The Discovery of Natural Products in Actinobacteria

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Shuning Xia

aus Guizhou, China

Tübingen

2026

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	08.06.2026
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Nadine Ziemert
2. Berichterstatter/-in:	Prof. Dr. Heike Brötz-Oesterhelt

Die vorliegende Arbeit wurde im Interfakultäres Institut für Mikrobiologie und Infektion-smedizin der Eberhard Karls Universität Tübingen von Dec 2020 bis Mar 2025 unter der Anleitung von Dr. Chambers C. Hughes durchgeführt.

ACKNOWLEDGEMENTS

As I complete my doctoral studies at the University of Tübingen, I would like to take this opportunity to express my sincere gratitude to all those who have supported me throughout this journey. The completion of this thesis would not have been possible without the guidance, encouragement, and assistance of many people around me.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Chambers C Hughes, for giving me the opportunity to pursue my doctoral research in the group I am especially grateful for your support and patience throughout my PhD journey. You are an exceptional chemist, from whom I learned how to begin research in natural products and gained not only invaluable knowledge, but also a broad scientific perspective. Beyond academia, you have been an inspiring supervisor and a wonderful person to work with. Chambers, I will always cherish and never forget the days when we discussed my projects and the conversations we shared about culture. Thank you for your constant encouragement and for giving me the freedom and trust to grow independently in my work.

I would also like to express my sincere thanks to Dr. Julia Moschny, who guided me at the beginning of my PhD project. Your help during the initial stage of my research was extremely important. Thank you for teaching me the relevant techniques and helping me build the foundation that allowed me to continue independently after your departure from the group. In addition, thank you for the delicious and truly memorable Donauwelle cakes you brought, which always brought a little extra joy to our days.

My sincere thanks also go to all members of the Hughes group and the Microbial Bioactive Compounds theme of IMIT for creating such a supportive and friendly working atmosphere. I feel truly fortunate to have had them as my colleagues throughout my phD journey. In particular, I would like to thank Sehee Jang, Marita Wurm, Dominika Gorniaková for the many helpful discussions, technical guidance, constant encouragement, and warm companionship over the years. Your kindness and support have brought me strength and comfort in ways I will always cherish. Working alongside you has made my PhD experience not only enjoyable, but also deeply meaningful and truly unforgettable. I would also like to thank Dr. Annika Esch, Dr. Giovanni Andrea Vitale, Luca Salvi, Max Knab, Annika Schulz, and Andreas Kulik for their helpful discussions, technical advice and daily help in the department.

In addition, I am deeply thankful to all my collaboration partners: Prof. Dr. Yvonne Mast, Alina Zimmermann, Dr. Juan Pablo Gomez-Escribano, Dr. Daniel Petras for their generous support, constructive suggestions, and inspiring discussions throughout this project. Working with you has been a truly rewarding experience. And I also extend my thanks to the members of the University of Tübingen and NMR department for measuring numerous my sample efficiently.

I also would like to thank my supervisor Prof. Dr. Nadine Ziemert and Prof. Dr. Heike Brötz-Oesterhelt, it is not possible for me to start my project without their support. Especially, I would

like to thank my supervisor Prof. Dr. Nadine Ziemert for the support and guidance provided before I arrived in Germany and during my first days here. I would also like to express my sincere gratitude to the China Scholarship Council for providing financial support for my doctoral studies and for the opportunity to study abroad.

In addition, I would like to extend my heartfelt thanks to all my friends in Tübingen (Ying Zhang, Xuefeng Yang, Menglei Zhou, Ying Xu, Xuanheng Hu, Ningna Li, Hao Xiong, Qingjun Jiang, Gaopeng Li, Pengfei Liu, Xin Yan, Xiaohang Dai, Jingyun Ding, Wenhua Sun…), who kept me going and gave me the courage to face challenges during moments of self-doubt and discouragement. You made my past years outside the lab truly joyful and meaningful. I am especially grateful to those who supported me during the corona time. It was an incredibly difficult time, but because of all of you, my life felt truly meaningful. Thank you for standing by me when I needed it most. I will never forget your kindness and support.

I would also like to thank my parents and all family members for their unconditional love, endless support, and constant encouragement. Their belief in me has been the foundation of everything I have achieved. I am also profoundly grateful to my partner, whose understanding, patience, and unwavering support have brought me strength and comfort throughout this journey.

Finally, I would like to thank myself for holding on, for persevering through all the difficulties, and for never giving up.

SUMMARY

The continuous emergence of antibiotic-resistant pathogens and the urgent need for novel antibacterial agents require innovative strategies for the discovery of new natural products. Actinobacteria have long been recognized as prolific producers of bioactive secondary metabolites; however, their vast chemical diversity remains far from fully explored. Expanding and facilitating the discovery of novel natural products from actinobacterial sources therefore represents an important challenge in contemporary drug discovery research.

In this thesis, novel natural products from actinobacteria of the DSMZ and Tübingen strain collections were investigated with the objective of expanding accessible chemical diversity for antibacterial drug discovery. A genome mining approach based on the phosphoenolpyruvate mutase (pepM) gene was applied to 940 actinomycete genomes, allowing the identification of putative phosphonate producers and the assessment of their phylogenetic distribution. Subsequently, 28 pepM-positive strains were cultivated under diverse growth conditions, and their metabolite profiles were analyzed by NMR spectroscopy. By combining bioinformatic analyses, targeted gene deletion, heterologous expression, and NMR-based structure elucidation, the minimal biosynthetic gene clusters required for phosphonate production were identified in *Kitasatospora fiedleri* DSM 114396 and *Streptomyces iranensis* DSM 41954. To enable the isolation and characterization of phosphonate compounds, several enrichment and purification strategies were developed, including chemical labeling reactions as well as size exclusion chromatography, hydrophilic interaction liquid chromatography (HILIC), and ion-exchange techniques. In addition, a multiplexed chemical metabolomics workflow (MCheM) employing post-column derivatization reactions was established for non-targeted LC-MS/MS analysis. Application of this workflow resulted in the isolation and structural characterization of 7-glycosyl oxazolomycin D from *Streptomyces libani subsp. rufus* DSM 41230, representing the first reported glycosylated member of the oxazolomycin family.

Furthermore, the metabolic potential of *Streptomyces aureocirculatus* DSM 40386 was explored, leading to the discovery of piperazic acid-containing peptides. Their mode of action was investigated using a bioreporter-based assay. Moreover, the presence of a brominated compound was detected, indicating additional unexplored chemical diversity, although its full characterization remains pending.

In summary, the results presented in this thesis demonstrate that the integration of genome mining, advanced metabolomics workflows and targeted biosynthetic investigations constitutes a robust strategy for the discovery and characterization of previously inaccessible natural products from actinobacteria, thereby providing a strong basis for future antibacterial lead discovery.

ZUSAMMENFASSUNG

Das ständige Auftreten antibiotikaresistenter Krankheitserreger und der dringende Bedarf an neuartigen antibakteriellen Wirkstoffen erfordern innovative Strategien zur Entdeckung neuer Naturstoffe. Actinobakterien sind seit langem als produktive Erzeuger bioaktiver Sekundärmetabolite bekannt; ihre enorme chemische Vielfalt ist jedoch noch bei weitem nicht vollständig erforscht. Die Erweiterung und Erleichterung der Entdeckung neuartiger Naturstoffe aus actinobakteriellen Quellen stellt daher eine wichtige Herausforderung in der zeitgenössischen Wirkstoffforschung dar.

In dieser Arbeit wurden neuartige Naturstoffe aus Actinobakterien der DSMZ- und Tübinger Stammsammlungen mit dem Ziel untersucht, die zugängliche chemische Vielfalt für die antibakterielle Wirkstoffforschung zu erweitern. Ein Genome-Mining-Ansatz basierend auf dem Phosphoenolpyruvat-Mutase-Gen (pepM) wurde auf 940 Actinomyceten-Genome angewandt, was die Identifizierung potenzieller Phosphonat-Produzenten und die Bewertung ihrer phylogenetischen Verteilung ermöglichte. Anschließend wurden 28 pepM-positive Stämme unter verschiedenen Wachstumsbedingungen kultiviert und ihre Metabolitenprofile mittels NMR-Spektroskopie analysiert. Durch die Kombination von bioinformatischen Analysen, gezielter Gendeletion, heterologer Expression und NMR-basierter Strukturaufklärung wurden die für die Phosphonatproduktion erforderlichen minimalen biosynthetischen Gencluster in *Kitasatospora fiedleri* DSM 114396 und *Streptomyces iranensis* DSM 41954 identifiziert. Um die Isolierung und Charakterisierung von Phosphonatverbindungen zu ermöglichen, wurden verschiedene Anreicherungs- und Reinigungsstrategien entwickelt, darunter Isotopenmarkierungsreaktionen sowie Größenausschlusschromatographie, Hydrophile Interaktions-Flüssigkeitschromatographie (HILIC) und Ionenaustauschtechniken. Zusätzlich wurde ein multiplexer chemischer Metabolomics-Workflow (MCheM) unter Verwendung von Post-Column-Derivatisierungsreaktionen für die nicht-zielgerichtete LC-MS/MS-Analyse etabliert. Die Anwendung dieses Workflows führte zur Isolierung und strukturellen Charakterisierung von 7-Glycosyloxazolomycin D aus *Streptomyces libani subsp. rufus* DSM 41230, was den ersten beschriebenen glycosylierten Vertreter der Oxazolomycin-Familie darstellt.

Darüber hinaus wurde das metabolische Potenzial von *Streptomyces aureocirculatus* DSM 40386 untersucht, was zur Entdeckung von Piperazinsäure-haltigen Peptiden führte. Ihr Wirkmechanismus wurde mithilfe eines Bioreporter-basierten Assays untersucht. Zudem wurde das Vorhandensein einer bromierten Verbindung nachgewiesen, was auf eine weitere unerschlossene chemische Vielfalt hindeutet, wenngleich deren vollständige Charakterisierung noch aussteht.

Zusammenfassend zeigen die in dieser Arbeit vorgestellten Ergebnisse, dass die Integration von Genome Mining, fortschrittlichen Metabolomics-Workflows und gezielten biosynthetischen Untersuchungen eine robuste Strategie zur Entdeckung und Charakterisierung bisher unzugänglicher Naturstoffe aus Actinobakterien darstellt und somit eine solide Grundlage für die zukünftige Entdeckung antibakterieller Leitstrukturen bietet.

LIST OF PUBLICATIONS

List of publications included in this thesis:

1. Zimmermann, A.; **Xia, S.-N.**; Moschny, J.; Gomez-Escribano, J. P.; Boldt, J.; Nübel, U.; Nouioui, I.; Krause, J.; Irle, M. K.; Metcalf, W. W.; et al. Expanding the actinomycetes landscape for phosphonate natural products through genome mining. *RSC Chemical Biology* **2026**, *7* (2), 298-312, 10.1039/D5CB00254K. DOI: 10.1039/D5CB00254K.

2. Vitale, G. A.; **Xia, S.-N.**; Dührkop, K.; Zare Shahneh, M. R.; Brötz-Oesterhelt, H.; Mast, Y.; Brungs, C.; Böcker, S.; Schmid, R.; Wang, M.; et al. Enhancing tandem mass spectrometry-based metabolite annotation with online chemical labeling. *Nature Communications* **2025**, *16* (1), 6911. DOI: 10.1038/s41467-025-61240-z.(Co-first author)

3. Gomez-Escribano, J. P.; Zimmermann, A.; **Xia, S.-N.**; Döppner, M.; Moschny, J.; Hughes, C. C.; Mast, Y. Application of a replicative targetable vector system for difficult-to-manipulate streptomycetes. *Applied Microbiology and Biotechnology* **2025**, *109* (1), 89. DOI: 10.1007/s00253-025-13477-3.

4. Nouioui, I.; Zimmermann, A.; Hennrich, O.; **Xia, S.**; Rössler, O.; Makitrynskyy, R.; Pablo Gomez-Escribano, J.; Pötter, G.; Jando, M.; Döppner, M.; et al. Challenging old microbiological treasures for natural compound biosynthesis capacity. *Frontiers in Bioengineering and Biotechnology* **2024**, *12*, Original Research. DOI: 10.3389/fbioe.2024.1255151.

List of publications not included in this thesis:

1. Stincone, P.; Pakkir Shah, A. K.; Schmid, R.; Graves, L. G.; Lambidis, S. P.; Torres, R. R.; **Xia, S.-N.**; Minda, V.; Aron, A. T.; Wang, M.; et al. Evaluation of Data-Dependent MS/MS Acquisition Parameters for Non-Targeted Metabolomics and Molecular Networking of Environmental Samples: Focus on the Q Exactive Platform. *Analytical Chemistry* **2023**, *95* (34), 12673-12682. DOI: 10.1021/acs.analchem.3c01202.

2. Zhang, L.; Esquembre, L. A.; **Xia, S.-N.**; Oesterhelt, F.; Hughes, C. C.; Brötz-Oesterhelt, H.; Teufel, R. Antibacterial Synnepyrroles from Human-Associated Nocardiopsis sp. Show Protonophore Activity and Disrupt the Bacterial Cytoplasmic Membrane. *ACS Chemical Biology* **2022**, *17* (10), 2836-2848. DOI: 10.1021/acschembio.2c00460.

Table of Contents

1 INTRODUCTION

1.1 Microbial-derived natural products

Natural products (NPs) are organic molecules generated through the biosynthetic processes of living organisms. They constitute an essential and invaluable source of inspiration for the discovery and development of novel therapeutics to address threats to human health, particularly in the field of drug discovery, including antibiotics, anticancer agents, immune-suppressants and so on.[1]Previous statistical analysis revealed that more than half of small molecule medicines approved by the U.S. Food and Drug Administration (FDA) between January 1981 and September 2019, were developed either directly or indirectly based on NPs, or mimicked their key pharmacophoric elements, particularly those sourced from plants, animals and microbes.[2,3]Among them, microbial-derived NPs have attracted increasing interest because of their immense structural diversity, unique mechanisms of action, and broad spectrum of biological properties. Around 5000 of compounds from microbes had been recognize, and the numbers are still rising.[4]

Microbes, also called microorganisms, are not only the oldest life presence on earth, but also prolific producers of bioactive secondary metabolites, which all have revolutionized human healthcare. For instance, penicillin, discovered in the crude extract of *Penicillum rubens* in 1928, is a famous and important beta-lactam antibiotic used for the treatment of a broad spectrum of infections caused by Gram-positive bacteria. The other representative example is clinic medicine anticancer doxorubicin, isolated from *streptomyces*. However, the golden era (1940s to 1960s) of the discovery of bioactive NPs especially antibiotics had ended due to repeatedly isolation of known compounds (**Figure 1**). In addition, other issues further speeded up this decline, including challenges associated with the stability, formulation, and resupply of structurally complex molecules; the limited compatibility of NPs with high-throughput screening approaches; and the tendency for screening libraries to compounds with reduced stereo-chemical complexity.[5]Although the gap between the huge research investment and the limitation of novel skeletons of NPs have prompted many pharmaceutical companies to chemically synthesize bioactive molecules instead of isolation of NPs from microbial extracts, the microbial-derived NPs still can provide a new insight to bio-mimicked synthesis. Furthermore, in most cases, NPs exhibit superior antibacterial activity compared with synthetic screening hits, which is not surprising given that antibiotic-producing strains have undergone long term co-evolution with other organisms while competing with same ecological niche for more than millions of years.[6]In recent years, the intersection of synthetic molecules bottlenecks and the rising incidence of drug-resistant pathogens, as well as technological advances and cognitive innovations, has persuaded pharmaceutical companies to redirect their focus towards NPs discovery field. A representative example is the 2014 announcement by Sanofi and Fraunhofer institute of the establishment of a NP Centre of Excellence aimed at the discovery and development of novel antibacterial compounds.[7]

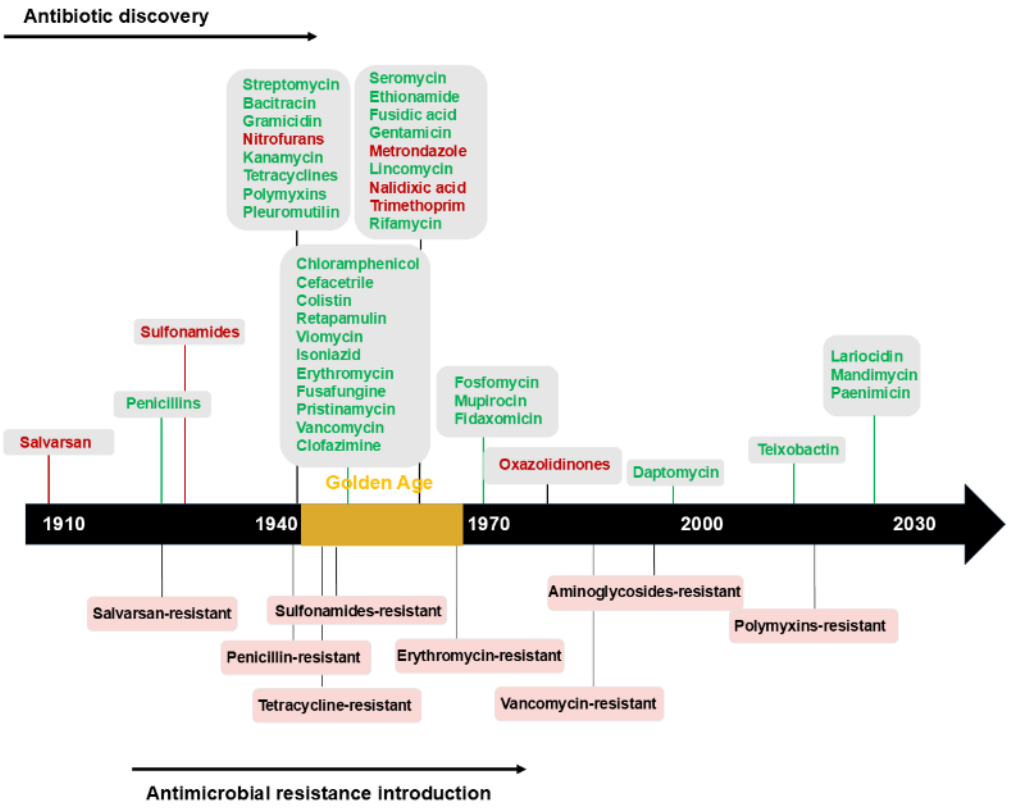


Figure 1. Timeline of antibiotic discovery and antimicrobial resistance development. The upper part shows the key compounds representing each identified antibiotic class. NPs are highlighted in green, while synthetic antibiotics are marked in red. The lower part shows the main antimicrobial resistance development after clinical introduction.

The capacity of microbes in the biosynthesis of NPs varies across taxonomic levels, leading to significant chemical diversity. This diversity is mainly attributed to the presence of biosynthetic gene clusters (BGCs), which encode the enzymatic pathways responsible for the formation of NPs. Approximately 30 BGCs are harbored in one microbial genome, each of which is theoretically capable of directing the production of a unique compound. However, experimentally identified NPs are disproportionately less than predicted based on BGCs. According to the computational genomic analysis, Gavriilidou *et al.* estimated that only 3% of the NPs potentially encoded within bacterial genomes have been experimentally characterized, a limitation that is mainly attributed to the transcriptional silence of numerous BGCs.[8] Consequently, the ecological significance and biotechnological value of microbes was underestimated due to their underexplored biosynthetic potential. Furthermore, nearly 99% of microbial species cannot be cultivated under standard laboratory methods, which significantly constrains access to their NPs repertoire. Thus, one way to uncover novel NPs is exploring unculturable microbes, which is well exemplified by the discovery of antibiotic teixobactin through *in situ* cultivation and product isolation.[9] Apart from that, exploring previously inaccessible microbial groups and their unique ecological niche also a valuable avenue for identifying novel bioactive compounds, as represented by recent studies where a new

antibiotic Darobactin was discovered from nematode symbionts *Photorhabdus*,[10]and Lugdunin, a first fibupeptide antibiotic against microbial infections like MRSA, was produced by nasal *Staphylococcus lugdunensis* strains.[11,12]In the context of the global antimicrobial resistance crisis, the renewed focus on microbes-derived NPs is of particular interest, as it is poised to reaffirm their critical role in drug discovery by offering the most promising way for the development of urgently needed therapies.

1.2 Actinobacteria as main sources of microbial-derived natural products

Actinobacteria constitutes one of the most dominant and evolutionarily significant bacterial phyla, encompassing an extensive diversity of genera, including *Tropheryma*, *Propionibacterium*, *Micromonospora*, *Salinispora*, *Mycobacterium*, *Nocardia*, *Corynebacterium*, *Gordonia*, *Rhodococcus*, *Leifsonia*, *Kitasatospora*, *Bifidobacterium*, *Gardnerella*, *Streptomyces*, *Frankia*, *Thermobifida*, among others[13](**Figure 2A**). This taxonomic diversity reflects the broad ecological and functional roles played by actinobacteria across natural and engineered environments. Members of this phylum are distributed across diverse ecological niches, ranging from terrestrial to aquatic ecosystems, and are particularly abundant in soil, where they play key roles in processes such as the decomposition of organic matter and the cycling of nutrients.[14,15]

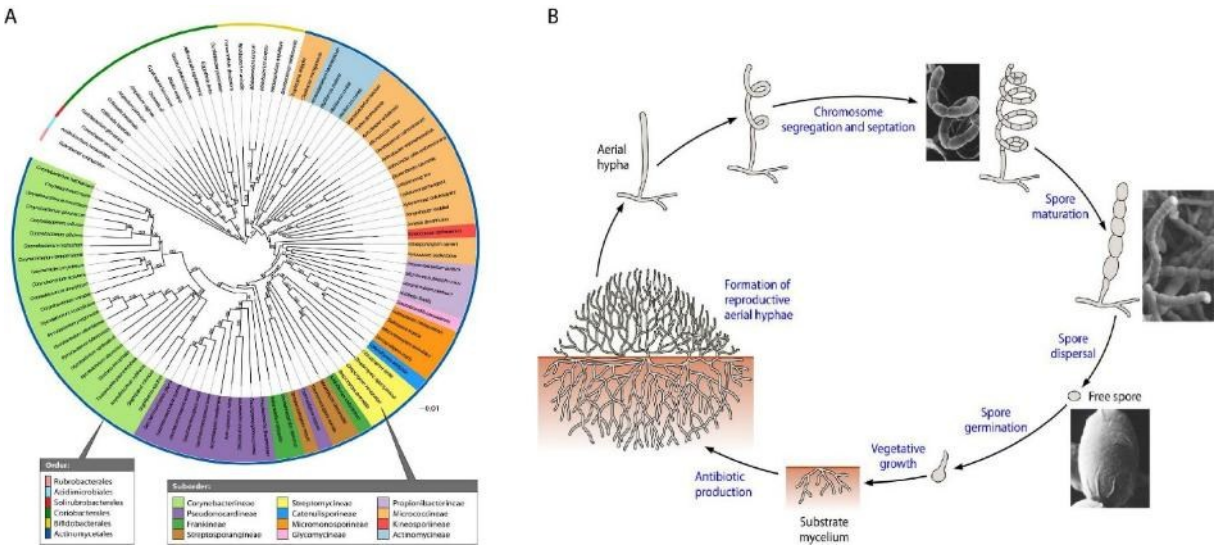


Figure 2. A genome-based phylogenetic tree based on 97 genome sequences of the phylum Actinobacteria(A) and schematic representation of the life cycle of sporulating actinomycetes(B). Modified with permission from Barka *et al*[13].

From a morphological perspective, actinobacterial growth and morphogenesis are characterized by a unique mode of hyphal tip elongation coupled with branching. This mode of development gives rise to filamentous structures resembling fungal hyphae, which enhance the ability of actinobacteria to effectively colonize and exploit complex substrates. (**Figure 2B**) Historically, actinobacteria were regarded as transitional organisms between fungi and

bacteria due to their fungal-like morphology and reproductive strategies. Like filamentous fungi, many actinobacteria form intricate networks of hyphae and reproduce through sporulation, a process commonly associated with fungal biology.[13]However, this superficial resemblance belies their fundamentally bacterial nature, as evidenced by several defining characteristics.

From the perspective of cell and molecule, actinobacteria exhibit hallmarks of bacterial organization. Their cells are thin and elongated, with chromosomes organized within prokaryotic nucleoids, distinguishing them from eukaryotic fungi, which possess membrane-bound nuclei. Additionally, the presence of peptidoglycan in their cell walls is a defining feature of bacteria, serving both structural and functional roles in maintaining cellular integrity. Another critical aspect underscoring their bacterial classification is their susceptibility to antimicrobial agents, a property that further differentiates them from fungi. This duality of fungal-like morphology and bacterial cellular organization underscores the unique evolutionary position of actinobacteria, highlighting their adaptability and ecological success.

Actinobacteria comprise a class of Gram-positive bacteria characterized by their high guanine-plus-cytosine (G+C) content in genomic DNA, often exceeding 60%.[13]This genomic composition is thought to confer stability to their DNA under environmental stress conditions, further contributing to their widespread distribution and ecological resilience. They are ubiquitous in terrestrial habitats, with a pronounced presence in soil, where they can be isolated from both surface layers and depths exceeding 2 meters. Their ability to thrive in diverse environments is attributed to their metabolic versatility and capacity for adapting to varying nutrients and oxygen availability.

Also, a significate feature of actinobacteria is their extraordinary potential to biosynthesize bioactive secondary metabolites. These compounds have profound ecological and biomedical significance, serving as mediators of microbial interactions and as the foundation for numerous therapeutics. Actinobacteria are particularly renowned for their role in the discovery and production of antibiotics, which have revolutionized modern medicine. Notably, they are responsible for synthesizing a wide array of antimicrobial drug classes.[16] Tetracycline, a polyketide broad-spectrum antibiotic with activity against a wide range of medically relevant Gram-positive and Gram-negative bacterial pathogens, is derived from *Streptomyces spp*. Similarly, erythromycin A, a representative macrolide antibiotic, has been widely used in clinical practice for the treatment of Gram-positive bacterial infections, was first isolated from *Saccharopolyspora erythraea*.[17] Another notable example is streptomycin, the first aminoglycoside antibiotic possessing broad antimicrobial properties, which was also produced by *Streptomyces griseus*.[18] For the management of infections associated with *Clostridium difficile* and other multidrug-resistant Gram-positive bacteria, daptomycin, a cyclic lipopeptide antibiotic, has emerged as a critical therapeutic agent. This compound is produced by *Streptomyces filamentosus*. Apart from their role as antibiotics, actinobacterial NPs have also contributed to multiple other fields of medicine. For instance, Doxorubicin (also called Adriamycin), an anthracycline-spectrum antitumor agent, is derived from *Streptomyces spp.* and remains widely used in clinical oncology. Additionally, certain terpene compounds like Lawsonone that isolated from *Streptomyces sp.* had shown potent immunomodulatory activity while polyether ionophores like Salinomycin from *Streptomyces albus* DSM 41398 had shown potent and selective anticancer activity toward cancer stem cells as well as cancer cell lines.[19,20,21]Even these examples highlight the unparalleled role of actinobacteria particularly

the genus *Streptomyces*, which is regarded as one of the most important producers of bioactive metabolites with therapeutic value, research on actinomycetes has still experienced a decline, largely as a consequence of the repeated isolation of already known metabolites, leading to the perception that these microorganisms have been overexploited.

Advances in biotechnological development and the post-genomic era have shown that the secondary metabolites BGCs in actinobacteria substantially exceeds conventional estimates. In addition, the characterized actinobacteria represent only a minor fraction in their global phylogenetic diversity, while a vast reservoir of taxa as mentioned before, particularly rare actinomycetes in inhabiting specialized ecological niches underexplored. To delineate the unknown chemical structure diversity of NPs, the prevailing strategies are mainly towards: 1. the optimization of cultivation parameters to mimic in situ environmental conditions, as represented by teixobactin and Lugdunomycin, two novel angucycline-derived compound were isolated from *Streptomyces sp.* QL37 by altering culture conditions;[22] 2. targeted the isolation of extremophilic strains from underexplored habitats. For example, Abyssomicin C, the first naturally occurring inhibitor of the *p*-aminobenzoate, was identified from the marine bacterium *Verrucosispora sp.* AB-18-032[23]and Schorn *et al.* revealed high potential of unique NPs biosynthetic gene clusters in rare marine actinomycetes;[24] 3. the activation of cryptic BGCs through synthetic biology approaches. Compound kutzneridine, a novel cyclic lipopeptide with antibacterial activities, was obtained by cloning the silent lipopeptide-encoding BGC from *Kutzneria sp.* and heterologous expressing in *Streptomyces coelicolor* M1152;[25] 4. the promoter engineering. Bauman *et al.* engineered and applied promotor cassettes to induce expression of the silent streptophenazine BGC in marine strain *Streptomyces sp.*, resulting in production of over 100 streptophenazines;[26] 5. combinatorial biosynthesis through modular reengineering of enzymatic domains. For example, Kaniusaite *et al.* redesigned peptide biosynthesis module *in vitro* from teicoplanin biosynthesis,[27]and and Yuzawa *et al.* engineered polyketide synthases and expressed the genes in *Streptomyces albus* to produce short-chain ketones.[28]

1.3 Approaches to the discovery of natural products in actinobacteria

1.3.1 Phenotypic screening-based approaches

Conventional phenotype-based screening, especially activity-guided isolation, is the most widespread method used in the discovery of new antimicrobial NPs, which essentially accounts for all the bioactive NPs identified in the golden era from microbes. In this approach, the pure cultures of actinobacteria are obtained usually from soil suspensions and fractionated, and the fractions then screened for the desired positive hit, which are fractionated further until a pure compound is isolated and identified with the desired biological activity. (**Figure 3**) This whole is usually carried out iteratively with different analytic methods.

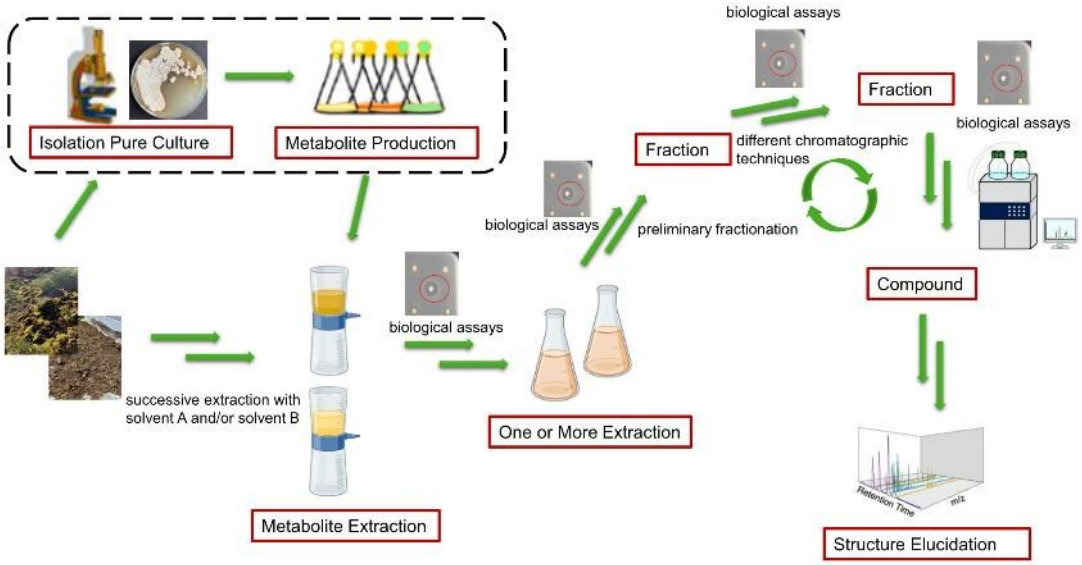


Figure 3. General workflow in activity-guided isolation procedures. Crude extracts are subjected to multiple rounds of extraction or fractionation, each guided by biological activity, until pure compound is obtained.

The advantage of activity-guided fractionation is that it can directly be isolated bioactive compounds with specific biological properties, that it is no surprise that it has its efficacy in the discovery of numerous antibiotics. And this activity-guided fractionation was main method to discover the bioactive NPs during the golden age of antibiotic discovery. The disadvantage of this approach is that it does not account for molecular structural features, leading to rediscovery of known compounds, and it also not able to access the full biosynthetic potential of microorganisms, not only can some compounds not be extracted using solvents like ethyl acetate, dichloromethane, butanol, but some taxa cannot be cultured using common methods. Even with the lack of a robust dereplication, it is still essential method for discovery of bioactive NPs, especially as a powerful and complementary method, for example Nothias *et al.* integrated bioactivity data with molecular networking to map the relationships between molecular structures and their biological activities, thereby rapidly identified bioactive compounds in complex mixtures.[29]In addition, advance in high throughput screening and chromatographic techniques also have further enhanced its efficiency and effectiveness.[30]

1.3.2 Genome mining-based approaches

With the advance in high-throughput genome sequencing, the availability of thousands of sequences has driven a paradigm shift in actinobacterial NPs discovery. Unlike traditional random screening approaches for novel compounds, genome mining provides a powerful strategy for predicting the NP biosynthetic capacity of actinobacteria prior to labor intensive wet-lab characterization[31](**Figure 4**).

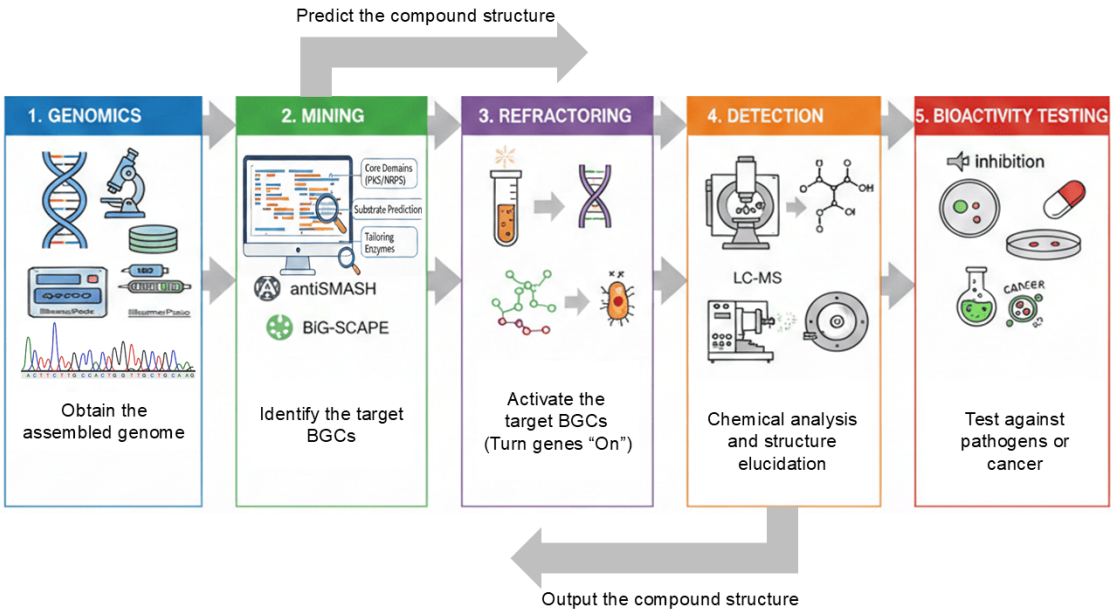


Figure 4. Workflow of genome mining-based natural products discovery. This approach starts from the genomic sequence of actinobacteria, which are used to identify target BGCs of putative NPs. The target BGCs are then activated using different tools, which is followed by compound purification, structure elucidation and bioactivity testing.

In addition, NPs sharing the same skeletons possess the similar BGCs, allowing for the discovery of structurally similar derivatives by means of genome mining, as exemplified by the isolation of lobosamides and related polyene macrolactams.[32] The discovery of NPs in actinobacteria can be shifted towards bioinformative analysis of the predicted BGCs encoded within the genome to expand chemical diversity.

1.3.2.1 Genes encoding core biosynthetic enzyme-based mining

Although NPs are typically generated through diverse biosynthetic reactions catalyzed by multiple specialized enzymes, their structural features and biosynthetic logics can be classified into several major groups that can be predicted through sequence similarity, including polyketides, nonribosomal peptides(NRPs), terpenoids, alkaloids, phosphonates, ribosomally synthesized and post-translationally modified peptides (RiPPs), and other hybrid forms.[33]

1.3.2.1.1 Polyketides

Polyketides are a structurally diverse family of NPs that exhibit a wide range of important biological activities, such as antibiotic erythromycin, immunosuppressant rapamycin, anticancer drug epothilone, all of which are biosynthesized by multifunctional enzyme complexes known as polyketide synthase (PKSs). Similar to fatty acids, polyketide biosynthesis begins with repeated Claisen condensation reactions between an activated acyl starter unit and malonyl-CoA-derived extender units, followed by modifications such as cyclization, oxidation, halogenation, alkylation, glycosylation etc., to form final complex polyketide compounds.[34]The whole process requires different domains including a β-

ketosynthase (KS), an acyl carrier protein (ACP), acyltransferase (AT), and one or more optional ketoreductase(KR), dehydratase(DH) and/or enoylreductase(ER). Typically, each domain catalyzes a specific reaction subsequentially during the formation of polyketides, which means the growing polyketide chain is initially attached to KS during each round of chain elongation. First, the ACP domain is converted into its active form through phosphopantetheinylation catalyzed by a 4'-Phosphopantetheinyl Transferase (PPTase), followed by a transformation of the malonyl group from malonyl-CoA to the ACP domain, which catalyzed by the AT domain. Second, the transferred extender unit is incorporated into the growing polyketide chain by the KS domain. In certain cases, the β-keto group may then be stereoselectively reduced to a β-hydroxyl group by the KR domain, followed by dehydration from an adjacent α-hydrogen catalyzed by the DH domain to yield an α, β-unsaturated intermediate. At last, the mature polyketide product is released from the ACP through hydrolysis of the thioester bond, a step mediated by Thioesterase (TE). Apart from that, the TE domain can also facilitate the intramolecular macrocyclization to generate lactones or lactams. Dan *et al.* recently presented an approach that replaces TE domain with terminal thioreductase, allowing the biosynthesis of diols, hydroxy acids and amino alcohols in *Streptomyces albus*,[35] and further widen the chemistry space in PKSs usage. From current knowledge of biosynthetic machinery architectures, PKSs are generally divided into three major classes in bacteria: Type I, Type II, and Type III PKSs.[36] Type I PKSs can be further subdivided into non-iterative and iterative form from a mechanistic perspective.

Within the non-iterative type I PKSs, two distinct subclasses are recognized: *cis*-AT PKSs and *trans*-AT PKSs, while enzymatic domains in *cis*-AT PKSs are incorporated in a co-linearity logic, such as KS-AT-DH-KR-ACP or KS-AT-KR-ACP. In addition, multiple *cis*-AT PKSs can be connected by intermodular linkers, forming a multi-modular PKS system, which can be exemplified by the biosynthesis of 6-deoxyerythronolide B, which is a key precursor in the formation of erythromycin A. This compound is synthesized by 6-Deoxyerythronolide B synthase (DEBS), which includes three proteins: DEBS1, DEBS2, and DEBS3[37](**Figure 5**). Each protein contains two modules, which executes a cycle of chain elongation and subsequent reduction independently.

Whereas all domains in the *trans*-AT PKSs are distributed across different polypeptide chains and assemble into functional complexes through specific interactions, such as docking domains. For example, the KS domain may be located on one polypeptide chain, while the AT and ACP domains are located on others. Furthermore, *trans*-AT PKSs are characterized by a discrete AT domain that transfers extender units onto the ACP domain.[38]*Trans*-AT PKSs have been identified across diverse bacterial taxa, including numerous lineages that remain largely unexplored in terms of their chemical diversity. Helfrich *et al.* developed a phylogenomic algorithm transPACT, providing comprehensive evidence that the diversification of *trans*-AT PKSs is widespread and associated with the replacement, insertion, or deletion of conserved PKS module blocks.[39]Apart from that, a key architectural distinction between the two systems lies in their genetic and functional organization. *Cis*-AT PKSs typically exhibit a colinear relationship between their genetic encoding and module arrangement, meaning the sequential order of modules on the DNA level directly corresponds to their operational sequence in the biosynthetic pathway. In contrast, *trans*-AT PKSs often display a noncolinear architecture,

where the genetic order of modules does not strictly align with their functional order, reflecting a more complex and evolutionarily adaptable system.

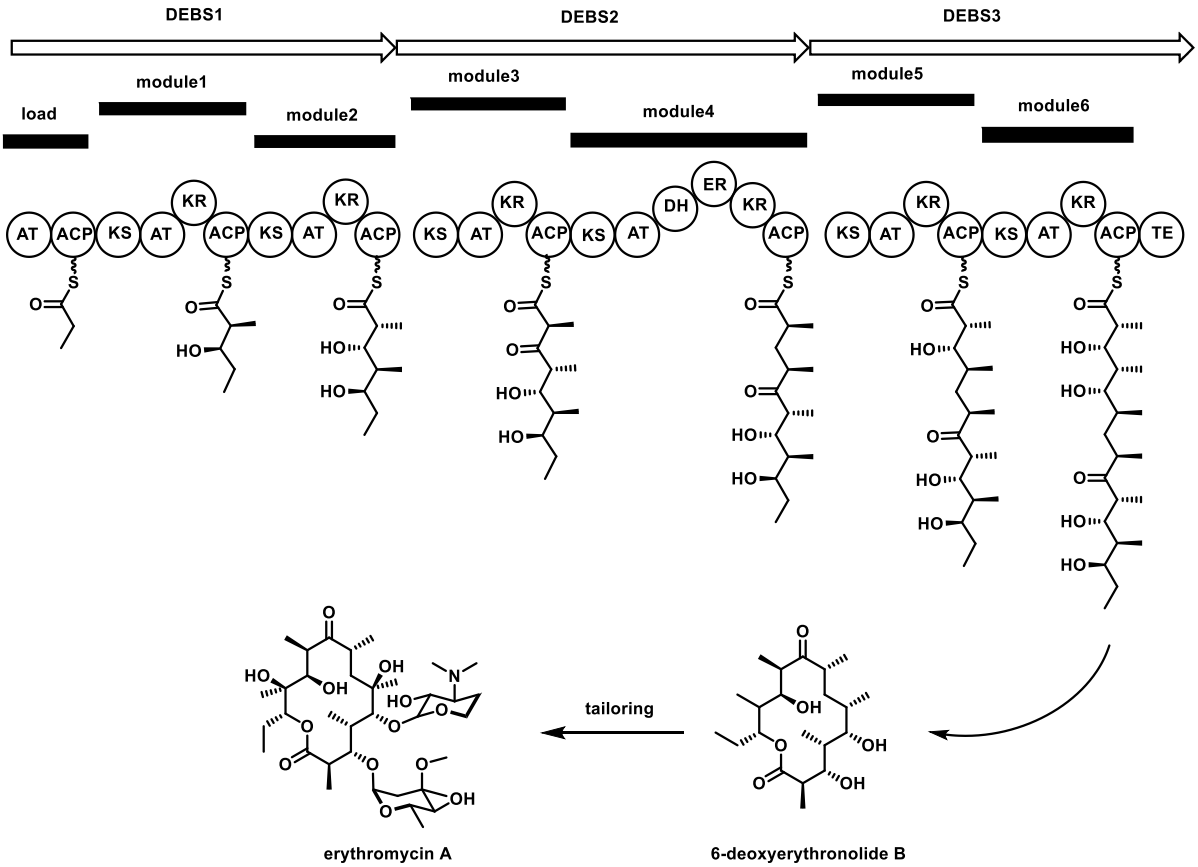


Figure 5. Example of polyketide biosynthesis model. The 6-deoxyerythronolide B was assembled on 6-deoxyerythronolide B synthase (DEBS). Erythromycin A is then produced via subsequent tailoring modifications.

Apart from the “canonical” modular PKSs,[40]iterative Type I PKSs are widely involved in highly reduced, partially reduced, or nonreduced polyketides, and are regarded as a hallmark of polyketide biosynthesis in fungi. Representative examples include aflatoxin B1, a DNA-alkylating secondary metabolite with prohepatocarcinogenic properties from *Aspergillus spp.*, whose biosynthesis is initiated by PksA,[41]and 6-methylsalicylic acid, which is assembled by 6-methylsalicylic acid synthase (6-MSAS) in *Penicillium patulum*.[42,43] Iterative type I PKSs are usually constituted by single modular that catalyzes intermediate chain elongation and elaboration.[44] Iterative monomodular type I PKSs were initially considered to be exclusive to fungi. However, Wang *et al.* discovered that they are unexpectedly abundant in *Streptomyces* by combining phylogenetic analysis and experimental verification.[45] Additional insights into architecture and evolutionary classify iterative type I PKSs into different subgroups.[40]These include bacterial iterative type I PKSs involved in the biosynthesis of aromatic compounds, such as the orsellinic acid moieties found in avilamycin[46]and polyketomycin,[47,48]as well as polyunsaturated fatty acid synthases responsible for the formation of polyunsaturated fatty acids,[49]polycyclic tetramate macrolactam synthases in the biosynthesis of frontalamides and maltophilin,[50,51]enediyne synthase for biosynthesis of C-1027 and calicheamicin[52,53]and other trans-AT type I PKSs.[40,54]

Type II PKSs represent multienzyme complexes formed by individual monofunctional proteins arranged in a nonmodular biosynthetic architecture, which synthesize many clinically useful aromatic polycyclic polyketides, such as tetracyclines and anthracyclines. Unlike type I assembly-line PKSs or type III PKSs, Type II PKSs are centered on two core proteins ketosynthase α-subunit (KS_α), and ketosynthase β-subunit (KS_β), also called as chain length factor (CLF).These proteins form a heterodimer and function in concert with an ACP to produce a poly-β-keto intermediate, which is highly unstable and readily undergoes spontaneous cyclization. The growing polyketide chain thus requires an environment to exclude spontaneous intramolecular reactions, and the biosynthetic enzymes must fix the orientation of the intermediates in a specific way to facilitate region-selective reactions controlled by various modifying enzymes, such as a KR and a cyclase.[55]

Type III PKSs are homodimeric KS-based enzymes that catalyze polyketide formation independently of ACP.[56]Although they were historically regarded as plant specific enzymes, genomic studies have revealed that type III PKSs are broadly distributed in microorganisms.[36] Representative examples include ArsB and ArsC from *Azotobacter vinelandii*, reported by Funa *et al.*,[57]which are responsible for the alkylresorcinols and alkylpyrones biosynthesis, respectively, and contribute to cyst formation through the production of phenolic lipids. Similarly, BpsA in *Bacillus subtilis* is involved in the biosynthesis of several aliphatic polyketides, including triketide pyrones, tetraketide pyrones and alkylresorcinols.[58]Moreover, Milke *et al.* revealed 96 putative type III PKS genes implicated in alkylresorcinol biosynthesis among known members of the phylum *Planctomycetota*.[59]

1.3.2.1.2 Nonribosomally synthesized peptides

Differ to peptides and proteins produced by ribosomal synthase, many nonribosomally synthesized peptides (NRPs) feature broad structural and biological activity, serving as antitumors, antibiotics, or immunosuppressants clinically, such as vancomycin, pristinamycin, bacitracin, cyclochlorotine, tentoxin. NRPs are biosynthesized by nonribosomal peptide synthases (NRPSs), modular enzymatic assembly lines that activate and incorporate specific monomeric building blocks into peptide products.[60,61,62,63](**Figure 6A**) A minimal NRPS module generally comprises three core domains: condensation (C), adenylation (A), and thiolation (T)domain. T domain also is called as peptidyl carrier protein (PCP) domain. Each module is generally responsible for the incorporation of one amino acid into the growing peptide chain.[64,65]The A domain recognizes and activates the corresponding amino acid substrate, then activated amino acid is covalently loaded onto phosphopantheteine (PPant) arm of a PCP domain as a thioester intermediate.[66]Peptide bond formation is subsequently catalyzed by the C domain, which links adjacent PCP-bound intermediates during chain elongation. Furthermore, tailoring domains such as oxidation(Ox), reduction(R), heterocyclization(Cy), epimerization(E) and methylation(M) further provide the structural diversification of NRPs[63, 67]Lastly, the completed NRPs is released from the assembly line either by hydrolysis or by intramolecular cyclization under the action of thioesterase (TE) domain. Notably, NRPs/PKs associated BGCs are widely distributed across bacterial genomes, especially among members of the phyla *Proteobacteria*, *Firmicutes*, *Actinobacteria*, and *Cyanobacteria*.[68]For example

Royer *et al* revealed *Xanthomonas spp.* contains specific NRPs genes which have potential to synthesize novel NRPs with structural diversity.[69]

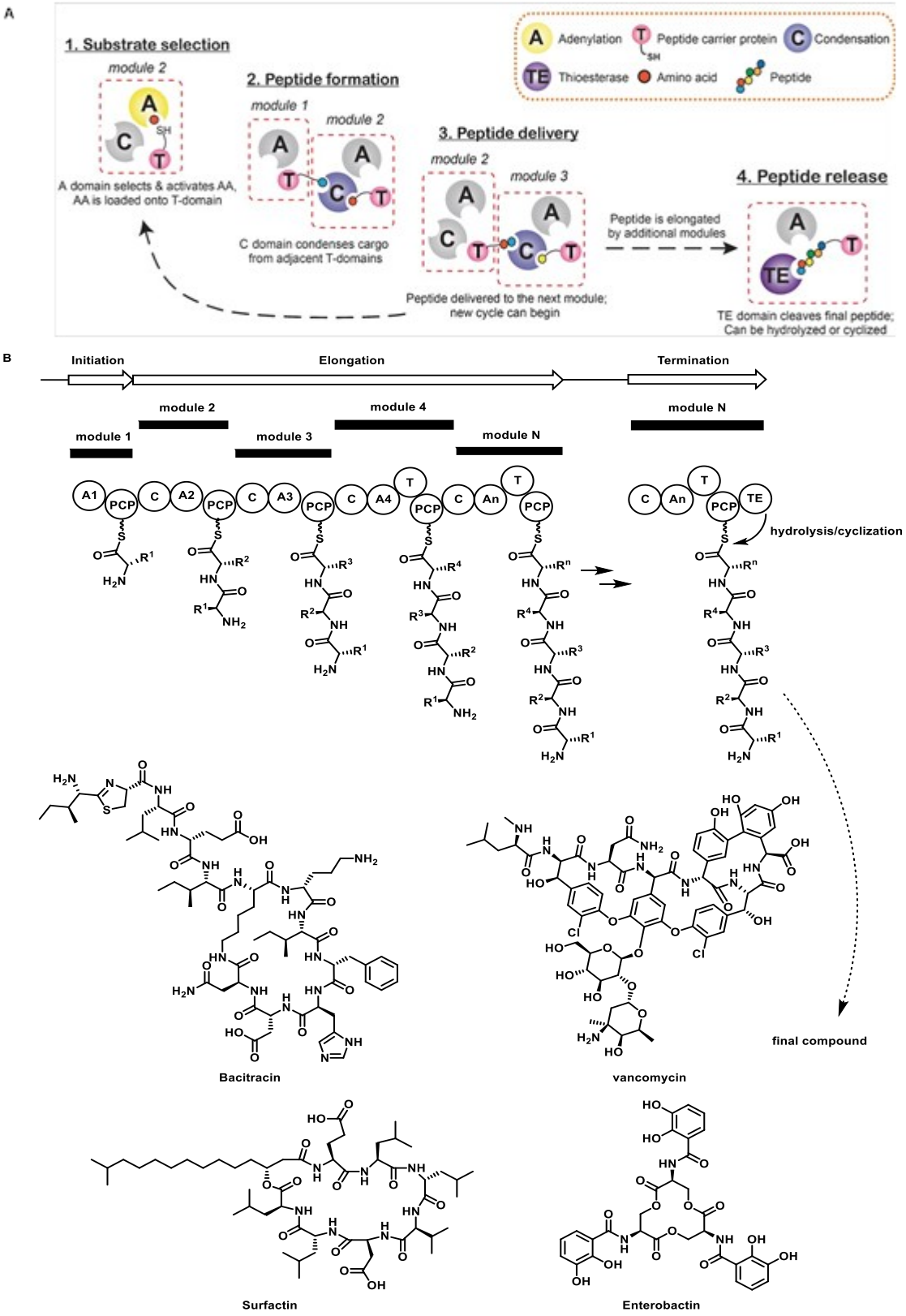


Figure 6. Biosynthetic model of NRPs(A) and chemical structure of examples (B). Adapted and modified with permission from Kittilä *et al*.[70]

Nevertheless, not all NRPs are synthesized modularly, for example BlmI type II PCP in the bleomycin-biosynthesis[71] and module/domain portable mechanism in pyrrolopyrazine biosynthesis.[72] A large scale genome mining analysis of 2699 bacterial genomes revealed more than half of BGCs associated with NRPs are nonmodular.[68]In addition, numerous hybrid biosynthetic systems comprising both NRPS and PKS modules have been reported.[66]

Although advance in genome mining tools like antiSMASH,[73]NRPSpredictor[74] and AdenPredictor,[75]the potential for large-scale discovery of NRPs has not yet been fully realized[76,77]due to a multitude of challenges such as noncanonical assembly lines, flexible A domains and tailoring domains. Behsaz *et al.*[77] recently presented a scalable modification-tolerant tool NRP miner, which designed to identify NRPs from large (meta) genomic and mass spectrometry datasets across different environments. This approach enables the detection of compounds produced by distinct non-canonical assembly lines, including surugamides, xenoinformycin, and lugdunin, as well as PAMs such as surfactins, arthrofactins, plipastatins, protegomycins, and PAX peptides. (**Figure 6B**)

1.3.2.1.3 Ribosomally synthesized and post-translationally modified peptides

Ribosomally synthesized and post-translationally modified peptides (RiPPs) have attracted extensive interest due to their antimicrobial activity and other therapeutic functions. Jangra *et al.*[78]recently reported Lariocidin, a first 30S ribosomal subunit-targeting lasso peptide that possesses broad-spectrum antibacterial activity against multiple bacterial pathogens, including *A. baumannii.* Xu *et al.* reported the discovery of Lexapeptide, a novel antibiotic with potent activity against Gram-positive bacteria, including MRSA and MRSE, through a bioactivity-guided functional genome mining strategy.[79]Richard *et al.* reported that Duramycin, a cyclic lantibiotic peptide, exhibits potent inhibitory activity against the entry of West Nile, dengue and Ebola viruses.[80]

The biosynthesis of RiPPs involves the stepwise processing of a precursor peptide that consists of both leader and core regions (**Figure 7A**). During this process, the core peptide undergoes a series of post-translational modifications to yield the final RiPP product, while the leader region mediates recognition of the precursor by corresponding tailoring enzymes and peptide export in many cases. To date, more than 40 RiPPs classes have been described such as Lanthipeptides, Cyanobactins, Lasso peptides, Glycocins, Microcins and so on, which categorized by common, class-defining post-translational modifications.[81,82]Due to the short sequence (often < 50 amino acids) of the precursor peptide, unclear annotation in genomes, and limited conserved features across class with varied tailoring enzymes, the discovery of novel RiPPs is a challenge.[83]Genome mining can overcome it, especially with increased knowledge about different RiPP biosynthetic mechanisms, exemplified by an interesting example the bacteria-derived tricyclic copper-binding lanthipeptides Noursin.[84] A putative class III lanthipeptide BGC (nor) was discovered in *Streptomyces noursei* ATCC 11455, whose precursor peptide NorA contains three Ser/Thr residues and one Cys residue. Heterologous expression of pNOR in *S. lividans*, combined with *in vitro* reconstitution experiments, revealed that a subgroup of lanthipeptide synthetases, named LanKCHbt, catalyzes the formations of both labionin and histidinobutyrine crosslinks in precursor peptides to yield noursin-like

compounds.[84] Another notable example is Daptides.[85]Using RRE-Finder and RODED in a class-independent manner, Ren *et al.* identified an unusual gene set associated with precursor peptides displaying atypical amino acid compositions. Direct cloning and heterologous expression of a putative BGC from *Microbacterium paraoxydans* DSM 15019 subsequently revealed a new RiPPs class bearing an unusual C-terminal (*S*)-N_2,N_2-dimethyl-1,2-propanediamine (Dmp) modification.[85](**Figure 7B**)

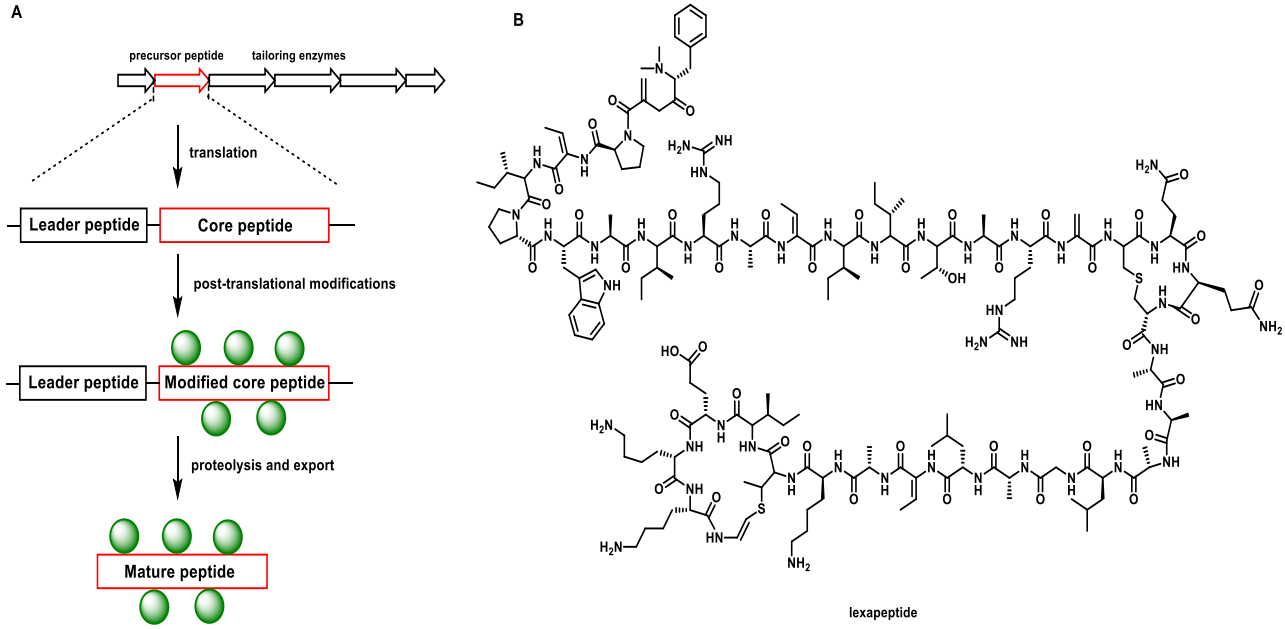


Figure 7. General biosynthetic scheme of ribosomally synthesized and post-translationally modified peptides (RiPPs, A) and chemical structure of lexapeptide (B).

1.3.2.1.4 Terpenes

Another a large family of NPs that has been studied well are terpenes, terpenoids, or isoprenoids, which are synthesized by terpene synthases. Terpene synthases can be broadly categorized into class I and class II enzymes. Class I terpene synthases contain the conserved DDxxD/E and NSE/DTE metal-binding motifs, which enable diphosphate activation following coordination with a trinuclear Mg^{2+} cluster, whereas class II enzyme rely on a conserved DXDD motif to initiate catalysis through protonation of the substrate.[86,87]The terpene precursors are assembled from the universal five carbon units dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP). These building blocks derive from two distinct pathways, the mevalonate (MVA) and the methylerythritol (MEP) pathway (**Figure 8**). The MVA pathway is mainly found in eukaryotes and archaea, while the MEP pathway occurs in most bacteria. Notably, plants possess both a cytosolic MVA pathway and a plastidial MEP pathway.
Terpenes have traditionally been regarded as metabolites primarily produced by plants and fungi. However, the genome mining work from Yamada et al[88] had uncovered a large number of terpene synthases in bacteria, in which 13 previously unreported cyclic terpenes were identified through the heterologous expression of several candidates terpene synthase genes.[89]

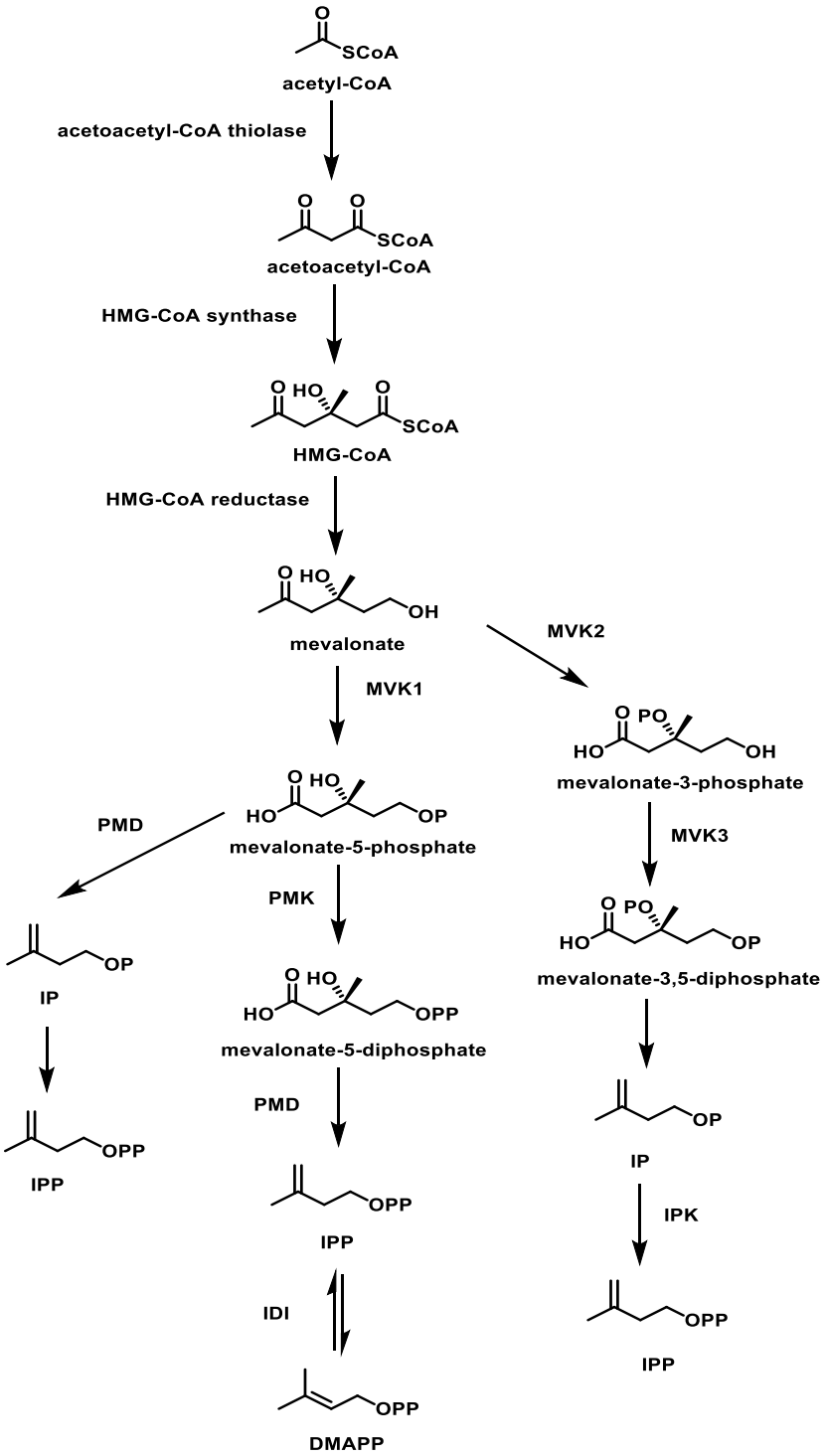


Figure 8. Mevalonate pathway towards IPP and DMAPP. MVK: mevalonate kinase, PMK: phosphomevalonate kinase, PMD: phosphomevalonate decarboxylase, IPK: isopentenyl pyrophosphate kinase, IDI: isopentenyl diphosphate isomerase.

1.3.2.1.5 Halogen-containing compounds

Halogen-containing compounds play an important role in the development of pharmaceuticals and agrochemicals due to their structural diversities and interesting bioactivities, such as anticancer candidate salinosporamides A, antifungal drug griseofulvin and antibiotic

aureomycin. Through their ability to participate in specific molecular interactions, halogens can enhance protein ligand recognition and consequently modulate the biological activity as well as the physicochemical characteristics of compounds.[90]A representative example is provided by halogenated antibiotic such as vancomycin,[91]chloramphenicol and balhimycin,[92]which exhibit stronger bioactivity than their dechlorinated derivatives, because of the effect of the chlorines for both the stability and the specificity of the targeted binding site. And halogenation of small natural molecular is carried out by halogenases, which can be divided into different groups according to the generation and utilization of activated halide: flavin-dependent halogenases, haloperoxidases, α-ketoglutarate-dependent halogenases and fluorinases.[93]As such, on the basis of the concept that structurally related NPs are typically biosynthesized by pathways harboring highly homologous enzymes, Hornung *et al.* used the conserved regions of six FADH2-dependent halogenases as marker sequences to examine 550 randomly selected actinomycete strains, resulting in the identification of 103 putative halogenase genes.[94]Heterologous expression of these candidates in *Streptomyces albus*, followed by HPLC-MS analysis, enabled the discovery of novel chlorinated xanthone antibiotic CBS40, which showed strong antibacterial activity against multiple Gram-positive bacteria and Gram-negative bacteria *E. coli* ATCC 10536.[94]Saha *et al.* mined four BGCs colocalized with lanthionine synthetase gene using SrpI, a flavin-dependent halogenase, as a query sequence, and demonstrated through heterologous expression and in vitro reconstitution that SrpI can catalyze the regioselective bromination of the indole moiety of a tryptophan residue in a macrocyclic peptide substrate.[95]This may provide a useful handle for the discovery of brominated Lanthipeptides. In addition, Dong *et al.* described 5'-fluoro-5'-deoxyadenosine synthase, the first native fluorination enzyme from *Streptomyces cattleya*, which can catalyze the formation of carbon-fluoride bonds.[96]Since the found of this fluorinase, Deng *et al.* employed genome mining and in vivo enzymatic assays to uncover three novel fluorometabolite BGCs with high sequence similarity from *Streptomyces sp* MA37, *Norcardia brasiliensis* and *Actinoplanes sp* N902-109.[97]This also means that the biosynthesis of fluorinated natural compounds still has a large chemical space waiting to be discovered.

1.3.2.2 Resistant gene-based mining

One effective strategy for the targeted discovery antimicrobial NPs is to mine BGCs based on self-resistant mechanisms encoded within microbial genomes.[98,99]NPs typically serve specific physiological roles that provide their producers with a competitive advantage within their ecological niches. However, while these NPs are effective in inhibiting competitors, they may also pose toxic risks to the producer themselves. To avoid this potential self-toxicity, producers that biosynthesize the NPs have evolved specialized resistance mechanisms for their self-protection. Such mechanisms include efflux pumps that actively transport metabolites out of the cell, exemplified by the LexABC efflux pump of *Lysobacter antibioticus* under the control of the transcriptional regulator LexR;[100]intracellular resistance proteins that sequester or inactive highly potent NPs; target proteins that allow essential housekeeping enzymes in the producing organism to escape inhibition.[99]From some examples like novobiocin,[101] platensin[102]has revealed that there often was the self-resistance gene, which directly encoded within the gene cluster. It enables the genome mining of NPs guided by resistant genes (**Figure**

9A). And microbial self-resistance mechanism could be used as a discriminating criterion for discovery of potential antibiotic producing strain, which was first demonstrated by Thaker et al. using this strategy, isolated pekiskomycin, a new glycopeptide antibacterials in 2013[103](**Figure 9B**).

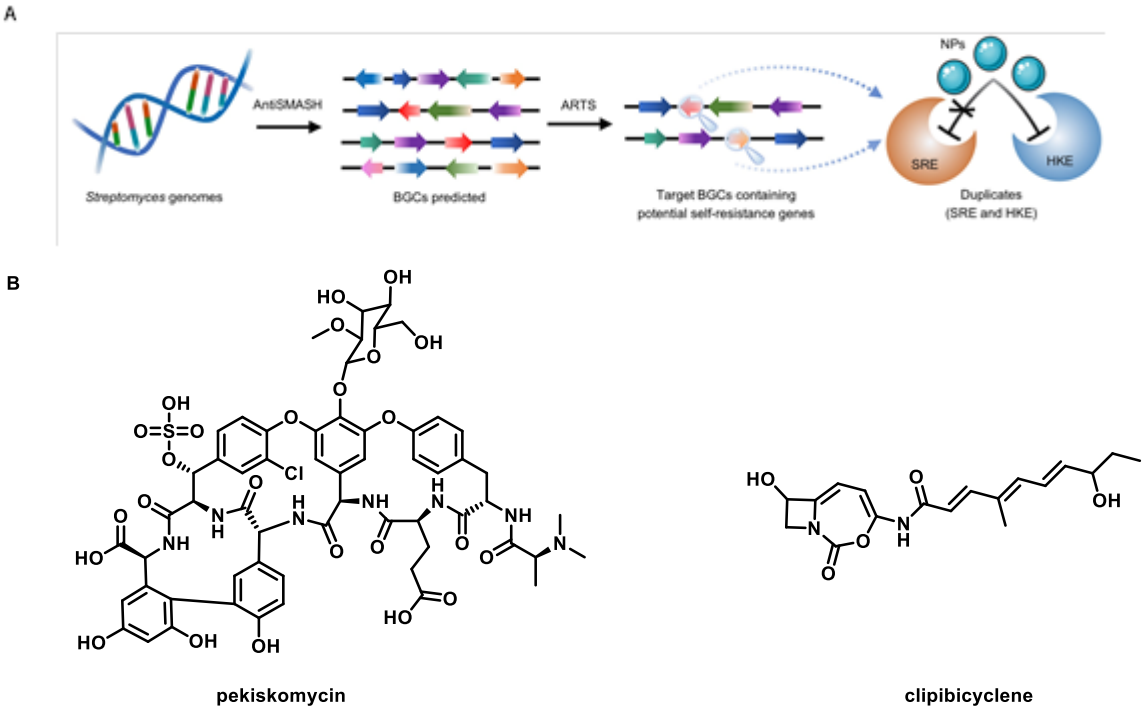


Figure 9. Discovery of target BGCs guided by self-resistance genes A) and examples of reported compounds B). Adapted and modified with permission from Yuan *et al*.[104]

Using a further target-directed genome mining approach, Tang et al. identified an unusual thiotetronic acid NPs in 2015 by screening for duplicated housekeeping genes associated with the orphan BGCs PKS4.[105]In order to facilitate to detect possible resistant housekeeping genes in actinobacterial genomes, Alanjary *et al.* introduced a web-based platform, Antibiotic Resistant Target Seeker (ARTS), designed for the rapid and efficient detection of duplicated essential gene within BGC-containing genomes, thereby enabling the prioritization of bacterial strains likely to encode novel antibiotics associated with resistant housekeeping genes.[106] Another interesting example concerns caseinolytic protease(ClpP) inhibitors.[107,108,109,110] ClpP is highly conserved protease found in bacteria as well as in eukaryotic mitochondria.[111] Organisms are killed when ClpP is inappropriately activated or inhibited. The mode of action of antibiotics acyldepsipeptides (ADEPs) is the most representative example.[107,112]Using this knowledge Culp *et al.*[113]used ClpP as a putative resistance marker to mine ClpP-associated clusters from genomes available in GenBank, leading to the prioritization of ten ClpP-associated gene cluster. Among them, one cluster cac, which contained bimodular NRPS with a type I polyketide synthase from *Streptomyces cattleya* DSM 46488 has attracted particular interest. Through heterologous expression in *S. coelicolor* and recombination, they identified a novel ClpP inhibitor, clipibicyclene, which can potently and selectively inactivate ClpP[113](**Figure 9B**). Nevertheless, only a limited number of bioactive NPs have been characterized from the vast reservoir of unexploited BGCs in actinobacteria. More recently,

Yuan *et al.* reported an automatically high-throughput platform, FAST-NPS,[104] which enables the efficient discovery of bioactive NPs produced from *Streptomyces* by integrating automated high-throughput direct cloning and heterologous expression of BGCs with self-resistance genes guided prioritization.

1.3.2.3 Phylogeny-based mining

The structural and functional diversity of NPs is thought to arise from the continuous evolution that involve different processes like duplication, rearrangement, transfer of genes or domains of microbial BGCs, enabling adaptation to ecological environments.[114]Therefore, investigating the evolution of gene and protein, along with their common ancestry and descent, provides valuable and fascinating insights into these complex evolutionary pathways. Phylogenetic tree construction is a widely displayed graphically approach to trace the evolutionary trajectories of specific genes and to distinguish the genes or proteins between orthologous and paralogous relationships (**Figure 10A**). With the increasing identification of biosynthetic enzymes and pathways, the genetic basis of NPs diversity has been elucidated through phylogeny-guided discovery strategies. This approach has facilitated the development of an evolutionary molecular network linking chemical structural diversity and novelty. Specifically, if a core gene of unknown function is phylogenetically distant from its known orthologs, its associated biosynthetic gene cluster may encode a novel compound with a unique structural backbone. Conversely, if a gene is closely related to others within an evolutionary branch, the products of the corresponding gene clusters are likely to exhibit high structural similarity.[114] For example, Kang *et al.* used short conserved β-ketosynthas (KS_β) sequences amplified from soil microbiomes as phylogenetic markers to establish correlation between genotype and chemotype, and thereby target the discovery of the novel pentangular polyphenols arixanthomycins A-C.[115]These compounds exhibited potent *in vitro* antiproliferative activity against colon cancer lines HCT-116 and WiDr and breast cancer cell line MDA-MB-231. Deng *et al.* constructed a phylogenetic tree using glycosyltransferase responsible for transferring mycosamine to the polyene macrolide scaffold, and used this enzyme as a sequence tag to identify an orphan clade not associated with any previously characterized polyene macrolide antibiotics[116] This approach led to the discovery of mandimycin, a novel polyene antifungal antibiotic from *Streptomyces netropsis* DSM 40259. Mandimycin displayed potent broad-spectrum antifungal activity against a range of multidrug-resistant fungal pathogens, including all *Candida* species listed on the WHO priority pathogen list for MDR fungi as well as *Aspergillus fumigatus* and *Cryptococcus neoformans*. It also exhibited a distinctive mechanism of action involving specific binding to multiple classes of phospholipids in fungal cell membranes. In another example, Yee *et al*.[117] identified a novel tRNA-dependent dipeptide synthase responsible for the biosynthesis of rare Arg-Xaa diketopiperazines by investigating a conserved fungal biosynthetic pathway (ank) that lacks a canonical core biosynthetic enzyme.

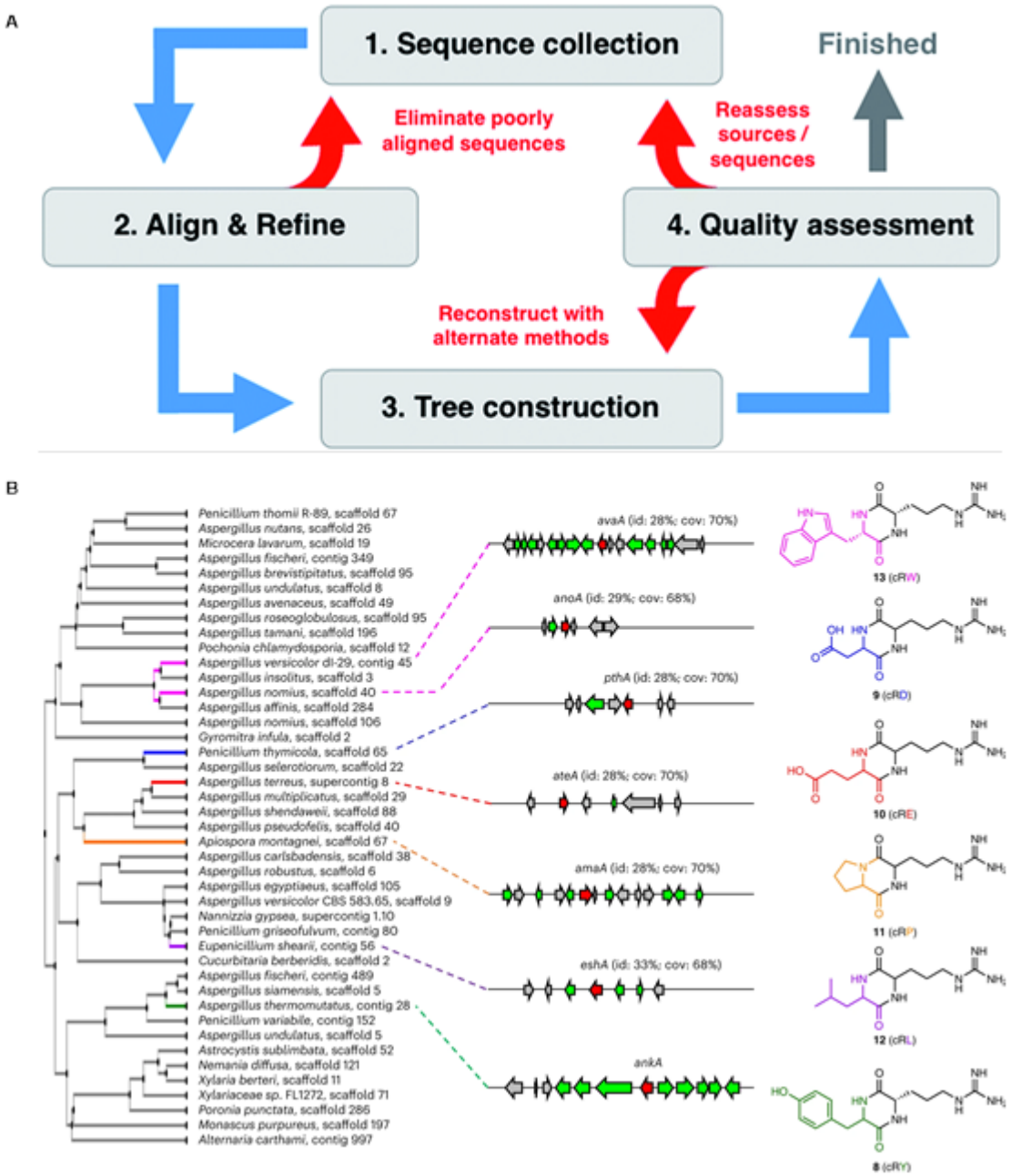


Figure 10. Workflow of construction of phylogenetic tree (A) and examples of reported compounds (B). Adapted and modified with permission from Adamek *et al.*[114] and Yee *et al.*[117]

1.3.3 Metabolomics-based approaches

Although genome mining offers a rational entry point for strain prioritization and for the targeted search of specific classes of NPs, genomic information alone does not demonstrate that a predicted metabolite is actually produced. The presence of a BGC does not necessarily correlate with pathway expression, metabolite accumulation, or detectability under laboratory conditions. Also, another main challenge encountered is determining how to prioritize molecules for novel compounds discovery from crude natural extracts. Progress in bioanalytical methodologies, coupled with expanding knowledge of microbial biosynthetic

systems and ecological evolution, has yielded fascinating molecular insights into these processes. Within this context, metabolomics has emerged as a relatively recent yet highly informative approach, which mainly used nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) to profile all small molecules involving metabolites, intermediates and products, which present in an organism and its lifecycle or under certain conditions and at a certain point in time, after genomics, transcriptomics and proteomics.[118,119,120,121]

1.3.3.1 MS-based metabolomics

Given the complexity of the NPs particularly unknown crude extract from microbes, MS can offer notable advantages such as exceptional sensitivity, rapid analytical performance, less sample consumption and wide application range as a useful and well-suited analysis technique in research on microbial NPs for structure elucidation.[122,123,124]It is typically coupled with high-performance liquid chromatography (HPLC) or gas chromatography (GC) to enhance detection coverage and facilitate high-throughput analysis of NPs, utilizing features such as exact mass, ion adducts and MS/MS fragmentation profiles.[125]However, a large amount of useful untargeted MS spectra was trapped in the private lab notebooks, databases or papers, even with some databases could be queried such as Dictionary of Natural Products, Coconuts, the natural products atlas,[126]and so on until Watrous et al. introduced molecular networking[127] in 2012. Molecular networking is a visual strategy for all of the chemical relationships of the ions of molecules that can be detected using MS.[128] Here, molecules with highly similar fragmentation pattern are calculated and grouped into molecular clusters or molecular families. In these clusters, each compound is visualized as a node, and the connections between them indicate their chemical relatedness. On the basis of molecular networking, Wang *et al.* established the Global Natural Products Social Molecular Networking (GNPS) platform in 2016.[129]GNPS is an open-access ecosystem for MS-based NPs analysis that supports untargeted metabolomics data in different fields through the Mass Spectrometry Interactive Virtual Environment (MassIVE) data repository.[130]By integrating MS/MS spectral information with clustering-based analysis, this platform enables online dereplication of complex raw MS data and accelerates the identification of potentially interesting compound.[129,131,132]Even with limited available MS spectra, the database of GNPS still is growing to make it serve as a powerful MS community. Also, the development of feature-based molecular networking (FBMN) provides enhanced analytical functions like distinguishing isomers producing similar MS/MS spectra, automated spectral annotation and statistical analysis.[133] The workflow of MS is **Figure 11.**

Notably, the performance and resulting data quality of MS-based analyses greatly depend on the ionization parameters chosen as well as the specific instrumentation used.[134]It means that samples that are difficult to ionize or unstable may be missed and poorly detected, leading thus to strong bias of structure identification, especially for novel or rare microbial products.

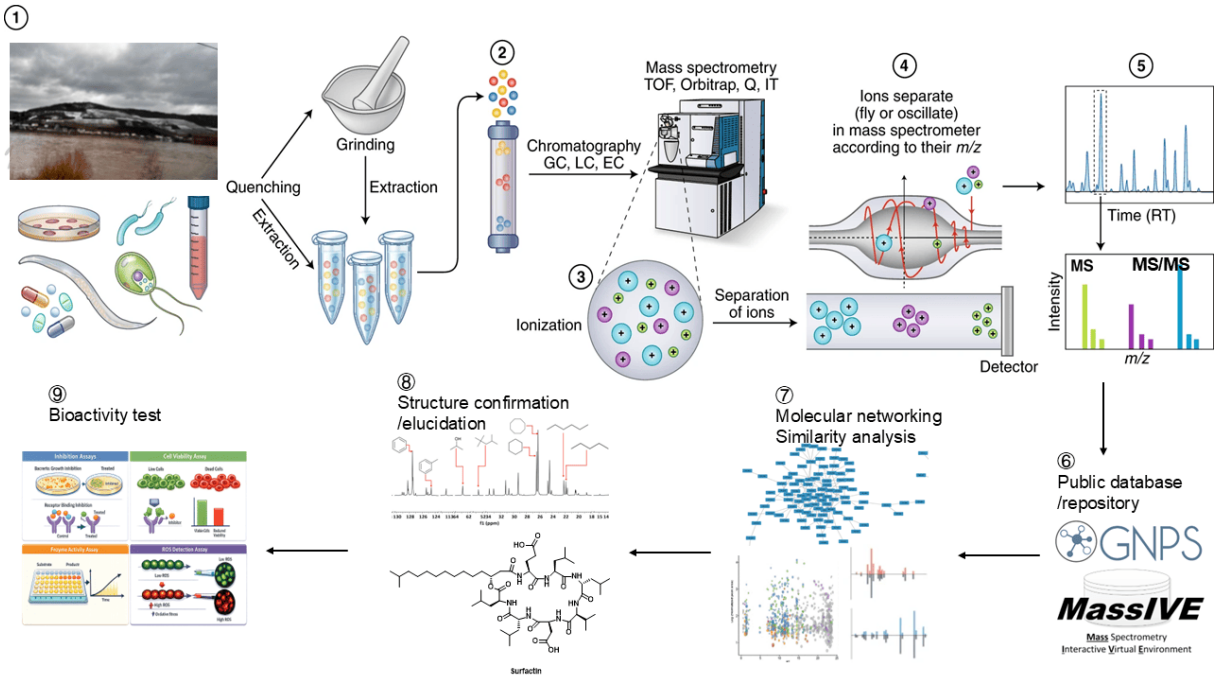


Figure 11. Workflow of MS based metabolomics. (1) sample preparation and extraction; (2) metabolite separation on a column (chromatography); (3) ionization of metabolites using an ion source; (4) separation by a mass analyzer as ions fly or oscillate; (5) detection of metabolites by combination of retention time and MS signature. (6) Data comparison in public database. (7) Construction of molecular networking and similarity for hits. (8) Structure elucidation. (9) Bioactivity test. Adapted and modified with permission from Alseekh *et al.*[125]

1.3.3.2 NMR-based metabolomics

Despite its lower sensitivity relative to MS, NMR continues to serve as an essential and highly informative tool for metabolome analysis and NPs discovery, largely due to its outstanding reproducibility, inherently quantitative nature, and unique capacity to provide unambiguous structural identification of previously unknown metabolite.[135,136]

In a typical NMR-based metabolomics workflow(**Figure 12**), extracted samples are analyzed using one-dimensional NMR like ^{1}H and ^{13}C NMR for determining and mapping the chemical environment of typical nuclei, as well as two-dimensional NMR such as COSY, HSQC, HMBC, TOCSY and NOESY for interpreting more detailed information about the chemical structure and the three-dimensional arrangement of molecules, followed by systematic spectral processing and multivariate statistical analysis. Variations among sample groups are translated into statistically significant spectral features, which can be traced back to specific metabolites or structural motifs. For example, stable isotopes such as ^{2}H, ^{13}C, and ^{15}N derived from labeled substrates can be followed through biochemical processes, providing a means to identify active metabolic pathways, and to quantify metabolic fluxes as well ae measure enzyme activity. Unlike mass spectrometry, NMR does not rely on ionization efficiency and is free from adduct formation and ion suppression effects, allowing a more unbiased representation of metabolite composition. In addition, and importantly, the direct relationship between signal intensity and molar concentration facilitates reliable relative quantification across samples.

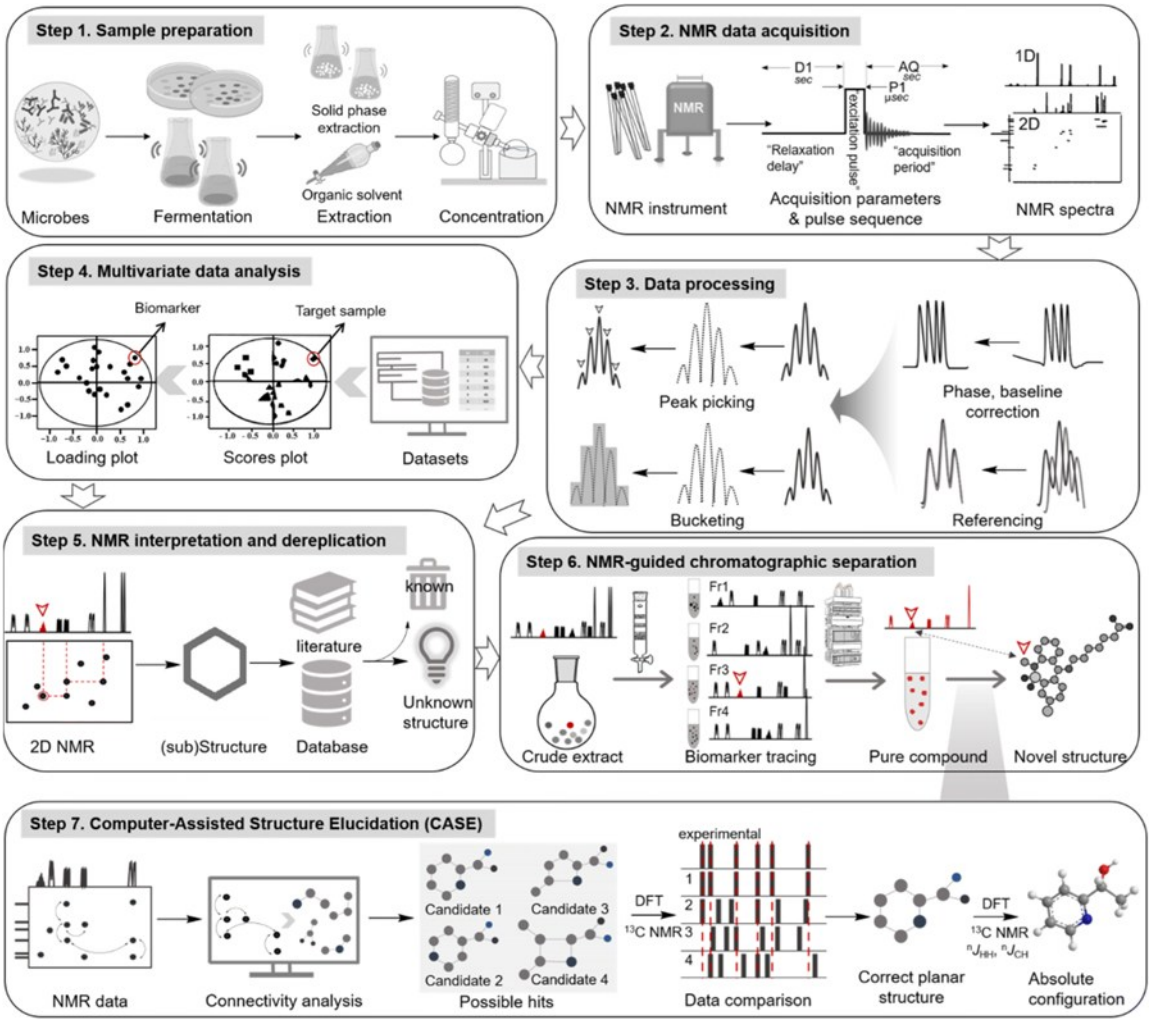


Figure 12. Workflow of NMR based metabolomics. Reproduced from Wang *et al*[137] with permission.

1.4 Aim and overview of this thesis

In this thesis, we aimed to identify novel NPs in actinobacteria from DSMZ and Tü collection for expanding and facilitating NPs chemical diversity discovery efforts that can potentially lead to new antibacterials. Although many antibiotics have been reported from the actinobacteria, its potential for chemical diversity remains largely unexplored.

To achieve this aim, we first turned our attention to phosphonate-containing compounds due to their nature and potential bioactivity. We analyzed and retrieved 940 actinomycetes genomes from the DSMZ and Tübingen strain collection using *pepM* gene, determined potential phosphonate producers and their phylogenetic distance. 28 pepM$^+$ strains were cultivated in different growth conditions. The extracts were analyzed using NMR. Bioinformatic analyses, gene deletion, heterologous expression and NMR defined the minimal and key gene cluster for phosphonate production from *Kitasatospora fiedleri* DSM 114396 and *Streptomyces iranensis* DSM41954. In addition, to isolate and characterize the phosphonate compounds from producers, we developed different labeling reactions and some attempts such as size exclusion, HILIC and ion exchange to obtain them, as described in chapter 2 of this thesis.

In chapter 3, we described a multiplexed chemical metabolomics workflow (MCheM) that leverages an array of post-column derivatization reactions for non-targeted LC-MS/MS analysis. By using this strategy, we isolated and characterized the first glycosylated member of the oxazolomycin family, 7-glycosyl oxazolomycin D, from *Streptomyces libani subsp. rufus* DSM 41230.
In chapter 4, we explored the potential of *S.aureocirculatus* DSM40386 and described the discovery of piperazic acid-containing peptides. Moreover, by analyzing the bioreporter we preliminarily characterized the mode of action for piperazic acid-containing peptides. More interestingly, we found that there is a brominated compound, but further characterization awaits completion.
In Chapter 5, we discuss all research efforts and address future perspectives.

2 DISCOVERY OF NOVEL PHOSPHONATE-CONTAINING NATURAL PRODUCTS USING GENOME-MINING IN BACTERIA

Part of sections 2.1.1, 2.1.2 and 2.1.5 in this chapter have already been published. All contents, including text, figures, and tables were reproduced and adapted with the permission of Zimmermann, A.; **Xia, S.-N.**; Moschny, J.; Gomez-Escribano, J. P.; Boldt, J.; Nübel, U.; Nouioui, I.; Krause, J.; Irle, M. K.; Metcalf, W. W.; et al. Expanding the actinomycetes landscape for phosphonate natural products through genome mining. *RSC Chemical Biology* **2025**, *7* (2), 298-312.See section 2.3.1

The author's contributions are as follows:
The study concept was conceived by Yvonne Mast and Chambers C. Hughes. *The genome mining of phosphonate BGCs for phosphonate producers was done by Alina Zimmermann, Juan Pablo Gomez-Escribano and Janina Krause. The methodology for screening phosphonate production of putative producers was designed and developed by close consultation between Julia Moschny and me. The cultivation, screening, extraction and analysis of all putative phosphonate producers from DSMZ was done by me. The genetic engineering of Kitasatospora fiedleri DSM114396 was done by Alina Zimmermann and Juan Pablo Gomez-Escribano. The phosphonate production validation of all resulting engineering strains was analyzed and completed jointly by Julia Moschny and me, including cultivation, screening, extraction and analysis. The initial manuscript was drafted by Alina Zimmermann, Juan Pablo Gomez-Escribano and Yvonne Mast. The data in this publication was curated by Alina Zimmermann, me, Juan Pablo Gomez-Escribano and Julia Moschny. All figures containing gene sequences were prepared by Alina Zimmermann and Juan Pablo Gomez-Escribano under supervision of Yvonne Mast. All figures containing ^{31}P NMR spectra were prepared by me under supervision of Chambers C. Hughes. William W. Metcalf gave scientific advice. All authors reviewed, edited and approved the final version of the manuscript.*

Part of section 2.1.3 in this chapter has already been published. All contents, including text, figures, and tables were reproduced and adapted with the permission of Nouioui, I.; Zimmermann, A.; Hennrich, O.; **Xia, S.**; Rössler, O.; Makitrynskyy, R.; Pablo Gomez-Escribano, J.; Pötter, G.; Jando, M.; Döppner, M.; *et al.* Challenging old microbiological treasures for

natural compound biosynthesis capacity. Frontiers in Bioengineering and Biotechnology **2024**, 12, Original Research. See section 2.3.2

The authors' contributions are as follows:
Imen Nouioui: conceptualization, data curation, formal analysis, investigation, supervision, writing–original draft, writing–review and editing, methodology, validation, and visualization. Alina Zimmermann: investigation, visualization, and writing–original draft. Oliver Hennrich: investigation and writing–original draft. Me: investigation, data generation, analysis and writing–original draft. Oona Rössler: investigation and writing–original draft. Roman Makitrynskyy: investigation, methodology, validation, visualization, writing–original draft, and supervision. Juan Pablo Gomez-Escribano: investigation, methodology, validation, writing–original draft, and supervision. Gabriele Pötter: writing–original draft and investigation. Marlen Jando: writing–original draft and investigation. Meike Döppner: writing–original draft, and investigation. Jacqueline Wolf: data curation, writing–review and editing, and investigation. Meina Neumann-Schaal: data curation, resources, writing–review and editing, and investigation. Chambers C. Hughes: investigation, supervision, writing–review and editing, data curation, resources, and visualization. Yvonne Mast: conceptualization, data curation, formal analysis, funding acquisition, investigation, project administration, resources, supervision, writing–original draft, writing–review and editing, methodology, and visualization.

Part of section 2.1.4 in this chapter have already been published. All contents, including text, figures, and tables were reproduced and adapted with the permission of Gomez-Escribano, J. P.; Zimmermann, A.; **Xia, S.-N.**; Döppner, M.; Moschny, J.; Hughes, C. C.; Mast, Y. Application of a replicative targetable vector system for difficult-to-manipulate streptomycetes. *Applied Microbiology and Biotechnology* **2025**, *109* (1), 89. See section 2.3.3

The authors' contributions are as follows:
The study concept was conceived by Yvonne Mast and Chambers C. Hughes. The related genetic engineering and the construction of vector pDS0007 experiments was conducted by Alina Zimmermann, Juan Pablo Gomez-Escribano and Meike Döppner. The phosphonate production validation of all resulting wildtype and engineering strains were conducted and completed by me, including cultivation, screening, extraction and analysis. Julia Moschny contributed new methods. All authors analyzed data. Juan Pablo Gomez-Escribano, Chambers C. Hughes and Yvonne Mast wrote the initial manuscript. All authors reviewed and contributed to the manuscript. All authors read and approved the manuscript.

The inappropriate and excessive use of antibiotics has accelerated the evolution of multidrug-resistant bacteria, particularly the ESKAPE pathogens, including *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa and Enterobacter spp.*,[138]GBD 2021 antimicrobial resistance collaborators estimated that drug-resistant infections were responsible for more than one million deaths annually from 1990 to 2021, and forecasted this may increase to 1.91 million attributable deaths and 8.22 million associated deaths by 2050,[139]thereby creating an urgent requirement for novel and highly efficacious antibiotics to tackle this emerging threat to public health. As described in Chapter 1, the expanding number of bioactive compounds and evidence to support that bacteria could still be most potential producer to produce novel natural products.

Phosphonates are a notable and underexploited class of NPs featuring carbon to phosphorus(C-P) bonds. This structural moiety, which enables potent molecular mimicry of essential metabolites (e.g., phosphate esters or carboxylates), is responsible for interesting and useful biological properties of these compounds, including antibacterial, antiparasitic, antiviral, insecticidal and herbicidal activities. For example, fosfomycin (also known as Monuril, Monurol, or Monural), is used clinically for the treatment of acute urinary tract infections.[140]By minicking phosphoenolpyruvate, it inhibits UDP-N-acetylglucosamine3-enolpyruvyltransferase and consequently blocks peptidoglycan biosynthesis. Notably, fosfomycin seems to be one of the most effective agents for combating infections caused by ESKAPE pathogens.[141] Fosmidomycin and its derivatives FR-900098, promisingly used to against malaria in human clinical trials, which acts through inhibition of 1-deoxy-d-xylulose 5-phosphate reductoisomerase and thereby blocks the non-mevalonate pathway of isoprenoid biosynthesis. In addition to antibacterial and antifungal activities, phosphinothricin, bialaphos and phosalacine, which acts as inhibitors of glutamine synthetase, mostly investigated as herbicides.[140,142]Even with these existing values and potential, less than 30 small molecule phosphonate-containing NPs have been reported thus far.[143]This is due to the distinctive characteristics of the phosphonate moiety that may have affected their detection and inclusion in standard NPs discovery, since most phosphonate compounds are highly polar, hydrophilic and lack strong UV absorption. Interestingly, insights into biosynthesis of all known phosphonates have revealed its genetic basis for C-P synthesis.[144] With the exception of K-26 and I5B2, all known phosphonates biosynthesis begin with the conversion of phosphoenolpyruvate (PEP) to phosphonopyruvate (PnPy) by phosphoenolpyruvate mutase (PepM)[144,143](**Figure 13**). Due to this conserved biosynthetic feature, which enables microbial genomes to be queried for the presence of pepM, allowing determination of their genetic capacity for biosynthesize phosphonates.[145,146]Using pepM gene as a molecular marker, Metcalf group developed a gene-based screening strategy and revealed that the approximately 5 % of randomly isolated actinomycetes possess the genetic capability to biosynthesize phosphonates.[147]

In this chapter, we use a genomics-guided approach to assess the distribution and diversity of phosphonate biosynthetic pathways in strains from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ), which further demonstrate the effectiveness of genome mining as a strategy for dereplicating known compounds and prioritizing interesting BGCs, and reveal many novel phosphonate biosynthetic potentials from actinomycetes. Subsequent culturing of all pepM$^+$ strains under OSMAC stratage identified potential phosphonate

producers, two of which were *Kitasatospora fiedleri* DSM 114396 and *Streptomyces iranensis* DSM41954 with unique gene clusters that prompted us to explore them further. And we tried to find the correct BGCs and use genetic engineering to increase phosphonate production so that further isolation and purification. Given the nature of phosphonate, we developed and optimized new methods that allow facile isolation of phosphonate.

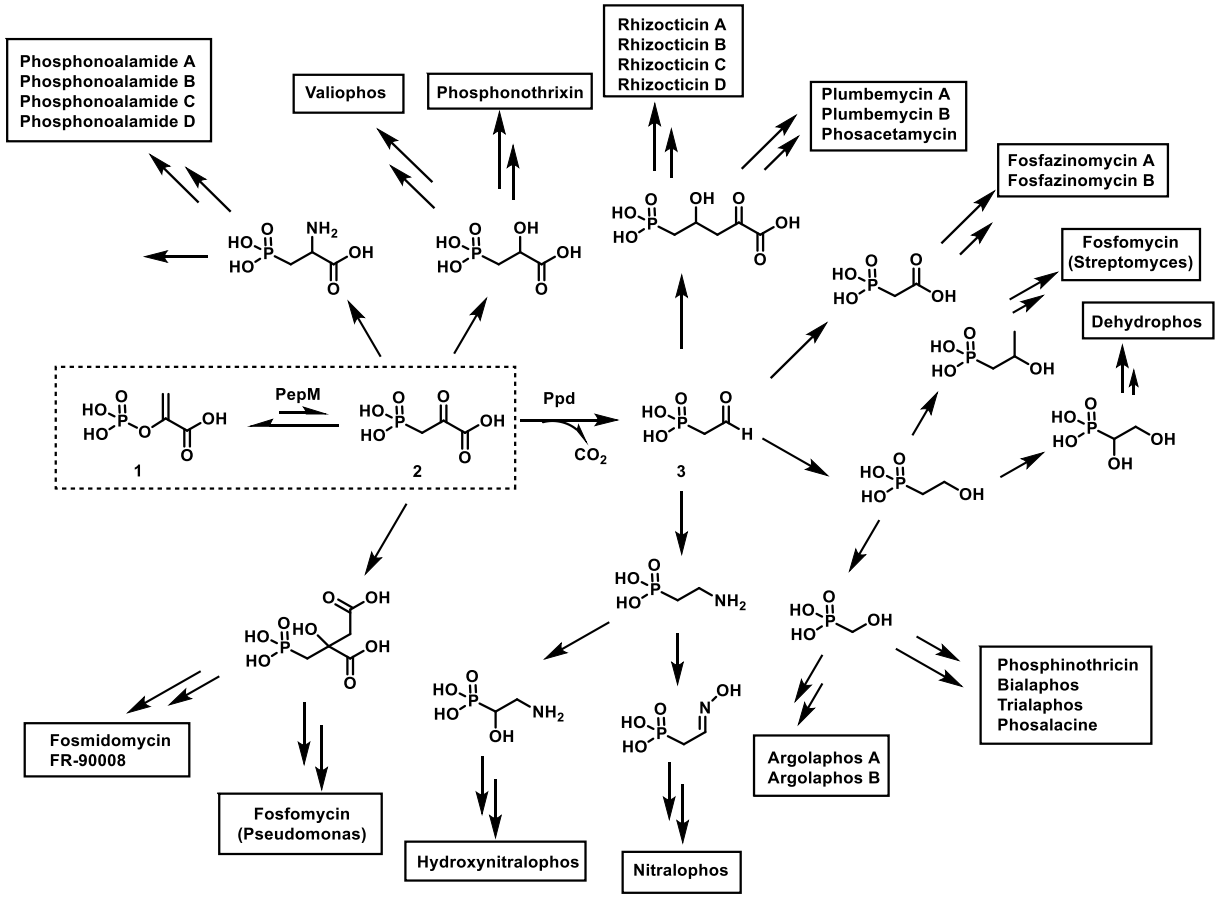


Figure 13. Overview of phosphonate biosynthesis.

2.1 Results and discussion

2.1.1 Genome mining reveals diversity of phosphonate biosynthetic pathways in DSMZ and Tü collection

The genome mining of phosphonate BGCs and genetic engineering has been done by Alina Zimmermann and Dr. Juan Pablo Gomez-Escribano at the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures in Braunschweig supervised by Prof. Dr. Yvonne Mast.

To identify potential phosphonate producers, we screened 940 actinomycetes genomes from the DSMZ (n = 781) and Tübingen (n = 159) strain collection using the pepM gene, as a

molecular marker. Due to the pepM gene is high conserved with an essential active site motif(EDKXXXXXNS),[147, 148]and also there is a strong correlation between PEP mutase amino acid sequences that encode phosphonate metabolic pathways and pepM gene neighborhood. Using this criterion, all putative phosphonate producers(pepM$^+$) were subsequently identified. To prioritize these pepM$^+$ srains, the phylogenetic tree of PepM amino acid sequences and cluster uniqueness was created and validated (**Figure 14**). This analysis not only classifies sequence diversiy but also provides a framework for estimating biosynthetic relatedness among strains.

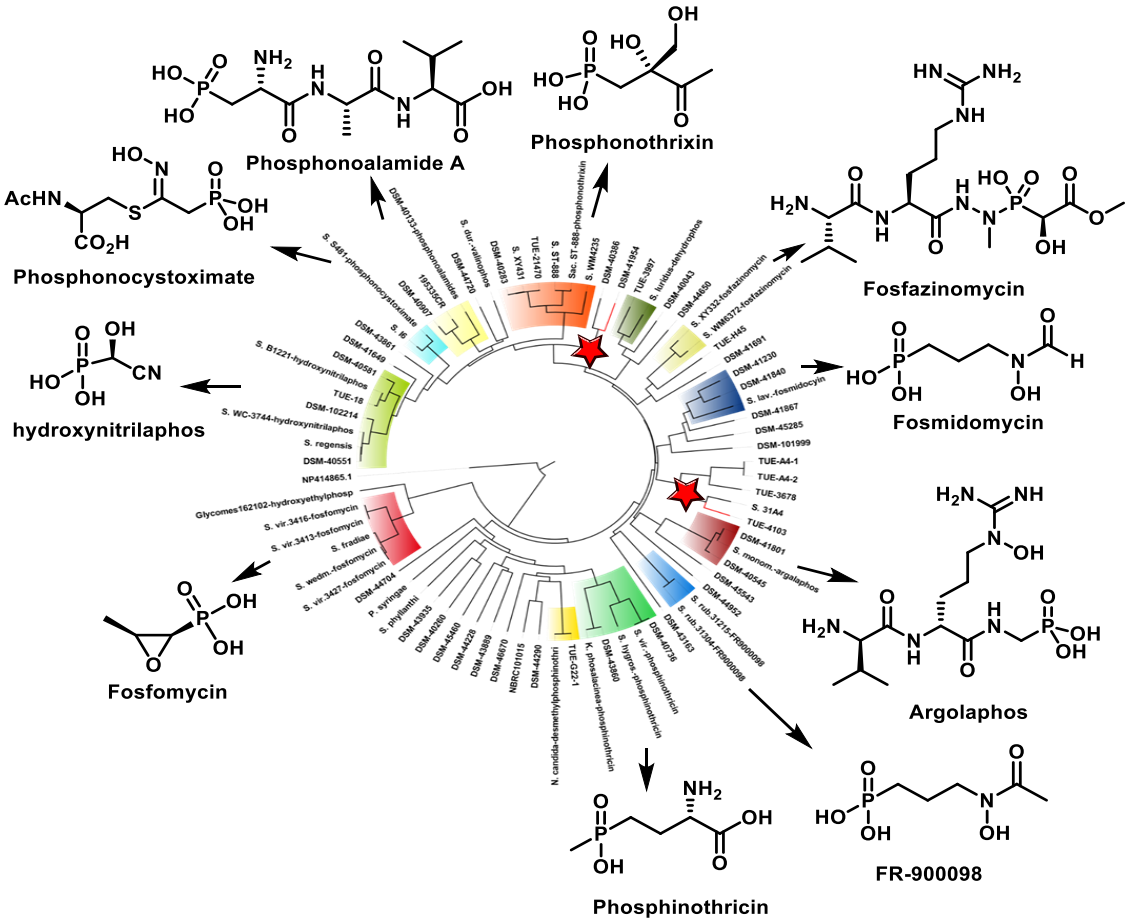


Figure 14. Maximum-likelihood phylogenetic tree of pepM amino acid sequences. Nodes corresponding to known phosphonate producers are color-coded, and the chemical structures of their known products are shown. Uncolored nodes represent putative phosphonate producers with unknown compounds. *K. fiedleri* DSM114396 and *S. iranensis* DSM41954 are highlighted with red stars.

Significantly, all phosphonate producers deposited in the DSMZ collection possess the potential to produce known phosphonate compounds, including producers of phosphonoalamides, argolaphos, fosmidomycin and hydroxynitrilaphos, as well as unknown phosphonate producers were discovered, validating the effectiveness and sensitivity of the approach based on PepM amino acid phylogeny. More importantly, phylogenetic reconstruction reveals a considerable extent of unexplored biosynthetic diversity (The simultaneous presence of color-coded nodes corresponding to characterized phosphonate producers and uncolored nodes representing strains for which no phosphonate compound has

yet been assigned, indicating that a substantial proportion of the observed PepM sequence space remains chemically uncharacterized.

Two of these strains, *Kitasatospora fiedleri* DSM114396 and *Streptomyces iranensis* DSM41954, were of particular interest first because the genomic information showed that their BGCs were unique and novel[149]. Nevertheless, the broader tree topology further suggests that phosphonate biosynthesis is widely distributed and evolutionarily diverse in DSMZ and Tü collection,[147] and there are still many strains carrying interesting gene clusters worth exploring, for example *Nocardia tenerifensis* DSM44704, *Goodfellowiella coeruleoviolacea* DSM43935, *Saccharopolyspora spinosa* DSM44228, which deserve further study.

2.1.2 Initial phosphonate production screening of pepM$^+$ strains

Although genome sequenced-based screening has proven effective in identifying potential phosphonate producers from the DSMZ collection, the actual production of phosphonate remains unknown. Therefore, it is crucial to figure out whether the pepM$^+$ strains could produce phosphonate. In contrast to many other NPs classes such as polyketides, NRPs, phosphonates have distinctive structural features that can be visualized by Nuclear Magnetic Resonance (NMR), which enables easy detection of phosphonate compounds when they are presented.[143] And the OSMAC approach is a classic and highly effective strategy for facilitating secondary metabolites by leveraging varied cultivation conditions to extend NPs chemical diversity. Therefore, we used semi-quantitative ^{31}P NMR spectroscopy and OSMAC method for linking genome sequenced-based screening results to phosphonate production(**Figure A 4-Figure A 18**). Initially, we examined the role of inorganic phosphate in phosphonate production for 28 pepM$^+$ strains.[150]

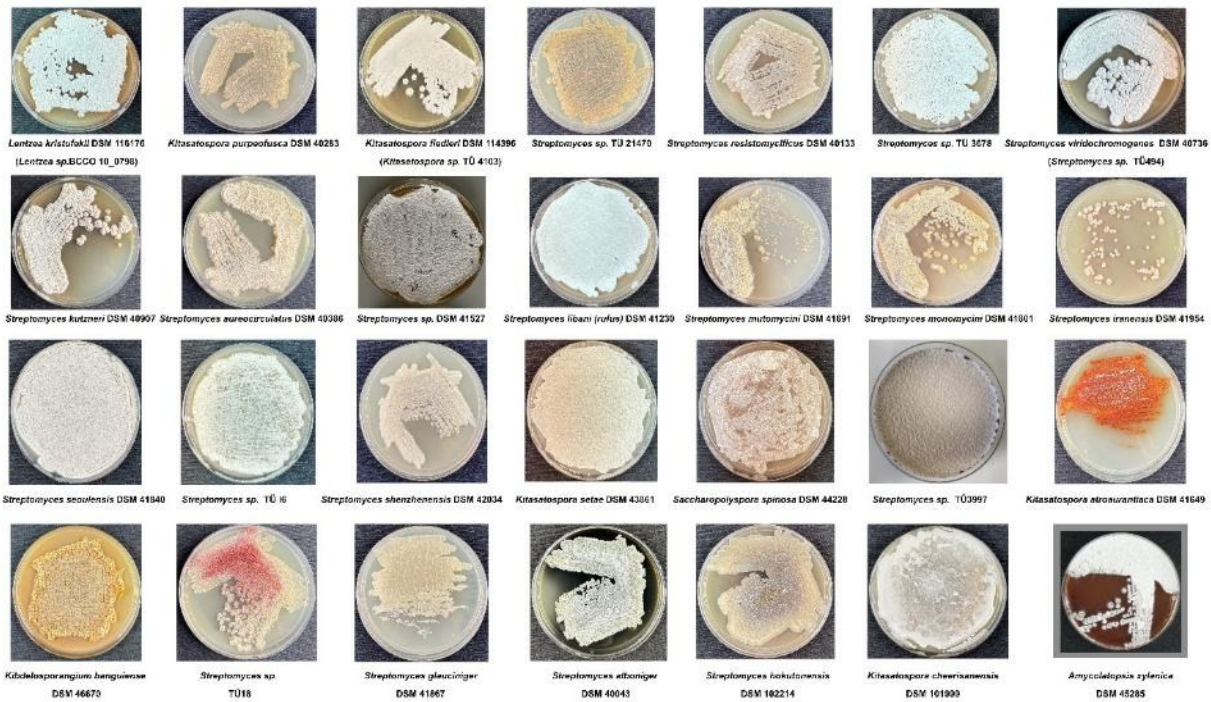


Figure 15. Images of all putative phosphonate producers in this study.

This analysis revealed inorganic phosphate is an essential precursor, providing the phosphorus-containing building blocks for phosphonate compounds. In addition, phosphonate production was influenced by the composition of carbon and nitrogen sources in the cultivation media. For example, under the addition of inorganic phosphate, the phosphonate compounds signals from *K.atroaurantiaca* DSM 41649 were detected only in spent OM and NL400 medium (**Figure A 13**). Similarly, *K.cheerisanensis* DSM 101999, the phosphonate compounds signals were observed only in spent HM and ISP4 medium with the addition of inorganic phosphate (**Figure A 7**). Likewise, the phosphonate compounds signal from *S. seoulensis* DSM41840 were detected only in spent HM and ISP4 medium with the addition of inorganic phosphate (**Figure A 8**). Furthermore, production screening of all 28 pepM$^+$ strains from the DSM and Tü collection by ^{31}P NMR spectroscopy and OSMAC method indicated that at least 4/5 produced phosphonates. Meanwhile, a tremendous wealth of phosphonate chemical diversity from DSM and Tü collection is yet to be explored.

2.1.3 *Streptomyces kutzneri* DSM40907 is a new phosphonoalamides producer

Even with the understanding of phosphonate biosynthetic capacity by genome mining and NMR,[145,146,151]it doesn't mean the phosphonate structure is clear. Therefore, linking phylogenetic placement with chemical output remains essential for validating biosynthetic predictions. From our phylogenetic tree, the two strains from DSM collection, *Streptomyces resistomycificus* DSM40133 and *Streptomyces kutzneri* DSM40907, were grouped into phosphonoalanine gene cluster family. Since *S.resistomycificus* DSM40133 was previously shown to produce phosphonoalamides based on comparative analysis of pepM genomic neighborhoods as reported by Kayrouz *et al.*[151,152]its close association with *S. kutzneri* DSM40907 suggested that the latter strain might encode a similar biosynthetic capability. Therefore, to further clearly verify the prediction of our phylogenetic tree, *S. kutzneri* DSM40907 was cultivated in ten different media to identify conditions that induced phosphonate production most effectively. Screening of the resulting extracts by ^{31}P NMR showed that phosphonate accumulation was highest after 7 days of growth in GUBC medium, indicating that metabolite production is strongly influenced by culture conditions (**Figure 16**). This analysis also highlights the importance of medium optimization when attempting to experimentally confirm cryptic or weakly expressed phosphonate pathways.
Subsequent LC-HRMS analysis of the culture supernatant detected several ions with masses and fragmentation features characteristic of phosphonoalamides. Specifically, the protonated molecular ions m/z 340.1270[M+H]$^+$, 370.1378 [M+H]$^+$, 354.1424[M+H]$^+$, and 368.1582 [M+H]$^+$, each of which matched the respective molecular formulas $C_{11}H_{23}N_3O_7P$, $C_{12}H_{25}N_3O_8P$, $C_{12}H_{25}N_3O_7P$, and $C_{13}H_{27}N_3O_7P$.[153]These data were consistent with previously reported MS/MS fragmentation patterns of phosphonoalamides, revealing that *S. kutzneri* DSM40907 is a new phosphonoalamides producer. And these findings not only expand the known distribution of phosphonoalamide biosynthesis within *Streptomyces*, but also further demonstrate that our phylogenetic analysis is functionally effective for uncovering novel phosphonate biosynthetic pathway.

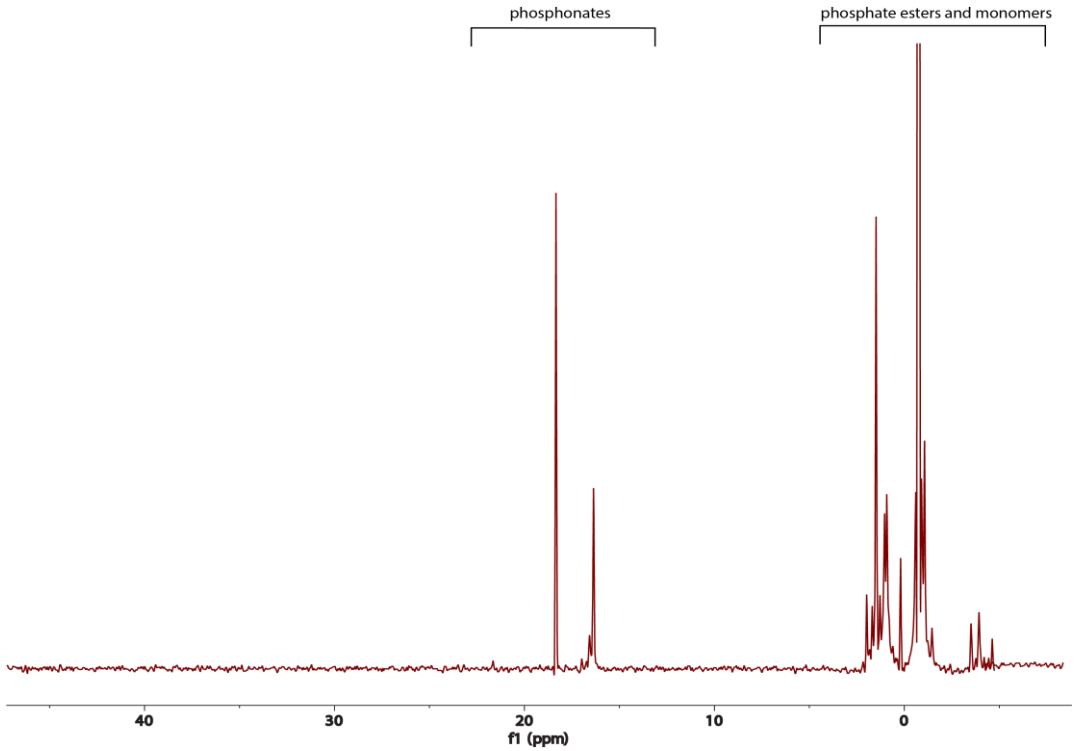


Figure 16. ^{31}P spectra of phosphonates from concentrated spent GUBC culture extract of *S. kutzneri* DSM40907. The P peaks with chemical shifts between 10 and 40 ppm are usually indicative of phosphonates and phosphinates, whereas P peaks between +10 to -20 ppm range are usually phosphate monomers, esters and pyrophosphates

2.1.4 Validating the phosphonate production in the *S. iranensis* DSM41954 based on a putative BGC using ^{31}P NMR analysis

As our phylogenetic tree shown, *S. iranensis* DSM41954 and *S. aureocirculatus* DSM 40386 was grouped together but without connection to any known phosphonate producers. Due to the uniqueness of BGC, *S. iranensis* DSM41954 was prioritised and investigated. The putative phosphonate biosynthetic gene cluster from antiSMASH 7.1[73] is depicted in **Figure 17**. The region 28.2 is reported by antiSMASH as a lager hybrid-BGC, which predicted multiple NP types, including phosphonate, acyl amino acids, butyrolactone, NRPS-like, T1PKS, and hserlactone. To idetify the boundaries of phosphonate BGC and to uncover its phosphonate biosynthesis, The specific vector pDS0007 was first constructed for mainipulation of *S. iranensis* DSM41954,[154]and sucessfully used to creat the pepM gene deletion of *S. iranensis* DSM41954. The function of *pepM* gene was confirmed by detecting the phosphonate signal in ^{31}P NMR. In addition, to further enhance the phosphonate production, we used two strategy including overexpression the *pepM* gene and deletion of putative repressor *lacI*.[154] Thus, the pepM-ppdAB overexpression mutants and the deletion mutant of *lacI* was created by conjugation. After cultivation of corresponding strains, distinct peak patterns appreared in compariison to the wildtype, with enhanced signal intensity observed in culture supernatant sample of the overexpression mutants, but no increased peak patterns detected in the *lacI* deletion mutant (**Figure A 19**). Therefore, the overexpression of genes encoding the first

enzymes of the pathway is useful strategy to enhance the phosphonate production, but *lacI* gene is not repressor in phosphonate biosynthesis as expected.

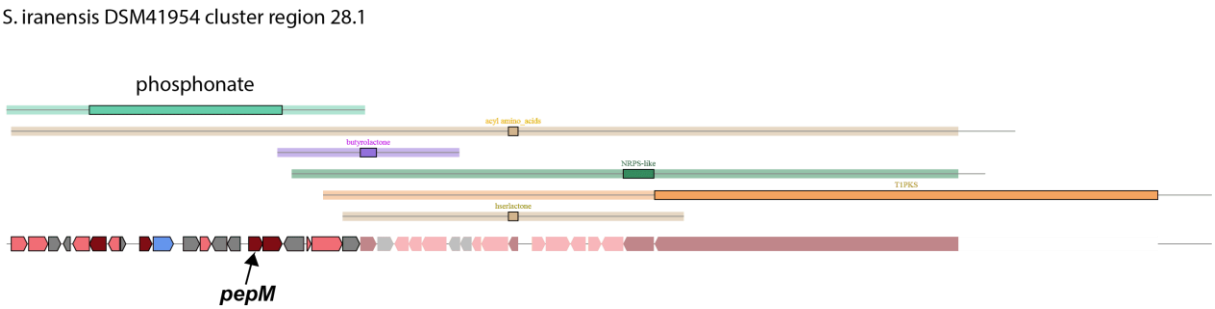


Figure 17. The cluster region 28.2 contains phosphonate BGC, which is a lager hybrid-BGC of *S. iranensis* DSM41954.

2.1.5 Validating the phosphonate production in the *K. fiedleri* DSM114396 based on a putative BGC using ^{31}P NMR analysis

The rare actinomycete *K. fiedleri* DSM114396 was described as a novel type strain,[155] and the phosphonate production in this strain was demonstrated by ^{31}P NMR previously. The phylogenetic analysis showed that *K. fiedleri* DSM114396 has conserved core enzyme PepM and further comparision of full gene cluster to other known phosphonate producers exhibited their different gene orders, thus *K. fiedleri* DSM114396 was further selected as a candidate. To clearly establish the BGC boundaries, NCBI BLAST and Clinker were used for analysis, and thus comfirmed the phosphonate BGC location from position 3401998 to position 3428393. Notably, we found a gene kfp24 encoding a putative transcriptional regulator of the LuxR family in the cluster, which may drive the phosphonate biosynthesis.

To test this hypothesis and futher investigate the phosphonate biosynthesis, we first deterimined the minimal gene cluster kfp02-kfp04 by heterologous expression in *Streptomyces albus* and *Streptomyces lividans*. Metabolites produced by the resulting strains were analyzed by ^{31}P NMR and LC-MS in comparision to the wildtype strain.[156]Phosphonate-specific chemical shifts were observed from either heterologous expression strain. Due to the instability of phosphonoacetaldehyde, which can be easily transformed to other products, it was detected by derivatived with dansyl group in the culture after incolubation of 3 days and 5 days(**Figure 18**, **Figure 19**).

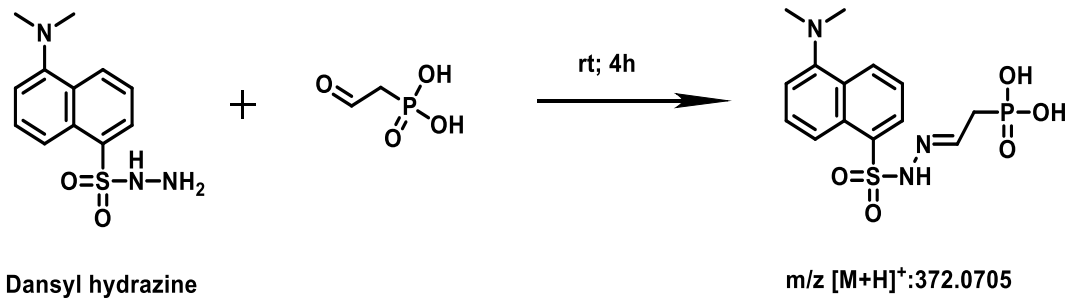


Figure 18. Chemical structure of the phosphonoacetaldehyde derivatized with dansyl hydrazine, and the product with an *m/z* value of 372.0.

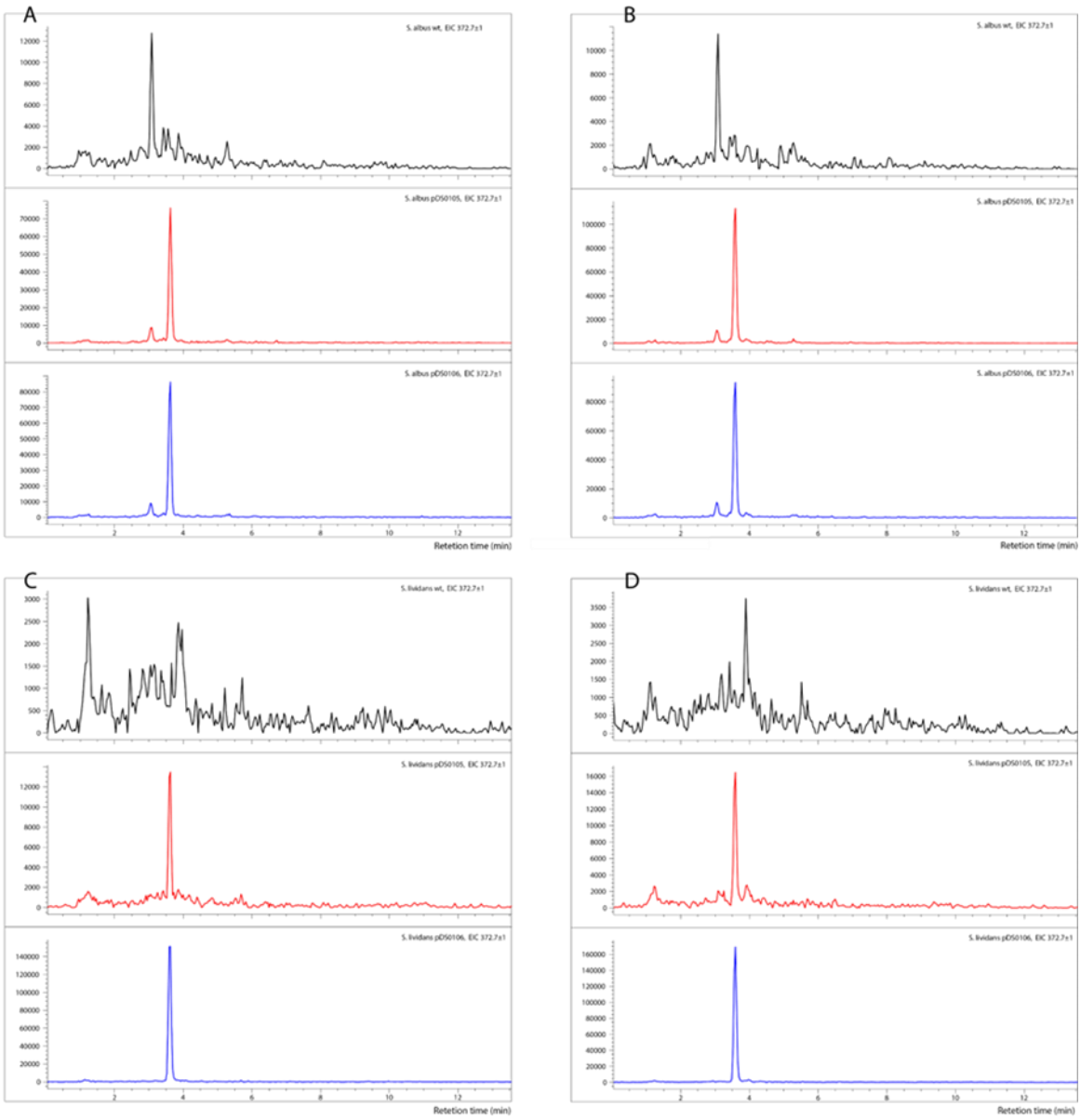


Figure 19. Phosphonoacetaldehyde production in the culture of heterologous expression strain. **A.** EIC spectrum of the sample from the culture of *S. albus* pDS0105 and *S. albus* pDS0105 after incubation of 3 days. **B.** EIC spectrum of the sample from the culture of *S. albus* pDS0105 and *S. albus* pDS0106 after incubation of 5 days. **C.** EIC spectrum of the sample from the culture of *S. albus* pDS0105 and *S. lividans* pDS0105 after incubation of 3 days. **D.** EIC spectrum of the sample from the culture of *S. lividans* pDS0106 and *S. albus* pDS0105 after incubation of 5 days

Meanwile, we further constructed the deletion mutant to determine if the genes kfp02-kfp04 are required to phosphonate biosynthesis in *K. fiedleri* DSM114396. Analysis of resulting strains by ^{31}P NMR revealed no significant phosphonate-specific peak patterns were observed, but phosphonate production restored after complementation of the deletion mutant with ectopic expression of kfp02-kfp04. These results were clearly demonstrated the essential role of kfp02-kfp04 in the phosphonate biosynthesis. In addition, overexpression of kfp02-kfp04 resulted in distinct peak patterns with stronger signals and new chemical shifts compared with the wildtype. As described above, we identified kfp24 as a gene encoding a putative pathway-specific

transcriptional activator belonging to the LuxR family. Therefore, we expected the overexpression of kfp24 should lead to more phosphonate production. To test our hypothesis, the kfp24-overexpression mutants were constructed and introduced into *K. fiedleri* wildtype. Analysis of resulting strains by ^{31}P NMR revealed more and new phosphonate-specific chemical shifts observed, confirming that the LuxR-family regulator Kfp24 is the pathway specific transcriptional activator of the phosphonate biosynthesis.

2.1.6 Production screening and pre-purification by size exclusion and ion exchange chromatography

One of the main goals in this thesis was to isolate and purify the unknown phosphonates from the spent culture media of *S. iranensis* DSM41954 and *K. fiedleri* DSM114396. This task was highly challenging because of the complex sample matrix, which contained high concentrations of inorganic salts, hydrophobic esters, and macromolecular constituents such as proteins and polysaccharides, as well as abundant polar metabolites including sugars and amino acids. Therefore, an efficient extraction and pre-purification strategy was required to reduce interference from these components as far as possible while preserving the target phosphonates for subsequent isolation.After comparing the intensity of target signals of ^{31}P NMR from the spent cuture sample of wildtype *S. iranensis* DSM41954 and *K. fiedleri* DSM114396, *K. fiedleri* DSM114396 was selected for scale-up production and purification of the unknown phosphonates. Based on the optimal growth conditions identified during the initial production screening, 13 L of *K. fiedleri* DSM114396 culture was prepared and concentrated by rotary evaporation to a final volume of 800 mL. The samples were monitored for the target compounds by detecting signals with chemical shifts of 8 ppm or higher in the ^{31}P NMR spectra. Given the complexity of the sample, additional extraction steps were required prior to further enrichment of the phosphonates. To remove undesirable constituents such as proteins, polysaccharides, colored metabolites, lipids, and other hydrophobic byproducts, methanol was added to a final concentration of 80%. The resulting suspension was filtered, evaporated to dryness, and the residue was reconstituted in water and analyzed by ^{31}P NMR (**Figure 20A**). Subsequently, 50 g of activated charcoal was added to further reduce matrix components that could interfere with further purification. Despite unavoidable losses during these steps, ^{31}P NMR analysis revealed an increase in the relative intensity of the target signals and a clearer appearance of phosphonate-specific chemical shifts (**Figure 20B**) In addition, an acid-treatment step was evaluated as an auxiliary purification measure, as similar procedures have been reported to facilitate phosphonate isolation and, in this case, likewise contributed to a reduction in undesirable compounds (**Figure A 22**). Interestingly, some phosphonates were extracted into MeCN, suggesting that variations in the side chain structures of the phosphonates produced by *K. fiedleri* DSM114396 result in distinct polarity properties (**Figure A 23**).

To futher enrich phosphonates, column chromatography with weak anion-exchange resin was applied in accordance with the acidic nature of phosphonate compounds. Initially, 100mg pretreated samples were reconstituted and loaded to test if phosphonates could be separated. By analyzing eluted samples by ^{31}P NMR, phosphonate singal was detected in elution with

concentrations of 10mM, 50mM and 100mM bicarbonate solution (**Figure 21**), indicated that column chromatography with weak anion-exchange resin was available for separating different phosphonates produced by *K. fiedleri* DSM114396 culture.

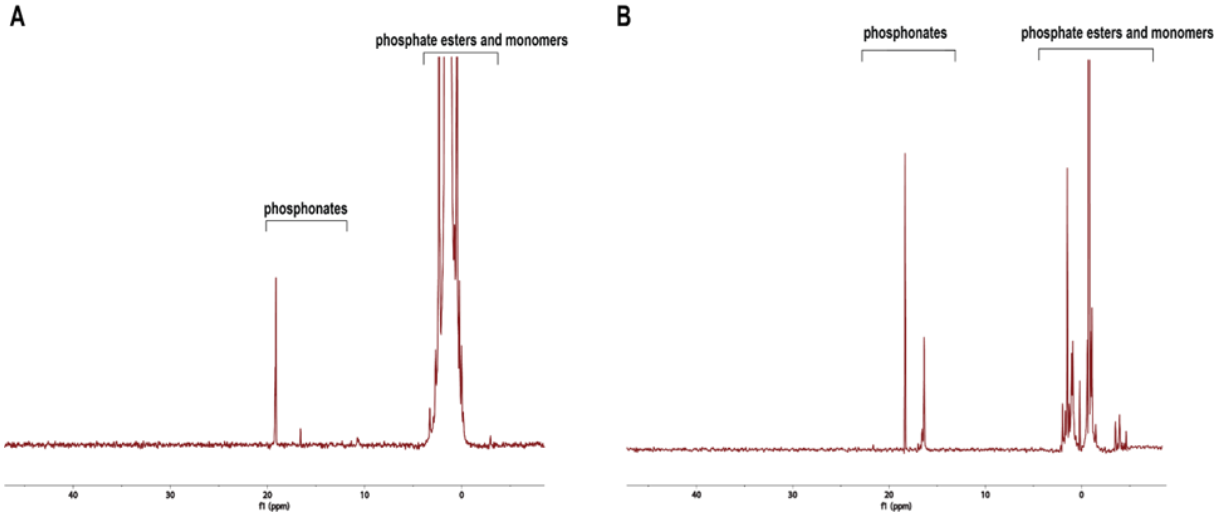


Figure 20. ^{31}P spectra of phosphonates from *K. fiedleri* DSM114396 **A.** Concentrated culture extract of *K. fiedleri DSM114396* shows the presence of phosphonates. **B.** Concentrated culture extract of *K. fiedleri DSM114396* after processing by activated charcoal shows the presence of phosphonates. The P peaks with chemical shifts between 10 and 40 ppm are usually indicative of phosphonates and phosphinates, whereas P peaks between +10 to -20 ppm range are usually phosphate monomers, esters and pyrophosphates

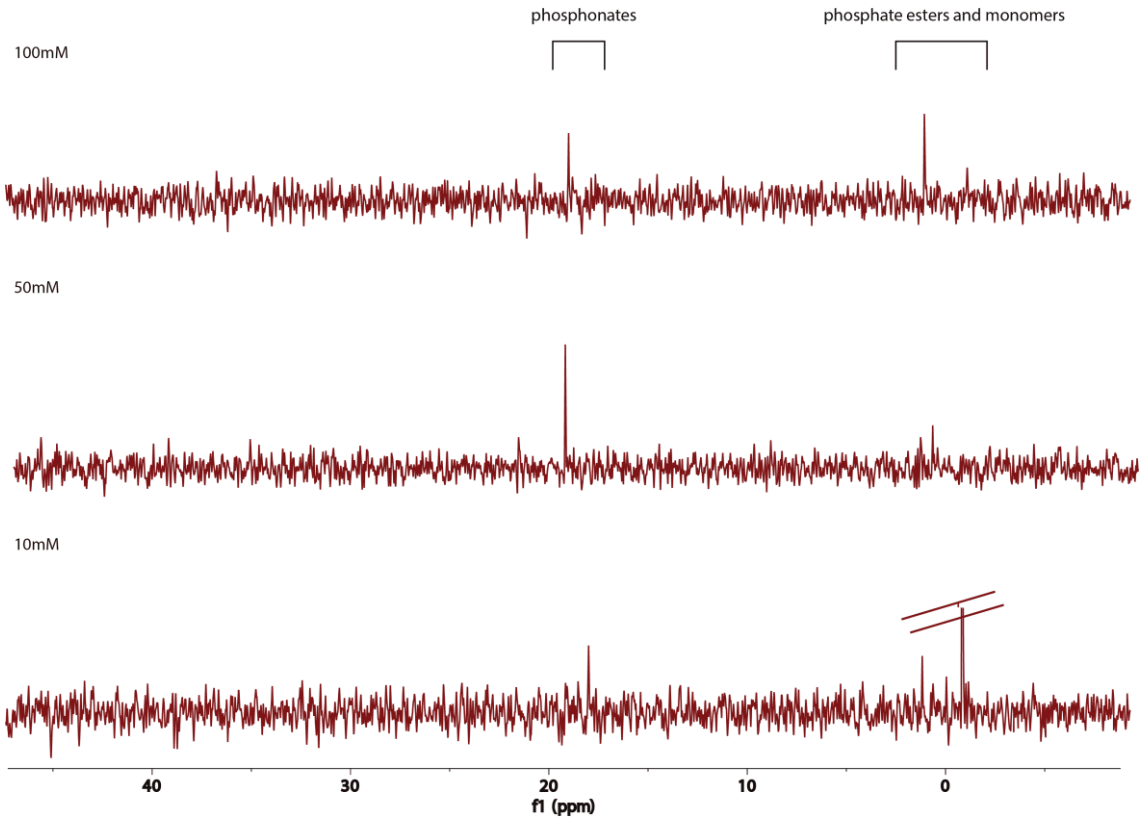


Figure 21. ^{31}P spectra of the concentrated elution fractions from *K. fiedleri* DSM114396 sample obtained by ion exchange resin separation. Phosphonate signals were detected in fractions eluted with 10mM, 50mM and 100mM bicarbonate solutions.

However, phosphonate signals were not detected with the increase of loading capacity and elution, revealed that tracing of phosphonate in ^{31}P NMR may be affected by excess bicarbonate, which accumulated during elution progressing. Although lyophilization has proven to remove the bicarbonate, the effectiveness was limited by the unavailability of a large-scale lyophilizer available. To overcome this limitation and to enable more reliable detection of the phosphonates in the obtained extracts, the different LCMS based detection and purification strategies were subsequently developed and explored.

2.1.7 Analytic strategies for phosphonate detection

A reliable analytical strategy is essential for tracking phosphonate-containing metabolites during isolation and purification. Although it would be logical to only perform ^{31}P NMR for the detection of phosphorus-containing compounds, this approach may not be complete and enough in further large-scale separations and purifications. ^{31}P NMR analysis requires relatively high sample amounts and is not well suited for the rapid screening of numerous fractions generated during multistep purification. In addition, the complexity of microbial extracts can complicate spectral interpretation and reduce analytical throughput. Challenges in phosphonate analysis is their occurrence as anionic species in microbial cultures, where they readily form complexes with metal ions such as Ca^{2+} and Mg^{2+}.[157]Such interactions may influence chromatographic separation and ionization efficiency during mass spectrometric analysis. Moreover, they are difficult to detect by conventional UV-based HPLC methods due to their feature with high polarity, low volatility, and the frequent absence of strong UV chromophores.[157,158]Their extreme hydrophilicity also limits their retention on standard reversed-phase columns such as C18, where they often elute at the begining. Thus, it is necessary to use chromatographic systems specifically designed for highly polar metabolites. To address these challenges, hydrophilic interaction liquid chromatography (HILIC) was evaluated as the primary chromatographic mode for phosphonate analysis. Since the first introdution of term “HILIC” in 1990, numerous stationary phases have been developed for HILIC, the majprity of which are derived from silanol-modified materials, including amino, amide, cyanopropyl, carbamate, diol, and polyol phases, as well as mixed-mode stationary phases.[159]Depending on the chromatographic conditions employed, HILIC can function in reversed-phase, aqueous normal-phase, and organic normal-phase modes, enabling the simultaneous separation of polar and nonpolar compounds such as sugars, polysaccharides, peptides, nucleic acids, amino acids, acidic compounds, and basic compounds. Given the strongly polar nature of phosphonates, HILIC was considered a preferred chromatographic analysis for method development.

To develop an effective LC-MS method, two HILIC columns, Luna® HILIC column and ZIC® HILIC column, were compared using a standard phosphonate mixture (SPM). Both columns allowed the detection of all phosphonates present in the standard mixture, confirming that HILIC is generally appropriate for phosphonate analysis. However, a clear difference was observed in chromatographic performance. The Luna® HILIC column provided better peak shapes in comparison to the ZIC® HILIC column. This is particularly important for reliable

extracted ion chromatogram (EIC) analysis and reproducible tracking of target compounds across fractions. Based on these results, the Luna® HILIC column was selected for subsequent analyses. As a verified phosphonoalamides producer, the culture of *S. kutzneri* DSM40907 was analyzed, EIC annalysis showed that phosphonoalamides A–D were retained on the Luna® HILIC column and could be detected directly from culture extracts (**Figure 22A**). This data further demonstrated that the method was suitable not only for phosphonate standards but also for structurally related phosphonate metabolites present in complex microbial culture. However, chromatographic retention and selectivity in HILIC are highly sensitive to mobile phase composition, including the buffer concentration, pH, and the ratio of aqueous to organic solvent. This can make method optimization difficult and may complicate further isolation and purification. For this reason, an alternative chromatographic approach was investigated to facilitate more convenient phosphonate detection.

A Hypercarb column, consisting of 100% porous graphitic carbon (PGC) as the stationary phase, was therefore evaluated as a complementary analytical method. Porous graphitic carbon differs fundamentally from silica-based stationary phases in both surface chemistry and retention mechanism. Due to its highly polarizable graphitic surface, PGC often shows enhanced retention of small, polar, and structurally challenging compounds, including analytes that are poorly retained on regular reversed phase columns such as C18. In addition, the mobile phase of hypercarb column is regular solvent, which may simplify the LC-MS analysis. To evaluate the performance of the Hypercarb column for phosphonate analysis, culture extracts of *S. kutzneri* DSM40907 were analyzed. The resulting extracted ion chromatograms (EICs) showed that phosphonoalamides A–D exhibited improved retention on the Hypercarb column (**Figure 22B**).In addition, standard phosphonates were analyzed under the same conditions, and the EIC profiles similarly demonstrated enhanced retention compared to the Luna® HILIC (**Figure A 25**). This improved retention is advantageous for distinguishing and separating analyte peaks from the fraction during further purification. Moreover, the use of a more conventional mobile phase made the Hypercarb method operationally simpler than HILIC, which is beneficial for further fraction analysis.

Although Hypercarb columns showed excellent analytical capability for phosphonates, the columns are costly for the purification of large sample volumes, especially from complex culture. In addition, the hypercarb SPE cartridges were used to test if solo compound can be sperate from the standard phosphonate mixture, results indicated that the expected compound separation was not achieved with the hypercarb SPE cartridges (**Figure A 24**). Although phosphonates are better fitted with LC analysis, their measurement using conventional detectors remains challenging due to the lack of adequate chemical moieties, such as chromophores and strong UV-absorbing groups.[160]For these reasons, three main labeling reactions were developed for further detection and isolation.

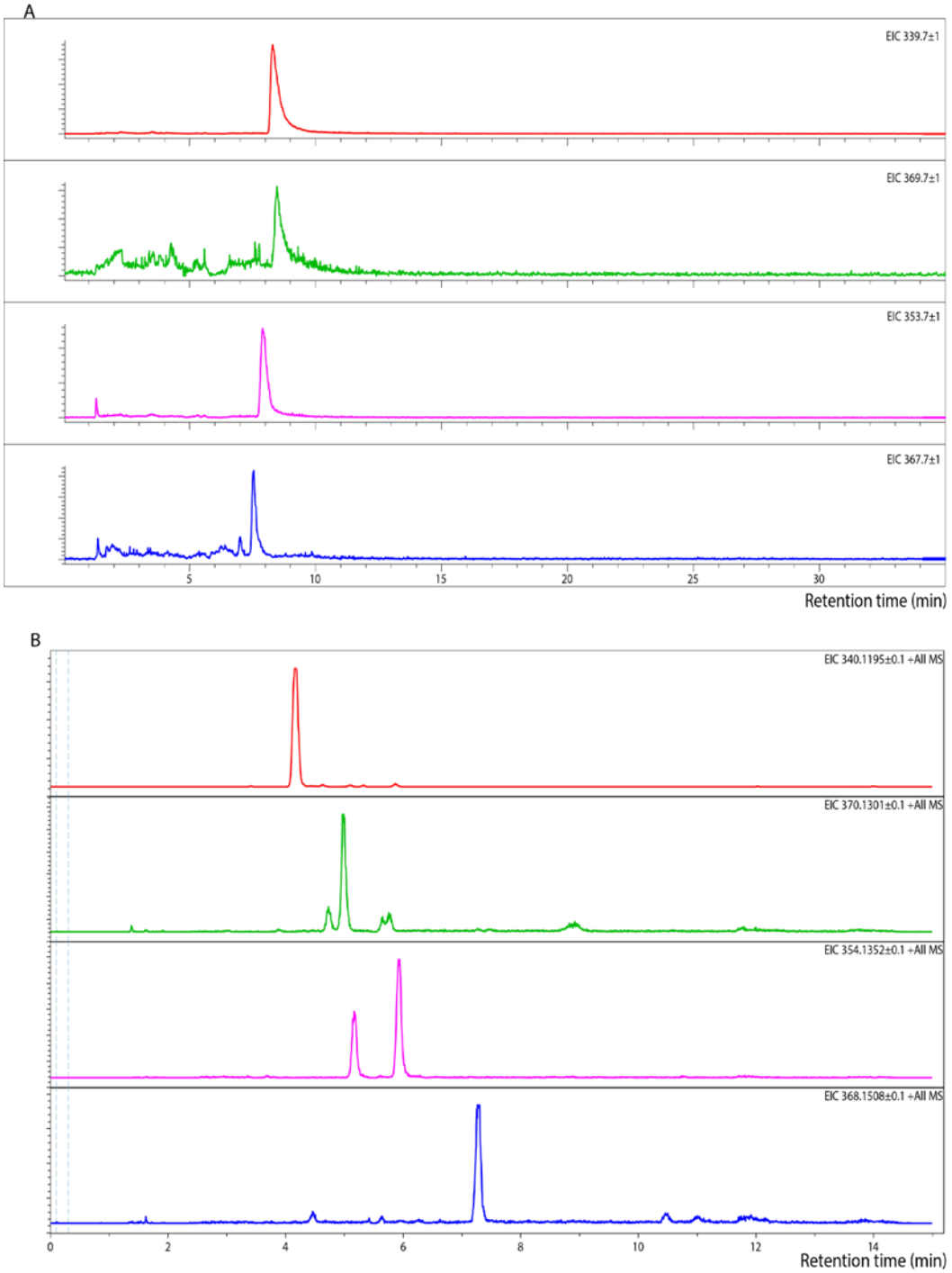


Figure 22. Detection of phosphonoalamides A-D in crude extract of *S. kutzneri* DSM40907. **A**. EIC spectrum showed that phosphonoalamides A-D were retained on the Luna® HILIC column and could be detected directly from the culture extract **B**. EIC spectrum of hypercarb column analysis, revealed improved retention of phosphonoalamides A-D.

2.1.8 Labeling of phosphonates with 9-fluorenylmethyl chloroformate

9-Fluorenylmethyl chloroformate (FMOC-Cl) derivatization was investigated as a selective approach for probing whether the unknown phosphonate produced by *K. fiedleri* DSM114396 contains amine functionality. FMOC-Cl is widely used as a pre-column derivatization reagent

because it reacts readily with primary and secondary amines under mild alkaline conditions.[161,160,162]The resulting 9-fluorenylmethyl derivatives are less polar and UV-active, which improves their detectability by LC-MS and facilitates enrichment or separation using reversed-phase chromatography.

To assess the applicability of this approach to phosphonate analysis and isolation, six representative standards, including aminomethylphosphonate (AMPA), phosphinothricin(PT), glyphosate (GP), phosphomycin (PM), methylphosphonate(MP) and 3-phosphonopropioinc acid(3-PPA) were test.

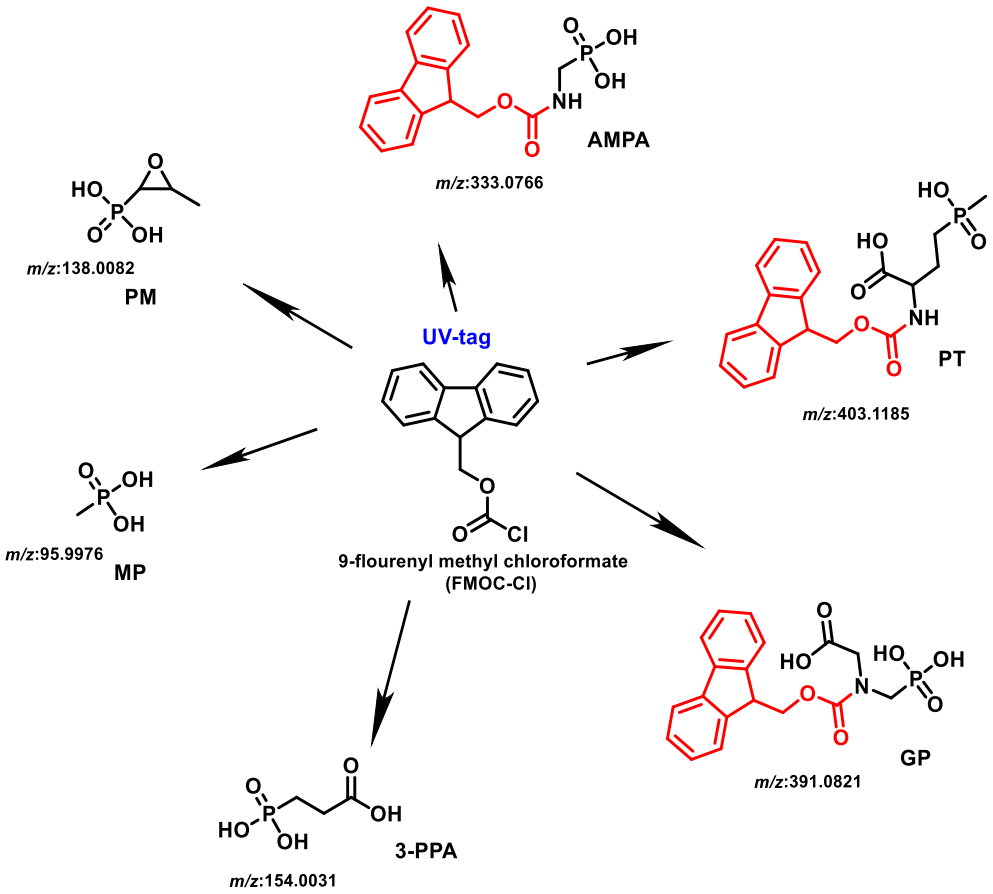


Figure 23. Chemical structures of aminomethylphosphonate(AMPA), phosphinothricin(PT), glyphosate (GP), phosphomycin(PM), methylphosphonate(MP) and 3-phosphonopropioinc acid(3-PPA) and derivatization reaction with FMOC. FOMC group is highlighted in red.

After derivatization, the stable FMOC derivatives were successfully yielded for AMP, GLY and PT (**Figure 23**). These products displayed clear UV absorbance at 254nm and were retained sufficiently on a C18 column, confirming successful derivatization of amine-containing phosphonates. (**Figure 24A**) In contrast, PM, MP, and 3-PPA failed to form detectable FMOC products, as expected from the absence of a primary or secondary amine group. (**Figure 24B**) These results confirmed that FMOC derivatization provides a useful chemical filter for distinguishing amino-phosphonates from non-amino phosphonates. Notably, hydrolyzed FMOC byproducts were consistently observed in reaction mixtures. Although this byproduct is unavoidable in aqueous derivatization systems and complicates sample analysis to some extent, it did not preclude assessment of whether derivatization had occurred.[163]

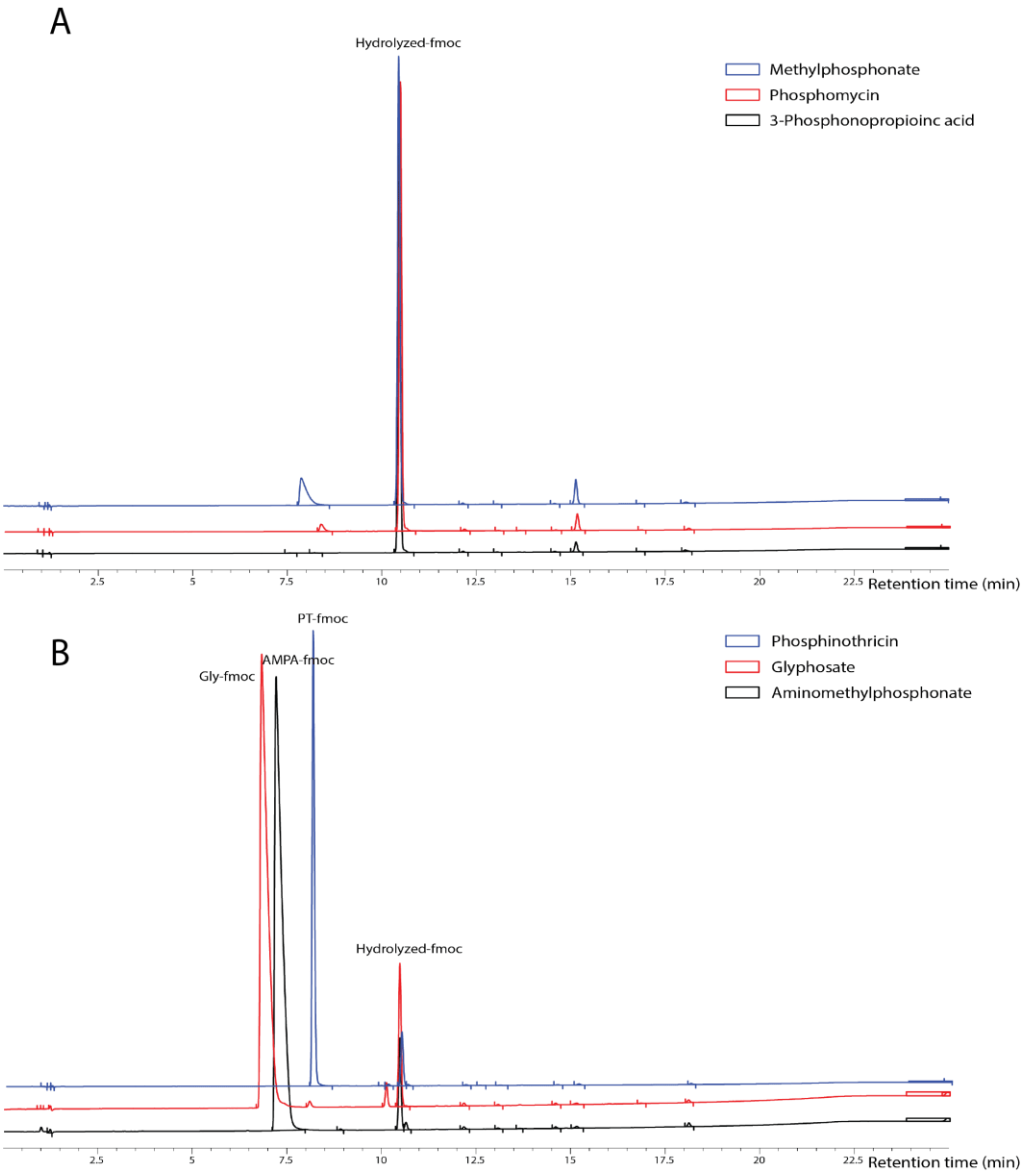


Figure 24. EIC chromatograms of FMOC derivatives of aminomethylphosphonate(AMP), phosphinothricin(PT), glyphosate (GP), phosphomycin(PM), methylphosphonate(MP) and 3-phosphonopropioinc acid(3-PPA) in 254nm. **A**) Unlabeled MP, PM and 3-PPA. **B**) Labeled PT, GP and AMP.

When this approach was applied to the crude extract of *K. fiedleri* DSM114396, no phosphonate-specific ^{31}P NMR signals appeared in the MeOH-eluted fraction after FMOC treatment, followed by C18 processing (**Figure 25**) This finding indicates that the unknown phosphonate was not retained in the derivatized fraction and therefore most likely did not react with FMOC-Cl.

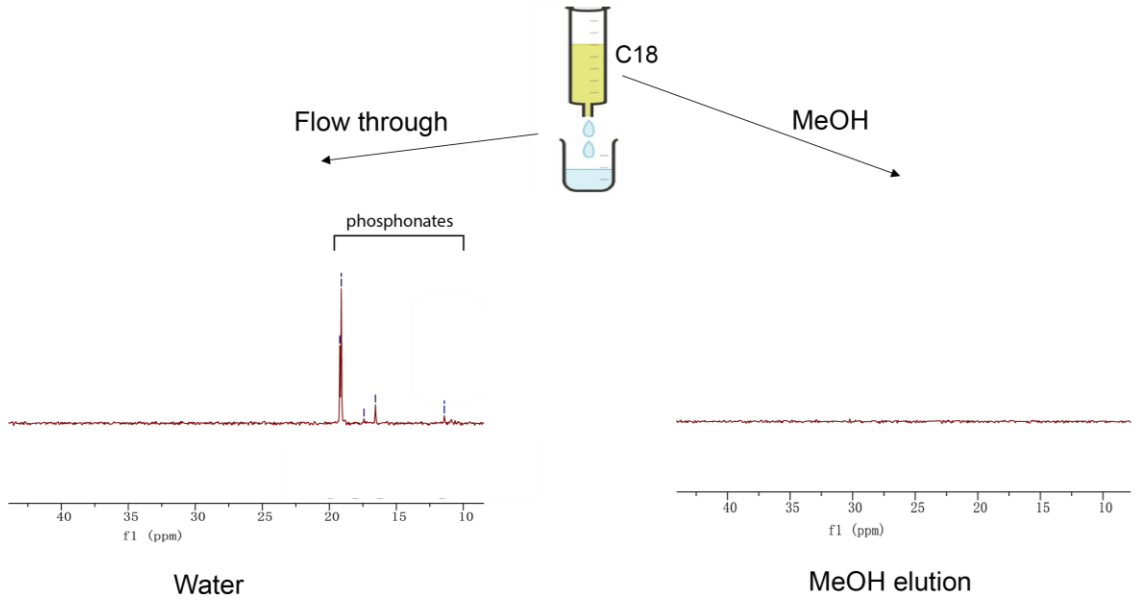


Figure 25. ^{31}P spectra of the labeled *K. fiedleri* DSM114396 sample. The detection of phosphonates produced by *K. fiedleri* DSM114396 in the MeOH elution fraction suggests that these compounds may not contain primary or secondary amine groups.

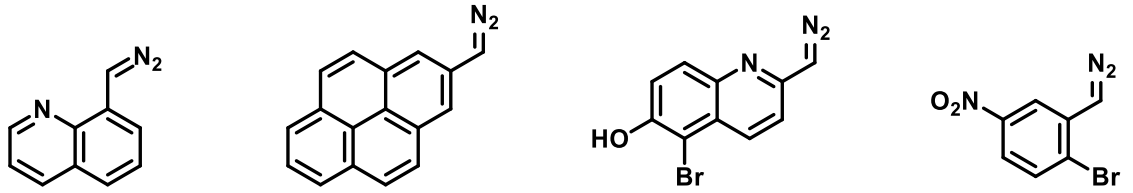
2.1.9 Labeling of phosphonates with diazo probe

Compared with FMOC, diazo probes show higher reactivity. The reaction between aryl diazo probe and acidic compounds is mediated by an electrophilic reaction activated by acid-base interaction. For this reaction, no additional catalysts are required and can proceed even under neutral conditions. Moreover, diazo probe have been successfully applied to derivatize the phosphate and carboxyl group.[164,165,166]For example, Liu *et al* used diazo compounds to derivatize phosphate group on ribonucleotides for the detection of endogenous ribonucleotides from cells.[167]Li *et al* used diazo compounds to label and determine derivatize twenty-two intermediates in central carbon metabolism pathway.[164] Kalie *et al* used diazo compounds to esterify the carboxyl group in the green fluorescent protein in an aqueous environment for cytosolic entry study.[168]

2.1.9.1 Characterizations of synthesized aryl diazo reagent

To achieve the analysis of phosphonates for facilitating further purification, we developed an aryl diazo-based labeling strategy with Br tag while preserving the integrity of the phosphonate moiety. Different diazo labeling reagents were synthesized and shown in **Figure 26**. Unfortunately, the synthesis experiment of **1** failed because quinoline was not dissolved in DCM, but successful case was reported involving the use of DSMO.[165]The compound **2** was

not good to detect under regular LC-MS analysis procedure. Finally, we got two diazo probes, the high-resolution mass spectrometry and NMR analysis demonstrated that these two Br-diazo compounds **3** and **4** were successfully synthesized. Both were determined as labeling reagent for further phosphonate labeling reactions but **4** was a priority. The introduction of Br brings the isotope peak, and the hydrophobic phenyl group enhances the retention of phosphonates on reversed-phase LC that can improve the chromatographic separation. Of note is that manganese dioxide from different sources affects the Br-diazo synthesis, our experiments revealed MnO_2 produced by electrolysis exhibits the most effective oxidation properties, and the reaction was completed in 2 hours. In addition, due to the instability of probe compound, it was directly used for the labeling of phosphonates.



1 2 3 4

Figure 26. Structures of synthesized Br-diazo in this study. The protocol and synthesis were kindly constructed by Dr. Chamber C. Hughes.

2.1.9.2 Labeling reaction of phosphonates

To evaluate the utility of our aryl diazo probe for labeling phosphonates, we used AMP, PT, GLY, PM, MP and 3-PPA, which were reacted with the probe compound and examined by LC-MS analysis. The results showed that the measured precursor ions and product ions of AMP, PT, GLY, PM, MP and 3-PPA-diazos were identical to the theretical values(**Figure 27**), suggesting that all the phosphonates successfully reacted with diazo probe compound and formed the desired derivatives of phosphonate diazos. The derivative peaks showed better retention and UV absorption, supporting the conclusion that diazo tagging enhances LC-MS performance by altering compound properties, thereby enable the phosphonate isolation. Notably, alkylated phosphate is the major product of the six analytes, although there is carboxyl group site available for alkylation in PT, GP and 3-PPA. In addition, the reaction with the phosphate diesters were barely observed. This result is qualitatively in agreement with the literature reported.[166]and the reactivity of diazo compounds were affected by pH and equivalents. The main product of diazo labeling is generated by the reaction with phosphonic acid group in pH 6.9 borate buffer.

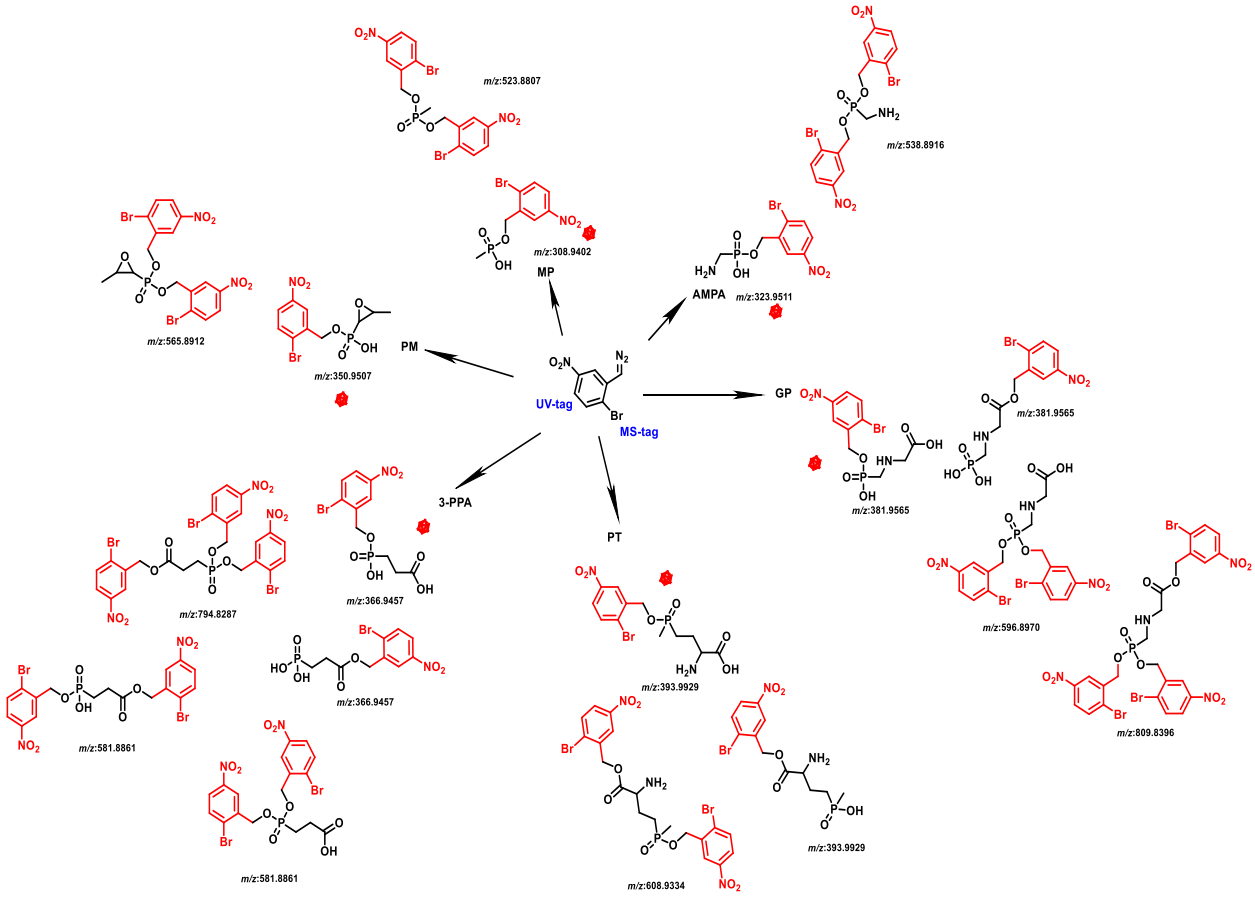


Figure 27. Structures theoretically of aminomethylphosphonate(AMPA), phosphinothricin(PT), glyphosate (GP), phosphomycin(PM), methylphosphonate(MP) and 3-phosphonopropioinc acid(3-PPA) after labeling with diazo reagent. The actual products obtained are marked with red diamonds.

In contrast, unlabeled phosphonates appeared close to the baseline at very early retention times, consistent with the high polarity and generally poor ESI response of the nature of phosphonates. In addition, the top peak eluting at approximately 6.0 min is primarily attributed to the formation of H_2O-diazo, since the hydrolysis is always predominant for reactions of diazo compounds in water (**Figure 28**). Nevertheless, the retention shows no obvious difference in the different analytes. Rather than interfering with labeled phosphonate detection, this peak eluted in a distinct chromatographic region and served as a useful indicator of reagent activity and reaction. And the absence of overlapping signals between H_2O-diazo and phosphonate-diazo derivatives demonstrated that the method maintains analytical clarity and utility despite excess derivatization. Additional peaks observed at retention times later than 8 min are likely associated with minor reaction by product or matrix-related components (**Figure 28**). These peaks do not overlap with the phosphonate diazo derivatives and therefore do not compromise analysis.

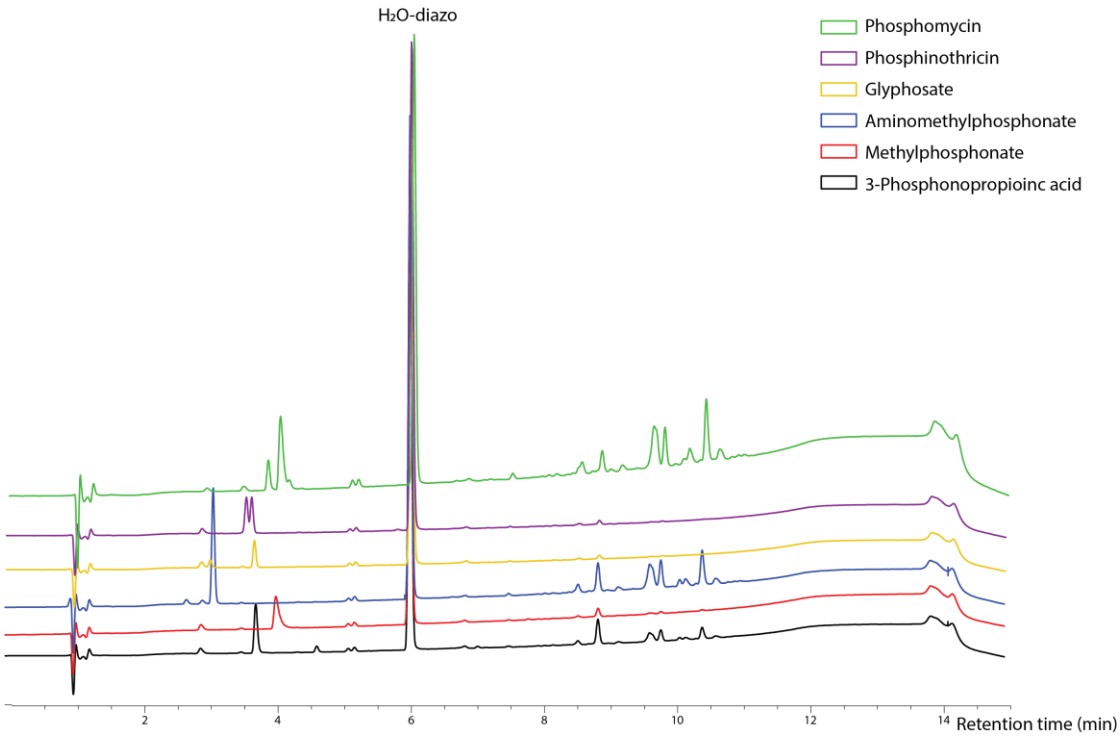


Figure 28. EIC chromatograms of diazo derivatives of aminomethylphosphonate, phosphinothricin, glyphosate, phosphomycin, methylphosphonate and 3-phosphonopropioinc acid at 254nm. Peaks with retention times between 2 and 6 min were assigned to the corresponding diazo derivatives, whereas peaks eluting after 8 min were attributed to other unknown products.

2.1.9.3 Optimization of labeling reaction

The conditions including buffer and pH were optimized to obtain high efficiency of the labeling reaction. To assess potential buffer effects on the labeling reaction, four buffer systems (borate, Bis-Tris, HEPES, and MOPS) were tested at three concentrations (250 mM, 500 mM, and 1 M). For each buffer and concentration, four pH conditions (pH 5.5, 6.0, 6.5, and 7.0) were examined for labeling AMP,PT,3-PPA. The results showed that the peak areas of AMP,PT,3-PPA derivatives presented slight variation.

2.1.10 Labeling of phosphonates with ethyl chloroformate

Ethyl chloroformate is widely used as a pre-column derivatization reagent across various fields for analysis due to several advantages.[169,170]Firstly, ester formation markedly reduces substrate polarity, which is particularly advantageous for the analysis of highly polar compounds. This benefit is especially evident when ECF labeling is performed in aqueous medium, as it avoids the need for complete water removal prior to derivatization and thereby simplifies sample preparation, especially for culture-derived samples. Secondly, the reaction proceeds rapidly at room temperature and typically affords high yields, while also enabling efficient esterification of carboxyl groups in multifunctional protic compounds such as amino acids.[170]Lastly, the esterification is highly fast, cost-effective and readily available reagents that can be easily removed by extraction or evaporation.[171] Based on this advantage and the

phosphonates that have been reported, we hypothesize that this chemical modification could increase the phosphonate hydrophobicity, thereby enable them retent on C18 columns.

To evaluate the labeling reaction and confirm our hypothesis, phosphinothricin(PT) and glyphosate(GP) was used to react with ECF as standard analytes. Due to ECF is reactive species with the ability to protect many protic functional groups, there are multiple sites can be reacted in PT and GP (**Figure 29**, **Figure 30**).

Under standard reversed phase LC-MS conditions, the underivatized compound exhibited minimal retention and eluted close to the solvent front. This is consistent with the high polarity and zwitterionic nature of PT and GP, which limits interaction with the hydrophobic stationary phase. After ECF derivatization, the chromatographic profile changed markedly. Several new peaks were observed at significantly longer retention times, as described.[170,172] And product ions were identical to the theoretical values (**Figure A 31**, **Figure A 32**). The obviously increase in retention time confirmed successful derivatization and demonstrated that introduction of ethyl groups effectively reduces polarity of PT and GP. Although such product heterogeneity may complicate absolute quantification, it does not compromise qualitative detection or chromato-graphic retention. These results demonstrate that ECF derivatization provides an effective and experimentally simple strategy for improving the reversed-phase retention of phosphonate compounds.

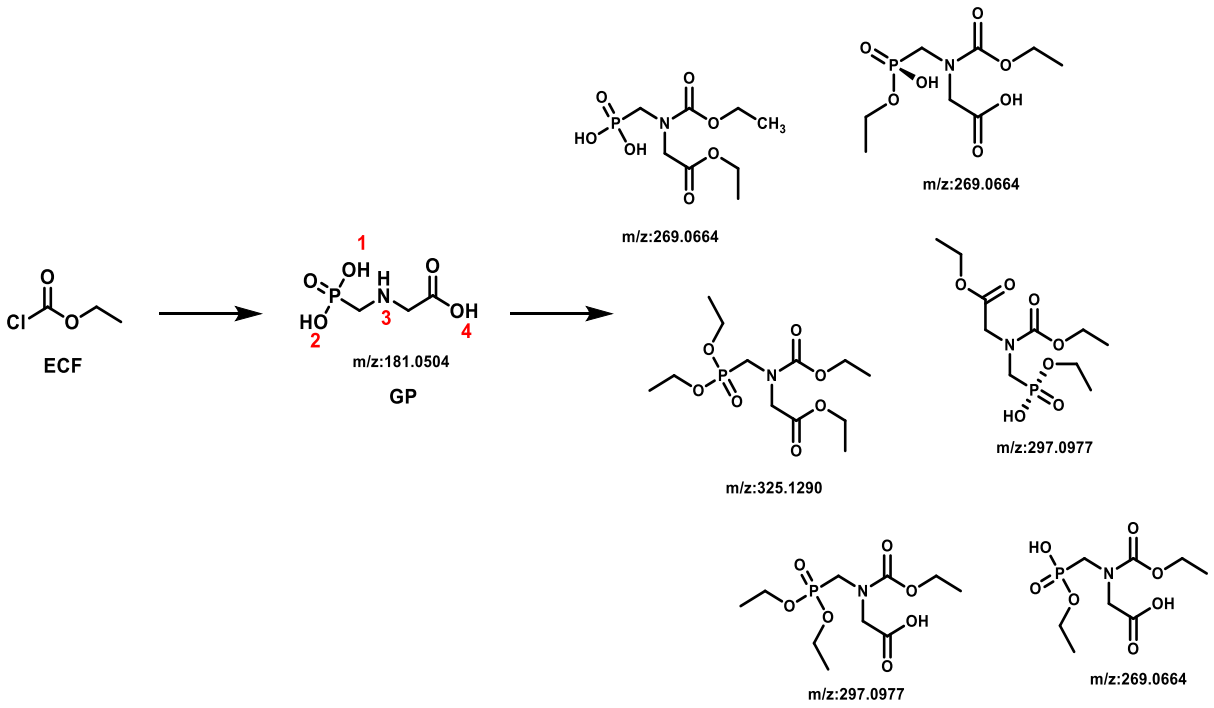


Figure 29. Chemical structures of glyphosate (GP) after labeling with ECF. Available labeling sites are highlighted in red.

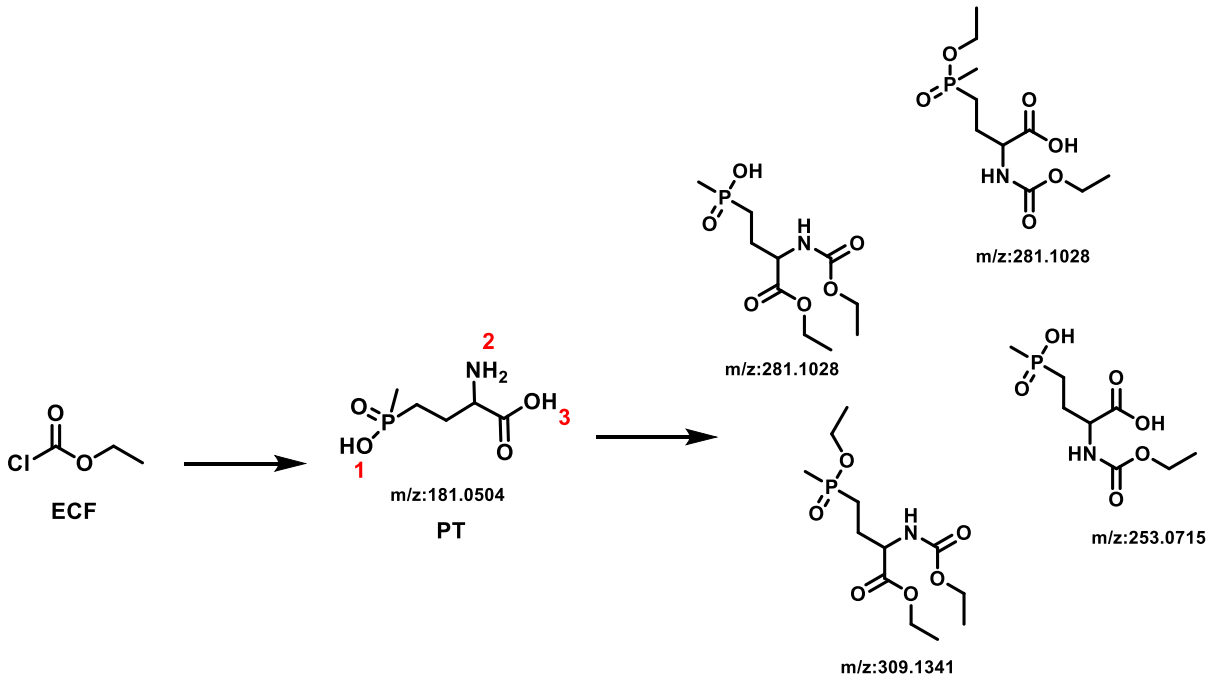


Figure 30. Chemical structures of phosphinothricin(PT) after labeling with ECF. Available labeling sites are highlighted in red.

Additional purification by HPLC indicated that the derivatized PT and GP products could be isolated, even using a small scale system. Unforunately, pure compounds were not obtained due to insufficient samples. Therefore, large pyridine and alcohol in excess to the labile ECF reagent is necessary to get the high ester yields in aqueous media as confirmed by Opekar *et al*,[170] since further purification needs enough sample. Neverthless, further study of labeling with ECF will be worth for phosphonate purification.

2.2 Materials and methods

For detailed materials and methods can be referred to the original publication 1-3, the following sections provide additional supplementary information for certain procedures.

2.2.1 Strains

All wildtype strains were from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). All strains used in this study were listed in **Table 1**.

Table 1. Strain list from DSMZ and Tübingen strain collection

No.	SPECIES	Isolation, sampling and environmental information from DSMZ
41867	*Streptomyces glauciniger*	aerobe, spore-forming, Gram-positive bacterium that was isolated from soil in a willow wood
41954	*Streptomyces iranensis*	aerobe, spore-forming, Gram-positive bacterium that builds an aerial mycelium and was isolated from rhizospherical soil at a depth of 10 cm.
42034	*Streptomyces shenzhenensis*	spore-forming bacterium that was isolated from mangrove soil
41649	*Kitasatospora atroaurantiaca*	spore-forming, Gram-positive bacterium that builds an aerial mycelium and was isolated from soil.
43861	*Kitasatospora setae*	spore-forming bacterium that builds an aerial mycelium and produces antibiotic compounds.
101999	*Kitasatospora cheerisanensis*	spore-forming, Gram-positive bacterium that was isolated from soil humus sample.
40043	*Streptomyces alboniger*	spore-forming bacterium that produces antibiotic compounds and was isolated from forest soil.
41230	*Streptomyces libani (rufus)*	spore-forming, Gram-positive bacterium that produces antibiotic compounds and was isolated from soil
40133	*Streptomyces resistomycificus*	spore-forming bacterium that builds an aerial mycelium and produces antibiotic compounds
41840	*Streptomyces seoulensis*	aerobe, spore-forming, Gram-positive bacterium that was isolated from soil
102214	*Streptomyces hokutonensis*	a bacterium that was isolated from sample of strawberry root rhizosphere
41801	*Streptomyces monomycini*	spore-forming, Gram-positive bacterium of the family Streptomycetaceae
41691	*Streptomyces mutomycini*	spore-forming bacterium that builds an aerial mycelium and produces antibiotic compounds
40386	*Streptomyces aureocirculatus*	spore-forming, Gram-positive bacterium that builds an aerial mycelium and was isolated from Soil.
44228	*Saccharopolyspora spinosa*	spore-forming, Gram-positive bacterium that builds an aerial mycelium and was isolated from soil; sugar mill rum still.
46670	*Kibdelosporangium banguiense*	spore-forming bacterium that builds an aerial mycelium and was isolated from soil

Table 1. Strain list from DSMZ and Tübingen strain collection (Continued)

No.	SPECIES	Isolation, sampling and environmental information from DSMZ
40907	*Streptomyces kutzneri*	unknown
45285	*Amycolatopsis xylanica*	aerobe, spore-forming, Gram-positive bacterium that was isolated from soil
40283	*Kitasatospora purpeofusca*	spore-forming, Gram-positive bacterium that produces antibiotic compounds and was isolated from soil
116176 (BCCO 100798)	*Lentzea kristufekii*	isolated from bare soil developing on Miocene lacustrine clay sediment
114396 (Tü 4103)	*Kitasatospora fiedleri*	isolated from soil in Indonesia Java
40736 (Tü 494)	*Streptomyces viridochromogenes*	unknown
Tü I6	*Streptomyces sp.*	isolated from an Indonesian mangrove sediment sample
Tü 18	*Streptomyces sp.*	isolated from soil
Tü 3678	*Streptomyces sp.*	isolated from soil
Tü 21470	*Streptomyces sp.*	isolated from soil
40386	*Streptomyces aureocirculatus*	spore-forming, Gram-positive bacterium that builds an aerial mycelium and was isolated from soil.

2.2.2 Medium

All media were prepared by dissolving the relevant ingredients (see **Table 2**) in deionized water and subsequent adjusted pH using NaOH or HCl, respectively. And 2 % (w/v) agar (Sigma-Aldrich, St. Louis, USA) were added for solid media. All media were autoclaved at 121℃.

Table 2. Compositions of media and buffer used.

Name	Ingredients	Concentration (g/L)	Volume per L medium (mL)	pH
M65	Yeast extract	4		7.0-7.3
	Malt extract	10		
	Glucose	4		
	CaCO3	2		
HM/ISP2	Yeast extract	4		7.0-7.3
	Malt extract	10		
	Glucose	4		
MS	Mannitol	20		7.0-7.3
	Vollfett soy flour	20		
OM/ISP3	Oatmeal	20		7.0-7.3
	Trace metal mix 2		5	
HM + 1 g/L K_2HPO_4	Yeast extract	4		7.0-7.3
	Malt extract	10		
	Glucose	4		
	K_2HPO_4	1		

Table 2. Compositions of media and buffer used. (Continued)

Name	Ingredients	Concentration (g/L)	Volume per L medium (mL)	pH
R5	Saccharose	103		7.0-7.3
	Glucose	10		
	K_2SO_4	0.25		
	$MgCl_2$	10.12		
	Casamino acids	0.1		
	Yeast extract	5		
	TES	5.73		
	Trace metal mix 2		2	
	dissolve in 955 mL H_2O, add after autoclaving:			
	1M $CaCl_2$		20	
	0.54% KH_2PO_4		10	
	20% L-Prolin		15	
GUBC	Saccharose	10		7.0-7.3
	Meat extract	5		
	Casamino acids	5		
	Glycerol	5		
	1M Na_2HPO_4 / KH_2PO_4 pH 7.3		5	
	Hunter's base		2	
	add after autoclaving			
	Balch's vitamins		10	

Table 2. Compositions of media and buffer used. (Continued)

Name	Ingredients	Concentration (g/L)	Volume per L medium (mL)	pH
ISP4	Solution I:			7.0-7.3
	10 g Soluble starch	10		
	dissolve in 500 mL H_2O			
	Solution II:			
	2 g $CaCO_3$	2		
	1 g K_2HPO_4 anhydrous	1		
	$MgCl_2$	1		
	NaCl	1		
	$(NH_4)SO_4$	2		
	Trace metal mix 1			
	dissolve in 500 mL H_2O			
	mix solution I and II and autoclave			
A1	Soluble starch	10		7.1
	Yeast extract	4		
	Bacto Peptone	2		
	Seasalts	33		
HM + 1M buffer	Yeast extract	4		7.0-7.3
	Malt extract	10		
	Glucose	4		
	1M Na_2HPO_4 / KH_2PO_4 pH 7.3		5	
MS + 1 g/L K_2HPO_4	Mannitol	20		7.0-7.3
	Vollfett soy flour	20		
	K_2HPO_4	1		

Table 2. Compositions of media and buffer used. (Continued)

Name	Ingredients	Concentration (g/L)	Volume per L medium (mL)	pH
MS + 1M buffer	Mannitol	20		7.0-7.3
	Vollfett soy flour	20		
	1M Na_2HPO_4 / KH_2PO_4 pH 7.3		5	
OM/ISP3 + 1 g/L K_2HPO_4	Oatmeal	20		7.0-7.3
	K2HPO4	1		
	Trace metal mix 2		5	
OM/ISP3 + 1M buffer	Oatmeal	20		7.0-7.3
	1M Na_2HPO_4 / KH_2PO_4 pH 7.3		5	
	Trace metal mix 2		5	
R5 + 1 g/L K_2HPO_4	Saccharose	103		7.0-7.3
	Glucose	10		
	K_2SO_4	0.25		
	$MgCl_2$	10.12		
	Casamino acids	0.1		
	Yeast extract	5		
	TES	5.73		
	Trace metal mix 2		2	
	dissolve in 955 mL H_2O			
	add after autoclaving			
	1M $CaCl_2$		20	
	0.54% KH_2PO_4		10	
	20% L-Prolin		15	

Table 2. Compositions of media and buffer used. (Continued)

Name	Ingredients	Concentration (g/L)	Volume per L medium (mL)	pH
R5 + 1M buffer	Saccharose	103		7.0-7.3
	Glucose	10		
	K_2SO_4	0.25		
	$MgCl_2$	10.12		
	Casamino acids	0.1		
	Yeast extract	5		
	TES	5.73		
	Trace metal mix 2		2	
	dissolve in 955 mL H_2O			
	add after autoclaving:			
	1M $CaCl_2$		20	
	0.54% KH_2PO_4		10	
	20% L-Prolin		15	
	1M Na_2HPO_4 / KH_2PO_4 pH 7.3		5	
NL200 + 1 g/L K_2HPO_4	Mannitol	20		7.0-7.3
	Cornsteep Powder	20		
	K_2HPO_4	1		
NL300 + 1 g/L K_2HPO_4	Mannitol	20		7.0-7.3
	Cotton Seed	20		
	K_2HPO_4	1		
SG	Glycerol	68		6.8
	Vollfett soy flour	20		
	KH_2PO_4	1		

Table 2. Compositions of media and buffer used. (Continued)

Name	Ingredients	Concentration (g/L)	Volume per L medium (mL)	pH
NL400 + 1 g/L K_2HPO_4	Glucose	10		7.0-7.3
	Soluble starch	20		
	Bacto Peptone	3		
	Meat extract	3		
	Yeast extract	5		
	$CaCO_3$	3		
	K_2HPO_4	1		
NL410 + 1 g/L K2HPO4	Glucose	10		7.0-7.3
	Glycerol	10		
	Oatmeal	5		
	Soy flour	10		
	Yeast extract	5		
	Casaminoacids	5		
	$CaCO_3$	1		
	K_2HPO_4	1		
Minimal medium	L-asparagine	0.5		7.0-7.2
	K2HPO4	0.5		
	MgSO4 x 7 H2O	0.2		
	FeSO4 x 7 H2O	0.010		
	add after autoclaving			
	50% glucose solution (autoclaved separately)		20	

Table 2. Compositions of media and buffer used. (Continued)

Name	Ingredients	Concentration (g/L)	Volume per L medium (mL)	pH
Hunter's concentrated base	$MgSO_4$ x 7 H_2O	59.3		6.8
	Nitrilotriacetic acid	20		
	$CaCl_2$ x 2 H2O	6.67		
	$(NH_4)6Mo_7O_{24} \cdot 4\ H_2O$	0.0185		
	$FeSO_4$ x 7 H_2O	0.198		
	Na-EDTA	0.25		
	$ZnSO_4$ x 7 H_2O	1.095		
	$FeSO_4$ x 7 H_2O	0.5		
	$MnSO_4$ × H_2O	0.154		
	$CuSO_4$ × 5 H_2O	0.0392		
	$Co(NO3)_2$ x 7 H2O	0.025		
	$Na_2B_4O_7 \cdot 10\ H_2O$	0.0177		
Balch's vitamins	p-amino benzoic aicd	0.005		7.0
	Folic acid	0.002		
	Biotin	0.002		
	Nicotinic acid	0.005		
	Calcium pantothenate	0.005		
	Riboflavin	0.005		
	Thiamin HCl	0.005		
	Pyridoxine HCl	0.010		
	Cyanocobalamin	0.0001		
	Thioctic (lipoic) acid	0.005		
LB	LB-Broth			

Table 2. Compositions of media and buffer used. (Continued)

Name	Ingredients	Concentration (g/L)	Volume per L medium (mL)	pH
Minimal medium	L-asparagine	0.5		7.0-7.2
	K_2HPO_4	0.5		
	$MgSO_4$ x 7 H_2O	0.2		
	$FeSO_4$ x 7 H_2O	0.010		
	add after autoclaving			
	50% glucose solution (autoclaved separately)		20	
Trace metal mix 2	$FeCl_3 \cdot 6\ H_2O$	0.2		
	$CuCl_2 \cdot 2\ H_2O$	0.01		
	10/8,2 mg $MnCl_2 \cdot 4\ H_2O$/ 2 H_2O			
	$(NH_4)6Mo_7O_{24} \cdot 4\ H_2O$	0.01		
	$Na_2B_4O_7 \cdot 10\ H_2O$	0.01		
Sporulation medium for *B. subtilis*	$FeCl_3$ x $6H_2O$	0.001		7.1
	$MnCl_2$ x $2H_2O$	0.0145		
	NH_4Cl	0.540		
	Na_2SO_4	0.105		
	KH_2PO_4	0.087		
	$CaCl_2$	0.1947		
	NH_4NO_3	0.096		
	$MgCl_2x6H_2O$	0.0083		
	Glucose	2		
	Sodium L-glutamate monohydrate	1.9		

Table 2. Compositions of media and buffer used. (Continued)

Name	Ingredients	Concentration (g/L)	Volume per L medium (mL)	pH
Antibiotic-test medium	KH_2PO_4		3	7.0-7.2
	K_2HPO_4		7	
	Tri-Na-Citrat x 2 H_2O		0.5	
	$MgSO_4$ x 7 H_2O		0.1	
	$(NH_4)_2SO_4$		1	
	Glucose		2	
LB+Glucose	LB-Broth			
	Glucose		1	

2.2.3 Instrument and analytical methods

2.2.3.1 LC-MS analysis

Low resolution analytical LC-MS was performed on an Agilent 1260 Infinity with different columns. For standard method: Kinetex 5 µm C18 100 Å, 100 x 4.6 mm column, G6125B LC/MSD. the flow rate was set at 0.50 mL/min with water as solvent A and acetonitrile (MeCN) as solvent B with 0.1% of formic acid in A and B. Standard conditions: Gradient 10-100% B over 10 minutes, hold 100% B for 2 minutes, then back to initial conditions for 1 minute, hold at initial conditions for 2 minutes. Long conditions: Gradient 10-100% B over 20 minutes, hold 100% B for 3 minutes, then back to initial conditions for 1 minute, hold at initial conditions for 1 minute.

High resolution analytical LC-MS(LC-HRMS/MS) were performed on 1290, Infinity II LC system (Agilent Technologies, France) coupled to an impact II mass spectrometer (Bruker, France) equipped with an electrospray ionization (ESI) source operated in the positive mode and negative mode. Chromatographic separations were developed with reverse phase, HILIC, or porous graphitic carbon columns. For reverse phase analysis, a Kinetex C18 (2.1x50 mm, 1.7 µm, Phenomenex) column was used at 30 ℃ with a flow rate of 0.50 mL/min. The mobile phase consisted of solvent A (water containing 0.1% formic acid) and solvent B (acetonitrile containing 0.1% formic acid). A 5 µL aliquot of sample was injected, and elution was performed using the following gradient: 5% B at initial conditions, increased linearly to 100% B over 10 min, held at 100% B for 2 min, returned to initial conditions over 1 min, and equilibrated for 2 min. For HILIC analysis, separations were performed on a XBridge® BEH Amide (4.6x150 mm, 5 µm, Waters), Luna HILIC (2.1x50 mm, 1.7 µm, Waters) and ZIC®HILIC (4.6x100 mm, 5 µm, Merck) column maintained at 30 ℃. The flow rate was set at 0.5 or 1.0 mL/min depending on the column used. The mobile phase consisted of solvent A (acetonitrile/50 mM ammonium formate, 9:1, v/v, pH 3.2) and solvent B (acetonitrile/water/50 mM ammonium formate, 5:4:1, v/v/v, pH 3.2). A 5 µL aliquot of sample was injected and separated using the following gradient: 90% B from 0 to 2.0 min, decreased to 40% B at 12.0 min, held at 40% B until 14.0 min, returned to 90% B at 14.5 min, maintained at 90% B until 20.0 min, followed by a 10 min re-equilibration step. For porous graphitic carbon analysis, a Hypercarb™ (4.6x100 mm, 5 µm, Thermo fisher scientific) column was used at 30 ℃. with a flow rate of 0.50 mL/min. The mobile phase consisted of solvent A (water containing 0.1% formic acid) and solvent B (acetonitrile containing 0.1% formic acid). A 5 µL aliquot of sample was injected, and separation was achieved using the following gradient: 5% B at initial conditions, increased linearly to 100% B over 10.0 min, held at 100% B until 12.0 min, returned to 5% B at 13.0 min, and equilibrated at 5% B until 15.0 min.

All mass spectrometric detection was performed in positive electrospray ionization (ESI) mode and negative ESI mode using auto-MS/MS acquisition. The positive-ion scan range was m/z 100-2000, whereas the negative-ion scan range was m/z 60-2000. Up to three precursor ions were selected in each cycle, with an absolute intensity threshold of 400. Dynamic exclusion was enabled, and the precursor cycle time was set to 0.30 s. Ion source parameters were set

as follows: capillary voltage 4.5 kV, an end plate offset of -500 V, nebulizer 2.2 bar, dry gas 10.0 L/min, desolvatation temperature 250 ℃ no collision in source. Nitrogen was used as drying gas in the source and for collision. The elemental compositions for all ions were determined with the instrument software DataAnalysis, the precision mass measurement was less than 5 ppm.

2.2.3.2 NMR analysis

The NMR experiments were kindly performed by NMR department at Eberhard Karls Universität Tübingen. The data was made available to me for further analysis.

^{31}P NMR spectra were recorded in 20% D_2O at 243 MHz on a Bruker Avance III HDX 600 MHz spectrometer fitted with a 5mm Prodigy BBO H&F CryoProbe. Phosphorus chemical shifts are reported in ppm relative to a 10 mM L-phosphinothricin HCl reference (δP 51.2) using a double-chamber coaxial NMR tube. 1H NMR spectra were recorded on a Bruker Avance III HDX 400 MHz, Bruker Avance III HDX 600 MHz. The deuterated solvents used in this study were from Cambridge Isotope Laboratories (Andover, MA). Spectra were collected in water supplemented with 20% D_2O as a lock solvent. The deuterated solvents used in this study were bought (Sigma-Aldrich, USA). NMR data were analyzed using MestReNova 14.3.0.

2.2.4 Sample preparation, isolation and purification

2.2.4.1 Preparation of cryopreserved spore suspensions

To ensure that sufficient active strains were available for all experiments, long-term spore suspensions stocks were prepared for each strain. Each strain was grown on agar plates until sporulation occurred. The spores were collected by suspending in 2mL sterilized 20% glycerol in water (v/v) with 0.1% Tween 80 (v/v), as previous described to maintain as a back-up. All spore suspensions were stored in 1.2 mL cryogenic vials at - 80 °C.

2.2.4.2 Preparation of pre-cultures of putative phosphonate producer

All culture agar of putative phosphonate producers (see **Table 1**) were prepared by inoculating spore suspension or the active culture agar from DSMZ. A small square block about 1cm^2 was picked with a sterilized toothpick from culture agar and transferred into sterilized 100mL Erlenmeyer baffled flasks (“mit Schikane”) with metal coil containing 30mL R5 or HM liquid medium. The pre-culture was grown at 150rpm and 29 ℃ on a platform shaker for three days.

2.2.4.3 Preparation of supernatant and agar sample for phosphonate compounds

For the liquid culture, the culture was transferred into 50mL centrifuge tubes or 750mL bottles. The cells were separated from the culture by centrifugation at 4000rpm for 20 min, and the

supernatant was collected, frozen and lyophilized. The dried materials were stored at -20 ℃ until further analysis.
For the agar, the culture plates were cut up into small pieces. Sliced culture was put into 250mL plastic beaker containing 100mL distilled water, and was frozen at -80 ℃ for overnight, then thawed to room temperature. Unfrozen agars were compressed by twisting the cheese cloth tightly to harvest liquid from each strain. The filtrate was centrifuged, collected, frozen and lyophilized. The dried materials were stored at -20 ℃ until further analysis.

2.2.4.4 Cultivation of *S. kutzneri* DSM40907

See publication 2

2.2.4.5 Processing of phosphonate compounds samples

The dried materials were melted and re-dissolved in 5mL Millipore water. Then concentrated liquid was centrifuged at 4000rpm for 20min to remove particles and applied to a 2g C18 solid phase extraction (SPE) cartridge, and aqueous phase was collected. C18 SPE cartridge was washed and equilibrated with MeOH and Millipore water. Then MeOH was added into aqueous phase to a final concentration of 80%MeOH and 20% H_2O. The precipitate was removed by filtration through filter paper. The soluble material (filtrate) was dried using a rotary evaporator and GeneVac.

2.2.4.6 Ion exchange treatment

The dried sample was redissolved in water and adjusted with acetic acid to 3. The sample was next subjected to fractionation on column chromatography with Dowex 1-X8 anion exchange resins (50g), which had been pre-equilibrated with 0.1% acetic acid. Under gravity flow, the column was sequentially washed with two column volumes of 0.1% acetic acid, followed by water. Bound components, including phosphonates, were then eluted with 1mM, 5mM, 10mM, 25mM, 50mM, 100mM, 250mM, 500mM and 1M bicarbonate solution. For Oasis MCX and WAX, the cartridges were conditioned and equilibrated with MeOH and water, then loaded with samples. After loading, the cartridges were washed with 2% formic acid to remove weakly retained interference. Then the elutions with 100% MeOH was performed to collect neutral compounds, followed by 5% NH_4OH in MeOH. For Oasis MAX and WCX, samples were prepared and loaded under basic conditions. Following loading, cartridges were washed with 5% NH_4OH to remove weakly retained compounds. Elution was carried out sequently with 100% MeOH for neutral compounds and followed by 2% formic acid in MeOH.

2.2.5 Bioassays

2.2.5.1 Preparation of E. *coli* K12, *E. coli* WM 6242 and *B. subtilis* 168 stocks

The initial strain E.coli K12 and E.coli WM6242 stocks was taken over from Dr. Julia Moschny. The initial strain B. subtilis 168 stock was obtained from Dr. Dardan Beqaj.

E. coli K12, *E. coli* WM 6242 was cultivated in the 100mL Erlenmeyer flask containing 30mL LB medium at 180rpm and 37 ℃ for overnight. Then the culture was diluted with 90% glycerol solution to OD_{600}=1 separately. The aliquots of 600 µl diluted culture was frozen at -20 ℃ as stock.

B. subtilis 168 was cultivated in the 500mL Erlenmeyer flask containing 100mL sporulation medium at 180rpm and 30 ℃ for 4-5 days. The culture was centrifuged at 4000rpm and 4℃ for 10min. Spores were collected and washed with sterilized water to OD_{546}=2. The remaining vegetative cells were killed by heating at 70 ℃ for 45min. The aliquots of 600 µl spore solution were frozen at -20 ℃ as stock. The optical density of *E. coli* K12, *E. coli* WM 6242 and *B. subtilis* 168 was determinated with OD_{600} and OD_{524}.

2.2.5.2 Preparation of cryopreserved spore suspensions

To ensure that sufficient active strains were available for all experiments, long-term spore suspensions stocks were prepared for each strain. Each strain was grown on agar plates until sporulation occurred. The spores were collected by suspending 2mL sterilized 20% glycerol in water (v/v) with 0.1% Tween 80 (v/v), as previously described to maintain as a back-up. All spore suspensions were stored in 1.2 mL cryogenic vials at - 80 ℃.

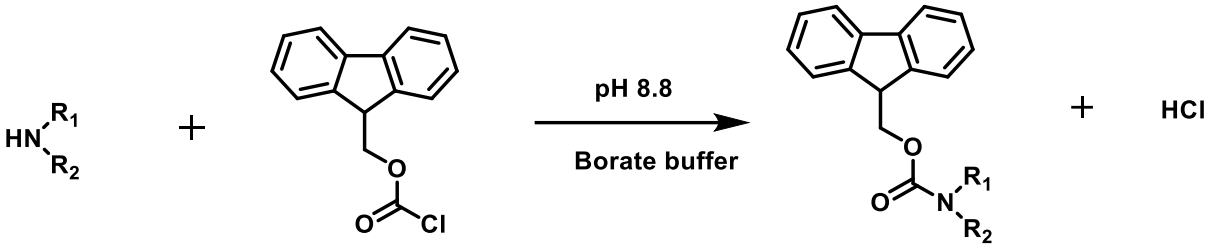
2.2.5.3 Agar/Disc diffusion assays

Agar diffusion assays against *E. coli* K12 and *E. coli* WM6242 were performed in square (120 x 120 mm) disposable petri dishes. *E. coli* WM6242 was an engineered *E. coli* strain that has phosphonate uptake transporter genes that are hypersensitive to bioactive phosphonate under the regulatory control of *Tac* promoter.[146] *E. coli* K12 was a wildtype strain that accounted for the presence of non-phosphonate-containing antibiotics. Petri dishes were filled with 50mL LB agar and mixed with 0.5mL stock of *E. coli* K12 or *E. coli* WM6242. The 0.1mL IPTG was added in petri dishes of *E. coli* WM6242. Disc diffusion assays against *B.subtilis* 168 were also performed in square (120 x 120 mm) disposable petri dishes.

2.2.6 Chemical labeling strategies

Phosphonate compounds are highly water-soluble and polar compounds. This results in a lack of sufficient hydrophobicity to allow them to enter organic solvents and a series of reverse chromatograms, such as C18, to separate them. Also, the cultures are usually complex in composition, containing large amounts of hydrocarbons such as sugars, amino acids, peptides and other metabolites. These pose a great challenge for the separation of phosphonate compounds.

2.2.6.1 Labeling of phosphonates with FMOC-Cl

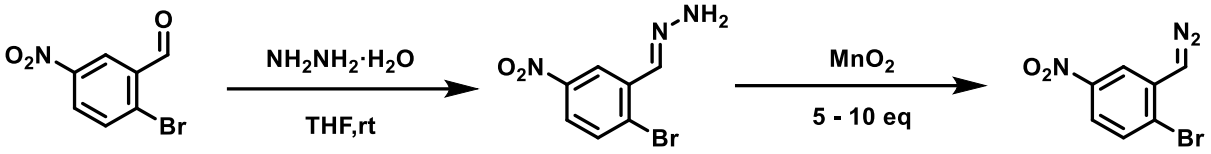


The labeling reaction of standard phosphonates was performed with Fmoc chloride. The reaction was carried out by mixing 200 µL of phosphonate standard solution, 200 µL of borate buffer (crystallized boric acid, 200 mM, pH 8.8) and 100 µL of Fmoc chloride (10mM in MeCN). The reaction was incubated at room temperature for 30 min, diluted and analyszed by LC-MS.

2.2.6.2 Labeling of phosphonates with a diazo probe

2.2.6.2.1 Diazo probe synthesis

The synthesis protocol was figured out by Dr. Chambers C. Hughes

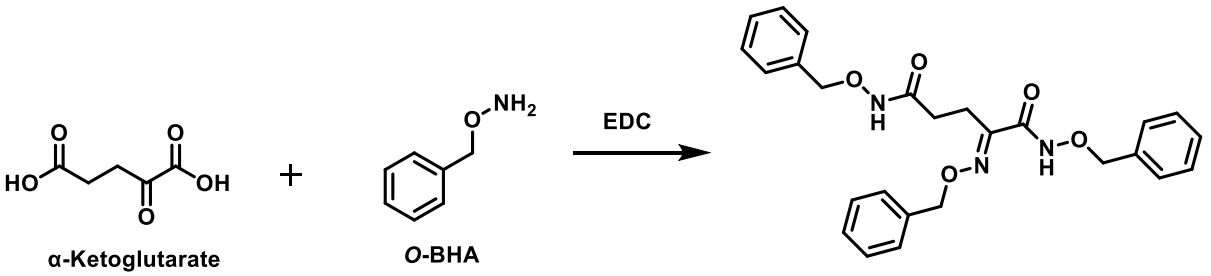


To a solution of the aldehyde in THF was added in 65% solution of hydrazine hydrate via syringe. And the reaction mixture was stirred overnight at room temperature, filtered to collect product. The solid was dried to obtain the desired hydrazone. The hydrazones were used in subsequent reactions without any further purification. Then electrolyzed MnO2 was added in the hydrazone in THF. After ~ 4 hours, the entire reaction was centrifuged to remove THF layer from MnO2 and filtered the reaction through glasswool. This THF solution was covered with aluminum foil and kept in the fridge away from light.

2.2.6.2.2 Labeling of phosphonates with a diazo probe

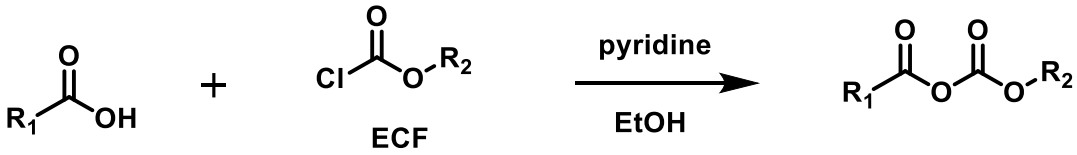
The reactions were performed in Eppendorf tubes. 200 µL phosphonate standard solution (2mg/mL solution of single phosphonate) and 100 µL diazo solution in THF was reacted with 200 µL borate buffer (pH 6.9). Different volumes of MeCN were added to it to make sure that everything was dissolved well. Then the reaction was dried under N2 and analyzed by LC-MS.

2.2.6.3 Labeling of phosphonates with O-benzylhydroxylamine(O-BHA)



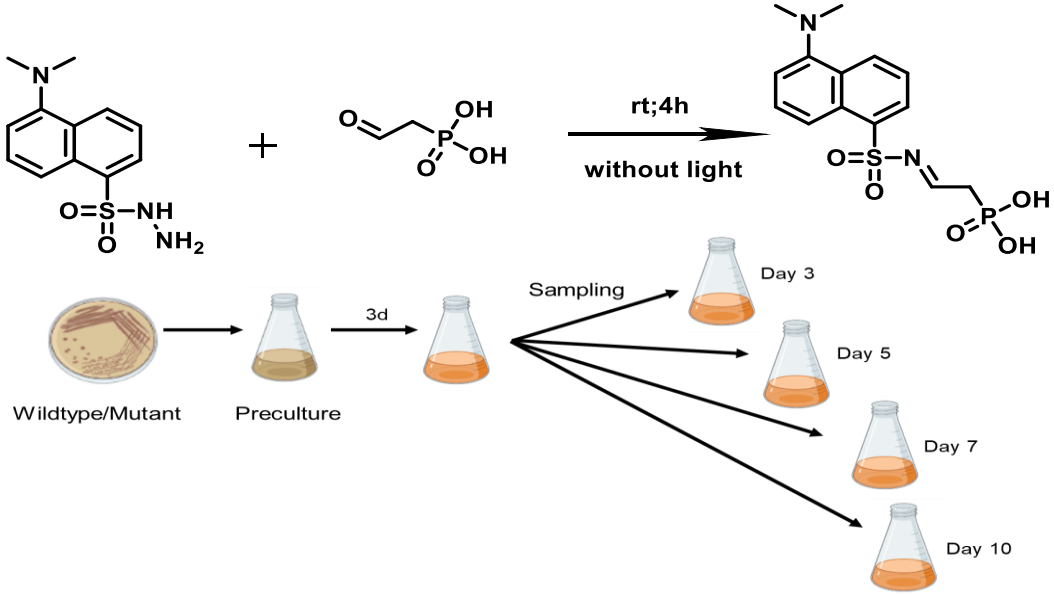
Pyridine buffer was prepared with 5.4mL of HCl (12.1M), 8.6mL of pyridine, and 86mL of water. The pH was measured with pH meter to be around 5.0. Reactions were performed in 8mL glass vials with caps. 500 µL phosphonate standard solution (1mg/mL) and α-ketoglutarate were mixed with 500 µL of 20mM O-BHA and 500 µL of 20mM EDC-HCl in the pyridine buffer. After 1h at room temperature, 1mL ethyl acetate was added and shaken. Then the organic layer was taken into another vial. The aqueous layer was extracted again with ethyl acetate. Then the organic layers were combined and dried under nitrogen. The residue was dissolved in 50% MeCN in H2O and analyzed by LC-MS.

2.2.6.4 Labeling of phosphonates with ethyl chloroformate (ECF)



Reactions were performed in 8mL glass vials with caps. 150 µL Ethanol-pyridine mixture (4:1, v/v) and 150 µL of chloroform- ECF (9:1, v/v) were vortexed with the 200 µL phosphonate standards solution (5mg/mL) to create emulsion. Then, 150 µL of 1.5N NaOH were added for neuralization first portion ECF and the content was vortexed. 150 µL of chloroform-ECF (9:1, v/v) was again added, followed by vortexing and centrifugation. The final mixture was treated with 150 µL of 3N HCl, mixed and centrifuged. The sample organic layer was transferred into another vial and evaporated to dry carefully under the N2 at room temperature. Then the dried residue was dissolved in MeOH and analyzed by LC-MS.

2.2.6.5 Carbonyl derivatization



The main culture was sampled after 3 days, 5 days, 7 days and 10 days of cultivation and centrifuged. Different supernatants were reacted with dansyl hydrazine in 8mL glass vials covering with aluminum foil at room temperature for 4 hours, diluted with MeCN and analyzed by LC-MS.

2.3 Publications

2.3.1 Publication 1

RSC Chemical Biology

ROYAL SOCIETY OF CHEMISTRY

PAPER

Check for updates

Cite this: *RSC Chem. Biol.*, 2026, 7, 298

Expanding the actinomycetes landscape for phosphonate natural products through genome mining

Alina Zimmermann,[ab] Shu-Ning Xia,[bc] Julia Moschny,[bc] Juan Pablo Gomez-Escribano,[a] Judith Boldt,[ad] Ulrich Nübel,[adef] Imen Nouioui,[a] Janina Krause,[c] Mattis Kreins Irle,[a] William W. Metcalf,[g] Chambers C. Hughes[bch] and Yvonne Mast[*abef]

Received 29th September 2025, Accepted 11th December 2025

DOI: 10.1039/d5cb00254k

rsc.li/rsc-chembio

Phosphonate natural products (P-NPs) represent a unique and underexplored class of bioactive compounds with significant pharmaceutical and biotechnological potential. Many novel P-NPs with promising bioactivities were identified in recent years by genome mining of actinomycetes. The DSMZ strain collection harbors more than 6000 actinobacterial strains including an increasing number of genome sequenced strains. In this study, 940 genome-sequenced actinomycetes from the DSMZ and University of Tübingen strain collections were screened for the presence of phosphonate biosynthetic gene clusters (P-BGCs) by searching for the conserved *pepM* gene. This effort led to the identification of 54 potential phosphonate producer strains. Subsequent bioassays with a phosphonate-sensitive *E. coli* test strain showed activity for 17 strains, and ^{31}P NMR spectroscopic analysis of culture supernatants confirmed phosphonate production for 21 strains, including the rare actinomycete *Kitasatospora fiedleri* DSM 114396^{T}. The functionality of the unique *K. fiedleri* P-BGC was verified by *pepM* gene deletion, which abolished phosphonate production in *K. fiedleri*, whereas overexpression of a cluster-situated LuxR-like regulator improved phosphonate production. These findings highlight the P-NP biosynthetic potential of actinomycetes and pave the way for discovering novel bioactive phosphonates.

Introduction

The rediscovery of known antibiotics and other specialized secondary metabolites poses a major challenge to the field of natural product (NP) discovery and development.[1] The majority of antibiotics, as well as a range of anticancer, antifungal, and immunosuppressive agents were identified between 1940 and 1970 by systematically screening soil microorganisms for production of bioactive substances.[2] This once successful pipeline is exhausted due to the frequent rate of rediscovery of known compounds.[3] The reasons for this are multifaceted and include the reliance on common sampling sites, the use of standard strain-isolation techniques, and conventional cultivation conditions, all of which often result in the re-isolation of known or closely related species with similar secondary metabolite biosynthesis profiles.[4–6] Furthermore, standard extraction protocols typically involve a limited range of organic solvents which creates a bias favoring the extraction of hydrophobic compounds.[7–9] In antimicrobial assays, extracts are usually tested against a panel of standard test organisms, such as *Escherichia coli*, *Bacillus subtilis*, *Staphylococcus aureus etc.*, which in turn also contributes to rediscovery.[10] Additionally, compounds that demonstrate prominent mass spectrometry (MS) patterns and ultra-violet (UV) spectroscopy signals are often prioritized even though they may represent "low hanging fruits" in discovery.[6,11] As a result, these interconnected practices create a cycle leading to the rediscovery of previously known NPs, accentuating the necessity of innovative approaches to overcome these challenges.

[a] *Leibniz Institute DSMZ -German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, 38124 Braunschweig, Germany. E-mail: yvonne.mast@dsmz.de*
[b] *German Center for Infection Research (DZIF), Partner Site Tübingen, 72076 Tübingen, Germany*
[c] *Department of Microbial Bioactive Compounds, Interfaculty Institute of Microbiology and Infection Medicine (IMIT), University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany*
[d] *German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, 38124 Braunschweig, Germany*
[e] *Braunschweig Integrated Centre of Systems Biology (BRICS), Rebenring 56, 38106 Braunschweig, Germany*
[f] *Technische Universität Braunschweig, Institut für Mikrobiologie, Spielmannstr. 7, 38106 Braunschweig, Germany*
[g] *Carl R. Woese Institute for Genomic Biology, 1206 West Gregory Drive, Urbana, IL 61801, USA*
[h] *Cluster of Excellence EXC 2124: Controlling Microbes to Fight Infection, University of Tübingen, 72076 Tübingen, Germany*

Despite these challenges, the remarkable potential of microorganisms for NP discovery remains undeniable. The most versatile and prolific antibiotic producers belong to the phylum *Actinobacteria*, particularly to the genus *Streptomyces*.[12–14] Most antimicrobial compounds commonly used clinically today are derived from these organisms.[2] In these microorganisms, all genes required for the biosynthesis of a particular NP, as well as genes involved in regulation, secretion, and self-resistance mechanisms, are typically located together in a continuous region of the genome, comprising the so called biosynthetic gene cluster (BGC), an arrangement first discovered in 1979 by Rudd and Hopwood for the genes responsible for actinorhodin biosynthesis in *Streptomyces coelicolor* A3(2).[3,15,16] Despite the extensive exploration of actinomycetes for NP production for the last 70 years, the advent of the genomics era has revealed that many novel compounds remain to be discovered.[17,18] Current estimates indicate that only about 3% of the genomically encoded bacterial biosynthetic diversity has been experimentally characterized, with streptomycetes in particular representing a largely untapped genetic reservoir for novel NP biochemistry.[18] Genome mining has played a pivotal role in bridging this gap by correlating genomic data with biosynthetic functions of encoded enzymes, and even in predicting the chemical structures of the resulting metabolic products.[5,19] However, only certain types of BGCs allow for straightforward and reliable prediction of biochemistry and metabolic products by bioinformatics analysis of genomic sequence, including BGCs encoding polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS).[20,21] For other kinds of biosynthetic pathways, specific marker genes can serve as molecular footprints, enabling the identification of distinct structural features and underexplored classes of specialized metabolites, thereby facilitating the discovery of novel NP chemistry.[22]

One underexplored class of specialized metabolites are phosphonate-containing NPs (P-NPs). The unifying feature of P-NPs is the direct carbon–phosphorus bond. Early discovery efforts using bioassay-guided fractionation for isolating P-NPs yielded a commercialization orders of magnitude higher than the estimated 0.1% for NPs overall,[2,40] highlighting the potential of small molecule P-NPs for pharmaceutical[41] and agricultural applications.[42] The global phosphonate market is set to grow, with many phosphonates demonstrating strong bioactivities suitable for various applications.[43,44] One example is FR900098, a derivate of fosmidomycin, which was originally isolated from *Streptomyces rubellomurinus*.[23,24] FR900098 is a candidate for malaria treatment due to its capability to inhibit the 1-deoxy-D-xylulose-5-phosphate (DXP) reductoisomerase in the non-mevalonate pathway for isoprenoid synthesis in *Plasmodium falciparum*.[25,26] Fosfomycin,[27–29] produced by *Streptomyces wedmorensis* DSM 41676^T, *Streptomyces fradiae*, and *Pseudomonas syringae*,[30,31] is a broad-spectrum antibiotic that has been used since the 1970s to treat various bacterial infections, including urinary tract and gastrointestinal infections.[32] Fosfomycin targets the initial step of peptidoglycan biosynthesis by mimicking phosphoenolpyruvate (PEP), the natural substrate of UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA).[33–35] P-NPs include phosphonates and phosphinates, with the latter featuring C–P–C or C–P–H bonds, and thereby phosphorus in an even lower redox state. Illustrative examples include the industrially used herbicide phosphinothricin tripeptide (PTT), also known as bialaphos, which is produced by *Streptomyces viridochromogenes* DSM 40736 and *Streptomyces mooreae* DSM 41527, formerly *S. hygroscopicus*.[36–38] PTT acts as an inhibitor of glutamine synthetase by releasing the unusual amino acid phosphinothricin (PT), which mimics the native substrate glutamic acid, leading to bactericidal, fungicidal, and herbicidal effects.[36,39] These compounds exemplify the diverse bioactivities of P-NPs, showcasing their potential as small molecular inhibitors.

Phosphonates and phosphinates frequently display structural similarities to primary metabolites such as phosphate esters, carboxylic acids, or other tetrahedral structures. This similarity underpins their bioactivity through "molecular mimicry", allowing phosphonates to bind the same enzymatic partners as their cognate primary metabolites and act as small molecular inhibitors across various pathways, due to the widespread occurrence of this structural feature in biology.[44] Furthermore, the unique P–C bond requires specialized catabolic pathways for the degradation into phosphates, preventing common degradation routes that involve the hydrolysis of phosphate esters or carboxylic acids.[45]

With a single exception,[46,47] all known natural phosphonate biosynthetic pathways have a common initial biosynthetic step, namely the isomerization of phosphoenolpyruvate (PEP) to phosphonopyruvate (PnPy) by the enzyme phosphoenolpyruvate mutase (PepM)[48] (Fig. 1). Due to the difference in P–O and P–C bond energies, this isomerization favors the formation of PEP at equilibrium. To generate the phosphonate product, the first reaction must be coupled to a subsequent enzymatic step that ensures thermodynamic favorability.[48] This reaction is most commonly carried out by the PnPy decarboxylase (Ppd), which converts PnPy to phosphonoacetaldehyde (PnAa) while releasing a carbon dioxide molecule[49] (Fig. 1). Phosphonate biosynthetic pathways show significant divergence after the first one or two biosynthetic steps.[49] This lack of a widely conserved biosynthetic route is also reflected in the diverse gene composition of P-BGCs. As a result, predicting phosphonate products solely based on gene sequence information is often challenging.[50] However, it has been found that the phylogeny of the conserved enzyme PepM strongly correlates with the *pepM* gene neighborhood, providing insights into phosphonate biosynthetic routes and products.[51] Conversely, PepM phylogeny does not align with organismal phylogenies, indicating that horizontal gene transfer is a significant driver for the acquisition of P-BGCs.[51] Due to the conserved nature of the initial step of phosphonate biosynthesis, together with the presence of the conserved motif $EDKX_5NS$ in the primary sequence of PepM, which distinguishes it from other members of the isocitrate lyase mutase protein family,[52–54] the *pepM* gene represents a suitable marker to screen for novel phosphonate producers.

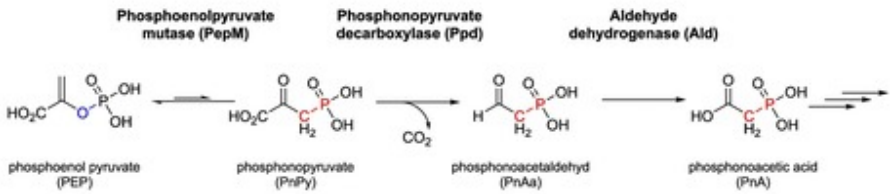


Fig. 1 Phosphoenolpyruvate mutase (PepM) catalyzes the initial step shared in phosphonate biosynthesis creating the carbon–phosphorus bond by isomerization of phosphoenolpyruvic acid (PEP) to phosphonopyruvic acid (PnPy). Subsequent steps leading to phosphonoacetic acid (PnA), a significant branchpoint in phosphonate biosynthesis, are depicted: Phosphonopyruvate decarboxylase (Ppd) converts PnPy to phosphonoacetaldehyde (PnAa), which is then oxidized to PnA by an aldehyde dehydrogenase (Adh).

PepM-based genome mining screening has been successfully applied to 10 000 actinomycetes, resulting in the discovery of 19 novel P-NPs, including valinophos, which originated from *Streptomyces durhamensis* DSM 40539^T,[55,56] argolaphos, an antimicrobial that was first isolated from *Streptomyces monomycini* DSM 41801^T,[55,57] and phosphonocystoximate.[55] Recently, additional novel P-NPs have been identified through PepM-based genome mining,[49,55–61] such as a novel phosphonopeptide from *Bacillus velezensis*,[61] a group of compounds termed phosphonoalamides isolated from *Streptomyces* sp. B-2790, *Streptomyces kutzneri* DSM 40907 and *Bacillus subtilis*,[49,60,62,63] the antibiotic phosacetamycin produced by *Streptomyces aureus*,[58] and the herbicide pantaphos from *Pantoea ananatis*.[59,64] With the growing knowledge of phosphonate biosynthetic pathways, combined with the characterization of P-BGCs and their modification through genetic engineering, a strong foundation has been established for synthetic biology approaches.[65,66] For example, the well-known herbicidal phosphonate glyphosate can now be obtained by fermenting *Streptomyces lividans* that expresses six genes from the argolaphos BGC to produce the precursor aminomethylphosphonate, which is then converted to glyphosate in one chemical step under aqueous conditions.[66]

Here, we present our results on the investigation of the phosphonate biosynthetic potential of 940 genome-sequenced actinomycetes from the DSMZ and University of Tübingen strain collections, focusing on an unexplored branch of phosphonate biosynthesis in the rare actinomycete strain *Kitasatospora fiedleri* DSM 114396^T as a representative producer.

Results and discussion

PepM screening leads to the identification of 45 unexplored phosphonate producers

In an effort to identify novel P-NPs, 940 actinomycetes genomes from the DSMZ (n = 781) and Tübingen (n = 159) strain collections were screened for the presence of *pepM*-like genes. Fifty-four strains (5.74%) from the two collections were found to contain PepM-encoding genes, defined as homologous proteins containing a perfect match for the conserved PEP mutase motif EDKX$_5$NS (Table S1). From this subset of strains, nine had already been described as producers of P-NPs, including the ones mentioned above along with the phosalacine producer *Kitasatospora phosalacinea* DSM 43860^T, 2-phosphinomethylmalic acid and desmethylphosphinothricin producer *Nonomuraea candida* DSM 45086^T, and *Stackebrandtia nassauensis* DSM 44728^T, a producer of phosphonoglycans. The remaining 45 strains did not show any association with known P-NPs, warranting further investigation.

Prioritization of potential novel P-NP producers

To prioritize potential producers of novel P-NPs, we analyzed the PepM amino acid sequences for phylogenetic diversity. By including sequences from known phosphonate producers, we were able to assign clusters of high similarity to known compounds (Table S1). Our analysis revealed distinct phylogenetic branches, with several branches containing sequences from known P-NP producers, suggesting that phylogenetic proximity might be used as a criterion for identifying producers of known P-NPs (Fig. 2). For instance, one branch encompassed the known phosphonoalamide producer strains *Streptomyces* sp. NRRL B-2790[49] and *Streptomyces kutzneri* DSM 40907,[62] as well as *Streptomyces resistomycificus* DSM 40133^T and *Streptomyces badungensis* DSM 114471, suggesting the latter two may also produce phosphonoalamides. Another branch includes the phosphonothrixin producers *Streptomyces* sp. WM4235[67] and *Saccharothrix* sp. ST-888,[68] as well as *Streptomyces* sp. TÜ 21470 and *Kitasatospora purpeofusca* DSM 40283^T; for the latter strain, an *ftx* gene cluster responsible for synthesizing phosphonothrixin has been reported.[67] A branch characterized by low genetic divergence in the tree, as evidenced by short branch lengths, contained the hydroxynitrilaphos producer *Streptomyces regensis* NRRL WC-3744,[69] along with five *Streptomyces* strains that are all likely hydroxynitrilaphos producers. The phylogenetic proximity of this branch to the phosphonocystoximate producer *Streptomyces* sp. NRRL S-481 aligns well with the previously reported similarity between the biosynthetic pathways leading to hydroxynitrilaphos and phosphonocystoximate and between their P-BGCs (Fig. 2).[55] Only one *pepM* gene was found per genome, except for three strains - *Streptomyces* sp. I6, *Nocardia tenerifensis* DSM 44704^T, and the pantaphos producer *P. ananatis* LMG 5342 - each containing two distinct P-BGCs with their own *pepM* gene, adding support to the hypothesis that PepM is a good marker for P-NP chemistry, since the same PepM does not seem to be shared by two different pathways in the same organism.

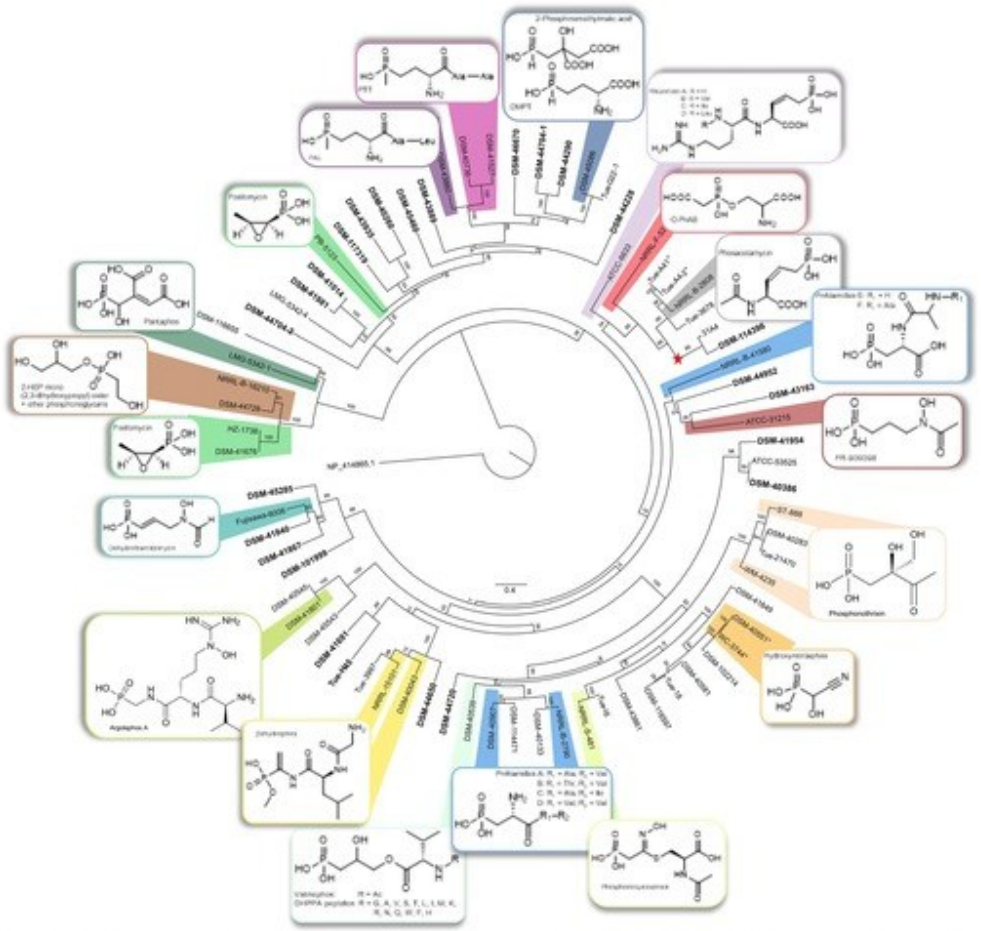


Fig. 2 Maximum-likelihood tree of PepM amino acid sequences, inferred under the LG + I + G4 + F model and 1000 bootstrap iterations using RAxML v. 8.2.12,[84,85,91] rooted by the outgroup 2-methylisocitrate lyase (NP_414865.1). The maximum likelihood (ML) phylogeny was inferred from a pairwise alignment generated using MAFFT v.7.[82,83] The branches are scaled in terms of the expected number of substitutions per site. The numbers above the branches are support values from ML bootstrapping which were calculated with branch lengths. Nodes with sequences from known P-NP producers are color-coded and chemical structures of known products are displayed. Strain designations highlighted in bold are likely members of novel and unique Gene Cluster Families (GCFs). This classification is based on their lack of association with sequences from known P-NPs producers, as determined by a similarity cutoff of 0.612 in the BiG-SCAPE analysis. DHPPA = 2,3-dihydroxypropylphosphonate; 2-HEP = 2-hydroxyethylphosphonate; *O*-PnAS = *O*-phosphonoacetic acid serine; PAL = phosalacine; PnAlamides = phosphonoalamides; PTT = phosphinothricin tripeptide. The node containing the PepM sequence of *K. fiedleri* DSM 114396^T, a putative novel producer of a phosphonoacetic acid-derived P-NP, was marked by a red star. *PepM sequence from TÜ A41 is identical to TÜ A4-2 and PepM sequence from DSM 40551^T is identical to WC 3744.

To support these findings and to further explore the uniqueness of the identified P-BGCs, a gene cluster network analysis was performed with BiG-SCAPE 2.0.[70] All P-BGCs predicted by antiSMASH 7.1.0[53] for the strains of the DSMZ and University of Tübingen collections, including 54 putative P-NP producers and 18 additional known producers previously reported in literature, were included in the analysis. Using a cutoff value of 0.612, we identified 45 gene cluster families (GCFs), of which 28 were singletons (Fig. S1 and Table S1). At this stringency, the resulting GCFs reflected established relationships among phosphonate-producing organisms: BGCs encoding structurally related phosphonates *via* similar biosynthetic pathways grouped together, whereas more divergent clusters remained separated. The largest cluster at this threshold comprised seven

P-BGCs (GCF01), some of which were closely related to BGCs in GCF02 and GCF03, reflecting the high similarity shared between the pathways.[55] Among the strains with these GCFs are the known hydroxynitrilaphos producer *Streptomyces regensis* NRRL WC-3744[69] (GCF01) and the phosphonocystoximate producer *Streptomyces* sp. NRRL S-481[55] (GCF02). The network analysis also enabled the detection of previously characterized P-BGCs in new species. For example, GCF12 linked the known pantaphos P-BGC from *P. ananatis* LMG 5342 with a previously unknown type-VI Hvr-like P-BGC in *Streptomyces thermoviolaceus* DSM 116655 (Fig. S1). In our PepM tree, the *S. thermoviolaceus* and *P. ananatis* sequences clustered as nearest neighbors, reflecting the limited distribution of *hvr* operon homologs in proteobacteria and actinobacteria, as also shown by Polidore *et al.*[59]

Five GCFs, each containing two P-BGCs, were not connected to any known producer in the network analysis. Among the 28 identified singletons, only ten corresponded to known P-NP producers, such as the valinophos producer *Streptomyces durhamensis* DSM 40539^T or the FR900098 producer *Streptomyces rubellomurinus* ATCC 31215.Overall, sequences from 24 strains across 22 GCFs/singletons, out of a total of 45 GCFs, showed no network connections to any known producers, suggesting they encode novel P-BGCs with low similarity to characterized ones. Therefore, these 24 strains were identified as likely producers of novel P-NPs (Fig. S1).

To assess P-NP production, 35 potential P-NP producers (the subset that had genome data available with the exclusion of biosafety level 2 strains) were cultivated on a panel of eight to ten different solid cultivation media, and antimicrobial assays by agar-block diffusion were conducted using the phosphonate-sensitive *Escherichia coli* strain WM6242,[24] as well as *E. coli* K12 (parental control) and *Kocuria rhizophila* DSM 348 as Gram-negative and Gram-positive test organisms, respectively (Table S2). *K. rhizophila*, formerly known as *Micrococcus luteus*, is commonly employed in antibiotic susceptibility testing and for the detection of antibiotic residues in food products. The fosfomycin producer *Streptomyces fradiae* DSM 40943 was included as a positive control. Culture supernatants from *S. fradiae* DSM 40943 exhibited stronger activity against the phosphonate-sensitive *E. coli* WM6242 strain than against *E. coli* K12, suggesting that the observed inhibition was likely due to the production of fosfomycin, expected to be more active against WM6242 due to the induced expression of a phosphonate uptake system.[24,71,72] Of the strains sharing a GCF with known phosphonate producers in the gene cluster network analysis, 19 were tested, of which 12 showed activity against *E. coli* WM6242, and seven of them additionally exhibited activity against *E. coli* K12. Given the prediction that these strains likely produce P-NPS similar to known compounds, this may reflect an investigation bias towards readily produced P-NPs. Of the 16 presumed novel P-NP strains included in the bioassay tests, five showed bioactivity against *E. coli* WM6242. Notably, all five strains that inhibited *E. coli* WM6242 also showed weaker activity against *E. coli* K12. This suggests either enhanced uptake of a single antimicrobial compound – such as a phosphonate – by *E. coli* WM6242, or the production of multiple bioactive metabolites, consistent with the high number of predicted BGCs per strain (36 on average, although some predictions may be influenced by contig fragmentation).

Eighteen strains showed no activity against the phosphonate-sensitive *E. coli* WM6242 under any tested conditions, suggesting that (i) the encoded P-NPs may lack antibiotic activity, (ii) were produced at low concentrations, (iii) the corresponding P-BGCs remained silent under the applied cultivation conditions, or (iv) that the phosphonate compound may not be active against *E. coli* WM6242 regardless of the overexpressed phosphonate uptake system (Table S2). Of the 18 strains, 12 exhibited antimicrobial activity against *K. rhizophila*, suggesting the production of non-phosphonate antimicrobials.

Confirmation of phosphonate production by 21 actinomycetes using ^{31}P NMR analysis

A subgroup of 26 strains was selected for phosphonate production analysis with ^{31}P NMR. Phosphonate-containing compounds show characteristic ^{31}P NMR chemical shifts above +8 ppm, in contrast to most other phosphorus-containing biomolecules, which typically resonate between −25 and +5 ppm. This distinct difference provides a reliable and convenient method to detect phosphonate compounds in complex culture samples.[40] The 26 strains were grown in 10 different cultivation media and supernatants were subjected to ^{31}P NMR analysis. 21 strains displayed spectra with chemical shifts consistent with phosphonates under at least two cultivation conditions. Confirmed phosphonate production included the known argolaphos producer *Streptomyces monomycini* DSM 41801^T, along with other established phosphonate producers and associated producers, such as the recently identified phosphonoalamide producers *Streptomyces kutzneri* DSM 40907[62] and *Streptomyces resistomycificus* DSM 40133^T,[49] as well as two strains from the largest cluster in the gene network analysis (GCF01-GCF03), likely associated with hydroxynitrilaphos or phosphonocystoximate production in *Kitasatospora setae* DSM 43861^T and *Kitasatospora atroaurantiaca* DSM 41649^T, respectively. Phosphonate production was detected for nine out of the ten tested strains from the likely novel P-NP producer group (24 likely novel P-NP producer strains), including *Kitasatospora fiedleri* DSM 114396^T, *Streptomyces aureocirculatus* DSM 40386^T, and *Streptomyces iranensis* DSM 41954^T (Fig. S2–S6). For the latter (Fig. S3), we have recently demonstrated that inactivation of the *pepM* gene using a newly developed cloning vector system results in the loss of phosphonate production, proving the functionality of the identified P-BGC in *S. iranensis* DSM 41954^T.[73]

Characterization of the phosphonate biosynthetic gene cluster in the rare actinomycete *Kitasatospora fiedleri* DSM 114396

The rare actinomycete *Kitasatospora fiedleri* DSM 114396^T demonstrated phosphonate production, evidenced by characteristic chemical shifts in ^{31}P NMR (Fig. S6). We recently described this strain as a novel type strain and published a high-quality genome assembly (NCBI accession GCA_948472415).[74] In the PepM-phylogenetic tree, the node containing the PepM amino

acid sequence from the *K. fiedleri* P-BGC did not cluster with any known P-NP producer (Fig. 2). Consistently, in the gene cluster network analysis, the *K. fiedleri* P-BGC clustered together with that from *Streptomyces* sp. 31A4[51,55] in GCF09 but showed no connection to a cluster of any known P-NP producer (Fig. S1).

A gene synteny analysis performed with NCBI BLAST against the type strains of *Kitasatospora cineracea* DSM 44780^T and *Kitasatospora niigatensis* DSM 44781^T, the most closely related type strains to *K. fiedleri* and lacking a P-BGC,[74] suggested that the P-BGC most likely spans from position 3401998 (stop codon of CDS *kfp*01 with locus_tag QMQ26_RS15690 encoding a putative aldehyde dehydrogenase) to position 3428393 (start codon of CDS *kfp*25 with locus_tag QMQ26_RS15810 encoding a putative NUDIX domain-containing protein) (Fig. S7 and S11). An analysis with clinker[75] of all putative P-BGCs highly similar to *K. fiedleri*'s (22 records in NCBI's "RefSeq Genome Database", from 21 strains, 3 of which belong to *K. fiedleri* species) confirmed the above mentioned boundaries of the P-BGC (Fig. S8). The presence of this P-BGC across phylogenetically diverse members of the *Streptomycetaceae*, coupled with its absence in the closest phylogenetic neighbors *K. cineracea* and *K. niigatensis*, suggests that *K. fiederli* likely acquired the cluster *via* horizontal gene transfer. A detailed analysis of the P-BGC DNA sequence supports the accuracy of both the sequence and its gene annotation (Fig. S8 and S9). All genes present in the suggested P-BGC run in the same orientation, and with little or no intergenic space, indicating they might form an operon (Fig. S9). The gene encoding the putative PepM enzyme is located between position 3404408 and 3403515 of the deposited chromosome sequence (*kfp*02, locus_tag QMQ26_RS15695). Notably, the cluster contains a gene *kfp*24, encoding a putative transcriptional regulator of the LuxR family (locus_tag QMQ26_RS15805) (Fig. 4 and Fig. S8). This gene is located immediately upstream of the BGC, shares the same transcriptional orientation as the downstream genes, and is separated from them by a 670 bp intergenic region, likely harboring the main promoter driving expression of the BGC. The *kfp*24 gene was also the only putative transcriptional regulatory gene identified within the entire genomic region proposed to encompass the complete P-BGC of *K. fiedleri*.

Genetic manipulation of the P-BGC confirms phosphonate biosynthesis in *Kitasatospora fiedleri* DSM 114396^T

To confirm the role of the identified *K. fiedleri* P-BGC in phosphonate biosynthesis, the genes *kfp*02-04, encoding the predicted first two steps of the pathway, *pepM-ppdAB*, were inactivated by replacement with a kanamycin resistance gene (*neo*) *via* homologous recombination with construct pDS0107 (Table S3). PCR analysis confirmed the gene replacement, verifying the successful generation of the deletion mutant *K. fiedleri* Δ*pepM-ppdAB*::*neo* (YM0173). Under the previously defined production conditions, YM0173 failed to produce any compounds with phosphonate-specific peak patterns in the ^{31}P NMR (Fig. 3B). Complementation of the deletion mutant with ectopic expression of *pepM-ppdAB* (*kfp*02-*kfp*04), using pRM4- (pDS0105) or pIJ10257-based constructs (pDS0106), restored phosphonate production, as measured by ^{31}P NMR (Fig. 3C). These results confirmed that the identified P-BGC is responsible for the production of P-NP compound(s) in *K. fiedleri*.

A common strategy to enhance metabolite production is the overexpression of genes encoding the first enzymes of the pathway. Thus, we mobilized the *pepM-ppdAB* overexpression constructs pDS0105 (pRM4_*kfp*02-*kfp*04) and pDS0106 (pIJ10257_*kfp*02-*kfp*04) to *K. fiedleri* by conjugation. Cultivation of the resulting strains led to a differential peak pattern in ^{31}P NMR, with increased signal intensity and the appearance of new chemical shifts relative to the wild-type strain (Fig. 3A and D). To establish the connection between these three enzymes and phosphonate production, we introduced pDS0105 and pDS0106 into the heterologous hosts *Streptomyces albus* and *Streptomyces lividans*, which do not inherently produce phosphonates. Following cultivation, phosphonate-specific chemical shifts were observed for the heterologous expression strains *S. albus* and *S. lividans* in ^{31}P NMR, confirming the role of the introduced genes in phosphonate biosynthesis (see SI Fig. S10).

As detailed above, we identified *kfp*24 as a gene encoding a putative pathway-specific transcriptional regulator of the LuxR-family. Members of this family typically function as transcriptional activators. Therefore, it was expected that overexpression of *kfp*24 should lead to overproduction of the metabolic product of the pathway. Thus, the gene *kfp*24 was cloned under the control of the constitutive promoter *ermE*p* in either pRM4 or pIJ10257 integrative vectors (resulting in pDS0101 and pDS0102, respectively) (Table S3). The *kfp*24-overexpression constructs pDS0101 and pDS0102 were mobilized to *K. fiedleri* wild-type to generate strains YM0107 and YM0108, respectively. Subsequently, the *kfp*02-*kfp*04 (*pepM-ppdAB*) overexpression constructs pDS0105 and pDS0106 were introduced into YM0107 and YM0108, respectively, resulting in strains YM0167 and YM0168, respectively (Tables S3 and S4). ^{31}P NMR analysis of culture supernatants from all these mutants resulted in increased and new phosphonate-specific chemical shifts in ^{31}P NMR, supporting that the LuxR-family regulator Kfp24 is the pathway-specific transcriptional activator of the P-BGC (Fig. 3E and F). Here, ^{31}P NMR signals are interpreted strictly as evidence for phosphonate biosynthesis, leaving the precise accumulation of the terminal product and the full extent of the Kfp24 regulon to be determined. Furthermore, we intentionally avoided making strong claims about bioactivity because the assays were performed on complex extracts, and it would be inappropriate to attribute activity to the phosphonate without unambiguous compound-level confirmation. While the precise structure and bioactivity of the phosphonate compound(s) produced by *K. fiedleri* remains under investigation, our combined genetic, bioinformatic, and ^{31}P NMR data unequivocally demonstrate the production of a phosphonate metabolite. Isolation and structure elucidation efforts are ongoing and will be reported separately.

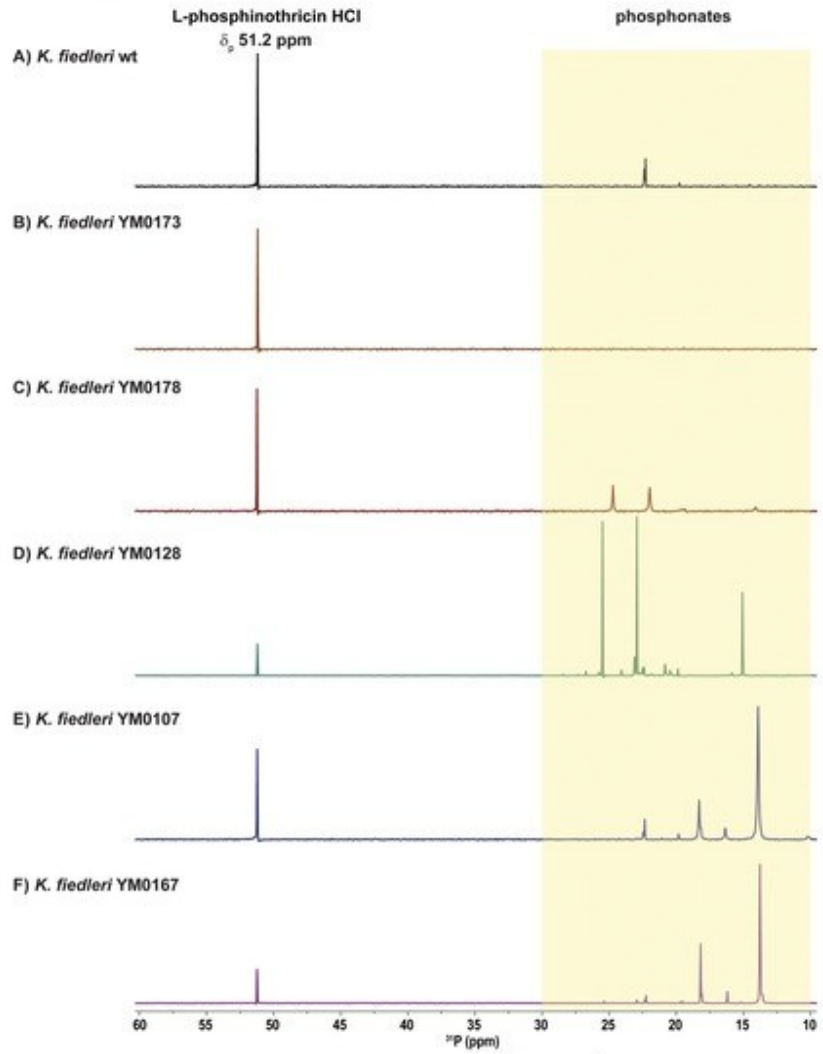


Fig. 3 ^{31}P NMR spectra of concentrated culture supernatants from *Kitasatospora fiedleri* DSM 114396^T and mutant strains: (A) wildtype *K. fiedleri* strain. (B) Deletion mutant *K. fiedleri* Δ*pepM-ppdAB*::*neo* (YM0173). (C) Deletion mutant complemented with *pepM-ppdAB* (YM0178). (D) *pepM-ppdAB* overexpression mutant (YM0128). (E) *luxR*-overexpression mutant (YM0107). (F) *pepM-ppdAB* and *luxR*-overexpression mutant (YM0167). L-Phosphinothricin HCl (10 mM in H_2O, δ_p 51.2 ppm) was used as a chemical shift reference.

K. fiedleri DSM 114396^T contains a unique P-BGC putatively linked to a phosphonoacetic acid-derived P-NP pathway

K. fiedleri DSM 114396^T harbors a unique P-BGC that network analysis placed alongside *Streptomyces* sp. 31A4 within GCF09 (Fig. S1). PepM from *Streptomyces* sp. 31A4 clustered very closely together with PepM from *Streptomyces* sp. MMG1121,[55] and all three P-BGCs share the same synteny with high sequence identity (Fig. 4 and Fig. S8A–D). All three P-BGCs carry genes encoding the canonical PepM accompanied by Ppd (and in all cases this enzyme is encoded by two separate genes). Notably, all three P-BGCs also carry a gene encoding a putative aldehyde dehydrogenase (Adh), which, in the case of *K. fiedleri* (gene *kfp*01), we

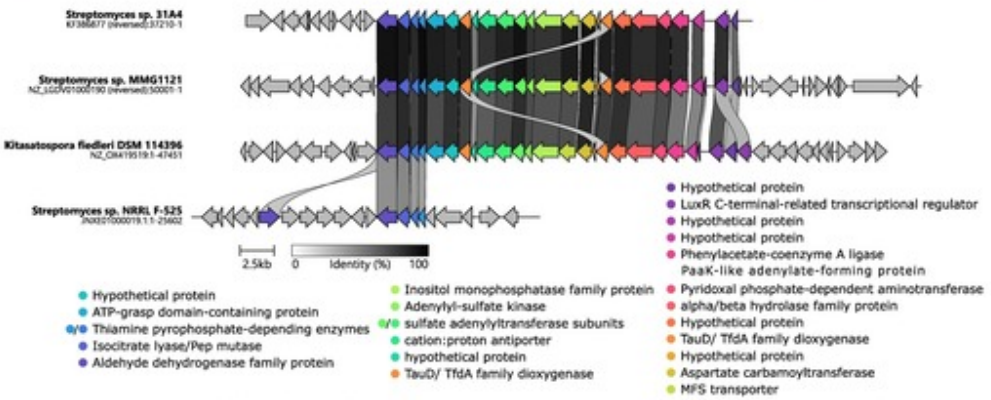


Fig. 4 Clinker alignment of the P-BGC of *K. fiedleri* DSM 114396^T and previously identified similar P-BGCs from Streptomyces sp. 31A4, Streptomyces sp. MMG112, and the well characterized P-BGC from Streptomyces sp. NRRL F-525. The orientation of the genes is given as for *K. fiedleri* chromosome sequence NZ_OX419519. The four first genes of the P-BGC, the aldehyde dehydrogenase, PEP mutase, and two encoding the phosphonopyruvate decarboxylase, are fully conserved in arrangement across all four P-BGCs. The second putative aldehyde dehydrogenase found in proximity in Streptomyces sp. NRRL F-525 is also highlighted due to the high identity (discussed with Fig. S12).

propose as one of the boundaries of the P-BGC, as detailed above and in the SI. The amino acid sequence of Kfp01 shares 82.32%, 81.88%, and 57.04% identity with the homologs from *Streptomyces* sp. 31A4, *Streptomyces* sp. MMG1121, and *Streptomyces* sp. NRRL F-525, respectively. The later has been investigated by Freestone *et al.*, 2017 which found the strain to be a producer of *O*-phosphonoacetic acid serine (*O*-PnAS) but the P-BGC that is shared between *K. fiedleri*, *S.* sp. 31A4 and *S.* sp. MMG1121 clearly extends beyond the P-BGC responsible for *O*-PnAS production in *S.* sp. NRRL F-525 (Fig. 4). All of these predicted enzymes have fully conserved catalytic and phosphonate-binding pocket residues, as compared to the well-characterized enzyme PhnY from *Sinorhizobium meliloti* 1021 (Fig. 4 and Fig. S6).[76] Based on this conservation, and as previously proposed for *Streptomyces* sp. 31A4,[65] we propose that the *K. fiedleri* DSM 114396^T P-BGC encodes the biosynthesis of a phosphonate natural product with phosphonoacetic acid as an intermediate, as depicted in Fig. 1.

A highly conserved subcluster encodes 3'-phosphoadenosine-5'-phosphosulfate metabolism

The gene products of *kfp*10 to *kfp*12 show high sequence similarity to widely conserved bacterial enzymes, including two putative sulfate adenylyltransferase subunits (*kfp*10, *kfp*11), and an adenylyl-sulfate kinase (*kfp*12). Along with *kfp*13, encoding a putative inositol monophosphatase family protein, this subcluster shows similarity to the *cysDNCQ* genes in *E. coli*[77] based on predicted amino acid sequence comparisons. The encoded enzymes are involved in synthesizing and recycling the universal sulfur donor 3′-phosphoadenosine-5′-phosphosulfate (PAPS), a key precursor for cysteine and methionine biosynthesis. The predicted functions for *kfp*10-13 indicate that all genes necessary for the synthesis, accumulation and recurrent utilization of PAPS are present in the *K. fiedleri* P-BGC. Although associated with primary metabolism, this set of enzymes may have adapted to function in specialized metabolism. Notably, components of the *cysDNCQ* genes have been identified in the BGCs of mitomycin C and azinomycin B, suggesting a role in specialized metabolism.[78,79] Such evolutionary transitions from primary to specialized metabolism have been reported before; for example, in *S. viridochromogenes* the aconitase-like enzyme Pmi has evolved to specialize in PTT synthesis, likely diverging from the typical TCA cycle function carried out by its counterpart, AcnA.[80]

Furthermore, BLASTp analysis of the amino acid sequences of four genes – *kfp*05, *kfp*07, *kfp*17, and *kfp*20 – in the *K. fiedleri* P-BGC predicted their function as putative enzymes, involved in previously characterized phosphonate biosynthetic pathways.[51,58,60,63,78] This included *kfp*05, which is predicted to code for ATP-grasp-domain-containing protein. ATP-grasp proteins are known for catalyzing ATP-dependent peptide bond formation between amino acids, as well as for mediating unique reactions such as esterification of valine onto a primary alcohol by VlpF in the valinophos producer *S. durhamensis*, resulting in the formation of 2,3-dihydroxypropylphosphonate-valine (DHPPA-Val) (Fig. S13A).[56,58,60,63] In the *K. fiedleri* P-BGC, the *kfp*07 and *kfp*17 genes encode putative TauD/TfdA family dioxygenases. Based on previous studies on *Burkholderia pseudomallei*, these enzymes are proposed to hydroxylate PnA to 2-hydroxy-phosphonoacetate (2-HPnA), suggesting a role of 2-HPnA as an alternative head group for phosphonolipids in *Burkholderia* species (Fig. S13B).[51] Additionally, *kfp*20 encodes a putative pyridoxal phosphate (PLP)-dependent aminotransferase. This type of enzyme is known to function as amino-group donors and has been implicated in phosphonopeptide synthesis, specifically catalyzing the transamination of PnPy to phosphonoalanine (PnAla) during phosphonoalamide

biosynthesis (Fig. S13C).[49,63] Altogether, the predicted functions of the *K. fiedleri* P-BGC genes suggest that the cluster encodes a phosphonopeptide biosynthesis pathway derived from PnA. However, for a substantial proportion of the BGC genes, the subsequent steps in biosynthesis remain unclear, suggesting that the cluster-encoded product may represent a novel P-NP.

Conclusion

Our investigation focused on uncovering novel P-NP producers from actinomycetes of the DSMZ and Tübingen strain collection. Screening 940 genome-sequences revealed 54 strains harboring *pepM*, of which 45 represent previously uncharacterized P-NP producers. Based on PepM phylogeny and gene cluster network analysis, we prioritized 24 strains with unique biosynthetic potential. Bioassays confirmed bioactivity for 17 of putative P-NP producer strains. ^{31}P NMR analysis of culture supernatants confirmed the production of phosphonates in 21 actinomycetes, nine of which identified as potential producers of novel P-NPs through bioinformatic analysis, supporting the effectiveness of our approach. Among these, the rare actinomycete *K. fiedleri* DSM 114396^{T} was selected for further analysis due to its unique P-BGC. Deletion of the *pepM*-like gene *kfp*02 BGC abolished phosphonate production, whereas genetic complementation restored it, as confirmed by ^{31}P NMR, thereby validating the functionality of the P-BGC in phosphonate biosynthesis. Furthermore, overexpression of a cluster-situated LuxR-like regulator gene (*kfp*24) and minimal P-BGC genes (*kfp*02-04) improved the phosphonate-specific signal pattern in ^{31}P NMR. Detailed cluster analysis points to a phosphonopeptide as the likely biosynthetic product.

Altogether, our results illustrate the effectiveness of combining genome-based phylogenetic analysis, gene cluster networking, and bioassays to prioritize and explore strains for novel phosphonate discovery. Phosphonate isolation and structure elucidation is ongoing and will be reported in due course.

Experimental

Identification of genome-sequenced actinomycetes harboring a P-BGC

A total of 940 genome sequences from the DSMZ and Tübingen actinomycetes strain collections were screened for the presence of P-BGCs using antiSMASH (versions 6.0 to 7.1.0;[53,81] the specific version deployed on the date of genome availability). Screening was performed manually as new sequencing data became available, with each genome reviewed for predicted P-BGCs and phosphonate-like clusters.

Identification of *pepM* genes

A total of 72 genome sequences of the studied strains (n = 54) and additional known phosphonate producers (n = 18) were subjected to antiSMASH analysis version 7.1.0,[53] using GenBank annotation files from the assemblies published with NCBI as input and relaxed strictness settings. Sequences from *Streptomyces* sp. 31A4 and *Streptomyces silvensis* ATCC 53525 were included in the analyses to explore branches of interest with low phylogenetic proximity to characterized P-NP producers. Neither strain is held by DSMZ or the University of Tübingen collections. The secondary metabolite clusters identified as phosphonate or phosphonate-like clusters were screened for the presence of the *pepM* gene using NCBI BlastP and evaluated to ensure that the encoded amino acid sequences conform to the consensus motif of PepM ($EDKX_5NS$).

Phylogenetic tree construction

A multiple sequence alignment of 74 PepM amino acid sequences and the sequence of 2-methylisocitrate lyase (NP_414865.1), which represents an outgroup to PepM, was created with MAFFT v.7 using the E-INS-I consistency-based method with default settings.[82,83] Maximum likelihood (ML) phylogeny was interfered from the alignment with RAxML v. 8.2.12.[84] For ML, rapid bootstrapping and subsequent search for the best tree was used. The best amino acid substitution matrix was determined with empirical frequencies with RAxML using a ML starting tree.[85]

Network analysis of P-BGCs

The P-BGC regions (n = 74) that were predicted by antiSMASH 7.1.0 to yield a phosphonic acid product were clustered into GCFs with BiG-SCAPE version 2.0[70] using a cut-off value of 0.612, including singletons in the output, and allowing the mixing of all classes of natural products (–mix). The BiG-SCAPE cutoff value (here, 0.612) sets the threshold for the maximal distance of links that group BGCs into families based on their overall similarity in domain content, arrangement, and sequence. The value was chosen to best reflect known relationships among P-NP producers—at this threshold, closely related clusters (with similar products and pathways) are grouped together. The gene cluster network obtained from BiG-SCAPE was visualized with Cytoscape v.3.10.3.

Bacterial strains, plasmids and cultivation conditions

Plasmid vectors and constructs, and bacterial strains, are listed in Tables S3 and S4 respectively. pGus21 and pRM4 and their DNA sequence were a gift from Günther Muth[86] (Eberhard Karls Universität Tübingen, Geschwister-Scholl-Platz, 72074 Tübingen, Germany). pIJ10257 was a gift from JIC StrepStrains (John Innes Centre, Norwich Research Park, NR4 7UH, Norwich, UK). pTC192km was as gift from Antonio Rodriguez (University of Leon, Spain). *Streptomyces albus* J1074 and *Escherichia coli* ET12567/pUZ8002[87,88] were a gift from JIC StrepStrains. *Escherichia coli* WM6242 was a gift from William Metcalf (University of Illinois at Urbana-Champaign, USA). *Streptomyces lividans* T7 was sourced from the Tübingen strain collection. DSM strains were sourced from the German Collection of Microorganisms and Cell Cultures (Leibniz-Institut DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Inhoffenstraße 7 B, 38124 Braunschweig, Germany).

E. coli cultivation and general molecular biology techniques were performed following established methods[89] and instructions provided by suppliers. *Streptomyces* strains were cultivated in SFM (MS) for preparation of spore stocks, and for mobilization of plasmid constructs to *Streptomyces* strains by conjugation from *E. coli*, all of which was done according to established methods.[90] Apramycin and kanamycin were used at 50 mg L^{-1}, fosfomycin at 20 mg L^{-1}, final concentration.

Actinomycetes strains were cultivated for phosphonate production in the following solid media, for 7 days at 28 °C or until confluent growth was reached: OM medium (20 g oat meal, 1 g K_2HPO_4, 20 g agar and 5 mL trace metal mix 1 dissolved in 1 L distilled water, pH adjusted to 7.3. Trace metal mix 1 : 2 g $FeCl_3 \cdot 6H_2O$, 0.1 g $CaCl_2 \cdot 2H_2O$, 0.1 g $MnCl_2 \cdot 2H_2O$, 0.1 g $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 0.1 g $Na_2B_4O_7 \cdot 10H_2O$, dissolved in 1 L distilled water), HM medium (4 g yeast extract, 10 g malt extract, 4 g glucose and 1 g K_2HPO_4 and 16 g agar dissolved in 1 L distilled water, pH adjusted to 7.3), SFM medium (20 g mannitol, 20 g soy flour and 1 g K_2HPO_4, 20 g agar dissolved in 1 L distilled water, pH adjusted to 7.3), R5 medium (103 g saccharose, 10 g glucose, 0.25 g K_2SO_4, 10.12 g $MgCl_2$, 0.1 g casamino acids, 5 g yeast extract, 5.73 g TES, 18 g agar and 2 mL trace metal mix 1 dissolved in 955 mL distilled water, pH adjusted to 7.3), NL200 medium (20 g mannitol, 20 g cornsteep powder and 1 g KH_2PO_4 and 16 g agar dissolved in 1 L distilled water, pH adjusted to 7.5), NL300 medium (20 g mannitol, 20 g cottonseed powder and 1 g KH_2PO_4 and 16 g agar dissolved in 1 L distilled water, pH adjusted to 7.5), NL400 medium (20 g mannitol, 20 g cottonseed powder and 1 g KH_2PO_4 and 16 g agar dissolved in 1 L distilled water, pH adjusted to 7.5), NL410 medium (10 g glucose, 10 g glycerol, 5 g oat meal, 10 g soy meal, 5 g yeast extract, 5 g casamino acids, 1 g $CaCO_3$, 1 g KH_2PO_4 and 16 g agar dissolved in 1 L distilled water, pH adjusted to 7.0), NL500 medium (10 g soluble starch, 10 g glucose, 10 g glycerol, 15 g fish meal, 10 g sea salts and 16 g agar dissolved in 1 L distilled water, pH adjusted to 8.0 and rechecked after 1 h) and NL800 medium (5 g glucose, 10 g glycerol, 10 g soluble starch, 5 g soy flour, 2 g yeast extract, 1 g NaCl, 1 g $CaCl_2$ and 16 g agar dissolved in 1 L distilled water, pH adjusted to 7.2).

Bioassay with phosphonate-sensitive *E. coli* strain

Antibiotic activity was analyzed in agar-block diffusion assays using *E. coli* WM6242, *E. coli* K12, and *Kocuria rhizophila* DSM 348 as indicator organisms. Test plates were prepared in L-agar medium (for all strains including non-induced *E. coli* WM6242) or LB-G medium supplemented with IPTG to a final concentration of 200 μM (for phosphonate transporter induced *E. coli* WM6242). For *E. coli* strains, an overnight liquid culture of the indicator organism in L medium was used to inoculate a day culture (also in L medium), which was incubated at 37 °C and 180 rpm orbital shaking until reaching an OD_{600} of around 0.8; this day culture was used to inoculate the molted agar media at 0.1% (v/v) concentration. For *K. rhizophila* test plates, molted L-agar was inoculated with overnight culture at 0.1% (v/v) concentration. Agar blocks with a diameter of 8 mm were excised from sections of the plates displaying mature, confluent growth and placed on the test plates inoculated with the test organisms, followed by incubation overnight at 37 °C. After incubation, the diameters of all observed inhibition zones were measured. Due to the large number of samples, and the qualitative nature of the experiment, only one biological replicate was prepared for each condition.

Molecular cloning

Restriction endonucleases, T4-DNA ligase, Q5 and One-Taq DNA polymerases, DNA Polymerase I-Large (Klenow) Fragment, were purchased from New England Biolabs (New England Biolabs GmbH, Brüningstraße 50; Geb. B852, (Industriepark Höchst), D-65926 Frankfurt am Main, Germany) and used according to provider's instructions. Sequences of oligonucleotides used in this study are given in Table S5.

Construction of pDS0105 and pDS0106 for overexpression of *kfp*02–*kfp*04

The coding sequences of *ppdB*, *ppdA*, and *pepM* (*kfp*02, *kfp*03, and *kfp*04) were PCR amplified as a continuous DNA fragment, with the same sequence and arrangement as in the wild-type genome (Table S6); optimal oligonucleotides JP555 and JP556 were used as PCR primers with gDNA as template, and the JP555–JP556 PCR product was used as template for a PCR with oligonucleotides JP549 and JP550 as primers. The JP549–JP550 PCR product was first blunt-end cloned in pBluescript II KS(+) (pKS) linearized with SmaI. The insert was then excised with NdeI/HindIII and ligated to the expression vectors pRM4 or pIJ10257 linearized with the same enzymes, obtaining constructs pDS0105 or pDS0106, respectively. The correctness of all plasmids was verified by Sanger sequencing.

Construction of pDS0107 for replacement of *kfp*02–*kfp*04 with *neo*

The upstream homologous region was PCR amplified with oligonucleotides JP553 and JP554 and blunt-end cloned in pKS linearized with SmaI; after confirmation of the correct insert by Sanger sequencing, a clone with the insert in orientation M13R-JP553 was selected to allow use the HindIII site of pKS. The downstream homologous region was PCR amplified with oligonucleotides JP551 and JP552, and blunt-end cloned in pKS linearized with SmaI; after confirmation of the correct insert by Sanger sequencing, a clone with the insert in orientation M13F-JP552 was selected. This construct was linearised with SpeI-XbaI and ligated with the upstream homologous region excised XbaI-HindIII and the kanamycin resistance marker neo excised with XbaI-HindIII from pTC192-km; the resulting cassette JP551-downstream-JP552_SpeI/XbaI_neo_-HindIII_JP553-upstream-JP554_XbaI was excised from pKS with BamHI (from JP551 and carried from pKS polylinker) and ligated to pGus21 linearised with BamHI, resulting in construct pDS0107.

Construction of pDS0101 and pDS0102 for overexpression of putative LuxR-like regulator gene *kfp*24 [pGE505 and pGE507]

*kfp*24 (putative LuxR-like regulator gene) was amplified by PCR with oligonucleotides JP525 and JP526, and blunt-end cloned in pKS linearized with SmaI. After confirmation of the correct insert by Sanger sequencing, it was excised with NdeI/HindIII and ligated to the expression vectors pIJ10257 or pRM4 linearized with the same enzymes, obtaining constructs pDS0101 or pDS0102, respectively.

Construction of *K. fiedleri* DSM 114396^T Δ*kfp*02*kfp*04::*neo* deletion mutant (YM0173) and complemented strains (YM0178)

pDS0107 was mobilized to *K. fiedleri* DSM 114396^T by conjugation from *E. coli* ET12567/pUZ8002 and exconjugants selected for resistance to kanamycin. Several exconjugants were cultivated on HM without antibiotic to facilitate segregation through sporulation. Spores from exconjugants were plated at low density (100–150 CFU per plate) on HM supplemented with kanamycin and X-Gluc; GusA-negative colonies were replicated on HM supplemented with apramycin and X-Gluc to confirm loss of vector, and on HM with kanamycin and X-Gluc to confirm presence of the *neo* marker gene. Kanamycin-resistant, apramycin-sensitive, GusA-negative clones were selected for genotype test. Mutant clones were verified by multiple PCR reactions: with oligonucleotides pepMupfw/pepMuprv, RTpepM2fw/RTpepM2rv, JP569/JP570, JP571/JP572, JP573/JP574, and JP575/JP576 as primer pairs (no amplicon expected for mutants), JP508/JP503 and JP555/JP556 (differential amplicon size), JP581/JP582 and JP583/JP584 (exclusive mutant, one primer anneals within *neo*, the other outside homologous regions) (Table S5); all PCR products were assessed in gel electrophoresis and verified by Sanger sequencing with the same oligonucleotides as primers.

To test genetic complementation of the mutants, *kfp*02–04 expression constructs pDS0106 was mobilized to selected mutant clones by conjugation from *E. coli* ET12567/pUZ8002 and exconjugants selected for resistance to hygromycin B.

Construction of *K. fiedleri* DSM 114396^T *kfp*24 overexpression strains (YM0107 and YM0108)

pDS0101 or pDS0102 were mobilized to *K. fiedleri* DSM 114396^T by conjugation from *E. coli* ET12567/pUZ8002 and exconjugants selected for resistance to hygromycin B or apramycin, respectively.

Construction of *K. fiedleri* DSM 114396^T *kfp*02-04 overexpression strains (YM0128 and YM0129)

pDS0105 or pDS0106 were mobilized to *K. fiedleri* DSM 114396^T by conjugation from *E. coli* ET12567/pUZ8002 and exconjugants selected for resistance to apramycin or hygromycin B, respectively.

Construction of *K. fiedleri* DSM 114396^T *kfp*24 (putative LuxR-like regulator gene) and *pepMppdAB* overexpression strains (YM0167 and YM0168)

pDS0105 or pDS0106 were mobilized to *Kitasatospora* sp. YM0107 or YM0108, respectively by conjugation from *E. coli* ET12567/pUZ8002 and exconjugants selected for resistance to apramycin or hygromycin B, respectively.

^{31}P NMR analysis for phosphonates detection

DSM 40043^T, DSM 40133^T, DSM 40283^T, DSM 40386^T, DSM 40581^T, DSM 40907, DSM 41649^T, DSM 41691^T, DSM 41801^T, DSM 41840^T, DSM 41867^T, DSM 41954^T, DSM 43861^T, DSM 44228^T, DSM 45285^T, DSM 46670^T, DSM 101999^T, DSM 102214^T, I6, DSM 40736, DSM 114396^T, TUE 18, TUE H45, TUE 3678, TUE 3997, TUE 21470 (for full species names see Table S1) were cultivated in 100 mL Erlenmeyer baffled flasks containing 30 mL of HM or R5 standard medium,[90] with orbital shaking (180 rpm) at 28 °C for 3–4 days. These cultures (5 mL) were used to inoculate 30 mL of OM, HM, SFM, R5, NL200, NL300, NL400, NL410 medium (prepared without agar according to the recipes stated above), GUBC medium (10 g saccharose, 5 g meat extract, 5 g casamino acids, 5 g glycerol, 5 mL 1 M Na_2HPO_4/KH_2PO_4 pH 7.3, 2 mL Hunter's base dissolved in 1 L distilled water, pH adjusted to 7.3, and 10 mL Balch's vitamins added after autoclaving. Hunter's concentrated base: 20 g nitrilotriacetic acid, 14 g KOH, 59.3 g $MgSO_4{\cdot}7H_2O$, 6.67 g $CaCl_2{\cdot}2H_2O$, 0.0185 g $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$, 0.198 g $FeSO_4{\cdot}7H_2O$, 0.25 g EDTA, 1.095 g $ZnSO_4{\cdot}7H_2O$, 0.5 g $FeSO_4{\cdot}7H_2O$, 0.154 g $MnSO_4{\cdot}H_2O$, 0.0392 g $CuSO_4{\cdot}5H_2O$, 0.025 g $Co(NO_3)_2{\cdot}7H_2O$, 0.0177 g $Na_2B_4O_7{\cdot}10H_2O$ dissolved in 1 L distilled water, pH adjusted to 6.8. Balch's vitamins: 5 mg *p*-aminobenzoic acid, 2 mg folic acid, 2 mg biotin, 5 mg nicotinic acid, 5 mg calcium pantothenate, 5 mg riboflavin, 5 mg thiamine HCl, 10 mg pyridoxine HCl (B6), 100 μg cyanocobalamin (B12), 5 mg thioctic acid (lipoic acid) dissolved in 1 L distilled water, pH adjusted to 7.0, sterilised by filtration), and ISP4 medium (10 g soluble starch, 2 g $CaCO_3$, 1 g K_2HPO_4, 1 g $MgCl_2$, 1 g NaCl, 2 g $(NH_4)SO_4$ and 1 mL trace metal mix 2 dissolved in 1 L distilled water, pH adjusted to 7.3. Trace metal mix 2 : 1 g $FeSO_4{\cdot}7H_2O$, 1 g $MnCl_2{\cdot}2H_2O$, 1 g $ZnSO_4{\cdot}7H_2O$ dissolved in 1 L distilled water) in 100 mL baffled flasks. After 7 days, cultures were harvested by centrifugation at 4000 rcf and 4 °C for 15 min. Supernatants were applied to C18 solid phase extraction (SPE) cartridges (1 g), and the aqueous flow-through was collected. Aqueous phases were diluted with methanol to an 80% final methanol concentration, filtered, and dried using a rotary evaporator. ^{31}P NMR spectra were recorded in 20% D_2O at 243 MHz on a Bruker Avance III HDX 600 MHz spectrometer fitted with a 5 mm Prodigy BBO H&F CryoProbe. Phosphorus chemical shifts are reported in ppm relative to a 10 mM L-phosphinothricin HCl reference (δ_P 51.2) using a double-chamber coaxial NMR tube. NMR data were analyzed using MestReNova 14.3.0.

Author contributions

YM and CCH conceived the study. AZ, SNX, JM, JPGE, JB, CCH and YM designed the research and developed the methodology. AZ, SNX, JM, JPGE, JK and MI conducted experiments. All authors carried out the investigation. AZ, SNX, JM, JPGE

curated the data. AZ, SNX and JPGE prepared the visualizations. AZ, JPGE and YM drafted the initial manuscript. JPGE, JB, UN, IN, WWM, CCH and YM supervised the work. WWM gave scientific advice. All authors reviewed, edited and approved the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

The genome accession numbers for all strains analyzed in this study are listed in the supplementary information (SI) (Table S1). Genome assemblies were obtained from NCBI. Strains are available from the DSMZ repository, with all strain-related metadata accessible through the BacDive metadatabase. Genome sequences were analyzed for biosynthetic gene clusters using antiSMASH (v6.0–7.1.0). Gene cluster families were generated using BiG-SCAPE (v2.0) and visualized with Cytoscape (v3.10.3); the clustering parameters and cutoff values are provided in the Methods section. The sequence of the *Kitasatospora fiedleri* DSM 114396^T genome, including the annotated phosphonate biosynthetic gene cluster (P-BGC), is available at NCBI under accession number GCA_948472415. *Kitasatospora fiedleri* phosphonate BGC description has been submitted to MIBiG (mibig.secondarymetabolites.org) and will be accessible upon the next database release with accession BGC0003187. All other supporting data, including raw ^{31}P NMR spectra, gene synteny alignments, and experimental protocols, are provided in the SI. See DOI: https://doi.org/10.1039/d5cb00254k.

Plasmids and strains generated in this study are available from the corresponding author upon request.

Acknowledgements

We gratefully acknowledge the funding received from the German Center for Infection Research (DZIF, TTU 09.826 and TTU 09.720). SNX is grateful for a PhD scholarship (202008330294) from the Chinese Scholarship Council. AZ was supported by a scholarship awarded by the DAAD (German Academic Exchange Service during a research stay at the Carl R. Woese Institute for Genomic Biology). CCH acknowledges infrastructural support from the Cluster of Excellence EXC 2124: Controlling Microbes to Fight Infection (CMFI, project ID 390838134). We thank Dr Sabine Gronow, Dr Rüdiger Pukall, Dr Christiane Baschien, and Dr Andrey Yurkov for providing us with strains from the DSMZ culture collection for antimicrobial bioassays. We gratefully acknowledge Meike Döppner for excellent technical assistance. We would also like to sincerely thank Prof. Dr Wolfgang Wohlleben (University of Tübingen) for his support and continued interest in the project.

References

1 M. Miethke, M. Pieroni, T. Weber, M. Brönstrup, P. Hammann and L. Halby, *et al.*, Towards the sustainable discovery and development of new antibiotics, *Nat. Rev. Chem.*, 2021, **5**(10), 726–749.
2 J. Bérdy, Thoughts and facts about antibiotics: Where we are now and where we are heading, *J. Antibiot.*, 2012, **65**(8), 385–395.
3 F. Hemmerling and J. Piel, Strategies to access biosynthetic novelty in bacterial genomes for drug discovery, *Nat. Rev. Drug Discovery*, 2022, 1–20.
4 B. Berdy, A. L. Spoering, L. L. Ling and S. S. Epstein, In situ cultivation of previously uncultivable microorganisms using the ichip, *Nat. Protoc.*, 2017, **12**(10), 2232–2242.
5 I. Handayani, H. Saad, S. Ratnakomala, P. Lisdiyanti, W. Kusharyoto and J. Krause, *et al.*, Mining Indonesian Microbial Biodiversity for Novel Natural Compounds by a Combined Genome Mining and Molecular Networking Approach, *Mar. Drugs*, 2021, **19**(6), 316.
6 I. Nouioui, J. Boldt, A. Zimmermann, R. Makitrynskyy, G. Pötter and M. Jando, *et al.*, Biotechnological and pharmaceutical potential of twenty-eight novel type strains of *Actinomycetes* from different environments worldwide, *Curr. Res. Microb. Sci.*, 2024, **7**, 100290.
7 M. Crüsemann, E. C. O'Neill, C. B. Larson, A. V. Melnik, D. J. Floros and R. R. da Silva, *et al.*, Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols, *J. Nat. Prod.*, 2017, **80**(3), 588–597.
8 V. Seidel, in Initial and Bulk Extraction of Natural Products Isolation, ed S. D. Sarker, L. Nahar, *Natural Products Isolation*, Humana Press, Totowa, NJ, 2012, p. 27–41.
9 S. D. Sarker, Z. Latif and A. I. Gray, in Natural Product Isolation, ed S. D. Sarker, Z. Latif, A. I. Gray, *Natural Products Isolation*, Humana Press, Totowa, NJ, 2005, p. 1–25.
10 L. Katz and R. H. Baltz, Natural product discovery: past, present, and future, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**(2–3), 155–176.
11 D. Panter F, C. Bader and R. Müller, Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics, *Chem. Sci.*, 2021, **12**(17), 5994–6010.
12 A. L. Demain, Importance of microbial natural products and the need to revitalize their discovery, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**(2), 185–201.
13 R. H. Baltz, Antimicrobials from actinomycetes: back to the future, *Microbe*, 2007, **2**, 6–7.
14 K. Alam, A. Mazumder, S. Sikdar, Y. M. Zhao, J. Hao and C. Song, *et al.*, *Streptomyces*: The biofactory of secondary metabolites, *Front. Microbiol.*, 2022, 13.
15 M. H. Medema, P. Cimermancic, A. Sali, E. Takano and M. A. Fischbach, A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis, *PLoS Comput. Biol.*, 2014, **10**(12), e1004016.
16 B. A. M. Rudd and D. A. Hopwood, Genetics of Actinorhodin Biosynthesis by *Streptomyces coelicolor* A3(2), *Microbiology*, 1979, **114**(1), 35–43.

17 M. Nett, H. Ikeda and B. S. Moore, Genomic basis for natural product biosynthetic diversity in the actinomycetes, *Nat. Prod. Rep.*, 2009, **26**(11), 1362–1384.

18 A. Gavriilidou, S. A. Kautsar, N. Zaburannyi, D. Krug, R. Müller and M. H. Medema, *et al.*, Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes, *Nat. Microbiol.*, 2022, **7**(5), 726–735.

19 M. Zerikly and G. L. Challis, Strategies for the Discovery of New Natural Products by Genome Mining, *ChemBioChem*, 2009, **10**(4), 625–633.

20 M. A. Fischbach and C. T. Walsh, Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms, *Chem. Rev.*, 2006, **106**(8), 3468–3496.

21 H. Machado, R. N. Tuttle and P. R. Jensen, Omics-based natural product discovery and the lexicon of genome mining, *Curr. Opin. Microbiol.*, 2017, **39**, 136–142.

22 N. Ziemert, M. Alanjary and T. Weber, The evolution of genome mining in microbes–a review, *Nat. Prod. Rep.*, 2016, **33**(8), 988–1005.

23 M. Okuhara, Y. Kuroda, T. Goto, M. Okamoto, H. Terano and M. Kohsaka, *et al.*, Studies on new phosphonic acid antibiotics studies on new phosphonic acid antibiotics, *J. Antibiot.*, 1980, **33**(1), 13–17.

24 A. C. Eliot, B. M. Griffin, P. M. Thomas, T. W. Johannes, N. L. Kelleher and H. Zhao, *et al.*, Cloning, Expression, and Biochemical Characterization of *Streptomyces rubellomurinus* Genes Required for Biosynthesis of Antimalarial Compound FR900098, *Chem. Biol.*, 2008, **15**(8), 765–770.

25 S. Takahashi, T. Kuzuyama, H. Watanabe and H. Seto, A 1-deoxy-d-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-d-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**(17), 9879–9884.

26 M. Rohmer and M. Rohmer, The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants, *Nat. Prod. Rep.*, 1999, **16**(5), 565–574.

27 E. O. Stapley, D. Hendlin, J. M. Mata, M. Jackson, H. Wallick and S. Hernandez, *et al.*, Phosphonomycin. I. Discovery and *in vitro* biological characterization, *Antimicrob. Agents Chemother.*, 1969, **9**, 284–290.

28 T. Hidaka, M. Goda, T. Kuzuyama, N. Takei, M. Hidaka and H. Seto, Cloning and nucleotide sequence of fosfomycin biosynthetic genes of *Streptomyces wedmorensis*, *Mol. Gen. Genet. MGG.*, 1995, **249**(3), 274–280.

29 R. D. Woodyer, Z. Shao, P. M. Thomas, N. L. Kelleher, J. A. V. Blodgett, W. W. Metcalf, W. A. van der Donk and H. Zhao, Heterologous Production of Fosfomycin and Identification of the Minimal Biosynthetic Gene Cluster, *Chem. Biol.*, 2006, **13**(11), 1171–1182.

30 S. Y. Kim, K. S. Ju, W. W. Metcalf, B. S. Evans, T. Kuzuyama and W. A. van der Donk, Different biosynthetic pathways to fosfomycin in *Pseudomonas syringae* and *Streptomyces* species, *Antimicrob. Agents Chemother.*, 2012, **56**(8), 4175–4183.

31 J. Shoji, T. Kato, H. Hinoo, T. Hattori, K. Hirooka and K. Matsumoto, *et al.*, Production of fosfomycin (phosphonomycin) by *Pseudomonas syringae*, *J. Antibiot.*, 1986, **39**(7), 1011–1012.

32 M. E. Falagas, K. P. Giannopoulou, G. N. Kokolakis and P. I. Rafailidis, Fosfomycin: Use Beyond Urinary Tract and Gastrointestinal Infections, *Clin. Infect. Dis.*, 2008, **46**(7), 1069–1077.

33 F. M. Kahan, J. S. Kahan, P. J. Cassidy and H. Kropp, The Mechanism of Action of Fosfomycin (phosphonomycin), *Ann. N. Y. Acad. Sci.*, 1974, **235**(1), 364–386.

34 M. Petek, Š. Baebler, D. Kuzman, A. Rotter, Z. Podlesek and K. Gruden, *et al.*, Revealing fosfomycin primary effect on *Staphylococcus aureus* transcriptome: modulation of cell envelope biosynthesis and phosphoenolpyruvate induced starvation, *BMC Microbiol.*, 2010, **10**(1), 159.

35 S. Eschenburg, M. Priestman and E. Schönbrunn, Evidence That the Fosfomycin Target Cys115 in UDP-*N*-acetylglucosamine Enolpyruvyl Transferase (MurA) Is Essential for Product Release, *J. Biol. Chem.*, 2005, **280**(5), 3757–3763.

36 E. Bayer, K. H. Gugel, K. Hägele, H. Hagenmaier, S. Jessipow and W. A. König, *et al.*, Stoffwechselprodukte von Mikroorganismen. 98. Mitteilung. Phosphinothricin und Phosphinothricyl-Alanyl-Alanin, *Helv. Chim. Acta*, 1972, **55**(1), 224–239.

37 D. Schwartz, N. Grammel, E. Heinzelmann, U. Keller and W. Wohlleben, Phosphinothricin tripeptide synthetases in *Streptomyces viridochromogenes* Tü494, *Antimicrob. Agents Chemother.*, 2005, **49**(11), 4598–4607.

38 D. Schwartz, S. Berger, E. Heinzelmann, K. Muschko, K. Welzel and W. Wohlleben, Biosynthetic Gene Cluster of the Herbicide Phosphinothricin Tripeptide from *Streptomyces viridochromogenes* Tü494, *Appl. Environ. Microbiol.*, 2004, **70**(12), 7093–7102.

39 C. Wendler, A. Putzer and A. Wild, Effect of Glufosinate (Phosphinothricin) and Inhibitors of Photorespiration on Photosynthesis and Ribulose-1,5-Bisphosphate Carboxylase Activity, *J. Plant Physiol.*, 1992, **139**(6), 666–671.

40 K. S. Ju, J. R. Doroghazi and W. W. Metcalf, Genomics-Enabled Discovery of Phosphonate Natural Products and their Biosynthetic Pathways, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**(2), 345–356.

41 M. E. Falagas, E. K. Vouloumanou, G. Samonis and K. Z. Fosfomycin Vardakas, *Clin. Microbiol. Rev.*, 2016, **29**(2), 321–347.

42 C. J. Thompson and H. Seto, in Chapter 6 - Bialaphos, ed Vining L. C., Stuttard C., *Genetics and Biochemistry of Antibiotic Production*, Butterworth-Heinemann, Boston, 1995, p. 197–222.

43 L. Widler, W. Jahnke and R. J. Green, The Chemistry of Bisphosphonates: From Antiscaling Agents to Clinical Therapeutics, *Adv. Anticancer Agents Med. Chem.*, 2012, **12**(2), 95–101.

44 W. W. Metcalf and W. A. van der Donk, Biosynthesis of Phosphonic and Phosphinic Acid Natural Products, *Annu. Rev. Biochem.*, 2009, **78**, 65–94.

45 G. P. Horsman and D. L. Zechel, Phosphonate Biochemistry, *Chem. Rev.*, 2017, **117**(8), 5704–5783.

46 I. Ntai, V. V. Phelan and B. O. Bachmann, Phosphonopeptide K-26 biosynthetic intermediates in *Astrosporangium hypotensionis*, *Chem. Commun.*, 2006, 4518–4520.

47 I. Ntai, M. L. Manier, D. L. Hachey and B. O. Bachmann, Biosynthetic Origins of C−P Bond Containing Tripeptide K-26, *Org. Lett.*, 2005, **7**(13), 2763–2765.

48 B. Elise, M. Michael and R. J. Barry, Dunaway-Mariano Debra. Catalysis and thermodynamics of the phosphoenolpyruvate/phosphonopyruvate rearrangement. Entry into the phosphonate class of naturally occurring organophosphorus compounds, *J. Am. Chem. Soc.*, 1988, **110**(16), 5575–5576.

49 C. M. Kayrouz, Y. Zhang, T. M. Pham and K. S. Ju, Genome Mining Reveals the Phosphonoalamide Natural Products and a New Route in Phosphonic Acid Biosynthesis, *ACS Chem. Biol.*, 2020, **15**(7), 1921–1929.

50 J. F. Villarreal-Chiu, J. P. Quinn and J. W. McGrath, The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment, *Front. Microbiol.*, 2012, 3.

51 X. Yu, J. R. Doroghazi, S. C. Janga, J. K. Zhang, B. Circello and B. M. Griffin, *et al.*, Diversity and abundance of phosphonate biosynthetic genes in nature, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**(51), 20759–20764.

52 C. C. H. Chen, Y. Han, W. Niu, A. N. Kulakova, A. Howard and J. P. Quinn, *et al.*, Structure and kinetics of phosphonopyruvate hydrolase from *Variovorax* sp. Pal2: new insight into the divergence of catalysis within the PEP mutase/isocitrate lyase superfamily, *Biochemistry*, 2006, **45**(38), 11491–11504.

53 K. Blin, S. Shaw, H. E. Augustijn, Z. L. Reitz, F. Biermann and M. Alanjary, *et al.*, antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation, *Nucleic Acids Res.*, 2023, gkad344.

54 E. Heinzelmann, G. Kienzlen, S. Kaspar, J. Recktenwald, W. Wohlleben and D. Schwartz, The phosphinomethylmalate isomerase gene pmi, encoding an aconitase-like enzyme, is involved in the synthesis of phosphinothricin tripeptide in *Streptomyces viridochromogenes*, *Appl. Environ. Microbiol.*, 2001, **67**(8), 3603–3609.

55 K. S. Ju, J. Gao, J. R. Doroghazi, K. K. A. Wang, C. J. Thibodeaux and S. Li, *et al.*, Discovery of phosphonic acid natural products by mining the genomes of 10 000 actinomycetes, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**(39), 12175–12180.

56 Y. Zhang, L. Chen, J. A. Wilson, J. Cui, H. Roodhouse and C. Kayrouz, *et al.*, Valinophos Reveals a New Route in Microbial Phosphonate Biosynthesis That Is Broadly Conserved in Nature, *J. Am. Chem. Soc.*, 2022, **144**(22), 9938–9948.

57 Y. Zhang, T. M. Pham, C. Kayrouz and K. S. Ju, Biosynthesis of Argolaphos Illuminates the Unusual Biochemical Origins of Aminomethylphosphonate and Nε-Hydroxyarginine Containing Natural Products, *J. Am. Chem. Soc.*, 2022, **144**(22), 9634–9644.

58 B. S. Evans, C. Zhao, J. Gao, C. M. Evans, K. S. Ju and J. R. Doroghazi, *et al.*, Discovery of the Antibiotic Phosacetamycin via a New Mass Spectrometry-Based Method for Phosphonic Acid Detection, *ACS Chem. Biol.*, 2013, **8**(5), 908–913.

59 A. L. A. Polidore, L. Furiassi, P. J. Hergenrother and W. W. Metcalf, A Phosphonate Natural Product Made by *Pantoea ananatis* is Necessary and Sufficient for the Hallmark Lesions of Onion Center Rot, *mBio*, 2021, **12**(1), e03402.

60 J. J. Cui, Y. Zhang and K. S. Ju, Phosphonoalamides Reveal the Biosynthetic Origin of Phosphonoalanine Natural Products and a Convergent Pathway for Their Diversification, *Angew. Chem., Int. Ed.*, 2024, **63**(32), e202405052.

61 J. Wilson, J. Cui, T. Nakao, H. Kwok, Y. Zhang and C. Kayrouz, *et al.*, Discovery of Antimicrobial Phosphonopeptide Natural Products from *Bacillus velezensis* by Genome Mining, *Appl. Environ. Microbiol.*, 2023, 89.

62 I. Nouioui, A. Zimmermann, O. Hennrich, S. Xia, O. Rössler and R. Makitrynskyy, *et al.*, Challenging old microbiological treasures for natural compound biosynthesis capacity, *Front. Bioeng. Biotechnol.*, 2024, **12**, 1255151.

63 J. Cui and K. S. Ju, Biosynthesis of *Bacillus* Phosphonoalamides Reveals Highly Specific Amino Acid Ligation, *ACS Chem. Biol.*, 2024, **19**(7), 1506–1514.

64 A. L. A. Polidore, A. D. Caserio, L. Zhu and W. W. Metcalf, Complete Biochemical Characterization of Pantaphos Biosynthesis Highlights an Unusual Role for a SAM-Dependent Methyltransferase, *Angew. Chem., Int. Ed.*, 2023, e202317262.

65 T. S. Freestone, K. S. Ju, B. Wang and H. Zhao, Discovery of a Phosphonoacetic Acid Derived Natural Product by Pathway Refactoring, *ACS Synth. Biol.*, 2017, **6**(2), 217–223.

66 L. Chu, X. Luo, T. Zhu, Y. Cao, L. Zhang and Z. Deng, *et al.*, Harnessing phosphonate antibiotics argolaphos biosynthesis enables a synthetic biology-based green synthesis of glyphosate, *Nat. Commun.*, 2022, **13**(1), 1736.

67 L. Bown, R. Hirota, M. N. Goettge, J. Cui, D. T. Krist and L. Zhu, *et al.*, A Novel Pathway for Biosynthesis of the Herbicidal Phosphonate Natural Product Phosphonothrixin Is Widespread in Actinobacteria, *J. Bacteriol.*, 2023, **0**(0), e00485.

68 E. Takahashi, T. Kimura, K. Nakamura, M. Arahira and M. Iida, Phosphonothrixin, a Novel Herbicidal Antibiotic Produced by *Saccharothrix* sp. ST-888 I. Taxonomy, Fermentation, Isolation and Biological Properties, *J Antibiot (Tokyo)*, 1995, **48**(10), 1124–1129.

69 J. P. Cioni, J. R. Doroghazi, K. S. Ju, X. Yu, B. S. Evans and J. Lee, *et al.*, Cyanohydrin Phosphonate Natural Product from *Streptomyces regensis*, *J. Nat. Prod.*, 2014, **77**(2), 243–249.

70 J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney, S. A. Kautsar, J. H. Tryon and E. I. Parkinson, *et al.*, A computational framework to explore large-scale biosynthetic diversity, *Nat. Chem. Biol.*, 2020, **16**(1), 60–68.

71 D. Hendlin, E. O. Stapley, M. Jackson, H. Wallick, A. K. Miller and F. J. Wolf, *et al.*, Phosphonomycin, a New Antibiotic Produced by Strains of *Streptomyces*, *Science*, 1969, **166**(3901), 122–123.

72 R. Stasi, H. I. Neves and B. Spira, Phosphate uptake by the phosphonate transport system PhnCDE, *BMC Microbiol.*, 2019, **19**(1), 79.

73 J. P. Gomez-Escribano, A. Zimmermann, S. N. Xia, M. Döppner, J. Moschny and C. C. Hughes, *et al.*, Application of a replicative targetable vector system for difficult-to-manipulate streptomycetes, *Appl. Microbiol. Biotechnol.*, 2025, **109**(1), 89.

74 A. Zimmermann, I. Nouioui, G. Pötter, M. Neumann-Schaal, J. Wolf and D. Wibberg, *et al.*, *Kitasatospora fiedleri* sp. nov., a novel antibiotic-producing member of the genus *Kitasatospora*, *Int. J. Syst. Evol. Microbiol.*, 2023, **73**(11), 006137.

75 C. L. M. Gilchrist and Y. H. Chooi, Clinker & clustermap.js: automatic generation of gene cluster comparison figures, *Bioinformatics*, 2021, **37**(16), 2473–2475.

76 V. Agarwal, S. C. Peck, J. H. Chen, S. A. Borisova, J. R. Chekan and W. A. van der Donk, *et al.*, Structure and function of phosphonoacetaldehyde dehydrogenase: the missing link in phosphonoacetate formation, *Chem. Biol.*, 2014, **21**(1), 125–135.

77 A. Badri, A. Williams, K. Xia, R. J. Linhardt and M. A. G. Koffas, Increased 3′-Phosphoadenosine-5′-phosphosulfate Levels in Engineered *Escherichia coli* Cell Lysate Facilitate the In Vitro Synthesis of Chondroitin Sulfate A, *Biotechnol. J.*, 2019, **14**(9), 1800436.

78 Q. Zhao, Q. He, W. Ding, M. Tang, Q. Kang and Y. Yu, *et al.*, Characterization of the Azinomycin B Biosynthetic Gene Cluster Revealing a Different Iterative Type I Polyketide Synthase for Naphthoate Biosynthesis, *Chem. Biol.*, 2008, **15**(7), 693–705.

79 Y. Mao, M. Varoglu and D. H. Sherman, Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lavendulae* NRRL 2564, *Chem. Biol.*, 1999, **6**(4), 251–263.

80 E. Schinko, K. Schad, S. Eys, U. Keller and W. Wohlleben, Phosphinothricin-tripeptide biosynthesis: An original version of bacterial secondary metabolism?, *Phytochemistry*, 2009, **70**(15), 1787–1800.

81 K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel and M. H. Medema, *et al.*, antiSMASH 6.0: improving cluster detection and comparison capabilities, *Nucleic Acids Res.*, 2021, **49**(W1), W29–W35.

82 S. Kuraku, C. M. Zmasek, O. Nishimura and K. Katoh, aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity, *Nucleic Acids Res.*, 2013, **41**(W1), W22–W28.

83 K. Katoh, J. Rozewicki and K. D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, *Brief Bioinform.*, 2019, **20**(4), 1160–1166.

84 D. Edler, J. Klein, A. Antonelli and D. Silvestro, raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML, *Methods Ecol. Evol.*, 2021, **12**(2), 373–377.

85 D. Darriba, D. Posada, A. M. Kozlov, A. Stamatakis, B. Morel and T. Flouri, ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models, *Mol. Biol. Evol.*, 2020, **37**(1), 291–294.

86 S. Sigle, N. Steblau, W. Wohlleben and G. Muth, Polydiglycosylphosphate Transferase PdtA (SCO2578) of *Streptomyces coelicolor* A3(2) Is Crucial for Proper Sporulation and Apical Tip Extension under Stress Conditions, *Appl. Environ. Microbiol.*, 2016, **82**(18), 5661–5672.

87 M. S. B. Paget, L. Chamberlin, A. Atrih, S. J. Foster and M. J. Buttner, Evidence that the Extracytoplasmic Function Sigma Factor σE Is Required for Normal Cell Wall Structure in *Streptomyces coelicolor* A3(2), *J. Bacteriol.*, 1999, **181**(1), 204–211.

88 D. J. MacNeil, K. M. Gewain, C. L. Ruby, G. Dezeny, P. H. Gibbons and T. MacNeil, Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector, *Gene*, 1992, **111**(1), 61–68.

89 J. Sambrook, E. F. Fritsch and T. Maniatis, *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989.

90 T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*. ed, Kieser T., Bibb M. J., Buttner M. J., Chater K. F., Hopwood D. A., John Innes Foundation, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, England, 2000. p. 613.

91 A. M. Kozlov, D. Darriba, T. Flouri, B. Morel and A. Stamatakis, RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference, *Bioinformatics.*, 2019, **35**(21), 4453–4455.

2.3.2 Publication 2

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 01 February 2024
DOI 10.3389/fbioe.2024.1255151

OPEN ACCESS

EDITED BY
Miguel Rocha,
University of Minho, Portugal

REVIEWED BY
Branislav T. Šiler,
University of Belgrade, Serbia
Chongxi Liu,
Northeast Agricultural University, China

*CORRESPONDENCE
Yvonne Mast,
yvonne.mast@dsmz.de

RECEIVED 08 July 2023
ACCEPTED 03 January 2024
PUBLISHED 01 February 2024

CITATION
Nouioui I, Zimmermann A, Hennrich O, Xia S, Rössler O, Makitrynskyy R, Pablo Gomez-Escribano J, Pötter G, Jando M, Döppner M, Wolf J, Neumann-Schaal M, Hughes C and Mast Y (2024), Challenging old microbiological treasures for natural compound biosynthesis capacity. *Front. Bioeng. Biotechnol.* 12:1255151. doi: 10.3389/fbioe.2024.1255151

Challenging old microbiological treasures for natural compound biosynthesis capacity

Imen Nouioui[1], Alina Zimmermann[1,2], Oliver Hennrich[1], Shuning Xia[2,3], Oona Rössler[1], Roman Makitrynskyy[1], Juan Pablo Gomez-Escribano[1], Gabriele Pötter[1], Marlen Jando[1], Meike Döppner[1], Jacqueline Wolf[1], Meina Neumann-Schaal[1,4], Chambers Hughes[2,3] and Yvonne Mast[1,2,4,5]*

[1]Department Bioresources for Bioeconomy and Health Research, Leibniz Institute DSMZ -German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, [2]German Center for Infection Research (DZIF), Partner Site Tübingen, Tübingen, Germany, [3]Department of Microbiology, Biotechnology, Faculty of Science, Interfaculty Institute of Microbiology and Infection Medicine, University of Tübingen, Tübingen, Germany, [4]Braunschweig Integrated Centre of Systems Biology (BRICS), Braunschweig, Germany, [5]Technische Universität Braunschweig, Institut für Mikrobiologie, Braunschweig, Germany

Strain collections are a treasure chest of numerous valuable and taxonomically validated bioresources. The Leibniz Institute DSMZ is one of the largest and most diverse microbial strain collections worldwide, with a long tradition of actinomycetes research. Actinomycetes, especially the genus *Streptomyces*, are renowned as prolific producers of antibiotics and many other bioactive natural products. In light of this, five *Streptomyces* strains, DSM 40971^T, DSM 40484^T, DSM 40713^T, DSM 40976^T, and DSM 40907^T, which had been deposited a long time ago without comprehensive characterization, were the subject of polyphasic taxonomic studies and genome mining for natural compounds based on *in vitro* and *in silico* analyses. Phenotypic, genetic, and phylogenomic studies distinguished the strains from their closely related neighbors. The digital DNA–DNA hybridization and average nucleotide identity values between the five strains and their close, validly named species were below the threshold of 70% and 95%–96%, respectively, determined for prokaryotic species demarcation. Therefore, the five strains merit being considered as novel *Streptomyces* species, for which the names *Streptomyces kutzneri* sp. nov., *Streptomyces stackebrandtii* sp. nov., *Streptomyces zähneri* sp. nov., *Streptomyces winkii* sp. nov., and *Streptomyces kroppenstedtii* sp. nov. are proposed. Bioinformatics analysis of the genome sequences of the five strains revealed their genetic potential for the production of secondary metabolites, which helped identify the natural compounds cinerubin B from strain DSM 40484^T and the phosphonate antibiotic phosphonoalamide from strain DSM 40907^T and highlighted strain DSM 40976^T as a candidate for regulator-guided gene cluster activation due to the abundance of numerous "*Streptomyces* antibiotic regulatory protein" (SARP) genes.

KEYWORDS

actinomycetes, *Streptomyces*, novel species, polyphasic taxonomy, biosynthetic gene cluster, antibiotic

1 Introduction

Microorganisms have long been recognized as a prolific source for valuable bioactive substances (Newman and Cragg, 2016). Especially, actinomycetes are well known for their remarkable biosynthetic potential, with the ability to produce a wide range of natural products, including antibiotics, immunosuppressants, anticancer agents, antifungal compounds, and many other bioactive molecules. Within the family *Streptomycetaceae*, *Streptomyces* is the most prominent genus with respect to the production of bioactive secondary metabolites, including many antibiotics. With over 700 validly named species, *Streptomyces* accounts for more than 50% of all clinically useful antibiotics, including tetracyclines, erythromycin, streptomycin, and vancomycin (Heul et al., 2018). Previous screening campaigns of soil-derived streptomycetes yielded many currently recognized drugs, such as the antibacterial substance streptomycin, the antifungal metabolite nystatin, and the anticancer compound doxorubicin [Schatz et al., 1944; Hazen and Brown, 1951; Acarmone et al., 1969; for review articles, see the work of Genilloud (2017), Baltz (2005), and Berdy (2005)]. However, in the last decades, the discovery rates of novel compounds have declined immensely, which is largely due to the high rediscovery rate of already known substances (Baltz, 2019; van Bergeijk et al., 2020). Classical screening attempts usually employ the so-called Waksman platform, where soil-derived microorganisms are screened for their antimicrobial activity against a panel of bacterial test strains. Re-identification of known substances is partly based on the experimental setup for compound isolation and detection. Standard procedures mostly concentrate on strains producing bioactive compounds readily and in high yields and focus on compounds that show characteristic mass spectrometry patterns (MS) and ultra-violet (UV) spectroscopy signals. Such compounds are commonly referred to as the 'low-hanging fruits' of antibiotics research (Panter et al., 2021). However, re-identification of known substances is also a matter of the phylogenetic uniqueness of the producer strain. Phylogenetically related strains tend to produce similar secondary metabolites (Handayani et al., 2021), and correspondingly, it has been shown that phylogenetic uniqueness is correlated with the diversity of novel natural compounds (Hoffmann et al., 2018). Consequently, re-identification of known substances is also a matter of dereplication of known producer strains. Thus, regarding novel compound discovery, it is expedient to focus on phylogenetically novel producer strains.

Even though actinomycetes have been extensively exploited for drug discovery over the past decades, they have continued to be valuable sources for novel antibiotics. Recent advances in genome sequencing technology and large-scale bioinformatics analyses have revealed an enormous genetic potential for the production of yet undiscovered natural compounds. It is expected that only 3% of the overall genomic potential for natural product biosynthesis has been discovered so far and that the genus *Streptomyces*, especially, represents a huge untapped reservoir for the production of novel secondary metabolites (Gavriilidou et al., 2022). The capability to produce antibiotics is genetically encoded in the actinomycetes' genomes, whereby the corresponding genes are organized as biosynthetic gene clusters (BGCs) (Medema et al., 2015). *Streptomyces* genomes harbor, on average, around 40 BGCs (Belknap et al., 2020), with the majority of all clusters (~90%) being proposed as cryptic or silent, which means that the encoded antibiotics are not known or the respective substances are not produced under standard lab conditions, respectively (Walsh and Fischbach, 2010). State-of-the-art discovery efforts make use of this genetic potential and apply genome mining and genetic engineering techniques to access the genetically encoded biosynthetic potential of microbial producer strains. In recent years, this has already led to the successful discovery of novel potent antibiotics, as recently demonstrated by the identification of the macolacins and cilagicins, which have been deduced from biosynthetic genes primarily (Wang et al., 2022a; b).

In the effort to find novel antibiotics, microbial culture collections represent a treasure trove of pure, well-curated, and freely accessible strains that can be used for bioprospecting. In general, microbial culture collections are dedicated to collecting, maintaining, and distributing microbial strains among microbiologists, as well as being committed to preserving microbial diversity. Thereby, strain collections are essential resources for microbiology, biotechnology, and many other fields and provide a valuable source of microorganisms for research, education, bioprospecting, and conservation. Researching large strain collections has the advantage of circumventing laborious strain isolation efforts and also allows access to many strains whose genome sequences are already available. The Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures is one of the largest and most diverse strain collections, housing bioresources from over 90 countries all over the world and 80% of all reported microbial type strains (https://www.dsmz.de/dsmz; https://www.sciencetheearth.com/latest-blog/the-deutsche-sammlung-von-mikrooganismen-und-zellkulturen-gmbh-dsmz-in-braunschweig-germany). The DSMZ subcollection "Actinobacteria" harbors more than 4,000 actinomycetes containing >2,000 type strains and including ~2,500 streptomycetes, with genome sequence data available for ~800 actinobacteria. Many strains included in culture collections have been deposited a long time ago without comprehensive characterization.

In this study, we report on the identification and description of five novel species from the DSMZ strain collection belonging to the genus *Streptomyces*. We analyzed their genetic potential for the biosynthesis of bioactive natural compounds and show empirical evidence for the production of known and novel bioactive natural products.

2 Materials and methods

2.1 Bacterial strains and cultivation conditions

Strains DSM 40907^T, DSM 40484^T, DSM 40976^T (=Gütt 467), and DSM 40971^T (= Sandoz 59283) were isolated from soil samples of unknown countries and deposited in the German Collection of Microorganisms and Cell Cultures (DSMZ) before the 1980s. Strain DSM 40713^T (=Tü 43 = ETH 21510) was isolated from the soil sample collected in Switzerland and deposited by Professor Dr. Hans Zähner (Eberhard Karls University of Tübingen, Germany) in the DSMZ open culture collection. Active culture of the strains and their

close phylogenetic neighbors, including *Streptomyces nojiriensis* DSM 41655T, *S. gardneri* DSM 40064T, *S. gardneri* DSM 40016T, *S. cacaoi asoensis* DSM 41440T, *S. marinus* DSM 41968T, and *S. tauricus* DSM 40560T, were maintained on medium DSMZ 65 (GYM, glucose, yeast, and malt extracts). The purity of the cultures was checked using light microscopy. For chemotaxonomic characterization, the cultures were prepared on ISP 2 broth [International *Streptomyces* Project (ISP); Shirling and Gottlieb, 1966] medium shaken at 200 rpm for 7 days at 28°C. The biomasses were harvested by centrifugation at 4,000 rpm for 15 min, washed with sterile distilled water, and freeze-dried. Fatty acid analysis was carried out from wet biomass collected from a 7-day-old culture prepared on ISP 2 broth medium.

2.2 Cultural and growth properties

The cultural properties of the strains were recorded on ISP 1 (DSMZ 1764), ISP 2 (DSMZ 987), ISP 3 (DSMZ 609), ISP 4 (DSMZ 547), ISP 5 (DSMZ 993), ISP 6 (DSM 1269), ISP 7 (DSM 1619), nutrient (DSMZ 1), Bennett's (DSMZ 548), and trypticase soy (DSMZ 535) agar media after 7 days of incubation at 28°C. The growth of the strains was tested under a wide range of temperatures (4°C, 10°C, 15°C, 20°C, 25°C, 28°C, 37°C, 42°C, and 45°C) and at various pH values (5.0, 5.5, 6.0, 7.5, 8.0, 8.5, 10.0, and 12.0) in DSMZ 65 medium. All these tests were carried out in duplicate using a bacterial suspension of 5 on the McFarland scale. The color of the aerial and substrate mycelium as well as the diffusible pigments were compared against color charts.

2.3 Phenotypic and chemotaxonomic properties

Chemotaxonomic markers of the strains and their close phylogenomic neighbors were determined using standard thin-layer chromatographic procedures. To this end, isomers of diaminopimelic acid (A2pm) (Schleifer and Kandler, 1972) and polar lipid (Bligh and Dyer, 1959; Tindall et al., 2007) patterns were carried out. Cellular fatty acids of the strains were extracted and analyzed by gas chromatography (Agilent 6890N) following the standard protocols of the microbial identification (MIDI) system (Sasser, 1990). Fatty acids were identified by a GC–MS run on an Agilent GC–MS 7000D instrument (Vieira et al., 2021). Isoprenoid quinones were extracted, separated by HPLC, and identified by using both a DAD and high-resolution mass spectrometer, according to the work of Schumann et al. (2021). Biochemical and enzymatic properties of the strains and their phylogenomic neighbors were determined using API-ZYM and API 20NE strips, as instructed by the manufacturer (bioMerieux, Lyon, France).

2.4 Molecular identification and genome sequencing

Wet biomass, harvested from a 7-day-old culture on ISP 2 broth medium, was used for genomic DNA preparation for single gene and genomic analyses, as reported previously (Zimmermann et al., 2023). PCR-mediated amplification of the 16S rRNA gene was performed with universal primers 7F (5′-AGA GTT TGATC(AC) TGG CTC AG-3′) and 1492R (5′-ACGG(CT)TAC CTT GTT ACG ACTT-3′), according to the work of Weisburg et al. (1991). The PCR products (~1520 bp) were sequenced using the fluorescence-based Sanger method (Sanger et al., 1977) with primers 7F and 1492R. The resulting 16S rRNA gene sequences were used for phylogenetic studies (below). The five strains described in this work were selected on the basis of 16S rRNA-based phylogenetic uniqueness for whole shotgun genome sequencing. Illumina genome sequencing and assembly, based on 250-bp paired-end reads from an ~500-bp insert library, was outsourced to MicrobesNG (Birmingham, United Kingdom). Draft genome sequences were annotated with the RAST-SEED webserver (https://rast.nmpdr.org/) (Overbeek et al., 2014).

2.5 Phylogeny and comparative genomic studies

The almost complete 16S rRNA gene sequence (>1,400 bp) of the strains and their close phylogenetic relatives were used in the present study. Pairwise 16S rRNA gene sequence similarities between the strains and their close phylogenetic neighbors were estimated under the setting recommended by Meier-Kolthoff et al. (2013a), which is implemented in the phylogeny web server available in Genome-to-Genome Distance Calculator (GGDC) 2.1 (http://ggdc.dsmz.de) (Meier-Kolthoff et al., 2022). The reference strains were retrieved from the EzBioCloud server (https://www.ezbiocloud.net/) (Yoon et a., 2017).

The maximum-likelihood (ML) trees based on 16S rRNA gene and genome sequences of the strains were carried out with the Type (Strain) Genome Server (TYGS) v. 1.0 (https://tygs.dsmz.de/), a free bioinformatics tool for whole-genome-based taxonomic analysis (Meier-Kolthoff and Göker, 2019). Phylogenetic classification using the TYGS is based on a genome database that contains the genomic, taxonomic, and nomenclatural data of all currently available type strains. The database is constantly updated. The TYGS platform allows phylogenetic analyses based on the full-length genome sequence of the strain of interest, which is compared with a database of type strain genomes. Thereby, the TYGS provides information on the similarity of the strain to its nearest related type strain with the help of digital DNA–DNA hybridization (dDDH) values calculated by the GGDC 2.1 (http://ggdc.dsmz.de) (Meier-Kolthoff et al., 2013b) and determines differences in genomic G + C contents. dDDH and average nucleotide identity (ortho ANI) between the strains and their phylogenomic relatives were determined using the GGDC webservers, with the recommended formula 4 and the ANI calculator from EzBioCloud (https://www.ezbiocloud.nettools/ani), respectively.

2.6 *In silico* screening for secondary metabolites

The antiSMASH web tool v. 6.0 (Blin et al., 2021) was used to analyze whole-genome sequences of the strains and their closest

phylogenomic neighbors for the presence of BGCs. Gene cluster similarity is displayed in percentage and indicates the number of genes similar to a known cluster. Genes are similar if a BLAST alignment yields an *e*-value < 1×10^{-5} and if sequence identity is >30%. In addition, the shortest alignment must encompass >25% of the sequence. If all genes of a known cluster can be found in the query cluster, the sequence similarity is 100%. The similarity decreases if fewer genes of the known cluster can be found in the query cluster (Medema et al., 2011). The default settings were used for all analyses.

2.7 Fermentation and preparation of culture extracts

The strains DSM 40484^T, DSM 40713^T, DSM 40907^T, DSM 40971^T, and DSM 40976^T were cultivated by inoculation from a GYM plate in 50 mL medium R5 (103 g L^{-1} sucrose, 10 g L^{-1} glucose, 0.25 g L^{-1} K_2SO_4, 10.12 g L^{-1} $MgCl_2$ x 6 H_2O, 0.1 g L^{-1} casamino acids, 5 g L^{-1} yeast extract, 5.73 g L^{-1} TES (pH 7.2), 2.94 g L^{-1} $CaCl_2$ x 2H_2O, 0.05 g L^{-1} KH_2PO_4, 3 g L^{-1} L-proline, and 2 mL trace element solution, according to the work of Handel et al. (2022); pH 7.4) at 28°C in 250 mL Erlenmeyer flasks on an orbital shaker (180 rpm). After 2 days of cultivation, 5 mL of the preculture was used to inoculate 250 mL Erlenmeyer flasks with 50 mL of production media R5 or NL 800 (5 g L^{-1} glucose, 10 g L^{-1} glycerol, 10 g L^{-1} soluble starch, 58 g L^{-1} oatmeal, 2 g L^{-1} yeast extract, 1 g L^{-1} NaCl, and 1 g L^{-1} $CaCO_3$; pH 7.2) as production cultures, which were incubated for 3 days at 28°C on an orbital shaker at 180 rpm. For extraction of organic compounds, 5 mL of culture was harvested after 3 days and extracted with 5 mL ethyl acetate (EtAc) for 3–6 h at room temperature under constant vertical rotation. After centrifugation at 5,000 rpm for 10 min, the organic phase was completely dried using a centrifugal evaporator (SP Genevac EZ-2, "Low BP" program). The concentrated extracts were dissolved in 0.25 mL 50% methanol (MeOH), resulting in a 20-fold concentrated EtAc extract sample. The crude extracts were used for bioassays and liquid chromatography–high-resolution mass spectrometry (LC–HRMS) analysis.

2.8 Bioassays

The crude extracts were analyzed for antimicrobial activities using agar-well diffusion assays against a panel of Gram-positive and Gram-negative bacteria, yeast, and fungal reference strains: *Staphylococcus aureus* DSM 18827, *Enterococcus faecium* DSM 20477^T, *Pseudomonas aeruginosa* DSM 1117, *Escherichia coli* DSM 1103, *Proteus vulgaris* DSM 2140, *Candida albicans* DSM 1386, and *Trichophyton rubrum* DSM 16111. Active culture of these strains was prepared under the growth conditions recommended by the DSMZ collection (https://www.dsmz.de/collection/catalogue). Bioassay test plates were prepared as reported previously (Handayani et al., 2021). A volume of 30 µL methanolic crude extract from three independent biological samples of the five *Streptomyces* strains DSM 40484^T, DSM 40713^T, DSM 40907^T, DSM 40971^T, and DSM 40976^T was pipetted in prepared agar wells of bioassay test agar plates inoculated with the test strains listed above. The plates were incubated overnight at 28°C and 37°C according to the optimal growth temperature of the bioassay test strains (*T. rubrum* bioassays plates were incubated at room temperature in darkness for several days, with daily observation). Antibiotic activity was estimated by measuring the diameter of the inhibition zone.

2.9 Chemical analyses for compound detection

Extracts were analyzed on an analytical 1290 Infinity II LC system coupled to a Bruker Impact II QTOF mass spectrometer. Gradient elution was performed through a C18 porous core-shell column (Phenomenex Kinetex C18, 100 × 2.1 mm, 1.7 µm, 100 Å) using a 10-min gradient from 5% MeCN to 100% MeCN in water, supplemented with 0.1% formic acid. The flow rate was 0.5 mL min^{-1}. QTOF parameters were as follows: 150–2,000 *m/z* scan range, 4,500 V capillary voltage, 500 V end plate offset, 2.8 bar nebulizer, 220°C drying temperature, and 10 L min^{-1} nitrogen drying gas.

Identification of cinerubin: Cinerubin B (t_R = 5.7 min) was initially identified using a combination of its UV/Vis spectrum and LC–HRMS. The UV/Vis spectrum (λ_{max} = 235, 255, 290, and 490 nm) matched cinerubin B in an "in-house" UV/Vis database. Its protonated molecular ion *m/z* 826.3278 $[M + H]^+$ (calculated (calcd) for $C_{42}H_{52}NO_{16}$, 826.3281, Δ 0.3 ppm) yielded a formula ($C_{42}H_{51}NO_{16}$) matching that of cinerubin B.

Identification of phosphonoalamide: Phosphonoalamides were identified using HRMS. A (t_R = 0.5 min): *m/z* 340.1270 $[M + H]^+$ (calcd for $C_{11}H_{23}N_3O_7P$, 340.1268, Δ −0.5 ppm); B (t_R = 0.5 min): *m/z* 370.1378 $[M + H]^+$ (calcd for $C_{12}H_{25}N_3O_8P$, 370.1374, Δ −1.0 ppm); C (t_R = 0.8 min): *m/z* 354.1424 $[M + H]^+$ (calcd for $C_{12}H_{25}N_3O_7P$, 354.1425, Δ 0.2 ppm); and D (t_R = 0.9 min): *m/z* 368.1582 $[M + H]^+$ (calcd for $C_{13}H_{27}N_3O_7P$, 368.1581, Δ −0.3 ppm). The MS/MS fragmentation data for the detected compounds matched the data reported by Kayrouz et al. (2020).

2.10 Inactivation of a cinerubin B biosynthesis gene in DSM 40484^T

To inactivate *ctg27_10* from the predicted cinerubin BGC, a 0.64-kb-long DNA fragment, harboring an internal part of the gene, was amplified from genomic DNA by PCR with primers RM_ctg27_10_kn_vn_for (AAAAAAGCTTTGAAACTGGAGGAGGAGTAC) and RM_ctg27_10_kn_vn_rev (AAAAAAGAATTCGTCTCGTGG ATGTCGTTCTG). The PCR product was digested with *Hind*III and *Eco*RI and then cloned into the respective restriction sites of pKC1132, a suicide vector containing an apramycin resistance gene, to generate pKC1132/ctg27_10_kn. The resulting plasmid was transferred to DSM 40484^T by conjugation using *E. coli* ET12567/pUZ8002 as a donor strain. Exconjugants were selected by resistance to apramycin (indicative of single crossover), leading to the creation of the mutant strain DSM 40484^T *ctg27_10::pKC1132/ctg27_10_kn* (*Mctg27_10*).

2.11 SARP overexpression in the strain DSM 40976^T

A detailed analysis of all BGCs from the five strains revealed the presence of numerous SARP genes in different BGCs of the strain DSM 40976^T. The latter was the subject of SARP overexpression studies. In this context, a conjugative, self-replicative *papR2* expression construct was generated. The plasmid pGM1190 was used as a vector, which contained the *tipA* promoter for the induction of gene transcription, and *oriT*, which is required for the intergenic conjugation from *E. coli* to *Streptomyces*. pGM1190 is a multi-copy and very stable plasmid, which does not require antibiotic selection in production cultures (Muth, 2018). The *papR2* gene was excised as an ~1 kb *Hin*dIII/*Nde*I-fragment from pRM4/*papR2* (Krause et al., 2020) and cloned into the *Hin*dIII/*Nde*I-linearized ~6.9 kb pGM1190 vector. This resulted in the plasmid pGM1190/*papR2-tipA*p, where *papR2* transcription is under the control of the thiostrepton-inducible *tipA* promoter. The pGM1190/*papR2-tipA*p plasmid was transferred to *Streptomyces* sp. DSM 40976^T by conjugation, as described by Kieser et al. (2000), resulting in the strain DSM 40976^T pGM1190/*papR2-tipA*p. Apramycin (50 μg/mL) was used for selection when appropriate. To test for the effect of *papR2* expression on secondary metabolite production, the strain DSM 40976^T pGM1190/*papR2-tipA*p and the WT strain were each cultivated in OM medium (20 g L^{-1} oat meal, 5 mL trace element solution, pH 7.3; according to Handel et al. (2022)) at 28°C; 20 mL of culture was harvested after 72 h of cultivation. For extraction of organic compounds, cultures were treated as described above. For the preparation of supernatant samples, 20–25 mL from the original 50 mL culture was transferred to a 50 mL Falcon tube and centrifuged at 5,000 rpm for 15 min. The supernatant was concentrated with the centrifugal evaporator (SP Genevac EZ-2, "aqueous" program) to 3–5 mL, resulting in a 5–7-fold concentrated supernatant sample. The crude extract and supernatant samples were analyzed in bioassays against *Micrococcus luteus* (valid name: *Kocuria rhizophila* DSM 11926^T), as described by Handayani et al. (2021). Experiments were carried out as three independent biological replicates.

3 Results and discussion

3.1 Polyphasic taxonomic studies

3.1.1 Cultural, morphological, and phenotypic properties

The strains showed phenotypic and morphological features consistent with their classification in the genus *Streptomyces* (Kämpfer, 2012). As shown in Supplementary Table S1, most of the strains grow well on ISP 3, ISP 7, and Bennett's agar media forming aerial hyphae, which varied from white to gray after 7 days of incubation at 28°C (https://www.dsmz.de/collection/catalogue). All the studied strains grew well on ISP3, ISP 7, DSMZ 65, R5, and Bennett's media. The strains were able to grow at an incubation temperature between 25°C and 28°C and at a pH of 6–8, and they grew optimally at 28°C and pH 6.5–7 (Supplementary Figure S1; Supplementary Table S1). No growth of the strains was observed at incubation temperature above 37°C and below 10°C. More details about the growth and cultural properties of the strains are provided in Supplementary Table S1. The strains can be distinguished from one another and from their closest phylogenomic relatives by a range of biochemical and enzymatic properties, as shown in Table 1. Strain DSM 40907^T was able to produce α-glucosidase and lipase C14, unlike its closest relative *S. nijoriensis* DSM 41655^T. Strain DSM 40713^T can be distinguished from the reference strain DSM 41440^T by the production of lipase C14, α-chymotrypsin, β-glucosidase, and n-acetyl-β-glucosaminidase. However, strain DSM 40971^T can be differentiated from its close neighbor, *S. marinus*, by its inability to produce α-galactosidase, β-galactosidase, and β-glucuronidase. Qualitative variation in the enzymatic profile of strain DSM 40976^T and its close relatives, *S. gardneri* and *S. narbonensis*, was noted, as shown in Table 1. Strain DSM 40484^T could be distinguished from the type strain DSM 40560^T by its ability to produce esterase (lipase) C8, lipase C14, trypsin, and α-chymotrypsin; further biochemical characteristics are displayed in Table 1. Excellence congruence was obtained for all the duplicated phenotypic tests.

3.1.2 Chemotaxonomic features

The studied strains and their close phylogenomic neighbors, *S. nojiriensis* DSM 41655^T, *S. cacaoi* subsp. *asoensis* DSM 41440^T, *S. gardneri* DSM 40064^T, *S. narbonensis* DSM 40016^T, *S. tauricus* DSM 40560^T, and *S. marinus* DSM 41968^T, have *LL*-A2pm as the diamino acid of the wall peptidoglycan. The polar lipid profile of the strains and their closest relatives contained diphosphatidylglycerol (DPG), phosphatidylethanolamine (PE), and phosphatidylinositol (PI), as well as unidentified lipids (L), phospholipids (PLs), aminolipids (ALs), glycolipids (GLs), and glycophospholipids (GPLs), as shown in Supplementary Figure S2. Quantitative and qualitative variations in the isoprenoid profiles were detected between the studied strains and their close phylogenetic relatives. The predominant menaquinone (≥15%) of the strains DSM 40907^T and DSM 40713^T and their closest neighbors, *S. nojiriensis* DSM 41655^T and *S. cacaoi* subsp. *asoensis* DSM 41440^T, was MK-9 H_4 and MK-9 H_6. Nevertheless, strains DSM 40976^T and its close relatives, *S. gardneri* DSM 40064^T and *S. narbonensis* DSM 40016^T, had MK-9 H_6 and MK-9 H_8. Strain DSM 40971^T had MK-9 H_4 and MK-9 H_6, while strain DSM 41968^T contained MK-9 H_6 and MK-9 H_8. The quinone pattern of strain DSM 40484^T was composed of MK-9 H_4 and MK-9 H_6 (Supplementary Table S2). The major fatty acid (>10%) of strain DSM 40907^T was $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, $C_{16:0}$ *iso*, and $C_{16:0}$, while *S. nojiriensis* DSM 41655^T, its closest phylogenomic neighbor, had $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, and $C_{16:0}$. Strain DSM 40976^T had $C_{17:0}$ *anteiso*, $C_{15:0}$ *anteiso*, and $C_{16:0}$. However, the type strains of *S. gardneri* and *S. narbonensis* had $C_{16:0}$ *iso* in addition. The fatty acid profile of strain DSM 40713^T consisted of $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, $C_{16:0}$ *iso*, $C_{16:0}$, and $C_{17:0}$ *anteiso*, while $C_{15:0}$ *iso*, $C_{16:0}$ *iso*, and $C_{17:0}$ *anteiso* were below the 10% for its close relative, *S. cacaoi* subsp. *asoensis* DSM 41440^T. Strain DSM 40971^T differed from its closest neighbor, *S. marinus* DSM 41968^T, by a fatty acid pattern composed of $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, $C_{16:0}$ *iso*, and $C_{16:0}$, whereas strain DSM 41968^T only showed traces of $C_{16:0}$. Strain DSM 40484^T had $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, $C_{16:0}$ *iso*, $C_{16:0}$, and $C_{17:0}$ *iso*, which is a fatty acid profile similar to that of its relative strain DSM 40560^T (Supplementary Table S3).

3.1.3 Phylogeny based on 16S rRNA gene and whole-genome sequences

The 16S rRNA gene sequence similarities between the strains and their closest relatives ranged from 99.3% to 100%

10.3389/fbioe.2024.1255151

TABLE 1 Phenotypic properties of the strains and their closest phylogenomic neighbors.

Substrates	DSM 40907^T	DSM 41655 ^{T}	DSM 40976 ^{T}	DSM 40064 ^{T}	DSM 40016^T	DSM 40713^T	DSM 41440^T	DSM 40484^T	DSM 40560^T	DSM 40971^T	DSM 41968^T
API Zym											
Esterase C4	+	+	+	+	+	+	+	+	+	+	+
Esterase (lipase) C8	+	+	+	+	+	+	+	+	-	+	-
Lipase C14	+	-	+	+	-	+	-	+	-	+	+
Trypsin	+	+	+	+	+	+	+	+	-	+	+
α-Chymotrypsin	+	+	+	+	-	+	-	+	-	+	+
Acid phosphatase	+	+	+	+	-	+	+	+	+	+	+
Naphthol-AS-BI-phosphohydrolase	+	+	+	+	-	+	+	+	+	+	+
α-Galactosidase	-	-	-	-	-	-	+	-	+	-	+
β-Galactosidase	-	-	+	+	-	+	+	+	+	-	+
β-Glucuronidase	-	-	-	-	-	-	-	-	-	-	+
α-Glucosidase	+	-	+	+	+	-	-	+	+	+	+
β-Glucosidase	+	+	+	+	+	+	-	+	+	+	+
n-Acetyl-β-glucosaminidase	+	+	+	+	+	+	-	+	+	+	+
α-Mannosidase	-	-	-	+	-	-	-	+	+	+	+
API 20NE											
Tryptophan	+	-	-	-	-	-	-	-	-	-	-
Urea	-	+	+	+	+	+	-	-	-	-	+
Gelatine	+	+	-	+	+	+	+	-	+	+	+
p-Nitro-phenyl-β-D-galactopyranosid	-	-	+	+	+	+	+	+	+	-	+
Glucose	+	+	+	+	+	+	+	+	+	-	+
Arabinose	+	+	+	+	-	+	+	+	+	-	+
Mannose	+	+	+	-	-	+	+	+	+	-	+
Mannitol	-	+	-	-	-	+	+	+	+	-	+
N-Acetylglucosamin	+	+	+	+	+	+	+	+	+	-	+

(Continued on following page)

10.3389/fbioe.2024.1255151

TABLE 1 (*Continued*) Phenotypic properties of the strains and their closest phylogenomic neighbors.

Substrates	DSM 40907^T	DSM 41655 ^{T}	DSM 40976 ^{T}	DSM 40064 ^{T}	DSM 40016^T	DSM 40713^T	DSM 41440^T	DSM 40484^T	DSM 40560^T	DSM 40971^T	DSM 41968^T
Maltose	-	+	+	+	+	+	+	-	-	-	+
Gluconate	+	+	+	+	-	+	+	+	+	-	+
Adipate	-	+	-	-	-	-	+	-	+	-	+
Malate	+	+	+	+	-	+	+	+	+	-	+
Citrate	-	+	-	+	-	+	+	+	+	-	+
Phenylacetate	+	-	+	+	-	-	+	-	-	-	-

All the strains were able to produce alkaline phosphatase, leucine arylamidase, valine arylamidase, and cystine arylamidase. All the strains were unable to produce α-fucosidase. All the strains had a positive reaction to esculin and a negative reaction to arginine, glucose, and cayratic.

(Supplementary Table S4), which were values well above the threshold of 98.65% for prokaryotic species delineation (Kim et al., 2014). Strain DSM 40907^T showed a 16S rRNA gene sequence similarity of 100% to that of the type strains of *Streptomyces spororaveus*, *S. xanthophaeus*, and *S. nojiriensis* species (Supplementary Table S4). Moreover, strain DSM 40976^T shared a similar sequence with *S. gardneri* NBRC 12865^T (99.9%), *S. narbonensis* NBRC 1280^T (99.9%), and *S. zaomyceticus* NBRC 13348^T (99.9%). Strains DSM 40484^T and DSM 40971^T showed a 16S rRNA gene sequence similarity value of 99.3% with *S. glomeroaurantiacus* NBRC 15418^T, *S. aurantiacus* NBRC 13017^T, and *S. nanshensis* SCSIO 01066^T. Strain DSM 40713^T had a high 16S rRNA gene sequence similarity to *S. humidus* NBRC 12877^T (99.7%) and *Streptomyces cacaoi* subsp. *asoensis* NRRL B-16592^T (99.5%). Low sequence similarities of 95.0%–98.7% were found among the studied strains.

These results were in accordance with the phylogenetic position of the strains in the ML tree, where the strains were distributed in different clusters (Figure 1). Strain DSM 40713^T formed a well-supported subcluster with *S. rishiriensis* NBRC 13407^T (99.4%) next to the representative strains of *S. humidus* and *S. cacaoi* subsp. *asoensis*. However, strain DSM 40484^T occupied a distinct branch loosely associated to a subcluster that encompassed *S. phaeochromogenes* (98.8%), *S. umbrinus* (98.8%), and *S. ederensis* (99.0%) and was next to the type strain of *S. glomeroaurantiacus* and *S. aurantiacus* species. In another well-supported lineage, strain DSM 40971^T was grouped with its close relative, *S. nanshensis*. Nevertheless, strain DSM 40976^T appeared in a distinct branch loosely associated to the type strain of *S. gardneri* and *S. zaomyceticus* and distant from *S. narbonensis*, which was inserted in another subcluster. Strain DSM 40907^T and its close neighbors mentioned above resided in the same branch and cannot be phylogenetically differentiated from one another. It is clear that the resolution of the 16S rRNA gene sequence does not allow reliably distinguishing closely related species (Nouioui et al., 2018; Nouioui and Sangal, 2022). To overcome this limitation, a phylogenomic tree on the whole-genome sequence was constructed, and the taxonomic status of these strains was resolved. Strain DSM 40907^T was grouped with *S. nojiriensis* and *S. spororaveus*, with the former as the closest phylogenomic neighbor. Strain DSM 40976^T was found to be closely related to a subcluster that contained the type strains of *S. gardneri* and *S. narbonensis* species. Strain DSM 40713^T formed a well-supported subcluster with *S. cacaoi* subsp. *asoensis* next to *S. humidus* and *S. rishiriensis*. Strain DSM 40484^T was found to be closely related to *'Streptomyces dioscori* A127' and *S. tauricus* JCM 4837^T. However, *'Streptomyces dioscori'* has no standing in nomenclature as the name has not been validly published since 2018. Therefore, only *S. tauricus* can be considered the closest phylogenomic neighbor to strain DSM 40484^T. Strain DSM 40971^T was closely related to *S. marinus*, together forming a well-supported subgroup associated with *S. daqingensis* CGMCC 4.7178^T (Figure 2).

3.1.4 Genomic features and comparative genomic analysis

The genomic features of the five studied strains were consistent with those of the genus *Streptomyces* (Kämpfer, 2012). The strains had a genome size between 7.2 and 9.2 Mbp, a G + C content of 70.8%–72.0%, a number of coding sequences from 6,369 to 8,959, and a number of RNAs between 58 and 84 (Supplementary Table

10.3389/fbioe.2024.1255151

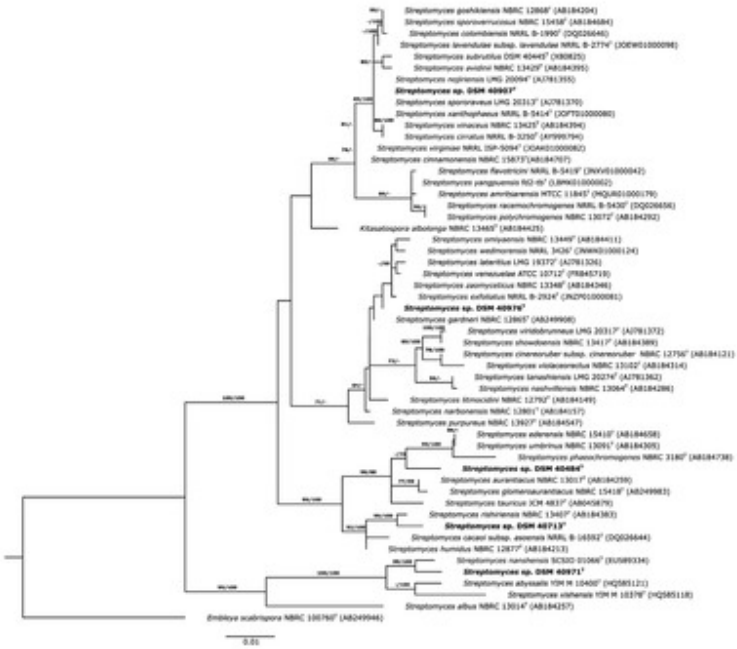


FIGURE 1
Maximum likelihood phylogenetic tree based on 16S rRNA gene sequences, showing the relationships between the strains and their closely related species. The tree was inferred with FastME from GBDP distances calculated from 16S rDNA gene sequences. The numbers above branches are GBDP pseudo-bootstrap support values > 60% from 100 replications. The tree was rooted at the midpoint.

S5). The dDDH values between the genome sequence of strains DSM 40907^T, DSM 40976^T, DSM 40713^T, DSM 40484^T, and DSM 40971^T were well below the 70% threshold established for prokaryotic species demarcation (Wayne et al., 1987), indicating the assignment of each of the strains to a new species as they could be distinguished from their close phylogenomic relatives on this basis (Table 2). The results were coherent with the ANI scores, which were below the cut-off point of 95%–96% for species delineation (Table 2) (Chun et al., 2018). These finding were in line with the phylogenomic studies revealing that the studied strains form five evolutionary lineages, which were distinct from those of the validly published species of *Streptomyces*. The strains can be differentiated from one another and from *Streptomyces* species validly named using the phenotypic, chemotaxonomic, phylogenomic, and genomic features.

3.2 Antimicrobial potential of the five novel *Streptomyces* species

3.2.1 *In silico* screening for secondary metabolites

The whole-genome-based phylogenetic classification of the five novel species and their closest neighbors revealed a clear grouping in five clades (Figure 2). The biosynthetic potential of the strains was unveiled by bioinformatics analysis of the genome sequences with antiSMASH. The strains within the five clades had a slightly different overall content of BGCs. The clade, which harbored DSM 40484^T, showed the highest overall BGC content, with, on average, ~37 BGCs, followed by the clades including DSM 40713^T (34 BGCs), DSM 40907^T (32 BGCs), DSM 40976^T (31 BGCs), and DSM 40971^T (27 BGCs) (Figure 2). Of all strains, *S. cacaoi* subsp. *asoensis* DSM 41440^T showed the highest BGC content, with 52 BGCs in total, while its closest phylogenetic relative, strain DSM 40713^T, had 31 BGCs. The individual BGC contents of the four residual novel *Streptomyces* species were 33 BGCs for DSM 40907^T, 32 BGCs for DSM 40976^T, 36 BGCs for DSM 40484^T, and 21 BGCs for DSM 40971^T. A comparison of the number and classes of BGCs identified in the genomes did not reveal significant differences within and across the five clades due to the large BGC number variation intra-clade. However, there are some notable findings worth highlighting: the clade including DSM 40976^T showed low intra-clade variation in BGC number, while strain DSM 10713^T had only approximately half the BGCs than the closest neighbor *S. cacaoi* subsp. *asoensis* DSM 41440^T. Finally, the clade to which DSM 40907^T belongs had an abundance of strains without any "ectoine" BGC, which may indicate a particular physiology or ecological niche of the respective microorganisms.

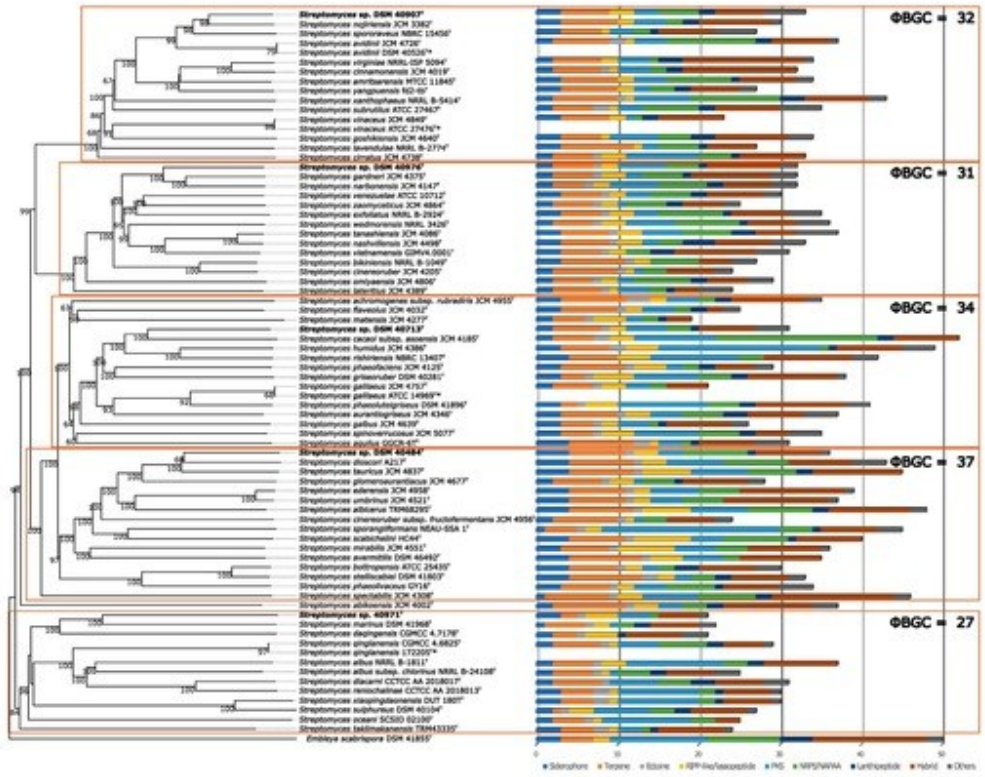


FIGURE 2
Whole-genome sequence tree generated with the TYGS web server for five novel *Streptomyces* species and closely related species. The tree was inferred with FastME from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d_4. Five distinct clades are highlighted with orange boxes. BGC content is shown at the right side of the tree, and BGC classes are indicated with colors according to the legend shown below. * indicates same strains.

3.2.2 *In vitro* screening for antimicrobial compounds

To assess the antibiotic production potential of the five novel *Streptomyces* species, extract samples were used to test for the antimicrobial activity in bioassays against a panel of selected strains from the WHO priority list, including Gram-positive and Gram-negative test bacteria, as well as against yeast and fungal test strain. Extracts obtained from cultures of DSM 40484^T, DSM 40971^T, and DSM 40976^T led to inhibition zones against the Gram-positive pathogens *S. aureus* DSM 18827 and *E. faecium* DSM 20477 and the Gram-negative strain *P. vulgaris* DSM 2140 (Figure 3). Extract samples from DSM 40713^T led to inhibition zones against *E. coli* DSM 1103 and *P. vulgaris* DSM 2140 and, thus, showed Gram-negative-specific antibiotic activity, whereas extract samples from DSM 40907^T led to inhibition against *S. aureus* DSM 18827 and *P. vulgaris* DSM 2140. Altogether, the culture extracts from the five novel species showed activity against at least two pathogenic test strains.

3.2.3 Antibiotic production profile of novel *Streptomyces* species

Strain DSM 40713^T is an old deposit from a well-studied strain collection (Tübingen actinomycetes strain collection) known to be a ferrimycin producer. A ferrimycin gene cluster was identified (Supplementary Figure S3, region 16.1), where 100% of genes showed similarity to a known desferrioxamine B BGC from *Streptomyces griseus* subsp. *griseus* NBRC 13350 (MIBiG accession BGC0000941; Ohnishi et al., 2008) (Supplementary Figure S4). It was assumed that the strain was thoroughly investigated for antibiotic production and, therefore, DSM 40713^T was excluded from further substance analysis. DSM 40971^T was not further analyzed due to the low overall BGC content, as outlined in Supplementary Figure S5.

3.2.4 Detection of cinerubin B in DSM 40484^T

HPLC analysis of ethyl acetate extract samples from DSM 40484^T revealed a prominent peak at 240 nm with a retention

TABLE 2 ANI and dDDH (d_4 formula) values between the whole-genome sequence of the strains and their closest phylogenomic relatives.

Query strain (T)	Subject strain	dDDH (%)	ANI score (%)	G + C content difference (in %)
Strain DSM 40907	*Streptomyces nojiriensis* JCM 3382^T	56.7	97.40	0.29
Strain DSM 40907	*Streptomyces sporoaveus* NBRC 15456^T	50.5	92.96	0.08
Strain DSM 40907	*Streptomyces avidinii* JCM 4726^T	45.0	91.67	0.20
Strain DSM 40907	*Streptomyces avidinii* DSM 40526^T	45.0	91.69	0.20
Strain DSM 40484	*Streptomyces dioscori* A217	45.0	91.57	0.09
Strain DSM 40484	*Streptomyces tauricus* JCM 4837^T	43.7	91.23	0.02
Strain DSM 40484	*Streptomyces glomeroaurantiacus* JCM 4677^T	35.2	87.72	0.51
Strain DSM 40976	*Streptomyces gardneri* JCM 4375^T	44.4	91.33	0.32
Strain DSM 40976	*Streptomyces narbonensis* JCM 4147^T	42.9	90.75	0.08
Strain DSM 40713	*Streptomyces cacaoi* subsp. *asoensis* JCM 4185^T	37.8	88.96	1.01
Strain DSM 40713	*Streptomyces rishiriensis* NBRC 13407^T	34.8	87.76	0.45
Strain DSM 40713	*Streptomyces humidus* JCM 4386^T	34.1	87.48	0.43
Strain DSM 40971	*Streptomyces marinus* DSM 41968^T	29.2	85.05	0.63
Strain DSM 40713	Strain DSM 40484^T	24.7	80.89	0.52
Strain DSM 40976	Strain DSM 40907^T	23.2	79.45	0.32
Strain DSM 40976	Strain DSM 40484^T	22.8	78.59	1.18
Strain DSM 40976	Strain DSM 40713^T	22.7	78.42	0.66
Strain DSM 40484	Strain DSM 40907^T	22.6	77.93	0.86
Strain DSM 40713	Strain DSM 40907^T	22.4	77.94	0.35
Strain DSM 40971	Strain DSM 40907^T	21.4	75.97	0.36
Strain DSM 40713	Strain DSM 40971^T	21.3	75.88	0.01
Strain DSM 40971	Strain DSM 40484^T	21.3	75.89	0.50
Strain DSM 40976	Strain DSM 40971^T	21.2	76.23	0.67

time (RT) of 12.75, with a characteristic UV/Vis spectrum (Figure 4A). LC–HRMS analysis of a bioactive fraction sample of the DSM 40484^T extract delivered an exact mass of 825.3278 (Figure 4B), which matched with a sum formula of $C_{42}H_{51}NO_{16}$ (Figure 4C). The characteristic UV/Vis spectrum (λ_{max} = 235, 255, 290, and 490 nm) matched with cinerubin B in an "in-house" UV/Vis database, which is a known glycosylated polyketide antibiotic (Ettlinger et al., 1959). Bioinformatics analysis of the genome sequence of DSM 40484^T led to the identification of cluster region 27.1, which showed similarity to a cinerubin BGC (Supplementary Figure S6). Sequence comparisons revealed that cluster region 27.1 contained all the genes suggested to be part of a functional cinerubin BGC, since 100% of the genes from the deposited cinerubin MIBiG reference sequence (BGC0000212) (Kersten et al., 2013) were also present in cluster region 27.1 (Supplementary Figure S7). To prove that the identified cluster region 27.1 from DSM 40484^T is responsible for cinerubin B biosynthesis, we inactivated the *ctg27_10* coding sequence, encoding a putative β-ketoacyl synthase, which was expected to be essential for aromatic polyketide biosynthesis (Supplementary Figure S7). Culture extracts from the *Streptomyces* sp. DSM 40484^T *Mctg27_10* mutant and the *Streptomyces* sp. DSM 40484^T wild-type (WT) strain were analyzed in comparative HPLC–MS and antibiotic activity tests against *K. rhizophila* DSM 11926^T as the test organism. In contrast to the WT strain, no antibiotic activity was detected from the extracts of the generated mutant strains (Figure 5A). Moreover, no mass peak corresponding to the molecular ion of cinerubin B (*m/z* $[M+H]^+$ = 826.3) was present in the extracts of DSM 40484^T *Mctg27_10* (Figure 5B), indicating that the inactivation of *ctg27_10* completely abolished cinerubin B production in DSM 40484^T. Until now, this is the first genetic evidence of the functionality of a cinerubin BGC by genetic knockout.

3.2.5 Detection of a phosphonate antibiotic in DSM 40907^T

Strain DSM 40907^T was identified as a potential phosphonate producer by BGC analysis (Supplementary Figure S8) due to the occurrence of a *pepM* coding sequence on region 3.1 (Supplementary Figure S9), which is indicative of phosphonate biosynthesis as it encodes a potential phosphoenolpyruvate mutase (PepM) known to be responsible for the first and essential biosynthetic step of phosphonate biosynthesis,

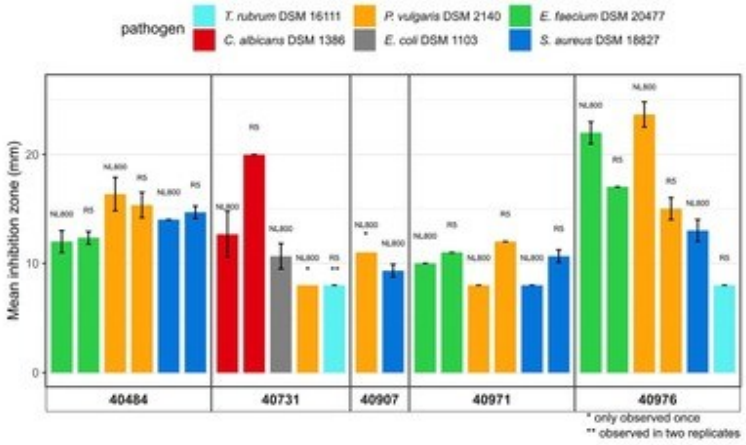


FIGURE 3
Antimicrobial bioassays with extract samples obtained from DSM 40484^T, DSM 40713^T, DSM 40907^T, DSM 40971^T, and DSM 40976^T grown for 72 h in RS and NL800 medium against pathogenic test microorganisms, including *Staphylococcus aureus* DSM 18827 (dark blue), *Enterococcus faecium* DSM 20477 (green), *Escherichia coli* DSM 1103 (gray), *Proteus vulgaris* DSM 2140 (orange), *Candida albicans* DSM 1386 (red), and *Trichophyton rubrum* DSM 16111 (light blue). Inhibition zone diameters of bioassays are given in mm. Data shown are as the result of three independent biological replicates.

converting phosphoenolpyruvate to phosphonopyruvate (Ju et al., 2015). Based on this assumption, culture supernatant samples from DSM 40907^T grown in GUBC medium (Parkinson et al., 2019) were analyzed with LC–HRMS. This led to the identification of characteristic mass fragmentations, including the protonated molecular ions *m/z* 340.1270 [M + H]$^+$, 370.1378 [M + H]$^+$, 354.1424 [M + H]$^+$, and 368.1582 [M + H]$^+$, which matched the sum formula of $C_{11}H_{23}N_3O_7P$, $C_{12}H_{25}N_3O_8P$, $C_{12}H_{25}N_3O_7P$, and $C_{13}H_{27}N_3O_7P$, respectively (Figures 6A–D, respectively). These data corresponded with the MS/MS fragmentation pattern reported for phosphonoalamides (Kayrouz et al., 2020), unveiling that DSM 40907^T produces phosphonoalamides as phosphonate antibiotics. Phosphonoalamides are phosphonoalanine-like natural compounds with good antibiotic activity against different Gram-positive and Gram-negative bacteria (Kayrouz et al., 2020), which is consistent with the observation that extract samples from DSM 40907^T showed bioactivity against *S. aureus* DSM 18827 and *P. vulgaris* DSM 2140 in bioassays, respectively (Figure 3).

3.2.6 Identification of numerous SARP genes in DSM 40976^T

The bioinformatics analysis of strain DSM 40976^T revealed another strain-specific feature, as it showed a comparatively large number of SARP (*Streptomyces* antibiotic regulatory protein) genes (eight in total) as part of BGCs (Supplementary Figure S10). Altogether, eight SARP genes were encoded in six of the 32 BGCs (Supplementary Figure S10). The six BGCs belonged to the following cluster types: type-I PKS, lanthipeptide-class, and three hybrid BGCs (Supplementary Figure S10). SARPs are transcriptional regulators, which act as pathway-specific activators of antibiotic biosynthesis (Bibb, 2005). In a previous study, we have shown that SARP regulators can be used as general activators of antibiotic biosynthesis in various actinomycetes when heterologously expressed (Krause et al., 2020). Due to the high abundance of SARP genes in a number of BGCs, DSM 40976^T was designated for heterologous SARP gene expression. For this purpose, the SARP-type regulator gene *papR2* from the pristinamycin producer *Streptomyces pristinaespiralis* was heterologously expressed in DSM 40976^T using the conjugatable, self-replicative plasmid pGM1190/*papR2-tipA*p. Both DSM 40976T pGM1190/*papR2-tipA*p and DSM 40976 WT strains were each grown in OM medium for 72 h. Ethyl acetate extracts and concentrated supernatant samples were obtained from DSM 40976^T cultures and were used for antimicrobial bioassays against *K. rhizophila* DSM 11926^T. Here, specifically supernatant samples from DSM 40976 pGM1190/*papR2-tipA*p cultures resulted in a significantly improved bioactivity against *K. rhizophila* DSM 11926^T in comparison to samples from the DSM 40976^T WT (Figure 7). This indicates that the SARP expression in strain DSM 40976^T activated BGC expression and, thus, improved antibiotic production. The gene products of the five SARP genes from DSM 40976^T showed amino acid similarity scores of ≥ 50% compared to PapR2. The highest similarity score of 58% was found for a SARP amino acid sequence, which is encoded by the locus Tag: "ctg45_9", located in region 45.1. Cluster region 45.1 represents a hybrid BGC, consisting of NRPS, NRPS-like, and type-I PKS core genes (Supplementary Table S6). Thus, gene region 45.1 is a potential BGC that may have been activated by PapR2 expression. However, due to the presence of several BGCs with *papR2* homologous genes in DSM 40976^T, a reliable prediction of

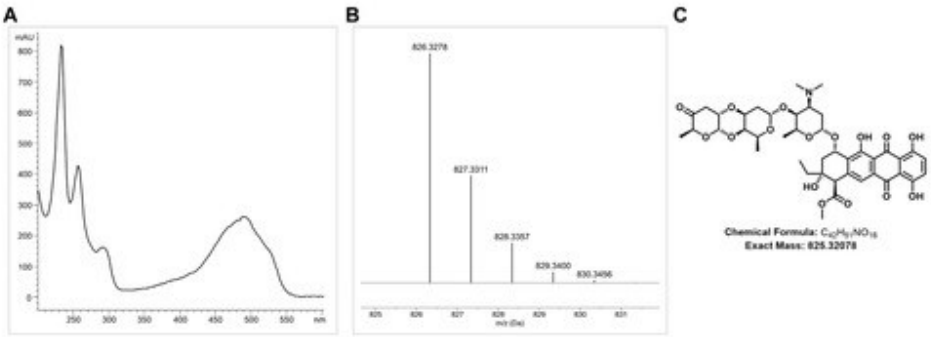


FIGURE 4
Detection of cinerubin B production in ethyl acetate extract samples of DSM 40484^T by LC-M analysis. Characteristic LC-HRMS analysis of cinerubin B **(A)**. HRMS data of the DSM 40484^T sample with a protonated molecular ion (*m/z*) of 825.3278 $(M + H)^+$ **(B)**. Chemical structure of cinerubin B ($C_{42}H_{51}NO_{16}$) **(C)**.

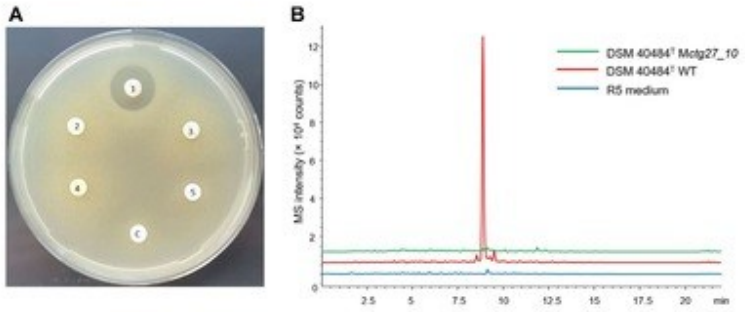


FIGURE 5
Inactivation of *ctg27_10* completely blocks cinerubin B production in DSM 40484^T. **(A)** *K. rhizophila* DSM 11926^T growth inhibition around paper discs saturated with extracts from different cinerubin B producers: DSM 40484^T WT (1), *Mctg27_10* (2–4), extract from the fermentation medium R5 (5), negative control, and methanol (C). **(B)** Overlaid extracted ion chromatograms for the protonated molecular ion of cinerubin B (*m/z* $[M+H]^+$ = 826.3) derived from DSM 40484^T strains grown in R5.

gene clusters that have been activated or bioactive compounds that were produced cannot be made. Definitely, the narrowing down to a few candidate gene clusters will help identify the bioactive compound in subsequent work.

3.2.7 Description of *Streptomyces kutzneri* sp. nov.

Streptomyces kutzneri (kutz'ne.ri N.L. gen. n. *kutzneri*, referring to Professor Hans-Jürgen Kutzner, who contributed significantly to prokaryotic systematics) is a Gram-stain-positive, aerobic, non-motile bacterium that forms light-gray aerial mycelium on oatmeal agar medium that turns white on GYM medium and white-grayish on Bennett's media. A dark-brown diffusible pigment is produced at an incubation temperature of 10°C and 15°C on GYM medium and at 28°C on ISP 7. The strain is able to grow from 10°C to 37°C, optimally at 28°C, and from pH 5–7.5, optimally at pH 7.0. Additional cultural and morphological properties are mentioned in Supplementary Table S1. It has *LL*-A2pm as the diamino acid of the cell wall peptidoglycan and diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol, a glycolipid, a lipid, two aminolipids, and four unidentified phospholipid components as the polar lipid profile. The fatty acid pattern (>4%) contains $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, $C_{16:0}$ *iso*, $C_{16:1}$ CIS 9, $C_{16:0}$, $C_{17:0}$ *iso*, and $C_{17:0}$ *anteiso*. The menaquinone pattern (≥4%) encompasses MK-9 H_2, MK-9 H_4, MK-9 H_6, and MK-9 H_8. The genome size of the strain is 9.2 Mbp, and its *in silico* G + C content is 71.7%.

The type strain DSM 40907^T was isolated from soil of an unknown origin and deposited in the DSMZ culture collection. The genome sequences of DSM 40907T have been deposited in the DDBJ/ENA/GenBank databases under the accession number JASTTI000000000. The BioProject accession number is PRJNA980006.

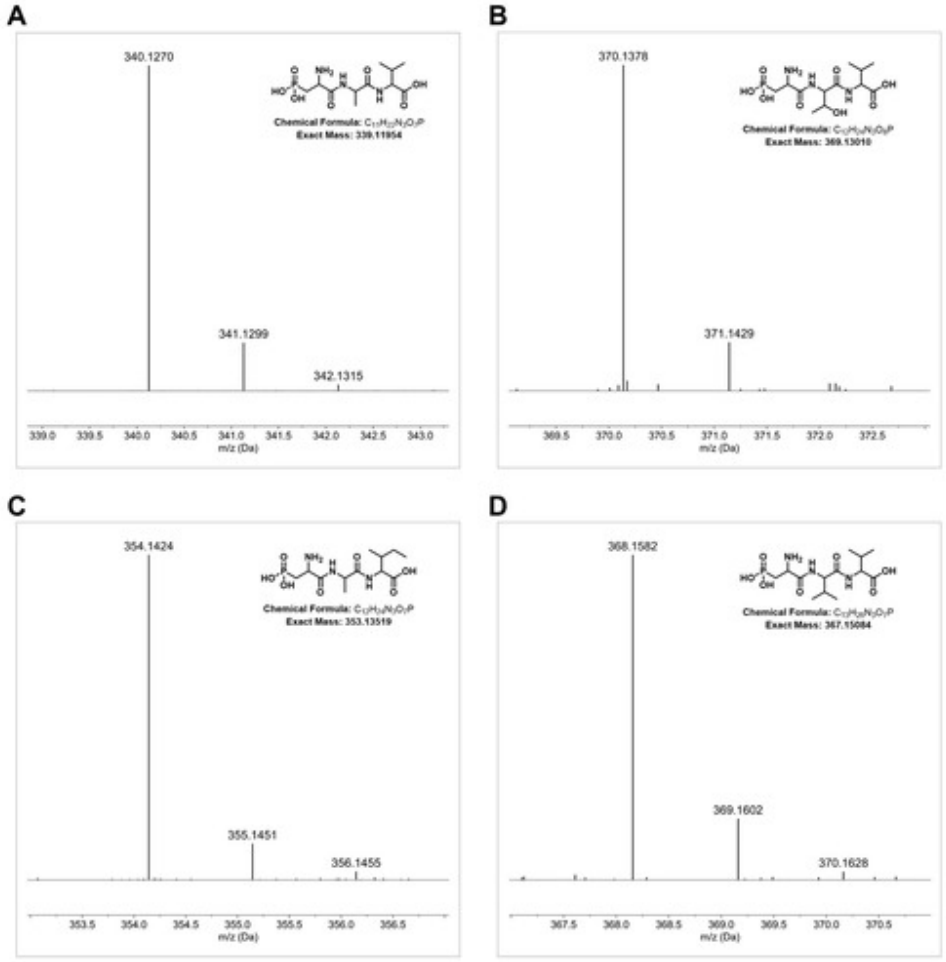


FIGURE 6
Detection of phosphonoalamides production in samples of DSM 40907T by HRMS analysis. HRMS data of the DSM 40907T sample with *m/z* 340.1268 $[M + H]^+$ ($C_{11}H_{23}N_3O_7P$) **(A)**, *m/z* 370.1378 $[M + H]^+$ ($C_{12}H_{25}N_3O_8P$) **(B)**, *m/z* 354.1424 $[M + H]^+$ ($C_{12}H_{25}N_3O_7P$) **(C)**, and *m/z* 368.1582 $[M + H]^+$ ($C_{13}H_{27}N_3O_7P$) **(D)**.

3.2.8 Description of *Streptomyces stackebrandtii* sp. nov.

Streptomyces stackebrandtii (stack.e.brandt'i.i. N.L. gen. n. *stackebrandtii*, referring to Prof. Erko Stackebrandt, a German microbiologist who has contributed significantly to the prokaryotic systematics) is a Gram-stain-positive, aerobic, non-motile bacterium that forms white–beige aerial mycelium on oatmeal agar and Bennett's media that turns white on ISP 5 and ISP 7 media. The strain is able to grow from 10°C to 37°C, optimally at 28°C, and from pH 5–7.5, optimally at pH 7.0. Additional cultural and morphological properties are mentioned in Supplementary Table S1. It has *LL*-A2pm as the diamino acid of the cell wall peptidoglycan and diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol, a glycolipid, two aminolipids and lipids, and three unidentified phospholipid components as the polar lipid profile. The fatty acid pattern (>4%) contains $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, $C_{16:0}$ *iso*, $C_{16:1}$ CIS 9, $C_{16:0}$, $C_{17:0}$ *iso*, $C_{17:0}$ *anteiso*, and $C_{17:0}$ *cyclo* CIS 9. The menaquinone pattern (≥4%) encompasses MK-9 H_6 and MK-9 H_8. The genome size of the strain is 8.7 Mbp, and its *in silico* G + C content is 72.0%.

The type strain DSM 40976^T (=Gütt 467) was isolated from soil of an unknown origin and deposited in the DSMZ culture collection.

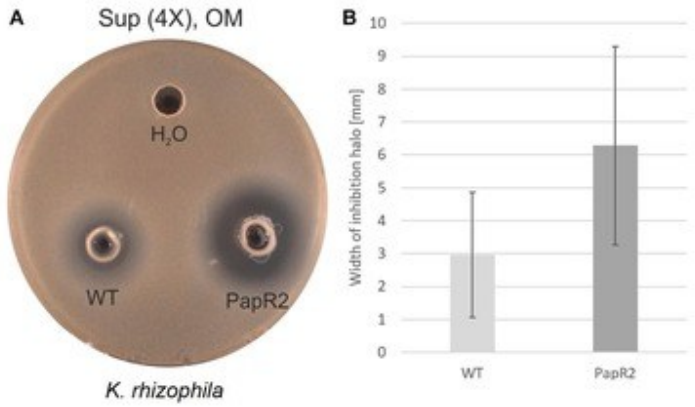


FIGURE 7
Bioassay results of *Streptomyces* sp. DSM 40976 WT and *papR2* overexpression strain (pGM1190/*papR2-tipAp*, "PapR2") after 72 h cultivation in OM. **(A)** Supernatant ("Sup," 4X-concentrated) was tested against *K. rhizophila* DSM 11926^T, with H_2O as the negative control. **(B)** Average width of inhibition halos against *K. rhizophila* DSM 11926^T. Black lines show the standard deviation (number of replicates = 3).

The genome sequence of DSM 40976T has been deposited in the DDBJ/ENA/GenBank databases under the accession number JASTTG000000000. The BioProject accession number is PRJNA979996.

3.2.9 Description of *Streptomyces zaehneri* sp. nov.

Streptomyces zaehneri (zaeh'ne.ri. N.L. gen. n. *zaehneri*, named after Professor Hans Zähner, a German microbiologist who contributed significantly to the field of natural products from microbial sources) is a Gram-stain-positive, aerobic, non-motile bacterium that forms gray-pinkish aerial mycelium on oatmeal agar that turns white–gray and gray on Bennett's and ISP 7 media, respectively. A dark-brown diffusible pigment is produced on ISP 7 and GYM media at 28°C at 15°C, respectively. The strain is able to grow from 15°C to 37°C, optimally at 28°C, and from pH 5–7.5, optimally at pH 7.0. Additional cultural and morphological properties are mentioned in Supplementary Table S1. It has *LL*-A2pm as the diamino acid of the cell wall peptidoglycan and diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol, a glycolipid, two aminolipids and lipid, and four unidentified phospholipid components as the polar lipid profile. The fatty acid pattern (>4%) contains $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, $C_{16:0}$ *iso*, $C_{16:1}$ CIS 9, $C_{16:0}$, $C_{17:0}$ *iso*, and $C_{17:0}$ *anteiso*. The menaquinone pattern (≥4%) encompasses MK-9 H_2, MK-9 H_4, MK-9 H_6, and MK-9 H_8. The genome size of the strain is 9.1 Mbp, and its *in silico* G + C content is 71.3%.

The type strain DSM 40713^T (= ETH 21510 = Tü 43) was isolated from soil from Firnhüttealp, Switzerland, and deposited by Prof. Hans Zähner in the DSMZ culture collection. The genome sequences of DSM 40713T have been deposited in the DDBJ/ENA/GenBank databases under the accession number JASTTJ000000000. The BioProject accession number is PRJNA980070.

3.2.10 Description of *Streptomyces kroppenstedtii* sp. nov.

Streptomyces kroppenstedtii (krop.pen.stedt'i.i. N.L. gen. n. *kroppenstedtii*, named after Professor Reiner M. Kroppenstedt, a German microbiologist who is well known for his contribution to the bacterial taxonomy and contributed significantly to the Actinobacteria collection in the DSMZ culture collection) is a Gram-stain-positive, aerobic, non-motile bacterium that forms gray–orange aerial mycelium on oatmeal agar medium that turns gray–beige and white on ISP 7 and Bennett's media, respectively. Bacteria diffusible pigment is produced at an incubation temperature of 10°C–28°C on GYM medium. The strain is able to grow from 10°C to 37°C, optimally at 28°C, and from pH 5–7.5, optimally at pH 7.0. Additional cultural and morphological properties are mentioned in Supplementary Table S1. It has *LL*-A2pm as the diamino acid of the cell wall peptidoglycan and diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol, an aminolipid, a glycolipid, a lipid, and four unidentified phospholipid components as the polar lipid profile. The fatty acid pattern (>4%) contains $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, $C_{16:0}$ *iso*, $C_{16:1}$ CIS 9, $C_{16:0}$, $C_{17:0}$ *iso*, and $C_{17:0}$ *anteiso*. The menaquinone pattern (≥4%) encompasses MK-9 H_2, MK-9 H_4, MK-9 H_6, and MK-9 H_8. The genome size of the strain is 9.2 Mbp, and its *in silico* G + C content is 70.8%.

The type strain DSM 40484^T was isolated from soil of an unknown origin and deposited in the DSMZ culture collection. The genome sequences of DSM 40484T have been deposited in the

DDBJ/ENA/GenBank databases under the accession number JASTTK000000000. The BioProject accession number is PRJNA980075.

3.2.11 Description of *Streptomyces winkii* sp. nov.

Streptomyces winkii (win'ki.i. N.L. gen. n. *winkii*, named after Joachim Wink, a German microbiologist who has made significant contributions to actinobacterial systematics and natural products research) is a Gram-stain-positive, aerobic, non-motile bacterium that forms white-grayish aerial mycelium on oatmeal agar medium that turns white and black–brown on ISP 6, ISP 7, and Bennett's media, respectively. A dark diffusible pigment is produced on ISP 6 and ISP 7 at 28°C. The strain is able to grow from 20°C to 37°C, optimally at 28°C, and from pH 5–7.5, optimally at pH 7.0. Additional cultural and morphological properties are mentioned in Supplementary Table S1. It has *LL*-A2pm as the diamino acid of the cell wall peptidoglycan and diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol, an aminolipid, a glycolipid, a lipid, and four unidentified phospholipid components as the polar lipid profile. The fatty acid pattern (>4%) contains $C_{15:0}$ *iso*, $C_{15:0}$ *anteiso*, $C_{16:0}$ *iso*, $C_{16:1}$ CIS 9, $C_{16:0}$, $C_{17:0}$ *iso*, and $C_{17:0}$ *anteiso*. The menaquinone pattern (≥4%) encompasses MK-9 H_2, MK-9 H_4, MK-9 H_6, MK-9 H_8, and MK-8 H_4. The genome size of the strain is 7.2 Mbp, and its *in silico* G + C content is 71.3%.

The type strain DSM 40971^T (= Sandoz 59283) was isolated from soil of an unknown origin and deposited in the DSMZ culture collection. The genome sequences of DSM 40974T have been deposited in the DDBJ/ENA/GenBank databases under the accession number JASTTH000000000. The BioProject accession number is PRJNA980003.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

IN: conceptualization, data curation, formal analysis, investigation, supervision, writing–original draft, writing–review and editing, methodology, validation, and visualization. AZ: investigation, visualization, and writing–original draft. OH: investigation and writing–original draft. SX: investigation and writing–original draft. OR: investigation and writing–original draft. RM: investigation, methodology, validation, visualization, writing–original draft, and supervision. JPG-E: investigation, methodology, validation, writing–original draft, and supervision. GP: writing–original draft and investigation. MJ: writing–original draft and investigation. MD: writing–original draft, and investigation. JW: data curation, writing–review and editing, and investigation. MN-S: data curation, resources, writing–review and editing, and investigation. CH: investigation, supervision, writing–review and editing, data curation, resources, and visualization. YM: conceptualization, data curation, formal analysis, funding acquisition, investigation, project administration, resources, supervision, writing–original draft, writing–review and editing, methodology, and visualization.

Funding

The authors gratefully acknowledge the funding received from the Leibniz Association (K445/2022) and the German Center of Infection Research (DZIF) TTU 09.826. SX is grateful for a PhD scholarship (202008330294) from the Chinese Scholarship Council.

Acknowledgments

The authors thank Sabine Gronow, Rüdiger Pukall, Christiane Baschien, and Andrey Yurkov for providing them with strains from the DSMZ culture collection for bioassays. They are grateful to Birgit Grün, Gesa Martens, and Anika Wasner (DSMZ-German culture collection) for excellent technical assistance. They thank Bernard Schink for assistance with etymology of the proposed names. Furthermore, they thank Antonia Kristin Heldmann, Felix Gonther, and Joshua Beck for their experimental contribution that helped generating the cinerubin B mutant strain DSM 40484^T *Mctg27_10*, in frame of the MSc course IB20B *Biotechnological aspects of Actinobacteria* 2023 of TU Braunschweig.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2024.1255151/full#supplementary-material

References

Acarmone, F., Cassinelli, G., Fantini, G., Grein, A., Orezzi, P., Po, C., et al. (1969). Adriamycin, 14-hydroxydaimomycin, a new antitumor antibiotic from *S. peucetius* var. *Caesius*. *Biotechnol. Bioeng.* 9, 1101–1110. doi:10.1002/bit.260110607

Baltz, R. H. (2005). Antibiotic discovery from actinomycetes: will a renaissance follow the decline and fall? *Sim News* 55, 186–196.

Baltz, R. H. (2019). Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. *J. Ind. Microbiol. Biotechnol.* 46, 281–299. doi:10.1007/s10295-018-2115-4

Belknap, K. C., Park, C. J., Barth, B. M., and Andam, C. P. (2020). Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci. Rep.* 10, 2003. doi:10.1038/s41598-020-58904-9

Berdy, J. (2005). Bioactive microbial metabolites. *J. Antibiotics* 58, 1–26. doi:10.1038/ja.2005.1

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* 8, 208–215. doi:10.1016/j.mib.2005.02.016

Bligh, E. G., and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiology* 37, 911–917. doi:10.1139/y59-099

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49, W29–W35. doi:10.1093/nar/gkab335

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi:10.1099/ijsem.0.002516

Ettlinger, L., Gäumann, E., Hütter, R., Keller-Schierlein, W., Kradolfer, F., Neipp, L., et al. (1959). Stoffwechselprodukte von Actinomyceten, XVI. Cinerubine. *Chem. Berichte* 92, 1867–1879. doi:10.1002/cber.19590920820

Gavriilidou, A., Kautsar, S. A., Zaburannyi, N., Krug, D., Müller, R., Medema, M. H., et al. (2022). Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* 7, 726–735. doi:10.1038/s41564-022-01110-2

Genilloud, O. (2017). Actinomycetes: still a source of novel antibiotics. *Nat. Prod. Rep.* 34, 1203–1232. doi:10.1039/C7NP00026J

Handayani, I., Saad, H., Ratnakomala, S., Lisdiyanti, P., Kusharyoto, W., Krause, J., et al. (2021). Mining Indonesian microbial biodiversity for novel natural compounds by a combined genome mining and molecular networking approach. *Mar. Drugs* 19, 316. doi:10.3390/md19060316

Handel, F., Kulik, A., Wex, K. W., Berscheid, A., Saur, J. S., Winkler, A., et al. (2022). Ψ-Footprinting approach for the identification of protein synthesis inhibitor producers. *NAR Genomics Bioinforma.* 4, lqac055. lqac055. doi:10.1093/nargab/lqac055

Hazen, E. L., and Brown, R. (1951). Fungicidin, an antibiotic produced by a soil actinomycete. *Proc. Soc. Exp. Biol. Med.* 76, 93–97. doi:10.3181/00379727-76-18397

Heul, H. U., van der, L., Bilyk, B., McDowall, K., Seipke, R., and Wezel, G. P. V. (2018). Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Nat. Product. Rep.* 35, 575–604. doi:10.1039/C8NP00012C

Hoffmann, T., Krug, D., Bozkurt, N., Duddela, S., Jansen, R., Garcia, R., et al. (2018). Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* 9, 803. doi:10.1038/s41467-018-03184-1

Ju, K.-S., Gao, J., Doroghazi, J. R., Wang, K.-K. A., Thibodeaux, C. J., Li, S., et al. (2015). Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc. Natl. Acad. Sci.* 112, 12175–12180. doi:10.1073/pnas.1500873112

Kayrouz, C. M., Zhang, Y., Pham, T. M., and Ju, K.-S. (2020). Genome mining reveals the phosphonoalamide natural products and a new route in phosphonic acid biosynthesis. *ACS Chem. Biol.* 15, 1921–1929. doi:10.1021/acschembio.0c00256

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). *Practical Streptomyces genetics*. Norwich: John Innes Foundation.

Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi:10.1099/ijs.0.059774-0

Krause, J., Handayani, I., Blin, K., Kulik, A., and Mast, Y. (2020). Disclosing the potential of the SARP-type regulator PapR2 for the activation of antibiotic gene clusters in streptomycetes. *Front. Microbiol.* 11, 225. doi:10.3389/fmicb.2020.00225

Mast, Y., Guezguez, J., Handel, F., and Schinko, E. (2015). A complex signaling cascade governs pristinamycin biosynthesis in *Streptomyces pristinaespiralis*. *Appl. Environ. Microbiol.* 81, 6621–6636. doi:10.1128/AEM.00728-15

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39, W339–W346. doi:10.1093/nar/gkr466

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* 11, 625–631. doi:10.1038/nchembio.1890

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013a). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinforma.* 14, 60. doi:10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., and Göker, M. (2022). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 50, D801–D807. doi:10.1093/nar/gkab902

Meier-Kolthoff, J. P., and Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* 10, 2182. doi:10.1038/s41467-019-10210-3

Meier-Kolthoff, J. P., Göker, M., Spröer, C., and Klenk, H.-P. (2013b). When should a DDH experiment be mandatory in microbial taxonomy? *Arch. Microbiol.* 195, 413–418. doi:10.1007/s00203-013-0888-4

Muth, G. (2018). The pSG5-based thermosensitive vector family for genome editing and gene expression in actinomycetes. *Appl. Microbiol. Biotechnol.* 102, 9067–9080. doi:10.1007/s00253-018-9334-5

Newman, D. J., and Cragg, G. M. (2016). Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661. doi:10.1021/acs.jnatprod.5b01055

Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J. P., Woyke, T., Kyrpides, N. C., et al. (2018). Genome-based taxonomic classification of the phylum Actinobacteria. *Front. Microbiol.* 9, 2007. doi:10.3389/fmicb.2018.02007

Nouioui, I., and Sangal, V. (2022). Advanced prokaryotic systematics: the modern face of an ancient science. *New Microbes New Infect.* 49, 101036. –50. doi:10.1016/j.nmni.2022.101036

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., et al. (2008). Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* 190, 4050–4060. doi:10.1128/jb.00204-08

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi:10.1093/nar/gkt1226

Panter, F., Bader, D., and Müller, R. (2021). Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Chem. Sci.* 12, 5994–6010. doi:10.1039/D0SC06919A

Parkinson, E. I., Erb, A., Eliot, A. C., Ju, K. S., and Metcalf, W. W. (2019). Fosmidomycin biosynthesis diverges from related phosphonate natural products. *Nat. Chem. Biol.* 15, 1049–1056. doi:10.1038/s41589-019-0343-1

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 74, 5463–5467. doi:10.1073/pnas.74.12.5463

Sasser, M. (1990). "MIDI technical note 101. Identification of bacteria by gas chromatography of cellular fatty acids," in *MIDI technical note 101* (Newark, Del: MIDI, Inc), 1–7.

Schatz, A., Bugle, E., and Waksman, S. A. (1944). Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Exp. Biol. Med.* 55, 66–69. doi:10.3181/00379727-55-14461

Schleifer, K. H., and Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* 36, 407–477. doi:10.1128/br.36.4.407-477.1972

Schumann, P., Kalensee, F., Cao, J., Criscuolo, A., Clermont, D., Köhler, J. M., et al. (2021). Reclassification of *Haloactinobacterium glacieicola* as *Occultella glacieicola* gen. nov., comb. nov., of *Haloactinobacterium album* as *Ruania alba* comb. nov, with an emended description of the genus *Ruania*, recognition that the genus names *Haloactinobacterium* and *Ruania* are heterotypic synonyms and description of *Occultella aeris* sp. nov., a halotolerant isolate from surface soil sampled at an ancient copper smelter. *Int. J. Syst. Evol. Microbiol.* 71, 004769. doi:10.1099/ijsem.0.004769

Shirling, E. B., and Gottlieb, D. (1966). Methods for characterization of *Streptomyces* species. *Int. J. Syst. Bacteriol.* 16, 313–340. doi:10.1099/00207713-16-3-313

Tindall, B. J., Sikorski, J., Smibert, R. A., and Krieg, N. R. (2007). "Phenotypic characterization and the principles of comparative systematics," in *Methods for general and molecular microbiology* (John Wiley and Sons, Ltd), 330–393. doi:10.1128/9781555817497.ch15

van Bergeijk, D. A., Terlouw, B. R., Medema, M. H., and van Wezel, G. P. (2020). Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nat. Rev. Microbiol.* 18, 546–558. doi:10.1038/s41579-020-0379-y

Vieira, S., Huber, K. J., Neumann-Schaal, M., Geppert, A., Luckner, M., Wanner, G., et al. (2021). *Usitatibacter rugosus* gen. nov., sp. nov. and *Usitatibacter palustris* sp. nov., novel members of *Usitatibacteraceae* fam. nov. within the order *Nitrosomonadales* isolated from soil. *Int. J. Syst. Evol. Microbiol.* 71, 004631. doi:10.1099/ijsem.0.004631

Walsh, C. T., and Fischbach, M. A. (2010). Natural products version 2.0: connecting genes to molecules. *J. Am. Chem. Soc.* 132, 2469–2493. doi:10.1021/ja909118a

Wang, Z., Koirala, B., Hernandez, Y., Zimmerman, M., and Brady, S. F. (2022a). Bioinformatic prospecting and synthesis of a bifunctional lipopeptide antibiotic that evades resistance. *Science* 376, 991–996. doi:10.1126/science.abn4213

Wang, Z., Koirala, B., Hernandez, Y., Zimmerman, M., Park, S., Perlin, D. S., et al. (2022b). A naturally inspired antibiotic to target multidrug-resistant pathogens. *Nature* 601, 606–611. doi:10.1038/s41586-021-04264-x

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., et al. (1987). Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Evol. Microbiol.* 37, 463–464. doi:10.1099/00207713-37-4-463

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* 173, 697–703. doi:10.1128/jb.173.2.697-703.1991

Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rrna gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi:10.1099/ijsem.0.001755

Zimmermann, A., Nouioui, I., Pötter, G., Neumann-Schaal, M., Wolf, J., Wibberg, D., et al. (2023). *Kitasatospora fiedleri* sp. nov., a novel antibiotic-producing member of the genus *Kitasatospora*. *Int. J. Syst. Evol. Microbiol.* 73. doi:10.1099/ijsem.0.006137

2.3.3 Publication 3

Applied Microbiology and Biotechnology (2025) 109:89
https://doi.org/10.1007/s00253-025-13477-3

RESEARCH

Application of a replicative targetable vector system for difficult-to-manipulate streptomycetes

Juan Pablo Gomez-Escribano[1] · Alina Zimmermann[1,2] · Shu-Ning Xia[2,3] · Meike Döppner[1] · Julia Moschny[2,3] · Chambers C. Hughes[2,3,4] · Yvonne Mast[1,2,5,6]

Received: 5 February 2025 / Revised: 31 March 2025 / Accepted: 2 April 2025 / Published online: 10 April 2025

Abstract

The low frequency of homologous recombination together with poor efficiency in introducing DNA into the cell are the main factors hampering genetic manipulation of some bacterial strains. We faced this problem when trying to construct mutants of *Streptomyces iranensis* DSM 41954, a strain in which conjugation is particularly inefficient, and suicidal vectors had failed to yield any exconjugants. In this work, we report the construction and application of a conjugative replicative vector, pDS0007, which allows selection of exconjugants even with poor conjugation efficiency. The persistence of the construct inside the cell for as long as required facilitates the homologous recombination events leading to single and double crossovers. While it was confirmed that the vector is frequently lost without selection, the recognition sequence for the I-SceI endonuclease was included in the backbone of pDS0007. The presence of a I-SceI recognition sequence would allow to force the loss of the vector and the appearance of double crossover recombinants by introducing a second construct (e.g. pIJ12742) that expresses a *Streptomyces* codon–optimised gene encoding the I-SceI endonuclease. To facilitate screening for vector-free clones, the construct also carries a *Streptomyces* codon–optimised *gusA* gene encoding the β-glucuronidase expressed from a constitutive promoter. We prove the usefulness of this vector and strategy with the strain *S. iranensis* DSM 41954, in which we could readily delete an essential gene of a newly discovered biosynthetic pathway for a phosphonate-containing natural product, which led to loss of phosphonate production according to ^{31}P NMR spectroscopy.

Key points

- *pDS0007 is a new vector for gene-targeting in difficult-to-manipulate streptomycetes.*
- *pDS0007 is self-replicative but easy to cure, targetable and allows visual screening.*
- *pDS0007 was used to prove the discovery of a novel phosphonate biosynthetic pathway.*

Keywords *Streptomyces* · Genetic manipulation · Homologous recombination · Knock-out · Genetic tools

Juan Pablo Gomez-Escribano and Alina Zimmermann contributed equally to the work.

✉ Yvonne Mast
yvonne.mast@dsmz.de

[1] Department Bioresources for Bioeconomy and Health Research, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, 38124 Braunschweig, Germany

[2] German Center for Infection Research (DZIF), Partner Site Tübingen, 72076 Tübingen, Germany

[3] Department of Microbial Bioactive Compounds, Interfaculty Institute of Microbiology and Infection Medicine (IMIT), University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany

[4] Cluster of Excellence EXC 2124: Controlling Microbes to Fight Infection, University of Tübingen, 72076 Tübingen, Germany

[5] Braunschweig Integrated Centre of Systems Biology (BRICS), Rebenring 56, 38106 Braunschweig, Germany

[6] Institute for Microbiology, Technical University Braunschweig, Rebenring 56, 38106 Braunschweig, Germany

Introduction

Streptomycetes and closely related filamentous actinomycetes are high mol % G + C Gram-positive bacteria that have gained the attention of molecular microbiologists for over half a century; first because of their very complex life cycle, resembling that of filamentous fungi, and mostly because of their unmatched capacity to produce bioactive natural products of outstanding importance for medicine, particularly antimicrobials (Katz and Baltz 2016).

Genetics studies based on recombinant DNA technologies have been extensively applied to the study of the biology of streptomycetes and specialised metabolism (Kieser et al. 2000; Baltz 2006). While novel technologies like CRISPR-Cas9 are now available for genetic manipulation in streptomycetes (Alberti and Corre 2019), gene targeting is still mostly based on homologous recombination, wherein a gene is replaced with a marker or mutated allele. To deliver the recombination construct into the *Streptomyces* cell, non-replicative (suicidal) vectors are the first choice to facilitate selection of at least single crossovers before proceeding to segregation and selection of double crossover mutants (Kieser et al. 2000).

Non-replicative vectors are however unsuitable for strains in which DNA introduction or homologous recombination is not efficient, which is a requirement for stabilising the construct in the genome by a first crossover event before the non-replicative construct is lost. Temperature-sensitive replicative vectors are then the choice, to facilitate the selection of transformants (or exconjugants) without the requirement of homologous recombination to maintain the construct. The stable presence of the construct inside the cell facilitates the required homologous recombination crossover events, while being able to counter-select non-recombinants by cultivation at the non-permissive higher temperature (Muth 2018). These vectors are however very stable and cultivation at the required temperature to abolish replication has been suggested to be mutagenic (Mo et al. 2019), if possible at all for many streptomycetes.

One possibility to address this problem is to employ vectors that are replicative but can be easily cured, either due to intrinsic instability or that carry genetic elements that facilitate forcing the loss of the vector. pIJ101 is a natural plasmid isolated from *Streptomyces lividans* ISP 5434 (Kieser et al. 1982) whose replication functionality has been extensively used in multiple vectors for genetic manipulation in *Streptomyces* (Kieser et al. 2000). pIJ101-*ori-rep*-based vectors have been constructed mostly for ectopic gene expression but also for cloning and gene replacement (Kieser et al. 1982; Kendall and Cohen 1988; Brasch and Cohen 1995). Particularly useful for gene replacement are pIJ101-derived vectors with decreased segregational stability (Sun et al. 2009). pIJ86 is an *Escherichia coli–Streptomyces* bifunctional conjugatable vector constructed by Helen Kieser (John Innes Centre, Norwich, UK; Mervyn Bibb, personal communication) as a gene expression vector with the constitutive promoter of the erythromycin resistance gene (*ermE**p), and has been successfully used as such for protein production in *Streptomyces* (Berini et al. 2020). pIJ86 can be mobilised to streptomycetes by RK2/RP4 conjugation from *E. coli*. pIJ86 relies on the origin of replication (*ori*) and replication protein (*rep*) gene from plasmid pIJ101 for replication in *Streptomyces* and it carries the selectable marker *aac(3)IV* for apramycin resistance (for both, *Streptomyces* and *E. coli*). Although it was designed for gene expression and therefore with stability in mind, pIJ86 has been used for CRISPR-Cas9-based genome editing and shown to be easily lost if antibiotic selection is not maintained (Gomez-Escribano et al. 2021).

During a project focused on the identification of novel strains producers of phosphonate-containing natural products (Zimmermann et al. 2023, 2024), we encountered the need for genetic manipulation of *Streptomyces iranensis* DSM 41954 (original designation *S. iranensis* HM 35). This is a biotechnologically interesting strain as it is an alternative rapamycin producer and with promising potential as source for novel bioactive natural products (Horn et al. 2014). However, genetic manipulation of this strain is particularly challenging because introduction of DNA is inefficient (Netzker et al. 2016). To address this, we developed a vector that would enable to create mutants in this strain, and other streptomycetes that are difficult to genetically manipulate.

In this study, we describe the construction of a gene-targeting vector based on pIJ86. pDS0007 vector includes the *gusA* gene encoding the β-glucuronidase for visual screening of plasmid presence. It also carries the recognition sequence for the endonuclease I-SceI to force the loss of the vector when not lost through segregation, and, more importantly, to force the DNA double-strand break repair by homologous recombination, thereby facilitating the appearance of gene-replacement mutants (Siegl et al. 2010; Fernández-Martínez and Bibb 2014). We prove that the vector is quickly lost without antibiotic selection despite being self-replicative, and investigate possible genetic causes of this rapid curing. We further test its usability and advantages by constructing a gene-replacement mutant in the phosphonate natural compound producer *S. iranensis* DSM 41954.

Materials and methods

Bacterial strains, plasmids and cultivation conditions

Plasmids and strains used or generated during this study are listed in Supplementary Tables S1 and S2 respectively.

pIJ86 (*aac(3)IV, RK2-oriT, ColE1-ori, ermE*p, pIJ101-ori-rep*) and pSET152 (Bierman et al. 1992) (*aac(3)IV, RK2-oriT, ColE1-ori, phiC31-int, phiC31-attP*) and their DNA sequences were a gift from JIC StrepStrains (jic.strepstrains@jic.ac.uk, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK). pGM1190 (Muth 2018) (*tsr, aac(3)IV, oriT, to* terminator *tipA*p, RBS, *fd* terminator, *sso*, rep_{ts}) and pGus21 (*aac(3)IV, oriT, ermE*p_gusA, oripMB1*; knock-out vector, non-replicative in *Streptomyces*) and their DNA sequence were a gift from Günther Muth (Sigle et al. 2016) (Eberhard Karls Universität Tübingen, Geschwister-Scholl-Platz, 72,074 Tübingen, Germany). Restriction endonucleases, T4-DNA ligase, Q5 and One-Taq DNA polymerases, DNA Polymerase I-Large (Klenow) Fragment, were purchased from New England Biolabs (New England Biolabs GmbH, Brüningstraße 50; Geb. B852 (Industriepark Höchst), D-65926 Frankfurt am Main, Germany) and used according to provider's instructions.

Streptomyces coelicolor strains M145 (Kieser et al. 2000) and M512 (Floriano and Bibb 1996), and *Escherichia coli* ET12567/pUZ8002 (MacNeil et al. 1992; Paget et al. 1999) were a gift from JIC StrepStrains; *Streptomyces iranensis* DSM 41954 (Hamedi et al. 2010) was sourced from the German Collection of Microorganisms and Cell Cultures (Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Inhoffenstraße 7 B, 38,124 Braunschweig, Germany); *E. coli* DH5α (Bethesda Research Laboratories 1986; Grant et al. 1990) was purchased from New England Biolabs (as NEB5-alpha) or sourced from DSMZ collection (strain DSM 6897).

E. coli cultivation and general molecular biology techniques were performed following established methods (Sambrook et al. 1989) and instructions provided by suppliers. *Streptomyces* strains were cultivated in SFM (MS) for preparation of spore stocks, and for mobilisation of plasmid constructs to *Streptomyces* strains by conjugation from *E. coli*, which were done according to established methods (Kieser et al. 2000) or with the modifications by Netzker and co-workers (Netzker et al. 2016) when indicated for *S. iranensis*. Apramycin and kanamycin were used at 50 mg/L, fosfomycin at 20 mg/L, final concentration for both *E. coli* and *Streptomyces*. For the comparative conjugation efficiency studies, the same culture of donor *E. coli* was split into two equal fractions and mixed with 0.5 mL of pregerminated spores of either *S. coelicolor* or *S. iranensis*; efficiency was calculated as number of colonies that appeared on the conjugation plates (full accurate count for *S. iranensis* as the number was very low, estimation for *S. coelicolor* as the number was very high) divided by the estimated number of viable spores used.

Genomic DNA was purified following the salting-out method (Kieser et al. 2000) from mycelium of *S. iranensis* cultivated in liquid S medium (10 g glucose, 10 g glycerol, 4 g peptone, 4 g yeast extract, 4 g K_2HPO_4, 2 g KH_2PO_4, 0.5 g $MgSO_4$ dissolved in 1 L distilled water) in baffled flasks and orbital shaking (160 rpm) at 28 °C for 7 days.

DNA sequencing and sequence analysis

DNA was sequenced using the Sanger sequencing (Sanger et al. 1977) and Oxford Nanopore Technologies-ONT (Oxford Nanopore Technologies plc. Gosling Building, Edmund Halley Road, Oxford Science Park, OX4 4DQ, UK). Whole Plasmid Sequencing services provided by Eurofins Genomics (Eurofins Genomics Europe Shared Services GmbH; Anzinger Str. 7a; 85,560 Ebersberg; Germany). Informatic DNA sequence analysis was performed with Clone Manager 11 (Sci Ed Software LLC, Westminster, Colorado, USA), ApE-A plasmid Editor v3.0.8 (Davis and Jorgensen 2022), Artemis release 17.0.1 (Rutherford et al. 2000) and Staden Package version 2.0.0b11 - 2016 (Staden et al. 1999; Bonfield and Whitwham 2010). Annotation of plasmid genetic elements was done with pLannotate (McGuffie and Barrick 2021). Search and annotation of putative biosynthetic gene clusters were performed with antiSMASH version 7 (Blin et al. 2023). Homology search was performed at NCBI BLAST server (Altschul et al. 1997) and locally with prfectBLAST (Santiago-Sotelo and Ramirez-Prado 2012).

pIJ86 de novo sequence was first obtained by Sanger sequencing; reads obtained with oligonucleotides in Supplementary Table S3 were assembled with GAP4 of Staden Package (Staden et al. 1999; Bonfield and Whitwham 2010) to generate a consensus. Independently, pIJ86 was sequenced with Oxford Nanopore Technologies (Eurofins Genomics) (which with a coverage of 854 × provided a reliable sequence that fully matched the manually curated Sanger assembly). The ONT sequence was annotated on first instance with pLannotate (McGuffie and Barrick 2021) and then manually curated. The KpnI site, most likely used for the cloning of *ermE**p, was used as the starting position for the circular sequence. The final annotated sequence was deposited at NCBI under accession number PQ361717.

β-glucuronidase assay

Detection of β-glucuronidase activity was performed based on the previously published methods (Myronovskyi et al. 2011) with modifications: X-Gluc, 5-Bromo- 4-chloro-3-indolyl-β-D-glucuronide, was purchased as powder (Carl Roth GmbH + Co. KG, catalogue number 0018.1; AppliChem GmbH catalogue number A1113,0001) and dissolved in N,N′-dimethylformamide to 40 mg/mL stock

solution, prepared just before using it (due to the short shelf-life of the stock, even at − 20 °C); X-Gluc was added to melted SFM medium at 55 °C at a final concentration of 40 mg/L.

Molecular cloning

Oligonucleotides used during this study are listed in Supplementary Table S3.

Construction of pDS0201 Two homologous DNA fragments of about two kilobases (kb) flanking *S. iranensis* DSM 41954 *pepM* (which lays between positions 128,908 and 129,582 of sequence with accession JAGGLR010000028.1) were PCR amplified with oligonucleotides AZ029-AZ030 (downstream) and AZ031-AZ033 (upstream) (Table S1) and Q5 high-fidelity DNA polymerase (New England Biolabs). Each PCR amplicon was first cloned as a blunt-ended DNA fragment in pBluescript II KS + linearized with SmaI, and their correctness and directionality in the cloning vector were assessed by Sanger sequencing with universal primers M13 F- 24 mer and M13R- 22 mer (Table S1). The upstream fragment was excised with EcoRI-HindIII (sites present in primers) and ligated to pGus21 linearized with the same enzymes. The kanamycin resistance gene *neo* was excised from pTC192-Km (Rodríguez-García et al. 2006) with XbaI and ligated to pBluescript-downstream fragment linearized with XbaI, and the desired orientation selected. The downstream-*neo* insert was then excised with HindIII and ligated to pGus21-upstream linearized with HindIII, and the correct orientation was selected. This resulted in the construct pDS0201.

Construction of pDS0202 The homologous recombination cassette was excised from pDS0201 as a 4.8 kb NruI fragment (sites for NruI were located at the edge of the homologous fragments) and ligated to pDS0007 linearized with EcoRV, yielding pDS0202.

Construction of pDS0204 A DNA fragment containing *S. iranensis pepM* was PCR amplified with oligonucleotides AZ072-AZ073 and cloned into pBluescript II KS + linearized with SmaI. This construct was used as template for a nested PCR with oligonucleotides AZ078-AZ079, which carry recognition sequence for NdeI and HindIII restriction enzymes; Q5 high-fidelity DNA polymerase (New England Biolabs) was used for all PCR reactions. The AZ078-AZ079 PCR product was then digested with these enzymes and ligated to pRM4 linearized with the same enzymes, leading to plasmid pDS0204. Selected clones were confirmed by Sanger sequencing with universal primers M13 F- 24 mer and M13R- 22 mer.

Construction of pDS0007 pIJ86 was cut with HindIII, ends refilled and cut with XbaI. pGus21 (Sigle et al. 2016) was cut with XhoI, ends refilled with DNA Polymerase I-Large (Klenow) Fragment and cut with XbaI; the blunt-XbaI DNA segment carrying *ermE**p/RBS/*gusA*/I-SceI-site was purified from an agarose gel after electrophoresis and ligated to the blunt-XbaI linearized pIJ86 vector to generate pDS0007. One of the clones that showed the expected arrangement on agarose electrophoresis was confirmed by whole plasmid sequencing with Oxford Nanopore Technologies (coverage 676x, Eurofins Genomics), annotated with pLannotate (McGuffie and Barrick 2021) and then manually curated. The unique XbaI site was used as the starting position for the circular sequence. The final annotated sequence of pDS0007 was deposited at NCBI with accession number PQ356203. pDS0007 was deposited at the DSMZ open collection (number DSM 119277).

^{31}P NMR analysis for phosphonate detection

S. iranensis strains were cultivated in 100-mL Erlenmeyer baffled flasks containing 30 mL of R5 medium (Kieser et al. 2000), with orbital shaking (180 rpm) at 28 °C for 3 to 4 days. These cultures (5 mL) were used to inoculate 30 mL of GUBC medium in 100-mL baffled flasks (GUBC medium: 10 g saccharose, 5 g meat extract, 5 g casamino acids, 5 g glycerol, 5 mL 1 M Na_2HPO_4/KH_2PO_4 pH 7.3, 2 mL Hunter's base dissolved in 1 L distilled water, pH adjusted to 7.3 and 10 mL Balch's vitamins added after autoclaving. Hunter's concentrated base: 20 g nitrilotriacetic acid, 14 g KOH, 59.3 g $MgSO_4 \cdot 7H_2O$, 6.67 g $CaCl_2 \cdot 2H_2O$, 0.0185 g $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 0.198 g $FeSO_4 \cdot 7H_2O$, 0.25 g EDTA, 1.095 g $ZnSO_4 \cdot 7H_2O$, 0.5 g $FeSO_4 \cdot 7H_2O$, 0.154 g $MnSO_4 \cdot H_2O$, 0.0392 g $CuSO_4 \cdot 5H_2O$, 0.025 g $Co(NO_3)_2 \cdot 7H_2O$, 0.0177 g $Na_2B_4O_7 \cdot 10H_2O$ dissolved in 1 L distilled water, pH adjusted to 6.8. Balch's vitamins: 5 mg p-aminobenzoic acid, 2 mg folic acid, 2 mg biotin, 5 mg nicotinic acid, 5 mg calcium pantothenate, 5 mg riboflavin, 5 mg thiamine HCl, 10 mg pyridoxine HCl (B6), 100 µg cyanocobalamin (B12), 5 mg thioctic acid (lipoic acid) dissolved in 1 L distilled water, pH adjusted to 7.0, sterilised by filtration). After 7 days, cultures were harvested by centrifugation at 4000 rcf and 4 °C for 15 min. Supernatants were applied to C18 solid-phase extraction (SPE) cartridges (1 g), and the aqueous flow-through was collected. Aqueous phases were diluted with methanol to an 80% final methanol concentration, filtered and dried using a rotary evaporator. ^{31}P NMR spectra were recorded in 20% D_2O at 243 MHz on a Bruker Avance III HDX 600 MHz spectrometer fitted with a 5-mm Prodigy BBO H&F CryoProbe. Phosphorus chemical shifts are reported in ppm relative to a 10 mM L-phosphinothricin HCl reference (δ_P 51.2) using a

double-chamber coaxial NMR tube. NMR data were analysed using MestReNova 14.3.0.

Results

Construction of pDS0007

Based on previous insights on the facile curing of pIJ86-based plasmids (Gomez-Escribano et al. 2021), we chose this vector as the backbone for our gene-targeting vector. To be able to visually screen colonies for the presence or absence of the vector, we chose to include the gene *gusA* in the backbone, which encodes the β-glucuronidase enzyme optimised for *Streptomyces* codon usage (Myronovskyi et al. 2011), expressed from the constitutive promoter *ermE**p (Bibb et al. 1994) and an optimised ribosome binding site (RBS) for *Streptomyces* (Kieser et al. 2000). To be able to force the loss of the vector, and the appearance of double crossovers, by directed double-strand DNA break, we added the recognition sequence for the yeast intron-encoded endonuclease I-SceI (Siegl et al. 2010; Fernández-Martínez and Bibb 2014).

The DNA segment carrying *ermE**p/RBS/*gusA*/I-SceI-site was excised from pGus21 (Sigle et al. 2016) and ligated to pIJ86 (substituting the original *ermE**p promoter from pIJ86) to generate vector pDS0007 (Fig. 1; see "Materials and methods" for details). While there is no remaining multi-cloning site, the deletion cassette can be cloned into several unique restriction sites that are not expected to interfere with plasmid functions: XbaI (used during the construction of pDS0007), HindIII (regenerated at the XhoI-HindIII scar), BglII, EcoRI, NruI or EcoRV (which we used in this study). A full de novo sequence was obtained (see "Materials and methods" and below) and deposited at NCBI with accession number PQ356203. The plasmid is available from the DSMZ open collection with number DSM 119277.

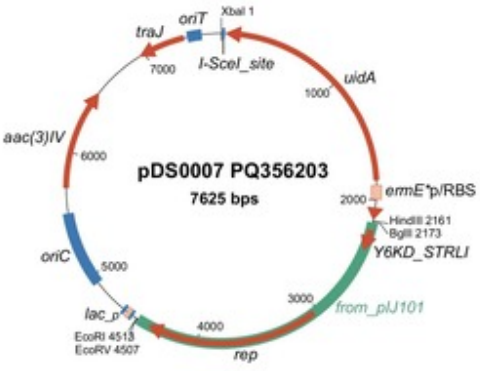


Fig. 1 Genetic map of pDS0007, including unique restriction sites for the addition of homologous recombination cassettes or other DNA fragments

Assessment of the replication region of pDS0007 derived from pIJ101

An alignment of the pIJ86 sequence (in silico, provided by JIC StrepStrains; Fig. S1) and pHZ1358 sequence (another pIJ101-derived vector, Genbank accession AY667410; Fig. S2) showed that the pIJ101 (Fig. S3) replication region present in pIJ86 is missing the *sti* locus (4366 to 4441 bp in pHZ1358 annotated sequence). A detailed analysis is given in Supplementary Information (Figs. S4 and S5). However, the available pIJ86 sequence was only an in silico compiled sequence with no further information, making it essential to obtain a de novo sequence of the full pIJ86 vector, which we obtained with both Sanger sequencing and ONT independently (see "Materials and methods"); we also obtained a de novo sequence of the replication region directly from pDS0007 (between position 2684 and 6814 bp of the NCBI PQ356203 sequence) by Sanger sequencing, and a whole plasmid sequence with Oxford Nanopore (see "Materials and methods"). Alignment of these sequences with the in silico sequence of pIJ86, and deposited ones of pHZ1358 (Genbank accession AY667410) and pIJ101 (Genbank accession M21778) revealed that the in silico sequence of pIJ86 is very accurate, and we confirmed that the pIJ101 replication region present in pIJ86 and pDS0007 lacks both the *sti* region and the entire sequence between *rep* and *sti* (Figs. S6 to S10; see Supplementary Information for full details). pIJ86 was known to lack the BamHI site present in the wild-type pIJ101 *rep*, but the actual sequence was unknown (filled with "Ns" in the original sequence). Analysis of the pIJ86 de novo sequence showed that the former BamHI site has the same sequence as in pH1358, which, along with the other single nucleotide mismatches identified (see Figs. S7 to S10), indicates that the pIJ101 *rep* region in pIJ86 and pHZ1358 was derived from the same or a closely related predecessor plasmid (one of those previously reported as a having lost the BamHI site by spontaneous mutation Kieser et al. 1982; Sun et al. 2009)).

Test of conjugation efficiency of pDS0007 and other vectors

Efficiency of plasmid vector mobilisation by conjugation to *S. iranensis* DSM 41954 has been described to be very low (Netzker et al. 2016). We were also unsuccessful obtaining any exconjugants with pDS0201 (described in "Materials and methods", discussed below), a suicidal vector carrying two homologous fragments, each two kilobases long,

regardless of the modifications to the conjugation protocol we applied, including the previously published protocol (Netzker et al. 2016). To test the conjugation efficiency of the newly generated pIJ86-derivative vector pDS0007, we undertook a comparative study with the model strain *S. coelicolor* M145. As control vectors, we used the phiC31-integrative vector pSET152 and the self-replicative vector pGM1190 (Muth 2018) based on the temperature-sensitive *Streptomyces* origin of replication pSG5 used previously in *S. iranensis* (Netzker et al. 2016). All three vectors are *E. coli–Streptomyces* bifunctional, conjugatable and selected for with apramycin in both *E. coli* and *Streptomyces*. Exconjugants were obtained from both strains with all three vectors, but while the conjugation plates were completely covered with exconjugants in the case of *S. coelicolor* M145 for all vectors (estimated efficiency $> 6 \times 10^{-4}$ exconjugants per spore), an efficiency of only 8.6×10^{-7} (pDS0007), 2.6×10^{-7} (pSET152) and 1.6×10^{-6} (pGM1190) exconjugants per spore was obtained for *S. iranensis*, i.e. between at least two and three orders of magnitude lower efficiency. The final conjugation protocol used was the standard developed for *S. coelicolor* (Kieser et al. 2000), since employing previously published modifications (Netzker et al. 2016) did not appear to have a significant and reproducible effect.

Test of stability and segregation properties of pDS0007

The aforementioned observed frequent loss of pIJ86-derived vectors (Gomez-Escribano et al. 2021) was tested in the model strain *S. coelicolor* M145 and in *S. iranensis* DSM 41954. The experimental design to test whether the newly generated pIJ86-derivative vector pDS0007 was stably maintained or readily lost without antibiotic selection, and in comparison to pGM1190, is depicted in Supplementary Information Fig. S11. Exconjugants were replicated on SFM with antibiotic selection for the vector (apramycin), followed by one sporulation round on SFM, either with apramycin or without, as to force the maintenance of the vector or to allow for its loss, respectively. After preparation of spore stocks from each condition, the same volume of a diluted spore stock was plated on SFM with or without apramycin. Colonies developed on each plate were counted after reaching about 2 mm diameter (i.e. allowing sufficient time for all possible colonies to appear). The results showed that when selecting with antibiotic, between 70 and 87% of the spores maintained the plasmids; when spores were developed without antibiotic selection, only 25–28% of spores maintained pGM1190 (with cultivation at the permissive temperature of 30 °C); with pIJ86-derived pDS0007, the plasmid was almost completely lost in *S. coelicolor* and *S. iranensis* (Figs. S12 to S15). Remarkably, in *S. coelicolor*, the vector was almost totally lost as the colonies developed before replica-plating (Fig. S13). An independent experiment to test plasmid loss during growth was a β-glucuronidase assay to visually discriminate colonies originating from spores that have maintained the plasmid (blue pigment production) versus colonies from spores that had lost the plasmid. This allowed screening on the same plate and therefore under exactly the same conditions. The results showed that pDS0007 was lost almost completely after just one round of sporulation without selection, in both *S. coelicolor* M145 and M512 (Figs. S12 to S14), as well as *S. iranensis* (Figs. 2 and S15). Thus, growing under antibiotic, selection forces the maintenance of the vector in most of the spores. It is worth noting that, in both experimental strategies, we were comparing the same spore preparations under either condition and calculating ratios, rather than absolute numbers, of colony count. Therefore, we do not expect our results being influenced by differences of plating efficiency between spore preparations or strains, and the results are a genuine representation of the poor segregational stability of pDS0007.

Construction of *S. iranensis* Δ*pepM::neo* by using pDS0007 for mutagenesis

As part of a study to screen for new phosphonate antibiotic producers (Zimmermann et al. 2024), *S. iranensis* has recently been identified as a potential phosphonate producer due to the presence of a *pepM* gene in the available draft genome sequence (Horn et al. 2014). *pepM* encodes the enzyme phosphoenolpyruvate mutase (PepM), which is the first and essential enzyme of phosphonate biosynthesis (Kayrouz et al. 2020; Li and Horsman 2022). The putative phosphonate biosynthetic gene cluster is depicted in Fig. 3a. To further study this gene cluster and to uncover its metabolic product, we decided to construct a *pepM* deletion mutant of *S. iranensis* DSM 41954.

For this purpose, a homologous recombination cassette was constructed, with the kanamycin resistance gene *neo* between two homologous DNA fragments of about two kilobases flanking *pepM* (see "Materials and methods" for details). This cassette was first cloned in the suicidal vector pGus21 (construct pDS0201; Fig. S15) and mobilised by conjugation to *S. iranensis* DSM 41954. However, no exconjugants were obtained with this construct under a variety of conditions including the standard ones for *Streptomyces* (Kieser et al. 2000) and those previously published as optimised for *S. iranensis* (Netzker et al. 2016). Thus, we devised the development of pDS0007. The recombination cassette from pDS0201 was transferred to pDS0007 (see "Materials and methods") and the resulting construct pDS0202 was mobilised to *S. iranensis* DSM 41954. Exconjugants selected for kanamycin resistance were readily obtained with the standard conjugation conditions (Kieser et al. 2000), replicated on SFM supplemented with

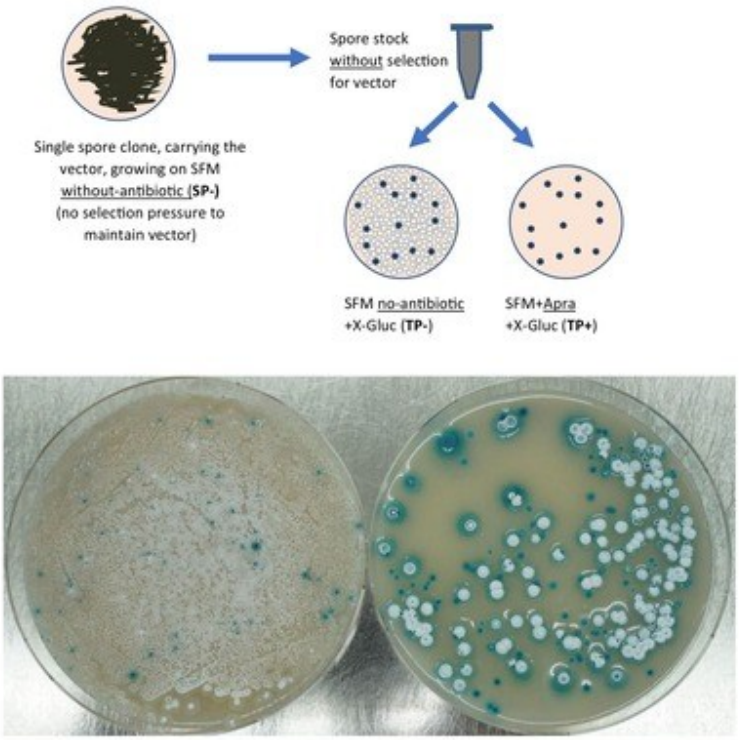
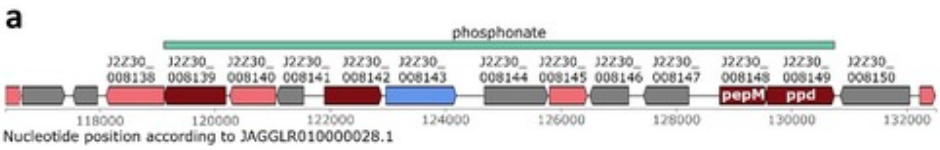


Fig. 2 Segregation stability test of pDS0007 in *S. iranensis* DSM 41954. The same volume of a spore preparation obtained from a plate without apramycin selection (i.e. an **SP-** plate in Fig. S11) was used to inoculate two SFM plates supplemented with either only X-Gluc (**TP-**, left) or X-Gluc and apramycin (**TP +**, right). The plate at the left shows confluent growth, indicative of a very high number of colony-forming units inoculated, but only some show conversion of X-Gluc to blue pigment, indicating the very low proportion of spores that have maintained the plasmid. The plate at the right only allows the growth of spores that maintain the plasmid (apramycin selection) and shows a number of individual colonies all of them showing conversion of X-Gluc to blue pigment

kanamycin and fosfomycin, and then re-streaked on SFM (without antibiotic selection) for sporulation. The expected phenotype of the mutants is kanamycin resistant, apramycin sensitive and β-glucuronidase negative, due to the integration of the *neo* cassette as a result from homologous recombination by a double crossover event, with simultaneous loss of the vector backbone and thus loss of the *gusA* gene and *apra* cassette. Spores of this "first round" were plated on SFM supplemented with X-Gluc and kanamycin only, allowing growth of both single and double crossover mutants, and enabling screening for presence of vector by the β-glucuronidase-driven conversion of X-Gluc to a blue pigment. Colonies that grew in the presence of kanamycin but did not show blue pigment production were replicated on SFM supplemented with X-Gluc and kanamycin, and then streaked on SFM for sporulation. This "second round" of segregation through sporulation was conducted to ensure that no single crossover or cells with the self-replicating vector were carried over. Spores from the "second round" were plated on SFM supplemented with X-Gluc and kanamycin only. None of the colonies showed blue pigment production nor grew upon replication on SFM with apramycin, clearly indicating they had completely lost the vector backbone, and therefore they were likely double crossovers and candidate mutants. It must be noted that we did not make use of the endonuclease I-SceI to force the loss of vector or second recombination event, since sufficient candidate clones were readily obtained. Ten clones were screened for the expected Δ*pepM::neo* mutant genotype by PCR and all of them showed the expected PCR result (Fig. S16). Two of the candidates were further assessed with additional PCR reactions and Sanger sequencing of the PCR products, and confirmed to have the expected Δ*pepM::neo* genotype (Fig. S17).

Functional proof of *pepM* deletion in *S. iranensis* by loss of phosphonate production

To prove that the *pepM* gene is involved in phosphonate production in *S. iranensis*, the S. *iranensis* wild-type strain, as well as the Δ*pepM::neo* mutant, was cultivated under phosphonate producing conditions and then analysed for phosphonate production using ^{31}P NMR spectroscopy. ^{31}P NMR studies revealed putative phosphonate signals (δ_p 33.4, δ_p 21.5, δ_p 17.0) for the *S. iranensis* wild-type



Identifier	Product	Length (nt)	Length (aa)
J2Z30_008137	Unknown	429	142
J2Z30_008138	SMCOG1079:oxidoreductase; PF01408.25 (Oxidoreductase family, NAD-binding Rossmann fold)	819	272
J2Z30_008139	SMCOG1056:DegT/DnrJ/EryC1/StrS aminotransferase	1011	336
J2Z30_008140	SMCOG1207:inositol monophosphatase	801	266
J2Z30_008141	PF00293.31 (NUDIX domain)	462	153
J2Z30_008142	PF11583.11 (P-aminobenzoate N-oxygenase AurF)	993	330
J2Z30_008143	SMCOG1020:major facilitator transporter	1221	406
J2Z30_008144	PF02195.21 (ParB/Sulfiredoxin domain)	1092	363
J2Z30_008145	SMCOG1115:HAD-superfamily hydrolase, subfamily IA	654	217
J2Z30_008146	PF01022.23 (Bacterial regulatory protein, arsR family	672	223
J2Z30_008147	Unknown	783	260
J2Z30_008148	Phosphoenolpyruvate phosphomutase, PEP_mutase, SMCOG1231:carboxyvinyl-carboxyphosphonate	678	225
J2Z30_008149	Phosphonopyruvate decarboxylase; phosphonate TPP_enzyme_C/TPP_enzyme_N; SMCOG1055:pyruvate oxidase/decarboxylase	1167	388
J2Z30_008150	PF02670.19 (1-deoxy-D-xylulose 5-phosphate reductoisomerase)	1200	399

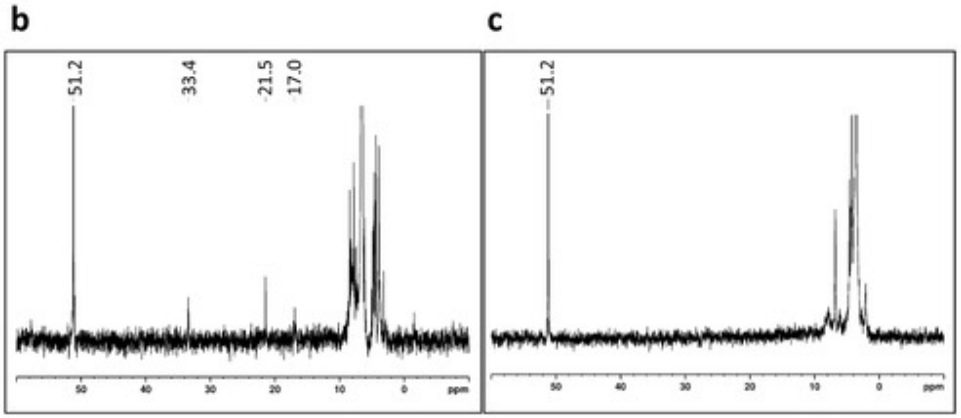


Fig. 3 Phosphonate BGC and production of *S. iranensis* DSM 41954. **a** Phosphonate BGC as annotated by antiSMASH on sequence with accession JAGGLR010000028.1. **b** ^{31}P NMR spectrum of *S. iranensis* DSM 41954 wild-type showing production of phosphonate-containing metabolites (δ_P 33.4, 21.5 and 17.0). **c** ^{31}P NMR spectrum of *S. iranensis* Δ*pepM::neo* mutant showing that production of phosphonate-containing metabolites has been abolished. L-phosphinothricin HCl (10 mM, δ_P 51.2) was used as a chemical shift reference

samples (Fig. 3b). These signals were absent in the sample of the Δ*pepM::neo* mutant (Fig. 3c) but they were restored upon genetic complementation by ectopic expression of *pepM* from the integrative construct pDS0204 (data not shown). These results show phosphonate production by the *S. iranensis* wild-type strain, which was abolished in the Δ*pepM::neo* mutant, demonstrating that *pepM* is essential for phosphonate production in *S. iranensis* and proving that a *pepM* mutant was successfully generated.

Discussion

Homology recombination-based genetic manipulation of streptomycetes is often hampered by low efficiency of DNA introduction into the cells combined with low recombinogenicity of the particular strain. Facilitating the persistence of recombination constructs inside the cell increases the chances for homologous recombination. Replicative vectors with temperature-sensitive replication mechanism have been widely adapted as gene-targeting vectors (Bierman et al. 1992; Muth 2018) but it is often not desirable or possible at all to cultivate streptomycetes strains at the non-permissive temperature. In this work, we have constructed a vector that, while being replicative and stably maintained with antibiotic selection, is quickly lost without antibiotic selection, therefore behaving like a suicidal vector.

The vector described in this work, pDS0007, carries the genetic functionality for replication in *E. coli* (colE1 *ori*) and *Streptomyces* (*sti*¯ pIJ101 *ori-rep*), transfer by conjugation from *E. coli* to *Streptomyces* (RK2 = RP4 *ori*), selection in both bacteria with apramycin (*aac(3)IV* marker) and visual screening for vector presence in *Streptomyces* by β-glucuronidase assay (*gusA* = *uidA* marker). The presence of the I-SceI recognition site allows to force the loss of the vector backbone, both when persisting as a self-replicative extrachromosomal plasmid or when integrated as a single crossover, thereby facilitating the appearance of double crossovers. This can be achieved by mobilising a second construct, such as pIJ12742, which expresses a *Streptomyces* codon–optimised gene encoding the I-SceI endonuclease (Fernández-Martínez and Bibb 2014). We have shown that the vector is quickly lost in two *Streptomyces* strains (two derivatives of the model strain *S. coelicolor* A(3)2 and *S. iranensis*) when antibiotic selection is not maintained, and single-spore clones cured from the vector are obtained at very high frequency after just one round of sporulation.

pDS0007 is based on the backbone of pIJ86. The details of the construction process of pIJ86 were never published and only an in silico sequence was available. The thorough analysis of a pIJ86 de novo sequence, together with the information provided by JIC StrepStrains, indicates that pIJ86 is likely a derivative of pSET152 through an intermediate construct named pIJ85, which was also never published (Mervyn Bibb, personal communication). The most likely construction step towards pIJ85 was the ligation of the SacII-SphI fragment of pSET152 (sequence accession AJ414670) that carries the pUC18 *oriC*, the apramycin resistance marker *aac(3)IV* and RP4 conjugation functions (lacking the phiC31 integrative functionality) with a fragment excised from a pIJ101-derived vector that carries the replications functions, including a BamHI-free *rep* gene (the SacII segment between the sites at position 8489 bp and 1970 bp as in pIJ101 sequence with accession M21778). The constitutive *ermE**p promoter (Bibb et al. 1994) was then added between the *oriT* and the pIJ101 replication fragment of pIJ85, as a KpnI-HindIII fragment, leading to pIJ86 (as deduced from the sequence information supplied by JIC StrepStrains). It is noticeable that pIJ86 does not contain an optimised ribosome binding site or an NdeI site for cloning of the gene to be expressed. Instead, there is a multi-cloning site with unique sites for BamHI, SphI, HindIII and BglII, that allows the cloning of the gene to be expressed. Another unique restriction site that can be used for cloning additional fragments is the EcoRV site downstream of *rep*, as used in this work for the cloning of the cassette for homologous recombination.

Vectors based on pIJ86 backbone had been previously shown to be segregationally unstable and easily lost without antibiotic selection (Gomez-Escribano et al. 2021) but the reasons remained unknown, and other parts of the construct could have been responsible for the observed high-frequent loss (e.g. CRISPR-Cas9 functionality). During a thorough study of the cloning gene-targeting vector pHZ1358, also based on pIJ101-*ori-rep*, Sun and co-workers showed that deletion of the *sti* locus ("site of second strand synthesis " (Sun et al. 2009) previously called "strong incompatibility locus"(Deng et al. 1988)), located downstream of the replication protein–encoding gene *rep*, results in increased instability and frequent plasmid loss, although its incidence seemed to be strain-dependent (Sun et al. 2009). Our thorough analysis with a newly obtained de novo sequence of pIJ86 shows that pIJ86 lacks the *sti* locus present in pIJ101 and pHZ1358. Sun and co-workers (Sun et al. 2009) referenced previously published work indicating that the lack of *sti* leads to a low plasmid copy number (Deng et al. 1988). The low copy number could be the reason for the quick loss of plasmid from mycelial cellular compartments and, since pIJ86 does not apparently carry any specific plasmid partitioning mechanism that ensures plasmid segregation to spores, the high-frequent number of spores that do not carry the plasmid. It is unlikely that the authors of pIJ86 were aware of the relevance of *sti*, since the pHZ1358 work was published in 2009 (Sun et al. 2009), while pIJ86 was constructed sometime between 2000 (the vector is not included in the Practical Streptomyces Genetics manual (Kieser et al.

2000)) and 2006 (first time we were aware of this vector). Still, it is not certain whether the lack of *sti* and the rest of sequence between *rep* and *sti* is responsible for the observed highly frequent loss of vector, a question that nevertheless goes beyond the aim of this work.

Finally, we have successfully used the vector to create targeted mutants by homologous recombination-based gene replacement in *S. iranensis*, a strain known to be very difficult to manipulate (Netzker et al. 2016). pDS007 was successfully used to delete the gene *pepM* from a newly identified phosphonate biosynthetic gene cluster in *S. iranensis*. The mutants showed the expected genotype and phenotype, with abolished phosphonate production. We show therefore that *S. iranensis* is a natural producer of phosphonate-containing specialised metabolites and that we have correctly identified the responsible biosynthetic gene cluster.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00253-025-13477-3.

Acknowledgements We thank Imen Nouioui for providing us with S. iranensis DSM 41954 from the DSMZ culture collection and Christine Rohde and Johannes Wittmann for the help in depositing the plasmid pDS0007 in the DSMZ culture collection. We are grateful to Günther Muth and to JIC StrepStrains for providing plasmids, strains, and sequence information. We thank Mervyn Bibb for valuable comments on the manuscript.

Author Contribution YM, and CCH conceived the study. JPGE, YM, and CCH designed research. JPGE, AZ, SNX, MD and JM conducted experiments. JM contributed new methods. All authors analyzed data. JPGE, CCH and YM wrote the initial manuscript. All authors reviewed and contributed to the manuscript. All authors read and approved the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. We gratefully acknowledge the funding received from the German Center for Infection Research (DZIF) TTU 09.826. SX is grateful for a PhD scholarship (202008330294) from the Chinese Scholarship Council.

Data availability DNA sequences of vectors pIJ86 and pDS0007 were deposited at NCBI under accession numbers PQ361717 and PQ356203 respectively. pDS0007 was deposited at the DSMZ open collection (number DSM 119277).

Declarations

Ethical approval No ethical approval was needed for this work.

Competing interests The authors declare no competing interests.

References

Alberti F, Corre C (2019) Editing streptomycete genomes in the CRISPR/Cas9 age. Nat Prod Rep 36:1237–1248

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Baltz RH (2006) Molecular engineering approaches to peptide, polyketide and other antibiotics. Nat Biotechnol 24:1533–1540

Berini F, Marinelli F, Binda E (2020) Streptomycetes: attractive hosts for recombinant protein production. Front Microbiol 11:1958

Bethesda Research Laboratories (1986) BRL pUC host: E. coli DH5alpha competent cells. Focus 8:9

Bibb MJ, White J, Ward JM, Janssen GR (1994) The mRNA for the 23S rRNA methylase encoded by the ermE gene of Saccharopolyspora erythraea is translated in the absence of a conventional ribosome-binding site. Mol Microbiol 14:533–545

Bierman M, Logan R, O'Brien K, Seno ET, Nagaraja Rao R, Schoner BE (1992) Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. Gene 116:43–49

Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M et al (2023) antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res 51(W1):W46–W50

Bonfield JK, Whitwham A (2010) Gap5—editing the billion fragment sequence assembly. Bioinformatics 26:1699–1703

Brasch MA, Cohen SN (1995) Sequences essential for replication of plasmid pIJ101 in Streptomyces lividans. Plasmid 33:191–197

Davis MW and Jorgensen EM (2022) ApE, a plasmid editor: a freely available DNA manipulation and visualization program. Front Bioinform 2

Deng Z, Kieser T, Hopwood DA (1988) "Strong incompatibility" between derivatives of the Streptomyces multi-copy plasmid pIJ101. Mol Gen Genet 214:286–294

Fernández-Martínez LT, Bibb MJ (2014) Use of the Meganuclease I-SceI of Saccharomyces cerevisiae to select for gene deletions in actinomycetes. Sci Rep 4:7100

Floriano B, Bibb M (1996) afsR is a pleiotropic but conditionally required regulatory gene for antibiotic production in Streptomyces coelicolor A3(2). Mol Microbiol 21:385–396

Gomez-Escribano JP, Algora Gallardo L, Bozhüyük KAJ, Kendrew SG, Huckle BD, Crowhurst NA et al (2021) Genome editing reveals that pSCL4 is required for chromosome linearity in Streptomyces clavuligerus. Microbial Genomics 7:000669

Grant SG, Jessee J, Bloom FR, Hanahan D (1990) Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. PNAS 87:4645–4649

Hamedi J, Mohammadipanah F, Klenk H-P, Pötter G, Schumann P, Spröer C et al (2010) Streptomyces iranensis sp. nov., isolated from soil. Int J Syst Evol Microbiol 60:1504–1509

Horn F, Schroeckh V, Netzker T, Guthke R, Brakhage AA, Linde J (2014) Draft genome sequence of Streptomyces iranensis. Genome Announc 2:e00616-e714

Katz L, Baltz RH (2016) Natural product discovery: past, present, and future. J Ind Microbiol Biotechnol 43:155–176

Kayrouz CM, Zhang Y, Pham TM, Ju K-S (2020) Genome mining reveals the phosphonoalamide natural products and a new route in phosphonic acid biosynthesis. ACS Chem Biol 15:1921–1929

Kendall KJ, Cohen SN (1988) Complete nucleotide sequence of the Streptomyces lividans plasmid pIJ101 and correlation of the sequence with genetic properties. J Bacteriol 170:4634–4651

Kieser T, Hopwood DA, Wright HM, Thompson CJ (1982) pIJ101, a multi-copy broad host-range Streptomyces plasmid: Functional analysis and development of DNA cloning vectors. Molec Gen Genet 185:223–238

Kieser T, Bibb MJ, Buttner MJ, Chater KF, and Hopwood DA (2000) Practical streptomyces genetics, Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (eds) John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, England: John Innes Foundation

Li S, Horsman GP (2022) An inventory of early branch points in microbial phosphonate biosynthesis. Microbial Genomics 8:000781

MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T (1992) Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. Gene 111:61–68

McGuffie MJ, Barrick JE (2021) pLannotate: engineered plasmid annotation. Nucleic Acids Res 49:W516–W522

Mo J, Wang S, Zhang W, Li C, Deng Z, Zhang L, Qu X (2019) Efficient editing DNA regions with high sequence identity in actinomycetal genomes by a CRISPR-Cas9 system. Synth Syst Biotechnol 4:86–91

Muth G (2018) The pSG5-based thermosensitive vector family for genome editing and gene expression in actinomycetes. Appl Microbiol Biotechnol 102:9067–9080

Myronovskyi M, Welle E, Fedorenko V, Luzhetskyy A (2011) β-Glucuronidase as a sensitive and versatile reporter in actinomycetes. Appl Environ Microbiol 77:5370–5383

Netzker T, Schroeckh V, Gregory MA, Flak M, Krespach MKC, Leadlay PF, Brakhage AA (2016) An efficient method to generate gene deletion mutants of the rapamycin-producing bacterium Streptomyces iranensis HM 35. Appl Environ Microbiol 82:3481–3492

Paget MSB, Chamberlin L, Atrih A, Foster SJ, Buttner MJ (1999) Evidence that the extracytoplasmic function sigma factor σE is required for normal cell wall structure in Streptomyces coelicolor A3(2). J Bacteriol 181:204–211

Rodríguez-García A, Santamarta I, Pérez-Redondo R, Martín JF, Liras P (2006) Characterization of a two-gene operon epeRA involved in multidrug resistance in Streptomyces clavuligerus. Res Microbiol 157:559–568

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B (2000) Artemis: sequence visualization and annotation. Bioinformatics 16:944–945

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74:5463–5467

Santiago-Sotelo P, Ramirez-Prado JH (2012) prfectBLAST: a platform-independent portable front end for the command terminal BLAST+ stand-alone suite. Biotechniques 53:299–300

Siegl T, Petzke L, Welle E, Luzhetskyy A (2010) I-SceI endonuclease: a new tool for DNA repair studies and genetic manipulations in streptomycetes. Appl Microbiol Biotechnol 87:1525–1532

Sigle S, Steblau N, Wohlleben W, Muth G (2016) Polydiglycosylphosphate transferase PdtA (SCO2578) of Streptomyces coelicolor A3(2) is crucial for proper sporulation and apical tip extension under stress conditions. Appl Environ Microbiol 82:5661–5672

Staden R, Beal KF, and Bonfield JK (1999) The Staden Package, 1998. In *Bioinformatics Methods and Protocols*. Methods in Molecular Biology™. Misener, S. and Krawetz, S.A. (eds). Totowa, NJ: Humana Press, pp. 115–130

Sun Y, He X, Liang J, Zhou X, Deng Z (2009) Analysis of functions in plasmid pHZ1358 influencing its genetic and structural stability in Streptomyces lividans 1326. Appl Microbiol Biotechnol 82:303–310

Zimmermann A, Nouioui I, Pötter G, Neumann-Schaal M, Wolf J, Wibberg D, Mast Y (2023) Kitasatospora fiedleri sp. nov., a novel antibiotic-producing member of the genus Kitasatospora. Int J Syst Evol Microbiol 73:006137

Zimmermann A, Nouioui I, and Mast Y (2024) Genome sequence of the bialaphos producer Streptomyces sp. DSM 41527 and two putative phosphonate antibiotic producers Streptomyces sp. DSM 41014 and DSM 41981 from the DSMZ strain collection. Access Microbiol 6: 000770.v3

3 DISCOVERY OF 7-GLYCOSYL OXAZOLOMYCIN D, A FIRST GLYCOSYLATED MEMBER OF OXAZOLOMYCIN FAMILY

This chapter are reproduced and adapted from the following published article, of which I am a co-first author: Vitale, G.A.*; **Xia, S.N.***; Dührkop, K; Shahneh, M.R.Z; Broetz-Oesterhelt, H.; Mast, Y.; Brungs, C.; Böcker, S.; Schmid R.; Wang, M.; Hughes, C.C.; Petras, D. Enhancing tandem mass spectrometry-based metabolite annotation with online chemical labeling. Nature Communications 16, 6911 (2025). *The text, figures, and page images included here are used with full acknowledgment of the original source and solely for academic and degree requirements.*

The author's contributions are as follows:

The study concept was conceived by Daniel Petras and Chambers C. Hughes. The compound isolation, purification, all related experiments and data analysis were done by me. I developed and optimized the cultivation conditions and isolation procedures. The NMR experiments and analysis were performed by me and Chambers C. Hughes. The derivatization reactions and LC-MS/MS experiments were performed by Giovanni Andrea Vitale, Chambers C. Hughes and Daniel Petras. The initial data in this publication was curated by Giovanni Andrea Vitale, me, Chambers C. Hughes and Daniel Petras. Kai Dührkop, Mohammad Reza Zare Shahneh, Corinna Brungs, Sebastian Böcker, Robin Schmid, and Mingxun Wang developed software. Giovanni Andrea Vitale, Kai Dührkop, Mohammad Reza Zare Shahneh and Daniel Petras performed MCheM data analysis. Heike Brötz-Oesterhelt and Yvonne Mast. provided natural product standards and strains. Yvonne Mast performed BGC analysis. Giovanni Andrea Vitale, Chambers C. Hughes and Daniel Petras wrote the initial manuscript. All authors edited and approved the manuscript.

* Contributed equally

The preceding chapter showed that genome mining can effectively prioritise pepM$^+$ strains and biosynthetic targets but also highlighted the difficulty of isolation and purification. In microbial fermentation systems, metabolite identification is often complicated by low compound abundance, the co-occurrence of structurally related analogues, and substantial interference arising from complex sample matrices. Under such conditions, conventional detection and isolation strategies may be insufficient to distinguish novel metabolites from background chemical complexity. These limitations reveal a persistent gap between biosynthetic inference and metabolite-level validation and indicate the need for more sensitive and structurally informative analytical approaches. As described in chapter 1, one of the major challenges is metabolite annotation in metabolomics, even with the technical improvement in high-resolution liquid chromatography tandem mass spectrometry (LC-MS/MS) instrumentation and computational approaches. In typical MS-based analyses, fewer than 10% of detected features can be assigned with high confidence.[173]Although some studies have shown that a considerable fraction of unannotated signals arises from ion adducts and in-source fragments,[174]evidence from metabolomics,[175,176,177,178]genome mining,[8,179,180]and NP discovery consistently indicates that metabolite chemical space of metabolites is vast and mostly uncharted. One major limitation in metabolite identification is the relatively narrow coverage of existing spectral libraries compared with the breadth of chemical diversity encountered in biological samples.[181] To address this gap, recent studies have expanded the use of large structural databases through *in silico* annotation methods,[182] as well as de novo structure annotation methods.[183] Nevertheless, confident metabolite identification remains difficult, highlighting the need for workflows that generate complementary chemical information to support annotation and improve accuracy.

Chemical derivatization has been broadly applied to improve structure elucidation and overcome limits of detection by exploiting the unique reactivity of certain chemical moieties[184]. In the field of targeted metabolomics, post-column derivatization is used to enhance the detection of specific compounds[185]. In contrast, derivatization strategies in non-targeted metabolomics has mainly been applied as an offline batch step before LC-MS/MS analysis[186,187,188.] Post-column derivatization offers considerable but underused potential in this context because it can provide feature-specific information about the functional groups present in unknown compounds. Unlike batch derivatization, post-column strategies preserve chromatographic co-elution relationships between precursor ions and their derivatization products, an advantage that is especially valuable when analyzing complex mixtures. The structural insight obtainable in non-targeted metabolomics could be greatly expanded by combining multiple orthogonal derivatization chemistries. Such an approach enables functional groups to be interrogated through sequential or parallel reactions and allows the resulting information to be integrated into a single analytical stream for in silico spectral annotation and open-modification searching.

In this chapter, we present a Multiplexed Chemical Metabolomics (MCheM) workflow that combines multiple post-column derivatization reactions with non-targeted LC-MS/MS analysis. In addition, we established a data analysis pipeline based on ion identity networking in mzmine[189]to post-process MCheM data to support downstream metabolite annotation from MCheM experiments[190]. By integrating CSI:FingerID in silico annotation[191] and GNPS2 open modification search[129], we observed annotation improvements of 31.9% for CSI:FingerID and

37.6% for GNPS2 over an experimental library of ~10K experimental compounds, and of 48.8% for CSI:FingerID and 20.4% for GNPS2 on a set of authentic NP standards. More importantly, we rapidly identified the first glycosylated member of the oxazolomycin family, 7-glycosyl oxazolomycin D, which was identified from *Streptomyces libani subsp. rufus* DSM 41230 using MCheM in a genome-guided natural product discovery.

3.1 Results and discussion

3.1.1 Construction and validation of MCheM

To enhance and improve the metabolite annotation, a customized LC-MS/MS configuration was constructed for MCheM data acquisition, integrating a make-up UHPLC pump, a T-splitter or reactor manifold, and a syringe pump (**Figure 31A**). By using this configuration, three post-column derivatization workflows for targeting different functional groups compatible with LC-MS were established (**Figure 31C**). Specifically, the first workflow employed L-cysteine as a labeling reagent for electrophilic moieties.[186]. Such motifs are widespread in NPs, likely reflecting their evolutionary role in forming covalent interactions with nucleophilic amino acid residues including serine, cysteine, and threonine in biomolecular targets. The second workflow focused on amine and phenolic moieties through derivatization with the commercially available reagent 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC). Due to efficient labeling requires basic conditions to maintain amines in the free-base state, a 0.5% trimethylamine solution was introduced between the column outlet and the AQC stream via the make-up pump. This adjustment maintained the eluent at approximately pH 5–6 throughout MS acquisition. The third workflow used commercially available hydroxylamine hydrochloride as the labeling reagent for aldehyde and ketone compounds.[192]. Simultaneously, the analytical performance of three workflows was assessed by measuring concentration-dependent response linearity and detection limits. To further validate our MCheM, 359 structurally diverse NP standards from the Tübingen Natural Compound Collection were derivatized with established reaction, resulting high selectivity and negligible false-positive rates revealed our MCheM had the robust chemical specificity. Interestingly, nearly 50% of the evaluated spectra exhibited improved ranking positions by utilizing the CSI: FingerID algorithm, with several candidates promoted to top-confidence annotations. In addition, analog library searches via GNPS2 benefited from functional-group filtering, resulting in increased structural similarity scores and improved ranking of relevant analogs, even in the absence of exact spectral matches in public databases.

To enable high-throughput analysis of multiple derivatization reactions in complex metabolomics samples and to combine these measurements into a unified multiplexed dataset, we implemented a dedicated “Online Reactivity” module in MZmine. This strategy uses the co-elution of precursor ions and their reaction products to create correlation-based links(**Figure 31B**), using ion identity networking[190] together with user-defined Δm/z values assigned to each derivatization reagent. The resulting MCheM output is a composite dataset that brings together

MS, MS/MS, and reactivity-derived information. In addition, it provides predicted functional groups or substructures encoded as SMILES Arbitrary Target Specification (SMARTS)[193]. These reactivity-informed annotations can be passed directly to downstream tools such as CSI:FingerID[182] or GNPS2[129] to constrain the candidate structural search space(**Figure 31B**).

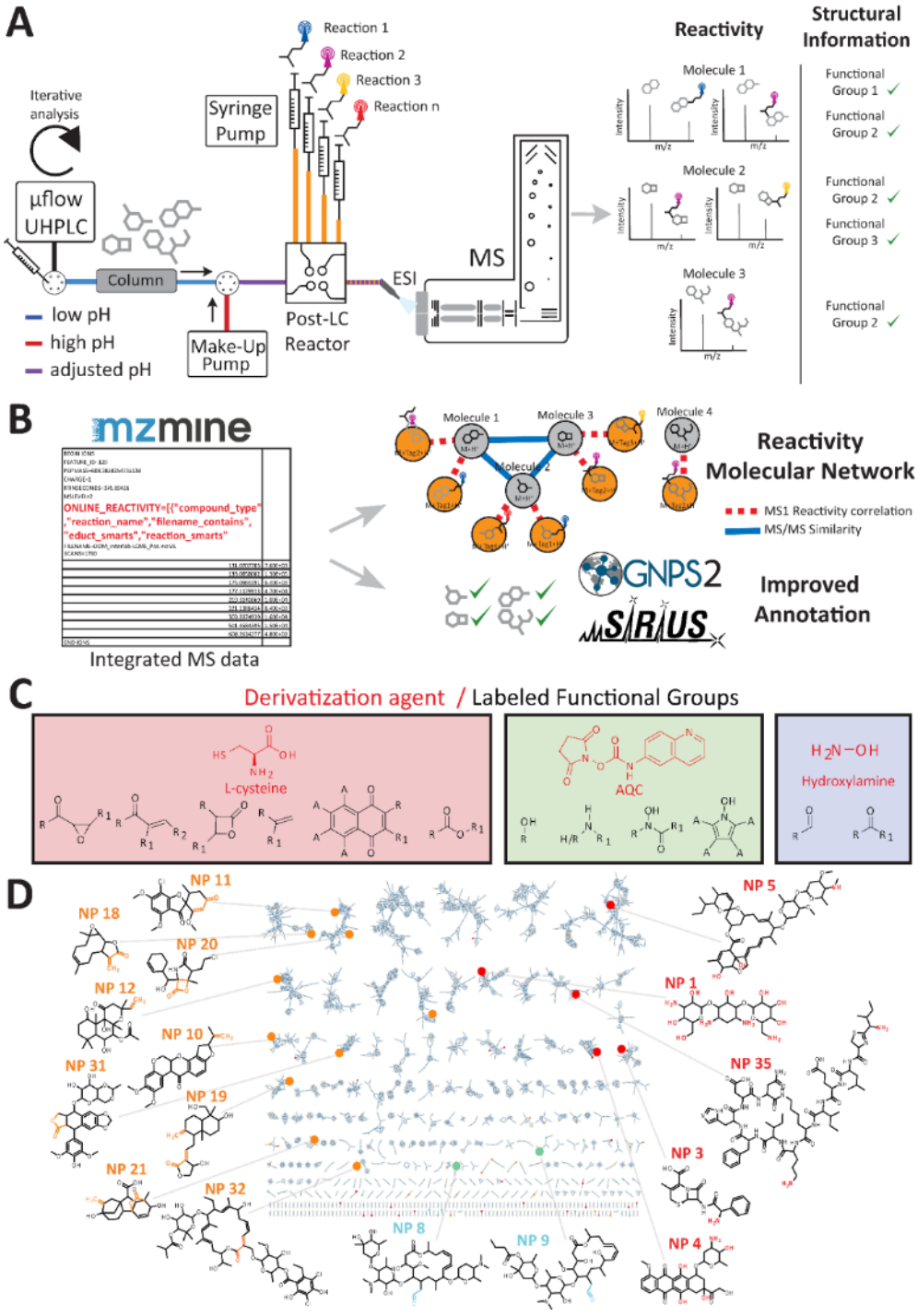


Figure 31. Multiplexed Chemical Metabolomics (MCheM) overview. **A)** The sample is injected and individual components are separated via the LC system. The derivatization agent is continuously infused post-column over the entire run and, when necessary (Reaction B), the effluent pH is adjusted with a buffer infused between the column and the derivatization agent injection port. The effluent is continuously analyzed via MS. **B)** The “Online Reactivity” module in mzmine automatically connects precursor/product pairs, integrating structural knowledge into the MS data, and allowing a reactivity

molecular network to be visualized. The integrated data are submitted to SIRIUS/CSI: FingerID and GNPS2 to improve annotation confidence. **C)** Reactions employed in the present study along with the reacting functional groups. **D)** Co-clustered reactivity network of a subset of the 359 NPs here tested, highlighting nodes reacting with L-cysteine (orange), AQC (red), and N-hydroxylamine (light blue). Reacting functional groups are marked with the corresponding colors. Adapted and modified from Vitale *et al.*[194]

3.1.2 MCheM allows for improved metabolite annotation of tandem mass spectra

We assessed the selectivity of the MCheM reactions using a panel of authentic NP standards (**Figure 31D**). Reaction A labeled electrophilic moieties. Reaction A successfully labeled electrophilic functional groups, including Michael acceptors[185,195], naphthoquinones[196,197], epoxyketones[186,198,] β-lactones[186,199], and macrocyclic esters (likely undergoing thioester formation)[200], as well as terminal alkenes[195]. Reaction B effectively labeled primary and secondary amines, phenols, and N-hydroxy groups[201,202]. Lastly, reaction C labeled aldehydes and ketones[192,203]. A total of 139 distinct derivatization events (including those from Reactions A–C) were detected across the 359 compounds, using their known structures as ground truth. Only five of these events (3.6%) were assigned as false positives, confirming the high specificity of the MCheM workflow.

We validated the performance of MCheM-enhanced annotation using CSI:FingerID. Initially, 208 spectra from the standard mixture that reacted with at least one derivatization reagent were searched against the SIRIUS biological structure database. For 180 of these spectra, the correct structures were represented in the database. Incorporation of MCheM information improved rankings across all Top-k levels, with 88 spectra (49%) showing a better overall placement. Of these, 20% moved into the top three candidates and 6% were reassigned to the top-ranked position. To estimate the improvement in metabolite annotation on a larger and more diverse set of molecules, we simulated MCheM-derived functional group information from 10,709 MS/MS known spectra from MassBank, MoNA and GNPS (CANOPUS dataset[204]), in which we added the SMARTS string to each spectra as ground truth. The associated molecular structures were not included in CSI: FingerID training. Also here, MCheM substantially improved the annotation rankings for 3297 spectra (32%), with 22% showing improved top 3 and 15% improved top 1 annotations. Next, we evaluated the impact of MCheM for open modification search (e.g., in case the target compound is not present in the library, but a structurally highly similar one is present). To do this, we removed the exact matching structures from the GNPS2 MS/MS libraries. Out of the experimental spectra from our 359 authentic standards, 189 yielded at least one hit during open modification spectrum library matching. Using the Tanimoto similarity[205] of molecular fingerprints[206] to evaluate structural similarity, the analysis focused on the queries with a potentially very similar structural analog in the libraries, higher than or equal to 0.5 Tanimoto similarity. This filtration step left 125 MS/MS queries. Comparison of rankings before and after MCheM-informed filtering revealed improvement in the average Tanimoto scores and the rank of the most structurally similar match. Of the 125 cases, the top 1 Tanimoto score improved in 27 (21.6%) and decreased in 15 (12%), raising the average Tanimoto score from 0.36 to 0.44. Among the top 5 hits, the best Tanimoto scores improved in 28 cases (22.4%) and decreased in 11 (8.8%), raising the average from 0.48 to

0.58. The rank of the highest Tanimoto score improved in 47 cases (37.9%) and declined in 19 (15.3%), with the average rank improving from 14.92 to 9.64 (Supplementary Figure 16). For the larger CANOPUS dataset, 7,248 of the 10,709 public MS/MS spectra yielded analog library matches. Among these, 861 spectra (11.9%) showed improved top 1 matches, while 562 scans (7.8%) decreased, increasing the average Tanimoto score from 0.44 to 0.52 (an 18% improvement). For the top 5,887 spectra (12.2%) improved, 446 (6.2%) decreased in ranking, and the average increased from 0.61 to 0.67 (Figure 3B and Supplementary Figures 14-15). The rank of the most similar matches improved in 2,194 cases (30.3%) and worsened in 669 (9.2%), with the average rank improving from 11.94 to 9.42 (Supplementary Figure 16).

3.1.3 MCheM facilitates the discovery of 7-glycosyl oxazolomycin D

To assess the effectiveness of MCheM for investigating uncharacterized bacterial extracts, we applied the workflow in a genome-guided NPs discovery study focused on specialized metabolites produced by *Streptomyces libani subsp. rufus* DSM 41230. Genome analysis revealed a BGC in *S. libani* genome features is similar to oxazolomycin B BGC reported from Streptomyces albus JA345339[207]. Thus, we used MCheM reaction A to target this functionality, since members of this metabolite family contain a reactive β-lactone group. Notably, conventional MS/MS annotations did not rank oxazolomycin among the leading candidates. In contrast, when CSI: FingerID results were reranked using MCheM data, oxazolomycin D emerged as the top annotation for the feature at m/z 700.3804, making it the strongest structural match. We then extended the analysis to include related, previously unassigned metabolites detected by molecular networking and flagged by MCheM reaction A using cysteine. In addition to oxazolomycin D, four abundant candidate analogs within the network also reacted with the cysteine probe. Their mass shifts and MS/MS fragmentation signatures suggested oxidation and cyclization events along the polyketide backbone, in line with structural changes previously described for oxazolomycin F[208]. A metabolite with m/z 862.434 displayed high spectral similarity to oxazolomycin D but differed by a Δm/z of +162.0528 Da. This delta mass corresponds to the expected mass of a hexose moiety ($C_6H_{10}O_5$), which is consistent with the presence of a putative glycosyltransferase gene (srufu_025110) in the *S. libani* oxazolomycin BGC, and this gene is absent from the corresponding cluster in *S. albus*. Notably, glycosylated oxazolomycins have not been described in the literature and are not present in natural product databases (GNPS, Dictionary of Natural Products) and other structure databases (e.g., PubChem). Importantly, glycosylated oxazolomycins are not represented in the structure database searched by CSI: FingerID. To verify the proposed structure, we isolated and purified the compound by flash chromatography followed by preparative HPLC and characterized it using orthogonal NMR experiments. These data established the metabolite as 7-glycosyl oxazolomycin D, identifying it as the first known glycosylated member of the oxazolomycin family.

3.2 Materials and methods

For full details of the materials and methods, see section 3.3.

3.3 Publication

nature communications

Article

https://doi.org/10.1038/s41467-025-61240-z

Enhancing tandem mass spectrometry-based metabolite annotation with online chemical labeling

Received: 13 January 2025

Accepted: 16 June 2025

Published online: 26 July 2025

Check for updates

Giovanni Andrea Vitale[1,12], Shu-Ning Xia[1,12], Kai Dührkop[2,3], Mohammad Reza Zare Shahneh[4], Heike Brötz-Oesterhelt[1,5,6], Yvonne Mast[7], Corinna Brungs[8,9], Sebastian Böcker[2], Robin Schmid[8,10], Mingxun Wang[4], Chambers C. Hughes[1,5,6] ✉ & Daniel Petras[6,11] ✉

Metabolite identification in non-targeted mass spectrometry-based metabolomics remains a major challenge due to limited spectral library coverage and difficulties in predicting metabolite fragmentation patterns. Here, we introduce Multiplexed Chemical Metabolomics (MCheM), which employs orthogonal post-column derivatization reactions integrated into a unified mass spectrometry data framework. MCheM generates orthogonal structural information that substantially improves metabolite annotation through in silico spectrum matching and open-modification searches, offering a powerful new toolbox for the structure elucidation of unknown metabolites at scale.

Despite rapid technical advances in high-resolution liquid-chromatography tandem mass spectrometry (LC-MS/MS) hardware and computational methods, metabolite annotation remains a challenging aspect of metabolomic research. On average, less than 10% of features are confidently annotated during MS analysis[1]. While previous studies have suggested that many unannotated features can be attributed to various ion adducts and in-source fragments[2], there is overwhelming evidence estimated by metabolomics[3–6], genome mining[7–9], and natural product discovery data[10–12] that the chemical space of metabolites is vast and mostly uncharted. A critical bottleneck in metabolite identification is the limited coverage of spectral libraries relative to the immense diversity of chemical space[13]. Recent efforts to address this bottleneck include exploiting structural databases, which are substantially larger than spectral libraries, through in silico annotation methods[14], as well as de novo structure annotation methods[15]. Even with advances in in silico annotation, high-confidence annotation of metabolites is still elusive, motivating the development of workflows to acquire orthogonal chemical data that aid annotation tools and increase accuracy.

Chemical derivatization has been broadly applied to improve structure elucidation and overcome limits of detection by exploiting the unique reactivity of certain chemical moieties[16]. In the field of targeted metabolomics, post-column derivatization is used to enhance the detection of specific compounds[17]. In contrast, derivatization strategies in non-targeted metabolomics have, for the most part, only been successfully implemented as batch processes prior to LC-MS/MS analysis[18–20]. Here, the use of post-column derivatization has enormous untapped potential, as it can be used to provide ion feature-resolved functional group information of unknown molecules. Post-column approaches provide an inherent advantage compared to batch derivatization, as chromatographic co-elution profiles maintain the link between precursors and their derivatization products, which is

[1]Department of Microbial Bioactive Compounds, Interfaculty Institute of Microbiology and Infection Medicine (IMIT), University of Tübingen, Tübingen, Germany. [2]Chair for Bioinformatics, Institute for Computer Science, Friedrich Schiller University Jena, Jena, Germany. [3]Bright Giant GmbH, Jena, Germany. [4]Department of Computer Science, University of California Riverside, Riverside, CA, USA. [5]German Center for Infection Research (DZIF), Partner Site Tübingen, Tübingen, Germany. [6]Cluster of Excellence EXC 2124: Controlling Microbes to Fight Infection, University of Tübingen, Tübingen, Germany. [7]Department Bioresources for Bioeconomy and Health Research, Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany. [8]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic. [9]Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Vienna, Austria. [10]mzio GmbH, Bremen, Germany. [11]Department of Biochemistry, University of California Riverside, Riverside, CA, USA. [12]These authors contributed equally: Giovanni Andrea Vitale, Shu-Ning Xia. ✉e-mail: chambers.hughes@uni-tuebingen.de; functionalmetabolomics@gmail.com

particularly important for complex mixtures. The gain of structural information in non-targeted chemical metabolomics experiments should increase drastically when orthogonal reactivities are considered. In this case, functional groups can be probed in iterative or parallel reactions, and further multiplexed in one data stream for in silico spectrum annotation and open-modification searches.

Here, we report the development of a Multiplexed Chemical Metabolomics workflow (MCheM) that leverages an array of post-column derivatization reactions for non-targeted LC-MS/MS analysis. To complement the new hardware setup (Fig. 1a), we developed an integrated data analysis pipeline that utilizes the computational concept of ion identity networking in mzmine[21] to post-process MCheM data for downstream metabolite annotation[22]. By integrating CSI:FingerID in silico annotation[23] and GNPS2 open modification search[24] (Fig. 1b), we observed annotation improvements of 31.9% for CSI:FingerID and 37.6% for GNPS2 over an experimental library of ~10 K experimental compounds, and of 48.8% for CSI:FingerID and 20.4% for GNPS2 on a set of authentic natural product (NP) standards. Finally, we demonstrated the utility of MCheM in a genome-guided natural product discovery case study, rapidly identifying novel oxazolomycin derivatives produced by *Streptomyces libani* subsp. *rufus* DSM 41230.

Results and discussion

MCheM hardware and software implementation

To implement MCheM data generation, we designed a custom LC-MS/MS hardware configuration consisting of a make-up UHPLC pump, a T-splitter or reactor manifold, and a syringe pump (Fig. 1a). Using this setup, we implemented three LC-MS-compatible post-column derivatization reactions, each targeting distinct functional groups (Fig. 2a and Supplementary Fig. 1). First, we applied an established derivatization reaction using *L*-cysteine to target electrophiles[18] (Reaction A). Electrophilic groups are fairly common in NPs, as electrophilic NPs often evolved to interact covalently with nucleophilic amino acid residues (e.g., serine, cysteine, threonine) present in biological targets. In our second reaction, we targeted amino and phenol groups using commercially available 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AQC) (Reaction B). As this reaction requires a basic pH to constrain the amino groups to their free-base form, we infused a 0.5% trimethylamine buffer

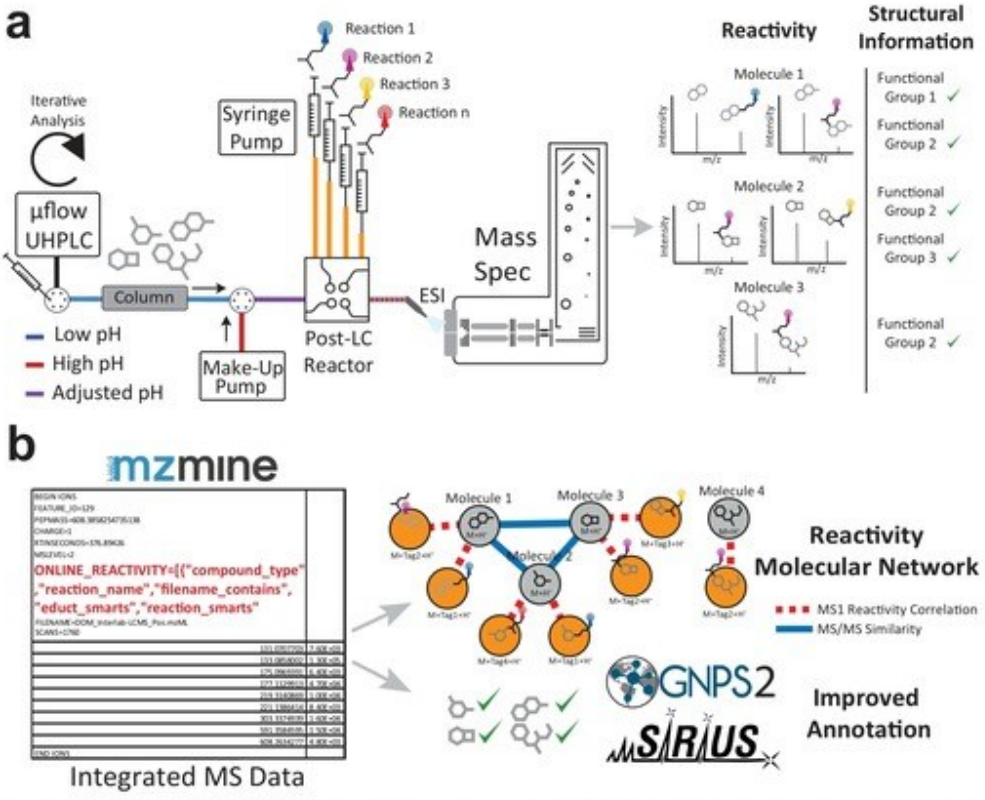


Fig. 1 | Overview of the Multiplexed Chemical Metabolomics (MCheM) workflow. a Samples are separated by liquid chromatography (LC), and derivatization agents are continuously infused post-column throughout the run. When necessary (e.g., Reaction B), the effluent pH is adjusted with a buffer infused between the column and the derivatization agent injection port. The effluent is continuously analyzed via MS. **b** The "Online Reactivity" module in mzmine automatically connects precursor/product pairs, integrating structural knowledge into the MS data, and allowing a reactivity molecular network to be visualized. The integrated data are submitted to SIRIUS/CSI:FingerID and GNPS2 to improve annotation confidence.

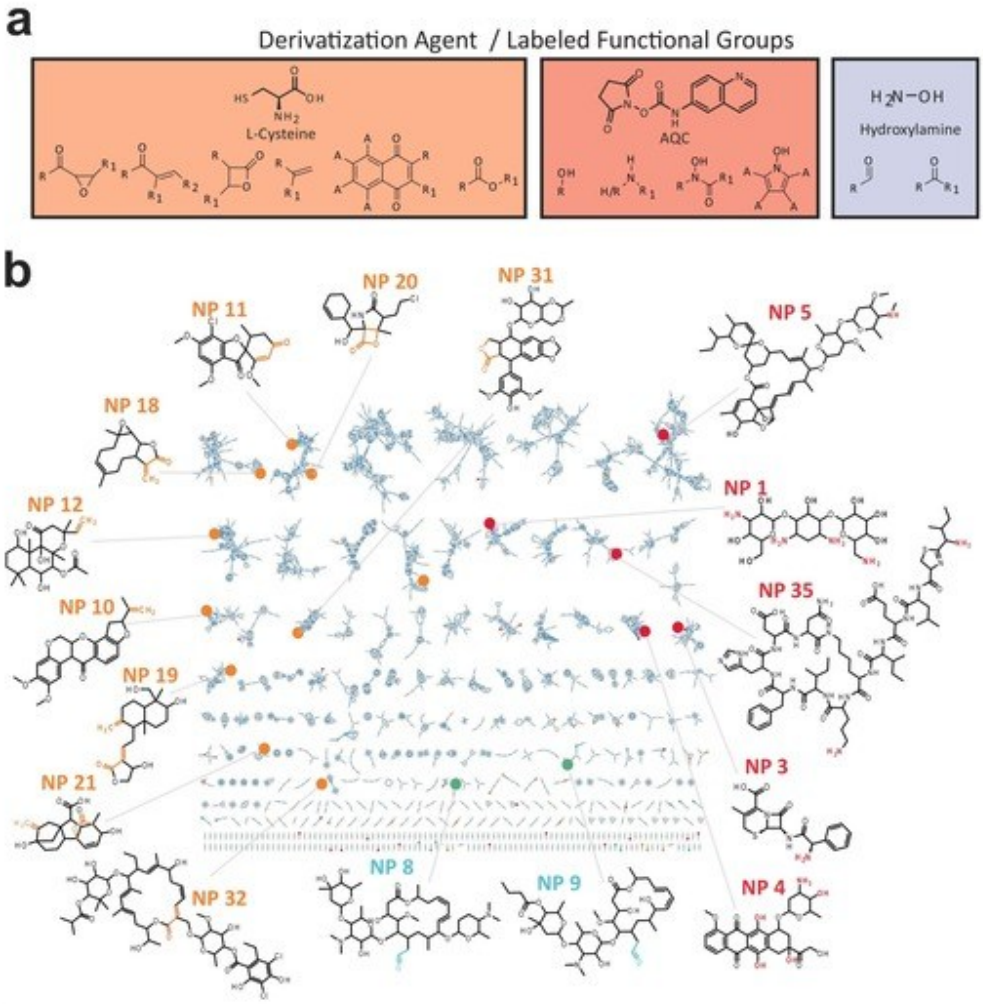


Fig. 2 | Overview of the Multiplexed Chemical Metabolomics (MCheM) derivatization chemistry and validation with a set of authentic natural product standards. a Reactions employed in the present study along with the reacting functional groups. **b** Co-clustered reactivity network of a subset of 32 representative natural products here tested, highlighting nodes reacting with cysteine (orange), AQC (red), and *N*-hydroxylamine (light blue). Reacting functional groups are marked with the corresponding colors. Source data is provided through the MassIVE and Zenodo repository (see data availability section and Supplementary Table 2).

between the column and AQC stream using the make-up pump, raising the effluent pH to the 5–6 range during the entire MS analysis. In our third reaction, we targeted aldehydes and ketones using commercially available hydroxylamine hydrochloride (Reaction C)[25]. Following initial test reactions using the three chemical labeling reactions with our post-column reactor setup, we determined concentration-dependent linearity and limits of detection (Supplementary Figs. 2–4). Subsequently, we experimentally validated all three reactions using 359 structurally diverse natural product standards from the Tübingen Natural Compound Collection (https://external.gnps2.org/gnpslibrary) (Supplementary Data 1).

To enable high-throughput analysis of all derivatization reactions in complex metabolomics samples and to merge the resulting data into a single multiplexed data stream, we developed a specialized "Online Reactivity" analysis module in mzmine. This computational strategy leverages the co-elution of precursors and products to establish correlation-based connections (Fig. 1b), using ion identity networking[22] in combination with user-defined $\Delta m/z$ values corresponding to each derivatization reagent. The resulting MCheM data output represents a hybrid dataset that integrates MS, MS/MS, and reactivity-based information. It also includes a list of predicted functional groups or substructures in the form of SMILES Arbitrary Target

Specification (SMARTS)[26] (Supplementary Figs. 9–11). This reactivity-resolved information can be directly used by downstream computational tools such as CSI:FingerID[24] or GNPS2[24] to constrain the molecular structure search space (Fig. 1b).

MCheM allows for improved metabolite annotation of tandem mass spectra

We validated the specificity of MCheM reactions using a set of authentic natural product standards, as shown in Fig. 2b. Reaction A successfully labeled electrophilic functional groups, including Michael acceptors[17,27], naphthoquinones[28,29], epoxyketones[18,30], β-lactones[19,31], and macrocyclic esters (likely undergoing thioester formation)[32], as well as terminal alkenes[27]. Reaction B effectively labeled primary and secondary amines, phenols, and *N*-hydroxy groups[33,34]. Lastly, reaction C labeled aldehydes and ketones[25,35]. A total of 139 distinct derivatization events (including those from Reactions A–C) were detected across the 359 compounds, using their known structures as ground truth. Of these, only five instances (3.6%) were classified as false positives, confirming the high specificity of the MCheM workflow.

We validated the performance of MCheM-enhanced annotation using CSI:FingerID. First, we analyzed 208 spectra from our standard mix that reacted with at least one derivatization reagent, querying them against the SIRIUS biological structure database. For 180 spectra the correct structures were present in the database. The ranking results for every Top k annotation improved due to MCheM, with 88 spectra having their overall rankings improved (49%). Notably, 20% of these spectra were promoted into the top 3, and 6% were reranked to the top 1 position (Fig. 3a). To estimate the improvement in metabolite annotation on a larger and more diverse set of molecules, we assessed the performance of MCheM by simulating the gained functional group information from 10,709 MS/MS known spectra from MassBank, MoNA and GNPS (CANOPUS dataset[36]), in which we added the SMARTS string to each spectra as ground truth. The corresponding molecular structures were not part of the CSI:FingerID training data. Also here, MCheM substantially improved the annotation rankings for 3297 spectra (32%), with 22% showing improved top 3 and 15% improved top 1 annotations (Fig. 3a). Next, we evaluated the impact of MCheM for open modification search (e.g., in case the target compound is not present in the library, but a structurally highly similar one is present). To do this, we removed the exact matching structures from the GNPS2 MS/MS libraries. Out of the experimental spectra from our 359 authentic standards, 189 yielded at least one hit during open modification spectrum library matching. Using the Tanimoto similarity[37] of molecular fingerprints[38] to evaluate structural similarity, the analysis focused on the queries with a potentially very similar structural analog in the libraries, higher than or equal to 0.5 Tanimoto similarity. This filtration step left 125 MS/MS queries. Comparison of rankings before and after MCheM-informed filtering revealed improvement in the average Tanimoto scores and the rank of the most structurally similar match, as shown in Fig. 3b and Supplementary Figs. 12 and 13. Of the 125 cases, the top 1 Tanimoto score improved in 27 (21.6%) and decreased in 15 (12%), raising the average Tanimoto score from 0.36 to 0.44. Among the top 5 hits, the best Tanimoto scores improved in 28 cases (22.4%) and decreased in 11 (8.8%), raising the average from 0.48 to 0.58. The rank of the highest Tanimoto score improved in 47 cases (37.9%) and declined in 19 (15.3%), with the average rank improving from 14.92 to 9.64 (Supplementary Fig. 16). For the larger CANOPUS dataset, 7248 of the 10,709 public MS/MS spectra yielded analog library matches. Among these, 861 spectra

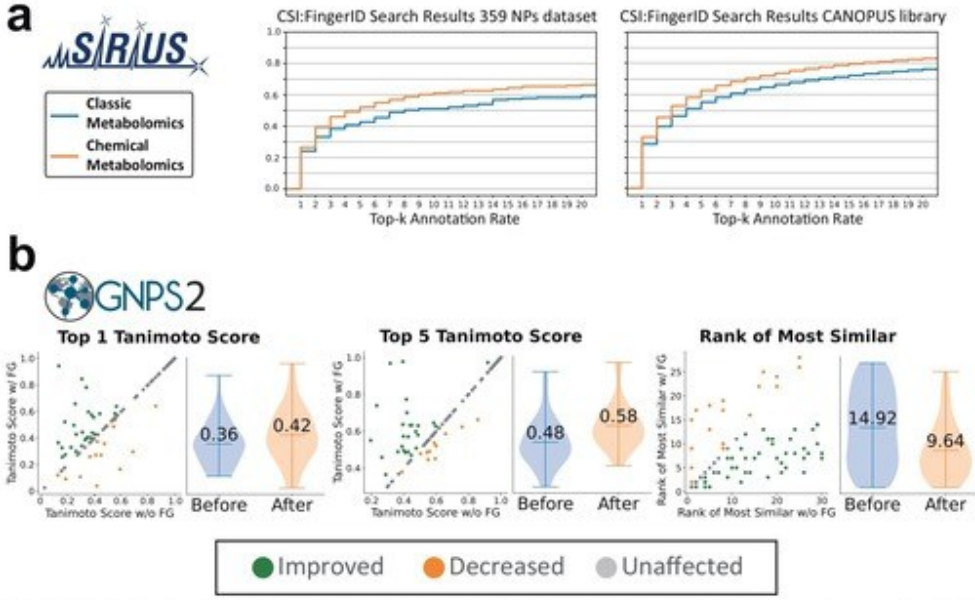


Fig. 3 | SIRIUS AND GNPS2 results of the Multiplexed Chemical Metabolomics (MCheM) workflow. a Impact of MCheM on CSI:FingerID search. The annotation rate obtained using only standard MS data (blue) increases when MCheM-enhanced MS data (orange) are used. The improvement is observed for both the experimental dataset and in silico on CANOPUS dataset for every Top k annotation. **b** Impact of MCheM on analog library search. MCheM improved average analogs similarity to the real structure in terms of Top 1 and Top 5 Tanimoto scores, elevating the average Tanimoto score in both cases and for both datasets. Datapoints from the scatter plots are visualized in the violin plots. Error bars indicate the spread of data, and the center line indicates the mean. Source data is provided in the supplemental information.

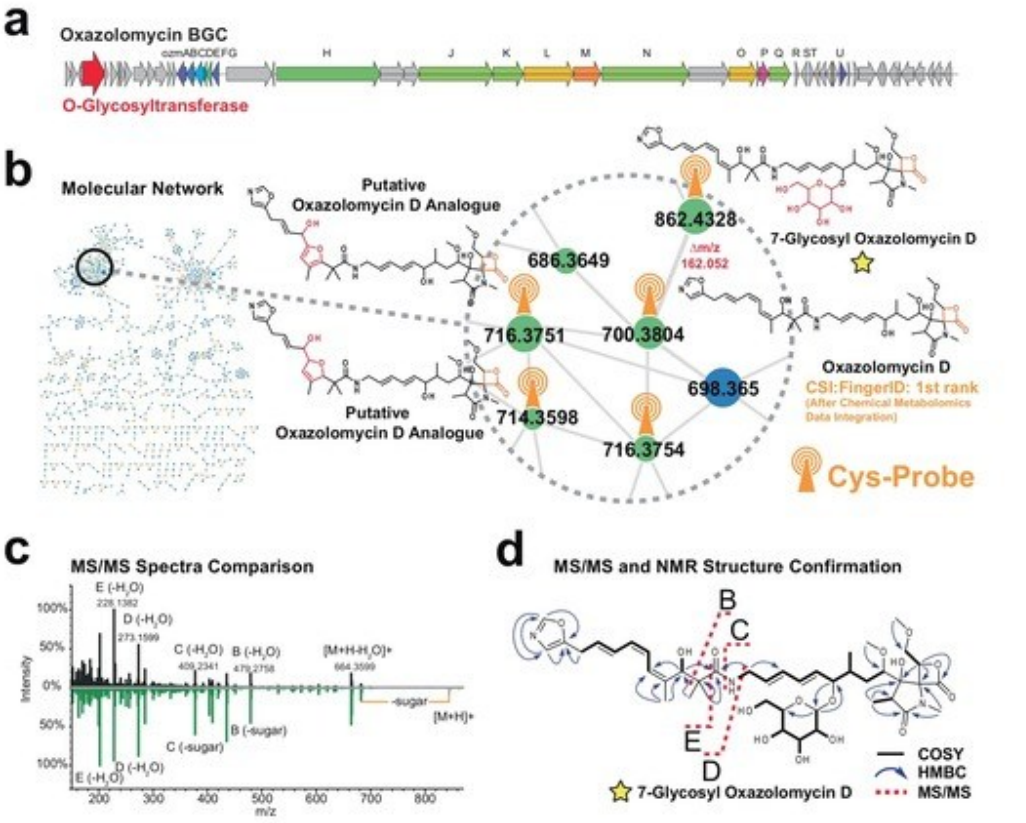


Fig. 4 | Application of Multiplexed Chemical Metabolomics (MCheM) for the discovery and structure elucidation of 7-glycosyl oxazolomycin D. **a** Oxazolomycin biosynthetic gene cluster BGC0001106 from *S. libani* subsp. *rufus* DSM 41230. **b** MCheM reactivity network of *S. libani* extract, with a zoomed view of oxazolomycin D cluster. **c** Mirror plot of MS/MS spectra of oxazolomycin D and 7-glycosyl oxazolomycin D. **d** NMR and MS/MS structure confirmation of 7-glycosyl oxazolomycin. Source data is provided through the nmrXiv, MassIVE, and Zenodo repositories (see data availability section and Supplementary Table 2).

(11.9%) showed improved top 1 matches, while 562 scans (7.8%) decreased, increasing the average Tanimoto score from 0.44 to 0.52 (an 18% improvement). For the top 5 matches, 887 spectra (12.2%) improved, 446 (6.2%) decreased in ranking, and the average increased from 0.61 to 0.67 (Fig. 3b and Supplementary Figs. 14 and 15). The rank of the most similar matches improved in 2194 cases (30.3%) and worsened in 669 (9.2%), with the average rank improving from 11.94 to 9.42 (Supplementary Fig. 16).

MCheM facilitates the discovery of 7-glycosyl oxazolomycin D

Finally, we evaluated the effectiveness of MCheM to explore uncharacterized bacterial extracts. We applied the approach in a genome-guided natural product discovery effort to investigate specialized metabolites produced by *Streptomyces libani* subsp. *rufus* DSM 41230. The *S. libani* genome features a biosynthetic gene cluster (BGC) similar to oxazolomycin B BGC from *Streptomyces albus* JA3453[39] (Fig. 4a). This natural product family features a reactive β-lactone moiety. In order to detect the β-lactone moiety, we leveraged MCheM reaction A. Strikingly, while oxazolomycin was not annotated as one of the top hits using regular MS/MS data, the MCheM-based reranking of CSI:FingerID search results pushed oxazolomycin D to the Top 1 annotation for node *m/z* 700.3804, rendering it the most likely structure (Fig. 4b).

Next, we extended our analysis to explore structurally related, previously uncharacterized derivatives identified through molecular networking and tagged via MCheM reaction A (cysteine). Among the most abundant features within our molecular network were four putative analogs in addition to oxazolomycin D that showed reactivity with the cysteine probe (Fig. 4b). Their mass differences and MS/MS fragmentation patterns indicated possible oxidation and cyclization events in the polyketide chain, consistent with modifications previously reported for oxazolomycin F (Supplementary Figs. 19 and 20)[40]. Another feature with *m/z* 862.434 showed high spectral similarity to oxazolomycin D with a *Δm/z* of +162.0528 Da (Fig. 4c). This delta mass corresponds to the expected mass of a hexose moiety ($C_6H_{10}O_5$), consistent with the presence of a putative glycosyltransferase gene (*srufu_025110*) in the *S. libani* oxazolomycin BGC, which is absent in the *S. albus* counterpart (Supplementary Fig. 17). Notably, glycosylated oxazolomycins have not been described in the literature and are not present in natural product databases (GNPS, Dictionary of Natural Products) and other structure databases (e.g., PubChem). Importantly, glycosylated oxazolomycins are not

represented in the structure database searched by CSI:FingerID. To confirm the structure, we purified the compound using flash chromatography and preparative HPLC, followed by a suite of orthogonal NMR experiments (Fig. 4d, Supplementary Figs. 22–27 and Supplementary Table 1). These analyses unambiguously identified the compound as 7-glycosyl oxazolomycin D, representing the first glycosylated member of the oxazolomycin family.

In summary, MCheM combines novel concepts for MS/MS data acquisition and analysis to improve metabolite annotation in non-targeted metabolomics. The workflow includes a scalable software solution in mzmine to detect MCheM-derived features, which can be integrated with downstream MS/MS annotation tools such as SIRIUS and GNPS2 for enhanced metabolite identification. Our results demonstrate that incorporating MCheM-derived reactivity information substantially improves metabolite annotation accuracy, in both in silico and open modification searches, as demonstrated in the natural product discovery of a novel glycosylated oxazolomycin.

A central aspect that underscores MCheM practicality is its simple hardware setup, which can be easily implemented on most commercial LC-MS/MS platforms using a secondary HPLC, a syringe pump, and a readily-available software pipeline. Notably, the improvements reported here were achieved using structural information from only three derivatization reactions. Additional reactions targeting other functional groups can similarly be integrated to further boost annotation. We anticipate that MCheM will be broadly applicable for expanding structural information in non-targeted metabolomics, including the integration of emerging computational approaches with the ultimate goal of providing complete MS/MS-based de novo structure elucidation.

Methods

Micro-flow UHPLC-MS/MS

For reactions A and C, LC-MS/MS analysis was performed on a 1290 Infinity II ultrahigh-performance liquid chromatography (UHPLC) system coupled to a Bruker Impact II QTOF mass analyzer. The instrument was operated in data-dependent analysis (DDA) mode with the following parameters: positive ion mode, capillary voltage = 4.5 kV, nebulizer gas pressure = 2.2 bar, dry gas flow rate = 10 L/min, drying temperature = 220 °C. The default mass range was 150–2000 *m/z* for reaction A and 200–2000 *m/z* for reaction C, both with a resolution of 50,000. The three most abundant precursor ions were selected for fragmentation with a collision energy (CE) of 20–50 eV. After three MS/MS spectra were acquired on a particular precursor ion, the ion was dynamically excluded from the fragmentation list for 0.3 min. For reaction B, LC-MS/MS analysis was performed on a Vanquish UHPLC equipped with an additional quaternary pump coupled to a Q Exactive HF mass spectrometer. The heated electrospray ionization (HESI) source was used with the auxiliary gas temperature = 400 °C, flow = 12 arbitrary units (AU), sweep gas flow = 1 L/min, sheath gas flow rate = 50 AU. The instrument was operated in data-dependent analysis (DDA) mode with the following parameters: positive ion mode, capillary voltage = 3.5 kV, drying temperature = 250 °C. The default mass range was 220–2000 *m/z* with a resolution of 120,000. The five most abundant precursor ions were selected for fragmentation with a stepped normalized collision energy (NCE) of 25, 35, and 45 eV, a resolution of 15,000, and an isolation window of 1 *m/z*. The same EVO C-18 column (1.7 µm, 100 Å, 100 ×1 mm) and chromatographic method were employed for all the analyses. The mobile phases consisted of A (H_2O + 0.1% formic acid) and B (ACN + 0.1% formic acid). The constant 150 µL/min flow rate was used following a linear gradient starting with 5% B and reaching 99% B in 10 min, followed by a washing segment between 10 and 12 min (99% B) and a subsequent re-equilibration segment between 13 and 15 min (5% B).

Chemical metabolomics reactions setup

The derivatization agent (DA) was constantly infused post-column for the whole analysis duration through a syringe pump and a PEEK T-splitter or manifold. Reaction A: L-cysteine, 0.1 mM (10 µL/min), mass range 150–2000 *m/z*. Reaction B: AQC, 100 mM (2 µL/min), mass range 220–1500 *m/z*, with 0.150 mL/min of a 0.5% trimethylamine solution in 1:1 H_2O:ACN infused from a second PEEKT-splitter located before the AQC splitter. Reaction C: hydroxylamine hydrochloride, 100 mM (10 µL/min), mass range 200–2000 *m/z*.

Standard mix

For chemical metabolomics method development, a 100 µg/mL stock mixture containing the following 32 standards in H_2O:ACN 1:1 was used: kanamycin (Sigma Aldrich), cephalexin hydrate (Alfa Aesar), doxorubicin hydrochloride (LC laboratories), emamectin (MedChemExpress), vancomycin hydrochloride (Sigma Aldrich), spiramycin (Tokyo Chemical Industry Co.), kitasamycin (AK Scientific), rotenone (MP Biomedicals), griseofulvin (Alfa Aesar), forskolin (LC laboratories), carfilzomib (LC laboratories), tetracycline hydrochloride (Sigma Aldrich), genistein (LC laboratories), erythromycin (Sigma Aldrich), parthenolide (EMD Milipore), andrographolide (Indofine Chemical Company), gibberellic acid (Acros Organics), monensin sodium salt (abcam), yohimbine hydrochloride (Sigma Aldrich), abamectin (abcamBiochemicals), paclitaxel (MedChemExpress), catharanthine (TSZ Chem), harmaline (Indofine Chemical Company), quinine (Sigma Aldrich), vinblastine sulfate (TSZ Chem), etoposide (MedChemExpress), fidaxomicin (ApexBio), mifepristone (Acros Organics), terbinafine hydrochloride (Acros Organics), bacitracin (Sigma Aldrich). Marinopyrrole was obtained via chemical synthesis[41]. Salinosporamide was purified from cultures of *Salinispora tropica*[42].

Titration experiments

The 32-mix stock (100 µg/mL) was diluted with H_2O:ACN 1:1 as a solvent to prepare a 50, 10, 1, and 0.1 µg/mL dilution series. For method B, the 32-mix stock solution was supplemented with L-phenylalanine as internal standard at the same concentration. For the analysis, 5 µL of each standard mix was injected and analyzed in duplicate with and without the chemical metabolomics procedure for each method. The reaction yields were calculated [yield = product area/(product area + educt area) × 100%] for each compound and each reaction and plotted against compound concentrations (Supplementary Figs. 2–4).

Chemical metabolomics analysis of an in-house natural products library. A total of 327 pure NPs (from the Tü NP Library) that we recently shared with the metabolomics community via GNPS were mixed into 17 different pools (Supplementary Data 1). For the analysis of each pool, 5 µL were injected and analyzed in duplicates with and without the chemical metabolomics procedure for each experiment, each sample was run in a randomized order. In addition, method A was also evaluated using an extract containing cystargolide A to specifically test the efficiency of β-lactone labeling[43,44]. The extract was obtained by extracting a 6-day old culture of *Kitasatospora cystarginea* NRRL B-16505 grown in KCM liquid medium (1.6 g dextrin, 0.8 g galactose, 0.8 g maltose, 0.8 g Bacto soytone, 0.4 g glucose, 0.3 g $(NH_4)_2SO_4$ dissolved in 1 L tap water) with ethyl acetate.

Data processing. Thermo raw data were converted into.mzML format using msConvert by Proteowizard[45], while Bruker raw data were converted into the same format through a script developed by the manufacturer. All the MS data were processed with mzmine (Version 4.1.0) to "clean up" and align the data, and reactive metabolites were automatically annotated through the "Online Reactivity" module. Here different theoretical Δ*m/z* values were set according to the expected products for each derivatization reaction: 121.0197 Da for reaction A, 170.0481 Da for reaction B, and 15.0109 Da for reaction C (Supplementary Figs. 5–7 and Supplementary Data 2). All metabolites were annotated, and all possible reactive substructures present across the metabolites were incorporated into a table in SMARTS format and

automatically integrated into the.mgf files for SIRIUS and GNPS (Supplementary Data 2). The correct SMARTS representation was tested using a script written in Python. Reacting metabolites but lacking any possible reactive substructure were listed as "false positives". Metabolites with a low conversion yield (<5%) in reaction A were excluded to refine the data. The parameters employed for processing each dataset and the output data are publicly accessible as mzmine batch files (Supplementary Table 2). The data were exported with the FBMN module as one file containing the MS and MS/MS features (.mgf), one quantification table (.csv), and an edges annotation table (.csv), and through the SIRIUS exporting module as one MS file (.mgf).

CANOPUS dataset and CSI:FingerID evaluation. The *CANOPUS dataset* consists of 8,569 tandem mass spectra from GNPS[24], 1417 tandem mass spectra from MoNA (https://mona.fiehnlab.ucdavis.edu), and 723 tandem mass spectra from MassBank[46], containing a total of 8553 unique structures[36]. For this dataset, results of the four MCheM reactions were determined using the known compound structures, assuming error-free chemistry. Annotations were performed using the latest version of CSI:FingerID that is part of SIRIUS 6, using a *worst-case* five-fold cross validation: the fingerprint for each spectrum was predicted using a machine learning model that has seen neither any spectrum of the same structure, nor any spectrum of a similar or derivatized structure. Chemically similarity was determined using the myopic Maximum Common Edge Subgraph distance[47].

GNPS2 data analysis. An analog library search[24] was conducted to retrieve the top 30 analogs for each reading. The search parameters were adjusted to be more relaxed than the GNPS default (minimum cosine = 0.5, minimum matched peaks = 4, fragment ion tolerance = 0.5, and precursor ion tolerance = 2.0) (https://gnps2.org/homepage). This adjustment leverages additional FG information to capture more structurally similar matches that might exhibit lower cosine values. For the analysis section of this work, duplicate hits and exact matches were removed from the library search results, and the search was expanded to maintain a complete set of analogs. This step ensures that the results consist only of analogs and not exact library matches, ensuring fair comparisons and eliminating potential information overlap. It should be noted that the remaining structures in the cleaned libraries may still show a Tanimoto similarity of 1 with some data structures, as the molecular fingerprint representation is lossy, meaning two structurally distinct molecules can produce the same fingerprint. The fingerprints were computed using the default settings of the RDKit (https://www.rdkit.org) fingerprint implementation. Next, the Tanimoto similarity between the fingerprint of each reading's structure and the fingerprints of its matches were calculated as the baseline measure of success. The matches were then ranked based on their modified cosine similarity score, referred to as the "ranking without functional groups." Finally, the structures of the matches were examined to determine if they contain the corresponding functional groups. With this information, the "ranking with functional groups" was calculated, where matches are first ordered by the presence of the functional group and then by their modified cosine score.

Genome analysis of DSM 41230. The genome sequence of *S. libani* subsp. *rufus* DSM 41230 (taxonomic correct name *Streptomyces platensis* DSM 41230) was analyzed with antiSMASH v. 7.0[48] to identify BGCs. Region 1.19 contained a set of genes similar to the oxazolomycin B BGC from *S. albus* JA3453. The analysis of synteny among clusters was performed with clinker[49].

Isolation and structure elucidation of 7-glycosyl oxazolomycin D. Strain *S. libani* subsp. *rufus* DSM 41230 was first cultivated in a 100 mL Erlenmeyer flask with 30 mL of R5 liquid medium (100 g sucrose, 10 g glucose, 0.25 g K_2SO_4, 10 g $MgCl_2$, 0.1 g casamino acids, 5 g yeast extract, 5.7 g TES buffer, and 2 mL trace metal mix in 1 L distilled water, pH 7.2). The culture was grown on a platform shaker at 100 rpm and 29 °C for three days. 5 mL of preculture was used to inoculate 500 mL flasks containing 150 mL fermentation medium (4 g yeast extract, 10 g soluble starch, 2 g Bacto peptone and 35 g sea salts in 1 L distilled water, pH 7.0). The culture was grown at 100 rpm and 29 °C for four days. The complete culture (10.5 L) was then centrifuged, and the supernatant was extracted three times with an equal volume of ethyl acetate (EtOAc). The EtOAc extract was dried over $NaSO_4$ and concentrated under reduced pressure to give 6.7 g of crude extract. The extract was fractionated on silica gel using a stepwise gradient elution with the following solvents: 100% hexanes, 30% EtOAc in hexanes, 100% EtOAc, 1% MeOH in DCM, 10% MeOH in DCM, 100% MeOH. Fraction 5 (eluted with 10% MeOH in DCM) was dried and subjected to further silica gel chromatography using the following elution profile: 100% DCM, 10% MeOH in DCM, and a linear gradient from 10-100% MeOH in DCM over 20 min, yielding multiple subfractions. Fractions containing the target compound were combined, dried, and redissolved in 7% MeOH in DCM for preparative HPLC purification (Phenomenex Luna silica (2), 10 µm, 250 × 21.2 mm) using an isocratic method (7% MeOH in DCM, 10 mL/min) and UV detection at 254 nm to yield 7-glycosyl oxazolomycin D (53.4 mg, t_R = 12 min): HRESIMS *m/z* $[M+H]^+$ = 862.4330, calculated for $C_{43}H_{64}N_3O_{15}^+$, 862.4332. The compound was characterized 1D and 2D NMR spectroscopy, including ^1H, ^{13}C, COSY, HSQC, HMBC, NOESY experiments, recorded on a Bruker Avance III HDX 700 MHz spectrometer equipped with a 5 mm Prodigy (^1H,^{19}F/^{13}C/^{15}N) TCI Cryoprobe (Supplementary Table 1 and Supplementary Figs. 22–27). ^1H NMR data were recorded at 700 MHz in DMSO-d_6 (2.50 ppm), and ^{13}C NMR data were recorded at 175 MHz in DMSO-d_6 (39.5 ppm). NMR data were processed using MestReNova (Mnova 14.3.0, Mestrelab Research) software[50].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All MS data acquired in this work as well as processed data, and processing batch files, are publicly available through both MASSIVE and Zenodo [https://zenodo.org]. All accession numbers and links are listed in the supplemental information in Supplementary Table 2. The CANOPUS dataset used to test the method can be downloaded from [https://bio.informatik.uni-jena.de/data/]. *Streptomyces libani* subsp. *rufus* DSM 41230 (strain NBRC 15424) genome is publicly accessible through NCBI Nucelotide Database with accession number AP023408. The GNPS2 FBMN job is publicly available at the following link: [https://gnps2.org/status?task=714b3ae18f8d4c499876d486658b9b30]. GNPS2 Chemical Metabolomics job for the experimental dataset can be accessed here: [https://gnps2.org/status?task=508a9676780d4e119cb6ac3bce011512]. NMR data has been deposited to nmrXiv with the https://doi.org/10.57992/NMRXIV.P114. Source data are provided with this paper.

Code availability

The code for the GNPS workflow and analysis is available on GitHub [https://github.com/Wang-Bioinformatics-Lab/Chemical_Metabolomics_with_functional_groups/tree/master] and Zenodo [https://zenodo.org/records/15469092]. A fully runnable instance of the workflow can be launched directly at GNPS2.org [https://gnps2.org/workflowinput?workflowname=Chemical_Metabolomics_With_Functional_Groups]. The mzmine source code and versions are available on GitHub [https://github.com/mzmine/mzmine] and mzio.io [https://mzio.io]. MCheM filtering is available in SIRIUS version 6.2, and source code and versions are available on GitHub [https://github.com/sirius-ms/sirius] and bright-giant.com [https://bright-giant.com]. The code for filtering SIRIUS

candidate lists via MCheM can be found at GitHub [https://github.com/kaibioinfo/functional_metabolomics]. The code for the SMARTS check is uploaded to GitHub [https://github.com/corinnabrungs/smarts_testing] and Zenodo [https://doi.org/10.5281/zenodo.15480298].

References

1. Bittremieux, W., Wang, M. & Dorrestein, P. C. The critical role that spectral libraries play in capturing the metabolomics community knowledge. *Metabolomics* **18**, 94 (2022).
2. Mahieu, N. G. & Patti, G. J. Systems-level annotation of a metabolomics data set reduces 25 000 features to fewer than 1000 unique metabolites. *Anal. Chem.* **89**, 10397–10406 (2017).
3. Metz, T. O. et al. Decoding the molecular universe–workshop report. *ArXiv* https://arxiv.org/abs/2311.11437 (2023).
4. Caesar, L. K. et al. Correlative metabologenomics of 110 fungi reveals metabolite–gene cluster pairs. *Nat. Chem. Biol.* **19**, 846–854 (2023).
5. Morehouse, N. J. et al. Annotation of natural product compound families using molecular networking topology and structural similarity fingerprinting. *Nat. Commun.* **14**, 308 (2023).
6. Papadopoulos Lambidis, S. et al. Two-dimensional liquid chromatography tandem mass spectrometry untangles the deep metabolome of marine dissolved organic matter. *Environ. Sci. Technol.* **58**, 19289–19304 (2024).
7. MohammadiPeyhani, H., Hafner, J., Sveshnikova, A., Viterbo, V. & Hatzimanikatis, V. Expanding biochemical knowledge and illuminating metabolic dark matter with ATLASx. *Nat. Commun.* **13**, 1560 (2022).
8. Bağcı, C. et al. BGC Atlas: a web resource for exploring the global chemical diversity encoded in bacterial genomes. *Nucleic Acids Res.* gkae953, https://doi.org/10.1093/nar/gkae953 (2024).
9. Gavriilidou, A. et al. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **7**, 726–735 (2022).
10. Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H. & Linington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl Acad. Sci.* **114**, 5601–5606 (2017).
11. Avalon, N. E. et al. Leptochelins A–C, cytotoxic metallophores produced by geographically dispersed leptothoe strains of marine cyanobacteria. *J. Am. Chem. Soc.* **146**, 18626–18638 (2024).
12. Neuhaus, G. F. et al. Environmental metabolomics characterization of modern stromatolites and annotation of ibhayipeptolides. *PLOS ONE* **19**, e0303273 (2024).
13. Theodoridis, G. et al. Ensuring fact-based metabolite identification in liquid chromatography–mass spectrometry-based metabolomics. *Anal. Chem.* **95**, 3909–3916 (2023).
14. Dührkop, K. et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **16**, 299–302 (2019).
15. Stravs, M. A., Dührkop, K., Böcker, S. & Zamboni, N. MSNovelist: de novo structure generation from mass spectra. *Nat. Methods* **19**, 865–870 (2022).
16. Zhao, S. & Li, L. Chemical derivatization in LC-MS-based metabolomics study. *TrAC Trends Anal. Chem.* **131**, 115988 (2020).
17. Hughes, C. C. Chemical labeling strategies for small molecule natural product detection and isolation. *Nat. Prod. Rep.* **38**, 1684–1705 (2021).
18. Castro-Falcón, G., Hahn, D., Reimer, D. & Hughes, C. C. Thiol probes to detect electrophilic natural products based on their mechanism of action. *ACS Chem. Biol.* **11**, 2328–2336 (2016).
19. Kaur, A. et al. Chemoselective bicyclobutane-based mass spectrometric detection of biological thiols uncovers human and bacterial metabolites. *Chem. Sci.* **14**, 5291–5301 (2023).
20. Lin, W. et al. Rapid and bifunctional chemoselective metabolome analysis of liver patient samples using the reagent 4-nitrophenyl-2H-azirine. *Angew. Chem. Int. Ed.* **63**, e202318579 (2024).
21. Schmid, R. et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* **41**, 447–449 (2023).
22. Schmid, R. et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nat. Commun.* **12**, 3832 (2021).
23. Dührkop, K., Shen, H., Meusel, M., Rousu, J. & Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl Acad. Sci.* **112**, 12580–12585 (2015).
24. Wang, M. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
25. Chen, D. et al. A liquid chromatography–mass spectrometry method based on post column derivatization for automated analysis of urinary hexanal and heptanal. *J. Chromatogr. A* **1493**, 57–63 (2017).
26. Daylight Chemical Information Systems, Inc. http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (2023).
27. Hoyle, C. E. & Bowman, C. N. Thiol–Ene click chemistry. *Angew. Chem. Int. Ed.* **49**, 1540–1573 (2010).
28. Kumagai, Y., Shinkai, Y., Miura, T. & Cho, A. K. The chemical biology of naphthoquinones and its environmental implications. *Annu. Rev. Pharmacol. Toxicol.* **52**, 221–247 (2012).
29. Olson, K. R. et al. Redox and nucleophilic reactions of naphthoquinones with small thiols and their effects on oxidization of H2S to inorganic and organic hydropolysulfides and thiosulfate. *Int. J. Mol. Sci.* **24**, 7516 (2023).
30. Wipf, P., Jeger, P. & Kim, Y. Thiophilic ring-opening and rearrangement reactions of epoxyketone natural products. *Bioorg. Med. Chem. Lett.* **8**, 351–356 (1998).
31. Lall, M. S., Karvellas, C. & Vederas, J. C. β-lactones as a new class of cysteine proteinase inhibitors: inhibition of hepatitis A virus 3C proteinase by N-Cbz-serine β-lactone. *Org. Lett.* **1**, 803–806 (1999).
32. Owen, T. C. Thiol detection, derivatization and tagging at micromole to nanomole levels using propiolates. *Bioorg. Chem.* **36**, 156–160 (2008).
33. Busto, O., Guasch, J. & Borrull, F. Determination of biogenic amines in wine after precolumn derivatization with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate. *J. Chromatogr. A* **737**, 205–213 (1996).
34. Motte, J. C., Windey, R. & Delafortrie, A. High-sensitivity fluorescence derivatization for the determination of hydroxy compounds in aqueous solution by high-performance liquid chromatography. *J. Chromatogr. A* **728**, 333–341 (1996).
35. Collins, J., Xiao, Z., Müllner, M. & Connal, L. A. The emergence of oxime click chemistry and its utility in polymer science. *Polym. Chem.* **7**, 3812–3826 (2016).
36. Dührkop, K. et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **39**, 462–471 (2021).
37. Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?. *J. Cheminformatics* **7**, 20 (2015).
38. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
39. Zhao, C. et al. Utilization of the methoxymalonyl-acyl carrier protein biosynthesis locus for cloning the oxazolomycin biosynthetic gene cluster from Streptomyces albus JA3453. *J. Bacteriol.* **188**, 4142–4147 (2006).
40. Mu, Y. et al. Oxazolomycins produced by Streptomyces glaucus and their cytotoxic activity. *RSC Adv.* **11**, 35011–35019 (2021).
41. Nicolaou, K., Simmons, N. L., Chen, J. S., Haste, N. M. & Nizet, V. Total synthesis and biological evaluation of marinopyrrole A and analogs. *Tetrahedron Lett.* **52**, 2041–2043 (2011).

42. Feling, R. H. et al. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus Salinospora. *Angew. Chem. Int. Ed.* **42**, 355–357 (2003).
43. Gill, K. A., Berrué, F., Arens, J. C., Carr, G. & Kerr, R. G. Cystargolides, 20S proteasome inhibitors isolated from Kitasatospora cystarginea. *J. Nat. Prod.* **78**, 822–826 (2015).
44. Illigmann, A. et al. Structure of Staphylococcus aureus ClpP bound to the covalent active-site inhibitor cystargolide A. *Angew. Chem. Int. Ed.* **63**, e202314028 (2024).
45. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
46. Horai, H. et al. MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **45**, 703–714 (2010).
47. Kretschmer, F., Seipp, J., Ludwig, M., Klau, G. W. & Böcker, S. Small molecule machine learning: all models are wrong, some may not even be useful. *bioRxiv* https://doi.org/10.1101/2023.03.27.534311 (2024).
48. Blin, K. et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).
49. Gilchrist, C. L. M. & Chooi, Y.-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).
50. Willcott, M. R. MestRe Nova. *J. Am. Chem. Soc.* **131**, 13180–13180 (2009).

Acknowledgements

We would like to thank Markus Fleischhauer (Bright Giant GmbH) for assistance with the MCheM code implementation in SIRIUS. This study was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) via the Cluster of Excellence EXC 2124: Controlling Microbes to Fight Infection (CMFI, project ID 390838134) to H.B.O., C.C.H., and D.P. and via TRR 261 (project ID 398967434) to H.B.O. and D.P., and a research grant to K.D. and S.B. (BO 1910/23). D.P. was supported by the Simons Foundation through an Simons Early Career Investigator in Aquatic Microbial Ecology and Evolution Award (SFI-LS-ECIAMEE-00013858). M.W. was supported by NIH 5U24DK133658-02 and by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. Additional funding was provided by the Deutsches Zentrum für Infektionsforschung (German Center for Infection Research, DZIF) to H.B.O., C.C.H., Y.M., project TTU 09.826. S.X. is grateful for a PhD scholarship (202008330294) from the Chinese Scholarship Council. C.B. was supported by the Czech Academy of Sciences PPLZ fellowship number L200552251. The mzmine project is funded by the European Union, the BAB - Funding Bank for Bremen and Bremerhaven, and the Senator of Economics, Ports and Transformation Bremen (65002459).

Author contributions

C.C.H. and D.P. conceptualized the MCheM approach and supervised the study. G.A.V., C.C.H. and D.P., performed the derivatization reactions and LC-MS/MS experiments. K.D., M.Z.S.H., C.B., S.B., R.S., and M.W. developed software. G.A.V., K.D., M.Z.S.H. and D.P. performed MCheM data analysis. H.B.O. and Y.M. provided natural product standards and strains. Y.M. performed BGC analysis. S.N.X. and C.C.H. performed natural product purification and NMR experiments. G.A.V., C.C.H. and D.P. wrote the manuscript. All authors edited and approved the manuscript.

Competing interests

M.W. is a co-founder of Ometa labs LLC. R.S. is a co-founder of mzio GmbH (Bremen, Germany). S.B. and K.D. are co-founders of Bright Giant GmbH (Jena, Germany). The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-61240-z.

Correspondence and requests for materials should be addressed to Chambers C. Hughes or Daniel Petras.

Peer review information *Nature Communications* thanks Eliseu Rodrigues and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

4 DISCOVERY OF PIPERAZIC ACID-CONTAINING NATURAL PRODUCTS FROM *STREPTOMYCES AUREOCIRCULATUS* DSM40386

Previous chapters addressed microbial NP discovery from the perspectives of biosynthetic prediction and metabolite level analysis. However, NP research also requires compound isolation, their structural elucidations, and the investigation of their biological properties. Such chemical investigation is essential for fully exploring the biosynthetic potential of individual strains and for supporting the discovery of novel therapeutic agents, particularly in the ongoing effort to combat multidrug resistance. In this context, the chemical investigation of selected strains remains an indispensable part of the overall workflow.

Although different phosphonate producers were identified in Chapter 2, these strains generally harbor additional BGCs that may give rise to NPs belonging to unrelated structural classes. To take this broader biosynthetic capacity into account, the pepM-positive strain collection was systematically examined with the aim of identifying bioactive non-phosphonate secondary metabolites. Through this screening, *Streptomyces aureocirculatus* DSM 40386 was identified as a suitable candidate for further chemical investigation.

By searching for publicly accessible strain databases and literature such as SciFinder, dictionary of natural procucts, the secondary metabolites had barely been reported and structurally characterized directly from *Streptomyces aureocirculatus* DSM 40386, although the genus *Streptomyces* is well known for its pronounced secondary metabolic potential. The information currently available for this strain is largely confined to taxonomic classification, cultivation data, strain synonymy, and genome-associated metadata, whereas compound-level chemical studies appear to be lacking. Interestingly, Liu *et al* reported that a cyclodipeptide synthase associated with *S. aureocirculatus* to generate *cyclo*-L-Trp-L-Ala and *cyclo*-L-Trp-L-Pro in a heterologous expression *E.coli* system.[209] However, these products were not obtained from cultures of *S. aureocirculatus* DSM 40386 itself and therefore cannot be regarded as chemically characterized metabolites of this strain in the strict sense. Therefore, secondary metabolite profile of *S. aureocirculatus* DSM 40386 has remained largely unexplored at the level of purified and structurally defined NPs.

This chapter describes the metabolic investigation of *S. aureocirculatus* DSM 40386, which led to the discovery of piperazic acid-containing peptides. The isolation and structural characterization of these compounds are presented, followed by an initial evaluation of their biological activity using selected bioreporter assays to obtain first insights into their cellular mode of action. In addition, the preliminary detection of a brominated metabolite is reported. Regrettably, this brominated metabolite was not isolated in the thesis.

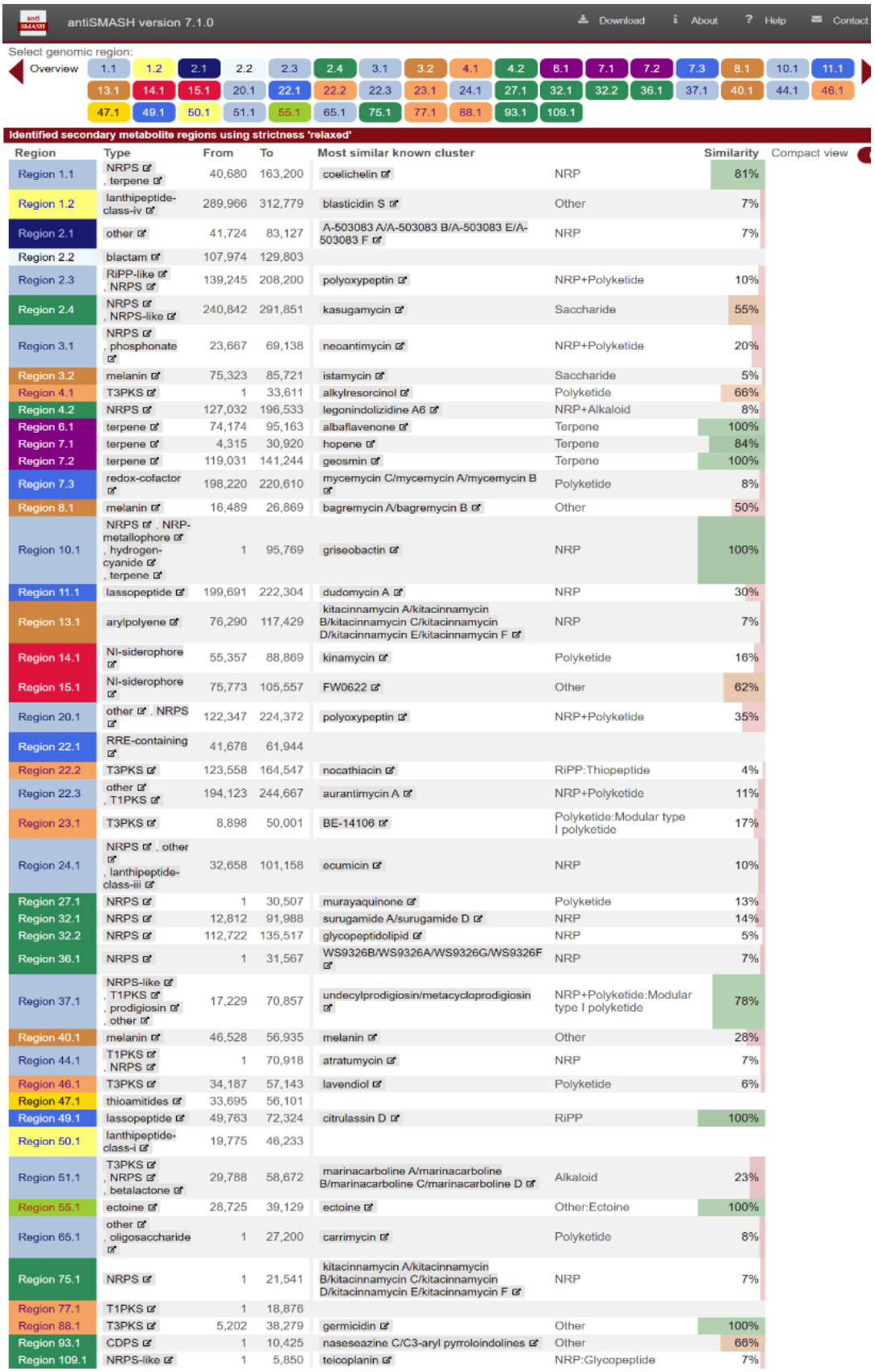
4.1 Results and discussion

4.1.1 AntiSMASH analysis of the *S.aureocirculatus* DSM40386 genome

To explore the potential of S. aureocirculatus DSM40386 as source of new antibiotics, the whole sequence of *S.aureocirculatus* DSM40386 was first retrieved from NCBI and analyzed using the antibiotic and Secondary Metabolite Analysis Shell (antiSMASH 7.0).[73]AntiSMASH analyses indicated the genome contains at least 45 BGCs potentially involved in the biosynthesis of diverse classes of secondary metabolites. Among them, six exhibited a similarity of 100% to the characterized BGC associated with the production of albaflavenone, geosmin, griseobactin, citrulassin D, ectoine, germicidin. More than half of the BGCs were predicted to produce PKS and NRPS, including PKS-NRPS hybrid-derived secondary metabolites, of them nine BGCs have genes that encode type I or III PKS systems. Six BGCs possess genes predicted to produce lanthipeptides, lassopeptide or thioamitides. And Several BGCs are predicted to be hybrid clusters that contain genes that encoding multiple types of scaffold-synthesizing enzymes (**Figure 31**). Thus, *S.aureocirculatus* DSM 40386 is a talented strain contains interesting biosynthetic diversity.

4.1.2 Bioactive metabolites screening with *S. aureocirculatus* DSM40386

The OSMAC approach is a highly effective and classic method in facilitating secondary metabolites by leveraging varied cultivation conditions to extend NPs chemical diversity. To investigate chemical diversity and evaluate the potential of *S. aureocirculatus* DSM40386 as a source of novel antibiotics, crude extracts were prepared from cultures grown under different cultivation conditions and tested for antibacterial activity against *Bacillus subtilis* 168. Of these extracts examined, the crude extract obtained from the self-made medium displayed strongest growth inhibition against *Bacillus subtilis*, as evidenced by the clear inhibition zone in the agar diffusion assay (**Figure 33 A**). This result indicates that the composition of the cultivation medium strongly influences the production of antibacterial metabolites by the strain. To further investigate the chemical basis of this activity, the active crude extract was analyzed by LC-MS and compared with the corresponding medium blank (**Figure 33 B**). The chromatographic profile of the extract showed several distinct peaks that were significantly enriched relative to the blank medium, indicating the presence of strain-derived metabolites. Notably, a distinct ion at *m/z* 551.1177 exhibited a characteristic bromine isotopic pattern, with nearly equal intensities of the M and M+2 peaks (**Figure 34**). HRMS analysis of the protonated molecular ion ($[M+H]^+$) supported the molecular formula assignment of $C_{28}H_{28}BrN_2O_5$ (calcd for $C_{28}H_{28}BrN_2O_5^+$, 551,1180), confirming the presence of a brominated metabolite. In addition, MS/MS analysis of precursor ion at *m/z* 551.1177 revealed a series of product ions at *m/z* 355.0278, 329.0492, 266.0180, and 179.0851 (**Figure A 62**), which is from the cleavage of substituent groups and backbone fragmentation.



Region	Type	From	To	Most similar known cluster		Similarity
Region 1.1	NRPS, terpene	40,680	163,200	coelichelin	NRP	81%
Region 1.2	lanthipeptide-class-iv	289,966	312,779	blasticidin S	Other	7%
Region 2.1	other	41,724	83,127	A-503083 A/A-503083 B/A-503083 E/A-503083 F	NRP	7%
Region 2.2	blactam	107,974	129,803			
Region 2.3	RiPP-like, NRPS	139,245	208,200	polyoxypeptin	NRP+Polyketide	10%
Region 2.4	NRPS, NRPS-like	240,842	291,851	kasugamycin	Saccharide	55%
Region 3.1	NRPS, phosphonate	23,667	69,138	neoantimycin	NRP+Polyketide	20%
Region 3.2	melanin	75,323	85,721	istamycin	Saccharide	5%
Region 4.1	T3PKS	1	33,611	alkylresorcinol	Polyketide	66%
Region 4.2	NRPS	127,032	196,533	legonindolizidine A6	NRP+Alkaloid	8%
Region 6.1	terpene	74,174	95,163	albaflavenone	Terpene	100%
Region 7.1	terpene	4,315	30,920	hopene	Terpene	84%
Region 7.2	terpene	119,031	141,244	geosmin	Terpene	100%
Region 7.3	redox-cofactor	198,220	220,610	mycemycin C/mycemycin A/mycemycin B	Polyketide	8%
Region 8.1	melanin	16,489	26,869	bagremycin A/bagremycin B	Other	50%
Region 10.1	NRPS, NRP-metallophore, hydrogen-cyanide, terpene	1	95,769	griseobactin	NRP	100%
Region 11.1	lassopeptide	199,691	222,304	dudomycin A	NRP	30%
Region 13.1	arylpolyene	76,290	117,429	kitacinnamycin A/kitacinnamycin B/kitacinnamycin C/kitacinnamycin D/kitacinnamycin E/kitacinnamycin F	NRP	7%
Region 14.1	NI-siderophore	55,357	88,869	kinamycin	Polyketide	16%
Region 15.1	NI-siderophore	75,773	105,557	FW0622	Other	62%
Region 20.1	other, NRPS	122,347	224,372	polyoxypeptin	NRP+Polyketide	35%
Region 22.1	RRE-containing	41,678	61,944			
Region 22.2	T3PKS	123,558	164,547	nocathiacin	RiPP:Thiopeptide	4%
Region 22.3	other, T1PKS	194,123	244,667	aurantimycin A	NRP+Polyketide	11%
Region 23.1	T3PKS	8,898	50,001	BE-14106	Polyketide:Modular type I polyketide	17%
Region 24.1	NRPS, other, lanthipeptide-class-iii	32,658	101,158	ecumicin	NRP	10%
Region 27.1	NRPS	1	30,507	murayaquinone	Polyketide	13%
Region 32.1	NRPS	12,812	91,988	surugamide A/surugamide D	NRP	14%
Region 32.2	NRPS	112,722	135,517	glycopeptidolipid	NRP	5%
Region 36.1	NRPS	1	31,567	WS9326B/WS9326A/WS9326G/WS9326F	NRP	7%
Region 37.1	NRPS-like, T1PKS, prodigiosin, other	17,229	70,857	undecylprodigiosin/metacycloprodigiosin	NRP+Polyketide:Modular type I polyketide	78%
Region 40.1	melanin	46,528	56,935	melanin	Other	28%
Region 44.1	T1PKS, NRPS	1	70,918	atratumycin	NRP	7%
Region 46.1	T3PKS	34,187	57,143	lavendiol	Polyketide	6%
Region 47.1	thioamitides	33,695	56,101			
Region 49.1	lassopeptide	49,763	72,324	citrulassin D	RiPP	100%
Region 50.1	lanthipeptide-class-i	19,775	46,233			
Region 51.1	T3PKS, NRPS, betalactone	29,788	58,672	marinacarboline A/marinacarboline B/marinacarboline C/marinacarboline D	Alkaloid	23%
Region 55.1	ectoine	28,725	39,129	ectoine	Other:Ectoine	100%
Region 65.1	other, oligosaccharide	1	27,200	carrimycin	Polyketide	8%
Region 75.1	NRPS	1	21,541	kitacinnamycin A/kitacinnamycin B/kitacinnamycin C/kitacinnamycin D/kitacinnamycin E/kitacinnamycin F	NRP	7%
Region 77.1	T1PKS	1	18,876			
Region 88.1	T3PKS	5,202	38,279	germicidin	Other	100%
Region 93.1	CDPS	1	10,425	naseseazine C/C3-aryl pyrroloindolines	Other	66%
Region 109.1	NRPS-like	1	5,850	teicoplanin	NRP:Glycopeptide	7%

Figure 32. AntiSMASH-predicted BGCs for S.aureocirculatus DSM40386

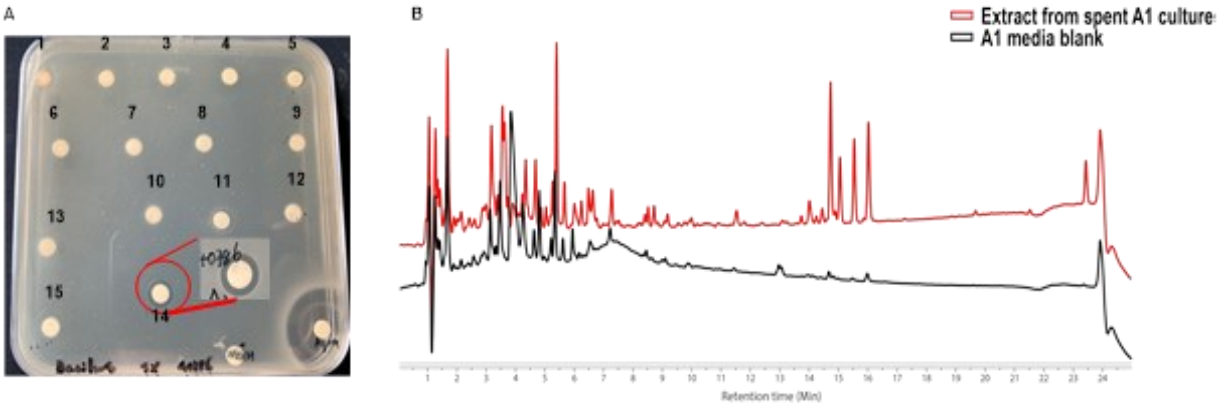


Figure 33. Result of bioactivity of different crude extracts against *Bacillus subtilis* 168(**A**) and stack chromatogram of A1 medium blank and the crude extract at 254 nm(**B**). The crude extract from spent self-made A1medium showed strongest growth inhibition and marked with circle. The chromatographic profile of the extract displayed several distinct peaks compared to A1 medium blank, suggesting the presence of multiple metabolites.

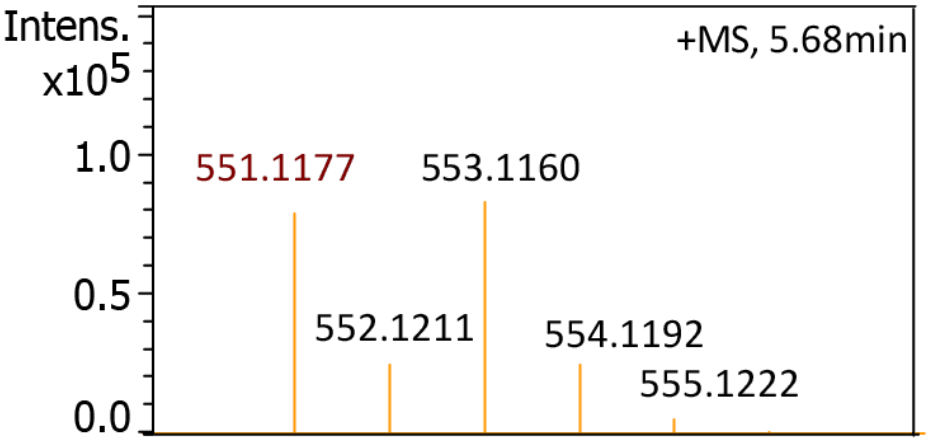


Figure 34. MS pattern of brominated metabolite.

The preservation of the bromine isotopic pattern in several fragment ions indicates that the bromine atom is retained within the core structure rather than attached on a labile side chain. These findings further demonstrate that the antibacterial activity of the crude extract may be attributable, at least in part, to one or more of these secondary metabolites.

4.1.3 Isolation and structure elucidation of PAC peptides

To isolate the main products for validating our hypothesis, the supernatants of 10L combined *S. aureocirculatus* DSM40386 cultures were extracted with EtOAc and subsequently purified by flash chromatography and preparative HPLC. UV and MS-guided fractionation followed by structure elucidation yielded 6.4 mg of PAC1, 1.2 mg of compound PAC2, 7.7 mg of compound PAC3, 2.5 mg of compound PAC4. The NMR spectra was shown in A4.
Analysis of the COSY data of 1 revealed at least four major spin systems. The first spin system consisted of H-1, H-2, H-3, and H-4, defining a contiguous fragment that, together with the

HMBC correlation from H-4 to C-5, was consistent with a proline-like five-membered nitrogen-containing ring. The second and third spin systems were highly similar and extended from H-8 to H-11 and from H-17 to H-20, respectively.

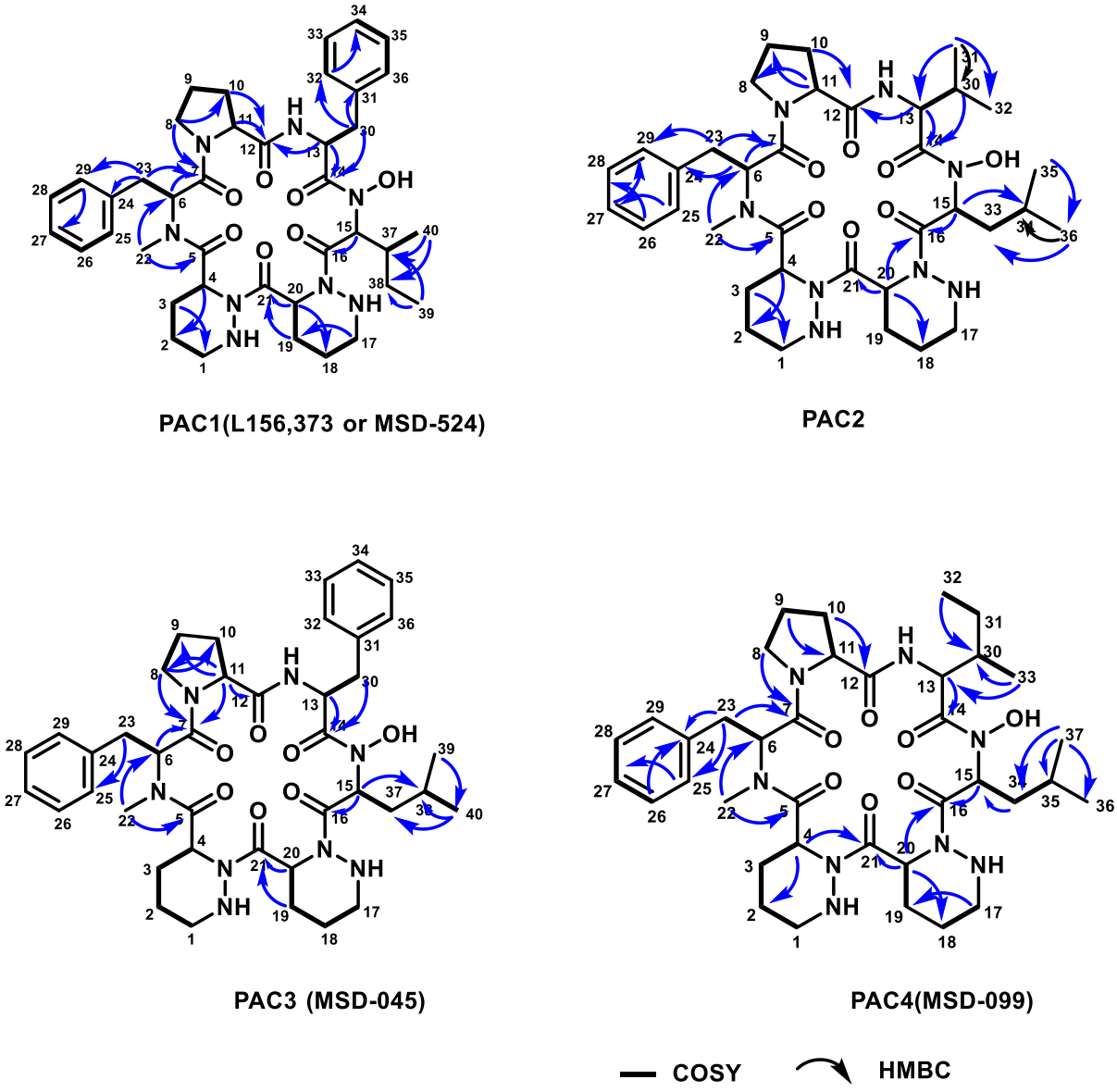


Figure 35. Structures of PAC1-4 and key HMBC and COSY correlations

Each of these fragments contained one α-methine and several methylene groups and showed HMBC correlations to adjacent carbonyl carbons, supporting the presence of two conserved six-membered nitrogen-containing cyclic amino acid-like units. The fourth spin system corresponded to a benzyl-aryl fragment involving C-23 and C-24-C-29, in which C-23 appeared as a benzylic methylene (δ_C 33.72, δ_H 3.12/2.92) and C-25-C-29 formed a monosubstituted phenyl ring. A second benzyl–aryl fragment was defined by C-30 and C-31-C-36, with C-30 resonating as a benzylic methylene (δ_C 37.06, δ_H 2.98/2.77), thereby indicating the presence of a second phenylalanine-derived aromatic side chain.

These partial structures were then connected using HMBC correlations to establish the structure of 1. The proline-like unit was linked into the peptide framework through correlations

to C-5. The N-methyl amino acid-derived aromatic fragment was secured by the HMBC correlations from H-22 to C-5/C-6 and the benzylic/aromatic correlations from H-23 to C-24-C-29. Likewise, the second aromatic side chain was established from correlations involving H-30 and aromatic carbons C-31-C-36. In addition, C-15 (δ_C 59.80, δ_H 5.31 (d, J = 8.4 Hz)) was connected to an aliphatic side chain composed of C-37-C-40, including one methine, one methylene, and two terminal methyl groups, forming a characteristic sec-butyl fragment. This pattern is most consistent with an isoleucine-derived side chain. The NMR data thus established compound PAC1 as a cyclic peptide congener containing one conserved N-methyl aromatic amino acid unit, one additional phenylalanine-derived side chain, and one isoleucine-derived aliphatic residue. After searching from NPs database, the structure of compound PAC1 was unfortunately reported, namely MSD-524 with bioactivity against *M. tuberculosis* and L156,737 as a selective oxytocin antagonist.[210,211]

Compound PAC2 was obtained as a white amorphous solid, and its formula was assigned as $C_{36}H_{54}N_8O_7$ by HRMS and ^1H and ^{13}C NMR spectroscopic data. The ^1H and multiplicity-edited HSQC NMR data in CD_3OD showed that Compound PAC2 multiple carbonyl signals in the region δ_C 169.75-173.22, the characteristic N-methyl singlet at C-22 (δ_C 31.14, δ_H 2.78 (s)) with HMBC correlations to C-5 and C-6, a proline-like spin system from C-1 to C-4, and two highly similar six-membered nitrogen-containing ring systems spanning C-8-C-11 and C-17-C-20. The benzyl-monosubstituted phenyl fragment at C-23-C-29 was also retained, indicating that the conserved N-methyl aromatic amino acid residue remained unchanged.

The main differences between 2 and 1 were observed at the side chains attached to C-13 and C-15. In compound 2, C-13 (δ_C 53.71, δ_H 4.71 (t, J = 8.5 Hz)) was connected to C-30 and two methyl groups, C-31 and C-32, giving a typical isopropyl pattern that is characteristic of a valine-derived side chain. In contrast, the side chain attached to C-15 consisted of C-33-C-36 and displayed an isobutyl pattern, with two methyl doublets attached to a methine/methylene unit, consistent with a leucine-derived side chain. With the remainder of the NMR data being essentially the same as those of 1, compound 2 was assigned as a closely related and new congener in which one aromatic side chain of 1 is replaced by a valine-derived substituent and the isoleucine-derived side chain is replaced by a leucine-derived one.

Compound PAC3 was obtained as a white amorphous solid, and its formula was assigned as $C_{40}H_{54}N_8O_7$ by HRMS and ^1H and ^{13}C NMR spectroscopic data. The ^1H and multiplicity-edited HSQC NMR data in CD_3OD showed that Compound PAC3 contained the same general set of carbonyl carbons, one N-methyl signal at C-22 (δ_C 31.15, δ_H 2.82 (s)), the conserved aromatic fragment at C-23-C-29, and the same proline-like and two six-membered nitrogen-containing ring substructures observed in 1 and 2.

Comparison of the side-chain region showed that compound 3 contains a second aromatic side chain at C-13. Specifically, C-13 (δ_C 50.51, δ_H 5.06 (m)) was connected to a benzylic methylene C-30 (δ_C 36.50, δ_H 2.97/2.84), which in turn showed HMBC correlations to an aromatic quaternary carbon C-31 and aromatic methine carbons C-32-C-36. These data defined a second phenylalanine-derived side chain. In contrast, the side chain attached to C-15 was composed of C-37-C-40 and showed an isobutyl pattern characteristic of a leucine-derived side chain. Thus, with the remainder of the NMR data essentially matching those of 1, compound 3 was assigned as a congener containing the same pair of aromatic residues as 1, but with a leucine-derived side chain in place of the isoleucine-derived residue.

Compound PAC4 was obtained as a white amorphous solid, and its formula was assigned as $C_{37}H_{54}N_8O_7$ by HRMS and ^1H and ^{13}C NMR spectroscopic data. The ^1H and multiplicity-edited HSQC NMR data in CD_3OD showed that Compound PAC4 as in compounds 1-3, the NMR data of 4 supported the same conserved cyclic peptide core, including multiple amide carbonyls, the characteristic N-methyl signal at C-22 (δ_C 31.13, δ_H 2.78 (s)), the proline-like fragment, the two six-membered nitrogen-containing ring systems, and the common N-methyl aromatic amino acid-derived benzyl/phenyl unit at C-23-C-29.
Both variable side chains in 4 were aliphatic. The side chain attached to C-13 (δ_C 52.36, δ_H 4.82) consisted of C-30-C-33 and displayed a sec-butyl pattern, which is consistent with an isoleucine-derived side chain. The side chain attached to C-15 [δ_C 56.20, δ_H 5.30] consisted of C-34-C-37 and showed an isobutyl pattern typical of a leucine-derived side chain. With the remainder of the structure unchanged, compound 4 was thus established as a congener carrying both isoleucine- and leucine-derived variable aliphatic residues.
Compound PAC1-4 thus were a related family of piperazic acid-containing cyclic peptides sharing a conserved macrocyclic backbone but differing in two variable amino acid derived side chains, indicating that they were NRPS-derived compounds with common biosynthetic origin.

4.1.4 Screening of putative PAC peptides biosynthetic gene clusters

Even though compound PAC1 was identified as L156,373 for long time, the complete biosynthesis pathway hasn't been confirmed. To elucidate the biosynthetic origin of the isolated piperazic acid-containing peptides, all NRPS-related regions were examined in detail. A putative NRPS gene cluster spanning approximately 100 kb of region 20.1 was identified. The BGC shows 35% similarity to that of aurantimycin A and polyoxypeptin, which contains piperazic acid moieties. And region 20.1 revealed at least three large biosynthetic genes directly involved in NRPS assembly, namely ctg20_139, ctg20_141, and ctg20_142.
According to the antiSMASH prediction, ctg20_139 encodes one adenylation domain, ctg20_141 encodes four adenylation domains, and ctg20_142 encodes one Piz-activating adenylation domain,[212] for a total of six A domains. This module number is in good agreement with the hexapeptidic macrocyclic scaffold of PAC1-4, strongly suggesting that this cluster can assemble a six-residue peptide backbone.
In addition to the core NRPS genes, region 20.1 also contains several auxiliary genes that may be involved in precursor supply and post-assembly tailoring. Notably, ctg20_135 is annotated as a cytochrome P450 protein, suggesting that this cluster has the capacity to catalyze oxidative modification. Given that the structure of PAC1 contains an N-OH-L-Ile unit, indicating that at least one oxidation step is required during biosynthesis, this P450 enzyme may participate in the oxidative tailoring of the target compounds. Also, region 20.1 encodes a number of regulatory and accessory elements, including an LmbU-type regulatory protein (ctg20_133), two SARP family regulators (ctg20_137 and ctg20_146), an MbtH-like protein (ctg20_143), and multiple ABC transporter-related genes. Such components are commonly associated with expression control, substrate activation, and product export in NRPS-derived natural product pathways. Their co-occurrence further supports the interpretation that this

region constitutes a relatively complete secondary metabolite biosynthetic locus. In particular, MbtH-like proteins are known in many NRPS systems to facilitate A-domain function and substrate recognition,[213,214] And its presence therefore provides additional support for the assignment of region 20.1 as an active NRPS biosynthetic cluster. In addition, the gene, ctg20_140, annotated as a halogenase, which is located adjacent to the principal NRPS genes within the core biosynthetic region rather than at a peripheral position, suggesting its possible involvement in product assembly or downstream tailoring. Interestingly, in addition to the four purified major compounds, LC-MS analysis indeed detects a low-abundance metabolite peak displaying a characteristic bromine-like isotopic pattern. Nevertheless, extra targeted genetic experiments and verification remain to be finished.

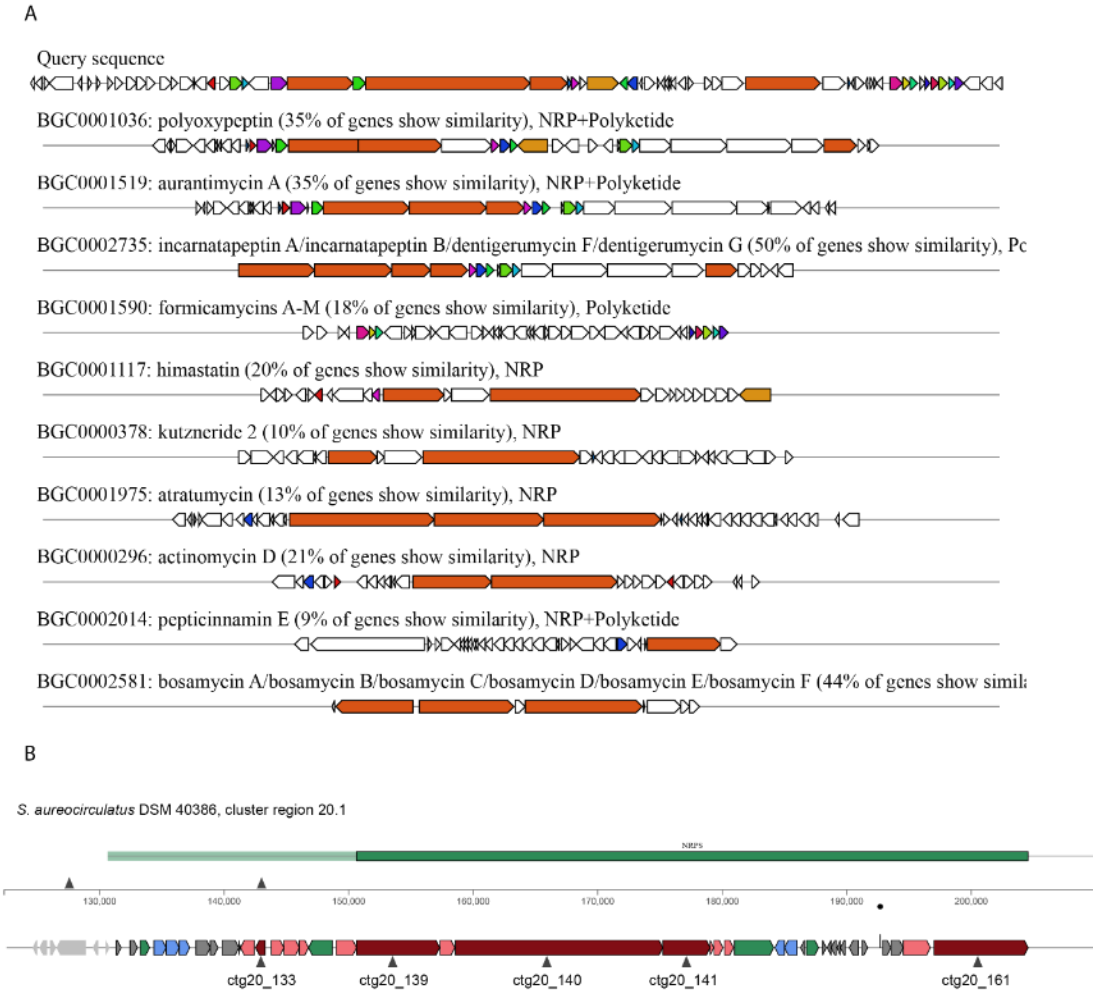


Figure 36. Result of Known Cluster Blast (A) and antiSMASH prediction about region 20.1(B). PAC BGC shows 35% similarity to that of aurantimycin A and polyoxypeptin.

4.1.5 Antibacterial activity and mode of action

The biological activities of PAC 1-4 were assessed with *S.aureus* USA 300 and *B.subtilis* 168. PAC1 exhibited strong antibacterial activity against *S.aureus* USA 300 and showed a minimum inhibitory concentration (MIC) value of 0.25 μg/ml. It showed strong activity against *B.subtilis* 168 with a MIC of 0.25 μg/ml. PAC3 exhibited strong antibacterial activity against *S.aureus*

USA 300 and showed a MIC value of 1 µg/ml. It showed strong activity against *B.subtilis* 168 with a MIC of 1 µg/ml. PAC4 exhibited strong antibacterial activity against *S.aureus* USA 300 and showed a MIC value of 8 µg/ml. It showed strong activity against *B.subtilis* 168 with a MIC of 8 µg/ml. Unfortunately, no enough PAC2 was obtained for full bioactivity experiments. Although the bioactivity experiments tested are limited, their inhibition against the human pathogen *Mycobacterium tuberculosis* has been determined.[210]

To further investigate the mode of action of these compounds, the agar-based reporter with different biomarkers (*clpE*, *yorB*, *yppS*, *bmrC*, *fabHB*, *ypuA*, and *liaI)* were specifically used.

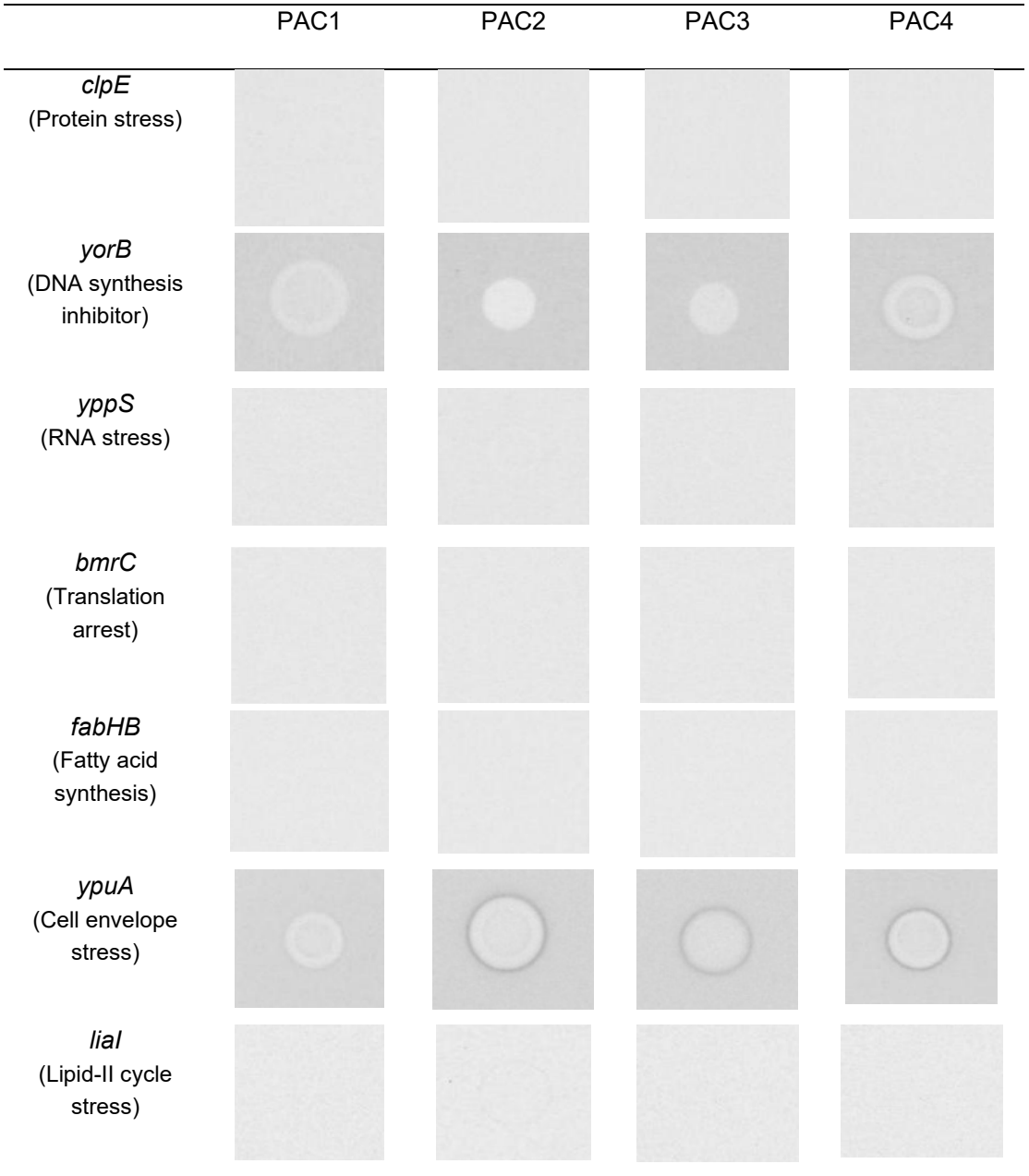


Figure 37. Reporter-based profiling of antibiotic stress biomarkers in response to PAC1-4. All four compounds induced the *ypuA* reporter, consistent with cell envelope-associated stress, while no substantial induction was observed for *clpE, yppS, bmrC, fabHB*, or *liaI*. PAC1 and PAC4 additionally induced *yorB*, suggesting a possible DNA synthesis or replication associated with stress response.

The stress responsive biomarker panel revealed a distinct induction profile for the four compounds. The biomarker *clpE*, *yorB*, *yppS*, *bmrC*, *fabHB*, *ypuA*, and *liaI* have previously been shown to respond specifically to different antibiotic-induced stress pathways, corresponding to protein misfolding, DNA synthesis inhibition, RNA, translation inhibition, fatty acid synthesis inhibition, cell envelope stress, and lipid-II cycle stress, respectively.[215]All four compounds induced a clear response in the *ypuA* reporter, whereas no evident induction was observed for *clpE*, *yppS*, *bmrC*, *fabHB*, or *liaI*. In addition, PAC1 and PAC4 also showed an additional induction of the DNA synthesis stress biomarker *yorB*. This response pattern indicates that compound PAC1-4 primarily trigger cell envelope associated processes, and in the case of PAC1 and PAC4, may additionally induce DNA replication or DNA synthesis associated stress. By contrast, the absence of induction of *clpE*, *yppS*, *bmrC*, *fabHB*, or *liaI* shows that these compounds are unlikely to act as canonical translation inhibitors, RNA synthesis inhibitors, fatty acid biosynthesis inhibitors, or direct Lipid II cycle-targeting cell wall antibiotics. These results provide valuable guidance for the design of subsequent experiments to further elucidate the underlying molecular target of PAC1-4.

4.2 Material and methods

4.2.1 Cultivation of *S. aureocirculatus* DSM 40386

The fresh agar was prepared. Then a small square block about 1cm^2 was picked with a sterilized toothpick from culture agar and transferred into sterilized 100mL Erlenmeyer baffled flasks (“mit Schikane”) with metal coil containing 30mL R5 or HM liquid medium. The pre-culture was grown with medium at 150rpm and 29 °C on a platform shaker. After three days, 5 mL of preculture was used to inoculate a 500 mL flask containing 150 mL A1 fermentation medium (in total 4L). The main culture was grown at 100 rpm and 29 °C for four days.

4.2.2 Extraction and purification of PAC peptides

The complete culture (4 L) was cultivated, centrifuged, and the supernatant was collected. The supernatant was then extracted three times with an equal volume of ethyl acetate. The EtOAc extract was dried over NaSO4 and concentrated under reduced pressure to give a crude extract. The extract was fractionated on silica gel using a stepwise gradient elution with 100% hexanes, 30% EtOAc in hexanes, 100% EtOAc, 1% MeOH in DCM, 10% MeOH in DCM, 100% MeOH. Fraction 5 (10% MeOH in DCM) was dried and further fractionated on C18 SPE using a stepwise gradient elution with 100% H_2O, 20% MeCN in H_2O, 40% MeCN in H_2O, 60% MeCN in H_2O, 80% MeCN in H_2O, 100% MeCN. All elution solvents contained 0.1% Trifluoroacetic

acid (TFA). Fraction 4(60% MeCN in H_2O) was dried and further purified by semi-HPLC Kintex® 5 µm C18 100 Å (250 X 21.2 mm), 65 % MeCN in H_2O (0.1% FA) ,13 ml/min, 220 nm. The compounds were collected as shown in Figure A61.

4.2.3 Structure elucidation of PAC peptides

The NMR experiments were kindly performed by NMR department at Eberhard Karls Universität Tübingen. The data was made available to me for further analysis.

The compounds were analyzed by LC-MS using with standard method. All compounds were characterized by 1H, 13C, and 2D (COSY, HSQC, HMBC, NOESY) NMR spectroscopy on a Bruker Avance III HDX 700 MHz spectrometer fitted with a 5 mm Prodigy (1H,19F/13C/15N) TCI Cryoprobe. 1H NMR data were recorded at 700 MHz in DMSO-d6 (2.50 ppm), and 13C NMR data were recorded at 175 MHz in DMSO-d6 (39.5 ppm). NMR spectra were processed using MestReNova 14.3.0.

4.2.4 Bioreporter assays

The bioreporter assays were kindly performed by Annika Schulz (Technical Assistant, AG Hughes) at IMIT. The data was made available to me for further analysis.

Reporter strains carrying lux fusions were grown overnight from cryostocks in 10 mL LB medium containing 5 µg/mL chloramphenicol at 37 °C and 190 rpm. Overnight cultures were diluted into 20 mL antibiotic-free LB to an initial OD_{600} of 0.05 and grown at 37 °C and 190 rpm to an OD_{600} of 0.8-1.3 (typically ~1.0). Soft agar was prepared in parallel, using LB soft agar for *yorB*, *ypuA, liaI, clpE,* and *fabHB*, and Belitzky soft agar for *yppS* and *yheI/bmrC*. Cells were used directly without centrifugation, and no $MgCl_2$ or X-Gal was added. The culture volume corresponding to 3×10^6 cells was calculated using the formula Volume (mL) = [6 / (OD_{600} × 19)] × 2.5. Cells were mixed with soft agar, poured onto plates (50 mL per large plate), and the plates were dried for ~30 min. Test compounds were then spotted onto the agar surface. Positive controls were ciprofloxacin for *yorB*, vancomycin for *ypuA*, daptomycin (1 µg) for *liaI*, erythromycin (0.1 µg) for *yheI/bmrC*, rifampicin (0.5 µg) for *yppS*, kanamycin for *clpE*, and triclosan (2.5 µg) for f*abHB*. Plates were incubated at 37 °C for 3–3.5 h, except for Belitzky agar plates, which were incubated for 4 h; *yppS* was evaluated after both 3 h and 4 h. Luminescence was recorded with a ChemiDoc system using an approximately 10 min exposure with 1 image acquired per minute. Plates were subsequently incubated overnight at 30 °C, and inhibition zones were assessed the next day.

5 CONCLUSION

5.1 Genome-guided identification and characterization of phosphonate biosynthesis in actinomycetes

This chapter demonstrates that a genome-based approach is an effective strategy for uncovering phosphonate biosynthetic potential in actinobacteria. Genome mining of the DSMZ and Tübingen strain collections using *pepM* as a marker revealed a broad and previously underexplored diversity of putative phosphonate biosynthetic gene clusters, confirming that these collections represent a valuable resource for phosphonate natural product discovery.
The OSMAC strategy, combined with ^{31}P NMR analysis, further showed that phosphonate production is widespread among *pepM*-positive strains but strongly dependent on medium composition and phosphate availability. These findings establish a direct link between genome-predicted biosynthetic capacity and experimentally detectable metabolite production, while also highlighting the importance of culture optimization for activating phosphonate pathways. In addition, phylogenetic analysis of *pepM* proved to be a practical tool for strain prioritization. On this basis, *S. kutzneri* DSM40907 was identified as a new producer of phosphonoalamides, and the phosphonate biosynthetic loci of *S. iranensis* DSM41954 and *K. fiedleri* DSM114396 were experimentally confirmed. Furthermore, genetic analyses showed that overexpression of early biosynthetic genes can enhance phosphonate production, and that the LuxR-family regulator Kfp24 functions as a pathway-specific activator in *K. fiedleri.*
Given the characteristics of phosphonates, several complementary approaches were explored. This included methanol precipitation, activated charcoal treatment, weak anion-exchange chromatography, HILIC, Hypercarbon chromatography, and three labeling strategies based on FMOC-Cl, an aryl diazo probe, and ECF. Although the complete isolation and structural elucidation of the target unknown phosphonates remain to be achieved, the extraction, enrichment, and LC-MS compatible detection approaches established here provide an important methodological basis for future work.

5.2 Enhancing tandem mass spectrometry-based metabolite with online chemical labeling

The present work demonstrates that MCheM is a useful strategy for MS/MS acquisition and data interpretation that strengthens metabolite annotation in untargeted metabolomics. The

approach is supported by a scalable implementation in MZmine for detecting MCheM-derived features, which can be coupled with downstream annotation platforms such as SIRIUS and GNPS2 to improve metabolite identification. Our findings show that adding reactivity-derived information from MCheM markedly increases annotation performance in both in silico and open-modification search workflows, as illustrated by the discovery of a previously unknown glycosylated oxazolomycin D.

An important feature contributing to the practical value of MCheM is the accessibility of its experimental setup. The method can be deployed on most commercial LC–MS/MS systems with only modest additions, including a second HPLC, a syringe pump, and an available software workflow. The gains reported here were obtained using just three derivatization reactions, highlighting the efficiency of the platform. In principle, further reactions directed at other functional groups could be incorporated in the same framework to extend annotation capabilities even more. We expect MCheM to serve as a broadly useful tool for enriching structural insight in untargeted metabolomics and to support future integration with emerging computational methods aimed at fully MS/MS-based de novo structure elucidation.

5.3 Structure characterization and antibacterial evaluation of cyclic peptides from *Streptomyces aureocirculatus* DSM40386

The present work demonstrates that *Streptomyces aureocirculatus* DSM40386 is a promising source of bioactive piperazic acid-containing natural products. Genome analysis revealed substantial biosynthetic capacity, including at least 45 predicted biosynthetic gene clusters, which justified its selection for detailed chemical investigation. Guided by cultivation-based induction, bioactivity screening, and LC–MS analysis, four related cyclic peptides, L156,373 and its derivatives, PAC1-4, were isolated and structurally characterized as a congeneric family sharing a conserved macrocyclic scaffold with variable amino acid-derived side chains, consistent with a common NRPS-dependent biosynthetic origin

Biological evaluation showed that PAC1, PAC3, and PAC4 possess antibacterial activity against *Staphylococcus aureus* USA300 and *Bacillus subtilis 168*, with PAC1 displaying the highest potency. Bioreporter assays further suggested that these compounds primarily induce cell envelope stress, while PAC1 and PAC4 additionally affect DNA synthesis-associated stress responses, indicating a mode of action distinct from that of several classical antibiotic classes.

In addition, the screening of a putative NRPS biosynthetic gene cluster in region 20.1 provides a plausible genetic basis for PAC biosynthesis. The presence of six adenylation domains, together with tailoring, regulatory, and transporter-associated genes, supports the assignment

of this locus to the production of the isolated hexapeptidic metabolites. The detection of a neighboring halogenase gene and a low-abundance brominated metabolite suggests that the metabolic potential of this strain may not yet be fully explored. The present work might provide a starting point for further minor congeners characterization, including the putative brominated derivative and biosynthetic mechanism studies to assess this compound family potential as a scaffold for further antibiotic development.

6 REFERENCES

(1) Newman, D. J.; Cragg, G. M. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *Journal of Natural Products* **2012**, *75* (3), 311-335. DOI: 10.1021/np200906s.
(2) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *Journal of Natural Products* **2020**, *83* (3), 770-803. DOI: 10.1021/acs.jnatprod.9b01285.
(3) Yu, H.; Chen, Y.; Wang, Y.; Fu, W.; Xu, R.; Liu, J.; Chen, Y.; Liu, X.; Wu, Y.; Xu, T. Integrated Thermal Proteome Profiling and Affinity Ultrafiltration Mass Spectrometry (iTPAUMS): A Novel Paradigm for Elucidating the Mechanism of Action of Natural Products. *Analytical Chemistry* **2024**. DOI: 10.1021/acs.analchem.4c03398.
(4) Bérdy, J. Bioactive Microbial Metabolites. *The Journal of Antibiotics* **2005**, *58* (1), 1-26. DOI: 10.1038/ja.2005.1.
(5) Walsh, C. T.; Fischbach, M. A. Natural Products Version 2.0: Connecting Genes to Molecules. *Journal of the American Chemical Society* **2010**, *132* (8), 2469-2493. DOI: 10.1021/ja909118a.
(6) Brötz-Oesterhelt, H.; Sass, P. Postgenomic Strategies in Antibacterial Drug Discovery. *Future Microbiology* **2010**, *5* (10), 1553-1579. DOI: 10.2217/fmb.10.119.
(7) Hutchings, M. I.; Truman, A. W.; Wilkinson, B. Antibiotics: past, present and future. *Current Opinion in Microbiology* **2019**, *51*, 72-80. DOI: 10.1016/j.mib.2019.10.008.
(8) Gavriilidou, A.; Kautsar, S. A.; Zaburannyi, N.; Krug, D.; Müller, R.; Medema, M. H.; Ziemert, N. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nature Microbiology* **2022**, *7* (5), 726-735. DOI: 10.1038/s41564-022-01110-2.
(9) Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; Mueller, A.; Schäberle, T. F.; Hughes, D. E.; Epstein, S.; et al. A new antibiotic kills pathogens without detectable resistance. *Nature* **2015**, *517* (7535), 455-459. DOI: 10.1038/nature14098.
(10) Imai, Y.; Meyer, K. J.; Iinishi, A.; Favre-Godal, Q.; Green, R.; Manuse, S.; Caboni, M.; Mori, M.; Niles, S.; Ghiglieri, M.; et al. A new antibiotic selectively kills Gram-negative pathogens. *Nature* **2019**, *576* (7787), 459-464. DOI: 10.1038/s41586-019-1791-1.
(11) Zipperer, A.; Konnerth, M. C.; Laux, C.; Berscheid, A.; Janek, D.; Weidenmaier, C.; Burian, M.; Schilling, N. A.; Slavetinsky, C.; Marschal, M.; et al. Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **2016**, *535* (7613), 511-516. DOI: 10.1038/nature18634.
(12) Ruppelt, D.; Trollmann, M. F. W.; Dema, T.; Wirtz, S. N.; Flegel, H.; Mönnikes, S.; Grond, S.; Böckmann, R. A.; Steinem, C. The antimicrobial fibupeptide lugdunin forms water-filled channel structures in lipid membranes. *Nature Communications* **2024**, *15* (1), 3521. DOI: 10.1038/s41467-024-47803-6.
(13) Barka Essaid, A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; van Wezel Gilles, P. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiology and Molecular Biology Reviews* **2015**, *80* (1), 1-43. DOI: 10.1128/mmbr.00019-15 (acccessed 2024/11/05).
(14) Kayiranga, A.; Isabwe, A.; Yao, H.; Shangguan, H.; Coulibaly, J. L. K.; Breed, M.; Sun, X. Distribution patterns of soil bacteria, fungi, and protists emerge from distinct assembly processes across subcommunities. *Ecology and Evolution* **2024**, *14* (7), e11672. DOI: 10.1002/ece3.11672.
(15) Wu, D.; Bai, H.; Zhao, C.; Peng, M.; Chi, Q.; Dai, Y.; Gao, F.; Zhang, Q.; Huang, M.; Niu, B. The characteristics of soil microbial co-occurrence networks across a high-latitude forested wetland ecotone in China. *Frontiers in Microbiology* **2023**, *Volume 14 - 2023*, Original Research. DOI: 10.3389/fmicb.2023.1160683.
(16) Genilloud, O. Actinomycetes: still a source of novel antibiotics. *Natural Product Reports* **2017**, *34* (10), 1203-1232, 10.1039/C7NP00026J. DOI: 10.1039/C7NP00026J.

(17) Staunton, J.; Wilkinson, B. Biosynthesis of Erythromycin and Rapamycin. *Chemical Reviews* **1997**, *97* (7), 2611-2630. DOI: 10.1021/cr9600316.
(18) Bruton, J.; Horner, W. H.; Russ, G. A. Biosynthesis of Streptomycin: IV. FURTHER STUDIES ON THE BIOSYNTHESIS OF STREPTIDINE AND N-METHYL-I-GLUCOSAMINE. *Journal of Biological Chemistry* **1967**, *242* (5), 813-818. DOI: 10.1016/S0021-9258(18)96198-3 (acccessed 2025/01/17).
(19) Yurkovich, M. E.; Tyrakis, P. A.; Hong, H.; Sun, Y.; Samborskyy, M.; Kamiya , K.; Leadlay, P. F. A Late-Stage Intermediate in Salinomycin Biosynthesis Is Revealed by Specific Mutation in the Biosynthetic Gene Cluster. *ChemBioChem* **2012**, *13* (1), 66-71.
DOI: 10.1002/cbic.201100590 .
(20) Gupta, P. B.; Onder, T. T.; Jiang, G.; Tao, K.; Kuperwasser, C.; Weinberg, R. A.; Lander, E. S. Identification of Selective Inhibitors of Cancer Stem Cells by High-Throughput Screening. *Cell* **2009**, *138* (4), 645-659. DOI: 10.1016/j.cell.2009.06.034.
(21) Fuchs, D.; Daniel, V.; Sadeghi, M.; Opelz, G.; Naujokat, C. Salinomycin overcomes ABC transporter-mediated multidrug and apoptosis resistance in human leukemia stem cell-like KG-1a cells. *Biochemical and Biophysical Research Communications* **2010**, *394* (4), 1098-1104. DOI: 10.1016/j.bbrc.2010.03.138.
(22) Wu, C.; van der Heul, H. U.; Melnik, A. V.; Lübben, J.; Dorrestein, P. C.; Minnaard, A. J.; Choi, Y. H.; van Wezel, G. P. Lugdunomycin, an Angucycline-Derived Molecule with Unprecedented Chemical Architecture. *Angewandte Chemie International Edition* **2019**, *58* (9), 2809-2814. DOI: 10.1002/anie.201814581.
(23) Keller, S.; Nicholson, G.; Drahl, C.; Sorensen, E.; Fiedler, H.-P.; Süssmuth, R. D. Abyssomicins G and H and atrop-Abyssomicin C from the Marine Verrucosispora Strain AB-18-032†. *The Journal of Antibiotics* **2007**, *60* (6), 391-394. DOI: 10.1038/ja.2007.54.
(24) Schorn, M. A.; Alanjary, M. M.; Aguinaldo, K.; Korobeynikov, A.; Podell, S.; Patin, N.; Lincecum, T.; Jensen, P. R.; Ziemert, N.; Moore, B. S. Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiology* **2016**, *162* (12), 2075-2086. DOI: 10.1099/mic.0.000386.
(25) Ortiz-López, F. J.; Oves-Costales, D.; Guerrero Garzón, J. F.; Gren, T.; Baggesgaard Sterndorff, E.; Jiang, X.; Sparholt Jørgensen, T.; Blin, K.; Fernández-Pastor, I.; Tormo, J. R.; et al. Genome-Led Discovery of the Antibacterial Cyclic Lipopeptide Kutzneridine A and Its Silent Biosynthetic Gene Cluster from Kutzneria Species. *Journal of Natural Products* **2024**, *87* (10), 2515-2522. DOI: 10.1021/acs.jnatprod.4c00633.
(26) Bauman, K. D.; Li, J.; Murata, K.; Mantovani, S. M.; Dahesh, S.; Nizet, V.; Luhavaya, H.; Moore, B. S. Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry. *Cell Chemical Biology* **2019**, *26* (5), 724-736.e727. DOI: 10.1016/j.chembiol.2019.02.004.
(27) Kaniusaite, M.; Goode, R. J. A.; Tailhades, J.; Schittenhelm, R. B.; Cryle, M. J. Exploring modular reengineering strategies to redesign the teicoplanin non-ribosomal peptide synthetase. *Chemical Science* **2020**, *11* (35), 9443-9458, 10.1039/D0SC03483E. DOI: 10.1039/D0SC03483E.
(28) Yuzawa, S.; Mirsiaghi, M.; Jocic, R.; Fujii, T.; Masson, F.; Benites, V. T.; Baidoo, E. E. K.; Sundstrom, E.; Tanjore, D.; Pray, T. R.; et al. Short-chain ketone production by engineered polyketide synthases in Streptomyces albus. *Nature Communications* **2018**, *9* (1), 4569. DOI: 10.1038/s41467-018-07040-0.
(29) Nothias, L.-F.; Nothias-Esposito, M.; da Silva, R.; Wang, M.; Protsyuk, I.; Zhang, Z.; Sarvepalli, A.; Leyssen, P.; Touboul, D.; Costa, J.; et al. Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. *Journal of Natural Products* **2018**, *81* (4), 758-767. DOI: 10.1021/acs.jnatprod.7b00737.
(30) Gaudêncio, S. P.; Bayram, E.; Lukić Bilela, L.; Cueto, M.; Díaz-Marrero, A. R.; Haznedaroglu, B. Z.; Jimenez, C.; Mandalakis, M.; Pereira, F.; Reyes, F.; et al. Advanced Methods for Natural Products Discovery: Bioactivity Screening, Dereplication, Metabolomics Profiling, Genomic Sequencing, Databases and Informatic Tools, and Structure Elucidation. *Marine Drugs* **2023**, *21* (5), 308.

(31) Ziemert, N.; Alanjary, M.; Weber, T. The evolution of genome mining in microbes – a review. *Natural Product Reports* **2016**, *33* (8), 988-1005, 10.1039/C6NP00025H. DOI: 10.1039/C6NP00025H.
(32) Schulze, C. J.; Donia, M. S.; Siqueira-Neto, J. L.; Ray, D.; Raskatov, J. A.; Green, R. E.; McKerrow, J. H.; Fischbach, M. A.; Linington, R. G. Genome-Directed Lead Discovery: Biosynthesis, Structure Elucidation, and Biological Evaluation of Two Families of Polyene Macrolactams against Trypanosoma brucei. *ACS Chemical Biology* **2015**, *10* (10), 2373-2381. DOI: 10.1021/acschembio.5b00308.
(33) Hemmerling, F.; Piel, J. Strategies to access biosynthetic novelty in bacterial genomes for drug discovery. *Nature Reviews Drug Discovery* **2022**, *21* (5), 359-378. DOI: 10.1038/s41573-022-00414-6.
(34) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angewandte Chemie International Edition* **2009**, *48* (26), 4688-4716. DOI: 10.1002/anie.200806121.
(35) Dan, Q.; Chiu, Y.; Lee, N.; Pereira, J. H.; Rad, B.; Zhao, X.; Deng, K.; Rong, Y.; Zhan, C.; Chen, Y.; et al. A polyketide-based biosynthetic platform for diols, amino alcohols and hydroxy acids. *Nature Catalysis* **2025**, *8* (2), 147-161. DOI: 10.1038/s41929-025-01299-5.
(36) Ridley, C. P.; Lee, H. Y.; Khosla, C. Evolution of polyketide synthases in bacteria. *Proceedings of the National Academy of Sciences* **2008**, *105* (12), 4595-4600. DOI: doi:10.1073/pnas.0710107105.
(37) Tsukamoto, N.; Chuck, J.-A.; Luo, G.; Kao, C. M.; Khosla, C.; Cane, D. E. 6-Deoxyerythronolide B Synthase 1 Is Specifically Acylated by a Diketide Intermediate at the β-Ketoacyl-Acyl Carrier Protein Synthase Domain of Module 2. *Biochemistry* **1996**, *35* (48), 15244-15248. DOI: 10.1021/bi961972f.
(38) Helfrich, E. J. N.; Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural Product Reports* **2016**, *33* (2), 231-316, 10.1039/C5NP00125K. DOI: 10.1039/C5NP00125K.
(39) Helfrich, E. J. N.; Ueoka, R.; Chevrette, M. G.; Hemmerling, F.; Lu, X.; Leopold-Messer, S.; Minas, H. A.; Burch, A. Y.; Lindow, S. E.; Piel, J.; et al. Evolution of combinatorial diversity in trans-acyltransferase polyketide synthase assembly lines across bacteria. *Nature Communications* **2021**, *12* (1), 1422. DOI: 10.1038/s41467-021-21163-x.
(40) Chen, H.; Du, L. Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Applied Microbiology and Biotechnology* **2016**, *100* (2), 541-557. DOI: 10.1007/s00253-015-7093-0.
(41) Crawford, J. M.; Dancy, B. C. R.; Hill, E. A.; Udwary, D. W.; Townsend, C. A. Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proceedings of the National Academy of Sciences* **2006**, *103* (45), 16728-16733. DOI: doi:10.1073/pnas.0604112103.
(42) Parascandolo, J. S.; Havemann, J.; Potter, H. K.; Huang, F.; Riva, E.; Connolly, J.; Wilkening, I.; Song, L.; Leadlay, P. F.; Tosin, M. Insights into 6-Methylsalicylic Acid Bio-assembly by Using Chemical Probes. *Angewandte Chemie* **2016**, *128* (10), 3524-3528.
(43) Grininger, M. The role of the iterative modules in polyketide synthase evolution. *Proceedings of the National Academy of Sciences* **2020**, *117* (16), 8680-8682. DOI: doi:10.1073/pnas.2004190117.
(44) Wang, J.; Deng, Z.; Liang, J.; Wang, Z. Structural enzymology of iterative type I polyketide synthases: various routes to catalytic programming. *Natural Product Reports* **2023**, *40* (9), 1498-1520, 10.1039/D3NP00015J. DOI: 10.1039/D3NP00015J.
(45) Wang, B.; Guo, F.; Huang, C.; Zhao, H. Unraveling the iterative type I polyketide synthases hidden in Streptomyces. *Proceedings of the National Academy of Sciences* **2020**, *117* (15), 8449-8454. DOI: doi:10.1073/pnas.1917664117.
(46) Weitnauer, G.; Mühlenweg, A.; Trefzer, A.; Hoffmeister, D.; Süßmuth, R. D.; Jung, G.; Welzel, K.; Vente, A.; Girreser, U.; Bechthold, A. Biosynthesis of the orthosomycin antibiotic avilamycin A: deductions from the molecular analysis of the avi biosynthetic gene cluster of Streptomyces viridochromogenes Tü57 and production of new antibiotics. *Chemistry & Biology* **2001**, *8* (6), 569-581. DOI: 10.1016/S1074-5521(01)00040-0.

(47) Daum , M.; Peintner , I.; Linnenbrink, A.; Frerich, A.; Weber, M.; Paululat, T.; Bechthold, A. Organisation of the Biosynthetic Gene Cluster and Tailoring Enzymes in the Biosynthesis of the Tetracyclic Quinone Glycoside Antibiotic Polyketomycin. *ChemBioChem* **2009**, *10* (6), 1073-1083. DOI: 10.1002/cbic.200800823.
(48) Zhang, Q.; Pang, B.; Ding, W.; Liu, W. Aromatic Polyketides Produced by Bacterial Iterative Type I Polyketide Synthases. *ACS Catalysis* **2013**, *3* (7), 1439-1447. DOI: 10.1021/cs400211x.
(49) Kaulmann, U.; Hertweck, C. Biosynthesis of Polyunsaturated Fatty Acids by Polyketide Synthases. *Angewandte Chemie International Edition* **2002**, *41* (11), 1866-1869. DOI: 10.1002/1521-3773(20020603)41:11<1866::AID-ANIE1866>3.0.CO;2-3.
(50) Blodgett, J. A. V.; Oh, D.-C.; Cao, S.; Currie, C. R.; Kolter, R.; Clardy, J. Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proceedings of the National Academy of Sciences* **2010**, *107* (26), 11692-11697. DOI: doi:10.1073/pnas.1001513107.
(51) Fisch, K. M. Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. *RSC Advances* **2013**, *3* (40), 18228-18247, 10.1039/C3RA42661K. DOI: 10.1039/C3RA42661K.
(52) Liu, W.; Christenson, S. D.; Standage, S.; Shen, B. Biosynthesis of the Enediyne Antitumor Antibiotic C-1027. *Science* **2002**, *297* (5584), 1170-1173. DOI: doi:10.1126/science.1072110.
(53) Ahlert, J.; Shepard, E.; Lomovskaya, N.; Zazopoulos, E.; Staffa, A.; Bachmann, B. O.; Huang, K.; Fonstein, L.; Czisny, A.; Whitwam, R. E.; et al. The Calicheamicin Gene Cluster and Its Iterative Type I Enediyne PKS. *Science* **2002**, *297* (5584), 1173-1176. DOI: doi:10.1126/science.1072105.
(54) Musiol, E. M.; Weber, T. Discrete acyltransferases involved in polyketide biosynthesis. *MedChemComm* **2012**, *3* (8), 871-886, 10.1039/C2MD20048A. DOI: 10.1039/C2MD20048A.
(55) Xie, S.; Zhang, L. Type II Polyketide Synthases: A Bioinformatics-Driven Approach. *ChemBioChem* **2023**, *24* (9), e202200775. DOI: 10.1002/cbic.202200775.
(56) Shimizu, Y.; Ogata, H.; Goto, S. Type III Polyketide Synthases: Functional Classification and Phylogenomics. *ChemBioChem* **2017**, *18* (1), 50-65. DOI: 10.1002/cbic.201600522.
(57) Funa, N.; Ozawa, H.; Hirata, A.; Horinouchi, S. Phenolic lipid synthesis by type III polyketide synthases is essential for cyst formation in Azotobacter vinelandii. *Proceedings of the National Academy of Sciences* **2006**, *103* (16), 6356-6361. DOI: doi:10.1073/pnas.0511227103.
(58) Nakano, C.; Ozawa, H.; Akanuma, G.; Funa, N.; Horinouchi, S. Biosynthesis of Aliphatic Polyketides by Type III Polyketide Synthase and Methyltransferase in Bacillus subtilis. *Journal of Bacteriology* **2009**, *191* (15), 4916-4923. DOI: doi:10.1128/jb.00407-09.
(59) Milke, L.; Kabuu, M.; Zschoche, R.; Gätgens, J.; Krumbach, K.; Carlstedt, K.-L.; Wurzbacher, C. E.; Balluff, S.; Beemelmanns, C.; Jogler, C.; et al. A type III polyketide synthase cluster in the phylum Planctomycetota is involved in alkylresorcinol biosynthesis. *Applied Microbiology and Biotechnology* **2024**, *108* (1), 239. DOI: 10.1007/s00253-024-13065-x.
(60) Sieber, S. A.; Marahiel, M. A. Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. *Chemical Reviews* **2005**, *105* (2), 715-738. DOI: 10.1021/cr0301191.
(61) Süssmuth, R. D.; Mainz, A. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angewandte Chemie International Edition* **2017**, *56* (14), 3770-3821. DOI: 10.1002/anie.201609079.
(62) Niquille, D. L.; Folger, I. B.; Basler, S.; Hilvert, D. Biosynthetic Functionalization of Nonribosomal Peptides. *Journal of the American Chemical Society* **2021**, *143* (7), 2736-2740. DOI: 10.1021/jacs.1c00925.
(63) Bozhüyük, K. A. J.; Präve, L.; Kegler, C.; Schenk, L.; Kaiser, S.; Schelhas, C.; Shi, Y.-N.; Kuttenlochner, W.; Schreiber, M.; Kandler, J.; et al. Evolution-inspired engineering of nonribosomal peptide synthetases. *Science* **2024**, *383* (6689), eadg4320. DOI: doi:10.1126/science.adg4320.

(64) Reimer, J. M.; Eivaskhani, M.; Harb, I.; Guarné, A.; Weigt, M.; Schmeing, T. M. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **2019**, *366* (6466), eaaw4388. DOI: doi:10.1126/science.aaw4388.
(65) Degen, A.; Mayerthaler, F.; Mootz, H. D.; Di Ventura, B. Context-dependent activity of A domains in the tyrocidine synthetase. *Scientific Reports* **2019**, *9* (1), 5119. DOI: 10.1038/s41598-019-41492-8.
(66) Bonhomme, S.; Contreras-Martel, C.; Dessen, A.; Macheboeuf, P. Architecture of a PKS-NRPS hybrid megaenzyme involved in the biosynthesis of the genotoxin colibactin. *Structure* **2023**, *31* (6), 700-712.e704. DOI: 10.1016/j.str.2023.03.012.
(67) Ishikawa, F.; Nakamura, S.; Nakanishi, I.; Tanabe, G. Recent progress in the reprogramming of nonribosomal peptide synthetases. *Journal of Peptide Science* **2024**, *30* (3), e3545. DOI: 10.1002/psc.3545.
(68) Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proceedings of the National Academy of Sciences* **2014**, *111* (25), 9259-9264. DOI: doi:10.1073/pnas.1401734111.
(69) Royer, M.; Koebnik, R.; Marguerettaz, M.; Barbe, V.; Robin, G. P.; Brin, C.; Carrere, S.; Gomez, C.; Hügelland, M.; Völler, G. H.; et al. Genome mining reveals the genus Xanthomonas to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. *BMC Genomics* **2013**, *14* (1), 658. DOI: 10.1186/1471-2164-14-658.
(70) Kittilä, T.; Mollo, A.; Charkoudian, L. K.; Cryle, M. J. New Structural Data Reveal the Motion of Carrier Proteins in Nonribosomal Peptide Synthesis. *Angewandte Chemie International Edition* **2016**, *55* (34), 9834-9840. DOI: 10.1002/anie.201602614.
(71) Du, L.; Shen, B. Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer Streptomyces verticillus ATCC 15003. *Chemistry & Biology* **1999**, *6* (8), 507-517. DOI: 10.1016/S1074-5521(99)80083-0.
(72) Li, S.; Wu, X.; Zhang, L.; Shen, Y.; Du, L. Activation of a Cryptic Gene Cluster in Lysobacter enzymogenes Reveals a Module/Domain Portable Mechanism of Nonribosomal Peptide Synthetases in the Biosynthesis of Pyrrolopyrazines. *Organic Letters* **2017**, *19* (19), 5010-5013. DOI: 10.1021/acs.orglett.7b01611.
(73) Blin, K.; Shaw, S.; Augustijn, H. E.; Reitz, Z. L.; Biermann, F.; Alanjary, M.; Fetter, A.; Terlouw, B. R.; Metcalf, W. W.; Helfrich, E. J. N.; et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research* **2023**, *51* (W1), W46-W50. DOI: 10.1093/nar/gkad344 (acccessed 3/4/2026).
(74) 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 Research* **2011**, *39* (suppl_2), W362-W367. DOI: 10.1093/nar/gkr323 (acccessed 3/26/2025).
(75) Mongia, M.; Baral, R.; Adduri, A.; Yan, D.; Liu, Y.; Bian, Y.; Kim, P.; Behsaz, B.; Mohimani, H. AdenPredictor: accurate prediction of the adenylation domain specificity of nonribosomal peptide biosynthetic gene clusters in microbial genomes. *Bioinformatics* **2023**, *39* (Supplement_1), i40-i46. DOI: 10.1093/bioinformatics/btad235 (acccessed 3/26/2025).
(76) Medema, M. H.; Fischbach, M. A. Computational approaches to natural product discovery. *Nature Chemical Biology* **2015**, *11* (9), 639-648. DOI: 10.1038/nchembio.1884.
(77) Behsaz, B.; Bode, E.; Gurevich, A.; Shi, Y.-N.; Grundmann, F.; Acharya, D.; Caraballo-Rodríguez, A. M.; Bouslimani, A.; Panitchpakdi, M.; Linck, A.; et al. Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery. *Nature Communications* **2021**, *12* (1), 3225. DOI: 10.1038/s41467-021-23502-4.
(78) Jangra, M.; Travin, D. Y.; Aleksandrova, E. V.; Kaur, M.; Darwish, L.; Koteva, K.; Klepacki, D.; Wang, W.; Tiffany, M.; Sokaribo, A.; et al. A broad-spectrum lasso peptide antibiotic targeting the bacterial ribosome. *Nature* **2025**. DOI: 10.1038/s41586-025-08723-7.
(79) Xu, M.; Zhang, F.; Cheng, Z.; Bashiri, G.; Wang, J.; Hong, J.; Wang, Y.; Xu, L.; Chen, X.; Huang, S.-X.; et al. Functional Genome Mining Reveals a Class V Lanthipeptide Containing a d-Amino Acid Introduced by an F420H2-Dependent Reductase. *Angewandte Chemie International Edition* **2020**, *59* (41), 18029-18035. DOI: 10.1002/anie.202008035.

(80) Richard, A. S.; Zhang, A.; Park, S.-J.; Farzan, M.; Zong, M.; Choe, H. Virion-associated phosphatidylethanolamine promotes TIM1-mediated infection by Ebola, dengue, and West Nile viruses. *Proceedings of the National Academy of Sciences* **2015**, *112* (47), 14682-14687. DOI: doi:10.1073/pnas.1508095112.
(81) 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. *Natural Product Reports* **2013**, *30* (1), 108-160, 10.1039/C2NP20085F. DOI: 10.1039/C2NP20085F.
(82) Nguyen, D. T.; Mitchell, D. A.; van der Donk, W. A. Genome Mining for New Enzyme Chemistry. *ACS Catalysis* **2024**, *14* (7), 4536-4553. DOI: 10.1021/acscatal.3c06322.
(83) Russell, A. H.; Truman, A. W. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Computational and Structural Biotechnology Journal* **2020**, *18*, 1838-1851. DOI: 10.1016/j.csbj.2020.06.032.
(84) Li, Y.; Ma, Y.; Xia, Y.; Zhang, T.; Sun, S.; Gao, J.; Yao, H.; Wang, H. Discovery and biosynthesis of tricyclic copper-binding ribosomal peptides containing histidine-to-butyrine crosslinks. *Nature Communications* **2023**, *14* (1), 2944. DOI: 10.1038/s41467-023-38517-2.
(85) Ren, H.; Dommaraju, S. R.; Huang, C.; Cui, H.; Pan, Y.; Nesic, M.; Zhu, L.; Sarlah, D.; Mitchell, D. A.; Zhao, H. Genome mining unveils a class of ribosomal peptides with two amino termini. *Nature Communications* **2023**, *14* (1), 1624. DOI: 10.1038/s41467-023-37287-1.
(86) Tao, H.; Lauterbach, L.; Bian, G.; Chen, R.; Hou, A.; Mori, T.; Cheng, S.; Hu, B.; Lu, L.; Mu, X.; et al. Discovery of non-squalene triterpenes. *Nature* **2022**, *606* (7913), 414-419. DOI: 10.1038/s41586-022-04773-3.
(87) Whitehead, J. N.; Leferink, N. G. H.; Johannissen, L. O.; Hay, S.; Scrutton, N. S. Decoding Catalysis by Terpene Synthases. *ACS Catalysis* **2023**, *13* (19), 12774-12802. DOI: 10.1021/acscatal.3c03047.
(88) Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H. Terpene synthases are widely distributed in bacteria. *Proceedings of the National Academy of Sciences* **2015**, *112* (3), 857-862. DOI: doi:10.1073/pnas.1422108112.
(89) Yamada, Y.; Arima, S.; Nagamitsu, T.; Johmoto, K.; Uekusa, H.; Eguchi, T.; Shin-ya, K.; Cane, D. E.; Ikeda, H. Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in an engineered Streptomyces host. *The Journal of Antibiotics* **2015**, *68* (6), 385-394. DOI: 10.1038/ja.2014.171.
(90) Ferreira de Freitas, R.; Schapira, M. A systematic analysis of atomic protein–ligand interactions in the PDB. *MedChemComm* **2017**, *8* (10), 1970-1981, 10.1039/C7MD00381A. DOI: 10.1039/C7MD00381A.
(91) Harris, C. M.; Kannan, R.; Kopecka, H.; Harris, T. M. The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. *Journal of the American Chemical Society* **1985**, *107* (23), 6652-6658. DOI: 10.1021/ja00309a038.
(92) Bister, B.; Bischoff, D.; Nicholson, G. J.; Stockert, S.; Wink, J.; Brunati, C.; Donadio, S.; Pelzer, S.; Wohlleben, W.; Süssmuth, R. D. Bromobalhimycin and Chlorobromobalhimycins—Illuminating the Potential of Halogenases in Glycopeptide Antibiotic Biosyntheses. *ChemBioChem* **2003**, *4* (7), 658-662. DOI: 10.1016/j.csbj.2020.06.032.
(93) Latham, J.; Brandenburger, E.; Shepherd, S. A.; Menon, B. R. K.; Micklefield, J. Development of Halogenase Enzymes for Use in Synthesis. *Chemical Reviews* **2018**, *118* (1), 232-269. DOI: 10.1021/acs.chemrev.7b00032.
(94) Hornung, A.; Bertazzo, M.; Dziarnowski, A.; Schneider, K.; Welzel, K.; Wohlert, S.-E.; Holzenkämpfer, M.; Nicholson, G. J.; Bechthold, A.; Süssmuth, R. D.; et al. A Genomic Screening Approach to the Structure-Guided Identification of Drug Candidates from Natural Sources. *ChemBioChem* **2007**, *8* (7), 757-766. DOI: 10.1002/cbic.200600375.
(95) Saha, N.; Vidya, F. N. U.; Luo, Y.; van der Donk, W. A.; Agarwal, V. Transformation-Guided Genome Mining Provides Access to Brominated Lanthipeptides. *Organic Letters* **2025**, *27* (4), 984-988. DOI: 10.1021/acs.orglett.4c04529.

(96) Dong, C.; Huang, F.; Deng, H.; Schaffrath, C.; Spencer, J. B.; O'Hagan, D.; Naismith, J. H. Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* **2004**, *427* (6974), 561-565. DOI: 10.1038/nature02280.
(97) Deng, H.; Ma, L.; Bandaranayaka, N.; Qin, Z.; Mann, G.; Kyeremeh, K.; Yu, Y.; Shepherd, T.; Naismith, J. H.; O'Hagan, D. Identification of Fluorinases from Streptomyces sp MA37, Norcardia brasiliensis, and Actinoplanes sp N902-109 by Genome Mining. *ChemBioChem* **2014**, *15* (3), 364-368. DOI: 10.1002/cbic.201300732.
(98) Scherlach, K.; Hertweck, C. Mining and unearthing hidden biosynthetic potential. *Nature Communications* **2021**, *12* (1), 3864. DOI: 10.1038/s41467-021-24133-5.
(99) Yan, Y.; Liu, N.; Tang, Y. Recent developments in self-resistance gene directed natural product discovery. *Natural Product Reports* **2020**, *37* (7), 879-892, 10.1039/C9NP00050J. DOI: 10.1039/C9NP00050J.
(100) Zhao, Y.; Xu, G.; Xu, Z.; Guo, B.; Liu, F. LexR Positively Regulates the LexABC Efflux Pump Involved in Self-Resistance to the Antimicrobial Di-N-Oxide Phenazine in Lysobacter antibioticus. *Microbiology Spectrum* **2023**, *11* (3), e04872-04822. DOI: doi:10.1128/spectrum.04872-22.
(101) Steffensky, M.; Mühlenweg, A.; Wang, Z.-X.; Li, S.-M.; Heide, L. Identification of the Novobiocin Biosynthetic Gene Cluster of Streptomyces spheroides NCIB 11891. *Antimicrobial Agents and Chemotherapy* **2000**, *44* (5), 1214-1222. DOI: doi:10.1128/aac.44.5.1214-1222.2000.
(102) Peterson, Ryan M.; Huang, T.; Rudolf, Jeffrey D.; Smanski, Michael J.; Shen, B. Mechanisms of Self-Resistance in the Platensimycin- and Platencin-Producing Streptomyces platensis MA7327 and MA7339 Strains. *Chemistry & Biology* **2014**, *21* (3), 389-397. DOI: 10.1016/j.chembiol.2014.01.005.
(103) Thaker, M. N.; Wang, W.; Spanogiannopoulos, P.; Waglechner, N.; King, A. M.; Medina, R.; Wright, G. D. Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nature Biotechnology* **2013**, *31* (10), 922-927. DOI: 10.1038/nbt.2685.
(104) Yuan, Y.; Huang, C.; Singh, N.; Xun, G.; Zhao, H. Self-resistance-gene-guided, high-throughput automated genome mining of bioactive natural products from Streptomyces. *Cell Systems* **2025**, *16* (3). DOI: 10.1016/j.cels.2025.101237 (acccessed 2025/04/15).
(105) Tang, X.; Li, J.; Millán-Aguiñaga, N.; Zhang, J. J.; O'Neill, E. C.; Ugalde, J. A.; Jensen, P. R.; Mantovani, S. M.; Moore, B. S. Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chemical Biology* **2015**, *10* (12), 2841-2849. DOI: 10.1021/acschembio.5b00658.
(106) Alanjary, M.; Kronmiller, B.; Adamek, M.; Blin, K.; Weber, T.; Huson, D.; Philmus, B.; Ziemert, N. The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Research* **2017**, *45* (W1), W42-W48. DOI: 10.1093/nar/gkx360 (acccessed 4/15/2025).
(107) Brötz-Oesterhelt, H.; Beyer, D.; Kroll, H.-P.; Endermann, R.; Ladel, C.; Schroeder, W.; Hinzen, B.; Raddatz, S.; Paulsen, H.; Henninger, K.; et al. Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nature Medicine* **2005**, *11* (10), 1082-1087. DOI: 10.1038/nm1306.
(108) Pahl, A.; Lakemeyer, M.; Vielberg, M.-T.; Hackl, M. W.; Vomacka, J.; Korotkov, V. S.; Stein, M. L.; Fetzer, C.; Lorenz-Baath, K.; Richter, K.; et al. Reversible Inhibitors Arrest ClpP in a Defined Conformational State that Can Be Revoked by ClpX Association. *Angewandte Chemie International Edition* **2015**, *54* (52), 15892-15896. DOI: 10.1002/anie.201507266.
(109) Raju, R. M.; Goldberg, A. L.; Rubin, E. J. Bacterial proteolytic complexes as therapeutic targets. *Nature Reviews Drug Discovery* **2012**, *11* (10), 777-789. DOI: 10.1038/nrd3846.
(110) Felix, J.; Weinhäupl, K.; Chipot, C.; Dehez, F.; Hessel, A.; Gauto, D. F.; Morlot, C.; Abian, O.; Gutsche, I.; Velazquez-Campoy, A.; et al. Mechanism of the allosteric activation of the ClpP protease machinery by substrates and active-site inhibitors. *Science Advances* **2019**, *5* (9), eaaw3818. DOI: doi:10.1126/sciadv.aaw3818.

(111) Gatsogiannis, C.; Balogh, D.; Merino, F.; Sieber, S. A.; Raunser, S. Cryo-EM structure of the ClpXP protein degradation machinery. *Nature Structural & Molecular Biology* **2019**, *26* (10), 946-954. DOI: 10.1038/s41594-019-0304-0.
(112) Brötz-Oesterhelt, H.; Sass, P. Bacterial caseinolytic proteases as novel targets for antibacterial treatment. *International Journal of Medical Microbiology* **2014**, *304* (1), 23-30. DOI: 10.1016/j.ijmm.2013.09.001.
(113) Culp, E. J.; Sychantha, D.; Hobson, C.; Pawlowski, A. C.; Prehna, G.; Wright, G. D. ClpP inhibitors are produced by a widespread family of bacterial gene clusters. *Nature Microbiology* **2022**, *7* (3), 451-462. DOI: 10.1038/s41564-022-01073-4.
(114) Adamek, M.; Alanjary, M.; Ziemert, N. Applied evolution: phylogeny-based approaches in natural products research. *Natural Product Reports* **2019**, *36* (9), 1295-1312, 10.1039/C9NP00027E. DOI: 10.1039/C9NP00027E.
(115) Kang, H.-S.; Brady, S. F. Arixanthomycins A–C: Phylogeny-Guided Discovery of Biologically Active eDNA-Derived Pentangular Polyphenols. *ACS Chemical Biology* **2014**, *9* (6), 1267-1272. DOI: 10.1021/cb500141b.
(116) Deng, Q.; Li, Y.; He, W.; Chen, T.; Liu, N.; Ma, L.; Qiu, Z.; Shang, Z.; Wang, Z. A polyene macrolide targeting phospholipids in the fungal cell membrane. *Nature* **2025**, *640* (8059), 743-751. DOI: 10.1038/s41586-025-08678-9.
(117) Yee, D. A.; Niwa, K.; Perlatti, B.; Chen, M.; Li, Y.; Tang, Y. Genome mining for unknown–unknown natural products. *Nature Chemical Biology* **2023**, *19* (5), 633-640. DOI: 10.1038/s41589-022-01246-6.
(118) Yuliana, N. D.; Khatib, A.; Choi, Y. H.; Verpoorte, R. Metabolomics for bioactivity assessment of natural products. *Phytotherapy Research* **2011**, *25* (2), 157-169. DOI: 10.1002/ptr.3258.
(119) Vitale, G. A.; Geibel, C.; Minda, V.; Wang, M.; Aron, A. T.; Petras, D. Connecting metabolome and phenotype: recent advances in functional metabolomics tools for the identification of bioactive natural products. *Natural Product Reports* **2024**, *41* (6), 885-904, 10.1039/D3NP00050H. DOI: 10.1039/D3NP00050H.
(120) Caesar, L. K.; Montaser, R.; Keller, N. P.; Kelleher, N. L. Metabolomics and genomics in natural products research: complementary tools for targeting new chemical entities. *Natural Product Reports* **2021**, *38* (11), 2041-2065, 10.1039/D1NP00036E. DOI: 10.1039/D1NP00036E.
(121) Zhang, H.-W.; Lv, C.; Zhang, L.-J.; Guo, X.; Shen, Y.-W.; Nagle, D. G.; Zhou, Y.-D.; Liu, S.-H.; Zhang, W.-D.; Luan, X. Application of omics- and multi-omics-based techniques for natural product target discovery. *Biomedicine & Pharmacotherapy* **2021**, *141*, 111833. DOI: 10.1016/j.biopha.2021.111833.
(122) Courant, F.; Antignac, J.-P.; Dervilly-Pinel, G.; Le Bizec, B. Basics of mass spectrometry based metabolomics. *PROTEOMICS* **2014**, *14* (21-22), 2369-2388. DOI: 10.1002/pmic.201400255.
(123) Sarkar, J.; Singh, R.; Chandel, S. Understanding LC/MS-Based Metabolomics: A Detailed Reference for Natural Product Analysis. *PROTEOMICS – Clinical Applications* **2025**, *19* (1), e202400048. DOI: 10.1002/prca.202400048.
(124) Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W. LC-MS-based metabolomics. *Molecular BioSystems* **2012**, *8* (2), 470-481, 10.1039/C1MB05350G. DOI: 10.1039/C1MB05350G.
(125) Alseekh, S.; Aharoni, A.; Brotman, Y.; Contrepois, K.; D'Auria, J.; Ewald, J.; C. Ewald, J.; Fraser, P. D.; Giavalisco, P.; Hall, R. D.; et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nature Methods* **2021**, *18* (7), 747-756. DOI: 10.1038/s41592-021-01197-1.
(126) 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 Central Science* **2019**, *5* (11), 1824-1833. DOI: 10.1021/acscentsci.9b00806.
(127) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; et al. Mass spectral molecular

networking of living microbial colonies. *Proceedings of the National Academy of Sciences* **2012**, *109* (26), E1743-E1752. DOI: doi:10.1073/pnas.1203689109.
(128) Quinn, R. A.; Nothias, L.-F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P. C. Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends in Pharmacological Sciences* **2017**, *38* (2), 143-154. DOI: 10.1016/j.tips.2016.10.011 (acccessed 2025/02/28).
(129) 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. *Nature Biotechnology* **2016**, *34* (8), 828-837. DOI: 10.1038/nbt.3597.
(130) Choi, M.; Carver, J.; Chiva, C.; Tzouros, M.; Huang, T.; Tsai, T.-H.; Pullman, B.; Bernhardt, O. M.; Hüttenhain, R.; Teo, G. C.; et al. MassIVE.quant: a community resource of quantitative mass spectrometry–based proteomics datasets. *Nature Methods* **2020**, *17* (10), 981-984. DOI: 10.1038/s41592-020-0955-0.
(131) Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J. M.; Sikora, N.; Vargas, F.; et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature Protocols* **2020**, *15* (6), 1954-1991. DOI: 10.1038/s41596-020-0317-5.
(132) Leao, T. F.; Clark, C. M.; Bauermeister, A.; Elijah, E. O.; Gentry, E. C.; Husband, M.; Oliveira, M. F.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Quick-start infrastructure for untargeted metabolomics analysis in GNPS. *Nature Metabolism* **2021**, *3* (7), 880-882. DOI: 10.1038/s42255-021-00429-0.
(133) 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. *Nature Methods* **2020**, *17* (9), 905-908. DOI: 10.1038/s41592-020-0933-6.
(134) Forseth, R. R.; Schroeder, F. C. NMR-spectroscopic analysis of mixtures: from structure to function. *Current Opinion in Chemical Biology* **2011**, *15* (1), 38-47. DOI: 10.1016/j.cbpa.2010.10.010.
(135) Nagana Gowda, G. A.; Raftery, D. Recent Advances in NMR-Based Metabolomics. *Analytical Chemistry* **2017**, *89* (1), 490-510. DOI: 10.1021/acs.analchem.6b04420.
(136) Moco, S. Studying Metabolism by NMR-Based Metabolomics. *Frontiers in Molecular Biosciences* **2022**, *Volume 9 - 2022*, Review. DOI: 10.3389/fmolb.2022.882487.
(137) Wang, D.-G.; Hu, J.-Q.; Wang, C.-Y.; Liu, T.; Li, Y.-Z.; Wu, C. Exploring microbial natural products through NMR-based metabolomics. *Natural Product Reports* **2025**, *42* (9), 1459-1488, 10.1039/D4NP00065J. DOI: 10.1039/D4NP00065J.
(138) Miller, W. R.; Arias, C. A. ESKAPE pathogens: antimicrobial resistance, epidemiology, clinical impact and therapeutics. *Nature Reviews Microbiology* **2024**, *22* (10), 598-616. DOI: 10.1038/s41579-024-01054-w.
(139) Naghavi, M.; Vollset, S. E.; Ikuta, K. S.; Swetschinski, L. R.; Gray, A. P.; Wool, E. E.; Robles Aguilar, G.; Mestrovic, T.; Smith, G.; Han, C.; et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *The Lancet* **2024**, *404* (10459), 1199-1226. DOI: 10.1016/S0140-6736(24)01867-1 (acccessed 2025/05/27).
(140) Silver, L. L. Fosfomycin: mechanism and resistance. *Cold Spring Harbor perspectives in medicine* **2017**, *7* (2), a025262.
(141) Voráčová, M.; Zore, M.; Yli-Kauhaluoma, J.; Kiuru, P. Harvesting phosphorus-containing moieties for their antibacterial effects. *Bioorganic & Medicinal Chemistry* **2023**, *96*, 117512. DOI: 10.1016/j.bmc.2023.117512.
(142) Cui, J. J.; Zhang, Y.; Ju, K.-S. Phosphonoalamides Reveal the Biosynthetic Origin of Phosphonoalanine Natural Products and a Convergent Pathway for Their Diversification. *Angewandte Chemie International Edition* **2024**, *63* (32), e202405052. DOI: 10.1002/anie.202405052.
(143) Ju, K.-S.; Doroghazi, J. R.; Metcalf, W. W. Genomics-enabled discovery of phosphonate natural products and their biosynthetic pathways. *Journal of Industrial Microbiology and*

Biotechnology **2014**, *41* (2), 345-356. DOI: 10.1007/s10295-013-1375-2 (acccessed 11/15/2024).
(144) Metcalf, W. W.; van der Donk, W. A. Biosynthesis of Phosphonic and Phosphinic Acid Natural Products. *Annual Review of Biochemistry* **2009**, *78* (Volume 78, 2009), 65-94. DOI:10.1146/annurev.biochem.78.091707.100215.
(145) Cioni, J. P.; Doroghazi, J. R.; Ju, K.-S.; Yu, X.; Evans, B. S.; Lee, J.; Metcalf, W. W. Cyanohydrin Phosphonate Natural Product from Streptomyces regensis. *Journal of Natural Products* **2014**, *77* (2), 243-249. DOI: 10.1021/np400722m.
(146) Ju, K.-S.; Gao, J.; Doroghazi, J. R.; Wang, K.-K. A.; Thibodeaux, C. J.; Li, S.; Metzger, E.; Fudala, J.; Su, J.; Zhang, J. K.; et al. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proceedings of the National Academy of Sciences* **2015**, *112* (39), 12175-12180. DOI: doi:10.1073/pnas.1500873112.
(147) Yu, X.; Doroghazi, J. R.; Janga, S. C.; Zhang, J. K.; Circello, B.; Griffin, B. M.; Labeda, D. P.; Metcalf, W. W. Diversity and abundance of phosphonate biosynthetic genes in nature. *Proceedings of the National Academy of Sciences* **2013**, *110* (51), 20759-20764. DOI: doi:10.1073/pnas.1315107110.
(148) Seidel, H. M.; Freeman, S.; Seto, H.; Knowles, J. R. Phosphonate biosynthesis: isolation of the enzyme responsible for the formation of a carbon–phosphorus bond. *Nature* **1988**, *335* (6189), 457-458. DOI: 10.1038/335457a0.
(149) Zimmermann, A.; Xia, S.-N.; Moschny, J.; Gomez-Escribano, J. P.; Boldt, J.; Nübel, U.; Nouioui, I.; Krause, J.; Irle, M. K.; Metcalf, W. W.; et al. Expanding the actinomycetes landscape for phosphonate natural products through genome mining. *RSC Chemical Biology* **2025**, *7* (2), 298-312. DOI: 10.1039/d5cb00254k.
(150) Krysenko, S.; Wohlleben, W. Role of Carbon, Nitrogen, Phosphate and Sulfur Metabolism in Secondary Metabolism Precursor Supply in Streptomyces spp. *Microorganisms* **2024**, *12* (8), 1571.
(151) Kayrouz, C. M.; Zhang, Y.; Pham, T. M.; Ju, K.-S. Genome Mining Reveals the Phosphonoalamide Natural Products and a New Route in Phosphonic Acid Biosynthesis. *ACS Chemical Biology* **2020**, *15* (7), 1921-1929. DOI: 10.1021/acschembio.0c00256.
(152) Wilson, J.; Cui, J.; Nakao, T.; Kwok, H.; Zhang, Y.; Kayrouz, C. M.; Pham, T. M.; Roodhouse, H.; Ju, K.-S. Discovery of Antimicrobial Phosphonopeptide Natural Products from Bacillus velezensis by Genome Mining. *Applied and Environmental Microbiology* **2023**, *89* (6), e00338-00323. DOI: doi:10.1128/aem.00338-23.
(153) Nouioui, I.; Zimmermann, A.; Hennrich, O.; Xia, S.; Rössler, O.; Makitrynskyy, R.; Pablo Gomez-Escribano, J.; Pötter, G.; Jando, M.; Döppner, M.; et al. Challenging old microbiological treasures for natural compound biosynthesis capacity. *Frontiers in Bioengineering and Biotechnology* **2024**, *Volume 12 - 2024*, Original Research. DOI: 10.3389/fbioe.2024.1255151.
(154) Gomez-Escribano, J. P.; Zimmermann, A.; Xia, S.-N.; Döppner, M.; Moschny, J.; Hughes, C. C.; Mast, Y. Application of a replicative targetable vector system for difficult-to-manipulate streptomycetes. *Applied Microbiology and Biotechnology* **2025**, *109* (1), 89. DOI: 10.1007/s00253-025-13477-3.
(155) Zimmermann, A.; Nouioui, I.; Pötter, G.; Neumann-Schaal, M.; Wolf, J.; Wibberg, D.; Mast, Y. Kitasatospora fiedleri sp. nov., a novel antibiotic-producing member of the genus Kitasatospora. *International Journal of Systematic and Evolutionary Microbiology* **2023**, *73* (11). DOI: 10.1099/ijsem.0.006137.
(156) Zimmermann, A.; Xia, S.-N.; Moschny, J.; Gomez-Escribano, J. P.; Boldt, J.; Nübel, U.; Nouioui, I.; Krause, J.; Irle, M. K.; Metcalf, W. W.; et al. Expanding the actinomycetes landscape for phosphonate natural products through genome mining. *RSC Chemical Biology* **2026**, *7* (2), 298-312, 10.1039/D5CB00254K. DOI: 10.1039/D5CB00254K.
(157) Wang, S.; Sun, S.; Shan, C.; Pan, B. Analysis of trace phosphonates in authentic water samples by pre-methylation and LC-Orbitrap MS/MS. *Water Research* **2019**, *161*, 78-88. DOI: 10.1016/j.watres.2019.05.099.
(158) Knepper, T. P. Synthetic chelating agents and compounds exhibiting complexing properties in the aquatic environment. *TrAC Trends in Analytical Chemistry* **2003**, *22* (10), 708-724. DOI: 10.1016/S0165-9936(03)01008-2.

(159) Berlinck, R. G. S.; Crnkovic, C. M.; Gubiani, J. R.; Bernardi, D. I.; Ióca, L. P.; Quintana-Bulla, J. I. The isolation of water-soluble natural products – challenges, strategies and perspectives. *Natural Product Reports* **2022**, *39* (3), 596-669, 10.1039/D1NP00037C. DOI: 10.1039/D1NP00037C.
(160) Ibáñez, M.; Pozo, Ó. J.; Sancho, J. V.; López, F. J.; Hernández, F. Residue determination of glyphosate, glufosinate and aminomethylphosphonic acid in water and soil samples by liquid chromatography coupled to electrospray tandem mass spectrometry. *Journal of Chromatography A* **2005**, *1081* (2), 145-155. DOI: 10.1016/j.chroma.2005.05.041.
(161) Ehling, S.; Reddy, T. M. Analysis of Glyphosate and Aminomethylphosphonic Acid in Nutritional Ingredients and Milk by Derivatization with Fluorenylmethyloxycarbonyl Chloride and Liquid Chromatography–Mass Spectrometry. *Journal of Agricultural and Food Chemistry* **2015**, *63* (48), 10562-10568. DOI: 10.1021/acs.jafc.5b04453.
(162) Kinoshita, M.; Saito, A.; Yamamoto, S.; Suzuki, S. A practical method for preparing fluorescent-labeled glycans with a 9-fluorenylmethyl derivative to simplify a fluorimetric HPLC-based analysis. *Journal of Pharmaceutical and Biomedical Analysis* **2020**, *186*, 113267. DOI: 10.1016/j.jpba.2020.113267.
(163) Jámbor, A.; Molnár-Perl, I. Amino acid analysis by high-performance liquid chromatography after derivatization with 9-fluorenylmethyloxycarbonyl chloride: Literature overview and further study. *Journal of Chromatography A* **2009**, *1216* (15), 3064-3077. DOI: 10.1016/j.chroma.2009.01.068.
(164) Li, S.; Chen, Y.-Y.; Ye, T.-T.; Zhu, Q.-F.; Feng, Y.-Q. Chemical isotope labeling assisted liquid chromatography-mass spectrometry method for simultaneous analysis of central carbon metabolism intermediates. *Journal of Chromatography A* **2023**, *1702*, 464083. DOI: 10.1016/j.chroma.2023.464083.
(165) Jiang, H.-P.; Xiong, J.; Liu, F.-L.; Ma, C.-J.; Tang, X.-L.; Yuan, B.-F.; Feng, Y.-Q. Modified nucleoside triphosphates exist in mammals. *Chemical Science* **2018**, *9* (17), 4160-4167, 10.1039/C7SC05472F. DOI: 10.1039/C7SC05472F.
(166) Fei, N.; Sauter, B.; Gillingham, D. The pKa of Brønsted acids controls their reactivity with diazo compounds. *Chemical Communications* **2016**, *52* (47), 7501-7504, 10.1039/C6CC03561B. DOI: 10.1039/C6CC03561B.
(167) Liu, F.-L.; Qi, C.-B.; Cheng, Q.-Y.; Ding, J.-H.; Yuan, B.-F.; Feng, Y.-Q. Diazo Reagent Labeling with Mass Spectrometry Analysis for Sensitive Determination of Ribonucleotides in Living Organisms. *Analytical Chemistry* **2020**, *92* (2), 2301-2309. DOI: 10.1021/acs.analchem.9b05122.
(168) Mix, K. A.; Lomax, J. E.; Raines, R. T. Cytosolic Delivery of Proteins by Bioreversible Esterification. *Journal of the American Chemical Society* **2017**, *139* (41), 14396-14398. DOI: 10.1021/jacs.7b06597.
(169) Leung, E. M. K.; Chan, W. A novel reversed-phase HPLC method for the determination of urinary creatinine by pre-column derivatization with ethyl chloroformate: comparative studies with the standard Jaffé and isotope-dilution mass spectrometric assays. *Analytical and Bioanalytical Chemistry* **2014**, *406* (6), 1807-1812. DOI: 10.1007/s00216-013-7592-8.
(170) Opekar, S.; Kvíčala, J.; Moos, M.; Pejchal, V.; Šimek, P. Mechanism of Alkyl Chloroformate-Mediated Esterification of Carboxylic Acids in Aqueous Media. *The Journal of Organic Chemistry* **2021**, *86* (23), 16293-16299. DOI: 10.1021/acs.joc.1c01546.
(171) Kvitvang, H. F. N.; Andreassen, T.; Adam, T.; Villas-Bôas, S. G.; Bruheim, P. Highly Sensitive GC/MS/MS Method for Quantitation of Amino and Nonamino Organic Acids. *Analytical Chemistry* **2011**, *83* (7), 2705-2711. DOI: 10.1021/ac103245b.
(172) You, J.; Shan, Y.; Zhen, L.; Zhang, L.; Zhang, Y. Determination of peptides and amino acids from wool and beer with sensitive fluorescent reagent 2-(9-carbazole)-ethyl chloroformate by reverse phase high-performance liquid chromotography and liquid chromotography mass spectrometry. *Analytical Biochemistry* **2003**, *313* (1), 17-27. DOI: 10.1016/S0003-2697(02)00398-6.
(173) Bittremieux, W.; Wang, M.; Dorrestein, P. C. The critical role that spectral libraries play in capturing the metabolomics community knowledge. *Metabolomics* **2022**, *18* (12), 94. DOI: 10.1007/s11306-022-01947-y.

(174) Mahieu, N. G.; Patti, G. J. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Analytical Chemistry* **2017**, *89* (19), 10397-10406. DOI: 10.1021/acs.analchem.7b02380.
(175) Metz, T. O.; Adkins, J. N.; Armentrout, P. B.; Chain, P.; Chu, F.; Corley, C. D.; Cort, J. R.; Denis, E.; Drell, D.; Duncan, K. R. Decoding the Molecular Universe--Workshop Report. *arXiv preprint arXiv:2311.11437* **2023**.
(176) Caesar, L. K.; Butun, F. A.; Robey, M. T.; Ayon, N. J.; Gupta, R.; Dainko, D.; Bok, J. W.; Nickles, G.; Stankey, R. J.; Johnson, D.; et al. Correlative metabologenomics of 110 fungi reveals metabolite–gene cluster pairs. *Nature Chemical Biology* **2023**, *19* (7), 846-854. DOI: 10.1038/s41589-023-01276-8.
(177) Morehouse, N. J.; Clark, T. N.; McMann, E. J.; van Santen, J. A.; Haeckl, F. P. J.; Gray, C. A.; Linington, R. G. Annotation of natural product compound families using molecular networking topology and structural similarity fingerprinting. *Nature Communications* **2023**, *14* (1), 308. DOI: 10.1038/s41467-022-35734-z.
(178) Papadopoulos Lambidis, S.; Schramm, T.; Steuer-Lodd, K.; Farrell, S.; Stincone, P.; Schmid, R.; Koester, I.; Torres, R.; Dittmar, T.; Aluwihare, L.; et al. Two-Dimensional Liquid Chromatography Tandem Mass Spectrometry Untangles the Deep Metabolome of Marine Dissolved Organic Matter. *Environmental Science & Technology* **2024**, *58* (43), 19289-19304. DOI: 10.1021/acs.est.4c07173.
(179) MohammadiPeyhani, H.; Hafner, J.; Sveshnikova, A.; Viterbo, V.; Hatzimanikatis, V. Expanding biochemical knowledge and illuminating metabolic dark matter with ATLASx. *Nature Communications* **2022**, *13* (1), 1560. DOI: 10.1038/s41467-022-29238-z.
(180) Bağcı, C.; Nuhamunada, M.; Goyat, H.; Ladanyi, C.; Sehnal, L.; Blin, K.; Kautsar, Satria A.; Tagirdzhanov, A.; Gurevich, A.; Mantri, S.; et al. BGC Atlas: a web resource for exploring the global chemical diversity encoded in bacterial genomes. *Nucleic Acids Research* **2024**, *53* (D1), D618-D624. DOI: 10.1093/nar/gkae953 (acccessed 3/6/2025).
(181) Theodoridis, G.; Gika, H.; Raftery, D.; Goodacre, R.; Plumb, R. S.; Wilson, I. D. Ensuring Fact-Based Metabolite Identification in Liquid Chromatography–Mass Spectrometry-Based Metabolomics. *Analytical Chemistry* **2023**, *95* (8), 3909-3916. DOI: 10.1021/acs.analchem.2c05192.
(182) 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. *Nature Methods* **2019**, *16* (4), 299-302. DOI: 10.1038/s41592-019-0344-8.
(183) Stravs, M. A.; Dührkop, K.; Böcker, S.; Zamboni, N. MSNovelist: de novo structure generation from mass spectra. *Nature Methods* **2022**, *19* (7), 865-870. DOI: 10.1038/s41592-022-01486-3.
(184) Zhao, S.; Li, L. Chemical derivatization in LC-MS-based metabolomics study. *TrAC Trends in Analytical Chemistry* **2020**, *131*, 115988. DOI: 10.1016/j.trac.2020.115988.
(185) Hughes, C. C. Chemical labeling strategies for small molecule natural product detection and isolation. *Natural Product Reports* **2021**, *38* (9), 1684-1705, 10.1039/D0NP00034E. DOI: 10.1039/D0NP00034E.
(186) Castro-Falcón, G.; Hahn, D.; Reimer, D.; Hughes, C. C. Thiol Probes To Detect Electrophilic Natural Products Based on Their Mechanism of Action. *ACS Chemical Biology* **2016**, *11* (8), 2328-2336. DOI: 10.1021/acschembio.5b00924.
(187) Kaur, A.; Lin, W.; Dovhalyuk, V.; Driutti, L.; Di Martino, M. L.; Vujasinovic, M.; Löhr, J. M.; Sellin, M. E.; Globisch, D. Chemoselective bicyclobutane-based mass spectrometric detection of biological thiols uncovers human and bacterial metabolites. *Chemical Science* **2023**, *14* (20), 5291-5301, 10.1039/D3SC00224A. DOI: 10.1039/D3SC00224A.
(188) Lin, W.; Gerullat, L.; Braadland, P. R.; Fournier, A.; Hov, J. R.; Globisch, D. Rapid and Bifunctional Chemoselective Metabolome Analysis of Liver Disease Plasma Using the Reagent 4-Nitrophenyl-2H-azirine. *Angewandte Chemie International Edition* **2024**, *63* (14), e202318579. DOI: 10.1002/anie.202318579.
(189) Schmid, R.; Heuckeroth, S.; Korf, A.; Smirnov, A.; Myers, O.; Dyrlund, T. S.; Bushuiev, R.; Murray, K. J.; Hoffmann, N.; Lu, M.; et al. Integrative analysis of multimodal mass

spectrometry data in MZmine 3. *Nature Biotechnology* **2023**, *41* (4), 447-449. DOI: 10.1038/s41587-023-01690-2.
(190) Schmid, R.; Petras, D.; Nothias, L.-F.; Wang, M.; Aron, A. T.; Jagels, A.; Tsugawa, H.; Rainer, J.; Garcia-Aloy, M.; Dührkop, K.; et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nature Communications* **2021**, *12* (1), 3832. DOI: 10.1038/s41467-021-23953-9.
(191) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences* **2015**, *112* (41), 12580-12585. DOI: doi:10.1073/pnas.1509788112.
(192) Chen, D.; Ding, J.; Wu, M.-K.; Zhang, T.-Y.; Qi, C.-B.; Feng, Y.-Q. A liquid chromatography–mass spectrometry method based on post column derivatization for automated analysis of urinary hexanal and heptanal. *Journal of Chromatography A* **2017**, *1493*, 57-63. DOI: 10.1016/j.chroma.2017.02.071.
(193) Daylight Chemical Information Systems.
http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (accessed Nov01, 2023).
(194) Vitale, G. A.; Xia, S.-N.; Dührkop, K.; Zare Shahneh, M. R.; Brötz-Oesterhelt, H.; Mast, Y.; Brungs, C.; Böcker, S.; Schmid, R.; Wang, M.; et al. Enhancing tandem mass spectrometry-based metabolite annotation with online chemical labeling. *Nature Communications* **2025**, *16* (1), 6911. DOI: 10.1038/s41467-025-61240-z.
(195) Hoyle, C. E.; Bowman, C. N. Thiol–Ene Click Chemistry. *Angewandte Chemie International Edition* **2010**, *49* (9), 1540-1573. DOI: 10.1002/anie.200903924.
(196) Kumagai, Y.; Shinkai, Y.; Miura, T.; Cho, A. K. The Chemical Biology of Naphthoquinones and Its Environmental Implications. *Annual Review of Pharmacology and Toxicology* **2012**, *52* (Volume 52, 2012), 221-247. DOI: 10.1146/annurev-pharmtox-010611-134517.
(197) Olson, K. R.; Clear, K. J.; Gao, Y.; Ma, Z.; Cieplik, N. M.; Fiume, A. R.; Gaziano, D. J.; Kasko, S. M.; Luu, J.; Pfaff, E.; et al. Redox and Nucleophilic Reactions of Naphthoquinones with Small Thiols and Their Effects on Oxidization of H2S to Inorganic and Organic Hydropolysulfides and Thiosulfate. *International Journal of Molecular Sciences* **2023**, *24* (8), 7516.
(198) Wipf, P.; Jeger, P.; Kim, Y. Thiophilic ring-opening and rearrangement reactions of epoxyketone natural products. *Bioorganic & Medicinal Chemistry Letters* **1998**, *8* (4), 351-356. DOI: 10.1016/S0960-894X(98)00026-2.
(199) Lall, M. S.; Karvellas, C.; Vederas, J. C. β-Lactones as a New Class of Cysteine Proteinase Inhibitors: Inhibition of Hepatitis A Virus 3C Proteinase by N-Cbz-serine β-Lactone. *Organic Letters* **1999**, *1* (5), 803-806. DOI: 10.1021/ol990148r.
(200) Owen, T. C. Thiol detection, derivatization and tagging at micromole to nanomole levels using propiolates. *Bioorganic Chemistry* **2008**, *36* (3), 156-160. DOI: 10.1016/j.bioorg.2008.02.003.
(201) Busto, O.; Guasch, J.; Borrull, F. Determination of biogenic amines in wine after precolumn derivatization with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate. *Journal of Chromatography A* **1996**, *737* (2), 205-213. DOI:10.1016/0021-9673(96)00022-2.
(202) Motte, J. C.; Windey, R.; Delafortrie, A. High-sensitivity fluorescence derivatization for the determination of hydroxy compounds in aqueous solution by high-performance liquid chromatography. *Journal of Chromatography A* **1996**, *728* (1), 333-341. DOI: 10.1016/0021-9673(95)01138-2.
(203) Collins, J.; Xiao, Z.; Müllner, M.; Connal, L. A. The emergence of oxime click chemistry and its utility in polymer science. *Polymer Chemistry* **2016**, *7* (23), 3812-3826. DOI: 10.1039/c6py00635c.
(204) 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. *Nature Biotechnology* **2021**, *39* (4), 462-471. DOI: 10.1038/s41587-020-0740-8.

(205) Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics* **2015**, *7* (1), 20. DOI: 10.1186/s13321-015-0069-3.
(206) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **2010**, *50* (5), 742-754. DOI: 10.1021/ci100050t.
(207) Zhao, C.; Ju, J.; Christenson, S. D.; Smith, W. C.; Song, D.; Zhou, X.; Shen, B.; Deng, Z. Utilization of the Methoxymalonyl-Acyl Carrier Protein Biosynthesis Locus for Cloning the Oxazolomycin Biosynthetic Gene Cluster from Streptomyces albus JA3453. *Journal of Bacteriology* **2006**, *188* (11), 4142-4147. DOI: doi:10.1128/jb.00173-06.
(208) Mu, Y.; Jiang, Y.; Qu, X.; Liu, B.; Tan, J.; Li, G.; Jiang, M.; Li, L.; Han, L.; Huang, X. Oxazolomycins produced by Streptomyces glaucus and their cytotoxic activity. *RSC Advances* **2021**, *11* (55), 35011-35019, 10.1039/D1RA06182H. DOI: 10.1039/D1RA06182H.
(209) Liu, J.; Yu, H.; Li, S.-M. Expanding tryptophan-containing cyclodipeptide synthase spectrum by identification of nine members from Streptomyces strains. *Applied Microbiology and Biotechnology* **2018**, *102* (10), 4435-4444. DOI: 10.1007/s00253-018-8908-6.
(210) Singh, S. B.; Odingo, J.; Bailey, M. A.; Sunde, B.; Korkegian, A.; O'Malley, T.; Ovechkina, Y.; Ioerger, T. R.; Sacchettini, J. C.; Young, K.; et al. Identification of cyclic hexapeptides natural products with inhibitory potency against Mycobacterium tuberculosis. *BMC Res Notes* **2018**, *11* (1), 416. DOI: 10.1186/s13104-018-3526-z From NLM.
(211) Pettibone, D. J.; Clineschmidt, B. V.; Anderson, P. S.; Freidinger, R. M.; Lundell, G. F.; Koupal, L. R.; Schwartz, C. D.; Williamson, J. M.; Goetz, M. A.; Hensens, O. D.; et al. A structurally unique, potent, and selective oxytocin antagonist derived from Streptomyces silvensis. *Endocrinology* **1989**, *125* (1), 217-222. DOI: 10.1210/endo-125-1-217 From NLM.
(212) Wei, Z.-W.; Niikura, H.; Morgan, K. D.; Vacariu, C. M.; Andersen, R. J.; Ryan, K. S. Free Piperazic Acid as a Precursor to Nonribosomal Peptides. *Journal of the American Chemical Society* **2022**, *144* (30), 13556-13564. DOI: 10.1021/jacs.2c03660.
(213) Zwahlen, R. D.; Pohl, C.; Bovenberg, R. A. L.; Driessen, A. J. M. Bacterial MbtH-like Proteins Stimulate Nonribosomal Peptide Synthetase-Derived Secondary Metabolism in Filamentous Fungi. *ACS Synthetic Biology* **2019**, *8* (8), 1776-1787. DOI: 10.1021/acssynbio.9b00106.
(214) Felnagle, E. A.; Barkei, J. J.; Park, H.; Podevels, A. M.; McMahon, M. D.; Drott, D. W.; Thomas, M. G. MbtH-Like Proteins as Integral Components of Bacterial Nonribosomal Peptide Synthetases. *Biochemistry* **2010**, *49* (41), 8815-8817. DOI: 10.1021/bi1012854.
(215) Wex, K. W.; Saur, J. S.; Handel, F.; Ortlieb, N.; Mokeev, V.; Kulik, A.; Niedermeyer, T. H. J.; Mast, Y.; Grond, S.; Berscheid, A.; et al. Bioreporters for direct mode of action-informed screening of antibiotic producer strains. *Cell Chemical Biology* **2021**, *28* (8), 1242-1252.e1244. DOI: 10.1016/j.chembiol.2021.02.022 (acccessed 2026/04/13).

APPENDIX

A.1 Abbreviations

AA	Amino acid
ACP	Acyl carrier protein
AT	Acyltransferase
A domain	Adenylation domain
ARTS	Antibiotic Resistant Target Seeker
ADEPs	Acyldepsipeptides
AQC	6-Aminoquinolyl-N-hydroxysuccinimidyl carbamate
BGCs	Biosynthetic gene clusters
Bis-tris	2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
Cis-AT PKSs	cis-Acyltransferase polyketide synthases
CLF	Chain length factor
C domain	Condensation domain
Cy	Heterocyclization
ClpP	Caseinolytic protease
COSY	Correlation Spectroscopy
C-P	Carbon to phosphorus
DNA	Deoxyribonucleic acid
DH	Dehydratase
DEBS	Deoxyerythronolide B synthase
DMAPP	Dimethylallyl pyrophosphate
DMSZ	Deutsche Managementsystem Zertifizierungsgesell-schaft
DMSO	Dimethyl sulfoxide
DCM	Dichloromethane
DAD	Diode array detection
EtoAc	Ethyl acetate
ER	Enoylreductase
E	Epimerization
E. coli	*Escherichia coli*
ESKAPE	*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa* and *Enterobacter spp.*
EIC	Extracted ion chromatogram
ECF	Ethyl chloroformate
ESI	Electrospray Ionization
FDA	Food and drug administration
Fig	Figure
FADH	Flavin-dependent halogenases
FBMN	Feature-based molecular networking
FMOC-Cl	Fluorenylmethyl Chloroformate

FA	Formic acid
G+C	Guanine-plus-cytosine
GC	Gas chromatography
GNPS	Global Natural Products Social Molecular Networking
HPLC	High-performance liquid chromatography
HSQC	Heteronuclear-single-quantum-coherence spectroscopy
HMBC	Heteronuclear Multiple Bond Correlation
HILIC	hydrophilic interaction liquid chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPP	Isopentenyl pyrophosphate
IPK	Isopentenyl pyrophosphate kinase
IDI	Isopentenyl diphosphate isomerase
KS	Ketosynthase
KR	Ketoreductase
LC-MS/MS	Liquid Chromatography-Tandem Mass Spectrometry
LC-HRMS	Liquid Chromatography-High Resolution Mass Spectrometry
MRSA	Methicillin-resistant *Staphylococcus aureus*
MSAS	Methylsalicylic acid synthase
M	Methylation
MRSE	Methicillin-Resistant *Staphylococcus Epidermidis*
MVA	Mevalonate
MEP	Methylerythritol
MVK	Mevalonate kinase
MS	Mass spectrometry
MS/MS	Tandem Mass Spectrometry
MassIVE	Mass Spectrometry Interactive Virtual Environment
MCheM	Multiplexed Chemical Metabolomics workflow
MOPS	3-(N-morpholino)propanesulfonic acid
MeCN	Acetonitrile
MeOH	Methanol
NPs	Natural products
NRPs	Nonribosomally synthesized peptides
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
Ox	Oxidation
OSMAC	One strain many compounds
PKSs	Polyketide synthase
PPTase	Phosphopantetheinyl Transferase
PCP	Peptidyl carrier protein
PPant	Phosphopantheteine
PMK	Phosphomevalonate kinase
PMD	Phosphomevalonate decarboxylase

PEP	Phosphoenolpyruvate
PnPy	Phosphonopyruvate
PCR	Polymerase Chain Reaction
RiPPs	Ribosomally synthesized and post-translationally modified peptides
R	Reduction
TE	Thioesterase
trans-AT PKSs	Trans-acyltransferase polyketide synthases
T domain	Thiolation domain
TOCSY	Total Correlation Spectroscopy
TFA	Trifluoroacetic acid
UV	Ultraviolet

A.2 Supplementary information to Chapter 2

The genome mining of phosphonate BGCs and genetic engineering has been done by Alina Zimmermann and Dr. Juan Pablo Gomez-Escribano at the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures in Braunschweig supervised by Prof. Dr. Yvonne Mast.

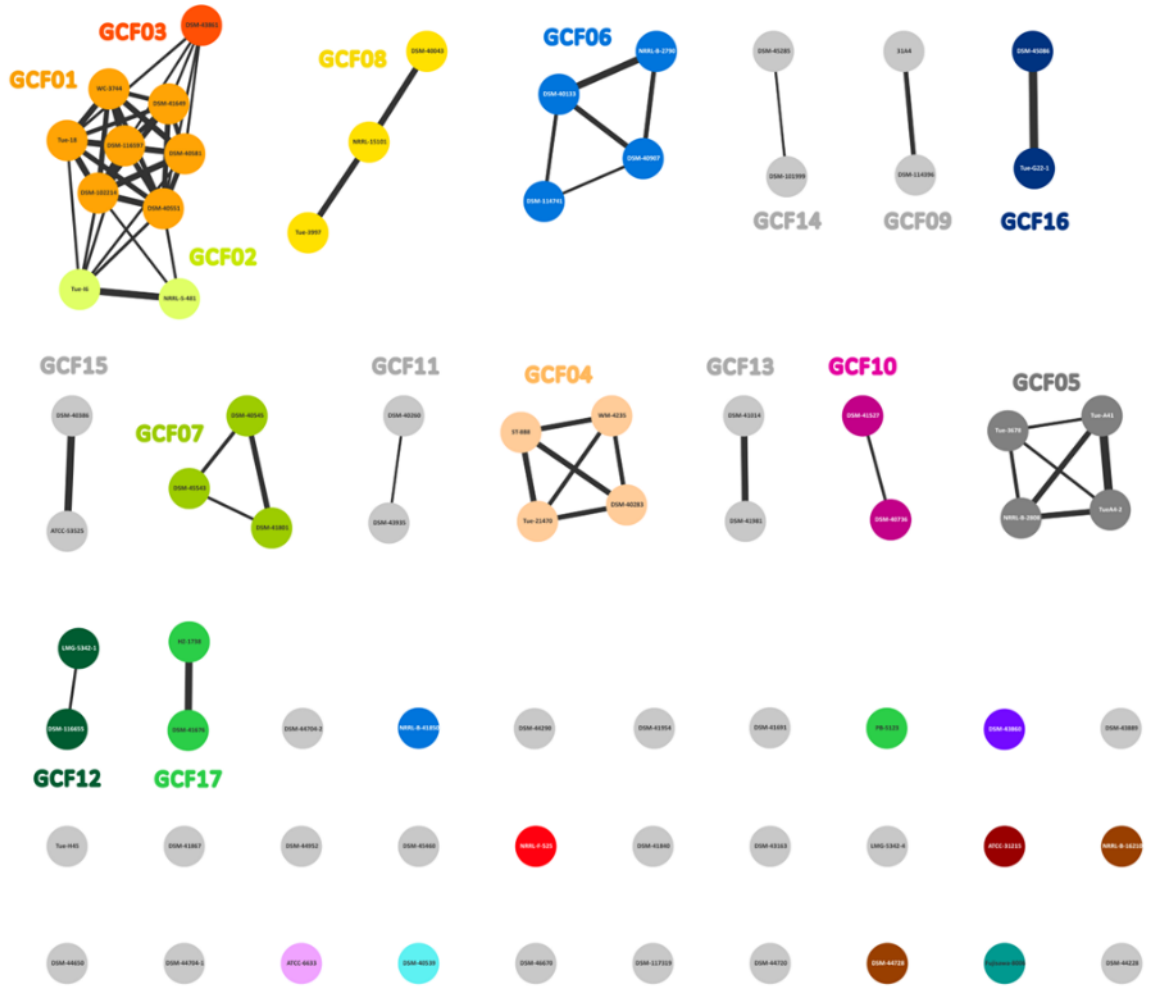


Figure A 1. Gene cluster families of identified and known phosphonate compounds BGCs. Gene cluster family (GCF) network as obtained from the BiG-SCAPE analysis and visualized with Cytoscape. Each node corresponds to a BGC. Colored spheres represent GCFs/singletons with sequences from previously identified phosphonate producers; light grey spheres are P-BGCs with unknown phosphonate products. Edges between two nodes represent a distance between the P-BGCs below a BiG-SCAPE cutoff threshold of 0.612 and distances are further represented by weighted lining of edges (lines get thicker with decreased distance).

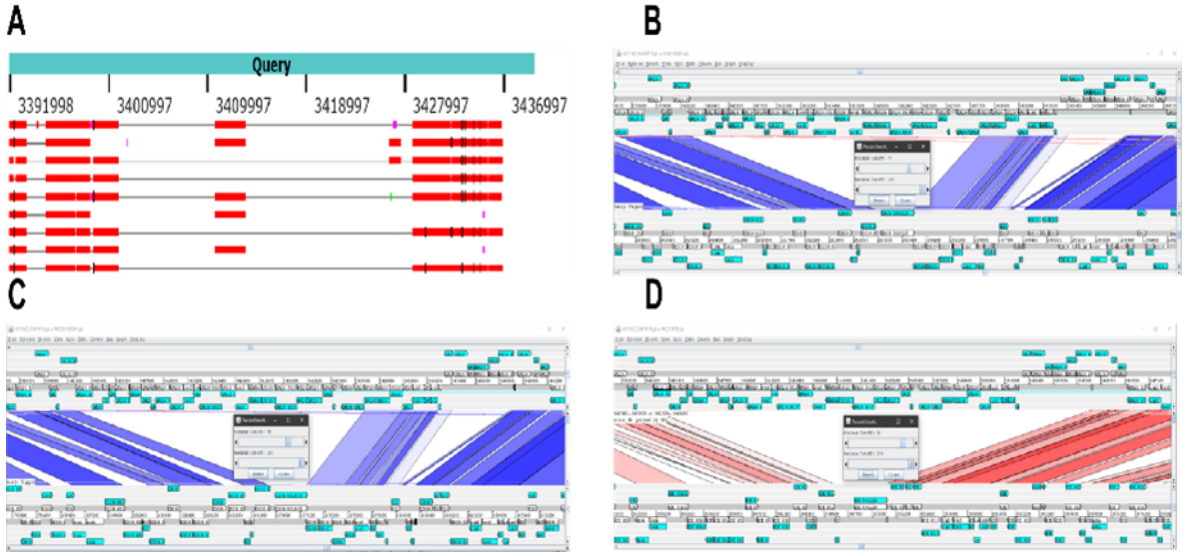
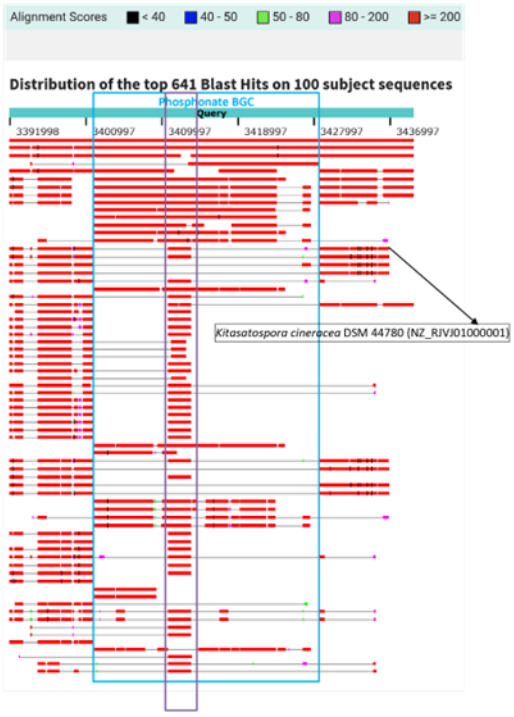


Figure A 2. Synteny analysis for defining the boundaries of the putative phosphonate BGC.**A.** Similarity search with NCBI BLAST blastn program, using as query *Kitasatospora fiedleri* DSM 114396 chromosome sequence with accession NZ_OX419519, limited to the range 3391998-3439448, against the NCBI Database "Refseq prokaryote representative genomes (contains refseq assembly) Update date 2025/05/12" limited to "*Kitasatospora cineracea* (taxid:88074)", the closest known species to DSM 114396 and for which several genome assemblies are available. The result shows the highly conserved regions surrounding the putative phosphonate biosynthetic gene cluster, therefore helping to determine the boundaries of the gene cluster. The highly conserved region around position 3409997 in the image represents the segment 3410635 to 3413339 and contains the genes with locus_tag QMQ26_RS15735, QMQ26_RS15740 and QMQ26_RS15745, all of them homologs of the highly conserved genes with locus_tag QMQ26_RS09285 (*cysN*), QMQ26_RS09280 (*cysD*), and QMQ26_RS09275 (*cysC*), that span from position 1960006 to 1962829 of the chromosome of DSM 114396, and encode the highly conserved enzymes sulphate adenylyltransferase subunit 1 CysN [EC:2.7.7.4; KEEG K00956], sulphate adenylyltransferase subunit 2 CysD [EC:2.7.7.4; KEEG K00957], and adenylylsulphate kinase CysC [EC:2.7.1.25; KEEG K00860] from sulphur metabolism in bacteria (see also Fig. S3 and S5). The upstream end (right in Figure S3) is supported by a synteny analysis with the type strains of *Kitasatospora cineracea* and *Kitasatospora niigatensis*, the most closely related species to *K. fiedleri* lacking a P-BGC[1] (Figure S3).The downstream end (left in the figure) is well-supported also by highly conserved homologous genes beyond the aldehyde dehydrogenase, with the first gene (encoding a chloride channel protein) showing 69-97% identity to *Kitasatospora* strains in the NCBI non-redundant protein database. **B and C.** Artemis Comparison Tool (ACT) visualisation of the NCBI BLAST analysis of DSM 114396 chromosome (NZ_OX419519) against *K. cineracea* DSM 44780 contig NZ_RJVJ01000001 (**B**) or against *Kitasatospora niigatensis* DSM 44781 contig NZ_RKQG01000001 (**C**), using NCBI megablast with default options. Note that the display of the bottom sequence has been flipped in both B and C to match the orientation of the top one. **D.** ACT visualisation of the NCBI BLAST analysis of NZ_OX419519 against *Kitasatospora setae* KM-6054 (the type strain of the genus) chromosome (NC_016109.1) showing that the synteny restarts only with the gene with locus_tag QMQ26_RS15885 encoding a tRNA-Thr, as a result of being *K. setae* a more distant species.

A



B

Strain source of PepM as compared against *K. fiedleri* DSM 114396 PepM	Identity	Similarity
Kitasatospora sp. NPDC088783_185_F4.ctg-0008 NZ_JBIVKJ010000008	1	1
Kitasatospora sp. SolWspMP NZ_QLLT01000023	1	1
Kitasatospora sp. NPDC090308_182_E10.ctg-0012 NZ_JBIVXY010000012	1	1
Streptomyces sp. ISL-100 ISL100_59 NZ_JAGGOK010000059	0.87	0.91
Streptomyces sp.NPDC051109_71_F2.ctg-0006 NZ_JBITFF010000006	0.86	0.91
Streptomyces sp. MMG1121 P433contig3.1 NZ_LGDV01000190	0.86	0.91
Streptomyces sp. MMG1121 KF386872	0.86	0.91
Streptomyces sp. NPDC091268_165_H6.ctg-0011 NZ_JBIVVE010000011	0.85	0.91
Streptomyces sp. NPDC053429_373_F9.ctg-0003 NZ_JBITXY010000003	0.85	0.9
Streptomyces sp. R35 CP163440	0.84	0.91
Streptomyces sp. NPDC060198_T5_A10.ctg-0004 NZ_JBHWZJ010000004	0.84	0.91
Streptomyces sp. 31A4 KF386877	0.84	0.9
Streptomyces sp. NBC_00247 NZ_CP108093	0.84	0.9
Kitasatospora sp. Root107 contig_18 NZ_LMCT01000010	0.83	0.9
Kitasatospora sp. Root187 contig_15 NZ_LMHX01000007	0.83	0.9
Streptomyces odontomachi strain ODS25 24 NZ_JAMJWG010000024	0.81	0.89
Streptomyces sp. NPDC019224_87_D10.ctg-0005 NZ_JBEYCF010000005	0.8	0.89
Streptomyces kanasensis strain ZX01 contig00030 NZ_LNSV01000030	0.8	0.88
Streptomyces changanensis strain HL-66 NZ_CP102332	0.8	0.88
Streptomyces sp. CSDS2 NODE_17 NZ_JAUEPB010000017	0.8	0.88
Streptomyces kanasensis NPDC016311 NZ_JBIBSF010000001	0.8	0.87
Streptomyces kanasensis strain NPDC086832 NZ_JBIVAV010000006	0.8	0.87

C

Figure A 3. An analysis at NCBI BLAST, using the range 3391998 to 3439448 of NZ_OX419519.1 was performed against the “RefSeq Genome Database”, limited to “Actinobacteria (taxid: 201174)”, and using the program “megablast” with default parameters. The graphical summary of the results is shown in panel **A**. The gene cluster, as delimited by synteny analysis with *Kitasatospora cineracea* DSM 44780 is framed in a blue box. All sequences that provided coverage of the BGC (and not just for the homologues of the highly conserved *cysCDN* genes for sulfur metabolism to the right of the 3409997 mark) originated from whole genome sequencing of microorganisms classified as *Streptomyces* or *Kitasatospora* species (table **B**); these alignments were thoroughly analysed, and the segments of the NCBI RefSeq records spanning 50 kb (or to the beginning or end in case the deposited sequence was not long enough) starting from 10 kb downstream of the aldehyde dehydrogenase gene, were downloaded as fully annotated GenBank files. These sequence files, together with the sequence comprising 3391998 to 3439448 of NZ_OX419519, were used for similarity analysis with CLINKER (with the default “Minimum alignment sequence identity” of 0.30). Table **B** shows the strains (with the accession numbers of the sequences used) included in CLINKER analysis, ordered according to CLINKER’s similarity of the PepM encoded by each strain as compared to *K. fiedleri* PepM. **C.** Legend for the colour used in the CLINKER plot to group homologous genes.

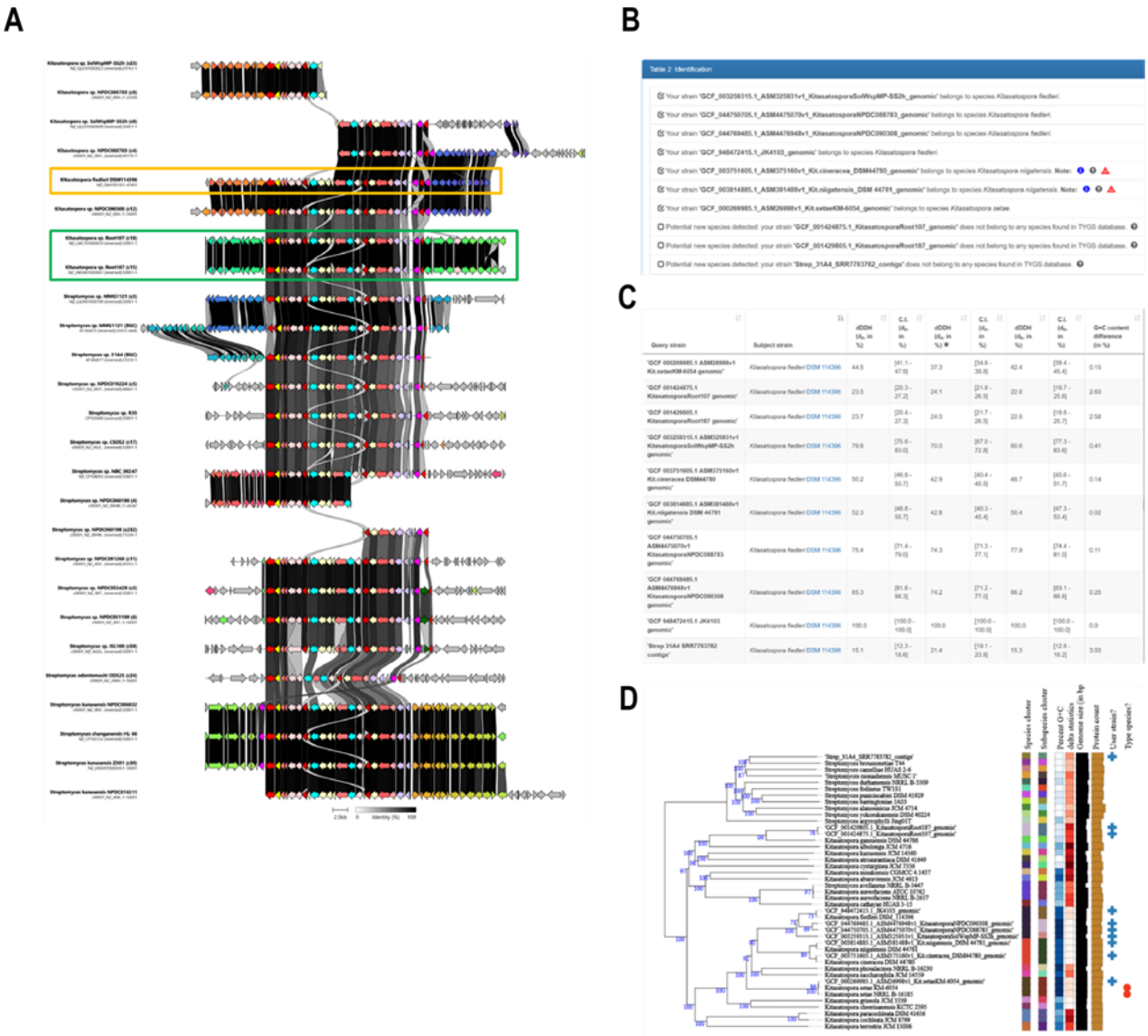


Figure A 4. Clinker analysis for defining the boundaries of the putative phosphonate BGC. CLINKER plot(**A**), with BGCs sorted according to, first, phylogeny proximity (*Kitasatospora* strains at the top), and finally the order was adapted to highlight similarity and synteny along the entire BGC and surrounding genes; the plot is aligned at the aldehyde dehydrogenase gene (in red). A phylogenetic analysis was performed with TYGS server to assess phylogenetic relationship among the *Kitasatospora* strains that carry a similar P-BGC to *K. fiedleri*. The full genome assemblies were downloaded from NCBI and submitted as queries to TYGS. The strains included in the analysis and the taxonomical judgment provided by TYGS(**B**). The similarity statistics(**C**). The full-genome phylogenetic tree(**D**).

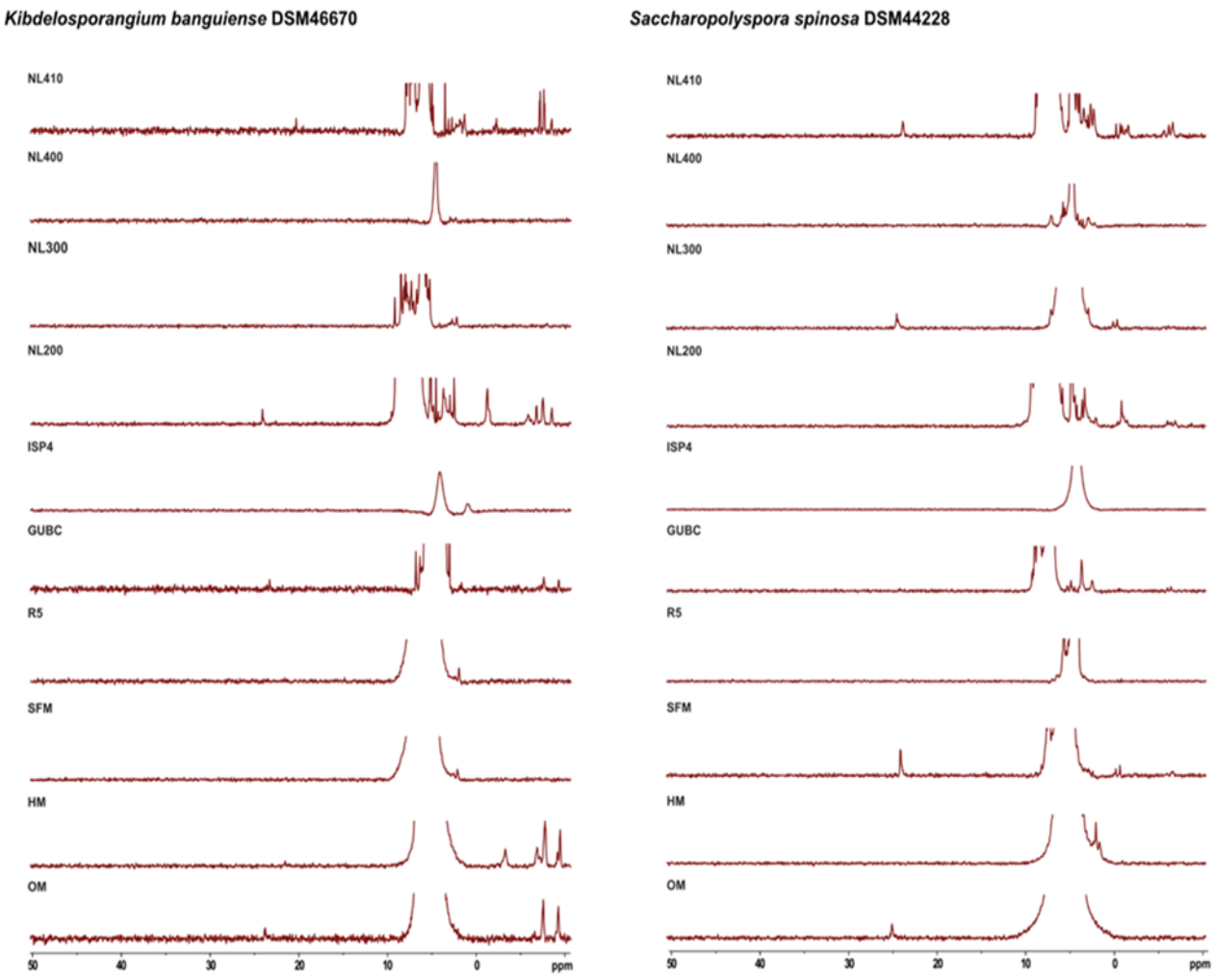


Figure A 5. ^{31}P NMR of culture extracts of *Kibdelosporangium banguiense* DSM46670 and *Saccharopolyspora spinosa* DSM 44228 grown in different media

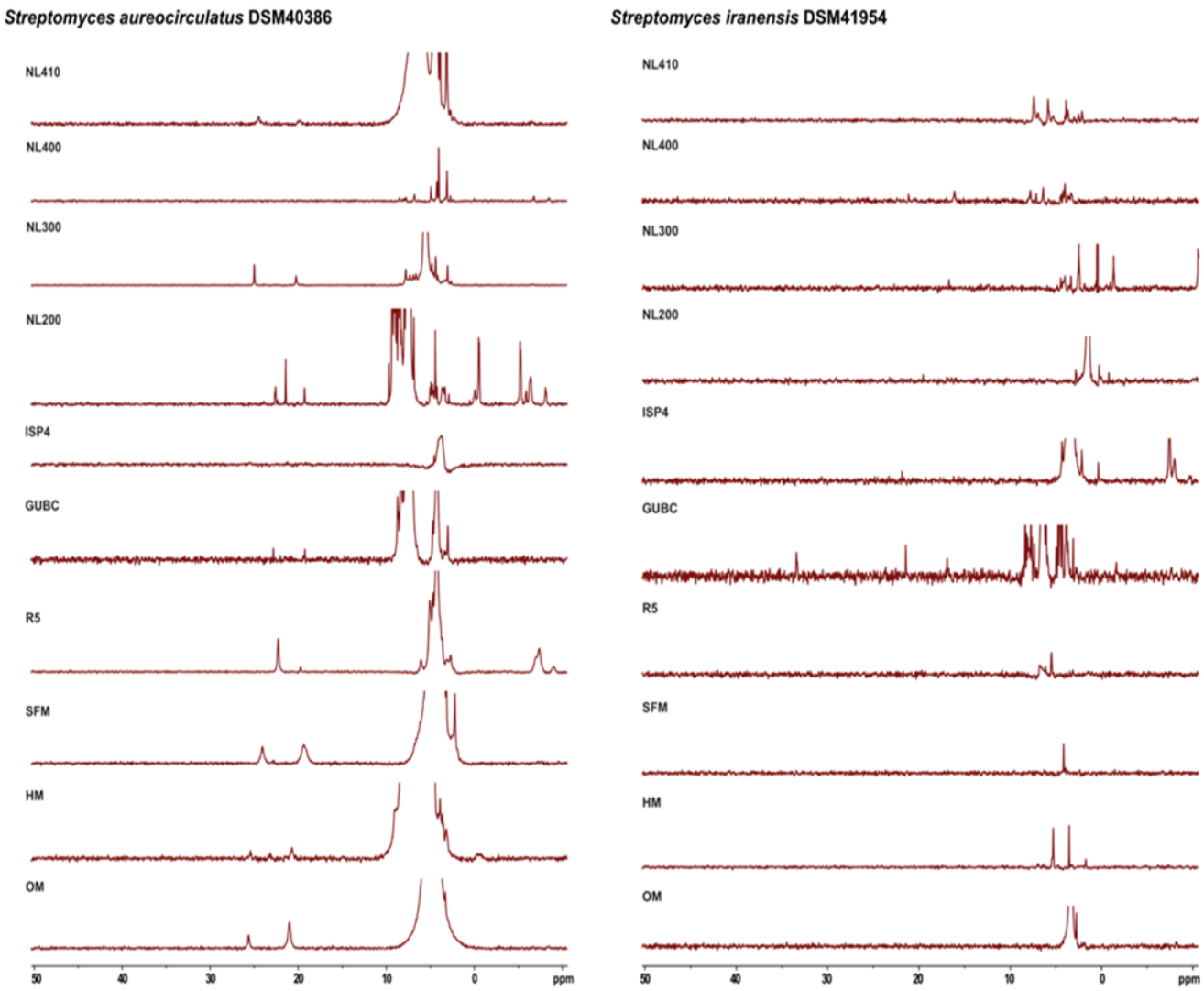


Figure A 6. ^{31}P NMR of culture extracts of *Streptomyces aureocirculatus* DSM40386 and *Streptomyces iranensis* DSM 41954 grown in different media

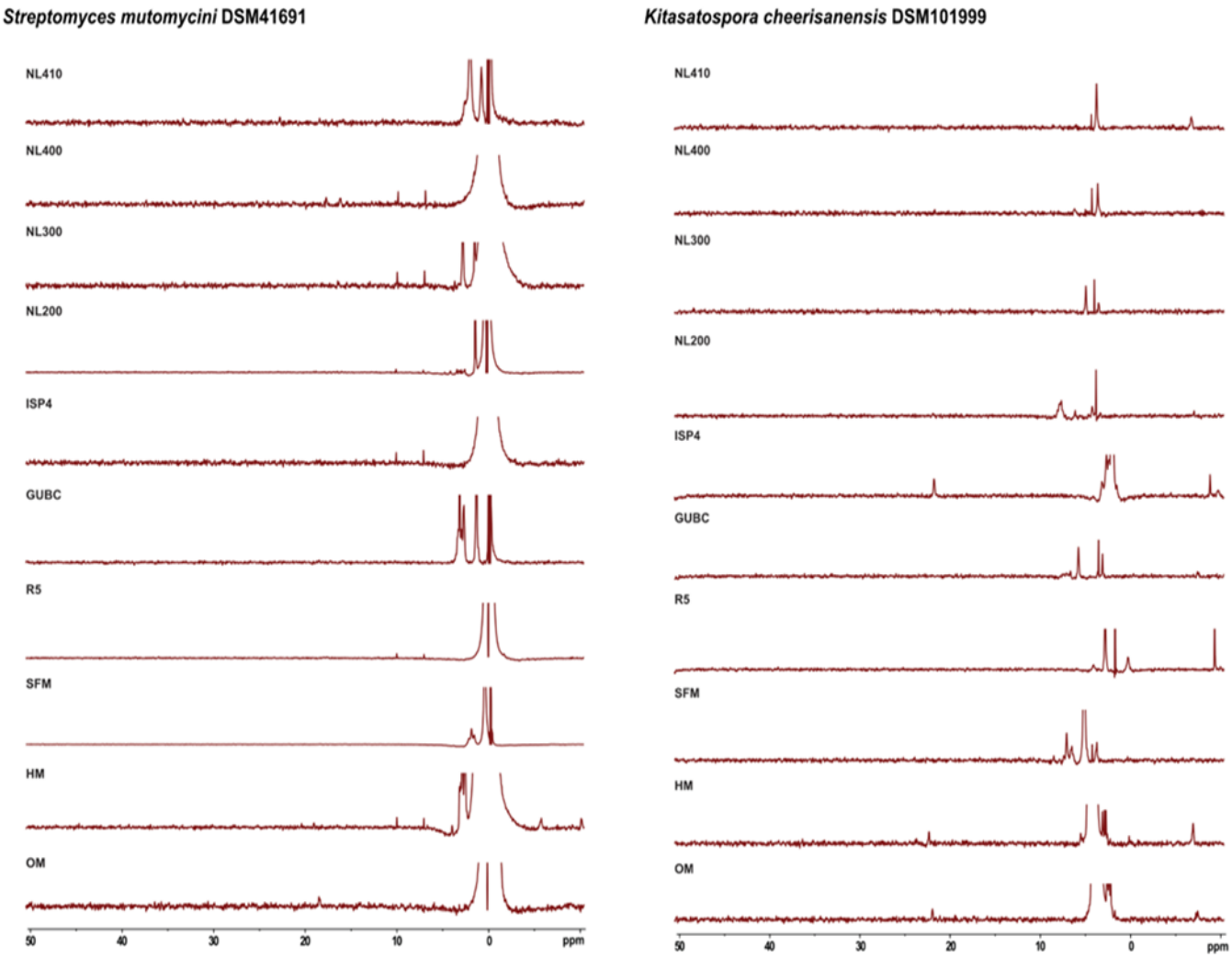


Figure A 7. ^{31}P NMR of culture extracts of *Streptomyces mutomycini* DSM41691 and *Kitasatospora cheerisanensis* DSM 101999 grown in different media

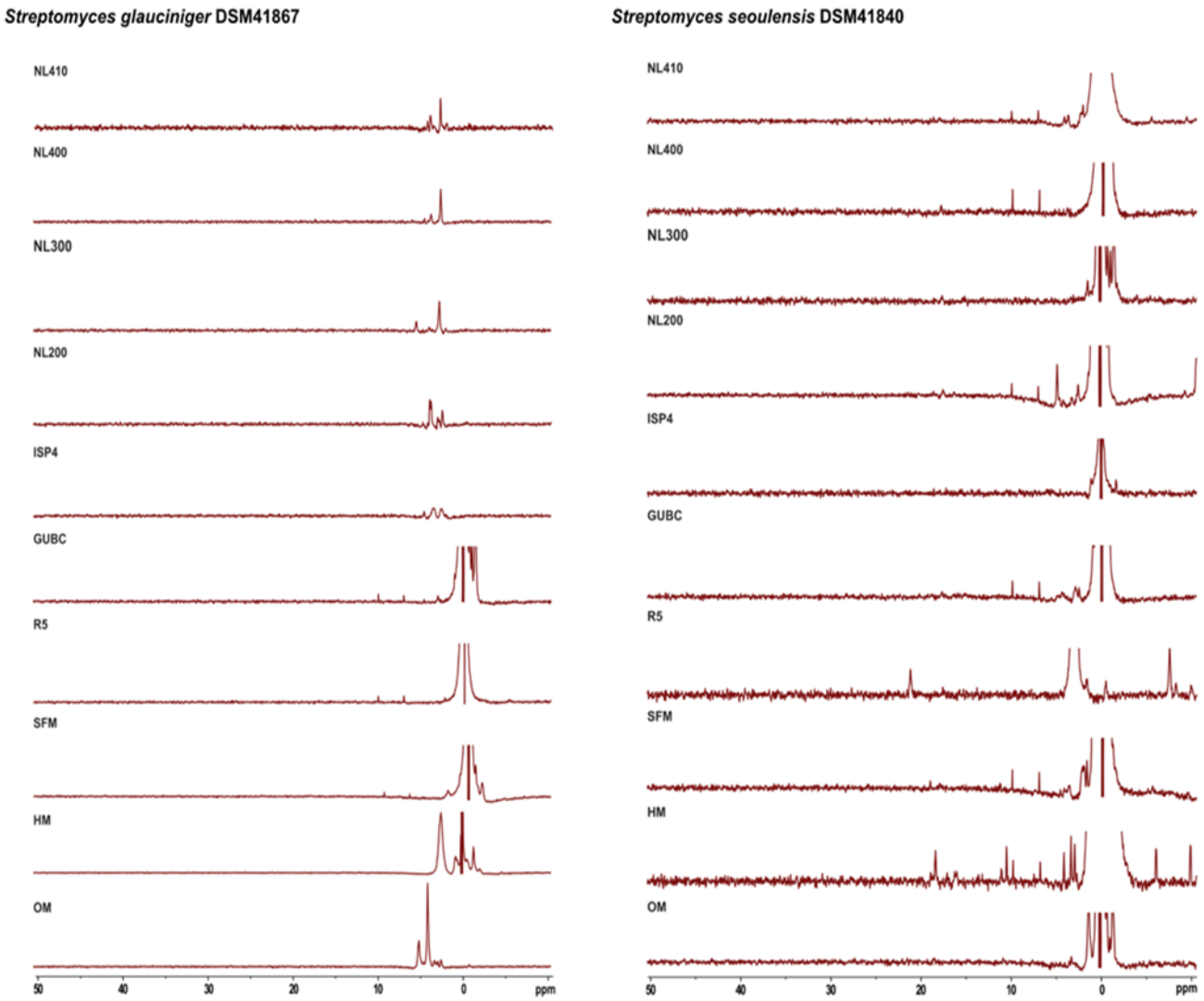


Figure A 8. ^{31}P NMR of culture extracts of *Streptomyces glauciniger* DSM 41867 and *Streptomyces seoulensis* DSM 41840 grown in different media

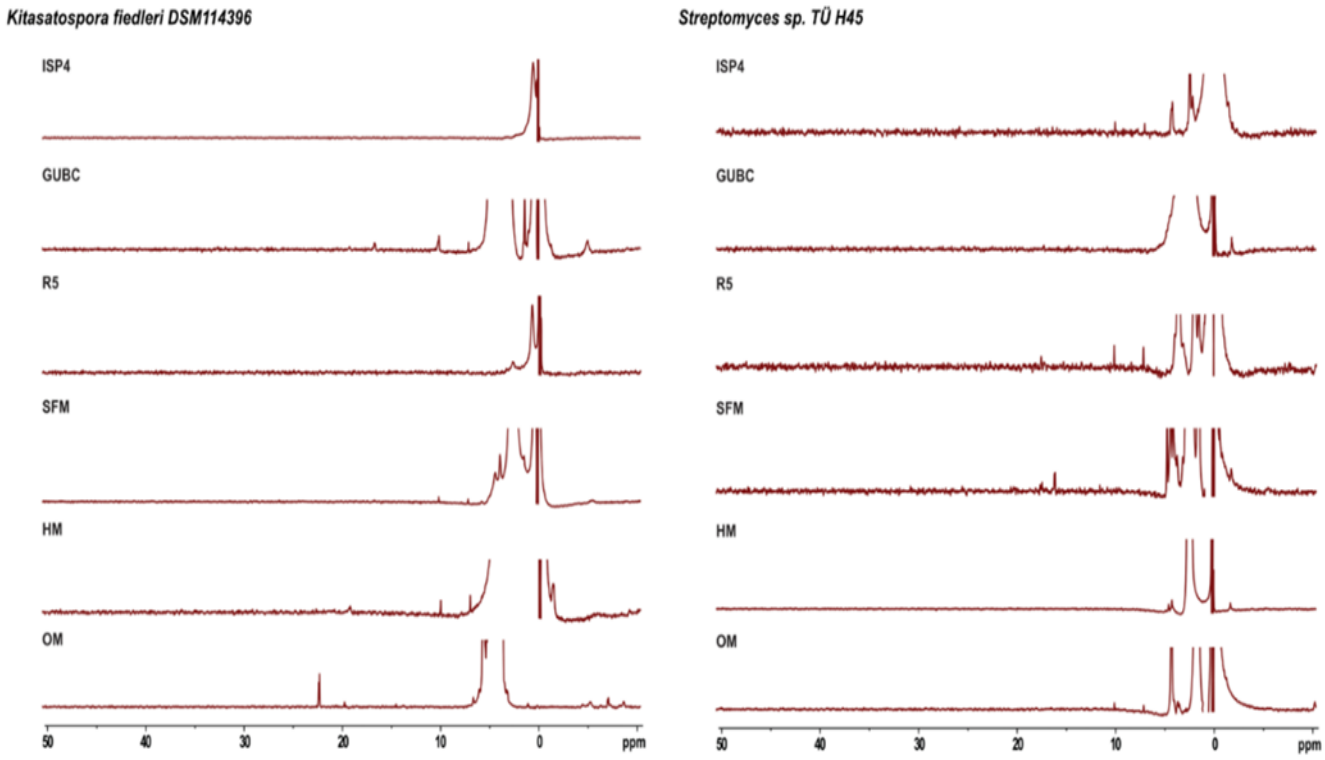


Figure A 9. ^{31}P NMR of culture extracts of *Kitasatospora fiedleri* DSM 114396 and *Streptomyces sp.* TÜ H45 grown in different media

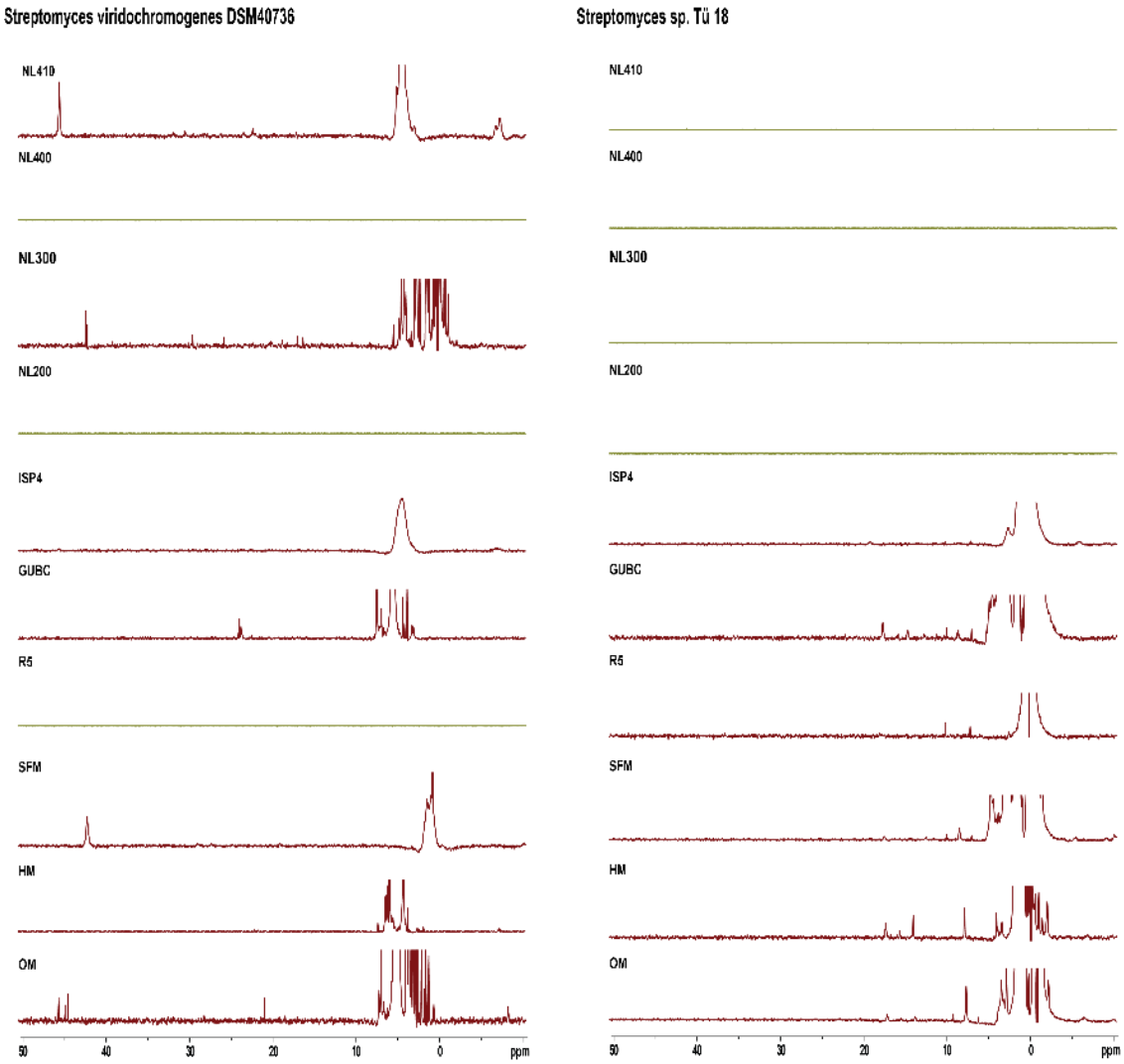


Figure A 10. ^{31}P NMR of culture extracts of *Streptomyces viridochromogenes* DSM 40736 and *Streptomyces sp.* TÜ 18 grown in different media

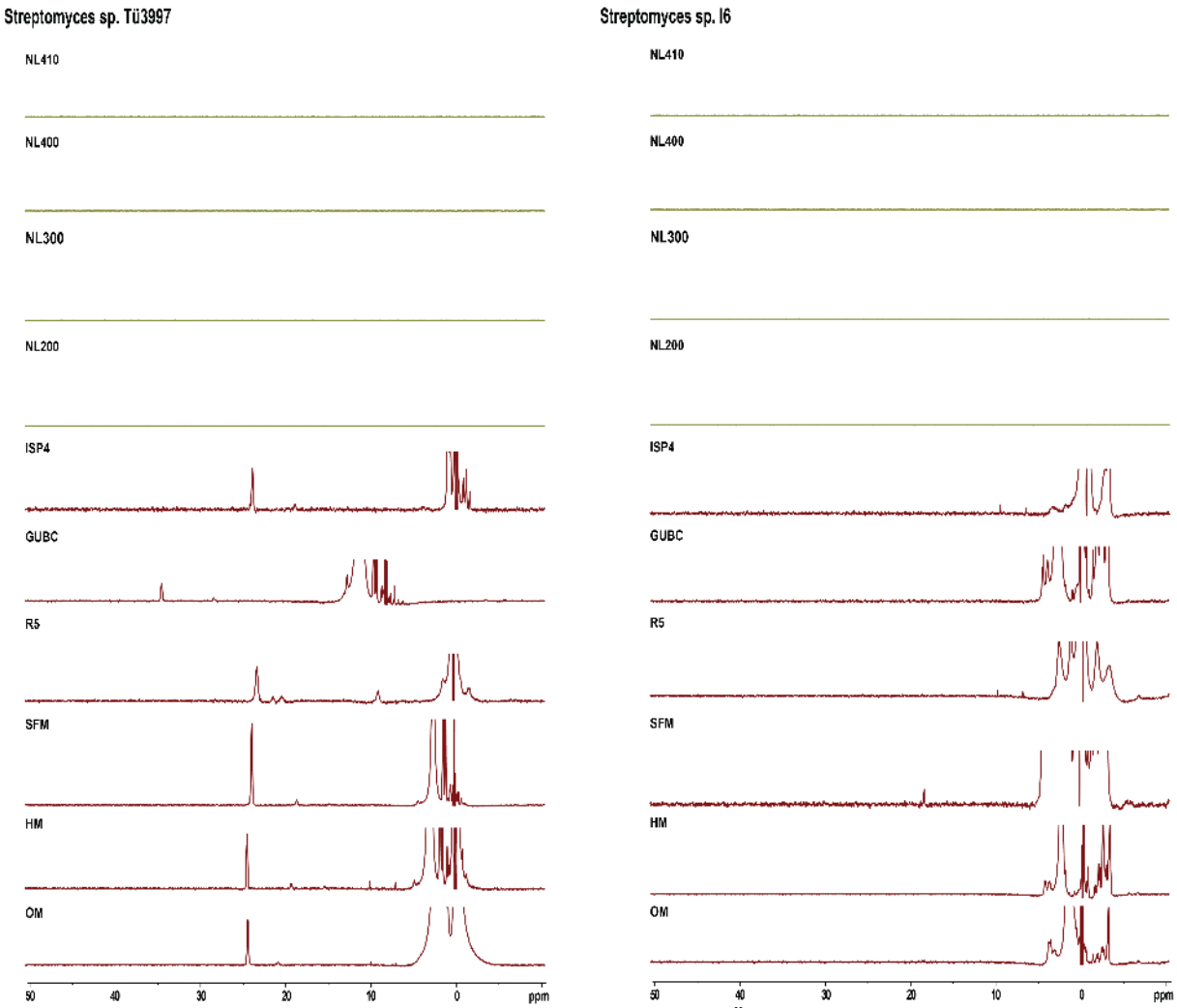


Figure A 11. ^{31}P NMR of culture extracts of *Streptomyces sp.* TÜ 3997 and *Streptomyces sp.* I6 grown in different media

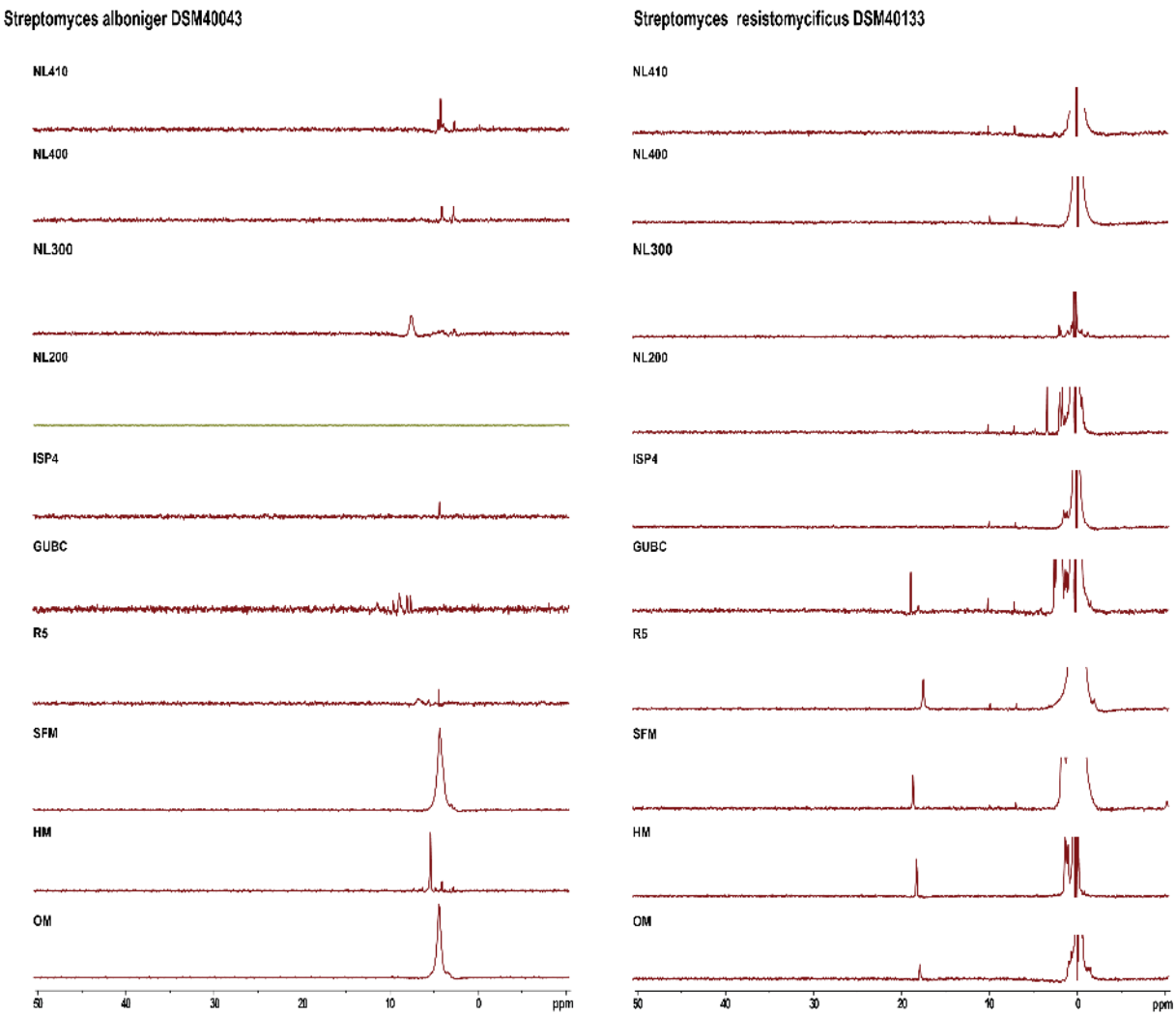


Figure A 12. ^{31}P NMR of culture extracts of *Streptomyces alboniger* DSM 40043 and *Streptomyces resistomycificus* DSM 40133 grown in different media.

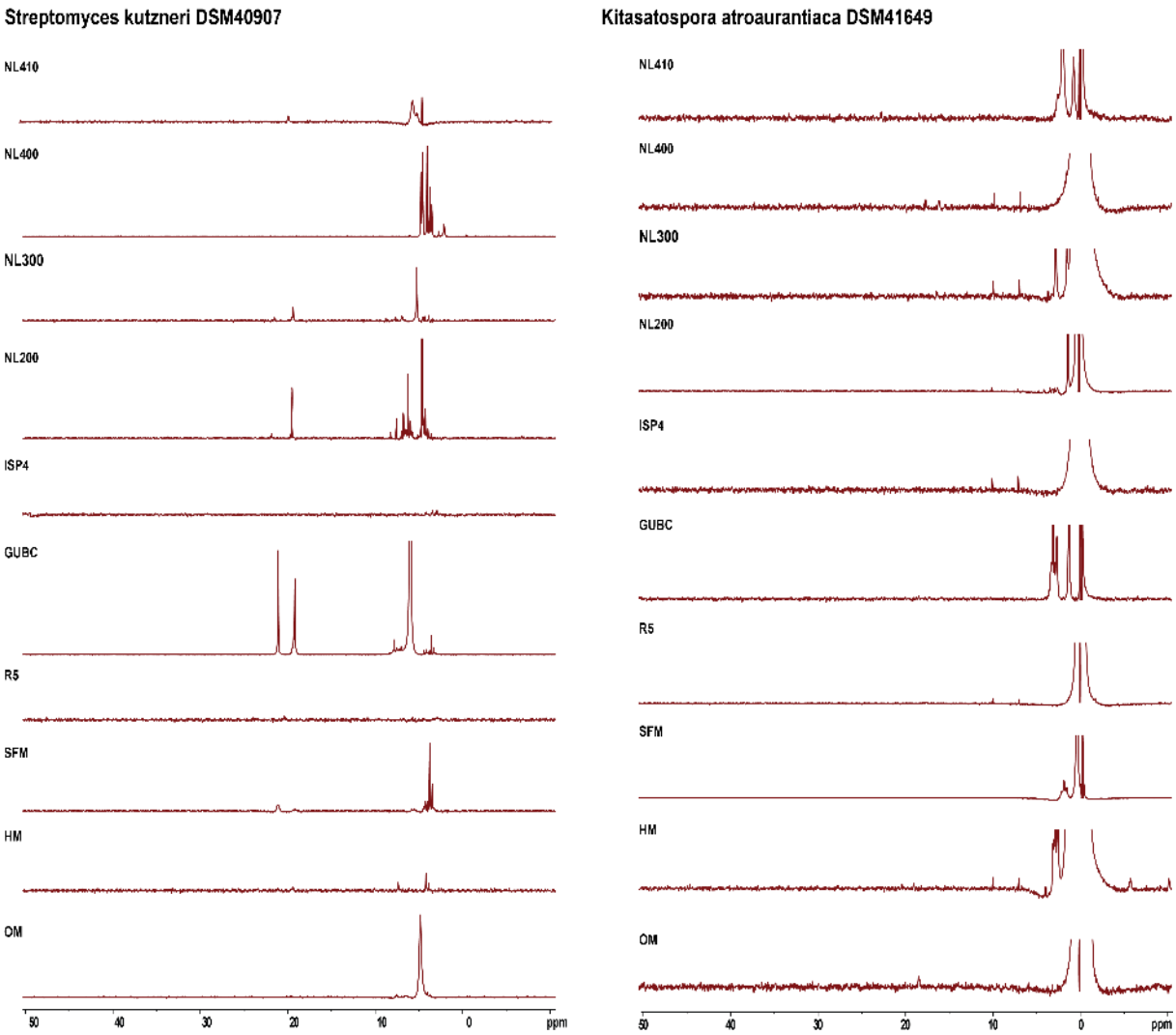


Figure A 13. ^{31}P NMR of culture extracts of *Streptomyces kutzneri* DSM 40907 and *Kitasatospora atroaurantiaca* DSM 41649 grown in different media.

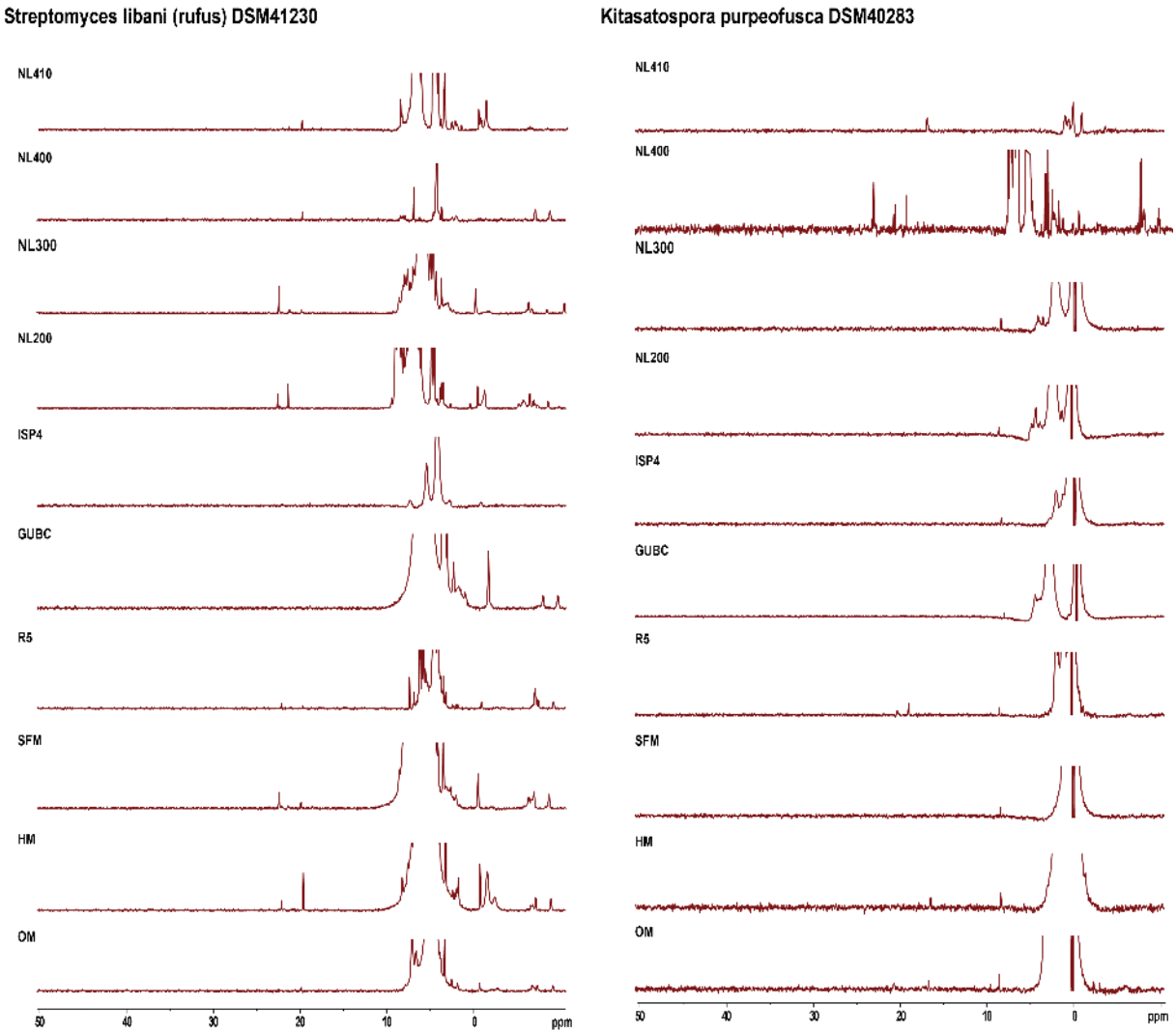


Figure A 14. ^{31}P NMR of culture extracts of *Streptomyces libani (rufus)*DSM 41230 and *Kitasatospora purpeofusca* DSM 40283 grown in different media.

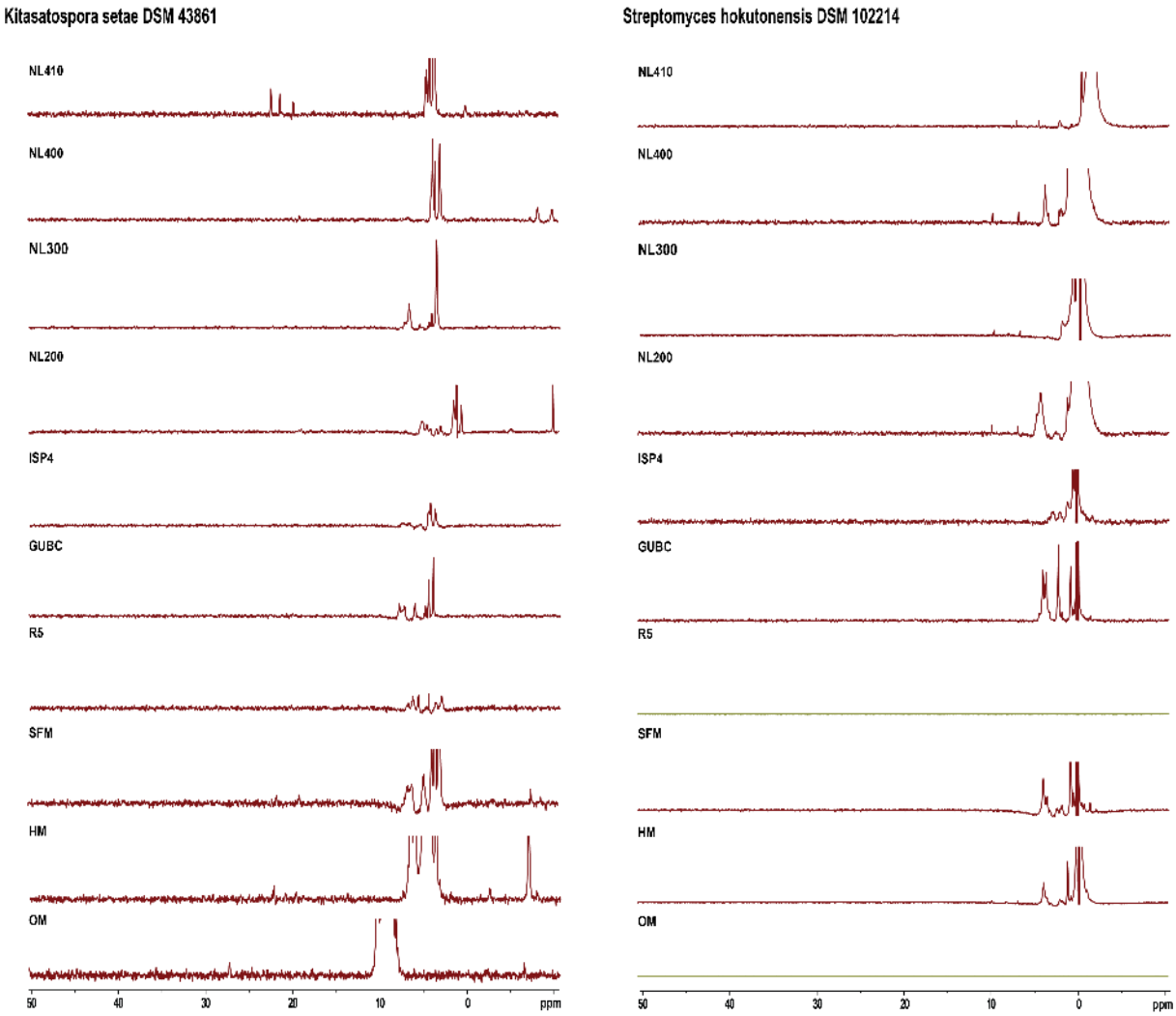


Figure A 15. ^{31}P NMR of culture extracts of *Kitasatospora setae* DSM 43861 and *Streptomyces hokutonensis* DSM 102214 grown in different media.

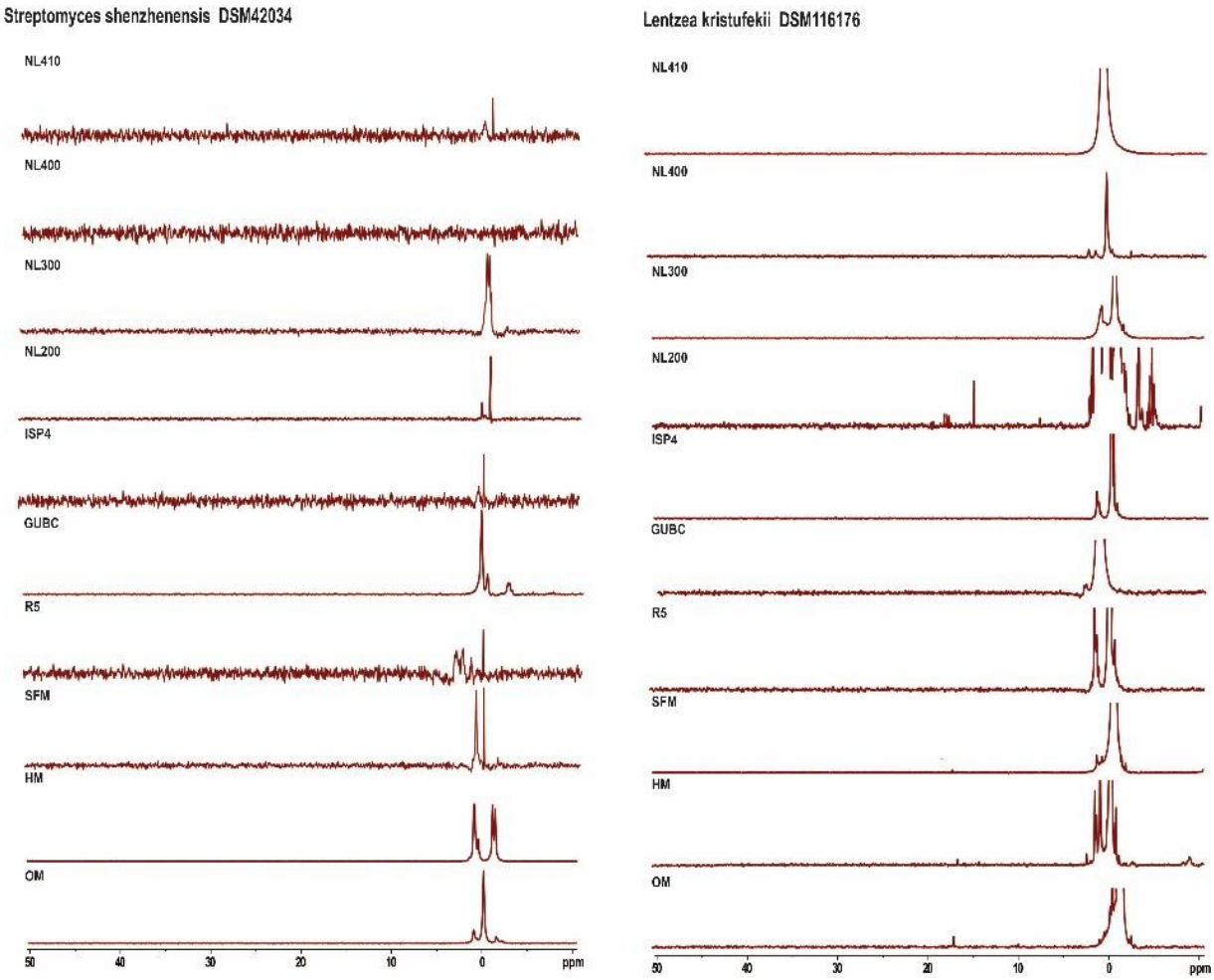


Figure A 16. ^{31}P NMR of culture extracts of *Streptomyces shenzhenensis* DSM 42034 and *Lentzea kristufekii* DSM 116176 grown in different media.

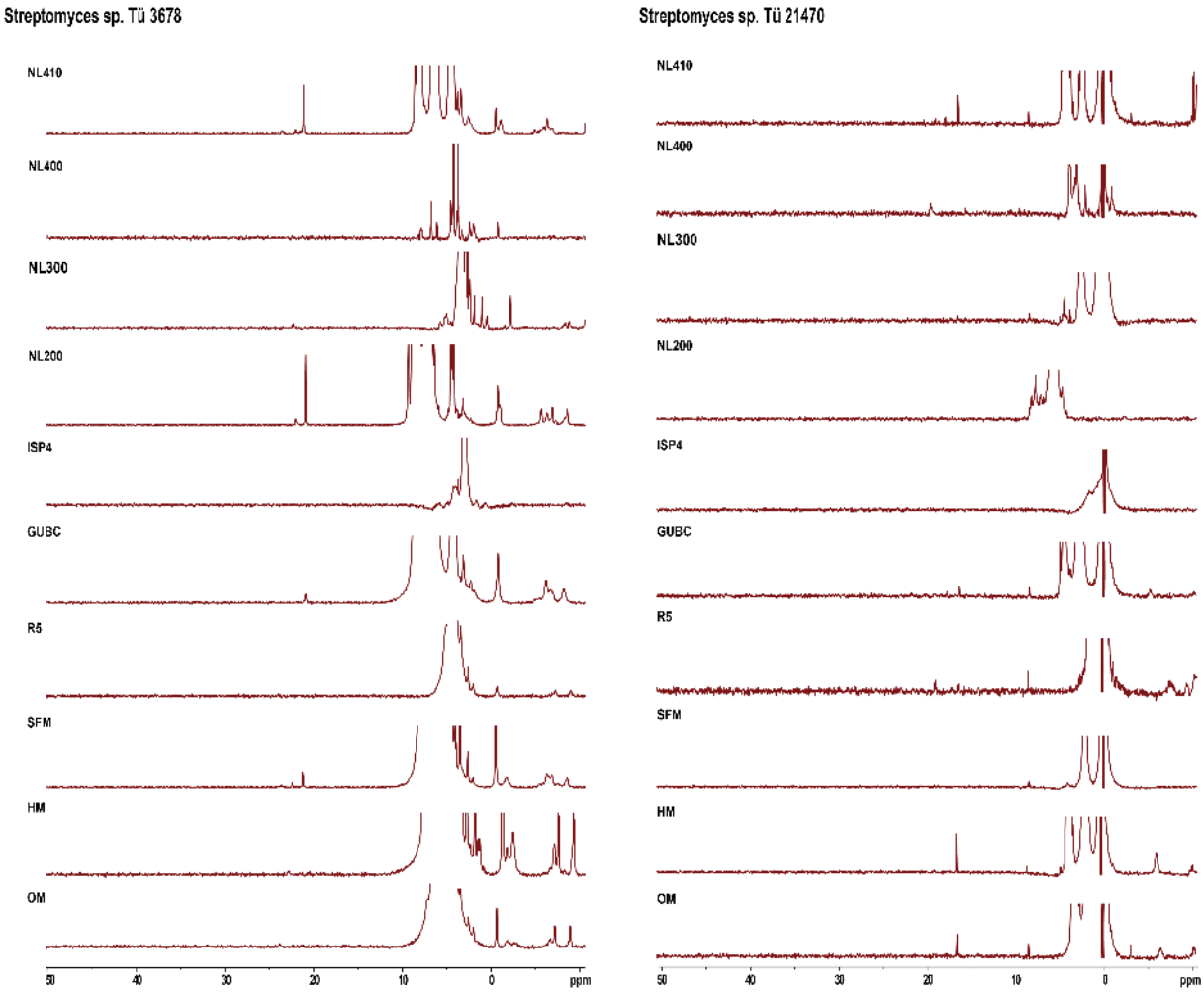


Figure A 17. ^{31}P NMR of culture extracts of *Streptomyces sp.* TÜ 3678 and *Streptomyces sp.* TÜ 21470 grown in different media

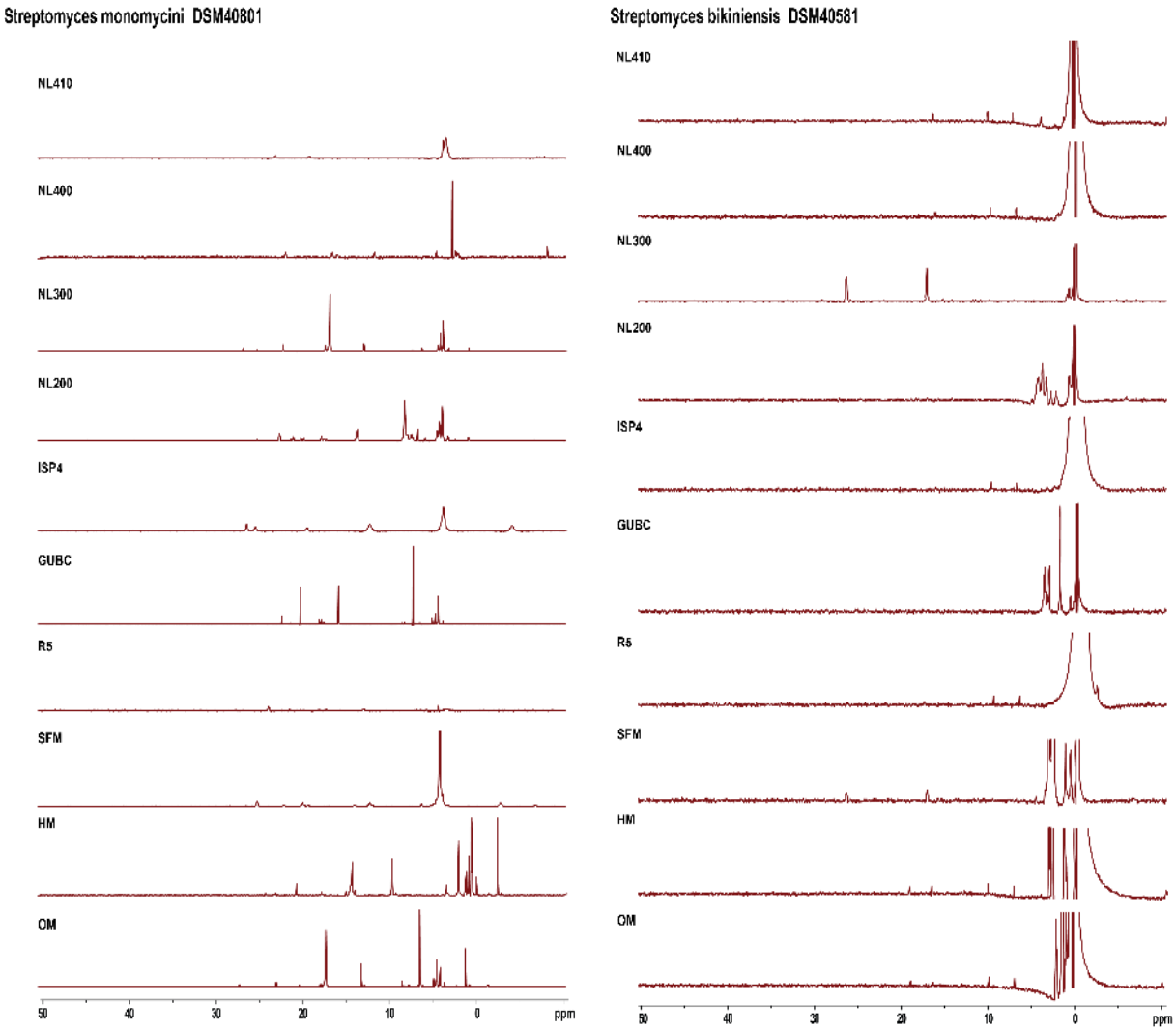


Figure A 18. ^{31}P NMR of culture extracts of *Streptomyces monomycini* DSM 40801 and *Streptomyces bikiniensis.* DSM40581 grown in different media

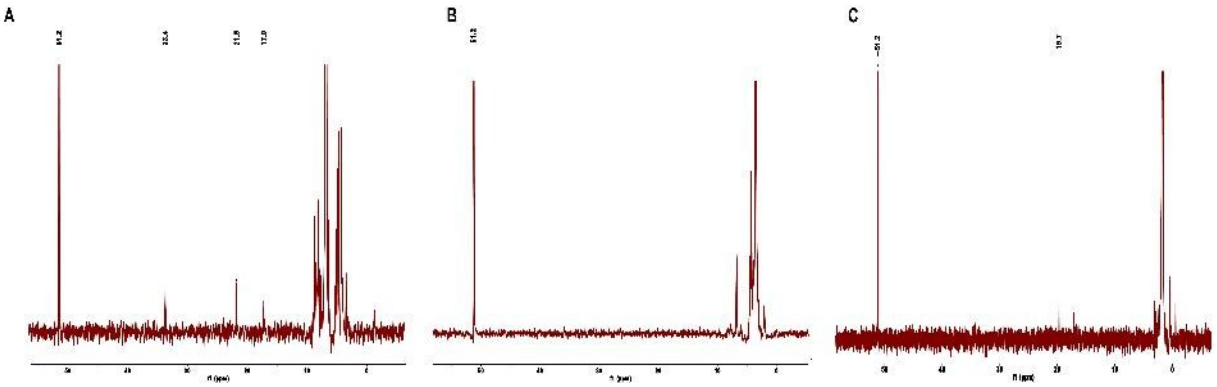


Figure A 19. ^{31}P NMR spectrum of concentrated culture supernatant from *S. iranensis* DSM 41954. A. Wild-type strain sample showing production of phosphonate-containing metabolites (δP 33.4, 21.5 and 17.0). B. Mutant strain sample of deletion *pepM* showing that production of phosphonate-containing

metabolites has been abolished. C. Mutant strain sample of deletion *lac* showing that production of phosphonate-containing metabolites (δP 33.4). L-phosphinothricin HCl (10 mM, δP 51.2) was used as a chemical shift reference

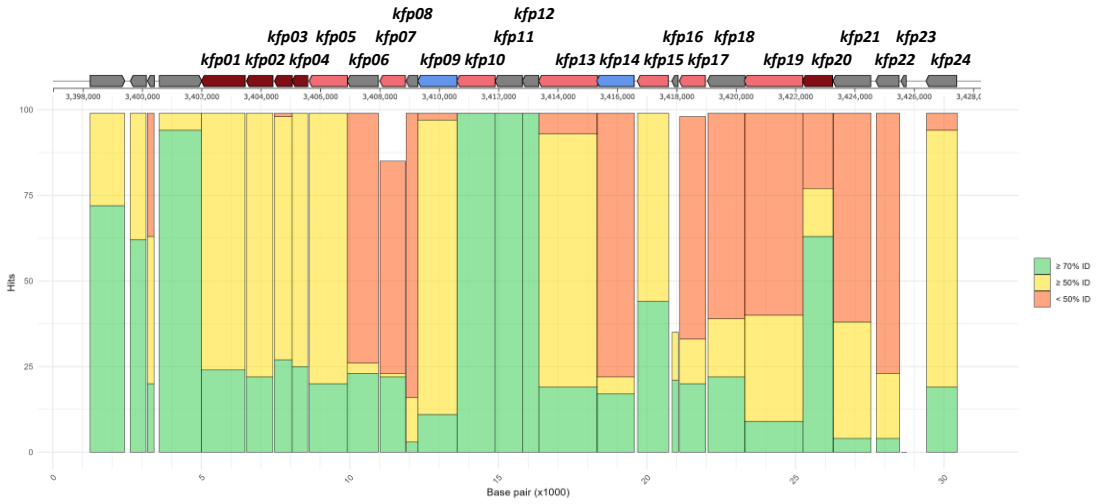


Figure A 20. BLAST analysis of AA sequences from the *Kitasatospora fiedleri* DSM *114396* phosphonate BGC

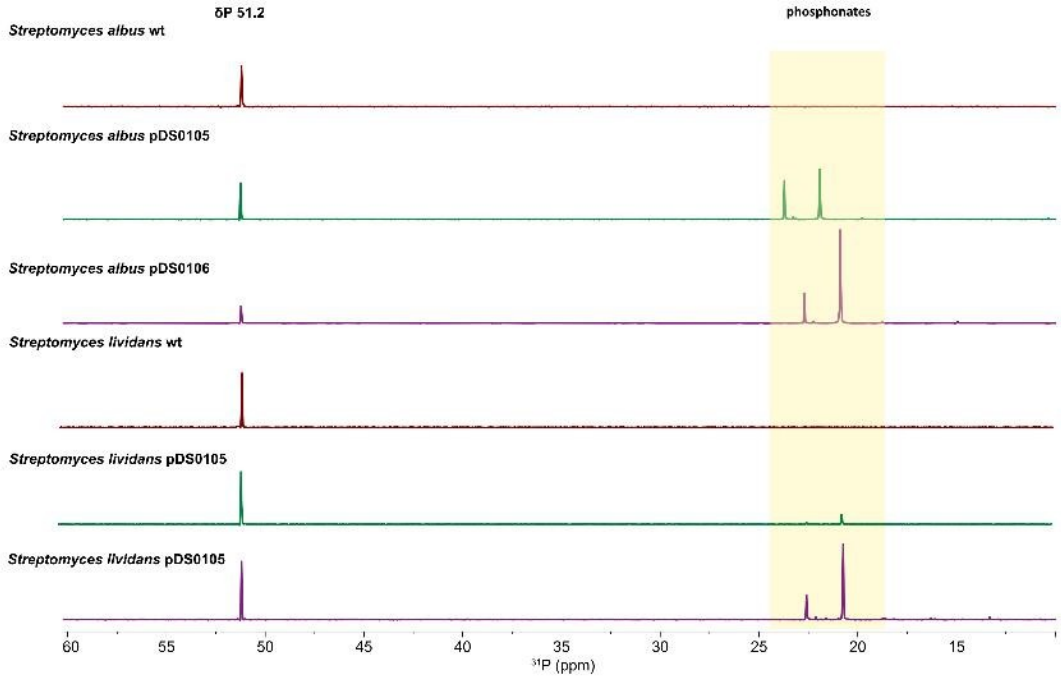


Figure A 21. ^{31}P NMR spectrum of cultures from *Kitasatospora fiedleri* DSM 114396 (minimal gene cluster kfp02-kfp04) heterologous expression in *S. albus* and *S. lividans*

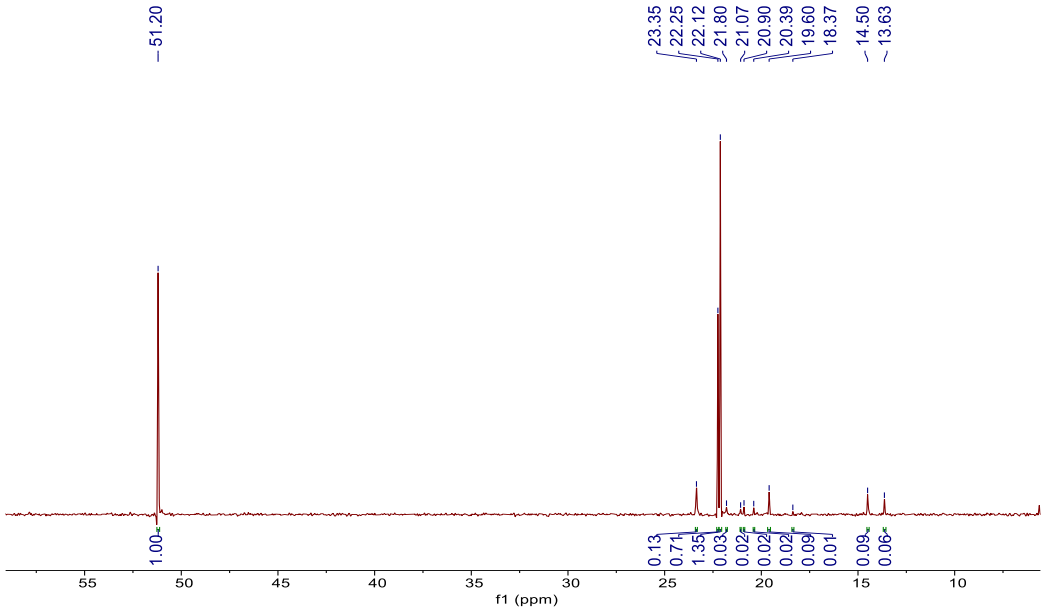


Figure A 22. ^{31}P NMR spectrum of cultures from *Kitasatospora fiedleri* DSM 114396 after MeOH and acetic acid

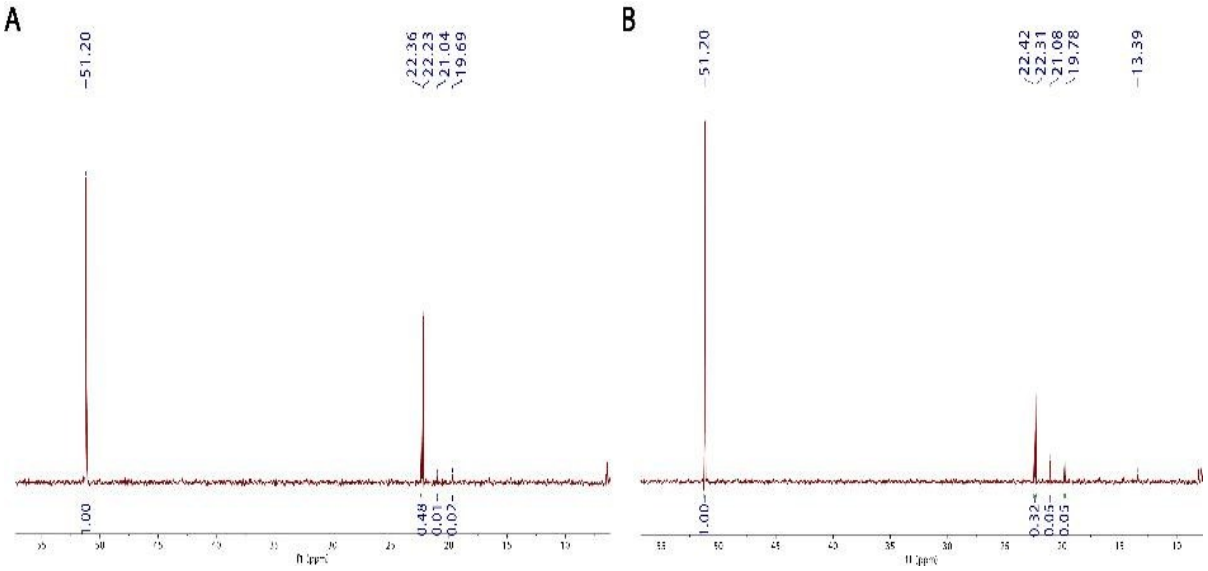


Figure A 23. ^{31}P NMR spectrum of cultures from *Kitasatospora fiedleri* DSM 114396 after MeCN extraction. A. ^{31}P NMR spectrum of concentrated MeCN layer. B. ^{31}P NMR spectrum of concentrated MeCN layer

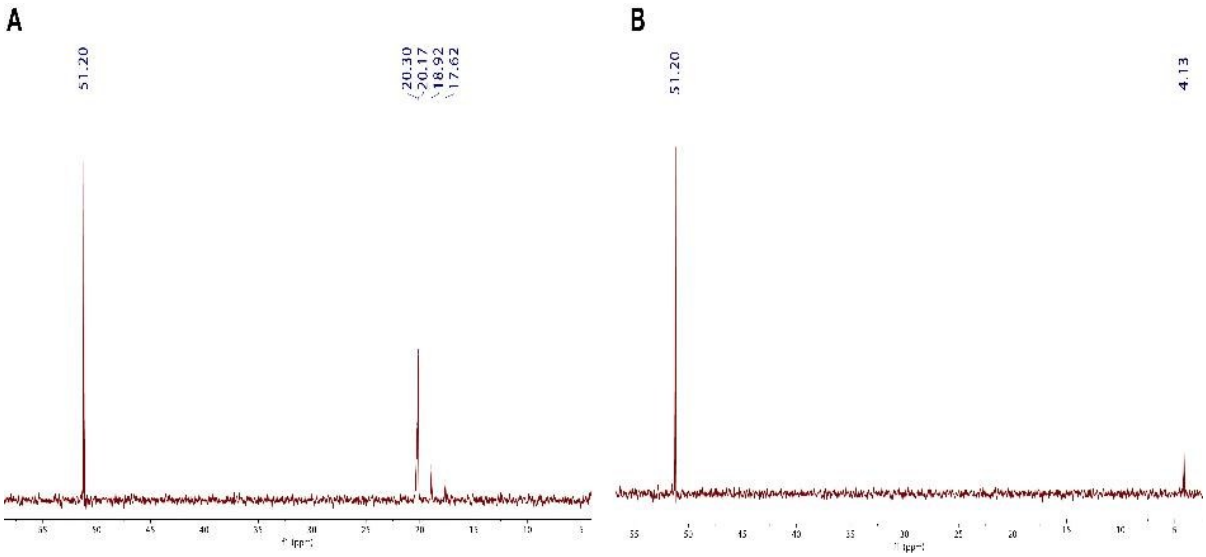


Figure A 24. ^{31}P NMR spectrum of *Kitasatospora fiedleri* DSM 114396 extracts following Hypercarb SPE fractionation. A. ^{31}P NMR spectrum of concentrated water elution fraction. B. ^{31}P NMR spectrum of concentrated MeCN elution fraction.

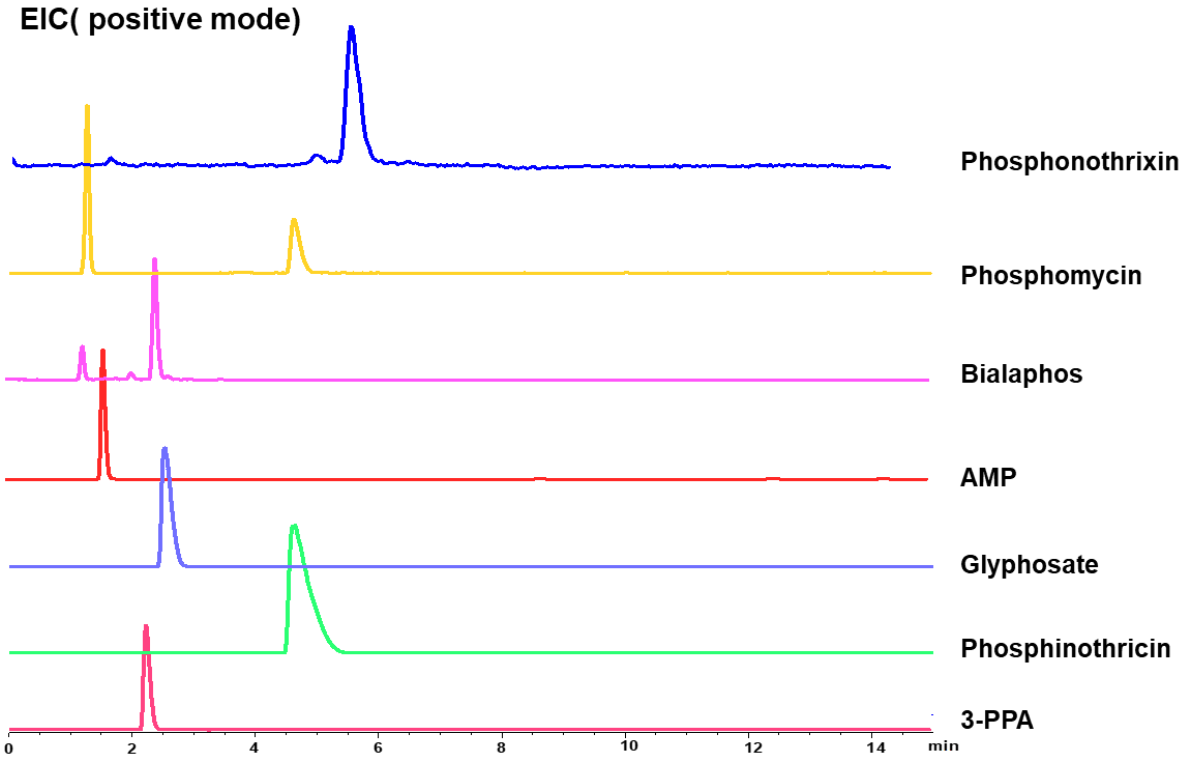


Figure A 25. EIC chromatograms of standard phosphonates (phosphonothrixin, phosphomycin, biolaphos, AMP, glyphosate, phosphinothricin and 3-PPA) under the column: Hypercarb™ column.

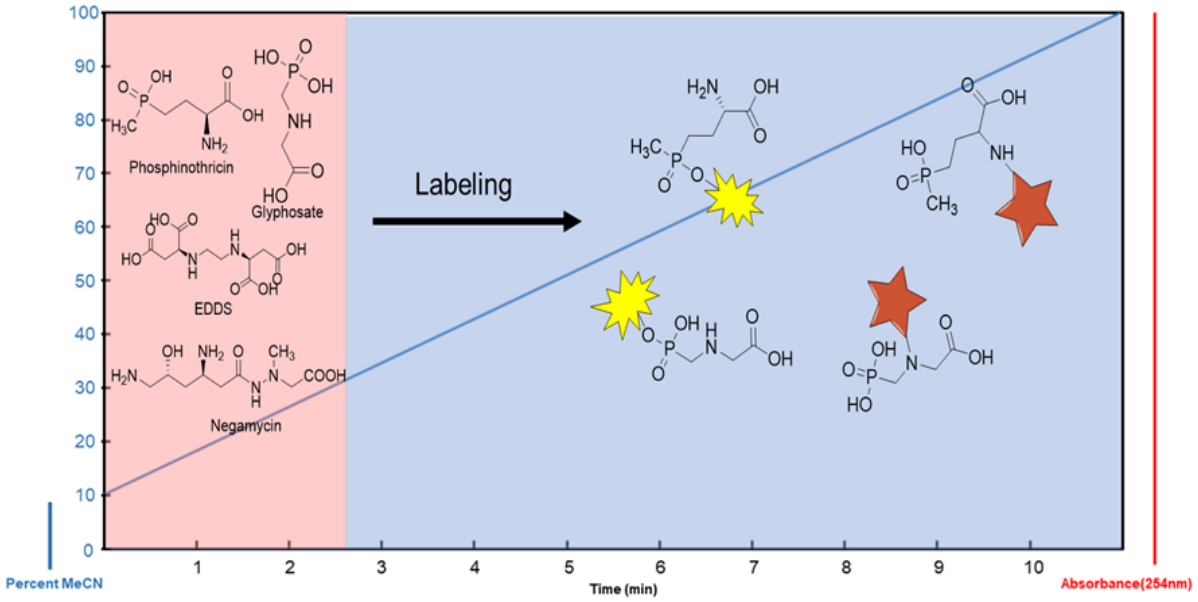


Figure A 26. Schematic diagram of chemical labeling strategies

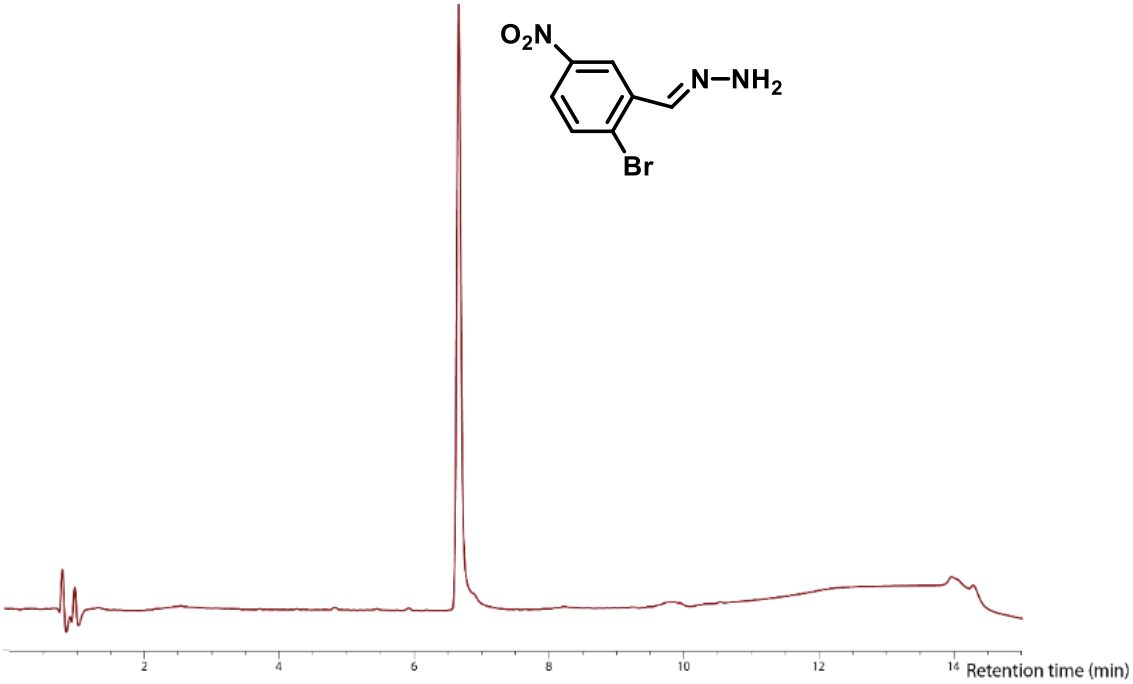


Figure **A 27**. Chromatograms of pure diazo compound **4** at 254nm

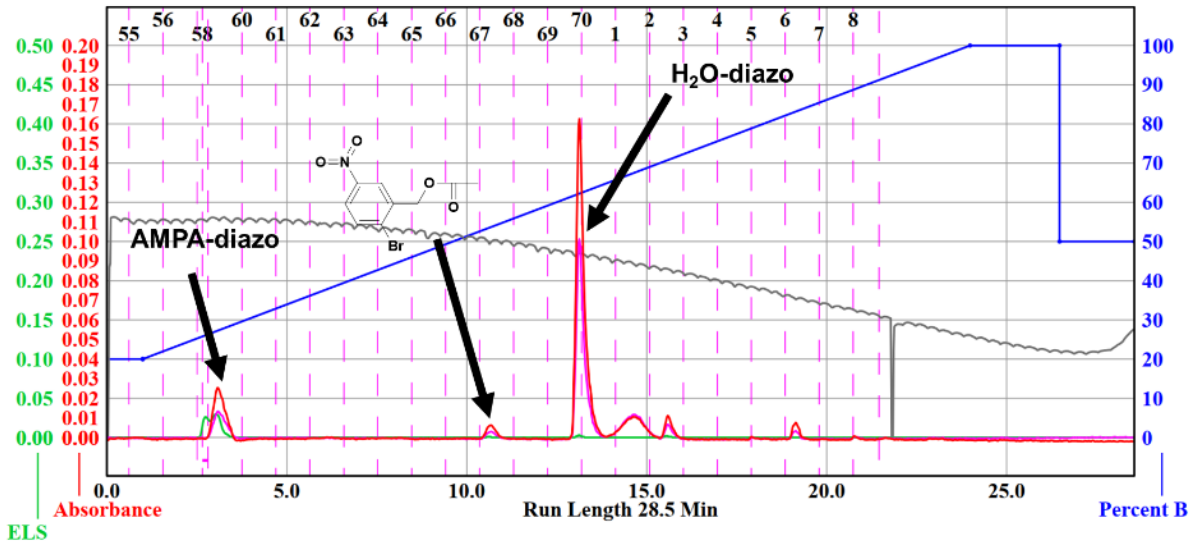


Figure A 28. Flash chromatograms of AMPA diazo compound with C18 column at 254nm

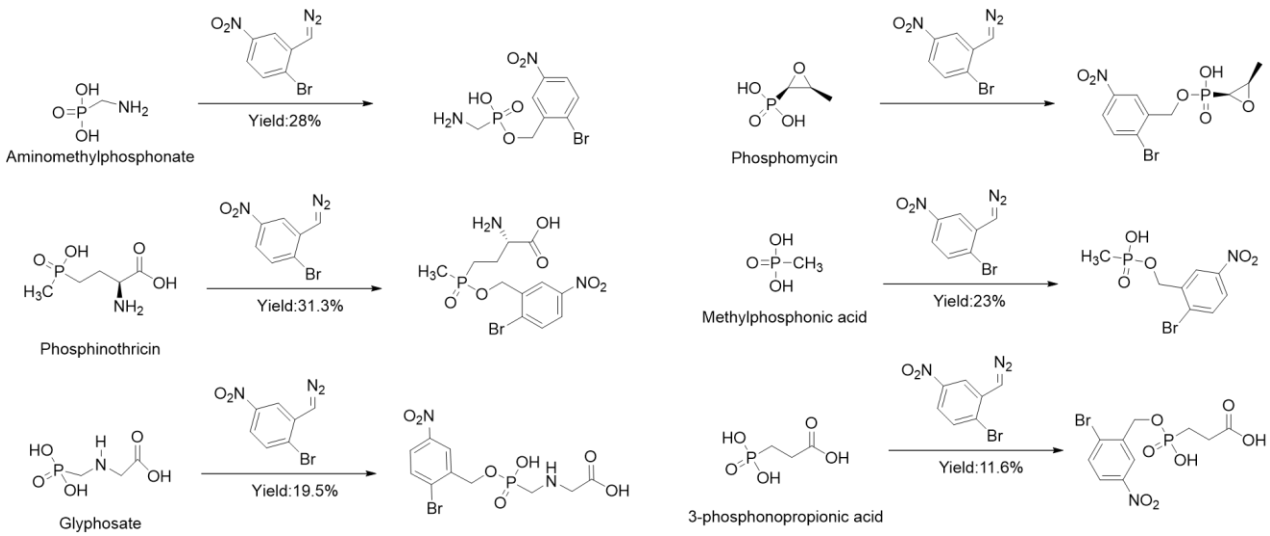


Figure A 29. Yields of standard phosphonates.

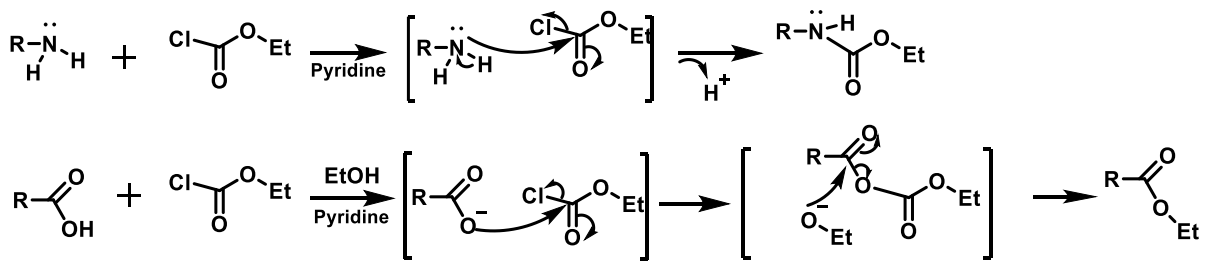


Figure A 30. Reaction mechanism of ECF labeling

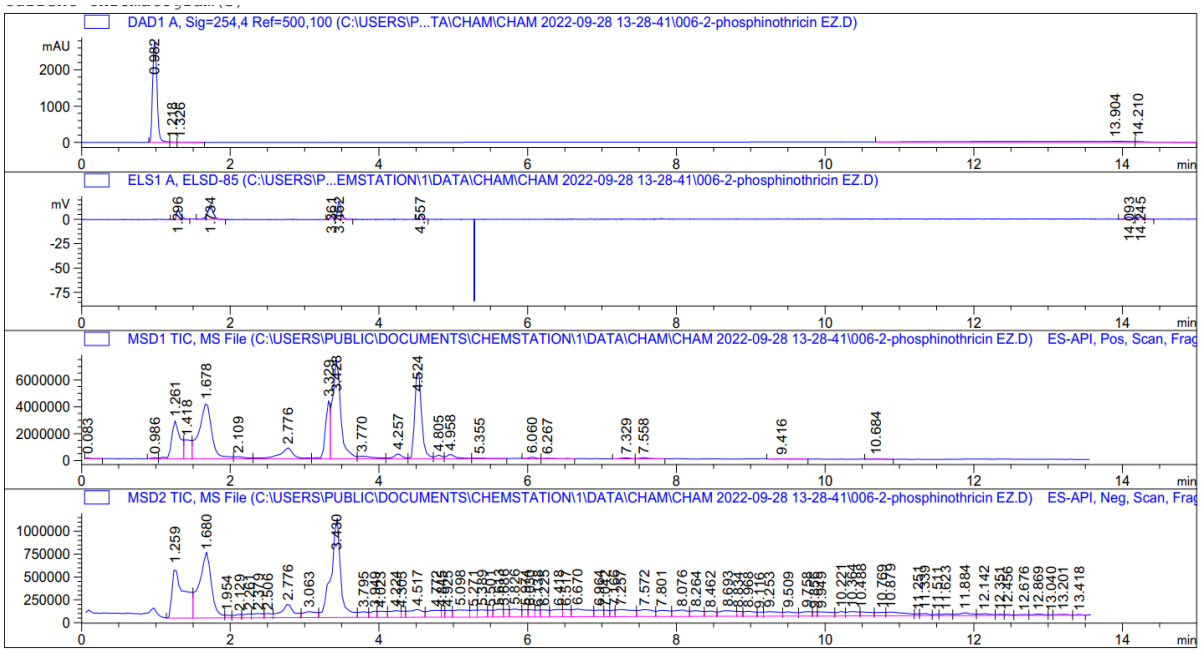


Figure A 31. Chromatograms of esterificated PT sample

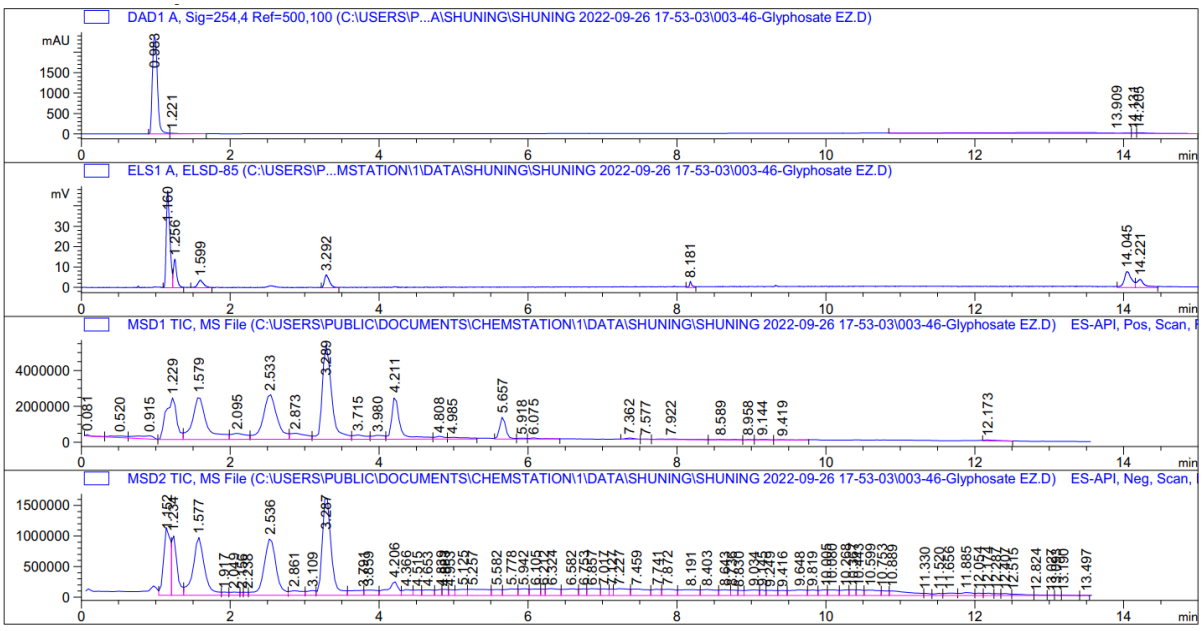


Figure A 32. Chromatograms of esterificated GP sample

A.3 Supplementary information to Chapter 3

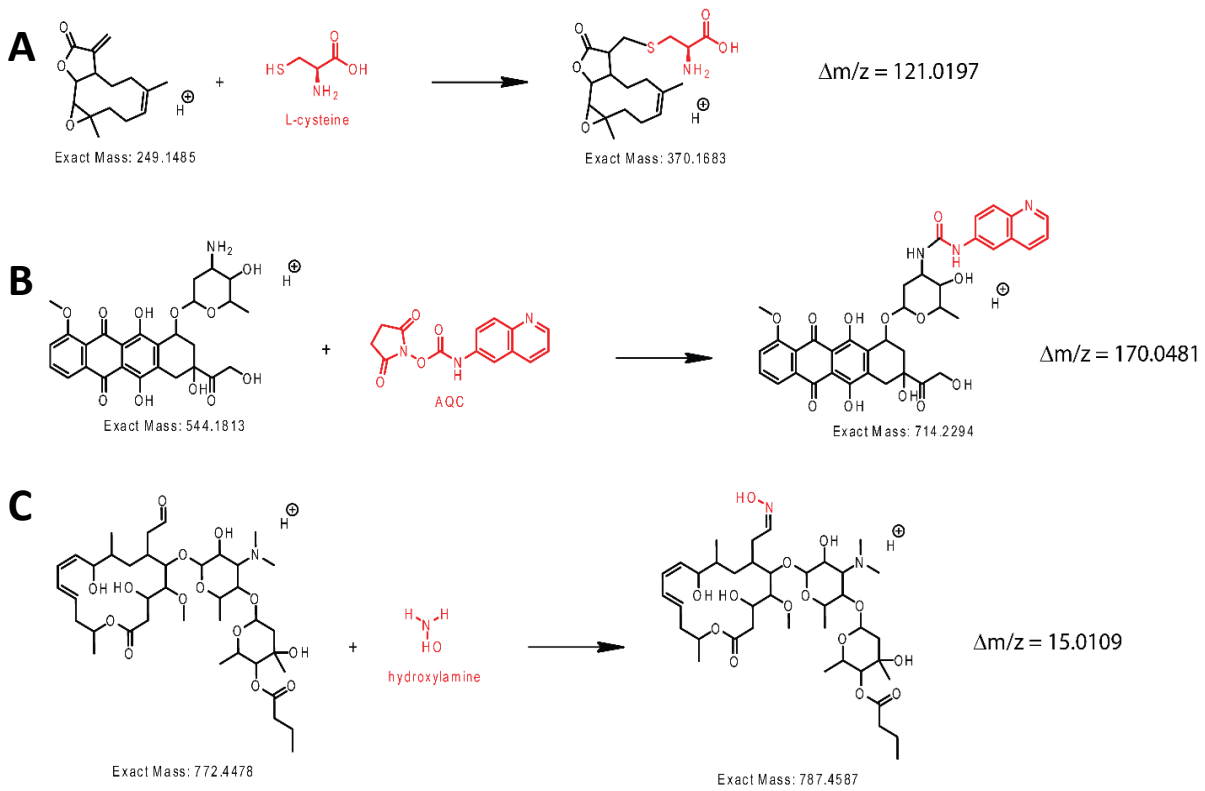


Figure A 33. Examples of the three derivatization reactions described in this manuscript with precursor and product *m/z* values and the corresponding expected mass shift of the [M+H]+ form. A) reaction A performed on parthenolide, B) reaction B performed on doxorubicin, and C) reaction C performed on kitasamycin with respective products.

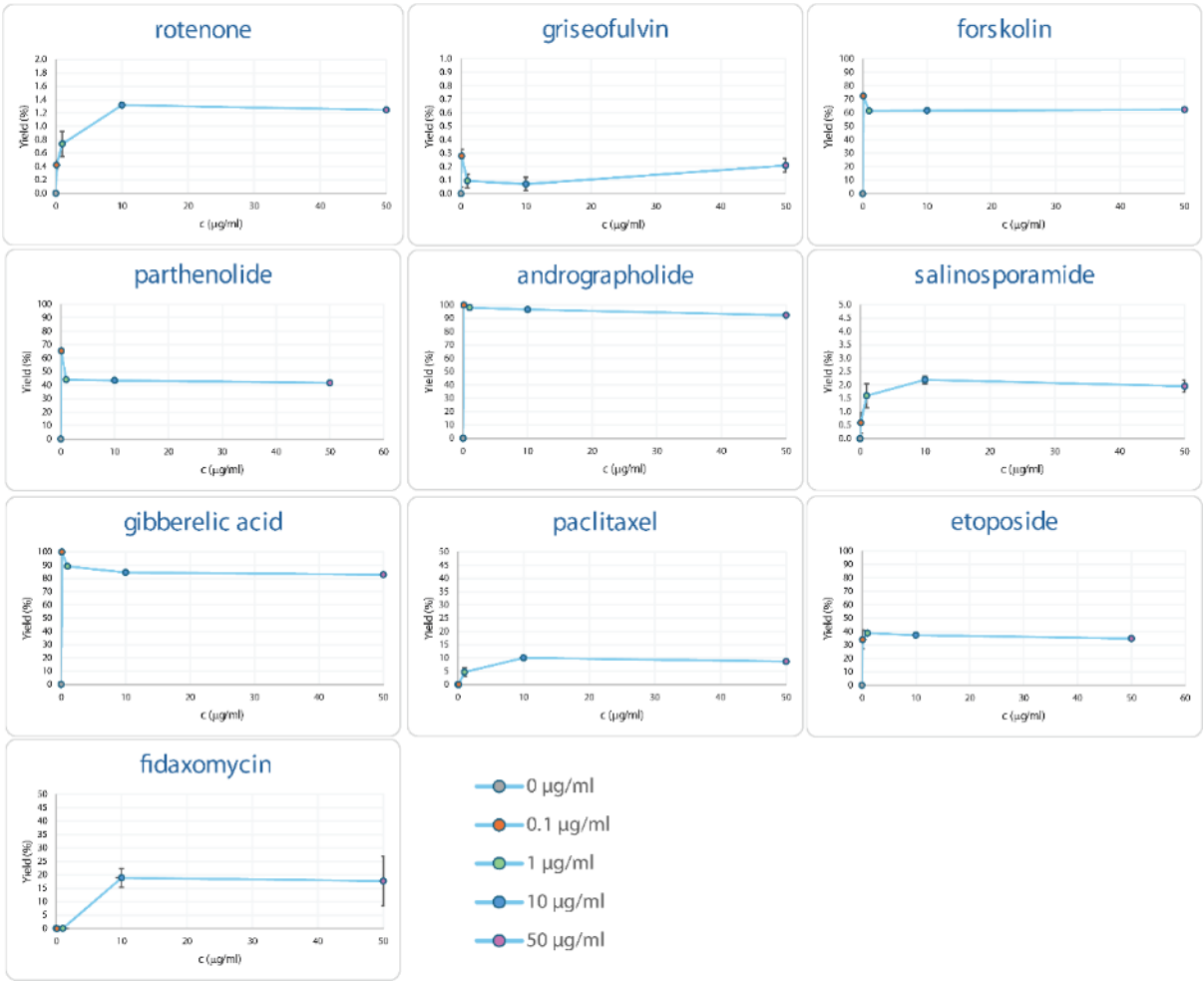


Figure A 34. Titration plots of the 32 standards reacting with L-cysteine 0.1 mM at different substrate concentrations as shown in the color chart. On the Y axis, it is reported the yield calculated on the [M+H]+ ions of the educts and products (except for fidaxomycin where [M+Na]+ (educt)/ [M+H]+(product) correlation was detected), while on the X axis the concentration in µg/ml of each standard in the mix. Error bars were drawn using the standard deviation between the two measurements.

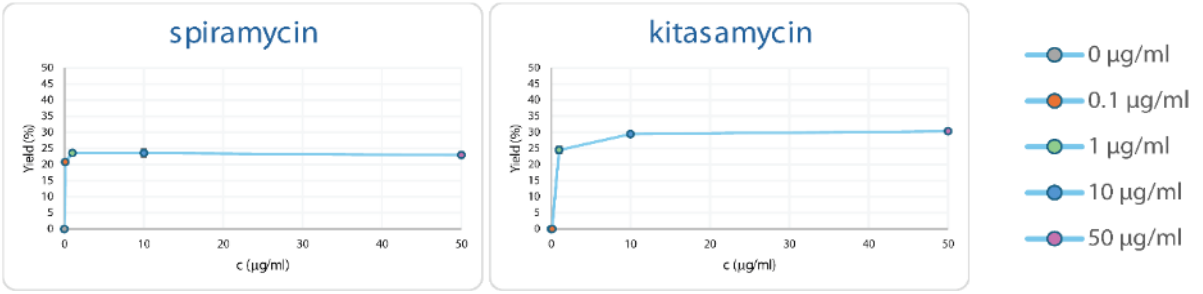


Figure A 35. Titration plots of 32 mix molecules reacting with hydroxylamine 10 mM at different concentrations (50, 10, 1 and 0.1 µg/ml). On the Y axis, it is reported the yield calculated on the [M+H] + ions of the unmodified and the derivatized molecule, while on the X axis the concentration in µg/ml of each metabolite in the mix.

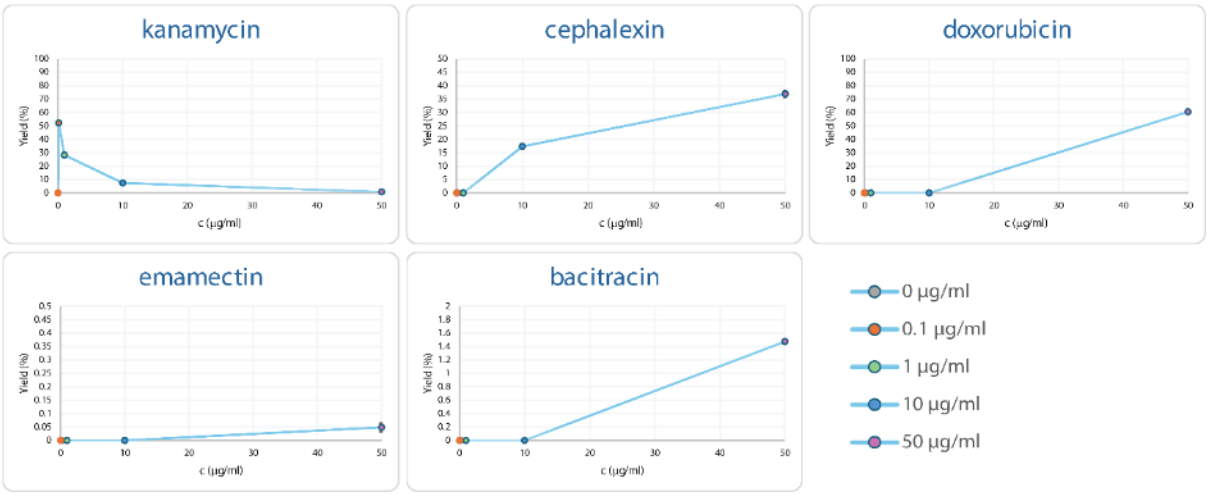


Figure A 36. Titration plots of 32 mix molecules reacting AQC 100 mM at different substrate concentrations as shown in the color chart. On the Y axis, it is reported the yield calculated on the $[M+H]^+$ ions of the educts and products (except for cephalexin where [M+Na]+ (educt)/ [M+H]+(product) correlation was detected), while on the X axis the concentration in µg/ml of each standard in the mix. Error bars were calculated using the standard deviation between the two measurements.

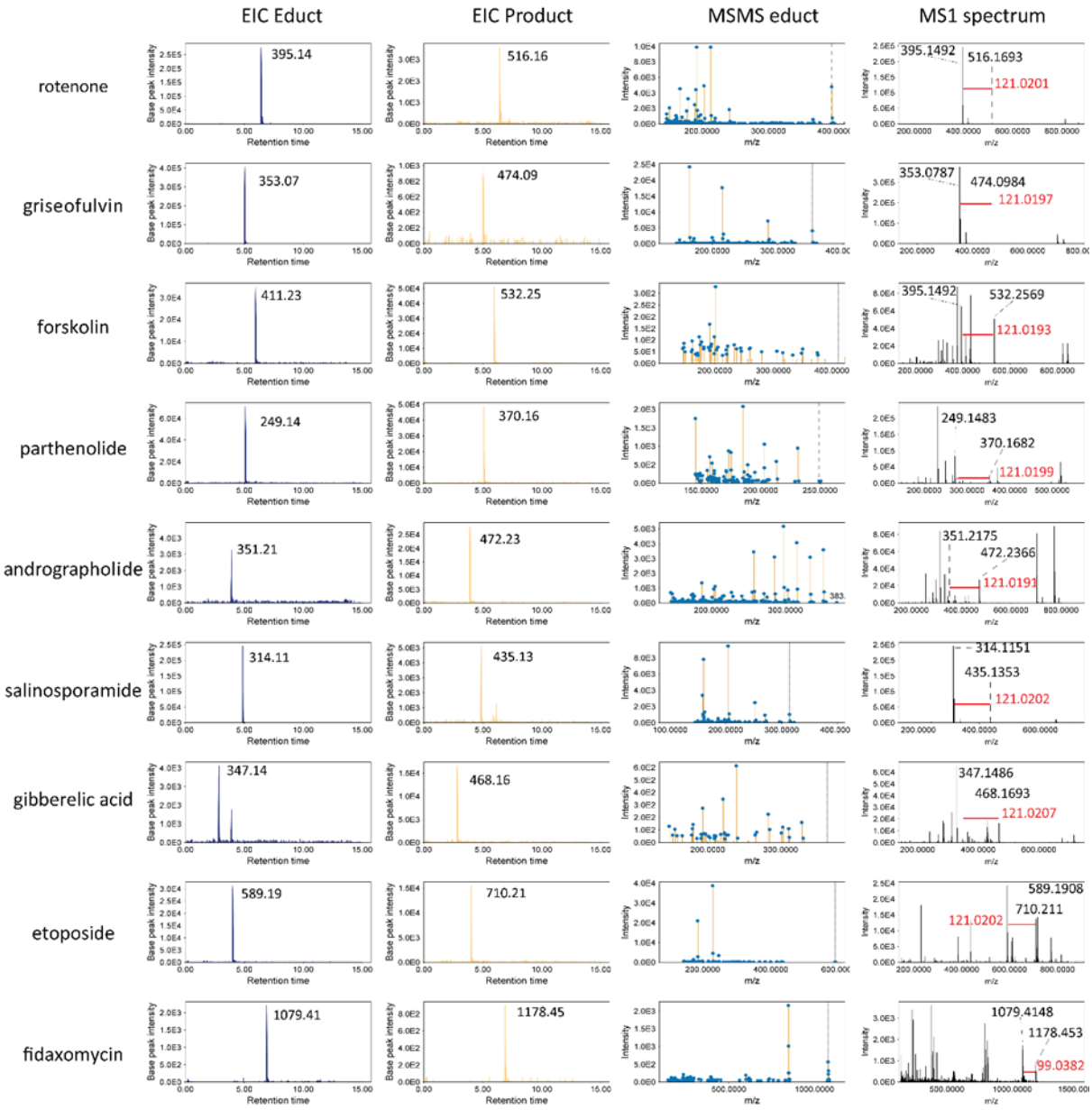


Figure A 37. EIC of the standards reacting with cysteine (Reaction A) in the 32 standards mixture, EIC of respective products correlated appearing at the same retention time in the treated sample, MSMS spectrum of the educts, and MS1 spectrum of the educts/products pairs observed at the same retention time in the treated samples, and connected via the peculiar reaction Δ*m/z* through the Online Reactivity. For all the molecules the reactivity was detected via the correlation between precursor and product [M+H]+ ions (calcld. Δ*m/z* = 121.0197), except for fidaxomycin (at the bottom) where [M+Na]+ (educt)/ [M+H]+(product) correlation was detected (calcld. Δ*m/z* = 99.03781).

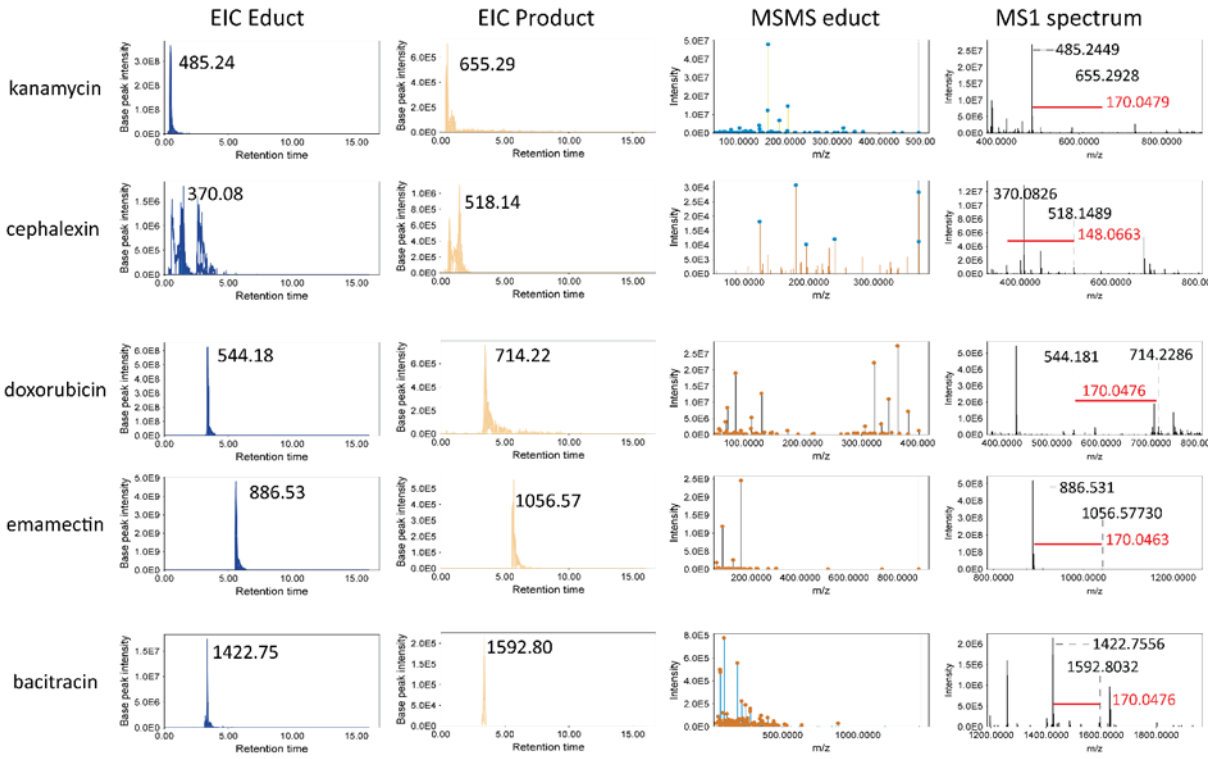


Figure A 38. EIC of molecules reacting with AQC (Reaction B) in the 32 standards mixture, respective products appearing at the same RT, MSMS spectrum of the educts, and MS1 spectrum of the educts/products pairs observed at the same retention time in the treated samples, and connected via the peculiar reaction Δ*m/z* through the Online Reactivity. For all the molecules the reactivity was detected thanks to the correlation between educt and product [M+H]+ ions (calcld. Δ*m/z* = 170.0481), except for cephalexin where [M+Na]+ (educt)/ [M+H]+(product) correlation was detected (calcld. Δ*m/z* = 148.0661).

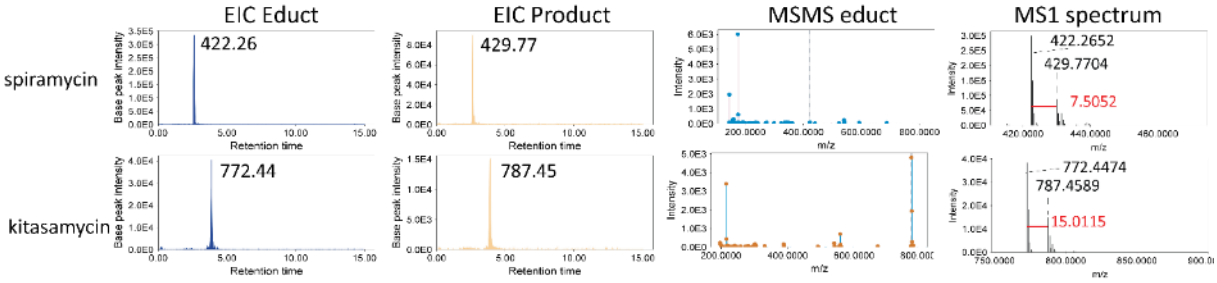


Figure A 39. EIC of the molecules reacting with hydroxylamine (Reaction C) in the 32 standards mixture, respective products appearing at the same RT, MSMS spectrum of the educts, and MS1 spectrum of the educts/products pairs observed at the same retention time in the treated samples, and connected via the peculiar reaction Δ*m/z* through the Online Reactivity. For kitasamycin the reactivity was detected thanks to the correlation between educt and product [M+H] + ions (calcld. Δ*m/z* = 15.0109), for spiramycin the correlation between the two [M+2H]2+ was detected (calcld. Δ*m/z* = 7.5054).

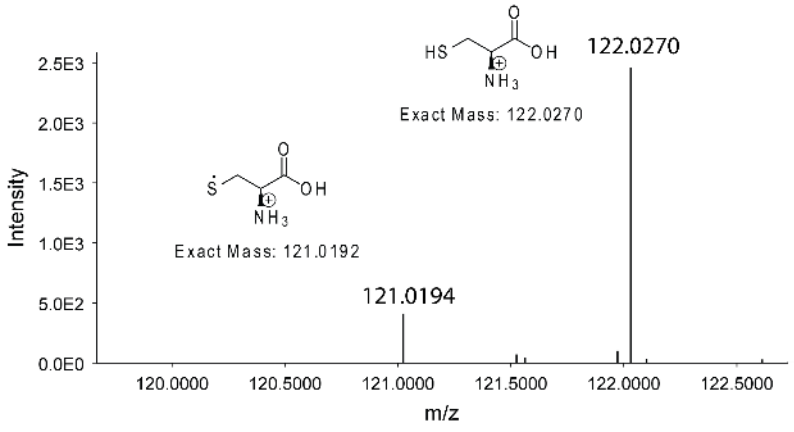


Figure A 40. Mass spectrum and structures of cysteine cation radical [M+H] + and protonated [M+H]+ forms, both were observed when during blank (MeOH) run with a continuous post-column injection of L-cysteine 0.1 mM (*m/z* range 100-300). MeOH without post-column injection of L-cysteine was also run, and none of the two ions was observed.

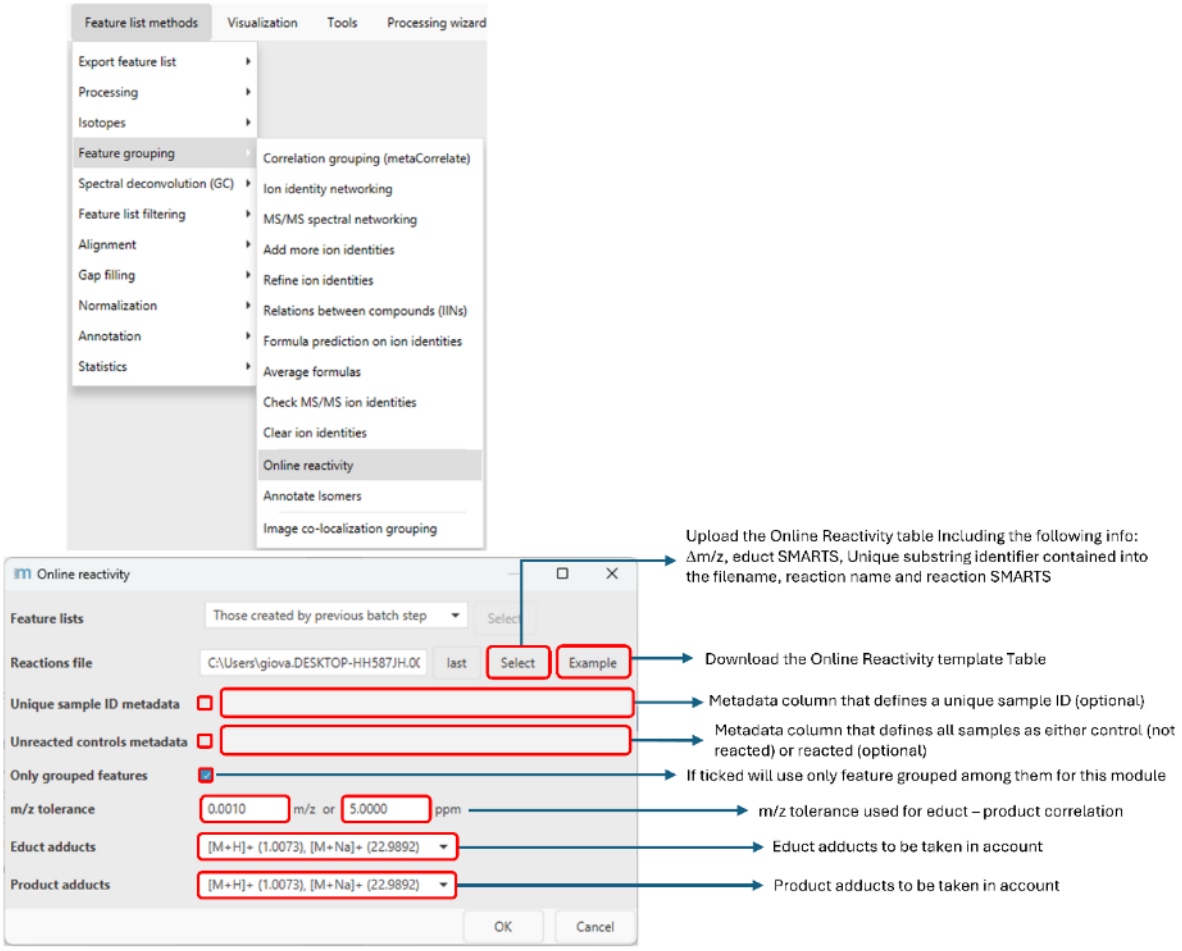


Figure A 41. **Mzmine Online Reactivity module.** Visualization of the Online Reactivity module of Mzmine4 built to automatically connect educt and products during an online reaction based on the typical D*m/z*. By clicking on the example button a template of the input table with the essential parameters for this module is provided, here the different reaction features are defined (I.e. Δ*m/z*, educt SMARTS, reaction SMARTS). If ticked the “Unique sample ID” and the “Reaction sample Type” can be used. The user can make use of the correlation produced by the meta correlation module if run before this one by ticking “Only grouped features”, moreover he can define the *m/z* tolerance and the allowed adducts for both educt and products.

Figure A 42. **Features visualization in Mzmine 4 Online Reactivity.** In the feature list, it is possible to activate the visualization of the online reaction attributes, such as the observed online reaction (including the ions involved) with the respective Δ*m/z*, educt, and product IDs, the status of the feature we are observing (educt or product) the putative educt substructure or the entire reaction (in the form of SMARTS). Here are shown as an example A) parthenolide [M+H]+ ion row and B) parthenolide derivatization product [M+H]+ with cysteine row, automatically connected by the Online reactivity module. By clicking on the feature, it is possible to visualize the relative networks where the cyan node represents the reaction product. C) Reactivity network of parthenolide.

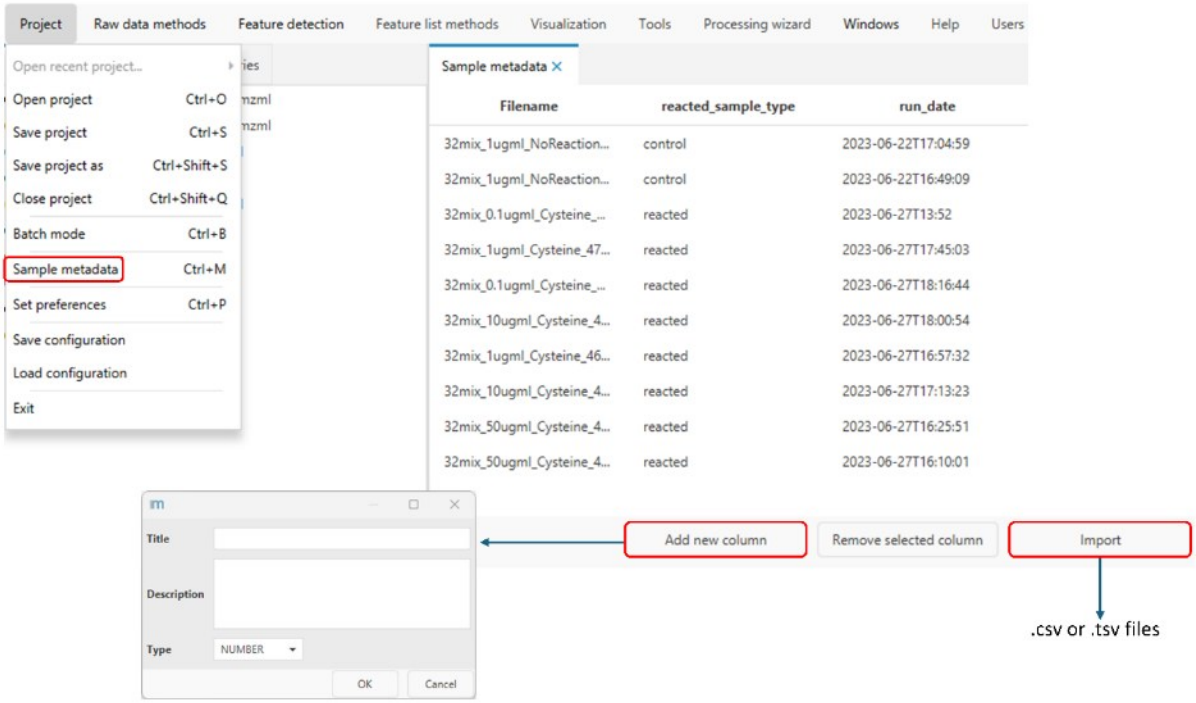


Figure A 43. Mzmine 4 Metadata. Once a dataset is uploaded a simple Sample metadata table is automatically created and can be accessed by clicking on Project/Sample metadata. This metadata file originally contains the Filename and the run data columns, however, it is possible to either import a metadata table prepared offline in the .csv and .tsv format or to manually add additional columns.

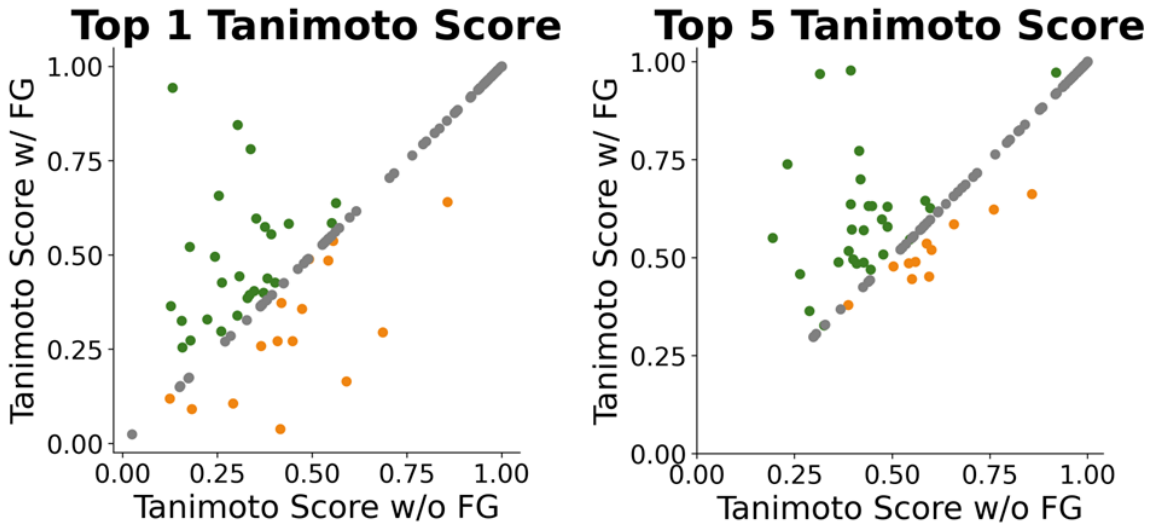


Figure A 44. **Impact of MCheM on GNPS2 analog search for the experimental dataset.** The MCheM method improved the structural similarity of the Top 1, Top 5 analogs as reflected by the improved Tanimoto scores (green dots on the scatter plots), while a lower number of worsened features were retrieved (orange dots on the scatter plots), leading to an improved average Tanimoto score after FG-filtering compared with not filtered data (violin plots Figure 2 in the main text).

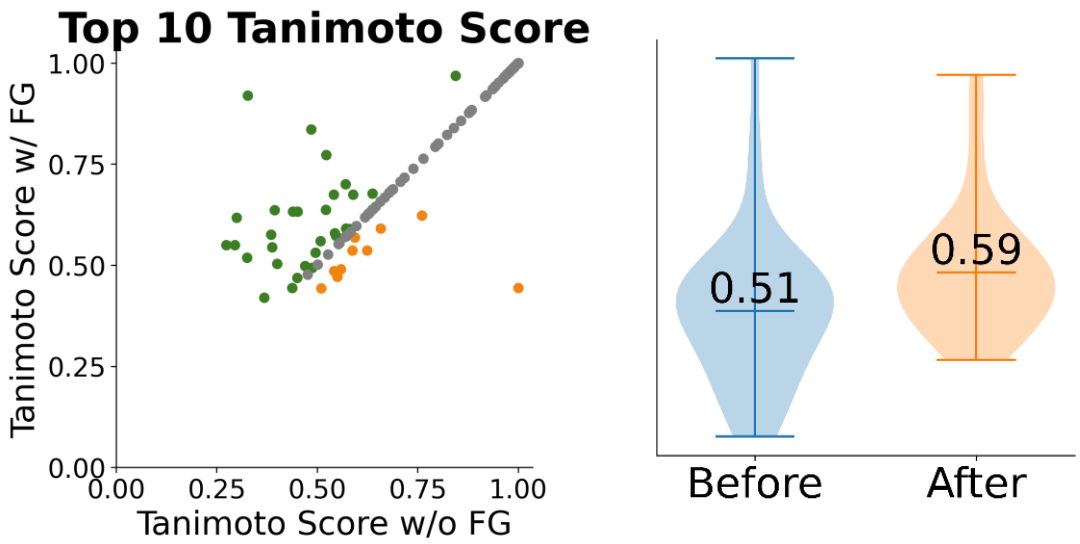


Figure A 45. **Impact of MCheM on Top 10 analog search on GNPS2 for the experimental dataset.** The MCheM method improved the similarity of Top 10 analogs with respect to the actual structure, as shown by the improved features reported as green dots in the scatter plot, while worsened annotations are reported as orange dots. Overall, MCheM showed an improvement in the average Tanimoto similarity score when compared to the classic annotation as depicted in the violin plot with the median value going from 0.51 to 0.59.

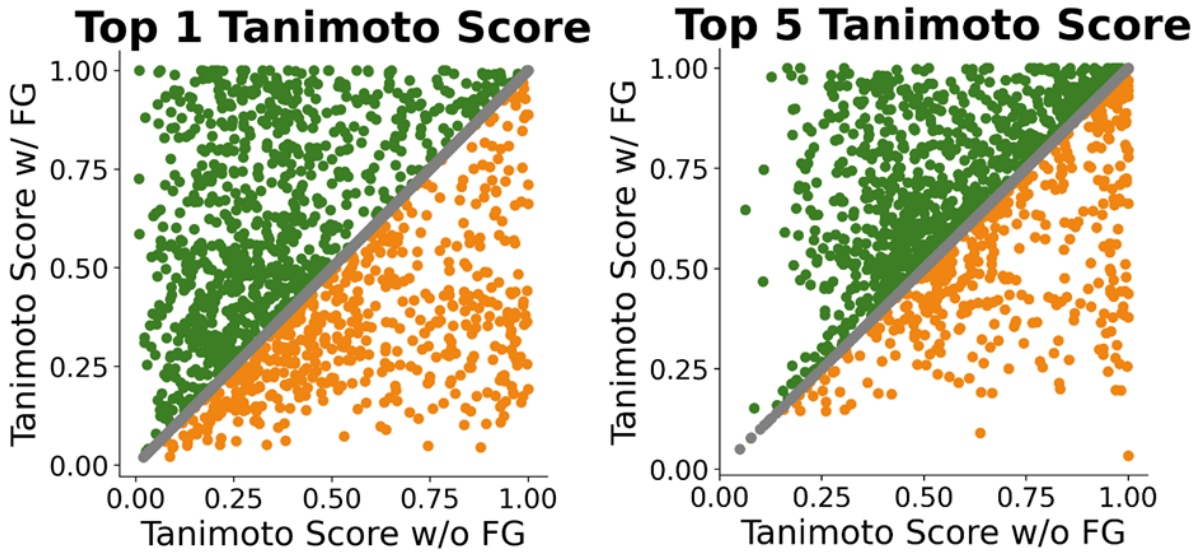


Figure A 46. **Impact of MCheM on GNPS2 analog search for the CANOPUS dataset.** The MCheM method improved the structural similarity of the Top 1, Top 5 analogs as reflected by the improved Tanimoto scores (green dots on the scatter plots), while a lower number of worsened features were retrieved (orange dots on the scatter plots), leading to an improved average Tanimoto score after FG-filtering compared with not filtered data violin plots Figure Y in the main text)

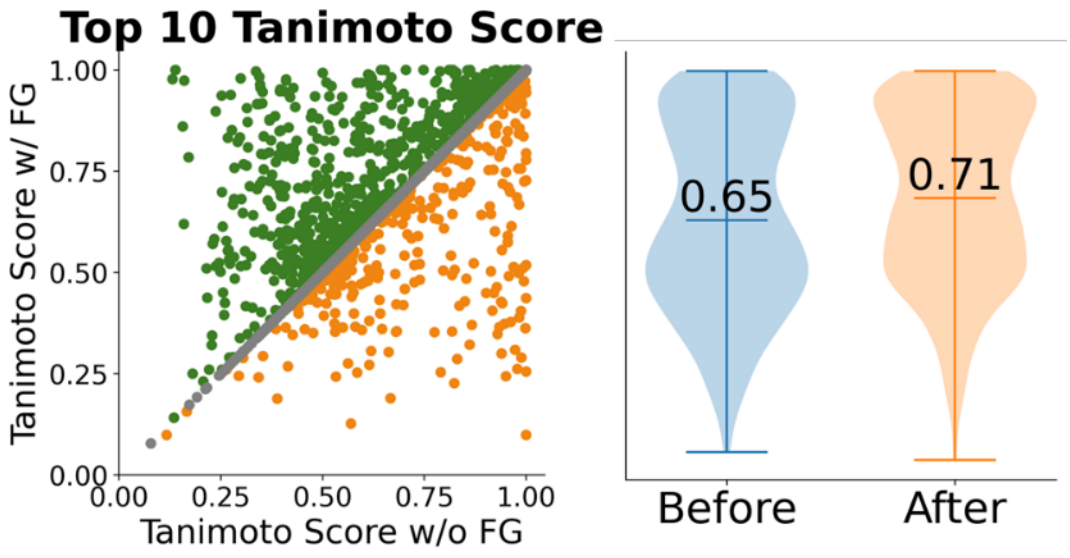


Figure A 47. Impact of MCheM on Top 10 analog search on GNPS2 for the CANOPUS dataset. The MCheM method improved the similarity of Top 10 analogs with respect to the actual structure, as shown by the improved features reported as green dots in the scatter plot, while worsened annotations are reported as orange dots. Overall, MCheM showed an improvement in the average Tanimoto similarity score when compared with the original annotation as depicted in the violin plot with median value going from 0.65 to 0.71.

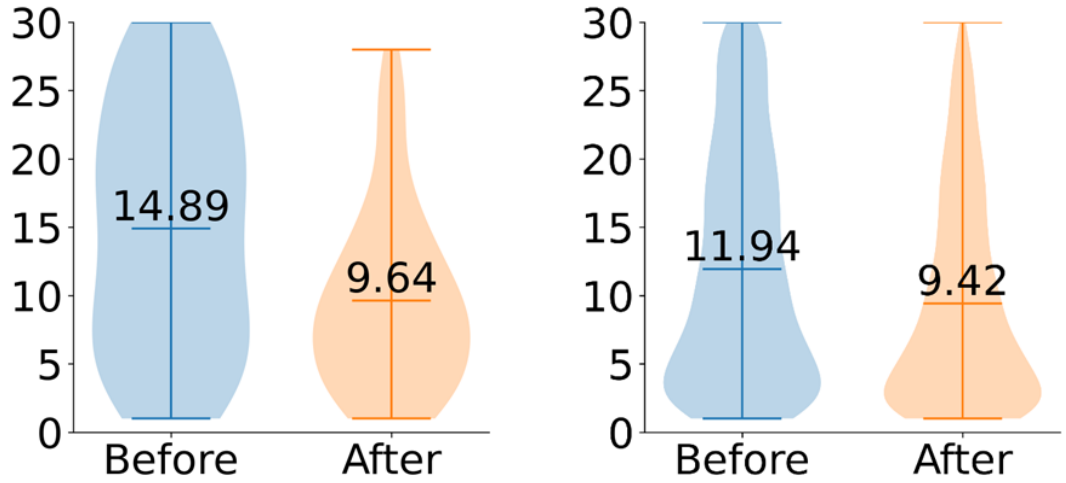


Figure A 48. Impact of MCheM on the average ranking of the most similar structure. The MCheM method improved the average ranking of the most similar structure (orange violin) compared with the classic data (blue violin) both in the experimental (left plot) and CANOPUS (right plot) datasets. The median values improved from 14.89 to 9.64 and from 11.94 to 9.42, and the shape also became narrower at the bottom for both datasets, which means a greater concentration of values closer to 1.

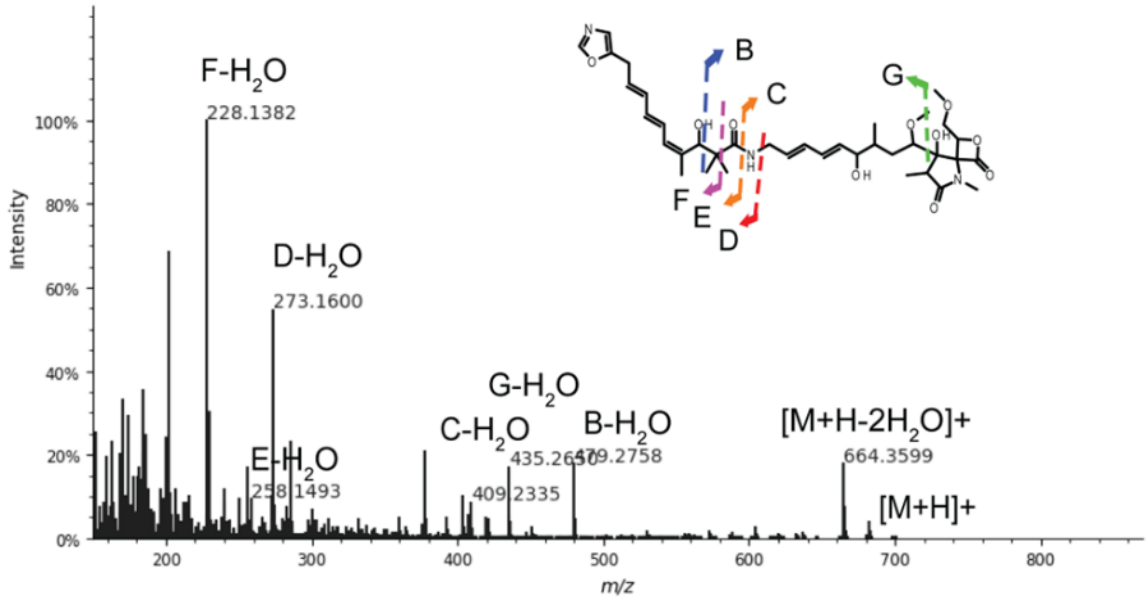


Figure A 49. MSMS spectrum of putative oxazolomycin D (ID 1911) with its diagnostic fragments.

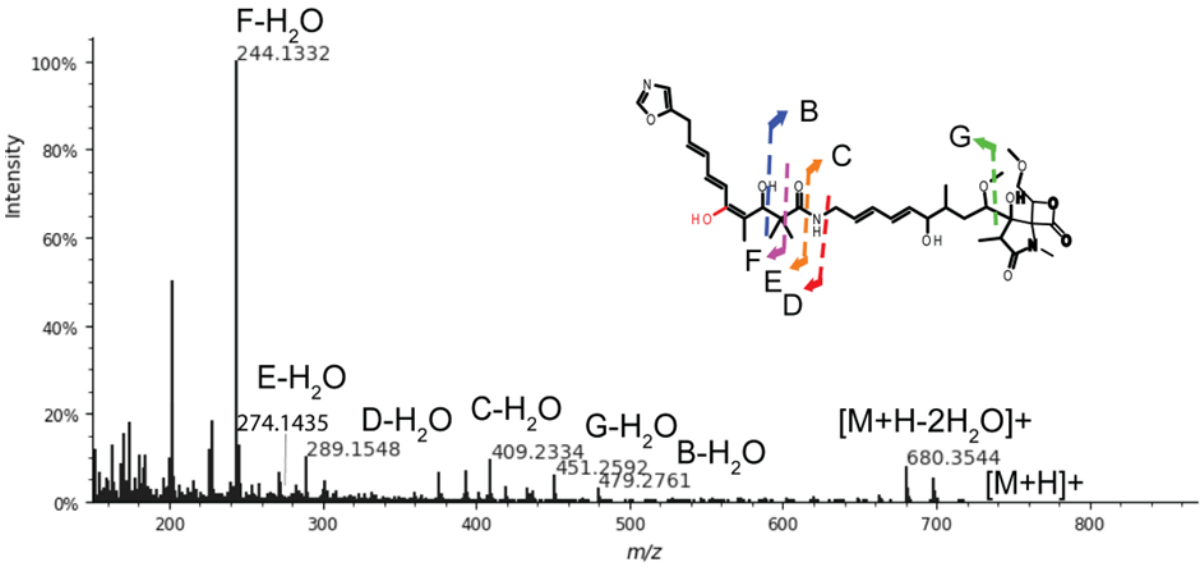


Figure A 50. MSMS spectrum of oxazolomycin D derivative with *m/z* 716.3754 (ID 1562) with its diagnostic fragments. Unshifted C and B fragments allowed us to predict the modification on the left part of the molecule, as confirmed by a shift of F, E and D fragments.

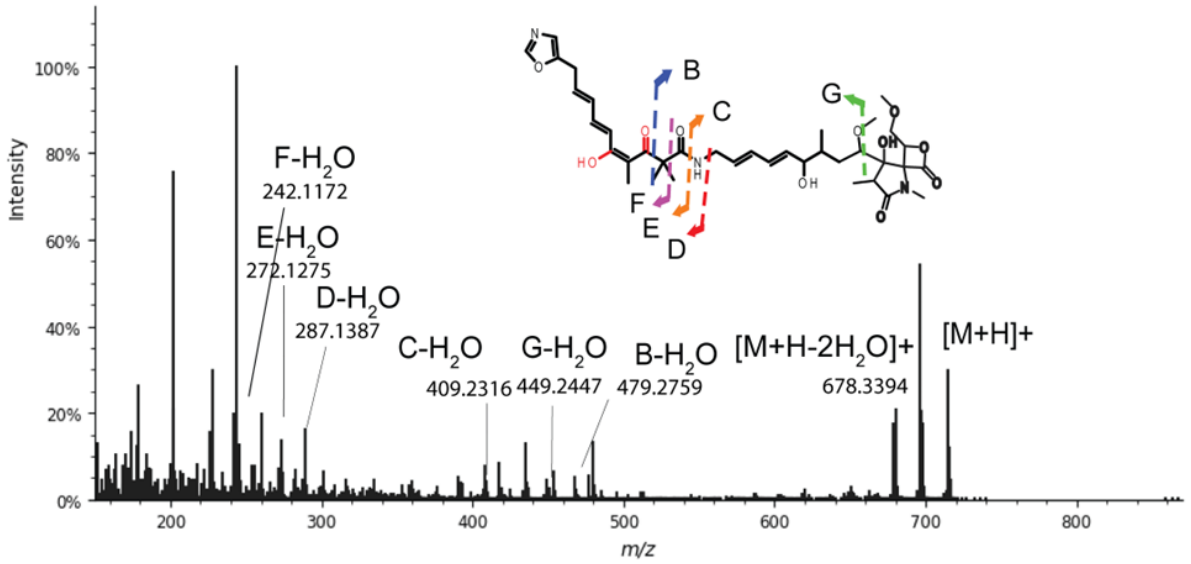


Figure A 51. MSMS spectrum of oxazolomycin D derivative with *m/z* 714.3598 (ID 1661) with its diagnostic fragments. Unshifted C and B fragments allowed us to predict the modification on the left part of the molecule, as confirmed by a shift of F, E and D fragments.

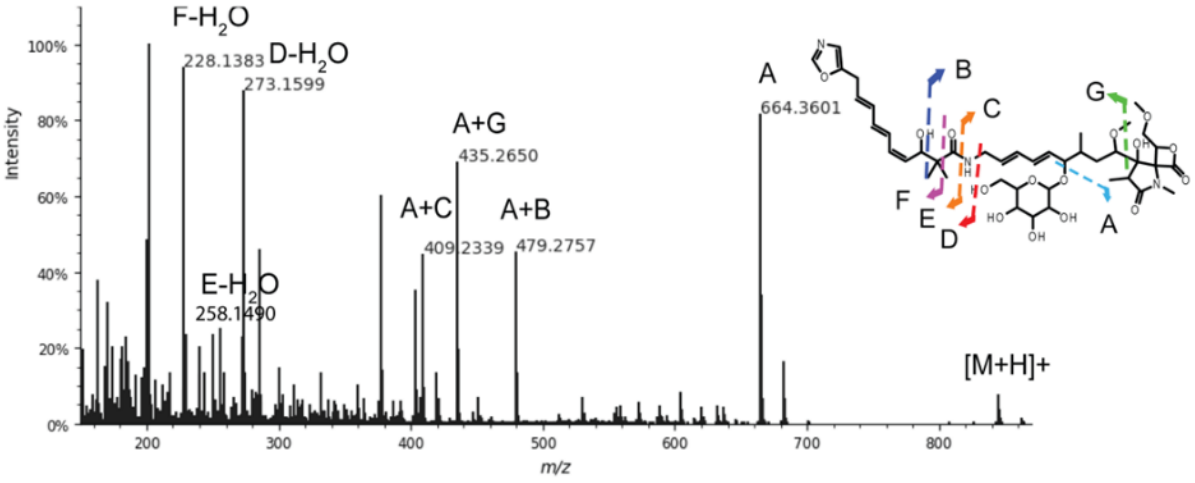


Figure A 52. MSMS spectrum of 7-glycosyl oxazolomycin (1) (ID 1569) with its diagnostic fragments.

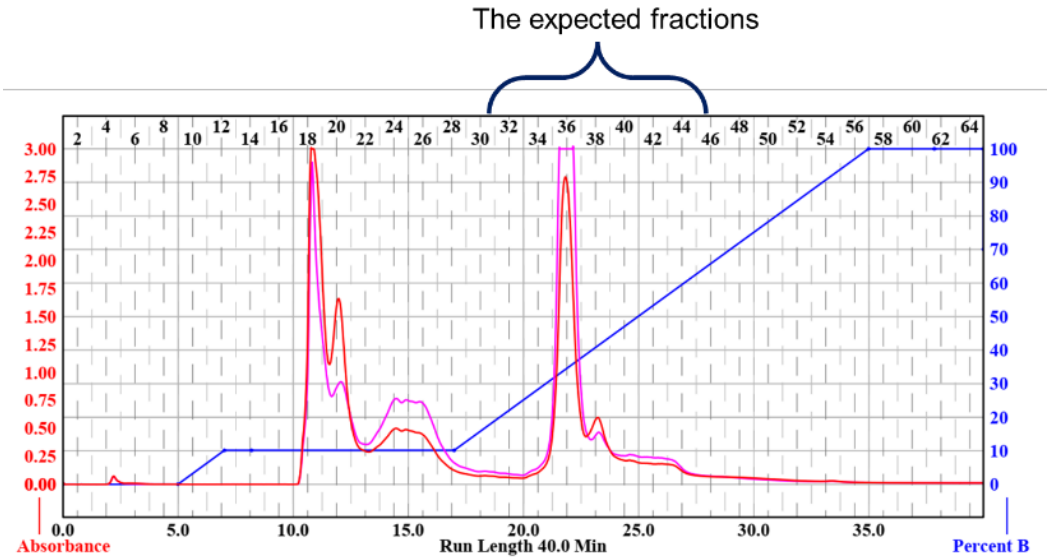


Figure A 53. Flash Chromatography of the extract with silica RediSep column.

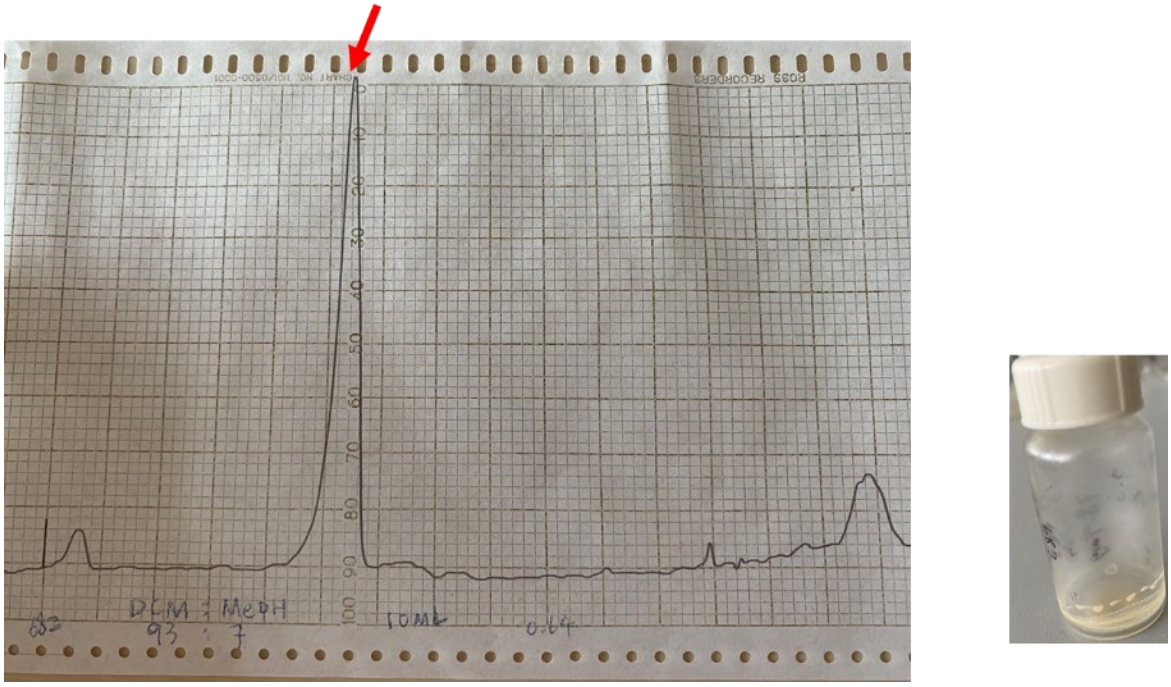
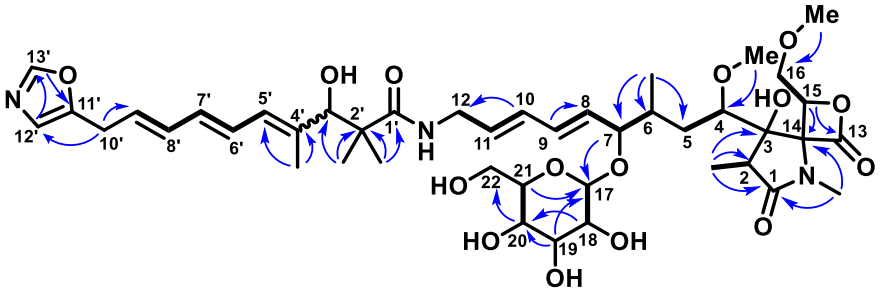


Figure A 54. Chromatogram of the concentrated fractions in 254nm.The target peak is pointed by red arrow. The substance on the right is pure compounds.

Table 3. NMR spectral data for 7-glycosyl oxazolomycin D in DMSO-d_6 at 700 MHz



carbon #	δ_C	δ_H, mult (*J* in Hz)	COSY	HMBC
1	174.2	--	--	--
2	43.3	2.43, q (7.3)	2-Me	1,2-Me
3	80.8	--	--	--
4	82.4	3.27, t (4.9)	5	4-OMe
5a	31.4	1.81, m	4	6-Me
5b		1.14, m	4	6-Me
6	32.7	1.95, m	7	--
7	80.9	4.16, m	6,8	6-Me,17
8	130.0	5.62, m	7,9	--
9	131.7	6.20, m	8	8
10	130.2	6.14, m	11	12
11	130.5	5.65, m	10,12	10,12
12	40.5	3.72, m	11	--
13	170.2	--	--	--
14	85.1	--	--	--
15	79.0	4.96, dd (8.9, 2.7)	16	3,13,14,16
16a	71.5	4.14, dd (12.1, 9.0)	15	15,16-OMe
16b		4.00, dd (12.3, 2.7)	15	--
17	100.3	4.20, d (7.6)	18	7
18	73.7	2.95, m	17	17,19,20
19	77.0	3.13, s	--	17,18,20
20	69.9	3.08, m	21	18,19,22
21	76.9	2.99, m	20	17,20,22
22a	60.8	3.58, dd (11.9, 2.4)	21	20,21
22b		3.44, m	21	20,21
1-NMe	26.1	2.83, s	--	1,14
2-Me	9.6	1.00, d (7.4)	--	1,2,3
4-OMe	56.0	3.15, s	--	4
6-Me	14.5	0.88, dd (7.0, 2.8)	--	5,6,7
16-OMe	58.5	3.29, s	--	16
1’	176.1	--	--	--
2’	45.9	--	--	--
3’	73.1	4.63, d (5.3)	--	--
4’	140.0	--	--	--
5’	123.5	6.40, brd (12.1)	6’	--
6’	124.5	6.32, t(11.4)	5’,7’	--
7’	127.1	5.93, t(11.4)	6’,8’	--
8’	128.1	6.75, m	7’,9’	--
9’	129.0	5.79, m	8’	--
10’	28.3	3.55, d(6.9)	--	9’,11’,12’
11’	150.6	--	--	--
12’	122.0	6.89, m	--	11’,13’
13’	151.4	8.23, s	--	12’,11’
2’-Mea	21.5	0.97, m	--	1’,2’,2’-Meb,3’
2’-Meb	24.7	1.10, m	--	3’,2’,2’-Mea,3’
4’-Me	20.0	1.73, s	--	3’,4’,5’

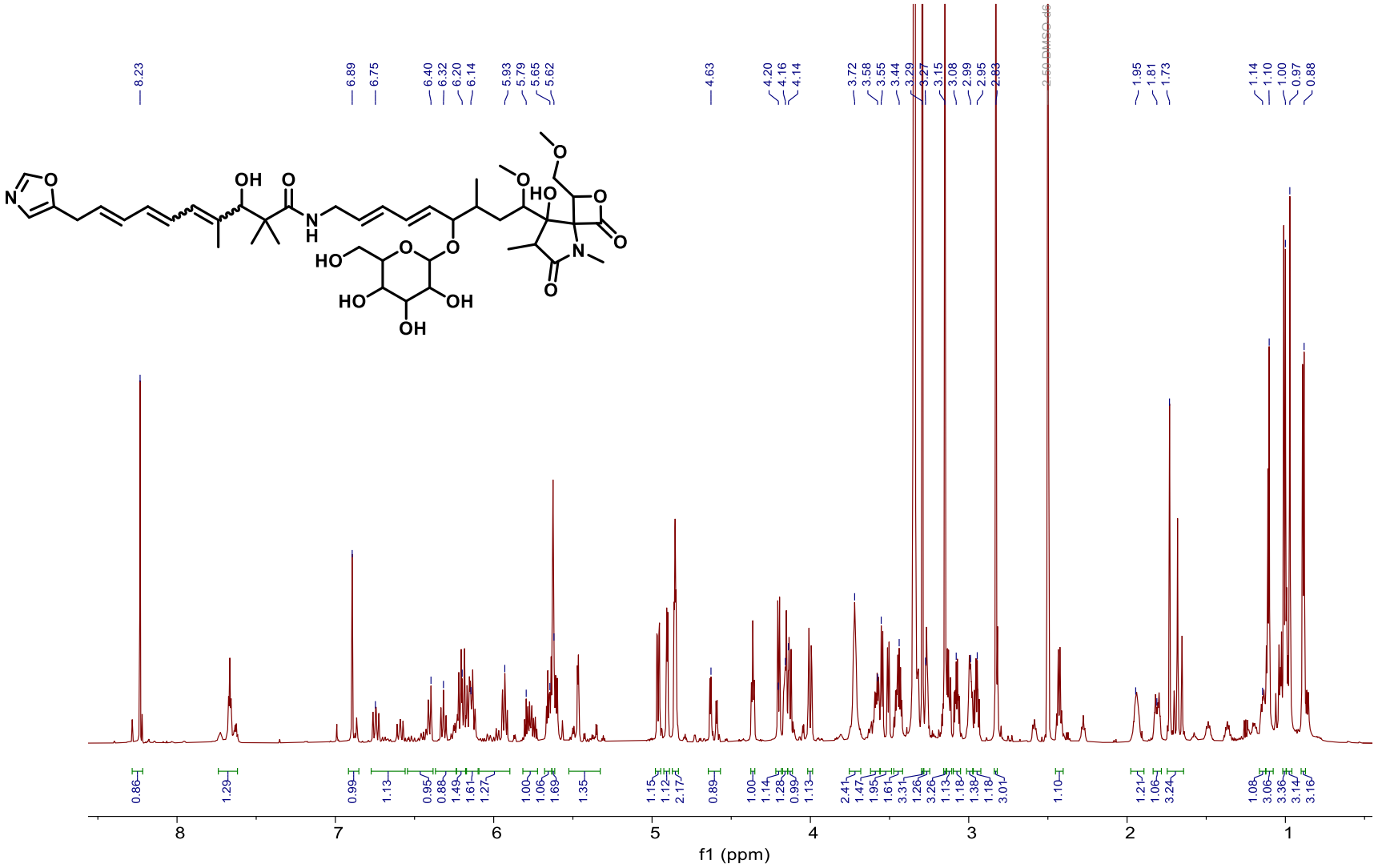


Figure A 55. ^{1}H NMR (DMSO-d_6, 700 MHz) of 7-glycosyl oxazolomycin D

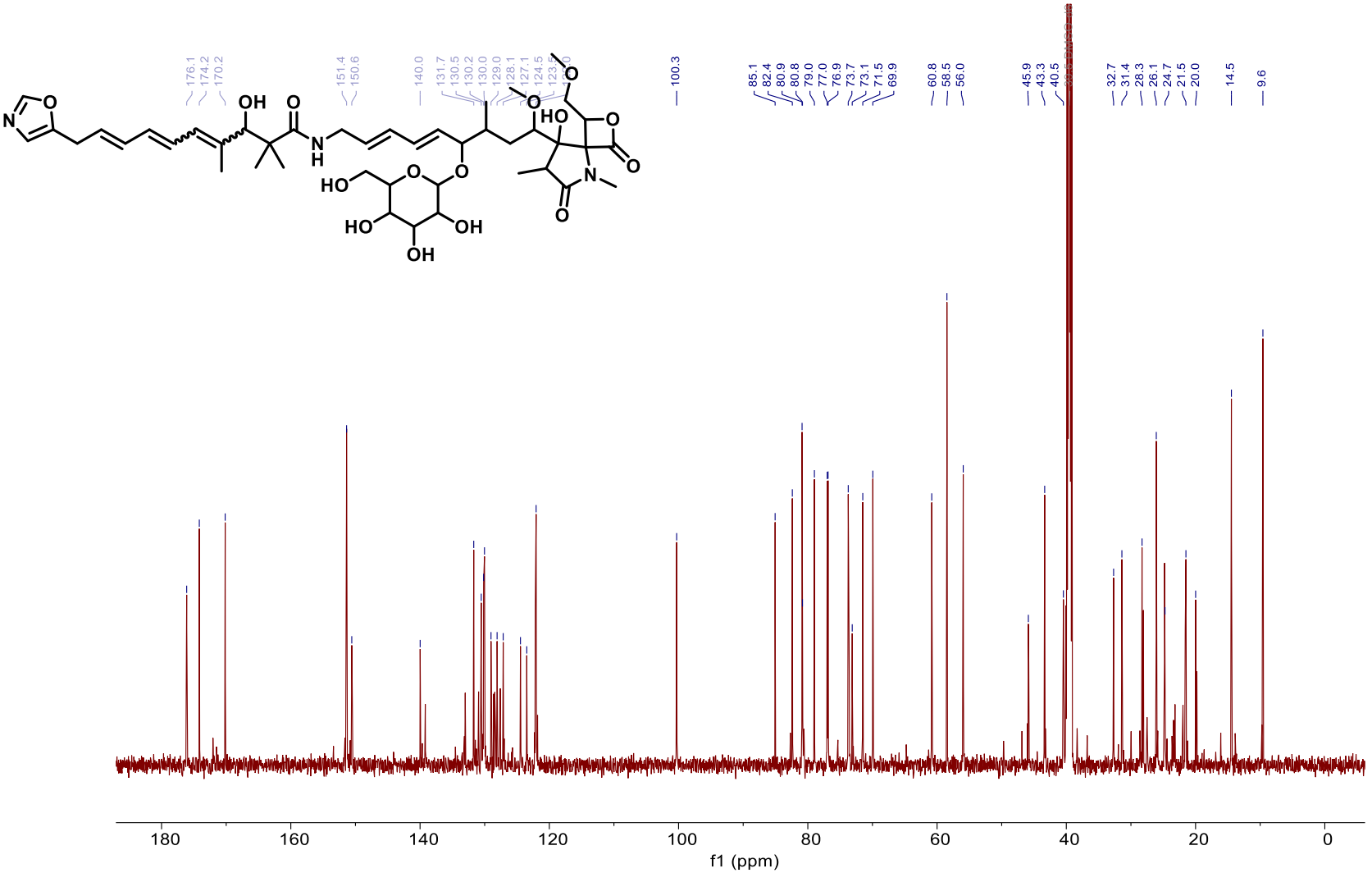


Figure A 56. ^{13}C NMR (DMSO-d_6, 700 MHz) of 7-glycosyl oxazolomycin D

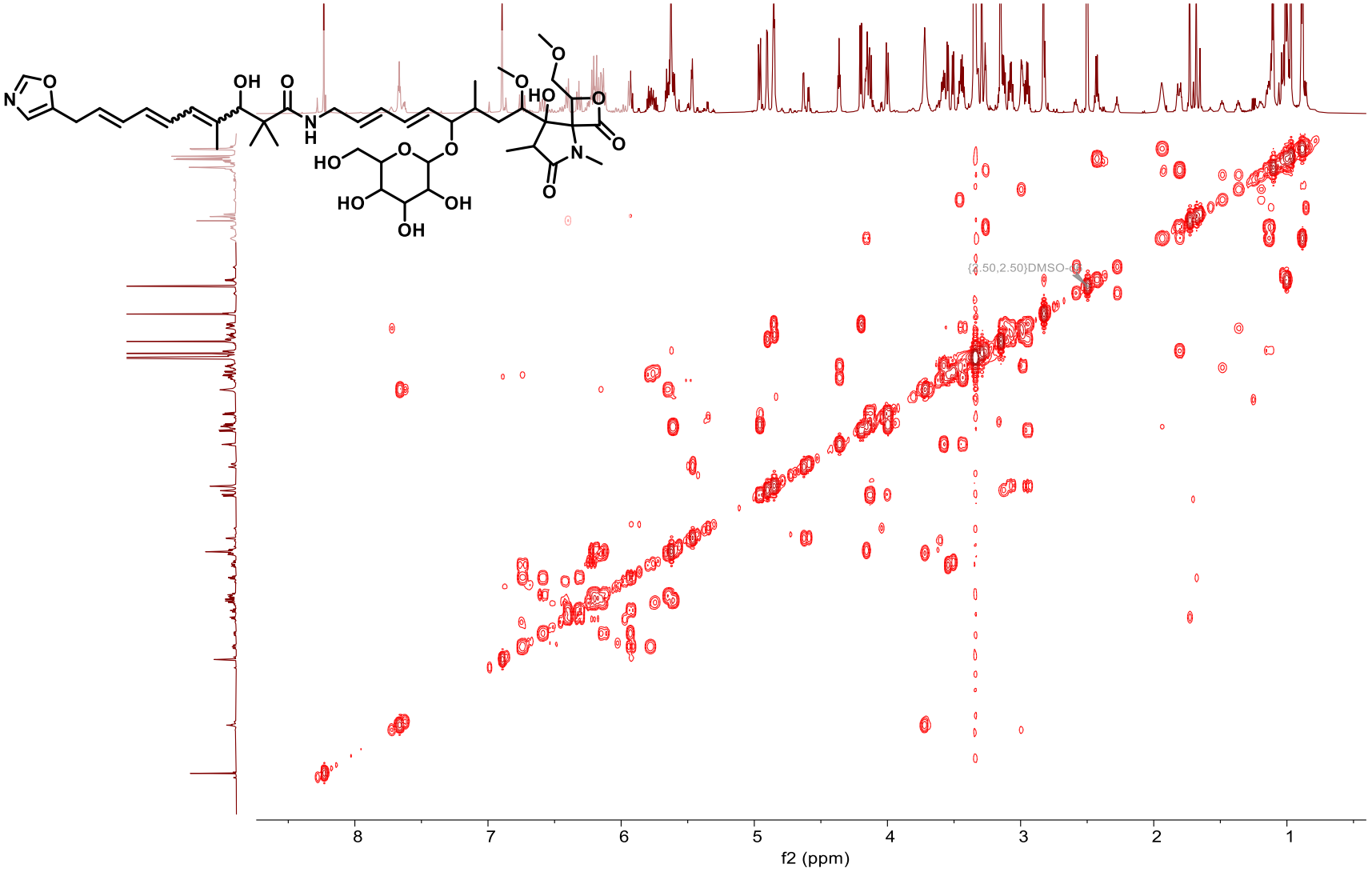


Figure A 57. COSY NMR (DMSO-d_6, 700 MHz) of 7-glycosyl oxazolomycin D

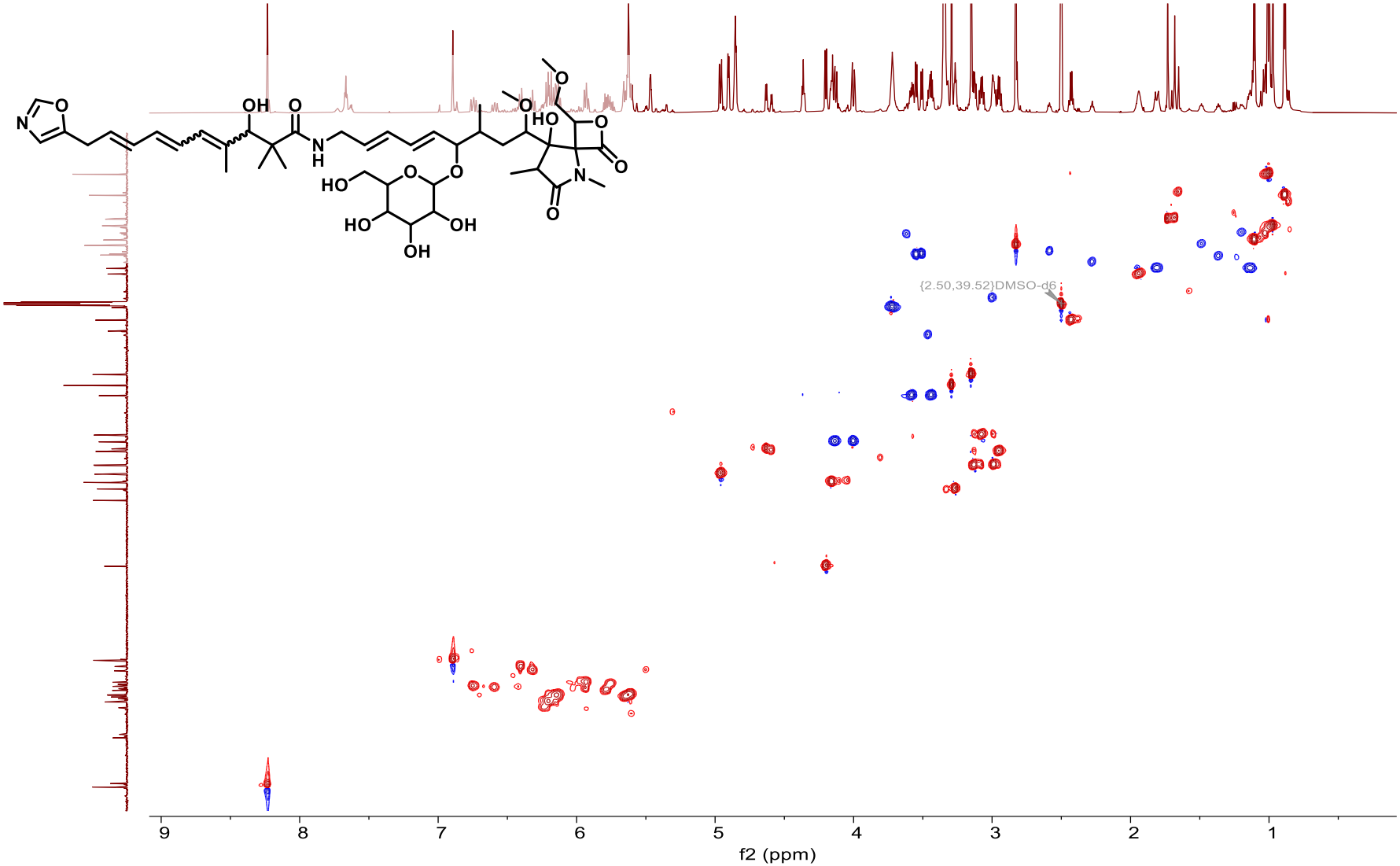


Figure A 58. HSQC NMR (DMSO-d_6, 700 MHz) of 7-glycosyl oxazolomycin D

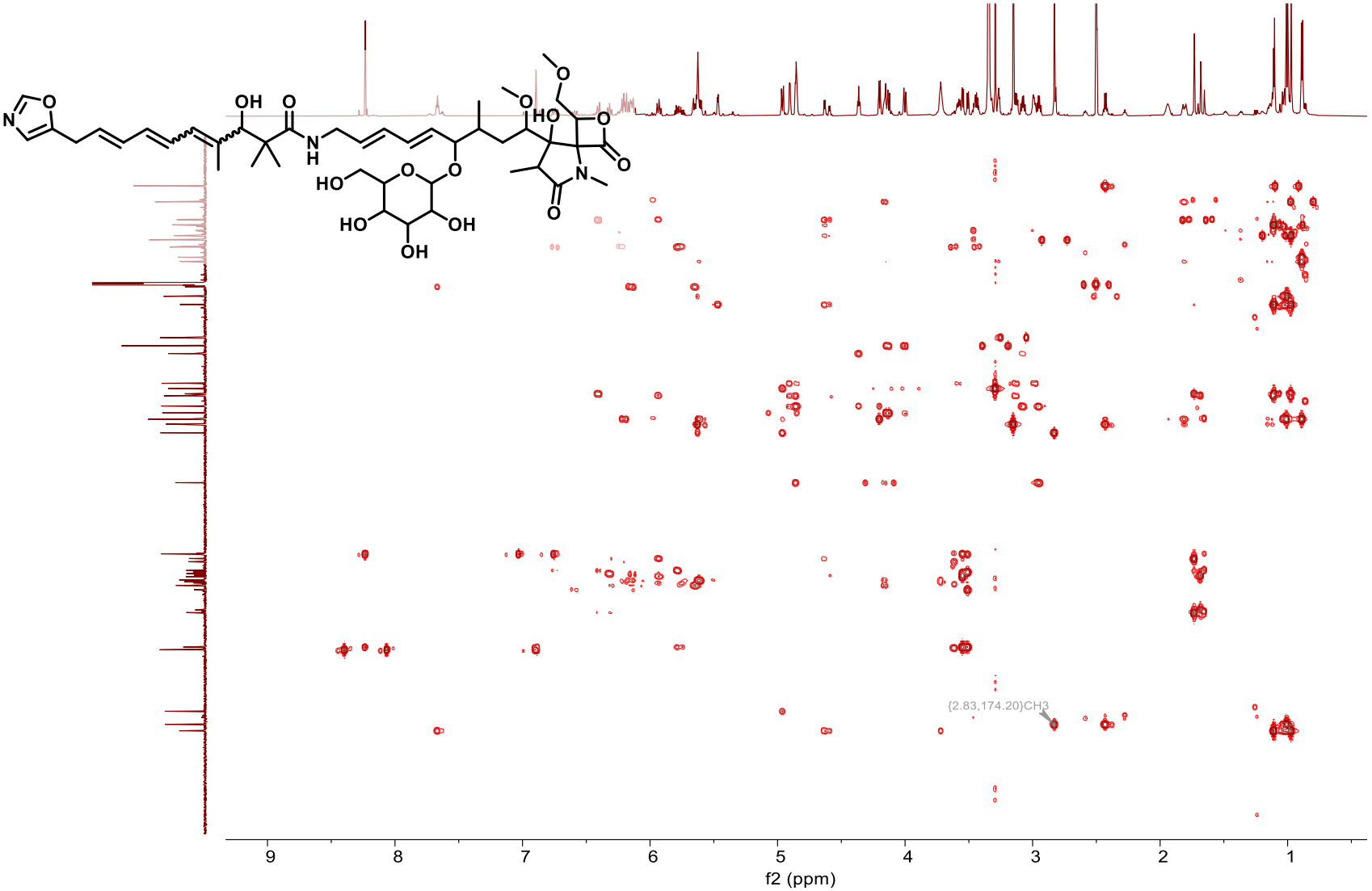


Figure A 59. HMBC NMR (DMSO-d_6, 700 MHz) of 7-glycosyl oxazolomycin D

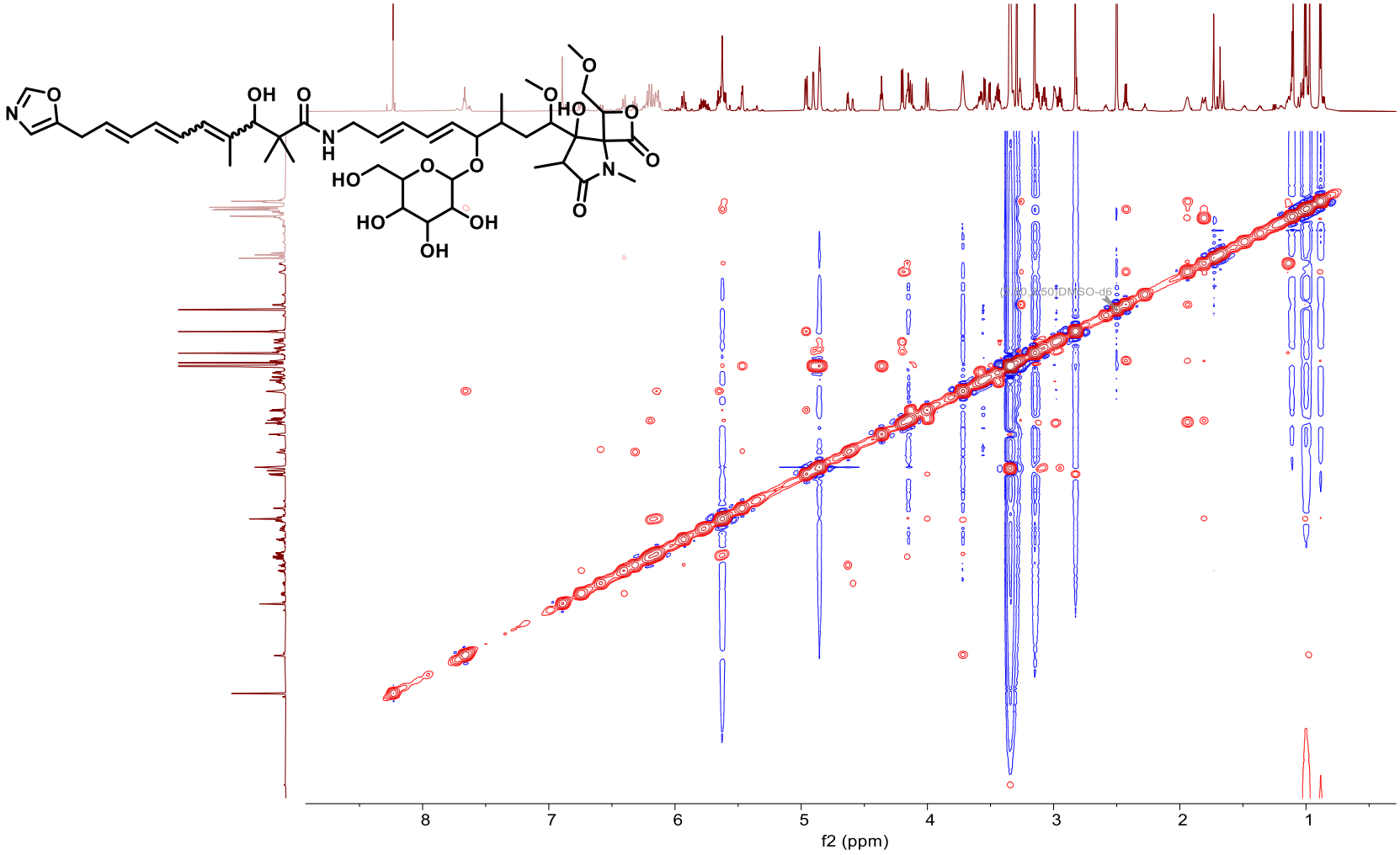


Figure A 60. NOESY NMR (DMSO-d_6, 700 MHz) of 7-glycosyl oxazolomycin D.

A.4 Supplementary information to Chapter 4

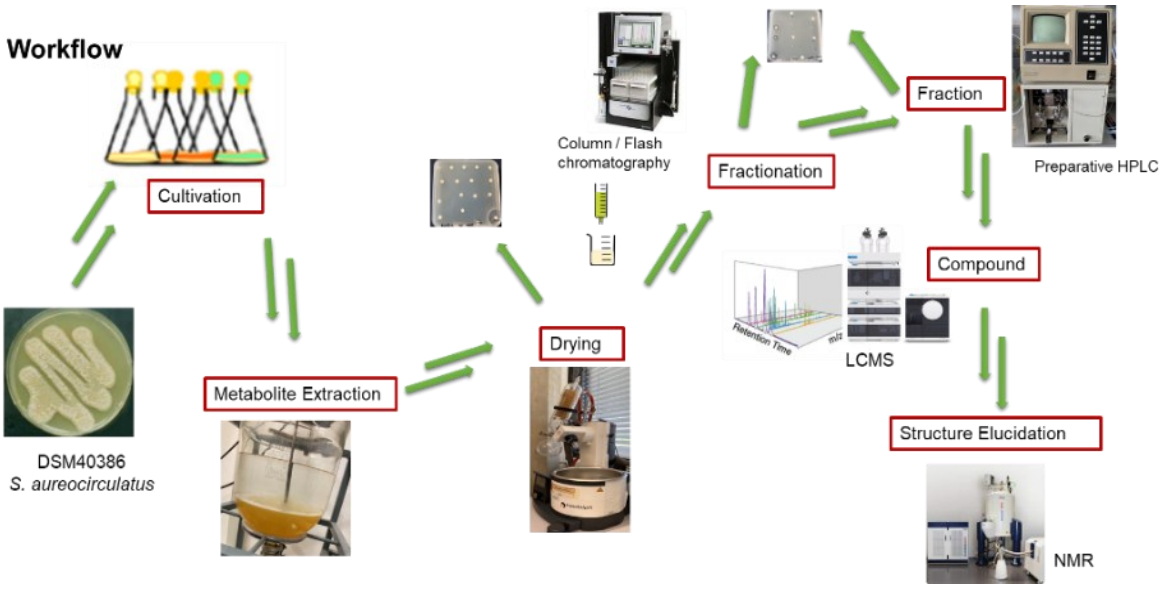


Figure A 61. Workflow for isolation of compounds PAC1-4.

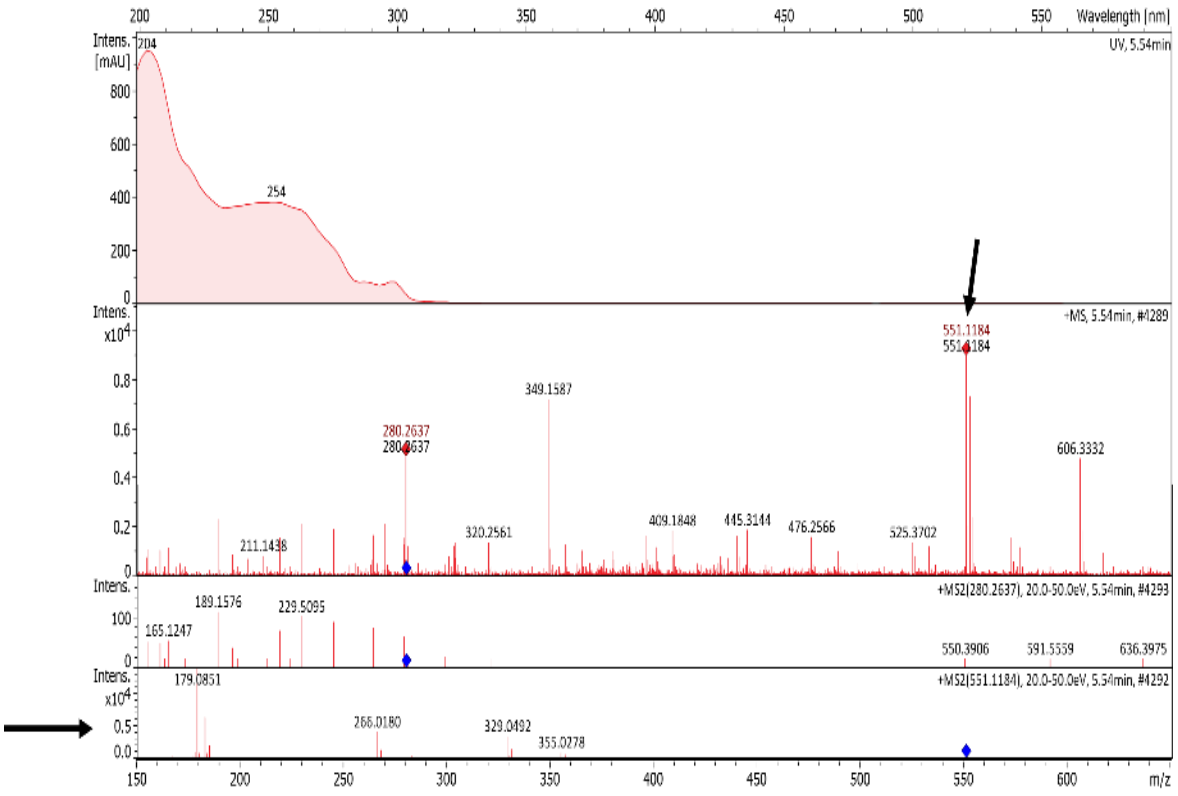


Figure A 62. Chromatogram of brominated signal, including UV, MS and MS/MS.

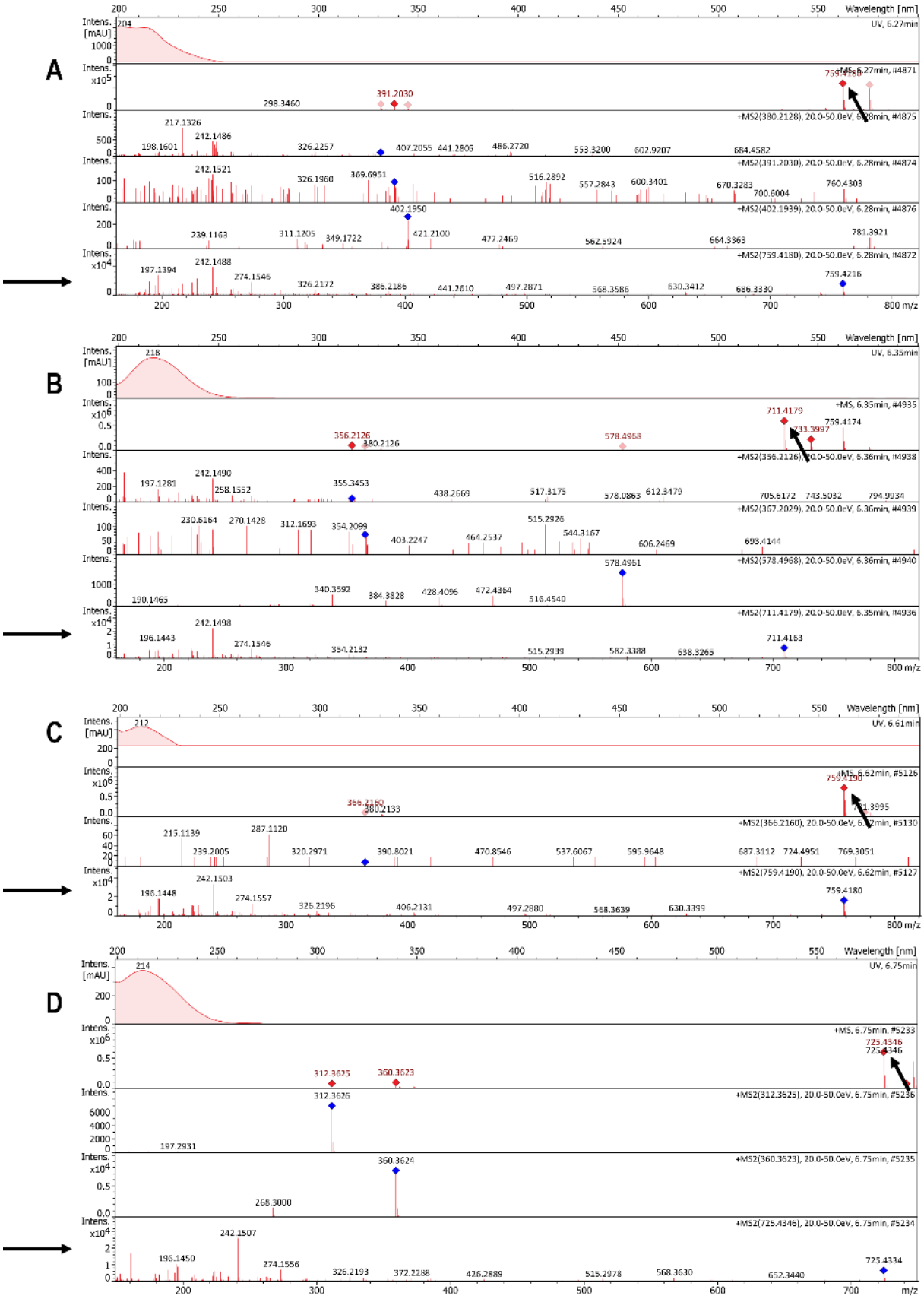


Figure A 63. Chromatogram of the target compounds.

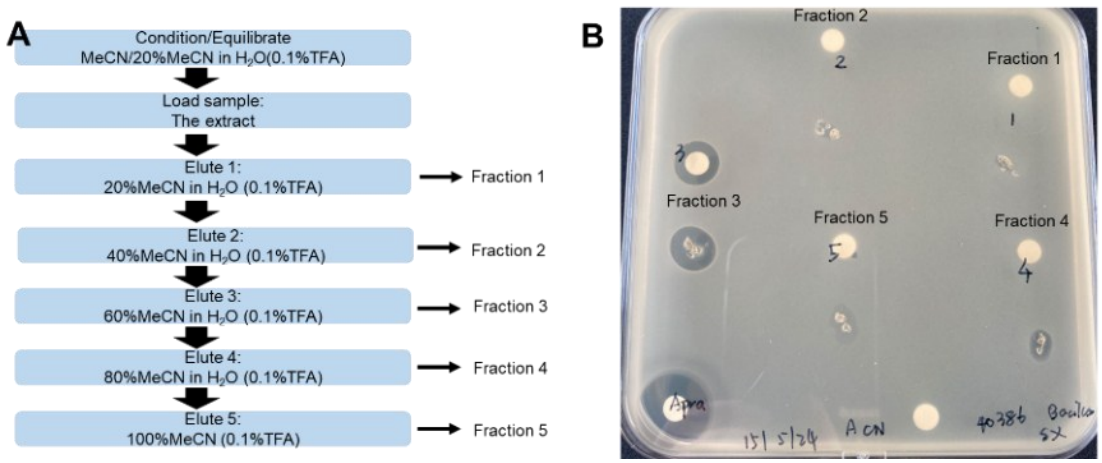


Figure A 64. C18 fractionation workflow to generate five fractions (A). Bioassay results against *Bacilus* (B).

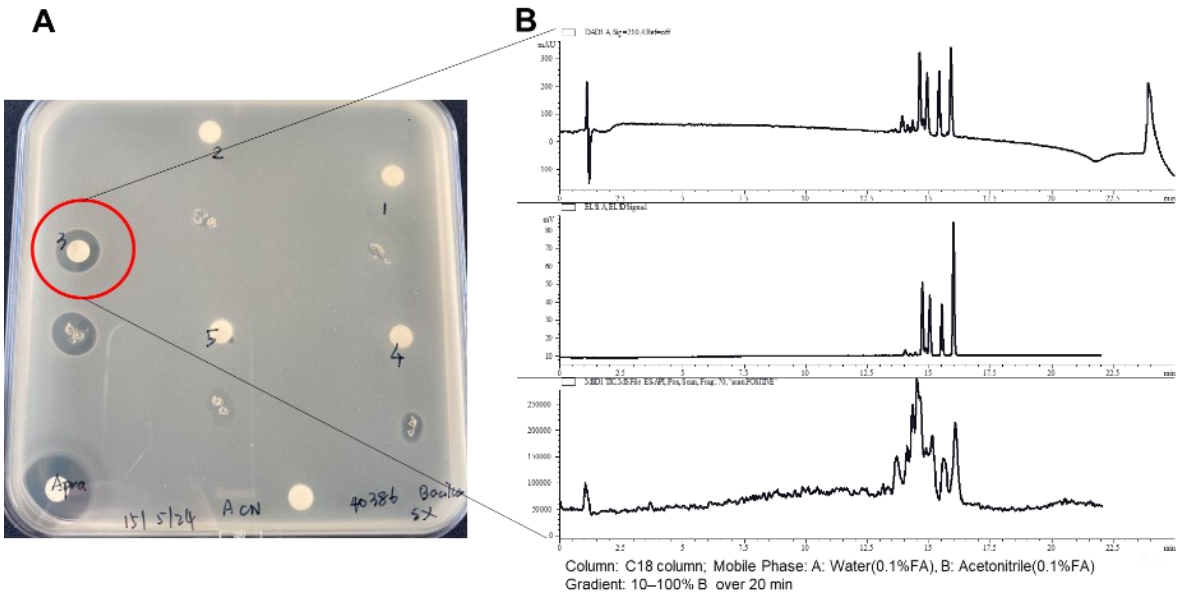


Figure A 65. Chromatogram of the fraction 3.

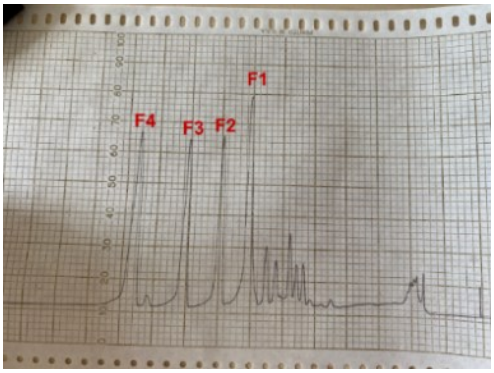


Figure A 66. Chromatogram of the concentrated fraction 3 at 215nm. The target peak is F1, F2, F3 and F4.

Table 4. NMR spectral data for compound PAC1 in DMSO-d6 at 700 MHz

carbon #	δ_C	δ_H, mult (J in Hz)	COSY	HMBC
1	47.0	2.99 (m)	2	3
		2.59 (m)	2	–
2	19.5	1.73 (m)	1, 3	1, 2
		1.28 (m)	1, 3	–
3	23.7	1.60 (m)	3, 4	4
		1.32 (m)	4	1
4	45.8	5.10 (m)	3	2, 5
5	170.4	–	–	–
6	54.9	5.39 (dd, *J* = 6.1, 9.8)	23	5, 7
7	169.9	–	–	–
8	46.2	3.43 (m)	9	–
		3.38 (m)	9	7, 9, 10, 11
9	24.3	1.90 (m)	8, 10	8, 11
		1.77 (m)	8, 10	8, 11
10	27.5	2.03 (m)	9, 11	8, 11, 12
		1.75 (m)	9, 11	8, 11, 12
11	59.7	4.39 (dd, *J* = 2.8, 8.0)	10	8, 9, 10, 12
12	170.3	–	–	–
13	50.2	5.07 (m)	30	12, 14, 30, 31
14	171.3	–	–	–
15	59.8	5.31 (d, *J* = 8.4)	37	16, 37, 38, 40
16	171.9	–	–	–
17	46.5	3.02 (m)	18	19
		2.56 (m)	18	–
18	20.6	1.48 (m, 2H)	17, 19	17, 19, 20
19	24.9	1.89 (m)	18, 20	17, 20
		1.73 (m)	18, 20	17, 20, 21
20	46.6	5.52 (dd, *J* = 1.1, 6.6)	19	18, 19, 21
21	172.6	–	–	–
22	31.2	2.98 (s)	–	5, 6
23	33.7	3.12 (dd, *J* = 6.0, 14.1)	6	6, 7, 24, 25, 29
		2.92 (dd, *J* = 9.4, 14.1)	6	6, 7, 24, 25, 29
24	137.4	–	–	–
25	130.0	7.24 (d, *J* = 7.5)	26	23, 24, 26, 27
26	128.1	7.27 (m)	25, 27	24, 25, 27, 28
27	126.4	7.20 (m)	26, 28	25, 26, 28, 29
28	128.1	7.27 (m)	27, 29	24, 26, 27 ,29
29	129.0	7.24 (d, *J* = 7.5)	28	23, 24, 27, 28
30	37.1	2.98 (m)	13	13, 14, 31, 32
		2.77 (dd, *J* = 5.8, 12.7)	13	13, 14, 31, 32
31	137.5	–	–	–
32	129.3	7.14 (d, *J* = 7.6)	33	30, 31, 33, 34, 36
33	128.0	7.23 (m)	32, 34	31, 32, 34, 35
34	126.2	7.17 (m)	33, 35	32, 33, 35, 36
35	128.0	7.23 (m)	34, 36	31, 33, 34, 36
36	129.3	7.14 (d, *J* = 7.6)	35	30, 31, 32, 34, 35
37	32.9	2.02 (m)	15, 38, 40	15, 38, 39, 40
38	25.5	1.28 (m)	37, 39	15, 37, 39, 40
		1.03 (m)	37, 39	15, 37, 39, 40
39	11.5	0.76 (t, *J* = 7.3)	38	37, 38
40	15.5	0.83 (d, *J* = 6.9)	37	15, 37, 38

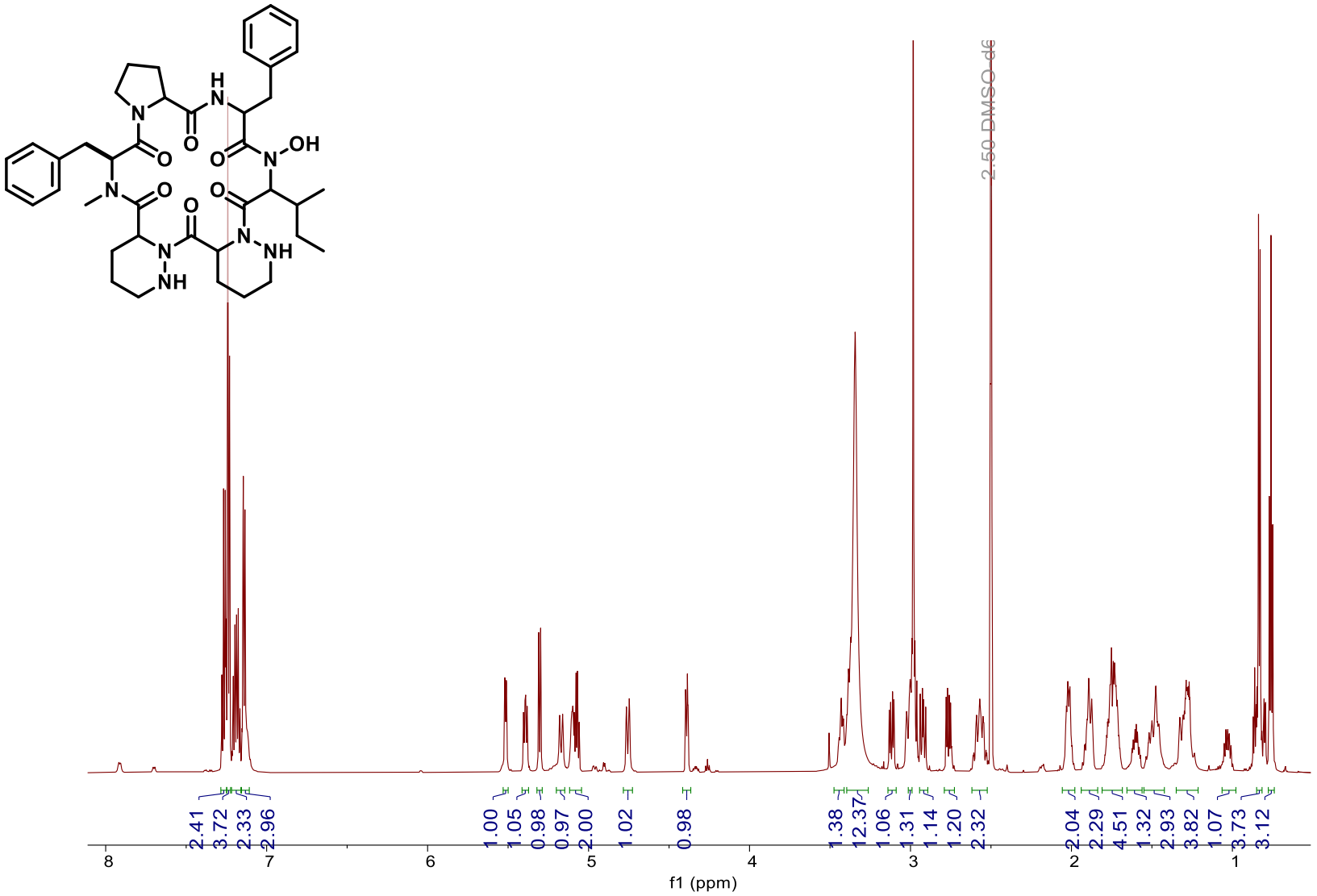


Figure A 67. ^{1}H NMR (DMSO-d_6, 700 MHz) of compound PAC1.

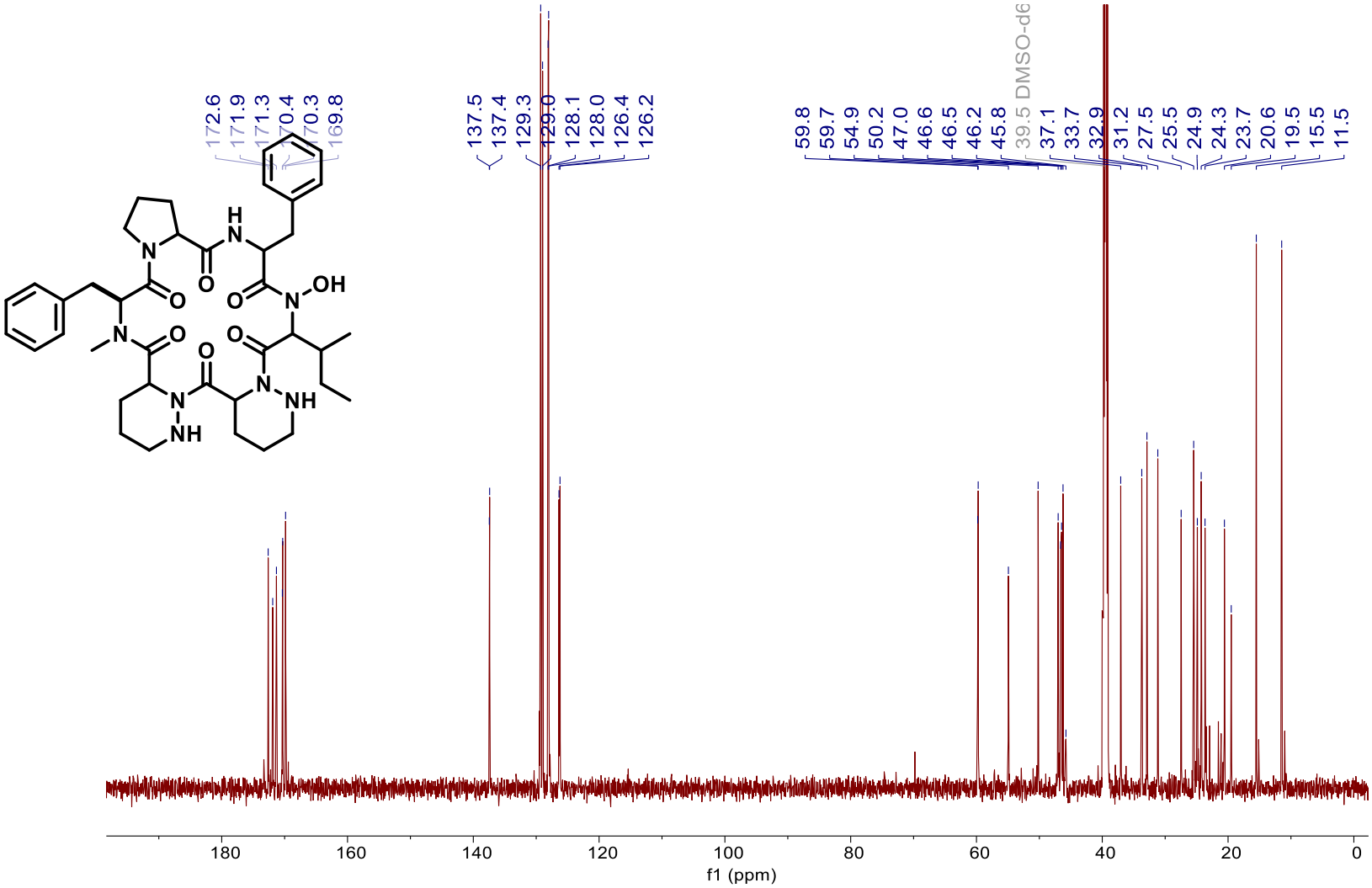


Figure A 68. ^{13}C NMR (DMSO-d_6, 700 MHz) of PAC1

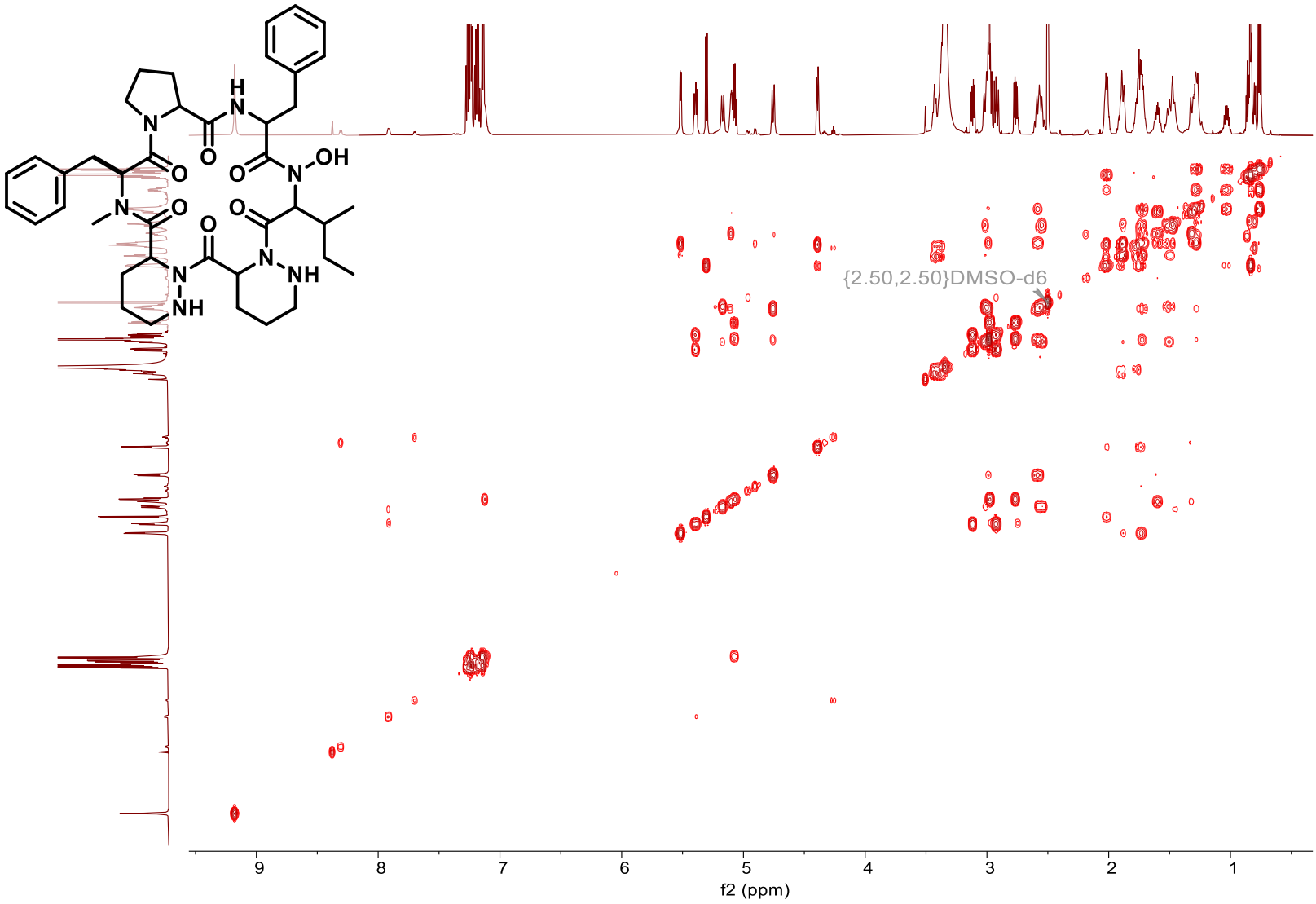


Figure A 69. COSY NMR (DMSO-d_6, 700 MHz) of PAC1

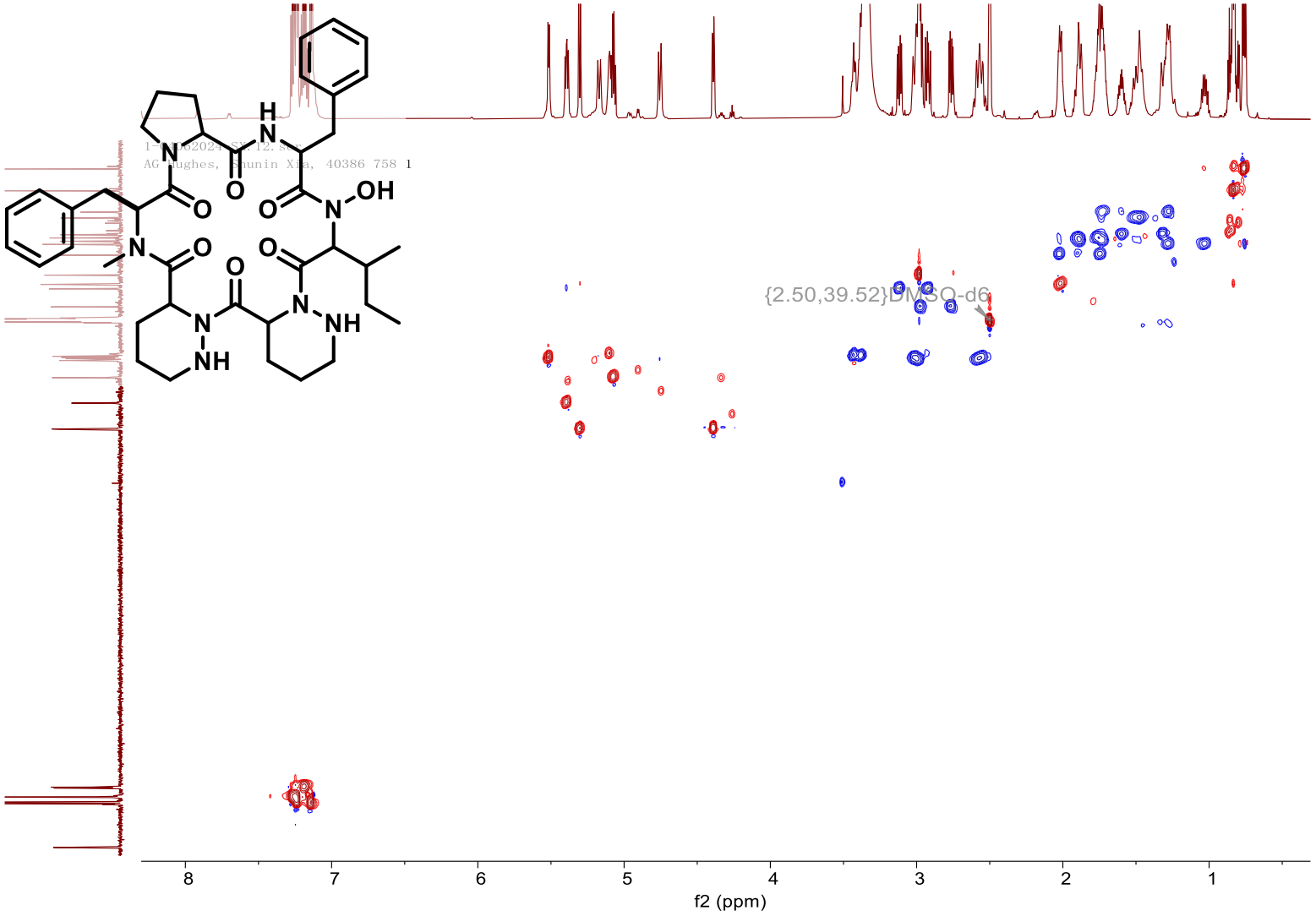


Figure A 70. HSQC NMR (DMSO-d_6, 700 MHz) of PAC1

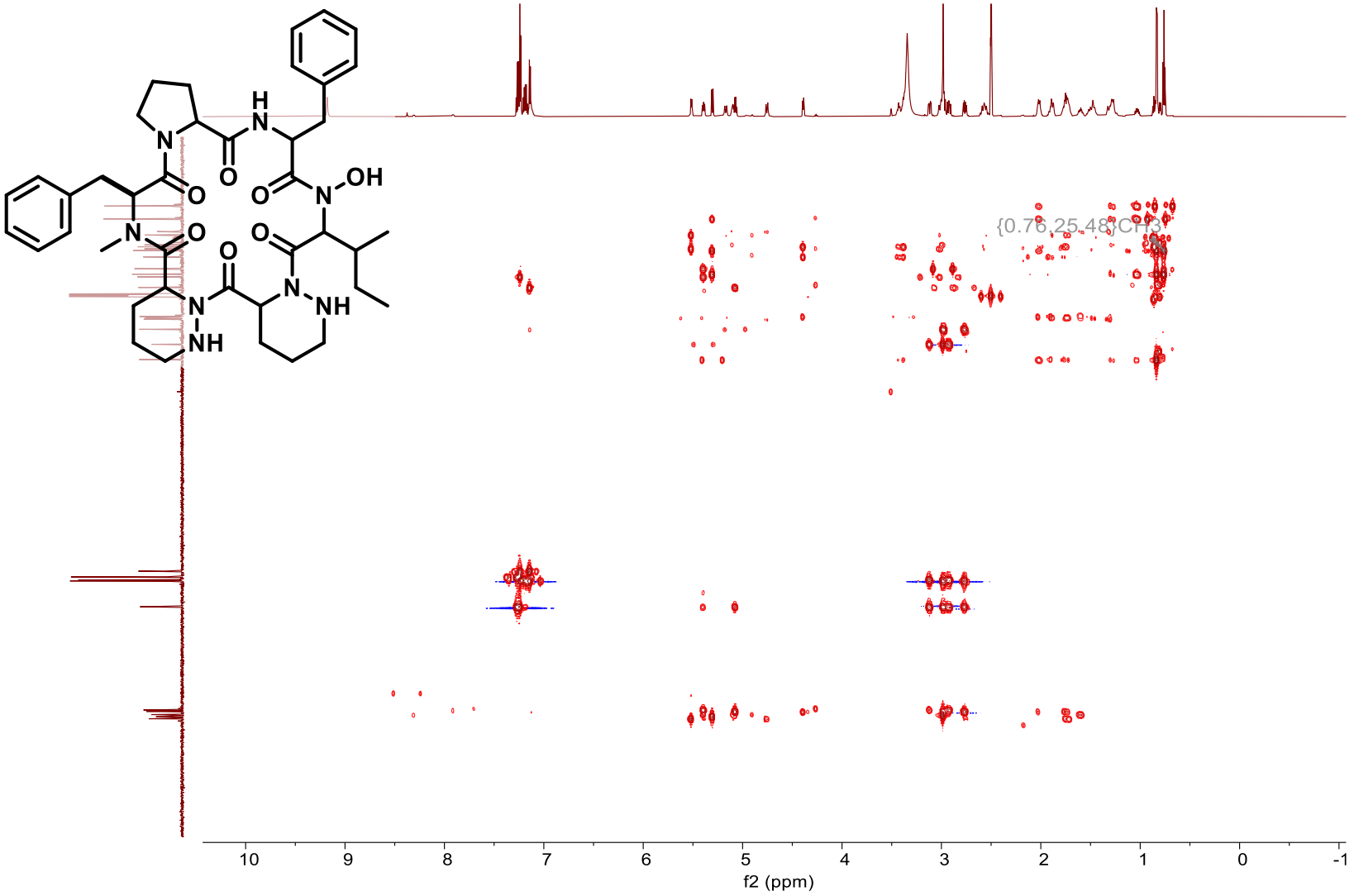


Figure A 71. HMBC NMR (DMSO-d_6, 700 MHz) of PAC1

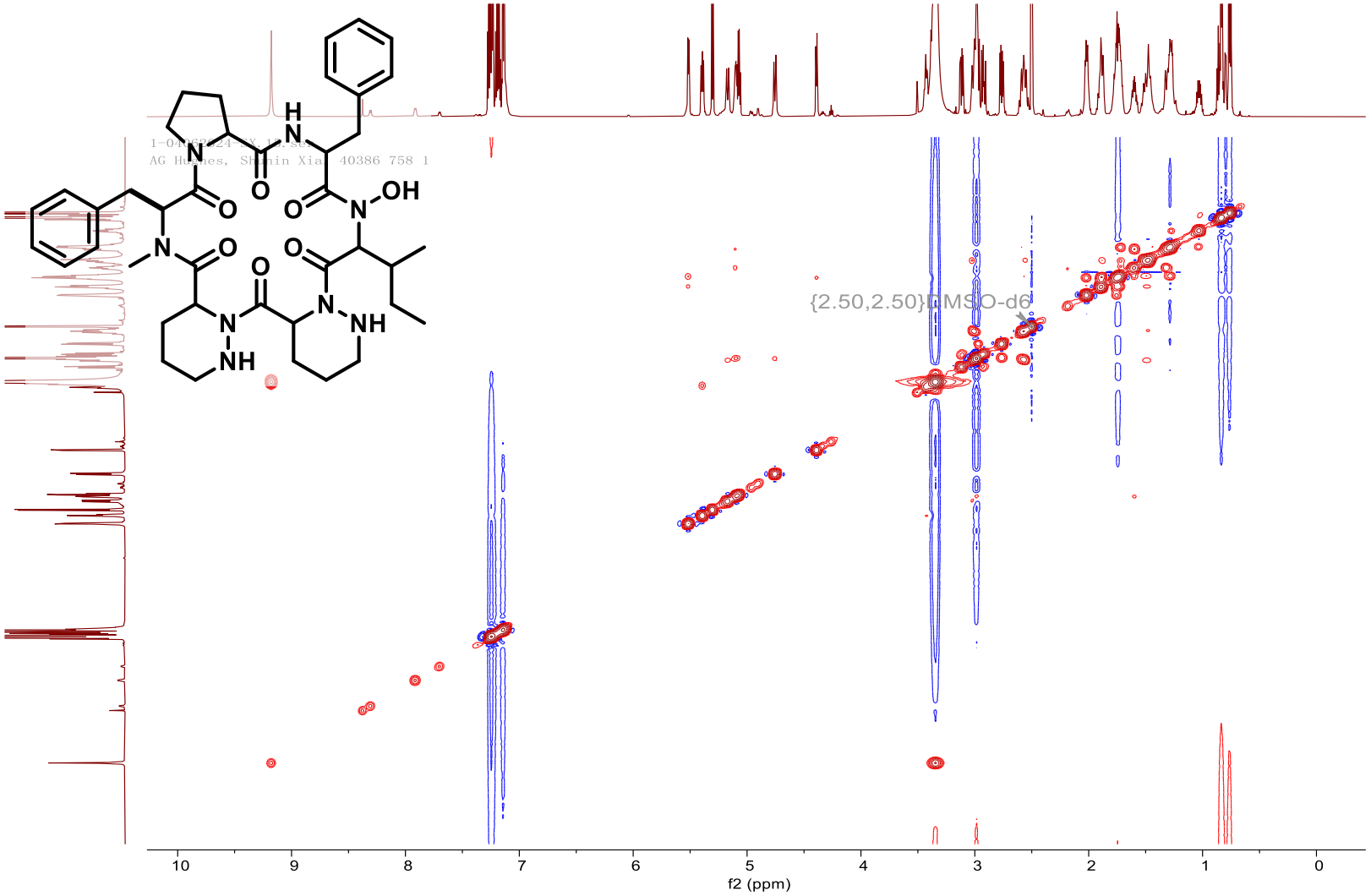


Figure A 72. NOESY NMR (DMSO-d_6, 700 MHz) of PAC1.

Table 5. NMR spectral data for compound PAC2 in DMSO-d6 at 700 MHz

carbon #	δ_C	δ_H, mult (J in Hz)	COSY	HMBC
1	47.7	3.01 (br d, *J* = 12.7)	2	3
		2.56 (dd, *J* = 2.3, 12.8)	2	–
2	19.6	1.94 (m)	1, 3	–
		1.29 (m)	1, 3	–
3	23.9	1.55 (m)	2, 4	2, 4, 5
		1.49 (m)	2, 4	–
4	46.0	5.19 (d, *J* = 6.2)	3	2, 3, 5
5	170.0	–	–	–
6	56.3	5.32 (dd, *J* = 7.6, 8.8)	23	7, 23, 24
7	169.8	–	–	–
8	46.1	3.35 (m)	9	9, 11
		3.28 (m)	9	–
9	24.3	1.91 (m)	8, 10	8
		1.78 (m)	8, 10	–
10	27.0	2.10 (m)	9, 11	8
		1.75 (m)	9, 11	9, 12
11	60.0	4.47 (dd, *J* = 3.0, 7.8)	10	8, 9, 10, 12
12	170.5	–	–	–
13	53.7	4.71 (t, *J* = 8.5)	30	14, 30, 31, 32
14	172.3	–	–	–
15	56.1	5.34 (dd, *J* = 3.5, 10.9)	33	16, 33, 34
16	173.2	–	–	–
17	46.3	3.06 (br d, *J* = 12.5)	18	–
		2.63 (dd, *J* = 2.5, 12.2)	18	–
18	19.6	1.94 (m)	17, 19	–
		1.29 (m)	17, 19	–
19	23.9	1.55 (m)	18, 20	17, 18, 21
		1.49 (m)	18, 20	–
20	46.6	5.40 (dd, *J* = 1.9, 6.7)	19	16, 18, 21
21	172.7	–	–	–
22	31.1	2.78 (s)	–	5, 6
23	33.9	3.13 (dd, *J* = 6.7, 13.6)	6	6, 7, 24, 25, 29
		2.81 (dd, *J* = 8.1, 13.5)	6	6, 7, 24, 25, 29
24	137.8	–	–	–
25	129.1	7.24 (m)	26	23, 24, 26, 27, 29
26	128.0	7.25 (m)	25, 27	24, 25, 27, 28
27	126.3	7.18 (m)	26, 28	25, 29
28	128.0	7.25 (m)	27, 29	24, 25, 26, 27
29	129.1	7.24 (m)	28	23, 24, 25, 27, 28
30	29.9	1.91 (m)	13, 31, 32	13, 14, 31, 32
31	18.4	0.89 (d, *J* = 6.5)	30	13, 30, 32
32	19.1	0.82 (d, *J* = 6.7)	30	13, 30, 31
33	35.5	1.82 (m)	15, 34	15, 34, 35, 36
		1.44 (t, *J* = 3.6)	15, 34	15, 34, 35, 36
34	25.1	1.65 (m)	33, 35, 36	15, 33
35	23.2	0.90 (d, *J* = 6.5)	34	33, 34, 36
36	20.8	0.85 (d, *J* = 6.5)	34	33, 34, 35

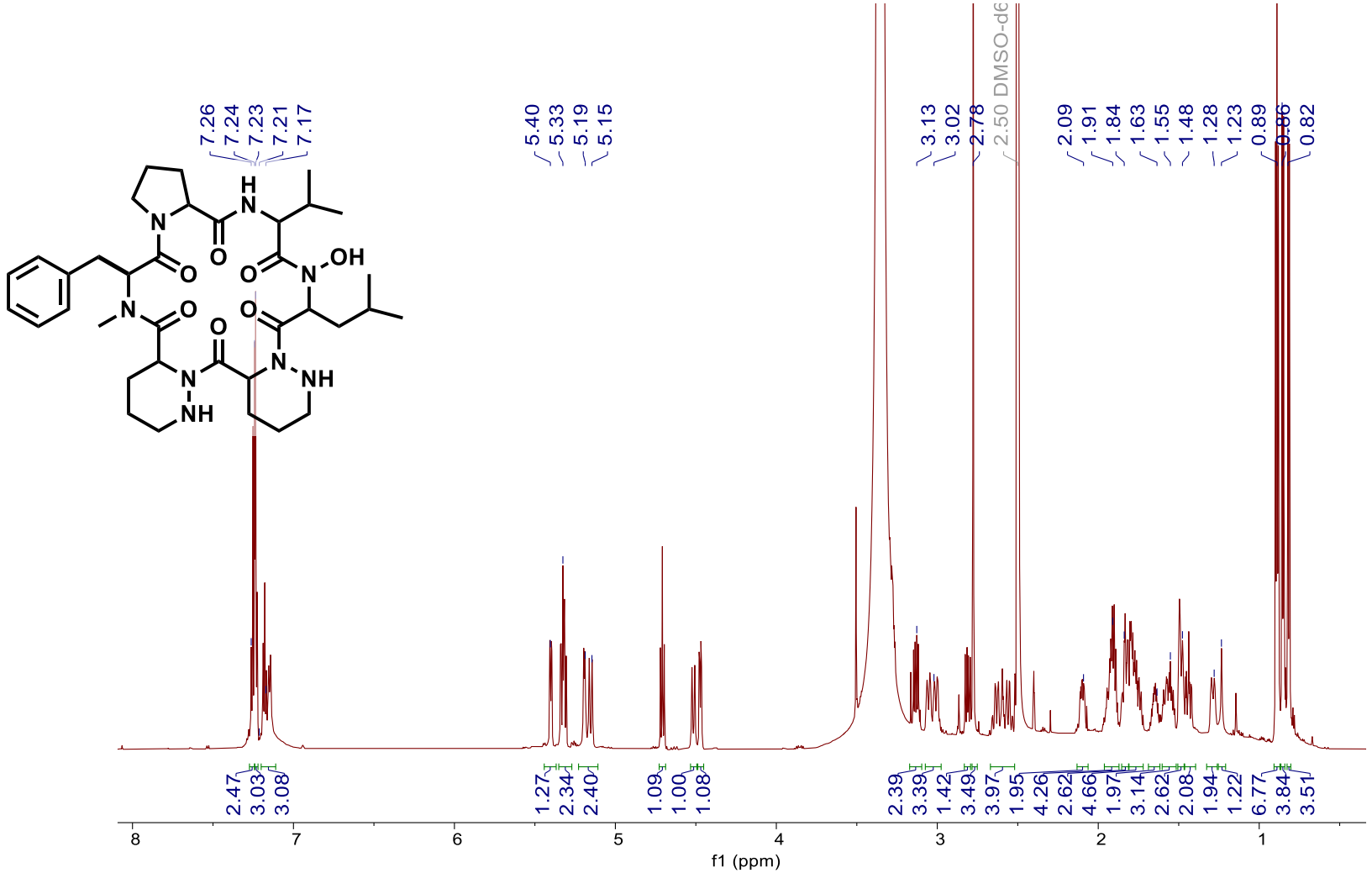


Figure A 73. ^{1}H NMR (DMSO-d_6, 700 MHz) of ccompound PAC2.

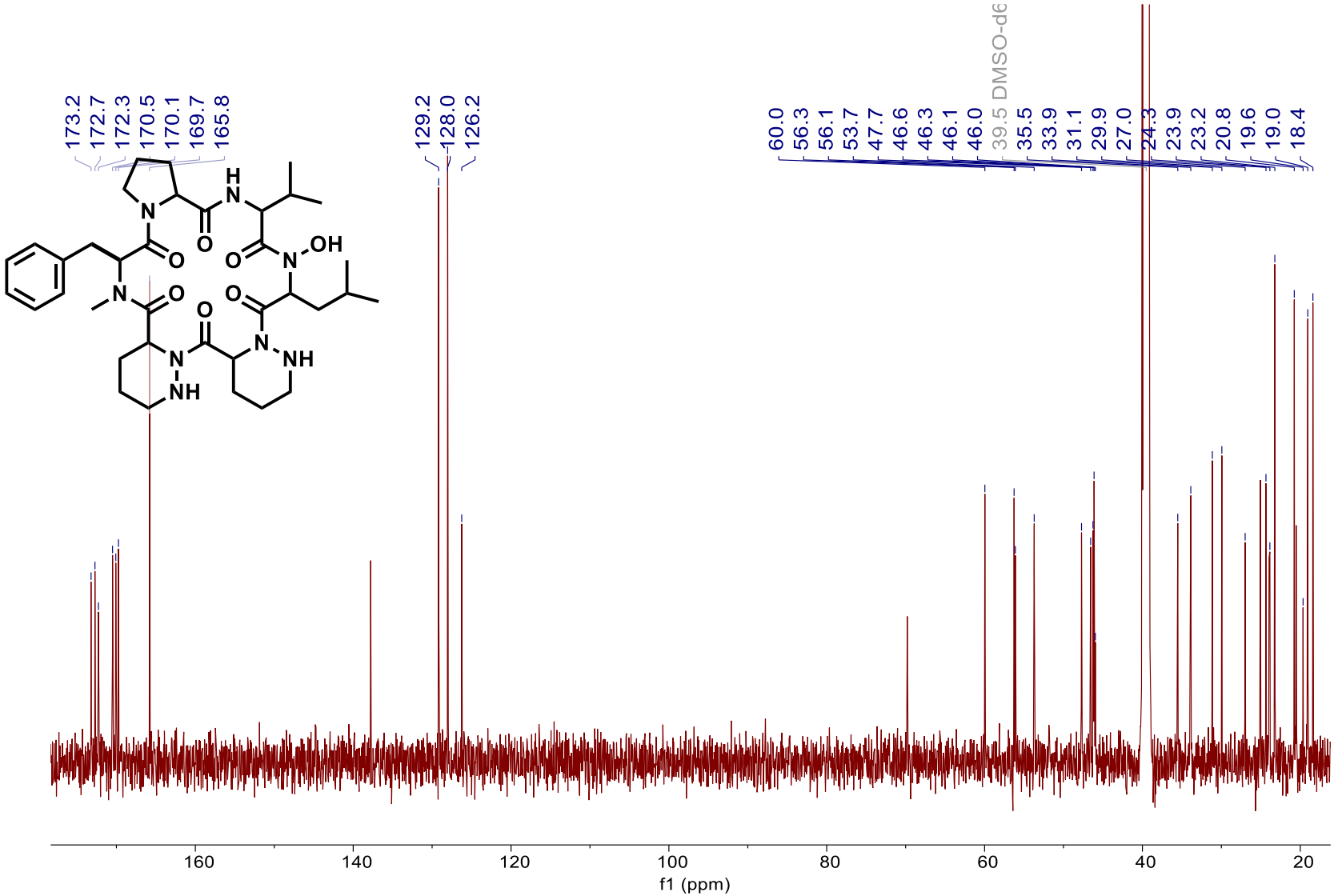


Figure A 74. ^{13}C NMR (DMSO-d_6, 700 MHz) of ccompound PAC2.

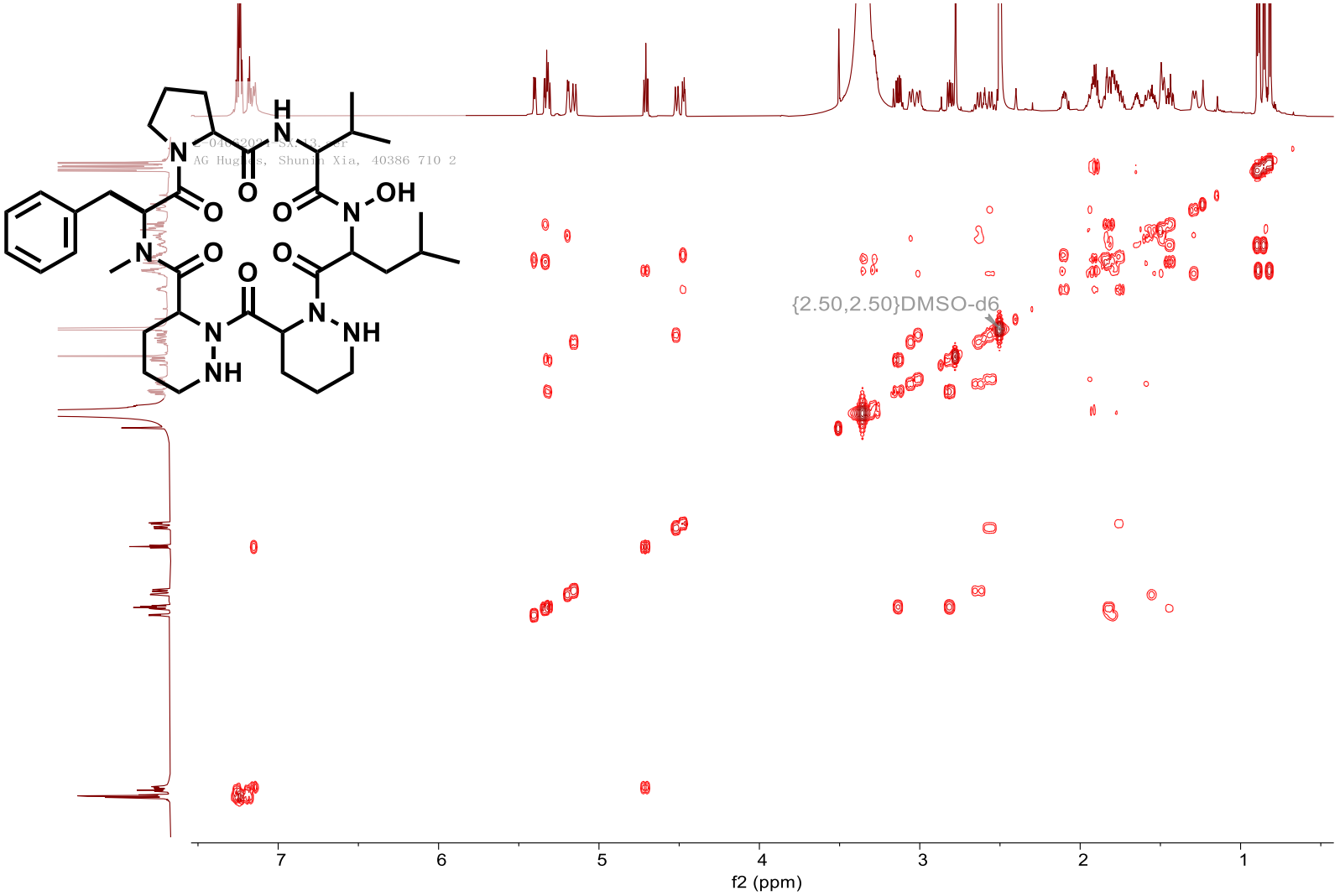


Figure A 75. COSY (DMSO-d_6, 700 MHz) of ccompound PAC2.

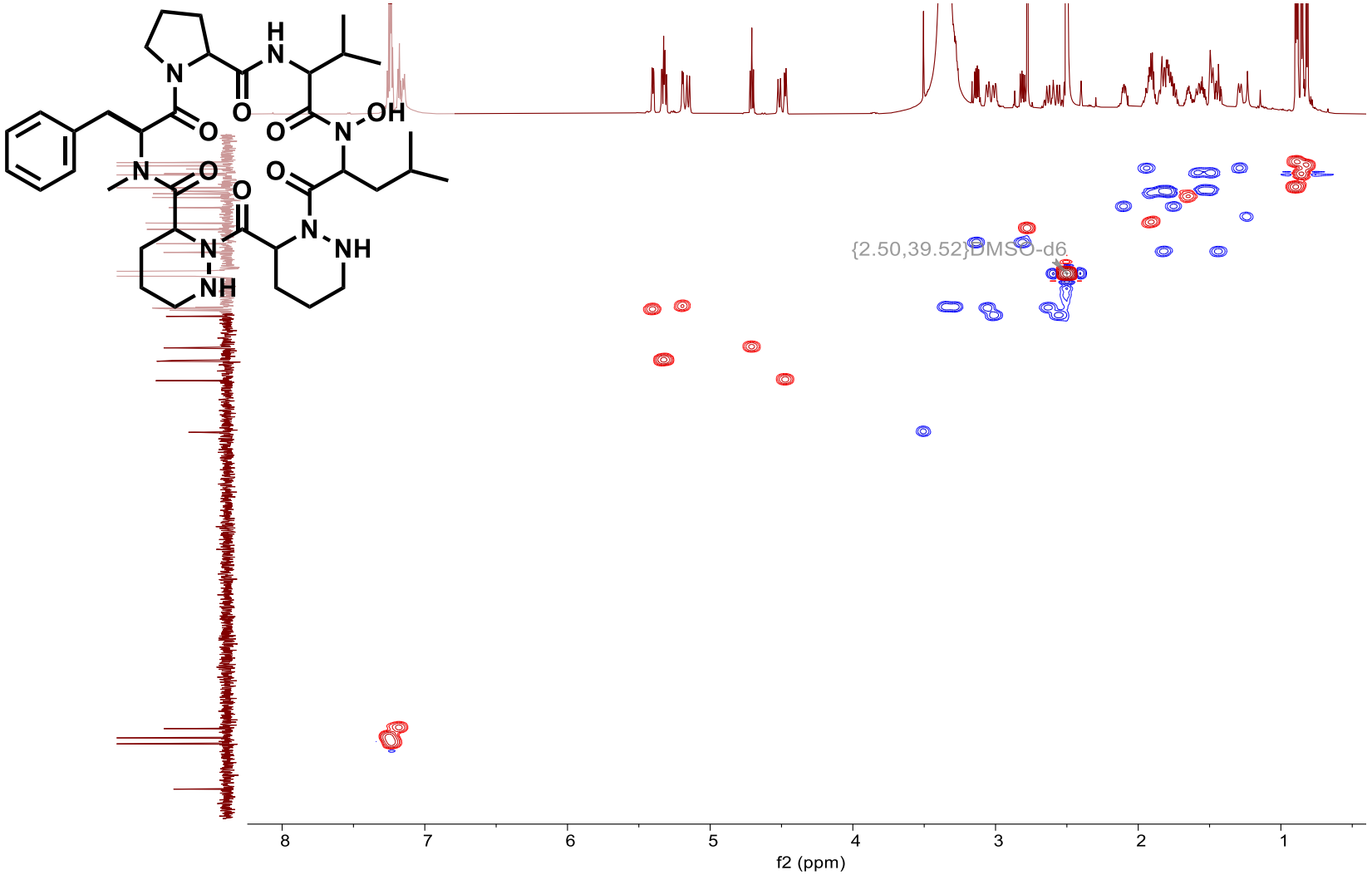


Figure A 76. HSQC NMR (DMSO-d_6, 700 MHz) of PAC2

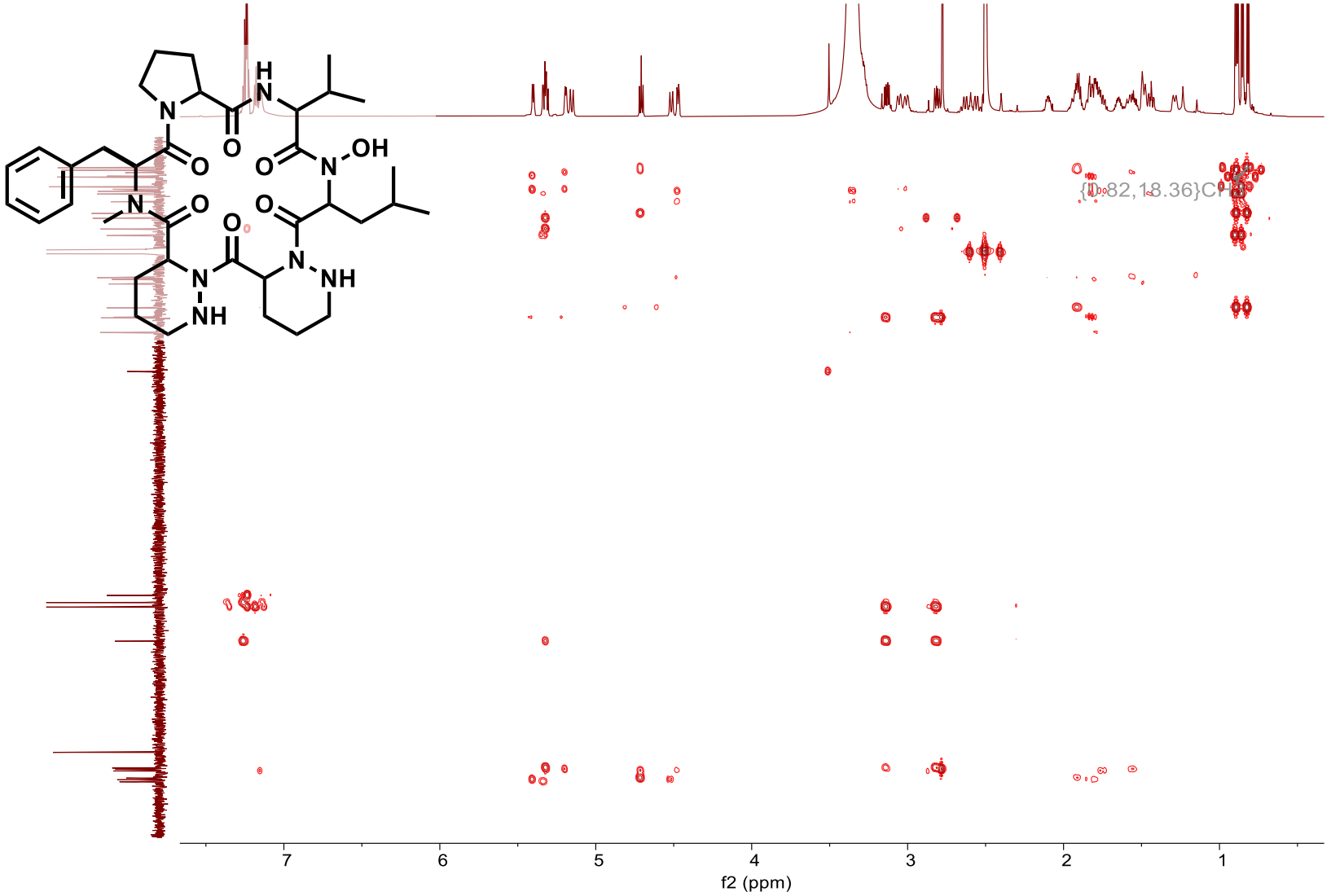


Figure A 77. HMBC NMR (DMSO-d_6, 700 MHz) of PAC2

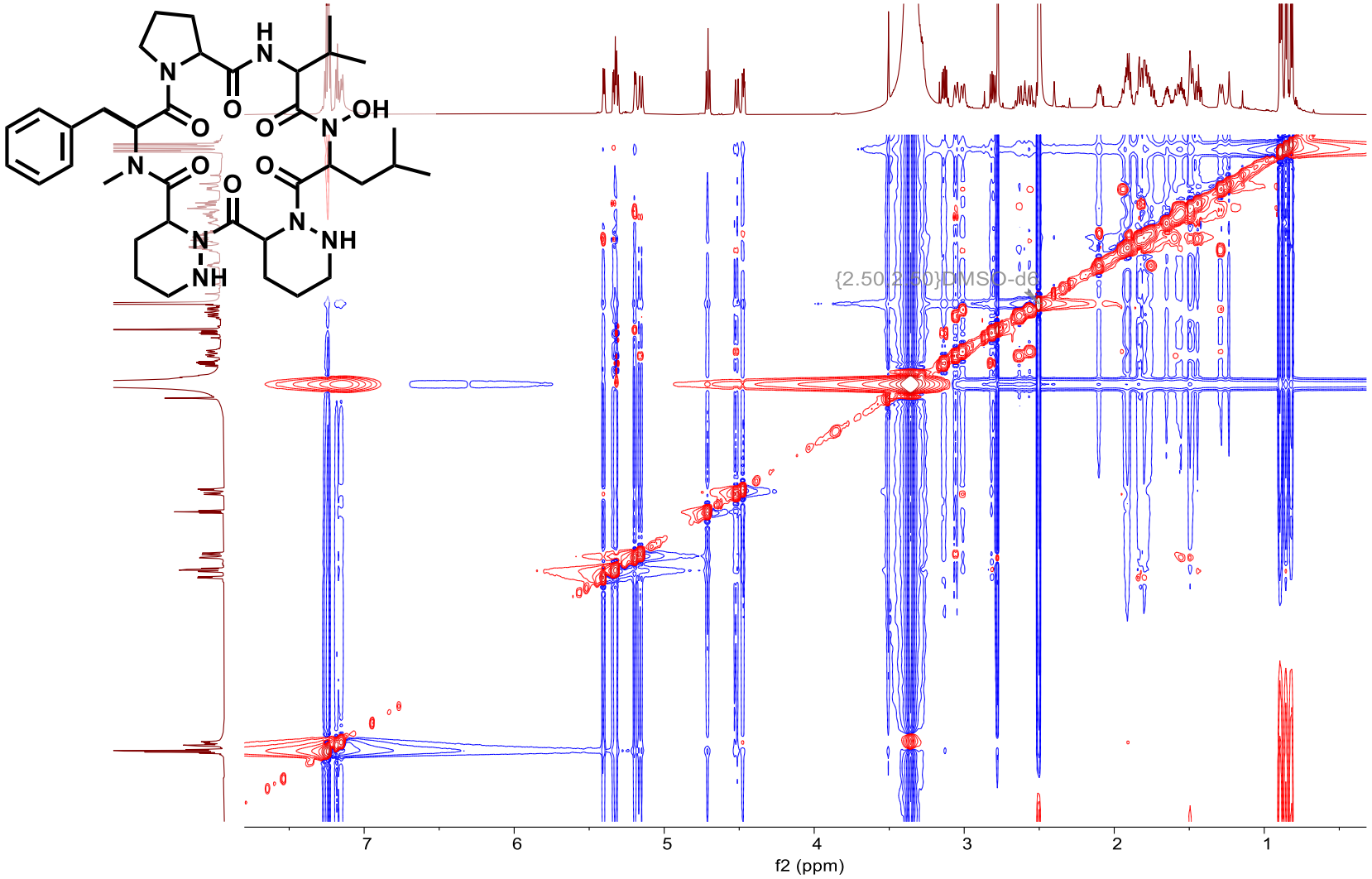


Figure A 78. NOESY NMR (DMSO-d_6, 700 MHz) of PAC2

Table 6. NMR spectral data for compound PAC3 in DMSO-d6 at 700 MHz

carbon #	δ_C	δ_H, mult (J in Hz)	COSY	HMBC
1	47.6	3.00 (m)	2	3
		2.56 (dd, *J* = 2.4, 12.6)	2	–
2	19.6	1.88 (m)	1	–
		1.28 (m)	1	4
3	23.5	1.56 (m)	2, 4	2, 4
		1.43 (m)	2, 4	–
4	45.8	5.19 (d, *J* = 13.1)	3	2, 3, 5
5	170.5	–	–	–
6	56.0	5.33 (m)	–	7, 22, 23, 24
7	169.6	–	–	–
8	46.2	3.37 (m, 2H)	9	7, 9, 10, 11
9	24.5	1.90 (m)	8	10, 11
		1.79 (m)	8	8, 11
10	27.1	2.04 (m)	9, 11	8, 12
		1.79 (m)	11	8, 11, 12
11	60.2	4.38 (dd, *J* = 4.1, 8.4)	10	8, 9, 10, 12
12	170.4	–	–	–
13	50.5	5.06 (m)	30	14, 30, 31
14	172.1	–	–	–
15	56.8	5.24 (m)	37	16, 37, 38
16	173.3	–	–	–
17	46.2	3.03 (m)	18	19
		2.62 (dd, *J* = 2.3, 11.6)	18	–
18	20.6	1.54 (m)	17, 19	–
		1.48 (m)	19	19, 20
19	23.9	1.83 (m)	18, 20	17, 20, 21
		1.80 (m)	18, 20	17, 20, 21
20	47.2	5.34 (m)	19	18, 19, 21
21	172.7	–	–	–
22	31.2	2.82 (s)	–	5, 6
23	33.8	3.15 (dd, *J* = 6.6, 13.2)	6	6, 7, 24, 25, 29
		2.84 (m)	6	6, 7, 24, 25, 29
24	137.8	–	–	–
25	129.1	7.24 (m)	26, 28	24, 26, 27, 28
26	128.1	7.25 (m)	25, 27, 28	24, 25, 27, 29
27	126.3	7.18 (m)	26, 28	25, 26, 28, 29
28	128.1	7.25 (m)	25, 27, 28	24, 25, 27, 29
29	129.1	7.24 (m)	26, 28	24, 26, 27, 28
30	36.5	2.97 (dd, *J* = 8.1, 13.0)	13	13, 14, 31, 32, 36
		2.84 (m)	13	13, 14, 31, 32, 36
31	137.4	–	–	–
32	129.2	7.15 (d, *J* = 7.6)	33	30, 31, 33, 34
33	128.1	7.25 (m)	32, 34	31, 32, 34, 35
34	126.3	7.19 (m)	33, 36	32, 33, 35, 36
35	128.1	7.25 (m)	34, 36	31, 33, 34, 36
36	129.2	7.15 (d, *J* = 7.6)	35	30, 31, 34, 35
37	32.5	1.75 (m)	15	15, 38, 39
		1.41 (m)	15	38, 39
38	24.8	1.49 (m)	39, 40	37
39	20.8	0.79 (d, *J* = 6.6)	38	37, 38, 40
40	23.2	0.83 (d, *J* = 6.6)	38	37, 38, 39

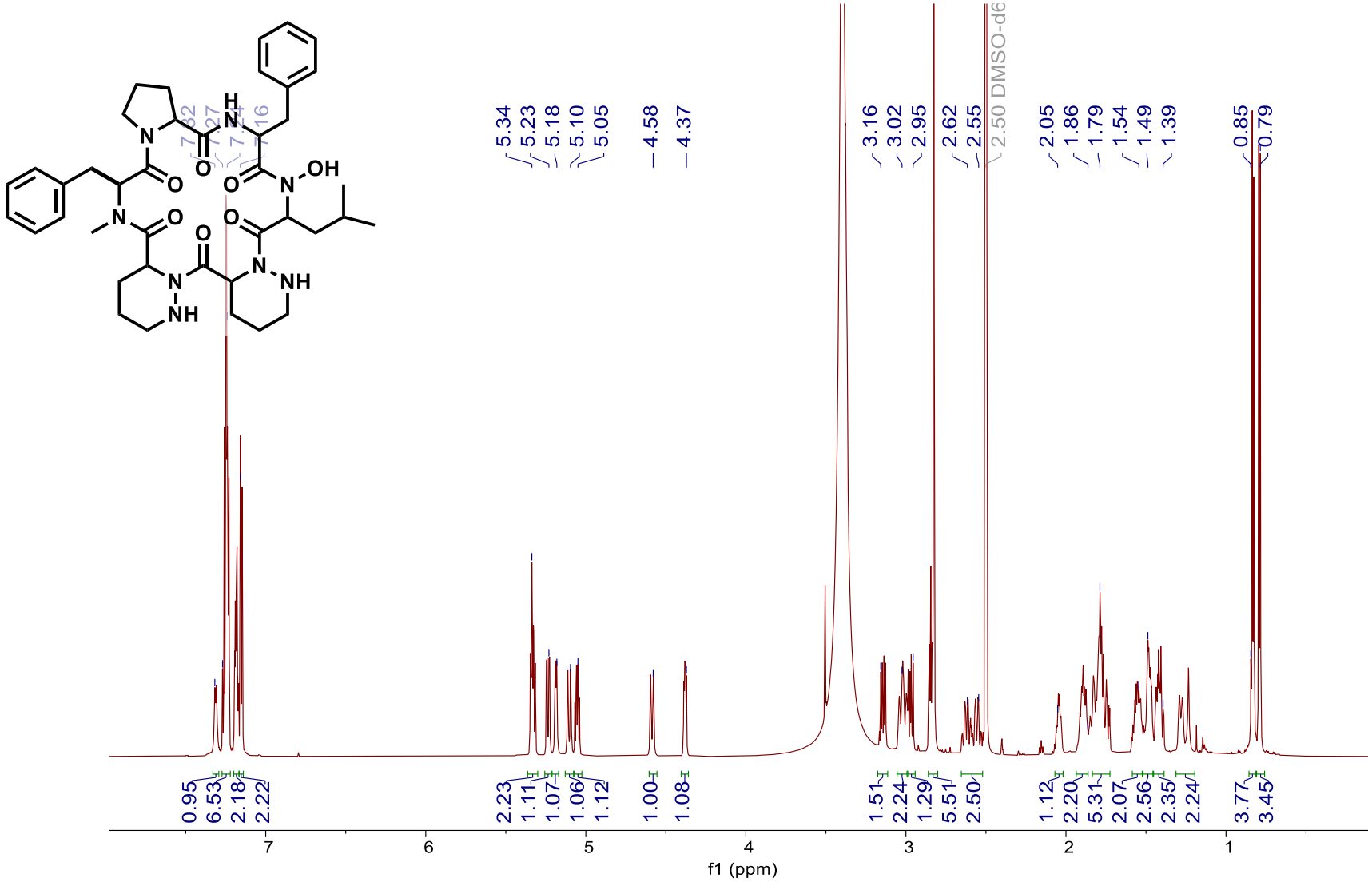


Figure A 79. ^{1}H NMR (DMSO-d_6, 700 MHz) of PAC3

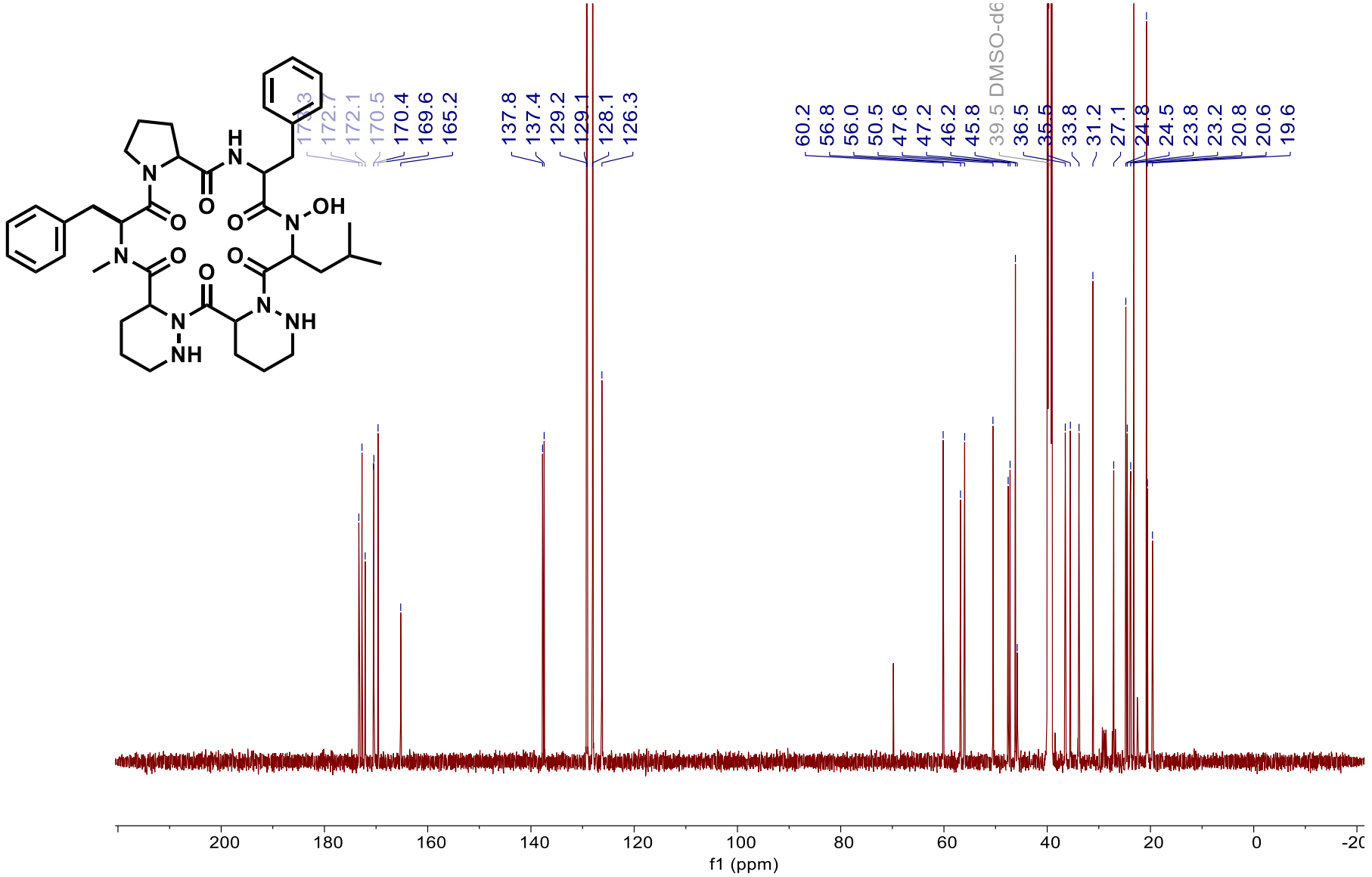


Figure A 80. ^{13}C NMR (DMSO-d_6, 700 MHz) of PAC3

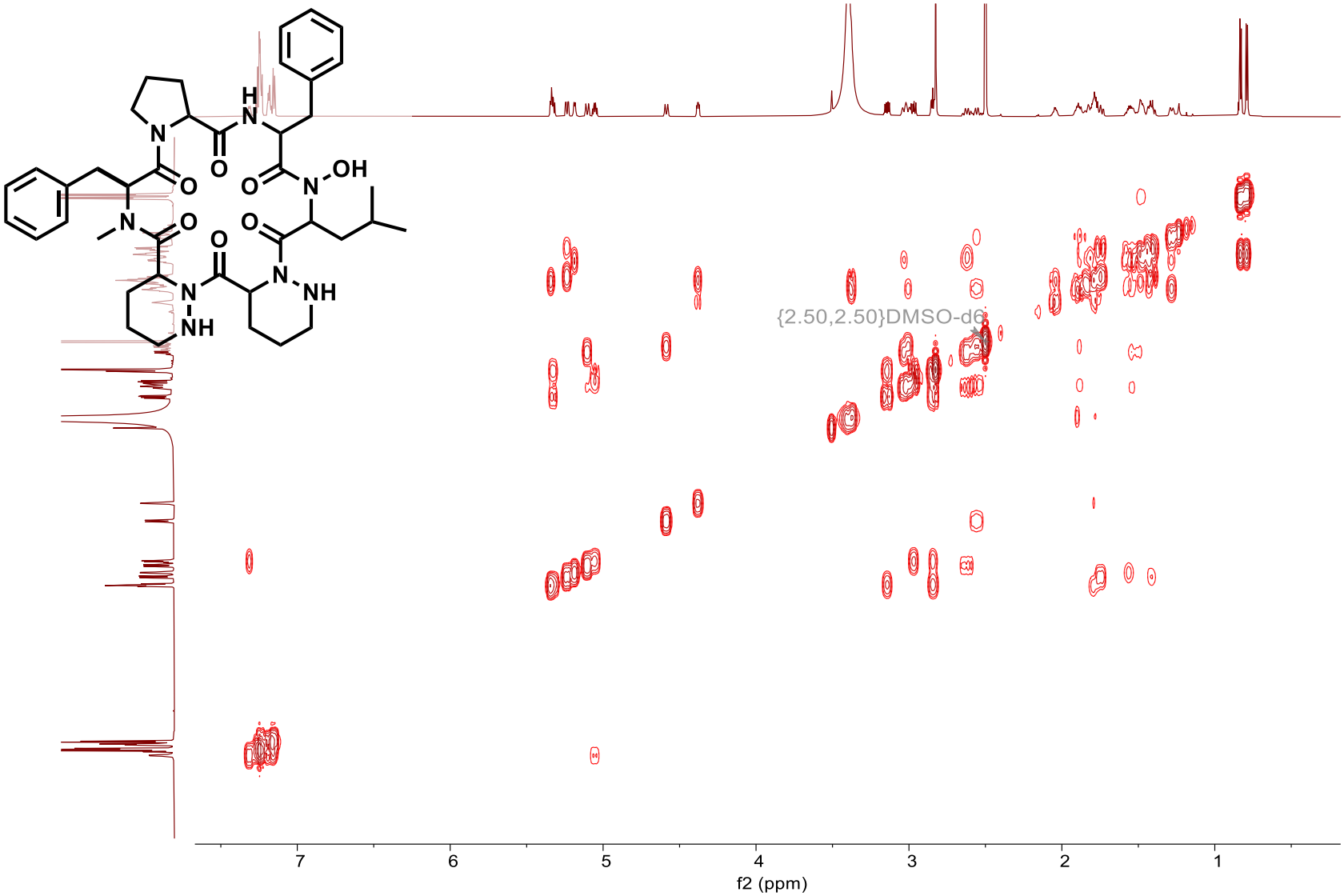


Figure A 81. COSY NMR (DMSO-d_6, 700 MHz) of PAC3

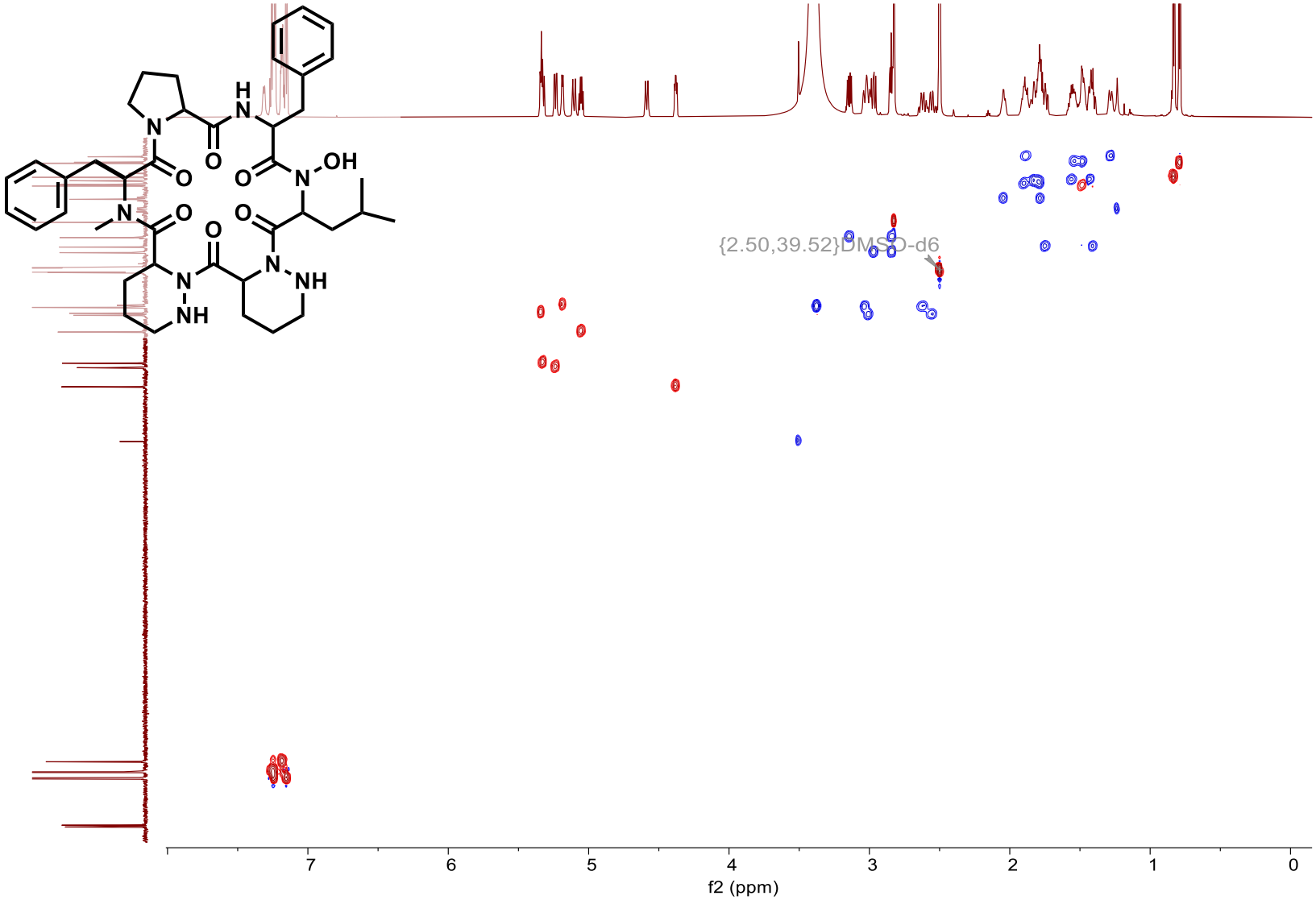


Figure A 82. HSQC NMR (DMSO-d_6, 700 MHz) of PAC3

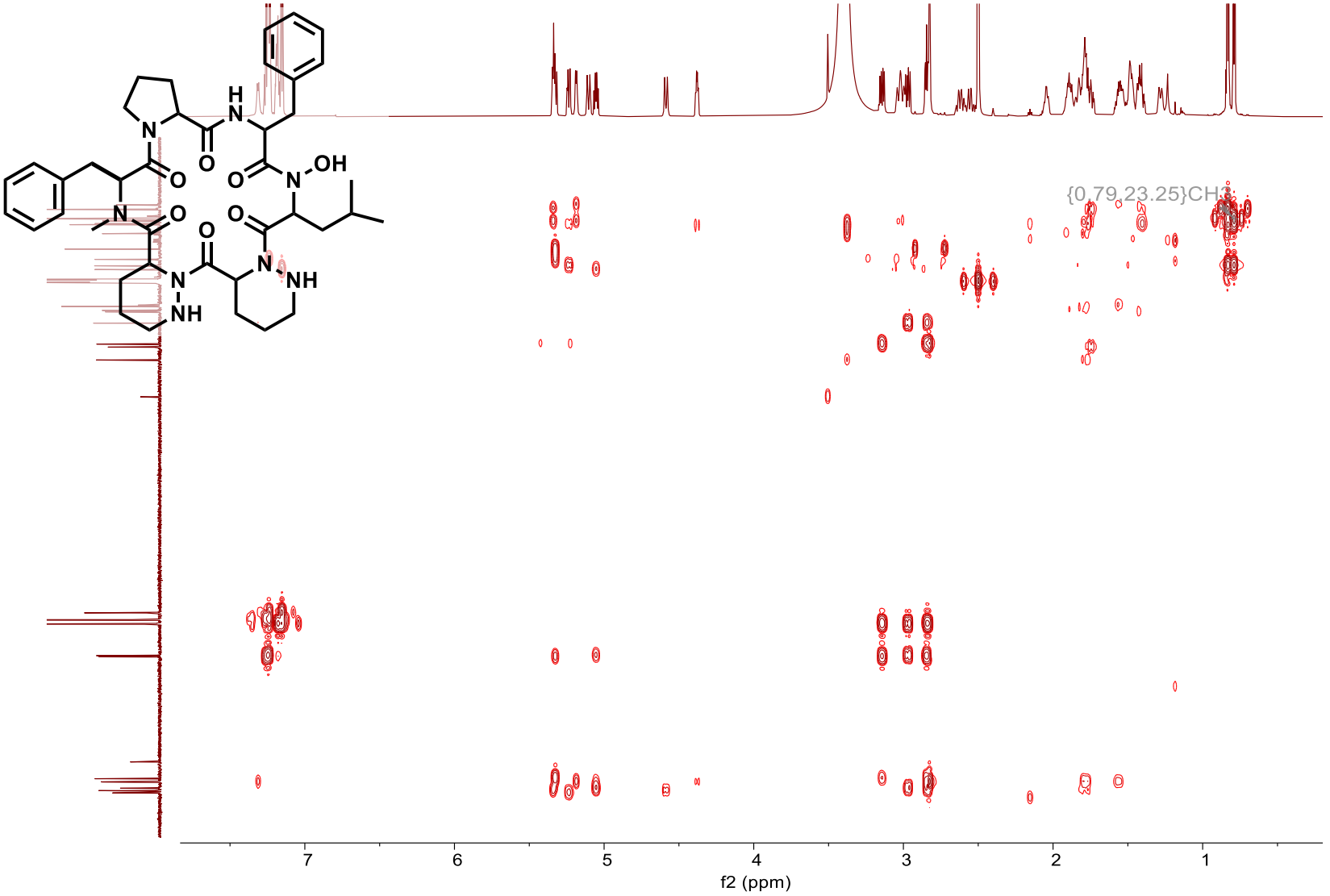


Figure A 83. HMBC NMR (DMSO-d_6, 700 MHz) of PAC3

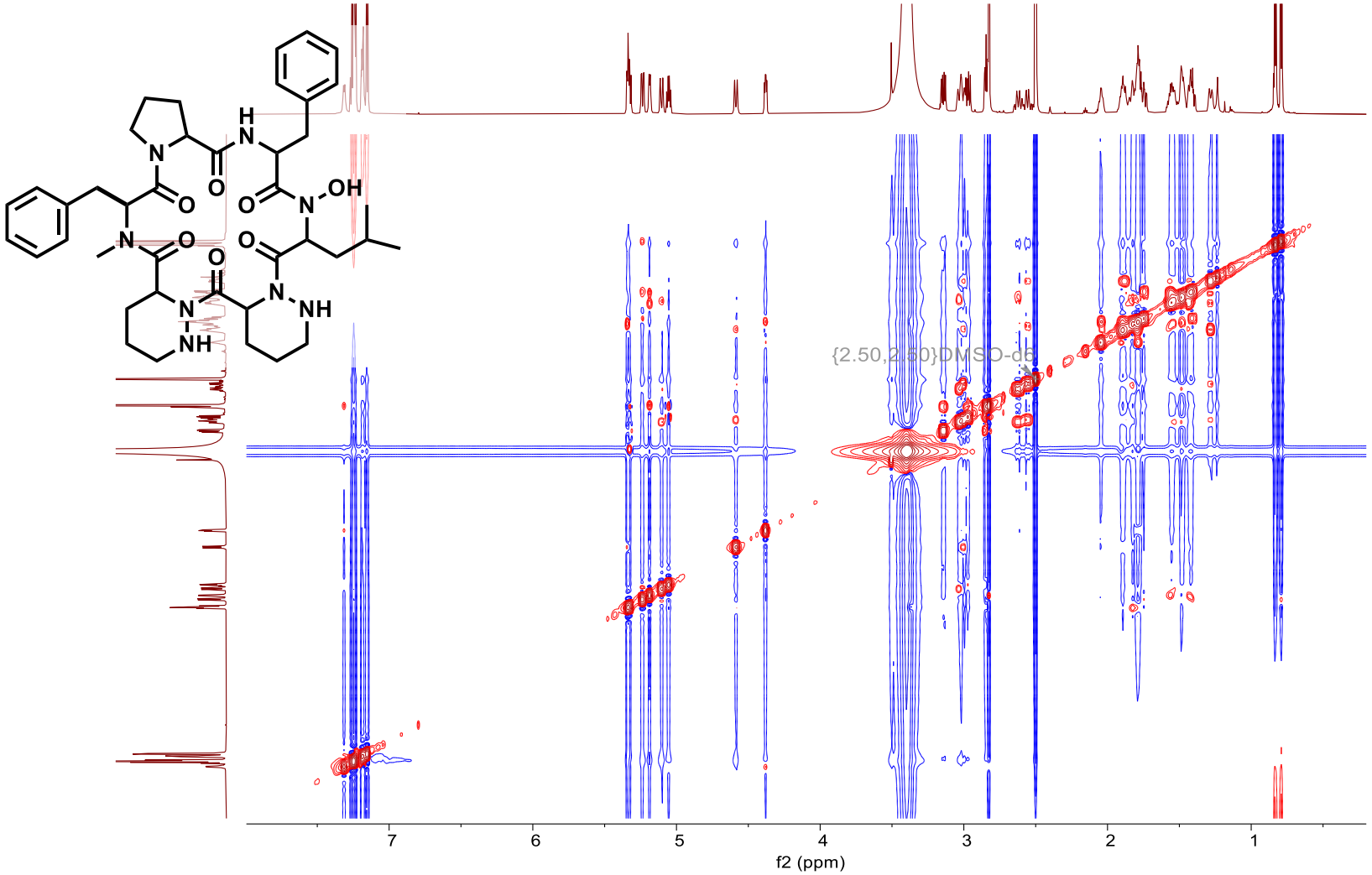


Figure A 84. NOESY NMR (DMSO-d_6, 700 MHz) of PAC3

Table 7. NMR spectral data for compound PAC4 in DMSO-d6 at 700 MHz

carbon #	δ_C	δ_H, mult (J in Hz)	COSY	HMBC
1	47.7	3.00 (d, *J* = 14.2)	2	2, 3
		2.55 (m)	2	–
2	19.6	1.95 (m)	2, 3	3, 4
		1.29 (m)	2,3	3, 4
3	23.9	1.83 (m)	–	–
		1.54 (m)	2, 4	–
4	45.9	5.19 (d, *J* = 5.7)	3	5, 21
5	170.7	–	–	–
6	56.2	5.33 (dd, *J* = 7.4, 13.7)	23	23
7	169.8	–	–	–
8	46.1	3.35 (m)	9	7
		3.27 (m)	9	–
9	24.3	1.91 (m)	10	8, 11
		1.79 (m)	–	8, 11
10	26.9	2.11 (m)	9	8, 11, 12
		1.74 (m)	–	11, 12
11	59.9	4.48 (dd, *J* = 2.9, 8.2)	10	8, 9, 10, 12
12	170.5	–	–	–
13	52.4	4.82 (t, *J* = 8.1)	30	14, 30, 31, 33
14	172.4	–	–	–
15	56.2	5.30 (dd, *J* = 3.6, 10.9)	34	16, 34, 35
16	173.1	–	–	–
17	46.3	3.01(d, *J* = 12.3)	18	19
		2.66 (dd, *J* = 2.5, 12.3)	18	19
18	20.5	1.59 (m)	17, 19	17, 20
		1.48 (m)	17	17
19	24.0	1.79 (m)	18, 20	20
		1.49 (m)	–	–
20	46.4	5.42 (dd, *J* = 1.8, 6.7)	19	18, 19, 21
21	172.7	–	–	–
22	31.1	2.78 (s)	–	5, 6
23	33.9	3.13 (dd, *J* = 6.7, 13.7)	6	6, 24, 25
		2.81 (dd, *J* = 8.2, 14.1)	6	6, 24, 25
24	137.8	–	–	–
25	129.2	7.23 (m)	26	27
26	128.0	7.25 (m)	25, 27	24, 25
27	126.3	7.18 (m)	26, 28	25
28	128.0	7.25 (m)	25, 27	24, 25
29	129.2	7.23 (m)	26	27
30	36.2	1.73 (m)	13, 33	13, 31, 32, 33
31	25.8	1.30 (m)	32	30, 32, 33
		1.05 (m)	30, 32	30, 32, 33
32	11.1	0.846 (s)	31	30, 31
33	14.6	0.86 (s)	30	13, 30, 31
34	35.5	1.81 (m)	15	15, 35, 36, 37
		1.44 (m)	15, 35	35
35	25.1	1.64 (m)	34, 36, 37	34
36	23.2	0.89 (d, *J* = 6.7)	35	34, 35, 37
37	20.8	0.846 (d, *J* = 6.7)	35	34, 35, 36

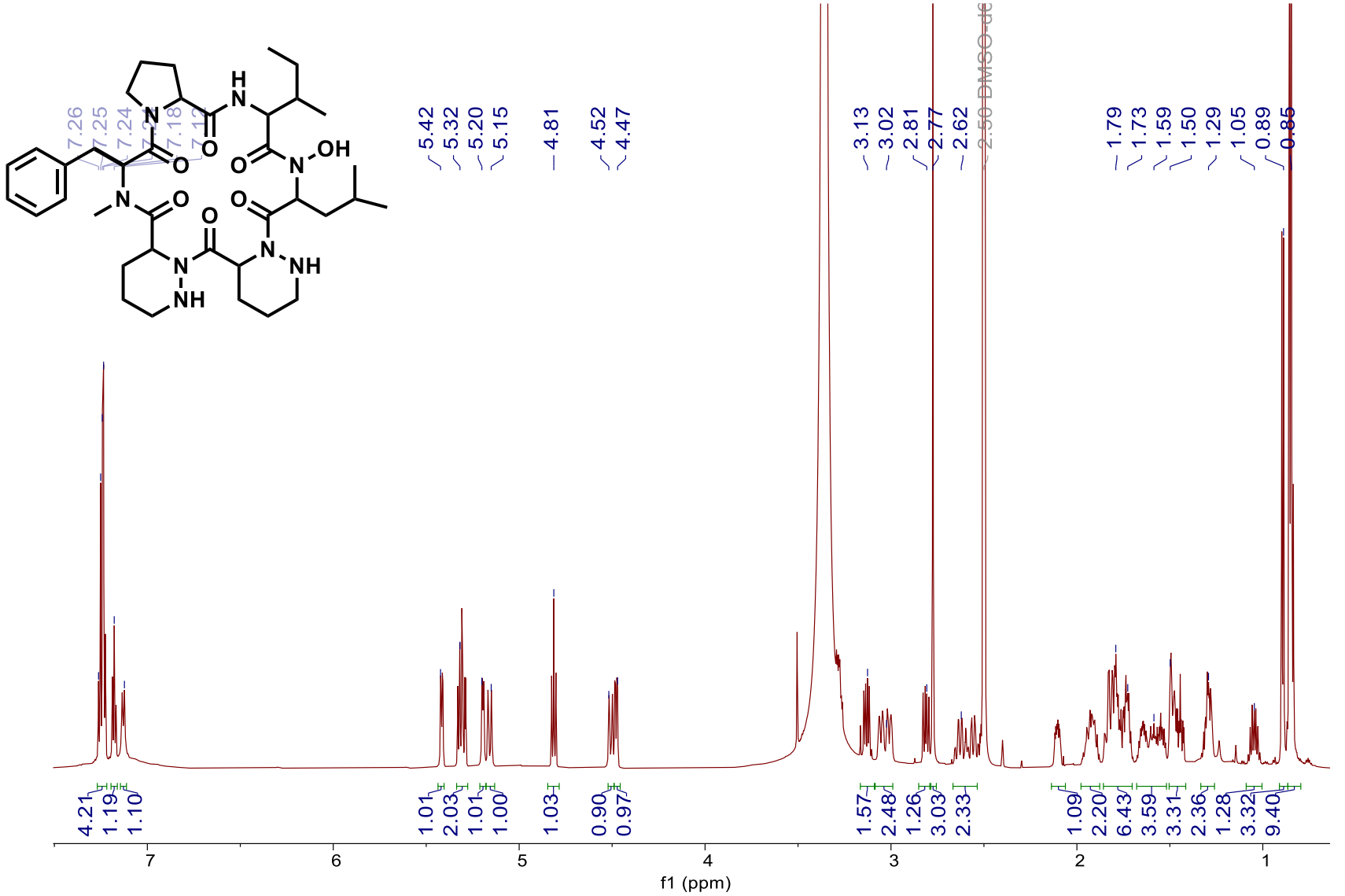


Figure A 85. ^{1}H NMR (DMSO-d_6, 700 MHz) of PAC4

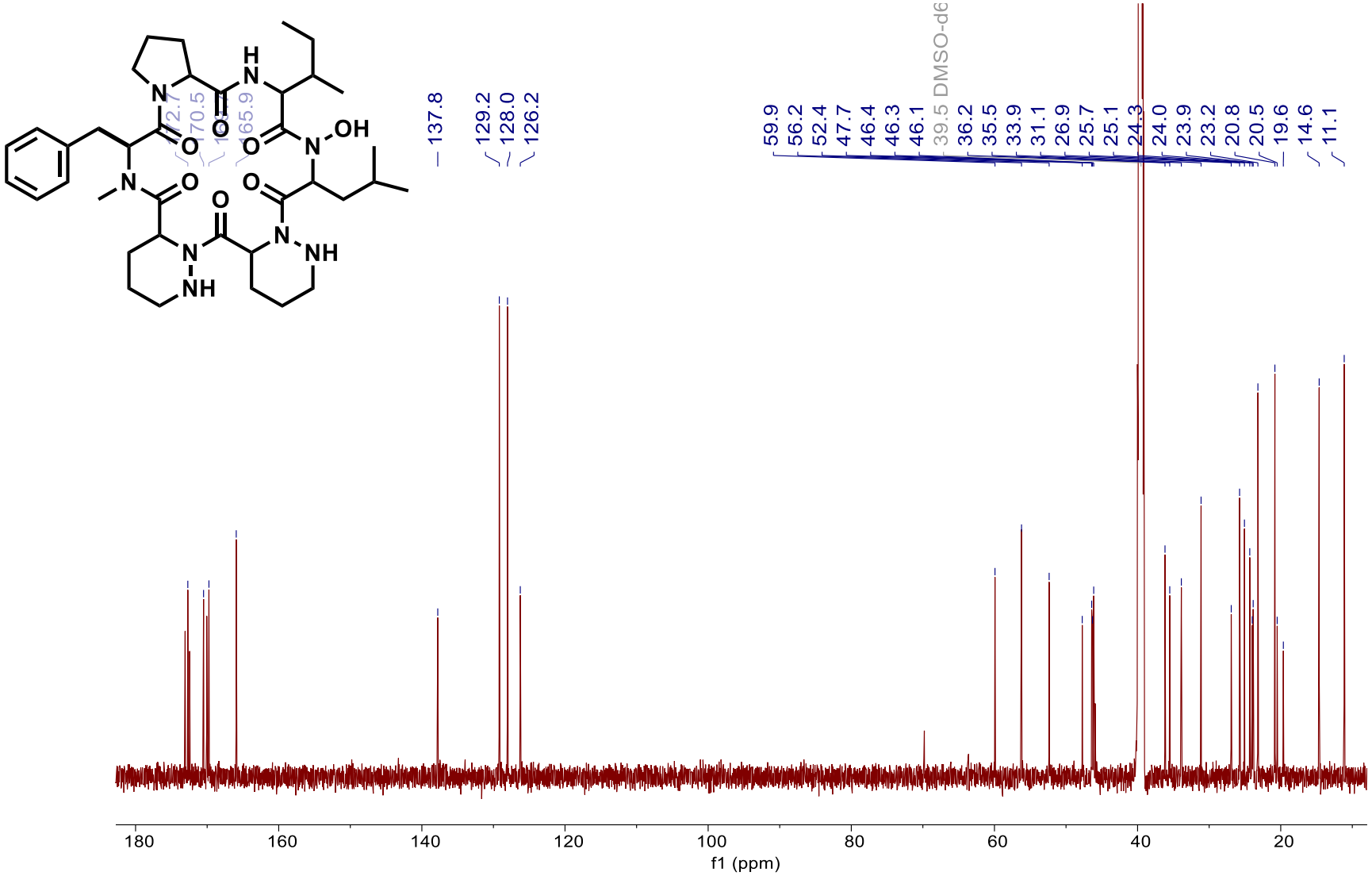


Figure A 86. ^{13}C NMR (DMSO-d_6, 700 MHz) of PAC4

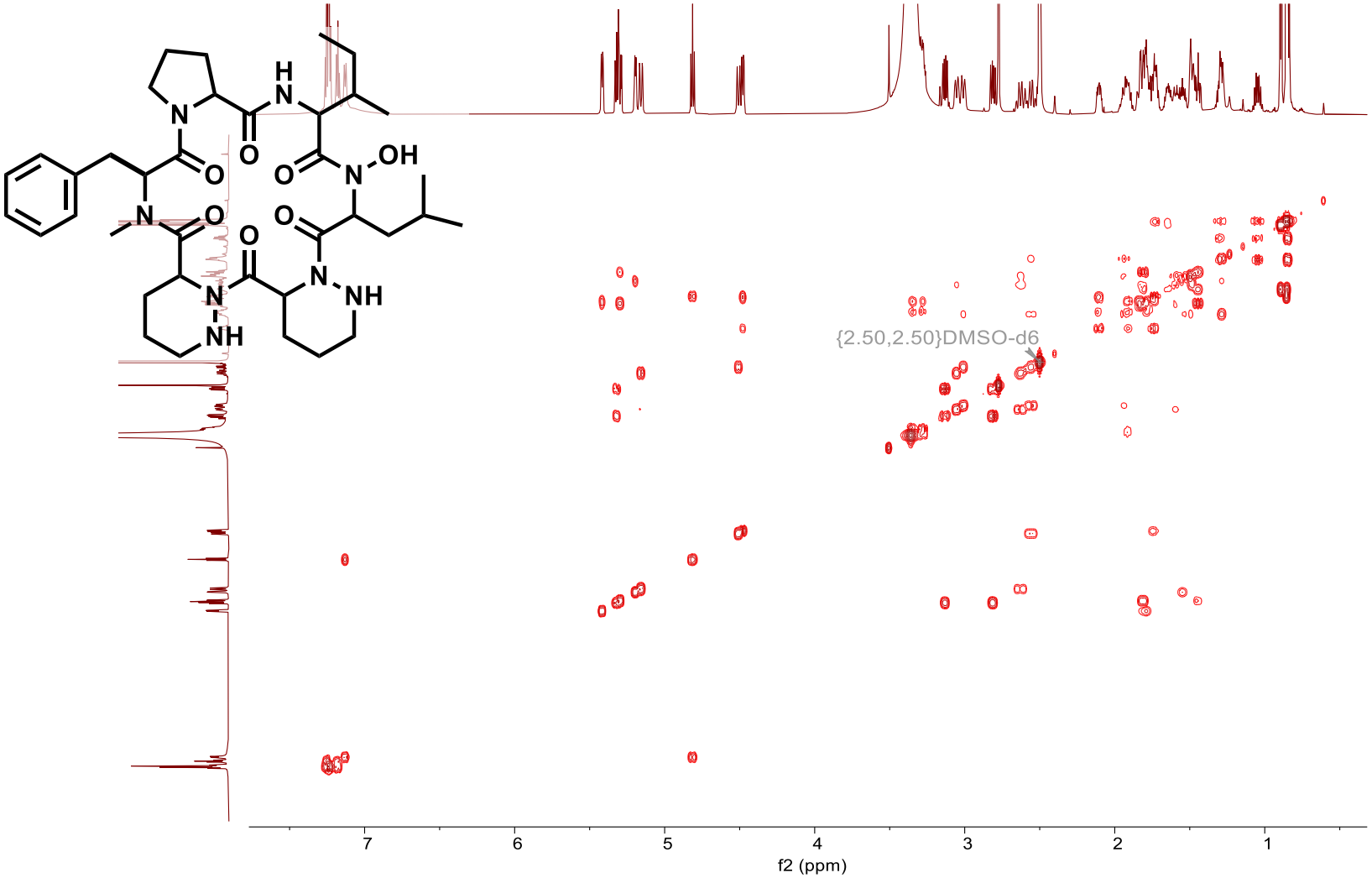


Figure A 87. COSY NMR (DMSO-d_6, 700 MHz) of PAC4

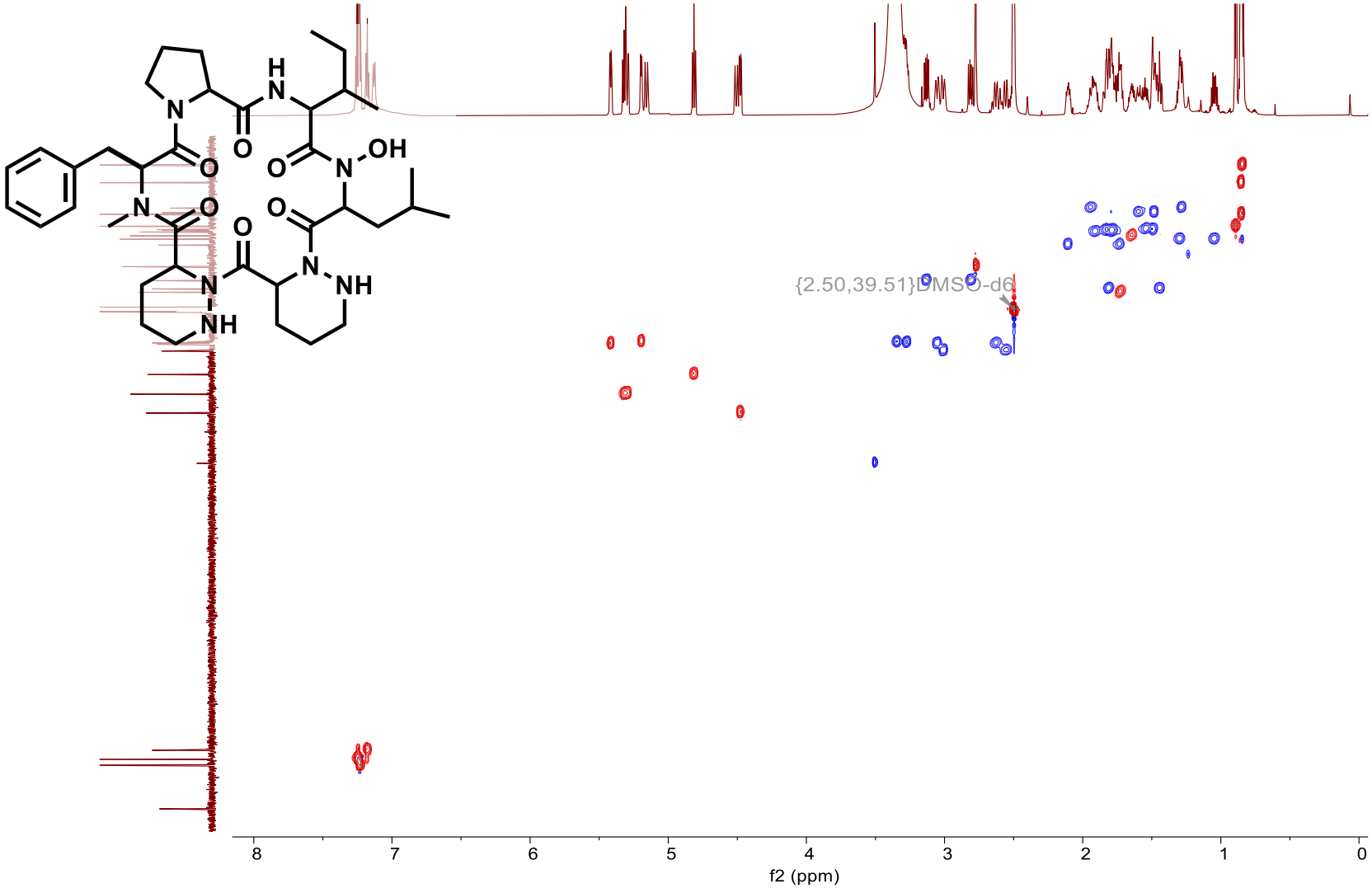


Figure A 88. HSQC NMR (DMSO-d_6, 700 MHz) of PAC4

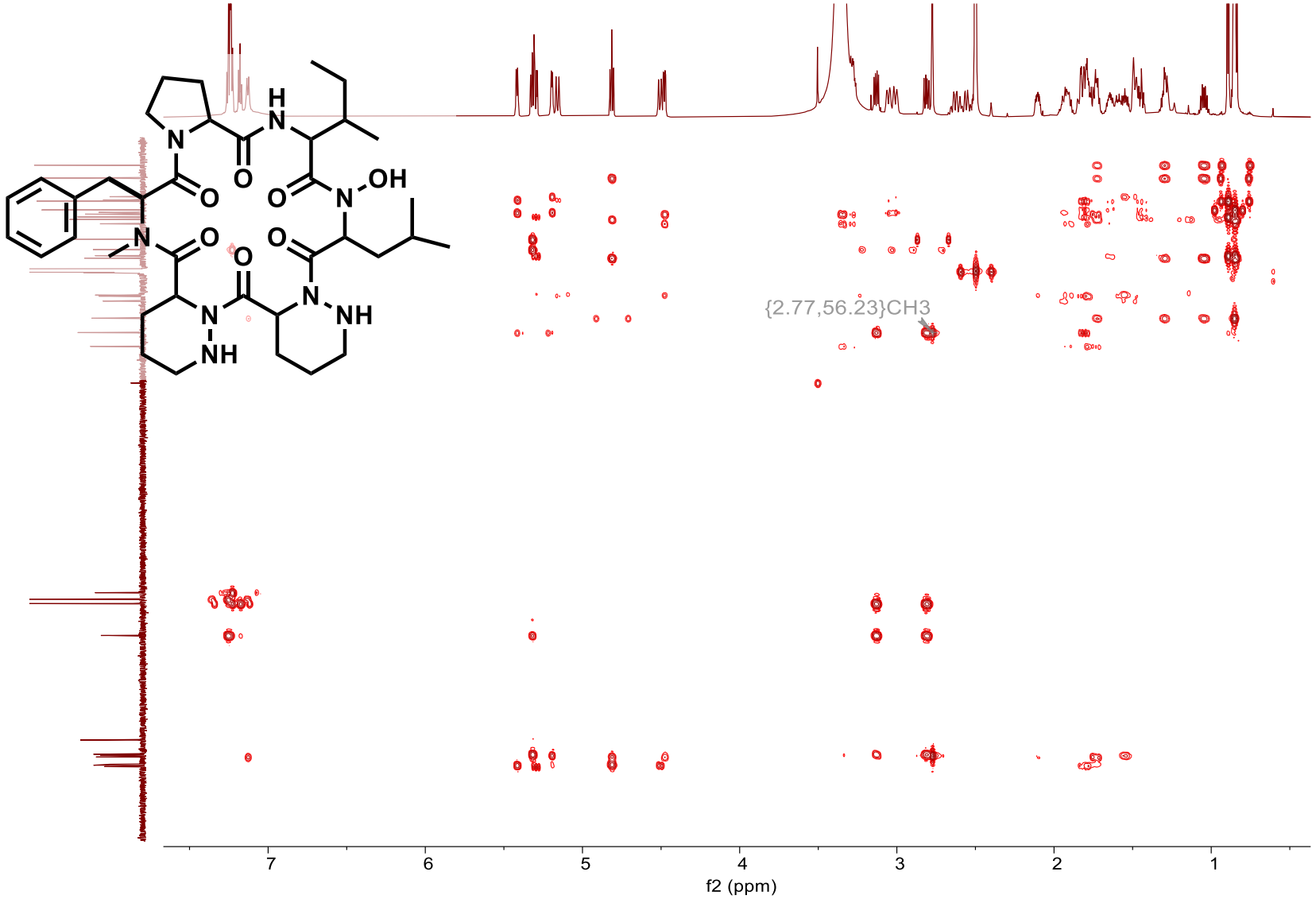


Figure A 89. HMBC NMR (DMSO-d_6, 700 MHz) of PAC4

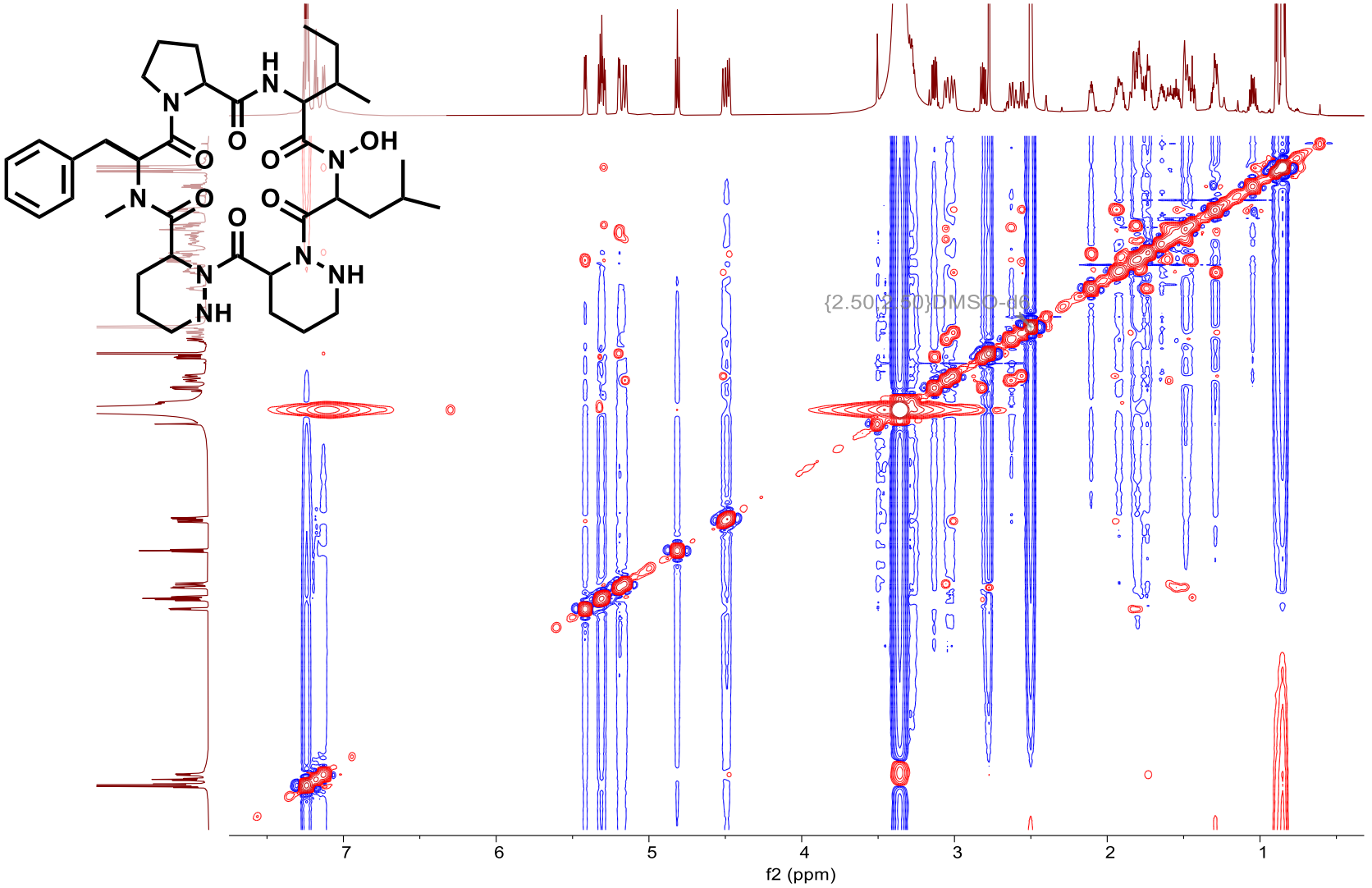


Figure A 90. NOESY NMR (DMSO-d_6, 700 MHz) of PAC4.

A.5 Copyright licenses

Document A 1. Copyright permission for Figure 2

CCC Marketplace

This is a License Agreement between University of Tübingen ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	09-Apr-2026	Type of Use	Republish in a thesis/dissertation
Order License ID	1717791-1		
ISSN	1098-5557	Publisher	AMERICAN SOCIETY FOR MICROBIOLOGY
		Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Microbiology and molecular biology reviews : MMBR	Publication Type	e-Journal
		Start Page	1
Article Title	Taxonomy, Physiology, and Natural Products of Actinobacteria.	End Page	43
		Issue	1
		Volume	80
Author / Editor	American Society for Microbiology.	URL	http://mmbr.asm.org
Date	01/01/1997		
Language	English		
Country	United States of America		
Rightsholder	American Society for Microbiology - Journals		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	Distribution	Worldwide
Number of Charts / Graphs / Tables / Figures Requested	2	Translation	Original language of publication
		Copies for the Disabled?	No
Format (select all that apply)	Print, Electronic	Minor Editing Privileges?	Yes
		Incidental Promotional Use?	No
Who Will Republish the Content?	Author of requested content		
		Currency	EUR
Duration of Use	Life of current edition		
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product, any product related to main product, and other compilations/derivative products		

NEW WORK DETAILS

Title	Discovery of Natural Product in Actinobacteria	Institution Name	University of Tuebingen
		Expected Presentation Date	2026-06-01
Instructor Name	Prof. Dr. Nadine Ziemert		

ADDITIONAL DETAILS

Order Reference Number	N/A	**The Requesting Person / Organization to Appear on the License**	University of Tübingen

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Thesis_Figure 2	**Title of the Article / Chapter the Portion Is From**	Taxonomy, Physiology, and Natural Products of Actinobacteria.
Editor of Portion(s)	Barka, Essaid Ait; Vatsa, Parul; Sanchez, Lisa; Gaveau-Vaillant, Nathalie; Jacquard, Cedric; Klenk, Hans-Peter; Clément, Christophe; Ouhdouch, Yder; van Wezel, Gilles P.	**Author of Portion(s)**	Barka, Essaid Ait; Vatsa, Parul; Sanchez, Lisa; Gaveau-Vaillant, Nathalie; Jacquard, Cedric; Klenk, Hans-Peter; Clément, Christophe; Ouhdouch, Yder; van Wezel, Gilles P.
Volume / Edition	80	**Issue, if Republishing an Article From a Serial**	1
Page or Page Range of Portion	1-43	**Publication Date of Portion**	2016-03-01

RIGHTSHOLDER TERMS AND CONDITIONS

We ask that you review the copyright information associated with the article that you are interested in, as some articles in the ASM Journals are covered by creative commons licenses that would permit free reuse with attribution. Please note that ASM cannot grant permission to reuse figures or images that are credited to publications other than ASM journals. For images credited to non-ASM journal publications, you will need to obtain permission from the journal referenced in the figure or table legend or credit line before making any use of the image(s) or table(s). ASM is unable to grant permission to reuse supplemental material as copyright for the supplemental material remains with the author. All requestors must include the copyright and permission notice in connection with any reproduction of the licensed material, i.e. Journal name, year, volume, page numbers, DOI and reproduced/amended with permission from American Society for Microbiology. Translations The translation must be of high quality and match the standard of the original article. The translation must be a true reproduction word for word. All drug names must be generic; no proprietary/trade names may be substituted. No additional text, tables or figures may be added to the translated text. The integrity of the article should be preserved, i.e., no advertisements will be printed on the article. Articles translated in their entirety must honor the ASM embargo period. The translated version of ASM material must also carry a disclaimer in English and in the language of the translation. The two versions (English and other language) of the disclaimer MUST appear on the inside front cover or at the beginning of the translated material as follows: "The American Society for Microbiology takes no responsibility for the accuracy of the translation from the published English original and is not liable for any errors which may occur. No responsibility is assumed, and responsibility is hereby disclaimed, by the American Society for Microbiology for any injury and/or damage to persons or property as a matter of product liability, negligence or otherwise, or from any use or operation of methods, products, instructions or ideas presented in the Journal. Independent verification of diagnosis and drug dosages should be made. Discussions, views, and recommendations as to medical procedures, choice of drugs and drug dosages are the responsibility of the authors."
This journal title may publish some Open Access articles, which provide specific user rights for reuse. Please refer to the article copyright line to determine whether the content you wish to use is Open Access. For Open Access content, permission to reuse is granted subject to the terms of the License under which the work was published. Please check the License conditions for the work which you wish to reuse. Full and appropriate attribution must be given. This permission does not cover any third party copyrighted material which may appear in the work requested. Please contact the publisher with any questions.

SPECIAL RIGHTSHOLDER TERMS AND CONDITIONS

Hello: Please indicate that the images are "Modified with permission from in the image legend. Thank you.

Marketplace Permissions General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is unlawful, including without limitation if such use would violate applicable sanctions laws or regulations, would be defamatory, violate the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. Any unlawful use will render any licenses hereunder null and void. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is unlawful, including without limitation if such use would violate applicable sanctions laws or regulations, would be defamatory, violate the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. Any unlawful use will render any licenses hereunder null and void. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) ***Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).*** For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) Books and Records; Right to Audit. As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) ***Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).*** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content,** which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to "clip" or reformat the requested material for purposes of time or content management or ease of delivery, provided that such "clipping" or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Conformation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;

ii) the input of Works or reproductions thereof into any computerized database;

iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;

iv) reproduction for resale to anyone other than a specific customer of User;

v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet).*** For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.

ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified without additional express permission, and a Work consisting of audiovisual content may, if necessary, be "clipped" or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or "clipping" does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC's privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

Last updated October 2022

Document A 2. Copyright permission for Figure 9

ELSEVIER LICENSE
TERMS AND CONDITIONS

Apr 09, 2026

This Agreement between Shuning Xia/University of Tübingen ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	6244750665715
License date	Apr 09, 2026
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell Systems
Licensed Content Title	Self-resistance-gene-guided, high-throughput automated genome mining of bioactive natural products from Streptomyces
Licensed Content Author	Yujie Yuan,Chunshuai Huang,Nilmani Singh,Guanhua Xun,Huimin Zhao
Licensed Content Date	Mar 19, 2025
Licensed Content Volume	16
Licensed Content Issue	3
Licensed Content Pages	1
Start Page	101237
End Page	0
Type of Use	reuse in a thesis/dissertation

Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title of new work	Discovery of Natural Product in Actinobacteria
Institution name	University of Tübingen
Expected presentation date	Jun 2026
Portions	Figure 2
The Requesting Person / Organization to Appear on the License	Shuning Xia/University of Tübingen
Requestor Location	Shuning Xia Auf der Morgenstelle 28 Tübingen, 72076 Germany
Publisher Tax ID	GB 494 6272 12
Total	0.00 EUR

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your RightsLink account and that are available at any time at https://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given. The material may not be reproduced or used in any other way, including use in combination with an artificial intelligence tool (including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool), or to create any derivative work and/or service (including resulting from the use of artificial intelligence tools).

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier's permissions helpdesk here). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all

claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.11

Questions? customercare@copyright.com.

Document A 3. Copyright permission for Figure 10 A

CCC Marketplace

This is a License Agreement between University of Tübingen ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	09-Apr-2026	Type of Use	Republish in a thesis/dissertation
Order License ID	1717803-1		
ISSN	1460-4752	Publisher	Royal Society of Chemistry
		Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Natural product reports	Publication Type	e-Journal
Article Title	Applied evolution: phylogeny-based approaches in natural products research.	Start Page	1295
		End Page	1312
		Issue	9
Author / Editor	Royal Society of Chemistry (Great Britain)	Volume	36
		URL	http://firstsearch.oclc.org/journal=0265-0568;screen=info;ECOIP
Date	01/01/1984		
Language	English		
Country	United Kingdom of Great Britain and Northern Ireland		
Rightsholder	Royal Society of Chemistry		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	Distribution	Worldwide
Number of Charts / Graphs / Tables / Figures Requested	1	Translation	Original language of publication
		Copies for the Disabled?	No
Format (select all that apply)	Print, Electronic	Minor Editing Privileges?	Yes
Who Will Republish the Content?	Author of requested content	Incidental Promotional Use?	No
		Currency	EUR
Duration of Use	Life of current edition		
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product, any product related to main product, and other compilations/derivative products		

NEW WORK DETAILS

Title	The Discovery of Natural Products in Actinobacteria	Institution Name	University of Tübingen
		Expected Presentation Date	2026-06-01
Instructor Name	Prof. Dr. Nadine Ziemert		

ADDITIONAL DETAILS

Order Reference Number	N/A	The Requesting Person / Organization to Appear on the License	University of Tübingen

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Thesis_Figure 10	Title of the Article / Chapter the Portion Is From	Applied evolution: phylogeny-based approaches in natural products research.
Editor of Portion(s)	Adamek, Martina; Alanjary, Mohammad; Ziemert, Nadine	Author of Portion(s)	Adamek, Martina; Alanjary, Mohammad; Ziemert, Nadine
Volume / Edition	36		
Page or Page Range of Portion	1295-1312	Issue, if Republishing an Article From a Serial	9
		Publication Date of Portion	2019-09-01

Marketplace Permissions General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is unlawful, including without limitation if such use would violate applicable sanctions laws or regulations, would be defamatory, violate the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. Any unlawful use will render any licenses hereunder null and void. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach

not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) ***Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).*** For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) Books and Records; Right to Audit. As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) ***Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).*** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content,** which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to "clip" or reformat the requested material for purposes of time or content management or ease of delivery, provided that such "clipping" or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Conformation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

ii) Unless expressly set forth in the relevant Order Confirmation, no License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied or, in the case of Works subject to Sections 14(b)(1)(B) or (C) above, as described in such Sections) (ii) permit "publishing ventures" where any particular course materials would be systematically marketed at multiple institutions.

iii) Subject to any further limitations determined in the Rightsholder Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the electronic course content pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to not more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular materials, whether photocopied or electronic, at more than one institution of learning;

E) electronic access to material which is the subject of an electronic-use permission must be limited by means of electronic password, student identification or other control permitting access solely to students and instructors in the class;

F) User must ensure (through use of an electronic cover page or other appropriate means) that any person, upon gaining electronic access to the material, which is the subject of a permission, shall see:

- a proper copyright notice, identifying the Rightsholder in whose name CCC has granted permission,
- a statement to the effect that such copy was made pursuant to permission,
- a statement identifying the class to which the material applies and notifying the reader that the material has been made available electronically solely for use in the class, and
- a statement to the effect that the material may not be further distributed to any person outside the class, whether by copying or by transmission and whether electronically or in paper form, and User must also

ensure that such cover page or other means will print out in the event that the person accessing the material chooses to print out the material or any part thereof.

G) any permission granted shall expire at the end of the class and, absent some other form of authorization, User is thereupon required to delete the applicable material from any electronic storage or to block electronic access to the applicable material.

iv) Uses of separate portions of a Work, even if they are to be included in the same course material or the same university or college class, require separate permissions under the electronic course content pay-per-use Service. Unless otherwise provided in the Order Confirmation, any grant of rights to User is limited to use completed no later than the end of the academic term (or analogous period) as to which any particular permission is granted.

v) Books and Records; Right to Audit. As to each permission granted under the electronic course content Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any electronic copies used by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this license for any reason.

c) ***Pay-Per-Use Permissions for Certain Reproductions (Academic photocopies for library reserves and interlibrary loan reporting) (Non-academic internal/external business uses and commercial document delivery).*** The License expressly excludes the uses listed in Section (c)(i)-(v) below (which must be subject to separate license from the applicable Rightsholder) for: academic photocopies for library reserves and interlibrary loan reporting; and non-academic internal/external business uses and commercial document delivery.

i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;

ii) the input of Works or reproductions thereof into any computerized database;

iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;

iv) reproduction for resale to anyone other than a specific customer of User;

v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet).*** For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.

ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified without additional express permission, and a Work consisting of audiovisual content may, if necessary, be "clipped" or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or "clipping" does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the

prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC's privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

Last updated October 2022

Document A 4. Copyright permission for Figure 10 B

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Apr 09, 2026

This Agreement between Shuning Xia/University of Tübingen ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	6244750108133
License date	Apr 09, 2026
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Chemical Biology
Licensed Content Title	Genome mining for unknown–unknown natural products
Licensed Content Author	Danielle A. Yee et al
Licensed Content Date	Jan 26, 2023
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Would you like a high resolution image with your order?	no

Will you be translating?	no
Circulation/distribution	50000 or greater
Author of this Springer Nature content	no
Title of new work	Discovery of Natural Product in Actinobacteria
Institution name	University of Tübingen
Expected presentation date	Jun 2026
Portions	Figure 2
The Requesting Person / Organization to Appear on the License	Shuning Xia/University of Tübingen
Requestor Location	Shuning Xia Auf der Morgenstelle 28 Tübingen, 72076 Germany
Payment Type	Invoice
Email Address	shuning.xia@student.uni-tuebingen.de
Billing Address	Shuning Xia Auf der Morgenstelle 28 Tübingen, Germany 72076
Total	0.00 EUR
Terms and Conditions	

Springer Nature Customer Service Centre GmbH Terms and Conditions

The following terms and conditions ("Terms and Conditions") together with the terms specified in your [RightsLink] constitute the License ("License") between you as

Licensee and Springer Nature Customer Service Centre GmbH as Licensor. By clicking 'accept' and completing the transaction for your use of the material ("Licensed Material"), you confirm your acceptance of and obligation to be bound by these Terms and Conditions.

1. Grant and Scope of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, non-sublicensable, revocable, world-wide License to reproduce, distribute, communicate to the public, make available, broadcast, electronically transmit or create derivative works using the Licensed Material for the purpose(s) specified in your RightsLink Licence Details only. Licenses are granted for the specific use requested in the order and for no other use, subject to these Terms and Conditions. You acknowledge and agree that the rights granted to you under this License do not include the right to modify, edit, translate, include in collective works, or create derivative works of the Licensed Material in whole or in part unless expressly stated in your RightsLink Licence Details. You may use the Licensed Material only as permitted under this Agreement and will not reproduce, distribute, display, perform, or otherwise use or exploit any Licensed Material in any way, in whole or in part, except as expressly permitted by this License.

1. 2. You may only use the Licensed Content in the manner and to the extent permitted by these Terms and Conditions, by your RightsLink Licence Details and by any applicable laws.

1. 3. A separate license may be required for any additional use of the Licensed Material, e.g. where a license has been purchased for print use only, separate permission must be obtained for electronic re-use. Similarly, a License is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the License.

1. 4. Any content within the Licensed Material that is owned by third parties is expressly excluded from the License.

1. 5. Rights for additional reuses such as custom editions, computer/mobile applications, film or TV reuses and/or any other derivative rights requests require additional permission and may be subject to an additional fee. Please apply to journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

2. Reservation of Rights

Licensor reserves all rights not expressly granted to you under this License. You acknowledge and agree that nothing in this License limits or restricts Licensor's rights in or use of the Licensed Material in any way. Neither this License, nor any act, omission, or statement by Licensor or you, conveys any ownership right to you in any Licensed Material, or to any element or portion thereof. As between Licensor and you, Licensor owns and retains all right, title, and interest in and to the Licensed Material subject to the license granted in Section 1.1. Your permission to use the Licensed Material is expressly conditioned on you not impairing Licensor's or the applicable copyright owner's rights in the Licensed Material in any way.

3. Restrictions on use

3. 1. Minor editing privileges are allowed for adaptations for stylistic purposes or formatting purposes provided such alterations do not alter the original meaning or intention of the Licensed Material and the new figure(s) are still accurate and representative of the Licensed Material. Any other changes including but not

limited to, cropping, adapting, and/or omitting material that affect the meaning, intention or moral rights of the author(s) are strictly prohibited.

3. 2. You must not use any Licensed Material as part of any design or trademark.

3. 3. Licensed Material may be used in Open Access Publications (OAP), but any such reuse must include a clear acknowledgment of this permission visible at the same time as the figures/tables/illustration or abstract and which must indicate that the Licensed Material is not part of the governing OA license but has been reproduced with permission. This may be indicated according to any standard referencing system but must include at a minimum 'Book/Journal title, Author, Journal Name (if applicable), Volume (if applicable), Publisher, Year, reproduced with permission from SNCSC'.

4. STM Permission Guidelines

4. 1. An alternative scope of license may apply to signatories of the STM Permissions Guidelines ("STM PG") as amended from time to time and made available at https://www.stm-assoc.org/intellectual-property/permissions/permissions-guidelines/.

4. 2. For content reuse requests that qualify for permission under the STM PG, and which may be updated from time to time, the STM PG supersede the terms and conditions contained in this License.

4. 3. If a License has been granted under the STM PG, but the STM PG no longer apply at the time of publication, further permission must be sought from the Rightsholder. Contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

5. Duration of License

5. 1. Unless otherwise indicated on your License, a License is valid from the date of purchase ("License Date") until the end of the relevant period in the below table:

Reuse in a medical communications project	Reuse up to distribution or time period indicated in License
Reuse in a dissertation/thesis	Lifetime of thesis
Reuse in a journal/magazine	Lifetime of journal/magazine
Reuse in a book/textbook	Lifetime of edition
Reuse on a website	1 year unless otherwise specified in the License. If you wish to reuse the content on your website for longer than 1 year, please make this clear in the 'additional information' field and the License will include a 'Special Term' to reflect your duration choice.
Reuse in a presentation/slide kit/poster	Lifetime of presentation/slide kit/poster. Note: publication whether electronic or in print of presentation/slide kit/poster may require further permission.
Reuse in conference proceedings	Lifetime of conference proceedings

Reuse in an annual report	Lifetime of annual report
Reuse in training/CME materials	Reuse up to distribution or time period indicated in License
Reuse in newsmedia	Lifetime of newsmedia
Reuse in coursepack/classroom materials	Reuse up to distribution and/or time period indicated in license

6. Acknowledgement

6. 1. The Licensor's permission must be acknowledged next to the Licensed Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract and must be hyperlinked to the journal/book's homepage.

6. 2. Acknowledgement may be provided according to any standard referencing system and at a minimum should include "Author, Article/Book Title, Journal name/Book imprint, volume, page number, year, Springer Nature".

7. Reuse in a dissertation or thesis

7. 1. Where 'reuse in a dissertation/thesis' has been selected, the following terms apply: Print rights of the Version of Record are provided for; electronic rights for use only on institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/) and only up to what is required by the awarding institution.

7. 2. For theses published under an ISBN or ISSN, separate permission is required. Please contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

7. 3. Authors must properly cite the published manuscript in their thesis according to current citation standards and include the following acknowledgement: *'Reproduced with permission from Springer Nature'.*

8. License Fee

You must pay the fee set forth in the License Agreement (the "License Fees"). All amounts payable by you under this License are exclusive of any sales, use, withholding, value added or similar taxes, government fees or levies or other assessments. Collection and/or remittance of such taxes to the relevant tax authority shall be the responsibility of the party who has the legal obligation to do so.

9. Warranty

9. 1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. **You are solely responsible for ensuring that the material you wish to license is original to the Licensor and does not carry the copyright of another entity or third party (as credited in the published version).** If the credit line on any part of the Licensed Material indicates that it was reprinted or adapted with permission from another source, then you should seek additional permission from that source to reuse the material.

9. 2. EXCEPT FOR THE EXPRESS WARRANTY STATED HEREIN AND TO THE EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR PROVIDES THE LICENSED MATERIAL "AS IS" AND MAKES NO OTHER

REPRESENTATION OR WARRANTY. LICENSOR EXPRESSLY DISCLAIMS ANY LIABILITY FOR ANY CLAIM ARISING FROM OR OUT OF THE CONTENT, INCLUDING BUT NOT LIMITED TO ANY ERRORS, INACCURACIES, OMISSIONS, OR DEFECTS CONTAINED THEREIN, AND ANY IMPLIED OR EXPRESS WARRANTY AS TO MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, PUNITIVE, OR EXEMPLARY DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE LICENSED MATERIAL REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION APPLIES NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

10. Termination and Cancellation

10. 1. The License and all rights granted hereunder will continue until the end of the applicable period shown in Clause 5.1 above. Thereafter, this license will be terminated and all rights granted hereunder will cease.

10. 2. Licensor reserves the right to terminate the License in the event that payment is not received in full or if you breach the terms of this License.

11. General

11. 1. The License and the rights and obligations of the parties hereto shall be construed, interpreted and determined in accordance with the laws of the Federal Republic of Germany without reference to the stipulations of the CISG (United Nations Convention on Contracts for the International Sale of Goods) or to Germany's choice-of-law principle.

11. 2. The parties acknowledge and agree that any controversies and disputes arising out of this License shall be decided exclusively by the courts of or having jurisdiction for Heidelberg, Germany, as far as legally permissible.

11. 3. This License is solely for Licensor's and Licensee's benefit. It is not for the benefit of any other person or entity.

Questions? For questions on Copyright Clearance Center accounts or website issues please contact springernaturesupport@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777. For questions on Springer Nature licensing please visit https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution

Other Conditions:

Version 1.5 - June 2025

Questions? customercare@copyright.com.

Document A 5. Copyright permission for Figure 11

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Mar 31, 2026

This Agreement between University of Tübingen ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	6239540270169
License date	Mar 31, 2026
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Methods
Licensed Content Title	Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices
Licensed Content Author	Saleh Alseekh et al
Licensed Content Date	Jul 8, 2021
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no

Circulation/distribution	50000 or greater
Author of this Springer Nature content	no
Title of new work	Discovery of Natural Product in Actinobacteria
Institution name	University of Tübingen
Expected presentation date	Jun 2026
Portions	Figure 1
The Requesting Person / Organization to Appear on the License	University of Tübingen
Requestor Location	Shuning Xia Auf der Morgenstelle 28 Tübingen, 72076 Germany
Order reference number	Thesis_Figure 11
Payment Type	Invoice
Email Address	shuning.xia@student.uni-tuebingen.de
Billing Address	Shuning Xia Auf der Morgenstelle 28 Tübingen, Germany 72076
Total	0.00 EUR

Terms and Conditions

Springer Nature Customer Service Centre GmbH Terms and Conditions

The following terms and conditions ("Terms and Conditions") together with the terms specified in your [RightsLink] constitute the License ("License") between you as Licensee and Springer Nature Customer Service Centre GmbH as Licensor. By

clicking 'accept' and completing the transaction for your use of the material ("Licensed Material"), you confirm your acceptance of and obligation to be bound by these Terms and Conditions.

1. Grant and Scope of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, non-sublicensable, revocable, world-wide License to reproduce, distribute, communicate to the public, make available, broadcast, electronically transmit or create derivative works using the Licensed Material for the purpose(s) specified in your RightsLink Licence Details only. Licenses are granted for the specific use requested in the order and for no other use, subject to these Terms and Conditions. You acknowledge and agree that the rights granted to you under this License do not include the right to modify, edit, translate, include in collective works, or create derivative works of the Licensed Material in whole or in part unless expressly stated in your RightsLink Licence Details. You may use the Licensed Material only as permitted under this Agreement and will not reproduce, distribute, display, perform, or otherwise use or exploit any Licensed Material in any way, in whole or in part, except as expressly permitted by this License.

1. 2. You may only use the Licensed Content in the manner and to the extent permitted by these Terms and Conditions, by your RightsLink Licence Details and by any applicable laws.

1. 3. A separate license may be required for any additional use of the Licensed Material, e.g. where a license has been purchased for print use only, separate permission must be obtained for electronic re-use. Similarly, a License is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the License.

1. 4. Any content within the Licensed Material that is owned by third parties is expressly excluded from the License.

1. 5. Rights for additional reuses such as custom editions, computer/mobile applications, film or TV reuses and/or any other derivative rights requests require additional permission and may be subject to an additional fee. Please apply to journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

2. Reservation of Rights

Licensor reserves all rights not expressly granted to you under this License. You acknowledge and agree that nothing in this License limits or restricts Licensor's rights in or use of the Licensed Material in any way. Neither this License, nor any act, omission, or statement by Licensor or you, conveys any ownership right to you in any Licensed Material, or to any element or portion thereof. As between Licensor and you, Licensor owns and retains all right, title, and interest in and to the Licensed Material subject to the license granted in Section 1.1. Your permission to use the Licensed Material is expressly conditioned on you not impairing Licensor's or the applicable copyright owner's rights in the Licensed Material in any way.

3. Restrictions on use

3. 1. Minor editing privileges are allowed for adaptations for stylistic purposes or formatting purposes provided such alterations do not alter the original meaning or intention of the Licensed Material and the new figure(s) are still accurate and representative of the Licensed Material. Any other changes including but not limited to, cropping, adapting, and/or omitting material that affect the meaning,

intention or moral rights of the author(s) are strictly prohibited.

3. 2. You must not use any Licensed Material as part of any design or trademark.

3. 3. Licensed Material may be used in Open Access Publications (OAP), but any such reuse must include a clear acknowledgment of this permission visible at the same time as the figures/tables/illustration or abstract and which must indicate that the Licensed Material is not part of the governing OA license but has been reproduced with permission. This may be indicated according to any standard referencing system but must include at a minimum 'Book/Journal title, Author, Journal Name (if applicable), Volume (if applicable), Publisher, Year, reproduced with permission from SNCSC'.

4. STM Permission Guidelines

4. 1. An alternative scope of license may apply to signatories of the STM Permissions Guidelines ("STM PG") as amended from time to time and made available at https://www.stm-assoc.org/intellectual-property/permissions/permissions-guidelines/.

4. 2. For content reuse requests that qualify for permission under the STM PG, and which may be updated from time to time, the STM PG supersede the terms and conditions contained in this License.

4. 3. If a License has been granted under the STM PG, but the STM PG no longer apply at the time of publication, further permission must be sought from the Rightsholder. Contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

5. Duration of License

5. 1. Unless otherwise indicated on your License, a License is valid from the date of purchase ("License Date") until the end of the relevant period in the below table:

Reuse in a medical communications project	Reuse up to distribution or time period indicated in License
Reuse in a dissertation/thesis	Lifetime of thesis
Reuse in a journal/magazine	Lifetime of journal/magazine
Reuse in a book/textbook	Lifetime of edition
Reuse on a website	1 year unless otherwise specified in the License. If you wish to reuse the content on your website for longer than 1 year, please make this clear in the 'additional information' field and the License will include a 'Special Term' to reflect your duration choice.
Reuse in a presentation/slide kit/poster	Lifetime of presentation/slide kit/poster. Note: publication whether electronic or in print of presentation/slide kit/poster may require further permission.
Reuse in conference proceedings	Lifetime of conference proceedings
Reuse in an annual report	Lifetime of annual report

Reuse in training/CME materials	Reuse up to distribution or time period indicated in License
Reuse in newsmedia	Lifetime of newsmedia
Reuse in coursepack/classroom materials	Reuse up to distribution and/or time period indicated in license

6. Acknowledgement

6. 1. The Licensor's permission must be acknowledged next to the Licensed Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract and must be hyperlinked to the journal/book's homepage.

6. 2. Acknowledgement may be provided according to any standard referencing system and at a minimum should include "Author, Article/Book Title, Journal name/Book imprint, volume, page number, year, Springer Nature".

7. Reuse in a dissertation or thesis

7. 1. Where 'reuse in a dissertation/thesis' has been selected, the following terms apply: Print rights of the Version of Record are provided for; electronic rights for use only on institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/) and only up to what is required by the awarding institution.

7. 2. For theses published under an ISBN or ISSN, separate permission is required. Please contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

7. 3. Authors must properly cite the published manuscript in their thesis according to current citation standards and include the following acknowledgement: *'Reproduced with permission from Springer Nature'.*

8. License Fee

You must pay the fee set forth in the License Agreement (the "License Fees"). All amounts payable by you under this License are exclusive of any sales, use, withholding, value added or similar taxes, government fees or levies or other assessments. Collection and/or remittance of such taxes to the relevant tax authority shall be the responsibility of the party who has the legal obligation to do so.

9. Warranty

9. 1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. **You are solely responsible for ensuring that the material you wish to license is original to the Licensor and does not carry the copyright of another entity or third party (as credited in the published version).** If the credit line on any part of the Licensed Material indicates that it was reprinted or adapted with permission from another source, then you should seek additional permission from that source to reuse the material.

9. 2. EXCEPT FOR THE EXPRESS WARRANTY STATED HEREIN AND TO THE EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR PROVIDES THE LICENSED MATERIAL "AS IS" AND MAKES NO OTHER REPRESENTATION OR WARRANTY. LICENSOR EXPRESSLY DISCLAIMS ANY LIABILITY FOR ANY CLAIM ARISING FROM OR OUT OF THE

CONTENT, INCLUDING BUT NOT LIMITED TO ANY ERRORS, INACCURACIES, OMISSIONS, OR DEFECTS CONTAINED THEREIN, AND ANY IMPLIED OR EXPRESS WARRANTY AS TO MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, PUNITIVE, OR EXEMPLARY DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE LICENSED MATERIAL REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION APPLIES NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

10. Termination and Cancellation

10. 1. The License and all rights granted hereunder will continue until the end of the applicable period shown in Clause 5.1 above. Thereafter, this license will be terminated and all rights granted hereunder will cease.

10. 2. Licensor reserves the right to terminate the License in the event that payment is not received in full or if you breach the terms of this License.

11. General

11. 1. The License and the rights and obligations of the parties hereto shall be construed, interpreted and determined in accordance with the laws of the Federal Republic of Germany without reference to the stipulations of the CISG (United Nations Convention on Contracts for the International Sale of Goods) or to Germany's choice-of-law principle.

11. 2. The parties acknowledge and agree that any controversies and disputes arising out of this License shall be decided exclusively by the courts of or having jurisdiction for Heidelberg, Germany, as far as legally permissible.

11. 3. This License is solely for Licensor's and Licensee's benefit. It is not for the benefit of any other person or entity.

Questions? For questions on Copyright Clearance Center accounts or website issues please contact springernaturesupport@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777. For questions on Springer Nature licensing please visit https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution

Other Conditions:

Version 1.5 - June 2025

Questions? customercare@copyright.com.

Document A 6. Copyright permission for Figure 12

CCC Marketplace

This is a License Agreement between University of Tübingen ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	31-Mar-2026	Type of Use	Republish in a thesis/dissertation
Order License ID	1714981-1		
ISSN	1460-4752	Publisher	Royal Society of Chemistry
		Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Natural product reports	Publication Type	e-Journal
Article Title	Exploring microbial natural products through NMR-based metabolomics	Start Page	1459
		End Page	1488
		Issue	9
Author / Editor	Royal Society of Chemistry (Great Britain)	Volume	42
Date	01/01/1984	URL	http://firstsearch.oclc.org/journal=0265-0568;screen=info;ECOIP
Language	English		
Country	United Kingdom of Great Britain and Northern Ireland		
Rightsholder	Royal Society of Chemistry		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	Distribution	Worldwide
Number of Charts / Graphs / Tables / Figures Requested	1	Translation	Original language of publication
		Copies for the Disabled?	No
Format (select all that apply)	Print, Electronic	Minor Editing Privileges?	No
Who Will Republish the Content?	Author of requested content	Incidental Promotional Use?	No
Duration of Use	Life of current edition	Currency	EUR
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product, any product related to main product, and other compilations/derivative products		

NEW WORK DETAILS

Title	Discovery of Natural Product in Actinobacteria	Institution Name	University of Tübingen
		Expected Presentation Date	2026-06-01
Instructor Name	Prof. Dr. Nadine Ziemert		

ADDITIONAL DETAILS

Order Reference Number	N/A	The Requesting Person / Organization to Appear on the License	University of Tübingen

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Thesis_Figure 12	Title of the Article / Chapter the Portion Is From	Exploring microbial natural products through NMR-based metabolomics
Editor of Portion(s)	Wang, De-Gao; Hu, Jia-Qi; Wang, Chao-Yi; Liu, Teng; Li, Yue-Zhong; Wu, Changsheng	Author of Portion(s)	Wang, De-Gao; Hu, Jia-Qi; Wang, Chao-Yi; Liu, Teng; Li, Yue-Zhong; Wu, Changsheng
Volume / Edition	42	Issue, if Republishing an Article From a Serial	9
Page or Page Range of Portion	1459-1488	Publication Date of Portion	2025-09-17

Marketplace Permissions General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is unlawful, including without limitation if such use would violate applicable sanctions laws or regulations, would be defamatory, violate the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. Any unlawful use will render any licenses hereunder null and void. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach

not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) ***Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).*** For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) Books and Records; Right to Audit. As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) ***Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).*** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content,** which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to "clip" or reformat the requested material for purposes of time or content management or ease of delivery, provided that such "clipping" or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Conformation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

ii) Unless expressly set forth in the relevant Order Confirmation, no License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied or, in the case of Works subject to Sections 14(b)(1)(B) or (C) above, as described in such Sections) (ii) permit "publishing ventures" where any particular course materials would be systematically marketed at multiple institutions.

iii) Subject to any further limitations determined in the Rightsholder Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the electronic course content pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to not more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular materials, whether photocopied or electronic, at more than one institution of learning;

E) electronic access to material which is the subject of an electronic-use permission must be limited by means of electronic password, student identification or other control permitting access solely to students and instructors in the class;

F) User must ensure (through use of an electronic cover page or other appropriate means) that any person, upon gaining electronic access to the material, which is the subject of a permission, shall see:

- a proper copyright notice, identifying the Rightsholder in whose name CCC has granted permission,
- a statement to the effect that such copy was made pursuant to permission,
- a statement identifying the class to which the material applies and notifying the reader that the material has been made available electronically solely for use in the class, and
- a statement to the effect that the material may not be further distributed to any person outside the class, whether by copying or by transmission and whether electronically or in paper form, and User must also

ensure that such cover page or other means will print out in the event that the person accessing the material chooses to print out the material or any part thereof.

G) any permission granted shall expire at the end of the class and, absent some other form of authorization, User is thereupon required to delete the applicable material from any electronic storage or to block electronic access to the applicable material.

iv) Uses of separate portions of a Work, even if they are to be included in the same course material or the same university or college class, require separate permissions under the electronic course content pay-per-use Service. Unless otherwise provided in the Order Confirmation, any grant of rights to User is limited to use completed no later than the end of the academic term (or analogous period) as to which any particular permission is granted.

v) Books and Records; Right to Audit. As to each permission granted under the electronic course content Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any electronic copies used by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this license for any reason.

c) ***Pay-Per-Use Permissions for Certain Reproductions (Academic photocopies for library reserves and interlibrary loan reporting) (Non-academic internal/external business uses and commercial document delivery).*** The License expressly excludes the uses listed in Section (c)(i)-(v) below (which must be subject to separate license from the applicable Rightsholder) for: academic photocopies for library reserves and interlibrary loan reporting; and non-academic internal/external business uses and commercial document delivery.

i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;

ii) the input of Works or reproductions thereof into any computerized database;

iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;

iv) reproduction for resale to anyone other than a specific customer of User;

v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet).*** For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.

ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified without additional express permission, and a Work consisting of audiovisual content may, if necessary, be "clipped" or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or "clipping" does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the

prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC's privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

Last updated October 2022