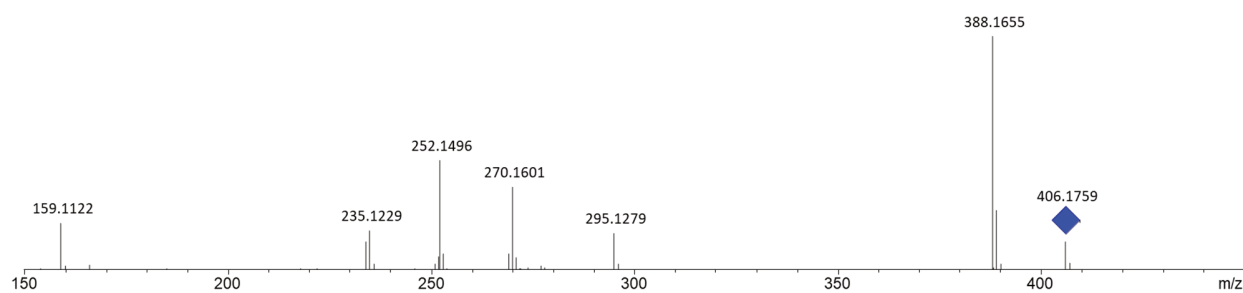


Figure S23. Tandem MS spectrum of 9g.

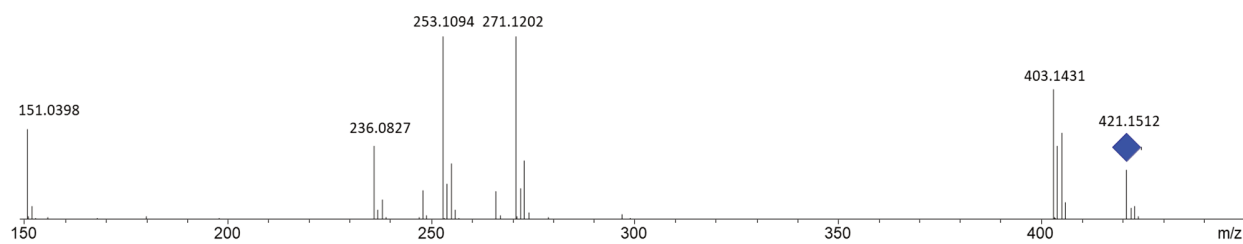


Figure S24. Tandem MS spectrum of 9h.

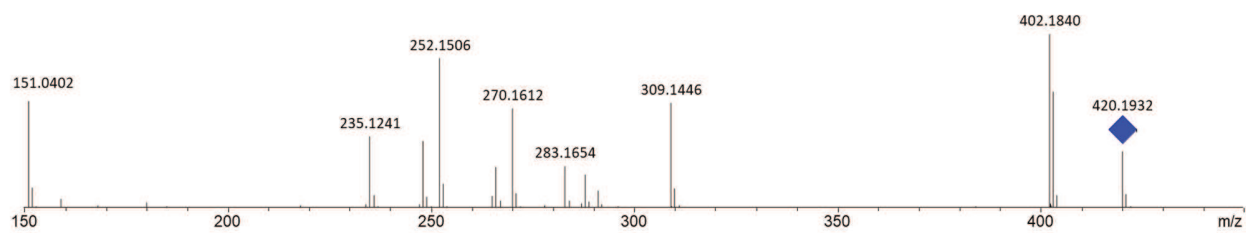


Figure S25. Tandem MS spectrum of **9i**.

### S 4.1.3 Circular Dichroism Spectra

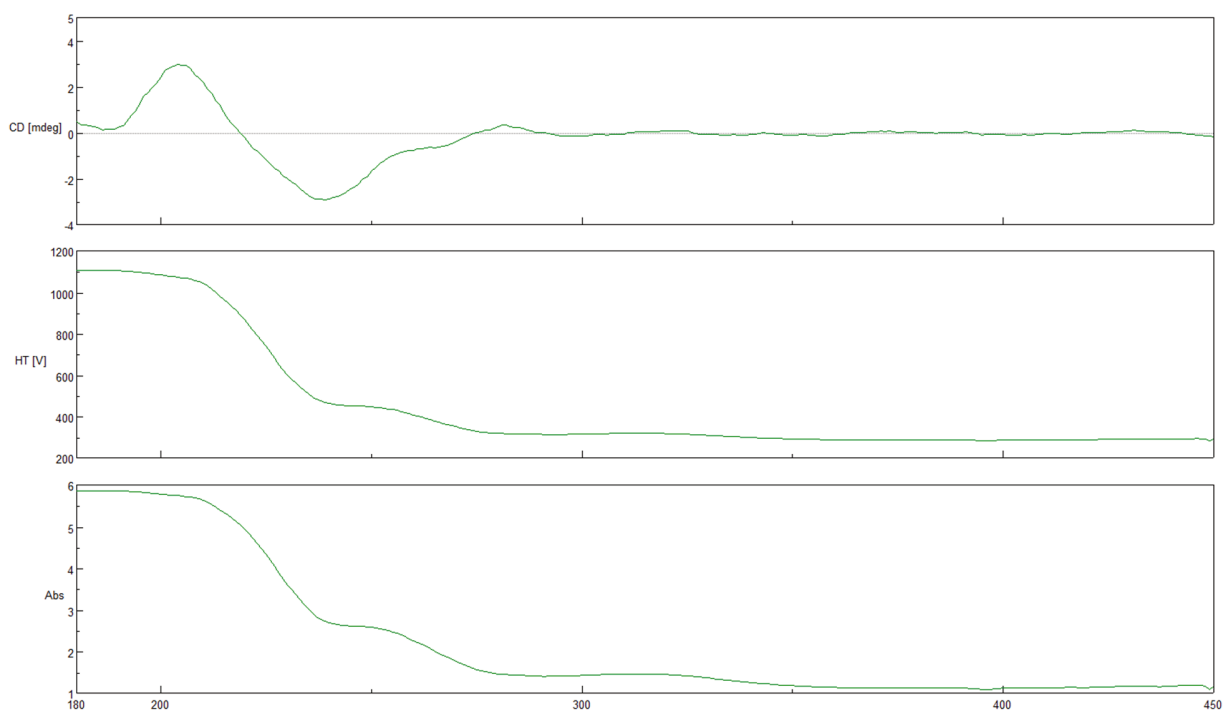
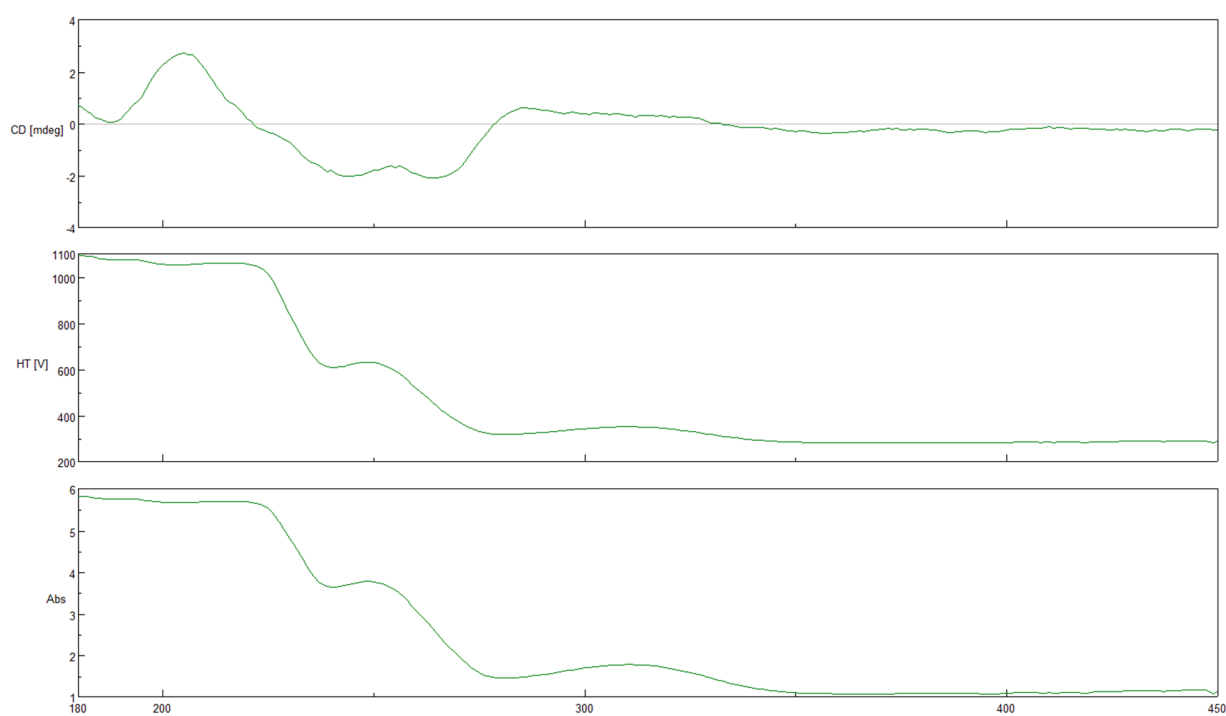
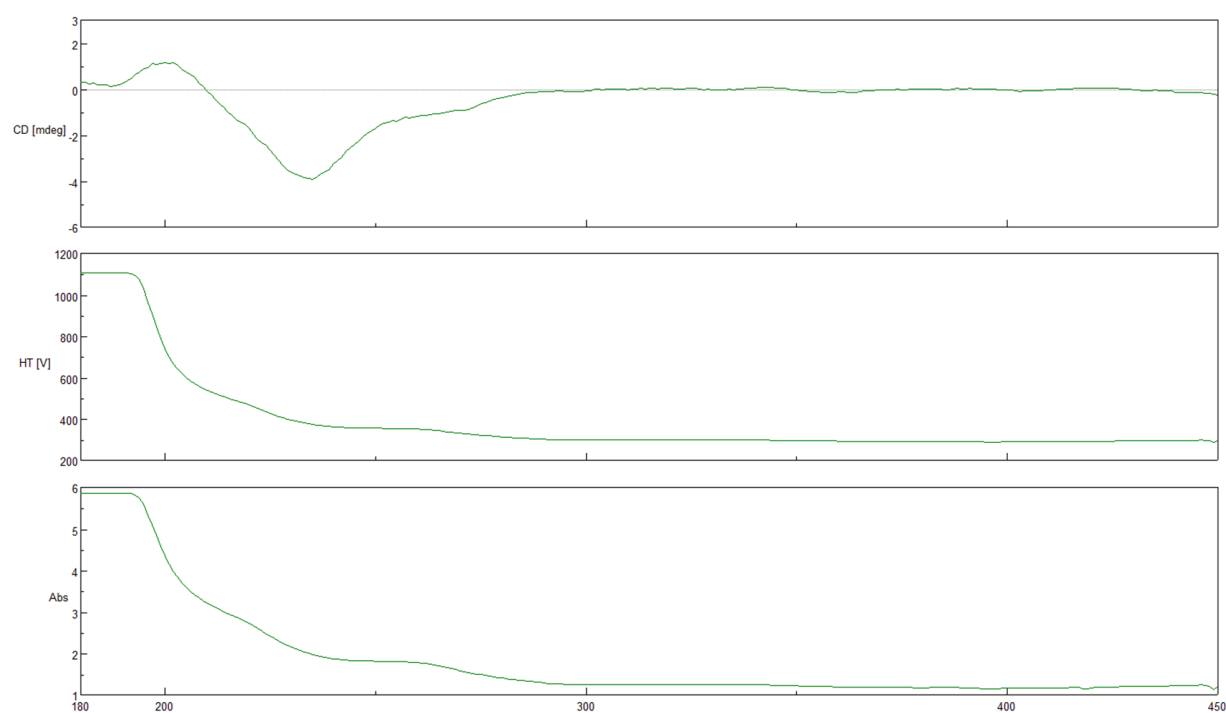


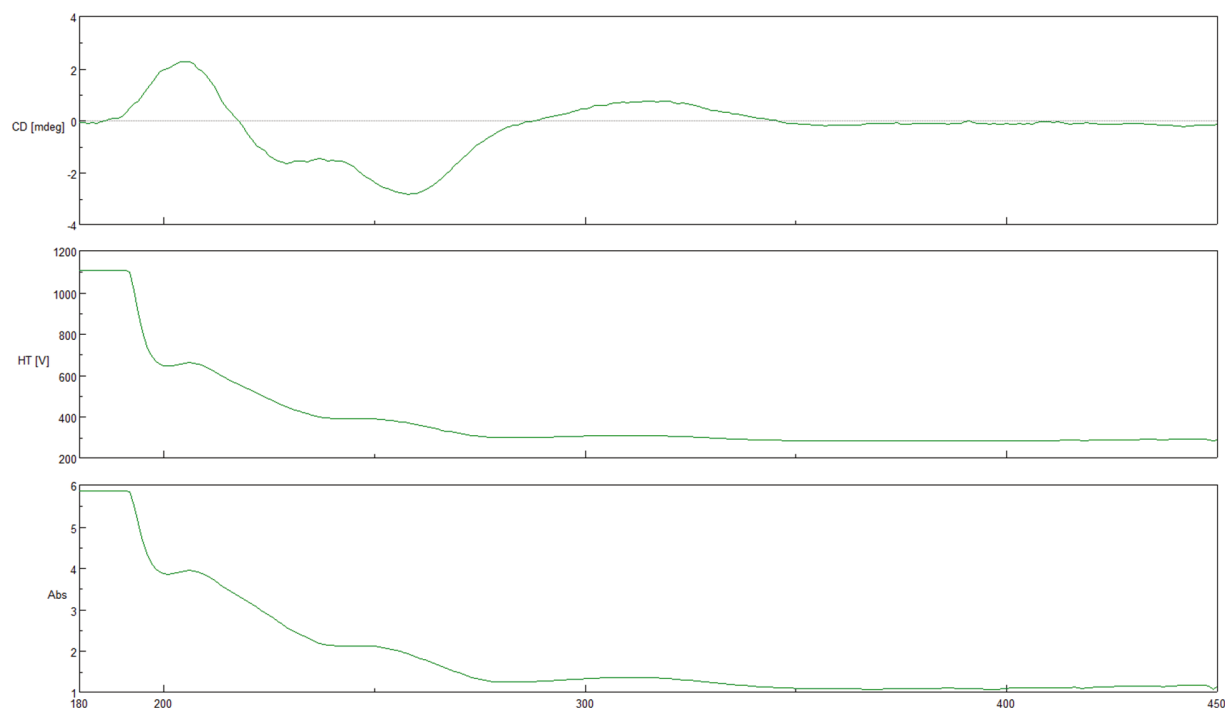
Figure S26. CD spectrum of 1 mg/mL **1** (isolated) in methanol in the area 180–450 nm.



**Figure S27.** CD spectrum of 1mg/mL **9a = 1** (synthetic) in methanol in the area 180–450 nm.

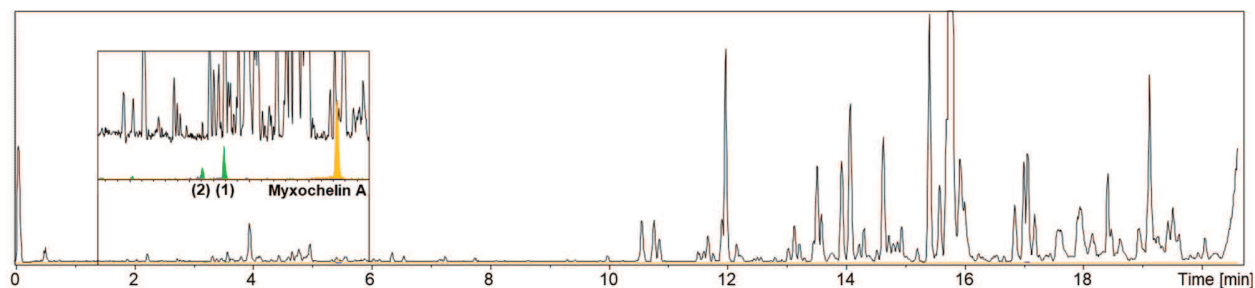


**Figure S28.** CD spectrum of 1mg/mL **2** (isolated) in methanol in the area 180–450 nm.

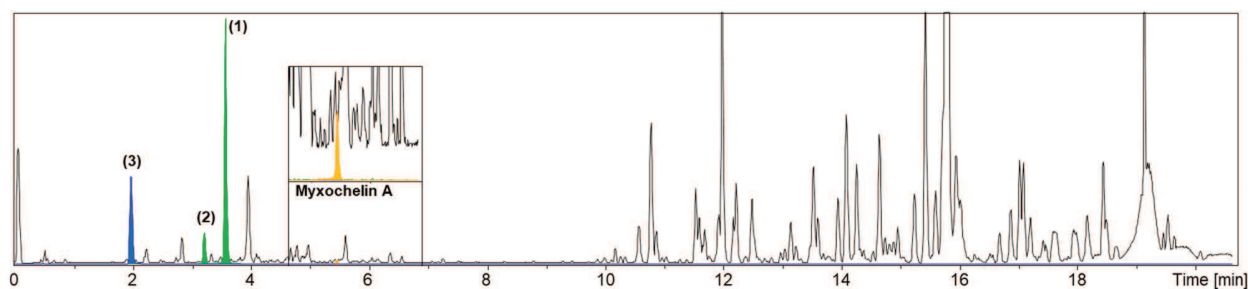


**Figure S29.** CD spectrum of 1mg/mL **9b = 2** (synthetic) in methanol in the area 180–450 nm.

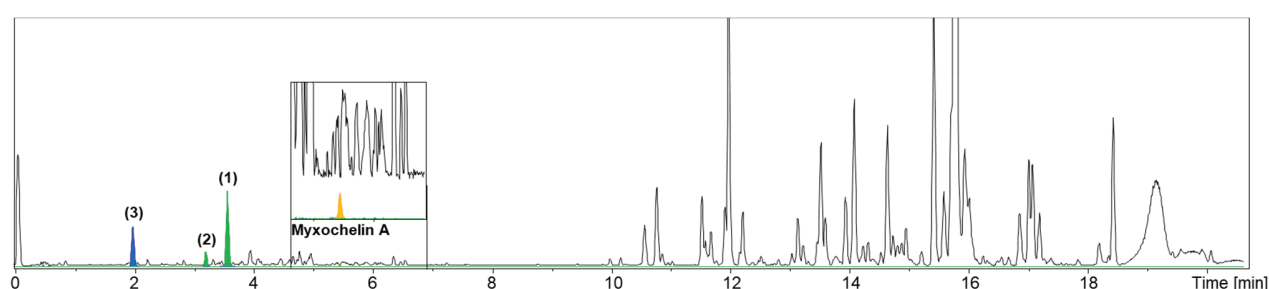
#### S 4.1.4 HPLC-MS Chromatograms of Crude Extracts



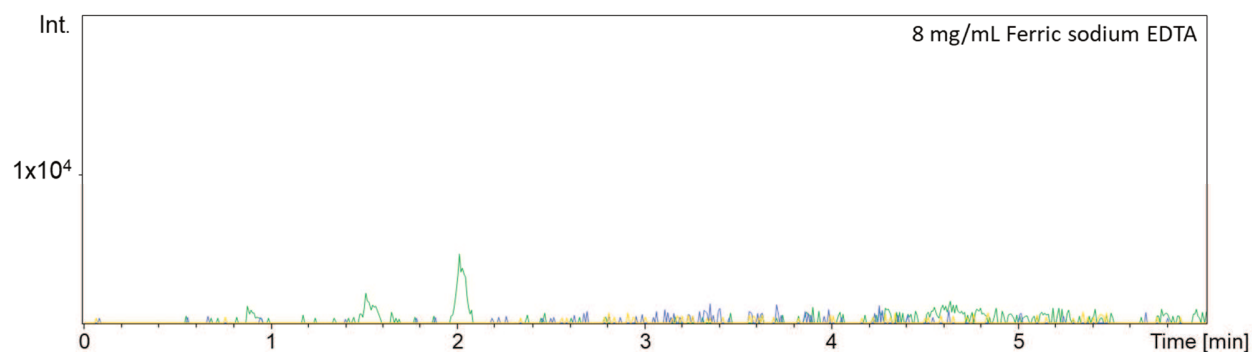
**Figure S30.** HPLC-MS base peak chromatogram (BPC) of *Corallocooccus* sp. MCy9049 in VY/2S medium with highlighted extracted ion chromatograms (EICs) of 374.17105 m/z (green: **1**, **2**), 343.17647 m/z (blue, **3**) and 405.16565 (orange: myxochelin A), with a width of 0.005 m/z.



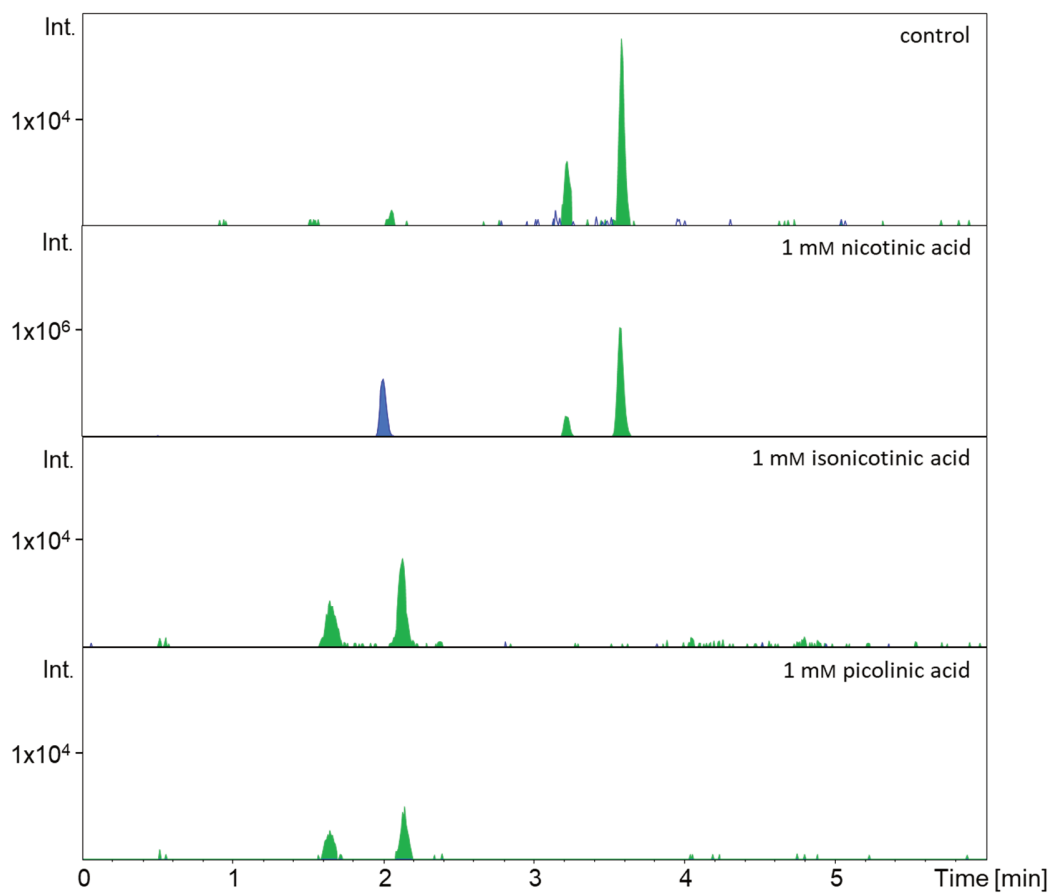
**Figure S31.** HPLC-MS base peak chromatogram (BPC) of *Corallocooccus* sp. MCy9049 in VY/2S medium with supplementation of 1 mM nicotinamide with highlighted extracted ion chromatograms (EICs) of 374.17105 m/z (green: **1**, **2**), 343.17647 m/z (blue, **3**) and 405.16565 (orange: myxochelin A), with a width of 0.005 m/z.



**Figure S32.** HPLC-MS base peak chromatogram (BPC) of *Corallocooccus* sp. MCy9049 in VY/2S medium with supplementation of 1 mM nicotinic acid with highlighted extracted ion chromatograms (EICs) of 374.17105 m/z (green: **1**, **2**), 343.17647 m/z (blue, **3**) and 405.16565 (orange: myxochelin A), with a width of 0.005 m/z.



**Figure S33.** HPLC-MS extracted ion chromatograms (EICs) of 374.17105 m/z (green: **1**, **2**), 343.17647 m/z (blue, **3**) and 405.16565 (orange: myxochelin A), with a width of 0.005 m/z from crude extracts of *Corallocooccus* sp. MCy9049 in VY/2S medium with supplementation of ferric sodium EDTA.



**Figure S34.** HPLC-MS extracted ion chromatograms (EICs) of 374.17105 m/z (green: **1**, **2**) and 343.17647 m/z (blue, **3**), with a width of 0.005 m/z from crude extracts of *Coralloccoccus* sp. MCy9049 in VY/2S medium with supplementation of different pyridinecarboxylic acids.

## S 4.1.5 HPLC-MS Chromatograms of the Synthetic Compounds

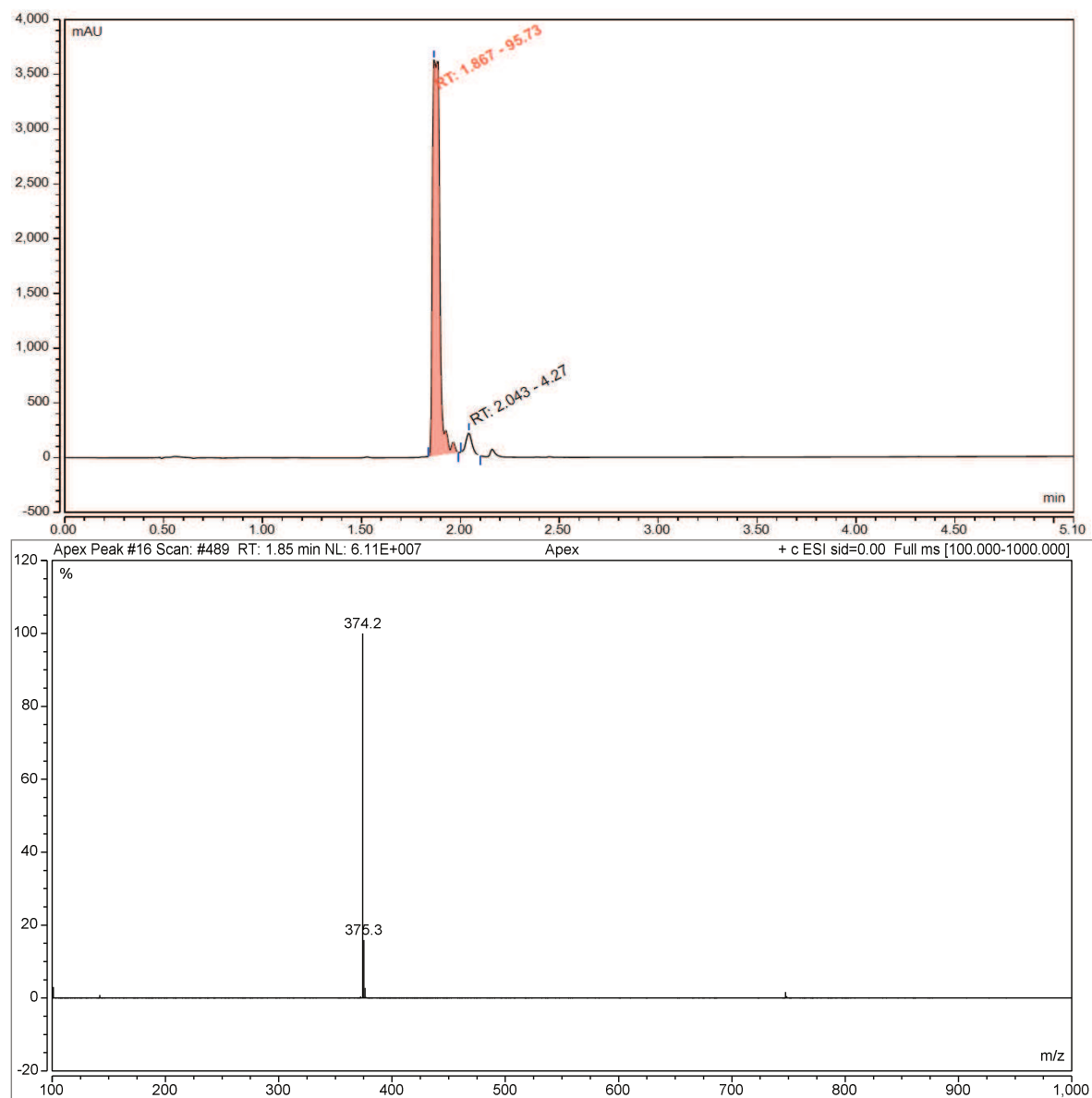


Figure S35. HPLC-MS UV chromatogram and positive ESI-MS of **9a**.

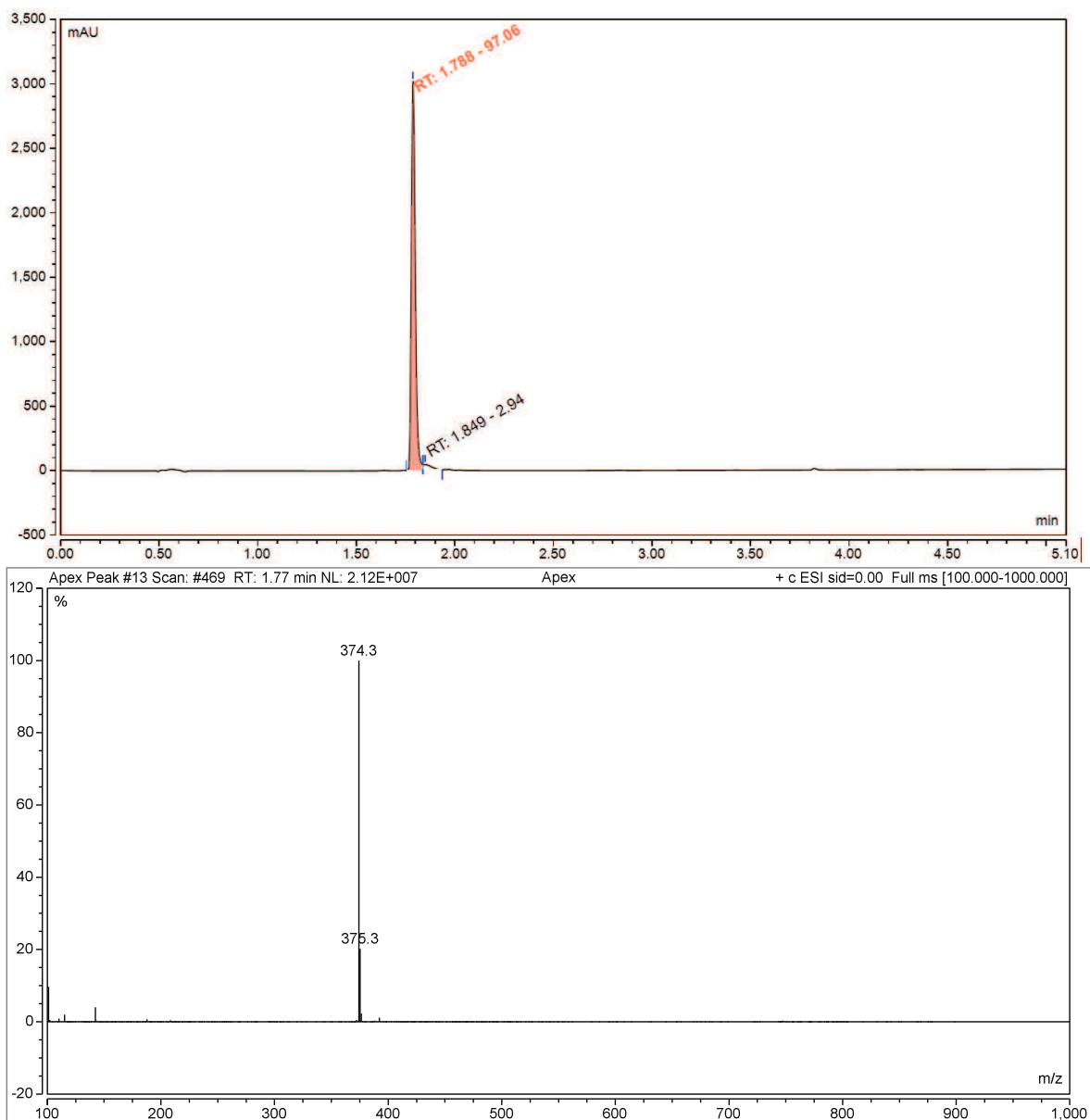


Figure S36. HPLC-MS UV chromatogram and positive ESI-MS of **9b**.



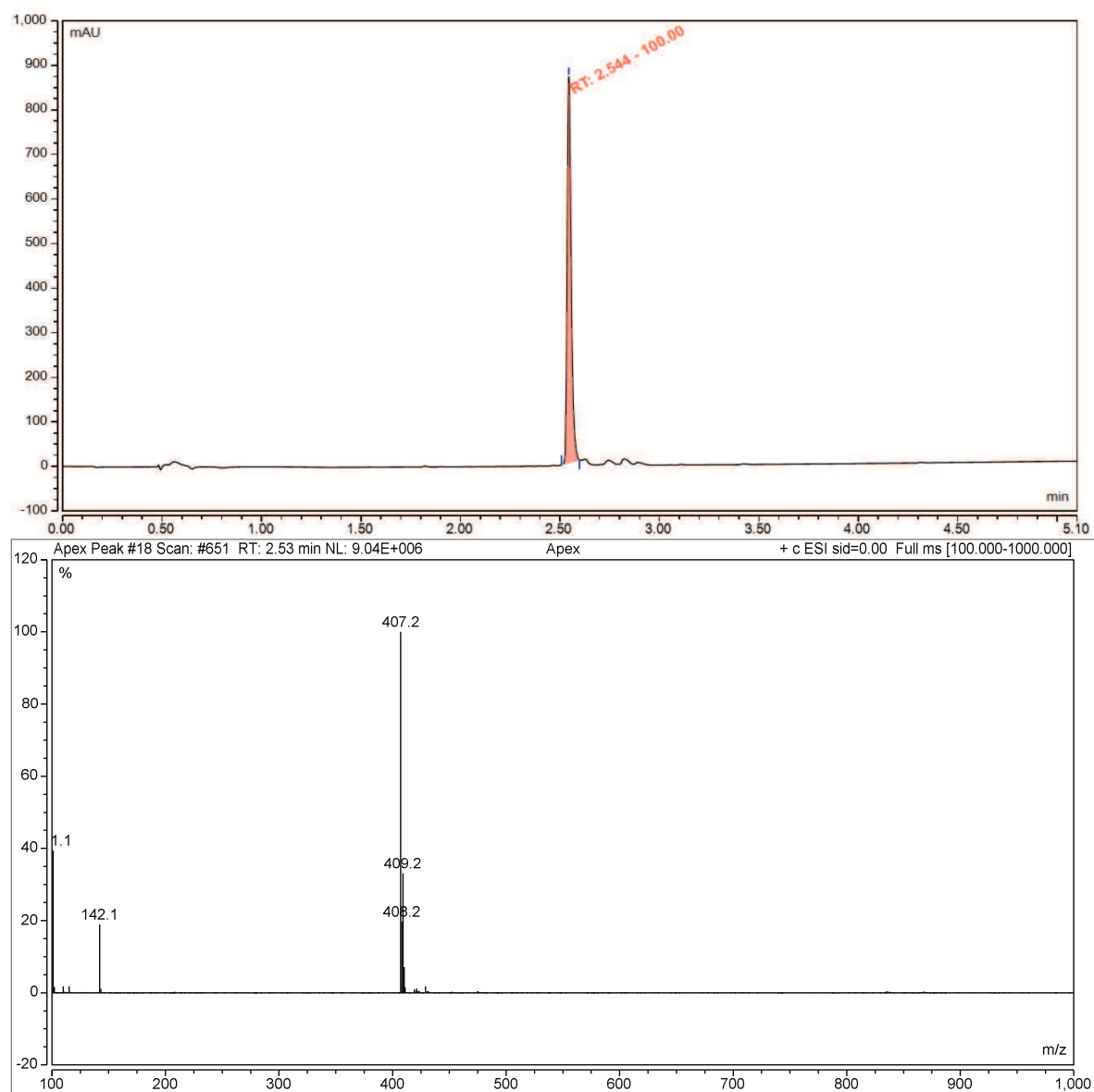


Figure S37. HPLC-MS UV chromatogram and positive ESI-MS of **9c**.

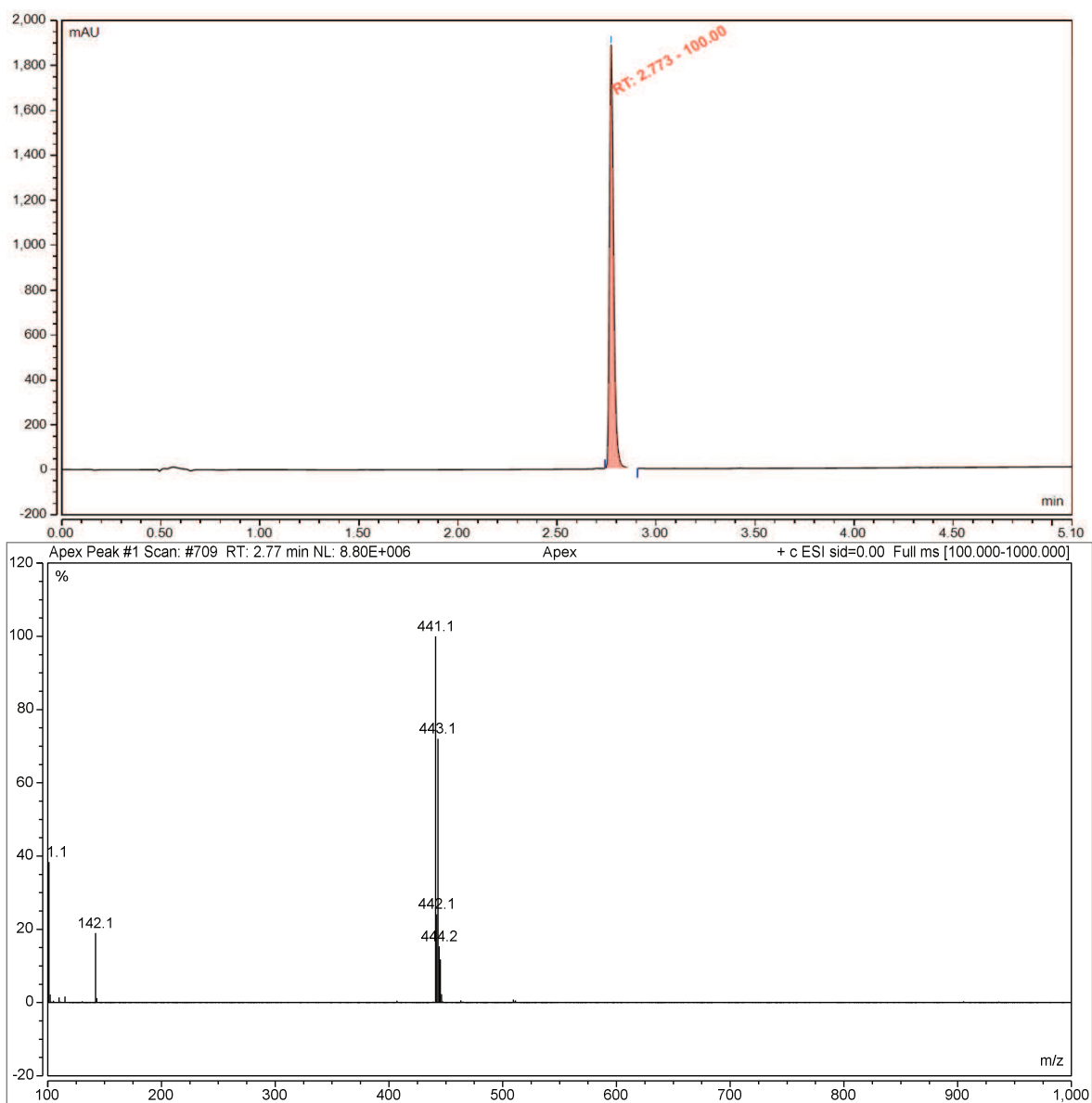


Figure S38. HPLC-MS UV chromatogram and positive ESI-MS of **9d**.

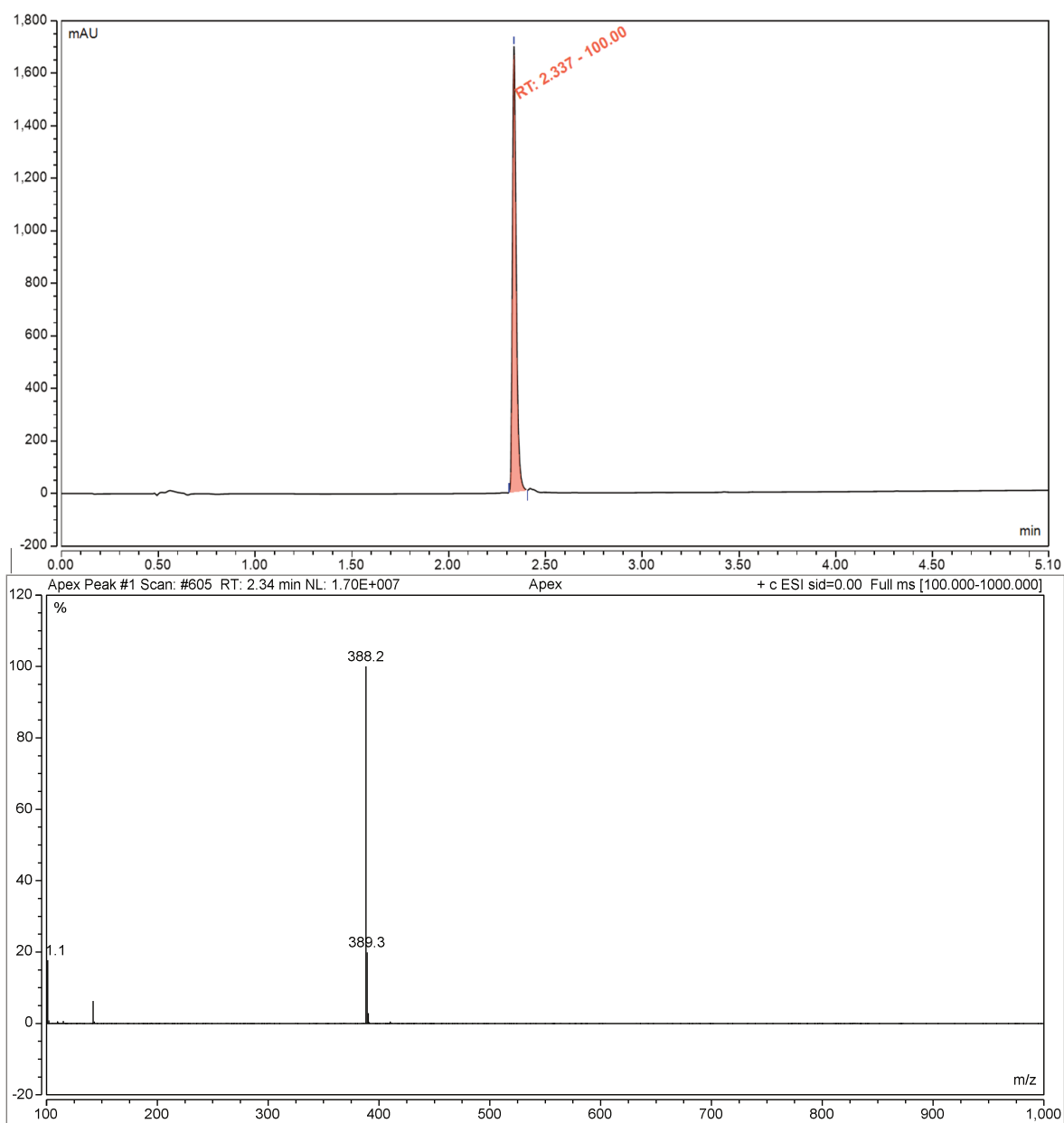


Figure S39. HPLC-MS UV chromatogram and positive ESI-MS of **9e**.

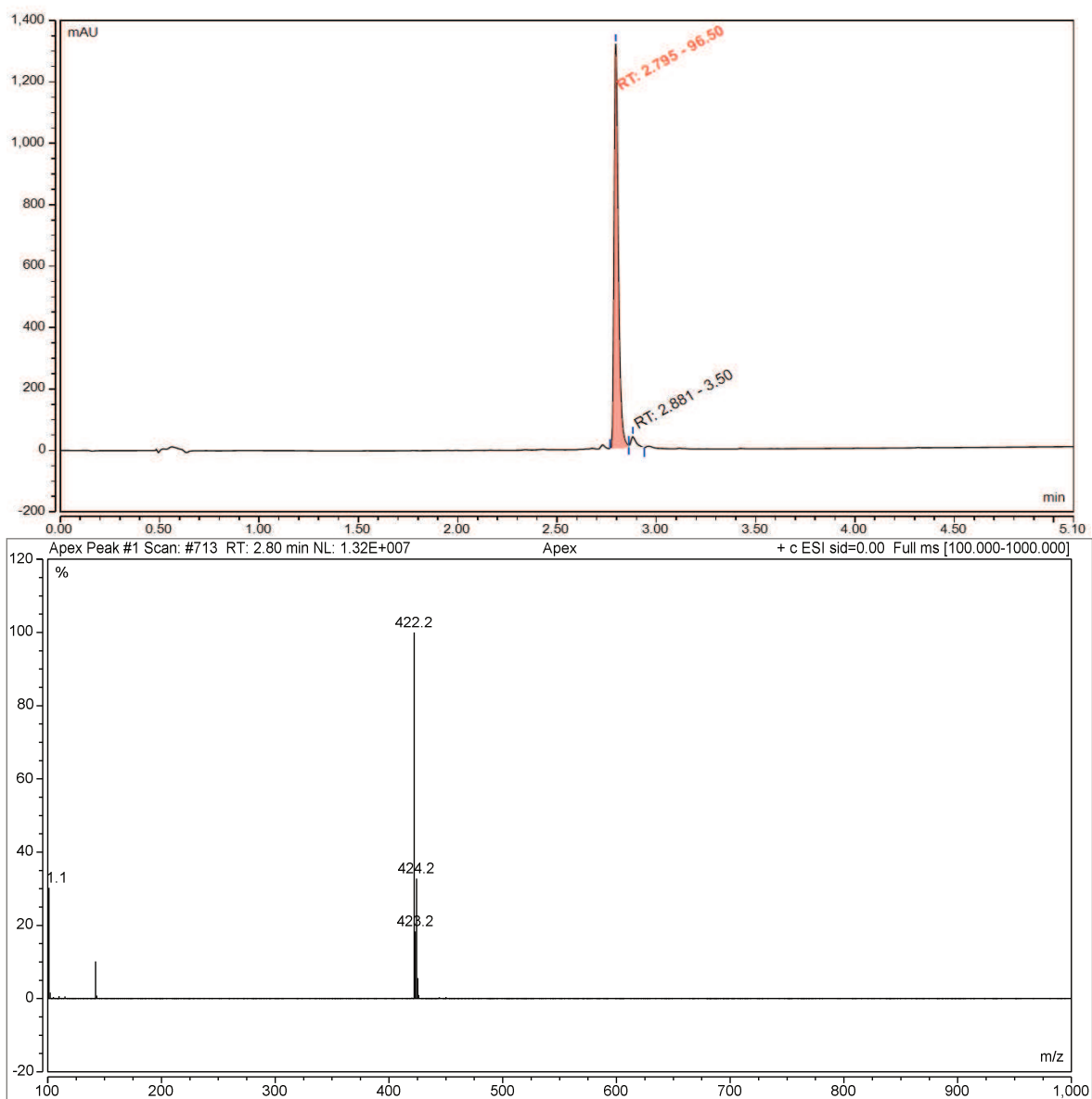


Figure S40. HPLC-MS UV chromatogram and positive ESI-MS of **9f**.

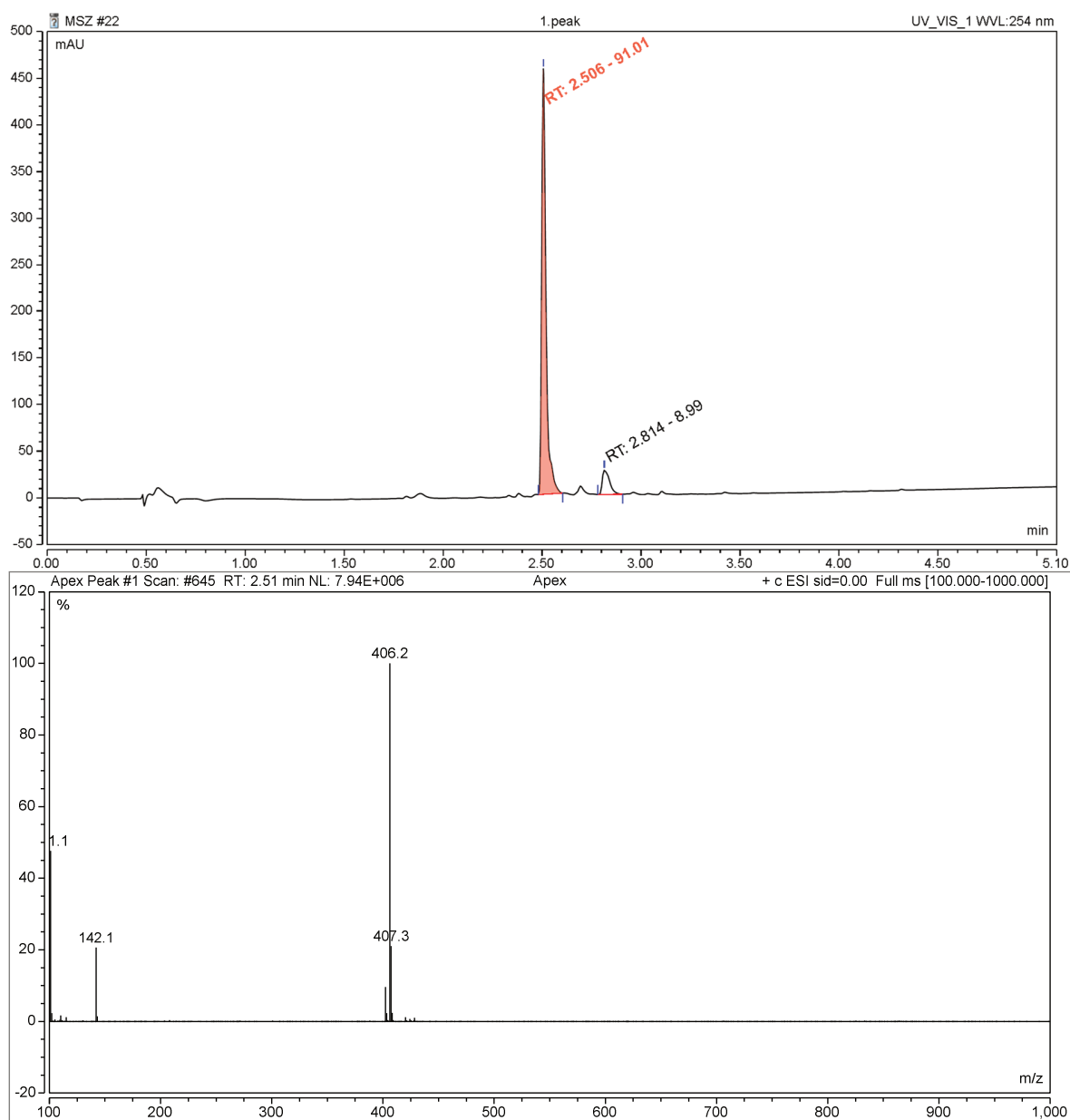


Figure S41. HPLC-MS UV chromatogram and positive ESI-MS of **9g**.

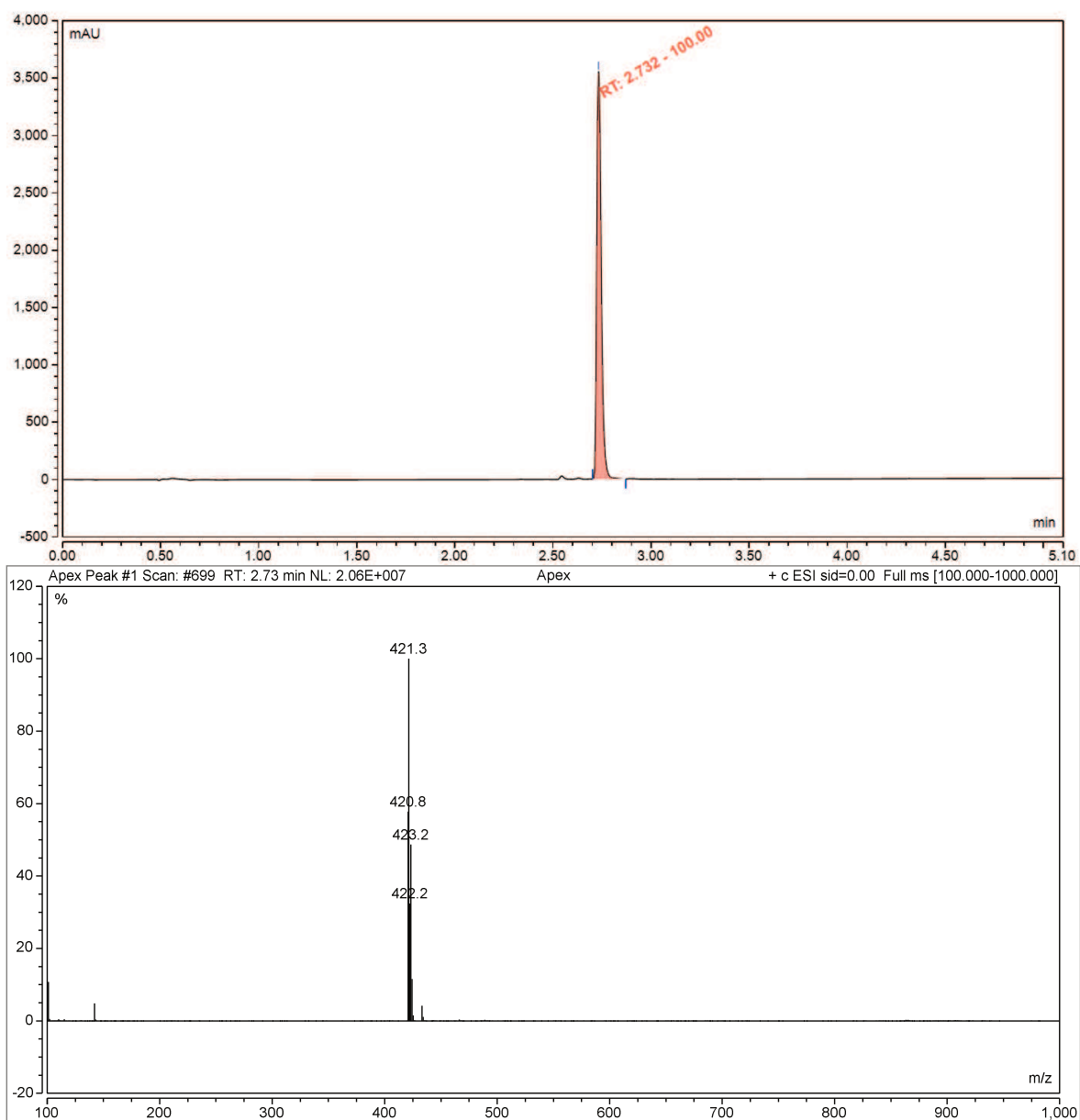
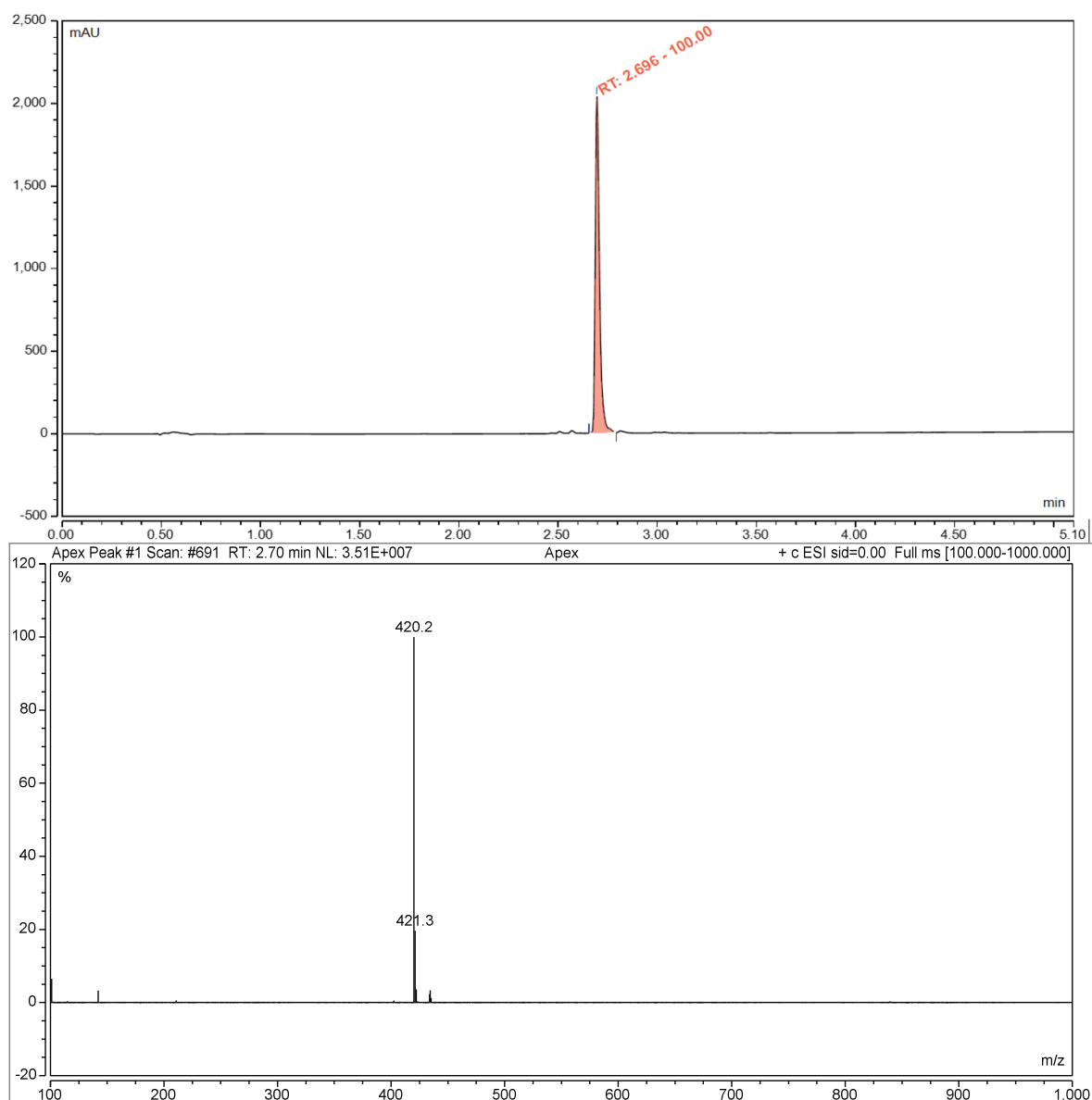


Figure S42. HPLC-MS UV chromatogram and positive ESI-MS of 9h.



**Figure S43.** HPLC-MS UV chromatogram and positive-ESI MS of **9i**.

#### S 4.1.6 Synthesis of Aryl Substituents with an Fmoc-protected Amino Group

**N-Fmoc-2-aminobenzoic acid (SI-1).** To 2-aminobenzoic acid (274 mg, 2.0 mmol), Fmoc-Cl (517 mg, 2.0 mmol) was added, and the mixture was dissolved in dioxane (5 mL). Then, saturated aqueous sodium bicarbonate solution (0.1 mL) was added to the solution, and stirred overnight. The solution was extracted with saturated aqueous ammonium chloride solution and ethyl acetate (3 x 50 mL). The organic layers were combined dried over sulfate and filtered. Then, the solvent was removed under reduced pressure.

The compound was obtained as white solid and it was used without further purification (650 mg, 90%).  $^1\text{H}$  NMR (500 MHz,  $\text{DMSO-}d_6$ ):  $\delta$  11.21 (s, 1H), 8.14 (d,  $J = 8.6$  Hz, 1H), 7.97 (dd,  $J = 7.9, 1.7$  Hz, 1H), 7.92 (d,  $J = 7.5$  Hz, 2H), 7.69 (d,  $J = 7.5$  Hz, 2H), 7.53 (t,  $J = 7.8$  Hz, 1H), 7.46–7.39 (m, 2H), 7.35 (td,  $J = 7.5, 1.2$  Hz, 2H), 7.08 (t,  $J = 7.6$  Hz, 1H), 4.48 (d,  $J = 6.8$  Hz, 2H), 4.35 (t,  $J = 6.8$  Hz, 1H); ESI-MS ( $m/z$ ): calcd for  $\text{C}_{22}\text{H}_{16}\text{NO}_4$  358.11, found 358.24  $[\text{M} - \text{H}]^-$ .

**N-Fmoc-2-amino-3-chlorobenzoic acid (SI-2).** To 2-amino-3-chlorobenzoic acid (348 mg, 2.0 mmol) Fmoc-Cl (517 mg, 2.0 mmol) was added and the mixture was dissolved in dioxane (5 mL). Then, saturated aqueous sodium bicarbonate solution (0.1 mL) was added to the solution, and stirred overnight. The solution was extracted with saturated aqueous ammonium chloride solution and ethyl acetate (3 x 50 mL). The organic layers were combined and dried over sodium sulfate and filtered. Then, the solvent was removed under reduced pressure. The compound was obtained as white solid and it was used without further purification (660 mg, 84%).  $^1\text{H}$  NMR (500 MHz,  $\text{DMSO-}d_6$ ):  $\delta$  13.04 (s, 1H), 7.96–7.80 (m, 2H), 7.83–7.70 (m, 2H), 7.71–7.53 (m, 1H), 7.51–7.14 (m, 6H), 4.36–4.29 (m, 2H), 4.29–4.23 (m, 1H); ESI-MS ( $m/z$ ): calcd for  $\text{C}_{22}\text{H}_{17}\text{ClNO}_4$  394.08, found 394.02  $[\text{M} + \text{H}]^+$ .

**N-Fmoc-2-amino-6-fluorobenzoic acid (SI-3).** To 2-amino-6-fluorobenzoic acid (310 mg, 2.0 mmol) Fmoc-Cl (517 mg, 2.0 mmol) was added and the mixture was dissolved in dioxane (5 mL). Then, saturated aqueous sodium bicarbonate solution (0.1 mL) was added to the solution, and stirred overnight. The solution was extracted with saturated aqueous ammonium chloride solution and ethyl acetate (3 x 50 mL). The organic layers were combined and dried over sodium sulfate and filtered. Then, the solvent was removed under reduced pressure. The compound was obtained as white solid and it was used without further purification (671 mg, 89%). No  $^1\text{H}$  NMR spectrum was recorded; ESI-MS ( $m/z$ ): calcd for  $\text{C}_{22}\text{H}_{15}\text{FNO}_4$  376.11, found 376.24  $[\text{M} - \text{H}]^-$ .



## S 4.1.7 NMR-based Structure Elucidation

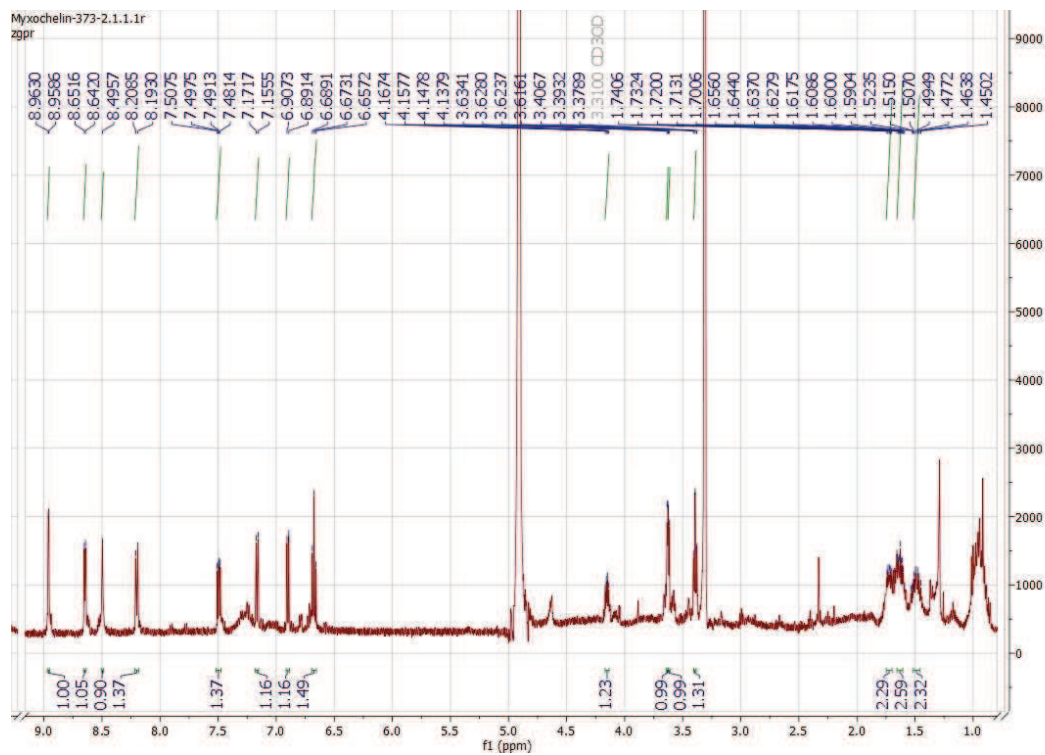
S 4.1.7.1  $^1\text{H}$  and  $^{13}\text{C}$  Spectra of Compounds 1 and 3

Figure S44.  $^1\text{H}$  NMR spectrum of **1** in methanol- $d_4$ .

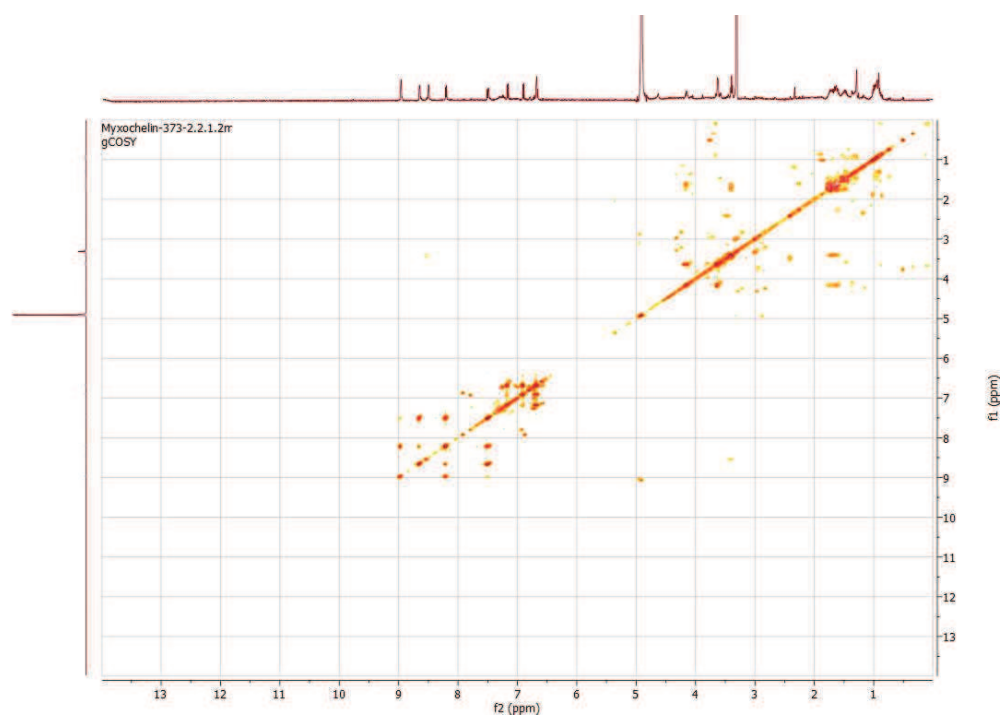


Figure S45.  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of **1** in methanol- $d_4$ .

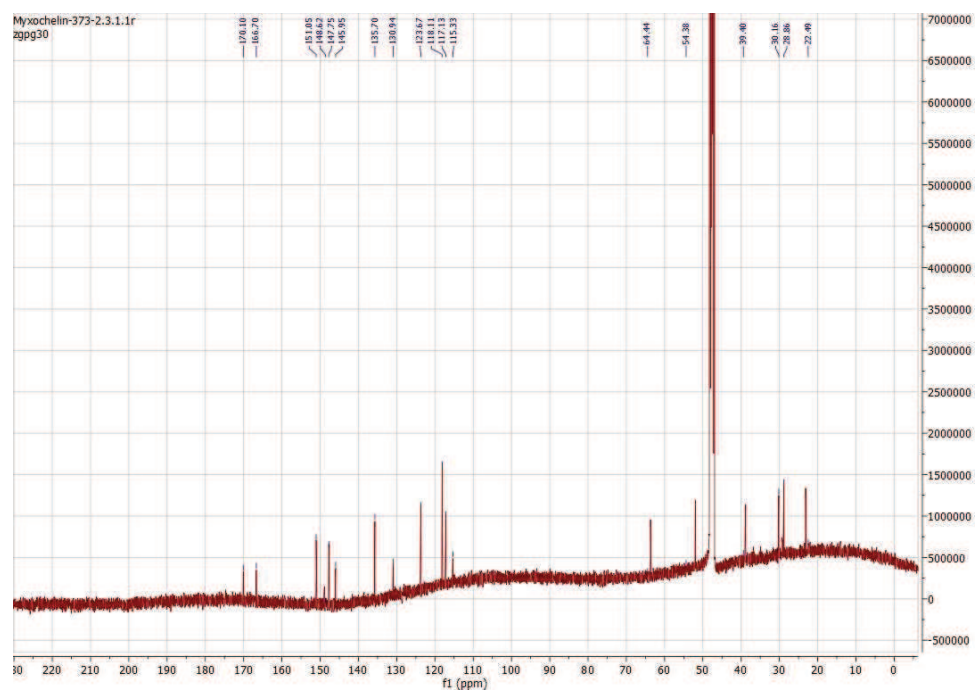


Figure S46.  $^{13}\text{C}$  NMR spectrum of **1** in methanol- $d_4$ .

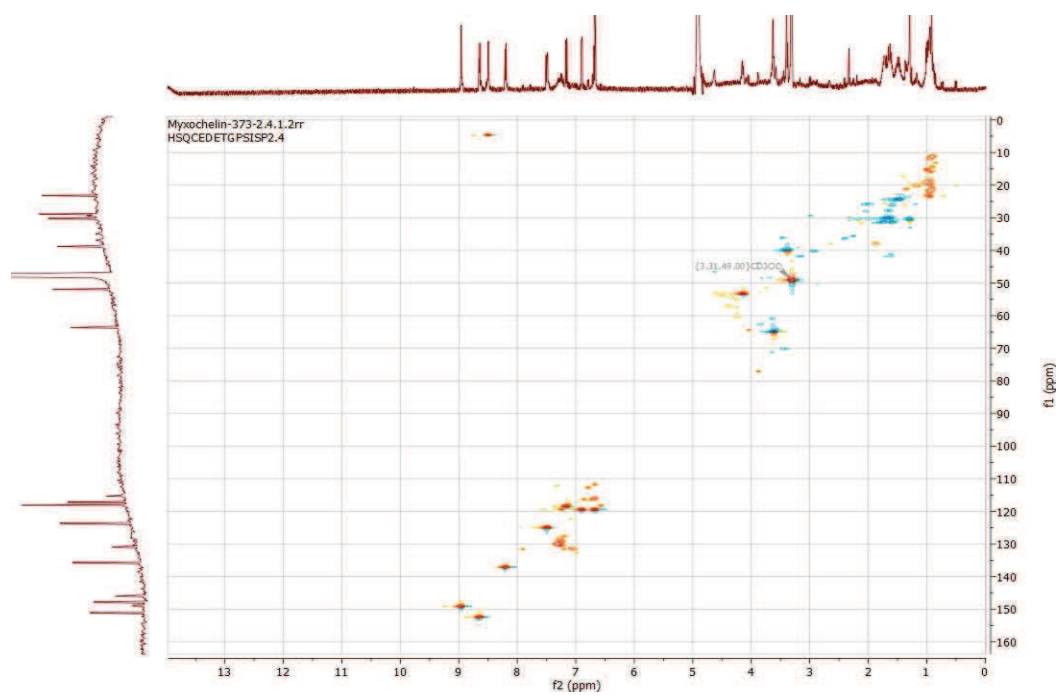


Figure S47. HSQC spectrum of **1** in methanol-*d*<sub>4</sub>.

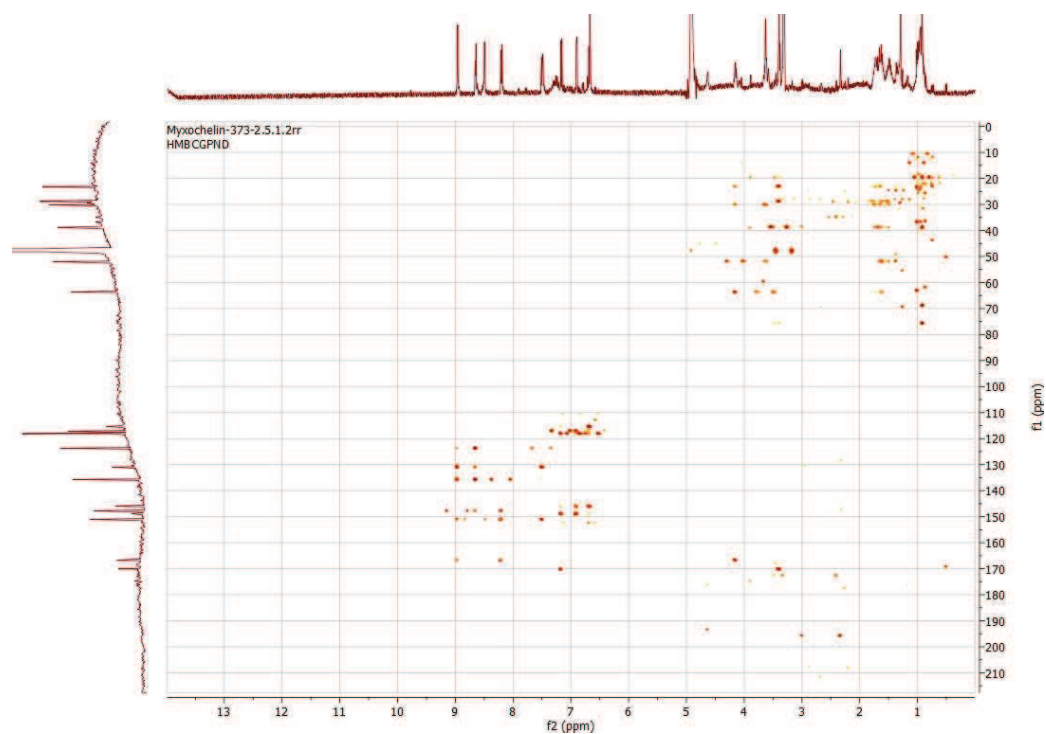


Figure S48. HMBC spectrum of **1** in methanol-*d*<sub>4</sub>.

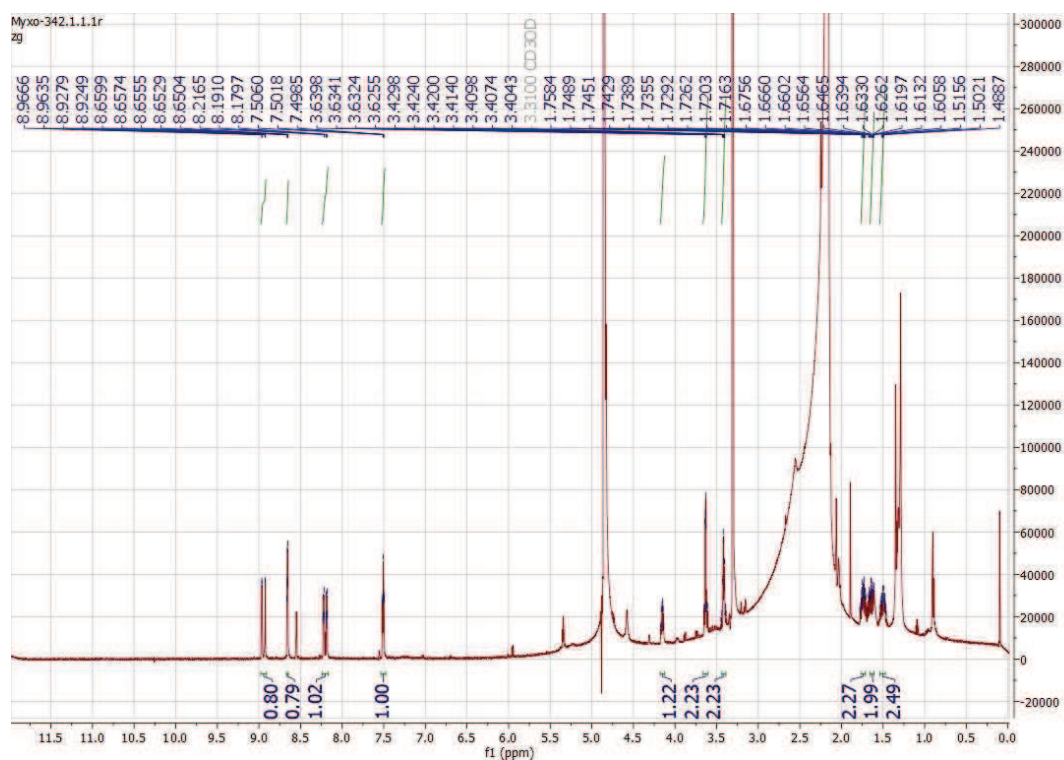


Figure S49.  $^1\text{H}$  NMR spectrum of **3** in methanol- $d_4$ .

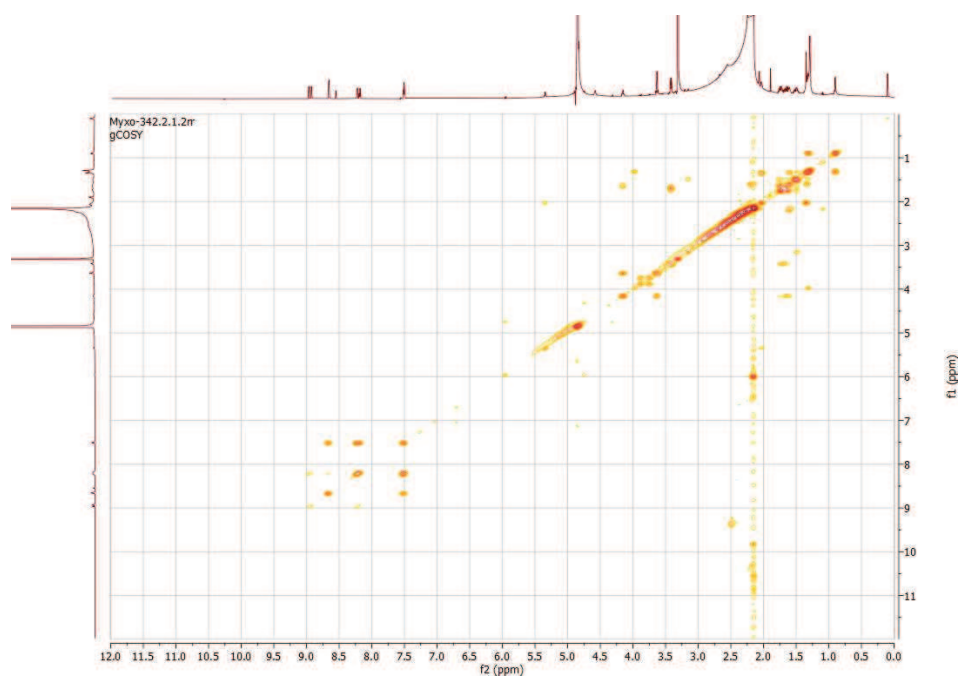


Figure S50.  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of **3** in methanol- $d_4$ .

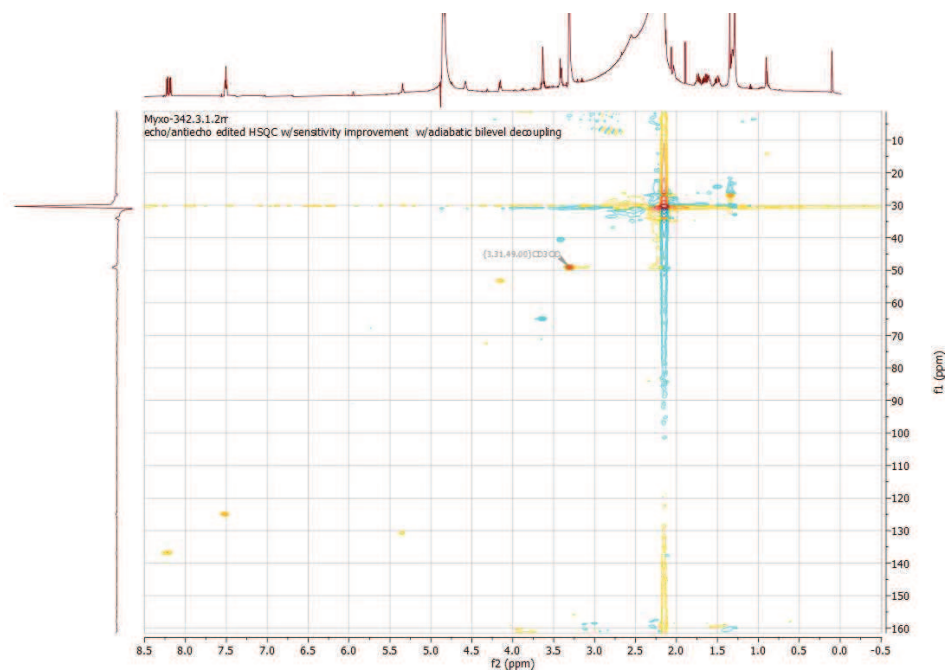


Figure S51. HSQC spectrum of **3** in methanol- $d_4$ .

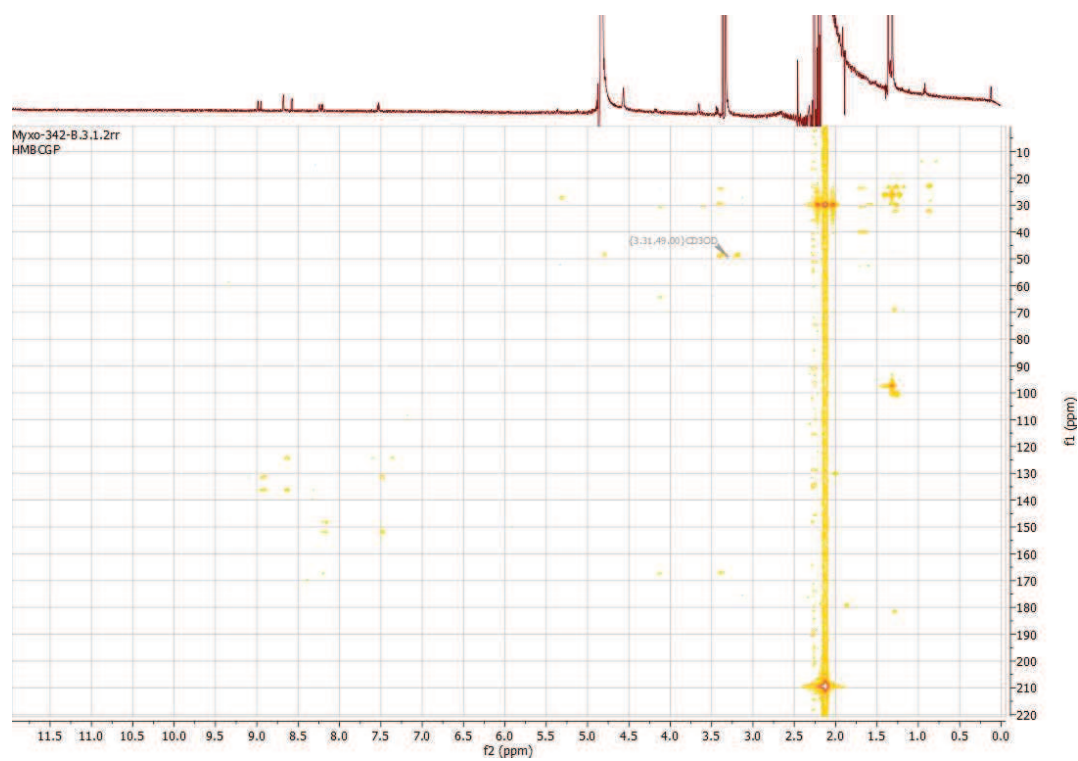


Figure S52. HMBC spectrum of **3** in methanol- $d_4$ .

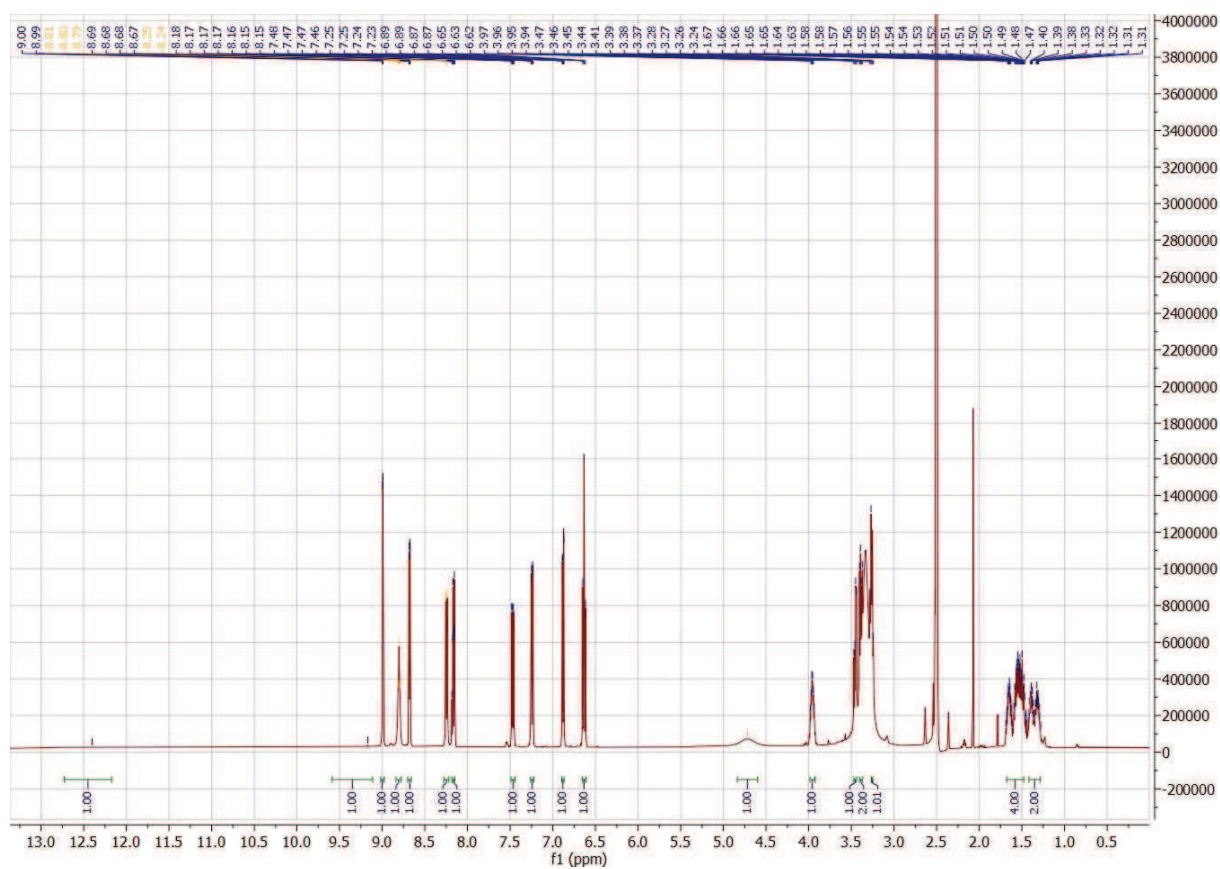
S 4.1.7.2  $^1\text{H}$  and  $^{13}\text{C}$  Spectra of the Synthetic Compounds

Figure S53.  $^1\text{H}$  NMR spectrum of **9a** in  $\text{DMSO-d}_6$ .

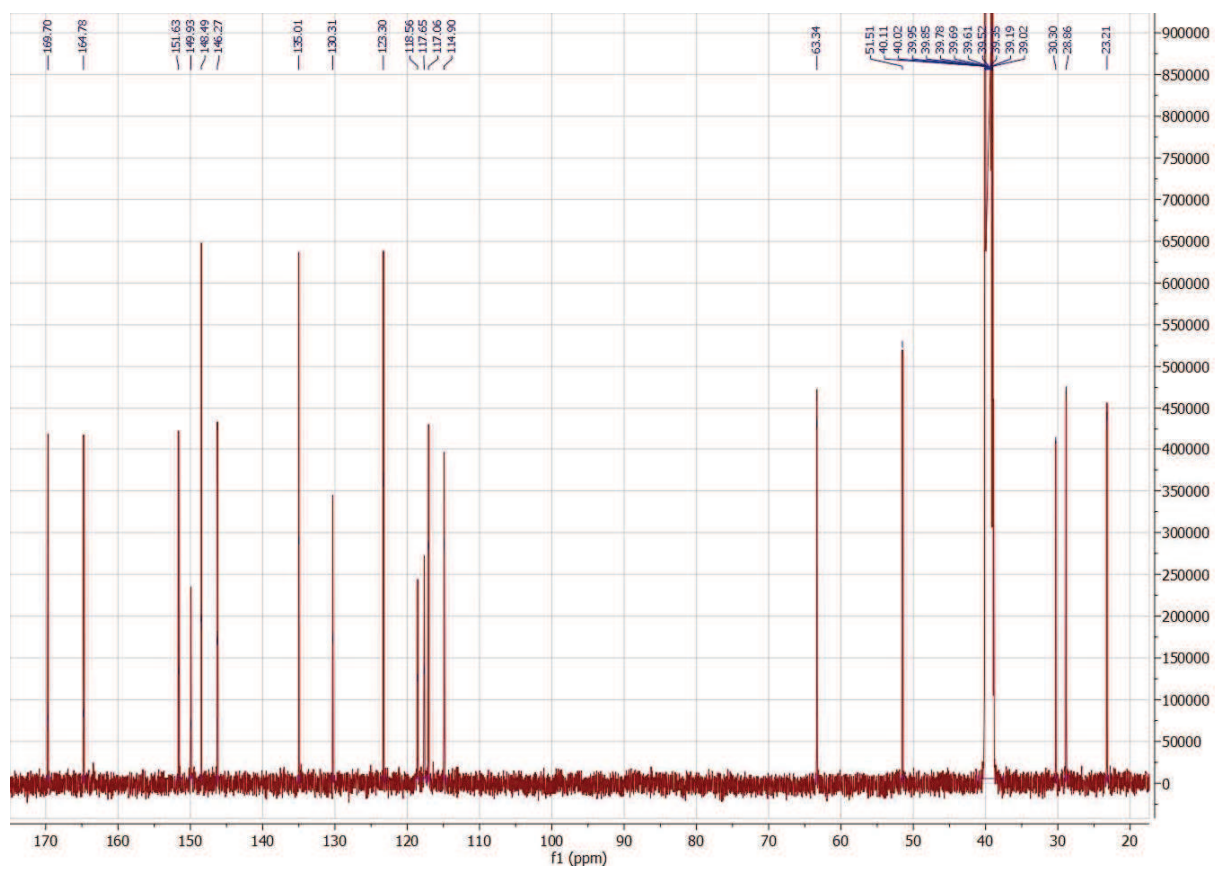


Figure S54.  $^{13}\text{C}$  NMR spectrum of **9a** in  $\text{DMSO-d}_6$ .

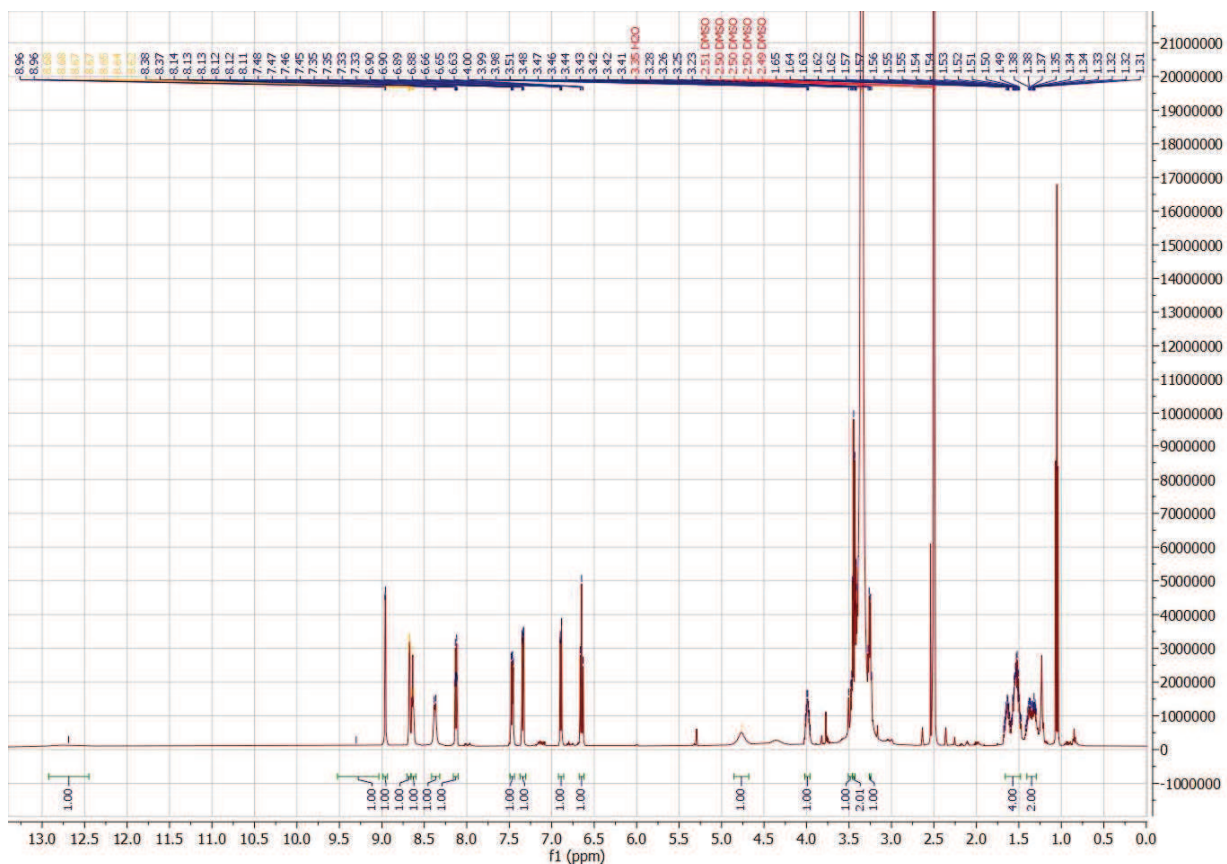


Figure S55.  $^1\text{H}$  NMR spectrum of **9b** in  $\text{DMSO-d}_6$ .



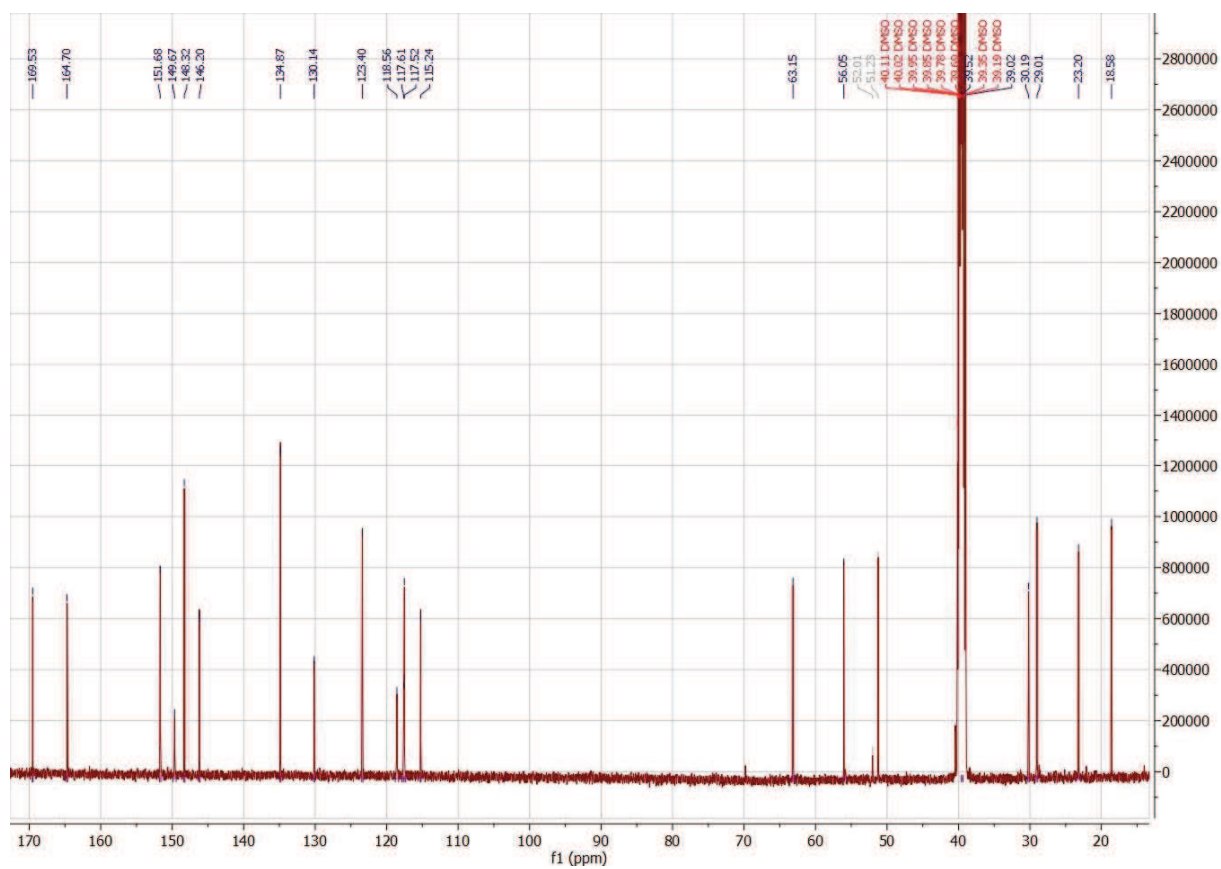


Figure S56.  $^{13}\text{C}$  NMR spectrum of **9b** in  $\text{DMSO-d}_6$ .

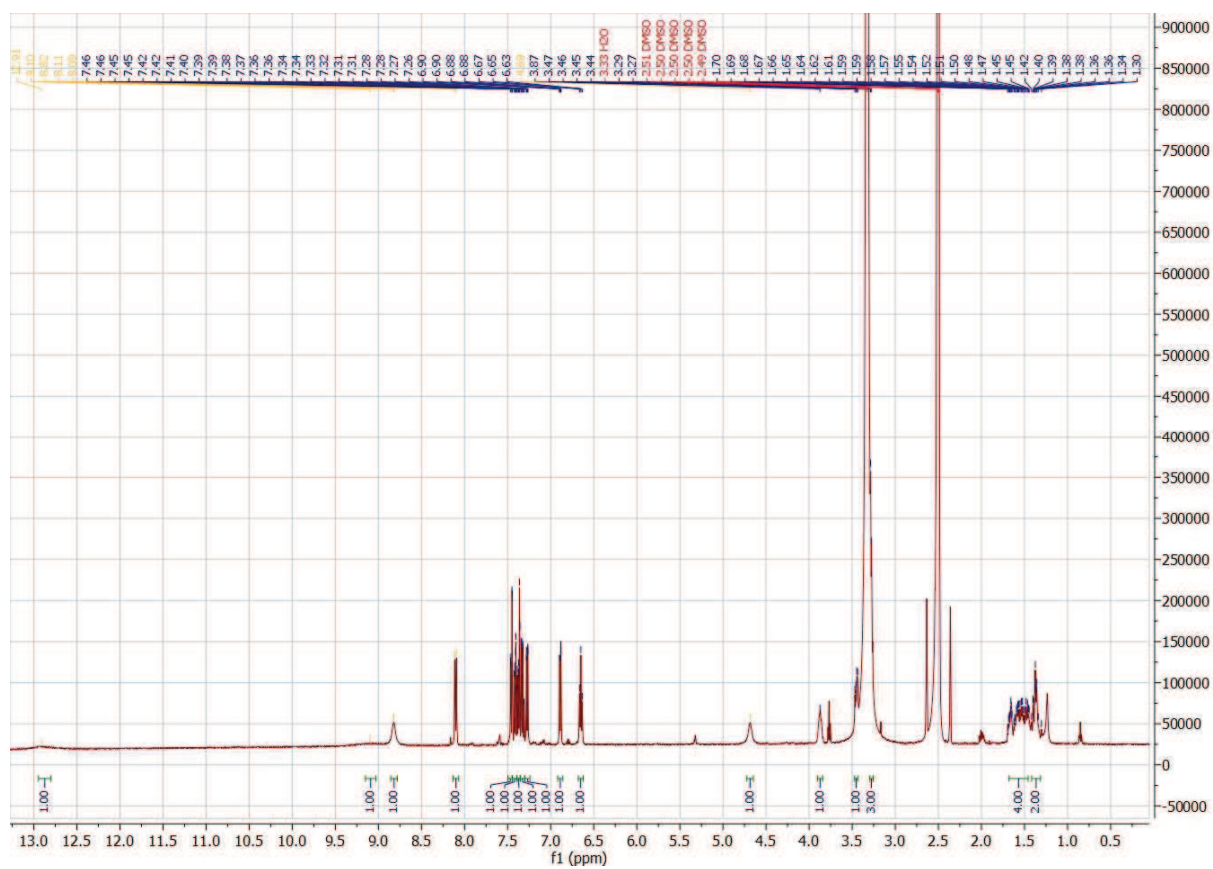


Figure S57.  $^1\text{H}$  NMR spectrum of **9c** in  $\text{DMSO-d}_6$ .

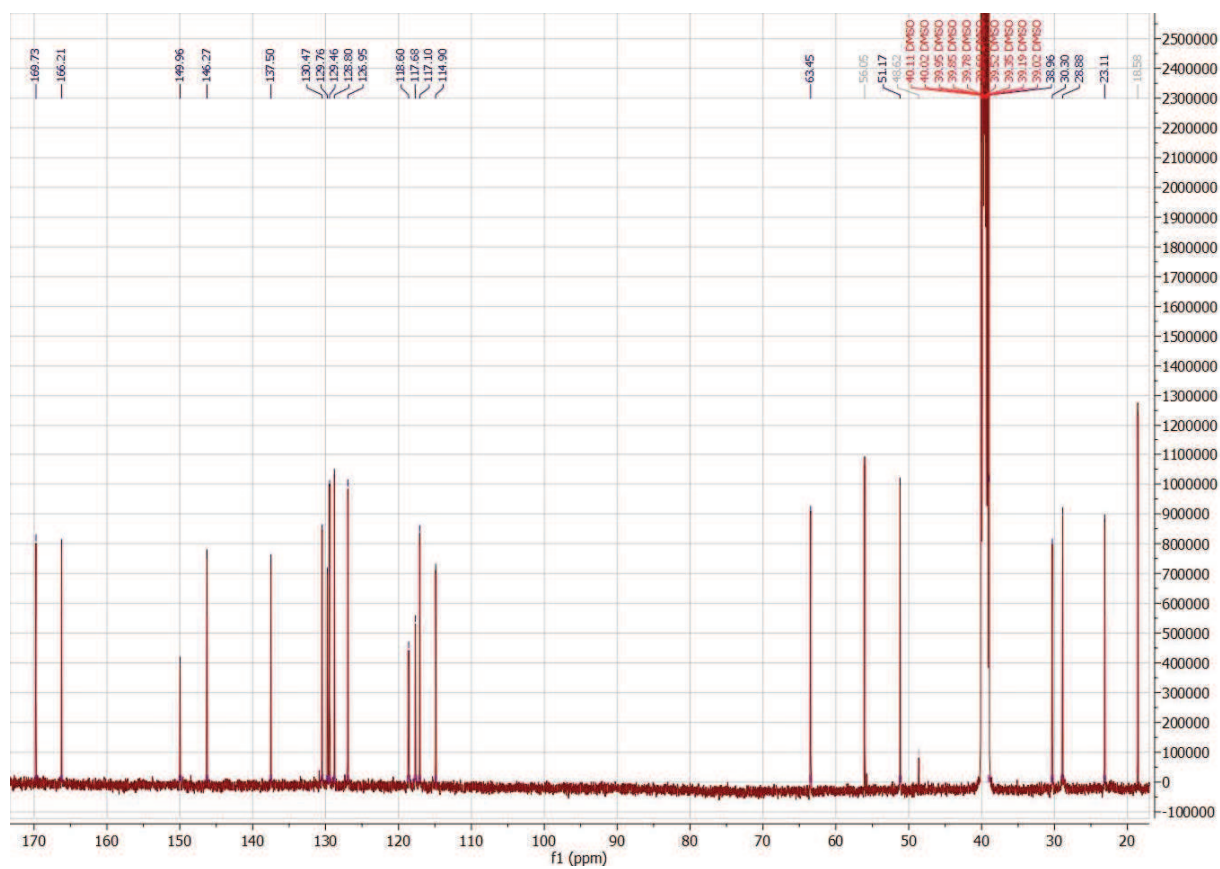


Figure S58.  $^{13}\text{C}$  NMR spectrum of **9c** in  $\text{DMSO-d}_6$ .

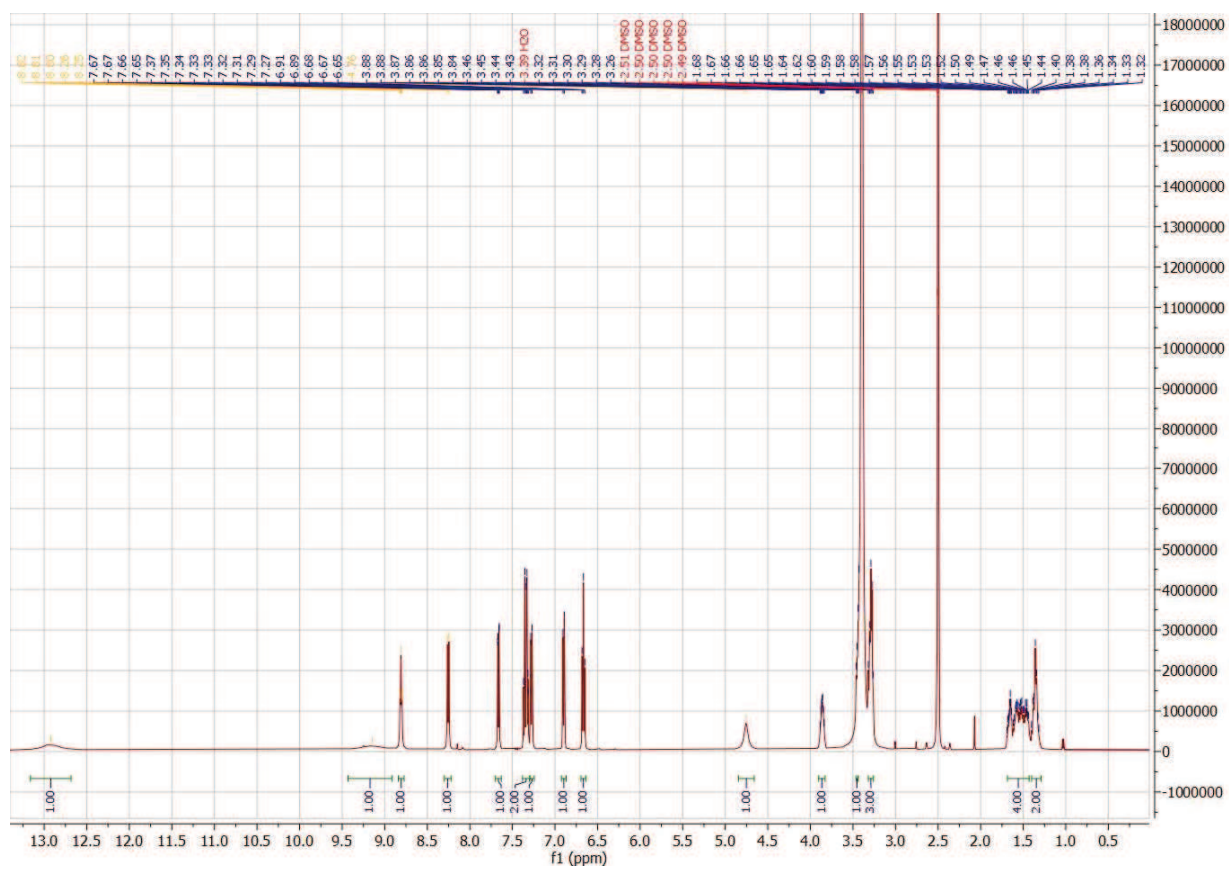
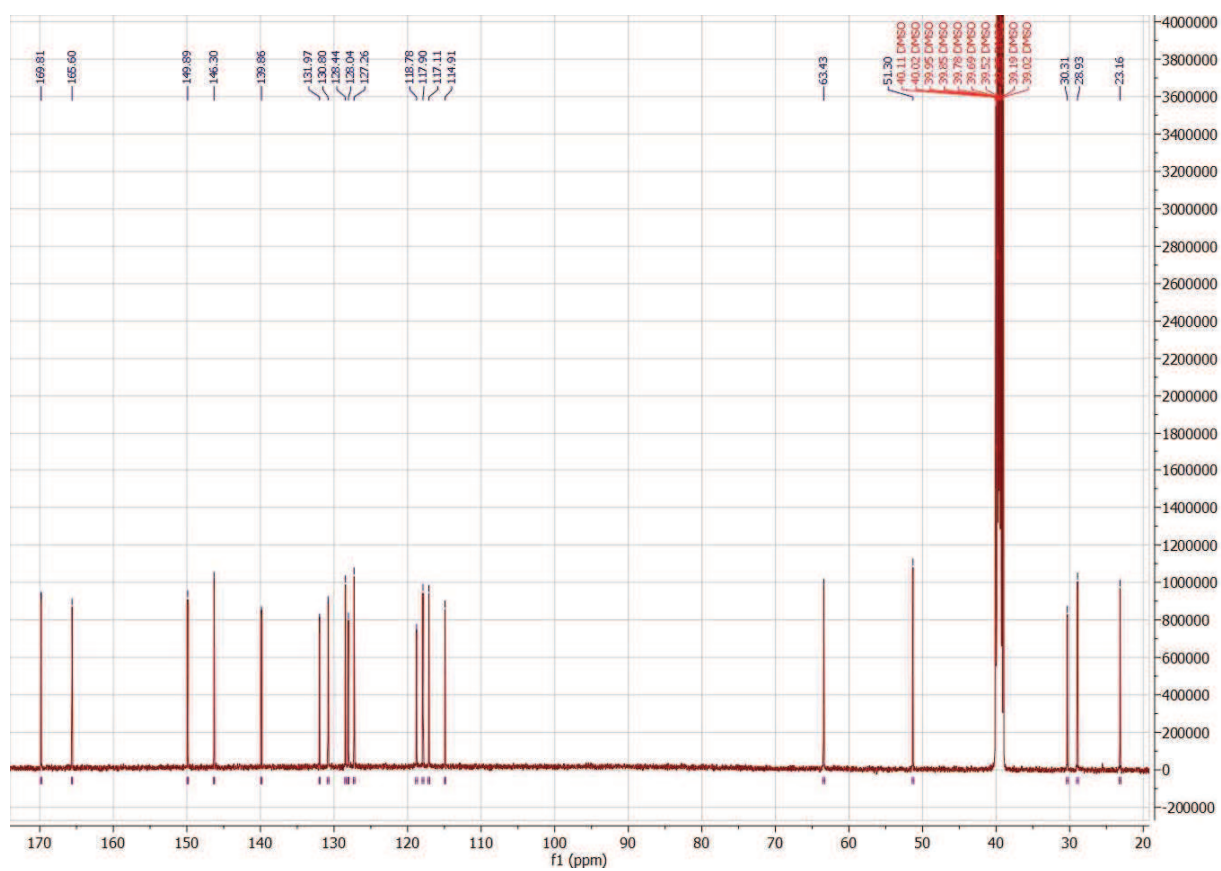


Figure S59.  $^1\text{H}$  NMR spectrum of **9d** in  $\text{DMSO-d}_6$ .



**Figure S60.**  $^{13}\text{C}$  NMR spectrum of **9d** in  $\text{DMSO-d}_6$ .

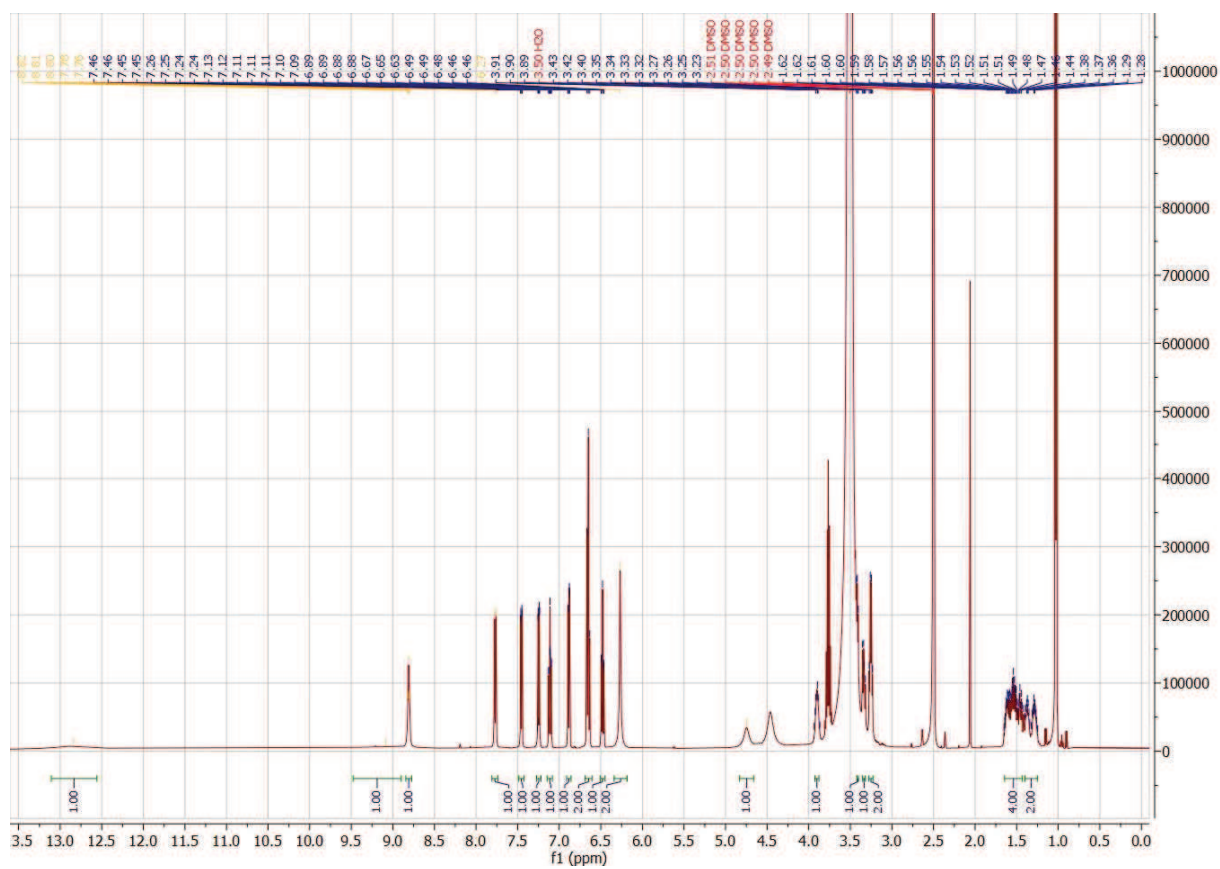


Figure S61.  $^1\text{H}$  NMR spectrum of **9e** in  $\text{DMSO-d}_6$ .

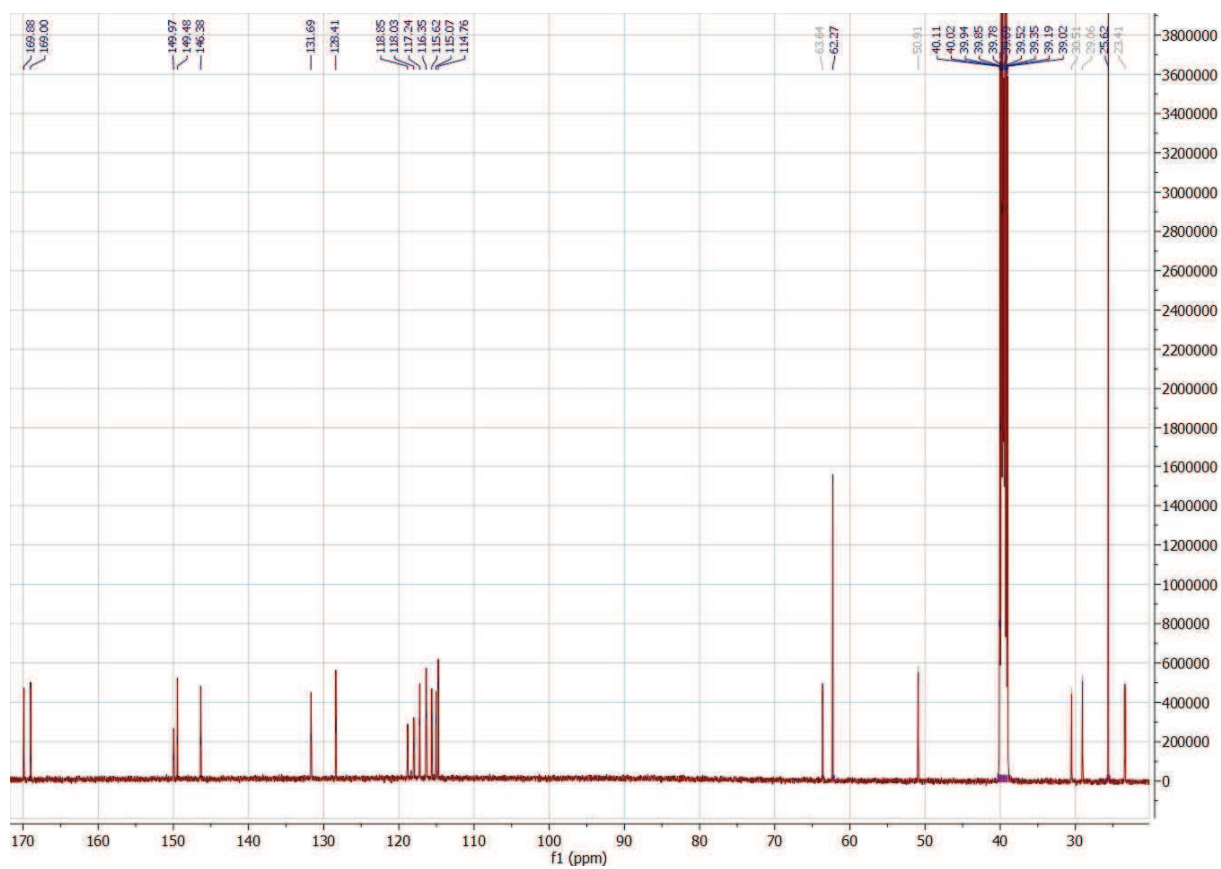


Figure S62.  $^{13}\text{C}$  NMR spectrum of **9e** in  $\text{DMSO-d}_6$ .

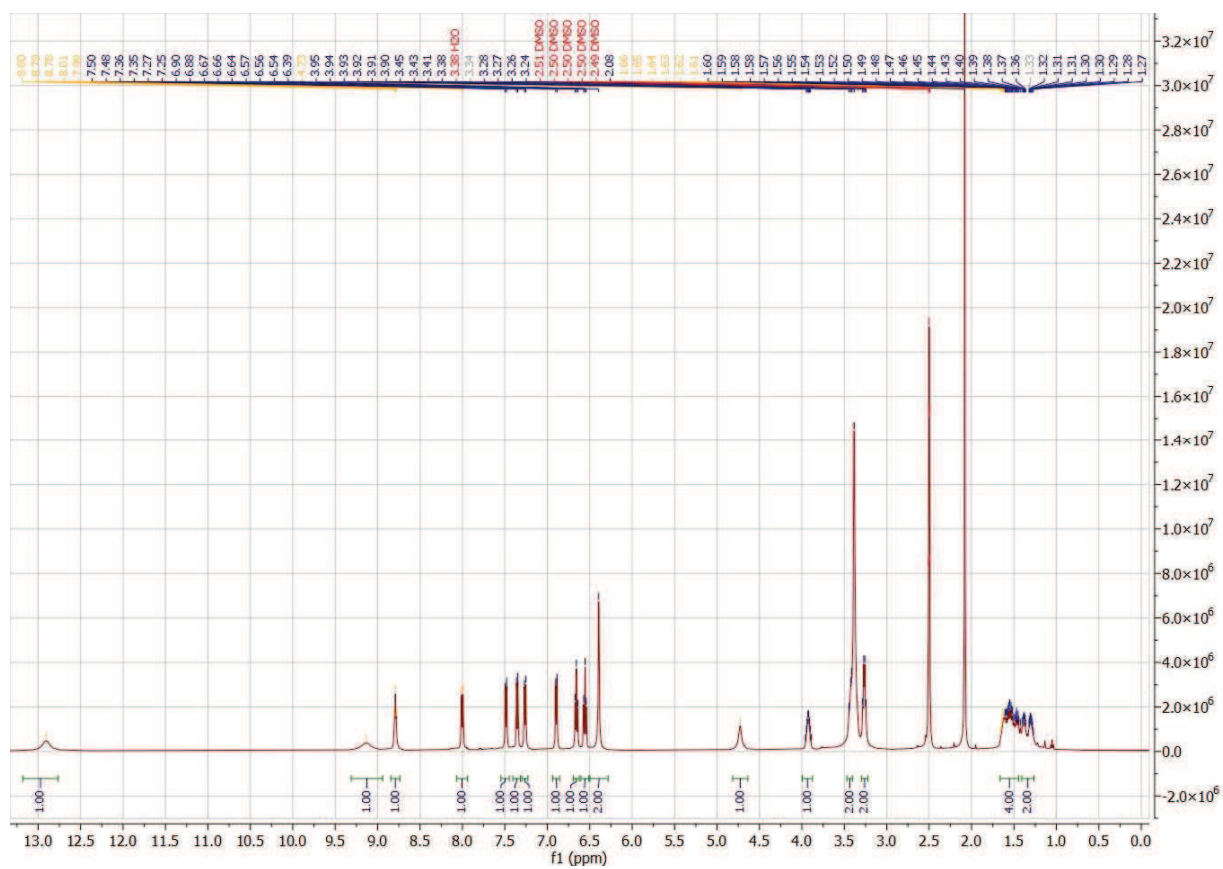


Figure S63.  $^1\text{H}$  NMR spectrum of **9f** in  $\text{DMSO-d}_6$ .



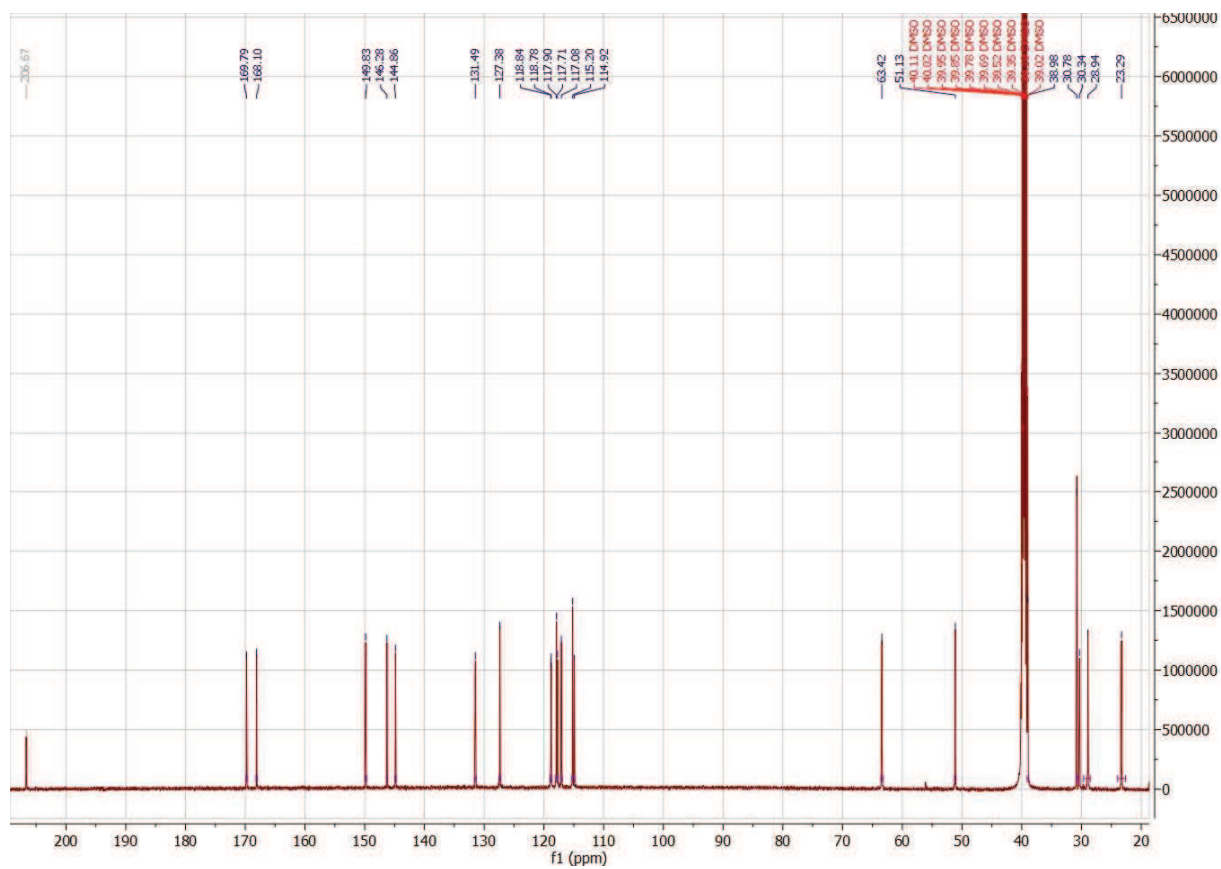


Figure S64.  $^{13}\text{C}$  NMR spectrum of **9f** in  $\text{DMSO-d}_6$ .

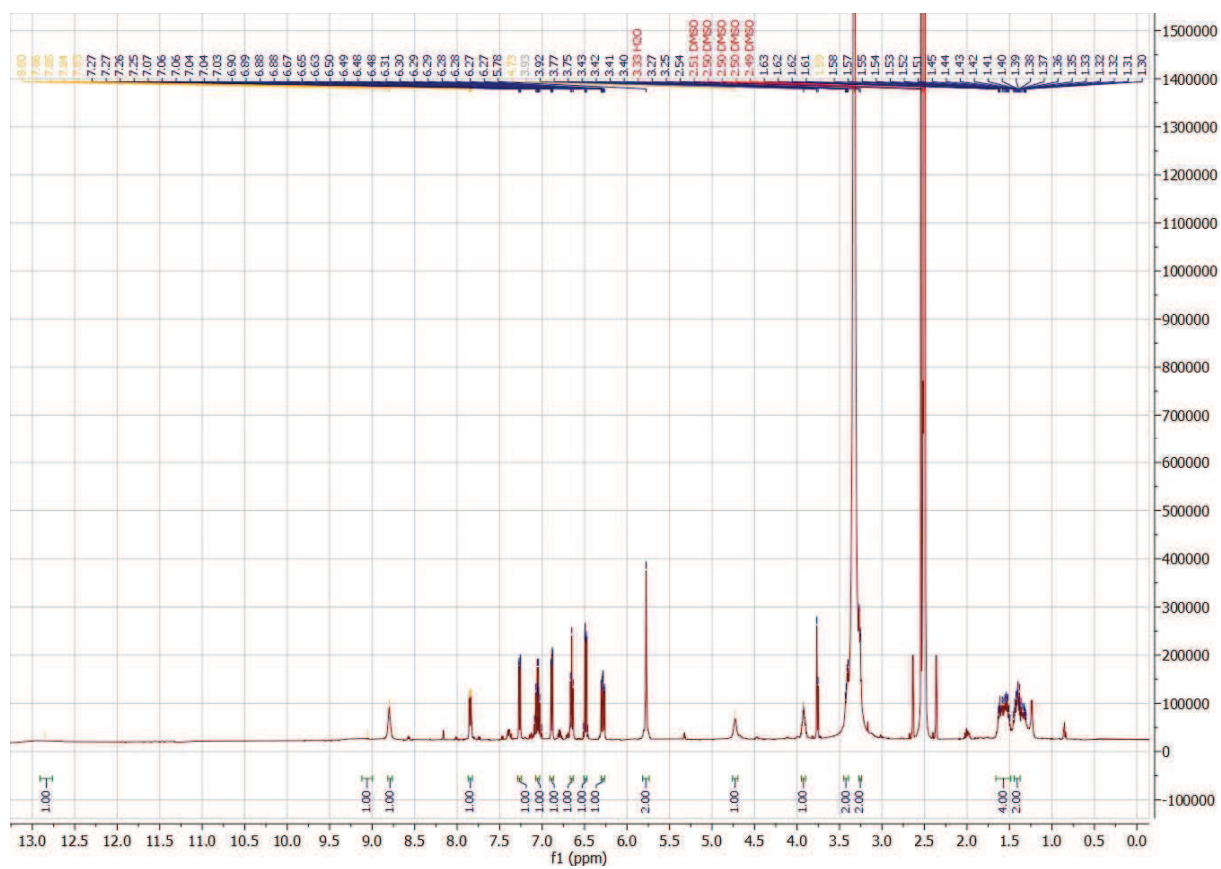


Figure S65.  $^1\text{H}$  NMR spectrum of **9g** in  $\text{DMSO-d}_6$ .

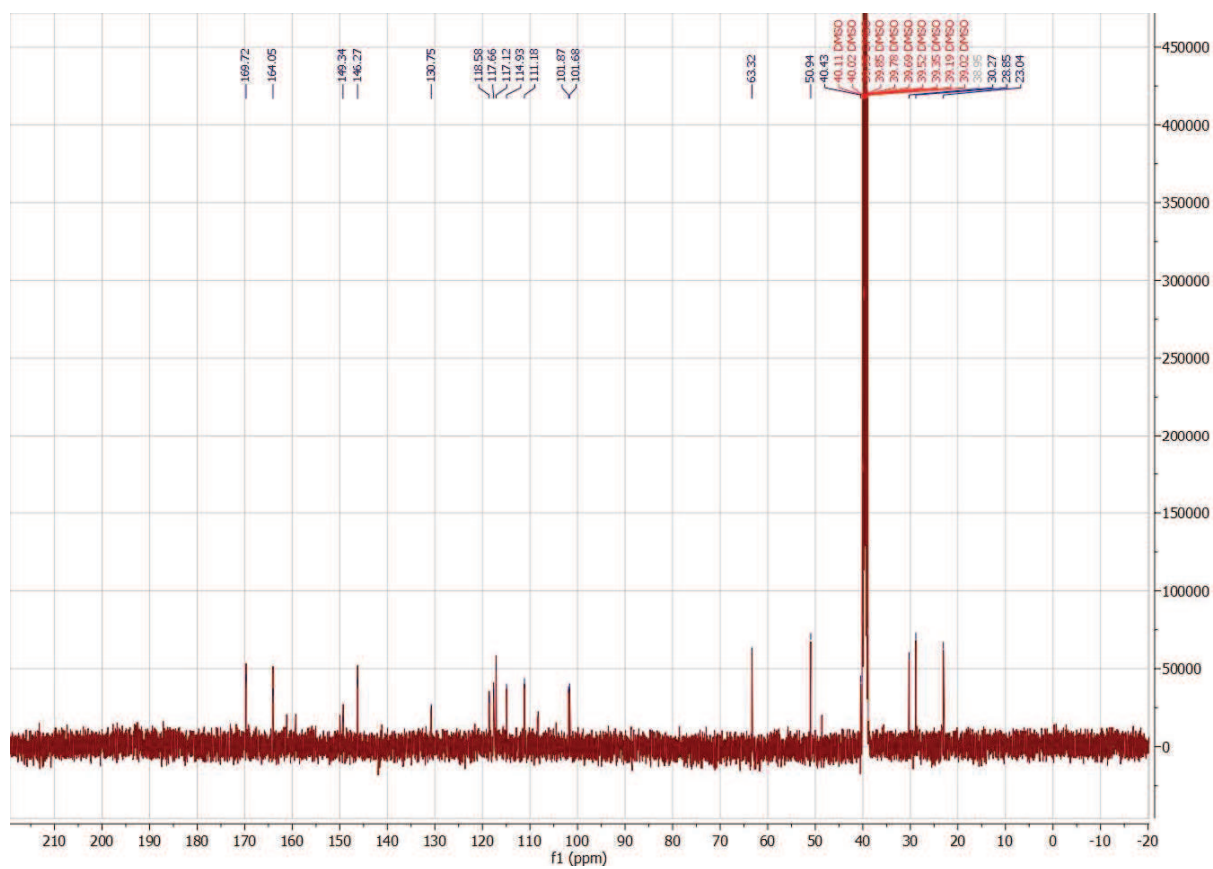


Figure S66.  $^{13}\text{C}$  NMR spectrum of **9g** in  $\text{DMSO-d}_6$ .

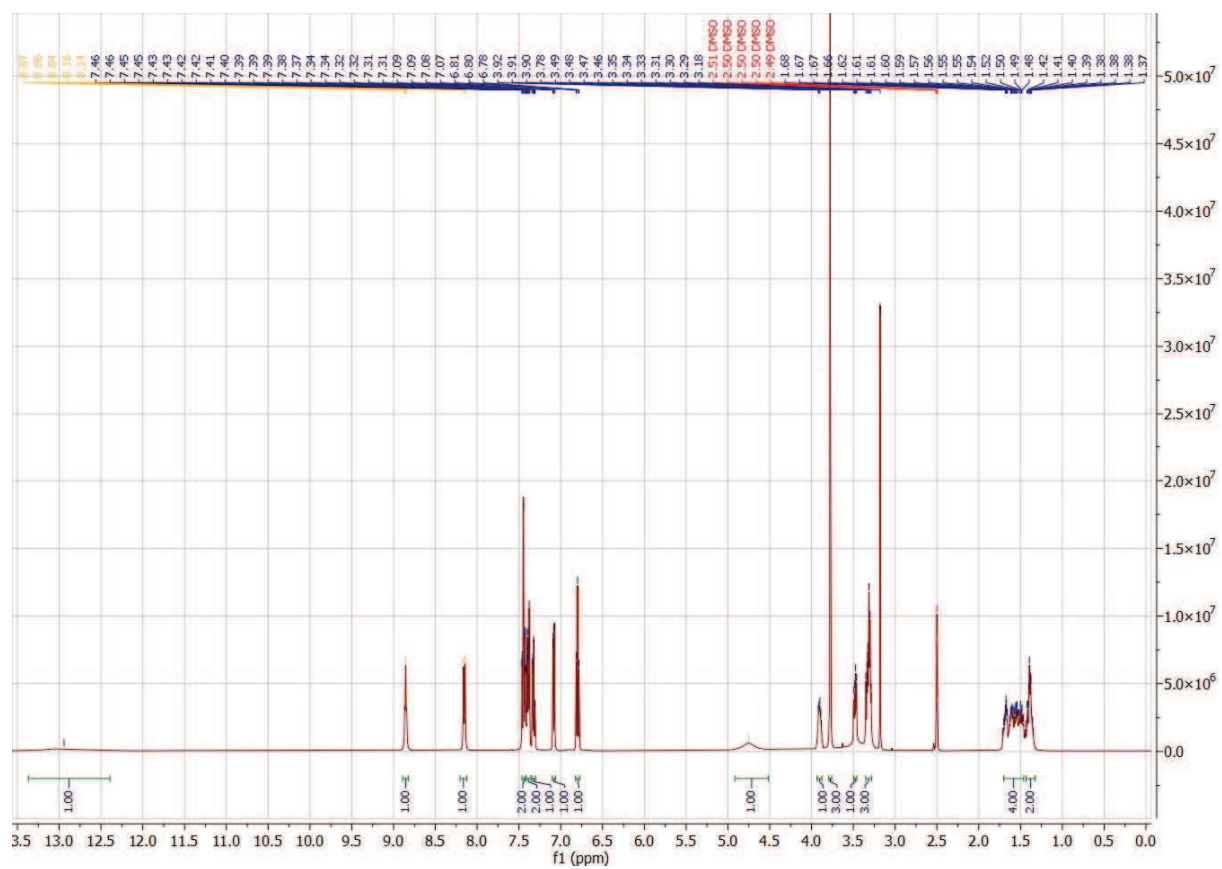


Figure S67.  $^1\text{H}$  NMR spectrum of **9h** in  $\text{DMSO}-d_6$ .

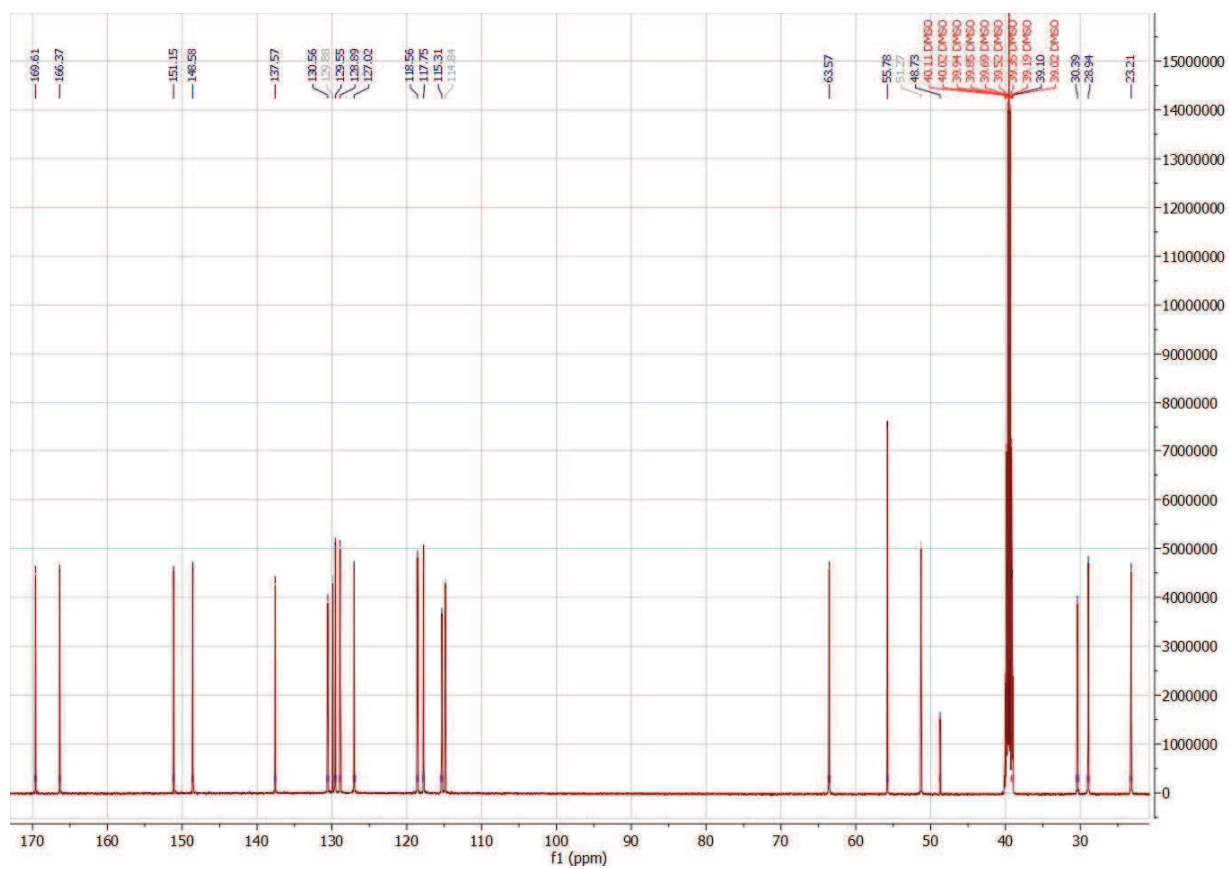


Figure S68.  $^{13}\text{C}$  NMR spectrum of **9h** in  $\text{DMSO-d}_6$ .

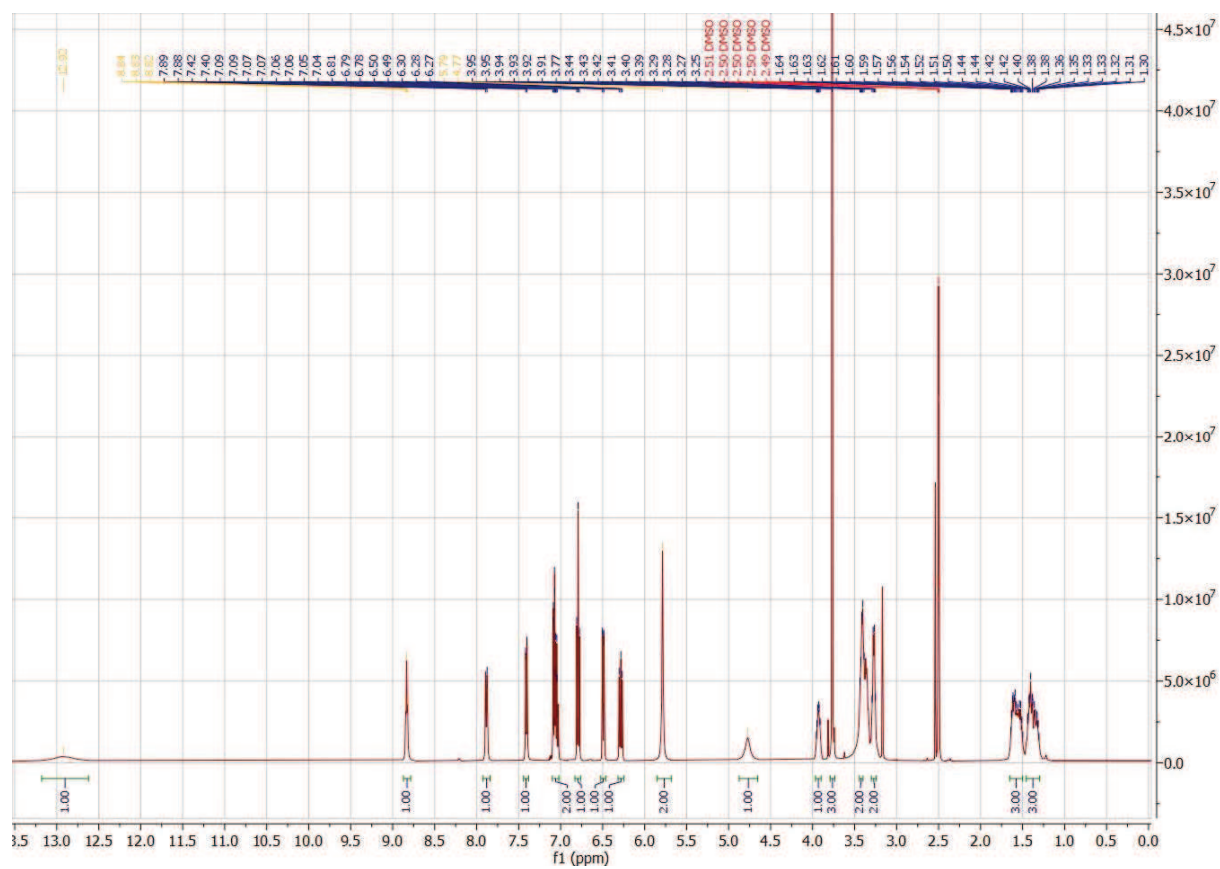


Figure S69.  $^1\text{H}$  NMR spectrum of **9i** in  $\text{DMSO-d}_6$ .

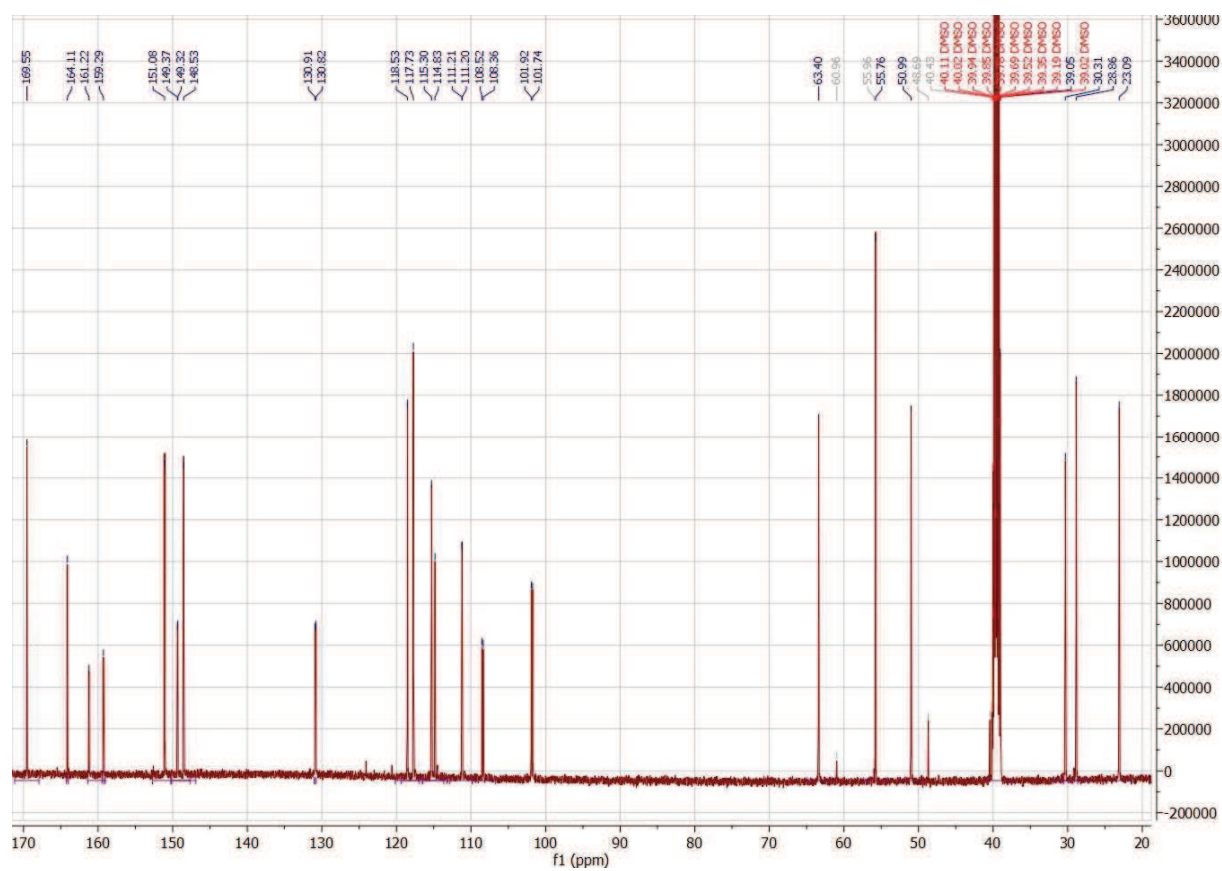
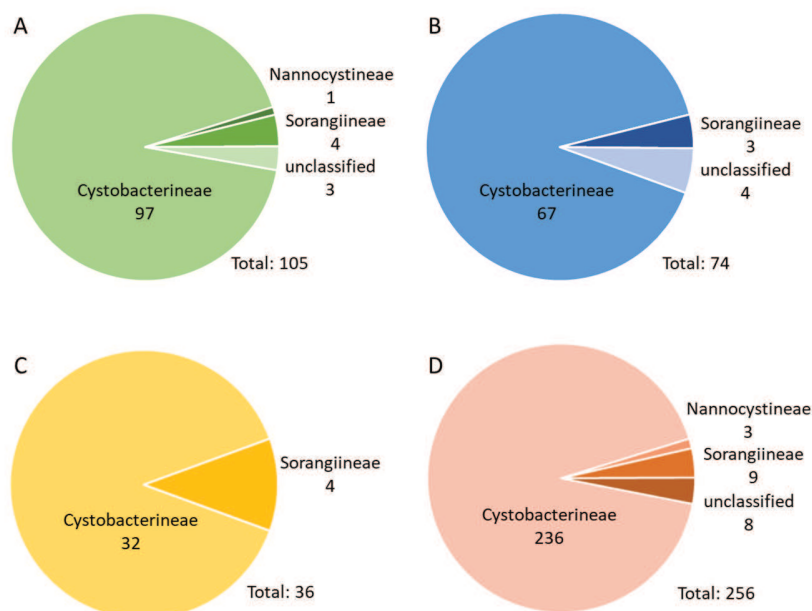


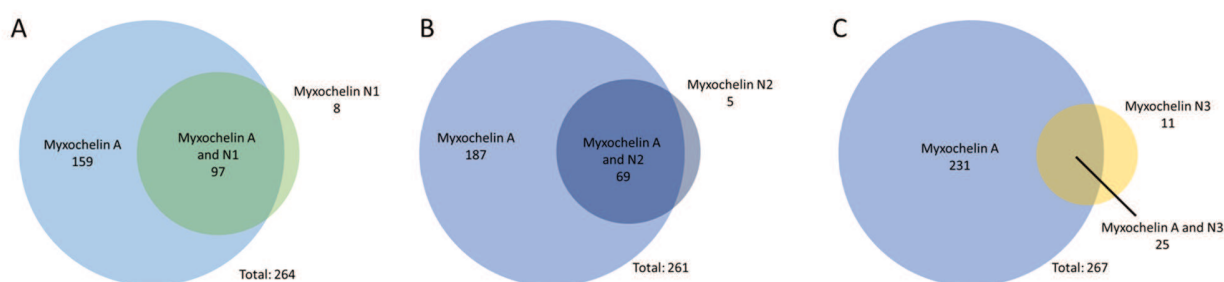
Figure S70.  $^{13}\text{C}$  NMR spectrum of **9i** in  $\text{DMSO-d}_6$ .

## S 4.2 In-silico Analysis of the Myxochelin N1–N3 Biosynthesis

### S 4.2.1 Metabolome Database Search of 1–3



**Figure S71.** Alternative producers of **1–3** (A–C) and myxochelin A (D) from our in-house metabolome database sorted by suborder. Search parameters: exact mass deviation <5ppm, retention time deviation <0.3min, area threshold  $3 \cdot 10^4$ . The decreasing number of hits from **1** to **3** is most likely due to the threshold of the peak finding algorithm.



**Figure S72.** Venn Diagrams of alternative producers of **1–3** (A–C) and of Myxochelin A from our in-house metabolome database. Search parameters: exact mass deviation <5ppm, retention time deviation <0.3min, area threshold  $3 \cdot 10^4$ .

The strains that were found to produce **1**, **2**, and/or **3**, but not myxochelin A were manually screened for presence of **1–3** and myxochelin A. The 24 hits refer to 22 myxobacterial metabolomes in total. Out of these, 10 hits were found to produce myxochelin A and **1**, **2**, and/or **3** and 12 hits were found to produce neither **1–3** nor myxochelin A. No strain was observed to produce **1**, **2**, and/or **3** in absence of myxochelin A.



## S 4.2.2 Analysis of the Myxochelin BGC

Every coding sequence in the myxochelin biosynthesis gene cluster of MCy9049 was extracted translated and searched with the blastp algorithm against the NCBI RefSeq non-redundant protein sequence database [1]. The genes encoding MxcA (NADP-dependent alcohol oxidoreductase) and MxcB (removal of iron from siderophore complex) are found outside the myxochelin regulon.

**Table S1.** Tabulated blastP results for the CDS regions present in the MCy9049 myxochelin biosynthetic gene cluster.

CDS Name	Length [AA]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Proposed function	Accession Nr.
<i>mxuC</i>	280	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase [Coralloccoccus]	97.3 / 257		WP_120546994
<i>mxuD</i>	429	isochorismate synthase Dhbc [Coralloccoccus sp. NCSPR001]	95.6 / 429		WP_206795025
<i>mxuE</i>	544	(2,3-dihydroxybenzoyl)adenylate synthase [Coralloccoccus interemptor]	98.3 / 544		WP_121769586
<i>mxuF</i>	301	isochorismatase family protein [Coralloccoccus sp. AB049A]	98.0 / 301	aryl carrier protein	WP_121723307
<i>mxuG</i>	1576	myxochelin non-ribosomal peptide synthetase MxcG [Coralloccoccus sp. AB049A]	92.6 / 913		WP_121723308
<i>aroAAS</i>	426	3-deoxy-7-phosphoheptulonate synthase [Coralloccoccus sp. AB049A]	99.3 / 426	DAHP synthase	WP_121723309
<i>mxuH</i>	856	TonB-dependent siderophore myxochelin receptor MxcH [Coralloccoccus interemptor]	97.0 / 856		WP_121769582
<i>mxuI</i>	422	hypothetical protein [Coralloccoccus interemptor]	92.9 / 422	unknown	WP_121769581
<i>mxuK</i>	405	myxochelin export MFS transporter MxcK [Coralloccoccus sp. AB049A]	96.3 / 405		WP_121725177
<i>mxuL</i>	427	myxochelin B biosynthesis transaminase MxcL [Coralloccoccus interemptor]	98.1 / 427		WP_121769587

**Table S2.** A-domain specificity codes for *mxxE* homologs from NRPSpredictor2 [2]. Prediction for all hits: dihydroxybenzoic acid or salicylic acid.

CDS Name	Organism of origin	8 angstrom signature code	Stachelhaus code
<i>entE</i>	<i>Escherichia coli</i> K12	SRSIYAMSSPGGALQVGGAAQQVFGMAEGLVNYTR	AMPAQGVVNK
<i>dhbE</i>	<i>Bacillus subtilis</i> <i>subsp.</i> 168	SRSLYPLSSPGGALQVGGAAQQVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Corallocooccus sp.</i> MCy9049	SASLFPLSSPGGALQVGGAAQQVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Archangium gephyra</i> DSM 2261	SASLFPLSSPGGALQVGGAAQQVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Archangium sp.</i> Cb G35	SGSLFPLSSPGGALQVGGAAQQVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Corallocooccus coralloides</i> DSM 2259	SASLFPMSSPGGALQVGGAAQQVFGMAEGLVNYTR	PMPAQGVVNK
<i>mxxE</i>	<i>Corallocooccus sp.</i> H22C18031201	SASLFPLSSPGGALQVGGAAQQVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Cystobacter fuscus</i> DSM 2262	SGSLFPLSSPGGALQVGGAAQQVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Cystobacter violaceus</i> Cb vi76	SGSLFPLSSPGGALQVGGAAQQVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Haliangium ochraceum</i> DSM 14365	SASLFTLSSPGGALQVGGAAQQVFGMAEGLVNYTR	TLPAQGVVNK
<i>mxxE</i>	<i>Hyalangium minutum</i> DSM 14724	SASLFPMSSPGGALQVGGAAQQVYVFGMAEGLVNYTR	PMPAQGVVNK
<i>mxxE</i>	<i>Myxococcus fulvus</i> 124B02	SASLFPMSSPGGALQVGGAAQQVFGMAEGLVNYTR	PMPAQGVVNK
<i>mxxE</i>	<i>Myxococcus fulvus</i> HW-1	SGSLFPLSSPGGAVQVGGAAQQVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Myxococcus stipitatus</i> DSM 14675	SASLFPMSSPGGALQVGGAAQQVFGMAEGLVNYTR	PMPAQGVVNK
<i>mxxE</i>	<i>Myxococcus xanthus</i> DK 1622	SGSLFPLSSPGGAVQVGGAAQQVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Polyangium fumosum</i> DSM 14668	SASLFTLSSPGGSVQVGGSSQQVFGMAEGLVNYTR	TLPSQGVVNK
<i>mxxE</i>	<i>Polyangium solediatum</i> DSM 14670	SASLFTLSSPGGSVQVGGSSQQVFGMAEGLVNYTR	TLPSQGVVNK
<i>mxxE</i>	<i>Polyangium sp.</i> SDU3-1	SASLFTLSSPGGAVQVGGAAQQVFGMAEGLVNYTR	TLPAQGVVNK
<i>mxxE</i>	<i>Sandaracinus amylolyticus</i> DSM 53668	SASLFPLSSPGGALQVGGAAQQVFGMAEGLVCYTR	PLPAQGVVCK
<i>mxxE</i>	<i>Stigmatella aurantiaca</i> DW4/3-1	SGSLFPLSSPGGALQVGGAAQQVYVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Stigmatella aurantiaca</i> S15	SGSLFPLSSPGGALQVGGAAQQVYVFGMAEGLVNYTR	PLPAQGVVNK
<i>mxxE</i>	<i>Vitiosangium sp.</i> GDMCC 1.1324	SASLFPMSSPGGALQVGGAAQQVFGMAEGLVNYTR	PMPAQGVVNK



**Figure S73.** Consensus sequences for 8 angstrom signature code (upper) and Stachelhaus code (lower) of myxobacterial mxcE translations.

### S 4.3 Assessment of Biological Activities

**Table S3.** Minimum inhibitory concentrations (MICs) for the synthetic myxochelins against a panel of microorganisms and MICs of reference drugs in combination with **9a** (128  $\mu\text{g}/\text{mL}$ ). -: not determined.

compound	MIC [ $\mu\text{g}/\text{mL}$ ]				
	<i>S. aureus</i> str. Newman	<i>E. coli</i> BW25113	<i>E. coli</i> K12 $\Delta\text{tolC}$	<i>C. albicans</i> DSM 1665	<i>M. hiemalis</i> DSM 2656
<b>9a</b>	> 128	> 128	> 128	> 128	> 128
<b>9b</b>	> 128	> 128	> 128	> 128	> 128
<b>9c</b>	> 128	> 128	> 128	> 128	> 128
<b>9d</b>	> 128	> 128	> 128	> 128	> 128
<b>9e</b>	> 128	> 128	> 128	> 128	> 128
<b>9f</b>	> 128	> 128	> 128	> 128	> 128
<b>9g</b>	> 128	> 128	> 128	> 128	> 128
<b>9h</b>	> 128	> 128	> 128	> 128	> 128
<b>9i</b>	> 128	> 128	> 128	> 128	> 128
Ciprofloxacin	0.125	-	-	-	-
Linezolid	2	-	-	-	-
Gentamicin	0.5	-	-	-	-
Daptomycin	1	-	-	-	-
Ciprofloxacin + <b>9a</b>	0.125	-	-	-	-
Linezolid + <b>9a</b>	2	-	-	-	-
Gentamicin + <b>9a</b>	0.5	-	-	-	-
Daptomycin + <b>9a</b>	1	-	-	-	-

## S 4.4 References

1. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733-45, doi:10.1093/nar/gkv1189.
2. Röttig, M.; Medema, M.H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **2011**, *39*, W362-W367, doi:10.1093/nar/gkr323.

## Chapter 5

# SAR Studies and Biological Characterization of Sandacrabin Derivatives as Antiviral RNA Polymerase Inhibitors

# Contributions and Acknowledgments

## Author's effort:

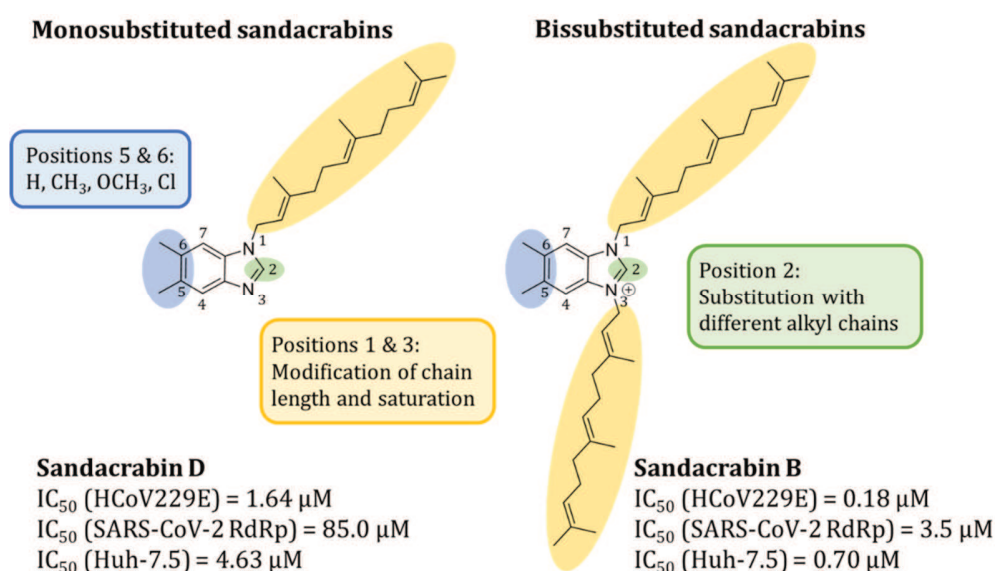
The author significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The laboratory and *in silico* work regarding synthesis and purification of mono- and bisubstituted sandacrabins derivatives were performed by the author. Furthermore, the author contributed significantly to conceiving and writing this chapter.

## Contributions by others:

Chantal Bader contributed to conception of this study, performed synthesis, purification and structure confirmation and contributed to conceiving, writing and editing of the manuscript. Alexander Voltz performed synthesis, purification and structure confirmation. Matthias Götte and Egor Tchesnokov contributed through inhibition assays against the Lassa virus and SARS-CoV-2 RdRp complex. Ralf Bartenschlager and Heeyoung Kim provided antiviral activities against DENV and ZIKV and cytotoxicity against Huh7/Lunet T7 cells. Ravindra Jumde contributed through docking studies of sandacrabins A-E and DEREK analysis. Andreas Kany performed the preliminary ADMET profiling. Rolf Müller contributed by supervision of the project and conceiving, editing and proofreading of this chapter.

## 5.1 Abstract

The sandacrabins are a recently discovered class of farnesylated 5,6-dimethylbenzimidazoles isolated from the myxobacterium *Sandaracinus defensii* MSr10575. These natural products are inhibitors of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA-dependent RNA polymerase (RdRp) complex, consisting of the nonstructural proteins Nsp7, Nsp8 and Nsp12 (RdRp), with promising  $IC_{50}$  values in low micromolar range. Furthermore, they exhibit promising antiviral activity against the human corona virus HCoV229E with comparably weak cytotoxicity against several cell lines, which qualifies them for further drug development. In this work, we conducted a computational docking study revealing that the sandacrabins are most likely binders of an allosteric site of Nsp12. We evaluated the broad spectrum antiviral activity and acquired preliminary data on metabolic stability and toxicity of the five natural sandacrabins. Finally, we synthesized 61 novel sandacrabin congeners featuring modifications at position 2, at the two nitrogen atoms in positions 1 and 3 as well as at positions 5 and 6 of the benzimidazole core scaffold. While the newly synthesized derivatives are currently under investigation for their inhibition of the SARS-CoV-2 RdRp complex, antiviral activity and safety profile, the aim of this study is to develop improved sandacrabin derivatives with increased bioactivity and reduced cytotoxicity and gain further insights into structure-activity relationship of this natural product class. The study presented in this chapter represents a good entry point into future drug development of the sandacrabins, which could ultimately reach clinical trials.



**Graphical abstract.** The naturally occurring sandacrabins D and B show promising  $IC_{50}$  values against the human pathogenic corona virus HCoV229E and the SARS-CoV-2 RNA-dependent RNA polymerase complex [1]. The derivatives synthesized in this study feature modifications in positions 5 and 6 (blue), 2 (green), as well as 1 and 3 (yellow) aiming to improve antiviral activity while reducing cytotoxicity.

## 5.2 Introduction

Historically, natural products (NPs) have been the most successful source of potential drug leads as they provide unprecedented structural diversity in combination with intriguing biological activities [2,3]. In treatment of infections caused by bacteria, viruses, fungi and parasites as well as in cancer treatment, bacterial NPs have proven to be particularly advantageous and therefore provided many natural product based drug leads [4,5]. Bacteriophages are the most abundant viruses on the planet, co-existing with bacteria for billions of years, which lead to the evolution of various antiviral defense mechanisms [6,7]. As many genes encoding for key proteins in viral replication, morphogenesis and capsid formation are shared by numerous RNA and DNA viruses, bacterial antiviral NPs are also considered a rich source of novel lead compounds against human pathogenic viruses [6,8]. Myxobacteria, an order of Gram—negative  $\delta$ -proteobacteria, represent an especially prolific source for structurally diverse NPs, including several candidate compounds with promising antiviral activity [9–11], such as the broad-spectrum antiviral soraphens that effectively inhibit Hepatitis C Virus (HCV) and Human Immunodeficiency Virus (HIV) replication [12,13]. The sandacrabins are a class of recently discovered farnesylated dimethylbenzimidazoles isolated from the terrestrial myxobacterium *Sandaracinus defensii* MSr10575 that exhibit promising antiviral activities against the human pathogenic coronavirus HCoV229E [1]. The sandacrabins were additionally found to inhibit the RNA-dependent RNA polymerase (RdRp) complex of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with  $IC_{50}$  values in the low micromolar range. This finding revealed the inhibition of viral RNA synthesis as a mode of action although the discrepancy between the reported antiviral activity and the significantly lower inhibition of the SARS-CoV-2 RdRp complex indicates the presence of additional targets. Interestingly however, the sandacrabins show only weak cytotoxicity, opening an application window represented by an up to 11-fold difference in cytotoxic and antiviral activities as reported for sandacrabins C. Their intriguing biological activities combined with a straight forward total synthesis routes for the naturally occurring sandacrabins make these compounds a valuable starting point for drug development. In the work described in this chapter, we adapted the total synthesis routes described by Bader *et al.* [1] to generate 61 synthetic derivatives aiming to reduce cytotoxicity while sustaining potency as well as ADMET properties and potentially improving target selectivity. Like their naturally occurring congeners, the synthetic compounds are divided into two series featuring mono- and bisubstituted benzimidazoles. The study described in this chapter targets to evaluate the influence of variations of the farnesyl side chains as well as modifications of the benzimidazole scaffold at positions 5 and 6. Furthermore, we substituted position 2 of the benzimidazole core with different alkyl chains in order to determine the influence on the weak antimicrobial activity



against *B. subtilis* that was reported for sandacrabins A [1]. In the course of this study, all synthetic derivatives will be tested for activity against *B. subtilis*, Huh-7.5 and HepG2 cells and tested in the HCoV229E assay to evaluate their antiviral activity. Most promising hits derived from these screenings will then be tested for their *in vitro* inhibition of the SARS-CoV2 RdRp complex.

## 5.3 Results

### 5.3.1 Evaluation of Broad-spectrum Antiviral Activity

In previous screenings, Bader *et al.* detected a significant reduction in viral replication of the human pathogenic corona virus HCoV229E for the sandacrabins B-E with IC<sub>50</sub> values in the low micromolar range. Furthermore, they presented that the sandacrabins B-E inhibit the SARS-CoV-2 RdRp complex revealing inhibition of viral RNA synthesis as a plausible target. Cytotoxic activity against Huh-7.5 and U-2 OS cells appeared to be 3-11 fold lower than antiviral activity against HCoV229E, opening a potential application window. Additionally, they determined minimal inhibitory concentrations (MICs) against a broad panel of Gram-positive and Gram-negative bacteria as well as fungal pathogens revealing a weak antimicrobial activity of sandacrabins A against *Bacillus subtilis* and *Staphylococcus aureus* [1].

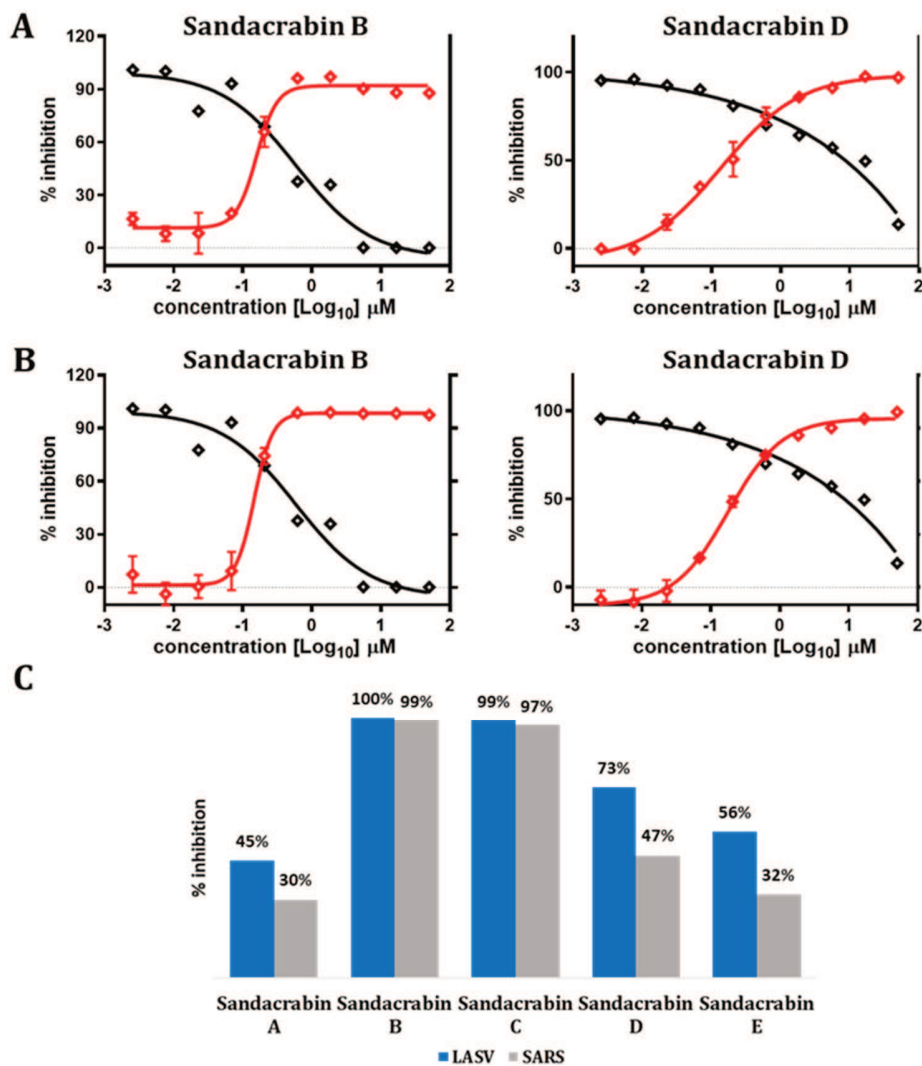
To further elaborate on the antiviral results and to determine the broad-spectrum activity of the sandacrabins, the antiviral activity against Dengue virus DENV-R2A and Zika virus ZIKV-R2A were assessed in an *in vitro* infection model. As already observed previously for HCoV229E, sandacrabins A was found to possess only very weak antiviral bioactivity. Interestingly though, sandacrabins B-E caused a strong reduction in viral replication, with sandacrabins B and D being particularly active with IC<sub>50</sub> values ranging from 0.14-0.18 μM against both flaviviruses. Since the bisfarnesylated sandacrabins B and C also exhibited strong cytotoxic activities of 0.59-1.00 μM against the Huh7/Lunet T7 host cells, the antiviral activities observed might however be aberrated. The monofarnesylated congeners D and E appeared to be 10-fold less cytotoxic, opening a window for further development. The detected cytotoxicity against Huh7/Lunet T7 cells aligns well with the antiproliferative activities previously observed for U-2 OS cells [1].

Additionally, inhibition of the Lassa virus RdRp complex was assessed *in vitro* as reduction in RNA synthesis products corresponding to the full template length and compared to the inhibition of the SARS-CoV-2 RdRp complex. At a concentration of 100 μM, all sandacrabins showed similar or slightly higher inhibition of the Lassa virus than the SARS-CoV-2 RdRp complex with sandacrabins B and C being the most active derivatives. The biological activity results are summarized in Table 1, Figure 1 and the supporting information.

**Table 1.** Antiviral activities of the sandacrabins. Activities against HCoV229E have previously been reported in [1].

Test organism	Sandacrabin A	Sandacrabin B	Sandacrabin C	Sandacrabin D	Sandacrabin E	Positive control
HCoV229E	> 10	0.18	0.34	1.64	1.91	0.0056 (Remdesivir)
DENV-R2A	> 10	0.16*	0.50*	0.14	0.80	2.33 (Ribavirin)
ZIKV-R2A	> 10	0.15*	0.32*	0.18	0.92	2.51 (Ribavirin)

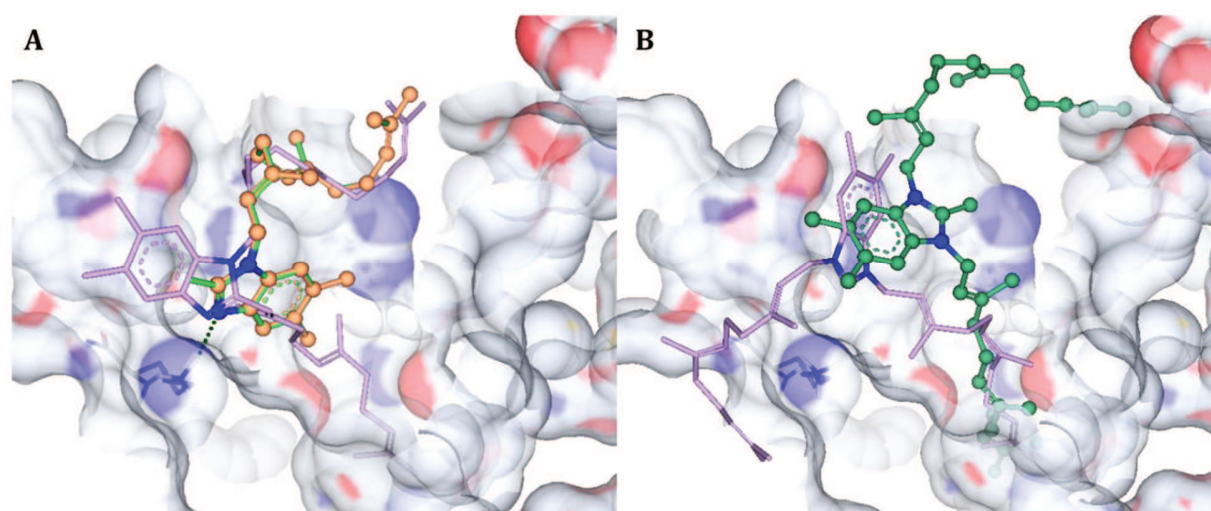
\*potentially aberrated bioactivities due to cytotoxicity against the host cells



**Figure 1.** Antiviral activities of sandacrabins B and D against dengue virus DENV-R2A (A) and zika virus ZIKV-R2A (B) displayed as reduction in viral replication (red trace, % inhibition) with simultaneous determination of the cell viability of the Huh7/Lunet T7 host cells (black trace, % viability). (C) Sandacrabin-dependent inhibition of the LASV (blue) and SARS-CoV-2 (gray) RdRp complex as reduction in RNA synthesis products corresponding to the full template length at a concentration of 100 μM.

### 5.3.2 *In Silico* Docking Studies with the SARS-CoV-2 RdRp Complex

In order to explore the putative binding site of the sandacrabins in the SARS-CoV-2 RdRp complex, a computational docking study was performed using SeeSAR [14]. Initially, DoGSiteScorer [15] was used to detect different putative binding sites of the energy minimized structure of the SARS-CoV-2 RdRp complex (PDB: 7BV2 [16]) and to subsequently predict their druggability using calculated geometric and physico-chemical properties. The two highest-scoring binding sites — the active site and one putative allosteric site — were selected and the orientation and binding affinity of the sandacrabins A-E were evaluated.

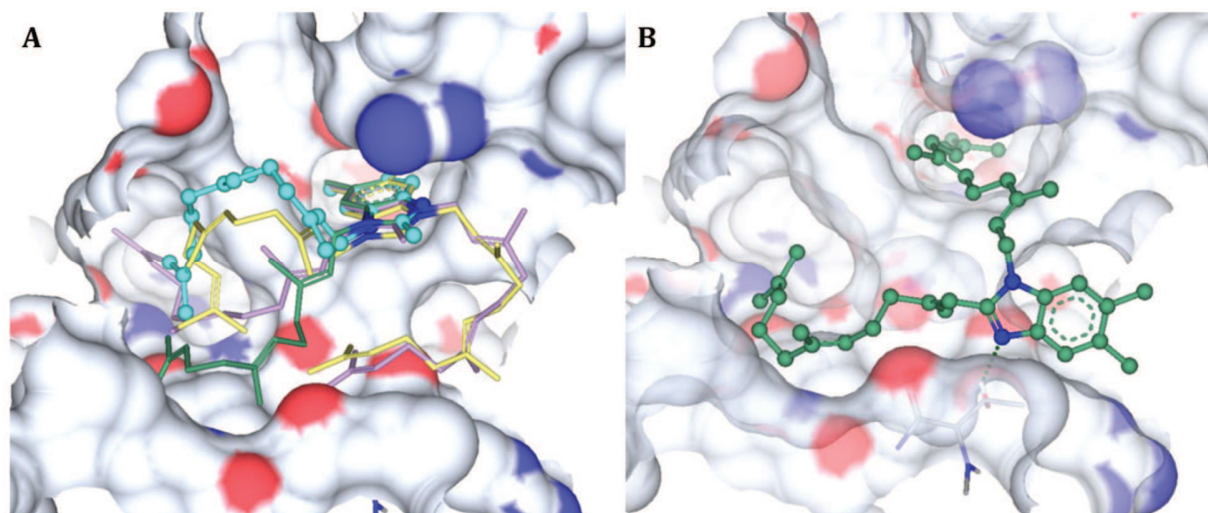


**Figure 2.** Comparative orientations of (A) sandacrabin A (violet), D (brown) and E (green) and (B) sandacrabin B (violet) and C (green) with calculated highest binding affinity in the active site of SARS-CoV2 Nsp12 (PDB: 7BV2 [16]).

Despite high structural similarity between the compounds, the predicted orientation and conformation of the sandacrabins in the active site differ substantially from each other (Figure 2). While sandacrabins D and E presumably bind identically in the active site, sandacrabins B and C – also differing by just a methyl group – are differently oriented in the active site. Like sandacrabin A, sandacrabins D and E interact with residue Lys545, while sandacrabins B and C lack this interaction. Besides the differences in orientation and conformation, the calculated binding affinities for all derivatives are located in a high micromolar to millimolar range, which is significantly lower than the  $IC_{50}$  values previously reported by Bader *et al.* [1].

When the sandacrabins were, however, docked into the allosteric binding site of the RdRp complex, the orientation of the sandacrabins B-E was predicted to be very similar with the benzimidazole moiety positioned identically (Figure 3). Only sandacrabin A acquires a distinctly different orientation than the other derivatives due to an additional interaction with Thr395. Overall, with calculated binding affinities from high nanomolar to micromolar range, all sandacrabins have higher binding affinities in the allosteric

site compared to the active site that also fit well to the previously determined  $IC_{50}$  values reported by Bader *et al.* [1]. Needless to say, the predicted binding modes and calculated binding affinities still need to be confirmed experimentally by three-dimensional electron or X-ray crystallography and binding assays such as surface plasmon resonance (SPR) spectroscopy or isothermal titration calorimetry (ITC).



**Figure 3.** Comparative orientations of (A) sandacrabin B (yellow), C (violet), D (green) and E (blue) and (B) sandacrabin A (green) with calculated highest binding affinity in the allosteric site of SARS-CoV2 Nsp12 (PDB: 7BV2 [16]).

### 5.3.3 Assessment of Metabolic Stability and Toxicity Prediction

Besides pharmaceutical activity and selectivity, the so-called ADMET (absorption, distribution, metabolism, excretion and toxicity) properties are of high importance for lead development [17]. For this reason, sandacrabins C and E were chosen for ADMET profiling to represent the most promising mono- and bisfarnesylated derivative, respectively. While the assay is still ongoing, preliminary data already shows that both compounds are stable in mouse and human plasma over 240 minutes without any observable degradation. Both tested derivatives are metabolized by mouse liver S9 fractions, while sandacrabin E is metabolized faster than sandacrabin C. Experiments conducted with human liver S9 fractions show increased half-life in comparison to mouse liver S9 fraction for sandacrabin E. Metabolic stability determinations of sandacrabin C have to be repeated due to insufficient extraction with standard solvents. The preliminary data is summarized in Figure 3 and Table 1. Toxicity predictions were carried out using the expert knowledge-based DEREK Nexus platform. The sandacrabins were predicted to be non-mutagenic, but they might be skin sensitizing. No conclusion could be drawn about additional 53 toxicity hazards, where no evidence for or against was found for similar compounds in the database (Table S3), thus the origin of the observed cytotoxicity of the sandacrabins remains open.

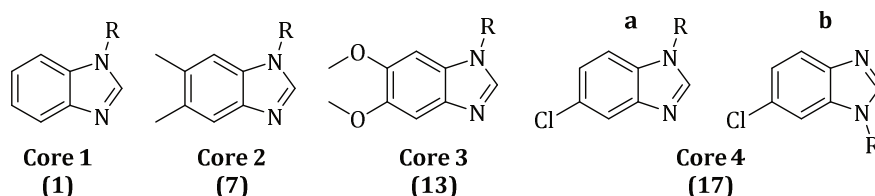
**Table 2.** Preliminary ADMET properties of sandacrabins C and E.

	Sandacrabin C	Sandacrabin E
Mouse Plasma $t_{1/2}$ [min]	> 240	> 240
Mouse Plasma % after 2.5h	111.3 ± 19.3	91.8 ± 16.9
Human Plasma $t_{1/2}$ [min]	> 240	> 240
Human Plasma % after 2.5h	98.1 ± 37.7	97.1 ± 31.5
Mouse Liver S9 Fraction $t_{1/2}$ [min]	21.9 ± 4.2	2.9 ± 0.7
Mouse Liver S9 Fraction $Cl_{int}$ [ $\mu$ l/mg/min]	32.5 ± 5.6	247 ± 54
Human Liver S9 Fraction $t_{1/2}$ [min]	<i>n.d.</i> *	13.7 ± 2.1
Human Liver S9 Fraction $Cl_{int}$ [ $\mu$ l/mg/min]	<i>n.d.</i> *	51.5 ± 7.8

\* Human Liver S9 Fraction  $t_{1/2}$  and  $Cl_{int}$  could not be calculated due to low signals with high deviation.

### 5.3.4 Modifications of the Benzimidazole Core at Positions 5 and 6

The first series of sandacrabin analogs focused on modifications of the benzimidazole core scaffold at positions 5 and 6, utilizing commercially available benzimidazole cores 1-4 (**1**, **7**, **13**, **17**). Substitution of the respective core scaffold was achieved with general procedure B (Section 5.5.6.2) using farnesyl bromide, dodecyl bromide, geranyl bromide, prenyl bromide and citronellyl bromide as alkylating agents. The reaction of core 2 (**7**) with farnesyl bromide generated the previously characterized sandacrabin D (**8**) that is produced by *Sandaracinus defensii* MSr10575 [1]. Reactions of core 3 (**13**) with dodecyl bromide and prenyl bromide were unsuccessful, only yielding the desired product in trace amounts. Monosubstitution of core 4 (**17**) resulted in the generation of two regioisomers that could be separated by HPLC and are referred to as **18-22 a** and **b**, respectively.



**Figure 4.** Benzimidazole cores with different substitution at positions 5 and 6 that were used to generate the first sandacrabin derivative series.

**Table 3.** Monosubstituted sandacrabins generated in the first series.

Compound #	Core	R
2	Core 1 (1)	Farnesyl-
3	Core 1 (1)	Dodecyl-
4	Core 1 (1)	Geranyl-
5	Core 1 (1)	Prenyl-
6	Core 1 (1)	Citronellyl-
8 (= Sandacrabins D)	Core 2 (7)	Farnesyl-
9	Core 2 (7)	Dodecyl-
10	Core 2 (7)	Geranyl-
11	Core 2 (7)	Prenyl-
12	Core 2 (7)	Citronellyl-
14	Core 3 (13)	Farnesyl-
15	Core 3 (13)	Geranyl-
16	Core 3 (13)	Citronellyl-
18a	Core 4 (17)	Farnesyl-
18b	Core 4 (17)	Dodecyl-
19a	Core 4 (17)	Geranyl-
19b	Core 4 (17)	Prenyl-
20a	Core 4 (17)	Citronellyl-
20b	Core 4 (17)	Farnesyl-
21a	Core 4 (17)	Dodecyl-
21b	Core 4 (17)	Geranyl-
22a	Core 4 (17)	Prenyl-
22b	Core 4 (17)	Citronellyl-

### 5.3.5 Modifications of the Benzimidazole Core at Position 2

The second series of sandacrabins focused on modifications in position 2 of the DMB core scaffold. Additionally, commercially available benzimidazole core 5 (**23**) was evaluated. Cores 6-13 (**29**, **35**, **37**, **39**, **41**, **43**, **45**, **48**) were synthesized from 4,5-dimethylbenzene-1,2-diamine (**47**) and the respective carboxylic acid using general procedure A (Section 5.5.6.1). All cores were substituted using farnesyl bromide, and cores 5 (**23**), 6 (**29**) and 13 (**48**) were additionally substituted with the respective dodecyl, geranyl, citronellyl and prenyl bromide according to general procedure B (Section 5.5.6.2). The reaction of core 13 (**48**) with farnesyl bromide produced the previously characterized sandacrabins E (**49**) [1].

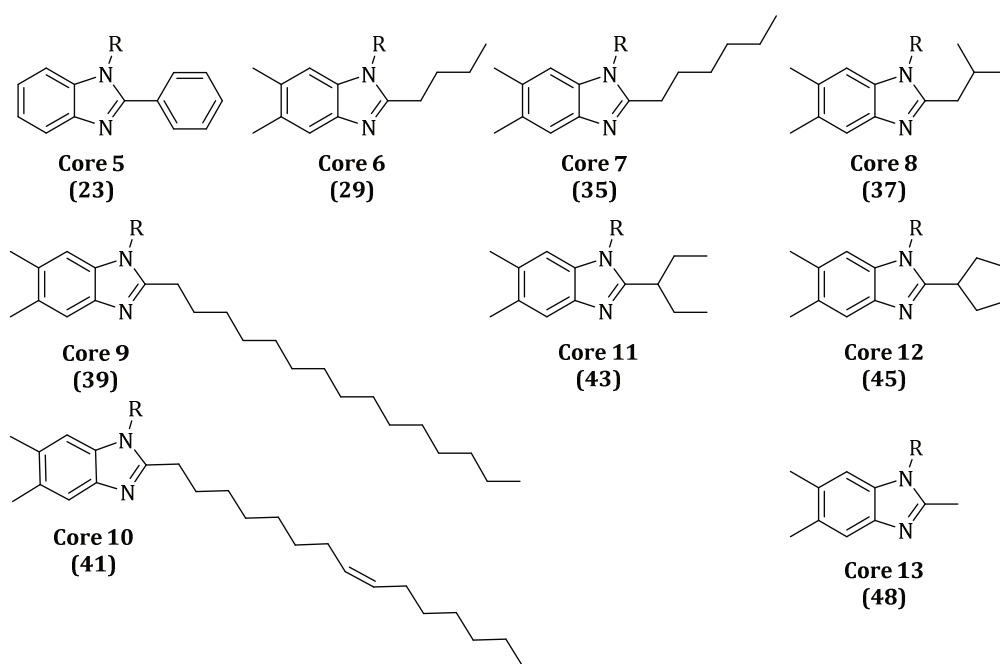


Figure 5. Benzimidazole cores with different substitution at position 2 that were used in the second series.

Table 4. Monosubstituted sandacrabins generated in the second series.

Compound #	Core	R
24	Core 5 (23)	Farnesyl-
25	Core 5 (23)	Dodecyl-
26	Core 5 (23)	Geranyl-
27	Core 5 (23)	Prenyl-
28	Core 5 (23)	Citronellyl-
30	Core 6 (29)	Farnesyl-
31	Core 6 (29)	Dodecyl-
32	Core 6 (29)	Geranyl-
33	Core 6 (29)	Prenyl-
34	Core 6 (29)	Citronellyl-
36	Core 7 (35)	Farnesyl-
38	Core 8 (37)	Farnesyl-
40	Core 9 (39)	Farnesyl-
42	Core 10 (41)	Farnesyl-
44	Core 11 (43)	Farnesyl-
46	Core 12 (45)	Farnesyl-
49 (= Sandacrabins E)	Core 13 (48)	Farnesyl-
50	Core 13 (48)	Dodecyl-

51	Core 13 (48)	Geranyl-
52	Core 13 (48)	Prenyl-
53	Core 13 (48)	Citronellyl-

### 5.3.6 Bissubstitution of the Modified Benzimidazole Cores

The third series of sandacrabins analogs covers the bissubstitution of the modified benzimidazole and DMB cores described in the first and second series. All cores were bissubstituted using farnesyl bromide and cores 2 (7) and 13 (48) were additionally substituted with the respective dodecyl, geranyl, citronellyl and prenyl bromide according to general procedure C (Section 5.5.6.3) or D (Section 5.5.6.4). This series also features the synthesis of the two previously described compounds sandacrabins B (55) and C (70) by farnesylation of core 2 (7) and 13 (48) respectively. Bisfarnesylated core 3 (60) was already obtained as a side product from the monosubstitution reaction as described in section 5.3.3 and was thus not resynthesized in this series.

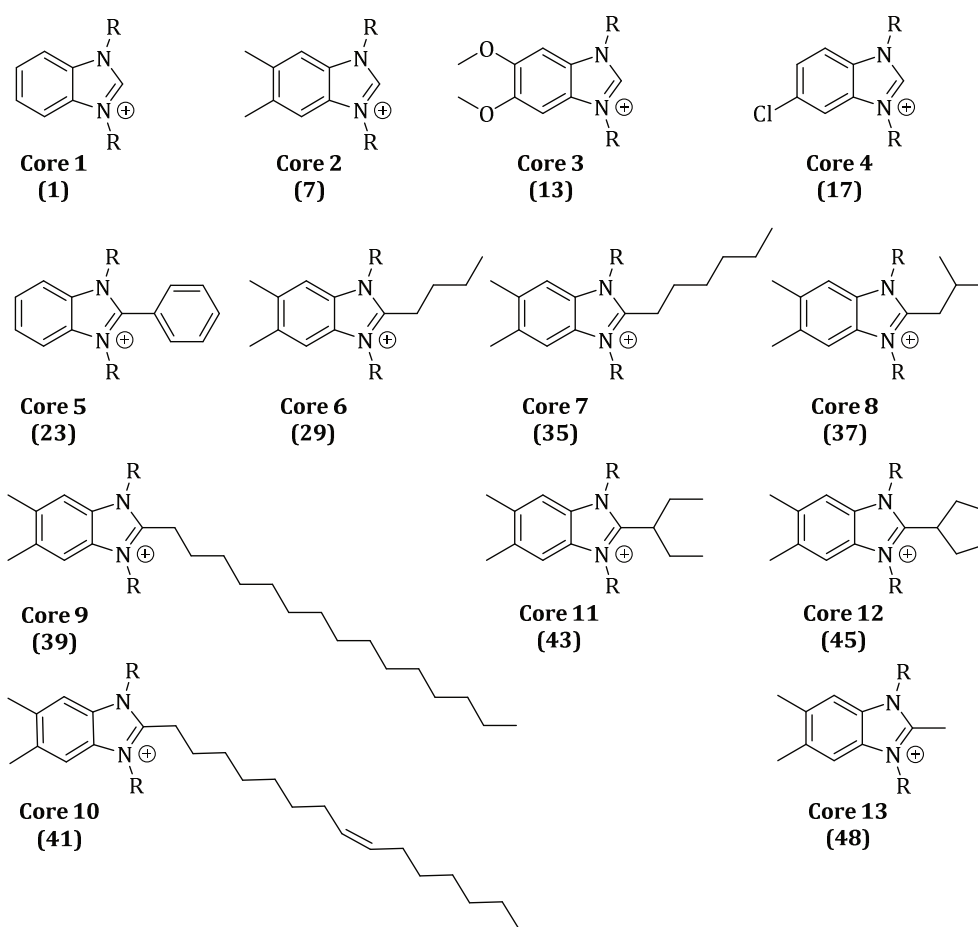


Figure 6. Benzimidazole cores that were bissubstituted in the third series.



Table 5. Bissubstituted sandacrabin analogs generated in the third series.

Compound #	Core	R
54*	Core 1 (1)	Farnesyl-
55 (= Sandacrabin B)	Core 2 (23)	Farnesyl-
56	Core 2 (23)	Dodecyl-
57*	Core 2 (23)	Geranyl-
58*	Core 2 (23)	Prenyl-
59*	Core 2 (29)	Citronellyl-
60*	Core 3 (29)	Farnesyl-
61*	Core 4 (29)	Farnesyl-
62*	Core 5 (29)	Farnesyl-
63*	Core 6 (29)	Farnesyl-
64*	Core 7 (35)	Farnesyl-
65*	Core 8 (37)	Farnesyl-
66	Core 9 (39)	Farnesyl-
67*	Core 10 (41)	Farnesyl-
68*	Core 11 (43)	Farnesyl-
69*	Core 12 (45)	Farnesyl-
70 (= Sandacrabin C)	Core 13 (48)	Farnesyl-
71*	Core 13 (48)	Dodecyl-
72*	Core 13 (48)	Geranyl-
73*	Core 13 (48)	Prenyl-
74*	Core 13 (48)	Citronellyl-

\*confirmed by high-resolution LC-MS; NMR in progress

## 5.4 Discussion and Outlook

In the course of the project described in this chapter, a total of 65 sandacrabin derivatives, including the previously described sandacrabins B-E, was synthesized. Most of the compounds in this ongoing project were already purified and characterized by MS and NMR, however several representatives – especially of the bissubstituted derivative in series 3 – are still in the process of compound purification. The positions that were modified were mainly those that also vary in the naturally occurring representatives, namely position 2 as well as substitution of the two nitrogen atoms in positions 1 and 3. Furthermore, positions 5 and 6 were selected, which carry methyl groups in the natural products.

The docking studies revealed that the sandacrabins are most likely interacting with an allosteric binding site of the SARS-CoV-2 RdRp complex rather than the active site. This conclusion is based on the uniform orientation of the bioactive molecules in the binding site, as well as the significantly better predicted binding affinities compared to the active site. While sandacrabins B-E, it only shows weak RdRp inhibition in previously conducted assays. One explanation for this finding might be its different orientation in the allosteric site, resulting in less conformational changes to the protein retaining its function. This however would need to be confirmed by crystallization experiments, as computational docking studies only give a rough estimation about the binding affinity and are far from accurate. In order to confirm the predicted binding position and orientation of the sandacrabins, the most active congener from previous RdRp inhibition assays [1], sandacrabins C, was selected for cryogenic electron microscopy (cryo-EM) structure determination in complex with the SARS-CoV-2 RdRp complex, which is work in progress.

Besides potency and selectivity, knowledge of the metabolic stability of new compounds is a prerequisite for predicting pharmacokinetic parameters that are further used to eliminate unfavorable drug candidates and pave the way for subsequent *in vivo* studies [18]. The sandacrabins were found to be stable in mouse and human plasma over 240 min without observable degradation and are metabolized by mouse and human liver S9 fraction. While these results are very promising considering a future application in human medicine, there are nevertheless some constraints remaining. First, the naturally occurring as well as the newly synthesized derivatives are highly lipophilic compounds with poor aqueous solubility, which might restrict oral bioavailability or require tailored formulation aids such as cyclodextrins [19,20]. Therefore, it would be desirable if the antiviral activity of the less lipophilic derivatives such as geranyl and prenyl substituted sandacrabins – especially from the monosubstituted series – would be retained. The main drawback of the sandacrabins however is their cytotoxicity against several human cell lines. The knowledge-based toxicity prediction by DEREK analysis could unfortunately not provide an apparent rationale for the observed toxicity. As toxicology and clinical safety are among the major causes of attrition in clinical studies accounting for approximately 30% of failures, it is particularly important to examine the upcoming cytotoxicity data against human liver cancer cell lines HepG2 and Huh-7.5 for a putative structure-cytotoxicity relationship and to evaluate potential off-target effects such as the inhibition of human RNA polymerases [21,22].

Because the synthetic compounds presented in this chapter have not yet been evaluated for their bioactivity against the human pathogenic corona virus HCoV229E or any of the other viral pathogens presented in this chapter, it is not yet possible to make a statement about a structure-activity relationship. The most active hit compounds against HCoV229E will be further tested for their *in vitro* inhibition of the

SARS-CoV-2 RdRp complex. Besides the antiviral activity, bioactivity is also evaluated against *Bacillus subtilis*, where sandacrabins A showed a mediocre MIC of 32 µg/mL [1]. The aim of this study is to develop improved sandacrabins derivatives with increased bioactivity against HCoV229E, the SARS-CoV-2 RdRp complex and *Bacillus subtilis*, while simultaneously reducing cytotoxicity. In particular, the comparison of the mono- and bisubstituted derivatives regarding their antiviral and cytotoxic activities is of special interest. Among the natural occurring derivatives, sandacrabins B and C had better antiviral activity against HCoV229E, but also increased cytotoxicity than the monosubstituted congeners D and E. In the Zika and Dengue virus *in vitro* infection model, sandacrabins D and E were the most promising derivatives as they showed strong antiviral bioactivities combined with lower cytotoxicity than the bisubstituted compounds.

In fact, many quaternary ammonium compounds (QACs) such as the bisubstituted sandacrabins are known to exhibit broad antimicrobial and cytotoxic activities mediated through cell membrane disruption [23]. However, the lack of any antimicrobial activity below 64 µg/mL in initial screenings – especially against fungal pathogens – suggests this might only occur in very high concentrations of the sandacrabins. Additional experiments such as hemolysis assays using human erythrocytes could be used to further investigate putative effects of the bisubstituted sandacrabins on the integrity of cell membranes. Either way, future experiments will be needed to uncover the mode of cytotoxic activity and to design compounds that retain their promising broad-spectrum antiviral activity, while eliminating the cytotoxicity.

Summing up, the fairly easy and fast routes for sandacrabins synthesis in combination with their promising antiviral activity and metabolic stability qualify the sandacrabins for further screenings and raises hope for the development of new lead structures with improved pharmacodynamic and pharmacokinetic properties.

## 5.5 Materials and Methods

### 5.5.1 UHPLC-hrMS Measurements

UHPLC-qTOF measurements were carried out on a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Bruker maXis 4G ultra-high-resolution quadrupole time-of-flight (UHR-qTOF) MS equipped with a high-resolution electrospray ionization (HRESI) source (Bruker Daltonics, Billerica, MA, USA). The separation of 1 µL sample was achieved with a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH<sub>2</sub>O with 0.1% formic acid on an ACQUITY BEH C18 column (50 mm × 2.1 mm, 1.7 µm d<sub>p</sub>) (Waters, Eschborn, Germany) equipped

with a Waters VanGuard BEH C18 1.7  $\mu\text{m}$  guard column at a flow rate of 0.6 mL/min and 45  $^{\circ}\text{C}$  for 6 min (monosubstituted compounds and core scaffolds) or 9 min (bissubstituted compounds) with detection by a diode array detector at 200–600 nm as well as by mass spectrometry. The LC flow was split into 75  $\mu\text{L}/\text{min}$  before entering the mass spectrometer. Mass spectrograms were acquired in centroid mode ranging from 150–2500  $m/z$  at an acquisition rate of 2 Hz in positive MS mode. Source parameters were set to 500 V end-plate offset; 4000 V capillary voltage; 1 bar nebulizer gas pressure; 5 L/min dry gas flow; and 200  $^{\circ}\text{C}$  dry gas temperature. Ion transfer and quadrupole parameters were set to 350  $V_{\text{pp}}$  funnel RF; 400  $V_{\text{pp}}$  multipole RF; 5 eV ion energy; and 120  $m/z$  low-mass cut-off. Collision cell was set to 5.0 eV and pre-pulse storage was set to 5  $\mu\text{s}$ . Calibration was conducted automatically before every HPLC-MS run by the injection of sodium formate and calibration on the respective clusters formed in the ESI source. All MS analyses were acquired in the presence of the lock masses  $\text{C}_{12}\text{H}_{19}\text{F}_{12}\text{N}_3\text{O}_6\text{P}_3$ ,  $\text{C}_{18}\text{H}_{19}\text{F}_{24}\text{N}_3\text{O}_6\text{P}_3$  and  $\text{C}_{24}\text{H}_{19}\text{F}_{36}\text{N}_3\text{O}_6\text{P}_3$ , which generate the  $[\text{M} + \text{H}]^+$  ions of 622.0289, 922.0098 and 1221.9906. The HPLC-MS system was operated by HyStar 5.1 (Bruker Daltonics, Billerica, MA, USA), and LC chromatograms as well as UV spectra and mass spectrograms were analyzed with DataAnalysis 5.3 (Bruker Daltonics, Billerica, MA, USA).

### 5.5.2 NMR Measurements

NMR spectra were recorded on a Bruker UltraShield 500 or a Bruker Ascend 700 spectrometer equipped with a 5 mm TCI cryoprobe ( $^1\text{H}$  at 500 and 700 MHz,  $^{13}\text{C}$  at 125 and 175 MHz). All observed chemical shift values ( $\delta$ ) are given in ppm and coupling constant values ( $J$ ) in Hz. Standard pulse programs were used for HMBC, HSQC, and gCOSY experiments. HSQC experiments were optimized for  $1J_{\text{C-H}} = 145$  Hz and HMBC experiments were optimized for  $^{2,3}J_{\text{C-H}} = 6$  Hz. For the spectra recorded in methanol- $d_4$  the chemical shifts of the solvent signals at  $\delta_{\text{H}}$  3.31 ppm and  $\delta_{\text{C}}$  49.2 ppm and for measurements in chloroform- $d$  the solvent signals at  $\delta_{\text{H}}$  7.24 ppm and  $\delta_{\text{C}}$  77.2 ppm were used as reference signals for spectra calibration.

### 5.5.3 Computational Docking Studies

For modelling studies, the energy minimized X-ray crystal structure of the SARS-CoV-2 Nsp12-Nsp7-Nsp8 complex bound to the template-primer RNA and triphosphate form of Remdesivir (PDB: 7BV2, [16]) was used. The DoGSiteScorer module implemented in SeeSAR 9.1 (BiosolveIT GmbH, Sankt Augustin, Germany) was used to identify different putative binding sites, calculate geometric and physico-chemical properties and to subsequently predict their druggability. According to their druggability score, the two highest-scoring binding sites – the active site and one allosteric site – were selected for further docking

studies. The active site of Nsp12 (RdRp complex) was defined by selecting residues 545-548, 553-558, 616-624, 680-693 and 757-762. The allosteric site was defined by selecting residues 172, 173, 176, 243, 246, 247, 249, 315, 318, 319, 350, 394-396, 456, 457, 459-462, 627-629, 675-677, 788 and 791. The sandacrabins A-E were used as ligands for docking using the FlexX docking module implemented in SeeSAR 9.1 (BiosolveIT GmbH, Sankt Augustin, Germany). The top two solutions were retained and subsequently post-scored using the HYDE module implemented in SeeSAR, while the maximal number of poses was set to 30. After manual visualization and exclusion of poses with unfavorable conformations or significant inter- or intramolecular clash terms, the resulting poses were ranked according to their predicted binding energies. The top-ranked pose from each molecule was used for comparison and generation of the figures presented in the main text.

## 5.5.4 Biological Assays

### 5.5.4.1 Antiviral Activity

Antiviral activity against Dengue virus DENV-R2A and Zika virus ZIKV-R2A were assessed according to the procedure described by Fischl and Bartenschlager [24,25]. Huh-7 /Lunet T7 cells constitutively expressing a firefly luciferase reporter gene were infected with DENV-R2A or ZIKA-R2A, both of which are Renilla luciferase reporter viruses. 48 h after the inoculation, cells were lysed in PBS/1% Triton-X100. Firefly luciferase activity was measured as surrogate marker for cell number and viability and Renilla luciferase activity was used as marker for infection efficiency and viral replication. The data were normalized to solvent treated controls. Ribavirin was used as a positive control.

### 5.5.4.2 RNA Synthesis Inhibition Assay

RNA synthesis inhibition was assessed using the procedure described by Gordon, Lee, Tchesnokov *et al.* [26]. The assay consisted of mixing (final concentrations) 0.02  $\mu\text{M}$  of Lassa virus RNA-directed RNA polymerase L (LASV L protein) or 0.1  $\mu\text{M}$  of SARS-CoV-2 RdRp complex, 100  $\mu\text{M}$  of the respective compound, 25 mM Tris-HCl pH 8, 200  $\mu\text{M}$  RNA primer, 2  $\mu\text{M}$  RNA template, 0.1  $\mu\text{M}$  [ $\alpha$ - $^{32}\text{P}$ ]-GTP, and 0.1  $\mu\text{M}$  NTP on ice. The reaction mixtures were incubated at 30  $^{\circ}\text{C}$  for 10 min, followed by the addition of 5 mM  $\text{MgCl}_2$  to initiate nucleotide incorporation. The reactions (15  $\mu\text{L}$ ) were incubated for 30 min and then stopped by the addition of 15  $\mu\text{L}$  formamide/EDTA (50 mM) mixture and incubated at 95  $^{\circ}\text{C}$  for 10 min. The reaction products were resolved through 20% PAGE and the [ $\alpha$ - $^{32}\text{P}$ ]-generated signal was scanned through phosphorimaging (Amersham Typhoon 5, Cytivia, Marlborough, MA, USA). The signal corresponding to the full-template length product of RNA-synthesis in the presence of a compound was

normalized (%) to the signal obtained in absence of the compound. The data were analysed using GraphPad Prism 7.0 (GraphPad Software Inc, San Diego, CA, USA).

### 5.5.5 Assessment of Metabolic Stability

#### 5.5.5.1 Metabolic Stability Tests in Human Liver S9 Fraction

For the evaluation of combined phase I and phase II metabolic stability, the compound (1  $\mu\text{M}$ ) was incubated with 1 mg/mL pooled liver S9 fraction (Corning Life Sciences B.V., Amsterdam, The Netherlands), 2 mM NADPH, 1 mM UDPGA, 10 mM  $\text{MgCl}_2$ , 5 mM GSH and 0.1 mM PAPS at 37 °C for 0, 5, 15, 30 and 60 min. The metabolic stability of Testosterone (1  $\mu\text{M}$ ), diclofenac (1  $\mu\text{M}$ ) and propranolol (1  $\mu\text{M}$ ) were determined in parallel to confirm the enzymatic activity of the S9 fraction. The incubation was stopped by precipitation of S9 enzymes with 2 volumes of cold acetonitrile containing internal standard (150 nM Diphenhydramine). Samples were stored on ice for 10 min and precipitated protein was removed by centrifugation (15 min, 4°C, 4,000 rpm). The remaining test compound at different time points was analyzed by LC-MS/MS (TSQ Quantum Access MAX, Thermo Fisher, Dreieich, Germany) and used to determine half-life ( $t_{1/2}$ ).

#### 5.5.5.2 Metabolic Stability Tests in Mouse Liver S9 Fraction.

For the evaluation of combined phase I and phase II metabolic stability, the compound (1  $\mu\text{M}$ ) was incubated with 1 mg/mL pooled liver S9 fraction (SEKISUI XenoTech LLC, Kansas City, KS, USA), 2 mM NADPH, 1 mM UDPGA, 10 mM  $\text{MgCl}_2$ , 5 mM GSH and 0.1 mM PAPS at 37 °C for 0, 5, 15, 30 and 60 min. The metabolic stability of testosterone (1  $\mu\text{M}$ ), verapamil (1  $\mu\text{M}$ ) and ketoconazol (1  $\mu\text{M}$ ) were determined in parallel to confirm the enzymatic activity of the S9 fraction. The incubation was stopped by precipitation of S9 enzymes with 2 volumes of cold acetonitrile containing internal standard (150 nM Diphenhydramine). Samples were stored on ice for 10 min and precipitated protein was removed by centrifugation (15 min, 4°C, 4,000 rpm). The remaining test compound at different time points was analyzed by LC-MS/MS (TSQ Quantum Access MAX, Thermo Fisher, Dreieich, Germany) and used to determine half-life ( $t_{1/2}$ ).

### 5.5.6 Chemical Synthesis of Sandacrabins

General procedures A, B and C were adapted from the previously reported procedures for the chemical synthesis of sandacrabins B-E by Bader *et al.* [1].

#### 5.5.6.1 General Procedure A: Synthesis of the Benzimidazole Cores

4,5-Dimethylbenzene-1,2-diamine (1.0 equiv) was added to the respective carboxylic acid (10.0 equiv) in a crimp vial equipped with a magnetic stir bar. The crimp vial was sealed and heated at 145 °C for 2 h. The reaction was allowed to cool to room temperature and is then poured into 20 mL of saturated NaHCO<sub>3</sub> solution. The mixture was extracted with EtOAc twice, before the combined organic layers were washed with brine and dried over Na<sub>2</sub>SO<sub>4</sub>. The products were purified by flash chromatography on silica gel columns with visualization by analytical TLC on pre-coated UV-254 silica gel plates and detection with UV light.

#### 5.5.6.2 General Procedure B: Monosubstitution

The benzimidazole core (1.0 equiv) in *N,N*-Dimethylformamide (DMF) was added dropwise to a mixture of DMF and a 90% oil suspension of NaH (1.0 equiv) at 5-10 °C. The resulting mixture was stirred for 1 h, before the respective alkyl bromide (1.0 equiv) in DMF was added dropwise at 5-10 °C. The 0.1 M solution was stirred overnight at room temperature. After complete conversion (LC-MS), the reaction was quenched with water. The mixture was extracted with EtOAc twice, before the combined organic layers were washed with brine and dried over Na<sub>2</sub>SO<sub>4</sub>. The EtOAc was evaporated under reduced pressure and the residue was re-dissolved in methanol before purification by semipreparative HPLC.

#### 5.5.6.3 General Procedure C: Bissubstitution

2 mL of DMF was added to a crimp vial equipped with a magnetic stir bar followed by the addition of the benzimidazole core (1 equiv), the respective alkyl bromide (4 equiv) and finally potassium carbonate (1 equiv). This suspension was subsequently heated at 80 °C for 20 h. The reaction was quenched with water. The mixture was extracted with EtOAc twice, before the combined organic layers were washed with brine and dried over Na<sub>2</sub>SO<sub>4</sub>. The EtOAc was evaporated under reduced pressure and the residue was re-dissolved in methanol before purification by semipreparative HPLC.

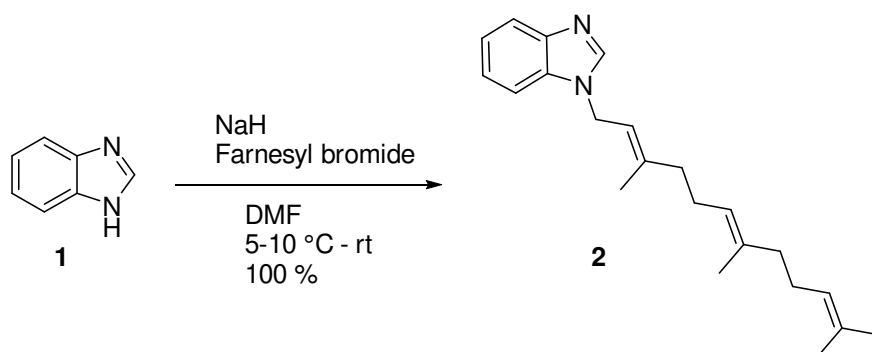
#### 5.5.6.4 General Procedure D: Bissubstitution

Bisfarnesylation of the respective benzimidazole core structure was alternatively achieved adapted from the procedure described by Gonell *et al.* [27]. The benzimidazole core (0.05 M, 1 equiv) was dissolved in acetonitrile (ACN) containing NaHCO<sub>3</sub> (2 equiv). The mixture was stirred at 90 °C for 1 h. After cooling down the mixture to room temperature, the respective alkyl bromide (5 equiv) was added and the mixture was further heated at 90 °C for 12 h. The mixture was extracted with EtOAc twice, before the combined

organic layers were washed with brine and dried over  $\text{Na}_2\text{SO}_4$ . The EtOAc was evaporated under reduced pressure and the residue was re-dissolved in methanol before purification by semipreparative HPLC.

#### 5.5.6.5 Experimental Procedures and Spectral Data

##### 1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**2**)



According to **GP B**, 1*H*-benzo[*d*]imidazole (**1**) (11.8 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**2**) (32.4 mg, 0.1 mmol) in quantitative yields as a yellow orange oil.

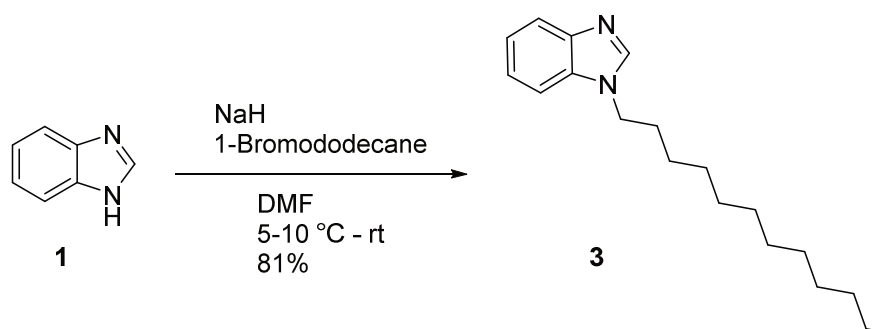
**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C)  $\delta$  = 1.54 (s, 5 H) 1.62 (s, 3 H) 1.87 (s, 3 H) 1.89 (br d,  $J$ =8.09 Hz, 2 H) 1.97 (q,  $J$ =14.20 Hz, 2 H) 2.03 - 2.19 (m, 5 H) 4.84 - 4.91 (m, 4 H) 4.99 - 5.07 (m, 2 H) 5.40 (br t,  $J$ =6.94 Hz, 1 H) 7.24 - 7.31 (m, 2 H) 7.48 (d,  $J$ =7.93 Hz, 1 H) 7.66 (d,  $J$ =7.32 Hz, 1 H) 8.12 (s, 1 H) ppm

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C)  $\delta$  = 144.3, 144.2, 142.8, 136.7, 135.0, 132.2, 125.5, 124.9, 124.3, 124.0, 123.7, 120.2, 120.0, 116.3, 111.8, 44.1, 40.9, 40.6, 27.9, 27.8, 27.1, 26.0, 17.9, 16.7, 16.3 ppm

**HRMS** (ESI):  $m/z$  calculated 323.2482 for  $\text{C}_{22}\text{H}_{31}\text{N}_2$  [ $\text{M}+\text{H}$ ]<sup>+</sup> found 323.2476  $\Delta$ -1.7 ppm

**UHPLC** (6 min):  $R_t$  = 4.65 min



1-dodecyl-1*H*-benzo[*d*]imidazole (**3**)

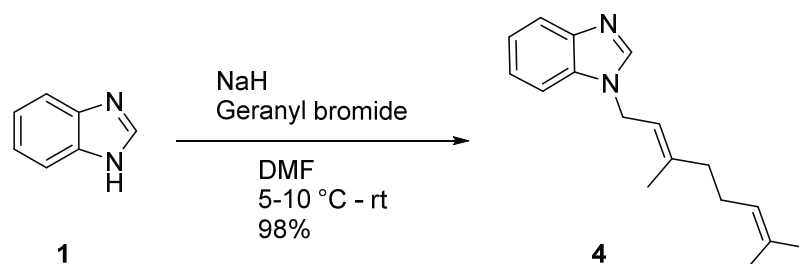
According to **GP B**, 1*H*-benzo[*d*]imidazole (**1**) (11.8 mg, 0.1 mmol) was reacted with 1-Bromododecane (24.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 1-dodecyl-1*H*-benzo[*d*]imidazole (**3**) (23.3 mg, 0.08 mmol) with 81% yield as a colorless oil.

<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 8.14 (s, 1H), 7.67 (d, *J* = 7.9 Hz, 1H), 7.54 (d, *J* = 8.1 Hz, 1H), 7.24-7.33 (m, 2H), 4.26 (t, *J* = 7.1 Hz, 2H), 1.78-1.92 (m, 2H), 1.18-1.34 (m, 19H), 0.88 ppm (t, *J* = 6.9 Hz, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 144.8, 144.2, 135.1, 124.3, 123.9, 123.6, 120.3, 111.6, 46.2, 33.2, 31.0, 30.9, 30.7, 30.7, 30.6, 30.3, 27.9, 23.9, 14.6 ppm

HRMS (ESI): *m/z* calculated 287.2482 for C<sub>19</sub>H<sub>31</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 287.2484 Δ-0.9 ppm

UHPLC (6 min): Rt = 4.79 min

*(E)*-1-(3,7-dimethylocta-2,6-dien-1-yl)-1*H*-benzo[*d*]imidazole (**4**)

According to **GP B**, 1*H*-benzo[*d*]imidazole (**1**) (11.8 mg, 0.1 mmol) was reacted with Geranyl bromide (21.6 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in *(E)*-1-(3,7-dimethylocta-2,6-dien-1-yl)-1*H*-benzo[*d*]imidazole (**4**) (25.1 mg, 0.1 mmol) in quantitative yields as a pale orange oil.

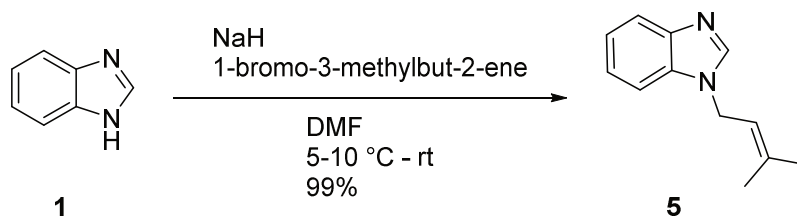
$^1\text{H NMR}$  (500 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 8.06-8.12 (m, 1H), 7.66 (d,  $J$  = 7.8 Hz, 1H), 7.45-7.51 (m, 1H), 7.22-7.32 (m, 2H), 5.34-5.42 (m, 1H), 4.99-5.06 (m, 1H), 4.80-4.90 (m, 3H), 2.05-2.17 (m, 5H), 1.84 (s, 3H), 1.59 (s, 3H), 1.54 ppm (s, 3H)

$^{13}\text{C NMR}$  (126 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 144.3, 144.2, 142.8, 135.1, 132.9, 125.0, 124.7, 124.3, 123.7, 120.2, 119.9, 111.8, 44.0, 40.5, 27.2, 26.0, 17.9, 16.6 ppm

**HRMS** (ESI):  $m/z$  calculated 255.1856 for  $\text{C}_{17}\text{H}_{23}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 255.1851  $\Delta$ 1.9 ppm

**UHPLC** (6 min):  $R_t$  = 3.68 min

1-(3-methylbut-2-en-1-yl)-1H-benzo[d]imidazole (**5**)



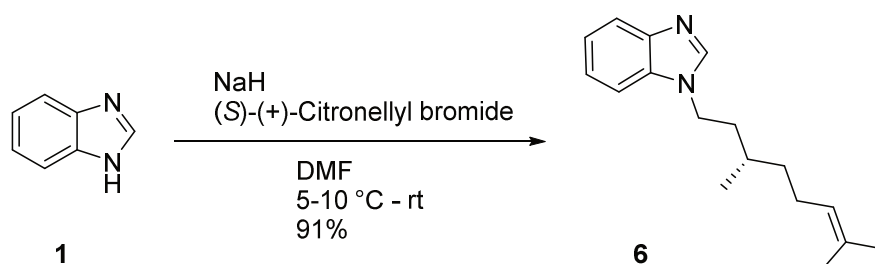
According to **GP B**, 1H-benzo[d]imidazole (**1**) (23.6 mg, 0.2 mmol) was reacted with 1-bromo-3-methylbut-2-ene (29.6 mg, 0.2 mmol) utilizing NaH (4.6 mg, 0.2 mmol) in 2 mL DMF. The reaction resulted in 1-(3-methylbut-2-en-1-yl)-1H-benzo[d]imidazole (**5**) (36.9 mg, 0.2 mmol) in quantitative yields as a pale yellow oil.

$^1\text{H NMR}$  (500 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 8.00 (br s, 1H), 7.57 (d,  $J$  = 7.8 Hz, 1H), 7.37 (br d,  $J$  = 7.5 Hz, 1H), 7.14-7.21 (m, 2H), 5.29 (br d,  $J$  = 5.8 Hz, 1H), 4.70-4.77 (m, 2H), 1.76 (br s, 3H), 1.67 ppm (br s, 3H)

$^{13}\text{C NMR}$  (126 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 142.8, 142.7, 137.8, 133.5, 122.7, 122.4, 122.1, 118.7, 118.3, 110.2, 42.5, 24.4, 16.7 ppm

**HRMS** (ESI):  $m/z$  calculated 187.1230 for  $\text{C}_{12}\text{H}_{15}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 187.1227  $\Delta$ 1.7 ppm

**UHPLC** (6 min):  $R_t$  = 2.36 min

*(S)*-1-(3,7-dimethyloct-6-en-1-yl)-1*H*-benzo[*d*]imidazole (**6**)

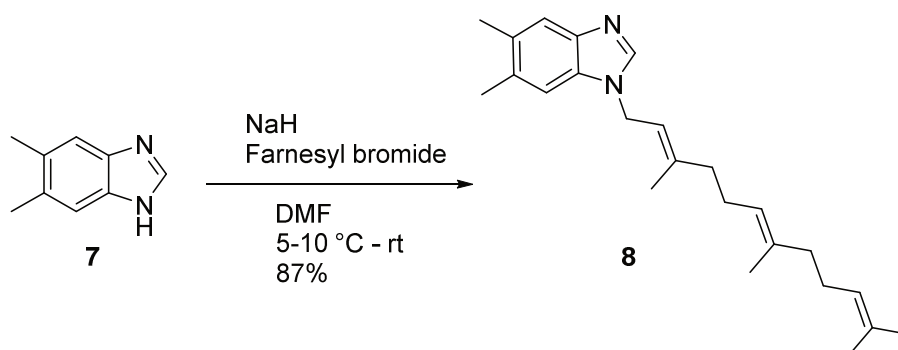
According to **GP B**, 1*H*-benzo[*d*]imidazole (**1**) (11.8 mg, 0.1 mmol) was reacted with (*S*)-(+)-citronellyl bromide (21.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in (*S*)-1-(3,7-dimethyloct-6-en-1-yl)-1*H*-benzo[*d*]imidazole (**6**) (23.3 mg, 0.09 mmol) with 91% yield as a pale yellow oil.

<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 8.13 (br s, 1H), 7.67 (d, *J* = 7.9 Hz, 1H), 7.51 (br d, *J* = 7.8 Hz, 1H), 7.24-7.32 (m, 2H), 4.86-4.91 (m, 1H), 4.21-4.34 (m, 2H), 1.84-2.03 (m, 3H), 1.63-1.68 (m, 1H), 1.61 (s, 3H), 1.51 (s, 3H), 1.33-1.43 (m, 2H), 1.14-1.25 (m, 1H), 0.97 ppm (br d, *J* = 6.4 Hz, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 144.7, 144.2, 135.0, 132.5, 125.6, 124.4, 123.6, 120.3, 111.6, 44.3, 38.0, 37.9, 31.1, 26.4, 26.0, 19.9, 17.9 ppm

HRMS (ESI): *m/z* calculated 257.2012 for C<sub>17</sub>H<sub>25</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 257.2016 Δ-1.3 ppm

UHPLC (6 min): *R*<sub>t</sub> = 3.81 min

5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**8**)

According to **GP B**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (14.6 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**8**) (30.6 mg, 0.09 mmol) in 87% yield as a yellow orange oil.

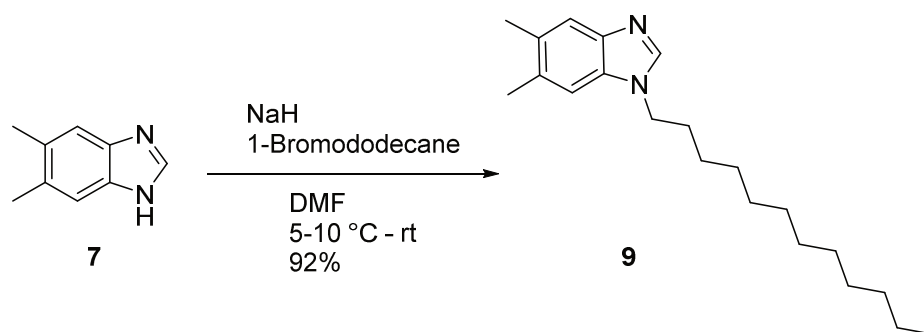
**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.96 (s, 1H), 7.41 (s, 1H), 7.22 (s, 1H), 5.37 (br t, J = 6.7 Hz, 1H), 4.96-5.06 (m, 2H), 4.78-4.80 (m, 1H), 4.79 (s, 1H), 2.37 (s, 4H), 2.33-2.35 (m, 4H), 2.07-2.17 (m, 5H), 1.91-1.95 (m, 2H), 1.84-1.87 (m, 5H), 1.62 (s, 3H), 1.53 ppm (s, 6H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 143.4, 142.8, 142.7, 142.5, 136.7, 133.6, 133.1, 132.7, 132.1, 125.6, 124.9, 120.2, 120.1, 116.2, 111.8, 43.9, 40.9, 40.5, 27.8, 27.1, 26.0, 20.8, 20.6, 20.5, 17.9, 16.6, 16.3 ppm

**HRMS** (ESI): *m/z* calculated 351.2795 for C<sub>24</sub>H<sub>35</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 351.2975 Δ0.0 ppm

**UHPLC** (6 min): Rt = 4.74 min

1-dodecyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**9**)



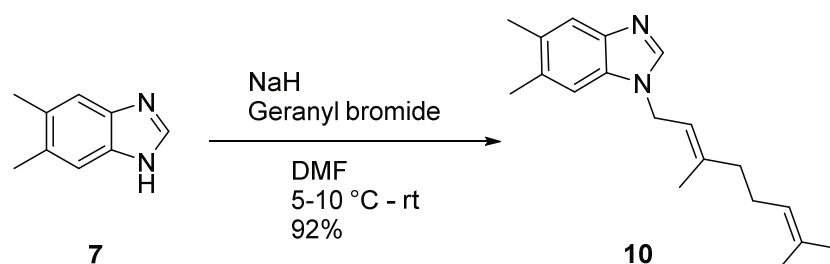
According to **GP B**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (14.6 mg, 0.1 mmol) was reacted with 1-Bromododecane (24.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 1-dodecyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**9**) (29 mg, 0.09 mmol) with 92% yield as a white solid.

**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.87 (s, 1H), 7.31 (s, 1H), 7.17 (s, 1H), 4.04-4.09 (m, 2H), 2.27 (s, 3H), 2.24 (s, 4H), 1.69-1.76 (m, 2H), 1.05-1.23 (m, 20H), 0.78 ppm (t, J = 7.0 Hz, 3H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 142.3, 141.2, 132.1, 132.2, 132.1, 131.0, 118.7, 110.1, 44.5, 31.7, 29.4, 29.3, 29.2, 29.2, 29.1, 28.8, 26.3, 22.3, 19.2, 19.1, 19.0, 13.1 ppm

**HRMS** (ESI): *m/z* calculated 315.2795 for C<sub>21</sub>H<sub>35</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 315.2787 Δ2.3 ppm

**UHPLC** (6 min): Rt = 4.94 min

*(E)*-1-(3,7-dimethylocta-2,6-dien-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**10**)

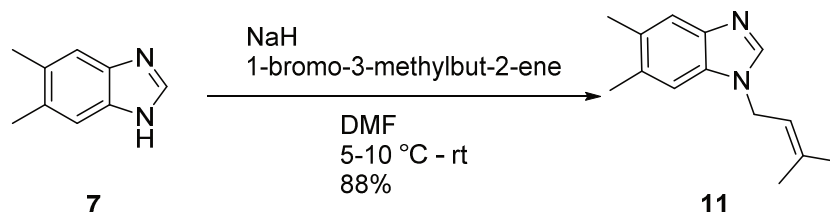
According to **GP B**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (14.6 mg, 0.1 mmol) was reacted with Geranyl bromide (21.6 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in (*E*)-1-(3,7-dimethylocta-2,6-dien-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**10**) (26.1 mg, 0.09 mmol) in 92% yield as a pale yellow solid.

<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.83 (s, 1H), 7.30 (s, 1H), 7.11 (s, 1H), 5.21-5.30 (m, 1H), 4.88-4.95 (m, 1H), 4.79 (s, 1H), 4.65 (d, *J* = 7.0 Hz, 2H), 2.26 (s, 3H), 2.23 (s, 4H), 1.93-2.07 (m, 4H), 1.73 (s, 3H), 1.48 (s, 3H), 1.43 ppm (s, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 141.8, 141.3, 141.1, 132.1, 132.0, 131.4, 131.1, 123.4, 118.6, 110.2, 47.9, 47.7, 42.3, 39.0, 25.7, 24.4, 19.3, 19.0, 19.0, 16.3, 15.1 ppm

HRMS (ESI): *m/z* calculated 283.2169 for C<sub>19</sub>H<sub>27</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 283.2168 Δ0.1 ppm

UHPLC (6 min): Rt = 3.99 min

1-(3-methylbut-2-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**11**)

According to **GP B**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (29.2 mg, 0.2 mmol) was reacted with 1-bromo-3-methylbut-2-ene (29.6 mg, 0.2 mmol) utilizing NaH (4.6 mg, 0.2 mmol) in 2 mL DMF. The reaction resulted in 1-(3-methylbut-2-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**11**) (38 mg, 0.18 mmol) with 88% yield as a pale yellow oil.

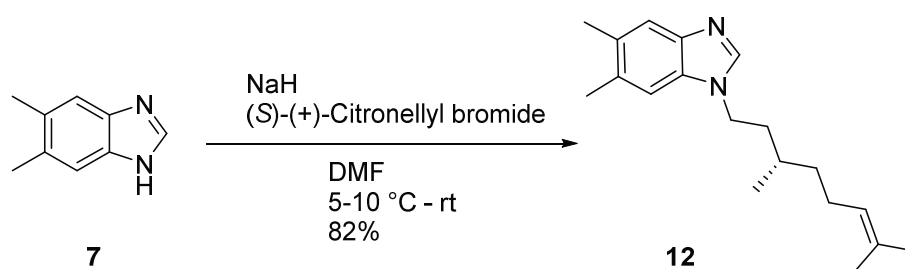
$^1\text{H NMR}$  (500 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 7.84 (s, 1H), 7.31 (s, 1H), 7.10 (s, 1H), 5.24-5.29 (m, 1H), 4.62 (d,  $J$  = 7.0 Hz, 2H), 2.26 (s, 3H), 2.23 (s, 3H), 1.75 (s, 3H), 1.67 ppm (s, 3H)

$^{13}\text{C NMR}$  (126 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 141.8, 141.3, 137.6, 132.1, 132.0, 131.1, 118.6, 118.5, 110.1, 42.4, 24.3, 19.2, 18.9, 16.7 ppm

**HRMS** (ESI):  $m/z$  calculated 215.1543 for  $\text{C}_{14}\text{H}_{19}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 215.1547  $\Delta$ -1.8 ppm

**UHPLC** (6 min):  $R_t$  = 2.95 min

(*S*)-1-(3,7-dimethyloct-6-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**12**)



According to **GP B**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (14.6 mg, 0.1 mmol) was reacted with (*S*)-(+)-citronellyl bromide (21.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in (*S*)-1-(3,7-dimethyloct-6-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**12**) (23.4 mg, 0.08 mmol) with 82% yield as a white oil.

$^1\text{H NMR}$  (500 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 7.98 (s, 1H), 7.41 (s, 1H), 7.26 (s, 1H), 4.96-5.04 (m, 1H), 4.89 (s, 1H), 4.14-4.26 (m, 2H), 2.38 (s, 3H), 2.34 (s, 3H), 1.82-2.05 (m, 3H), 1.61-1.67 (m, 1H), 1.60 (s, 3H), 1.52 (s, 3H), 1.33-1.42 (m, 2H), 1.13-1.24 (m, 1H), 0.96 ppm (d,  $J$  = 6.6 Hz, 3H)

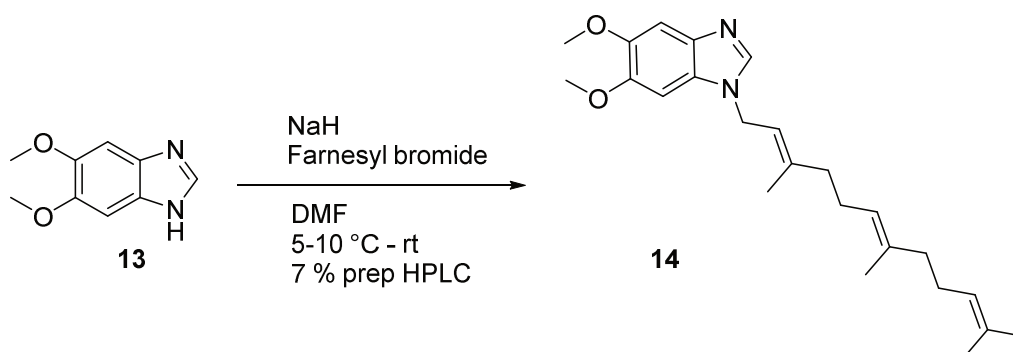
$^{13}\text{C NMR}$  (126 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 142.2, 141.2, 132.1, 132.1, 131.1, 130.9, 124.1, 118.7, 110.0, 47.9, 42.6, 36.4, 36.3, 29.6, 24.9, 24.9, 24.4, 19.2, 18.9, 18.4, 16.3 ppm

**HRMS** (ESI):  $m/z$  calculated 285.2325 for  $\text{C}_{19}\text{H}_{29}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 285.2325  $\Delta$ -0.1 ppm

**UHPLC** (6 min):  $R_t$  = 4.10 min

1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-5,6-dimethoxy-1*H*-benzo[*d*]imidazole

(**14**)



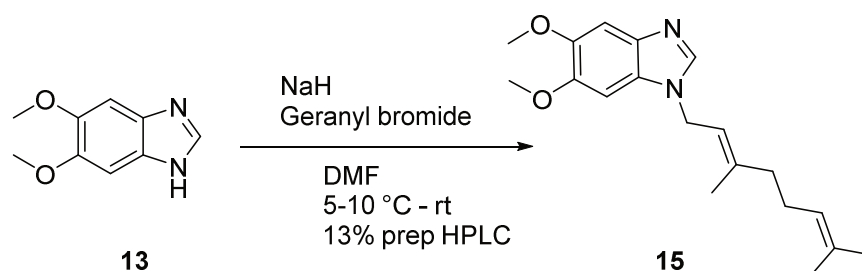
According to **GP B**, 5,6-dimethoxy-1*H*-benzo[*d*]imidazole (**13**) (19.4 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.5 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-5,6-dimethoxy-1*H*-benzo[*d*]imidazole (**14**) (2.69 mg, 7  $\mu$ mol) with 7% yield after preparative HPLC as a yellow orange oil.

**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C):  $\delta$  = 7.92 (s, 1H), 7.16 (s, 1H), 7.01 (s, 1H), 5.36-5.43 (m, 1H), 4.92-5.07 (m, 2H), 4.81 (d, *J* = 7.1 Hz, 2H), 3.87 (s, 3H), 3.85 (s, 3H), 2.06-2.19 (m, 4H), 1.78-1.95 (m, 7H), 1.60 (s, 2H), 1.52 ppm (d, *J* = 4.2 Hz, 5H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C):  $\delta$  = 147.6, 147.1, 141.3, 141.1, 141.0, 136.2, 135.2, 130.6, 127.5, 124.0, 123.3, 118.7, 101.1, 93.4, 55.4, 48.1, 47.9, 47.7, 47.5, 42.4, 39.3, 39.0, 26.2, 25.5, 24.5, 16.3, 15.1, 14.7 ppm

**HRMS** (ESI): *m/z* calculated 383.2693 for C<sub>24</sub>H<sub>35</sub>N<sub>2</sub>O<sub>2</sub> [M+H]<sup>+</sup> found 383.2690  $\Delta$ 0.7 ppm

**UHPLC** (6 min): Rt = 4.61 min

*(E)*-1-(3,7-dimethylocta-2,6-dien-1-yl)-5,6-dimethoxy-1*H*-benzo[*d*]imidazole (**15**)

According to **GP B**, 5,6-dimethoxy-1*H*-benzo[*d*]imidazole (**13**) (17.8 mg, 0.1 mmol) was reacted with Geranyl bromide (21.6 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in *(E)*-1-(3,7-dimethylocta-2,6-dien-1-yl)-5,6-dimethoxy-1*H*-benzo[*d*]imidazole (**15**) (4 mg, 0.01 mmol) with 13% yield after preparative HPLC as a pale yellow oil.

**<sup>1</sup>H NMR** <sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 7.92 (s, 1H), 7.10 (s, 1H), 6.96 (s, 1H), 5.32 (br t, *J* = 6.7 Hz, 1H), 4.91-4.99 (m, 1H), 4.71-4.79 (m, 2H), 3.85-3.88 (m, 1H), 3.79 (s, 3H), 3.75 (s, 3H), 3.50-3.59 (m, 1H), 2.00-2.08 (m, 4H), 1.78 (s, 3H), 1.48 (s, 3H), 1.45 ppm (s, 3H)

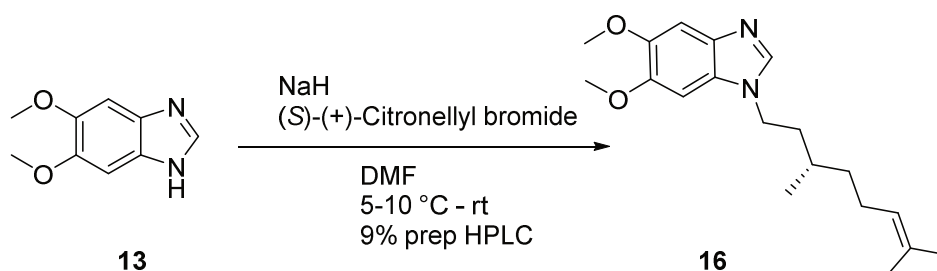
**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 147.8, 147.3, 141.6, 141.0, 135.5, 131.4, 127.3, 123.4, 118.3, 100.7, 93.5, 55.4, 55.4, 48.2, 48.0, 47.9, 42.6, 39.0, 25.8, 24.4, 16.3, 15.0 ppm

**HRMS** (ESI): *m/z* calculated 315.2067 for C<sub>19</sub>H<sub>27</sub>N<sub>2</sub>O<sub>2</sub> [M+H]<sup>+</sup>; found 315.2070 Δ-0.8 ppm

**UHPLC** (6 min): Rt = 3.55 min



(*S*)-1-(3,7-dimethyloct-6-en-1-yl)-5,6-dimethoxy-1*H*-benzo[*d*]imidazole (**16**)



According to **GP B**, 5,6-dimethoxy-1*H*-benzo[*d*]imidazole (**13**) (17.8 mg, 0.1 mmol) was reacted with (*S*)-(+)-citronellyl bromide (21.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in (*S*)-1-(3,7-dimethyloct-6-en-1-yl)-5,6-dimethoxy-1*H*-benzo[*d*]imidazole (**16**) (2.7 mg, 0.009 mmol) with 9% yield after preparative HPLC as a pale yellow oil.

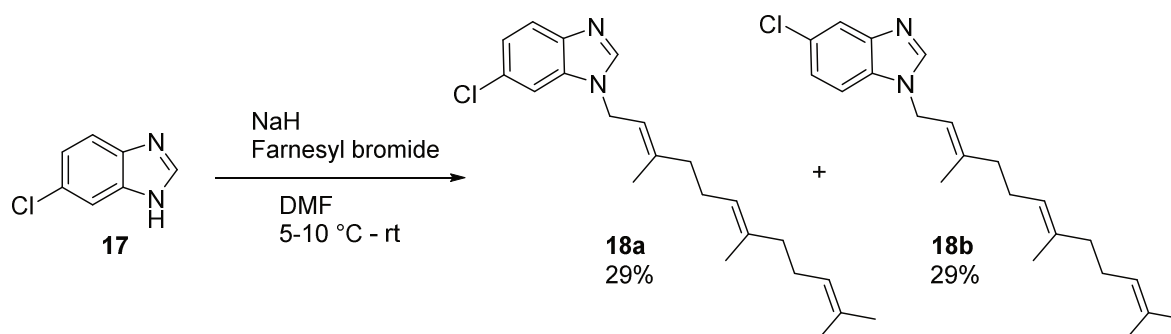
**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C):  $\delta$  = 8.12 (s, 1H), 7.24 (s, 1H), 7.13-7.19 (m, 1H), 4.98-5.07 (m, 1H), 4.20-4.38 (m, 3H), 3.94 (s, 3H), 3.87 (s, 2H), 3.85 (s, 3H), 3.63-3.72 (m, 1H), 1.88-2.03 (m, 4H), 1.66-1.74 (m, 2H), 1.63 (s, 4H), 1.54 (s, 3H), 1.38-1.45 (m, 3H), 1.28-1.35 (m, 1H), 1.21-1.31 (m, 3H), 1.01 ppm (d, *J* = 6.4 Hz, 4H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C):  $\delta$  = 148.0, 147.4, 141.6, 135.2, 130.9, 127.3, 124.0, 100.7, 93.4, 55.6, 55.4, 47.4, 42.9, 36.4, 36.3, 29.5, 29.4, 24.9, 24.4, 18.3, 16.3 ppm

**HRMS** (ESI): *m/z* calculated 317.2224 for C<sub>19</sub>H<sub>29</sub>N<sub>2</sub>O<sub>2</sub> [M+H]<sup>+</sup>; found 317.2226  $\Delta$ -0.7 ppm

**UHPLC** (6 min): Rt = 3.68 min

6-Chloro-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**18a**) and 5-Chloro-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**18b**)



According to **GP B**, 5-chloro-1*H*-benzo[*d*]imidazole (**17**) (15.2 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in a mixture of 6-chloro-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**18a**) and 5-chloro-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**18b**) as pale orange oils. Preparative HPLC yielded 29% (10.2 mg, 0.03 mmol) of each isomer.

#### **18a:**

<sup>1</sup>**H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 8.04 (s, 1H), 7.51 (d, *J* = 8.5 Hz, 1H), 7.42 (d, *J* = 2.0 Hz, 1H), 7.15 (dd, *J* = 8.5, 2.0 Hz, 1H), 5.30 (td, *J* = 7.0, 0.9 Hz, 1H), 4.88-4.97 (m, 2H), 4.77 (s, 5H), 4.74-4.76 (m, 1H), 2.01-2.10 (m, 4H), 1.85 (q, *J* = 6.9 Hz, 2H), 1.76-1.80 (m, 5H), 1.52 (s, 3H), 1.44 ppm (d, *J* = 4.6 Hz, 6H)

<sup>13</sup>**C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 144.0, 141.7, 141.5, 135.2, 134.2, 130.6, 128.5, 124.0, 123.3, 122.7, 119.8, 118.2, 110.4, 48.2, 48.1, 47.9, 47.7, 42.6, 39.3, 39.0, 26.3, 25.5, 24.5, 16.3, 15.1, 14.7 ppm

**HRMS** (ESI): *m/z* calculated 357.2092 for C<sub>22</sub>H<sub>30</sub>ClN<sub>2</sub> [M+H]<sup>+</sup> found 357.2084 Δ2.2 ppm

**UHPLC** (6 min): Rt = 5.71 min

#### **18b:**

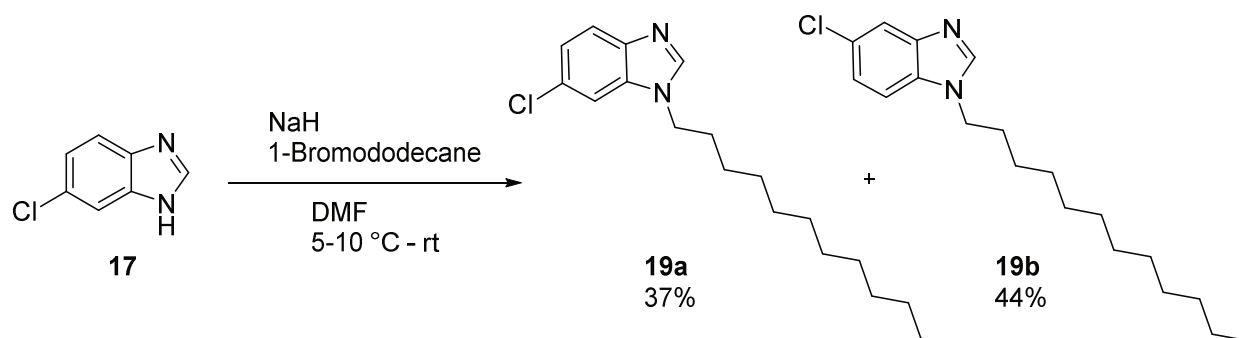
<sup>1</sup>**H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 8.03-8.11 (m, 1H), 7.54 (d, *J* = 1.8 Hz, 1H), 7.37 (d, *J* = 8.7 Hz, 1H), 7.17 (d, *J* = 8.5 Hz, 1H), 5.26-5.33 (m, 1H), 4.89-4.96 (m, 2H), 4.78 (s, 1H), 4.77 (s, 4H), 2.00-2.08 (m, 4H), 1.85 (q, *J* = 7.2 Hz, 2H), 1.77 (s, 4H), 1.73-1.77 (m, 1H), 1.53 (s, 3H), 1.44 (s, 5H), 0.70-0.92 ppm (m, 1H)

<sup>13</sup>**C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 144.3, 143.7, 141.6, 135.2, 132.3, 130.7, 127.8, 124.0, 123.3, 123.1, 118.3, 118.2, 111.4, 48.2, 48.1, 47.9, 47.7, 42.7, 39.3, 39.0, 26.3, 25.5, 24.5, 16.3, 15.1, 14.7 ppm

**HRMS** (ESI): *m/z* calculated 357.2092 for C<sub>22</sub>H<sub>30</sub>ClN<sub>2</sub> [M+H]<sup>+</sup> found 357.2084 Δ2.4 ppm

**UHPLC** (6 min): Rt = 5.75 min

6-chloro-1-dodecyl-1*H*-benzo[*d*]imidazole (**19a**) and 5-chloro-1-dodecyl-1*H*-benzo[*d*]imidazole (**19b**)



According to **GP B**, 5-chloro-1*H*-benzo[*d*]imidazole (**17**) (15.2 mg, 0.1 mmol) was reacted with 1-Bromododecane (24.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in a mixture of 6-chloro-1-dodecyl-1*H*-benzo[*d*]imidazole (**19a**) and 5-chloro-1-dodecyl-1*H*-benzo[*d*]imidazole imidazole (**19b**) as colorless oils. Preparative HPLC yielded 37% of **19a** (11.8 mg, 0.04 mmol) and 44% of **19b** (14.2 mg, 0.04 mmol).

#### **19a:**

**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 8.09 (s, 1H), 7.53 (s, 1H), 7.51 (s, 1H), 7.16 (dd, *J* = 8.5, 2.0 Hz, 1H), 4.16 (t, *J* = 7.1 Hz, 2H), 1.77 (quin, *J* = 7.1 Hz, 2H), 1.08-1.32 (m, 21H), 0.79 ppm (t, *J* = 6.9 Hz, 3H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 144.4, 141.4, 134.3, 128.6, 122.6, 119.8, 110.3, 48.1, 47.9, 47.7, 47.1, 44.7, 31.7, 29.3, 29.2, 29.1, 29.0, 28.7, 26.2, 22.3, 13.0 ppm

**HRMS** (ESI): *m/z* calculated 321.2092 for C<sub>19</sub>H<sub>30</sub>ClN<sub>2</sub> [M+H]<sup>+</sup>; found 321.2084 Δ2.4 ppm

**UHPLC** (6 min): Rt = 5.94 min

#### **19b:**

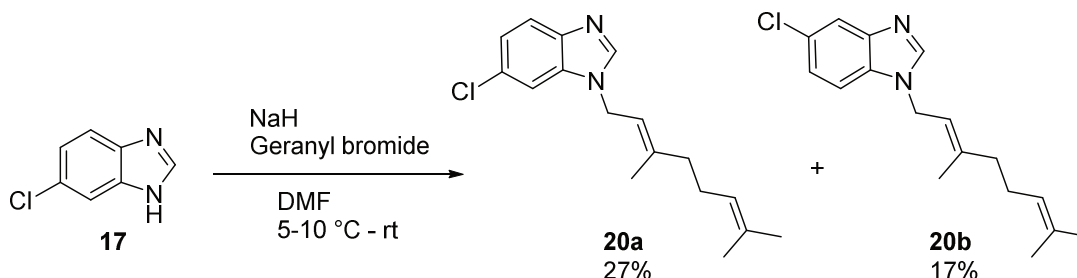
**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 8.11 (s, 1H), 7.55 (d, *J* = 1.8 Hz, 1H), 7.45 (d, *J* = 8.7 Hz, 1H), 7.20 (d, *J* = 8.7 Hz, 1H), 4.17 (t, *J* = 7.1 Hz, 2H), 1.77 (quin, *J* = 6.9 Hz, 2H), 1.08-1.30 (m, 21H), 0.79 ppm (t, *J* = 6.9 Hz, 3H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 144.8, 143.5, 132.4, 127.8, 123.1, 118.4, 111.3, 48.1, 47.9, 47.7, 44.8, 31.7, 29.4, 29.3, 29.3, 29.2, 29.1, 29.0, 28.7, 26.3, 22.3, 13.0 ppm

**HRMS** (ESI): *m/z* calculated 321.2092 for C<sub>19</sub>H<sub>30</sub>ClN<sub>2</sub> [M+H]<sup>+</sup>; found 321.2087 Δ1.6 ppm

**UHPLC** (6 min): Rt = 5.96 min

6-chloro-(*E*)-1-(3,7-dimethylocta-2,6-dien-1-yl)-1*H*-benzo[*d*]imidazole (**20a**) and 5-chloro-(*E*)-1-(3,7-dimethylocta-2,6-dien-1-yl)-1*H*-benzo[*d*]imidazole (**20b**)



According to **GP B**, 5-chloro-1*H*-benzo[*d*]imidazole (**17**) (15.2 mg, 0.1 mmol) was reacted with Geranyl bromide (21.6 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in a mixture of 6-chloro-(*E*)-1-(3,7-dimethylocta-2,6-dien-1-yl)-1*H*-benzo[*d*]imidazole (**20a**) and 5-chloro-(*E*)-1-(3,7-dimethylocta-2,6-dien-1-yl)-1*H*-benzo[*d*]imidazole (**20b**) as colorless oils. Preparative HPLC yielded 27% of **20a** (7.9 mg, 0.03 mmol) and 17% of **20b** (5 mg, 0.02 mmol).

**20a:**

<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 18°C): δ = 8.15 (s, 1H), 7.62 (d, *J* = 8.7 Hz, 1H), 7.54 (d, *J* = 1.8 Hz, 1H), 7.27 (dd, *J* = 8.5, 2.0 Hz, 1H), 5.40 (br d, *J* = 1.2 Hz, 1H), 5.01-5.06 (m, 1H), 4.87 (d, *J* = 7.0 Hz, 2H), 2.09-2.19 (m, 4H), 1.87 (s, 3H), 1.58 (s, 3H), 1.54-1.57 ppm (m, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 18°C): δ = 145.5, 143.3, 143.1, 135.8, 133.0, 130.0, 124.9, 124.2, 121.3, 119.6, 112.0, 44.1, 40.5, 27.2, 26.0, 17.9, 16.6 ppm

HRMS (ESI): *m/z* calculated 289.1466 for C<sub>17</sub>H<sub>22</sub>ClN<sub>2</sub> [M+H]<sup>+</sup>; found 289.1469 Δ-1.0 ppm

UHPLC (6 min): Rt = 4.59 min

**20b:**

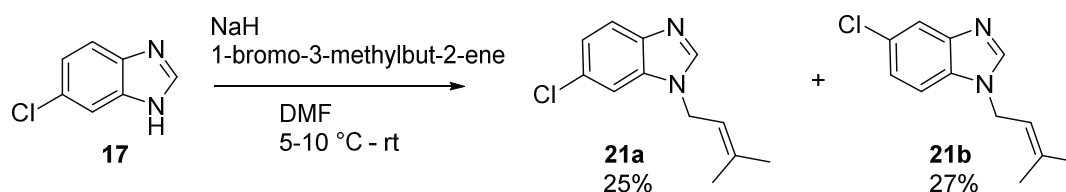
<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 18°C): δ = 8.07 (s, 1H), 7.55 (d, *J* = 1.8 Hz, 1H), 7.39 (d, *J* = 8.5 Hz, 1H), 7.19 (d, *J* = 8.6 Hz, 1H), 5.31 (t, *J* = 7.0 Hz, 1H), 4.91-4.96 (m, 1H), 4.76-4.80 (m, 2H), 3.52-3.55 (m, 1H), 1.97-2.07 (m, 4H), 1.77 (s, 3H), 1.47-1.51 (m, 3H), 1.45 ppm (s, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 144.3, 143.6, 141.8, 132.3, 131.4, 127.8, 123.4, 123.1, 122.7, 118.3, 118.1, 111.5, 48.2, 48.0, 47.9, 47.7, 42.7, 42.6, 39.0, 25.7, 24.4, 24.4, 16.3, 15.0, 15.0 ppm

HRMS (ESI): *m/z* calculated 289.1466 for C<sub>17</sub>H<sub>22</sub>ClN<sub>2</sub> [M+H]<sup>+</sup>; found 289.1469 Δ-1.0 ppm

UHPLC (6 min): Rt = 4.63 min

6-chloro-1-(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazole (**21a**) and 5-chloro-1-(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazole (**21b**)



According to **GP B**, 5-chloro-1*H*-benzo[*d*]imidazole (**17**) (30.4 mg, 0.2 mmol) was reacted with 1-bromo-3-methylbut-2-ene (29.6 mg, 0.12 mmol) utilizing NaH (4.6 mg, 0.2 mmol) in 2 mL DMF. The reaction resulted in a mixture of 6-chloro-1-(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazole (**21a**) and 5-chloro-1-(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazole (**21b**) as colorless oils. Preparative HPLC yielded 25% of **21a** (11.1 mg, 0.05 mmol) and 27% of **21b** (12.1 mg, 0.05 mmol).

**21a:**

<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 8.18 (s, 1H), 7.63 (d, *J* = 8.7 Hz, 1H), 7.55 (d, *J* = 1.8 Hz, 1H), 7.28 (dd, *J* = 8.5, 2.0 Hz, 1H), 5.40-5.45 (m, 1H), 4.87 (d, *J* = 7.0 Hz, 2H), 1.90 (s, 3H), 1.82 ppm (s, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 144.0, 141.5, 138.4, 134.2, 128.5, 122.6, 119.7, 118.0, 110.4, 48.1, 47.9, 42.7, 24.4, 16.7 ppm

HRMS (ESI): *m/z* calculated 221.0840 for C<sub>12</sub>H<sub>14</sub>ClN<sub>2</sub> [M+H]<sup>+</sup>; found 221.08436 Δ1.6 ppm

UHPLC (6 min): Rt = 3.10 min

**21b:**

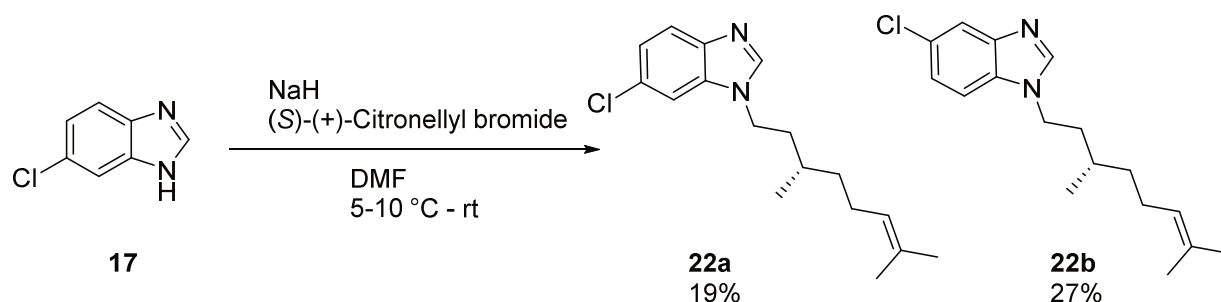
<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 8.19 (s, 2H), 7.66 (d, *J* = 2.0 Hz, 2H), 7.49 (d, *J* = 8.7 Hz, 2H), 7.30 (dd, *J* = 8.6, 1.9 Hz, 2H), 5.40-5.44 (m, 2H), 4.88 (d, *J* = 7.0 Hz, 4H), 1.90 (s, 6H), 1.80-1.82 ppm (m, 6H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 144.3, 143.6, 138.3, 132.3, 127.8, 123.1, 118.3, 118.0, 111.4, 48.1, 47.9, 42.7, 24.4, 16.7 ppm

HRMS (ESI): *m/z* calculated 221.0840 for C<sub>12</sub>H<sub>14</sub>ClN<sub>2</sub> [M+H]<sup>+</sup>; found 221.0837 Δ1.5 ppm

UHPLC (6 min): Rt = 3.18 min

6-chloro-(*S*)-1-(3,7-dimethyloct-6-en-1-yl)-1*H*-benzo[*d*]imidazole (**22a**) and 5-chloro-(*S*)-1-(3,7-dimethyloct-6-en-1-yl)-1*H*-benzo[*d*]imidazole (**22b**)



According to **GP B**, 5-chloro-1*H*-benzo[*d*]imidazole (**17**) (15.2 mg, 0.1 mmol) was reacted with (*S*)-(+)-citronellyl bromide (21.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in a mixture of 6-chloro-(*S*)-1-(3,7-dimethyloct-6-en-1-yl)-1*H*-benzo[*d*]imidazole (**22a**) and 5-chloro-(*S*)-1-(3,7-dimethyloct-6-en-1-yl)-1*H*-benzo[*d*]imidazole (**22b**) as colorless oils. Preparative HPLC yielded 19% of **22a** (5.4 mg, 0.02 mmol) and 27% of **22b** (8 mg, 0.03 mmol).

#### **22a:**

<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 8.22 (s, 1H), 7.64-7.66 (m, 1H), 7.64 (s, 1H), 7.29 (d, *J* = 8.5 Hz, 1H), 5.00-5.05 (m, 1H), 4.26-4.38 (m, 2H), 1.90-2.10 (m, 3H), 1.66-1.77 (m, 1H), 1.63 (d, *J* = 0.9 Hz, 3H), 1.55 (s, 3H), 1.36-1.45 (m, 2H), 1.21-1.29 (m, 1H), 1.02 ppm (d, *J* = 6.6 Hz, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 144.4, 141.4, 134.2, 131.0, 128.6, 124.0, 122.7, 119.8, 110.2, 48.1, 47.9, 47.7, 42.9, 36.3, 36.2, 29.5, 24.9, 24.4, 18.3, 16.3 ppm

HRMS (ESI): *m/z* calculated 291.1623 for C<sub>17</sub>H<sub>24</sub>ClN<sub>2</sub> [M+H]<sup>+</sup>; found 291.1629 Δ-2.3 ppm

UHPLC (6 min): Rt = 4.77 min

#### **22b:**

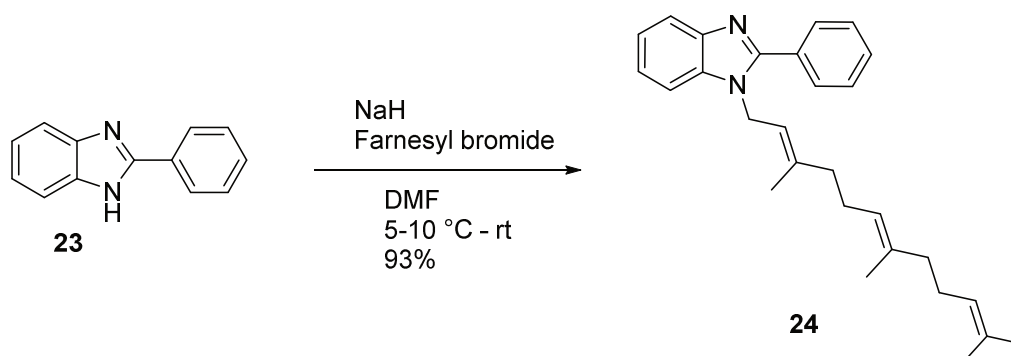
<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 8.22-8.26 (m, 1H), 7.67 (d, *J* = 1.8 Hz, 1H), 7.57 (d, *J* = 8.5 Hz, 1H), 7.33 (dd, *J* = 8.6, 1.9 Hz, 1H), 5.02 (tt, *J* = 7.2, 1.3 Hz, 1H), 4.26-4.39 (m, 2H), 1.89-2.11 (m, 3H), 1.64-1.84 (m, 1H), 1.63 (s, 3H), 1.54 (s, 3H), 1.36-1.45 (m, 2H), 1.20-1.34 (m, 2H), 1.01 ppm (d, *J* = 6.4 Hz, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 144.7, 143.5, 132.3, 131.0, 127.8, 124.0, 123.1, 118.4, 111.3, 47.9, 47.7, 47.1, 43.0, 36.4, 36.3, 36.3, 29.5, 24.9, 24.4, 18.3, 16.3 ppm

HRMS (ESI): *m/z* calculated 291.1623 for C<sub>17</sub>H<sub>24</sub>ClN<sub>2</sub> [M+H]<sup>+</sup>; found 291.1628 Δ-1.7 ppm

UHPLC (6 min): Rt = 4.80 min

2-phenyl-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**24**)



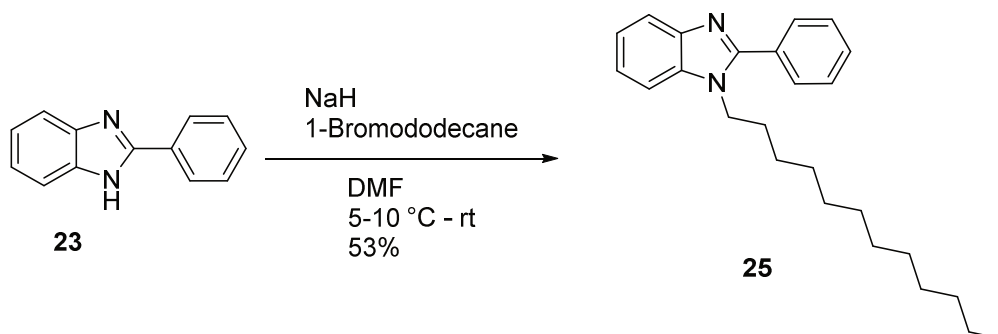
According to **GP B**, 2-phenyl-1*H*-benzo[*d*]imidazole (**23**) (19.4 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 2-phenyl-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**24**) (37.2 mg, 0.09 mmol) with 93% yield as a yellow orange oil.

**<sup>1</sup>H NMR** (500 MHz, CHLOROFORM-*d*, 19°C):  $\delta$  = 8.02-8.06 (m, 1H), 7.71-7.77 (m, 1H), 7.65 (d,  $J$  = 6.1 Hz, 2H), 7.38-7.41 (m, 3H), 7.30-7.33 (m, 1H), 7.21-7.23 (m, 1H), 7.07-7.13 (m, 1H), 5.26-5.36 (m, 1H), 4.96-5.03 (m, 2H), 4.75 (br d,  $J$  = 6.1 Hz, 2H), 1.97-2.09 (m, 4H), 1.91-1.97 (m, 2H), 1.83-1.89 (m, 2H), 1.65 (s, 3H), 1.58 (s, 3H), 1.50 (s, 3H), 1.49 ppm (s, 3H)

**<sup>13</sup>C NMR** (126 MHz, CHLOROFORM-*d*, 18°C):  $\delta$  = 153.7, 142.9, 139.6, 135.7, 131.4, 130.1, 129.8, 129.5, 128.8, 128.7, 126.9, 124.2, 123.4, 122.8, 122.6, 119.7, 119.5, 110.4, 43.5, 39.7, 39.4, 26.7, 26.2, 25.8, 17.7, 16.6, 16.1 ppm

**HRMS** (ESI):  $m/z$  calculated 399.2795 for C<sub>28</sub>H<sub>35</sub>N<sub>2</sub> [M+H]<sup>+</sup> found 399.2789  $\Delta$ 1.4 ppm

**UHPLC** (6 min): Rt = 5.30 min

1-dodecyl-2-phenyl-1*H*-benzo[*d*]imidazole (**25**)

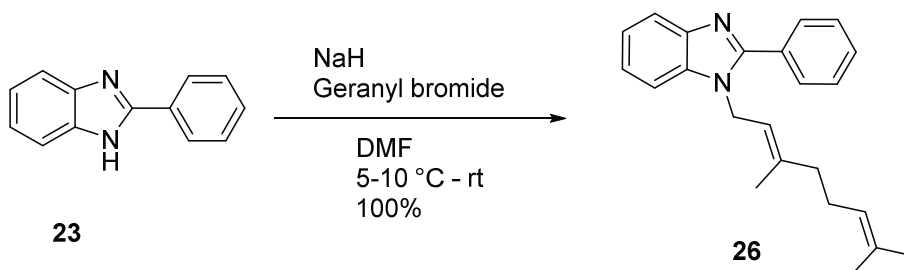
According to **GP B**, 2-phenyl-1*H*-benzo[*d*]imidazole (**23**) (19.4 mg, 0.1 mmol) was reacted with 1-Bromododecane (24.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 1-dodecyl-2-phenyl-1*H*-benzo[*d*]imidazole (**25**) (19.1 mg, 0.05 mmol) with 53% yield as a yellow orange oil.

**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.98 (d, *J* = 7.4 Hz, 1H), 7.55-7.60 (m, 2H), 7.48-7.51 (m, 1H), 7.35-7.46 (m, 4H), 7.11-7.23 (m, 2H), 4.16 (t, *J* = 7.4 Hz, 2H), 1.65-1.73 (m, 1H), 1.58 (br t, *J* = 6.9 Hz, 2H), 1.04-1.22 (m, 12H), 0.99 (br s, 5H), 0.77 ppm (t, *J* = 7.0 Hz, 3H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 153.6, 141.9, 135.1, 129.9, 129.1, 128.8, 128.6, 126.4, 122.8, 122.5, 122.4, 118.4, 110.5, 47.7, 47.1, 44.1, 33.0, 32.6, 31.7, 29.4, 29.2, 29.1, 29.1, 29.2, 29.0, 28.5, 28.5, 27.8, 26.0, 22.4, 13.1 ppm

**HRMS** (ESI): *m/z* calculated 363.2795 for C<sub>25</sub>H<sub>35</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 363.2796 Δ-0.2 ppm

**UHPLC** (6 min): Rt = 5.44 min

*(E)*-1-(3,7-dimethylocta-2,6-dien-1-yl)-2-phenyl-1*H*-benzo[*d*]imidazole (**26**)

According to **GP B**, 2-phenyl-1*H*-benzo[*d*]imidazole (**23**) (19.4 mg, 0.1 mmol) was reacted with Geranyl bromide (21.6 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in (*E*)-



1-(3,7-dimethylocta-2,6-dien-1-yl)-2-phenyl-1*H*-benzo[*d*]imidazole (**26**) (33.2 mg, 0.1 mmol) in quantitative yields as a yellow orange oil.

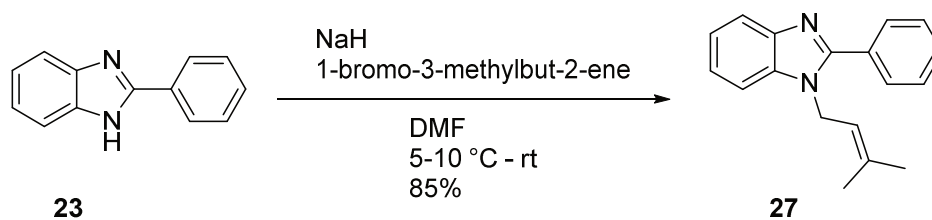
**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 8.09 (d, *J* = 7.3 Hz, 1H), 7.68-7.72 (m, 3H), 7.59-7.63 (m, 1H), 7.36-7.57 (m, 5H), 7.23-7.34 (m, 3H), 5.23-5.43 (m, 1H), 4.95-5.02 (m, 1H), 4.84 (d, *J* = 6.3 Hz, 2H), 1.96-2.16 (m, 4H), 1.64 (s, 2H), 1.58 (s, 2H), 1.52 ppm (s, 2H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 153.5, 141.9, 139.6, 135.2, 131.3, 130.0, 129.9, 129.8, 129.6, 129.2, 128.8, 128.5, 126.4, 123.4, 122.8, 122.5, 122.5, 119.1, 118.4, 110.5, 47.9, 47.7, 42.8, 38.8, 25.7, 24.5, 16.4, 15.1 ppm

**HRMS** (ESI): *m/z* calculated 331.2169 for C<sub>23</sub>H<sub>27</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 331.2169 Δ-0.1 ppm

**UHPLC** (6 min): Rt = 4.40 min

1-(3-methylbut-2-en-1-yl)-2-phenyl-1*H*-benzo[*d*]imidazole (**27**)



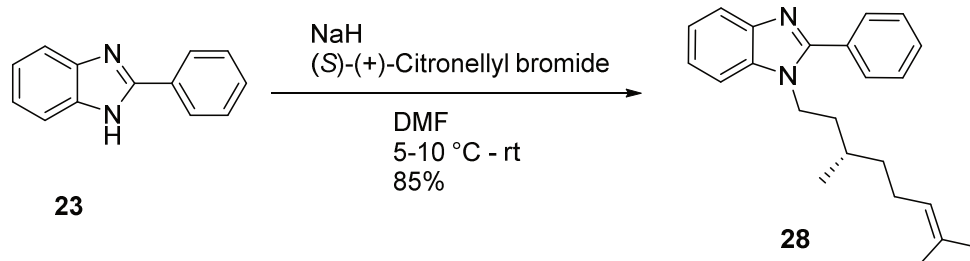
According to **GP B**, 2-phenyl-1*H*-benzo[*d*]imidazole (**23**) (38.8 mg, 0.2 mmol) was reacted with 1-bromo-3-methylbut-2-ene (29.6 mg, 0.2 mmol) utilizing NaH (4.6 mg, 0.2 mmol) in 2 mL DMF. The reaction resulted in 1-(3-methylbut-2-en-1-yl)-2-phenyl-1*H*-benzo[*d*]imidazole (**27**) (44.9 mg, 0.17 mmol) with 85% yield as a pale yellow oil.

**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 8.09 (d, *J* = 7.2 Hz, 1H), 7.66-7.78 (m, 9H), 7.58-7.63 (m, 1H), 7.38-7.57 (m, 12H), 7.23-7.33 (m, 7H), 5.22-5.30 (m, 3H), 4.89-4.92 (m, 1H), 4.81 (br d, *J* = 6.4 Hz, 5H), 1.70 (s, 9H), 1.65 ppm (s, 8H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 153.5, 141.9, 136.3, 135.2, 130.0, 129.9, 129.7, 129.2, 128.8, 128.5, 126.4, 122.9, 122.5, 122.5, 119.0, 118.4, 110.5, 42.8, 24.3, 24.2, 16.8 ppm

**HRMS** (ESI): *m/z* calculated 263.1543 for C<sub>18</sub>H<sub>19</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 263.1528 Δ1.1 ppm

**UHPLC** (6 min): Rt = 3.30 min

*(S)*-1-(3,7-dimethyloct-6-en-1-yl)-2-phenyl-1*H*-benzo[*d*]imidazole (**28**)

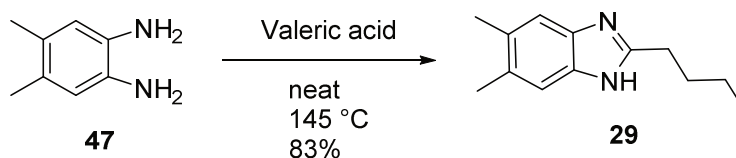
According to **GP B**, 2-phenyl-1*H*-benzo[*d*]imidazole (**23**) (19.4 mg, 0.1 mmol) was reacted with (*S*)-(+)-citronellyl bromide (21.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in (*S*)-1-(3,7-dimethyloct-6-en-1-yl)-2-phenyl-1*H*-benzo[*d*]imidazole (**28**) (28.3 mg, 0.09 mmol) with 85% yield as a pale yellow oil.

**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.97 (d, *J* = 7.2 Hz, 1H), 7.55-7.61 (m, 5H), 7.44-7.50 (m, 6H), 7.26-7.44 (m, 4H), 7.16-7.24 (m, 4H), 7.11-7.15 (m, 1H), 4.81-4.82 (m, 1H), 4.79 (br d, *J* = 1.5 Hz, 1H), 4.11-4.24 (m, 4H), 1.58-1.77 (m, 5H), 1.50 (s, 5H), 1.32-1.48 (m, 7H), 1.02-1.24 (m, 4H), 0.91-0.99 (m, 2H), 0.64 ppm (d, *J* = 6.6 Hz, 5H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 153.6, 142.0, 135.1, 130.8, 130.0, 130.0, 129.9, 129.1, 128.8, 128.6, 126.4, 124.0, 122.9, 122.6, 122.4, 118.5, 110.4, 47.9, 47.7, 42.4, 36.3, 35.9, 29.6, 24.8, 24.5, 18.3, 16.4 ppm

**HRMS** (ESI): *m/z* calculated 333.2325 for C<sub>23</sub>H<sub>29</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 333.2330 Δ-1.3 ppm

**UHPLC** (6 min): Rt = 4.49 min

2-butyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**29**)

According to **GP A**, 4,5-dimethyl-1,2-phenylenediamine (**47**) (267.3 mg, 1.96 mmol) was reacted with valeric acid (2.00 mL, 18.36 mmol). The reaction resulted in 2-butyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**29**) (328 mg, 1.62 mmol) in 83% yield as an orange solid.

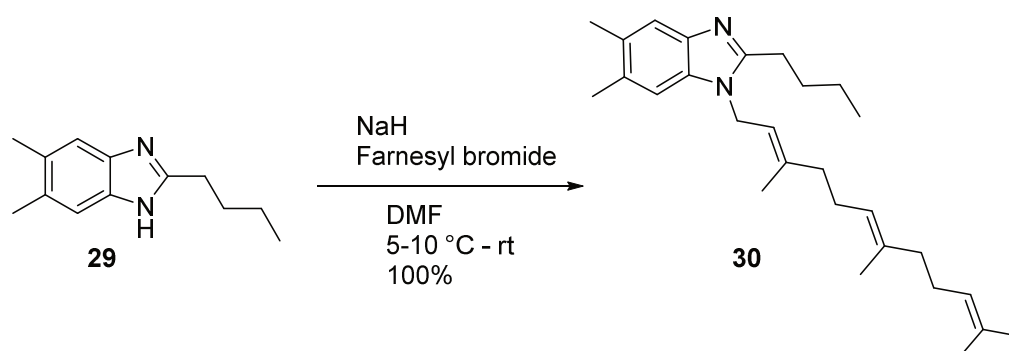
$^1\text{H NMR}$  (500 MHz, CHLOROFORM-*d*, 25°C):  $\delta$  = 7.33 (s, 2H), 2.91 (t,  $J$  = 7.7 Hz, 2H), 2.35 (s, 6H), 1.82 (quin,  $J$  = 7.7 Hz, 2H), 1.35-1.44 (m, 2H), 0.89 ppm (t,  $J$  = 7.4 Hz, 3H)

$^{13}\text{C NMR}$  (126 MHz, CHLOROFORM-*d*, 25°C):  $\delta$  = 154.6, 130.8, 114.8, 30.4, 29.0, 22.4, 20.3, 13.7 ppm

HRMS (ESI):  $m/z$  calculated 203.1543 for  $\text{C}_{13}\text{H}_{19}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 203.1544  $\Delta$ -0.6 ppm

UHPLC (6 min):  $R_t$  = 2.70 min

2-butyl-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**30**)



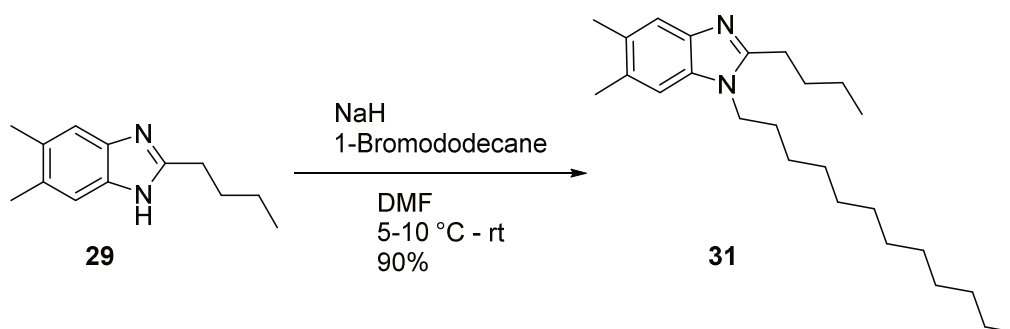
According to **GP B**, 2-butyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**29**) (20.2 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 2-butyl-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**30**) (43.0 mg, 0.1 mmol) in quantitative yields as a yellow orange oil.

$^1\text{H NMR}$  (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C):  $\delta$  = 7.31-7.37 (m, 1H), 7.11-7.15 (m, 1H), 5.10-5.18 (m, 1H), 4.94-5.06 (m, 2H), 4.77 (br d,  $J$  = 6.6 Hz, 2H), 2.86 (br t,  $J$  = 7.8 Hz, 2H), 2.34-2.39 (m, 3H), 2.33 (s, 4H), 2.02-2.17 (m, 4H), 1.91-1.96 (m, 2H), 1.90 (br s, 1H), 1.88 (s, 2H), 1.80-1.87 (m, 2H), 1.73-1.80 (m, 3H), 1.61 (s, 3H), 1.52 (br s, 2H), 1.52 (br s, 2H), 1.39-1.47 (m, 2H), 0.97 ppm (t,  $J$  = 7.4 Hz, 4H)

$^{13}\text{C NMR}$  (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C):  $\delta$  = 155.6, 141.1, 140.7, 136.6, 134.6, 132.8, 132.3, 132.2, 125.5, 124.8, 120.9, 119.1, 111.5, 50.0, 42.9, 40.9, 40.4, 31.6, 31.3, 29.4, 28.0, 27.9, 27.0, 26.0, 25.1, 23.8, 23.5, 20.8, 20.5, 20.5, 17.9, 16.9, 16.3, 14.3, 14.2 ppm

HRMS (ESI):  $m/z$  calculated 407.3421 for  $\text{C}_{28}\text{H}_{43}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 407.3414  $\Delta$ 1.6 ppm

UHPLC (6 min):  $R_t$  = 5.38 min

2-butyl-1-dodecyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**31**)

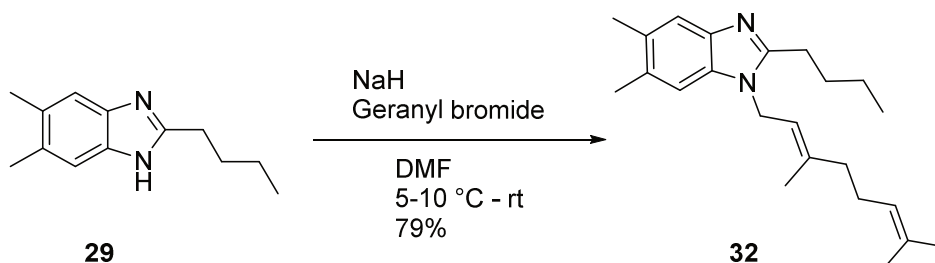
According to **GP B**, 2-butyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**29**) (20.2 mg, 0.1 mmol) was reacted with 1-Bromododecane (24.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 2-butyl-1-dodecyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**31**) (33.4 mg, 0.09 mmol) with 90% yield as a yellow orange oil.

<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.32 (s, 1H), 7.30-7.35 (m, 1H), 7.17 (s, 1H), 4.11 (t, *J* = 7.4 Hz, 2H), 2.80-2.88 (m, 2H), 2.36 (s, 3H), 2.31-2.34 (m, 4H), 1.73-1.83 (m, 5H), 1.40-1.50 (m, 2H), 1.32 (br d, *J* = 4.3 Hz, 4H), 1.25 (s, 11H), 0.95-1.02 (m, 4H), 0.85-0.92 ppm (m, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 155.7, 141.6, 134.7, 132.6, 132.1, 119.3, 111.4, 44.6, 33.2, 31.7, 31.4, 31.0, 30.9, 30.7, 30.7, 30.6, 30.4, 29.6, 27.9, 27.9, 23.9, 23.7, 23.5, 20.7, 20.5, 20.5, 14.6, 14.3 ppm

HRMS (ESI): *m/z* calculated 371.3421 for C<sub>25</sub>H<sub>43</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 371.3414 Δ1.7 ppm

UHPLC (6 min): Rt = 5.54 min

*(E)*-2-butyl-1-(3,7-dimethylocta-2,6-dien-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**32**)

According to **GP B**, 2-butyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**29**) (20.2 mg, 0.1 mmol) was reacted with Geranyl bromide (21.6 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted

in (*E*)-2-butyl-1-(3,7-dimethylocta-2,6-dien-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**32**) (26.9 mg, 0.08 mmol) with 79% yield as a yellow orange oil.

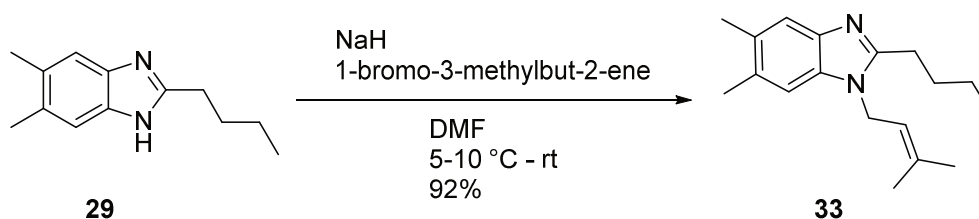
<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.33 (s, 1H), 7.13 (s, 1H), 5.11-5.16 (m, 1H), 4.95-5.03 (m, 1H), 4.70-4.82 (m, 3H), 2.85 (br t, *J* = 7.8 Hz, 3H), 2.31-2.37 (m, 8H), 2.01-2.14 (m, 5H), 1.87 (s, 3H), 1.73-1.79 (m, 3H), 1.56 (s, 3H), 1.52 (s, 3H), 1.38-1.48 (m, 3H), 0.97 ppm (t, *J* = 7.3 Hz, 4H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 155.7, 141.3, 140.7, 134.6, 132.8, 132.7, 132.3, 124.9, 120.8, 119.1, 111.5, 49.8, 49.6, 49.4, 49.3, 42.9, 40.4, 31.3, 28.0, 27.3, 25.9, 23.8, 20.8, 20.5, 17.9, 16.8, 14.3 ppm

HRMS (ESI): *m/z* calculated 339.2795 for C<sub>23</sub>H<sub>35</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 339.2791 Δ0.1 ppm

UHPLC (6 min): Rt = 4.70 min

2-butyl-5,6-dimethyl-1-(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazole (**33**)



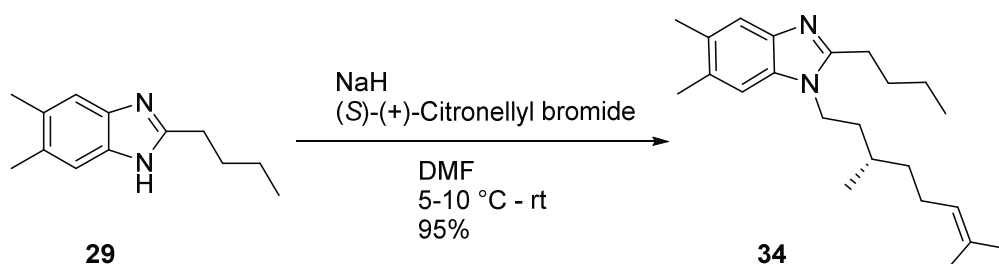
According to **GP B**, 2-butyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**29**) (20.2 mg, 0.1 mmol) was reacted with 1-bromo-3-methylbut-2-ene (14.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 2-butyl-5,6-dimethyl-1-(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazole (**33**) (25 mg, 0.09 mmol) with 92% yield as a pale yellow oil.

<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.34 (s, 1H), 7.15 (s, 1H), 5.18 (ddt, *J* = 7.9, 5.2, 1.4 Hz, 1H), 4.77 (d, *J* = 6.6 Hz, 1H), 2.78-2.94 (m, 2H), 2.48 (s, 1H), 2.33-2.40 (m, 5H), 1.90 (s, 2H), 1.69-1.88 (m, 4H), 1.37-1.60 (m, 2H), 0.85-1.07 ppm (m, 3H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 154.2, 139.9, 135.8, 133.1, 131.1, 130.6, 119.2, 117.6, 109.8, 48.1, 47.8, 47.4, 47.6, 47.1, 41.3, 29.8, 26.4, 24.3, 22.2, 18.9, 16.8, 12.7 ppm

HRMS (ESI): *m/z* calculated 271.2169 for C<sub>18</sub>H<sub>27</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 271.2163 Δ2.1 ppm

UHPLC (6 min): Rt = 3.82 min

*(S)*-2-butyl-1-(3,7-dimethyloct-6-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**34**)

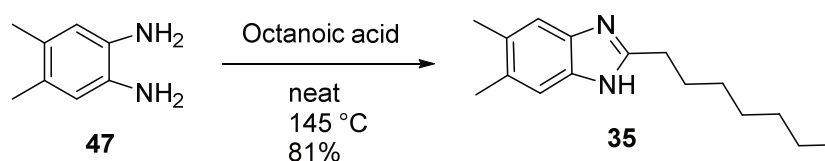
According to **GP B**, 2-butyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**29**) (20.2 mg, 0.1 mmol) was reacted with (*S*)-(+)-Citronellyl bromide (21.8 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in (*S*)-2-butyl-1-(3,7-dimethyloct-6-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**34**) (32.4 mg, 0.1 mmol) with 95% yield as a yellow orange oil.

<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.32 (s, 1H), 7.15 (s, 1H), 5.02-5.09 (m, 1H), 4.88 (s, 3H), 4.12 (qdd, *J* = 14.7, 9.2, 5.9 Hz, 2H), 2.80-2.89 (m, 3H), 2.37 (s, 3H), 2.33 (s, 3H), 2.31 (s, 2H), 1.91-2.05 (m, 2H), 1.78-1.84 (m, 2H), 1.78 (br s, 2H), 1.63-1.66 (m, 3H), 1.56 (s, 4H), 1.46 (br d, *J* = 7.6 Hz, 2H), 1.38 (br d, *J* = 5.6 Hz, 1H), 1.20-1.29 (m, 2H), 1.02 (d, *J* = 6.4 Hz, 3H), 0.97-1.01 ppm (m, 2H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 155.6, 141.6, 134.6, 132.7, 132.5, 132.1, 125.6, 119.3, 111.3, 42.9, 38.1, 37.9, 31.6, 31.4, 27.9, 26.5, 26.0, 23.7, 20.7, 20.5, 20.2, 17.9, 14.3 ppm

HRMS (ESI): *m/z* calculated 341.2951 for C<sub>23</sub>H<sub>37</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 341.2944 Δ2.1 ppm

UHPLC (6 min): Rt = 4.83 min

2-heptyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**35**)

According to **GP A**, 4,5-dimethyl-1,2-phenylenediamine (**47**) (302.0 mg, 2.22 mmol) was reacted with octanoic acid (3.5 mL, 22.09 mmol). The reaction resulted in 2-heptyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**35**) (441 mg, 1.80 mmol) in 81% yield as a yellow orange solid.

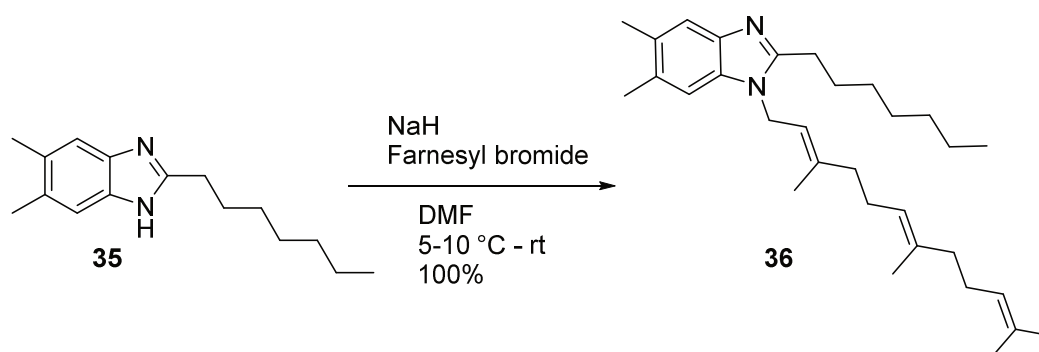
<sup>1</sup>H NMR (500 MHz, CHLOROFORM-*d*, 25°C): δ = 7.32 (s, 2H), 2.89 (t, *J* = 7.8 Hz, 2H), 2.35 (s, 6H), 1.83 (quin, *J* = 7.6 Hz, 2H), 1.18-1.41 (m, 9H), 0.85 ppm (t, *J* = 6.9 Hz, 3H)

$^{13}\text{C}$  NMR (126 MHz, CHLOROFORM-*d*, 25°C):  $\delta$  = 154.3, 130.9, 114.8, 31.7, 29.3, 29.3, 29.0, 28.3, 22.6, 20.3, 14.0 ppm

HRMS (ESI):  $m/z$  calculated 245.2012 for  $\text{C}_{16}\text{H}_{25}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 245.2013  $\Delta$ -0.1 ppm

UHPLC (6 min):  $R_t$  = 3.53 min

2-heptyl-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**36**)



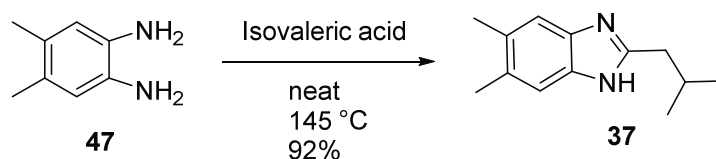
According to **GP B**, 2-heptyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**35**) (24.4 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 2-heptyl-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**30**) (47 mg, 0.1 mmol) in quantitative yield as a yellow orange oil.

$^1\text{H}$  NMR (500 MHz, CHLOROFORM-*d*, 19°C):  $\delta$  = 7.47 (br s, 1H), 7.03 (br s, 1H), 5.17 (br s, 1H), 5.06 (br s, 2H), 4.68 (br s, 2H), 2.75-2.88 (m, 2H), 2.37 (br s, 7H), 2.07-2.16 (m, 1H), 1.99-2.12 (m, 6H), 1.90-1.99 (m, 3H), 1.86 (br s, 6H), 1.63-1.71 (m, 4H), 1.58 (br s, 6H), 1.41 (br s, 4H), 1.29 (br s, 7H), 0.81-0.96 ppm (m, 4H)

$^{13}\text{C}$  NMR (126 MHz, CHLOROFORM-*d*, 18°C):  $\delta$  = 154.3, 141.2, 139.0, 135.6, 133.5, 131.3, 130.6, 130.2, 124.2, 123.4, 119.5, 119.2, 109.6, 41.7, 39.6, 39.3, 31.7, 29.5, 29.1, 27.9, 27.6, 26.7, 26.2, 25.7, 22.6, 20.6, 20.2, 17.7, 16.6, 16.0, 14.1 ppm

HRMS (ESI):  $m/z$  calculated 449.3890 for  $\text{C}_{31}\text{H}_{49}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 449.3885  $\Delta$ 1.2 ppm

UHPLC (6 min):  $R_t$  = 5.54 min

2-isobutyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**37**)

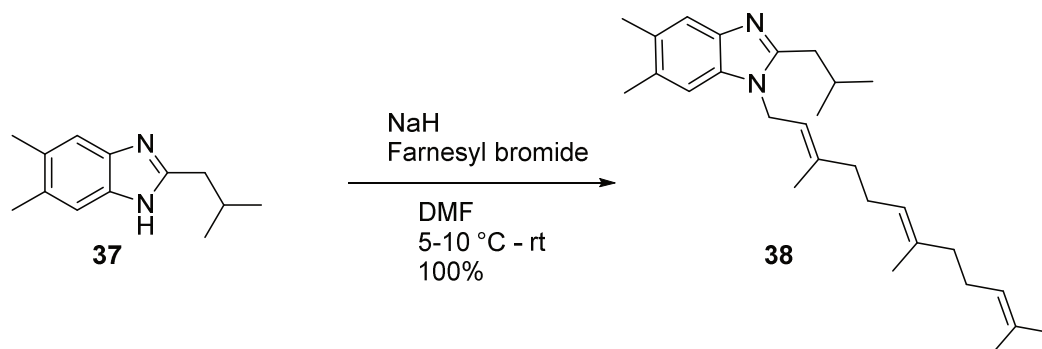
According to **GP A**, 4,5-dimethyl-1,2-phenylenediamine (**47**) (267.1 mg, 1.96 mmol) was reacted with isovaleric acid (2.0 mL, 18.11 mmol). The reaction resulted in 2-isobutyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**35**) (365.0 mg, 1.80 mmol) in 92% yield as a yellow orange solid.

**<sup>1</sup>H NMR** (500 MHz, CHLOROFORM-*d*, 25°C):  $\delta$  = 7.33 (s, 2H), 2.78 (d,  $J$  = 7.3 Hz, 2H), 2.36 (s, 6H), 2.15-2.30 (m, 1H), 0.98 ppm (d,  $J$  = 6.7 Hz, 6H)

**<sup>13</sup>C NMR** (126 MHz, CHLOROFORM-*d*, 25°C):  $\delta$  = 153.6, 130.8, 114.8, 38.5, 28.6, 22.5, 20.3 ppm

**HRMS** (ESI):  $m/z$  calculated 203.1543 for C<sub>13</sub>H<sub>19</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 203.1545  $\Delta$ -1.1 ppm

**UHPLC** (6 min): Rt = 2.64 min

2-isobutyl-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**38**)

According to **GP B**, 2-isobutyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**37**) (20.2 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 2-isobutyl-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**38**) (40.7 mg, 0.1 mmol) in quantitative yield as a yellow oil.

**<sup>1</sup>H NMR** (500 MHz, CHLOROFORM-*d*, 25°C):  $\delta$  = 7.49 (s, 1H), 7.04 (s, 1H), 5.14-5.19 (m, 1H), 5.02-5.09 (m, 1H), 4.69 (br d,  $J$  = 6.0 Hz, 2H), 2.72 (br d,  $J$  = 7.3 Hz, 2H), 2.37 (s, 3H), 2.36 (s, 3H), 2.20-2.27 (m, 1H), 2.07-



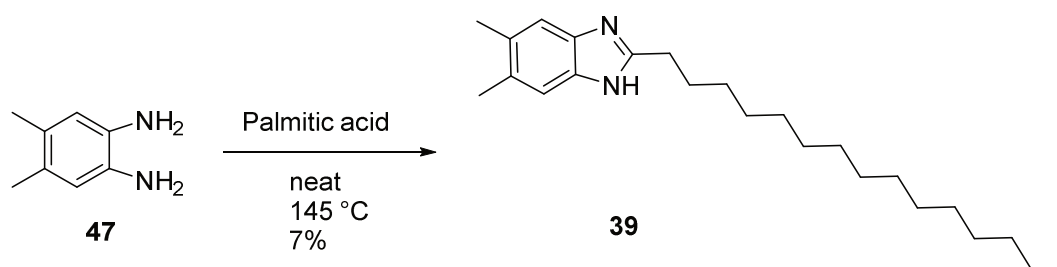
2.13 (m, 2H), 1.98-2.06 (m, 3H), 1.95 (br d,  $J = 8.2$  Hz, 2H), 1.86 (s, 2H), 1.67 (s, 3H), 1.59 (s, 3H), 1.57 (s, 2H), 1.25-1.30 (m, 1H), 1.01 ppm (d,  $J = 6.6$  Hz, 7H)

$^{13}\text{C}$  NMR (126 MHz, CHLOROFORM-*d*, 25°C):  $\delta = 153.4, 140.8, 139.0, 135.6, 133.4, 131.3, 130.8, 130.5, 124.2, 123.3, 119.4, 119.1, 109.8, 41.8, 39.6, 39.3, 36.3, 28.2, 26.7, 26.2, 25.7, 22.6, 20.5, 20.2, 17.6, 16.6, 16.0$  ppm

HRMS (ESI):  $m/z$  calculated 407.3421 for  $\text{C}_{28}\text{H}_{43}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 407.3432  $\Delta$ -2.8 ppm

UHPLC (6 min):  $R_t = 5.12$  min

5,6-dimethyl-2-pentadecyl-1*H*-benzo[*d*]imidazole (**39**)



According to **GP A**, 4,5-dimethyl-1,2-phenylenediamine (**47**) (301.5 mg, 2.21 mmol) was reacted with palmitic acid (5.65 g, 22.03 mmol). The reaction resulted in 5,6-dimethyl-2-pentadecyl-1*H*-benzo[*d*]imidazole (**39**) (52 mg, 0.15 mmol) in 7% yield as a pale yellow solid.

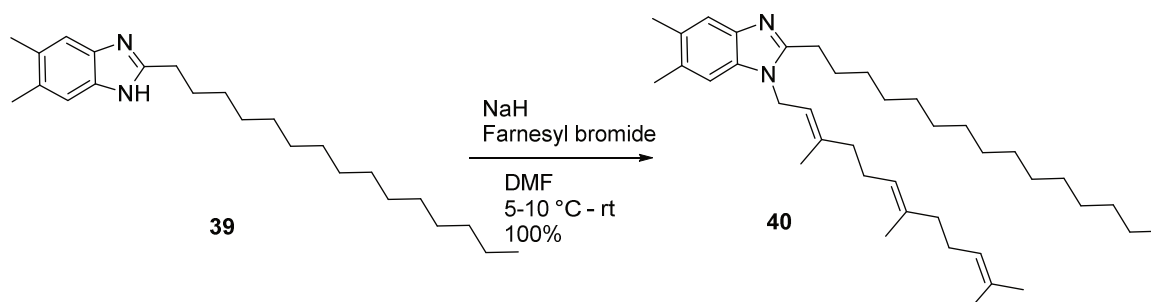
$^1\text{H}$  NMR (500 MHz, CHLOROFORM-*d*, 25°C):  $\delta = 7.32$  (s, 2H), 2.88 (t,  $J = 7.7$  Hz, 2H), 2.35 (s, 6H), 1.82 (quin,  $J = 7.6$  Hz, 2H), 1.19-1.44 (m, 30H), 0.83-0.96 ppm (m, 5H)

$^{13}\text{C}$  NMR (126 MHz, CHLOROFORM-*d*, 25°C):  $\delta = 154.2, 131.0, 114.8, 31.9, 29.7, 29.6, 29.5, 29.3, 29.3, 29.3, 28.3, 23.9, 22.7, 20.3, 14.1$  ppm

HRMS (ESI):  $m/z$  calculated 357.3264 for  $\text{C}_{24}\text{H}_{41}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 357.3264  $\Delta$ 0.2 ppm

UHPLC (6 min):  $R_t = 5.27$  min

5,6-dimethyl-2-pentadecyl-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**40**)



According to **GP B**, 5,6-dimethyl-2-pentadecyl-1*H*-benzo[*d*]imidazole (**39**) (35.6 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 5,6-dimethyl-2-pentadecyl-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**40**) (28.5 mg, 0.1 mmol) in quantitative yield as a yellow orange amorphous solid.

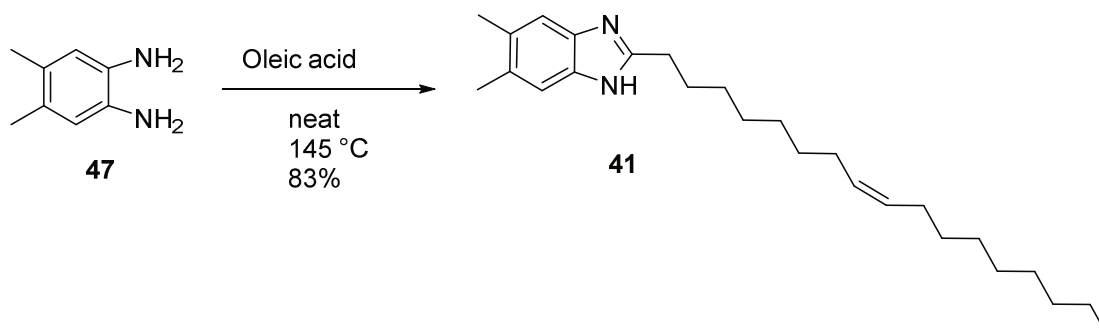
**<sup>1</sup>H NMR** (500 MHz, CHLOROFORM-*d*, 25°C):  $\delta$  = 7.49 (br s, 1H), 7.03 (s, 1H), 5.17 (br t,  $J$  = 5.6 Hz, 1H), 4.98-5.15 (m, 2H), 4.68 (br d,  $J$  = 6.0 Hz, 2H), 2.79-2.88 (m, 3H), 2.37 (s, 3H), 2.35 (s, 3H), 2.06-2.13 (m, 2H), 1.98-2.06 (m, 3H), 1.91-1.96 (m, 2H), 1.86 (s, 2H), 1.83 (br d,  $J$  = 7.6 Hz, 3H), 1.67 (s, 2H), 1.59 (s, 3H), 1.57 (s, 2H), 1.25 (br s, 33H), 0.89 ppm (t,  $J$  = 6.9 Hz, 5H)

**<sup>13</sup>C NMR** (126 MHz, CHLOROFORM-*d*, 25°C):  $\delta$  = 154.1, 140.4, 139.2, 135.6, 133.3, 131.3, 131.0, 130.7, 124.2, 123.3, 119.3, 119.0, 109.7, 41.8, 39.6, 39.3, 31.9, 29.7, 29.6, 29.5, 29.4, 28.0, 27.4, 26.7, 26.2, 25.7, 22.7, 20.5, 20.2, 17.7, 16.7, 16.0, 14.1 ppm

**HRMS** (ESI):  $m/z$  calculated 561.5142 for C<sub>39</sub>H<sub>65</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 561.5144  $\Delta$ -0.3 ppm

**UHPLC** (6 min): Rt = 6.51 min

(*Z*)-2-(heptadec-8-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**41**)



According to **GP A**, 4,5-dimethyl-1,2-phenylenediamine (**47**) (202.0 mg, 1.48 mmol) was reacted with oleic acid (4.6 mL, 14.49 mmol). The reaction resulted in (*Z*)-2-(heptadec-8-en-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**41**) (472 mg, 1.23 mmol) in 83% yield as an orange oil.

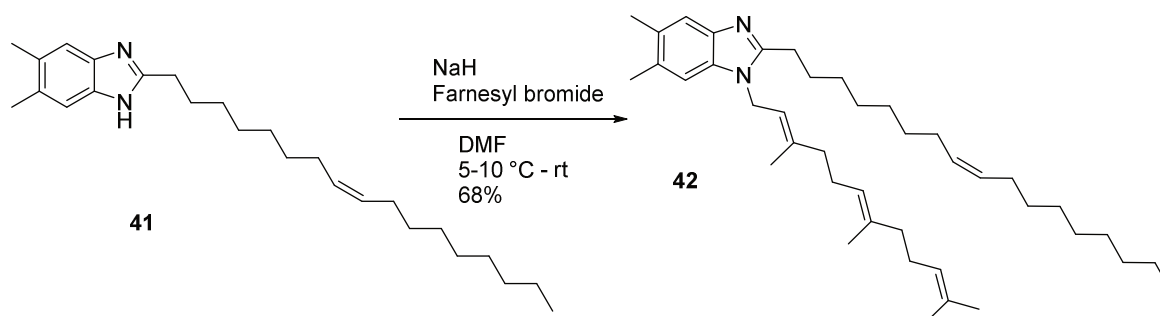
<sup>1</sup>H NMR (500 MHz, CHLOROFORM-*d*, 25°C): δ = 7.32 (s, 2H), 5.27-5.36 (m, 2H), 4.14 (q, *J* = 7.2 Hz, 2H), 2.91 (t, *J* = 7.6 Hz, 2H), 2.34 (s, 6H), 2.06 (s, 3H), 1.93-2.02 (m, 4H), 1.83 (quin, *J* = 7.6 Hz, 2H), 1.19-1.42 (m, 25H), 0.81-1.01 ppm (m, 4H)

<sup>13</sup>C NMR (126 MHz, CHLOROFORM-*d*, 25°C): δ = 154.7, 130.6, 129.9, 129.6, 114.7, 60.4, 31.8, 29.7, 29.7, 29.5, 29.3, 29.3, 29.1, 28.4, 27.1, 27.1, 22.6, 21.0, 20.2, 14.1, 14.0 ppm

HRMS (ESI): *m/z* calculated 383.3421 for C<sub>26</sub>H<sub>43</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 383.3414 Δ1.7 ppm

UHPLC (6 min): Rt = 5.35 min

2-((*Z*)-heptadec-8-en-1-yl)-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**42**)



According to **GP B**, (*Z*)-2-(heptadec-8-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**41**) (38.2 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL

DMF. The reaction resulted in 2-((*Z*)-heptadec-8-en-1-yl)-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**42**) (40.2 mg, 0.07 mmol) with 68% yield as a yellow orange oil.

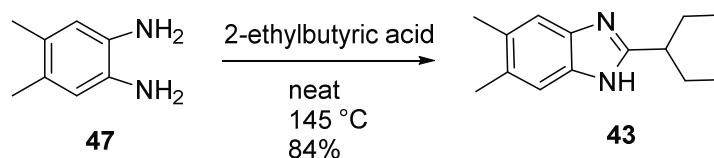
<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 7.23 (s, 1H), 7.15 (s, 1H), 7.02 (s, 1H), 5.22 (t, *J* = 4.9 Hz, 2H), 5.04 (br t, *J* = 6.3 Hz, 1H), 4.90 (q, *J* = 1.0 Hz, 2H), 4.65 (br d, *J* = 6.4 Hz, 2H), 3.25 (s, 7H), 3.20-3.22 (m, 1H), 2.89 (s, 1H), 2.71-2.77 (m, 3H), 2.25 (s, 3H), 2.23 (s, 5H), 1.86-2.03 (m, 10H), 1.76-1.79 (m, 1H), 1.65-1.85 (m, 9H), 1.51 (s, 3H), 1.42 (d, *J* = 2.6 Hz, 5H), 1.12-1.32 (m, 27H), 0.72-0.82 ppm (m, 4H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 154.4, 154.0, 139.7, 139.0, 136.2, 135.0, 133.1, 131.1, 130.8, 130.7, 130.6, 129.5, 129.5, 129.4, 124.0, 123.3, 119.4, 117.6, 114.0, 109.9, 48.5, 48.1, 47.9, 47.7, 47.1, 41.3, 39.4, 38.9, 31.7, 31.7, 29.5, 29.4, 29.4, 29.3, 29.1, 28.9, 29.2, 29.0, 28.8, 28.2, 27.9, 27.6, 26.8, 26.7, 26.7, 26.4, 25.5, 24.6, 22.4, 19.3, 19.0, 19.0, 16.4, 15.4, 14.8, 13.1, 13.1 ppm

HRMS (ESI): *m/z* calculated 587.5299 for C<sub>41</sub>H<sub>67</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 587.5287 Δ2.1 ppm

UHPLC (6 min): Rt = 6.54 min

#### 5,6-dimethyl-2-(pentan-3-yl)-1*H*-benzo[*d*]imidazole (**43**)



According to **GP A**, 4,5-dimethyl-1,2-phenylenediamine (**47**) (299.2 mg, 2.20 mmol) was reacted with 2-ethylbutyric acid (2.8 mL, 22.18 mmol). The reaction resulted in 5,6-dimethyl-2-(pentan-3-yl)-1*H*-benzo[*d*]imidazole (**43**) (398 mg, 1.84 mmol) in 84% yield as a yellow orange solid.

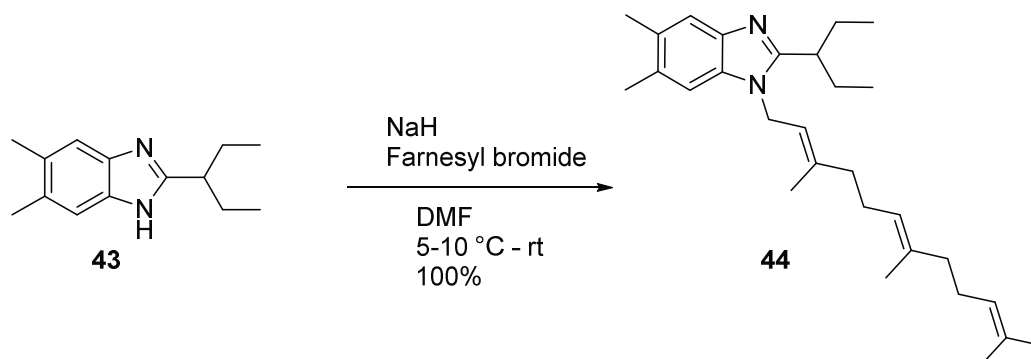
<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.25 (s, 2H), 2.67-2.73 (m, 1H), 2.33 (s, 6H), 1.73-1.83 (m, 4H), 0.83 ppm (t, *J* = 7.4 Hz, 6H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 159.0, 132.0, 115.4, 45.5, 28.8, 20.6, 12.6 ppm

HRMS (ESI): *m/z* calculated 217.1699 for C<sub>14</sub>H<sub>21</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 217.1699 Δ0.0 ppm

UHPLC (6 min): Rt = 2.74 min

5,6-dimethyl-2-(pentan-3-yl)-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**44**)



According to **GP B**, 5,6-dimethyl-2-(pentan-3-yl)-1*H*-benzo[*d*]imidazole (**43**) (21.6 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 5,6-dimethyl-2-(pentan-3-yl)-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**44**) (42.1 mg, 0.1 mmol) with 100% yield as an orange oil.

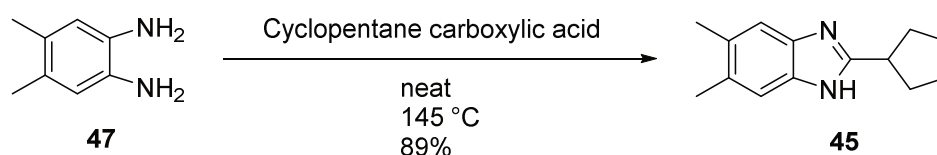
<sup>1</sup>H NMR (500 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 7.26 (s, 1H), 7.02-7.07 (m, 1H), 5.01-5.06 (m, 1H), 4.87-4.93 (m, 2H), 4.73 (br d, *J* = 6.3 Hz, 2H), 3.23-3.27 (m, 3H), 2.70-2.84 (m, 1H), 2.26 (s, 4H), 2.24 (s, 4H), 1.92-2.03 (m, 5H), 1.79-1.90 (m, 5H), 1.78 (br d, *J* = 7.5 Hz, 1H), 1.67-1.77 (m, 6H), 1.51 (s, 3H), 1.42 (d, *J* = 6.6 Hz, 6H), 0.73 ppm (t, *J* = 7.4 Hz, 7H)

<sup>13</sup>C NMR (126 MHz, METHANOL-*d*<sub>4</sub>, 19°C): δ = 157.0, 139.8, 138.8, 135.0, 132.6, 131.1, 130.9, 130.9, 130.6, 124.0, 123.3, 119.6, 117.5, 114.0, 110.1, 47.9, 41.2, 40.4, 39.4, 38.9, 27.4, 26.3, 25.5, 24.5, 19.3, 19.0, 19.0, 16.4, 15.4, 14.7, 11.1 ppm

HRMS (ESI): *m/z* calculated 421.3577 for C<sub>29</sub>H<sub>45</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 421.3584 Δ-1.5 ppm

UHPLC (6 min): Rt = 5.19 min

2-cyclopentyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**45**)



According to **GP A**, 4,5-dimethyl-1,2-phenylenediamine (**47**) (298.3 mg, 2.19 mmol) was reacted with Cyclopentane carboxylic acid (2.5 mL, 23.04 mmol). The reaction resulted in 2-cyclopentyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**45**) (419 mg, 1.96 mmol) in 89% yield as an orange brown solid.

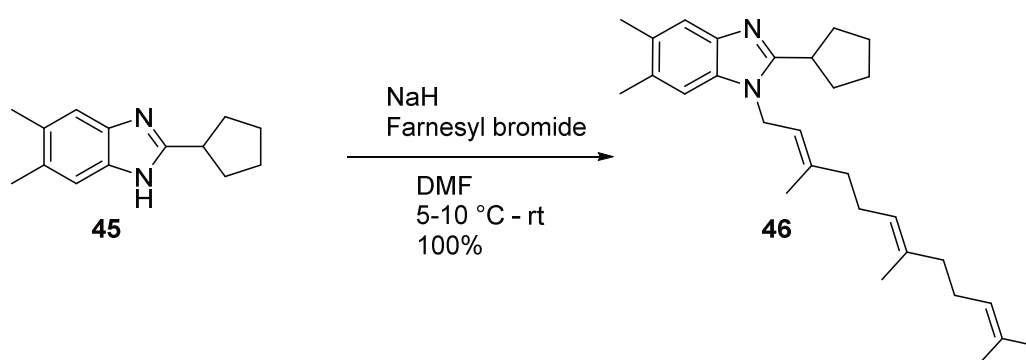
$^1\text{H NMR}$  (500 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 7.23 (s, 2H), 3.17-3.29 (m, 1H), 2.31 (s, 6H), 2.09-2.19 (m, 2H), 1.80-1.92 (m, 4H), 1.67-1.78 ppm (m, 2H)

$^{13}\text{C NMR}$  (126 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 159.5, 132.0, 115.6, 115.6, 41.0, 33.4, 26.6, 20.5 ppm

**HRMS** (ESI):  $m/z$  calculated 215.1543 for  $\text{C}_{14}\text{H}_{19}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 215.1539  $\Delta$ 1.7 ppm

**UHPLC** (6 min):  $R_t$  = 2.64 min

2-cyclopentyl-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**46**)



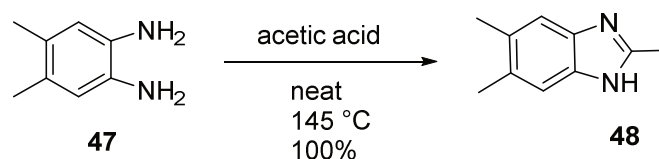
According to **GP B**, 2-cyclopentyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**45**) (21.4 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 2-cyclopentyl-5,6-dimethyl-1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**46**) (41.9 mg, 0.1 mmol) with 100% yield as a yellow oil.

$^1\text{H NMR}$  (500 MHz, METHANOL- $d_4$ , 19°C):  $\delta$  = 7.87 (s, 1H), 7.24 (s, 1H), 7.01 (s, 1H), 5.00-5.09 (m, 1H), 4.86-4.95 (m, 2H), 4.69 (d,  $J$  = 6.3 Hz, 2H), 3.23-3.28 (m, 1H), 2.88 (s, 2H), 2.75 (s, 2H), 2.20-2.26 (m, 7H), 1.91-2.06 (m, 8H), 1.77-1.89 (m, 1H), 1.74-1.80 (m, 6H), 1.66-1.75 (m, 1H), 1.47-1.55 (m, 5H), 1.42 ppm (d,  $J$  = 4.3 Hz, 5H)

$^{13}\text{C NMR}$  (126 MHz, METHANOL- $d_4$ , 19°C):  $\delta$  = 163.4, 157.5, 139.5, 138.9, 135.0, 133.1, 131.2, 130.8, 130.6, 124.0, 123.3, 119.6, 117.5, 109.9, 41.3, 39.4, 39.1, 38.9, 36.8, 31.9, 31.9, 26.3, 25.5, 25.3, 24.5, 19.3, 19.0, 19.0, 16.4, 15.3, 14.8 ppm

**HRMS** (ESI):  $m/z$  calculated 419.3421 for  $\text{C}_{29}\text{H}_{43}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 419.3418  $\Delta$ 0.8 ppm

**UHPLC** (6 min):  $R_t$  = 5.15 min

2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**)

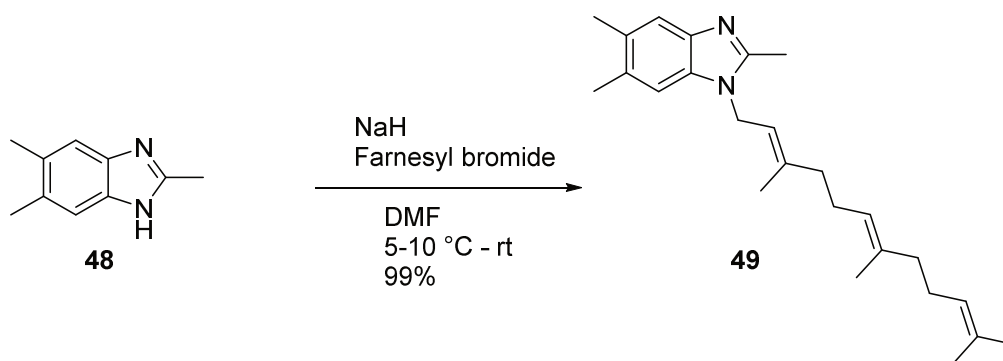
According to **GP A**, 4,5-dimethyl-1,2-phenylenediamine (**47**) (1.02 g, 7.36 mmol) was reacted with acetic acid (20 mL, 73.6 mmol). The reaction resulted in 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (1.18 g, 7.36 mmol) in quantitative yields as a light brown solid.

**<sup>1</sup>H NMR** (500 MHz, DMSO-*d*<sub>6</sub>, 25°C): δ = 11.96 (br s, 1H, NH), 7.20 (br s, 2H), 2.43 (s, 3H), 2.26 ppm (s, 6H)

**<sup>13</sup>C NMR** (126 MHz, DMSO-*d*<sub>6</sub>, 25°C): δ = 150.3, 142.2, 133.0, 129.2, 118.1, 110.9, 19.9, 14.6 ppm

**HRMS** (ESI): *m/z* calculated 161.1073 for C<sub>10</sub>H<sub>13</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 161.1077 Δ-0.4 ppm

**UHPLC** (6 min): Rt = 2.05 min

2,5,6-trimethyl-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**49**)

According to **GP B**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (16.0 mg, 0.1 mmol) was reacted with farnesyl bromide (28.5 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 2,5,6-trimethyl-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**49**) (36.1 mg, 0.1 mmol) with 99% yield as a pale orange oil.

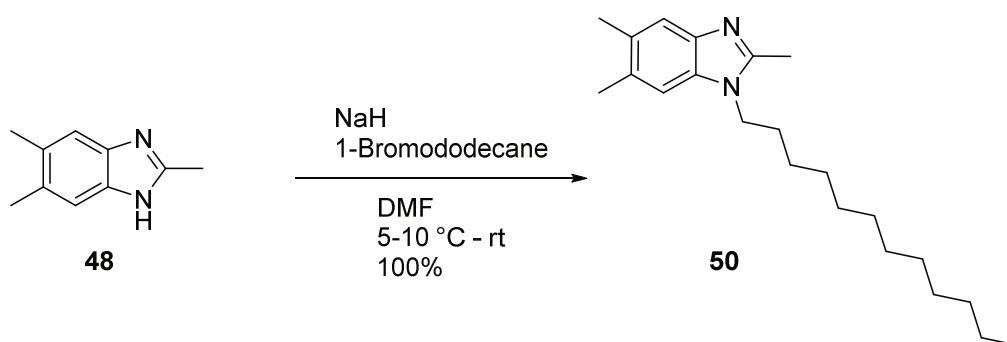
**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C): δ = 7.29 (s, 1H), 7.12 (s, 1H), 5.15 (br t, *J* = 6.48 Hz, 1H), 5.04-5.07 (m, 1H), 5.00 (br t, *J* = 5.87 Hz, 1H), 4.72 (br d, *J* = 6.56 Hz, 2H), 2.51 (s, 3H), 2.34 (s, 3H), 2.31 (s, 3H), 2.02-2.17 (m, 4H), 1.94 (m, 2H), 1.87 (s, 3H), 1.83 (m, 2H), 1.61 (s, 3H), 1.52 (s, 3H), 1.51 ppm (s, 3H)

$^{13}\text{C}$  NMR (126 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 152.0, 141.0, 140.8, 136.6, 134.6, 132.6, 131.7, 131.1, 125.5, 124.8, 120.4, 118.9, 111.4, 42.9, 40.9, 40.5, 27.9, 27.1, 26.0, 20.8, 20.5, 17.9, 16.8, 16.3, 13.5 ppm

HRMS (ESI):  $m/z$  calculated 365.2951 for  $\text{C}_{25}\text{H}_{37}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 365.2956  $\Delta$ -1.3 ppm

UHPLC (6 min):  $R_t$  = 5.17 min

1-dodecyl-2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**50**)



According to **GP B**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (16.0 mg, 0.1 mmol) was reacted with 1-Bromododecane (24.9 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 1-dodecyl-2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**50**) (32.9 mg, 0.1 mmol) in quantitative yields as a pale orange solid.

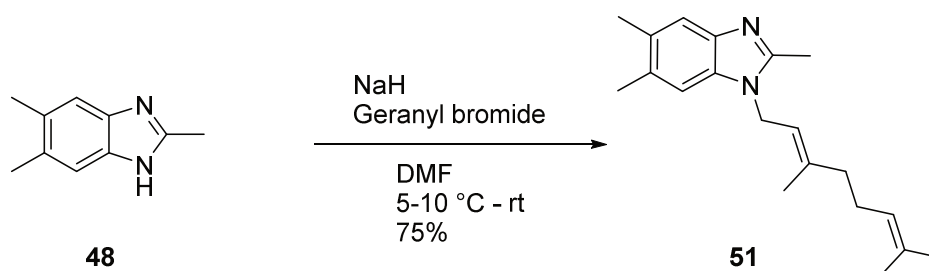
$^1\text{H}$  NMR (500 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 7.30 (s, 1H), 7.16 (s, 1H), 4.09 (t,  $J$ =7.25 Hz, 2H), 2.51 (s, 3H), 2.32 (br s, 3H), 2.31 (s, 3H), 1.78-1.82 (m, 2H), 1.27 (m, 18H), 0.89 ppm (s, 3H)

$^{13}\text{C}$  NMR (126 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 151.9, 141.0, 134.7, 132.7, 132.3, 119.0, 111.3, 44.8, 30.9, 30.9, 30.0, 30.0, 30.0, 30.0, 29.4, 29.4, 27.9, 23.9, 20.7, 20.5, 14.6, 13.5 ppm

HRMS (ESI):  $m/z$  calculated 329.2951 for  $\text{C}_{22}\text{H}_{37}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 329.2954  $\Delta$ -0.8 ppm

UHPLC (6 min):  $R_t$  = 5.17 min



*(E)*-1-(3,7-dimethylocta-2,6-dien-1-yl)-2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**51**)

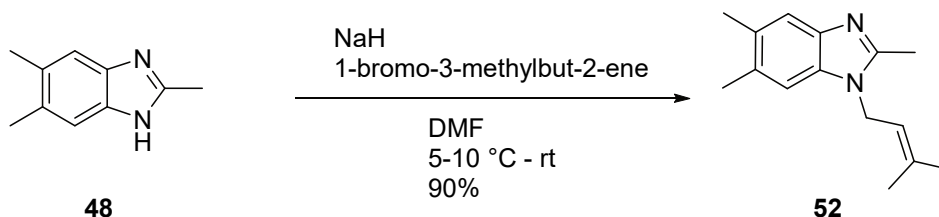
According to **GP B**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (16.0 mg, 0.1 mmol) was reacted with Geranyl bromide (21.7 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in (*E*)-1-(3,7-dimethylocta-2,6-dien-1-yl)-2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**51**) (22.4 mg, 0.08 mmol) with 75% yield as a pale orange oil.

**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C):  $\delta$  = 7.29 (s, 1H), 7.12 (s, 1H), 5.14 (br t, *J* = 6.56 Hz, 1H), 4.99 (br t, *J* = 6.56 Hz, 1H), 4.71 (d, *J* = 6.56 Hz, 2H), 2.50 (s, 3H), 2.35 (s, 3H), 2.32 (s, 3H), 2.02-2.11 (m, 4H), 1.85 (s, 3H), 1.56 (s, 3H), 1.52 ppm (s, 3H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C):  $\delta$  = 152.1, 141.1, 141.0, 134.7, 132.8, 132.6, 132.3, 132.2, 125.0, 120.4, 119.0, 111.3, 42.8, 40.5, 27.3, 25.9, 20.7, 20.4, 17.9, 16.7, 13.5

**HRMS** (ESI): *m/z* calculated 297.2325 for C<sub>20</sub>H<sub>29</sub>N<sub>2</sub> [M+H]<sup>+</sup>; found 297.2329  $\Delta$ -1.3 ppm

**UHPLC** (6 min): Rt = 4.34 min

2,5,6-trimethyl-1-(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazole (**52**)

According to **GP B**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (16.0 mg, 0.1 mmol) was reacted with 1-bromo-3-methylbut-2-ene (14.9 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 2,5,6-trimethyl-1-(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazole (**52**) (20.7 mg, 0.09 mmol) with 90% yield as a pale yellow oil.

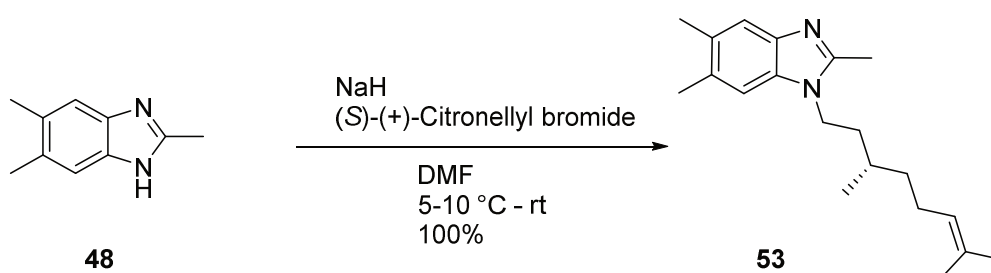
$^1\text{H NMR}$  (500 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 7.29 (s, 1H), 7.11 (s, 1H), 5.16 (br t,  $J$ =6.18 Hz, 1H), 4.69 (br d,  $J$  = 6.71 Hz, 2H), 2.49 (s, 3H), 2.34 (s, 3H), 2.31 (s, 3H), 1.87 (s, 3H), 1.74 (s, 3H)

$^{13}\text{C NMR}$  (126 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 152.1, 141.2, 137.6, 134.7, 132.7, 132.2, 120.3, 119.0, 111.3, 42.9, 25.8, 20.7, 20.5, 18.3, 13.5

**HRMS** (ESI):  $m/z$  calculated 229.1699 for  $\text{C}_{15}\text{H}_{21}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 229.1700  $\Delta$ -0.4 ppm

**UHPLC** (6 min):  $R_t$  = 3.27 min

(*S*)-2-butyl-1-(3,7-dimethyloct-6-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**53**)



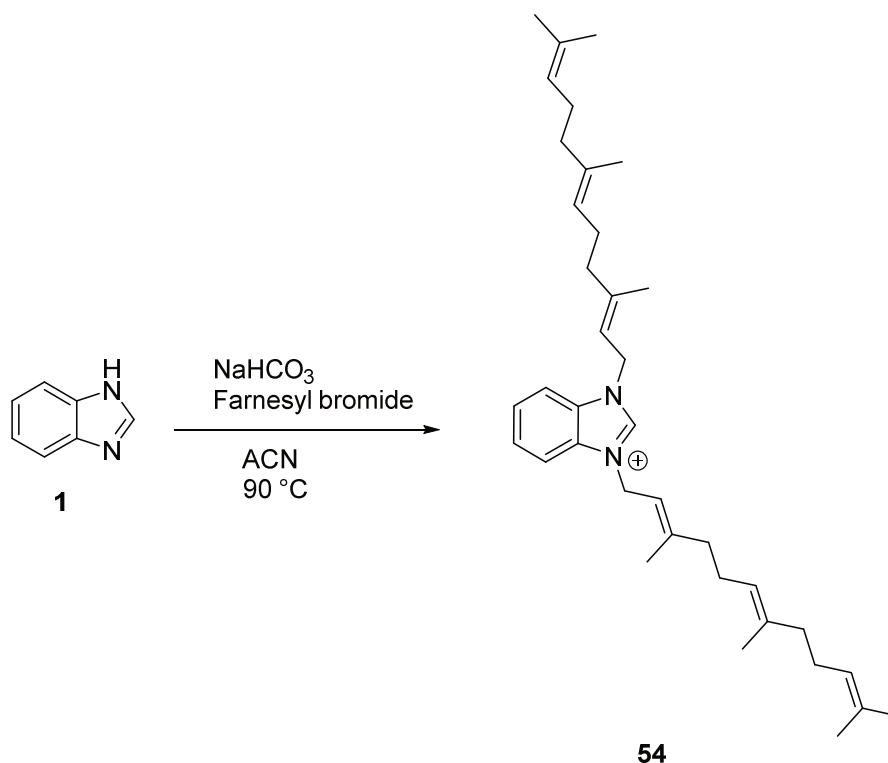
According to **GP B**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (16.0 mg, 0.1 mmol) was reacted with (*S*)-(+)-Citronellyl bromide (21.9 mg, 0.1 mmol) utilizing NaH (2.3 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in (*S*)-1-(3,7-dimethyloct-6-en-1-yl)-2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**53**) (29.9 mg, 0.1 mmol) in quantitative yields as a pale orange oil.

$^1\text{H NMR}$  (500 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 7.29 (s, 1H), 7.12 (s, 1H), 5.00-5.06 (m, 1H), 4.00-4.13 (m, 2H), 2.51 (s, 3H), 2.36 (s, 3H), 2.32 (s, 3H), 1.88-2.04 (m, 3H), 1.70-1.78 (m, 1H), 1.63 (s, 3H), 1.55 (s, 3H), 1.33-1.41 (m, 2H), 1.19-1.25 (m, 1H), 1.00 ppm (d,  $J$  = 6.41 Hz, 3H)

$^{13}\text{C NMR}$  (126 MHz, METHANOL- $d_4$ , 25°C):  $\delta$  = 152.0, 141.2, 134.6, 132.7, 132.5, 132.2, 125.6, 119.1, 111.2, 43.0, 38.0, 37.6, 31.5, 26.5, 26.0, 20.7, 20.5, 20.1, 17.9, 13.4

**HRMS** (ESI):  $m/z$  calculated 299.2482 for  $\text{C}_{20}\text{H}_{31}\text{N}_2$   $[\text{M}+\text{H}]^+$ ; found 299.2485  $\Delta$ -1.2 ppm

**UHPLC** (9 min):  $R_t$  = 4.46 min

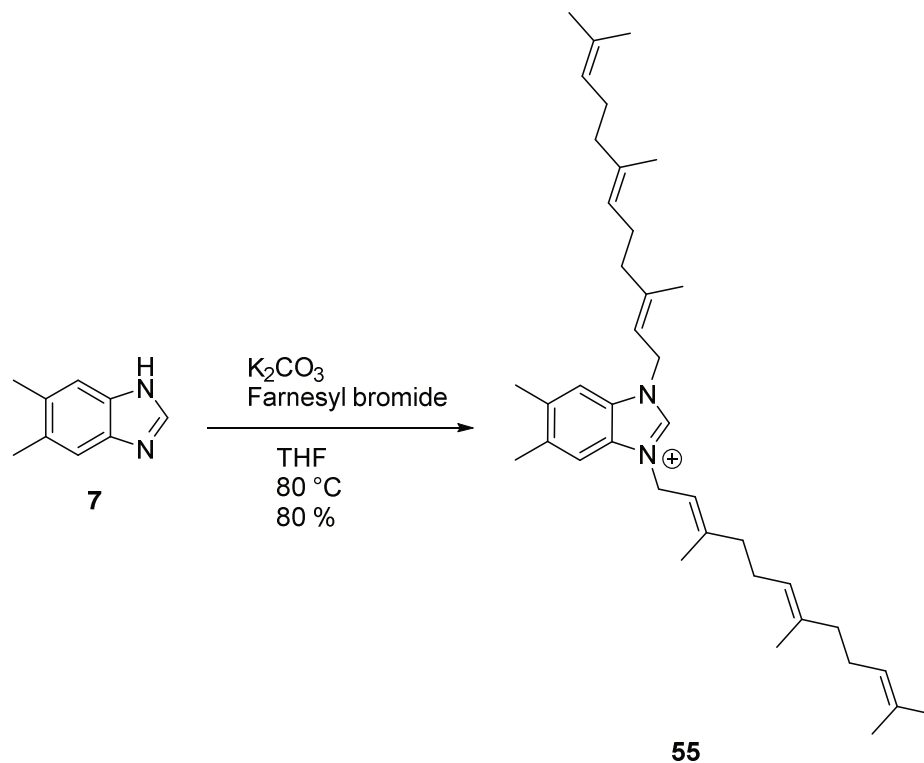
1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**54**)

According to **GP D**, 1*H*-benzo[*d*]imidazole (**1**) (20.0 mg, 0.17 mmol) was reacted with farnesyl bromide (241.4 mg, 0.85 mmol) utilizing NaHCO<sub>3</sub> (28.4 mg, 0.34 mmol) in 3.38 mL ACN. The reaction resulted in 1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**54**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 527.4360 for C<sub>37</sub>H<sub>59</sub>N<sub>2</sub> [M]<sup>+</sup> found 527.4365  $\Delta$ -1.1 ppm

**UHPLC** (9 min): Rt = 7.95 min

5,6-dimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**55**) = Sandacrabin B



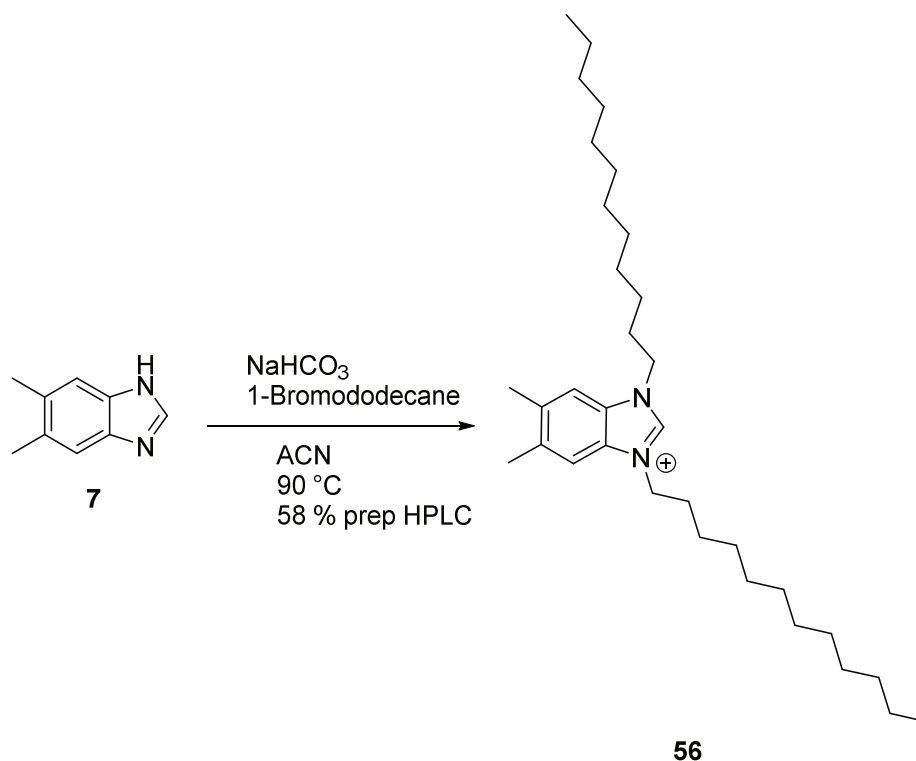
According to **GP C**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (6.6 mg, 0.05 mmol) was reacted with farnesyl bromide (50.0 mg, 0.18 mmol) utilizing  $K_2CO_3$  (6.2 mg, 0.05 mmol) in 2 mL THF. The reaction resulted in 5,6-dimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**55**) (14.0 mg, 0.04 mmol) with 80% yield as a pale yellow oil.

$^1H$  NMR (500 MHz, METHANOL- $d_4$ , 25°C)  $\delta$  = 9.27 (s, 1H), 7.64 (s, 2H), 5.45-5.52 (m, 2H), 5.08 (d,  $J$  = 7.1 Hz), 5.02-5.06 (m, 2H), 4.99-5.03 (m, 2H), 2.48 (s, 6H), 2.15-2.21 (m, 8H), 1.93-1.98 (m, 4H), 1.93 (br s, 6H), 1.83-1.88 (m, 4H), 1.64 (m, 6H), 1.54-1.57 ppm (m, 12H)

$^{13}C$  NMR (126 MHz, METHANOL- $d_4$ , 25°C)  $\delta$  = 146.1, 140.9, 138.8, 136.9, 132.3, 131.7, 125.5, 124.7, 117.5, 114.5, 46.3, 40.9, 40.6, 27.9, 27.1, 26.1, 20.8, 17.9, 16.9, 16.3 ppm

HRMS (ESI):  $m/z$  calculated 555.4673 for  $C_{39}H_{59}N_2$  [ $M$ ] $^+$  found 555.4679  $\Delta$ -1.1 ppm

UHPLC (9 min):  $R_t$  = 8.39 min

1,3-didodecyl-5,6-dimethyl-1*H*-benzo[*d*]imidazolium (**56**)

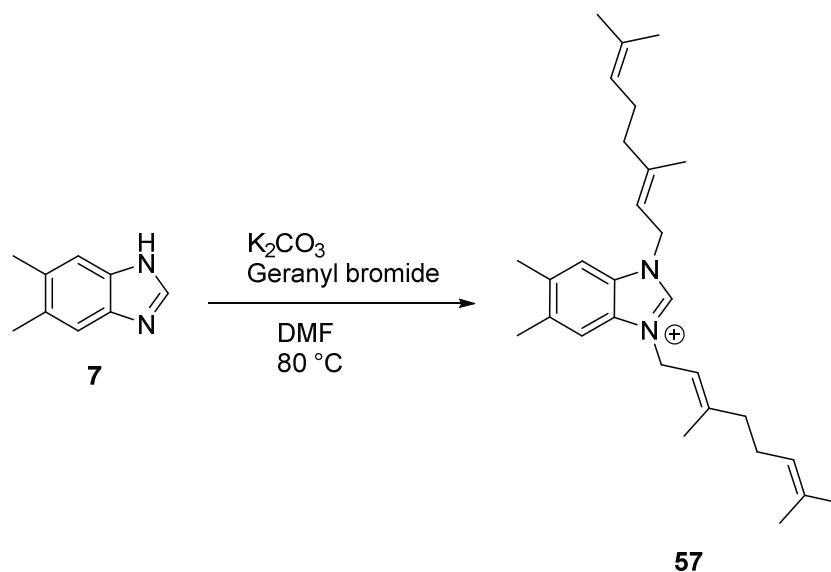
According to **GP D**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (20.0 mg, 0.1 mmol) was reacted with 1-bromododecane (168.7 mg, 0.7 mmol) utilizing NaHCO<sub>3</sub> (23.0 mg, 0.3 mmol) in 2.74 mL ACN. The reaction resulted in 1,3-didodecyl-5,6-dimethyl-1*H*-benzo[*d*]imidazolium (**56**) as a pale yellow oil. Preparative HPLC yielded 58% of **56** (38.2 mg, 0.08 mmol).

**<sup>1</sup>H NMR** (500 MHz, CHLOROFORM-*d*, 25°C)  $\delta$  = 8.50 (s, 1H), 7.15 (s, 1H), 4.37 (br t, *J* = 7.32 Hz, 4H), 2.33 (s, 6H), 1.04-1.31 (m, 40H), 0.74 ppm (t, *J* = 6.94 Hz, 6H)

**<sup>13</sup>C NMR** (126 MHz, CHLOROFORM-*d*, 25°C)  $\delta$  = 142.5, 136.9, 129.9, 112.6, 47.3, 31.8, 29.5, 29.4, 29.3, 29.3, 29.3, 28.9, 26.4, 22.6, 20.6, 14.0 ppm

**HRMS** (ESI): *m/z* calculated 483.4673 for C<sub>33</sub>H<sub>59</sub>N<sub>2</sub> [M]<sup>+</sup> found 483.4678  $\Delta$ -1.1 ppm

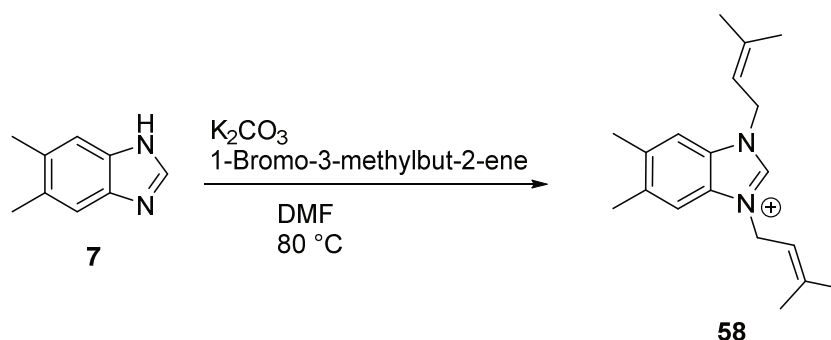
**UHPLC** (9 min): *Rt* = 8.61 min

1,3-bis((*E*)-3,7-dimethylocta-2,6-dien-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazolium (**57**)

According to **GP C**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (10.5 mg, 0.07 mmol) was reacted with geranyl bromide (62.1 mg, 0.29 mmol) utilizing  $K_2CO_3$  (9.8 mg, 0.07 mmol) in 2 mL DMF. The reaction resulted in 1,3-bis((*E*)-3,7-dimethylocta-2,6-dien-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazolium (**57**) as a yellow oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 419.3420 for  $C_{29}H_{43}N_2$  [ $M$ ] $^+$  found 419.3417  $\Delta$ 0.8 ppm

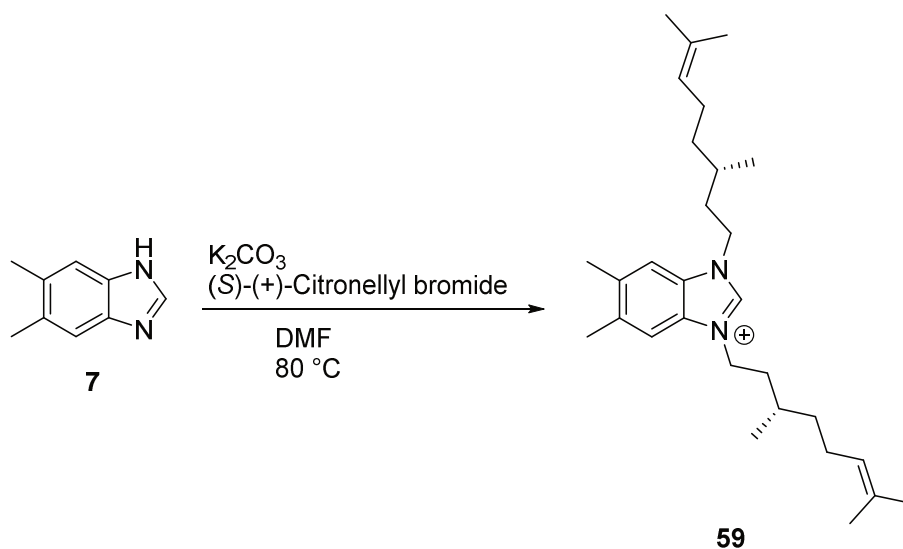
**UHPLC** (9 min):  $R_t$  = 6.92 min

5,6-dimethyl-1,3-bis(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazolium (**58**)

According to **GPC**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (15.5 mg, 0.1 mmol) was reacted with 1-bromo-3-methylbut-2-ene (63.1 mg, 0.4 mmol) utilizing  $K_2CO_3$  (14.6 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 5,6-dimethyl-1,3-bis(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazolium (**58**) as a yellow oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 283.2169 for  $C_{19}H_{27}N_2$  [ $M$ ]<sup>+</sup> found 283.2168  $\Delta$ 0.2 ppm

**UHPLC** (9 min):  $R_t$  = 4.57 min

1,3-bis((*S*)-3,7-dimethyloct-6-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazolium (**59**)

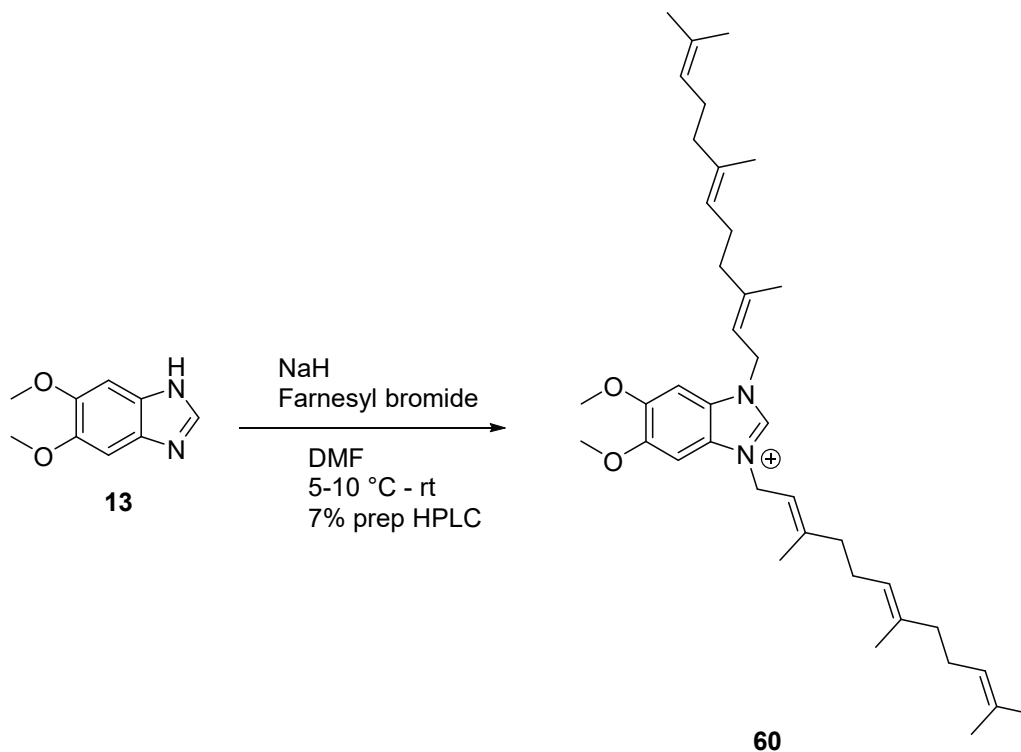
According to **GP C**, 5,6-dimethyl-1*H*-benzo[*d*]imidazole (**7**) (10.4 mg, 0.07 mmol) was reacted with (*S*)-(+)-citronellyl bromide (62.1 mg, 0.28 mmol) utilizing  $K_2CO_3$  (9.9 mg, 0.07 mmol) in 2 mL DMF. The reaction resulted in 1,3-bis((*S*)-3,7-dimethyloct-6-en-1-yl)-5,6-dimethyl-1*H*-benzo[*d*]imidazolium (**59**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 423.3734 for  $C_{29}H_{47}N_2$  [ $M$ ]<sup>+</sup> found 423.3736  $\Delta$ -0.5 ppm

**UHPLC** (9 min):  $R_t$  = 7.12 min



5,6-dimethoxy-1,3-bis((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**60**)



According to **GP B**, 5,6-dimethoxy-1*H*-benzo[*d*]imidazole (**13**) (19.4 mg, 0.1 mmol) was reacted with farnesyl bromide (28.4 mg, 0.1 mmol) utilizing NaH (2.5 mg, 0.1 mmol) in 2 mL DMF. The reaction resulted in 1-((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-5,6-dimethoxy-1*H*-benzo[*d*]imidazolium (**60**) (5.0 mg, 8.5  $\mu$ mol) with 7% yield after preparative HPLC as a yellow orange oil.

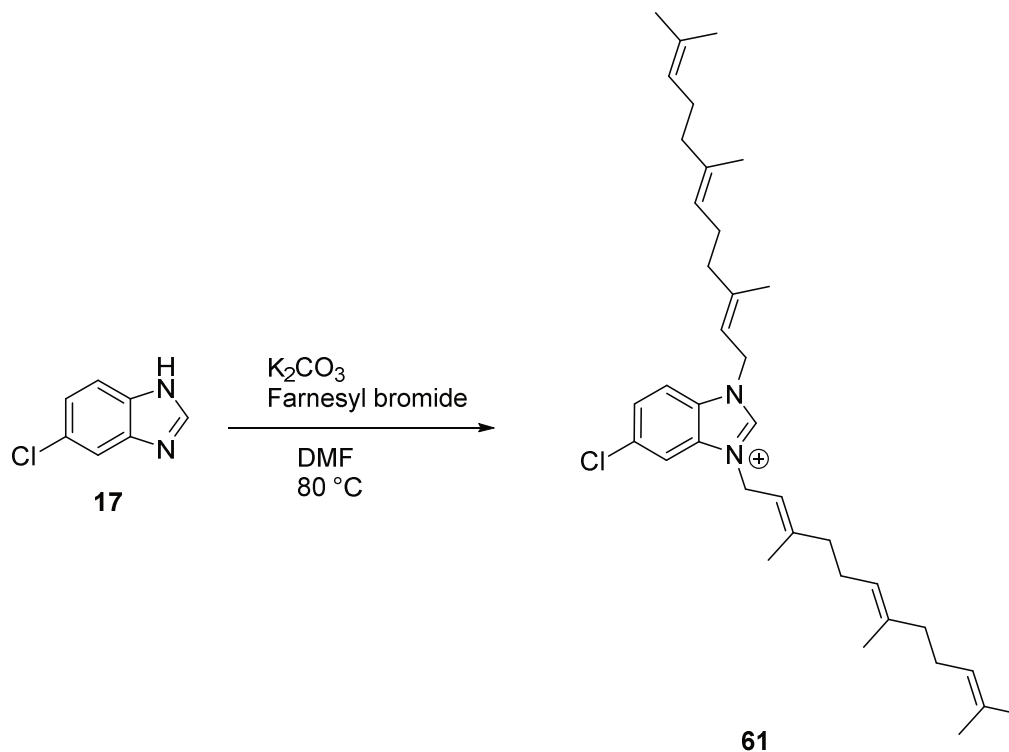
**<sup>1</sup>H NMR** (500 MHz, METHANOL-*d*<sub>4</sub>, 25°C)  $\delta$  = 7.32 (s, 2H), 5.51 (br t, *J*=6.87 Hz, 2H), 5.07 (br d, *J*=7.17 Hz, 4H), 4.99 - 5.06 (m, 4H), 3.96 (s, 6H), 2.16 - 2.21 (m, 8H), 1.89 - 1.99 (m, 14H), 1.64 (s, 6H), 1.54 - 1.58 ppm (m, 12H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C)  $\delta$  = 152.1, 146.3, 136.9, 132.4, 127.1, 125.4, 124.7, 117.5, 96.2, 57.2, 46.2, 40.9, 40.6, 27.9, 27.1, 26.1, 17.9, 16.9, 16.3 ppm

**HRMS** (ESI): *m/z* calculated 587.4571 for C<sub>39</sub>H<sub>59</sub>N<sub>2</sub>O<sub>2</sub> [M]<sup>+</sup> found 587.4574  $\Delta$ -0.4 ppm

**UHPLC** (9 min): *Rt* = 8.25 min

5-chloro-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium  
(**61**)

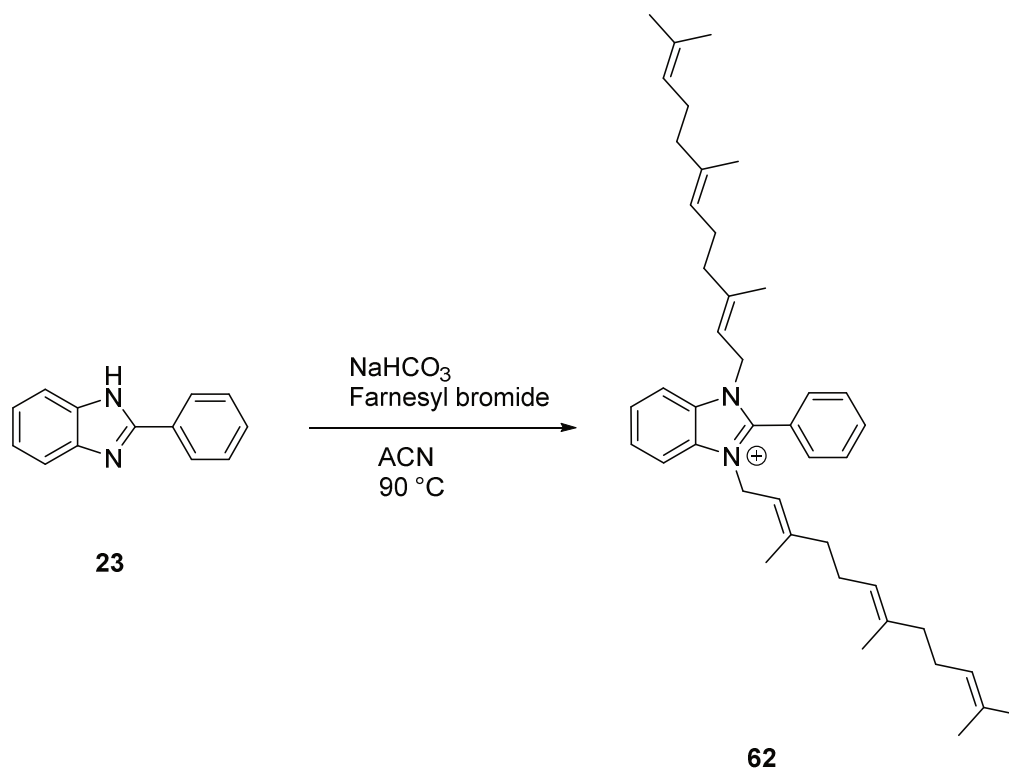


According to **GP C**, 5-chloro-1*H*-benzo[*d*]imidazole (**17**) (8.0 mg, 0.05 mmol) was reacted with farnesyl bromide (60.5 mg, 0.2 mmol) utilizing  $\text{K}_2\text{CO}_3$  (7.32 mg, 0.05 mmol) in 2 mL DMF. The reaction resulted in 5-chloro-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**61**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 561.3970 for  $\text{C}_{37}\text{H}_{54}\text{ClN}_2$  [ $\text{M}$ ]<sup>+</sup> found 561.3969  $\Delta 0.2$  ppm

**UHPLC** (9 min):  $R_t = 8.92$  min

2-phenyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (62)

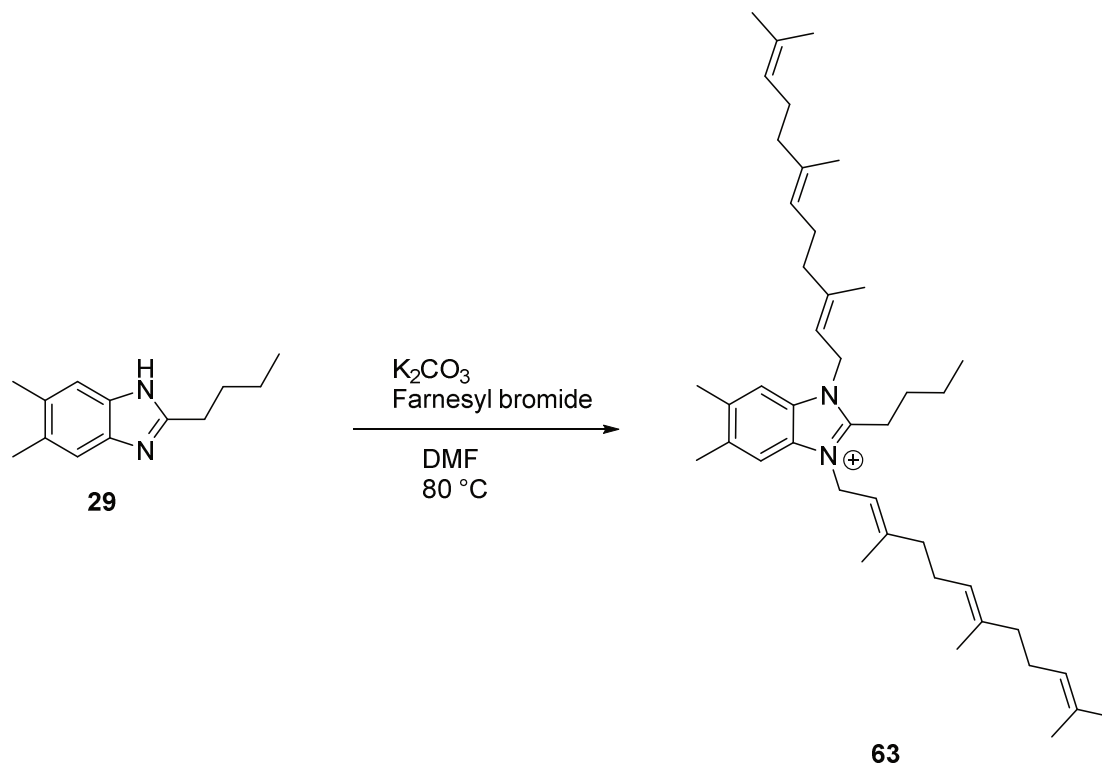


According to **GP D**, 2-phenyl-1*H*-benzo[*d*]imidazole (**23**) (20.0 mg, 0.1 mmol) was reacted with farnesyl bromide (147.6 mg, 0.5 mmol) utilizing NaHCO<sub>3</sub> (17.4 mg, 0.2 mmol) in 2.06 mL ACN. The reaction resulted in 2-phenyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**62**) as yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 603.4673 for C<sub>43</sub>H<sub>59</sub>N<sub>2</sub> [M]<sup>+</sup> found 603.4671 Δ0.4 ppm

**UHPLC** (9 min): Rt = 8.43 min

2-butyl-5,6-dimethyl-1,3-bis((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**63**)

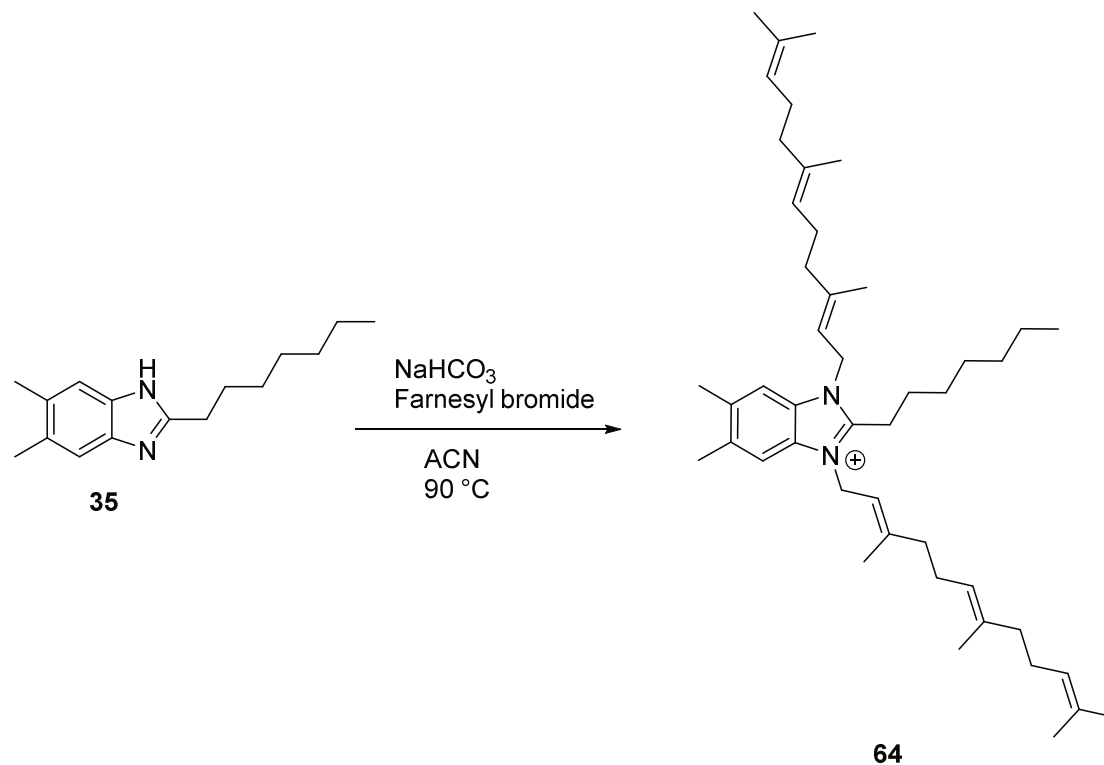


According to **GP C**, 2-butyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**29**) (5.0 mg, 0.03 mmol) was reacted with farnesyl bromide (28.0 mg, 0.1 mmol) utilizing  $K_2CO_3$  (3.4 mg, 0.02 mmol) in 2 mL DMF. The reaction resulted in 2-butyl-5,6-dimethyl-1,3-bis((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**63**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 611.5299 for  $C_{43}H_{67}N_2$  [ $M$ ]<sup>+</sup> found 611.5302  $\Delta$ -0.5 ppm

**UHPLC** (9 min):  $R_t$  = 8.73 min

2-heptyl-5,6-dimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**64**)

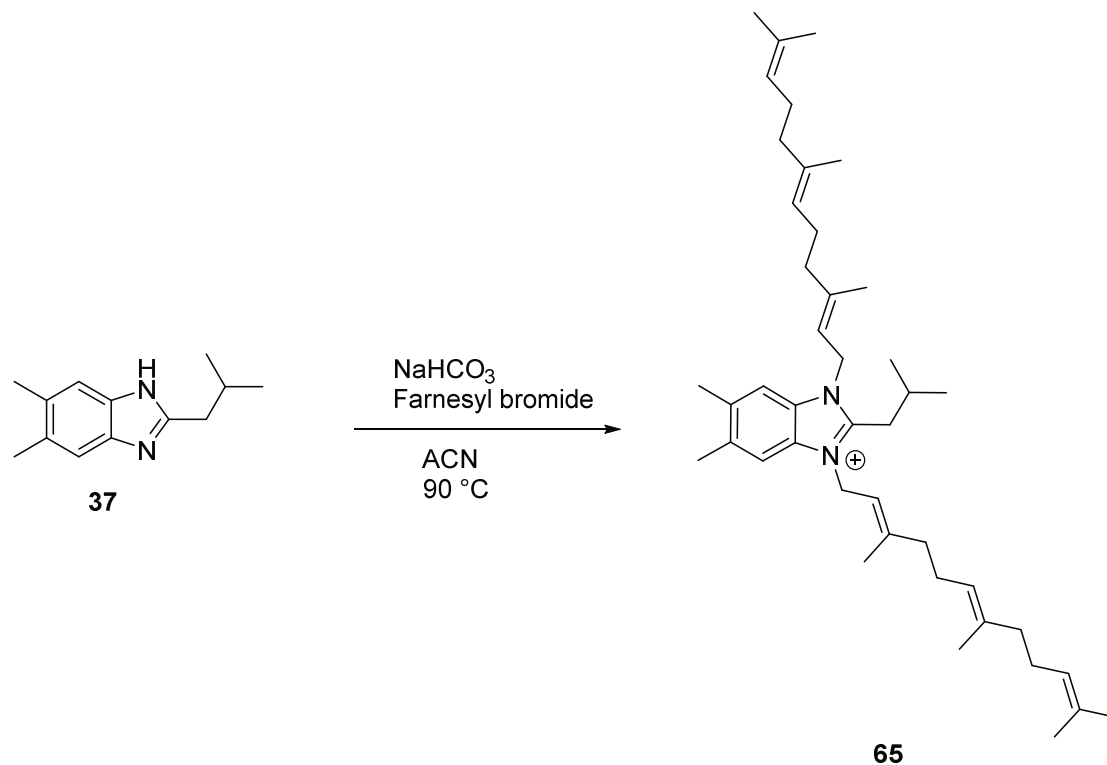


According to **GP D**, 2-heptyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**35**) (20.0 mg, 0.08 mmol) was reacted with farnesyl bromide (116.8 mg, 0.4 mmol) utilizing NaHCO<sub>3</sub> (13.8 mg, 0.16 mmol) in 1.64 mL ACN. The reaction resulted in 2-heptyl-5,6-dimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**64**) as yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 653.5768 for C<sub>46</sub>H<sub>73</sub>N<sub>2</sub> [M]<sup>+</sup> found 653.5777  $\Delta$ -1.3 ppm

**UHPLC** (9 min): Rt = 9.06 min

2-isobutyl-5,6-dimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**65**)

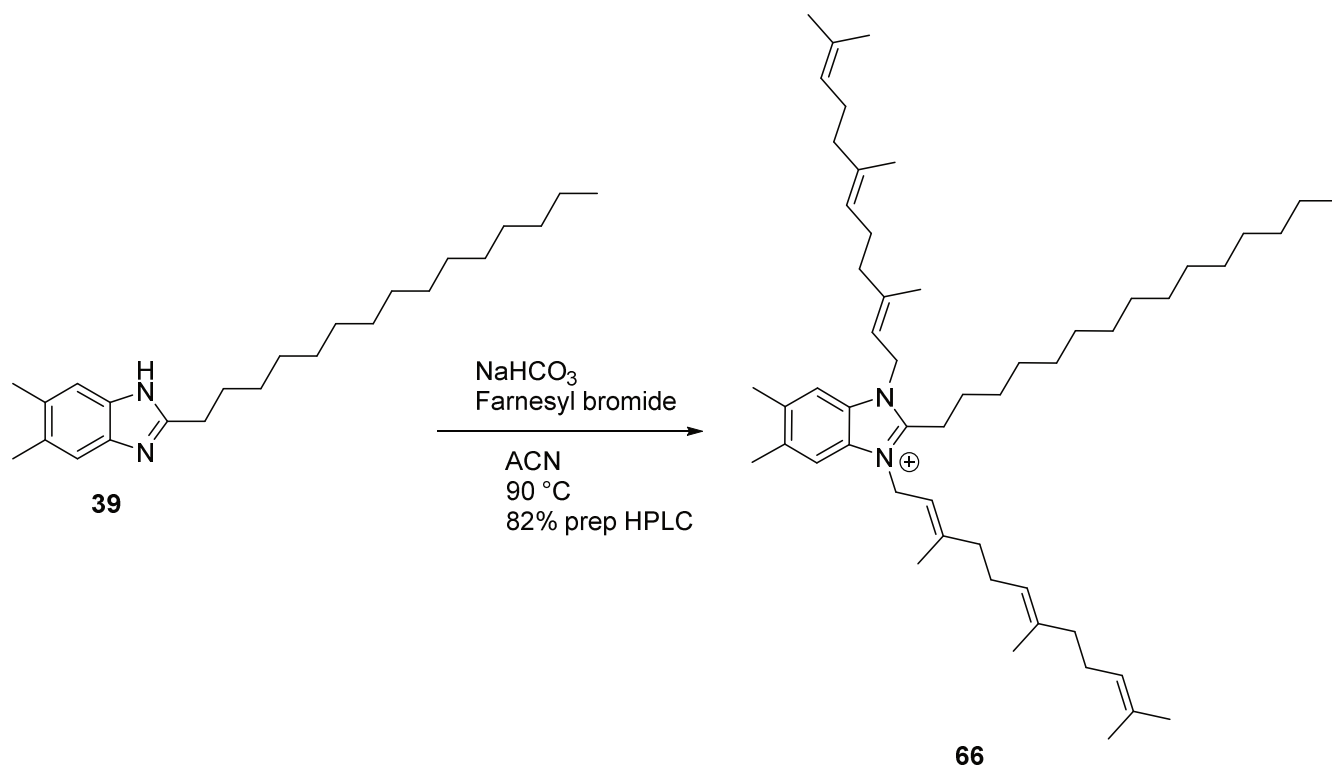


According to **GP D**, 2-isobutyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**37**) (20.0 mg, 0.1 mmol) was reacted with farnesyl bromide (141.2 mg, 0.5 mmol) utilizing NaHCO<sub>3</sub> (16.6 mg, 0.2 mmol) in 2.0 mL ACN. The reaction resulted in 2-isobutyl-5,6-dimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**65**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 611.5299 for C<sub>43</sub>H<sub>67</sub>N<sub>2</sub> [M]<sup>+</sup> found 611.5301 Δ-0.4 ppm

**UHPLC** (9 min): Rt = 8.76 min

5,6-dimethyl-2-pentadecyl-1,3-bis((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**66**)



According to **GP D**, 5,6-dimethyl-2-pentadecyl-1*H*-benzo[*d*]imidazole (**39**) (4.0 mg, 0.01 mmol) was reacted with farnesyl bromide (16.0 mg, 0.05 mmol) utilizing NaHCO<sub>3</sub> (1.9 mg, 0.02 mmol) in 1.0 mL ACN. The reaction resulted in 5,6-dimethyl-2-pentadecyl-1,3-bis((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**66**) (7.1 mg, 0.009 mmol) with 82% yield after preparative HPLC as a yellow oil.

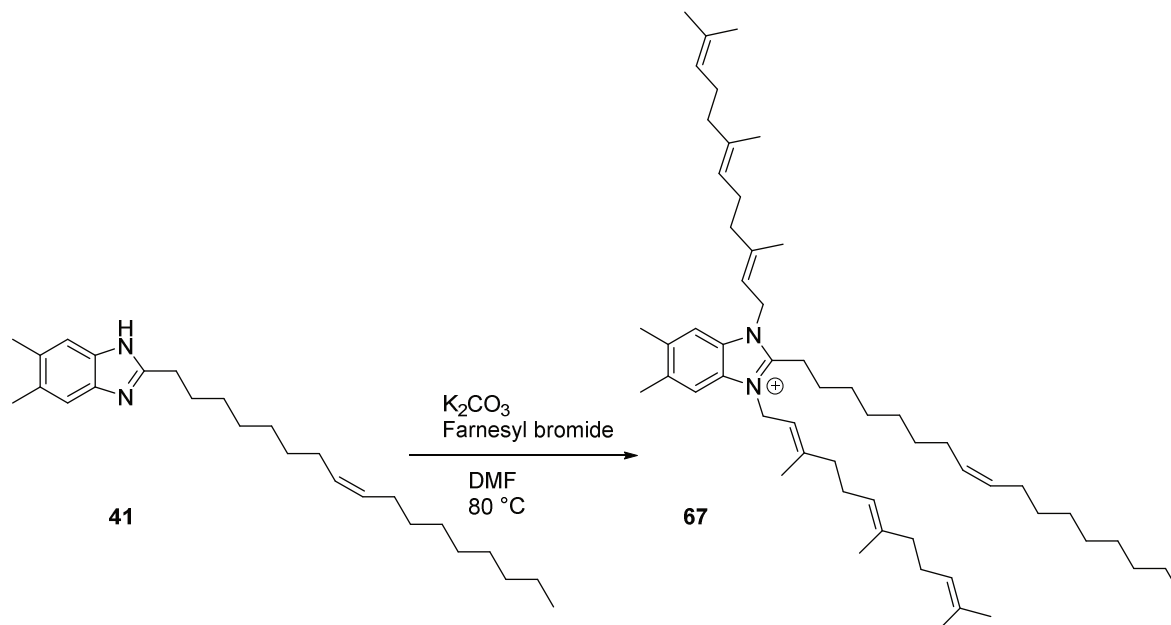
**<sup>1</sup>H NMR** (500 MHz, CHLOROFORM-*d*, 25°C)  $\delta$  = 7.20-7.25 (m, 2H), 5.12-5.17 (m, 2H), 5.02 (br s, 12H), 4.58-4.68 (m, 4H), 3.32 (br s, 2H), 2.35-2.40 (m, 6H), 2.04 (br s, 8H), 1.92-1.99 (m, 6H), 1.85-1.90 (m, 9H), 1.62 (br s, 6H), 1.53 (br s, 12H), 1.21 (br s, 26H), 0.81-0.86 ppm (m, 3H)

**<sup>13</sup>C NMR** (126 MHz, METHANOL-*d*<sub>4</sub>, 25°C)  $\delta$  = 152.4, 142.8, 136.5, 136.4, 135.9, 131.4, 129.6, 124.1, 123.0, 116.5, 112.6, 77.0, 44.5, 39.7, 39.4, 31.9, 29.7, 29.7, 29.6, 29.4, 29.4, 29.3, 27.9, 26.7, 26.1, 25.7, 24.7, 22.7, 20.6, 17.9, 17.7, 17.1, 16.0, 14.1 ppm

**HRMS** (ESI):  $m/z$  calculated 765.7020 for C<sub>54</sub>H<sub>89</sub>N<sub>2</sub> [M]<sup>+</sup> found 765.7022  $\Delta$ -0.2 ppm

**UHPLC** (9 min): Rt = 9.90 min

2-((*Z*)-heptadec-8-en-1-yl)-5,6-dimethyl-1,3-bis((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**67**)



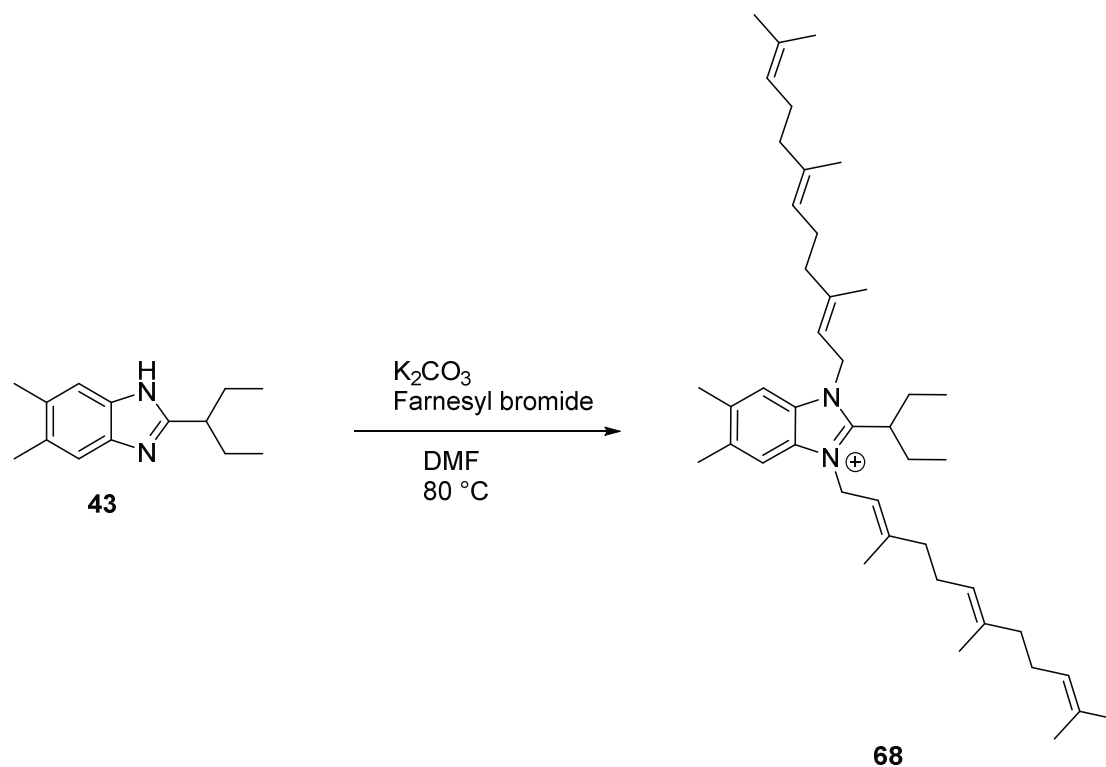
According to **GPC**, (*Z*)-2-(heptadec-8-en-1-yl)-5,6-dimethyl-1*H*-beno[*d*]imidazole (**41**) (7.2 mg, 0.02 mmol) was reacted with farnesyl bromide (21.6 mg, 0.08 mmol) utilizing  $K_2CO_3$  (2.6 mg, 0.02 mmol) in 2 mL DMF. The reaction resulted in 2-((*Z*)-heptadec-8-en-1-yl)-5,6-dimethyl-1,3-bis((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**67**) as an orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 791.7177 for  $C_{56}H_{91}N_2$  [ $M$ ]<sup>+</sup> found 791.7180  $\Delta$ -0.4 ppm

**UHPLC** (9 min):  $R_t$  = 9.90 min



5,6-dimethyl-2-(pentan-3-yl)-1,3-bis((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**68**)

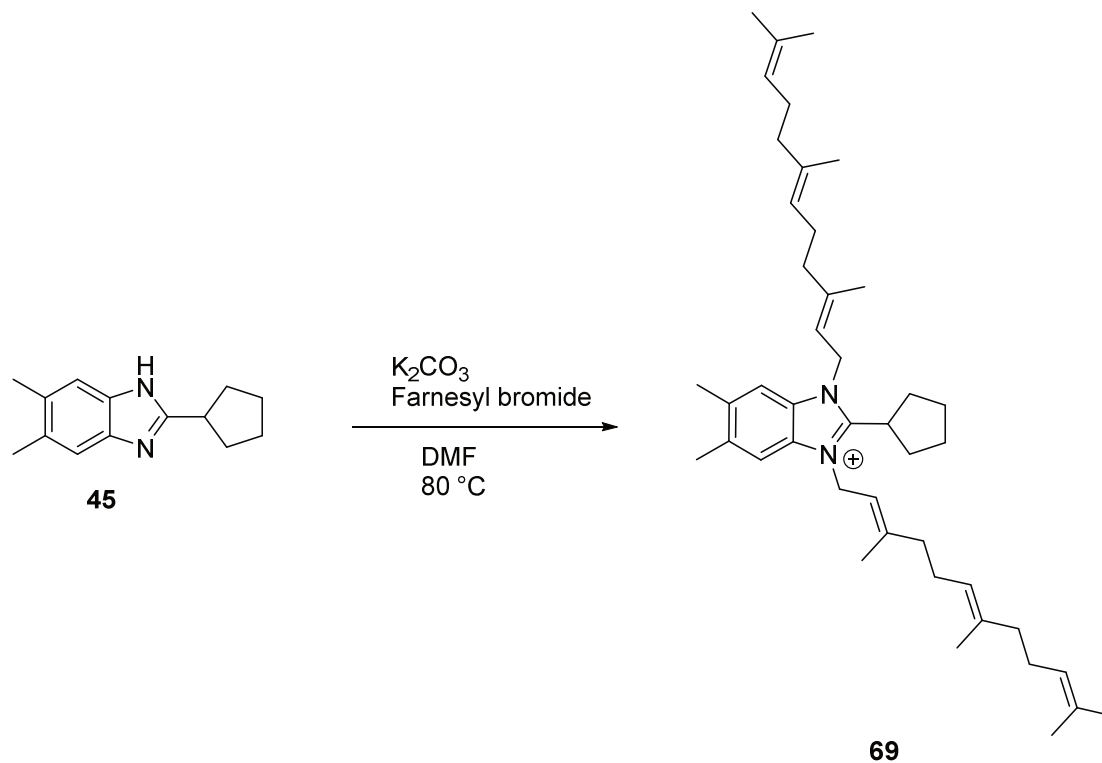


According to **GP C**, 5,6-dimethyl-2-(pentan-3-yl)-1*H*-benzo[*d*]imidazole (**43**) (5.2 mg, 0.02 mmol) was reacted with farnesyl bromide (27.3 mg, 0.1 mmol) utilizing  $K_2CO_3$  (3.3 mg, 0.02 mmol) in 2 mL DMF. The reaction resulted in 5,6-dimethyl-2-(pentan-3-yl)-1,3-bis((*2E,6E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**68**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 625.5455 for  $C_{44}H_{69}N_2$   $[M+H]^+$  found 625.5458  $\Delta$ -0.4 ppm

**UHPLC** (9 min):  $R_t$  = 8.78 min

2-cyclopentyl-5,6-dimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**69**)

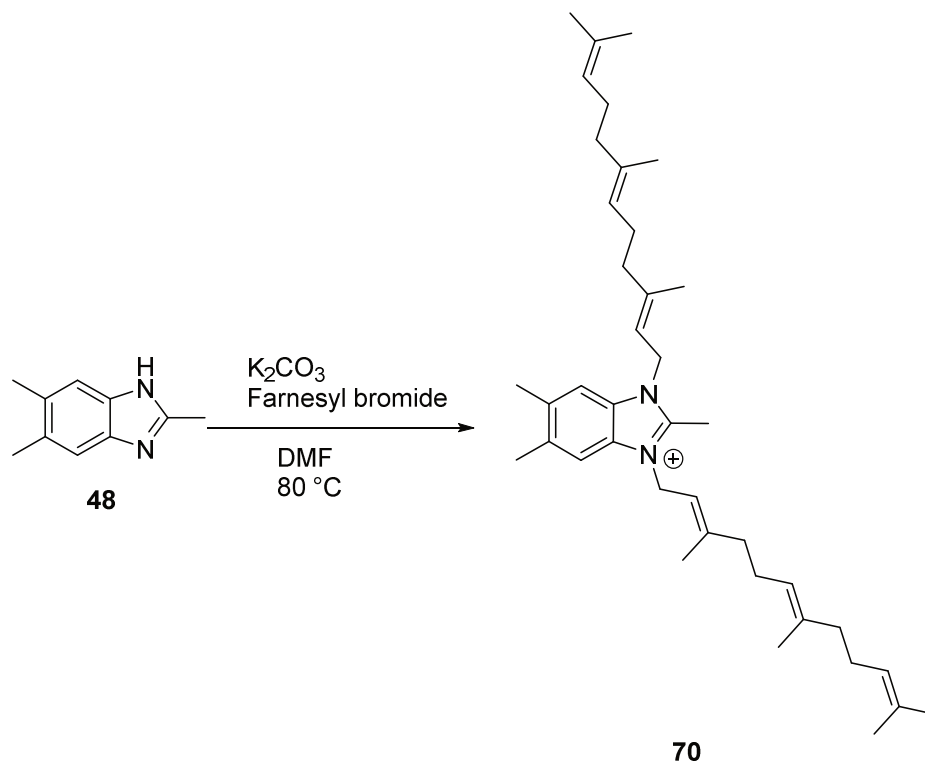


According to **GPC**, 2-cyclopentyl-5,6-dimethyl-1*H*-benzo[*d*]imidazole (**45**) (5.2 mg, 0.02 mmol) was reacted with farnesyl bromide (27.4 mg, 0.1 mmol) utilizing  $K_2CO_3$  (3.3 mg, 0.02 mmol) in 2 mL DMF. The reaction resulted in 2-cyclopentyl-5,6-dimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**69**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 623.5299 for  $C_{44}H_{67}N_2$   $[M]^+$  found 623.5304  $\Delta$ -0.8 ppm

**UHPLC** (9 min):  $R_t$  = 8.75 min

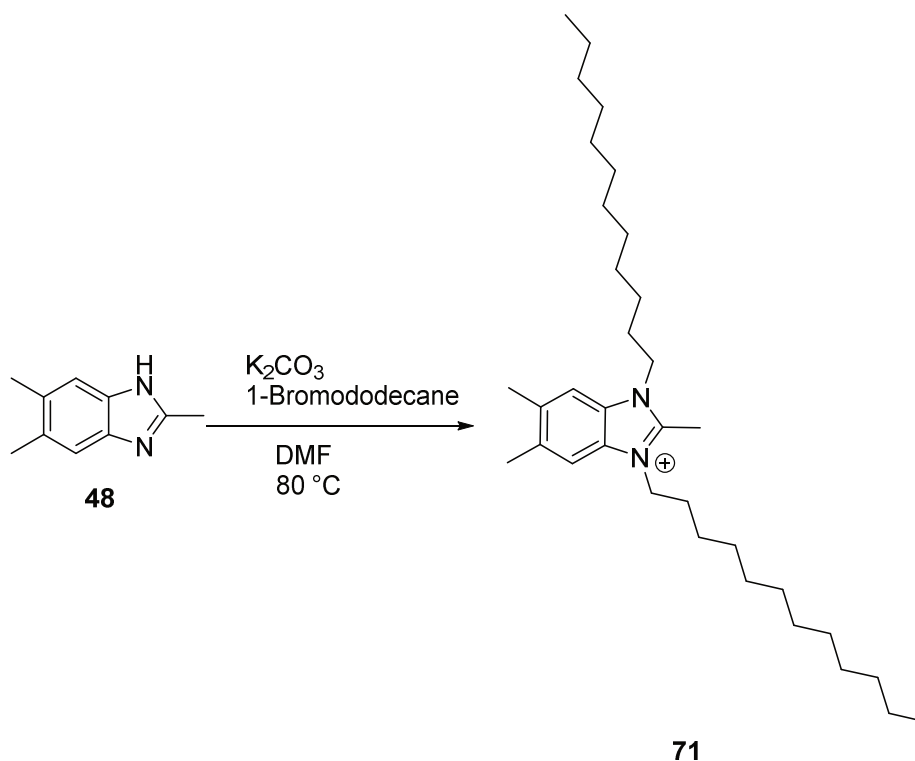
2,5,6-trimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**70**) = Sandacrabins C



According to **GP C**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (8.0 mg, 0.05 mmol) was reacted with farnesyl bromide (57.0 mg, 0.2 mmol) utilizing  $K_2CO_3$  (6.9 mg, 0.05 mmol) in 2 mL DMF. The reaction resulted in 2,5,6-trimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazolium (**70**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 569.4829 for  $C_{40}H_{61}N_2$  [ $M$ ] $^+$  found 569.4834  $\Delta$ -0.7 ppm

**UHPLC** (9 min):  $R_t$  = 8.77 min

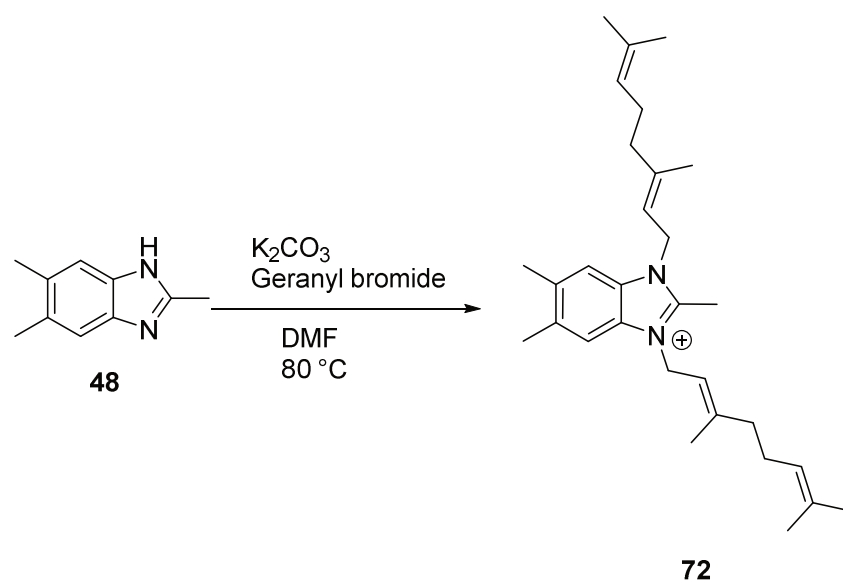
1,3-didodecyl-2,5,6-trimethyl-1*H*-benzo[*d*]imidazolium (**71**)

According to **GP C**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (8.0 mg, 0.05 mmol) was reacted with 1-bromododecane (49.8 mg, 0.2 mmol) utilizing  $K_2CO_3$  (6.9 mg, 0.05 mmol) in 2 mL DMF. The reaction resulted in 1,3-didodecyl-2,5,6-trimethyl-1*H*-benzo[*d*]imidazolium (**71**) as a pale yellow oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 497.4829 for  $C_{34}H_{61}N_2$  [ $M$ ]<sup>+</sup> found 497.4831  $\Delta$ -0.3 ppm

**UHPLC** (9 min):  $R_t$  = 9.03 min

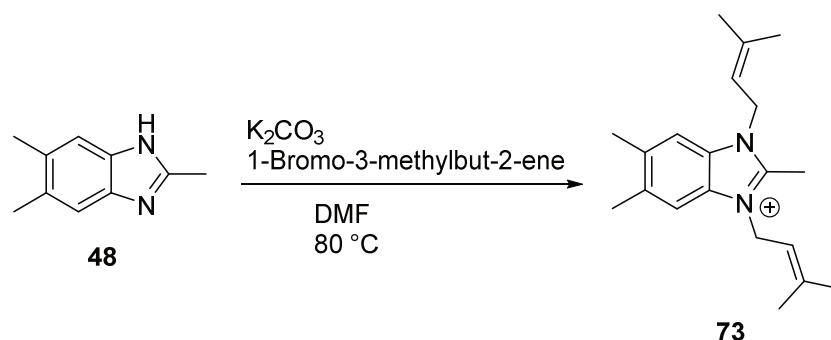
1,3-bis((*E*)-3,7-dimethylocta-2,6-dien-1-yl)-2,5,6-trimethyl-1*H*-benzo[*d*]imidazolium (**72**)



According to **GP C**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (8.0 mg, 0.05 mmol) was reacted with geranyl bromide (43.4 mg, 0.2 mmol) utilizing  $K_2CO_3$  (6.9 mg, 0.05 mmol) in 2 mL DMF. The reaction resulted in 1,3-bis((*E*)-3,7-dimethylocta-2,6-dien-1-yl)-2,5,6-trimethyl-1*H*-benzo[*d*]imidazolium (**72**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 433.3577 for  $C_{30}H_{45}N_2$  [ $M$ ]<sup>+</sup> found 433.3579  $\Delta$ -0.4 ppm

**UHPLC** (9 min):  $R_t$  = 7.38 min

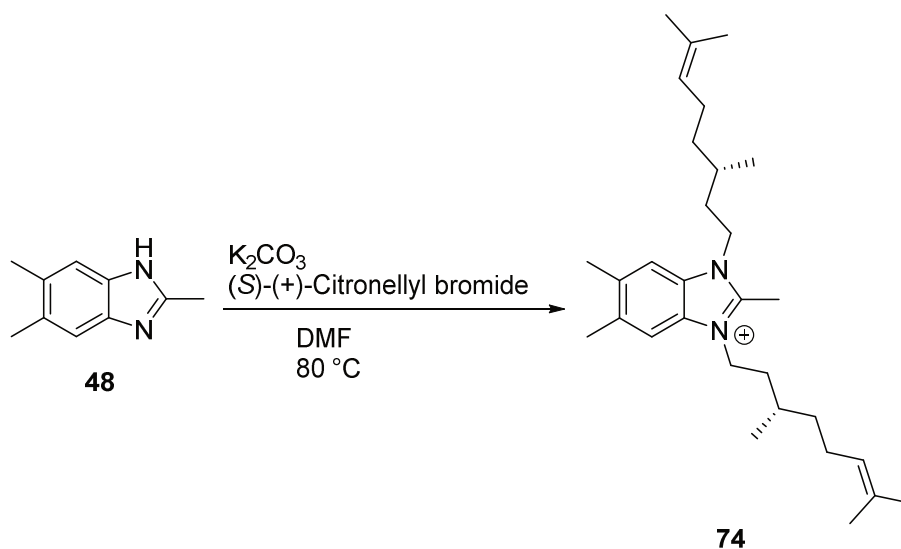
2,5,6-trimethyl-1,3-bis(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazolium (**73**)

According to **GP C**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (8.0 mg, 0.05 mmol) was reacted with 1-bromo-3-methylbut-2-ene (29.8 mg, 0.2 mmol) utilizing  $K_2CO_3$  (6.9 mg, 0.05 mmol) in 2 mL DMF. The reaction resulted in 2,5,6-trimethyl-1,3-bis(3-methylbut-2-en-1-yl)-1*H*-benzo[*d*]imidazolium (**73**) as a yellow orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 297.2325 for  $C_{20}H_{29}N_2$  [ $M$ ]<sup>+</sup> found 297.2330  $\Delta$ -1.5 ppm

**UHPLC** (9 min):  $R_t$  = 5.14 min

1,3-bis((*S*)-3,7-dimethyloct-6-en-1-yl)-2,5,6-trimethyl-1*H*-benzo[*d*]imidazolium (**74**)



According to **GP C**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**48**) (8.0 mg, 0.05 mmol) was reacted with (*S*)-(+)-citronellyl bromide (43.8 mg, 0.2 mmol) utilizing  $K_2CO_3$  (6.9 mg, 0.05 mmol) in 2 mL DMF. The reaction resulted in 1,3-bis((*S*)-3,7-dimethyloct-6-en-1-yl)-2,5,6-trimethyl-1*H*-benzo[*d*]imidazolium (**74**) as an orange oil. Since the purification by semi-preparative HPLC is in progress, no yield or NMR data are reported so far.

**HRMS** (ESI):  $m/z$  calculated 437.3890 for  $C_{30}H_{49}N_2$  [ $M$ ]<sup>+</sup> found 437.3894  $\Delta$ -0.8 ppm

**UHPLC** (9 min):  $R_t$  = 7.59 min

## 5.6 References

1. Bader, C.D.; Panter, F.; Garcia, R.; Tchesnokov, E.P.; Haid, S.; Walt, C.; Spröer, C.; Kiefer, A.F.; Götte, M.; Overmann, J.; et al. Sandacrabins - Structurally Unique Antiviral RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry – A European Journal* **2022**, *28*, e202104484, doi:10.1002/chem.202104484.
2. Dias, D.A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2*, 303–336, doi:10.3390/metabo2020303.
3. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803, doi:10.1021/acs.jnatprod.9b01285.
4. Miethke, M.; Pieroni, M.; Weber, T.; Brönstrup, M.; Hammann, P.; Halby, L.; Arimondo, P.B.; Glaser, P.; Aigle, B.; Bode, H.B.; et al. Towards the sustainable discovery and development of new antibiotics. *Nat. Rev. Chem.* **2021**, doi:10.1038/s41570-021-00313-1.
5. Thomford, N.E.; Senthebane, D.A.; Rowe, A.; Munro, D.; Seele, P.; Maroyi, A.; Dzobo, K. Natural Products for Drug Discovery in the 21<sup>st</sup> Century: Innovations for Novel Drug Discovery. *Int. J. Mol. Sci.* **2018**, *19*, doi:10.3390/ijms19061578.
6. Koonin, E.V.; Senkevich, T.G.; Dolja, V.V. The ancient Virus World and evolution of cells. *Biol Direct* **2006**, *1*, 1–27, doi:10.1186/1745-6150-1-29.
7. Bernheim, A.; Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol* **2020**, *18*, 113–119, doi:10.1038/s41579-019-0278-2.
8. El Sayed, K.A. Natural Products as Antiviral Agents. In *Studies in natural products chemistry*; Rahman, A.-u., Ed.; Elsevier: Amsterdam, London, 1988-; pp 473–572, ISBN 9780444506436.
9. Herrmann, J.; Fayad, A.A.; Muller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2016**, doi:10.1039/c6np00106h.
10. Weissman, K.J.; Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **2010**, *27*, 1276–1295, doi:10.1039/c001260m.
11. Mulwa, L.S.; Stadler, M. Antiviral Compounds from Myxobacteria. *Microorganisms* **2018**, *6*, doi:10.3390/microorganisms6030073.
12. Koutsoudakis, G.; Romero-Brey, I.; Berger, C.; Pérez-Vilaró, G.; Perin, P.M.; Vondran, Florian Wolfgang Rudolf; Kalesse, M.; Harmrolfs, K.; Müller, R.; Martinez, J.P.; et al. Soraphen A: a broad-spectrum antiviral natural product with potent anti-hepatitis C virus activity. *J. Hepatol.* **2015**, doi:10.1016/j.jhep.2015.06.002.



13. Singaravelu, R.; Desrochers, G.F.; Srinivasan, P.; O'Hara, S.; Lyn, R.K.; Müller, R.; Jones, D.M.; Russell, R.S.; Pezacki, J.P. Soraphen A: A Probe for investigating the role of de novo lipogenesis during viral infection. *ACS Infect. Dis.* **2015**, *1*, 130–134, doi:10.1021/acsinfecdis.5b00019.
14. SeeSAR; BioSolveIT GmbH, Sankt Augustin, Germany.
15. Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **2012**, *28*, 2074–2075, doi:10.1093/bioinformatics/bts310.
16. Yin, W.; Mao, C.; Luan, X.; Shen, D.-D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **2020**, *368*, 1499–1504, doi:10.1126/science.abc1560.
17. Waring, M.J.; Arrowsmith, J.; Leach, A.R.; Leeson, P.D.; Mandrell, S.; Owen, R.M.; Pairaudeau, G.; Pennie, W.D.; Pickett, S.D.; Wang, J.; et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov* **2015**, *14*, 475–486, doi:10.1038/nrd4609.
18. Kathleen M. Knights; David M. Stresser; John O. Miners; Charles L. Crespi. In Vitro Drug Metabolism Using Liver Microsomes. *Current Protocols in Pharmacology* **2016**, *74*, 7.8.1–7.8.24, doi:10.1002/cpph.9.
19. Chen, X.-Q.; Ziembra, T.; Huang, C.; Chang, M.; Xu, C.; Qiao, J.X.; Wang, T.C.; Finlay, H.J.; Salvati, M.E.; Adam, L.P.; et al. Oral Delivery of Highly Lipophilic, Poorly Water-Soluble Drugs: Self-Emulsifying Drug Delivery Systems to Improve Oral Absorption and Enable High-Dose Toxicology Studies of a Cholesteryl Ester Transfer Protein Inhibitor in Preclinical Species. *Journal of Pharmaceutical Sciences* **2018**, *107*, 1352–1360, doi:10.1016/j.xphs.2018.01.003.
20. Veber, D.F.; Johnson, S.R.; Cheng, H.-Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623, doi:10.1021/jm020017n.
21. Kola, I.; Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **2004**, *3*, 711–715, doi:10.1038/nrd1470.
22. Segall, M.D.; Barber, C. Addressing toxicity risk when designing and selecting compounds in early drug discovery. *Drug Discovery Today* **2014**, *19*, 688–693, doi:10.1016/j.drudis.2014.01.006.
23. Jennings, M.C.; Minbiole, K.P.C.; Wuest, W.M. Quaternary Ammonium Compounds: An Antimicrobial Mainstay and Platform for Innovation to Address Bacterial Resistance. *ACS Infect. Dis.* **2015**, *1*, 288–303, doi:10.1021/acsinfecdis.5b00047.
24. Fischl, W.; Bartenschlager, R. High-throughput screening using dengue virus reporter genomes. *Methods Mol. Biol.* **2013**, *1030*, doi:10.1007/978-1-62703-484-5\_17.

25. Münster, M.; Płaszczycza, A.; Cortese, M.; Neufeldt, C.J.; Goellner, S.; Long, G.; Bartenschlager, R. A Reverse Genetics System for Zika Virus Based on a Simple Molecular Cloning Strategy. *Viruses* **2018**, *10*, doi:10.3390/v10070368.
26. Gordon, C.J.; Lee, H.W.; Tchesnokov, E.P.; Perry, J.K.; Feng, J.Y.; Bilello, J.P.; Porter, D.P.; Götte, M. Efficient incorporation and template-dependent polymerase inhibition are major determinants for the broad-spectrum antiviral activity of remdesivir. *Journal of Biological Chemistry* **2022**, *298*, 101529, doi:10.1016/j.jbc.2021.101529.
27. Gonell, S.; Peris, E.; Poyatos, M. Structural Features of Mono- and Dimetallic Complexes of Palladium Combining Two Types of Aromatic NHC Ligands. *Eur. J. Inorg. Chem.* **2019**, *2019*, 3776–3781, doi:10.1002/ejic.201900684.

## Supporting Information

# SAR Studies and Biological Characterization of Sandacrabin Derivatives as Antiviral RNA Polymerase Inhibitors

## S 5.1 Antiviral activities and estimated binding affinities

**Table S1.** Summary of sandacrab-in-dependent inhibition of viral RNA-dependent RNA polymerase (RdRp) complexes.  $IC_{50}$  values for the SARS-CoV-2 RdRp complex have previously been reported in [1].

Compound	Lassa Virus (LASV) RdRp complex Inhibition at 100 $\mu$ M	SARS-CoV-2 RdRp complex Inhibition at 100 $\mu$ M	SARS-CoV2 RdRp complex $IC_{50}$ [ $\mu$ M]
Sandacrab-in A	45%	30%	108
Sandacrab-in B	100%	99%	3.5
Sandacrab-in C	99%	97%	6.8
Sandacrab-in D	73%	47%	85.0
Sandacrab-in E	56%	32%	43.0
Remdesivir*	<i>n.d.</i>	<i>n.d.</i>	0.014

*n.d.*: not determined, \*: positive control

**Table S2.** Summary of antiviral activities of the sandacrab-ins.  $IC_{50}$  values for HCoV229E have previously been reported in [1].

Compound	Human pathogenic corona virus HCoV229E $IC_{50}$ [ $\mu$ M]	Dengue virus DENV-R2A $IC_{50}$ [ $\mu$ M]	Zika virus ZIKV-R2A $IC_{50}$ [ $\mu$ M]
	Sandacrab-in A	> 10	> 10
Sandacrab-in B	0.18	0.16**	0.15**
Sandacrab-in C	0.34	0.50**	0.32**
Sandacrab-in D	1.64	0.14	0.18
Sandacrab-in E	1.91	0.80	0.92
Remdesivir*	0.0056	<i>n.d.</i>	<i>n.d.</i>
Ribavirin*	<i>n.d.</i>	2.33	2.51

*n.d.*: not determined, \*: positive control, \*\*: the bioactivities of sandacrab-ins B and C against Dengue and Zika virus are potentially aberrated because of cytotoxicity against the Huh7 / Lunet T7 host cells

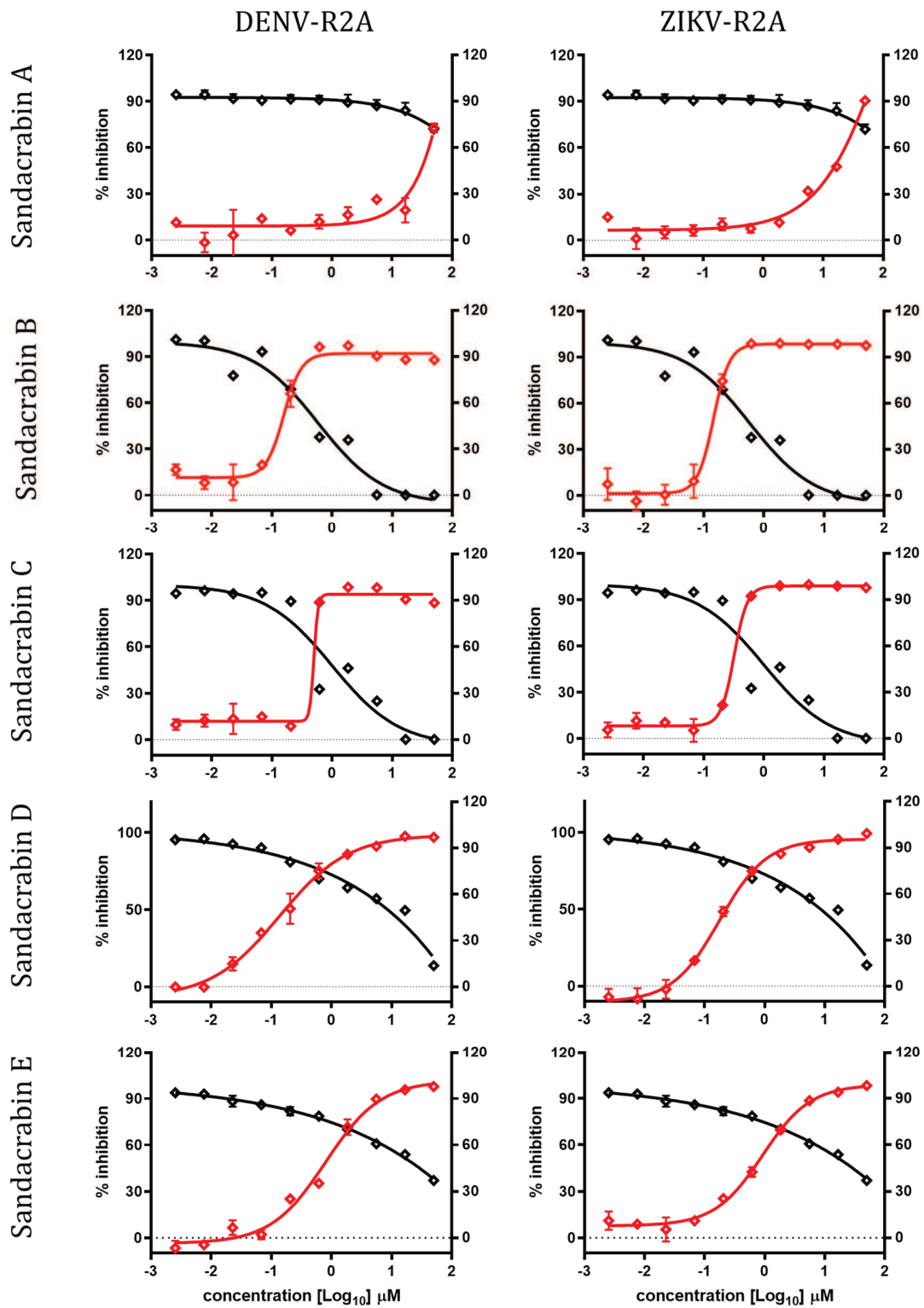
**Table S3.** Summary of cytotoxicity of the sandacrabins.  $CC_{50}$  values for Huh7.5 FLuc and U-2 OS have previously been reported in [1].

Compound	Huh7.5 FLuc $CC_{50}$ [ $\mu$ M]	U-2 OS $CC_{50}$ [ $\mu$ M]	Huh7 / Lunet T7 $CC_{50}$ [ $\mu$ M]
Sandacrabin A	> 37	> 37	> 50
Sandacrabin B	0.70	0.51	0.59
Sandacrabin C	3.61	1.22	1.00
Sandacrabin D	4.63	16.90	> 10
Sandacrabin E	6.72	13.70	> 10
Doxorubicin	0.037	0.015	<i>n.d.</i>
Ribavirin	<i>n.d.</i>	<i>n.d.</i>	ca. 200

*n.d.*: not determined

**Table S4.** Calculated binding affinities of the sandacrabins to SARS-CoV2 RdRp complex using SeeSAR. Minimum and maximum values derive from different orientations in the respective binding site.

Compound	Estimated binding affinity active site	Estimated binding affinity allosteric site
Sandacrabin A	high $\mu$ M to mM	$\mu$ M to mM
Sandacrabin B	high $\mu$ M to mM	low $\mu$ M to mM
Sandacrabin C	mM	high nM to mM
Sandacrabin D	$\mu$ M to mM	low $\mu$ M to mM
Sandacrabin E	$\mu$ M to mM	low $\mu$ M to mM



**Figure S1.** Antiviral activities of the sandacrabins against dengue virus DENV-R2A and zika virus ZIKV-H/PF/2013 (red trace, % inhibition) and cell viability of Huh7/Lunet T7 cells (black trace, % viability).

## S 5.2 Derek analysis

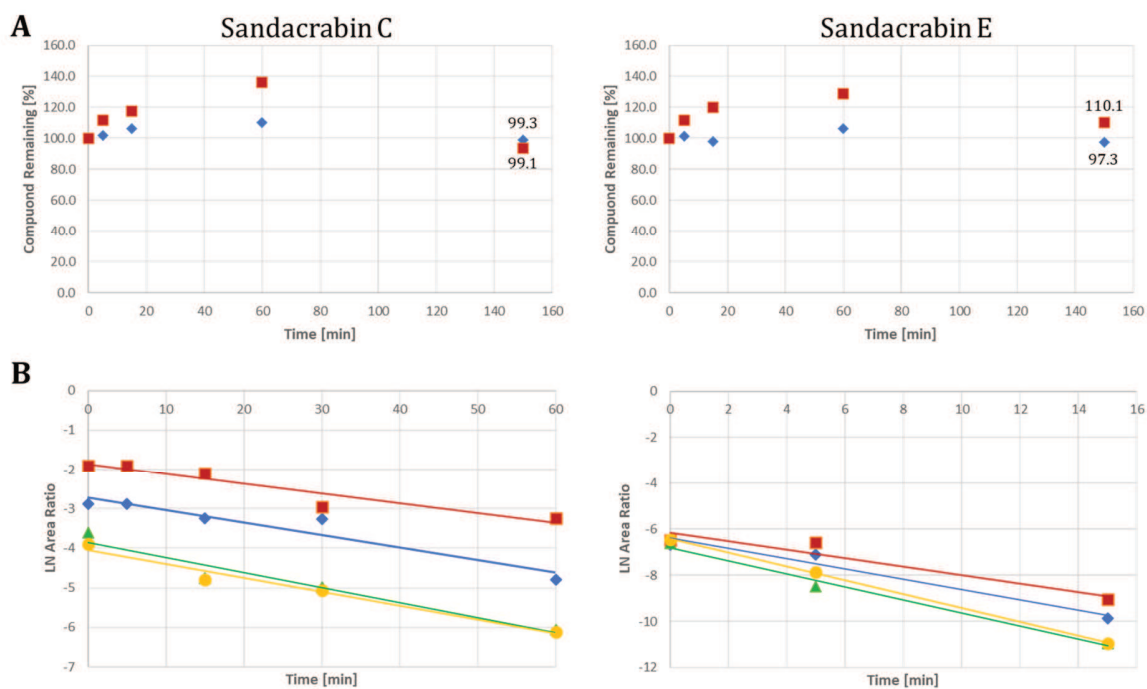
**Table S5.** Toxicity predictions of the sandacrabins from Derek analysis [2,3]. Toxicity predictions decrease in confidence in the following order: certain > probable > plausible > equivocal. Predictions against toxicity increase in confidence in the following order: inactive (unclassified/misclassified) < inactive < improbable. If no evidence of toxicity is found, then 'no report' is returned.

Name	Sandacrabin A	Sandacrabin B	Sandacrabin C	Sandacrabin D	Sandacrabin E
<b>Mutagenicity</b>	inactive	inactive	inactive	inactive	inactive
<b>Skin sensitization</b>	equivocal	equivocal	equivocal	equivocal	equivocal
<b>Carcinogenicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Photocarcinogenicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Chromosome damage</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Photo-induced chromosome damage</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Photomutagenicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Non-specific genotoxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Photo-induced non-specific genotoxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Irritation (of the eye)</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Irritation (of the gastrointestinal tract)</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Irritation (of the respiratory tract)</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Irritation (of the skin)</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Lachrymation</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>5alpha-Reductase inhibition</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Anaphylaxis</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Blood in urine</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Cerebral oedema</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Chloracne</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Cyanide-type effects</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>High acute toxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Methaemoglobinaemia</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Oestrogen receptor modulation</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Oestrogenicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>

<b>Peroxisome proliferation</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Phospholipidosis</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Phototoxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Mitochondrial dysfunction</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Uncoupler of oxidative phosphorylation</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Cholinesterase inhibition</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Neurotoxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Adrenal gland toxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Hepatotoxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Ocular toxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Pulmonary toxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Splenotoxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Thyroid toxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Urolithiasis</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Bladder disorders</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Bladder urothelial hyperplasia</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Bone marrow toxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Cumulative effect on white cell count and immunology</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Bradycardia</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Cardiotoxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>HERG channel inhibition</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Kidney disorders</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Kidney function-related toxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Nephrotoxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>alpha-2-mu-Globulin nephropathy</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Developmental toxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Teratogenicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Testicular toxicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Occupational asthma</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Respiratory sensitisation</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>
<b>Photoallergenicity</b>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>	<i>no report</i>



## S 5.3 ADMET profiling



**Figure S2.** (A) Plasma stability assay of sandacrabins C (left) and sandacrabins E (right) in mouse plasma shows no degradation over 150 minutes. (B) Metabolic stability assay of sandacrabins C (left) and sandacrabins E (right) in mouse liver S9 fraction. Different colors and markers refer to different individual replicates.

## S 5.4 NMR Spectra

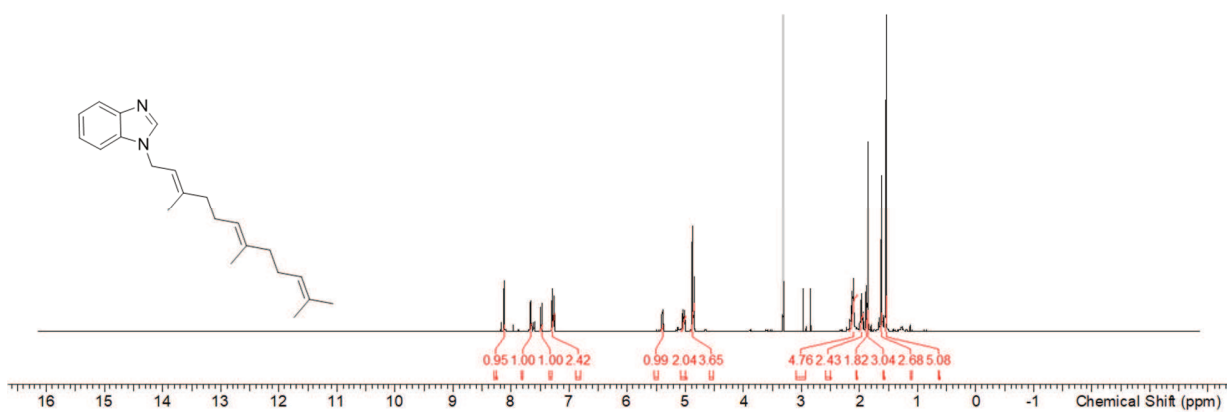


Figure S3.  $^1\text{H}$  spectrum of **2** in methanol- $d_4$  at 500 MHz.

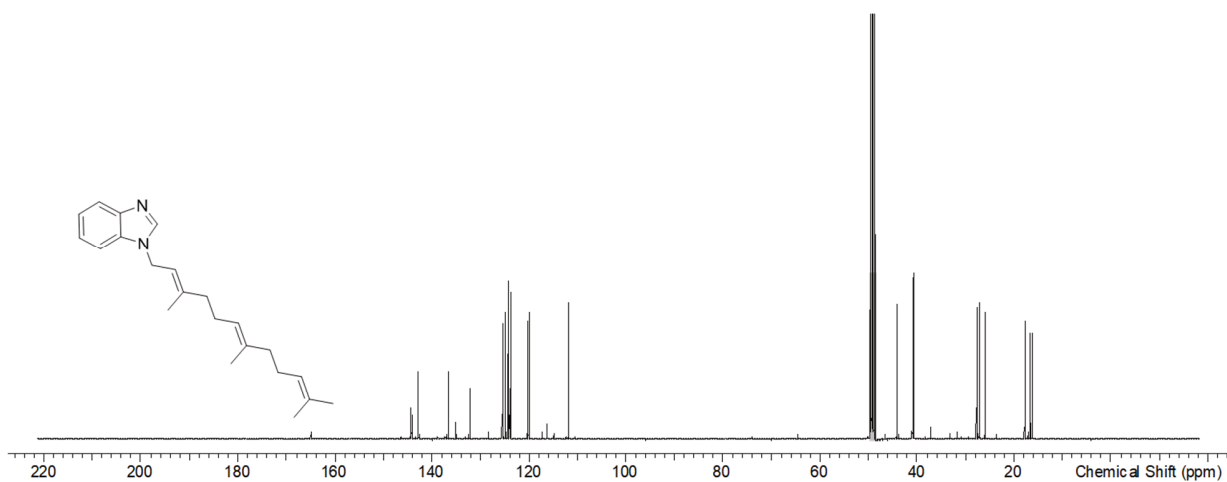
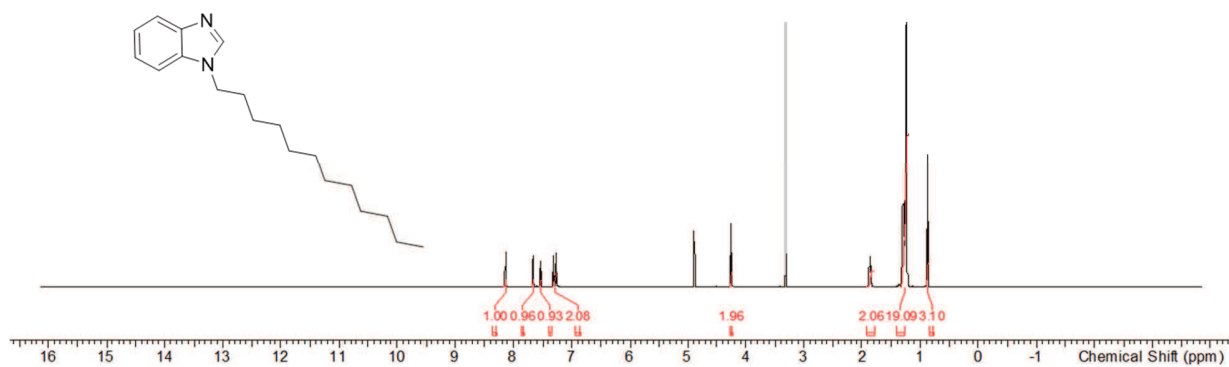
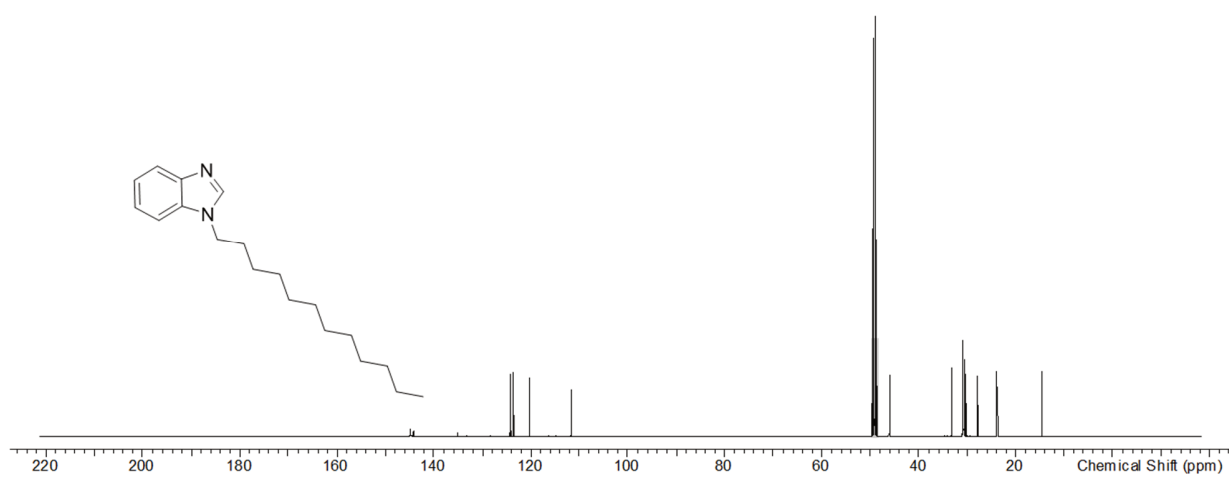


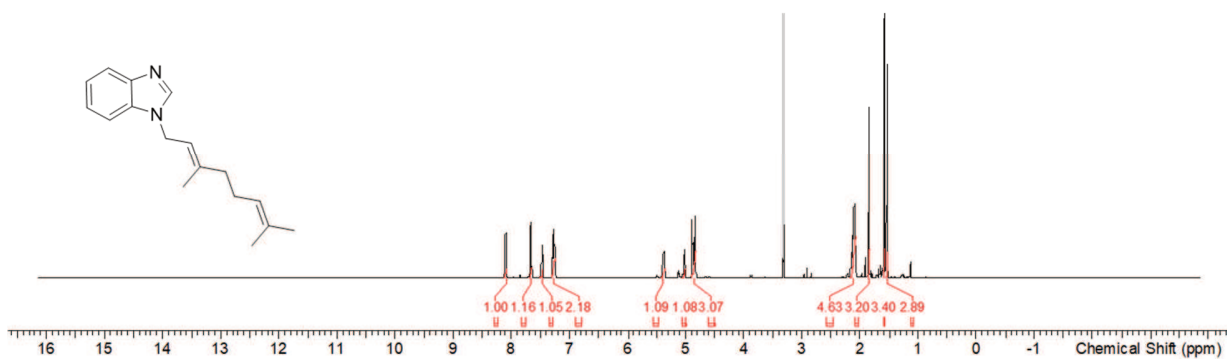
Figure S4.  $^{13}\text{C}$  spectrum of **2** in methanol- $d_4$  at 126 MHz.



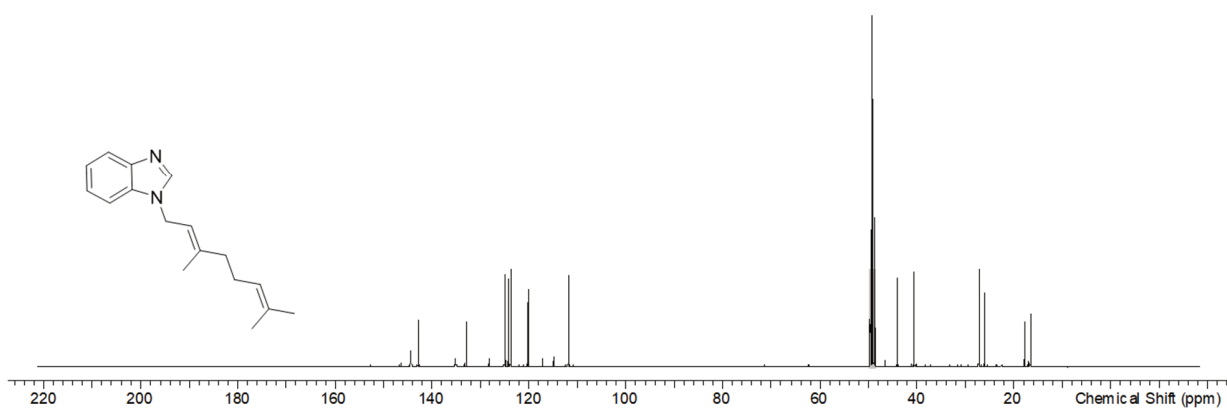
**Figure S5.**  $^1\text{H}$  spectrum of **3** in methanol- $d_4$  at 500 MHz.



**Figure S6.**  $^{13}\text{C}$  spectrum of **3** in methanol- $d_4$  at 126 MHz.



**Figure S7.**  $^1\text{H}$  spectrum of **4** in methanol- $d_4$  at 500 MHz.



**Figure S8.**  $^{13}\text{C}$  spectrum of **4** in methanol- $d_4$  at 126 MHz.

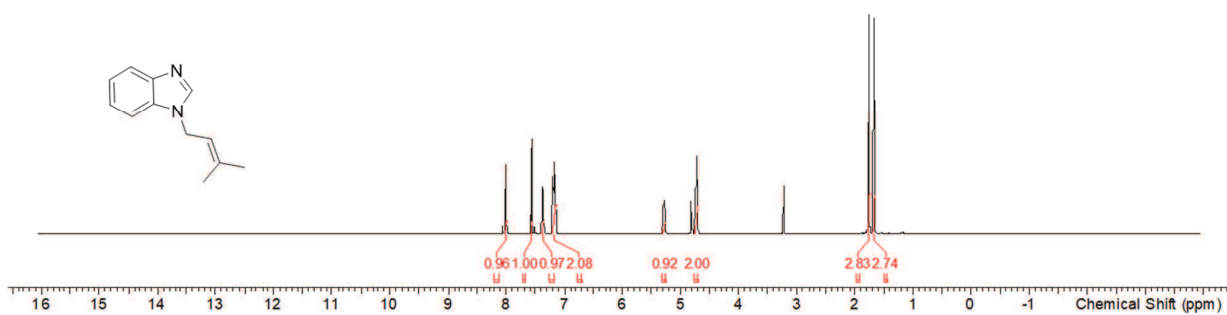


Figure S9. <sup>1</sup>H spectrum of 5 in methanol-d<sub>4</sub> at 500 MHz.

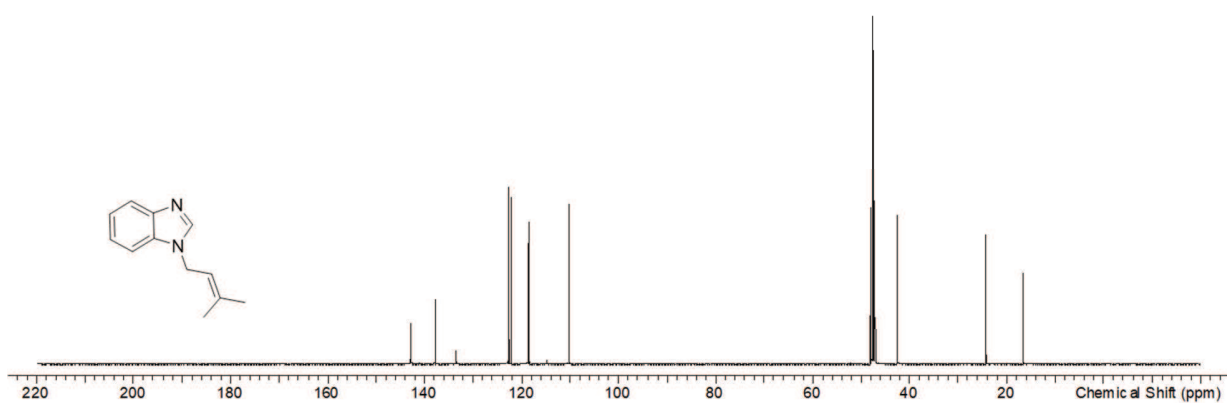
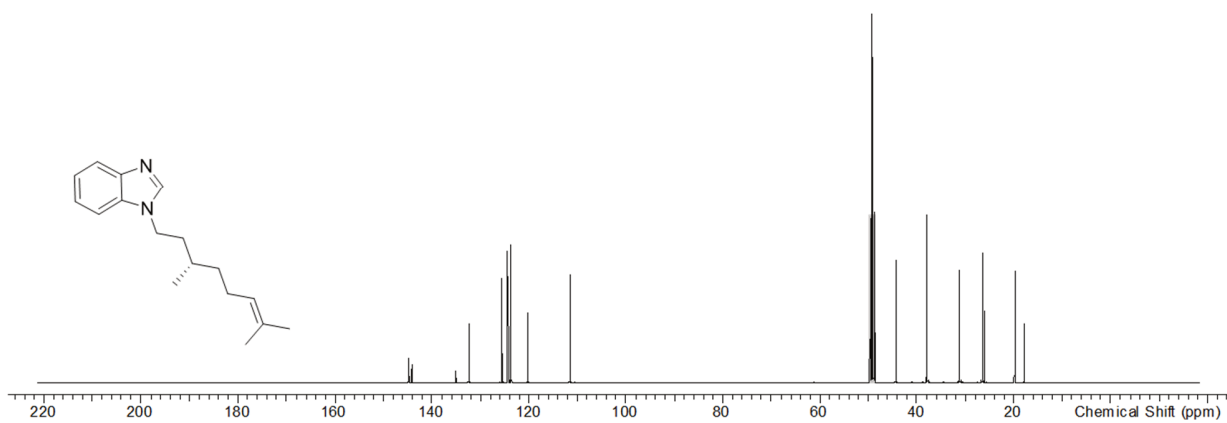
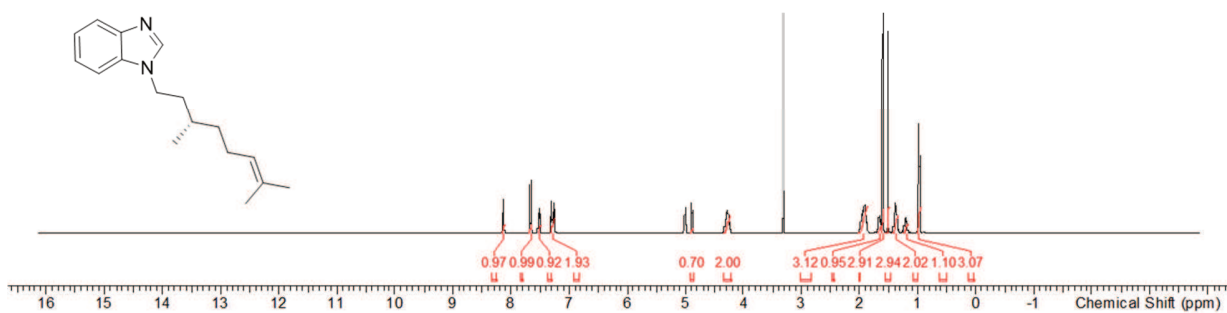
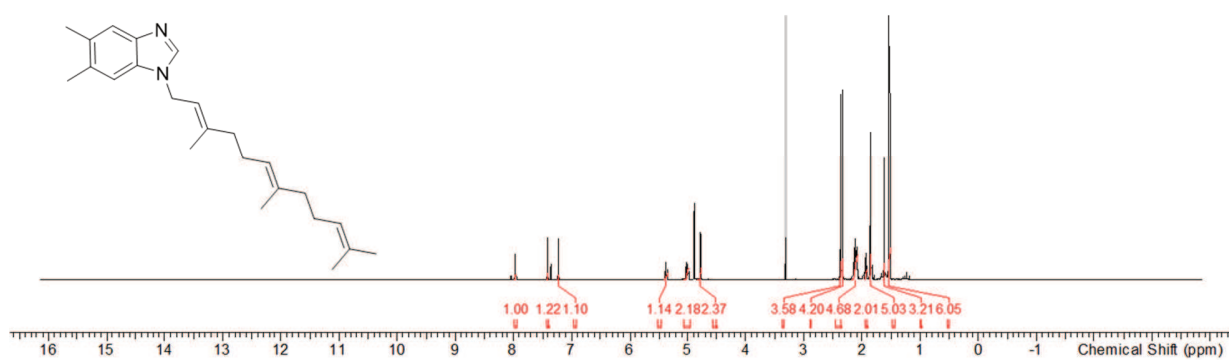
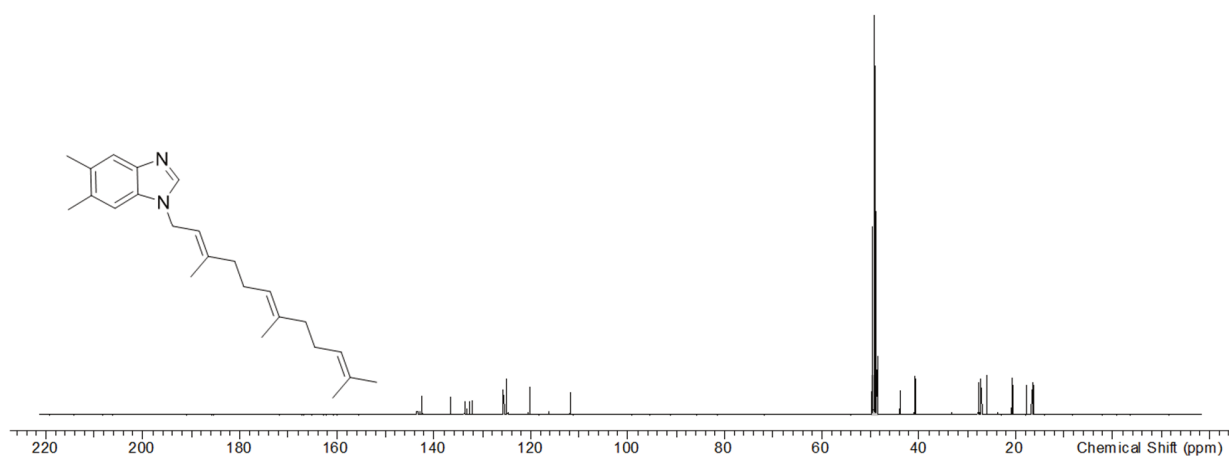


Figure S10. <sup>13</sup>C spectrum of 5 in methanol-d<sub>4</sub> at 126 MHz.





**Figure S13.**  $^1\text{H}$  spectrum of **8** in methanol- $d_4$  at 500 MHz.



**Figure S14.**  $^{13}\text{C}$  spectrum of **8** in methanol- $d_4$  at 126 MHz.

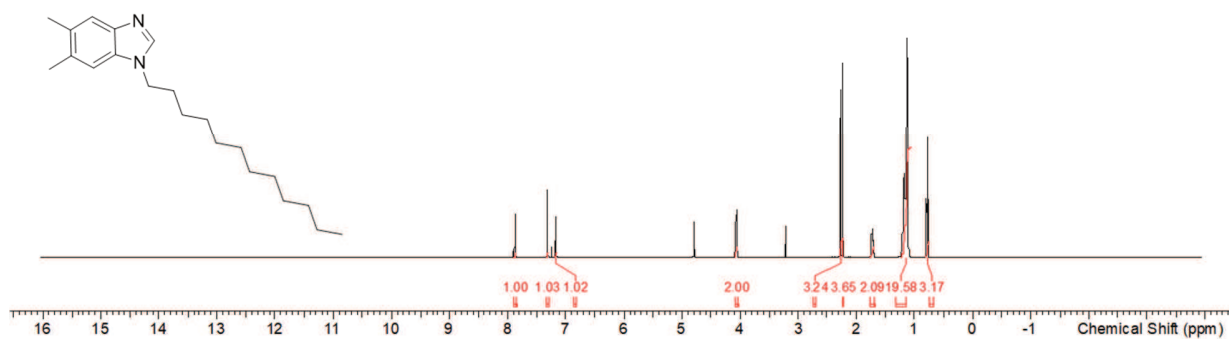


Figure S15.  $^1\text{H}$  spectrum of **9** in  $\text{methanol-d}_4$  at 500 MHz.

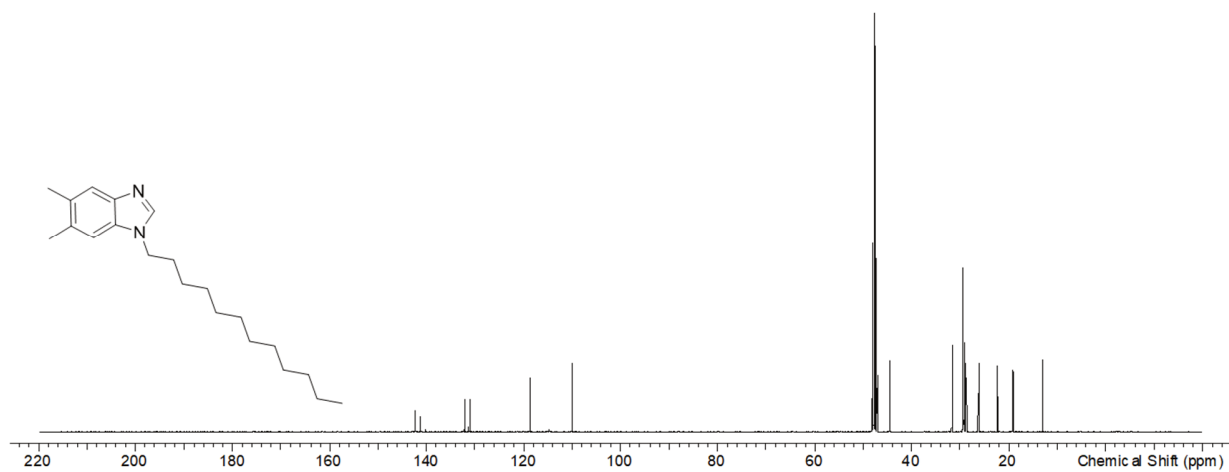
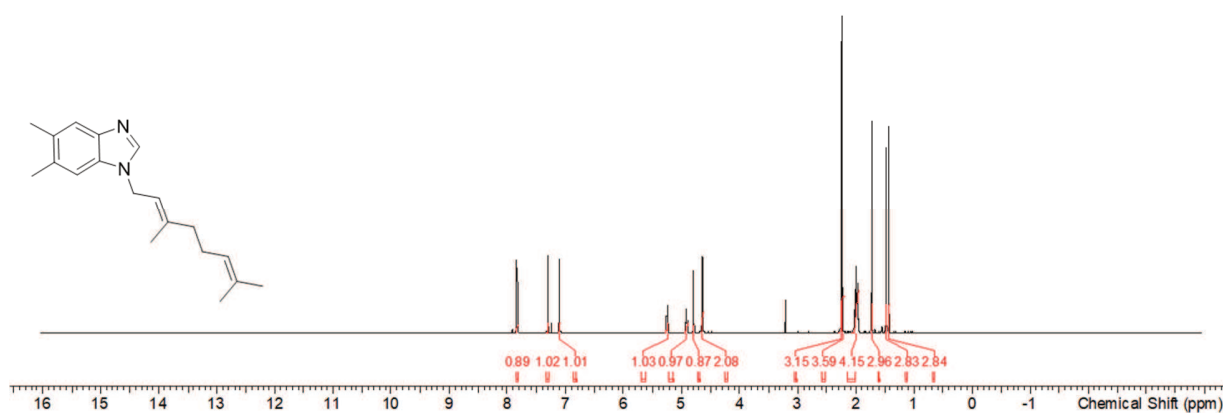
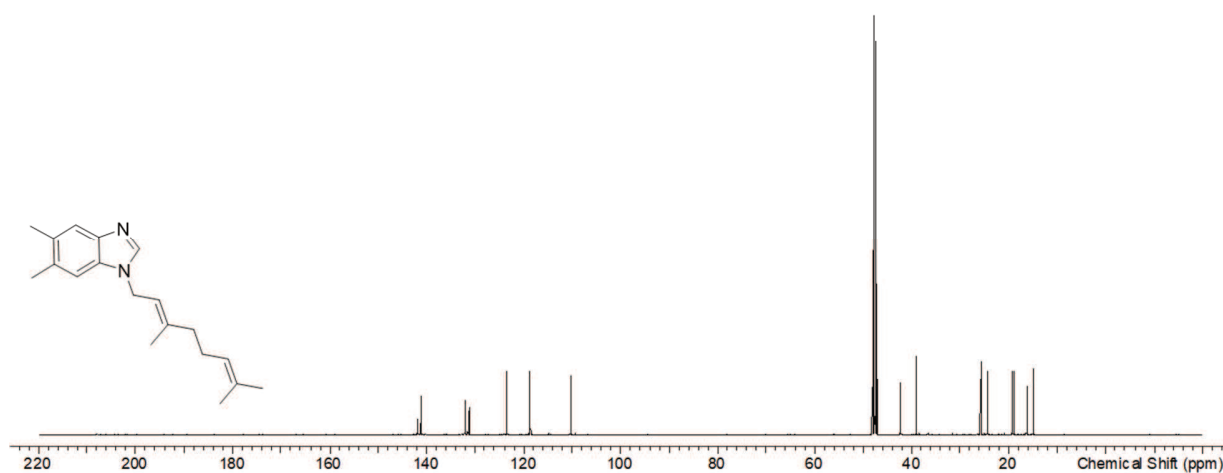


Figure S16.  $^{13}\text{C}$  spectrum of **9** in  $\text{methanol-d}_4$  at 126 MHz.





**Figure S17.** <sup>1</sup>H spectrum of **10** in methanol-*d*<sub>4</sub> at 500 MHz.



**Figure S18.** <sup>13</sup>C spectrum of **10** in methanol-*d*<sub>4</sub> at 126 MHz.

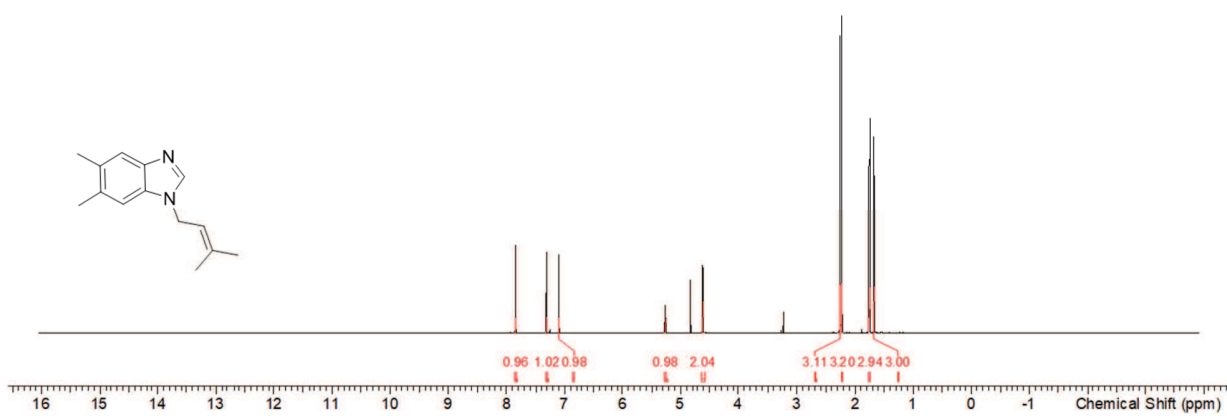


Figure S19.  $^1\text{H}$  spectrum of **11** in  $\text{methanol-}d_4$  at 500 MHz.

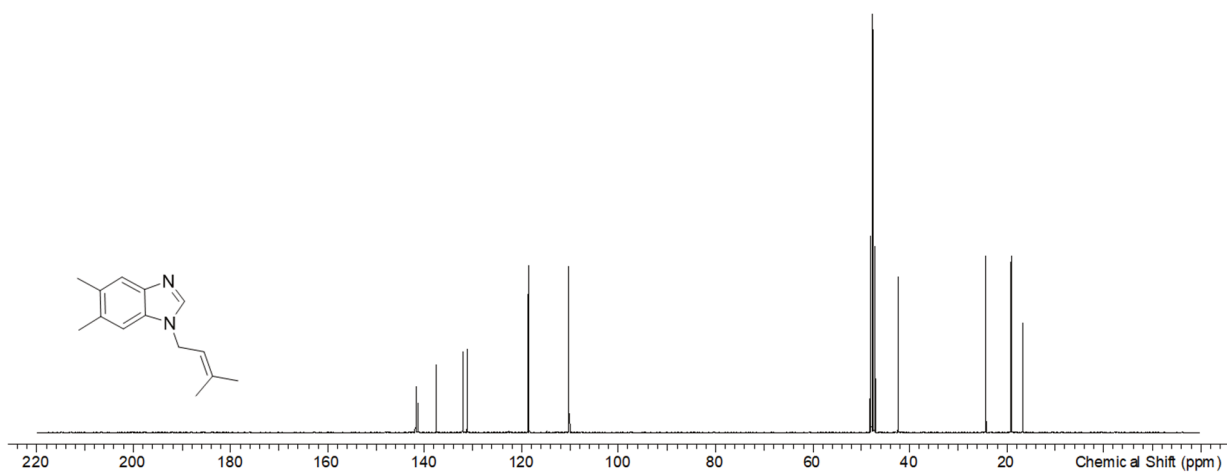


Figure S20.  $^{13}\text{C}$  spectrum of **11** in  $\text{methanol-}d_4$  at 126 MHz.

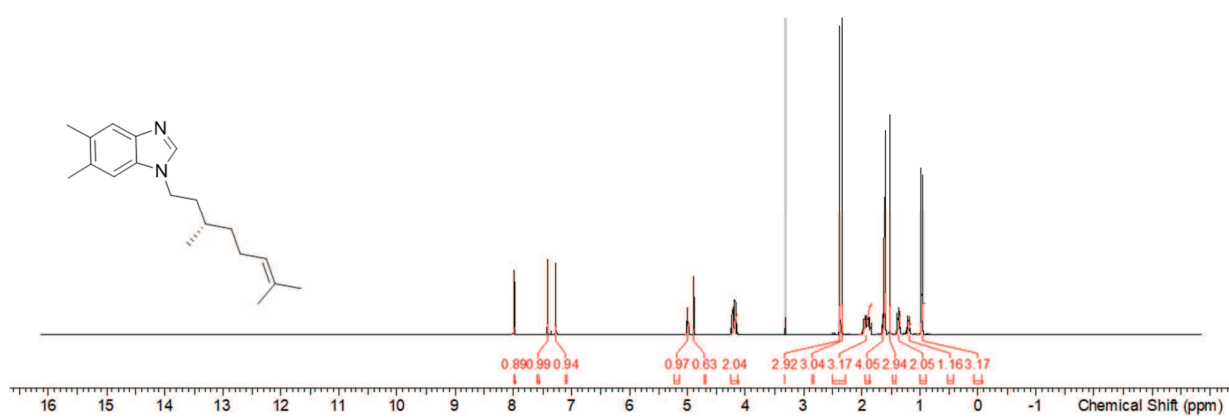


Figure S21.  $^1\text{H}$  spectrum of **12** in methanol- $d_4$  at 500 MHz.

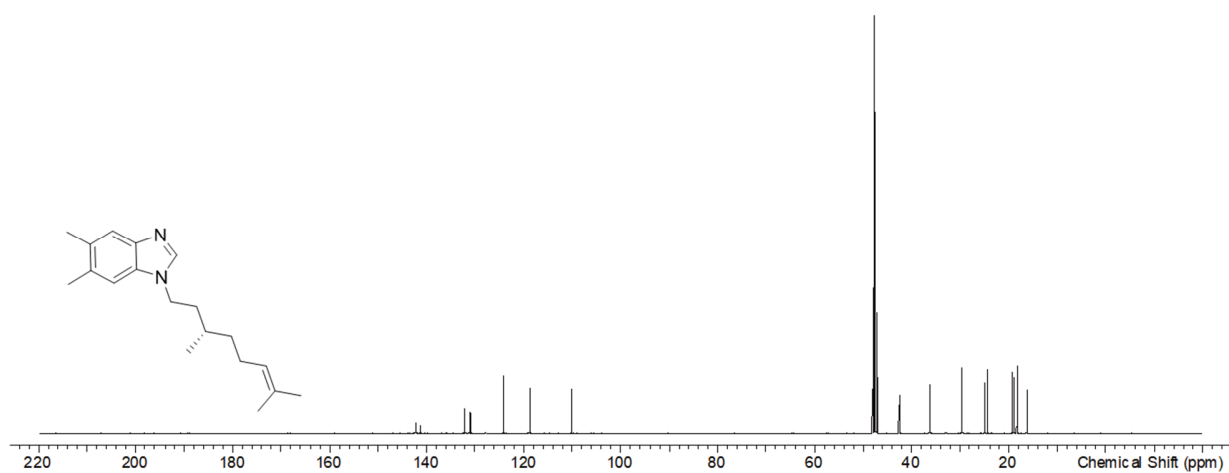
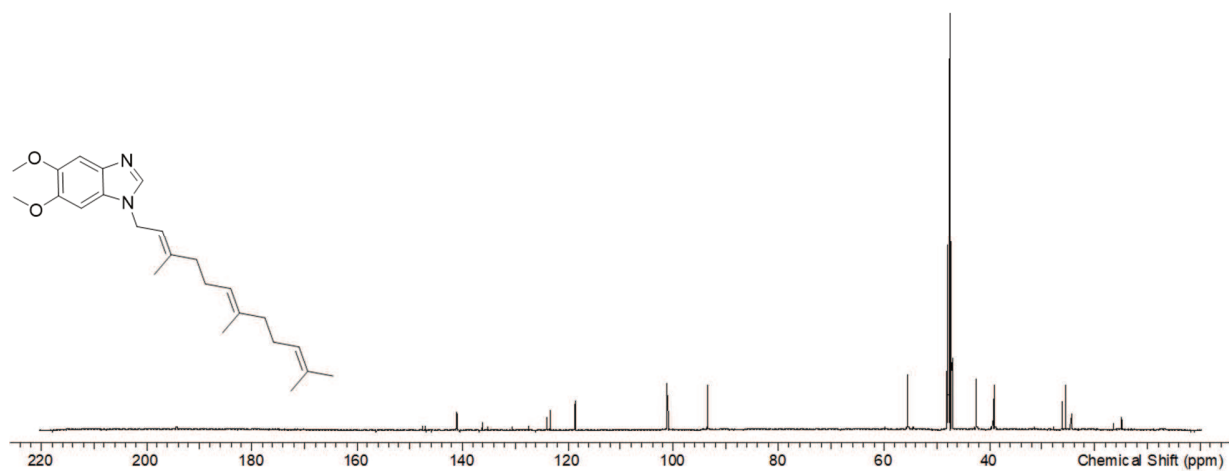
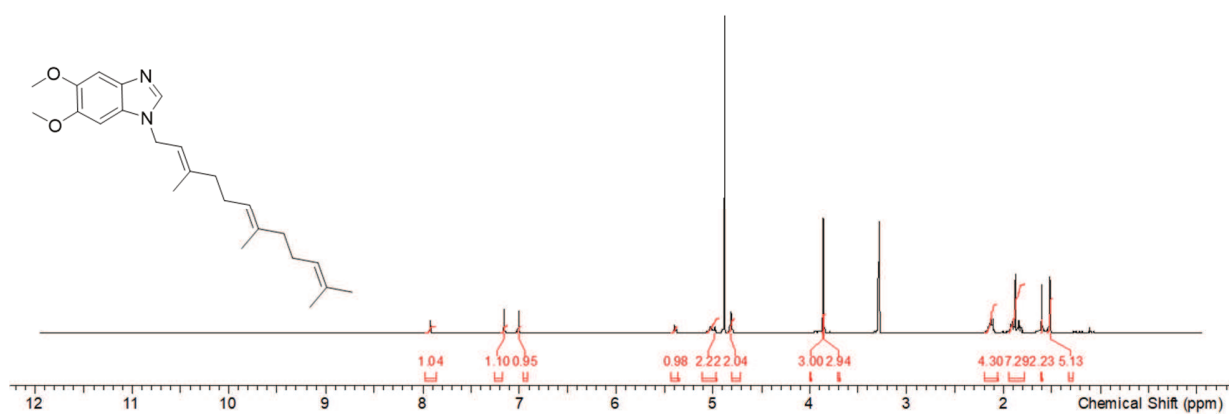


Figure S22.  $^{13}\text{C}$  spectrum of **12** in methanol- $d_4$  at 126 MHz.



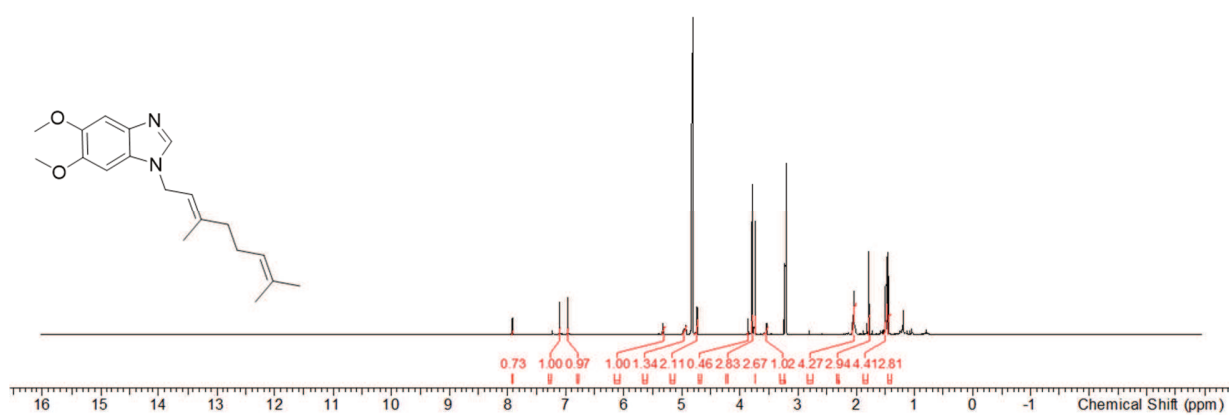


Figure S25. <sup>1</sup>H spectrum of **15** in methanol-d<sub>4</sub> at 500 MHz.

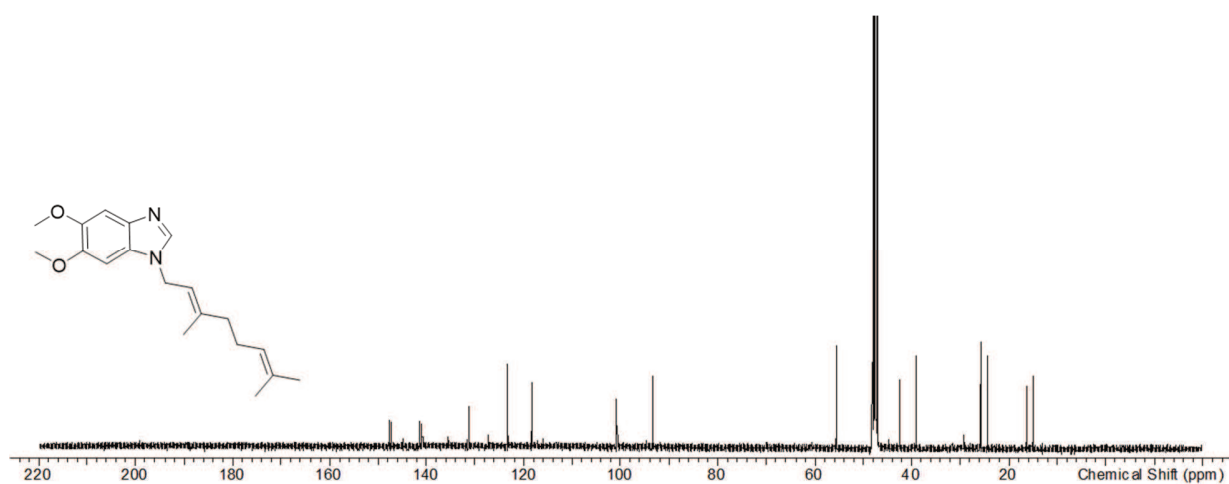


Figure S26. <sup>13</sup>C spectrum of **15** in methanol-d<sub>4</sub> at 126 MHz.

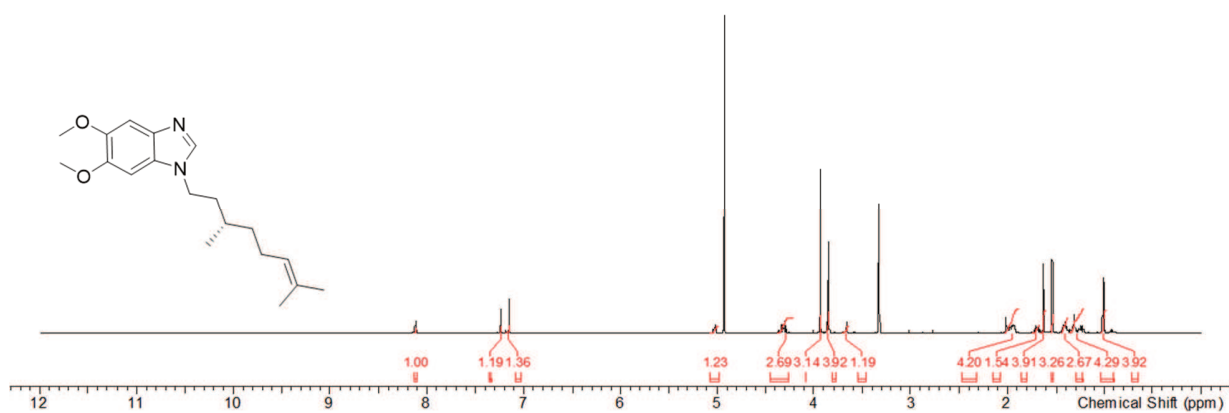


Figure S27.  $^1\text{H}$  spectrum of **16** in methanol- $d_4$  at 500 MHz.

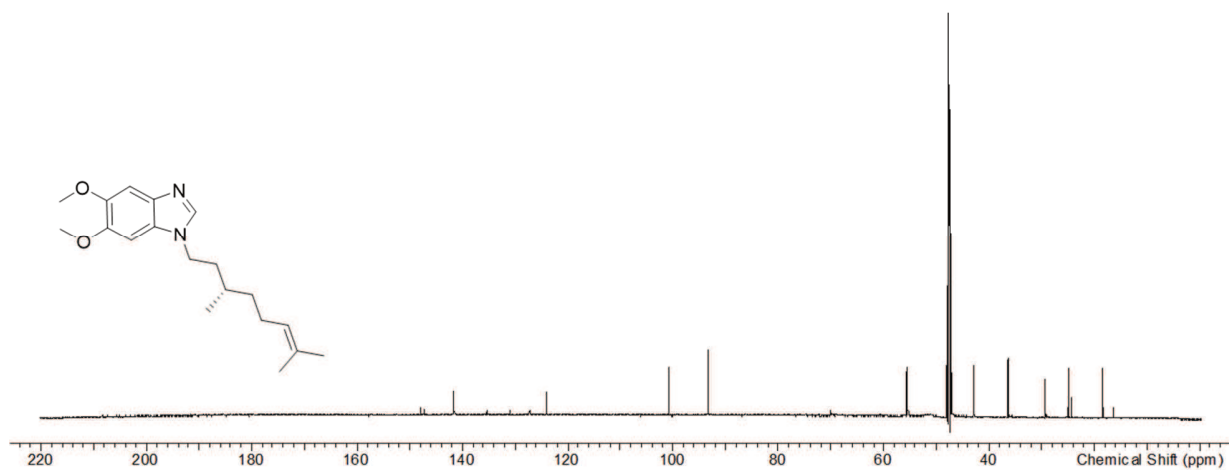
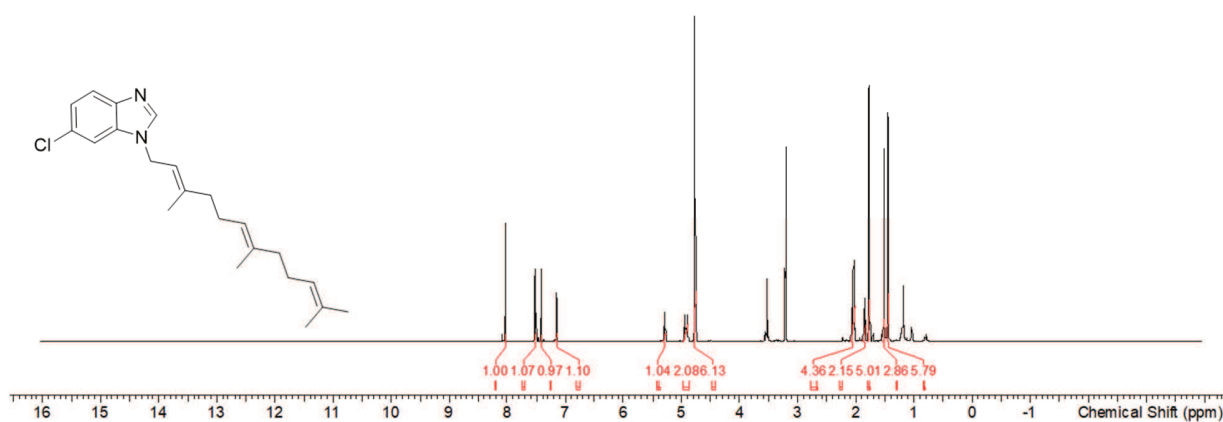
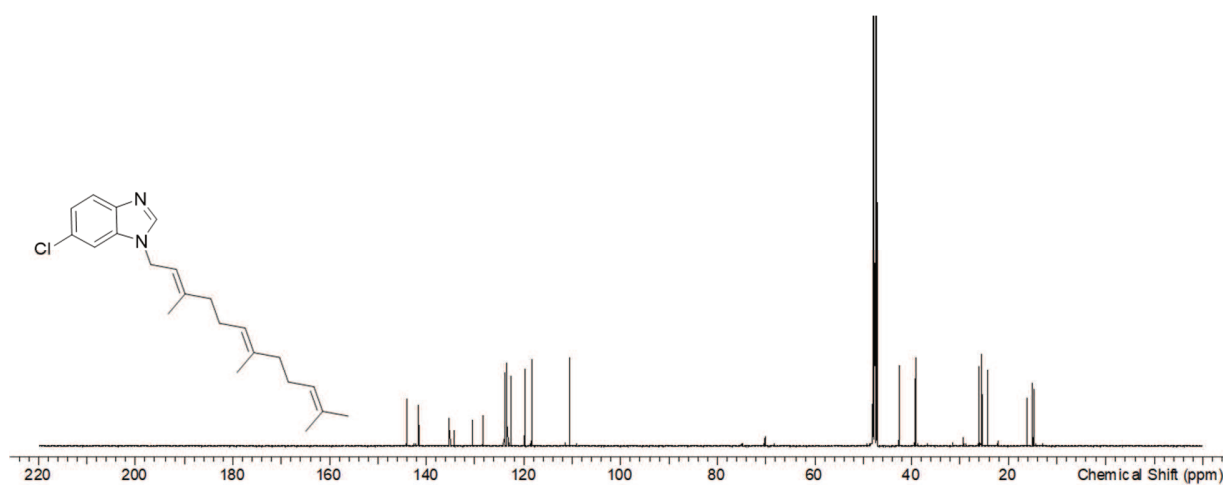


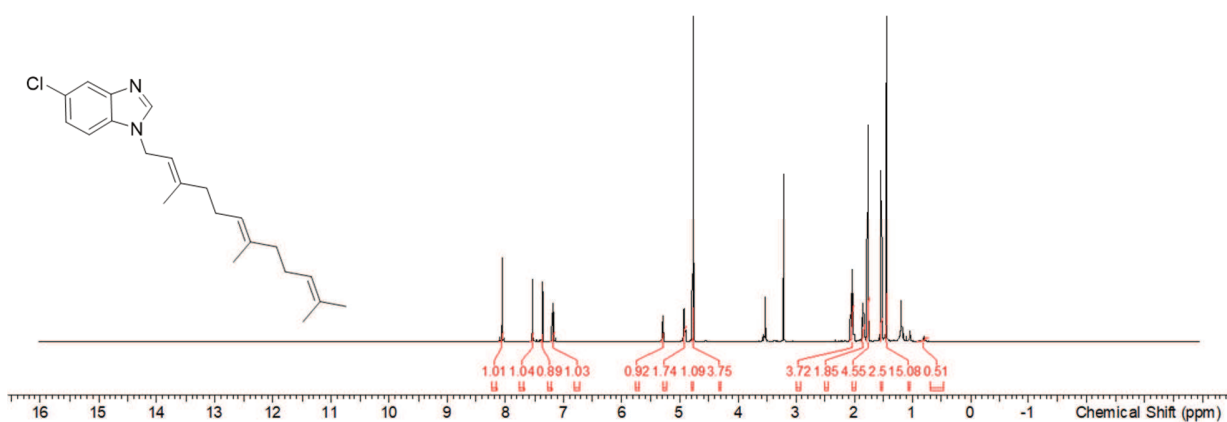
Figure S28.  $^{13}\text{C}$  spectrum of **16** in methanol- $d_4$  at 126 MHz.



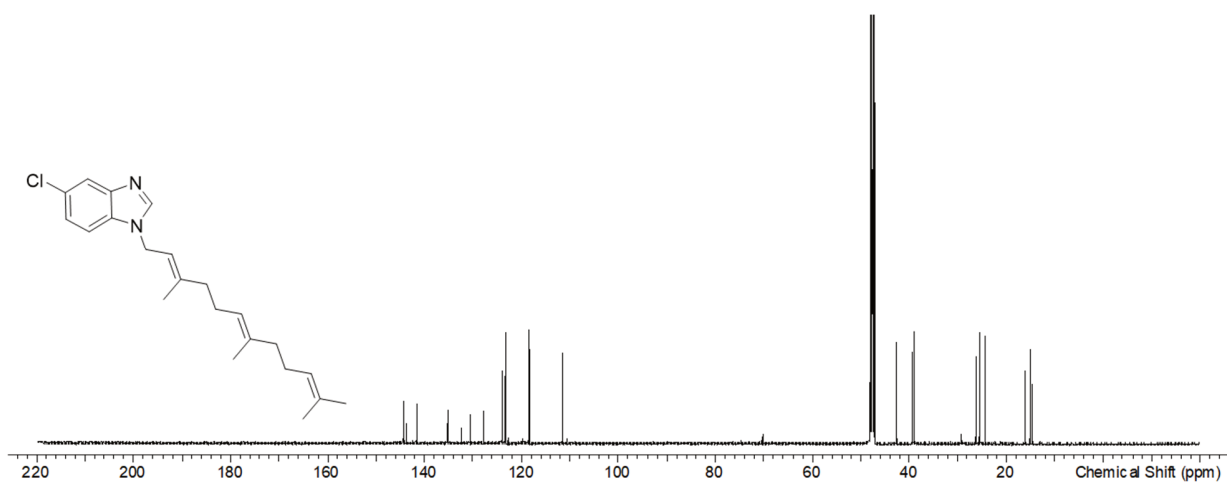
**Figure S29.**  $^1\text{H}$  spectrum of **18a** in methanol- $d_4$  at 500 MHz.



**Figure S30.**  $^{13}\text{C}$  spectrum of **18a** in methanol- $d_4$  at 126 MHz.



**Figure S31.**  $^1\text{H}$  spectrum of **18b** in methanol- $d_4$  at 500 MHz.



**Figure S32.**  $^{13}\text{C}$  spectrum of **18b** in methanol- $d_4$  at 126 MHz.



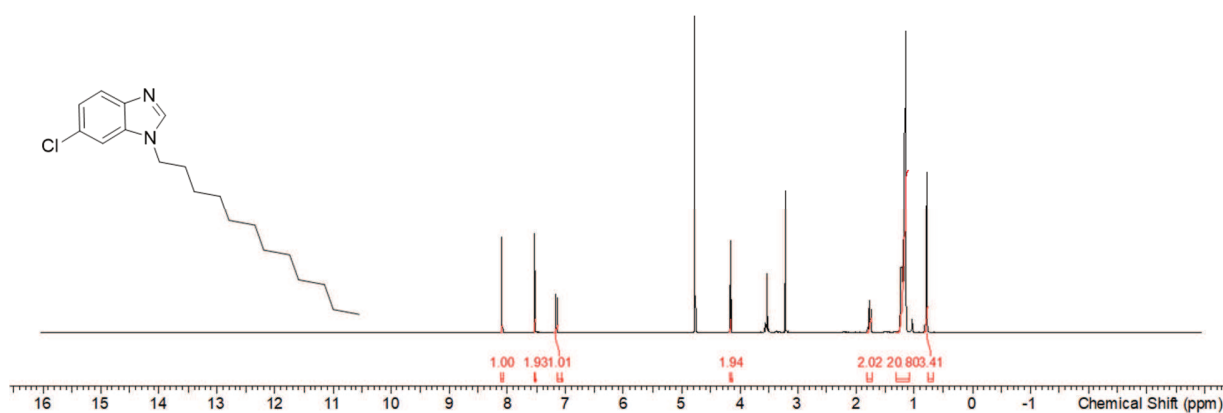


Figure S33.  $^1\text{H}$  spectrum of **19a** in methanol- $d_4$  at 500 MHz.

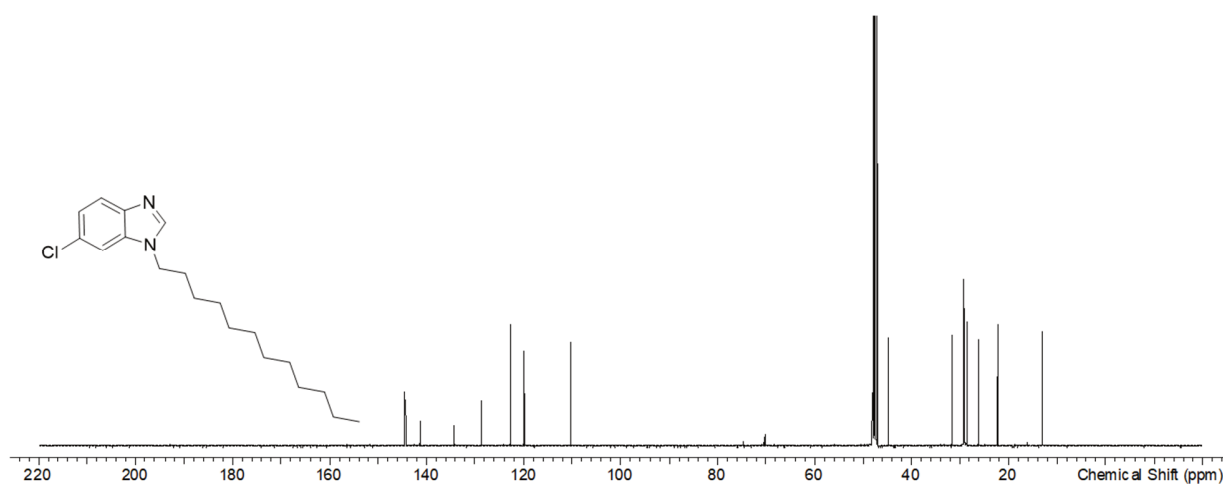


Figure S34.  $^{13}\text{C}$  spectrum of **19a** in methanol- $d_4$  at 126 MHz.

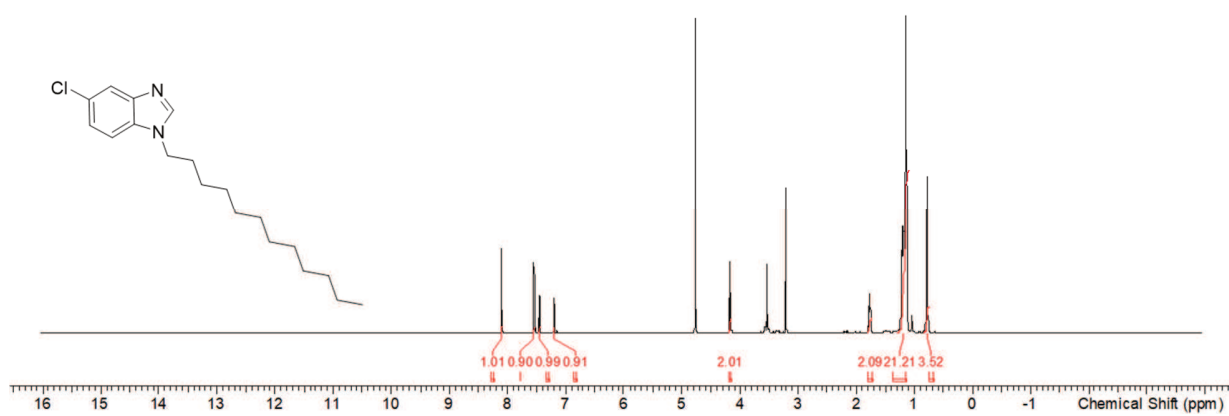


Figure S35.  $^1\text{H}$  spectrum of **19b** in methanol- $d_4$  at 500 MHz.

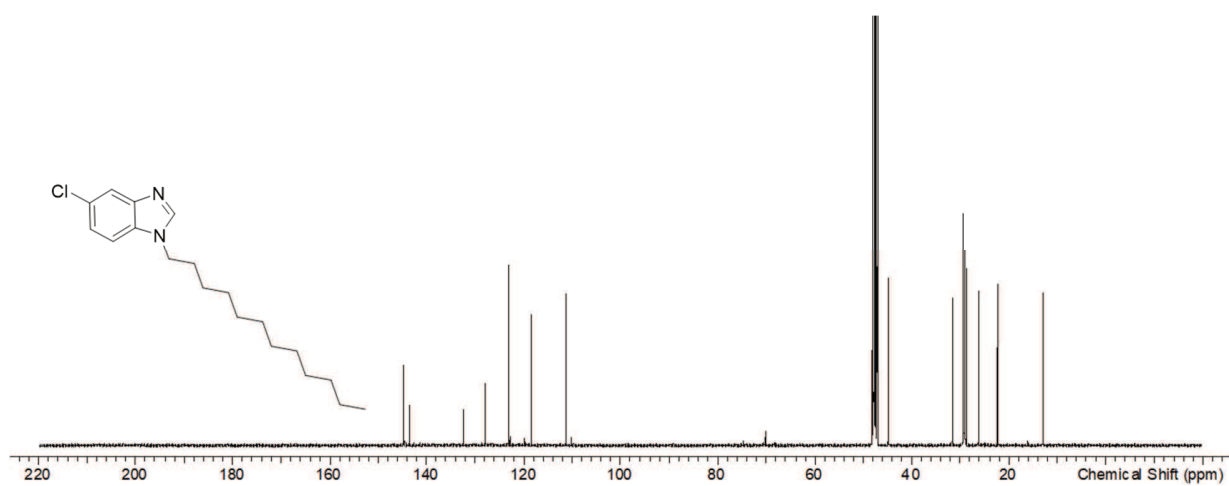
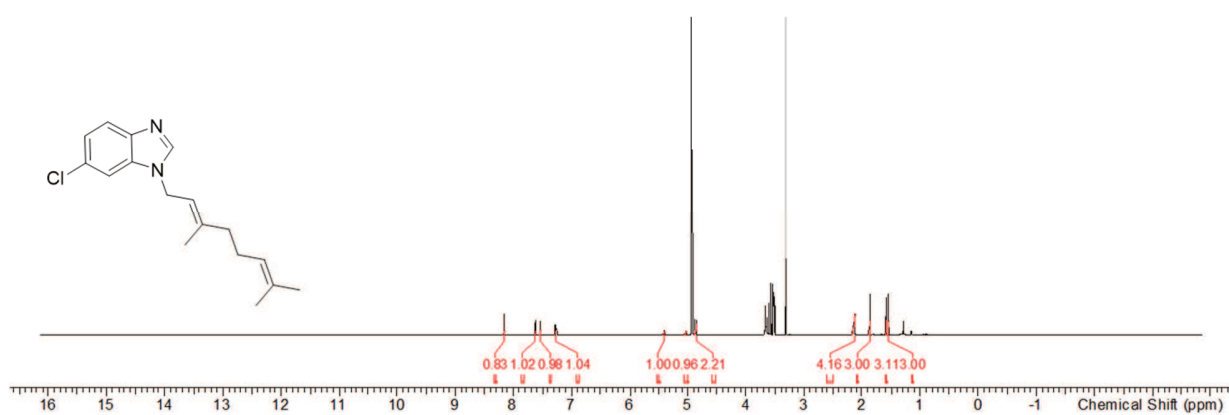
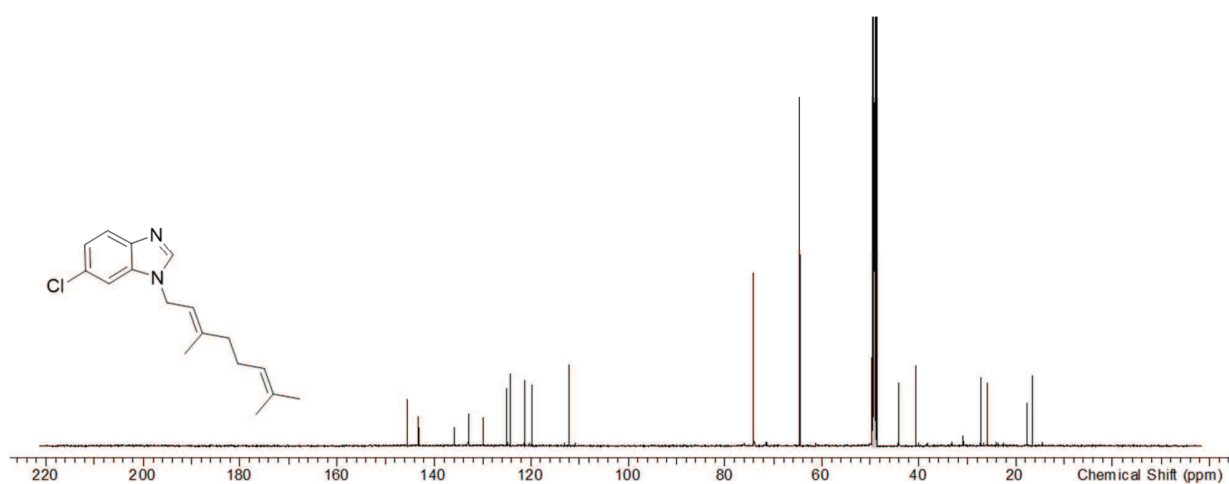


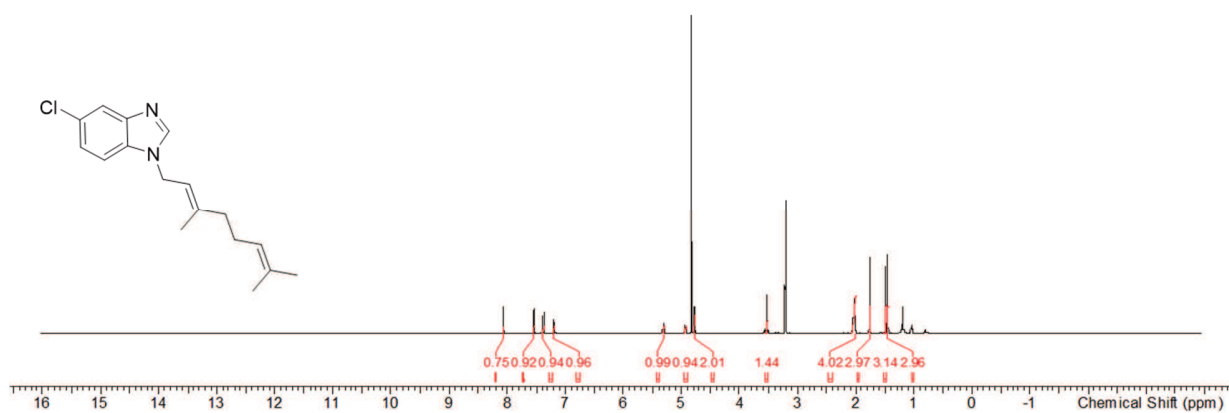
Figure S36.  $^{13}\text{C}$  spectrum of **19b** in methanol- $d_4$  at 126 MHz.



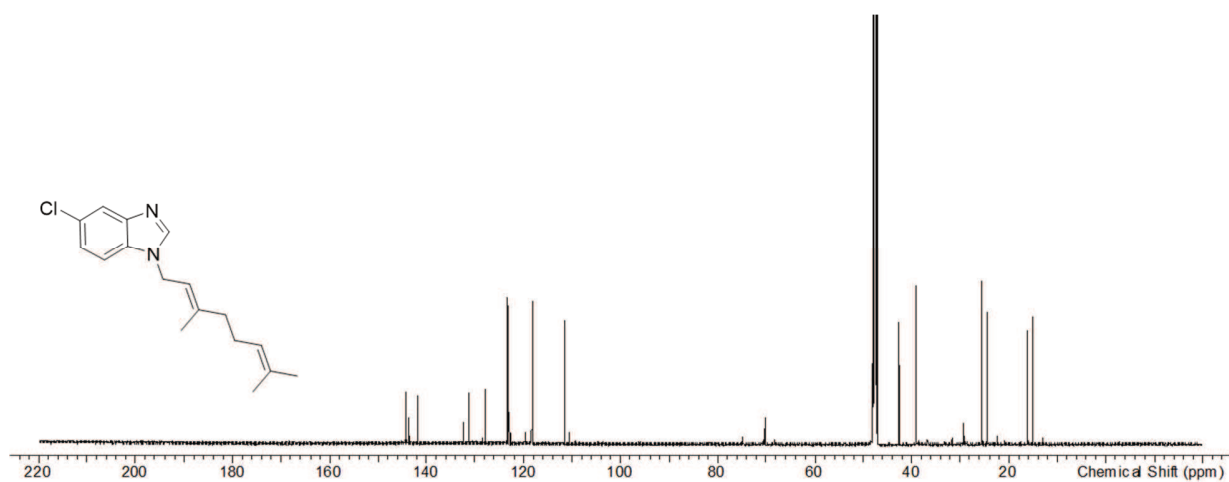
**Figure S37.**  $^1\text{H}$  spectrum of **20a** in methanol- $d_4$  at 500 MHz.



**Figure S38.**  $^{13}\text{C}$  spectrum of **20a** in methanol- $d_4$  at 126 MHz.



**Figure S39.**  $^1\text{H}$  spectrum of **20b** in methanol- $d_4$  at 500 MHz.



**Figure S40.**  $^{13}\text{C}$  spectrum of **20b** in methanol- $d_4$  at 126 MHz.

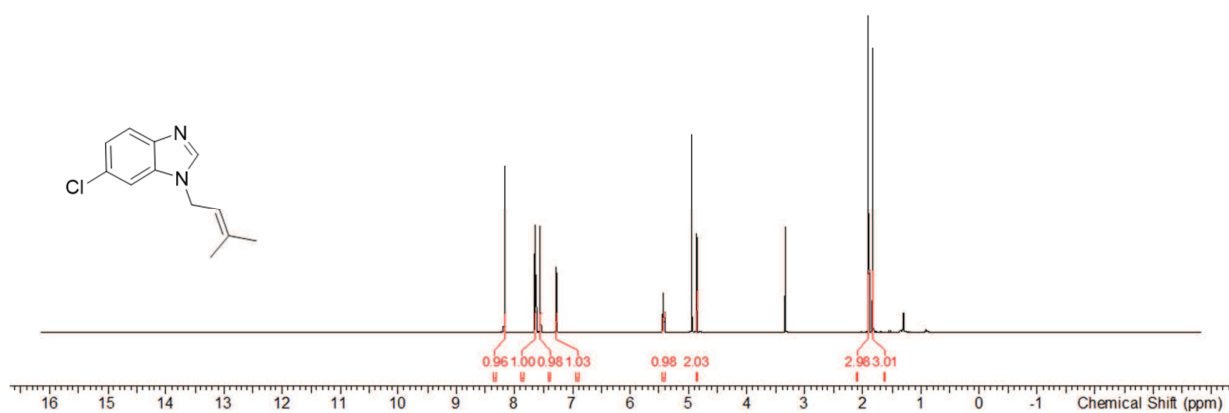


Figure S41. <sup>1</sup>H spectrum of **21a** in methanol-d<sub>4</sub> at 500 MHz.

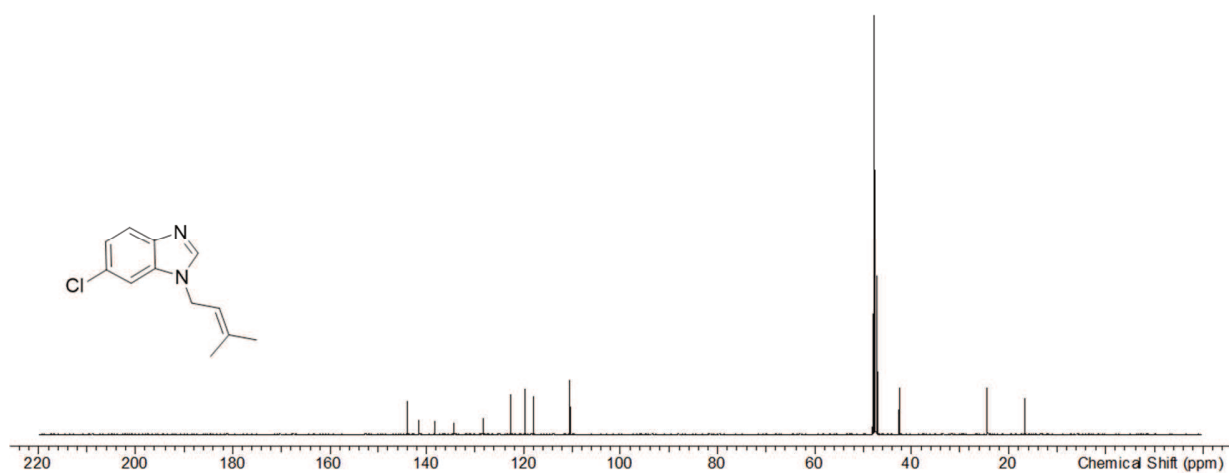


Figure S42. <sup>13</sup>C spectrum of **21a** in methanol-d<sub>4</sub> at 126 MHz.

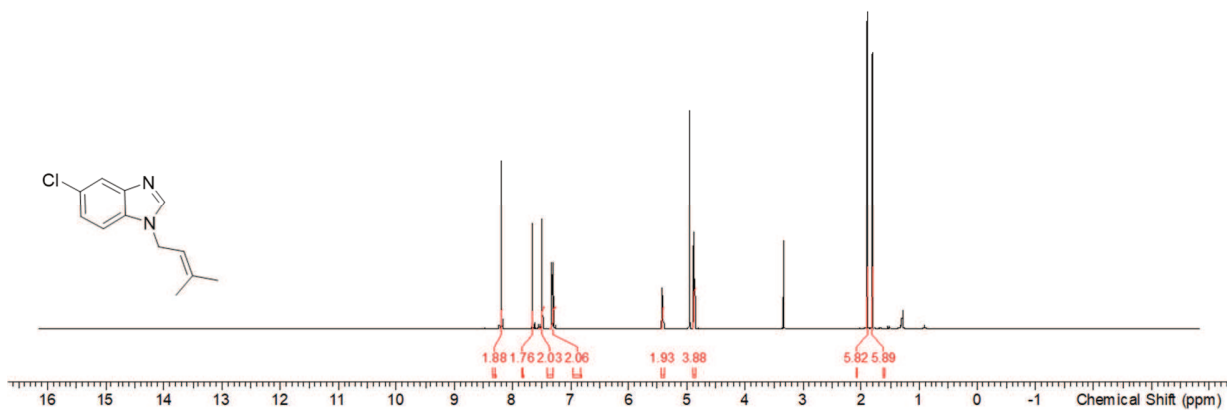


Figure S43.  $^1\text{H}$  spectrum of **21b** in methanol- $d_4$  at 500 MHz.

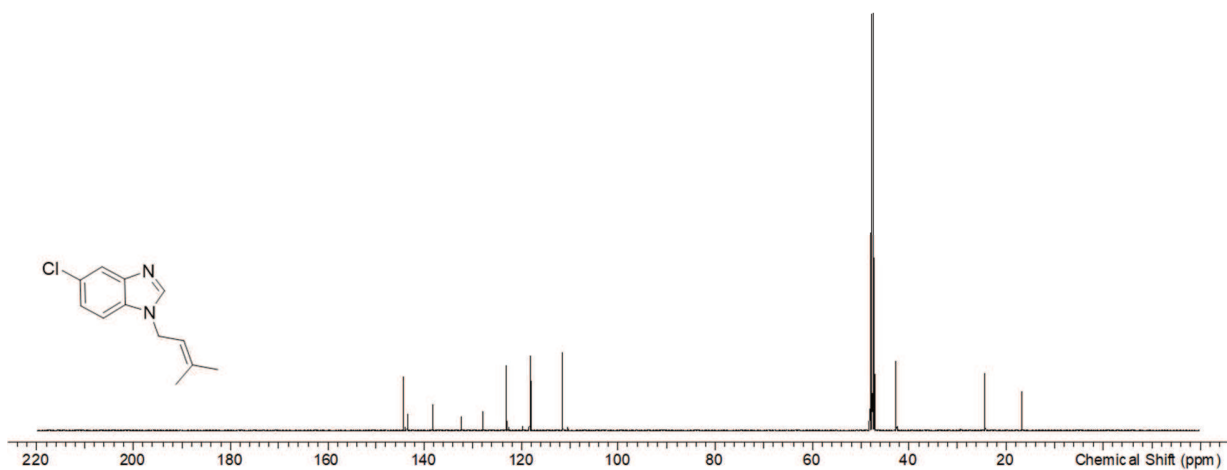


Figure S44.  $^{13}\text{C}$  spectrum of **21b** in methanol- $d_4$  at 126 MHz.

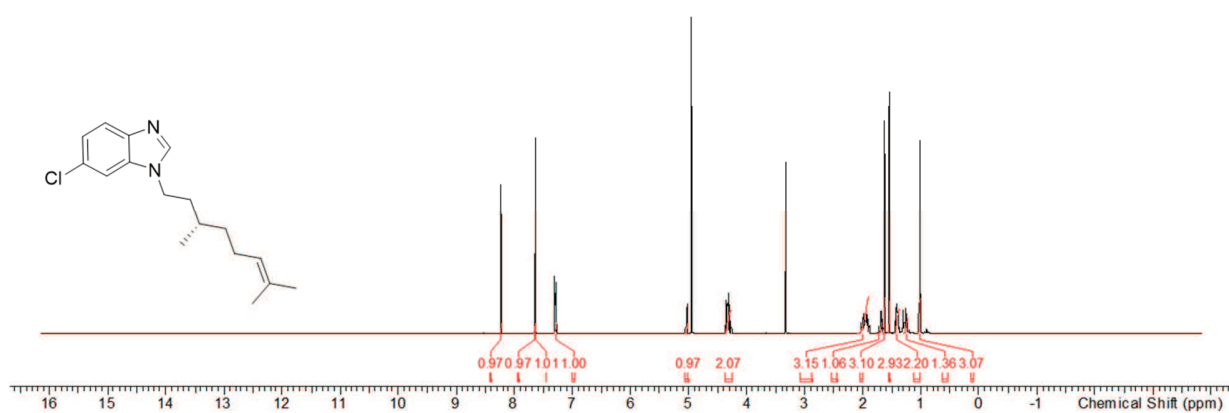


Figure S45. <sup>1</sup>H spectrum of **22a** in methanol-*d*<sub>4</sub> at 500 MHz.

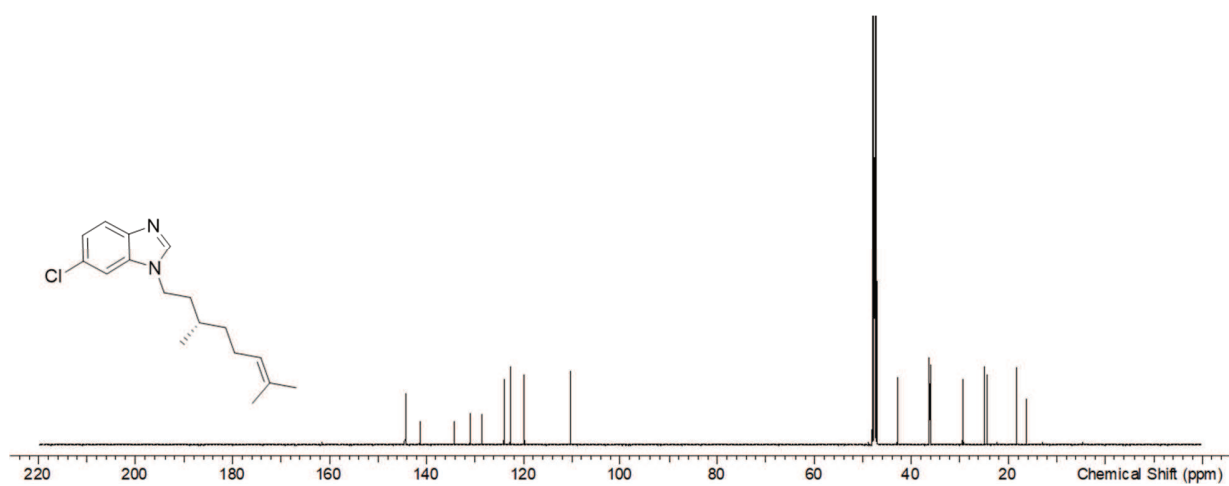


Figure S46. <sup>13</sup>C spectrum of **22a** in methanol-*d*<sub>4</sub> at 126 MHz.

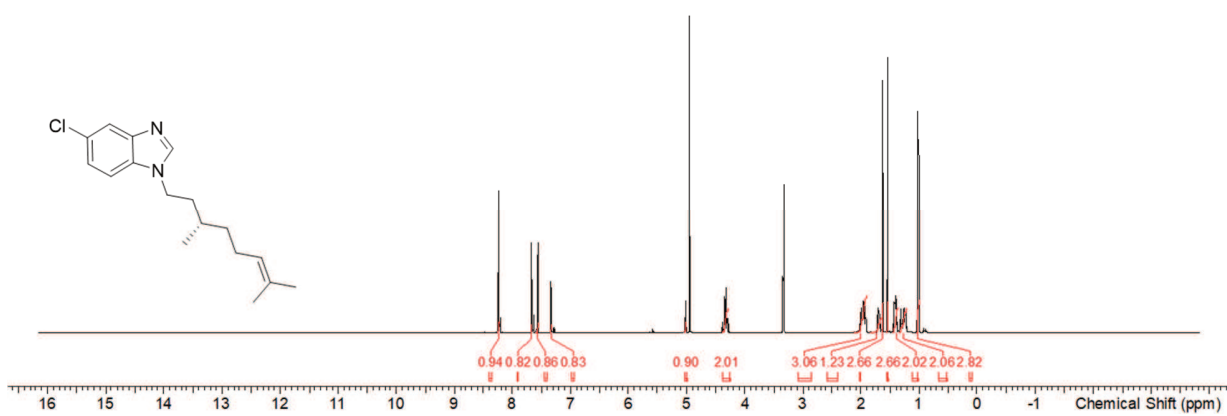


Figure S47. <sup>1</sup>H spectrum of **22b** in methanol-d<sub>4</sub> at 500 MHz.

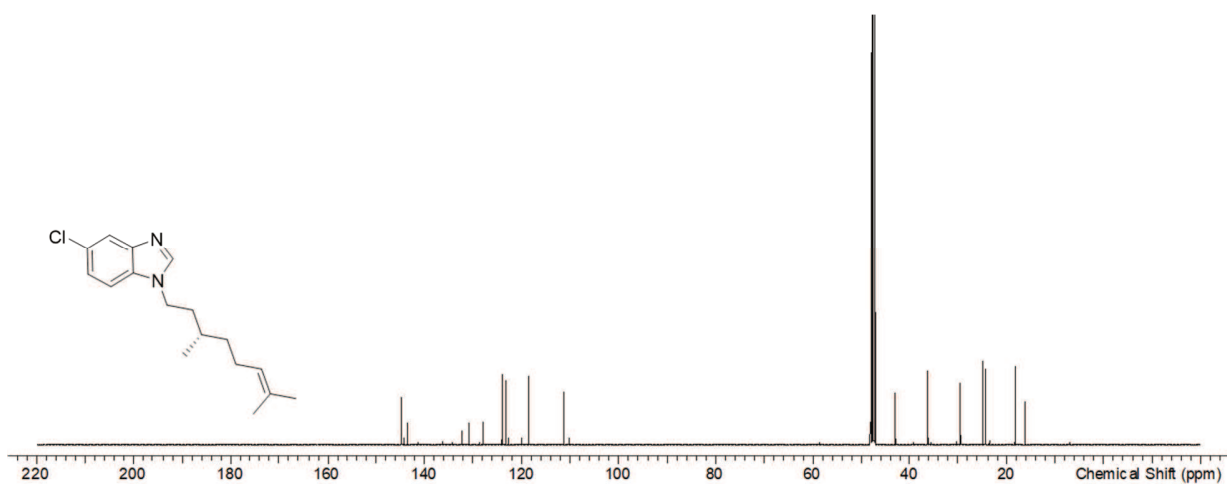
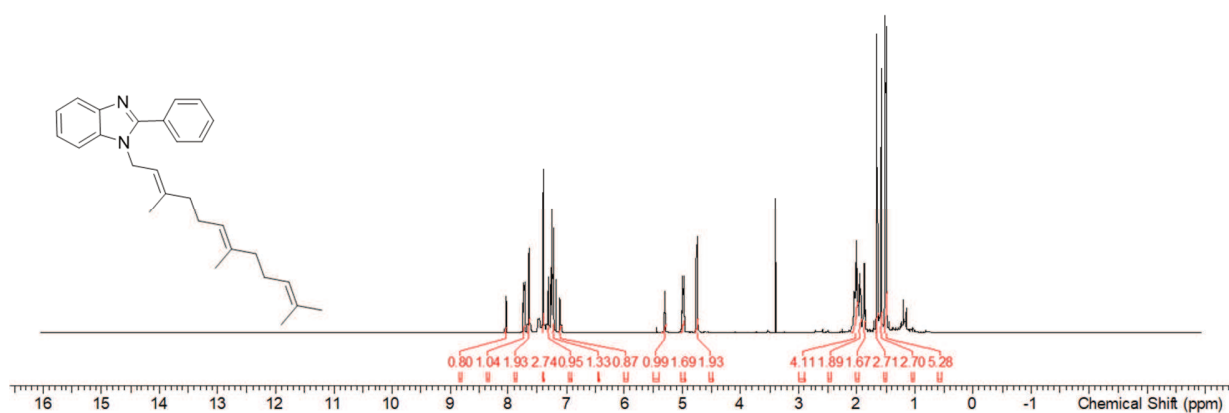
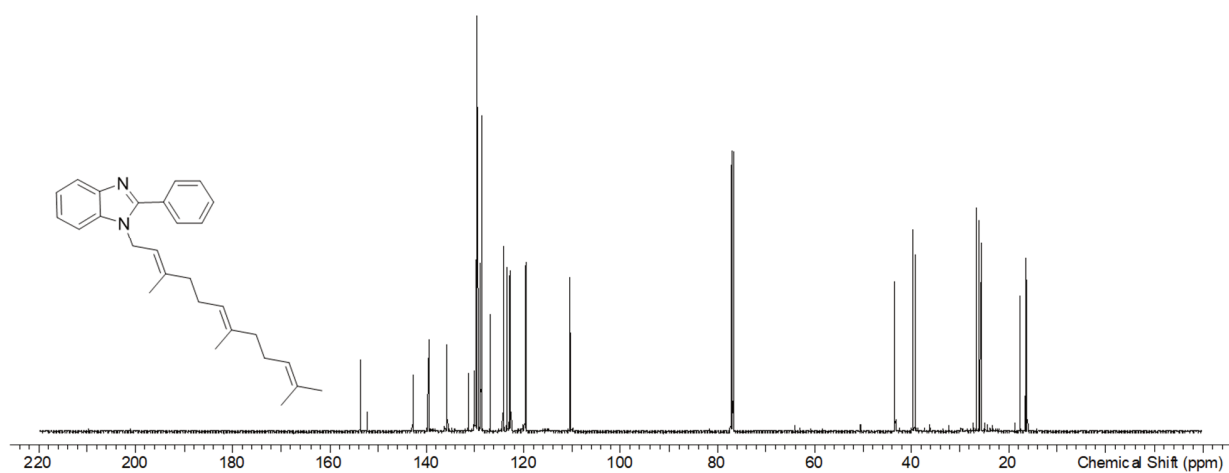


Figure S48. <sup>13</sup>C spectrum of **22b** in methanol-d<sub>4</sub> at 126 MHz.





**Figure S49.**  $^1\text{H}$  spectrum of **24** in chloroform-*d* at 500 MHz.



**Figure S50.**  $^{13}\text{C}$  spectrum of **24** in chloroform-*d* at 126 MHz.

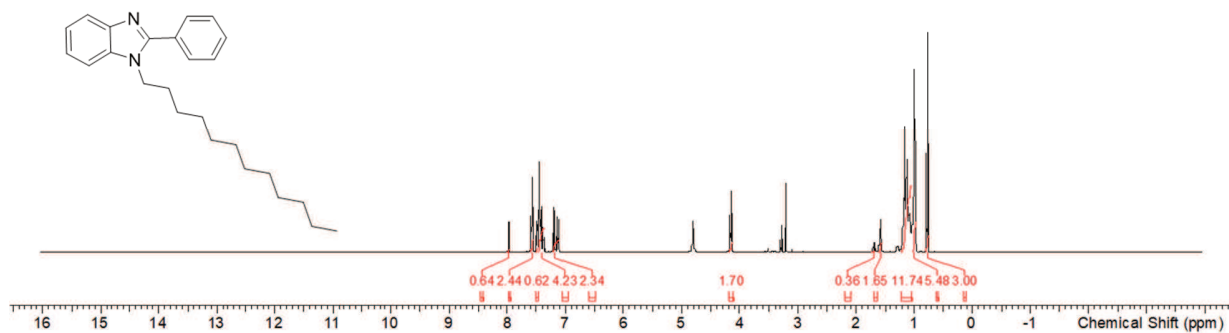


Figure S51.  $^1\text{H}$  spectrum of **25** in methanol- $d_4$  at 500 MHz.

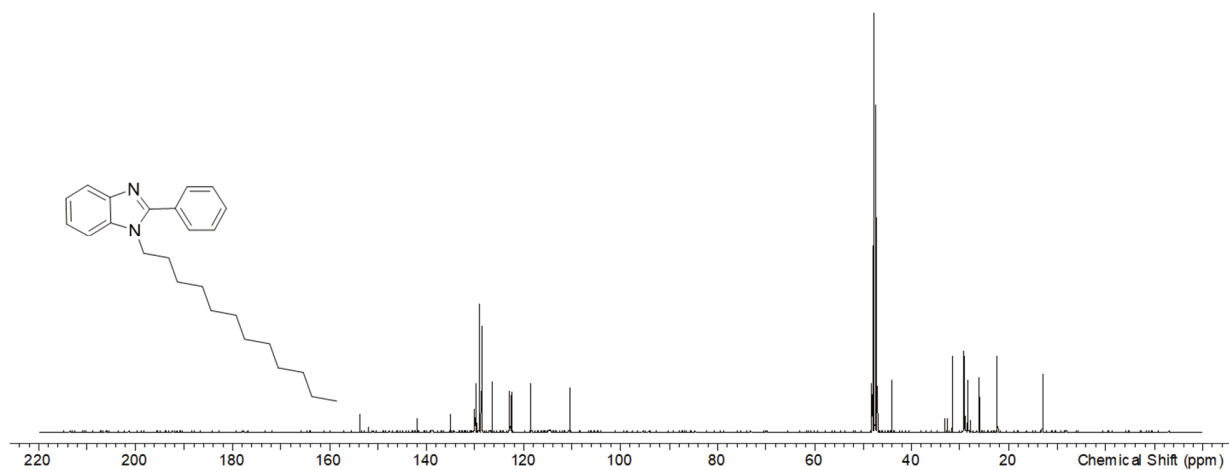


Figure S52.  $^{13}\text{C}$  spectrum of **25** in methanol- $d_4$  at 126 MHz.

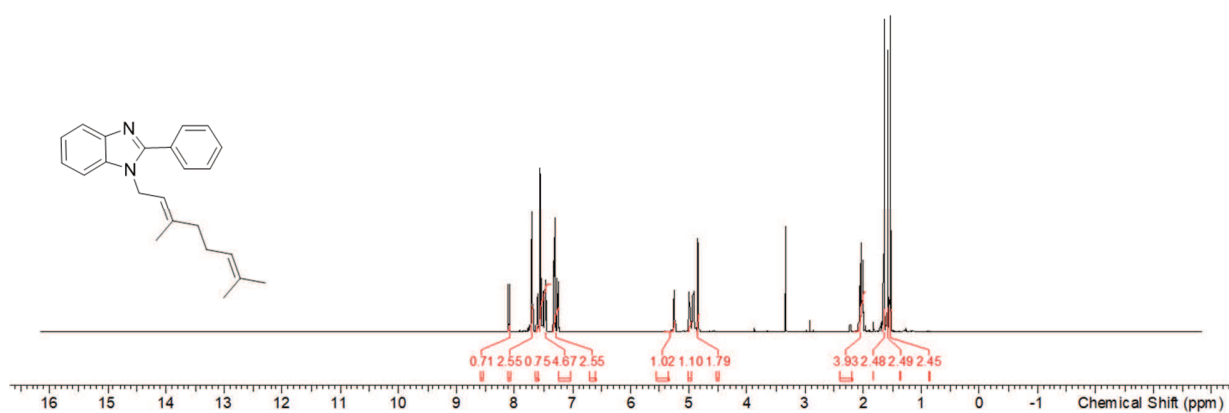


Figure S53. <sup>1</sup>H spectrum of **26** in methanol-*d*<sub>4</sub> at 500 MHz.

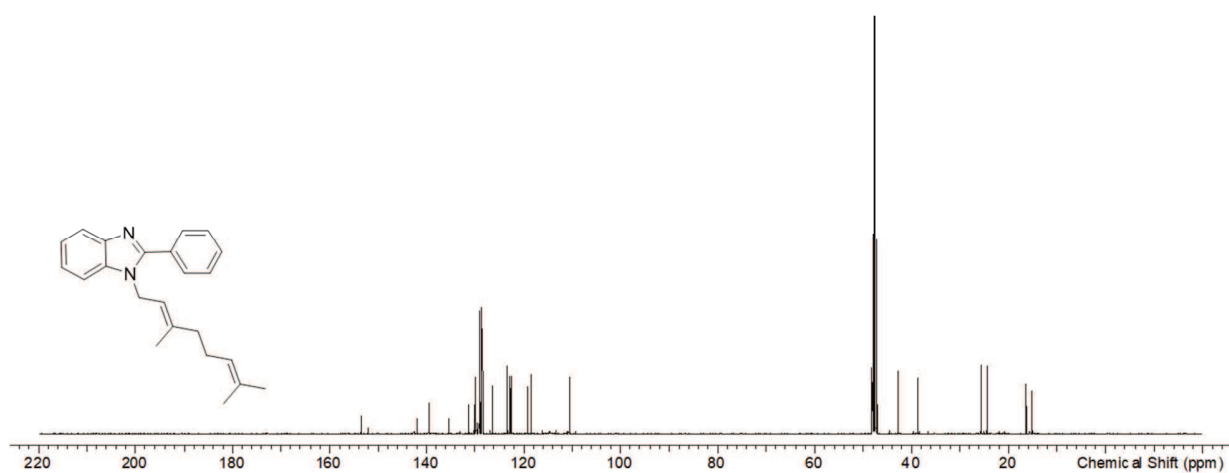


Figure S54. <sup>13</sup>C spectrum of **26** in methanol-*d*<sub>4</sub> at 126 MHz.

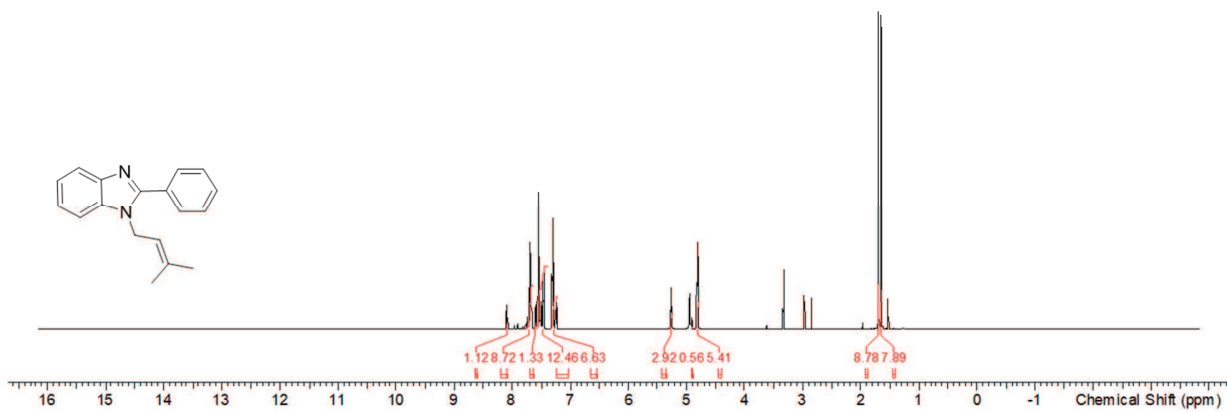


Figure S55. <sup>1</sup>H spectrum of **27** in methanol-*d*<sub>4</sub> at 500 MHz.

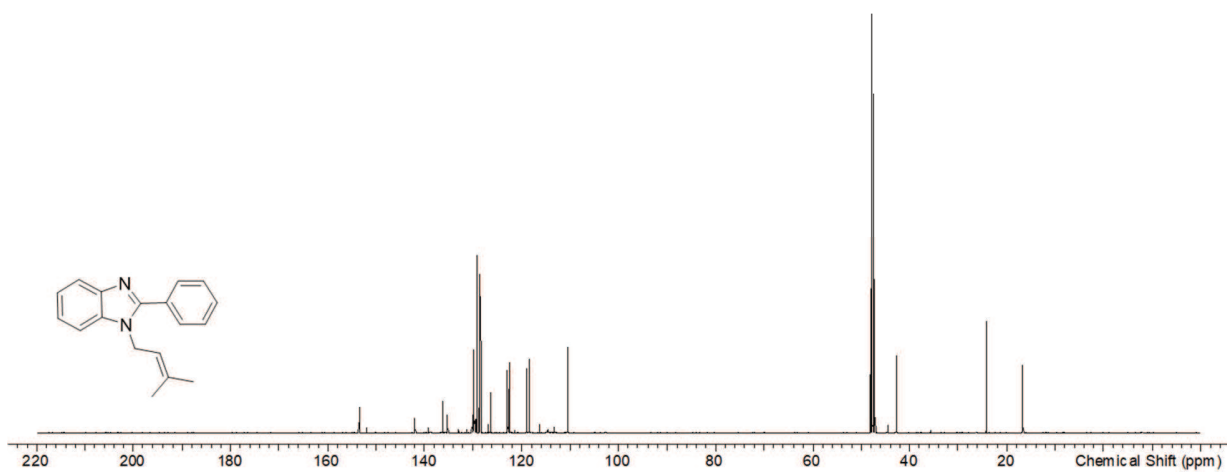
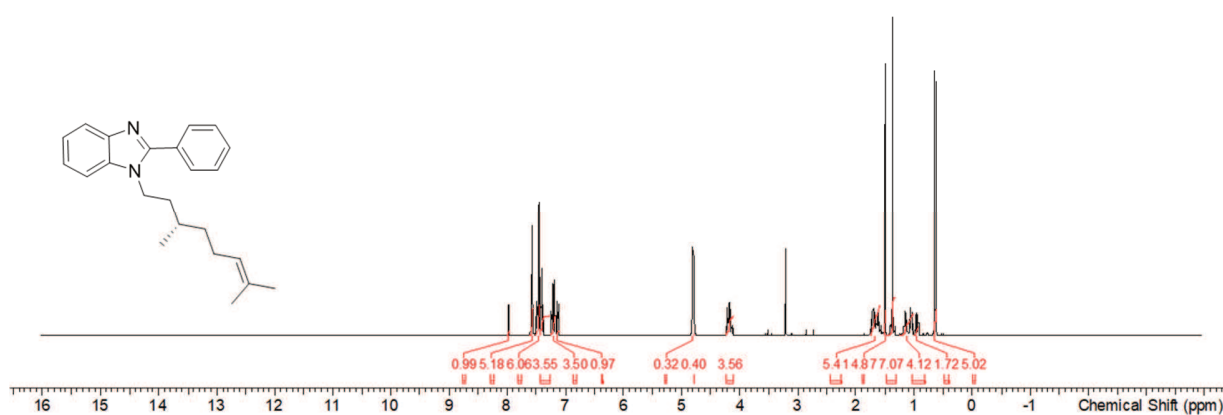
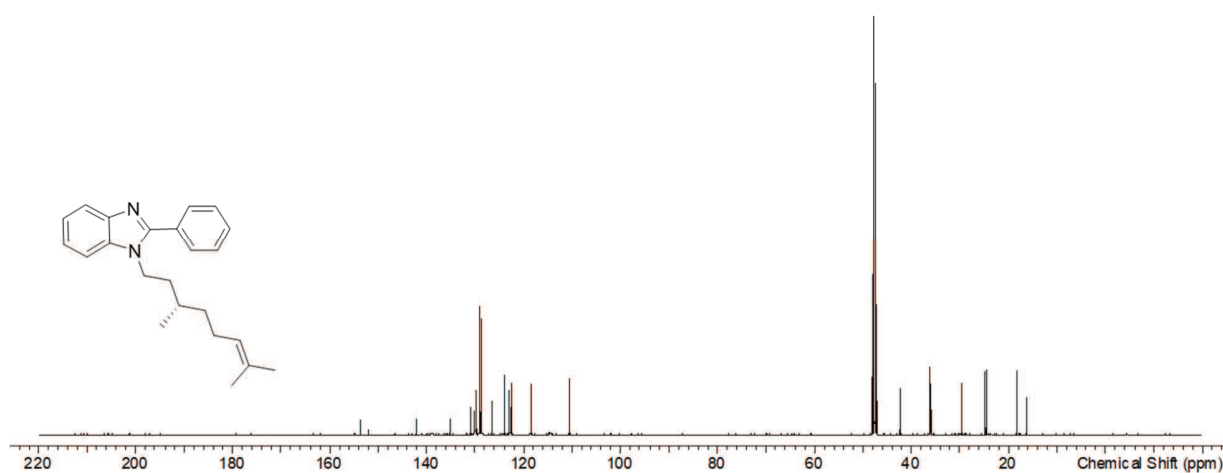


Figure S56. <sup>13</sup>C spectrum of **27** in methanol-*d*<sub>4</sub> at 126 MHz.



**Figure S57.**  $^1\text{H}$  spectrum of **28** in methanol- $d_4$  at 500 MHz.



**Figure S58.**  $^{13}\text{C}$  spectrum of **28** in methanol- $d_4$  at 126 MHz.

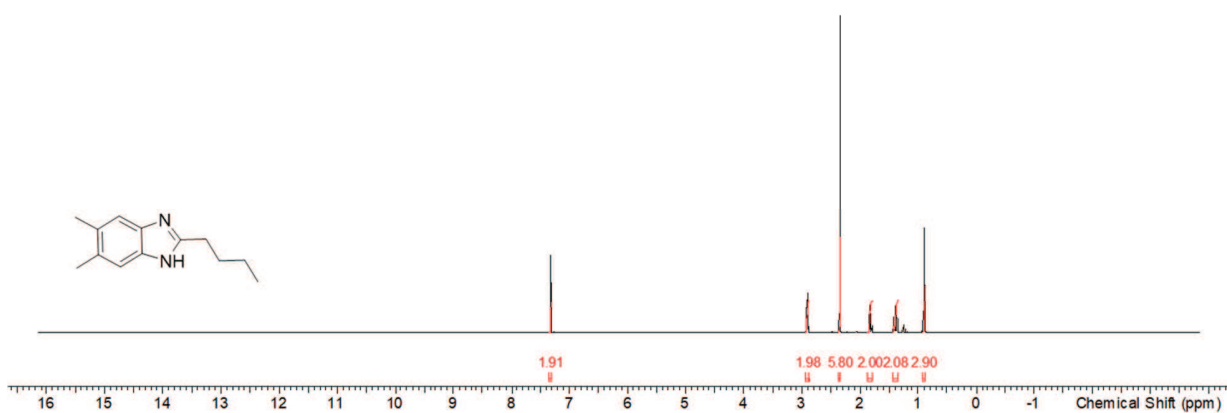


Figure S59.  $^1\text{H}$  spectrum of **29** in chloroform-*d* at 500 MHz.

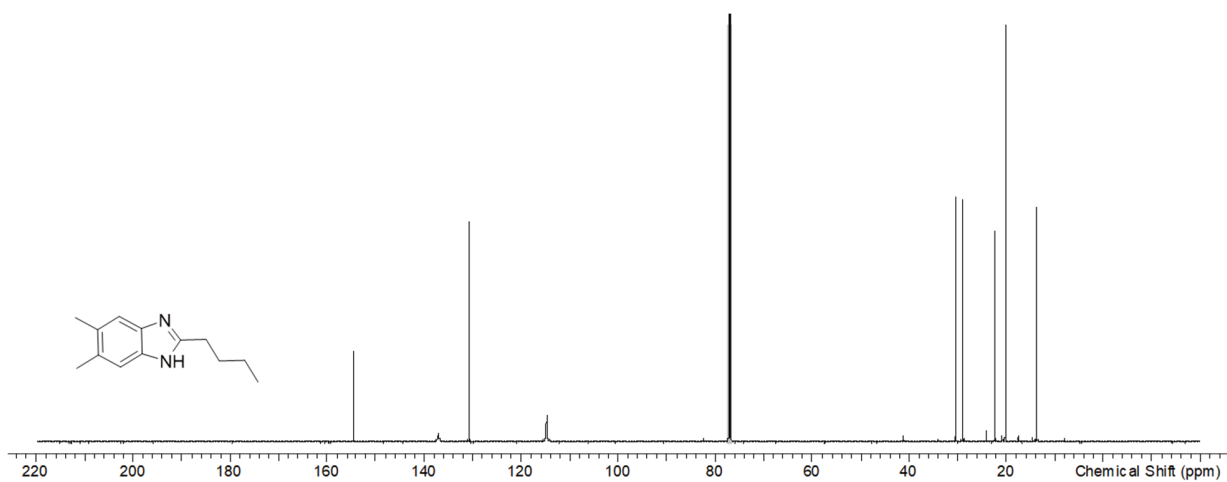


Figure S60.  $^{13}\text{C}$  spectrum of **29** in chloroform-*d* at 126 MHz.

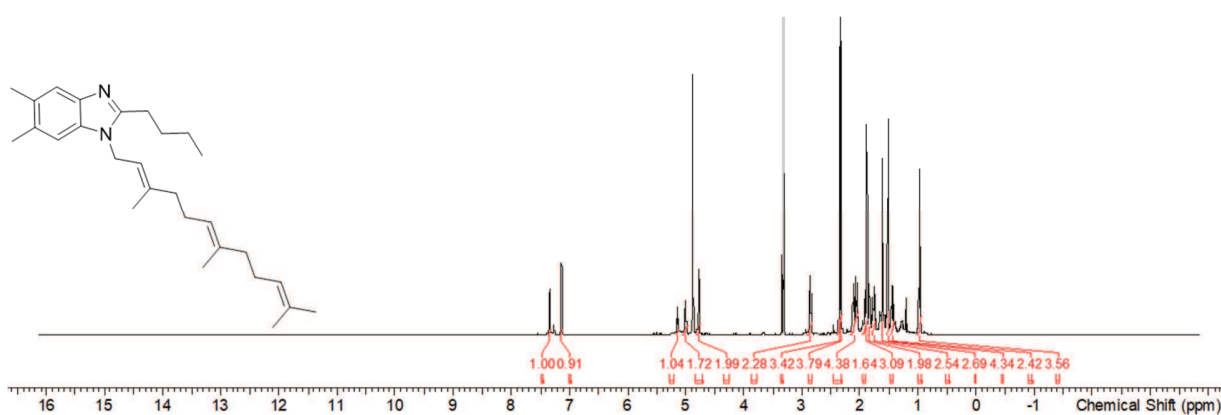


Figure S61. <sup>1</sup>H spectrum of **30** in methanol-*d*<sub>4</sub> at 500 MHz.

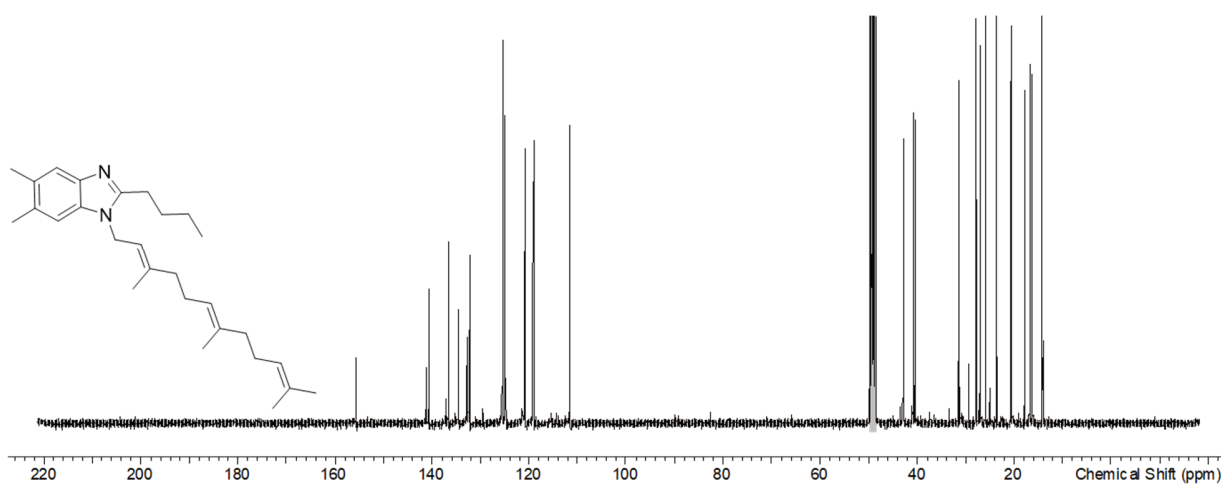
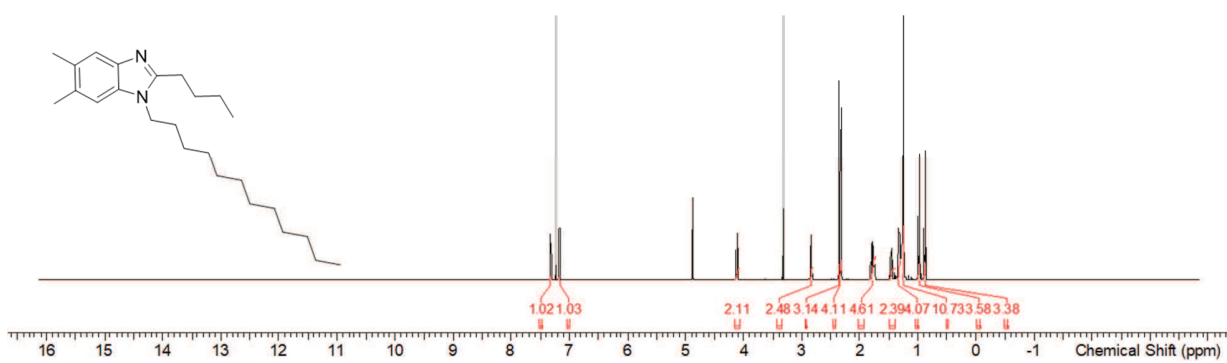
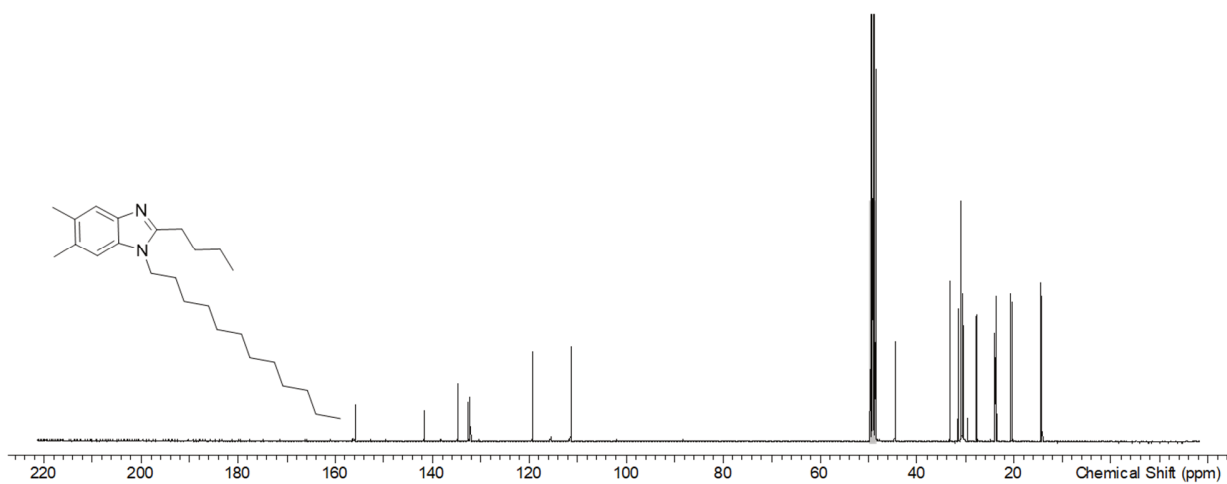


Figure S62. <sup>13</sup>C spectrum of **30** in methanol-*d*<sub>4</sub> at 126 MHz.

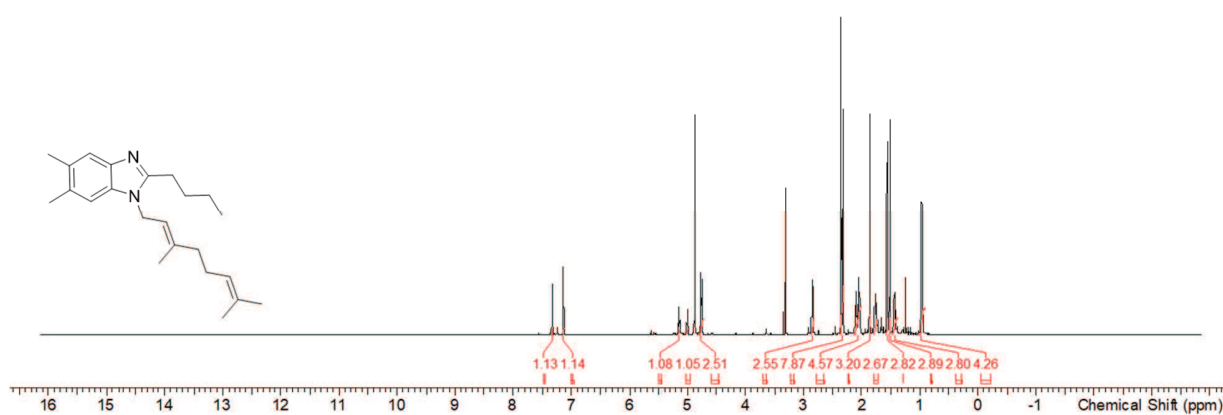


**Figure S63.**  $^1\text{H}$  spectrum of **31** in methanol- $d_4$  at 500 MHz.

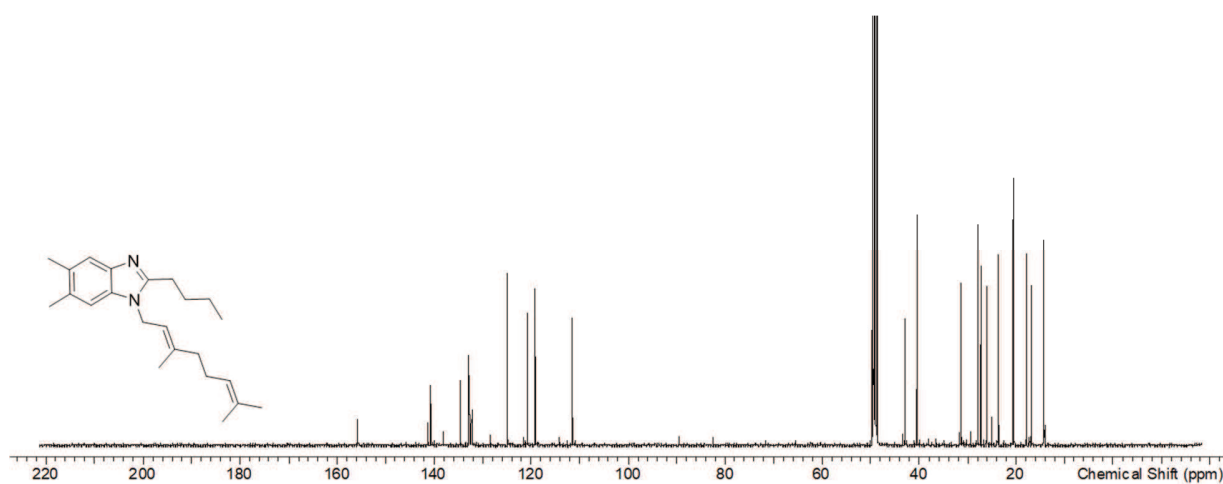


**Figure S64.**  $^{13}\text{C}$  spectrum of **31** in methanol- $d_4$  at 126 MHz.

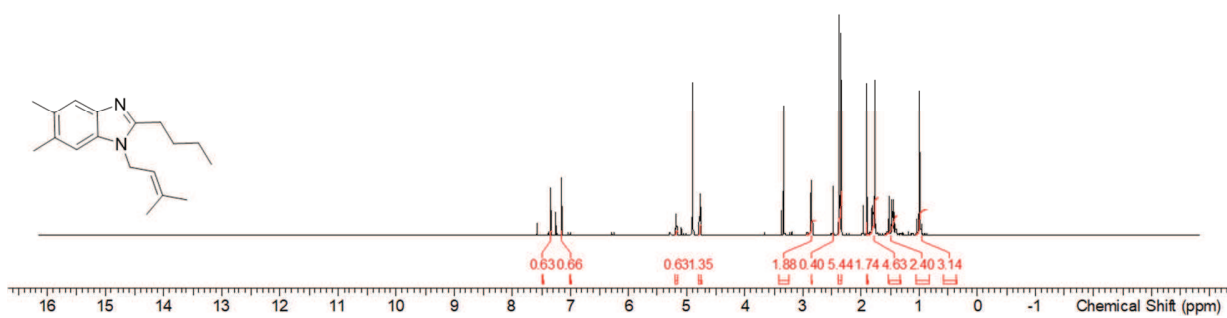




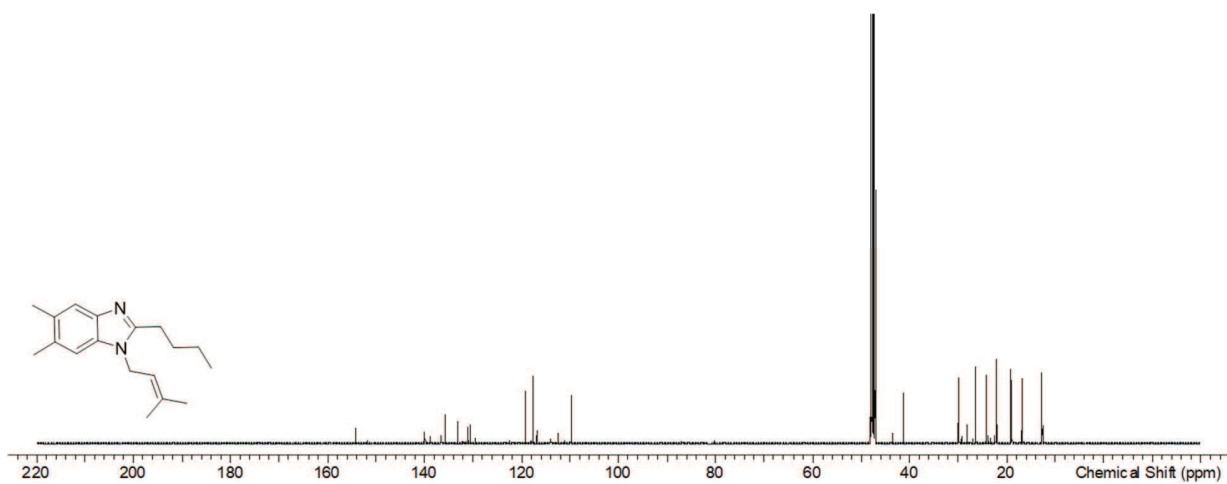
**Figure S65.**  $^1\text{H}$  spectrum of **32** in methanol- $d_4$  at 500 MHz.



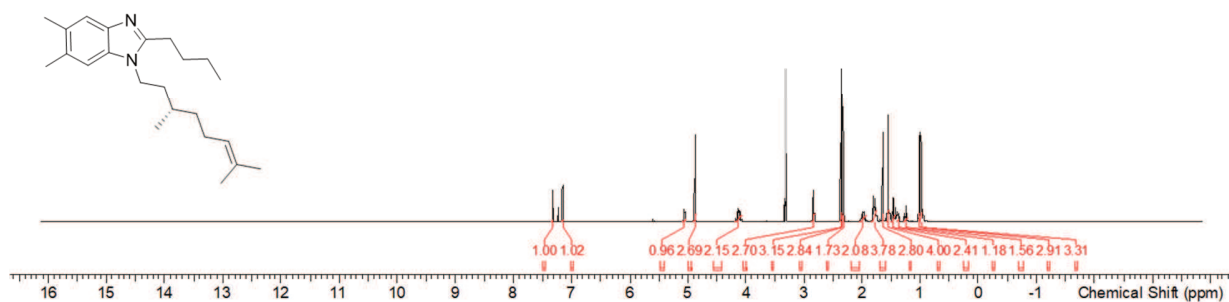
**Figure S66.**  $^{13}\text{C}$  spectrum of **32** in methanol- $d_4$  at 126 MHz.



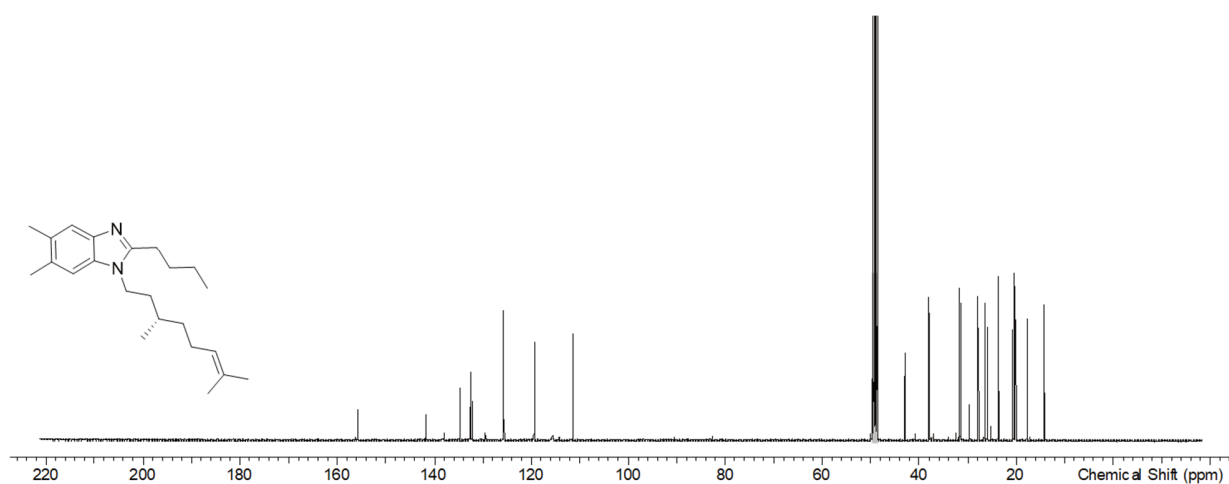
**Figure S67.** <sup>1</sup>H spectrum of **33** in methanol-*d*<sub>4</sub> at 500 MHz.



**Figure S68.** <sup>13</sup>C spectrum of **33** in methanol-*d*<sub>4</sub> at 126 MHz.



**Figure S69.**  $^1\text{H}$  spectrum of **34** in methanol- $d_4$  at 500 MHz.



**Figure S70.**  $^{13}\text{C}$  spectrum of **34** in methanol- $d_4$  at 126 MHz.

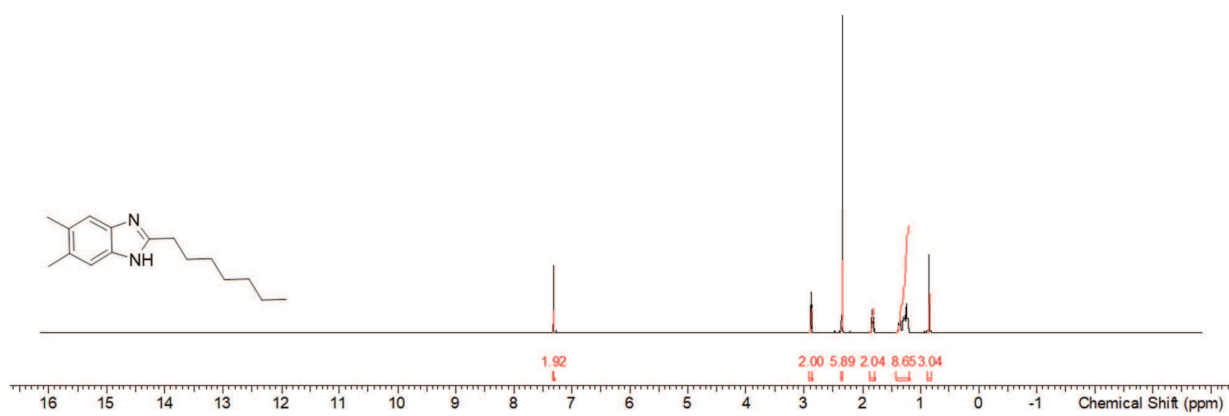


Figure S71. <sup>1</sup>H spectrum of **35** in chloroform-d at 500 MHz.

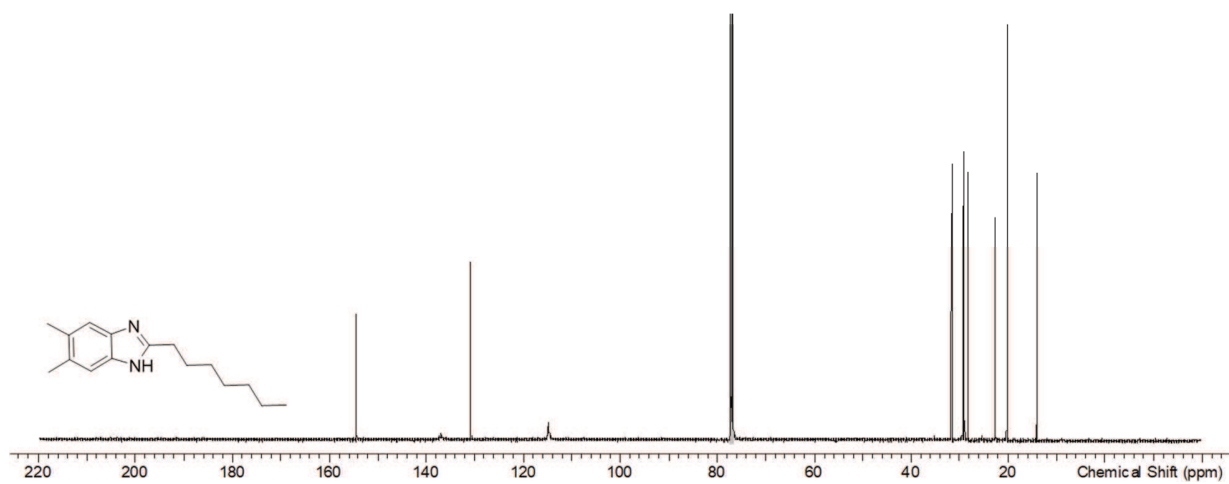
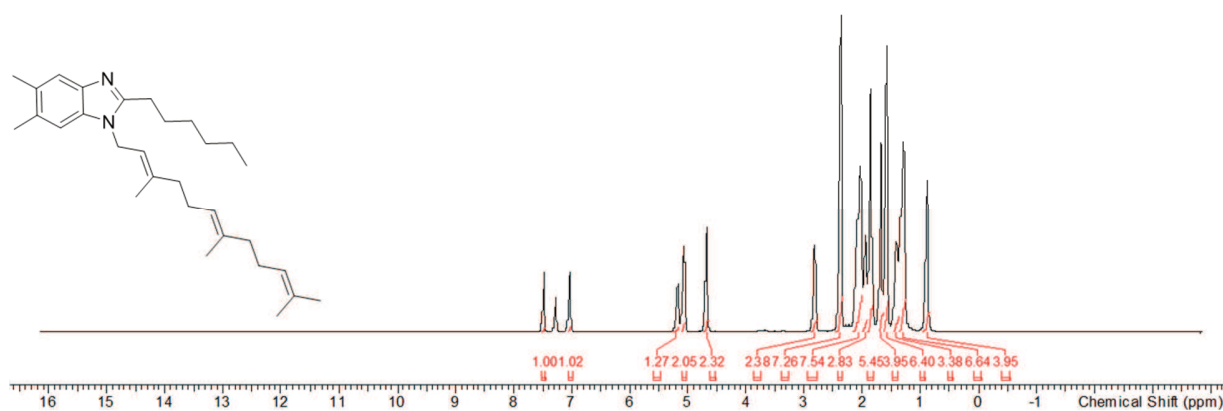
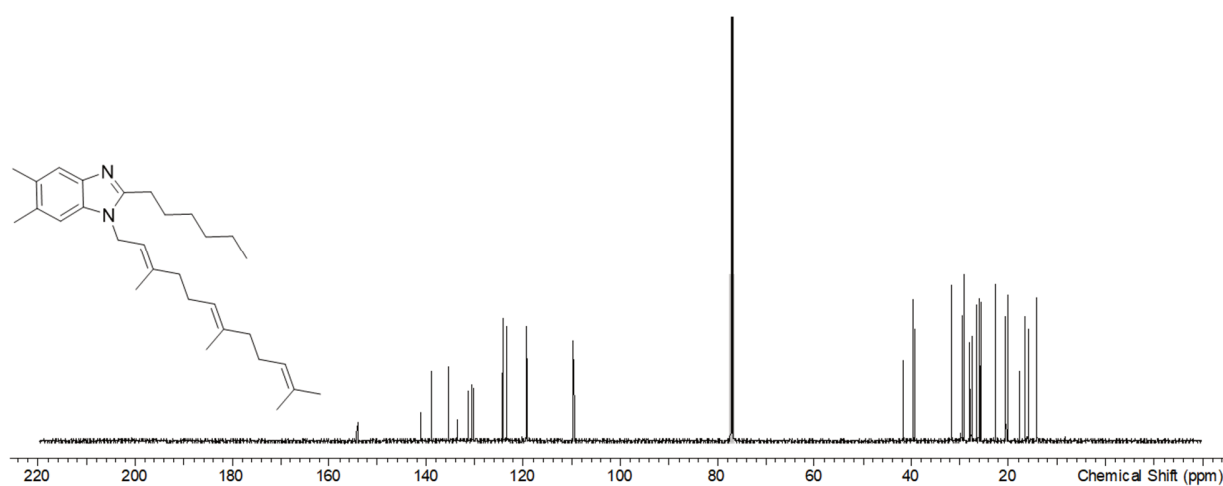


Figure S72. <sup>13</sup>C spectrum of **35** in chloroform-d at 126 MHz.



**Figure S73.**  $^1\text{H}$  spectrum of **36** in chloroform-d at 500 MHz.



**Figure S74.**  $^{13}\text{C}$  spectrum of **36** in chloroform-d at 126 MHz.

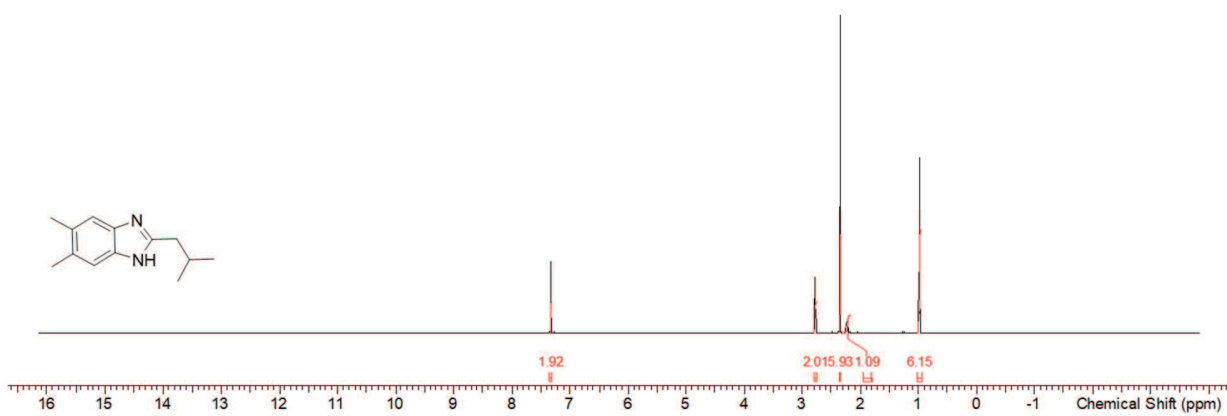


Figure S75. <sup>1</sup>H spectrum of **37** in chloroform-d at 500 MHz.

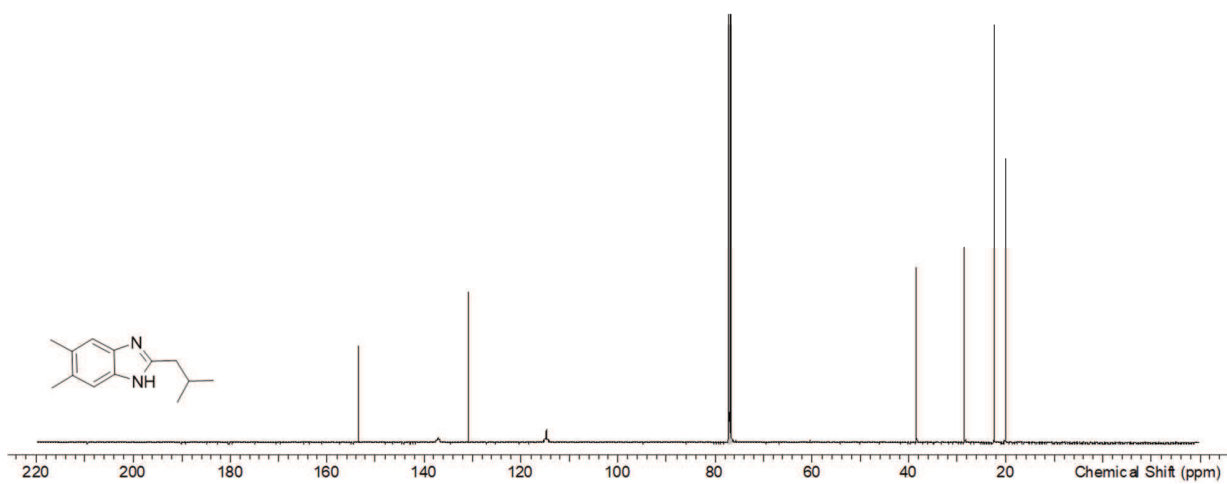
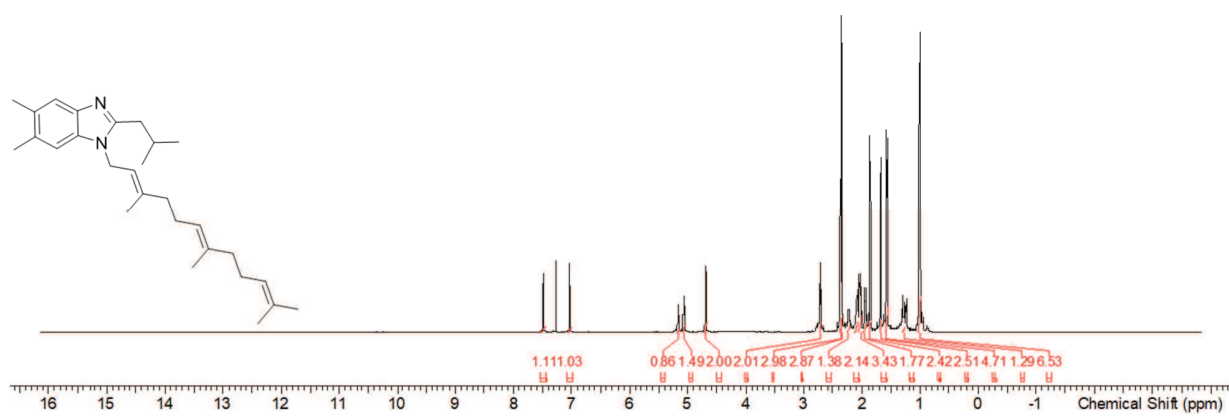
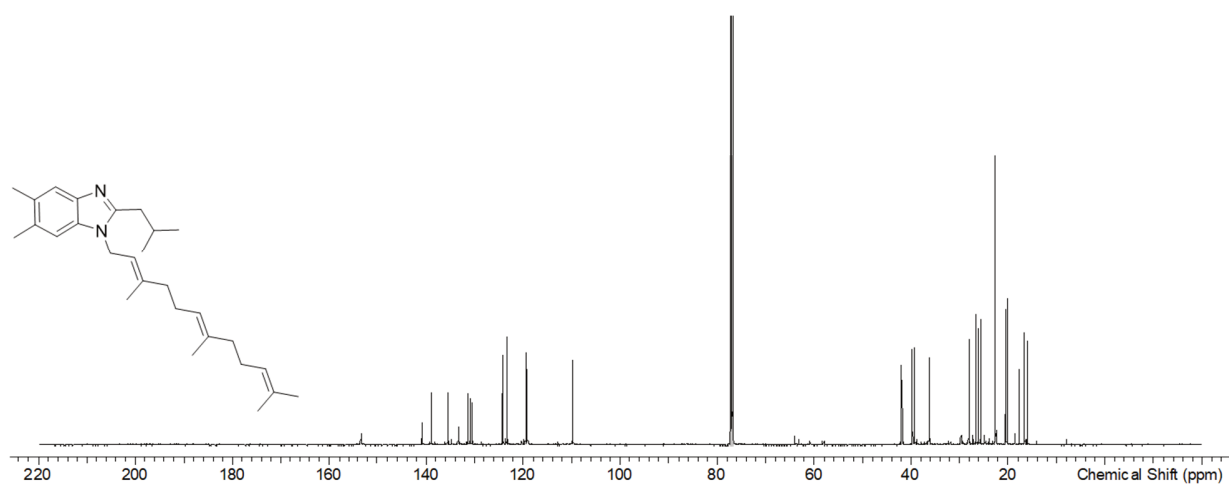


Figure S76. <sup>13</sup>C spectrum of **37** in chloroform-d at 126 MHz.



**Figure S77.**  $^1\text{H}$  spectrum of **38** in chloroform-*d* at 500 MHz.



**Figure S78.**  $^{13}\text{C}$  spectrum of **38** in chloroform-*d* at 126 MHz.

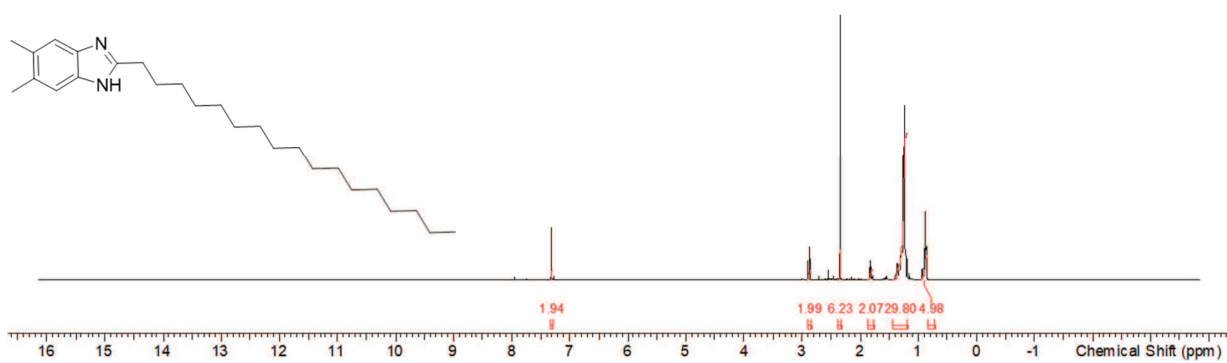


Figure S79.  $^1\text{H}$  spectrum of **39** in chloroform-*d* at 500 MHz.

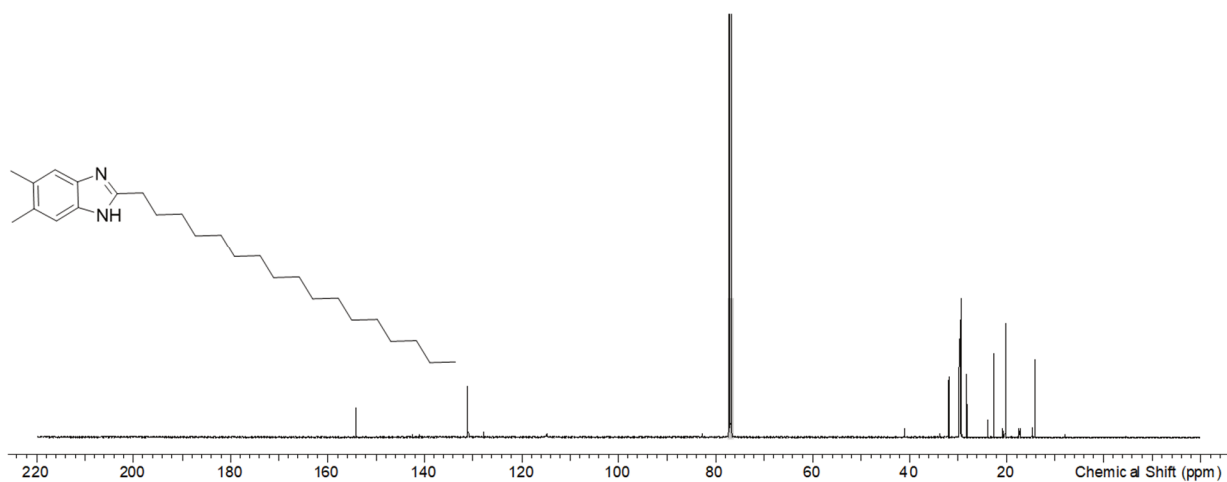
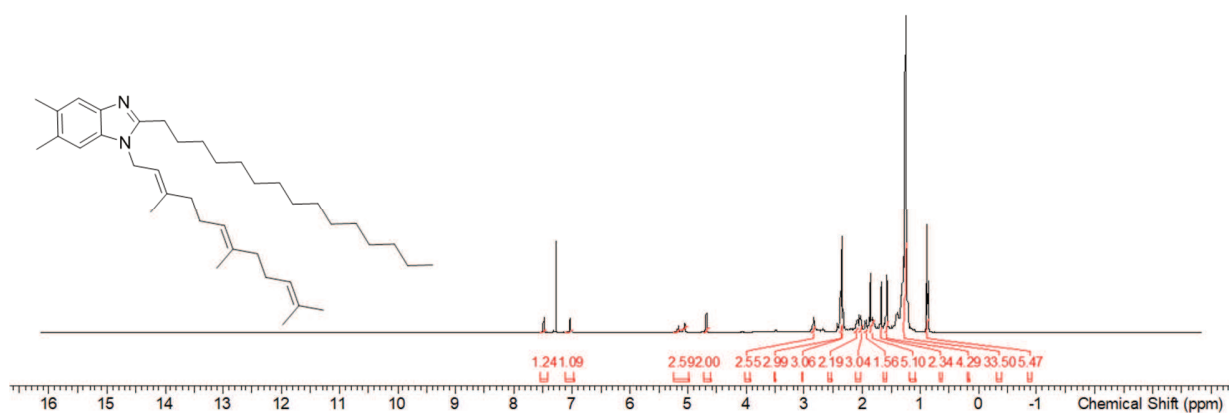
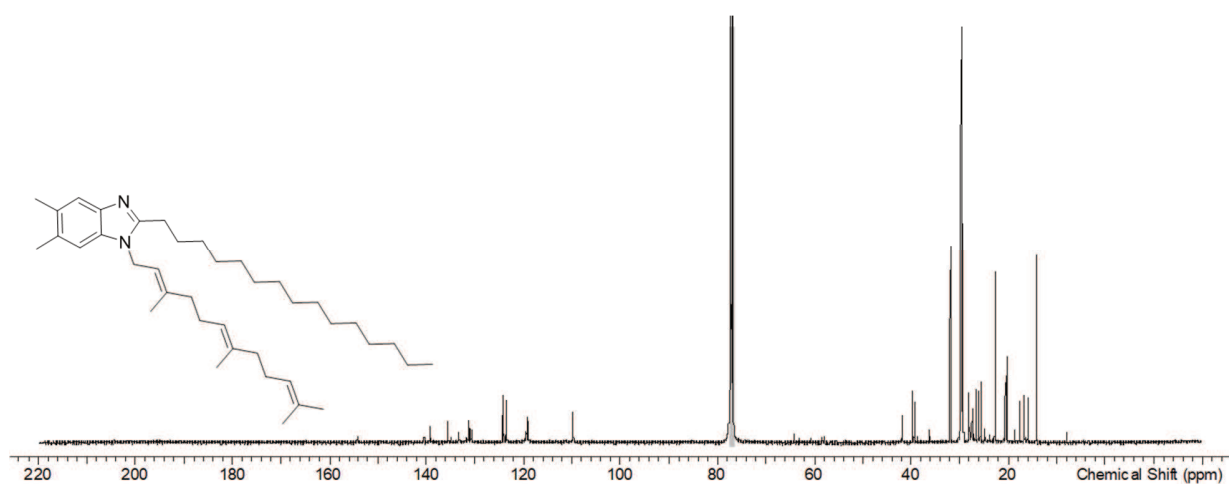


Figure S80.  $^{13}\text{C}$  spectrum of **39** in chloroform-*d* at 126 MHz.





**Figure S81.**  $^1\text{H}$  spectrum of **40** in chloroform-*d* at 500 MHz.



**Figure S82.**  $^{13}\text{C}$  spectrum of **40** in chloroform-*d* at 126 MHz.

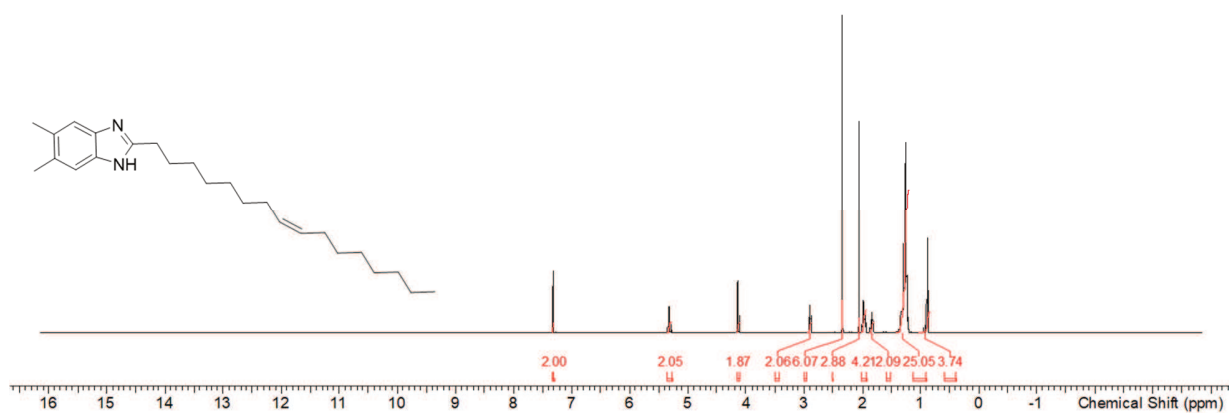


Figure S83.  $^1\text{H}$  spectrum of **41** in chloroform-*d* at 500 MHz.

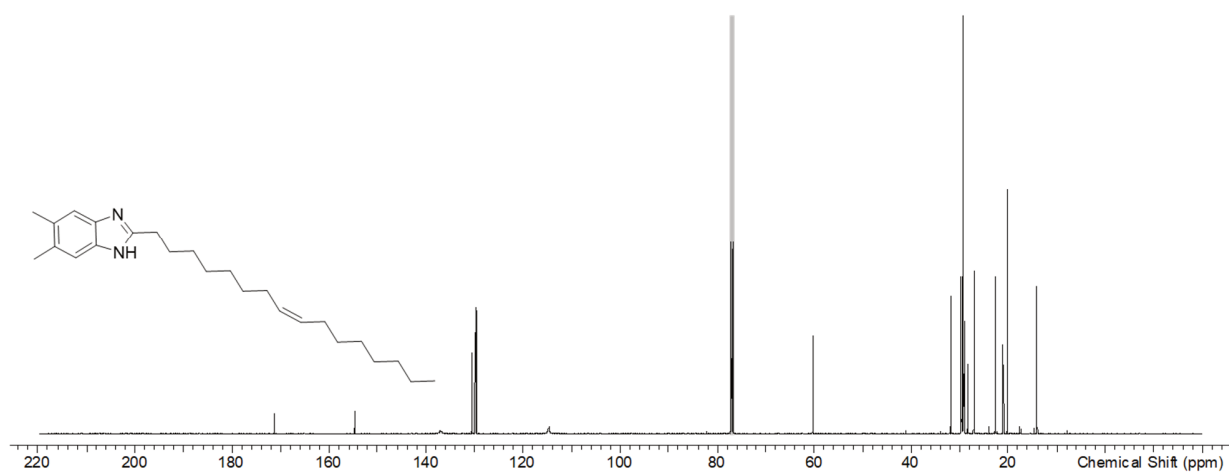
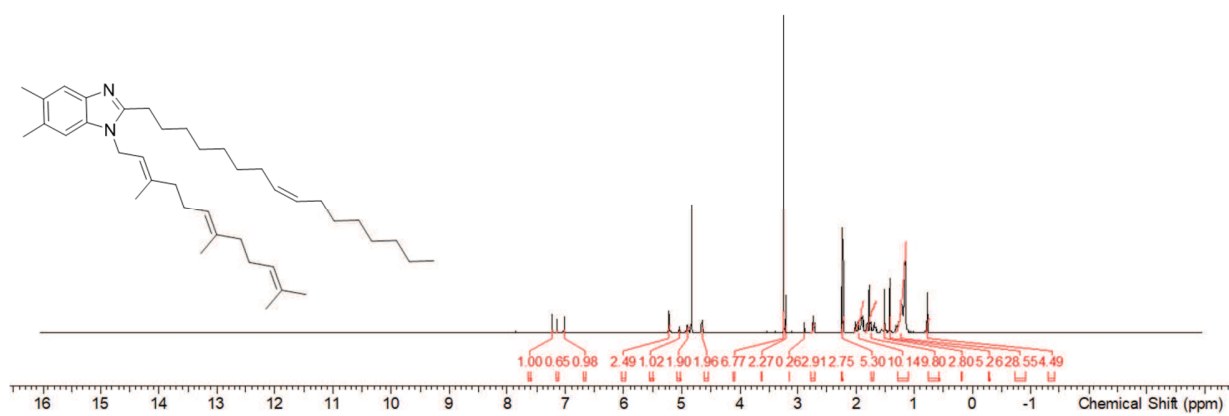
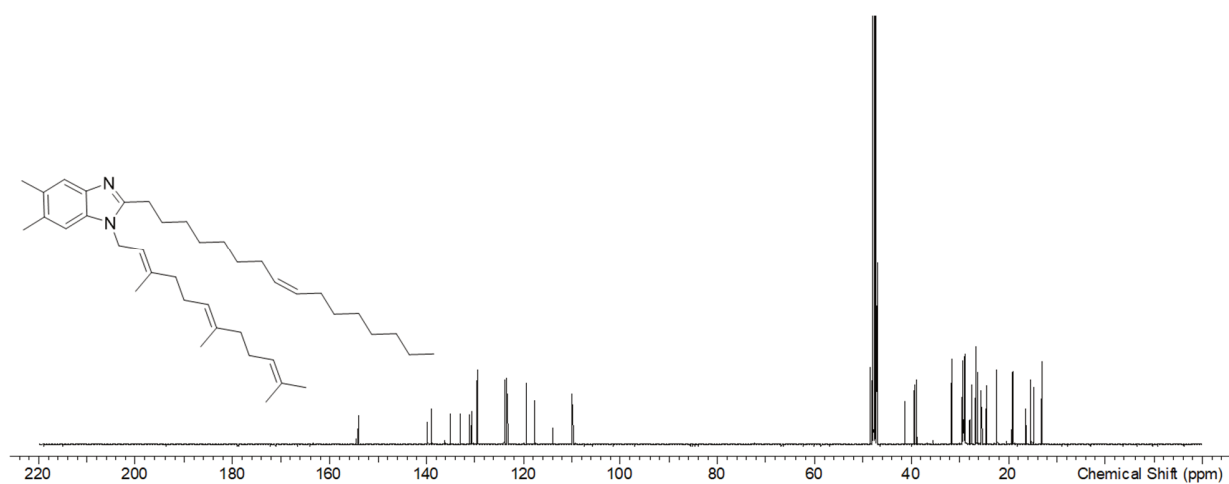


Figure S84.  $^{13}\text{C}$  spectrum of **41** in chloroform-*d* at 126 MHz.



**Figure S85.**  $^1\text{H}$  spectrum of **42** in methanol- $d_4$  at 500 MHz.



**Figure S86.**  $^{13}\text{C}$  spectrum of **42** in methanol- $d_4$  at 126 MHz.

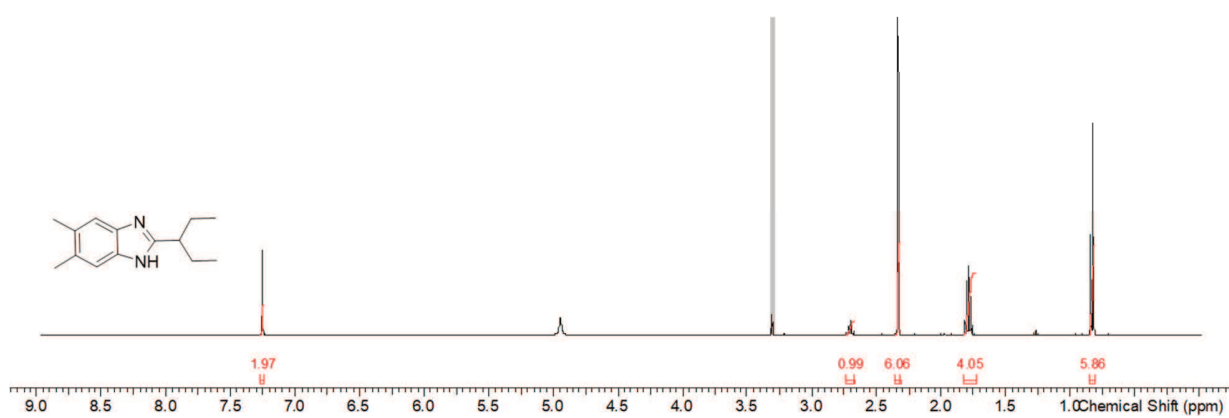


Figure S87.  $^1\text{H}$  spectrum of **43** in methanol- $d_4$  at 500 MHz.

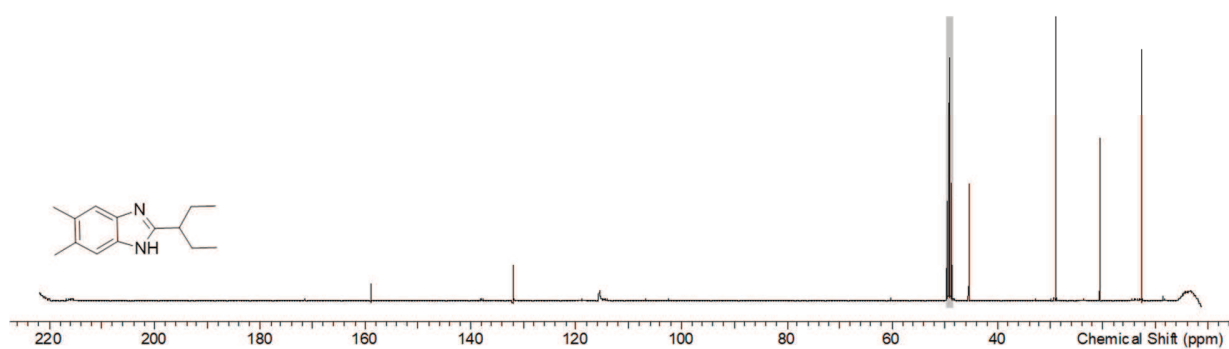


Figure S88.  $^{13}\text{C}$  spectrum of **43** in methanol- $d_4$  at 126 MHz.

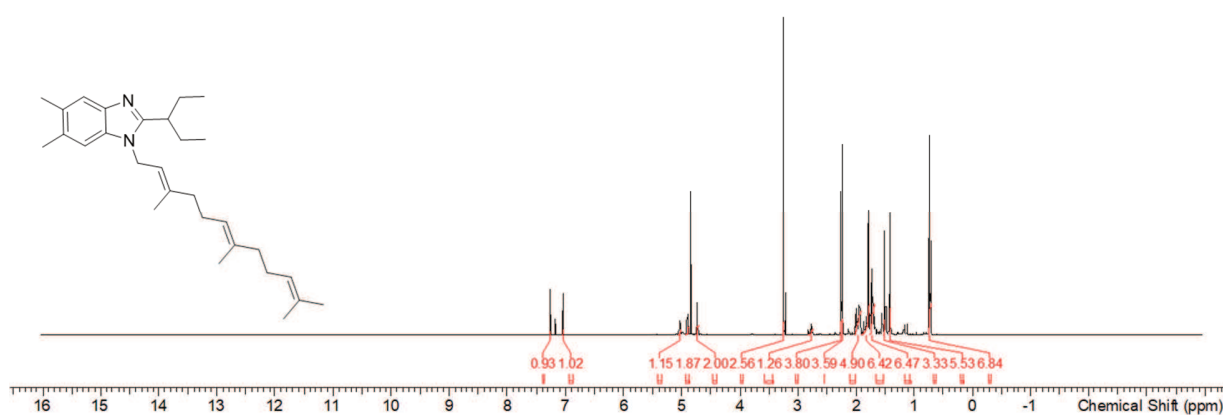


Figure S89.  $^1\text{H}$  spectrum of **44** in methanol- $d_4$  at 500 MHz.

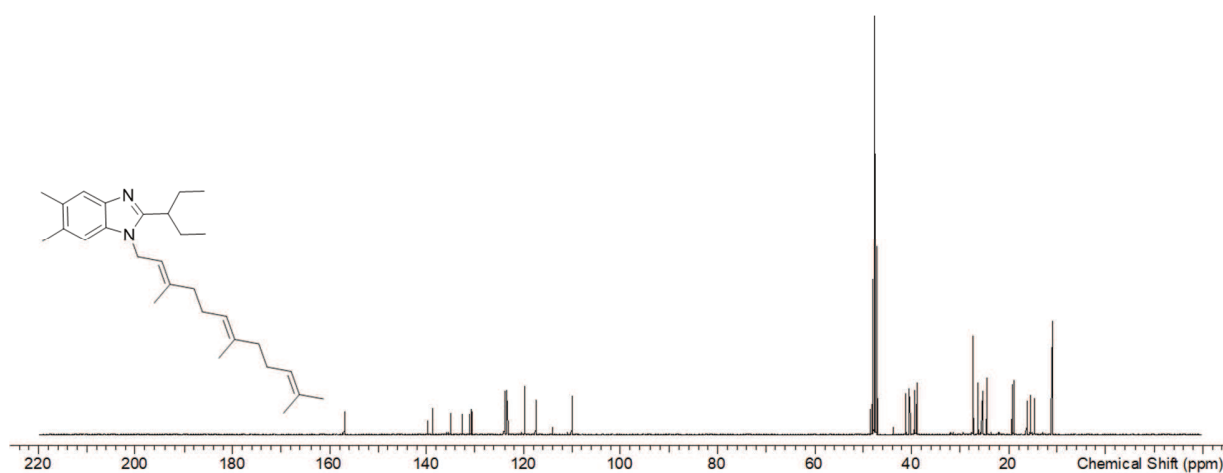


Figure S90.  $^{13}\text{C}$  spectrum of **44** in methanol- $d_4$  at 126 MHz.

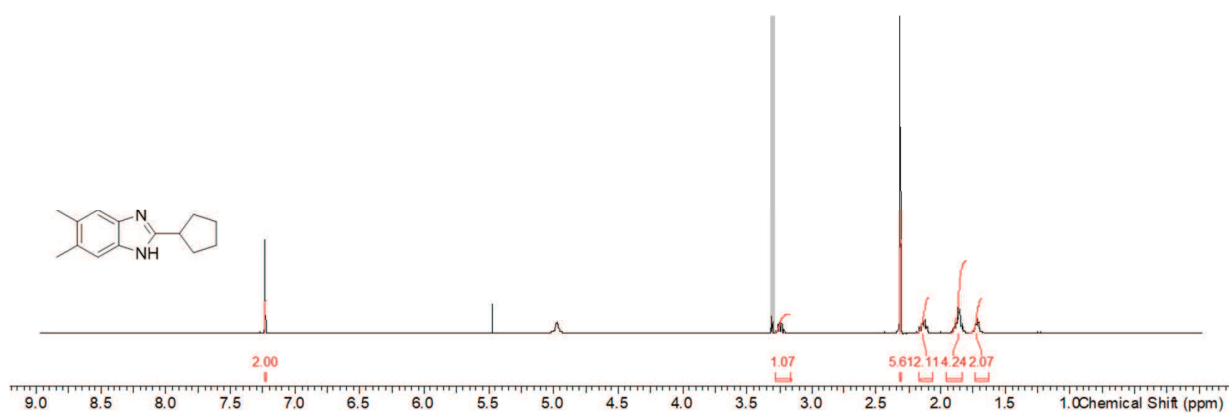


Figure S91.  $^1\text{H}$  spectrum of **45** in  $\text{methanol-}d_4$  at 500 MHz.

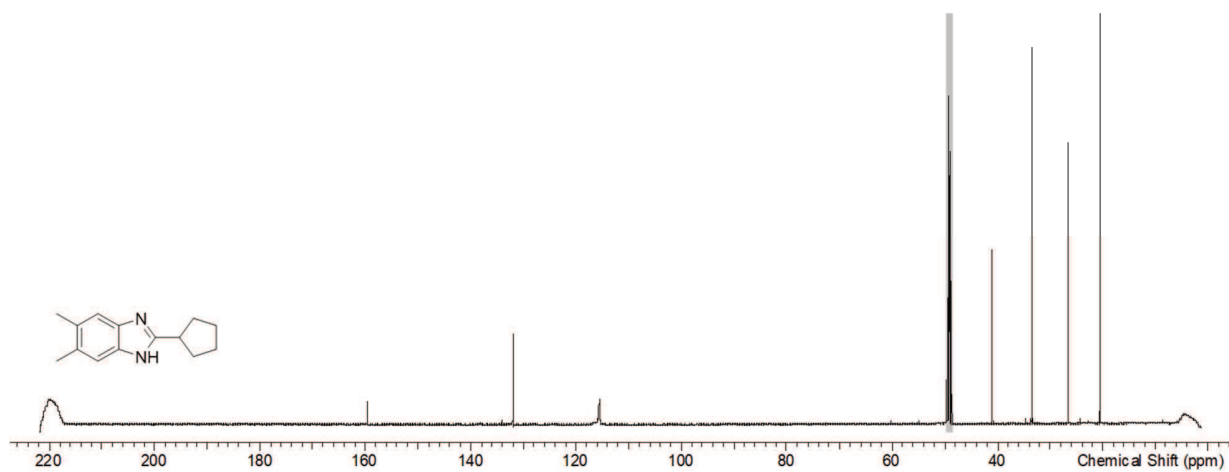
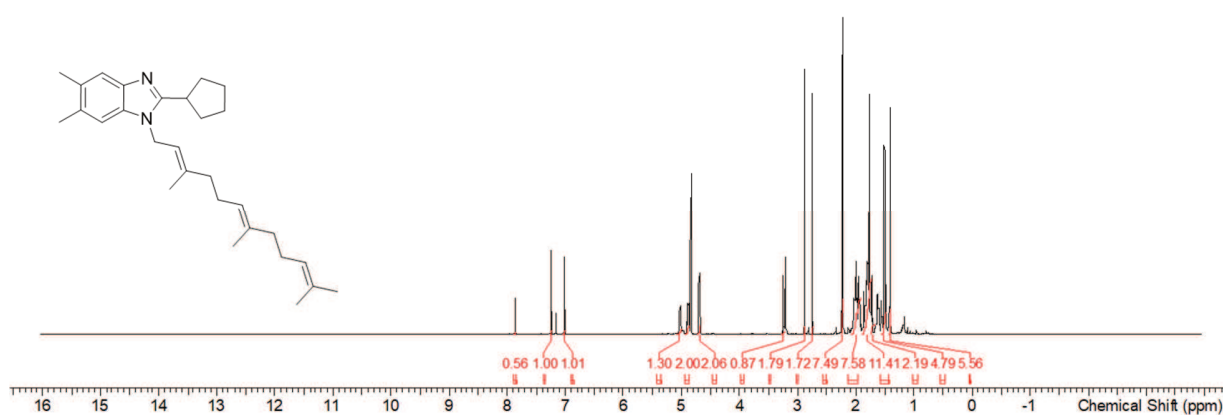
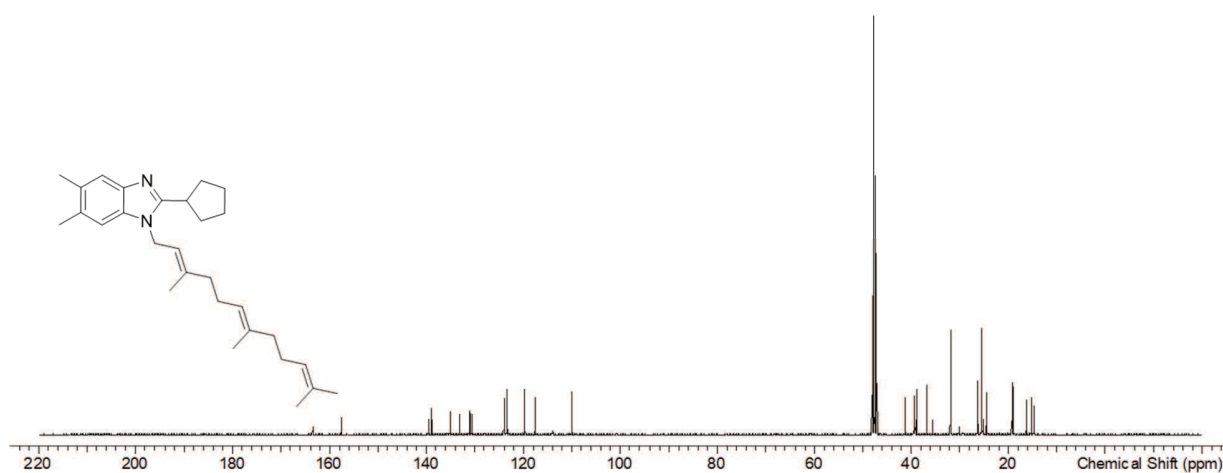


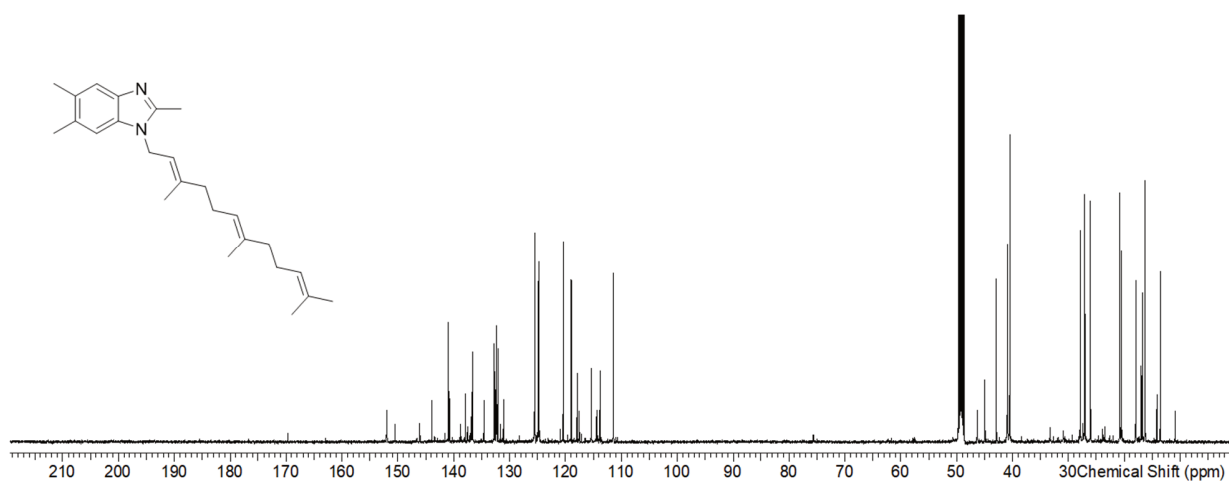
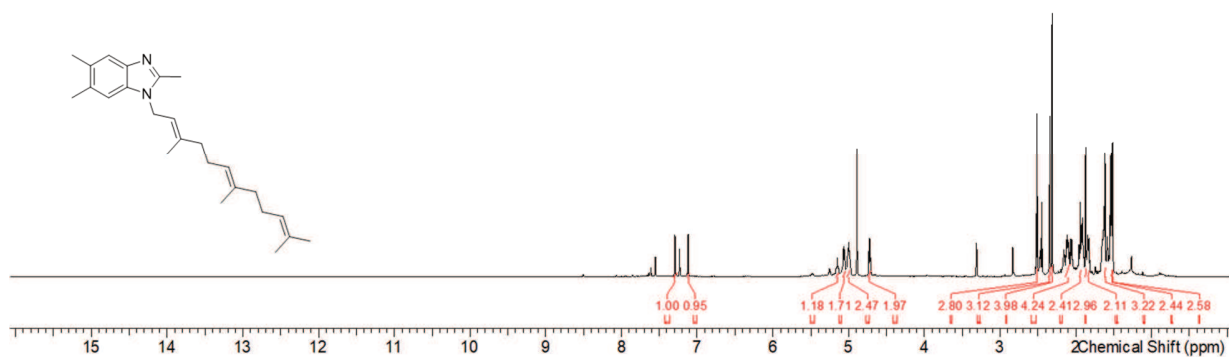
Figure S92.  $^{13}\text{C}$  spectrum of **45** in  $\text{methanol-}d_4$  at 126 MHz.



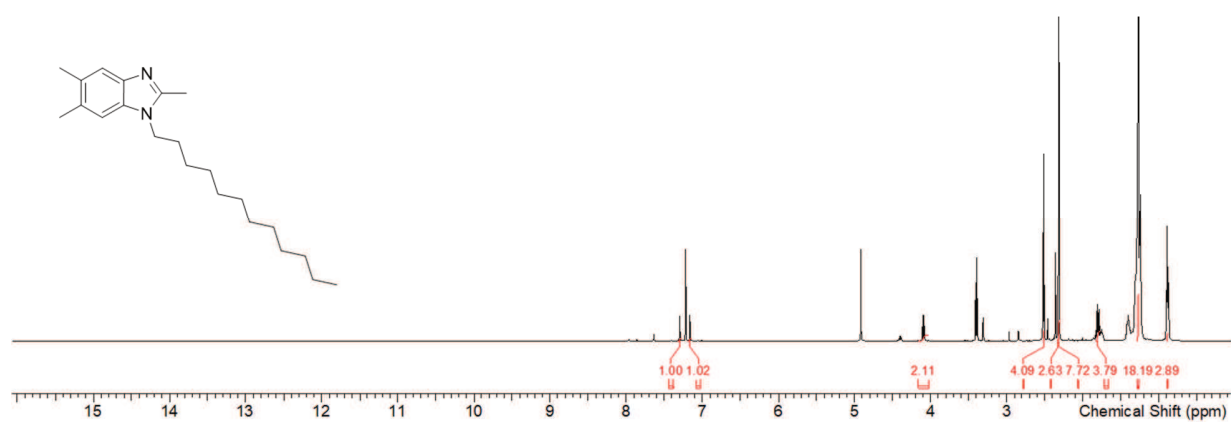
**Figure S93.**  $^1\text{H}$  spectrum of **46** in methanol- $d_4$  at 500 MHz.



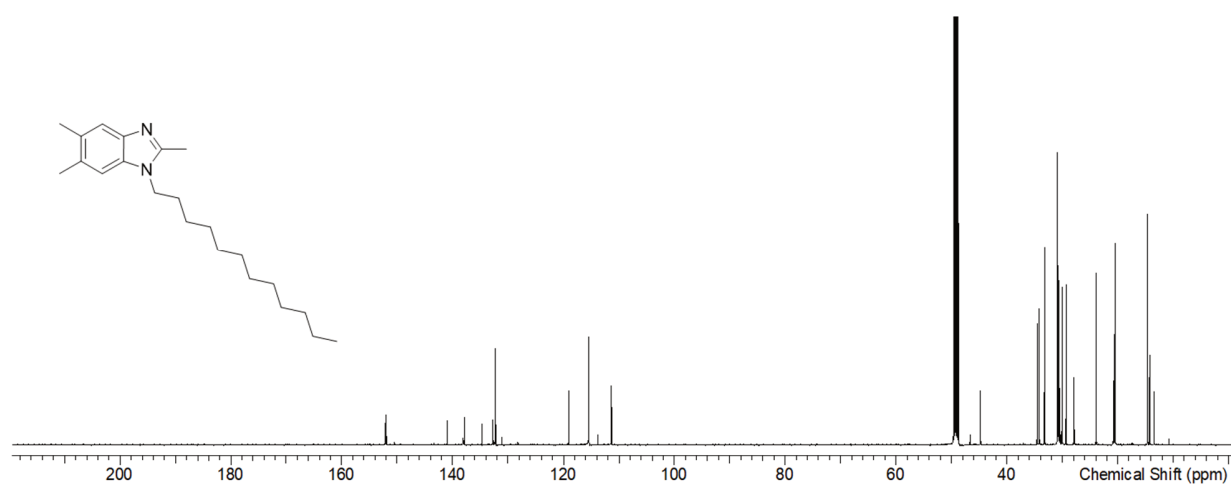
**Figure S94.**  $^{13}\text{C}$  spectrum of **46** in methanol- $d_4$  at 126 MHz.







**Figure S97.**  $^1\text{H}$  spectrum of **50** in methanol- $d_4$  at 500 MHz.



**Figure S98.**  $^{13}\text{C}$  spectrum of **50** in methanol- $d_4$  at 500 MHz.

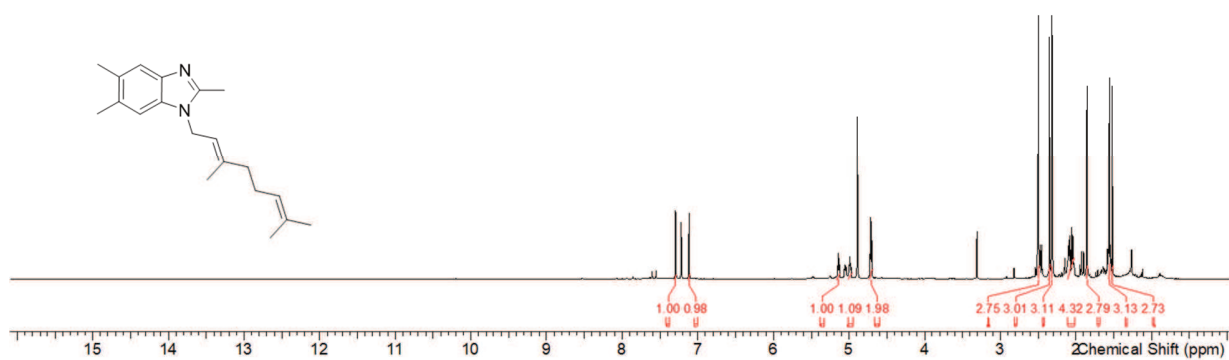


Figure S99.  $^1\text{H}$  spectrum of **51** in methanol- $d_4$  at 500 MHz.

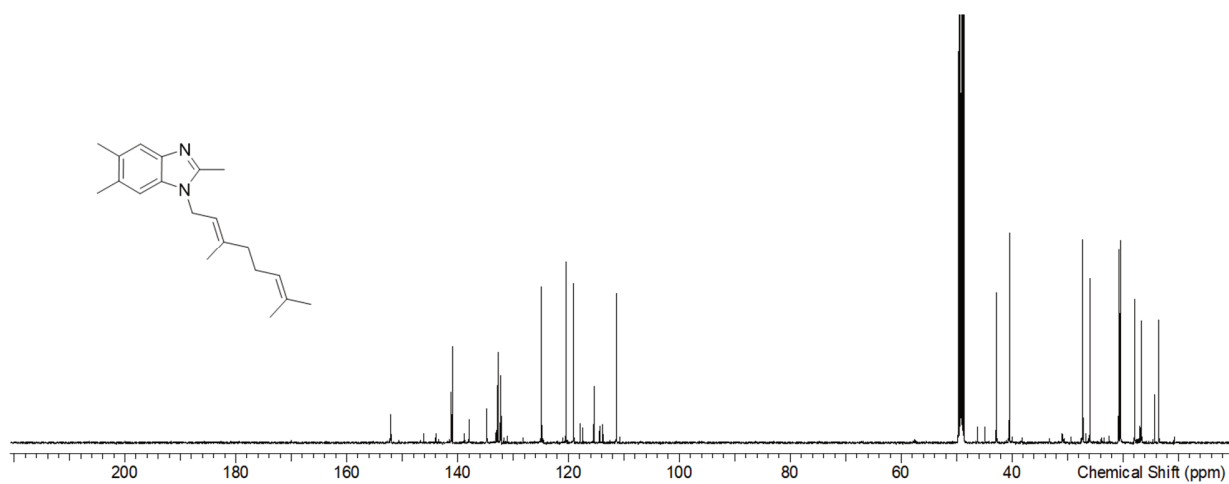


Figure S100.  $^{13}\text{C}$  spectrum of **51** in methanol- $d_4$  at 500 MHz.

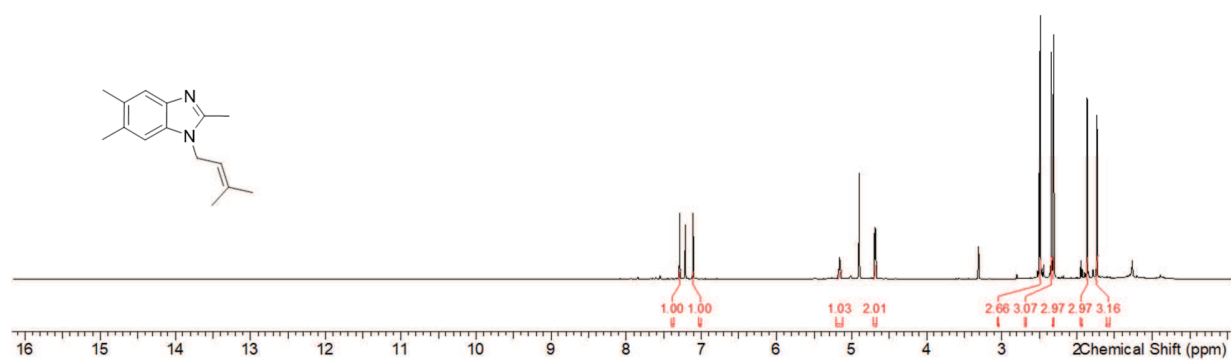


Figure S101.  $^1\text{H}$  spectrum of **52** in  $\text{methanol-d}_4$  at 500 MHz.

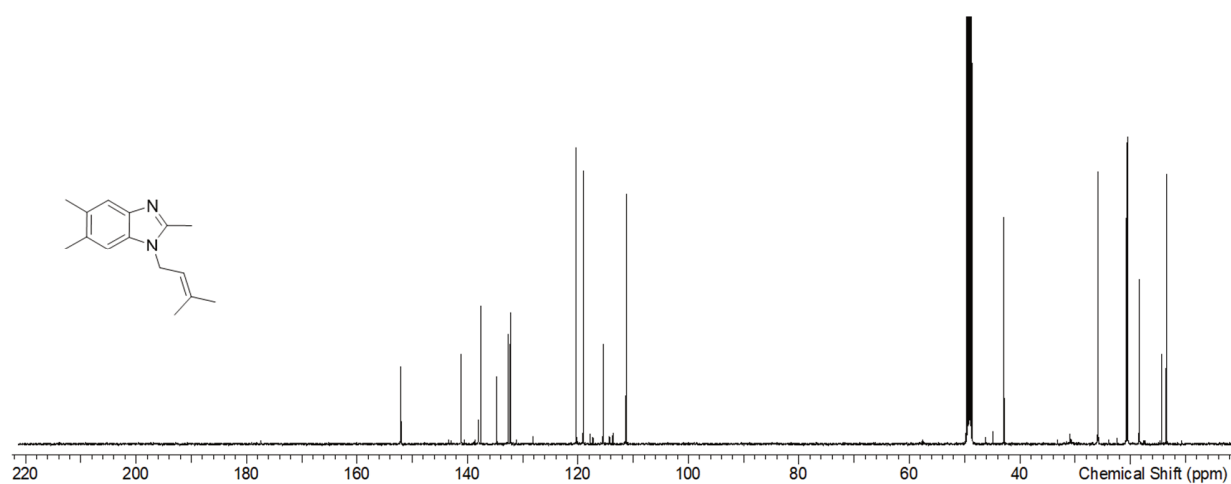
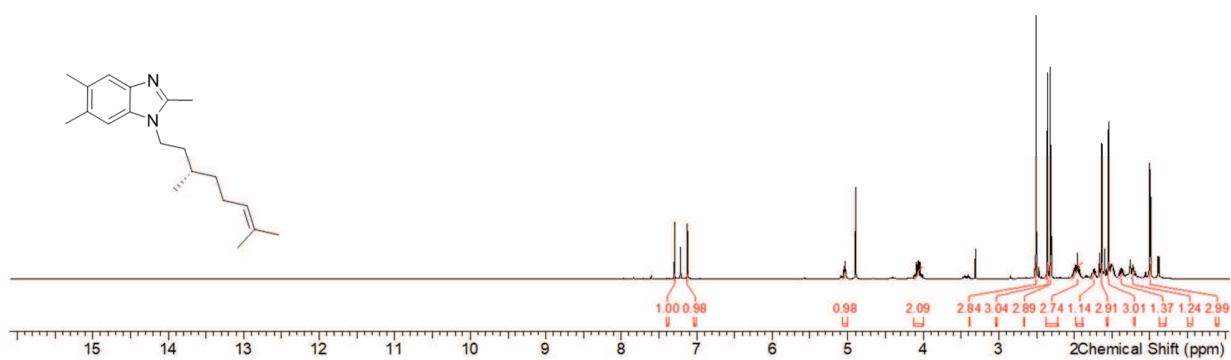
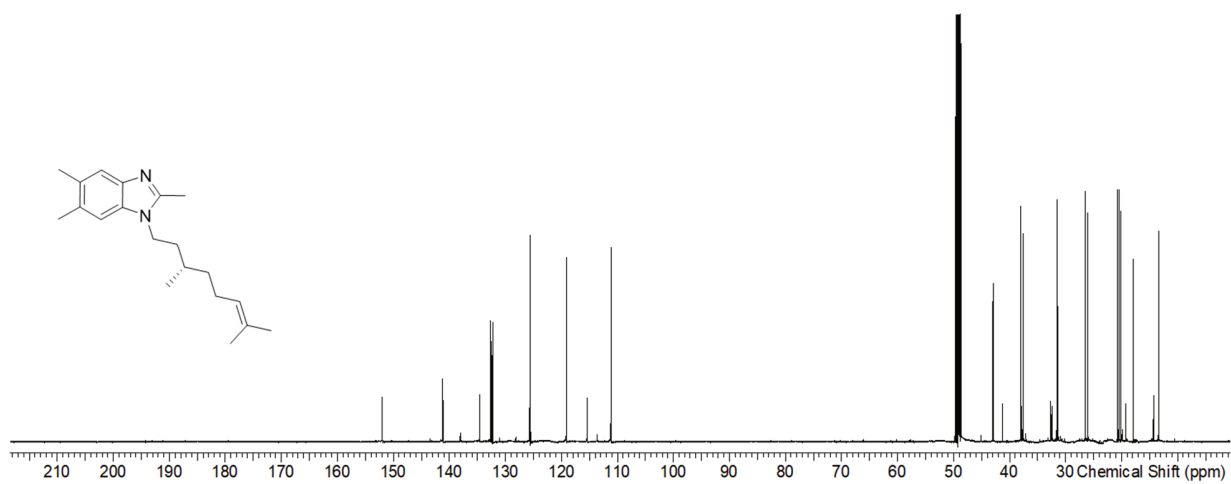


Figure S102.  $^{13}\text{C}$  spectrum of **52** in  $\text{methanol-d}_4$  at 500 MHz.



**Figure S103.**  $^1\text{H}$  spectrum of **53** in methanol- $d_4$  at 500 MHz.



**Figure S104.**  $^{13}\text{C}$  spectrum of **53** in methanol- $d_4$  at 500 MHz.

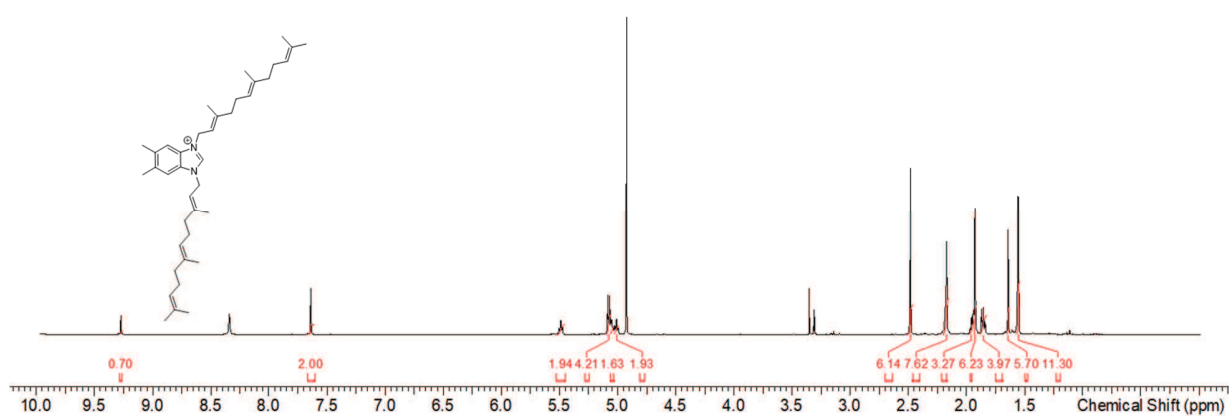


Figure S105.  $^1\text{H}$  spectrum of **55** in  $\text{methanol-d}_4$  at 500 MHz.

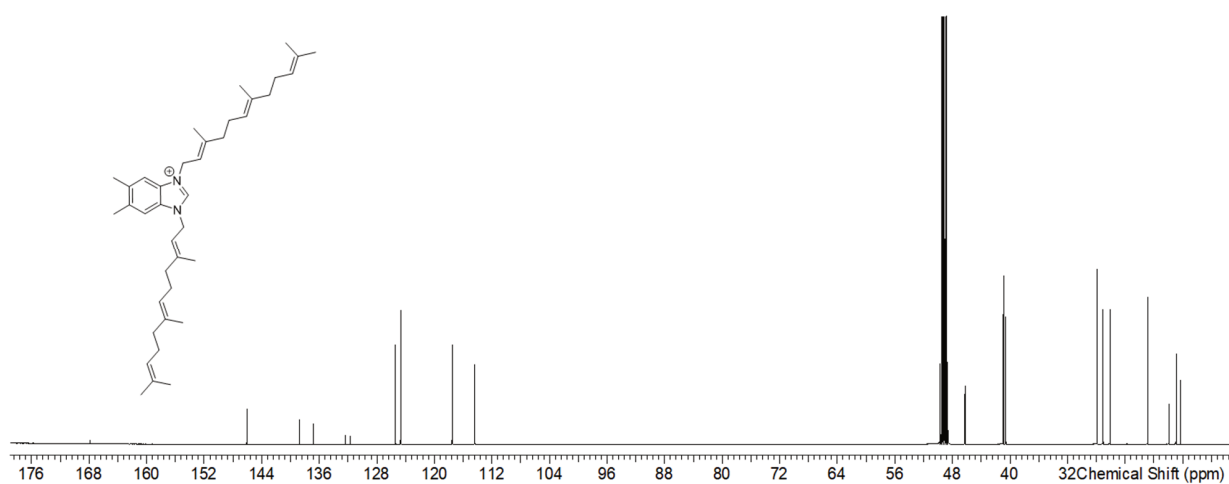
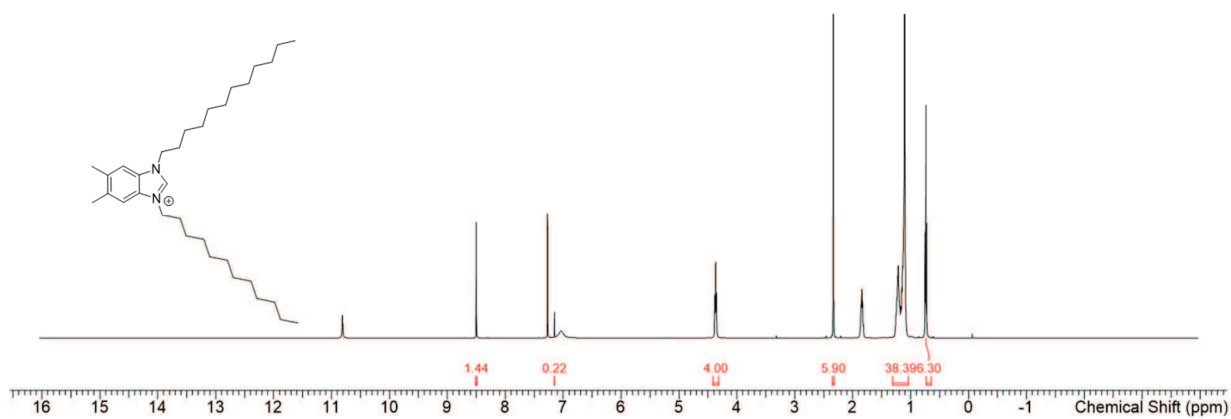
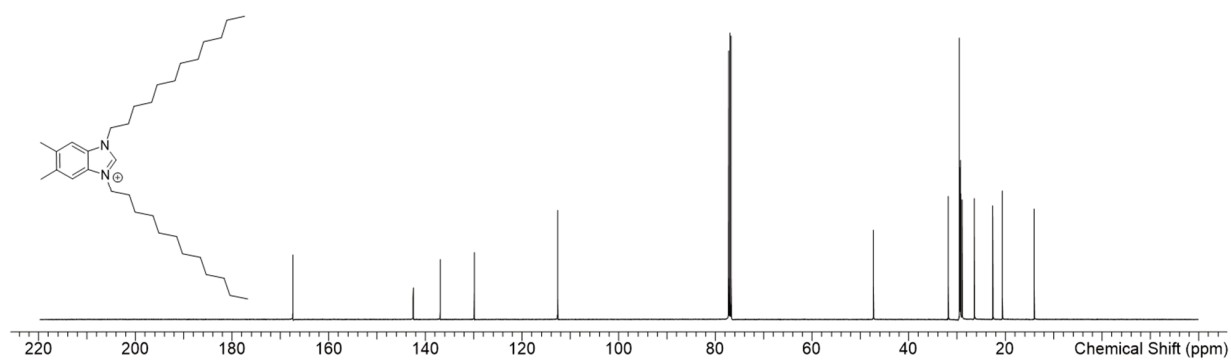


Figure S106.  $^{13}\text{C}$  spectrum of **55** in  $\text{methanol-d}_4$  at 126 MHz.



**Figure S107.**  $^1\text{H}$  spectrum of **56** in chloroform-*d* at 500 MHz.



**Figure S108.**  $^{13}\text{C}$  spectrum of **56** in chloroform-*d* at 126 MHz.

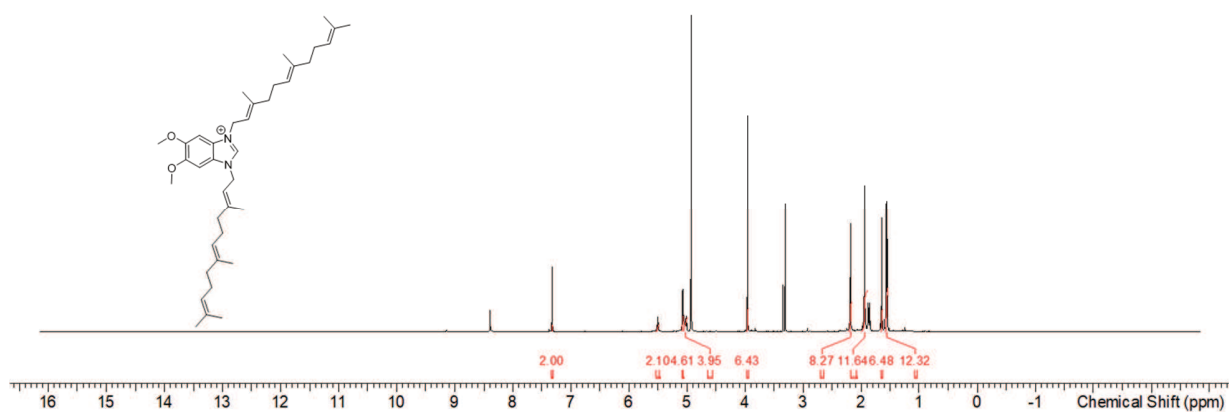


Figure S109. <sup>1</sup>H spectrum of **60** in methanol-d<sub>4</sub> at 500 MHz.

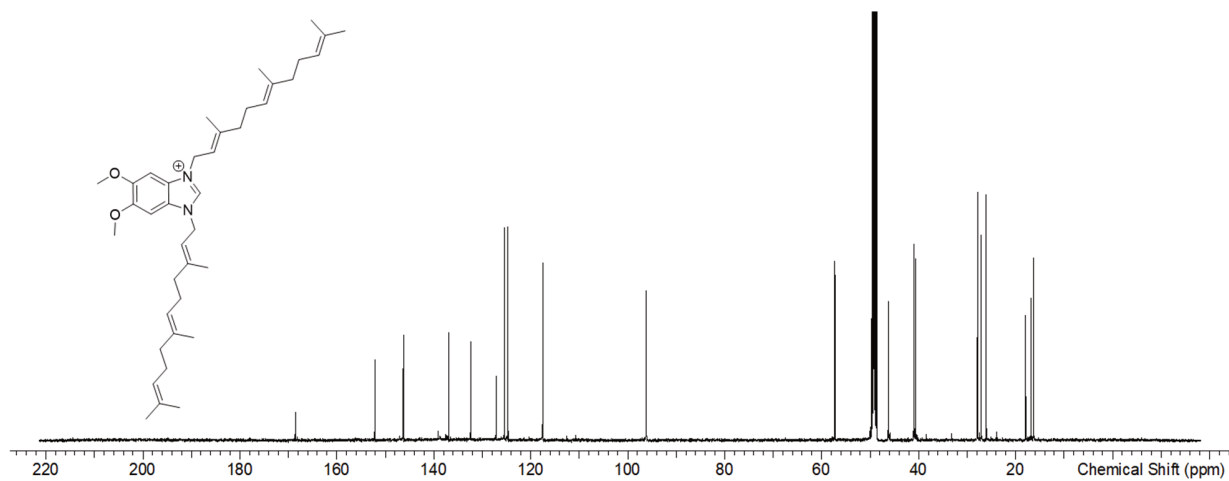
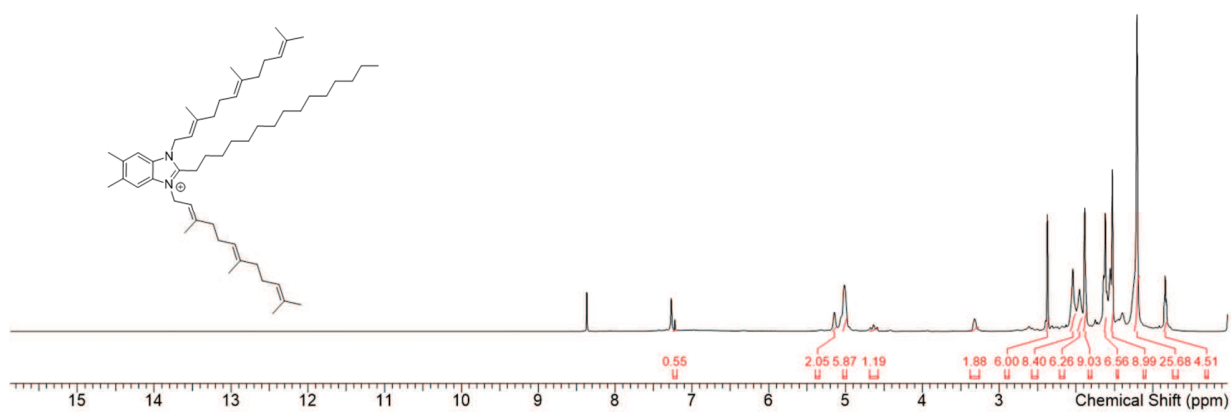
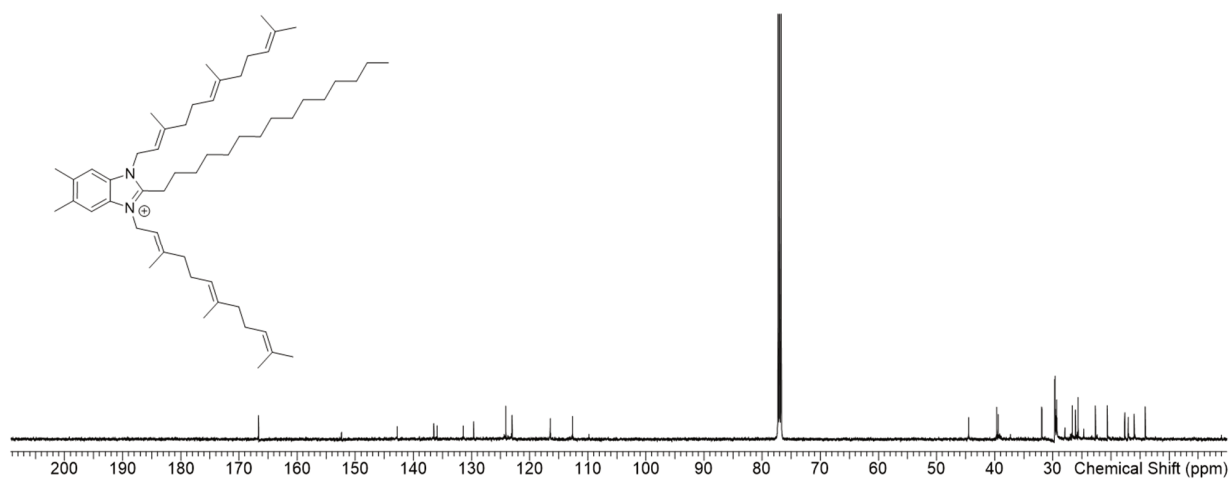


Figure S110. <sup>13</sup>C spectrum of **60** in methanol-d<sub>4</sub> at 126 MHz.



**Figure S111.**  $^1\text{H}$  spectrum of **66** in chloroform-*d* at 500 MHz.



**Figure S112.**  $^{13}\text{C}$  spectrum of **66** in chloroform-*d* at 126 MHz.



## S 5.5 References

1. Bader, C.D.; Panter, F.; Garcia, R.; Tchesnokov, E.P.; Haid, S.; Walt, C.; Spröer, C.; Kiefer, A.F.; Götte, M.; Overmann, J.; et al. Sandacrabins - Structurally Unique Antiviral RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry – A European Journal* **2022**, *28*, e202104484, doi:10.1002/chem.202104484.
2. Marchant, C.A.; Briggs, K.A.; Long, A. In Silico Tools for Sharing Data and Knowledge on Toxicity and Metabolism: Derek for Windows, Meteor, and Vitic. *Toxicology Mechanisms and Methods* **2008**, *18*, 177–187, doi:10.1080/15376510701857320.
3. Sanderson, D.M.; Earnshaw, C.G. Computer prediction of possible toxic action from chemical structure; the DEREK system. *Hum. Exp. Toxicol.* **1991**, *10*, 261–273, doi:10.1177/096032719101000405.

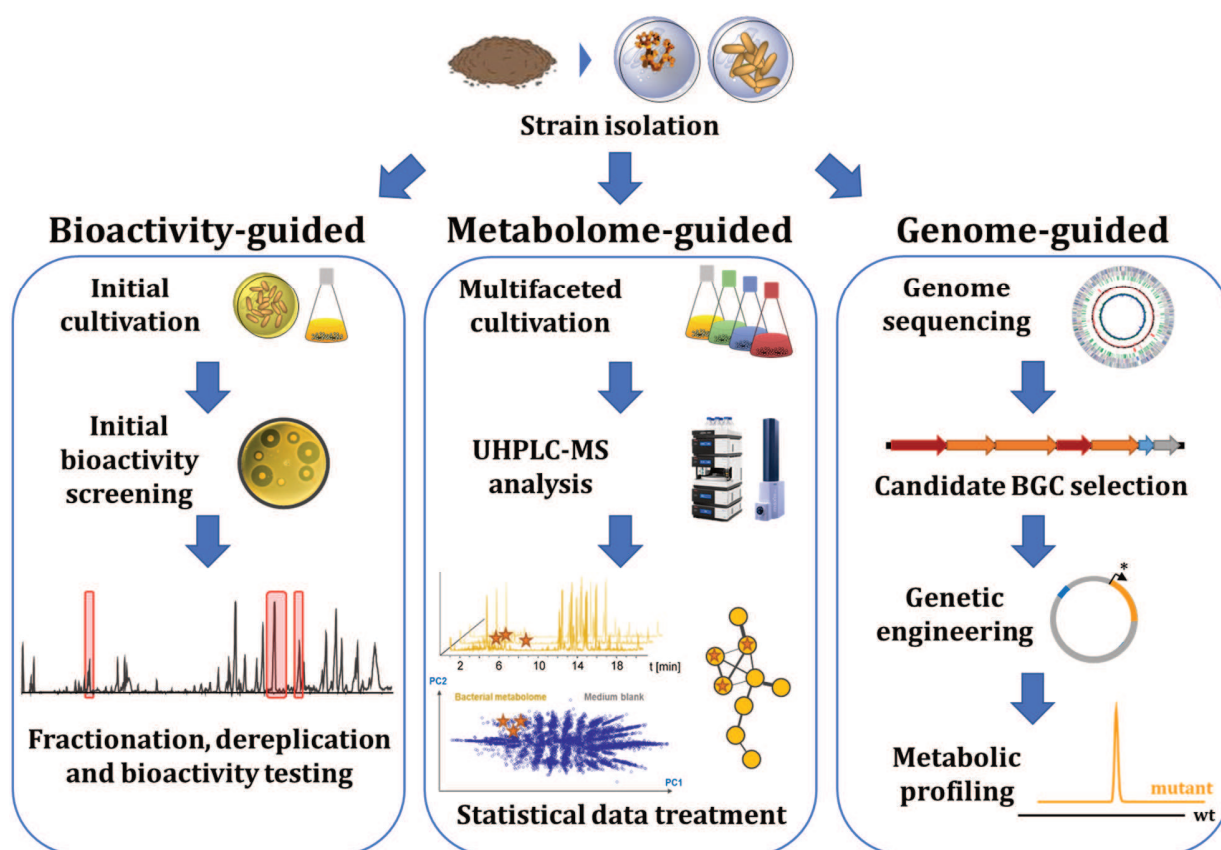
## Chapter 6

### Discussion and Outlook

## 6.1 Challenges and Opportunities in Natural Products Discovery

Because of their versatile biological activities, natural products represent a rich source of novel structures for therapeutic application in humans [1]. Historically, most antimicrobial natural products, especially those from the so-called 'golden era of antibiotics', have been discovered by bioactivity-guided isolation approaches [2]. Since antibiotics produced in large quantities by well-characterized strains have already been largely harvested, natural products research nowadays relies on alternative approaches that access the so-called 'hidden metabolome', which comprises also low-yield compounds and such entities not easily discovered in primary assays. These include, for instance, compounds whose bioactivity could not be detected in primary assays due to poor productivity under the cultivation conditions used, or owing to the microbial test panel. In addition this includes secondary metabolites that could not be detected by instrumental analytics because of low abundance or inappropriate analysis systems and methods. Oftentimes, the hidden metabolome also represents the discrepancy between *in silico* identified biosynthetic gene clusters (BGCs) and actually detected and isolated compounds. The myxobacterial model strain *Myxococcus xanthus* DK1622 for example harbors 25 BGCs according to antiSMASH 6.0 [3] prediction, of which to date only 8 could be assigned to known secondary metabolites [4,5]. Interestingly, although the majority of natural product BGCs in bacterial genomes lacks connected small molecule compounds, transcriptomics and proteomics analyses suggest that most BGCs are expressed to some degree in wild-type strains under standard laboratory conditions [6,7]. Consequently, it is of utmost importance to advance methods for the discovery of novel natural products from both metabolome- and genome-guided approaches. A major opportunity for advancing natural products research towards exploiting the untapped reservoir of new chemical scaffolds and innovative targets lies particularly in the interplay between the different disciplines, namely microbiology, molecular biology and analytical chemistry [8]. Throughout the chapters in this thesis, the application of improved analytical techniques as well as genome-guided approaches enabled the discovery of several novel natural products of myxobacterial origin that would probably not have been identified using classical bioactivity-guided isolation workflows. While myxarylin was discovered by a genome-driven approach (Chapter 2), the light-sensitive compounds from *Coralloccoccus* sp. MCy9049 (Chapter 3) as well as the nicotinic acid-incorporating myxochelins (Chapter 4) and the naturally occurring sandacrabins (Chapter 5) have been identified using in-depth metabolome analysis of the respective myxobacterial strains. A key feature common to all chapters in this thesis is that comprehensive *in silico* analyses of both genomics and metabolomics data played a major role in dereplication and subsequent prioritization of strains, BGCs and

compounds and demonstrated the possibilities and the limitations of contemporary natural product discovery workflows.



**Figure 1.** Typical contemporary natural products discovery workflows featuring the three fundamental compound identification approaches guided by bioactivity, metabolome or genome data.

### 6.1.1 Expanding the Detectable Metabolome from Bacterial Crude Extracts

Due to their broad detection coverage and high resolving power, reversed-phase (RP) ultra-high performance liquid chromatography (UHPLC) coupled to electrospray ionization (ESI) high resolution mass spectrometers (*hrMS*) emerged as the most widely used technique in secondary metabolomics [9]. While these setups are easy to use, cost-efficient and have proven to be very successful in the past, they still come with several constraints [10]. Very polar natural products like myxarylin (Chapter 2) have almost no retention on RP-HPLC columns and thus might get lost in the void peak where the signal is suppressed by co-eluting salts or primary metabolites [11]. For myxarylin, this challenge was resolved in a comparatively simple way by using the ion-pairing agent trifluoroacetic acid (TFA) as a mobile phase modifier, which shifted the retention time of myxarylin from 0.7 min to 8.2 min on a standard 18 min gradient. Nevertheless, TFA is not universally suited for ESI-MS applications as it tremendously lowers sensitivity by

signal suppression [12] and modification of the standardized HPLC-MS setup in general hinders automated database-assisted dereplication as measured retention times and formed adducts are prone to change. Very non-polar compounds like many of the synthetic bisubstituted sandacrabins (Chapter 5) on the other hand show too much affinity to RP-HPLC columns and do not elute at all from the commonly used C<sub>18</sub>-columns in a short standardized gradient. Furthermore, highly hydrophobic natural products like many steroids, terpenes or flavonoids are oftentimes only poorly soluble in HPLC eluents and difficult to protonate in liquid, which makes them often undetectable by standard UHPLC-ESI-MS workflows. In these cases, a fundamentally different analytical setup is required, such as gas chromatography (GC) or supercritical fluid chromatography (SFC) instead of classical HPLC, or atmospheric pressure chemical ionization (APCI) replacing the commonly used ESI as ionization technique [9,13]. A recently published study using myxobacterial secondary metabolomes as an example also suggests that direct infusion (DI) measurements on Fourier transform ion cyclotron resonance (FTICR) spectrometers are capable of providing new insights into the bacterial metabolome that are widely complementary to UHPLC-ESI-MS data [14]. Summing up, it can be stated that while UHPLC-ESI-MS is probably the most versatile technique employed in natural products discovery, it becomes more and more clear that it is not an exhaustive technique in a drug discovery context and likely to miss many secondary metabolites due to the constraints mentioned above.

### 6.1.2 Statistical Treatment of UHPLC-MS Data

The continuous development of LC-MS instruments and methods inevitably leads to ever increasing amounts and complexity of data that can no longer be analyzed manually [15]. In order to allow the efficient mining for new natural products, it is necessary to reduce complexity and data size by automated extraction of the bacterial metabolome data from the vast amounts of 'unimportant' data consisting of matrix features and spectral background [16]. This procedure requires very versatile, but highly standardized and tightly controlled LC-MS conditions that generate reproducible  $m/z$  values and retention times. The most commonly used setup for this application is UHPLC-ESI-MS as described in the previous section. An important step in separating the bacterial metabolome from matrix features is the subdivision of the measured data into molecular features, which in this thesis was performed using the Metaboscape T-ReX 3D framework (Bruker Daltonics, Bremen, Germany). The Metaboscape software framework not only provides feature detection, but also allows for automated database-assisted dereplication as well as statistical data treatment such as analysis of variance (ANOVA) or principal component analysis (PCA). This setup could for instance successfully be used to identify the nicotinic acid-incorporating myxochelins (Chapter 4) as well as the light-sensitive compounds from *Corallococcus* sp. MCy9049 (Chapter 3). The

chance of successfully identifying novel chemical scaffolds can be further improved by adding an additional way of prioritization as for example prioritizing strains by taxonomy [17] or dereplication and prioritization of molecular features by early structural information obtained from tandem MS fragmentation data [18].

### 6.1.3 Extracting Early-Stage Structural Information from Tandem MS Data

Due to the availability of improved instrumentation and methods, detection of presumably new natural products has become a cost-efficient and fast process, however isolation and nuclear magnetic resonance (NMR) spectroscopy-based structure elucidation remain expensive and time-consuming efforts effectively limiting natural products discovery workflows. This finding is mainly related to the fact that stable and sufficient production of natural products, as well as their subsequent purification, are often sophisticated tasks. The molecules described throughout the chapters of this thesis all demonstrated several restraints that needed tailored approaches ultimately leading towards successful structure elucidation or confirmation. Considering the tremendous endeavor, it is of utmost importance to evaluate compounds early on according to their presumed structural novelty and prioritize them accordingly as not to lose time and resources on isolation of unpromising compounds. While NMR experiments require significant amounts of highly pure compounds, tandem MS spectrometry can be used to obtain structural information using only small amounts of unpurified mixtures such as bacterial crude extracts. Specific neutral loss fragments in tandem MS spectra already give strong hints towards incorporated building blocks as for example observed for myxarylin (Chapter 2), where such a fragment could be assigned to leucine/isoleucine, which was later on confirmed by NMR data [19]. Moreover, spectral networking tools such as Global Natural Products Social Molecular Networking (GNPS) not only enable similarity-based clustering of fragmentation data, but furthermore allow comparison with the community-curated GNPS database, a large-scale data repository tandem MS data of known natural products [20,21]. That makes it possible to match fragmentation data with published compounds at an early stage, even if they were recorded using other MS platforms. The annotation with spectral libraries as exemplified by GNPS enables additional dereplication on structural data by distinguishing undescribed derivatives or degradation products of known compounds and media compounds from members of new natural product families. The molecular network generated from extracts of *Coralloccoccus* sp. MCy9049 by GNPS, for example, was used to identify and prioritize the family of light-sensitive compounds described in Chapter 3. The nicotinic acid-containing myxochelins (Chapter 4) on the other hand did not cluster with any of the previously known myxochelins - or each other - despite very similar fragmentation patterns. This is presumably the case because they don't create enough fragments for the similarity matching algorithm. Nevertheless, the distinct and very similar fragmentation patterns of the myxochelins N1-N3 allowed an early assignment of

the newly discovered compounds to the myxochelin natural product family [22]. Last but not least, there are several examples where spectral networking was effectively used to identify and further describe minor derivatives of compound families too low abundant for isolation [23]. The recent development of Parallel Accumulation-Serial Fragmentation (PASEF), a technique that drastically improves MS fragmentation speed and coverage, even allows acquisition and subsequent analysis of fragmentation spectra without the need of previous statistical data reduction thus again boosting the efficiency of future natural product discovery and prioritization workflows [24].

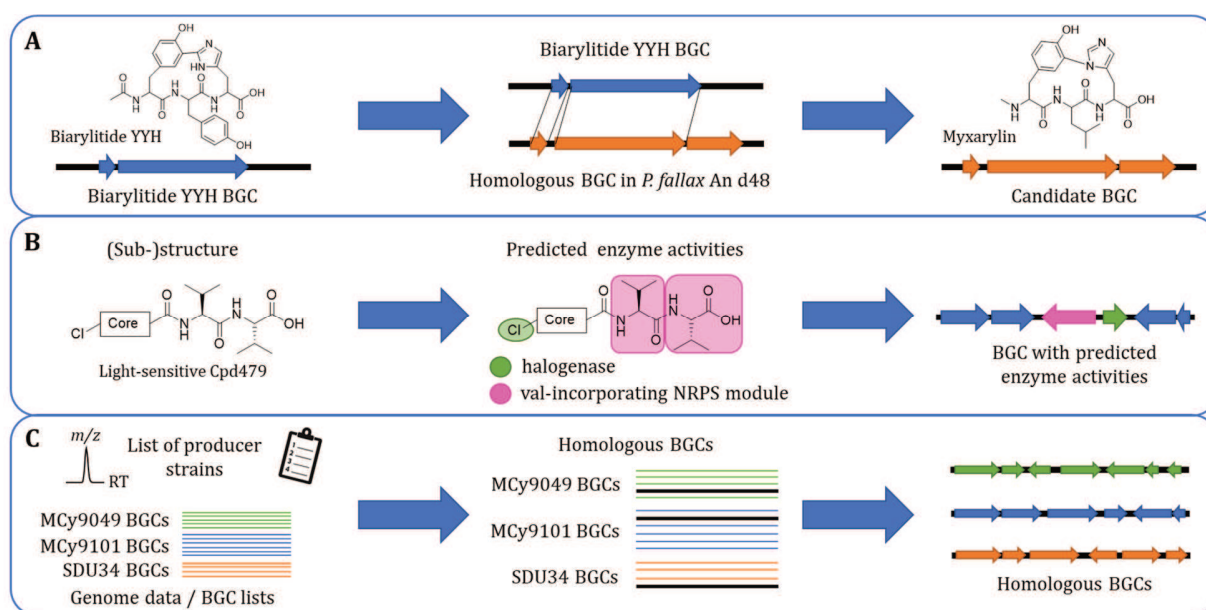
#### 6.1.4 Genome Mining

The development of next generation sequencing techniques such as illumina sequencing, single molecule real time (SMRT) sequencing or nanopore-based sequencing recently enabled fast and cheap full genome sequencing from only minimal amounts of DNA [25]. Along with DNA extraction from environmental samples, these technologies even provide access to genomic data from strains that have not yet been cultivated under laboratory conditions [26]. The ever-increasing amount of publicly available genomic data has led to the development of several *in silico* tools that can automatically detect coding DNA sequences (CDS) and predict their function based on their similarity to reference genes enabling high-throughput full genome annotation [27,28]. Additionally, bioinformatics tools like antiSMASH or PRISM allow efficient prediction and annotation of biosynthetic genes and biosynthetic gene clusters (BGC) and can furthermore predict putative structures of the natural product encoded by a BGC [3,29]. Since the prediction of BGCs is based on similarity of the biosynthetic genes to already known genes, the predictive power varies significantly between different cluster types. Notably, nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) gene clusters encode large enzymatic complexes containing catalytic domains with high sequence homology that make these BGCs relatively easy to detect. Other types of BGCs however, like ribosomally synthesized and post-translationally modified peptide (RiPP) BGC are oftentimes encoded by small and poorly conserved open reading frames (ORFs), which notably hinders BGC identification and annotation [30]. This is well reflected by several recently published natural products such as the cittelins, myxarylin (Chapter 2) and the actinobacterial biarylites, whose BGCs are not detected by any of the genome mining tools applied [19,31,32]. The biosynthetic machinery of the sandacrabins (Chapter 5), which belong to the class of terpenoid-alkaloids, even involves genes that are not clustered in a single BGC at all [33]. Similar observations have already been made for several other natural products such as the imidacins, where the biosynthetic genes were found distributed on at least three distinct chromosomal loci [34]. Analogous to the LC-MS metabolome data discussed in the previous sections, detection capabilities continue to evolve in the genome domain. Likewise, the genome mining

community relies on community-curated data repositories of BGC data such as the Minimum Information about a Biosynthetic Gene Cluster (MIBiG) data repository for reliable annotation and dereplication of genome data. Sharing manually curated genome and related biosynthesis data with the community is a key step to enable efficient evaluation of the biosynthetic potential of newly isolated microorganisms as well as to further understand the complex biosynthesis of bacterial secondary metabolites.

### 6.1.5 Connecting Genome to Metabolome Data

Correlation of genomics and metabolomics data is a crucial step in natural products research that can, for example, enable researchers to formulate a biosynthesis hypothesis or help in interpretation of NMR spectra and *de novo* structure elucidation [35]. In this work, three different *in silico* strategies were employed to establish this link starting from a whole genome sequence or a selected natural product (Figure 2).



**Figure 2.** In silico strategies used in this thesis to connect biosynthetic gene clusters (BGCs) to their corresponding natural products. (A) Compound-based homology search using the biarylittide YYH BGC as template was used to identify the myxarylin BGC. (B) A retrobiosynthesis approach was applied aiming to find a BGC for the light-sensitive compounds (Chapter 4), but eventually failed. (C) Comparative genomics/metabolomics was used for the light-sensitive compounds (Chapter 4) to narrow down the number of candidate BGCs. Strategies adapted from [36,37].

For myxarylin (Chapter 2), initial genome to metabolome correlation was achieved by using a homology search against our in-house genome database. In this case, the published BGC of the closely-related compound biarylittide YYH [32] was used as template to preselect strains based on the presence of a homologous cluster (Figure 2A). The selected myxobacterial strains were subsequently cultured under



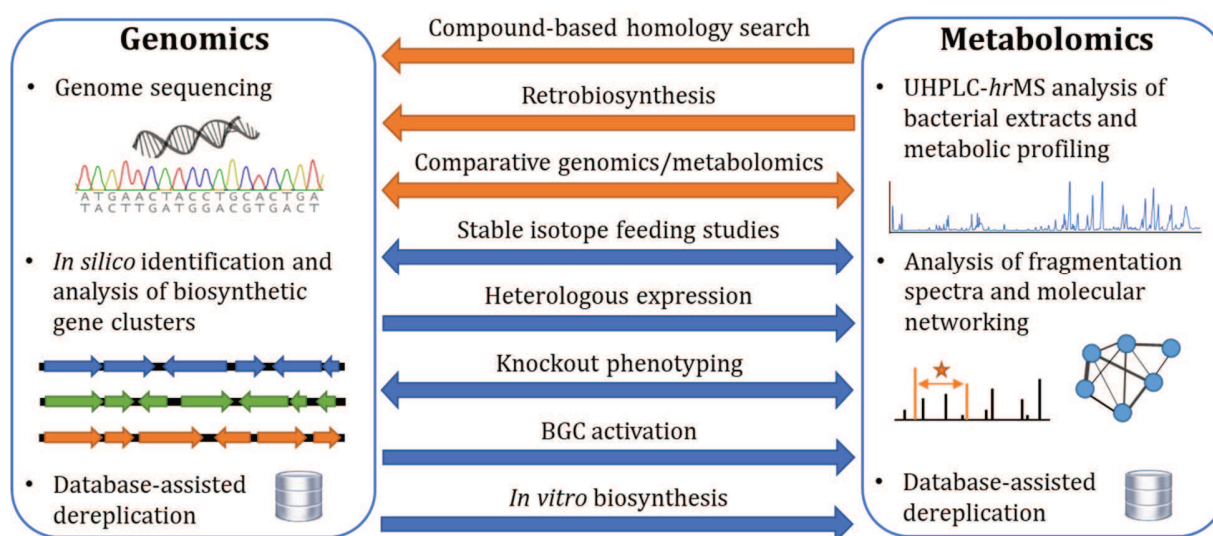
several different growth conditions leading to the identification of a target mass corresponding to the candidate BGC in *P. fallax* An d48. Finally, the putative BGC was confirmed experimentally by heterologous expression in *Myxococcus xanthus* DK1622. While this strategy only works if a similar natural product including the corresponding BGC have already been characterized and published previously, it is still a commonly employed approach [36], which was used, for instance, to identify the novofumigatonin BGC based on homology to the BGCs of austinol, terretonin and andrastin A, which use parts of the same biosynthetic pathway [38]. A similar homology-based approach was also used to link the nicotinic acid-containing myxochelins from *Corallocooccus* sp. MCy9049 (Chapter 4) to their corresponding BGC. Due to the very similar tandem MS fragmentation patterns as well as their exclusive presence together with the already known myxochelin A, the novel compounds were identified relatively quickly as new derivatives of the well-described myxochelin compound family. Homology search in the genome of MCy9049 revealed one myxochelin-like BGC that was found to be likely responsible for the biosynthesis of the novel compounds, especially since it had previously been shown that the promiscuity of MxcE allows the activation and loading of several other benzoic acid derivatives to the myxochelin assembly line [39,40]. Interestingly, another group recently identified the same nicotinic acid-containing myxochelins in the metabolome of *Archangium* sp. SDU34, a strain that is supposedly not able to biosynthesize myxochelin A, even when fed with the precursor 2,3-dihydroxybenzoic acid [41]. While these results are contradictory to the findings from this thesis in so far that in our data the myxochelins N1-N3 were always found paired with the occurrence of myxochelin A, Wang, Niu, *et al.* could not further substantiate the molecular basis for their divergent observation. Like all the myxochelin BGCs from *Archangium* presented in their work, SDU34 lacks *mxcl*, the gene encoding the aminotransferase essential for the biosynthesis of myxochelin B. Besides that, the biosynthetic genes *mxcc-G*, that are indispensable for myxochelin A biosynthesis, were uniformly present and supposedly functional. In the cited study the allegedly altered substrate specificity of the SDU34-MxcE homolog was not correlated to any differences in the sequence compared to other myxochelin producers.

A different strategy that was used with the aim to find a BGC for the light-sensitive compounds from *Corallocooccus* sp. MCy9049 (Chapter 3), is *in silico* retrobiosynthesis (Figure 2B). In this approach, the structure or partial structure of a natural product is used to deduce potential enzyme activities that could lead to the respective moieties. For the light-sensitive compounds, all BGCs of MCy9049 were manually screened for the presence of genes encoding for a putative halogenase in combination with two NRPS modules specific for valine or a RiPP core peptide containing two adjacent valines. Although this strategy failed to identify a BGC for the light-sensitive compounds, it is still a commonly used approach that has been used, for example, to identify the BGCs of scopularide A [42] and usnic acid [43].

An experimental way to correlate metabolome to genome data, which is closely related to the *in silico* retrobiosynthesis approach, is through feeding experiments with stable isotope-labeled precursor molecules such as amino acids for nonribosomal peptides and RiPPs or acyl building blocks for polyketides [44]. In this context, an observed mass shift or change in the isotopic pattern directly hints towards incorporation of the respectively fed building block into a microbial metabolite. Since stable isotopes are comparably expensive, they are mostly used in later stages to confirm matched BGCs, predicted substrate specificities of adenylation and acyltransferase domains and proposed biosynthesis routes. Additionally, feeding of stable isotopes can also help with structure elucidation, especially in combination with tandem MS experiments, and can furthermore be used to generate a comprehensive biosynthesis proposal. In Chapter 3, for instance, several  $^{13}\text{C}$ -,  $^{15}\text{N}$ - and  $^2\text{H}$ -labeled amino acids and other building blocks were fed to liquid cultures of *Coralloccoccus* sp. MCy9049 in order to gain insights into the structures of the yet unknown molecules as well as into the putative biosynthetic machinery. Especially when working with promising novel strains that are taxonomically distant to the well-characterized type strains and harbor many presumably novel BGCs, early on feeding of stable isotopes might certainly be useful to gain insights on the full biosynthetic potential.

Another method that was used for the light-sensitive compounds (Chapter 3) is based on comparative genomics and metabolomics. The starting point of this approach is a list of genome sequenced producer strains that was, in this case, obtained by mining our in-house metabolome database using the exact mass and retention time of the previously identified metabolites. Subsequently, the genome sequences of these strains were used to search for homologous BGCs that are present in all of the producer's genomes. However, as most of the producer strains were closely related species sharing many BGCs, this approach could finally only be used to filter BGCs rather than to identify a single putative BGC. Generally speaking, comparative genomics/metabolomics represents a beneficial *in silico* method that is easy to use and does not require any knowledge about the BGC or chemical structure of the produced compounds. This strategy was successfully applied to link gene clusters to several other compounds such as griseofulvin [45], notoamide [46] or malformin [47]. Nevertheless, when correlating the presence of a BGC in the genome of a strain with the presence of a specific metabolite in its extracts, it needs to be considered that non-producers provide only little information as BGCs could possibly be silent and that an *in silico* correlation only provides an indication that still needs to be verified experimentally [36]. Despite all the *in silico* efforts, the BGC putatively encoding for the light-sensitive compounds from *Coralloccoccus* sp. MCy9049 was finally identified by a rather classical single crossover-mediated gene disruption approach in the alternative producer *Cystobacter* sp. MCy9101.

Although all of the *in silico* strategies presented in this section theoretically allow a genome to compound link independently, a combination of the approaches is often highly recommended. For instance, comparative genomics/metabolomics could be used initially to narrow down the number of candidate BGCs, while *in silico* retrobiosynthesis could be employed to further select a single candidate BGC from the predetermined selection. Nevertheless, it should be kept in mind that these approaches only give a hint of the corresponding BGC and usually require additional experimental proof (Figure 3).



**Figure 3.** Summary of different *in silico* (orange arrows) and experimental (blue arrows) approaches to link genome to metabolome data. Strategies adapted from [36,37,48].

### 6.1.6 Conclusions for Contemporary Natural Products Discovery

As presented in this thesis, natural products research has evolved into a highly interdisciplinary field combining microbiology, analytical chemistry, molecular biology and bioinformatics. Improvements are constantly taking place in all of these areas, which, when strategically combined, also have synergistic influences on all other disciplines [8]. To avoid rediscoveries and unnecessary work, careful dereplication of genomics and metabolomics data is inevitable. As discussed before, these dereplication steps mostly rely on huge community-curated data repositories such as GenBank [28] or MIBiG [49] for genomics data, GNPS [21] or SIRIUS [50] for tandem MS data and the dictionary of natural products (DNP) [51] or NPAtlas [52] for published natural products including analytical and taxonomic data. Additional parameters such as retention time can be used for dereplication of LC-MS data by building in-house databases such as the Myxobase containing analytic data obtained using tightly controlled and highly standardized UHPLC-*hrMS* conditions. Bioinformatics tools like antiSMASH [3] or PRISM [29] enable comprehensive analysis of full genome sequences including detection of BGCs, functional predictions of the genes involved in natural

product biosynthesis and even create predicted structures for the natural products encoded by a putative BGC. Nevertheless, even with all these highly beneficial tools, it remains a tedious process to combine genomics and metabolomics data as these tools rely on several assumptions such as regular and predictable fragmentation of the natural products and collinearity during biosynthesis. Additionally, even after a successful correlation between a molecular feature and a BGC, the follow-up work including reproducible and sufficient production, isolation and full structure elucidation adds additional challenges that oftentimes still lead to ambiguous or incomplete results. This is well exemplified by the light-sensitive compounds described in Chapter 3, where even the combination of BGC data as well as LC-MS, tandem MS and stable isotope feeding data provided only very little indication on the structure of the core scaffold. In BGCs that do not follow the collinearity rule such as the RiPP myxarylin (Chapter 2), it is furthermore very difficult to predict the order of enzymatic reactions without any knowledge about substrate specificities of the biosynthetic enzymes. The construction of a comprehensive biosynthesis proposal in these cases would need the complete biosynthesis to be conducted *in vitro*, which is again very time-consuming.

Summing up, it can be noted that there is constant development in all of the interdisciplinary fields involved in natural products research. Nevertheless, many processes are still far away from being automated or conducted in a high-throughput manner. It is therefore necessary for natural product researchers to share their data with the scientific community in order to constantly improve the various tools used for effective prioritization of secondary metabolites and thus to avoid unnecessary work. Furthermore, leaving the “comfort zone” of traditional and well-established methodologies can help to avoid rediscovery and explore yet untapped parts of the natural products reservoir.

## 6.2 Organic Synthesis in Natural Products Discovery, Isolation and Characterization Workflows

### 6.2.1 Total Synthesis of Natural Products

Total synthesis is a process that is widely used in natural products research. Advantages include the elimination of biological production systems, which avoids many correlated constraints like long growing times, danger of contamination of the cultures, differences in between batches and yield limitations by self-toxicity. Furthermore, a well-established total synthesis route enables comparably simple modification of natural products. Nevertheless, natural products tend to contain large scaffolds with low hydrophobicity, are rich in stereocenters – including difficult to synthesize stereogenic quaternary carbons

– and harbor an astonishing molecular complexity [53,54]. To make the effort worthwhile, total synthesis is usually carried out only when a target natural product has already been characterized in detail and evaluated as promising. At least an initial planar structure hypothesis of a natural product is obviously the prerequisite for designing a total synthesis route. However, particularly in the case of easy-to-implement total synthesis routes, it can be beneficial to employ total synthesis at an early state. For example, for the nicotinic acid-containing myxochelins (Chapter 4), a previously published synthesis route for myxochelin A could be modified and was employed to elucidate the stereochemistry of the new derivatives and to produce additional non-natural congeners [22]. For the sandacrabins, a total synthesis route was established in order to achieve higher yields necessary for follow-up experiments and to confirm the proposed structures of the monofarnesylated minor derivatives sandacrabins D and E [33]. As a result, the synthesis route only needed to be slightly adapted to advance directly into generation of a small synthetic sandacrabins library used for structure-activity relationship (SAR) studies (Chapter 5).

### 6.2.2 Semi-synthetic Approaches in Natural Products Isolation and Characterization Workflows

In many instances, such as the determination of the stereochemistry, performing *de novo* total synthesis of natural products is not essential and can be replaced by semi-synthetic methods. For example, the absolute stereochemistry of various primary amines like amino acids can be assessed by derivatization with D- respective L-(1-fluoro-2,4-dinitrophenyl-5-leucine amide) (FDLA), commonly known as Marfey's reagent [55,56]. This method was for example used to determine the configuration of all three amino acids of myxarylin (Chapter 2). Derivatization with (*R*)- respective (*S*)-3,3,3-trifluoro-2-methoxy-2-phenylpropanoic acid (MTPA), also known as the Mosher's method, represents another small-scale synthetic method to determine the absolute stereochemistry of natural products that is mostly applied to secondary alcohols, but can also be used for amines or carboxylic acids [57,58].

Apart from frequent application on pure compounds, semi-synthesis on complex mixtures – for example in natural products isolation workflows – is only found very rarely. Among the few examples described in literature are mostly chemical labeling strategies for small molecules that aim to facilitate the detection of metabolites containing certain functional groups or structural motifs by using derivatizing agents that introduce easily ionizable groups, chromophores or chlorine and bromine atoms that are easy to dereplicate [59]. Although moieties like stereocenters, that are difficult to assess in total synthesis, do not concern here, such reactions are nevertheless challenging for several reasons. First, the reactions must be specific, robust and able to proceed even in the presence of a large number of interfering molecules such as nucleophiles, electrophiles or redox-active compounds as biological mixtures are highly complex.

Many natural products are furthermore sensitive to several reaction conditions such as high temperatures or strong acids and bases. Additionally, numerous natural products are highly hydrophilic and might not be sufficiently soluble in various organic solvents as for example observed for myxarylin (Chapter 2) or the light-sensitive compounds presented in chapter 3, which were found to be only soluble in water and methanol. Last but not least, the reactants need to be comparably inexpensive as reactions on crude mixtures such as bacterial extracts or even fractions consume many equivalents until sufficient conversion of the desired natural products is achieved. Nevertheless, semi-synthetic approaches to facilitate the isolation of difficult-to-handle natural products might be beneficial. In chapter 3, we performed catalytic hydrogenation using hydrogen and palladium over activated carbon in order to suppress compound isomerization and light-induced degradation. Although the goal of stabilizing the sensitive natural products to facilitate the isolation was ultimately not achieved, it was still shown that conversion to the hydrogenated derivatives could be achieved. However, such reactions have a significant disadvantage as they are not reversible and thus the “native” natural product cannot easily be recovered. Protecting groups are widely used in small-scale organic synthesis to obtain chemoselectivity in following reactions and can afterwards be cleaved off easily [60]. Since the pool of protecting groups for most functional groups is enormous, the choice can be adapted to the stability and properties of the target natural product, especially considering the deprotection reaction to recover the native metabolite. The isolation of zwitterionic compounds such as relebactam (previously known as MK-7655) or many peptides containing primary or secondary amine as well as carboxyl groups often presents significant challenges because of their unique properties [61]. Alike myxarylin (Chapter 2), the  $\beta$ -lactamase inhibitor relebactam was found to be only soluble in water and polar aprotic solvents such as dimethyl sulfoxide (DMSO) or *N*-methyl-2-pyrrolidone (NMP), which effectively hindered compound purification. However, using a tailored deprotection strategy, the compound could be isolated fairly easily via its *N*-*tert*-butyloxycarbonyl (*N*-Boc) protected congener [61]. Likewise, for myxarylin (Chapter 2), a Boc protection group was introduced at the *N*-terminus to increase retention on an RP-HPLC column and thus to enable efficient purification using semi-preparative HPLC. Deprotection of the purified compound was finally achieved by using the strong acid trifluoroacetic acid. An alternative group for protection of the *N*-terminus of peptides is 9-fluorenylmethyloxycarbonyl (Fmoc), which is widely used in solid phase peptide synthesis and usually cleaved off by using a base such as piperidine [62]. Although semi-synthesis in natural products isolation workflows is definitely not essential or beneficial for all compounds, the projects presented in this thesis demonstrate that the semi-synthesis-assisted purification of natural products is indeed a possibility if conventional purification methods fail to retrieve sufficient amounts of pure compound.

### 6.2.3 Conclusions on the Use of Organic Synthesis in Natural Products Research

Organic synthesis is a widely used method in natural products research. Especially total synthesis of bioactive natural products is of great importance as it has the potential to provide access to large yields and paves the way for SAR studies and following hit-to-lead development. Also from a synthetic point of view, total synthesis of natural products represents interesting challenges as secondary metabolites are often highly complex molecules equipped with multiple stereocenters and reactive functional groups that raise many problems in terms of regio- and stereoselectivity [63,64]. In the course of this thesis, total synthesis approaches played a role to elucidate the stereochemistry of the myxochelins N1-N3 (Chapter 4) as well as to fully characterize the natural products and to generate additional synthetic congeners. Furthermore, a small library of sandacrabins derivatives was generated in order to conduct an SAR study aiming to enhance antiviral activities while reducing cytotoxicity (Chapter 5). It should be noted, however, that unlike many other natural products, both the myxochelins and sandacrabins are comparatively easily synthetically accessible, which made early stage total synthesis feasible in the first place.

Small-scale semi-synthesis represents a method that is also frequently used to determine the absolute stereochemistry of natural products. Commonly used reactions include the derivatization of secondary alcohols with Mosher's acid or primary amines with Marfey's reagent. The latter was for example used to elucidate the stereochemistry of the three amino acids building up myxarylin (Chapter 2). In contrast, organic synthesis is rarely used in earlier steps such as the isolation of natural products. While for the light-sensitive compounds from *Coralloccoccus* sp. MCy9049, catalytic hydrogenation failed to stabilize the compounds and enable purification (Chapter 3), for Myxarylin, the introduction of a Boc protecting group enabled isolation by RP-HPLC and subsequent deprotection with TFA allowed to recover the original natural product (Chapter 2). As many of the so-called low-hanging fruits in natural products research have already widely been isolated and characterized, semi-synthesis-assisted isolation of natural products might represent a possibility to access natural products that have previously been dismissed because of properties effectively hindering isolation approaches. Although it remains to be established whether such methods will become widespread at some point, they can still be of great advantage in some instances.

## 6.3 Accessing the Function of Natural Products

The biosynthesis of natural products is a highly complex process that consumes a lot of energy. Therefore, it can be assumed that the producing organism somehow benefits from the production of natural products and thus gains an evolutionary advantage over competing microorganisms [65].

Especially in genome- and metabolome-guided natural products isolation workflows, it is alleged that some of the compounds that are beneficial to their producing organism are also in some way useful for human application. Typically, this aims at antimicrobial, antiviral or cytotoxic compounds synthesized for defense, which can be used in human medicine to treat bacterial, viral and fungal infections or find their application in cancer treatment [66]. For instance, the sandacrabins (Chapter 5), which were also detected by metabolome mining, show promising antiviral activities against several human pathogenic viruses and might eventually find an application in human medicine. There are numerous theories on why many natural products have proven pharmacological effects that are beneficial for humans although these compounds are certainly not designed to interact with human proteins or to inhibit human pathogens. One prominent theory answers this question from a structural biology perspective. It is estimated that the theoretically infinite number of protein structures falls into a very limited number of 1,000-1,700 folding patterns for all naturally occurring protein structures, which hints towards a prevalent evolutionary “recycling” of protein binding sites. As a consequence, drug targets may contain similar structural domains to the targets with which natural products have evolved to interact [67–69]. However, many natural products like myxarylin (Chapter 2), the light-sensitive compounds from MCy9049 (Chapter 3) and the nicotinic acid-containing myxochelins (Chapter 4) do not have a biological function assigned so far. Besides potential defense roles against competing microorganisms, natural products can exhibit diverse other functions such as pigments, signaling molecules or iron-chelating siderophores [70]. Additionally, natural products are evolutionary optimized compounds that tend to display highly specific interactions with their respective target molecules. Consequently, a lack of bioactivity from initial screening with a standard panel of pathogens does not necessarily mean that the tested natural product is inactive against any microorganism or cannot be beneficial for humans. Indeed, there are several examples in the literature, where natural products have initially been isolated without any observed bioactivity, but were later still found active like crocagin A, which showed no antimicrobial or cytotoxic activity in initial screenings, but was later found to be an inhibitor of carbon storage regulator protein A (CsrA) [71]. Therefore, it is particularly important to also publish so far ‘inactive’ secondary metabolites and make them available to the scientific community by sharing data and integrating these compounds into natural products compound libraries. This allows the compounds to be used by other research groups for computational studies or virtual screening using new target structures, for bioactivity screenings against pathogens not included in the initial panel, or assays focusing on other functions beneficial for human use. Ultimately, as observed for many other cryptic natural products isolated in the past, the yet elusive function of several natural products discovered in the course of this thesis might be found, whether or not it is suitable to be developed towards an application beneficial for humans.



## 6.4 Concluding Remarks

Myxobacteria have proven to be prolific producers of a multitude of natural products with a broad variety of chemical scaffolds as well as biological activities. In the course of this thesis, genome and metabolome-guided approaches were followed, leading to the discovery of several structurally diverse metabolites from diverse biosynthetic origins. Furthermore, it could be demonstrated that organic synthesis is not only a valuable tool to prove the assignment of structures, improve yields and advance hit to lead development, but can also be highly beneficial in earlier stages of natural products research such as compound isolation workflows. Taken together, the findings presented in this thesis support the notion that the majority of the natural product reservoir is still untapped, but the continuous development of technologies and exchange of knowledge among the key disciplines in natural products research will help to advance from only scratching the surface to exploiting the full potential at some point in the future.

## 6.5 References

1. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803, doi:10.1021/acs.jnatprod.9b01285.
2. Lyddiard, D.; Jones, G.L.; Greatrex, B.W. Keeping it simple: lessons from the golden era of antibiotic discovery. *FEMS Microbiol. Lett.* **2016**, *363*, fnw084, doi:10.1093/femsle/fnw084.
3. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* **2021**, *49*, W29-W35, doi:10.1093/nar/gkab335.
4. Wenzel, S.C.; Muller, R. Myxobacteria—‘microbial factories’ for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, *5*, 567–574, doi:10.1039/b901287g.
5. Bader, C.D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **2020**, *39*, 107480, doi:10.1016/j.biotechadv.2019.107480.
6. Amos, G.C.A.; Awakawa, T.; Tuttle, R.N.; Letzel, A.-C.; Kim, M.C.; Kudo, Y.; Fenical, W.; Moore, B.; Jensen, P.R. Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E11121-E11130, doi:10.1073/pnas.1714381115.
7. Schley, C.; Altmeyer, M.O.; Swart, R.; Müller, R.; Huber, C.G. Proteome analysis of *Myxococcus xanthus* by off-line two-dimensional chromatographic separation using monolithic poly-(styrene-

- divinylbenzene) columns combined with ion-trap tandem mass spectrometry. *J. Proteome Res.* **2006**, *5*, 2760–2768, doi:10.1021/pr0602489.
8. Hug, J.J.; Bader, C.D.; Remškar, M.; Cirnski, K.; Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **2018**, *7*, 44, doi:10.3390/antibiotics7020044.
  9. Panter, F.; Bader, C.D.; Müller, R. Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Chem. Sci.* **2021**, 5994–6010, doi:10.1039/D0SC06919A.
  10. Chen, F.; Wang, M.; Cheng, K.-W. Liquid Chromatography-Mass Spectrometry in Natural Product Research. In *Bioactive natural products: Detection, isolation, and structural determination / edited by Steven M. Colegate, Russell J. Molyneux*, 2<sup>nd</sup> ed.; Colegate, S.M., Molyneux, R.J., Eds.; CRC Press: Boca Raton, 2008; pp 245–2651, ISBN 978-0-8493-7258-2.
  11. Annesley, T.M. Ion Suppression in Mass Spectrometry. *Clin. Chem.* **2003**, *49*, 1041–1044, doi:10.1373/49.7.1041.
  12. Nshanian, M.; Lakshmanan, R.; Chen, H.; Ogorzalek Loo, R.R.; Loo, J.A. Enhancing Sensitivity of Liquid Chromatography-Mass Spectrometry of Peptides and Proteins Using Supercharging Agents. *International Journal of Mass Spectrometry* **2018**, *427*, 157–164, doi:10.1016/j.ijms.2017.12.006.
  13. Ma, Y.-C.; Kim, H.-Y. Determination of steroids by liquid chromatography/mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 1010–1020, doi:10.1016/S1044-0305(97)00122-0.
  14. Bader, C.D.; Haack, P.A.; Panter, F.; Krug, D.; Müller, R. Expanding the Scope of Detectable Microbial Natural Products by Complementary Analytical Methods and Cultivation Systems. *J. Nat. Prod.* **2021**, doi:10.1021/acs.jnatprod.0c00942.
  15. Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **2014**, *86*, 10780–10788, doi:10.1021/ac502805w.
  16. Ito, T.; Masubuchi, M. Dereplication of microbial extracts and related analytical technologies. *J. Antibiot* **2014**, *67*, 353–360, doi:10.1038/ja.2014.12.
  17. Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **2018**, *9*, 803, doi:10.1038/s41467-018-03184-1.
  18. Dührkop, K.; Nothias, L.-F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M.A.; Petras, D.; Gerwick, W.H.; Rousu, J.; Dorrestein, P.C.; et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **2020**, doi:10.1038/s41587-020-0740-8.

19. Hug, J.J.; Frank, N.A.; Walt, C.; Šenica, P.; Panter, F.; Müller, R. Genome-Guided Discovery of the First Myxobacterial Biarylittide Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis. *Molecules* **2021**, *26*, 7483, doi:10.3390/molecules26247483.
20. Vargas, F.; Weldon, K.C.; Sikora, N.; Wang, M.; Zhang, Z.; Gentry, E.C.; Panitchpakdi, M.W.; Caraballo-Rodríguez, A.M.; Dorrestein, P.C.; Jarmusch, A.K. Protocol for Community-created Public MS/MS Reference Spectra Within the Global Natural Products Social Molecular Networking Infrastructure. *Rapid Commun. Mass Spectrom.* **2020**, e8725, doi:10.1002/rcm.8725.
21. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837, doi:10.1038/nbt.3597.
22. Frank, N.A.; Széles, M.; Akone, S.H.; Rasheed, S.; Hüttel, S.; Frewert, S.; Hamed, M.M.; Herrmann, J.; Schuler, S.M.M.; Hirsch, A.K.H.; et al. Expanding the Myxochelin Natural Product Family by Nicotinic Acid Containing Congeners. *Molecules* **2021**, *26*, 4929, doi:10.3390/molecules26164929.
23. Panter, F.; Krug, D.; Müller, R. Novel Methoxymethacrylate Natural Products Uncovered by Statistics-Based Mining of the *Myxococcus fulvus* Secondary Metabolome. *ACS Chem. Biol.* **2019**, *14*, 88–98, doi:10.1021/acscchembio.8b00948.
24. Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M.A.; Raether, O.; Mann, M. Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **2015**, *14*, 5378–5387, doi:10.1021/acs.jproteome.5b00932.
25. Slatko, B.E.; Gardner, A.F.; Ausubel, F.M. Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology* **2018**, *122*, doi:10.1002/cpmb.59.
26. Garza; Dutilh, B.E. From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems. *Cell. Mol. Life Sci.* **2015**, *72*, doi:10.1007/s00018-015-2004-1.
27. Delcher, A.L.; Bratke, K.A.; Powers, E.C.; Salzberg, S.L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **2007**, *23*, 673–679, doi:10.1093/bioinformatics/btm009.
28. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2013**, *41*, D36–42, doi:10.1093/nar/gks1195.
29. Skinnider, M.A.; Johnston, C.W.; Gunabalasingam, M.; Merwin, N.J.; Kieliszek, A.M.; MacLellan, R.J.; Li, H.; Ranieri, M.R.M.; Webster, A.L.H.; Cao, M.P.T.; et al. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Commun.* **2020**, *11*, 6058, doi:10.1038/s41467-020-19986-1.

30. Jong, A. de; van Hijum, S.A.F.T.; Bijlsma, J.J.E.; Kok, J.; Kuipers, O.P. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res* **2006**, *34*, W273-W279, doi:10.1093/nar/gkl237.
31. Hug, J.J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. Biosynthesis of Cittelins, Unusual Ribosomally Synthesized and Post-translationally Modified Peptides from *Myxococcus xanthus*. *ACS Chem. Biol.* **2020**, *15*, 2221–2231, doi:10.1021/acscchembio.0c00430.
32. Zdouc, M.M.; Alanjary, M.M.; Zarazúa, G.S.; Maffioli, S.I.; Crüsemann, M.; Medema, M.H.; Donadio, S.; Sosio, M. A biaryl-linked tripeptide from Planomonospora reveals a widespread class of minimal RiPP gene clusters. *Cell Chem. Biol.* **2020**, doi:10.1016/j.chembiol.2020.11.009.
33. Bader, C.D.; Panter, F.; Garcia, R.; Tchesnokov, E.P.; Haid, S.; Walt, C.; Spröer, C.; Kiefer, A.F.; Götte, M.; Overmann, J.; et al. Sandacrabins - Structurally Unique Antiviral RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry – A European Journal* **2022**, *28*, e202104484, doi:10.1002/chem.202104484.
34. Krug, D. Natural product biosynthesis in myxobacteria: Studies on enzymatic versatility and secondary metabolite diversity. Dissertation; Saarland University, Saarbrücken, 2009.
35. Doroghazi, J.R.; Albright, J.C.; Goering, A.W.; Ju, K.-S.; Haines, R.R.; Tchalukov, K.A.; Labeda, D.P.; Kelleher, N.L.; Metcalf, W.W. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **2014**, *10*, 963–968, doi:10.1038/nchembio.1659.
36. Kjærboelling, I.; Mortensen, U.H.; Vesth, T.; Andersen, M.R. Strategies to establish the link between biosynthetic gene clusters and secondary metabolites. *Fungal Genet. Biol.* **2019**, *130*, 107–121, doi:10.1016/j.fgb.2019.06.001.
37. van der Hooft, J.J.J.; Mohimani, H.; Bauermeister, A.; Dorrestein, P.C.; Duncan, K.R.; Medema, M.H. Linking genomics and metabolomics to chart specialized metabolic diversity. *Chem. Soc. Rev.* **2020**, doi:10.1039/d0cs00162g.
38. Matsuda, Y.; Bai, T.; Phippen, C.B.W.; Nørdvig, C.S.; Kjærboelling, I.; Vesth, T.C.; Andersen, M.R.; Mortensen, U.H.; Gotfredsen, C.H.; Abe, I.; et al. Novofumigatonin biosynthesis involves a non-heme iron-dependent endoperoxide isomerase for orthoester formation. *Nat Commun* **2018**, *9*, 1–10, doi:10.1038/s41467-018-04983-2.
39. Korp, J.; König, S.; Schieferdecker, S.; Dahse, H.-M.; König, G.M.; Werz, O.; Nett, M. Harnessing Enzymatic Promiscuity in Myxochelin Biosynthesis for the Production of 5-Lipoxygenase Inhibitors. *ChemBioChem* **2015**, *16*, 2445–2450, doi:10.1002/cbic.201500446.
40. Gaitatzis, N.; Kunze, B.; Müller, R. Novel insights into siderophore formation in myxobacteria. *ChemBioChem* **2005**, *6*, 365–374, doi:10.1002/cbic.200400206.

41. Wang, D.-G.; Niu, L.; Lin, Z.-M.; Wang, J.-J.; Gao, D.-F.; Sui, H.-Y.; Li, Y.-Z.; Wu, C. The Discovery and Biosynthesis of Nicotinic Myxochelins from an Archangium sp. SDU34. *J. Nat. Prod.* **2021**, doi:10.1021/acs.jnatprod.1c00524.
42. Lukassen, M.B.; Saei, W.; Sondergaard, T.E.; Tamminen, A.; Kumar, A.; Kempken, F.; Wiebe, M.G.; Sørensen, J.L. Identification of the Scopularide Biosynthetic Gene Cluster in *Scopulariopsis brevicaulis*. *Mar. Drugs* **2015**, *13*, 4331–4343, doi:10.3390/md13074331.
43. Abdel-Hameed, M.; Bertrand, R.L.; Piercey-Normore, M.D.; Sorensen, J.L. Putative identification of the usnic acid biosynthetic gene cluster by de novo whole-genome sequencing of a lichen-forming fungus. *Fungal Biology* **2016**, *120*, 306–316, doi:10.1016/j.funbio.2015.10.009.
44. McCaughey, C.S.; van Santen, J.A.; van der Hooft, J.J.J.; Medema, M.H.; Lington, R.G. An isotopic labeling approach linking natural products with biosynthetic gene clusters. *Nat Chem Biol* **2022**, *18*, 295–304, doi:10.1038/s41589-021-00949-6.
45. Chooi, Y.-H.; Cacho, R.; Tang, Y. Identification of the viridicatumtoxin and griseofulvin gene clusters from *Penicillium aethiopicum*. *Chem. Biol.* **2010**, *17*, 483–494, doi:10.1016/j.chembiol.2010.03.015.
46. Li, S.; Srinivasan, K.; Tran, H.; Yu, F.; Finefield, J.M.; Sunderhaus, J.D.; McAfoos, T.J.; Tsukamoto, S.; Williams, R.M.; Sherman, D.H. Comparative analysis of the biosynthetic systems for fungal bicyclo[2.2.2]diazaoctane indole alkaloids: the (+)/(-)-notoamide, paraherquamide and malbrancheamide pathways. *Med. Chem. Commun.* **2012**, *3*, 987–996, doi:10.1039/c2md20029e.
47. Theobald, S.; Vesth, T.C.; Rendsvig, J.K.; Nielsen, K.F.; Riley, R.; Abreu, L.M. de; Salamov, A.; Frisvad, J.C.; Larsen, T.O.; Andersen, M.R.; et al. Uncovering secondary metabolite evolution and biosynthesis using gene cluster networks and genetic dereplication. *Sci. Rep.* **2018**, *8*, 17957, doi:10.1038/s41598-018-36561-3.
48. Sekurova, O.N.; Schneider, O.; Zotchev, S.B. Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. *Microbial Biotechnology* **2019**, *12*, 828–844, doi:10.1111/1751-7915.13398.
49. Kautsar, S.A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J.C.; Terlouw, B.R.; van der Hooft, Justin J J; van Santen, J.A.; Tracanna, V.; Suarez Duran, H.G.; Pascal Andreu, V.; et al. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* **2020**, *48*, D454–D458, doi:10.1093/nar/gkz882.
50. Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A.A.; Melnik, A.V.; Meusel, M.; Dorrestein, P.C.; Rousu, J.; Böcker, S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **2019**, *16*, 299–302, doi:10.1038/s41592-019-0344-8.

51. CRC Press. Dictionary of Natural Products 30.2. Available online: <https://dnp.chemnetbase.com> (accessed on 3 January 2022).
52. van Santen, J.A.; Poynton, E.F.; Iskakova, D.; McMann, E.; Alsup, T.A.; Clark, T.N.; Fergusson, C.H.; Fewer, D.P.; Hughes, A.H.; McCadden, C.A.; et al. The Natural Products Atlas 2.0: a database of microbially-derived natural products. *Nucleic Acids Res.* **2022**, *50*, D1317-D1323, doi:10.1093/nar/gkab941.
53. Stratton, C.F.; Newman, D.J.; Tan, D.S. Cheminformatic comparison of approved drugs from natural product versus synthetic origins. *Bioorganic & Medicinal Chemistry Letters* **2015**, *25*, 4802–4807, doi:10.1016/j.bmcl.2015.07.014.
54. Peterson, E.A.; Overman, L.E. Contiguous stereogenic quaternary carbons: a daunting challenge in natural products synthesis. *PNAS* **2004**, *101*, 11943–11948, doi:10.1073/pnas.0402416101.
55. Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res. Commun.* **1984**, *49*, 591–596, doi:10.1007/BF02908688.
56. Bhushan, R.; Bruckner, H. Marfey's reagent for chiral amino acid analysis: A review. *Amino Acids* **2004**, *27*, 231–247, doi:10.1007/s00726-004-0118-0.
57. Dale, J.A.; Mosher, H.S. Nuclear magnetic resonance enantiomer reagents. Configurational correlations via nuclear magnetic resonance chemical shifts of diastereomeric mandelate, O-methylmandelate, and .alpha.-methoxy-.alpha.-trifluoromethylphenylacetate (MTPA) esters. *J. Am. Chem. Soc.* **1973**, *95*, 512–519, doi:10.1021/ja00783a034.
58. Hoye, T.R.; Jeffrey, C.S.; Shao, F. Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nat. Protoc.* **2007**, *2*, 2451–2458, doi:10.1038/nprot.2007.354.
59. Chambers C. Hughes. Chemical labeling strategies for small molecule natural product detection and isolation. *Nat. Prod. Rep.* **2021**, *38*, 1684–1705, doi:10.1039/D0NP00034E.
60. Kocienski, P.J. *Protecting groups*, 3<sup>rd</sup> ed.; Georg Thieme Verlag: Stuttgart, 2005, ISBN 9783131841735.
61. Liu, Z.; Yasuda, N.; Simeone, M.; Reamer, R.A. N-Boc Deprotection and Isolation Method for Water-Soluble Zwitterionic Compounds. *J. Org. Chem.* **2014**, *79*, 11792–11796, doi:10.1021/jo502319z.
62. Chan, W.C.; White, P.D. *Fmoc solid phase peptide synthesis: A practical approach / edited by Weng C. Chan and Peter D. White*; Oxford University Press: Oxford, 2000, ISBN 0199637253.
63. Roberts, A.A.; Ryan, K.S.; Moore, B.S.; Gulder, T.A.M. Total (bio)synthesis: Strategies of nature and of chemists. *Natural Products Via Enzymatic Reactions* **2010**, *297*, 149–203, doi:10.1007/128\_2010\_79.

64. K. C. Nicolaou; Dionisios Vourloumis; Nicolas Winssinger; Phil S. Baran. The Art and Science of Total Synthesis at the Dawn of the Twenty-First Century. *Angewandte Chemie International Edition* **2000**, 39, 44–122, doi:10.1002/(SICI)1521-3773(20000103)39:1<44:AID-ANIE44>3.0.CO;2-L.
65. Davies, J. Specialized microbial metabolites: functions and origins. *J. Antibiot.* **2013**, 66, 361–364, doi:10.1038/ja.2013.61.
66. Berdy, J. Bioactive microbial metabolites. *J. Antibiot.* **2005**, 58, 1–26, doi:10.1038/ja.2005.1.
67. Gerstein, M.; Hegyi, H. Comparing genomes in terms of protein structure: surveys of a finite parts list. *FEMS Microbiol. Rev.* **1998**, 22, 277–304, doi:10.1111/j.1574-6976.1998.tb00371.x.
68. Chothia, C. One thousand families for the molecular biologist. *Nature* **1992**, 357, 543–544, doi:10.1038/357543a0.
69. Grabowski, K.; Schneider, G. Properties and Architecture of Drugs and Natural Products Revisited. *Current Chemical Biology* **2007**, 1, 115–127, doi:10.2174/187231307779814066.
70. Weissman, K.J.; Müller, R. A brief tour of myxobacterial secondary metabolism. *Bioorg. Med. Chem.* **2009**, 17, 2121–2136, doi:10.1016/j.bmc.2008.11.025.
71. Viehrig, K.; Surup, F.; Volz, C.; Herrmann, J.; Abou Fayad, A.; Adam, S.; Kohnke, J.; Trauner, D.; Müller, R. Structure and biosynthesis of crocagins: polycyclic postranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem.* **2017**, 1–5, doi:10.1002/anie.201612640.