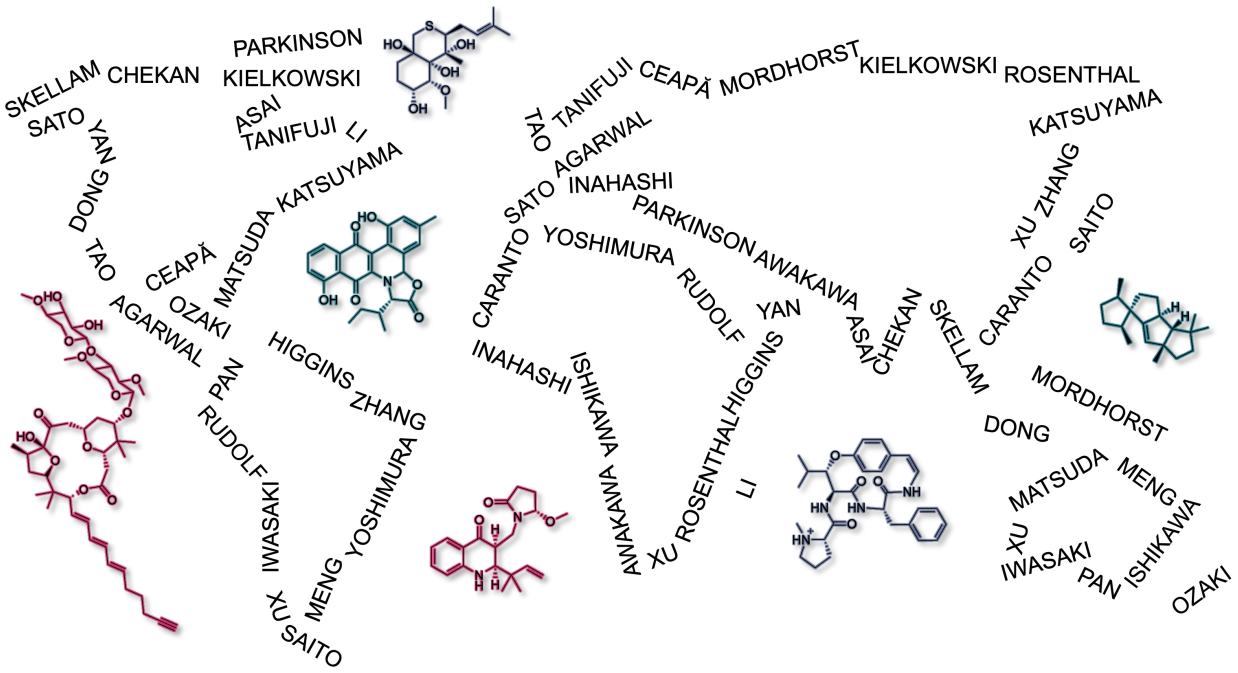
BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Young investigators in natural products chemistry, biosynthesis, and enzymology

Edited by Jeffrey D. Rudolf, Lena Barra and Takayoshi Awakawa




Generated on 01 August 2026, 04:31

# Imprint

Beilstein Journal of Organic Chemistry
www.bjoc.org
ISSN 1860-5397
Email: journals-support@beilstein-institut.de

The *Beilstein Journal of Organic Chemistry* is published by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften.

Beilstein-Institut zur Förderung der Chemischen Wissenschaften
Trakehner Straße 7–9
60487 Frankfurt am Main
Germany
www.beilstein-institut.de

The copyright to this document as a whole, which is published in the *Beilstein Journal of Organic Chemistry*, is held by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften. The copyright to the individual articles in this document is held by the respective authors, subject to a Creative Commons Attribution license.

The cover image, copyright 2024 Jeffrey D. Rudolf, Lena Barra and Takayoshi Awakawa, is licensed under the Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0). The reuse, redistribution or reproduction requires that the author, source and license are credited.The cover image highlights the invited young investigators who contributed to the success of this thematic issue.

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Young investigators in natural products chemistry, biosynthesis, and enzymology

Jeffrey D. Rudolf[*1], Lena Barra[*2] and Takayoshi Awakawa[*3]

Editorial

Open Access

Address:
[1]Department of Chemistry, University of Florida, PO Box 117200, Gainesville, FL 32611, USA, [2]Fachbereich Chemie, Universität Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany and [3]RIKEN Center for Sustainable Resource Science, Wako, Saitama 351-0198, Japan

Email:
Jeffrey D. Rudolf[*] - jrudolf@chem.ufl.edu;
Lena Barra[*] - lena.barra@uni-konstanz.de;
Takayoshi Awakawa[*] - takayoshi.awakawa@riken.jp

* Corresponding author

*Beilstein J. Org. Chem.* **2024,** *20,* 2720–2721.
https://doi.org/10.3762/bjoc.20.229

Received: 16 September 2024
Accepted: 15 October 2024
Published: 29 October 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Associate Editor: J. D. Rudolf; Guest Editors: L. Barra and T. Awakawa

© 2024 Rudolf et al.; licensee Beilstein-Institut.
License and terms: see end of document.

There is just something about natural products. When staring at these structurally complex molecules produced by diverse organisms on our planet, one may wonder why a particular natural product is made, how it is made, how I can make it, or how I can use it. The isolation and structural elucidation of a single natural product can inspire a wide array of scientific disciplines. Natural products, first as bioactive metabolites in traditional medicines and now as isolated or (bio)synthesized metabolites, are one of the most important sources of therapeutics historically and today. Their privileged structures have led organic and bioorganic chemists to develop methods to construct them. Our fundamental knowledge in enzymology is continually expanded by enzymes involved in natural products biosynthesis, as their production requires evolved enzymes to perform chemical reactions distinct from those seen in primary metabolism.

Natural products have inspired generations of scientists and continue to do so today. This thematic issue highlights young investigators in natural products chemistry, biosynthesis, and enzymology. As early-career researchers ourselves, we are honored to present this collection of 19 Full Research Papers, three Letters, and nine Review articles from our colleagues around the world. Topics covered include natural product discovery efforts from a variety of organisms, genome and transcriptome mining and bioinformatics of natural products, biosynthetic gene clusters, and enzymes, development of chemical probes, biocatalysis and chemoenzymatic total synthesis, enzymatic mechanisms, and computational investigations of chemical structures and reactions. All of the major classes of natural products are represented here: nonribosomal peptides, ribosomally-synthesized and posttranslationally modified peptides, polyketides, alkaloids, aromatics, and terpenoids. This generation of young investigators will continue to shape the field of natural products for years to come.

We pass on our greatest thanks to all the contributors to this special thematic issue. We also humbly thank Prof. Jeroen Dickschat for inviting each of us to guest-edit this wonderful issue featuring rising stars in the field of natural products. Finally, we thank the staff of the *Beilstein Journal of Organic Chemistry* for making this issue possible. As you browse this

issue, we hope that you are inspired, not only by the science but by the young researchers contributing to this field. Keep an eye on these early-career researchers; the future of natural products research is in very capable hands!

Jeff Rudolf, Lena Barra, and Takayoshi Awakawa

Gainesville, Konstanz, Wako, September 2024

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed.

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.229

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Unraveling the role of prenyl side-chain interactions in stabilizing the secondary carbocation in the biosynthesis of variexenol B

Moe Nakano[1], Rintaro Gemma[1] and Hajime Sato[*1,2]

Letter

Open Access

Address:
[1]Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 4-4-37 Takeda, Kofu, Yamanashi 400-8510, Japan and [2]PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332–0012, Japan

Email:
Hajime Sato[*] - hsato@yamanashi.ac.jp

* Corresponding author

Keywords:
biosynthesis; carbocation; cation–π interaction; DFT; terpene

*Beilstein J. Org. Chem.* **2023,** *19,* 1503–1510.
https://doi.org/10.3762/bjoc.19.107

Received: 30 June 2023
Accepted: 07 September 2023
Published: 28 September 2023

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf



© 2023 Nakano et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Terpene cyclization reactions involve a number of carbocation intermediates. In some cases, these carbocations are stabilized by through-space interactions with π orbitals. Several terpene/terpenoids, such as sativene, santalene, bergamotene, ophiobolin and mangicol, possess prenyl side chains that do not participate in the cyclization reaction. The role of these prenyl side chains has been partially investigated, but remains elusive in the cyclization cascade. In this study, we focus on variexenol B that is synthesized from iso-GGPP, as recently reported by Dickschat and co-workers, and investigate the possibility of through-space interactions with prenyl side chains using DFT calculations. Our calculations show that (i) the unstable secondary carbocation is stabilized by the cation–π interaction from prenyl side chains, thereby lowering the activation energy, (ii) the four-membered ring formation is completed through bridging from the exomethylene group, and (iii) the annulation from the exomethylene group proceeds in a barrier-free manner.

## Introduction

Terpene/terpenoids are most abundant natural products in nature, more than 180,000 terpenoid compounds have been reported to date [1-4]. One of the most intriguing point is that all diversified structures are synthesized from common starting materials, isoprenoids. Reactions that generate complex cyclic structures and multiple stereocenters from linear achiral precursors offer many valuable insights from a fundamental organic chemistry perspective.

The terpene cyclization cascade generally involves a multistep domino-type reaction. Therefore, it is challenging to reveal the detailed reaction mechanism solely by an experimental method.

To address this issue, computational chemistry including DFT [5-9], QM/MM [10-16] and QM/MM MD [14-17] calculations have been used for the biosynthetic studies of terpene/terpenoids [18].

Terpene-forming reactions, which involve various types of carbocation species stabilized by hyperconjugative interactions, through-space interactions, and C–H–π interactions, have been intensively investigated by Tantillo and co-workers, who have contributed greatly to revealing the intriguing nature of carbocations [7,19,20].

We have also elucidated various new insights of carbocation chemistry, such as the C–H–π interaction between the carbocation intermediate and the Phe residue of terpene cyclase in the biosynthesis of sesterfisherol [21], and the intricated rearrangement reaction mechanism promoted by the equilibrium state of the homoallyl cation and the cyclopropylcarbinyl cation in the biosynthesis of trichobrasilenol [22], by combined methods of computational and experimental chemistry.

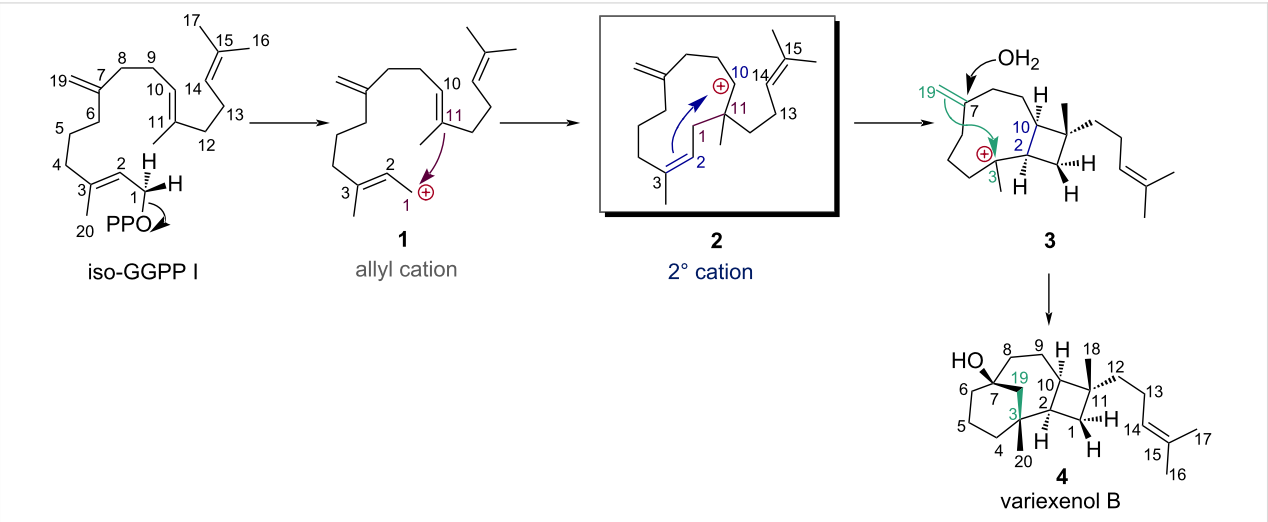
Recently, Dickschat et al. reported the synthesis of a novel diterpene compound, variexenol B, using a substrate analogue called iso-GGPP (Scheme 1) [23]. This biosynthetic pathway has two interesting aspects. First, this cyclization cascade involves a prenyl side chain that do not participate in the cyclization cascade. This type of terpene compounds has already been reported, such as santalene, bergamotene, mangicol, etc. The idea that the reaction mechanism changes due to differences in the prenyl side chains has been studied by Tantillo and co-workers [24-26]. They reported that the carbocation intermediates traversed in pinene/camphene and ylangene/sativene biosynthesis change depending on the presence or absence of prenyl side chains. In their study, it was argued that the extent of hyperconjugation determines whether the reaction proceeds in a stepwise or concerted manner.

The second interesting aspect of the biosynthesis of variexenol B is that the biosynthetic pathway involves an intermediate with an exomethylene group. A terpene with an exomethylene group as a starting material is rare. Several terpene cyclizations with an exomethylene group are known, such as with caryolene and crotinsulidane diterpenoids, and the reaction mechanisms have been analyzed [27-30]. It would be interesting to see how the exomethylene group reacts in the cyclization of variexenol B. In this study, we investigated the biosynthetic pathways using DFT calculations to validate the above-mentioned aspects.

## Results and Discussion

The detailed structures of the intermediates and transition states were elucidated by computational analysis. Interestingly, we have found an interaction between the secondary carbocation and the prenyl side chain. Figure 1 shows the computed biosynthetic pathway and energy diagram without cation–π interaction, while Figure 2 shows the computed biosynthetic pathway including cation–π interaction from the prenyl side chain.

Our research started with the application of DFT calculations to the putative biosynthetic pathway of variexenol B (Figure 1). It was revealed that the variexenol B biosynthetic pathway undergoes a two-step reaction process. Contrary to the putative biosynthetic pathway, the formation of the C1–C11 and C2–C10 bonds was found to be concerted, due to the formation of a secondary carbocation at the C10 position. Then, the tertiary carbocation formed at the C3 position undergoes virtu-




**Scheme 1:** Proposed biosynthetic pathway for variexenol B.

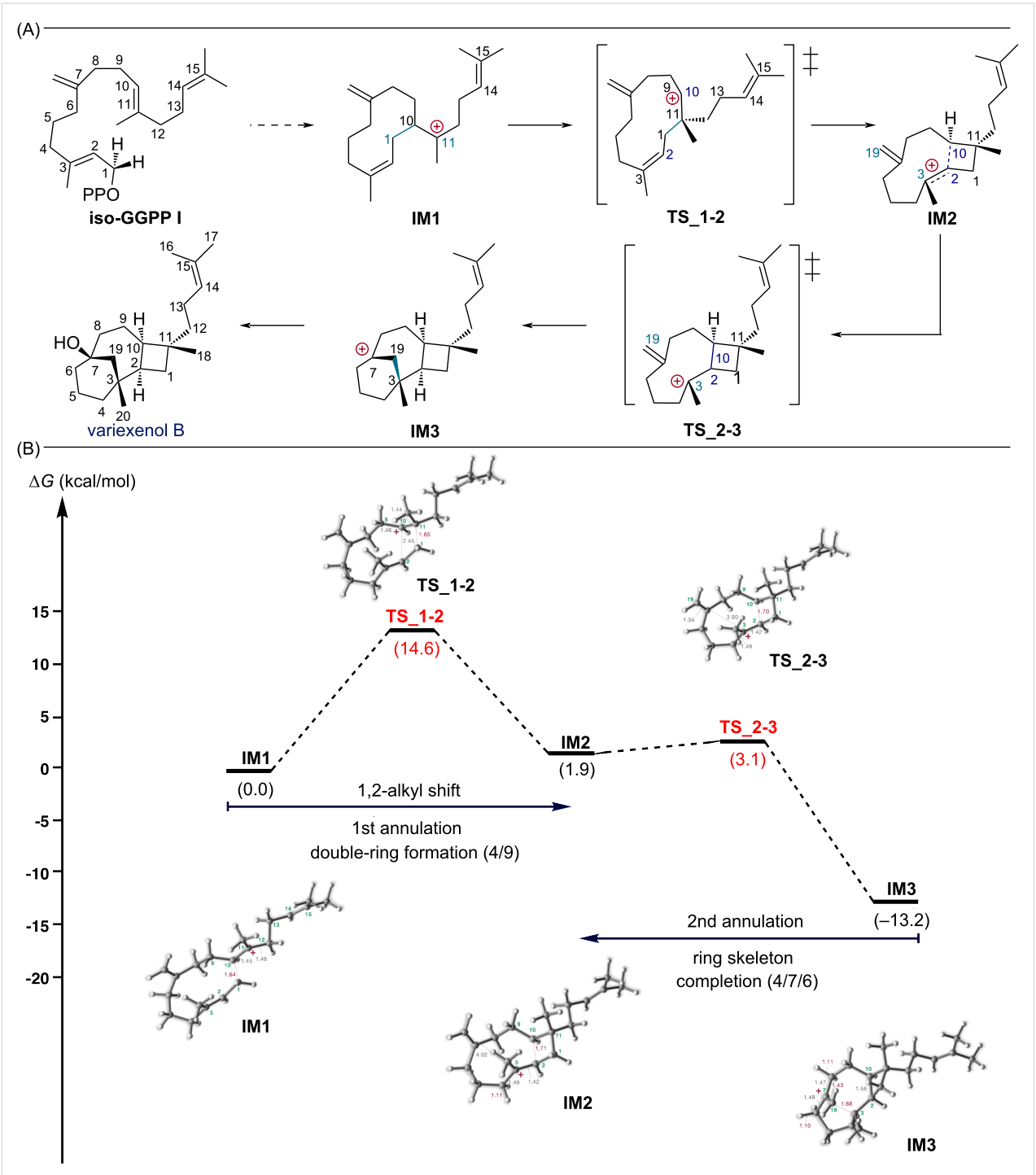



**Figure 1:** (A) Results of DFT evaluation of the whole pathway of variexenol B without cation–π interaction. (B) Energy diagram of variexenol B without cation–π interaction. IM means intermediate and TS transition state. Potential energies (kcal/mol, Gibbs free energies calculated at the mPW1PW91/6-31+G(d,p)//M06-2X/6-31+G(d,p) level) relative to **IM1** are shown in parentheses.

ally barrier-free cyclization from the exomethylene group to yield **IM3**.

We next investigated the effect of the prenyl side chain in the biosynthesis of variexenol B. Although several terpene compounds with prenyl side chains have been reported, it remains unclear whether these prenyl side chains are located inside or outside the active site during the cyclization process. Therefore, we searched for conformations in which the side chain is closer to the carbocation center and performed calculations.

It was found that the structure with the prenyl side chain containing the C14=C15 double bond positioned inwards was more advantageous than the pathway shown in Figure 1. Calculations based on the specified structure are shown in Figure 2. In this pathway, the C14=C15 double bond interacts with the secondary carbocation at C10, reducing the activation energy of the

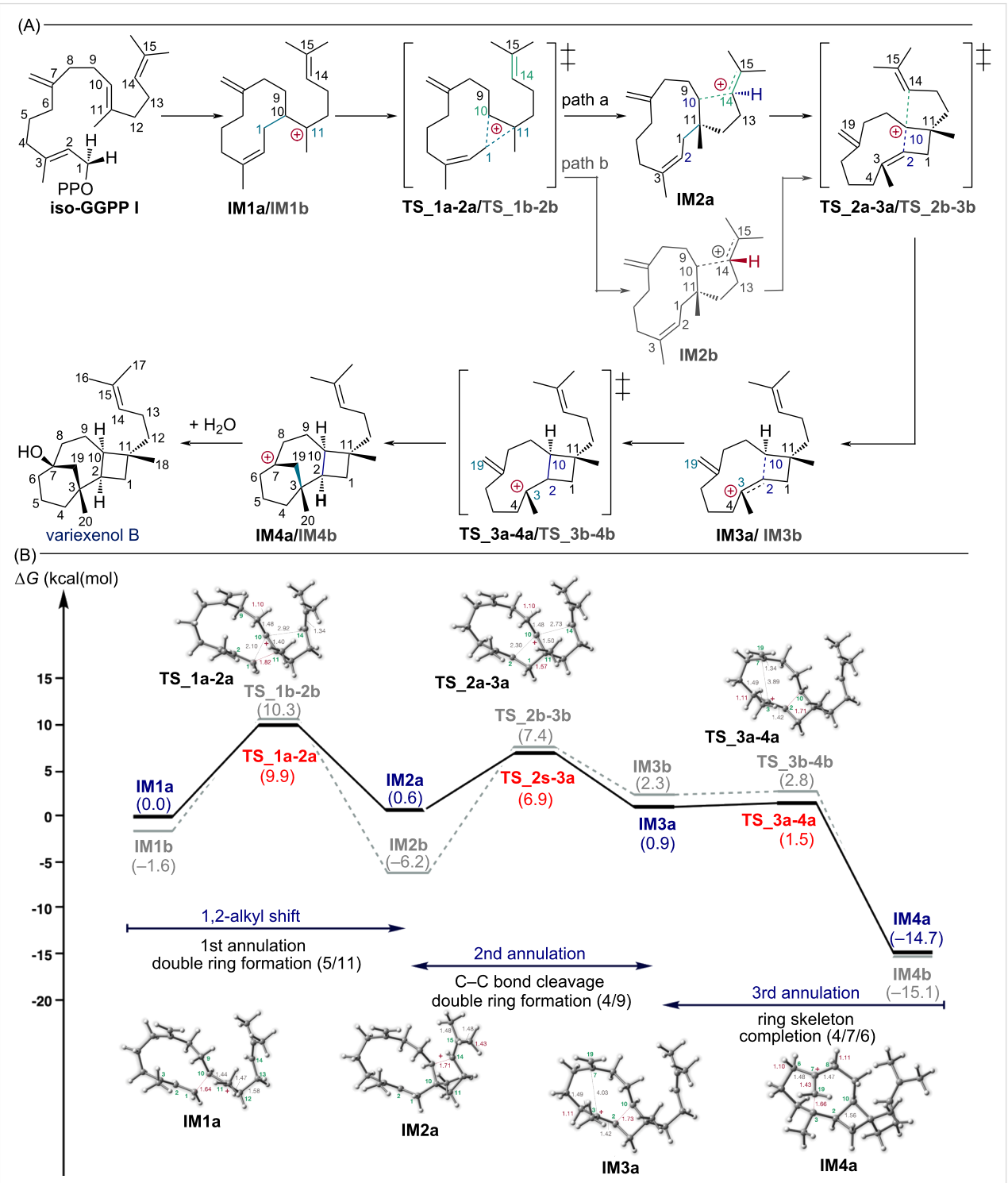



**Figure 2:** (A) Results of the DFT evaluation of the whole pathway of variexenol B including cation–π interaction from the prenyl side chain. Path a has an α-hydrogen at the C14 position in **IM2**, while path b has the opposite orientation. (B) Energy diagram of variexenol B with consideration of cation–π interaction. Potential energies (kcal/mol, Gibbs free energies calculated at the mPW1PW91/6-31+G(d,p)//M06-2X/6-31+G(d,p) level) relative to **IM1** are shown in parentheses.

first step by approximately 4.7 kcal/mol. Moreover, due to the stabilization of the secondary carbocation-like intermediate **IM2**, the reaction proceeds stepwise rather than concertedly [7]. It was found that the final cyclization reaction from the exomethylene group proceeds without a barrier, similar to the previous pathway.

Regarding the orientation of the prenyl side chain, two pathways can be considered depending on whether the hydrogen at C14 is pointing; α-hydrogen (path a) or β-hydrogen (path b). Both pathways follow similar reaction mechanisms, however, when comparing path a and path b, the most striking energy difference is in the step from **IM2a/b** to **IM3a/b** (Figure 2B). The energy barrier of this step is 6.3 kcal/mol for path a, whereas 13.6 kcal/mol for path b, with a difference of 7.3 kcal/mol. Although the stabilization of the intermediate **IM2b** is greater in path b, the activation energy suggests that path a is more favorable.

Generally, the activation energies for terpene cyclization reactions are often below 10 kcal/mol. However, in the case of complex rearrangement reactions involving secondary carbocations, which we recently discovered, reactions with activation energies around 16 kcal/mol have been reported [22]. In the pathway shown in Figure 1, the highest energy barrier was 14.6 kcal/mol. Conversely, in Figure 2, path a had an energy barrier of 9.9 kcal/mol and path b 13.6 kcal/mol. From these results, it can be concluded that although all three pathways have the potential to advance the reaction, the most energetically favorable pathway is path a, as shown in this study.

To the best of our knowledge, the interaction from the prenyl side chain towards the carbocation center have not been reported. Systems with secondary carbocations on rings bearing prenyl side chains are commonly observed in steroid biosynthesis. These type of cyclization reactions have been vigorously studied by Hess [31-36] and Wu [37,38]. In these systems, the secondary carbocation and the double bond of the neighboring prenyl side chain interacts and promptly induce C–C bond formation. There have been no reports published where, as in our case, the cation is stabilized without bond formation. We have also considered other transannular cation–π interactions in this system. In this case, the interaction between the secondary carbocation at C10 and the C2=C3 double bond or the exomethylene group at C7 should be considered. However, moving it closer to the C2=C3 bond would result in **IM2** as shown in Figure 1 and a C–C bond would be formed. The exomethylene group at C7 is also very reactive, so if it gets close, it would easily form a C–C bond. Therefore, we believe that no other transannular cation–π interactions need to be considered in this system.

In systems without cation–π interactions, such as in the biosynthesis of variediene [39] and spiroviolene [40], bonds around the secondary carbocation are strongly influenced by hyperconjugation. In particular, C–C bonds containing a secondary cation are shortened to about 1.45 Å, showing a slight double bond character. On the other hand, in intermediates such as **IM2a** and **IM2b**, which have cation–π interactions, the surrounding bonds are hardly affected by hyperconjugation (C9–C10: 1.54 Å, C10–C11: 1.54 Å, C11–C1: 1.57 Å). We have also done a comparative analysis of the charge distribution in scenarios with and without cation–π interactions. In cases where the interaction is absent, the cationic character at C10 is pronounced. Conversely, in the presence of the cation–π interaction, the cation is delocalized, resulting in a decrease in cationicity at C10 and a corresponding increase in cationicity at C15.

Note that the interconversion of **TS_2a-3a** to **TS_2b-3b** requires a significant conformational change, such as a 180 degree rotation of the iPr group. However, such a large conformational change is unlikely to occur within the enzyme. Therefore, the Curtin–Hammett principle is not applicable to this system.

To investigate the details of carbocations and hyperconjugations in the variexenol B biosynthetic pathway, we carried out a bond length change analysis on the bonds that contribute most to the reaction from **IM1a** to **IM4a** (Figure 3A).

In the process from **IM1a** to **TS_1a–2a**, the C1–C10 bond ruptures as C1 shifts towards C11. Subsequently, in the transition towards **IM2a**, a complete formation of the C1–C11 bond occurs. At this point, the vacant orbital of the carbocation at C10 interacts with the π orbital of the C14=C15 double bond.

The distance between C10 and C14 is 1.71 Å, which is hardly to recognize as a single C–C bond, since the distance is greatly elongated. Moreover, the bond length of C14=15 is 1.43 Å, which is close to the double bond length. Judging from the bond length alone, it is not impossible to conclude that the C10–C14 bond is formed, but considering the rational mechanism of organic reactions, bond cleavage does not occur immediately after the bond is formed.

On the other hand, the C10–C14 bond length of **IM2b** is 1.64 Å, which is the bond length when hyperconjugated and is commonly observed in terpene-forming reactions. This relatively short bond length appears to contribute to the stability of **IM2b**. The energy difference between **IM2a** and **IM2b** appears to be due to small conformational differences caused by the stereochemistry of H14.

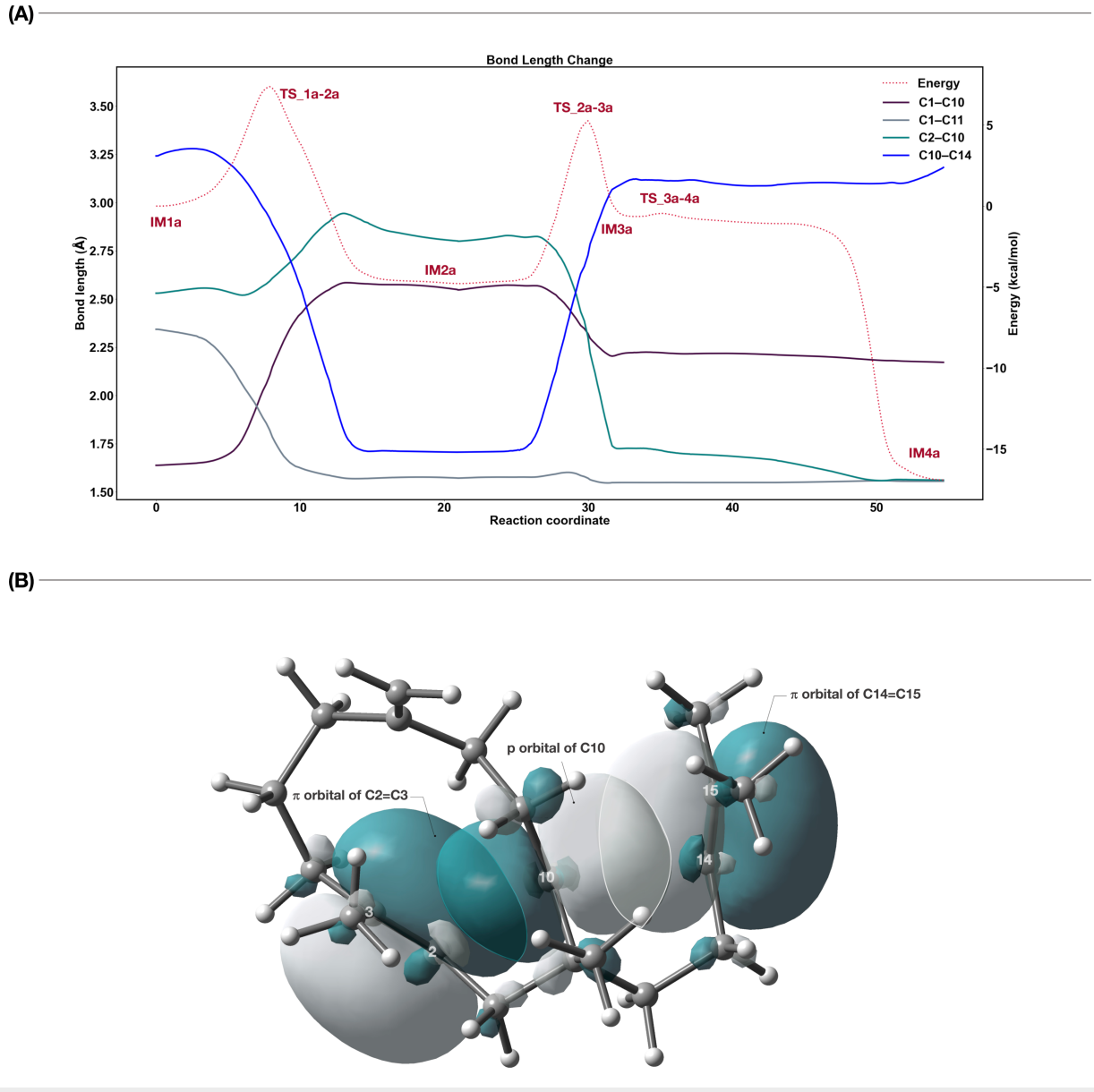



**Figure 3:** (A) A representative example of the evolution of key bond lengths in the conversion of path a. (B) Key representative orbitals of **TS_2a–3a** computed by DFT calculations.

In **TS_2a–3a**, the C10 secondary carbocation is stabilized and sandwiched between the two π orbitals of C2=C3 and C14=C15. The status of this orbital interaction is depicted in Figure 3B. This interaction forms the C2–C10 bond and the reaction proceeds to **IM3a**.

Regarding the 4-membered ring formation, the C2–C10 bond in **IM3a** is 1.73 Å long, which is hard to recognize as a single bond. However, it is well consistent with the previously reported hyperconjugation in 4-membered ring formation [7]. Then, the C2–C10 bond became 1.56 Å and the 4-membered ring bond is completed (Figure 3B) when the hyperconjugation effect is eliminated by the removal of the C3 carbocation through annulation from the exomethylene group. Based on the key bond analysis, we have successfully elucidated the details of variexenol B biosynthesis. Note that there is an interaction between the empty p orbital of C10 and the π orbital of C14=C15 in Figure 3B, although the bond is not shown. The presence or absence of a bond in the GaussView depends only on the distance between the atoms and may differ from the actual bonding. If a bond is stretched due to hyperconjugation etc., it often happens that the bond is not displayed correctly. Therefore, we performed bond length analysis and NBO analysis to understand the state of the carbocation and bonding.

## Conclusion

In conclusion, we have investigated the detailed reaction mechanism of the biosynthesis of variexenol. We have revealed three new insights: (i) the possibility of stabilization of the secondary carbocation by the prenyl side chain of the intermediate, (ii) the four-membered ring formation is completed by the bridging reaction, and (iii) the annulation from the exomethylene group is a barrier-free process.

To date, when constructing the computational model, we have sometimes truncated the prenyl side chains that do not participate in the cyclization cascade in order to reduce the computational cost [35,41]. However, as demonstrated in this study, the possibility of cation–π interactions lowering the activation energy of annulation requires caution when constructing computational models in the future. Furthermore, future research is expected to determine whether there is space in the enzyme active site for these prenyl side chains to fold and approach the reaction center, as seen in X-ray crystallographic analysis.

## Experimental

All calculations were carried out using the Gaussian 16 package [42]. Structure optimizations were done with the M06-2X [43] density functional theory method and the 6-31+G(d,p) basis set without any symmetry restrictions. M06-2X was selected because of its accuracy in calculating terpene-forming reactions and its proven track record being used in previous studies of reaction mechanism analysis [16,44]. Vibrational frequency calculations at the same level of theory with optimization were performed to verify that each local minimum has no imaginary frequency and that each **TS** has only a single imaginary frequency. Conformational search was done with conflex program [45-47]. Intrinsic reaction coordinate (IRC) calculations [48-51] for all **TS**s were performed with GRRM11 [52] based on Gaussian 16. Single-point energies were calculated at the mPW1PW91/6-31+G(d,p) level based on the optimized structure by using the M06-2X method. The utility of relative Gibbs free energies ($G_{rel}$) based on single-point energy at the mPW1PW91 level has been previously validated for a wide variety of terpene-forming reactions [22,41,53].

## Supporting Information

### Supporting Information File 1

IRC plot, 3D representations of all computed structures, cartesian coordinates, energies, and imaginary frequencies.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-19-107-S1.pdf]

## Acknowledgements

We would like to express our deepest gratitude to Dr. Makoto Obata for his insightful feedback and continuous mentorship that significantly contributed to the success of this research.

## Funding

This work was supported by a JSPS KAKENHI Grant-in-Aid for Early-Career Scientists (No. 22K14791(H.S.)), a JSPS KAKENHI Grant-in-Aid for Transformative Research Areas (No. 22H05125 (H.S.)), a MEXT grant for Leading Initiative for Excellent Young Researchers (No. JPMXS0320200422 (H.S.)), JST PRESTO (No. JPMJPR21D5 (H.S.)), the Uehara Memorial Foundation (No. 202110117 (H.S.)), the Terumo Life Science Foundation (No. 21-III4030 (H.S.)), Inamori foundation, and Astellas Foundation for Research on Metabolic Disorders (H.S.).

## ORCID® iDs

Moe Nakano - https://orcid.org/0009-0002-0245-0075
Rintaro Gemma - https://orcid.org/0009-0006-9662-6607
Hajime Sato - https://orcid.org/0000-0001-5185-096X

## Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2023.27.v1

## References

1. Rudolf, J. D.; Alsup, T. A.; Xu, B.; Li, Z. *Nat. Prod. Rep.* **2021,** *38,* 905–980. doi:10.1039/d0np00066c
2. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
3. Zeng, T.; Chen, Y.; Jian, Y.; Zhang, F.; Wu, R. *New Phytol.* **2022,** *235,* 662–673. doi:10.1111/nph.18133
4. Dickschat, J. S. *Nat. Prod. Rep.* **2016,** *33,* 87–110. doi:10.1039/c5np00102a
5. Tantillo, D. J. *Nat. Prod. Rep.* **2011,** *28,* 1035–1053. doi:10.1039/c1np00006c
6. Tantillo, D. J. *Nat. Prod. Rep.* **2013,** *30,* 1079–1086. doi:10.1039/c3np70028c
7. Tantillo, D. J. *Chem. Soc. Rev.* **2010,** *39,* 2847–2854. doi:10.1039/b917107j
8. Tantillo, D. J. *Angew. Chem., Int. Ed.* **2017,** *56,* 10040–10045. doi:10.1002/anie.201702363
9. Hong, Y. J.; Tantillo, D. J. *Chem. Soc. Rev.* **2014,** *43,* 5042–5050. doi:10.1039/c3cs60452g
10. Weitman, M.; Major, D. T. *J. Am. Chem. Soc.* **2010,** *132,* 6349–6360. doi:10.1021/ja910134x
11. Major, D. T.; Weitman, M. *J. Am. Chem. Soc.* **2012,** *134,* 19454–19462. doi:10.1021/ja308295p
12. Raz, K.; Levi, S.; Gupta, P. K.; Major, D. T. *Curr. Opin. Biotechnol.* **2020,** *65,* 248–258. doi:10.1016/j.copbio.2020.06.002
13. Das, S.; Shimshi, M.; Raz, K.; Nitoker Eliaz, N.; Mhashal, A. R.; Ansbacher, T.; Major, D. T. *J. Chem. Theory Comput.* **2019,** *15,* 5116–5134. doi:10.1021/acs.jctc.9b00366

14. Ansbacher, T.; Freud, Y.; Major, D. T. *Biochemistry* **2018,** *57,* 3773–3779. doi:10.1021/acs.biochem.8b00452
15. Dixit, M.; Weitman, M.; Gao, J.; Major, D. T. *ACS Catal.* **2018,** *8,* 1371–1375. doi:10.1021/acscatal.7b02823
16. Dixit, M.; Weitman, M.; Gao, J.; Major, D. T. *ACS Catal.* **2017,** *7,* 812–818. doi:10.1021/acscatal.6b02584
17. Lou, T.; Li, A.; Xu, H.; Pan, J.; Xing, B.; Wu, R.; Dickschat, J. S.; Yang, D.; Ma, M. *J. Am. Chem. Soc.* **2023,** *145,* 8474–8485. doi:10.1021/jacs.3c00278
18. Sato, H.; Saito, K.; Yamazaki, M. *Front. Plant Sci.* **2019,** *10,* 802. doi:10.3389/fpls.2019.00802
19. Hong, Y. J.; Tantillo, D. J. *Chem. Sci.* **2013,** *4,* 2512–2518. doi:10.1039/c3sc50571e
20. Hong, Y. J.; Tantillo, D. J. *Org. Lett.* **2015,** *17,* 5388–5391. doi:10.1021/acs.orglett.5b02740
21. Sato, H.; Narita, K.; Minami, A.; Yamazaki, M.; Wang, C.; Suemune, H.; Nagano, S.; Tomita, T.; Oikawa, H.; Uchiyama, M. *Sci. Rep.* **2018,** *8,* 2473. doi:10.1038/s41598-018-20916-x
22. Sato, H.; Hashishin, T.; Kanazawa, J.; Miyamoto, K.; Uchiyama, M. *J. Am. Chem. Soc.* **2020,** *142,* 19830–19834. doi:10.1021/jacs.0c09616
23. Li, H.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2022,** *61,* e202211054. doi:10.1002/anie.202211054
24. Hong, Y. J.; Tantillo, D. J. *Chem. Commun.* **2012,** *48,* 1571–1573. doi:10.1039/c1cc14414f
25. Hong, Y. J.; Tantillo, D. J. *Org. Biomol. Chem.* **2010,** *8,* 4589–4600. doi:10.1039/c0ob00167h
26. Zi, J.; Matsuba, Y.; Hong, Y. J.; Jackson, A. J.; Tantillo, D. J.; Pichersky, E.; Peters, R. J. *J. Am. Chem. Soc.* **2014,** *136,* 16951–16953. doi:10.1021/ja508477e
27. Nguyen, Q. N. N.; Tantillo, D. J. *Beilstein J. Org. Chem.* **2013,** *9,* 323–331. doi:10.3762/bjoc.9.37
28. Ortega, D. E.; Nguyen, Q. N. N.; Tantillo, D. J.; Toro-Labbé, A. *J. Comput. Chem.* **2016,** *37,* 1068–1081. doi:10.1002/jcc.24294
29. Campos, R. B.; Tantillo, D. J. *J. Org. Chem.* **2018,** *83,* 1073–1076. doi:10.1021/acs.joc.7b02904
30. Hong, Y. J.; Tantillo, D. J. *J. Am. Chem. Soc.* **2010,** *132,* 5375–5386. doi:10.1021/ja9084786
31. Hess, B. A., Jr. *Eur. J. Org. Chem.* **2004,** 2239–2242. doi:10.1002/ejoc.200300684
32. Smentek, L.; Hess, B. A., Jr. *J. Am. Chem. Soc.* **2010,** *132,* 17111–17117. doi:10.1021/ja1039133
33. Hess, B. A., Jr. *J. Am. Chem. Soc.* **2002,** *124,* 10286–10287. doi:10.1021/ja026850r
34. Hess, B. A., Jr. *Org. Lett.* **2003,** *5,* 165–167. doi:10.1021/ol027449c
35. Hess, B. A., Jr.; Smentek, L. *Org. Lett.* **2004,** *6,* 1717–1720. doi:10.1021/ol0496125
36. Hess, B. A., Jr.; Smentek, L. *Angew. Chem., Int. Ed.* **2013,** *52,* 11029–11033. doi:10.1002/anie.201302886
37. Chen, N.; Wang, S.; Smentek, L.; Hess, B. A., Jr.; Wu, R. *Angew. Chem., Int. Ed.* **2015,** *54,* 8693–8696. doi:10.1002/anie.201501986
38. Zhang, F.; Wang, Y.; Yue, J.; Zhang, R.; Hu, Y.-e.; Huang, R.; Ji, A.-j.; Hess, B. A., Jr.; Liu, Z.; Duan, L.; Wu, R. *Sci. Adv.* **2023,** *9,* eadh1418. doi:10.1126/sciadv.adh1418
39. Sato, H.; Li, B.-X.; Takagi, T.; Wang, C.; Miyamoto, K.; Uchiyama, M. *JACS Au* **2021,** *1,* 1231–1239. doi:10.1021/jacsau.1c00178
40. Hong, Y. J.; Tantillo, D. J. *Aust. J. Chem.* **2017,** *70,* 362–366. doi:10.1071/ch16504
41. Sato, H.; Takagi, T.; Miyamoto, K.; Uchiyama, M. *Chem. Pharm. Bull.* **2021,** *69,* 1034–1038. doi:10.1248/cpb.c21-00536
42. *Gaussian 16,* Revision C.01; Gaussian, Inc.: Wallingford, CT, 2016.
43. Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008,** *120,* 215–241. doi:10.1007/s00214-007-0310-x
44. Zev, S.; Gupta, P. K.; Pahima, E.; Major, D. T. *J. Chem. Theory Comput.* **2022,** *18,* 167–178. doi:10.1021/acs.jctc.1c00746
45. Goto, H.; Osawa, E. *J. Am. Chem. Soc.* **1989,** *111,* 8950–8951. doi:10.1021/ja00206a046
46. Gotō, H.; Ōsawa, E. *J. Chem. Soc., Perkin Trans. 2* **1993,** 187–198. doi:10.1039/p29930000187
47. *CONFLEX 9;* CONFLEX Corporation: Tokyo, Japan, 2021.
48. Fukui, K. *Acc. Chem. Res.* **1981,** *14,* 363–368. doi:10.1021/ar00072a001
49. Page, M.; Doubleday, C.; McIver, J. W., Jr. *J. Chem. Phys.* **1990,** *93,* 5634–5642. doi:10.1063/1.459634
50. Ishida, K.; Morokuma, K.; Komornicki, A. *J. Chem. Phys.* **1977,** *66,* 2153–2156. doi:10.1063/1.434152
51. Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990,** *94,* 5523–5527. doi:10.1021/j100377a021
52. Maeda, S.; Ohno, K.; Morokuma, K. *Phys. Chem. Chem. Phys.* **2013,** *15,* 3683–3701. doi:10.1039/c3cp44063j
53. Matsuyama, T.; Togashi, K.; Nakano, M.; Sato, H.; Uchiyama, M. *JACS Au* **2023,** *3,* 1596–1603. doi:10.1021/jacsau.3c00039

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.19.107

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Identification of the *p*-coumaric acid biosynthetic gene cluster in *Kutzneria albida*: insights into the diazotization-dependent deamination pathway

Seiji Kawai[1], Akito Yamada[1], Yohei Katsuyama[*1,2] and Yasuo Ohnishi[1,2]

Full Research Paper

Open Access

Address:
[1]Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan and [2]Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan

Email:
Yohei Katsuyama[*] - aykatsuhko@g.ecc.u-tokyo.ac.jp

* Corresponding author

Keywords:
actinomycetes; avenalumic acid; biosynthesis; *p*-coumaric acid; polyketides

*Beilstein J. Org. Chem.* **2024,** *20,* 1–11.
https://doi.org/10.3762/bjoc.20.1

Received: 06 October 2023
Accepted: 13 December 2023
Published: 02 January 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: T. Awakawa

© 2024 Kawai et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Recently, we identified the biosynthetic gene cluster of avenalumic acid (*ava* cluster) and revealed its entire biosynthetic pathway, resulting in the discovery of a diazotization-dependent deamination pathway. Genome database analysis revealed the presence of more than 100 *ava* cluster-related biosynthetic gene clusters (BGCs) in actinomycetes; however, their functions remained unclear. In this study, we focused on an *ava* cluster-related BGC in *Kutzneria albida* (*cma* cluster), and revealed that it is responsible for *p*-coumaric acid biosynthesis by heterologous expression of the *cma* cluster and in vitro enzyme assays using recombinant Cma proteins. The ATP-dependent diazotase CmaA6 catalyzed the diazotization of both 3-aminocoumaric acid and 3-aminoavenalumic acid using nitrous acid in vitro. In addition, the high efficiency of the CmaA6 reaction enabled us to perform a kinetic analysis of AvaA7, which confirmed that AvaA7 catalyzes the denitrification of 3-diazoavenalumic acid in avenalumic acid biosynthesis. This study deepened our understanding of the highly reducing type II polyketide synthase system as well as the diazotization-dependent deamination pathway for the production of avenalumic acid or *p*-coumaric acid.

## Introduction

The genomes of microorganisms possess diverse biosynthetic gene clusters (BGCs) to produce natural products [1]. In particular, actinomycetes, a major source of bioactive compounds, have numerous BGCs in their genomes, but the majority of them are thought to be silent under laboratory conditions [2]. Because these orphan BGCs are often related to the production of unknown natural products, genome mining has been vigorously used to search for novel compounds in recent years due to the rapid improvement of DNA sequence technologies and computational approaches to analyze BGCs [1,3].

Our research group previously identified the secondary metabolite-specific nitrous acid biosynthetic pathway, named ANS (aspartate-nitrosuccinate), from the study on cremeomycin biosynthesis [4,5]. The ANS pathway is composed of two enzymes, CreE (FAD-dependent monooxygenase) and CreD (lyase), to synthesize nitrous acid from L-aspartate and the nitrous acid is used to synthesize the diazo group of cremeomycin [4]. After the discovery of the ANS pathway, it has been shown that the ANS pathway is involved in the nitrogen–nitrogen (N–N) bond formation in the biosynthesis of several natural products [6-8]. Enzymes that catalyze N–N bond formation by using nitrous acid from the ANS pathway have also been characterized in several recent studies [9-14]. Most of them belong to the adenylate-forming enzyme superfamily (ANL superfamily) and utilize ATP to activate nitrous acid by AMPylation, with the only exception being AzpL in alazopeptin biosynthesis, which is a membrane protein that catalyzes diazotization, presumably in an ATP-independent manner [9-14]. Moreover, the genome database analysis indicated that there are many orphan BGCs containing genes encoding the ANS pathway, which implies that the biosynthesis of many unknown natural products requires nitrous acid derived from the ANS pathway [4].

To further understand the role of the ANS pathway in secondary metabolism, we recently identified the BGC for avenalumic acid (*ava* cluster, see the lower right corner of Figure 1B for its structure) by genome mining targeting the ANS pathway in *Streptomyces* sp. RI-77, and revealed its entire biosynthetic pathway [13]. In this pathway, 3-amino-4-hydroxybenzoic acid (3,4-AHBA, **1**), synthesized by AvaH and AvaI, is loaded onto AvaA3 (carrier protein) by AvaA1 (AMP-dependent ligase), resulting in 3,4-AHBA-AvaA3. A highly reducing type II polyketide synthase (PKS) system [15,16] (AvaA2, A4, A5, and A8) then forms a diene moiety using two malonyl units to synthesize 3-aminoavenalumic acid (3-AAA, **7**) from 3,4-AHBA-AvaA3. The amino group of 3-AAA is diazotized by AvaA6 using nitrous acid derived from the ANS pathway (AvaE and AvaD) to form 3-diazoavenalumic acid (3-DAA, **8**), and this diazo group is finally substituted for a hydride derived from NADPH by AvaA7. Interestingly, there are more than 100 BGCs that possess a set of *ava* gene homologs in the genome database, suggesting that these BGCs are responsible for the biosynthesis of avenalumic acid or its derivatives [13].

In this study, we focused on an *ava* cluster-related BGC in *Kutzneria albida* JCM 3240. We showed that this BGC is involved in *p*-coumaric acid biosynthesis by heterologous expression in *Streptomyces albus* J1074 and several in vitro biochemical experiments using recombinant proteins. CmaA6 was shown to catalyze the diazotization of 3-aminocoumaric acid (3-ACA, **3**) and 3-AAA (**7**) with considerably higher efficiency than AvaA6 in avenalumic acid biosynthesis. We also performed kinetic analysis of AvaA7, which catalyzes the denitrification of 3-DAA (**8**) in avenalumic acid biosynthesis, using the highly efficient diazotase CmaA6, for the first time. These results provided new insights into highly reducing type II PKSs and microbial production of *p*-coumaric acid using the ANS pathway, and strengthened our previous proposal for the biosynthesis of avenalumic acid.

# Results

## Bioinformatic analysis of *ava* and *cma* clusters

In our previous study, we found more than 100 BGCs that have a similar gene component to that of the *ava* cluster in the genome database [13]. In this study, to obtain and analyze *ava* cluster-related BGCs, we again searched the genome database for gene clusters that contain (i) genes responsible for 3,4-AHBA synthesis (AvaH and AvaI homologs), (ii) genes encoding a KS-CLF complex (AvaA4 and AvaA5 homologs), and (iii) a gene encoding a diazotase (AvaA6 homolog). As a result, we discovered 134 BGCs and analyzed them using antiSMASH [17] and BiG-SCAPE (Figures S1 and S2 in Supporting Information File 1) [18]. Interestingly, approximately half of them (72 BGCs) do not have *avaE* and *avaD* homologs for the ANS pathway, suggesting three possibilities: (i) *avaE* and *avaD* homologs are encoded apart from the cluster, (ii) the strains have an alternative route to synthesize nitrous acid (as reported for alanosine biosynthesis [19]), and (iii) the strains use exogenous nitrous acid [13]. The BiG-SCAPE analysis divided the BGCs into seven groups (Figures S1 and S2 in Supporting Information File 1) and the *ava* cluster belongs to FAM_00127 (Figure S1). Interestingly, most of the BGCs in FAM_00091, FAM_00111, FAM_00133, and FAM_00125 (approximately 40 BGCs) do not have *avaA8* (FabG-like ketoreductase) and *avaC* (major facilitator superfamily transporter) homologs, but have a *cmaG*-like gene encoding an FMN-dependent oxidoreductase and a *cmaR*-like gene encoding a LysR family transcriptional regulator (Figure 1, Figures S1 and S2 in Supporting Information File 1, and Table 1). We anticipated that they would be a subtype of *ava*-like BGCs and might produce compounds other than avenalumic acid. Among these clusters, we focused on a cluster present in the genome of a rare actinomycete, *Kutzneria albida*, and named it as *cma* cluster because the *cma* cluster was later shown to be responsible for the biosynthesis of *p*-coumaric acid (Figure 1 and Table 1, NCBI accession number CP007155.1).

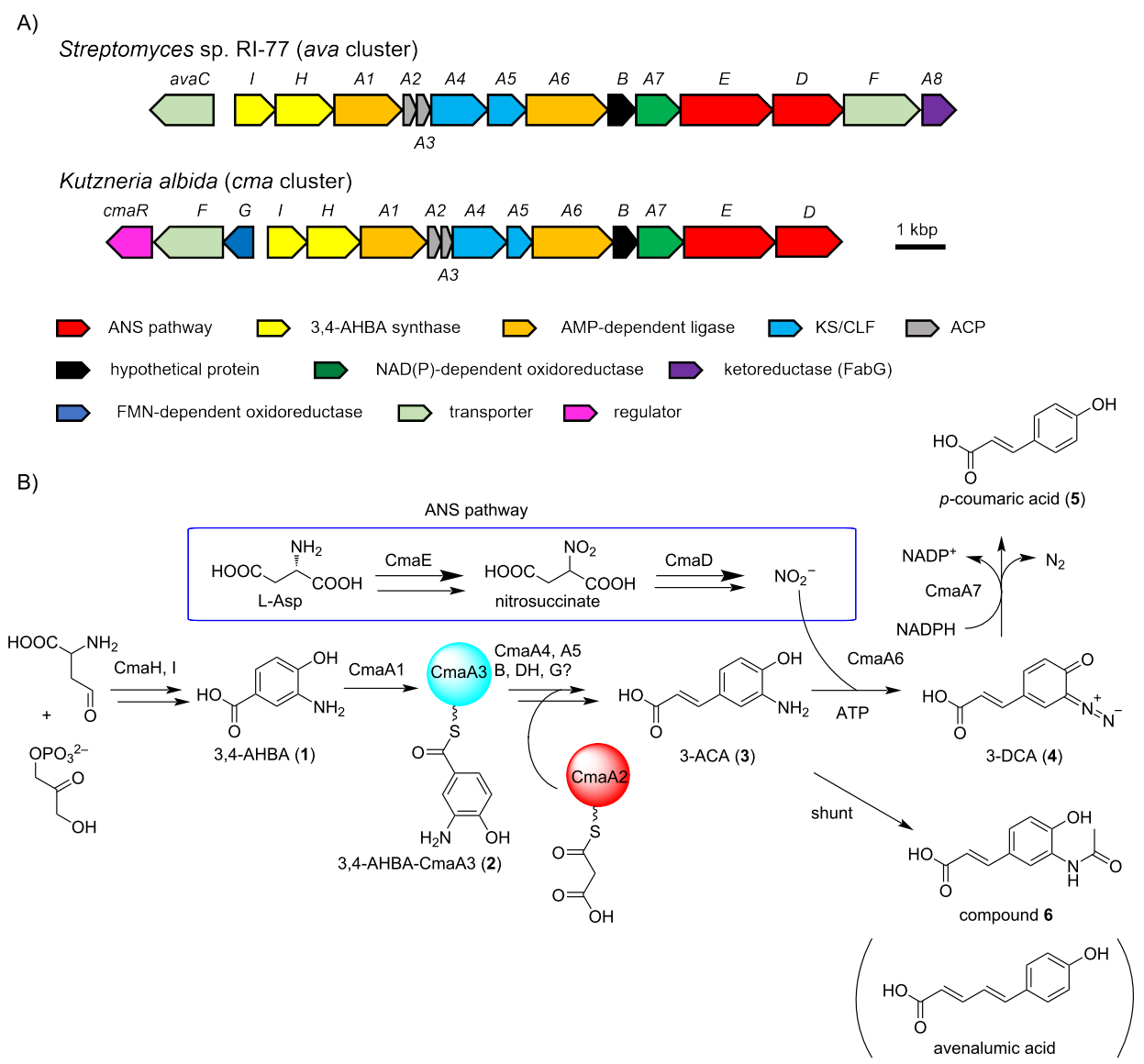



**Figure 1:** Comparison of *ava* and *cma* clusters and the biosynthetic pathway of *p*-coumaric acid. A) Schematic representation of *ava* and *cma* clusters from *Streptomyces* sp. RI-77 and *Kutzneria albida*, respectively. They harbor similar gene components except for the genes encoding oxidoreductases (*avaA8* and *cmaG*), transporter (*avaC*), and regulator (*cmaR*). B) The proposed biosynthetic pathway of *p*-coumaric acid by the *cma* cluster.

**Table 1:** Comparison of *cma* and *ava* gene clusters.[a]

| Genes in *cma* cluster | Genes in *ava* cluster | Amino acid sequence identity (%) | Predicted function of encoded protein |
|---|---|---|---|
| *cmaR* | – | | LysR family transcriptional regulator |
| *cmaF* | *avaF* | 45.9 | major facilitator superfamily transporter |
| *cmaG* | – | | FMN-dependent oxidoreductase |
| – | *avaC* | | major facilitator superfamily transporter |
| *cmaI* | *avaI* | 65.4 | 3,4-AHBA synthase |
| *cmaH* | *avaH* | 66.2 | 3,4-AHBA synthase |
| *cmaA1* | *avaA1* | 61.4 | acyl-ACP ligase |
| *cmaA2* | *avaA2* | 50.5 | acyl carrier protein |
| *cmaA3* | *avaA3* | 51.9 | acyl carrier protein |

**Table 1:** Comparison of *cma* and *ava* gene clusters.[a] (continued)

| | | | |
|---|---|---|---|
| *cmaA4* | *avaA4* | 62.8 | ketosynthase |
| *cmaA5* | *avaA5* | 35.4 | chain length factor |
| *cmaA6* | *avaA6* | 65.3 | ATP-dependent diazotase |
| *cmaB* | *avaB* | 50.8 | hypothetical protein |
| *cmaA7* | *avaA7* | 58.0 | NAD(P)-dependent denitrification enzyme |
| *cmaE* | *avaE* | 57.6 | aspartate monooxygenase (nitrosuccinate synthase) |
| *cmaD* | *avaD* | 64.7 | nitrosuccinate lyase |
| – | *avaA8* | | ketoreductase (FabG) |

[a]"–" denotes no counterpart.

## Heterologous expression of *cma* cluster

To examine whether the *cma* cluster is responsible for the biosynthesis of avenalumic acid or another natural product, we performed heterologous expression of the *cma* cluster. Plasmids named pHKO4-*cmaI-D* and pTYM3a-*cmaG* (Figure 2A) were introduced into *S. albus* by conjugation, resulting in *S. albus-cma* [20]. *S. albus-cma* possesses all the *cma* genes, except for the genes encoding a transporter (*cmaF*) and a regulator (*cmaR*), under the control of *tipA* promoter (Figure 2A). When we cultured *S. albus-cma* and analyzed its metabolites by liquid chromatography–mass spectrometry (LC–MS), formation of avenalumic acid was not observed. Instead, production of *p*-coumaric acid (**5**) was detected (Figure 2B and Figure S3A,D in Supporting Information File 1). In addition, *S. albus-cma* produced compound **6**, which showed $[M + H]^+$ ion at $m/z$ = 222, as well as other putative 3,4-AHBA derivatives, among which compound **9** was indicated to be *N*-acetyl-3,4-AHBA (**9**) by its mass and UV spectra (Figure 2B and Figure S3B,C,E,F in Supporting Information File 1). The control strain of *S. albus*, which harbors two empty vectors, did not produce *p*-coumaric acid and compound **6**, suggesting that they are the biosynthetic products of the *cma* cluster (Figure 2B). The yield of compound **6** decreased when *cmaG* was removed from *S. albus-cma*, indicating that CmaG is involved in the production of **6**, although it is not essential (Figure 2B). In addition, the Δ*cmaG* strain produced a higher amount of compound **9** than *S. albus-cma*. Because compound **9** seems to be a shunt product derived from 3,4-AHBA (**1**), which is the starter substrate of Cma PKS, this result indicated that CmaG is a component of Cma PKS. From NMR analysis and high-resolution (HR)MS analysis ($[M + H]^+$ ion at $m/z$ = 222.0766, which corresponds to $C_{11}H_{12}NO_4^+$, calcd. 222.0761) of purified **6**, the structure of **6** was determined as *N*-acetyl-3-aminocoumaric acid (Figures S9–S13 and Table S1 in Supporting Information File 1). The production of **6** is consistent with our previous work in which *N*-acetyl-3-aminoavenalumic acid was produced by *S. albus-ava* (an recombinant *S. albus* strain expressing the *ava* cluster for 3-aminoavenalumic acid produc-

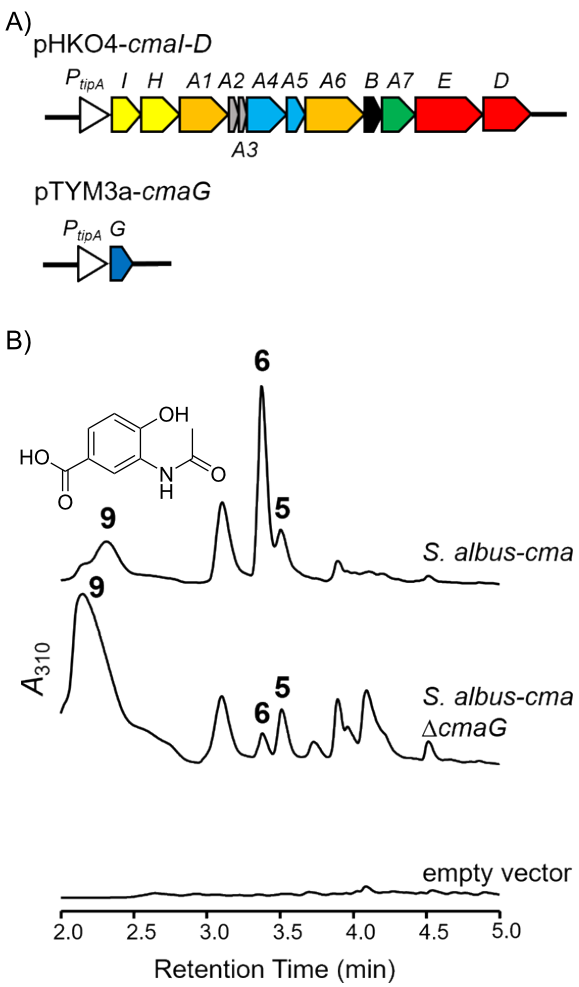



**Figure 2:** Heterologous expression of the *cma* cluster in *S. albus*. A) Schematic representation of the gene sets on heterologous expression plasmids to construct *S. albus-cma*. The *tipA* promoter was used for gene expression. B) Metabolite analysis of *S. albus-cma*. UV chromatograms at 310 nm are shown. The production yield of **6** was decreased in *S. albus-cma* Δ*cmaG*. Several metabolites produced by *S. albus-cma* and *S. albus-cma* Δ*cmaG* are presumed to be shunt products derived from 3,4-AHBA including *N*-acetyl-3,4-AHBA (**9**). These compounds were not produced by a control strain harboring empty vectors (pHKO4 and pTYM3a).

tion) as a shunt product [13]. From these results, we confirmed that the *cma* cluster is responsible for *p*-coumaric acid (**5**) biosynthesis, in which the number of condensations of the malonyl units was one time less than that in the biosynthesis of avenalumic acid.

## In vitro analysis of recombinant Cma proteins

To further understand the biosynthetic machinery encoded by the *cma* cluster, we performed in vitro enzyme analyses using recombinant Cma proteins. First, we obtained recombinant CmaA1, *holo*-CmaA3, and CmaA6 by expressing the corresponding genes in *Escherichia coli* BL21(DE3) (Figure S4 in Supporting Information File 1). Note that *holo*-CmaA3 was obtained by co-expressing with the *sfp* gene to obtain the *holo*-form of CmaA3 [21].

We tested whether CmaA1 and *holo*-CmaA3 are involved in the initial reaction in polyketide synthesis similar to AvaA1 and AvaA3; AvaA1 loads 3,4-AHBA (**1**) onto *holo*-AvaA3 using ATP in the biosynthetic pathway of avenalumic acid. As expected, the formation of 3,4-AHBA-CmaA3 (**2**) was verified by LC–MS analysis of the reaction mixture (Figure 3A). Furthermore, we confirmed that AvaA1 can recognize *holo*-CmaA3 to synthesize 3,4-AHBA-CmaA3 (**2**) and that CmaA1 can recognize *holo*-AvaA3 to synthesize 3,4-AHBA-AvaA3 (Figure S5, Supporting Information File 1).

Next, we analyzed the function of CmaA6, which was predicted to catalyze the diazotization of 3-aminocoumaric acid (3-ACA, **3**) in *p*-coumaric acid biosynthesis. As expected, 3-diazocoumaric acid (3-DCA, **4**) was synthesized by CmaA6 in the presence of ATP and sodium nitrite (Figure 3B). Because we supposed that CmaA6 has a promiscuous substrate specificity, we tested whether CmaA6 could also catalyze the diazotization of 3-aminoavenalumic acid (3-AAA, **7**). As expected, CmaA6 also accepted **7** as a substrate to synthesize 3-DAA (**8**) (Figure 3C). Note that the efficiency of diazotization catalyzed by CmaA6 was much higher than that of AvaA6; almost 100% of 3-ACA and 3-AAA were converted to corresponding aromatic diazo compounds **4** and **8**, respectively (Figure 3C).

## Kinetic analysis of AvaA7 catalyzing denitrification of 3-DAA

The high conversion efficiency of 3-AAA (**7**) to 3-DAA (**8**) catalyzed by CmaA6 motivated us to perform the kinetic analysis of AvaA7, which we could not achieve in our previous study. First, we attempted to determine the kinetic values toward NADH and NADPH using compound **8** synthesized by CmaA6. The absorbance at 435 nm ($A_{435}$) exhibited by **8** was measured during the AvaA7-catalyzed reaction (Figure S6D in Supporting Information File 1), and the decrease of $A_{435}$ was tracked to determine the initial velocity of AvaA7. Although the kinetic values toward NADPH could be calculated, the kinetic values toward NADH could not be obtained probably because the $K_m$ value is very high, suggesting that NADH is not the physiological cofactor of AvaA7 (Figure S6 in Supporting Information File 1 and Table 2). Furthermore, we carried out kinetic analysis of AvaA7 toward **8** in the presence of saturated NADPH (Table 2). The $k_{cat}/K_m$ values toward NADH and **8** were reasonable as an NAD(P)-dependent oxidoreductase, suggesting that the denitrification activity of AvaA7 is not an artifact and that AvaA7-catalyzed denitrification is a genuine reaction in bacterial cells (Figure S6 in Supporting Information File 1 and Table 2).

**Table 2:** Kinetic analysis of AvaA7.

| | vs NADPH | vs 3-DAA (**8**) |
|---|---|---|
| $K_m$ [μM] | 138 ± 10 | 152 ± 31 |
| $k_{cat}$ [$min^{-1}$] | 266 ± 2.0 | 615 ± 55 |
| $k_{cat}/K_m$ [$min^{-1}$ $\mu M^{-1}$] | 1.9 | 4.0 |

# Discussion

In the present study, we showed that the *cma* cluster is responsible for the production of *p*-coumaric acid using heterologous expression experiments and in vitro enzyme assays. We propose the biosynthetic pathway of *p*-coumaric acid as follows. First, 3,4-AHBA (**1**) synthesized by CmaH and CmaI is loaded onto *holo*-CmaA3 to form 3,4-AHBA-CmaA3 (**2**) by CmaA1. 3-ACA (**3**) is then synthesized from 3,4-AHBA-CmaA3 and malonyl-CmaA2 by the highly reducing type II PKS composed of four Cma proteins (CmaA4, CmaA5, CmaB, and CmaG), a dehydratase from fatty acid synthase, and an unknown thioesterase. CmaB should also stimulate the reaction in an unknown manner as its homolog, AvaB [13]. Next, CmaA6 catalyzes the diazotization of 3-ACA (**3**) using nitrous acid generated by the ANS pathway (CmaE and CmaD) to form 3-DCA (**4**). Finally, CmaA7 catalyzes the denitrification of 3-DCA (**4**) using NADPH to synthesize *p*-coumaric acid (**5**).

Although we have not yet investigated the activity of CmaG (a putative FMN-dependent oxidoreductase), it is expected to catalyze the reduction of the keto group of an β-ketoacyl intermediate, similar to the ketoreductase of fatty acid synthase (FAS) and highly reducing type II PKS. This hypothesis is supported by the fact that *cmaG* exists in the *cma* cluster instead of *avaA8*, which encodes a putative FabG-like ketoreductase, and by the observation that the yield of compound **6** decreased considerably when *cmaG* was removed from the heterologous expression system (Figure 1A and Figure 2B, and Figures S1 and S2

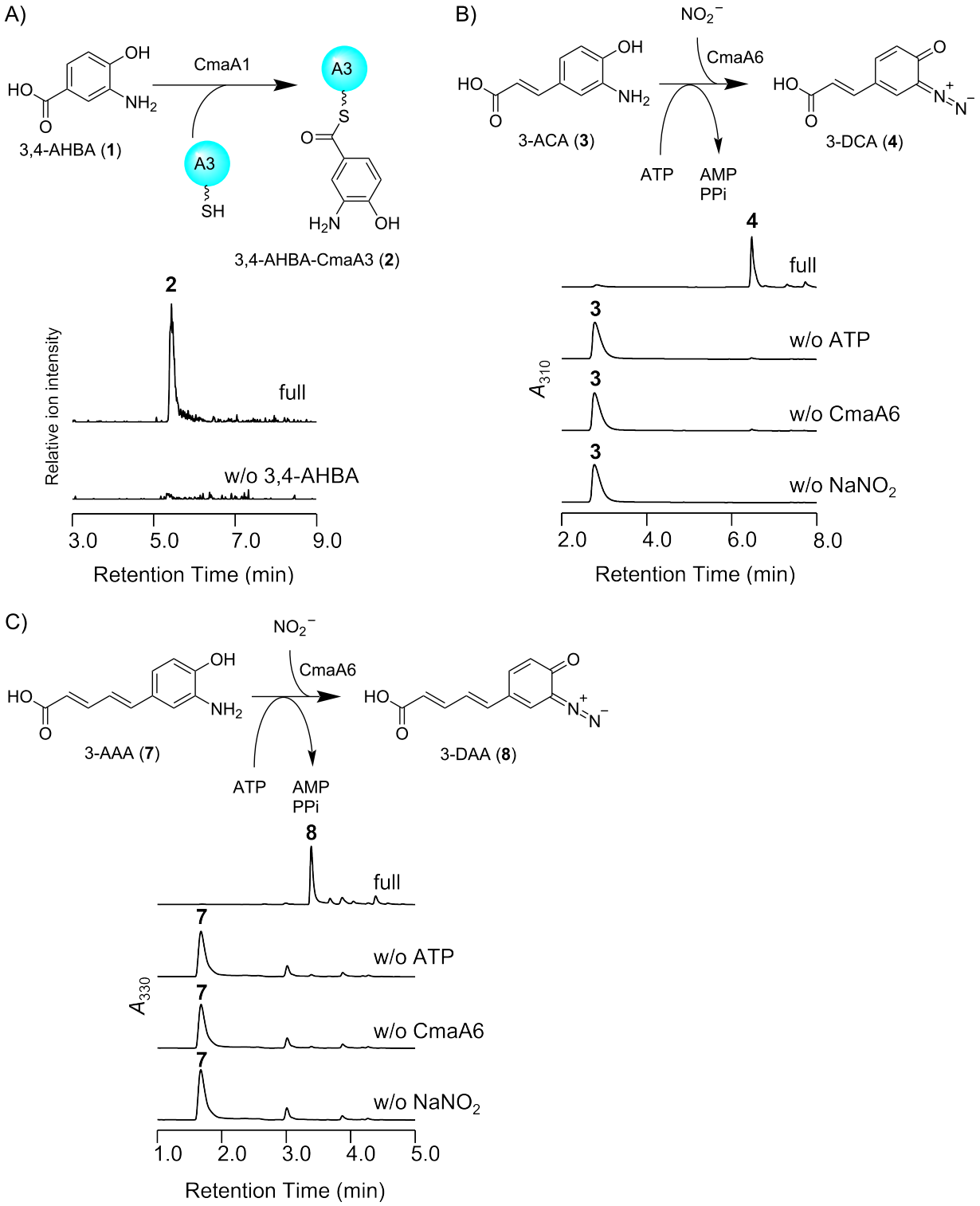



**Figure 3:** In vitro analysis of Cma proteins. A) In vitro analysis of CmaA1 and CmaA3. Extracted ion chromatograms at *m/z* 1087.9, which corresponds to $[M + 10H]^{10+}$ of 3,4-AHBA-CmaA3 (**2**) under positive ion mode, are shown. B, C) In vitro analysis of CmaA6. UV chromatograms at 310 nm and 310 nm are shown. CmaA6 catalyzed diazotization toward 3-ACA (**3**) (B) and 3-AAA (**7**) (C) to synthesize 3-DCA (**4**) and 3-DAA (**8**), respectively, in a high conversion ratio.

in Supporting Information File 1). It should be emphasized that the production of **5** and **6** was not completely abolished in *S. albus-cma* Δ*cmaG*. This is because FAS's ketoreductase probably interacts with the Cma system in the absence of CmaG.

The loading of 3,4-AHBA onto ACP (CmaA3), which is the initial reaction of Cma PKS, is catalyzed by CmaA1, as shown by in vitro analysis. Interestingly, both *ava* and *cma* clusters have two ACP genes. These ACPs are expected to have different roles in each biosynthetic pathway; AvaA3 and CmaA3 carry the starter substrate (3,4-AHBA) and AvaA2 and CmaA2 carry the extender unit (malonyl moiety) [13]. CmaA2 and CmaA3 showed relatively high amino acid sequence similarity to AvaA2 and AvaA3, respectively (≈50% identity; see Table S2 in Supporting Information File 1). In contrast, low similarities were observed between CmaA2 and AvaA3, and between CmaA3 and AvaA2 (Table S2, Supporting Information File 1). Because AvaA1 was shown not to recognize AvaA2, but to recognize AvaA3 in vitro, the homologs of these proteins (CmaA1, CmaA2, and CmaA3) are expected to have the same features. The mechanism by which AvaA1 and CmaA1 distinguish between two different ACPs (AvaA2 and AvaA3 for AvaA1 and CmaA2 and CmaA3 for CmaA1) is a matter of interest. The fact that AvaA1 and CmaA1 recognized CmaA3 and AvaA3, respectively, as partner ACP suggests that there are some important residues conserved between AvaA3 and CmaA3 for the ligase–ACP interaction. Indeed, two amino acid residues (Trp38 and His41 of CmaA3), which seem to be important for the formation of the CmaA1–CmaA3 complex in the structural model predicted by AlphaFold2 [22,23], are conserved in AvaA3 and CmaA3, but not in AvaA2 and CmaA2 (Figure S7 in Supporting Information File 1). Further analysis of these systems would provide important insights into how highly reducing type II PKSs control starter substrate incorporation.

The only difference between avenalumic acid and *p*-coumaric acid is the chain length. The CLF CmaA5 shows 35.4% identity to AvaA5, which is the lowest similarity score among Ava and Cma protein pairs (Table 1). This is reasonable because CLF is involved in the chain length control of type II PKSs [24,25]. Phylogenetic analysis of CLFs encoded in the homolog gene clusters of *ava* cluster-related BGCs indicated that they can be divided into three large clades (Figure S8 in Supporting Information File 1). AvaA5 and CmaA6 belong to different clades, and the proteins belonging to the third clade have a low sequence similarity to both AvaA5 and CmaA6. Thus, the *ava* cluster-related BGCs that encode a CLF belonging to the third clade likely produce analog compounds with different chain lengths.

*p*-Coumaric acid is an intermediate in flavonoid biosynthesis and is much more ubiquitous than avenalumic acid. In general, *p*-coumaric acid is synthesized by phenylalanine ammonia-lyase and cinnamate 4-hydroxylase from phenylalanine or by a tyrosine ammonia-lyase from tyrosine [26]. However, the Cma system synthesizes *p*-coumaric acid through a completely different pathway, which requires at least 12 enzymes and two carrier proteins if two primary metabolites (dihydroxyacetone phosphate and aspartate-4-semialdehyde) are considered as starting materials. Thus, the Cma system appears to be more complicated than the general *p*-coumaric acid synthesis. Therefore, it is interesting to understand why actinomycetes synthesize *p*-coumaric acid using such a specialized biosynthetic pathway including diazotization and denitrification. It should be noted that only three ammonia-lyases for *p*-coumaric acid biosynthesis have been characterized in actinomycetes [27,28]. The Cma system may have some advantages over the general *p*-coumaric acid biosynthetic pathways when actinomycetes evolutionally develop the *p*-coumaric acid biosynthetic pathway in secondary metabolism.

We identified CmaA6 as an ATP-dependent 3-ACA diazotase that showed a very high conversion efficiency. CmaA6 also catalyzed the diazotization of 3-AAA to synthesize 3-DAA with a much higher conversion efficiency than AvaA6 in avenalumic acid biosynthesis; CmaA6 converted almost all 3-ACA (**3**) to 3-DCA (**4**), whereas AvaA6 converted approximately 10% of 3-ACA (**3**) to 3-DCA (**4**) within 1 h. This property is probably due to the high stability of CmaA6. The high activity of CmaA6 makes it a useful tool for biochemical experiments and chemoenzymatic synthesis. Indeed, the discovery of CmaA6 allowed us to analyze the kinetics of the denitrification reaction catalyzed by AvaA7, which had been difficult to achieve in our previous studies because of the low enzymatic activity of AvaA6. The kinetic values for AvaA7 are in a reasonable range for an enzyme of this family (e.g., some NADPH-dependent sugar epimerases reported to have $k_{cat}/K_m$ values of 0.79 and 0.86 $min^{-1}$ $\mu M^{-1}$) and strongly support our previous observation that AvaA7 showed a preference for NADPH as a cofactor [13,29]. In addition, CmaA6 could be an attractive target for understanding the reaction mechanism of ATP-dependent diazotase. CmaA6 could also be an ancestor for generating useful biocatalysts to synthesize diazo group-containing compounds through directed evolution.

Finally, this study is important because of the success of genome mining for BGCs in a rare actinomycete, *K. albida*. Although 47 BGCs were indicated in the *K. albida* genome by the antiSMASH analysis, only aculeximycin was reported as a natural product produced by *K. albida* [30]. Although rare

actinomycetes have been expected to be a source of novel natural products, reports of natural product isolation from rare actinomycetes are limited because of the difficulty in cultivation and genetic manipulation. This study demonstrates that heterologous expression of BGCs combined with promoter swapping is a powerful tool for the discovery of new natural products from rare actinomycetes.

## Conclusion

In summary, we identified a *cma* gene cluster for *p*-coumaric acid biosynthesis in the *K. albida* genome. We demonstrated the following two biosynthetic reactions in the Cma system using in vitro enzyme assays. (i) CmaA1 loads 3,4-AHBA (**1**) onto *holo*-CmaA3 using ATP. (ii) CmaA6 catalyzes the diazotization of 3-ACA (**3**) to produce 3-DCA (**4**) in the presence of ATP and sodium nitrite. By analogy with the Ava system, we concluded that the Cma system biosynthesizes *p*-coumaric acid via diazotization-dependent deamination. The diazotase CmaA6 had a high catalytic activity, which enabled us to determine the kinetics of the denitrification reaction catalyzed by AvaA7. Furthermore, careful comparison between *ava* and *cma* clusters provided insights into (i) the mechanism of protein–protein interaction between the carrier protein and AMP-dependent ligase and (ii) the chain length control of highly reducing type II PKSs.

## Experimental

### Strains, chemicals, and enzymes

*E. coli* JM109 was used for DNA manipulation, and *E. coli* BL21(DE3) was used for expressing recombinant proteins. *E. coli* S17-1 was used for conjugation. *Streptomyces albus* J1074 was used for heterologous expression. *Kutzneria albida* JCM 3240 was purchased from the Japan Collection of Microorganisms. Enzymes used for DNA manipulation, including polymerase and restriction enzymes, were purchased from TaKaRa Bio Inc. (Shiga, Japan). Primers were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Chemicals for the enzymatic assay were purchased from Sigma-Aldrich (St. Louis, MO, USA). Chemicals for organic synthesis were purchased from Tokyo Chemical Industry (Tokyo, Japan). 3-Aminocoumaric acid (**3**) and 3-aminoavenalumic acid (**7**) were synthesized in our previous study [13].

### Construction of heterologous expression vectors

First, pHKO4-*cmaI-D* was constructed by cloning the *cmaI-H-A1-A2-A3-A4-A5-A6-B-A7-E-D* operon (which was amplified by PCR using *K. albida* genomic DNA as the template) into the NdeI and HindIII sites of pHKO4 using In-Fusion (TaKaRa Bio Inc.) [20]. pTYM3a-*cmaG* was constructed by cloning *cmaG* (which was amplified by PCR using *K. albida* genomic DNA as the template) into the NdeI and XbaI sites of pTYM3a using In-Fusion (TaKaRa Bio Inc.). The primers used for plasmid construction are listed in Table S3 of Supporting Information File 1.

### Heterologous expression of the *cma* cluster in *S. albus*

pHKO4-*cmaI-D* and pTYM3a-*cmaG* were introduced into *S. albus* to construct *S. albus-cma* in the same way to construct *S. albus-ava* described in our previous study [13].

*S. albus-cma* was inoculated into 10 mL of Tryptone Soya Broth (TSB) medium with kanamycin (50 mg/L) and incubated with shaking (300 rpm) at 30 °C for 2 days. One milliliter of this preculture medium was inoculated into 100 mL of Waksman medium (5 g/L Bacto™ peptone, 5 g/L Bacto™ yeast extract, 20 g/L glucose, 5 g/L meat extract, 5 g/L NaCl, and 3 g/L $CaCO_3$) with kanamycin (50 mg/L), and incubated with shaking (120 rpm) at 30 °C for 2 days. Thiostrepton (15 mg/L) was then added to the culture, and the culture was continued for another 1 day.

After incubation, brine (0.5 mL) was added to 5 mL of the culture, and the metabolites were extracted with 5 mL of ethyl acetate after adjusting the pH to approximately 4 by adding 6 M HCl. The ethyl acetate layer was collected, and it was washed with an equal volume of distilled water to remove the compounds that can be dissolved in water. The ethyl acetate layer was then collected and evaporated completely in vacuo. The residual materials were dissolved in 200 μL of methanol. The obtained samples were analyzed by liquid chromatography–electrospray ionization mass spectrometry (LC–ESIMS) using an LC-2040C 3D Plus system (Shimazu Corp., Kyoto, Japan) equipped with a COSMOCORE 2.6C18 Packed column (2.1 mm ID × 100 mm, Nacalai Tesque) coupled with a model LCMS-8040 liquid chromatography–mass spectrometer (LC–MS) (Shimazu Corp.). The compounds were eluted with a linear gradient of water/acetonitrile containing 0.1% formic acid.

### Isolation and structural determination of compound **6**

*S. albus-cma* was inoculated into 10 mL of TSB medium with kanamycin (50 mg/L) and incubated with shaking (300 rpm) at 30 °C for 2 days. The whole culture was transferred into 1 L of Waksman medium with kanamycin (50 mg/L) and incubated with shaking (120 rpm) at 30 °C for 2 days. Then, thiostrepton (15 mg/L) was added to the culture, and the culture was continued for another 1 day. The pH was adjusted to 4 by adding 6 M HCl to the culture, and the metabolites were extracted by ethyl acetate. After the ethyl acetate was

evaporated, the metabolites were absorbed for 5.0 g silica gel 60 (0.040–0.063 mm, Merck Millipore, Burlington, MA, USA). The silica gel holding the metabolites was applied to a normal-phase medium-pressure liquid chromatography system (MPLC, Shoko Scientific, Kanagawa, Japan) equipped with a silica column (Purif-Pack, Shoko Scientific), and the metabolites were eluted with a linear gradient of chloroform/methanol. Fractions containing compound **6** were concentrated by evaporation. The residual materials were desorbed in 1 mL DMSO and applied to a reversed-phase high-performance liquid chromatography (HPLC, Shimadzu Corp.) equipped with a COSMOCORE Packed column 5C18-AR-II (10 mm ID × 250 mm, Nacalai Tesque), and the metabolites were eluted with a linear gradient of water/acetonitrile containing 0.1% formic acid. Fractions containing compound **6** were concentrated by evaporation. Compound **6** was then desorbed in DMSO-$d_6$, and the structure was determined by the JNM-A600 NMR system (JEOL, Tokyo, Japan) (Table S1 and Figures S9–S13 in Supporting Information File 1).

## Production and purification of recombinant proteins

pColdI-*cmaA1*, pColdI-*cmaA3*, and pColdI-*cmaA6* were constructed by cloning each gene (which was amplified by PCR using pHKO4-*cmaI-D* as a template) into the NdeI and XhoI sites of pColdI using In-Fusion (TaKaRa Bio Inc.). The primers used for plasmid construction are listed in Table S4 of Supporting Information File 1.

Each protein expression plasmid was introduced into *E. coli* BL21(DE3). To obtain *holo*-CmaA3, pACYC-*sfp* [21] was co-introduced with pColdI-*cmaA3* into *E. coli* BL21(DE3). The obtained strain was cultured in 1 L of Terrific broth (24 g/L yeast extract, 12 g/L tryptone, 0.8% glycerol, 0.94% $K_2HPO_4$, and 0.22% $KH_2PO_4$) with ampicillin and chloramphenicol (if necessary for *sfp* expression) and incubated with shaking (150 rpm) at 37 °C until the $OD_{600}$ reached 0.8. After the culture was cooled on ice for 30 min, isopropyl β-D-thiogalactopyranoside (IPTG) was added to the culture (for CmaA1 and CmaA3 expression, final concentration was 100 μM; for CmaA6 expression, IPTG was not added) and incubated with shaking (150 rpm) at 15 °C for 20 h. Cells were harvested by centrifugation and suspended in lysis buffer (20 mM HEPES-NaOH, 10% glycerol, and 200 mM $NaCl_2$; pH 8.0). The cells were disrupted by sonication on ice, and cell debris was removed by centrifugation. The recombinant protein was purified using His60 Ni Superflow Resin (TaKaRa Bio Inc.). The protein was eluted using a stepwise gradient of imidazole in lysis buffer (20–500 mM imidazole). The buffer was replaced with lysis buffer using an Amicon Ultra centrifugal filter with a suitable molecular mass cutoff (Merck Millipore).

Recombinant AvaA1 and AvaA3 obtained in our previous study [13] were used in the current experiment.

## In vitro analysis of CmaA1, CmaA3, AvaA1, and AvaA3

A reaction mixture (50 μL) containing 4 μM CmaA1 or AvaA1, 40 μM CmaA3 or AvaA3, 2.5 mM 3,4-AHBA, 1 mM ATP, 5 mM $MgCl_2$, 20 mM HEPES-NaOH (pH 8.0), 10% glycerol, and 200 mM NaCl was prepared and incubated at 30 °C for 1 h. After centrifugation, the supernatant was analyzed using LC–MS equipped with a BioResolve RP mAb polyphenyl column (2.1 mm ID × 100 mm, Waters, Milford, MA, USA). The compounds were eluted using a linear gradient of water/acetonitrile containing 0.1% formic acid.

## In vitro analysis of CmaA6

A reaction mixture (50 μL) containing 5 μM CmaA6, 0.5 mM 3-ACA (**3**), 1 mM ATP, 5 mM $NaNO_2$, 5 mM $MgCl_2$, 20 mM HEPES-NaOH (pH 8.0), 10% glycerol, and 200 mM NaCl was prepared and incubated at 30 °C for 1 h. The reaction was quenched by adding 50 μL of methanol. After centrifugation, the supernatant was analyzed by LC–MS equipped with a COSMOCORE 2.6Pbr Packed column (2.0 mm ID × 100 mm, Nacalai Tesque). The compounds were eluted with a linear gradient of water/acetonitrile containing 0.1% formic acid.

A reaction mixture (50 μL) containing 5 μM CmaA6, 0.5 mM 3-AAA (**7**), 1 mM ATP, 5 mM $NaNO_2$, 5 mM $MgCl_2$, 20 mM HEPES-NaOH (pH 8.0), 10% glycerol, and 200 mM NaCl was prepared and incubated at 30 °C for 1 h. The reaction was quenched by adding 50 μL methanol. After centrifugation, the supernatant was analyzed by LC–MS, following the same method used for "Heterologous expression of *cma* cluster using *S. albus*."

## Kinetic analysis of AvaA7

To measure the kinetic parameters of NADPH or NADH, the reaction mixture (90 μL) containing 10 μM CmaA6, 1.0 mM 3-AAA (**7**), 2 mM ATP, 5 mM $NaNO_2$, 2.5 mM $MgCl_2$, 20 mM HEPES-NaOH (pH 8.0), 10% glycerol, and 200 mM NaCl was prepared and incubated at 30 °C for 1 h. Then, 5.0 μL of 4.0 μM AvaA7 (final concentration: 0.2 μM) and 5.0 μM of different concentrations of NADPH or NADH (final concentration: 25–800 μM) were added to the reaction mixture (final reaction volume: 100 μL). The initial velocity of the reaction catalyzed by AvaA7 was estimated by monitoring the decrease in absorbance at 435 nm for 3-DAA using a SpectraMax M2 microplate reader (Molecular Devices, San Jose, CA, USA).

The kinetic parameters ($V_{max}$, $k_{cat}$, and $K_m$) were calculated by fitting the substrate concentration [S]–initial velocity ($v$) plot using the equation $v = V_{max} \times [S] / (K_m + [S])$.

To measure the kinetic parameters of 3-DAA, the reaction mixture (90 μL) containing 10 μM CmaA6, 1.0 mM 3-AAA (**7**), 2 mM ATP, 5 mM $NaNO_2$, 2.5 mM $MgCl_2$, 20 mM HEPES-NaOH (pH 8.0), 10% glycerol, and 200 mM NaCl was prepared and incubated at 30 °C for 1 h. Then, the concentration of 3-DAA (**8**) in the reaction mixture was calculated by measuring the absorbance at 435 nm for 3-DAA (**8**) using a SpectraMax M2 microplate reader. The initial velocity of the reaction catalyzed by AvaA7 was calculated following the same method as mentioned above, after diluting the reaction mixture to a final concentration of 25 to 400 μM and adding AvaA7 (final concentration: 0.2 μM) and NADPH (final concentration: 800 μM).

## Supporting Information

### Supporting Information File 1

Additional experimental data and NMR spectra.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-1-S1.pdf]

## Funding

This work was supported by Grants-in-Aid for Scientific Research on Innovative Areas (19H04645 to Y.K., 19H05685 to Y.O.), Grant-in-Aid for Transformative Research Areas (22H05130 to Y.K.), Grant-in-Aid for JSPS Fellows (22KJ1046 to S.K.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the World-leading Innovative Graduate Study Program in Proactive Environmental Studies, The University of Tokyo (to S.K.), and JSPS A3 Foresight Program grant (to Y.O.).

## ORCID® iDs

Seiji Kawai - https://orcid.org/0000-0003-3989-2547
Akito Yamada - https://orcid.org/0009-0002-8789-1901
Yohei Katsuyama - https://orcid.org/0000-0003-4905-5320
Yasuo Ohnishi - https://orcid.org/0000-0001-7633-9236

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Blin, K.; Kim, H. U.; Medema, M. H.; Weber, T. *Briefings Bioinf.* **2019,** *20,* 1103–1113. doi:10.1093/bib/bbx146
2. Nett, M.; Ikeda, H.; Moore, B. S. *Nat. Prod. Rep.* **2009,** *26,* 1362–1384. doi:10.1039/b817069j
3. Albarano, L.; Esposito, R.; Ruocco, N.; Costantini, M. *Mar. Drugs* **2020,** *18,* 199. doi:10.3390/md18040199
4. Sugai, Y.; Katsuyama, Y.; Ohnishi, Y. *Nat. Chem. Biol.* **2016,** *12,* 73–75. doi:10.1038/nchembio.1991
5. Hagihara, R.; Katsuyama, Y.; Sugai, Y.; Onaka, H.; Ohnishi, Y. *J. Antibiot.* **2018,** *71,* 911–919. doi:10.1038/s41429-018-0088-1
6. Katsuyama, Y.; Matsuda, K. *Curr. Opin. Chem. Biol.* **2020,** *59,* 62–68. doi:10.1016/j.cbpa.2020.05.002
7. He, H.-Y.; Niikura, H.; Du, Y.-L.; Ryan, K. S. *Chem. Soc. Rev.* **2022,** *51,* 2991–3046. doi:10.1039/c7cs00458c
8. Chen, L.; Deng, Z.; Zhao, C. *ACS Chem. Biol.* **2021,** *16,* 559–570. doi:10.1021/acschembio.1c00052
9. Kawai, S.; Sugaya, Y.; Hagihara, R.; Tomita, H.; Katsuyama, Y.; Ohnishi, Y. *Angew. Chem., Int. Ed.* **2021,** *60,* 10319–10325. doi:10.1002/anie.202100462
10. Waldman, A. J.; Balskus, E. P. *J. Org. Chem.* **2018,** *83,* 7539–7546. doi:10.1021/acs.joc.8b00367
11. Ma, G.-L.; Candra, H.; Pang, L. M.; Xiong, J.; Ding, Y.; Tran, H. T.; Low, Z. J.; Ye, H.; Liu, M.; Zheng, J.; Fang, M.; Cao, B.; Liang, Z.-X. *J. Am. Chem. Soc.* **2022,** *144,* 1622–1633. doi:10.1021/jacs.1c10369
12. Del Rio Flores, A.; Twigg, F. F.; Du, Y.; Cai, W.; Aguirre, D. Q.; Sato, M.; Dror, M. J.; Narayanamoorthy, M.; Geng, J.; Zill, N. A.; Zhai, R.; Zhang, W. *Nat. Chem. Biol.* **2021,** *17,* 1305–1313. doi:10.1038/s41589-021-00895-3
13. Kawai, S.; Hagihara, R.; Shin-Ya, K.; Katsuyama, Y.; Ohnishi, Y. *Angew. Chem., Int. Ed.* **2022,** *61,* e202211728. doi:10.1002/anie.202211728
14. Kawai, S.; Yamada, A.; Du, D.; Sugai, Y.; Katsuyama, Y.; Ohnishi, Y. *ACS Chem. Biol.* **2023,** *18,* 1821–1828. doi:10.1021/acschembio.3c00248
15. Du, D.; Katsuyama, Y.; Shin-ya, K.; Ohnishi, Y. *Angew. Chem., Int. Ed.* **2018,** *57,* 1954–1957. doi:10.1002/anie.201709636
16. Grammbitter, G. L. C.; Schmalhofer, M.; Karimi, K.; Shi, Y.-M.; Schöner, T. A.; Tobias, N. J.; Morgner, N.; Groll, M.; Bode, H. B. *J. Am. Chem. Soc.* **2019,** *141,* 16615–16623. doi:10.1021/jacs.8b10776
17. Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2021,** *49,* W29–W35. doi:10.1093/nar/gkab335
18. Navarro-Muñoz, J. C.; Selem-Mojica, N.; Mullowney, M. W.; Kautsar, S. A.; Tryon, J. H.; Parkinson, E. I.; De Los Santos, E. L. C.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; Roeters, A.; Lokhorst, W.; Fernandez-Guerra, A.; Cappelini, L. T. D.; Goering, A. W.; Thomson, R. J.; Metcalf, W. W.; Kelleher, N. L.; Barona-Gomez, F.; Medema, M. H. *Nat. Chem. Biol.* **2020,** *16,* 60–68. doi:10.1038/s41589-019-0400-9
19. Wang, M.; Ryan, K. S. *J. Am. Chem. Soc.* **2023,** *145,* 16718–16725. doi:10.1021/jacs.3c04447
20. Tsutsumi, H.; Katsuyama, Y.; Izumikawa, M.; Takagi, M.; Fujie, M.; Satoh, N.; Shin-ya, K.; Ohnishi, Y. *J. Am. Chem. Soc.* **2018,** *140,* 6631–6639. doi:10.1021/jacs.8b02769
21. Taguchi, T.; Awakawa, T.; Nishihara, Y.; Kawamura, M.; Ohnishi, Y.; Ichinose, K. *ChemBioChem* **2017,** *18,* 316–323. doi:10.1002/cbic.201600589

22. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature* **2021,** *596,* 583–589. doi:10.1038/s41586-021-03819-2
23. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. *Nat. Methods* **2022,** *19,* 679–682. doi:10.1038/s41592-022-01488-1
24. Tang, Y.; Tsai, S.-C.; Khosla, C. *J. Am. Chem. Soc.* **2003,** *125,* 12708–12709. doi:10.1021/ja0378759
25. Du, D.; Katsuyama, Y.; Horiuchi, M.; Fushinobu, S.; Chen, A.; Davis, T. D.; Burkart, M. D.; Ohnishi, Y. *Nat. Chem. Biol.* **2020,** *16,* 776–782. doi:10.1038/s41589-020-0530-0
26. Jendresen, C. B.; Stahlhut, S. G.; Li, M.; Gaspar, P.; Siedler, S.; Förster, J.; Maury, J.; Borodina, I.; Nielsen, A. T. *Appl. Environ. Microbiol.* **2015,** *81,* 4458–4476. doi:10.1128/aem.00405-15
27. Cui, P.; Zhong, W.; Qin, Y.; Tao, F.; Wang, W.; Zhan, J. *Bioprocess Biosyst. Eng.* **2020,** *43,* 1287–1298. doi:10.1007/s00449-020-02325-5
28. Zhu, Y.; Liao, S.; Ye, J.; Zhang, H. *Biotechnol. Lett.* **2012,** *34,* 269–274. doi:10.1007/s10529-011-0755-9
29. Ghosh, M. K.; Xiang, D. F.; Thoden, J. B.; Holden, H. M.; Raushel, F. M. *Biochemistry* **2022,** *61,* 2036–2048. doi:10.1021/acs.biochem.2c00364
30. Rebets, Y.; Tokovenko, B.; Lushchyk, I.; Rückert, C.; Zaburannyi, N.; Bechthold, A.; Kalinowski, J.; Luzhetskyy, A. *BMC Genomics* **2014,** *15,* 885. doi:10.1186/1471-2164-15-885

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.1

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Discovery of unguisin J, a new cyclic peptide from *Aspergillus heteromorphus* CBS 117.55, and phylogeny-based bioinformatic analysis of UngA NRPS domains

Sharmila Neupane[‡1,2], Marcelo Rodrigues de Amorim[‡1,2,3] and Elizabeth Skellam[*1,2]

Full Research Paper

Open Access

Address:
[1]Department of Chemistry, University of North Texas, 1155 Union Circle, Denton, TX 76203, USA, [2]BioDiscovery Institute, University of North Texas, 1155 Union Circle, Denton, TX 76203, USA and [3]Instituto de Química de São Carlos, Universidade de São Paulo, CP 780, CEP 13560-970, São Carlos, SP, Brazil

Email:
Elizabeth Skellam[*] - Elizabeth.skellam@unt.edu

* Corresponding author ‡ Equal contributors

Keywords:
adenylation domain; condensation domain; fungal non-ribosomal peptide synthetase; heptapeptide; unguisin biosynthesis

*Beilstein J. Org. Chem.* **2024,** *20,* 321–330.
https://doi.org/10.3762/bjoc.20.32

Received: 15 December 2023
Accepted: 07 February 2024
Published: 19 February 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: T. Awakawa

© 2024 Neupane et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Several under-explored *Aspergillus* sp. produce intriguing heptapeptides containing a γ-aminobutyric acid (GABA) residue with as yet unknown biological functions. In this study, a new GABA-containing heptapeptide – unguisin J (**1**) – along with known unguisin B (**2**) were isolated from a solid culture of *Aspergillus heteromorphus* CBS 117.55. The structure of compound **1** was elucidated by extensive 1D and 2D NMR spectroscopic analysis including HSQC, HMBC, COSY, and 2D NOESY as well as HRESIMS. The stereochemistry of **1** and **2** was determined by Marfey's method. A biosynthetic gene cluster (BGC) encoding unguisins B and J was compared to characterized BGCs in other *Aspergillus* sp. Since the unguisin family of heptapetides incorporate different amino acid residues at different positions of the peptide, the A and C domains of the UngA NRPS were analyzed in an attempt to understand the lack of substrate specificity observed.

## Introduction

Unguisins are a small family of fungal cyclic heptapeptides isolated predominantly from *Aspergillus* sp. [1-8]. Distinctive features of these cyclic peptides include the non-proteinogenic amino acid γ-aminobutyric acid (GABA) and the incorporation of up to five D-amino acids (Figure 1) [1-8]. The amino acids at positions 1 (D-Ala) and 7 (GABA) are conserved but there is considerable variability at positions 2–6, including the incorporation of additional non-proteinogenic amino acids β-methylphenylalanine (βMePhe) and kynurenine (Kyn) [3,4]. So far, no significant biological activities have been reported for these small peptides [3,4,9], however, unguisin A has been shown to bind a series of anions [10].

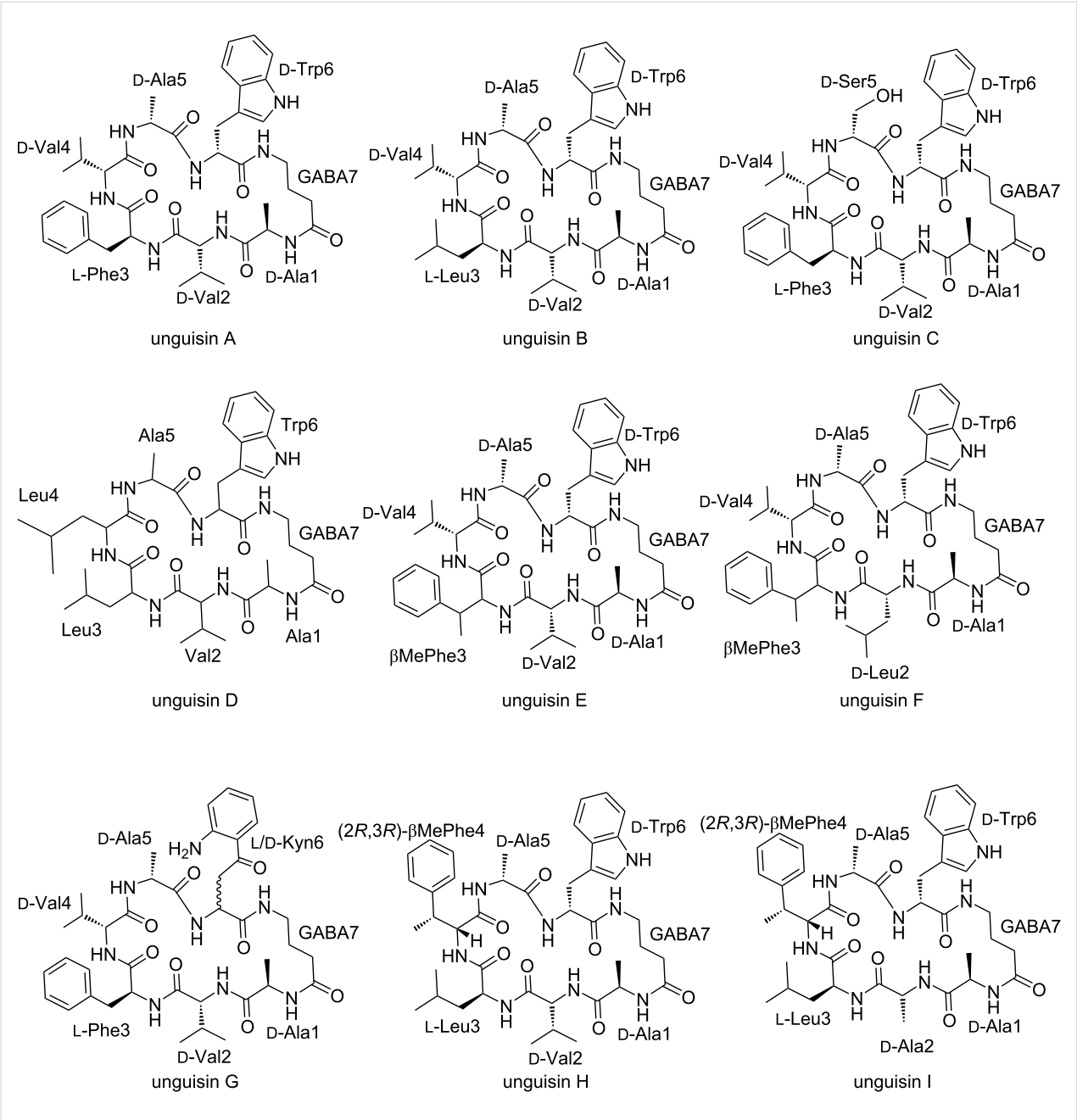



**Figure 1:** Structures of unguisins.

Recently the biosynthesis of unguisins A and B from *Aspergillus violaceofuscus* CBS 115571 was reported [5]. A seven module non-ribosomal peptide synthetase (NRPS; UngA) was heterologously expressed in *Aspergillus oryzae* NSAR1 yielding both unguisins A and B, which differ by the incorporation of D-Phe and D-Leu at position 3, respectively. The highly conserved D-Ala at position 1 was shown to be synthesized from L-alanine via the PLP-dependent alanine racemase UngC [5]. A hydrolase/peptidase (UngD) was also discovered that linearized the cyclic unguisins to linear peptides during in vitro investigations, although the linear peptides were not detected from the fungal cultures.

NRPS enzymes are large multifunctional enzymes that often synthesize very important bioactive molecules [11,12]. These enzymes consist of several catalytic domains organized into modules. Typically, a module possesses an adenylation (A) domain for selecting and activating amino- or keto acids, a thiolation (T) domain for shuttling intermediates between catalytic domains, and a condensation (C) domain that catalyzes amide

or ester bond formation. Additional common domains include epimerization (E) domains for converting naturally occurring L-amino acids to D-amino acids, methyltransferase (MT) domains that typically methylate specific N atoms, and terminal condensation ($C_T$) domains which cyclize the growing peptide chain and facilitate release from the NRPS. Of the fungal NRPS studied to date, many appear to have some tolerance for the range of amino acids incorporated by the A domains and the C domain has been highlighted as a gatekeeper [13].

Here, we describe the isolation of unguisin B, and a new congener named unguisin J, from *Aspergillus heteromorphus* CBS 117.55. We also perform bioinformatic analysis of the A and C domains of the UngA NRPS enzymes involved in their biosynthesis to try and rationalize the relaxed substrate specificity observed in this family of heptapeptides.

## Results and Discussion

The cultivation of *A. heteromorphus* CBS 117.55 on rice solid medium yielded an organic-soluble extract, which was subjected to fractionation using preparative HPLC-PDA-ELSD and purification by semipreparative HPLC-PDA; this led to the isolation of a new cyclic peptide **1**, along with unguisin B (**2**, Figure 2). The structure of the new compound **1** was elucidated by 1D and 2D NMR and HRESIMS/MS. Unguisin B was identified by the $^1$H and $^{13}$C NMR data with the reported data [1,5].

Compound **1** was obtained as a white amorphous solid optically active, with $[\alpha]_D^{22}$ +23.4 (*c* 0.1, MeOH). Its molecular formula was established as $C_{41}H_{56}N_8O_7$ by HRMS ($[M + H]^+$ at *m/z* 773.4338, calculated for $C_{41}H_{57}N_8O_7^+$, *m/z* 773.4345, Δ 0.9 ppm; $[M + Na]^+$ at *m/z* 795.4162, calculated for $C_{41}H_{56}N_8O_7Na^+$, *m/z* 795.4164, Δ 0.3 ppm) and NMR data analysis, corresponding to eighteen indices of hydrogen deficiency. Its UV spectrum exhibited absorption maxima at $\lambda_{max}$ 219 and 279 nm.

The $^1$H and $^{13}$C NMR spectra of **1** revealed the presence of seven amide NH signals between $\delta_H$ 7.43 and 8.44 ppm supported by the amide carbonyl signals at $\delta_C$ 173.1, 172.6, 172.6, 172.1, 172.1, 171.1 and 171.0 ppm (Table 1). An additional NH signal at $\delta_H$ 10.82 ppm and four aromatic signals at $\delta_H$ 7.50, 7.33, 7.07 and 6.97 ppm, exhibiting key $^1$H,$^1$H-COSY and HMBC correlations, suggested a tryptophan aromatic amino acid portion (Figure 3). The other six amino acid residues were assigned based on 2D NMR spectra ($^1$H-$^1$H COSY, HSQC and HMBC) as Ala (2 equiv), Phe (1 equiv), Leu (1 equiv), Val (1 equiv), and γ-aminobutyric acid (GABA) (1 equiv). In addition to the COSY and HMBC correlations, the NOESY experiment showed important interactions between the NH signals corroborating with the peptide sequence defined to be Ala-1, Val-2, Leu-3, Phe-4, Ala-5, Trp-6, and GABA-7 (Figure 3).

Analysis of the NMR data of **1** allowed identifying characteristic $^1$H and $^{13}$C signals very similar to those of unguisin B (**2**) [1,5], the difference being the replacement of the Phe-4 in **1** by Val-4 in **2**. This assignment was confirmed by observation of HMBC correlations from $\delta_H$ 7.98 (NH) and $\delta_H$ 8.44 (NH) to C=O ($\delta_C$ 173.1) and from $\delta_H$ 2.93 and 3.01 ($H_2$-β) to C=O ($\delta_C$ 171.1) (Figure 3), together with key NOESY interactions between the NH signals at $\delta_H$ 7.98↔8.44↔8.14.

The absolute configuration of **1** was assigned by Marfey's method [14]. Comparison of the retention time by LC–MS between the derivatized **1** as well as the authentic amino acid samples determined the structure of **1** as cyclo(D-alanine-D-valine-L-leucine-D-phenylalanine-D-alanine-D-tryptophan-GABA). Compound **1** was named as unguisin J.

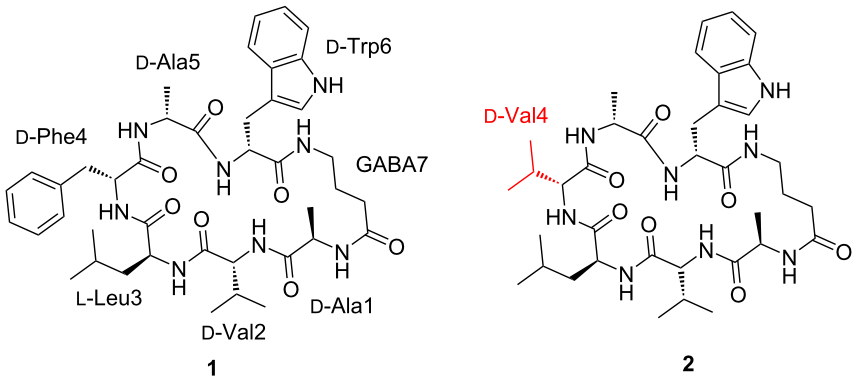



**Figure 2:** Chemical structures of unguisin J (**1**) and unguisin B (**2**).

**Table 1:** 1D and 2D NMR data for **1** ($^1$H: 500 MHz, $^{13}$C: 125 MHz; DMSO-$d_6$).

| **1** | | | | | | |
|---|---|---|---|---|---|---|
| residue/position | | $\delta_C$ | $\delta_H$ (mult., *J* in Hz) | NOESY | HMBC | COSY |
| alanine | NH | – | 8.35 (d, 4.2) | 2.14, 1.98, 7.42 | 172.1, 50.2 | 3.83 |
| | Cα | 50.2 | 3.83 (m) | | 17.0 | 8.35, 1.13 |
| | Cβ | 17.0 | 1.13 (d, 6.9) | | 172.6, 50.2 | 3.83 |
| | C=O | 172.1 | – | – | – | – |
| valine | NH | – | 7.43 (d 8.9) | 8.35, 8.14 | 172.1 | 3.98 |
| | Cα | 58.4 | 3.98 (t 8.7) | | 30.0, 18.2, 172.1 | 7.43, 1.98 |
| | Cβ | 30.0 | 1.98 (m) | | 19.0, 58.4, 172.6 | 3.98, 0.68 |
| | Cγ | 18.2 | 0.68 (d, 6.7) | | 58.4, 30.0, 19.0 | 1.98, 0.78 |
| | Cγ | 19.0 | 0.78 (d, 6.7) | | 58.4, 30.0, 18.2 | 0.68 |
| | C=O | 172.1 | – | – | – | – |
| leucine | NH | – | 8.14 (d, 7.6) | 7.43 | 171.1, 51.8 | 3.88 |
| | Cα | 51.8 | 3.88 (m) | | 39.4, 172.1 | 8.14, 1.36 |
| | Cβ | 39.4 | 1.36 (m) | | 23.6, 23.3, 20.8 | 3.88, 0.98 |
| | Cγ | 23.6 | 0.98 (m) | | – | 1.36, 0.56 |
| | Cδ | 20.8 | 0.56 (d, 6.6) | | 39.4, 23.6, 23.3 | 0.98, 0.69 |
| | Cδ | 23.3 | 0.69 (d, 6.6) | | 39.4, 23.6, 20.8 | 0.98, 0.56 |
| | C=O | 171.1 | – | – | – | – |
| phenylalanine | NH | – | 8.44 (d, 4.7) | 8.14, 7.98 | 173.1 | 4.15 |
| | Cα | 55.8 | 4.15 (m) | | 171.1 | 8.44, 2.93, 3.01 |
| | Cβ | 36.1 | 2.93 (m), 3.01 (m) | | 55.8, 129.2, 137.2 | 4.15 |
| | C-1 | 137.2 | – | – | – | – |
| | C-2, C-6 | 129.2 | 7.16 (dd, 7.5, 1.3) | | 126.4 | 7.24 |
| | C-3, C-5 | 128.3 | 7.24 (d, 7.5) | | 137.2, 128.3 | 7.16, 7.19 |
| | C-4 | 126.4 | 7.19 (dd, 7.5, 1.3) | | 129.2 | 7.24 |
| | C=O | 173.1 | – | – | – | – |
| alanine | NH | – | 7.98 (d, 5.1) | 4.06, 1.12 | 172.6 | 4.06 |
| | Cα | 48.7 | 4.06 (m) | | 173.1, 17.5 | 7.98, 4.06 |
| | Cβ | 17.5 | 1.12 (d, 6.9) | | 173.1, 48.7 | 4.06 |
| | C=O | 172.6 | – | – | – | – |
| tryptophan | NH | – | 7.86 (d, 6.9) | 7.61 | 172.6 | 4.01 |
| | Cα | 55.3 | 4.01 (m) | | 172.6, 171.1, 110.7, 25.0 | 7.86, 3.22 |
| | Cβ | 25.0 | 3.22 (m) | | 127.2, 123.7, 110.7, 55.3 | 4.01 |
| | NH | – | 10.82 (d, 1.7) | | 136.4, 127.2, 123.7, 110.7 | 7.06 |
| | C-2 | 123.7 | 7.06 (br s) | | 110.7, 127.2, 136.4 | 10.82 |
| | C-3 | 110.7 | – | – | – | – |
| | C3a | 127.2 | – | – | – | – |
| | C-4 | 118.3 | 7.50 (br d, 8.0) | | 136.4, 121.1, 127.2w, 110.7w | 6.97 |
| | C-5 | 118.4 | 6.97 (dt, 8.0, 0.9) | | 111.5, 127.2 | 7.50, 6.97 |
| | C-6 | 121.1 | 7.07 (dt, 8.0, 1.0) | | 136.4, 118.4, 118.3 | 6.97, 7.33 |
| | C-7 | 111.5 | 7.33 (dt, 8.0, 0.9) | | 127.2, 118.4, 118.3 | 7.07 |
| | C-7a | 136.4 | – | – | – | – |
| | C=O | 171.0 | – | – | – | – |
| GABA | NH | – | 7.61 (dd, 5.4, 4.3) | 7.86, 4.01 | 171.0 | 3.07, 2.99 |
| | Cα | 38.8 | 3.07 (m), 2.99 (m) | | 25.6 | 7.61, 1.69, 1.60 |

**Table 1:** 1D and 2D NMR data for **1** ($^1$H: 500 MHz, $^{13}$C: 125 MHz; DMSO-$d_6$). (continued)

| | | | | | |
|---|---|---|---|---|---|
| Cβ | 25.6 | 1.69 (m), 1.60 (m) | | 38.8w | 3.07, 2.99, 1.69, 1.60 |
| Cγ | 32.9 | 2.14 (m), 1.98 (m) | | 25.6, 38.8 | 1.69, 1.60 |
| C=O | 172.6 | – | – | – | |

w. weak.

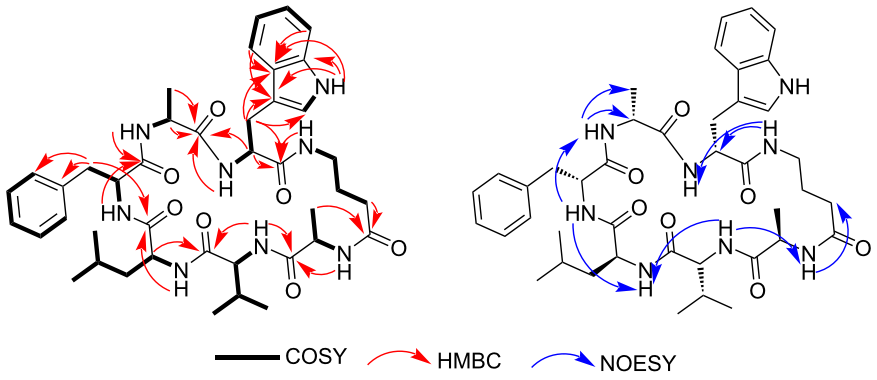



**Figure 3:** Key *g*HMBC and *g*COSY correlations, and NOESY interactions of **1**.

A second peptide was isolated from the same culture of *A. heteromorphus* CBS 117.55. Compound **2** was obtained as an amorphous white powder, $[\alpha]_D^{22}$ +37 (*c* 0.1, EtOH) [lit +40 (*c* 1.0, EtOH)] [5]; for the $^1$H and $^{13}$C NMR spectroscopic data, see Table S2 in Supporting Information File 1. By comparison with literature data this compound was identified as unguisin B (**2**) [1,5], further corroborating the identification of the new unguisin J (**1**).

To the best of our knowledge these are the first metabolites reported from *A. heteromorphus* CBS 117.55.

The co-isolation of unguisins B and J indicates that module 4 of the NRPS is able to accept two different amino substrates and so may possess subtle differences to UngA from *A. violaceofuscus* CBS 115571 which has relaxed substrate specificity in module 3. We performed genome mining of the publicly available *A. heteromorphus* CBS 117.55 (accession number MSFL00000000.1) [15] using fungiSMASH and identified a four gene BGC encoding a seven module NRPS, an alanine-racemase, a hydrolase, and a transporter. We named this BGC *ung''* to distinguish it from the *ung* BGC present in *A. violaceofuscus* and the *ung'* BGC in *A. campestris* IBT 28561 which encodes unguisins H and I [5]. Clinker analysis with the *ung* BGCs from *A. violaceofuscus* CBS 115571 and *A. campestris* IBT 28561 indicated a high level of homology (Figure 4). The biosynthesis of unguisins B and J therefore is proposed to arise from this single BGC, similar to the biosynthesis

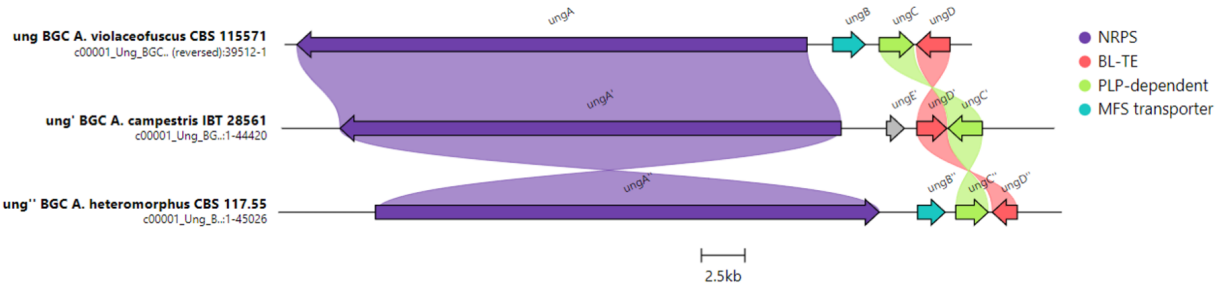



**Figure 4:** Clinker analysis of identified unguisin-encoding BGCs. UngE' is a methyltransferase that methylates phenylalanine and appears only in the *A. campestris* BGC.

of unguisins A and B in *A. violaceofuscus* CBS 115571 (Scheme 1).

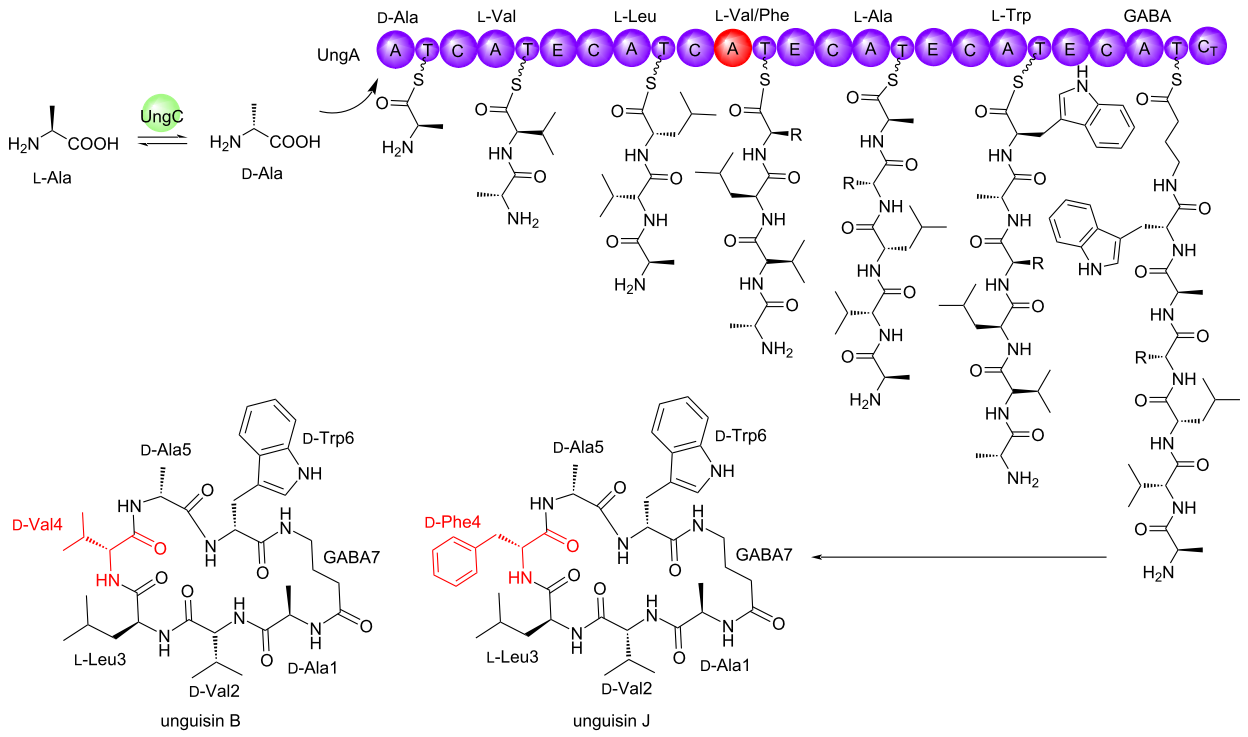
Within the unguisin family, there is variability in the amino acids incorporated at positions 2–6 (Figure 1), however, there are usually only one or two residue differences between molecules that are co-isolated from each source, e.g., A and B from *A. violaceofuscus* CBS 115571 [5]; A, B, and C from *Emericella unguis* [1]; A, E, F and G from *Aspergillus candidus* NF2412 [4]; H and I from *A. campestris* IBT 28561 [5]; and B and J from *A. heteromorphus* CBS 117.55 (Figure 1). This implies that only one or two modules per NRPS possesses a noticeable level of relaxed substrate specificity. To explore this observation, the A and C domains were identified in UngA, UngA' and UngA'' and phylogenetic analysis of the A and C domains was performed (Figure 5 and Figure 6).

The A domains do not clade according to substrate specificity – instead they clade according to which module they were extracted from. The A domains from modules 2, 3, and 4, which have relaxed substrate specificity, do appear to have evolved differently than A domains from modules 1, 5, 6, and 7 (Figure 5). Perhaps unsurprisingly the domains from UngA and UngA'' which both synthesize unguisin B, were more closely related than those from UngA' despite differences in substrate specificity in modules 3 and 4. Previously Matsuda et al. had compared the putative non-ribosomal codes for the UngA and UngA' A domains and also observed that conventional approaches are inadequate to understand or predict the specificity of fungal A domains [5].

The clades formed by the C domains showed higher divergence than the A domains with the $C_T$ domains forming their own branch and C domain from modules 1 and 3 clearly distinct to those from modules 2, 4, 5, and 6 (Figure 6). This separation of the non-terminal C domains could be due to modules 1 and 3 lacking an E domain. Again, the domains from UngA and UngA'' were more closely related than those from UngA' regardless of which two amino acids were condensed.

## Conclusion

In this study unguisins B and J were isolated from *A. heteromorphus* CBS 117.55 which has not been extensively investigated for secondary metabolite production. A BGC encoding the unguisins was identified by genome mining with high homology to *ung* BGCs from other *Aspergillus* sp. Phylogenetic analysis of the A and C domains extracted from the UngA NRPS indicates that domains within modules are more closely related – even when substrate specificity differs – than domains within other modules that accept the same substrates.




**Scheme 1:** Proposed biosynthesis of unguisins B and J in *A. heteromorphus* CBS 117.55.

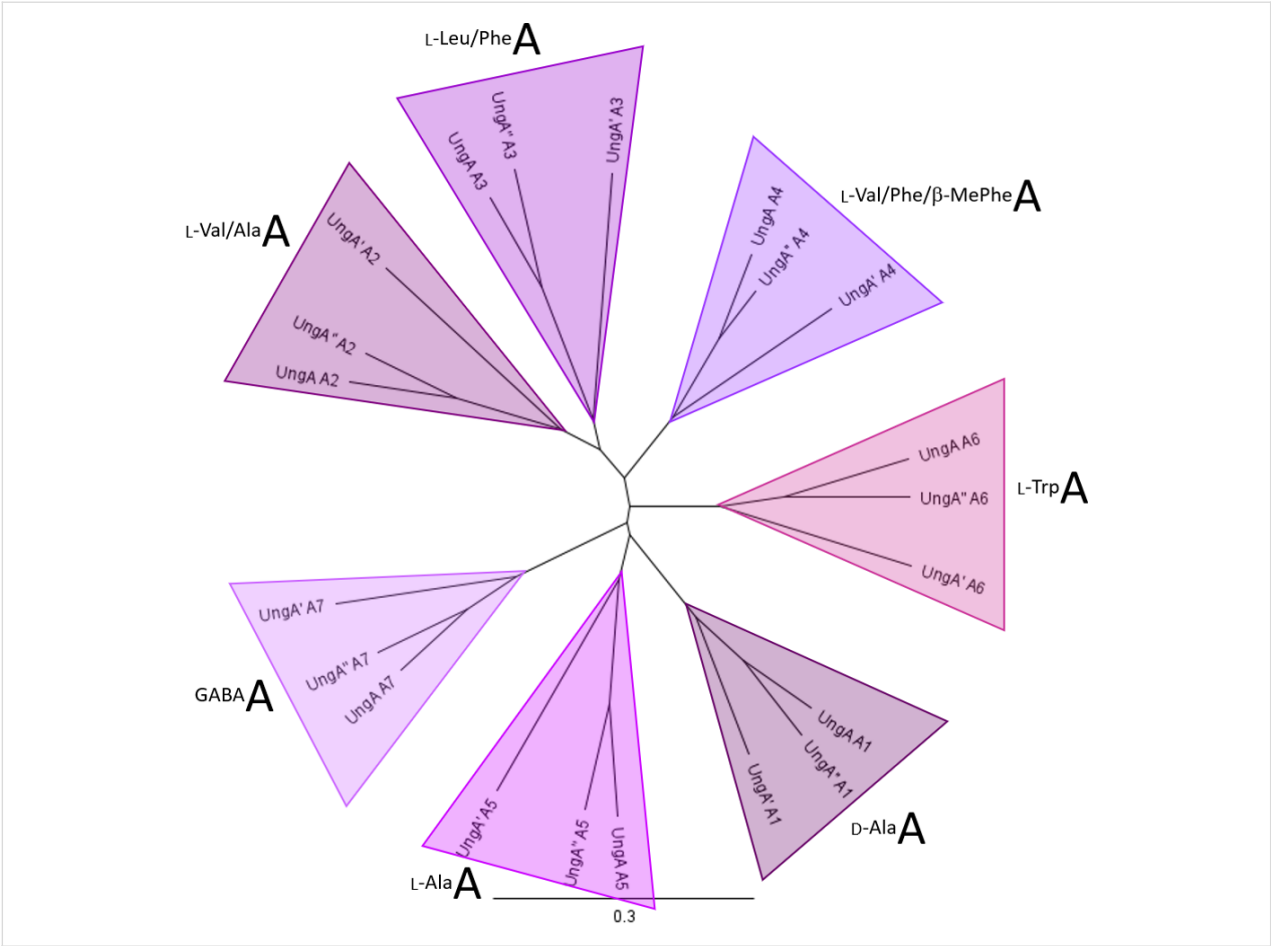



**Figure 5:** Phylogenetic analysis of A domains extracted from UngA NRPS. The substrate of the A domain is indicated for each clade.

# Experimental

## General experimental procedures

*A. heteromorphus* CBS 117.55 (also known as A. heteromorphus NRRL 4747) was purchased from the ARS Culture Collection. Rice solid medium was purchased from RiceSelect Organic Texmati. All solvents used for conducting LC analysis were purchased from Fisher Scientific. DMSO-$d_6$ NMR solvent was purchased from Sigma-Aldrich. $N^{\alpha}$-(5-Fluoro-2,4-dinitrophenyl)-D-leucinamide was purchased from TCI Chemicals. Authentic amino acid samples were purchased from Thermo Scientific. Sodium bicarbonate purchased from Fisher Chemical.

1D and 2D NMR experiments were recorded on a Varian INOVA 500 instrument ($^1$H: 500 MHz; $^{13}$C: 125 MHz). The chemical shifts (δ) were expressed in ppm and recorded with reference to solvent signals ($^1$H NMR: DMSO-$d_6$ 2.50 ppm; $^{13}$C NMR: DMSO-$d_6$ 39.5 ppm). Optical rotation was measured on a Jasco P-2000 polarimeter with a path length of 100 mm. Analytical HPLC-PDA-MS system was a Shimadzu instrument (LC2030C 3D Plus Prominence) coupled to a Shimadzu LCMS-2020 mass spectrometer. Analyses were performed using a Phenomenex Kinetex $RP_{18}$ column (100 mm × 4.6 mm i.d., 2.6 µm) along with the Security Guard $RP_{18}$ protective guard column (4.6 mm i.d.) and eluting with $H_2O$ + 0.1% formic acid and MeCN + 0.1% formic acid using a gradient from 90:10 to 10:90 of $H_2O$/MeCN over 15 min, maintaining in 10:90 $H_2O$/MeCN for 3 min, from 10:90 to 90:10 in 1 min, and maintaining at 90:10 for 1 min, using a flow rate of 1.0 mL/min. The PDA detector scanned between λ = 190 and 700 nm. The MS was optimized using the following conditions: interface voltage 4.5 kV; interface temperature 350 °C; DL temperature 250 °C; heat block 200 °C; ESI mode, acquisition range 100 to 1000 Da; nebulizing gas 1.5 L min$^{-1}$; drying gas flow 15 L min$^{-1}$. The fractionation of the sample was performed on a Shimadzu LC-20AP preparative liquid chromatograph (SCL-40 System Controler Deliver and LH-40 Liquid Handler) coupled to a Shimadzu SPD-M40 Photo Diode Array Detector (PDA) system using a RP-18 column (Phenomenex, Kinetex 250 × 30 mm i.d., 5 µm, flow rate of 18.0 mL min$^{-1}$). The purification of compounds was performed on a Shimadzu LC-20AD liquid chromatography (CBM-20A Communication

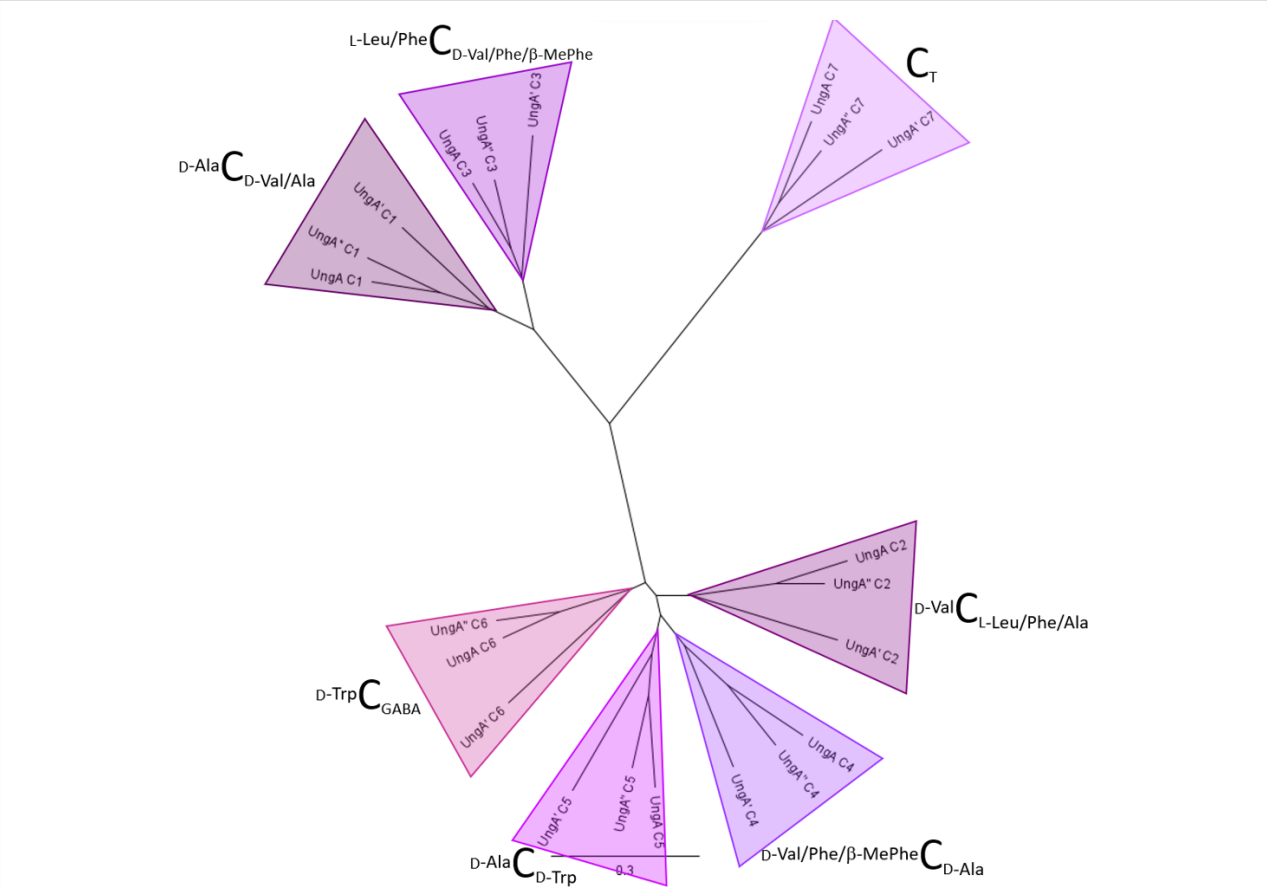



**Figure 6:** Phylogenetic analysis of C domains extracted from UngA NRPS. The substrates condensed by each C domain is indicated.

Bus Module, CTO-20A column oven, DGU-20A Degassing Unit and SIL-20A AutoSampler) coupled to a Shimadzu SPD-20A UV–vis Detector system using a RP-18 column (Shimadzu, Premier 250 × 10 mm i.d., 5 µm, flow rate of 3.0 mL $min^{-1}$). High-resolution mass spectra were recorded on an ABSciex TripleTOF 6600+ mass spectrometer. Direct infusion of compounds **1** and **2** through the high-resolution mass spectrometry (HRMS) was performed using a flow rate of 10 µL $min^{-1}$ which the samples were diluted at 10 ppm with a solution of MeCN/$H_2O$ (50:50; v/v) containing 0.1% formic acid. The parameters such as declustering and entrance potentials remained constant for MS and MS/MS were set up at 150 V and 10 V, respectively. Collision energy for MS and $MS^2$ scan surveys was 10 V and 45 V, respectively, with a collision energy spread of 12 V for $MS^2$ scan survey. Precursor ion was impacted with three different collision energies (33, 45, 57 V), and the resulting $MS^2$ spectra were combined into one final $MS^2$ spectrum. The mass spectra were acquired using Turbo Spray Ionization set to 5.5 kV in positive ion mode with an accumulation time of 100 ms. The mass ranges for MS and $MS^2$ scan surveys were 500–800 amu and 30–800 amu, respectively. The curtain gas (nitrogen), nebulizing and heating gas were fixed at 25 psi, 20 psi and 15 psi, respectively. The temperature of the source was 25 °C. MS spectra were acquired and processed using Analyst TF 1.8.1 software.

## Fungal growth and extraction

*A. heteromorphus* CBS 117.55 was cultivated in 2 Erlenmeyer flasks (500 mL), each containing 90 g of rice and 150 mL of $H_2O$ [16]. The medium was autoclaved at 121 °C for 20 min. After sterilization, the medium was inoculated with the spore solution of *A. heteromorphus* (1 mL) and incubated in static mode at 25 °C for 21 days. The following day, the cultured mass in the flasks was ground and extracted with ethyl acetate (EtOAc, 3 × 100 mL). The EtOAc fraction was dried using a rotary evaporator and then dissolved in $CH_3CN$ for defatting with hexane by partitioning. The $CH_3CN$ fraction was evaporated, yielding 0.601 g of soluble-organic extract.

## Fractionation and isolation of unguisins J and B

The soluble-organic extract was fractionated by preparative HPLC-PDA using Kinetex RP18 column (250 mm × 30 mm i.d., 5 µm) and UV detector at $\lambda_{max}$ = 254 nm. The mobile

phase consisted of $H_2O$ + 0.05% formic acid (eluent A) and MeCN + 0.05% formic acid (eluent B), which was eluted of 20–100% of B with flow rate of 18 mL $min^{-1}$, yielding 20 fractions.

Fractions Fr13 and Fr15 were subjected to a purification by semipreparative HPLC-UV using a Premier $RP_{18}$ column (250 mm × 10 mm i.d., 5 µm) and UV detector at $\lambda_{max}$ = 210 nm. The gradient elution consisted from 65:35 to 35:65 of $H_2O$/MeCN over 20 min, using a flow rate of 3.0 mL $min^{-1}$. Fractions Fr13 and Fr15 resulted in the isolation of **1** (11.5 mg) and **2** (14.1 mg), respectively.

**Unguisin J (1).** Amorphous white powder, $[\alpha]_D^{22}$ +23.4 (*c* 0.1, MeOH); UV (photodiode array, MeCN/$H_2O$) $\lambda_{max}$ = 219 and 279 nm; HRESIMS *m/z*: $[M + H]^+$ calcd for $C_{41}H_{57}N_8O_7^+$, 773.4345; found, 773.4338, $[M + Na]^+$ calcd for $C_{41}H_{56}N_8O_7Na^+$, 795.4164; found, 795.4162. For the $^1H$ and $^{13}C$ NMR spectroscopic data, see Table 1.

### Bioinformatics

The *A. heteromorphus* CBS 117.55 genome was initially screened using fungiSMASH to identify scaffolds/contigs encoding secondary metabolites. Scaffold MSFL01000005.1 was further investigated using FGENESH [17] to further refine gene boundaries, introns, and resulting protein sequence (Table S1, Supporting Information File 1). Comparative genomics/Clinker analysis was performed using Cagecat [18,19]. A and C domains were identified via Scan Prosite [20] and fungiSMASH [21]. Sequence alignments (Figures S1 and S2, Supporting Information File 1) and phylogenetic trees were generated using Geneious™. Functional domain motifs were visualized using Web Logo [22] and compared to literature motifs (Figures S3 and S4, Supporting Information File 1).

### Marfey's method to determine the absolute configuration of **1** and **2**

Samples **1** and **2** (0.3 mg each) were hydrolyzed in 0.5 mL of 6 N HCl at 115 °C for 20 h. After cooling, the reaction mixture was evaporated under nitrogen gas flow. The residue was dissolved in 0.5 mL of $H_2O$, and dried using speedvac to remove the residual HCl. The hydrolysates of **1** and **2** as well as authentic amino acid samples were treated with 200 µL of 1 N sodium bicarbonate solution and 100 µL of 1% Marfey's reagent ($N^{\alpha}$-(5-fluoro-2,4-dinitrophenyl)-D-leucinamide) in acetone [14]. The samples were incubated for 1 h at 60 °C followed by neutralization with 100 µL of 2 N HCl. The HCl was removed by using $N_2$ gas. Then, the samples were diluted by 2 mL of MeCN/$H_2O$ (1:1, v/v) solution, filtered with 0.22 µm filter and analyzed by LC–MS for comparison of the retention times.

## Supporting Information

### Supporting Information File 1

Spectroscopic and spectrometric data of **1** and **2**. Bioinformatic data of the biosynthetic gene clusters.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-32-S1.pdf]

## Acknowledgements

We would like to thank Dr. Hongjun Pan for his kindness and teachings in the acquisition of NMR spectra as well as Jean Christophe Cocuron from the Bioanalytical Facilities (BAF) for HRMS data.

## Funding

This project was supported by UNT Department of Chemistry and BioDiscovery Institute start-up funds, the National Science Foundation (NSF; award number 2048347) and the São Paulo Research Foundation (FAPESP) via a post-doctoral scholarship awarded to M.R.A. (2022/15570-6).

## Author Contributions

Sharmila Neupane: conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft; writing – review & editing. Marcelo Rodrigues de Amorim: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; supervision; validation; visualization; writing – original draft; writing – review & editing. Elizabeth Skellam: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; visualization; writing – original draft; writing – review & editing.

## ORCID® iDs

Elizabeth Skellam - https://orcid.org/0000-0003-4087-0708

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2023.62.v1

## References

1. Malmstrøm, J. *J. Nat. Prod.* **1999,** *62,* 787–789. doi:10.1021/np980539z

2. Malmstrøm, J.; Ryager, A.; Anthoni, U.; Nielsen, P. H. *Phytochemistry* **2002,** *60,* 869–872. doi:10.1016/s0031-9422(02)00150-4
3. Liu, S.; Shen, Y. *Chem. Nat. Compd.* **2011,** *47,* 786–788. doi:10.1007/s10600-011-0059-2
4. Li, W.; Jiao, F.-W.; Wang, J.-Q.; Shi, J.; Wang, T.-T.; Khan, S.; Jiao, R.-H.; Tan, R.-X.; Ge, H.-M. *Tetrahedron Lett.* **2020,** *61,* 152322. doi:10.1016/j.tetlet.2020.152322
5. Wei, X.; Chan, T. K.; Kong, C. T. D.; Matsuda, Y. *J. Nat. Prod.* **2023,** *86,* 416–422. doi:10.1021/acs.jnatprod.2c01065
6. Anh, C. V.; Yoon, Y. D.; Kang, J. S.; Lee, H.-S.; Heo, C.-S.; Shin, H. J. *Mar. Drugs* **2022,** *20,* 217. doi:10.3390/md20030217
7. Wang, Y.; Glukhov, E.; He, Y.; Liu, Y.; Zhou, L.; Ma, X.; Hu, X.; Hong, P.; Gerwick, W. H.; Zhang, Y. *Antibiotics (Basel, Switz.)* **2022,** *11,* 513. doi:10.3390/antibiotics11040513
8. Yang, W.-C.; Bao, H.-Y.; Liu, Y.-Y.; Nie, Y.-Y.; Yang, J.-M.; Hong, P.-Z.; Zhang, Y. *Molecules* **2018,** *23,* 2245. doi:10.3390/molecules23092245
9. Akone, S. H.; Daletos, G.; Liu, W.; Proksch, P. *Z. Naturforsch., C: J. Biosci.* **2016,** *71,* 15–19. doi:10.1515/znc-2015-0137
10. Ariawan, A. D.; Webb, J. E. A.; Howe, E. N. W.; Gale, P. A.; Thordarson, P.; Hunter, L. *Org. Biomol. Chem.* **2017,** *15,* 2962–2967. doi:10.1039/c7ob00316a
11. Süssmuth, R. D.; Mainz, A. *Angew. Chem., Int. Ed.* **2017,** *56,* 3770–3821. doi:10.1002/anie.201609079
12. Zhang, L.; Wang, C.; Chen, K.; Zhong, W.; Xu, Y.; Molnár, I. *Nat. Prod. Rep.* **2023,** *40,* 62–88. doi:10.1039/d2np00036a
13. Steiniger, C.; Hoffmann, S.; Süssmuth, R. D. *Cell Chem. Biol.* **2019,** *26,* 1526–1534.e2. doi:10.1016/j.chembiol.2019.08.005
14. Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K.-i. *Anal. Chem. (Washington, DC, U. S.)* **1997,** *69,* 5146–5151. doi:10.1021/ac970289b
15. Vesth, T. C.; Nybo, J. L.; Theobald, S.; Frisvad, J. C.; Larsen, T. O.; Nielsen, K. F.; Hoof, J. B.; Brandl, J.; Salamov, A.; Riley, R.; Gladden, J. M.; Phatale, P.; Nielsen, M. T.; Lyhne, E. K.; Kogle, M. E.; Strasser, K.; McDonnell, E.; Barry, K.; Clum, A.; Chen, C.; LaButti, K.; Haridas, S.; Nolan, M.; Sandor, L.; Kuo, A.; Lipzen, A.; Hainaut, M.; Drula, E.; Tsang, A.; Magnuson, J. K.; Henrissat, B.; Wiebenga, A.; Simmons, B. A.; Mäkelä, M. R.; de Vries, R. P.; Grigoriev, I. V.; Mortensen, U. H.; Baker, S. E.; Andersen, M. R. *Nat. Genet.* **2018,** *50,* 1688–1695. doi:10.1038/s41588-018-0246-1
16. de Amorim, M. R.; Hilário, F.; dos Santos, F. M., Junior.; Batista, J. M., Junior.; Bauab, T. M.; Araújo, A. R.; Carlos, I. Z.; Vilegas, W.; dos Santos, L. C. *Planta Med.* **2019,** *85,* 957–964. doi:10.1055/a-0853-7793
17. Solovyev, V. Statistical approaches in Eukaryotic gene prediction. In *Handbook of Statistical genetics,* 3rd ed.; Balding, D.; Cannings, C.; Bishop, M., Eds.; John Wiley & Sons: Chichester, UK, 2007; pp 97–159. doi:10.1002/9780470061619.ch4
18. Gilchrist, C. L. M.; Chooi, Y.-H. *Bioinformatics* **2021,** *37,* 2473–2475. doi:10.1093/bioinformatics/btab007
19. Implemented tools. https://cagecat.bioinformatics.nl/tools/ (accessed Jan 28, 2024).
20. de Castro, E.; Sigrist, C. J. A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P. S.; Gasteiger, E.; Bairoch, A.; Hulo, N. *Nucleic Acids Res.* **2006,** *34,* W362–W365. doi:10.1093/nar/gkl124
21. 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.; van Wezel, G. P.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2023,** *51,* W46–W50. doi:10.1093/nar/gkad344
22. Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. *Genome Res.* **2004,** *14,* 1188–1190. doi:10.1101/gr.849004

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.32

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Development of a chemical scaffold for inhibiting nonribosomal peptide synthetases in live bacterial cells

Fumihiro Ishikawa[*,‡1], Sho Konno[‡2], Hideaki Kakeya[2] and Genzoh Tanabe[*1]

Full Research Paper

Open Access

Address:
[1]Faculty of Pharmacy, Kindai University, 3-4-1 Kowakae, Higashi-Osaka, Osaka 577-8502, Japan and [2]Department of System Chemotherapy and Molecular Sciences, Division of Medicinal Frontier Sciences, Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo, Kyoto 606-8501, Japan

Email:
Fumihiro Ishikawa[*] - ishikawa@phar.kindai.ac.jp; Genzoh Tanabe[*] - g-tanabe@phar.kindai.ac.jp

* Corresponding author ‡ Equal contributors

Keywords:
adenylation domain inhibitor; gramicidin S synthetase; natural product; nonribosomal peptide; nonribosomal peptide synthetase

*Beilstein J. Org. Chem.* **2024,** *20,* 445–451.
https://doi.org/10.3762/bjoc.20.39

Received: 14 December 2023
Accepted: 12 February 2024
Published: 26 February 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf

© 2024 Ishikawa et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

The adenylation (A) domain is essential for non-ribosomal peptide synthetases (NRPSs), which synthesize various peptide-based natural products, including virulence factors, such as siderophores and genotoxins. Hence, the inhibition of A-domains could attenuate the virulence of pathogens. 5'-*O*-*N*-(Aminoacyl or arylacyl)sulfamoyladenosine (AA-AMS) is a bisubstrate small-molecule inhibitor of the A-domains of NRPSs. However, the bacterial cell permeability of AA-AMS is typically a problem owing to its high hydrophilicity. In this study, we investigated the influence of a modification of 2′-OH in the AMS scaffold with different functional groups on binding to target enzymes and bacterial cell penetration. The inhibitor **7** with a cyanomethyl group at 2′-OH showed desirable inhibitory activity against both recombinant and intracellular gramicidin S synthetase A (GrsA) in the gramicidin S-producer *Aneurinibacillus migulanus* ATCC 9999, providing an alternative scaffold to develop novel A-domain inhibitors.

## Introduction

Nonribosomal peptides (NRPs) exhibit various biological activities and have been used as therapeutic agents, such as antibiotics, anticancer agents, and immunosuppressants [1]. Additionally, NRPs function as virulence factors, such as siderophores and genotoxins [2]. Therefore, inhibiting their biosynthesis by using small molecules can help to elucidate their natural functions and their potential as therapeutic targets. NRPs are synthesized by large, versatile, and multifunctional proteins called nonribosomal peptide synthetases (NRPSs), which are composed of multiple modules and subdivided domains (Figure 1) [3]. The adenylation (A) domain in NRPSs is responsible for the selection and activation of amino acids, hydroxy acids, and aryl acids upon ATP consumption (Figure 2a) [4]. The activated aminoacyladenosine monophosphate (AMP) is transferred to the thiol

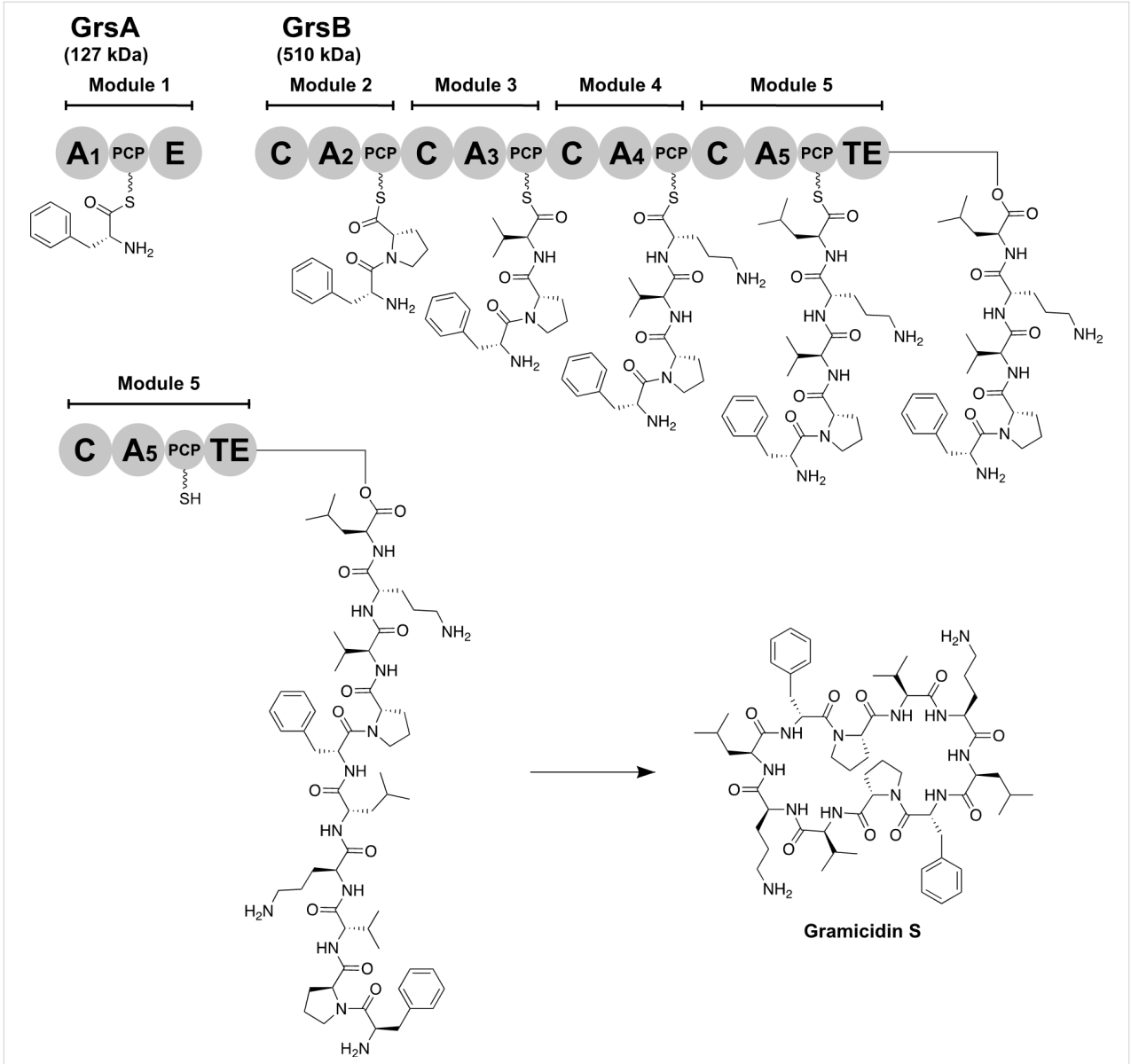



**Figure 1:** Biosynthesis of gramicidin S. Modules comprise the PCP, A, E, C, and TE domains. PCP, peptidyl carrier protein; A1, L-Phe-specific A-domain; A2, L-Pro-specific A-domain; A3, L-Val-specific A-domain; A4, L-Orn-specific A-domain; A5, L-Leu-specific A-domain; E, epimerization domain; C, condensation domain; TE, thioesterase domain.

group of a phosphopantetheine prosthetic arm in an adjacent peptidyl carrier protein (PCP). The amino acid loaded on the PCP then undergoes coupling with the amino acid loaded on the downstream PCP in the condensation (C) domain. Finally, the linear peptide on the PCP in the last module is either hydrolyzed or cyclized by a thioesterase (TE) domain, consequently resulting in the formation of the final products.

Inhibitors that target each domain of NRPSs are valuable for elucidating the biosynthetic pathways associated with bioactive NRPs and for developing antibiotic molecules. Burkart et al. reported a systematic strategy for inhibiting modular synthases [5]. They used the inhibitors of individual domains to investigate the biosynthetic pathway of blue pigment synthetase A, which produces the blue pigment indigoidine, and demonstrated that their results complement the proposed biosynthetic pathway. Furthermore, among the catalytic domain inhibitors of NRPSs, A-domain inhibitors have been widely developed as potential therapeutic agents for treating infectious diseases.

Aryl acid A-domains are involved in the synthesis of several bacterial siderophores such as vibriobactin from *Vibrio cholera*, yersiniabactin from *Yersinia pestis*, and mycobactin from *Mycobacterium tuberculosis* [6]. 5′-*O*-Sulfamoyladenosine

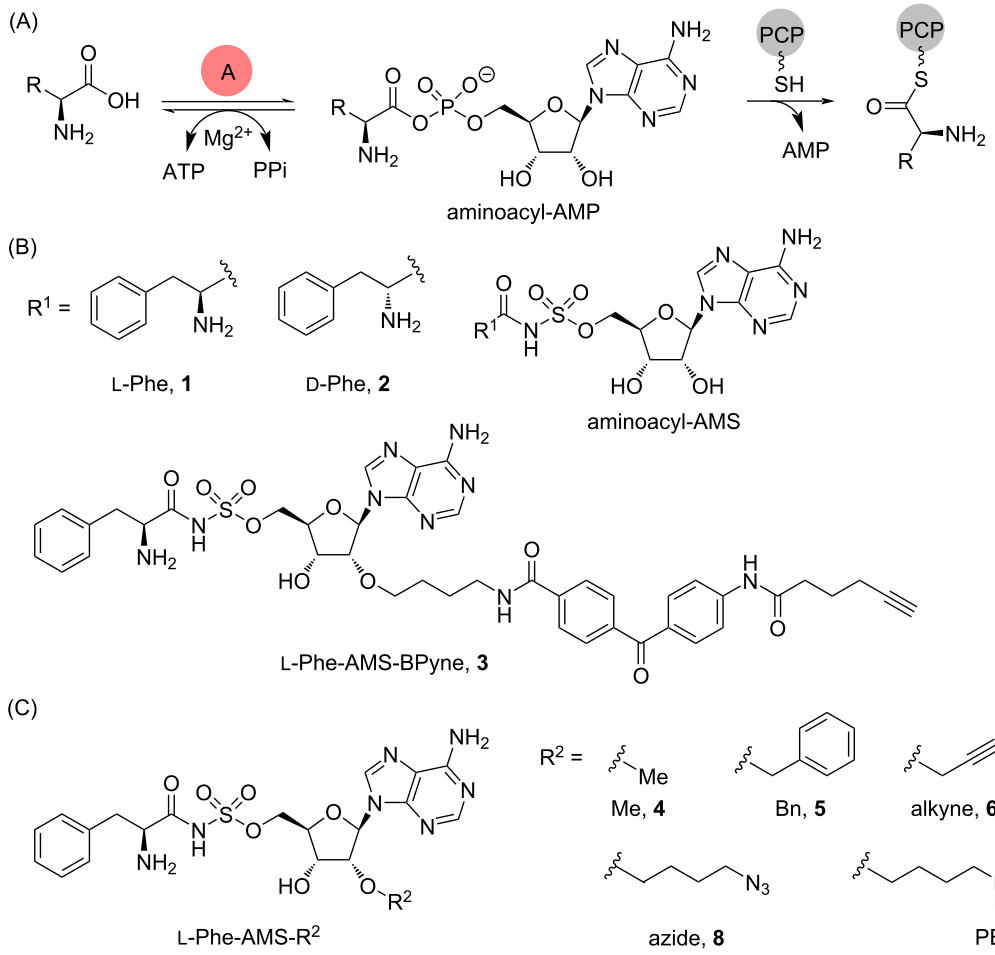



**Figure 2:** (A) Adenylation reaction in a nonribosomal peptide synthetase. (B) Structures of aminoacyl-AMS inhibitors and the L-Phe-AMS-BPyne probe. (C) Structures of 2′-OH-substituted L-Phe-AMS derivatives for A-domains. PCP, peptidyl carrier protein; AMS, 5′-*O*-sulfamoyladenosine.

(AMS), a bioisosteric analog of an AMP intermediate, has been used as a non-hydrolysable scaffold for developing A-domain inhibitors. Moreover, 5′-*O*-[*N*-(salicyl)sulfamoyl]adenosine (Sal-AMS) and its derivatives show potent inhibitory activities against the A-domain of MbtA, a component of mycobactin synthetase and antimicrobial activities against *M. tuberculosis* [7]. In addition, aminoacyl (AA)-AMS has been designed to inhibit the amino acid-activating A-domains in NRPSs and has been found to be a tight-binding inhibitor (Figure 2b) [8]. Moreover, the intracellular concentrations of a series of AMS derivatives in *Escherichia coli*, *Bacillus subtilis*, and *Mycobacterium smegmatis* have been investigated by Tan et al., demonstrating non-obvious correlations between the chemical structure and permeability among various bacteria, owing to the differences in the composition of cell membranes and presence of efflux pumps [9].

Specific protein labeling using a chemical probe can help to identify, characterize, and visualize target proteins [10,11]. The first chemical probe used for A-domains in NRPSs was reported by Marahiel et al. [8]. They introduced a pegylated biotin linker at the 2′-OH group of L-Phe-AMS and confirmed that the probe retains the binding activities toward the A-domain of GrsA, a gramicidin S synthetase. Aldrich et al. developed a Sal-AMS-based activity-based probe (ABP) to profile MbtA in *M. tuberculosis* [12]. In contrast, we previously described an activity-based protein profiling (ABPP) strategy for NRPSs using ABPs that target A-domains (Figure 2b) [13-15]. The probes comprise an aminoacyl-AMS ligand and a photoaffinity group with clickable alkyne functionality appended to the 2′-OH group of adenosine. A complex structure of the GrsA A-domain with L-Phe and AMP revealed that the 2′-OH of the adenosine skeleton is oriented toward the outside of the active site of the GrsA A-domain, suggesting that chemical modification at the 2′-OH group of the adenosine skeleton would be tolerated [16] (Figure 2b). Moreover, these probes (AA-AMS-BPyne) can selectively label the A-domains corresponding to the amino acid of the ligand in both recombinant enzymes and proteomes. We

recently reported that these probes can be used to label the A-domains of endogenous NRPSs in live bacterial cells [17-19]. The intracellular labeling of the enterobactin synthetase EntF with Sal-AMS-BPyne requires carbonyl cyanide *m*-chlorophenylhydrazone, which collapses the proton motive force used in most efflux pumps [17]. Under the same conditions, the competitive inhibition of labeling using excess Sal-AMS is not observed, suggesting that the modification at the 2′-OH group of the adenosine in AA-AMS might affect the cell permeability of the compounds. Given the high hydrophilicity of the bisubstrate AMS scaffold, it is reasonable to conclude that the BPyne component increases hydrophobicity and facilitates accumulation in live *E. coli*. In the present study, we investigated the influence of the introduction of several functional groups at the 2′-OH group of the AMS scaffold on both the binding affinities for A-domains and cell permeability (Figure 2c and Figure 3). We selected the L-Phe-selective A-domain of the gramicidin S synthetase GrsA, which was previously demonstrated to be selectively labeled with L-Phe-AMS-BPyne (**3**) in *Aneurinibacillus migulanus* ATCC 9999 [19]. Intracellular competitive ABPP of GrsA using L-Phe-AMS-BPyne was also performed to reveal the cell permeability of 2′-OH-modified L-Phe-AMS derivatives.

## Results and Discussion

To investigate the influence of the introduction of different alkyl groups at the 2′-OH, we prepared L-Phe-AMS derivatives **4**–**9**. As the compounds **6**, **8**, and **9** were synthesized previously [20], we designed and synthesized three new L-Phe-AMS derivatives containing methyl (**4**), benzyl (**5**), and cyanomethyl (**7**) groups at the 2′-OH. The synthetic routes to compounds **4**–**9** are shown in Scheme 1. The 2′-OH of adenosine was alkylated with several alkyl halides in the presence of sodium hydride (NaH). Both the 3′-OH and 5′-OH groups of compounds **10a**–**e** were protected by a TBS group, followed by the selective deprotection of the 5′-OH group using 25% trifluoroacetic acid in tetrahydrofuran. Subsequently, the 5′-OH group of compounds **12a**–**e** were reacted with sulfamoyl chloride in the presence of NaH. Compounds **13a**–**e** were coupled to pre-activated Boc-L-Phe-OSu in the presence of $Cs_2CO_3$. Removal of the Boc and TBS groups of compounds **14a**–**e** yielded the desired L-Phe-AMS derivatives **4**, **5**, and **7**.

We first determined the binding affinities of L-Phe-AMS derivatives **4**–**9** for the A-domain of GrsA using a previously developed competitive enzyme-linked immunosorbent assay technique for A-domains in NRPSs [14], which allowed us to measure the dissociation constant ($K_d$) of the test compounds (Figure S1a, Supporting Information File 1). The L-Phe-AMS-biotin probe was immobilized on a streptavidin-coated 96-well plate and incubated with recombinant $His_6$-tagged GrsA in the presence or absence of inhibitors (Figure S1b, Supporting Information File 1). After washing, the wells were incubated with an anti-$His_6$ tag antibody and subsequently with a horseradish peroxidase-conjugated secondary antibody. The amount of GrsA bound to L-Phe-AMS-biotin was determined by measuring the absorbance of *o*-phenylenediamine dihydrochloride at 492 nm. The $K_d$ values of the L-Phe-AMS derivatives are listed in Table 1. Compared with the binding affinity of L-Phe-AMS **1** ($K_d$ value, 11.4 ± 3.4 nM) for the A-domain of GrsA, all tested compounds showed slightly decreased binding affinities. Among them, inhibitors **4** (23.9 ± 0.7 nM), **7** (16.6 ± 0.6 nM), and **8** (30.2 ± 4.2 nM) displayed $K_d$ values comparable to that of L-Phe-AMS **1**, suggesting that the substitution of small functional groups such as methyl and cyanomethyl groups has a minor effect on the $K_d$ values (Table 1 and Figures S2 and S3, Supporting Information File 1). In contrast, inhibitors **6** and **9** exhibited lower affinity than inhibitors **1**, **2**, **4**, **5**, and **7**. This result suggests that the incorporation of the pegylated functionality at the 2′-OH group of the adenosine skeleton particularly perturbs binding via steric hindrance.

Next, we investigated whether L-Phe-AMS derivatives could competitively inhibit the labeling of recombinant GrsA using

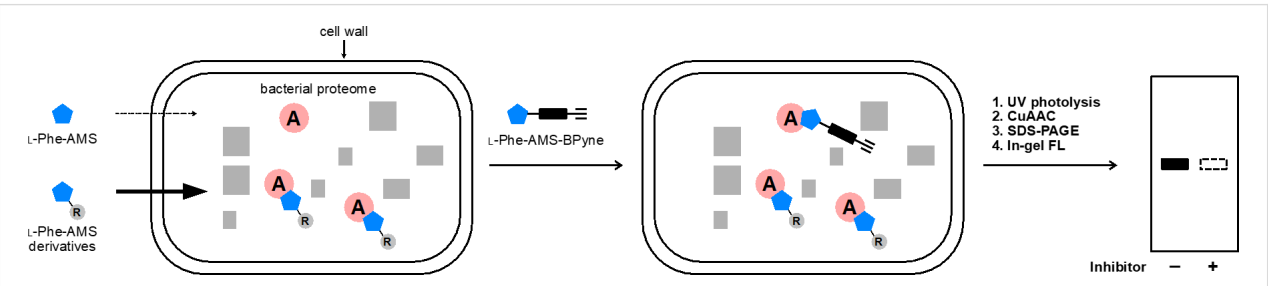
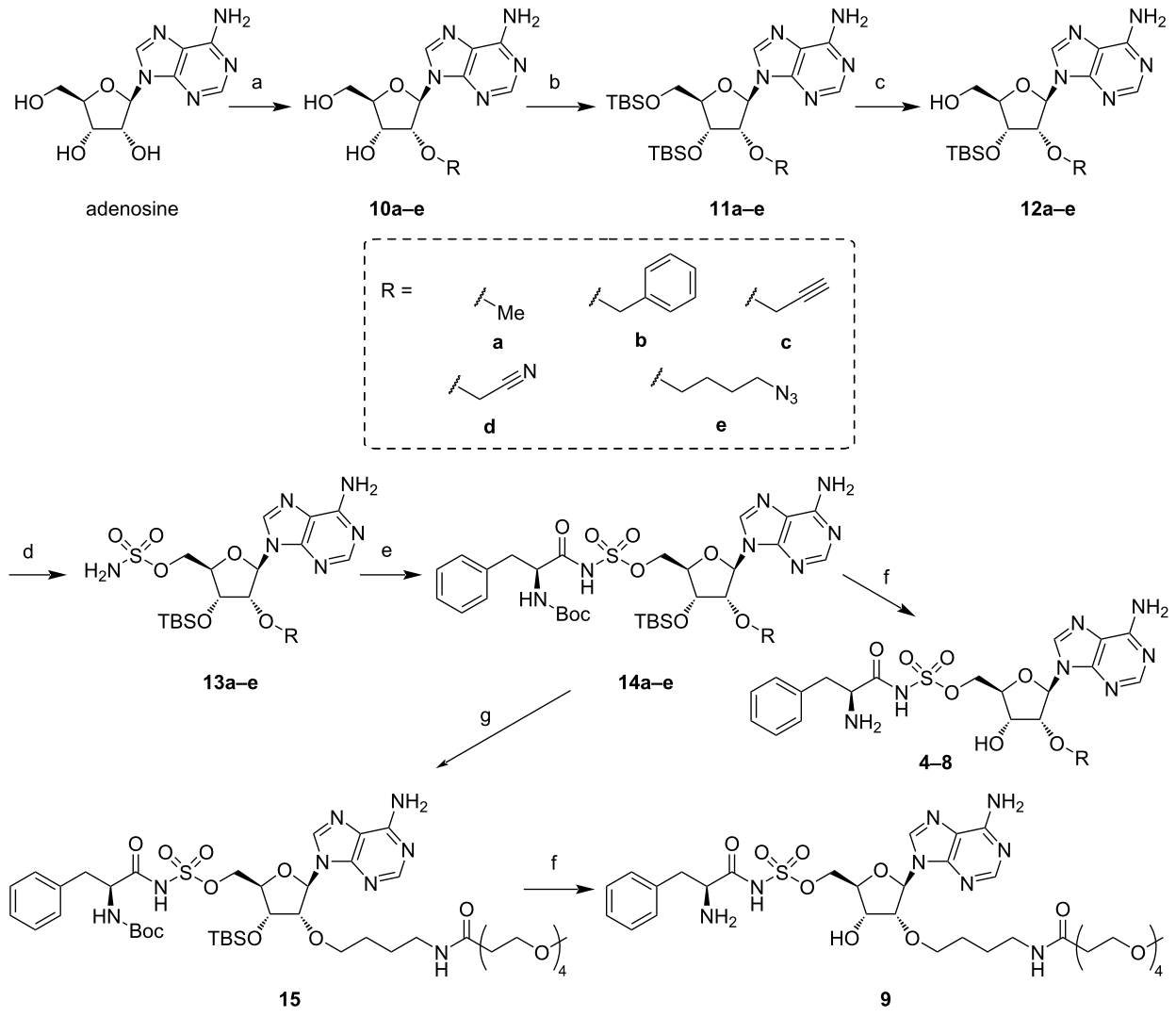



**Figure 3:** Illustration of the inhibition of A-domains by aminoacyl-AMS derivatives in live bacterial cells. Cell-permeable L-Phe-AMS derivatives can penetrate cells and interact with A-domains in live bacterial cells, resulting in the competitive inhibition of the labeling by L-Phe-AMS-BPyne. The substituent group (R; gray) of L-Phe-AMS derivatives could facilitate their cell penetration. After UV irradiation (365 nm), the labeled proteins are treated with a TAMRA-$N_3$ under copper(I)-catalyzed azide–alkyne cycloaddition conditions, followed by SDS-PAGE coupled with in-gel fluorescence scanning. AMS, 5′-*O*-sulfamoyladenosine.




**Scheme 1:** Synthesis of 2′-OH-substituted L-Phe-AMS derivatives. Reagents and conditions: (a) NaH, TBAI, R–X (a: MeI, b: BnBr, c: $BrCH_2CCH_3$, d: $BrCH_2CN$, e: $Br(CH_2)_4N_3$), DMF, rt, 45% (**10a**), 59% (**10b**), 45% (**10c**), 21% (**10d**), and 45% (**10e**); (b) TBSCl, imidazole, DMAP, DMF, rt, 83% (**11a**), 78% (**11b**), 57% (**11c**), 62% (**11d**), and 65% (**11e**); (c) 25% TFA in THF aq, 0 °C, 77% (**12a**), 77% (**12b**), 50% (**12c**), 96% (**12d**), and 75% (**12e**); (d) NaH, $NH_2SO_2Cl$, DME, 0 °C to rt, 86% (**13a**), 81% (**13b**), 92% (**13c**), 78% (**13d**), and 62% (**13e**); (e) Boc-L-Phe-OSu, $Cs_2CO_3$, DMF, rt.; (f) 80% aqueous TFA, rt, two steps 95% (**4**), 94% (**5**), 56% (**6**), 88% (**7**), 58% (**8**), 69% (**9**); (g) methyl-$Peg_4$-NHS ester, $Cs_2CO_3$, DMF, rt.

the ABP L-Phe-AMS-BPyne (**3**). GrsA (1 μM) was incubated with probe **3** (1 μM) in the absence or presence of either inhibitor **1**, **2**, or **4**–**9** (100 μM). The mixtures were exposed to ultraviolet light at 365 nm to form a covalent bond between GrsA and the benzophenone moiety in probe **3**. The samples were then reacted with TAMRA-$N_3$ (structure shown in Figure S4, Supporting Information File 1) under copper(I)-catalyzed azide–alkyne cycloaddition conditions [21] and subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis coupled with in-gel fluorescence scanning (excitation wavelength, 532 nm; emission wavelength, 580 nm). The labeling of GrsA by probe **3** was completely suppressed by the addition of inhibitors **1**, **2**, **4**, **5**, and **7** (Figure 4a). In contrast, inhibitors **6**, **8**, and **9** moderately inhibited labeling, which was consistent with the decreased $K_d$ values of these compounds. Subsequently, we conducted competitive labeling experiments for endogenous GrsA in the proteome of the gramicidin S-producer *A. miglanus* ATCC 9999. The cellular lysates of strain ATCC 9999 were treated with probe **3** (1 μM) in the absence or presence of inhibitor **1**, **2**, or **4**–**9** (100 μM), irradiated at 365 nm, and subjected to the click reaction with TAMRA-$N_3$. In-gel fluorescence scanning revealed that inhibitors **1**, **2**, **4**, and **7** completely suppressed GrsA labeling by probe **3** in the proteomic environment (Figure 4b). Unlike the results obtained for the labeling of purified recombinant GrsA, inhibitor **5** moderately inhibited labeling at 100 μM.

**Table 1:** $K_d$ values for the A-domain of gramicidin S synthetase[a].

| Compounds | $K_d$ [nM] |
|---|---|
| L-Phe-AMS[b] (**1**) | 11.4 ± 3.4 |
| D-Phe-AMS[b] (**2**) | 26.4 ± 4.1 |
| L-Phe-AMS-Me (**4**) | 23.9 ± 0.7 |
| L-Phe-AMS-Bn (**5**) | 33.6 ± 0.9 |
| L-Phe-AMS-alkyne[b] (**6**) | 60.0 ± 3.6 |
| L-Phe-AMS-CN (**7**) | 16.6 ± 0.6 |
| L-Phe-AMS-azide[b] (**8**) | 30.2 ± 4.2 |
| L-Phe-AMS-peg[b] (**9**) | 85.5 ± 10.7 |

[a]$K_d$ values were determined via competitive enzyme-linked immunosorbent assays using L-Phe-AMS-biotin at room temperature in 20 mM Tris (pH 8.0), 1 mM $MgCl_2$, 1 mM TCEP, and 0.0025% NP-40. [b]$K_d$ values were obtained from the study of Ishikawa et al. [20]. The assay conditions were identical to those of the present study.

Finally, we investigated whether the inhibitors could penetrate cells and inhibit intracellular GrsA labeling in *A. miglanus* ATCC 9999. In this study, *A. migulanus* ATCC 9999 cells were grown at 37 °C in YPG medium for 24 h, harvested, and washed with phosphate-buffered saline. The bacterial suspension was then treated with probe **3** (10 μM) in either the absence or presence of inhibitors **1**, **2**, or **4**–**9** (10 or 100 μM). Inhibitors **1**, **2**, **4**, **5**, **7**, and **8** completely inhibited the labeling of endogenous GrsA at high concentrations (100 μM), suggesting that these inhibitors can penetrate cells (Figure 4c). Notably, inhibitors **1** and **7** efficiently inhibited the labeling of GrsA at 10 μM. The incorporation of the nitrile group at the 2'-OH group of the adenosine skeleton is expected to provide chemical properties that would allow the compound to retain its binding affinity and cell permeability. Overall, these results indicate that a 2′-OH modification with a cyanomethyl group represents a useful AMS scaffold for intracellular NRPS inhibition.

## Conclusion

In this study, we investigated the effect of a 2′-OH modification in an AMS scaffold on the binding affinities for A-domains and cell permeability. Our experiments demonstrated that inhibitor **7**, harboring a cyanomethyl group at the 2′-OH, showed a $K_d$ value comparable to that of the original L-Phe-AMS **1**. In addition, intracellular competitive ABPP suggested that the inhibitor **7** can penetrate cells. The application of this new scaffold to

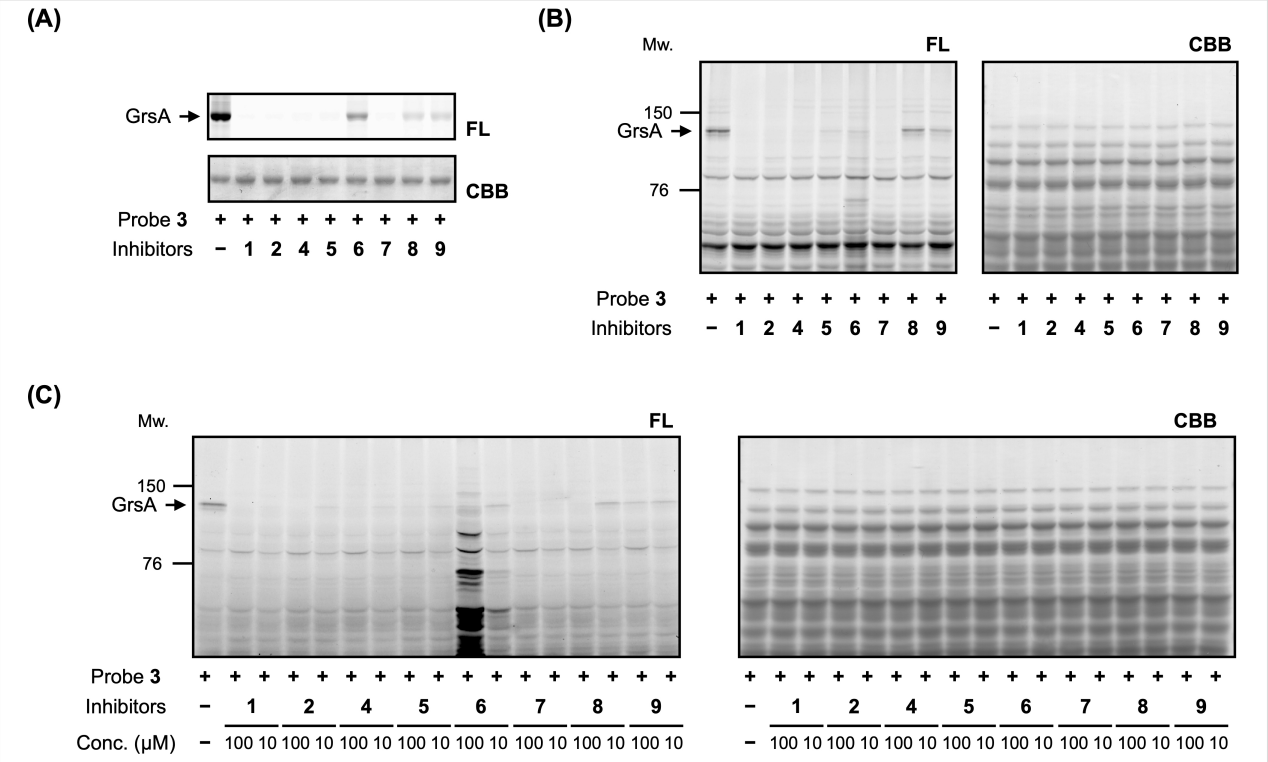



**Figure 4:** Competitive labeling experiments of GrsA using probe **3** in the presence of L-Phe-AMS inhibitors. (A) Labeling of recombinant GrsA with probe **3** in either the absence or presence of L-Phe-AMS inhibitors. Recombinant GrsA (1 μM) was pre-incubated with or without L-Phe-AMS inhibitors **1**, **2**, or **4**–**9** (100 μM) and then treated with probe **3** (1 μM). (B) Competitive labeling assay of endogenous GrsA with probe **3** using the lysates of *Aneurinibacillus migulanus* ATCC 9999. The lysates were pre-incubated with or without L-Phe-AMS inhibitors **1**, **2,** or **4**–**9** (100 μM) and then treated with probe **3** (1 μM). (C) Intracellular competitive labeling of GrsA in *A. migulanus* ATCC 9999. *A. migulanus* ATCC 9999 in YPG medium was treated with probe **3** (10 μM) in either the absence or presence of L-Phe-AMS inhibitors **1**, **2**, or **4**–**9** (10 or 100 μM). GrsA, gramicidin S synthetase; AMS, 5′-*O*-sulfamoyladenosine; FL, fluorescent gel; CBB; Coomassie brilliant blue.

NRPS inhibitors involved in the production of virulence factors could thus facilitate the development of new antibiotics.

## Supporting Information

### Supporting Information File 1

Additional Figures, experimental part and NMR spectra.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-39-S1.pdf]

## Funding

This study was supported by grants from the Institute Foundation for Science, Osaka (IFO) to F.I. This work was partially supported by the Japan Foundation for Applied Enzymology and the Research Foundation for Pharmaceutical Sciences to F.I. and a Grant-in Aid for Scientific Research on Innovative Areas (17H06401 to H.K.) and for Transformative Research Area (A) (23H04882 to H.K.) from the Ministry of Education Culture, Sports, Science, and Technology (MEXT), Japan. This work was also supported by the 2023 Kindai University Research Enhancement Grant (IP004). Finally, we are grateful for financial support provided by the Anti-aging Project for Private Universities.

## ORCID® iDs

Fumihiro Ishikawa - https://orcid.org/0000-0002-8681-9396
Sho Konno - https://orcid.org/0000-0002-5851-4130
Hideaki Kakeya - https://orcid.org/0000-0002-4293-7331
Genzoh Tanabe - https://orcid.org/0000-0002-7954-8874

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Walsh, C. T. *Science* **2004,** *303,* 1805–1810. doi:10.1126/science.1094318
2. Gulick, A. M. *Nat. Prod. Rep.* **2017,** *34,* 981–1009. doi:10.1039/c7np00029d
3. Süssmuth, R. D.; Mainz, A. *Angew. Chem., Int. Ed.* **2017,** *56,* 3770–3821. doi:10.1002/anie.201609079
4. Finking, R.; Marahiel, M. A. *Annu. Rev. Microbiol.* **2004,** *58,* 453–488. doi:10.1146/annurev.micro.58.030603.123615
5. Vickery, C. R.; McCulloch, I. P.; Sonnenschein, E. C.; Beld, J.; Noel, J. P.; Burkart, M. D. *Bioorg. Med. Chem. Lett.* **2020,** *30,* 126820. doi:10.1016/j.bmcl.2019.126820
6. Ferreras, J. A.; Ryu, J.-S.; Di Lello, F.; Tan, D. S.; Quadri, L. E. N. *Nat. Chem. Biol.* **2005,** *1,* 29–32. doi:10.1038/nchembio706
7. Neres, J.; Labello, N. P.; Somu, R. V.; Boshoff, H. I.; Wilson, D. J.; Vannada, J.; Chen, L.; Barry, C. E., III; Bennett, E. M.; Aldrich, C. C. *J. Med. Chem.* **2008,** *51,* 5349–5370. doi:10.1021/jm800567v
8. Finking, R.; Neumüller, A.; Solsbacher, J.; Konz, D.; Kretzschmar, G.; Schweitzer, M.; Krumm, T.; Marahiel, M. A. *ChemBioChem* **2003,** *4,* 903–906. doi:10.1002/cbic.200300666
9. Davis, T. D.; Gerry, C. J.; Tan, D. S. *ACS Chem. Biol.* **2014,** *9,* 2535–2544. doi:10.1021/cb5003015
10. Shiraiwa, K.; Cheng, R.; Nonaka, H.; Tamura, T.; Hamachi, I. *Cell Chem. Biol.* **2020,** *27,* 970–985. doi:10.1016/j.chembiol.2020.06.016
11. Deng, H.; Lei, Q.; Wu, Y.; He, Y.; Li, W. *Eur. J. Med. Chem.* **2020,** *191,* 112151. doi:10.1016/j.ejmech.2020.112151
12. Duckworth, B. P.; Wilson, D. J.; Nelson, K. M.; Boshoff, H. I.; Barry, C. E., III; Aldrich, C. C. *ACS Chem. Biol.* **2012,** *7,* 1653–1658. doi:10.1021/cb300112x
13. Konno, S.; Ishikawa, F.; Suzuki, T.; Dohmae, N.; Burkart, M. D.; Kakeya, H. *Chem. Commun.* **2015,** *51,* 2262–2265. doi:10.1039/c4cc09412c
14. Ishikawa, F.; Konno, S.; Suzuki, T.; Dohmae, N.; Kakeya, H. *ACS Chem. Biol.* **2015,** *10,* 1989–1997. doi:10.1021/acschembio.5b00097
15. Kasai, S.; Konno, S.; Ishikawa, F.; Kakeya, H. *Chem. Commun.* **2015,** *51,* 15764–15767. doi:10.1039/c5cc04953a
16. Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P. *EMBO J.* **1997,** *16,* 4174–4183. doi:10.1093/emboj/16.14.4174
17. Ishikawa, F.; Konno, S.; Takashima, K.; Kakeya, H.; Tanabe, G. *Org. Biomol. Chem.* **2021,** *19,* 8906–8911. doi:10.1039/d1ob01112j
18. Ishikawa, F.; Konno, S.; Uchida, C.; Suzuki, T.; Takashima, K.; Dohmae, N.; Kakeya, H.; Tanabe, G. *Cell Chem. Biol.* **2022,** *29,* 145–156.e8. doi:10.1016/j.chembiol.2021.05.014
19. Ishikawa, F.; Konno, S.; Uchiyama, Y.; Kakeya, H.; Tanabe, G. *Philos. Trans. R. Soc., B* **2023,** *378,* 20220026. doi:10.1098/rstb.2022.0026
20. Ishikawa, F.; Kakeya, H. *ChemBioChem* **2016,** *17,* 474–478. doi:10.1002/cbic.201500553
21. Speers, A. E.; Cravatt, B. F. *Chem. Biol.* **2004,** *11,* 535–546. doi:10.1016/j.chembiol.2004.03.012

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.39

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Pseudallenes A and B, new sulfur-containing ovalicin sesquiterpenoid derivatives with antimicrobial activity from the deep-sea cold seep sediment-derived fungus *Pseudallescheria boydii* CS-793

Zhen Ying[1,2], Xiao-Ming Li[1], Sui-Qun Yang[1], Hong-Lei Li[1], Xin Li[1], Bin-Gui Wang[*1,2,3,§] and Ling-Hong Meng[*1,2,¶]

Full Research Paper

Open Access

Address:
[1]CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Nanhai Road 7, Qingdao 266071, China, [2]University of Chinese Academy of Sciences, Yuquan Road 19A, Beijing 100049, China and [3]Laboratory of Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Wenhai Road 1, Qingdao 266237, China

Email:
Bin-Gui Wang[*] - wangbg@ms.qdio.ac.cn; Ling-Hong Meng[*] - menglh@ms.qdio.ac.cn

* Corresponding author
§ Phone: +86-532-82898553
¶ Phone: +86-532-82898890

Keywords:
antimicrobial activity; cold seep; *Pseudallescheria boydii*; sesquiterpenoid

*Beilstein J. Org. Chem.* **2024,** *20,* 470–478.
https://doi.org/10.3762/bjoc.20.42

Received: 02 December 2023
Accepted: 16 February 2024
Published: 28 February 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra

© 2024 Ying et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Pseudallenes A and B (**1** and **2**), the new and rare examples of sulfur-containing ovalicin derivatives, along with three known analogues **3**–**5**, were isolated and identified from the culture extract of *Pseudallescheria boydii* CS-793, a fungus obtained from the deep-sea cold seep sediments. Their structures were established by detailed interpretation of NMR spectroscopic and mass spectrometric data. X-ray crystallographic analysis confirmed and established the structures and absolute configurations of compounds **1**–**3**, thus providing the first characterized crystal structure of an ovalicin-type sesquiterpenoid. In the antimicrobial assays, compounds **1**–**3** showed broad-spectrum inhibitory activities against several plant pathogens with MIC values ranging from 2 to 16 μg/mL.

## Introduction

Marine cold seeps are typical of chemosynthetically driven ecosystems, characterized by methane-rich fluid emissions and unique sulfur oxidation–reduction reactions [1]. Due to the unique habitat, microorganisms surviving in the deep-sea cold seeps may serve as promising sources of secondary metabolites with functional and structural diversity [2]. In particular, two

indole diketopiperazine alkaloids containing an unprecedented spiro[bicyclo[2.2.2]octane-diketopiperazine] skeleton, chevalinulins A and B, and the first cytochalasin homodimer containing a thioether bridge, verruculoid A, were identified from deep-sea cold seeps-derived fungi and were described to have potent proangiogenic and antimicrobial activities [3,4].

As part of our continuing search for bioactive metabolites from deep-sea-derived fungi [3-6], the fungal strain *Pseudallescheria boydii* CS-793, which was obtained from sediments collected at the deep-sea cold seep area in the Northeast of the South China Sea, attracted our attention. Several meroterpenoids, alkaloids, polyketides, and sesquiterpenoids from the species displayed various biological properties including anti-inflammatory, antimicrobial, and cytotoxic activities [7-9]. In the present work, two rare new examples of sulfur-containing ovalicin sesquiterpenoids (**1**, **2**), together with three known related analogs (**3**–**5**) [10-13] have been isolated and identified from the bioactive fraction of *P. boydii* CS-793. Details of the isolation and purification, structure elucidation, and biological evaluation of compounds **1**–**5** are described herein.

## Results and Discussion

For chemical investigation, the solvent EtOAc was used to extract the fermentation culture of the fungus *P. boydii* CS-793 to afford an organic extract. Isolation and purification of the crude extract with a combination of column chromatography (CC) by Lobar LiChroprep RP-18, silica gel, Sephadex LH-20, and semi-preparative HPLC, yielded compounds **1**–**5** (Figure 1).

### Structure elucidations

Pseudallene A (**1**), initially obtained as colorless amorphous powder, was assigned a molecular formula of $C_{16}H_{28}O_5S$ with three indices of hydrogen deficiency according to the HRESIMS data. The $^1$H NMR spectrum (Table 1 and Figure S1 in Supporting Information File 1) for compound **1** showed well-dispersed signals over a wide field range, and aided by HSQC experiment, these signals were attributable to four singlet methyls [$\delta_H$ 1.65 (3H, s, $H_3$-12), 1.55 (3H, s, $H_3$-13), 1.39 (3H, s, $H_3$-15), and 3.37 (3H, s, 2-OMe)], four aliphatic methylenes, four methines (with two oxygenated and one olefinic), and four exchangeable protons. The $^{13}$C NMR spectroscopic data (Table 1) displayed all 16 resonances which were classified by DEPT experiments into the categories of four methyls (including one methoxy), four methylenes, four methines (with one olefinic and two oxygenated), and four non-protonated carbons (with one olefinic and three bonded to oxygen).

Detailed interpretation of the COSY spectrum of compound **1** revealed the presence of two discrete proton spin-coupling systems corresponding to a –CH–CH(OH)–$CH_2$–$CH_2$– unit (C-2 to C-5), and a –CH–$CH_2$–CH= moiety (C-8 to C-10) (Figure 2). HMBC correlations from H-4 to C-2 and C-6, from H-5 to C-1 and C-3, from 1-OH to C-2 and C-6, and from 6-OH to C-1, C-5, and C-14, led to the construction of the cyclohexane ring for **1**, while HMBC correlations from H-9 to C-11, and from H-12 and H-13 to C-10 and C-11 constructed the isopentenyl group. Further HMBC correlations from H-15 to C-1, C-7, and C-8 connected the cyclohexane ring and isopentenyl from C-1 and C-8 via the non-protonated carbon C-7. The assignment of the thio-ether bond between C-8 and C-14 was supported by the HMBC correlation from H-14 to C-8, as well as the molecular formula combined with chemical shifts ($\delta_C$ 35.9, $CH_2$-14, and $\delta_C$ 45.8, CH-8) [12]. Furthermore, HMBC correlations from the protons of methoxy to C-2 attached the methoxy group to C-2.

The relative configuration of compound **1** was determined by analysis of NOESY data (Figure 3). NOE cross peaks from H-15 to H-2, 6-OH, and from H-3 to 6-OH indicated the cofacial orientation of these groups, while correlations from 1-OH to H-8, 7-OH, 2-OMe, and 3-OH suggested the opposite position of these groups. To unambiguously clarify the structure of compound **1**, crystals suitable for X-ray crystal analysis were obtained by slow evaporation of the solvent, which could be analyzed by X-ray diffraction analysis using Cu Kα radiation (Figure 4). The resulting Flack parameter, 0.019(6), allowed the

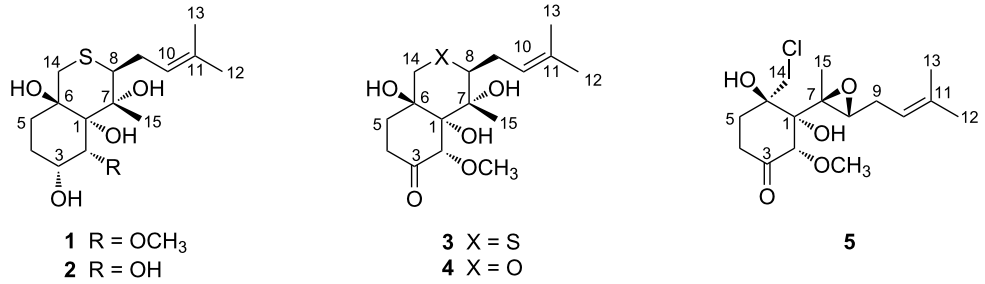



**Figure 1:** Chemical structures of compounds **1**–**5** isolated from *P. boydii* CS-793.

**Table 1:** $^{1}$H (500 MHz) and $^{13}$C NMR (125 MHz) data for compounds **1**–**3**.

| no. | **1** | | **2** | | **3** | |
|---|---|---|---|---|---|---|
| | $\delta_C$[a] | $\delta_H$[b], mult (*J* in Hz) | $\delta_C$[a] | $\delta_H$[b], mult (*J* in Hz) | $\delta_C$[c] | $\delta_H$[d], mult (*J* in Hz) |
| 1 | 78.4, C | | 76.8, C | | 82.4, C | |
| 2 | 79.2, CH | 3.75 d (3.1) | 75.9, CH | 3.89 d (8.7) | 87.4, CH | 4.67, s |
| 3 | 65.0, CH | 4.22 dq (6.1, 3.1) | 70.0, CH | 3.57 ddd (11.1, 8.7, 5.5) | 207.7, C | |
| 4 | 26.8, $CH_2$ | α 1.61 overlap<br>β 1.87 overlap | 28.2, $CH_2$ | α 1.56 overlap<br>β 1.60 overlap | 36.7, $CH_2$ | α 2.78 td (13.7, 6.9)<br>β 2.40 ddd (13.7, 5.4, 1.7) |
| 5 | 30.0, $CH_2$ | α 1.03 dt (11.6, 2.7)<br>β 1.97 overlap | 33.7, $CH_2$ | α 1.09 ddd (13.2, 4.3, 2.7)<br>β 1.77 td (13.2, 4.8) | 35.4, $CH_2$ | α 1.72 ddd (13.5, 6.9, 1.7)<br>β 2.26 td (13.5, 5.4) |
| 6 | 73.5, C | | 72.7, C | | 73.1, C | |
| 7 | 75.8, C | | 76.8, C | | 76.8, C | |
| 8 | 45.8, CH | 2.93 dd (11.5, 2.4) | 45.7, CH | 2.94 overlap | 47.5, CH | 2.91 dd (11.5, 2.5) |
| 9 | 26.2, $CH_2$ | α 1.68 overlap<br>β 2.55 dd (15.2, 6.3) | 26.1, $CH_2$ | α 1.67 overlap<br>β 2.53 overlap | 26.6, $CH_2$ | α 1.85 ddd (14.8, 11.5, 7.9)<br>β 2.63 dd (15.5, 6.3) |
| 10 | 123.2, CH | 5.14 t (6.9) | 123.2, CH | 5.12 t (6.4) | 121.8, CH | 5.22 t (7.3) |
| 11 | 131.4, C | | 131.6, C | | 134.1, C | |
| 12 | 25.5, $CH_3$ | 1.65 s | 25.6, $CH_3$ | 1.65 s | 26.0, $CH_3$ | 1.75 s |
| 13 | 17.6, $CH_3$ | 1.55 s | 17.8, $CH_3$ | 1.55 s | 18.1, $CH_3$ | 1.64 s, |
| 14 | 35.9, $CH_2$ | α 1.93 d (13.7)<br>β 3.01 d (13.7) | 36.1, $CH_2$ | α 1.92 d (13.6)<br>β 2.95 overlap | 37.5, $CH_2$ | α 2.20 d (13.9)<br>β 3.27 d (13.9) |
| 15 | 19.8, $CH_3$ | 1.39 s | 19.3, $CH_3$ | 1.43 s | 19.9, $CH_3$ | 1.54 s |
| 1-OH | | 5.28 s | | 3.93 s | | |
| 2-$OCH_3$/OH | 55.1, $CH_3$ | 3.37 s | | 4.50 brs | 59.3, $CH_3$ | 3.55 s |
| 3-OH | | 5.29 d (6.0) | | 5.61 brs | | |
| 6-OH | | 4.32 s | | 4.16 s | | 3.71 s |
| 7-OH | | 4.79 s | | 3.34 | | 4.28 s |

[a]Data collected at 125 MHz in DMSO-$d_6$. [b]Data collected at 500 MHz in DMSO-$d_6$. [c]Data collected at 125 MHz in $CDCl_3$. [d]Data collected at 500 MHz in $CDCl_3$.

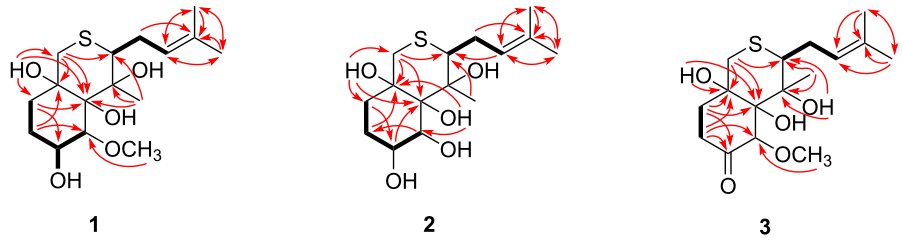



**Figure 2:** Key $^{1}$H-$^{1}$H COSY (bond lines), and HMBC (red arrows) correlations of **1**–**3**.

assignment of the absolute configurations of all the stereogenic centers in compound **1** as 1*R*, 2*R*, 3*R*, 6*R*, 7*R*, 8*S*. This is likely the first characterized crystal structure of an ovalicin-type sesquiterpenoid.

The molecular formula of pseudallene B (**2**) was assigned as $C_{15}H_{26}O_5S$ (three unsaturations), with one $CH_2$ unit less than **1**, based on positive HRESIMS data. Its NMR spectroscopic data were similar to compound **1**. However, the obvious differences from compound **1** was the absence of the methoxy signals resonating at $\delta_C/\delta_H$ 55.1/3.37 (2-$OCH_3$) in compound **2**, which suggested the replacement of an $OCH_3$ in **1** by an OH in compound **2**. This deduction was further verified by the COSY and HMBC correlations (Figure 2).

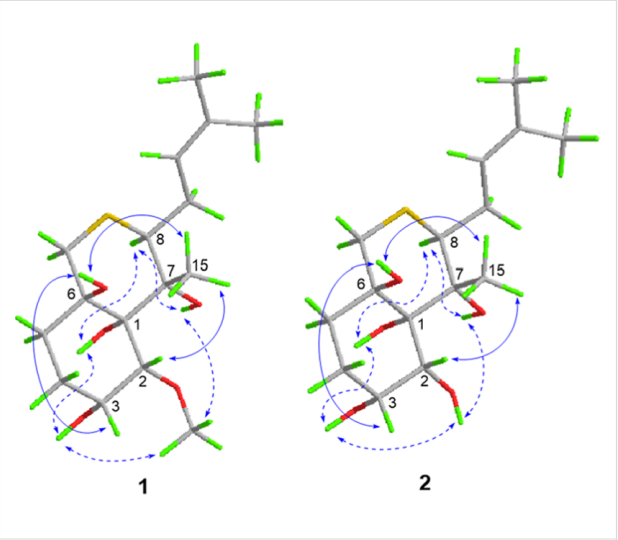



**Figure 3:** NOE correlations of compounds **1** and **2** (solid line indicates β-orientation and dashed lines represent α-orientation).

The structure and relative configuration of compound **2** were deduced the same as for **1** by NOE correlations (Figure 3). Moreover, the absolute configurations of compound **2** were unambiguous determined by X-ray diffraction with the refined Flack parameter of 0.048(7), which suggested all the stereogenic centers in compound **2** as 1*R*, 2*R*, 3*R*, 6*R*, 7*R*, 8*S*.

In addition to compounds **1** and **2**, three related ovalicin-type sesquiterpenoid derivatives **3** [10], **4** [11], and chlovalicin (**5**) [12,13] were isolated and identified from the fungus *P. boydii* CS-793. It should be mentioned that compound **3** was the first sulfur-containing ovalicin sesquiterpenoid, which was previously isolated from *Sporothrix sp.* FO-4649, but its absolute configuration was not explicitly represented, and their $^{1}$H and $^{13}$C NMR data were incomplete [10]. Thus, a full assignment of the NMR data for compound **3** was conducted (Table 1) and its absolute configurations were assigned as 1*R*, 2*S*, 6*R*, 7*R*, 8*S* by single-crystal X-ray diffraction analysis with a Flack parameter of 0.024(6) (Figure 4).

A plausible biosynthetic pathway for compounds **1**–**5** is proposed as shown in Scheme 1. In this pathway, the bergamotene sesquiterpenoid (**I**) is presumed to be a key intermediate cyclized from farnesyl diphosphate (FPP) via nerolidyl diphosphate (NPP) followed by a bisabolyl cation [14]. Subsequent oxidation (bishydroxylation) catalyzed by some oxygenase such as P450 would afford the key intermediate **II**, which could be transferred to **III** by cyclization and epoxidation. Oxidation and methylation of intermediate **III** would produce **IV**. Compounds **1**–**4** could be obtained by nucleophilic attack at C-8 with the hydroxy or thiol group from **IV** via intermediate **V**, followed by oxidation and cyclization (pathway b), while nucleophilic attack at C-14 of intermediate **IV** by a chloride could generate compound **5** (pathway a). In addition, compound **5** might also be derived from intermediate **IV** by cleavage of the ester bond at C-2 to form the intermediate **VI** [15], followed by chlorination (pathway c).

Compounds **1**–**3** were tested against seven human- and marine-derived aquatic pathogenetic bacteria (*Edwardsiella tarda*, *Escherichia coli*, *Micrococcus luteus*, *Pseudomonas aeruginosa*, *Vibrio anguillarum*, *Vibrio harveyi*, and *Vibrio vulnificus*), and six plant pathogenic fungi (*Alternaria brassicae*, *Colletotrichum gloeosporioides*, *Coniothyrium diplodiella*, *Curvularia spicifera*, *Fusarium proliferatum*, and *Penicillium digitatum*) (Table 2). In the antibacterial screening, none of the compounds displayed potent activity against the tested strains (MIC ≥ 32 μg/mL). The antifungal assays showed

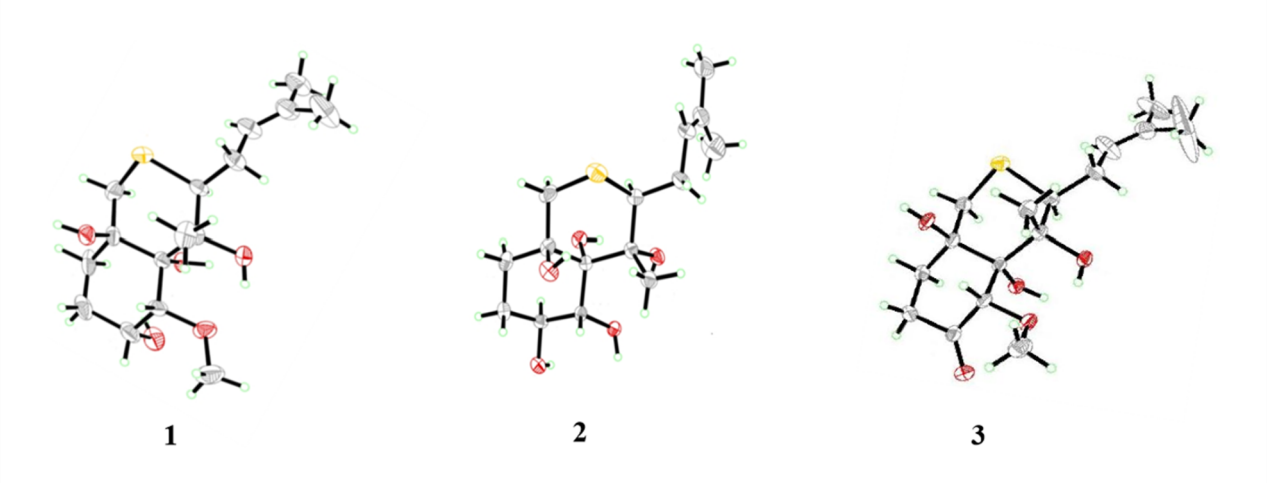
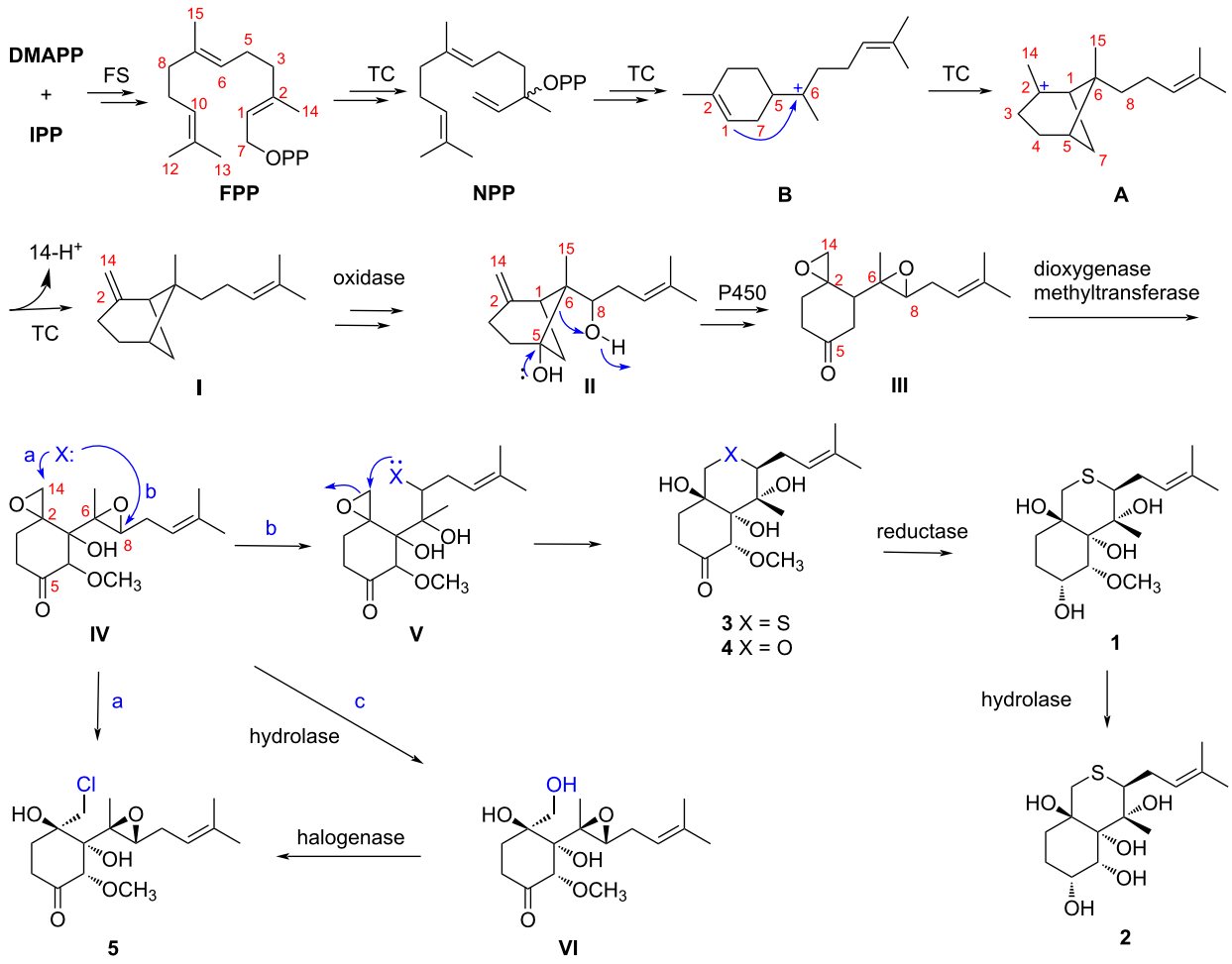



**Figure 4:** X-ray crystal structure of compounds **1**–**3** (with a thermal ellipsoid probability of 50%).




**Scheme 1:** Proposed biosynthetic pathway for compounds **1**–**5**.

**Table 2:** Antimicrobial activities of compounds **1**–**3** (MIC).[a]

| strains | 1 | | 2 | | 3 | | positive control | |
|---|---|---|---|---|---|---|---|---|
| | μg/mL | μM | μg/mL | μM | μg/mL | μM | μg/mL | μM |
| *E. tarda*[b] | 32 | 96.38 | – | – | – | – | 2 | 6.19 |
| *V. anguillarum*[b] | 32 | 96.38 | 64 | 201.26 | 32 | 96.96 | 0.5 | 1.55 |
| *V. harveyi*[b] | – | – | – | – | – | – | 0.5 | 1.55 |
| *E. coli*[b] | – | – | – | – | – | – | 0.5 | 1.55 |
| *V. vulnificus*[b] | – | – | – | – | – | – | 2 | 6.19 |
| *P. aeruginosa*[b] | – | – | – | – | – | – | 2 | 6.19 |
| *M. luteus*[b] | – | – | – | – | – | – | 0.5 | 1.55 |
| *C. diplodiella*[c] | 8 | 24.10 | 4 | 12.58 | 2 | 6.06 | 0.5 | 1.55 |
| *P. digitatum*[c] | 8 | 24.10 | 8 | 25.16 | 8 | 24.24 | 0.5 | 1.55 |
| *A. brassicae*[c] | 16 | 48.19 | 16 | 50.31 | 8 | 24.24 | 0.5 | 1.55 |
| *C. spicifera*[c] | 4 | 12.05 | 4 | 12.58 | 4 | 12.12 | 0.5 | 1.55 |
| *F. proliferatum*[c] | 2 | 6.02 | 4 | 12.58 | 4 | 12.12 | 0.5 | 1.55 |
| *C. gloeosporioides*[c] | 4 | 12.05 | 4 | 12.58 | 16 | 48.48 | 1 | 3.10 |

[a]"–": MIC > 64 μg/mL. [b]Chloramphenicol as positive control. [c]Amphotericin B as positive control.

that compounds **1**–**3** exhibited potent activities against the plant pathogenic fungi. Compound **1** exhibited prominent activity against *F. proliferatum*, *C. spicifera*, *and C. gloeosporioides* with MIC values of 2, 4, and 4 μg/mL, respectively, while compound **3** showed considerable activities against *C. diplodiella*, *C. spicifera* and *F. proliferatum* with MIC values of 2, 4, and 4 μg/mL, respectively.

## Conclusion

In conclusion, two new ovalicin sesquiterpenoid derivatives with a thioether bond, namely, pseudallenes A (**1**) and B (**2**), together with three known analogues **3**–**5**, were isolated and identified from the cold seep-derived fungus *P. boydii* CS-793. To date, only three sulfur-containing ovalicin sesquiterpenoids have been reported [10,12]. Moreover, the first characterized crystal structure of an ovalicin-type sesquiterpenoid was obtained, which further confirmed the structures and absolute configurations of compounds **1**–**3**. Biological evaluation revealed that compounds **1**–**3** exhibit potent antifungal activities against the plant pathogenic fungi *C. diplodiella*, *P. digitatum*, *A. brassicae*, *C. spicifera*, *F. proliferatum*, and *C. gloeosporioides,* with MIC values ranging from 2 to 16 μg/mL.

## Experimental

**General experimental procedures.** Melting points were determined by an SGW X-4 micro-melting point apparatus. Optical rotations were measured in MeOH using an Optical Activity AA-55 polarimeter. UV spectra were obtained with a PuXi TU-1810 UV–visible spectrophotometer. NMR spectra data were recorded on a Bruker Avance 500 or 600 MHz spectrometer using solvent chemical shifts (DMSO: $\delta_H/\delta_C$ 2.50/39.52) as reference. HRESIMS data were measured using an API QSTAR Pulsar 1 mass spectrometer. HPLC was performed on a Dionex HPLC system equipped with a P680 pump, an ASI-100 automated sample injector, and a UVD340U multiple-wavelength detector controlled by Chromeleon software (version 6.80). LC–MS were obtained with a Bruker maXis plus Q-TOF. Column chromatography was carried out using commercially available silica gel (200–300 mesh, Qingdao Haiyang Chemical Co.), Lobar LiChroprep RP-18 (40–63 μm, Merck), and Sephadex LH-20 (Pharmacia). Thin-layer chromatography (TLC) was performed with precoated Si gel $GF_{254}$ plates (Merck, Darmstadt, Germany). Solvents used for extraction and purification were distilled prior to use. Peptone from yeast extract was purchased from Sigma-Aldrich. Rice, monosodium glutamate, and corn steep liquor were purchased from China Oil & Foodstuffs Corporation.

**Fungal material.** The fungus *Pseudallescheria boydii* CS-793 was isolated from the cold seep sediment at the Northeast of the South China Sea, collected in September 2020. The fungal strain was identified as *P. boydii* according to the ITS (internal transcript spacer) region sequence, which is the same (98.98%) as that of *P. boydii* (accession no. OW986361). The sequence data of CS-793 have been deposited in GenBank with the accession no. OQ390095. The strain is preserved at the Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences (IOCAS).

**Fermentation, extraction, and isolation.** Fermentation and extraction were performed in a manner analogous to reference [6]. For chemical investigations, rice solid medium containing rice (100 g/flask), peptone from animal (0.3 g/flask), yeast extract (0.5 g/flask), corn steep liquor (0.2 g/flask), monosodium glutamate (0.1 g/flask), and naturally sourced and filtered seawater (acquired from the Huiquan Gulf of the Yellow Sea near the campus of IOCAS, 100 mL/flask) were autoclaved at 120 °C for 20 min before inoculation. The fresh mycelia of the fungus *P. boydii* CS-793 were incubated in a shaker on PDB medium at 28 °C for five days, which were then inoculated into the noted rice solid medium in 1 L Erlenmeyer flasks and static cultivation for 30 days at room temperature. After fermentation, it was fragmented mechanically and extracted thoroughly with EtOAc. The combined extracts were filtered and concentrated under reduced pressure to give 123.6 g of an organic extract.

The EtOAc extract was subjected to Si gel VLC (vacuum liquid chromatography) and fractionated using solvent mixtures of increasing polarity consisting of petroleum ether (PE) and EtOAc 20:1 to 1:1 and finally with $CH_2Cl_2$/MeOH 20:1 to 1:1 to yield nine fractions (Frs. 1–9). Fr.4 (2.1 g) was further purified by reversed-phase column chromatography (CC) over Lobar LiChroprep RP-18 with a MeOH/$H_2O$ gradient (from 1:9 to 10:0) to afford ten subfractions (Fr. 4.1–Fr. 4.10). Compound **3** (58.7 mg, $t_R$ = 12.0 min) was isolated by semipreparative HPLC (Elite ODSBP column, 5 μm; 10 × 250 mm; 75% MeOH/$H_2O$, 3 mL/min) from Fr. 4.3 (5.8 g). Fr. 5 (6.5 g) was further fractionated by CC over Lobar LiChroprep RP-18 eluting with a MeOH/$H_2O$ gradient (from 1:9 to 10:0) to yield 10 subfractions (Frs. 5.1–5.10). Fr. 5.3 (258 mg) was further purified by CC on silica gel eluting with a $CH_2Cl_2$/MeOH gradient (from 200:1 to 100:1) and then by preparative TLC (plate: 20 × 20 cm, developing solvents: ether/acetone 2:1) to afford compound **4** (8.6 mg). Fr. 5.4 (538 mg) was separated by CC on Si gel and Sephadex LH-20 (MeOH), after that compounds **1** (12.5 mg, $t_R$ = 14.0 min) and **5** (6.0 mg, $t_R$ = 15.0 min) were isolated by semipreparative HPLC (Elite ODSBP column, 5 μm; 10 × 250 mm; 71% MeOH/$H_2O$, 3 mL/min). Fr. 6 (10.5 g) was fractionated by CC over Lobar LiChroprep RP-18 eluting with a MeOH/$H_2O$ gradient (from 1:9 to 10:0) to yield 10 subfractions (Frs. 6.1–6.10). Then, compound **2** (13.7 mg) was isolated by CC on Si gel ($CH_2Cl_2$/MeOH, 250:1 to 50:1)

and preparative TLC (plate: 20 × 20 cm, developing solvent: ether/acetone 2:1) from Fr. 6.3 (578 mg).

Pseudallene A (**1**): colorless crystals (MeOH); mp 115–117 °C; $[\alpha]_D^{25}$ +20.0 (*c* 0.4, MeOH); $^1$H and $^{13}$C NMR data, see Table 2; HRESIMS (*m/z*): $[M + H]^+$ calcd for $C_{16}H_{29}O_5S$, 333.1730; found: 333.1733).

Pseudallene B (**2**): colorless crystals (MeOH); mp 171–175 °C; $[\alpha]_D^{25}$ +53.3 (*c* 0.3, MeOH); $^1$H and $^{13}$C NMR data, see Table 2; HRESIMS (*m/z*): $[M + H]^+$ calcd for $C_{15}H_{27}O_5S$, 319.1573; found: 319.1568..

**X-ray crystallographic analysis of compounds 1–3** [16]. Compounds **1**–**3** were crystallized in MeOH. All crystallographic data were collected on a Bruker Smart-1000 or Bruker D8 Venture CCD diffractometer using Cu Kα radiation (λ = 1.54178 Å). The data were corrected for absorption by using the program SADABS [17]. The structures were solved by direct methods using the SHELXTL software package [18,19]. All non-hydrogen atoms were refined anisotropically. The H atoms were located by geometrical calculations, and their positions and thermal parameters were fixed during the structure refinement. The absolute structures were determined by refinement of the Flack parameters [20], based on anomalous scattering. The structures were optimized by full matrix least-squares techniques.

Crystal data for compound **1**: $C_{16}H_{28}O_5S$, fw = 332.15; orthorhombic space group *P*2(1)2(1)2(1), unit cell dimensions *a* = 15.0176(3) Å, *b* = 24.2644(4) (11) Å, *c* = 29.3542(5) (4) Å, *V* = 4749.6 (3) Å$^3$, α = β = γ = 90°, *Z* = 4, $d_{calcd}$ = 1.237 mg/m$^3$, crystal dimensions 0.18 × 0.16 × 0.12 mm, μ = 1.783 mm$^{-1}$, *F*(000) = 4312. The 67427 measurements yielded 19396 independent reflections after equivalent data were averaged. The final refinement gave $R_1$ = 0.0512 and w$R_2$ = 0.1339 [$I > 2\sigma(I)$]. The absolute structure parameter was 0.019(6).

Crystal data for compound **2**: $C_{15}H_{26}O_5S$, fw = 318.15, triclinic space group *P*1, unit cell dimensions *a* = 6.50160(10) Å, *b* = 7.7295(2) Å, *c* = 17.5958(4) Å, *V* = 794.00(3) Å$^3$, α = 85.4270(10)°, β = 81.7990(10)°, γ = 65.1510(10)°, *Z* = 1, $d_{calcd}$ = 1.332 mg/m$^3$, crystal dimensions 0.18 × 0.15 × 0.12 mm, μ = 1.979 mm$^{-1}$, *F*(000) = 344. The 18694 measurements yielded 5468 independent reflections after equivalent data were averaged. The final refinement gave $R_1$ = 0.0287 and w$R_2$ = 0.0714 [$I > 2\sigma(I)$]. The absolute structure parameter was 0.048(7).

Crystal data for compound **3**: $C_{16}H_{26}O_5S$, fw = 330.43, triclinic space group *P*6(1), unit cell dimensions *a* = 13.99630(10) Å, *b* = 13.99630(10) Å, *c* = 30.1665(6) Å, *V* = 5117.78(13) Å$^3$, α = 90°, β = 90°, γ = 120°, *Z* = 12, $d_{calcd}$ = 1.287 mg/m$^3$, crystal dimensions 0.16 × 0.14 × 0.12 mm, μ = 1.863 mm$^{-1}$, *F*(000) = 2136. The 65189 measurements yielded 6250 independent reflections after equivalent data were averaged. The final refinement gave $R_1$ = 0.0317 and w$R_2$ = 0.0848 [$I > 2\sigma(I)$]. The absolute structure parameter was 0.024(6).

**Antibacterial assay.** The antibacterial activities against human pathogenic bacteria (*Escherichia coli* QDIO-1 and *Pseudomonas aeruginosa* QDIO-4) and aquatic pathogens (*Edwardsiella tarda* QDIO-2, *Micrococcus luteus* QDIO-3, *V. anguillarum* QDIO-6, *Vibrio harveyi* QDIO-7, and *V. vulnificus* QDIO-10), as well as plant pathogenic fungi (*Colletotrichum gloeosporioides* QDAU-2, *Alternaria brassicae* QDAU-11, *Penicillium digitatum* QDAU-14, *Coniothyrium diplodiella* QDAU-15, *Curvularia spicifera* QDAU-18, and *Fusarium proliferatum* QDAU-19) were determined by a serial dilution technique using 96-well microtiter plates [21] with minor modifications as our previously report [22]. Briefly, the bacteria were cultivated in the LB broth medium at 37 °C for the human pathogenic bacteria, while the temperature was 28 °C for the aquatic pathogens, and they were prepared at a concentration of 1.5 × 10$^8$ CFU/mL. Tested compounds and positive control (chloramphenicol) were dissolved in DMSO to give a stock solution with DMSO as a negative control. Then, 95 μL of thus prepared bacteria and 5 μL of the above tested compounds or chloramphenicol (with the final concentration of 0.5, 1, 2, 4, 8, 16, 32, 64 μg/mL) were added into the 96-well plates and cultivated at 37 °C or 28 °C for 24 h. Optical density at 600 nm was read by a multi-detection microplate reader (Infinite M1000 Pro, Tecan). The human pathogenic bacteria and aquatic pathogenic strains were offered by the Institute of Oceanology, Chinese Academy of Sciences.

## Supporting Information

### Supporting Information File 1

Selected 1D and 2D NMR, and HRESIMS spectra of compounds **1** and **2**, and 1D NMR spectra of compounds **3**–**5**.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-42-S1.pdf]

### Supporting Information File 2

X-ray crystallographic files of compounds **1–3**.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-42-S2.zip]

## Acknowledgements

We thank Lingling Li from the Instrumental Analysis Center of Shanghai Jiao Tong University for X-ray crystallographic analysis and Quancai Peng from Institute of Oceanology, Chinese Academy of Sciences for LC–MS assay.

## Funding

This work was supported by the National Natural Science Foundation of China (41976090, U2006203, and 42176115), the National Key R&D Program of China (2022YFC2804700), the Natural Science Foundation of Jiangsu Province (BK20201211), and the Senior User Project of RV *KEXUE* (KEXUE2020GZ02). B.-G.W. acknowledges the support of the Oceanographic Data Center at IOCAS (for CPU time) and the RV *KEXUE* of the National Major Science and Technology Infrastructure from the Chinese Academy of Sciences (for sampling).

## Author Contributions

Zhen Ying: investigation; project administration; writing – original draft. Xiao-Ming Li: data curation; methodology. Sui-Qun Yang: resources. Hong-Lei Li: formal analysis. Xin Li: resources. Bin-Gui Wang: supervision; writing – review & editing. Ling-Hong Meng: funding acquisition; resources; supervision; writing – review & editing.

## ORCID® iDs

Zhen Ying - https://orcid.org/0000-0002-5989-8420
Bin-Gui Wang - https://orcid.org/0000-0003-0116-6195
Ling-Hong Meng - https://orcid.org/0000-0001-6884-0346

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Cong, M.; Pang, X.; Zhao, K.; Song, Y.; Liu, Y.; Wang, J. *Mar. Drugs* **2022,** *20,* 404. doi:10.3390/md20060404
2. Jin, E.; Li, H.; Liu, Z.; Xiao, F.; Li, W. *J. Nat. Prod.* **2021,** *84,* 2606–2611. doi:10.1021/acs.jnatprod.1c00411
3. Yan, L.-H.; Li, P.-H.; Li, X.-M.; Yang, S.-Q.; Liu, K.-C.; Wang, B.-G.; Li, X. *Org. Lett.* **2022,** *24,* 2684–2688. doi:10.1021/acs.orglett.2c00781
4. Hu, X.-Y.; Wang, C.-Y.; Li, X.-M.; Yang, S.-Q.; Li, X.; Wang, B.-G.; Si, S.-Y.; Meng, L.-H. *J. Nat. Prod.* **2021,** *84,* 3122–3130. doi:10.1021/acs.jnatprod.1c00907
5. Hu, X.; Li, X.; Yang, S.; Li, X.; Wang, B.; Meng, L. *Chin. Chem. Lett.* **2023,** *34,* 107516. doi:10.1016/j.cclet.2022.05.030
6. Chi, L.-P.; Li, X.-M.; Wan, Y.-P.; Li, X.; Wang, B.-G. *J. Nat. Prod.* **2020,** *83,* 3652–3660. doi:10.1021/acs.jnatprod.0c00860
7. Lan, W.-J.; Wang, K.-T.; Xu, M.-Y.; Zhang, J.-J.; Lam, C.-K.; Zhong, G.-H.; Xu, J.; Yang, D.-P.; Li, H.-J.; Wang, L.-Y. *RSC Adv.* **2016,** *6,* 76206–76213. doi:10.1039/c6ra06661e
8. Staerck, C.; Landreau, A.; Herbette, G.; Roullier, C.; Bertrand, S.; Siegler, B.; Larcher, G.; Bouchara, J.-P.; Fleury, M. J. J. *FEMS Microbiol. Lett.* **2017,** *364,* fnx223. doi:10.1093/femsle/fnx223
9. Chang, Y.-C.; Deng, T.-S.; Pang, K.-L.; Hsiao, C.-J.; Chen, Y.-Y.; Tang, S.-J.; Lee, T.-H. *J. Nat. Prod.* **2013,** *76,* 1796–1800. doi:10.1021/np400192q
10. Hayashi, M.; Kim, Y.-P.; Takamatsu, S.; Preeprame, S.; Komiya, S.; Masuma, R.; Taknaka, H.; Komiyama, K.; Omura, S. *J. Antibiot.* **1996,** *49,* 631–634. doi:10.7164/antibiotics.49.631
11. Bollinger, P.; Sigg, H.-P.; Weber, H.-P. *Helv. Chim. Acta* **1973,** *56,* 819–830. doi:10.1002/hlca.19730560303
12. Liu, D.-H.; Sun, Y.-Z.; Kurtán, T.; Mándi, A.; Tang, H.; Li, J.; Su, L.; Zhuang, C.-L.; Liu, Z.-Y.; Zhang, W. *J. Nat. Prod.* **2019,** *82,* 1274–1282. doi:10.1021/acs.jnatprod.8b01053
13. Jenssen, M.; Kristoffersen, V.; Motiram-Corral, K.; Isaksson, J.; Rämä, T.; Andersen, J. H.; Hansen, E. H.; Hansen, K. Ø. *Molecules* **2021,** *26,* 7560. doi:10.3390/molecules26247560
14. Wang, W.; Shi, Y.; Liu, Y.; Zhang, Y.; Wu, J.; Zhang, G.; Che, Q.; Zhu, T.; Li, M.; Li, D. *Mar. Drugs* **2022,** *20,* 338. doi:10.3390/md20050338
15. Sorres, J.; Hebra, T.; Elie, N.; Leman-Loubière, C.; Grayfer, T.; Grellier, P.; Touboul, D.; Stien, D.; Eparvier, V. *Molecules* **2022,** *27,* 1182. doi:10.3390/molecules27041182
16. Crystallographic data of compounds **1–3** have been deposited in the Cambridge Crystallographic Data Centre as CCDCs 2309363 (https://www.ccdc.cam.ac.uk/structures/Search?access=referee&ccdc=2309363&Author=Zhen+Ying), 2309407 (https://www.ccdc.cam.ac.uk/structures/Search?access=referee&ccdc=2309407&Author=Zhen+Ying), and 2309408 (https://www.ccdc.cam.ac.uk/structures/Search?access=referee&ccdc=2309408&Author=Zhen+Ying), respectively. These data can be obtained free of charge via http://www.ccdc.cam.ac.uk (or from the CCDC, 12 Union Road, Cambridge CB21EZ, U.K.; fax: +44-1223-336-033; e-mail: deposit@ccdc.cam.ac.uk).
17. *SADABS, Software for Empirical Absorption Correction;* University of Göttingen: Göttingen, Germany, 1996.
18. *SHELXL, Structure Determination Software Programs;* Bruker Analytical X-ray System Inc.: Madison, WI, USA, 1997.
19. *SHELXL, Program for the refinement of crystal structures;* University of Göttingen: Göttingen, Germany, 2014.
20. Parsons, S.; Flack, H. D.; Wagner, T. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2013,** *69,* 249–259. doi:10.1107/s2052519213010014
21. Pierce, C. G.; Uppuluri, P.; Tristan, A. R.; Wormley, F. L., Jr.; Mowat, E.; Ramage, G.; Lopez-Ribot, J. L. *Nat. Protoc.* **2008,** *3,* 1494–1500. doi:10.1038/nprot.2008.141
22. Hu, X.-Y.; Li, X.-M.; Yang, S.-Q.; Liu, H.; Meng, L.-H.; Wang, B.-G. *Mar. Drugs* **2020,** *18,* 194. doi:10.3390/md18040194

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.42

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# A new analog of dihydroxybenzoic acid from *Saccharopolyspora* sp. KR21-0001

Rattiya Janthanom[1,2], Yuta Kikuchi[3], Hiroki Kanto[3], Tomoyasu Hirose[2,3], Arisu Tahara[4], Takahiro Ishii[4], Arinthip Thamchaipenet[*1,§] and Yuki Inahashi[*2,3,¶]

Full Research Paper

Open Access

Address:
[1]Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngamwongwan Road, Ladyao, Chatuchak, Bangkok 10900, Thailand, [2]Ōmura Satoshi Memorial Institute, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan, [3]Graduate School of Infection Control Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan and [4]Faculty of Agriculture, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903-0213, Japan

Email:
Arinthip Thamchaipenet[*] - arinthip.t@ku.ac.th;
Yuki Inahashi[*] - y-ina@lisci.kitasato-u.ac.jp

* Corresponding author
§ Tel: +66-0-2562-5444 ext. 646044, Fax: +66 0-2578-5528
¶ Tel/Fax: +81-3-5791-6133

Keywords:
antioxidant activity; dihydroxybenzoic acid analog; rare actinomycetes

*Beilstein J. Org. Chem.* **2024,** *20,* 497–503.
https://doi.org/10.3762/bjoc.20.44

Received: 15 December 2023
Accepted: 14 February 2024
Published: 29 February 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra



© 2024 Janthanom et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Actinomycetes are well-known as the main producers of bioactive compounds such as antibiotics, anticancers, and immunosuppressants. Screening of natural products from actinomycetes has been an essential part of several drug discovery programs. Finding such novel biologically active metabolites is immensely important because of their beneficial health effects. Recently, the discovery of new compounds has diverted attention to rare actinomycetes, since they are rich sources of natural products. In this study, a collection of rare actinomycetes at Kitasato University has been screened for potential novel compound producers. Among the rare actinomycetes, *Saccharopolyspora* sp. KR21-0001 isolated from soil on Ōha Island, Okinawa, Japan was selected as a potential producer. The strain was cultured in 20 L of production medium in a jar fermenter and the culture broth was extracted. Further purification revealed the presence of a new compound designated KR21-0001A (**1**). The structure was elucidated by NMR, and the absolute stereochemistry was determined by advanced Marfey's method. The results indicated that **1** is a new analog of dihydroxybenzoic acid. **1** has no antimicrobial activity against bacteria and fungi but showed potent antioxidant activity.

## Introduction

Actinomycetes are Gram-positive bacteria with high GC content in their genome. They are well-known as the main producers of bioactive compounds such as antibiotic, anticancer, antitumor, antiinflammatory, antihyperglycemic, antiapoptotic, immunosuppressant, and antioxidant compounds [1-3]. According to natural product screening since the 1920s,

about 30,000 compounds derived from microbial sources have been reported [4]. Over 10,000 compounds are produced by actinomycetes, with 70% from *Streptomyces* and the rest from rare actinomycetes (non-*Streptomyces*) [5]. Rare actinomycetes are defined as actinomycete strains with low isolation rates when compared with the isolation of *Streptomyces* [6]. Currently, the discovery of new natural compounds is focusing on rare actinomycetes, which are believed to be rich sources of interestingly new compounds.

Bioactive compounds are vastly beneficial in medicine, pharmaceutical industry, and agriculture [7,8]. The thoughtful exploration of novel sources for acquiring new compounds is crucial, and the strategies employed in this pursuit are equally significant. One of the efficient approaches for finding new secondary metabolites from microorganisms is physicochemical (PC) screening. This approach involves the screening of the physicochemical properties of potential compounds, such as UV spectroscopy, mass spectrometry, and color reaction [9]. This technique has been used to find several new compounds from actinomycetes [10-13].

In the course of our PC screening for new natural products from rare actinomycetes, *Saccharopolyspora* sp. KR21-0001 produced a new compound **1**, whose physicochemical properties, accurate mass and UV spectrum, did not match any compound in the Dictionary of Natural Products database (version 32.1). Fermentation, isolation, structural elucidation, and biological activity of **1** are described in the following.

## Results and Discussion

*Saccharopolyspora* sp. KR21-0001 was isolated from soil in Ōha island, Okinawa, Japan. The 16S rRNA gene sequence (GenBank accession number: PP237390) showed the highest similarity values with *Saccharopolyspora shandongensis* 88$^T$ (similarity: 99.6%) and *Saccharopolyspora elongate* 7K502$^T$ (similarity: 99.0%). Phylogenetic analysis of the 16S rRNA gene sequences showed that the strain KR21-0001 was clustered with them (Figure 1). The similarity of 16S rRNA gene sequence and phylogenetic analysis indicated that strain KR21-0001 belongs to the genus *Saccharopolyspora*.

*Saccharopolyspora* sp. KR21-0001 was incubated in a jar fermenter containing 20 L of production medium for 7 days (Scheme 1). KR21-0001A (**1**) was purified from the culture broth guided by the ion peak (*m*/*z* 316) using LC–MS. The supernatant of the culture broth underwent HP20 column chromatography, but **1** was not retained in the resin and eluted in the

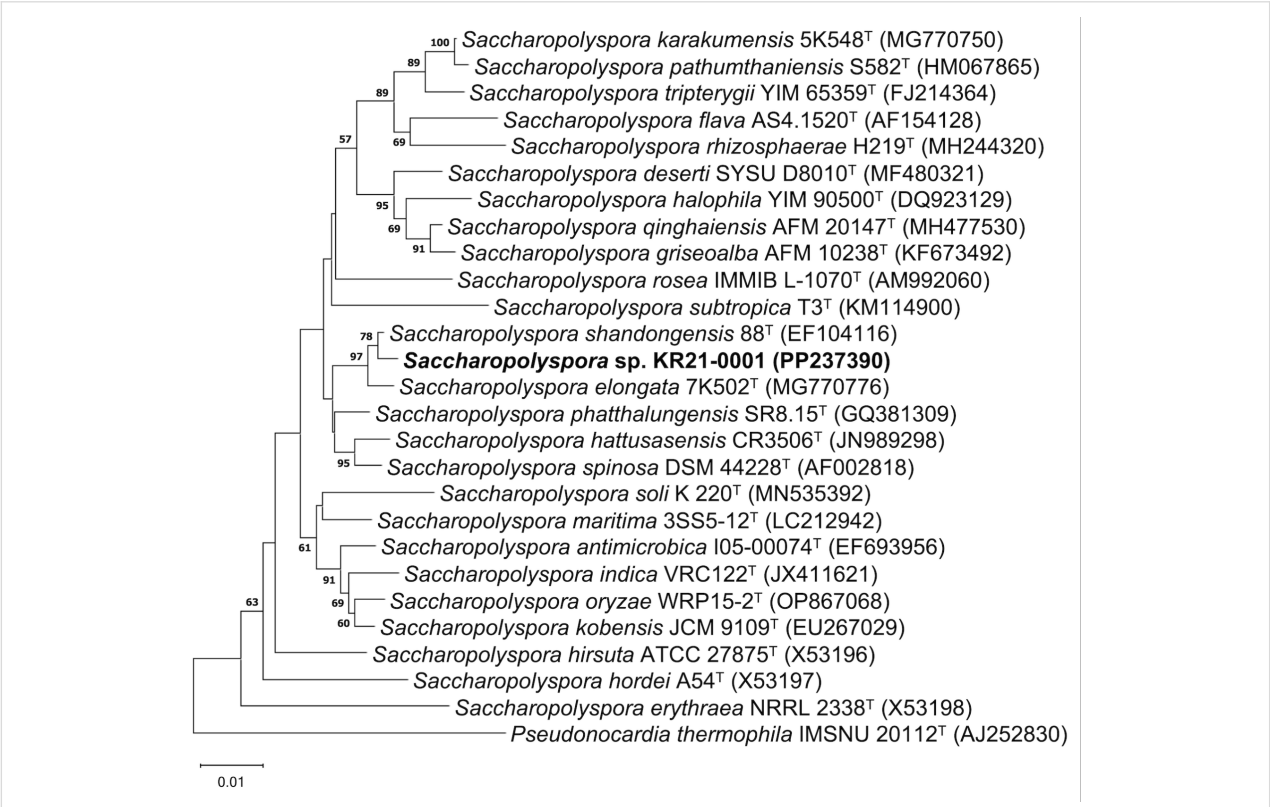
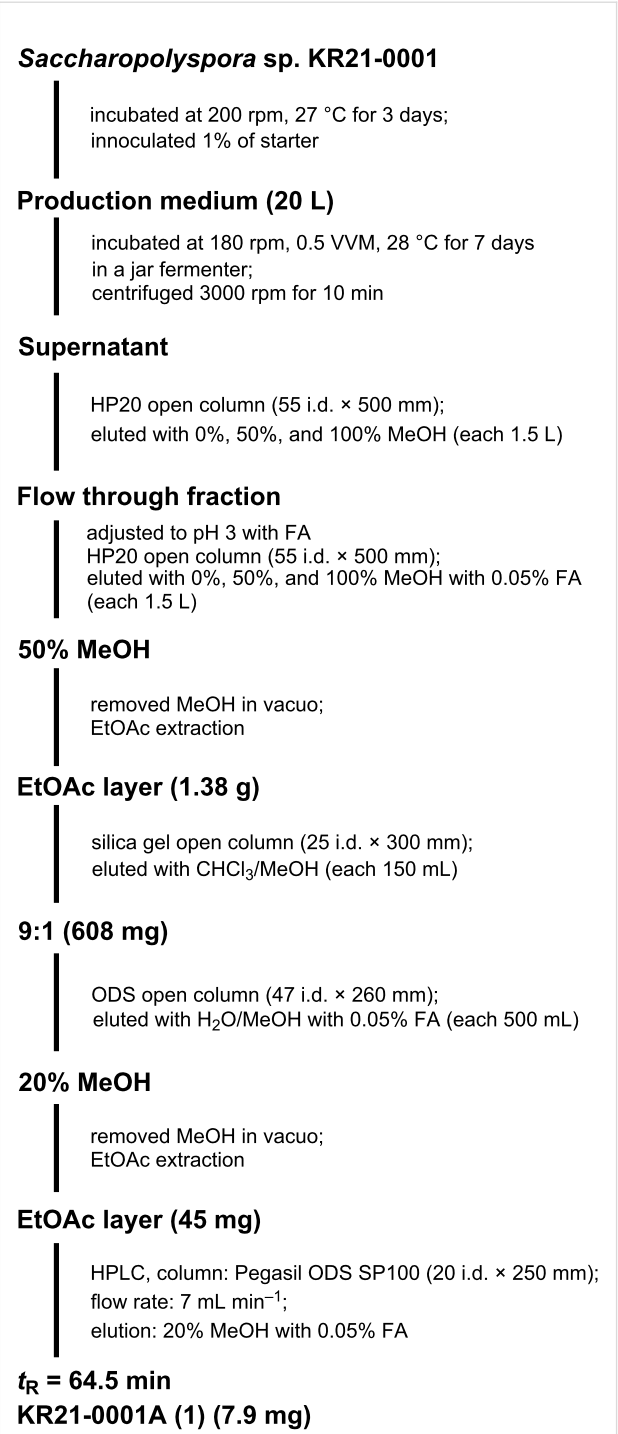



**Figure 1:** Neighbor-joining tree based on 16S rRNA gene sequences between KR21-0001 and members of the genus *Saccharopolyspora*. Only bootstrap values above 50% (percentages of 1000 replications) are shown.




**Scheme 1:** Fermentation of *Saccharopolyspora* sp. KR21-0001 and isolation procedure of KR21-0001A (**1**).

flowthrough fraction. Since **1** was an acidic compound, the pH of the flowthrough fraction was adjusted to 3 with formic acid (FA) and chromatographed again on HP20 column. Now, **1** was retained in the resin and eluted by 50% MeOH with 0.05% FA. Then **1** was purified by silica gel and ODS column chromatography, and ethyl acetate extraction under acidic conditions. As the result of preparative HPLC of the crude extract, 7.9 mg of **1** was obtained (Scheme 1).

Table S1 (Supporting Information File 1) shows the physicochemical properties of **1**, which is a yellow oil soluble in MeOH and DMSO. The UV absorption maximum of **1** was at 286 nm ($\varepsilon$ = 10238 $M^{-1}{\cdot}cm^{-1}$). The molecular formula of **1** was determined as $C_{12}H_{13}NO_7S$ (requiring seven degrees of unsaturation) from the $[M + H]^+$ ion at *m*/*z* 316.0484 (calcd. for $C_{12}H_{14}NO_7S$, 316.0485) by HRESIMS. The $^1H$ and $^{13}C$ NMR spectral data of **1** are listed in Table 1. The $^1H$ NMR and heteronuclear single quantum coherence (HSQC) data indicated the presence of two $sp^2$ methines, one $sp^3$ methine, one $sp^3$ methylene, and one methyl group. The $^{13}C$ NMR data showed the resonances of twelve carbons, which were classified into six olefinic carbons (including two oxygenated carbons: $\delta_C$ 151.0 and 145.2), three carbonyl carbons, one $sp^3$ methine carbon, one $sp^3$ methylene carbon, and one methyl carbon. A spin system H-6 ($\delta_H$ 6.85)/H-7 ($\delta_H$ 7.33) with an *ortho*-coupling constant (8.5 Hz) was observed using homonuclear correlation spectrometry (COSY). The heteronuclear multiple bond correlation (HMBC) cross-peaks from H-6 to C-2 ($\delta_C$ 112.4) and C-4 ($\delta_C$ 145.2), and from H-7 to C-1 ($\delta_C$ 173.8), C-3 ($\delta_C$ 151.0), and C-5 ($\delta_C$ 129.4) indicated that **1** has a 2,3-dihydroxybenzoic acid moiety (Table 1 and Figure 2a; Figure S3, Supporting Information File 1). A spin system H-9 ($\delta_H$ 4.58)/$H_2$-10 ($\delta_H$ 3.54 and 3.23) was observed in $^1H,^1H$ COSY (Figure S5, Supporting Information File 1). The HMBC cross-peaks from H-9 to C-8 ($\delta_C$ 173.5) and C-11 ($\delta_C$ 173.4), from $H_2$-10 to C-8 ($\delta_C$ 173.5) and C-9 ($\delta_C$ 53.6), and from $H_3$-12 ($\delta_H$ 1.94) to C-11 ($\delta_C$ 173.4)

**Table 1:** NMR spectroscopic data of KR21-0001A (**1**) in $CD_3OD$.

| position | $\delta_C$, type | $\delta_H$ (*J* in Hz) | HMBC |
|---|---|---|---|
| 1 | 173.8, C | | |
| 2 | 112.4, C | | |
| 3 | 151.0, C | | |
| 4 | 145.2, C | | |
| 5 | 129.4, C | | |
| 6 | 120.3, CH | 6.85 d (8.5) | 145.2, 112.4 |
| 7 | 121.5, CH | 7.33 d (8.5) | 173.8, 151.0, 129.4 |
| 8 | 173.5, C | | |
| 9 | 53.6, CH | 4.58 dd (8.4, 4.4) | 173.4, 173.5 |
| 10 | 34.2, $CH_2$ | 3.54 dd (13.9, 4.4)<br>3.23 dd (13.9, 8.4) | 173.4, 129.4<br>173.4, 129.4, 53.2 |
| 11 | 173.4, C | | |
| 12 | 22.3, $CH_3$ | 1.94 s | 173.4 |

indicated that **1** has a *N*-acetylcysteine moiety (Figure 2a). The HMBC cross-peak from H-10 ($\delta_H$ 3.23) to C-5 revealed the linkage between C-5 ($\delta_C$ 129.4) and C-10 through a sulfur atom. Based on these results, the structure of **1** was elucidated as 4-((2-acetamido-2-carboxyethyl)thio)-2,3-dihydroxybenzoic acid.

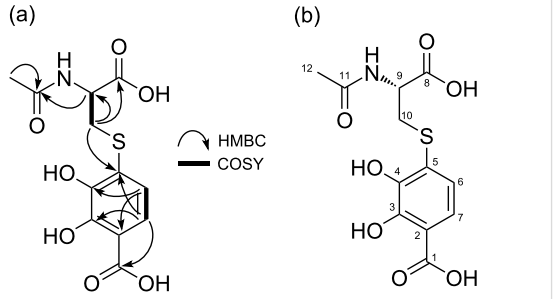



**Figure 2:** Structure of KR21-0001A (**1**). (a) $^1$H,$^1$H COSY and HMBC correlations and (b) absolute configuration.

The absolute configuration of **1** was determined by advanced Marfey's analysis. **1** was de-sulfurized with skeletal Ni and then hydrolyzed with HCl. The obtained amino acid (alanine) and authentic amino acids were modified using $N^{\alpha}$-(5-fluoro-2,4-dinitrophenyl)-D-leucinamide (D-FDLA). The LC–MS analysis of D-FDLA derivatives revealed that the chiral center of the amino acid in **1** had L configuration (Figure S7, Supporting Information File 1).

Antioxidant activity of **1** was measured via the DPPH radical. **1** showed potent DPPH radical scavenging activity with an $IC_{50}$ value of 5.0 μg·mL$^{-1}$, which is lower than that of trolox ($IC_{50}$; 7.5 μg·mL$^{-1}$) (Table 2). In contrast, **1** did not show antimicrobial activity against Gram-positive and Gram-negative bacteria and fungi at 20 μg per disc in the paper disc diffusion method.

**Table 2:** $IC_{50}$ value of DPPH radical scavenging activity.

| Sample | $IC_{50}$ (μg·mL$^{-1}$) |
|---|---|
| KR21-0001A | 5.0 |
| Trolox | 7.5 |

Dihydroxybenzoic acid (DHBA) plays a role in anti-inflammatory, antihyperglycemic, antiapoptotic, and antioxidant processes [14]. 2,3-DHBA is found in nature and is produced by various plants (e.g., *Gentiana rigerscense* and *Vinca minor*), fungi (e.g., *Aspergillus sojae* and *Penicillium roquefortii*), and bacteria (e.g., *Acinetobacter calcoaceticus, Brucella abortus,* and *Bacillus* sp.) [15]. It is also known to be a component of some natural products, such as enterobactin, showing strong radical scavenging activity or antioxidant activity [16]. KR21-0001A (**1**) is a new analog of 2,3-DHBA connected to *N*-acetylcysteine (Figure 2b). **1** has a stronger antioxidant activity than trolox, which is a water-soluble analog of the free radical scavenger α-tocopherol [17,18]. **1** shows no antimicrobial activity against bacteria and fungi.

## Conclusion

Rare actinomycetes are excellent sources of novel bioactive compounds, since they are less explored for secondary metabolites than the more common strains of *Streptomyces* [19,20]. The compounds from this group often have unique structures that may exhibit novel biological activities and could be applied in various industries, such as pharmaceutics, agriculture, and environmental remediation [21,22]. However, the discovery of bioactive compounds from rare actinomycetes comes with challenges, such as the difficulty of isolation from environments. In this study, we searched for new secondary metabolites from the rare actinomycete *Saccharopolyspora* sp. KR21-0001. As a result, KR21-0001A (**1**), a new analog of dihydroxybenzoic acid, was discovered as a potent antioxidant. In conclusion, exploring rare actinomycetes for bioactive compound discovery represents a promising way in the quest for new drugs and biotechnological innovations. Continued research in this area has the potential to uncover valuable compounds that contribute to human health, pharmaceutical industry, and environmental sustainability.

## Experimental

### Taxonomic analysis

*Saccharopolyspora* sp. KR21-0001 was isolated from soil in Ōha Island, Ou, Kumejima, Shimajiri District, Okinawa, Japan. Genomic DNA was prepared, and the 16S rRNA gene was amplified by PCR using the method of Inahashi and co-workers [23]. The sequencing analysis was performed by Eurofins Genomics. Similarity of 16S rRNA gene sequence was computed by using EzBioCloud [24]. A phylogenetic tree was constructed based on the neighbor-joining method [25] by using MEGAX [26].

### Fermentation

*Saccharopolyspora* sp. KR21-0001 was inoculated into 100 mL of seed medium (2.4% soluble starch, 0.1% glucose, 0.3% peptone, 0.3% meat extract, 0.3% yeast extract, and 0.4% $CaCO_3$; pH 7.0) as the starter and incubated at 180 rpm in a rotary shaker at 27 °C for 4 days. The 1% portion of the starter was transferred into 20 L of production medium (2% glycerol, 2% soybean meal, and 0.3% NaCl) in a jar fermenter and further incubated at 150 rpm, 27 °C, 0.5 VVM (volume per volume per minute) for 7 days.

## Purification and extraction

The 7-day culture broth of *Saccharopolyspora* sp. KR21-0001 was centrifuged at 3000 rpm for 10 min. The supernatant fraction was eluted stepwise with 1.5 L of MeOH/$H_2O$ (0%, 50%, and 100%) through a HP20 column (55 i.d. × 500 mm), and all fractions were analyzed by LC–MS. The flow-through and 0% fractions were mixed and adjusted to pH 3 by adding formic acid (FA). Then, the mixture was purified by eluting with MeOH/$H_2O$ (0%, 50%, and 100%) with 0.05% of FA through a HP20 column (55 i.d. × 500 mm). The 50% fraction was concentrated in vacuo to remove MeOH and extracted with ethyl acetate. The ethyl acetate layer was purified using a silica gel column (25 i.d. × 300 mm) by eluting with $CHCl_3$/MeOH (100:0, 100:1, 100:2, 10:1, 1:1, and 0:100). The 10:1 fraction was purified with an ODS column (47 i.d. × 260 mm) by eluting with MeOH/$H_2O$ (0%, 10%, 20%, 30%, 40%, 50%, 60%, 80%, and 100%) with 0.05% FA. The 20% fraction was purified by HPLC (Shimadzu, Columbia, MD, U.S.A.) with an ODS column (Pegasil ODS SP100, 20 i.d. × 250 mm) at a flow rate of 7 mL·min$^{-1}$ and eluted with 20% MeOH with 0.05% FA (Scheme 1). The fraction with a retention time of 64.5 min was concentrated in vacuo to yield **1** (7.9 mg).

## LC–MS analysis

Electrospray ionization-mass spectrometry (ESIMS) data was collected using a Triple TOF 5600+ LC–MS/MS System (AB Sciex, Framingham, MA, USA) with an ODS column (Capcell Core C18, 3.0 i.d. × 100 mm, 40 °C) (Osaka Soda, Japan) at a flow rate of 0.5 mL·min$^{-1}$ and gradient elution with MeOH/$H_2O$ with 0.1% FA.

## Structure elucidation

Spectra from $^1$H NMR at 500 MHz and $^{13}$C NMR at 125 MHz were measured in $CD_3OD$ using a JNM-ECA500 (JEOL Ltd., Tokyo Japan). Chemical shifts were referenced to $CD_3OD$ (3.31 ppm) in the $^1$H NMR spectra and $CD_3OD$ (49.0 ppm) in the $^{13}$C NMR spectra. Infrared (IR) spectroscopy was carried out using an FT-4600 Fourier transform infrared spectrometer (JASCO P-2200 Polarimeter, Japan). The UV spectra were measured using a SpectraMax QuickDrop micro-volume spectrophotometer (Molecular Devices, LLC., San Jose, USA) at wavelengths of 200 to 800 nm. Optical rotation was measured using a P-2200 polarimeter (JASCO).

## De-sulfurization and hydrolyzation

KR21-0001A (**1**) (1 mg) was dissolved in EtOH (0.5 mL) and treated with skeletal Ni (1 mg). The reaction mixture was stirred for 13 h at room temperature under $H_2$ (Figure S6, Supporting Information File 1). Then, the mixture was filtered through celite pad, and the filtrate was concentrated in vacuo. A volume of 100 µL of 1 M HCl was added and heated at 100 ºC for 3 h. After removing HCl in vacuo, the hydrolysate of de-sulfurized derivative of **1** was dissolved in 1 mL of $H_2O$.

## Advanced Marfey's analysis

A volume of 20 µL of 1 M $NaHCO_3$ was added to each 50 µL of the hydrolysate of the de-sulfurized derivative of **1** and 1 mM standard amino acids (L- and D-alanine). Then, 50 µL of D-FDLA (10 mg·mL$^{-1}$ in acetone) was added to the mixtures and mixed well (the color changed from yellow to red). Next, the mixtures were incubated at 37 °C for 1 h. After incubation, 25 µL of 1 M HCl was added to each sample for neutralization. The mixtures were concentrated to dryness in vacuo, dissolved in acetonitrile, and analyzed by LC–MS.

## Biological activity

### Antioxidant activity

The antioxidant activity of KR21-0001A (**1**) was measured using 2,2-diphenyl-1-picrylhydrazyl (DPPH) (Antioxidant Assay Kit, DOJINDO© laboratory, Japan) according to the manufacturer's protocol.

### Antimicrobial activity

KR21-0001A (**1**) was investigated for antibacterial and antifungal activities with two strains of Gram-positive bacteria, *Kocuria rhizophila* ATCC 9341 and *Bacillus subtilis* ATCC 6633; two strains of Gram-negative bacteria, *Escherichia coli* Europe NIHJ and *Xanthomona campestis* pv. *oryzae* KB 88; and two strains of fungi, *Candida albicans* ATCC 64548 and *Mucor racemosus* IFO 4581. All strains were tested by paper disc diffusion assay at 20 µg on each paper disc and incubated at 37 °C for 24 h for *M. luteus*, *B. subtilis*, and *E. coli*; and 27 °C for 24 h for *X. oryzae*, *C. albicans*, and *M. racemosus*.

## Supporting Information

### Supporting Information File 1

Additional data and NMR spectra.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-44-S1.pdf]

## Acknowledgements

We are grateful to Distinguished Emeritus Professor Satoshi Ōmura (Kitasato University) for his helpful support and valuable suggestions.

## Funding

This research was financially supported by Royal Golden Jubilee PhD (RGJPHD), Thailand Research Fund, and National Research Council of Thailand (RGJ-NRCT) under grant no.

PHD/0150/2557. This study was partially supported by a Grant-in-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports, Science, and Technology of Japan under grant no. 19H05685, Okinawan Create Leading Projects in Growing Fields 2017-2019 and 2020-2021, OKINAWA Prefectural Government, and the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research; BINDS) from the Japan Agency for Medical Research & Development (AMED) under grant no. JP19am0101096 (Phase I) and JP22ama121035 (Phase II).

## Author Contributions

Rattiya Janthanom: funding acquisition; investigation; writing – original draft. Yuta Kikuchi: investigation. Hiroki Kanto: investigation. Tomoyasu Hirose: funding acquisition; investigation; writing – review & editing. Arisu Tahara: investigation. Takahiro Ishii: resources; writing – review & editing. Arinthip Thamchaipenet: conceptualization; writing – review & editing. Yuki Inahashi: conceptualization; funding acquisition; project administration; supervision; writing – original draft.

## ORCID® iDs

Takahiro Ishii - https://orcid.org/0000-0003-4180-9917
Arinthip Thamchaipenet - https://orcid.org/0000-0002-8749-0414
Yuki Inahashi - https://orcid.org/0000-0001-7436-2037

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Alenazi, A. M.; Anwar, Y.; Abo-Aba, S. E. M.; Bataweel, N. M. *Novel Res. Microbiol. J.* **2023,** *7,* 1918–1931. doi:10.21608/nrmj.2023.294180
2. Prudence, S. M. M.; Addington, E.; Castaño-Espriu, L.; Mark, D. R.; Pintor-Escobar, L.; Russell, A. H.; McLean, T. C. *Microbiology (London, U. K.)* **2020,** *166,* 683–694. doi:10.1099/mic.0.000944
3. Tiwari, K.; Gupta, R. K. *Crit. Rev. Biotechnol.* **2012,** *32,* 108–132. doi:10.3109/07388551.2011.562482
4. Saito, S.; Kato, W.; Ikeda, H.; Katsuyama, Y.; Ohnishi, Y.; Imoto, M. *J. Antibiot.* **2020,** *73,* 203–210. doi:10.1038/s41429-020-0279-4
5. Ding, T.; Yang, L.-J.; Zhang, W.-D.; Shen, Y.-H. *RSC Adv.* **2019,** *9,* 21964–21988. doi:10.1039/c9ra03579f
6. Parra, J.; Beaton, A.; Seipke, R. F.; Wilkinson, B.; Hutchings, M. I.; Duncan, K. R. *Curr. Opin. Microbiol.* **2023,** *76,* 102385. doi:10.1016/j.mib.2023.102385
7. Haruna, A.; Yahaya, S. M. *Chem. Afr.* **2021,** *4,* 231–248. doi:10.1007/s42250-020-00213-9
8. Katz, L.; Baltz, R. H. *J. Ind. Microbiol. Biotechnol.* **2016,** *43,* 155–176. doi:10.1007/s10295-015-1723-5
9. Nakashima, T.; Takahashi, Y.; Ōmura, S. *Biochem. Pharmacol.* **2017,** *134,* 42–55. doi:10.1016/j.bcp.2016.09.026
10. Koomsiri, W.; Inahashi, Y.; Kimura, T.; Shiomi, K.; Takahashi, Y.; Ōmura, S.; Thamchaipenet, A.; Nakashima, T. *J. Antibiot.* **2017,** *70,* 1142–1145. doi:10.1038/ja.2017.113
11. Matsuo, H.; Nonaka, K.; Nagano, Y.; Yabuki, A.; Fujikura, K.; Takahashi, Y.; Ōmura, S.; Nakashima, T. *Biosci., Biotechnol., Biochem.* **2018,** *82,* 1323–1326. doi:10.1080/09168451.2018.1467264
12. Nakashima, T.; Kimura, T.; Miyano, R.; Matsuo, H.; Hirose, T.; Kimishima, A.; Nonaka, K.; Iwatsuki, M.; Nakanishi, J.; Takahashi, Y.; Ōmura, S. *J. Biosci. Bioeng.* **2017,** *123,* 765–770. doi:10.1016/j.jbiosc.2017.01.011
13. Inahashi, Y.; Iwatsuki, M.; Ishiyama, A.; Matsumoto, A.; Hirose, T.; Oshita, J.; Sunazuka, T.; Panbangred, W.; Takahashi, Y.; Kaiser, M.; Otoguro, K.; Omura, S. *Org. Lett.* **2015,** *17,* 864–867. doi:10.1021/ol5037216
14. Moazzen, A.; Öztinen, N.; Ak-Sakalli, E.; Koşar, M. *Heliyon* **2022,** *8,* e10467. doi:10.1016/j.heliyon.2022.e10467
15. Ito, T. *Appl. Environ. Microbiol.* **1993,** *59,* 2343–2345. doi:10.1128/aem.59.7.2343-2345.1993
16. Peralta, D. R.; Adler, C.; Corbalán, N. S.; Paz García, E. C.; Pomares, M. F.; Vincent, P. A. *PLoS One* **2016,** *11,* e0157799. doi:10.1371/journal.pone.0157799
17. Giordano, M. E.; Caricato, R.; Lionetto, M. G. *Antioxidants* **2020,** *9,* 1058. doi:10.3390/antiox9111058
18. Poljšak, B.; Raspor, P. *J. Appl. Toxicol.* **2008,** *28,* 183–188. doi:10.1002/jat.1264
19. Ezeobiora, C. E.; Igbokwe, N. H.; Amin, D. H.; Enwuru, N. V.; Okpalanwa, C. F.; Mendie, U. E. *Future J. Pharm. Sci.* **2022,** *8,* 23. doi:10.1186/s43094-022-00410-y
20. Jose, P. A.; Sivakala, K. K.; Rajeswari, P.; Jebakumar, S. R. D. *Sci. World J.* **2014,** 456070. doi:10.1155/2014/456070
21. Oyedoh, O. P.; Yang, W.; Dhanasekaran, D.; Santoyo, G.; Glick, B. R.; Babalola, O. O. *Agronomy (Basel, Switz.)* **2023,** *13,* 666. doi:10.3390/agronomy13030666
22. Oyedoh, O. P.; Yang, W.; Dhanasekaran, D.; Santoyo, G.; Glick, B. R.; Babalola, O. O. *Biotechnol. Adv.* **2023,** *67,* 108205. doi:10.1016/j.biotechadv.2023.108205
23. Inahashi, Y.; Matsumoto, A.; Danbara, H.; Ōmura, S.; Takahashi, Y. *Int. J. Syst. Evol. Microbiol.* **2010,** *60,* 2652–2658. doi:10.1099/ijs.0.016477-0
24. Yoon, S.-H.; Ha, S.-M.; Kwon, S.; Lim, J.; Kim, Y.; Seo, H.; Chun, J. *Int. J. Syst. Evol. Microbiol.* **2017,** *67,* 1613–1617. doi:10.1099/ijsem.0.001755
25. Saitou, N.; Nei, M. *Mol. Biol. Evol.* **1987,** *4,* 406–425. doi:10.1093/oxfordjournals.molbev.a040454
26. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. *Mol. Biol. Evol.* **2018,** *35,* 1547–1549. doi:10.1093/molbev/msy096

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.44

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Recent developments in the engineered biosynthesis of fungal meroterpenoids

Zhiyang Quan and Takayoshi Awakawa*

Review

Open Access

Address:
RIKEN Center for Sustainable Resource Science, Wako, Saitama 351-0198, Japan

Email:
Takayoshi Awakawa* - takayoshi.awakawa@riken.jp

* Corresponding author

Keywords:
αKG-dependent dioxygenases; enzyme engineering; fungal meroterpenoids; synthetic biology; terpene cyclases

*Beilstein J. Org. Chem.* **2024,** *20,* 578–588.
https://doi.org/10.3762/bjoc.20.50

Received: 15 December 2023
Accepted: 21 February 2024
Published: 13 March 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Associate Editor: K. N. Allen

© 2024 Quan and Awakawa; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Meroterpenoids are hybrid compounds that are partially derived from terpenoids. This group of natural products displays large structural diversity, and many members exhibit beneficial biological activities. This mini-review highlights recent advances in the engineered biosynthesis of meroterpenoid compounds with C15 and C20 terpenoid moieties, with the reconstruction of fungal meroterpenoid biosynthetic pathways in heterologous expression hosts and the mutagenesis of key enzymes, including terpene cyclases and α-ketoglutarate (αKG)-dependent dioxygenases, that contribute to the structural diversity. Notable progress in genome sequencing has led to the discovery of many novel genes encoding these enzymes, while continued efforts in X-ray crystallographic analyses of these enzymes and the invention of AlphaFold2 have facilitated access to their structures. Structure-based mutagenesis combined with applications of unnatural substrates has further diversified the catalytic repertoire of these enzymes. The information in this review provides useful knowledge for the design of biosynthetic machineries to produce a variety of bioactive meroterpenoids.

## Introduction

Meroterpenoids are complex natural products with intricate skeletal structures, and are partially derived from terpenoids [1]. Many fungal meroterpenoids are composed of polyketide and terpenoid moieties. Examples of fungal meroterpenoids include mycophenolic acid (Figure 1, **1**), which shows immunosuppressive activity and cell differentiation-inducing activity by inhibiting the IMP dehydrogenase involved in inosinic acid metabolism [2]; ascofuranone (Figure 1, **2**), which exhibits antiparasitic activity by selectively inhibiting the oxidase of *Plasmodium trypanosoma* [3]; and pyripyropene (Figure 1, **3**), which displays hyperlipidemia treatment activity through the inhibition of acyl-CoA: cholesterol acyltransferase [4]. In this mini-review, we focus on the fungal meroterpenoids biosynthesis, especially terpenoid cyclizations and post-cyclization

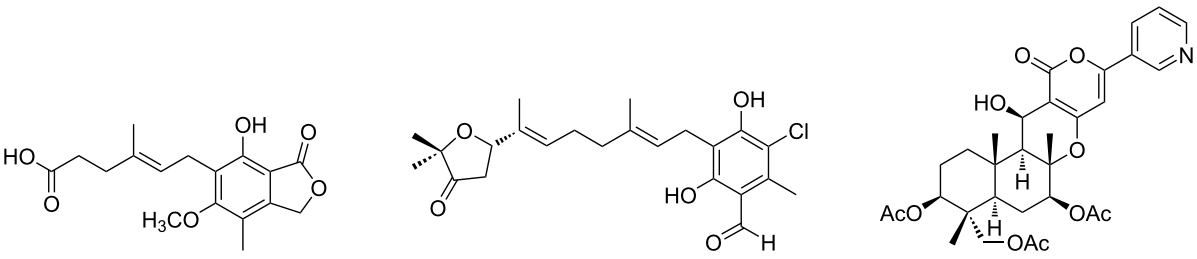



**Figure 1:** Examples of bioactive fungal meroterpenoids.

modifications, which mostly contribute to the skeletal diversity. Several terpenoid cyclases and αKG-dependent dioxygenases will be discussed as examples of engineering biosynthetic pathways and key enzymes involved in fungal meroterpenoid biosynthesis. Furthermore, a construction of the artificial biosynthetic pathway composed of the fungal meroterpenoids pathway and the pathway from other species, in fungal host *Aspergillus oryzae*, will be also introduced.

# Review

## Standard reactions of fungal meroterpenoid biosynthesis

The skeletal diversity within this group of compounds arises from polyketide synthesis, prenyl transfer, terpenoid cyclizations, and post-cyclization modifications [5]. In these reactions, terpenoid cyclizations and post-cyclization modifications are especially important in fungal meroterpenoid biosynthesis, contributing to the structural diversity of this class of compounds. Fungal meroterpenoid cyclases (CYCs) are seven-membrane-spanning transporter-like enzymes that regulate the conformations of cationic intermediates, leading to the terpenoid structure of each product [6,7]. Particular attention has been paid to the biosynthesis of the compounds derived from farnesyl-DMOA (**5**) composed of 3,5-dimethylorsellinic acid (DMOA, **4**) and the C15 terpenoid moiety due to their structural diversity (Figure 2).

In 2012, the biosynthesis of terretonin, a meroterpenoid derived from compound **5** and produced by the filamentous fungus *Aspergillus terreus*, was investigated [8,9] using the heterologous expression host *Aspergillus oryzae* strain NSAR1, a filamentous fungus developed by multiple mutations of metabolic genes for ease of gene transfer and high substance production capabilities [10,11].

The expression of *trt4* (polyketide synthase, PKS), *trt2* (prenyltransferase, PT), *trt5* (methyltransferase, MT), *trt8* (flavin-dependent monooxygenase, FMO), and *trt1* (meroterpenoid cyclase, CYC) in *A. oryzae* NSAR1 led to the production of preterretonin A (**7**) (Figure 2) [9]. These data indicated that Trt1 protonates the epoxide of (10'*R*)-epoxyfarnesyl-DMOA-3,5-methyl ester (**6**), leading to the cyclization of the terpenoid moiety in the chair–chair conformation, and catalyzes the deprotonation of H-9' of the resulting cation intermediate at C-4' to induce an acyl shift, forming the steroid-like structure of **7** with a 6-6-6-5 ring (Figure 2).

## Swapping terpenoid cyclases in heterologous expression systems

A search of the genome database for Trt1-homolog CYCs revealed the enzyme AusL (41% identity with Trt1) in *Aspergillus nidulans*, within the same phylogenetic clade. Expression of this enzyme in place of Trt1 resulted in the formation of the 6-6-6-6-membered ring protoaustinoid A (**8**) (Figure 2) [9]. In addition, the expression of the cyclase AdrI (38% identity with Trt1) from *Penicillium chrysogenum* produced the 6-6-6-5-membered andrastin E (**9**) (Figure 2) [8]. Like Trt1, AusL and AdrI create the common cation intermediate from **6**, but they deprotonate the cationic intermediate from H-1' and H-11, respectively [12]. The differences in the structural bases of Trt1, AusL, and AdrI are quite intriguing, in that they have similar substrate binding modes but different regional specificities.

The rapid increase in genomic information has further expanded the number of available enzyme genes, thus increasing the potential for diversifying DMOA-derived meroterpenoid biosynthesis. Recently, a similar genome mining approach led to the discovery of new CYCs, InsA7 (32% identity with Trt1) and InsB2 (40% identity with Trt1) from *Aspergillus insuetus* [13].

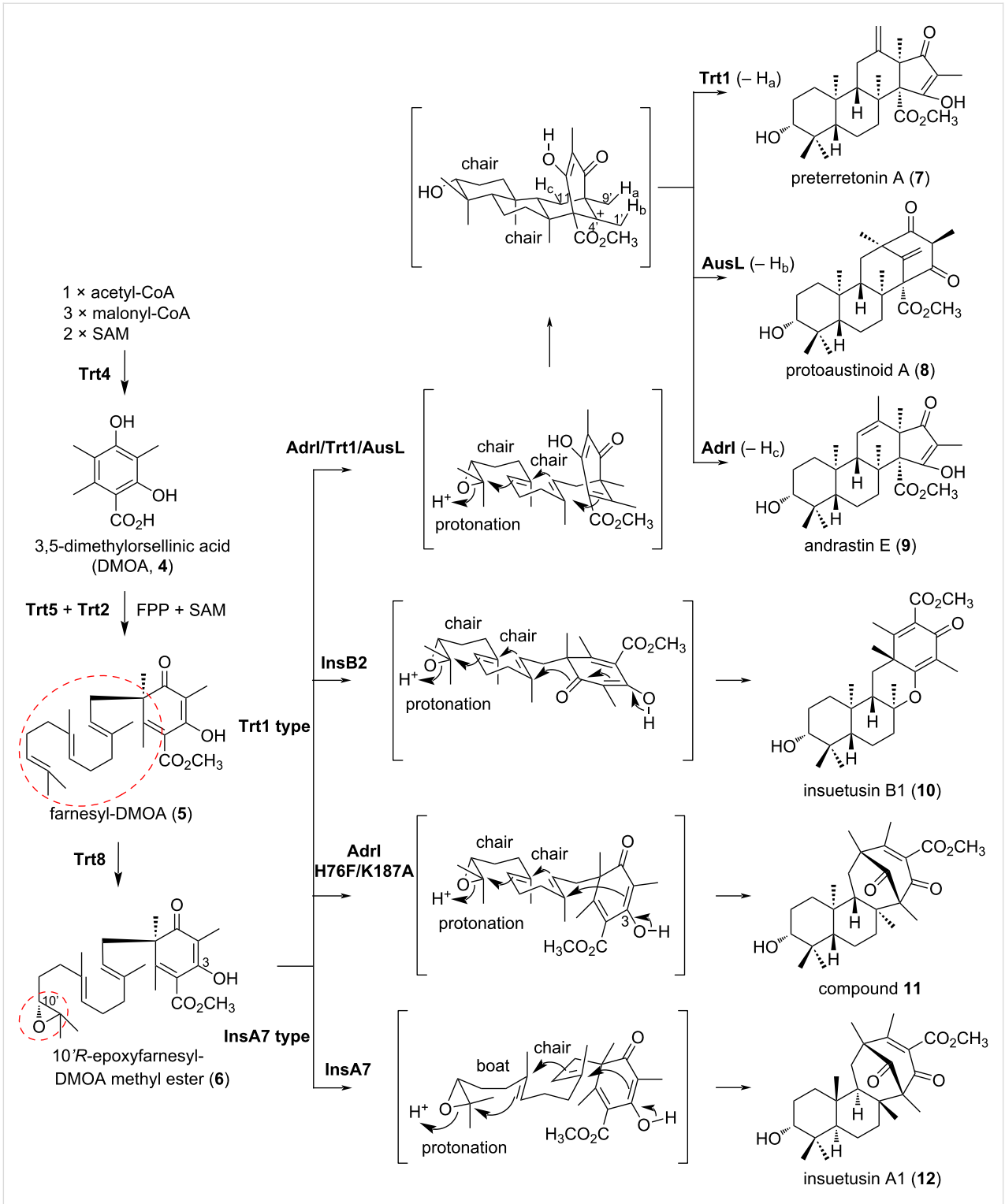



**Figure 2:** The diversity of DMOA-derived meroterpenoid biosyntheses.

These enzymes catalyze the formation of the terpenoid skeleton from **6**, adopting chair–boat and chair–chair conformations to create two distinct 6-6-6-6-membered ring meroterpenoids: insuetusin A1 (**12**) and insuetusin B1 (**10**), respectively (Figure 2) [9]. Like other Trt1-type enzymes, InsB2 catalyzes the protonation of the epoxide, the formation of two six-membered rings in a chair–chair conformation, but the reaction finishes with the deprotonation of the hydroxy group at C-3 to

produce compound **10**. In contrast, InsA7 commonly initiates and terminates the reaction with the protonation of the epoxide and the deprotonation of OH-3, respectively, but it produces product **12** via a boat–chair conformation. Since all of the other Trt1-like cyclases catalyze the reaction with chair–chair stereocontrol, the structure basis of InsA7 should be unique. A structural comparison of InsB2 and InsA7 will help to clarify the differences in their structural bases.

Recent advances in artifical intelligence (AI) have even made the structural modeling of membrane-bound proteins possible. In a recent study, a structural model of AdrI was constructed by AlphaFold2 [14], and its product **9** was docked into the model [15]. In the complex structure, H76 and K187 are located close to the D ring of **9**, and are expected to determine the selectivity of the product. The AdrI H76F/K187A variant was created, and produced the new compound **11**, as expected. In this mutant, deprotonation occurs from the OH-3 of the folded intermediate, and the altered cyclization mode produces **11** (Figure 2).

## Reconstitution based on the several meroterpenoid pathways in heterologous expression systems

Recently, the combinatorial biosynthesis of diterpene pyrones, derived from a pyrone skeleton and a C20 terpenoid skeleton, has been achieved by combining multiple biosynthetic pathways in a fungal heterologous expression host [16]. First, *subA* (PKS), *sub*C (PT), *subD* (GGPP synthase), *subE* (FMO), and *subB* (CYC) were expressed in *A. oryzae* to produce the intermediate **15**. Next, genes encoding a short-chain oxidoreductase (SDR), methyltransferase (MT), cytochrome P450 oxygenase (P450), and FMO from various fungi were additionally expressed in *A. oryzae* expessing *subABCDE* to produce 22 bioactive meroterpenoids, including the known antitumor compound subglutinol A (**16**) (Figure 3A and B) [17]. Among the novel compounds isolated from the production system, some exhibited intruiging pharmacological activities, such as antitumor (**16**), anti-HIV activity (**17**), and anti-Alzheimer's disease properties (**18**). Furthermore, FMO EsdpE, which epoxidizes the C-14, C-15 double bond of **13**, was employed to produce **19**, and the CYC EsdpB cyclizes in a chair–chair–chair conformation to form the 6-6-6 ring structure of **20** [18,19]. This difference in the epoxidation position represents a new point of biosynthetic diversity that will receive keen attention in the future.

## Production of pharmacologically active compounds by genetic engineering of terpenoid cyclization pathways

A single filamentous fungus reportedly synthesizes two meroterpenoid compounds via two different cyclization schemes. In the filamentous fungus *Acremonium egypticum*, both asochlorin (**22**) and ascofuranone (**23**) are synthesized from a common intermediate, ilicicolin A epoxide (**21**), but the backbones of the terpenoid moiety are different (Figure 4) [20]. In ascochlorin biosynthesis, intermediate **21** is accepted by AscF to form the hexanone ring of **22**. The cyclization process is initiated by the protonation of the epoxide, leading to the formation of a cation at C-7. A cascade of migrations then occurs to form the final backbone: H-6 migrates to the cation C-7, the methyl group at C-13 shifts to C-6, H-10 migrates to C-11, and finally, deprotonation of the hydroxy group terminates the cyclization [20,21]. In ascofuranone biosynthesis, compound **21** is initially hydroxylated at C-8, and then the hydroxylated product is cyclized via epoxidation by AscI to form the tetrahydrofuran ring of **23**. The biosynthetic pathways of **22** and **23** were elucidated through heterologous expression, gene disruption, and in vitro reactions. The *ascF* gene was disrupted with the aim toward the mass-production of **23**, and a system for producing quantities greater than 500 mg/L was successfully established, thus achieving an industrial-level substance production system [20].

## αKG-dependent dioxygenases – important enzymes involved in post-cyclization modifications

After terpenoid cyclization, skeletal modifications are performed by P450 monooxygenases, α-ketoglutarate (αKG)-dependent dioxygenases, isomerases, and acyltransferases [5,22-25]. In particular, the functions of the αKG-dependent dioxygenase have been analyzed in detail due to its relatively small molecular weight and the low costs of its cofactors: αKG, ascorbic acid, and iron ions. The αKG reacts with iron and molecular oxygen to form the highly reactive Fe(IV)=O via oxidative decarboxylation. This active molecular species withdraws a hydrogen atom, and the generated radical induces various reactions such as hydroxylation, unsaturation, epoxidation, halogenation, endoperoxidation, and C–C bond reconstruction, leading to the formation of diverse chemical structures [22,26-31].

### Structure-based mutagenesis of αKG-dependent dioxygenases

The first αKG-dependent dioxygenases characterized in fungal meroterpenoid biosynthesis, AusE and PrhA, share high identity (76% identity to each other) and accept the meroterpenoid preaustinoid A1 (**24**) to yield the products preaustinoid A3 (**27**) and berkeleydione (**28**), respectively (Figure 5) [32,33]. In the initial reaction, AusE abstracts H-2 of **24**, while PrhA abstracts H-5 of **24**. PrhA and AusE form a double bond to yield preaustinoid A2 (**25**) and berkeleyone B (**26**), respectively (Figure 5). AusE abstracts H-5 of **25**, forming a radical that

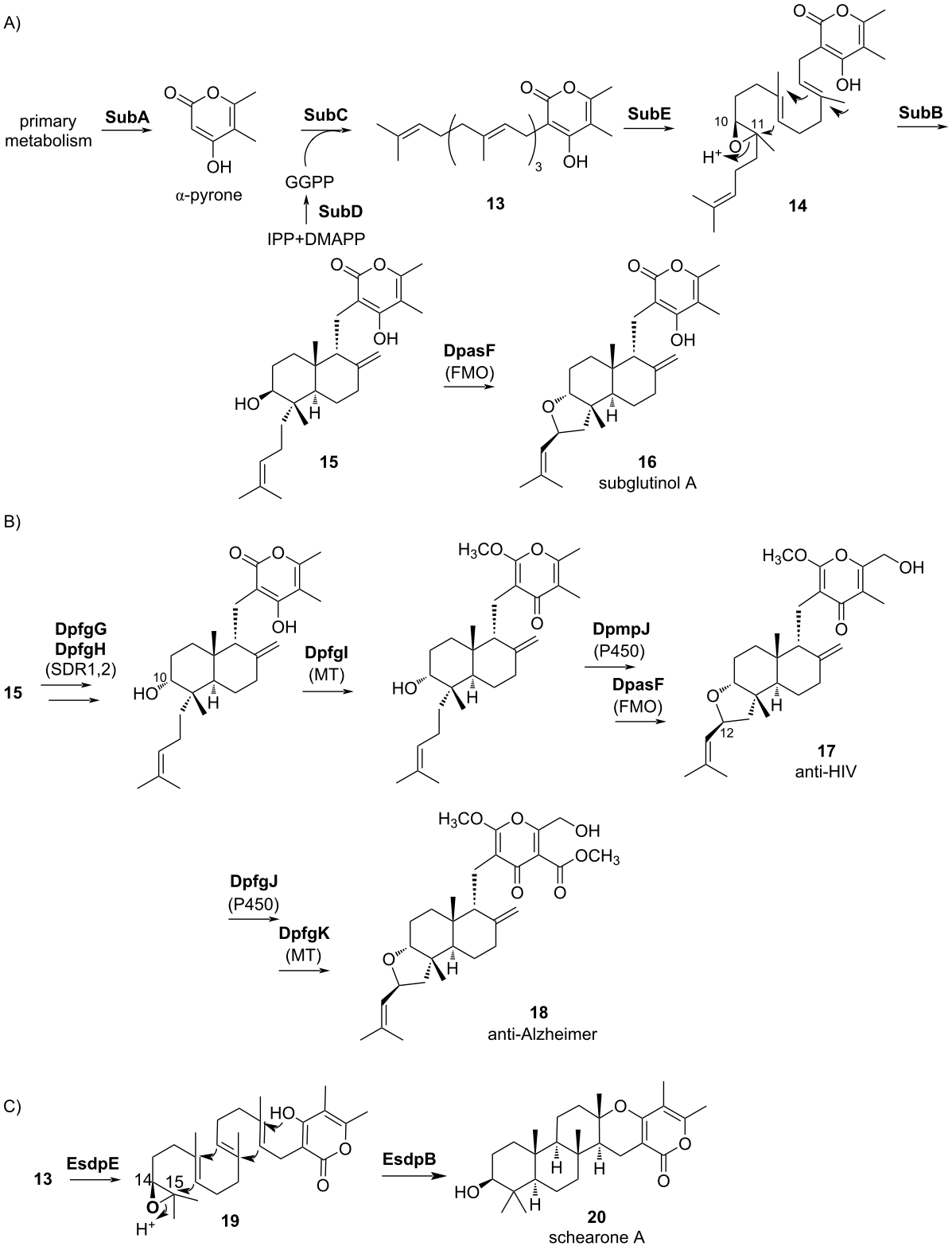



**Figure 3:** The combinatorial biosynthesis of diterpene pyrone meroterpenoids. The production of subglutinol A by SubABCDE + DpasF (A), the production of novel diterpene pyrones by SubABCDE + various tailoring enzymes (B), and the production of schearone A by SubACD + EsdpE + EsdpB (C).

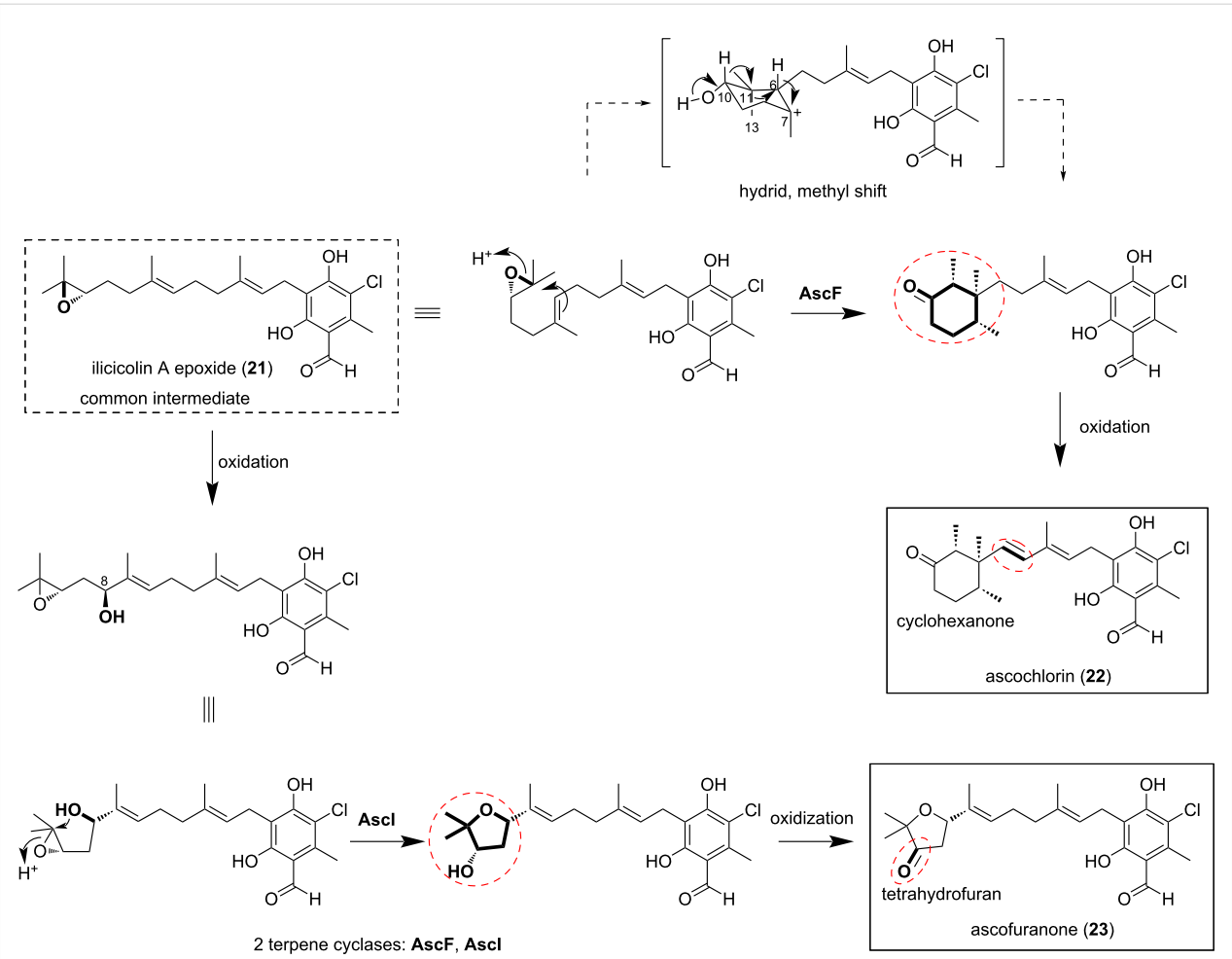



**Figure 4:** The biosynthetic reaction from the common intermediate **21** to ascochlorin (**22**) and ascofuranone (**23**), respectively.

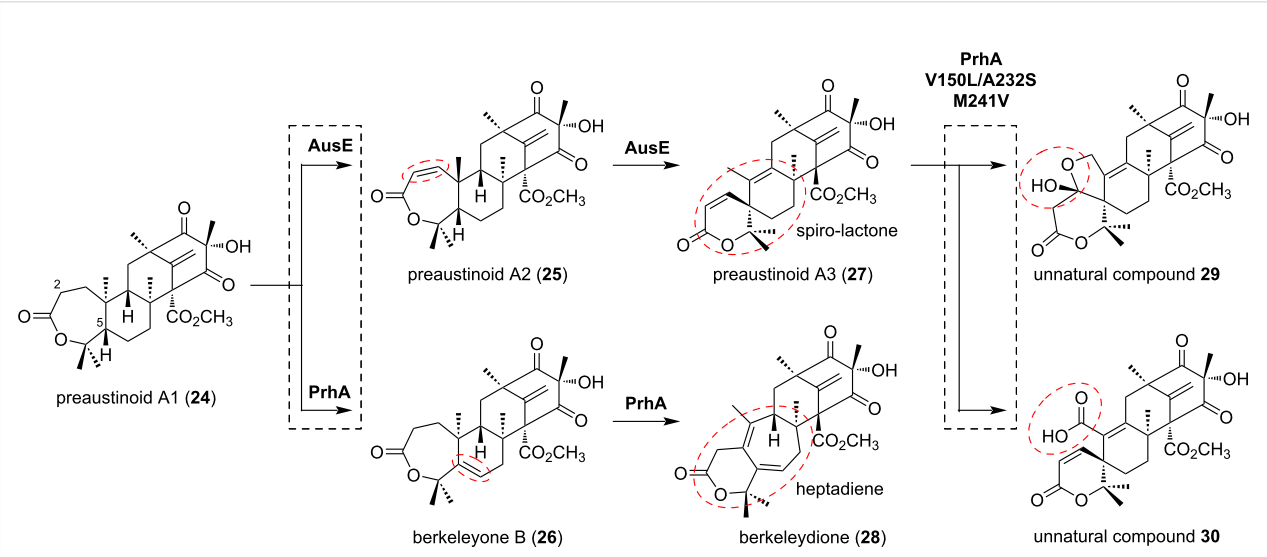



**Figure 5:** The multistep oxidations catalyzed by AusE and PrhA from the common intermediate **24**.

initiates a sequential reaction to produce the spirolactone **27**. In contrast, PrhA dehydrogenates H-1 of **26** to form the heptadiene **28**. Comparing the active centers, only three amino acid residues differ between AusE and PrhA: L150(AusE)/V150(PrhA), S232(AusE)/A232(PrhA), and V241(AusE)/M241(PrhA). By swapping these residues, the catalytic activities of these enzymes were exchanged [34]. From **24**, PrhA V150L/A232S produced **27**, and AusE L150V/S232A produced **28**. Surprisingly, PrhA V150L/A232S/M241V catalyzed additional oxidations of **27** to produce compounds **29** and **30**, as unnatural products. As demonstrated in these reactions, the terpenoid skeleton undergoes significant structural changes due to radical formation through oxidase-induced hydrogen atom abstraction. Close examinations of the substrate complex structures of these αKG-dependent dioxygenases involved in these meroterpenoid oxidations revealed that the terpenoid moiety is held in a substrate pocket consisting of hydrophobic amino acids, and the polyketide moiety is retained by a loop structure serving as a lid. The conformation of the terpenoid moiety can be easily altered by mutations to the enzyme's substrate retention site, thereby changing the position of hydrogen atom abstraction and the subsequent skeletal reorganization reaction mode, resulting in the synthesis of a variety of skeletons. Thus, the meroterpenoid αKG-dependent dioxygenase is a highly potent biosynthetic enzyme, and the structure of the product can be easily altered by mutagenesis.

## Application of diverse unnatural substrates to αKG-dependent dioxygenases

SptF, an α-KG dioxygenase, represents another example of structure-based engineering to create multiple products from a single enzyme reaction [35]. The natural substrates of SptF, andiconin D (**31**) and andilesin C (**32**) derived from andiconin, can be converted into emervaridone B (**33**) and anditomin (**34**), respectively (Figure 6). Subsequently, **33** can be further modified into emervaridone C (**35**), which can be oxidized to **37** or **38** (Figure 6). In contrast, **34** is converted to compounds **36a** and **36b** (Figure 6). These multistep reactions with several substrates illustrate the broad substrate specificity of SptF.

To apply the high potential of SptF to generate a variety of oxidation products, meroterpenoids (**39**, **41**, **42**, **45**, and **46**) and terpenoids (**49** and **52**) were also used as unnatural substrates to construct a series of new compounds (**40**, **43**, **44**, **47**, **48**, **50**, **51**, and **53**–**55**) (Figure 7A). In addition, a structure-based mutagenesis study of SptF was performed to further amplify its catalytic potential. Firstly, the hydrophobic residues Ile63, Phe133, and Ile231, which compose the substrate binding site of SptF, were mutated. As a result, SptF F133A oxidized **31** into **56**, while I231A oxidized the same substrate **31** into compound **57** (Figure 7B). Then, the hydrophilic residues Asn65, Ser114, Thr148, and Asn150, which also line the cavity, were also mutated. Compound **31** was converted into **38** by SptF N150A,

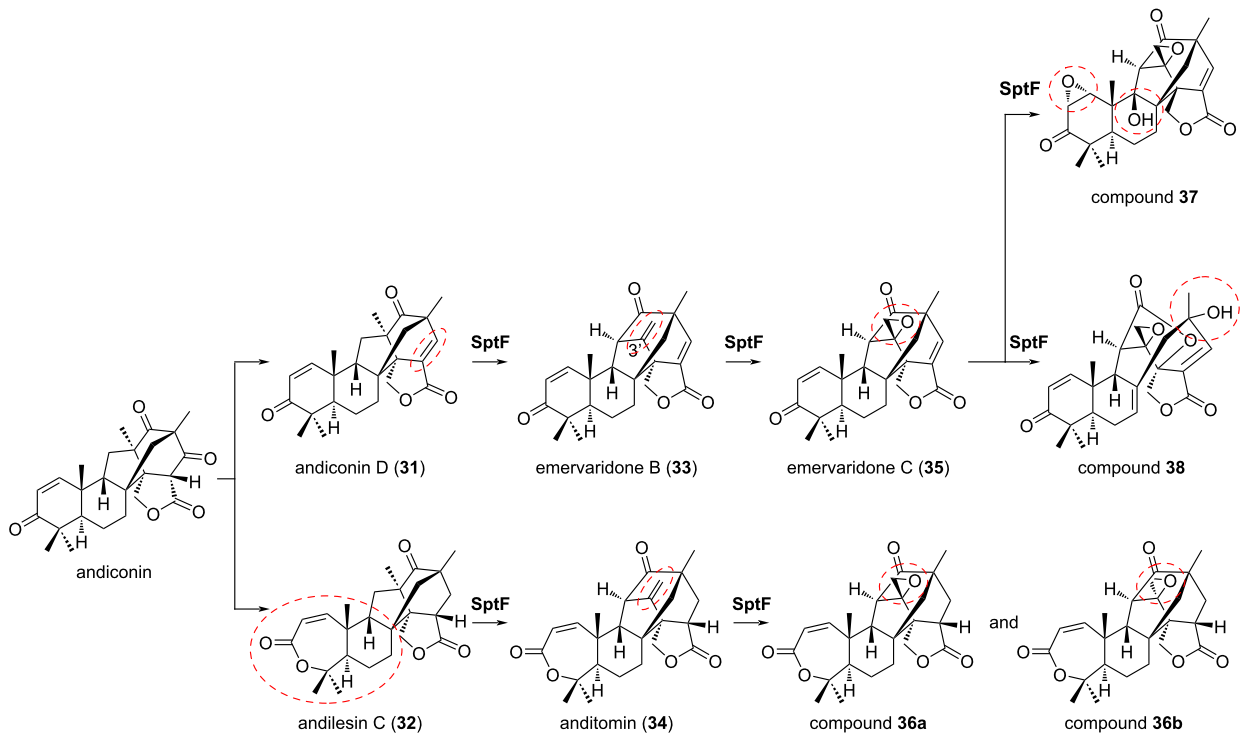



**Figure 6:** Reactions of SptF with native substrates **31** and **32**.

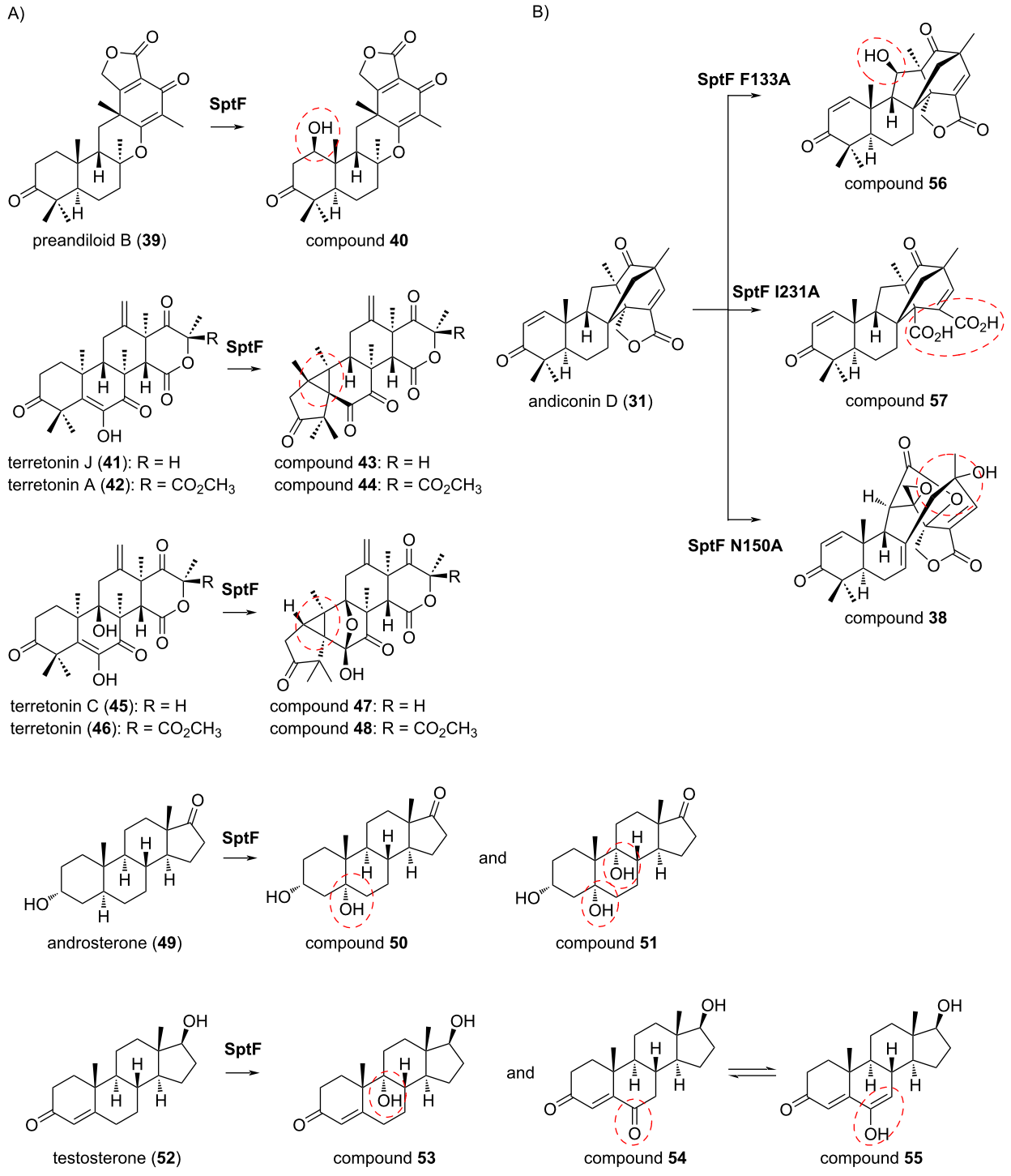



**Figure 7:** A) Reactions of SptF with unnatural substrates. B) Reactions of SptF variants with **31**.

while the WT enzyme converted **31** into **37** and **38** (Figure 7B). These results exemplify the production of oxidized terpenoids and meroterpenoids through the use of non-natural substrates and the application of structure-based mutagenesis.

## Evolution of αKG-dependent dioxygenase with saturated mutagenesis

As meroterpenoids readily change their reaction products depending on the conformation of the terpenoid skeleton, the

regiospecificity of the oxidation reaction can be modified by introducing random mutations in the substrate-binding site of αKG-dependent dioxygenase. The αKG-dependent dioxygenase AndA withdraws H-12 of preandiloid C (**58**), and generates a radical leading to the construction of the bicyclo[2.2.2]octane skeleton of andiconin (**59**) (Figure 8) [36,37]. Two amino acid pairs, M119/N121 and A228/A230, in its substrate-binding pocket were mutated by saturation mutagenesis using degenerate primers, resulting in an enzyme that oxidizes C-5, thereby reacting with the electron between C-1 and C-2 [38]. In the future, such mutagenesis strategies will be applied to a wider range of oxidases to alter their reactivities and create an expanded range of products.

### Construction of artificial fungi–plant hybrid biosynthetic pathways

Furthermore, there are examples of the enzymatic synthesis of plant-derived pharmaceutical meroterpenoids through the heterologous expression of a combination of fungal and plant biosynthetic enzymes. The biosynthetic enzymes StbA (polyketide synthase, PKS) and StbC (prenyltransferase, PT) from the filamentous fungus *Stachbotrys bisbyi* were produced in *A. oryzae* to synthesize grifolic acid (**60**) (Figure 9) [39]. In addition, the gene encoding DCAS, a prenyl group oxidase from the plant *Rhododendron dauricum*, was additionally expressed in *A. oryzae* expressing *stbAC* after codon-optimization to produce daurichromenic acid (**61**) (Figure 9), a plant-derived meroterpenoid that exhibits anti-HIV activity [40]. Furthermore, by expressing AscD, a fungal halogenase from the ascochlorin biosynthetic pathway, the authors succeeded in biosynthesizing **62**, an unnatural meroterpneoid, chloro-daurichromenic acid, halogenated at C-3 (Figure 9). Today, with the improvement of gene synthesis and expression technologies, genes from a wide variety of species are available, further increasing the potential for applications of biogenetic resources. In the future, advances in synthetic biology will enable the biosynthesis of bioactive compounds with still greater skeletal diversity.

## Conclusion

In summary, many novel terpene cyclases and αKG-dependent dioxygenases were discovered by recent developments in genome mining approaches. Unique meroterpenoids have been generated by integrating these enzymes into meroterpenoid biosynthetic pathways, enzymatic engineering, and feeding with

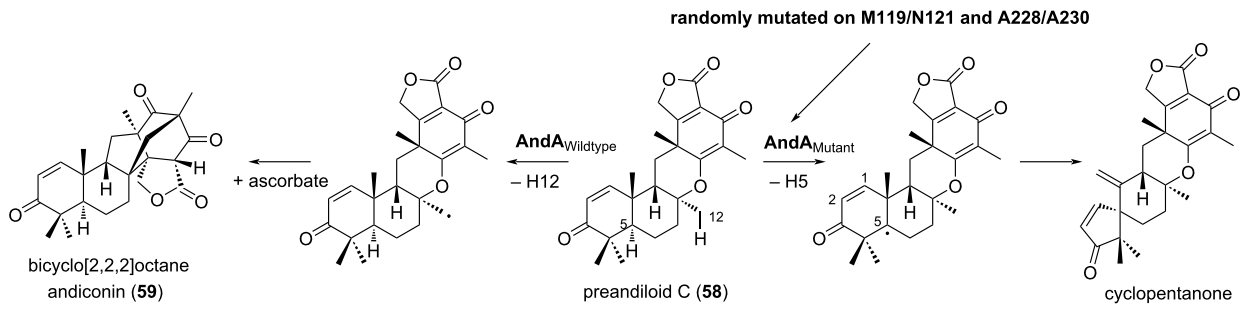



**Figure 8:** The reaction of the αKG enzyme AndA and its variants generated via saturated mutagenesis.

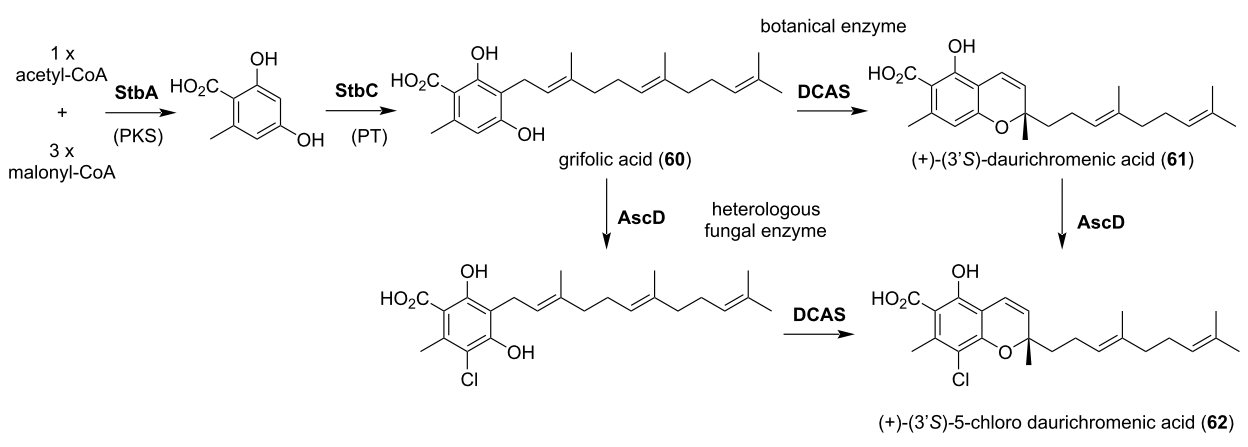



**Figure 9:** The synthetic biological production of daurichromenic acid and its halogenated derivative.

unnatural substrates. Enzyme crystal structure analysis and AI structure prediction also facilitated these investigations. Besides terpene cyclases and αKG-dependent dioxygenases, P450 monooxygenases and UbiA-type prenyltransferases have also expanded the variety of meroterpenoids [41,42]. These untapped membrane enzymes will be applicable for engineered biosynthesis, with structural models providing useful information for mutagenesis, as in the engineering of meroterpenoid cyclases [11] and fungal P450 oxygenases catalyzing cross-coupling between two aromatic rings [43]. In the future, further developments in bioinformatics, structural biology, and AI techniques will enable the design of biosynthetic enzymes and pathways to produce desired bioactive compounds, although understanding the chemistry catalyzed by individual enzymes will remain important. Furthermore, highly productive substance production systems will be developed through host, metabolome, and enzyme improvements by genome editing technology.

## Funding

This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (JSPS KAKENHI Grant Number JP21H02636 and JP22H05123), AMED (Grant Number JP21ak0101164) from Japan Science and Technology Agency, The Naito foundation, Nagase Science Technology Foundation, Yamada Science foundation, and The Mitsubishi Foundation.

## ORCID® iDs

Zhiyang Quan - https://orcid.org/0009-0009-9507-8794
Takayoshi Awakawa - https://orcid.org/0000-0003-2575-3175

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

1. Geris, R.; Simpson, T. J. *Nat. Prod. Rep.* **2009,** *26,* 1063–1094. doi:10.1039/b820413f
2. Wu, J. C. *Perspect. Drug Discovery Des.* **1994,** *2,* 185–204. doi:10.1007/bf02171743
3. Nakamura, K.; Fujioka, S.; Fukumoto, S.; Inoue, N.; Sakamoto, K.; Hirata, H.; Kido, Y.; Yabu, Y.; Suzuki, T.; Watanabe, Y.-i.; Saimoto, H.; Akiyama, H.; Kita, K. *Parasitol. Int.* **2010,** *59,* 560–564. doi:10.1016/j.parint.2010.07.006
4. Tomoda, H.; Kim, Y. K.; Nishida, H.; Masuma, R.; Ōmura, S. *J. Antibiot.* **1994,** *47,* 148–153. doi:10.7164/antibiotics.47.148
5. Matsuda, Y.; Abe, I. *Nat. Prod. Rep.* **2016,** *33,* 26–53. doi:10.1039/c5np00090d
6. Itoh, T.; Tokunaga, K.; Matsuda, Y.; Fujii, I.; Abe, I.; Ebizuka, Y.; Kushiro, T. *Nat. Chem.* **2010,** *2,* 858–864. doi:10.1038/nchem.764
7. Barra, L.; Abe, I. *Nat. Prod. Rep.* **2021,** *38,* 566–585. doi:10.1039/d0np00056f
8. Itoh, T.; Tokunaga, K.; Radhakrishnan, E. K.; Fujii, I.; Abe, I.; Ebizuka, Y.; Kushiro, T. *ChemBioChem* **2012,** *13,* 1132–1135. doi:10.1002/cbic.201200124
9. Matsuda, Y.; Awakawa, T.; Itoh, T.; Wakimoto, T.; Kushiro, T.; Fujii, I.; Ebizuka, Y.; Abe, I. *ChemBioChem* **2012,** *13,* 1738–1741. doi:10.1002/cbic.201200369
10. Jin, F.; Maruyama, J.; Juvvadi, P.; Arioka, M.; Kitamoto, K. *Biosci., Biotechnol., Biochem.* **2004,** *68,* 656–662. doi:10.1271/bbb.68.656
11. Awakawa, T.; Abe, I. *J. Fungi* **2021,** *7,* 486–495. doi:10.3390/jof7060486
12. Matsuda, Y.; Awakawa, T.; Abe, I. *Tetrahedron* **2013,** *69,* 8199–8204. doi:10.1016/j.tet.2013.07.029
13. Tang, J.; Matsuda, Y. *Chem. Sci.* **2022,** *13,* 10361–10369. doi:10.1039/d2sc02994d
14. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature* **2021,** *596,* 583–589. doi:10.1038/s41586-021-03819-2
15. Tang, J.; Matsuda, Y. *Angew. Chem., Int. Ed.* **2023,** e202306046. doi:10.1002/anie.202306046
16. Tsukada, K.; Shinki, S.; Kaneko, A.; Murakami, K.; Irie, K.; Murai, M.; Miyoshi, H.; Dan, S.; Kawaji, K.; Hayashi, H.; Kodama, E. N.; Hori, A.; Salim, E.; Kuraishi, T.; Hirata, N.; Kanda, Y.; Asai, T. *Nat. Commun.* **2020,** *11,* 1830. doi:10.1038/s41467-020-15664-4
17. Lee, J. C.; Lobkovsky, E.; Pliam, N. B.; Strobel, G.; Clardy, J. *J. Org. Chem.* **1995,** *60,* 7076–7077. doi:10.1021/jo00127a001
18. Morishita, Y.; Tsukada, K.; Murakami, K.; Irie, K.; Asai, T. *J. Nat. Prod.* **2022,** *85,* 384–390. doi:10.1021/acs.jnatprod.1c00973
19. Xiao, Z.-H.; Dong, J.-Y.; Li, A.; Dai, J.-M.; Li, Y.-P.; Hu, Q.-F.; Shao, L.-D.; Matsuda, Y.; Wang, W.-G. *Org. Chem. Front.* **2022,** *9,* 1837–1843. doi:10.1039/d2qo00055e
20. Araki, Y.; Awakawa, T.; Matsuzaki, M.; Cho, R.; Matsuda, Y.; Hoshino, S.; Shinohara, Y.; Yamamoto, M.; Kido, Y.; Inaoka, D. K.; Nagamune, K.; Ito, K.; Abe, I.; Kita, K. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116,* 8269–8274. doi:10.1073/pnas.1819254116
21. Quan, Z.; Awakawa, T.; Wang, D.; Hu, Y.; Abe, I. *Org. Lett.* **2019,** *21,* 2330–2334. doi:10.1021/acs.orglett.9b00616
22. Awakawa, T.; Mori, T.; Ushimaru, R.; Abe, I. *Nat. Prod. Rep.* **2023,** *40,* 46–61. doi:10.1039/d2np00014h
23. Zhang, X.; Guo, J.; Cheng, F.; Li, S. *Nat. Prod. Rep.* **2021,** *38,* 1072–1099. doi:10.1039/d1np00004g
24. Mori, T.; Iwabuchi, T.; Hoshino, S.; Wang, H.; Matsuda, Y.; Abe, I. *Nat. Chem. Biol.* **2017,** *13,* 1066–1073. doi:10.1038/nchembio.2443
25. Matsuda, Y.; Awakawa, T.; Mori, T.; Abe, I. *Curr. Opin. Chem. Biol.* **2016,** *31,* 1–7. doi:10.1016/j.cbpa.2015.11.001
26. Gao, S.-S.; Naowarojna, N.; Cheng, R.; Liu, X.; Liu, P. *Nat. Prod. Rep.* **2018,** *35,* 792–837. doi:10.1039/c7np00067g
27. Nakamura, H.; Matsuda, Y.; Abe, I. *Nat. Prod. Rep.* **2018,** *35,* 633–645. doi:10.1039/c7np00055c
28. Ushimaru, R.; Abe, I. *ACS Catal.* **2023,** *13,* 1045–1076. doi:10.1021/acscatal.2c05247

29. Matsuda, Y.; Bai, T.; Phippen, C. B. W.; Nødvig, C. S.; Kjærbølling, I.; Vesth, T. C.; Andersen, M. R.; Mortensen, U. H.; Gotfredsen, C. H.; Abe, I.; Larsen, T. O. *Nat. Commun.* **2018,** *9,* 2587. doi:10.1038/s41467-018-04983-2
30. Mori, T.; Zhai, R.; Ushimaru, R.; Matsuda, Y.; Abe, I. *Nat. Commun.* **2021,** *12,* 4417. doi:10.1038/s41467-021-24685-6
31. Li, X.; Awakawa, T.; Mori, T.; Ling, M.; Hu, D.; Wu, B.; Abe, I. *J. Am. Chem. Soc.* **2021,** *143,* 21425–21432. doi:10.1021/jacs.1c11548
32. Matsuda, Y.; Awakawa, T.; Wakimoto, T.; Abe, I. *J. Am. Chem. Soc.* **2013,** *135,* 10962–10965. doi:10.1021/ja405518u
33. Matsuda, Y.; Iwabuchi, T.; Fujimoto, T.; Awakawa, T.; Nakashima, Y.; Mori, T.; Zhang, H.; Hayashi, F.; Abe, I. *J. Am. Chem. Soc.* **2016,** *138,* 12671–12677. doi:10.1021/jacs.6b08424
34. Nakashima, Y.; Mori, T.; Nakamura, H.; Awakawa, T.; Hoshino, S.; Senda, M.; Senda, T.; Abe, I. *Nat. Commun.* **2018,** *9,* 104. doi:10.1038/s41467-017-02371-w
35. Tao, H.; Mori, T.; Chen, H.; Lyu, S.; Nonoyama, A.; Lee, S.; Abe, I. *Nat. Commun.* **2022,** *13,* No. 95. doi:10.1038/s41467-021-27636-3
36. Matsuda, Y.; Wakimoto, T.; Mori, T.; Awakawa, T.; Abe, I. *J. Am. Chem. Soc.* **2014,** *136,* 15326–15336. doi:10.1021/ja508127q
37. Nakashima, Y.; Mitsuhashi, T.; Matsuda, Y.; Senda, M.; Sato, H.; Yamazaki, M.; Uchiyama, M.; Senda, T.; Abe, I. *J. Am. Chem. Soc.* **2018,** *140,* 9743–9750. doi:10.1021/jacs.8b06084
38. Mori, T.; Nakashima, Y.; Chen, H.; Hoshino, S.; Mitsuhashi, T.; Abe, I. *Chem. Commun.* **2022,** *58,* 5510–5513. doi:10.1039/d2cc00736c
39. Li, C.; Matsuda, Y.; Gao, H.; Hu, D.; Yao, X. S.; Abe, I. *ChemBioChem* **2016,** *17,* 904–907. doi:10.1002/cbic.201600087
40. Okada, M.; Saito, K.; Wong, C. P.; Li, C.; Wang, D.; Iijima, M.; Taura, F.; Kurosaki, F.; Awakawa, T.; Abe, I. *Org. Lett.* **2017,** *19,* 3183–3186. doi:10.1021/acs.orglett.7b01288
41. Qi, B.-w.; Li, N.; Zhang, B.-b.; Zhang, Z.-k.; Wang, W.-j.; Liu, X.; Wang, J.; Awakawa, T.; Tu, P.-f.; Abe, I.; Shi, S.-p.; Li, J. *Org. Lett.* **2022,** *24,* 2526–2530. doi:10.1021/acs.orglett.2c00684
42. Zhang, K.; Zhang, G.; Hou, X.; Ma, C.; Liu, J.; Che, Q.; Zhu, T.; Li, D. *Org. Lett.* **2022,** *24,* 2025–2029. doi:10.1021/acs.orglett.2c00495
43. Zetzsche, L. E.; Yazarians, J. A.; Chakrabarty, S.; Hinze, M. E.; Murray, L. A. M.; Lukowski, A. L.; Joyce, L. A.; Narayan, A. R. H. *Nature* **2022,** *603,* 79–85. doi:10.1038/s41586-021-04365-7

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.50

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# A *myo*-inositol dehydrogenase involved in aminocyclitol biosynthesis of hygromycin A

Michael O. Akintubosun and Melanie A. Higgins*

Full Research Paper

Open Access

Address:
Department of Biological Sciences, The University of Alabama, 3314 Science and Engineering Complex, Tuscaloosa, AL 35487, USA

Email:
Melanie A. Higgins* - mahiggins1@ua.edu

* Corresponding author

Keywords:
aminocyclitol; biosynthesis; hygromycin A; inositol dehydrogenase; *myo*-inositol

*Beilstein J. Org. Chem.* **2024,** *20,* 589–596.
https://doi.org/10.3762/bjoc.20.51

Received: 15 December 2023
Accepted: 07 March 2024
Published: 14 March 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf

© 2024 Akintubosun and Higgins; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Hygromycin A is a broad-spectrum antibiotic that contains a furanose, cinnamic acid, and aminocyclitol moieties. The biosynthesis of the aminocyclitol has been proposed to proceed through six enzymatic steps from glucose 6-phosphate through *myo*-inositol to the final methylenedioxy-containing aminocyclitol. Although there is some in vivo evidence for this proposed pathway, biochemical support for the individual enzyme activities is lacking. In this study, we verify the activity for one enzyme in this pathway. We show that Hyg17 is a *myo*-inositol dehydrogenase that has a unique substrate scope when compared to other *myo*-inositol dehydrogenases. Furthermore, we analyze sequences from the protein family containing Hyg17 and discuss genome mining strategies that target this protein family to identify biosynthetic clusters for natural product discovery.

## Introduction

Hygromycin A is a natural product that was discovered in the 1950s and is produced by the soil bacterium *Streptomyces hygroscopicus* [1]. It has broad spectrum antibiotic activity, antitreponemal activity against the pathogen that causes swine dysentery, and selective activity against the spirochete that causes Lyme disease [1-3]. It binds the large 50S ribosomal subunit to inhibit ribosomal peptidyl transferase activity [4,5] and contains three distinctive functional groups: furanose, cinnamic acid, and aminocyclitol (Figure 1). The cinnamic acid and aminocyclitol moieties directly restrict access of the amino-tRNA to inhibit peptidyl transferase activity while the furanose group does not appear to be important for target inhibition [6,7]. In addition, hygromycin A contains a unique methylenedioxy group found on the aminocyclitol that is not required for ribosome interaction and in vitro inhibition [8]. Instead, it is essential for in vivo antimicrobial activity suggesting a distinct biological function independent of ribosome binding.

The hygromycin A biosynthetic gene cluster has been identified and the biosynthesis of the aminocyclitol has been pro-

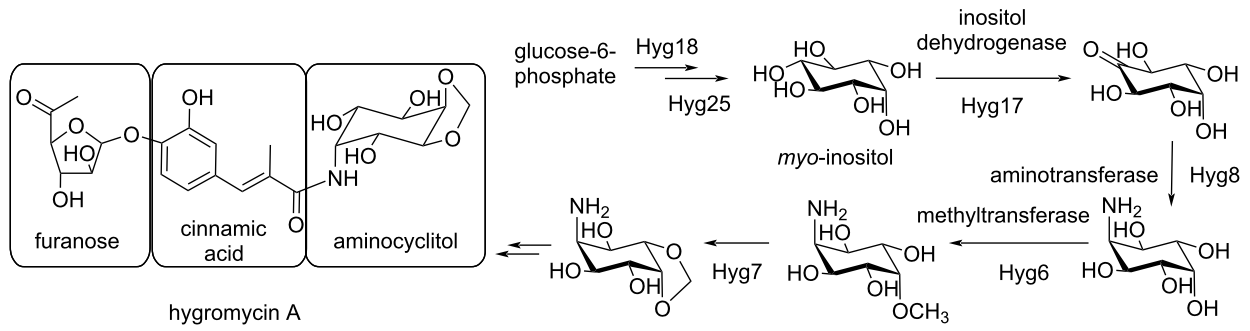



**Figure 1:** Proposed biosynthetic pathway for the aminocyclitol from hygromycin A.

posed (Figure 1) [8,9]. Starting from glucose-6-phosphate, the pathway is hypothesized to proceed through six steps to the final methylenedioxy-containing aminocyclitol. First, glucose-6-phosphate is thought to be converted to *myo*-inositol by the *myo*-inositol-1-phosphate synthase Hyg18 and phosphatase Hyg25. *Myo*-inositol is then proposed to be oxidized by the dehydrogenase Hyg17 to form *neo*-inosose, followed by a transamination to *neo*-inosamine-2 by the aminotransferase Hyg8. The methyltransferase Hyg6 would then install a methyl group which would set up cyclization of the methylenedioxy group by Hyg7. This biosynthetic pathway has been proposed based on gene annotations and in vivo studies [8]. However, validation by in vitro approaches or biochemical analysis of the individual enzymes is lacking. Here, we verify that Hyg17 is a *myo*-inositol dehydrogenase and show that it has a distinct substrate scope. In addition, we use sequence similarity networks to compare Hyg17 sequences with other members of the oxidoreductase family and inositol dehydrogenases and discuss specialized genome mining approaches using these sequences to identify new natural product biosynthetic clusters.

# Results and Discussion

## Hyg17 enzyme activity

We found that Hyg17 formed inclusion bodies when recombinantly produced by various *E. coli* expression strains. However, we were able to obtain pure soluble protein when expressing Hyg17 in a *Rhodococcus* expression system (Supporting Information File 1, Figure S1) [10,11]. According to the proposed hygromycin A biosynthetic pathway, Hyg17 is a *myo*-inositol dehydrogenase. These types of enzymes typically use $NAD^+$ as a cofactor [12,13]. So, we first tested Hyg17 with *myo*-inositol and $NAD^+$ and found that it was able to produce NADH, suggesting it can function as a *myo*-inositol dehydrogenase (Figure 2a). Since this assay tests for the formation of NADH, we are assuming the formation of a ketone product. However, further experiments are required to verify this assumption and determine which inositol position is being oxidized. When we tested Hyg17 with *myo*-inositol and NADP, we found no activity, showing that Hyg17 is $NAD^+$-dependent (Figure 2b). Although this is consistent with native *myo*- and *scyllo*-dehydrogenases LcIDH1 and LcIDH2 from *Lactobacillus casei* [12], there have been reports of a genetically engineered NAD-specific dehydrogenase that has been converted to an efficient NADP-dependent enzyme [14]. Next, we determined the optimal pH of Hyg17. We observed increased activity as the pH increased, with the greatest activity observed at pH 10.5–11 (Figure 2c and Supporting Information File 1, Figure S2). Similarly, LcIDH1 and LcIDH2 have an optimum pH of 9.3 and 9.5, respectively, while the *myo*-inositol dehydrogenase from *Bacillus subtilis*, BsIDH, has an optimal pH between 9.5–10 [12,13]. We also compared product formation between reactions with Hyg17 or BsIDH and *myo*-inositol using thin-layer chromatography (Supporting Information File 1, Figure S3). We found that both enzymes generated a ketone product with identical retention factors. Further studies are required to determine if Hyg17 oxidizes the same position of *myo*-inositol as BsIDH and other *myo*-inositol dehydrogenases.

We then tested the substrate scope for Hyg17. Activity was best using *myo*-inositol as substrate further validating that Hyg17 is a *myo*-inositol dehydrogenase (Figure 2a). We also found there was reduced activity with *scyllo*-inositol, minimal activity with L-*chiro*-inositol and no activity with D-glucose, D-*chiro*-inositol, *epi*-inositol, *muco*-inositol, and *allo*-inositol. By comparison, other *myo*-inositol dehydrogenases typically do not have activity on *scyllo*-inositol but can have reduced activity on D-*chiro*-inositol, D-glucose, D-xylose and 4-*O*-benzyl-*myo*-inositol [12,13,15-17]. However, *scyllo*-inositol dehydrogenases are active on *scyllo*-inositol and *myo*-inositol to a lesser extent [16,18]. Altogether, this suggests that Hyg17 can accommodate different substrates when compared to known *myo*-inositol dehydrogenases and has a more similar substrate scope to *scyllo*-inositol dehydrogenases.

We performed kinetics analysis for Hyg17 with *myo*- and *scyllo*-inositol (Table 1 and Figure 2d,e). The $K_M$ value

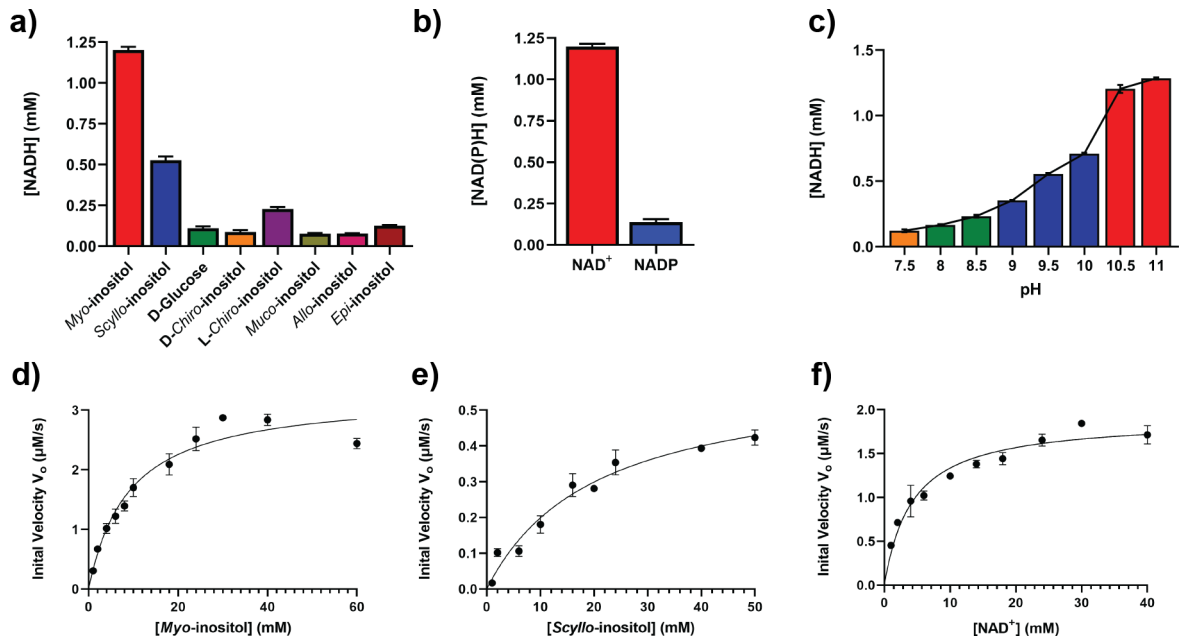



**Figure 2:** Hyg17 activity. Reactions with Hyg17 and (a) various inositols with $NAD^+$, (b) *myo*-inositol with $NAD^+$ or NADP, and (c) *myo*-inositol at different pH values (orange is HEPEs, green is Tris, blue is CHES, and red is CAPS). NAD(P)H concentrations were measured after 20 minutes. Michaelis–Menten plots for Hyg17 using varying concentrations of (d) *myo*-inositol, (e) *scyllo*-inositol, and (f) $NAD^+$. Reactions were monitored for NADH.

**Table 1:** Kinetic parameters for Hyg17.

| Substrate/cofactor | $k_{cat}$, $s^{-1}$ | $K_M$, mM | $k_{cat}/K_M$, $M^{-1}$ $s^{-1}$ |
|---|---|---|---|
| *myo*-inositol | 3.3 ± 0.13 | 9.0 ± 1.1 | 366.7 ± 46.96 |
| *scyllo*-inositol | 0.60 ± 0.044 | 20.3 ± 3.3 | 29.6 ± 5.28 |
| $NAD^+$ | 1.9 ± 0.057 | 4.2 ± 0.49 | 452.4 ± 54.49 |

for Hyg17 with *myo*-inositol was 9.0 ± 1.1 mM, which is relatively high. However, this is similar to other *myo*-inositol dehydrogenases whose $K_M$ values are also in the mM range [12,16]. Hyg17 showed a higher catalytic efficiency of 366.7 ± 46.96 $M^{-1}$ $s^{-1}$ with *myo*-inositol as compared to the 29.6 ± 5.28 $M^{-1}$ $s^{-1}$ for *scyllo*-inositol. This reduced catalytic efficiency can be attributed to the reduced $k_{cat}$ and increased $K_M$ for *scyllo*-inositol over *myo*-inositol. We also found that the catalytic efficiency for $NAD^+$ was 452.4 ± 54.49 $M^{-1}$ $s^{-1}$ (Table 1 and Figure 2f).

## Sequence similarity network

We generated a sequence similarity network (SSN) for the protein family PF01408, in which Hyg17 is a member (Figure 3). PF01408 is classified as an oxidoreductase with NAD-binding Rossmann fold family and contains over 340,000 sequences. Many of the family members act as sugar dehydrogenases with diverse sugar substrates (Supporting Information File 1, Table S1). These enzymes can have distinct biological functions, such as sugar metabolism and LPS biosynthesis [19-23]. By contrast, other members are involved in natural product biosynthetic pathways similar to Hyg17 [24-26]. Although many of the PF01408 enzymes have reported activities, the SSN shows there are still several clusters whose activity remain unknown, suggesting potential for new enzyme discovery within this family of enzymes.

Not surprising, Hyg17 is found in an SSN cluster with known inositol dehydrogenases. To further analyze the relationship between Hyg17 and these inositol dehydrogenases, a second SSN was generated using only sequences from this cluster (Figure 4a). Enzymes with verified activities are grouped in three of the main clusters. However, Hyg17 is found in a small cluster independent of the other sequences, which could help to explain its unique substrate scope. Interestingly, this cluster is separated into two groups. All the sequences that

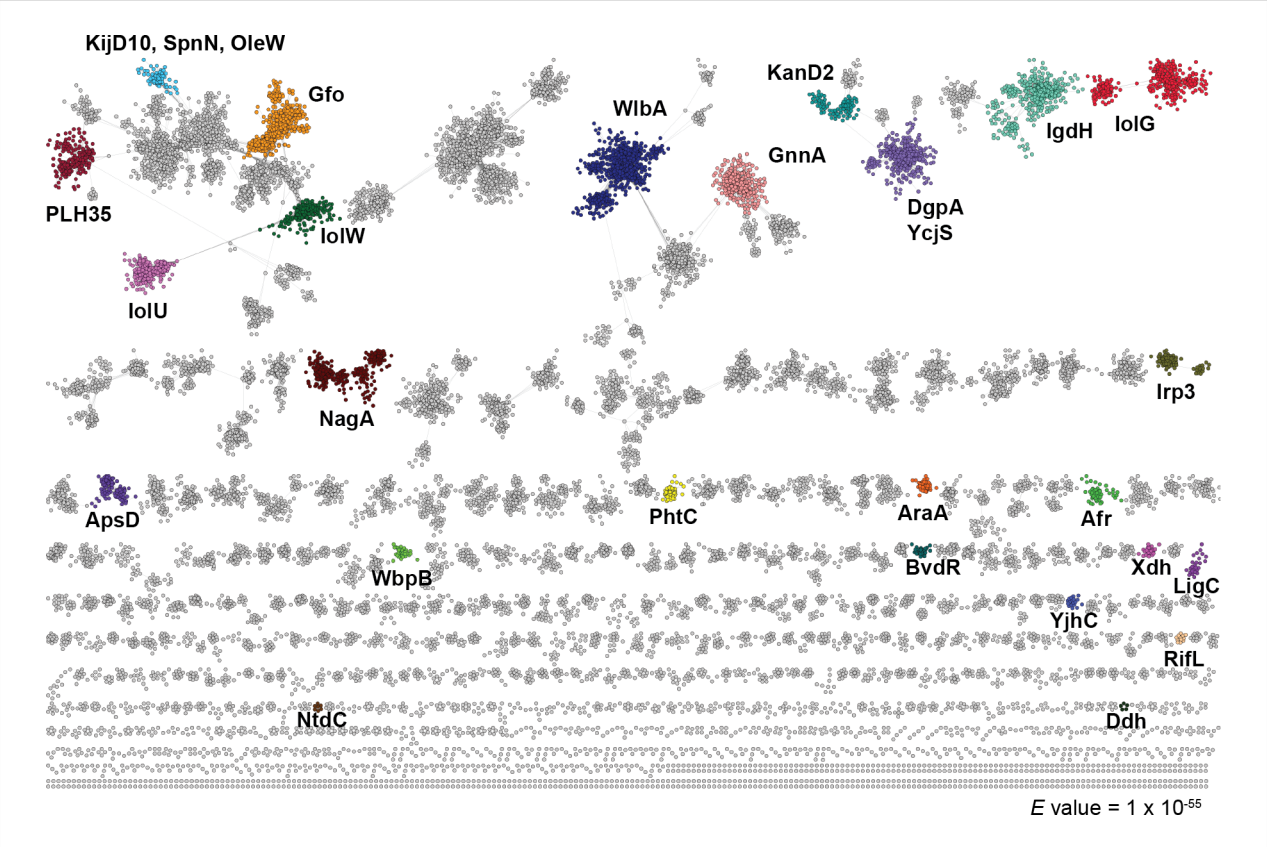



**Figure 3:** SSN for PF01408. Clusters with characterized enzymes are shown in different colors and labeled with the protein name. Functional details for characterized enzymes in PF01408 can be found in Supporting Information File 1, Table S1.

group with Hyg17 are found in identical hygromycin A biosynthetic clusters. The second group has 12 of the 29 Hyg genes with greater than 30% amino acid identity (Figure 4b and Supporting Information File 1, Table S2). The strains that contain these biosynthetic gene clusters could be producing a hygromycin A-like compound and further investigation is required.

## Genome mining for natural product biosynthetic clusters

Since some enzymes from PF01408 are involved in natural product biosynthetic pathways, we wanted to see if sequences from this PFAM could be used to mine for potentially new natural product biosynthetic clusters. We analyzed the genomic neighborhood of these sequences and searched for domains commonly associated with natural product biosynthetic enzymes (Supporting Information File 1, Table S3). We observed that 584 sequences were near an acyl synthase domain, 340 sequences by an acyl carrier protein domain, and 1,193 sequences by a thioesterase domain. In addition, Hyg17 works together with the aminotransferase Hyg8 to replace a hydroxy group with an amine generating an aminocyclitol from *myo*-inositol. We noticed that some other enzymes from this PFAM are also paired with aminotransferases to generate amino sugars found in natural products [25,26], so we looked to see if aminotransferases are commonly found by PF01408 sequences. We observed three separate aminotransferase PFAMs near many of the PF01408 sequences. DegT/DnrJ/EryCq/StrS aminotransferases (PF01041) were most frequently associated with PF01408 sequences with 15,139 occurrences. Class III (PF00202) and class I and II (PF00155) aminotransferases were significantly less frequent, with 1,236 and 1,318 occurrences, respectively. Furthermore, the presence of a resistance gene within a biosynthetic gene cluster can indicate that it produces an antimicrobial compound [27]. We searched for annotated resistance genes in the surrounding genomic neighborhood of PF01408 sequences. We found 1,166 PF01408 sequences were associated with nearby glyoxalase/bleomycin resistance protein/dioxygenase superfamily (PF00903) sequences, 81 associated with CrcB-like protein, camphor resistance (PF02537) sequences, 22 associated with aminoglycoside antibiotic resistance kinase (PF04655) sequences, and 2 associated with putative multidrug resistance efflux protein (PF13536) sequences. Together, this suggests that PF01408 sequences may be found in a significant number of uncharacterized biosynthetic gene clusters. However, a more detailed analysis of the genomic

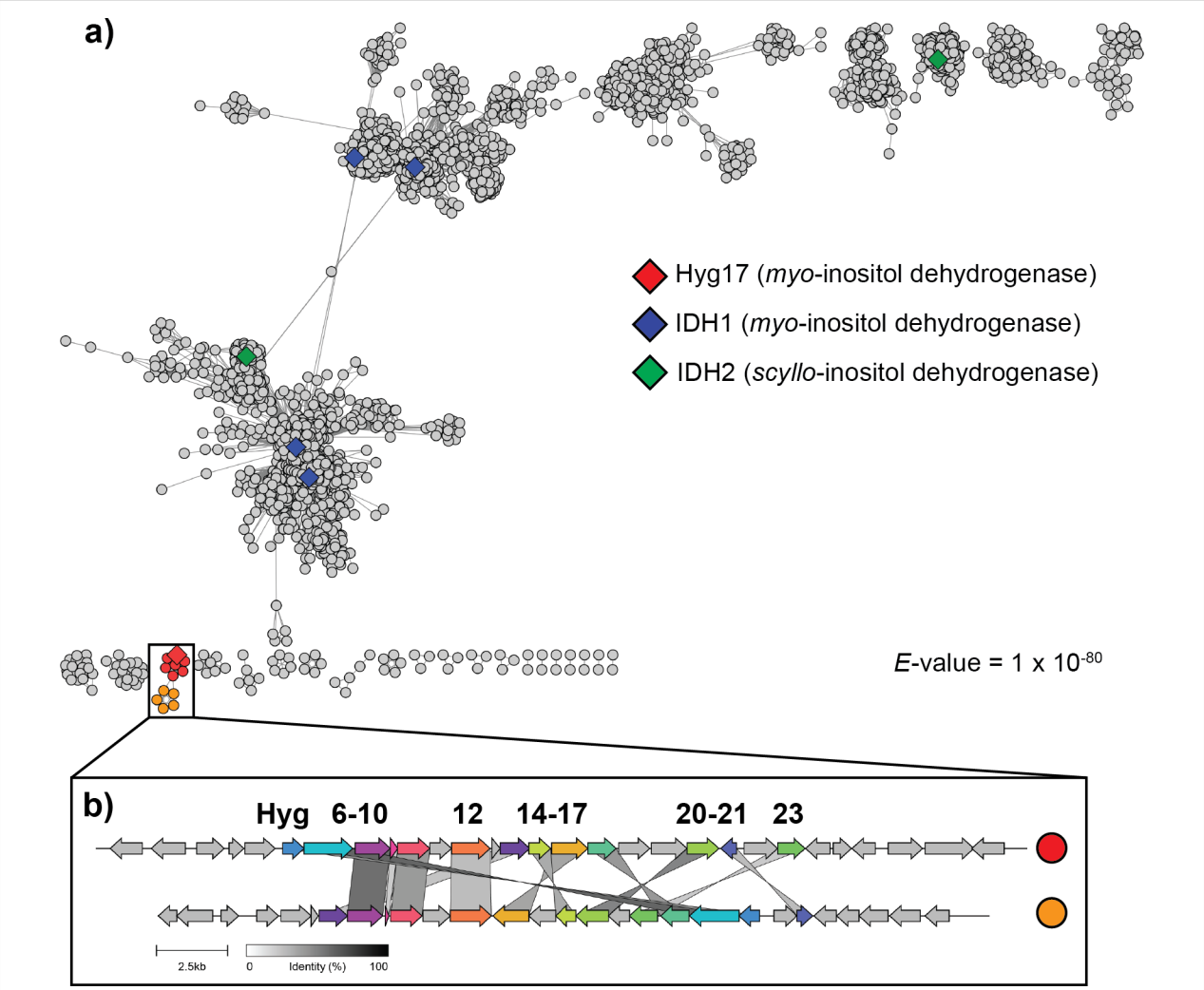



**Figure 4:** (a) SSN for inositol dehydrogenases. (b) Comparison of the hygromycin A (red) and hygromycin A-like (orange) biosynthetic gene clusters. A more detailed comparison can be found in Supporting Information File 1, Table S2.

neighborhoods is needed to assess their promising potential in natural product discovery.

## Conclusion

Hygromycin A is a broad-spectrum antibiotic that contains an aminocyclitol moiety essential for bioactivity [6,7]. The biosynthesis of the aminocyclitol has been proposed to proceed through six steps starting from glucose-6-phosphate through a *myo*-inositol intermediate to the final methylenedioxy aminocyclitol [8]. However, none of the enzymes in the pathway have been characterized in vitro. In this study, we show that Hyg17 is an $NAD^+$-dependent *myo*-inositol dehydrogenase with an optimal pH of 10.5–11. When tested on a range of inositol substrate, we show that Hyg17 has reduced activity on *scyllo*-inositol and L-*chiro*-inositol substrates, which is unusual for a *myo*-inositol dehydrogenase. Hyg17 belongs to PF01408, which is a protein family whose members have a broad range of activities and substrates. However, there are a variety of PF01408 members involved in various natural product biosynthetic pathways. Genome neighborhood analysis shows that many PF01408 sequences can be found near known biosynthetic enzymes, aminotransferases, and resistance genes suggesting that these enzymes may be promising candidates for targeted gene mining strategies to discover novel natural products.

## Experimental

### Cloning, expression, and purification

*Streptomyces leeuwenhoekii* NRRL B-24963 [28] was used as a template to amplify the *hyg17* (GenBank CQR59633) with the primers 5'-GTTAGCCATATGACGGTCGCCGTCGTGGGC-3' and 5'-GTAATGCTCGAGCGGCGCCACCGGCACCGA-3'. *hyg17* was cloned into pTip-QC1 [10] using NdeI and XhoI restriction sites and verified by DNA sequencing (Eurofins Genomics). pTip-QC1-*hyg17* plasmid [10] was transformed

into *Rhodococcus jostii* RHA1 [11]. Cultures were grown in Luria Bertani (LB) media supplemented with 34 μg $mL^{-1}$ chloramphenicol at 30 °C while shaking at 200 rpm for 48 h reaching an $OD_{600}$ of ≈1.4 then induced with 50 μL of 20 mg $mL^{-1}$ thiostrepton and grown for another 24 h at 30 °C while shaking at 200 rpm. Cells were then harvested by centrifugation, resuspended in binding buffer (500 mM NaCl, 20 mM Tris pH8.0) and disrupted by sonication using a Branson Sonifier 450 (5 rounds of 3 s/3 s on/off cycles for 5 min at a duty cycle of 50). Initial purification was performed with gravity filtration using nickel-nitrilotriacetic acid (Ni-NTA) resin (GE Healthcare). Cell lysate was loaded onto the Ni-NTA resin and washed with binding buffer. Protein was eluted using increasing concentrations of imidazole in binding buffer (10, 20, 50, and 500 mM imidazole). Fractions were run on an ExpressPlus™ PAGE Gel (GenScript) and protein was visualized using Coomassie Brilliant Blue G-250 (VWR). Fractions containing Hyg17 were pooled and further purified using a HiLoad 16/600 Superdex 75 pg size exclusion chromatography column (Cytiva) equilibrated in 50 mM NaCl, 20 mM Tris pH 8.0 and run on a Bio-Rad NGC chromatography system. Protein-containing fractions were identified by UV absorbance at 280 nm and were pooled and concentrated using a centrifugal device with a 10 kDa molecular weight cut-off (Pall Corporation).

*bsIDH* (Uniprot ID P26935) was gene synthesized with codon optimization and cloned into pET28a by Twist Bioscience. pET28a-*bsIDH* was transformed into *E. coli* BL21 star (DE3) (Agilent). For overnight cultures, 5–6 colonies were inoculated into 50 mL of LB supplemented with 10 μg $mL^{-1}$ ampicillin and 1.5 μg $mL^{-1}$ tetracycline and grown at 200 rpm and 37 °C overnight. One liter of autoinduction media (20 g $L^{-1}$ tryptone, 10 g $L^{-1}$ yeast extract, 50 mM $NH_4Cl$, 2 mM $MgSO_4$, 0.5% glycerol, 17 mM $KH_2PO_4$, 72 mM $K_2HPO_4$, 0.05% glucose, and 0.2% lactose) supplemented with 10 μg $mL^{-1}$ ampicillin and 1.5 μg $mL^{-1}$ tetracycline was inoculated with 20 mL of overnight culture and incubated at 200 rpm and 37 °C for 2 h. After 2 h, the culture was grown at 150 rpm and 16 °C for three nights. Cultures were harvested by centrifugation at 5,000 rpm for 10 min. The pellet was disrupted using a chemical lysis method. Specifically, the pellet was resuspended in 40 mL of sucrose solution (25% sucrose, 50 mM Tris pH 8) using continuous stirring. Then, 10 mg of lysozyme (Bio Basic) was added and stirred at room temperature for 10 min, followed by the addition of 80 mL deoxycholate solution (1% deoxycholate, 1% Triton X-100, 100 mM NaCl, 20 mM Tris pH 7.5) for another 10 min. Then $MgSO_4$ (to ≈1 mM) and 0.2 mg DNase (Thermo Fisher Scientific) were added and stirred for 10 min. The lysate was centrifuged to remove debris for 45 min at 10,000 rpm and 4 °C, and the lysate was collected into a beaker. Purification was then performed using the same protocol as described above for Hyg17 except that a HiLoad 26/600 Superdex 200 pg size-exclusion chromatography column (Cytiva) equilibrated in 50 mM NaCl, 20 mM Tris pH 8.0 was used.

Protein concentration was determined by UV absorbance at 280 nm using the calculated extinction coefficient 33,460 $M^{-1}$ $cm^{-1}$ and 38,350 $M^{-1}$ $cm^{-1}$ for Hyg17 and BsIDH, respectively [29]. Protein was flash frozen in liquid nitrogen at 100 μM in 20 mM Tris pH 8, 50 mM NaCl, and 10% glycerol and stored at −80 °C for biochemical assays.

## Enzyme assays

Generally, 100 μL reactions contained 10 mM inositol substrate, 10 mM $NAD^+$ cofactor, 1 μM Hyg17 enzyme, 100 mM CAPS, 50 mM NaCl, pH 10.5 buffer. The production of NADH was monitored by measuring the absorbance at 340 nm using the SpectraMax M2 UV–vis spectrophotometer over 20 min at 25 °C. The optimal pH was determined by performing reactions with varying buffers (Tris-HCl, HEPES, CHES and CAPS) from pH 7.5 to 11. Different inositol substrates, including *myo*-inositol, *scyllo*-inositol, *muco*-inositol, *epi*-inositol, D-*chiro*-inositol, L-*chiro*-inositol, and D-glucose were tested. Inositol substrates were purchased from TCI Chemicals or Santa Cruz Biotechnology. In addition, $NAD^+$ and NADP cofactors were also tested in reactions with *myo*-inositol. $NAD^+$ and NADP were purchased from Research Products International. Results were analyzed using GraphPad Prism version 9.5.1 for Windows, GraphPad Software, Boston, Massachusetts USA, https://www.graphpad.com.

Reactions for thin-layer chromatography analysis were carried out using the same general reaction conditions reported above. Six μL of each reaction were spotted on the TLC silica gel 60 $F_{254}$ (Sigma) and run with a solvent system containing butanol/ethanol/water 5:4:3. The TLC was stained using a solution containing 3.7 mL *p*-anisaldehyde, 135 mL ethanol, 5 mL sulfuric acid, and 1.5 mL glacial acetic acid. Light pink spots were observed following heating at 105 °C.

Kinetics assays were carried out by varying concentrations of substrate (*myo*-inositol or *scyllo*-inositol) from 1 mM to 60 mM with 1 μM Hyg17 and 10 mM $NAD^+$ in 100 mM CAPS, 50 mM NaCl, pH 10.5. Additional kinetics assays were carried out by varying concentrations of $NAD^+$ from 1 mM to 40 mM with 1 μM Hyg17 and 10 mM *myo*-inositol in 100 mM CAPS, 50 mM NaCl, pH 10.5. The production of NADH was monitored and initial reaction rates calculated by determining the slope of the reaction from 0 to 5 min. Results were analyzed using Microsoft Excel and GraphPad Prism version 9.5.1 for Windows, GraphPad Software, Boston, Massachusetts USA, https://www.graphpad.com.

## SSN network generation and genome mining

The sequence similarity network was generated using the enzyme function initiative (EFI-EST) [30] web tool. An SSN was created by inputting the Hyg17 sequence into the Sequence BLAST function and retrieving the top 500 sequence hits along with sequences from the protein family PF01408 in which Hyg17 is found. The sequences were filtered for bacterial sequences and the UniRef50 function was used along with an *E*-value cut off of $1 \times 10^{-55}$ and node network with 80% ID. Clusters with a single node were removed for simplicity. This produced an SSN with 28,698 nodes. A second SSN was generated for the IolG cluster from the PF01408 SSN. The Uniprot IDs for all sequences found in the IolG cluster were used to generate an SSN using the EFI-EST web tool. The *E*-value cut off was set to $1 \times 10^{-80}$ and the node network with 80% ID was used. Clusters with a single node were removed for simplicity which generated an SSN with 5339 nodes. Genomic neighborhood analysis was completed using the enzyme function initiative (EFI-GNT) web tool [31,32]. Gene clusters were compared using CAGECAT [33]. SSNs and GNNs were visualized using Cytoscape [34].

## Supporting Information

### Supporting Information File 1

Additional tables and figures.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-51-S1.pdf]

## Acknowledgements

The authors would like to thank Lindsay Eltis (University of British Columbia, Canada) for providing the *Rhodococcus* expression strain and Tomohiro Tamura (National Institute of Advanced Industrial Science and Technology, Japan) for providing the pTip-QC1 expression vector.

## Funding

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM151137.

## ORCID® iDs

Michael O. Akintubosun - https://orcid.org/0000-0003-4129-6752
Melanie A. Higgins - https://orcid.org/0000-0003-1485-1378

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Pittenger, R. C.; Wolfe, R. N.; Hoehn, M. M.; Marks, P. N.; Daily, W. A.; McGuire, J. M. *Antibiot. Chemother.* **1953,** *3,* 1268–1278.
2. Mann, R. L.; Gale, R. M.; Van Abeele, F. R. *Antibiot. Chemother.* **1953,** *3,* 1279–1282.
3. Leimer, N.; Wu, X.; Imai, Y.; Morrissette, M.; Pitt, N.; Favre-Godal, Q.; Iinishi, A.; Jain, S.; Caboni, M.; Leus, I. V.; Bonifay, V.; Niles, S.; Bargabos, R.; Ghiglieri, M.; Corsetti, R.; Krumpoch, M.; Fox, G.; Son, S.; Klepacki, D.; Polikanov, Y. S.; Freliech, C. A.; McCarthy, J. E.; Edmondson, D. G.; Norris, S. J.; D'Onofrio, A.; Hu, L. T.; Zgurskaya, H. I.; Lewis, K. *Cell* **2021,** *184,* 5405–5418.e16. doi:10.1016/j.cell.2021.09.011
4. Guerrero, M. d. C.; Modolell, J. *Eur. J. Biochem.* **1980,** *107,* 409–414. doi:10.1111/j.1432-1033.1980.tb06044.x
5. Poulsen, S. M.; Kofoed, C.; Vester, B. *J. Mol. Biol.* **2000,** *304,* 471–481. doi:10.1006/jmbi.2000.4229
6. Kaminishi, T.; Schedlbauer, A.; Fabbretti, A.; Brandi, L.; Ochoa-Lizarralde, B.; He, C.-G.; Milón, P.; Connell, S. R.; Gualerzi, C. O.; Fucini, P. *Nucleic Acids Res.* **2015,** gkv975. doi:10.1093/nar/gkv975
7. Jaynes, B. H.; Elliott, N. C.; Schicho, D. L. *J. Antibiot.* **1992,** *45,* 1705–1707. doi:10.7164/antibiotics.45.1705
8. Palaniappan, N.; Dhote, V.; Ayers, S.; Starosta, A. L.; Wilson, D. N.; Reynolds, K. A. *Chem. Biol.* **2009,** *16,* 1180–1189. doi:10.1016/j.chembiol.2009.10.013
9. Palaniappan, N.; Ayers, S.; Gupta, S.; Habib, E.-S.; Reynolds, K. A. *Chem. Biol.* **2006,** *13,* 753–764. doi:10.1016/j.chembiol.2006.05.013
10. Nakashima, N.; Tamura, T. *Appl. Environ. Microbiol.* **2004,** *70,* 5557–5568. doi:10.1128/aem.70.9.5557-5568.2004
11. Round, J. W.; Robeck, L. D.; Eltis, L. D. *ACS Synth. Biol.* **2021,** *10,* 2383–2395. doi:10.1021/acssynbio.1c00292
12. Aamudalapalli, H. B.; Bertwistle, D.; Palmer, D. R. J.; Sanders, D. A. R. *Biochim. Biophys. Acta, Proteins Proteomics* **2018,** *1866,* 1115–1124. doi:10.1016/j.bbapap.2018.08.011
13. Ramaley, R.; Fujita, Y.; Freese, E. *J. Biol. Chem.* **1979,** *254,* 7684–7690. doi:10.1016/s0021-9258(18)36000-9
14. Zheng, H.; Bertwistle, D.; Sanders, D. A. R.; Palmer, D. R. J. *Biochemistry* **2013,** *52,* 5876–5883. doi:10.1021/bi400821s
15. Rodionova, I. A.; Leyn, S. A.; Burkart, M. D.; Boucher, N.; Noll, K. M.; Osterman, A. L.; Rodionov, D. A. *Environ. Microbiol.* **2013,** *15,* 2254–2266. doi:10.1111/1462-2920.12096
16. Ramp, P.; Pfleger, C.; Dittrich, J.; Mack, C.; Gohlke, H.; Bott, M. *Microbiol. Spectrum* **2022,** *10,* e01950-22. doi:10.1128/spectrum.01950-22
17. Morinaga, T.; Ashida, H.; Yoshida, K.-i. *Microbiology (London, U. K.)* **2010,** *156,* 1538–1546. doi:10.1099/mic.0.037499-0
18. Kang, D.-M.; Tanaka, K.; Takenaka, S.; Ishikawa, S.; Yoshida, K.-i. *Biosci., Biotechnol., Biochem.* **2017,** *81,* 1026–1032. doi:10.1080/09168451.2016.1268043
19. Watanabe, Y.; Iga, C.; Watanabe, Y.; Watanabe, S. *FEBS Lett.* **2019,** *593,* 1257–1266. doi:10.1002/1873-3468.13424
20. Mihasan, M.; Stefan, M.; Hritcu, L.; Artenie, V.; Brandsch, R. *Res. Microbiol.* **2013,** *164,* 22–30. doi:10.1016/j.resmic.2012.10.003
21. Bell, A.; Severi, E.; Lee, M.; Monaco, S.; Latousakis, D.; Angulo, J.; Thomas, G. H.; Naismith, J. H.; Juge, N. *J. Biol. Chem.* **2020,** *295,* 13724–13736. doi:10.1074/jbc.ra120.014454
22. Manissorn, J.; Sitthiyotha, T.; Montalban, J. R. E.; Chunsrivirot, S.; Thongnuek, P.; Wangkanont, K. *ACS Chem. Biol.* **2020,** *15,* 3235–3243. doi:10.1021/acschembio.0c00791

23. Thoden, J. B.; Holden, H. M. *Biochemistry* **2010,** *49,* 7939–7948. doi:10.1021/bi101103s
24. Kubiak, R. L.; Holden, H. M. *Biochemistry* **2011,** *50,* 5905–5917. doi:10.1021/bi200514b
25. Vetter, N. D.; Langill, D. M.; Anjum, S.; Boisvert-Martel, J.; Jagdhane, R. C.; Omene, E.; Zheng, H.; van Straaten, K. E.; Asiamah, I.; Krol, E. S.; Sanders, D. A. R.; Palmer, D. R. J. *J. Am. Chem. Soc.* **2013,** *135,* 5970–5973. doi:10.1021/ja4010255
26. Kudo, F.; Kitayama, Y.; Miyanaga, A.; Hirayama, A.; Eguchi, T. *Biochemistry* **2020,** *59,* 1470–1473. doi:10.1021/acs.biochem.0c00204
27. O'Neill, E. C.; Schorn, M.; Larson, C. B.; Millán-Aguiñaga, N. *Crit. Rev. Microbiol.* **2019,** *45,* 255–277. doi:10.1080/1040841x.2019.1590307
28. Busarakam, K.; Bull, A. T.; Girard, G.; Labeda, D. P.; van Wezel, G. P.; Goodfellow, M. *Antonie van Leeuwenhoek* **2014,** *105,* 849–861. doi:10.1007/s10482-014-0139-y
29. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. *Nucleic Acids Res.* **2003,** *31,* 3784–3788. doi:10.1093/nar/gkg563
30. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. *Biochim. Biophys. Acta, Proteins Proteomics* **2015,** *1854,* 1019–1037. doi:10.1016/j.bbapap.2015.04.015
31. Oberg, N.; Zallot, R.; Gerlt, J. A. *J. Mol. Biol.* **2023,** *435,* 168018. doi:10.1016/j.jmb.2023.168018
32. Zallot, R.; Oberg, N.; Gerlt, J. A. *Biochemistry* **2019,** *58,* 4169–4182. doi:10.1021/acs.biochem.9b00735
33. Gilchrist, C. L. M.; Chooi, Y.-H. *Bioinformatics* **2021,** *37,* 2473–2475. doi:10.1093/bioinformatics/btab007
34. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003,** *13,* 2498–2504. doi:10.1101/gr.1239303

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.51

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Chemical and biosynthetic potential of *Penicillium shentong* XL-F41

Ran Zou[1,2], Xin Li[1], Xiaochen Chen[1], Yue-Wei Guo[1,3] and Baofu Xu[*1,3]

Full Research Paper

Open Access

Address:
[1]Shandong Laboratory of Yantai Drug Discovery, Bohai Rim Advanced Research Institute for Drug Discovery, Yantai, Shandong 264117, China, [2]School of Life Sciences, Ludong University, Yantai 264025, China and [3]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zu Chong Zhi Road, Zhangjiang Hi-Tech Park, Shanghai, 201203, China

Email:
Baofu Xu[*] - bfxu@simm.ac.cn

* Corresponding author

Keywords:
genome analysis; indole terpene alkaloid; natural products; *Penicillium*; structure elucidation

*Beilstein J. Org. Chem.* **2024,** *20,* 597–606.
https://doi.org/10.3762/bjoc.20.52

Received: 07 December 2023
Accepted: 07 March 2024
Published: 15 March 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra

© 2024 Zou et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

*Penicillium* strains are renowned for producing diverse secondary metabolites with unique structures and promising bioactivities. Our chemical investigations, accompanied by fermentation media optimization, of a newly isolated fungus, *Penicillium shentong* XL-F41, led to the isolation of twelve compounds. Among these are two novel indole terpene alkaloids, shentonins A and B (**1** and **2**), and a new fatty acid **3**. Shentonin A (**1**) is distinguished by an unusual methyl modification at the oxygen atom of the typical succinimide ring, a feature not seen in the structurally similar brocaeloid D. Additionally, shentonin A (**1**) exhibits a *cis* relationship between H-3 and H-4, as opposed to the *trans* configuration in brocaeloid D, suggesting a divergent enzymatic ring-expansion process in their respective fungi. Both shentonins A (**1**) and B (**2**) also feature a reduction of a carbonyl to a hydroxy group within the succinimide ring. All isolated compounds were subjected to antimicrobial evaluations, and compound **12** was found to have moderate inhibitory activity against *Candia albicans*. Moreover, genome sequencing of *Penicillium shentong* XL-F41 uncovered abundant silent biosynthetic gene clusters, indicating the need for future efforts to activate these clusters and unlock the full chemical potential of the fungus.

## Introduction

*Penicillium*, a genus within the *Ascomycota* phylum, is a type of critical saprophytic fungus with over 400 strains identified in diverse environments such as mountains, oceans, and the human gut [1]. After the first antibiotic mycophenolic acid originally isolated by Gosio in the 1890s [2], the important antibiotic penicillin was characterized more than one decade after Fleming discovered the antibacterial activity of a *Penicillium* extract, and since then, *Penicillium* has been an important target in drug de-

velopment. Researchers have identified numerous compounds with anticancer properties, including mycophenolic acid, brefeldin A, and wortmannin [3], as well as compounds with antibacterial properties like xestodecalactones A–C, penicifurans A, and anthraquinone-citrinin [4]. From 2010 to 2022, researchers have identified over 260 secondary metabolites from *Penicillium* [5], exhibiting not only antibacterial and anticancer activities but also potent antioxidant properties, inhibition of GSK-3β and α-glucosidase activities, and interaction with the pregnane X receptor (PXR). These compounds are categorized into polyketides, alkaloids, sterol derivatives, terpenoids, and macrolides, with polyketides and alkaloids comprising 40% and 32% of the total, respectively.

Alkaloids are a diverse group of compounds with multiple pharmacological activities, including anti-inflammatory, antibacterial, antiviral, insecticidal, and anticancer properties [6,7]. Historically, most alkaloids were isolated from higher plants, with a significant number found in the *Apocynaceae* family. Notable examples such as vinblastine, vinorelbine, vincristine, and vindesine have gained prominence as effective anticancer drugs [6-10]. Recent studies have revealed that certain fungi are also prolific sources of indole alkaloids, which are among the largest classes of nitrogen-containing secondary metabolites. Characterized by at least one indole moiety and derived from tryptophan or tryptamine, indole alkaloids are known for their diverse structures, electron-donating capabilities, and excellent biocompatibility, contributing to their potent antibacterial and anticancer activities [9,10]. Over 4000 species [8] producing indole alkaloids have been identified, and many of these compounds are now successfully employed in clinical applications.

Despite the extensive catalog of secondary metabolites discovered, the pace of new findings has decelerated. However, the advent of bioinformatics analysis tools has reinvigorated the search for fungal secondary metabolites. The estimated number of non-redundant clusters in *Penicillium* is around 25,000 [1], yet the number of isolated compounds is significantly lower, indicating the presence of many unexpressed gene clusters. This suggests a wealth of undiscovered compounds with potentially novel structures and significant biological activities. To stimulate the expression of biosynthetic gene clusters (BGCs), several methods can be utilized, for instance, epigenetic regulation, co-culture, precursor feeding, heterologous expression, and changing fermentation parameters [11-14].

In the present study, we focused on a newly identified *Penicillium* strain, *Penicillium shentong* XL-F41. To activate the BGCs of this strain, we employed a combination of elicitors in our fermentation media, including histone deacetylase inhibitors and DNA methyltransferase inhibitors. We developed two specialized media, XISR I and XISR III, which outperformed the traditional potato dextrose broth (PDB) in stimulating the production of a greater number of metabolite peaks, as shown in Figure 1. Scaled-up fermentation allowed us to isolate and characterize two new indole terpene alkaloids, shentonins A and B (**1** and **2**), a new fatty acid **3**, and nine previously identified compounds **4**–**12**, among which were gram quantities of curvularin analogs. Our bioactivity assays identified one compound, **12**, with promising antimicrobial properties. Subsequent genome sequencing analysis pinpointed the likely BGCs associated with our isolated compounds and suggested a vast potential for the production of additional compounds, given the application of suitable activation techniques.

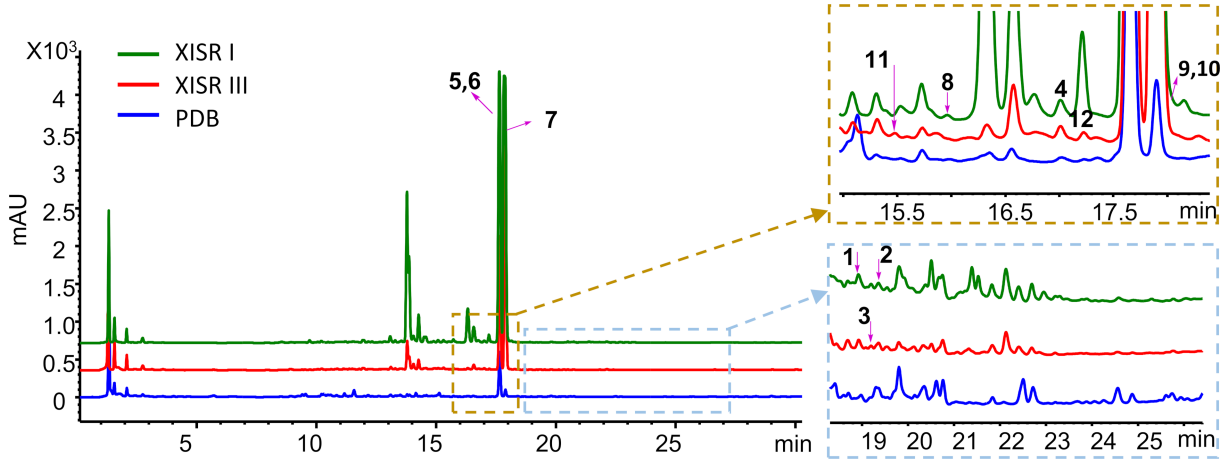



**Figure 1:** HPLC analysis of small-scale fermentation with different media. More details of media, XISR I and XISR III can be found in the methods section.

# Results and Discussion

## Compound isolation and structure elucidation

To activate the silent BGCs in *Penicillium shentong* XL-F41, we conducted small-scale fermentations using various media. Analysis revealed that HPLC peaks, which correspond to fermentation products, showed a lower number and abundance in the PDB medium than in the XISR I and XISR III media, as illustrated in Figure 1. Consequently, we chose XISR I and XISR III media for further fermentation.

The fermentation broth was exhaustively extracted with ethanol, after which the ethanol extract was partitioned between EtOAc and $H_2O$. The EtOAc fraction was chromatographed repeatedly over silica gel and reversed-phase high-performance liquid chromatography (RP-HPLC), resulting in the isolation of pure compounds **1**–**12** (Figure 2). According to literature reports of known compounds, some of them were identified as fusarindoles B (**4**) [15], dehydrocurvularin (**5**) [16], hydroxycurvularin (**6**) [17], curvularin (**7**) [18], curvulopyran (**8**) [19], (*S*)-6-(*sec*-butyl)-3-isobutylpyrazin-2(1*H*)-one (**9**) [20], 3,6-di-*sec*-butyl-2(1*H*)-pyrazinone (**10**) [20], daidzein (**11**) [21], and genistein (**12**) [22]. Notably, compound **7**, corresponding to the major peak in our optimized fermentation (Figure 1), was obtained at the gram level.

Compound **1** (shentonin A) was obtained as a light green solid with a chemical formula of $C_{20}H_{26}N_2O_3$, as determined by HRMS *m*/*z* 365.1828 $[M + Na]^+$ (calcd for $C_{20}H_{26}N_2O_3Na^+$, 365.1835) and HRMS *m*/*z* 341.1862 $[M - H]^-$ (calcd for $C_{20}H_{25}N_2O_3$, 341.1870). Spectroscopic analysis, including $^1H$ NMR, $^{13}C$ NMR (Table 1), and DEPT, revealed that compound **1** contains three methyl groups, one of which is oxygenated, four methines, three saturated non-protonated carbons, and two ketone carbonyl carbons ($\delta_C$ 175.94, $\delta_C$ 194.36). Its NMR data closely resemble those of brocaeloid D [23], with the notable addition of a methoxy group ($\delta_H$ 3.20/$\delta_C$ 53.92). HMBC correlations confirmed the presence of a reversed prenyl group and differentiated compound **1** from brocaeloid D by the substitution of a succinimide substructure at C-14 with a methine at C-16, indicated by the methoxy group. The position of the methoxy substituent was established by HMBC correlations, and the $^{13}C$ NMR data suggested that compound **1** includes a 4-oxo-2,3-dihydro-(1*H*)-quinolin-3-yl fragment. The planar structure was established from HMBC correlations linking three different fragments.

Compound **1** features three stereogenic centers at C-2, C-3, and C-16. The relative configuration of C-2 and C-3 was determined as (2*R**,3*R**) by $^1H$-$^1H$ NOESY correlations (Figure 3),

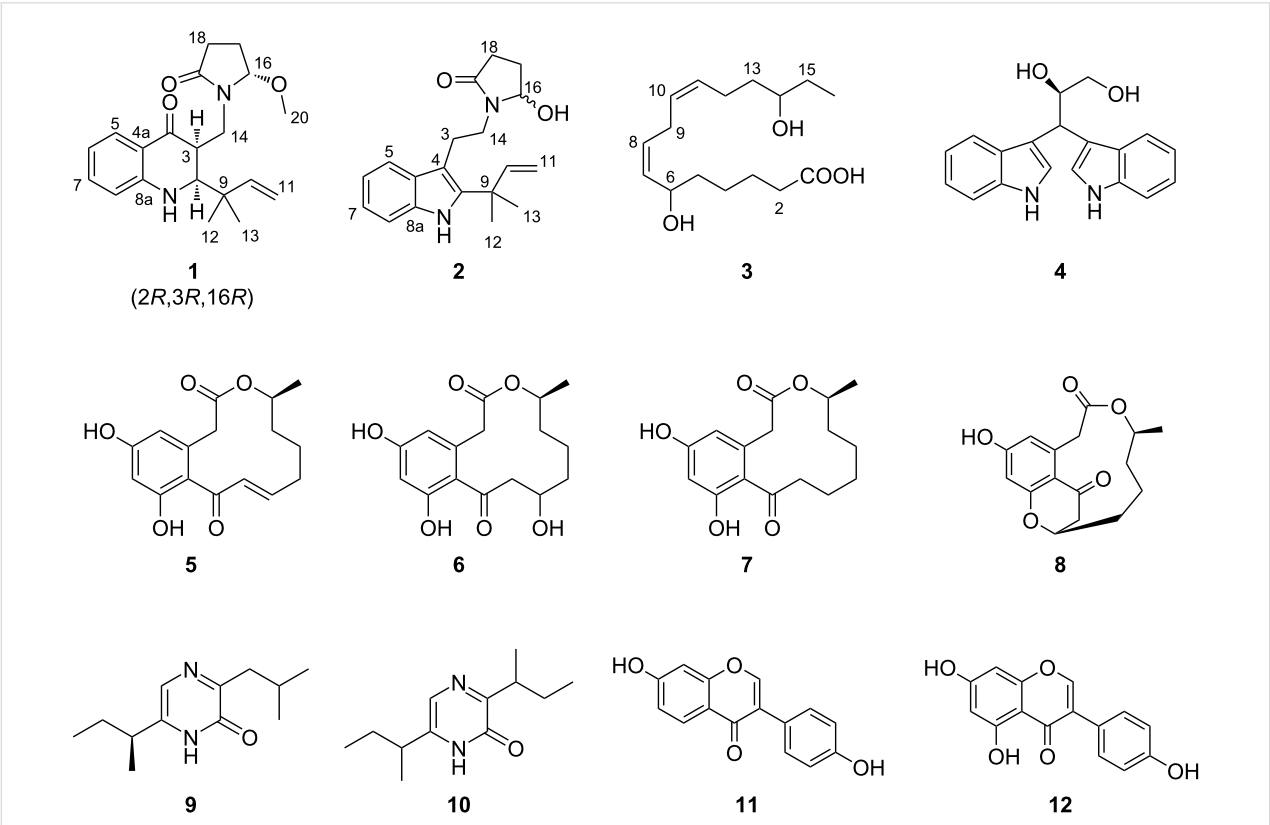



**Figure 2:** Chemical structures of compounds **1–12**.

**Table 1:** $^{1}H$ and $^{13}C$ data of compound **1** (recorded in $CDCl_3$).

| | $\delta_H$ mult (*J* in Hz) | $\delta_C$ mult |
|---|---|---|
| 1 | 4.66 (s) | NH |
| 2 | 3.09 (dd, 3.9, 1.0) | 61.4, CH |
| 3 | 2.92 (t, 7.6) | 45.2, CH |
| 4 | – | 194.3, qC |
| 4a | – | 116.9, qC |
| 5 | 7.68 (dd, 7.9, 1.6) | 127.2, CH |
| 6 | 6.61, m | 114.9, CH |
| 7 | 7.28 (d, 1.7) | 136.0, CH |
| 8 | 6.61, m | 116.7, CH |
| 8a | – | 149.9, qC |
| 9 | – | 43.1, qC |
| 10 | 5.65 (dd, 17.5, 10.8) | 144.3, CH |
| 11 | 5.02, m | 114.3, $CH_2$ |
| 12 | 0.97 (d, 7.2) | 23.2, $CH_3$ |
| 13 | 0.97 (d, 7.2) | 23.5, $CH_3$ |
| 14a | 3.16, m | 41.8, $CH_2$ |
| 14b | 3.90 (dd, 13.9, 8.8) | – |
| 15 | – | N |
| 16 | 4.77 (dd, 6.1, 1.2) | 89.7, CH |
| 17a | 1.99 (ddd,13.5,9.5) | 24.3, $CH_2$ |
| 17b | 2.12, m | – |
| 18a | 2.37 (ddd, 17.2, 9.9) | 28.9, $CH_2$ |
| 18b | 2.51 (dt, 17.8, 9.2) | – |
| 19 | – | 175.9, qC |
| 20 | 3.20, s | 53.9, $CH_3$ |

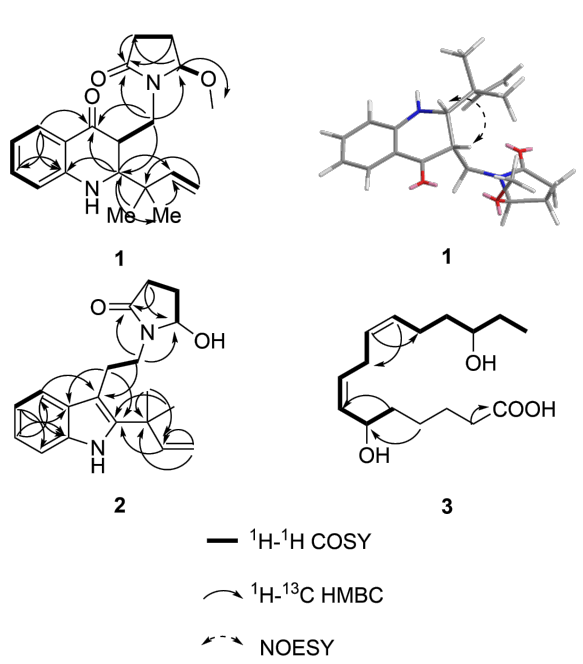



**Figure 3:** Key 2D NMR correlations of compounds **1–3**.

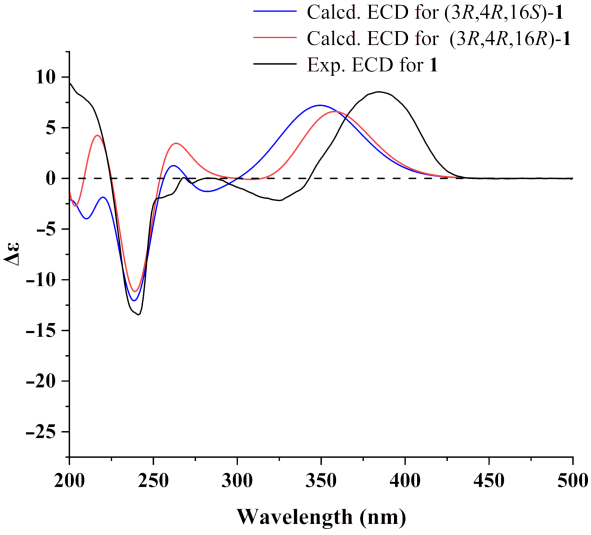



**Figure 4:** Experimental and calculated ECD spectra at the CAM-B3LYP/6-311G(d,p) level of theory for compound **1**.

while the relative configuration of C-16 remains unresolved due to the inapplicability of the NOESY experiment. To establish the absolute configuration of compound **1**, electron circular dichroism (ECD) calculations were conducted using the time-dependent density functional theory (TDDFT) approach at a B3LYP/6-311G (d,p) (IEFPCM) level (Figure 4). Considering the uncertainty of the relative configuration of C-16, both ECD spectra of 2*R*,3*R*,16*R*-**1** and 2*R*,3*R*,16*S*-**1** were calculated, and compared with experimental ECD. Both calculated spectra displayed almost identical curves compared to the experimental one, which was suggestive of the 2*R*,3*R* absolute configuration for compound **1**. The calculated spectrum of 2*R*,3*R*,16*R*-**1** matched better with the experimental one, so the absolute configuration of **1** was tentatively determined to be 2*R*,3*R*,16*R*, namely shentonin A.

Compound **2** (shentonin B) was isolated as a light green solid. Its chemical formula, $C_{19}H_{24}N_2O_2$, was confirmed by HRMS with *m*/*z* 335.1719 $[M + Na]^+$ (calcd for $C_{19}H_{24}N_2O_2Na^+$, 335.1730) and *m*/*z* 311.1755 $[M - H]^-$ (calcd for $C_{19}H_{23}N_2O_2$, 311.1765). Spectroscopic analysis using $^{1}H$ NMR, $^{13}C$ NMR, and DEPT (Table 2) indicated that compound **2** comprises two methyl groups, five methines, five saturated non-protonated carbons, and one ketone carbonyl carbon ($\delta_C$ 174.6). Its NMR profile is similar to brocaeloid C [24], with the distinction of an added succinimide substructure at N-15, where the ketone carbonyl carbon at C-16 is replaced by a hydroxy carbon. The isoprene group is consistent with that in compound **1**. HMBC cross-peaks from H-10 to C-2 and H-12 to C2 connect the indole and isoprene units, while HMBC correlations from H-14

**Table 2:** $^{1}H$ and $^{13}C$ NMR data of compound **2** (recorded in $CDCl_3$).

| | $\delta_H$ mult (*J* in Hz) | $\delta_C$ mult |
|---|---|---|
| 1 | 7.90, s | NH |
| 2 | – | 140.0, qC |
| 3 | 3.09 (ddd, 14.1, 9.3, 5.6) | 23.9, $CH_2$ |
| 3a | 3.16 (ddd, 14.1, 9.5, 6.5) | – |
| 4 | – | 108.1, qC |
| 4a | – | 129.6, qC |
| 5 | 7.61 (dp, 7.8, 0.7) | 118.3, CH |
| 6 | 7.09 (ddd, 8.1, 7.0, 1.1) | 119.6, CH |
| 7 | 7.14 (ddd, 8.1, 7.0, 1.2) | 121.7, CH |
| 8 | 7.29 (dt, 8.0, 1.0) | 110.6, CH |
| 8a | – | 134.2, qC |
| 9 | – | 39.1, qC |
| 10 | 6.14 (dd, 17.4, 10.5) | 146.0, CH |
| 11 | 5.16 (d, 1.1) | 112.2, $CH_2$ |
| 11a | 5.18 (dd, 2.5, 1.1) | – |
| 12 | 1.56 (d, 1.5) | 27.7, $CH_3$ |
| 13 | – | 27.8, $CH_3$ |
| 14 | 3.51 (m) | 41.5, $CH_2$ |
| 14a | 3.67 (ddd, 13.7, 9.5, 5.6) | – |
| 15 | – | N |
| 16 | 4.98 (s) | 84.3, CH |
| 17 | 2.30 (ddd, 17.1, 10.1, 4.3) | 28.9, $CH_2$ |
| 17a | 2.54 (ddd, 16.9, 9.7, 7.2) | – |
| 18 | 1.76 (dddd, 13.8, 9.7, 4.3, 2.4) | 28.7, $CH_2$ |
| 18a | 2.19 (dddd, 13.7, 10.1, 7.3, 6.4) | – |
| 19 | – | 174.6, qC |

to C-4, C-16, and C-19, and from H-3 to C-4a and C-9, elucidate the connectivity of three fragments. These data collectively establish the planar structure of compound **2**.

Compound **3** was isolated as a transparent oily liquid, and its chemical formula, $C_{16}H_{28}O_4$, was confirmed by LC–MS with *m/z* 283.2 $[M - H]^-$ (calcd for $C_{16}H_{27}O_4$, 283.2) (Figure S25 in Supporting Information File 1). Spectroscopic analyses, including $^{1}H$ NMR, $^{13}C$ NMR, DEPT, HSQC, COSY, and HMBC (Table 3, Figure 3), identified compound **3** as a sixteen-carbon fatty acid. Notably, two methylene carbons overlapped in the $^{13}C$ NMR spectrum. The COSY correlations facilitated the determination of the carbon chain fragments from C-11 to C-16 and C-2 to C-10, despite two methylene signals overlapping. The carboxyl group's position at C-1 was confirmed by HMBC correlations from H-2/3. Furthermore, HMBC cross-peaks from H-12 to C-10, H-11 to C-9, and H-10 to C-12 indicated that the fragments are connected through C-11 and C-10, establishing the structure of compound **3**.

**Table 3:** $^{1}H$ and $^{13}C$ NMR data of compound **3** (recorded in $CDCl_3$).

| | $\delta_H$ mult (*J* in Hz) | $\delta_C$ mult |
|---|---|---|
| 1 | – | 177.2, qC |
| 2 | 2.41, m; 1.96 (dp, 12.9, 9.3) | 29.3, $CH_2$ |
| 3 | 2.57, m | 29.0, $CH_2$ |
| 4 | 2.57, m | 29.0, $CH_2$ |
| 5 | 2.41, m; 1.96 (dp, 12.9, 9.3) | 29.3, $CH_2$ |
| 6 | 5.30, m | 76.4, CH |
| 7 | 5.47, m | 127.5, CH |
| 8 | 5.66 (dt, 11.0, 7.5) | 134.0, CH |
| 9 | 2.86 (dt, 15.3, 7.3) | 26.3, $CH_2$ |
| 9a | 2.96 (dt, 13.5, 7.6) | – |
| 10 | 5.36, m | 126.9, CH |
| 11 | 5.47, m | 131.0, CH |
| 12 | 2.19 (q, 7.6) | 23.7, $CH_2$ |
| 13 | 1.49, m | 36.5, $CH_2$ |
| 14 | 3.53 (tt, 8.3, 4.4) | 72.8, CH |
| 15 | 1.49, m | 30.5, $CH_2$ |
| 16 | 0.94 (t, 7.5) | 10.0, $CH_3$ |

## Biological activities

In our bioassays, we evaluated the inhibitory activity of all isolated compounds against a panel of microorganisms (Table 4), including *Escherichia coli* (ATCC 25922), *Candida albicans* (ATCC 76485), *Staphylococcus aureus* (ATCC 27154), *Pseudomonas fulva* (CGMCC 1.15147), and *Enterobacter hormaechei* (CGMCC 1.10608). The results indicated that compounds **3**, **5**, **6**, **7**, and **12** were active against *Candida albicans*. Notably, compound **12** showed particularly promising inhibitory activity against this fungal pathogen.

## Genome sequencing analysis

The genome sequencing yielded 7,118,236 reads with an average read length of 1,858.7 bp. The assembled genome is 34,621,366 bp long, comprising 9 contigs with a mean contig length of 3,846,818.44 bp, and the longest contig is 5,975,444 bp. The genome's GC content is 46.43%. Annotation of the genome sequence of *Penicillium shentong* XL-F41 identified 11,235 coding sequences and 172 tRNA genes.

Upon utilizing the fungal version of antiSMASH 7.0 [25] software for the analysis of the *Penicillium shentong* XL-F41 genome, we identified 46 BGCs. These include 13 NRPS-like fragments, 6 NRPS, 13 type I PKS, 2 PKS/NRPS hybrids, 1 NI-siderophore, 2 NRP-metallophore/NRPS hybrids, 1 fungal RiPP with POP or UstH peptidase types, 1 fungal-RiPP-like/T1PKS, 1 betalactone, 1 PKS type I/NRPS/indole hybrid, 1 fungal-RiPP-like/T1PKS hybrid, 1 NRP-metallophore/NRPS

**Table 4:** Antimicrobial activity of compounds **1–12**. Minimum inhibitory concentrations were shown in μg/mL.

| No. | *Escherichia coli* | *Candida albicans* | *Staphylococcus aureus* | *Pseudomonas Fulva* | *Enterobacter hormaechei* |
|---|---|---|---|---|---|
| **1** | – | – | >100 | – | – |
| **3** | – | 50–100 | – | >100 | >100 |
| **4** | – | >100 | – | – | – |
| **5** | >100 | 25–50 | >100 | >100 | >100 |
| **6** | >100 | 25–50 | >100 | >100 | >100 |
| **7** | >100 | 64–128 | >100 | >100 | >100 |
| **8** | – | – | – | >100 | >100 |
| **9** | >100 | >100 | – | >100 | >100 |
| **11** | – | – | – | – | – |
| **12** | >100 | 12.5–25 | >100 | >100 | >100 |

hybrid, NRPS-like/terpene/phosphonate hybrids, 3 terpenes, and 1 indole-related cluster (Table 5).

BGC 7.3, identified as an indole-type gene cluster, includes genes for cytochrome P450, pyridoxal-dependent decarboxylase, glutamine synthase, and tryptophan dimethyltransferase (Figure 5). These genes are likely crucial for the biosynthesis of the newly isolated alkaloids, **1** and **2**. In examining the XL-F41 genome for methyltransferase domain-containing BGCs, we found a methyltransferase near BGC 7.3, suggesting its involvement in adding a methoxy group at the C16 position of compound **1**. From these key enzyme genes, we propose a hypothetical biosynthetic pathway (Figure 5).

**Table 5:** Biosynthetic gene clusters of the *Penicillium shentong* XL-F41.

| BGC | Type | Putative product |
|---|---|---|
| 1.1 | NRPS-like | |
| 1.2 | NRPS-like | |
| 1.3 | NI-siderophore | |
| 1.4 | NPR-metallophore, NRPS | |
| 1.5 | NRPS | |
| 1.6 | PKS type I | |
| 1.7 | PKS type I | |
| 1.8 | PKS type I | |
| 1.9 | PKS type I | |
| 1.10 | NRPS-like | |
| 2.1 | NRPS-like | |
| 2.2 | NRPS-like | |
| 2.3 | PKS type I | |
| 2.4 | NRPS-like | |
| 3.1 | PKS type I | |
| 3.2 | terpene | |
| 3.3 | fungal-RiPP | |
| 3.4 | PKS type I | |
| 3.5 | PKS type I | |
| 4.1 | PKS type I | |
| 4.2 | NRPS-like | |
| 4.3 | NRPS | |
| 4.4 | NRPS-like | |
| 5.1 | PKS type I, NRPS | |
| 5.2 | NRPS-like | |
| 5.3 | PKS type I, NRPS, indole | |
| 5.4 | PKS type I | |
| 5.5 | betalactone | |
| 6.1 | NRPS | |
| 6.2 | fungal-RiPP-like, T1PKS | |
| 6.3 | PKS type I | |
| 6.4 | terpene | |
| 6.5 | NRP-metallophore, NRPS | |
| 7.1 | NRPS | |
| 7.2 | NRPS-like, terpene, phosphonate | |
| 7.3 | indole | shentonins A and B |
| 7.4 | terpene | |
| 7.5 | NRPS-like | |
| 8.1 | NRPS-like | |
| 8.2 | NRPS, PKS type I | |
| 9.1 | PKS type I | dehydrocurvularin |
| 9.2 | NRPS | |
| 9.3 | NRPS-like | |
| 9.4 | NRPS | |
| 9.5 | NRPS-like | |
| 9.6 | PKS type I | |

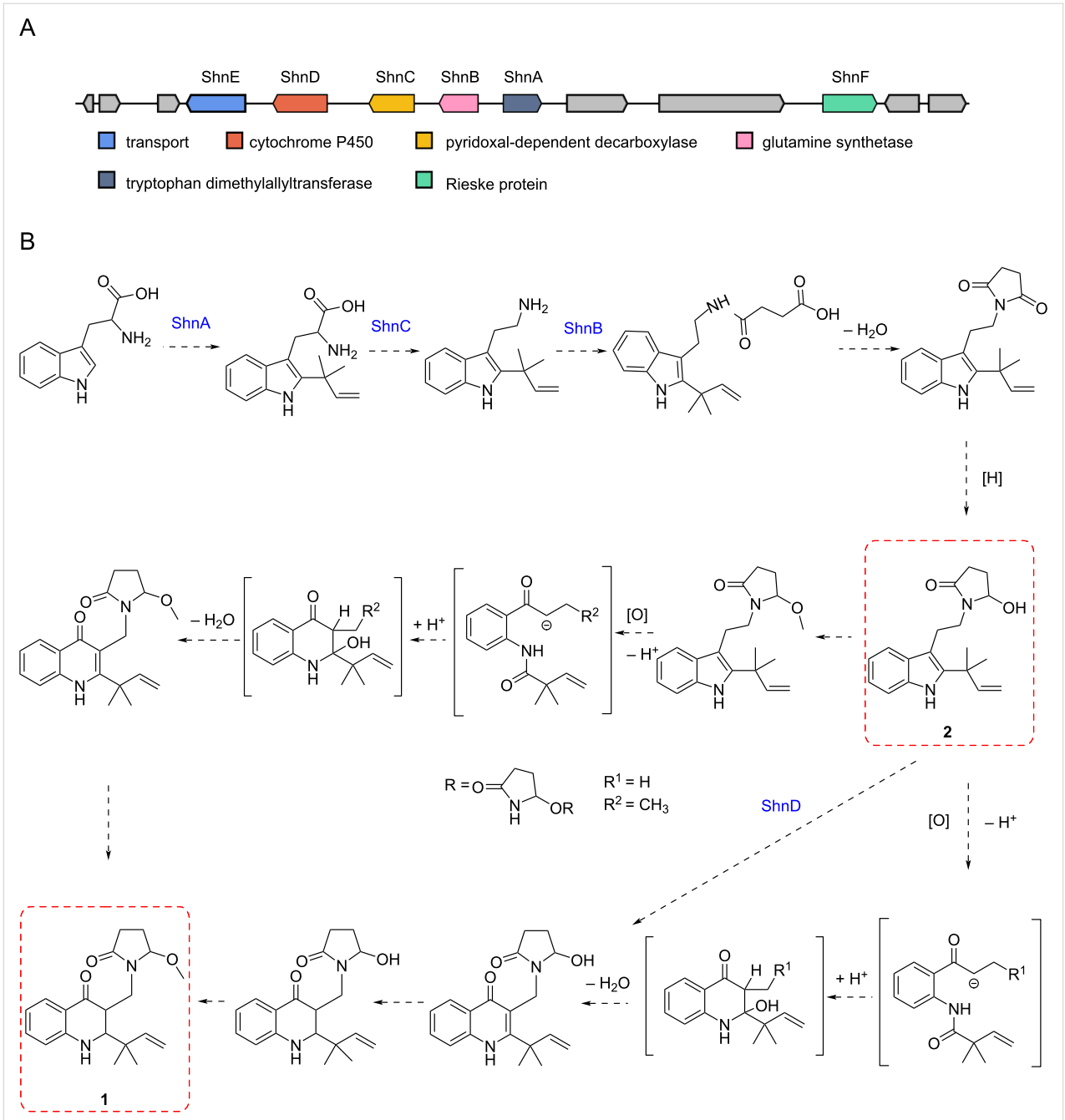



**Figure 5:** Biosynthetic exploration of compounds **1** and **2**. A: The schematic presents the biosynthetic gene cluster for compounds **1** and **2**, highlighting ShnA, ShnB, ShnC, ShnD, and ShnE as core genes. B: The diagram proposes biosynthetic pathways for compounds **1** and **2**, detailing three potential mechanisms that could convert the five-membered ring structure of compound **2** into the six-membered ring structure of compound **1**.

Compounds **1** and **2** are hypothesized to be synthesized from a tryptophan precursor via a shared biosynthetic pathway (Figure 5). Briefly, the prenyl group is attached to tryptophan through a prenylation reaction catalyzed by ShnA, followed by the decarboxylation of the carboxy group by ShnC. Subsequently, compound **2** is formed by the addition of succinimide to N15 in **1** via a reaction catalyzed by ShnB. The transformation of the five-membered pyrrole ring in compound **2** to the six-membered ring in compound **1** is particularly intriguing. For this transformation, three hypotheses are considered. One suggests that the methyl modification at the oxygen atom of the succinimide ring occurs first, which is then followed by a ring-opening rearrangement. Alternatively, it is proposed that the ring-opening rearrangement precedes the methyl modification at the oxygen atom of the succinimide ring.

We aim to confirm the initial step of this pathway, where tryptophan and DMAPP are catalyzed by the enzyme ShnA to form a reverse prenylated tryptophan. However, attempts to express the protein in various *Escherichia coli* hosts were unsuccessful, suggesting that eukaryotic hosts might be more suitable for future studies. We plan to conduct further experiments to substantiate the hypothesis regarding the biosynthetic pathways in the future.

## Conclusion

In the present study, we fermented *Penicillium shentong* XL-F41 by adding a series of elicitors in the medium, which led us to identify twelve compounds, including two new indole alkaloids, shentonins A and B (**1** and **2**), and a new fatty acid (**3**).

Notably, compound **1** differs from the known brocaeloid D by the addition of a methyl group, and there is a change in the relative stereochemistry at C2 and C3. In addition, the conversion of the five-membered pyrrole ring in compound **2** to the six-membered piperidine ring in compound **1** is intriguing.

Although there have been many reports on the biosynthetic pathways from tryptophan to quinoline rings. Through our analysis, we discovered that tryptophan in the primary metabolic pathway is primarily catalyzed by indoleamine 2,3-dioxygenase and tryptophan 2,3-dioxygenase (IDO and TDO), as well as canine urinary tryptophan 3-monooxygenase, to form quinoline rings [26]. Quinine is frequently cited as one of the primary forms of quinoline rings in secondary metabolic pathways. Francesco Trenti et al. [27] studied some of the biosynthesis processes of quinine, in which enzymes involved are much more complex than primary metabolism, such as medium-chain alcohol dehydrogenase (*Cp*DCS), esterase (*Cp*DCE), P450 and O-methyltransferase (*Cp*OMT1) (Figure S37). Through genome excavation and analysis of *Penicillium shentong* XL-F41, a significant difference was discovered between the key enzymes involved in the formation of product compound **1** and those previously reported. This suggests that the formation of the quinoline ring in compound **1** may represent a new and unreported biosynthetic pathway.

Moreover, to address the low yields that hindered the determination of absolute stereochemistry, we attempted to boost the production of compounds **1** and **2** by supplementing the medium with the precursor tryptophan. Contrary to our expectations, this approach did not increase their production. Our next step is to plan the heterologous expression of core genes in proper fungal hosts to improve production and investigate the biosynthetic pathways of compounds **1** and **2**.

Furthermore, we conducted a genome sequencing analysis of *Penicillium shentong* XL-F41, which allowed us to pinpoint the biosynthetic gene clusters (BGCs) associated with our isolated compounds and reveal the biosynthetic capabilities of this strain. Despite the addition of various elicitors to the fermentation medium, numerous BGCs remain uncharacterized, indicating that additional strategies are required to fully elucidate the compounds encoded by all BGCs.

## Experimental

### General experimental procedures

HRESIMS spectra were acquired using a Waters ACQUITY UPLC I-Class-Vion IMS Q-Tof liquid chromatograph mass spectrometer. For NMR analysis, we utilized an AVANCE II 600 spectrometer, referencing residual solvent peaks at $\delta_H/\delta_C$ 7.27/77.0 ppm in $CDCl_3$ for chemical shift calibration. We utilized commercial silica gel from Yantai Xinnuo New Material Technology Co., Ltd, Yantai, China, available in 100–200 and 200–300 mesh sizes. Reversed-phase HPLC analyses were conducted on an Agilent 1260 instrument equipped with a DAD detector and an Agilent ZORBAX SB-C18 column (5 μm, 4.6 × 150 mm). The solvents used for HPLC were supplied by Yantai Huisente New Material Technology Co., Ltd, Yantai, China.

### Fungus isolation and characterization

The fungus *Penicillium shentong* XL-F41 was isolated from soil collected in Shentong Mountain, Shandong Province, China, in June 2022. It was cultured on potato dextrose agar (PDA) at 28 °C. Sequencing and comparison with the GenBank database confirmed its identification as *Penicillium* sp. This strain is preserved at the Shandong Laboratory of Yantai Drug Discovery.

### Fermentation in shaking flasks

For large-scale fermentations, fresh mycelia of *Penicillium shentong* XL-F41 were first cultivated in liquid potato dextrose broth at 28 °C for 2 days. Subsequently, they were inoculated into XISR I and XISR III liquid media (total volume 30 L) with a 20% inoculum dose. The cultures were further fermented for 14 days at 28 °C and 200 rpm.

#### Media recipes

XISR I medium (yeast extract 4 g/L; malt extract 10 g/L; glucose 4 g/L; $MgCl_2$ 1 μM; $FeSO_4$ 1 μM; KI 2 g/L, KCl 2 g/L, KBr 2 g/L, $NaNO_2$ 2 g/L; $H_2O_2$ 20 μM; (methyl jasmonate) MeJA 10 μM).

XISR III medium (yeast extract 4 g/L; soy flour 10 g/L; glucose 30 g/L; $MgCl_2$ 1 μM; $FeSO_4$ 1 μM; KI 2 g/L, KCl 2 g/L, KBr 2 g/L, $NaNO_2$ 2 g/L; $H_2O_2$ 20 μM; MeJA 10 μM; 5-azacyti-

dine 6 μM; suberoylanilide hydroxamic acid 6 μM; sodium butyrate 6 μM).

## Extraction

The mycelium was separated from the fermentation broth using a centrifuge and subsequently extracted with ethanol in a 1:1 ratio using ultrasound, three times for 20 minutes each. The combined organic solvents were dried with a rotary evaporator to yield an ethanol extract. This extract was further processed with ethyl acetate (EtOAc) three times. The combined EtOAc phase was then dried using a rotary evaporator to obtain the EtOAc extract, which was stored at −80 °C until further purification process.

The above fermentation broth was adsorbed onto macroporous resin for 4 hours or left overnight. It was then eluted with deionized water and ethanol through a chromatography column. The ethanol eluate was concentrated to dryness using a rotary evaporator. Subsequently, the ethanol extract underwent a triple extraction with EtOAc. The combined EtOAc extracts were dried using a rotary evaporator to obtain the final EtOAc extract, which was stored at −80 °C until further isolation process.

## Compounds purification

The EtOAc extract (3.5 g) obtained from the fermentation broth of XISR I medium underwent column chromatography on silica using a gradient of petroleum ether–ethyl acetate (PE–EA) and ethyl acetate–methanol (EA–MeOH) to yield 22 fractions (Fr.A–Fr.N) as determined by TLC analysis. Fraction Fr.J (62.32 mg) was further purified using silica gel column chromatography with a PE–EtOAc gradient to obtain nine subfractions (Fr.J1–Fr.J9). Subfraction Fr.J7 (17 mg) was then subjected to RP-HPLC with an acetonitrile–water gradient, resulting in the isolation of compounds **1** (1.36 mg) and **2** (0.36 mg).

Similarly, the EtOAc extract (9.3 g) from the fermentation broth of XISR III medium was fractionated by column chromatography on silica with PE–EA and EA–MeOH gradients to yield 22 fractions (Fr.1–Fr.22) based on TLC analysis. Fraction Fr.12 (200 mg) was purified using a PE–EtOAc gradient to produce seven subfractions (Fr.a–Fr.g). Subfraction Fr.e (23.19 mg) was processed using RP-HPLC with an acetonitrile–water gradient, leading to the isolation of compound **3** (2.19 mg).

### Physical and spectroscopic data of compounds **1–3**

Shentonin A (**1**): light green solid; $[\alpha]_D^{20}$ +40.0 (*c* 0.17, $CH_3OH$); UV ($CH_3OH$) $\lambda_{max}$, nm (log ε): 400 (3.45), 240 (4.24) nm; IR (KBr) $\nu_{max}$: 3347, 2960, 2926, 1688, 1654, 1612, 1260, 1078, 1021, 797 $cm^{-1}$; for $^1H$ NMR ($CDCl_3$, 600 MHz) and $^{13}C$ NMR ($CDCl_3$, 125 MHz) spectral data, see Table 1; HRESIMS (*m/z*): $[M - H]^-$ calcd for 341.1870; found, 341.1862.

Shentonin B (**2**): light green solid; $[\alpha]_D^{20}$ +5.3 (*c* 0.07, $CH_3OH$); UV ($CH_3OH$) $\lambda_{max}$m nm (log ε): 280 (3.59), 220 (4.30) nm; IR (KBr) $\nu_{max}$: 3315, 2969, 1667, 1461, 1159, 1138, 1061, 981, 914, 742 $cm^{-1}$; for $^1H$ NMR ($CDCl_3$, 600 MHz) and $^{13}C$ NMR ($CDCl_3$, 125 MHz) spectral data, see Table 1; HRESIMS (*m/z*): $[M - H]^-$ calcd for 311.1765; found, 311.1755.

Compound **3**: transparent oily liquid; $[\alpha]_D^{20}$ −46.7 (*c* 0.18, $CH_3OH$); IR (KBr) $\nu_{max}$: 3432, 2960, 2923, 1762, 1450, 1260, 1180, 1016, 800 $cm^{-1}$; for $^1H$ NMR ($CDCl_3$, 600 MHz) and $^{13}C$ NMR ($CDCl_3$, 125 MHz) spectral data, see Table 1; LC–MS (*m/z*): $[M - H]^-$ calcd for 283.2; found, 283.2.

## Antimicrobial activity evaluation

All isolated compounds were dissolved in 1% DMSO and introduced to pathogenic bacteria or fungi in LB or PDB media. The 96-well plates were incubated at 37 °C for 18 hours, with 1% DMSO serving as the negative control. After incubation, the $OD_{600}$ of the bacterial cultures was measured using a Microplate Reader.

## Supporting Information

### Supporting Information File 1

NMR and mass spectra of isolated compounds.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-52-S1.pdf]


# Acknowledgements

We thank Congcong Guo and the Pharmaceutical Analysis and Quality Research Platform of Bohai Rim Advanced Research Institute for Drug Discovery for all the technical support in NMR data collection.

# Funding

B.X. thanks for the financial support of the National Key Research and Development Program of China (No. 2022YFC2804100), the Taishan Scholars Program, the Shandong Laboratory Program (SYS202205), the Shanghai Pujiang Program (23PJ1415300), and the start-up funding provided by the Shanghai Institute of Materia Medica, Chinese Academy of Sciences.

## Author Contributions

Ran Zou: investigation; writing – original draft. Xin Li: investigation. Xiaochen Chen: investigation. Yue-Wei Guo: visualization; writing – review & editing. Baofu Xu: conceptualization; funding acquisition; methodology; project administration; supervision; visualization; writing – review & editing.

## ORCID® iDs

Baofu Xu - https://orcid.org/0000-0002-0144-2455

## Data Availability Statement

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

## Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2023.58.v1

## References

1. El Hajj Assaf, C.; Zetina-Serrano, C.; Tahtah, N.; El Khoury, A.; Atoui, A.; Oswald, I. P.; Puel, O.; Lorber, S. *Int. J. Mol. Sci.* **2020,** *21,* 9462. doi:10.3390/ijms21249462
2. Bentley, R. *Chem. Rev.* **2000,** *100,* 3801–3826. doi:10.1021/cr990097b
3. Koul, M.; Singh, S. *Anticancer Drugs* **2017,** *28,* 11–30. doi:10.1097/cad.0000000000000423
4. Ma, H.-G.; Liu, Q.; Zhu, G.-L.; Liu, H.-S.; Zhu, W.-M. *J. Asian Nat. Prod. Res.* **2016,** *18,* 92–115. doi:10.1080/10286020.2015.1127230
5. Zhang, X.; Yin, Q.; Li, X.; Liu, X.; Lei, H.; Wu, B. *Fitoterapia* **2022,** *163,* 105349. doi:10.1016/j.fitote.2022.105349
6. Mondal, A.; Gandhi, A.; Fimognari, C.; Atanasov, A. G.; Bishayee, A. *Eur. J. Pharmacol.* **2019,** *858,* 172472. doi:10.1016/j.ejphar.2019.172472
7. Qiu, S.; Sun, H.; Zhang, A. H.; Xu, H. Y.; Yan, G. L.; Han, Y.; Wang, X. J. *Chin. J. Nat. Med.* **2014,** *12,* 401–406. doi:10.1016/s1875-5364(14)60063-7
8. Rosales, P. F.; Bordin, G. S.; Gower, A. E.; Moura, S. *Fitoterapia* **2020,** *143,* 104558. doi:10.1016/j.fitote.2020.104558
9. Song, J.; Zhang, B.; Li, M.; Zhang, J. *Fitoterapia* **2023,** *165,* 105430. doi:10.1016/j.fitote.2023.105430
10. Liu, Y.; Cui, Y.; Lu, L.; Gong, Y.; Han, W.; Piao, G. *Arch. Pharm. (Weinheim, Ger.)* **2020,** *353,* 2000120. doi:10.1002/ardp.202000120
11. Zou, R.; Chen, B.; Sun, J.; Guo, Y.-W.; Xu, B. *Fitoterapia* **2023,** *167,* 105503. doi:10.1016/j.fitote.2023.105503
12. Brakhage, A. A. *Nat. Rev. Microbiol.* **2013,** *11,* 21–32. doi:10.1038/nrmicro2916
13. Brakhage, A. A.; Schroeckh, V. *Fungal Genet. Biol.* **2011,** *48,* 15–22. doi:10.1016/j.fgb.2010.04.004
14. Kumar, A.; Asthana, M.; Gupta, A.; Nigam, D.; Mahajan, S. Secondary Metabolism and Antimicrobial Metabolites of *Penicillium*. *New and Future Developments in Microbial Biotechnology and Bioengineering;* Elsevier: Amsterdam, Netherlands, 2018; pp 47–68. doi:10.1016/b978-0-444-63501-3.00003-x
15. Dai, X.-M.; Pan, H.-L.; Lan, W.-J.; Chen, L.-P.; Feng, G.-K.; Deng, R.; Zhu, X.-F.; Li, H.-J. *Phytochemistry* **2022,** *204,* 113456. doi:10.1016/j.phytochem.2022.113456
16. Arai, K.; Rawlings, B. J.; Yoshizawa, Y.; Vederas, J. C. *J. Am. Chem. Soc.* **1989,** *111,* 3391–3399. doi:10.1021/ja00191a042
17. Liu, Y.; Li, Z.; Vederas, J. C. *Tetrahedron* **1998,** *54,* 15937–15958. doi:10.1016/s0040-4020(98)01003-5
18. Ghisalberti, E. L.; Hockless, D. C. R.; Rowland, C. Y.; White, A. H. *Aust. J. Chem.* **1993,** *46,* 571–575. doi:10.1071/ch9930571
19. Choochuay, J.; Xu, X.; Rukachaisirikul, V.; Guedduaythong, P.; Phongpaichit, S.; Sakayaroj, J.; Chen, J.; Shen, X. *Phytochem. Lett.* **2017,** *22,* 122–127. doi:10.1016/j.phytol.2017.09.011
20. Shaala, L. A.; Youssef, D. T. A.; Badr, J. M.; Harakeh, S. M. *Molecules* **2016,** *21,* 1116. doi:10.3390/molecules21091116
21. Lee, S. J.; Baek, H. J.; Lee, C. H.; Kim, H. P. *Arch. Pharmacal Res.* **1994,** *17,* 31–35. doi:10.1007/bf02978244
22. Selepe, M. A.; Drewes, S. E.; van Heerden, F. R. *J. Nat. Prod.* **2010,** *73,* 1680–1685. doi:10.1021/np100407n
23. Zhang, J.; Wang, Z.; Song, Z.; Karthik, L.; Hou, C.; Zhu, G.; Jiang, L.; Han, J.; Ma, R.; Li, L.; Zhang, L.; Liu, X.; Hsiang, T. *Synth. Syst. Biotechnol.* **2019,** *4,* 173–179. doi:10.1016/j.synbio.2019.09.001
24. Zhang, P.; Meng, L.-H.; Mándi, A.; Kurtán, T.; Li, X.-M.; Liu, Y.; Li, X.; Li, C.-S.; Wang, B.-G. *Eur. J. Org. Chem.* **2014,** 4029–4036. doi:10.1002/ejoc.201400067
25. Blin, K.; Shaw, S.; Augustijn, H. E.; Reitz, Z. L.; Biermann, F.; Alanjary, M.; Fetter, A.; Terlouw, B. R.; Metclaf, W. W.; Helfrich, E. J. N.; van Wezel, G. P.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2023,** *51,* W46–W50. doi:10.1093/nar/gkad344
26. Červenka, I.; Agudelo, L. Z.; Ruas, J. L. *Science* **2017,** *357,* eaaf9794. doi:10.1126/science.aaf9794
27. Trenti, F.; Yamamoto, K.; Hong, B.; Paetz, C.; Nakamura, Y.; O'Connor, S. E. *Org. Lett.* **2021,** *23,* 1793–1797. doi:10.1021/acs.orglett.1c00206

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at: https://doi.org/10.3762/bjoc.20.52

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Production of non-natural 5-methylorsellinate-derived meroterpenoids in *Aspergillus oryzae*

Jia Tang, Yixiang Zhang and Yudai Matsuda*

Letter

Open Access

Address:
Department of Chemistry, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China

Email:
Yudai Matsuda* - ymatsuda@cityu.edu.hk

* Corresponding author

Keywords:
biosynthesis; meroterpenoids; natural products; pathway engineering; terpene cyclases

*Beilstein J. Org. Chem.* **2024,** *20,* 638–644.
https://doi.org/10.3762/bjoc.20.56

Received: 16 October 2023
Accepted: 15 March 2024
Published: 20 March 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra

© 2024 Tang et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Fungal meroterpenoids are diverse structurally intriguing molecules with various biological properties. One large group within this compound class is derived from the aromatic precursor 3,5-dimethylorsellinic acid (DMOA). In this study, we constructed engineered metabolic pathways in the fungus *Aspergillus oryzae* to expand the molecular diversity of meroterpenoids. We employed the 5-methylorsellinic acid (5-MOA) synthase FncE and three additional biosynthetic enzymes for the formation of (6*R*,10′*R*)-epoxyfarnesyl-5-MOA methyl ester, which served as a non-native substrate for four terpene cyclases from DMOA-derived meroterpenoid pathways. As a result, we successfully generated six unnatural 5-MOA-derived meroterpenoid species, demonstrating the effectiveness of our approach in the generation of structural analogues of meroterpenoids.

## Introduction

Meroterpenoids are a class of natural products partially biosynthesized from a terpenoid pathway; their non-terpenoid portions can be polyketides, indole, or shikimate-derived compounds [1-3]. Their hybrid nature significantly contributes to their structural diversity and wide range of biological activities. Although meroterpenoids are found ubiquitously in nature, as both primary and secondary metabolites, filamentous fungi stand out as the most prolific producers of meroterpenoids [1-3]. Representative fungal meroterpenoids of medicinal importance include pyripyropene A, a cholesterol acyltransferase inhibitor [4]; fumagillin, an antimicrobial agent [5]; and mycophenolic acid, a strong inosine 5-monophosphate dehydrogenase inhibitor [6]. The biosynthesis of fungal meroterpenoids has garnered interest in the organic chemistry field due to their structural complexity and associated intriguing enzymatic reactions and has thus been extensively researched for over a decade, providing a general understanding of their biosynthesis [7,8].

Polyketide–terpenoid hybrids are among the largest families of meroterpenoids. Orsellinic acid, an aromatic polyketide, and its

analogues have been commonly identified as polyketide components in fungal meroterpenoids. Notably, 3,5-dimethylorsellinic acid (DMOA) serves as a precursor for a wide array of structurally diverse meroterpenoid species [7,8]. The fully substituted nature of DMOA leads to dearomatizing prenylation during the biosynthesis of DMOA-derived meroterpenoids (Figure 1A), facilitating the rearrangement reactions of the polyketide moiety, contributing to structural diversification [9,10]. By contrast, in the biosynthesis of meroterpenoids derived from orsellinic acid and 5-methylorsellinic acid (5-MOA), the prenylation reaction typically occurs at the non-substituted carbon atom and thus preserves the aromaticity of the polyketide portion [11-14]. One exception has been found in funiculolide biosynthesis, in which a 5-MOA-derived phthalide undergoes dearomatizing prenylation catalyzed by the UbiA-like prenyltransferase FncB (Figure 1B) [15]. In addition to prenyltransferases, transmembrane terpene cyclases play a key role in diversifying the structures of fungal meroterpenoids [16]. For example, (6*R*,10′*R*)-epoxyfarnesyl-DMOA methyl ester, a common intermediate with a linear terpenoid moiety, is known to be recognized by five different enzymes, namely Trt1, AusL, AdrI, InsA7, and InsB2, resulting in conversion into distinct cyclized products (Figure 1C) [17-19]. In addition, a recent study has demonstrated that some of these transmembrane terpene cyclases can accept synthetic substrate analogues to yield several unnatural meroterpenoid molecules [20]. By

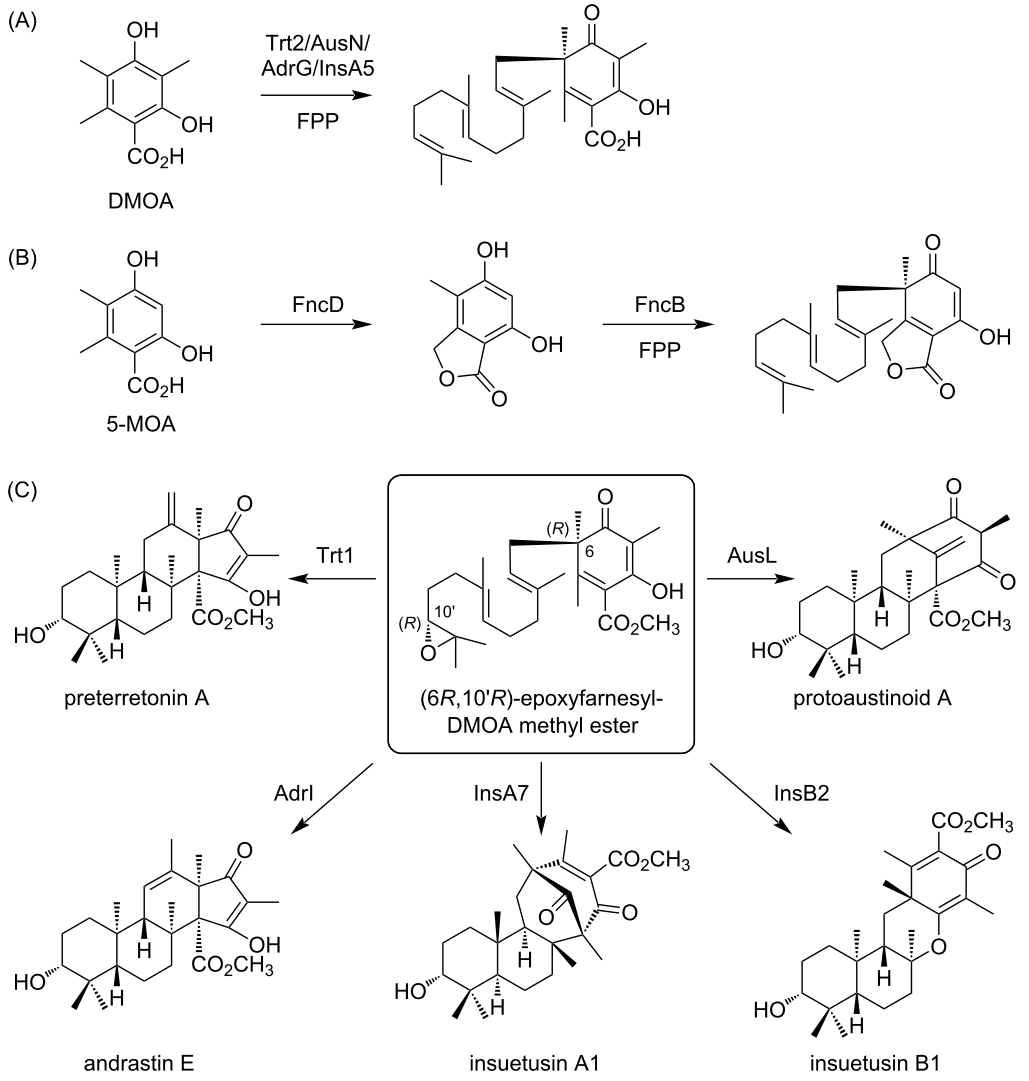



**Figure 1:** Biosynthesis of selected fungal meroterpenoids. (A, B) Dearomatizing prenylation reactions in the biosynthesis of (A) 3,5-dimethylorsellinic acid (DMOA)- and (B) 5-methylorsellinic acid (5-MOA)-derived meroterpenoids. (C) Reactions catalyzed by the terpene cyclases involved in DMOA-derived meroterpenoid pathways.

mimicking nature's strategy to synthesize diverse meroterpenoids, we can access meroterpenoids that have not yet been reported.

In this study, we aimed to expand the structural diversity of meroterpenoids through artificial pathway construction. To achieve this, we constructed a series of artificial metabolic pathways by harnessing genes from the funiculolide biosynthetic pathway and those from DMOA-derived meroterpenoid pathways, successfully yielding six unnatural 5-MOA-derived meroterpenoid species.

## Results and Discussion

To synthesize unnatural meroterpenoid molecules, we sought to generate a series of 5-MOA-derived meroterpenoids by utilizing the terpene cyclases responsible for DMOA-derived compounds. To achieve this goal, we first aimed to establish a production system for the 4-desmethyl analogue of (6*R*,10′*R*)-epoxyfarnesyl-DMOA methyl ester by utilizing the polyketide synthase FncE, the prenyltransferase FncB, the *O*-methyltransferase InsA1, and the FAD-dependent monooxygenase InsA4. In this engineered pathway, FncE first synthesizes 5-MOA, which then undergoes farnesylation by FncB, methyl ester formation by InsA1, and epoxidation of the terminal olefin in the farnesyl moiety by InsA4 (Figure 2A). Thus, we heterologously expressed the genes encoding these four enzymes in the *Aspergillus oryzae* NSARU1 strain [19]. Consequently, the *A. oryzae* transformant yielded two metabolites **1** and **2**, which were absent in the host strain (Figure 2B, traces i and ii). Although we were unable to isolate compounds **1** and **2** because of their instability, high-resolution mass spectrometry (HRMS) analysis revealed the molecular formula of **1** to be $C_{25}H_{36}O_5$, corresponding to the 4-desmethyl form of (6*R*,10′*R*)-epoxyfarnesyl-DMOA methyl ester (Figure 2C). Furthermore, the molecular formula of **2** was determined to be $C_{25}H_{38}O_6$, indicating that **2** is formed by the hydrolysis of the epoxide ring in **1** (Figure 2B). Taken together, this observation suggests that the desired production system was successfully established. Notably, although the native substrate of FncB is believed to be the phthalide form of 5-MOA (Figure 1B) [15], this result demonstrates that it can also efficiently accept 5-MOA as a substrate for dearomatizing farnesylation.

Subsequently, we introduced five terpene cyclase genes involved in DMOA-derived meroterpenoid biosynthesis, namely, *adrI*, *trt1*, *ausL*, *insA7*, and *insB2*, individually into the *A. oryzae* transformant that already expresses the four genes constructed earlier. We then analyzed the metabolites from the resulting transformants using high-performance liquid chromatography (HPLC), which revealed that all of the enzymes, except AdrI, accepted **1** and produced 5-MOA-derived meroterpenoids (Figure 2B, traces iii to vii). Since the transformation plasmid with *adrI* used in this study was constructed in our previous study, in which the product of AdrI was clearly detected [21], the inability of AdrI to yield a cyclized product is not likely to be caused by an inactive protein. The *trt1*-transformed strain produced a new compound **3** (molecular formula: $C_{25}H_{36}O_5$). After large-scale cultivation, **3** was isolated and subjected to nuclear magnetic resonance (NMR) analysis, which suggested that **3** is the C-5′ desmethyl form of preterretonin A [17]. However, several missing signals in the $^{13}C$ NMR spectrum, likely due to keto–enol tautomerization in the D-ring, hindered the complete structural determination of **3**. To overcome this challenge in structural determination, we obtained a single crystal of **3** and performed X-ray diffraction analysis, which unambiguously established the structure of **3** as 5′-desmethylpreterretonin A (Figure 2C and Figure S1 in Supporting Information File 1; CCDC: 2300693). The *A. oryzae* transformant with *ausL* yielded two products **4** and **5**. The major product **4** was identified as the C-5′ desmethyl analogue of protoaustinoid A and thus named 5′-desmethylprotoaustinoid A (Figure 2C) [17,22]. Meanwhile, the minor product **5** was determined as the C-5′ desmethyl form of the product from the K187A variant of AdrI, which was created during the in-depth functional analysis of terpene cyclases involved in DMOA-derived meroterpenoid biosynthesis [21], through NMR and single-crystal X-ray crystallographic analyses (Figure 2C and Figure S1 in Supporting Information File 1; CCDC: 2300694). Interestingly, compound **5** was also detected as a major product from the *A. oryzae* transformant harboring *insB2*, whereas the desmethyl analogue of the original InsB2 product [19], 5′-desmethylinsuetusin B1 (**6**) (Figure 2C), was only obtained as a minor product. Finally, the *A. oryzae* strain expressing *insA7* produced two major metabolites **7** and **8**. Compound **7** was determined to be the C-5′ desmethyl form of insuetusin A1 [19] using NMR and single-crystal X-ray diffraction analyses (Figure 2C and Figure S1 in Supporting Information File 1; CCDC: 2300695) and was designated as 5′-demethylinsuetusin A1. Unlike compounds **3**–**7**, the molecular formula of **8** was determined to be $C_{25}H_{38}O_5$, which corresponds to the hydrogenated form of **7**. NMR analysis revealed that the double bond at C-2′/C-3′ of **7** was reduced to a single bond in **8** and that **8** contained the enol functionality instead of the C-4′ carbonyl group (Figure 2C). It is unlikely that **8** is the direct product of InsA7; thus, we hypothesized that an endogenous enzyme in *A. oryzae* is responsible for the reduction, with an enoylreductase first reducing the C-2′/C-3′ double bond of **7** and the resulting product undergoing keto–enol tautomerization to form **8** (Figure S2, Supporting Information File 1). Incubation of **7** with the host *A. oryzae* strain indeed resulted in the formation of **8** (Figure S3, Supporting Information File 1), confirming that InsA7 is not involved in the enoylreduction.

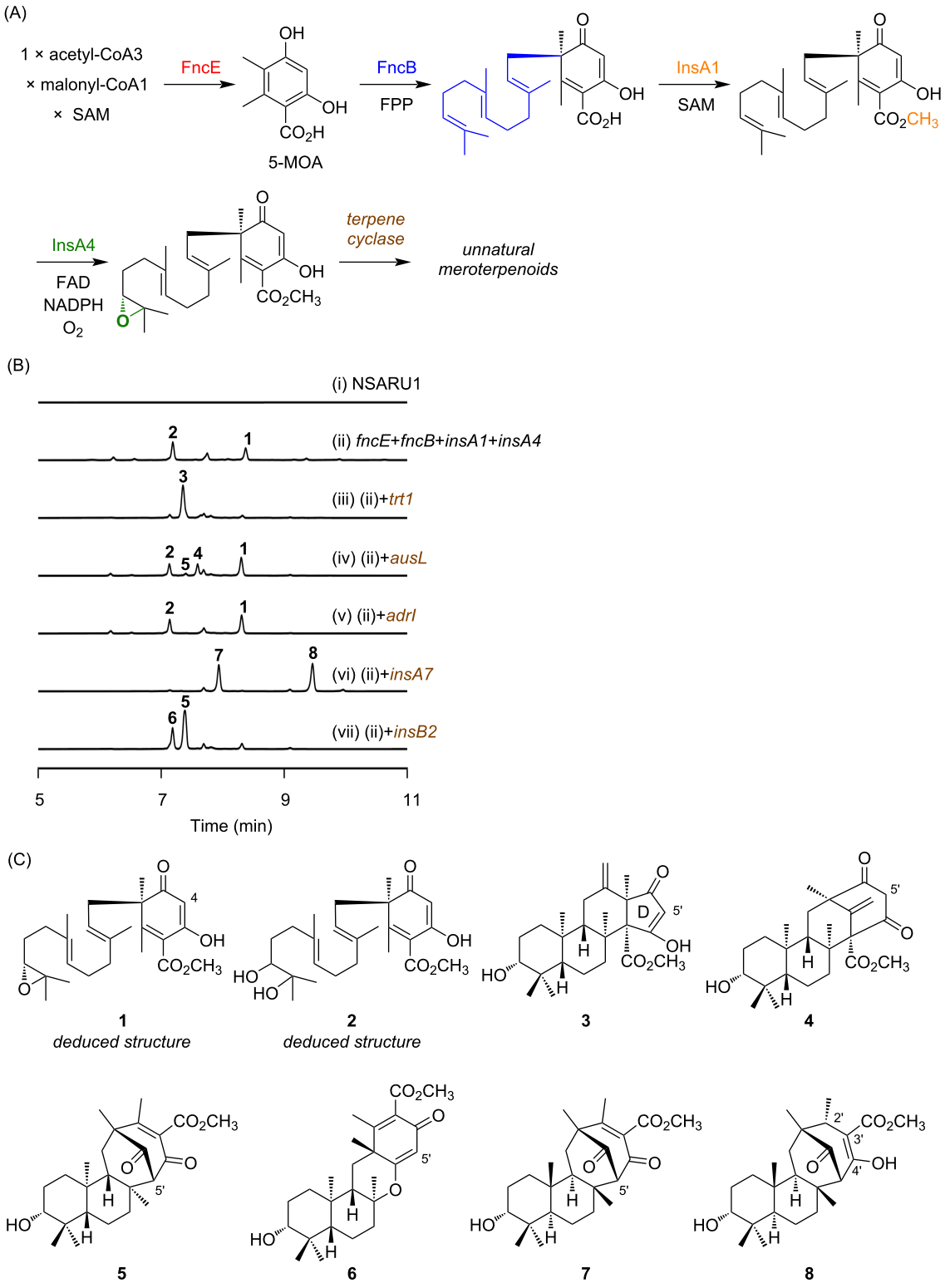



**Figure 2:** Generation of unnatural 5-MOA-derived meroterpenoids. (A) Working concept to synthesize unnatural 5-MOA-derived meroterpenoids. SAM: *S*-adenosyl-L-methionine; FPP: farnesyl pyrophosphate; FAD: flavin adenine dinucleotide; NADPH: nicotinamide adenine dinucleotide phosphate. (B) HPLC profiles of the metabolites from *Aspergillus oryzae* transformants. The chromatograms were extracted at 254 nm. (C) Structures of metabolites detected or isolated in this study. Note that the structures of **1** and **2** were deduced based on their HRMS spectra and the predicted biosynthetic pathway.

Our findings revealed that four of the five tested terpene cyclases could accept the desmethyl form of their native substrate, although the cyclization by AusL occurs less efficiently than by the other three enzymes. The cyclized products obtained in this study are all previously undescribed meroterpenoid species. The cyclization of **1** appears to occur in a manner similar to that of (6*R*,10′*R*)-epoxyfarnesyl-DMOA methyl ester (Figure 3). Although the major products from Trt1, AusL, and InsA7 are the C-5′ desmethyl forms of their original products, InsB2 displayed an altered cyclization preference toward the desmethyl analogue **1**, indicating that the missing methyl group significantly influences the folding mode of the substrate in InsB2. It remains unclear why AdrI was unable to accept **1**, but it is possible that the C-4 methyl group of (6*R*,10′*R*)-epoxyfarnesyl-DMOA methyl ester plays a critical role in substrate recognition by AdrI.

Finally, we evaluated the antibacterial activities of the compounds obtained in this study. As a result, only compound **3** exhibited weak activity against *Staphylococcus aureus* ATCC 6538 and *Bacillus cereus* with the minimum inhibitory concentration (MIC) of 500 μg/mL, respectively. Preterretonin A, the DMOA-derived counterpart of **3**, displayed weaker activity against the two bacterial strains and was not active at the con-

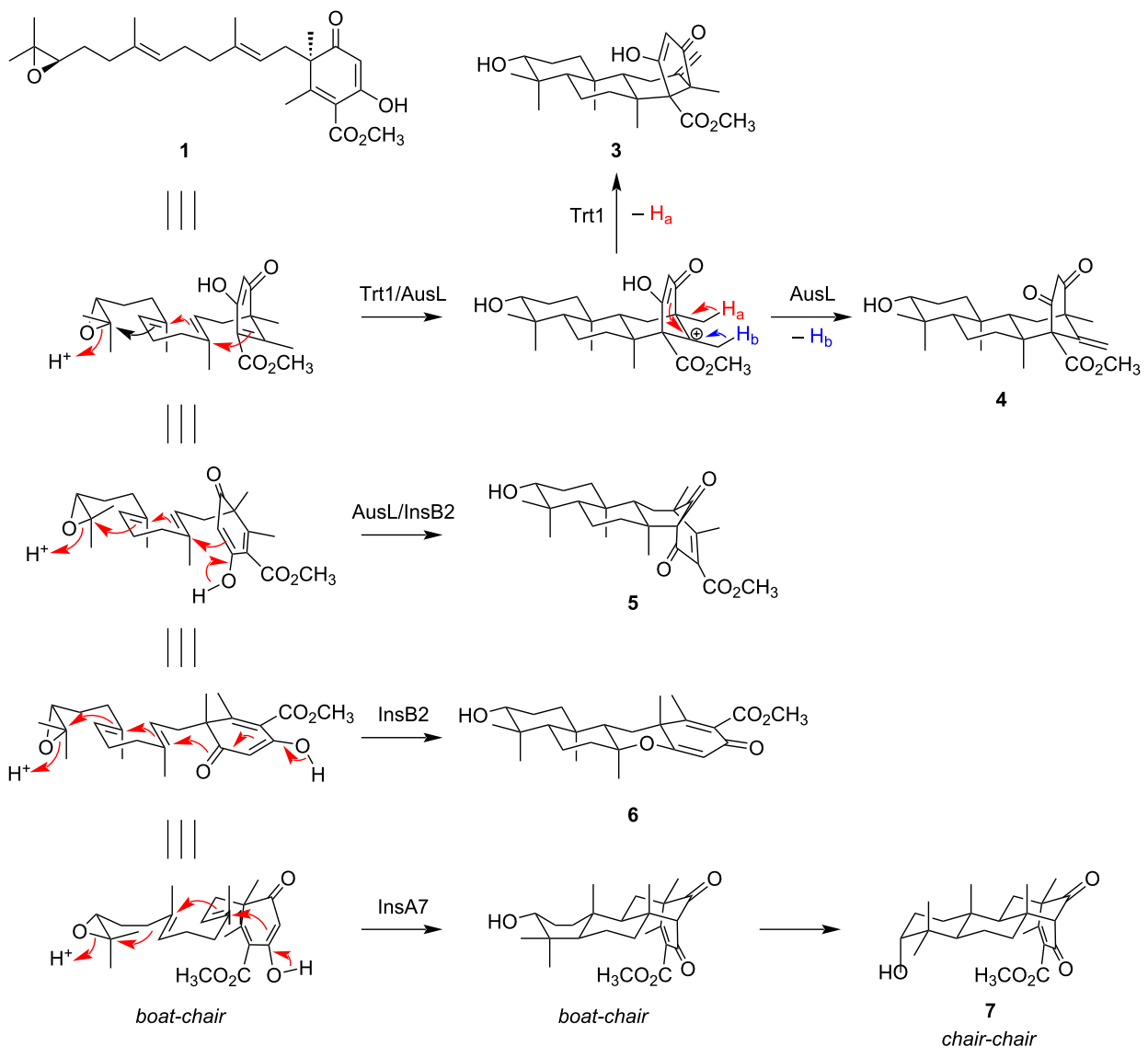



**Figure 3:** Biosynthetic mechanisms of the 5-MOA-derived meroterpenoids obtained in this study. In the reaction by InsA7, the cyclization should proceed via a pre-boat-chair conformation, and the conformational change after cyclization would afford the chair-chair conformation of **7** [19].

centration of 500 µg/mL. Thus, the lack of the C-5′ methyl group somehow contributes to the biological activity of **3**.

## Conclusion

We employed a pathway engineering approach to create a series of "unnatural" natural products by rationally combining the biosynthesis of DMOA- and 5-MOA-derived meroterpenoids. To the best of our knowledge, our study provides the first examples in which meroterpenoid species are generated via the direct dearomatizing prenylation of 5-MOA. Despite its simplicity, this strategy yielded six previously unreported meroterpenoids, demonstrating the effectiveness of our methodology in discovering new natural products. Given the increasing elucidation of fungal meroterpenoid pathways in recent years, similar approaches could be applied to other meroterpenoid biosynthetic processes, as reported in a recent study that yielded new DMOA-derived meroterpenoids with a monocyclic terpenoid moiety through pathway engineering [23]. Furthermore, the meroterpenoids produced in this study could be accepted by downstream enzymes in each pathway, further expanding the structural diversity of fungal meroterpenoids and potentially aiding the functional characterization of these tailoring enzymes.

## Supporting Information

### Supporting Information File 1

Experimental details, analytical data, tables of primer sequences, constructed plasmids, and *A. oryzae* transformants and figures showing the X-ray crystal structures, the biosynthetic pathway, and NMR data and spectra.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-56-S1.pdf]

## Acknowledgements

We thank Prof. Katsuya Gomi (Tohoku University), Prof. Katsuhiko Kitamoto (University of Tokyo), and Prof. Jun-ichi Maruyama (University of Tokyo) for providing the expression vectors and fungal strain. We are grateful to Dr. Man-Kit Tse (City University of Hong Kong) and Dr. Shek-Man Yiu (City University of Hong Kong) for their assistance in NMR spectra acquisition and X-ray diffraction data collection and analysis, respectively.

## Funding

This work was supported by a General Research Fund grant from the Research Grants Council of Hong Kong (Project No. 11301321).

## ORCID® iDs

Jia Tang - https://orcid.org/0000-0001-8612-4140
Yudai Matsuda - https://orcid.org/0000-0001-5650-4732

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Geris, R.; Simpson, T. J. *Nat. Prod. Rep.* **2009,** *26,* 1063–1094. doi:10.1039/b820413f
2. Jiang, M.; Wu, Z.; Liu, L.; Chen, S. *Org. Biomol. Chem.* **2021,** *19,* 1644–1704. doi:10.1039/d0ob02162h
3. Zhao, M.; Tang, Y.; Xie, J.; Zhao, Z.; Cui, H. *Eur. J. Med. Chem.* **2021,** *209,* 112860. doi:10.1016/j.ejmech.2020.112860
4. Obata, R.; Sunazuka, T.; Tomoda, H.; Harigaya, Y.; Ōmura, S. *Bioorg. Med. Chem. Lett.* **1995,** *5,* 2683–2688. doi:10.1016/0960-894x(95)00451-x
5. Molina, J.-M.; Tourneur, M.; Sarfati, C.; Chevret, S.; de Gouvello, A.; Gobert, J.-G.; Balkan, S.; Derouin, F. *N. Engl. J. Med.* **2002,** *346,* 1963–1969. doi:10.1056/nejmoa012924
6. Shaw, L. M.; Sollinger, H. W.; Halloran, P.; Morris, R. E.; Yatscoff, R. W.; Ransom, J.; Tsina, I.; Keown, P.; Holt, D. W.; Lieberman, R.; Jaklitsch, A.; Potter, J. *Ther. Drug Monit.* **1995,** *17,* 690–699. doi:10.1097/00007691-199512000-00025
7. Matsuda, Y.; Abe, I. *Nat. Prod. Rep.* **2016,** *33,* 26–53. doi:10.1039/c5np00090d
8. Matsuda, Y.; Abe, I. Fungal meroterpenoids. In *Comprehensive Natural Products III;* Liu, H.-W.; Begley, T. P., Eds.; Elsevier: Oxford, UK, 2020; pp 445–478. doi:10.1016/b978-0-12-409547-2.14663-3
9. Itoh, T.; Tokunaga, K.; Radhakrishnan, E. K.; Fujii, I.; Abe, I.; Ebizuka, Y.; Kushiro, T. *ChemBioChem* **2012,** *13,* 1132–1135. doi:10.1002/cbic.201200124
10. Matsuda, Y.; Wakimoto, T.; Mori, T.; Awakawa, T.; Abe, I. *J. Am. Chem. Soc.* **2014,** *136,* 15326–15336. doi:10.1021/ja508127q
11. Li, C.; Matsuda, Y.; Gao, H.; Hu, D.; Yao, X. S.; Abe, I. *ChemBioChem* **2016,** *17,* 904–907. doi:10.1002/cbic.201600087
12. Araki, Y.; Awakawa, T.; Matsuzaki, M.; Cho, R.; Matsuda, Y.; Hoshino, S.; Shinohara, Y.; Yamamoto, M.; Kido, Y.; Inaoka, D. K.; Nagamune, K.; Ito, K.; Abe, I.; Kita, K. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116,* 8269–8274. doi:10.1073/pnas.1819254116
13. Zhang, W.; Du, L.; Qu, Z.; Zhang, X.; Li, F.; Li, Z.; Qi, F.; Wang, X.; Jiang, Y.; Men, P.; Sun, J.; Cao, S.; Geng, C.; Qi, F.; Wan, X.; Liu, C.; Li, S. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116,* 13305–13310. doi:10.1073/pnas.1821932116
14. Awakawa, T.; Liu, W.; Bai, T.; Taniguchi, T.; Abe, I. *Philos. Trans. R. Soc., B* **2023,** *378,* 20220037. doi:10.1098/rstb.2022.0037
15. Yan, D.; Matsuda, Y. *Org. Lett.* **2021,** *23,* 3211–3215. doi:10.1021/acs.orglett.1c00951
16. Barra, L.; Abe, I. *Nat. Prod. Rep.* **2021,** *38,* 566–585. doi:10.1039/d0np00056f
17. Matsuda, Y.; Awakawa, T.; Itoh, T.; Wakimoto, T.; Kushiro, T.; Fujii, I.; Ebizuka, Y.; Abe, I. *ChemBioChem* **2012,** *13,* 1738–1741. doi:10.1002/cbic.201200369
18. Matsuda, Y.; Awakawa, T.; Abe, I. *Tetrahedron* **2013,** *69,* 8199–8204. doi:10.1016/j.tet.2013.07.029

19. Tang, J.; Matsuda, Y. *Chem. Sci.* **2022,** *13,* 10361–10369. doi:10.1039/d2sc02994d
20. Mitsuhashi, T.; Barra, L.; Powers, Z.; Kojasoy, V.; Cheng, A.; Yang, F.; Taniguchi, Y.; Kikuchi, T.; Fujita, M.; Tantillo, D. J.; Porco, J. A., Jr.; Abe, I. *Angew. Chem., Int. Ed.* **2020,** *59,* 23772–23781. doi:10.1002/anie.202011171
21. Tang, J.; Matsuda, Y. *Angew. Chem., Int. Ed.* **2023,** *62,* e202306046. doi:10.1002/anie.202306046
22. Lo, H.-C.; Entwistle, R.; Guo, C.-J.; Ahuja, M.; Szewczyk, E.; Hung, J.-H.; Chiang, Y.-M.; Oakley, B. R.; Wang, C. C. C. *J. Am. Chem. Soc.* **2012,** *134,* 4709–4720. doi:10.1021/ja209809t
23. Jiang, J.; Li, X.; Mori, T.; Awakawa, T.; Abe, I. *Chem. Pharm. Bull.* **2021,** *69,* 444–446. doi:10.1248/cpb.c21-00123

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.56

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Isolation and structure determination of a new analog of polycavernosides from marine *Okeania* sp. cyanobacterium

Kairi Umeda[1], Naoaki Kurisawa[1], Ghulam Jeelani[2], Tomoyoshi Nozaki[2], Kiyotake Suenaga[*1] and Arihiro Iwasaki[*3]

Full Research Paper

Open Access

Address:
[1]Department of Chemistry, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan, [2]Department of Biomedical Chemistry, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan and [3]Department of Applied Chemistry, Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Email:
Kiyotake Suenaga[*] - suenaga@chem.keio.ac.jp; Arihiro Iwasaki[*] - aiwasaki686@g.chuo-u.ac.jp

* Corresponding author

Keywords:
macrolide glycoside; marine cyanobacterium; marine natural products; polycavernosides; terminal alkyne

*Beilstein J. Org. Chem.* **2024,** *20,* 645–652.
https://doi.org/10.3762/bjoc.20.57

Received: 14 December 2023
Accepted: 08 March 2024
Published: 21 March 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra



© 2024 Umeda et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Polycavernoside E (**1**), a new polycavernoside analog, was isolated from a marine *Okeania* sp. cyanobacterium. The relative configuration was elucidated primarily by analyzing the two dimensional nuclear magnetism resonance (2D NMR) data. The absolute configuration was clarified by comparing the electronic circular dichroism (ECD) data of **1** with those of known analogs. Polycavernoside E (**1**) exhibited moderate antitrypanosomal activity against *Trypanosoma brucei rhodesiense*. Furthermore, the isolation of polycavernoside E (**1**) from marine cyanobacteria provides additional evidence that marine cyanobacteria, and not red algae, are responsible for the biosynthesis of polycavernosides.

## Introduction

In 1991, an outbreak of food poisoning caused by a species of red algae known as '*Polycavernosa tsudai*' occurred in Guam, which resulted in killing of three people. Two novel macrolide glycosides, polycavernosides A (**2**) and B (**3**), were reported as the causative compounds for the illness [1]. After that, the second fatal food poisoning incidents occurred in the Philippines caused by the ingestion of polycavernoside A (**2**)-contaminated red algae [2]. Subsequently, polycavernoside analogs such as polycavernoside C (**4**) were isolated from red algae [3,4]. In 2015, Navarro et al. isolated polycavernoside D (**5**) from a marine *Okeania* sp. cyanobacterium [5]. They suggested that polycavernosides were produced by marine cyanobacteria based on their high content and structural similarity to other cyanobacterial metabolites. In this study, polycavernoside E (**1**),

a new polycavernoside analog, was isolated from a marine *Okeania* sp. cyanobacterium obtained from Okinawa Prefecture, Japan (Figure 1). This finding provides additional evidence that polycavernosides are secondary metabolites derived from marine *Okeania* sp. cyanobacteria.

## Results and Discussion

The EtOH extract of marine *Okeania* sp. cyanobacterium (340 g, wet weight) collected from Akuna Beach, Okinawa, Japan, was partitioned between EtOAc and $H_2O$. The EtOAc fraction was further partitioned into 90% aqueous MeOH and hexane. The aqueous MeOH portion was purified by reversed-phase column chromatography (ODS silica gel, MeOH/$H_2O$), automated flash chromatography (hexane/EtOAc), and repeated reversed-phase HPLC to give polycavernoside E (**1**, 0.5 mg as a colorless oil). The isolation of compound **1** was directed by its characteristic UV absorption around 270 nm.

The molecular formula of **1** was determined to be $C_{44}H_{66}O_{15}$ based on the HRESIMS data. The NMR data for **1** are summarized in Table 1. The $^1$H NMR spectrum of compound **1** was similar to those of known polycavernosides but matched none of them, suggesting that **1** was a new analog of polycavernosides [1,3-5]. A detailed analysis of the NMR data revealed the planar structure of **1**, as shown in Figure 2. COSY and HMQC spectral analyses revealed several partial structures, indicated by the bold bonds in Figure 2. Four HMBC were observed from singlet methyl signals: $\delta_H$ 0.85 (H-28)/$\delta_C$ 19.4 (C-29), $\delta_H$ 0.86 (H-29)/$\delta_C$ 17.8 (C-28), $\delta_H$ 0.94 (H-30)/$\delta_C$ 13.9 (C-31), and $\delta_H$ 0.90 (H-31)/$\delta_C$ 22.2 (C-30). These correlations elucidated the presence of two *gem*-dimethyl groups. Moreover, three HMBC, $\delta_H$ 4.03 (H-5a')/$\delta_C$ 106.1 (C-1'), $\delta_H$ 3.61 (H-6')/$\delta_C$ 83.8 (C-2'), and $\delta_H$ 3.45 (H-7')/$\delta_C$ 78.5 (C-4'), revealed the presence of a 2,4-di-*O*-methylpyranose substructure. Furthermore, an HMBC, $\delta_H$ 3.48 (H-6")/$\delta_C$ 78.7 (C-4"), along with typical chemical shifts and coupling constants from C-1" to C-5" obtained in $CD_3OD$ (Table 2), indicated the presence of a 4-*O*-methylpyranose substructure. The HMBC, $\delta_H$ 3.64 (H-3')/$\delta_C$ 103.0 (C-1"), indicated that these two sugar structures were connected through a glycosidic bond.

The geometries of the two olefins at C-16 and C-20 were determined to be *trans* based on the large coupling constants, $^3J_{H\text{-}16/H\text{-}17}$ 15.0 Hz and $^3J_{H\text{-}20/H\text{-}21}$ 15.0 Hz, respectively. The geometry of the remaining double bond at C-18 was established to be *trans* by comparing the $^{13}$C NMR chemical shifts at C-16 and C-21 between **1** and polycavernoside D (**5**) (Table S1 in Supporting Information File 1) [5]. In addition, a $^4J$ long-range coupling between $\delta_H$ 1.95 (H-26) and $\delta_H$ 2.18 (H-24) and three HMBC $\delta_H$ 1.62 (H-23)/$\delta_C$ 84.6 (C-25), $\delta_H$ 2.18 (H-24)/$\delta_C$ 84.6 (C-25), and $\delta_H$ 2.18 (H-24)/$\delta_C$ 68.6 (C-26) revealed a terminal alkyne structure. Additionally, COSY correlations

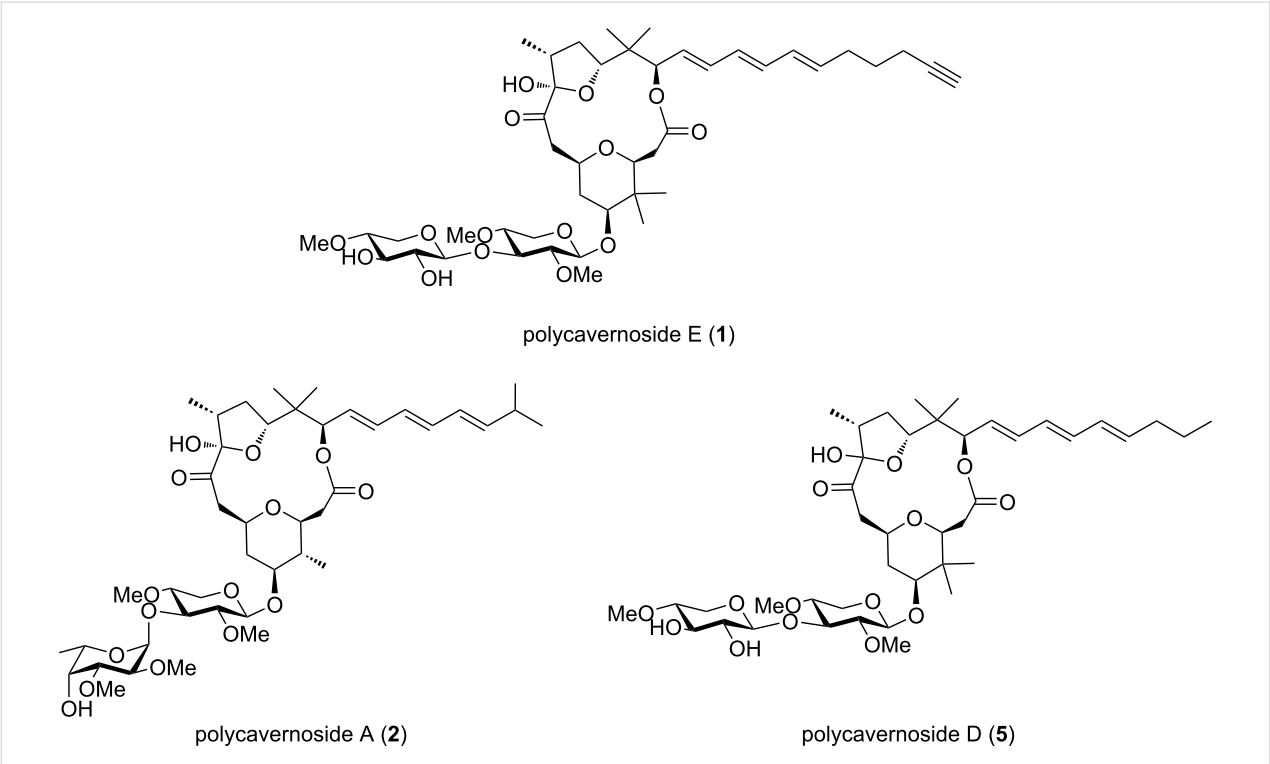



**Figure 1:** Structures of compounds **1**, **2**, and **5**.

**Table 1:** NMR data for polycavernoside E (**1**) in $CDCl_3$.

| position | $\delta_C$, type[a] | $\delta_H$[b] (*J* in Hz) | COSY | selected HMBC |
|---|---|---|---|---|
| 1 | 171.9, C | | | |
| 2 | 35.6, $CH_2$ | 2.29, m | 3 | 1 |
| 3 | 82.0, CH | 3.43, m | 2 | |
| 4 | 38.3, C | | | |
| 5 | 85.3, CH | 3.32, m | 6a, 6b | |
| 6a | 37.7, $CH_2$ | 1.95, m | 5, 6b, 7 | |
| 6b | | 1.61, m | 5, 6a, 7 | |
| 7 | 83.8, CH | 3.07, m | 6a, 6b, 8a, 8b | |
| 8a | 42.1, $CH_2$ | 3.08, m | 7, 8b | 9 |
| 8b | | 2.00, m | 7, 8a | |
| 9 | 206.9, C | | | |
| 10 | 103.0, C | | | |
| 11 | 39.7, CH | 2.74, m | 12a, 12b, 27 | |
| 12a | 33.6, $CH_2$ | 2.01, m | 11, 12b, 13 | |
| 12b | | 1.70, m | 11, 12a | |
| 13 | 83.5, CH | 4.18, dd (11.3, 5.0) | 12a, 12b | |
| 14 | 39.8, C | | | |
| 15 | 78.4, CH | 5.17, d (8.2) | 16 | 1 |
| 16 | 127.4, CH | 5.55, dd (8.2, 15.0) | 15, 17 | |
| 17 | 135.4, CH | 6.26, m | 16, 18 | |
| 18 | 130.1, CH | 6.09, m | 17, 19 | |
| 19 | 133.9, CH | 6.13, m | 18, 20 | |
| 20 | 131.2, CH | 6.08, m | 19, 21 | |
| 21 | 134.6, CH | 5.67, dt (15.0, 7.3) | 20, 22 | |
| 22 | 31.8, $CH_2$ | 2.19, m | 21, 23 | |
| 23 | 28.1, $CH_2$ | 1.62, m | 22, 24 | 25 |
| 24 | 17.9, $CH_2$ | 2.18, m | 23, 26 | 25, 26 |
| 25 | 84.6, C | | | |
| 26 | 68.6, CH | 1.95, t (2.7) | 24 | |
| 27 | 13.3, $CH_3$ | 0.99, d (6.8) | 11 | 10 |
| 28 | 17.8, $CH_3$ | 0.85, s | | 13, 14, 15, 29 |
| 29 | 19.4, $CH_3$ | 0.86, s | | 13, 14, 15, 28 |
| 30 | 22.2, $CH_3$ | 0.94, s | | 3, 4, 5, 31 |
| 31 | 13.9, $CH_3$ | 0.90, s | | 3, 4, 5, 30 |
| 32 | OH | 4.47, s | | 9, 10, 11 |
| 1’ | 106.1, CH | 4.27, d (7.7) | 2’ | 5 |
| 2’ | 83.8, CH | 3.07, m | 1’, 3’ | |
| 3’ | 79.9, CH | 3.64, m | 2’, 4’ | 1” |
| 4’ | 78.5, CH | 3.27, m | 3’, 5a’, 5b’ | |
| 5a’ | 63.2, $CH_2$ | 4.03, dd (11.3, 5.0) | 4’, 5b’ | 1’ |
| 5b’ | | 3.12, m | 4’,5a’ | |
| 6’ | 61.1, $CH_3$ | 3.61, s | | 2’ |
| 7’ | 58.8, $CH_3$ | 3.45, s | | 4’ |
| 1” | 103.0, CH | 4.87, d (4.5) | 2” | |
| 2” | 71.7, CH | 3.53, m | 1”, 3” | |
| 3” | 71.0, CH | 3.75, m | 2”, 4” | |
| 4” | 78.7, CH | 3.34, m | 3”, 5a”, 5b” | |
| 5a” | 60.1, $CH_2$ | 4.23, dd (12.2, 3.2) | 5b”, 4” | |
| 5b” | | 3.46, m | 5a”, 4” | |
| 6” | 58.1, $CH_3$ | 3.48, s | | 4” |

**Table 1:** NMR data for polycavernoside E (**1**) in $CDCl_3$. (continued)

| | | |
|---|---|---|
| 7” | - | OH |
| 8” | - | OH |

[a]Measured at 400 MHz. [b]Measured at 100 MHz.

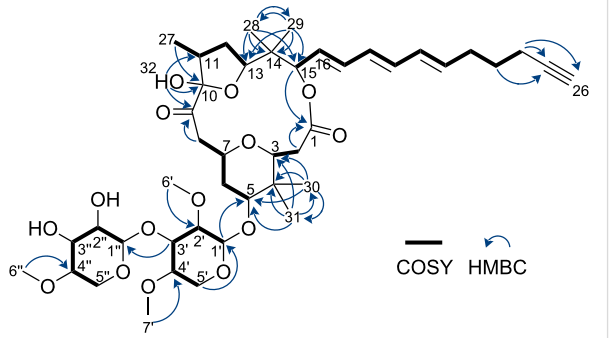



**Figure 2:** Planar structure of polycavernoside E (**1**) based on 2D NMR analysis.

shown in Figure 2 revealed the side chain structure of **1** containing a terminal alkyne and a conjugated *trans* triene (C-15 to C-26).

We then focused on the macrolide structure of **1**. Six HMBC, $\delta_H$ 0.94 (H-30)/$\delta_C$ 82.0 (C-3), $\delta_H$ 0.94 (H-30)/$\delta_C$ 38.3 (C-4), $\delta_H$ 0.94 (H-30)/$\delta_C$ 85.3 (C-5), $\delta_H$ 0.90 (H-31)/$\delta_C$ 82.0 (C-3), $\delta_H$ 0.90 (H-31)/$\delta_C$ 38.3 (C-4), and $\delta_H$ 0.90 (H-31)/$\delta_C$ 85.3 (C-5), along with COSY correlations shown in Figure 2, revealed a chain structure from C-2 to C-8. In addition, eight HMBC, $\delta_H$ 5.17 (H-15)/$\delta_C$ 171.9 (C-1), $\delta_H$ 2.29 (H-2)/$\delta_C$ 171.9 (C-1), $\delta_H$ 0.85 (H-28)/$\delta_C$ 83.5 (C-13), $\delta_H$ 0.85 (H-28)/$\delta_C$ 39.8 (C-14), $\delta_H$ 0.85 (H-28)/$\delta_C$ 78.4 (C-15), $\delta_H$ 0.86 (H-29)/$\delta_C$ 83.5 (C-13), $\delta_H$ 0.86 (H-29)/$\delta_C$ 39.8 (C-14), and $\delta_H$ 0.86 (H-29)/$\delta_C$ 78.4 (C-15), and COSY correlations shown in Figure 2, clarified the connection of C-1 to C-8 and C-15 to C-11(-C27) through an ester bond. Furthermore, five HMBC, $\delta_H$ 0.99 (H-27)/$\delta_C$ 103.0 (C-10), $\delta_H$ 4.47 (H-32)/$\delta_C$ 206.9 (C-9), $\delta_H$ 4.47 (H-32)/$\delta_C$ 103.0 (C-10), $\delta_H$ 4.47 (H-32)/$\delta_C$ 39.7 (C-11), and $\delta_H$ 3.08 (H-8a)/$\delta_C$ 206.9 (C-9) connected C-11 and C-8 through a ketone carbonyl carbon at C-9 and hemiacetal carbon at C-10, revealing the 16-membered macrolide structure of **1**. The HMBC, $\delta_H$ 4.27 (H-1’)/$\delta_C$ 85.3 (C-5), revealed that the disaccharide moiety was connected to C-5. Finally, considering the molecular formula of **1** and the chemical shifts of known polycavernosides, we established the presence of a THP ring containing C-3 to C-7 and a THF ring containing C-10 to C-13 in the macrolide structure. Consequently, we established the planar structure of **1**, as shown in Figure 2.

The relative configuration of compound **1** was determined based on the NMR data obtained in $CD_3OD$ and $CDCl_3$ (Table 1 and Table 2). The relative configuration of the THP ring and the disaccharide moiety of **1** was determined by analyzing the proton coupling constants and NOESY correlations (Figure 3). The two coupling constants in $CD_3OD$, $^3J_{H\text{-}5/H\text{-}6b}$ (11.9 Hz) and $^3J_{H\text{-}6b/H\text{-}7}$ (11.9 Hz), indicated that H-5, H-6b, and H-7 were in the axial position. The two NOESY correlations in $CD_3OD$, $\delta_H$ 1.80 (H-6b)/$\delta_H$ 0.91 (H-31) and $\delta_H$ 2.36 (H-2)/$\delta_H$ 0.91 (H-31), revealed that H-6b, C-31, and

**Table 2:** NMR data for polycavernoside E (**1**) in $CD_3OD$.

| position | $\delta_C$, type[a] | $\delta_H$[b] (*J* in Hz) | COSY | selected HMBC |
|---|---|---|---|---|
| 1 | 174.3, C | | | |
| 2 | 36.5, $CH_2$ | 2.36, d (7.7) | 3 | 1 |
| 3 | 83.2, CH | 3.40, m | 2 | |
| 4 | 39.4, C | | | |
| 5 | 86.1, CH | 3.40, m | 6a, 6b | |
| 6a | 38.0, $CH_2$ | 1.91, m | 5, 6b, 7 | |
| 6b | | 1.80, ddd (11.9, 11.9, 11.9) | 5, 6a, 7 | |
| 7 | 76.6, CH | 3.65, m | 6a, 6b, 8a, 8b | |
| 8a | 42.3, $CH_2$ | 2.85, m | 7, 8b | 9 |
| 8b | | 2.37, m | 7, 8a | |
| 9 | 207.4, C | | | |
| 10 | 104.8, C | | | |
| 11 | 39.7, CH | 2.82, m | 12a, 12b, 27 | |

**Table 2:** NMR data for polycavernoside E (**1**) in $CD_3OD$. (continued)

| | | | | |
|---|---|---|---|---|
| 12a | 34.6, $CH_2$ | 1.98, m | 11, 12b, 13 | |
| 12b | | 1.62, m | 11, 12a | |
| 13 | 83.8, CH | 4.12, dd (11.6, 4.7) | 12a, 12b | |
| 14 | 40.6, C | | | |
| 15 | 80.3, CH | 5.10, d (8.1) | 16 | 1 |
| 16 | 128.5, CH | 5.61, dd (8.1, 15.1) | 15, 17 | 18 |
| 17 | 136.4, CH | 6.21, m | 16, 18 | |
| 18 | 131.1, CH | 6.13, m | 17 | |
| 19 | 135.0, CH | 6.19, m | | |
| 20 | 132.4, CH | 6.12, m | 21 | |
| 21 | 135.6, CH | 5.72, dt (15.3, 7.1) | 20, 22 | 19 |
| 22 | 32.7, $CH_2$ | 2.22, m | 21, 23 | |
| 23 | 29.3, $CH_2$ | 1.60, quint (7.2) | 22, 24 | 25 |
| 24 | 18.5, $CH_2$ | 2.17, m | 23 | 25, 26 |
| 25 | 84.7, C | | | |
| 26 | 69.8, CH | 2.21, m | | |
| 27 | 13.6, $CH_3$ | 0.98, d (6.8) | 11 | 10 |
| 28 | 18.0, $CH_3$ | 0.868, s | | 13, 14, 15, 29 |
| 29 | 19.5, $CH_3$ | 0.870, s | | 13, 14, 15, 28 |
| 30 | 22.3, $CH_3$ | 0.98, s | | 3, 4, 5, 31 |
| 31 | 14.1, $CH_3$ | 0.91, s | | 3, 4, 5, 30 |
| 1’ | 107.1, CH | 4.33, d (7.8) | 2’ | 5 |
| 2’ | 85.5, CH | 3.08, dd (7.8, 9.0) | 1’, 3’ | |
| 3’ | 81.9, CH | 3.63, m | 2’, 4’ | 1” |
| 4’ | 79.4, CH | 3.27, m | 3’, 5a’, 5b’ | |
| 5a’ | 64.3, $CH_2$ | 3.96, dd (11.5, 5.1) | 4’, 5b’ | 1’ |
| 5b’ | | 3.16, m | 4’, 5a’ | |
| 6’ | 61.3, $CH_3$ | 3.61, s | | 2’ |
| 7’ | 59.5, $CH_3$ | 3.46, s | | 4’ |
| 1” | 105.0, CH | 4.62, d (7.3) | 2” | |
| 2” | 75.3, CH | 3.21, dd (7.3, 9.1) | 1”, 3” | |
| 3” | 76.5, CH | 3.39, m | 2”, 4” | |
| 4” | 81.0, CH | 3.18, m | 3”, 5a”, 5b” | |
| 5a” | 64.1, $CH_2$ | 4.07, dd (11.0, 4.4) | 5b”, 4” | |
| 5b” | | 3.13, m | 5a”, 4” | |
| 6” | 58.9, $CH_3$ | 3.47, s | | 4” |

[a]Measured at 400 MHz. [b]Measured at 100 MHz.

C-2 were located in the same face of the THP ring as shown in Figure 3. Consequently, the relative configuration of the THP ring was determined to be 3*S**,5*S**,7*S**. For the 2,4-di-*O*-methylpyranose moiety, the two large coupling constants in $CD_3OD$, $^3J_{H\text{-}1'/H\text{-}2'}$ (7.8 Hz) and $^3J_{H\text{-}2'/H\text{-}3'}$ (9.0 Hz), indicated that H-1’, H-2’, and H-3’ were in the axial position. The two NOESY correlations in $CDCl_3$, $\delta_H$ 3.45 (H-7’)/$\delta_H$ 4.87 (H-1”) and $\delta_H$ 3.45 (H-7’)/$\delta_H$ 4.23 (H-5a”), revealed that the methoxy group at C-4’ was in the equatorial position and H-4’ was in the axial position. The 2,4-di-*O*-methylpyranose moiety was identified as 2,4-di-*O*-methylxylose. For the 4-*O*-methylpyranose moiety, the two large coupling constants in $CD_3OD$, $^3J_{H\text{-}1''/H\text{-}2''}$ (7.3 Hz) and $^3J_{H\text{-}2''/H\text{-}3''}$ (9.1 Hz), indicated that H-1”, H-2”, and H-3” were in the axial position. NOESY correlations in $CDCl_3$, $\delta_H$ 3.75 (H-3”)/$\delta_H$ 3.48 (H-6”), revealed that the methoxy group at C-4” was in the equatorial position and H-4” was in the axial position. The 4-*O*-methylpyranose moiety was identified as 4-*O*-methylxylose. The relationship between the relative configuration of the 2,4-di-*O*-methylxylose moiety and 4-*O*-methylxylose moiety was identified using three NOESY correlations in $CDCl_3$, $\delta_H$ 3.45 (H-7’)/$\delta_H$ 4.87 (H-1”), $\delta_H$ 3.45 (H-7’)/$\delta_H$ 4.23 (H-5a”), and $\delta_H$ 3.64 (H-3’)/$\delta_H$ 4.87 (H-1”), as shown in Figure 3. Furthermore, the relationship of the relative configuration between the disaccharide moiety and the THP ring was

revealed by two NOESY correlations in $CDCl_3$, $\delta_H$ 4.27 (H-1')/$\delta_H$ 3.32 (H-5) and $\delta_H$ 4.27 (H-1')/$\delta_H$ 0.94 (H-30), shown in Figure 3. The validity of the relative configurations shown in Figure 3 is further substantiated by the good agreement with the corresponding chemical shifts of polycavernoside D (**5**), both of which possess the same disaccharide moiety attached to a THP ring (Tables S1 and S2 in Supporting Information File 1) [5].

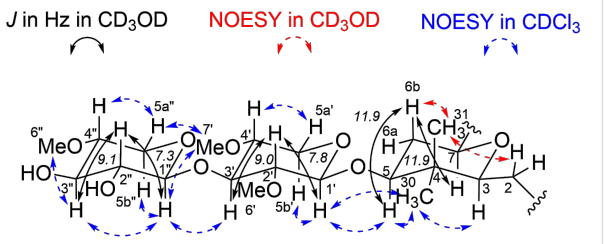



**Figure 3:** Relative configuration of the THP ring and disaccharide moiety of **1**.

The remaining relative configuration of **1** was determined by a comparison of the carbon chemical shifts between **1** and **5** in $CDCl_3$ [5]. As shown in Table S1 (Supporting Information File 1), the two sets of data were in good agreement, indicating that the relative configurations of compounds **1** and **5** were identical.

To reveal the absolute configuration, we recorded the ECD spectrum of **1** (Figure 4) and compared it with those of **2** and **5** reported in previous papers [5,6]. We detected a Cotton effect of negative sign at around 280 nm corresponding to the n–π* transition of a ketone group as same as the literature data for **2** and **5**. As a result, the absolute configuration of polycavernoside E was determined to be **1**.

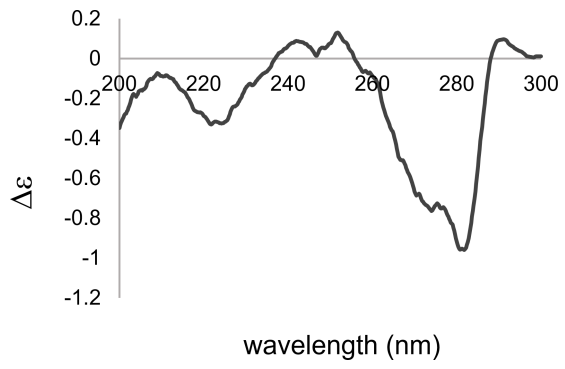



**Figure 4:** ECD spectrum of **1** in MeOH.

Next, we examined the antitrypanosomal activity of **1** against the bloodstream form of *Trypanosoma brucei rhodesiense* IL-1501 (the causative organism of African trypanosomiasis) (Table 3). As a result, **1** showed moderate growth-inhibitory activities against *Trypanosoma brucei rhodesiense* ($IC_{50}$: 9.9 μM). In addition, we examined the growth-inhibitory activity of **1** on normal human fibroblasts WI-38 (Table 3). In summary, **1** showed a moderate activity against *Trypanosoma brucei rhodesiense*.

**Table 3:** Growth-inhibitory activities of polycavernoside E (**1**).

| Compounds | $IC_{50}$ values (μM) | |
|---|---|---|
| | *T. b. rhodesiense* | WI-38 cells |
| polycavernoside E (**1**) | 9.9 ± 1.5 | 27 |
| pentamidine[a] | 0.006 ± 0.002 | – |

[a]Positive control.

## Conclusion

In conclusion, we isolated a new polycavernoside analog, named polycavernoside E (**1**), from a marine *Okeania* sp. cyanobacterium. The relative configuration was elucidated mainly by analyzing the 2D NMR data. The absolute configuration was determined based on a comparison of the ECD data for **1** and its known analogs. Polycavernoside E (**1**) showed selective antitrypanosomal activity against *Trypanosoma brucei rhodesiense* with an $IC_{50}$ value of 9.9 μM. This discovery provides additional evidence that polycavernosides, previously thought to be derived from red algae, are produced by the marine *Okeania* sp. cyanobacterium. In the field, this type of cyanobacterium that produces the analog of human lethal toxin is often observed with macroalgae and shells, and, therefore, it can be a potential risk for food poisoning of fishery resources.

## Experimental

### General experimental procedures

Optical rotations were measured with a JASCO DIP-1000 polarimeter. UV spectra were recorded on a UV-3600. ECD spectra were measured with JASCO J-1100. IR spectra were recorded on a Bruker ALPHA instrument. All NMR data were recorded on a JEOL ECX-400/ECS-400 spectrometer for $^1H$ (400 MHz) and $^{13}C$ (100 MHz). $^1H$ NMR chemical shifts (referenced to the residual solvent signal of $CHD_2OD$: δ 3.31, $CHCl_3$: δ 7.26) were assigned using a combination of data from COSY and HMQC experiments. Similarly, $^{13}C$ NMR chemical shifts (referenced to the solvent signal $CD_3OD$: δ 49.0, $CDCl_3$: δ 77.16) were assigned based on HMBC and HMQC experiments. HRESIMS spectra were obtained on a Bruker timsTOF mass spectrometer. For reversed-phase column chromatography, ODS silica gel Cosmosil 75$C_{18}$-OPN (Nacalai Tesque) was used. For medium pressure column chromatography, AFCS (Smart Flash AKROS, Yamazen) consisting of a pump and a

UV detector was used. HPLC analysis was conducted using a pump (model PU-2080, Jasco) and a UV detector (model UV-2075, Jasco). All chemicals and solvents used in this study were the best grade available and obtained from a commercial source (Nacalai Tesque).

## Collection and identification of the sample

The marine cyanobacterium producing polycavernoside E (**1**) was collected in March 2022 at the coast in Akuna beach, Yonashiromiyagi, Uruma city, Okinawa, Japan. It was classified into *Okeania* sp. based on the phylogenetic analysis as described in the previous paper (accession no. LC771053) [7].

## Isolation of polycavernoside E (**1**)

In a manner analogous to [7], the collected cyanobacterium (340 g) was extracted with EtOH (0.5 L) for 10 days at room temperature (rt). The extract was filtered, and the residue was homogenized with a blender and re-extracted with EtOH (0.5 L) at room temperature for one day. The extract was filtered, and the combined filtrates were concentrated. The residue was partitioned between EtOAc (3 × 300 mL) and $H_2O$ (300 mL). The combined organic layers were concentrated, and the residue was partitioned between 90% aqueous MeOH (300 mL) and hexane (3 × 300 mL). The aqueous MeOH layer was concentrated, and the obtained residue (673 mg) was separated by column chromatography on ODS (7 g) eluted with 35 mL of 40%, 60%, 80%, and 90% aqueous MeOH, followed by 35 mL of MeOH and 70 mL of $CHCl_3$/MeOH 1:1. The fraction (244.4 mg) eluted with 80% MeOH was subjected to AFCS [Ø 11 × 300 mm; flow rate 5 mL/min; detection at 254 nm; solvent gradient condition, hexane/EtOAc 28:72 → 7:93] to give a fraction that contained compound **1** (17.5 mg, $t_R$ = 32.0 min). The fraction that contained **1** was further purified by HPLC [Cosmosil 5$C_{18}$-MS-II (Ø 20 mm × 250 mm); solvent MeOH/$H_2O$ 85:15; flow rate 5 mL/min; detection UV 215 nm] to give a fraction that contained **1** (8.6 mg, last collected fraction). The fraction that contained **1** was further separated by HPLC [Cosmosil Cholester (Ø 20 mm × 250 mm); solvent MeCN/$H_2O$ 75:25; flow rate 5 mL/min; detection UV 254 nm] to give a fraction that contained **1** (0.6 mg, $t_R$ = 32.8 min). The fraction that contained **1** was further separated by HPLC [Cosmosil 5PE-MS (Ø 20 mm × 250 mm); solvent MeOH/$H_2O$ 85:15; flow rate 5 mL/min; detection UV 270 nm] to give **1** (0.5 mg, $t_R$ = 37.3 min).

Polycavernoside E (**1**): colorless oil; $[\alpha]_D^{26}$ −19 (*c* 0.04, MeOH); UV (MeOH) $\lambda_{max}$, nm (log ε): 281 (2.27), 270 (2.96), 260 (2.40); ECD (100 µg/mL; MeOH), $\lambda_{max}$, nm (Δε): 226 (−0.31), 274 (−0.76), 282 (−0.96); IR (neat): 3443, 2965, 2925, 2896, 1646, 1457, 1086 $cm^{-1}$; HRESIMS (*m*/*z*): $[M + Na]^+$ calcd for $C_{44}H_{66}O_{15}Na^+$, 857.4294; found, 857.4294.

## In vitro antitrypansomal assay

The bloodstream form of *Trypanosoma brucei rhodesiense* strain IL-1501 was cultured at 37 °C under a humidified 5% $CO_2$ atmosphere in HMI-9 medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) [8,9]. For in vitro studies, compounds were dissolved in DMSO and diluted in culture medium prior to being assayed. The maximum DMSO concentration in the in vitro assays was 1%. The compounds were tested in an AlamarBlue serial drug dilution assay to determine the 50% inhibitory concentrations ($IC_{50}$) [10]. Serial drug dilutions were prepared in 96-well microtiter plates, containing 50 µL of culture medium. Subsequently, 50 µL of a parasite suspension with a concentration of $4.0 \times 10^4$ cells/mL was introduced into each well. Cultures were incubated for 69 h at 37 °C under a humidified 5% $CO_2$ atmosphere. After this time, 10 µL of resazurin (12.5 mg resazurin (Sigma) dissolved in 100 mL phosphate-buffered saline) was added to each well. The plates were incubated for an additional 3 h. The plates were read in a SpectraMax Gemini XS microplate fluorescence scanner (Molecular Devices) using an excitation wavelength of 536 nm and an emission wavelength of 588 nm.

## WI-38 cells assay

In a manner analogous to [7], WI-38 cells were cultured at 37 °C with 5% $CO_2$ in DMEM (Nissui) supplemented with 10% heat-inactivated FBS, 100 units/mL penicillin, 100 µg/mL streptomycin, 0.25 µg/mL amphotericin, 300 µg/mL L-glutamine, and 2.25 mg/mL $NaHCO_3$. Cells were seeded at $4 \times 10^3$ cells/well in 96-well plates (Iwaki) and cultured overnight. Various concentrations of compounds were then added, and cells were incubated for 72 h. Cell proliferation was measured by the MTT assay.

# Supporting Information

### Supporting Information File 1

NMR data for polycavernoside E (**1**).
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-57-S1.pdf]

# Funding

This work was supported by JSPS KAKENHI (Grant Numbers 21K14747 and 20H02870) and The Naito Foundation.

# ORCID® iDs

Naoaki Kurisawa - https://orcid.org/0000-0002-3204-7693
Arihiro Iwasaki - https://orcid.org/0000-0002-3775-5066

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Yotsu-Yamashita, M.; Haddock, R. L.; Yasumoto, T. *J. Am. Chem. Soc.* **1993,** *115,* 1147–1148. doi:10.1021/ja00056a048
2. Yotsu-Yamashita, M.; Yasumoto, T.; Yamada, S.; Bajarias, F. F. A.; Formeloza, M. A.; Romero, M. L.; Fukuyo, Y. *Chem. Res. Toxicol.* **2004,** *17,* 1265–1271. doi:10.1021/tx0498556
3. Yotsu-Yamashita, M.; Seki, T.; Paul, V. J.; Naoki, H.; Yasumoto, T. *Tetrahedron Lett.* **1995,** *36,* 5563–5566. doi:10.1016/0040-4039(95)01052-j
4. Yotsu-Yamashita, M.; Abe, K.; Seki, T.; Fujiwara, K.; Yasumoto, T. *Tetrahedron Lett.* **2007,** *48,* 2255–2259. doi:10.1016/j.tetlet.2007.02.003
5. Navarro, G.; Cummings, M. E.; Lee, J.; Moss, N.; Glukhov, E.; Valeriote, F. A.; Gerwick, L.; Gerwick, W. H. *Environ. Sci. Technol. Lett.* **2015,** *2,* 166–170. doi:10.1021/acs.estlett.5b00116
6. Blakemore, P. R.; Browder, C. C.; Hong, J.; Lincoln, C. M.; Nagornyy, P. A.; Robarge, L. A.; Wardrop, D. J.; White, J. D. *J. Org. Chem.* **2005,** *70,* 5449–5460. doi:10.1021/jo0503862
7. Umeda, K.; Iwasaki, A.; Taguchi, R.; Kurisawa, N.; Jeelani, G.; Nozaki, T.; Suenaga, K. *J. Nat. Prod.* **2023,** *86,* 2529–2538. doi:10.1021/acs.jnatprod.3c00742
8. Kuboki, N.; Inoue, N.; Sakurai, T.; Di Cello, F.; Grab, D. J.; Suzuki, H.; Sugimoto, C.; Igarashi, I. *J. Clin. Microbiol.* **2003,** *41,* 5517–5524. doi:10.1128/jcm.41.12.5517-5524.2003
9. Räz, B.; Iten, M.; Grether-Bühler, Y.; Kaminsky, R.; Brun, R. *Acta Trop.* **1997,** *68,* 139–147. doi:10.1016/s0001-706x(97)00079-x
10. Huber, W.; Koella, J. C. *Acta Trop.* **1993,** *55,* 257–261. doi:10.1016/0001-706x(93)90083-n

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.57

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# New variochelins from soil-isolated *Variovorax* sp. H002

Jabal Rahmat Haedar, Aya Yoshimura and Toshiyuki Wakimoto*§

Full Research Paper

Open Access

Address:
Faculty of Pharmaceutical Sciences, Hokkaido University, Kita 12, Nishi 6, Sapporo 060-0812, Japan

Email:
Toshiyuki Wakimoto* - wakimoto@pharm.hokudai.ac.jp

* Corresponding author
§ Phone: +81-11-706-3239

Keywords:
antimicrobial activity; siderophore; variochelin; *Variovorax*

*Beilstein J. Org. Chem.* **2024,** *20,* 692–700.
https://doi.org/10.3762/bjoc.20.63

Received: 05 January 2024
Accepted: 08 March 2024
Published: 02 April 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra



© 2024 Haedar et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

The soil bacterial genus *Variovorax* produce distinct photoreactive siderophores that may play a crucial role in the iron cycle within the rhizosphere. This study focused on exploring the natural products of the soil-isolated *Variovorax* sp. H002, leading to the isolation of variochelins A–E (**1**–**5**), a series of lipohexapeptide siderophores. NMR and MS/MS analyses revealed that these siderophores share a common core structure – a linear hexapeptide with β-hydroxyaspartate and hydroxamate functional groups, serving in iron-binding coordination. Three new variochelins C–E (**3**–**5**) were characterized by varied fatty acyl groups at their *N*-termini; notably, **4** and **5** represent the first variochelins with *N*-terminal unsaturated fatty acyl groups. Furthermore, the variochelin biosynthetic gene cluster was identified through draft genome sequencing and gene knockout experiments. Compounds **1**–**5** exhibited antimicrobial activities against Gram-negative bacteria, including several soil-isolated plant pathogens.

## Introduction

Almost all organisms require iron as a crucial cofactor of enzymes essential for their physiological functions, encompassing primary and secondary metabolisms [1]. However, in an aerobic environment, iron predominantly exists in the form of insoluble ferric oxide, diminishing its bioavailability. To mitigate this problem, many microorganisms utilize specialized metabolites known as siderophores to efficiently sequester iron. Siderophores are abundantly produced under iron-deficient growth conditions and secreted through dedicated transporters to form the siderophore Fe(III) complex extracellularly [2,3]. Subsequently, the stable and water soluble siderophore Fe(III) complex is transported across the cell membrane into the cytoplasm, where the iron is released from the complex by reduction, generating free Fe(II) ions that are accessible for the cell [2,3].

In addition to the siderophores forming the stable Fe(III) complexes described above, certain siderophores create Fe(III) complexes with the ability to release Fe(II) ions in a light-responsive manner within the extracellular environment. The released Fe(II) ions in the ocean serve as a nutrient source for phytoplankton and other organisms [4]. The distinctive structural feature of these siderophores is the α-hydroxy carboxylic acid (i.e., β-hydroxyaspartate), relevant to a photoreactive functional group [2,4]. In 2016, using the gene encoding a hydroxylase specific to the aspartate residue as a query, Nett et al. per-

formed genome mining of photoreactive siderophores and isolated two such siderophores, variochelins A and B, from the soil bacterium *Variovorax boronicumulans* [5].

The *Variovorax* spp. belonging to plant growth-promoting rhizobacteria (PGPR) enhance plant development across several conditions, such as increased levels of auxin plant hormone (IAA) [6] and reduced ethylene levels [7]. They also play a crucial ecological role in the degradation of environmentally harmful pollutants [8-10]. Moreover, these bacteria produce unique photoreactive siderophores other than variochelins, including vacidobactin from *V. paradoxus* S110, variobactin from *V. paradoxus* P4B, and imaqobactin from *Variovorax* sp. RKJM285, suggesting their significant role in interactions with both plants and other microbes in near-surface soil through iron photocycling [5,11,12].

In this study, we isolated three new congeners of variochelin-type siderophores, variochelins C–E (**3**–**5**), along with two known compounds, variochelins A (**1**) and B (**2**), from *Variovorax* sp. Their structures were elucidated by a combination of NMR, ESIMS/MS, and chemical derivatization. The analysis of the draft genome sequence of the H002 strain identified the variochelin biosynthetic gene cluster (*var*), which encodes PKS (polyketide synthase) and NRPS (non-ribosomal peptide synthetase) genes. Finally, the siderophores isolated in this study exhibited antibacterial activity against several bacteria, including soil-isolated plant pathogens.

## Results and Discussion

### Isolation and structure elucidation

In this study, we explored the secondary metabolite potential of the soil-isolated *Variovorax* sp. H002, domesticated from the Medicinal Plant Garden of the Faculty of Pharmaceutical Sciences, Hokkaido University. To stimulate siderophore production, H002 was grown in modified minimal medium containing casamino acids and glucose as nitrogen and carbon sources, respectively, in the presence of HP-20 resin. Upon cultivation for one week, the non-polar resin was collected and washed thoroughly with water to remove salt and carbohydrate residues. After the retained siderophores were extracted with methanol, the obtained crude extract was further fractionated through several rounds of column chromatography, guided by a colorimetric-based chrome azurol sulfonate siderophore assay (CAS assay) [13], to yield five compounds (**1**–**5**, Figure 1). The production of these siderophores was high under iron-starvation conditions with 0.1 μM $FeCl_3$ in the media, and began to decrease when the concentration of $Fe^{3+}$

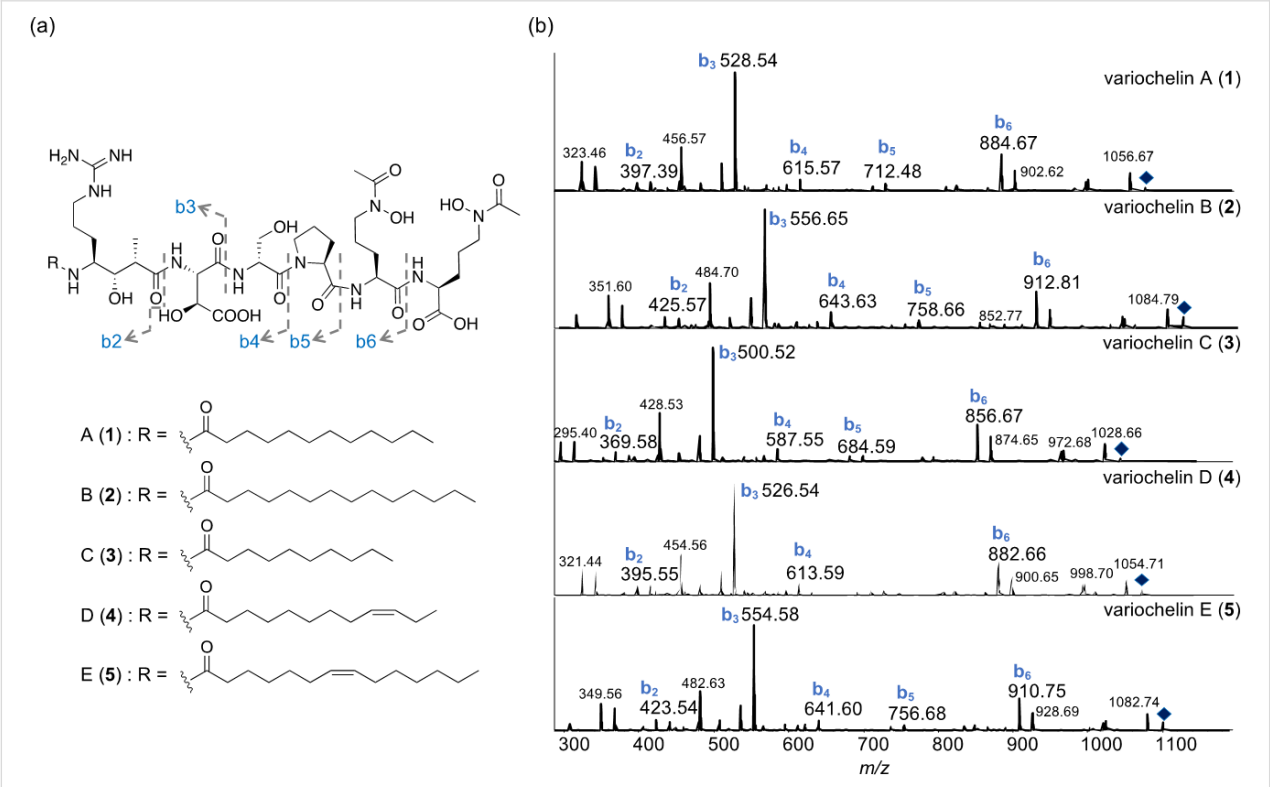



**Figure 1:** (a) Chemical structures and (b) ESIMS/MS of variochelins A–E (**1**–**5**) isolated from *Variovorax* sp. H002. Parent ions are *m/z* 1075.08 $[M + H]^+$ (**1**), *m/z* 1103.18 $[M + H]^+$ (**2**), *m/z* 1047.06 $[M + H]^+$ (**3**), *m/z* 1072.95 $[M + H]^+$ (**4**), and *m/z* 1101.14 $[M + H]^+$ (**5**). Parent ions are represented by diamonds.

ions reached 0.5 μM (Figure S1, Supporting Information File 1).

The HRESIMS data of the major compound **1** showed a molecular weight of *m*/*z* 535.7912 for the $[M - 2H]^{2-}$ ion, inferring a chemical formula of $C_{47}H_{83}O_{17}N_{11}$ (calcd 535.7911 for the double-negative ion). A combination of $^1$H NMR and COSY analyses revealed a peptidic structure comprising two $N^{\delta}$-acetyl-$N^{\delta}$-hydroxyornithine residues, a proline (Pro), a serine (Ser), a β-hydroxyaspartic acid (Hya), and a unique γ-amino acid biosynthetically originating from an arginine and a methyl malonate (4-amino-7-guanidino-3-hydroxy-2-methylheptanoic acid), with the amino group acylated by a saturated fatty acid. Subsequent HMBC and MS/MS analyses indicated that the planar structure of **1** is consistent with the known siderophore variochelin A, originally reported from *V. boronicumulans* BAM-48 by Nett et al. [5].

The NMR spectra of **2** are superimposable on those of **1**. However, the MS/MS of **2** differs in the fragmentation ion corresponding to a 4-acylamino-7-guanidino-3-hydroxy-2-methylheptanoate, indicating that the *N*-terminal fatty acyl group is a tetradecanoic acid in **2** (Figure 1 and Figure S10 in Supporting Information File 1). Therefore, compound **2** is concluded to be the known lipohexapeptide variochelin B [5].

Compound **3** has the chemical formula of $C_{45}H_{79}O_{17}N_{11}$, as suggested by the HRESIMS data (*m*/*z* 1044.5603 for the $[M - H]^-$ ion), indicating the loss of two methylene groups ($C_2H_4$) from **1**. In addition, the MS/MS fragmentation profile of **3** was consistent with those of **1** and **2**, except for the loss of 28 mass units in the *N*-terminal $N^{\gamma}$-acyl-γ-amino acid moiety (Figure 1 and Figure S18 in Supporting Information File 1). Subsequently, to determine the acyl chain, the hydrolysate of **3** was treated with 2-amino-2-methyl-1-propanol to yield a 4,4-dimethyloxazoline (DMOX) derivative of the fatty acid, which was then subjected to GC–MS [14,15]. The EIMS of the derivatized acyl moiety of **3** displayed two typical prominent ions at *m*/*z* 113 and *m*/*z* 126, which are diagnostic for the DMOX derivative [14-16] (Figure 2). The profile, with a molecular ion at *m*/*z* 225 $[M]^+$ and the loss of each successive methylene group by *m*/*z* 196, 182, 168, 154, and 140, allowed us to deduce the fatty acyl moiety in **3** as an decanoic acid. Thus, we determined **3** to be a new siderophore termed variochelin C, with an *N*-terminal decanoate.

Compound **4** has the chemical formula of $C_{47}H_{81}O_{17}N_{11}$, as suggested by HRESIMS (*m*/*z* 534.7837 for $[M - 2H]^{2-}$ ion), inferring the loss of two hydrogen atoms from **1**. The $^1$H NMR spectrum showed additional olefinic proton signals at $\delta_H$ 5.35, as compared with that of **1**. In line with this, two $sp^2$ carbon signals at $\delta_C$ 130.6 and 131.0 were detected in the $^{13}$C NMR spectrum, demonstrating the presence of an olefin in **4**. The comparative analysis of the MS/MS profiles of **1** and **4** (Figure 1 and Figure S26 in Supporting Information File 1) revealed that all detected *b* ions in **4** were two mass units smaller than those in **1**, suggesting that the *N*-terminal acyl moiety is an unsaturated dodecenoic acid. To further confirm

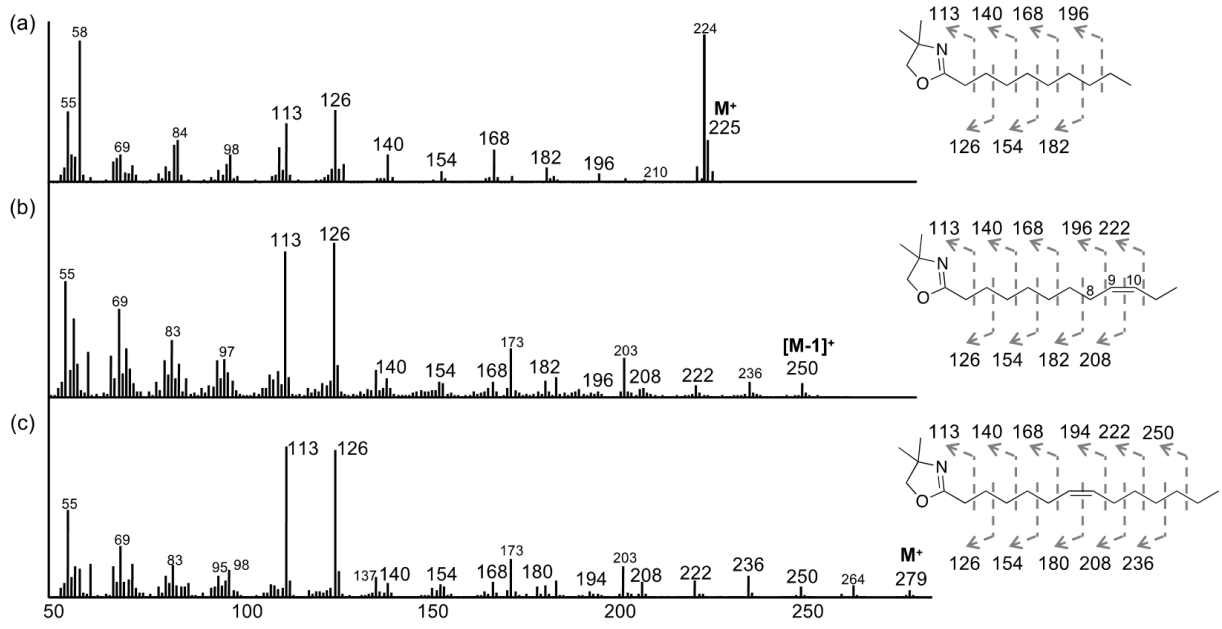



**Figure 2:** EIMS of DMOX derivatized free fatty acids derived from variochelins C–E (**3**–**5**). (a) **3**, (b) **4**, and (c) **5**.

the acyl group, the DMOX derivative prepared from **4** was subjected to GC–MS analysis. The EIMS spectrum showed two typical prominent ions at *m*/*z* 113 and *m*/*z* 126, as well as two ions at *m*/*z* 222 and *m*/*z* 208, corresponding to the loss of methylene groups (14 mass units). More importantly, a gap of 12 mass units between *m*/*z* 196 (C8) and 208 (C9) was clearly detected, which indicated the position of the double bond between carbons C9 and C10 [14,15] and established the *N*-terminal acyl moiety in **4** as 9-dodecenoic acid. Therefore, **4** is a new analog, variochelin D, possessing an unsaturated fatty acyl moiety.

Finally, based on HRESIMS data, the chemical formula of **5** was deduced as $C_{49}H_{85}O_{17}N_{11}$ (1098.6071 $[M - H]^-$, calcd. 1098.6052), indicating the loss of two hydrogen atoms from variochelin B (**2**). Similar to **4**, compound **5** also has a double bond in its fatty acyl moiety, as suggested by the MS/MS analysis (Figure 1 and Figure S34 in Supporting Information File 1). Subsequent GC–MS analysis of the DMOX derivative allowed us to deduce the fatty acyl chain in **5** as a 7-tetradecenoic acid, confirming **5** as another new variochelin analog, variochelin E.

The absolute configurations of the newly isolated congeners (**3**–**5**) were determined by Marfey's analysis [17,18]. The absolute configurations of the amino acid constituents, Pro, Ser and two modified ornithine residues, in **3**–**5** were assigned as L, D, and L, respectively (Figures S25, S33, and S41, Supporting Information File 1). The geometry of the olefins in **4** and **5** was determined to be *Z* based on the $^{13}C$ chemical shifts of allylic carbons [19].

Unfortunately, we were unable to determine the absolute configurations of Hya and three stereocenters in 4-amino-7-guanidino-3-hydroxy-2-methylheptanoic acid. L-*threo*-Hya is reportedly present in **1**, based on the elution order of FDAA ($N^{\alpha}$-(5-fluoro-2,4-dinitrophenyl)-L-alaninamide) derivatives [5]. In addition, a combination of NOESY and bioinformatics analyses proposed all three stereocenters of 4-amino-7-guanidino-3-hydroxy-2-methylheptanoic acid as *S* configured [5]. The stereochemistry of these moieties in **3**–**5** are discussed in the next section, regarding the variochelin biosynthetic gene cluster.

## Biosynthetic pathway of variochelins

Nett et al. reported the variochelin biosynthetic gene cluster (NCBI accession number KT900023) encoded in *V. boronicumulans*, although it has not yet been experimentally validated. In the draft genome sequence of *Variovorax* sp. H002, we also identified a *var* gene cluster containing NRPS and PKS genes: the domain organizations of NRPS and PKS, and the adjacently encoded modification enzymes, were comparable to those of the gene cluster reported by Nett et al. with 92–99% identity at the protein level [5] (Figure 3a and Figure S42 in Supporting Information File 1). To investigate whether the identified *var* biosynthetic gene cluster is responsible for variochelin production, we substituted *varG* encoding the polyketide synthase (PKS) module with a chloramphenicol resistance (*cmR*) gene cassette, using homologous recombination (Figure 3b). As a result, the *varG* knock-out mutant no longer produced **1**–**5**, validating that the *var* biosynthetic gene cluster is indispensable for variochelin production in *Variovorax* sp. H002 (Figure 3b). As **1**–**5** share a common biosynthetic pathway, the Hya and the 4-amino-7-guanidino-3-hydroxy-2-methylheptanoic acid moieties in **3**–**5** have the same absolute configurations as those of **1**. In particular, the β-hydroxy group at the Asp residue is introduced by the hydroxylase domain (TauD) embedded in VarG. This type of hydroxylase reportedly generates L-*erythro*-Hya, in agreement with the proposed structures of variochelins [20].

## Biological activity

The antimicrobial activities of **1**–**5** were evaluated against several Gram-negative and Gram-positive bacteria (Table 1). As a result, **1**–**5** inhibited the growth of the Gram-negative bacteria *Escherichia coli* and *Burkholderia multivorans*, but lacked activity against *Bacillus cereus* and *Kocuria rhizophila*. We subsequently investigated the antimicrobial properties of **1** against several known Gram-negative plant pathogens isolated from the same collection site as *Variovorax* sp. H002. Remarkably, variochelin A (**1**) moderately inhibited seven out of eight tested plant pathogens, suggesting that *Variovorax* sp. H002 modulates the local rhizosphere microbiome by utilizing variochelins (Table 2).

# Conclusion

In this study, we reported the isolation of new variochelin congeners **3**–**5**, as well as the known variochelins **1** and **2**, from soil-isolated *Variovorax* sp. H002. The lipopeptide siderophores **1**–**5** share the same amino acid residues containing a combination of two hydroxamates and α-hydroxycarboxylate functional groups, which are capable of forming a photoreactive Fe(III) complex [5]. The common peptide headgroup with the additional repertoire of hydrophobic tails at the *N*-termini (**3**–**5**) resembles those of other amphiphilic marine siderophores such as marinobactins and aquachelins exhibiting the membrane affinity [21] as well as the self-assembling ability to form iron-containing vesicles [22]. This study presents another suite of amphiphilic peptidic siderophores originating in the terrestrial environment. Whether these siderophores play a role in iron acquisition through functions similar to or different from their marine counterparts is an intriguing question that will be answered by future studies.

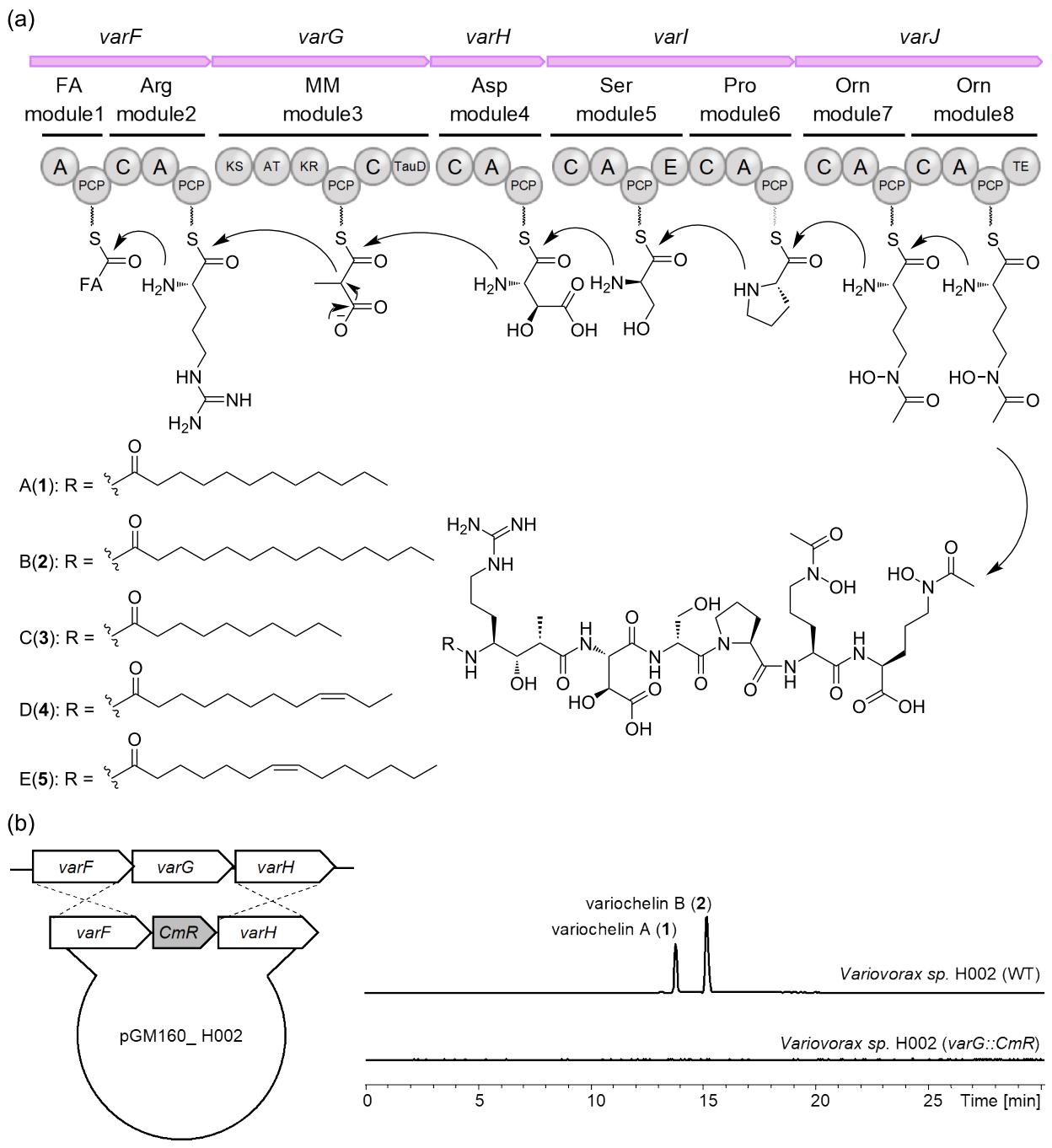



**Figure 3:** (a) Plausible biosynthetic pathway of variochelin A–E (**1**–**5**). The circles represent domains: A: adenylation domain, C: condensation domain, PCP: peptide carrier protein, E: epimerization domain, TE: thioesterase domain, KS: ketosynthase domain, AT: acyl transferase domain, KR: ketoreductase domain, TauD: taurine dioxygenase domain. Substrate specificity of A and KS domains is presented at the top of each module (FA: fatty acyl, Arg: arginine, MM: methyl malonyl, Asp: aspartic acid, Ser: serine, Pro: proline, Orn: ornithine). (b) Validation of *var* biosynthetic gene cluster. The plasmid constructed for the generation of the *varG* null-mutant strain (*varG::cmR*) using the *cmR* gene cassette (left), and HPLC-profile comparison of the *Variovorax* sp. H002 wild type strain and the *varG::cmR* strain monitored at 210 nm (right).

# Experimental

## General experiments

All chemicals were purchased from Wako Pure Chemical Industries (Osaka, Japan), Kanto Chemicals (Tokyo, Japan) and Nacalai Tesque (Kyoto, Japan).

Optical rotations were measured on a JASCO P-1030 polarimeter. NMR spectra were recorded on a JEOL ECA 500 (500 MHz) spectrometer. Chemical shifts are denoted in δ (ppm) relative to residual solvent peaks as internal standards ($CD_3OD$: $\delta_H$ 3.31, $\delta_C$ 49.0). LC–MS experiments and ESI–TOF

**Table 1:** Antibacterial activities of variochelins A (**1**), B (**2**), C (**3**), and E (**5**) against representative Gram-negative and Gram-positive bacteria.

| | MIC (μg/mL) | | | |
|---|---|---|---|---|
| Compound | *E.coli* | *B. multivorans* | *B. cereus* | *K. rhizophila* |
| **1** | 6.15 | 9.98 | ≥125 | ≥125 |
| **2** | 4.47 | 14.66 | ≥125 | ≥125 |
| **3** | 10.14 | 17.84 | ≥125 | ≥125 |
| **5** | 11.77 | 19.23 | ≥125 | ≥125 |

**Table 2:** Antibacterial activities of variochelin A (**1**) against some known soil-isolated plant pathogens.

| Bacteria | MIC (μg/mL) |
|---|---|
| *Burkholderia plantarii* | 8.49 |
| *Burkholderia multivorans* | 10.28 |
| *Agrobacterium tumefaciens* | 16.45 |
| *Erwinia rhapontici* | ≥125 |
| *Pseudomonas syringae* | 25.46 |
| *Pantoea agglomerans* | 21.47 |
| *Ralstonia basilensis* | 34.39 |
| *Paraburkholderia* Sp. 60 | 29.03 |

MS/MS analyses were performed with an amaZon SL-NPC (Bruker Daltonics) or LCMS-2020 (Shimadzu) spectrometer coupled with a Shimadzu HPLC system equipped with an LC-20AD intelligent pump. GC–MS experiments were performed with a Shimadzu-QP2010 Ultra system. HRESIMS analyses were performed with a Thermo Scientific LTQ Orbitrap XL (LTQ XL coupled with LTQ QR) spectrometer.

## Isolation of variochelins A–E (**1**–**5**)

An overnight seed culture of *Variovorax* sp. H002 was inoculated into 100 mL freshly prepared modified minimal M9 media, containing M9 salts (0.7% $K_2HPO_4$, 0.2% $KH_2PO_4$, 0.01% $MgSO_4$, 0.1% $(NH_4)_2SO_4$, 0.06% NaCl), 1% glucose, 1% casamino acids, and HP20 resin (Diaion, Mitsubushi Chemical), in forty baffled acrylic flasks. After growth at 30 °C/140 rpm for seven days, the cultures were filtered through cotton to obtain the HP20 resin, which was packed into a column for chromatography. The packed resin was washed with water several times and then eluted with an excess of methanol. The collected methanol fraction was concentrated in vacuo to obtain a crude extract. Subsequently, the crude extract was subjected to Sephadex LH-20 column chromatography, with methanol as the mobile phase. After monitoring the fractions by a CAS assay, the positive fractions were concentrated and purified by semi-preparative HPLC on a Cosmosil $5C_{18}$-MS-II column (10 mm ID × 250 mm, Nacalai Tesque), using a gradient system with MeCN/$H_2O$ (3:7 to 7:3) as the mobile phase for 30 minutes to yield variochelins A (**1**, 10.0 mg), B (**2**, 20.1 mg), C (**3**, 3.0 mg), D (**4**, 2.1 mg), and E (**5**, 5.3 mg).

Variochelin A (**1**) HRESIMS *m*/*z*: $[M - 2H]^{2-}$, calcd for $C_{47}H_{83}O_{17}N_{11}$, 535.7911; found, 535.7912, Δ 0.2 ppm for $C_{47}H_{83}O_{17}N_{11}$; $[\alpha]_D^{25}$ −7.5, (*c* 1.78, MeOH), $^1H$ and $^{13}C$ NMR chemical shifts see Table S1 (Supporting Information File 1).

Variochelin B (**2**) HRESIMS *m*/*z*: $[M - H]^-$ calcd for $C_{49}H_{87}O_{17}N_{11}$, 1100.6209; found, 1100.6218, Δ 0.8 ppm for; $C_{49}H_{87}O_{17}N_{11}$; $[\alpha]_D^{25}$ −11.1, (*c* 0.23, MeOH), $^1H$ and $^{13}C$ NMR chemical shifts see Table S1 (Supporting Information File 1).

Variochelin C (**3**) HRESIMS *m*/*z*: $[M - H]^-$ calcd for $C_{45}H_{79}O_{17}N_{11}$, 1044.5583; found, 1044.5603, Δ 1.9 ppm for $C_{45}H_{79}O_{17}N_{11}$; $[\alpha]_D^{25}$ −13.8, (*c* 0.25, MeOH), $^1H$ and $^{13}C$ NMR chemical shifts see Table S1 (Supporting Information File 1).

Variochelin D (**4**) HRESIMS *m*/*z*: $[M - 2H]^{2-}$ calcd for $C_{47}H_{81}O_{17}N_{11}$, 534.7833; found, 534.7837, Δ 0.8 ppm for $C_{47}H_{81}O_{17}N_{11}$; $[\alpha]_D^{25}$ −20.9, (*c* 0.19, MeOH), $^1H$ and $^{13}C$ NMR chemical shifts see Table S1 (Supporting Information File 1).

Variochelin E (**5**) HRESIMS *m*/*z*: $[M - H]^-$ calcd for $C_{49}H_{85}O_{17}N_{11}$, 1098.6052; found, 1098.6071, Δ 1.7 ppm for $C_{49}H_{85}O_{17}N_{11}$; $[\alpha]_D^{25}$ −19.7, (*c* 0.21, MeOH), $^1H$ and $^{13}C$ NMR chemical shifts see Table S1 (Supporting Information File 1).

## Preparation of fatty acid methyl ester (FAME) derivatives

0.1 mg of **1** or **2** were hydrolyzed with 6 M HCl at 110 °C overnight. The obtained hydrolysates were freeze dried, resuspended in 100 μL of methanol, and then treated with 100 μL of

trimethylsilyldiazomethane, and the mixtures were kept at room temperature for one hour. The reaction mixtures were concentrated in vacuo, dissolved in acetone, and then injected into a GC–MS system, Zebron ZB-WAX (30 m × 0.25 mm ID × 0.25 μm, Phenomenex), using He as the gas source at a flow rate of 1.41 mL/min. The initial temperature (120 °C) was held for one minute, increased to 230 °C at a rate of 10 °C/min, and then held at 230 °C for six minutes.

## Preparation of 2-alkenyl-4,4-dimethyloxazoline (DMOX) fatty acid derivatives

0.2 mg of **3**–**5** were hydrolyzed with 6 M HCl at 110 °C overnight. The obtained hydrolysates were treated with 250 μL of 2-amino-2-methylpropanol in screw-capped vials. After flushing with nitrogen, the vials were sealed and heated at 180 °C for 18 hours. The reaction mixtures were then cooled down, resuspended in 3 mL of dichloromethane, and washed three times with 1 mL of water. The organic layer was dried with anhydrous $Na_2SO_4$ and concentrated in vacuo. The samples were redissolved in acetone, and then injected into the aforementioned GC–MS system.

## Determination of amino acid configuration by Marfey's method

A 100 μg portion of each compound (**3**–**5**) was hydrolyzed with 6 M HCl at 110 °C overnight. The lyophilized hydrolysates were dissolved in water (160 μL), and treated with saturated $NaHCO_3$ (200 μL) and 1% L-FDLA in acetone (160 μL). The mixtures were heated at 40 °C for 30 minutes, quenched with 1 M HCl (200 μL), and concentrated in vacuo. The residue was dissolved in DMSO (400 μL) and passed through a membrane filter, and then injected into an LC–MS system, Cosmosil 5$C_{18}$-MS-II (2 mm ID × 100 mm, Nacalai Tesque) (oven 40 °C; flowrate 0.2 mL/min) using a gradient of MeCN/$H_2O$ (1:9 to 6:4) containing 0.1% formic acid for 45 minutes. The standards were processed in the same manner.

## Whole genome analysis of *Variovorax* sp. H002

*Variovorax* sp. H002 was streaked onto a 2R2A agar plate and grown at 30 °C overnight. The resulting colonies were inoculated into 5 mL of LB in a 15 mL Falcon tube. After growth at 30 °C/140 rpm for 1 day, the culture was pelleted by centrifugation. The cell pellet was incubated at 37 °C for 1 h with 10 mg/mL lysozyme. After 10% SDS was added to the tube, the mixture was incubated at room temperature, 60 °C, and then 0 °C, for 5 min durations. AcOK (5 M) and phenol/$CHCl_3$/isoamyl alcohol (25:24:1) were then added, and the resulting solution was gently mixed by inversion and centrifuged at 7,000*g* for 10 min. The upper aqueous phase was collected in a new tube. This partitioning step was repeated twice. Isopropanol was added to the aqueous phase obtained after AcOK treatment, and the solution was mixed by inversion and centrifuged at 7,000*g* for 10 min. The pellet was washed with 70% EtOH, dried, and dissolved in TE buffer. The extracted genomic DNA was quantified by a Qubit v3.0 fluorometer (Life Technologies, Thermo Fisher Scientific, Inc.).

The *Variovorax* sp. H002 genome was sequenced by a DNBSEQ for short-lead sequencing and an Oxford nanopore GridION X5 for long-lead sequencing.

For the short-lead sequencing, the library was prepared using 100 ng of the gDNA with an MGIEasy FS DNA Library Prep Set (MGI), following the manufacturer's protocol. The gDNA was fragmented enzymatically to approximately 400 bp. As a result of sequencing (150 bp × 2) and quality filtering (Q > 30), a total of 1.4 Gbp reads were obtained. The genome size of *Variovorax* sp. H002 was roughly estimated to be 8 Mb.

For the long-lead sequencing, 1,000 ng of the gDNA was treated with Short Read Eliminator XS (Circulomics) to remove fragments shorter than 10 kb. The resultant gDNA was subjected to library preparation with a Ligation Sequencing Kit (SQK-LSK 109), following the 1D genomic DNA by ligation protocol. The library was applied to a MinION flowcell (FLO MIN106 R9.41revD) operated by the MinKNOW (20.06.9) software, and then processed by Guppy basecaller (4.0.11) in the high accuracy mode. As a result, a total of 0.4 Gbp reads were obtained.

The reads were subjected to hybrid de novo assembly using SPAdes (v0.4.8) [23]. The assemblies were subjected to BGC prediction using antiSMASH (6.0.1) [24].

## Generation of Δ*varG* mutant

The 1 kb upstream and downstream regions flanking *varG* were amplified by PCR, using *Variovorax* sp. H002 genomic DNA as the template. The chloramphenicol-resistance gene (*cmR*) and a linearized recipient plasmid pRED were amplified by PCR, using pRED as the template. The primer sets used to amplify the above-mentioned four fragments are listed in Table S2 (Supporting Information File 1). The four fragments were combined by Gibson Assembly Reaction (New England Biolabs) to generate a disruption cassette: upstream-*cmR*-downstream. After the transformation of *E. coli* DH5α with the cassette-pRED, a *HindIII* site was inserted on the end of the cassette sequence to yield the disruption plasmid, pGM160-Δ*varG*. The cassette was sequenced for validation. The disruption plasmid was introduced into *E coli* S17-1 λpir by electroporation.

Both *Variovorax* sp. H002 and *E. coli* S17-1 λpir were cultured overnight at 30 °C in 2xR2A and LB, respectively. The overnight cultures were washed with 2xR2A three times, and the pellets were resuspended with 500 mL of 2xR2A. The $OD_{600}$ values of the donor (*E. coli* S17-1 λpir) and recipient (*Variovorax* sp. H002) cells were set to 2.0 and mixed in a 1:1 ratio. The mixtures were spotted onto 2xR2A plates and incubated at 30 °C for two days. The colonies were collected and spread onto 2xR2A agar plates with chloramphenicol (30 μg/mL) and kanamycin (50 μg/mL). The plates were incubated at 30 °C until colonies formed. The disruption of *varG* was confirmed by PCR amplification.

### Effect of ferric ion ($Fe^{3+}$) concentration on production of variochelins

An overnight culture of *Variovorax* sp. H002 in 2xR2A liquid media was transferred to small baffled acrylic flasks containing 15 mL of freshly prepared modified M9 media fortified with 0, 0.5, 1, 5, 10, or 50 μM (final concentration) of $FeCl_3{\cdot}6H_2O$. Upon cultivation at 30 °C/140 rpm for seven days, the cultures were centrifuged, and the supernatants were applied to pre-washed Sep-Pak $C_{18}$ cartridges, which were then washed with water and eluted with methanol. The collected methanolic fractions were dried and resuspended in 0.5 mL of methanol. The particulate-free sample solutions (up to 100 μL) were injected into a Cosmosil $5C_{18}$-MS-II column (10 mm ID × 250 mm, Nacalai Tesque) attached to an HPLC with an autosampler, at a flow rate of 0.8 mL/min at 40 °C. Each analysis was performed using an aqueous acetonitrile mobile phase.

### Biological activities of variochelins

Assays for all compounds were performed in triplicate, according to the Clinical Laboratory Standards Institute testing standard, in 96-well plate microbroth dilution assays. Compounds were tested against Gram-negative bacteria (*Escherichia coli* JW5503, *Bukholderia multivorans* NBRC 102086 and *Bukholderia plantarii* NBRC 104884) and Gram-positive bacteria *(Bacillus cereus* NBRC 15305 and *Kocuria rhizophila* NBRC 12708). In addition, several in-house collections of plant pathogens including *Erwinia rhapontici*_D020, *Pantoea agglomerans*_G054_1, *Agrobacterium tumefaciens*_K021_1, *Pseudomonas syringae*_E029, *Ralstonia basilensis*_M004 and *Paraburkholderia* Sp. 60 isolated from the same location (Botanical Garden of Faculty of Pharmacy Hokkaido University) with *Variovorax* sp. H002 were selected as tested bacteria against variochelin A (**1**). Briefly, the overnight seed cultures of tested bacteria in Mueller–Hinton (MH) media were diluted with a fresh MH media to obtain a final $OD_{600}$ of 0.05. The diluted bacteria were subsequently transferred into a 96-well plate (100 μL/well) and exposed to an appropriate concentration of variochelins. After an incubation at 30 °C for 22 h, the optical density was recorded at 600 nm using plate reader, to determine percent growth inhibition.

## Supporting Information

### Supporting Information File 1

Additional figures and tables, NMR and MS spectra.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-63-S1.pdf]

## Acknowledgements

We thank H. Ikeda (Kitasato University) for providing vectors used for genetic manipulation of *Variovorax* sp. H002.

## Funding

This research was supported by Hokkaido University, Global Facility Center (GFC), Pharma Science Open Unit (PSOU), funded by MEXT under “Support Program for Implementation of New Equipment Sharing System”, Global Station for Biosurfaces and Drug Discovery, a project of Global Institution for Collaborative Research and Education in Hokkaido University, the Japan Agency for Medical Research and Development (AMED, JP23gm1610007, JP23ak0101163, JP23ama121039), and the Japanese Society for the Promotion of Science (JSPS), Kakenhi (16H06448, 18H02581, 21H02635, and 22H05128).

## Author Contributions

Jabal Rahmat Haedar: investigation; validation; visualization; writing – original draft. Aya Yoshimura: conceptualization; investigation; project administration; visualization; writing – review & editing. Toshiyuki Wakimoto: conceptualization; funding acquisition; project administration; resources; supervision; writing – review & editing.

## ORCID® iDs

Aya Yoshimura - https://orcid.org/0000-0001-7665-1027
Toshiyuki Wakimoto - https://orcid.org/0000-0003-2917-1797

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Gurzau, E. S.; Neagu, C.; Gurzau, A. E. *Ecotoxicol. Environ. Saf.* **2003,** *56,* 190–200. doi:10.1016/s0147-6513(03)00062-9
2. Sandy, M.; Butler, A. *Chem. Rev.* **2009,** *109,* 4580–4595. doi:10.1021/cr9002787
3. Hider, R. C.; Kong, X. *Nat. Prod. Rep.* **2010,** *27,* 637–657. doi:10.1039/b906679a

4. Barbeau, K.; Rue, E. L.; Bruland, K. W.; Butler, A. *Nature* **2001,** *413,* 409–413. doi:10.1038/35096545
5. Kurth, C.; Schieferdecker, S.; Athanasopoulou, K.; Seccareccia, I.; Nett, M. *J. Nat. Prod.* **2016,** *79,* 865–872. doi:10.1021/acs.jnatprod.5b00932
6. Finkel, O. M.; Salas-González, I.; Castrillo, G.; Conway, J. M.; Law, T. F.; Teixeira, P. J. P. L.; Wilson, E. D.; Fitzpatrick, C. R.; Jones, C. D.; Dangl, J. L. *Nature* **2020,** *587,* 103–108. doi:10.1038/s41586-020-2778-7
7. Satola, B.; Wübbeler, J. H.; Steinbüchel, A. *Appl. Microbiol. Biotechnol.* **2013,** *97,* 541–560. doi:10.1007/s00253-012-4585-z
8. Larke-Mejía, N. L.; Crombie, A. T.; Pratscher, J.; McGenity, T. J.; Murrell, J. C. *Front. Microbiol.* **2019,** *10,* 2700. doi:10.3389/fmicb.2019.02700
9. Heine, V.; Meinert-Berning, C.; Lück, J.; Mikowsky, N.; Voigt, B.; Riedel, K.; Steinbüchel, A. *PLoS One* **2019,** *14,* e0211876. doi:10.1371/journal.pone.0211876
10. Dhakar, K.; Zarecki, R.; van Bommel, D.; Knossow, N.; Medina, S.; Öztürk, B.; Aly, R.; Eizenberg, H.; Ronen, Z.; Freilich, S. *Front. Bioeng. Biotechnol.* **2021,** *9,* 602464. doi:10.3389/fbioe.2021.602464
11. Johnston, C. W.; Skinnider, M. A.; Wyatt, M. A.; Li, X.; Ranieri, M. R. M.; Yang, L.; Zechel, D. L.; Ma, B.; Magarvey, N. A. *Nat. Commun.* **2015,** *6,* 8421. doi:10.1038/ncomms9421
12. Robertson, A. W.; McCarville, N. G.; MacIntyre, L. W.; Correa, H.; Haltli, B.; Marchbank, D. H.; Kerr, R. G. *J. Nat. Prod.* **2018,** *81,* 858–865. doi:10.1021/acs.jnatprod.7b00943
13. Schwyn, B.; Neilands, J. B. *Anal. Biochem.* **1987,** *160,* 47–56. doi:10.1016/0003-2697(87)90612-9
14. Luthria, D. L.; Sprecher, H. *Lipids* **1993,** *28,* 561–564. doi:10.1007/bf02536089
15. Christie, W. W.; Robertson, G. W.; McRoberts, W. C.; Hamilton, J. T. G. *Eur. J. Lipid Sci. Technol.* **2000,** *102,* 23–29. doi:10.1002/(sici)1438-9312(200001)102:1<23::aid-ejlt23>3.0.co;2-r
16. Hamilton, J. T. G.; Christie, W. W. *Chem. Phys. Lipids* **2000,** *105,* 93–104. doi:10.1016/s0009-3084(99)00133-4
17. Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K.-i. *Anal. Chem. (Washington, DC, U. S.)* **1997,** *69,* 5146–5151. doi:10.1021/ac970289b
18. Fujii, K.; Ikai, Y.; Mayumi, T.; Oka, H.; Suzuki, M.; Harada, K.-i. *Anal. Chem. (Washington, DC, U. S.)* **1997,** *69,* 3346–3352. doi:10.1021/ac9701795
19. Rossi, R.; Carpita, A.; Quirici, M. G.; Veracini, C. A. *Tetrahedron* **1982,** *38,* 639–644. doi:10.1016/0040-4020(82)80205-6
20. Reitz, Z. L.; Hardy, C. D.; Suk, J.; Bouvet, J.; Butler, A. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116,* 19805–19814. doi:10.1073/pnas.1903161116
21. Xu, G.; Martinez, J. S.; Groves, J. T.; Butler, A. *J. Am. Chem. Soc.* **2002,** *124,* 13408–13415. doi:10.1021/ja026768w
22. Martinez, J. S.; Zhang, G. P.; Holt, P. D.; Jung, H.-T.; Carrano, C. J.; Haygood, M. G.; Butler, A. *Science* **2000,** *287,* 1245–1247. doi:10.1126/science.287.5456.1245
23. Antipov, D.; Korobeynikov, A.; McLean, J. S.; Pevzner, P. A. *Bioinformatics* **2016,** *32,* 1009–1015. doi:10.1093/bioinformatics/btv688
24. Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2021,** *49,* W29–W35. doi:10.1093/nar/gkab335

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.63

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Genome mining of labdane-related diterpenoids: Discovery of the two-enzyme pathway leading to (–)-sandaracopimaradiene in the fungus *Arthrinium sacchari*

Fumito Sato, Terutaka Sonohara, Shunta Fujiki, Akihiro Sugawara, Yohei Morishita, Taro Ozaki* and Teigo Asai*

Full Research Paper

Open Access

Address:
Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai 980-8578, Japan

Email:
Taro Ozaki* - taro.ozaki.d3@tohoku.ac.jp; Teigo Asai* - teigo.asai.c8@tohoku.ac.jp

* Corresponding author

Keywords:
diterpenoids; fungi; genome mining; labdane; terpene cyclase

*Beilstein J. Org. Chem.* **2024,** *20,* 714–720.
https://doi.org/10.3762/bjoc.20.65

Received: 11 January 2024
Accepted: 18 March 2024
Published: 03 April 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf

© 2024 Sato et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Labdane-related diterpenoids (LRDs) in fungi are a pharmaceutically important, but underexplored family of natural products. In the biosynthesis of fungal LRDs, bifunctional terpene cyclases (TCs) consisting of αβγ domains are generally used to synthesize the polycyclic skeletones of LRDs. Herein, we conducted genome mining of LRDs in our fungal genome database and identified a unique pair of TCs, AsPS and AsCPS, in the fungus *Arthrinium sacchari*. AsPS consists of catalytically active α and inactive β domains, whereas AsCPS contains βγ domains and a truncated α domain. Heterologous expression in *Aspergillus oryzae* and biochemical characterization of recombinant proteins demonstrated that AsCPS synthesized copalyl diphosphate and that AsPS then converted it to (−)-sandaracopimaradiene. Since AsPS and AsCPS have distinct domain organizations from those of known fungal TCs and are likely generated through fusion or loss of catalytic domains, our findings provide insight into the evolution of TCs in fungi.

## Introduction

Terpenoids are a structurally diverse family of natural products, including more than 80,000 compounds [1]. In the biosynthesis of terpenoids, terpene cyclases (TCs) add structural diversity and complexity. TCs are generally classified into two main classes, class I and class II. Class I TCs initiate the cyclization by heterolytic cleavage of substrates to generate a diphosphate anion and an allylic carbocation, and class II enzymes start cyclization by protonating a double bond or an epoxide existing in their substrates. From a structural point of view, TCs consist of α, β, and γ domains in various combinations. As these TCs

are assumed to be evolved by fusion and loss of domains [2-4], the functional analysis of terpene synthases of unique domain organization is of great significance to understand the evolutionary traits of TCs.

Among terpenoids, labdane-related diterpenoids (LRDs) are an important class which includes biologically active molecules such as plant hormone gibberellins (Figure 1A). In their biosynthesis, class II TCs often synthesize copalyl diphosphate (CPP) or its stereoisomers, *ent*-CPP and *syn*-CPP. Class II TCs can also generate further structural diversity through hydride shifts, methyl shifts, and/or skeletal rearrangement of the labdadienyl$^+$ diphosphate intermediate which is formed by the initial bicyclization. Following this class II TC-mediated cyclization, class I TCs subsequently catalyze the second cyclization to construct the polycyclic scaffold of natural products [5]. In plants, two independent αβγ tri-domain TCs, *ent*-CPP synthase (CPS) and *ent*-kaurene synthase (KS), are often used for this conversion [6], and a single bifunctional enzyme that successively catalyzes these reactions is also known [7]. Bacteria also use two enzyme systems for the biosynthesis of LRDs, but the domain organization of the corresponding TCs is different from those of plant enzymes. In bacteria, the class II enzymes with βγ domains and the class I enzyme with a single α domain are employed [8]. In fungi, only bifunctional enzymes consisting of αβγ tri-domains have been identified to date [9-13]. In the

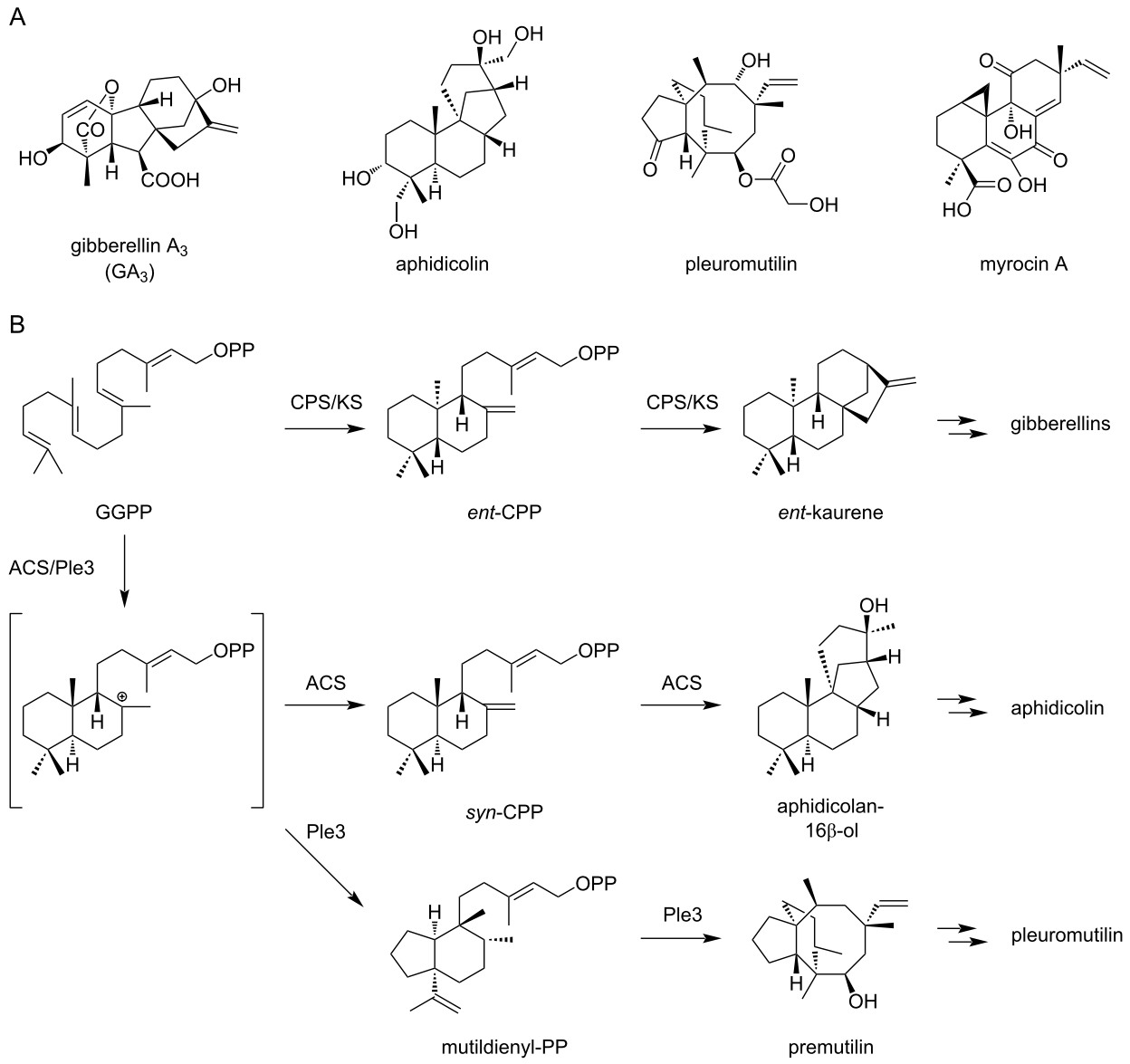



**Figure 1:** LRDs in fungi. A) Chemical structures of representative fungal LRDs. B) Reactions catalyzed by selected fungal bifunctional TCs.

evolutionary aspects, fungal bifunctional TCs are proposed to have been acquired from plants by a horizontal gene transfer event [14] and eukaryotic tri-domain TCs are proposed to be derived from the early fusion of bacterial class I and II enzymes [3,15].

Fungal genomes contain a number of biosynthetic genes that are attractive targets for genome mining of novel natural products and biosynthetic enzymes [16-18]. In our research group, genome-guided exploration of fungal natural products and biosynthetic machineries have been conducted, leading to the discovery of various new natural products and enzymes [19-27]. Although fungal LRDs are a biologically and pharmaceutically important class of natural products including DNA polymerase α inhibitor aphidicolin [28] and a ribosome-targeting antibiotic pleuromutilin [29] (Figure 1), they remained as underexplored targets in genome-mining approaches. In this study, we examined the biosynthetic genes for LRDs and identified two TCs consisting of αβ and αβγ domains. Heterologous expression in *Aspergillus oryzae* and biochemical characterization of recombinant enzymes unveiled a fungal two-enzyme pathway to isopimaradiene, which is unprecedented for fungal LRDs.

## Results and Discussion

To explore the fungal LRDs, we searched our in-house database by BLAST using known fungal bifunctional TCs such as aphidicolan-16β-ol synthase (ACS) as queries. We identified a pair of genes encoding TCs, which we later defined as AsPS (isopimaradiene synthase) and AsCPS, in the genome of *Arthrinium sacchari* Kumo-3 and *Ar. sacchari* MPU169. AsPS and AsCPS are clustered with the putative geranylgeranyl diphosphate synthase gene (AsGGS), suggesting that they are related to the biosynthesis of diterpenes (Figure 2A). AsPS and AsCPS consist of αβ and αβγ domains, respectively. Further sequence analysis revealed that the β domain of AsPS contains a mutation at the catalytic aspartic acid residue, indicating that this enzyme functions as class I TC (Figure 2B and Figure S1 in Supporting Information File 1). AsCPS was assumed to be a class II TC because its α domain is truncated and is likely nonfunctional. Given these domain organizations of identified TCs, we hypothesized that AsPS and AsCPS would likely cooperate to synthesize a diterpene skeleton: After the initial cyclization of GGPP possibly catalyzed by a putative class II enzyme AsCPS, the resulting labdane or related skeleton would undergo cyclization by a putative class I enzyme AsPS. To assess this hypothesis, we conducted the functional characterization of AsPS and AsCPS as described below.

First, we examined the heterologous expression of *AsPS* and *AsCPS* in the heterologous host *Aspergillus oryzae* NSAR1. These two genes along with *AsGGS* were amplified and integrated into the pUARA2 vector, yielding the plasmid pUARA2-*AsGGS*/*AsCPS*/*AsPS* (Table S1 in Supporting Information File 1). The constructed plasmid was introduced into *A. oryzae*

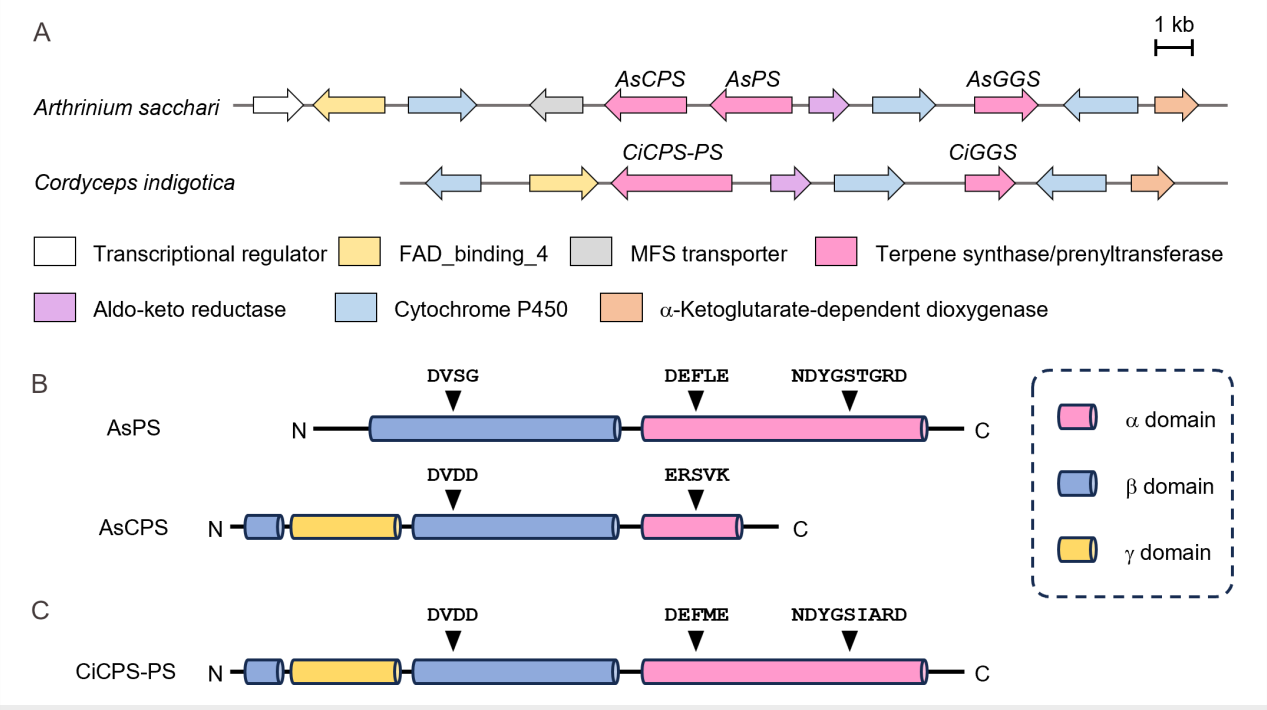



**Figure 2:** Sequence analysis of AsPS and AsCPS. A) Biosynthetic gene clusters containing AsCPS and AsPS. B) Domain organization of AsPS and AsCPS. C) Domain organization of CiCPS-PS.

to give the transformant AO-*AsGGS*/*AsCPS*/*AsPS*. We then analyzed the metabolite of this transformant by GC–MS. By comparing the metabolites extracted from *A. oryzae* NSAR1, we identified compound **1** as a product (Figure 3A and Figure S2 in Supporting Information File 1). We also examined the expression of several oxidoreductases in the flanking region in the same host strain, but the possible products other than **1** were not clearly observed (data not shown). Further characterization of oxidative modifications is in progress and will be reported elsewhere.

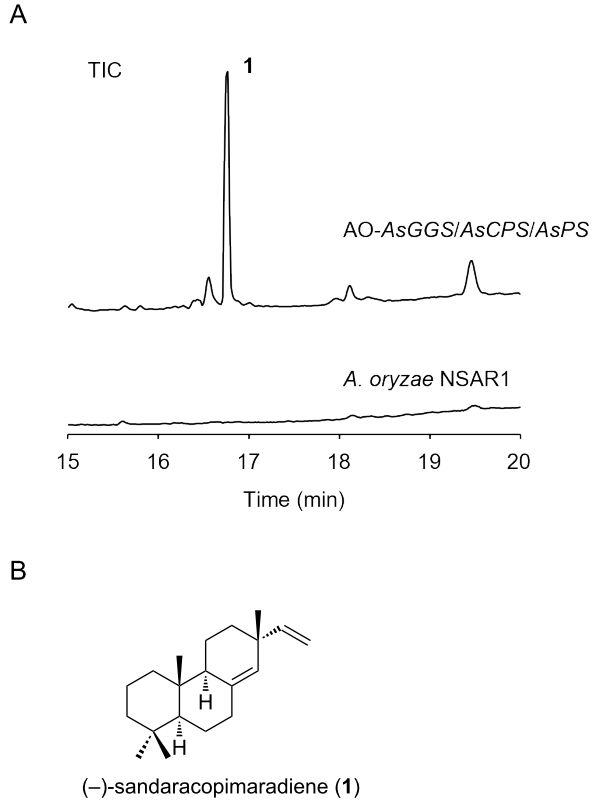



**Figure 3:** GC–MS analysis of the *A. oryzae* transformant. A) TIC of the extracts obtained from AO-*AsGGS*/*AsCPS*/*AsPS* and *A. oryzae* NSAR1. Compound **1** was identified to be a product, while the other minor peaks were not related to diterpenoids and were likely derived from the host strain. B) Chemical structure of (–)-sandaracopimaradiene (**1**)).

We successfully purified compound **1** from the mycelial extract of the *A. oryzae* transformant harboring *AsGGS*, *AsPS*, and *AsCPS* with a yield of 10.1 mg. The chemical formula of **1** was elucidated to be $C_{20}H_{32}$ due to the monoisotopic ion peak at *m*/*z* 272.2494 $[M]^{+\cdot}$ (calcd for $C_{20}H_{32}$, 272.2499) in the HREIMS spectrum. The $^{1}H$ and $^{13}C$ NMR spectra of **1** were identical to those of sandaracopimaradiene [30,31]. As the specific rotation of compound **1** was in good agreement with the reported data ($[\alpha]_D^{24}$ −14.3 (*c* 0.97, $CHCl_3$) in this study; $[\alpha]_D^{25}$ −10.29 (*c* 0.65, $CHCl_3$) in the literature [31]), compound **1** was determined to be (−)-sandaracopimaradiene (Figure 3B).

We then turned our attention to the individual function of these enzymes. With this aim, AsPS and AsCPS were expressed in *Escherichia coli* as N-terminal hexa-histidine-tagged proteins and purified by Ni-affinity chromatography (see Supporting Information File 1, Figure S3). The purified proteins were incubated in the presence of GGPP and $Mg^{2+}$. GC–MS analysis of the hexane extract revealed the enzyme-dependent formation of compound **1** (Figure 4A and Figure S2 in Supporting Information File 1). The formation of **1** was not observed when either AsPS or AsCPS was omitted from the reaction mixture, showing that both enzymes were indispensable for the biosynthesis of compound **1**.

Based on the abovementioned results and domain organizations of AsPS and AsCPS, we hypothesized that compound **1** was synthesized via *normal*-CPP. We incubated AsPS with the cell-free extract of *E. coli* expressing ObCPS_11g, which synthesizes *normal*-CPP [32]. As expected, the formation of **1** was observed (Figure 4A and Figure S2 in Supporting Information File 1), indicating that the product of AsCPS was the same as that of ObCPS_11g. In addition,we incubated AsCPS in the presence of GGPP, followed by hydrolysis by acid phosphatase. GC–MS analysis revealed that AsCPS synthesizes the same product to ObCPS_11g (see Supporting Information File 1, Figure S4A and B). It should be noted that the stereochemistry of CPP needs further verification due to the lack of chiral resolution of our GC–MS conditions. Based on these results, we proposed that AsCPS synthesizes CPP and that AsPS subsequently converted it to compound **1**.

Notably, *Ar. sacchari* MPU169 is known as a producer of myrocins (Figure 1A), which are known to exhibit antiangiogenic activity [33]. Based on the chemical structure and absolute configuration of **1**, we assumed this compound could be a biosynthetic precursor of myrocins. In the flanking regions of *AsPS* and *AsCPS*, genes encoding oxidoreductases such as cytochrome P450 are present. These genes might be involved in the oxidation of **1** in the biosynthesis of myrocins. A similar gene cluster was also found in the genome of *Cordyceps indigotica* (Figure 2A). In this cluster, one bifunctional TC, which we later defined as CiCPS-PS, was present instead of separate TCs, such as AsCPS and AsPS. Although we could identify similar BGCs containing one bifunctional TCs in the public database, there are no similar BGCs with separate type TCs (Figure S5 in Supporting Information File 1). Heterologous expression of CiCPS-PS with AsGGS also gave **1** as a product (Figure 5A and Figure S2 in Supporting Information File 1), showing that CiCPS-PS retains the same activity as those of

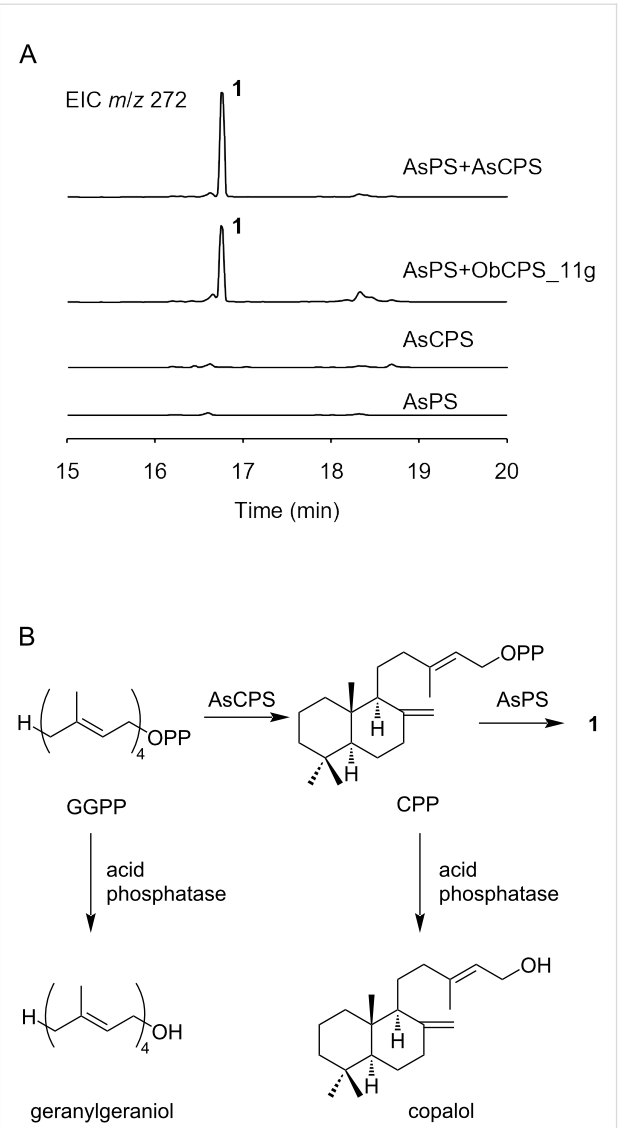



**Figure 4:** GC–MS analysis of the enzymatic reactions. A) Extracted ion chromatograms (EICs) at *m/z* 272 of the extracts from the reaction mixture of AsPS and AsCPS; AsPS and ObCPS_11g are shown. B) Proposed reactions catalyzed by AsPS, AsCPS, and acid phosphatase.

AsCPS and AsPS. Interestingly, CiCPS-PS exhibits the relatively low sequence identity to AnCPS-PS (34%) and NfCPS-PS (39%), both of which synthesize **1** [13]. This observation may indicate that evolutional origins of CiCPS-PS and these enzymes are different or not closely related. As mentioned above, TCs are considered to be evolved by fusion and loss of domains. Following this model, CiCPS-PS might be ancestral and its duplication and subsequent loss of domains likely generate separate type TCs, such as AsCPS and AsPS (Figure 5B, model a). Alternatively, translocation and clustering of two bifunctional TCs instead of gene duplication TCs can also be considered (Figure 5B, model b).

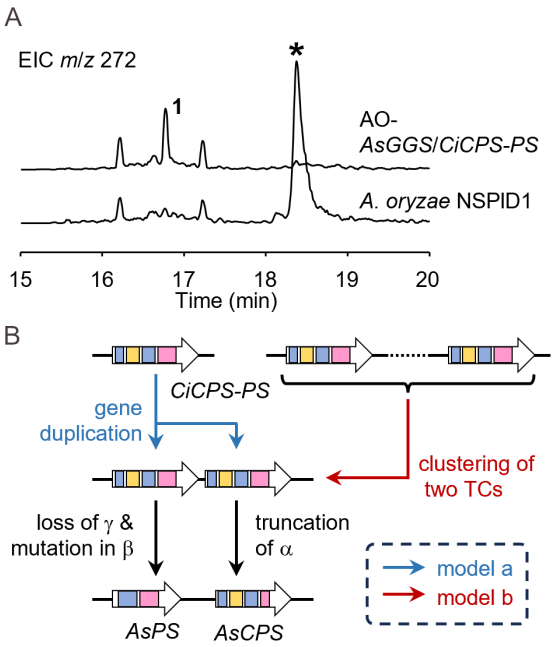



**Figure 5:** Characterization of CiCPS-PS. A) GC–MS analysis of the transformant expressing *CiCPS-PS* and *AsGGS*. EICs at *m/z* 272 are shown. The peak marked by the asterisk is probably derived from the host strain. B) Putative models of evolutionary relationships of CiCPS-PS, AsPS, and AsCPS.

## Conclusion

Through genome mining of LRDs in fungi, we identified a unique pair of TCs in which domain organizations are distinct from those of known fungal TCs involved in the related pathways. Heterologous expression in *A. oryzae* and biochemical characterization of recombinant proteins showed that the identified enzymes are responsible for the biosynthesis of (−)-**1**. We also identified the bifunctional TC that could synthesize **1**. Since a set of enzymes retaining the same activity but having different domain organizations were discovered, these enzymes would be useful to discuss and further analyze the evolution of TCs in fungi.

## Supporting Information

### Supporting Information File 1

Experimental part.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-65-S1.pdf]

## Acknowledgements

We thank Prof. Kazunori Okada (The University of Tokyo) and Prof. Tomonobu Toyomasu (Yamagata University) for providing the expression plasmid of ObCPS_11g, Prof. Kaoru

Kinoshita (Meiji Pharmaceutical University) for providing *Ar. sacchari* MPU169, Prof. Tomohisa Kuzuyama (The University of Tokyo) for providing GGPP, and Prof. Hideaki Oikawa (Hokkaido University), Prof. Atsushi Minami (Hokkaido University), and Prof. Jun-ichi Maruyama (The University of Tokyo) for providing the genome editing system of *A. oryzae*.

## Funding

This work was supported by JSPS KAKENHI (22H02775 [T.A.], 23H04538 [T.O.], 22K19095 [T.O.]); AMED-CREST (23gm1610007 [T.A.]); Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research [BINDS]) from AMED [JP23ama121038 (T.A.) and JP23ama121040 (T.O.)].

## ORCID® iDs

Fumito Sato - https://orcid.org/0009-0009-5912-2532
Akihiro Sugawara - https://orcid.org/0000-0003-1266-7497
Yohei Morishita - https://orcid.org/0000-0001-8442-8642
Taro Ozaki - https://orcid.org/0000-0002-4673-4478
Teigo Asai - https://orcid.org/0000-0002-8903-392X

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
2. Gao, Y.; Honzatko, R. B.; Peters, R. J. *Nat. Prod. Rep.* **2012,** *29,* 1153–1175. doi:10.1039/c2np20059g
3. Cao, R.; Zhang, Y.; Mann, F. M.; Huang, C.; Mukkamala, D.; Hudock, M. P.; Mead, M. E.; Prisic, S.; Wang, K.; Lin, F.-Y.; Chang, T.-K.; Peters, R. J.; Oldfield, E. *Proteins: Struct., Funct., Bioinf.* **2010,** *78,* 2417–2432. doi:10.1002/prot.22751
4. Moosmann, P.; Ecker, F.; Leopold-Messer, S.; Cahn, J. K. B.; Dieterich, C. L.; Groll, M.; Piel, J. *Nat. Chem.* **2020,** *12,* 968–972. doi:10.1038/s41557-020-0515-3
5. Peters, R. J. *Nat. Prod. Rep.* **2010,** *27,* 1521–1530. doi:10.1039/c0np00019a
6. Toyomasu, T.; Shenton, M. R.; Okada, K. *Plant Cell Physiol.* **2020,** *61,* 1850–1859. doi:10.1093/pcp/pcaa106
7. Hayashi, K.-i.; Kawaide, H.; Notomi, M.; Sakigi, Y.; Matsuo, A.; Nozaki, H. *FEBS Lett.* **2006,** *580,* 6175–6181. doi:10.1016/j.febslet.2006.10.018
8. Smanski, M. J.; Yu, Z.; Casper, J.; Lin, S.; Peterson, R. M.; Chen, Y.; Wendt-Pienkowski, E.; Rajski, S. R.; Shen, B. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108,* 13498–13503. doi:10.1073/pnas.1106919108
9. Yamane, M.; Minami, A.; Liu, C.; Ozaki, T.; Takeuchi, I.; Tsukagoshi, T.; Tokiwano, T.; Gomi, K.; Oikawa, H. *ChemBioChem* **2017,** *18,* 2317–2322. doi:10.1002/cbic.201700434
10. Oikawa, H.; Toyomasu, T.; Toshima, H.; Ohashi, S.; Kawaide, H.; Kamiya, Y.; Ohtsuka, M.; Shinoda, S.; Mitsuhashi, W.; Sassa, T. *J. Am. Chem. Soc.* **2001,** *123,* 5154–5155. doi:10.1021/ja015747j
11. Toyomasu, T.; Kawaide, H.; Ishizaki, A.; Shinoda, S.; Otsuka, M.; Mitsuhashi, W.; Sassa, T. *Biosci., Biotechnol., Biochem.* **2000,** *64,* 660–664. doi:10.1271/bbb.64.660
12. Toyomasu, T.; Niida, R.; Kenmoku, H.; Kanno, Y.; Miura, S.; Nakano, C.; Shiono, Y.; Mitsuhashi, W.; Toshima, H.; Oikawa, H.; Hoshino, T.; Dairi, T.; Kato, N.; Sassa, T. *Biosci., Biotechnol., Biochem.* **2008,** *72,* 1038–1047. doi:10.1271/bbb.70790
13. Xu, M.; Hillwig, M. L.; Tiernan, M. S.; Peters, R. J. *J. Nat. Prod.* **2017,** *80,* 328–333. doi:10.1021/acs.jnatprod.6b00764
14. Fischer, M. J.; Rustenhloz, C.; Leh-Louis, V.; Perrière, G. *BMC Microbiol.* **2015,** *15,* 221. doi:10.1186/s12866-015-0564-8
15. Smanski, M. J.; Peterson, R. M.; Huang, S.-X.; Shen, B. *Curr. Opin. Chem. Biol.* **2012,** *16,* 132–141. doi:10.1016/j.cbpa.2012.03.002
16. Vesth, T. C.; Nybo, J. L.; Theobald, S.; Frisvad, J. C.; Larsen, T. O.; Nielsen, K. F.; Hoof, J. B.; Brandl, J.; Salamov, A.; Riley, R.; Gladden, J. M.; Phatale, P.; Nielsen, M. T.; Lyhne, E. K.; Kogle, M. E.; Strasser, K.; McDonnell, E.; Barry, K.; Clum, A.; Chen, C.; LaButti, K.; Haridas, S.; Nolan, M.; Sandor, L.; Kuo, A.; Lipzen, A.; Hainaut, M.; Drula, E.; Tsang, A.; Magnuson, J. K.; Henrissat, B.; Wiebenga, A.; Simmons, B. A.; Mäkelä, M. R.; de Vries, R. P.; Grigoriev, I. V.; Mortensen, U. H.; Baker, S. E.; Andersen, M. R. *Nat. Genet.* **2018,** *50,* 1688–1695. doi:10.1038/s41588-018-0246-1
17. Inglis, D. O.; Binkley, J.; Skrzypek, M. S.; Arnaud, M. B.; Cerqueira, G. C.; Shah, P.; Wymore, F.; Wortman, J. R.; Sherlock, G. *BMC Microbiol.* **2013,** *13,* 91. doi:10.1186/1471-2180-13-91
18. Sanchez, J. F.; Somoza, A. D.; Keller, N. P.; Wang, C. C. C. *Nat. Prod. Rep.* **2012,** *29,* 351–371. doi:10.1039/c2np00084a
19. Morishita, Y.; Zhang, H.; Taniguchi, T.; Mori, K.; Asai, T. *Org. Lett.* **2019,** *21,* 4788–4792. doi:10.1021/acs.orglett.9b01674
20. Kaneko, A.; Morishita, Y.; Tsukada, K.; Taniguchi, T.; Asai, T. *Org. Biomol. Chem.* **2019,** *17,* 5239–5243. doi:10.1039/c9ob00807a
21. Morishita, Y.; Tsukada, K.; Murakami, K.; Irie, K.; Asai, T. *J. Nat. Prod.* **2022,** *85,* 384–390. doi:10.1021/acs.jnatprod.1c00973
22. Morishita, Y.; Sonohara, T.; Taniguchi, T.; Adachi, K.; Fujita, M.; Asai, T. *Org. Biomol. Chem.* **2020,** *18,* 2813–2816. doi:10.1039/d0ob00519c
23. Homma, Y.; Sugawara, A.; Morishita, Y.; Tsukada, K.; Ozaki, T.; Asai, T. *Org. Lett.* **2022,** *24,* 3504–3509. doi:10.1021/acs.orglett.2c01172
24. Furumura, S.; Ozaki, T.; Sugawara, A.; Morishita, Y.; Tsukada, K.; Ikuta, T.; Inoue, A.; Asai, T. *J. Nat. Prod.* **2023,** *86,* 398–405. doi:10.1021/acs.jnatprod.2c01027
25. Shi, Y.; Ozaki, T.; Sugawara, A.; Morishita, Y.; Pei Tan, Y.; Shivas, R. G.; Davis, R. A.; Asai, T. *Tetrahedron Lett.* **2024,** *134,* 154865. doi:10.1016/j.tetlet.2023.154865
26. Morishita, Y.; Aoki, Y.; Ito, M.; Hagiwara, D.; Torimaru, K.; Morita, D.; Kuroda, T.; Fukano, H.; Hoshino, Y.; Suzuki, M.; Taniguchi, T.; Mori, K.; Asai, T. *Org. Lett.* **2020,** *22,* 5876–5879. doi:10.1021/acs.orglett.0c01975
27. Tsukada, K.; Shinki, S.; Kaneko, A.; Murakami, K.; Irie, K.; Murai, M.; Miyoshi, H.; Dan, S.; Kawaji, K.; Hayashi, H.; Kodama, E. N.; Hori, A.; Salim, E.; Kuraishi, T.; Hirata, N.; Kanda, Y.; Asai, T. *Nat. Commun.* **2020,** *11,* 1830. doi:10.1038/s41467-020-15664-4
28. Ikegami, S.; Taguchi, T.; Ohashi, M.; Oguro, M.; Nagano, H.; Mano, Y. *Nature* **1978,** *275,* 458–460. doi:10.1038/275458a0

29. Kavanagh, F.; Hervey, A.; Robbins, W. J. *Proc. Natl. Acad. Sci. U. S. A.* **1951,** *37,* 570–574. doi:10.1073/pnas.37.9.570
30. Guajardo Touché, E. M.; Lopez, E. G.; Reyes, A. P.; Sánchez, H.; Honecker, F.; Achenbach, H. *Phytochemistry* **1997,** *45,* 387–390. doi:10.1016/s0031-9422(96)00807-2
31. Tungcharoen, P.; Wattanapiromsakul, C.; Tansakul, P.; Nakamura, S.; Matsuda, H.; Tewtrakul, S. *Phytother. Res.* **2020,** *34,* 612–623. doi:10.1002/ptr.6549
32. Toyomasu, T.; Goda, C.; Sakai, A.; Miyamoto, K.; Shenton, M. R.; Tomiyama, S.; Mitsuhashi, W.; Yamane, H.; Kurata, N.; Okada, K. *Biochem. Biophys. Res. Commun.* **2018,** *503,* 1221–1227. doi:10.1016/j.bbrc.2018.07.028
33. Tsukada, M.; Fukai, M.; Miki, K.; Shiraishi, T.; Suzuki, T.; Nishio, K.; Sugita, T.; Ishino, M.; Kinoshita, K.; Takahashi, K.; Shiro, M.; Koyama, K. *J. Nat. Prod.* **2011,** *74,* 1645–1649. doi:10.1021/np200108h

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.65

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Chemoenzymatic synthesis of macrocyclic peptides and polyketides via thioesterase-catalyzed macrocyclization

Senze Qiao[1], Zhongyu Cheng[1] and Fuzhuo Li[*1,2]

Review

Open Access

Address:
[1]Department of Natural Medicine, School of Pharmacy, Fudan University, Shanghai, 201203, China and [2]Key Laboratory of Smart Drug Delivery (Ministry of Education), State Key Laboratory of Medical Neurobiology, Fudan University, Shanghai, 201203, China

Email:
Fuzhuo Li* - lifuzhuo@fudan.edu.cn

* Corresponding author

Keywords:
biosynthesis; chemoenzymatic synthesis; macrocyclic peptides; macrocyclic polyketides; thioesterase

*Beilstein J. Org. Chem.* **2024,** *20,* 721–733.
https://doi.org/10.3762/bjoc.20.66

Received: 14 December 2023
Accepted: 21 March 2024
Published: 04 April 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf

© 2024 Qiao et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

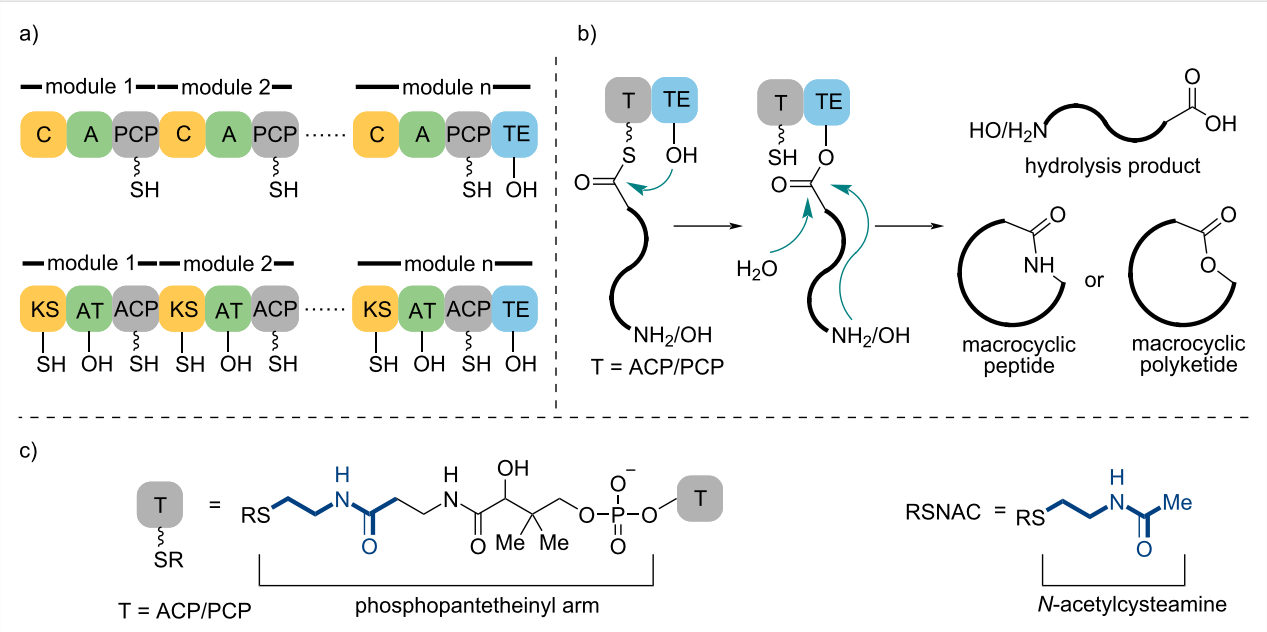
Chemoenzymatic strategies that combine synthetic and enzymatic transformations offer efficient approaches to yield target molecules, which have been increasingly employed in the synthesis of bioactive natural products. In the biosynthesis of macrocyclic nonribosomal peptides, polyketides, and their hybrids, thioesterase (TE) domains play a significant role in late-stage macrocyclization. These domains can accept mimics of native substrates in vitro and exhibit potential for use in total synthesis. This review summarizes the recent advances of TE domains in the chemoenzymatic synthesis for these natural products that aim to address the common issues in classical synthetic approaches and increase synthetic efficiencies, which have the potential to facilitate further pharmaceutical research.

## Introduction

Nonribosomal peptides, polyketides, and their hybrids exhibit significant diversity and a broad spectrum of bioactivities [1-3]. Particularly, macrocycles from these three categories of natural products are vital resources for developing pharmaceuticals and drug candidates [4]. Many clinical pharmaceuticals with high market value, including the immunosuppressants cyclosporin and rapamycin, the antibiotics daptomycin and erythromycin, and the antitumor agent epothilone, all belong to these categories. Therefore, the rising demand for corresponding therapeutics requires effective and economical preparation methods [5]. In the synthesis of these natural products and their analogs, macrocyclization through linear precursors, the key step in the general routes, was typically accomplished via conventional chemical methodologies [6,7], keeps presenting an obstacle. Developing more efficient and diverse macrocyclization strategies is urgently needed to overcome issues such as insufficient regioselectivity, intermolecular oligomerization, the overuse of protective groups, and other drawbacks [8].

In the biosynthetic logic, these natural products are produced by the large and multifunctional enzymatic assembly, nonribosomal peptide synthases (NRPS), polyketide synthases (PKS), and hybrid NRPS/PKS systems, which are organized into sets of functional domains known as modules and function through a similar mechanism [9-12]. Each NRPS module is composed of three essential domains, namely adenylation (A), condensation (C), and peptidyl carrier protein (PCP). Each type I PKS module consists of three core domains containing acyltransferase (AT), ketosynthase (KS), and acyl carrier protein (ACP). PCP and ACP are collectively called thiolation domain (T). The sequence of monomers in the final product intimately correlates with the order of modules in the assembly line (Scheme 1a). Beyond several additional domains, including ketoreductase (KR), dehydratase (DH), enoyl reductase (ER), and methyltransferase (MT) domains and epimerase (E) domains, which are responsible for the structural diversity of natural products, both NRPS and PKS contain thioesterase (TE) domains in the final elongation module, which contribute to terminating biosynthesis [13,14]. Typically, TE domains cleave the thioester bond between the last PCP or ACP domain and the intermediate of polyketides or peptides, and form an ester bond. Then, they catalyze either intramolecular macrocyclization to give macrolactones or macrolactams with attacking of internal nucleophiles (alcohols or amine), or hydrolysis to release linear acids or peptides (Scheme 1b). Although TE domains may display cyclization and hydrolytic activities that are not easily predictable, related mechanism studies indicated that the pre-reaction states of the enzyme and substrate are critical for selectivity [15,16]. Thus, both the mutation of key residues in the active pocket and the addition of a nonionic detergent can increase the ratio of intramolecular nucleophilic attack, resulting in macrocyclic products via preorganization of substrate and enzyme in an active conformation [17,18].

Chemoenzymatic strategies, which merge practical enzymatic transformations with modern organic synthetic methods to increase the efficiency of synthetic approaches, have already shown a growing influence in the synthesis of bioactive natural products, pharmaceutical components, and other valuable molecules with the development of microbial genetics and enzyme engineering [19-22]. The comprehensive investigation of TE domains presents a significant opportunity to establish more efficient and environmentally friendly approaches toward bioactive macrocyclic peptides and macrolides [23]. Nevertheless, the native substrates of TE domains are tethered with PCP or ACP in biosynthetic pathways, which have to be simplified to chemically synthetic mimics before developing the enzymatic transformation. Due to *N*-acetylcysteamine (NAC) having a substructure to the phosphopantetheinyl arm of the carrier protein [24,25], the corresponding thioester can be recognized by TE domains and has become the most common substrate in enzymatic macrocyclizations (Scheme 1c). NAC thioester and other related mimics (such as coenzyme A (CoA), phosphopantetheine, and thiophenol) span the gap between the chemical synthesis and biosynthesis languages and expand the substrate promiscuity of TE domains. This bridge makes the in vitro TE-catalyzed macrocyclization possible and provides a poten-




**Scheme 1:** Brief introduction of thioesterase (TE) domain. (a) NRPS and PKS assembly lines. (b) Mechanism of TE domain-catalyzed macrocyclization and product release. (c) The common phosphopantetheinyl arm and its mimic (*N*-acetylcysteamine).

tial to construct the analogs library of these bioactive macrocycles for further biological investigations.

This review presents representative examples of chemoenzymatic approaches for macrocyclic peptides, polyketides, and their hybrids employing TE domains, and particular attention is given to the strategies of mimics formation to demonstrate how biocatalysis provides an elegant link between chemistry and biology.

# Review

## Macrocyclic peptides

Since first being reported in the 1960s [26], solid-phase peptide synthesis (SPPS) has been an invaluable tool for preparing numerous peptides and even small proteins. In the chemoenzymatic synthesis of macrocyclic peptides, SPPS strategies provide highly efficient routes to access linear precursors, accelerating the development of enzymatic macrocyclization.

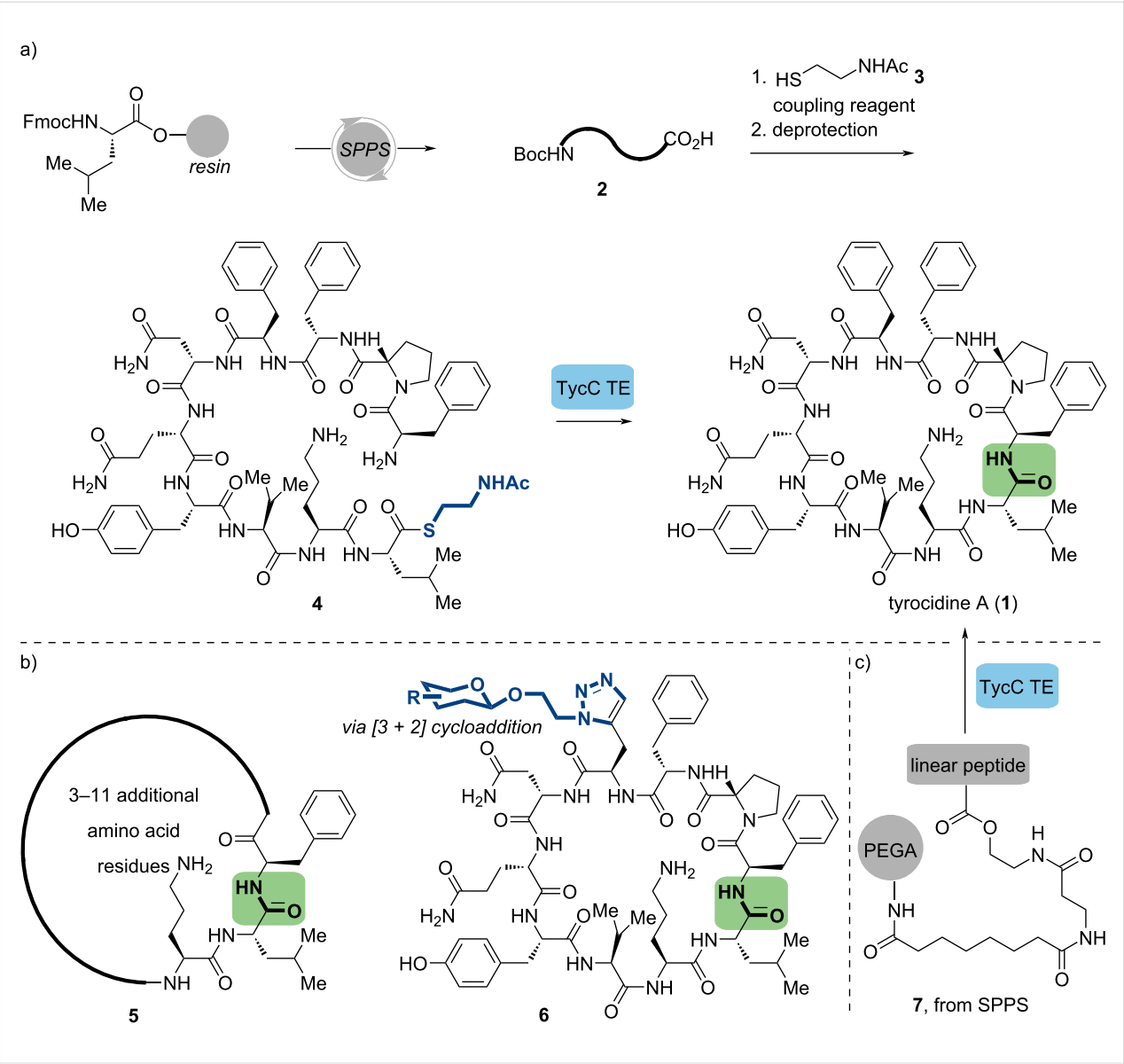
### The tyrocidines

Tyrocidine A (**1**), a cyclic decapeptide isolated from *Bacillus brevis* [27], can penetrate the lipid phase of a Gram-positive inner cell membrane [28,29]. Despite exhibiting high antimicrobial activity, this compound also disrupts the membranes of higher mammalian cells, as evidenced by their pronounced hemolytic activity [30]. Establishing a concise and diverse method to produce analog libraries is critical for structure–activity relationship studies to enhance its specificity. Since the first total synthesis by Ohno and Izumiya in 1966 [31], tyrocidine A and its analogs have been synthesized by several groups employing viable strategies over the past half-century [32-35].

Biosynthetically, the corresponding cluster consists of three NRPS, TycA-C, and at the C-terminus of TycC, the TE domain can catalyze a head-to-tail macrocyclization and deliver tyrocidines [30]. With a comprehensive understanding of its biosynthetic mechanism, Walsh and co-workers developed an elegant chemoenzymatic route in 2000 [36]. The synthesis commenced to construct linear decapeptide **2** by global SPPS method on 2-chlorotrityl resin. After coupling **2** with *N*-acetylcysteamine (NAC, **3**), the mimic of peptide-*S*-PCP, peptide-SNAC **4** was prepared. When incubated with purified TycC TE, the precursor was effectively converted into the macrocyclic tyrocidine A (**1**), exhibiting a low rate of substrate hydrolysis (Scheme 2a). In addition, TycC TE demonstrated a broad range of substrate tolerance, as it can cyclize a series of decapeptide-NACs that contain non-native residues in several positions and also form 6–14 residue cyclic peptides [37,38]. It should be noted that TycC TE was more sensitive to the amino acid changes near the site of ring closure. The alkyne-containing analogs were conjugated to a variety of azido sugars via copper(I)-catalyzed cycloaddition to obtain the corresponding carbohydrates modified tyrocidine derivatives [39], two of which exhibited a 6-fold better therapeutic index than the natural tyrocidine (Scheme 2b).

This seminal study illustrates that an isolated TE domain retains cyclization activity when peptide-SNAC is utilized to replace peptide-S-PCP. This strategy has been widely employed in the TE domains characterization and chemoenzymatic synthesis of other bioactive macrocyclic peptides, such as surfactin [40], streptogramin B [41], cereulide [42], seongsanamide E [43], etc. In 2005, Marahiel and co-workers accomplished the chemoenzymatic synthesis of type B streptogramin variants [41], including pristinamycin IE (**8**), which belongs to the class of depsipeptide antibiotics. Similarly, the linear peptide-SNAC was prepared through the SPPS method on 2-chlorotrityl resin, and macrolactonization was catalyzed by SnbDE TE, a thioesterase from pristinamycin I NRPS. This TE domain showed activity for hydroxy groups and amines to form either lactone or lactam, and the broad substrate scope made this strategy potent for modifying the bioactivity of streptogramin antibiotics. In 2007, the same laboratory identified the interactive TE domain of the gramicidin S synthetase GrsB [44]. Combined with the peptidyl carrier protein, GrsB PCP-TE was tested by using corresponding pentapeptides NAC thioester and thiophenol thioester, which led to the formation of the desired cyclic decapeptide lactam gramicidin S (**9**) through a sequential dimerization and cyclization process. Most recently, the synthesis of monocyclic depsipeptide, seongsanamide E (**10**), was reported by Boddy and co-workers via two different strategies [43]. On the one hand, the regular chemical approach, attempting Yonemitsu's conditions to macrolactonize the seco-acid, was unsuccessful. On the other hand, the chemoenzymatic process using purified Sgd TE from its biosynthesis and a linear peptide SNAC substrate gave the macrocycle in an acceptable yield without epimerization (Scheme 3).

Although NAC-containing thioesters were widely employed, as described above, it had several limitations, such as possible Cα epimerization [45] during SNAC coupling and essential HPLC purification, which was generally difficult and time-consuming. Developing other different methods, exceptionally more straightforward approaches to access activated substrates, would solve this inevitable bottleneck and promote the utilization of TE domains as biocatalysts with tremendous potential. To this purpose, Walsh and co-workers developed the 2nd generation of tyrocidine chemoenzymatic approach utilizing a linear peptide immobilized on the solid-phase support poly(ethylene glycol) acrylate (PEGA) with a biomimetic linker to imitate peptide-S-PCP [46], which not only employed in the efficient cyclization of tyrocidine A (**1**) but also worked on

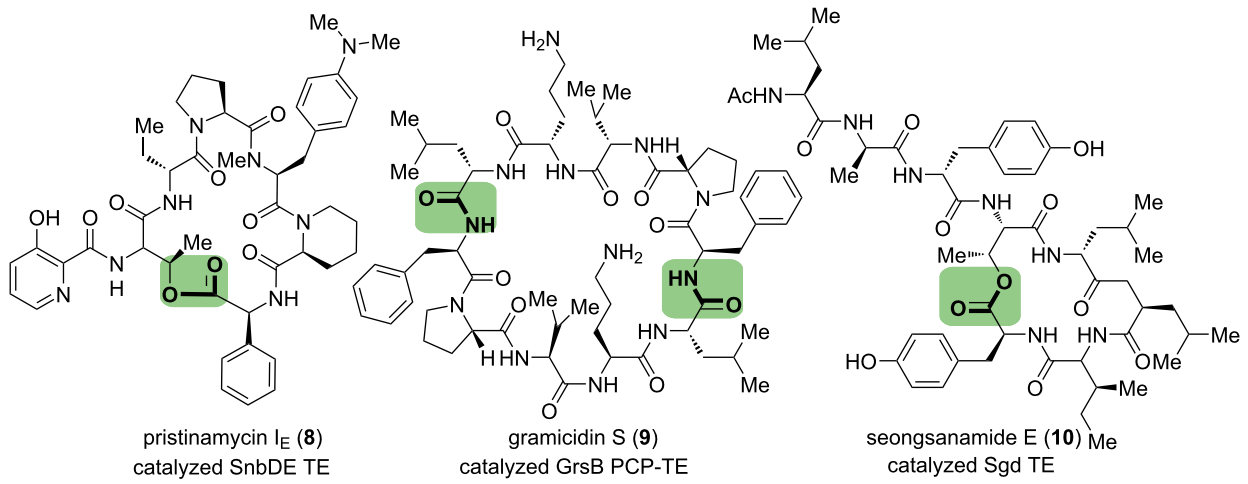



**Scheme 2:** Chemoenzymatic synthesis of tyrocidine A and its analogs. (a) First-gen chemoenzymatic synthesis of tyrocidine A. (b) The analogs preparation catalyzed by TycC TE. (c) Second-gen chemoenzymatic synthesis of tyrocidine A.

hundreds of other linear substrates, some of which exhibited broad-spectrum activity against both Gram-positive and Gram-negative organisms. By combining natural-product biosynthesis and combinatorial solid-phase chemistry, this strategy has expanded the sequence space of macrocyclic peptides significantly (Scheme 2c).

## The daptomycins

The calcium-dependent antibiotic (CDA, **11**), daptomycin (**12**), and A54145 are acidic lipopeptides isolated from *Streptomycetes*, which produce over 67% of naturally occurring antibiotics [47]. Notably, daptomycin, branded as Cubicin, was approved by the FDA as a last-resort antibiotic in 2003 for the treatment of infections caused by numerous Gram-positive bacterial strains [48], including methicillin-resistant *S. aureus* (MRSA) and vancomycin-resistant *S. aureus* (VRSA). However, the recent discovery of daptomycin-resistant *Enterococcus* and *S. aureus* provided the impetus to develop novel derivatives that enable more comprehensive structure–activity relationship (SAR) and resistance mechanism studies. Multiple approaches have been developed to address this challenge to produce daptomycin and its derivatives. These approaches include biosynthetic [49], chemoenzymatic [50], solid-phase [51], and solution-phase methods [52], but most only encompass modifications of the lipid chain and specific amino acid mutations.




**Scheme 3:** Representative examples of NAC-activated thioesters-mediated biocatalytic macrolactamization.

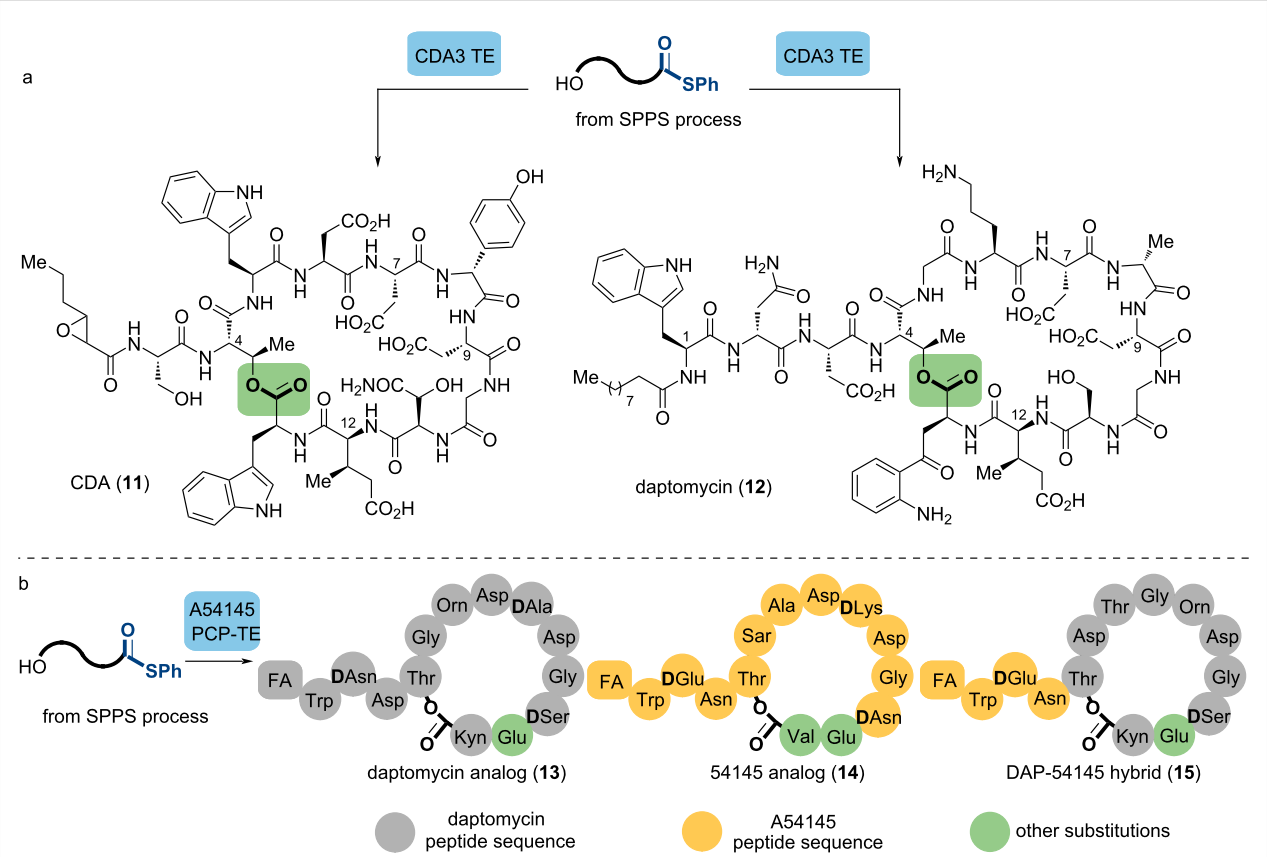
Learning from the biosynthesis of these acidic lipopeptides, Marahiel and co-workers accomplished a chemoenzymatic synthesis of the calcium-dependent antibiotic (CDA,**11**) utilizing CDA3 TE, a cyclase derived from CDA synthetase [53]. To simulate the native peptide-S-PCP substrate, they evaluated four leaving groups: SNAC, coenzyme A (CoA), phosphopantetheine, and thiophenol. The thiophenol thioesters exhibited the highest cyclization rates, suggesting that chemical reactivity precedes cofactor recognition [54]. Moreover, CDA3 TE had a broad substrate spectrum, even indicating activity to cyclize daptomycin and its analogs, resulting in daptomycin (**12**) formation with a ratio of cyclization to hydrolysis of 3:1 (Scheme 4a). The significance of single amino acids for daptomycin bioactivity was evaluated, for instance, the substitution of L-3-MeGlu$_{12}$ by L-Glu$_{12}$ in Dap yielded a 7-fold increase of the MIC against *B. subtilis*.

However, this enzymatic macrocyclization can be limited by low yields due to competing hydrolysis or the fact that specific recognition elements intrinsic to cognate substrates can be needed for efficient substrate cyclization. To overcome this drawback, the Marahiel lab characterized two additional TE domains along with their associated peptidyl carrier proteins (PCPs): daptomycin and A54145 PCP-TE [55]. A series of thiophenol-activated precursors were tolerated by these enzymes to produce daptomycin derivatives, A54145 as well as hybrid molecules of the two compounds, which pushed forward the better understanding of the acidic lipopeptide structure–activity relationship (Scheme 4b).

## Surugamide B

The cyclic octapeptides surugamides were isolated from several *Streptomyces* sp. and shown to be cathepsin B inhibitors [56-58]. According to a biosynthetic viewpoint, the corresponding modules consist of four sequential NRPS genes. However, none of them contain the thioesterase domain, which is essential for late-stage cyclization [59]. In 2018, Wakimoto, Kuranaga and co-workers reported the first total synthesis of surugamide B (**16**) in 34% overall yield through the general SPPS process and the macrocyclization at the biomimetic position (**17a**), which not only alleviated the epimerization in the macrolactamization process compared to other positions, but also enabled investigation of its biosynthetic pathway [60]. They also identified a stand-alone enzyme known as SurE, which is classified as a penicillin-binding protein (PBP) family and plays a role in chain termination and macrocyclization in the biosynthesis of surugamides. This PBP-type discrete TE was utilized in the chemoenzymatic synthesis of surugamide B with corresponding peptidyl-SNAC thioester (**17b**).

Most recently, it was observed that SurE exhibits a cyclization activity against a peptide methyl ester that is feeble but readily detectable [61]. This finding indicates that SurE has a high tolerance for leaving groups. In the light of this property, Wakimoto, Matsuda, and co-workers discovered that ethylene glycol (EG) can act as a linker on the resin before the SPPS, as well as a leaving group in further enzymatic cyclization (**17c**). Utilizing this approach, the overall yield of surugamide B (82.8%) was greater than the SNAC-based peptides with the same sequence used in the previous study (30%, Scheme 5a). In addition to investigating the high tolerance for different ring sizes, the sequential explorations of homologous wild-type enzymes and rational protein engineering have broadened the scope of the enzymatic macrolactamization [62]. Antibiotics, wollamide B1 (**18**) and desprenylagaramide (**19**), were prepared efficiently using the same manner catalyzed by homolog WolJ [63] and




**Scheme 4:** Chemoenzymatic synthesis of CDA, daptomycin and their analogs. (a) Biocatalytic macrocyclization of CDA and daptomycin mediated by thiophenol-activated esters. (b) Macrocyclization of daptomycin and A54145 catalyzed by A54145 PCP-TE.

SurE G235L (Scheme 5b). Additionally, the above mentioned type I TE, TycC TE, can also tolerate ethylene glycol as a leaving group and gave tyrocidine A (**1**) in 70% yield, indicating that this convenient bifunctional linker may have a comparable applied range to *N*-acetylcysteamine.

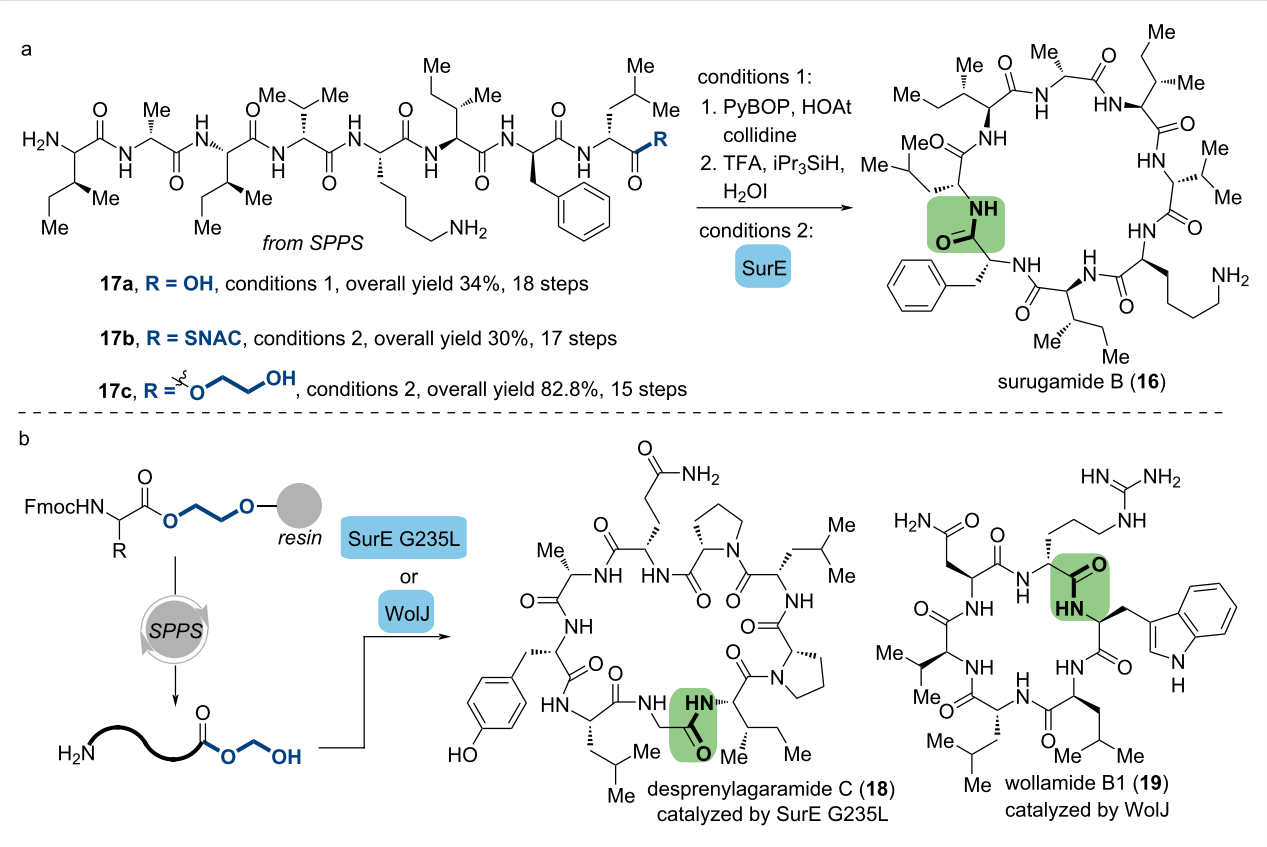
Via bioinformatics analysis, Parkinson and co-workers most recently reported the characterization of Ulm16, the PBP-TE predicted to cyclize ulleungymycin. Compared with previously studied PBP-TEs, Ulm16 showed much higher efficiency and broader substrate scope, producing a variety of ullemgymycin-like hexapeptides and also working on the cyclization of penta- and tetrapeptides [64]. The findings illustrated that PBP-type discrete TEs would become potent tools for the construction of a noncanonical macrocyclic peptides library.

## Macrocyclic polyketides and PKS/NRPS hybrids

In contrast to the aforementioned NPRS macrocycles, the synthesis of macrocylic polyketides and PKS/NRPS hybrids is more challenging due to the absence of a streamlined preparation strategy such as solid-phase peptide synthesis (SPPS). In enzymology studies, it was common to hydrolyze the cyclic natural products in order to obtain the linear predecessors. Nevertheless, this method was unable to acquire substrate derivatives, restricting the exploration of the substrate scope for TE domains. Hence, it is necessary to produce linear precursors using concise stereoselective methodologies to facilitate the investigation of biocatalytic cyclization, a crucial feature for the chemoenzymatic synthesis of macrolides and PKS/NRPS hybrids.

### The pikromycins

Methymycin (**20**) and pikromycin (**21**) are 12- and 14-membered macrolide antibiotics both isolated from *Streptomyces venezuelae* ATCC15439. The Kang lab reported the total synthesis of pikromycin and the aglycones in this family, 10-deoxymethynolide (**24**) and norbonolide (**25**), using asymmetric aldol reaction, Yamaguchi esterification, and ring-closing metathesis as key steps [65,66]. Nevertheless, the inherent complexity of these natural products demands high step counts, leading to low overall yield. According to the biosynthetic approach, these macrolides are produced by the type I PKS system, including thioesterase (TE)-catalyzed cycli-

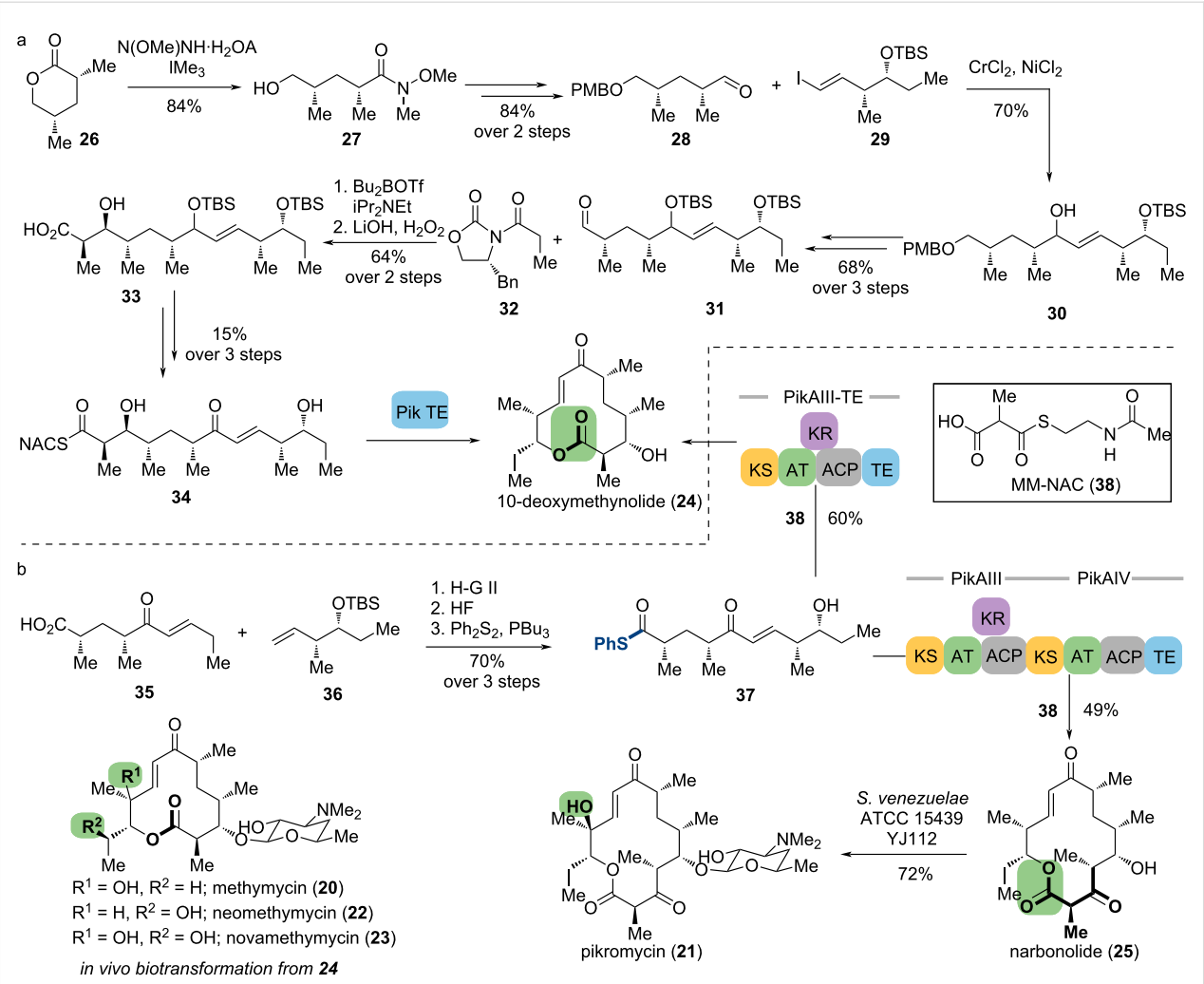



**Scheme 5:** Chemoenzymatic synthesis of surugamide B and related natural products. (a) Three synthetic strategies of surugamide B. (b) Biocatalytic macrocyclization of desprenylagaramide C and wollamide B1 mediated by ethylene glycol-linked esters.

zation of the linear hexa- and heptaketide intermediates, post-PKS oxidation, and glycosylation [67]. Cane and co-workers reported that Pik TE, the TE domain found in pikromycin biosynthesis, has broad substrate tolerance for chain length variation, suggesting this enzyme can be a potent tool in the chemoenzymatic synthesis of macrolides [68].

In 2005, Sherman and co-workers accomplished the total synthesis of 10-deoxymethynolide (**24**), via a late-stage TE-catalyzed marcolactonization [69]. The synthesis of linear peptide **34** commenced with the lactone opening of **26** to afford Weinreb amide **27**. Following primary alcohol protection and amide reduction, the aldehyde **28** was coupled with iodide **29** to afford **30** via Nozaki–Hiyama–Kishi coupling, which was then transformed into aldehyde **31** through several protecting group adjustments and the corresponding alcohol and Ley oxidation. After the preparation of **33** using Evans *syn*-aldol condensation as a critical step, **34** was produced by thioester formation, desilylation, and allylic oxidation. Incubating **34** with the purified Pik TE afforded 10-deoxymethynolide (**24**) as the exclusive product (Scheme 6a). Notably, using the corresponding C7-hydroxy NAC-hexaketide as substrate in this transformation resulted in exclusive hydrolysis to produce a seco-acid, indicating that Pik TE exhibits sensitivity to minor functional group changes of its natural substrates.

To increase the efficiency of pikromycins preparation, the Sherman lab developed a chemoenzymatic synthesis method through sequential propionate extension and marcocyclization catalyzed by fused PikAIII-TE and PikAIII-PikAIV modules. Based on this study, they established a preparative-scale approach toward the pikromycins family and their aglycones in 2013 [70]. The preparation of activated pentaketides (**37**) using asymmetric α-alkylation and cross metathesis as key reactions reduced the step counts from 14 to 11 steps. Replacing the extender unit from methylmalonyl-coenzyme A to its mimic MM-NAC (**38**) [68], and the substrate from NAC thioester to thiophenol-activated **37**, the PKS-mediated conversion proceeded with modest yield to 10-deoxymethynolide (60% yield) and acetylnarbonolide (49% yield) at preparative scale (>1 mmol), generating about 250 mg of both macrolactones. Using engineered variants of *S. venezuelae* ATCC 15439 designated strains DHS200141 [71] and YJ11242 [72], **24** and **25** were transformed to the corresponding macrolides through whole cell biotransformation to append D-desosamine and perform C–H oxidation(s) by the PikC monooxygenase



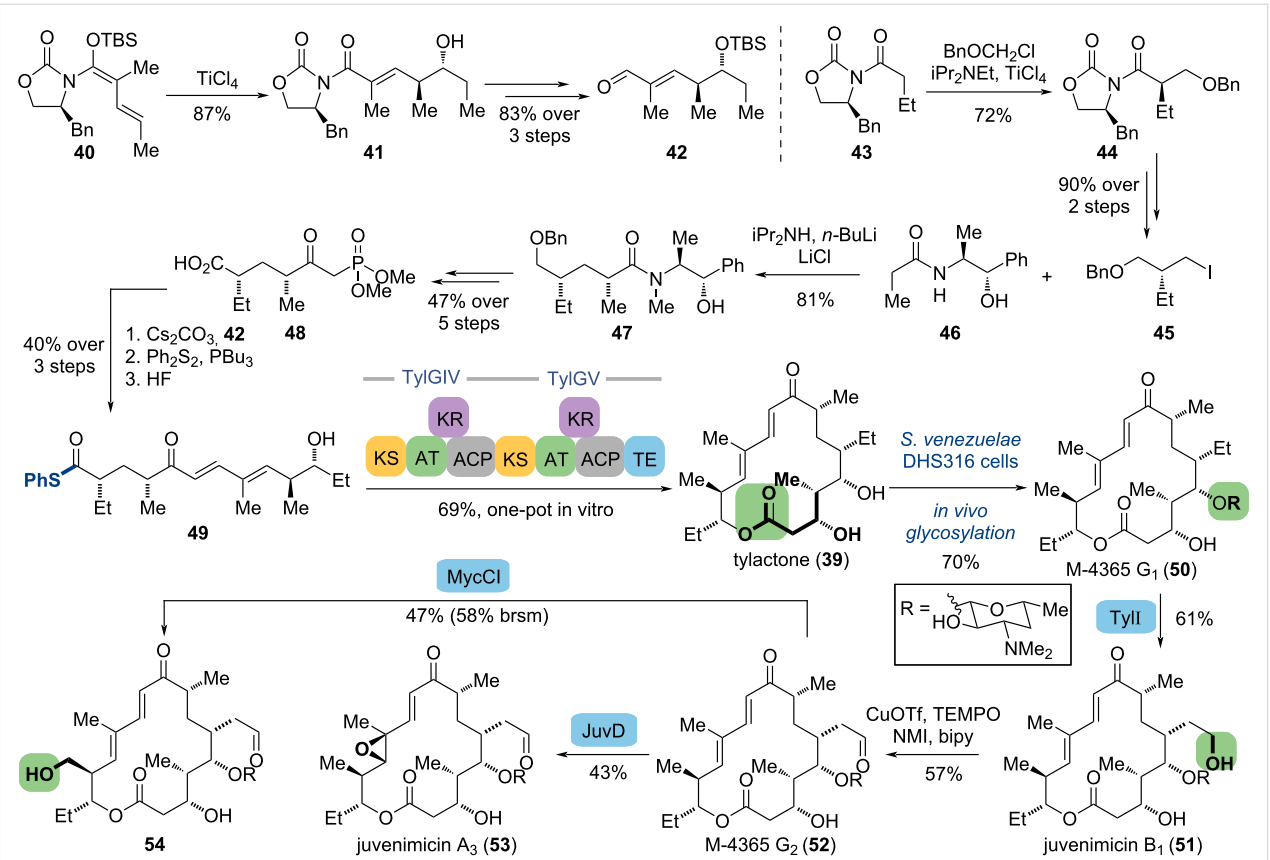

**Scheme 6:** Chemoenzymatic synthesis of the pikromycins. (a) Macrocyclization of 10-deoxymethynolide catalyzed by Pik TE. (b) Biocatalytic synthesis of pikromycin, methymycin, and related natural products.

(Scheme 6b). In contrast to Kang's chemical synthesis route, this biotransformation provided a more efficient and productive strategy for the desoaminylation of macrolide aglycones. Combining in vitro and in vivo enzymatic reactions together, this chemoenzymatic platform exhibits the potential to access a broader range of unnatural macrolides with similar skeletons.

## The juevnimicins

Juvenimicins belong to a family of broad-spectrum macrolide antibiotics [73], playing an essential role in veterinary medicine, isolated from *Micromonospora chalcea* and *Micromonospora capillata*. They contain a 16-membered macrolide aglycone, tylactone (**39**), and a dimethylamino sugar, which are synthetically challenging. Therefore, the synthesis and evaluation of tylosin-related macrolides are hot topics in medicinal chemistry [66,74]. With the experience in pikromycins synthesis, Sherman and co-workers investigated the capabilities of two terminal polyketide synthases (PKSs) in juvenimicin biosynthesis in 2017 [75], which presented a chance to accomplish the chemoenzymatic total syntheses of tylactone and the juvenimicins (Scheme 7).

To generate an appropriately activated tylactone hexaketide intermediate **49**, two key fragments, aldehyde **42** and phosphonate **48**, were synthesized, employing Evans' vinylogous aldol and Myers' auxiliary-mediated alkylation reactions as key steps. By utilizing these two fragments, a Horner–Wadsworth–Emmons olefination, followed by thioester formation and desilylation, produced several different activated tylactone hexaketides, such as NAC thioester and thiophenol ester **49**, which serves as the precursor for downstream enzymatic assembly. Thiophenol ester **49** was accepted by the two terminal of tylosin PKS modules (TylGIV and TylGV) in vitro, which are responsible for the last four carbon atoms assembly




**Scheme 7:** Chemoenzymatic synthesis of the juevnimicins.

and macrolactonization, leading to the formation of tylactone (**39**) in 69% yield. Furthermore, the *Streptomyces* strain *S. venezuelae* DHS316 [76] performed an in vivo glycosylation resulting in M-4365 $G_1$ (**50**) in 15 linear steps and 4.6% overall yield from commercial resources. With regio- and stereoselective C–H hydroxylation and epoxidation using three P450s (TylI, JuvD and MycCI) involved in the biosynthesis of several different macrolides, eight additional macrolides were achieved from **50**, including juvenimicin B1, M-4365 $G_2$, and juvenimicin $A_3$. In the light of this approach, the following bioactive assay demonstrated that some of them exhibit comparable activities to the clinically approved antibiotics against Gram-positive strains while also enhancing activities against Gram-negative pathogens.

## The cryptophycins

The cryptophycins are a large family of 16-membered ring depsipeptide natural products, which exhibit a potent ability to induce tubulin depolymerization [77], originally isolated from the cyanobacteria *Nostoc* sp. ATCC 53789 [78]. Notably, the cryptophycins cannot serve as substrates for *P*-glycoprotein and multiple drug resistance-associated proteins, making them attractive as chemotherapeutic options for treating vinca alkaloid- and taxol-resistant cancers [79]. Therefore, the pharmaceutical investigation of these natural products started for the first time when they were isolated in the early 1990s and has lasted until the present. A synthetic analog, cryptophycin 52, completed phase I clinical trials for the treatment of non-small-cell lung cancer and platinum-resistant ovarian cancer, but was halted in phase II due to dose-limiting peripheral neuropathy and limited efficacy in vivo [80]. However, this family of depsipeptides remains of therapeutic significance and has recently been explored as prospective payloads for antibody-drug conjugation [81,82].

Numerous synthetic approaches have been devised to deliver the cryptophytes skeleton, indicating that the most challenging steps are the regio- and stereospecific macrocyclization and epoxidation [83]. To address these problems, in 2005, Sherman and co-workers reported a chemoenzymatic approach through the stereospecific macrocyclization based on the identification of the thioesterase domain (CrpTE) from the cryptophycin biosynthetic pathway, which demonstrated that the CrpTE has both high efficiency in generating the 16-membered depsipeptide ring and broad tolerance for structural variation [84]. To simplify the synthetic process and expedite the comprehensive

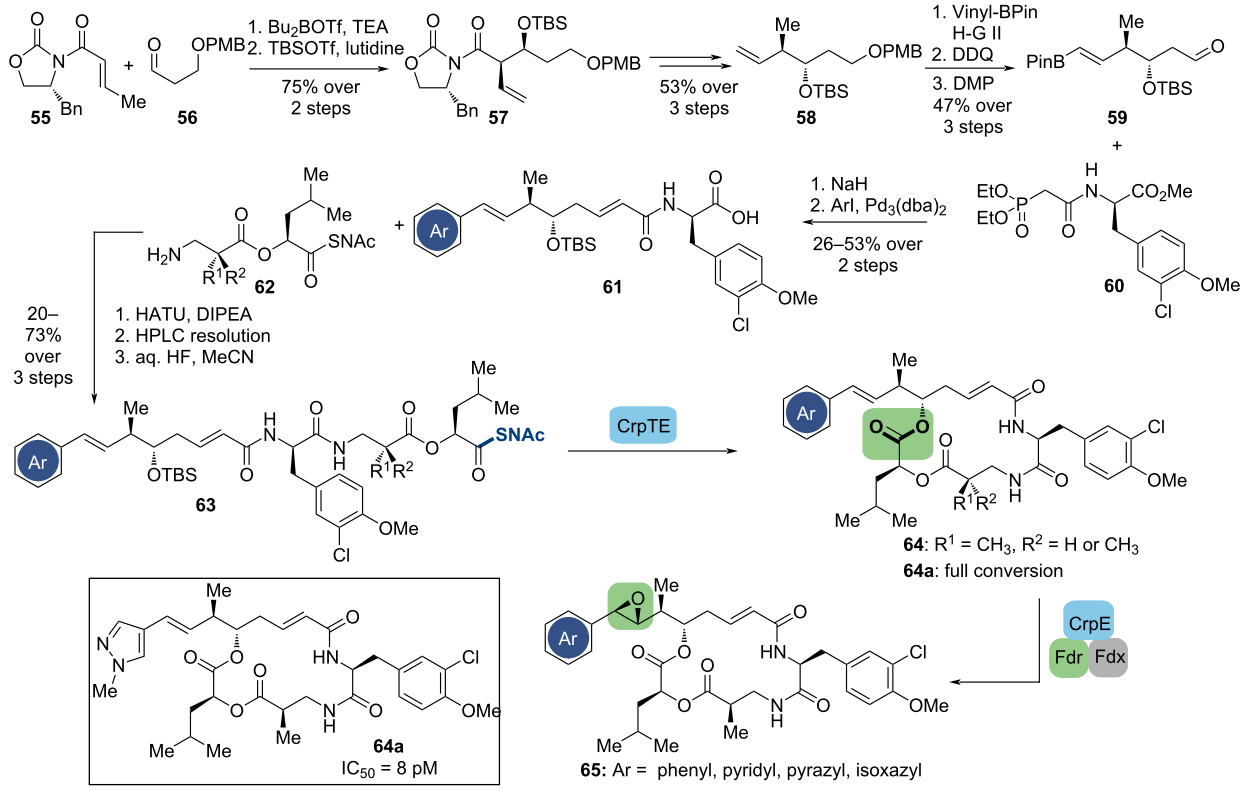
structure–activity relationship analysis, they modified the preparation of the linear chain elongation intermediate and conjugated the late-stage P450-catalyzed selective epoxidation with enzymatic macrocyclization in 2020 as shown in Scheme 8 [85].

According to their previous report [86], the production of fragments **61** was initiated by Evans' asymmetric aldol and alcohol protection to generate **57**. Six-step route transformations, including cross metathesis, afforded aldehyde **59**, which was reacted with phosphonate **60** through Horner–Wadsworth–Emmons (HWE) olefination. Afterward, the coupling of **61** and **62**, followed by removing the silyl groups, gave the desired linear precursors **63**. The investigation of the enzymatic macrocyclization suggested that CrpTE is able to accept a diverse range of heteroaromatics. For instance, incorporation of a 4-methylpyrazole ring (**64a**) showed nearly complete conversion to product with no measurable starting substrate or hydrolytic byproducts. After producing these cryptophycin analogs utilizing the CrpTE, the selective epoxidation using wild-type CrpE P450 was examined with the assistance of the spinach reductase system, which provided a series of more complex analogs. Through biological evaluation, one of the most potent cryptophycin analogs (**64a**) to date has been identified, exhibiting significant potency against HCT-116 human colorectal cancer with an $IC_{50}$ value of 8 pM [85].

## Conclusion

Macrocyclic peptides, polyketides, and their hybrids are natural products often used in different therapeutic areas. In the synthesis of these natural products and their analogs, more efficient macrocyclization strategies need to be developed to address the current issues, such as insufficient regioselectivity, intermolecular oligomerization, and the overuse of protective groups. The biosynthetic studies demonstrated that thioesterase (TE) domains exhibit a high level of chemoselectivity and regioselectivity in late-stage macrocyclizations. This review summarizes recent advances in combining thioesterase-catalyzed macrocyclization and typical chemical approaches in the rapid generation of these complex cyclic natural products and their analogs with exquisite biological activity. Moreover, multistep enzyme cascades simplify synthesis by reducing step counts, increasing yields, and minimizing waste generation as they couple different biotransformations in sequential reactions.

Although TE-mediated chemoenzymatic synthesis is becoming a prospective strategy, many challenges still need to be resolved, such as limited reaction solvents, enzyme stability,




**Scheme 8:** Chemoenzymatic synthesis of the cryptophycins.

etc. Emerging research methods on bioinformatics, computational modeling, deep learning, protein engineering, and high-throughput screening will accelerate the pace of enzyme discovery to provide a broader platform of tools for employing chemoenzymatic strategies [64,87-89]. More chemoenzymatic approaches involving TE-catalyzed macrocyclization will keep expanding in scope and depth to explore previously inaccessible chemical space for discovering important therapeutically active natural product drug leads.

## Funding

This work is financially supported by the National Natural Science Foundation of China (22301041), the Natural Science Foundation of Shanghai Municipality (23ZR1412900), and the start-up funding of Fudan University.

## Author Contributions

Senze Qiao: writing – original draft; writing – review & editing. Zhongyu Cheng: writing – review & editing. Fuzhuo Li: supervision; writing – original draft; writing – review & editing.

## ORCID® iDs

Fuzhuo Li - https://orcid.org/0000-0002-2710-7743

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

1. Muttenthaler, M.; King, G. F.; Adams, D. J.; Alewood, P. F. *Nat. Rev. Drug Discovery* **2021,** *20,* 309–325. doi:10.1038/s41573-020-00135-8
2. Zorzi, A.; Deyle, K.; Heinis, C. *Curr. Opin. Chem. Biol.* **2017,** *38,* 24–29. doi:10.1016/j.cbpa.2017.02.006
3. Kanoh, S.; Rubin, B. K. *Clin. Microbiol. Rev.* **2010,** *23,* 590–615. doi:10.1128/cmr.00078-09
4. Marsault, E.; Peterson, M. L. *J. Med. Chem.* **2011,** *54,* 1961–2004. doi:10.1021/jm1012374
5. Itoh, H.; Inoue, M. *Chem. Rev.* **2019,** *119,* 10002–10031. doi:10.1021/acs.chemrev.9b00063
6. Parenty, A.; Moreau, X.; Niel, G.; Campagne, J.-M. *Chem. Rev.* **2013,** *113,* PR1–PR40. doi:10.1021/cr300129n
7. Martí-Centelles, V.; Pandey, M. D.; Burguete, M. I.; Luis, S. V. *Chem. Rev.* **2015,** *115,* 8736–8834. doi:10.1021/acs.chemrev.5b00056
8. White, C. J.; Yudin, A. K. *Nat. Chem.* **2011,** *3,* 509–524. doi:10.1038/nchem.1062
9. Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006,** *106,* 3468–3496. doi:10.1021/cr0503097
10. Hertweck, C. *Angew. Chem., Int. Ed.* **2009,** *48,* 4688–4716. doi:10.1002/anie.200806121
11. Strieker, M.; Tanović, A.; Marahiel, M. A. *Curr. Opin. Struct. Biol.* **2010,** *20,* 234–240. doi:10.1016/j.sbi.2010.01.009
12. Horsman, M. E.; Hari, T. P. A.; Boddy, C. N. *Nat. Prod. Rep.* **2016,** *33,* 183–202. doi:10.1039/c4np00148f
13. Kopp, F.; Marahiel, M. A. *Nat. Prod. Rep.* **2007,** *24,* 735–749. doi:10.1039/b613652b
14. Walsh, C. T. *Nat. Prod. Rep.* **2016,** *33,* 127–135. doi:10.1039/c5np00035a
15. Schwarzer, D.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **2001,** *8,* 997–1010. doi:10.1016/s1074-5521(01)00068-0
16. Chen, X.-P.; Shi, T.; Wang, X.-L.; Wang, J.; Chen, Q.; Bai, L.; Zhao, Y.-L. *ACS Catal.* **2016,** *6,* 4369–4378. doi:10.1021/acscatal.6b01154
17. Liu, L.; Tao, W.; Bai, L.; Kim, E.-S.; Zhao, Y.-L.; Shi, T. *Catal. Sci. Technol.* **2019,** *9,* 6391–6403. doi:10.1039/c9cy01355e
18. Yeh, E.; Lin, H.; Clugston, S. L.; Kohli, R. M.; Walsh, C. T. *Chem. Biol.* **2004,** *11,* 1573–1582. doi:10.1016/j.chembiol.2004.09.003
19. Li, J.; Amatuni, A.; Renata, H. *Curr. Opin. Chem. Biol.* **2020,** *55,* 111–118. doi:10.1016/j.cbpa.2020.01.005
20. Stout, C. N.; Wasfy, N. M.; Chen, F.; Renata, H. *J. Am. Chem. Soc.* **2023,** *145,* 18161–18181. doi:10.1021/jacs.3c03422
21. Chakrabarty, S.; Romero, E. O.; Pyser, J. B.; Yazarians, J. A.; Narayan, A. R. H. *Acc. Chem. Res.* **2021,** *54,* 1374–1384. doi:10.1021/acs.accounts.0c00810
22. Li, F.; Deng, H.; Renata, H. *Nat. Synth.* **2023,** *2,* 708–718. doi:10.1038/s44160-023-00358-8
23. Kopp, F.; Marahiel, M. A. *Curr. Opin. Biotechnol.* **2007,** *18,* 513–520. doi:10.1016/j.copbio.2007.09.009
24. Aggarwal, R.; Caffrey, P.; Leadlay, P. F.; Smith, C. J.; Staunton, J. *J. Chem. Soc., Chem. Commun.* **1995,** 1519–1520. doi:10.1039/c39950001519
25. Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. *Chem. Biol.* **1999,** *6,* 117–125. doi:10.1016/s1074-5521(99)80008-8
26. Merrifield, R. B. *J. Am. Chem. Soc.* **1963,** *85,* 2149–2154. doi:10.1021/ja00897a025
27. Hotchkiss, R. D.; Dubos, R. J. *J. Biol. Chem.* **1941,** *141,* 155–162. doi:10.1016/s0021-9258(18)72830-5
28. Ruttenberg, M. A.; Mach, B. *Biochemistry* **1966,** *5,* 2864–2869. doi:10.1021/bi00873a013
29. Shai, Y. *Pept. Sci.* **2002,** *66,* 236–248. doi:10.1002/bip.10260
30. Mootz, H. D.; Marahiel, M. A. *J. Bacteriol.* **1997,** *179,* 6843–6850. doi:10.1128/jb.179.21.6843-6850.1997
31. Ohno, M.; Izumiya, N. *J. Am. Chem. Soc.* **1966,** *88,* 376–378. doi:10.1021/ja00954a046
32. Okamoto, K.; Yonezawa, H.; Izumiya, N. *J. Chromatogr.* **1974,** *92,* 147–156. doi:10.1016/s0021-9673(00)97199-1
33. Okamoto, K.; Nonaka, K.; Izumiya, N. *Bull. Chem. Soc. Jpn.* **1977,** *50,* 231–236. doi:10.1246/bcsj.50.231
34. Ösapay, G.; Profit, A.; Taylor, J. W. *Tetrahedron Lett.* **1990,** *31,* 6121–6124. doi:10.1016/s0040-4039(00)97003-5
35. Hu, H.; Xue, J.; Swarts, B. M.; Wang, Q.; Wu, Q.; Guo, Z. *J. Med. Chem.* **2009,** *52,* 2052–2059. doi:10.1021/jm801577r
36. Trauger, J. W.; Kohli, R. M.; Mootz, H. D.; Marahiel, M. A.; Walsh, C. T. *Nature* **2000,** *407,* 215–218. doi:10.1038/35025116
37. Kohli, R. M.; Trauger, J. W.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T. *Biochemistry* **2001,** *40,* 7099–7108. doi:10.1021/bi010036j
38. Trauger, J. W.; Kohli, R. M.; Walsh, C. T. *Biochemistry* **2001,** *40,* 7092–7098. doi:10.1021/bi010035r
39. Lin, H.; Walsh, C. T. *J. Am. Chem. Soc.* **2004,** *126,* 13998–14003. doi:10.1021/ja045147v
40. Wagner, B.; Sieber, S. A.; Baumann, M.; Marahiel, M. A. *ChemBioChem* **2006,** *7,* 595–597. doi:10.1002/cbic.200500495

41. Mahlert, C.; Sieber, S. A.; Grünewald, J.; Marahiel, M. A. *J. Am. Chem. Soc.* **2005,** *127,* 9571–9580. doi:10.1021/ja051254t
42. Heberlig, G. W.; Boddy, C. N. *J. Nat. Prod.* **2020,** *83,* 1990–1997. doi:10.1021/acs.jnatprod.0c00333
43. Brazeau-Henrie, J. T.; Paquette, A. R.; O'Rourke, A. Q.; Darnowski, M. G.; Boddy, C. N. *Org. Lett.* **2022,** *24,* 6369–6373. doi:10.1021/acs.orglett.2c02271
44. Hoyer, K. M.; Mahlert, C.; Marahiel, M. A. *Chem. Biol.* **2007,** *14,* 13–22. doi:10.1016/j.chembiol.2006.10.011
45. Teruya, K.; Tanaka, T.; Kawakami, T.; Akaji, K.; Aimoto, S. *J. Pept. Sci.* **2012,** *18,* 669–677. doi:10.1002/psc.2452
46. Kohli, R. M.; Walsh, C. T.; Burkart, M. D. *Nature* **2002,** *418,* 658–661. doi:10.1038/nature00907
47. Baltz, R. H.; Miao, V.; Wrigley, S. K. *Nat. Prod. Rep.* **2005,** *22,* 717–741. doi:10.1039/b416648p
48. Raja, A.; LaBonte, J.; Lebbos, J.; Kirkpatrick, P. *Nat. Rev. Drug Discovery* **2003,** *2,* 943–944. doi:10.1038/nrd1258
49. Robbel, L.; Marahiel, M. A. *J. Biol. Chem.* **2010,** *285,* 27501–27508. doi:10.1074/jbc.r110.128181
50. Cottagnoud, P.; Pfister, M.; Acosta, F.; Cottagnoud, M.; Flatz, L.; Kühn, F.; Müller, H.-P.; Stucki, A. *Antimicrob. Agents Chemother.* **2004,** *48,* 3928–3933. doi:10.1128/aac.48.10.3928-3933.2004
51. Barnawi, G.; Noden, M.; Taylor, R.; Lohani, C.; Beriashvili, D.; Palmer, M.; Taylor, S. D. *Pept. Sci.* **2019,** *111,* e23094. doi:10.1002/bip.23094
52. Lam, H. Y.; Zhang, Y.; Liu, H.; Xu, J.; Wong, C. T. T.; Xu, C.; Li, X. *J. Am. Chem. Soc.* **2013,** *135,* 6272–6279. doi:10.1021/ja4012468
53. Grünewald, J.; Sieber, S. A.; Mahlert, C.; Linne, U.; Marahiel, M. A. *J. Am. Chem. Soc.* **2004,** *126,* 17025–17031. doi:10.1021/ja045455t
54. Grünewald, J.; Sieber, S. A.; Marahiel, M. A. *Biochemistry* **2004,** *43,* 2915–2925. doi:10.1021/bi036140d
55. Kopp, F.; Grünewald, J.; Mahlert, C.; Marahiel, M. A. *Biochemistry* **2006,** *45,* 10474–10481. doi:10.1021/bi0609422
56. Takada, K.; Ninomiya, A.; Naruse, M.; Sun, Y.; Miyazaki, M.; Nogi, Y.; Okada, S.; Matsunaga, S. *J. Org. Chem.* **2013,** *78,* 6746–6750. doi:10.1021/jo400708u
57. Mohimani, H.; Gurevich, A.; Mikheenko, A.; Garg, N.; Nothias, L.-F.; Ninomiya, A.; Takada, K.; Dorrestein, P. C.; Pevzner, P. A. *Nat. Chem. Biol.* **2017,** *13,* 30–37. doi:10.1038/nchembio.2219
58. Xu, F.; Nazari, B.; Moon, K.; Bushin, L. B.; Seyedsayamdost, M. R. *J. Am. Chem. Soc.* **2017,** *139,* 9203–9212. doi:10.1021/jacs.7b02716
59. Ninomiya, A.; Katsuyama, Y.; Kuranaga, T.; Miyazaki, M.; Nogi, Y.; Okada, S.; Wakimoto, T.; Ohnishi, Y.; Matsunaga, S.; Takada, K. *ChemBioChem* **2016,** *17,* 1709–1712. doi:10.1002/cbic.201600350
60. Kuranaga, T.; Matsuda, K.; Sano, A.; Kobayashi, M.; Ninomiya, A.; Takada, K.; Matsunaga, S.; Wakimoto, T. *Angew. Chem., Int. Ed.* **2018,** *57,* 9447–9451. doi:10.1002/anie.201805541
61. Matsuda, K.; Zhai, R.; Mori, T.; Kobayashi, M.; Sano, A.; Abe, I.; Wakimoto, T. *Nat. Catal.* **2020,** *3,* 507–515. doi:10.1038/s41929-020-0456-7
62. Kobayashi, M.; Fujita, K.; Matsuda, K.; Wakimoto, T. *J. Am. Chem. Soc.* **2023,** *145,* 3270–3275. doi:10.1021/jacs.2c11082
63. Booth, T. J.; Bozhüyük, K. A. J.; Liston, J. D.; Batey, S. F. D.; Lacey, E.; Wilkinson, B. *Nat. Commun.* **2022,** *13,* 3498. doi:10.1038/s41467-022-30950-z
64. Budimir, Z. L.; Patel, R. S.; Eggly, A.; Evans, C. N.; Rondon-Cordero, H. M.; Adams, J. J.; Das, C.; Parkinson, E. I. *Nat. Chem. Biol.* **2024,** *20,* 120–128. doi:10.1038/s41589-023-01495-z
65. Xuan, R.; Oh, H.-S.; Lee, Y.; Kang, H.-Y. *J. Org. Chem.* **2008,** *73,* 1456–1461. doi:10.1021/jo702384d
66. Oh, H.-S.; Kang, H.-Y. *J. Org. Chem.* **2012,** *77,* 1125–1130. doi:10.1021/jo201158q
67. Xue, Y.; Wilson, D.; Zhao, L.; Liu, H.-w.; Sherman, D. H. *Chem. Biol.* **1998,** *5,* 661–667. doi:10.1016/s1074-5521(98)90293-9
68. Pohl, N. L.; Gokhale, R. S.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1998,** *120,* 11206–11207. doi:10.1021/ja9830290
69. Aldrich, C. C.; Venkatraman, L.; Sherman, D. H.; Fecik, R. A. *J. Am. Chem. Soc.* **2005,** *127,* 8910–8911. doi:10.1021/ja0504340
70. Hansen, D. A.; Rath, C. M.; Eisman, E. B.; Narayan, A. R. H.; Kittendorf, J. D.; Mortison, J. D.; Yoon, Y. J.; Sherman, D. H. *J. Am. Chem. Soc.* **2013,** *135,* 11232–11238. doi:10.1021/ja404134f
71. Jung, W. S.; Lee, S. K.; Hong, J. S. J.; Park, S. R.; Jeong, S. J.; Han, A. R.; Sohng, J. K.; Kim, B. G.; Choi, C. Y.; Sherman, D. H.; Yoon, Y. J. *Appl. Microbiol. Biotechnol.* **2006,** *72,* 763–769. doi:10.1007/s00253-006-0318-5
72. Jung, W. S.; Jeong, S. J.; Park, S. R.; Choi, C. Y.; Park, B. C.; Park, J. W.; Yoon, Y. J. *Appl. Environ. Microbiol.* **2008,** *74,* 1972–1979. doi:10.1128/aem.02296-07
73. Kishi, T.; Harada, S.; Yamana, H.; Miyake, A. *J. Antibiot.* **1976,** *29,* 1171–1181. doi:10.7164/antibiotics.29.1171
74. Matsumoto, T.; Maeta, H.; Suzuki, K.; Gen-ichi Tsuchihashi, I. *Tetrahedron Lett.* **1988,** *29,* 3575–3578. doi:10.1016/0040-4039(88)85296-1
75. Lowell, A. N.; DeMars, M. D., II; Slocum, S. T.; Yu, F.; Anand, K.; Chemler, J. A.; Korakavi, N.; Priessnitz, J. K.; Park, S. R.; Koch, A. A.; Schultz, P. J.; Sherman, D. H. *J. Am. Chem. Soc.* **2017,** *139,* 7913–7920. doi:10.1021/jacs.7b02875
76. DeMars, M. D., II; Sheng, F.; Park, S. R.; Lowell, A. N.; Podust, L. M.; Sherman, D. H. *ACS Chem. Biol.* **2016,** *11,* 2642–2654. doi:10.1021/acschembio.6b00479
77. Smith, C. D.; Zhang, X.; Mooberry, S. L.; Patterson, G. M. L.; Moore, R. E. *Cancer Res.* **1994,** *54,* 3779–3784.
78. Schwartz, R. E.; Hirsch, C. F.; Sesin, D. F.; Flor, J. E.; Chartrain, M.; Fromtling, R. E.; Harris, G. H.; Salvatore, M. J.; Liesch, J. M.; Yudin, K. *J. Ind. Microbiol.* **1990,** *5,* 113–123. doi:10.1007/bf01573860
79. Breier, A.; Barancik, M.; Sulova, Z.; Uhrik, B. *Curr. Cancer Drug Targets* **2005,** *5,* 457–468. doi:10.2174/1568009054863636
80. D'Agostino, G.; del Campo, J.; Mellado, B.; Izquierdo, M. A.; Minarik, T.; Cirri, L.; Marini, L.; Perez-Gracia, J. L.; Scambia, G. *Int. J. Gynecol. Cancer* **2006,** *16,* 71–76. doi:10.1111/j.1525-1438.2006.00276.x
81. Su, D.; Kozak, K. R.; Sadowsky, J.; Yu, S.-F.; Fourie-O'Donohue, A.; Nelson, C.; Vandlen, R.; Ohri, R.; Liu, L.; Ng, C.; He, J.; Davis, H.; Lau, J.; Del Rosario, G.; Cosino, E.; Cruz-Chuh, J. d.; Ma, Y.; Zhang, D.; Darwish, M.; Cai, W.; Chen, C.; Zhou, H.; Lu, J.; Liu, Y.; Kaur, S.; Xu, K.; Pillow, T. H. *Bioconjugate Chem.* **2018,** *29,* 1155–1167. doi:10.1021/acs.bioconjchem.7b00785
82. Verma, V. A.; Pillow, T. H.; DePalatis, L.; Li, G.; Phillips, G. L.; Polson, A. G.; Raab, H. E.; Spencer, S.; Zheng, B. *Bioorg. Med. Chem. Lett.* **2015,** *25,* 864–868. doi:10.1016/j.bmcl.2014.12.070
83. Weiss, C.; Sammet, B.; Sewald, N. *Nat. Prod. Rep.* **2013,** *30,* 924–940. doi:10.1039/c3np70022d
84. Beck, Z. Q.; Aldrich, C. C.; Magarvey, N. A.; Georg, G. I.; Sherman, D. H. *Biochemistry* **2005,** *44,* 13457–13466. doi:10.1021/bi051140u
85. Schmidt, J. J.; Khatri, Y.; Brody, S. I.; Zhu, C.; Pietraszkiewicz, H.; Valeriote, F. A.; Sherman, D. H. *ACS Chem. Biol.* **2020,** *15,* 524–532. doi:10.1021/acschembio.9b00998

86. Bolduc, K. L.; Larsen, S. D.; Sherman, D. H. *Chem. Commun.* **2012,** *48,* 6414–6416. doi:10.1039/c2cc32417b
87. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. *Science* **2021,** *373,* 871–876. doi:10.1126/science.abj8754
88. Mazurenko, S.; Prokop, Z.; Damborsky, J. *ACS Catal.* **2020,** *10,* 1210–1223. doi:10.1021/acscatal.9b04321
89. Jacques, P.; Béchet, M.; Bigan, M.; Caly, D.; Chataigné, G.; Coutte, F.; Flahaut, C.; Heuson, E.; Leclère, V.; Lecouturier, D.; Phalip, V.; Ravallec, R.; Dhulster, P.; Froidevaux, R. *Bioprocess Biosyst. Eng.* **2017,** *40,* 161–180. doi:10.1007/s00449-016-1690-x

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.66

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Research progress on the pharmacological activity, biosynthetic pathways, and biosynthesis of crocins

Zhongwei Hua[‡1], Nan Liu[‡2] and Xiaohui Yan[*1]

Review

Open Access

Address:
[1]State Key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, No. 10 Poyang Lake Road, Tuanbo New Town, Jinghai District, Tianjin 301617, China and [2]School of Chinese Materia Medica, Tianjin University of Traditional Chinese Medicine, No. 10 Poyang Lake Road, Tuanbo New Town, Jinghai District, Tianjin 301617, China

Email:
Xiaohui Yan[*] - yanxh@tjutcm.edu.cn

* Corresponding author ‡ Equal contributors

Keywords:
biosynthetic pathway; crocetin; crocins; metabolic engineering; pharmacological activity

*Beilstein J. Org. Chem.* **2024,** *20,* 741–752.
https://doi.org/10.3762/bjoc.20.68

Received: 13 December 2023
Accepted: 22 March 2024
Published: 09 April 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: T. Awakawa



© 2024 Hua et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Crocins are water-soluble apocarotenoids isolated from the flowers of crocus and gardenia. They exhibit various pharmacological effects, including neuroprotection, anti-inflammatory properties, hepatorenal protection, and anticancer activity. They are often used as coloring and seasoning agents. Due to the limited content of crocins in plants and the high cost of chemical synthesis, the supply of crocins is insufficient to meet current demand. The biosynthetic pathways for crocins have been elucidated to date, which allows the heterologous production of these valuable compounds in microorganisms by fermentation. This review article provides a comprehensive overview of the chemistry, pharmacological activity, biosynthetic pathways, and heterologous production of crocins, aiming to lay the foundation for the large-scale production of these valuable natural products by using engineered microbial cell factories.

## Introduction

Crocins are hydrophilic apocarotenoids mainly isolated from the fruit of *Gardenia jasminoides* Ellis and the stigma tissue of *Crocus sativus* L. Crocins are the main active ingredients of *C. sativus*, a precious medicinal plant known as the "gold of spices". They are also responsible for the characteristic red color of saffron. Compounds of the crocin family are the monoglycosyl, diglycosyl, or triglycosyl products of 8,8′-diapocarotene-8,8′-dioic acid (crocetin, **1**) or derivatives thereof. In addition to *C. sativus* and *G. jasminoides*, *Buddleja officinalis* and *Buddleja davidii* from the Loganiaceae family [1,2], *Nyctanthes arbor-tristis* from the Oleaceae family [3], and *Arctium lappa* from the Compositae family also contain crocins [4]. The content of crocins varies significantly among different plant species and different organs of the same plant.

Crocins are utilized extensively as colorants and flavoring agents in the food, cosmetics, and pharmaceutical industries [5]. They exhibit a broad spectrum of pharmacological activity, including antitumor [6], anti-inflammatory [7], hepatoprotective [8], nephroprotective [9], antidepressant [10], antioxidant [11], and antidiabetic [12] properties. The Saffron Multi-glycoside Tablet (Xihonghua Zonggan Pian) ameliorates myocardial ischemia or myocardial infarction induced by coronary ligation. Although crocin and crocetin derivatives exhibit a diverse scope of biological activity, they are characterized by poor stability and low bioavailability, which has significantly impeded clinical developments. Recently, novel pharmaceutical systems, such as liposomes, microcapsules, and nanoparticles, were adopted to enhance the stability and bioavailability of these compounds [13-15].

To find a sustainable way of supplying crocins, researchers have developed various approaches to produce them, including plant cell suspension culture, heterologous biosynthesis, and total synthesis [16,17]. Crocins can be obtained from plant cell culture, but the production is prone to epigenetic silencing and toxic intermediates. The chemical synthesis of crocins is challenging due to the presence of numerous chiral centers, which complicates matching supply to demand. Engineering microbial strains for the heterologous production of rare natural products has emerged as a promising approach [18,19]. With the elucidation of the biosynthetic pathways for crocins in plants, the heterologous production of crocin and crocetin derivatives in microorganisms has been achieved by various teams.

This article comprehensively reviews the research progress on the extraction, separation, pharmacological activity, biosynthesis, and synthetic biology of crocins. The biosynthesis of crocins is depicted in detail to shed light on the efficient heterologous production of crocins in microorganisms to break the current bottleneck of a sustainable supply. It provides the basis for the further development of crocins in the food, cosmetic, and pharmaceutical industries.

# Review

## Chemical properties and distribution of crocins

More than ten crocins have been identified to date (Table 1). Crocin-I (**2a**), crocin-II (**2b**), crocin-III (**2c**), crocin-IV (**2d**), and crocin-V (**2e**) are the most common crocins, among which **2a** is

**Table 1:** Structure of crocin and crocetin derivatives. A, SG, G, GB, and GT represent the common substituents of the crocin skeleton shown in Figure 1.

| compound | $R^1$ | $R^2$ | molecular formula |
|---|---|---|---|
| crocetin (**1**) | H | H | $C_{20}H_{24}O_4$ |
| crocin-I (**2a**) | GB | GB | $C_{44}H_{64}O_{24}$ |
| crocin-II (**2b**) | G | GB | $C_{38}H_{54}O_{19}$ |
| crocin-III (**2c**) | G | G | $C_{32}H_{44}O_{14}$ |
| crocin-IV (**2d**) | H | GB | $C_{32}H_{44}O_{14}$ |
| crocin-V(**2e**) | G | H | $C_{26}H_{34}O_9$ |
| crocin-6 (**2f**) | GB | GT | $C_{50}H_{74}O_{29}$ |
| crocin-7 (**2g**) | GT | GT | $C_{56}H_{84}O_{34}$ |
| *trans*-crocin-4' (**2h**) | G | GT | $C_{44}H_{64}O_{24}$ |
| neocrocin C (**2i**) | GB | A at 4′′ | $C_{48}H_{60}O_{22}$ |
| neocrocin F (**2j**) | SG | H | $C_{43}H_{54}O_{18}$ |
| neocrocin G (**2k**) | SG | GB | $C_{55}H_{74}O_{28}$ |
| neocrocin H (**2l**) | GB | $CH_2CH_3$ | $C_{34}H_{48}O_{14}$ |
| neocrocin J (**2m**) | β-D-xylosyl-(1→6)-β-D-glucosyl | H | $C_{31}H_{42}O_{13}$ |
| 13-*cis*-crocetin (**3**) | H | H | $C_{20}H_{24}O_4$ |
| *cis*-crocin-2 (**4a**) | GB | H | $C_{32}H_{44}O_{14}$ |
| *cis*-crocin-3 (**4b**) | GB | G | $C_{38}H_{54}O_{19}$ |
| 13-*cis*-crocin (**4c**) | GB | GB | $C_{44}H_{64}O_{24}$ |
| neocrocin D (**4d**) | GB | A at 4′′ | $C_{48}H_{60}O_{22}$ |
| neocrocin E (**4e**) | A at 4′′ | GB | $C_{48}H_{60}O_{22}$ |
| 13-*cis*-crocetin-8-*O*-β-D-glucopyranosyl-8′-*O*-β-D-gentiobioside (**4f**) | H | GB | $C_{32}H_{44}O_{14}$ |
| 13-*cis*-crocetin-8-*O*-β-D-glucopyranosyl-8′-*O*-β-D-gentiobioside (**4g**) | G | GB | $C_{38}H_{54}O_{19}$ |

the most abundant in nature. Compound **2a** is a diester derivative of crocetin (**1**), with a gentiobiose unit attached to each carboxyl group. Crocin-II (**2b**) consists of one glucose and one gentiobiose unit, crocin-III (**2c**) comprises two glucose molecules, crocin-IV (**2d**) contains only one gentiobiose moiety, and crocin-V (**2e**) has only one glucose moiety (Figure 1 and Table 1). Pure crocins form reddish-brown acicular crystals with a slight odor. They are highly soluble in hot water but slightly soluble in anhydrous ethanol and ether. The presence of multiple conjugated double bonds in crocins makes them susceptible to degradation when exposed to certain conditions, such as high temperature, the presence of metal ions or light, certain pH values, etc. For example, during the harvest of *C. sativus*, a high drying temperature leads to the cleavage of the glycosidic bonds [20-22]. Crocins are stable in alkaline and neutral solutions but are labile under acidic solutions.

Crocins can be detected in the flowers, fruits, stigmas, leaves, and roots of plants. The content varies significantly among different plant species and different parts within the same plant. For instance, in *C. sativus*, crocins are predominantly accumulated in the stigma, but in *G. jasminoides,* they are primarily stored in the pulp. The traditional methods for extracting crocins include ultrasound-assisted extraction (UAE), supercritical fluid extraction, enzyme-linked extraction, and microwave-assisted extraction. Among these methods, the UAE exhibits a higher extraction yield [23,24]. Recently, Fiorito et al. developed a technique that utilizes cost-effective and environmentally friendly adsorbents for the selective adsorption of crocins from an aqueous solution of saffron pollen [25].

## Pharmacological activity of crocins

Various pharmacological studies have shown that crocins are effective against diseases of the central nervous system (neurodegenerative diseases, epilepsy, convulsion, and insomnia) and cardiovascular diseases (hypertension, hyperlipidemia, and atherosclerosis). In addition, they also have anticancer, anti-inflammatory, antioxidative, liver- and kidney-

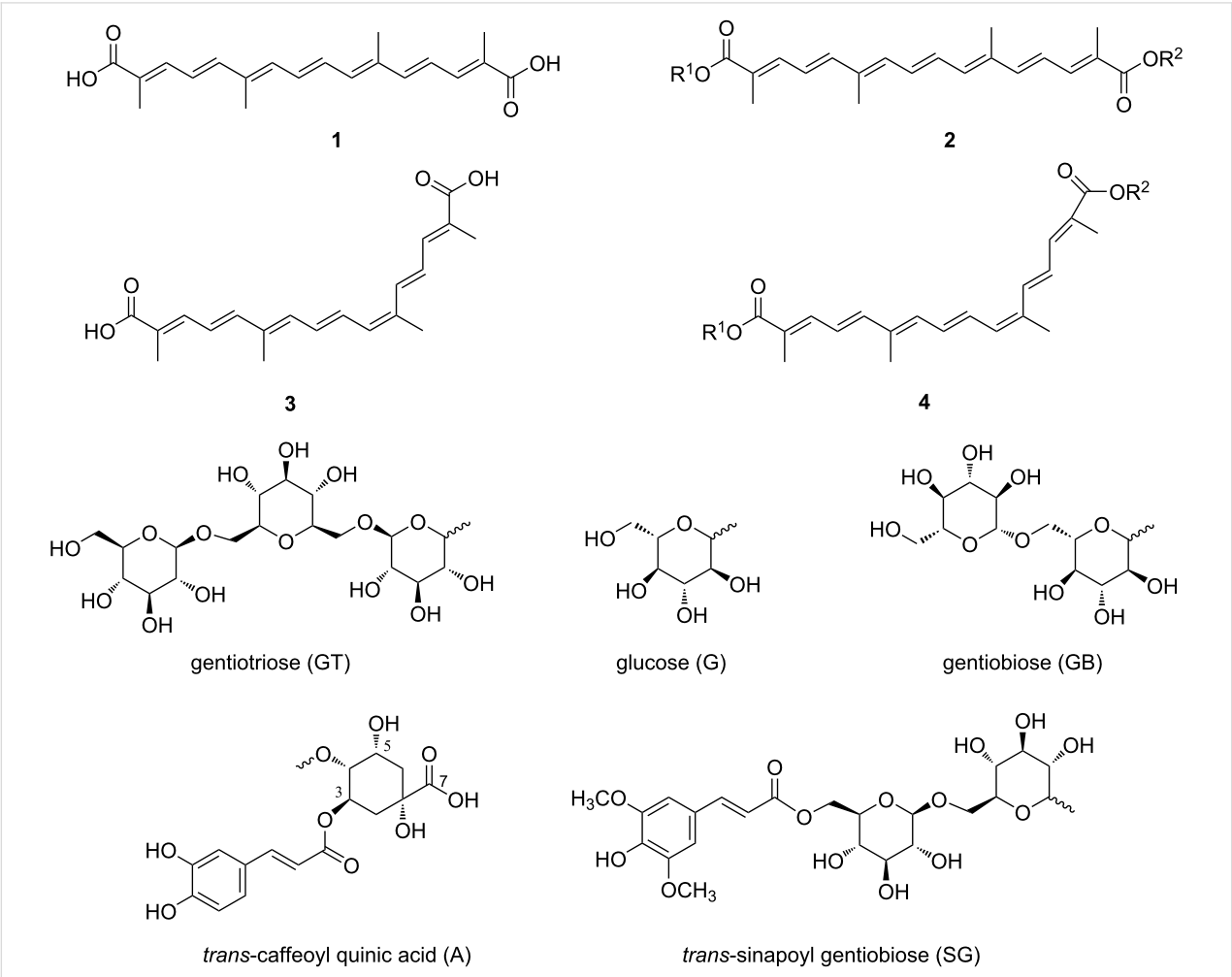



**Figure 1:** Principal structure of crocin and crocetin derivatives, including common substituents of the crocetin skeleton.

protective, antidepressant, and antidiabetic properties (Figure 2).

## Neuroprotection

Pharmacokinetic investigations have shown that crocins can be hydrolyzed to crocetin (**1**) in the gastrointestinal tract of mice. Crocetin derivatives can penetrate the blood–brain barrier to exert therapeutic effects on neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, retinal diseases, and epileptic disorders [26,27].

Four mechanisms have been reported for the treatment of Alzheimer's disease with crocins. Crocins function as antioxidants that slow down the progression of the disease by increasing the L-glutathione (GSH) level and reducing the presence of reactive oxygen species (ROSs) [28]. Crocins can inhibit the activity of acetylcholinesterase and increase the acetylcholine concentration, thus improving the learning and memory ability of the brain [29]. Moreover, crocins prevent the abnormal aggregation of amyloid β-protein (Aβ), microtubule-associated protein tau, and α-synuclein (αS), thereby alleviating the apoptosis of neuronal cells and cognitive impairment [30-33]. Lastly, the extract of *C. sativus* can also improve the chronic stress-induced loss of learning and memory by reducing the ROS level or regulating the synthesis of superoxide dismutase (SOD) and glutathione peroxidase (GPx) [34].

The current clinical treatment of epilepsy relies on the use of antiepileptic drugs. However, the occurrence of drug-resistant epilepsy is a top-priority problem [35-37]. It was found that intravenous injection of 50 μg crocin could delay epileptiform activity in mice and enhance the antiepileptic effect of diazepam. The mechanism of action involved γ-aminobutyric acid (GABA) [38-40]. The brain-derived neurotrophic factor (BDNF)–tropomyosin receptor kinase B (TrkB) pathway is closely associated with epilepsy. Crocin can increase the BDNF level in the cortex of temporal lobe epilepsy and reduce the content of tumor necrosis factor α (TNF-α) in the hippocampus, thereby playing an antiepileptic role [41,42].

## Antihypertension

Imenshahidi et al. demonstrated that crocins, safranal (**12**), and the extract of *C. sativus* can reduce the mean arterial pressure in both healthy and hypertensive rats depending on the dose [43]. In an angiotensin II-induced acute hypertension model, crocins could probably improve cardiovascular function by reducing the

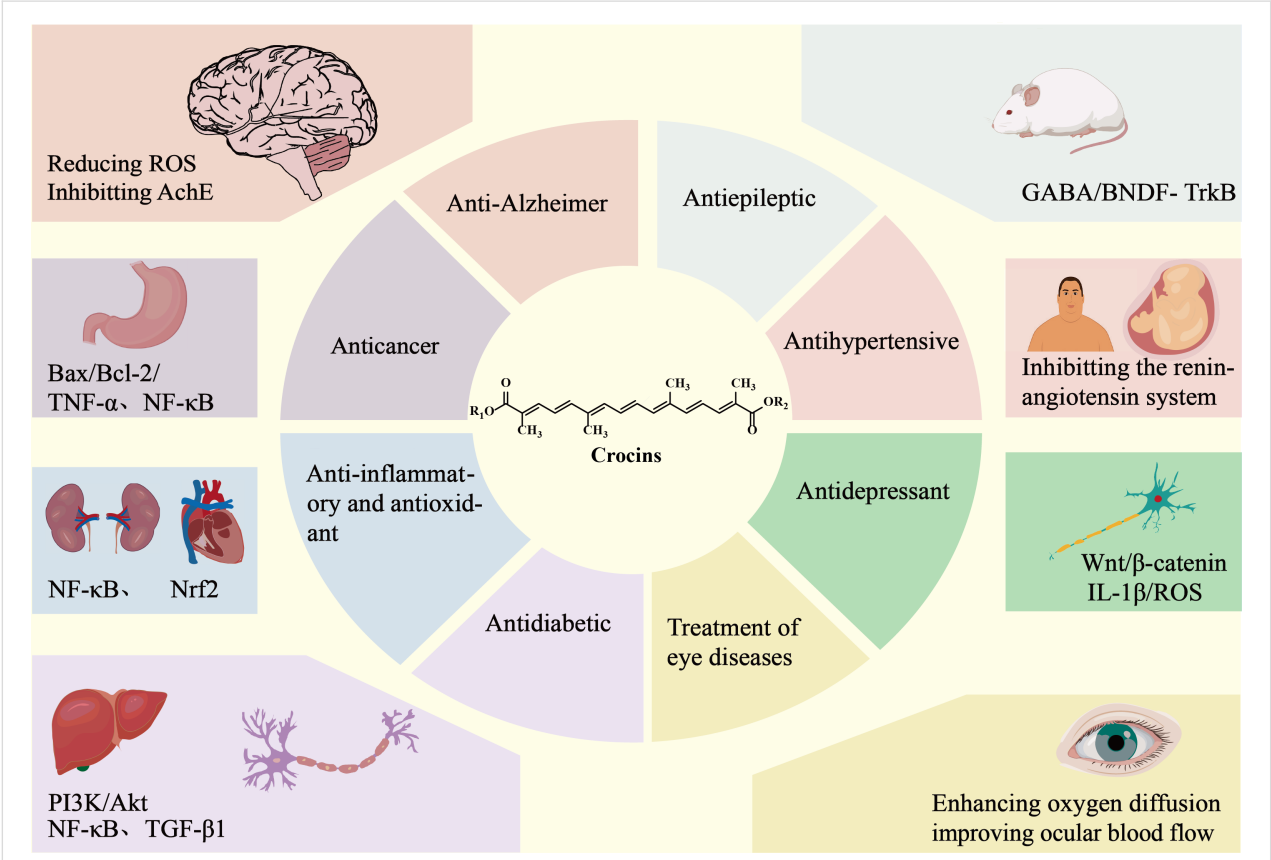



**Figure 2:** The pharmacological activity and mechanisms of action of crocins.

ROS level [44]. Chen et al. found that crocins could reduce inflammation and oxidative stress in the rat placenta and alleviate gestational hypertension. Therefore, crocins have the potential to prevent pregnancy-induced hypertension [45].

### Anticancer

Crocins exhibit potent anticancer activity against various cancer cell lines. Jiang et al. found that crocins reduce the survival and activity of cervical cancer cells [46]. Mollaei et al. observed an increased Bax/Bcl-2 ratio in crocin-treated cancer cells. Therefore, crocins are proposed to exert anticancer activity by promoting apoptosis of cancer cells [47]. This mechanism is consistent with the results of Hoshyar et al., who observed the apoptosis-promoting activity of crocins in gastric adenocarcinoma cells [48]. In addition to the cervical cancer cell lines and gastric cancer cell lines, crocins also exhibit anticancer effects on breast, prostate, liver, colon, and leukemia cancer cell lines [49-53]. Besides regulating the Bax/Bcl-2 ratio, crocins were also reported to inhibit cell invasion and metastasis [54,55].

### Anti-inflammation and antioxidation

Crocins exhibit anti-inflammation properties by scavenging free radicals and regulating the expression of antioxidant enzymes. Crocins can inhibit the NF-κB signaling pathway, thus downregulating the expression of inducible nitric oxide synthase (iNOS) and cyclooxygenase 2 (COX2) [11]. Crocins can also regulate the Nrf2 signaling pathway to protect the brain, liver, and other organs and decrease the expression of pro-inflammatory factors, such as TNF-α, prostaglandin E2 (PGE2), interleukin-1β (IL-1β), and interleukin-6 (IL-6) [56-58]. The combination of crocins with doxorubicin and curcumin also showed a good anti-inflammation effect [56,59].

### Antidepression

Alsanie et al. found that the therapeutic effect of *C. sativus* extract on mild and moderate depressive disorder is comparable to antidepressants, such as imipramine and fluoxetine [10,60,61]. The antidepressant activity of *C. sativus* extract was exerted by regulating the serotonin, norepinephrine, and dopamine levels in the brain [62-64]. Further studies confirmed that crocins are the key antidepressant agents in *C. sativus* extract [65-67]. Crocins ameliorate depression-like behavior in rats by mitigating neuroinflammation and reducing IL-1β, ROS, and malondialdehyde [10,68]. Tao et al. found that crocins exert an antidepressant effect by enhancing adult hippocampal neurogenesis (AHN) levels and activating the Wnt/β-catenin pathway [69].

### Other functions

In addition to the effects mentioned above, crocins also possess other pharmacological properties, such as antidiabetic, antiatherosclerotic, and organ-protective effects. They also have a role in immune modulation and the treatment of glaucoma and diabetic retinopathy [70-80].

### Biosynthesis

The biosynthetic pathways of crocins have recently been studied extensively. Crocin biosynthesis can be divided into three stages: 1) biosynthesis of lycopene (**5**) from simple carbon resources, 2) cleavage of lycopene (**5**), β-carotene (**6**), or zeaxanthin (**7**) by a carotenoid cleavage dioxygenase (CCD) to form crocetin aldehyde (**8**) and, after oxidation, **1**, and 3) glycosylation of **1** to generate crocins (Figure 3). Since the biosynthetic pathways of **5** in plants and microorganisms have been elucidated and reviewed, we will only elaborate the last two stages in this review [81,82].

## Biosynthesis of crocetin (**1**)

In higher plants, the cyclization of lycopene (**5**) by lycopene ε-cyclase (ε-LCYC) and β-LCYC is a critical branch point in carotenoid biosynthesis. In one route, β-LCYC catalyzes the formation of the β-rings at both ends of **5**, leading to the formation of β-carotene (**6**). In the other branch, both ε-LCYC and β-LCYC are involved in introducing one β- and one ε-ring into **5** to form α-carotene. β-Carotene (**6**) is hydroxylated by β-carotene hydroxylase (BCH) to form **7**. Zeaxanthin (**7**) possesses strong antioxidative activity and is efficacious in preventing macular degeneration, cataracts, and cardiovascular diseases [83]. Many microorganisms, such as bacteria, cyanobacteria, and microalgae, can also produce **7** [84]. The biosynthesis of **7** in bacteria is analogous to that in plants. The difference is that in bacteria, **7** is formed through a single hydroxylation step of β-carotene (**6**), but it requires two hydroxylation steps in plants [84].

In the crocin biosynthetic pathways, lycopene (**5**), β-carotene (**6**), and zeaxanthin (**7**) are cleaved by different CCDs to form crocetin dialdehyde (**8**). *Cs*CCD2 from *C. sativus* could break the 7,8 and 7',8' double bonds of **7** to yield **8** and 3-OH-β-cyclocitral (**9**). ALDH then catalyzes the conversion of **8** into crocetin (**1**). Notably, *Gj*CCD4a from *G. jasminoides* catalyzes the cleavage of zeaxanthin (**7**), lycopene (**5**), and β-carotene (**6**) to generate **8** and conversion of **8** into **1** [85].

### Biosynthesis of crocins

The late steps in crocin biosynthesis are the glycosylation of crocetin (**1**) by various UGTs [86]. The crocin biosynthetic pathways in *G. jasminoides* have been characterized in detail, and several *Gj*UGTs were identified. Attributed to different substrate specificities, the glucose or gentiobiose moieties are attached to the terminal carboxyl groups of **1** sequentially or si-

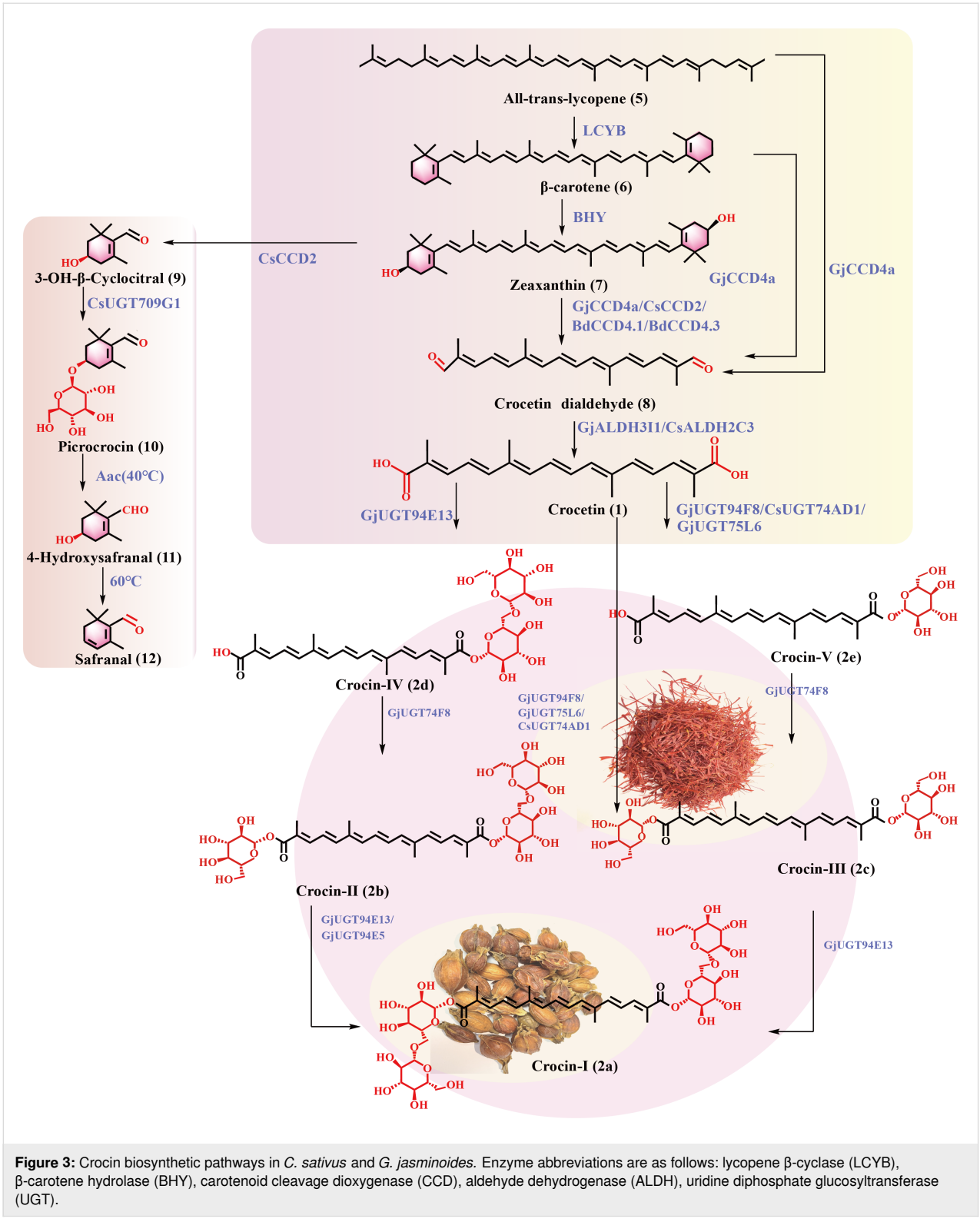



**Figure 3:** Crocin biosynthetic pathways in *C. sativus* and *G. jasminoides*. Enzyme abbreviations are as follows: lycopene β-cyclase (LCYB), β-carotene hydrolase (BHY), carotenoid cleavage dioxygenase (CCD), aldehyde dehydrogenase (ALDH), uridine diphosphate glucosyltransferase (UGT).

multaneously. Besides *Gj*UGTs, a few UGTs from *C. sativus* were also characterized. In addition to **1**, **9** is also glycosylated to form picrocrocin (**10**), which is further transformed into safranal (**12**) by β-glucosidase [87].

## Key enzymes in crocin biosynthetic pathways

**CCDs:** Crocin biosynthesis initiates with the catalytic cleavage of the C–C double bond by CCDs at various sites of the carotenoid skeleton. The cleavage generates a carbonyl group at

the end of the formed carotenoid derivatives [88]. The CCD family can be categorized into the CCD subfamily and the 9-*cis*-epoxycarotenoid dioxygenase (NCED) subfamily. Five enzymes have been identified within the CCD subfamily: CCD1, CCD2, CCD4, CCD7, and CCD8. In animals, CCDs are involved in retinoid biosynthesis. In plants, CCD1 and CCD7 cleave the 9,10-double bond of carotenoids, while NCEDs cleave the 11,12-double bond [89]. Rubio et al. identified four CCDs, *Cs*CCD1a, *Cs*CCD1b, *Cs*CCD4a, and *Cs*CCD4b, from *C. crocus*. The expression patterns of *Cs*CCD4a and *Cs*CCD4b were associated with the accumulation of β-ionone during stigma development. However, no enzyme involved in zeaxanthin (**7**) cleavage was identified. Subsequently, *Cs*CCD2 was discovered in the cytoplasm alongside CCD1. *Cs*CCD2 was expressed at the early stage of stigma development, consistent with the accumulation of **1** initially. Sequence alignment of various CCDs from *C. crocus* and *Arabidopsis thaliana* revealed that *Cs*CCD2 belongs to a distinct branch within the CCD family. In vitro experiments confirmed that *Cs*CCD2 catalyzed the cleavage of **7** between the 7,8- and 7',8'-double bonds, resulting in the formation of crocetin dialdehyde (**8**) and 3-OH-β-cyclocitral (**9**) [90]. Recently, Liang et al. designed a variant of *Cs*CCD2 that exhibits broader substrate specificity and higher catalytic efficiency through computational modeling. Heterologous expression of the S323A mutant of *C*sCCD2 in yeast led to a 4-fold enhancement in the production of **1**, with a crocetin (**1**) titer of 107 mg/L in a 5-liter fed-batch fermentation. The *Cs*CCD2(S323A) mutant could also catalyze the conversion of β-carotene (**6**) to **1** [91]. *Bd*CCD4.1 and *Bd*CCD4.3, the homologs of CsCCD2 from *B. davidii*, were discovered by bioinformatic analysis. In vitro and in vivo experiments demonstrated that these two enzymes could cleave zeaxanthin (**7**), but the enzyme activity was lower than *Cs*CCD2 [2,92].

*Bo*CCD4-3 from *Bixa orellana* was revealed to cleave various carotenoids, lycopene (**5**), β-carotene (**6**), and zeaxanthin (**7**), to form crocetin dialdehyde (**8**) by in vitro assay [93,94]. In contrast, *Cs*CCD2, *Bd*CCD4.1, and *Bd*CCD4.3 could only cleave zeaxanthin (**7**) [95]. Xu et al. identified *Gj*CCD4a from *G. jasminoides*. Through prokaryotic expression and in vitro characterization, *Gj*CCD4a was found to cleave the 7,8-/7',8'-double bonds of lycopene (**5**), β-carotene (**6**), and zeaxanthin (**7**). Further investigation revealed that *Gj*CCD4a originated from tandem duplication events during genome evolution [96]. When expressed in the callus of *Nicotiana tabacum*, *Gj*CCD4a exhibited a higher efficiency in carotenoid cleavage than *Cs*CCD4a [85].

**ALDHs:** ALDH utilizes $NAD^+$ or $NADP^+$ as cofactor to catalyze the oxidation of acetaldehyde into carboxylic acids. In 2018, Demurtas et al. identified six ALDHs from *C. sativus* [97]. Among them, *Cs*ALDH3I1 oxidized crocetin dialdehyde (**8**) to crocetin (**1**). Subcellular localization and transgene expression analysis showed that *Cs*ALDH3I1 is located in the endoplasmic reticulum [97]. Gómez-Gómez et al. screened four genes (*CsALDH3898*, *CsALDH20158*, *CsALDH54788*, and *CsALDH11367*) from *C. sativus* and showed that all four enzymes could convert crocetin dialdehyde (**8**) into crocetin (**1**) [98]. Subsequently, Tan et al. used *Cs*ALDH3 to construct an engineered *Saccharomyces cerevisiae* strain for the heterologous production of crocetin (**1**) [99]. In 2020, Xu et al. conducted a multiomics analysis of 19 ALDH genes from *G. jasminoides*. Among them, *GjALDH2C3* was highly expressed in the fruit and flower of *G. jasminoides* and showed a coexpression pattern with *Gj*CCD4a [96]. Subsequently, Diretto et al. identified *Bd*ALDH from *B. davidii* and confirmed that it was able to convert crocetin dialdehyde (**8**) into crocetin (**1**) in vitro [100]. *Bd*ALDH and *Cs*ALDH3I1 belong to the ALDH3 family, one of the most diverse groups of the ALDH superfamily.

**UGTs:** Glycosylation of crocetin (**1**) improves the solubility significantly [101,102]. Moraga et al. cloned two UGT genes, *CsUGT2* and *CsUGT3*, from the stigmas of *C. sativus* and expressed them in *E. coli*. In vitro experiments showed that *Cs*UGT2 could glycosylate crocetin (**1**), crocin-V (**2e**), and crocinIV (**2d**) [86]. Demurtas et al. revealed that *Cs*UGT74AD1 could convert crocetin (**1**) into crocin-I (**2a**) and crocin-II (**2b**). Through subcellular localization analysis, they proposed that the biosynthesis of crocins initiates in the plastids. The metabolites are transported to the endoplasmic reticulum and cytoplasm and are stored in the vacuole. Moreover, they confirmed that the ABC transporter is involved in crocin accumulation in the vacuole within the stigmas of *C. sativus* [97]. López-Jimenez et al. screened the *UGT91P3* gene from *C. sativus* and showed that coexpression of *Cs*UGT74AD1 and *Cs*UGT91P3 in *N. benthamiana* produced various crocins. These results confirmed that *Cs*UGT91P3 plays a pivotal role in the biosynthesis of crocins in *C. sativus* [103].

Nagatoshi et al. identified *Gj*UGT75L6 and *Gj*UGT94E5 from *G. jasminoides*. *Gj*UGT75L6 could catalyze the glycosylation of crocetin (**1**) to produce crocetin glycosyl esters **2c** and **2e**, and *Gj*UGT94E5 catalyzed downstream glycosylation reactions [104]. Unexpectedly, the expression patterns of these two UGTs did not align with the accumulation of crocins, suggesting that additional enzymes may participate in crocin biosynthesis in *G. jasminoides*. Xu et al. screened 173 *Gj*UGTs from *G. jasminoides*. Through coexpression analysis with *GjCCD4a*, *Gj*UGT74F8 and *Gj*UGT94E13 were identified [105]. Specifically, *Gj*UGT74F8 catalyzed the conversion of crocetin (**1**) into

crocin-V (**2e**) and crocin-III (**2c**), as well as the conversion of crocin-IV (**2d**) to crocin-II (**2b**). On the other hand, *Gj*UGT94E13 could convert crocin-II (**2b**) and crocin-III (**2c**) to crocin-I (**2a**) [105]. Diretto et al. screened *Bd*UGT74BC1 and *Bd*UGT74BC2 from *B. davidii* and found that both enzymes could glycosylate crocetin (**1**) to yield crocin-I (**2a**) and crocin-II (**2b**). They also isolated *Bd*UGT94AA3 from the flower of *B. davidii* and identified a role in the conversion of crocin-III (**2c**) into crocin-IV (**2d**) [100]. Ding et al. discovered two UGTs, Bs-GT and Bc-GTA, from *Bacillus subtilis*168 and *Bacillus cereus* WQ9-2, respectively. Bs-GT and Bc-GTA both exhibited UGT activity towards crocetin (**1**) [106].

## Heterologous production of crocins

Crocetin (**1**) and crocins have shown great value in the food, cosmetic, and pharmaceutical industries. However, the insufficient supply of crocins impedes the applications. Various strategies have been utilized to construct engineered microorganisms for the large-scale production of crocetin (**1**) and crocins (Table 2).

### Heterologous synthesis of crocetin (**1**)

Chai et al. first achieved the heterologous production of crocetin (**1**) in *S. cerevisiae* [107]. Through combining and screening the BCHs, CCDs, and ALDHs and optimizing the fermentation conditions, 1219 μg/L and 6278 μg/L of crocetin (**1**) was produced in the shake flask and a 5 L fermenter, respectively. Followed by enhancing acetyl-CoA supply, overexpressing *Ps*BCH-*Cs*CCD2, and optimizing the fermentation medium, the yield of crocetin (**1**) reached 12.43 ± 0.62 mg/L in a 5 L bioreactor [115]. While CCD has a higher catalytic activity at low temperature, the biosynthesis of zeaxanthin (**7**) is more suited to a higher temperature. As such, Liu et al. used CRISPR-Cas9 to create a temperature-responsive strain of *S. cerevisiae* for crocetin (**1**) production. By increasing the copy number of *CCD2* and *ALDH*, the conversion of zeaxanthin (**7**) was increased by 77% [108]. Wang et al. screened the *ALD* genes from *C. sativus* and *Neurospora crassa* and showed that the strain expressing *Nc*ALD8 could produce 4.42 mg/L of crocetin (**1**) [109]. In addition to the de novo biosynthesis of crocetin (**1**), Shan et al. reported the biotransformation of crocetin (**1**) from zeaxanthin (**7**). When the 3-OH-β-apo-8'-carotenoic acid (3-HACA) byproduct was eliminated by adjusting the molar ratio of *Cs*CCD2-M1 and *Syn*ALD in the system, the yield of crocetin (**1**) reached 5.92 mg/L with a conversion rate of 80.8% [110].

### Heterologous biosynthesis of crocins

The heterologous production of crocins has been achieved in various hosts, including *E. coli*, *N. benthamiana*, and *S. lycop-*

**Table 2:** Heterologous production of crocetin (**1**) and crocins.

| host | production strategy | product | titer or yield | Reference |
|---|---|---|---|---|
| *S. cerevisiae* | combining and screening of *BCH-CCD*, overexpression of ALD, Δ*CIT2*, and *Ps*BCH-*Cs*CCD2 | **1** | 12.43 ± 0.62 mg/L in fermenter | [107] |
| *S. cerevisiae* | introduction of CrtE, CrtYB, CrtI, and CrtZ in yeast, overexpression of *Cs*CCD2-*Cs*ALDH, overexpression of *Ps*BCH-*Cs*CCD2 | **1** | 139.67 ± 2.24 μg/g DCW | [108] |
| *E. coli* | introduction of *Cs*CCD2-*Gj*UGT94E15-*Gj*UGT75L6-*Trc*ALD8 | **1** | 4.42 mg/L | [109] |
| *E. coli* | adjusting the molar ratio of *Cs*CCD2 and *Syn*ALD | **1** | 1.48 mg/L/h | [110] |
| *E. coli* | introduction of *Cs*CCD2-*Gj*UGT94E15-*Gj*UGT75L6-*Trc*ALD8, YjiC, YdhE, and YojK from *B. subtilis* | **2e** | — | [109] |
| *E. coli* | introduction of *Gj*UGT94E13-*Gj*UGT74F8, overexpression of PGM and GalU | crocins | — | [105] |
| *N. benthamiana* | introduction of *Cs*CCD2L and *Pa*CrtB | crocins | 3.493 mg/g DCW | [92] |
| *N. benthamiana* | introduction of *Cs*CCD2L-*Os*BCH-*Tp*CartB | crocins | 1.61 mg/g | [85] |
| *N. tabacum* and *N. glauca* | introduction of *At*OrMut-*Br*CrtZ-*Cs*CCD2L | crocins | 400 μg/g DW | [111] |
| *S. lycopersicum* | introduction of *Cs*CCD2L-*Gs*UGT709G1-*Cs*UGT2 | crocins | 14.48 mg/g | [112] |
| *N. benthamiana* | overexpression of *Cs*ABCC4a-*Cs*CCD2 | crocins | — | [113] |
| *N. benthamiana* | overexpression of *Cs*CCD2 in the zeaxanthin (**7**)-producing strain | crocins | 800 μg/g DW | [97] |
| *N. benthamiana* | expression of *GjCCD4a-GjALDH2C3-GjUGT74F8-GjUGT94E13* multigene expression vector | crocins | 78 μg/g FW | [114] |

*ersicum*. Wang et al. introduced YjiC and YojK, two UGTs from *B. subtilis* into a crocetin (**1**)-producing *E. coli* strain to realize the biosynthesis of crocin-V (**2e**) [109]. Pu et al. biosynthesized different crocins by introducing *Gj*UGT74F8 and *Gj*UGT94E13 from *G. jasminoides* [105]. However, the titers of the produced crocins were not reported. Heterologous biosynthesis of crocins was also attempted in *Nicotiana tabacum* and *Lycopersicon esculentum* [85]. Because *Cs*CCD2L exhibited high CCD activity against zeaxanthin (**7**), Martí et al. used a virus-driven transduction system to transfer *CsCCD2L*, *PaCrtB* (the gene encoding the carotenoid synthase from *Pantea ananatis*), and *BCH2* from *C. sativus* into *N. benthamiana*. The content of crocins in the transgenic *N. benthamiana* reached 3.493 ± 0.325 mg/g dry weight [92]. Zheng et al. introduced *GjCCD4a*, *PSY*, and *BCH* into *N. benthamiana* and achieved a crocin content of 1.61 mg/g in the leaves [85]. It was reported that CrtZ catalyzes the rate-limiting reaction in zeaxanthin (**7**) biosynthesis, and the ORANGE protein regulates PSY expression to increase carotenoid accumulation. Ahrazem et al. coexpressed *AtOrMut* from *A. thaliana*, *BrCrtZ* from a *Brevundimonas* sp., and *CsCCD2L* into *N. tabacum* and achieved a crocin content of 400 μg/g (dry weight) in the engineered plant. Compared to transgenic *N. tabacum* that expressed *Cs*CCD2L alone, the crocin yield was increased 3.5-fold [111]. Ahrazem et al. also introduced *CsCCD2L*, *CsUGT709G1*, and *CsUGT2* into *L. esculentum* to produce crocins. The transgenic plants exhibited a dry weight yield of 14.48 mg/g for crocins and 2.92 mg/g for picrocrocin (**10**), representing the highest yield of crocins in transgenic plants [111,112]. Augustine et al. characterized the transporter ABCC4a for crocin transportation through transcriptomic and metabolomic analysis. Coexpression of ABCC4a with CCD2 in the leaves of *N. benthamiana* increased the yield of crocins [113].

Crocin-I (**2a**) and crocin-II (**2b**) are the main active crocin components, but the proportion of these two compounds in the total crocin content was usually low in heterologous hosts. Xie et al. introduced a multigene expression vector (*GjCCD4a-GjALDH2C3-GjUGT74F8-GjUGT94E13*) into *N. benthamiana* by using the fusion and 2A polypeptides strategy and obtained a crocin yield of 78,362 ng/g (fresh weight). The content of crocin-I (**2a**) and crocin-II (**2b**) accounted for 99% of the total crocin content [114]. The supply of UDP-glucose is a rate-limiting step in crocin biosynthesis. Glucose is first phosphorylated to generate glucose-6-phosphate by hexokinase and then converted into glucose-1-phosphate by phosphoglucomutase (PGM1). Afterwards, UTP-glucose-1-phosphate uridylyltransferase (GalU) converts glucose-1-phosphate into UDP-glucose. Pu et al. introduced *Gj*UGT74F8 from *G. jasminoides* into an *E. coli* strain overexpressing *PGM* and *GalU* and achieved the production of crocins [105].

## Conclusion

Crocins possess the characteristic polyene core structure of carotenoids and exhibit broad pharmacological properties. They are promising drugs for treating cardiovascular and liver diseases. Despite numerous pharmacological properties, the exact mechanisms of action of crocins remain elusive. The therapeutic potential of crocins is impeded by the limited bioavailability and rapid clearance in vivo. Only a few plants can produce crocins, and the content of crocins in these plants is very low. Due to the numerous chiral centers, the total synthesis of crocins is challenging. Therefore, heterologous biosynthesis of crocins utilizing the synthetic biology strategy holds great potential to solve the supply issue of crocins.

Significant progress has been made in the heterologous production of carotenoids in microorganisms in recent decades. The biosynthetic pathways of crocins have been extensively elucidated, and many enzymes in these pathways have been identified. The CCDs, ALDHs, and UGTs are the key players in crocin biosynthesis. Empowered by the genomic, transcriptomic, and metabolomic analysis and biochemical characterization, many such enzymes have been identified. Some proteins, exemplified by *Gj*CCD4a and *Gj*UGT94E13, exhibit broad substrate promiscuity and high catalytic activity. Screening of novel enzymes for crocin biosynthesis and the directed evolution of known enzymes should facilitate the high-titer production of the key intermediates **5–7** etc. and the efficient conversion of these precursors into crocetin (**1**) and crocins.

In conclusion, crocins are high-value plant natural products, and their biosynthetic pathways have been extensively studied. The construction of efficient hosts for the heterologous production of crocins will be a practical solution for the industrial production of these important chemicals.

## Acknowledgments

We thank all co-workers who contributed to this review.

## Funding

This work was supported by the Natural Key Research and Development Program of China (2020YFA0907900) and the Science and Technology Project of Haihe Laboratory of Modern Chinese Medicine (22HHZYJC00004).

## Author Contributions

Zhongwei Hua: writing – original draft; writing – review & editing. Nan Liu: writing – original draft; writing – review & editing. Xiaohui Yan: project administration; supervision; validation; writing – review & editing.

## ORCID® iDs

Zhongwei Hua - https://orcid.org/0009-0000-6508-9972
Nan Liu - https://orcid.org/0000-0002-9095-6430
Xiaohui Yan - https://orcid.org/0000-0003-4180-8543

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

1. Yan, X.; Hong, L.; Pei, S.; Hamilton, A.; Sun, H.; Yang, R.; Liu, A.; Yang, L. *Ind. Crops Prod.* **2021,** *171,* 113968. doi:10.1016/j.indcrop.2021.113968
2. Ahrazem, O.; Diretto, G.; Argandoña, J.; Rubio-Moraga, Á.; Julve, J. M.; Orzáez, D.; Granell, A.; Gómez-Gómez, L. *J. Exp. Bot.* **2017,** *68,* 4663–4677. doi:10.1093/jxb/erx277
3. Varadkar, M.; Gadgoli, C. *J. Tradit. Complementary Med.* **2022,** *12,* 354–360. doi:10.1016/j.jtcme.2021.10.002
4. Lou, Z.; Li, C.; Kou, X.; Yu, F.; Wang, H.; Smith, G. M.; Zhu, S. *J. Food Prot.* **2016,** *79,* 1404–1409. doi:10.4315/0362-028x.jfp-15-576
5. José Bagur, M.; Alonso Salinas, G. L.; Jiménez-Monreal, A. M.; Chaouqi, S.; Llorens, S.; Martínez-Tomé, M.; Alonso, G. L. *Molecules* **2017,** *23,* 30. doi:10.3390/molecules23010030
6. Colapietro, A.; Mancini, A.; D'Alessandro, A. M.; Festuccia, C. *Anti-Cancer Agents Med. Chem.* **2019,** *19,* 38–47. doi:10.2174/1871520619666181231112453
7. Wang, Y.; Han, T.; Zhu, Y.; Zheng, C.-J.; Ming, Q.-L.; Rahman, K.; Qin, L.-P. *J. Nat. Med.* **2010,** *64,* 24–30. doi:10.1007/s11418-009-0360-6
8. Liu, Y.; Yao, C.; Wang, Y.; Liu, X.; Xu, S.; Liang, L. *J. Surg. Res.* **2021,** *261,* 301–309. doi:10.1016/j.jss.2020.12.027
9. Yin, Q.; Xiong, H. *Eur. J. Histochem.* **2022,** *66,* 3541. doi:10.4081/ejh.2022.3541
10. Alsanie, W. F.; Alamri, A. S.; Abdulaziz, O.; Salih, M. M.; Alamri, A.; Asdaq, S. M. B.; Alhomrani, M. H.; Alhomrani, M. *Molecules* **2022,** *27,* 5462. doi:10.3390/molecules27175462
11. Hussain, M. A.; Abogresha, N. M.; AbdelKader, G.; Hassan, R.; Abdelaziz, E. Z.; Greish, S. M. *Oxid. Med. Cell. Longevity* **2021,** *2021,* 8841726. doi:10.1155/2021/8841726
12. Radmehr, V.; Ahangarpour, A.; Mard, S. A.; Khorsandi, L. *Iran. J. Basic Med. Sci.* **2022,** *25,* 179186.
13. Puglia, C.; Santonocito, D.; Musumeci, T.; Cardile, V.; Graziano, A. C. E.; Salerno, L.; Raciti, G.; Crascì, L.; Panico, A. M.; Puglisi, G. *Planta Med.* **2019,** *85,* 258–265. doi:10.1055/a-0732-5757
14. Bari, E.; Perteghella, S.; Rassu, G.; Gavini, E.; Petretto, G. L.; Bonferoni, M. C.; Giunchedi, P.; Torre, M. L. *Front. Pharmacol.* **2023,** *14,* 1129882. doi:10.3389/fphar.2023.1129882
15. Nasrpour, S.; Yousefi, G.; Niakosari, M.; Aminlari, M. *J. Food Sci.* **2022,** *87,* 1148–1160. doi:10.1111/1750-3841.16052
16. Moradi, A.; Zarinkamar, F.; De Domenico, S.; Mita, G.; Di Sansebastiano, G. P.; Caretto, S. *Plants* **2020,** *9,* 949. doi:10.3390/plants9080949
17. Liu, T.; Yu, S.; Xu, Z.; Tan, J.; Wang, B.; Liu, Y.-G.; Zhu, Q. *Comput. Struct. Biotechnol. J.* **2020,** *18,* 3278–3286. doi:10.1016/j.csbj.2020.10.019
18. Wan, X.; Zhou, X.-R.; Moncalian, G.; Su, L.; Chen, W.-C.; Zhu, H.-Z.; Chen, D.; Gong, Y.-M.; Huang, F.-H.; Deng, Q.-C. *Prog. Lipid Res.* **2021,** *81,* 101083. doi:10.1016/j.plipres.2020.101083
19. Liu, T.; Huang, Y.; Jiang, L.; Dong, C.; Gou, Y.; Lian, J. *Commun. Biol.* **2021,** *4,* 1089. doi:10.1038/s42003-021-02617-w
20. Carmona, M.; Zalacain, A.; Pardo, J. E.; López, E.; Alvarruiz, A.; Alonso, G. L. *J. Agric. Food Chem.* **2005,** *53,* 3974–3979. doi:10.1021/jf0404748
21. Aghaei, Z.; Jafari, S. M.; Dehnad, D. *Plant Foods Hum. Nutr. (N. Y., NY, U. S.)* **2019,** *74,* 171–178. doi:10.1007/s11130-019-00729-7
22. Cid-Pérez, T. S.; Nevárez-Moorillón, G. V.; Ochoa-Velasco, C. E.; Navarro-Cruz, A. R.; Hernández-Carranza, P.; Avila-Sosa, R. *Molecules* **2021,** *26,* 6954. doi:10.3390/molecules26226954
23. Cai, X.; Zhang, R.; Guo, Y.; He, J.; Li, S.; Zhu, Z.; Liu, G.; Liu, Z.; Yang, J. *Food Funct.* **2015,** *6,* 2194–2204. doi:10.1039/c5fo00205b
24. Sarfarazi, M.; Rajabzadeh, Q.; Tavakoli, R.; Ibrahim, S. A.; Jafari, S. M. *Ultrason. Sonochem.* **2022,** *86,* 105971. doi:10.1016/j.ultsonch.2022.105971
25. Fiorito, S.; Palumbo, L.; Epifano, F.; Collevecchio, C.; Cardellini, F.; Bastianini, M.; Spogli, R.; Genovese, S. *Food Chem.* **2022,** *377,* 132040. doi:10.1016/j.foodchem.2022.132040
26. Hatziagapiou, K.; Kakouri, E.; Lambrou, G. I.; Bethanis, K.; Tarantilis, P. A. *Curr. Neuropharmacol.* **2019,** *17,* 377–402. doi:10.2174/1570159x16666180321095705
27. Ahmed, S.; Hasan, M. M.; Heydari, M.; Rauf, A.; Bawazeer, S.; Abu-Izneid, T.; Rebezov, M.; Shariati, M. A.; Daglia, M.; Rengasamy, K. R. *Food Chem. Toxicol.* **2020,** *145,* 111739. doi:10.1016/j.fct.2020.111739
28. Wang, C.; Cai, X.; Hu, W.; Li, Z.; Kong, F.; Chen, X.; Wang, D. *Int. J. Mol. Med.* **2019,** *43,* 956–966.
29. Geromichalos, G. D.; Lamari, F. N.; Papandreou, M. A.; Trafalis, D. T.; Margarity, M.; Papageorgiou, A.; Sinakos, Z. *J. Agric. Food Chem.* **2012,** *60,* 6131–6138. doi:10.1021/jf300589c
30. Hadipour, M.; Kaka, G.; Bahrami, F.; Meftahi, G. H.; Pirzad Jahromi, G.; Mohammadi, A.; Sahraei, H. *Synapse* **2018,** *72,* e22026. doi:10.1002/syn.22026
31. Asadi, F.; Jamshidi, A. H.; Khodagholi, F.; Yans, A.; Azimi, L.; Faizi, M.; Vali, L.; Abdollahi, M.; Ghahremani, M. H.; Sharifzadeh, M. *Pharmacol., Biochem. Behav.* **2015,** *139,* 47–58. doi:10.1016/j.pbb.2015.10.011
32. Inoue, E.; Shimizu, Y.; Masui, R.; Hayakawa, T.; Tsubonoya, T.; Hori, S.; Sudoh, K. *J. Nat. Med.* **2018,** *72,* 274–279. doi:10.1007/s11418-017-1150-1
33. Karakani, A. M.; Riazi, G.; Mahmood Ghaffari, S.; Ahmadian, S.; Mokhtari, F.; Jalili Firuzi, M.; Zahra Bathaie, S. *Iran. J. Basic Med. Sci.* **2015,** *18,* 485492.
34. Ghadrdoost, B.; Vafaei, A. A.; Rashidy-Pour, A.; Hajisoltani, R.; Bandegi, A. R.; Motamedi, F.; Haghighi, S.; Sameni, H. R.; Pahlvan, S. *Eur. J. Pharmacol.* **2011,** *667,* 222–229. doi:10.1016/j.ejphar.2011.05.012
35. Chen, B.; Choi, H.; Hirsch, L. J.; Katz, A.; Legge, A.; Buchsbaum, R.; Detyniecki, K. *Epilepsy Behav.* **2017,** *76,* 24–31. doi:10.1016/j.yebeh.2017.08.039
36. Ryvlin, P.; Rheims, S.; Hirsch, L. J.; Sokolov, A.; Jehi, L. *Lancet Neurol.* **2021,** *20,* 1038–1047. doi:10.1016/s1474-4422(21)00300-8
37. Remy, S.; Beck, H. *Brain* **2006,** *129,* 18–35. doi:10.1093/brain/awh682
38. Tamaddonfard, E.; Gooshchi, N. H.; Seiednejad-Yamchi, S. *Pharmacol. Rep.* **2012,** *64,* 94–101. doi:10.1016/s1734-1140(12)70735-1

39. Rana, A.; Musto, A. E. *J. Neuroinflammation* **2018,** *15,* 144. doi:10.1186/s12974-018-1192-7
40. Nikpour Moghaddam, S.; Qujeq, D.; Rastegari Efahani, A. A. *Avicenna J. Med. Biochem.* **2016,** *4,* 4-29429. doi:10.17795/ajmb.29429
41. Wang, X.; Tang, O.; Ye, Y.; Zheng, M.; Hu, J.; Chen, Z.; Zhong, K. *Zhejiang Daxue Xuebao, Yixueban* **2017,** *46,* 714.
42. Zhong, K.; Qian, C.; Lyu, R.; Wang, X.; Hu, Z.; Yu, J.; Ma, J.; Ye, Y. *Front. Pharmacol.* **2022,** *13,* 757729. doi:10.3389/fphar.2022.757729
43. Imenshahidi, M.; Hosseinzadeh, H.; Javadpour, Y. *Phytother. Res.* **2010,** *24,* 990–994. doi:10.1002/ptr.3044
44. Shafei, M. N.; Faramarzi, A.; Khajavi Rad, A.; Anaeigoudari, A. *Avicenna J. Phytomed.* **2017,** *7,* 345352.
45. Chen, X.; Huang, J.; Lv, Y.; Chen, Y.; Rao, J. *Hypertens. Res.* **2021,** *44,* 642–650. doi:10.1038/s41440-020-00609-7
46. Jiang, Z.; Gu, M.; Liu, J.; Li, H.; Peng, J.; Zhang, Y. *J. Photochem. Photobiol., B* **2018,** *180,* 118–124. doi:10.1016/j.jphotobiol.2018.01.013
47. Mollaei, H.; Safaralizadeh, R.; Babaei, E.; Abedini, M. R.; Hoshyar, R. *Biomed. Pharmacother.* **2017,** *94,* 307–316. doi:10.1016/j.biopha.2017.07.052
48. Hoshyar, R.; Bathaie, S. Z.; Sadeghizadeh, M. *DNA Cell Biol.* **2013,** *32,* 50–57. doi:10.1089/dna.2012.1866
49. Mostafavinia, S. E.; Khorashadizadeh, M.; Hoshyar, R. *DNA Cell Biol.* **2016,** *35,* 340–347. doi:10.1089/dna.2015.3208
50. Festuccia, C.; Mancini, A.; Gravina, G. L.; Scarsella, L.; Llorens, S.; Alonso, G. L.; Tatone, C.; Di Cesare, E.; Jannini, E. A.; Lenzi, A.; D'Alessandro, A. M.; Carmona, M. *BioMed Res. Int.* **2014,** 135048. doi:10.1155/2014/135048
51. Abdu, S.; Juaid, N.; Amin, A.; Moulay, M.; Miled, N. *Antioxidants* **2022,** *11,* 1645. doi:10.3390/antiox11091645
52. Bakshi, H. A.; Quinn, G. A.; Nasef, M. M.; Mishra, V.; Aljabali, A. A. A.; El-Tanani, M.; Serrano-Aroca, Á.; Webba Da Silva, M.; McCarron, P. A.; Tambuwala, M. M. *Cells* **2022,** *11,* 1502. doi:10.3390/cells11091502
53. Moradzadeh, M.; Kalani, M. R.; Avan, A. *J. Cell. Biochem.* **2019,** *120,* 4732–4738. doi:10.1002/jcb.27525
54. Kim, B.; Lee, K. Y.; Park, B. *J. Cell. Biochem.* **2017,** *118,* 3290–3298. doi:10.1002/jcb.25980
55. Kim, B.; Park, B. *Oncol. Rep.* **2018,** *39,* 1883–1891. doi:10.3892/or.2018.6232
56. Abdulkareem Aljumaily, S. A.; Demir, M.; Elbe, H.; Yigitturk, G.; Bicer, Y.; Altinoz, E. *Environ. Sci. Pollut. Res.* **2021,** *28,* 65802–65813. doi:10.1007/s11356-021-15409-w
57. Khoshandam, A.; Razavi, B. M.; Hosseinzadeh, H. *Iran. J. Basic Med. Sci.* **2022,** *25,* 789798.
58. Baradaran Rahim, V.; Khammar, M. T.; Rakhshandeh, H.; Samzadeh-Kermani, A.; Hosseini, A.; Askari, V. R. *Pharmacol. Rep.* **2019,** *71,* 1228–1234. doi:10.1016/j.pharep.2019.07.007
59. Rashidi, K.; Korani, M.; Nemati, H.; Shahraki, R.; Korani, S.; Abbasifard, M.; Majeed, M.; Jamialahmadi, T.; Sahebkar, A. *Curr. Med. Chem.* **2024,** in press. doi:10.2174/0929867330666230409003744
60. Noorbala, A. A.; Akhondzadeh, S.; Tahmacebi-Pour, N.; Jamshidi, A. H. *J. Ethnopharmacol.* **2005,** *97,* 281–284. doi:10.1016/j.jep.2004.11.004
61. Akhondzadeh, S.; Fallah-Pour, H.; Afkham, K.; Jamshidi, A.-H.; Khalighi-Cigaroudi, F. *BMC Complementary Altern. Med.* **2004,** *4,* 12. doi:10.1186/1472-6882-4-12
62. Moshiri, M.; Vahabzadeh, M.; Hosseinzadeh, H. *Drug Res. (Stuttgart, Ger.)* **2015,** *65,* 287–295. doi:10.1055/s-0034-1375681
63. Siddiqui, S. A.; Ali Redha, A.; Snoeck, E. R.; Singh, S.; Simal-Gandara, J.; Ibrahim, S. A.; Jafari, S. M. *Molecules* **2022,** *27,* 2076. doi:10.3390/molecules27072076
64. Moragrega, I.; Ríos, J. L. *Planta Med.* **2021,** *87,* 656–685. doi:10.1055/a-1338-1011
65. Kashani, L.; Esalatmanesh, S.; Eftekhari, F.; Salimi, S.; Foroughifar, T.; Etesam, F.; Safiaghdam, H.; Moazen-Zadeh, E.; Akhondzadeh, S. *Arch. Gynecol. Obstet.* **2018,** *297,* 717–724. doi:10.1007/s00404-018-4655-2
66. Akhondzadeh, S.; Mostafavi, S.-A.; Keshavarz, S. A.; Mohammadi, M. R.; Hosseini, S.; Eshraghian, M. R. *J. Clin. Pharm. Ther.* **2020,** *45,* 134–143. doi:10.1111/jcpt.13040
67. Tabeshpour, J.; Sobhani, F.; Sadjadi, S. A.; Hosseinzadeh, H.; Mohajeri, S. A.; Rajabi, O.; Taherzadeh, Z.; Eslami, S. *Phytomedicine* **2017,** *36,* 145–152. doi:10.1016/j.phymed.2017.10.005
68. Xiao, Q.; Xiong, Z.; Yu, C.; Zhou, J.; Shen, Q.; Wang, L.; Xie, X.; Fu, Z. *Physiol. Behav.* **2019,** *212,* 112699. doi:10.1016/j.physbeh.2019.112699
69. Tao, W.; Ruan, J.; Wu, R.; Zhao, M.; Zhao, T.; Qi, M.; Yau, S. S. Y.; Yao, G.; Zhang, H.; Hu, Y.; Chen, G. *J. Adv. Res.* **2023,** *43,* 219–231. doi:10.1016/j.jare.2022.02.015
70. Samaha, M. M.; Said, E.; Salem, H. A. *Environ. Toxicol. Pharmacol.* **2019,** *72,* 103238. doi:10.1016/j.etap.2019.103238
71. Qiu, Y.; Jiang, X.; Liu, D.; Deng, Z.; Hu, W.; Li, Z.; Li, Y. *Front. Pharmacol.* **2020,** *11,* 541. doi:10.3389/fphar.2020.00541
72. Ali Hammood Keelo, R. M.; Elbe, H.; Bicer, Y.; Yigitturk, G.; Koca, O.; Karayakali, M.; Acar, D.; Altinoz, E. *Chem.-Biol. Interact.* **2022,** *351,* 109733. doi:10.1016/j.cbi.2021.109733
73. Yang, X.; Huo, F.; Liu, B.; Liu, J.; Chen, T.; Li, J.; Zhu, Z.; Lv, B. *J. Mol. Neurosci.* **2017,** *61,* 581–589. doi:10.1007/s12031-017-0899-8
74. Rajaei, Z.; Hadjzadeh, M.-A.-R.; Nemati, H.; Hosseini, M.; Ahmadi, M.; Shafiee, S. *J. Med. Food* **2013,** *16,* 206–210. doi:10.1089/jmf.2012.2407
75. Kakouri, E.; Agalou, A.; Kanakis, C.; Beis, D.; Tarantilis, P. A. *Molecules* **2020,** *25,* 5223. doi:10.3390/molecules25225223
76. Liu, H.; Chen, Y.-F.; Li, F.; Zhang, H.-Y. *J. Asian Nat. Prod. Res.* **2013,** *15,* 94–110. doi:10.1080/10286020.2012.723203
77. Karayakali, M.; Altinoz, E.; Elbe, H.; Koca, O.; Onal, M. O.; Bicer, Y.; Demir, M. *Environ. Sci. Pollut. Res.* **2023,** *30,* 47670–47684. doi:10.1007/s11356-023-25766-3
78. Sepahi, S.; Ghorani-Azam, A.; Hossieni, S. M.; Mohajeri, S. A.; Khodaverdi, E. *Curr. Neuropharmacol.* **2021,** *19,* 392–401. doi:10.2174/1570159x18666200507083346
79. Boskabady, M. H.; Farkhondeh, T. *Phytother. Res.* **2016,** *30,* 1072–1094. doi:10.1002/ptr.5622
80. Manochkumar, J.; Singh, A.; Efferth, T.; Ramamoorthy, S. *Phytomedicine* **2022,** *104,* 154286. doi:10.1016/j.phymed.2022.154286
81. Xie, Y.; Chen, S.; Xiong, X. *Front. Microbiol.* **2021,** *12,* 699235. doi:10.3389/fmicb.2021.699235
82. Gupta, I.; Adin, S. N.; Panda, B. P.; Mujeeb, M. *Biotechnol. Appl. Biochem.* **2022,** *69,* 2517–2529. doi:10.1002/bab.2301
83. Jia, Y.-P.; Sun, L.; Yu, H.-S.; Liang, L.-P.; Li, W.; Ding, H.; Song, X.-B.; Zhang, L.-J. *Molecules* **2017,** *22,* 610. doi:10.3390/molecules22040610

84. Graham, J. E.; Bryant, D. A. *J. Bacteriol.* **2009,** *191,* 3292–3300. doi:10.1128/jb.00050-09
85. Zheng, X.; Mi, J.; Balakrishna, A.; Liew, K. X.; Ablazov, A.; Sougrat, R.; Al-Babili, S. *Plant Biotechnol. J.* **2022,** *20,* 2202–2216. doi:10.1111/pbi.13901
86. Moraga, A. R.; Nohales, P. F.; Pérez, J. A. F.; Gómez-Gómez, L. *Planta* **2004,** *219,* 955–966. doi:10.1007/s00425-004-1299-1
87. Diretto, G.; Ahrazem, O.; Rubio-Moraga, Á.; Fiore, A.; Sevi, F.; Argandoña, J.; Gómez-Gómez, L. *New Phytol.* **2019,** *224,* 725–740. doi:10.1111/nph.16079
88. Ahrazem, O.; Gómez-Gómez, L.; Rodrigo, M. J.; Avalos, J.; Limón, M. C. *Int. J. Mol. Sci.* **2016,** *17,* 1781. doi:10.3390/ijms17111781
89. Auldridge, M. E.; McCarty, D. R.; Klee, H. J. *Curr. Opin. Plant Biol.* **2006,** *9,* 315–321. doi:10.1016/j.pbi.2006.03.005
90. Frusciante, S.; Diretto, G.; Bruno, M.; Ferrante, P.; Pietrella, M.; Prado-Cabrero, A.; Rubio-Moraga, A.; Beyer, P.; Gomez-Gomez, L.; Al-Babili, S.; Giuliano, G. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111,* 12246–12251. doi:10.1073/pnas.1404629111
91. Liang, N.; Yao, M.-D.; Wang, Y.; Liu, J.; Feng, L.; Wang, Z.-M.; Li, X.-Y.; Xiao, W.-H.; Yuan, Y.-J. *J. Agric. Food Chem.* **2021,** *69,* 11626–11636. doi:10.1021/acs.jafc.1c04588
92. Martí, M.; Diretto, G.; Aragonés, V.; Frusciante, S.; Ahrazem, O.; Gómez-Gómez, L.; Daròs, J.-A. *Metab. Eng.* **2020,** *61,* 238–250. doi:10.1016/j.ymben.2020.06.009
93. Carballo-Uicab, V. M.; Cárdenas-Conejo, Y.; Vallejo-Cardona, A. A.; Aguilar-Espinosa, M.; Rodríguez-Campos, J.; Serrano-Posada, H.; Narváez-Zapata, J. A.; Vázquez-Flota, F.; Rivera-Madrid, R. *PeerJ* **2019,** *7,* e7064. doi:10.7717/peerj.7064
94. Bouvier, F.; Dogbo, O.; Camara, B. *Science* **2003,** *300,* 2089–2091. doi:10.1126/science.1085162
95. Frusciante, S.; Demurtas, O. C.; Sulli, M.; Mini, P.; Aprea, G.; Diretto, G.; Karcher, D.; Bock, R.; Giuliano, G. *Plant Physiol.* **2022,** *188,* 1469–1482. doi:10.1093/plphys/kiab583
96. Xu, Z.; Pu, X.; Gao, R.; Demurtas, O. C.; Fleck, S. J.; Richter, M.; He, C.; Ji, A.; Sun, W.; Kong, J.; Hu, K.; Ren, F.; Song, J.; Wang, Z.; Gao, T.; Xiong, C.; Yu, H.; Xin, T.; Albert, V. A.; Giuliano, G.; Chen, S.; Song, J. *BMC Biol.* **2020,** *18,* 63. doi:10.1186/s12915-020-00795-3
97. Demurtas, O. C.; Frusciante, S.; Ferrante, P.; Diretto, G.; Azad, N. H.; Pietrella, M.; Aprea, G.; Taddei, A. R.; Romano, E.; Mi, J.; Al-Babili, S.; Frigerio, L.; Giuliano, G. *Plant Physiol.* **2018,** *177,* 990–1006. doi:10.1104/pp.17.01815
98. Gómez-Gómez, L.; Pacios, L. F.; Diaz-Perales, A.; Garrido-Arandia, M.; Argandoña, J.; Rubio-Moraga, Á.; Ahrazem, O. *Int. J. Mol. Sci.* **2018,** *19,* 1409. doi:10.3390/ijms19051409
99. Tan, H.; Chen, X.; Liang, N.; Chen, R.; Chen, J.; Hu, C.; Li, Q.; Li, Q.; Pei, W.; Xiao, W.; Yuan, Y.; Chen, W.; Zhang, L. *J. Exp. Bot.* **2019,** *70,* 4819–4834. doi:10.1093/jxb/erz211
100. Diretto, G.; López-Jiménez, A. J.; Ahrazem, O.; Frusciante, S.; Song, J.; Rubio-Moraga, Á.; Gómez-Gómez, L. *J. Exp. Bot.* **2021,** *72,* 3200–3218. doi:10.1093/jxb/erab053
101. Dufresne, C.; Cormier, F.; Dorion, S. *Planta Med.* **1997,** *63,* 150–153. doi:10.1055/s-2006-957633
102. Côté, F.; Cormier, F.; Dufresne, C.; Willemot, C. *J. Plant Physiol.* **2001,** *158,* 553–560. doi:10.1078/0176-1617-00305
103. López-jimenez, A. J.; Frusciante, S.; Niza, E.; Ahrazem, O.; Rubio-Moraga, Á.; Diretto, G.; Gómez-Gómez, L. *Int. J. Mol. Sci.* **2021,** *22,* 8815. doi:10.3390/ijms22168815
104. Nagatoshi, M.; Terasaka, K.; Owaki, M.; Sota, M.; Inukai, T.; Nagatsu, A.; Mizukami, H. *FEBS Lett.* **2012,** *586,* 1055–1061. doi:10.1016/j.febslet.2012.03.003
105. Pu, X.; He, C.; Yang, Y.; Wang, W.; Hu, K.; Xu, Z.; Song, J. *ACS Synth. Biol.* **2020,** *9,* 1160–1168. doi:10.1021/acssynbio.0c00039
106. Ding, F.; Liu, F.; Shao, W.; Chu, J.; Wu, B.; He, B. *J. Agric. Food Chem.* **2018,** *66,* 11701–11708. doi:10.1021/acs.jafc.8b04274
107. Chai, F.; Wang, Y.; Mei, X.; Yao, M.; Chen, Y.; Liu, H.; Xiao, W.; Yuan, Y. *Microb. Cell Fact.* **2017,** *16,* 54. doi:10.1186/s12934-017-0665-1
108. Liu, T.; Dong, C.; Qi, M.; Zhang, B.; Huang, L.; Xu, Z.; Lian, J. *Front. Bioeng. Biotechnol.* **2020,** *8,* 653. doi:10.3389/fbioe.2020.00653
109. Wang, W.; He, P.; Zhao, D.; Ye, L.; Dai, L.; Zhang, X.; Sun, Y.; Zheng, J.; Bi, C. *Microb. Cell Fact.* **2019,** *18,* 120. doi:10.1186/s12934-019-1166-1
110. Shan, M.; Yao, M.; Liang, N.; Wang, H.; Wu, N.; Wang, Y.; Xiao, W.; Yuan, Y.-j. *ACS Sustainable Chem. Eng.* **2023,** *11,* 8615–8623. doi:10.1021/acssuschemeng.3c01533
111. Ahrazem, O.; Zhu, C.; Huang, X.; Rubio-Moraga, A.; Capell, T.; Christou, P.; Gómez-Gómez, L. *Front. Plant Sci.* **2022,** *13,* 861140. doi:10.3389/fpls.2022.861140
112. Ahrazem, O.; Diretto, G.; Rambla, J. L.; Rubio-Moraga, Á.; Lobato-Gómez, M.; Frusciante, S.; Argandoña, J.; Presa, S.; Granell, A.; Gómez-Gómez, L. *Hortic. Res.* **2022,** *9,* uhac074. doi:10.1093/hr/uhac074
113. Augustine, R. C. *Plant Cell* **2019,** *31,* 2558. doi:10.1105/tpc.19.00720
114. Xie, L.; Luo, Z.; Jia, X.; Mo, C.; Huang, X.; Suo, Y.; Cui, S.; Zang, Y.; Liao, J.; Ma, X. *Int. J. Mol. Sci.* **2023,** *24,* 14139. doi:10.3390/ijms241814139
115. Song, T.; Wu, N.; Wang, C.; Wang, Y.; Chai, F.; Ding, M.; Li, X.; Yao, M.; Xiao, W.; Yuan, Y. *Front. Bioeng. Biotechnol.* **2020,** *8,* 578005. doi:10.3389/fbioe.2020.578005


## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.68

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Methodology for awakening the potential secondary metabolic capacity in actinomycetes

Shun Saito* and Midori A. Arai*

Review

Open Access

Address:
Department of Biosciences and Informatics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan

Email:
Shun Saito* - ssaito@bio.keio.ac.jp; Midori A. Arai* - midori_arai@bio.keio.ac.jp

* Corresponding author

Keywords:
actinomycete; co-culture; heat shock metabolites (HSMs); secondary metabolites; silent genes

*Beilstein J. Org. Chem.* **2024,** *20,* 753–766.
https://doi.org/10.3762/bjoc.20.69

Received: 15 December 2023
Accepted: 27 March 2024
Published: 10 April 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: T. Awakawa

© 2024 Saito and Arai; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Secondary metabolites produced by actinomycete strains undoubtedly have great potential for use in applied research areas such as drug discovery. However, it is becoming difficult to obtain novel compounds because of repeated isolation around the world. Therefore, a new strategy for discovering novel secondary metabolites is needed. Many researchers believe that actinomycetes have as yet unanalyzed secondary metabolic activities, and the associated undiscovered secondary metabolite biosynthesis genes are called “silent” genes. This review outlines several approaches to further activate the metabolic potential of actinomycetes.

## Introduction

Research clearly indicates that compounds derived from natural products have contributed greatly to enhancing the welfare of human society. For example, the structures of compounds produced by microorganisms are very diverse, facilitating their application in the development of therapeutic drugs [1], agrochemicals [2], veterinary medicines [3], medium supplements for selective microbial/cell culture [4-6], food preservatives/colorings [7,8], and biochemical reagents for pharmacological/chemical biology studies [9]. This review has focused particularly on compounds produced by actinomycetes. Secondary metabolites produced by actinomycetes, representing a class of compounds with a diverse chemical space, have promising potential as sources of bioactive substances [10]. Due to his discovery of streptomycin from *Streptomyces griseus*, Waksman was awarded the Nobel Prize in Physiology and Medicine in 1952 [11]. Ōmura, who discovered avermectin from *Streptomyces avermitilis*, received the same award in 2015 [12]. While there are many such brilliant achievements, researchers are finding it increasingly difficult to isolate novel compounds from secondary metabolites produced by actinomycetes [13,14]. This is the result of more than 50 years of effort to scrutinize substances produced by actinomycetes. However, many researchers have questioned whether the potential secondary metabolic capacity of actinomycetes has been fully analyzed and have continued to search for compounds from actinomycetes based on original ideas. For example, in searching for compounds based on biological activity, a proper screening system is critical [15]. Screening systems are broadly

classified as either forward screening (searching for compounds that induce phenotypic changes) or reverse screening (searching for compounds that act directly on specific molecules). Imoto and colleagues constructed a reverse screening system that targets a malignant factor of prostate cancer and thereby identified several novel compounds from actinomycete strains [16-19].

Many researchers have focused on identifying secondary metabolites from actinomycetes of the genus *Streptomyces* due to the prominence of species of this genus in surface soils and the difficulty of isolating members of other genera from natural environments. For this reason, actinomycetes other than those of the genus *Streptomyces* are referred to as "rare actinomycetes", and the secondary metabolites produced by these species have thus attracted increasing attention. Indeed, an increasing proportion of newly identified antibiotics are reportedly produced by rare actinomycetes [20]. Igarashi and colleagues have also focused on various rare actinomycetes, analyzing their secondary metabolic activity and demonstrating the importance of studying new genera and species [21-25]. On the other hand, it is now thought that secondary metabolic activities remain to be analyzed in actinomycetes, including *Streptomyces* strains, which were considered to have been completely characterized. This may be because the number of discovered compounds is small compared to the number of secondary metabolite biosynthesis genes harbored by actinomycetes [26-28]. For example, in *Streptomyces avermitilis*, 38 secondary metabolite biosynthetic gene clusters have been identified by genome sequencing. Interestingly, only 16 secondary metabolites of *S. avermitilis* have been identified to date, even though the species is widely used as a standard actinomycete. This means that only 40% of the secondary metabolites from the 38 gene clusters have been identified. These as yet undiscovered secondary metabolite biosynthesis genes are called "silent genes", because they are either not expressed or their expression levels are low under normal culture conditions. A number of studies have reported methods to activate these genes, and many new compounds have been discovered. In this review, we outline the silent gene activation methods, including the authors' efforts (Figure 1).

# Review

## Artificial methods

Several target-oriented methods focusing on specific biosynthetic genes and regulatory factors and artificial methods have been reported to activate secondary metabolism in actinomycetes. The technique known as "heterologous expression", in which a biosynthetic gene cluster of a secondary metabolite is introduced into another organism, is a useful method for producing silent secondary metabolites (Figure 2a). For example, the cryptic trans-acyltransferase polyketide synthase biosyn-

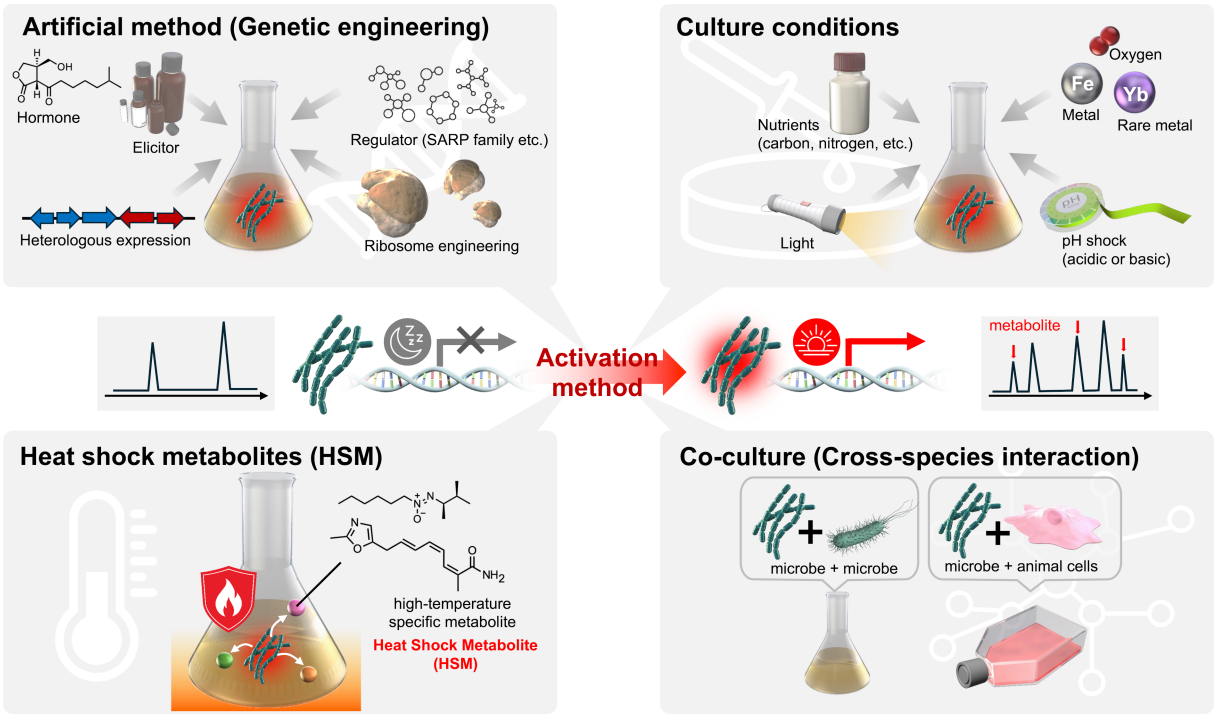



**Figure 1:** Schematic diagram of methods to activate silent genes in actinomycetes as presented in this review. This figure was created by Science Graphics Co., Ltd. This content is not subject to CC BY 4.0.

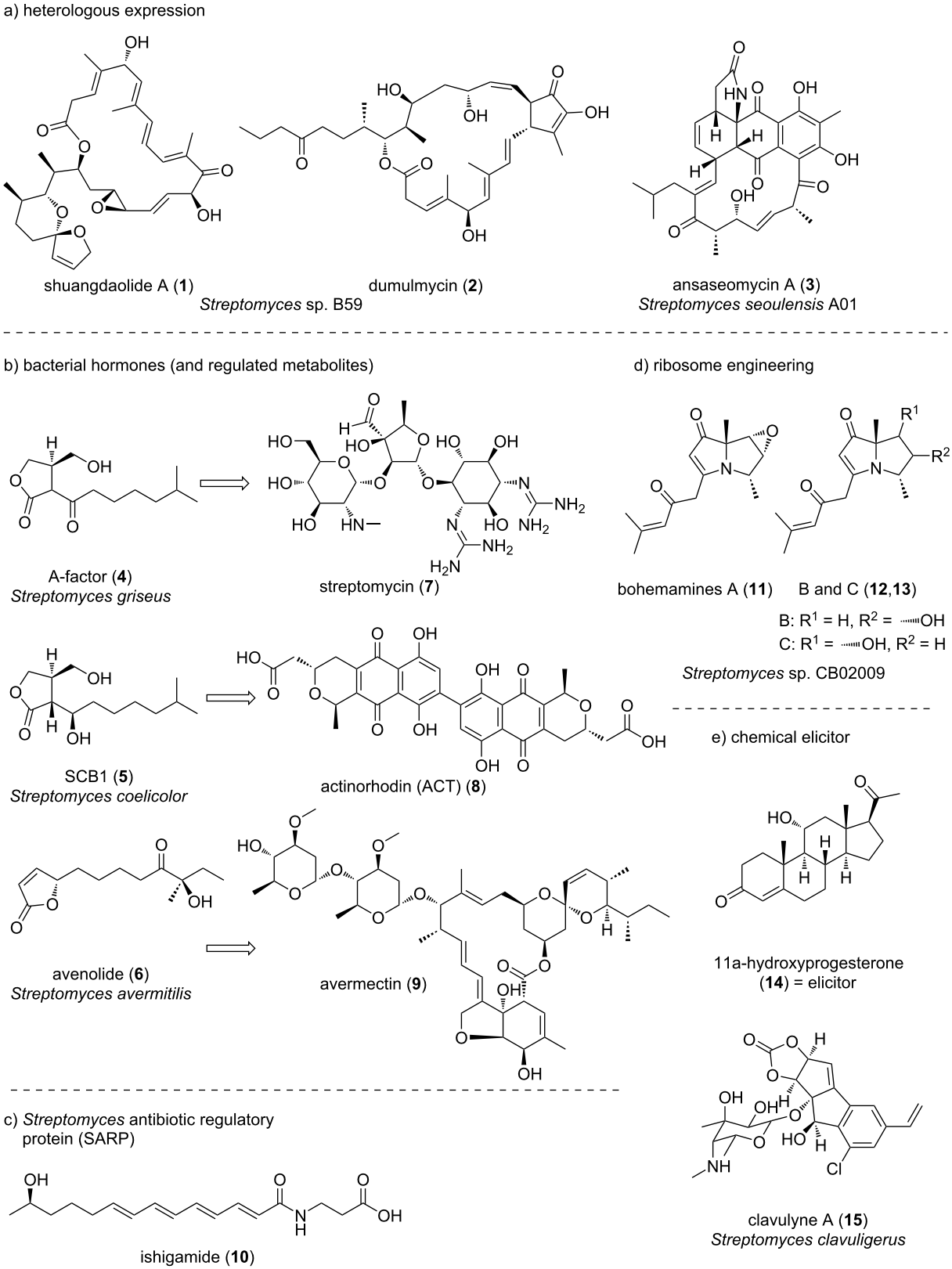



**Figure 2:** Structures of secondary metabolites obtained from actinomycetes using artificial methods.

thetic gene cluster *sdl* (80 kb) from *Streptomyces* sp. B59 was cloned and transferred into a heterologous host, *Streptomyces albus* J1074, resulting in the production of a class of polycyclic macrolide shuangdaolides A (**1**), B, and D and dumulmycin (**2**) [29]. Furthermore, genome mining of the marine actinomycete *Streptomyces seoulensis* A01 enabled the identification of a giant type I polyketide synthase gene cluster (*asm*) [30]. Heterologous expression of the cryptic *asm* cluster using a bacterial artificial chromosome vector in a heterologous host led to the production of ansaseomycins A (**3**) and B.

A number of factors associated with the biosynthetic gene clusters of secondary metabolites regulate their expression; thus, controlling these factors can also be used to activate secondary metabolism (Figure 2b). For example, the production of some secondary metabolites in actinomycetes is regulated by various bacterial hormones, such as A-factor (**4**) [31] and SCB1 (**5**) [32]. Normally, the tetR regulator negatively regulates the expression of biosynthetic genes in actinomycetes, but bacterial hormones release this repression [33,34]. A-factor, SCB1, and avenolide (**6**) induce the production of streptomycin (**7**), actinorhodin (**8**), and avermectin (**9**), respectively [35-37]. The production of these metabolites is regulated by the respective tetR regulators ArpA, ScbR, and AvaR1, respectively [38-40]. Therefore, increasing the levels of bacterial hormones present or disruption of the tetR regulator have been reported in several studies as methods to improve target-substance production capacity. Interestingly, Wilbanks et al. reported the structure–activity relationships of bacterial hormones in secondary metabolite biosynthesis by diversifiable synthesis [41]. The ability to synthesize stereochemically pure bacterial hormones would greatly expedite studies of these repressors and ultimately lead to the discovery of regulatory systems for many secondary metabolites.

Overexpression of transcription factors present in biosynthetic gene clusters can also be employed to increase the production of secondary metabolites. For example, overexpression of *Streptomyces* antibiotic regulatory protein (SARP) family transcriptional activators led to the discovery of many new secondary metabolites [42]. Du et al. searched for SARP family transcriptional activators from the draft genome of *Streptomyces* sp. MSC090213JE08, and constructed recombinant strains with forced expression of each of the seven SARP genes. When these were cultured in four different media and examined for metabolic profiles, a total of nine unknown metabolites were found in the four SARP-expressing strains under specific culture conditions. Du et al. identified a new compound (designated ishigamide (**10**)) as 3-([2*E*,4*E*,6*E*,8*E*]-13-hydroxytetradeca-2,4,6,8-tetraenamido)propanoic acid and elucidated the biosynthetic machinery (Figure 2c) [43-45]. Seo et al. synthesized ishigamide and determined its absolute configuration [46].

Finally, several studies have reported the modulation of secondary metabolic activation in actinomycetes using small molecules. For example, Ochi et al. proposed the "ribosome engineering" technique, in which actinomycetes are cultured with antibiotics to activate secondary metabolism by inducing resistance, and this approach has contributed to the isolation of many new compounds [47]. The ribosome engineering method is based on the principle that a mutation in the ribosomal S12 protein, which is associated with the acquisition of antibiotic resistance, activates secondary metabolism. The ribosomal S12 protein mutation results in increased expression of translation factors, which leads to enhanced protein synthesis in secondary metabolism [47]. Liu et al. reported activation of the production of bohemamines **11**–**13**, bacterial alkaloids containing a pyrrolizidine core with two unusual methyl groups, using a ribosome engineering approach in *Streptomyces* sp. CB02009 (Figure 2d) [48].

Seyedsayamdost et al. developed the high-throughput elicitor screening (HiTES) approach, a forward chemical genetics method that identifies small-molecule inducers of silent natural products (Figure 2e) [49]. Han et al. applied a fluorescence-based DNA cleavage assay coupled with HiTES to *Streptomyces clavuligerus* and identified the steroid 11α-hydroxyprogesterone (**14**) as an effective elicitor and characterized 10 cryptic enediyne-derived natural products, designated clavulynes A (**15**) and B–J with unusual carbonate and terminal olefin functionalities [50]. Thus, artificial methods of genetic engineering and chemistry also play an important role in the identification of novel secondary metabolites.

## Culture conditions

### Medium composition, pH, oxygen supply, light

Simply modifying the culture conditions is an important means of activating silent genes in actinomycetes. Because the cultivation conditions play a key role in metabolite production by actinomycetes, changing these parameters can significantly alter secondary metabolite production patterns. For example, OSMAC (one strain-multi compound), developed by Zeeck and co-workers in the early 2000s, is a method in which the target bacteria are cultured under various conditions (medium composition, temperature, pH, oxygen supply, light quality and quantity, addition of precursors and enzyme inhibitors, etc.) and all metabolites obtained from them are analyzed (Figure 3a) [51]. Zeeck's group has successfully isolated metabolites consisting of diverse structural classes from only a few microbial species by their OSMAC approach [52-54]. By applying this method, Rateb et al. and Sproule et al. discovered 22-membered macro-

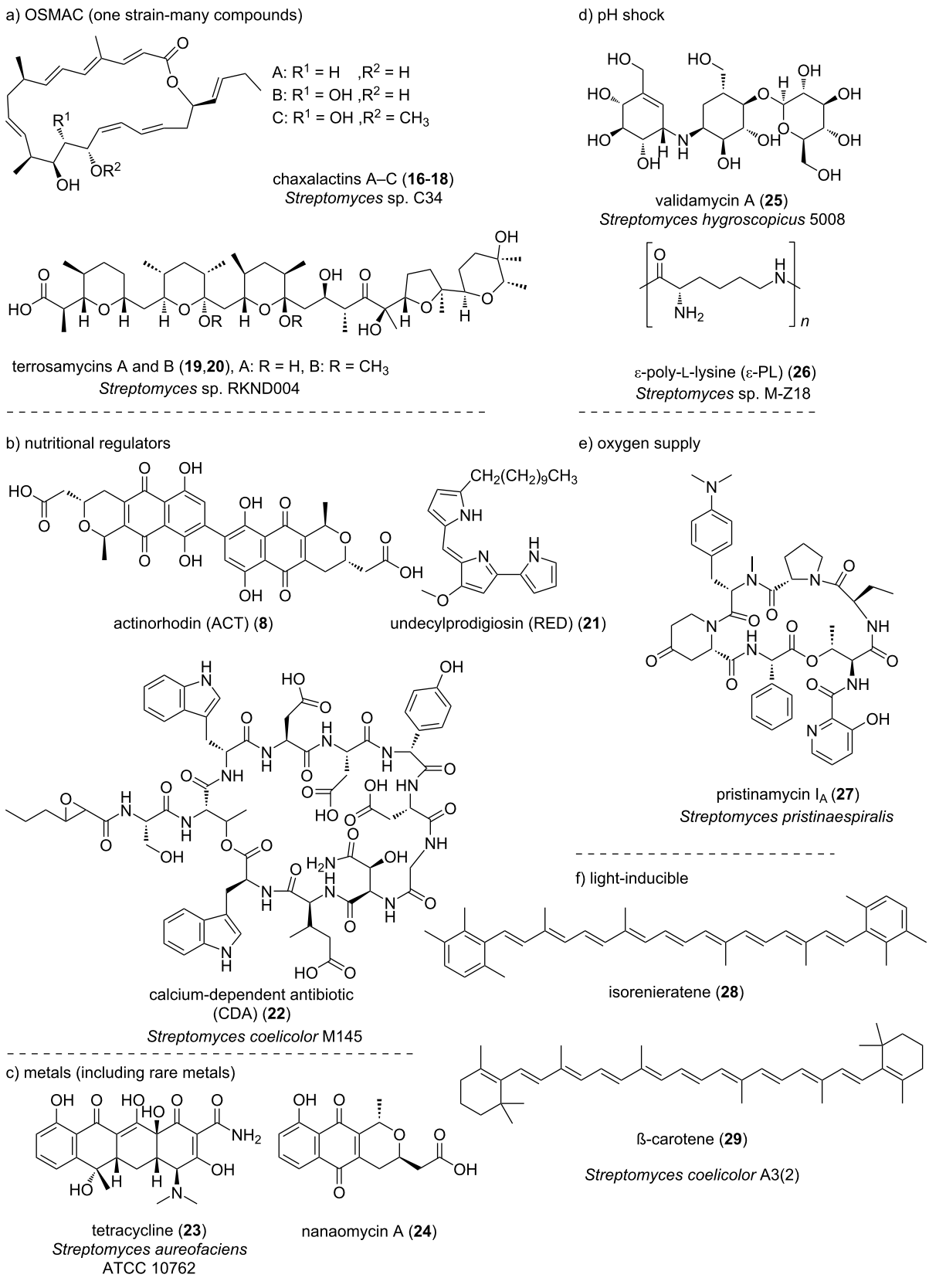



**Figure 3:** Structures of secondary metabolites obtained from actinomycetes by adjusting culture conditions.

lactone polyketides designated chaxalactins A–C (**16**–**18**), and polycyclic polyether natural products designated terrosamycins A and B (**19**, **20**), respectively [55,56]. Thus, the OSMAC strategy has been successfully applied over the last 10 years to generate new secondary metabolites from a single microbial strain.

Sensor proteins for medium components are also present in actinomycetes, and a regulatory system for secondary metabolism has been reported [57]. For example, Świątek et al. proposed that Rok7B7, a transcriptional regulator, acts as the central control protein for carbon control in *Streptomyces* [58,59]. Rok7B7 also controls the production of antibiotics such as actinorhodin (ACT, **8**), undecylprodigiosin (RED, **21**), and calcium-dependent antibiotic (CDA, **22**) in *Streptomyces coelicolor* M145 (Figure 3b). Chen et al. and Shu et al. reported that under stress associated with high concentrations of glutamate, AfsQ1/Q2 is important not only for maintaining the metabolic homeostasis of nutrient utilization but also for biosynthesis of antibiotics such as ACT, RED, CDA and the yellow pigment coelimycin P2 in *Streptomyces coelicolor* strains M145 [60,61].

Depletion of a metal component essential for growth activates the production of secondary metabolites with chelating activity, which enables the actinomycetes to acquire that metal (Figure 3c). Béchet and Blondeau reported that *Streptomyces aureofaciens* ATCC 10762 produces substantial amounts of tetracycline (**23**) only when the defined medium is deprived of iron [62]. Although not a compound obtained using approaches for activating silent genes, Wang et al. discovered SF2768, a metabolite produced by *Streptomyces thioluteus* that chelates copper as a chalkophore [63]. However, it has been suggested that activation of secondary metabolism in actinomycetes may depend on various metals. Indeed, it has been reported that adding to the culture medium rare metals not normally used as medium components can activate secondary metabolite biosynthetic genes. For example, Kamijo et al. found that *Streptomyces* sp. YB-1 produced reddish-purple pigment(s) only in the presence of a rare earth metal, ytterbium (Yb), but not in the presence of other metals. Structural analyses revealed that this pigment is a type of naphthoquinone similar to nanaomycin (**24**) [64]. Furthermore, Tanaka et al. showed that low concentrations of the rare earth elements scandium (Sc) and/or lanthanum (La) markedly activate (2.5 to 12-fold increase) the expression of nine secondary metabolite biosynthetic gene clusters in *Streptomyces coelicolor* A3(2) [65].

Furthermore, it has been reported that changing the pH of the culture medium can also result in the production of new substances (Figure 3d). Jiang et al. achieved a higher yield of validamycin A (**25**) from *Streptomyces hygroscopicus* 5008 together with more rapid protein synthesis and sugar consumption by subjecting cultures to one or more NaOH shocks [66]. Ren et al. obtained enhanced ε-poly-L-lysine (ε-PL, **26**) production by acidic pH shock of *Streptomyces* sp. M-Z18, achieving 52.5% higher production than controls without pH shock [67]. Conversely, the present authors were able to increase the production of antarlides, which have a 22-membered ring macrocyclic structure, by adding disodium hydrogen phosphate to the culture medium of *Streptomyces* sp. BB47 to prevent the pH of the culture medium from becoming too acidic [17]. In addition, Mehmood et al. reported that the oxygen supply controls the production of pristinamycins **27** in *Streptomyces pristinaespiralis* (Figure 3e) [68]. Takano et al. discovered the light-induced production of carotenoid pigments, including isorenieratene (**28**) and β-carotene (**29**), in *Streptomyces coelicolor* A3(2) (Figure 3f) [69]. Two transcriptional regulators, LitR and LitS, have been proposed as playing central roles in this light-induction mechanism in actinomycetes.

Thus, each element in the culture environment is important for the activation of silent genes. The authors' group has been working on methods to activate these genes by increasing the culture temperature above the normal level.

## Temperature – heat-shock metabolites (HSMs) – metabolites for which production is activated by high-temperature culture

Several examples of high-temperature culture of actinomycetes have been reported prior to the authors' examples as studies attempting to increase the production of the targeted metabolites (Figure 4a). For example, Doull et al. achieved an increase in the production of jadomycin B (**30**) produced by *Streptomyces venezuelae* ISP5230 by heat shock at 42 °C for 1 h [70]. Similarly, James et al. increased the production of granaticin (**31**) by *Streptomyces thermoviolaceus* NCIB 10076 by changing the culture temperature [71]. They showed that the yield was best at 45 °C, whereas the rate of synthesis was highest at 37 °C. However, there have been no exhaustive analyses of high-temperature culture of actinomycetes. Thus, Saito et al. decided to test on a large scale whether high-temperature culture of actinomycetes can activate silent genes.

First, an in-house library of 3,160 strains of actinomycetes widely isolated from Japanese soil was searched to identify actinomycete strains capable of growing at high temperatures (45 °C). This screening identified 57 actinomycete strains capable of growing at 45 °C. Comparative analysis of metabolites produced by these 57 actinomycetes at room temperature or high temperature revealed that almost half of the strains produced metabolites selectively at 45 °C. As these results were

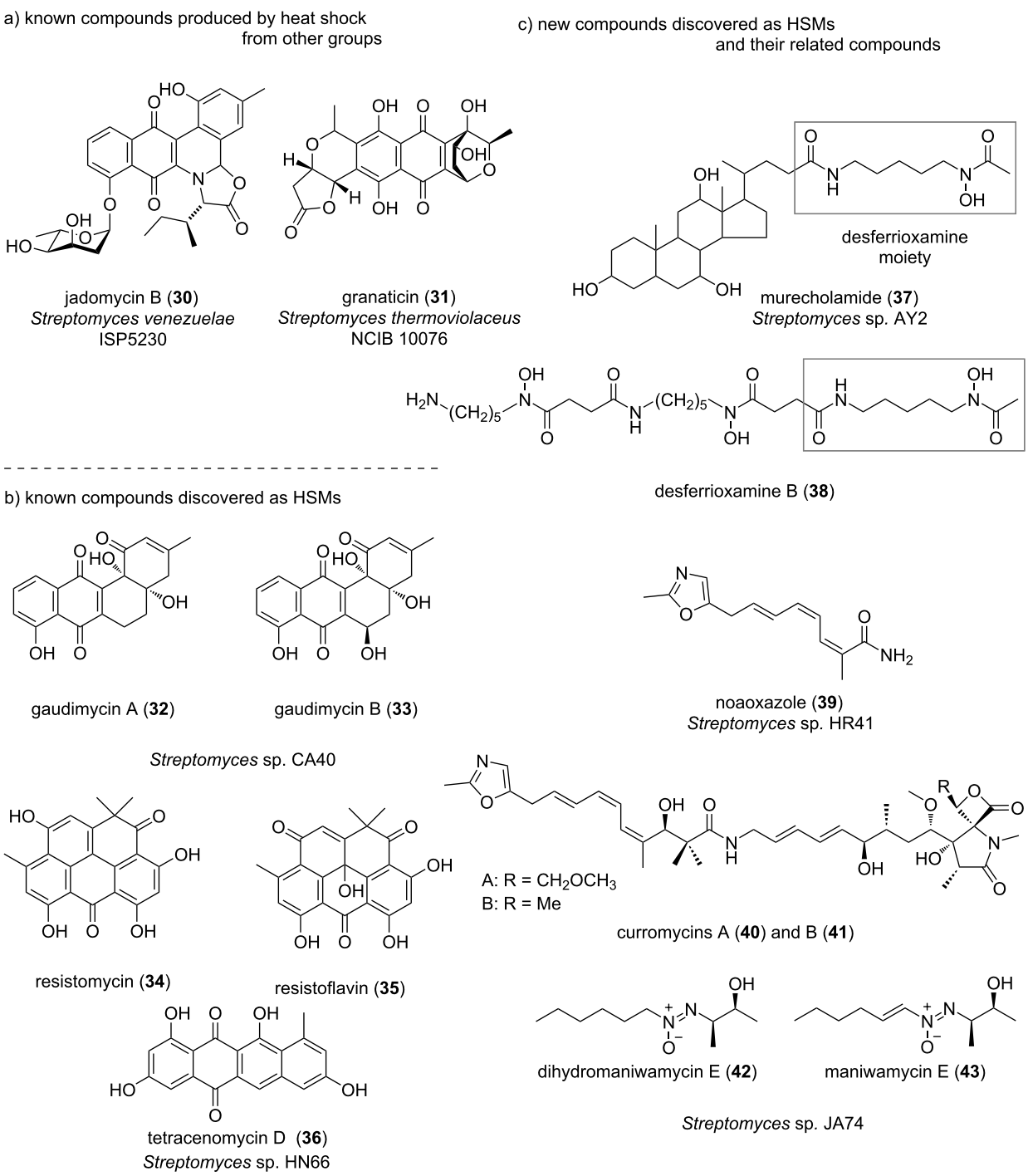



**Figure 4:** Structures of secondary metabolites obtained by high-temperature culture of actinomycetes.

obtained using one type of medium, a set number of incubation days, and one analytical condition using LCMS, it is possible that additional metabolites could be detected by examining a variety of conditions.

Sixteen of the abovementioned strains were randomly selected and cultured on five different media to obtain metabolic profile data according to culture temperature. This analysis resulted in the detection of newly isolated metabolites, metabolites exhibiting enhanced production, metabolites exhibiting diminished production, and metabolites in which production ceased. Plotting the newly identified metabolites according to retention time and molecular weight determined by LCMS, it was expected that most of the metabolites would tend to be rela-

tively polar and of low molecular weight. The results suggested that high-temperature culture induces the production of silent secondary metabolites in certain actinomycetes, and Saito et al. therefore designated these compounds HSMs [72]. Then, the structures of these HSMs are attempted to elucidate (Figure 4b,c).

Gaudimycins A and B (**32**, **33**) were identified as HSMs from *Streptomyces* sp. CA40. Palmu et al. discovered gaudimycins by heterologous expression of the biosynthetic gene clusters *pga* or *cab*, which were silent in the producing strain *Streptomyces* sp. PGA64 or *Streptomyces* sp. H021 relative to another host, *Streptomyces lividans* TK24 [73]. Resistomycin (**34**) [74], resistoflavin (**35**) [75], and tetracenomycin D (**36**) [76] were also identified as HSMs from *Streptomyces* sp. HN66. Carlson et al. reported the production of resistomycin by co-cultivating *Streptomyces* sp. B033 and *Burkholderia vietnamiensis* ATCC BAA-248 [77]. These results suggest that high-temperature culture of actinomycetes activates the production of silent secondary metabolites.

Further investigation of HSMs led to the discovery of three new compounds from different actinomycete strains. The first HSM is murecholamide (**37**), produced by *Streptomyces* sp. AY2 [72]. This substance has a skeleton in which a substructure of desferrioxamine (DFO, **38**), a well-known secondary metabolite [78], is condensed at the end of the steroid skeleton. It is thought that DFO is produced as a result of the activation of a silent gene, whereas the steroid skeleton may be a component of the culture medium. At a concentration of 182 μM, murecholamide exhibits biological activity in inhibiting the migration of HT29 cancer cells. The second HSM is noaoxazole (**39**), which was discovered in *Streptomyces* sp. HR41 [79]. This substance has a chemical structure with a methylated oxazole ring at the end of a chain-like structure. The structure of noaoxazole is similar to that of inthomycin and oxazolomycin [80,81], but the methylated oxazole ring is a relatively valuable structure found elsewhere only in curromycins A and B (**40**, **41**) and neocurromycin [82,83]. As with curromycin A, the end product of noaoxazole is expected to be different, but its discovery has not been reported to date. Noaoxazole has an activating effect on Notch signaling, which may result in inhibition of the malignant transformation of some cancers [84,85]. The third HSM is dihydromaniwamycin E (**42**), which was discovered in *Streptomyces* sp. JA74 [86]. Maniwamycin E (**43**) [87] was also discovered from this strain as a metabolite for which production is enhanced by high-temperature culture. These substances have an azoxy structure, which is relatively rare in the natural environment [88]. The stereochemistry of dihydromaniwamycin E was determined by total synthesis. The biological activities of these maniwamycins were also evaluated, and they showed antiviral activity against influenza (H1N1) virus and the causative agent of COVID-19, SARS-CoV-2. This was the first report of such activity for a natural product having an azoxy structure.

Saito et al. are now interested in determining how and why these HSMs are produced. First, when actinomycetes sense heat stress, repression mediated by heat shock sensor proteins such as HrcA, HspR, and RheA is released. The expression of various genes, such as those encoding heat shock proteins (HSPs), is then activated [89-91]. Lu et al. reported that the heat shock sensor HspR regulates the production of avermectin in *Streptomyces avermitilis* [92]. Further investigation is needed to determine if HSM production is induced after short periods of thermal stress and whether this production is regulated by such heat shock sensors. Although long-term high-temperature culture experiments are employed in this study, reactive oxygen species (ROS) may exert a stronger effect than high temperature. In actinomycetes, several sensor proteins for ROS have been reported (e.g., SoxR and OxyR), which upon sensing ROS, release their repression and activate the expression of various genes [93-95]. Wei et al. reported that the production of validamycin A (**25**) by *Streptomyces hygroscopicus* 5008 could be activated at the transcriptional level by simply adding hydrogen peroxide ($H_2O_2$; which generates ROS) to the culture medium [96]. In addition, there are also reports of cases in which denatured proteins are folded by molecular chaperones in high-temperature environments, and such activity may affect the production of secondary metabolites [97]. Saito et al. are currently attempting to determine how HSMs are produced.

Some studies have reported that secondary metabolites produced by actinomycetes are used physiologically by the producing strains and for communication with other strains. For example, Tenconi et al. reported that under certain conditions, the biosynthesis of undecylprodigiosin (**21**) is always triggered in the dying zone of the mycelial network of *Streptomyces coelicolor* M145 prior to morphological differentiation, right after an initial round of cell death [98]. They hypothesized that under certain circumstances, production of antiproliferative agents could play a role in the genetically programmed death of the producing organism. In addition, Nishiyama et al. suggested that actinorhodin (**8**) produced by *Streptomyces coelicolor* A3(2) functions as an organocatalyst to kill bacteria by catalyzing the production of toxic levels of $H_2O_2$ [99]. They also suggested that living organisms produce and use such metabolites as biocatalysts in nature. Therefore, it is thought that the production of HSMs has some physiological significance. For example, other microbial species reportedly activate energy metabolism and strengthen the cell membrane or walls

in high-temperature environments as a means of ensuring thermotolerance [100-102]. As this suggests that HSMs may support thermotolerance, Saito et al. are also currently trying to determine why HSMs are produced.

## Co-culture

It is also important to maintain the culture environment in a state close to that of the natural environment in order to activate silent genes in actinomycetes. Actinomycetes mainly inhabit soils, but they are widely symbiotic with plants [103], insects [104], and other organisms [105,106], and there are reports of their secondary metabolites controlling their own physiological functions or functioning as communication molecules [107,108]. Therefore, various studies have attempted to activate secondary metabolite biosynthetic genes under culture conditions that mimic the natural environment. In particular, the co-culture method, in which multiple microorganisms are cultured simultaneously, has been used in many studies examining silent gene activation (Figure 5a).

For example, Krespach et al. reported that *Streptomyces iranensis* secretes azalomycin F3a (**44**) in the presence of *Aspergillus nidulans* [109]. Interestingly, *A. nidulans* also reacted to the presence of azalomycin F3a by producing orsellinic acid, lecanoric acid, and the compounds F-9775A and F-9775B. Lee et al. discovered that secondary metabolism in *Streptomyces coelicolor* A3(2) M145 is activated by iron competition during co-culture with *Myxococcus xanthus* [110]. Co-culture promotes the production of the siderophore myxochelin in *M. xanthus*, which enhances the iron-scavenging capacity of *M. xanthus* and places *S. coelicolor* A3(2) M145 under iron starvation conditions. This chemical interaction activates the actinorhodin (**8**) biosynthetic gene cluster in *S. coelicolor* A3(2) M145. Onaka's group reported that co-cultivation of *Streptomyces* and mycolic acid-containing bacteria (MACB) activates secondary metabolism [111]. In their method, direct attachment by live MACB cells is thought to alter secondary metabolism in the *Streptomyces* cells [112]. Using this method, Onaka et al. identified a variety of new compounds, such as catenulobactins A and B (**45**, **46**) [113], dracolactams A and B (**47**, **48**) [114], and chojalactones A–C (**49**–**51**) [115]. Finally, although not an example of co-culture, Chevrette et al. found that the potential secondary metabolic capacity differs between soil-derived and insect-derived actinomycetes [116]. Insect-associated *Streptomyces* inhibit antimicrobial-resistant pathogens to a greater degree than soil *Streptomyces*. This suggests that the evolutionary trajectories of *Streptomyces* from the insect microbiome influence their biosynthetic potential and ability to inhibit resistant pathogens. Against this background, a unique case of activation of secondary metabolite biosynthetic genes by co-culture of actinomycetes and animal cells has been reported, which we outline in detail below.

If microorganisms and animal cells are used, one might first imagine the interaction between humans and intestinal bacteria. Indeed, several examples of co-culture of human cells and intestinal bacteria have been reported, and the changes in gene expression on the host side during such co-culture experiments have been well studied [117,118]. On the other hand, unlike intestinal bacteria, pathogenic microorganisms that adversely affect human health also interact with host cells when they invade the body [119,120]. Noting this, Ishibashi and Arai et al., hypothesized that when pathogenic microorganisms invade the body, they are subjected to xenocidal stress from immune cells, which activates silent genes and produces defensive compounds as a countermeasure. Therefore, the pathogenic actinomycetes were co-cultured with J774.1 mouse macrophage-like cells (Figure 5b).

The first compound isolated was nocarjamide (**52**), a cyclic nonapeptide, discovered by co-culturing *Nocardia tenerifensis* IFM 10554$^T$ with J774.1 cells [121]. Nocarjamide exhibited biological activity by activating Wnt signaling at an effective concentration of 20–40 μM. Wnt signaling plays an important role in a number of vital processes, such as the formation of various tissues and the differentiation/proliferation of eukaryotic cells [122]. It is interesting that substances that affect Wnt signaling can be produced in response to the interaction between animal cells and microorganisms. In addition, dehydropropylpantothenamide (**53**) was also identified using the same co-culture combination [123]. The second group of compounds identified included uniformides A and B (**54**, **55**), which were discovered by co-culturing *Nocardia uniformis* IFM0856$^T$ and J774.1 cells [124,125]. Uniformides were shown to suppress the production of nitric oxide, IL-6, and IL-1β by inhibiting the NF-κB pathway. Because NF-κB signaling plays a central role in the immune response [126], it is interesting that substances that regulate the pathway were produced by co-culture.

Thus, co-culture of animal cells and microorganisms may also be useful as a method for activating silent genes. However, it is important to determine what types of interactions between animal cells and microorganisms activate secondary metabolism in order to understand communication within the human body. For example, co-culture experiments using certain microorganisms have shown that physical contact between live microorganisms is important for the activation of secondary metabolism [112]. In addition, as mentioned above, various substances and changes in the concentrations of metal ions can affect secondary metabolism, and secondary metabolite production may thus be activated by these factors [110].

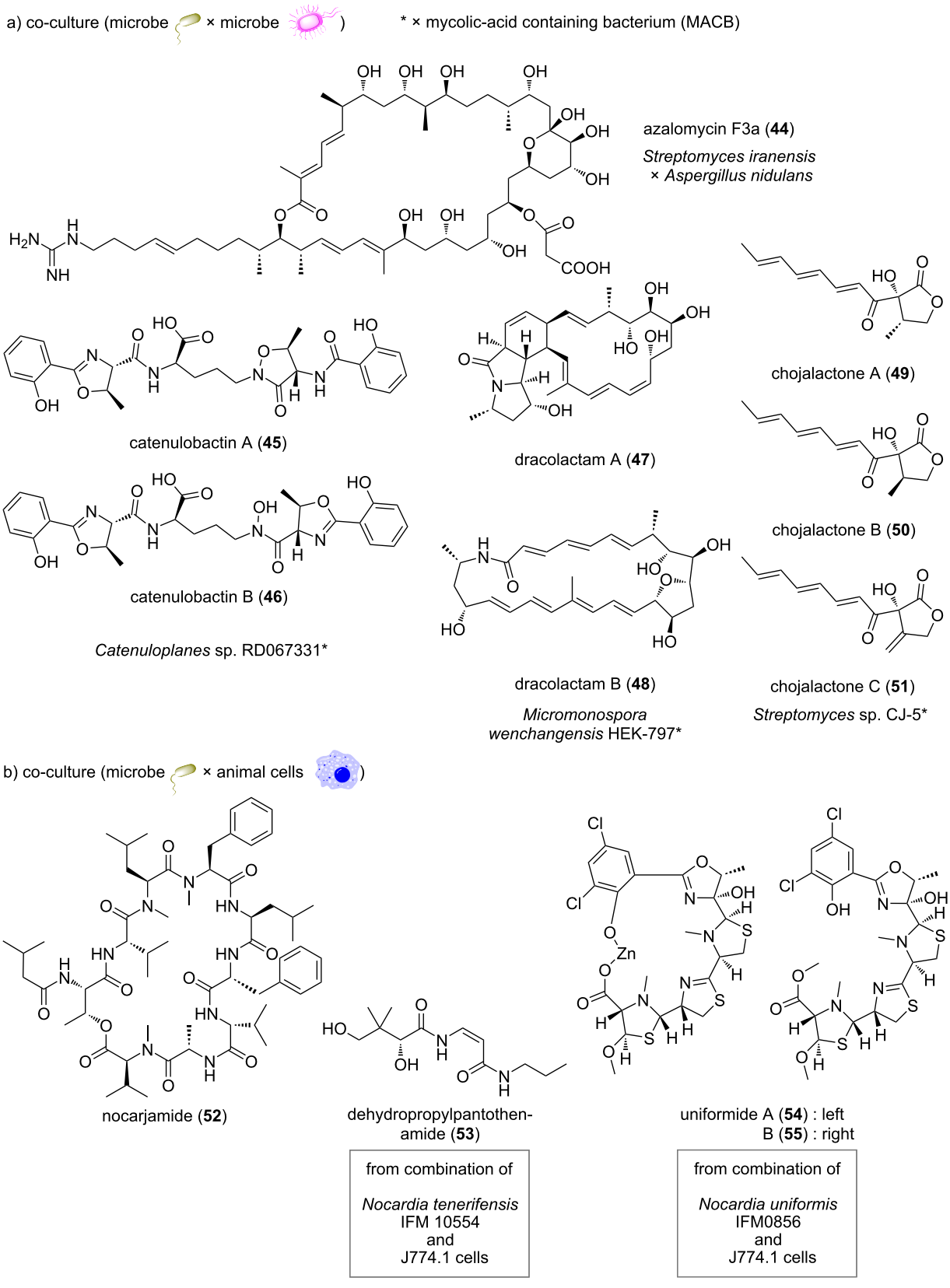



**Figure 5:** Structures of secondary metabolites obtained by co-culture of actinomycetes with other microorganisms or animal cells.

## Conclusion

Natural products hold great potential as leads in drug development and as useful tools in chemical biology research. Therefore, attempts to obtain novel secondary metabolites by activating the latent metabolic potential of actinomycetes are gaining momentum. In this review, methods used to search for new secondary metabolites, focusing mainly on high-temperature culture of actinomycetes and the co-culture of microorganisms and animal cells, were outlined. The mechanisms by which these silent secondary metabolites are produced should be specific to each method. Elucidation of these mechanisms will directly lead to the development of new silent gene activation methods. In these approaches, there are some disadvantages associated with limited adaptability of microbial species (high-temperature culture) and somewhat complex experimental manipulations (co-culture method). It is important to overcome these disadvantages, broaden the range of adaptable microbial species, and develop a simple method exhibiting high activation efficiency.

The metabolites produced by the activation of silent genes are expected to be physiologically significant. In the case of high-temperature culture or co-culture methods, these compounds may play roles in thermotolerance or communication between different species, respectively. Analyses of the functionality of such metabolites may also lead to the elucidation of new life phenomena. However, in recent years, it has not only become increasingly difficult to obtain novel secondary metabolites from natural products, but applied research using natural products, such as drug discovery studies, has been losing momentum. In drug discovery studies in the field of chemical biology, it may be important to investigate the compounds obtained by silent gene activation. It is expected that compounds useful in a variety of fields will be discovered by making full use of such silent gene activation methods.

## Acknowledgements

The graphical abstract was created by Science Graphics Co., Ltd. This content is not subject to CC BY 4.0.

## Funding

This study was supported by Grants-in-Aid for Early Career Scientists (nos. 20K15461 and 23K13897 to S.S.); a Grant-in-Aid for Scientific Research (B) (no. 21H02639 to M.A.A.) from the Japan Society for the Promotion of Science (JSPS); and Grants-in-Aid for Transformative Research Area (A) “Latent Chemical Space” (nos. 23H04880 and 23H04884 to M.A.A.) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. M.A.A. thanks the Mitsubishi Foundation. M.A.A. was the recipient of a JSBBA Innovative Research Program Award. S.S. was the recipient of a Noda Institute for Scientific Research (NISR) Young Investigator Research Grant.

## ORCID® iDs

Shun Saito - https://orcid.org/0000-0001-9270-4272
Midori A. Arai - https://orcid.org/0000-0003-0254-9550

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

1. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2020,** *83,* 770–803. doi:10.1021/acs.jnatprod.9b01285
2. Demain, A. L.; Sanchez, S. *J. Antibiot.* **2009,** *62,* 5–16. doi:10.1038/ja.2008.16
3. Gonzalez Ronquillo, M.; Angeles Hernandez, J. C. *Food Control* **2017,** *72B,* 255–267. doi:10.1016/j.foodcont.2016.03.001
4. Zimbro, M. J.; Power, D. A.; Miller, S. M.; Wilson, G. E.; Johnson, J. A., Eds. *DifcoTM & BBLTM Manual. Manual of Microbiological Culture Media,* 2nd ed.; Becton Dickinson and Company: Sparks, MD, USA, 2009.
5. Ryu, A. H.; Eckalbar, W. L.; Kreimer, A.; Yosef, N.; Ahituv, N. *Sci. Rep.* **2017,** *7,* 7533. doi:10.1038/s41598-017-07757-w
6. Cell Culture Antibiotic Selection Guide. https://www.sigmaaldrich.com/life-science/cell-culture/learning-center/antibiotic-selector.html (accessed Dec 7, 2023).
7. Villalobos-Delgado, L. H.; Nevárez-Moorillon, G. V.; Caro, I.; Quinto, E. J.; Mateo, J. Natural antimicrobial agents to improve foods shelf life. In *Food Quality and Shelf Life;* Galanakis, C. M., Ed.; Elsevier: Amsterdam, Netherlands, 2019; pp 125–157. doi:10.1016/b978-0-12-817190-5.00004-5
8. Srianta, I.; Ristiarini, S.; Nugerahani, I.; Sen, S. K.; Zhang, B. B.; Xu, G. R.; Blanc, P. J. *Int. Food Res. J.* **2014,** *21,* 1–12.
9. Purich, D. *The Inhibitor Index: A Desk Reference on Enzyme Inhibitors, Receptor Antagonists, Drugs, Toxins, Poisons, Biologics, and Therapeutic Leads,* 1st ed.; CRC Press: Boca Raton, FL, USA, 2017. doi:10.1201/9781315184289
10. Tyc, O.; Song, C.; Dickschat, J. S.; Vos, M.; Garbeva, P. *Trends Microbiol.* **2017,** *25,* 280–292. doi:10.1016/j.tim.2016.12.002
11. Schatz, A.; Bugle, E.; Waksman, S. A. *Proc. Soc. Exp. Biol. Med.* **1944,** *55,* 66–69. doi:10.3181/00379727-55-14461
12. Burg, R. W.; Miller, B. M.; Baker, E. E.; Birnbaum, J.; Currie, S. A.; Hartman, R.; Kong, Y.-L.; Monaghan, R. L.; Olson, G.; Putter, I.; Tunac, J. B.; Wallick, H.; Stapley, E. O.; Oiwa, R.; Ōmura, S. *Antimicrob. Agents Chemother.* **1979,** *15,* 361–367. doi:10.1128/aac.15.3.361
13. Watve, M. G.; Tickoo, R.; Jog, M. M.; Bhole, B. D. *Arch. Microbiol.* **2001,** *176,* 386–390. doi:10.1007/s002030100345
14. Lacey, H. J.; Rutledge, P. J. *Molecules* **2022,** *27,* 887. doi:10.3390/molecules27030887
15. Tashiro, E.; Imoto, M. *Arch. Pharmacal Res.* **2015,** *38,* 1651–1660. doi:10.1007/s12272-015-0633-4
16. Kawamura, T.; Fujimaki, T.; Hamanaka, N.; Torii, K.; Kobayashi, H.; Takahashi, Y.; Igarashi, M.; Kinoshita, N.; Nishimura, Y.; Tashiro, E.; Imoto, M. *J. Antibiot.* **2010,** *63,* 601–605. doi:10.1038/ja.2010.98

17. Saito, S.; Fujimaki, T.; Panbangred, W.; Igarashi, Y.; Imoto, M. *Angew. Chem., Int. Ed.* **2016,** *55,* 2728–2732. doi:10.1002/anie.201510079
18. Fujimaki, T.; Saito, S.; Imoto, M. *J. Antibiot.* **2017,** *70,* 328–330. doi:10.1038/ja.2016.162
19. Saito, S.; Fujimaki, T.; Panbangred, W.; Sawa, R.; Igarashi, Y.; Imoto, M. *J. Antibiot.* **2017,** *70,* 595–600. doi:10.1038/ja.2017.6
20. Tiwari, K.; Gupta, R. K. *Crit. Rev. Biotechnol.* **2012,** *32,* 108–132. doi:10.3109/07388551.2011.562482
21. Saito, S.; Atsumi, K.; Zhou, T.; Fukaya, K.; Urabe, D.; Oku, N.; Karim, M. R. U.; Komaki, H.; Igarashi, Y. *Beilstein J. Org. Chem.* **2020,** *16,* 1100–1110. doi:10.3762/bjoc.16.97
22. Saito, S.; Indo, K.; Oku, N.; Komaki, H.; Kawasaki, M.; Igarashi, Y. *Beilstein J. Org. Chem.* **2021,** *17,* 2939–2949. doi:10.3762/bjoc.17.203
23. Saito, S.; Oku, N.; Igarashi, Y. *J. Antibiot.* **2022,** *75,* 44–47. doi:10.1038/s41429-021-00474-7
24. Saito, S.; Xiaohanyao, Y.; Zhou, T.; Nakajima-Shimada, J.; Tashiro, E.; Triningsih, D. W.; Harunari, E.; Oku, N.; Igarashi, Y. *J. Nat. Prod.* **2022,** *85,* 1697–1703. doi:10.1021/acs.jnatprod.2c00137
25. Igarashi, Y. *J. Antibiot.* **2023,** *76,* 365–383. doi:10.1038/s41429-023-00625-y
26. Ohnishi, Y.; Ishikawa, J.; Hara, H.; Suzuki, H.; Ikenoya, M.; Ikeda, H.; Yamashita, A.; Hattori, M.; Horinouchi, S. *J. Bacteriol.* **2008,** *190,* 4050–4060. doi:10.1128/jb.00204-08
27. Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M.-A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A. *Nature* **2002,** *417,* 141–147. doi:10.1038/417141a
28. Ikeda, H.; Shin-ya, K.; Omura, S. *J. Ind. Microbiol. Biotechnol.* **2014,** *41,* 233–250. doi:10.1007/s10295-013-1327-x
29. Liu, Y.; Zhou, H.; Shen, Q.; Dai, G.; Yan, F.; Li, X.; Ren, X.; Sun, Q.; Tang, Y.-J.; Zhang, Y.; Bian, X. *Org. Lett.* **2021,** *23,* 6967–6971. doi:10.1021/acs.orglett.1c02589
30. Liu, S. H.; Wang, W.; Wang, K. B.; Zhang, B.; Li, W.; Shi, J.; Jiao, R. H.; Tan, R. X.; Ge, H. M. *Org. Lett.* **2019,** *21,* 3785–3788. doi:10.1021/acs.orglett.9b01237
31. Khokhlov, A. S.; Tovarova, I. I.; Borisova, L. N.; Pliner, S. A.; Shevchenko, L. N.; Kornitskaia, E. Y.; Ivkina, N. S.; Rapoport, I. A. *Dokl. Akad. Nauk SSSR* **1967,** *177,* 232–235.
32. Takano, E.; Nihira, T.; Hara, Y.; Jones, J. J.; Gershater, C. J. L.; Yamada, Y.; Bibb, M. *J. Biol. Chem.* **2000,** *275,* 11010–11016. doi:10.1074/jbc.275.15.11010
33. Onaka, H.; Ando, N.; Nihira, T.; Yamada, Y.; Beppu, T.; Horinouchi, S. *J. Bacteriol.* **1995,** *177,* 6083–6092. doi:10.1128/jb.177.21.6083-6092.1995
34. Takano, E.; Chakraburtty, R.; Nihira, T.; Yamada, Y.; Bibb, M. J. *Mol. Microbiol.* **2001,** *41,* 1015–1028. doi:10.1046/j.1365-2958.2001.02562.x
35. HARA, O.; BEPPU, T. *J. Antibiot.* **1982,** *35,* 349–358. doi:10.7164/antibiotics.35.349
36. Butler, M. J.; Takano, E.; Bruheim, P.; Jovetic, S.; Marinelli, F.; Bibb, M. J. *Appl. Microbiol. Biotechnol.* **2003,** *61,* 512–516. doi:10.1007/s00253-003-1277-8
37. Kitani, S.; Miyamoto, K. T.; Takamatsu, S.; Herawati, E.; Iguchi, H.; Nishitomi, K.; Uchida, M.; Nagamitsu, T.; Omura, S.; Ikeda, H.; Nihira, T. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108,* 16410–16415. doi:10.1073/pnas.1113908108
38. Ohnishi, Y.; Furusho, Y.; Higashi, T.; Chun, H. K.; Furihata, K.; Sakuda, S.; Horinouchi, S. *J. Antibiot.* **2004,** *57,* 218–223. doi:10.7164/antibiotics.57.218
39. Takano, E.; Kinoshita, H.; Mersinias, V.; Bucca, G.; Hotchkiss, G.; Nihira, T.; Smith, C. P.; Bibb, M.; Wohlleben, W.; Chater, K. *Mol. Microbiol.* **2005,** *56,* 465–479. doi:10.1111/j.1365-2958.2005.04543.x
40. Sultan, S. P.; Kitani, S.; Miyamoto, K. T.; Iguchi, H.; Atago, T.; Ikeda, H.; Nihira, T. *Appl. Microbiol. Biotechnol.* **2016,** *100,* 9581–9591. doi:10.1007/s00253-016-7781-4
41. Wilbanks, L. E.; Hennigan, H. E.; Martinez-Brokaw, C. D.; Lakkis, H.; Thormann, S.; Eggly, A. S.; Buechel, G.; Parkinson, E. I. *ACS Chem. Biol.* **2023,** *18,* 1624–1631. doi:10.1021/acschembio.3c00241
42. Barajas, J. F.; Blake-Hedges, J. M.; Bailey, C. B.; Curran, S.; Keasling, J. D. *Synth. Syst. Biotechnol.* **2017,** *2,* 147–166. doi:10.1016/j.synbio.2017.08.005
43. Du, D.; Katsuyama, Y.; Onaka, H.; Fujie, M.; Satoh, N.; Shin-ya, K.; Ohnishi, Y. *ChemBioChem* **2016,** *17,* 1464–1471. doi:10.1002/cbic.201600167
44. Du, D.; Katsuyama, Y.; Shin-ya, K.; Ohnishi, Y. *Angew. Chem., Int. Ed.* **2018,** *57,* 1954–1957. doi:10.1002/anie.201709636
45. Du, D.; Katsuyama, Y.; Horiuchi, M.; Fushinobu, S.; Chen, A.; Davis, T. D.; Burkart, M. D.; Ohnishi, Y. *Nat. Chem. Biol.* **2020,** *16,* 776–782. doi:10.1038/s41589-020-0530-0
46. Seo, M.; Du, D.; Katsuyama, Y.; Katsuta, R.; Yajima, A.; Nukada, T.; Ohnishi, Y.; Ishigami, K. *Biosci., Biotechnol., Biochem.* **2021,** *85,* 148–153. doi:10.1093/bbb/zbaa023
47. Ochi, K.; Okamoto, S.; Tozawa, Y.; Inaoka, T.; Hosaka, T.; Xu, J.; Kurosawa, K. *Adv. Appl. Microbiol.* **2004,** *56,* 155–184. doi:10.1016/s0065-2164(04)56005-7
48. Liu, L.; Li, S.; Sun, R.; Qin, X.; Ju, J.; Zhang, C.; Duan, Y.; Huang, Y. *Org. Lett.* **2020,** *22,* 4614–4619. doi:10.1021/acs.orglett.0c01224
49. Seyedsayamdost, M. R. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111,* 7266–7271. doi:10.1073/pnas.1400019111
50. Han, E. J.; Lee, S. R.; Townsend, C. A.; Seyedsayamdost, M. R. *ACS Chem. Biol.* **2023,** *18,* 1854–1862. doi:10.1021/acschembio.3c00281
51. Bode, H. B.; Bethe, B.; Höfs, R.; Zeeck, A. *ChemBioChem* **2002,** *3,* 619–627. doi:10.1002/1439-7633(20020703)3:7<619::aid-cbic619>3.0.co;2-9
52. Schiewe, H.-J.; Zeeck, A. *J. Antibiot.* **1999,** *52,* 635–642. doi:10.7164/antibiotics.52.635
53. Puder, C.; Krastel, P.; Zeeck, A. *J. Nat. Prod.* **2000,** *63,* 1258–1260. doi:10.1021/np0001373
54. Puder, C.; Loya, S.; Hizi, A.; Zeeck, A. *J. Nat. Prod.* **2001,** *64,* 42–45. doi:10.1021/np000377i
55. Rateb, M. E.; Houssen, W. E.; Harrison, W. T. A.; Deng, H.; Okoro, C. K.; Asenjo, J. A.; Andrews, B. A.; Bull, A. T.; Goodfellow, M.; Ebel, R.; Jaspars, M. *J. Nat. Prod.* **2011,** *74,* 1965–1971. doi:10.1021/np200470u
56. Sproule, A.; Correa, H.; Decken, A.; Haltli, B.; Berrué, F.; Overy, D. P.; Kerr, R. G. *Mar. Drugs* **2019,** *17,* 347. doi:10.3390/md17060347
57. Martín, J. F.; Liras, P. *Front. Microbiol.* **2020,** *10,* 3120. doi:10.3389/fmicb.2019.03120

58. Swiatek, M. A.; Gubbens, J.; Bucca, G.; Song, E.; Yang, Y.-H.; Laing, E.; Kim, B.-G.; Smith, C. P.; van Wezel, G. P. *J. Bacteriol.* **2013,** *195,* 1236–1248. doi:10.1128/jb.02191-12
59. Kim, S. H.; Traag, B. A.; Hasan, A. H.; McDowall, K. J.; Kim, B.-G.; van Wezel, G. P. *Antonie van Leeuwenhoek* **2015,** *108,* 201–213. doi:10.1007/s10482-015-0479-2
60. Shu, D.; Chen, L.; Wang, W.; Yu, Z.; Ren, C.; Zhang, W.; Yang, S.; Lu, Y.; Jiang, W. *Appl. Microbiol. Biotechnol.* **2009,** *81,* 1149–1160. doi:10.1007/s00253-008-1738-1
61. Chen, S.; Zheng, G.; Zhu, H.; He, H.; Chen, L.; Zhang, W.; Jiang, W.; Lu, Y. *FEMS Microbiol. Lett.* **2016,** *363,* fnw160. doi:10.1093/femsle/fnw160
62. Béchet, M.; Blondeau, R. *J. Appl. Microbiol.* **1998,** *84,* 889–894. doi:10.1046/j.1365-2672.1998.00428.x
63. Wang, L.; Zhu, M.; Zhang, Q.; Zhang, X.; Yang, P.; Liu, Z.; Deng, Y.; Zhu, Y.; Huang, X.; Han, L.; Li, S.; He, J. *ACS Chem. Biol.* **2017,** *12,* 3067–3075. doi:10.1021/acschembio.7b00897
64. Kamijo, M.; Suzuki, T.; Kawai, K.; Fujii, T.; Murase, H. *J. Biosci. Bioeng.* **1999,** *87,* 340–343. doi:10.1016/s1389-1723(99)80042-5
65. Tanaka, Y.; Hosaka, T.; Ochi, K. *J. Antibiot.* **2010,** *63,* 477–481. doi:10.1038/ja.2010.53
66. Jiang, J.; Sun, Y.-F.; Tang, X.; He, C.-N.; Shao, Y.-L.; Tang, Y.-J.; Zhou, W.-W. *Bioresour. Technol.* **2018,** *249,* 234–240. doi:10.1016/j.biortech.2017.10.012
67. Ren, X.-D.; Chen, X.-S.; Zeng, X.; Wang, L.; Tang, L.; Mao, Z.-G. *Bioprocess Biosyst. Eng.* **2015,** *38,* 1113–1125. doi:10.1007/s00449-015-1354-2
68. Mehmood, N.; Olmos, E.; Goergen, J.-L.; Blanchard, F.; Ullisch, D.; Klöckner, W.; Büchs, J.; Delaunay, S. *Biotechnol. Bioeng.* **2011,** *108,* 2151–2161. doi:10.1002/bit.23177
69. Takano, H.; Obitsu, S.; Beppu, T.; Ueda, K. *J. Bacteriol.* **2005,** *187,* 1825–1832. doi:10.1128/jb.187.5.1825-1832.2005
70. Doull, J. L.; Ayer, S. W.; Singh, A. K.; Thibault, P. *J. Antibiot.* **1993,** *46,* 869–871. doi:10.7164/antibiotics.46.869
71. James, P. D. A.; Edwards, C. *J. Gen. Microbiol.* **1989,** *135,* 1997–2003. doi:10.1099/00221287-135-7-1997
72. Saito, S.; Kato, W.; Ikeda, H.; Katsuyama, Y.; Ohnishi, Y.; Imoto, M. *J. Antibiot.* **2020,** *73,* 203–210. doi:10.1038/s41429-020-0279-4
73. Palmu, K.; Ishida, K.; Mäntsälä, P.; Hertweck, C.; Metsä-Ketelä, M. *ChemBioChem* **2007,** *8,* 1577–1584. doi:10.1002/cbic.200700140
74. Brockmann, H.; Schmidt-Kastner, G. *Naturwissenschaften* **1951,** *38,* 479–480. doi:10.1007/bf00622090
75. Eckardt, K.; Bradler, G.; Fritzsche, H.; Tresselt, D. *Z. Chem.* **1970,** *10,* 221. doi:10.1002/zfch.19700100607
76. Yue, S.; Motamedi, H.; Wendt-Pienkowski, E.; Hutchinson, C. R. *J. Bacteriol.* **1986,** *167,* 581–586. doi:10.1128/jb.167.2.581-586.1986
77. Carlson, S.; Tanouye, U.; Omarsdottir, S.; Murphy, B. T. *J. Nat. Prod.* **2015,** *78,* 381–387. doi:10.1021/np500767u
78. Barona-Gómez, F.; Wong, U.; Giannakopulos, A. E.; Derrick, P. J.; Challis, G. L. *J. Am. Chem. Soc.* **2004,** *126,* 16282–16283. doi:10.1021/ja045774k
79. Saito, S.; Suzuki, S.; Arai, M. A. *J. Antibiot.* **2022,** *75,* 509–513. doi:10.1038/s41429-022-00551-5
80. Henkel, T.; Zeeck, A. *Liebigs Ann. Chem.* **1991,** 367–373. doi:10.1002/jlac.199119910164
81. Mori, T.; Takahashi, K.; Kashiwabara, M.; Uemura, D.; Katayama, C.; Iwadare, S.; Shizuri, Y.; Mitomo, R.; Nakano, F.; Matsuzaki, A. *Tetrahedron Lett.* **1985,** *26,* 1073–1076. doi:10.1016/s0040-4039(00)98515-0
82. Ogura, M.; Nakayama, H.; Furihata, K.; Shimazu, A.; Seto, H.; Otake, N. *J. Antibiot.* **1985,** *38,* 669–673. doi:10.7164/antibiotics.38.669
83. Hayakawa, Y.; Yaguchi, R.; Akimoto, M.; Kimata, S.; Shin-ya, K. *J. Antibiot.* **2020,** *73,* 790–793. doi:10.1038/s41429-020-0339-9
84. Rampias, T.; Vgenopoulou, P.; Avgeris, M.; Polyzos, A.; Stravodimos, K.; Valavanis, C.; Scorilas, A.; Klinakis, A. *Nat. Med.* **2014,** *20,* 1199–1205. doi:10.1038/nm.3678
85. Loganathan, S. K.; Schleicher, K.; Malik, A.; Quevedo, R.; Langille, E.; Teng, K.; Oh, R. H.; Rathod, B.; Tsai, R.; Samavarchi-Tehrani, P.; Pugh, T. J.; Gingras, A.-C.; Schramek, D. *Science* **2020,** *367,* 1264–1269. doi:10.1126/science.aax0902
86. Saito, S.; Funayama, K.; Kato, W.; Okuda, M.; Kawamoto, M.; Matsubara, T.; Sato, T.; Sato, A.; Otsuguro, S.; Sasaki, M.; Orba, Y.; Sawa, H.; Maenaka, K.; Shindo, K.; Imoto, M.; Arai, M. A. *J. Nat. Prod.* **2022,** *85,* 2583–2591. doi:10.1021/acs.jnatprod.2c00550
87. Fukumoto, A.; Murakami, C.; Anzai, Y.; Kato, F. *J. Antibiot.* **2016,** *69,* 395–399. doi:10.1038/ja.2015.126
88. Wibowo, M.; Ding, L. *J. Nat. Prod.* **2020,** *83,* 3482–3491. doi:10.1021/acs.jnatprod.0c00725
89. Grandvalet, C.; Servant, P.; Mazodier, P. *Mol. Microbiol.* **1997,** *23,* 77–84. doi:10.1046/j.1365-2958.1997.1811563.x
90. Grandvalet, C.; Rapoport, G.; Mazodier, P. *J. Bacteriol.* **1998,** *180,* 5129–5134. doi:10.1128/jb.180.19.5129-5134.1998
91. Servant, P.; Rapoport, G.; Mazodier, P. *Microbiology (London, U. K.)* **1999,** *145,* 2385–2391. doi:10.1099/00221287-145-9-2385
92. Lu, X.; Wang, Q.; Yang, M.; Chen, Z.; Li, J.; Wen, Y. *Appl. Environ. Microbiol.* **2021,** *87,* e00473-21. doi:10.1128/aem.00473-21
93. Hahn, J.-S.; Oh, S.-Y.; Chater, K. F.; Cho, Y.-H.; Roe, J.-H. *J. Biol. Chem.* **2000,** *275,* 38254–38260. doi:10.1074/jbc.m006079200
94. Hahn, J.-S.; Oh, S.-Y.; Roe, J.-H. *J. Bacteriol.* **2002,** *184,* 5214–5222. doi:10.1128/jb.184.19.5214-5222.2002
95. Dela Cruz, R.; Gao, Y.; Penumetcha, S.; Sheplock, R.; Weng, K.; Chander, M. *J. Bacteriol.* **2010,** *192,* 6428–6438. doi:10.1128/jb.00916-10
96. Wei, Z.-H.; Bai, L.; Deng, Z.; Zhong, J.-J. *Bioresour. Technol.* **2011,** *102,* 1783–1787. doi:10.1016/j.biortech.2010.08.114
97. Kudo, H.; Arakawa, G.-y.; Shirai, S.; Ogawa, M.; Shindo, H.; Hosaka, M.; Tokuoka, M. *J. Biosci. Bioeng.* **2022,** *133,* 235–242. doi:10.1016/j.jbiosc.2021.12.001
98. Tenconi, E.; Traxler, M. F.; Hoebreck, C.; van Wezel, G. P.; Rigali, S. *Front. Microbiol.* **2018,** *9,* 1742. doi:10.3389/fmicb.2018.01742
99. Nishiyama, T.; Hashimoto, Y.; Kusakabe, H.; Kumano, T.; Kobayashi, M. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111,* 17152–17157. doi:10.1073/pnas.1417941111
100. Sun, Q.-l.; Sun, Y.-y.; Zhang, J.; Luan, Z.-d.; Lian, C.; Liu, S.-q.; Yu, C. *J. Proteomics* **2019,** *203,* 103380. doi:10.1016/j.jprot.2019.103380
101. Péter, M.; Gudmann, P.; Kóta, Z.; Török, Z.; Vígh, L.; Glatz, A.; Balogh, G. *Int. J. Mol. Sci.* **2021,** *22,* 13272. doi:10.3390/ijms222413272
102. Wang, L.; Li, A.; Fang, J.; Wang, Y.; Chen, L.; Qiao, L.; Wang, W. *Int. J. Mol. Sci.* **2023,** *24,* 11822. doi:10.3390/ijms241411822
103. Ayswaria, R.; Vasu, V.; Krishna, R. *Crit. Rev. Microbiol.* **2020,** *46,* 750–758. doi:10.1080/1040841x.2020.1828816
104. Van Moll, L.; De Smet, J.; Cos, P.; Van Campenhout, L. *Crit. Rev. Microbiol.* **2021,** *47,* 562–579. doi:10.1080/1040841x.2021.1907302
105. Sarmiento-Vizcaíno, A.; García, L. A.; Blanco, G. *Arch. Microbiol.* **2023,** *205,* 81. doi:10.1007/s00203-023-03422-1

106. Chen, J.; Xu, L.; Zhou, Y.; Han, B. *Mar. Drugs* **2021,** *19,* 629. doi:10.3390/md19110629
107. Takano, H.; Nishiyama, T.; Amano, S.-i.; Beppu, T.; Kobayashi, M.; Ueda, K. *J. Ind. Microbiol. Biotechnol.* **2016,** *43,* 143–148. doi:10.1007/s10295-015-1680-z
108. Salwan, R.; Sharma, V. *Microbiol. Res.* **2020,** *231,* 126374. doi:10.1016/j.micres.2019.126374
109. Krespach, M. K. C.; Stroe, M. C.; Netzker, T.; Rosin, M.; Zehner, L. M.; Komor, A. J.; Beilmann, J. M.; Krüger, T.; Scherlach, K.; Kniemeyer, O.; Schroeckh, V.; Hertweck, C.; Brakhage, A. A. *Nat. Microbiol.* **2023,** *8,* 1348–1361. doi:10.1038/s41564-023-01382-2
110. Lee, N.; Kim, W.; Chung, J.; Lee, Y.; Cho, S.; Jang, K.-S.; Kim, S. C.; Palsson, B.; Cho, B.-K. *ISME J.* **2020,** *14,* 1111–1124. doi:10.1038/s41396-020-0594-6
111. Onaka, H.; Mori, Y.; Igarashi, Y.; Furumai, T. *Appl. Environ. Microbiol.* **2011,** *77,* 400–406. doi:10.1128/aem.01337-10
112. Asamizu, S.; Ozaki, T.; Teramoto, K.; Satoh, K.; Onaka, H. *PLoS One* **2015,** *10,* e0142372. doi:10.1371/journal.pone.0142372
113. Hoshino, S.; Ozeki, M.; Awakawa, T.; Morita, H.; Onaka, H.; Abe, I. *J. Nat. Prod.* **2018,** *81,* 2106–2110. doi:10.1021/acs.jnatprod.8b00261
114. Hoshino, S.; Okada, M.; Awakawa, T.; Asamizu, S.; Onaka, H.; Abe, I. *Org. Lett.* **2017,** *19,* 4992–4995. doi:10.1021/acs.orglett.7b02508
115. Hoshino, S.; Wakimoto, T.; Onaka, H.; Abe, I. *Org. Lett.* **2015,** *17,* 1501–1504. doi:10.1021/acs.orglett.5b00385
116. Chevrette, M. G.; Carlson, C. M.; Ortega, H. E.; Thomas, C.; Ananiev, G. E.; Barns, K. J.; Book, A. J.; Cagnazzo, J.; Carlos, C.; Flanigan, W.; Grubbs, K. J.; Horn, H. A.; Hoffmann, F. M.; Klassen, J. L.; Knack, J. J.; Lewin, G. R.; McDonald, B. R.; Muller, L.; Melo, W. G. P.; Pinto-Tomás, A. A.; Schmitz, A.; Wendt-Pienkowski, E.; Wildman, S.; Zhao, M.; Zhang, F.; Bugni, T. S.; Andes, D. R.; Pupo, M. T.; Currie, C. R. *Nat. Commun.* **2019,** *10,* 516. doi:10.1038/s41467-019-08438-0
117. Zhang, J.; Hernandez-Gordillo, V.; Trapecar, M.; Wright, C.; Taketani, M.; Schneider, K.; Chen, W. L. K.; Stas, E.; Breault, D. T.; Carrier, R. L.; Voigt, C. A.; Griffith, L. G. *Nat. Protoc.* **2021,** *16,* 3874–3900. doi:10.1038/s41596-021-00562-w
118. Nie, J.; Liao, W.; Zhang, Z.; Zhang, M.; Wen, Y.; Capanoglu, E.; Sarker, M. M. R.; Zhu, R.; Zhao, C. *Curr. Res. Food Sci.* **2023,** *6,* 100402. doi:10.1016/j.crfs.2022.11.021
119. Toyotome, T.; Saito, S.; Koshizaki, Y.; Komatsu, R.; Matsuzawa, T.; Yaguchi, T. *J. Infect. Chemother.* **2020,** *26,* 321–323. doi:10.1016/j.jiac.2019.09.002
120. Hamada, M.; Enomoto, N.; Yamashita, T.; Shimojima, M.; Tanno, D.; Ohana, N.; Toyokawa, M.; Takahashi, H.; Yaguchi, T. *Int. J. Syst. Evol. Microbiol.* **2023,** *73,* 005935. doi:10.1099/ijsem.0.005935
121. Hara, Y.; Arai, M. A.; Toume, K.; Masu, H.; Sato, T.; Komatsu, K.; Yaguchi, T.; Ishibashi, M. *Org. Lett.* **2018,** *20,* 5831–5834. doi:10.1021/acs.orglett.8b02522
122. Cadigan, K. M.; Liu, Y. I. *J. Cell Sci.* **2006,** *119,* 395–402. doi:10.1242/jcs.02826
123. Hara, Y.; Arai, M. A.; Sakai, K.; Ishikawa, N.; Gonoi, T.; Yaguchi, T.; Ishibashi, M. *J. Nat. Med.* **2018,** *72,* 280–289. doi:10.1007/s11418-017-1161-y
124. Hara, Y.; Watanabe, K.; Takaya, A.; Manome, T.; Yaguchi, T.; Ishibashi, M. *Org. Lett.* **2022,** *24,* 4998–5002. doi:10.1021/acs.orglett.2c02092
125. Hara, Y.; Watanabe, K.; Takaya, A.; Ebihara, I.; Manome, T.; Arai, M. A.; Yaguchi, T.; Ishibashi, M. *Org. Lett.* **2022,** *24,* 5867. doi:10.1021/acs.orglett.2c02441
126. Mussbacher, M.; Derler, M.; Basílio, J.; Schmid, J. A. *Front. Immunol.* **2023,** *14,* 1134661. doi:10.3389/fimmu.2023.1134661

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.69

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Discovery and biosynthesis of bacterial drimane-type sesquiterpenoids from *Streptomyces clavuligerus*

Dongxu Zhang[1], Wenyu Du[1], Xingming Pan[1], Xiaoxu Lin[1], Fang-Ru Li[1], Qingling Wang[1], Qian Yang[1], Hui-Min Xu[2] and Liao-Bin Dong[*1]

Full Research Paper

Open Access

Address:
[1]State Key Laboratory of Natural Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing 211198, China and [2]The Public Laboratory Platform, China Pharmaceutical University, Nanjing 211198, China

Email:
Liao-Bin Dong[*] - ldong@cpu.edu.cn

* Corresponding author

Keywords:
bacterial terpenoid; cytochrome P450s; drimane-type sesquiterpenoid; *Streptomyces clavuligerus*; terpenoid biosynthesis

*Beilstein J. Org. Chem.* **2024,** *20,* 815–822.
https://doi.org/10.3762/bjoc.20.73

Received: 22 December 2023
Accepted: 04 April 2024
Published: 16 April 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra

© 2024 Zhang et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Drimane-type sesquiterpenoids (DMTs) are characterized by a distinctive 6/6 bicyclic skeleton comprising the A and B rings. While DMTs are commonly found in fungi and plants, their presence in bacteria has not been reported. Moreover, the biosynthetic pathways for DMTs have been primarily elucidated in fungi, with identified P450s only acting on the B ring. In this study, we isolated and characterized three bacterial DMTs, namely 3β-hydroxydrimenol (**2**), 2α-hydroxydrimenol (**3**), and 3-ketodrimenol (**4**), from *Streptomyces clavuligerus*. Through genome mining and heterologous expression, we identified a *cav* biosynthetic gene cluster responsible for the biosynthesis of DMTs **2**–**4**, along with a P450, CavA, responsible for introducing the C-2 and C-3 hydroxy groups. Furthermore, the substrate scope of CavA revealed its ability to hydroxylate drimenol analogs. This discovery not only broadens the known chemical diversity of DMTs from bacteria, but also provides new insights into DMT biosynthesis in bacteria.

## Introduction

Terpenoids, encompassing over 11,000 compounds (http://dnp.chemnbase.com), are the most diverse group of natural products found in nature [1]. All terpenoids are biosynthesized from $C_5$ carbon units, which are sourced from the isoprenoid building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) from the mevalonate (MVA) and 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathways [2,3]. Sesquiterpenoids, synthesized from farnesyl diphosphate (FPP) which is composed of three $C_5$ units, hold significant biological and industrial value, particularly in the realm of perfumery [4]. Among these, drimane-type sesquiterpenoids (DMTs) are distinct due to their chemical structures, which feature a decahydronaphthalene core adorned with methyl groups, mirroring the A/B rings found in labdane-derived diterpenoids [5,6] (Figure 1a). DMTs exhibit significant biological activities, such as those found in polygodial, which serves as an

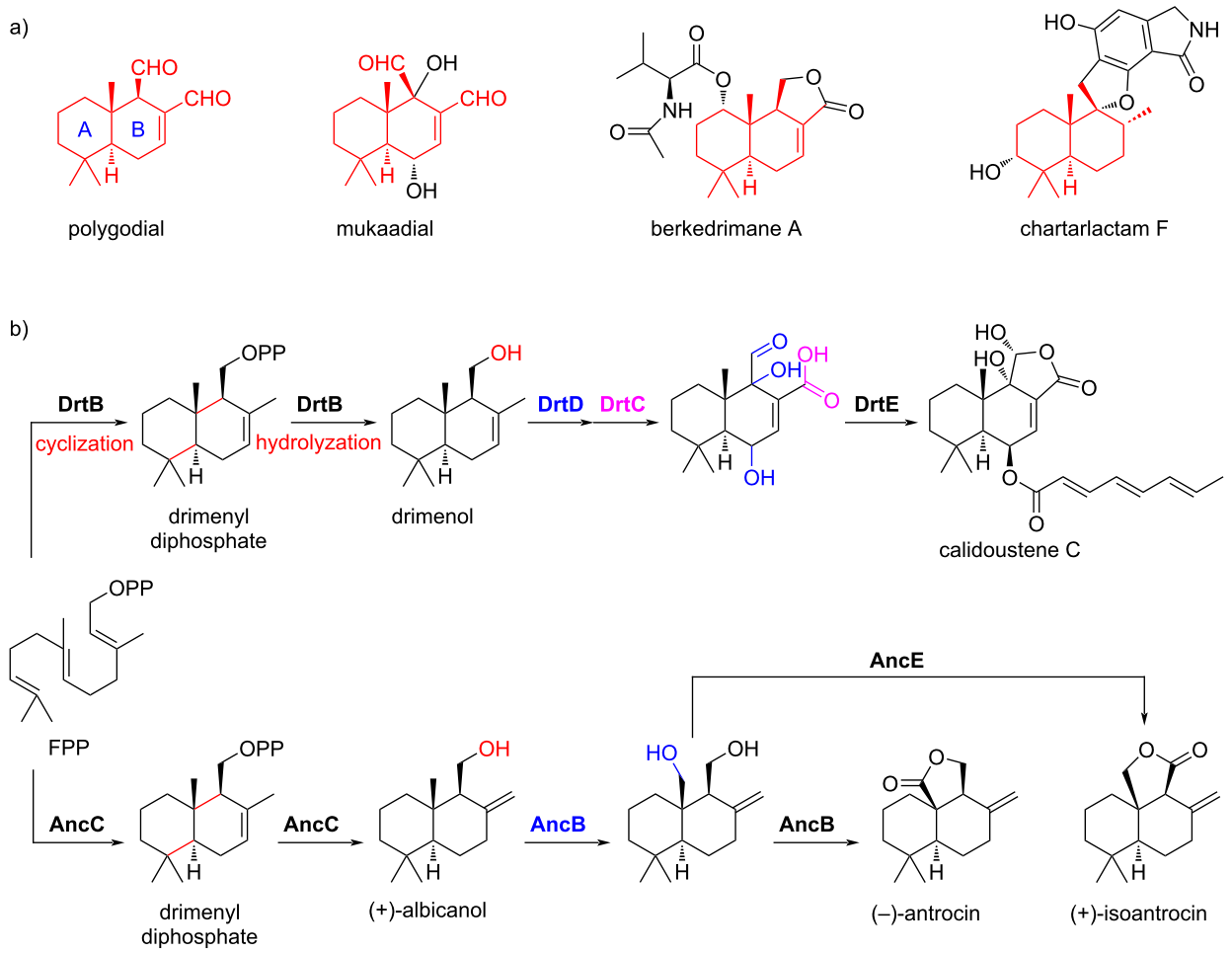



**Figure 1:** (a) Representative bioactive drimane-type sesquiterpenoids (DMTs). (b) Reported biosynthetic pathways of DMTs from fungi.

antifeedant for various herbivorous insects [7]. Additionally, berkedrimane A, mukaadial, and chartarlactam F demonstrate caspase-3 inhibition, antifungal, and antihyperlipidemic activities, respectively [6,8-12]. Collectively, these examples emphasize the potential of DMTs as substances with broad and significant biological activities.

The biosynthetic pathways for DMTs, especially for calidoustene C, (+)-isoantrocin, and (−)-antrocin, have been extensively elucidated [6,13-20] (Figure 1b). In the biosynthesis of calidoustene C, DrtB from *Aspergillus calidoustus* functions as a dual-functional enzyme, comprising two domains: a HAD-like hydrolase domain fused with a terpene cyclase domain. Initially, FPP is cyclized into the drimenyl diphosphate in a class II terpene cyclase manner, which is then processed by the hydrolase domain to form drimenol [21]. Subsequently, this molecule undergoes further modifications by two cytochrome P450s monooxygenases (P450s, DrtC and DrtD) and one acyltransferase (DrtE) to produce calidoustene C [13]. This cyclization is similar with the course of AncC from *Antrodia cinnamomea*, which performs a similar role to DrtB, engaging two P450s (AncE and AncB) to synthesize (+)-isoantrocin and (−)-antrocin [18]. Notably, the P450s identified in fungi and plants predominantly modify the B-ring of DMTs [15,16,18].

DMTs are commonly found in plants and fungi [6,9,13,15]. While enzymes associated with DMT biosynthesis have been identified in bacteria, the corresponding natural DMTs have not been discovered [17]. In this study, we isolated and characterized three drimenol congeners (**2**–**4**) from *Streptomyces clavuligerus* (Figure 2a). In the genome of *S. clavuligerus*, we identified a *cav* biosynthetic gene cluster (BGC) that typically includes three P450s, one class II drimenyl diphosphate synthases (DMS), and one characteristic Nudix hydrolase responsible for the DMTs biosynthesis. Among the three P450s, CavA was the only oxygenase acting on drimenol and versatile in oxidizing C-2 and C-3 positions of the A ring of drimenol by heterologous expression and gene knockout. Furthermore, we

tested the substrate promiscuity of CavA with drimenol analogs. CavA exhibited the ability to accept albicanol (**5**) and drim-8-ene-11-ol (**6**) as substrates, leading to the formation of the hydroxylated derivatives **7**–**9**. This study marks the first discovery of natural DMTs from bacteria and unveils the role of CavA in a novel late-stage modification pathway, expanding DMT biosynthesis beyond fungi and plants.

## Results and Discussion

### Discovery of three DMTs in *S. clavuligerus*

*S. clavuligerus* is notably significant in industrial applications, renowned for its production of diverse natural products with chemical structures and bioactivities, such as cephamycin C, clavulanic acid, and isopenicillin N [22-24]. Genomic sequencing of *S. clavuligerus* has revealed 48 potential secondary metabolite BGCs, and our analysis indicated that 16 of them are potential terpene-producing BGCs, indicative of its extensive capacity for diverse metabolite production [25]. In our large-scale fermentation of *S. clavuligerus*, aimed at uncovering novel bacterial terpenoids, we observed that specific culture conditions are crucial for metabolite production. Different culture media can lead to the biosynthesis of varied compounds. Specifically, in this study, *S. clavuligerus* was cultivated in two terpene favored media, XTM and YMS [26,27], resulting in three drimenol (**1**) congeners from the extracts: 3β-hydroxydrimenol (**2**), 2α-hydroxydrimenol (**3**), and 3-ketodrimenol (**4**) (Figure 2a). HPLC analysis of metabolites from different culture media showed that YMS medium was more conducive to produce compound **3** (Figure 2b and Table S1 in Supporting Information File 1).

The chemical structures of these isolated compounds were mainly elucidated through comparative $^{1}H$ and $^{13}C$ NMR spectra and by comparing analytical data with existing literature reports [28,29] (see Supporting Information File 1, Figures S1–S7). The $^{1}H$ NMR spectrum of compound **2** exhibited characteristic signals for one olefinic proton at $\delta_H$ (5.48) and four methyl groups at $\delta_H$ (0.83, 0.87, 0.98 and 1.77). Its $^{13}C$ and

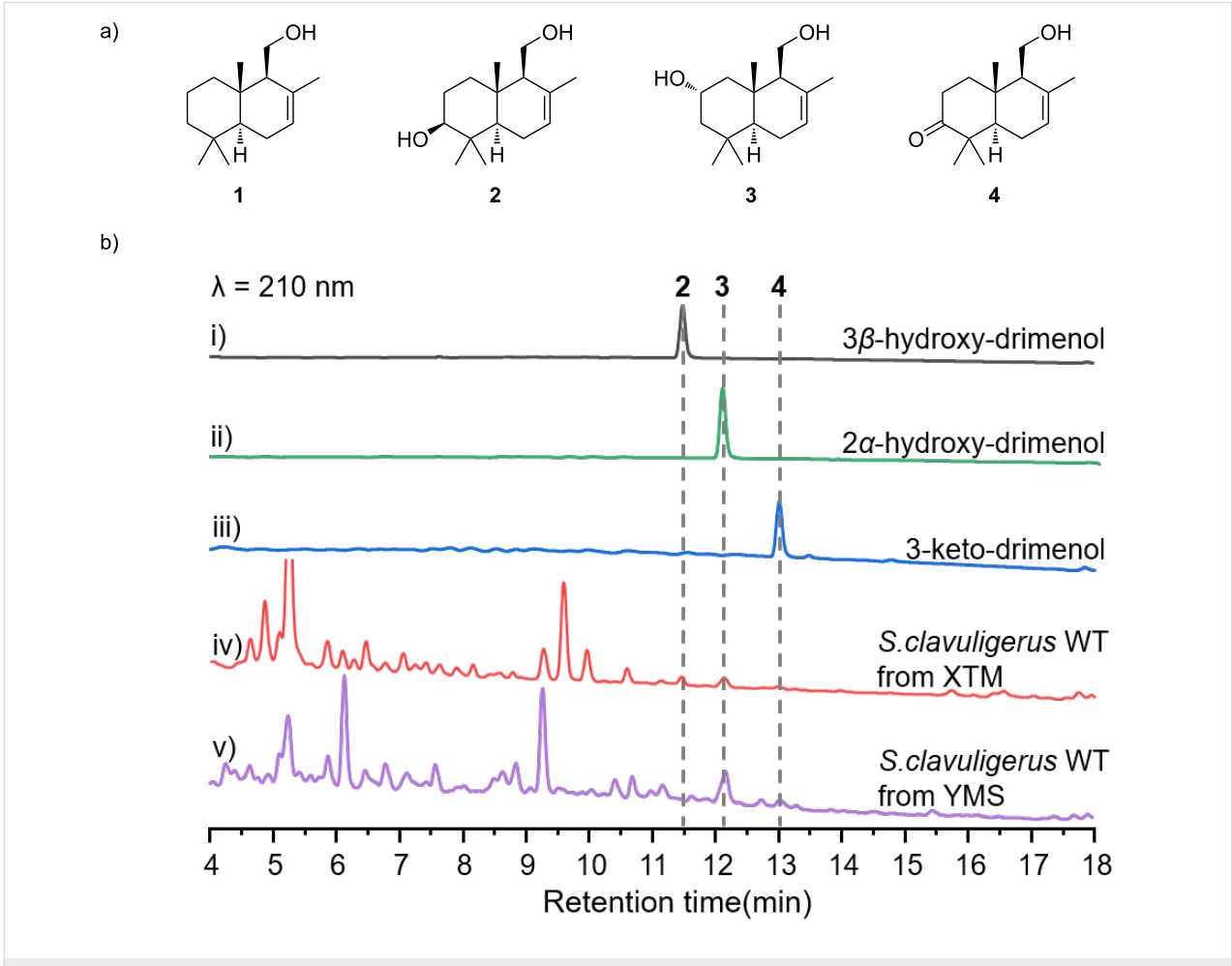



**Figure 2:** Chemical structures and HPLC analysis. (a) Chemical structures of isolated sesquiterpenes **2**–**4** from *S. clavuligerus*. (b) HPLC chromatograms comparing standards with metabolites extracted from *S. clavuligerus* cultured in YMS and XTM media.

DEPT NMR spectra showed 15 carbon resonances, including four methyl groups, four methylenes, four methines, and three quaternary carbon atoms. This information suggests that it may be a drimenol congener and the structure was further supported by comparison of analytical data with a literature report [29]. The $^1$H NMR spectrum of compound **3** is similar to that of **2**, showing four methyl groups and one olefinic methyl singlet, indicating structural resemblances with minor variations. It was ultimately determined through literature comparison that the only difference between compounds **3** and **2** is the position of the hydroxy group [28]. Analyses of the NMR data of compound **4** concluded that it is an analogue of **2**. In comparison with **2**, compound **4** has a ketone carbonyl signal at $\delta_C$ 219.1 and we finally confirmed its structure by comparing it with literature data [29]. This marks the first discovery of natural DMTs from bacteria, surpassing previous findings of bacterial DMSs which focused only on the enzyme itself without reporting natural DMT or exploring associated BGCs [17]. Additionally, while drimentines, bacterial meroterpenoids, share the drimanyl scaffold, their biosynthetic pathways, particularly the terpene cyclases synthesizing the drimanyl structures, show substantial divergence from DMT pathways [21,30-32]. This fundamental difference in biosynthetic mechanisms serves to categorize DMSs and drimentines into distinct subfamilies within the sesquiterpenoids. Our discovery significantly broadens the scope for future exploration of bacterial DMTs.

## Discovery and identification of a DMS in the *cav* BGC in *S. clavuligerus*

To explore the BGC responsible for DMTs production in *S. clavuligerus*, we performed bioinformatics research utilizing the EFI-genome neighborhood tool (EFI-GNT), sequence alignment, and manual BLAST analysis [33]. Leveraging our previous discovery of SsDMS from *Streptomyces showdoensis*, we identified a homologous terpene cyclase (sclav_p0068) in *S. clavuligerus*. This enzyme shares a 50% sequence similarity with SsDMS, including a conserved "DXDD" motif, indicating its potential as a class II sesquiterpene synthase (Figure S8, Supporting Information File 1) [20]. Further analysis of adjacent genes revealed a Nudix hydrolase located upstream of the putative DMS, similar to the reported situation with SsDMS [19]. Inspection of the BGC identified additional three P450s. Thus, the entire cluster mainly encompasses a DMS (CavC), a Nudix hydrolase (CavB), a class I terpene cyclase (CavF), and three P450s (CavA, CavE, and CavG), and it was designated as the *cav* cluster (Figure 3a and Table S2 in Supporting Information File 1).

To validate the function of the presumed DMS, we cloned its gene into the pETDuet-1 vector to form plasmid pLD10050, which was transformed into our truncated artificial FPP-overproduction system in *E. coli* to yield strain DL10092 [34]. Analysis of the fermentation extracts revealed that DL10092 displayed a new peak in its HPLC profile compared to the control of wild-type *E. coli* BL21 (DE3) (see Figure S9 in Supporting Information File 1). The fermented product appeared as an alcohol, likely due to the activity of the endogenous hydrolase in *E. coli*. The retention time of the new peak was identical with the drimenol standard, suggesting successful synthesis of drimenol. This confirms CavC as a DMS, and the *cav* BGC is likely responsible for the biosynthesis of compounds **2**–**4**.

## CavA, a P450, responsible for the formation of C-2 and C-3 hydroxy groups

It was hypothesized that compound **1** serves as a precursor for derivatives **2**–**4**. The biosynthetic pathway is likely to involve enzymatic modifications at the C-2 and C-3 positions, leading to the formation of diols **2** and **3**, with additional oxidation steps producing ketone **4** (Figure 3b). This supports the presence of three P450s in the *cav* cluster. We initially cloned and expressed the complete *cav* BGC in three model *Streptomyces* species (Tables S3–S5 in Supporting Information File 1). Notably, *Streptomyces lividans* TK64 DL10081, containing the entire *cav* BGC, produced only compound **2** in three media (PTMM, XTM, and ISM3; Figure S10 in Supporting Information File 1). Interestingly, *S. avermitilis* SUKA22 DL10085 can produce **2** and **3** in XTM medium (Figure 3c). The subsequent 7.5 L fermentation of strain DL10085 in XTM medium resulted in a significant production of compounds **2** (14 mg) and **3** (5 mg), highlighting the crucial role of P450s in the biosynthesis. However, compound **4** was not detected in the heterologous expression products, potentially due to the involvement of a dehydrogenase outside the *cav* BGC.

An in-depth study of the specific roles of the P450 genes was conducted by creating single-gene deletion mutants from strain *S. avermitilis* SUKA22 DL10085. The DL10086 strain, with the deletion of the *cavA* gene, failed to produce **2** or **3**, suggesting the vital role of *cavA* in *cav* BGC. To further substantiate the hypothesis, *cavB* and *cavC* were co-expressed with each of the three P450s in *S. avermitilis* SUKA22, creating three recombinant strains (DL10089, DL10090, and DL10091; Tables S3–S5 in Supporting Information File 1). Notably, the DL10089 strain, co-expressing drimenol synthase and CavA, successfully produced **2** and **3**. This result not only corroborates our initial findings but also confirms the CavA role in hydroxylating the C-2 and C-3 positions of drimenol (Figure 3d). However, the remaining two P450 enzymes, CavE and CavG, appear to be non-functional either in the native *S. clavuligerus* or heterologous expression systems. Given the detection of two terpene cyclases (CavC and CavF) and the exclusive DMTs generated,

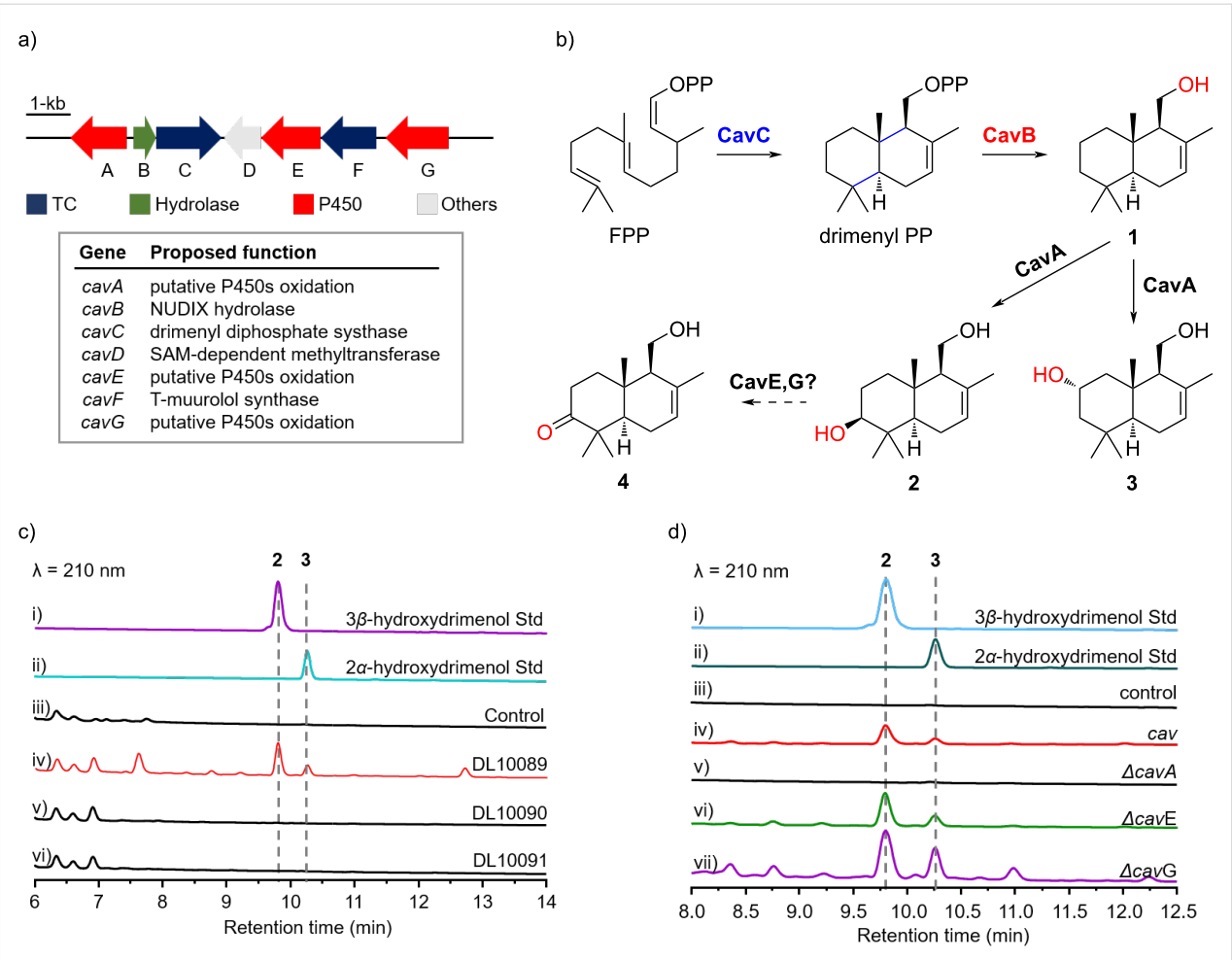



**Figure 3:** Biosynthesis of drimenol congeners. (a) The *cav* BGC. (b) Proposed biosynthetic pathway for drimenol congeners **2**–**4**. (c) HPLC analysis of metabolites from genetically engineered *Streptomyces* strains, with *S. avermitilis* SUKA22 harboring the empty pSET152 vector as the control. (d) HPLC analysis of metabolites co-expressed with drimenyl diphosphate synthases (CavC), Nudix hydrolase (CavB), and one of the P450s, either CavA (DL10089), CavE (DL10090), or CavG (DL10091).

we also speculated that the characterized *cav* BGC may include two separate BGCs situated closely together. Future efforts will be focused on uncovering the roles of the other three enzymes (CavF, CavE, and CavG).

## Substrate scope of CavA with drimenol analogs

P450s are renowned for their remarkable versatility as biocatalysts, capable of catalyzing a wide range of reactions and processing a variety of substrates [35,36]. They effectively recognize different structural types, including aromatic compounds, polyketides, terpenes, peptides, and carbohydrates [37-40]. To investigate the substrate spectrum of CavA more closely, eight compounds were chosen for this study (Figure 4a). Due to the insolubility of CavA in *E. coli*, the study employed in vivo substrate screening. The *S. avermitilis* SUKA22 DL10089 was cultivated in XTM medium with the supplement of various substrates for five days. HPLC analysis indicated that CavA specifically oxidizes albicanol (**5**) and drim-8-ene-11-ol (**6**) to synthesize **7**–**9**, while other substrates showed no reactivity with CavA (Figure 4c). Following this, we increased the substrate concentration to enrich and isolate these products, elucidating their structures through NMR analysis (Figure 4b). In these experiments, CavA was found to hydroxylate drim-8-ene-11-ol (**6**) at the C-3 and C-7 positions to produce **8** and **9**, respectively, while albicanol (**5**) was modified only at the C-3 position. Compound **7** is a known natural product, and its chemical structure was determined by comparing the $^{1}H$ and $^{13}C$ NMR data with the literature (Figures S11 and S12 in Supporting Information File 1) [41]. Although the chemical structure of compound **8** has been documented, its $^{1}H$ and $^{13}C$ NMR data were not fully reported [42]. Therefore, we conducted comprehensive 1D and 2D NMR experiments on it (Figures S13–S18, Supporting Information File 1) and summarized the resulting data in Table

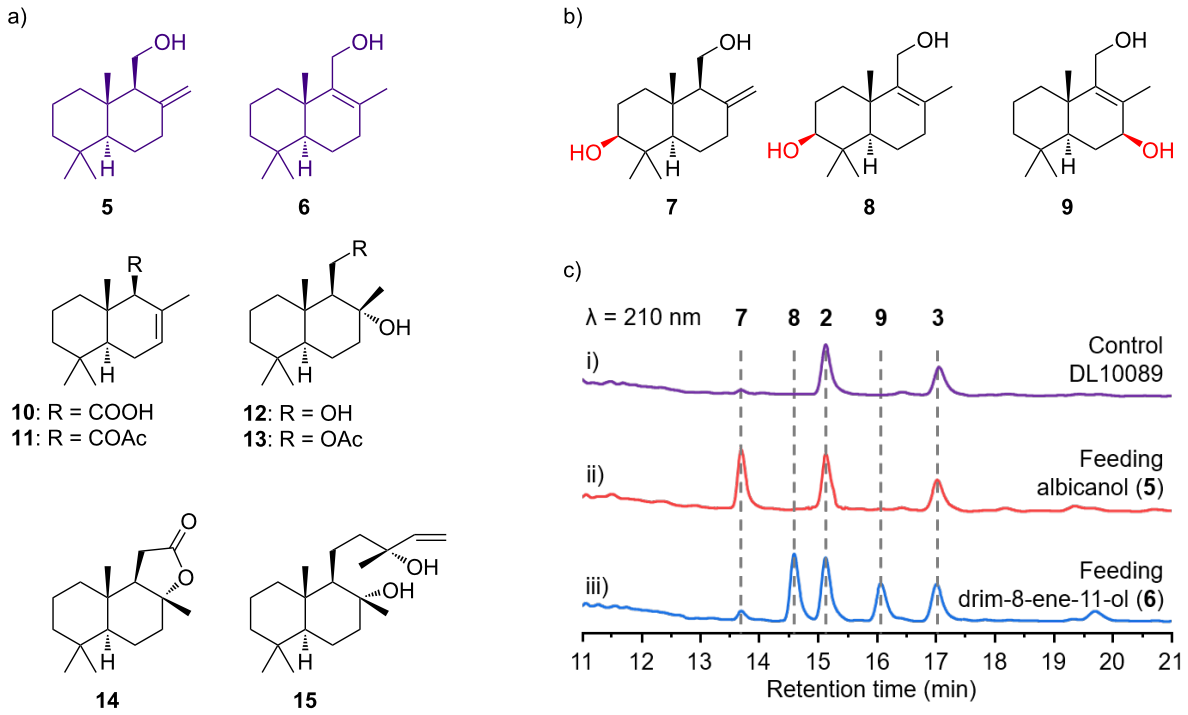



**Figure 4:** Substrate scope of CavA. (a) The chemical structures of eight natural products that were used for feeding experiments. (b) The three hydroxylated products generated from feeding experiments. (c) HPLC analysis of metabolites from *S. avermitilis* SUKA22 strain DL10089 cultured in XTM medium with albicanol (**5**) and drim-8-ene-11-ol (**6**) supplementation.

S6 (Supporting Information File 1). For compound **9**, despite sharing an identical planar structure with diaporol E, the configurations of the C7 hydroxy groups are opposite, as evidenced by significant chemical shift differences in $^{13}C$ NMR ($\Delta\delta_C$ = 3.3) (Figures S19–S25, Supporting Information File 1) [43]. Moreover, the key correlation observed between H5 and H7 in the ROESY spectrum provides additional support for the β-configuration of this hydroxy group (Figure S26 and Table S7 in Supporting Information File 1). These findings from both the heterologous expression and substrate spectrum analysis indicate the particular affinity of CavA for the C-3 position, which might be used for chemoenzymatic synthesis of C-3 hydroxylated drimanyl meroterpenoids using drimenol and its analogues as substrates. However, it is important to note that CavA exhibits limited substrate promiscuity, predominantly targeting the drimenol skeleton with minor variations. This selectivity may be attributed to the structural configuration of the enzyme, which appears to be finely tuned to recognize and interact with specific features of the drimenol framework.

## Conclusion

In this study, we isolated and characterized three bacterial DMTs (**2**–**4**) derived from *S. clavuligerus*. This finding challenges the traditional focus on extracting DMTs exclusively from fungi and plants, establishing a new avenue for the discovery of DMTs. Subsequently, we identified the *cav* BGC responsible for the production of compounds **2** and **3** and elucidated its biosynthetic pathway by heterologous expression. Specifically, our study unveils a P450, CavA, capable of oxidizing the C-2 and C-3 positions of drimenol, representing the first native P450 that activates the A-ring of DMTs. The in vivo experiments further expanded the substrate scope of CavA to include albicanol (**5**) and drim-8-ene-11-ol (**6**), showcasing the enzyme's biocatalytic potential. This study establishes a foundation for biocatalysts targeting the A-ring of drimenol, which might be beneficial for the efficient chemoenzymatic synthesis of DMTs or drimanyl indoles [44].

## Supporting Information

### Supporting Information File 1

Experimental part and supplementary figures and tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-73-S1.pdf]

## Funding

This work is supported in part by the National Natural Science Foundation of China Grant 82073746 (L.-B.D.), the Thousand

Youth Talents Program of China (L.-B.D.), the Double First-Class University Project (CPUQNJC22 04, L.-B.D.), and the Jiangsu Specially Appointed Professor Program (L.-B.D.).

## ORCID® iDs

Liao-Bin Dong - https://orcid.org/0000-0002-2943-1299

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Huang, Z.-Y.; Taizoumbe, K. A.; Liang, C.; Goldfuss, B.; Xu, J.-H.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2023,** e202315659. doi:10.1002/anie.202315659
2. Dairi, T.; Kuzuyama, T.; Nishiyama, M.; Fujii, I. *Nat. Prod. Rep.* **2011,** *28,* 1054–1086. doi:10.1039/c0np00047g
3. Frank, A.; Groll, M. *Chem. Rev.* **2017,** *117,* 5675–5703. doi:10.1021/acs.chemrev.6b00537
4. Maffei, M. E.; Gertsch, J.; Appendino, G. *Nat. Prod. Rep.* **2011,** *28,* 1359–1380. doi:10.1039/c1np00021g
5. Salman, K.; Zhu, H.; Sun, Z.; Li, Y.; Wang, L.; Wang, R.; Guo, Z.; Jiao, R. *Chin. J. Nat. Med.* **2023,** *21,* 58–64. doi:10.1016/s1875-5364(23)60385-1
6. Du, W.; Yang, Q.; Xu, H.-M.; Dong, L. *Chin. J. Nat. Med.* **2022,** *20,* 737–748. doi:10.1016/s1875-5364(22)60190-0
7. Szallasi, A.; Bíró, T.; Modarres, S.; Garlaschelli, L.; Petersen, M.; Klusch, A.; Vidari, G.; Jonassohn, M.; De Rosa, S.; Sterner, O.; Blumberg, P. M.; Krause, J. E. *Eur. J. Pharmacol.* **1998,** *356,* 81–89. doi:10.1016/s0014-2999(98)00514-7
8. Burgos, V.; Paz, C.; Saavedra, K.; Saavedra, N.; Foglio, M. A.; Salazar, L. A. *Food Chem. Toxicol.* **2020,** *146,* 111775. doi:10.1016/j.fct.2020.111775
9. Huang, Y.; Valiante, V. *ChemBioChem* **2022,** *23,* e202200173. doi:10.1002/cbic.202200173
10. Stierle, D. B.; Stierle, A. A.; Girtsman, T.; McIntyre, K.; Nichols, J. *J. Nat. Prod.* **2012,** *75,* 262–266. doi:10.1021/np200528n
11. Li, Y.; Wu, C.; Liu, D.; Proksch, P.; Guo, P.; Lin, W. *J. Nat. Prod.* **2014,** *77,* 138–147. doi:10.1021/np400824u
12. Chen, T.-H.; Lin, H.-C. *ChemBioChem* **2023,** *24,* e202300518. doi:10.1002/cbic.202300518
13. Kwon, M.; Cochrane, S. A.; Vederas, J. C.; Ro, D.-K. *FEBS Lett.* **2014,** *588,* 4597–4603. doi:10.1016/j.febslet.2014.10.031
14. Shinohara, Y.; Takahashi, S.; Osada, H.; Koyama, Y. *Sci. Rep.* **2016,** *6,* 32865. doi:10.1038/srep32865
15. Henquet, M. G. L.; Prota, N.; van der Hooft, J. J. J.; Varbanova-Herde, M.; Hulzink, R. J. M.; de Vos, M.; Prins, M.; de Both, M. T. J.; Franssen, M. C. R.; Bouwmeester, H.; Jongsma, M. *Plant J.* **2017,** *90,* 1052–1063. doi:10.1111/tpj.13527
16. Huang, Y.; Hoefgen, S.; Valiante, V. *Angew. Chem., Int. Ed.* **2021,** *60,* 23763–23770. doi:10.1002/anie.202108970
17. Vo, N. N. Q.; Nomura, Y.; Kinugasa, K.; Takagi, H.; Takahashi, S. *ACS Chem. Biol.* **2022,** *17,* 1226–1238. doi:10.1021/acschembio.2c00163
18. Chen, T.-H.; Chen, C.-T.; Lee, C.-F.; Huang, R.-J.; Chen, K.-L.; Lu, Y.-C.; Liang, S.-Y.; Pham, M.-T.; Rao, Y. K.; Wu, S.-H.; Chein, R.-J.; Lin, H.-C. *Angew. Chem., Int. Ed.* **2023,** *62,* e202215566. doi:10.1002/anie.202215566
19. Kreuzenbeck, N. B.; Dhiman, S.; Roman, D.; Burkhardt, I.; Conlon, B. H.; Fricke, J.; Guo, H.; Blume, J.; Görls, H.; Poulsen, M.; Dickschat, J. S.; Köllner, T. G.; Arndt, H.-D.; Beemelmanns, C. *Commun. Chem.* **2023,** *6,* 79. doi:10.1038/s42004-023-00871-z
20. Pan, X.; Du, W.; Zhang, X.; Lin, X.; Li, F.-R.; Yang, Q.; Wang, H.; Rudolf, J. D.; Zhang, B.; Dong, L.-B. *J. Am. Chem. Soc.* **2022,** *144,* 22067–22074. doi:10.1021/jacs.2c09412
21. Pan, X.; Rudolf, J. D.; Dong, L.-B. *Nat. Prod. Rep.* **2024,** *41,* 402–433. doi:10.1039/d3np00033h
22. Mendelovitz, S.; Aharonowitz, Y. *Antimicrob. Agents Chemother.* **1982,** *21,* 74–84. doi:10.1128/aac.21.1.74
23. Reading, C.; Cole, M. *Antimicrob. Agents Chemother.* **1977,** *11,* 852–857. doi:10.1128/aac.11.5.852
24. Leskiw, B. K.; Aharonowitz, Y.; Mevarech, M.; Wolfe, S.; Vining, L. C.; Westlake, D. W. S.; Jensen, S. E. *Gene* **1988,** *62,* 187–196. doi:10.1016/0378-1119(88)90557-4
25. Medema, M. H.; Trefzer, A.; Kovalchuk, A.; van den Berg, M.; Müller, U.; Heijne, W.; Wu, L.; Alam, M. T.; Ronning, C. M.; Nierman, W. C.; Bovenberg, R. A. L.; Breitling, R.; Takano, E. *Genome Biol. Evol.* **2010,** *2,* 212–224. doi:10.1093/gbe/evq013
26. Kim, S.-H.; Lu, W.; Ahmadi, M. K.; Montiel, D.; Ternei, M. A.; Brady, S. F. *ACS Synth. Biol.* **2019,** *8,* 109–118. doi:10.1021/acssynbio.8b00361
27. Dong, L.-B.; Rudolf, J. D.; Deng, M.-R.; Yan, X.; Shen, B. *ChemBioChem* **2018,** *19,* 1727–1733. doi:10.1002/cbic.201800285
28. Rasser, F.; Anke, T.; Sterner, O. *Tetrahedron* **2002,** *58,* 7785–7789. doi:10.1016/s0040-4020(02)00943-2
29. Xu, D.; Sheng, Y.; Zhou, Z.-Y.; Liu, R.; Leng, Y.; Liu, J.-K. *Chem. Pharm. Bull.* **2009,** *57,* 433–435. doi:10.1248/cpb.57.433
30. Che, Q.; Zhu, T.; Qi, X.; Mándi, A.; Kurtán, T.; Mo, X.; Li, J.; Gu, Q.; Li, D. *Org. Lett.* **2012,** *14,* 3438–3441. doi:10.1021/ol301396h
31. Yao, T.; Liu, J.; Liu, Z.; Li, T.; Li, H.; Che, Q.; Zhu, T.; Li, D.; Gu, Q.; Li, W. *Nat. Commun.* **2018,** *9,* 4091. doi:10.1038/s41467-018-06411-x
32. Rudolf, J. D.; Alsup, T. A.; Xu, B.; Li, Z. *Nat. Prod. Rep.* **2021,** *38,* 905–980. doi:10.1039/d0np00066c
33. Oberg, N.; Zallot, R.; Gerlt, J. A. *J. Mol. Biol.* **2023,** *435,* 168018. doi:10.1016/j.jmb.2023.168018
34. Li, F.-R.; Lin, X.; Yang, Q.; Tan, N.-H.; Dong, L.-B. *Beilstein J. Org. Chem.* **2022,** *18,* 881–888. doi:10.3762/bjoc.18.89
35. Guo, J.; Li, F.; Cheng, F.; Ma, L.; Liu, X.; Durairaj, P.; Zhang, G.; Tang, D.; Long, X.; Zhang, W.; Du, L.; Zhang, X.; Li, S. *J. Org. Chem.* **2021,** *86,* 14563–14571. doi:10.1021/acs.joc.1c01407
36. Pan, Y.; Li, G.; Liu, R.; Guo, J.; Liu, Y.; Liu, M.; Zhang, X.; Chi, L.; Xu, K.; Wu, R.; Zhang, Y.; Li, Y.; Gao, X.; Li, S. *Nat. Commun.* **2023,** *14,* 1669. doi:10.1038/s41467-023-37288-0
37. Peschke, M.; Haslinger, K.; Brieke, C.; Reinstein, J.; Cryle, M. J. *J. Am. Chem. Soc.* **2016,** *138,* 6746–6753. doi:10.1021/jacs.6b00307
38. Wang, Z.; Yang, Q.; He, J.; Li, H.; Pan, X.; Li, Z.; Xu, H.-M.; Rudolf, J. D.; Tantillo, D. J.; Dong, L.-B. *Angew. Chem., Int. Ed.* **2023,** *62,* e202312490. doi:10.1002/anie.202312490
39. Zheng, Q.; Wang, S.; Liao, R.; Liu, W. *ACS Chem. Biol.* **2016,** *11,* 2673–2678. doi:10.1021/acschembio.6b00419

40. Devine, A. J.; Parnell, A. E.; Back, C. R.; Lees, N. R.; Johns, S. T.; Zulkepli, A. Z.; Barringer, R.; Zorn, K.; Stach, J. E. M.; Crump, M. P.; Hayes, M. A.; van der Kamp, M. W.; Race, P. R.; Willis, C. L. *Angew. Chem., Int. Ed.* **2023,** *62,* e202213053. doi:10.1002/anie.202213053
41. Otaka, J.; Araya, H. *Phytochem. Lett.* **2013,** *6,* 598–601. doi:10.1016/j.phytol.2013.07.010
42. Mori, K.; Komatsu, N. *Tetrahedron* **1987,** *43,* 3409–3412. doi:10.1016/s0040-4020(01)81631-8
43. Zang, L. Y.; Wei, W.; Guo, Y.; Wang, T.; Jiao, R. H.; Ng, S. W.; Tan, R. X.; Ge, H. M. *J. Nat. Prod.* **2012,** *75,* 1744–1749. doi:10.1021/np3004112
44. Li, J.; Li, F.; King-Smith, E.; Renata, H. *Nat. Chem.* **2020,** *12,* 173–179. doi:10.1038/s41557-019-0407-6

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.73

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Activity assays of NnlA homologs suggest the natural product *N*-nitroglycine is degraded by diverse bacteria

Kara A. Strickland‡[1], Brenda Martinez Rodriguez‡[1], Ashley A. Holland[1], Shelby Wagner[1], Michelle Luna-Alva[1], David E. Graham[2] and Jonathan D. Caranto*[1]

Full Research Paper

Open Access

Address:
[1]Department of Chemistry, University of Central Florida, Orlando, FL 32816, USA and [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Email:
Jonathan D. Caranto* - jonathan.caranto@ucf.edu

* Corresponding author ‡ Equal contributors

Keywords:
enzymology; natural products; nitramine; N–N bond

*Beilstein J. Org. Chem.* **2024,** *20,* 830–840.
https://doi.org/10.3762/bjoc.20.75

Received: 16 December 2023
Accepted: 04 April 2024
Published: 17 April 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: T. Awakawa



© 2024 Strickland et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Linear nitramines (R–N(R′)$NO_2$; R′ = H or alkyl) are toxic compounds, some with environmental relevance, while others are rare natural product nitramines. One of these natural product nitramines is *N*-nitroglycine (NNG), which is produced by some *Streptomyces* strains and exhibits antibiotic activity towards Gram-negative bacteria. An NNG degrading heme enzyme, called NnlA, has recently been discovered in the genome of *Variovorax sp.* strain JS1663 (*Vs* NnlA). Evidence is presented that NnlA and therefore, NNG degradation activity is widespread. To achieve this objective, we characterized and tested the NNG degradation activity of five *Vs* NnlA homologs originating from bacteria spanning several classes and isolated from geographically distinct locations. *E. coli* transformants containing all five homologs converted NNG to nitrite. Four of these five homologs were isolated and characterized. Each isolated homolog exhibited similar oligomerization and heme occupancy as *Vs* NnlA. Reduction of this heme was shown to be required for NnlA activity in each homolog, and each homolog degraded NNG to glyoxylate, $NO_2^-$ and $NH_4^+$ in accordance with observations of *Vs* NnlA. It was also shown that NnlA cannot degrade the NNG analog 2-nitroaminoethanol. The combined data strongly suggest that NnlA enzymes specifically degrade NNG and are found in diverse bacteria and environments. These results imply that NNG is also produced in diverse environments and NnlA may act as a detoxification enzyme to protect bacteria from exposure to NNG.

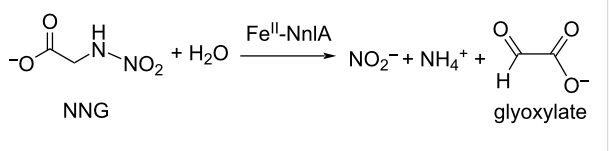
## Introduction

Degradation of nitramines (R–N(R′)$NO_2$; R′ = H or alkyl) has been well studied in the context of the environmental degradation of explosive cyclic nitramines [1,2]. The cyclic nitramines hexahydro-1,3,5-trinitro-1,3,5-triazine (commonly called RDX), octogen (HMX), and hexanitrohexaazaisowurtzitane (CL-20) are compounds found in military grade explosives and propel-

lants. Contamination of these cyclic nitramines in soil and groundwater is concerning due to their toxicity and potential carcinogenicity [3-8]. Biotic and abiotic degradation of cyclic nitramines often produce linear nitramine byproducts. For example, degradation of RDX and HMX by microbes or alkaline hydrolysis forms the linear nitramine 4-nitro-2,4-diazabutanal (NDAB) [9-12]. Linear nitramines are also produced during the process of amine-based carbon dioxide capture technologies [13,14]. These linear nitramines from these reactions pose their own health and environmental consequences [13]. Therefore, strategies to remediate linear nitramines are needed.

Compared to cyclic nitramines [1,2,15,16], there is far less known regarding the environmental biodegradation pathways of linear nitramine contaminants. Biodegradations of NDAB by the fungus *Phanerochaete chrysosporium* and the bacterium *Methylobacterium sp.* strain JS178 have been reported [17,18]. Initiation of the *P. chrysosporium* degradation was attributed to a manganese peroxidase, however, the mechanism of degradation is unclear. Linear nitramines, produced by carbon capture, were shown to be biodegraded in soil and water [19]. Nitramines with hydroxy groups were best degraded in this study, including diethylnitramine, 2-methyl-2-(nitroamino)-1-propanol, and 2-nitroaminoethanol (2-NAE). While little is known regarding the degradation of these anthropogenic linear nitramines, we can glean insight into their reactivities from recent studies regarding the enzymatic degradation of *N*-nitroglycine (NNG), a naturally occurring linear nitramine.

An enzyme, *N*-nitroglycine lyase A (NnlA), from the bacterium *Variovorax sp.* strain JS1663 (*Vs* NnlA) was recently shown to degrade NNG. This strain was enriched from sludge from the Holston Army Ammunition Plant using selective growth media containing NNG as the only carbon and nitrogen source [20]. The discovery of the *nnlA* gene resulted from screening a JS1663 genomic library by monitoring for *E. coli* transformants that produced nitrite ($NO_2^-$) in the presence of NNG. Analysis of the mass balance via in vitro experiments showed that NnlA degraded NNG into $NO_2^-$, ammonium ($NH_4^+$), and glyoxylate (Scheme 1) [20,21]. *Vs* NnlA contains a Per-Arnt-Sim (PAS) domain – protein domains that often bind heme and function as gas or redox sensors [20]. Indeed, *Vs* NnlA was shown to contain a heme cofactor [21]. Mutagenesis of a predicted histidine ligand to this heme resulted in loss of the heme and the variant could not degrade NNG. Additionally, this heme must be reduced to the ferrous ($Fe^{II}$) state to initiate NNG degradation. Therefore, the heme is critical for NnlA's NNG degradation activity.




**Scheme 1:** NNG degradation by reduced NnlA.

While the activity of NnlA is established less is known about its physiological function and, for that matter, the physiological function of its substrate NNG. This compound is one of the few known nitramine natural products and the only one produced by bacteria instead of fungi [22]. Its only known natural sources are strains of *Streptomyces* bacteria [23,24]. The abundance and distribution of these NNG producers and of NNG is unknown. Additionally, NNG's physiological function is unknown, but it is toxic to plants, mice, and Gram-negative bacteria [25,26]. While there is no direct evidence of the mechanism of this toxicity, NNG has been shown to competitively inhibit succinate dehydrogenase, a component enzyme of the Krebs cycle [25]. Therefore, NNG may be a toxin released to kill or outcompete nearby bacteria or other organisms for limited resources. In such a context, the physiological function of NnlA could be to protect bacteria from toxic NNG exposure. Alternatively, NnlA could be a promiscuous nitramine degrader that allows bacteria to use alternative nitrogen sources. In fact, the RDX-degrading enzyme, XplA, is remarkably conserved amongst RDX-degrading microbes (>99% identity across several species) [16,27,28]. Based on these observations, it has been proposed that XplA evolved within the past 100 years in response to the rise in RDX contamination. Given that *Variovorax sp.* strain JS1663 was isolated from a nitramine-contaminated sludge, it should be considered if NnlA is promiscuous and can degrade both natural nitramines, such as NNG, and anthropogenic nitramines, such as 2-NAE.

We have previously identified several *Vs* NnlA homologs in sequence databases [20], however, the NNG degradation activities of these homologs have not been tested. Doing so will differentiate between these two hypotheses by testing if NnlA homologs with NNG degradation activity are highly conserved or if they are found in bacteria found in widespread classes and environments. Herein, we report the characterization of five NnlA homologs. It is shown that all five homologs exhibited NNG degradation activity. Isolation and characterization of four of these homologs showed that all contain heme, the reduction of which is required for NNG degradation activity. In addition, we show that NnlA cannot degrade 2-NAE. Combined with previous substrate scope studies, this result strongly suggests that NnlA is specific for NNG. The implications of our results in understanding the environmental abundance and physiological function of NNG are discussed below.

# Results

## Screening of *Vs* NnlA homologs for NNG degradation activity

To select *Vs* NnlA homologs to test for NNG degradation activity, we performed a BLAST search of the *Vs* NnlA amino acid sequence in the NCBI database. This search resulted in retrieval of 99 homologous amino acid sequences with an E-value of less than $1 \times 10^{-10}$. A subset of 55 of these sequences was selected with a query cover of greater than 88%. From these sequences, we selected five homologs of the *nnlA* gene [*Pseudovibrio denitrificans* JCM 1230 (*Pd*), *Pseudovibrio japonicus* strain KCTC 12861 (*Pj*), *Pseudonocardia spinosispora* DSM 44797 (*Ps*), *Mycobacterium sp.* 1465703.0 (*Ms*), *Microbispora rosea subsp. nonnitritogenes* strain NRRL B-2631 (*Mr*)], which were synthesized and cloned into *E. coli* recombinant expression vectors. These homologs ranged in amino acid sequence identity from 46 to 76% compared to *Vs* NnlA. Additionally, these homologs along with *Vs* NnlA are found in bacteria that span a wide range of bacterial classes (Alphaproteobacteria, Betaproteobacteria, and Actinomycetia).

A preliminary screen of these homologs for NNG degradation activity was performed to identify homologs for further characterization. The H73A *Vs* NnlA variant, previously shown to lack NNG degradation activity, was used as a negative control [21]. *E. coli* transformants containing the expression vectors were incubated at 37 °C overnight in diluted lysogeny broth (LB) containing NNG and IPTG, the latter component was used to induce NnlA expression.

Overnight cultures of *Vs* NnlA and all five of the selected homologs exhibited nitrite formation as measured by the Griess assay (Figure 1). By contrast, cultures expressing H73A *Vs* NnlA lacked $NO_2^-$. This result strongly suggests that the $NO_2^-$ observed in the experimental samples resulted from NNG degradation activity by the recombinantly expressed NnlA homologs. We conclude from these results that all five of the selected NnlA homologs exhibit NNG degradation activity.

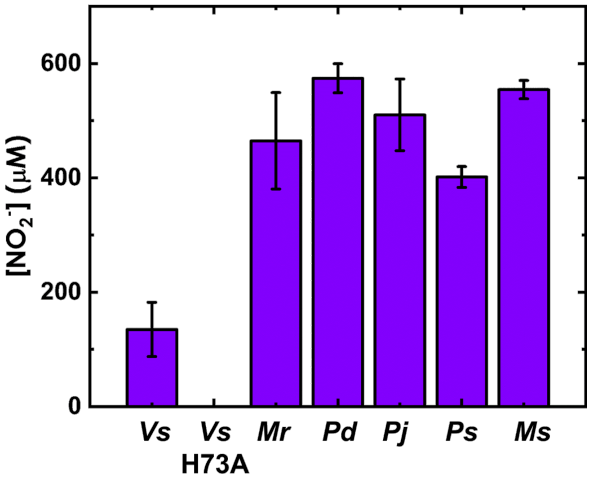



**Figure 1:** Nitrite concentrations observed in cultures of *E. coli* transformed with NnlA homologs or variants grown in the presence of NNG. Cells were incubated overnight in 1/5 LB, 20 μM IPTG, 3 mM NNG with appropriate antibiotics, incubated overnight at 37 °C.

## NnlA homologs exhibit similar heme and iron occupancy as *Vs* NnlA

To better compare the NnlA homologs to *Vs* NnlA, each homolog was recombinantly expressed in *E. coli* and purified. One of the five homologs, *Pj* NnlA, was found to be substantially insoluble and thus, could not be isolated. The remaining four homologs were isolated by immobilized metal affinity chromatography. While the most prominent band observed in the SDS-PAGE gel of each homolog is consistent with the expected monomer molecular masses of approximately 21 kDa, several other bands appear (Figure 2A). The banding patterns of each of the homologs are similar to those of *Vs* NnlA. These higher molecular weight bands are likely not contaminants but are either undissociated higher oligomer states or are oligomers whose formation is induced by SDS treatment, which has been observed for other proteins [29,30].

To characterize these oligomer states of native protein, analytical size exclusion chromatography data were collected (Figure 2B). As previously reported, *Vs* NnlA exhibited two major peaks, a lower molecular weight peak consistent with a dimer and a second peak consistent with a large oligomer [21]. By contrast, the chromatograms of the purified *Ps*, *Mr*, *Ms*, and *Pd* NnlA were dominated by a single peak ranging in molecular mass from 35.9 to 49.0 kDa, masses consistent with dimers (Table S2 in Supporting Information File 1). The higher oligomer peak was absent in all of these samples. As observed for *Vs* NnlA, there is no evidence for a significant population of monomer in any of these samples. From these data we conclude that these homologs exist mostly as dimers in solution.

Next, the heme incorporation of the isolated homologs was measured. UV–vis absorption spectra showed that each NnlA homolog exhibited characteristic Soret absorption features consistent with heme binding to the protein (Figure 3). In addition, the $A_{412\ nm}/A_{280\ nm}$ ratio for each homolog was greater than 1.0, consistent with high occupancy of heme incorporation in the proteins. Iron analyses of each of the homologs were consistent with this conclusion; the heme iron concentrations per protein were consistent with stoichiometric or nearly stoichiometric heme occupancy (Table 1).

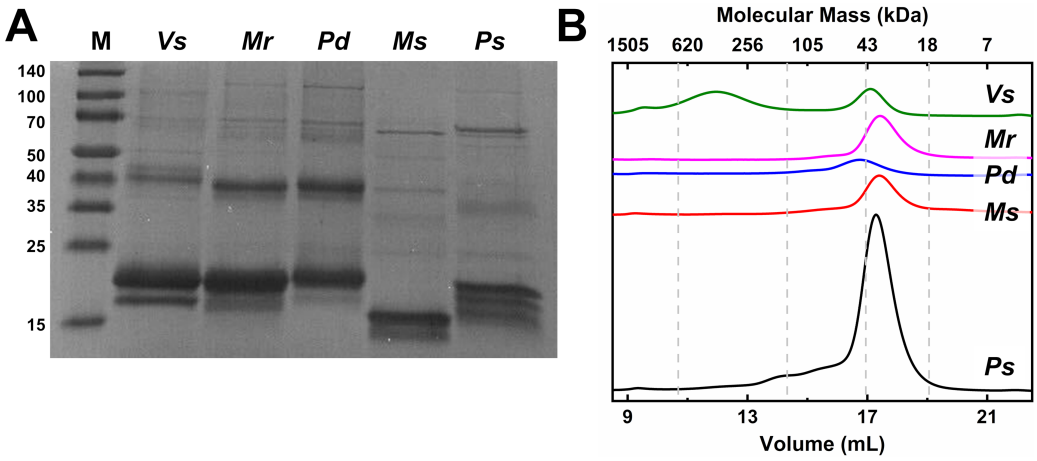



**Figure 2:** Molecular mass determination of purified NnlA homologs by A) SDS-PAGE or B) analytical size exclusion chromatography. Homolog labeled in figure. Mobile phase and sample buffer is 100 mM tricine 100 mM NaCl buffer at pH 7.5. Dashed grey lines represent elution volumes of molecular mass standards. Theoretical molecular weights are as follows *Vs* NnlA: 21,869 Da; *Mr* NnlA: 20638 Da; *Pd* NnlA: 18,473 Da; *Ms* NnlA : 18,239 Da; and *Ps* NnlA: 19,042 Da.

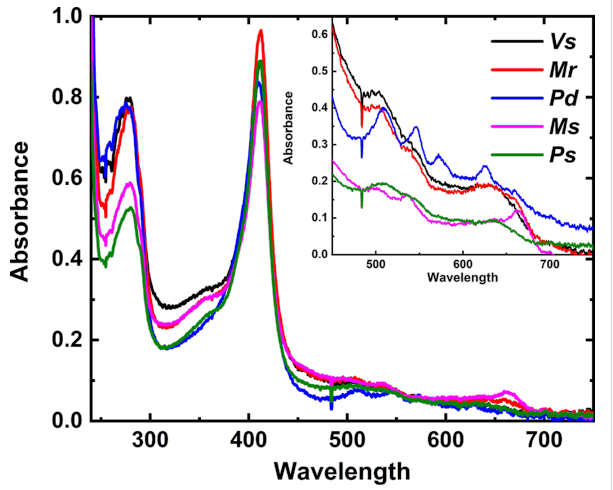



**Figure 3:** UV–vis absorption spectra of purified NnlA homologs. All spectra were measured in 100 mM tricine, 100 mM NaCl buffer at pH 7.5. NnlA homolog concentrations were *Vs* (10 μM), *Mr* (19 μM), *Pd* (18 μM), *Ms* (13 μM), and *Ps* (4 μM). Inset: Q-band region of the UV–vis spectra for concentrated NnlA homolog samples: *Vs* (90 μM), *Mr* (170 μM), *Pd* (160 μM), *Ms* (38 μM), and *Ps* (28 μM).

**Table 1:** Iron analyses of purified NnlA homologs.

| Sample | [Fe] (μM) | [NnlA] (μM) | [Fe] / [NnlA] |
|---|---|---|---|
| *Vs* | 260 ± 30 | 300 ± 60 | 0.87 ± 0.21[a] |
| *Ms* | 22.7 ± 1.8 | 16.7 ± 0.5 | 1.36 ± 0.08 |
| *Mr* | 141.8 ± 12.6 | 168.8 ± 12.1 | 0.84 ± 0.10 |
| *Pd* | 71.9 ± 12.1 | 158.9 ± 6.1 | 0.45 ± 0.08 |
| *Ps* | 67.4 ± 11.2 | 70.7 ± 0.6 | 0.95 ± 0.16 |

[a]Ref. [21]

## All homologs degrade NNG to glyoxylate, $NH_4^+$, and $NO_2^-$

Previous work showed that reduction of the *Vs* NnlA heme was required to activate NNG degradation. To test this requirement for the homologs, reduced samples of 5 μM of each NnlA homolog containing 350 μM NNG in deoxygenated 30 mM tricine buffer at pH 7.5 were incubated for one hour at 21 °C in an anaerobic glove box. The samples were analyzed by LC–MS to measure final glyoxylate and NNG concentrations. The extracted ion chromatograms (EICs) showed that NNG was completely consumed and glyoxylate accumulated within the incubation time (Figure 4).

The nitrogenous products in these samples were quantified by enzymatic and colorimetric assays to verify the nitrogen mass balance (Table 2). The data show that nearly stoichiometric concentrations of $NH_4^+$ and $NO_2^-$ are produced per mole of NNG as previously reported for *Vs* NnlA. Samples containing the as-isolated NnlA homologs without any reductant produced negligible concentrations of $NH_4^+$ and $NO_2^-$ (Table S3 in Supporting Information File 1). The combined results indicate each of the four purified NnlA homologs require reduction of the heme cofactor to initiate degradation of NNG to glyoxylate, $NH_4^+$, and $NO_2^-$.

## NnlA homologs do not degrade 2-NAE

Given the similarities in structure between NNG and 2-NAE, we sought to test if NnlA could also degrade 2-NAE. To test if NnlA could degrade 2-NAE, *E. coli* transformed with *Vs, Mr, Pd, Pj, Ps,* or *Ms* NnlA were incubated overnight at 37 °C in diluted LB containing 300 μM 2-NAE. Post-incubation treat-

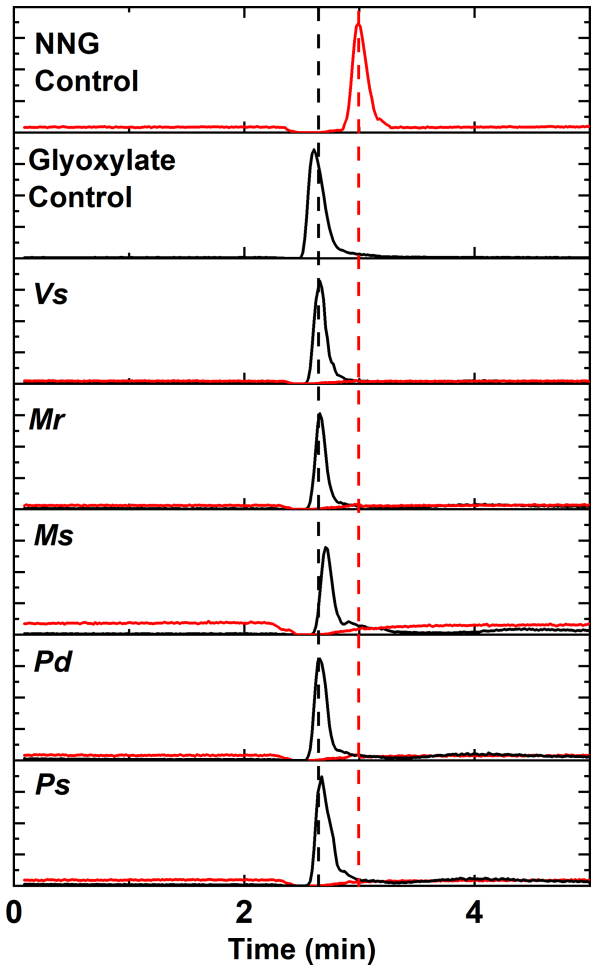



**Figure 4:** Representative LC–MS EICs monitoring molecular anions of NNG (*m/z* 119.01 ± 100 ppm) and glyoxylate (*m/z* 72.99 ± 100 ppm) in samples containing 350 μM NNG, 10 μM dithionite, and 5 μM of indicated NnlA homolog. Samples were incubated for approximately one hour at 21 °C in deoxygenated 30 mM tricine buffer at pH 7.5.

**Table 2:** Nitrogen mass balance resulting from NNG degradation by NnlA.

| NnlA[a] | $[NNG]_{final}$ (μM) | $[NH_4^+]_{final}$ (μM) | $[NO_2^-]_{final}$ (μM) |
|---|---|---|---|
| *Vs* | ND | 270 ± 30 | 240 ± 10 |
| *Mr* | ND | 300 ± 30 | 250 ± 10 |
| *Pd* | ND | 290 ± 30 | 260 ± 10 |
| *Ps* | ND | 320 ± 10 | 250 ± 10 |
| *Ms* | ND | 260 ± 10 | 250 ± 20 |

[a]Reaction conditions: 5 μM NnlA, 10 μM sodium dithionite, 350 μM NNG in 30 mM tricine buffer at pH 7.5 and room temperature in anaerobic glovebox. Mean values with standard deviations from triplicate reactions are shown.

ment of the samples with Griess assay revealed that all the cultures lacked $NO_2^-$ (Figure S3, Supporting Information File 1). Combined, these results show that none of the NnlA homologs can degrade 2-NAE.

To ensure the lack of 2-NAE degradation was unrelated to *E. coli* being unable to uptake 2-NAE, in vitro experiments with purified, reduced *Vs* NnlA incubated with 2 mM 2-NAE at pH 7.5 were performed. The LC–MS EICs monitoring 2-NAE (*m/z* 105.03 ± 100 ppm) show the prominent peak characteristic for 2-NAE. The intensity of this peak does not change in samples containing reduced *Vs* NnlA compared with samples without *Vs* NnlA. Additionally, $NO_2^-$ was not formed in these samples (Table S4, Supporting Information File 1).

## Alpha-fold model of NnlA

A model of the *Vs* NnlA dimer was produced using AlphaFold2, which allows for predicting the structure of oligomeric proteins [31]. The highest ranked model is shown in Figure 5. AlphaFold predicted a canonical α/β fold with high confidence, between residues Arg16 and Gly146 (Figure 5A). Structural clustering using Foldseek Cluster identified substantial similarities to other PAS domain proteins [32].

We sought to identify the heme binding site, but AlphaFold does not model this. However, this AlphaFold model was predicted to bind a heme cofactor by the consensus modeling tool COACH [33]. This protein–ligand model exhibited steric clashes with the heme and protein side chains (data not shown), limiting the use of this model to predict the heme environment. Nevertheless, this protein–ligand model heme binding between the β-sheet and an α-helix based on similarity to the oxygen-sensing dimeric DosH protein [34]. DosH is also a heme-binding PAS-domain containing protein, further validating the assignment of NnlA as a heme-binding PAS domain protein.

As previously reported for a structural homology model of *Vs* NnlA, the heme position was estimated by overlaying the AlphaFold model with the structure of *Pseudomonas aeruginosa* Aer2 (Figure 5B). By this method, the His73 is located near the heme, likely acting as its proximal ligand. As described above, the H73A *Vs* NnlA variant lacked NNG degradation activity and had reduced iron content [21]. The AlphaFold model also predicts a distal pocket composed of Tyr99, Ala101, Ile109, Val111, and Gln127. These amino acids are conserved in an alignment of orthologous protein sequences and could facilitate NNG hydrolysis in the active site (Figure S4, Supporting Information File 1). The function of these residues are being investigated. There are also several nearby conserved basic residues. The significance of these residues will be further dis-

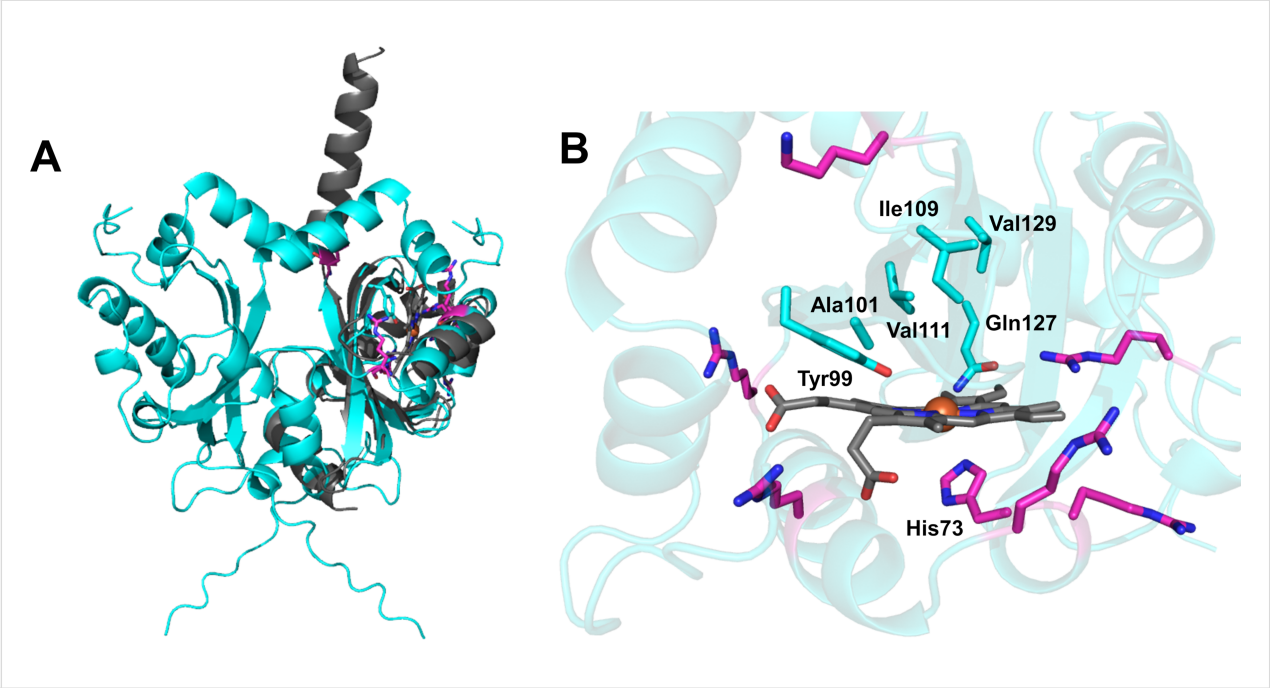



**Figure 5:** Alpha-fold model of *Vs* NnlA dimer (cyan) overlayed with *Pseudomonas aeruginosa* (grey; PDB: 3VOL [35,36]) to estimate the position and orientation of the heme cofactor. Conserved residues in the distal heme pocket are labeled. Conserved basic residues outside of the heme pocket are colored in magenta. Nitrogen, oxygen, and iron atoms are colored blue, red, and orange, respectively. Figure generated using PyMOL.

cussed below. Other conserved positions could be required for subunit association in the active homodimer.

## Discussion

The combined activity and characterization data indicate that each of the five homologs were similar to *Vs* NnlA in terms of oligomerization (Figure 2B), heme occupancy (Table 1 and Figure 3), NNG degradation activity (Figure 1, Figure 4 and Table 2) and the requirement for heme reduction to initiate NnlA activity (Table S3 in Supporting Information File 1). These similar protein characteristics persist despite the wide range of amino acid sequence identity between the tested homologs (46 to 76%). Therefore, NnlA is unlikely to have recently evolved to exploit anthropogenic nitramine contaminants in a similar fashion as RDX [16,28].

With evidence supporting that all of the homologs were able to degrade NNG, homologs of the *Variovorax nnlA* gene, including those described here, were identified by sequence similarity searches and used to infer phylogenetic relationships. An alignment of 11 amino acid sequences with 28 to 87% identity was prepared resulting in a maximum likelihood tree (Figure 6) [37]. This tree identifies clusters of sequences within taxonomic lineages, suggesting that the gene has been laterally transferred several times, within and among the Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, and Actinomycetes lineages. The combined data suggest that NNG degradation activity is found in diverse bacteria. Additionally, these bacteria were isolated from geographically distinct locations (Table S5 in Supporting Information File 1). These results strongly suggest that NnlA, and therefore NNG degradation activity is widespread amongst bacteria and in the environment.

There is no apparent conservation in the gene neighborhoods surrounding the NnlA homologs (Figure S5, Supporting Information File 1). Notably, the YjgF protein previously proposed to aid in deamination of imines – a proposed direct product of NNG degradation by NnlA – is absent [20]. This observation suggests this protein is not needed to aid in glyoxylate formation. Therefore, imine hydrolysis to glyoxylate may occur non-enzymatically or is catalyzed within the NnlA active site. However, an imine product has not yet been observed and further investigations of the NNG degradation mechanism are needed.

Given the potential widespread presence of NnlA, it is possible that NnlA could mediate previously reported nitramine biodegradations. However, prior work showed that NnlA was incapable of degrading nitroguanidine, cyclic nitramines (RDX, HMX), and the linear nitramines, NDAB, and *N*-nitroethylenediamine. As discussed above, several linear nitramines with hydroxy groups were shown to be biodegraded, including the carbon capture byproduct 2-NAE [19]. However, our results indicate that 2-NAE was not degraded either in our cell assays or overnight in the presence of isolated reduced NnlA (see Sup-

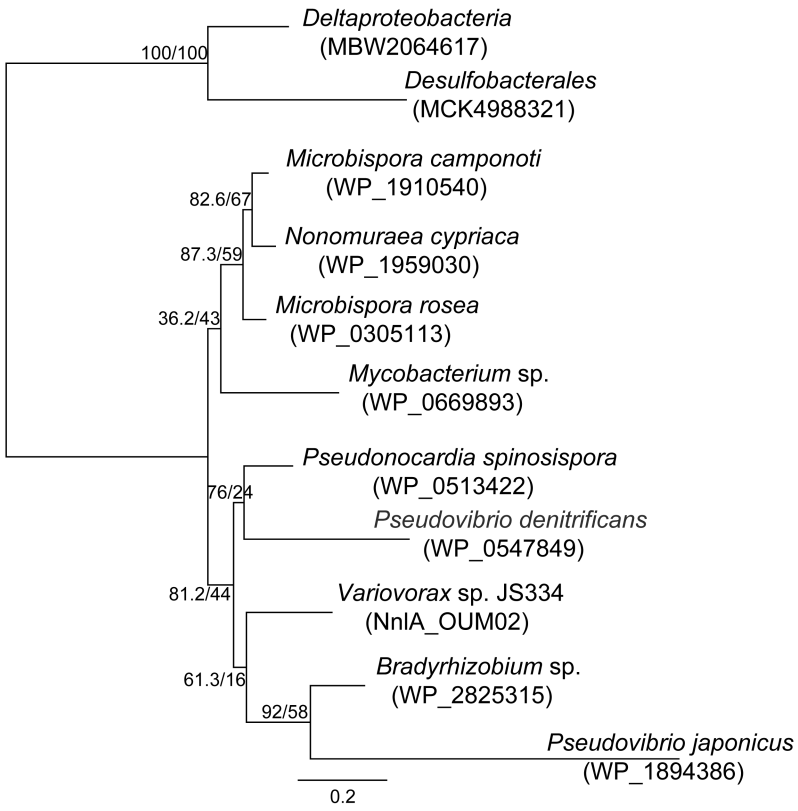


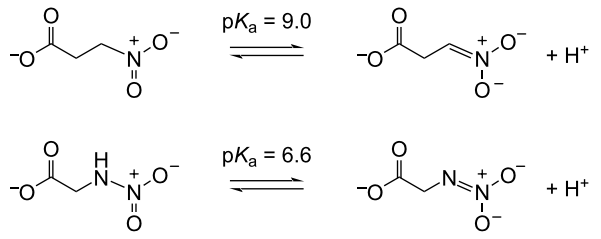

**Figure 6:** Phylogenetic tree of NnlA homologs with accession numbers. Branch lengths correspond to amino acid substitutions per position. Numbers at nodes indicate Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT) support (%) and ultrafast bootstrap support (%).

porting Information File 1, Figure S3 and Table S4). Therefore, NnlA appears to be specific for NNG degradation.

It is somewhat surprising that NnlA cannot degrade 2-NAE given its structural similarities with NNG. The structural difference between the two molecules is the replacement of the α-carboxylate of NNG with a hydroxy group in 2-NAE. While a decrease in substrate affinity might be expected, there was no activity even in the presence of millimolar concentrations of 2-NAE (Table S4, Supporting Information File 1). This observation suggests that binding in the substrate pocket is dependent on an electrostatic interaction with the NNG α-carboxylate, most likely from a lysine, arginine, histidine side chain or the N-terminus. An alignment of orthologous protein sequences reveals several conserved basic residues (Figure S4, Supporting Information File 1). The only conserved histidine residue is H73, which has been assigned as a heme ligand. Conserved arginine and lysine residues are colored magenta in the *Vs* NnlA model in Figure 5A. Several of these residues are nearby the predicted heme binding site, which may suggest their importance in NNG binding near the heme. However, many of these residues are either far from the heme or would not orient NNG towards the distal pocket of the heme. Future mutagenesis and kinetic experiments or crystallization of the active homodimer will be required to resolve the catalytic mechanism.

If NnlA is specific for NNG as suggested by these results, it is worth speculating about potential functions of NNG and other nitramine natural products. Bacterial natural products often exhibit antibiotic activity and it has been shown that NNG exhibits antibiotic activity towards Gram-negative bacteria (0.18 to 25 μg/mL) [24]. Moreover, the nitramine functional group has potential to serve as a potent warhead in an antibiotic. For example, a cytochrome P450 homolog, XplA, reductively decomposes the nitramine functionality of RDX to form $^{\bullet}NO_2$ [38], a toxic reactive nitrogen species. Additionally, NNG is a structural analog of another natural product 3-nitropropionate (3NP) found in plants and fungi [39]. This highly toxic compound inhibited succinate dehydrogenase and other metabolic enzymes. In addition, it has been shown to irreversibly inhibit isocitrate lyase 1 (ICL1) from *Mycobacterium tuberculosis* [40], and key metabolic protein for these pathogens [41]. Isocitrate lyases convert isocitrate to glyoxylate and succinate. Deprotonation of

3NP ($pK_a$ = 9.0) results in the formation of propionate 3-nitronate (P3N) as a conjugate base (Scheme 2) [39]. It is P3N that directly reacts with a cysteine in the ICL1 active site, forming a thiohydroxamate adduct that inhibits ICL1 turnover [40]. Additionally, the nitronate form of nitro acids has been proposed to behave as a transition analog of carboxylate groups, resulting in nitro compounds also acting as tight-binding reversible inhibitors [42]. Deprotonation of NNG also results in formation of the corresponding nitronate, albeit with a $pK_a$ of 6.6 [24], far lower than that for 3-NP (Scheme 2). Therefore, a much larger portion of NNG would be expected to exist as the inhibitory nitronate form at physiological pH, suggesting another potential role for nitramine groups as potent warheads in antibiotics.




**Scheme 2:** Acid–base equilibrium of 3-nitropropionate (3-NP) vs *N*-nitroglycine (NNG).

This antibiotic activity may also require further modification of NNG or its incorporation into a larger natural product. NNG is a non-proteinogenic amino acid, similar to other such N–N containing compounds such as piperazic acid and hydrazinoacetic acid [43]. These precursors are incorporated into larger NPs by non-ribosomal peptide synthases or polyketide synthases, and NNG may have a similar fate [44,45].

Another possibility is that NNG has several physiological functions and fates. For example, a natural product nitronate intermediate was recently shown to have two fates within *Streptomyces achromogenes var. streptozoticus* NRRL 3125 [46]. This nitronate intermediate was shown to be *O*-methylated to form *O*-methylnitronate, and subsequently incorporated into enteromycin. Alternatively, the intermediate could be denitrified by a nitronate monooxygenase (NMO) to produce $NO_2^-$. NnlA could replace NMO as the denitrifying enzyme in NNG producing bacteria, however, a BLAST search of NnlA in the genome of *Streptomyces noursei*, an NNG-producing bacterium, did not reveal any NnlA homologs. Interestingly, four NMOs are annotated in the *S. noursei* genome. These enzymes could protect *S. noursei* from NNG toxicity during its biosynthesis. Meanwhile, we posit that NnlA protects non-NNG producing bacteria from exposure. In vivo experiments comparing the toxicity of NNG towards wild-type cells expressing NnlA and NnlA knockout mutant strains would test this hypothesis.

## Conclusion

In this study the NNG degradation activity of five *Vs* NnlA homologs was screened in *E. coli* transformants, providing evidence that all five degrade NNG. Of these, four were fully isolated and characterized. Each isolated homolog exhibited similar oligomerization and heme occupancy as *Vs* NnlA. In addition, we confirmed by in vitro assays that initiation of NNG degradation activity by the NnlA requires reduction of the heme, verifying the necessity of the heme for NnlA activity. The nitrogen mass balance was consistent with NNG degradation to $NO_2^-$ and $NH_4^+$ as shown for *Vs* NnlA. It was also shown that NnlA cannot degrade the hydroxylated linear nitramine, 2-NAE. The combined data indicate that NnlA homologs specific for NNG degradation activity are found in diverse bacteria and environments. These results suggest the natural product NNG may also be found in diverse environments. The reactivity of the nitramine functionality begs for further studies confirming the natural abundance and physiological functions of NNG and other nitramine natural products.

## Experimental

### General reagents and protocols

Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 5-aminolevulinic acid (5-ALA) were purchased from Gold Biotechnology. NNG was purchased from AAblocks. 2-NAE was purchased from Toronto Research Chemicals. General buffers and media components were purchased from Fisher Scientific or VWR. Stock dithionite concentrations were determined by UV–vis absorbance at 318 nm ($\varepsilon_{318}$ = 8000 $M^{-1}cm^{-1}$). Water used for all solutions was of 18.2 MΩ·cm resistivity from a Barnstead Nanopure (Thermo Fisher Scientific). Solvents for LC–MS experiments were of at least HPLC grade and contained 0.1% vol/vol formic acid.

### Protein expression and purification

The vectors to express *Vs* NnlA and H73A *Vs* NnlA were previously reported [21]. *Pd*, *Mr*, *Ms*, *Ps*, and *Pj nnla* genes were synthesized as *E. coli* codon-optimized constructs and cloned into the NdeI and XhoI restriction sites of pET-28a(+)-TEV by GenScript.

For protein expression, plasmids were electroporated into *E. coli* BL21(DE3) cells and protein was expressed and purified by immobilized metal affinity chromatography (IMAC) as previously described for *Vs* NnlA [21]. The only modification was that plasmids using pET-28a(+)-TEV required 50 μg/mL

kanamycin instead of 100 μg/mL ampicillin. IMAC purified and concentrated protein were exchanged into 100 mM tricine, 100 mM NaCl buffer at pH 7.5 and stored at −60 °C.

## Protein characterization

Total iron concentrations in protein samples were quantified using an iron assay that allows for release and subsequent detection of heme-ligated iron [47]. Protein concentration was determined using bicinchoninic acid protein quantification assay (Pierce). The oligomeric state was determined by processing the protein through Superdex 200 Increase 10/300 GL analytical size exclusion column with 100 mM tricine with 100 mM NaCl at pH 7.5 as the mobile phase. Protein size exclusion chromatography standards (BioRad) were used to determine molecular masses.

## Nitramine degradation assays

LC–MS analysis was performed using an Agilent 1260 LC stack equipped with a Zorbax RX-C18 column (5 μm, 4.6 × 150 mm) and connected to an Agilent 6230 TOF mass spectrometer with electrospray ionization (ESI). Analyses used an isocratic mixture containing 65% water, 25% acetonitrile, and 10% isopropanol at a flow rate of 0.5 mL/min. The mass spectrometer was run in the negative ion mode with a probe voltage of 4,500 V and a fragmentation voltage of 175 V. To monitor NNG, 2-NAE, and glyoxylate, extracted ion chromatograms were obtained at *m/z* 119.0, 105.0, and 73.0, respectively.

Ammonium concentrations were determined using a glutamate dehydrogenase assay (Sigma-Aldrich) kit using the manufacturer's instructions. Nitrite concentrations were determined by reacting 25 μL aliquots of reaction sample with 25 μL of deoxygenated Griess reagent R1 (1% sulfanilamide in 5% $H_3PO_4$) followed by addition of 25 μL of deoxygenated Griess reagent R2 (0.1% naphthylethylenediamine dihydrochloride in water). The absorbance was read at 548 nm using an Infinite M200 Plate Reader (Tecan). Nitrite concentrations were determined by comparison of $A_{548\ nm}$ to a nitrite standard curve.

## Screening of *E. coli* transformants for NNG or 2-NAE degradation activity

Transformation of NnlA homologs were obtained as described above. The cells were then plated on LB agar plates containing ampicillin (100 μg/mL) or kanamycin (50 μg/mL) as appropriate and incubated overnight at 37 °C. Three colonies from each plate were picked with a toothpick and then resuspended in 0.2 mL of sterile water. Thirty microliter aliquots of suspended cells were used to inoculate 100 μL of selective growth media (1/5 LB media, 20 μM isopropyl β-D-thiogalactopyranoside (IPTG), 300 μM 2-NAE or 3 mM NNG, and antibiotic) in a 96-well plate. After overnight incubation at 37 °C, the cells were pelleted by centrifugation and the nitrite quantified by Griess assay as described above.

## Preparation of NNG and 2-NAE degradation samples

Triplicate samples containing 2 mM 2-NAE or 350 μM NNG, 40 μM titanium citrate with or without 20 μM *Vs* NnlA in deoxygenated 23 mM tricine at pH 7.5 were incubated overnight at 21 °C.

## Phylogenetic tree

Homologs of the *Variovorax nnlA* gene, including those described here, were identified by sequence similarity searches and their predicted amino acid sequences were used to infer phylogenetic relationships. An alignment of 11 amino acid sequences with 28 to 87% identity was prepared using MUSCLE software (ver. 5.1) [48] and trimmed to 148 positions in conserved blocks using Gblocks (ver. 0.91b) [49]. A maximum likelihood tree was inferred using IQ-TREE (ver. 2.2.2.6) with the LG+G4 substitution model [37].

# Supporting Information

### Supporting Information File 1

Additional Figures and Tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-75-S1.pdf]

# Acknowledgements

The world map in the graphical abstract was supplied by simplemaps.com. This content is not subject to CC BY 4.0.

# Funding

JDC and AAH were supported by funding from the United States Army Research Office award #W911NF2010286. BMR and ML were supported by a supplement to this parent grant provided by the Army Education Outreach Program (AEOP). This work was supported in part by the Strategic Environmental Research and Development Program (SERDP) under project WP20-1151. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

# ORCID® iDs

Kara A. Strickland - https://orcid.org/0009-0001-0565-6477
Brenda Martinez Rodriguez - https://orcid.org/0009-0001-2333-4295
David E. Graham - https://orcid.org/0000-0001-8968-7344
Jonathan D. Caranto - https://orcid.org/0000-0002-9196-5275

## Data Availability Statement

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

## Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2023.63.v1

## References


1. Crocker, F. H.; Indest, K. J.; Fredrickson, H. L. *Appl. Microbiol. Biotechnol.* **2006,** *73,* 274–290. doi:10.1007/s00253-006-0588-y
2. Pichtel, J. *Appl. Environ. Soil Sci.* **2012,** 617236. doi:10.1155/2012/617236
3. Abadin, H.; Ingerman, L.; Smith, C. *Toxicological profile for RDX;* Agency for Toxic Substances and Disease Registry (US): Atlanta, GA, USA, 2012.
4. Bachmann, W. E.; Sheehan, J. C. *J. Am. Chem. Soc.* **1949,** *71,* 1842–1845. doi:10.1021/ja01173a092
5. Lapointe, M.-C.; Martel, R.; Diaz, E. *J. Environ. Qual.* **2017,** *46,* 1444–1454. doi:10.2134/jeq2017.02.0069
6. D'Amico, L.; Blessinger, T.; Subramaniam, R.; Brinkerhoff, C. *Toxicological Review of Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX);* National Center for Environmental Assessment: Washington, DC, USA, 2018.
7. Maleh, H. K.; Carvalho-Knighton, K. M.; Martin, D. F. *Fla. Sci.* **2009,** *72,* 249–265.
8. Stone, W. J.; Paletta, T. L.; Heiman, E. M.; Bruce, J. I.; Knepshield, J. H. *Arch. Intern. Med.* **1969,** *124,* 726–730. doi:10.1001/archinte.1969.00300220078015
9. Balakrishnan, V. K.; Halasz, A.; Hawari, J. *Environ. Sci. Technol.* **2003,** *37,* 1838–1843. doi:10.1021/es020959h
10. Chatterjee, S.; Deb, U.; Datta, S.; Walther, C.; Gupta, D. K. *Chemosphere* **2017,** *184,* 438–451. doi:10.1016/j.chemosphere.2017.06.008
11. Fournier, D.; Halasz, A.; Thiboutot, S.; Ampleman, G.; Manno, D.; Hawari, J. *Environ. Sci. Technol.* **2004,** *38,* 4130–4133. doi:10.1021/es049671d
12. Sabir, D. K.; Grosjean, N.; Rylott, E. L.; Bruce, N. C. *FEMS Microbiol. Lett.* **2017,** *364,* fnx144. doi:10.1093/femsle/fnx144
13. Låg, M.; Lindeman, B.; Instanes, C.; Brunborg, G.; Schwarze, P. *Health effects of amines and derivatives associated with CO2 capture;* The Norwegian Institute of Public Health: Oslo, Norway, 2011.
14. Yu, K.; Mitch, W. A.; Dai, N. *Environ. Sci. Technol.* **2017,** *51,* 11522–11536. doi:10.1021/acs.est.7b02597
15. Paquet, L.; Monteil-Rivera, F.; Hatzinger, P. B.; Fuller, M. E.; Hawari, J. *J. Environ. Monit.* **2011,** *13,* 2304–2311. doi:10.1039/c1em10329f
16. Rylott, E. L.; Jackson, R. G.; Sabbadin, F.; Seth-Smith, H. M. B.; Edwards, J.; Chong, C. S.; Strand, S. E.; Grogan, G.; Bruce, N. C. *Biochim. Biophys. Acta, Proteins Proteomics* **2011,** *1814,* 230–236. doi:10.1016/j.bbapap.2010.07.004
17. Fournier, D.; Halasz, A.; Spain, J.; Spanggord, R. J.; Bottaro, J. C.; Hawari, J. *Appl. Environ. Microbiol.* **2004,** *70,* 1123–1128. doi:10.1128/aem.70.2.1123-1128.2004
18. Fournier, D.; Trott, S.; Hawari, J.; Spain, J. *Appl. Environ. Microbiol.* **2005,** *71,* 4199–4202. doi:10.1128/aem.71.8.4199-4202.2005
19. Brakstad, O. G.; Sørensen, L.; Zahlsen, K.; Bonaunet, K.; Hyldbakk, A.; Booth, A. M. *Int. J. Greenhouse Gas Control* **2018,** *70,* 157–163. doi:10.1016/j.ijggc.2018.01.021
20. Mahan, K. M.; Zheng, H.; Fida, T. T.; Parry, R. J.; Graham, D. E.; Spain, J. C. *Appl. Environ. Microbiol.* **2017,** *83,* e00457-17. doi:10.1128/aem.00457-17
21. Strickland, K. A.; Holland, A. A.; Trudeau, A.; Szlamkowicz, I.; Beazley, M. J.; Anagnostopoulos, V. A.; Graham, D. E.; Caranto, J. D. *Appl. Environ. Microbiol.* **2022,** *88,* e0102322. doi:10.1128/aem.01023-22
22. Parry, R.; Nishino, S.; Spain, J. *Nat. Prod. Rep.* **2011,** *28,* 152–167. doi:10.1039/c0np00024h
23. Graham, D. E.; Spain, J. C.; Parry, R. J.; Hettich, R. L.; Mahan, K. M.; Klingeman, D. M.; Giannone, R. J.; Gulvick, C. A.; Fida, T. T. *Nitration Enzyme Toolkit for the Biosynthesis of Energetic Materials;* Oak Ridge National Lab: Oak Ridge, TN, USA, 2018.
24. Miyazaki, Y.; Kono, Y.; Shimazu, A.; Takeuchi, S.; Yonehara, H. *J. Antibiot.* **1968,** *21,* 279–282. doi:10.7164/antibiotics.21.279
25. Alston, T. A.; Seitz, S. P.; Porter, D. J. T.; Bright, H. J. *Biochem. Biophys. Res. Commun.* **1980,** *97,* 294–300. doi:10.1016/s0006-291x(80)80167-7
26. Borer, K.; Hardy, R.; Lindsay, W.; Spratt, D.; Mees, G. *J. Exp. Bot.* **1966,** *17,* 378–389. doi:10.1093/jxb/17.2.378
27. Andeer, P. F.; Stahl, D. A.; Bruce, N. C.; Strand, S. E. *Appl. Environ. Microbiol.* **2009,** *75,* 3258–3262. doi:10.1128/aem.02396-08
28. Chong, C. S.; Sabir, D. K.; Lorenz, A.; Bontemps, C.; Andeer, P.; Stahl, D. A.; Strand, S. E.; Rylott, E. L.; Bruce, N. C. *Appl. Environ. Microbiol.* **2014,** *80,* 6601–6610. doi:10.1128/aem.01818-14
29. Matsui, T.; Kamata, S.; Ishii, K.; Maruno, T.; Ghanem, N.; Uchiyama, S.; Kato, K.; Suzuki, A.; Oda-Ueda, N.; Ogawa, T.; Tanaka, Y. *Sci. Rep.* **2019,** *9,* 2330. doi:10.1038/s41598-019-38861-8
30. Verhagen, M. F. J. M.; Voorhorst, W. G. B.; Kolkman, J. A.; Wolbert, R. B. G.; Hagen, W. R. *FEBS Lett.* **1993,** *336,* 13–18. doi:10.1016/0014-5793(93)81599-u
31. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. *Nat. Methods* **2022,** *19,* 679–682. doi:10.1038/s41592-022-01488-1
32. Barrio-Hernandez, I.; Yeo, J.; Jänes, J.; Mirdita, M.; Gilchrist, C. L. M.; Wein, T.; Varadi, M.; Velankar, S.; Beltrao, P.; Steinegger, M. *Nature* **2023,** *622,* 637–645. doi:10.1038/s41586-023-06510-w
33. Yang, J.; Roy, A.; Zhang, Y. *Bioinformatics* **2013,** *29,* 2588–2595. doi:10.1093/bioinformatics/btt447
34. Park; Suquet, C.; Satterlee, J. D.; Kang, C. *Biochemistry* **2004,** *43,* 2738–2746. doi:10.1021/bi035980p
35. Sawai, H.; Sugimoto, H.; Shiro, Y.; Aono, S. X-ray Crystal Structure of PAS-HAMP Aer2 in the CN-bound Form. https://doi.org/10.2210/pdb3vol/pdb. doi:10.1093/molbev/msaa015
36. Sawai, H.; Sugimoto, H.; Shiro, Y.; Ishikawa, H.; Mizutani, Y.; Aono, S. *Chem. Commun.* **2012,** *48,* 6523–6525. doi:10.1039/c2cc32549g
37. Minh, B. Q.; Schmidt, H. A.; Chernomor, O.; Schrempf, D.; Woodhams, M. D.; von Haeseler, A.; Lanfear, R. *Mol. Biol. Evol.* **2020,** *37,* 1530–1534. doi:10.1093/molbev/msaa015
38. Halasz, A.; Manno, D.; Perreault, N. N.; Sabbadin, F.; Bruce, N. C.; Hawari, J. *Environ. Sci. Technol.* **2012,** *46,* 7245–7251. doi:10.1021/es3011964
39. Francis, K.; Smitherman, C.; Nishino, S. F.; Spain, J. C.; Gadda, G. *IUBMB Life* **2013,** *65,* 759–768. doi:10.1002/iub.1195

40. Ray, S.; Kreitler, D. F.; Gulick, A. M.; Murkin, A. S. *ACS Chem. Biol.* **2018,** *13,* 1470–1473. doi:10.1021/acschembio.8b00225
41. McKinney, J. D.; zu Bentrup, K. H.; Muñoz-Elías, E. J.; Miczak, A.; Chen, B.; Chan, W.-T.; Swenson, D.; Sacchettini, J. C.; Jacobs, W. R., Jr.; Russell, D. G. *Nature* **2000,** *406,* 735–738. doi:10.1038/35021074
42. Alston, T. A.; Porter, D. J. T.; Bright, H. J. *Acc. Chem. Res.* **1983,** *16,* 418–424. doi:10.1021/ar00095a005
43. Hedges, J. B.; Ryan, K. S. *Chem. Rev.* **2020,** *120,* 3161–3209. doi:10.1021/acs.chemrev.9b00408
44. Wei, Z.-W.; Niikura, H.; Morgan, K. D.; Vacariu, C. M.; Andersen, R. J.; Ryan, K. S. *J. Am. Chem. Soc.* **2022,** *144,* 13556–13564. doi:10.1021/jacs.2c03660
45. Morgan, K. D.; Andersen, R. J.; Ryan, K. S. *Nat. Prod. Rep.* **2019,** *36,* 1628–1653. doi:10.1039/c8np00076j
46. He, H.-Y.; Ryan, K. S. *Nat. Chem.* **2021,** *13,* 599–606. doi:10.1038/s41557-021-00656-8
47. Fish, W. W. *Methods Enzymol.* **1988,** *158,* 357–364. doi:10.1016/0076-6879(88)58067-9
48. Edgar, R. C. *BMC Bioinf.* **2004,** *5,* 113. doi:10.1186/1471-2105-5-113
49. Castresana, J. *Mol. Biol. Evol.* **2000,** *17,* 540–552. doi:10.1093/oxfordjournals.molbev.a026334

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.75

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Confirmation of the stereochemistry of spiroviolene

Yao Kong[‡1], Yuanning Liu[‡1], Kaibiao Wang[1], Tao Wang[1], Chen Wang[1], Ben Ai[1], Hongli Jia[1], Guohui Pan[2], Min Yin[3] and Zhengren Xu[*1]

Letter

Open Access

Address:
[1]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China; Ningbo Institute of Marine Medicine, Peking University, Ningbo 315010, China, [2]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; University of Chinese Academy of Sciences, Beijing 100049, China and [3]School of Medicine, Yunnan University, 2 North Cui Hu Road, Kunming 650091, China

Email:
Zhengren Xu[*] - zhengrenxu@bjmu.edu.cn

* Corresponding author ‡ Equal contributors

Keywords:
boron migration; diterpene; spiroviolene; stereochemistry

*Beilstein J. Org. Chem.* **2024,** *20,* 852–858.
https://doi.org/10.3762/bjoc.20.77

Received: 28 December 2023
Accepted: 10 April 2024
Published: 18 April 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra



© 2024 Kong et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

We confirm the previously revised stereochemistry of spiroviolene by X-ray crystallographically characterizing a hydrazone derivative of 9-oxospiroviolane, which is synthesized by hydroboration/oxidation of spiroviolene followed by oxidation of the resultant hydroxy group. An unexpected thermal boron migration occurred during the hydroboration process of spiroviolene that resulted in the production of a mixture of 1α-hydroxyspiroviolane, 9α- and 9β-hydroxyspiroviolane after oxidation. The assertion of the *cis*-orientation of the 19- and 20-methyl groups provided further support for the revised cyclization mechanism of spiroviolene.

## Introduction

Terpenes represent one of the most fascinating families of natural products due to their structural complexity and diversity, as well as their indispensable biological functions that would be potentially applied as fragrances, pharmaceuticals etc. Until now, more than 80,000 terpenoid structures have been reported, which are found in all domains of life [1-3]. Despite their remarkable chemodiversity, the biosynthetic logic of terpenes is straightforward [4]. All terpenes are originated from two key C5 building blocks, namely isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), which are biosynthesized via either the methylerythritol phosphate (MEP) pathway or the mevalonic acid (MVA) pathway by using the primary metabolites. Different numbers of IPP and DMAPP are assembled by prenyltransferases to afford oligoprenyl pyrophosphates, such as farnesyl pyrophosphate (FPP, 3 × C5) and geranylgeranyl pyrophosphate (GGPP, 4 × C5), with varied C5 units. The linear oligoprenyl pyrophosphates are typically converted by terpene synthases in a chemo- and stereoselective process to form complex terpene skeletons, normally with multiple stereocenters. In this context, the 3D-defined cyclization products retain the rich information of the complex cyclization process. Thus, assignment of the stereochemistry of the

terpene skeleton with high confidence is crucial for proposing a reasonable cyclization mechanism [5].

Spiroviolene (**1**, Figure 1) was identified by Dickschat and co-workers as a nascent cyclization product of spiroviolene synthase (SvS), the coding gene of which was cloned from *Streptomyces violens* NRRL ISP-5597 [6]. Its unique spiro-fused linear triquinane to cyclopentane skeleton, as well as its stereochemistry, was originally elucidated as **1'** as shown in Figure 1, on the basis of detailed analysis of 1D and 2D NMR spectroscopy. Spiroviolene was also found to be produced by several bacterial strains harboring SvS homologs [6,7], as well as putative ancestors of SvS generated by ancestral sequence reconstruction [8,9]. Related natural products with the same 5-5-5-5 tetracyclic ring system, including spirograterpene A (**2**) from *Penicillium granulatum* MCCC 3A00475 [10], and GJ1012A (**3**) from an engineered *E. coli* strain harboring FgGS (FgJ07623) cloned from *Fusarium graminearum* GJ1012 [11], have been reported almost at the same time. The discrepancy of the stereochemistry at C3 between **1'** and **2** was first noticed by Snyder and co-workers [12]. The conserved stereochemistry of spiroviolene and **2** at C3 was later confirmed by the conversion of a synthetic intermediate of **2** to spiroviolene. By taking advantage of the DFT transition state analysis of the hydroboration reaction of a key intermediate, as well as NOE correlation analysis of the resultant product, Snyder and co-workers have reassigned the right structure of spiroviolene to **1**. However, direct evidences such as single-crystal X-ray diffraction results were not reported in their study.

The reassignment of the stereochemistry at C3 has resulted in the revision of the proposed cyclization mechanism [12-14]. The revised mechanism resembled the cyclization process for the formation of deoxyconidiogenol (**4**, Scheme 1A) by several terpene cyclases from fungus (PcCS, PchDS, PrDS) [15,16], which involves a 1,11-10,14 cyclization of GGPP, followed by 1,2-alkyl shift and a 2,10-cyclization, to give the key C3 cationic intermediate **IM-1**. A key 1,2-hydride shift from C2 to C3, which was observed in the isotope labeling experiments [6], followed by a 2,7-cyclization, afforded C6 cationic intermediate **IM-3** with cyclopiane skeleton. Quench of the cation **IM-3** with water would give **4**, while upon two 1,2-alkyl shifts of **IM-3**, followed by deprotonation of cation **IM-4**, would give spiroviolene (**1**).

On the other hand, the originally proposed cyclization mechanism (Scheme 1B) involves a 3,6-cyclization of cation **IM-1** through a conformation shown as **IM-1A** to generate cation **IM-5**, which was proposed to undergo a dyotropic rearrangement, followed by a 1,2-alkyl shift of cation **IM-6** to yield the spirocyclic cation **IM-7**. A key 1,3-hydride shift of **IM-7** from the β-face, followed by deprotonation of the formed C2-cation **IM-8**, would deliver the originally proposed structure **1'** [6]. However, no related natural products that would be derived from the intermediates of this pathway have been found so far.

A third cyclization mechanism (Scheme 1C) leading to the same spirocyclic skeleton as spiroviolene with an altered stereochemistry at C7 found in GJ1012A (**3**) could be proposed [11,17,18]. A 3,6- or 3,7-cyclization of cation **IM-1** through a conformation shown as **IM-1B** with β-oriented 20-methyl group, would generate either **IM-9** or **IM-11** cations. A direct dyotropic rearrangement, or two stepwise 1,2-alkyl migrations of **IM-9**, are possible pathways en route to cation **IM-10**. The presence of these intermediates **IM-9, -10, -11** could be inferred by the identification of GJ1012B/D (**5/6**, Scheme 1D) [11],

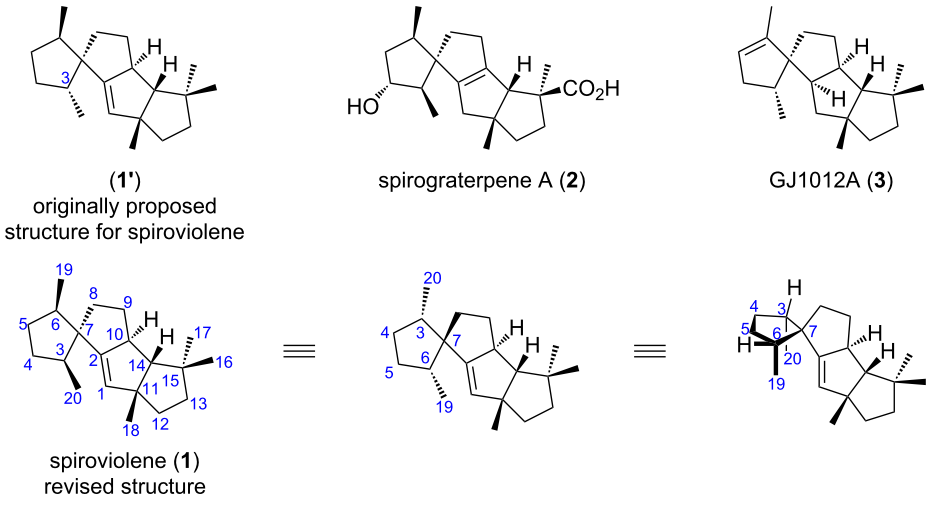
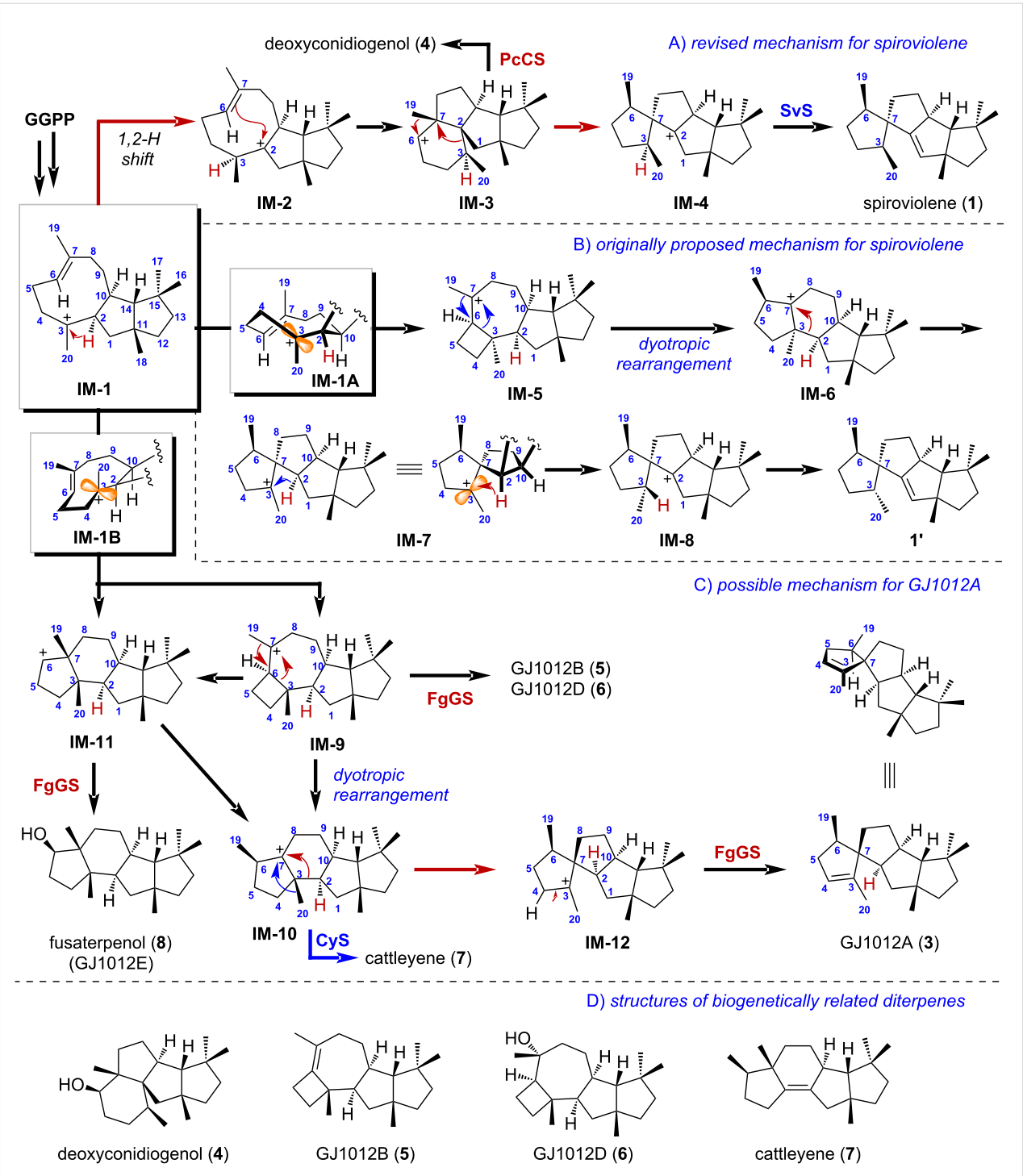



**Figure 1:** Structures of spiroviolene and related natural products.




**Scheme 1:** Possible cyclization mechanisms for spiroviolene (**1**) and related natural products. A) Revised cyclization mechanism for **1** that is related to deoxycondiogenol (**4**). B) Originally proposed cyclization mechanism for the misassigned structure **1'**. C) Possible cyclization mechanism for GJ1012A (**3**). D) Structures of biogenetically related diterpenes during the proposed cyclization process.

cattleyene (**7**) [19,20], and fusaterpenol (**8**, GJ1012E) [17]. A similar 1,2-alkyl shift of **IM-10**, followed by deprotonation of the formed spirocyclic cation **IM-12**, afforded **3**. Although previous isotope labeling experiments did not support this pathway for spiroviolene cyclization, it should be noted that a subtle alteration of stereochemical assignment of spiroviolene would have consequences for a different mechanistic proposal. We herein report the production of spiroviolene (**1**) in a hetero-

logous host by taking advantage of an artificial isopentenol utilization pathway [21-26], and confirm its stereochemistry by X-ray crystallography using a hydrazone derivative of **1**.

## Results and Discussion

Our work commenced with the heterologous production of spiroviolene by *E. coli* using a recently developed isopentenol utilization pathway for the efficient supply of two C5 precursors for terpene biosynthesis (Figure 2) [21-26]. In this artificially generated pathway, DMAPP and IPP could be easily generated from prenol and isoprenol, respectively, by the effect of two kinases, such as hydroxyethylthiazole kinase from *E. coli* (*Ec*ThiM) and isopentenyl phosphate kinase from *Methanocaldococcus jannaschii* (*Mj*IPK). Thus, we have cloned genes coding *Ec*ThiM, *Mj*IPK, isopentenyl diphosphate isomerase of *E. coli* (IDI), farnesyl diphosphate synthase of *E. coli* (IspA) and geranylgeranyl diphosphate synthase of *Pantoea agglomerans* (CrtE) into the multiple cloning site-2 of pCDFDeut-1 to give pCDFDeut-TIIAE for GGPP production. Also, we have cloned the SvS-coding gene directly from *Streptomyces violens* CGMCC 4.1786 (= NRRL ISP-5597) into pET28a to give pET28a-svs. The resultant two plasmids were then co-transformed into commercially available *E. coli* BL21(DE3) for diterpene production. Spiroviolene could be produced by feeding prenol and/or isoprenol to the fermentation broth after *E. coli* being induced by IPTG, and fermented at 18 °C for 72 h. In our hand, feeding a mixture of prenol and isoprenol in a 1:1 ratio would give the best yield of spiroviolene. GC–MS analysis (Figure 2A) of the EtOAc extract of the fermentation broth gave a single peak, whose EIMS spectrum (Figure 2B) matches that of spiroviolene. We then carried out a large-scale fermentation using 2 L-shake flasks, and each flask contains 1.0 L of modified TB medium. We have isolated 201 mg (40 mg/L) of spiroviolene from 5.0 L of the fermentation broth. The physicochemical data of the isolated material are consistent with those reported for spiroviolene [6].

With sufficient amount of spiroviolene in hand, we next attempted to obtain a crystalline compound suitable for X-ray diffraction by introducing functional groups (e.g., a hydroxy group, or a keto group) for further derivatization. Spiroviolene was not transformed when subjected to conditions for allylic oxidation ($SeO_2$) even at elevated temperature [27], and the starting material was fully recovered.

We have also tried hydroboration/oxidation conditions for transforming the double bond in a congested environment of spiroviolene (Scheme 2). Low conversion was observed when a tetrahydrofuran (THF) solution of **1** was treated with $BH_3$·THF at ambient temperature. The hydroboration reaction could be driven to synthetically useful yield when **1** was directly dissolved in 1 M $BH_3$·THF in THF, and heated at 60 °C for 3 days. After oxidative treatment of the resultant alkylborane products with $NaOH/H_2O_2$, we have obtained three derivatives **9**–**11** with one hydroxy group in 37%, 30% and 21% isolated yield, respectively, as well as recovery of 10% of the starting material. After detailed analysis of the NMR spectra, we have found that besides the normal hydroboration/oxidation product

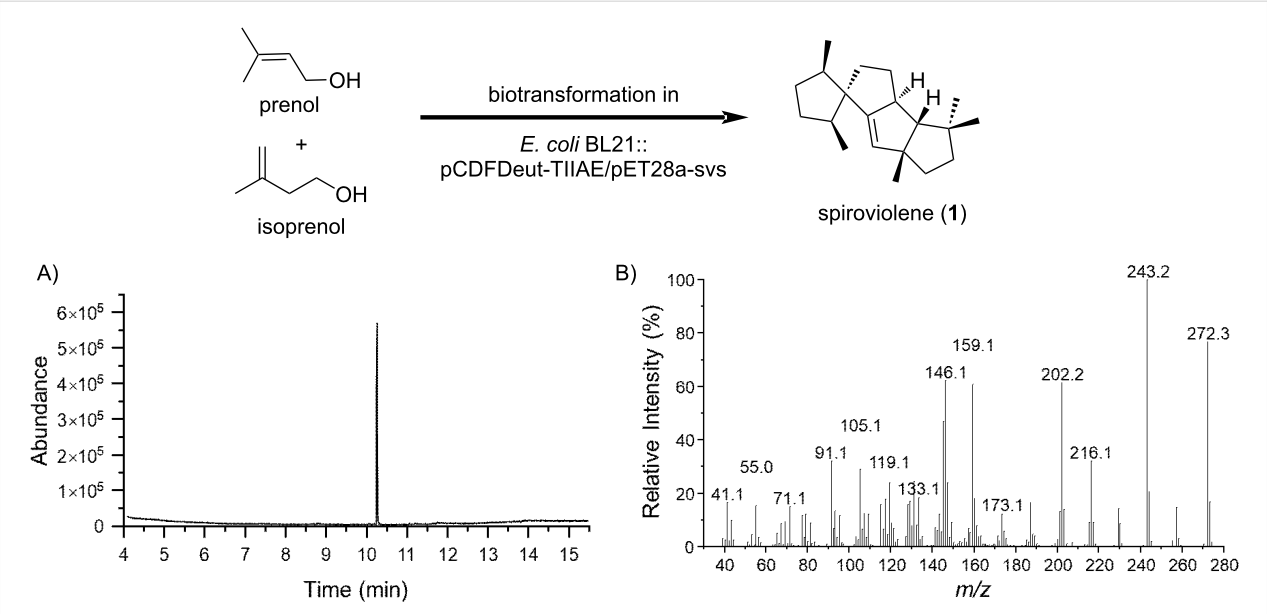
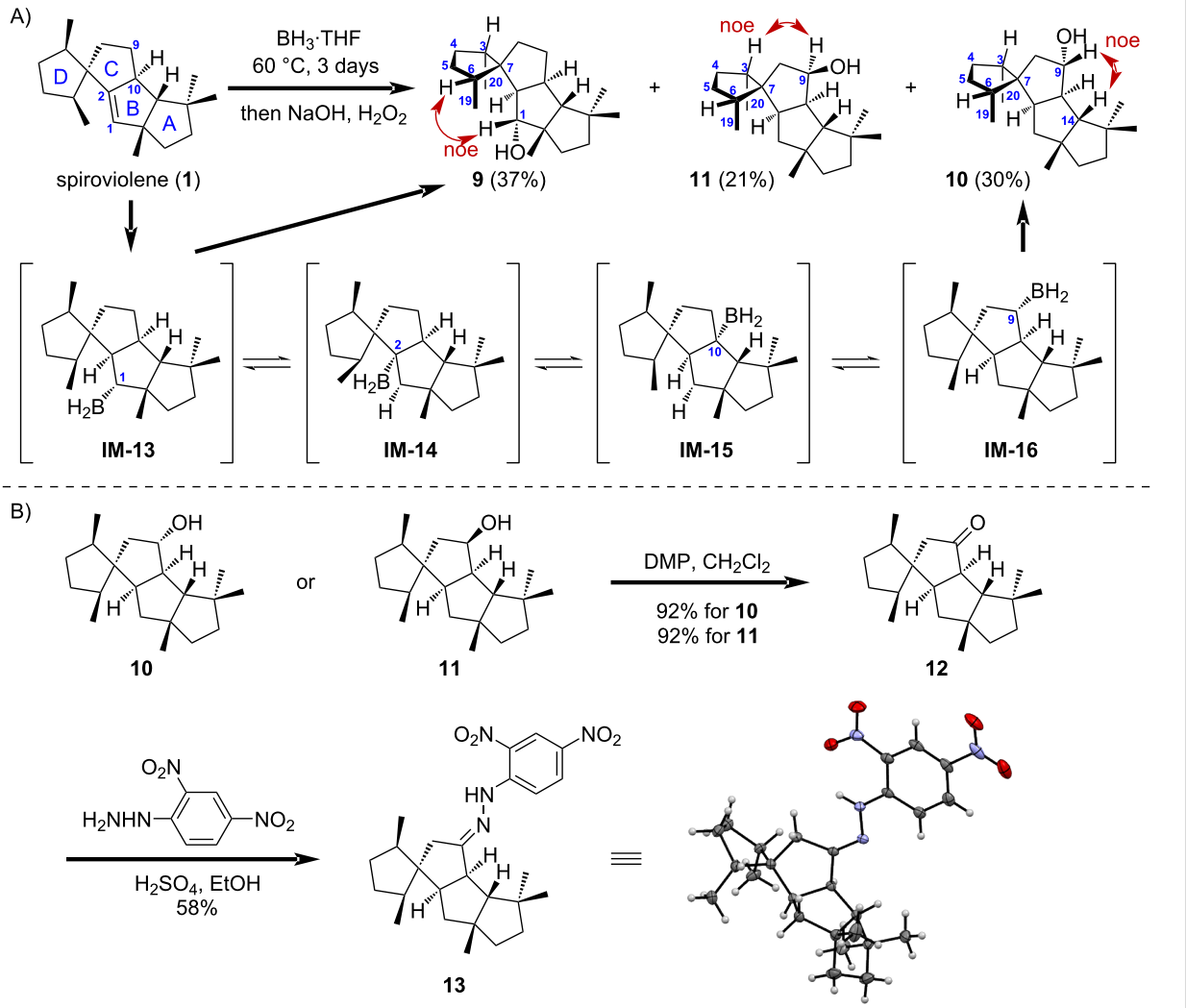



**Figure 2:** Heterologous production of spiroviolene using the isopentenol utilization pathway. A) Gas chromatogram of the EtOAc extract of the fermentation broth. B) EI mass spectrum of spiroviolene.




**Scheme 2:** Derivatization of spiroviolene for X-ray crystallography. A) Hydroboration/oxidation reaction of spiroviolene involving a borane migration process. B) Synthesis of the hydrazone derivative of spiroviolene for single-crystal X-ray diffraction.

1α-hydroxyspiroviolane (**9**), 9α- (**10**) and 9β-hydroxyspiroviolane (**11**) resulting from a formal boration at the homoallylic C9 position were also produced. The stereochemistry of the newly generated stereocenters was elucidated on the basis of their NOESY spectra. Thus, the key NOE correlations of H-1/H-6, H-1/$H_3$-19, and H-2/$H_3$-20 of **9** allowed to assign the 1-OH to be α-oriented, while correlations of H-9/H-14 of **10**, and H-9/H-3 of **11**, supported the assignment of 9α- and 9β-oriented hydroxy groups, respectively.

The formation of all three products **9–11** can be explained as follows (Scheme 2A) [28-31]: Due to the favorable formation of *cis*-5,5-fused B/C ring system, the borane reagent is preferred to approach the double bond of **1** from the α-face, to give either a secondary 1-organoborane intermediate **IM-13**, or a tertiary 2-organoborane intermediate **IM-14**. Oxidation of **IM-13** with $H_2O_2$/NaOH could stereoselectively furnish the normal hydroboration/oxidation product **9**. On the other hand, the unstable tertiary 2-organoborane **IM-14** would undergo two consecutive thermal 1,2-boron migrations to give the 9-organoborane intermediate **IM-16** probably through borane–olefin complexes. The suprafacial nature of the boron migration allowed the boron to be α-oriented in intermediate **IM-16**, which would give **10** with retention of the configuration after NaOH/$H_2O_2$ oxidation. The formation of a significant amount of the C-9 epimerization product **11** might be explained by a competing radical oxidation of the organoborane intermediate **IM-16** by oxygen when the oxidation step is opened to air [32,33]. On the other hand, an alternative mechanism for the formation of **11** involving the oxidation of the 9-*epi*-**IM-16** organoborane intermediate, which was formed by elimination of $BH_3$ from **IM-16** followed by re-addition of $BH_3$ from the opposite β-face to the proposed C8–C9 double bond intermediate, would also be possible.

To further advance the intermediate to crystalline hydrazone product (Scheme 2B), we have found that both **10** and **11** can be oxidized to the same 9-oxospiroviolane (**12**) in the same 92% isolated yield, hence confirming the structural assignment of **10** and **11**. By reacting with 2,4-dinitrophenylhydrazine [34], ketone **12** was further converted to hydrazone derivative **13**, which gave a brownish-yellow crystal suitable for X-ray diffraction [35]. The crystal structure of **13** clearly showed that the 19- and 20-methyl groups are *cis*-oriented in the D-ring which is consistent with that of spirograterpene A. This structural data reaffirms the revised structure of spiroviolene, and further support the unified cyclization process of fungi-derived deoxyconidiogenol and bacteria-derived spiroviolene by sharing the common C6-cation intermediate **IM-3** with cyclopiane skeleton (Scheme 1A).

## Conclusion

We have unambiguously confirmed the structural revision of spiroviolene with *cis*-oriented 19- and 20-methyl groups by obtaining a suitable crystal of the hydrazone derivative of 9-oxospiroviolane, which was synthesized by hydroboration/oxidation of spiroviolene followed by converting the resultant hydroxy group to a keto group, for single-crystal X-ray diffraction. The similar skeleton and conserved stereochemistry of spiroviolene and spirograterpene A therefore indicated that there must exist two terpene cyclases of different origins (Actinomycetes and Fungus) that were able to carry out similar chemical processes for the formation of the intriguing 5-5-5-5 tetracyclic ring system. Also, our study supports the proposed unified cyclization processes of spiroviolene and deoxyconidiogenol that bifurcate at the C6-cation intermediate **IM-3** with a cyclopiane skeleton. Thus, further mutational studies of these related terpene cyclases would give us more insights into the complex cyclization processes.

## Supporting Information

### Supporting Information File 1

Materials, synthetic methods, and copies of NMR spectra for all compounds.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-77-S1.pdf]

### Supporting Information File 2

Crystallographic information file of compound **13**.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-77-S2.cif]

## Funding

We are grateful to the National Natural Science Foundation of China (No. 81973197, 81991525, 22107008), Beijing Natural Science Foundation (No. Z210007), and Peking University Medicine Fund for world's leading discipline (BMU2022DJXK002) for financial support.

## ORCID® iDs

Tao Wang - https://orcid.org/0009-0001-9095-5374
Guohui Pan - https://orcid.org/0000-0002-8839-3886
Min Yin - https://orcid.org/0000-0003-2540-9016
Zhengren Xu - https://orcid.org/0000-0001-7302-6488

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2023.68.v1

## References

1. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
2. Minami, A.; Ozaki, T.; Liu, C.; Oikawa, H. *Nat. Prod. Rep.* **2018,** *35,* 1330–1346. doi:10.1039/c8np00026c
3. Rudolf, J. D.; Alsup, T. A.; Xu, B.; Li, Z. *Nat. Prod. Rep.* **2021,** *38,* 905–980. doi:10.1039/d0np00066c
4. Christianson, D. W. *Curr. Opin. Chem. Biol.* **2008,** *12,* 141–150. doi:10.1016/j.cbpa.2007.12.008
5. Nicolaou, K. C.; Snyder, S. A. *Angew. Chem., Int. Ed.* **2005,** *44,* 1012–1044. doi:10.1002/anie.200460864
6. Rabe, P.; Rinkel, J.; Dolja, E.; Schmitz, T.; Nubbemeyer, B.; Luu, T. H.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2017,** *56,* 2776–2779. doi:10.1002/anie.201612439
7. Lauterbach, L.; Rinkel, J.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2018,** *57,* 8280–8283. doi:10.1002/anie.201803800
8. Hendrikse, N. M.; Charpentier, G.; Nordling, E.; Syrén, P.-O. *FEBS J.* **2018,** *285,* 4660–4673. doi:10.1111/febs.14686
9. Schriever, K.; Saenz-Mendez, P.; Rudraraju, R. S.; Hendrikse, N. M.; Hudson, E. P.; Biundo, A.; Schnell, R.; Syrén, P.-O. *J. Am. Chem. Soc.* **2021,** *143,* 3794–3807. doi:10.1021/jacs.0c10214
10. Niu, S.; Fan, Z.-W.; Xie, C.-L.; Liu, Q.; Luo, Z.-H.; Liu, G.; Yang, X.-W. *J. Nat. Prod.* **2017,** *80,* 2174–2177. doi:10.1021/acs.jnatprod.7b00475
11. Bian, G.; Han, Y.; Hou, A.; Yuan, Y.; Liu, X.; Deng, Z.; Liu, T. *Metab. Eng.* **2017,** *42,* 1–8. doi:10.1016/j.ymben.2017.04.006
12. Chi, H. M.; Cole, C. J. F.; Hu, P.; Taylor, C. A.; Snyder, S. A. *Chem. Sci.* **2020,** *11,* 10939–10944. doi:10.1039/d0sc04686h
13. Xu, H.; Dickschat, J. S. *ChemBioChem* **2021,** *22,* 850–854. doi:10.1002/cbic.202000682
14. Sato, H.; Takagi, T.; Miyamoto, K.; Uchiyama, M. *Chem. Pharm. Bull.* **2021,** *69,* 1034–1038. doi:10.1248/cpb.c21-00536
15. Mitsuhashi, T.; Kikuchi, T.; Hoshino, S.; Ozeki, M.; Awakawa, T.; Shi, S.-P.; Fujita, M.; Abe, I. *Org. Lett.* **2018,** *20,* 5606–5609. doi:10.1021/acs.orglett.8b02284

16. Shiina, T.; Nakagawa, K.; Fujisaki, Y.; Ozaki, T.; Liu, C.; Toyomasu, T.; Hashimoto, M.; Koshino, H.; Minami, A.; Kawaide, H.; Oikawa, H. *Biosci., Biotechnol., Biochem.* **2019,** *83,* 192–201. doi:10.1080/09168451.2018.1536518
17. Rinkel, J.; Steiner, S. T.; Bian, G.; Chen, R.; Liu, T.; Dickschat, J. S. *ChemBioChem* **2020,** *21,* 486–491. doi:10.1002/cbic.201900462
18. Dickschat, J. S. *Angew. Chem., Int. Ed.* **2019,** *58,* 15964–15976. doi:10.1002/anie.201905312
19. Rinkel, J.; Steiner, S. T.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2019,** *58,* 9230–9233. doi:10.1002/anie.201902950
20. Xing, B.; Xu, H.; Li, A.; Lou, T.; Xu, M.; Wang, K.; Xu, Z.; Dickschat, J. S.; Yang, D.; Ma, M. *Angew. Chem., Int. Ed.* **2022,** *61,* e202209785. doi:10.1002/anie.202209785
21. Chatzivasileiou, A. O.; Ward, V.; Edgar, S. M.; Stephanopoulos, G. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116,* 506–511. doi:10.1073/pnas.1812935116
22. Clomburg, J. M.; Qian, S.; Tan, Z.; Cheong, S.; Gonzalez, R. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116,* 12810–12815. doi:10.1073/pnas.1821004116
23. Lund, S.; Hall, R.; Williams, G. J. *ACS Synth. Biol.* **2019,** *8,* 232–238. doi:10.1021/acssynbio.8b00383
24. Couillaud, J.; Rico, J.; Rubini, A.; Hamrouni, T.; Courvoisier-Dezord, E.; Petit, J.-L.; Mariage, A.; Darii, E.; Duquesne, K.; de Berardinis, V.; Iacazio, G. *ACS Omega* **2019,** *4,* 7838–7849. doi:10.1021/acsomega.9b00561
25. Johnson, L. A.; Dunbabin, A.; Benton, J. C. R.; Mart, R. J.; Allemann, R. K. *Angew. Chem., Int. Ed.* **2020,** *59,* 8486–8490. doi:10.1002/anie.202001744
26. Couillaud, J.; Leydet, L.; Duquesne, K.; Iacazio, G. *Genes* **2021,** *12,* 1974. doi:10.3390/genes12121974
27. Warpehoski, M. A.; Chabaud, B.; Sharpless, K. B. *J. Org. Chem.* **1982,** *47,* 2897–2900. doi:10.1021/jo00136a017
28. Brown, H. C.; Zweifel, G. *J. Am. Chem. Soc.* **1960,** *82,* 1504–1505. doi:10.1021/ja01491a058
29. Lhermitte, F.; Knochel, P. *Angew. Chem., Int. Ed.* **1998,** *37,* 2459–2461. doi:10.1002/(sici)1521-3773(19981002)37:18<2459::aid-anie2459>3.0.co;2-o
30. Laaziri, H.; Bromm, L. O.; Lhermitte, F.; Gschwind, R. M.; Knochel, P. *J. Am. Chem. Soc.* **1999,** *121,* 6940–6941. doi:10.1021/ja9843977
31. Hupe, E.; Denisenko, D.; Knochel, P. *Tetrahedron* **2003,** *59,* 9187–9198. doi:10.1016/j.tet.2003.09.043
32. Pelter, A.; Smith, K. Oxidation of carbon-boron bonds. In *Comprehensive Organic Synthesis;* Trost, B. M., Ed.; Pergamon Press: Oxford, UK, 1991; Vol. 7, pp 593–611. doi:10.1016/b978-0-08-052349-1.00203-1
33. Brown, H. C.; Midland, M. M.; Kabalka, G. W. *J. Am. Chem. Soc.* **1971,** *93,* 1024–1025. doi:10.1021/ja00733a040
34. Hearn, M. J.; Lebold, S. A.; Sinha, A.; Sy, K. *J. Org. Chem.* **1989,** *54,* 4188–4193. doi:10.1021/jo00278a036
35. Deposition number CCDC 2274944 for **13** contain the supplementary crystallographic data for this paper. These data are provided free of charge by the Cambridge Crystallographic Data Centre.

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.77

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Enhancing structural diversity of terpenoids by multisubstrate terpene synthases

Min Li[1,2,3] and Hui Tao*[1,2,3]

Review

Open Access

Address:
[1]Department of Otolaryngology, Zhongnan Hospital of Wuhan University, School of Pharmaceutical Sciences, Wuhan University, Wuhan, Hubei 430071, China, [2]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education, Wuhan University, Wuhan, Hubei 430071, China and [3]TaiKang Center for Life and Medical Sciences, Wuhan University, Wuhan, Hubei 430071, China

Email:
Hui Tao* - thui@whu.edu.cn

* Corresponding author

Keywords:
noncanonical terpene; substrate promiscuity; synthetic biology; terpene synthase; terpenoid

*Beilstein J. Org. Chem.* **2024,** *20,* 959–972.
https://doi.org/10.3762/bjoc.20.86

Received: 15 December 2023
Accepted: 09 April 2024
Published: 30 April 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf

© 2024 Li and Tao; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Terpenoids are one of the largest class of natural products with diverse structures and activities. This enormous diversity is embedded in enzymes called terpene synthases (TSs), which generate diverse terpene skeletons via sophisticated cyclization cascades. In addition to the many highly selective TSs, there are many promiscuous TSs that accept multiple prenyl substrates, or even noncanonical ones, with 6, 7, 8, 11, and 16 carbon atoms, synthesized via chemical approaches, *C*-methyltransferases, or engineered lepidopteran mevalonate pathways. The substrate promiscuity of TSs not only expands the structural diversity of terpenes but also highlights their potential for the discovery of novel terpenoids via combinatorial biosynthesis. In this review, we focus on the current knowledge on multisubstrate terpene synthases (MSTSs) and highlight their potential applications.

## Introduction

Terpenoids constitute the largest class of natural products with more than 80000 known structures [1] and a broad range of bioactivities [2,3]. Despite their stunning diversity, all terpenes are biosynthetically derived from two general isomeric $C_5$ building blocks, dimethylallyl diphosphate (DMAPP, **1**) and isopentenyl diphosphate (IPP, **2**), via the mevalonate (MVA) or methylerythritol 4-phosphate (MEP) pathways. These two isomeric $C_5$ precursors are further condensed by prenyltransferases (PTs) in successive elongation reactions, resulting in geranyl diphosphate (GPP, **3**), farnesyl diphosphate (FPP, **4**), geranylgeranyl diphosphate (GGPP, **5**), and longer prenyl diphosphates. The acyclic precursors are then converted into (poly)cyclic skeletons, including hemiterpenes, monoterpenes, sesquiterpenes, diterpenes, sesterterpenes, and triterpenes,

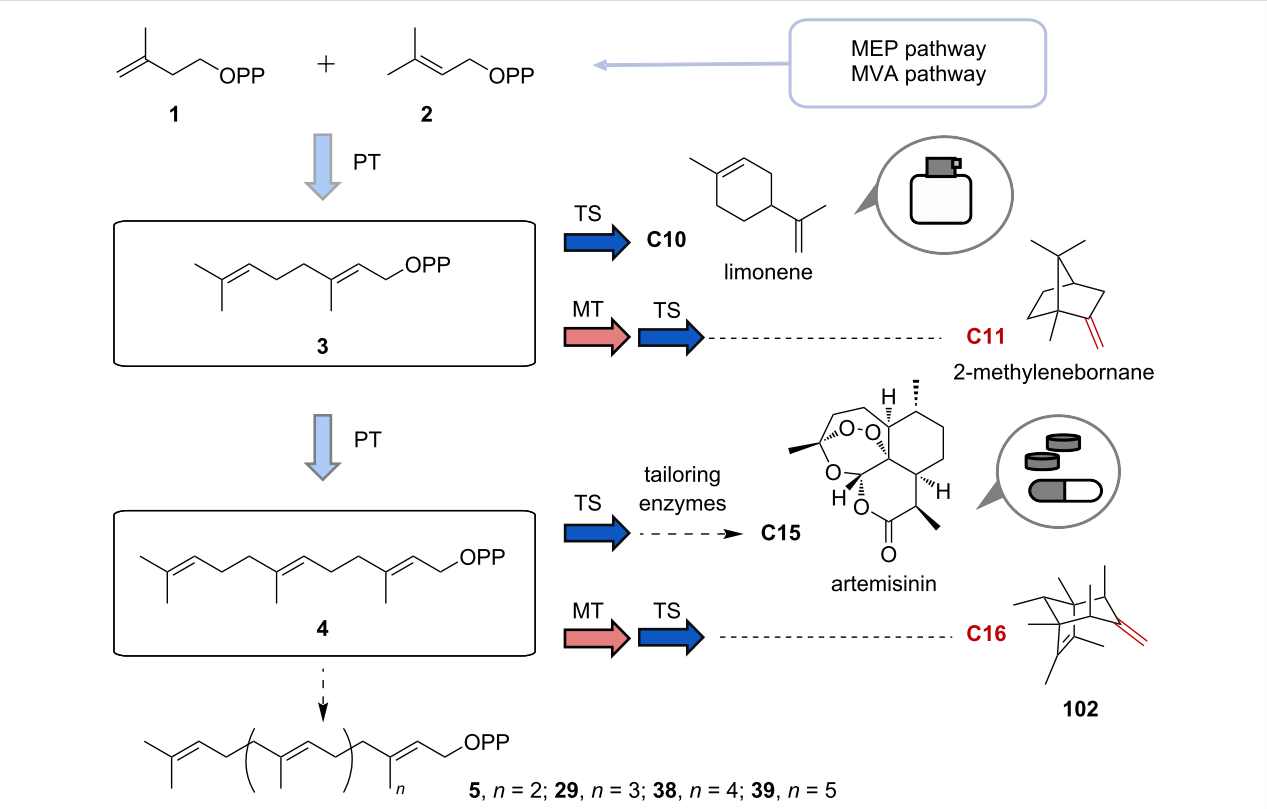



**Figure 1:** Biosynthetic pathway of terpenoids. Valuable terpenoids, noncanonical $C_{11}$ and $C_{16}$ terpenes are shown. MEP: methylerythritol 4-phosphate; MVA: mevalonate; PT: prenyltransferase; TS: terpene synthase; MT: methyltransferase.

by a large class of enzymes called terpene synthases (TSs) (Figure 1).

The reactions of TSs are one of the most important factors contributing to terpene diversity, as they often synthesize multiple products from a single substrate through complex cyclization cascades [4-10]. Based on the mechanism of initial carbocation generation, TSs generally fall into two main classes. Class I TSs generate an allylic cation from a prenyl substrate by depyrophosphorylation, whereas class II TSs utilize a general acid (a key Asp residue) to protonate the terminal C=C bond or epoxide group to yield a tertiary carbocation. The highly reactive carbocation is then converted to different carbocation intermediates, facilitated by the hydrophobic pocket of the TSs, which often results in multiple terpene products from a single prenyl substrate. In addition to their distinct mechanisms, the two major classes of TSs are classified according to their sequences, structures, and functions. For instance, class I TSs often have conserved sequence motifs, DDXXD and NSE/DTE, that bind trinuclear magnesium clusters for diphosphate abstraction, whereas class II TSs have a DXDD motif that acts as the catalytic acid. Recently, several novel unconventional TSs that share low sequence and structural similarities with classical TSs have been discovered and comprehensively reviewed [11,12].

In addition to the capability to generate multiple products using a single substrate, a growing number of TSs called multisubstrate terpene synthases (MSTSs) are capable of utilizing prenyl precursors with different chain lengths or configurations to synthesize diverse terpenoid products. Notably, MSTSs can also convert noncanonical prenyl substrates, including chemically synthesized analogs and bio-originated 6-, 7-, 8-, 11-, and 16-carbon substrates generated by methyltransferases or engineered lepidopteran mevalonate pathways. The multisubstrate features of these enzymes have often been characterized using in vitro assays. The in vivo activities of MSTSs were revealed by the development of an efficient precursor-providing chassis. The inherent features of MSTSs not only increase the structural diversity of terpenoids but also underscore their potential for generating new terpenoids through combinatorial biosynthesis. An important review published previously comprehensively addressed the transformation of synthetic prenyl-substrate analogs by TSs as well as TS-mimicking chemical transformations [13]. In this review, we discuss representative MSTSs originating from different species that use canonical prenyl substrates. We also highlight recent advances in the production of novel terpenoids by MSTSs using synthetic prenyl substrates. Finally, we focused on MSTSs that catalyze the transformation of naturally occurring noncanonical prenyl substrates.

# Review

## MSTSs using canonical prenyl diphosphate substrates

### MSTSs from plants

Substrate-promiscuous TSs are widely spread in plants, which mainly produce linear terpenoids such as linalool (**6**), (*E*)-nerolidol (**7**) and (*E,E*)-α-farnesene (**8**) (Figure 2) [14,15]. Most plant MSTSs accept two prenyl substrates: $C_5$ and $C_{10}$ [16], $C_{10}$ and $C_{15}$ [17-20], and $C_{15}$ and $C_{20}$ [18]. For instance, *Pam*Tps1 from *Plectranthus amboinicus* (Lour.) Spreng has been characterized as bifunctional in converting compounds **3** and **4**, respectively, to **6** and **7** both in vivo and in vitro (Table 1) [17]. In addition to the bifunctional plant TSs, a few plant MSTSs have

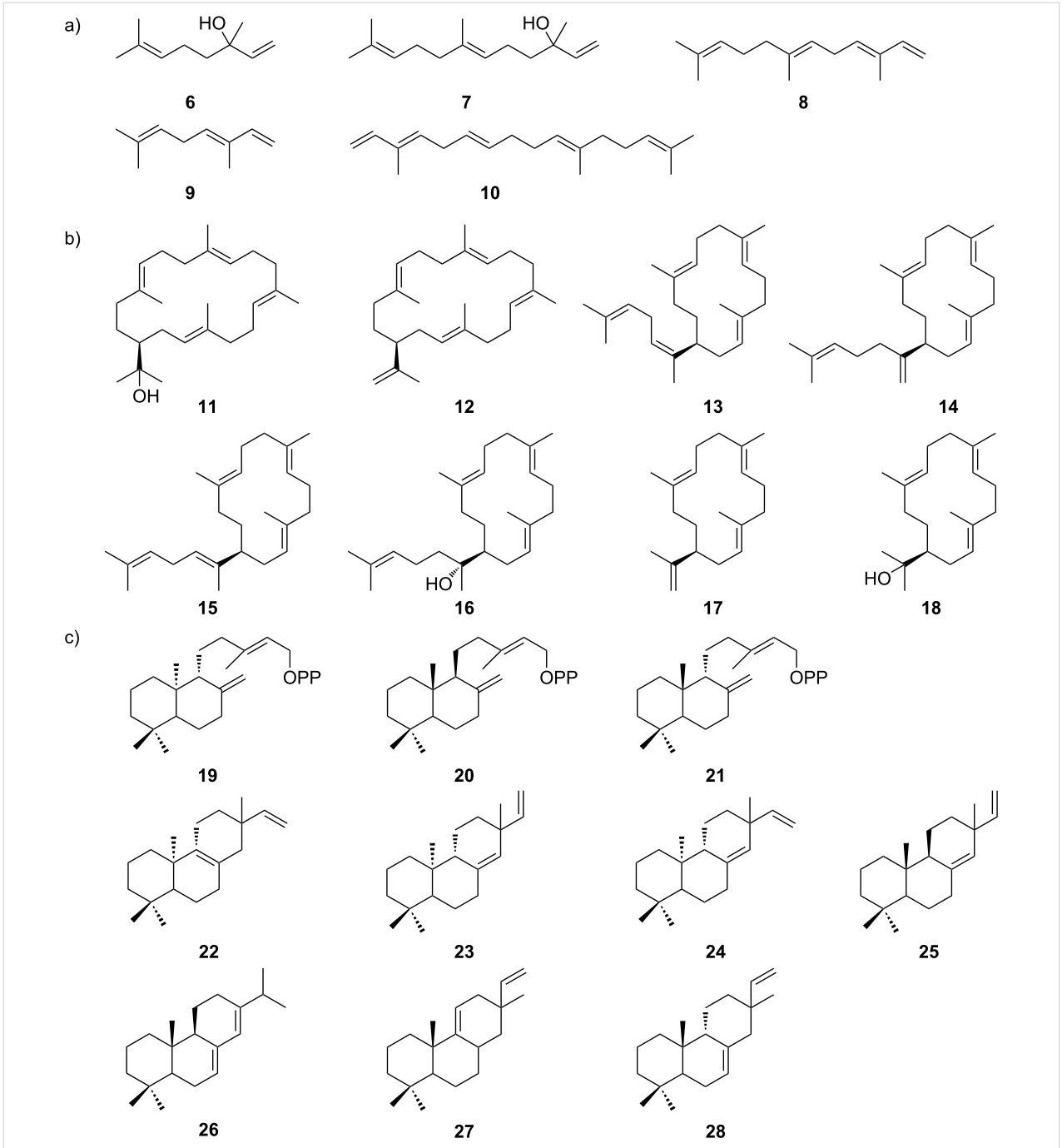



**Figure 2:** Representative terpenoids produced by plant MSTSs. a) **6** and **7** are products of *Pam*Tps1 from *Plectranthus amboinicus*, **8**–**10** are products of CoTPS5 from *Cananga odorata*; b) **11**–**18** products of LcTPS from *Leucosceptrum canum*; c) **22**–**28** are products of SiTPS from *Setaria italicais* using substrates **19**–**21**.

**Table 1:** MSTSs in plants.

| Enzymes | Organism | Substrates[a] | Number of characterized products | Reference |
|---|---|---|---|---|
| *Pam*Tps1 | *Plectranthus amboinicus* | GPP, FPP | 2 | [17] |
| PvTPS2 | *Prunella vulgaris* | NPP, FPP, (*Z,Z*)-FPP, GGPP, NNPP | 15 | [18] |
| PvHVS | | NPP, FPP, (*Z,Z*)-FPP, GGPP | 9 | [18] |
| PvTPS4 | | GPP, NPP, FPP, (*Z,Z*)-FPP | 10 | [18] |
| PvTPS5 | | GPP, NPP, FPP, (*Z,Z*)-FPP, NNPP | 11 | [18] |
| CoTPS5 | *Cananga odorata* | GPP, FPP, GGPP | 3 | [22] |
| AdAFS1 | *Actinidia deliciosa* | GPP, FPP | 2 | [23] |
| *Lc*TPS2 | *Leucosceptrum canum* | GPP, FPP | 8 | [19] |
| SiTPS | *Setaria italica* | FPP, *ent*-CPP, (+)-CPP, *syn*-CPP | 10 | [20] |

[a]GPP: geranyl diphosphate; FPP: farnesyl diphosphate; NPP: nerylpyrophosphate; GGPP: geranylgeranyl diphosphate; NNPP: nerylneryl diphosphate; CPP: copalyl pyrophosphate.

been characterized using multiple prenyl substrates. Recently, four TSs belonging to the TPS-a subfamily [21] from common self-healing (*Prunella vulgaris*) were characterized in vitro to accept **3**, neryl diphosphate (NPP), **4**, (*Z,Z*)-FPP, **5**, and/or nerylneryl diphosphate (NNPP) using purified recombinant enzymes (Table 1) [18]. A TPS-f subfamily enzyme CoTPS5 from *Cananga odorata* has been characterized to convert **3** to (*E*)-β-ocimene (**9**), **4** to **8**, and **5** to diterpene α-springene (**10**) (Figure 2) [22]. Both in vitro assays and in vivo transgenic expression of CoTPS5 confirmed the absence of side products, indicating that CoTPS5 is highly selective for individual prenyl substrates and that the reaction is tightly regulated under different conditions. Notably, CoTPS5 was the first plant TS to produce **10** [22].

Although many MSTSs exhibit a broad substrate scope in vitro, their product profiles may be altered in vivo owing to the subcellular localization of enzymes and the availability of substrates in different intracellular compartments [23]. For instance, two nerolidol/linalool synthases from *Antirrhinum majus* (AmNES/LIS-1, -2) both synthesize **6** and **7** in vitro, but cytosol-localized AmNES/LIS-1 produces only **7**, while plastid-localized AmNES/LIS-2 synthesizes **6** [24]. Similarly, in the case of CoTPS5, the transient expressed cytosol CoTPS5 in *N. benthamiana* only generated **8**, while the plastid-localized CoTPS5 yielded **9** and **10** other than **8** (Table 1). These studies indicate that redirecting MSTSs to different subcellular compartments may facilitate the generation of multiple terpenoids in plants.

Recently, in addition to linear terpenoid-producing TSs, MSTSs that form cyclic terpenoids have been discovered in plants, further increasing our understanding of chemodiversity and biosynthesis of plant terpenoids. *Lc*TPS2 from *Leucosceptrum canum* was characterized as a versatile TS that generated six macrocyclic sesterterpenoids (**11**–**16**) and two macrocyclic diterpenoids (**17**,**18**), representing the first macrocyclic terpenoids isolated from plants (Table 1, Figure 2) [19]. In addition to linear prenyl substrates, MSTSs can also accept partially cyclized substrates. A class I diterpene synthase SiTPS8 from *Setaria italicais* is capable of utilizing three copalyl pyrophosphate (CPP) stereoisomers that were generated by different class II TSs, including *ent*-CPP (**19**), (+)-CPP (**20**), and *syn*-CPP (**21**), to generate different diterpene skeletons **22**–**28**, which were further converted to diterpenoids by a P450 monooxygenase (CYP99A17) (Table 1, Figure 2) [20]. These findings will enable further investigation of the functions of terpenoids in plants and crops.

## MSTSs from fungi

Fungi are also prolific producers of terpenoids with diverse cyclic structures and important biological activities, which are of great interest. However, the number of known fungal MSTSs is currently limited, and researchers have focused on the promiscuity of their products rather than substrates [25-27]. Unlike plant MSTSs, fungal MSTSs convert natural substrates into cyclic skeletons. According to a phylogenetic tree constructed using 51 well-characterized class I TSs, clade III is of particular interest because most characterized di- and sester-TSs are enriched in this clade [28]. Two clade III TSs, FgMS and FgGS, from *Fusarium graminearum* J1-012 were characterized as promiscuous TSs with broad substrate specificities both in vitro and in vivo, indicating that TSs in clade III are more likely to be promiscuous. Using an efficient precursor-provid-

ing chassis, 50 terpenoids were generated via combinatorial biosynthesis using only two TSs and three PTs to generate **4**, **5** or geranylfarnesyl diphosphate (GFPP, **29**, Figure 1), representative products **30**–**33** are shown in Figure 3a [28]. Notably, FgMS is a chimeric enzyme (PTTS) consisting of an N-terminal class I TS domain and a C-terminal GFPP synthase domain. Therefore, to block the generation of **29**, a variant of FgMS-D510A with an inactive PT domain, rather than wild-type FgMS, was used in combinatorial biosynthesis. Furthermore, critical residues controlling substrate specificity were identified using site-directed mutagenesis. Interestingly, when the aromatic residue Phe65 was replaced with Ala, the resulting variant F65A produced a novel 5/8/6/6 tetracyclic sesterterpene in the presence of **29** [28]. Domain swapping is another useful approach for changing the PTTS product profile. For example, EvVS from *Emericella variecolor* majorly produced diterpene variediene (**34**) with a minor production of sesterterpene (2*E*)-α-cericerene (**33**) in vitro (Figure 3a) [29]. By replacing the PT domain of EvVS with that of sester-TS EvSS, the resulting variant generated **33**, which was not produced in vivo as the major product, both in vitro and in vivo. These studies revealed that altering and enhancing the supply of prenyl substrates can significantly change the product profile of promiscuous TSs, thereby generating terpenes with novel structures.

### MSTSs from bacteria

According to previous studies, plants and fungi are the major producers of terpenoids [30-33]. Recently, an increasing number of TSs have been discovered in bacteria [32,33]. VenA from *Streptomyces venezuelae* ATCC 15439 was characterized as a promiscuous class I TS that converts **3** to geraniol (24.2% yield), **4** to seven sesquiterpenes (24.6% yield), and **5** to four diterpenes (31.2% yield), with venezuelaene A (**35**, Figure 3b) as the predominant product in vitro. Notably, compound **35** has an unprecedented 5/5/6/7 tetracyclic skeleton [34]. In addition to *Actinomyces*, MSTSs have been discovered in *Bacillus*. The large-TS BclTS from *Bacillus clausii* generated β-geranylfarnesene (**36**) and β-hexaprene (**37**) from **29** and hexaprenyl diphosphate (HexPP, **38**, Figure 1), respectively. Similarly, a related TS, BalTS [35,36] from *Bacillus alcalophilus* was discovered to convert $C_{25}$, $C_{30}$, and $C_{35}$ prenyl diphosphates (**39**, Figure 1) into the corresponding β-prenes (**36**, **37**, and **40**, Figure 3b) [37]. Although BalTS shows no conserved motifs and distinct primary structures with class I TSs, its crystal structure reveals a similar overall structure of BalTS to the α-domain of class I TSs and therefore was proposed as class IB, a new subclass of TSs [37]. The discovery of TSs from bacteria not only expands the diversity of terpene skeletons but also the repertoire of TSs from nature.

## TSs using noncanonical prenyl diphosphate substrates

### Chemically synthesized noncanonical prenyl substrates

Noncanonical prenyl diphosphates are analogs of natural prenyl diphosphates. Most noncanonical prenyl diphosphate substrates are chemically synthesized. Classically, these prenyl analogs have been used as co-crystallization ligands [38], inhibitors of

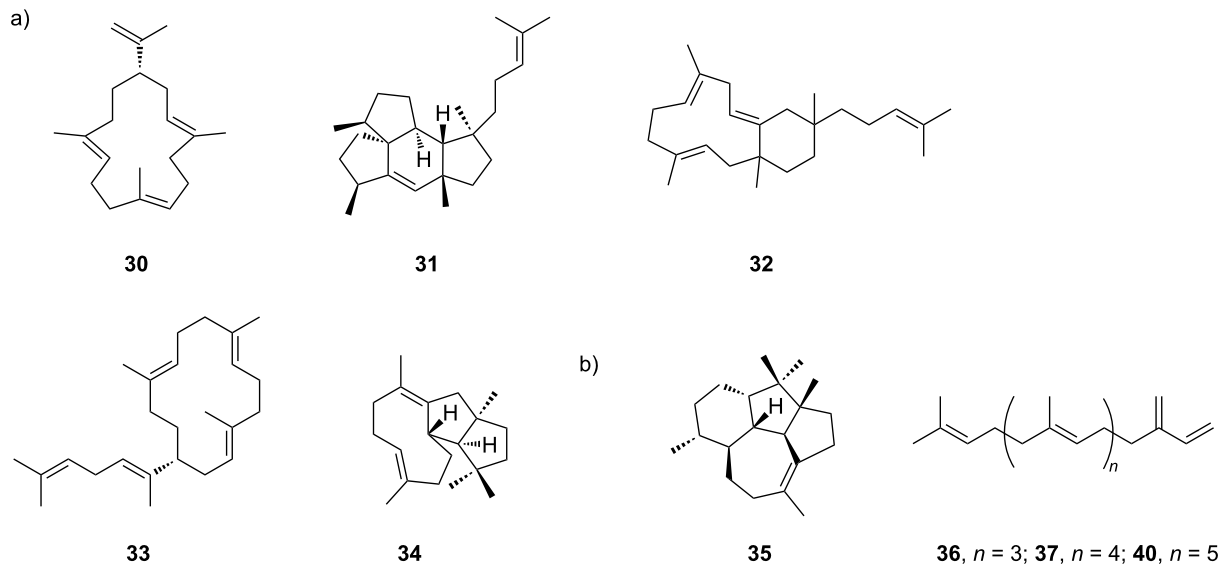



**Figure 3:** The structure of representative terpene products of MSTSs. a) From fungi: compounds **30**–**33** are produced by the fungal TS FgMS, **34** is the product of wild-type EvVS, and **33** is a new product of an EvVS variant with a swapped PT domain. b) From bacteria: compound **35** is a representative product of bacteria MSTSs VenA; compouns **36**, **37**, and **40** are products of two long β-prene TSs BclTS and BalTS.

specific TSs [39], and tools to study the reaction mechanisms of cyclization cascades [40,41] which have been comprehensively addressed in important previous reviews [8,13]. Currently, noncanonical prenyl analogs have been synthesized to act as actual substrates of TSs to generate novel terpene skeletons, introduce reaction handles, and produce value-added compounds. A previous review has covered the advances of TS-catalyzed transformations of synthetic substrate analogs up to 2019 [13,42]. Here, we provide updated examples on this topic.

Recently, novel sesquiterpene backbones **41**–**44** were synthesized by feeding presilphiperfolan-8β-ol synthase (BcBOT2) with methyl-shifted FPP analogs **45**–**47** (Figure 4a) [43]. Three new homosesquiterpenes **48**–**50** were produced after the

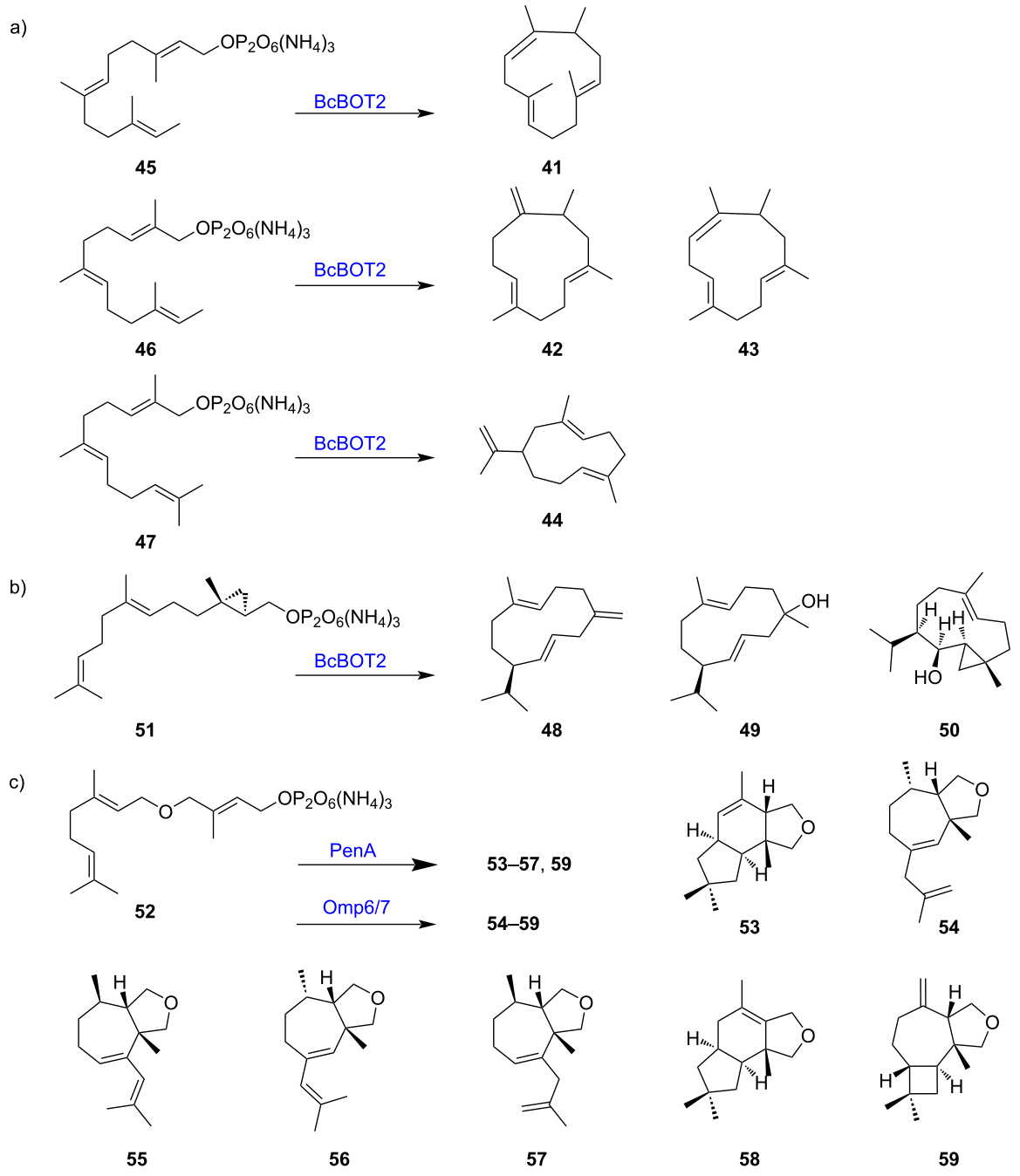



**Figure 4:** Terpenoid products of TSs using chemically synthesized noncanonical prenyl substrates. a) Products of BcBOT2 using iso-FPPs **45**–**47**. b) Biotransformation of **51** by BcBOT2. c) Products of two TSs PenA and Omp6/7 using FPP ether derivative **52**. PenA generated compounds **53**–**57** and **59**, whereas Omp6/7 produced **54**–**59**.

biotransformation of BcBOT2 with the FPP cyclopropylmethyl analog **51** (Figure 4b) [44]. With FPP ether derivative **52**, pentalenene synthase (PenA) and the $\Delta^6$-protoilludene synthases (Omp6/7) from *Omphalotus olearius* several new tetrahydrofurano terpenoids **53**–**58** were obtained, some of them accompanied with pronounced olfactoric properties (Figure 4c) [45]. Notably, one tetrahydrofuranoterpenoid **59** is also formed as a major product in the BcBoT2 reaction, despite the low sequence similarity between these sesqui-TSs, which could be explained by similar active-site conformations to stabilize prenyl substrates. Similarly, limonene synthase (CLS) from *Cannabis sativa* and 5-*epi*-aristolochene synthase (TEAS) from *Nicotiana tabacum* were incubated with 11 synthetic prenyl analogs with ether, thioether, alkyne, or phenyl groups, and six of them (**60**–**65**) were converted into several novel monoterpenoids **66**–**71** [46] (Figure 5). In addition to regular chemically synthesized prenyl analogs, chemoenzymatic approaches have been developed to obtain new sesquiterpenes from prenols, providing an alternative approach for accessing FPP analogs [47]. These studies demonstrate the potential of TSs to utilize noncanonical synthetic prenyl analogs to yield unusual terpenoid skeletons and new value-added terpenoids.

In addition to expanding the repertoire of terpenoids, the biotransformation of noncanonical prenyl substrates by TSs provides insights into the mechanisms of cyclization reactions. β-Himachalene synthase (HcS) and (*Z*)-γ-bisabolene synthase (BbS) from *Cryptosporangium arvum*, and germacrene A synthase (SmTS6) from *Streptomyces mobaraensis* were chosen to convert four FPP analogs **72**–**75**, which not only generated several new terpenoids (**76**–**79**), but also revealed the cyclization mechanisms of selected TSs [40] (Figure 6a). Similarly, two GGPP analogues **80** and **81** with shifted double bonds were synthesized to study the stereochemistry and cyclization mechanism of casbene synthase (CS) from the castor bean (*Ricinus communis*), which indicated a stereochemical course in accordance with the reported absolute configuration of casbene [41] (Figure 6b). The same GGPP isomers (**80**, **81**) were employed to generate novel diterpene derivatives and revealed the cyclization mechanisms of 12 di-TSs [48]. Similarly, dihydro-GGPP (**82**) and dihydro-GFPP (**83**) have been synthesized for biotransformation using several di- and sester-TSs. The conversion of analogues **82** and **83** by TSs led to the production of ruptenes including compounds **84**–**90**, which revealed the structure of the proposed intermediates for the cyclization reactions and there-

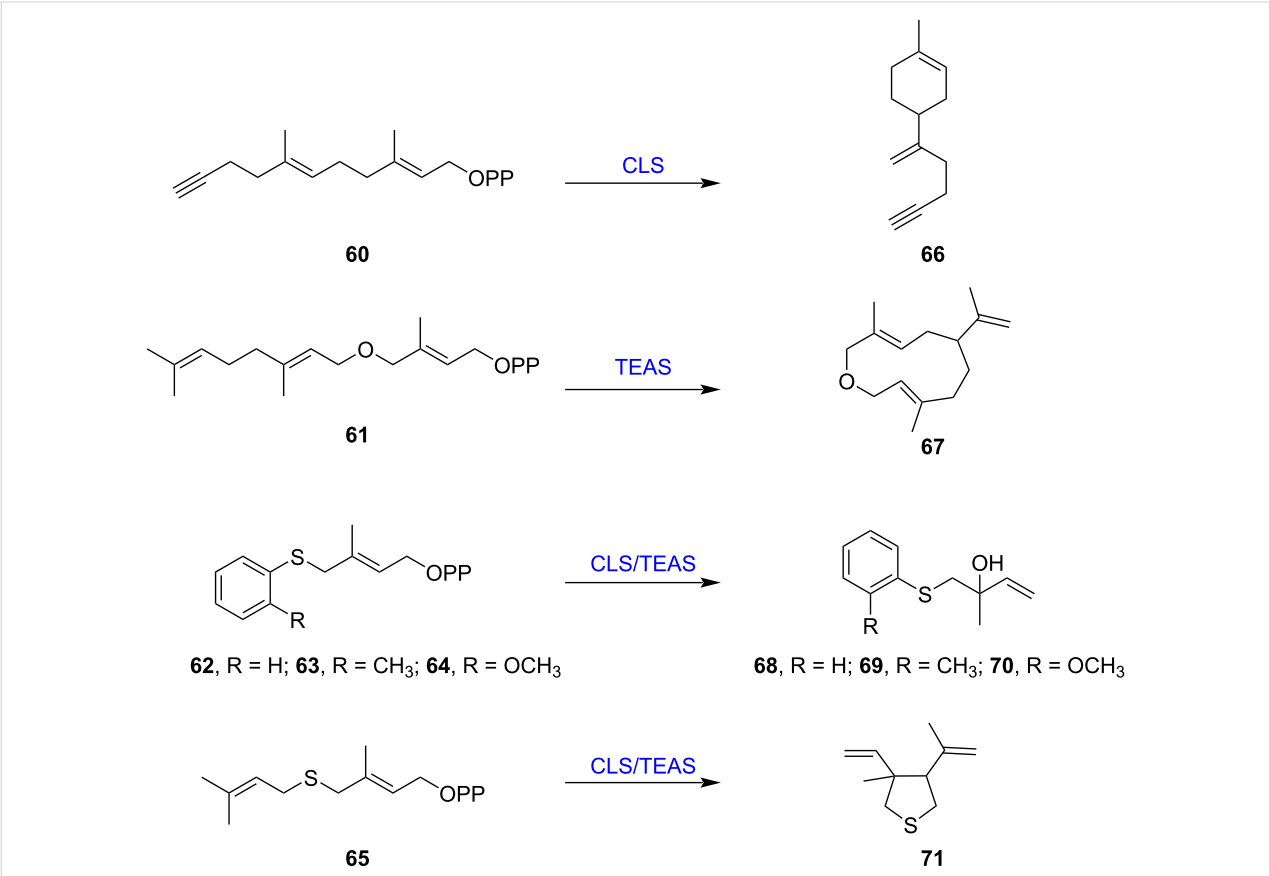



**Figure 5:** Biotransformation of noncanonical prenyl analogs **60**–**65** using two terpene synthases, limonene synthase (CLS) and 5-*epi*-aristolochene synthase (TEAS).

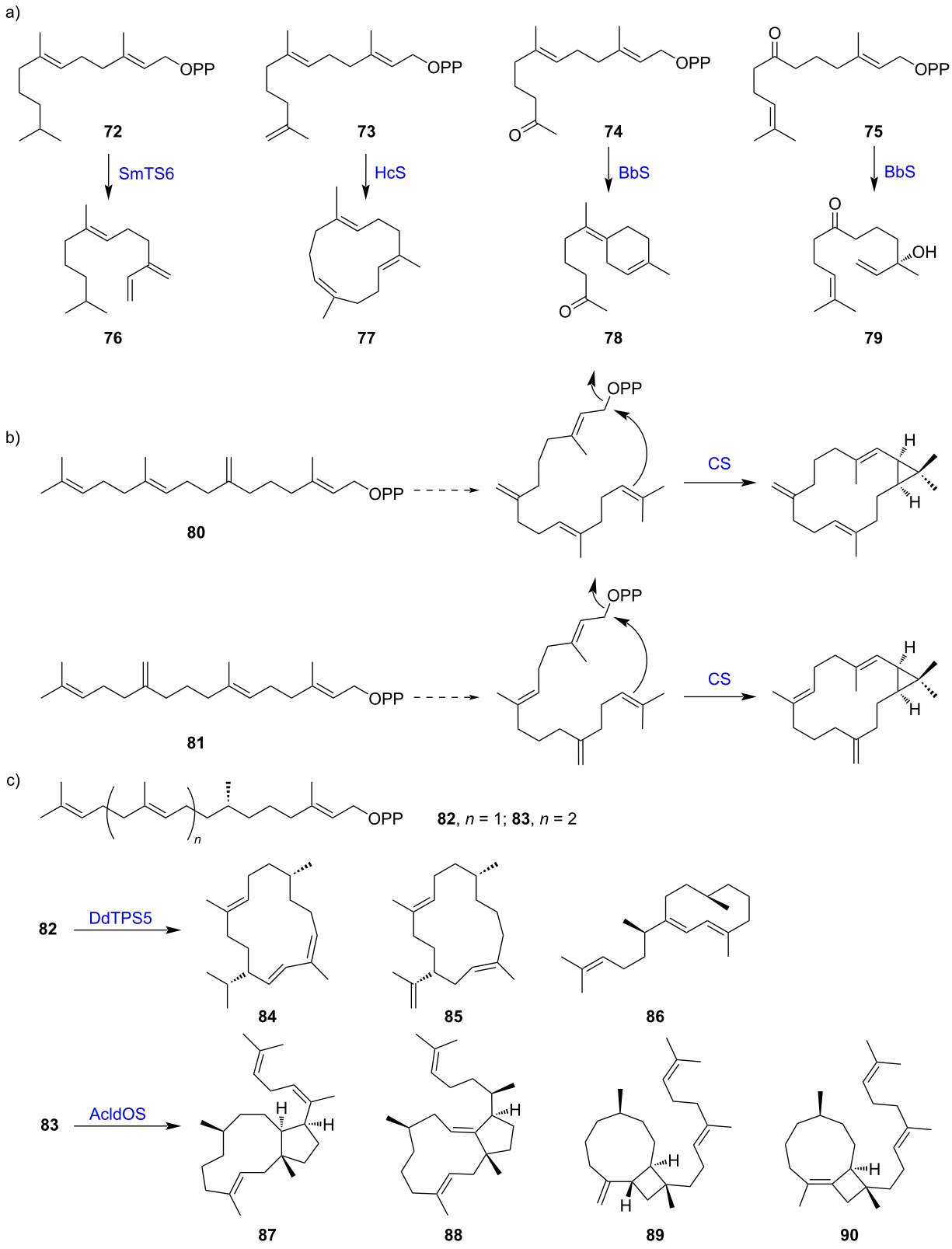



**Figure 6:** Noncanonical substrates for mechanism studies and their conversion to new terpenoids. a) New terpenoids from FPP analogs by β-himachalene synthase (HcS), (Z)-γ-bisabolene synthase (BbS), and germacrene A synthase (SmTS6). b) Mechanism study of casbene synthase (CS) by two GGPP isomers. c) Biotransformation of **82** and **83** to **84**–**90** by di-TS DdTPS5 and sester-TS AcldOS.

fore provided important insights into the reaction mechanism [49] (Figure 6c). With the aid of artificial prenyl analogs, a new route was developed to access a pool of unnatural terpenoids. It is worth noting that the rational design and synthesis of analogs play a valuable role in elucidating the cyclization mechanism of TSs, which further broadens our knowledge of the biosynthesis of terpenoids.

## Naturally occurring noncanonical prenyl substrates

Most of the terpene biosynthesis is well defined by the 'isoprene rule' to form natural products by the polymerization of $C_5$ isoprene. Although terpenes with irregular carbon atoms ($C_6$, $C_7$, $C_{11}$, $C_{12}$, $C_{16}$, and $C_{17}$) have been characterized, they are thought to be synthesized by modifications after the formation of the terpene skeletons [50]. Recently, additional routes have been discovered for the production of noncanonical terpenoids, whose biosynthesis requires *C*-methyltransferases from bacteria. IPP/DMAPP methyltransferases have been shown to convert $C_5$ prenyl substrates to irregular $C_6$ (**91**–**95**), $C_7$ (**96**–**100**), and $C_8$ diphosphates (**101**), which could serve as building blocks for the generation of new terpenoids [51] (Figure 7). Furthermore, a series of noncanonical $C_{11}$, $C_{12}$, $C_{16}$, and $C_{17}$ prenyl substrates were synthesized in *Escherichia coli* harboring heterologously expressed IPP methyltransferase (IPPMT) from *Streptomyces monomycini*. Notably, polymethylated $C_{41}$, $C_{42}$, and $C_{43}$ carotenoids were produced by combining the endogenous terpene biosynthesis pathway and IPPMT, demonstrating the potential of this approach to expand the terpene structural space [52].

In addition to methylation of the elongation unit IPP, noncanonical prenyl substrates can also be prepared by modifying the prenyl substrate of TSs. For instance, the heterologous expression of GPP C2-methyltransferases with $C_{11}$-TSs and mevalonate biosynthesis enzymes in *E. coli* yielded 35 $C_{11}$ terpenes and 11 $C_{16}$ terpenes [53]. By introducing a GPP C2-methyltransferase from *Pseudanabaena limnetica* to yeast together with an engineered $C_{11}$-specific TS, 40 $C_{11}$ terpene scaffolds were produced, which significantly increased the chemical space of terpenoids [54]. More recently, the GPP C6-methyltransferase BezA was discovered in *Streptomyces coelicolor* [55] (Figure 8a). Further structure-based engineering of BezA successfully repurposed it to catalyze the unprecedented C6-methylation of FPP by a single residue substitution in its substrate-binding pocket [55]. Moreover, efforts have also been made to engineer the TSs to modulate their product selectivity with the noncanonical prenyl substrates. To enable the biotechnological synthesis of irregular terpenes, the product selectivity

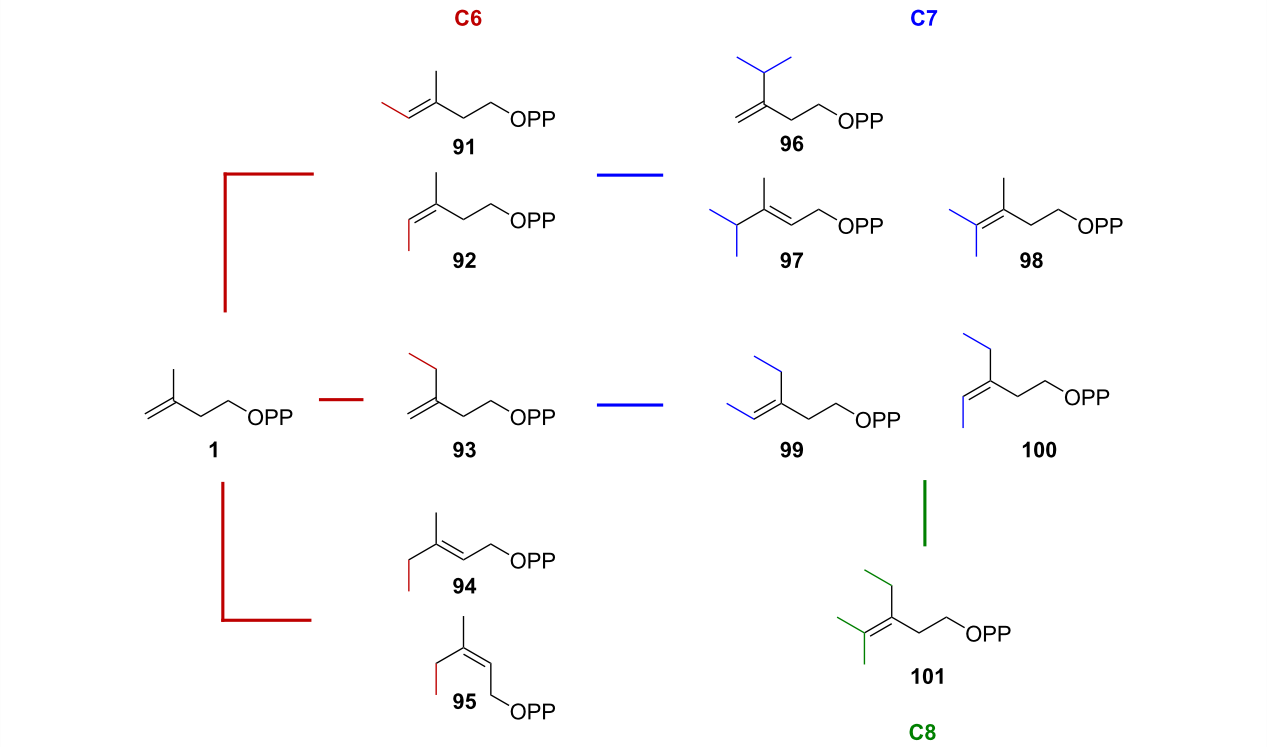



**Figure 7:** A series of prenyl diphosphate building blocks produced by different IPP methyltransferases. humMT from *Micromonospora humi* produced compounds **92**, **93**, and **99**–**101**; catMT from *Streptomyces catenulae* produced compounds **93**–**95** and **98**; argMT from *Streptomyces argenteolus* produced compounds **91**–**94** and **96**; azuMT from *Amycolatopsis azurea* produced **91**; fasMT from *Rhodococcus fascians* produced **93** and **94**; fraMT from *Frankia* sp. Produced **94** and **95**; monMT from *Streptomyces monomycini* produced compounds **91**, **93**, **94**, and **96**–**98**.

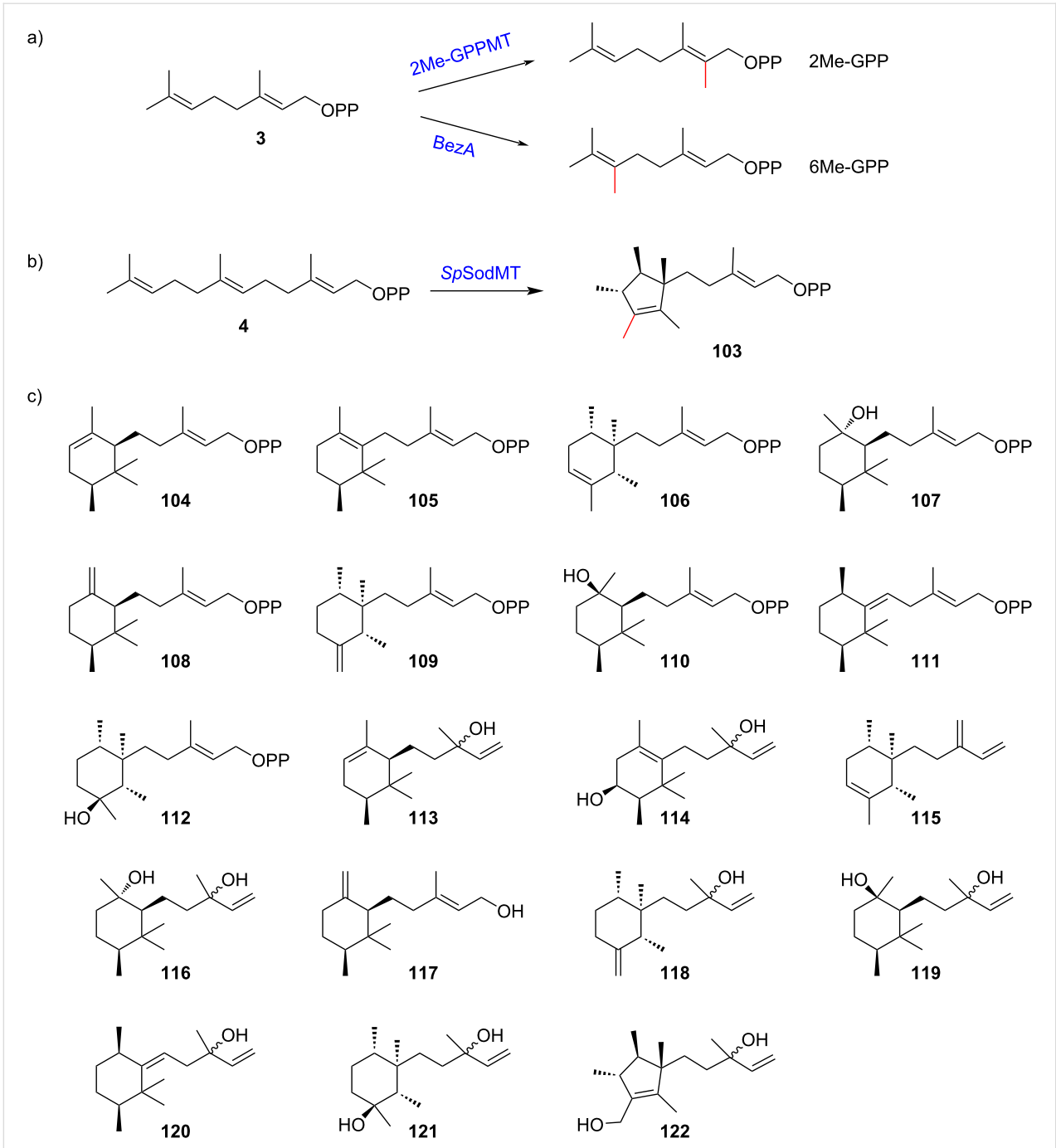



**Figure 8:** The structure of noncanonical prenyl substrates generated by *C*-methyltransferases and variants. a) 2Me-GPP and 6Me-GPP are produced by GPP *C*-methyltransferases. b) Compound **103** is produced by FPP *C*-methyltransferases. c) Compounds **104**–**112** are new $C_{16}$ building blocks synthesized by *Sp*SodMT variants, and **113**–**122** are selected typical products yielded by terpentetriene synthase and kolavelool synthase with further modifications by a cytochrome P450 CYP720B1.

of 2-methylenebornane synthase from *Pseudomonas fluorescenes* was altered using a semi-rational engineering approach [56].

In contrast to GPP methylation, modification of FPP is catalyzed by the *C*-methyltransferase *Sp*SodMT. In 2018, the biosynthesis of an unusual homosesquiterpene, sodorifen (**102**, Figure 1), from *Serratia plymuthica* 4RX13 was elucidated [57]. The in vitro and in vivo results revealed that a SAM-dependent-*C*-methyltransferase catalyzed methylation and cyclization reactions to form pre-sodorifen (**103**, Figure 8b), which was subsequently converted to **102** by TS [57]. Key residues

lining the catalytic cavity of *Sp*SodMT, Q57, F58, N219, V273, and L302, were found to affect product outcomes, and mutagenesis of these residues resulted in new $C_{16}$-prenyl substrates [58]. Selcted *Sp*SodMT variants provided ten $C_{16}$ building blocks, including **103**, plymuthenyl diphosphate (**104**), thorvaldsenyl diphosphate (**105**), weylandtenyl diphosphate (**106**), blixenyl diphosphate (**107**), kimlarsenyl diphosphate (**108**), serratinyl diphosphate (**109**), jacobsenyl diphosphate (**110**), hammershoyl diphosphate (**111**), and ancheryl diphosphate (**112**) (Figure 8c). Subsequently, terpentetriene synthase from *Kitasatospora griseola* (Cyc2) and kolavelool synthase from *Herpetosiphon aurantiacus* (HaKS) were identified as capable of converting these building blocks into $C_{16}$ scaffolds, whereas other selective sesqui-TSs failed to transform **104**–**112**. After further modification with cytochrome P450 CYP720B1 in the yeast host, 28 noncanonical terpenoids were generated, 10 of them are shown in Figure 8c (**113**–**122**) [58]. Notably, the widespread biosynthesis of $C_{16}$ terpenoids was reported in a recent study in which biosynthetic gene clusters for $C_{16}$ terpenoids were identified and grouped into four types according to the number of MTs and TSs in the gene cluster [59]. A subset of methyltransferase genes was functionally characterized using engineered yeast, which has an enhanced supply of **4** (strain AM109) [60] and the main product of these enzymes was compound **103** [59]. Subsequently, 35 selective TSs in these gene clusters were characterized using a yeast chassis, and 47 noncanonical terpenoids were produced with 13 of them being elucidated (**123**–**135**, Figure 9), which enabled further studies on their functions and prompted the discovery of new types of terpenoids [59].

In addition to methylation of prenyl substrates, the lepidopteran mevalonate (LMVA) pathway can also form six-carbon homoisopentenyl pyrophosphate (HIPP, **136**). The lepidopteran (butterflies and moths) pathway incorporates 3-ketovaleryl-CoA (**137**), instead of acetoacetyl-CoA (**138**), into the normal MVA pathway to generate compound **136** (Figure 10a). By heterologously expressing the LMVA pathway with a propionyl-CoA ligase and TSs, researchers have successfully produced several novel terpenes containing 16 carbon atoms in *E. coli* albeit at low titers [61]. To increase $C_{16}$ terpene titers, a later study redirected the 3-ketovaleryl-CoA formation step from the previous thiolase-dependent LMVA pathway to a β-oxidation LMVA pathway, and combined it with a promiscuous phosphatase, NudB, to produce $C_6$- and $C_7$-isoprenol. Notably, the final products, 3-ethyl-3-buten-1-ol (**139**) and 3-propyl-3-buten-1-ol (**140**), have potent fuel properties, highlighting the potential of this strategy for producing isoprenol analogs as next-generation biofuels (Figure 10b) [62].

## Conclusion

During terpenoid biosynthesis, most TSs have strict substrate selectivity; nevertheless, some promiscuous TSs accept multiple prenyl substrates and produce various products. In nature, the biosynthesis of prenyl substrates may have subcellular locations, and the available types of prenyl substrates are limited, especially for noncanonical substrates in living cells. Therefore, the potential of TSs to generate terpenoids has been underestimated. With the development of synthetic biology technologies, an efficient precursor-providing chassis was constructed.

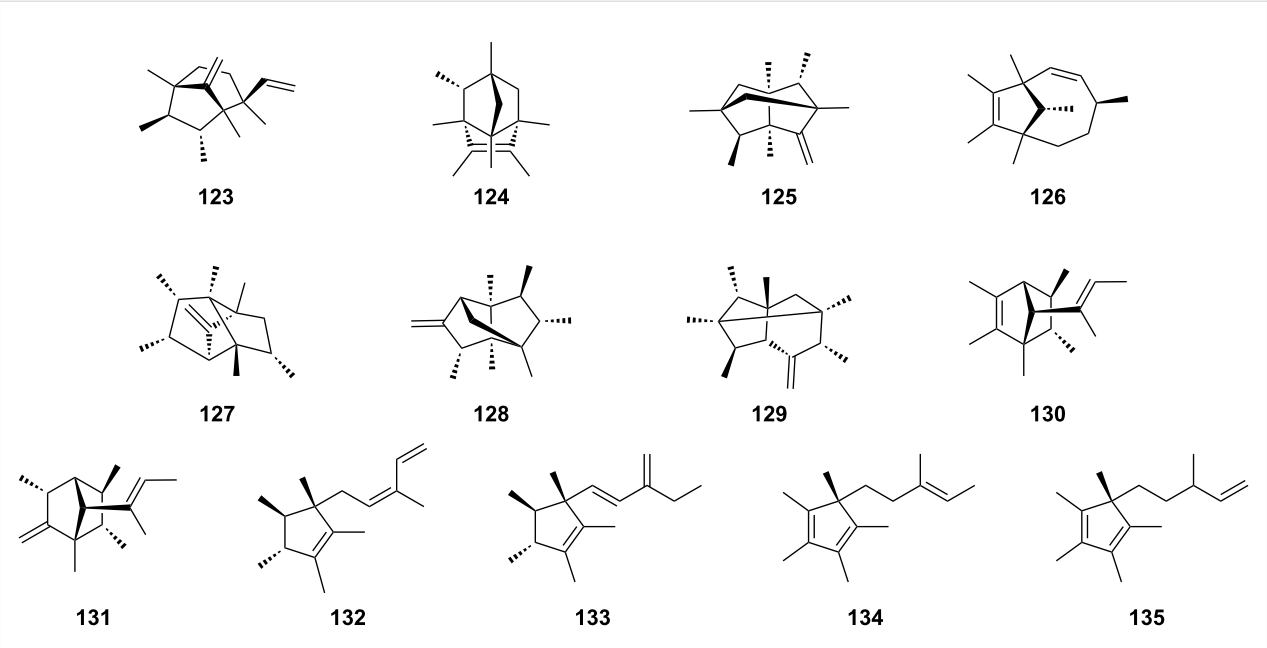



**Figure 9:** Structures of $C_{16}$ terpenes identified via genome mining of $C_{16}$ biosynthetic gene clusters from bacteria.

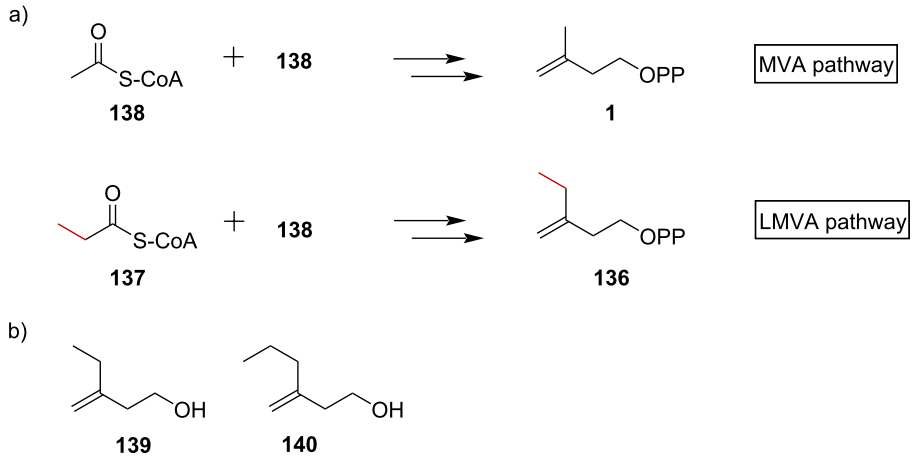



**Figure 10:** a) Precursors and final products of the MVA pathway and LMVA pathway. b) The structure of $C_6$- and $C_7$-isoprenols **139** and **140**.

Together with the accumulation of genome sequencing data, we systematically evaluated the function of TSs and discovered new terpenoids via genome mining. Nevertheless, for drug development, the accumulated terpene skeletons still require further functionalization, which requires additional genome-mining efforts for the discovery of tailored enzymes. Researchers have successfully expanded the chemical space of terpenoid biosynthesis using noncanonical prenyl substrates, which were synthesized using chemical approaches or via biosynthetic pathways. Many new terpenoids have been derived from chemically prepared prenyl analogs for decades, and until recently, the conversion of enzymatically modified noncanonical substrates has been utilized. New building blocks with irregular carbon numbers broaden the diversity of terpenoid structures. However, more systematic studies on noncanonical terpenoids are needed to study their biological activities.

## Acknowledgements

The authors thank Dr. Yuchen Zhang from the University of Tokyo for critical reading and helpful discussion.

## Funding

This work was financially supported by the National Key R&D Program of China (2023YFA0916200), the Fundamental Research Funds for the Central Universities (2042024kf1039), the Young Elite Scientists Sponsorship Program by CAST (2022QNRC001), the National Natural Science Foundation of China, and TaiKang Center for Life and Medical Sciences.

## Author Contributions

Min Li: investigation; writing – original draft. Hui Tao: conceptualization; investigation; supervision; validation; writing – review & editing.

## ORCID® iDs

Hui Tao - https://orcid.org/0000-0001-7978-5063

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

1. Dictionary of Natural Products 31.2. https://dnp.chemnetbase.com/ (accessed Dec 11, 2023).
2. Gershenzon, J.; Dudareva, N. *Nat. Chem. Biol.* **2007,** *3,* 408–414. doi:10.1038/nchembio.2007.5
3. Tetali, S. D. *Planta* **2019,** *249,* 1–8. doi:10.1007/s00425-018-3056-x
4. Sacchettini, J. C.; Poulter, C. D. *Science* **1997,** *277,* 1788–1789. doi:10.1126/science.277.5333.1788
5. Christianson, D. W. *Chem. Rev.* **2006,** *106,* 3412–3442. doi:10.1021/cr050286w
6. Christianson, D. W. *Science* **2007,** *316,* 60–61. doi:10.1126/science.1141630
7. Degenhardt, J.; Köllner, T. G.; Gershenzon, J. *Phytochemistry* **2009,** *70,* 1621–1637. doi:10.1016/j.phytochem.2009.07.030
8. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
9. Zhang, Q.; Tiefenbacher, K. *Nat. Chem.* **2015,** *7,* 197–202. doi:10.1038/nchem.2181
10. Eaton, S. A.; Christianson, D. W. *Biochemistry* **2023,** *62,* 2301–2313. doi:10.1021/acs.biochem.3c00247
11. Rudolf, J. D.; Chang, C.-Y. *Nat. Prod. Rep.* **2020,** *37,* 425–463. doi:10.1039/c9np00051h
12. Yang, Y.; Zhang, Y.; Zhang, S.; Chen, Q.; Ma, K.; Bao, L.; Tao, Y.; Yin, W.; Wang, G.; Liu, H. *J. Nat. Prod.* **2018,** *81,* 1089–1092. doi:10.1021/acs.jnatprod.7b01033
13. Harms, V.; Kirschning, A.; Dickschat, J. S. *Nat. Prod. Rep.* **2020,** *37,* 1080–1097. doi:10.1039/c9np00055k
14. Pazouki, L.; Niinemets, Ü. *Front. Plant Sci.* **2016,** *7,* 1019. doi:10.3389/fpls.2016.01019

15. Karunanithi, P. S.; Zerbe, P. *Front. Plant Sci.* **2019,** *10,* 1166. doi:10.3389/fpls.2019.01166
16. Sharkey, T. D.; Gray, D. W.; Pell, H. K.; Breneman, S. R.; Topper, L. *Evolution* **2013,** *67,* 1026–1040. doi:10.1111/evo.12013
17. Ashaari, N. S.; Ab. Rahim, M. H.; Sabri, S.; Lai, K. S.; Song, A. A.-L.; Abdul Rahim, R.; Wan Abdullah, W. M. A. N.; Ong Abdullah, J. *PLoS One* **2020,** *15,* e0235416. doi:10.1371/journal.pone.0235416
18. Johnson, S. R.; Bhat, W. W.; Sadre, R.; Miller, G. P.; Garcia, A. S.; Hamberger, B. *New Phytol.* **2019,** *223,* 323–335. doi:10.1111/nph.15778
19. Chen, Y.-G.; Li, D.-S.; Ling, Y.; Liu, Y.-C.; Zuo, Z.-L.; Gan, L.-S.; Luo, S.-H.; Hua, J.; Chen, D.-Y.; Xu, F.; Li, M.; Guo, K.; Liu, Y.; Gershenzon, J.; Li, S.-H. *Angew. Chem., Int. Ed.* **2021,** *60,* 25468–25476. doi:10.1002/anie.202110842
20. Karunanithi, P. S.; Berrios, D. I.; Wang, S.; Davis, J.; Shen, T.; Fiehn, O.; Maloof, J. N.; Zerbe, P. *Plant J.* **2020,** *103,* 781–800. doi:10.1111/tpj.14771
21. Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. *Plant J.* **2011,** *66,* 212–229. doi:10.1111/j.1365-313x.2011.04520.x
22. Dhandapani, S.; Tjhang, J. G.; Jang, I.-C. *Metab. Eng.* **2020,** *61,* 397–405. doi:10.1016/j.ymben.2020.08.002
23. Nieuwenhuizen, N. J.; Wang, M. Y.; Matich, A. J.; Green, S. A.; Chen, X.; Yauk, Y.-K.; Beuning, L. L.; Nagegowda, D. A.; Dudareva, N.; Atkinson, R. G. *J. Exp. Bot.* **2009,** *60,* 3203–3219. doi:10.1093/jxb/erp162
24. Nagegowda, D. A.; Gutensohn, M.; Wilkerson, C. G.; Dudareva, N. *Plant J.* **2008,** *55,* 224–239. doi:10.1111/j.1365-313x.2008.03496.x
25. Quin, M. B.; Flynn, C. M.; Schmidt-Dannert, C. *Nat. Prod. Rep.* **2014,** *31,* 1449–1473. doi:10.1039/c4np00075g
26. Schmidt-Dannert, C. Biosynthesis of Terpenoid Natural Products in Fungi. In *Biotechnology of Isoprenoids;* Schrader, J.; Bohlmann, J., Eds.; Advances in Biochemical Engineering/Biotechnology, Vol. 148; Springer: Cham, Switzerland, 2015; pp 19–61. doi:10.1007/10_2014_283
27. González-Hernández, R. A.; Valdez-Cruz, N. A.; Macías-Rubalcava, M. L.; Trujillo-Roldán, M. A. *World J. Microbiol. Biotechnol.* **2023,** *39,* 194. doi:10.1007/s11274-023-03635-y
28. Bian, G.; Han, Y.; Hou, A.; Yuan, Y.; Liu, X.; Deng, Z.; Liu, T. *Metab. Eng.* **2017,** *42,* 1–8. doi:10.1016/j.ymben.2017.04.006
29. Qin, B.; Matsuda, Y.; Mori, T.; Okada, M.; Quan, Z.; Mitsuhashi, T.; Wakimoto, T.; Abe, I. *Angew. Chem., Int. Ed.* **2016,** *55,* 1658–1661. doi:10.1002/anie.201509263
30. Cane, D. E.; Ikeda, H. *Acc. Chem. Res.* **2012,** *45,* 463–472. doi:10.1021/ar200198d
31. Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112,* 857–862. doi:10.1073/pnas.1422108112
32. Dickschat, J. S. *Nat. Prod. Rep.* **2016,** *33,* 87–110. doi:10.1039/c5np00102a
33. Reddy, G. K.; Leferink, N. G. H.; Umemura, M.; Ahmed, S. T.; Breitling, R.; Scrutton, N. S.; Takano, E. *PLoS One* **2020,** *15,* e0232220. doi:10.1371/journal.pone.0232220
34. Li, Z.; Jiang, Y.; Zhang, X.; Chang, Y.; Li, S.; Zhang, X.; Zheng, S.; Geng, C.; Men, P.; Ma, L.; Yang, Y.; Gao, Z.; Tang, Y.-J.; Li, S. *ACS Catal.* **2020,** *10,* 5846–5851. doi:10.1021/acscatal.0c01575
35. Sato, T.; Yamaga, H.; Kashima, S.; Murata, Y.; Shinada, T.; Nakano, C.; Hoshino, T. *ChemBioChem* **2013,** *14,* 822–825. doi:10.1002/cbic.201300035
36. Ueda, D.; Yamaga, H.; Murakami, M.; Totsuka, Y.; Shinada, T.; Sato, T. *ChemBioChem* **2015,** *16,* 1371–1377. doi:10.1002/cbic.201500138
37. Fujihashi, M.; Sato, T.; Tanaka, Y.; Yamamoto, D.; Nishi, T.; Ueda, D.; Murakami, M.; Yasuno, Y.; Sekihara, A.; Fuku, K.; Shinada, T.; Miki, K. *Chem. Sci.* **2018,** *9,* 3754–3758. doi:10.1039/c8sc00289d
38. Köksal, M.; Chou, W. K. W.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2013,** *52,* 5247–5255. doi:10.1021/bi400797c
39. Vattekkatte, A.; Gatto, N.; Schulze, E.; Brandt, W.; Boland, W. *Org. Biomol. Chem.* **2015,** *13,* 4776–4784. doi:10.1039/c5ob00506j
40. Hou, A.; Dickschat, J. S. *Chem. – Eur. J.* **2021,** *27,* 15644–15649. doi:10.1002/chem.202103049
41. Li, H.; Dickschat, J. S. *Org. Chem. Front.* **2022,** *9,* 795–801. doi:10.1039/d1qo01707a
42. Oberhauser, C.; Harms, V.; Seidel, K.; Schröder, B.; Ekramzadeh, K.; Beutel, S.; Winkler, S.; Lauterbach, L.; Dickschat, J. S.; Kirschning, A. *Angew. Chem., Int. Ed.* **2018,** *57,* 11802–11806. doi:10.1002/anie.201805526
43. Harms, V.; Schröder, B.; Oberhauser, C.; Tran, C. D.; Winkler, S.; Dräger, G.; Kirschning, A. *Org. Lett.* **2020,** *22,* 4360–4365. doi:10.1021/acs.orglett.0c01345
44. Tran, C. D.; Dräger, G.; Struwe, H. F.; Siedenberg, L.; Vasisth, S.; Grunenberg, J.; Kirschning, A. *Org. Biomol. Chem.* **2022,** *20,* 7833–7839. doi:10.1039/d2ob01279k
45. Harms, V.; Ravkina, V.; Kirschning, A. *Org. Lett.* **2021,** *23,* 3162–3166. doi:10.1021/acs.orglett.1c00882
46. Weigel, B.; Ludwig, J.; Weber, R. A.; Ludwig, S.; Lennicke, C.; Schrank, P.; Davari, M. D.; Nagia, M.; Wessjohann, L. A. *ChemBioChem* **2022,** *23,* e202200211. doi:10.1002/cbic.202200211
47. Johnson, L. A.; Dunbabin, A.; Benton, J. C. R.; Mart, R. J.; Allemann, R. K. *Angew. Chem., Int. Ed.* **2020,** *59,* 8486–8490. doi:10.1002/anie.202001744
48. Li, H.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2022,** *61,* e202211054. doi:10.1002/anie.202211054
49. Gu, B.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2023,** *62,* e202307006. doi:10.1002/anie.202307006
50. Piechulla, B.; Zhang, C.; Eisenschmidt-Bönn, D.; Chen, F.; Magnus, N. *FEMS Microbiol. Rev.* **2021,** *45,* fuab024. doi:10.1093/femsre/fuab024
51. Drummond, L.; Haque, P. J.; Gu, B.; Jung, J. S.; Schewe, H.; Dickschat, J. S.; Buchhaupt, M. *ChemBioChem* **2022,** *23,* e202200091. doi:10.1002/cbic.202200091
52. Drummond, L.; Kschowak, M. J.; Breitenbach, J.; Wolff, H.; Shi, Y.-M.; Schrader, J.; Bode, H. B.; Sandmann, G.; Buchhaupt, M. *ACS Synth. Biol.* **2019,** *8,* 1303–1313. doi:10.1021/acssynbio.8b00525
53. Kschowak, M. J.; Wortmann, H.; Dickschat, J. S.; Schrader, J.; Buchhaupt, M. *PLoS One* **2018,** *13,* e0196082. doi:10.1371/journal.pone.0196082
54. Ignea, C.; Pontini, M.; Motawia, M. S.; Maffei, M. E.; Makris, A. M.; Kampranis, S. C. *Nat. Chem. Biol.* **2018,** *14,* 1090–1098. doi:10.1038/s41589-018-0166-5
55. Tsutsumi, H.; Moriwaki, Y.; Terada, T.; Shimizu, K.; Shin-ya, K.; Katsuyama, Y.; Ohnishi, Y. *Angew. Chem., Int. Ed.* **2022,** *61,* e202111217. doi:10.1002/anie.202111217
56. Kschowak, M. J.; Maier, F.; Wortmann, H.; Buchhaupt, M. *ACS Synth. Biol.* **2020,** *9,* 981–986. doi:10.1021/acssynbio.9b00432
57. von Reuss, S.; Domik, D.; Lemfack, M. C.; Magnus, N.; Kai, M.; Weise, T.; Piechulla, B. *J. Am. Chem. Soc.* **2018,** *140,* 11855–11862. doi:10.1021/jacs.8b08510

58. Ignea, C.; Raadam, M. H.; Koutsaviti, A.; Zhao, Y.; Duan, Y.-T.; Harizani, M.; Miettinen, K.; Georgantea, P.; Rosenfeldt, M.; Viejo-Ledesma, S. E.; Petersen, M. A.; Bredie, W. L. P.; Staerk, D.; Roussis, V.; Ioannou, E.; Kampranis, S. C. *Nat. Commun.* **2022,** *13,* 5188. doi:10.1038/s41467-022-32921-w
59. Duan, Y.-T.; Koutsaviti, A.; Harizani, M.; Ignea, C.; Roussis, V.; Zhao, Y.; Ioannou, E.; Kampranis, S. C. *Nat. Chem. Biol.* **2023,** *19,* 1532–1539. doi:10.1038/s41589-023-01445-9
60. Ignea, C.; Trikka, F. A.; Kourtzelis, I.; Argiriou, A.; Kanellis, A. K.; Kampranis, S. C.; Makris, A. M. *Microb. Cell Fact.* **2012,** *11,* 162. doi:10.1186/1475-2859-11-162
61. Eiben, C. B.; de Rond, T.; Bloszies, C.; Gin, J.; Chiniquy, J.; Baidoo, E. E. K.; Petzold, C. J.; Hillson, N. J.; Fiehn, O.; Keasling, J. D. *ACS Synth. Biol.* **2019,** *8,* 2238–2247. doi:10.1021/acssynbio.9b00230
62. Pang, B.; Li, J.; Eiben, C. B.; Oksen, E.; Barcelos, C.; Chen, R.; Englund, E.; Sundstrom, E.; Keasling, J. D. *Metab. Eng.* **2021,** *68,* 210–219. doi:10.1016/j.ymben.2021.10.007

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.86

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# A Diels–Alder probe for discovery of natural products containing furan moieties

Alyssa S. Eggly[1], Namuunzul Otgontseren[1], Carson B. Roberts[1], Amir Y. Alwali[1], Haylie E. Hennigan[1] and Elizabeth I. Parkinson[*1,2]

Full Research Paper

Open Access

Address:
[1]Department of Chemistry, Purdue University, West Lafayette, Indiana 47906, United States and [2]Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, Indiana 47906, United States

Email:
Elizabeth I. Parkinson[*] - eparkins@purdue.edu

* Corresponding author

Keywords:
Diels–Alder reaction; furans; methylenomycin furan hormones; natural products; reactivity-based probes

*Beilstein J. Org. Chem.* **2024,** *20,* 1001–1010.
https://doi.org/10.3762/bjoc.20.88

Received: 01 January 2024
Accepted: 15 April 2024
Published: 02 May 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf

© 2024 Eggly et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Natural products (NPs) are fantastic sources of inspiration for novel pharmaceuticals, oftentimes showing unique bioactivity against interesting targets. Specifically, NPs containing furan moieties show activity against a variety of diseases including fungal infections, and cancers. However, it is challenging to discover and isolate these small molecules from cell supernatant. The work described herein showcases the development of a molecular probe that can covalently modify furan moieties via a [4 + 2] Diels–Alder cycloaddition, making them easily identifiable on liquid chromatography–mass spectrometry (LC–MS). The molecular probe, which undergoes this reaction with a variety of furans, was designed with both a UV-tag and a mass tag to enable easy identification. The probe has been tested with a variety of purified furans, including natural products, methylenomycin furan (MMF) hormones, and MMF derivatives. Moreover, the molecular probe has been tested in crude supernatants of various *Streptomyces* strains and enables identification of MMFs.

## Introduction

To date, natural products and their derivatives have served as the foundation of many pharmaceuticals used today. These compounds have activity against various diseases such as cancer, bacterial infections, and fungal infections, as well as other indications [1]. Specifically, natural products containing furan moieties (Figure 1A) are of interest as they have shown promising biological activities.

One well-studied example is the antifungal flufuran, which was isolated from *Aspergillus flavus* [2]. Another example is plakorsin D, an anticancer polyketide natural product which was isolated from a sea sponge, *Plakortis simplex* [3,4]. Wortmannin, which also has a furan, was isolated from *Penicillium wortmannin* and shows anticancer activity via potent inhibition of phosphoinositide 3-kinase [5]. These are just a handful of the

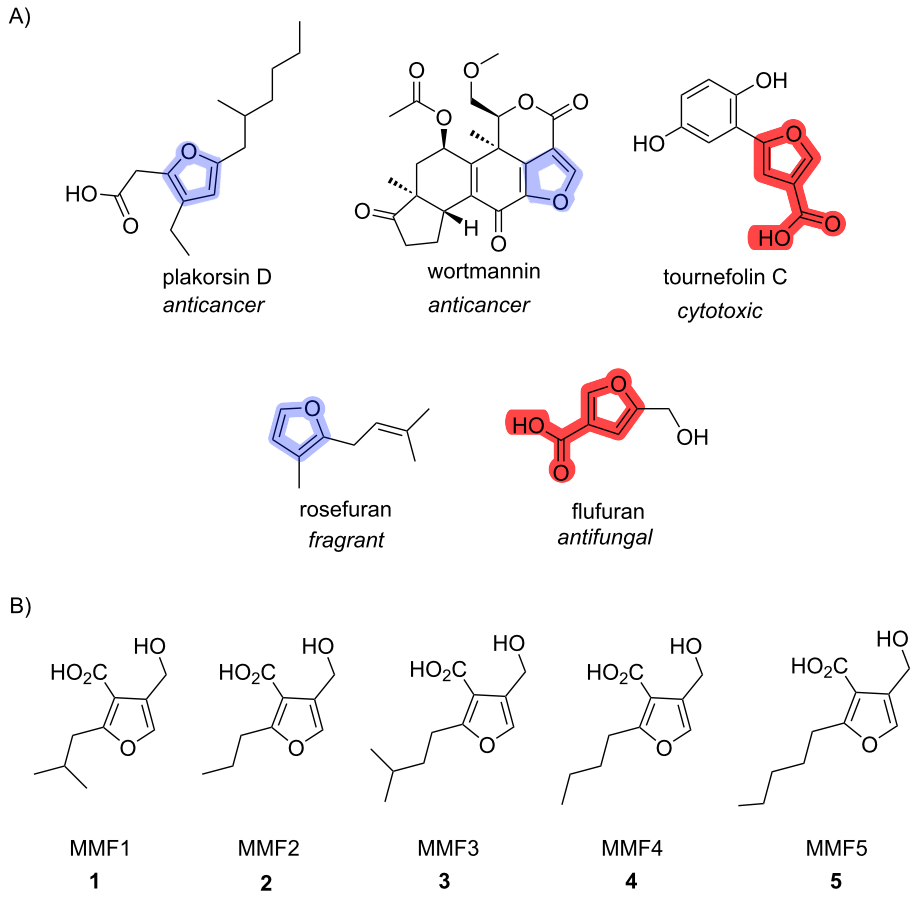



**Figure 1:** A) Bioactive natural products containing a furan ring (blue) or 3-furoic acid moiety (red): plakorsin D, wortmannin, tournefolin C, rosefuran, and flufuran and their respective biological activities. B) MMF hormones isolated to date.

bioactive natural products containing furans that exist. Natural products with furan moieties can also be signaling hormones. Methylenomycin furans (MMFs) are naturally occurring secondary metabolites that are produced by *Streptomyces coelicolor*, a soil dwelling bacterium. These molecules are important as they induce the production of the antibiotic methylenomycin A. Specifically, an MMF binds to the TetR family transcriptional repressor (TFTR) resulting in the complex being released from the DNA ultimately allowing for gene transcription and production of enzymatic machinery necessary for the biosynthesis of methylenomycin A [6,7]. To date, there have been five natural MMFs isolated and characterized, with all of the MMFs being isolated from *Streptomyces coelicolor* W75 (Figure 1B). These compounds are 2,3,4-trisubstituted furans, and all five contain a carboxylic acid at the three position and a hydroxymethyl group at the four position. They differ solely based on which alkyl chain is present at the two position [7]. Based on genomic data, it is hypothesized that other secondary metabolites similar to the MMFs exist in other strains of actinobacteria. However, it is time-intensive and challenging to isolate and elucidate novel natural products. Hormones such as MMFs are particularly challenging to isolate because they are often very potent (often active at picomolar concentrations) and thus produced at low levels. To date, MMFs have only been isolated from *Streptomyces coelicolor* W75, with 1 mg of MMF-1 being isolated from approximately 1 L of agar plates (1 mg/L titer or 5 μM). However, this initial isolation was enabled by heterologous expression of their biosynthetic genes from *Streptomyces coelicolor* to produce higher titers of the MMF molecules compared to the native producer [7].

Given the challenges of isolating furans generally and the MMFs specifically, we have developed a chemical probe that is capable of covalently binding to natural products containing furan moieties. Molecular probes are molecules that covalently bind to a compound of interest in order to make them easier to identify from the complex milieu of the cell supernatant. There are two main types of molecular probes based on their detec-

tion methods: imaging and UV–vis (Figure 2A). An imaging probe contains a mass spectrometry (MS) tag, oftentimes a halogen with a distinct isotopic ratio such as chlorine or bromine. A UV–vis probe contains a UV-tag, such as an aromatic ring, to make the compound UV-active. Previously developed probes containing both MS and UV–vis tags have proven to be successful in cell supernatant in identifying a number of functional groups including epoxides, enones, citrulline, conjugated alkenes, and others [8-16]. One probe has been developed capable of undergoing a Diels–Alder reaction with a peptide containing a terminal furan [17]. However, to date, no one has developed probes that have been demonstrated to work well with natural products containing furans. Therefore, we have developed a probe that can undergo a [4 + 2] Diels–Alder cycloaddition to identify furan moieties within complex cell supernatants.

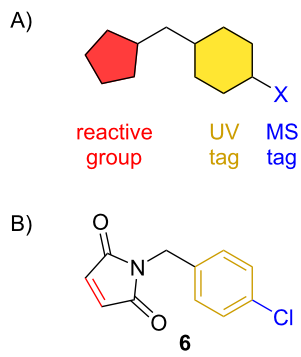



**Figure 2:** A) A diagram of an example of a molecular probe. B) The maleimide-based probe (**6**) for furans.

## Results and Discussion

When designing the reactive group of the furan probe **6**, we chose to focus on a well-known reaction with furans, the Diels–Alder cyclization. While the Diels–Alder cyclization strategy has been previously used in probe development [9,16,17], no one has reported it previously with furans. This is likely because utilizing a furan ring as the diene source can be quite challenging. Furans are aromatic molecules, resulting in these five-membered ring systems being relatively unreactive. Specifically, breaking their aromaticity, which is required for the Diels–Alder reaction to occur, is unfavorable and results in Diels–Alder reactions with furans often being easily reversible [18]. Fortunately, although difficult, it is not impossible to perform a Diels–Alder reaction with furans. While the frontier molecular orbital theory can be applied to this cycloaddition reaction, the thermodynamics must also be taken into consideration. Most Diels–Alder reactions with furans as the diene follow normal-electron-demand with the diene HOMO reacting with the dienophile LUMO due to the oxygen atom in the ring destabilizing the diene HOMO. Thus, when it comes to substituents on a furan ring, it is better to place electron-withdrawing groups in the three and/or five positions of the furan to aid in the natural flow of the electrons [19]. Due to this reactivity, we expected that a Diels–Alder probe would work well for the MMFs as well as other natural products containing electron-withdrawing groups at the three position such as flufuran (**23**). With this in mind, we designed the maleimide-based molecular probe (Figure 2B). Compound **6** was easily synthesized over three steps by adapting an existing route [20] and is inexpensive ($1.80/g of starting material) to make in large quantities. It also contains a chlorine group as its MS tag, which can be easily identified via its 3:1 isotopic ratio when run on a UPLC-MS, and a phenyl ring, which allows it to be UV visible at 214 nm and 254 nm.

Once the probe was synthesized in good yields, the optimization of the Diels–Alder reaction was undertaken. The Diels–Alder reaction (Figure 3A) was optimized utilizing 3-furoic acid (**14**) as it fits the requirement of having an electron-withdrawing group in the three position and was an inexpensive, commercially available option. Initial conditions were based on previously reported work with similar structures [21,22]. We planned to perform the reaction using water as the solvent, but due to solubility challenges with the probe, this was

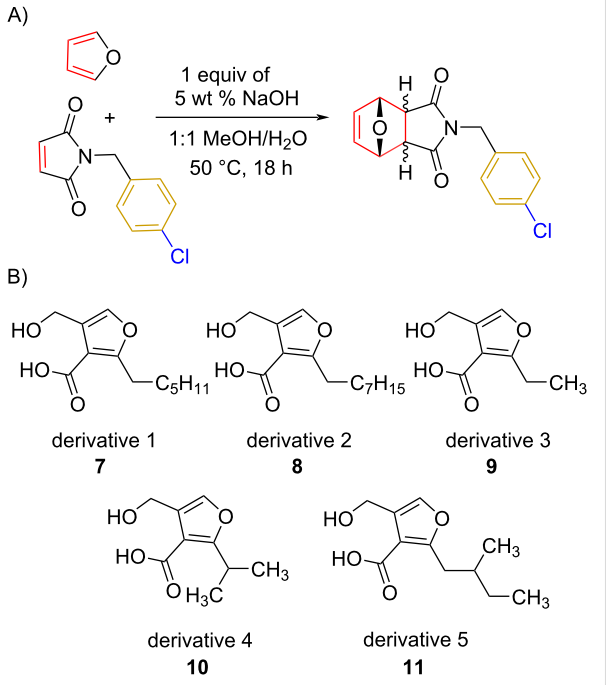



**Figure 3:** A) General Diels–Alder reaction scheme. B) MMF derivatives explored in this reaction.

not a viable option. After completing a solvent screen, it was determined that a ratio of 1:1 MeOH/$H_2O$ should be used as this solubilized the chemical probe most efficiently. Additionally, others have previously noted that polar solvents tend to benefit Diels–Alder reactions between furoic acids and maleimides [19,23]. Next, we looked at how different additives would affect the conversion. Others have observed that sodium hydroxide or salts can have beneficial effects on a Diels–Alder reaction with furans [18,19]. For this reason, we explored both sodium hydroxide and sodium chloride as additives. When 1 equivalent of a 5 percent solution of sodium hydroxide was used, the reaction proceeded well with 78% conversion when monitored by both MS and NMR spectroscopy. However, when sodium chloride was used instead of sodium hydroxide, no conversion was observed. Temperature was also taken into consideration; temperatures above 55 °C were not considered because we aim to use these same conditions in cell extracts to probe for small molecules that may be unstable or volatile at higher temperatures.

To test the chemical probe and determine the substrate scope of the reaction, we tested a small library of furans (Table 1). After each reaction was run, the samples were diluted and run on a UPLC-MS. Cycloadduct formation was verified by the presence of the 3:1 isotopic ratio of the chlorine on the probe and the correct *m*/*z* ratio. The conversion was calculated based on the remaining starting material, which was quantified using a standard curve. In general, very minimal to no side products were observed via MS or NMR, thus validating the use of starting material for quantification.

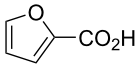
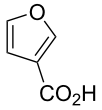
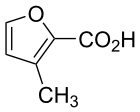
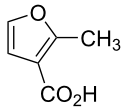
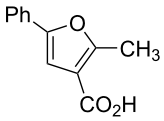
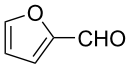
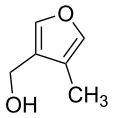
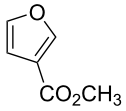
From this data, a few generalized conclusions can be drawn. Overall, the substrates with higher conversions were **14** and **16**, which is in line with our hypothesis that having an electron-withdrawing group in the three position will push the reaction forward. With a less electron-withdrawing group, such as **20**, the reaction still proceeds but we see less conversion. When we place an electron-withdrawing group in the two-position, as in **13** or **15**, conversion decreases compared to **14** or **16**, respectively.

In addition to the substrates that did undergo the Diels–Alder [4 + 2] cycloaddition, some substrates were unable to react with our chemical probe. For example, **18** did covalently attach to the probe, but there is no **18** leftover to quantify the amount of conversion. Additionally, the mass spectrum for **18** was more complicated, showing many more side products compared to other tested substrates (see Figure S7 in Supporting Information File 1). We hypothesize this is due to **18** undergoing a reaction in aqueous solutions to generate a geminal diol in place of an aldehyde, as has been previously described [24]. Thiophene (**22**) also did not produce any product, which may be due to the differences in the electronics of the sulfur or the lack of electron-withdrawing groups on the molecule. The phenyl-substituted compound **17** was tested but ultimately showed no conversion as well. Although it does have the carboxylic acid in the three position, the phenyl group is most likely too electron-donating, as well as potentially being too sterically bulky. Finally, pyrrole **21** was tested and ultimately resulted in an inconclusive amount of conversion. Compound **21** does not ionize well on our LC–MS and thus we are unable to quantify the amount of product formed. However, we can see the mass of the product in the total ion chromatogram as well as in the UV trace (Figure S10 in Supporting Information File 1).

Once the initial substrate scope was completed, we moved forward with testing the chemical probe with known natural products. To do this, the five naturally occurring MMFs (**1–5**) were synthesized by following a previous route [22]. In addition to the naturally occurring MMFs, a library of derivatives (Figure 3B) was also tested to determine if changes in the alkyl chain would affect conversion. These derivatives contained shorter side chains, longer side chains, and different branching patterns than the naturally occurring MMFs. Given that other well-studied actinobacteria hormones such as the γ-butyrolactones have varying alkyl chains [25-28], we expect that derivatives of the MMFs with alternative alkyl chains likely exist in other bacterial strains. Lastly, the probe was tested with flufuran (**23**) to show that it can also identify a natural product that is not an MMF.

The maleimide probe was tested with the natural MMF molecules and the library of MMF derivatives in the same fashion as previously described for the synthetic molecules (Table 2). The results from these experiments are very promising (Figure 4), as almost all of the naturally occurring MMF molecules had essentially quantitative conversion at the 18 h timepoint. Compound **1** was the lowest converting of all of the natural MMFS, but it still had an impressive 94% conversion. The five MMF derivatives that were tested were also able to undergo the desired Diels–Alder reaction. Compound **9**, which contains the ethyl group as its side chain showed quantitative conversion. However, when the side chain was increased to an eight-carbon chain shown with **8**, conversion dropped to an average of 86%. This is likely because the bulkier side chain sterically hinders the molecules from aligning in the optimal orientation for the HOMO and LUMO to interact. Similar high yields were observed with **10** and **11**. The maleimide probe also worked well with flufuran (**23**), having a 92% conversion. The slightly lower conversion is most likely due to the loss of the hydroxy group at the four position and is not surprising as it aligns with the results of 3-furoic acid (**14**).

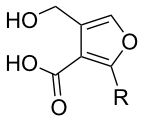
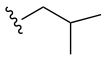
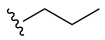
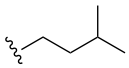
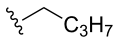
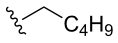
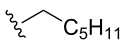
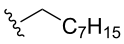
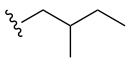
**Table 1:** Conversion of various synthetic substrates when reacted with the maleimide probe.[a]

| Substrate | Structure | Average % conversion[b] ± standard deviation |
|---|---|---|
| furan[c] (**12**) |  | N/A[d] |
| 2-furoic acid (**13**) |  | 21 ± 1 |
| 3-furoic acid (**14**) |  | 79 ± 8 |
| 3-methyl-2-furoic acid (**15**) |  | 81 ± 6 |
| 2-methyl-3-furoic acid (**16**) |  | 87 ± 8 |
| 2-methyl-5-phenyl-3-furoic acid (**17**) |  | no conversion |
| furfural (**18**) |  | inconclusive |
| (4-methylfuran-3-yl)methanol (**19**) |  | 7 ± 7 |
| methyl 3-furan carboxylate (**20**) |  | 57 ± 8 |
| 1*H*-pyrrole (**21**) |  | inconclusive |
| thiophene (**22**) |  | no conversion |

[a]Representative LC–MS traces can be found in Figures S1–S12 in Supporting Information File 1. [b]The average conversion is calculated based on the amount of starting material remaining in the crude reaction. Reactions run in triplicate. [c]Furan is too volatile to react at 50 °C. [d]Not available.

After demonstrating the probe could undergo the Diels–Alder reaction with natural products containing furan moieties, it was tested with crude cell supernatants. Experiments were first performed to determine the limit of detection of the cycloadduct in the cell extracts when run on a UPLC-MS (Figure S23 in Supporting Information File 1). Specifically, a cell supernatant from a strain that does not produce MMFs (*Streptomyces coelicolor* M145) was spiked with varying amounts of compound **14**. The sample was then split, with half being reacted with the maleimide probe and the other half being exposed to the same conditions just lacking the probe. The samples were then used to determine the limit of detection for both the starting material **14**, as well as the cycloadduct product in the complete cell supernatant. The reactions were run at various concentrations of compound **14** ranging from 71 mM to 17 µM. Gratifyingly, the cycloadduct has 260-fold increase in sensitivity compared to **14**

**Table 2:** Conversion of various natural product substrates when reacted with the maleimide probe.[a]

| Substrate | Average % conversion[b] ± standard deviation |  R = |
|---|---|---|
| MMF1 (**1**) | 94 ± 1 |  |
| MMF2 (**2**) | 98 ± 1 |  |
| MMF3 (**3**) | 98 ± 2 |  |
| MMF4 (**4**) | 98 ± 2 |  |
| MMF5 (**5**) | 99 ± 1 |  |
| MMF derivative 1 (**7**) | 99 ± 1 |  |
| MMF derivative 2 (**8**) | 86 ± 4 |  |
| MMF derivative 3 (**9**) | 98 ± 1 |  |
| MMF derivative 4 (**10**) | 99 ± 7 |  |
| MMF derivative 5 (**11**) | 96 ± 1 |  |
| flufuran (**23**) | 93 ± 2[c] | N/A[d] |

[a]Representative traces can be found in Figures S13–S23 in Supporting Information File 1. [b]The average conversion is calculated based on the amount of starting material remaining in the crude reaction. The average is based on three replicates. [c]The average is based on five replicates. [d]Not applicable.

(69 μM and 18 mM, respectively), suggesting that the probe will be highly useful for detecting lower concentration furans. The maleimide probe was then tested for its ability to detect the naturally produced MMF compounds (Figure 5). Two strains of *Streptomyces coelicolor* were tested: *S. coelicolor* W75, a strain known to produce the MMF molecules was examined, and *S. coelicolor* M145, a control that does not contain the biosynthetic machinery necessary to produce MMFs but is otherwise genetically identical to *S. coelicolor* W75. Controls containing just the cell supernatant, as well as reactions of the cell supernatant with the maleimide probe, were run on the UPLC-MS after reacting for 18 h. As expected, samples from *S. coelicolor* M145 did not show any MMFs present in either the control sample or the probe-treated sample, even after concentrating the sample 25-fold. Compound **5** and its cycloadduct were both observed in *Streptomyces* W75. Gratifyingly, the cycloadduct was much more easily visualized, with an ion intensity that was approximately 15-fold higher than that of **5** alone. This demonstrates that our chemical probe is able to label naturally occurring furan natural products within cell supernatants and that it can be utilized as a tool to more efficiently uncover furan-containing natural products. However, the ability of the probe to detect furans in cell extracts will depend both on their structures, as described in Table 1, and their titers. The cycloadduct is observed in *S. coelicolor* W75 treated with the probe at both 5-time (≈25 μM) and 25-time (≈125 μM) concentrated samples. This is consistent with our limit of detection of **14**. However, *S. coelicolor* W75 is a heterologous expression strain that

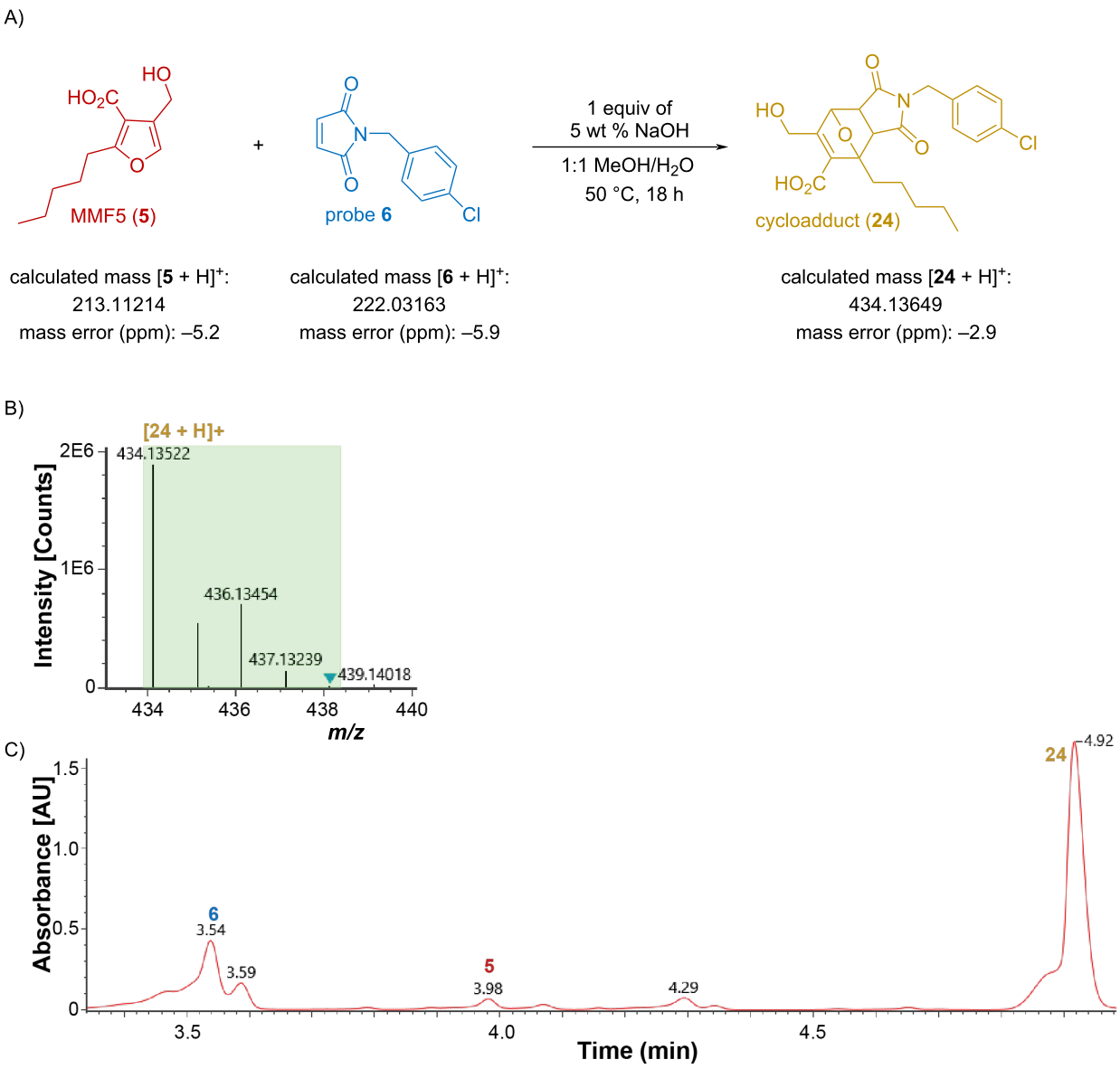



**Figure 4:** A. Reaction scheme. B. Representative mass spectra of MMF5 cycloadduct **24** with 3:1 isotopic ratio present. C. Representative UV trace of the reaction products; probe **6**, MMF5 (**5**), and MMF5 cycloadduct **24**. All reactions were performed a minimum of three times with similar results being observed in all three replicates.

produces higher levels of MMFs than expected in native strains. We unfortunately do not know the titers of MMFs in native producing strains. However, if titers are similar to that of γ-butyrolactones and butenolides quorum sensing molecules, we would expect them to be between 80 and 2.5 nM [29-31]. Given these very low concentrations, cultures would need to be greatly concentrated (≈1000 times) to observe the molecules in their native producers. While this is not ideal, it is still significantly better than the concentration needed without probe (≈260,000 times).

## Conclusion

MMFs are important signaling hormones that induce the production of methylenomycin A. Bioinformatics data suggests that there are more natural products containing furan moieties that are waiting to be discovered. However, the discovery of natural products can be tedious because they are often produced at very low levels. We have developed a chemical probe that can alleviate some of these challenges. Our chemical probe contains a UV-tag and MS-tag for easy identification utilizing LC–MS. Additionally, it is capable of covalently attaching to a

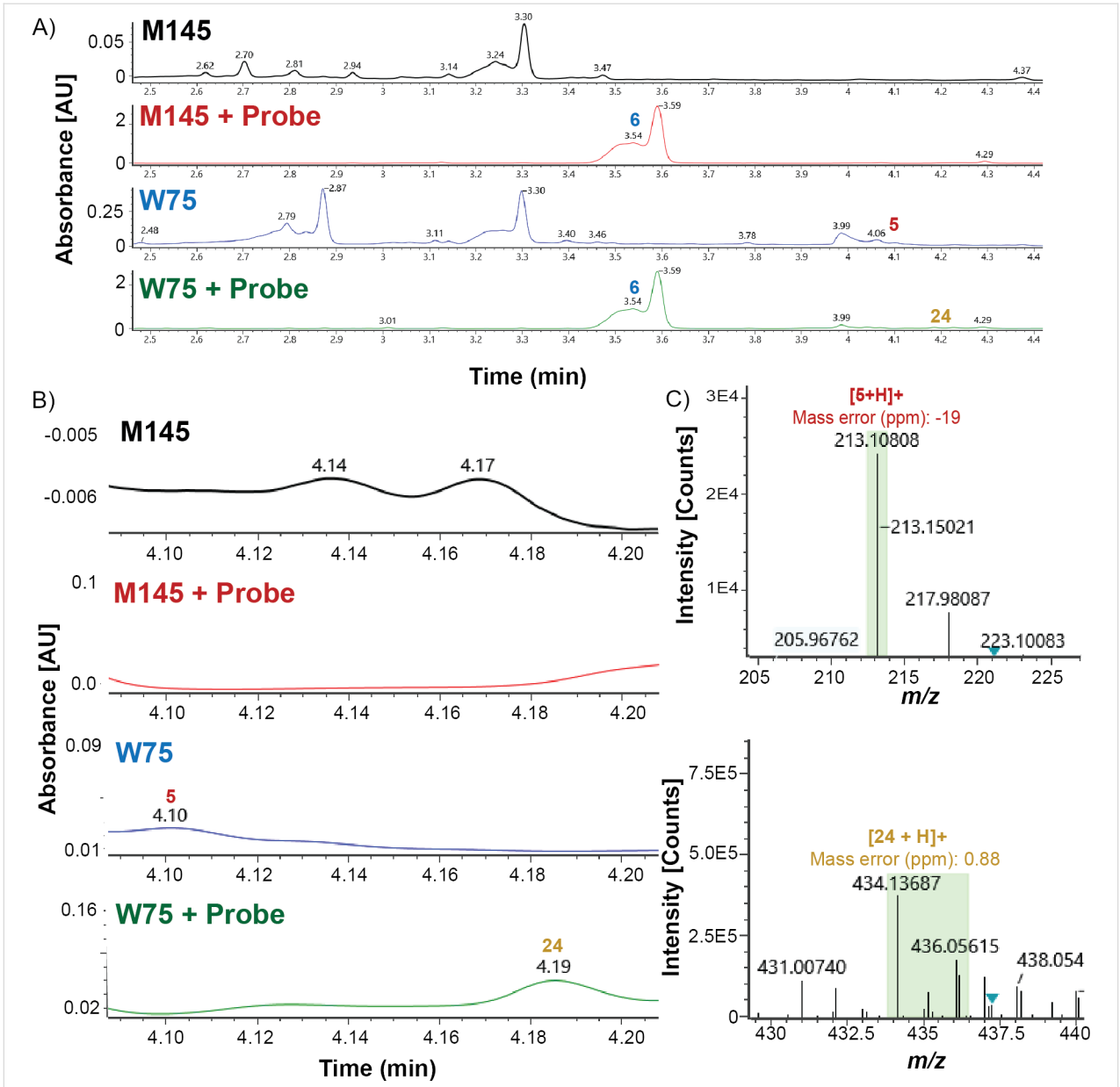



**Figure 5:** A. The UV trace (254 nm) for M145 control, M145 + probe **6**, W75 control, and W75 + probe **6**. M145 control shows no MMFs and M145 + probe shows only the maleimide probe **6**. W75 control shows MMF5 (**5**) present. W75 + probe shows both MMF5 cycloadduct (**24**) and excess maleimide probe **6**. B. Zoom in of the UV trace (254 nm) for M145 control, M145 + probe, W75 control, and W75 + probe. M145 control shows no MMFs and M145 + probe shows nothing significant in the zoomed in region. W75 control shows MMF5 (**5**) and W75 shows the MMF5 cycloadduct **24**. C. Mass spectra of W75 control and W75 + probe showing mass of MMF5 (**5**) and mass of the MMF5 cycloadduct **24**, respectfully. All samples shown in this figure were run at 25-times the extracted concentration. Representative data from three experiments is shown.

variety of furan rings via a [4 + 2] Diels–Alder cycloaddition in relatively mild reaction conditions. It reacts with naturally occurring MMFs as well as their derivatives and other natural products such as flufuran. Finally, the maleimide probe has proven that it can identify furan-containing natural products in complex cell supernatants, further demonstrating its potential utility in the discovery of novel furan-containing natural products. Future work will focus on utilizing the probe in additional strains to probe for novel natural products similar to the MMFs.

## Supporting Information

### Supporting Information File 1

Experimental materials and methods, supplemental figures, and supplemental spectra.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-88-S1.pdf]

## Acknowledgements

This work was supported in part by the Research Instrumentation Center in the Department of Chemistry at Purdue University. We thank G. Challis (University of Warwick), L. Alkhalaf (University of Warwick), and C. Corre (University of Warwick) for providing us with the *Streptomyces* strains.

## Funding

This work was supported by a grant from the National Institutes of Health (1R35GM138002-01 to E.I.P.). The authors acknowledge the support from the Purdue Center for Cancer Research, NIH grant P30 CA023168.

## Author Contributions

Alyssa S. Eggly: conceptualization; data curation; formal analysis; investigation; methodology; visualization; writing – original draft; writing – review & editing. Namuunzul Otgontseren: investigation; validation; writing – review & editing. Carson B. Roberts: investigation; validation; writing – review & editing. Amir Y. Alwali: formal analysis; methodology. Haylie E. Hennigan: resources. Elizabeth I. Parkinson: conceptualization; data curation; formal analysis; funding acquisition; project administration; resources; supervision; visualization; writing – original draft; writing – review & editing.

## ORCID® iDs

Namuunzul Otgontseren - https://orcid.org/0009-0002-3266-5157
Elizabeth I. Parkinson - https://orcid.org/0000-0002-3665-5522

## Data Availability Statement

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

## References

1. Atanasov, A. G.; the International Natural Product Sciences Taskforce; Zotchev, S. B.; Dirsch, V. M.; Supuran, C. T. *Nat. Rev. Drug Discovery* **2021,** *20,* 200–216. doi:10.1038/s41573-020-00114-z
2. Masi, M.; Di Lecce, R.; Marsico, G.; Linaldeddu, B. T.; Maddau, L.; Superchi, S.; Evidente, A. *J. Nat. Prod.* **2021,** *84,* 2600–2605. doi:10.1021/acs.jnatprod.1c00365
3. Chen, V. Y.; Kwon, O. *Angew. Chem., Int. Ed.* **2021,** *60,* 8874–8881. doi:10.1002/anie.202015232
4. Zhang, J.; Tang, X.; Li, J.; Li, P.; de Voogd, N. J.; Ni, X.; Jin, X.; Yao, X.; Li, P.; Li, G. *J. Nat. Prod.* **2013,** *76,* 600–606. doi:10.1021/np300771p
5. Ng, S. S.; Tsao, M. S.; Nicklee, T.; Hedley, D. W. *Clin. Cancer Res.* **2001,** *7,* 3269–3275.
6. Zhou, S.; Bhukya, H.; Malet, N.; Harrison, P. J.; Rea, D.; Belousoff, M. J.; Venugopal, H.; Sydor, P. K.; Styles, K. M.; Song, L.; Cryle, M. J.; Alkhalaf, L. M.; Fülöp, V.; Challis, G. L.; Corre, C. *Nature* **2021,** *590,* 463–467. doi:10.1038/s41586-021-03195-x
7. Corre, C.; Song, L.; O'Rourke, S.; Chater, K. F.; Challis, G. L. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 17510–17515. doi:10.1073/pnas.0805530105
8. Harris, L. A.; Saint-Vincent, P. M. B.; Guo, X.; Hudson, G. A.; DiCaprio, A. J.; Zhu, L.; Mitchell, D. A. *ACS Chem. Biol.* **2020,** *15,* 3167–3175. doi:10.1021/acschembio.0c00685
9. Castro-Falcón, G.; Millán-Aguiñaga, N.; Roullier, C.; Jensen, P. R.; Hughes, C. C. *ACS Chem. Biol.* **2018,** *13,* 3097–3106. doi:10.1021/acschembio.8b00598
10. Castro-Falcón, G.; Hahn, D.; Reimer, D.; Hughes, C. C. *ACS Chem. Biol.* **2016,** *11,* 2328–2336. doi:10.1021/acschembio.5b00924
11. Odendaal, A. Y.; Trader, D. J.; Carlson, E. E. *Chem. Sci.* **2011,** *2,* 760. doi:10.1039/c0sc00620c
12. Carlson, E. E.; Cravatt, B. F. *Nat. Methods* **2007,** *4,* 429–435. doi:10.1038/nmeth1038
13. Cox, C. L.; Tietz, J. I.; Sokolowski, K.; Melby, J. O.; Doroghazi, J. R.; Mitchell, D. A. *ACS Chem. Biol.* **2014,** *9,* 2014–2022. doi:10.1021/cb500324n
14. Harris, L. A.; Mitchell, D. A. *Methods Enzymol.* **2022,** *665,* 177–208. doi:10.1016/bs.mie.2021.11.018
15. Back, D.; Shaffer, B. T.; Loper, J. E.; Philmus, B. *J. Nat. Prod.* **2022,** *85,* 105–114. doi:10.1021/acs.jnatprod.1c00798
16. Guo, X.; Zhang, J.; Li, X.; Xiao, E.; Lange, J. D.; Rienstra, C. M.; Burke, M. D.; Mitchell, D. A. *ACS Cent. Sci.* **2021,** *7,* 781–791. doi:10.1021/acscentsci.1c00148
17. Hoogewijs, K.; Buyst, D.; Winne, J. M.; Martins, J. C.; Madder, A. *Chem. Commun.* **2013,** *49,* 2927. doi:10.1039/c3cc40588e
18. Spicer, C. D.; Pashuck, E. T.; Stevens, M. M. *Chem. Rev.* **2018,** *118,* 7702–7743. doi:10.1021/acs.chemrev.8b00253
19. Cioc, R. C.; Smak, T. J.; Crockatt, M.; van der Waal, J. C.; Bruijnincx, P. C. A. *Green Chem.* **2021,** *23,* 5503–5510. doi:10.1039/d1gc01535d
20. Schweizer, E.; Hoffmann-Röder, A.; Olsen, J. A.; Seiler, P.; Obst-Sander, U.; Wagner, B.; Kansy, M.; Banner, D. W.; Diederich, F. *Org. Biomol. Chem.* **2006,** *4,* 2364–2375. doi:10.1039/b602585d
21. Galkin, K. I.; Sandulenko, I. V.; Polezhaev, A. V. *Processes* **2021,** *10,* 30. doi:10.3390/pr10010030
22. Cioc, R. C.; Crockatt, M.; van der Waal, J. C.; Bruijnincx, P. C. A. *Angew. Chem., Int. Ed.* **2022,** *61,* e202114720. doi:10.1002/anie.202114720
23. Kumar, A. *Chem. Rev.* **2001,** *101,* 1–20. doi:10.1021/cr990410+
24. Cioc, R. C.; Lutz, M.; Pidko, E. A.; Crockatt, M.; van der Waal, J. C.; Bruijnincx, P. C. A. *Green Chem.* **2021,** *23,* 367–373. doi:10.1039/d0gc03558k
25. Daniel-Ivad, M.; Pimentel-Elardo, S.; Nodwell, J. R. *Annu. Rev. Microbiol.* **2018,** *72,* 25–48. doi:10.1146/annurev-micro-022618-042458
26. Wilbanks, L. E.; Hennigan, H. E.; Martinez-Brokaw, C. D.; Lakkis, H.; Thormann, S.; Eggly, A. S.; Buechel, G.; Parkinson, E. I. *ACS Chem. Biol.* **2023,** *18,* 1624–1631. doi:10.1021/acschembio.3c00241
27. Hsiao, N.-H.; Nakayama, S.; Merlo, M. E.; de Vries, M.; Bunet, R.; Kitani, S.; Nihira, T.; Takano, E. *Chem. Biol.* **2009,** *16,* 951–960. doi:10.1016/j.chembiol.2009.08.010
28. Wang, W.; Zhang, J.; Liu, X.; Li, D.; Li, Y.; Tian, Y.; Tan, H. *J. Biol. Chem.* **2018,** *293,* 20029–20040. doi:10.1074/jbc.ra118.005667

29. Kitani, S.; Miyamoto, K. T.; Takamatsu, S.; Herawati, E.; Iguchi, H.; Nishitomi, K.; Uchida, M.; Nagamitsu, T.; Omura, S.; Ikeda, H.; Nihira, T. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108,* 16410–16415. doi:10.1073/pnas.1113908108
30. Takano, E.; Nihira, T.; Hara, Y.; Jones, J. J.; Gershater, C. J. L.; Yamada, Y.; Bibb, M. *J. Biol. Chem.* **2000,** *275,* 11010–11016. doi:10.1074/jbc.275.15.11010
31. Khokhlov, A. S. Problems of studies of specific cell autoregulators (on the example of substances produced by some Actinomycetes). In *Proceedings of the International Symposium on Frontiers of Bioorganic Chemistry and Molecular Biology,* Sept 25–Oct 2, 1978; Ananchenko, S. N., Ed.; Pergamon Press: Moscow and Tashkent, USSR, 1980; pp 201–210. doi:10.1016/b978-0-08-023967-5.50030-9

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.88

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Cofactor-independent C–C bond cleavage reactions catalyzed by the AlpJ family of oxygenases in atypical angucycline biosynthesis

Jinmin Gao[‡1,2], Liyuan Li[‡1], Shijie Shen[1,2], Guomin Ai[1], Bin Wang[1,2], Fang Guo[1], Tongjian Yang[1,2], Hui Han[1], Zhengren Xu[3], Guohui Pan[*1,2,§] and Keqiang Fan[*1,¶]

Full Research Paper

Open Access

Address:
[1]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, No. 1 Beichen West Road, Beijing 100101, China, [2]University of Chinese Academy of Sciences, No. 1 Yanqihu East Road, Beijing 101408, China and [3]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China

Email:
Guohui Pan[*] - panguohui@im.ac.cn; Keqiang Fan[*] - fankq@im.ac.cn

* Corresponding author ‡ Equal contributors
§ Tel: 86-10-64806056, Fax: 86-10-64806056
¶ Tel: 86-10-64807800, Fax: 86-10-64807800

Keywords:
angucyclines; aromatic polyketide; biosynthesis; cofactor-independent oxygenase; oxidative rearrangement reaction

*Beilstein J. Org. Chem.* **2024,** *20,* 1198–1206.
https://doi.org/10.3762/bjoc.20.102

Received: 28 December 2023
Accepted: 07 May 2024
Published: 23 May 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: T. Awakawa



© 2024 Gao et al.; licensee Beilstein-Institut.
License and terms: see end of document.

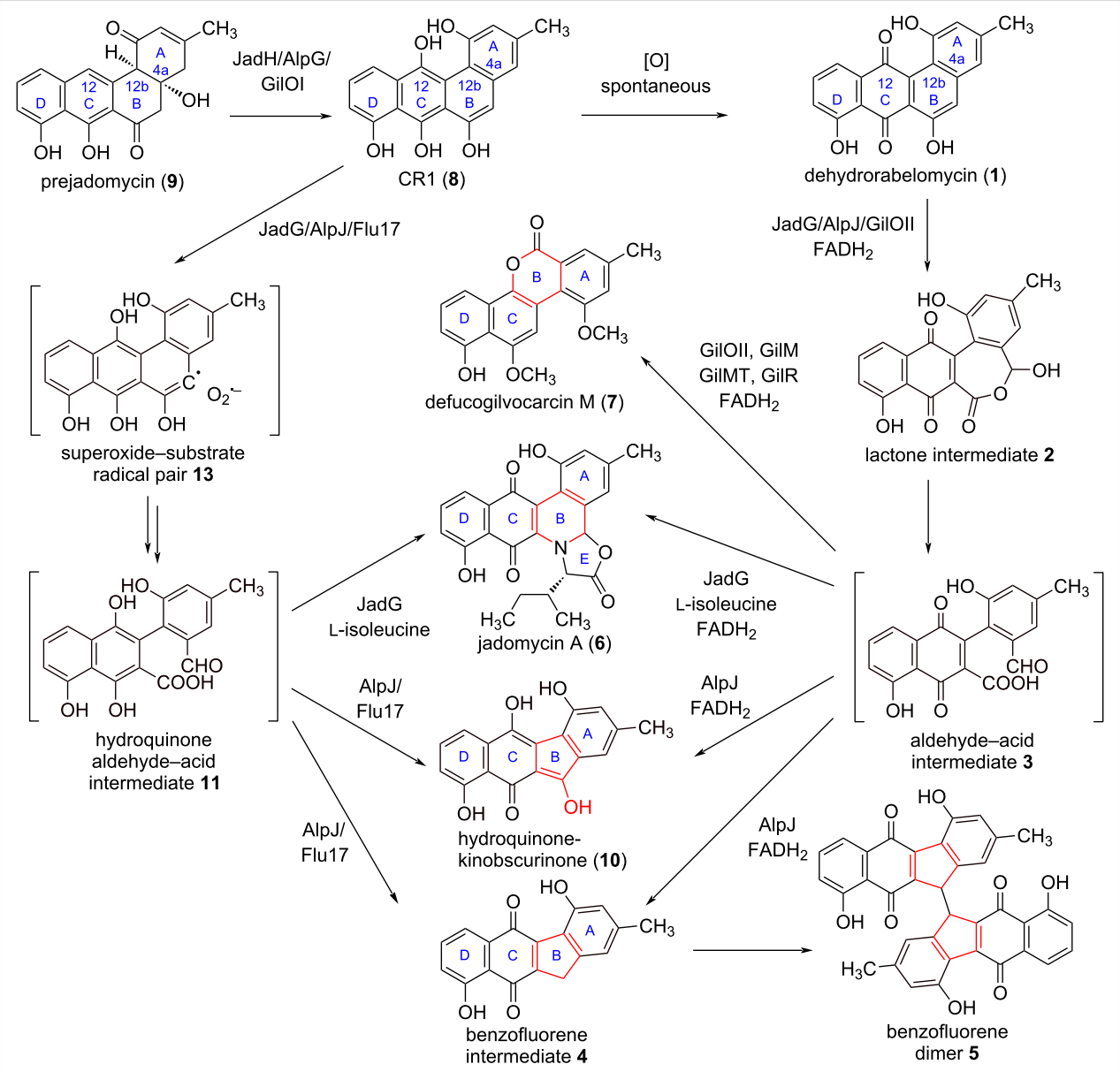
## Abstract

Biosynthesis of atypical angucyclines involves unique oxidative B-ring cleavage and rearrangement reactions, which are catalyzed by AlpJ-family oxygenases, including AlpJ, JadG, and GilOII. Prior investigations established the essential requirement for $FADH_2$/$FMNH_2$ as cofactors when utilizing the quinone intermediate dehydrorabelomycin as a substrate. In this study, we unveil a previously unrecognized facet of these enzymes as cofactor-independent oxygenases when employing the hydroquinone intermediate CR1 as a substrate. The enzymes autonomously drive oxidative ring cleavage and rearrangement reactions of CR1, yielding products identical to those observed in cofactor-dependent reactions of AlpJ-family oxygenases. Furthermore, the AlpJ- and JadG-catalyzed reactions of CR1 could be quenched by superoxide dismutase, supporting a catalytic mechanism wherein the substrate CR1 reductively activates molecular oxygen, generating a substrate radical and the superoxide anion $O_2^{\bullet-}$. Our findings illuminate a substrate-controlled catalytic mechanism of AlpJ-family oxygenases, expanding the realm of cofactor-independent oxygenases. Notably, AlpJ-family oxygenases stand as a pioneering example of enzymes capable of catalyzing oxidative reactions in either an $FADH_2$/$FMNH_2$-dependent or cofactor-independent manner.

## Introduction

The angucyclines represent a large class of natural products biosynthesized by type II polyketide synthases. Within this category, a distinctive subset of compounds, including jadomycin, gilvocarcin, kinamycin, fluostatin, and lomaiviticin, arises from typical angucycline intermediates via oxidative C–C bond cleavage and subsequent ring rearrangement reactions. These atypical angucyclines exhibit intriguing chemical structures and various biological activities [1,2]. Notably, kinamycins demonstrate both antibacterial and antitumor activities, while lomaiviticin emerges as a potent antitumor agent, displaying remarkable $IC_{50}$ values ranging from 0.007–72 nM against 25 cultured human cancer cell lines [3-7].

The pivotal B-ring cleavage and rearrangement in atypical angucycline biosynthesis are catalyzed by the AlpJ-family oxygenases, encompassing AlpJ, JadG, and GilOII. Biochemical analyses reveal the capability to transform the shared angucycline intermediate, dehydrorabelomycin (DHR, **1**), into products with unique chemical scaffolds (Scheme 1) [1,2,8-14]. Noteworthy is the prerequisite for $FADH_2$/$FMNH_2$, provided by *E. coli* Fre or flavin reductases in corresponding biosynthetic pathways, to facilitate these oxidation reactions. Experimental data support that **1** was firstly oxidized to a lactone intermediate **2** via Baeyer–Villiger oxidation, followed by hydrolysis to yield another crucial aldehyde/acid intermediate **3**




**Scheme 1:** Ring cleavage and ring rearrangement reactions in the biosynthesis of atypical angucyclines.

[11,15]. Commencing from **3**, diverse ring rearrangement reactions can occur, leading to the formation of distinct products.

In the AlpJ-catalyzed reaction, compound **3** undergoes ring contraction reactions, yielding a benzofluorene intermediate **4** and the dimer **5**, both featuring a kinamycin skeleton (Scheme 1) [11,12]. Recent investigations unveiled the catalytic activity of the O-methyltransferase-like protein AlpH, which catalyzes a unique SAM-independent coupling of L-glutamylhydrazine and **3** through a rare Mannich reaction. This results in the formation of prekinamycin harboring a diazo group [14]. Given the presence of AlpJ-like oxygenases and L-glutamylhydrazine biosynthetic enzymes in the gene clusters of lomaiviticin, nenestatin, and fluostatins [16-20], it is plausible that the corresponding biosynthesis involves analogous oxidative ring opening and rearrangement steps. On the other side, in JadG-catalyzed reactions, compound **3** reacts with L-isoleucine or other amino acids, leading to the formation of jadomycin A (**6**) or analogues, respectively [10]. In the gilvocarcin biosynthesis, the conversion of **3** to defucogilvocarcin M (**7**), featuring a benzo[*d*]naphtho[1,2-*b*]pyran-6-one backbone, necessitates the collaborative actions of GilOII and three additional tailoring enzymes, namely GilM, GilMT, and GilR [9].

To date, all available biochemical data consistently demonstrate the $FADH_2$/$FMNH_2$-dependent nature of AlpJ-family oxygenases when utilizing **1** as the substrate. Intriguingly, these enzymes exhibit notable similarities to cofactor-independent anthrone oxygenases, exemplified by TcmH and ActVA-Orf6, in both protein sequences and structures (see Figure S1, Supporting Information File 1) [21,22]. In this study, we reveal the previously undisclosed facet that AlpJ-family oxygenases can function as cofactor-independent oxygenases when the hydroquinone intermediate CR1 (**8**) serves as the substrate. In this context, the enzymes autonomously catalyze oxidative C–C bond cleavage, ring opening, and rearrangement reactions, yielding the respective products. Furthermore, the reactions of **8** catalyzed by JadG and AlpJ could be quenched by superoxide dismutase (SOD), supporting a catalytic mechanism involving the generation of a substrate radical and the superoxide anion $O_2^{\bullet-}$. This discovery sheds more light on the biosynthesis of atypical angucyclines and extends the repertoire of cofactor-independent oxygenases.

## Results and Discussion

### Cofactor-independent oxidative ring cleavage and rearrangement reactions catalyzed by AlpJ with CR1 (**8**) as substrate

Given the resemblance of AlpJ-family oxygenases to cofactor-independent anthrone oxygenases rather than to the conventional $FADH_2$-dependent counterparts, we aimed to investigate the potential for catalyzing oxidation reactions in a cofactor-independent manner. For cofactor-independent oxygenases such as TcmH and ActVA-Orf6, the substrates were able to reductively activate molecular oxygen, thereby enabling subsequent oxidation reactions [21-24]. However, in the documented reactions of AlpJ-family oxygenases, the substrate **1** appeared insufficient in providing the requisite reducing power, considering the quinone structure. Consequently, we redirected our focus towards alternative, more electron-rich substrates. In a preceding investigation, we identified that the bifunctional enzyme JadH proficiently converted prejadomycin (**9**) to **8**, a compound demonstrated to spontaneously oxidize to **1** under aerobic conditions [25]. This observation not only affirmed **8** as an intermediate in jadomycin biosynthesis but also suggested a role as a more electron-rich substrate with the potential for direct activation of molecular oxygen.

We first confirmed the generation of **8** in the biosynthetic pathway of kinamycin. The analysis of the *alp* gene cluster from *Streptomyces ambofaciens* revealed AlpG as the homolog of JadH. The N-terminal $His_6$-tagged construct of AlpG was expressed and purified to homogeneity in *E. coli* (Figure S2, Supporting Information File 1). The purified AlpG displayed a light yellow color, and the associated prosthetic group was identified as FAD through HPLC analysis of the denatured AlpG supernatant (Figure S3, Supporting Information File 1). As anticipated, AlpG exhibited the capacity to convert **9** to **8** and **1** in the presence of NADPH and FAD (Figure 1, trace b). This result underscores **8** as a biosynthetic intermediate in kinamycin biosynthesis, establishing AlpG as a bifunctional hydroxylase/dehydratase same to JadH (Figure 1, trace c) [25].

Next, employing the AlpG-catalyzed reaction of **9** as an in situ generation system of **8**, we introduced AlpJ (in the absence of Fre) to the reaction. Apart from the spontaneous formation of a limited quantity of **1** derived from **8**, we detected the generation of products, including the previously characterized benzofluorene intermediate **4** and benzofluorene dimer **5**, alongside a novel compound, designated as **10** (Figure 1, trace d). High-resolution mass spectrometry (HRMS) analysis of **10** ($[M - H]^-$ calcd for $C_{18}H_{12}O_5$, 307.0612; found, 307.0607, Figure S4, Supporting Information File 1) suggested a possible identity as hydroquinone–kinobscurinone [11]. However, attempts to elucidate the chemical structure of **10** were not successful due to the inherent instability. These findings indicate that AlpJ, in the absence of $FADH_2$, is capable of catalyzing the ring cleavage and contraction reactions of **8**. However, it cannot be definitively ruled out that AlpG in the reaction might contribute trace amounts of $FADH_2$ to AlpJ.

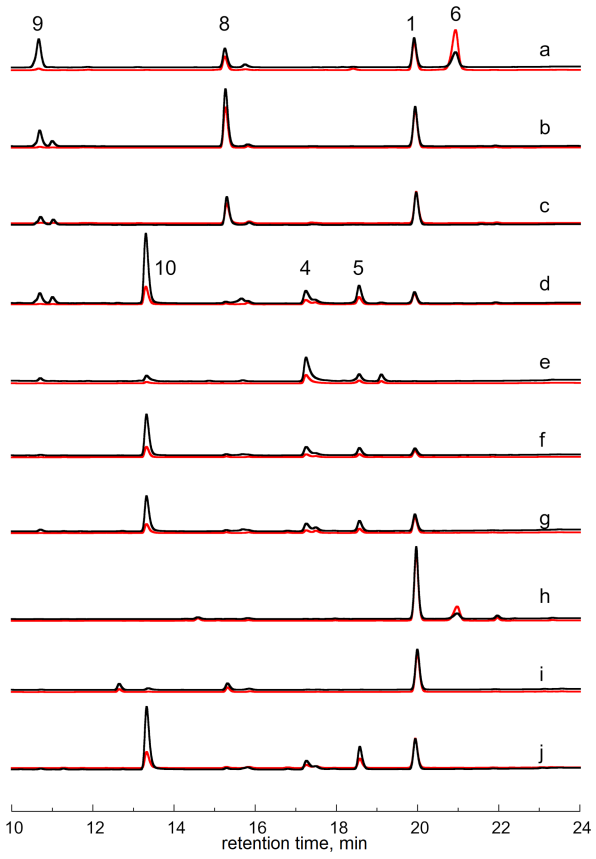



**Figure 1:** HPLC traces of reaction mixtures of AlpG, AlpJ, Flu17, and JadG. (a) standards of prejadomycin (**9**), CR1 (**8**), dehydrorabelomycin (**1**), and jadomycin A (**6**); (b) **9** + AlpG + NADPH + FAD; (c) **9** + JadH + NADPH + FAD; (d) **9** + AlpG + AlpJ + NADPH + FAD; (e) **1** + AlpJ + Fre + NADPH + FAD; (f) **8** + AlpJ; (g) **8** + Flu17; (h) **8** + JadG + L-isoleucine; (i) **8** + JadG; (j) **8** + AlpJ + L-isoleucine. UV absorptions at 266 nm (black line) and 313 nm (red line) are displayed for each reaction.

Subsequently, we carefully isolated **8** from a scaled-up reaction of AlpG with **9** as the substrate [25]. When **8** was incubated with AlpJ in the absence of any cofactors, the formation of **4**, **5**, and **10** was observed (Figure 1, trace f). Additionally, we conducted a reaction using **1** and AlpJ in the presence of $FADH_2$ generated by Fre. As anticipated, compounds **4** and **5** were detected, and notably, a small amount of **10** was also observed (Figure 1, trace e). As expected, in the negative control reactions, **8** was spontaneously oxidized to **1** without any enzyme added, and no **4**, **5**, or **10** was observed; **1** remained unchanged in the reaction mixture with Fre, NADH, and FAD but without AlpJ (Figure S5, Supporting Information File 1). These results demonstrate that AlpJ does not necessitate any cofactors for catalyzing the oxidation reactions of **8**, yielding products identical to those observed in the $FADH_2$-dependent reactions of AlpJ using **1** as the substrate.

## Cofactor-independent oxidative ring cleavage and rearrangement reactions of CR1 (**8**) catalyzed by homologous proteins of AlpJ

To extend the generality of our findings, we assessed the cofactor-independent catalytic capabilities of other homologous proteins of AlpJ. Previous investigations indicated that the biosynthesis of fluostatins involves analogous B-ring cleavage and contraction steps as observed in kinamycin biosynthesis [2,13,19,26]. A sequence analysis of a reported fluostatin biosynthetic gene cluster in reassembled environmental DNA identified Flu17 as the putative ring opening oxygenase [8]. The N-terminal $His_6$-tagged Flu17 was expressed and purified to homogeneity in *E. coli* BL21 (Figure S2, Supporting Information File 1). Significantly, Flu17 alone demonstrated the ability to convert **8** into products **4**, **5**, and **10** (Figure 1, trace g). Furthermore, incubating **8** with JadG and L-isoleucine yielded **6** (Figure 1, trace h). Consistent with expectations, omission of L-isoleucine from the reaction resulted in the absence of detectable **6** (Figure 1, trace i). The addition of L-isoleucine to the reaction of AlpJ and **8** did not alter the product profile (Figure 1, trace j).

Our findings demonstrate the capacity of AlpJ-family oxygenases to catalyze the ring cleavage and rearrangement reactions of **8** in a cofactor-independent manner. Moreover, these enzymes efficiently convert **8** into products featuring distinct chemical skeletons, aligning with outcomes observed when catalyzing the oxidation of **1** with $FADH_2$ as a cofactor [10-13]. The parallel behavior of Flu17 and AlpJ in the reaction with **8** provides additional support for the shared oxidative B-ring cleavage and contraction steps in the biosynthesis of fluostatins and kinamycins. Nevertheless, the mechanism by which a B-ring-contracted intermediate with a benzo[*b*]fluorene skeleton undergoes further conversion into fluostatins with a benzo[*a*]fluorene skeleton remains to be elucidated [27,28].

## Inhibition of cofactor-independent catalytic activities of AlpJ-family oxygenases by superoxide dismutase

There are many reported cofactor-independent oxygenases, including anthrone oxygenases such as ActVA-Orf6 and nogalamycin monooxygenase (NMO or SnoaB) [21,23,29-32], TnmJ and TnmK2 [24], 1-*H*-3-hydroxy-4-oxoquinaldine-2,4-dioxygenase (HOD) [33,34], and urate oxidase [35,36]. Investigations into NMO and HOD revealed the utilization of substrates to activate molecular oxygen, leading to the generation of a substrate radical and the superoxide anion $O_2^{\bullet-}$ [29,33,37]. The well-established superoxide trapping agent superoxide dismutase (SOD) has demonstrated significant inhibition of the NMO-catalyzed monooxygenation reaction [29].

To probe the potential involvement of a radical-mediated catalytic mechanism in AlpJ-family oxygenases, we investigated the impact of supplementing the reaction mixture of **8** and AlpJ, Flu17, or JadG with SOD. The results revealed that adding 5,000 U/mL SOD quenched the tested reactions, with no discernible products detected for AlpJ, Flu17, or JadG (Figure 2). Furthermore, a dose-dependent inhibition of the JadG-catalyzed reaction by SOD was observed (Figure S6, Supporting Information File 1). These findings support the presence of the superoxide anion $O_2^{\bullet-}$ and underscore the indispensable role in the reactions catalyzed by AlpJ-family oxygenases with **8**.

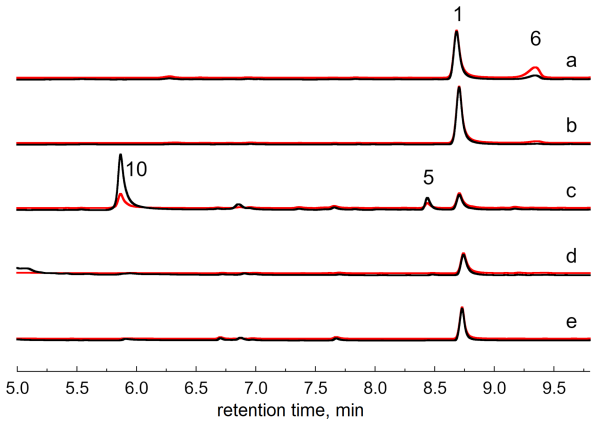



**Figure 2:** HPLC traces of reactions of JadG, AlpJ, or Flu17 quenched by SOD. (a) **8** + JadG + L-isoleucine; (b) **8** + JadG + L-isoleucine + SOD (5,000 U/mL); (c) **8** + AlpJ; (d) **8** + AlpJ + SOD (5,000 U/mL); (e) **8** + Flu17 + SOD (5,000 U/mL). UV absorption at 266 nm (black line) and 313 nm (red line) are displayed for each reaction.

## Proposed catalytic mechanism of cofactor-independent AlpJ-family oxygenases

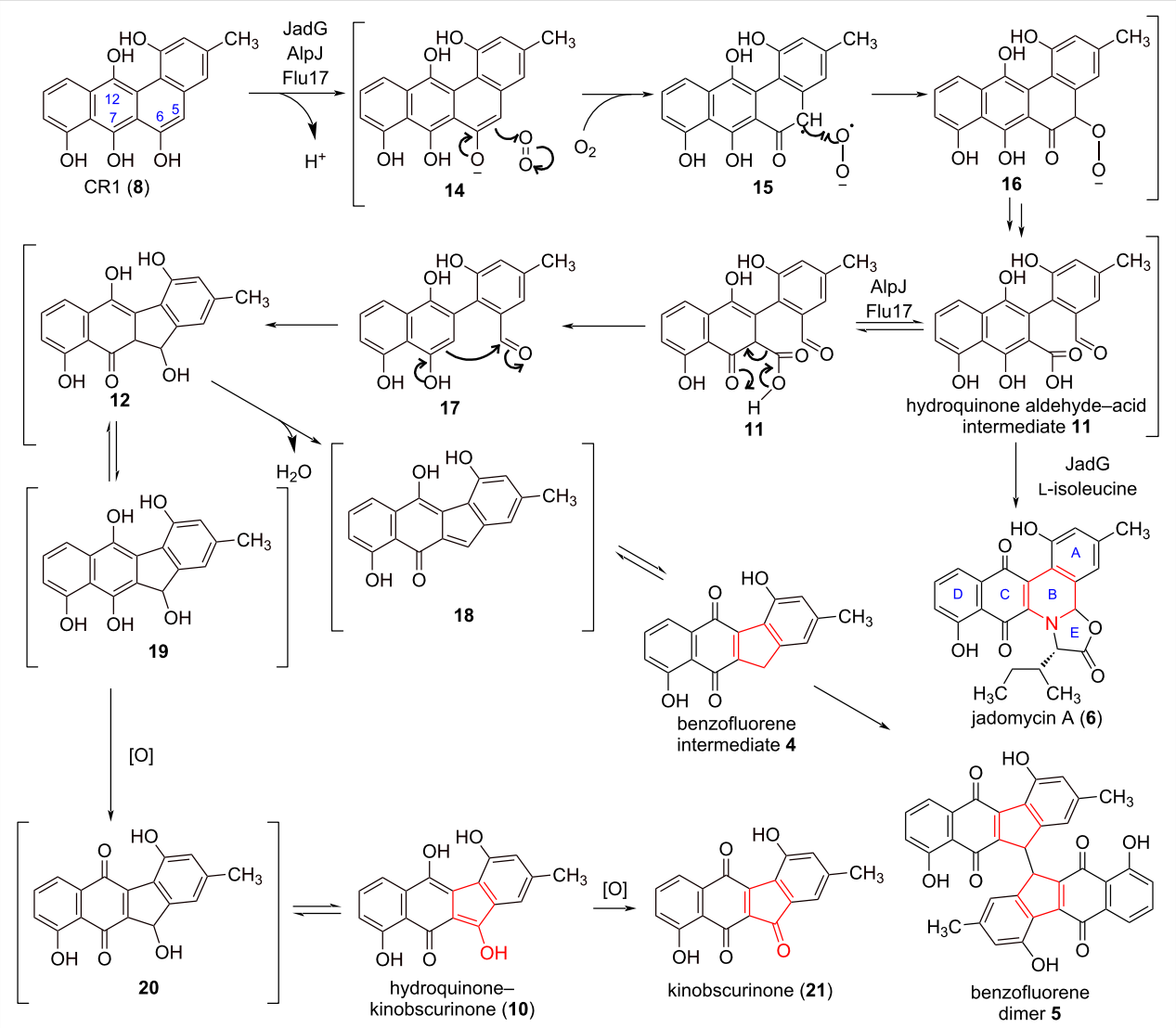
The class of cofactor-independent oxygenases presents a captivating group of enzymes that harness substrates to directly activate molecular oxygen. A recent investigation into the cofactor-independent enzyme NMO, employing dithranol as the substrate, unveiled the presence of a dithranyl radical and superoxide anion pair [29]. The production of the superoxide anion was captured using 1-hydroxy-3-methoxycarbonyl-2,2,5,5-tetramethylpyrrolidine (CMH) and quantified by detecting the stable $CM^{\bullet}$ radical through electron paramagnetic resonance. Based on previous studies of cofactor-independent oxygenases and our results, we propose a radical-mediated catalytic mechanism for the cofactor-independent AlpJ-family oxygenases (Scheme 2).

The catalytic cascade was initiated by deprotonation of **8** to form a substrate anion, which activated molecular oxygen to generate a substrate radical and the superoxide anion $O_2^{\bullet-}$. The ensuing radical rebound reaction led to the formation of a peroxide intermediate, which in turn underwent a ring opening reaction to generate the hydroquinone aldehyde–acid intermediate **11**. In JadG-catalyzed reactions, compound **11** participated in a reaction with L-isoleucine to yield **6**. In contrast, in AlpJ- or Flu17-catalyzed reactions, **11** underwent decarboxylation and an aldol reaction, giving rise to intermediate **12**. Subsequent dehydration of **12** led to the formation of product **4**, which through a presumed radical-mediated reaction could spontaneously convert to the dimer **5**. Additionally, compound **12** had the potential to be oxidized to hydroquinone–kinobscurinone **10**, a proposed precursor of kinobscurinone.

When employing **1** as the substrate, such cofactor-independent reactions failed to occur. It is plausible that **1** lacked the capability to reductively activate molecular oxygen owing to the quinone structure. In these instances, $FADH_2$/$FMNH_2$ were deemed necessary as cofactors for AlpJ-family oxygenases. These findings uncover a noteworthy substrate-controlled catalytic mechanism employed by AlpJ-family oxygenases, wherein the nature of the substrate dictates whether a cofactor-dependent or -independent reaction ensues. While it is not entirely unexpected that these enzymes could exhibit cofactor-independent oxygenase activity, given our prior discovery that the monomeric structure of AlpJ shares similarities with the dimeric cofactor-independent anthrone oxygenases, such as ActVA-Orf6 [13], the structural intricacies governing the accommodation of these two distinct mechanisms warrant further investigation.

The current findings indicate that both **8** and **1** can serve as direct substrates for AlpJ-family oxygenases. However, the dominant biosynthetic pathway in vivo remains unclear. Given that **8** has been experimentally validated as the authentic product of the JadH/AlpG-catalyzed reaction, it is recognized as a bona fide intermediate in angucycline biosynthesis [25].

On the other hand, the origin of **1**, spontaneously derived from **8** under aerobic conditions, raises uncertainties about the actual production in vivo. Furthermore, the significance of the C12 hydroxy group in the hydroquinone intermediate **11** is evident in subsequent reactions, including the formation of a hemiacetal intermediate in gilvocarcin biosynthesis [9,38,39] and the AlpH-catalyzed L-glutamylhydrazine adduct formation [14]. Notably, this C12 hydroxy group is absent in the quinone intermediate **3** of the cofactor-dependent pathway. Collectively, these observations suggest that the cofactor-independent pathway is likely the favored biosynthetic route in the angucycline biosynthesis process.




**Scheme 2:** Proposed catalytic mechanism of cofactor-independent AlpJ-family oxygenases.

## Conclusion

In this study, we showed that AlpJ-family oxygenases, traditionally recognized as $FADH_2$/$FMNH_2$-dependent enzymes, exhibit cofactor-independent oxygenase activity when **8** is employed as the substrate. AlpJ or the homologs Flu17 and JadG autonomously catalyze the oxidative ring cleavage and rearrangement reactions of **8**, yielding products identical to those observed in $FADH_2$-dependent reactions orchestrated by AlpJ-family oxygenases using **1** as the substrate. The results allowed us to propose a catalytic mechanism for cofactor-independent AlpJ-family oxygenases, wherein the substrate **8** reductively activates molecular oxygen, generating a substrate radical and the superoxide anion $O_2^{\bullet-}$. These findings unveil a captivating substrate-controlled catalytic mechanism employed by AlpJ-family oxygenases. Notably, AlpJ-family oxygenases stand as a pioneering example of enzymes capable of catalyzing oxidative reactions in either an $FADH_2$/$FMNH_2$-dependent or -independent manner. Up to date, many oxygenases from diverse evolutionary families, featuring varied protein folds and quaternary arrangements, have been identified as cofactor-independent oxygenases [24,29,37,40,41]. Our discoveries broaden the landscape of cofactor-independent oxygenases.

## Experimental

### Materials, culture conditions, and DNA manipulations

The substrates **8** and **1** were purified from the reaction mixture of AlpG and the cultures of *S. ambofaciens* ΔΔ*alpJW*, respectively, as previously described [11,25]. *E. coli* strains were grown in lysogeny broth (LB) [42]. Restriction enzymes, T4 DNA ligase, and KOD DNA polymerase were purchased from

New England BioLabs. FAD and NADPH were purchased from Sigma-Aldrich. DNA manipulation, competent cell preparation, and transformation were performed as described previously [42].

## Protein expression and purification

The *alpG* gene was amplified by PCR from *S. ambofaciens* ΔΔ*alpW* genomic DNA using primers 5'-ggaattccatATGGAAGGGACAACGGCGGACAC-3' and 5'-cccaagcttTCAGCGGGCGGGGCCGAA-3', digested with *Nde*I and *Hin*dIII, and cloned into pET28a to construct the expression plasmid of N-terminal $His_6$-tagged AlpG. The codon-optimized *flu17* gene was synthesized by Synbio Technologies, digested with *Nde*I and *Hin*dIII, and cloned into pET28a to afford the expression plasmid of N-terminal $His_6$-tagged Flu17. Both plasmids were confirmed by sequencing and transformed into *E. coli* BL21(DE3). Protein overproduction was induced with 0.1 mM isopropyl β-D-thiogalactopyranoside at 16 °C for 15 h. Cells were collected by centrifugation (8,000*g*, 15 min) at 4 °C. The pellets were resuspended in ice-cold 50 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS) buffer (pH 7.5), and the cells were disrupted by ultrasonication to obtain the cell extract. Cell debris was removed by centrifugation (14,000*g*, 15 min). The proteins were purified by Ni-NTA agarose chromatography, desalted, and concentrated by centrifugation (8,000*g*, 30 min) in 10 kDa ultrafiltration tubes (Centriplus YM series, Merck Millipore). Protein concentration was quantified by the Bradford method. AlpJ, JadG, JadH, and *E. coli* flavin reductase (Fre) were overproduced and purified by Ni-NTA agarose chromatography as described previously [10,11,43,44].

## In vitro enzymatic assay

All reactions were performed in 50 mM Tris-HCl buffer at pH 8.0. For conversion of **9** by AlpG or JadH, the reaction mixtures (50 µL) consisting of 10 µM FAD, 125 µM NADPH, 100 µM **9**, and 25 µM AlpG or JadH were incubated at 30 °C for 5 min and analyzed by HPLC directly. For conversion of **8** by AlpJ or Flu17, the reaction mixtures (50 µL) consisting of 200 µM **8** and 50 µM AlpJ or Flu17 were incubated at 30 °C for 5 min and analyzed by HPLC directly. For conversion of **8** by JadG, the reaction mixtures (500 µL) consisting of 200 µM **8**, 50 mM L-isoleucine, and 500 µM JadG were incubated at 30 °C for 30 min. Reactions were terminated with 2 M HCl (2 µL) and then extracted with ethyl acetate (500 µL). The extracts were dried using a miVac Duo concentrator (SP Scientific), redissolved in 20 µL ethanol, and then analyzed by HPLC.

Reaction mixtures of AlpG, AlpJ, Flu17, or JadG were analyzed on a Waters HPLC system equipped with a 1525 pump and a 2487 dual λ UV detector using an Agilent ZORBAX SB-C18 reversed-phase column (5 µm, 4.6 × 250 mm) or a SilGreen C18AB reversed-phase column (5 µm, 4.6 × 250 mm). The elution solvents were water with 0.1% (v/v) trifluoroacetic acid (solvent A) and acetonitrile with 0.1% (v/v) trifluoroacetic acid (solvent B). A 20-min linear gradient from 25% (v/v) to 100% (v/v) solvent B was used with a flow rate of 1.0 $mL \cdot min^{-1}$. Absorbance at 266 and 313 nm was monitored. For SOD quench experiments, the reaction mixtures were analyzed on an Agilent 1260 Infinity HPLC system with a diode array detector using an Agilent ZORBAX SB-C18 rapid resolution HT column (1.8 µm, 4.6 × 100 mm). A 10 min linear gradient from 20% (v/v) to 100% (v/v) solvent B was used with a flow rate of 1.5 $mL \cdot min^{-1}$. Full spectra from 190 to 640 nm were monitored.

# Supporting Information

### Supporting Information File 1

Sequence comparison results and phylogenetic tree of AlpJ-family enzymes and anthrone oxygenases, crystal structures of AlpJ and ActVA-Orf6, SDS-PAGE of purified proteins, HPLC traces of prosthetic group identification in AlpG, negative controls of enzymatic reactions, SOD inhibition reactions, and HRMS spectrum of hydroquinone–kinobscurinone **10**.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-102-S1.pdf]


# Acknowledgements

This article is devoted to our beloved mentor Professor Keqian Yang.

# Funding

This work was supported by the National Key R&D Program of China (2022YFC2303100) and the National Natural Science Foundation of China (Grant 32070067).


# Author Contributions

Jinmin Gao: investigation; methodology; writing – original draft; writing – review & editing. Liyuan Li: investigation; validation; writing – review & editing. Shijie Shen: investigation. Guomin Ai: investigation. Bin Wang: conceptualization; methodology; writing – review & editing. Fang Guo: methodology. Tongjian Yang: methodology. Hui Han: resources. Zhengren Xu: writing – review & editing. Guohui Pan: conceptualization; funding acquisition; supervision; writing – original draft; writing – review & editing. Keqiang Fan: conceptualization; funding acquisition; supervision; visualization; writing – original draft; writing – review & editing.

## ORCID® iDs

Hui Han - https://orcid.org/0000-0003-0772-5461
Zhengren Xu - https://orcid.org/0000-0001-7302-6488
Guohui Pan - https://orcid.org/0000-0002-8839-3886
Keqiang Fan - https://orcid.org/0000-0003-2420-1889

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Kharel, M. K.; Pahari, P.; Shepherd, M. D.; Tibrewal, N.; Nybo, S. E.; Shaaban, K. A.; Rohr, J. *Nat. Prod. Rep.* **2012,** *29,* 264–325. doi:10.1039/c1np00068c
2. Fan, K.; Zhang, Q. *Synth. Syst. Biotechnol.* **2018,** *3,* 275–282. doi:10.1016/j.synbio.2018.11.001
3. Gould, S. J. *Chem. Rev.* **1997,** *97,* 2499–2510. doi:10.1021/cr9600215
4. Hata, T.; Omura, S.; Iwai, Y.; Nakagawa, A.; Otani, M.; Ito, S.; Matsuya, T. *J. Antibiot.* **1971,** *24,* 353–359. doi:10.7164/antibiotics.24.353
5. Omura, S.; Nakagawa, A.; Yamada, H.; Hata, T.; Furusaki, A.; Watanabe, T. *Chem. Pharm. Bull.* **1973,** *21,* 931–940. doi:10.1248/cpb.21.931
6. He, H.; Ding, W.-D.; Bernan, V. S.; Richardson, A. D.; Ireland, C. M.; Greenstein, M.; Ellestad, G. A.; Carter, G. T. *J. Am. Chem. Soc.* **2001,** *123,* 5362–5363. doi:10.1021/ja010129o
7. Woo, C. M.; Beizer, N. E.; Janso, J. E.; Herzon, S. B. *J. Am. Chem. Soc.* **2012,** *134,* 15285–15288. doi:10.1021/ja3074984
8. Feng, Z.; Kim, J. H.; Brady, S. F. *J. Am. Chem. Soc.* **2010,** *132,* 11902–11903. doi:10.1021/ja104550p
9. Tibrewal, N.; Pahari, P.; Wang, G.; Kharel, M. K.; Morris, C.; Downey, T.; Hou, Y.; Bugni, T. S.; Rohr, J. *J. Am. Chem. Soc.* **2012,** *134,* 18181–18184. doi:10.1021/ja3081154
10. Fan, K.; Pan, G.; Peng, X.; Zheng, J.; Gao, W.; Wang, J.; Wang, W.; Li, Y.; Yang, K. *Chem. Biol.* **2012,** *19,* 1381–1390. doi:10.1016/j.chembiol.2012.09.009
11. Wang, B.; Ren, J.; Li, L.; Guo, F.; Pan, G.; Ai, G.; Aigle, B.; Fan, K.; Yang, K. *Chem. Commun.* **2015,** *51,* 8845–8848. doi:10.1039/c5cc01986a
12. Wang, P.; Hong, G. J.; Wilson, M. R.; Balskus, E. P. *J. Am. Chem. Soc.* **2017,** *139,* 2864–2867. doi:10.1021/jacs.6b10586
13. Pan, G.; Gao, X.; Fan, K.; Liu, J.; Meng, B.; Gao, J.; Wang, B.; Zhang, C.; Han, H.; Ai, G.; Chen, Y.; Wu, D.; Liu, Z.-J.; Yang, K. *ACS Chem. Biol.* **2017,** *12,* 142–152. doi:10.1021/acschembio.6b00621
14. Zhao, Y.; Liu, X.; Xiao, Z.; Zhou, J.; Song, X.; Wang, X.; Hu, L.; Wang, Y.; Sun, P.; Wang, W.; He, X.; Lin, S.; Deng, Z.; Pan, L.; Jiang, M. *Nat. Commun.* **2023,** *14,* 5372. doi:10.1038/s41467-023-41062-7
15. Rix, U.; Zheng, J.; Remsing Rix, L. L.; Greenwell, L.; Yang, K.; Rohr, J. *J. Am. Chem. Soc.* **2004,** *126,* 4496–4497. doi:10.1021/ja031724o
16. Wang, K.-K. A.; Ng, T. L.; Wang, P.; Huang, Z.; Balskus, E. P.; van der Donk, W. A. *Nat. Commun.* **2018,** *9,* 3687. doi:10.1038/s41467-018-06083-7
17. Jiang, X.; Zhang, Q.; Zhu, Y.; Nie, F.; Wu, Z.; Yang, C.; Zhang, L.; Tian, X.; Zhang, C. *Tetrahedron* **2017,** *73,* 3585–3590. doi:10.1016/j.tet.2017.03.054
18. Huang, C.; Yang, C.; Fang, Z.; Zhang, L.; Zhang, W.; Zhu, Y.; Zhang, C. *Mar. Drugs* **2019,** *17,* 150. doi:10.3390/md17030150
19. Jin, J.; Yang, X.; Liu, T.; Xiao, H.; Wang, G.; Zhou, M.; Liu, F.; Zhang, Y.; Liu, D.; Chen, M.; Cheng, W.; Yang, D.; Ma, M. *Mar. Drugs* **2018,** *16,* 87. doi:10.3390/md16030087
20. Liu, X.; Liu, D.; Xu, M.; Tao, M.; Bai, L.; Deng, Z.; Pfeifer, B. A.; Jiang, M. *J. Nat. Prod.* **2018,** *81,* 72–77. doi:10.1021/acs.jnatprod.7b00652
21. Kendrew, S. G.; Hopwood, D. A.; Marsh, E. N. *J. Bacteriol.* **1997,** *179,* 4305–4310. doi:10.1128/jb.179.13.4305-4310.1997
22. Shen, B.; Hutchinson, C. R. *Biochemistry* **1993,** *32,* 6656–6663. doi:10.1021/bi00077a019
23. Li, X.; Li, X.; Zhang, Q.; Lv, P.; Jia, Y.; Wei, D. *Org. Biomol. Chem.* **2022,** *20,* 5525–5534. doi:10.1039/d2ob00848c
24. Gui, C.; Kalkreuter, E.; Liu, Y.-C.; Li, G.; Steele, A. D.; Yang, D.; Chang, C.; Shen, B. *Nat. Chem. Biol.* **2024,** *20,* 243–250. doi:10.1038/s41589-023-01476-2
25. Chen, Y.; Fan, K.; He, Y.; Xu, X.; Peng, Y.; Yu, T.; Jia, C.; Yang, K. *ChemBioChem* **2010,** *11,* 1055–1060. doi:10.1002/cbic.201000178
26. Yang, C.; Huang, C.; Zhang, W.; Zhu, Y.; Zhang, C. *Org. Lett.* **2015,** *17,* 5324–5327. doi:10.1021/acs.orglett.5b02683
27. Janso, J. E.; Haltli, B. A.; Eustáquio, A. S.; Kulowski, K.; Waldman, A. J.; Zha, L.; Nakamura, H.; Bernan, V. S.; He, H.; Carter, G. T.; Koehn, F. E.; Balskus, E. P. *Tetrahedron* **2014,** *70,* 4156–4164. doi:10.1016/j.tet.2014.03.009
28. Yang, C.; Huang, C.; Fang, C.; Zhang, L.; Chen, S.; Zhang, Q.; Zhang, C.; Zhang, W. *J. Org. Chem.* **2021,** *86,* 11019–11028. doi:10.1021/acs.joc.0c02517
29. Machovina, M. M.; Usselman, R. J.; DuBois, J. L. *J. Biol. Chem.* **2016,** *291,* 17816–17828. doi:10.1074/jbc.m116.730051
30. Grocholski, T.; Koskiniemi, H.; Lindqvist, Y.; Mäntsälä, P.; Niemi, J.; Schneider, G. *Biochemistry* **2010,** *49,* 934–944. doi:10.1021/bi901985b
31. Sciara, G.; Kendrew, S. G.; Miele, A. E.; Marsh, N. G.; Federici, L.; Malatesta, F.; Schimperna, G.; Savino, C.; Vallone, B. *EMBO J.* **2003,** *22,* 205–215. doi:10.1093/emboj/cdg031
32. Machovina, M. M.; Ellis, E. S.; Carney, T. J.; Brushett, F. R.; DuBois, J. L. *J. Biol. Chem.* **2019,** *294,* 3661–3669. doi:10.1074/jbc.ra118.006144
33. Thierbach, S.; Bui, N.; Zapp, J.; Chhabra, S. R.; Kappl, R.; Fetzner, S. *Chem. Biol.* **2014,** *21,* 217–225. doi:10.1016/j.chembiol.2013.11.013
34. Hernández-Ortega, A.; Quesne, M. G.; Bui, S.; Heyes, D. J.; Steiner, R. A.; Scrutton, N. S.; de Visser, S. P. *J. Am. Chem. Soc.* **2015,** *137,* 7474–7487. doi:10.1021/jacs.5b03836
35. Bui, S.; von Stetten, D.; Jambrina, P. G.; Prangé, T.; Colloc'h, N.; de Sanctis, D.; Royant, A.; Rosta, E.; Steiner, R. A. *Angew. Chem., Int. Ed.* **2014,** *53,* 13710–13714. doi:10.1002/anie.201405485
36. Wei, D.; Huang, X.; Qiao, Y.; Rao, J.; Wang, L.; Liao, F.; Zhan, C.-G. *ACS Catal.* **2017,** *7,* 4623–4636. doi:10.1021/acscatal.7b00901
37. Bui, S.; Steiner, R. A. *Curr. Opin. Struct. Biol.* **2016,** *41,* 109–118. doi:10.1016/j.sbi.2016.06.015
38. Liu, T.; Fischer, C.; Beninga, C.; Rohr, J. *J. Am. Chem. Soc.* **2004,** *126,* 12262–12263. doi:10.1021/ja0467521
39. Kharel, M. K.; Zhu, L.; Liu, T.; Rohr, J. *J. Am. Chem. Soc.* **2007,** *129,* 3780–3781. doi:10.1021/ja0680515
40. Fetzner, S. *Appl. Microbiol. Biotechnol.* **2002,** *60,* 243–257. doi:10.1007/s00253-002-1123-4
41. Fetzner, S.; Steiner, R. A. *Appl. Microbiol. Biotechnol.* **2010,** *86,* 791–804. doi:10.1007/s00253-010-2455-0

42. Sambrook, J.; Russel, D. W. *Molecular Cloning: a Laboratory Manual,* 3rd ed.; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2001.
43. Lin, S.; Van Lanen, S. G.; Shen, B. *J. Am. Chem. Soc.* **2007,** *129,* 12432–12438. doi:10.1021/ja072311g
44. Chen, Y.-H.; Wang, C.-C.; Greenwell, L.; Rix, U.; Hoffmeister, D.; Vining, L. C.; Rohr, J.; Yang, K.-Q. *J. Biol. Chem.* **2005,** *280,* 22508–22514. doi:10.1074/jbc.m414229200

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.102

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Computation-guided scaffold exploration of 2*E*,6*E*-1,10-*trans/cis*-eunicellanes

Zining Li‡[1], Sana Jindani‡[2,3], Volga Kojasoy[2], Teresa Ortega[2], Erin M. Marshall[1,4], Khalil A. Abboud[1], Sandra Loesgen[1,4], Dean J. Tantillo*[2] and Jeffrey D. Rudolf*[1]

Full Research Paper

Open Access

Address:
[1]Department of Chemistry, University of Florida, PO Box 117200, Gainesville, FL 32611, USA, [2]Department of Chemistry, University of California–Davis, 1 Shields Ave., Davis, CA 95616, USA, [3]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, 201002, India and [4]Whitney Laboratory for Marine Bioscience, University of Florida, 9505 N Ocean Shore Blvd., St. Augustine, FL 32080, USA

Email:
Dean J. Tantillo* - djtantillo@ucdavis.edu; Jeffrey D. Rudolf* - jrudolf@chem.ufl.edu

* Corresponding author ‡ Equal contributors

Keywords:
atropisomer; Cope rearrangement; DFT calculations; diterpene; electrophilic cyclization; eunicellane

*Beilstein J. Org. Chem.* **2024,** *20,* 1320–1326.
https://doi.org/10.3762/bjoc.20.115

Received: 03 February 2024
Accepted: 21 May 2024
Published: 07 June 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Associate Editor: S. Bräse

© 2024 Li et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Eunicellane diterpenoids are a unique family of natural products containing a foundational 6/10-bicyclic framework and can be divided into two main classes, *cis* and *trans*, based on the configurations of their ring fusion at C1 and C10. Previous studies on two bacterial diterpene synthases, Bnd4 and AlbS, revealed that these enzymes form *cis*- and *trans*-eunicellane skeletons, respectively. Although the structures of these diterpenes only differed in their configuration at a single position, C1, they displayed distinct chemical and thermal reactivities. Here, we used a combination of quantum chemical calculations and chemical transformations to probe their intrinsic properties, which result in protonation-initiated cyclization, Cope rearrangement, and atropisomerism. Finally, we exploited the reactivity of the *trans*-eunicellane skeleton to generate a series of 6/6/6 gersemiane-type diterpenes via electrophilic cyclization.

## Introduction

The eunicellane diterpenoids are a family of nearly 400 natural products that present a conserved 6/10-bicyclic hydrocarbon framework [1-3]. Mostly known from soft corals [2], but with a growing number of family members in bacteria and plants [4], these diterpenoids have four main structural differences: the number and location of oxidized carbons, the absence or presence of transannular ether bridges, the configuration (*cis* or *trans*) of the bicyclic ring fusion, and the presence and configuration (*E* or *Z*) of alkenes in the 10-membered ring (Figure 1A). During biosynthesis, the eunicellane skeleton is first constructed by terpene synthases that cyclize the diterpene precursor geranylgeranyl diphosphate (Figure 1B) [5-9]. The

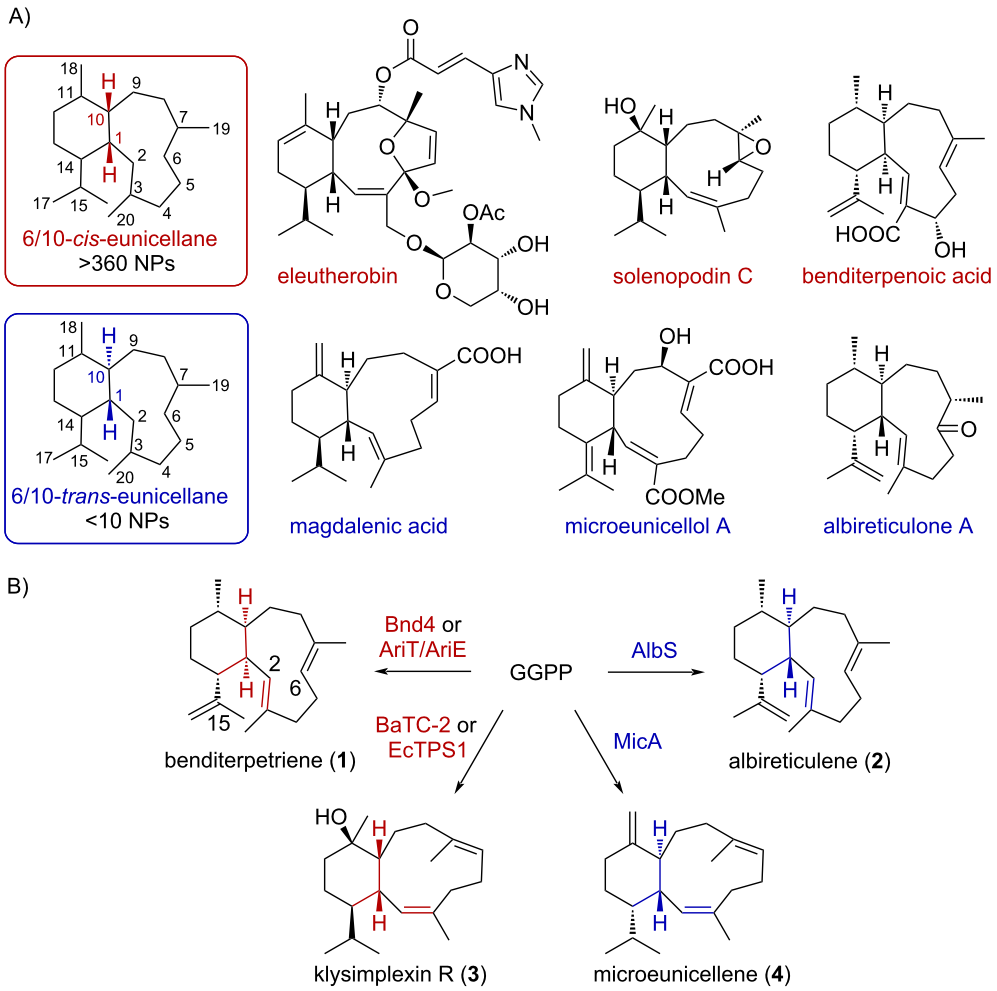



**Figure 1:** Eunicellane diterpenoids and their biosyntheses. (A) The 6/10-bicyclic hydrocarbon framework is conserved in eunicellane diterpenoids. Selected natural products consisting of *cis*- (red) and *trans*-eunicellane skeletons (blue) are shown. (B) Four types of diterpene synthases are known to form the eunicellane skeleton, each with differing configurations at the ring-fused carbons and the 2,3-alkene.

latter two structural elements, the configurations of the 6/10 system and the C2–C3 alkene, are instilled by these terpene synthases.

Four types of eunicellane synthases are known (Figure 1B). The first eunicellane synthase identified, Bnd4 from the biosynthesis of benditerpenoic acid in *Streptomyces* sp. (CL12-4) [5], forms a *cis*-eunicellane named benditerpetriene (**1**) [6]. In **1**, the C2–C3 and C6–C7 alkenes are *E*-configured, with the latter alkene configuration being conserved in all known eunicellane cyclization mechanisms. The first *trans*-eunicellane synthase, AlbS from the biosynthesis of albireticulone in *Streptomyces albireticuli* [10], was also identified from bacteria and forms albireticulene (**2**), a C1 diastereomer of **1** that also features the 2*E* alkene [7]. Two coral enzymes, BaTC-2 and EcTPS1, were found to form klysimplexin R (**3**), a 2*Z*-*cis*-eunicellane [8,9]. Recently, a third bacterial version, MicA, was identified as producing the 2*Z*-*trans*-eunicellane (**4**, microeunicellene) necessary for microeunicellol biosynthesis in *Micromonospora* sp. HM134 [11].

During our work on Bnd4, AlbS, and their enzymatic products **1** and **2**, we noticed that the *cis* and *trans*-eunicellane skeletons behaved differently under certain conditions. For example, we successfully collected NMR data of **1** in chloroform [5], but when we dissolved **2** in chloroform for NMR, it cyclized into two 6/6/6-tricyclic diterpenes (**5** and **6**) [7]. We discovered that **2** was much more sensitive to acid than **1** and eventually took advantage of its reactivity to determine its absolute configuration [7]. Here, we sought to understand the molecular basis of chemical and thermal reactivities of these diterpene skeletons. We also took advantage of their intrinsic chemical properties to transform the eunicellanes in functionalized tricyclic skeletons of natural product importance.

## Results and Discussion

### Ring fusion configuration does not affect protonation-induced cyclization

Our initial observation that albireticulene (**2**) was unstable in chloroform, resulting in two tricyclic isomers gersemienes A (**5**) and B (**6**) [7], while benditerpetriene (**1**) was stable [5], led us to investigate the protonation-induced cyclization of the eunicellane skeleton (Figure 2). When **1** and **2** were individually subjected to acidic conditions (i.e., TFA in $CHCl_3$), both cyclized into 6/6/6-tricyclic skeletons via a selective C2–C7 cyclization after protonation at C6 to give *trans*-BC ring systems (i.e., *cis*,*trans*-6/6/6 for **1** and *trans*,*trans*-6/6/6 for **2**; Figure 2A). Only a single isomer, with the exocyclic olefin on ring C, of *cis*,*trans*-gersemiene was found, which matched a previous report of **1** cyclization in aqueous 0.1 M HCl into gersemiene C (**7**, Figures S1–S3, Supporting Information File 1) [12].

To assess these chemically-induced cationic cyclization mechanisms, we performed quantum chemical calculations [mPW1PW91/6–31+G(d/p)/SMD(chloroform)] [13-20] to obtain the relative free energies of the cationic intermediates and transition states that interconvert them, as well as the relative free energies of the neutral products. In the cyclization of **2**, protonation at C6, results in a C7 tertiary cationic intermediate (**A2**$^+$) where C2 is only 1.65 Å away from C7 (Figure 2B); this

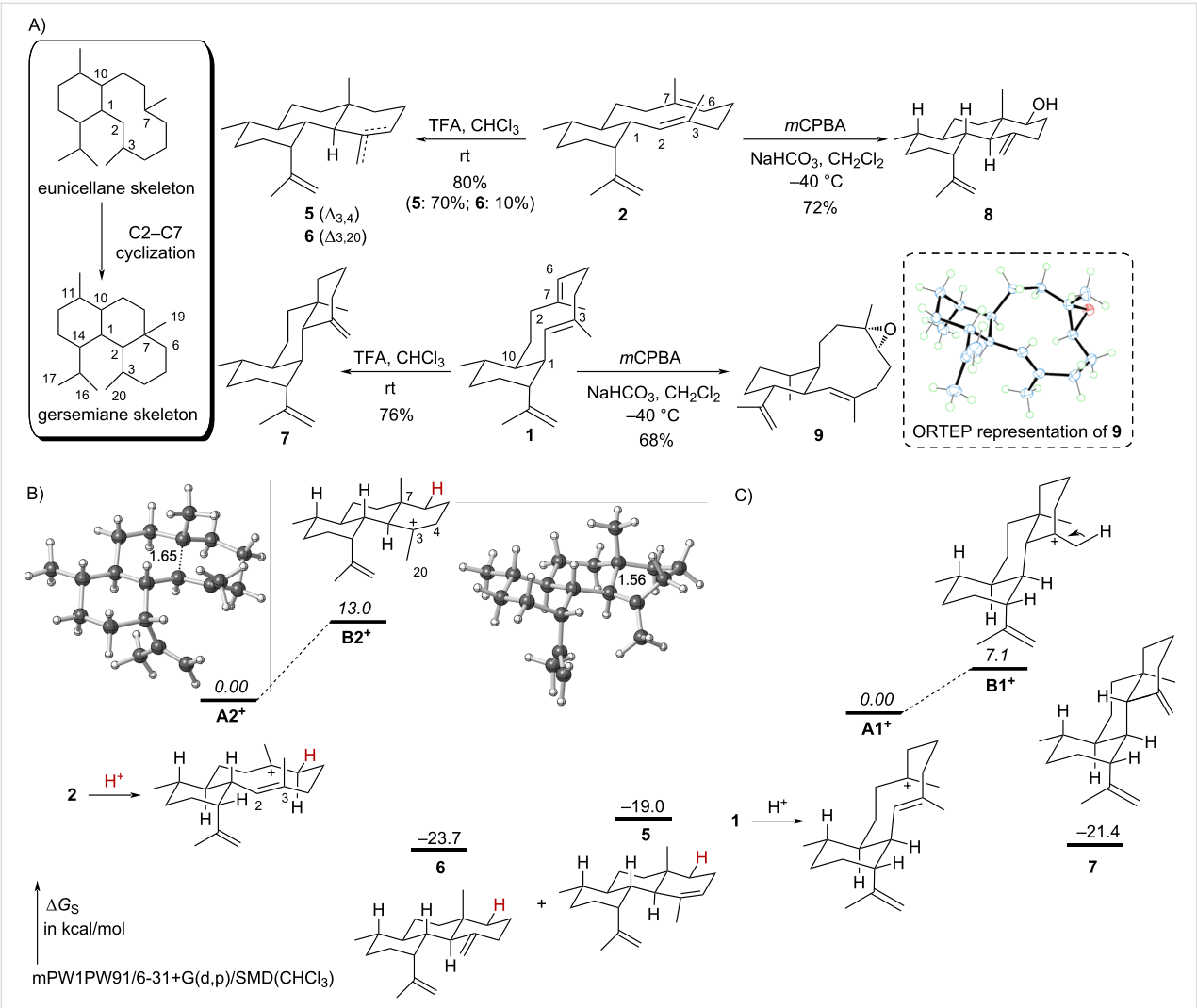



**Figure 2:** Protonation-mediated cyclization of *trans*- and *cis*-eunicellanes. (A) The 2*E*-*trans*- and 2*E*-*cis*-eunicellane skeletons form 6/6/6-tricyclic gersemiene skeletons in the presence of acid. Epoxidation of **1** with *m*CPBA yields the epoxide **9**, but the similar reaction with **2** yields gersemienol **8.** Isolation yields are provided. (B and C) Results of DFT calculations on the protonation-induced cyclizations of **1** and **2**. The energies of the cationic intermediates (italicized values) are not on the same energy scale as for the substrates and products (i.e., the free energy of **B2**$^+$ is relative to **A2**$^+$, while those of **5** and **6** are relative to that of **2**; the free energy of **B1**$^+$ is relative to **A1**$^+$, while that of **7** is relative to **1**; the energies given in B and C are not relative to each other). Note that the other rotamer of **7** is predicted to be 3.5 kcal mol$^{-1}$ higher in energy than the one shown, an energy difference that is larger than expected based on experiment.

structure can be viewed as protonated **5** or **6** with a strongly hyperconjugated C2–C7 bond [21]. Reducing this hyperconjugation via bending the exocyclic methyl group such that the formal *p*-orbital at C3 no longer aligns with the C2–C7 bond results in the 6/6/6-tricyclic C3 tertiary cation **B2**$^+$, which is 13.0 kcal mol$^{-1}$ higher in energy. In principle, either **A2**$^+$ or **B2**$^+$ could be deprotonated to form products. A similar energy profile is seen for protonation-induced cyclization of **1**, although the 6/6/6-tricyclic C3 tertiary cation is 7.1 kcal mol$^{-1}$ higher in energy than its preceding intermediate (Figure 2C). A lower energy conformer of **1** exists, but we have been unable to find the corresponding cation **A1b**$^+$, which implies that protonation and cyclization may be concerted for that conformer (Figure S4, Supporting Information File 1).

We previously transformed **2** into the *trans*,*trans*-6/6/6-tricyclic C6 alcohol **8** using *m*CPBA to epoxidize the C6–C7 alkene [7]. Under the same conditions, epoxidation of **1** yielded the 6,7-epoxide **9**, which readily crystallized (Figure 2A and Figures S5–S10, Table S1 in Supporting Information File 1). Similarly, the 6,7-epoxy derivatives of klysimplexin R (**3**) and microeunicellene (**4**) were recently synthesized and isolated [11,21]; **3** cyclized to the 6/6/6-scaffold after the addition of acid [22]. An X-ray structure of **9** allowed us to solve its absolute configuration and measure the C2–C7 distance to be 3.31 Å. This distance matches reasonably well with the calculated distance, 3.36 Å, of the lowest energy conformer of **9**. Calculation of the proposed 6,7-epoxy derivative of **2**, which was never isolated, gave a similar estimated C2–C7 distance of 3.57 Å, suggesting that a factor other than distance controls any subsequent or concomitant cyclization.

## Cope rearrangement is facile for *trans*-eunicellanes

During our study of the albireticulene (**2**) biosynthetic gene cluster, we found that **2** easily undergoes Cope rearrangement at 90 °C to generate the stereospecific 6/6-bicyclic product **10** in 96% yield (Figure 3A) [10]. This Cope rearrangement product was found as two inseparable atropisomers (**10a**/**10b**) at room temperature, which coalesced into a single conformer at 130 °C [10]. We anticipated that **1** would similarly undergo Cope rearrangement. However, when we heated **1** up to 200 °C for 5 h, we did not observe any Cope rearrangement products and we were able to recover >90% of **1**. DFT calculations

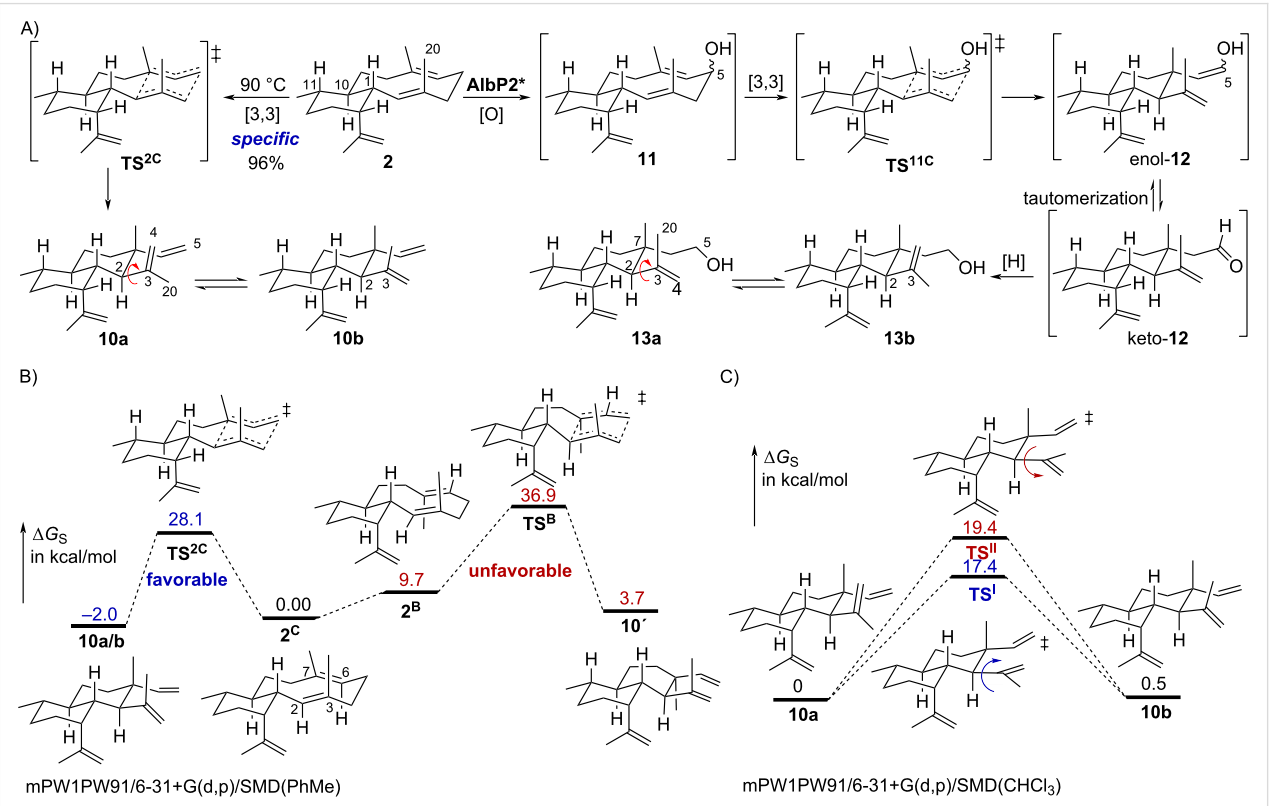



**Figure 3:** Cope rearrangement in *trans*-eunicellanes. (A) The 2*E*-*trans*-eunicellane undergoes thermal Cope rearrangement, yielding atropisomers **10a**/**10b**; an isolation yield of 96% was previously reported [10]. A biosynthetically related analog is similarly proposed to undergo oxy-Cope rearrangement to yield **13a**/**13b**. (B) Results of DFT calculations indicate that the Cope rearrangement of **2** is more favorable via a chair–chair transition state than a chair–boat transition state. (C) Results of DFT calculations on the atropisomers **10a**/**10b** and the free energy barriers for clockwise (blue) and counter-clockwise (red) rotations.

[mPW1PW91/6–31+G(d/p)/SMD(toluene)] on the Cope rearrangement of **2** revealed a free energy barrier through the chair–chair transition state of 28.1 kcal mol$^{-1}$ and an overall $\Delta G$ from **2** to **10a**/**10b**, which have the same predicted energy, of −2.0 kcal mol$^{-1}$ (Figure 3B). The relative free energy barrier through the chair–boat transition state was calculated to be 36.9 kcal mol$^{-1}$, but would require a prior chair-to-boat conformational change of 9.7 kcal mol$^{-1}$, providing a similar activation barrier for the Cope rearrangement step itself (27.2 kcal mol$^{-1}$). For **1**, a potential Cope rearrangement was predicted to be 32.5 kcal mol$^{-1}$, an overall higher energetic barrier compared to **2** (Figure S11, Supporting Information File 1). The lowest transition state for a Cope rearrangement, a chair–chair structure at 29.3 kcal mol$^{-1}$, originates from a DD (down–down orientations of the methyl groups on the 10-membered ring) conformer that is 3.2 kcal mol$^{-1}$ higher in free energy than the most dominant conformer, leading to the predicted overall barrier of 32.5 kcal mol$^{-1}$ and consistent with the diminished reactivity observed experimentally. The two lowest energy conformers of **1** (−3.2 and −3.0 kcal mol$^{-1}$ relative to the DD conformer mostly likely to undergo Cope rearrangement) would both require significantly higher free energy barriers of 38.0 and 38.8 kcal mol$^{-1}$, respectively, to react through their associated chair–boat Cope transition states (Figure S11, Supporting Information File 1). In addition, the resulting products were predicted to be approximately 10 kcal mol$^{-1}$ higher in free energies than the reactants. We note here, circumstantially, that no Cope rearrangement products of *cis*-eunicellanes have been reported. Overall, the *trans* ring fusion of the 2*E*,6*E*-*trans*-eunicellanes appears to control whether Cope rearrangement is energetically possible.

In the same biosynthetic study [10], we also identified what appeared to be an oxy-Cope product (**13a**/**13b**), which would logically originate from a 5-hydroxyalbireticulene analog (**11**, Figure 3A). Because this product was isolated directly from the producing bacterium as inseparable atropisomers (i.e., no heat was applied), we speculated that a lower activation energy may allow for oxy-Cope rearrangement at 28 °C. DFT calculations (in toluene) of 5-hydroxyalbireticulene for both the pseudo-axial and pseudo-equatorial conformations of the C5 hydroxy moiety suggest that there is not a significant difference in the free energy barriers for the oxy-Cope rearrangement (27.7 kcal mol$^{-1}$; Figure S12, Supporting Information File 1) of uncharged **11** compared to that of **2** (Figure 3B). The free energy barrier value for **11** is likely too high to be non-enzymatic at 28 °C and may at least require deprotonation of the hydroxy group or at least H-bonding with solvent [23].

In regard to the atropisomerism of **10a**/**10b**, the two atropisomers are approximately equivalent in energy (0.5 kcal mol$^{-1}$ difference) and require at least 17.4 kcal mol$^{-1}$ to undergo conversion (Figure 3C). Interestingly, rotation in different directions requires barriers that differ by 2 kcal mol$^{-1}$.

## Scaffold exploration of the eunicellane skeleton

During protonation-induced cyclization of **1** and **2**, the C6–C7 alkene showed higher nucleophilicity than either of the other two double bonds likely due to the unique conformation of the eunicellane skeleton. This selective reactivity was further supported when the C6–C7 alkene of **2** was oxidized by *m*CPBA oxidation, resulting in formation of the gersemianol derivative **8** [7]. Inspired by the reactivity of the *trans*-eunicellane skeleton, we conducted a series of chemical transformations to convert **2** into *trans*/*trans*-6/6/6 bicyclic skeletons with various functional groups at C6 (Figure 4). The goals were to evaluate the scope of electrophile-mediated cyclization and diversify the gersemiane skeleton for bioactivity assays. A structural similarity search of **5** and **6** revealed that several natural products, including the coral gersemiols [24] and plagicosin N from liverwort [25], possess a similar 6/6/6-skeleton and likely originate from a *cis*-eunicellane skeleton.

We first independently tested **2** with halogen electrophiles NCS and NBS (1.2 equivalents). NCS and NBS provided the C6-monohalogenated 6/6/6-tricyclic products **14** and **15** in 78% and 74% yields (Figure 4 and Figures S13–S23 in Supporting Information File 1). The $^1$H NMR chemical shifts for H6 of 6-chlorogersemiene (**14**), 3.81 ppm ($J$ = 12.1, 4.8 Hz), and 6-bromogersemiene (**15**), 4.03 ppm ($J$ = 12.4, 4.7 Hz), supported their assignments. The large coupling constants (>10 Hz) of H6 support its axial position and therefore the equatorial nature of the halogen. These configurations were also verified by NOESY correlations between H6 and H2 in both compounds. Some minor isomers of the halogenated derivatives were detected by LC–MS, but were deemed too low in abundance to purify. Two gersemiane derivatives (**16** and **17**), in a ratio of 3:1 in 82% overall yield, with a phenylselenyl groups at C6 were obtained when **2** was reacted with PhSeBr in acetonitrile overnight (Figure 4 and Figures S24–S36 in Supporting Information File 1). A similar reaction with PhSCl was also tested but yielded only several minor products and mostly starting material, hence we decided these products were not worth further purification. The minor isomers of the halogenated derivatives were presumably detected but in too low yields to purify. Since 6/6/6-tricyclic diterpenes are known to show cytotoxicity [26], we tested our seven analogs for cytotoxicity against human colon carcinoma HCT-116 cells. Unfortunately, none of the compounds up to 10 μM showed any activity when tested in MTT-based cell viability assays.

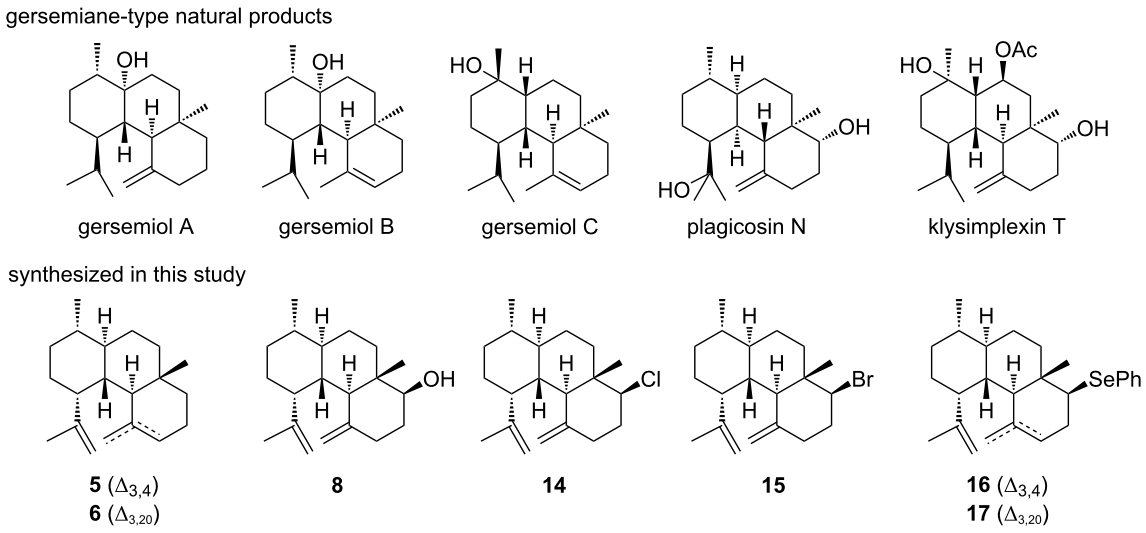



**Figure 4:** Scaffold exploration of the 2*E-trans*-eunicellane skeleton.

## Conclusion

Eunicellane diterpenoids have been known for over 50 years, but it was not until recently that their biosynthetic precursors, the eunicellane skeletons, were identified and isolated. Currently, four different eunicellane skeletons are known: 2*E-cis* (**1**), 2*E- trans* (**2**), 2*Z-cis* (**3**), and 2*Z-trans* (**4**). Based on the current distribution of eunicellane diterpenoids isolated, the coral-derived 2*Z-cis*-eunicellane appears to be the dominant form in nature (≈98%). Perhaps the inherent reactivity seen for 2*E-trans*-eunicellanes contributes to their presumed rarity in nature, as certain conditions or transformations, either enzymatic or spontaneous, may alter the hydrocarbon skeleton. This idea is supported by the recent discovery of the aridacins, which transforms a 2*E-cis*-eunicellane into a 6/7/5-tricyclic diterpene via a cytochrome P450 [12]. Coincidentally, these reactive skeletons also provide chemists the ability to access synthetically challenging scaffolds with simple electrophilic cyclizations. With genome mining of bacterial eunicellane synthases suggesting there are a significant number of both *cis*- and *trans*-eunicellane diterpenoids remaining to be discovered [5,10-12], we expect continuing synthetic and biosynthetic advancements in eunicellane diterpenoids in the near future.

## Supporting Information

### Supporting Information File 1

Experimental methods, NMR and MS spectra, and crystallographic information.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-115-S1.pdf]

## Acknowledgements

We acknowledge the University of Florida Center for Nuclear Magnetic Resonance Spectroscopy and Ion Ghiviriga for excellent NMR support. We acknowledge the University of Florida Mass Spectrometry Research and Education Center (MSREC), which is supported by the NIH (S10 OD021758-01A1), and Jodie Johnson for GC–MS support. We thank Wenbo Ning for measuring the melting point of **9**.

## Funding

This work was funded in part by NIH grant R35 GM142574 (to J.D.R.), NSF grant CHE-1856416 (to D.J.T.), and NSF CHE-2020110 (to S.L.). Computational resources were provided by NSF ACCESS program. S. J. was supported in part by the Institute of International Education (IIE) and the United States-India Education Foundation (USIEF) as part of a Fulbright Fellowship.

## ORCID® iDs

Zining Li - https://orcid.org/0000-0002-6921-8417
Sana Jindani - https://orcid.org/0000-0002-5224-6198
Volga Kojasoy - https://orcid.org/0000-0003-4832-0479
Erin M. Marshall - https://orcid.org/0009-0000-0681-0600
Sandra Loesgen - https://orcid.org/0000-0003-1090-564X
Dean J. Tantillo - https://orcid.org/0000-0002-2992-8844
Jeffrey D. Rudolf - https://orcid.org/0000-0003-2718-9651

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article. A data set collection of computational results is available in the ioChem-BD repository and can be accessed via https://doi.org/10.19061/iochem-bd-6-334. Crystal

data for compound 9 was deposited into the CCDC, deposition number 2326275.

## References


1. Welford, A. J.; Collins, I. *J. Nat. Prod.* **2011,** *74,* 2318–2328. doi:10.1021/np200125v
2. Li, G.; Dickschat, J. S.; Guo, Y.-W. *Nat. Prod. Rep.* **2020,** *37,* 1367–1383. doi:10.1039/d0np00016g
3. Dictionary of Natural Products. http://dnp.chemnetbase.com (accessed Dec 12, 2023).
4. Li, Z.; Rudolf, J. D. *J. Ind. Microbiol. Biotechnol.* **2023,** *50,* kuad027. doi:10.1093/jimb/kuad027
5. Zhu, C.; Xu, B.; Adpressa, D. A.; Rudolf, J. D.; Loesgen, S. *Angew. Chem., Int. Ed.* **2021,** *60,* 14163–14170. doi:10.1002/anie.202102453
6. Xu, B.; Tantillo, D. J.; Rudolf, J. D. *Angew. Chem., Int. Ed.* **2021,** *60,* 23159–23163. doi:10.1002/anie.202109641
7. Li, Z.; Xu, B.; Kojasoy, V.; Ortega, T.; Adpressa, D. A.; Ning, W.; Wei, X.; Liu, J.; Tantillo, D. J.; Loesgen, S.; Rudolf, J. D. *Chem* **2023,** *9,* 698–708. doi:10.1016/j.chempr.2022.12.006
8. Scesa, P. D.; Lin, Z.; Schmidt, E. W. *Nat. Chem. Biol.* **2022,** *18,* 659–663. doi:10.1038/s41589-022-01027-1
9. Burkhardt, I.; de Rond, T.; Chen, P. Y.-T.; Moore, B. S. *Nat. Chem. Biol.* **2022,** *18,* 664–669. doi:10.1038/s41589-022-01026-2
10. Li, Z.; Xu, B.; Alsup, T. A.; Wei, X.; Ning, W.; Icenhour, D. G.; Ehrenberger, M. A.; Ghiviriga, I.; Giang, B.-D.; Rudolf, J. D. *J. Am. Chem. Soc.* **2023,** *145,* 22361–22365. doi:10.1021/jacs.3c09446
11. Li, J.; Chen, B.; Fu, Z.; Mao, J.; Liu, L.; Chen, X.; Zheng, M.; Wang, C.-Y.; Wang, C.; Guo, Y.-W.; Xu, B. *ChemRxiv* **2023**. doi:10.26434/chemrxiv-2023-jzr1p
12. Wang, Z.; Yang, Q.; He, J.; Li, H.; Pan, X.; Li, Z.; Xu, H.-M.; Rudolf, J. D.; Tantillo, D. J.; Dong, L.-B. *Angew. Chem., Int. Ed.* **2023,** *62,* e202312490. doi:10.1002/anie.202312490
13. Adamo, C.; Barone, V. *J. Chem. Phys.* **1998,** *108,* 664–675. doi:10.1063/1.475428
14. All structures were fully optimized (and confirmed as minima or transition state structures through vibrational analysis). Free energies were computed at this level of theory, which has been validated previously for terpene-forming carbocation cyclization/rearrangements; see, for example [15-18].
15. Tantillo, D. J. Exploring Terpenoid Biosynthesis With Quantum Chemical Computations. In *Comprehensive Natural Products III: Chemistry and Biology;* Liu, H.; Begley, T. P., Eds.; Elsevier: Amsterdam, Netherlands, 2020; pp 644–653. doi:10.1016/b978-0-12-409547-2.14656-6
16. Matsuda, S. P. T.; Wilson, W. K.; Xiong, Q. *Org. Biomol. Chem.* **2006,** *4,* 530–543. doi:10.1039/b513599k
17. Tantillo, D. J. *Nat. Prod. Rep.* **2011,** *28,* 1035–1053. doi:10.1039/c1np00006c
18. Tantillo, D. J. *Angew. Chem., Int. Ed.* **2017,** *56,* 10040–10045. doi:10.1002/anie.201702363
19. *Gaussian 16,* Revision C.01; Gaussian, Inc.: Wallingford, CT, 2016.
20. A data set collection of computational results is available in the ioChem-BD repository and can be accessed via https://doi.org/10.19061/iochem-bd-6-334. doi:10.19061/iochem-bd-6-334.
21. Tantillo, D. J. *Chem. Soc. Rev.* **2010,** *39,* 2847–2854. doi:10.1039/b917107j
22. Scesa, P. D.; Schmidt, E. W. *Angew. Chem., Int. Ed.* **2023,** *62,* e202311406. doi:10.1002/anie.202311406
23. Schneider, C.; Weise, C. F. Cope, Oxy-Cope, and Anionic Oxy-Cope Rearrangements. *Comprehensive Organic Synthesis,* 2nd ed.; Elsevier: Amsterdam, Netherlands, 2014; Vol. 5, pp 867–911. doi:10.1016/b978-0-08-097742-3.00520-6
24. Angulo-Preckler, C.; Genta-Jouve, G.; Mahajan, N.; de la Cruz, M.; de Pedro, N.; Reyes, F.; Iken, K.; Avila, C.; Thomas, O. P. *J. Nat. Prod.* **2016,** *79,* 1132–1136. doi:10.1021/acs.jnatprod.6b00040
25. Qiao, Y.-N.; Jin, X.-Y.; Zhou, J.-C.; Zhang, J.-Z.; Chang, W.-Q.; Li, Y.; Chen, W.; Ren, Z.-J.; Zhang, C.-Y.; Yuan, S.-Z.; Lou, H.-X. *J. Nat. Prod.* **2020,** *83,* 1766–1777. doi:10.1021/acs.jnatprod.9b00895
26. Matias, D.; Nicolai, M.; Saraiva, L.; Pinheiro, R.; Faustino, C.; Diaz Lanza, A.; Pinto Reis, C.; Stankovic, T.; Dinic, J.; Pesic, M.; Rijo, P. *ACS Omega* **2019,** *4,* 8094–8103. doi:10.1021/acsomega.9b00512


## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.115

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Bioinformatic prediction of the stereoselectivity of modular polyketide synthase: an update of the sequence motifs in ketoreductase domain

Changjun Xiang[1,2], Shunyu Yao[2], Ruoyu Wang[2] and Lihan Zhang[*2,3,4]

Full Research Paper

Open Access

Address:
[1]Department of Chemistry, Fudan University, Shanghai 200433, China, [2]Key Laboratory of Precise Synthesis of Functional Molecules of Zhejiang Province, Department of Chemistry, School of Science and Research Center for Industries of the Future, Westlake University, Hangzhou 310030, China, [3]Institute of Natural Sciences, Westlake Institute for Advanced Study, Hangzhou 310024, China and [4]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou 310030, China

Email:
Lihan Zhang[*] - zhanglihan@westlake.edu.cn

* Corresponding author

Keywords:
bioinformatics; conserved motifs; ketoreductase; polyketide synthase; stereocontrol

*Beilstein J. Org. Chem.* **2024,** *20,* 1476–1485.
https://doi.org/10.3762/bjoc.20.131

Received: 18 April 2024
Accepted: 21 June 2024
Published: 02 July 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra

© 2024 Xiang et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Polyketides are a major class of natural products, including bioactive medicines such as erythromycin and rapamycin. They are often rich in stereocenters biosynthesized by the ketoreductase (KR) domain within the polyketide synthase (PKS) assembly line. Previous studies have identified conserved motifs in KR sequences that enable the bioinformatic prediction of product stereochemistry. However, the reliability and applicability of these prediction methods have not been thoroughly assessed. In this study, we conducted a comprehensive bioinformatic analysis of 1,762 KR sequences from *cis*-AT PKSs to reevaluate the residues involved in conferring stereoselectivity. Our findings indicate that the previously identified fingerprint motifs remain valid for KRs in β-modules from actinobacteria, but their reliability diminishes for KRs from other module types or taxonomic origins. Additionally, we have identified several new motifs that exhibit a strong correlation with the stereochemical outcomes of KRs. These updated fingerprint motifs for stereochemical prediction not only enhance our understanding of the enzymatic mechanisms governing stereocontrol but also facilitate accurate stereochemical prediction and genome mining of polyketides derived from modular *cis*-AT PKSs.

## Introduction

Type I modular polyketide synthases (PKSs) are large enzyme complexes that play a crucial role in the biosynthesis of bacterial polyketides, including many important clinical drugs such as erythromycin (antibiotic), epothilone (anticancer), ivermectin (antiparasitic), and spinosyn (insecticide) [1,2]. Modular PKSs consist of multiple modules that catalyze one round of chain

extension by domains with different functions and are divided into *cis*-acyltransferase (*cis*-AT) and *trans*-AT PKSs depending on whether AT domain is embedded in the assembly line or not. All *cis*-AT PKS modules contain a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP) to produce β-keto-intermediates, and some modules contain additional β-processing domains such as ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) [3,4]. The building blocks for PKS biosynthesis often include malonyl-CoA or methylmalonyl-CoA, which are loaded onto the ACP by the AT domain. Subsequently, the KS domain catalyzes the decarboxylative Claisen condensation between the ACP-tethered extender unit and the KS-tethered growing chain. The elongated growing chain may undergo further processing by KR, DH, and ER domains, generating β-hydroxy, α,β-alkene, and saturated β-methylene groups, respectively (Figure 1a).

KR domains have garnered significant attention from researchers due to their ability to control the stereochemistry at the α- and β-positions of a polyketide chain [5-7]. KRs are classified into three types: A-type and B-type denote KRs that form L- and D-configured β-hydroxy products, respectively, while C-type denotes KRs incapable of reducing the β-keto group [8,9]. KRs are further classified into subtypes 1 and 2 when the substrate contains an α-substition. Subtype 1 (A1, B1, and C1) KRs retain the original α-D-configuration, while subtype 2 (A2, B2, and C2) KRs epimerize the α-carbon to yield α-L-configured products (Figure 1b). The stereocontrol of KR has been found to correlate with several conserved sequence motifs identified through bioinformatic and structural analyses (Figure 1c) [8-11]. A pioneering work by Caffrey identified that A-type KRs possess the conserved "W" motif but lack the "LDD" motif, whereas B-type KRs possess the "LDD" motif but lack the "W" motif, based on the sequence alignment of 68 KRs from 17 modular *cis*-AT PKSs [8]. In addition, Keatinge–Clay reported a conserved "H" motif in the sequence of A2-type KRs and a "P" motif in B2-type KRs as markers to distinguish them from the non-epimerizing A1/B1-type KRs [9]. The presence of the catalytic "Y" motif and the absence of the NADPH binding motif can be used to predict C2-type KRs [9]. These conserved motifs have been widely used to predict the stereochemical outcome of modular *cis*-AT PKSs and have facilitated bioinformatics-guided structural determination of complex polyketides [12-17].

However, despite being widely adopted, the prediction accuracy and the applicable range of these conserved motifs remain elusive, and not a few exceptions to these prediction rules have been reported (Figure 1c). For instance, the stereoselectivity of KR7 of ibomycin [12], KR15 of neaumycin B [14], and KR7 and KR10 of rifamycin [18] PKS cannot be accurately predicted due to the absence of both the LDD and W motifs. Moreover, KR3 of neaumycin B [14], KR19 of caniferolide A [15], KR16 of epemicin B [16], and KR26 of gargantulide B [17] show discrepancies between bioinformatic prediction and the determined product configuration. There are also instances where KRs contain or lack both of the LDD and W motifs, making it difficult to predict stereochemistry (Figure 1c).

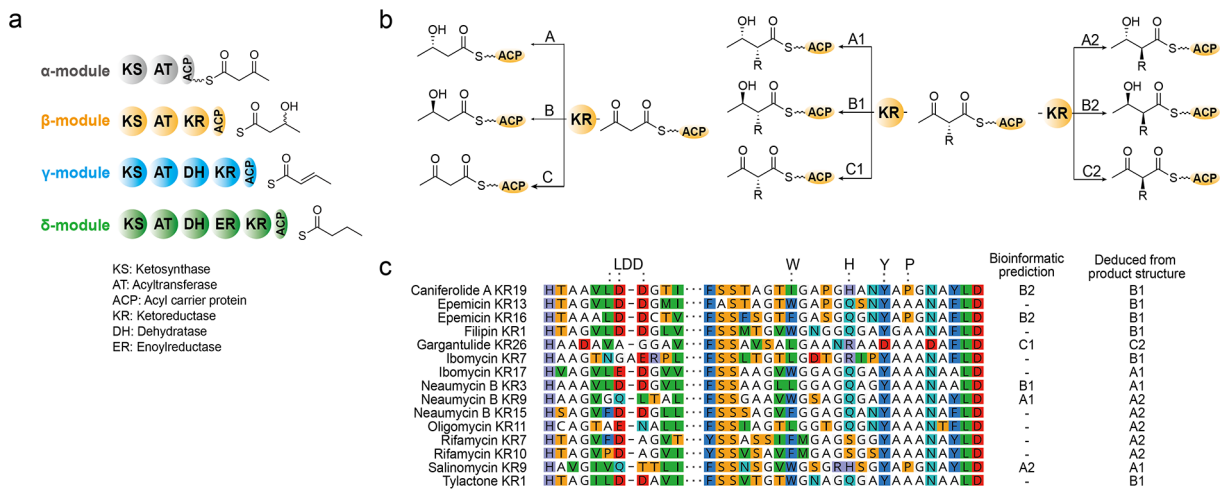



**Figure 1:** (a) Domain compositions and the products of α-, β-, γ- and δ-modules in *cis*-AT PKSs. (b) Stereocontrol of KR domains. KRs can be classified into A-type (β-L-hydroxy), B-type (β-D-hydroxy), and C-type (reduction-incompetent; β-keto) depending on the product structure and into subtypes 1 (non-epimerizing, α-D-substitution product) and 2 (epimerizing, α-L-substitution product). (c) Sequence alignment of KR domains whose stereoselectivity cannot be predicted based on the reported motifs. The labels show the product name and the order of the module where KR is located in the PKS assembly line (starting from the module with condensation function). KR types predicted based on the fingerprint motifs and deduced from product structure identified by chemical methods are shown on the right.

With the increasing availability of whole genome sequencing, a growing number of PKS gene clusters and their products have been characterized [19]. To assess and evaluate the sequence–stereoselectivity relationship of KRs, here we have collected 1,762 KR sequences from modular *cis*-AT PKS gene clusters, whose product structures have been verified using spectroscopic and/or chemical methods. We reveal that the previously identified conserved motifs are best applicable to KRs in the β-module of actinobacterial PKSs but less applicable for other types of KRs. Moreover, we have identified additional fingerprint residues that improve stereochemical prediction. These fingerprint residues also suggest potential mode of interactions among KR, ACP, and DH domains, deepening our understanding of the stereocontrol of PKSs.

# Results and Discussion

## Preparation of KR sequence dataset

We first curated the amino acid sequences of KR domains from characterized bacterial *cis*-AT PKSs recorded in MIBiG database [20] and by manual literature review. In total, 1,762 KRs whose product structures were experimentally determined, such as by crystallography, nuclear magnetic resonance (NMR) analysis or by chemical synthesis, were obtained for further analysis.

The modules in PKSs were categorized as α-module (containing KS-AT-ACP tridomain), β-module (KS-AT-KR-ACP), γ-module (KS-AT-DH-KR-ACP), and δ-module (KS-AT-DH-ER-KR-ACP) based on their domain composition regardless of the domain activity [2] (Figure 1a). While the stereoselectivity of a KR in a β-module can be directly inferred from the product hydroxy group, the stereochemical outcome of a KR from a γ- and δ-module is obscured by the following dehydration by the DH domain. It is widely believed that an α,β-*trans* (*E*) double bond is generated from a D-β-hydroxy intermediate produced by a B-type KR, whereas a *cis* (*Z*) double bond is from an L-β-hydroxy intermediate generated by an A-type KR via *syn* elimination [21]. However, based on the *syn*-elimination mechanism, the L-α-methyl,L-β-hydroxy substrate produced by A2-type KR can also result in a *trans* (*E*) double bond [18,22], and some DH domains are reported to have epimerase activity on the α-substitution [23]. Thus, the stereochemical outcomes of KRs in γ- and δ-modules cannot be directly inferred from the final product structure, and we discuss these KRs separately based on their module types.

The curated 1,762 KR sequences were grouped into different types based on their product structure and the taxonomy of their host strains (Figure 2a). Among the KR sequences we collected, 90% are from *Actinomycetota*, and less than 10% are from *Myxococcota* and *Cyanobacteriota*. Among actinobacterial KRs, γ-modules accounted for more than half, of which approximately 78% resulting in a *trans* (*E*) double bond, 5% c*is* (*Z*) double bond, and the remaining 17% possessing an inactive DH domain. Among actinobacterial KRs from β-modules, A-type KRs (52%) are the most abundant, followed by B-type KRs (36%), and C-type KRs (12%).

## Comparison between $KR_S$ and $KR_C$ subdomains

KR domains are structurally divided into two subdomains: a catalytic subdomain ($KR_C$) with an intact Rossmann fold where the active site resides, and a structural subdomain ($KR_S$) with a truncated Rossmann fold that lacks NADPH binding sites and solely provides structural support by forming a heterodimer with $KR_C$ [24-26]. In δ-modules, an ER domain is inserted between $KR_S$ and $KR_C$. Phylogenetic analyses of $KR_C$ and $KR_S$ revealed that both subdomains exhibited a cladogram pattern dependent on taxonomy, with cyanobacterial and myxobacterial KRs forming an ancestral clade followed by actinobacterial KRs (Figure 2b, and Supporting Information File 1, Figures S1 and S2). Such cladogram pattern was also observed in their KS domains [27], suggesting that they have diverged at an early point during evolution. Furthermore, both $KR_S$ and $KR_C$ trees showed that KRs belonging to γ- and δ-modules formed a separate clade distinct from those in β-modules, indicating that the existence of DH domain significantly influences the sequence of both KR subdomains [13]. This suggests a strong domain–domain interaction between KR and DH, as observed in the structure of KR-DH-ER tridomain (Figure 2c) [24], and indicates that KRs in β-modules and γ/δ-modules may not be generally interchangeable for domain swapping.

However, the phylogenetic cladograms of the two subdomains were significantly different within the clade formed by β-modules. In the $KR_C$ tree, stereoselectivity-dependent clades were formed, but such clades were not observed in the $KR_S$ tree (Figure 2b and Figure S1 in Supporting Information File 1). This indicates that the stereoselectivity of KR is controlled solely by $KR_C$, and $KR_S$ does not influence its catalytic selectivity. Moreover, A2- and B2-type $KR_C$s formed a separatable clade from A1- and B1-type $KR_C$, respectively, suggesting that phylogenetic analysis can be used for stereochemical prediction. It is noteworthy that A0- (0 denotes as product without α-substitution) and A1-type KRs, and B0- and B1-type KRs cannot be distinguished in the phylogenetic trees of both subdomains. This implies that KRs likely exhibit promiscuity for α-substitution, and the structural differences in the α-position should not affect the original stereoselectivity of a KR. Such promiscuity can also be observed in the MycA KR in mycolactone biosynthesis, which accepts both α-substituted and α-unsubstituted

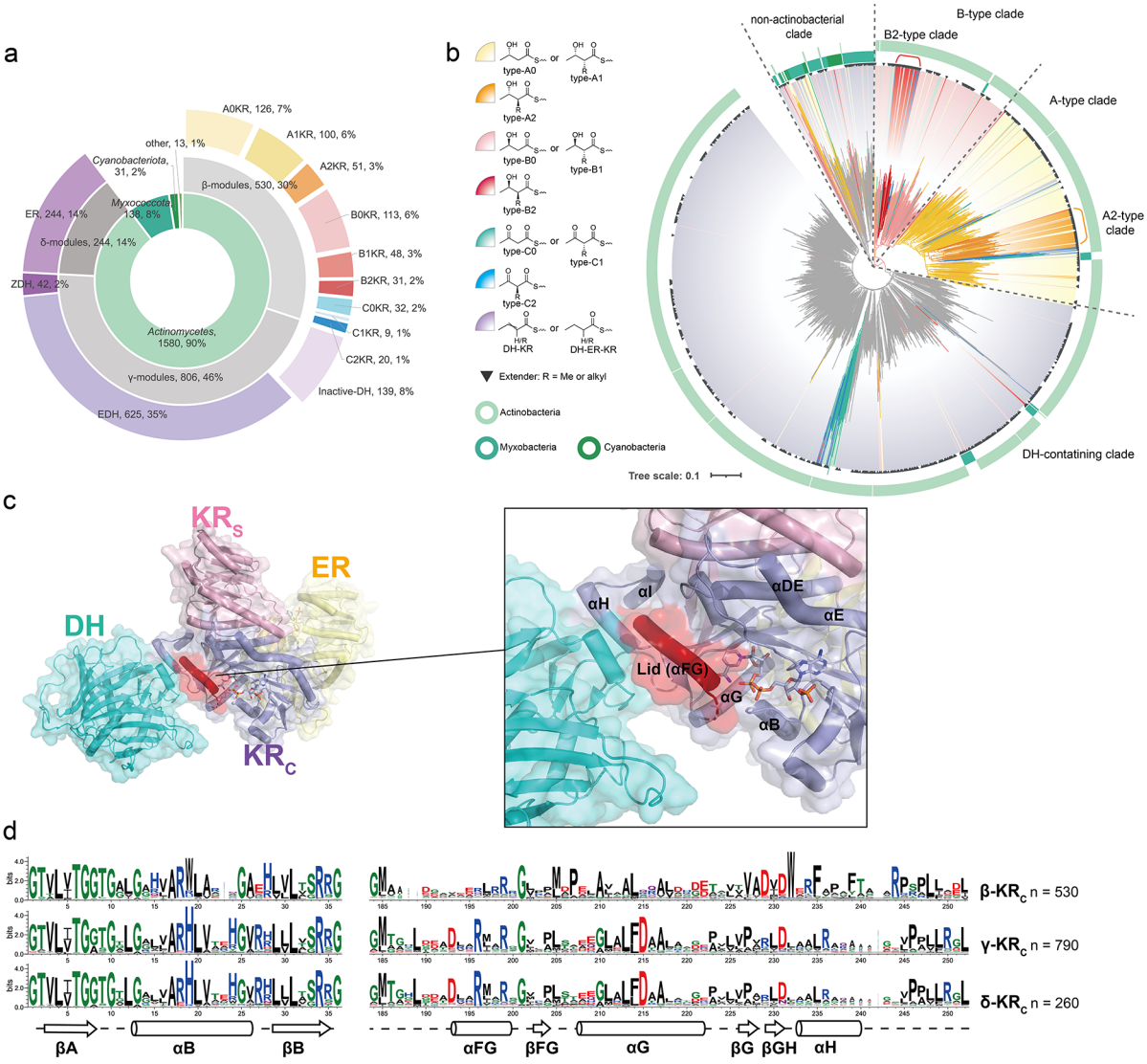



**Figure 2:** (a) Classification and distribution of the collected KR sequences according to host taxonomy, module types, and stereoconfigurations of the product. Here, "EDH" and "ZDH" respectively denote the *E*- (*trans*-) and *Z*- (*cis*-) configurations of a double bond. "ER" represents an α,β-alkane product. The numbers indicate the quantity of each KR type. (b) Phylogenetic analysis of the $KR_C$ subdomains. Product types were labeled by the inner color ranges, and the taxonomic origins were labeled by outer color rings. The triangles indicate a product having α-substituents. (c) Crystal structure of DH-ER-KR tridomain (PDB ID 8G7W) [24]. DH, $KR_S$, ER, $KR_C$ are colored in cyan, pink, yellow, and purple, respectively. The lid of KR catalytic pocket is colored in red. The co-crystallized cofactor is represented as sticks. (d) Sequence logo comparation of actinobacterial $KR_C$s sorted by module types. The overall height of a stack in the logo indicates the level of conservation, while the width represents the number of amino acid residues at each position. For full-length $KR_C$ logos, see Figure S3 in Supporting Information File 1.

substrates while retaining the same stereoselectivity for the β-hydroxy group [28].

To further investigate the sequence features of each module type, we performed sequence logo analyses of $KR_C$ subdomains from β-, γ- and δ-modules. Consistent with the phylogenetic cladogram, $KR_C$s from γ- and δ-modules showed nearly identical logo features each other but were significantly different from the logo of β-modules. Major differences were observed in two regions of $KR_C$, the *N*-terminal helix αB and the lid region αFG located at the *C*-terminal of $KR_C$ (Figure 2d). Indeed, this lid region exhibits a direct interaction with the DH and DH-KR linker (Figure 2c). However, such interaction was not observed between the *N*-terminal helix αB and DH in the crystal structure of DH-ER-KR tridomain (Figure 2c). This finding suggests that the helix αB may have an allosteric interaction with DH or potential large conformational changes between DH and KR during catalysis.

## Sequence logo analysis of $KR_C$ from β-modules

Based on the clear stereoselectivity-dependent clades observed in the $KR_C$ tree of actinobacterial β-modules, we next carried out sequence logo analyses to identify key amino acid residues associated with stereoselectivity. Comparison of sequence logos among different types of $KR_C$s revealed that the previously reported motifs [8,9], such as LDD (**2**) for B-type, W (**7**) for A-type, H (**8**) for A2-type, and P (**10**) for B2-type KR are highly conserved (Figure 3 and Figure S3 in Supporting Information File 1). In KRs from *Myxococcota* and *Cyanobacteriota*, while the LDD and W motif were mostly conserved, the H and P motifs characteristic for A2- and B2-type KRs, respectively, were not conserved (Figure S4). Thus, more characterized sequences are needed to better predict the stereochemistry at the α-position derived from non-actinobacterial KRs.

In actinobacterial KRs, we identified several new fingerprint motifs associated with stereoselectivity. First, A-type KRs and B-type KRs exhibited distinct variations around the αDE helix (highlighted in pink shadow area in Figure 3b). The difference is across the entire region, with the conserved R (**3**) in B-type KRs being most representative. This αDE helix region is located at the entrance of the predicted B-type KR substrate channel (Figure 3a) [9], suggesting its potential contribution for regulating substrate entry. On the αFG lid helix, a RLXR (**12**) motif

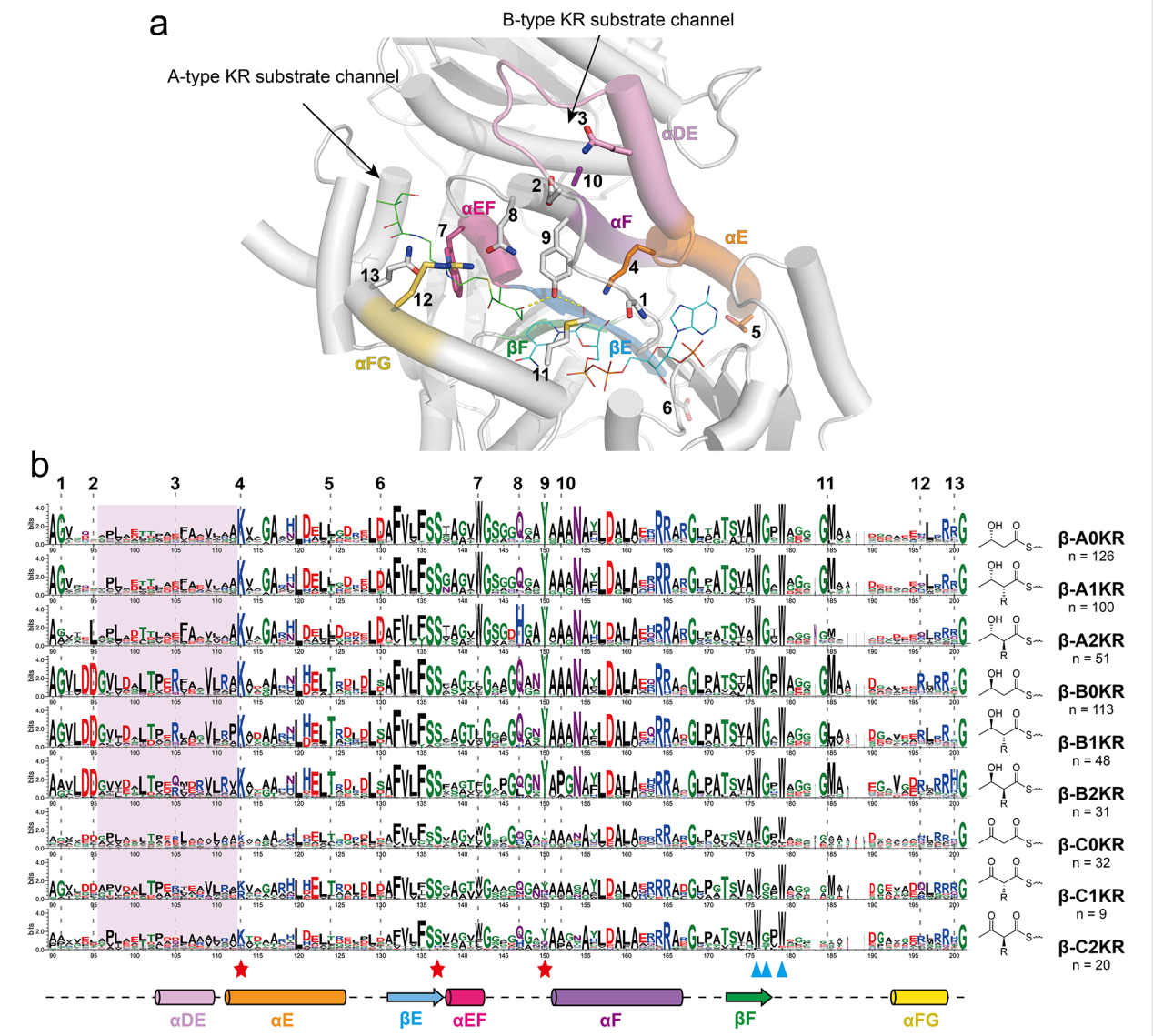



**Figure 3:** (a) The position of fingerprint motifs in the KR structure (AmpKR2, A1-type, PDB ID 5XWV) [29]. Fingerprints are shown as sticks. The co-crystallized substrate mimic and co-factor are shown as lines. (b) Sequence logo comparation of the core moiety of β-module $KR_C$ based on the classification of their products. The key catalytic residues are marked by red stars, and the NADPH-binding residues (partially) are marked by blue triangles. The numbers at the top indicate the fingerprint motifs.

was found in B-type KRs, while a XLXR motif was found in A-type KRs. A highly conserved T (**5**) was also identified in B-type KRs, whereas it was replaced by a more hydrophobic L in A-type KRs. On the loop between αE and βE, a conserved D (**6**) was found in A-type KRs.

For A2-type KRs, we identified a highly conserved L (**2**) in the position corresponding to the LDD motif (Figure 3). This residue may participate in interactions with the lid helix to adjust the pocket for epimerization of α-substituted intermediates. The lack of the conserved GM (**11**) motif preceding the αFG helix, which is highly conserved in other types of reductase-competent KRs, was also observed in A2-type KRs. The absence of this region in the electron density map of the A2-type AmpKR11 [30] may indicate a more flexible structure of the A2-type lid helix.

In B2-type KRs, we identified a conserved H (**13**) located at the end of lid helix (αFG) and the previously identified P (**10**) as signature motifs (Figure 3). Slight differences were observed in the loop region between αEF and αF, such as the more conserved P and less conserved Q, preceding the catalytic Y (**9**). Furthermore, some differences were noted in the αDE region downstream of the LDD motif, such as the R to Q mutation (**3**) and a conserved (D/E)RVLR between the R (**3**) and K (**4**) motifs, which may also regulate the structure of the loop forming the catalytic groove.

The sequence logo of C0/C1-type KRs showed that some of them possess the catalytic Y (**9**), but such KRs instead possess mutation in NADPH binding site (Figure S3 in Supporting Information File 1). In C2-type KRs, we found that most C2-type KRs possessed the catalytic Y (**9**), but some possessed Q at this residue. This substitution can be reasonable, as it serves as a hydrogen bond donor to activate the β-keto moiety of a substrate [31,32]. Additionally, the presence of K (**4**), which activates the catalytic tyrosine, was more conserved in C2-type KRs. In general, C2-KRs show similarity to either A2- or B2-KRs, but are likely to contain mutations in NADPH binding site [31].

## Sequence logo analysis of $KR_C$ from γ- and δ-modules

In the actinobacterial γ- and δ-modules, the stereochemical outcome of a KR is obscured by further dehydration catalyzed by DH. The dehydration mechanism by DH has been investigated in several DH domains, which all follow the *syn*-elimination mechanism [5,21,33]. Based on the product configuration, we classified DHs into two types, *E*-type DHs (*trans*-double bond) and *Z*-type DHs (*cis*-double bond) and analyzed sequence logos of the DH-associated KRs.

The sequence logo analysis of these $KR_C$s revealed that the LDD (**2**) motif representative of B-type KR is highly conserved in KRs associated with *E*-type DHs (Figure 4 and Figure S5 in Supporting Information File 1), which is in agreement with previous proposal that the *trans* (*E*) double bond is formed from the D-β-hydroxyacyl intermediate produced by B-type KR, and a *cis* (*Z*) double bond is formed from the L-β-hydroxyacyl intermediate [5]. However, only 5 out of 42 KRs associated with *Z*-type DHs contain the conserved W (**7**) motif found in A-type KRs. Additionally, some KRs associated with *Z*-type DHs possessed the LDD (**2**) motif. These observations suggest several possibilities: (1) the stereoselectivity prediction for KRs in γ- and δ-modules is inaccurate, (2) a *cis* (*Z*) double bond can also be produced from D-β-hydroxyacyl intermediates due to a different substrate binding mode in DH [7], or (3) a *cis* (*Z*) double bond is formed during post-PKS modification [22,34].

To gain a better understanding of the stereoselectivity of KRs in γ- and δ-modules, we focused on KRs associated with an inactive DH domain ($DH^0$) that still produce a β-hydroxy intermediate. These KRs are phylogenetically mixed together with those in γ- and δ-modules with an active DH (Supporting Information File 1, Figure S2), making them good models for analyzing the stereoselectivity of DH-associated KRs. Based on product structures, we found eleven A-type KRs and as many as 102 B-type KRs from γ- and δ-modules, similar to the ratio of *Z*- and *E*-DHs (Figure 4). However, all A-type KRs associated with an inactive DH lacked the diagnostic W (**7**) motif, and other A-type characteristic motifs such as L (**5**) and D (**6**) were also missing or less conserved (Figure 4). Contrary, nearly all B-type KRs associated with an inactive DH possessed the second D in the LDD (**2**) and the T (**5**) motifs. B2-type KRs showed slightly less conserved LD in the LDD motif, which was also reported as a feature of KRs in *trans*-AT PKSs [11]. In γ-A2-type KR, the characteristic L (**2**) and H (**8**) motifs found in β-A2-type KR were absent, and γ- and δ-B2-type KR lacked the characteristic P (**10**) and H (**13**) features. Therefore, the situation for KRs in γ- and δ-modules is complicated: a KR containing the second D in the LD**D** motif is likely to be B-type, but the possibility of it being A-type cannot be excluded, and the stereochemistry of the α-carbon cannot be predicted.

## Towards better bioinformatic prediction of stereochemical outcomes

Based on the above analyses, here we provide an update in the scope and motifs used for bioinformatic prediction of the α-methyl (or -alkyl) and β-hydroxy stereochemistry in modular *cis*-AT PKSs. First, actinobacterial β-module KRs are good targets for bioinformatic stereochemical prediction by the previously reported motifs, and newly identified motifs can further support the prediction. However, other KRs cannot be pre-

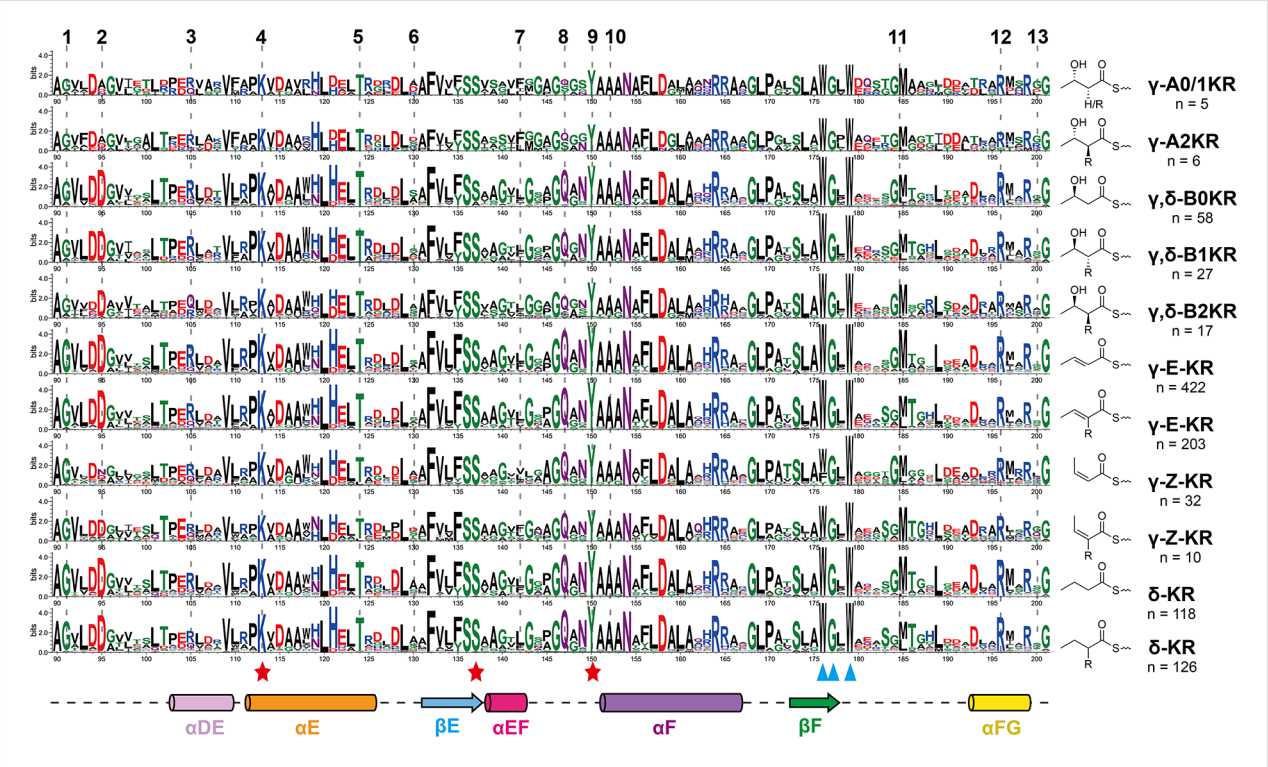



**Figure 4:** Sequence logo comparation of γ- and δ-module $KR_C$ based on the classification of their products. Top five rows show KRs associated with an inactive DH that produces the hydroxy products. The key catalytic residues are marked by red stars, and the NADPH-binding residues (partially) are marked by blue triangles. The motif numbers at the top are corresponding to the location of fingerprints in Figure 3. For full-length sequence logo, see Figure S5 in Supporting Information File 1.

dicted with high accuracy, especially for the stereochemistry at the α-position. Second, phylogenetic analysis of $KR_C$ subdomain is an effective means to predict stereoselectivity of KRs from actinobacterial β-modules. Specifically, our results indicate that the stereochemistry can be predicted well using the fingerprint motifs if the KR falls within the A- and B-type clades by phylogenetic analysis of $KR_C$ subdomain (and most of actinobacterial β-module KRs do fall within there). Overall, the motifs having correlation with stereoselectivity are summarized in Table S1 in Supporting Information File 1. Below we provide a detailed guideline for stereochemical prediction of KRs.

For non-actinobacterial PKSs, stereochemistry can be predicted by the presence of LDD motif for B-type. The absence of the LDD and the presence of the W motif are indicative for A-type, but with moderate accuracy (Figure 5).

For actinobacterial β-module KRs, stereochemistry can be predicted by using the LDD and W, as well as the newly identified R/Q (3), T (5), and D (6) motifs as summarized in Figure 5. About 87% to 98% of the KRs belonging to each product type satisfied our prediction criteria, verifying the accuracy of stereochemical prediction within this group of KRs.

In contrast, for actinobacterial γ- and δ-module KRs, the newly identified motifs cannot guide the prediction, and only A- and B-type can be roughly predicted by the presence and absence of the LDD motif. However, phylogenetic analysis can support bioinformatic prediction if a KR falls within A1/A2 or B1/B2 clades: for example, fostriecin KR2, which produces a *cis* double bond together with a DH, locates in the A1-type clade, in agreement with its stereoselectivity [35].

Lastly, we applied our prediction criteria for the previously mispredicted or unpredictable β-module KRs, for example due to having both LDD and W motifs, for validation. Among nine KRs we analyzed, five can be accurately predicted by either conserved motif analysis or by phylogenetic analysis, illustrating the advantage of using multiple strong predictive motifs (Table S2 in Supporting Information File 1). It is also worth noting that some of the KRs analyzed here may contain wrong stereochemical assignment of the product, potentially derived from misinterpretation of NMR analyses [14,36,37]. Nevertheless, many of the previously reported unpredictable KRs are from non-actinobacteria or from γ- and δ-module KRs, and bioinformatic prediction of these KRs would require further studies and more characterized sequences.

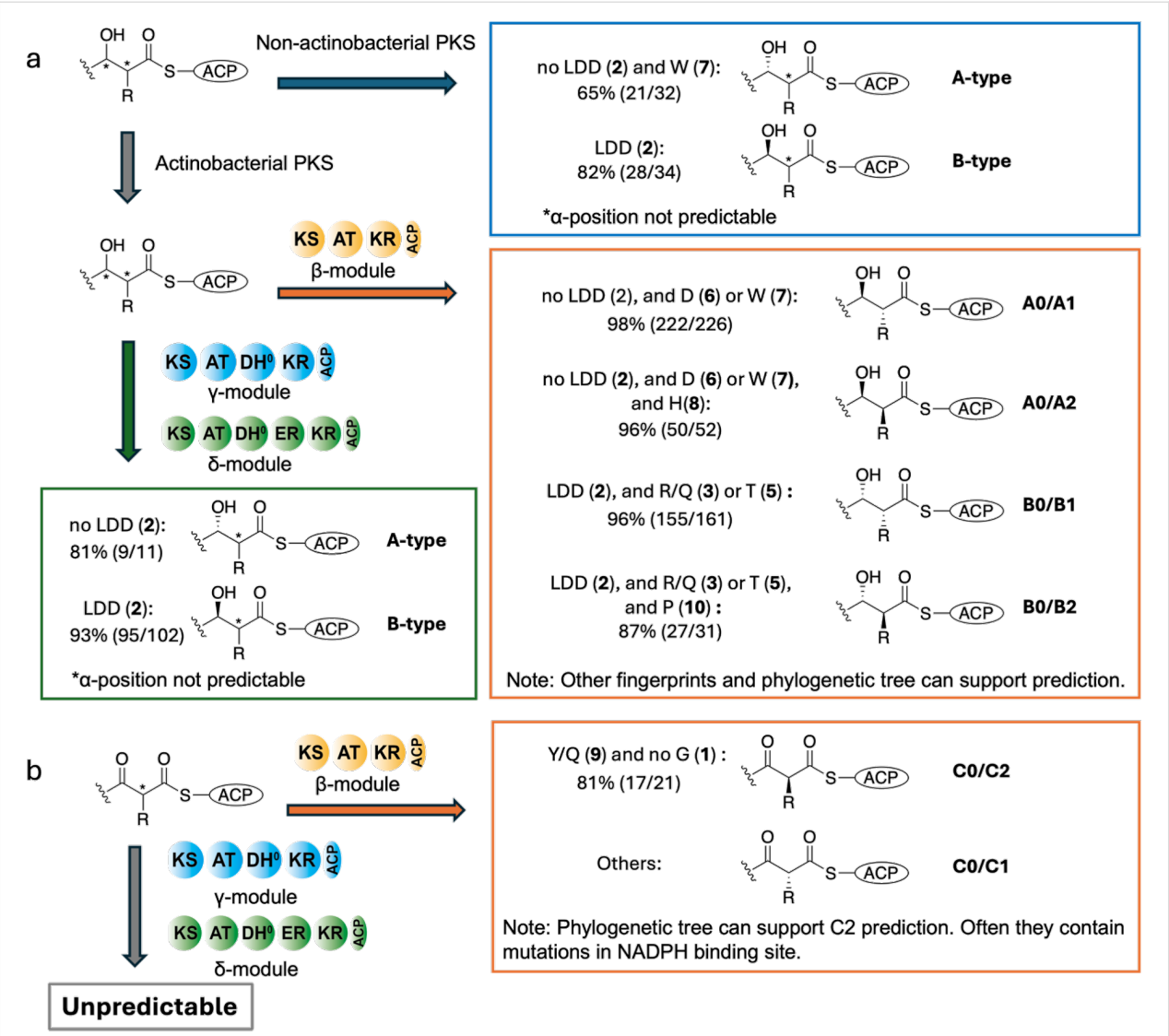



**Figure 5:** Summary of the updated fingerprints sorted by the taxonomic origin and the module type. Percentage numbers show KRs meeting the fingerprint description in our curated dataset. (a) Motifs useful for the stereochemical prediction of β-hydroxyl products. (b) Motifs useful for the stereochemical prediction of α-methyl,β-keto products.

## Conclusion

In summary, we systematically assessed the stereoselectivity-associated sequence motifs from 1,762 KR sequences in bacterial modular *cis*-AT PKSs whose product structures were experimentally determined. Our analyses revealed that different KRs, by taxonomic origin or by module types, have differences in the fingerprint motifs, affecting the prediction accuracy by conserved motif analysis. We identified several additional fingerprint motifs in the $KR_C$ subdomain that can be used to better predict the stereochemistry of KRs from actinobacterial β-modules. The identified motifs reside at the interface of domain–domain or domain–substrate interactions, shedding light on the enzyme mechanism for stereocontrol. Our work provides an overview for current bioinformatic prediction of stereoselectivity of KR in *cis*-AT PKS, expands the understanding of the stereocontrol of PKS from bioinformatics perspective, and will facilitate accurate stereochemical prediction and genome mining of complex polyketides.

## Experimental

### Sequence collection and screening

All *cis*-AT PKS and PKS-NRPS amino acid sequences recorded in MIBiG, as well as the PKS sequences reported in literatures we have curated were targeted for further analysis. Sequences with multiple modules on a peptide chain were cut into single-module sequences based on the boundaries annotated of modules. These sequences were then filtered according to the following rules: (a) The PKS sequences that produce the

same final product but originate from different strains of bacteria only retain the sequences from one of the sources. For example, only the erythromycin PKS from *Saccharopolyspora erythraea* were retained. (b) To associate sequence information with reliable absolute configuration of the product, the stereochemical configurations of collected polyketide products were manually checked by literature searches. Absolute configurations determined entirely by chemical methods, such as crystal structures, NMR, chemical degradation and derivatization, were considered reliable. Alternatively, relative configurations determined by NMR methods that corresponded exactly to the results predicted by bioinformatics were also considered reliable. Compounds of which only the relative configurations were elucidated were excluded from the dataset. (c) Sequences for which it was impossible to infer the stereochemistry of KR product were removed, such as M1 to M3 of rifamycin PKS whose product experiences later aromatization. (d) Modules with unconventional module compositions (e.g., modules with two KR domains), modules without KR domains, and loading modules were excluded from the data. After organizing and filtering, the $KR_S$ and $KR_C$ subdomain sequences were extracted for further bioinformatic analysis.

### Bioinformatic analysis

The multiple sequence alignment, sequence logo analysis, and tree building were performed on Geneious Prime 2023.2 (https://www.geneious.com) with options shown below: multiple sequence alignments were generated by MAFFT 7.450 (E-INS-i option), Neighbor-Joining trees were built by Geneious Tree Builder using Jukes-Cantor distance model. Sequence logo visualization was conducted by WebLogo 3. Protein structural analysis was conducted by PyMOL.

## Supporting Information

### Supporting Information File 1

Supporting Figures and Tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-131-S1.pdf]

### Supporting Information File 2

Supporting dataset.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-131-S2.xlsx]

## Acknowledgements

We thank the Zhang lab members Zilei Huang, Shengling Xie, Zhihan Zhang, Shanchong Chen, Cui Yang, Fanzhong Zhang, for KR sequence retrieval from public database.

## Funding

This research was supported by the “Pioneer” and “Leading Goose” R&D Program of Zhejiang (2023SDXHDX0007) and National Natural Science Foundation of China General Program (22177092) to L.Z.

## Author Contributions

Changjun Xiang: data curation; formal analysis; investigation; writing – original draft. Shunyu Yao: data curation. Ruoyu Wang: data curation. Lihan Zhang: conceptualization; data curation; formal analysis; funding acquisition; supervision; writing – review & editing.

## ORCID® iDs

Ruoyu Wang - https://orcid.org/0009-0004-3368-0265
Lihan Zhang - https://orcid.org/0000-0003-4364-6242

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## References

1. Hertweck, C. *Angew. Chem., Int. Ed.* **2009,** *48,* 4688–4716. doi:10.1002/anie.200806121
2. Keatinge-Clay, A. T. *Chem. Rev.* **2017,** *117,* 5334–5366. doi:10.1021/acs.chemrev.6b00683
3. Keatinge-Clay, A. T. *Nat. Prod. Rep.* **2012,** *29,* 1050–1073. doi:10.1039/c2np20019h
4. Grininger, M. *Nat. Chem. Biol.* **2023,** *19,* 401–415. doi:10.1038/s41589-023-01277-7
5. Keatinge-Clay, A. T. *Nat. Prod. Rep.* **2016,** *33,* 141–149. doi:10.1039/c5np00092k
6. Drufva, E. E.; Spengler, N. R.; Hix, E. G.; Bailey, C. B. *ChemBioChem* **2021,** *22,* 1122–1150. doi:10.1002/cbic.202000613
7. Weissman, K. J. *Beilstein J. Org. Chem.* **2017,** *13,* 348–371. doi:10.3762/bjoc.13.39
8. Caffrey, P. *ChemBioChem* **2003,** *4,* 654–657. doi:10.1002/cbic.200300581
9. Keatinge-Clay, A. T. *Chem. Biol.* **2007,** *14,* 898–908. doi:10.1016/j.chembiol.2007.07.009
10. Kitsche, A.; Kalesse, M. *ChemBioChem* **2013,** *14,* 851–861. doi:10.1002/cbic.201300063
11. Piasecki, S. K.; Zheng, J.; Axelrod, A. J.; Detelich, E. M.; Keatinge-Clay, A. T. *Proteins: Struct., Funct., Bioinf.* **2014,** *82,* 2067–2077. doi:10.1002/prot.24561
12. Robbins, N.; Spitzer, M.; Wang, W.; Waglechner, N.; Patel, D. J.; O'Brien, J. S.; Ejim, L.; Ejim, O.; Tyers, M.; Wright, G. D. *Cell Chem. Biol.* **2016,** *23,* 1383–1394. doi:10.1016/j.chembiol.2016.08.015
13. Zhang, L.; Hashimoto, T.; Qin, B.; Hashimoto, J.; Kozone, I.; Kawahara, T.; Okada, M.; Awakawa, T.; Ito, T.; Asakawa, Y.; Ueki, M.; Takahashi, S.; Osada, H.; Wakimoto, T.; Ikeda, H.; Shin-ya, K.; Abe, I. *Angew. Chem., Int. Ed.* **2017,** *56,* 1740–1745. doi:10.1002/anie.201611371

14. Kim, M. C.; Machado, H.; Jang, K. H.; Trzoss, L.; Jensen, P. R.; Fenical, W. *J. Am. Chem. Soc.* **2018,** *140,* 10775–10784. doi:10.1021/jacs.8b04848
15. Pérez-Victoria, I.; Oves-Costales, D.; Lacret, R.; Martín, J.; Sánchez-Hidalgo, M.; Díaz, C.; Cautain, B.; Vicente, F.; Genilloud, O.; Reyes, F. *Org. Biomol. Chem.* **2019,** *17,* 2954–2971. doi:10.1039/c8ob03115k
16. Kontou, E. E.; Gren, T.; Ortiz-López, F. J.; Thomsen, E.; Oves-Costales, D.; Díaz, C.; de la Cruz, M.; Jiang, X.; Jørgensen, T. S.; Blin, K.; Charusanti, P.; Reyes, F.; Genilloud, O.; Weber, T. *ACS Chem. Biol.* **2021,** *16,* 1456–1468. doi:10.1021/acschembio.1c00318
17. Carretero-Molina, D.; Ortiz-López, F. J.; Gren, T.; Oves-Costales, D.; Martín, J.; Román-Hurtado, F.; Sparholt Jørgensen, T.; de la Cruz, M.; Díaz, C.; Vicente, F.; Blin, K.; Reyes, F.; Weber, T.; Genilloud, O. *Org. Chem. Front.* **2022,** *9,* 462–470. doi:10.1039/d1qo01480c
18. You, Y.-O.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2013,** *135,* 7406–7409. doi:10.1021/ja4014776
19. Kishore, S.; Khosla, C. *Open Biol.* **2023,** *13,* 230096. doi:10.1098/rsob.230096
20. Terlouw, B. R.; Blin, K.; Navarro-Muñoz, J. C.; Avalon, N. E.; Chevrette, M. G.; Egbert, S.; Lee, S.; Meijer, D.; Recchia, M. J. J.; Reitz, Z. L.; van Santen, J. A.; Selem-Mojica, N.; Tørring, T.; Zaroubi, L.; Alanjary, M.; Aleti, G.; Aguilar, C.; Al-Salihi, S. A. A.; Augustijn, H. E.; Avelar-Rivas, J. A.; Avitia-Domínguez, L. A.; Barona-Gómez, F.; Bernaldo-Agüero, J.; Bielinski, V. A.; Biermann, F.; Booth, T. J.; Carrion Bravo, V. J.; Castelo-Branco, R.; Chagas, F. O.; Cruz-Morales, P.; Du, C.; Duncan, K. R.; Gavriilidou, A.; Gayrard, D.; Gutiérrez-García, K.; Haslinger, K.; Helfrich, E. J. N.; van der Hooft, J. J. J.; Jati, A. P.; Kalkreuter, E.; Kalyvas, N.; Kang, K. B.; Kautsar, S.; Kim, W.; Kunjapur, A. M.; Li, Y.-X.; Lin, G.-M.; Loureiro, C.; Louwen, J. J. R.; Louwen, N. L. L.; Lund, G.; Parra, J.; Philmus, B.; Pourmohsenin, B.; Pronk, L. J. U.; Rego, A.; Rex, D. A. B.; Robinson, S.; Rosas-Becerra, L. R.; Roxborough, E. T.; Schorn, M. A.; Scobie, D. J.; Singh, K. S.; Sokolova, N.; Tang, X.; Udwary, D.; Vigneshwari, A.; Vind, K.; Vromans, S. P. J. M.; Waschulin, V.; Williams, S. E.; Winter, J. M.; Witte, T. E.; Xie, H.; Yang, D.; Yu, J.; Zdouc, M.; Zhong, Z.; Collemare, J.; Linington, R. G.; Weber, T.; Medema, M. H. *Nucleic Acids Res.* **2023,** *51,* D603–D610. doi:10.1093/nar/gkac1049
21. Yin, Z.; Dickschat, J. S. *Nat. Prod. Rep.* **2021,** *38,* 1445–1468. doi:10.1039/d0np00091d
22. Gay, D.; You, Y.-O.; Keatinge-Clay, A.; Cane, D. E. *Biochemistry* **2013,** *52,* 8916–8928. doi:10.1021/bi400988t
23. Xie, X.; Garg, A.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2017,** *139,* 9507–9510. doi:10.1021/jacs.7b05502
24. McCullough, T. M.; Dhar, A.; Akey, D. L.; Konwerski, J. R.; Sherman, D. H.; Smith, J. L. *Structure* **2023,** *31,* 1109–1120.e3. doi:10.1016/j.str.2023.05.019
25. Keatinge-Clay, A. T.; Stroud, R. M. *Structure* **2006,** *14,* 737–748. doi:10.1016/j.str.2006.01.009
26. Zheng, J.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T. *Nat. Chem. Biol.* **2012,** *8,* 615–621. doi:10.1038/nchembio.964
27. Hirsch, M.; Fitzgerald, B. J.; Keatinge-Clay, A. T. *ACS Chem. Biol.* **2021,** *16,* 2515–2526. doi:10.1021/acschembio.1c00598
28. Piasecki, S. K.; Taylor, C. A.; Detelich, J. F.; Liu, J.; Zheng, J.; Komsoukaniants, A.; Siegel, D. R.; Keatinge-Clay, A. T. *Chem. Biol.* **2011,** *18,* 1331–1340. doi:10.1016/j.chembiol.2011.07.021
29. Liu, C.; Yuan, M.; Xu, X.; Wang, L.; Keatinge-Clay, A. T.; Deng, Z.; Lin, S.; Zheng, J. *J. Struct. Biol.* **2018,** *203,* 135–141. doi:10.1016/j.jsb.2018.04.001
30. Zheng, J.; Piasecki, S. K.; Keatinge-Clay, A. T. *ACS Chem. Biol.* **2013,** *8,* 1964–1971. doi:10.1021/cb400161g
31. Zheng, J.; Keatinge-Clay, A. T. *J. Mol. Biol.* **2011,** *410,* 105–117. doi:10.1016/j.jmb.2011.04.065
32. Xie, X.; Garg, A.; Keatinge-Clay, A. T.; Khosla, C.; Cane, D. E. *Biochemistry* **2016,** *55,* 1179–1186. doi:10.1021/acs.biochem.6b00024
33. Keatinge-Clay, A. *J. Mol. Biol.* **2008,** *384,* 941–953. doi:10.1016/j.jmb.2008.09.084
34. Kandziora, N.; Andexer, J. N.; Moss, S. J.; Wilkinson, B.; Leadlay, P. F.; Hahn, F. *Chem. Sci.* **2014,** *5,* 3563–3567. doi:10.1039/c4sc00883a
35. Shah, D. D.; You, Y.-O.; Cane, D. E. *J. Am. Chem. Soc.* **2017,** *139,* 14322–14330. doi:10.1021/jacs.7b08896
36. Zhang, S.; Zhang, S.; Fan, Y.; Zhang, X.; Chen, J.; Jin, C.; Chen, S.; Wang, L.; Zhang, Q.; Chen, Y. *Angew. Chem., Int. Ed.* **2023,** *62,* e202313186. doi:10.1002/anie.202313186
37. Ding, J.; Smith, A. B., III. *J. Am. Chem. Soc.* **2023,** *145,* 18240–18246. doi:10.1021/jacs.3c06573

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.131

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Mining raw plant transcriptomic data for new cyclopeptide alkaloids

Draco Kriger‡, Michael A. Pasquale‡, Brigitte G. Ampolini‡ and Jonathan R. Chekan*

Full Research Paper

Open Access

Address:
Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, NC, USA

Email:
Jonathan R. Chekan* - jrchekan@uncg.edu

* Corresponding author ‡ Equal contributors

Keywords:
burpitides; natural products; plants; RiPPs; transcriptome mining

*Beilstein J. Org. Chem.* **2024,** *20,* 1548–1559.
https://doi.org/10.3762/bjoc.20.138

Received: 16 March 2024
Accepted: 24 June 2024
Published: 11 July 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf



© 2024 Kriger et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

In recent years, genome and transcriptome mining have dramatically expanded the rate of discovering diverse natural products from bacteria and fungi. In plants, this approach is often more limited due to the lack of available annotated genomes and transcriptomes combined with a less consistent clustering of biosynthetic genes. The recently identified burpitide class of ribosomally synthesized and post-translationally modified peptide (RiPP) natural products offer a valuable opportunity for bioinformatics-guided discovery in plants due to their short biosynthetic pathways and gene encoded substrates. Using a high-throughput approach to assemble and analyze 700 publicly available raw transcriptomic data sets, we uncover the potential distribution of split burpitide precursor peptides in Streptophyta. Metabolomic analysis of target plants confirms our bioinformatic predictions of new cyclopeptide alkaloids from both known and new sources.

## Introduction

Plants are prolific producers of cyclic peptide natural products, making 1000s of different molecules [1]. While the orbitide [2] and cyclotide [3] classes of peptides are well known, it has been recently discovered that a new class of molecules called burpitides are also prevalent in plants [4]. Like all known plant peptides, burpitides fall under the ribosomally synthesized and post-translationally modified peptides (RiPPs) superclass of natural products (Figure 1). The typical pathway for a RiPP starts with a precursor peptide made by the ribosome that undergoes post-translational modifications by specific tailoring enzymes [5,6]. This precursor peptide substrate can be subdivided into multiple segments including 1) an N-terminal leader or recognition sequence used for binding by the tailoring enzymes and 2) a core peptide that is targeted for modification by the biosynthetic enzymes. Ultimately proteolysis releases the modified core peptide as the mature RiPP natural product [5,6].

In the case of the newly described burpitide family of RiPPs, the defining feature is the presence of amino acid side-chain crosslinks installed by a copper-dependent burpitide cyclase [4].

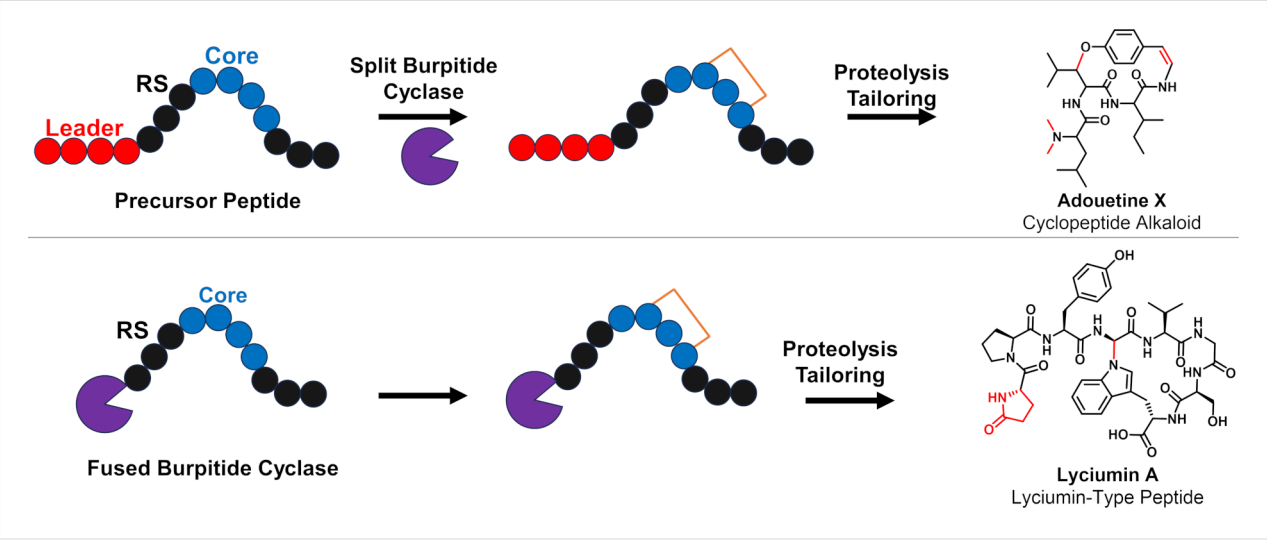



**Figure 1:** Biosynthetic scheme for the formation of burpitides using a split (top) or fused (bottom) pathway. RS is recognition sequence. Peptide modifications are shown in red.

These RiPP natural products encompass a wide range of structural scaffolds including cyclopeptide alkaloids that contain a phenolic ether-linkage and the lyciumin-type peptides that are composed of a crosslink between the Trp-indole-N and the carbon in another amino acid side chain or peptide backbone (Figure 1). Recent enzymatic reconstitution has demonstrated that the burpitide cyclases can function either autocatalytically (fused) or as traditional stand-alone proteins with separate free peptide substrates (split) [4,7-11]. When fully matured, burpitides possess a wide range of bioactivities including analgesic, sedative, and cytotoxic [12-14].

In order to identify new potential burpitides, we searched the NCBI Sequence Read Archive (SRA) for Streptophyta (all land plants and most green algae) transcriptomes that could be assembled and examined with a custom hidden Markov model (HMM) for split burpitide precursor peptides. Using this approach, we assembled 700 transcriptomes and used predicted precursor peptide diversity for target discovery of new cyclopeptide alkaloids. Additionally, metabolomic analysis revealed the tissue distribution of these molecules across select members of the Rubiaceae and Rhamnaceae families. Ultimately, our results lay the groundwork for the rapid discovery of new burpitides across diverse plant species.

## Results and Discussion

### Assembly and analysis of 700 public transcriptomes

In order to efficiently download, assemble, and analyze raw SRA data in a high throughput manner, we created an automated pipeline. Raw Illumina sequencing data were downloaded using the NCBI SRA Toolkit, the transcriptome was then assembled with rnaSPAdes [15], and the open reading frames were detected with TransDecoder [16,17]. To search for putative precursor peptides for split burpitide pathways, we developed a custom HMM (see Supporting Information File 1, Figure S1). This approach gave us the speed, sensitivity, and customization needed to identify diverse sequences. We targeted a diversity of plants from every available order in Streptophyta for a total of 700 transcriptomes from 647 species (Table in Supporting Information File 2).

Potential split burpitide precursor peptide transcripts were found in 66% of the assemblies using the custom HMM. However, due to the only recent discovery of split burpitides, the rules that dictate the presence of a candidate precursor peptide and observation of a corresponding small molecule are unclear. For example, precursors rich in S/TxY core motifs are prevalent in nature, but do not correspond to a known molecule [7]. Therefore, we sought to rapidly identify the most promising candidates. This presented a challenge as the publicly available data was generated using different protocols, from different plant tissues, and with different levels of sequencing depth. Ultimately, we reasoned that the best candidate plants for prolific production of burpitides should have numerous unique precursor peptide transcripts which may arise from multiple gene copies, multiple core sequences, or multiple RNA splicing events. Therefore, we counted the number of translated transcripts from each assembly that scored better than an E-value of 0.1 using our custom precursor peptide HMM. To select for only the split precursor peptides, we removed any transcript that had contained a BURP domain as determined by the public HMM model (PFAM 03181) with an E-value of 0.1 inclusion

threshold. The BURP domain itself is named after the four founding members BNM2, USP-like, RD22, and PG1β, which were bioinformatically described in 1998 [18]. This domain is typically around 300 amino acids in length and has a conserved $CHX_{10}CHX_{25\text{-}27}CHX_{25\text{-}26}CH$ motif which comprises the active site of the burpitide cyclase. The resulting number of stand-alone transcripts from this filtering step were mapped onto a cladogram of the 647 species we surveyed (Figure 2 and Supporting Information File 3). For species with multiple transcriptomes, the number of precursor peptides was averaged (Table in Supporting Information File 2). We further mapped known producers with a corresponding precursor peptide onto

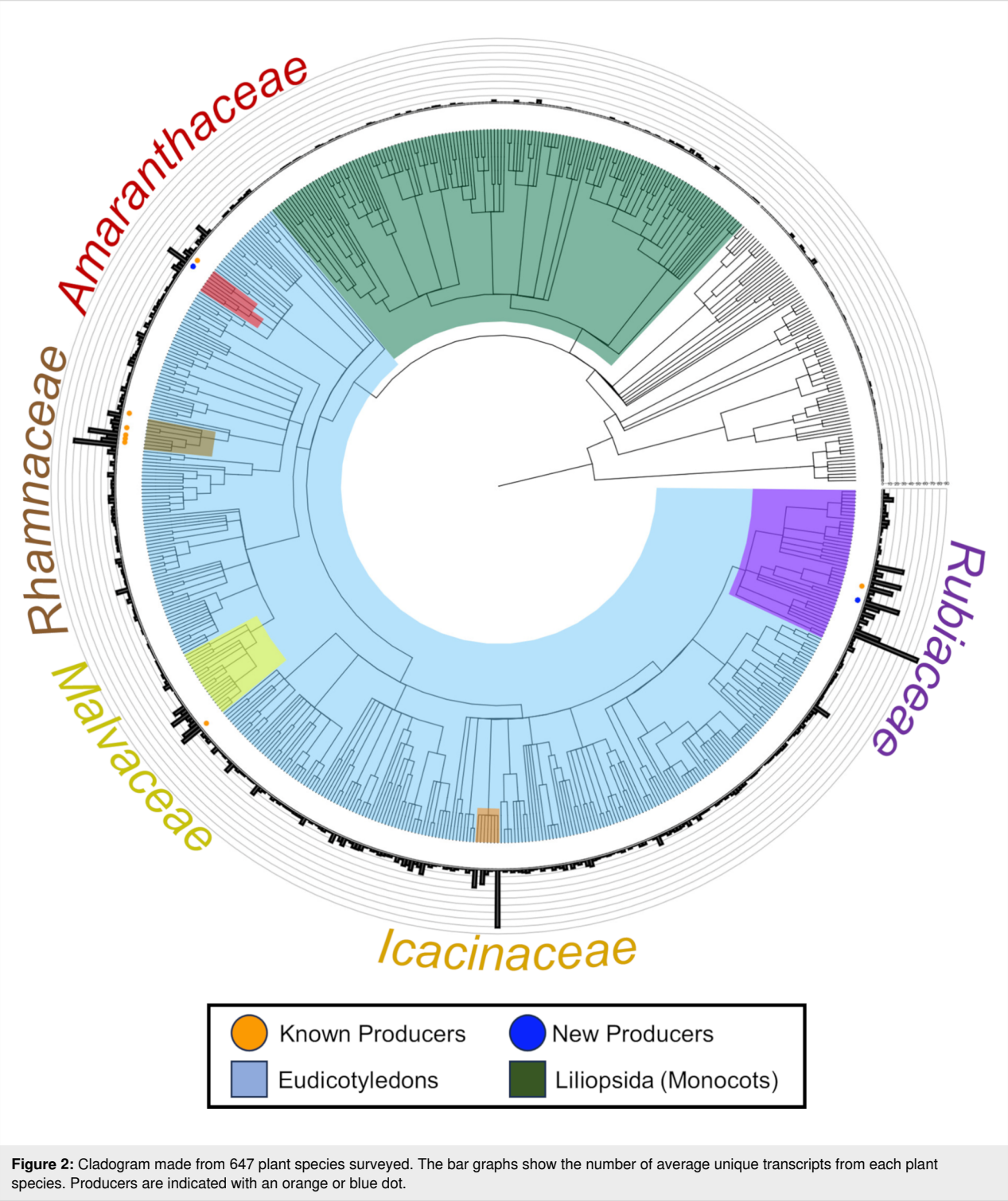



**Figure 2:** Cladogram made from 647 plant species surveyed. The bar graphs show the number of average unique transcripts from each plant species. Producers are indicated with an orange or blue dot.

the cladogram (orange dots). Using a cut-off of 30 unique transcripts, there were five clear phylogenetic hot spots for the production of burpitides. These families were examined in detail for their composition of core peptide sequences.

## Rhamnaceae family

The Rhamnaceae family is amongst the most iconic producers of cyclopeptide alkaloids, containing both *Ziziphus* and *Ceanothus* genera. In addition to the subclass defining Tyr-phenol-O to carbon linkage, the cyclopeptide alkaloids of this family are typically oxidatively decarboxylated and *N*-methylated (Figure 1, adouetine X and Figure 3, ceanothine B) [19-21]. Our transcriptome analysis supports the prevalence of cyclopeptide alkaloids in Rhamnaceae with multiple species transcribing 20+ potential split precursor peptides. Alignment of the predicted recognition and core sequences shows the core peptides are 4–5 amino acids in length and appear to exclusively encode for cyclopeptide alkaloids due to the invariant C-terminal Tyr residue (Figure 3).

## Rubiaceae family

Our previous results revealed *Coffea arabica* (Arabica coffee plant), to be a cyclopeptide alkaloid producer [7]. In contrast to those from the Rhamnaceae family, these cyclopeptide alkaloids were only cyclized and lack an oxidative decarboxylation of the C-terminus (Figure 3, arabipeptin A). Our transcriptome

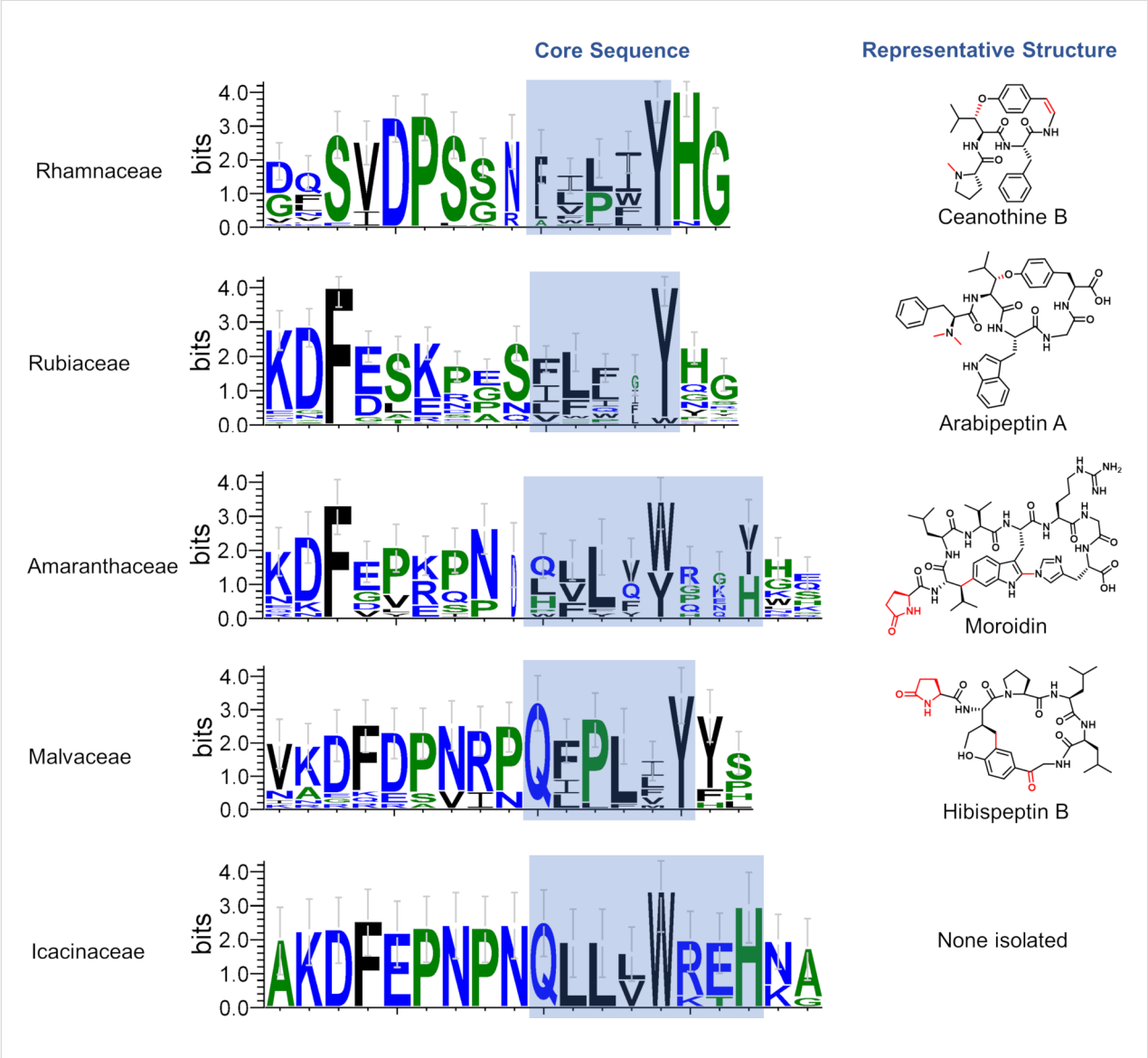



**Figure 3:** Weblogos generated with aligned recognition and core sequences from the six different families discussed. Up to three unique sequences were taken from each assembly. See Supporting Information File 2 for exact sequences identified and used. The highlighted region is the predicted core peptide motif.

mining results further support the prevalence of cyclopeptide alkaloids in the Rubiaceae family. *Chiococca alba* (Snowberry), *Cinchona calisaya* (Yellow Cinchona), *Gardenia jasminoides* (Cape Jasmine), *C. arabica*, and *Coffea eugenioides* (Eugenioides coffee plant) all contained >30 unique transcripts for potential precursor peptides. The core sequences themselves were typically 4 or 5 amino acids in length and most often contained a tyrosine for cyclization. However, a few sequences appeared to have a tryptophan at the position for cyclization, indicating they may correspond to the stephanotic acid-type burpitides like moroidin (Trp-indole-C to carbon crosslink, Figure 3) [4].

## Amaranthaceae family

The Amaranthaceae family is home to the known moroidin producer, *Celosia argentea* var. *cristata* (Cockscomb and previously *Celosia cristata*) (Figure 3) [22]. Transcriptome analysis revealed that many other species in this family and the recently merged Chenopodioideae subfamily [23] have the capacity to produce burpitides as well. For example, *Alternanthera bettzickiana* (Calico Plant) was rich in 31 potential unique precursor peptide sequences. Examination of possible cores suggested that the producers closely related to *C. argentea* (*A. bettzickiana* and *Amaranthus tricolor* [edible amaranth] are likely to make bicyclic burpitides but may vary in the cycle size and identity of the cyclized residues (Figure 3 and Supporting Information File 2).

## Malvaceae family

The Malvaceae family contains *Hibiscus syriacus* (Rose of Sharon), the only known producer of hibispeptin-type burpitides (Figure 3, hibispeptin B) [24,25]. These molecules contain a C–C linkage between the phenol derived from tyrosine and the γ-carbon of isoleucine. As expected from previous genomic analysis [7], the transcriptome of *H. syriacus* contained precursor peptides corresponding to both hibispeptin A and B (Figure 3 and Supporting Information File 2). The closely related *Hibiscus cannabinus* possessed similar transcripts, while *Hibiscus mutabilis* (Confederate rose) appeared to code for core peptides that were only four amino acids in length and not clearly hibispeptins. Outside of the *Hibiscus* genera, *Quararibea asterolepisn, Kosteletzkya pentacarpos* (Coastal Mallow), and *Taliparitі hamabo* all contained core peptides indicating hibispeptin-type burpitide production.

## Icacinaceae family

In addition to these known families of producers, our transcriptomic results suggest production of burpitides from the Icacinaceae family. In particular, *Merrilliodendron megacarpum* produced 81 unique transcripts, while *Pyrenacantha malvifolia* (Monkey chair) produced more than 20. Analysis of the predicted core sequence seemed to suggest moroidin-like molecules (Figure 3), but no cyclic peptides have been isolated from any member of this family. Unfortunately, sourcing members of this small family proved challenging and we were unable to evaluate these hypotheses.

## Iconic moroidin producers contain split precursor peptides

The moroidin family of plant peptides has previously been shown to be biosynthesized from a fused burpitide pathway using a transcriptome mining approach [9]. However, precursor peptides for the well-known producers of moroidin (*Dendrocnide moroides* [Gympie stinger] and *C. argentea*) [22,26] and the structurally related celogentin molecules (*C. argentea* var. *cristata*) [27,28] remained elusive. Using the publicly available sequencing data from *D. moroides* and *C. argentea* var. *cristana*, we could identify core sequences that perfectly matched the structure of moroidin (Figure 4). Furthermore, manually searching for the "QLLVWRGH" core sequence in the full transcriptome only revealed the stand-alone precursor peptides, further suggesting the biosynthetic pathways contain split burpitide cyclases in these plants. This seems to suggest that both split and fused burpitide biosynthetic pathways may be responsible for producing the stephanotic acid sub-family of RiPPs that is characterized by the Trp-indole-C to carbon linkage [4].

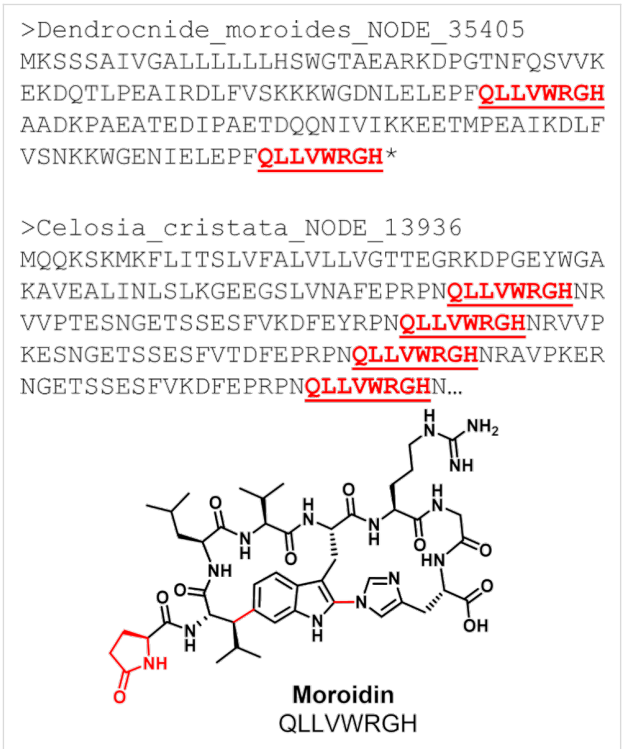



**Figure 4:** Core peptide sequences from the putative precursor peptides map to moroidin.

## Molecular networking of known and potential producers

The transcriptome analysis indicated that many additional plant species may be responsible for the production of burpitides. To evaluate this, we examined *G. jasminoides* and *A. bettzickiana* from the Rubiaceae and Amaranthaceae families, respectively. We analyzed methanol extracts of these plants along with the known cyclopeptide alkaloid producer *C. ameranicus* using UHPLC-HRMS/MS. Finally, we generated a global natural product social (GNPS) network to show the correlation between metabolites from these different plant species (Figure 5) [29].

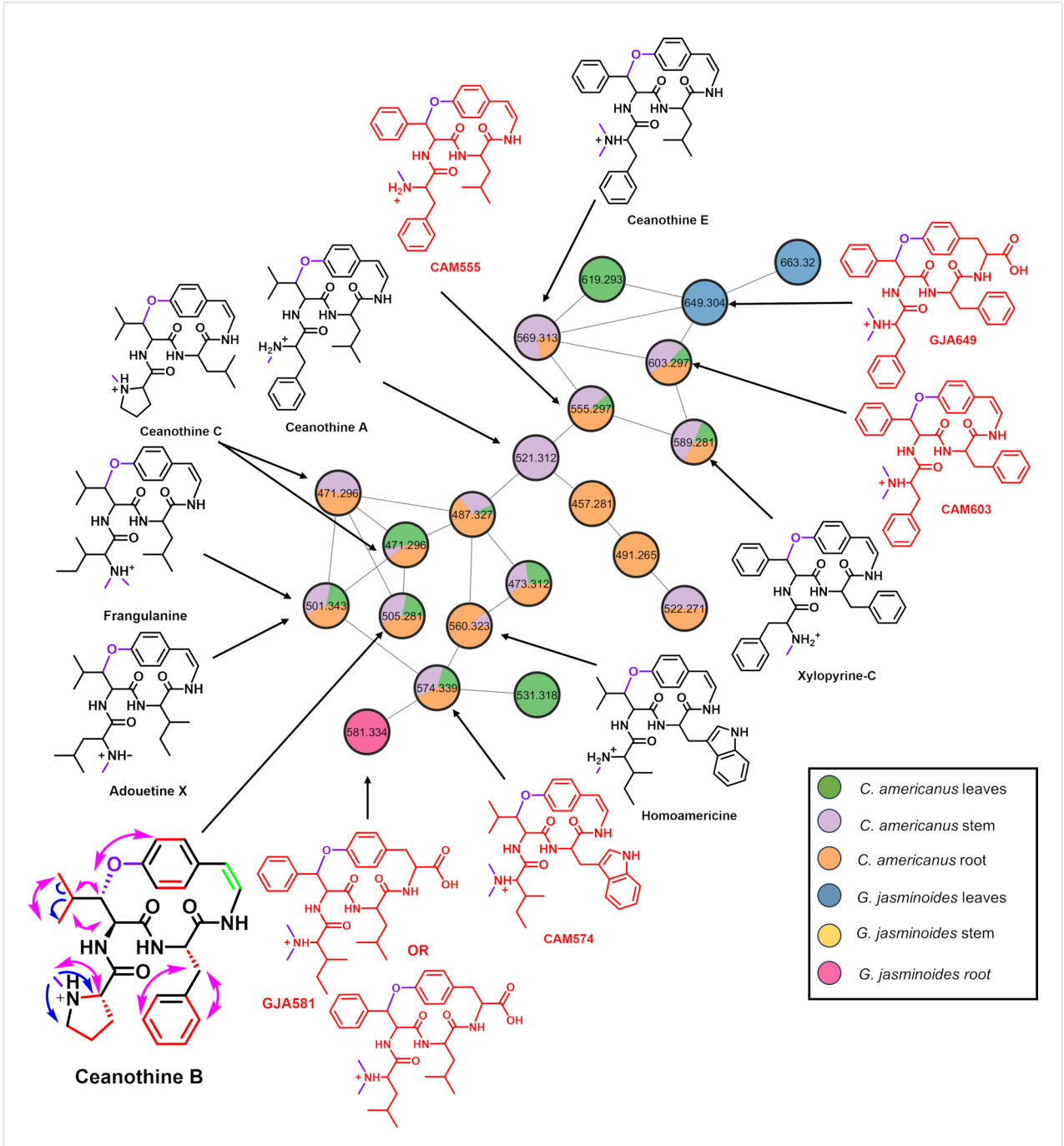



**Figure 5:** GNPS network created from extracts of *C. americanus* and *G. jasminoides*. The nodes are color-coded to show localization of various molecules. Molecules in red are new to this study and predicted by MS/MS fragmentation. Ceanothine B was isolated and its structure confirmed by 1D and 2D NMR. Correlations are indicated as follows: red lines – COSY, green lines – TOCSY, pink arrows – NOSEY, blue arrows – HMBC. Peptide modifications are shown in purple.

### Cyclopeptides in *Ceanothus americanus*

The GNPS network showed the presence of multiple features from *C. americanus* that appeared to match known molecules such as ceanothine B (505.2809 *m/z*), ceanothine C (471.2966 *m/z*), ceanothine A (521.3122 *m/z*), ceanothine E (569.3128 *m/z*) fragulanine/adouetine X (501.3435 *m/z*), and homoamericine (560.3231 *m/z*). Despite being first isolated over 50 years ago [30-32], the majority of the proposed structures for these known cyclopeptide alkaloids are simply predicted by MS/MS fragmentation. Therefore, we isolated ceanothine B and completed full structural characterization using NMR, Marfey's analysis, and MS/MS fragmentation (Supporting Information File 1, Figure S2, Figures S16–S26, Table S1). Indeed, our results support the proposed structure of ceanothine B while also assigning absolute stereochemistry consistent with all L-amino acids (Figure 5).

We next explored the GNPS network for new cyclopeptide alkaloids from *C. americanus*. We previously noted the presence of a feature as CAM603 (603.29 *m/z*) [7]. Fragmentation indicated a core sequence of FFFY that was observed in precursor peptides found in the transcriptome (Supporting Information File 1, Figure S3). To further support this assignment, we partially purified the compound and completed Marfey's analysis, which confirmed the presence of all L-Phe in the structure (Supporting Information File 1, Figure S27). Beyond this, multiple other features in the GNPS network appeared to represent new molecules for *C. americanus* (Supporting Information File 1, Figures S4–S11). Xylopyrine-C was previously isolated from *Zizyphus xylopyra* [33] and matches the exact mass and fragmentation of the 589.2820 *m/z* feature. The 555.2953 *m/z* feature (CAM555) appears to be a mono-methylated version of ceanothine E. Likewise, the 574.3396 *m/z* (CAM574) feature is consistent with a methylated analogue of homoamericine.

### New cyclopeptide alkaloids in *Gardenia jasminoides*

The GNPS network also suggested the presence of new cyclopeptide alkaloids from *G. jasminoides*, a previously unknown producer (Figure 5). This data corroborates the transcriptome mining data where *G. jasminoides* was shown to possess an average of 50 unique precursor peptide transcripts. The 649.3047 *m/z* feature, which we named GJA649, had an exact mass and fragmentation suggested a molecule with an FFFY core (Supporting Information File 1, Figure S12). Indeed, the transcript data assembled as part of this study identified multiple transcripts from *G. jasminoides* that directly match this FFFY sequence (Figure 6A). Like other members of the Rubiaceae family [7,34], the tyrosine that forms the ether linkage remains intact, and not oxidatively decarboxylated as seen in members of the Rhamnaceae family. It is also predicted to contain a dimethylation at the N-terminal Phe residue. Additional 663.3192 and 581.3345 *m/z* features were also present in the GNPS network, but MS/MS fragmentation could not definitively support a specific structure.

For a better understanding of the biosynthesis of burpitides in *G. jasminoides*, we explored the genomic context by using the

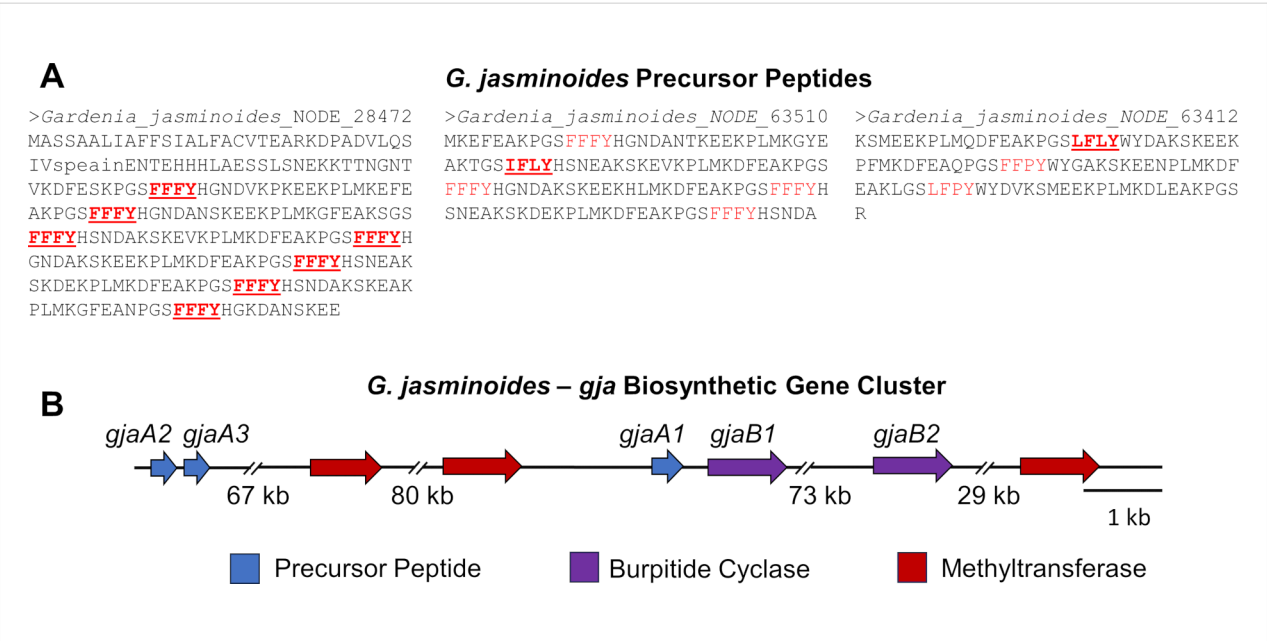



**Figure 6:** A) Precursor peptide sequences found in the transcriptome of *G. jasminoides* containing the cores FFFY, IFLY, and LFLY. A) GJA649 gene cluster from *G. jasminoides*.

burpitide cyclase sequence and precursor peptides found during transcriptome assembly by performing a tblastn search against the unannotated *G. jasminoides* genome (NCBI GenBank assembly: GCA_013103745.1). Gene matches were primarily found on chromosome 4 (NCBI GenBank assembly: CM023098.1) with the location of the biosynthetic gene cluster (BGC) on the chromosome between base pairs 8,210,000 and 8,480,000. The BGC contained precursor peptides with the core FFFY (*gjaA1)*, FLFY (*gjaA2)*, and FLLY (*gjaA3),* three potential methyltransferases, and two complete burpitide cyclases all within 266,286 base pairs (Figure 6B). While the proposed gene cluster is missing a potential peptidase, it contains all the other proteins necessary for the biosynthesis of GJA649 and the other putative *G. jasminoides* cyclopeptide alkaloids.

## New cyclopeptide alkaloids in *Alternanthera bettzickiana*

Based on the assembled transcriptomic data sets, it appeared that *A. bettzickiana* had the capacity to produce new cyclopeptide alkaloids. To evaluate this, a methanol extract of *A. bettzickiana* was prepared and analyzed by UHPLC-HRMS/MS. Subsequent GNPS networks did not show any clustering with cyclopeptide alkaloids from either *C. americanus* or *G. jasminoides*. Instead, the MS/MS data was searched for fragmentation patterns representing peptides, namely iminium ions [8]. Both the exact mass and fragment ions show evidence for three potential compounds with alpha-keto groups that matched putative core sequences (Figure 7 and Supporting Information File 1, Figures S13–S15). While unusual, this motif could be biosynthetically related to the alpha-hydroxy modification seen in vignatic acid [35]. These potential compounds are bicyclic which resemble the bicyclic motif found in moroidin from the closely related *C. argentea* (Figure 4). Unlike moroidins which are part of the stephanotic acid subclass of burpitides defined by the Trp-indole-C to carbon crosslink, the molecules from the *A. bettzickiana* seems to be cyclopeptide alkaloids with two tyrosine derived ether linkages that resemble selanine A isolated from African clubmoss (*Selaginella kraussiana*) [8]. To confirm these proposed structures from *A. bettzickiana*, full elucidation by NMR will be required.

## Localization of cyclopeptide alkaloids in plants

Traditionally, cyclopeptide alkaloids have been isolated from the roots and bark of the producing plants [1,19-21], likely due to the abundance of these materials. With more sensitive metabolomics techniques, we can better identify where these molecules are located in the plant. For the described GNPS network, live samples of *C. americanus* and *G. jasminoides* were processed by separating the leaves, stems, and roots. With few exceptions, cyclopeptides alkaloids from *C. americanus* were found distributed throughout multiple tissues in the plant and not only localized to the roots (Figure 5). In contrast, the cyclopeptide alkaloids in G. *jasminoides* appear to be exclusively present in either the leaves or root. As the biological functions of these molecules are currently unknown, their localization may give insights into their physiological purpose.

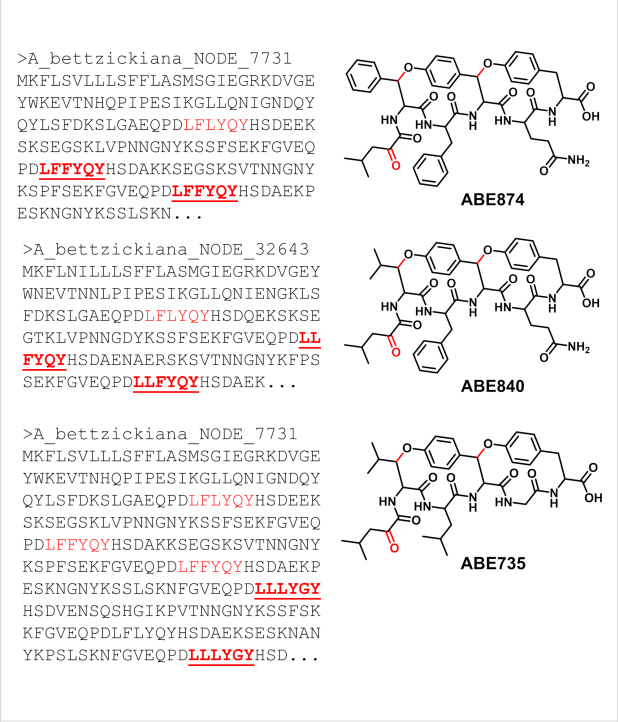



**Figure 7:** Core sequences from the putative precursor peptides map to three predicted products from *A. bettzickiana* observed by LC–MS/MS.

# Conclusion

Overall, this study illustrates the ability to rapidly assemble and analyze public transcriptomic data for the targeted discovery of new burpitide natural products from plants. Phylogenetic visualization of the potential precursor peptides validated families known to make these cyclic peptides, but also suggested new producers. Finally, metabolomic analyses supported the predicted new sources of burpitides, aiding in our understanding of the precursor rulesets for the biosynthesis of this emerging class of plant natural products.

# Experimental

## Construction of precursor peptide HMM

A custom hidden Markov model designed to target the precursor peptides from split burpitide pathways was created to favor producers of known burpitides. Sequences containing cores matching known molecules such as hibispeptins, moroidins,

jubanines, and lyciumins were selected for the new model (see Supporting Information). The sequences were aligned and the hmm build function of HMMER v3.3.2 was used to generate the model [36]. The HMM is available in Supporting Information File 4 compressed archive.

## Transcriptome mining and cladogram construction

Representatives of every available order of Streptophyta were selected from the NCBI Sequence Read Archive. The full list of accessions and species is available in the Supporting Table (Supporting Information File 2). rnaSPAdes was used to assemble the raw reads using the default settings [15]. This software package was chosen primarily for its speed (≈1.5 h per transcriptome), which was essential for this large analysis effort. Our custom HMM was used to search for potential precursor peptides using an E-value inclusion threshold of 0.1. These results are present in Supporting Information File 4 compressed archive file. To remove fused burpitide cyclases, the HMM for the BURP domain (PFAM 03181) was also run against the assembled transcriptomes using the same 0.1 E-value inclusion threshold. Transcriptomes that indicated a BURP domain were removed from the putative precursor peptide list. The resulting precursor peptides were counted and mapped as bar graphs onto a cladogram of the species phylogeny. phyloT was used to generate the cladogram and R-studio and ggplot2 were used to visualize it with the number of transcripts [37,38]. For our analysis, the previously distinct Chenopodioideae family was viewed as members of the Amaranthaceae family based on the most up to date Angiosperm Phylogeny Group classification (APG IV) [23].

## Isolation of burpitides from *C. americanus*

Three separate extractions were performed overnight with methanol using 200 g of ground root powder (Mountain Rose Herb Lot# 26356) and 1.5 L of methanol. After overnight extraction, vacuum filtration was used to separate the solids from the liquids. The methanol extract was dried and saved. The solids were extracted with methanol two more times. After the extractions were finished and dried by rotary evaporation, the extract was resuspended in basic water (pH 10 using NaOH) and vacuum filtered to remove insoluble particulates. Liquid–liquid extraction was performed three times with basic water and DCM. The DCM layer was collected and dried. Afterward, the extract was resuspended in 0.1% formic acid in water.

A Combiflash EZ prep system was used to purify the burpitides from *C. americanus*. Water with 0.1% formic acid (A) and methanol (B) were used as solvents for a reversed phase column (C18 21 × 250 mm Kinetix 5 µm C18 100 Å (00G-4601-P0-AX5) with a 30–80% B gradient over 50 min at 10 mL/min. Roughly 100 mg of sample was loaded using 5 mL of a 25% MeOH/$H_2O$ solution. The fractions were tested using UHPLC-HRMS/MS for purity and presence of desired burptidies. Ceanothine B was successfully isolated using this approach. For CAM603, an additional purification step using an HPLC system (Agilent) with a C18 column (C18 250 × 4.6 mm Luna Omega 5 5 µm C18 100 Å) and gradient of 55–70% B was employed.

## Structure elucidation of ceanothine B

Ceanothine B was analyzed by 1D and 2D NMR in methanol-$d_4$ in 5 mm NMR tubes (Wilmad LabGlass). NMR instrumentation was a JEOL ECA-500 MHz NMR spectrometer (JEOL Ltd.). $^1$H NMR (500 MHz, methanol-$d_4$) δ 7.17 (m, 2H), 7.12 (m, 2H), 7.09 (m, 2H), 7.04 (m, 1H), 7.01 (m, 1H), 6.98 (m, 1H), 6.96 (m, 1H), 6.73 (d, $J$ = 7.5 Hz, 1H), 6.19 (d, $J$ = 7.0 Hz, 1H), 4.80 (dd, $J$ = 2.0 Hz, $J$ = 8.0 Hz, 1H), 4.42 (d, $J$ = 8.0 Hz, 1H), 4.42 (m, 1H), 3.05 (m, 1H), 2.85 (d, $J$ = 13.5 Hz, 1H), 2.74 (dd, $J$ = 4.0 Hz, $J$ = 10.0 Hz, 1H), 2.59 (t, $J$ = 12.5 Hz, 1H), 2.33 (m, 1H), 2.16 (m, 1H), 1.97 (s, 1H), 1.95 (m, 1H), 1.79 (m, 1H), 1.68 (m, 2H), 1.15 (d, $J$ = 7.0 Hz, 1H), 0.926 (d, $J$ = 7.0 Hz, 1H)

Ceanothine B macrocyclization between Tyr-3 and Leu-β is supported by HMBC correlation between Tyr-C3 (δ 156.3 ppm) and Tyr-C4 (δ120.5 ppm) and NOESY correlation between Tyr4-H (δ 7.01 ppm, 1H) and Leuβ-H (δ 4.80 ppm, 1H). Similar to other characterized cyclopeptide alkaloids, $J$-coupling constant ($J_{\alpha\text{-}\beta}$ = 8.0 Hz) supports *S*-stereochemistry at the β-carbon of leucine ether linkage. Structurally related cyclopeptide alkaloids have similar $J$-values ($J_{\alpha\text{-}\beta}$ ≈ 8.0 Hz) and have been assigned as L-erythro (*anti*) conformation [20,39].

## Metabolomic analysis and GNPS networking

Live samples of *C. americanus* and *G. jasminoides* were separated into leaves, stems, and roots and lyophilized for extraction. *G. jasminoides* samples were prepared by extracting 89.24 g of dried leaf material over three days with 1 L of methanol each day yielding 19.10 g of crude extract. The root and stem extracts were prepared by extracting 180.6 mg and 145.2 mg of plant material, respectively, in methanol overnight. *A. bettzickiana* was lyophilized as the whole plant. Each sample was extracted separately in methanol overnight and the extract was dried by rotary evaporation. All the residues were diluted in methanol to 2 mg/mL, except for the *G. jasminoides* leaves (1.5 mg/mL). UHPLC-HRMS/MS data was collected using an Q Exactive Plus (Thermo Scientific) coupled to a Acquity UPLC (Waters), and Acquity UPLC BEH 1.7 µm C18

reversed phase 130 Å 2.1 × 50 mm column. The following gradient was used at 1 mL/min: 0 min, 5% B; 1 min, 5% B; 11 min 100% B; 12 min, 100% B; 12.1 min, 5% B; 13 min, 5% B where solvent A was water with 0.1% of formic acid and solvent B was acetonitrile with 0.1% of formic acid. A top 5 HCD method at 25 NCE (200–2000 *m*/*z*) was used on each plant tissue extract. See Supporting Information File 1, Table S2 for complete acquisition parameters. The mass spectra were used to generate a GNPS network using the following parameters: 0.2 Da precursor ion mass tolerance, 0.5 Da fragment ion tolerance, and a minimum of 8 matched fragment ions.

## Marfey's analysis

To determine the stereochemistry of the new molecules, Marfey's analysis was used. Standards were prepared to match the core amino acids of the predicted ceanothine B and CAM603 cyclopeptide alkaloids. To make each amino acid standard, 0.2 mg of the L- and D-amino acid was aliquoted into separate reaction vials. To each vial, 50 µL of water, 20 µL of 1 M $NaHCO_3$, and 100 µL 1% Marfey's reagent (sodium (2,4-dinitro-5-fluorophenyl)-L-alaninamide) in acetone were added. The reaction mixture was incubated at 40 ºC for 1 h with periodic agitation. The reactions were quenched by adding 10 µL of 2 M HCl and dried under a nitrogen stream. Dried samples were dissolved in 1.7 mL of MeOH and injected both individually and as a combined mixture (5 µL) onto UPLC-MS (Thermo Orbitrap Elite mass spectrometer coupled to a Dionex Ultimate 3000 UPLC), with a gradient of 15–50% B over 10 min on a Kinetex 1.7 µm C18 100 Å 50 × 2.1 mm LC column using solvents of A: 0.1% formic acid in water and B: 0.1% formic acid in acetonitrile. Products were observed at 340 nm and monitored by their expected $[M + H]^+$ ion.

To hydrolyze the isolated burpitides, 0.2 mg of sample was aliquoted into reaction vial. To each vial, 0.5 mL of 6 M HCl was added. The sample was incubating for 24 h at 90 ºC and dried under a nitrogen stream. Afterward, 25 µL of water, 10 µL of 1 M $NaHCO_3$ and 50 µL of 1% Marfey's reagent in acetone were added to the reaction vial. The reaction was incubated for 1 h at 40 ºC with periodic agitation. The reaction was quenched by adding 5 µL of 2 M HCl and dried under a nitrogen stream. The dried peptides were resuspended in 200 µL of MeOH. The sample was injected (6 µL) onto UPLC-MS using the same conditions and instrument as the amino acids standards. To normalize the retention time, 1 µM of fluorescein was added to the standard and peptide samples before injection onto the UPLC-MS. The retention time for the fluorescein sample was adjusted to 6.30 min to correct for run to run drifts in retention time.

## Supporting Information

The sequencing data used in this study is openly available in the NCBI SRA.

### Supporting Information File 1

Additional details and figures including NMR spectra and MS/MS fragmentation.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-138-S1.pdf]

### Supporting Information File 2

Specific dataset used are listed in this Supporting Table File along with core peptide sequence alignments.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-138-S2.xlsx]

### Supporting Information File 3

The full cladogram with species names as a high-resolution pdf.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-138-S3.pdf]

### Supporting Information File 4

The split burpitide precursor peptide HMM and output from all 700 transcriptomes as a compressed archive.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-138-S4.zip]

## Acknowledgements

We like to acknowledge Dr. Manuel Rangel-Grimaldo for his contribution to the structural elucidation of the cyclopeptide alkaloids from *C. americanus* and data collection for *G. jasminoides*.

## Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (R35GM147439 to J.R.C.).

## Author Contributions

Draco Kriger: data curation; investigation; writing – original draft; writing – review & editing. Michael A. Pasquale: conceptualization; investigation; writing – original draft; writing – review & editing. Brigitte G. Ampolini: conceptualization; software; writing – original draft; writing – review & editing. Jonathan R. Chekan: conceptualization; data curation; funding acquisition; software; supervision; visualization; writing – original draft; writing – review & editing.

## ORCID® iDs

Michael A. Pasquale - https://orcid.org/0000-0002-5480-6355
Brigitte G. Ampolini - https://orcid.org/0000-0002-6328-3306
Jonathan R. Chekan - https://orcid.org/0000-0003-4651-0594

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article. The data used for/in this study is openly available in NCBI SRA. Specific dataset used are listed in Supporting Table.

## References

1. Tan, N.-H.; Zhou, J. *Chem. Rev.* **2006,** *106,* 840–895. doi:10.1021/cr040699h
2. Shim, Y. Y.; Young, L. W.; Arnison, P. G.; Gilding, E.; Reaney, M. J. T. *J. Nat. Prod.* **2015,** *78,* 645–652. doi:10.1021/np500802p
3. de Veer, S. J.; Kan, M.-W.; Craik, D. J. *Chem. Rev.* **2019,** *119,* 12375–12421. doi:10.1021/acs.chemrev.9b00402
4. Chekan, J. R.; Mydy, L. S.; Pasquale, M. A.; Kersten, R. D. *Nat. Prod. Rep.* **2024,** in press. doi:10.1039/d3np00042g
5. Montalbán-López, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; van Heel, A. J.; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; Grande Burgos, M. J.; Hill, C.; Kim, S.; Koehnke, J.; Latham, J. A.; Link, A. J.; Martínez, B.; Nair, S. K.; Nicolet, Y.; Rebuffat, S.; Sahl, H.-G.; Sareen, D.; Schmidt, E. W.; Schmitt, L.; Severinov, K.; Süssmuth, R. D.; Truman, A. W.; Wang, H.; Weng, J.-K.; van Wezel, G. P.; Zhang, Q.; Zhong, J.; Piel, J.; Mitchell, D. A.; Kuipers, O. P.; van der Donk, W. A. *Nat. Prod. Rep.* **2021,** *38,* 130–239. doi:10.1039/d0np00027b
6. 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.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013,** *30,* 108–160. doi:10.1039/c2np20085f
7. Lima, S. T.; Ampolini, B. G.; Underwood, E. B.; Graf, T. N.; Earp, C. E.; Khedi, I. C.; Pasquale, M. A.; Chekan, J. R. *Angew. Chem., Int. Ed.* **2023,** *62,* e202218082. doi:10.1002/anie.202218082
8. Chigumba, D. N.; Mydy, L. S.; de Waal, F.; Li, W.; Shafiq, K.; Wotring, J. W.; Mohamed, O. G.; Mladenovic, T.; Tripathi, A.; Sexton, J. Z.; Kautsar, S.; Medema, M. H.; Kersten, R. D. *Nat. Chem. Biol.* **2022,** *18,* 18–28. doi:10.1038/s41589-021-00892-6
9. Kersten, R. D.; Mydy, L. S.; Fallon, T. R.; de Waal, F.; Shafiq, K.; Wotring, J. W.; Sexton, J. Z.; Weng, J.-K. *J. Am. Chem. Soc.* **2022,** *144,* 7686–7692. doi:10.1021/jacs.2c00014
10. Lee, Y.-Y.; Guler, M.; Chigumba, D. N.; Wang, S.; Mittal, N.; Miller, C.; Krummenacher, B.; Liu, H.; Cao, L.; Kannan, A.; Narayan, K.; Slocum, S. T.; Roth, B. L.; Gurevich, A.; Behsaz, B.; Kersten, R. D.; Mohimani, H. *Nat. Commun.* **2023,** *14,* 4219. doi:10.1038/s41467-023-39905-4
11. Mydy, L. S.; Hungerford, J.; Chigumba, D. N.; Konwerski, J. R.; Jantzi, S. C.; Wang, D.; Smith, J. L.; Kersten, R. D. *Nat. Chem. Biol.* **2024,** *20,* 530–540. doi:10.1038/s41589-024-01552-1
12. Trevisan, G.; Maldaner, G.; Velloso, N. A.; da Silva Sant'Anna, G.; Ilha, V.; de Campos Velho Gewehr, C.; Rubin, M. A.; Morel, A. F.; Ferreira, J. *J. Nat. Prod.* **2009,** *72,* 608–612. doi:10.1021/np800377y
13. Kaleem, W. A.; Muhammad, N.; Qayum, M.; Khan, H.; Khan, A.; Aliberti, L.; De Feo, V. *Fitoterapia* **2013,** *91,* 154–158. doi:10.1016/j.fitote.2013.08.024
14. Xu, X.; Jiang, N.; Liu, S.; Jin, Y.; Cheng, Y.; Xu, T.; Wang, X.; Liu, Y.; Zhang, M.; Du, S.; Fan, J.; Zhang, A. *J. Nat. Prod.* **2022,** *85,* 1918–1927. doi:10.1021/acs.jnatprod.1c01215
15. Bushmanova, E.; Antipov, D.; Lapidus, A.; Prjibelski, A. D. *GigaScience* **2019,** *8,* giz100. doi:10.1093/gigascience/giz100
16. Prjibelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. *Curr. Protoc. Bioinf.* **2020,** *70,* e102. doi:10.1002/cpbi.102
17. Haas, B. J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P. D.; Bowden, J.; Couger, M. B.; Eccles, D.; Li, B.; Lieber, M.; MacManes, M. D.; Ott, M.; Orvis, J.; Pochet, N.; Strozzi, F.; Weeks, N.; Westerman, R.; William, T.; Dewey, C. N.; Henschel, R.; LeDuc, R. D.; Friedman, N.; Regev, A. *Nat. Protoc.* **2013,** *8,* 1494–1512. doi:10.1038/nprot.2013.084
18. Hattori, J.; Boutilier, K. A.; Lookeren Campagne, M. M.; Miki, B. L. *Mol. Gen. Genet.* **1998,** *259,* 424–428. doi:10.1007/s004380050832
19. Gournelis, D. C.; Laskaris, G. G.; Verpoorte, R. Cyclopeptide Alkaloids. *Fortschritte der Chemie organischer Naturstoffe/Progress in the chemistry of organic natural products.;* Springer: Vienna, Austria, 1998; Vol. 75, pp 1–179. doi:10.1007/978-3-7091-6507-2_1
20. Gournelis, D. C.; Laskaris, G. G.; Verpoorte, R. *Nat. Prod. Rep.* **1997,** *14,* 75–82. doi:10.1039/np9971400075
21. Tuenter, E.; Exarchou, V.; Apers, S.; Pieters, L. *Phytochem. Rev.* **2017,** *16,* 623–637. doi:10.1007/s11101-016-9484-y
22. Morita, H.; Shimbo, K.; Shigemori, H.; Kobayashi, J. *Bioorg. Med. Chem. Lett.* **2000,** *10,* 469–471. doi:10.1016/s0960-894x(00)00029-9
23. The Angiosperm Phylogeny Group; Chase, M. W.; Christenhusz, M. J. M.; Fay, M. F.; Byng, J. W.; Judd, W. S.; Soltis, D. E.; Mabberlay, D. J.; Sennikov, A. N.; Soltis, P. S.; Stevens, P. F. *Bot. J. Linn. Soc.* **2016,** *181,* 1–20. doi:10.1111/boj.12385
24. Yun, B.-S.; Ryoo, I.-J.; Lee, I.-K.; Yoo, I.-D. *Tetrahedron Lett.* **1998,** *39,* 993–996. doi:10.1016/s0040-4039(97)10797-3
25. Yun, B.-S.; Ryoo, I.-J.; Lee, I.-K.; Yoo, I.-D. *Tetrahedron* **1998,** *54,* 15155–15160. doi:10.1016/s0040-4020(98)00967-3
26. Christina Leung, T.-W.; Williams, D. H.; Barna, C. J.; Foti, S.; Oelrichs, P. B. *Tetrahedron* **1986,** *42,* 3333–3348. doi:10.1016/s0040-4020(01)87397-x
27. Kobayashi, J.; Suzuki, H.; Shimbo, K.; Takeya, K.; Morita, H. *J. Org. Chem.* **2001,** *66,* 6626–6633. doi:10.1021/jo0103423
28. Suzuki, H.; Morita, H.; Iwasaki, S.; Kobayashi, J. *Tetrahedron* **2003,** *59,* 5307–5315. doi:10.1016/s0040-4020(03)00762-2

29. 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.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. *Nat. Biotechnol.* **2016,** *34,* 828–837. doi:10.1038/nbt.3597
30. Servis, R. E.; Kosak, A. I.; Tschesche, R.; Frohberg, E.; Fehlhaber, H. W. *J. Am. Chem. Soc.* **1969,** *91,* 5619–5624. doi:10.1021/ja01048a032
31. Warnhoff, E. W.; Pradhan, S. K.; Ma, J. C. N. *Can. J. Chem.* **1965,** *43,* 2594–2602. doi:10.1139/v65-356
32. Klein, F. K.; Rapoport, H. *J. Am. Chem. Soc.* **1968,** *90,* 2398–2404. doi:10.1021/ja01011a033
33. Singh, A. K.; Pandey, M. B.; Singh, V. P.; Pandey, V. B. *J. Asian Nat. Prod. Res.* **2008,** *10,* 715–718. doi:10.1080/10286020802015828
34. Onozawa, T.; Kitajima, M.; Kogure, N.; Peerakam, N.; Santiarworn, D.; Takayama, H. *J. Nat. Prod.* **2017,** *80,* 2156–2160. doi:10.1021/acs.jnatprod.7b00290
35. Sugawara, F.; Ishimoto, M.; Le-Van, N.; Koshino, H.; Uzawa, J.; Yoshida, S.; Kitamura, K. *J. Agric. Food Chem.* **1996,** *44,* 3360–3364. doi:10.1021/jf960166c
36. Finn, R. D.; Clements, J.; Arndt, W.; Miller, B. L.; Wheeler, T. J.; Schreiber, F.; Bateman, A.; Eddy, S. R. *Nucleic Acids Res.* **2015,** *43,* W30–W38. doi:10.1093/nar/gkv397
37. Letunic, I. phyloT : a phylogenetic tree generator. https://phylot.biobyte.de/ (accessed June 23, 2022).
38. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis;* Springer: Berlin, Germany, 2016.
39. González Sierra, M.; Mascaretti, O. A.; Diaz, F. J.; Rúveda, E. A.; Chang, C.-J.; Hagaman, E. W.; Wenkert, E. *J. Chem. Soc., Chem. Commun.* **1972,** 915–916. doi:10.1039/c39720000915

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.138

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Polymer degrading marine *Microbulbifer* bacteria: an un(der)utilized source of chemical and biocatalytic novelty

Weimao Zhong[1] and Vinayak Agarwal[*1,2,§]

Review

Open Access

Address:
[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA and [2]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

Email:
Vinayak Agarwal[*] - vagarwal@gatech.edu

* Corresponding author
§ Phone (+1)404-385-3798

Keywords:
bacteria; hydrolases; marine natural products

*Beilstein J. Org. Chem.* **2024,** *20,* 1635–1651.
https://doi.org/10.3762/bjoc.20.146

Received: 19 February 2024
Accepted: 02 July 2024
Published: 17 July 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf



© 2024 Zhong and Agarwal; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

*Microbulbifer* is a genus of halophilic bacteria that are commonly detected in the commensal marine microbiomes. These bacteria have been recognized for their ability to degrade polysaccharides and other polymeric materials. Increasingly, *Microbulbifer* genomes indicate these bacteria to be an untapped reservoir for novel natural product discovery and biosynthetic novelty. In this review, we summarize the distribution of *Microbulbifer* bacteria, activities of the various polymer degrading enzymes that these bacteria produce, and an up-to-date summary of the natural products that have been isolated from *Microbulbifer* strains. We argue that these bacteria have been hiding in plain sight, and contemporary efforts into their genome and metabolome mining are going to lead to a proliferation of *Microbulbifer*-derived natural products in the future. We also describe, where possible, the ecological interactions of these bacteria in marine microbiomes.

## Introduction

Microbial natural products represent an important reservoir for drug discovery and along with their (semi)synthetic derivatives, constitute a major component of the contemporary pharmaceutical arsenal. Their thusly validated utility is juxtaposed against the barriers to natural product discovery; high rediscovery rates, reliance on largely serendipitous response in bioactivity assays, and the resources and expertise required for their structure determination being the primary impediments [1,2]. With this background, accessing biological sources that have not yet been extensively mined for natural product discovery is a promising route, one which, at the very least, promises to ameliorate the problem of continued rediscovery of known natural products. This review outlines the recent progress that has been realized using bacteria of the genus *Microbulbifer*. Together with the emerging field of natural product discovery from *Microbulbifer* bacteria, this genus of obligate marine bacteria is well validated to be a valuable source of biopolymer degrading enzymes. In this review, we outline the discovery of *Microbulbifer* enzymes

that have substantial biocatalytic potential as it pertains to biopolymer degradation, followed by a description of natural product discovery from this genus.

*Microbulbifer* is a genus of rod-shaped, aerobic Gram-negative *Alteromonadaceae* family γ-proteobacteria. They are halophilic obligate marine bacteria that are frequently isolated from diverse marine holobionts such as sponges [3,4], sediments [5], algae [6], and corals [7] using media that mimic the native salt water habitat. Strains of this genus are detected together with *Pseudovibrio* and *Ruegeria* strains that grow in identical culture conditions, and hence obtaining axenic *Microbulbifer* strains may require repeated rounds of serial enrichment. This genus was first described by Gonzales et al. in 1997 [5]. More than 130 strains have been reported from regions of wide geographical dispersion (Table 1, Figure 1). *Microbulbifer* bacteria are well known for their capacity to degrade not only a wide variety of polysaccharides including cellulose, agar, chitin, alginate, and xylan [8-11], but also plastics [4,12]. Progressing from their isolation and cultivation, natural product discovery from *Microbulbifer* bacteria is starting to gather steam [6,13,14]. The major classes of natural products isolated from *Microbulbifer* to date include alkaloids [15], fatty acid and polyketides [16,17], and non-ribosomally synthesized peptides [3,4,7]. These findings validate the potential for secondary metabolite discovery from *Microbulbifer* genus. This review will cover the distributions and origins of *Microbulbifer* bacteria strains, degradation enzymes, and secondary metabolite discoveries. A focus is placed on the novel chemical structures reported with reference to their biological activities and the biosynthetic studies they have inspired.

**Table 1:** Summary of marine sources, geographical locations, and key characteristics of *Microbulbifer* strains reported in literature.

| | | Origin | Location | Key characteristics | References |
|---|---|---|---|---|---|
| 1 | *Microbulbifer hydrolyticus* gen. nov. | marine pulp mill | USA | produces hydrolytic enzymes, polyethylene degradation | [5,18,19] |
| 2 | *Microbulbifer* degradans. 2-40 | salt marsh | USA | degrades polysaccharides, polyserine linkers, chitinase | [20-22] |
| 3 | *Microbulbifer salipaludis* sp. nov. | salt marsh | Korea | –[a] | [8] |
| 4 | *Microbulbifer arenaceous* sp. nov. | red sandstone | Scotland | hydrolyzes chitin, esculin, gelatin, and starch | [23] |
| 5 | *Microbulbifer elongates* comb. nov. | –[a] | Germany | –[a] | [24] |
| 6 | *Microbulbifer* sp. JAMB-A7 | sediment | Japan | β-agarase | [25] |
| 7 | *Microbulbifer* sp. JAMB-A94 | sediment | Japan | β-agarase, ι-carrageenase | [10,26-29] |
| 8 | *Microbulbifer* sp. JAMB-A3 | sediment | Japan | β-agarase | [10,30] |
| 9 | *Microbulbifer maritimus* sp. nov. | sediment | Korea | –[a] | [31] |
| 10 | *Microbulbifer* sp. A4B-17 | ascidian | Palau | produces 4-HBA and parabens | [6] |
| 11 | *Microbulbifer* sp. A4A-72 | sponge | Palau | produces 4-HBA | [6] |
| 12 | *Microbulbifer* sp. V296 | sponge | Yap | produces 4-HBA | [6] |
| 13 | *Microbulbifer* sp. A4B-20 | algae | Palau | produces 4-HBA | [6] |
| 14 | *Microbulbifer* sp. A2C-113 | sponge | Palau | produces 4-HBA | [6] |
| 15 | *Microbulbifer* sp. A4A-79 | ascidian | Palau | produces 4-HBA | [6] |
| 16 | *Microbulbifer* sp. AM222 | sediment | Palau | produces 4-HBA | [6] |
| 17 | *Microbulbifer* sp. AM292 | sediment | Palau | produces 4-HBA | [6] |
| 18 | *Microbulbifer* sp. A4A-68 | sponge | Palau | produces 4-HBA | [6] |
| 19 | *Microbulbifer* sp. A2D-17 | sponge | Palau | produces 4-HBA | [6] |
| 20 | *Microbulbifer* sp. I1876 | algae | Japan | produces 4-HBA | [6] |
| 21 | *Microbulbifer* sp. C91 | algae | Yap | produces 4-HBA | [6] |
| 22 | *Microbulbifer* sp. HG111 | sponge | Japan | produces 4-HBA | [6] |
| 23 | *Microbulbifer* sp. F-104 | algae | Palau | produces 4-HBA | [6] |
| 24 | *Microbulbifer* sp. V32 | sponge | Yap | produces 4-HBA | [6] |
| 25 | *Microbulbifer* sp. HG125 | sponge | Japan | produces 4-HBA | [6] |
| 26 | *Microbulbifer* sp. A3G-2 | ascidian | Palau | produces 4-HBA | [6] |
| 27 | *Microbulbifer* sp. A3P2-23-3-2 | ascidian | Palau | produces 4-HBA | [6] |
| 28 | *Microbulbifer* sp. AM220 | sediment | Japan | produces 4-HBA | [6] |

**Table 1:** Summary of marine sources, geographical locations, and key characteristics of *Microbulbifer* strains reported in literature. (continued)

| | | | | | |
|---|---|---|---|---|---|
| 29 | *Microbulbifer* sp. YM3-0188 | sponge | Palau | produces 4-HBA | [6] |
| 30 | *Microbulbifer* sp. A2D-15 | sediment | Palau | produces 4-HBA | [6] |
| 31 | *Microbulbifer* sp. HG869 | ascidian | Palau | produces 4-HBA | [6] |
| 32 | *Microbulbifer* sp. HG279 | sponge | Japan | produces 4-HBA | [6] |
| 33 | *Microbulbifer* sp. ssthio04PA2-36a | sponge | Palau | produces 4-HBA | [6] |
| 34 | *Microbulbifer celer* sp. nov. | saltern | Korea | –[a] | [32,33] |
| 35 | *Microbulbifer halophilus* sp. nov. | saline soil | China | –[a] | [34] |
| 36 | *Microbulbifer* sp. CMC-5 | seaweed | India | degrades polysaccharides, β-agarase | [35-37] |
| 37 | *Microbulbifer variabilis* sp. nov. | algae | Japan | –[a] | [38] |
| 38 | *Microbulbifer epialgicus* sp. nov. | algae | Japan | –[a] | [38] |
| 39 | *Microbulbifer* sp. L4-n2 | sponge | France | produces parabens | [13] |
| 40 | *Microbulbifer donghaiensis* sp. nov. | sediment | China | –[a] | [39] |
| 41 | *Microbulbifer chitinilyticus* sp. nov. | mangrove | Japan | chitin-degrading | [9] |
| 42 | *Microbulbifer okinawensis* sp. nov. | mangrove | Japan | chitin-degrading | [9] |
| 43 | *Microbulbifer* sp. SD-1 | seawater | Korea | degrades agar | [40] |
| 44 | *Microbulbifer* sp. JAM-3301 | sediment | Japan | inulinase producer | [41] |
| 45 | *Microbulbifer maritimus* | seaweed | India | agarase producer | [42] |
| 46 | *Microbulbifer* sp. 6532A | seaweed | Japan | alginate lyase producer | [43,44] |
| 47 | *Microbulbifer marinus* sp. nov. | sediment | China | –[a] | [45] |
| 48 | *Microbulbifer yueqingensis* sp. nov. | sediment | China | –[a] | [45] |
| 49 | *Microbulbifer gwangyangensis* sp. nov. | tidal flat | Korea | –[a] | [46] |
| 50 | *Microbulbifer pacificus* sp. nov. | sponge | Korea | –[a] | [46] |
| 51 | *Microbulbifer mangrove* sp. nov. | mangrove | India | polysaccharide-degrading | [11] |
| 52 | *Microbulbifer elongatus* HZ11 | seawater | China | seaweed-degrading | [47] |
| 53 | *Microbulbifer thermotolerans* DAU221 | sediment | Korea | produces carbohydrate esterase, α-amylase, esterase, chitinase, GH3 -glucosidase polysaccharide lyase | [48-53] |
| 54 | *Microbulbifer elongatus* sp. A13 | seaweeds | India | agarase producer | [54] |
| 55 | *Microbulbifer rhizosphaerae* sp. nov. | halophytic plant | Spain | –[a] | [55,56] |
| 56 | *Microbulbifer* sp. ALW1 | algae | China | produces alginate lyase, laminarinase, β-glucosidase, chondroitinase | [12,57-62] |
| 57 | *Microbulbifer flavimaris* sp. WRN-8 | sediment | China | –[a] | [63] |
| 58 | *Microbulbifer mangrovistrain* DD-13 | mangrove | India | degrades polysaccharides | [64] |
| 59 | *Microbulbifer echini* sp. AM134 | sea urchin | Korea | –[a] | [65] |
| 60 | *Microbulbifer aggregans* sp. CCB-MM1 | sediment | Malaysia | produces sulfite reductase | [66-68] |
| 61 | *Microbulbifer aestuariivivens* sp. GHTF-23 | tidal flat | Korea | –[a] | [69] |
| 62 | *Microbulbifer* sp. Q7 | sea cucumber | China | produces β-agarase, alginate lyases | [70-72] |
| 63 | *Microbulbifer* sp. 127CP7-12 | sponge | Korea | produces violacein | [73] |
| 64 | *Microbulbifer hydrolyticus* IRE-31-192 | chemical mutation | China | polyethylene degradation | [74] |
| 65 | *Microbulbifer* sp. GL-2 | blackfish | Japan | cellulase | [75] |
| 66 | *Microbulbifer agarilyticus*. GP101 | invertebrate turbo cornutus | Korea | –[a] | [76] |
| 67 | *Microbulbifer* sp. BN3 | sediment | China | β-agarase, chitinase | [77-79] |
| 68 | *Microbulbifer* sp. WMMC-695 | tunicate | USA | produces antibacterial acetamido-4-hydroxybenzoate esters | [14] |
| 69 | *Microbulbifer* sp. BH-1 | seawater soil | China | β-agarase | [80] |

**Table 1:** Summary of marine sources, geographical locations, and key characteristics of *Microbulbifer* strains reported in literature. (continued)

| | | | | | |
|---|---|---|---|---|---|
| 70 | *Microbulbifer* sp. C4-6 | coral | Japan | produces fatty acid with weak growth inhibition against *Saccharomyces cerevisiae* | [16] |
| 71 | *Microbulbifer* sp. AG1 | mangrove soil | China | β-agarase producer | [81] |
| 72 | *Microbulbifer* sp. KIT–19 | seawater | Japan | alginate lyase producer | [82] |
| 73 | *Microbulbifer harenosus* sp. nov. | coastal sand | China | alginate-degrading | [83] |
| 74 | *Microbulbifer* sp. DC3-6 | coral | Japan | produces antimicrobial and cytotoxic alkanoyl imidazoles | [15] |
| 75 | *Microbulbifer* sp. SH-1 | coastal soil | China | alginate lyase producer | [84] |
| 76 | *Microbulbifer elongates*. PORT2 | seawater | Indonesia | agarase producer | [85,86] |
| 77 | *Microbulbifer hainanensis* sp. nov. | sediment | China | –[a] | [87] |
| 78 | *Microbulbifer* sp. YNDZ01 | macroalgae | Indonesia | lipase, ι-carrageenase | [88,89] |
| 79 | *Microbulbifer* sp. CL37 | sediment | Malaysia | xylanase producer | [90] |
| 80 | *Microbulbifer* sp. SOL66 | soil | Korea | degrades poly(3-hydroxybutyrate) | [91,92] |
| 81 | *Microbulbifer* sp. SOL03 | soil | Korea | degrades poly(3-hydroxybutyrate) | [91,93] |
| 82 | *Microbulbifer* sp. SOL51 | soil | Korea | degrades poly(3-hydroxybutyrate) | [91] |
| 83 | *Microbulbifer* sp. SOL55 | soil | Korea | degrades poly(3-hydroxybutyrate) | [91] |
| 84 | *Microbulbifer* sp. SOL84 | soil | Korea | degrades poly(3-hydroxybutyrate) | [91] |
| 85 | *Microbulbifer* sp. GL-1 | blackfish | Japan | cellulase-producing | [94] |
| 86 | *Microbulbifer* sp. GL-2 | blackfish | Japan | cellulase-producing | [94] |
| 87 | *Microbulbifer* sp. GL-3 | blackfish | Japan | cellulase-producing | [94] |
| 88 | *Microbulbifer* sp. YPW1 | sediments | China | –[a] | [95] |
| 89 | *Microbulbifer* sp. SH-1 | soil | China | alginate lyase producer | [96] |
| 90 | *Microbulbifer okhotskensis* sp. KMM 9862 | sediment | Russia | –[a] | [97] |
| 91 | *Microbulbifer* sp. BY17 | gracilaria | China | alginate lyase producer | [98] |
| 92 | *Microbulbifer* sp. YX04 | mangrove | China | algicidal | [99] |
| 93–130 | *Microbulbifer* sp. | sponge | USA | –[a] | [100] |
| 131 | *Microbulbifer* sp. C10-1 | coral | Japan | produces cyclic hexapeptides | [7] |
| 132 | *Microbulbifer* sp. RZ01 | coastal environment | China | algicidal | [17] |
| 133 | *Microbulbifer* sp. TB12003 | coastal environment | China | nonalgicidal | [17] |
| 134 | *Microbulbifer* sp. MLAF003 | sponge | USA | produces cyclic hexapeptides | [4] |
| 135 | *Microbulbifer* sp. VAAF005 | sponge | USA | produces cyclic hexapeptides | [4] |
| 136 | *Microbulbifer* sp. MKSA007 | sponge | USA | produces linear peptides | [3] |

[a]No information available.

# Review

## Biopolymer degrading enzymes from *Microbulbifer*

### Agarases

Agarose is a natural polymer consisting of a linear chain of alternating residues of 3-*O*-linked β-D-galactopyranose and 4-*O*-linked 3,6-anhydro-α-L-galactose. Agarose is a major component of agar, a polysaccharide present in the cell walls of some red algae [101]. Agarases are glycoside hydrolases (GHs) that degrade agarose to smaller oligosaccharides [102]. Marine bacteria of diverse genera produce agarases [103-105]. Based on the reactions that they catalyze, bacterial agarases are grouped as α- and β-agarases wherein they hydrolyze α-1,3 linkages and β-1,4 linkages in agarose, respectively (Figure 2). However, based on sequence similarity, agarases are alternatively classified into different families of glycoside hydrolases (GHs): GH-96 and GH-117 for α-agarases, and GH-16, GH-50, GH-86, and GH-118 for β-agarases [106]. All agarases characterized from *Microbulbifer* bacteria belong to β-agarases.

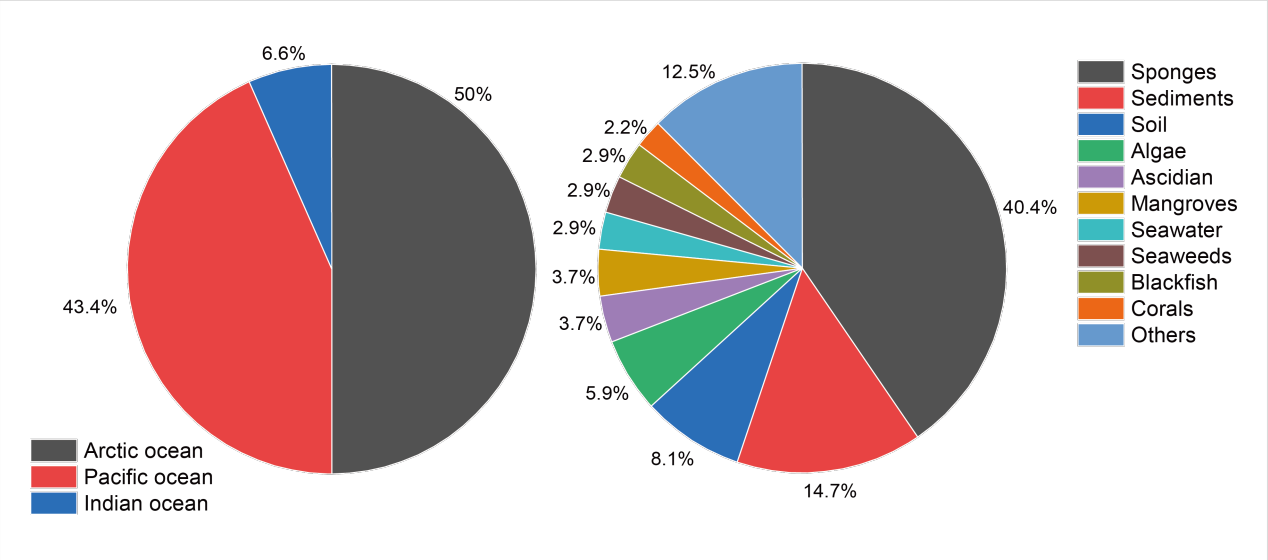



**Figure 1:** Oceanic distribution and marine holobiont sources of *Microbulbifer* strains described in the literature.

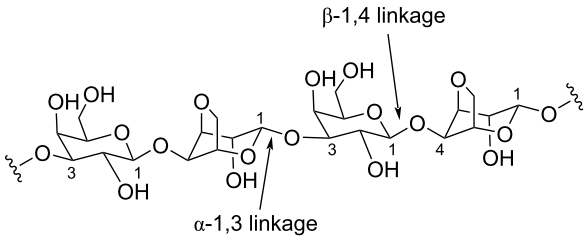



**Figure 2:** The chemical structure of agarose with the key β-1,4 linkage denoted.

**GH-16 agarases:** The first report for the characterization of an agarase from the *Microbulbifer* genus was from the strain *Microbulbifer* sp. JAMB-A7 isolated from deep-sea sediment [25] wherein the agarase encoding gene *agaA7* was sequenced and cloned. The recombinant enzyme RagaA7 was produced in a *Bacillus subtilis* host and characterized to be a neoagarotetraose-producing GH-16 family endo-type β-agarase with a pH and temperature optima being 7.0 and 50 °C, respectively. Another thermostable neoagarotetraose-producing GH-16 endo-type β-agarase rAgaA was identified and cloned from deep-sea-derived *Microbulbifer* sp. JAMB-A94 with pH and temperature optima being 7.0 and 55 °C, respectively [26]. The recombinant enzyme was likewise produced using a *B. subtilis* host. The crystal structure of the catalytic domain was determined to show a β-jelly roll fold with structural similarity to two other β-agarases, ZgAgaA and ZgAgaB, from *Zobellia galactanivorans*, which is a model marine bacterium for the bioconversion of algal biomass [29,107-109]. The crystal structure suggested that the thermostability of rAgaA likely derives from the numerous surface salt bridges, relatively short length of the surface loops, and the increased number of Pro and Arg residues [29].

A truncated β-agarase gene without the carbohydrate-binding modules (Aga16A-ΔCBM) from the marine *Microbulbifer* sp. BH-1 was cloned and expressed in *Escherichia coli* with pH and temperature being 8.0 and 55 °C, respectively. The biocatalytic utility of the enzyme was also evaluated; it was used for neoagarooligosaccharide production in high yield [80].

Five additional GH-16 agarases AgaA3 produced in *B. subtilis*, ID2563 produced in *E. coli*, N3-1 produced in *Pichia pastoris*, AG1 produced in *E. coli*, and AgaF16A produced in *E. coli*, were cloned and expressed from marine *Microbulbifer* bacteria *Microbulbifer* sp. JAMB-A3 [30], *Microbulbifer* sp. Q7 [70], *Microbulbifer* sp. BN3 [77], *Microbulbifer* sp. AG1 [81], and *Microbulbifer elongatus* PORT2 [86], respectively. These agarases similarly degraded agarose into neoagarotetraose. With a view towards their biotechnological utility, AgaA3 was stable to high concentrations of surfactants.

**GH-50 agarases:** The strain *Microbulbifer elongatus* PORT2 was reported to be isolated in Indonesia [85]. In the sequenced genome of this bacterium were detected three agarases to be encoded which belonged to the GH-50 agarase family. The enzymes were named AgaA50, AgaB50, and AgaC50. Biochemical characterization of these three enzymes revealed that AgaA50 and AgaC50 generated neoagarobiose products implying cleavage within the agarose polymer (endolytic hydrolysis). The activity of AgaB50 was different, in that this enzyme cleaved at the external periphery of the agarose polymer to generate neagarosetetraose and neoagarobiose.

**GH-86 agarases:** The *Microbulbifer* sp. JAMB-A94 was shown to produce a GH-16 agarase (rAgaA) [26]. A gene encoding (*agaO*) for a novel GH-86 family agarase was also reported from the same bacterium [27]. The recombinant enzyme (rAgaO) was expressed in *B. subtilis* cells to demonstrate that this enzyme was an endo-type β-agarase and degraded agarose to neoagarohexaose as the main product [27]. The enzyme rAgaO represents the first and only report on the characterization of a purified agarase belonging to the family GH-86 from *Microbulbifer* bacteria.

## Alginate lyases

The polysaccharide alginate is produced by various seaweeds and bacteria [44]. Two building blocks, β-D-mannuronic acid (M) and α-L-guluronic acid (G), are used to assemble alginate in homo- and heteropolymeric forms. The 1,4-*O*-linkage in alginate is cleaved by alginate lyases, in a β-elimination manner (Figure 3). Alginate lyases can be either endolyases, or exolyases with preference for polyM or polyG present in the alginate matrix [110]. Alginate lyases have been isolated from seaweed epiphytic microorganisms, mostly marine bacteria, including the *Microbulbifer* genus [111].

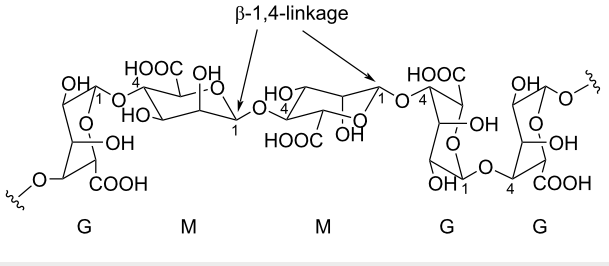



**Figure 3:** The chemical structure of the biopolymer alginate.

An alginate lyase encoding gene *AlgMsp* was discovered from the *Microbulbifer* sp. 6532A which was isolated from seaweed surfaces [44]. Recombinant AlgMsp expressed in *E. coli* was characterized to a slight preference for polyG over polyM substrates [44]. Recombinant alginate lyase AlyM expressed in *E. coli* from *Microbulbifer* sp. Q7 exhibited preference for polyG [71]. AlyM preferably degraded the glycosidic bond at the G–X linkage in alginate [71]. Further studies focused on the improvement in thermal stability by introduction of disulfide bonds [112,113].

Two new alginate lyases, AlgL17 and AlgL6, both expressed in *E. coli*, were characterized from *Microbulbifer* sp. ALW1 isolated from rotten brown algae [61,62]. AlgL17 preferentially degraded polyM with poor activity towards polyG, indicating it to be a polyM-specific alginate lyase. AlgL17 was an exotype alginate lyase based on its ability of producing 4-deoxy-L-erythro-5-hexoseulose uronic acid (DEH) from sodium alginate [62]. AlgL6 degraded polyM, polyG, and sodium alginate in an exolytic manner. It exhibited good stability in the presence of some nonionic detergents [61].

Additional alginate lyases AlgSH17, BY17PV7, and MtAl138 were cloned and expressed in *E. coli* from *Microbulbifer* sp. SH-1 [96], *Microbulbifer* sp. BY17 [98], and *M. thermotolerans* DAU221 [52], respectively. Recombinant AlgSH17 exhibited both exolytic and endolytic activities [96]. BY17PV7 demonstrated wide substrate tolerance and good degradation effects for both polyM and polyG [98]. MtAl138 belongs to the polysaccharide lyase family 7 and is an endotype enzyme that produces di-, tri-, or tetrasaccharides from polyG and polyM [52].

## Chitinases

Chitin is a linear β-1,4-linked homopolymer of *N*-acetyl-β-D-glucosamine (GlcNAc), the second most abundant biomass on earth after cellulose (Figure 4). Approximately $10^{11}$ tons of chitin are produced annually in nature, about 70% of which originate from the ocean. Most of the chitin is recycled by bacteria and fungi as carbon and nitrogen sources [114,115]. Chitin is traditionally degraded with environmentally unfriendly concentrated acids or alkalis. Chitinases are hydrolytic enzymes that degrade glycosidic bonds between chitin polymers and belong to the GH-18 and GH-19 family [116]. Chitinases can be also classified as endo- or exochitinases.

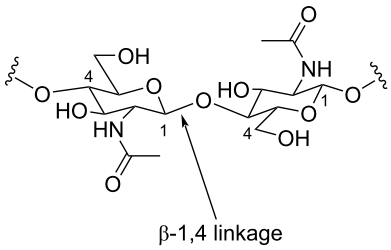



**Figure 4:** The chemical structure of chitin.

Genomic analysis of a marine *Microbulbifer degradans* 2-40 revealed three chitin depolymerases (ChiA, ChiB, and ChiC) [117]. ChiB was cloned and expressed in *E. coli* [22]. It is a modular protein that is predicted to contain two GH-18 catalytic domains, two polyserine domains, and an acidic repeat domain. It functions as an exochitinase. The two catalytic domains have different activities on chitooligosaccharides. Each domain was maximally active from 30 °C to 37 °C and from pH 7.2 to 8.0. Both domains function cooperatively to degrade chitin [22]. It should be noted that *M. degradans* 2-40 was reclassified as *Saccharophagus degradans* 2-40 [118].

Two additional GH18 chitinases, MtCh509 [50] and rChi1602 [79] were characterized from *Microbulbifer thermotolerans* DAU221 and *Microbulbifer* sp. BN3, respectively. MtCh509 was expressed in *E. coli*. Some organic solvents (benzene, DMSO, hexane, isoamyl alcohol, isopropyl alcohol, and toluene) increased the reactivity of MtCh509 relative to the aqueous system, representing the first solvent-tolerant chitinase from *Microbulbifer* species and its potential applications in industrial processes [50]. rChi1602 exhibited maximal activity at 60 °C and over a broad pH range between 5.0 and 9.0 and was highly expressed in *Pichia pastoris* [79].

## Carrageenase

Carrageenans are sulfated saccharide polymers with a wide range of commercial applications. The differential sulfation patterns allow for carrageenans to be differentiated into three basic forms – κ-, ι-, and λ-carrageenans (Figure 5) [28]. Carrageenans find utility in drug and food industries because of their antioxidant and antiviral activities [119-121].

In the sequenced genome of a marine-derived *Microbulbifer thermotolerans* JAMB-A94$^T$ was detected an ι-carrageenase encoded by gene *cgiA* [28]. The recombinant CgiA enzyme was produced in *B. subtilis* and its biochemical characterization revealed it exhibited maximal activity at 50 °C. It can break ι-carrageenan into tetrasaccharide at high proportion. A conserved Glu351, rather than an Asp, was identified as the catalytic residue by site-directed mutagenesis experiments.

Another ι-carrageenase, Car1293, was identified, expressed in *E. coli,* and characterized from macroalgae-associated *Microbulbifer* sp. YNDZ01 [89].

## β-Glucosidases

The β-glucosidases are enzymes that can hydrolyze the β-D-glycosidic bonds to release glucose at the non-reducing end of oligosaccharides. Based on their substrate specificity, β-glucosidases can be divided into aryl-β-glucosidases that act on aryl glucosides, cellobiases that hydrolyze the disaccharide cellobiose, and other glucosidases [122]. They can also be classified into GH-1 and GH-3 families based on sequence similarities [123].

The first β-glucosidase characterized from *Microbulbifer* bacteria is the GH-3 MtBgl85 derived from the *Microbulbifer thermotolerans* isolated from deep sea sediment [51]. It was expressed in *E. coli.* A gene encoding the GH-1 β-glucosidase MaGlu1A was cloned and expressed in *E. coli* from a marine *Microbulbifer* sp. ALW1 [59]. Enzyme structure and site-directed mutagenesis led to the identification of key residues in the enzyme active site that participated in the hydrolytic activity [59].

## Carbohydrate esterases

Carbohydrate esterases (CEs) catalyze the *O*- or *N*-deacylation of substituted saccharides. The vast diversity among CEs is organized into 16 different families [123]. Members of the CE6 family are typical serine-type esterases. A cold-adapted CE6 enzyme, CEST, was cloned and expressed in *E. coli* from *Microbulbifer thermotolerans* DAU221 with a temperature optimum of 15 °C [48]. The three-dimensional structure of CEST revealed that it belongs to the α/β-class of proteins consisting of a central six-stranded β-sheet flanked by eight α-helices. Site-directed mutagenesis indicated that a Ser-His-Glu catalytic triad was essential for the enzyme activity [48]. Another cold-adapted CE was isolated from a fosmid genomic library of *Microbulbifer thermotolerans* DAU221 with a similar catalytic triad and expressed in *E. coli* [56].

## Other enzymes

Chondroitinases (ChSases) are enzymes that digest chondroitin sulfate chains to generate disaccharides. They can be classified into hydrolases and lyases, based on their enzymatic mechanism. ChSase B6 is a chondroitinase originally identified in marine *Microbulbifer* sp. ALW1 [60] and was cloned and expressed in *E. coli.* ChSase B6 could digest chondroitin sulfate B into disaccharides with good thermostability and stability against surfactants. Structural analysis and site-directed mutage-

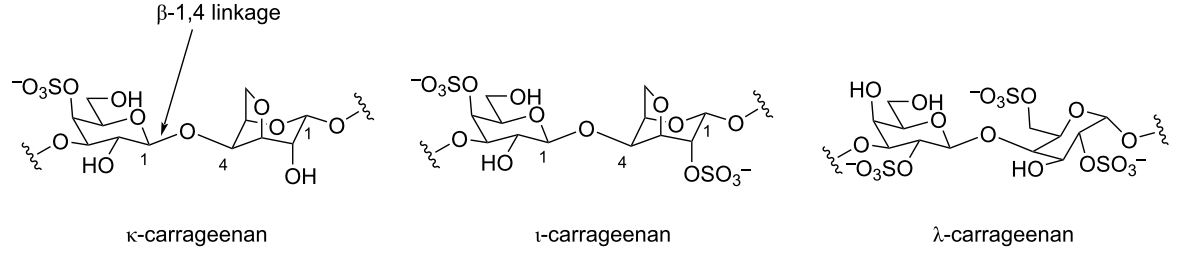



**Figure 5:** Chemical structures of sulfated polysaccharides κ-, ι-, and λ-carrageenans.

nesis identified residues important for the catalytic activity of ChSase B6 [60]. In addition, a novel laminarinase, MaLamNA, was characterized and expressed in *Pichia pastoris* from the same bacterium. It functioned exclusively towards the substrate laminarin and represented the first β-1,3-glucanase from the genus of *Microbulbifer* [58].

*Microbulbifer* sp. JAM-3301 is an inulinase-producing bacterium isolated from deep sea sediment. An inulin operon that contained three open reading frames was cloned and sequenced. One inulinase and one β-fructofuranosidase were expressed in *E. coli* [41]. Both enzymes can work together to effectively degrade inulin.

Lipases are enzymes that can hydrolyze various fatty esters and find applications in the detergent, oil, and pharmaceutical industries. The *lip4346* is a novel lipase-encoding gene cloned from the macroalgae-derived *Microbulbifer* sp. YNDZ01 and expressed in *E coli* [88]. The recombinant enzyme Lip4346 showed high stability at high temperatures and in alkaline conditions, and tolerance to organic solvents.

α-Amylases are glycoside hydrolases that catalyze the hydrolysis of internal α-1,4-*O*-glycosidic bonds in starch and maltodextrins. An α-amylase AmyA from *Microbulbifer thermotolerans* DAU221 was heterologously expressed in *E. coli,* purified, and characterized [49]. It represents the first α-amylase reported from the *Microbulbifer* genus [49].

## Natural products from *Microbulbifer* genus

### 4-Hydroxybenzoate esters

The first group of secondary metabolites described from the genus *Microbulbifer* were 4-hydroxybenzoate (4HBA, **1**) and three alkyl esters (butyl, heptyl, and nonyl, **2**–**4**, Figure 6). Commonly known as parabens, these molecules were isolated from a tropical ascidian-derived *Microbulbifer* sp. A4B-17, the crude extract of which demonstrated modest antimicrobial activity against several Gram-positive bacteria, molds, and yeasts [6]. The study further characterized another 23 *Microbulbifer* bacteria derived from different marine sources. Although they were able to produce 4HBA in lower amounts, other parabens were not detected.

Parabens are used as preservatives in pharmaceutical, food, and cosmetic industries due to their low toxicity, stability, and activity against a broad spectrum of microorganisms [124]. Industrially, parabens are prepared starting from 4HBA, which in turn, is a petrochemical. *Microbulbifer* sp. A4B-17 constitutes the first microorganism to accumulate 4HBA and parabens, most probably as a metabolite derived from the shikimate pathway [125].

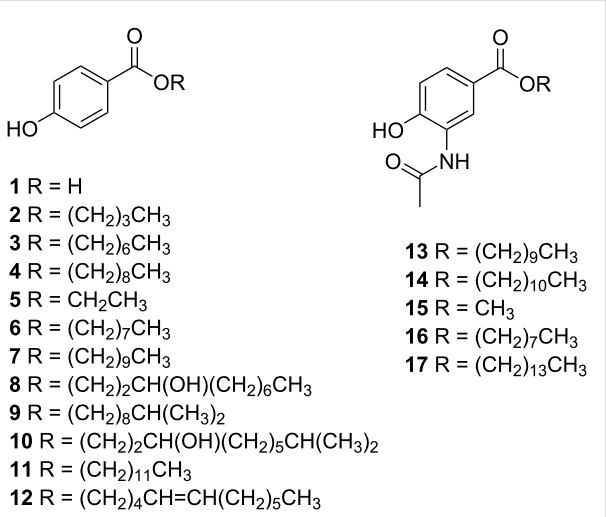
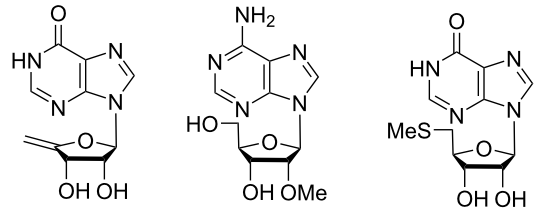



**Figure 6:** Chemical structures of 4HBA (**1**) and parabens (**2**–**14**) isolated from *Microbulbifer* strains, and synthetic analogus (**15**–**17**).

A variety of parabens were isolated and characterized from a marine *Microbulbifer* sp. L4-n2 (Figure 6) [13]. In additon to four known parabens (**2**, **5**, **6**, and **11**), five new ones (**7**–**10**, and **12**) were identified. Their structures were elucidated by high-resolution mass spectrometry (HRMS) and 1D and 2D nuclear magnetic resonance (NMR). The configuration in **8** and **10** remained undetermined. Antimicrobial activities of compounds **2** and **5**–**12** against *Staphlyococcus aureus* indicated that **7** and **12** were bacteriostatic, while **9** and **10** were bactericidal. Compound **9** was also active towards another two commensal Gram-positive bacteria *Bacillus* sp. and *Planococcus* sp. isolated from the host sponge. Compound **9** can be detected in situ in the sponge collected at different times, suggesting it may serve an ecological role in the balance between bacterial community and the sponge host [13].

Jayanetti et al. investigated 16 *Microbulbifer* strains isolated from marine tunicate *Ecteinascidia turbinate*. Among these, one strain *Microbulbifer* sp. WMMC-695 showed antimicrobial activity against *E. coli* [14]. Bioassay-guided fractionation led to the isolation of two new acetamidobenzoate esters: bulbiferate A (**13**) and bulbiferate B (**14**, Figure 6). Their structures were established on the basis of comprehensive spectroscopic and spectrometric data. Their structures differ from common parabens by the presence of an acetamide group *ortho* to the phenolic hydroxy group [14]. They exhibited weak activity against *E. coli* and methicillin-sensitive *Staphylococcus aureus* (MSSA). The three analogues **15**–**17** (Figure 6) with different ester chain lengths were synthesized and tested for antimicrobial activity. However, no clear trend was found between the ester length and bioactivity. In addition to the paraben derivatives, three known nucleosides, 4′,5′-didehydro-5′-deoxyinosine (**18**),

2′-*O*-methyladenosine (**19**), and 5′-methylthioinosine (**20**) were also identified from this bacterium by comparison of their $^1$H NMR and MS data with those reported in literature (Figure 7).




**18** **19** **20**

**Figure 7:** Chemical structures of nucleosides **18**–**20** isolated from *Microbulbifer* strains.

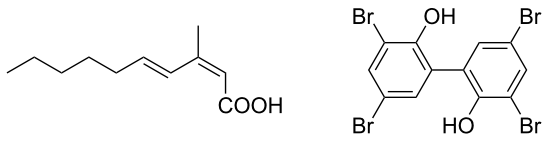
## Alkaloids

Three new alkanoyl imidazoles **21**–**23** (Figure 8) were isolated from a culture extract of a *Microbulbifer* bacterium isolated from a scleractinian coral [15]. Their structures were determined using a combination of NMR spectroscopy and chemical derivatization experiments, adding new members to this class of imidazole-containing natural products such as the nocarimidazoles A and B reported from a marine-derived actinomycete *Nocardiopsis* sp. before [126]. Compound **21** was determined to be a pair of enantiomers with a ratio of 9% *R* and 91% *S* using the Ohrui–Akasaka method [127]. Structurally, compound **21** represents the second example with a (*R*)-*anteiso* enantiomer in addition to nocapyrone L [128]. An antimicrobial activity assay showed that compounds **21**–**23** could inhibit the growth of Gram-positive bacteria *Kocuria rhizophila* and *S. aureus*, Gram-negative bacterium *Tenacibaculum maritimum*, and a panel of fungi. These molecules also demonstrated moderate cytotoxicity against P388 murine leukemia cells.

The bisindole alkaloid violacein (**24**, Figure 8) was isolated and characterized from a sponge-derived *Microbulbifer* strain [73]. This deep purple pigment has been isolated from several other bacterial genera [129], including but not limited to the epiphytic commensal marine bacteria of the genus *Pseudoalteromonas* where the biosynthetic pathway was eventually established [130,131]. The fact that *Microbulbifer* bacteria produce natural products that have been detected to be formed by other marine bacterial genera supports the hypothesis that biosynthetic gene clusters encoding production of natural products with wide ranging implications in bacterial physiology or chemical interactions in the marine environment may be broadly shared [132]. Further examples of shared natural products are delineated below.

## Aliphatic and aromatic hydrocarbons

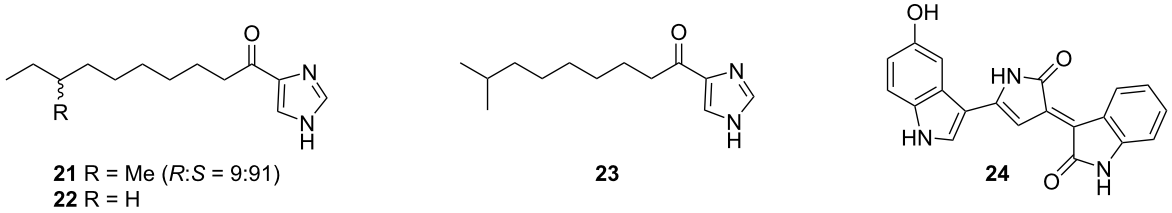
(2*Z*,4*E*)-3-Methyl-2,4-decadienoic acid (**25**, Figure 9), an unsaturated fatty acid with unusual methylation pattern was isolated and identified from a coral-associated *Microbulbifer* sp. [16]. The molecule was previously known as a synthetic compound; this is the first report of it being detected as a secondary metabolite from a microorganism. Compound **25** is weakly active against *Saccharomyces cerevisiae*. The position of the branching methyl group in **25** is rare. Feeding [1-$^{13}$C]acetate to the culture resulted the enrichment of C1, C3, C5, C7, and C9 carbons, which verified the origin of the carbon skeleton as being fatty acid-derived. Feeding experiments with L-[*methyl*-$^{13}$C]methionine yielded only enrichment of C11 methyl, indicating that this branching methyl could be derived from via *S*-adenosylmethionine (SAM).




**25** **26**

**Figure 9:** Chemical structures of (2*Z*,4*E*)-3-methyl-2,4-decadienoic acid (**25**) and 4-BP (**26**) natural products isolated from *Microbulbifer* strains.

Phytoplankton and bacteria have had a long-term coexistence and complex interactions ranging from mutualism, antagonism,




**21** R = Me (*R*:*S* = 9:91)
**22** R = H

**23** **24**

**Figure 8:** Chemical structures of alkaloids **21**–**24** isolated from *Microbulbifer* strains.

and competition to parasitism in the ocean [133]. Algicidal bacteria often emerge in the late stage of an algal bloom and are thought to be involved in the cessation of the bloom [134]. However, molecular algicidal strategies and their global prevalence remain understudied. Recently, Zhang, et al. reported a marine *Microbulbifer* sp. RZ01 which demonstrated broad algicidal activity against 11 cultivated algae [17]. Further investigation indicated that the bioactive substance was mainly located in the cell-free supernatant of *Microbulbifer* sp. RZ01. Fractionation and spectroscopic characterization led to the isolation and identification of the extracellularly secreted molecule, 3,3´,5,5´-tetrabromo-2,2-biphenyldiol (4-BP, **26**, Figure 9).

An in situ algal bloom was simulated to test the algicidal activity of compound **26**. These experiments demonstrated that even low concentrations of **26** can cause changes in overall phytoplankton community. Mechanistic investigations suggested that compound **26** could disrupt electron transport in the photosynthetic light reactions by interfering with the synthesis of plastoquinone-9 [17]. Global genome sequence analyses confirmed the ubiquitous presence of biosynthetic gene clusters (BGCs) encoding the production of **26** in diverse marine bacterial species, suggesting it to be a potent bacterial tool to mediate bacterial–algal antagonistic relationships. Indeed, prior to isolation from a *Microbulbifer* strain, molecule **26** had been isolated from marine *Pseudoalteromonas* and *Marinomonas* genera, and the accumulation and biotransformation of this bacterial natural product in marine mammal metabolomes were postulated [135-137]. In the same vein, before *Microbulbifer*, BGCs encoding production of **26** were also first identified and experimentally characterized from *Pseudoalteromonas* and *Marinomonas* bacteria; genes encoding biosynthetic enzymes that assembled **26** were clustered together with genes encoding the production of polybrominated pyrroles [138-140]. The *Microbulbifer* BGCs lack genes for brominated pyrrole biosynthesis and polybrominated pyrroles have not been reported to be produced by *Microbulbifer* bacteria. Future studies looking at the evolutionary relationship between polybrominated phenol biosynthetic enzymes in diverse marine microbial genera will undoubtedly complement the wide distribution of these natural products in the marine metabolome. All marine bacteria discussed in this section are members of commensal or symbiotic microbiomes of marine invertebrates and plants, perhaps hinting at these molecules being widely used alphabets in inter-organismal chemical cross talk.

## Non-ribosomally synthesized peptidic natural products

Assembly line non-ribosomal peptide synthetases (NRPSs) are routinely detected to be encoded in *Microbulbifer* genomes [141]. An NRPS-derived macrolactam ureidopeptide, bulbiferamide (**27**, Figure 10), was isolated and characterized from a coral-derived *Microbulbifer* sp. C10-1 by Igarashi and co-workers [7]. The ureido linkage did not allow for the peptide main chain in **27** to possess a terminal primary amine and none of the side chains possessed a nucleophilic primary amine. Curiously, the macrocyclizing amide bond was found to exist between a Trp side chain indole nitrogen and the C-terminal carboxylate. To the best of our knowledge, this was the first example of the involvement of a Trp indole in macrolactam formation in bacterial natural products, the only other example being the fungal psychrophilins [142-144]. It should be noted that the Trp indole nitrogen does participate in amide bond formation in other fungal natural products as well, but not in the context of macrocycle assembly [145]. Compound **27** inhibited the growth of *T. cruzi* epimastigotes with an $IC_{50}$ value of 4.1 μM, which is more potent than the approved drug benznidazole ($IC_{50}$ = 20 μM) [7].

Concomitant with Igarashi, we reported the isolation of **27** from a sponge-derived *Microbulbifer* sp. MLAF003 [4]. Three analogues (**28**–**30**, Figure 10) were detected and characterized by NMR and MS from another sponge-derived *Microbulbifer* sp. VAAF005. Additional congeners were detected in the extracts of yet another sponge-derived *Microbulbifer* sp. 22VTAC004. These data demonstrate that the potential for bulbiferamide production is widely spread among *Microbulbifer* bacteria and is not limited to *Microbulbifer* strains isolated from either coral or sponge microbiomes. Imaging mass spectrometry demonstrated that bulbiferamides are excreted by *Microbulbifer* bacterial colonies into the extracellular media.

Based on the characteristic mass spectrometric fragmentations of the Phe-ureido moiety in **27**–**30**, we detected additional ureidopeptidic natural products in the extracts of the sponge-derived strain *Microbulbifer* sp. MKSA007. Subsequent isolation and structure elucidation efforts yielded a new group of ureidopeptides **31**–**35** (Figure 10) which we named pseudobulbiferamides [3]. Congruent with the theme of *Microbulbifer* bacteria sharing natural products with other marine bacteria, compounds **31**–**34** are similar to pseudovibriamides, ureidopeptide natural products that are produced by numerous strains of *Pseudovibrio* bacteria [146]. Moreover, BGCs encoding production of pseudobulbiferamides in *Microbulbifer* and pseudovibriamides in *Pseudovibrio* are quite similar and are located on plasmids, rather than chromosomal DNA. With the observation that *Pseudovibrio* and *Microbulbifer* co-inhabit commensal microbiomes of marine animals, it is tantalizing to speculate that this plasmid borne BGC has been shared. Unlike the bulbiferamides mentioned above, the pseudobulbiferamides were not exclusively excreted out of the bacterial colonies.

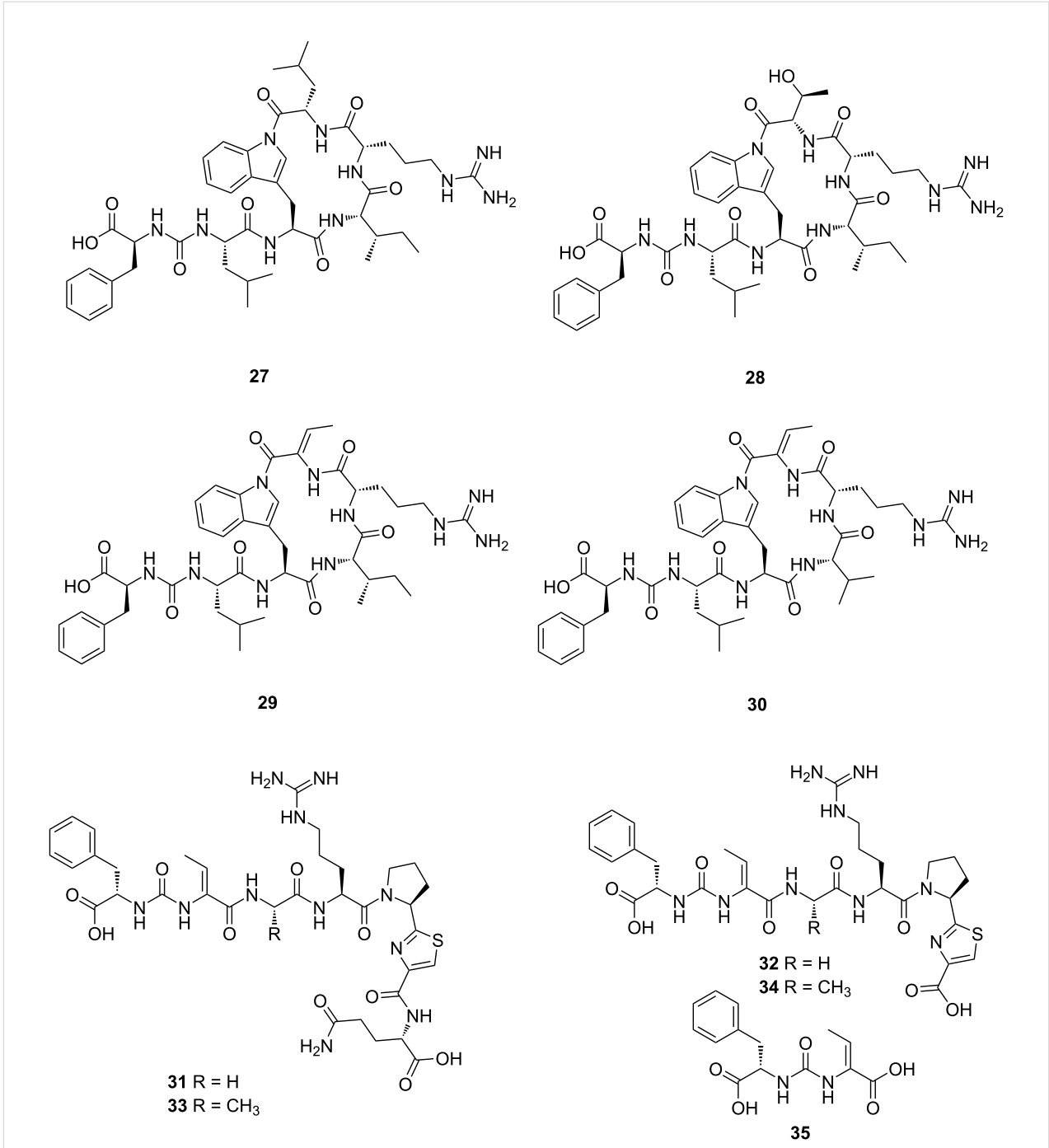



**Figure 10:** Chemical structures of bulbiferamides **27**–**30** and pseudobulbiferamides **31**–**35**.

## Biosynthesis

The first genome sequence of a *Microbulbifer* bacterium was reported by Howard et al. in 2003 [117]. Now, more than 70 *Microbulbifer* genomes are publicly available which are typically 5 Mbp in size. Computational mining these genomes with antiSMASH reveals that these bacteria, though not endowed with the biosynthetic prowess of actinomycetes, do still possess BGCs encoding NRPS-derived peptides, polyketides, RiPPs, and siderophores [4,147]. A typical *Microbulbifer* genome contains less than ten BGCs, as identified by antiSMASH, which implies that a large fraction of their genome is not devoted to natural product biosynthesis. The bulbiferamides and the pseudobulbiferamides represent the first and to date the only examples of peptidic natural products isolated from *Micro-*

*bulbifer* (the corresponding BGCs and NRPS assembly lines are identified as denoted in Figure 11) [3,4]. Thus, it is immediately apparent that the potential for novel compound discovery and for the heterologous production of cryptic natural products from this genus is high. Correlating *Microbulbifer* genomes and metabolomes will drive advances in these directions.

The genome sequence of *Microbulbifer* sp. strain HZ11 revealed the presence of an alkylquinolone (AQ)-producing BGC; however, no AQs were detected to be produced by this strain. This BGC was unique to strain HZ11 among the available *Microbulbifer* genomes [148]. Feeding AQ substrates 2-heptyl-1*H*-quinolin-4-one (**36**, HHQ, Figure 12), 2-heptyl-1-hydroxyquinolin-4-one (**37**, HQHO) to *Microbulbifer* sp. strain HZ11 yielded the brominated AQ products BrHHQ **38** and

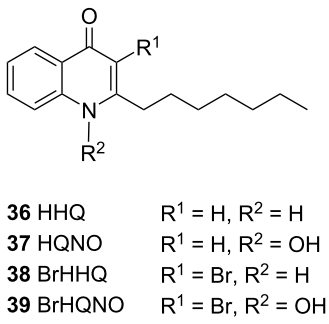



**Figure 12:** Chemical structures of 2-heptyl-1*H*-quinolin-4-one (**36**, HHQ), 2-heptyl-1-hydroxyquinolin-4-one (**37**, HQHO), and their brominated products BrHHQ **38** and BrHQHO **39**.

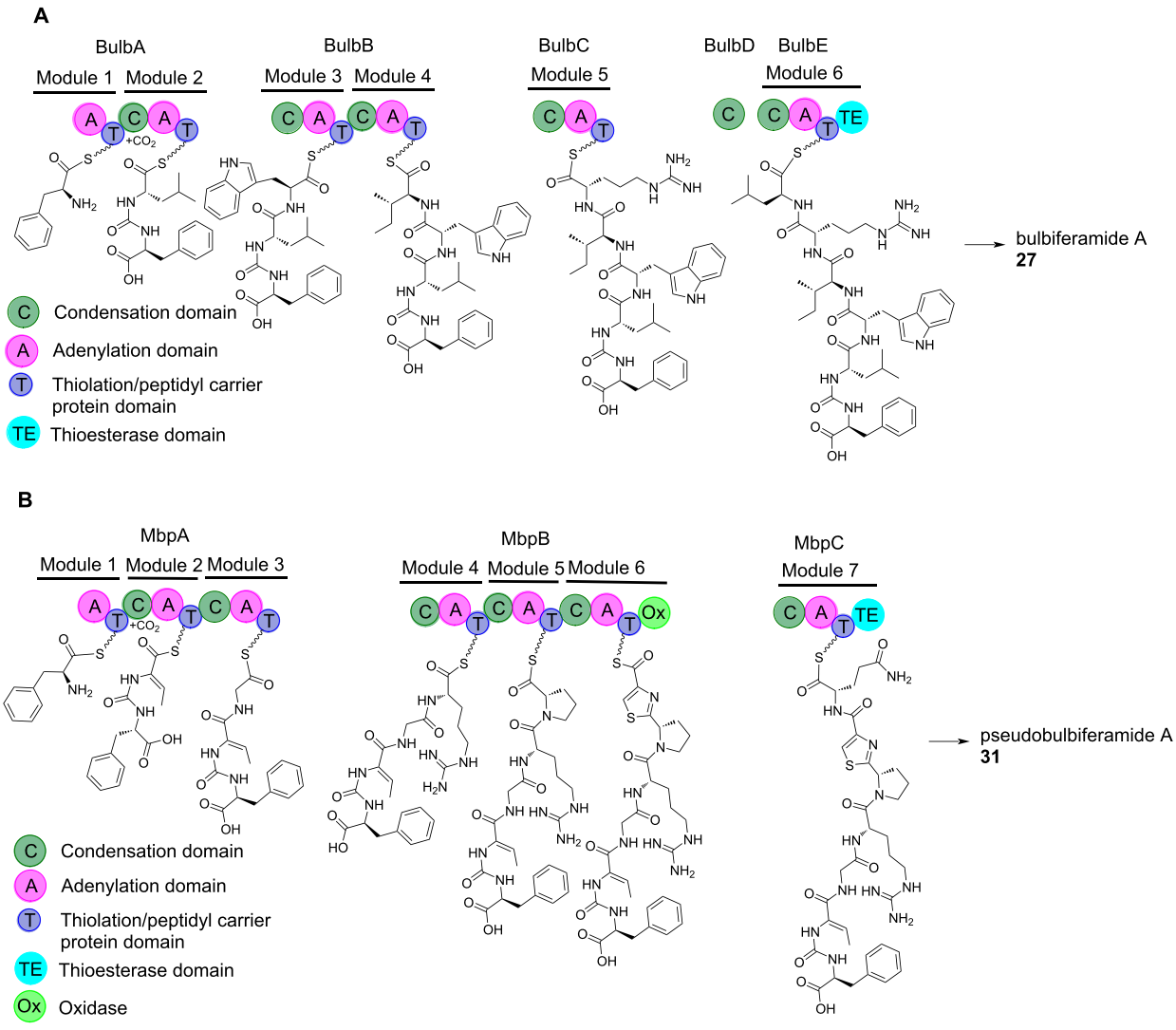



**Figure 11:** Proposed NRPS assembly lines for the biosynthesis of (**A**) bulbiferamide A (**27**) and (**B**) pseudobulbiferamide A (**31**).

BrHQHO **39**. A gene encoding a vanadium-depenent haloperoxidase (VHPO) was detected in the genome and recombinantly produced VHPO was shown to brominate AQs. Besides, the brominated AQs demonstrated increased antibiotic activity against *Staphylococcus aureus* and other marine bacteria [148]. In addition to the antimicrobial activity demonstrated herein, AQs are fundamentally signaling molecules and enzymatic modification of AQs by *Microbulbifer* could represent a mechanism of ecological cross-talk among marine microbiomes. Enzymatic halogenation of other signaling molecules, such as that of acyl homoserine lactones, has been postulated to modulate bacterial interactomes [149,150].

## Conclusion

Geographically, halophilic *Microbulbifer* bacteria are widely distributed with marine sponges (40.4%) and sediments (14.7%) being the most prolific sources for isolation (Figure 1, Table 1). Twelve groups of degradation enzymes have been expressed and characterized. A total of 32 natural products, including NRPSs, PKS, parabens, and alkaloids, have been isolated and identified from *Microbulbifer* bacteria. Based on the growing genomic data, it is unquestionable that the *Microbulbifer* genus is an untapped resource for natural products. The number of molecules described thus so far is not comparable to the expected metabolites envisioned from bioinformatics analyses of the *Microbulbifer* genomes. Contemporary and in-development tools and technologies for data mining, synthetic biology, and strain manipulation will have a transformative effect on future natural product discovery from the *Microbulbifer* genus.

## Funding

The authors are grateful to the National Science Foundation (NSF) for grant support (CHE-2238650) to V.A.) which has enabled their work on peptidic natural products from marine bacteria.

## Author Contributions

Weimao Zhong: conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; writing – original draft; writing – review & editing. Vinayak Agarwal: conceptualization; data curation; formal analysis; funding acquisition; investigation; project administration; supervision; writing – original draft; writing – review & editing.

## ORCID® iDs

Vinayak Agarwal - https://orcid.org/0000-0002-2517-589X

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

1. Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; the International Natural Product Sciences Taskforce; Supuran, C. T. *Nat. Rev. Drug Discovery* **2021,** *20,* 200–216. doi:10.1038/s41573-020-00114-z
2. Hifnawy, M. S.; Fouda, M. M.; Sayed, A. M.; Mohammed, R.; Hassan, H. M.; AbouZid, S. F.; Rateb, M. E.; Keller, A.; Adamek, M.; Ziemert, N.; Abdelmohsen, U. R. *RSC Adv.* **2020,** *10,* 20939–20959. doi:10.1039/d0ra04025h
3. Zhong, W.; Aiosa, N.; Deutsch, J. M.; Garg, N.; Agarwal, V. *J. Nat. Prod.* **2023,** *86,* 2414–2420. doi:10.1021/acs.jnatprod.3c00595
4. Zhong, W.; Deutsch, J. M.; Yi, D.; Abrahamse, N. H.; Mohanty, I.; Moore, S. G.; McShan, A. C.; Garg, N.; Agarwal, V. *ChemBioChem* **2023,** *24,* e202300190. doi:10.1002/cbic.202300190
5. Gonzalez, J. M.; Mayer, F.; Moran, M. A.; Hodson, R. E.; Whitman, W. B. *Int. J. Syst. Bacteriol.* **1997,** *47,* 369–376. doi:10.1099/00207713-47-2-369
6. Peng, X.; Adachi, K.; Chen, C.; Kasai, H.; Kanoh, K.; Shizuri, Y.; Misawa, N. *Appl. Environ. Microbiol.* **2006,** *72,* 5556–5561. doi:10.1128/aem.00494-06
7. Lu, S.; Zhang, Z.; Sharma, A. R.; Nakajima-Shimada, J.; Harunari, E.; Oku, N.; Trianto, A.; Igarashi, Y. *J. Nat. Prod.* **2023,** *86,* 1081–1086. doi:10.1021/acs.jnatprod.2c01083
8. Yoon, J.-H.; Kim, I.-G.; Shin, D.-Y.; Kang, K. H.; Park, Y.-H. *Int. J. Syst. Evol. Microbiol.* **2003,** *53,* 53–57. doi:10.1099/ijs.0.02342-0
9. Baba, A.; Miyazaki, M.; Nagahama, T.; Nogi, Y. *Int. J. Syst. Evol. Microbiol.* **2011,** *61,* 2215–2220. doi:10.1099/ijs.0.024158-0
10. Miyazaki, M.; Nogi, Y.; Ohta, Y.; Hatada, Y.; Fujiwara, Y.; Ito, S.; Horikoshi, K. *Int. J. Syst. Evol. Microbiol.* **2008,** *58,* 1128–1133. doi:10.1099/ijs.0.65507-0
11. Vashist, P.; Nogi, Y.; Ghadi, S. C.; Verma, P.; Shouche, Y. S. *Int. J. Syst. Evol. Microbiol.* **2013,** *63,* 2532–2537. doi:10.1099/ijs.0.042978-0
12. Li, Z.; Du, Z.; Li, H.; Chen, Y.; Zheng, M.; Jiang, Z.; Du, X.; Ni, H.; Zhu, Y. *AMB Express* **2022,** *12,* 139. doi:10.1186/s13568-022-01482-y
13. Quévrain, E.; Domart-Coulon, I.; Pernice, M.; Bourguet-Kondracki, M.-L. *Environ. Microbiol.* **2009,** *11,* 1527–1539. doi:10.1111/j.1462-2920.2009.01880.x
14. Jayanetti, D. R.; Braun, D. R.; Barns, K. J.; Rajski, S. R.; Bugni, T. S. *J. Nat. Prod.* **2019,** *82,* 1930–1934. doi:10.1021/acs.jnatprod.9b00312
15. Karim, M. R. U.; Harunari, E.; Oku, N.; Akasaka, K.; Igarashi, Y. *J. Nat. Prod.* **2020,** *83,* 1295–1299. doi:10.1021/acs.jnatprod.0c00082
16. Sharma, A. R.; Harunari, E.; Zhou, T.; Trianto, A.; Igarashi, Y. *Beilstein J. Org. Chem.* **2019,** *15,* 2327–2332. doi:10.3762/bjoc.15.225
17. Zhang, Z.; Li, D.; Xie, R.; Guo, R.; Nair, S.; Han, H.; Zhang, G.; Zhao, Q.; Zhang, L.; Jiao, N.; Zhang, Y. *ISME J.* **2023,** *17,* 1979–1992. doi:10.1038/s41396-023-01510-0
18. Li, Z.; Wei, R.; Gao, M.; Ren, Y.; Yu, B.; Nie, K.; Xu, H.; Liu, L. *J. Environ. Manage.* **2020,** *263,* 110402. doi:10.1016/j.jenvman.2020.110402
19. Arens, K.; Liu, S. *J. Bioprocess Eng. Biorefinery* **2013,** *2,* 262–270. doi:10.1166/jbeb.2013.1060
20. González, J. M.; Weiner, R. M. *Int. J. Syst. Evol. Microbiol.* **2000,** *50,* 831–834. doi:10.1099/00207713-50-2-831

21. Howard, M. B.; Ekborg, N. A.; Taylor, L. E.; Hutcheson, S. W.; Weiner, R. M. *Protein Sci.* **2004,** *13,* 1422–1425. doi:10.1110/ps.03511604
22. Howard, M. B.; Ekborg, N. A.; Taylor, L. E., II; Weiner, R. M.; Hutcheson, S. W. *J. Bacteriol.* **2004,** *186,* 1297–1303. doi:10.1128/jb.186.5.1297-1303.2004
23. Tanaka, T.; Yan, L.; Burgess, J. G. *Curr. Microbiol.* **2003,** *47,* 412–416. doi:10.1007/s00284-003-4006-8
24. Yoon, J.-H.; Kim, H.; Kang, K. H.; Oh, T.-K.; Park, Y.-H. *Int. J. Syst. Evol. Microbiol.* **2003,** *53,* 1357–1361. doi:10.1099/ijs.0.02464-0
25. Ohta, Y.; Hatada, Y.; Nogi, Y.; Miyazaki, M.; Li, Z.; Akita, M.; Hidaka, Y.; Goda, S.; Ito, S.; Horikoshi, K. *Appl. Microbiol. Biotechnol.* **2004,** *64,* 505–514. doi:10.1007/s00253-004-1573-y
26. Ohta, Y.; Nogi, Y.; Miyazaki, M.; Li, Z.; Hatada, Y.; Ito, S.; Horikoshi, K. *Biosci., Biotechnol., Biochem.* **2004,** *68,* 1073–1081. doi:10.1271/bbb.68.1073
27. Ohta, Y.; Hatada, Y.; Nogi, Y.; Li, Z.; Ito, S.; Horikoshi, K. *Appl. Microbiol. Biotechnol.* **2004,** *66,* 266–275. doi:10.1007/s00253-004-1757-5
28. Hatada, Y.; Mizuno, M.; Li, Z.; Ohta, Y. *Mar. Biotechnol.* **2011,** *13,* 411–422. doi:10.1007/s10126-010-9312-0
29. Takagi, E.; Hatada, Y.; Akita, M.; Ohta, Y.; Yokoi, G.; Miyazaki, T.; Nishikawa, A.; Tonozuka, T. *Biosci., Biotechnol., Biochem.* **2015,** *79,* 625–632. doi:10.1080/09168451.2014.988680
30. Ohta, Y.; Hatada, Y.; Nogi, Y.; Li, Z.; Zhang, H.-M.; Ito, S.; Horikoshi, K. *J. Appl. Glycosci.* **2004,** *51,* 203–210. doi:10.5458/jag.51.203
31. Yoon, J.-H.; Kim, I.-G.; Oh, T.-K.; Park, Y.-H. *Int. J. Syst. Evol. Microbiol.* **2004,** *54,* 1111–1116. doi:10.1099/ijs.0.02985-0
32. Yoon, J.-H.; Jung, S.-Y.; Kang, S.-J.; Oh, T.-K. *Int. J. Syst. Evol. Microbiol.* **2007,** *57,* 2365–2369. doi:10.1099/ijs.0.65184-0
33. Wang, J.; Jin, L.; Wang, J.; Chan, Z.; Zeng, R.; Wu, J.; Qu, W. *Mar. Genomics* **2022,** *62,* 100931. doi:10.1016/j.margen.2022.100931
34. Tang, S.-K.; Wang, Y.; Cai, M.; Lou, K.; Mao, P.-H.; Jin, X.; Jiang, C.-L.; Xu, L.-H.; Li, W.-J. *Int. J. Syst. Evol. Microbiol.* **2008,** *58,* 2036–2040. doi:10.1099/ijs.0.65519-0
35. Jonnadula, R.; Verma, P.; Shouche, Y. S.; Ghadi, S. C. *Curr. Microbiol.* **2009,** *59,* 600–607. doi:10.1007/s00284-009-9480-1
36. Jonnadula, R.; Ghadi, S. C. *Biotechnol. Bioprocess Eng.* **2011,** *16,* 513–519. doi:10.1007/s12257-010-0399-y
37. Jonnadula, R.; Imran, M.; Poduval, P. B.; Ghadi, S. C. *Biotechnol. Rep.* **2018,** *17,* 93–96. doi:10.1016/j.btre.2017.12.008
38. Nishijima, M.; Takadera, T.; Imamura, N.; Kasai, H.; An, K.-D.; Adachi, K.; Nagao, T.; Sano, H.; Yamasato, K. *Int. J. Syst. Evol. Microbiol.* **2009,** *59,* 1696–1707. doi:10.1099/ijs.0.006452-0
39. Wang, C.-S.; Wang, Y.; Xu, X.-W.; Zhang, D.-S.; Wu, Y.-H.; Wu, M. *Int. J. Syst. Evol. Microbiol.* **2009,** *59,* 545–549. doi:10.1099/ijs.0.004531-0
40. Kim, D.-K.; Jang, Y.-R.; Kim, K.-H.; Lee, M.-N.; Kim, A.-R.; Jo, E.-J.; Byun, T.-H.; Jeong, E.-T.; Kwon, H.-J.; Kim, B.-W.; Lee, E.-W. *Fish. Aquat. Sci.* **2011,** *14,* 186–191. doi:10.5657/fas.2011.0186
41. Kobayashi, T.; Uchimura, K.; Deguchi, S.; Horikoshi, K. *Appl. Environ. Microbiol.* **2012,** *78,* 2493–2495. doi:10.1128/aem.07442-11
42. Vijayaraghavan, R.; Rajendran, S. *J. Basic Microbiol.* **2012,** *52,* 705–712. doi:10.1002/jobm.201100315
43. Wakabayashi, M.; Sakatoku, A.; Noda, F.; Noda, M.; Tanaka, D.; Nakamura, S. *Biodegradation* **2012,** *23,* 93–105. doi:10.1007/s10532-011-9489-6
44. Swift, S. M.; Hudgens, J. W.; Heselpoth, R. D.; Bales, P. M.; Nelson, D. C. *PLoS One* **2014,** *9,* e112939. doi:10.1371/journal.pone.0112939
45. Zhang, D.-S.; Huo, Y.-Y.; Xu, X.-W.; Wu, Y.-H.; Wang, C.-S.; Xu, X.-F.; Wu, M. *Int. J. Syst. Evol. Microbiol.* **2012,** *62,* 505–510. doi:10.1099/ijs.0.027714-0
46. Jeong, S. H.; Yang, S.-H.; Jin, H. M.; Kim, J. M.; Kwon, K. K.; Jeon, C. O. *Int. J. Syst. Evol. Microbiol.* **2013,** *63,* 1335–1341. doi:10.1099/ijs.0.042606-0
47. Sun, C.; Chen, Y.-j.; Zhang, X.-q.; Pan, J.; Cheng, H.; Wu, M. *Mar. Genomics* **2014,** *18,* 83–85. doi:10.1016/j.margen.2014.05.009
48. Lee, Y.-S.; Heo, J. B.; Lee, J.-H.; Choi, Y.-L. *J. Microbiol. Biotechnol.* **2014,** *24,* 925–935. doi:10.4014/jmb.1402.02033
49. Lee, Y.-S.; Park, D.-J.; Choi, Y.-L. *Appl. Microbiol. Biotechnol.* **2015,** *99,* 3901–3911. doi:10.1007/s00253-014-6186-5
50. Lee, H.-J.; Lee, Y.-S.; Choi, Y.-L. *Biotechnol. Biofuels* **2018,** *11,* 303. doi:10.1186/s13068-018-1299-1
51. Pyeon, H.-M.; Lee, Y.-S.; Choi, Y.-L. *PeerJ* **2019,** *7,* e7106. doi:10.7717/peerj.7106
52. Jeong, H.-R.; Yoo, J.-S.; Choi, Y.-L.; Jang, Y.-S.; Lee, Y.-S. *Int. J. Biol. Macromol.* **2021,** *169,* 452–462. doi:10.1016/j.ijbiomac.2020.12.138
53. Lee, Y.-S.; Choi, Y.-L. *J. Biotechnol.* **2016,** *229,* 31–32. doi:10.1016/j.jbiotec.2016.05.002
54. Kandasamy, K. P.; Subramanian, R. K.; Srinivasan, R.; Ragunath, S.; Balaji, G.; Gracy, M.; Latha, K. *Access Microbiol.* **2020,** *2,* acmi000170. doi:10.1099/acmi.0.000170
55. Camacho, M.; del Carmen Montero-Calasanz, M.; Redondo-Gómez, S.; Rodríguez-Llorente, I.; Schumann, P.; Klenk, H.-P. *Int. J. Syst. Evol. Microbiol.* **2016,** *66,* 1844–1850. doi:10.1099/ijsem.0.000955
56. Lee, C. S.; Kim, M.; Lee, C.; Yu, Z.; Lee, J. *Front. Microbiol.* **2016,** *7,* 1826. doi:10.3389/fmicb.2016.01826
57. Zhu, Y.; Wu, L.; Chen, Y.; Ni, H.; Xiao, A.; Cai, H. *Microbiol. Res.* **2016,** *182,* 49–58. doi:10.1016/j.micres.2015.09.004
58. Hu, Q.; Yin, X.; Li, H.; Wang, X.; Jiang, Z.; Li, L.; Ni, H.; Li, Q.; Zhu, Y. *Int. J. Food Sci. Technol.* **2021,** *56,* 4129–4138. doi:10.1111/ijfs.15041
59. Jiang, Z.; Long, L.; Liang, M.; Li, H.; Chen, Y.; Zheng, M.; Ni, H.; Li, Q.; Zhu, Y. *Microbiol. Res.* **2021,** *251,* 126840. doi:10.1016/j.micres.2021.126840
60. Mou, M.; Hu, Q.; Li, H.; Long, L.; Li, Z.; Du, X.; Jiang, Z.; Ni, H.; Zhu, Y. *Int. J. Mol. Sci.* **2022,** *23,* 5008. doi:10.3390/ijms23095008
61. Long, L.; Hu, Q.; Wang, X.; Li, H.; Li, Z.; Jiang, Z.; Ni, H.; Li, Q.; Zhu, Y. *Enzyme Microb. Technol.* **2022,** *161,* 110109. doi:10.1016/j.enzmictec.2022.110109
62. Jiang, Z.; Guo, Y.; Wang, X.; Li, H.; Ni, H.; Li, L.; Xiao, A.; Zhu, Y. *Protein Expression Purif.* **2019,** *161,* 17–27. doi:10.1016/j.pep.2019.03.015
63. Xiong, Q.; Wang, D.; Dong, X.; Liu, D.; Liu, Y.; Li, P.; Wu, G.; Luo, Y.; Zhang, R.; Liu, S.; Zhang, G. *Int. J. Syst. Evol. Microbiol.* **2019,** *69,* 1135–1141. doi:10.1099/ijsem.0.003279
64. Imran, M.; Pant, P.; Shanbhag, Y. P.; Sawant, S. V.; Ghadi, S. C. *Mar. Biotechnol.* **2017,** *19,* 116–124. doi:10.1007/s10126-017-9737-9
65. Lee, J.-Y.; Kim, P. S.; Hyun, D.-W.; Kim, H. S.; Shin, N.-R.; Jung, M.-J.; Yun, J.-H.; Kim, M.-S.; Whon, T. W.; Bae, J.-W. *Int. J. Syst. Evol. Microbiol.* **2017,** *67,* 998–1004. doi:10.1099/ijsem.0.001731

66. Moh, T. H.; Furusawa, G.; Amirul, A. A.-A. *Int. J. Syst. Evol. Microbiol.* **2017,** *67,* 4089–4094. doi:10.1099/ijsem.0.002258
67. Diyana, T.; Furusawa, G. *J. Basic Microbiol.* **2021,** *61,* 1124–1132. doi:10.1002/jobm.202100198
68. Moh, T. H.; Lau, N.-S.; Furusawa, G.; Amirul, A.-A. A. *Stand. Genomic Sci.* **2017,** *12,* 36. doi:10.1186/s40793-017-0248-0
69. Park, S.; Yoon, S. Y.; Ha, M.-J.; Yoon, J.-H. *Int. J. Syst. Evol. Microbiol.* **2017,** *67,* 1436–1441. doi:10.1099/ijsem.0.001831
70. Su, Q.; Jin, T.; Yu, Y.; Yang, M.; Mou, H.; Li, L. *AMB Express* **2017,** *7,* 220. doi:10.1186/s13568-017-0525-8
71. Yang, M.; Yu, Y.; Yang, S.; Shi, X.; Mou, H.; Li, L. *Front. Microbiol.* **2018,** *9,* 2894. doi:10.3389/fmicb.2018.02894
72. Yang, M.; Yu, Y.; Jin, T.; Mou, H.; Li, L. *Mar. Genomics* **2018,** *39,* 7–10. doi:10.1016/j.margen.2017.07.003
73. Won, N.-I.; Lee, G.-E.; Ko, K.; Oh, D.-C.; Na, Y. H.; Park, J.-S. *Microbiol. Biotechnol. Lett.* **2017,** *45,* 124–132. doi:10.4014/mbl.1702.02002
74. Liu, H.; Zeng, L.; Jin, Y.; Nie, K.; Deng, L.; Wang, F. *Appl. Biochem. Biotechnol.* **2019,** *188,* 741–749. doi:10.1007/s12010-018-02948-1
75. Sugimoto, Y.; Ohnishi, K.-i.; Suzuki, S. *Microbiol. Resour. Announce.* **2020,** *9,* e00746–20. doi:10.1128/mra.00746-20
76. Jung, J.; Bae, S. S.; Chung, D.; Baek, K. J. *J. Microbiol. (Seoul, Repub. Korea)* **2018,** *54,* 299–301. doi:10.7845/kjm.2018.8047
77. Li, R.-K.; Chen, Z.; Ying, X.-J.; Ng, T. B.; Ye, X.-Y. *Int. J. Biol. Macromol.* **2018,** *119,* 1164–1170. doi:10.1016/j.ijbiomac.2018.08.053
78. Li, R. K.; Ying, X. J.; Chen, Z. L.; Ng, T. B.; Zhou, Z. M.; Ye, X. Y. *Catalysts* **2020,** *10,* 885. doi:10.3390/catal10080885
79. Li, R. K.; Hu, Y. J.; He, Y. J.; Ng, T. B.; Zhou, Z. M.; Ye, X. Y. *Biotechnol. Appl. Biochem.* **2021,** *68,* 1076–1085. doi:10.1002/bab.2027
80. Ma, J.; Yan, Q.; Yi, P.; Yang, S.; Liu, H.; Jiang, Z. *Process Biochem. (Oxford, U. K.)* **2019,** *87,* 119–127. doi:10.1016/j.procbio.2019.08.021
81. Zhu, Y.; Gao, H.; Li, H.; Ni, H.; Jiang, Z.; Li, L.; Xiao, A. *Acta Oceanol. Sin.* **2019,** *38,* 96–106. doi:10.1007/s13131-019-1349-y
82. Hakamada, Y.; Syouyama, S. *J. Biol. Macromol.* **2020,** *20,* 9–16. doi:10.14533/jbm.20.9
83. Huang, H.; Mo, K.; Hu, Y.; Liu, M.; Zhu, J.; Zou, X.; Bao, S. *Int. J. Syst. Evol. Microbiol.* **2020,** *70,* 1639–1643. doi:10.1099/ijsem.0.003945
84. Yang, J.; Cui, D.; Chen, D.; Chen, W.; Ma, S.; Shen, H. *Mar. Drugs* **2020,** *18,* 184. doi:10.3390/md18040184
85. Anggraeni, S. R.; Ansorge-Schumacher, M. B. *Mar. Biotechnol.* **2021,** *23,* 809–820. doi:10.1007/s10126-021-10065-0
86. Anggraeni, S. R.; Ansorge-Schumacher, M. B. *Squalen Bull. Mar. Fish. Postharvest Biotechnol.* **2022,** *17,* 65–73. doi:10.15578/squalen.653
87. Cheng, Y.; Zhu, S.; Guo, C.; Xie, F.; Jung, D.; Li, S.; Zhang, W.; He, S. *Antonie van Leeuwenhoek* **2021,** *114,* 1033–1042. doi:10.1007/s10482-021-01574-y
88. Gu, X.; Gui, Y.; Li, J.; Zhang, X.; Zilda, D. S. *J. Ocean Univ. China* **2021,** *20,* 181–188. doi:10.1007/s11802-021-4506-9
89. Zhang, Q.; Gui, Y.; Zhao, L.; Zhang, A.; Fu, L.; Cao, Z.; Li, J. *J. Sci. Food Agric.* **2023,** *103,* 6095–6104. doi:10.1002/jsfa.12705
90. Mah, M. H.; Lam, M. Q.; Tokiman, L.; Kamaroddin, M. F.; Ibrahim, Z.; Shahir, S.; Chong, C. S. *Arabian J. Sci. Eng.* **2022,** *47,* 6795–6805. doi:10.1007/s13369-021-06400-1
91. Park, S. L.; Cho, J. Y.; Kim, S. H.; Bhatia, S. K.; Gurav, R.; Park, S.-H.; Park, K.; Yang, Y.-H. *Polymers (Basel, Switz.)* **2021,** *13,* 4257. doi:10.3390/polym13234257
92. Cho, J. Y.; Kim, S. H.; Jung, H. J.; Cho, D. H.; Kim, B. C.; Bhatia, S. K.; Ahn, J.; Jeon, J.-M.; Yoon, J.-J.; Lee, J.; Yang, Y.-H. *Polymers (Basel, Switz.)* **2022,** *14,* 3625. doi:10.3390/polym14173625
93. Park, S. L.; Cho, J. Y.; Kim, S. H.; Lee, H.-J.; Kim, S. H.; Suh, M. J.; Ham, S.; Bhatia, S. K.; Gurav, R.; Park, S.-H.; Park, K.; Kim, Y.-G.; Yang, Y.-H. *J. Microbiol. Biotechnol.* **2022,** *32,* 27–36. doi:10.4014/jmb.2109.09005
94. Tanaka, D.; Ohnishi, K.-i.; Watanabe, S.; Suzuki, S. *J. Gen. Appl. Microbiol.* **2021,** *67,* 47–53. doi:10.2323/jgam.2020.05.001
95. Wang, D.; Wang, J.; Shui, B.; Zhu, L.; Wang, J.; Jin, L.; Qu, W. *Arch. Microbiol.* **2021,** *203,* 6143–6151. doi:10.1007/s00203-021-02578-y
96. Yang, J.; Cui, D.; Ma, S.; Chen, W.; Chen, D.; Shen, H. *Int. J. Biol. Macromol.* **2021,** *169,* 551–563. doi:10.1016/j.ijbiomac.2020.12.196
97. Romanenko, L.; Kurilenko, V.; Otstavnykh, N.; Velansky, P.; Isaeva, M.; Mikhailov, V. *Arch. Microbiol.* **2022,** *204,* 548. doi:10.1007/s00203-022-03161-9
98. Yan, F.; Chen, J.; Cai, T.; Zhong, J.; Zhang, S. *J. Sci. Food Agric.* **2022,** *102,* 4522–4531. doi:10.1002/jsfa.11807
99. Zhu, X.; Chen, S.; Luo, G.; Zheng, W.; Tian, Y.; Lei, X.; Yao, L.; Wu, C.; Xu, H. *Microbiol. Spectrum* **2022,** *10,* e00934–21. doi:10.1128/spectrum.00934-21
100. Deutsch, J. M.; Green, M. O.; Akavaram, P.; Davis, A. C.; Diskalkar, S. S.; Du Plessis, I. A.; Fallon, H. A.; Grason, E. M.; Kauf, E. G.; Kim, Z. M.; Miller, J. R., II; Neal, A. L.; Riera, T.; Stroeva, S.-E.; Tran, J.; Tran, V.; Coronado, A. V.; Coronado, V. V.; Wall, B. T.; Yang, C. m.; Mohanty, I.; Abrahamse, N. H.; Freeman, C. J.; Easson, C. G.; Fiore, C. L.; Onstine, A. E.; Djeddar, N.; Biliya, S.; Bryksin, A. V.; Garg, N.; Agarwal, V. *Mar. Drugs* **2023,** *21,* 53. doi:10.3390/md21010053
101. Duckworth, M.; Yaphe, W. *Carbohydr. Res.* **1971,** *16,* 189–197. doi:10.1016/s0008-6215(00)86113-3
102. Fu, X. T.; Kim, S. M. *Mar. Drugs* **2010,** *8,* 200–218. doi:10.3390/md8010200
103. Ha, J.-C.; Kim, G.-T.; Kim, S.-K.; Oh, T.-K.; Yu, J.-H.; Kong, I.-S. *Biotechnol. Appl. Biochem.* **1997,** *26,* 1–6. doi:10.1111/j.1470-8744.1997.tb00439.x
104. Vera, J.; Alvarez, R.; Murano, E.; Slebe, J. C.; Leon, O. *Appl. Environ. Microbiol.* **1998,** *64,* 4378–4383. doi:10.1128/aem.64.11.4378-4383.1998
105. Zhong, Z.; Toukdarian, A.; Helinski, D.; Knauf, V.; Sykes, S.; Wilkinson, J. E.; O'Bryne, C.; Shea, T.; DeLoughery, C.; Caspi, R. *Appl. Environ. Microbiol.* **2001,** *67,* 5771–5779. doi:10.1128/aem.67.12.5771-5779.2001
106. Chi, W.-J.; Chang, Y.-K.; Hong, S.-K. *Appl. Microbiol. Biotechnol.* **2012,** *94,* 917–930. doi:10.1007/s00253-012-4023-2
107. Allouch, J.; Jam, M.; Helbert, W.; Barbeyron, T.; Kloareg, B.; Henrissat, B.; Czjzek, M. *J. Biol. Chem.* **2003,** *278,* 47171–47180. doi:10.1074/jbc.m308313200
108. Allouch, J.; Helbert, W.; Henrissat, B.; Czjzek, M. *Structure* **2004,** *12,* 623–632. doi:10.1016/j.str.2004.02.020

109. Hehemann, J.-H.; Correc, G.; Thomas, F.; Bernard, T.; Barbeyron, T.; Jam, M.; Helbert, W.; Michel, G.; Czjzek, M. *J. Biol. Chem.* **2012,** *287,* 30571–30584. doi:10.1074/jbc.m112.377184
110. Gacesa, P. *Int. J. Biochem.* **1992,** *24,* 545–552. doi:10.1016/0020-711x(92)90325-u
111. Wong, T. Y.; Preston, L. A.; Schiller, N. L. *Annu. Rev. Microbiol.* **2000,** *54,* 289–340. doi:10.1146/annurev.micro.54.1.289
112. Yang, M.; Li, N.; Yang, S.; Yu, Y.; Han, Z.; Li, L.; Mou, H. *Appl. Microbiol. Biotechnol.* **2019,** *103,* 807–817. doi:10.1007/s00253-018-9502-7
113. Yang, M.; Yang, S.-X.; Liu, Z.-M.; Li, N.-N.; Li, L.; Mou, H.-J. *Mar. Drugs* **2019,** *17,* 378. doi:10.3390/md17060378
114. Steinfeld, L.; Vafaei, A.; Rösner, J.; Merzendorfer, H. Chitin Prevalence and Function in Bacteria, Fungi and Protists. In *Targeting chitin-containing organisms;* Yang, Q.; Fukamizo, T., Eds.; Springer: Singapore, 2019; Vol. 1142, pp 19–59. doi:10.1007/978-981-13-7318-3_3
115. Tharanathan, R. N.; Kittur, F. S. *Crit. Rev. Food Sci. Nutr.* **2003,** *43,* 61–87. doi:10.1080/10408690390826455
116. Nguyen, S. T. C.; Freund, H. L.; Kasanjian, J.; Berlemont, R. *Appl. Microbiol. Biotechnol.* **2018,** *102,* 1629–1637. doi:10.1007/s00253-018-8778-y
117. Howard, M. B.; Ekborg, N. A.; Taylor, L. E.; Weiner, R. M.; Hutcheson, S. W. *J. Bacteriol.* **2003,** *185,* 3352–3360. doi:10.1128/jb.185.11.3352-3360.2003
118. Ekborg, N. A.; Gonzalez, J. M.; Howard, M. B.; Taylor, L. E.; Hutcheson, S. W.; Weiner, R. M. *Int. J. Syst. Evol. Microbiol.* **2005,** *55,* 1545–1549. doi:10.1099/ijs.0.63627-0
119. Kloareg, B.; Quatrano, R. *Oceanogr. Mar. Biol.* **1988,** *26,* 259–315.
120. Talarico, L. B.; Zibetti, R. G. M.; Faria, P. C. S.; Scolaro, L. A.; Duarte, M. E. R.; Noseda, M. D.; Pujol, C. A.; Damonte, E. B. *Int. J. Biol. Macromol.* **2004,** *34,* 63–71. doi:10.1016/j.ijbiomac.2004.03.002
121. Grassauer, A.; Weinmuellner, R.; Meier, C.; Pretsch, A.; Prieschl-Grassauer, E.; Unger, H. *Virol. J.* **2008,** *5,* 107. doi:10.1186/1743-422x-5-107
122. Srivastava, N.; Rathour, R.; Jha, S.; Pandey, K.; Srivastava, M.; Thakur, V. K.; Sengar, R. S.; Gupta, V. K.; Mazumder, P. B.; Khan, A. F.; Mishra, P. K. *Biomolecules* **2019,** *9,* 220. doi:10.3390/biom9060220
123. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009,** *37,* D233–D238. doi:10.1093/nar/gkn663
124. Haag, T.; Loncrini, D. *Cosmet. Sci. Technol. Ser.* **1984,** *1,* 63–77.
125. Shende, V. V.; Bauman, K. D.; Moore, B. S. *Nat. Prod. Rep.* **2024,** *41,* 604–648. doi:10.1039/d3np00037k
126. Leutou, A. S.; Yang, I.; Kang, H.; Seo, E. K.; Nam, S.-J.; Fenical, W. *J. Nat. Prod.* **2015,** *78,* 2846–2849. doi:10.1021/acs.jnatprod.5b00746
127. Akasaka, K.; Meguro, H.; Ohrui, H. *Tetrahedron Lett.* **1997,** *38,* 6853–6856. doi:10.1016/s0040-4039(97)01616-x
128. Kim, Y.; Ogura, H.; Akasaka, K.; Oikawa, T.; Matsuura, N.; Imada, C.; Yasuda, H.; Igarashi, Y. *Mar. Drugs* **2014,** *12,* 4110–4125. doi:10.3390/md12074110
129. Choi, S. Y.; Yoon, K.-h.; Lee, J. I.; Mitchell, R. J. *BioMed Res. Int.* **2015,** 465056. doi:10.1155/2015/465056
130. Balibar, C. J.; Walsh, C. T. *Biochemistry* **2006,** *45,* 15444–15457. doi:10.1021/bi061998z
131. Füller, J. J.; Röpke, R.; Krausze, J.; Rennhack, K. E.; Daniel, N. P.; Blankenfeldt, W.; Schulz, S.; Jahn, D.; Moser, J. *J. Biol. Chem.* **2016,** *291,* 20068–20084. doi:10.1074/jbc.m116.741561
132. Chevrette, M. G.; Gutiérrez-García, K.; Selem-Mojica, N.; Aguilar-Martínez, C.; Yañez-Olvera, A.; Ramos-Aboites, H. E.; Hoskisson, P. A.; Barona-Gómez, F. *Nat. Prod. Rep.* **2020,** *37,* 566–599. doi:10.1039/c9np00048h
133. Seymour, J. R.; Amin, S. A.; Raina, J.-B.; Stocker, R. *Nat. Microbiol.* **2017,** *2,* 17065. doi:10.1038/nmicrobiol.2017.65
134. Meyer, N.; Bigalke, A.; Kaulfuß, A.; Pohnert, G. *FEMS Microbiol. Rev.* **2017,** *41,* 880–899. doi:10.1093/femsre/fux029
135. Fehér, D.; Barlow, R.; McAtee, J.; Hemscheidt, T. K. *J. Nat. Prod.* **2010,** *73,* 1963–1966. doi:10.1021/np100506z
136. Isnansetyo, A.; Kamei, Y. *Antimicrob. Agents Chemother.* **2003,** *47,* 480–488. doi:10.1128/aac.47.2.480-488.2003
137. Malmberg, T.; Athanasiadou, M.; Marsh, G.; Brandt, I.; Bergman, Å. *Environ. Sci. Technol.* **2005,** *39,* 5342–5348. doi:10.1021/es050574+
138. Agarwal, V.; El Gamal, A. A.; Yamanaka, K.; Poth, D.; Kersten, R. D.; Schorn, M.; Allen, E. E.; Moore, B. S. *Nat. Chem. Biol.* **2014,** *10,* 640–647. doi:10.1038/nchembio.1564
139. El Gamal, A.; Agarwal, V.; Diethelm, S.; Rahman, I.; Schorn, M. A.; Sneed, J. M.; Louie, G. V.; Whalen, K. E.; Mincer, T. J.; Noel, J. P.; Paul, V. J.; Moore, B. S. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* 3797–3802. doi:10.1073/pnas.1519695113
140. Agarwal, V.; Blanton, J. M.; Podell, S.; Taton, A.; Schorn, M. A.; Busch, J.; Lin, Z.; Schmidt, E. W.; Jensen, P. R.; Paul, V. J.; Biggs, J. S.; Golden, J. W.; Allen, E. E.; Moore, B. S. *Nat. Chem. Biol.* **2017,** *13,* 537–543. doi:10.1038/nchembio.2330
141. Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006,** *106,* 3468–3496. doi:10.1021/cr0503097
142. Zhao, M.; Lin, H.-C.; Tang, Y. *J. Antibiot.* **2016,** *69,* 571–573. doi:10.1038/ja.2016.33
143. Dalsgaard, P. W.; Larsen, T. O.; Frydenvang, K.; Christophersen, C. *J. Nat. Prod.* **2004,** *67,* 878–881. doi:10.1021/np0303714
144. Dalsgaard, P. W.; Larsen, T. O.; Christophersen, C. *J. Antibiot.* **2005,** *58,* 141–144. doi:10.1038/ja.2005.16
145. Xu, W.; Gavia, D. J.; Tang, Y. *Nat. Prod. Rep.* **2014,** *31,* 1474–1487. doi:10.1039/c4np00073k
146. Ióca, L. P.; Dai, Y.; Kunakom, S.; Diaz-Espinosa, J.; Krunic, A.; Crnkovic, C. M.; Orjala, J.; Sanchez, L. M.; Ferreira, A. G.; Berlinck, R. G. S.; Eustáquio, A. S. *Angew. Chem., Int. Ed.* **2021,** *60,* 15891–15898. doi:10.1002/anie.202017320
147. 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.; van Wezel, G. P.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2023,** *51,* W46–W50. doi:10.1093/nar/gkad344
148. Ritzmann, N. H.; Mährlein, A.; Ernst, S.; Hennecke, U.; Drees, S. L.; Fetzner, S. *Environ. Microbiol.* **2019,** *21,* 2595–2609. doi:10.1111/1462-2920.14654
149. Borchardt, S. A.; Allain, E. J.; Michels, J. J.; Stearns, G. W.; Kelly, R. F.; McCoy, W. F. *Appl. Environ. Microbiol.* **2001,** *67,* 3174–3179. doi:10.1128/aem.67.7.3174-3179.2001
150. Syrpas, M.; Ruysbergh, E.; Stevens, C. V.; De Kimpe, N.; Mangelinckx, S. *Beilstein J. Org. Chem.* **2014,** *10,* 2539–2549. doi:10.3762/bjoc.10.265

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.146

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Methyltransferases from RiPP pathways: shaping the landscape of natural product chemistry

Maria-Paula Schröder‡, Isabel P.-M. Pfeiffer‡ and Silja Mordhorst*

Review

Open Access

Address:
Pharmaceutical Institute, Department of Pharmaceutical Biology, University of Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany

Email:
Silja Mordhorst* - silja.mordhorst@uni-tuebingen.de

* Corresponding author ‡ Equal contributors

Keywords:
biocatalysis; methylation; post-translational modifications; ribosomal peptides; SAM-dependent enzymes

*Beilstein J. Org. Chem.* **2024,** *20,* 1652–1670.
https://doi.org/10.3762/bjoc.20.147

Received: 15 February 2024
Accepted: 28 June 2024
Published: 18 July 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf

© 2024 Schröder et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

This review article aims to highlight the role of methyltransferases within the context of ribosomally synthesised and post-translationally modified peptide (RiPP) natural products. Methyltransferases play a pivotal role in the biosynthesis of diverse natural products with unique chemical structures and bioactivities. They are highly chemo-, regio-, and stereoselective allowing methylation at various positions. The different possible acceptor regions in ribosomally synthesised peptides are described in this article. Furthermore, we will discuss the potential application of these methyltransferases as powerful biocatalytic tools in the synthesis of modified peptides and other bioactive compounds. By providing an overview of the various methylation options available, this review is intended to emphasise the biocatalytic potential of RiPP methyltransferases and their impact on the field of natural product chemistry.

## Introduction

In the complex landscape of natural product biosynthesis, ribosomally synthesised and post-translationally modified peptides (RiPPs) stand out as a fascinating class of compounds with both structural diversity and unique bioactivities [1]. RiPP natural products are typically encoded in a biosynthetic gene cluster (BGC) and produced following defined biosynthetic rules. Initially, a precursor peptide encompassing a core peptide, a leader part and/or a follower part, is synthesised at the ribosome. Next, distinct maturases encoded in the same BGC as the precursor peptide install post-translational modifications in the core peptide. Finally, a protease releases the modified core peptide, creating the mature natural product [2].

The transfer of a methyl group is a common post-translational modification in RiPP biosynthesis; this reaction is catalysed by *S*-adenosylmethionine (SAM or AdoMet)-dependent methyl-

transferases (MTs). These enzymes are known for their excellent chemo-, regio-, and stereoselectivity [3-6]. Their ability to introduce methyl groups at specific positions within ribosomal peptides equips the resulting natural products with a spectrum of chemical complexity, contributing to the broad range of biological activities exhibited by RiPPs.

Bioactivities attributed to RiPPs include a wide range of effects, such as antibiotic, antifungal, antiviral, antiparasitic, antitumour, analgesic, anti-inflammatory, antidiabetic, antihypertensive, and anti-parkinsonian [7,8]. This diverse profile makes RiPPs particularly valuable for various therapeutic and medicinal applications. Generally, peptide natural products are promising drug development candidates due to their intermediate molecular weight, which falls between small molecules and large biologicals such as antibodies. Small molecules typically weigh less than 900 Dalton (Da), while the molecular weight of large biologicals ranges from approximately 5,000 to 150,000 Da. Although some small RiPPs may be classified as small molecules, the majority of RiPPs have a molecular weight between 1,000 and 5,000 Da. Peptide natural products exhibit high specificity and binding affinity to their corresponding targets [9,10]. The inhibition of protein–protein interactions is an emerging strategy in the development of novel therapeutics [11]. Binding to proteins allows peptides to disturb protein–protein interactions, which is a challenging task for small molecules.

The principles of RiPP biosynthesis make the RiPP technology an ideal platform for the generation of designed peptides and allow tailored derivatisation of these peptides. The RiPP technology can also be employed to mimic non-ribosomal peptides [12,13], since it is substantially more challenging to engineer non-ribosomal peptide synthetase (NRPS) pathways. Non-ribosomal peptides are often produced by giant multi-modular enzyme complexes (type I NRPS) and they are most commonly involved in the specialised metabolism of bacteria and fungi. Each NRPS module is minimally composed of a condensation (C) domain, an adenylation (A) domain, and a peptidyl carrier protein (PCP or thiolation/T domain), and one module typically incorporates one amino acid into the final natural product. In contrast to ribosomal peptide biosynthesis, many non-proteinogenic amino acids can be incorporated as building blocks. However, RiPP pathways are more flexible and they represent interesting biotechnological production routes.

In this review, we will focus on post-translational methylation reactions. A single methyl group has the ability to modulate a drug's efficacy by orders of magnitude; this occurrence is also called the 'magic methyl effect' [14,15]. Compared to chemical methylation reactions, a biocatalytic methylation often offers several advantages. MTs exhibit high selectivity and are environmentally friendly, whereas chemical methylating agents are usually toxic and unselective [3]. Nonetheless, a drawback is that MTs rely on the costly cofactor SAM, which is required in stoichiometric amounts. Furthermore, it is unstable in physiological conditions and produces *S*-adenosylhomocysteine (SAH) as a co-product, which strongly inhibits various MTs [16-19]. To overcome these limitations, numerous in situ SAM supply and SAM regeneration systems have been developed [4,20-26], thereby enabling preparative biocatalytic scale methylation reactions [27]. Such regeneration systems have been widely applied to small molecules [28-30], as well as polyketide natural products [31], and RNA [32]. To the best of our knowledge, these systems have not yet been applied to peptide natural products.

By providing a comprehensive overview of the various methylation options available within RiPP biosynthetic pathways, this review aims to emphasise the biocatalytic potential of RiPP MTs and their transformative impact on the field of natural product chemistry. Instead of providing an exhaustive list of all identified RiPP MTs to date, we will focus on selected examples for varying acceptor regions to demonstrate the broad range of methylated positions in ribosomally synthesised peptides (Figure 1). This article also highlights the potential application of RiPP MTs in biocatalysis and their ability to enable the production of bioactive compounds with tailored chemical functionalities.

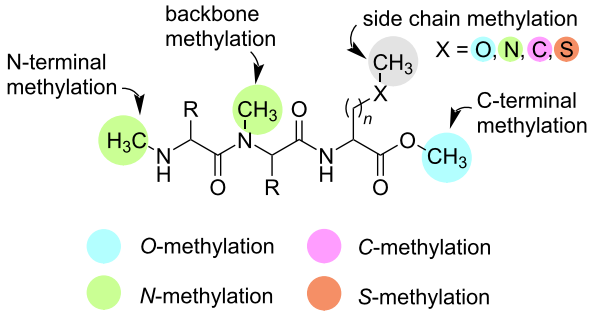



**Figure 1:** Schematic representation of the different acceptor regions for the methylation of RiPPs discussed in this article.

# Review

## Methylation strategies

The main section of this paper discusses different RiPP MTs involved in the biosynthesis of methylated peptide natural products. This section serves as an introduction and outlines additional synthetic strategies for methylated amino acids and peptides. There are three main routes for the synthesis of

methylated peptides: chemical synthesis [33,34], in vitro ribosomal synthesis [35], and enzymatic synthesis.

**Chemical synthesis**. Three main categories of reactions are commonly used: reductive amination, reductive ring openings, and the use of methylating agents [34]. In reductive amination, the substrate is usually an aldehyde or amine. After the formation of the iminium ion, it is reduced with the appropriate reagent to form the *N*-methylated amino acid. Different methods have been established using for example benzaldehyde as a protection group, sodium cyanoborohydride as a mild reducing agent, and paraformaldehyde as a methylating agent [36]. Methanol can be used as the methylating reagent in other methods. Here, a palladium on carbon (Pd/C) catalyst processes the dehydrogenation of the alcohol to form the corresponding aldehyde. The subsequently formed imine intermediate is then reduced with $H_2$ [37]. When synthesising via a reductive ring opening, typically, an *N*-Fmoc-protected amino acid is condensed with formaldehyde in the presence of *p*-toluic acid in refluxing toluene to yield the 5-oxazolidinone. Reductive ring opening can be achieved by using an excess of triethylsilane in TFA-$CH_3Cl$, resulting in the *N*-methylated amino acid as the final product [38]. The third method for chemical *N*-methylation involves the use of protection groups that also enhance the reactivity of the primary amine (Figure 2). Once the amine is deprotonated, an electrophilic methylation reagent donates its methyl group to yield the mono-*N*-methylated amino acid. Common reagents for this reaction include iodomethane (MeI), cyanamide ($CH_2N_2$), and dimethyl sulphate ($(CH_3O)_2SO_2$). A similar logic can be followed to methylate the $C_\alpha$-carbon, with reaction conditions ranging from mild to very harsh. Additionally, stereoselectivity can be achieved by using potassium bis(trimethylsilyl)amide for deprotonation and methylation with MeI [33,39].

Any of these routes would yield methylated amino acids that can then be incorporated into an oligopeptide in solid-phase peptide synthesis (SPPS). Additionally, *N*-methylation concurrently to peptide synthesis is also possible. Commonly used methods for SPPS of *N*-methylated peptides involve activating the amino group with 2-nitrobenzenesulfonamide and methylating agents, such as dimethyl sulphate [40-42]. Directly

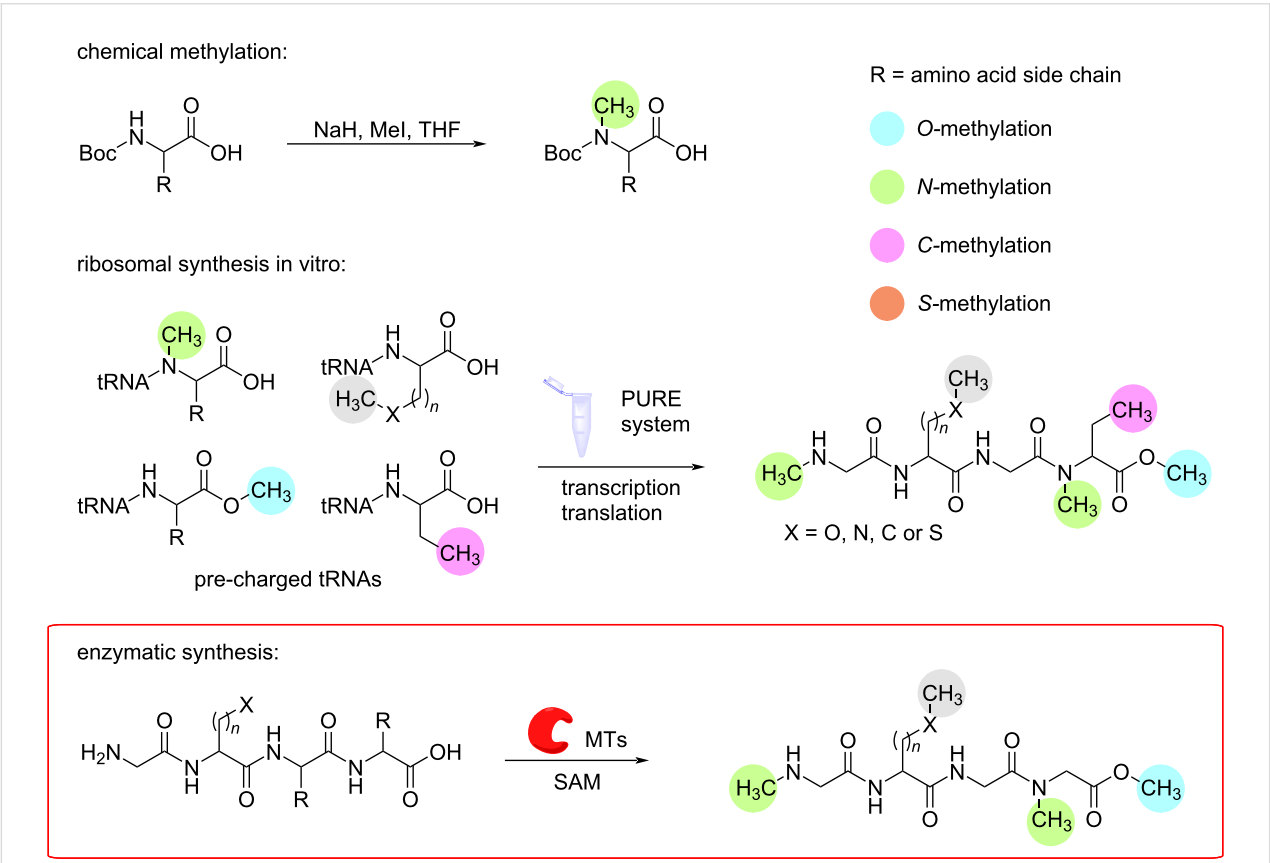



**Figure 2:** Schematic overview of different methylation strategies for amino acids and peptides. There are several options for chemical methylation (see text), but only one example is shown. Enzymatic synthesis is highlighted in red as this is the focus of this highlight. PURE = protein synthesis using recombinant elements, MT = methyltransferase.

methylating peptides is challenging. Multiple amide bonds and primary amines of the amino acid chains are available for methylation, leading to poor site selectivity of any methylating reaction [39]. Therefore, either pre-methylated amino acids or short methylated peptides are incorporated into the full-length peptide. Direct *O*-methylation of an *O*-desmethyl coibamide A in solution has been demonstrated by Sable et al. [43]. The use of diazomethane in diethyl ether as methylating agent yielded the methylated synthetic coibamide A. All of the aforementioned methods are robust and well-established. However, they do have their drawbacks. The application of toxic agents such as MeI, TFA, or sodium cyanoborohydride can be harmful to both health and environment. Additionally, the need for constant protection and deprotection steps is time-consuming and results in low atomic economy, while also increasing the use of solvents, energy, and chemicals. Furthermore, side reaction such as racemisation, fragmentation, deletions, and double or triple methylations can lead to lower yields and hinder the coupling of further amino acids after *N*-methylation [44].

**Ribosomal synthesis in vitro**. A cell-free translation system termed "protein synthesis using recombinant elements" (PURE) was developed by Shimizu et al. [45]. PURE contains all the necessary components of the *Escherichia coli* ribosomal machinery, such as the ribosome, initiation, elongation, and release factors, and pre-charged tRNAs. In the first step, mRNA or a plasmid containing the desired sequence is introduced into the system. After transcription and translation, the final peptide is purified and can be used for downstream applications (Figure 2) [45,46]. The incorporation of methylated amino acids relies upon the supply of the correct tRNAs. In this case, the tRNAs are prepared in advance and are not recycled in the system. The amino acids are methylated either prior to tRNA binding or after. The efficiency of incorporating unnatural amino acids is lower, but Merryman and Green have demonstrated that an excess of tRNA can overcome this problem [47]. When attempting to incorporate both methylated and unmethylated versions of the same amino acid, simply supplying both tRNAs is not sufficient, as the ribosome has a higher selectivity for the tRNA charged with the natural amino acid. This can be addressed by synthesising tRNA for a stop codon charged with the desired modified amino acid. This approach is limited to the addition of two unnatural amino acids, as the third stop codon is needed for termination. Another option is the reassignment of sense codons and supplying the desired chemically charged tRNAs. This has been demonstrated to work for up to 12 unnatural amino acid analogues [48-52]. The PURE system is a well-established method for rapidly expressing functional proteins. However, synthesising peptides with unnatural amino acids remains a labour-intensive process. tRNAs charged with unnatural or modified amino acids cannot be easily replenished within the system and must be prepared beforehand. The incorporation of unnatural amino acids by the ribosome is less efficient and leads to incomplete translation, resulting in shunt products. Additionally, ribosome recycling efficiency and synthesis of full length proteins declines with length of the primary sequence; Li et al. demonstrated this for proteins of up to 224 kDa [35]. And even when full-length peptides are synthesised, they may not always be active due to the absence of chaperones that aid in the proper folding of the nascent peptide chain [35,44].

**Enzymatic synthesis**. Another option is enzymatic peptide methylation. Biocatalysis provides an efficient and sustainable alternative to chemical synthesis strategies. Both NRPS and RiPP biosynthetic machineries can be used for cell-based or in vitro strategies [53]. Especially well suited as biocatalysts are MTs from RiPP pathways, as they demonstrate high substrate flexibility.

SAM is the universal methyl donor of biological methylation reactions, and it is the main outcome of the methionine cycle. Following transfer of the methyl group, SAH is formed, which is then hydrolysed to L-homocysteine and adenosine by an SAH hydrolase. L-Homocysteine is remethylated to L-methionine by the enzyme methionine synthase, which utilises 5-methyltetrahydrolate (5-MTHF) as a methyl donor. 5-MTHF is a constituent of the folate cycle, which encompasses the intermediates tetrahydrofolate (THF) and 5,10-methylene-THF (5,10-$CH_2$-THF). 5,10-$CH_2$-THF is formed from THF and L-serine, which provides the methyl group in this case. Another methyl donor that provides a C1 building block for L-methionine and subsequent SAM synthesis is betaine. Betaine-homocysteine MT remethylates L-homocysteine to L-methionine. Non-SAM-dependent MTs catalyse $B_{12}$-dependent methylations using the cofactors adenosylcobalamin (AdoCbl) and methylcobalamin (MeCbl). In such biological processes, MTHF can serve as a methyl donor, but also smaller molecules such as methanol and methanethiol [54,55]. Further low-molecular methyl donors are known from biological systems, including chloromethane ($CH_3Cl$), bromomethane ($CH_3Br$) [56], and *S*-methylmethionine [57] (Figure 3).

As mentioned above, RiPP MTs use SAM as a methyl donor in their enzymatic methylation reactions. The following section provides an overview of various RiPP maturases that methylate different functional groups.

## Methyltransferases from RiPP pathways

Methyltransferases can be classified based on various factors, such as their substrates (small molecule MTs, protein MTs, or RNA/DNA MTs), the atom that accepts the methyl group

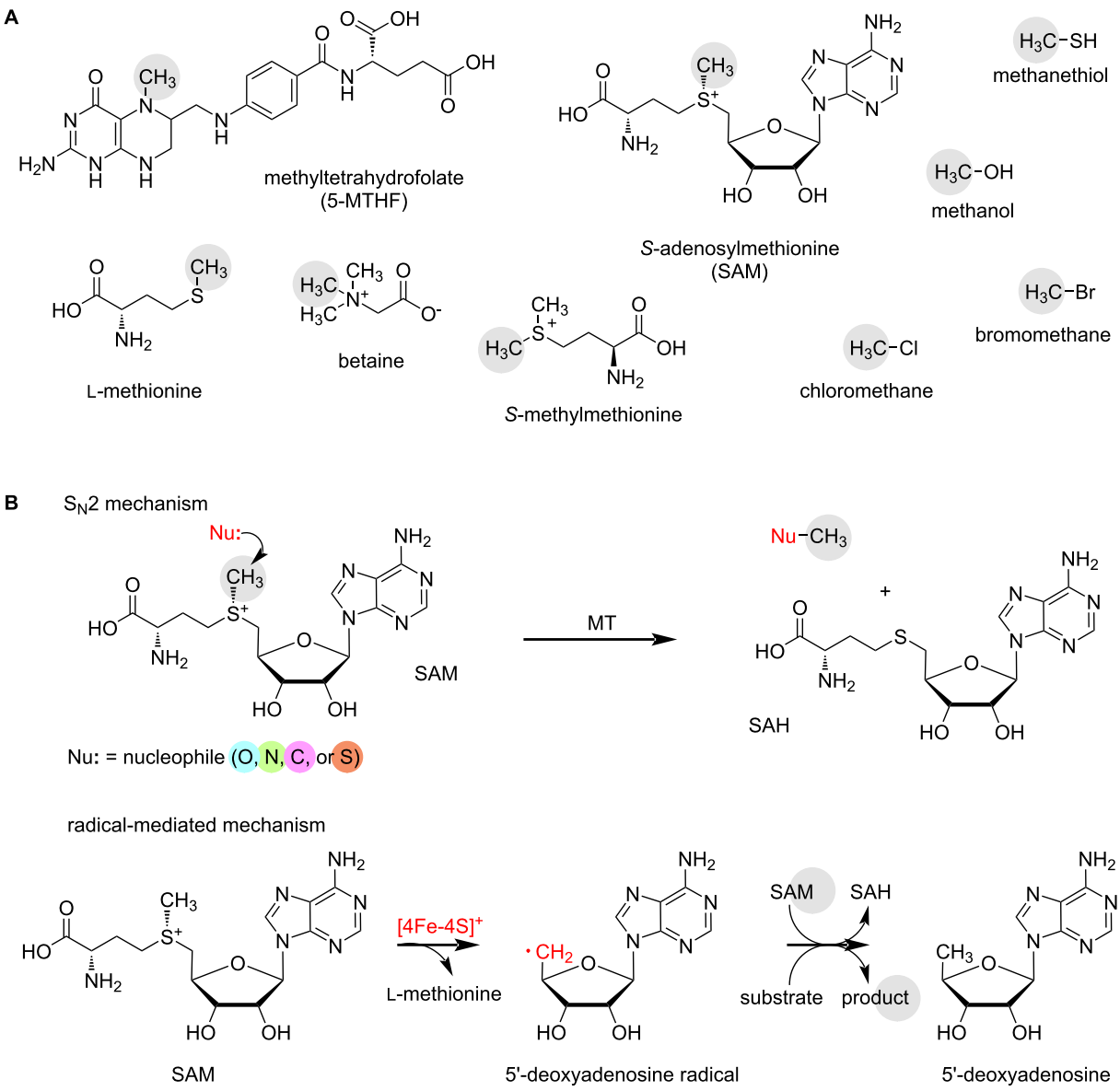



**Figure 3:** Biological methylation. A) Methyl donors from biological systems. The transferred methyl group is highlighted in grey. B) General reaction schemes of biological methylation reactions.

(oxygen = *O*-MTs, nitrogen = *N*-MTs, carbon = *C*-MTs, sulphur = *S*-MTs, or halide = H-MTs), metal or cofactor dependence ($Mg^{2+}$-dependent, metal-independent, cobalamin-dependent), common structural folds (class I–V, with class I being the largest group, characterised by the Rossmann fold) [58], or catalytic mechanism ($S_N2$ mechanism, radical mechanism, Figure 3) [59]. This review categorises RiPP MTs based on the acceptor atom, describing *O*-, *N*-, *C*-, and *S*-MTs; halide MTs have not (yet) been identified in RiPP pathways. The enzymes described below are either conventional SAM-dependent MTs or radical SAM (rSAM) MTs; rSAM MTs are one subfamily of the large rSAM enzyme superfamily, which encompasses enzymes catalysing a broad range of reactions [60]. There are different rSAM MT classes, description of the classes follows below in the *C*-MT section. Unlike conventional MTs, which follow an $S_N2$ mechanism, rSAM MTs can methylate non-activated carbons.

## *O*-Methyltransferases

The predominant group of MTs are *O*-MTs, which perform oxygen-directed methylation [59]. A range of natural products are known to undergo *O*-methylation [61], improving mem-

brane permeability, absorption, and oral bioavailability [61,62]. The most extensively studied *O*-MTs are those that methylate catecholic hydroxy groups [63]. Several *O*-methylation modifications have been observed during the maturation of different RiPPs.

One example is the biosynthesis of thiopeptide GE2270, which is produced by the actinomycete *Planobispora rosea* ATCC53733 as a complex of related metabolites that differ in the number of methyl groups installed on the macrocyclic core. The *pbt* cluster, responsible for the synthesis of GE2270, encodes four MTs. Among these, only PbtM4 is an *O*-MT. It methylates thiazole D, resulting in a *C*-methylene-*O*-methylthiazole. The other MTs of the *pbt* cluster, one *N*-MT (PbtM1) and two *C*-MTs (PbtM2 and PbtM3) [64], will be discussed in later sections.

*Myxococcus xanthus* DK1622 and other strains of *M. xanthus* produce the peptide cittilin A (Figure 4). The novel RiPP class of cittilins was established based on its unique bicyclic structure. Experimental evidence suggests that *O*-methylation occurs in the last biosynthetic step by the MT CitC [65]. Further *O*-MTs can be classified based on a common acceptor group of modification, namely C-terminal carboxy-MTs and peptide/protein L-aspartyl *O*-MTs (PAMTs).

## C-terminal carboxymethyltransferases

Carboxy-MTs catalyse the formation of a methyl ester by methylating the oxygen atom within the hydroxy group of a carboxylic acid. SAM is used as a donor of the methyl group [67-69]. The linear cyanobactin aeruginosamide B, produced by *Microcystis aeruginosa* PCC 9432, was the first RiPP reported to be *O*-methylated at its C-terminus [70].

$LahS_B$ is a SAM-dependent MT encoded in the *lah* BGC of *Lachnospiraceae bacterium* C6A11. It belongs structurally to class I MTs and was the first MT observed acting on the C-terminus of lanthipeptides. Furthermore, it is the second occurrence of an *O*-MT acting on lanthipeptides, alongside the $LanS_A$-type *O*-MTs [71]. $LahS_B$ methylates four out of seven LahA precursor peptides at the C-terminal carboxy group of different amino acids, namely methionine, isoleucine, and valine. This demonstrates its tolerance to various substrates. $LahS_B$ does not require the leader peptide or other modifying enzymes to carry out its function [72], making it a promising tool for biocatalytic applications. The co-crystallisation of $LahS_B$ with bound SAH provides important details about the structure–function relationship, the substrate–enzyme interaction, and the cofactor binding site (Figure 5). This structural information, including the residues involved in binding, is essential for enzyme engineering and rational design [73,74].

*Lentzea kentuckyensis* sp. produces lassomycin (Figure 4), a lasso peptide, with a methyl ester at the C-terminus. The *las* cluster encodes LasF, which is predicted to function as an *O*-MT. Since no other putative MT is encoded in the *las* cluster, LasF is proposed to methylate lassomycin [77].

## PAMTs – peptide/protein L-aspartyl *O*-methyltransferases

PAMTs share a conserved C-terminal domain, substantial sequence homology, and a common enzymatic activity. They catalyse the conversion between aspartate and isoaspartate with aspartyl-*O*-methyl ester and aspartimide intermediates (Figure 6). The activity of most PAMTs depends on a cyclised precursor peptide. Modification of the aspartate/isoaspartate residue was found only in a hairpin-like or the macrocyclic region [71,78]. The underlying purpose of this structural requirement of PAMTs has not yet been demonstrated.

$OlvS_A$, an *O*-MT encoded in the *olv* cluster from *Streptomyces olivaceus* NRRL B-3009 (Figure 5), is involved in the biosynthesis of the class I lanthipeptide $OlvA(BCS_A)$. It catalyses the conversion of a highly conserved L-aspartate to L-isoaspartate by methylation of a side chain carboxylate group of L-aspartate, resulting in an L-aspartyl-*O*-methyl ester. The L-succinimide intermediate that is formed is non-enzymatically hydrolysed to the final isoaspartate. Experimental proof demonstrated that $OlvS_A$ operates without the need for leader peptide recognition; instead, it relies on an already cyclised core peptide to exert its effects on the precursor peptide. Furthermore, SAM was shown to be an obligate co-substrate for $OlvS_A$ [71]. Crystallographic data of $OlvS_A$ is not yet available. The predicted three-dimensional structure displays substantial differences from the crystal structure of $LahS_B$ [RMSD: 8.152 Å (109 to 109 atoms)], demonstrating the structural diversity among *O*-MTs (Figure 5).

*Streptomyces bottropensis* produces various bottromycins, including bottromycin A2 (Figure 4) [79]. The precursor peptide consists of an N-terminal core and a rare C-terminal follower region. The biosynthesis of bottromycin involves several methylations, catalysed by four different MTs encoded in the BGC. Gomez-Escribano et al. named the BGC in which the MTs are encoded *bmb* and proposed their putative function [80]. Huo et al. named the same BGC *bot* and experimentally determined the function of the MTs. BmbA/BotOMT is the only *O*-MT involved in the biosynthesis of bottromycin [81]. It *O*-methylates the aspartate by transforming its carboxylic acid group to a methyl ester in the ultimate step in biosynthesis [79]. The remaining methylations are catalysed by *C*-MTs, which will be addressed in a following section.

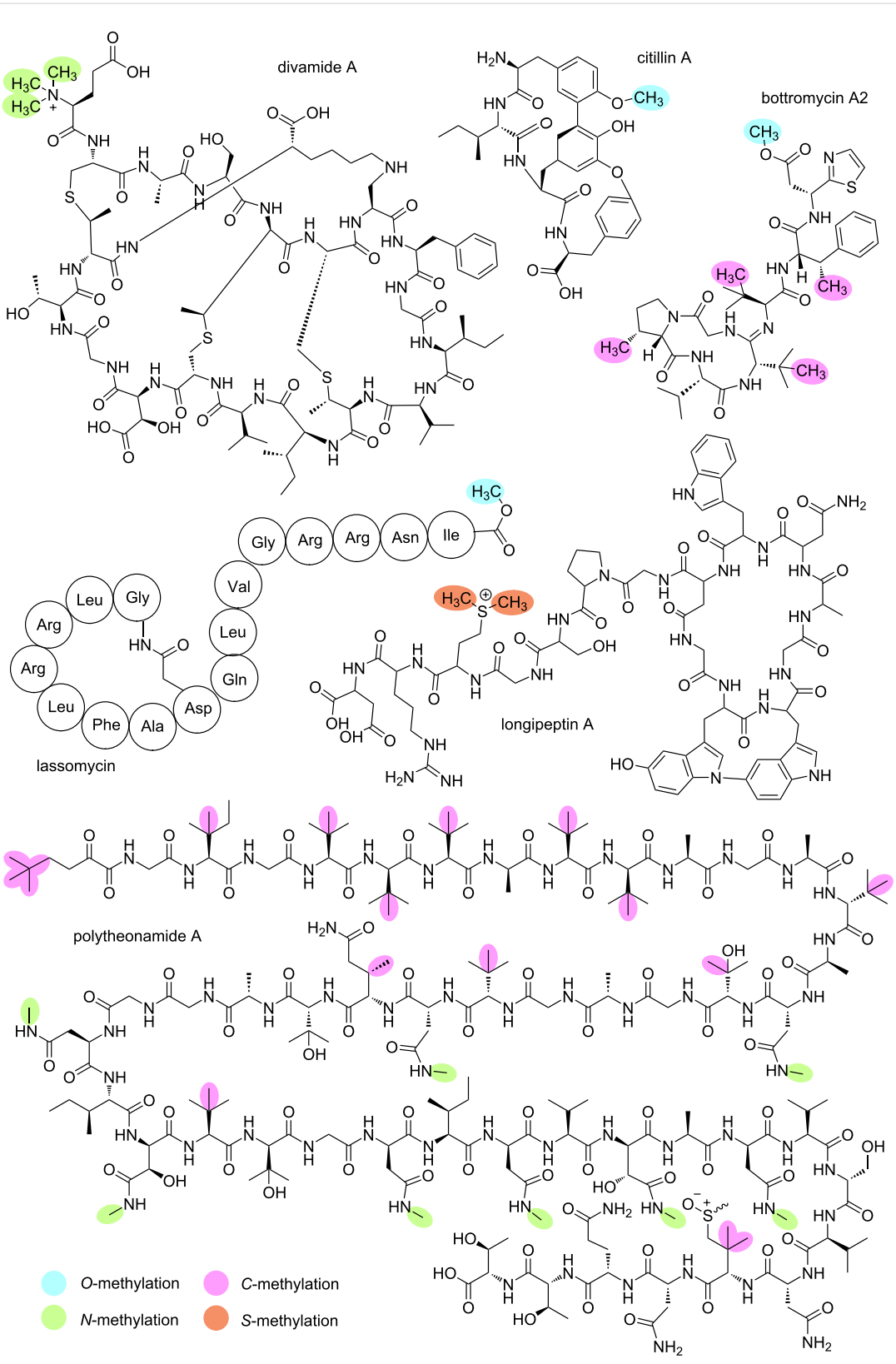



**Figure 4:** Chemical structures of RiPPs with diverse *O*-, *N*-, *C*-, and *S*-methylations. Amino acids of lassomycin are shown in the three-letter code. For longipeptin A, the stereochemistry has not been fully determined yet, one of two proposed positional isomers of the biaryl link between Trp2 and Trp3 is depicted [66].

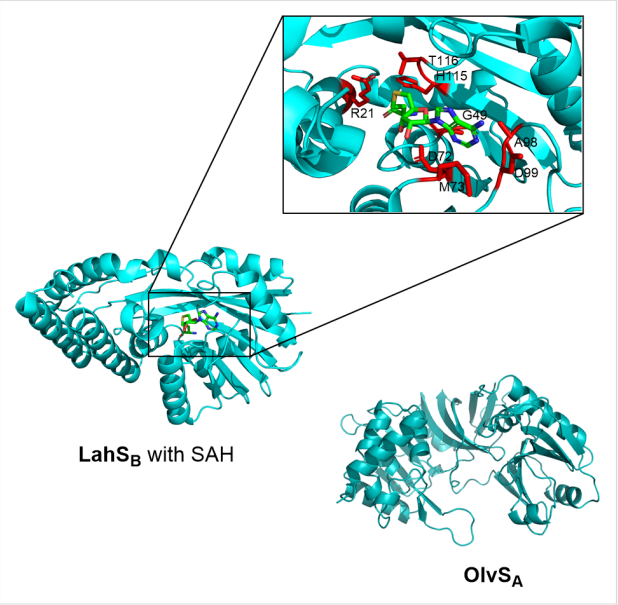



**Figure 5:** The three-dimensional structures of the conventional *O*-MTs $OlvS_A$ (model structure calculated by Colabfold [75] and visualised by PyMOL [76]) and $LahS_B$ with bound SAH, PDB ID: 6UAK (https://doi.org/10.2210/pdb6UAK/pdb), [72]. The residues involved in SAH binding are depicted in red.

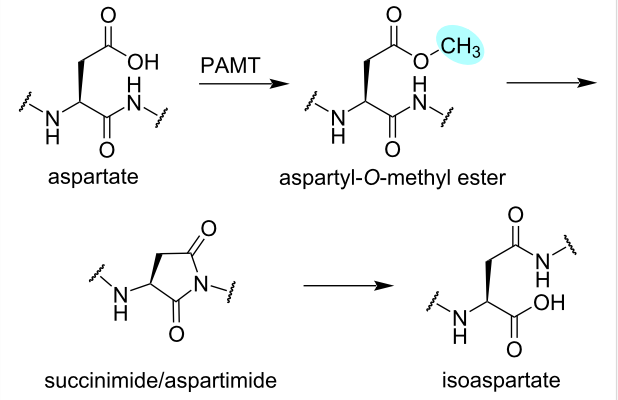



**Figure 6:** Reaction scheme of the PAMT´s catalysis, leading to the enzymatic conversion of aspartate to aspartyl-*O*-methyl ester, followed by a two-step non-enzymatic conversion via succinimide/aspartimide to isoaspartate. The methyl group is highlighted in blue. The reverse reaction with isoaspartate as substrate of PAMTs is also known.

## *N*-Methyltransferases

*N*-Methylations of peptides offer several advantages for peptide therapeutics, including conformational modulation, influence on receptor subtype specificity, increased proteolytic resistance, and improved oral bioavailability and cell permeability [82,83]. *N*-Methylation can occur at the N-terminal amide, the peptide backbone, as well as on amino acid side chain amides (Figure 1). While traditional synthesis of *N*-methylated peptides is challenging and time consuming, installation via RiPP maturases provides an easier, quicker, and greener approach to confer these attributes to different peptide natural products. Although certain *N*-methylations may affect selectivity and bioactivity, a study on somatostatin derivatives showed that *N*-methylating the backbone improved oral bioavailability without affecting biological activity and selectivity [84].

### α-*N*-Methyltransferases/ *N*-terminal methyltransferases

The α-amino group of N-terminal amino acid residues can either be mono-, di-, or trimethylated. This methylation step occurs after the core peptide is cleaved from the leader and can enhance the metabolic stability of the final peptide natural product [85].

Crocagin A is the first RiPP to be isolated from Myxobacteria. It was specifically isolated from *Chondromyces crocatus* Cm C5. The core peptide is composed of only three amino acids: isoleucine, tyrosine, and tryptophan. It forms a tetracyclic core system that contains a tetrahydropyrroloindoline, which is unprecedented in bacterial natural products. Crocagin biosynthesis starts with formation of the central ring structure, followed by cleavage of the core peptide by peptidase CgnD. The final steps involve *N*-methylation by CgnL and carbamoylation by CgnI. It is unclear which modification is installed first. CgnL is a class I SAM-dependent MT. The closest structural homologues with defined biosynthetic pathways are CcbJ and KedS8. These enzymes *N*-methylate the antibiotic lincosamide celesticetin and the anticancer protein kedaricidin, respectively. Crocagin A is a potent inhibitor of the RNA binding protein CsrA, which is a promising target for new anti-infective compounds. However, drug development is hampered by poor cellular uptake [86,87].

In the biosynthesis of cypemycin in *Streptomyces* sp. OH-4156, CypM adds two methyl groups to the N-terminal alanine residue. Cypemycin has a very narrow activity spectrum and loses all activity when unmethylated. Although the mode of action is unclear, the methylations are essential to the bioactivity. CypM methylates many oligopeptides that mimic the cypemycin N-terminal sequence, and unrelated peptide antibiotics. Additionally, CypM methylates short oligopeptides with N-terminal alanine, serine, threonine, glycine, and methionine residues, as well as the antibiotics nisin and haloduracin with good efficiency. Nisin was methylated on its N-terminal isoleucine and on the side chain of Lys12, where two methylations occurred. It is important to note that N-terminally methylated nisin retains its antibiotic activity, with MIC (minimal inhibitory concentration) values of 1 μM and 4 μM against *Lactococcus lactis* and *Bacillus subtilis*, respectively. Haloduracin is a two-component lanthibiotic that is methylated at both cyclised N-termini by CypM. These experiments demonstrate

that CypM has a relaxed substrate specificity, which makes it interesting for bioengineering applications [88].

Other examples of terminal *N*-methylation can be found in the dikaritins, a family of fungal RiPPs. Among the first members of this family to be discovered were the ustiloxins. Ustiloxins were first identified in 1994, but their BGC was not identified until 2013 [89,90]. The biosynthesis of ustiloxin starts with the UstA precursor, which contains a 16-fold repeat of the Tyr-Ala-Ile-Gly core. Following the cleavage of the precursor into tetrapeptides, the tetrapeptides are cyclised via the tyrosine and isoleucine side chains, and then *N*-methylation is installed by UstM on the N-terminal tyrosine residue [91,92]. Another member of the dikaritin family is phomopsin. All phomopsins contain at least one *N*-methylated tyrosine. In vivo studies have attributed the MT activity to PhonM, and in vitro studies have also demonstrated MT activity on phomopsin A, resulting in the dimethylated phomopsin E [93].

In the biosynthesis of plantazolicin, PznL (BamL) catalyses a reaction that is highly analogous to that of PhonM, involving the α-*N*-dimethylation of arginine. Plantazolicin was isolated from *Bacillus amyloliquefaciens* FZB42 and it mainly exhibits bactericidal activity against *Bacillus anthracis* [94-96]. The dimethylation is essential for its bioactivity; desmethyl plantazolicin showed MIC values >128 µg/mL for *B. anthracis*, while the dimethyl species showed MIC values between 2–4 µg/mL [96]. In contrast to CypM, PznL has higher substrate specificity towards its peptide substrate. It required more than just an N-terminal arginine as a recognition element. PznL cannot accept peptide substrates that lack the polyheterocyclic peptide backbone modifications of plantazolicin [94].

Divamide A (Figure 4) was first isolated from a marine tunicate, *Didemnum molle*, extract, based on its anti-HIV activity in a bioassay-guided fractionation. Further investigation on the identification of the BGC revealed that the tunicate symbiont, *Prochloron didemni*, is actually responsible for producing divamide A. Divamide A contains a trimethylated glutamate at the N-terminus (Figure 4), which is installed by DivMT as the final step in divamide A biosynthesis. The presence of the cyclic lysinoalanine moiety is required for this post-translational modification. The addition of other N-terminal amino acids is not tolerated by DivMT. It appears that the MT has a high specificity for the overall scaffold of divamide A [97].

## Backbone-*N*-methyltransferases

Borosins are a particular interesting RiPP family, and the first family containing maturases capable of backbone-*N*-methylation. Omphalotin A is the founding member of this RiPP class. It is biosynthesised by *Omphalotus olearius* and has nematocidal properties. In types I through III of borosins, the precursor peptide is fused to the MT domain; OphMA designates the fused precursor-MT from the omphalotin A BGC (Figure 7). After auto-modifying the core, the precursor is cleaved and the MT domain is presumably degraded without catalysing any further methylating reactions [98]. The OphMA fusion functions as a homodimer, with the active site of the interlocking

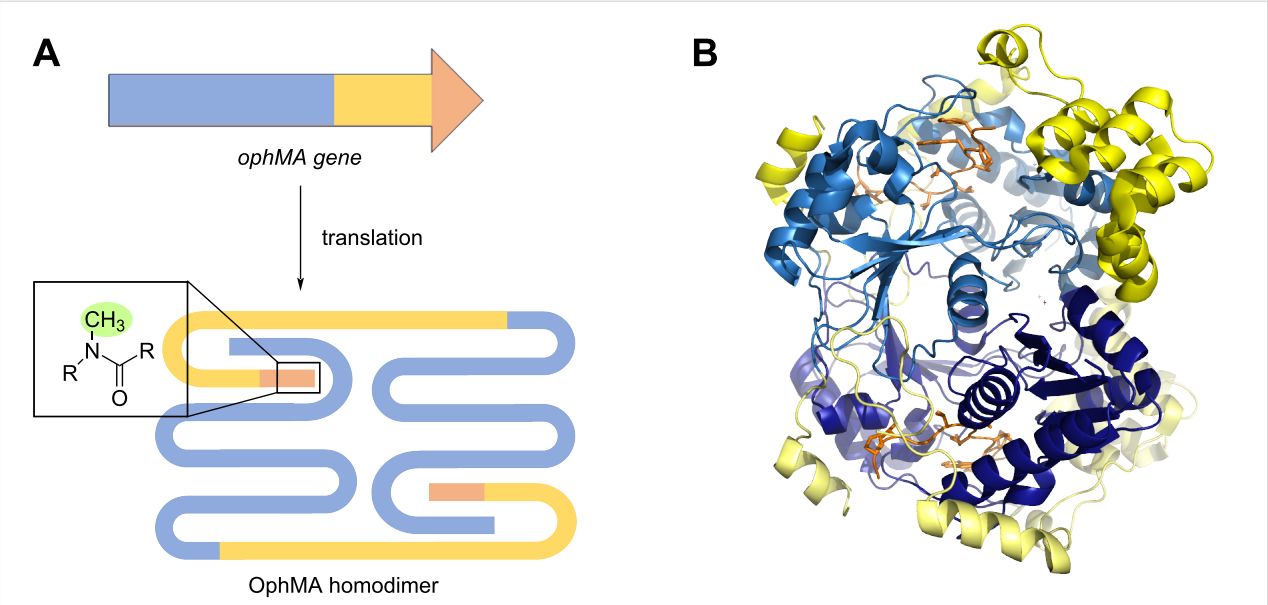



**Figure 7:** Structural organisation of the OphMA homodimer. A) Schematic representation. The MT domain is coloured in blue; the leader peptide is coloured in yellow, and the core peptide is coloured in orange. Methylation of the backbone is highlighted in green. B) Three-dimensional structure of the OphMA homodimer, PDB ID: 6TCS (https://doi.org/10.2210/pdb6tsc/pdb) [99].

dimers modifying the substrate peptide (Figure 7) [99]. The MT domain of OphMA is especially interesting due to its relaxed substrate specificity. In experiments where the precursor was replaced with unrelated peptide natural products amino acid sequences, such as cyclosporin and dictyonamide, the MT domain methylated the backbone up to 5 times, but the pattern did not correspond to that of the natural modified core peptide of the precursor domain. A subsequent study demonstrated that the MT domain can also *N*-methylate the amide bonds of non-natural amino acids fused to the already folded MT domain of OphMA. These results show that the MT domain of OphMA is highly permissive in terms of core peptide sequences [100,101].

Type IV borosins are more distantly related homologues found in bacteria; although it was originally believed that borosins occur mainly in fungi, Cho et al. and Imani et al. both demonstrated that type IV borosin BGCs are widespread in bacteria and have additionally defined classes V through X (Figure 8) [102,103]. The type IV MT SonM was first described and characterised by Miller et al. using *Shewanella oneidensis* MR-1 as a model system [104]. While the MT still acts as a homodimer, the precursor SonA bears a borosin binding domain, which allows SonM-SonA dimers to arrange into a heterotetramer for methylation of the backbone. SonM installs two methylations on the peptide backbone of SonA at Leu63 and Ile65. The borosin binding domain could potentially be exploited for engineering of other precursors and synthesising of novel peptides [104]. It was demonstrated that SahM from *Shewanella* sp. HN-41 has a substrate flexibility similar to that of the MT domain of OphMA, thus expanding the scope of type IV borosins. SahM was able to methylate different alanine-substituted precursors, as well as completely unnatural core sequences of OphMA and cyclosporin A comparable to the native precursor.

Over 50 putative borosin MTs have been identified in a study by Quijano et al. [105]. Heterologous expression of 11 precursor–MT pairs highlighted different methylation patterns. All of the MTs tested appear to methylate in an N-to-C-terminal direction on primarily hydrophobic core peptides [105]. If other MTs display similarly lax specificity to their core peptides as the MT domain of OphMA and SahM, a varied toolbox of MTs could be envisioned for use in the synthesis of peptide therapeutics [102].

EreM is a backbone-MT which does not belong to the borosin family. It is part of the pythonamide biosynthesis in *Candidatus* Eudoremicrobium spp. Pythonamide belongs to the proteusin RiPP family, it has a 46-amino acid core peptide on which up to six *N*-methylations are installed. EreM is an FkbM-like MT and catalyses this reaction on different valine residues of the core peptide. While the first methylation can be installed

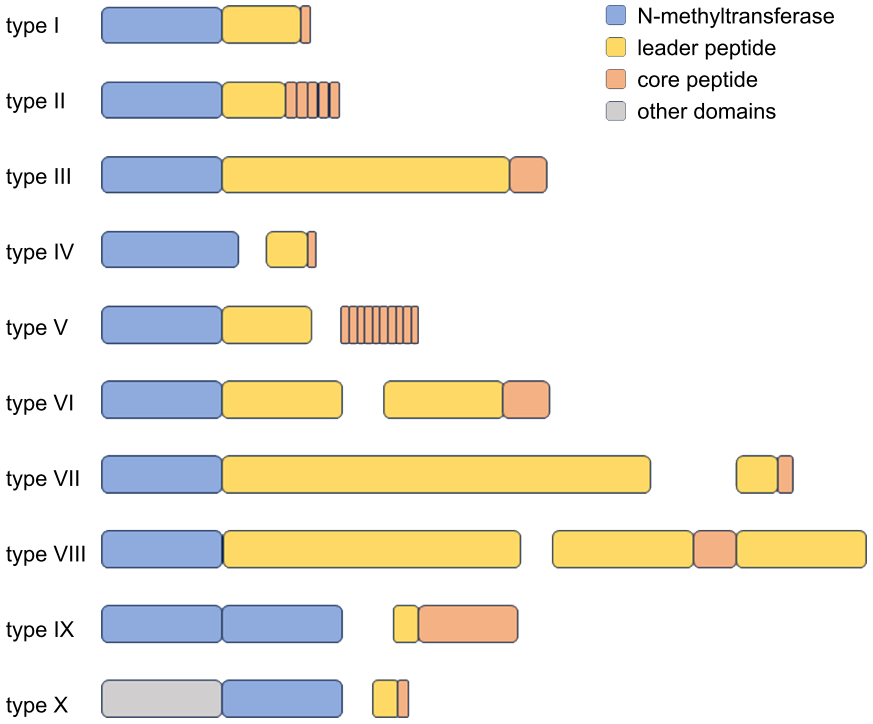



**Figure 8:** Overview of the protein architectures and core peptide compositions of borosin *N*-MTs as defined by Imani et al. Types I–III are of fungal origin, types IV–X are of bacterial origin [103].

independently of other modifications, the other five are only installed when EreM is co-expressed with the oxygenase EreI [106].

### *N*-Methylation on amino acid side chains

The polytheonamides are heavily modified peptides that belong to the proteusin family of RiPPs. They are produced by the tunicate symbiont *Candidatus* Entotheonella factor and exhibit membrane activity against Gram-positive bacteria and are highly cytotoxic. Cytotoxicity was tested with murine leukaemia cells; polytheonamide A, B, and C showed $IC_{50}$ values between 68 and 78 pg/mL [107-109]. While originally thought to be of NRPS origin due to their extensive modifications, Freeman et al. discovered in 2012 that polytheonamides are actually of ribosomal origin [108]. Polytheonamides A and B contain eight methylated asparagine residues (Figure 4), all of which are installed by the MT PoyE. Interestingly, two of these methylated asparagine residues also carry a hydroxy group at the β-carbon, which does not appear to affect the methylations. Renevey and Riniker demonstrated that the *N*-methylated asparagine residues of polytheonamide B are essential for the stability of its β-helical structure; reversion of the *N*-methylation led to complete conformation loss. The methylated asparagine residues at positions 16, 22, and 28 form an intramolecular chain of hydrogen bonds along the outside of the β-helix. Hydrogen atoms from the amide methyl group form hydrogen bonds with the oxygen of the neighbouring *N*-methylated asparagine amino acid side chain. This chain of hydrogen bonds stabilises the β-helical structure of polytheonamide B [110]. PoyE acts independently of the dehydratase and epimerase activities of other maturases in the cluster, but the MT is more efficient on D-asparagine compared to L-asparagine. PoyE exhibits a degree of tolerance towards precursor variations. The insertion of a GNINVN motif into the precursor, which represents one turn of the helical structure, did not affect the activity of PoyE: this elongated precursor was methylated by PoyE with comparable efficiency to the natural precursor. However, simple swapping of amino acids on the precursor sequence yielded mixed results. These findings suggest that PoyE has a plastic yet regiospecific activity that could potentially be exploited in bioengineering applications [111].

Aeronamide A is another member of the channel-forming, cytotoxic proteusins. It is biosynthesised by *Microvirgula aerodenitrificans*. Like polytheonamide B, aeronamide A shows cytotoxic activity against HeLa cells, with an $IC_{50}$ of 1.48 nM. AerE, the PoyE homologue, installs five methylations on the amides of Asn25, Asn31, Asn37, Asn41, and Asn43 (putative localisation) [112]. The activity of AerE is independent of the leader peptide; nevertheless, it only methylates D-asparagine residues. Additionally, AerE is independent of secondary structure, in co-crystallisation experiments the precursor is bound in an extended, unstructured configuration to AerE [113]. Otherwise, the biosynthetic machinery of aeronamide A seems to be permissive to other core sequences. Swapping the core sequence with other proteusin core structures, such as the polytheonamides core and a *Rhodospirillaceae* sp. proteusin precursor, demonstrated a relaxed substrate specificity for all maturases in the *aer* BGC: All hybrid precursors were extensively modified with dehydrations, *C*-methylation, *N*-methylations, and epimerisations comparable to the AerA precursor [112,113].

PbtM1 from the thiopeptide GE2270 BGC is another example of a side chain amide-*N*-MT [114]. PbtM1 methylates Asn4 in the GE2270 precursor peptide. Substrate specificity appears to be pronounced, as amino acid substitutions in the core were not tolerated by the biosynthetic machinery. PtbM1 acts independently of other modifications since linear intermediates after enzyme deletion almost always show Asn4 methylation. Further investigation into the substrate scope of PbtM1 would be interesting to determine its potential in peptide engineering [64].

An example for lysine *N*-methylation can be found in epichloëcyclins. This family of RiPPs is found in endophytic fungi of the genus *Epichloë*. Epichloëcyclins A–E are cyclic heptameric peptides, all produced from the same precursor GigA. They carry a dimethylated lysine residue; this modification is installed as a final step in the biosynthesis by GigC. Genome mining showed that analogues of the GigA precursor are widespread in fungal genomes, exploration of these BGCs may lead to the discovery of further *N*-MTs [115,116].

## *C*-Methyltransferases

*C*-MTs are less prevalent than *O*- or *N*-directed MTs; they transfer a methyl group to a carbon atom [117]. *C*-Methylation of drugs can enhance their potency enormously [14]. Several *C*-MTs are involved in RiPP maturation, all *C*-MTs described in the following section are classified as rSAM *C*-MTs.

### Radical SAM *C*-methyltransferases

Radical SAM enzymes typically contain a conserved CxxxCxxC motif. The cysteine residues of this motif coordinate a [4Fe-4S] cluster. rSAM MTs can be classified based on their domain structure. Class A rSAM MTs hold solely an rSAM domain, while classes B and C each contain an additional functional domain [118]. The activity of class B rSAM MTs is $B_{12}$-/cobalamin-dependent. The proposed structure of $B_{12}$/rSAM-MTs consists of a didomain protein consisting of a cobalamin binding domain (CBD) at the N-terminus and an rSAM domain at the C-terminus (Figure 9) [119].

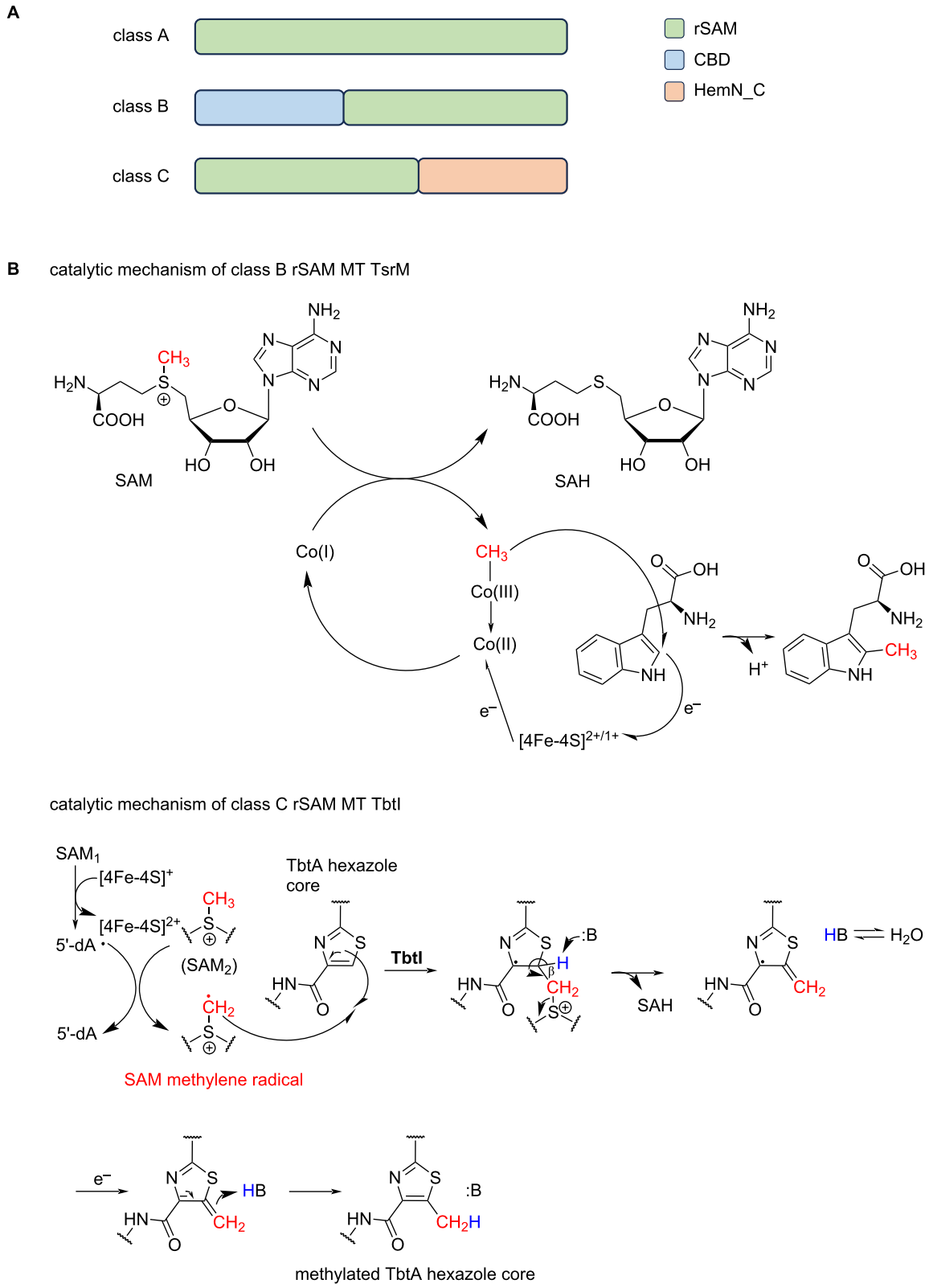



**Figure 9:** Radical SAM *C*-methyltransferases. A) The different rSAM MT classes containing different functional domains. Class B includes a cobalamin binding domain (CBD) and class C a domain similar to HemN_C, which is an oxygen-independent coproporphyrinogen-III oxidase (CPO). B) Catalytic mechanisms of rSAM-mediated methylation are depicted exemplary for class B (TsrM) and class C (TbtI) RiPP MTs.

In addition to its different enzyme structure, a distinct mechanism of methylation than the classical $S_N2$ mechanism is employed. Reductive cleavage of SAM generates a 5'-deoxyadenosyl (dAdo) radical, which putatively produces a radical intermediate by abstracting a hydrogen atom from the peptide substrate. Methylation occurs by reaction of methylcob(III)alamin (MeCbl) with the substrate. Then, cob(II)alamin is released, reduced to cob(I)alamin, and methylated by a second SAM molecule to restore MeCbl [118,120-122].

The biosynthesis of bottromycin involves various *C*-methylations catalysed by three different enzymes, in addition to the *O*-methylation mentioned in a previous section. Specifically, phenylalanine is methylated by BmbB/BotRMT1, two of the three valines present are methylated by BmbF/BotRMT2, and proline is methylated by BmbJ/BotRMT3 (Figure 4) [81]. Three derivatives of bottromycin were obtained through fermentation of the natural producer. The proline residue methylation pattern varies, with bottromycin A2 being monomethylated, bottromycin B2 being unmethylated, and bottromycin C2 being dimethylated [81,123].

Polytheonamides are originally derived from the marine sponge *Theonella swinhoei*, which is colonised by various symbiotic bacteria. The polytheonamide encoding BGC *poy* includes two $B_{12}$-dependent rSAM *C*-MTs, namely PoyB and PoyC. In silico modelling and alignment of the structures of PoyB and PoyC indicate their high structural similarity with a RMSD of 0.907 Å (552 to 552 atoms) and a sequence similarity of 64.56% (Figure 10). PoyC is the first enzyme in the class B rSAM MTs to perform $C_\beta$-methylations, and exclusive methylates L-configured amino acids. Furthermore, PoyC contains a unique Cx7Cx2C rSAM motif. In the polytheonamide biosynthesis, PoyC methylates Ile3 and valine 5, 10, and 21. The activity of PoyC and PoyB is spatially separated, as PoyC only operates on residues 1 to 21, while PoyB operates only on residues 23 to 49. PoyB methylates Val31 and Met45, albeit with low efficiency. PoyB was proposed to produce the N-terminal *tert*-butyl group of the polytheonamide A (Figure 4), but later PoyC was demonstrated to be responsible for this modification [111,124]. This structural feature conveys hydrophobicity to the N-terminus and is in line with the decreasing hydrophobicity towards the C-terminus, named "axial amphipathicity". Thus, this methylation is thought to play an important role in correct target membrane insertion and the associated cytotoxic effect [125]. TsrM is another cob(II)alamin-dependent *C*-MT (Figure 10). This enzyme is involved in the biosynthesis of thiostrepton A, which is known to be produced by different *Streptomyces* strains [126,127]. During the maturation of thiostrepton A, a quinaldic moiety is formed from tryptophan. TsrM catalyses the initial step by transferring a methyl group from SAM to the electrophilic carbon atom C2 of tryptophan. In the process of biosynthesis, the quinaldic moiety is eventually integrated into the precursor peptide TsrA. Besides its unique methylation reaction, TsrM has a high turnover rate and shows a great substrate ambiguity [128]. TsrM was crystallised, revealing its structure, and enabling the localisation of bound cobalamin and the [4Fe-4S] cluster (Figure 10). The structural variation among *C*-MTs becomes apparent when comparing the crystal structure of TsrM with the predicted structures of PoyB and PoyC.

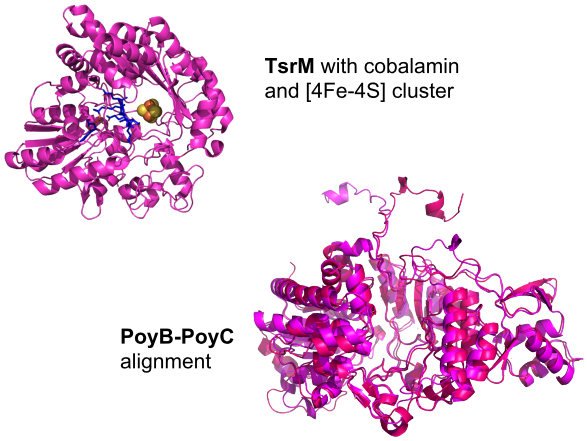



**Figure 10:** The three-dimensional structures of the rSAM *C*-MTs TsrM with bound cobalamin and [4Fe-4S] cluster (PDB ID: 6WTE (https://doi.org/10.2210/pdb6WTE/pdb), [129]) and the structural alignment of PoyB and PoyC (model structures calculated by Colabfold [75] and visualised by PyMOL [76]). The RMSD of TsrM–PoyB is 4.032 Å (214 to 214 atoms), of TsrM–PoyC is 7.067 Å (329 to 329 atoms).

Class C rSAM MTs are enzymes with a HemN-like didomain structure, consisting of an rSAM core domain and a C-terminal domain similar to HemN. HemN is an oxygen-independent coproporphyrinogen-III oxidase (CPO) and is assigned as rSAM dependent [119,130]. The HemN-like domain is hypothesised to be involved in substrate binding. Most of the residues responsible for binding two SAM molecules in HemN are conserved in all class C rSAM MTs. Therefore, it is suggested that one SAM molecule is bound for the production of a dAdo radical, while a second SAM molecule serves as a donor of the methyl group. The precise mechanism of class C rSAM MTs remains elusive [118].

Previous sections already described the *O*- and *N*-MT encoded in the *pbt* cluster of *P. rosea*. The thiopeptide GE2270 encoded by *pbt* undergoes further regioselective modification through *C*-methylation of thiazoles. The two rSAM MTs PbtM2 and PbtM3 methylate thiazole E and thiazole D, respectively [64]. PbtM2 and PbtM3 show substantial sequence similarities to the

*C*-MTs of other thiopeptides, such as Tbtl (biosynthesis of thiomuracin) and NosN (involved in the maturation of nosiheptide) [131,132].

Thiomuracin is produced by *Nonomuraea* sp. and encoded in the *tbt* cluster [132]. During maturation it is methylated by the rSAM *C*-MT Tbtl, forming a 5-methylthiazole through *C*-methylation on an unactivated $sp^2$ carbon centre of thiazole, likely thiazole 4. The analysis of Tbtl´s activity revealed its predominant independence from the leader peptide. Only one residue, Asn3, in the precursor peptide was shown to be crucial for activity [133].

*Streptomyces actuosus* contains the BGC *nos*, which encodes nosiheptide. The protein NosN encoded in *nos*, is responsible for catalysing the *C*-methylation of the peptide at C4 of the indolyl moiety. A unique characteristic of NosN, which distinguishes it from other known rSAM MTs, is that is does not generate SAH as a byproduct. This observation suggests a novel mechanism for indole C4 methylation by NosN [131,134].

## *S*-Methyltransferases

*S*-MTs are rare in RiPP biosynthesis and in natural product synthesis in general. There are few examples, and in most cases, the characterisation of the MT does not go beyond deletion studies to establish function, such as thiol *S*-methylation by SchX in the biosynthesis of the thiopeptide Sch40832, if it is identified at all [135,136]. In this section two *S*-MTs will be discussed.

The first MT is involved in the biosynthesis of an unidentified metabolite in *Candidatus* Entotheonella factor. The MT PtyS is part of a proteusin BGC *pty* and catalyses the *S*-methylation of up to four out of five cysteine residues in the precursor PtyA. In heterologous co-expressions of PtyA and PtyS, methylations were localised via LC–MS analysis to Cys9 and Cys15, with Cys9 being the preferred target. In the heterologous expression system, *S*-methylation of the cysteine residues can co-occur with lanthionine formation. When PtyM and PtyS are co-expressed, only one methylation of cysteine and three dehydrations could be observed. This may be due to the installed modifications inhibiting each other or due to limitations of the expression system [137].

LopH is a second example of RiPP *S*-MTs, and it is part of the longipeptin A biosynthesis (Figure 3). LopH is predicted to be responsible for *S*-methylation of methionine, yielding a sulphonium moiety on Met13. While *S*-methylation was shown via LC–MS and NMR analysis of the natural product, no knockout studies were performed to identify which enzyme is responsible for the *S*-methylation. The authors propose LopH as the most likely enzyme, as it shares modest sequence and structural similarity with other MTs. This also hints at a possible new MT family with LopH as its founding member. Future experimental work will be necessary to unequivocally assign this function to LopH [66].

## Future prospects: biocatalytic potential of RiPP methyltransferases

The substrate scope of several of the RiPP MTs described above has been investigated, and some show promising potential for biocatalytic applications, accepting a wide range of different precursor substrates. In many cases, these substrate scope analyses are performed by co-expressing different precursor peptides with the MT – in such in vivo systems, SAM supply or SAM regeneration is not required. However, using SAM supply or regeneration systems may be useful if the peptide substrate cannot be synthesised ribosomally, e.g., if it contains non-proteinogenic amino acids that cannot be produced by post-translational modifications of RiPP pathways.

SAM supply or SAM regeneration systems typically synthesise SAM from L-methionine and adenosine 5'-triphosphate (ATP) using a methionine adenosyltransferase. The MT then utilises SAM for the methylation reaction and the resulting SAH is degraded to adenine and *S*-ribosylhomocysteine, or to adenosine and L-homocysteine [4,21,25]. To achieve complete SAM regeneration, adenine or adenosine can be recycled to ATP, and L-homocysteine can be remethylated to L-methionine [21,25]. The SAM regeneration system is not only applicable to conventional SAM-dependent MTs but also to rSAM MTs and other SAM-dependent enzymes, such as amino-propyltransferases [26]. Furthermore, alternative alkyl groups such as ethyl, propargyl, allyl, benzyl, and nitrobenzyl groups can be transferred [31,138-142]. Such alkylation strategies have been successfully applied to small molecules such as coumarin [141], anthranilate [21], and pyrazole [24], as well as larger acceptor molecules such as the polyketide rapamycin [31] or DNA [143].

The use of RiPP MTs with SAM supply or SAM regeneration systems has the potential to greatly increase chemical diversity of RiPP natural products. The addition of alternative alkyl groups onto peptide substrates could further expand the range of products. RiPP MTs exhibit excellent specificity in the late-stage alkylation of peptide natural products.

# Conclusion

RiPP pathways are a valuable source of novel MTs that enable methylations at various positions of peptide substrates. MTs from RiPP pathways can supplement the MTs currently used for biocatalytic methylation reactions. If RiPP MTs, like many other MTs, also accept SAM-derivatives with larger alkyl

groups as substrates, they can transfer various groups such as ethyl, allyl, propyl, and benzyl moieties. In vitro systems can be used to install specific modifications in peptide structures. The remaining challenges include combining various post-translational modifications and developing a medium to high throughput screening approach to subject interesting peptide drug candidates to various MTs and perform SAR studies (structure–activity relationship). Overall, RiPP MTs will be a useful tool for drug development; it will be intriguing to observe what the future holds.

## Acknowledgements

We would like to thank Jennifer N. Andexer and Philipp Germer (both Institute of Pharmaceutical Sciences, University of Freiburg) for many discussions on SAM-dependent methyltransferases and critical proofreading of the manuscript.

## Funding

Research in the Mordhorst lab is supported by the Baden-Württemberg Stiftung through the Postdoctoral fellowship for leading early career researchers and by the Federal Ministry of Education and Research (BMBF) and the Baden-Württemberg Ministry of Science as part of the Excellence Strategy of the German Federal and State Governments.

## Author Contributions

Maria-Paula Schröder: visualization; writing – original draft; writing – review & editing. Isabel P.-M. Pfeiffer: visualization; writing – original draft; writing – review & editing. Silja Mordhorst: conceptualization; funding acquisition; supervision; visualization; writing – original draft; writing – review & editing.

## ORCID® iDs

Maria-Paula Schröder - https://orcid.org/0009-0000-7887-0514
Isabel P.-M. Pfeiffer - https://orcid.org/0009-0008-9765-8103
Silja Mordhorst - https://orcid.org/0000-0001-7274-7451

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

1. Montalbán-López, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; van Heel, A. J.; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; Grande Burgos, M. J.; Hill, C.; Kim, S.; Koehnke, J.; Latham, J. A.; Link, A. J.; Martínez, B.; Nair, S. K.; Nicolet, Y.; Rebuffat, S.; Sahl, H.-G.; Sareen, D.; Schmidt, E. W.; Schmitt, L.; Severinov, K.; Süssmuth, R. D.; Truman, A. W.; Wang, H.; Weng, J.-K.; van Wezel, G. P.; Zhang, Q.; Zhong, J.; Piel, J.; Mitchell, D. A.; Kuipers, O. P.; van der Donk, W. A. *Nat. Prod. Rep.* **2021,** *38,* 130–239. doi:10.1039/d0np00027b
2. 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.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013,** *30,* 108–160. doi:10.1039/c2np20085f
3. Struck, A.-W.; Thompson, M. L.; Wong, L. S.; Micklefield, J. *ChemBioChem* **2012,** *13,* 2642–2655. doi:10.1002/cbic.201200556
4. Siegrist, J.; Aschwanden, S.; Mordhorst, S.; Thöny-Meyer, L.; Richter, M.; Andexer, J. N. *ChemBioChem* **2015,** *16,* 2576–2579. doi:10.1002/cbic.201500410
5. Sommer-Kamann, C.; Fries, A.; Mordhorst, S.; Andexer, J. N.; Müller, M. *Angew. Chem., Int. Ed.* **2017,** *56,* 4033–4036. doi:10.1002/anie.201609375
6. Jalali, E.; Thorson, J. S. *Curr. Opin. Biotechnol.* **2021,** *69,* 290–298. doi:10.1016/j.copbio.2021.02.010
7. Ongpipattanakul, C.; Desormeaux, E. K.; DiCaprio, A.; van der Donk, W. A.; Mitchell, D. A.; Nair, S. K. *Chem. Rev.* **2022,** *122,* 14722–14814. doi:10.1021/acs.chemrev.2c00210
8. Pfeiffer, I. P.-M.; Schröder, M.-P.; Mordhorst, S. *Nat. Prod. Rep.* **2024,** in press. doi:10.1039/d3np00057e
9. Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M. *Drug Discovery Today* **2010,** *15,* 40–56. doi:10.1016/j.drudis.2009.10.009
10. Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. *Chem. Biol. Drug Des.* **2013,** *81,* 136–147. doi:10.1111/cbdd.12055
11. Kahan, R.; Worm, D. J.; de Castro, G. V.; Ng, S.; Barnard, A. *RSC Chem. Biol.* **2021,** *2,* 387–409. doi:10.1039/d0cb00205d
12. Zhao, X.; Li, Z.; Kuipers, O. P. *Cell Chem. Biol.* **2020,** *27,* 1262–1271.e4. doi:10.1016/j.chembiol.2020.07.005
13. Mordhorst, S.; Morinaka, B. I.; Vagstad, A. L.; Piel, J. *Angew. Chem., Int. Ed.* **2020,** *59,* 21442–21447. doi:10.1002/anie.202008990
14. Schönherr, H.; Cernak, T. *Angew. Chem., Int. Ed.* **2013,** *52,* 12256–12267. doi:10.1002/anie.201303207
15. de Sena Murteira Pinheiro, P.; Franco, L. S.; Fraga, C. A. M. *Pharmaceuticals* **2023,** *16,* 1157. doi:10.3390/ph16081157
16. Parks, L. W.; Schlenk, F. *J. Biol. Chem.* **1958,** *230,* 295–305. doi:10.1016/s0021-9258(18)70564-4
17. Deguchi, T.; Barchas, J. *J. Biol. Chem.* **1971,** *246,* 3175–3181. doi:10.1016/s0021-9258(18)62211-2
18. Parveen, N.; Cornell, K. A. *Mol. Microbiol.* **2011,** *79,* 7–20. doi:10.1111/j.1365-2958.2010.07455.x
19. Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. *Chem. Rev.* **2014,** *114,* 4229–4317. doi:10.1021/cr4004709
20. Eustáquio, A. S.; Pojer, F.; Noel, J. P.; Moore, B. S. *Nat. Chem. Biol.* **2008,** *4,* 69–74. doi:10.1038/nchembio.2007.56
21. Mordhorst, S.; Siegrist, J.; Müller, M.; Richter, M.; Andexer, J. N. *Angew. Chem., Int. Ed.* **2017,** *56,* 4037–4041. doi:10.1002/anie.201611038

22. Liao, C.; Seebeck, F. P. *Nat. Catal.* **2019,** *2,* 696–701. doi:10.1038/s41929-019-0300-0
23. Mordhorst, S.; Andexer, J. N. *Nat. Prod. Rep.* **2020,** *37,* 1316–1333. doi:10.1039/d0np00004c
24. Bengel, L. L.; Aberle, B.; Egler-Kemmerer, A.-N.; Kienzle, S.; Hauer, B.; Hammer, S. C. *Angew. Chem., Int. Ed.* **2021,** *60,* 5554–5560. doi:10.1002/anie.202014239
25. Popadić, D.; Mhaindarkar, D.; Dang Thai, M. H. N.; Hailes, H. C.; Mordhorst, S.; Andexer, J. N. *RSC Chem. Biol.* **2021,** *2,* 883–891. doi:10.1039/d1cb00033k
26. Gericke, L.; Mhaindarkar, D.; Karst, L. C.; Jahn, S.; Kuge, M.; Mohr, M. K. F.; Gagsteiger, J.; Cornelissen, N. V.; Wen, X.; Mordhorst, S.; Jessen, H. J.; Rentmeister, A.; Seebeck, F. P.; Layer, G.; Loenarz, C.; Andexer, J. N. *ChemBioChem* **2023,** *24,* e202300133. doi:10.1002/cbic.202300133
27. Jockmann, E.; Subrizi, F.; Mohr, M. K. F.; Carter, E. M.; Hebecker, P. M.; Popadić, D.; Hailes, H. C.; Andexer, J. N. *ChemCatChem* **2023,** *15,* e202300930. doi:10.1002/cctc.202300930
28. Stecher, H.; Tengg, M.; Ueberbacher, B. J.; Remler, P.; Schwab, H.; Griengl, H.; Gruber-Khadjawi, M. *Angew. Chem., Int. Ed.* **2009,** *48,* 9546–9548. doi:10.1002/anie.200905095
29. Iwai, N.; Kitahara, Y.; Kitazume, T. *J. Mol. Catal. B: Enzym.* **2011,** *73,* 1–4. doi:10.1016/j.molcatb.2011.06.018
30. Sadler, J. C.; Humphreys, L. D.; Snajdrova, R.; Burley, G. A. *ChemBioChem* **2017,** *18,* 992–995. doi:10.1002/cbic.201700115
31. Law, B. J. C.; Struck, A.-W.; Bennett, M. R.; Wilkinson, B.; Micklefield, J. *Chem. Sci.* **2015,** *6,* 2885–2892. doi:10.1039/c5sc00164a
32. Ovcharenko, A.; Weissenboeck, F. P.; Rentmeister, A. *Angew. Chem., Int. Ed.* **2021,** *60,* 4098–4103. doi:10.1002/anie.202013936
33. Sagan, S.; Karoyan, P.; Lequin, O.; Chassaing, G.; Lavielle, S. *Curr. Med. Chem.* **2004,** *11,* 2799–2822. doi:10.2174/0929867043364108
34. Luisa Di Gioia, M.; Leggio, A.; Malagrinò, F.; Romio, E.; Siciliano, C.; Liguori, A. *Mini-Rev. Med. Chem.* **2016,** *16,* 683–690. doi:10.2174/1389557516666160322152457
35. Li, J.; Zhang, C.; Huang, P.; Kuru, E.; Forster-Benson, E. T. C.; Li, T.; Church, G. M. *Translation* **2017,** *5,* e1327006. doi:10.1080/21690731.2017.1327006
36. White, K. N.; Konopelski, J. P. *Org. Lett.* **2005,** *7,* 4111–4112. doi:10.1021/ol051441w
37. Xu, C.-P.; Xiao, Z.-H.; Zhuo, B.-Q.; Wang, Y.-H.; Huang, P.-Q. *Chem. Commun.* **2010,** *46,* 7834–7836. doi:10.1039/c0cc01487g
38. Freidinger, R. M.; Hinkle, J. S.; Perlow, D. S. *J. Org. Chem.* **1983,** *48,* 77–81. doi:10.1021/jo00149a016
39. Chen, Y. *Chem. – Eur. J.* **2019,** *25,* 3405–3439. doi:10.1002/chem.201803642
40. Miller, S. C.; Scanlan, T. S. *J. Am. Chem. Soc.* **1997,** *119,* 2301–2302. doi:10.1021/ja9635443
41. Naoum, J. N.; Chandra, K.; Shemesh, D.; Gerber, R. B.; Gilon, C.; Hurevich, M. *Beilstein J. Org. Chem.* **2017,** *13,* 806–816. doi:10.3762/bjoc.13.81
42. Biron, E.; Chatterjee, J.; Kessler, H. *J. Pept. Sci.* **2006,** *12,* 213–219. doi:10.1002/psc.711
43. Sable, G. A.; Park, J.; Kim, H.; Lim, S.-J.; Jang, S.; Lim, D. *Eur. J. Org. Chem.* **2015,** 7043–7052. doi:10.1002/ejoc.201500697
44. Li, X.; Wang, N.; Liu, Y.; Li, W.; Bai, X.; Liu, P.; He, C.-Y. *Bioorg. Chem.* **2023,** *141,* 106892. doi:10.1016/j.bioorg.2023.106892
45. Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001,** *19,* 751–755. doi:10.1038/90802
46. Shimizu, Y.; Kanamori, T.; Ueda, T. *Methods* **2005,** *36,* 299–304. doi:10.1016/j.ymeth.2005.04.006
47. Merryman, C.; Green, R. *Chem. Biol.* **2004,** *11,* 575–582. doi:10.1016/j.chembiol.2004.03.009
48. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001,** *292,* 498–500. doi:10.1126/science.1060077
49. Forster, A. C.; Tan, Z.; Nalam, M. N. L.; Lin, H.; Qu, H.; Cornish, V. W.; Blacklow, S. C. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100,* 6353–6357. doi:10.1073/pnas.1132122100
50. Josephson, K.; Hartman, M. C. T.; Szostak, J. W. *J. Am. Chem. Soc.* **2005,** *127,* 11727–11735. doi:10.1021/ja0515809
51. Kawakami, T.; Murakami, H.; Suga, H. *J. Am. Chem. Soc.* **2008,** *130,* 16861–16863. doi:10.1021/ja806998v
52. Kawakami, T.; Murakami, H.; Suga, H. *Chem. Biol.* **2008,** *15,* 32–42. doi:10.1016/j.chembiol.2007.12.008
53. Liu, D.; Rubin, G. M.; Dhakal, D.; Chen, M.; Ding, Y. *iScience* **2021,** *24,* 102512. doi:10.1016/j.isci.2021.102512
54. Jacoby, C.; Krull, J.; Andexer, J.; Jehmlich, N.; von Bergen, M.; Brüls, T.; Boll, M. *mBio* **2020,** *11,* e01259-20. doi:10.1128/mbio.01259-20
55. Gruber, K.; Puffer, B.; Kräutler, B. *Chem. Soc. Rev.* **2011,** *40,* 4346–4363. doi:10.1039/c1cs15118e
56. Coulter, C.; Hamilton, J. T. G.; McRoberts, W. C.; Kulakov, L.; Larkin, M. J.; Harper, D. B. *Appl. Environ. Microbiol.* **1999,** *65,* 4301–4312. doi:10.1128/aem.65.10.4301-4312.1999
57. Kocsis, M. G.; Ranocha, P.; Gage, D. A.; Simon, E. S.; Rhodes, D.; Peel, G. J.; Mellema, S.; Saito, K.; Awazuhara, M.; Li, C.; Meeley, R. B.; Tarczynski, M. C.; Wagner, C.; Hanson, A. D. *Plant Physiol.* **2003,** *131,* 1808–1815. doi:10.1104/pp.102.018846
58. Schubert, H. L.; Blumenthal, R. M.; Cheng, X. *Trends Biochem. Sci.* **2003,** *28,* 329–335. doi:10.1016/s0968-0004(03)00090-2
59. Abdelraheem, E.; Thair, B.; Varela, R. F.; Jockmann, E.; Popadić, D.; Hailes, H. C.; Ward, J. M.; Iribarren, A. M.; Lewkowicz, E. S.; Andexer, J. N.; Hagedoorn, P.-L.; Hanefeld, U. *ChemBioChem* **2022,** *23,* e202200212. doi:10.1002/cbic.202200212
60. Booker, S. J.; Lloyd, C. T. *ACS Bio Med Chem Au* **2022,** *2,* 538–547. doi:10.1021/acsbiomedchemau.2c00078
61. Liscombe, D. K.; Louie, G. V.; Noel, J. P. *Nat. Prod. Rep.* **2012,** *29,* 1238. doi:10.1039/c2np20029e
62. Koirala, N.; Thuan, N. H.; Ghimire, G. P.; Thang, D. V.; Sohng, J. K. *Enzyme Microb. Technol.* **2016,** *86,* 103–116. doi:10.1016/j.enzmictec.2016.02.003
63. Siegrist, J.; Netzer, J.; Mordhorst, S.; Karst, L.; Gerhardt, S.; Einsle, O.; Richter, M.; Andexer, J. N. *FEBS Lett.* **2017,** *591,* 312–321. doi:10.1002/1873-3468.12530
64. Tocchetti, A.; Maffioli, S.; Iorio, M.; Alt, S.; Mazzei, E.; Brunati, C.; Sosio, M.; Donadio, S. *Chem. Biol.* **2013,** *20,* 1067–1077. doi:10.1016/j.chembiol.2013.07.005
65. Hug, J. J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. *ACS Chem. Biol.* **2020,** *15,* 2221–2231. doi:10.1021/acschembio.0c00430
66. Saad, H.; Majer, T.; Bhattarai, K.; Lampe, S.; Nguyen, D. T.; Kramer, M.; Straetener, J.; Brötz-Oesterhelt, H.; Mitchell, D. A.; Gross, H. *Chem. Sci.* **2023,** *14,* 13176–13183. doi:10.1039/d3sc02380j
67. Sato, C. S.; Byerrum, R. U.; Ball, C. D. *J. Biol. Chem.* **1957,** *224,* 717–723. doi:10.1016/s0021-9258(18)64963-4

68. Kauss, H.; Swanson, A. L.; Hassid, W. Z. *Biochem. Biophys. Res. Commun.* **1967,** *26,* 234–240. doi:10.1016/0006-291x(67)90240-9
69. Ward, L. C.; McCue, H. V.; Carnell, A. J. *ChemCatChem* **2021,** *13,* 121–128. doi:10.1002/cctc.202001316
70. Leikoski, N.; Liu, L.; Jokela, J.; Wahlsten, M.; Gugger, M.; Calteau, A.; Permi, P.; Kerfeld, C. A.; Sivonen, K.; Fewer, D. P. *Chem. Biol.* **2013,** *20,* 1033–1043. doi:10.1016/j.chembiol.2013.06.015
71. Acedo, J. Z.; Bothwell, I. R.; An, L.; Trouth, A.; Frazier, C.; van der Donk, W. A. *J. Am. Chem. Soc.* **2019,** *141,* 16790–16801. doi:10.1021/jacs.9b07396
72. Huo, L.; Zhao, X.; Acedo, J. Z.; Estrada, P.; Nair, S. K.; van der Donk, W. A. *ChemBioChem* **2020,** *21,* 190–199. doi:10.1002/cbic.201900483
73. Ali, M.; Ishqi, H. M.; Husain, Q. *Biotechnol. Bioeng.* **2020,** *117,* 1877–1894. doi:10.1002/bit.27329
74. van der Meer, J.-Y.; Biewenga, L.; Poelarends, G. J. *ChemBioChem* **2016,** *17,* 1792–1799. doi:10.1002/cbic.201600382
75. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. *Nat. Methods* **2022,** *19,* 679–682. doi:10.1038/s41592-022-01488-1
76. *Pymol,* Version 2.0; The PyMOL Molecular Graphics System; Schrödinger LLC: New York, NY, USA.
77. Gavrish, E.; Sit, C. S.; Cao, S.; Kandror, O.; Spoering, A.; Peoples, A.; Ling, L.; Fetterman, A.; Hughes, D.; Bissell, A.; Torrey, H.; Akopian, T.; Mueller, A.; Epstein, S.; Goldberg, A.; Clardy, J.; Lewis, K. *Chem. Biol.* **2014,** *21,* 509–518. doi:10.1016/j.chembiol.2014.01.014
78. Lee, H.; Park, S. H.; Kim, J.; Lee, J.; Koh, M. S.; Lee, J. H.; Kim, S. *Adv. Sci.* **2024,** *11,* 2305946. doi:10.1002/advs.202305946
79. Franz, L.; Kazmaier, U.; Truman, A. W.; Koehnke, J. *Nat. Prod. Rep.* **2021,** *38,* 1659–1683. doi:10.1039/d0np00097c
80. Gomez-Escribano, J. P.; Song, L.; Bibb, M. J.; Challis, G. L. *Chem. Sci.* **2012,** *3,* 3522. doi:10.1039/c2sc21183a
81. Huo, L.; Rachid, S.; Stadler, M.; Wenzel, S. C.; Müller, R. *Chem. Biol.* **2012,** *19,* 1278–1287. doi:10.1016/j.chembiol.2012.08.013
82. Chatterjee, J.; Rechenmacher, F.; Kessler, H. *Angew. Chem., Int. Ed.* **2013,** *52,* 254–269. doi:10.1002/anie.201205674
83. Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. *Acc. Chem. Res.* **2008,** *41,* 1331–1342. doi:10.1021/ar8000603
84. Biron, E.; Chatterjee, J.; Ovadia, O.; Langenegger, D.; Brueggen, J.; Hoyer, D.; Schmid, H. A.; Jelinek, R.; Gilon, C.; Hoffman, A.; Kessler, H. *Angew. Chem., Int. Ed.* **2008,** *47,* 2595–2599. doi:10.1002/anie.200705797
85. Diaz, K.; Meng, Y.; Huang, R. *Curr. Opin. Chem. Biol.* **2021,** *63,* 115–122. doi:10.1016/j.cbpa.2021.02.017
86. Viehrig, K.; Surup, F.; Volz, C.; Herrmann, J.; Abou Fayad, A.; Adam, S.; Köhnke, J.; Trauner, D.; Müller, R. *Angew. Chem., Int. Ed.* **2017,** *56,* 7407–7410. doi:10.1002/anie.201612640
87. Adam, S.; Zheng, D.; Klein, A.; Volz, C.; Mullen, W.; Shirran, S. L.; Smith, B. O.; Kalinina, O. V.; Müller, R.; Koehnke, J. *Nat. Chem.* **2023,** *15,* 560–568. doi:10.1038/s41557-023-01153-w
88. Zhang, Q.; van der Donk, W. A. *FEBS Lett.* **2012,** *586,* 3391–3397. doi:10.1016/j.febslet.2012.07.050
89. Koiso, Y.; Li, Y.; Iwasaki, S.; Hanaka, K.; Kobayashi, T.; Sonoda, R.; Fujita, Y.; Yaegashi, H.; Sato, Z. *J. Antibiot.* **1994,** *47,* 765–773. doi:10.7164/antibiotics.47.765
90. Umemura, M.; Koike, H.; Nagano, N.; Ishii, T.; Kawano, J.; Yamane, N.; Kozone, I.; Horimoto, K.; Shin-ya, K.; Asai, K.; Yu, J.; Bennett, J. W.; Machida, M. *PLoS One* **2013,** *8,* e84028. doi:10.1371/journal.pone.0084028
91. Umemura, M.; Nagano, N.; Koike, H.; Kawano, J.; Ishii, T.; Miyamura, Y.; Kikuchi, M.; Tamano, K.; Yu, J.; Shin-ya, K.; Machida, M. *Fungal Genet. Biol.* **2014,** *68,* 23–30. doi:10.1016/j.fgb.2014.04.011
92. Ye, Y.; Minami, A.; Igarashi, Y.; Izumikawa, M.; Umemura, M.; Nagano, N.; Machida, M.; Kawahara, T.; Shin-ya, K.; Gomi, K.; Oikawa, H. *Angew. Chem., Int. Ed.* **2016,** *55,* 8072–8075. doi:10.1002/anie.201602611
93. Ding, W.; Liu, W.-Q.; Jia, Y.; Li, Y.; van der Donk, W. A.; Zhang, Q. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* 3521–3526. doi:10.1073/pnas.1522907113
94. Lee, J.; Hao, Y.; Blair, P. M.; Melby, J. O.; Agarwal, V.; Burkhart, B. J.; Nair, S. K.; Mitchell, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110,* 12954–12959. doi:10.1073/pnas.1306101110
95. Scholz, R.; Molohon, K. J.; Nachtigall, J.; Vater, J.; Markley, A. L.; Süssmuth, R. D.; Mitchell, D. A.; Borriss, R. *J. Bacteriol.* **2011,** *193,* 215–224. doi:10.1128/jb.00784-10
96. Molohon, K. J.; Saint-Vincent, P. M. B.; Park, S.; Doroghazi, J. R.; Maxson, T.; Hershfield, J. R.; Flatt, K. M.; Schroeder, N. E.; Ha, T.; Mitchell, D. A. *ACS Infect. Dis.* **2016,** *2,* 207–220. doi:10.1021/acsinfecdis.5b00115
97. Smith, T. E.; Pond, C. D.; Pierce, E.; Harmer, Z. P.; Kwan, J.; Zachariah, M. M.; Harper, M. K.; Wyche, T. P.; Matainaho, T. K.; Bugni, T. S.; Barrows, L. R.; Ireland, C. M.; Schmidt, E. W. *Nat. Chem. Biol.* **2018,** *14,* 179–185. doi:10.1038/nchembio.2537
98. Ramm, S.; Krawczyk, B.; Mühlenweg, A.; Poch, A.; Mösker, E.; Süssmuth, R. D. *Angew. Chem., Int. Ed.* **2017,** *56,* 9994–9997. doi:10.1002/anie.201703488
99. Song, H.; Fahrig-Kamarauskaitè, J.; Matabaro, E.; Kaspar, H.; Shirran, S. L.; Zach, C.; Pace, A.; Stefanov, B.-A.; Naismith, J. H.; Künzler, M. *ACS Chem. Biol.* **2020,** *15,* 1901–1912. doi:10.1021/acschembio.0c00237
100. van der Velden, N. S.; Kälin, N.; Helf, M. J.; Piel, J.; Freeman, M. F.; Künzler, M. *Nat. Chem. Biol.* **2017,** *13,* 833–835. doi:10.1038/nchembio.2393
101. Song, H.; Burton, A. J.; Shirran, S. L.; Fahrig-Kamarauskaitė, J.; Kaspar, H.; Muir, T. W.; Künzler, M.; Naismith, J. H. *Angew. Chem., Int. Ed.* **2021,** *60,* 14319–14323. doi:10.1002/anie.202100818
102. Cho, H.; Lee, H.; Hong, K.; Chung, H.; Song, I.; Lee, J.-S.; Kim, S. *Biochemistry* **2022,** *61,* 183–194. doi:10.1021/acs.biochem.1c00764
103. Imani, A. S.; Lee, A. R.; Vishwanathan, N.; de Waal, F.; Freeman, M. F. *ACS Chem. Biol.* **2022,** *17,* 908–917. doi:10.1021/acschembio.1c01002
104. Miller, F. S.; Crone, K. K.; Jensen, M. R.; Shaw, S.; Harcombe, W. R.; Elias, M. H.; Freeman, M. F. *Nat. Commun.* **2021,** *12,* 5355. doi:10.1038/s41467-021-25575-7
105. Quijano, M. R.; Zach, C.; Miller, F. S.; Lee, A. R.; Imani, A. S.; Künzler, M.; Freeman, M. F. *J. Am. Chem. Soc.* **2019,** *141,* 9637–9644. doi:10.1021/jacs.9b03690

106. Paoli, L.; Ruscheweyh, H.-J.; Forneris, C. C.; Hubrich, F.; Kautsar, S.; Bhushan, A.; Lotti, A.; Clayssen, Q.; Salazar, G.; Milanese, A.; Carlström, C. I.; Papadopoulou, C.; Gehrig, D.; Karasikov, M.; Mustafa, H.; Larralde, M.; Carroll, L. M.; Sánchez, P.; Zayed, A. A.; Cronin, D. R.; Acinas, S. G.; Bork, P.; Bowler, C.; Delmont, T. O.; Gasol, J. M.; Gossert, A. D.; Kahles, A.; Sullivan, M. B.; Wincker, P.; Zeller, G.; Robinson, S. L.; Piel, J.; Sunagawa, S. *Nature* **2022,** *607,* 111–118. doi:10.1038/s41586-022-04862-3
107. Freeman, M. F.; Vagstad, A. L.; Piel, J. *Curr. Opin. Chem. Biol.* **2016,** *31,* 8–14. doi:10.1016/j.cbpa.2015.11.002
108. Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H.-G.; Matsunaga, S.; Piel, J. *Science* **2012,** *338,* 387–390. doi:10.1126/science.1226121
109. Hamada, T.; Matsunaga, S.; Yano, G.; Fusetani, N. *J. Am. Chem. Soc.* **2005,** *127,* 110–118. doi:10.1021/ja045749e
110. Renevey, A.; Riniker, S. *Eur. Biophys. J.* **2017,** *46,* 363–374. doi:10.1007/s00249-016-1179-1
111. Freeman, M. F.; Helf, M. J.; Bhushan, A.; Morinaka, B. I.; Piel, J. *Nat. Chem.* **2017,** *9,* 387–395. doi:10.1038/nchem.2666
112. Bhushan, A.; Egli, P. J.; Peters, E. E.; Freeman, M. F.; Piel, J. *Nat. Chem.* **2019,** *11,* 931–939. doi:10.1038/s41557-019-0323-9
113. Cogan, D. P.; Bhushan, A.; Reyes, R.; Zhu, L.; Piel, J.; Nair, S. K. *Proc. Natl. Acad. Sci. U. S. A.* **2022,** *119,* e2116578119. doi:10.1073/pnas.2116578119
114. Selva, E.; Ferrari, P.; Kurz, M.; Tavecchia, P.; Colombo, L.; Stella, S.; Restelli, E.; Goldstein, B. P.; Ripamonti, F.; Denaro, M. *J. Antibiot.* **1995,** *48,* 1039–1042. doi:10.7164/antibiotics.48.1039
115. Johnson, R. D.; Lane, G. A.; Koulman, A.; Cao, M.; Fraser, K.; Fleetwood, D. J.; Voisey, C. R.; Dyer, J. M.; Pratt, J.; Christensen, M.; Simpson, W. R.; Bryan, G. T.; Johnson, L. J. *Fungal Genet. Biol.* **2015,** *85,* 14–24. doi:10.1016/j.fgb.2015.10.005
116. Zhang, W.; Forester, N. T.; Chettri, P.; Heilijgers, M.; Mace, W. J.; Maes, E.; Morozova, Y.; Applegate, E. R.; Johnson, R. D.; Johnson, L. J. *J. Agric. Food Chem.* **2023,** *71,* 13965–13978. doi:10.1021/acs.jafc.3c03073
117. Ansari, M. Z.; Sharma, J.; Gokhale, R. S.; Mohanty, D. *BMC Bioinf.* **2008,** *9,* 454. doi:10.1186/1471-2105-9-454
118. Zhang, Q.; van der Donk, W. A.; Liu, W. *Acc. Chem. Res.* **2012,** *45,* 555–564. doi:10.1021/ar200202c
119. Sofia, H. J.; Chen, G.; Hetzler, B. G.; Reyes-Spindola, J. F.; Miller, N. E. *Nucleic Acids Res.* **2001,** *29,* 1097–1106. doi:10.1093/nar/29.5.1097
120. Matthews, R. G. *Acc. Chem. Res.* **2001,** *34,* 681–689. doi:10.1021/ar0000051
121. Matthews, R. G.; Koutmos, M.; Datta, S. *Curr. Opin. Struct. Biol.* **2008,** *18,* 658–666. doi:10.1016/j.sbi.2008.11.005
122. Gagsteiger, J.; Jahn, S.; Heidinger, L.; Gericke, L.; Andexer, J. N.; Friedrich, T.; Loenarz, C.; Layer, G. *Angew. Chem., Int. Ed.* **2022,** *61,* e202204198. doi:10.1002/anie.202204198
123. Nakamura, S.; Yajima, T.; Lin, Y. C.; Umezawa, H. *J. Antibiot., Ser. A* **1967,** *20,* 1–5.
124. Parent, A.; Guillot, A.; Benjdia, A.; Chartier, G.; Leprince, J.; Berteau, O. *J. Am. Chem. Soc.* **2016,** *138,* 15515–15518. doi:10.1021/jacs.6b06697
125. Iwamoto, M.; Shimizu, H.; Muramatsu, I.; Oiki, S. *FEBS Lett.* **2010,** *584,* 3995–3999. doi:10.1016/j.febslet.2010.08.007
126. Bagley, M. C.; Dale, J. W.; Merritt, E. A.; Xiong, X. *Chem. Rev.* **2005,** *105,* 685–714. doi:10.1021/cr0300441
127. Trejo, W. H.; Dean, L. D.; Pluscec, J.; Meyers, E.; Brown, W. E. *J. Antibiot.* **1977,** *30,* 639–643. doi:10.7164/antibiotics.30.639
128. Benjdia, A.; Pierre, S.; Gherasim, C.; Guillot, A.; Carmona, M.; Amara, P.; Banerjee, R.; Berteau, O. *Nat. Commun.* **2015,** *6,* 8377. doi:10.1038/ncomms9377
129. Knox, H. L.; Chen, P. Y.-T.; Blaszczyk, A. J.; Mukherjee, A.; Grove, T. L.; Schwalm, E. L.; Wang, B.; Drennan, C. L.; Booker, S. J. *Nat. Chem. Biol.* **2021,** *17,* 485–491. doi:10.1038/s41589-020-00717-y
130. Layer, G.; Verfürth, K.; Mahlitz, E.; Jahn, D. *J. Biol. Chem.* **2002,** *277,* 34136–34142. doi:10.1074/jbc.m205247200
131. Ding, W.; Li, Y.; Zhao, J.; Ji, X.; Mo, T.; Qianzhu, H.; Tu, T.; Deng, Z.; Yu, Y.; Chen, F.; Zhang, Q. *Angew. Chem., Int. Ed.* **2017,** *56,* 3857–3861. doi:10.1002/anie.201609948
132. Morris, R. P.; Leeds, J. A.; Naegeli, H. U.; Oberer, L.; Memmert, K.; Weber, E.; LaMarche, M. J.; Parker, C. N.; Burrer, N.; Esterow, S.; Hein, A. E.; Schmitt, E. K.; Krastel, P. *J. Am. Chem. Soc.* **2009,** *131,* 5946–5955. doi:10.1021/ja900488a
133. Mahanta, N.; Zhang, Z.; Hudson, G. A.; van der Donk, W. A.; Mitchell, D. A. *J. Am. Chem. Soc.* **2017,** *139,* 4310–4313. doi:10.1021/jacs.7b00693
134. LaMattina, J. W.; Wang, B.; Badding, E. D.; Gadsby, L. K.; Grove, T. L.; Booker, S. J. *J. Am. Chem. Soc.* **2017,** *139,* 17438–17445. doi:10.1021/jacs.7b08492
135. Puar, M. S.; Chan, T. M.; Hegde, V.; Patel, M.; Bartner, P.; Ng, K. J.; Pramanik, B. N.; Macfarlane, R. D. *J. Antibiot.* **1998,** *51,* 221–224. doi:10.7164/antibiotics.51.221
136. Cheng, B.; Guo, H.; Wang, H.; Zhao, Q.; Liu, W. *J. Am. Chem. Soc.* **2021,** *143,* 13790–13797. doi:10.1021/jacs.1c05956
137. Helf, M. J.; Jud, A.; Piel, J. *ChemBioChem* **2017,** *18,* 444–450. doi:10.1002/cbic.201600594
138. Lu, Z. J.; Markham, G. D. *J. Biol. Chem.* **2002,** *277,* 16624–16631. doi:10.1074/jbc.m110456200
139. Singh, S.; Zhang, J.; Huber, T. D.; Sunkara, M.; Hurley, K.; Goff, R. D.; Wang, G.; Zhang, W.; Liu, C.; Rohr, J.; Van Lanen, S. G.; Morris, A. J.; Thorson, J. S. *Angew. Chem., Int. Ed.* **2014,** *53,* 3965–3969. doi:10.1002/anie.201308272
140. Dippe, M.; Brandt, W.; Rost, H.; Porzel, A.; Schmidt, J.; Wessjohann, L. A. *Chem. Commun.* **2015,** *51,* 3637–3640. doi:10.1039/c4cc08478k
141. Hoffmann, A.; Schülke, K. H.; Hammer, S. C.; Rentmeister, A.; Cornelissen, N. V. *Chem. Commun.* **2023,** *59,* 5463–5466. doi:10.1039/d3cc01036h
142. Mohr, M. K. F.; Saleem-Batcha, R.; Cornelissen, N. V.; Andexer, J. N. *Chem. – Eur. J.* **2023,** *29,* e202301503. doi:10.1002/chem.202301503
143. Michailidou, F.; Klöcker, N.; Cornelissen, N. V.; Singh, R. K.; Peters, A.; Ovcharenko, A.; Kümmel, D.; Rentmeister, A. *Angew. Chem., Int. Ed.* **2021,** *60,* 480–485. doi:10.1002/anie.202012623

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.147

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Chemo-enzymatic total synthesis: current approaches toward the integration of chemical and enzymatic transformations

Ryo Tanifuji* and Hiroki Oguri*

Review

Open Access

Address:
Department of Chemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Email:
Ryo Tanifuji* - tanifujir@g.ecc.u-tokyo.ac.jp; Hiroki Oguri* - hirokioguri@g.ecc.u-tokyo.ac.jp

* Corresponding author

Keywords:
chemo-enzymatic synthesis; late-stage modification; reactive biosynthetic intermediate; regio- and stereoselective (macro)cyclization; total synthesis

*Beilstein J. Org. Chem.* **2024,** *20,* 1693–1712.
https://doi.org/10.3762/bjoc.20.151

Received: 01 April 2024
Accepted: 02 July 2024
Published: 23 July 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: J. D. Rudolf



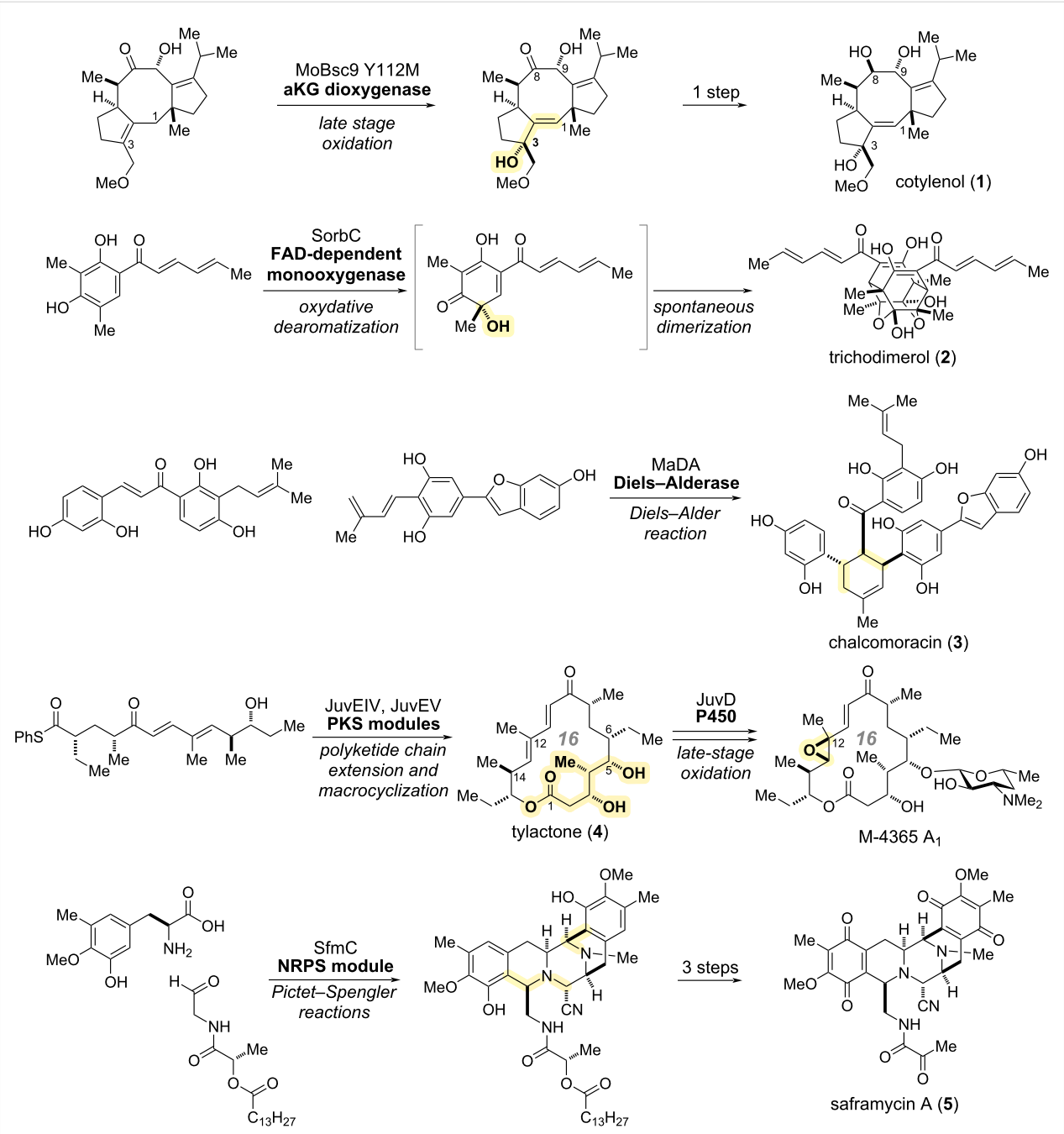
© 2024 Tanifuji and Oguri; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

A steadily increasing number of reports have been published on chemo-enzymatic synthesis methods that integrate biosynthetic enzymatic transformations with chemical conversions. This review focuses on the total synthesis of natural products and classifies the enzymatic reactions into three categories. The total synthesis of five natural products: cotylenol, trichodimerol, chalcomoracin, tylactone, and saframycin A, as well as their analogs, is outlined with an emphasis on comparing these chemo-enzymatic syntheses with the corresponding natural biosynthetic pathways.

## Introduction

Naturally occurring organic compounds with potent biological activities continue to be compelling lead candidates for drug discovery, with advancements in their synthesis and supply techniques progressing rapidly. Of particular note is the progress made in the "chemo-enzymatic approach" merging efficient enzymatic synthesis – traditionally employed in the biosynthesis of natural products by microorganisms and plants – with precise chemical synthesis conducted by chemists. Chemo-enzymatic total syntheses reported recently fall into three main categories based on the purpose for using an enzyme or at what stage in a synthesis the enzyme is employed: 1) regio- and stereoselective late-stage functionalization of core scaffolds, 2) in situ generation of highly reactive intermediates, and 3) the one-step construction of macrocyclic or fused multicyclic scaffolds via regio- and stereoselective cyclization reactions. This review aims to provide an overview of these approaches and parallel comparisons with original biosynthetic pathways by highlighting five examples of chemo-enzymatic total syntheses of natural products reported since 2017. The examples are the synthesis of cotylenol (**1**), trichodimerol (**2**), chalcomoracin (**3**), tylactone (**4**), and saframycin A (**5**), as well as a number of analogues of these natural products (Scheme 1). The overview of




**Scheme 1:** Targeted natural products and key enzymatic transformations in the chemo-enzymatic total syntheses featured in this review. Enzymatically formed bonds or units are highlighted in yellow.

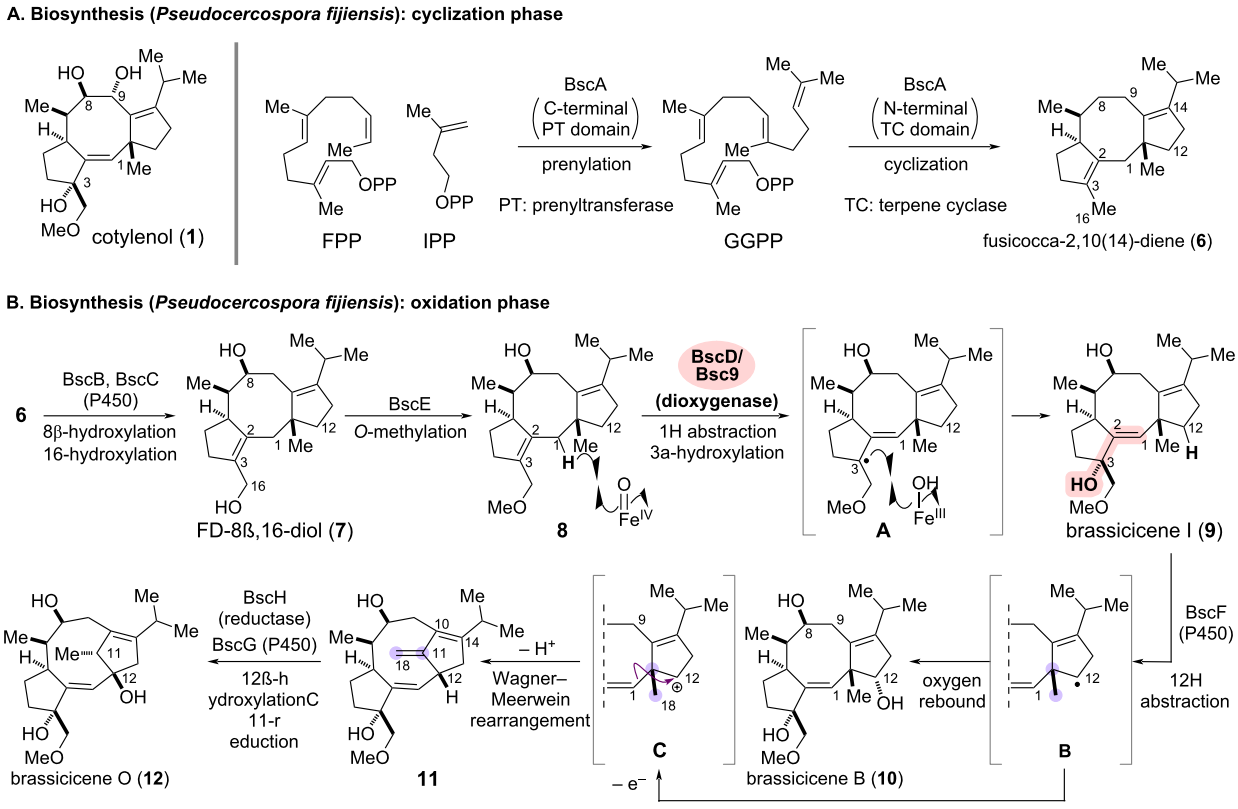
all five natural products begins with a description of the well-studied biosynthetic strategies evolved by microorganisms and plants. Biosynthetic pathways are described with a focus on the biosynthetic intermediates and enzymatic transformations that enable cascade reactions and pinpoint modifications. This focus highlights how these biosynthetic pathways are applicable to the development of streamlined chemo-enzymatic synthesis processes. The discussion will also encompass the design of biosynthetic intermediates and their analogs to achieve chemo-enzymatic total syntheses. Given our emphasis on natural products, this review does not cover the exquisite synthetic approaches involving biocatalysts for small-molecule pharmaceuticals. To gain a more comprehensive understanding of the chemo-enzymatic synthetic approach, we refer the reader to recent excellent reviews that provide multiple perspectives on the topic [1-10].

# Review

## Late-stage oxidative transformations of natural product scaffolds: cotylenol and brassicicenes

Chemo-enzymatic synthesis, featuring the late-stage enzymatic oxidation of chemically synthesized intricate scaffolds are attracting increasing attention. A collaboration between Stoltz and Arnold led to the pioneering accomplishment in the total synthesis of nigelladine A by exploiting P450 enzymes engineered through directed evolution [11]. P450 catalysis during the oxidation phase enabled the total synthesis of mitrephorone A [12], chevalone A [13], polysin [14], excolide B [15], and gedunin [16]. Fe(II)/2OG-dependent dioxygenases, such as FtmOx1, were employed as versatile catalysts in the synthesis of 13-oxoverruculogen [17]. The use of prenyltransferase NotF and flavin monooxygenase BvnB allowed the synthesis of eurotiumin A [18]. Although this review cannot cover the extensive recent progress in this type of chemo-enzymatic approach, we here provide an overview of the very recently reported chemo-enzymatic hybrid syntheses of cotylenol (**1**) and brassicicenes [19]. The key oxidative allylic rearrangement was conducted enzymatically, while the skeletal rearrangement originally mediated by P450 enzymes in the biosynthetic pathway was achieved through chemical transformation. Hence, this strategy can be considered a remarkable example of utilizing the complementarity between chemical and enzymatic transformations (Scheme 1 and Scheme 2).

The cotylenin and fusicoccin families comprise structurally related diterpene glucosides with a 5/8/5 fused tricyclic aglycon and a sugar moiety linked through the C9 hydroxy group of the aglycon. Cotylenin exhibits promising anticancer activity, and its diterpene aglycon, cotylenol (**1**), was isolated from the filamentous fungi such as *Phomopsis amygdali* and *Cladosporium* sp. 501-7W (Scheme 2A) [20-22]. Brassicicenes, differing in oxidation levels from **1**, have been isolated from *Alternaria brassicicola*, a phytopathogenic fungus that causes dark leaf spots in *Brassica* species. The biosynthesis process of brassicicenes was primarily elucidated through in vitro enzymatic transformations and heterologous expression using *Aspergillus oryzae*. The proposed biosynthetic pathway of brassicicenes in *Pseudocercospora fijiensis* is outlined based on an investigation by Oikawa and co-workers [23].




**Scheme 2:** Biosynthetic pathway to brassicicenes in *Pseudocercospora fijiensis* [14]. (A) Cyclization phase catalyzed by BscA. (B) Oxidation phase and scaffold rearrangements by modification enzymes.

The biosynthesis begins with the conversion of farnesyl pyrophosphate (FPP) and isopentenyl pyrophosphate (IPP) to geranylgeranyl pyrophosphate (GGPP), catalyzed by the prenyltransferase (PT) domain located at the C-terminus of BscA (Scheme 2A). Subsequently, the terpene cyclase (TC) domain at the N-terminal of BscA generates fusicocca-2,10(14)-diene (**6**), which bears the common 5/8/5 fused tricyclic scaffold common to this natural products family.

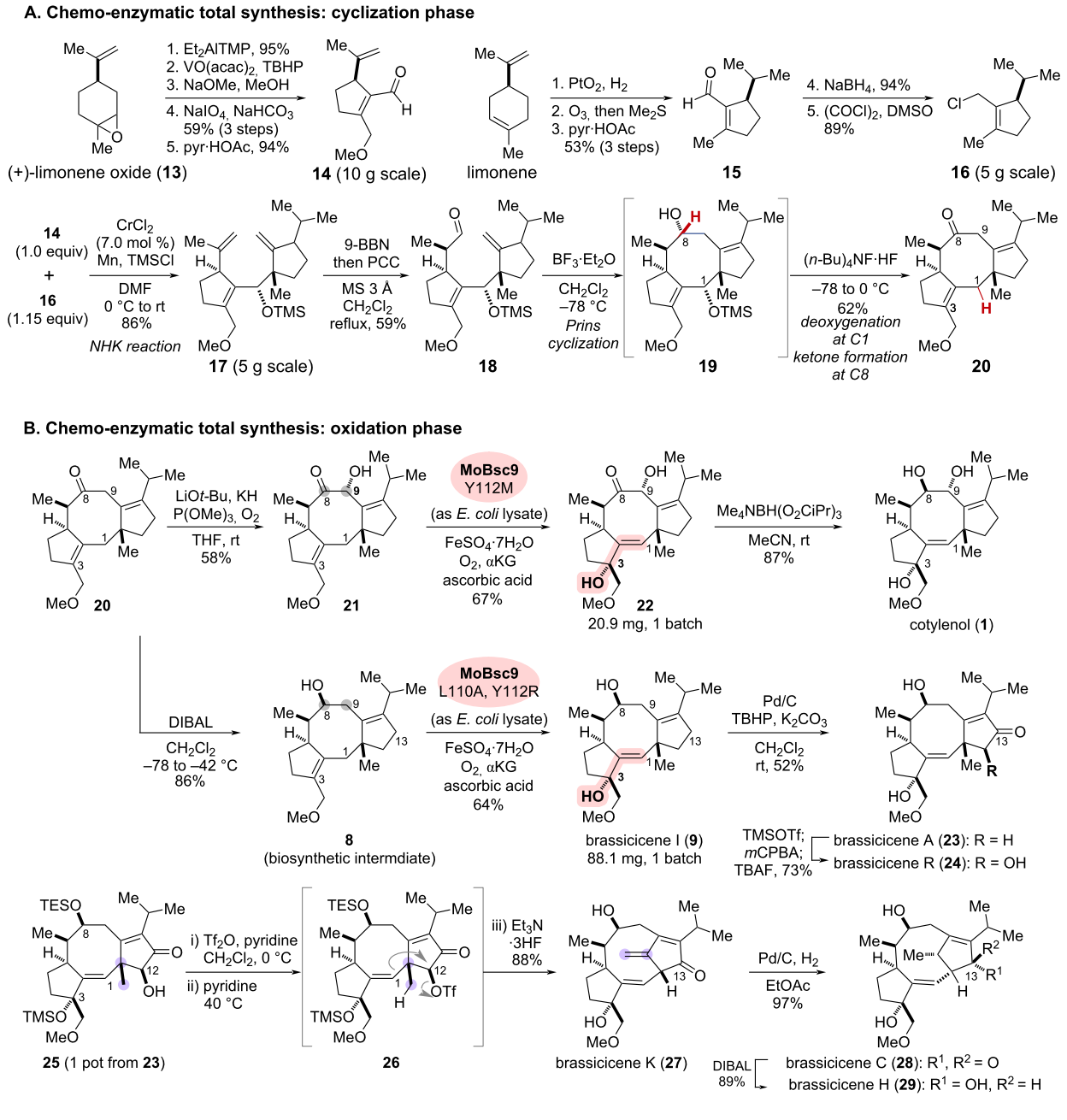
Sequential oxidative conversions of scaffold **6** yield a series of intermediates and natural products, including cotylenol (**1**) and brassicicenes I and B (**9** and **10**), as well as brassicicene O (**12**), which possesses a distinct scaffold resulting from a skeletal rearrangement. To the core scaffold **6**, the P450 enzymes, BscB and BscC, introduce hydroxy groups at C8 and C16 to produce FD-8β,16-diol (**7**), and BscE-catalyzed *O*-methylation generates the putative intermediate **8**. The subsequent oxidative allylic rearrangement (**8**→**9**), catalyzed by the nonheme iron(II) and 2-oxoglutarate (Fe(II)/2OG)-dependent dioxygenase BscD, was a key step toward developing a chemo-enzymatic synthetic process. Presumably, the reactive iron(IV)-oxo species in dioxygenase BscD abstracts an allylic hydrogen at C1 and generates intermediate **A**. Subsequent α-face-selective hydroxylation of the resulting allylic radical at the C3 position would yield brassicicene I (**9**). As a pioneering investigation to elucidate the mechanism of this essentially identical allylic oxidations by Fe(II)/2OG-dependent dioxygenase, Dairi and co-workers conducted in vitro enzymatic conversions with the homologous enzyme Bsc9, derived from *Alternaria brassicicola* ATCC96836 [24].

The P450 enzyme BscF is responsible for regioselective abstraction of a hydrogen at C12 and subsequent diastereoselective hydroxylation of the radical intermediate **B** to produce brassicicene B (**10**). Meanwhile, further single-electron oxidation of the intermediate **B** would trigger a Wagner–Meerwein-type skeletal rearrangement, providing the distinct skeleton of **11** via carbocation **C**. This rearrangement involves the preferential migration of an alkenyl group in **C** to the carbocation, followed by deprotonation at C18 to form an *exo*-olefin. β-face-selective hydroxylation at C12 in **11** by the P450 enzyme BscG, and subsequent reduction of the exomethylene at C11–C18 catalyzed by BscH yield brassicicene O (**12**).

Renata and co-workers successfully accomplished the chemoenzymatic total syntheses of cotylenol (**1**) and nine brassicicenes (Scheme 3) [19]. In the cyclization phase, a suitably functionalized 5/8/5 tricyclic scaffold was produced through scalable chemical synthesis. Subsequently, in vitro enzymatic oxidative functionalizations were carried out during the oxidation stage. The exploration of Bsc9 homologs and directed evolution expanded the scope of substrates of the dioxygenase beyond the natural biosynthetic intermediate **8** to its analogs **21**, enabling the chemo-enzymatic total synthesis of **1**.

The chemical synthesis commenced with the preparation of two fragments and their subsequent coupling to assemble the core 5/8/5 tricyclic scaffold (Scheme 3A). The left-half fragment, aldehyde **14**, was synthesized on a decagram scale in five steps from (+)-limonene oxide (**13**), involving epoxide manipulation, oxidative cleavage, and intramolecular aldol condensation. Similarly, the right-half fragment, allyl chloride **16**, was synthesized from limonene in five steps. Site-selective hydrogenation, oxidative cleavage, and intramolecular cyclization provided **15**, followed by functional group manipulation to yield **16**. The two segments **14** and **16** were assembled by Nozaki–Hiyama–Kishi (NHK) coupling while controlling the regio- and diastereoselectivities to afford intermediate **17** [25]. Site-selective hydroboration and subsequent oxidation yielded aldehyde **18**, a precursor for the intramolecular ring closure of the eight-membered ring. Upon treatment of **18** with $BF_3{\cdot}Et_2O$, diastereoselective Prins cyclization of **18** proceeded to generate secondary alcohol **19**. Subsequent one-pot treatment with $(n\text{-}Bu)_4NF{\cdot}HF$ resulted in unexpected conversions, including the formation of an allylic carbocation at C1, followed by transannular hydride transfer from C8 to afford ketone **20** in 62% yield.

With the 5/8/5 tricyclic scaffold **20** in hand, site- and diastereocontrolled C9 hydroxylation of **20** produced a substrate **21** for the enzymatic conversions, commencing with the oxidation phase (Scheme 3B). As shown in Scheme 2B, the two cognate biosynthetic enzymes, BscD and Bsc9, are responsible for the oxidative allylic rearrangement (**8**→**9**). While soluble BscD was not obtained, the overexpressed dioxygenase Bsc9 enabled to catalyze the regio- and diastereocontrolled hydroxylation at C3, along with transposition of the double bond, to form the desired product **22** in approximately 20% yield with 50% conversion. To enhance efficiency of the enzymatic oxidation, Renata and co-workers conducted two approaches: homolog screening and enzyme engineering. A Genome Neighborhood Diagram (GND) analysis [26] of dioxygenase Bsc9 identified five homologs, including a homolog MoBsc9 derived from *Magnaporthe oryzae,* which shares 54% sequence identity with Bsc9 and was found to be suitable for the oxidative allylic rearrangement (**21**→**22**). Further directed evolution using site-saturation mutagenesis targeting the putative active sites L110 and Y112, led to the variant MoBsc9 Y112M, which substantially improved the enzymatic conversion into **22**, achieving an isolated yield of up to 67%. Diastereoselective reduction of the C8 ketone was then achieved using the protocol of Nakada and co-workers [27], enabling the chemo-enzymatic total synthesis of cotylenol (**1**).




**Scheme 3:** Chemo-enzymatic total synthesis of cotylenol (**1**) and brassicicenes. (A) Chemical cyclization phase. (B) Chemical and enzymatic oxidation phase.

Similarly, diastereoselective reduction of the C8 ketone in **20** yielded the biosynthetic intermediate **8** for brassicicenes. Although substrate **8** has a different oxidation state at C8 and lacks the C9 secondary alcohol of **21**, the utilization of another variant, MoBsc9 L110A, Y112R, generated through the directed evolution of MoBsc9, facilitated the optimal conversion in the corresponding oxidative allylic rearrangement to afford brassicicene I (**9**) in 64% yield. Further palladium-catalyzed allylic oxidation with the incorporation of a ketone at the C13 position yielded brassicicene A (**23**). After conversion of **23** into the corresponding silyl enol ether, Rubottom oxidation allowed completion of the total synthesis of brassicicene R (**24**).

As an effort to explore the biomimetic rearrangement, analogous to the biosynthetic conversion of **9** into **11** (Scheme 2B), an α-hydroxylated ketone **25** with suitable protections of the C3 and C8 hydroxy groups was synthesized from **23** in one pot.

Conversion of the secondary alcohol in **25** into triflate **26** enabled the alkenyl shift from C1 to C12 followed by deprotonation to form exomethylene. Subsequent in situ removal of the silyl protecting groups led to the total synthesis of brassicicene K (**27**). Thus, the biosynthetic proposal featuring the Wagner–Meerwein-type skeletal rearrangement (**9**→**11**) through the catalysis of P450 enzyme BscF was successfully emulated by the chemical conversion.

Site- and diastereocontrolled hydrogenation of the resulting exomethylene in **27** yielded brassicicene C (**28**). Further diastereoselective reduction of the C13 ketone completed the total synthesis of brassicicene H (**29**). By integrating the convergent and scalable chemical synthesis of the 5/8/5 scaffold with enzymatic regio- and stereoselective oxidative conversions, Renata and co-workers successfully achieved the collective chemoenzymatic total synthesis of ten natural diterpenes, five with a 5/8/5 tricyclic skeleton and five with a rearranged scaffold including brassicicenes **27**–**29**.

## In situ generation of highly reactive intermediates: trichodimerol and the bisorbicillinoid family

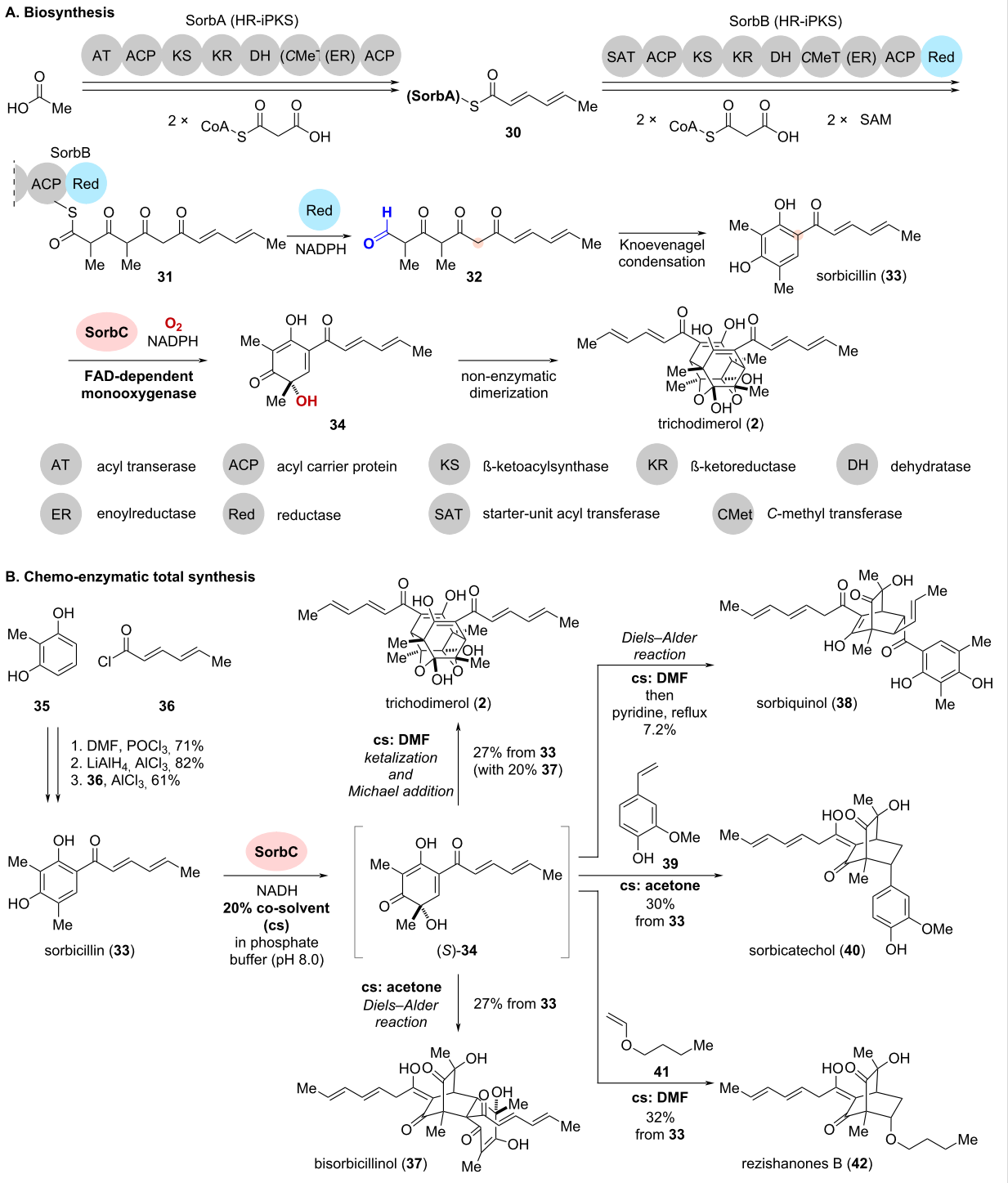
The bisorbicillinoid family, isolated from fungi such as *Penicillium chrysogenum*, has dimeric intricate scaffolds, as exemplified by trichodimerol (**2**) (Scheme 4) [28,29]. The biosynthetic pathway of these natural products involves late-stage skeletal diversification triggered by enzymatic oxidative dearomatization (Scheme 4A). Specifically, FAD-dependent monooxygenase catalyzes the oxidative dearomatization of the aromatic intermediate, leading to subsequent homo- and hetero-dimerization processes. The mechanism was initially postulated by Dreiding [30], and supported through isotope-labelling studies conducted by Abe and co-workers [31-33]. Subsequently, Cox et al. have elucidated the key enzyme SorbC responsible for the dearomatization process [34,35].

Scheme 4A illustrates the biosynthetic pathway as proposed by Cox, based on bioinformatic analyses of polyketide synthase (PKS) modules and in vitro studies. Initially, highly reducing iterative polyketide synthase (HR-iPKS) SorbA forms the thioester **30** on its acyl carrier protein (ACP) domain from acetate and two units of malonyl-CoA. Subsequently, the downstream HR-iPKS module, SorbB, catalyzes both the extension of polyketide chain from **30** using two additional malonyl-CoAs and *C*-methylations exploiting two units of SAM (*S*-adenosyl methionine) to generate thioester **31** on the ACP domain. Then, the reduction (Red) domain located at the C-terminus of SorbB catalyzes conversion of thioester **31** using NADPH to liberate aldehyde **32**, and subsequent Knoevenagel condensation-type cyclization leads to sorbicillin (**33**).

The FAD-dependent monooxygenase SorbC, utilized in the chemo-enzymatic total synthesis of **2**, catalyzes the oxidative dearomatization of **33** via enantioselective hydroxylation using molecular oxygen and generates cyclohexadienone **34**. As demonstrated by Corey [36] and Nicolaou [37], highly reactive intermediate **34** likely dimerizes non-enzymatically through stepwise reactions involving (1) an initial intermolecular Michael addition, (2) a second intramolecular Michael reaction forming a head to tail formal [4 + 4] cycloadduct, and (3) sequential hemiketal formations to furnish the densely-functionalized and complicated pentacyclic scaffold of trichodimerol (**2**). Alternatively, the 2,4-cyclohexadienone moiety of **34** undergoes intermolecular Diels–Alder reactions with **34** or other dienophiles, leading to a diverse array of bisorbicillinoids. Based on the crystal structure of TropB, a homologous enzyme of SorbC, Narayan and co-workers have intensively elucidated that the detailed mechanisms of the oxidative dearomatization catalyzed by the FAD-dependent monooxygenases [38,39].

Gulder and co-workers achieved the chemo-enzymatic total synthesis of bisorbicillinoids family members by merging chemical synthesis and the enzymatic oxidative dearomatization (Scheme 4B) [40,41]. The SorbC-catalyzed enzymatic oxidation of chemically synthesized **33** generated the highly reactive intermediate **34** in aqueous solvents under mild conditions [37,42]. In this process, the co-solvent (cs) allowed control of the dimerization modes via either Michael addition or Diels–Alder reactions, facilitating the systematic total synthesis of the bisorbicillinoid family.

Substrate **33** for SorbC-catalyzed enzymatic transformation was synthesized in 3 steps from phenol **35** via formylation and subsequent reduction to introduce a methyl group, followed by a Friedel–Crafts acylation with **36** (Scheme 4B) [39,40]. The in vitro enzymatic transformation of **33** by recombinant SorbC enantioselectively introduced a hydroxy group to generate the reactive intermediate (*S*)-**34** [34]. Cyclohexadienone (*S*)-**34** was relatively stable in aqueous reaction solvents, however, quenching the enzymatic reaction by the addition of organic solvents led to homo-dimerization. Based on these experimental results, Gulder and co-workers devised a strategy to control the dimerization modes by adjusting the polarity of the organic co-solvent to establish the divergent synthesis of dimeric scaffolds. Indeed, with 20% DMF in the SorbC-catalyzed enzymatic oxidative dearomatization, the Michael addition/ketalization cascades of generated (*S*)-**34** proceeded predominantly, achieving the chemoenzymatic total synthesis of trichodimerol (**2**) in 27% yield from **33**. In contrast, the use of acetone as the co-solvent resulted in the homodimerization of (*S*)-**34** via Diels–Alder reaction between the cyclohexadienone moieties,




**Scheme 4:** (A) Biosynthetic pathway for trichodimerol (**2**) in *Penicillium chrysogenum*. (B) Chemo-enzymatic total synthesis of **2** and bisorbicillinoids utilizing an FAD-dependent monooxygenase SorbC.

allowing the total synthesis of bisorbicillinol (**37**). Meanwhile, treatment of the resulting (*S*)-**34** under harsher conditions, such as reflux in pyridine, resulted in a distinct intermolecular Diels–Alder reaction between the cyclohexadienone moiety and the side chain of (*S*)-**34** enabling the first total synthesis of sorbiquinol (**38**). This chemo-enzymatic synthesis exploiting (*S*)-**34** generated through enzymatic transformation as a common and versatile intermediate, illustrates how judicious choice of the co-solvent can successfully diversify multicyclic complex scaffolds. The enhancement of organic solvent toler-

ance through enzyme mutagenesis should allow more precise control over scaffold-constructing reactions.

As with the related natural products, hybrid sorbicillinoids were also known to be biosynthesized through Diels–Alder reactions between diene **34** and non-sorbicillinoid-derived dienophiles [29]. Gulder and co-workers achieved the total synthesis of hybrid sorbicillinoids using the established chemo-enzymatic synthetic process, in addition to the synthesis of homodimers (Scheme 4B). Upon treatment of the extract of the enzymatic reaction mixture containing (*S*)-**34** with dienophile **39**, sorbicatechol (**40**) was obtained in 30% yield from **33**. The use of dienophile **41** instead of **39** led to the concise total synthesis of rezishanone B (**42**) in 32% yield. The same research group has also facilitated rapid access to bisorbicillinoid analogs, by using either various dienophiles or non-natural type substrates of SorbC [43].

Almost concurrently, Narayan and co-workers substantially expanded the application scope of synthetic methodologies by utilizing the FAD-dependent monooxygenase TropB, AzaH, and SorbC to catalyze oxidative dearomatization [38,39]. Through extensive exploration of substrate diversity, the research group achieved the synthesis of various dearomatized compounds, and the total synthesis of a member of the sorbicillinoid family.

## Enantioselective intermolecular Diels–Alder reaction to assemble core scaffolds: chalcomoracin and kuwanons

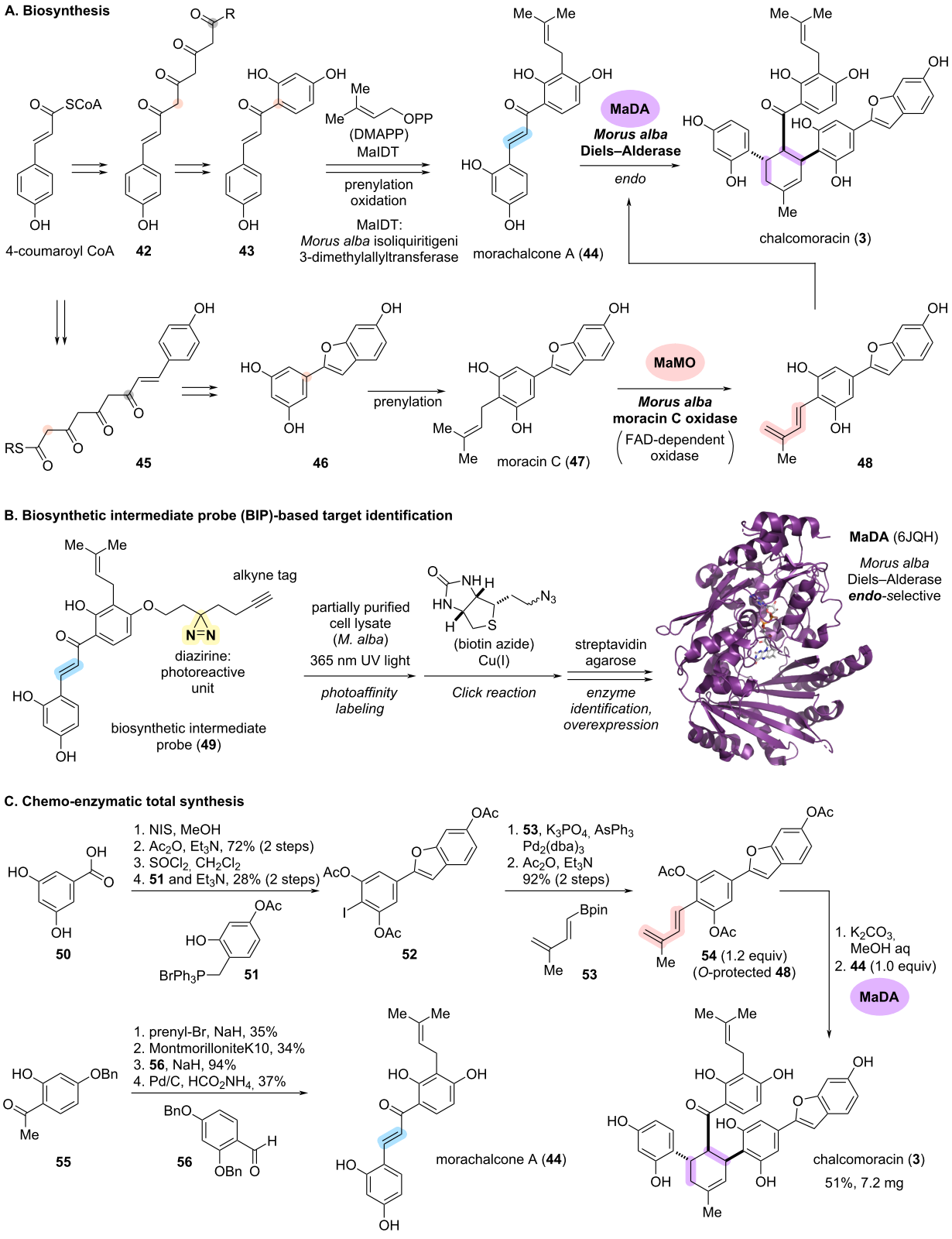
Chalcomoracin (**3**) and kuwanons bearing a highly substituted methyl cyclohexene core were isolated from plants of the *Moraceae* family as phytoalexins (Scheme 5) [44]. Since more than 160 related compounds have been discovered in an optically active form, Masamune, Takahashi, and colleagues postulated that the core scaffold is biosynthesized via Diels–Alder (DA) reactions [45]. In the 1980s, the research group led by Nomura supported this biosynthetic hypothesis by isotope labelling experiments using cell cultures of *Morus alba* [46-48]. Furthermore, the proposed DA reaction precursors to **3**, morachalcone A (**44**) and moracin C (**47**), were also isolated [47,49].

As shown in Scheme 5A, the proposed biosynthetic pathway of chalcomoracin (**3**) commence with the addition of three C2 units from the acetate pathway to 4-coumaroyl-CoA supplied from the shikimate pathway, leading to compound **42** [47]. Subsequent Claisen condensation, dehydration and aromatization provides chalcone **43**. Prenylation of the resultant aromatic ring of **43**, catalyzed by MaIDT (*Morus alba* isoliquiritigenin 3-dimethylallyltransferase), leads to morachalcone A (**44**) [50,51]. In parallel, benzofuran **46** was biosynthesized from 4-coumaroyl-CoA via thioester **45**. Further prenylation of phenol **46** yields moracin C (**47**). As the final steps in the biosynthesis of **3**, oxidation of the prenyl group of **47** to diene **48** and the subsequent DA reaction with dienophile **44** should theoretically be catalyzed by oxidase and Diels–Alderase, respectively. However, despite effort over several decades, these enzymes had not been discovered from *Morus alba*.

In 2020, Lei and co-workers successfully identified two key enzymes in the biosynthesis of **3**, MaMO (*M. alba* moracin C oxidase) and MaDA (*M. alba* Diels–Alderase), by developing a novel method termed "biosynthetic intermediate probe (BIP)-based target identification" (Scheme 5B) [52-54]. In addition, the utilization of MaDA allowed the chemo-enzymatic total synthesis of **3** and related natural products (Scheme 5C).

To identify the Diels–Alderase (MaDA), the research group initially demonstrated an in vivo enzymatic reaction by treating chemically synthesized **44** and **47** with cultured *M. alba* cells. The formation of both diene **48** and DA adduct **3** were observed, indicating that *M. alba* cells harbor the key enzymes MaMO and MaDA. Based on these results, synthetic analogs of intermediate **44** were designed and exposed to the *M. alba* cells to explore the substrate tolerance of MaDA for chemical probe design. After several attempts, Lei and co-workers designed and synthesized chemical probe **49** bearing a diazirine photoaffinity labelling unit with an alkyne tag on the phenolic hydroxy group of biosynthetic intermediate **44** (Scheme 5B). The treatment of fractionated cell cultures of *M. alba* with **49** was followed by irradiation with 365 nm light to generate reactive carbene from diazirine. This sequence allowed the formation of covalent bonds between the synthetic probe and binding proteins. The resulting mixture was subjected to a copper-catalyzed click reaction with biotin azide, which led to selective pull-down with streptavidin agarose and isolation of the probe–protein covalent complex. Proteomic analysis of the isolated proteins narrowed down the MaMO and MaDA candidates, including several berberine bridge enzyme (BBE)-like enzymes. This FAD-linked oxidase family is known to catalyze a variety of oxidative transformations critical for natural products biosynthesis [55]. Further transcriptome analysis of the candidate proteins led to the identification of two BBE-like enzymes, MaMO and MaDA, as key biosynthetic enzymes for **3**. These enzymes were subsequently overexpressed in insect cells and purified as soluble proteins, culminating in the elucidation of the X-ray crystallographic structure of MaDA (Scheme 5B, PDB: 6JQH).

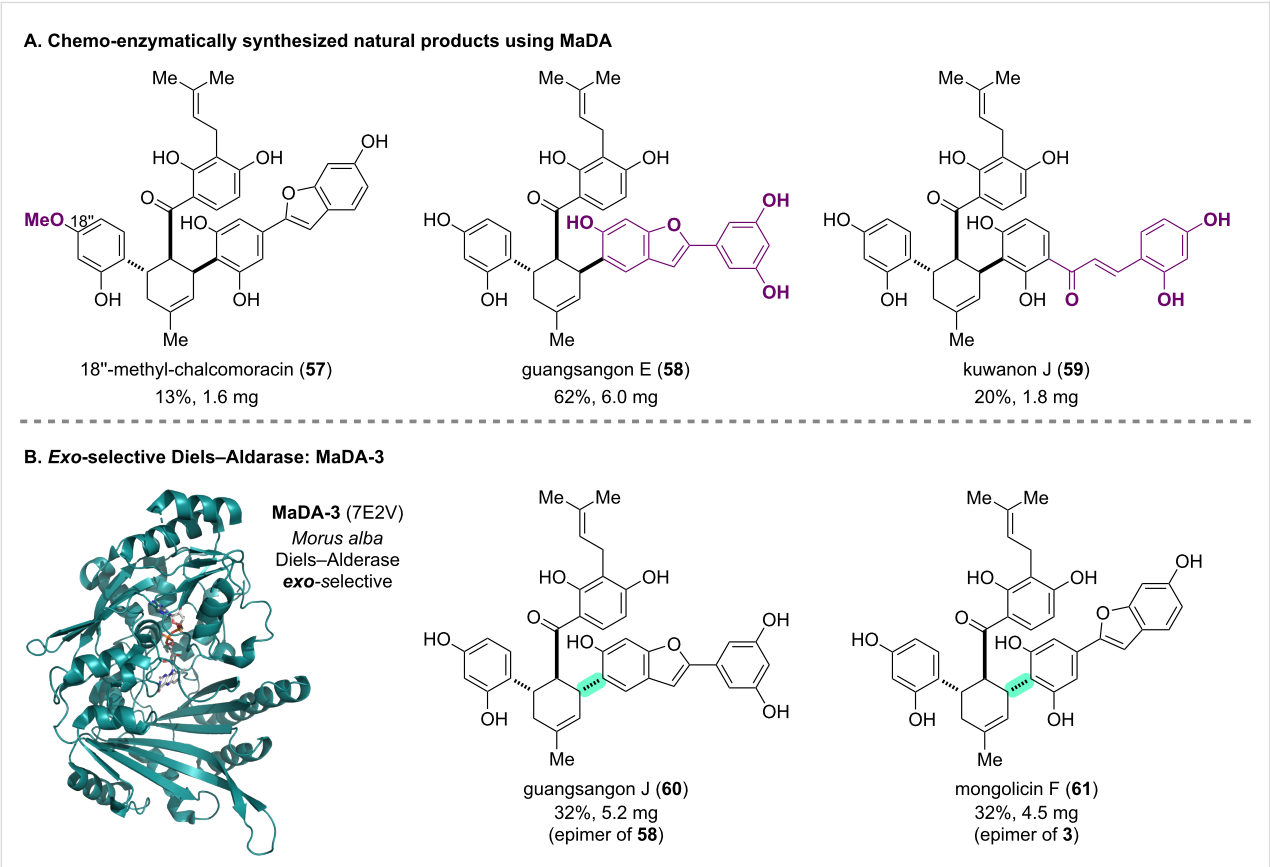
With these key enzymes in hand, Lei and co-workers conducted a chemo-enzymatic total synthesis of **3** and related




**Scheme 5:** (A) Proposed biosynthetic pathway for chalcomoracin (**3**) in *Morus alba*. (B) Outline of the biosynthetic intermediate probe (BIP)-based target identification strategy. (C) Chemo-enzymatic total synthesis of **3** utilizing an identified Diels–Alderase, MaDA.

natural products utilizing the successfully overexpressed Diels–Alderase, MaDA (Scheme 5C). The chemical synthesis of **54**, tri-*O*-acetylated precursor of the diene component **48**, commenced from phenol **50**. Iodination and *O*-acetylations of **50** followed by coupling with phosphorus ylide **51** afforded aryl iodide **52**. Subsequent Suzuki–Miyaura coupling with boronic ester **53** and *O*-acetylation furnished **54**. The dienophile component, morachalcone A (**44**), was synthesized from phenol **55** in four steps including *O*-prenylation and subsequent Claisen rearrangement, aldol condensation with **56**, and deprotection. The key chemo-enzymatic conversions, in situ generation of **48** by deprotection of **54**, followed by treatment with **44** in the presence of purified Diels–Alderase MaDA, facilitated an *endo*-selective DA reaction and led to the concise total synthesis of chalcomoracin (**3**) in 51% yield.

To achieve the systematic total synthesis of the natural product family sharing the highly substituted methyl cyclohexene moiety, Lei and co-workers synthesized analogs of biosynthetic intermediates **44** and **48**, and subjected them to the established chemo-enzymatic synthetic process (Scheme 6A) [56]. MaDA exhibited a relatively broad substrate tolerance, enabling the total synthesis of chiral natural products such as 18”-methyl-chalcomoracin (**57**), guangsangon E (**58**), and kuwanon J (**59**). Furthermore, the same research group identified a series of MaDA homologous enzymes in *Morus notabilis*, designated as MaDA-1–3, which showed distinct stereoselectivities in the Diels–Alder reactions (Scheme 6B). Notably, MaDA-3 exhibited high *exo*-selectivity (original MaDA: *endo*-selective) and enabled the enantioselective rapid total synthesis of guangsangon J (**60**) and mongolicin F (**61**), which are epimers of **58** and **3**, respectively. Comparative analysis of the X-ray crystallographic structures of MaDA (6JQH) and MaDA-3 (7E2V) provided insights into the mechanisms of different stereoselectivities in the Diels–Alder reactions [56]. In 2014, the same group led by Lei also reported the total synthesis of kuwanon J (**59**) and related natural products. In their earlier fully chemical approach, they used a chiral boron catalyst as a Lewis acid and achieved at best an *endo*/*exo* selectivity of 1.9:1 in a similar DA reaction. The use of Diels–Alderase in their recent work significantly improved the *endo*/*exo* selectivity under mild conditions in water, thereby highlighting the strengths of the chemo-enzymatic approach for synthesizing this family of natural products [57].

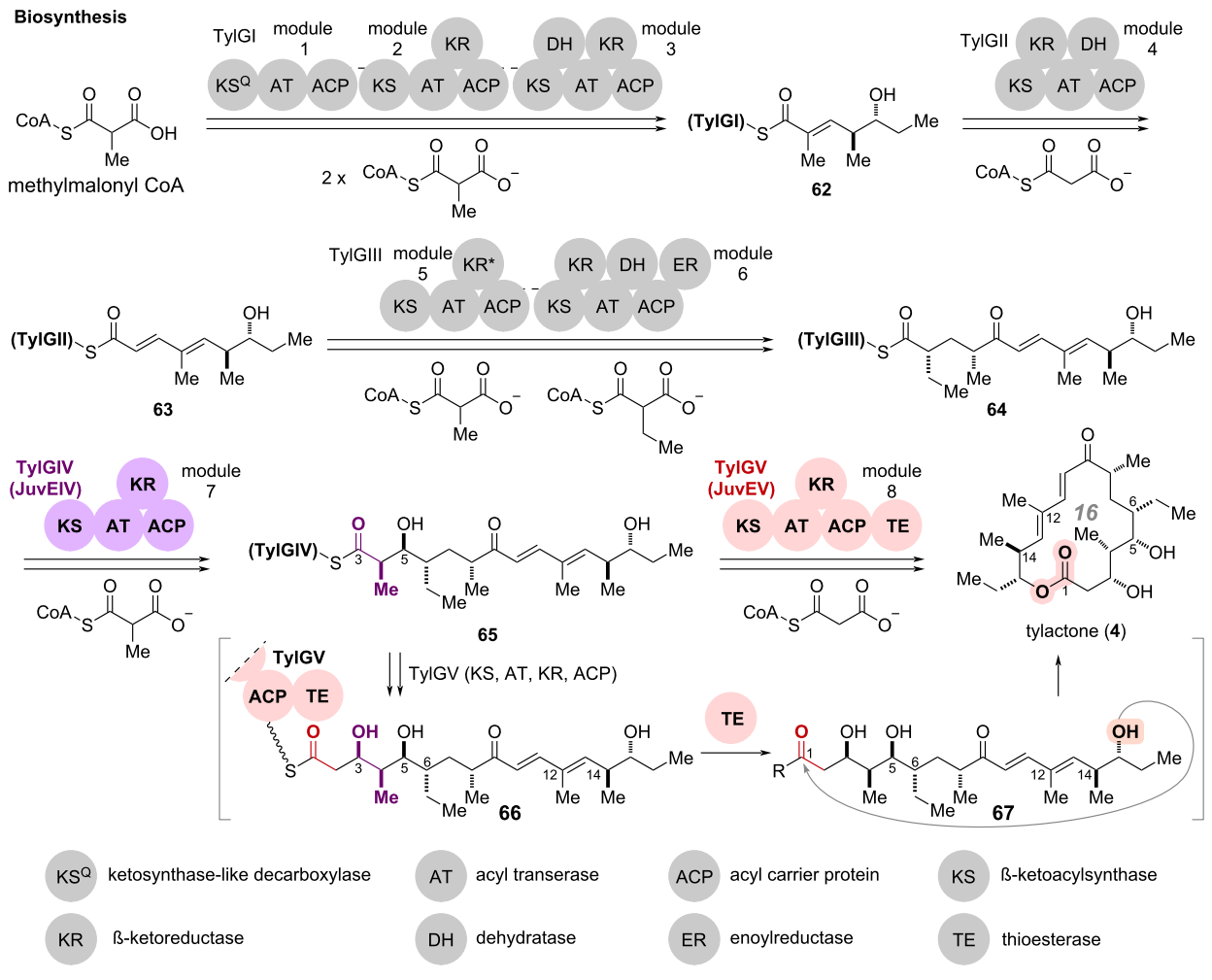



**Scheme 6:** (A) Chemo-enzymatically synthesized natural products by using the originally identified MaDA. (B) MaDA homologue enzyme MaDA-3: X-ray crystal structure and synthesized natural products by leveraging this enzyme.

## Regioselective macrocyclization: tylactone and juvenimicins

In the chemical synthesis of macrolides, macrolactone construction most generally involves the generation of activated esters from the corresponding carboxylic acids followed by intramolecular condensation. As a chemo-enzymatic complementary strategy, thioesterase (TE) domains of NRPS and PKS gene clusters have been utilized as enzymes for macrolactone formation to produce a wide range of natural/non-natural macrocyclic compounds [58-63]. In the field of chemo-enzymatic total synthesis utilizing polyketide synthase (PKS)-related macrocyclization enzymes, Xiang and co-workers recently reported the total synthesis of cylindrocyclophanes [64]. The cyanobacterial non-canonical enzyme CylK, discovered and characterized by Balskus and co-workers [65-67], facilitated an intramolecular Friedel–Crafts-type homo-dimerization and led to the efficient construction of the 22-membered paracyclophane scaffold. Of several sophisticated chemo-enzymatic strategies, this review highlights the total syntheses of tylactone (**4**) and juvenimicins (Scheme 7 and Scheme 8) [68]. This chemo-enzymatic total synthesis represents a pioneering approach, as it ingeniously integrates biosynthetic intermediate mimics, synthesized through multiple steps, into the biosynthetic pathway. A cascade of stereoselective carbon chain elongation and regioselective macrolactonization catalyzed by two massive PKS modules exemplifies a refined and innovative method in chemo-enzymatic synthesis.

Tylactone (**4**), a 16-membered macrolactone, was isolated from a *Streptomyces fradiae* mutant and characterized as the aglycone of the macrolide antibiotic tylosin (Scheme 7) [69]. Since the 1970s, the biosynthetic pathway of **4** has been investigated through isotope labelling and analysis of metabolites from *S. fradiae* mutants [70-73]. Heterologous production of **4** was also achieved by expression of elucidated biosynthetic gene cluster from *S. fradiae* in *Streptomyces venezuelae* [74-77].




**Scheme 7:** Proposed biosynthetic mechanism of tylactone (**4**) in *Streptomyces fradiae*.

In the proposed biosynthesis of **4** as illustrated in Scheme 7, five units of the type I polyketide synthase (PKS), TylGI–GV, comprising eight modules catalyze the polyketide chain extension reactions and a macrocyclization (Scheme 7) [78]. Initially, the acyltransferase (AT) domain of TylGI loads a methylmalonyl-CoA onto the acyl carrier protein (ACP) in module 1. The ketosynthase-like decarboxylase ($KS^Q$) domain catalyzes the decarboxylation of the loaded methylmalonyl moiety, and subsequent modules 2 and 3 extend the carbon chain using two molecules of malonyl-CoA. The β-ketoreductase (KR) and dehydratase (DH) domains, sequentially catalyze the carbonyl reduction and dehydration of the extended polyketide chains to provide thioester **62** connecting to the ACP domain of module 3. TylGII then iterates similar conversions including carbon chain extension reaction, carbonyl reduction, and dehydration to generate the thioester **63** on its ACP domain. Module 5 of TylGIII further extends the carbon chain with incorporation of methylmalonyl-CoA and ethylmalonyl-CoA. Then the KR, DH, and enoylreductase (ER) domains of module 6 catalyze the formation and reduction of enone to furnish thioester **64** on the ACP domain.

Subsequent carbon chain extension and macrolactonization catalyzed by two PKS modules, TylGIV and TylGV, are pivotal enzymatic transformations for the chemo-enzymatic total synthesis of **4**. TylGIV catalyzes the carbon chain extension reaction using a methylmalonyl-CoA and the stereoselective reduction of the C5 carbonyl group to form thioester **65**. The subsequent PKS module TylGV introduces two more carbons from malonyl-CoA and reduces the carbonyl group at the C3 position to produce thioester **66** on the ACP domain. The thioesterase (TE) domain then catalyzes the release of **67** from the ACP domain and regioselective macrolactonization to furnish the 16-membered tylactone (**4**).

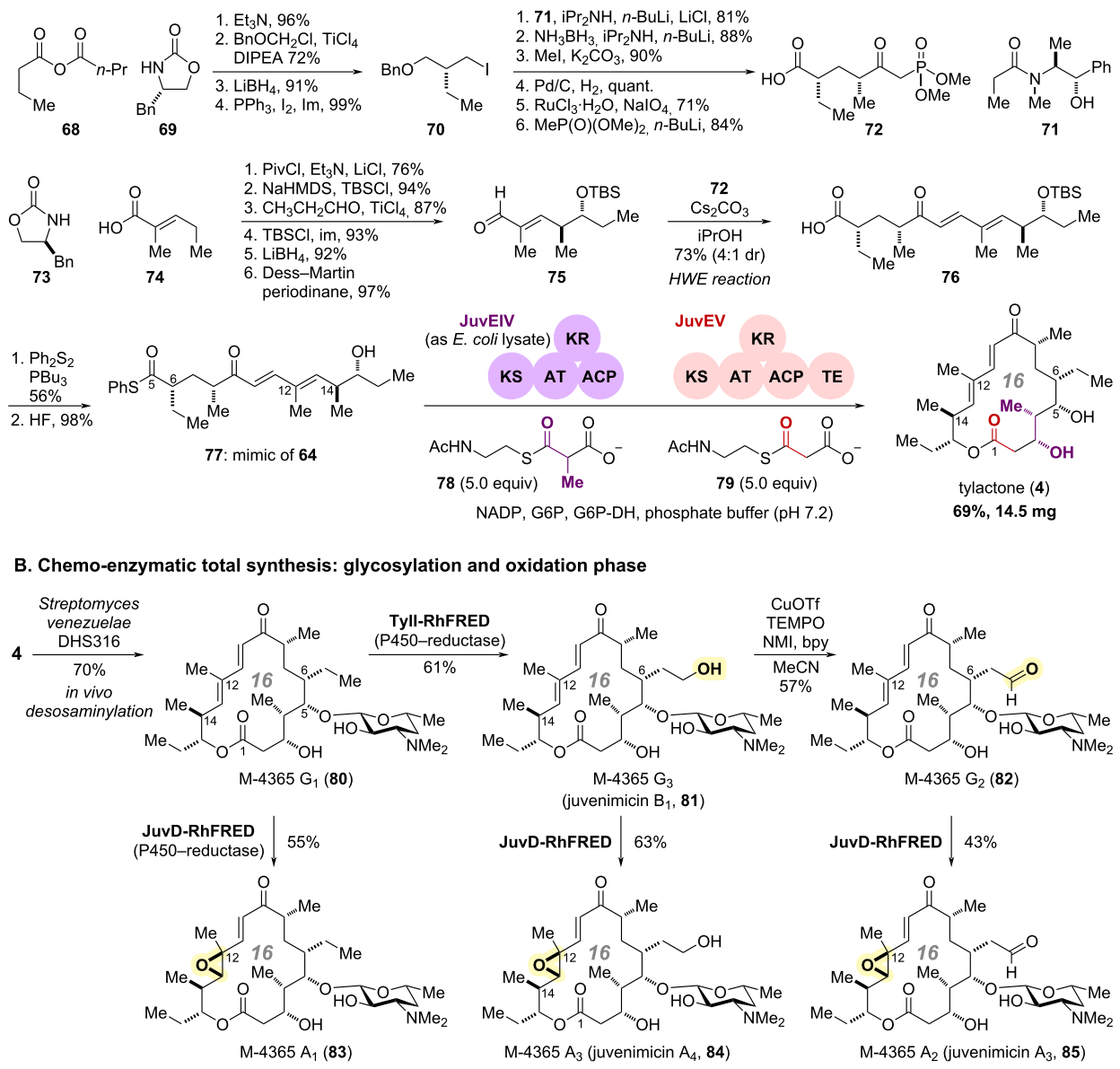
The research group led by Sherman accomplished the efficient chemo-enzymatic total synthesis of tylactone (**4**) and a series of M-4365 (juvenimicin) by merging consecutive enzymatic transformations by JuvEIV and JuvEV, homologous enzymes to TylGIV and TylGV (Scheme 8) [68]. Reconstruction of the JuvEIV–JuvEV-catalyzed cascade enzymatic conversions successfully furnished **4** in one pot from chemically synthesized thioester **77** mimicking biosynthetic intermediate **64**.

For the chemical synthesis of mimic **77**, the left-half segment **72** was prepared from mixed anhydride **68** (Scheme 8A). Condensation with chiral auxiliary **69** and subsequent diastereoselective alkylation, followed by reductive removal of the auxiliary and iodination of the resulting primary alcohol provided alkyl iodide **70** [78]. Subsequent six-step transformations including diastereoselective alkylation of (+)-pseudoephedrine derivative **71** [79] with the iodide **70**, sequential functional group manipulations, and installation of the β-ketophosphonate, provided the left-half segment **72** [80]. The right-half segment **75** was synthesized in six steps via condensation of **73** and **74**, followed by highly-diastereoselective vinylogous Mukaiyama aldol reaction and subsequent protecting group manipulations [81-83]. The separately synthesized left and right segments, **72** and **75**, were then assembled via Horner–Wadsworth–Emmons (HWE) olefination to give carboxylic acid **76** in a 4:1 diastereomeric ratio. Thioesterification and removal of the TBS group furnished the desired thioester **77** as a mimic of biosynthetic intermediate **64**.

With synthetic substrate **77** for enzymatic conversions in hand, Sherman and co-workers attempted to overexpress TylGIV and TylGV, however, they were faced with low expression levels of TylGV. To circumvent this problem, they employed the homologous enzymes JuvEIV and JuvEV from juvenimicin-producing *Micromonospora chalcea* subsp. *izumensis* [84,85]. The Juv gene cluster contained JuvEI–JuvEV with high homology to TylGI–TylGV, and P450 monooxygenases JuvC and JuvD for post-macrocyclization oxidative modifications.

Based on the successful cloning and high-level overexpression of the two key PKS modules, JuvEIV and JuvEV, cascade enzymatic conversions of **77** were conducted. The SNAC (*N*-acetylcysteamine) esters **78** and **79** were used as surrogates for methyl malonyl-CoA and malonyl-CoA in the enzymatic conversions. After careful and systematic optimization of the reaction conditions, such as the stoichiometry of the PKS modules and pH, along with the application of a NADPH recycling system, JuvEIV/JuvEV-catalyzed enzymatic conversions of **77**–**79** afforded tylactone (**4**) in one pot with 69% yield. By orchestrating the cooperation of two massive PKS modules (>150 kDa), this chemoenzymatic process enabled: (1) incorporation of fully synthetic hexaketide **77** as a substrate, (2) installation of the last four-carbon polyketide chain, (3) stereoselective introduction of three consecutive chiral centers, and (4) successful completion of the macrolactonization of the resulting octaketide to produce tylactone (**4**). This chemo-enzymatic process demonstrated highly efficient in vitro cascade transformations, underscoring the potential of integrating enzymatic catalysis with chemical synthesis [86].

By taking advantage of the chemo-enzymatically accessible **4**, Sherman and co-workers further implemented the systematic total synthesis of juvenimicins and the M-4365 series via enzymatic and chemical late-stage modifications (Scheme 8B) [68]. In vivo glycosylation utilizing *Streptomyces venezuelae* DHS316, developed by the same group, mediated the desosaminylation of **4** at C5 and afforded M-4365 $G_1$ (**80**) in



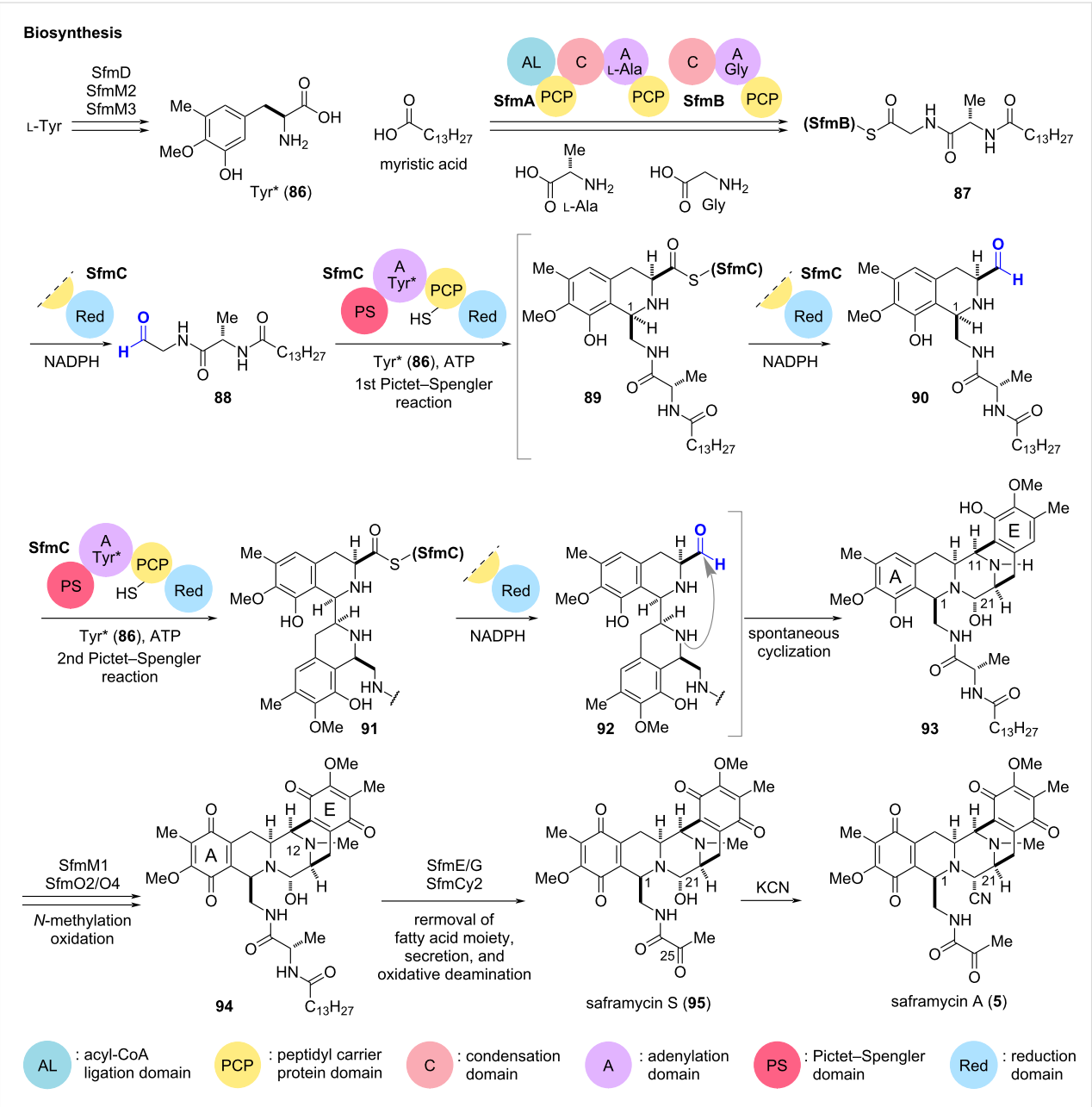

**Scheme 8:** (A) Chemical synthesis and cascade enzymatic transformations of cyclization precursors. (B) Late-stage modifications of chemo-enzymatically synthesized aglycone **4** utilizing in vivo/in vitro enzymatic transformations and chemical conversions.

70% yield [87]. For further oxidative modifications, they prepared the fusion proteins TylI-RhFRED and JuvD-RhFRED, comprising P450 monooxygenases (TylI, JuvD) with the P450 RhF reductase domain (RhFRED) [88]. TylI-RhFRED facilitated the in vitro site-selective enzymatic hydroxylation of the C6 ethyl substituent in **80**, yielding M-4365 $G_3$ (juvenimicin $B_1$, **81**). Subsequent chemoselective oxidation of the resulting primary alcohol furnished M-4365 $G_2$ (**82**) bearing an aldehyde at the C6 sidechain. Furthermore, the fusion protein, JuvD-RhFRED, enabled regio- and diastereoselective epoxidation of the C12 double bond of macrocycles **80**–**82**, culminating in the total synthesis of M-4365 $A_1$ (**83**), M-4365 $A_3$ (juvenimicin $A_4$, **84**), and M-4365 $A_2$ (juvenimicin $A_3$, **85**), respectively.

## Iterative Pictet–Spengler cyclizations: saframycin A and jorunnamycin A

Saframycin A (**5**) was isolated from *Streptomyces lavendulae*, and a number of related alkaloid families such as safracins and renieramycins have been identified from both soil and marine microorganisms [89,90]. These natural product families, known as bistetrahydroisoquinoline (THIQ) alkaloids, share a highly functionalized pentacyclic scaffold with different aromatic ring

oxidation states and sidechain structures [90-93]. The biosynthetic mechanism of **5** has been extensively studied by gene disruption and reconstruction of in vitro enzymatic conversions (Scheme 9) [94]. The biosynthesis of **5** begins with the conversion of L-tyrosine to tyrosine derivative **86** (Tyr*) by peroxygenase SfmD and the methyltransferases, SfmM2 and M3 [95,96]. Concurrently, two non-ribosomal peptide synthetase (NRPS) modules, SfmA and SfmB, catalyze the successive condensation of myristic acid, L-alanine, and glycine to furnish thioester **87** on the peptidyl carrier protein (PCP) domain of SfmB [94,97]. The downstream NRPS module, SfmC, then catalyzes sequential reactions with **86** and **87** to assemble pentacyclic scaffold **93** in a single stroke [97-99]. The reduction (Red) domain at the C-terminus of SfmC reduces the thioester **87** to release aldehyde **88** from SfmB, while the adenylation (A) domain activates **86** and loads it onto the PCP domain of SfmC. The Pictet–Spengler (PS) domain then catalyzes the first diastereoselective PS cyclization to form bicyclic thioester **89** with incorporation of a stereogenic center at C1. The Red domain liberates bicyclic aldehyde **90** by reducing the resulting thioester **89**, while the A domain again activates another molecule of tyrosine derivative **86**, facilitating its loading onto the




**Scheme 9:** Proposed biosynthetic mechanism of saframycin A (**5**) in *Streptomyces lavendulae*.

PCP domain. In the second cycle, the PS domain catalyzes the assembly of the tyrosine derivative and the liberated aldehyde **90** to form tetracyclic thioester **91** bearing an additional chiral center at C11. Subsequent reduction of **91** to aldehyde **92** by the Red domain and spontaneous cyclization would furnish **93**, the pentacyclic core scaffold of **5**. Overall, the single NRPS module SfmC is responsible for the construction of the highly functionalized scaffold **93** from the two simple amino acid derivatives **86** and **88**.

After NRPS-catalyzed scaffold assembly, SfmM1-mediated *N*-methylation at N12, and subsequent SfmO2/O4-promoted oxidation of phenols on A- and E-rings, yield bisquinone **94**. The membrane-bound peptidase SfmE excises the long-chain fatty acid moiety on the N-terminus of the C1 side chain. Secretion through the efflux pump SfmG, and extracellular oxidative deamination catalyzed by the berberine bridge enzyme (BBE)-like enzyme SfmCy2, lead to saframycin S (**95**) [100,101]. To convert the labile hemiaminal moiety at C21 to a more stable aminonitrile, the extracts of the microorganism containing **95** were treated with KCN, allowing the isolation as saframycin A (**5**) [89].

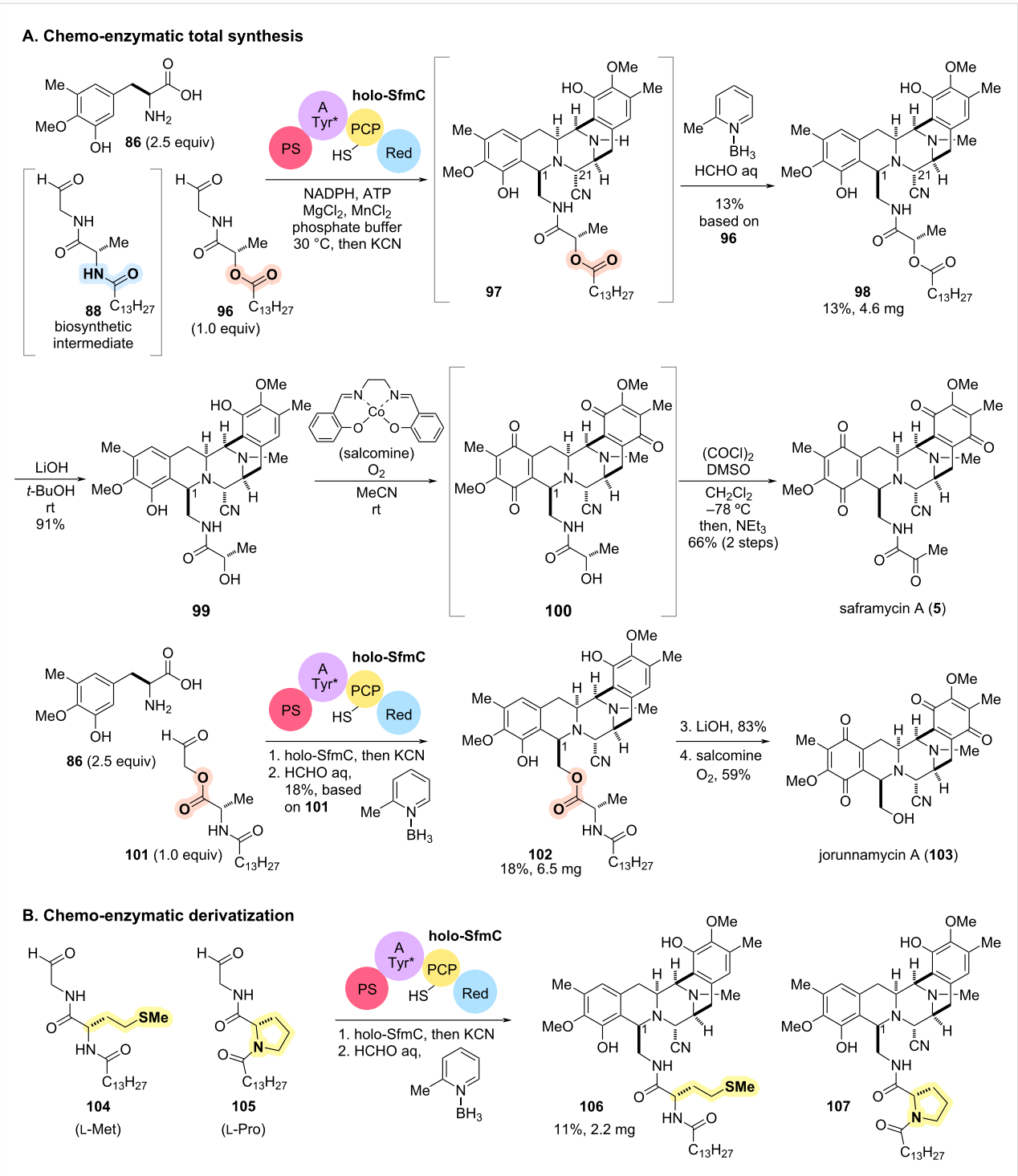
A notable feature of this biosynthetic machinery is the attachment and detachment of the fatty acid moiety. The myristic acid, introduced by SfmA at the beginning, is not attached in the final product **5**. However, detailed investigations by Oikawa and co-workers clearly demonstrated that this fatty acid is essential for SfmC-catalyzed scaffold assembly [97-99]. Furthermore, recent gene disruption studies by Tang and co-workers have also shown that the biosynthetic pathway of safracin, another member of the THIQ family, incorporates palmitic acid ($C_{16}$) at the initial stage of NRPS-catalyzed scaffold assembly [101].

To achieve the chemo-enzymatic synthesis of the bis-THIQ alkaloid family, Oguri and Oikawa utilized the NRPS module SfmC to construct a highly functionalized scaffold from amino acid derivatives (Scheme 10) [102]. The total synthesis of **5** from original biosynthetic intermediate **93** requires regioselective amide bond cleavage to remove the fatty acid moiety on the C1 side chain (Scheme 9). Although the enzymatic deacylation of **94** can be catalyzed by the membrane-bound peptidase SfmE, similar chemical transformations require harsh conditions and are difficult to perform in the presence of highly reactive functional groups. To circumvent this problem, the research group designed synthetic substrate analogs (**96**, **101**) that mimic the biosynthetic intermediate **88** (Scheme 10A). By replacing an amide linkage in **88** with an ester linkage, the fatty acid side chain could be removed under mild conditions after enzymatic construction of the pentacyclic scaffolds.

The designed and chemically synthesized substrate analog **96** was tolerated by the key NRPS module SfmC despite the structural modification (Scheme 10A). In vitro enzymatic conversion of **96** with tyrosine derivative **86** [103] followed by the addition of KCN concisely furnished the pentacyclic secondary amine **97** bearing the ester linker in the C1 side chain in one pot. After removal of SfmC by precipitation and centrifugation, the reaction mixture containing secondary amine **97** was subjected to the reductive amination using 2-picoline borane as a hydride source, yielding tertiary amine **98** [104]. Without isolation of any intermediate from the simple synthetic substrates **88** and **96**, rapid access to **98** was achieved in 13% yield over two pots based on peptidyl aldehyde **96**. From chemo-enzymatically constructed scaffold **98**, chemical manipulations of functional groups were investigated to achieve the total synthesis of **5** (Scheme 10A). Basic hydrolysis of the ester on the C1 side chain of **98** proceeded under mild conditions and furnished secondary alcohol **99**. Subsequent oxidation of the two phenol rings of **99**, catalyzed by the cobalt complex salcomine, afforded bisquinone **100**. The subsequent Swern oxidation of the resulting secondary alcohol **100** allowed the total synthesis of saframycin A (**5**) in five pots, six steps from the simple synthetic substrates, **86** and **96** [102,105].

By adopting the same strategy to another synthetic substrate analog **101**, the research group also achieved the concise chemo-enzymatic total synthesis of jorunnamycin A (**103**). SfmC-catalyzed enzymatic conversion followed by cyanation and *N*-methylation also converted substrate analog **101** to the corresponding pentacyclic tertiary amine **102** in 18% overall yield based on peptidyl aldehyde **101**. Subsequent simple chemical hydrolysis of the ester and oxidation of the phenol rings allowed concise access to jorunnamycin A (**103**) in just 4 pots from **86** and **101** [102,106]. In subsequent work, Tanifuji and Oguri designed and applied eight variants of the peptidyl aldehyde (e.g., **104**, **105**) bearing various L- and D-amino acids in place of L-alanine in **88** to the SfmC-catalyzed chemo-enzymatic process [107]. The key enzyme SfmC tolerated all the synthetic analogs of the biosynthetic intermediate **88** and facilitated the rapid synthesis of non-natural bis-THIQ alkaloid-like scaffolds such as **106**, **107** (Scheme 10B).

The essence of this chemo-enzymatic total synthesis was the strategic focus on peptidyl aldehyde **88**, the intermediate that transiently dissociates from the enzymes in the biosynthetic pathway. Leveraging the broad substrate tolerance of SfmC towards **88** facilitated the design of substrate analogs that streamlined enzymatic conversions and subsequent chemical transformations, culminating in the concise total synthesis. This rational and flexible synthetic strategy provided rapid access to the THIQ alkaloid family with a diverse array of side chains by




**Scheme 10:** (A) Chemo-enzymatic total synthesis of saframycin A (**5**) and jorunnamycin A (**103**). (B) Chemo-enzymatic structural diversification of the side chain at the C1 position of the bis-THIQ scaffolds.

manipulating simple substrate structures. As demonstrated by SfmC, the "catalytic promiscuity" would enhance the utility of enzymes as synthetic tools and facilitate rapid access to a diverse array of natural product analogs through integration with chemical synthesis [108].

## Conclusion

In this review, recent advancements in the field of chemo-enzymatic total synthesis were categorized into three distinct classifications based on the type of enzymatic conversions: 1) regio- and stereoselective late-stage functionalization of core scaf-

folds, 2) in situ generation of highly reactive intermediates, and 3) one-step construction of macrocyclic or fused multicyclic scaffolds. This classification, along with parallel discussions of the original biosynthetic pathways, helps organize the current state of the art and offers a comprehensive overview of how synthetic methodologies leverage the natural biosynthetic pathways.

The development and optimization of chemo-enzymatic synthetic processes relies heavily on enzyme selection and substrate design. The first strategy, "site- and stereoselective late-stage modification", was highlighted by demonstrating the systematic total synthesis of cotylenol (**1**) and its related natural products, brassicicenes (Scheme 2 and Scheme 3). Natural and engineered P450 enzymes catalyzed the site- and stereoselective oxidative modifications enabling further chemical transformations. Since the scaffolds of terpenes and ribosomally synthesized and post-translationally modified peptides (RiPPs) are first biosynthesized and then modified, the strategy of late-stage enzymatic functionalization of core scaffolds could be effective. Selectivity towards target biomolecules can be tailored by gradually increasing the oxidation level of the complex scaffold or by further site- and stereoselective modifications. Through rational enzyme engineering, the selectivity, efficiency, and robustness of these biocatalysts are expected to enhance their usefulness and generality as synthetic tools.

Apart from oxygenated terpenes and RiPPs, chemo-enzymatic approaches for polyketides and non-ribosomal peptides (NRPs) are considered more suitable for the second and third strategies due to the modular nature of their biosynthetic assembly lines. The second strategy, "in situ generation of highly reactive intermediates", led to the total synthesis of trichodimerol (**2**) and related bisorbicillinoids (Scheme 4). The enzymatic approach, which generates highly reactive, chiral intermediates from stable and achiral compounds, is complementary to chemical methods and is concise. In addition, adjusting the reaction environment with co-solvents and controlling the selectivity of subsequent dimerization may provide an avenue for the future design and application of more sophisticated artificial reaction environments for precise reactivity control.

The third strategy, "regio- and stereoselective scaffold construction", enabled concise access to highly functionalized complex core skeletons of natural products. From the discovery of unknown enzymes to the application of the discovered enzymes to synthetic intermediates, the chemo-enzymatic total synthesis of chalcomoracin (**3**) was comprehensively illustrated (Scheme 5 and Scheme 6). The biosynthetic pathway of natural products in plants still remains challenging due to the complexity and redundancy of their genes, making it difficult to identify the corresponding enzymes. The developed "biosynthetic intermediate probe (BIP)-based target identification" method, a chemical pull-down approach for identifying the target enzymes, would be applied and expanded to the chemo-enzymatic synthesis of other natural products.

In the total synthesis of tylactone (**4**), the combination of two large PKS modules with synthetic substrate mimics facilitated the multistep cascade reaction including carbon chain extension, installation of chiral centers, and macrolactonization in one pot (Scheme 7 and Scheme 8). Moreover, the "site- and stereoselective late-stage modification", presented as the first strategy in this review, was also applied, leading to the systematic chemo-enzymatic synthesis of the related natural products, M-4365 series and juvenimicins. The strategic use of enzymatic and chemical transformations, in accordance with substrate specificity, affords a novel paradigm for exploiting the inherent specificity and efficiency of enzymatic catalysts within synthetic sequences.

The integration of designed substrate analogs that can be transformed by promiscuous enzymes provides a versatile synthetic platform towards natural products and their derivatives. In the total synthesis of saframycin A (**5**) and jorunnamycin A (**103**), the application of a non-standard NRPS module to the synthetic substrate analogs realized the rapid construction of an appropriately functionalized complex scaffold in one pot, with precise control of regio- and diastereoselectivity (Scheme 9 and Scheme 10).

A major limitation of the chemo-enzymatic approach is the current difficulty of designing or evolving enzymes. The chemical synthesis of each substrate and the introduction of mutations into enzymes must be verified to increase their reactivity and selectivity. As the number of applications of chemo-enzymatic hybrid syntheses increases and we better understand the extent to which the structure of enzymes and the reactivity of synthetic substrates can be predicted, guidelines for the rational design of enzymes will likely be established, enabling the rapid identification of optimal substrate–enzyme combinations. The integrative approach of combining intriguing enzymes in ever-evolving biosynthetic research with chemical synthesis, would foster collaborative breakthroughs in the interdisciplinary landscape of natural product biosynthesis, total synthesis as well as function- and diversity-oriented synthesis.

## Author Contributions

Ryo Tanifuji: conceptualization; visualization; writing – original draft; writing – review & editing. Hiroki Oguri: supervision.

## ORCID® iDs

Ryo Tanifuji - https://orcid.org/0009-0001-9466-9084
Hiroki Oguri - https://orcid.org/0000-0001-8007-1631

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

1. Li, J.; Amatuni, A.; Renata, H. *Curr. Opin. Chem. Biol.* **2020,** *55,* 111–118. doi:10.1016/j.cbpa.2020.01.005
2. Moore, B. S.; Gulder, T. A. M. *Nat. Prod. Rep.* **2020,** *37,* 1292–1293. doi:10.1039/d0np90038a
3. Chakrabarty, S.; Romero, E. O.; Pyser, J. B.; Yazarians, J. A.; Narayan, A. R. H. *Acc. Chem. Res.* **2021,** *54,* 1374–1384. doi:10.1021/acs.accounts.0c00810
4. Wang, W.; Taber, D. F.; Renata, H. *Chem. – Eur. J.* **2021,** *27,* 11773–11794. doi:10.1002/chem.202101232
5. Gao, L.; Yang, J.; Lei, X. *Tetrahedron Chem* **2022,** *2,* 100013. doi:10.1016/j.tchem.2022.100013
6. Kaspar, F.; Schallmey, A. *Curr. Opin. Biotechnol.* **2022,** *77,* 102759. doi:10.1016/j.copbio.2022.102759
7. Vanable, E. P.; Habgood, L. G.; Patrone, J. D. *Molecules* **2022,** *27,* 6373. doi:10.3390/molecules27196373
8. Paulsel, T. Q.; Williams, G. J. *ChemBioChem* **2023,** *24,* e202300386. doi:10.1002/cbic.202300386
9. Li, F.; Deng, H.; Renata, H. *Nat. Synth.* **2023,** *2,* 708–718. doi:10.1038/s44160-023-00358-8
10. Stout, C. N.; Wasfy, N. M.; Chen, F.; Renata, H. *J. Am. Chem. Soc.* **2023,** *145,* 18161–18181. doi:10.1021/jacs.3c03422
11. Loskot, S. A.; Romney, D. K.; Arnold, F. H.; Stoltz, B. M. *J. Am. Chem. Soc.* **2017,** *139,* 10196–10199. doi:10.1021/jacs.7b05196
12. Zhang, X.; King-Smith, E.; Dong, L.-B.; Yang, L.-C.; Rudolf, J. D.; Shen, B.; Renata, H. *Science* **2020,** *369,* 799–806. doi:10.1126/science.abb8271
13. Li, J.; Li, F.; King-Smith, E.; Renata, H. *Nat. Chem.* **2020,** *12,* 173–179. doi:10.1038/s41557-019-0407-6
14. Li, F.; Renata, H. *J. Am. Chem. Soc.* **2021,** *143,* 18280–18286. doi:10.1021/jacs.1c08696
15. Li, J.; Chen, F.; Renata, H. *J. Am. Chem. Soc.* **2022,** *144,* 19238–19242. doi:10.1021/jacs.2c09048
16. Li, F.; Deng, H.; Renata, H. *J. Am. Chem. Soc.* **2022,** *144,* 7616–7621. doi:10.1021/jacs.2c02958
17. Yang, J.; Singh, B.; Cohen, G.; Ting, C. P. *J. Am. Chem. Soc.* **2023,** *145,* 19189–19194. doi:10.1021/jacs.3c07078
18. Kelly, S. P.; Shende, V. V.; Flynn, A. R.; Dan, Q.; Ye, Y.; Smith, J. L.; Tsukamoto, S.; Sigman, M. S.; Sherman, D. H. *J. Am. Chem. Soc.* **2022,** *144,* 19326–19336. doi:10.1021/jacs.2c06631
19. Jiang, Y.; Renata, H. *Nat. Chem.* **2024,** in press. doi:10.1038/s41557-024-01533-w
20. Sassa, T.; Tojyo, T.; Munakata, K. *Nature* **1970,** *227,* 379. doi:10.1038/227379a0
21. Sassa, T.; Negoro, T.; Ueki, H. *Agric. Biol. Chem.* **1972,** *36,* 2281–2285. doi:10.1080/00021369.1972.10860584
22. Sassa, T. *Agric. Biol. Chem.* **1972,** *36,* 2037–2039. doi:10.1080/00021369.1972.10860517
23. Tazawa, A.; Ye, Y.; Ozaki, T.; Liu, C.; Ogasawara, Y.; Dairi, T.; Higuchi, Y.; Kato, N.; Gomi, K.; Minami, A.; Oikawa, H. *Org. Lett.* **2018,** *20,* 6178–6182. doi:10.1021/acs.orglett.8b02654
24. Ono, Y.; Minami, A.; Noike, M.; Higuchi, Y.; Toyomasu, T.; Sassa, T.; Kato, N.; Dairi, T. *J. Am. Chem. Soc.* **2011,** *133,* 2548–2555. doi:10.1021/ja107785u
25. Kato, N.; Okamoto, H.; Takeshita, H. *Tetrahedron* **1996,** *52,* 3921–3932. doi:10.1016/s0040-4020(96)00059-2
26. Zallot, R.; Oberg, N.; Gerlt, J. A. *Biochemistry* **2019,** *58,* 4169–4182. doi:10.1021/acs.biochem.9b00735
27. Uwamori, M.; Osada, R.; Sugiyama, R.; Nagatani, K.; Nakada, M. *J. Am. Chem. Soc.* **2020,** *142,* 5556–5561. doi:10.1021/jacs.0c01774
28. Harned, A. M.; Volp, K. A. *Nat. Prod. Rep.* **2011,** *28,* 1790–1810. doi:10.1039/c1np00039j
29. Meng, J.; Wang, X.; Xu, D.; Fu, X.; Zhang, X.; Lai, D.; Zhou, L.; Zhang, G. *Molecules* **2016,** *21,* 715. doi:10.3390/molecules21060715
30. Trifonov, L. S.; Dreiding, A. S.; Hoesch, L.; Rast, D. M. *Helv. Chim. Acta* **1981,** *64,* 1843–1846. doi:10.1002/hlca.19810640616
31. Abe, N.; Sugimoto, O.; Tanji, K.-i.; Hirota, A. *J. Am. Chem. Soc.* **2000,** *122,* 12606–12607. doi:10.1021/ja003013l
32. Abe, N.; Yamamoto, K.; Hirota, A. *Biosci., Biotechnol., Biochem.* **2000,** *64,* 620–622. doi:10.1271/bbb.64.620
33. Abe, N.; Arakawa, T.; Yamamoto, K.; Hirota, A. *Biosci., Biotechnol., Biochem.* **2002,** *66,* 2090–2099. doi:10.1271/bbb.66.2090
34. Fahad, A. a.; Abood, A.; Fisch, K. M.; Osipow, A.; Davison, J.; Avramović, M.; Butts, C. P.; Piel, J.; Simpson, T. J.; Cox, R. J. *Chem. Sci.* **2014,** *5,* 523–527. doi:10.1039/c3sc52911h
35. Kahlert, L.; Bassiony, E. F.; Cox, R. J.; Skellam, E. J. *Angew. Chem., Int. Ed.* **2020,** *59,* 5816–5822. doi:10.1002/anie.201915486
36. Barnes-Seeman, D.; Corey, E. J. *Org. Lett.* **1999,** *1,* 1503–1504. doi:10.1021/ol991070h
37. Nicolaou, K. C.; Vassilikogiannakis, G.; Simonsen, K. B.; Baran, P. S.; Zhong, Y.-L.; Vidali, V. P.; Pitsinos, E. N.; Couladouros, E. A. *J. Am. Chem. Soc.* **2000,** *122,* 3071–3079. doi:10.1021/ja9942843
38. Baker Dockrey, S. A.; Lukowski, A. L.; Becker, M. R.; Narayan, A. R. H. *Nat. Chem.* **2018,** *10,* 119–125. doi:10.1038/nchem.2879
39. Benitez, A. R.; Tweedy, S. E.; Baker Dockrey, S. A.; Lukowski, A. L.; Wymore, T.; Khare, D.; Brooks, C. L., III; Palfey, B. A.; Smith, J. L.; Narayan, A. R. H. *ACS Catal.* **2019,** *9,* 3633–3640. doi:10.1021/acscatal.8b04575
40. Sib, A.; Gulder, T. A. M. *Angew. Chem., Int. Ed.* **2017,** *56,* 12888–12891. doi:10.1002/anie.201705976
41. Sib, A.; Gulder, T. A. M. *Angew. Chem., Int. Ed.* **2018,** *57,* 14650–14653. doi:10.1002/anie.201802176
42. Nicolaou, K. C.; Jautelat, R.; Vassilikogiannakis, G.; Baran, P. S.; Simonsen, K. B. *Chem. – Eur. J.* **1999,** *5,* 3651–3665. doi:10.1002/(sici)1521-3765(19991203)5:12<3651::aid-chem3651>3.0.co;2-y
43. Milzarek, T. M.; Gulder, T. A. M. *Commun. Chem.* **2023,** *6,* 187. doi:10.1038/s42004-023-00996-1
44. Tortora, C.; Pisano, L.; Vergine, V.; Ghirga, F.; Iazzetti, A.; Calcaterra, A.; Marković, V.; Botta, B.; Quaglio, D. *Molecules* **2022,** *27,* 7580. doi:10.3390/molecules27217580
45. Takasugi, M.; Nagao, S.; Masamune, T.; Shirata, A.; Takahashi, K. *Chem. Lett.* **1980,** *9,* 1573–1576. doi:10.1246/cl.1980.1573

46. Nomura, T.; Hano, Y.; Suzuko, S.; Kohno, H. *Heterocycles* **1988,** *27,* 75. doi:10.3987/com-87-4365
47. Hano, Y.; Nomura, T.; Ueda, S. *Chem. Pharm. Bull.* **1989,** *37,* 554–556. doi:10.1248/cpb.37.554
48. Hano, Y.; Nomura, T.; Ueda, S. *J. Chem. Soc., Chem. Commun.* **1990,** 610–613. doi:10.1039/c39900000610
49. Takasugi, M.; Nagao, S.; Ueno, S.; Masamune, T.; Shirata, A.; Takahashi, K. *Chem. Lett.* **1978,** *7,* 1239–1240. doi:10.1246/cl.1978.1239
50. Wang, R.; Chen, R.; Li, J.; Liu, X.; Xie, K.; Chen, D.; Yin, Y.; Tao, X.; Xie, D.; Zou, J.; Yang, L.; Dai, J. *J. Biol. Chem.* **2014,** *289,* 35815–35825. doi:10.1074/jbc.m114.608265
51. de Bruijn, W. J. C.; Levisson, M.; Beekwilder, J.; van Berkel, W. J. H.; Vincken, J.-P. *Trends Biotechnol.* **2020,** *38,* 917–934. doi:10.1016/j.tibtech.2020.02.006
52. Gao, L.; Su, C.; Du, X.; Wang, R.; Chen, S.; Zhou, Y.; Liu, C.; Liu, X.; Tian, R.; Zhang, L.; Xie, K.; Chen, S.; Guo, Q.; Guo, L.; Hano, Y.; Shimazaki, M.; Minami, A.; Oikawa, H.; Huang, N.; Houk, K. N.; Huang, L.; Dai, J.; Lei, X. *Nat. Chem.* **2020,** *12,* 620–628. doi:10.1038/s41557-020-0467-7
53. Li, C.; Huang, Z.; Hu, Y.; Liang, W.; Su, R.; Chen, M.; Zhou, L.; Wu, D.; Gao, G.; You, J. *Org. Lett.* **2021,** *23,* 3746–3750. doi:10.1021/acs.orglett.1c01156
54. Wang, Z.-X.; Gao, P.-C.; Lin, E.-Z.; Li, B.-J. *Angew. Chem., Int. Ed.* **2022,** *61,* e202200075. doi:10.1002/anie.202200075
55. Daniel, B.; Konrad, B.; Toplak, M.; Lahham, M.; Messenlehner, J.; Winkler, A.; Macheroux, P. *Arch. Biochem. Biophys.* **2017,** *632,* 88–103. doi:10.1016/j.abb.2017.06.023
56. Gao, L.; Zou, Y.; Liu, X.; Yang, J.; Du, X.; Wang, J.; Yu, X.; Fan, J.; Jiang, M.; Li, Y.; Houk, K. N.; Lei, X. *Nat. Catal.* **2021,** *4,* 1059–1069. doi:10.1038/s41929-021-00717-8
57. Han, J.; Li, X.; Guan, Y.; Zhao, W.; Wulff, W. D.; Lei, X. *Angew. Chem., Int. Ed.* **2014,** *53,* 9257–9261. doi:10.1002/anie.201404499
58. DeMars, M. D., II; Sheng, F.; Park, S. R.; Lowell, A. N.; Podust, L. M.; Sherman, D. H. *ACS Chem. Biol.* **2016,** *11,* 2642–2654. doi:10.1021/acschembio.6b00479
59. Matsuda, K.; Kobayashi, M.; Kuranaga, T.; Takada, K.; Ikeda, H.; Matsunaga, S.; Wakimoto, T. *Org. Biomol. Chem.* **2019,** *17,* 1058–1061. doi:10.1039/c8ob02867b
60. Matsuda, K.; Zhai, R.; Mori, T.; Kobayashi, M.; Sano, A.; Abe, I.; Wakimoto, T. *Nat. Catal.* **2020,** *3,* 507–515. doi:10.1038/s41929-020-0456-7
61. Kobayashi, M.; Fujita, K.; Matsuda, K.; Wakimoto, T. *J. Am. Chem. Soc.* **2023,** *145,* 3270–3275. doi:10.1021/jacs.2c11082
62. Budimir, Z. L.; Patel, R. S.; Eggly, A.; Evans, C. N.; Rondon-Cordero, H. M.; Adams, J. J.; Das, C.; Parkinson, E. I. *Nat. Chem. Biol.* **2024,** *20,* 120–128. doi:10.1038/s41589-023-01495-z
63. Lacerna, N., II; Cong, Y.; Schmidt, E. W. *ACS Synth. Biol.* **2024,** *13,* 394–401. doi:10.1021/acssynbio.3c00645
64. Chen, K.-Y.; Wang, H.-Q.; Yuan, Y.; Mou, S.-B.; Xiang, Z. *Angew. Chem., Int. Ed.* **2023,** *62,* e202307602. doi:10.1002/anie.202307602
65. Nakamura, H.; Schultz, E. E.; Balskus, E. P. *Nat. Chem. Biol.* **2017,** *13,* 916–921. doi:10.1038/nchembio.2421
66. Schultz, E. E.; Braffman, N. R.; Luescher, M. U.; Hager, H. H.; Balskus, E. P. *Angew. Chem., Int. Ed.* **2019,** *58,* 3151–3155. doi:10.1002/anie.201814016
67. Glasser, N. R.; Cui, D.; Risser, D. D.; Okafor, C. D.; Balskus, E. P. *Nat. Chem.* **2024,** *16,* 173–182. doi:10.1038/s41557-023-01390-z
68. Lowell, A. N.; DeMars, M. D., II; Slocum, S. T.; Yu, F.; Anand, K.; Chemler, J. A.; Korakavi, N.; Priessnitz, J. K.; Park, S. R.; Koch, A. A.; Schultz, P. J.; Sherman, D. H. *J. Am. Chem. Soc.* **2017,** *139,* 7913–7920. doi:10.1021/jacs.7b02875
69. Baltz, R. H.; Seno, E. T. *Antimicrob. Agents Chemother.* **1981,** *20,* 214–225. doi:10.1128/aac.20.2.214
70. Ōmura, S.; Nakagawa, A.; Takeshima, H.; Miyazawa, J.; Kitao, C.; Piriou, F.; Lukacs, G. *Tetrahedron Lett.* **1975,** *16,* 4503–4506. doi:10.1016/s0040-4039(00)91104-3
71. Seno, E. T.; Pieper, R. L.; Huber, F. M. *Antimicrob. Agents Chemother.* **1977,** *11,* 455–461. doi:10.1128/aac.11.3.455
72. Omura, S.; Takeshima, H.; Nakagawa, A.; Miyazawa, J.; Piriou, F.; Lukacs, G. *Biochemistry* **1977,** *16,* 2860–2866. doi:10.1021/bi00632a009
73. Huber, M. L.; Paschal, J. W.; Leeds, J. P.; Kirst, H. A.; Wind, J. A.; Miller, F. D.; Turner, J. R. *Antimicrob. Agents Chemother.* **1990,** *34,* 1535–1541. doi:10.1128/aac.34.8.1535
74. Fishman, S. E.; Cox, K.; Larson, J. L.; Reynolds, P. A.; Seno, E. T.; Yeh, W. K.; Van Frank, R.; Hershberger, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **1987,** *84,* 8248–8252. doi:10.1073/pnas.84.23.8248
75. Cundliffe, E. *Actinomycetologica* **1999,** *13,* 68–75. doi:10.3209/saj.13_68
76. Jung, W. S.; Lee, S. K.; Hong, J. S. J.; Park, S. R.; Jeong, S. J.; Han, A. R.; Sohng, J. K.; Kim, B. G.; Choi, C. Y.; Sherman, D. H.; Yoon, Y. J. *Appl. Microbiol. Biotechnol.* **2006,** *72,* 763–769. doi:10.1007/s00253-006-0318-5
77. Castonguay, R.; Valenzano, C. R.; Chen, A. Y.; Keatinge-Clay, A.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2008,** *130,* 11598–11599. doi:10.1021/ja804453p
78. Sefkow, M.; Neidlein, A.; Sommerfeld, T.; Sternfeld, F.; Maestro, M. A.; Seebach, D. *Liebigs Ann. Chem.* **1994,** 719–729. doi:10.1002/jlac.199419940713
79. Myers, A. G.; Yang, B. H.; Chen, H.; McKinstry, L.; Kopecky, D. J.; Gleason, J. L. *J. Am. Chem. Soc.* **1997,** *119,* 6496–6511. doi:10.1021/ja970402f
80. Aldrich, C. C.; Beck, B. J.; Fecik, R. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2005,** *127,* 8441–8452. doi:10.1021/ja042592h
81. Shirokawa, S.-i.; Shinoyama, M.; Ooi, I.; Hosokawa, S.; Nakazaki, A.; Kobayashi, S. *Org. Lett.* **2007,** *9,* 849–852. doi:10.1021/ol0630191
82. Nicolaou, K. C.; Sun, Y.-P.; Guduru, R.; Banerji, B.; Chen, D. Y.-K. *J. Am. Chem. Soc.* **2008,** *130,* 3633–3644. doi:10.1021/ja710485n
83. Guo, X.; Liu, T.; Valenzano, C. R.; Deng, Z.; Cane, D. E. *J. Am. Chem. Soc.* **2010,** *132,* 14694–14696. doi:10.1021/ja1073432
84. Hatano, K.; Higashide, E.; Shibata, M. *J. Antibiot.* **1976,** *29,* 1163–1170. doi:10.7164/antibiotics.29.1163
85. Kishi, T.; Harada, S.; Yamana, H.; Miyake, A. *J. Antibiot.* **1976,** *29,* 1171–1181. doi:10.7164/antibiotics.29.1171
86. Schmidt, J. J.; Khatri, Y.; Brody, S. I.; Zhu, C.; Pietraszkiewicz, H.; Valeriote, F. A.; Sherman, D. H. *ACS Chem. Biol.* **2020,** *15,* 524–532. doi:10.1021/acschembio.9b00998
87. Kinumaki, A.; Harada, K. I.; Suzuki, T.; Suzuki, M.; Okuda, T. *J. Antibiot.* **1977,** *30,* 450–454. doi:10.7164/antibiotics.30.450
88. Li, S.; Podust, L. M.; Sherman, D. H. *J. Am. Chem. Soc.* **2007,** *129,* 12940–12941. doi:10.1021/ja075842d
89. TAKAHASHI, K.; KUBO, A. *J. Antibiot.* **1977,** *30,* 1015–1018. doi:10.7164/antibiotics.30.1015

90. Scott, J. D.; Williams, R. M. *Chem. Rev.* **2002,** *102,* 1669–1730. doi:10.1021/cr010212u
91. Chrzanowska, M.; Grajewska, A.; Rozwadowska, M. D. *Chem. Rev.* **2016,** *116,* 12369–12465. doi:10.1021/acs.chemrev.6b00315
92. Saito, N. *Chem. Pharm. Bull.* **2021,** *69,* 155–177. doi:10.1248/cpb.c20-00820
93. Kim, A. N.; Ngamnithiporn, A.; Du, E.; Stoltz, B. M. *Chem. Rev.* **2023,** *123,* 9447–9496. doi:10.1021/acs.chemrev.3c00054
94. Li, L.; Deng, W.; Song, J.; Ding, W.; Zhao, Q.-F.; Peng, C.; Song, W.-W.; Tang, G.-L.; Liu, W. *J. Bacteriol.* **2008,** *190,* 251–263. doi:10.1128/jb.00826-07
95. Tang, M.-C.; Fu, C.-Y.; Tang, G.-L. *J. Biol. Chem.* **2012,** *287,* 5112–5121. doi:10.1074/jbc.m111.306316
96. Shin, I.; Davis, I.; Nieves-Merced, K.; Wang, Y.; McHardy, S.; Liu, A. *Chem. Sci.* **2021,** *12,* 3984–3998. doi:10.1039/d0sc06369j
97. Koketsu, K.; Watanabe, K.; Suda, H.; Oguri, H.; Oikawa, H. *Nat. Chem. Biol.* **2010,** *6,* 408–410. doi:10.1038/nchembio.365
98. Koketsu, K.; Minami, A.; Watanabe, K.; Oguri, H.; Oikawa, H. *Curr. Opin. Chem. Biol.* **2012,** *16,* 142–149. doi:10.1016/j.cbpa.2012.02.021
99. Koketsu, K.; Minami, A.; Watanabe, K.; Oguri, H.; Oikawa, H. *Methods Enzymol.* **2012,** *516,* 79–98. doi:10.1016/b978-0-12-394291-3.00026-5
100. Song, L.-Q.; Zhang, Y.-Y.; Pu, J.-Y.; Tang, M.-C.; Peng, C.; Tang, G.-L. *Angew. Chem., Int. Ed.* **2017,** *56,* 9116–9120. doi:10.1002/anie.201704726
101. Zhang, Y.-Y.; Shao, N.; Wen, W.-H.; Tang, G.-L. *Org. Lett.* **2022,** *24,* 127–131. doi:10.1021/acs.orglett.1c03741
102. Tanifuji, R.; Koketsu, K.; Takakura, M.; Asano, R.; Minami, A.; Oikawa, H.; Oguri, H. *J. Am. Chem. Soc.* **2018,** *140,* 10705–10709. doi:10.1021/jacs.8b07161
103. Tanifuji, R.; Oguri, H.; Koketsu, K.; Yoshinaga, Y.; Minami, A.; Oikawa, H. *Tetrahedron Lett.* **2016,** *57,* 623–626. doi:10.1016/j.tetlet.2015.12.110
104. Sato, S.; Sakamoto, T.; Miyazawa, E.; Kikugawa, Y. *Tetrahedron* **2004,** *60,* 7899–7906. doi:10.1016/j.tet.2004.06.045
105. Tanifuji, R.; Minami, A.; Oguri, H.; Oikawa, H. *Nat. Prod. Rep.* **2020,** *37,* 1098–1121. doi:10.1039/c9np00073a
106. Tanifuji, R.; Tsukakoshi, K.; Ikebukuro, K.; Oikawa, H.; Oguri, H. *Bioorg. Med. Chem. Lett.* **2019,** *29,* 1807–1811. doi:10.1016/j.bmcl.2019.05.009
107. Tanifuji, R.; Haraguchi, N.; Oguri, H. *Tetrahedron Chem* **2022,** *1,* 100010. doi:10.1016/j.tchem.2022.100010
108. Leveson-Gower, R. B.; Mayer, C.; Roelfes, G. *Nat. Rev. Chem.* **2019,** *3,* 687–705. doi:10.1038/s41570-019-0143-x


## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.151

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Discovery of antimicrobial peptides clostrisin and cellulosin from Clostridium: insights into their structures, co-localized biosynthetic gene clusters, and antibiotic activity

Moisés Alejandro Alejo Hernandez[1], Katia Pamela Villavicencio Sánchez[1], Rosendo Sánchez Morales[1], Karla Georgina Hernández-Magro Gil[1], David Silverio Moreno-Gutiérrez[1,2], Eddie Guillermo Sanchez-Rueda[2], Yanet Teresa-Cruz[3], Brian Choi[4], Armando Hernández Garcia[2], Alba Romero-Rodríguez[3], Oscar Juárez[5], Siseth Martínez-Caballero[6], Mario Figueroa[7] and Corina-Diana Ceapă*[1]

Full Research Paper

Open Access

Address:
[1]Laboratory of Microbiology, Institute of Chemistry, National Autonomous University of Mexico , Mexico City, 04510, Mexico, [2]Biomolecular Engineering and Bionanotechnology Laboratory, Institute of Chemistry, National Autonomous University of Mexico , Mexico City, 04510, Mexico, [3]Laboratory A-107, Biomedical Research Institute, National Autonomous University of Mexico , Mexico City, 04510, Mexico, [4]Department of Chemical and Biological Engineering, Chemistry and Molecular Biology, Princeton University, Princeton, NJ, USA, [5]Department of Biological Sciences, Illinois Institute of Technology, Chicago, Illinois 60616, [6]Institute of Chemistry, National Autonomous University of Mexico, Mexico City, 04510, Mexico and [7]Facultad de Química, Universidad Nacional Autónoma de México (UNAM), Ciudad de México, 04510, México

Email:
Corina-Diana Ceapă* - corina.ceapa@iquimica.unam.mx

* Corresponding author

Keywords:
antimicrobials; genome mining; lantibiotics; lanthipeptides; multi-drug resistant bacteria; natural products

*Beilstein J. Org. Chem.* **2024,** *20,* 1800–1816.
https://doi.org/10.3762/bjoc.20.159

Received: 04 December 2023
Accepted: 05 July 2024
Published: 30 July 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra



© 2024 Alejo Hernandez et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Antimicrobial resistance presents a substantial threat to global public health, demanding urgent attention and action. This study focuses on lanthipeptides, ribosomally encoded peptides that display significant structural diversity and hold promising potential as antibiotics. Genome mining was employed to locate biosynthetic gene clusters (BGCs) containing class II lanthipeptide synthetases encoded by *lanM* genes. A phylogenetic study analyzing homologous sequences of functional LanM sequences revealed a unique

evolutionary clade of 17 LanM proteins associated with 12 *Clostridium* bacterial genomes. In silico exploration identified nine complete BGCs, including one super-cluster containing two co-localized operons from *Clostridium cellulovorans* 743B, that encode for two new peptides named clostrisin and cellulosin. Each operon was heterologously expressed in *Escherichia coli*. Molecular weights associated with the expected post-translational modifications of the purified lanthipeptide were confirmed by MS–MS/MS analysis for cellulosin, while clostrisin was not post-translationally modified. Both peptides demonstrated antimicrobial activity against multidrug-resistant bacteria, such as a clinical strain of *Staphylococcus epidermidis* MIQ43 and *Pseudomonas aeruginosa* PA14. This is the first report of lanthipeptides from the *Clostridium* genus produced with its native biosynthetic machinery, as well as chemically and biologically characterized. This study showcases the immense potential of genome mining in identifying new RiPP synthetases and associated bioactive peptides.

## Introduction

Antimicrobial resistance (AMR) is a significant public health challenge. Only in 2019, there were 4.95 million deaths associated with AMR [1], a number expected to increase exponentially. One fundamental objective of the Global Action Plan on Antimicrobial Resistance by the World Health Organization (WHO) is the investment in developing new drugs, diagnostic tools, vaccines, and other interventions [2]. In this context, many antibiotics are derived from bacterial natural products (NPs), which have proven to be a valuable source of antimicrobial agents. During the latter part of the 20th century, the discovery of NPs was hindered by the limitation of traditional methods, which often led to the rediscovery of previously identified NPs. Next-generation whole genome sequencing technologies have a newfound ability to explore and identify biosynthetic gene clusters (BGCs) responsible for NPs production. This renewed focus offers the advantage of preventing redundant discovery and predicting novelty, resistance, and bioactivities. Furthermore, the increasing availability of genomic data has led to the development of bioinformatics tools, such as AntiSMASH [3], Bagel4 [4], and RiPPMiner [5], that have emerged to streamline the process of exploring and discovering BGCs in bacteria, known as genome mining.

Genome mining has emerged as a crucial research area in discovering novel antimicrobial compounds [6,7]. Due to its unique position between basic and applied research, it has become an essential tool for identifying compounds with potential therapeutic applications. Genetic and genomic data mining enables the study of the evolution of genes and genomes across diverse species and populations. This provides insights into the origins and evolution of genes that have evolved as "tools" in the ongoing biological battles in the microbial world. The biodiversity of specialized metabolites is strain-specific, meaning that even closely related organisms can have distinct metabolic capabilities. Consequently, researchers can unveil new clusters and metabolites by tracking the conserved genes participating in biosynthetic processes.

In recent years, notable achievements have been seen in genome mining in discovering antimicrobial compounds. For example, the discovery of teixobactin [8] has paved the way for developing antibiotics with innovative mechanisms of action. Similarly, the discovery of halicin, a molecule identified through a machine learning-based approach, has shown promise as a broad-spectrum antimicrobial agent including some resistant to existing antibiotics [9]. Thus, peptides, including ribosomally synthesized and post-translationally modified peptides (RiPPs), have been regarded as important sources of antibiotics, both historically and through recent discoveries [10].

Lanthipeptides are a class of RiPPs. Lanthipeptide biosynthetic genes have undergone complex evolutionary processes that have produced chemically diverse active peptides [11,12]. These genes are predominantly found in bacteria and have evolved through selective pressures driven by competition for resources and defense against predators [13]. The resulting peptides exhibit unique structural features due to thioether bridges between dehydrated serine or threonine (Dha/Dhb) and cysteine (Cys). These thioether bridges lead to cyclic peptides with the modified amino acid lanthionine (Dha-Cys) or methyllanthionine (Dhb-Cys). The strong antimicrobial activity of nisin, a lanthipeptide produced by certain strains of *Lactococcus lactis,* is dependent on thioether-ring formation [14]. Additional post-translational modifications contributing to their distinct biological activities were identified [15,16]. Intermolecular disulfide bridges were proved to spontaneously form in bovicin HJ50 to exert their natural activity and change spectra of antimicrobial effects when synthetically intramolecularly bound [17,18].

The biosynthesis of mature lanthipeptides involves the translation of *lanA*, which encodes a precursor peptide with two regions: a leader peptide that is recognized by post-translational enzymes and a core peptide where these enzymes produce the formation of Dha-Cys or Dhb-Cys residues. The

post-translational modification enzymes involved span different functional domains (a cyclase and a dehydratase), and their activity is tightly controlled at various levels to ensure the production of high-quality peptides [19]. As an example, subtilin production is autoregulated based on the extracellular concentration of the molecule. At the same time, subtilin expression is also controlled by the spaRK operon and the sigmaH ($\sigma^H$) global regulator, which leads to its production during mid-exponential and log-phase [19].

To date, five classes of lanthipeptide synthetases have been identified [20] and are the basis for classifying lanthipeptide gene clusters. For the class II lanthipeptide synthetases discussed in this work, multidomain LanM enzymes catalyze the formation of Dha-Cys or Dhb-Cys [21]. Moreover, *lanPt* encodes a membrane protein with two ABC transporter domains and a C39 peptidase domain. The protein cleaves the leader peptide to generate mature lanthipeptides and exports them to the extracellular environment [22].

The diversity of lanthipeptides produced by these biosynthetic pathways has led to the discovery of many new compounds with potential therapeutic applications during recent years, along with a deeper understanding of their evolution, ecological importance, and gene composition [23]. Lanthipeptides produced by class I and II lanthipeptide synthetases exhibit antimicrobial activity against various pathogens, including drug-resistant strains [23,24]. These molecules are relevant as they have strong antibiotic effects against pathogenic bacteria such as *Staphylococcus aureus*, *Enterococcus sp.*, *Clostridioides difficile*, and *Mycobacterium tuberculosis*. The effects of some of these peptides are attributed to their affinity for the lipid II component of Gram-positive bacterial cell walls [25]. Additionally, there have been reports of lantibiotics such as CMB001 displaying activity against resistant Gram-negative bacteria, including *Acinetobacter baumannii* [26], as supported by the current study. In addition, some lanthipeptides have been shown to have anticancer and immunomodulatory properties [11,27].

A low development of resistance to lanthipeptides has been observed, with nisin being the most well-known case in the food industry. While some strains exhibit resistance due to changes in the cell wall, biofilm formation, or the expression of resistance proteins such as ABC transporters or proteases [28], specific mutations in nisin have rendered previously resistant strains susceptible [29]. The structural diversity of these peptides, coupled with their successful production in *E. coli* [30], has driven research to consider them a source for new antibiotics, as supported by the present study. Identifying the genetic and biochemical mechanisms underlying the production of compounds is of utmost importance in developing effective antibiotics against drug-resistant pathogens. Although certain RiPPs with activity against *Clostridia* have been described [31], the characterization of complete gene clusters encoding for lantibiotics, using their native biosynthetic machinery and not that of nisin or of other non-native peptides, as well as the chemical characterization of the NPs that are produced and the modifications they suffer, from a *Clostridium* genus, have yet to be accomplished. This is due to the difficulty in isolating and maintaining such strains in the laboratory and the limited availability of genome sequences compared to other phylogenetic groups, with published sequences primarily focusing on pathogenic strains [32-34]. Despite these challenges, the study of environmental *Clostridium* has already yielded several interesting NPs, including polythioamides such as closthioamide, a new chemical class from *Clostridium cellulolyticum*, and clostrubin, a polyphenolic polyketide antibiotic and the first reported polyketide from anaerobic bacteria [35-37]. Therefore, the discovery of a lantibiotic with a *Clostridium* origin, reported in this study, has significant implications for developing novel antibiotics. The unique chemical structures of these lantibiotics make them promising candidates for treating drug-resistant pathogens, and the characterization of these compounds from *Clostridium* provides an opportunity to develop new antibiotics.

## Results and Discussion

### Genome mining of LanM enzymes' sequence diversity to localize novel evolutive clades

This study used 28 amino acid sequences of LanM enzymes from MIBiG [38] whose activity has been experimentally proven [24] to identify homologous sequences using BLASTP [39] (in BV-BRC [40]). They are considered our reference sequences. Two of these reference sequences are part of the biosynthetic machinery for the production of ruminococcin A [41] and Flv peptides [42] from the clostridial class. As a result, 315 homologous proteins were identified and included in the comparative phylogenetic study. A phylogenetic tree was built (Figure 1) with the reference LanM enzymes amino acid sequences from MIBiG and the homologs identified in our study. The taxonomy of the bacteria to which they belong is indicated in the figure. The selected superclade highlighted in Figure 1, containing sequences that mainly belong to Bacilli and *Clostridia*, was of particular interest because of the novelty that lanthipeptides from *Clostridium* could constitute, since no such BGCs with this taxonomic origin had been described before. The tree also indicates the presence of other yet unexplored LanM clades, such as from Bacilli, Cyanobacteria, and Actinobacteria, which could be interesting to study further due to their sequence uniqueness.

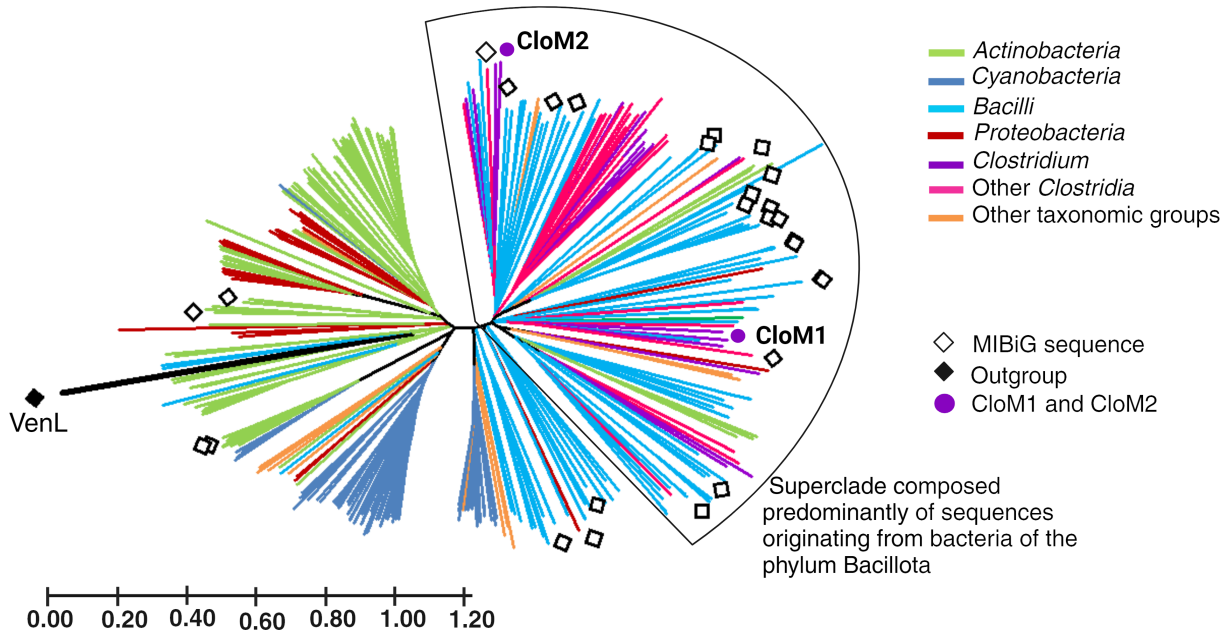



**Figure 1:** Phylogenetic trees of the LanM synthetase amino acid sequences. Unrooted phylogenetic tree of all the homologous LanM sequences identified by BLAST in our study and those in MIBiG (previously characterized). This analysis involved 343 amino acid sequences; there were a total of 1964 positions in the final dataset. The phylogenetic tree was built from the alignments using the Neighbor-Joining [43], and the evolutionary distances were calculated using the Poisson correction method. A LanL class IV lanthionine synthetase from the venezuelin cluster [44] was used as an outgroup. Square symbols refer to previously characterized LanM sequences from the MIBiG database.

The sequence for venezuelin class IV lanthionine synthetase (VenL – square black symbol) was used as an outgroup, due to the common presence of a LanM domain. The phylogenetic distance confirms the usefulness of the outgroup, confirming a higher similarity within the sequences resulting from the BLAST search.

## Identification and selection of CloA1 and CloA2 BGCs

For *Clostridia*, there were 37 predicted BGCs (15 of these with at least one precursor and one biosynthetic gene, Table S2 in Supporting Information File 1), with 76 precursor peptides associated. Of these, 2 LanM enzymes and 5 precursor peptides have previously been studied. We decided to focus on the results of the *Clostridium* genus, as there are no described lanthipeptides or associated gene clusters from this taxonomic group. The homology search for novel LanM biosynthetic enzymes from the *Clostridium* genus encountered 17 protein sequences associated with 12 genomes. After analyzing these genomes with AntiSMASH [3], we identified 150 different BGCs of specialized metabolites from different biosynthetic classes. The RiPPs accounted for 36% of these (54 clusters). A total of 14 BGCs containing homologs for class II lanthipeptide synthetases were identified, which in turn helped determine the presence of precursor peptides, gene duplications, paralogues, transporters, and resistance genes in these clusters. Three of these clusters contain more than one biosynthetic gene and one or several precursors, forming superclusters. The RiPPMiner platform [5] was used to predict in silico post-translational modifications of the mature lanthipeptides. Based on these predictions, several clusters did not contain precursor peptides with a predicted lanthipeptide heterocycle formation and were discarded. As compared to reference sequences from MIBiG, no known amino acid precursor sequences similar to previously characterized ones were detected. From these BGCs, 9 precursor peptides (out of a total of 32 precursors) were considered noteworthy of further investigation, based on their physicochemical characteristics, number of predicted cycles (one cycle or more was required), the presence of serine (Ser) and threonine (Thr) residues, and the characteristic cleavage site of the C39 peptidases domain (GG or GA) (Table S2 in Supporting Information File 1) obtained in RiPPMiner [5]. The nine biosynthetic clusters or superclusters these precursor peptides belonged to contained all the genes *lanA, lanM*, and *lanT*. These genes were found to be the minimal machinery required for lanthipeptide biosynthesis (Figure 2). Notably, transporter genes and regulatory elements were not considered in the second filter as their presence is not needed for biosynthesis, albeit it is for resistance. As a result, the *C. cellulovorans* 743 clusters were chosen because they were thought to be complete, novel, and

most interesting, primarily due to the presence of two different, apparently independent operons, each with its lanthipeptide precursor (Figure 2). A cluster analysis using BAGEL4 [4] confirmed the presence of terminators. The sequence information for CloA1 (precursor peptide), CloM1 (biosynthetic enzyme), CloPt1 (peptide domain), CloA2 (precursor peptide), CloM2 (biosynthetic enzyme), and CloPt2 (peptide domain) and the associated genes are reported in Table S1 (Supporting Information File 1). The mature peptides formed after the enzymatic removal of the leader peptide by the CloPt domain are called clostrisin (for CloA1) and cellulosin (for CloA2).

To confirm the novelty of the lanthipeptides from the cluster of *C. cellulovorans* 743, clostrisin and cellulosin were compared with known lanthipeptides, generating a similarity network made using the percent identity between the precursor peptides from all experimentally characterized lanthipeptides to date (Figure 3A). We found it noteworthy to examine the two peptides due to their relationship to known lanthipeptides. Despite being part of a diverse cluster of characterized lanthipeptides, these peptides maintain enough sequence divergence, having an identity percentage of no more than 60%. One of these peptides, clostrisin, occupies a hinge position between two divergent lanthipeptide clusters – one represented by lichenicidin and Flv peptides. The other peptide, cellulosin, is located next to belongs to the lichenicidin family. However, it has undergone a significant number of functional mutations. Studying these mutations can provide added data on function–structure relationships and how they impact biological activity. Since we have not found evidence of horizontal transfer of the *cloA1* and *cloA2* gene clusters, it is an interesting speculation that they could constitute the origin of this family and that, by duplicating one of these clusters, *C. cellulovurans* has directed their evolution to serve completely different biological activities.

The amino acid sequence of the precursors shares some similarity with the characterized peptides: for clostrisin with FlvA.2g [42], enterocin Wβ [45], plantaricin Wβ [46], and thusin α [47], and cellulosin with lichenicidin A1 [48], FlvA.1 [42], FlvA.2 [42], BhtA α [49], SmB β [50], Smb α [50], and thusin β [47] (Figure 3B). We find that, especially in the first 20 amino acids of the alignment or the first half of the mature peptide, there are significant sequence differences. Of particular interest, cellulosin forms two larger heterocycles (of seven and nine amino acids) in this sequence region, whilst most lanthipeptides only form small ones (up to five) or none at all. As for the second part of the alignment, where most lanthipeptides form various heterocycles, cellulosin shares a very similar profile to liquenicidin VK21, and as the experimental results confirm, despite the in silico predictions, clostrisin does not form heterocycles. These differences in primary and secondary structure could be responsible for the divergent biological activities we encountered.

## In silico characterization of the post-translational modification enzymes

The selected supercluster is predicted to be formed by two adjacent, biosynthetically complete transcriptional units, each with specific promoter and terminator sites (Figure 2) containing two LanM enzymes (CloM1 and CloM2), the precursor peptides CloA1 and CloA2, as well as two transporter protein peptidases, which we named CloT1 and CloT2 (Figure 2). The amino acid

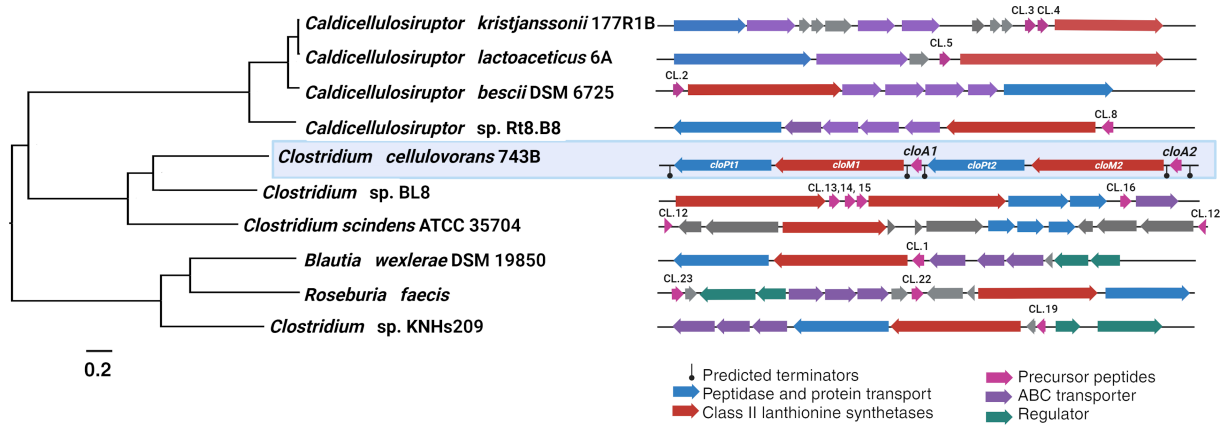



**Figure 2:** The phylogenetic tree was built by concatenating 1000 shared clostridial genes (left) between the studied genomes, and the lanthipeptide class II synthetase clusters from the selected BGCs of this taxonomic clade are presented (right). The highlighted cluster was chosen for heterologous expression.

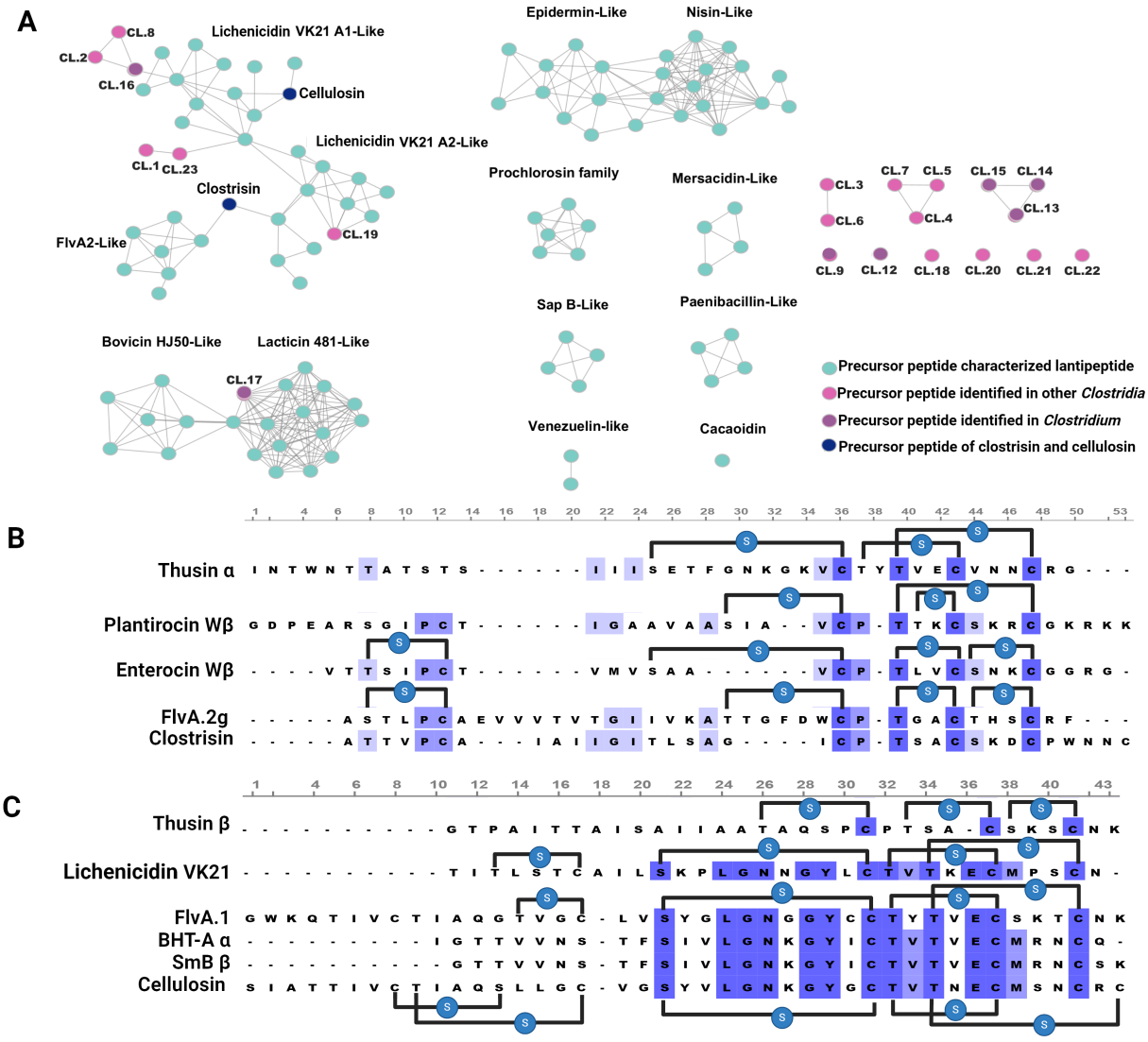



**Figure 3:** A. Similarity network created with the ESI web tool with the precursor peptide amino acid sequences. In the similarity network of precursor peptides, each blue node represents a characterized lanthipeptide, and each pink node represents the precursor peptides identified in this work. Each line represents a minimum of 27% of sequence identity (EFI – enzyme similarity tool [51]). B. Alignment of clostrisin with FlvA.2g [43], enterocin Wβ [44], plantaricin Wβ [45], and thusin α [46]. C. Alignment of cellulosin with lichenicidin A1 [47], FlvA.1 [43], FlvA.2 [43], BhtA α [48], SmB β [49], Smb α [49], and thusin β [46]. The lanthionine bridges placed on clostrisin and cellulosin are based on the experimental results of the $MS^2$ analyses.

sequences of the C39 peptidase domain of CloT1 and CloT2, further identified as CloPt1 and CloPt2, were subjected to BLAST analysis, revealing a strong primary sequence conservation with all homologous sequences from other clostridial species. A phylogenetic tree was made based on this sequence, showing various organisms within the same class (as detailed in Figure S5A in Supporting Information File 1). Additionally, structure models for CloPt1 and CloPt2 were generated using AlphaFold 2.0 [52] (Figure S5B and S5C in Supporting Information File 1) and compared to the protein PCAT1 [22] (the closest homologous protein with a structure resolved by crystallography, here presented in complex with its peptide ligand (PDB 6V9Z). The amino acid sequence identity between CloPt1 and PCAT1 was 29%, while for CloPt2, it was 24%. Structural alignment revealed RMSD values below 3 Å, meeting the minimum criteria for structural conservation (Table S3, Supporting Information File 1). Furthermore, the catalytic residues within the peptidase domains were confirmed. CloPt1 catalytic residues are Cys35 and His111; for CloPt2, they are Cys18 and His 92. These residues maintained a distance and structural positions like the catalytic residues in PCAT1 (where the catalytic residues are Cys21 and His99).

A similar procedure was carried out for CloM1 and CloM2, leading to the identification of the closest sequences within the clostridial class (Figure S6 in Supporting Information File 1). Structure models for CloM1 and CloM2 generated using AlphaFold 2.0 (Figure S6B and S6C, Supporting Information File 1) were compared with the CylM [53] protein (PDB 5DZT), as it remains the sole LanM enzyme characterized through crystallography to date. The amino acid sequence identity between CloM1 and CylM was 23%, while for CloM2, it was 24%. Structural alignment was performed, and the obtained RMSD values were below 3 Å (see Supporting Information File 1, Table S4). Our findings suggested that these clusters and their associated biosynthetic enzymes hold promise as novel entities with untapped bioactive potential.

## Expression and purification of the CloA1 precursor peptide, CloA2 precursor peptide, and C39 peptidase domain

We could not access the producing microorganisms, so the gene sequences reported in public databases (Supporting Information File 1, Table S1) were synthesized de novo, cloned, and checked by sequencing using a service (GenScript). First, the clostridial nucleotide sequences were codon adjusted for expression in *E. coli* using the GenScript tools, cloned in the pRSF-Duet vector in the MCS1 using the restriction enzymes BamHI and SpeI, and verified by sequencing (GenScript). To express the biosynthetic machinery for the two lanthipeptides, we transformed the *E. coli* strain NiCo21(DE3) (NEB) with the following vectors: pRSF-Duet_CloA1_CloM1, pRSF-Duet_CloA2_CloM2, and pRSF-Duet_CloPt1 (containing the C39 peptidase domain from the clostrisin cluster). The latter contains the peptidase domain of CloPt1, as the complete ABC transporter was deemed too large to synthesize and express in this system. This strategy was functional in previous studies and retained catalytic activity without the rest of the transporter domains [49]. As in the previous literature report, we could confirm that the precursor CloA2 could be cut by the non-native peptidase CloPt1, likely due to this domain's sequence and structure conservation.

SDS-PAGE gel electrophoresis confirmed the expression of all proteins. Each of the LanM enzymes was co-expressed with the corresponding precursor peptide in the same *E. coli* strain (expected size 150 kDa). The precursor peptides and the peptidase domain of CloPt1 were fused with a 6xHis tag at their N-terminal end for purification purposes (Figure S7 in Supporting Information File 1). After purification, we estimated mature peptide and peptidase yields based on SDS-PAGE densitometry between 0.8 and 1.5 g/L of culture for all products. The heat liability of the peptide products challenged the purification, and when trying to perform several purification cycles to increase their purity, we obtained only meager amounts of the peptides. This issue greatly limited the quantity and purity of peptides available for the biological assays.

## The activity of the C39 peptidase domain cleaving the leader peptide of CloA1 and CloA2 precursor peptides

The proteolytic activity was monitored by mixing the purified samples of the C39 peptidase domain with purified CloA1 precursor peptide and CloA2 precursor peptide. The SDS-PAGE electrophoretic pattern confirmed the proteolytic reaction (Figure S7D, Supporting Information File 1). This was further confirmed through HPLC–MS–MS/MS analysis.

### Mass spectrometric analysis of the CloA1 and CloA2 precursor peptides

Clostrisin and cellulosin were purified through non-native means using a $Ni^{2+}$ column due to a 6xHis tag at the N-terminal end, followed by elution via imidazole gradient (refer to Figures S7A and S7B in Supporting Information File 1 for SDS-PAGE analysis). The fraction containing the purest products underwent analysis via HPLC–MS–MS/MS experiments. Spectra of the cellulosin sample revealed a pattern of multicharged ions, with the most intense peak at an *m/z* of 1,140.64, consistent with the calculated average mass *m/z* associated with the CloA2 precursor peptide ion six times dehydrated $[M - 6H_2O + 8H]^{8+}$ (Figure 4A1 and Table S6 in Supporting Information File 1). Detected $MS^2$ ions for the cellulosin experiments are presented in Tables S6–S9 (Supporting Information File 1). Additionally, an ion with an *m/z* of 1,142.89 was identified, corresponding to the cellulosin precursor peptide five times dehydrated average mass (Figure 4A2 and Table S6 in Supporting Information File 1). For celullosin, a pattern of multicharged ions was observed, with the most intense peak at an *m/z* of 1,068.99, associated with an octuple-charged ion $[M + 8H]^{8+}$ linked to an unmodified precursor peptide average mass (Figure 4D, Figure S4 and Table S9 in Supporting Information File 1). Detected $MS^2$ ions for the clostrisin experiments are presented in Tables S10–S13.

Subsequently, a trypsin digestion assay was conducted on the purified cellulosin sample, revealing a pattern of multicharged ions with the most intense peak at *m/z* 1,333.71 associated with the fragment N23 to C64 $[M - 6H_2O + 5H]^{5+}$, in turn, associated with six dehydrations (Figure 4B1 and Table S7 in Supporting Information File 1). Similarly, an ion at *m/z* 1,337.31 $[M - 5H_2O + 5H]^{5+}$ associated with the same fragment five times dehydrated was identified. Cyclization reactions introduced thioether cross-links between dehydrated Thr/Ser residues and specific cysteine thiol groups. However, due to identical masses of dehydrated linear peptides (lacking thioether

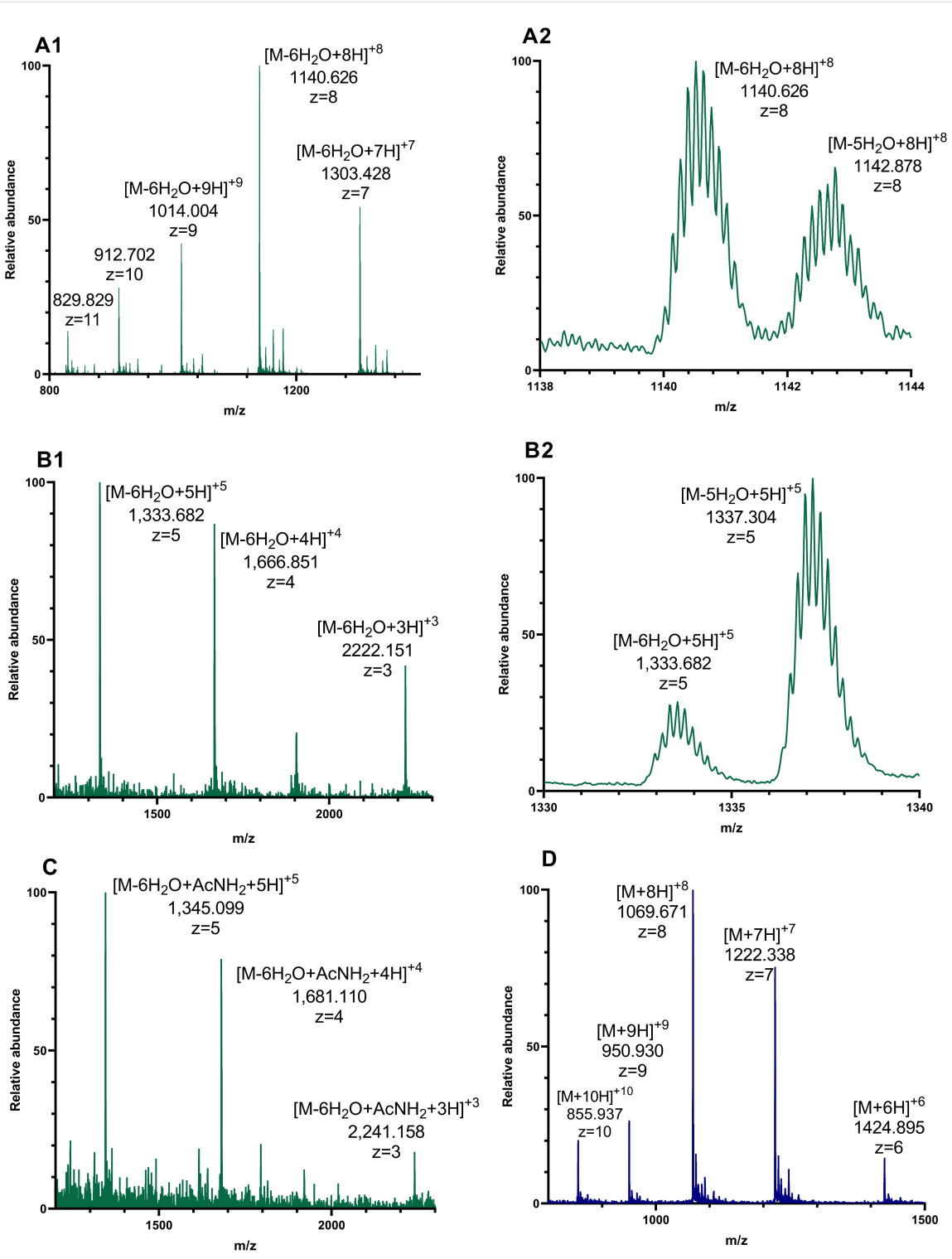



**Figure 4:** Mass-spectrometric analysis of purified clostrisin and cellulosin (ESIMS spectra). A1 and A2: CloA2, the precursor peptide of cellulosin with a 6xHis tag, was analyzed. A1 shows the ion pattern of CloA2. A2 displays the specific ions with 8 charges of CloA2. GSSHHHHHHSQDPNSSSARLQKNYEELFNEVNENASLQAELNGGSIATTIVCTIAQSLLGCVGSYVLGNKGYGCTVTNECMSNCRC, average mass: 9225.12 Da – $6H_2O$ 9147.12 Da. B1 and B2: The tryptic fragment from CloA2 was analyzed. B1 shows the ion pattern of the tryptic fragment. B2 displays the specific ions with 8 charges of the tryptic fragment. C shows the ion pattern of the tryptic fragment after iodoacetamide derivatization NYEELFNEVNENASLQAELNGGSIATTIVCTIAQSLLGCVGSYVLGNKGYGCTVTNECMSNCRC. Average mass: 6771.58 Da – $6H_2O$ 663.58 Da + IAA 6720.58 Da. D. CloA1, precursor peptide of clostrisin with 6xHis, was analyzed: GSSHHHHHHSQDPNSSSARLQQNYESKAGFISEMELDELVSNKTVGGATTVPCAIAIIGITLSAGICPTSACSKDCPWNNC. Isotopic mass: 5843.88 Da.

rings) and cyclic forms (containing Dhb and Dha rings), a derivatization assay with iodoacetamide followed by trypsin digestion was performed for cellulosin, revealing a pattern of multicharged ions with the most intense peak at $[M - 6H_2O^+AcNH_2 + 5H]^{5+}$ with an *m/z* of 1,345.16 associated with the fragment from N23 to C64, six times dehydrated, and exhibiting an increase of 57u due to cysteine alkylation (Figure 4C and Table S8 in Supporting Information File 1).

These analyses confirm the activity of the CloM2 enzyme, with the dehydratase domain activity forming a mixture of six and five times dehydrated lantipeptides, and the cyclase domain activity demonstrated by the iodoacetamide derivatization assay, where 5 out of 6 cysteines are involved in the formation of Dha and Dhb along the core peptide, preventing trypsin digestion.

We were unable to detect any modifications of the CloA1 precursor peptide by the CloM1 biosynthetic enzyme, despite previous in silico predictions suggesting that this should be a functional system. This lack of modification could be due to the amino acid sequence of the precursor, which differs from those described in the literature, or to mutations in the CloM1 enzyme. These mutations may require additional co-factors or specific biochemical conditions that we have not yet identified. At this point, based on the information we have gathered, we have been unable to explain the lack of post-translational modification. Although we attempted experiments to mix the CloM2 enzyme with the CloA1 precursor peptide and to mix the CloM1 enzyme with the CloA2 precursor, neither precursor underwent modifications, indicating a certain degree of peptide specificity in the biosynthetic enzymes of this gene supercluster.

## HPLC–MS–MS/MS experiment on the CloM2-modified CloA2 tryptic fragment

HPLC–MS–MS/MS experiments were performed on the CloM2-modified cellulosin tryptic fragment, featuring 5 and 6 net dehydrations, followed by the analysis of the sixth dehydrated fragment post-iodoacetamide derivatization to detect possible cross-links and dehydrated residues (Figure 4 and Supporting Information File 1, Figure S1, associated with ion lists in Tables S6–S9). These experiments provided crucial insights into the cellulosin lanthipeptide structure. High-confidence conclusions from the $MS^2$ experiment on the fragment with 6 net dehydrations highlighted the linearity of N1–V29. They identified T27 as Dhb, revealing macrocyclization of C30–C39 and suggesting Dhb and Dha residues at T31 and S35, respectively, with potential macrocyclization of S42 forming a lanthionine bridge. Additionally, medium-confidence observations hinted at macrocyclization of C64 (Figure 4 and Supporting Information File 1, Figure S2 associated with ion lists in Tables S6–S9). Furthermore, analysis of the $MS^2$ experiment post-iodoacetamide reaction indicated the absence of free cysteines between C30 and C39, supported by alkylated y35 and y24 ions, implying macrocyclization, and proposed the presence of three macrocycles between G41 and the C-terminus (Figure S3 in Supporting Information File 1). Moreover, findings from the $MS^2$ experiment on the fragment with 5 net dehydrations suggested one internal tryptic cleavage between G41 and the C-terminus, with K48 as the likely cleavage site if C64 is macrocyclized (Figure S3, Supporting Information File 1).

## Antimicrobial activity of clostrisin and cellulosin

Antimicrobial activity assays of clostrisin and cellulosin were performed using microplate bioassays with strains from the American type culture collection (ATCC): *S. aureus* 43300 (MRSA), *P. aeruginosa* ATCC 15442 (PA14), *E. coli* IM08B, and *A. baumannii* ATCC BAA 747 as well as clinical strains isolated from patients in Mexico: *S. epidermidis* MIQ43 (multidrug-resistant clinical sample) (internal code from the MicroIQ laboratory library), and *P. aeruginosa* MIQPA25 [54] (multidrug-resistant clinical sample isolated from cystic fibrosis patients), and *C. difficile* R20291. All strains have a multidrug-resistant profile (Table S5, Supporting Information File 1). These bacteria were chosen due to their classification in the ESKAPE group and the species of origin (in the case of clostridioides).

We tested the clostrisin and cellulosin alongside a mixture with the C39 peptidase domain. The approximate concentrations for the mature peptides were roughly estimated based on the SDS-PAGE using densitometry. Each assay was performed at the highest obtainable concentrations: 1.4, 2.8, and 5.6 µg/mL of clostrisin and 1.2, 2.4, and 4.8 µg/mL of cellulosin (Figure 5). We used the same concentrations of precursor peptides in the mixtures of CloA1 and CloA2. In our testing against *C. difficile*, we evaluated clostrisin at 5.6 µg/mL and cellulosin at 4.8 µg/mL. Additionally, to ensure that the observed activity was not due to the C39 peptidase domain protein, we tested these purified proteins on each bacterium at 10.5 µg/mL, where they showed no inhibition.

The peptides exhibited no antibiotic activity against *E. coli* ATCC IM08B, *C. difficile* R20291, *A. baumannii* ATCC BAA 747, and MRSA at the concentrations employed. It was expected that the lack of effect on *E. coli* would occur, as the prepeptide accumulation did not impact the expression host. This assumption was made due to reports of auto-proteolysis for other lanthipeptides during production in this cell system. This leads us to believe that at least a few mature peptides might have been present. Despite this, the growth curves of the pro-

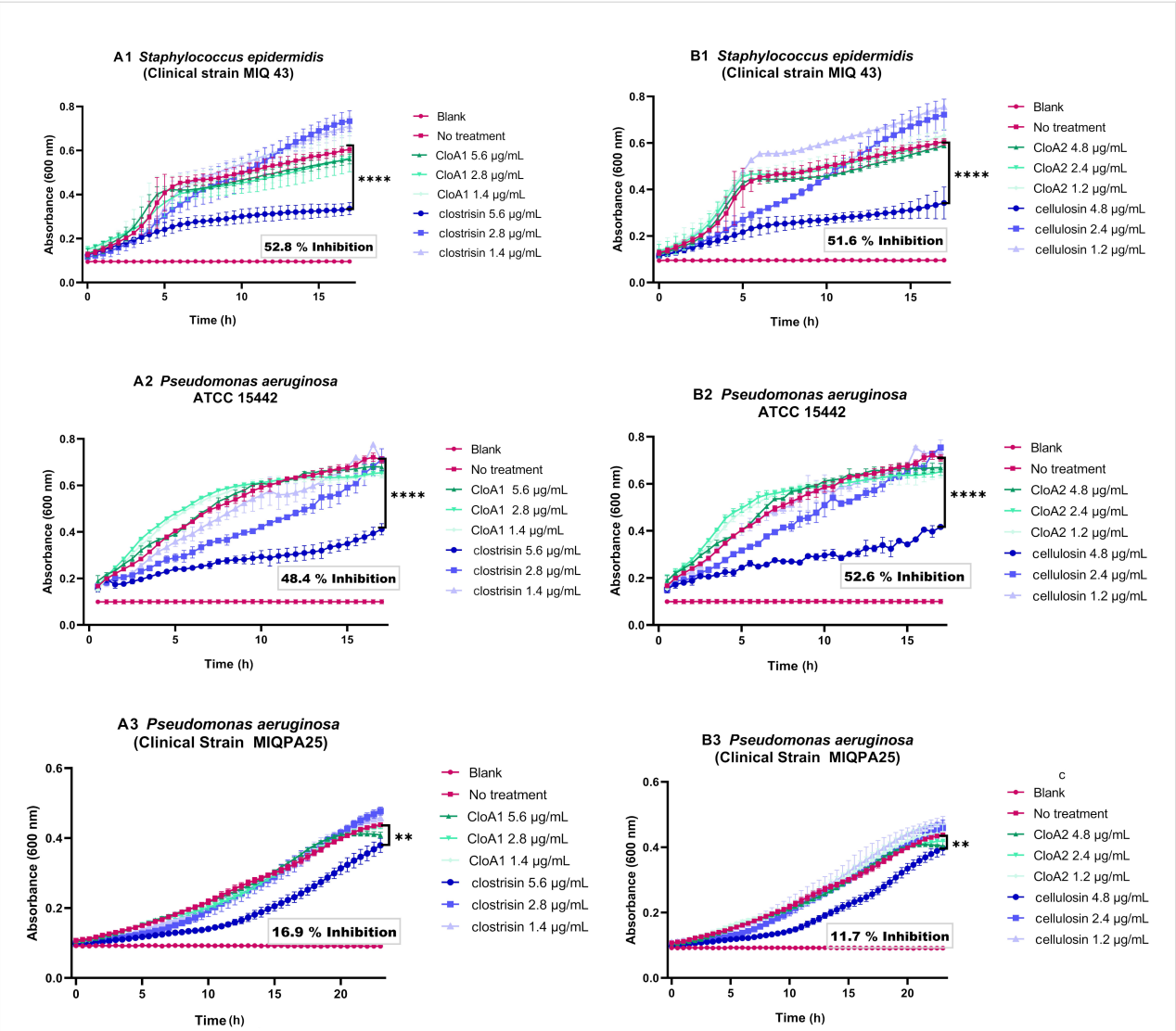



**Figure 5:** Growth curves of the strains with the bacterial activity of the samples. A. Precursor peptide for CloA1 and clostrisin and B. precursor peptide for CloA2 and cellulosin with the bacteria: 1. *S. epidermidis* MIQ43 and the 2. *P. aeruginosa* ATCC PA14 and 3. *P. aeruginosa* MIQPA25. Tests were performed in sterile polystyrene 96-well flat-bottom-shaped microtiter plates. Inoculum densities were adjusted at an $OD_{600nm}$ of 0.1. The optical density (OD) was measured at 600 nm after incubating the plates at 36 °C for 16 hours at 220 rpm. Three independent experiments were performed, each with three repetitions of every peptide concentration. The statistical significance was calculated with a t-test in GraphPad 8.0.2.

ducing strains were not distinct from the control ones, transformed with the empty vector. As a result, it was not surprising that there was no effect on the *E. coli* test strains, which confirmed previous observations.

For *S. epidermidis* MIQ43, *P. aeruginosa* ATCC PA14, and MIQPA25, both clostrisin and cellulosin samples displayed robust bacteriostatic activity at the highest estimated concentration of 5.6 µg/mL and 4.8 µg/mL, respectively, compared to the bacteria grown in LB media (control), and lower but statistically significant effects at lower concentrations (Figure 5). The immature peptides CloA1 precursor peptide and CloA2 precursor peptide showed no discernible effect at the same concentration. These findings highlight a change in the activity after the leader peptide's proteolysis and the presence of mature lanthipeptides (Figure 5). For clostrisin, as it has no post-translational modifications, we expect that the removal of the leader peptide can be the cause for the increase in its biological activity.

Finally, it was extremely interesting to test the lanthipeptides on *C. difficile,* due to the origin of these peptides in the clostridial clade. The strict ecological environments of *Clostridia* are the human and animal microbiota and soil for both anaerobic and

aerotolerant organisms. We interpret a lack of effect on *C. difficile,* an intestinal clinical isolate, due to either the presence of resistance mechanisms or the fact that these peptides evolved to be active against soil microbiota members. The habitat of *C. cellulovurans*, the strain of origin of the lanthipeptide clusters, appears to be plant-associated due to its metabolism and isolation environment (wood).

Our results raise numerous questions regarding the ecological functions of the lantibiotics we discovered. Our bioinformatics studies have not provided any clues about the resistance mechanisms of these peptides. Therefore, we plan to conduct further research to gain insights into these functions. This could help us to understand the potential for these peptides to display antimicrobial resistance in the future.

## Atomic force microscopy

The study employed atomic force microscopy (AFM) in air to observe the bacterial morphological changes on the surface of *S. epidermidis* triggered by the impact of lanthipeptides. The concentrations of peptides were estimated based on the SDS gel band intensity, and we expect there is some uncertainty in these estimations. The samples were treated with an estimated 22.4 µg/mL concentration for 1 and 5 hours. The AFM images of untreated *S. epidermidis* MIQ43 displayed characteristic clustered cocci (see Figure 6A and 6F). The treatment with clostrisin and cellulosin samples resulted in more pronounced morphological changes after 5 hours of treatment. The changes included the loss of the characteristic cocci structure, with the formation of blebs that increased the membrane's rugosity. Besides, cytosolic leakage and remnants of the outer membrane were identified (see Figure 6B, C, G, and H). The treatment with the mature peptides led to numerous blebs visible after 1 hour. Interestingly, no cocci structures were detected after 5 hours of treatment (see Figure 6). The precursor peptides, especially cellulosin, appear to elicit some effects on the membrane of *S. epidermidis*. The increase in peptide concentration can explain the increase in the cellular damage compared with the microplate assay, effects that could not be detected at lower concentrations. Also, the limitation we faced is that, since we were working with only partially purified peptides, our estimates for the concentrations could have a significant error range.

Based on the observations, it can be inferred that both clostrisin and cellulosin directly interact with the membrane of *S. epidermidis*. However, no significant differences were found in the samples regarding using the peptidase C39 peptidase domain. The findings of this study provide valuable insights into the effects of lanthipeptides on *S. epidermidis*, and confirm the previous results acquired during the antimicrobial activity assays. Further studies are needed to confirm the mechanism of action of these lanthipeptides and their molecular targets in *S. epidermidis,* a Gram-positive bacterium.

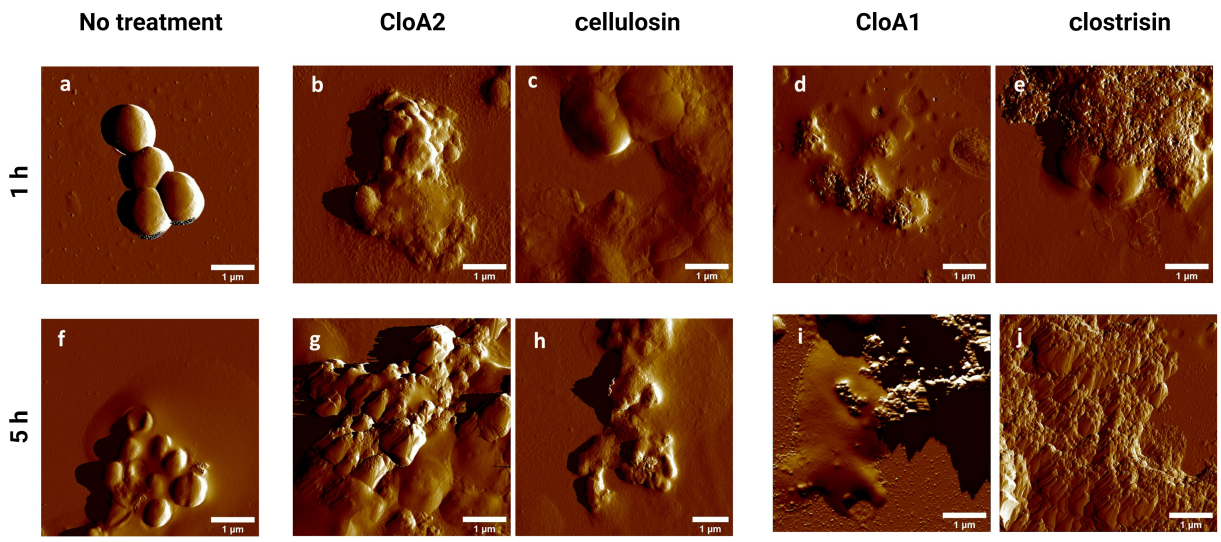



**Figure 6:** Atomic force microscopy images (peak force mode) of *S. epidermidis* MIQ43 incubated with the different samples for 1 hour and 5 hours. From left to right: no treatment, CloA2 precursor peptide, cellulosin, CloA1 precursor peptide, clostrisin. All samples were evaluated at an estimated concentration of 22.4 ng/µL. Scale bar: 1 µm.

## Conclusion

The pressing need for new antibiotics has made the search for them a top priority amongst the scientific community. In this study, we utilized genomic mining to forecast the uniqueness of two lanthipeptides derived from the *Clostridium* genus. These peptides were expressed in an *E. coli* host and produced in sufficient concentrations to perform bioactivity assays. The C39-associated domain was purified to facilitate the transformation from pre-peptide to the mature configuration. The mature clostrisin and cellulosin exhibited antimicrobial properties against *S. epidermidis* and *P. aeruginosa* resistant strains and failed to exhibit the same effects against *E. coli* and *C. difficile*. As a result, we present the discovery of two new lantibiotics that have demonstrated efficacy in combating both Gram-positive and Gram-negative pathogens.

## Methodology

### Bioinformatic analyses

#### Lanthipeptides identification

For the search of lanthipeptides, a LanM class II lanthionine synthetase enzyme gene was selected using a MIBiG [38] search of experimentally characterized lanthipeptide BGCs. 28 LanM enzyme amino acid sequences underwent analysis with BLASTP [39] on the BV-BRC platform [40], utilizing the "Reference and representative genomes-Proteins-(faa)" database with a maximum of 100 results and a minimum E-value of 0.0001. All results from each analysis were retained for subsequent studies. The identified homologous LanM amino acid sequences, alongside the 28 experimental MIBiG sequences, were processed in MEGA X [55]. The sequences were aligned using MUSCLE [56] with specific parameters: Open gap: −2.90, Extended gap: 0.00, hydrophobicity multiplier: 1.20, maximum interaction: 16, ensemble method (interactions 1,2): UPGMA, ensemble method (others interactions): UPGMA, and minimum diagnostic length (lambda): 24. Additionally, in MEGA X, phylogenetic trees were made using Neighbor-joining [43]. Evolutionary distances were calculated employing the Poisson correction method, measured in amino acid substitution units per site. Ambiguous positions were removed for each sequence pair using the pairwise deletion option. Bootstrap phylogeny testing was conducted with 1000 replicates, utilizing the Poisson model for amino acid substitution and uniform rates between sites while handling gaps and missing data through pairwise deletion. The genomes associated with LanM enzymes from the clostridial clade were downloaded from the BV-BRCPATRIC platform in FASTA format. Subsequently, the genomes were analyzed using AntiSMASH 6.0 [3] with the following parameters: Detection stringency: relaxed; additional features: all turned on, including BLAST with known sets, BLAST of sets, BLAST of subsets, comparison of MIBiG sets, active site finder, RREF finder, Pfam Cluster analysis, GO term annotation based on Pfam, and analysis with TIGRFam.

#### Precursor peptide analysis

After identifying the lanthipeptide BGCs from the genomes of the *Clostridium* genus, BGCs containing all genes needed for the biosynthesis of a mature peptide were manually selected. The core peptide sequences of the lanthipeptides were then analyzed using the ProtParam and RiPPMiner tools [5] to calculate their physicochemical properties, such as molecular weight, isoelectric point, and GRAVY score (Table S2 in Supporting Information File 1). For the similarity network, a database was created containing 145 precursor peptides of lanthipeptides reported in the MIBiG [38], RiPPMiner [5], and UniProt [57] databases. This input was utilized in the EFI-EST [51] web platform, employing the "FASTA" analysis option with the following parameters: E-value: 0.001; Fragments: disabled. Upon completion of the initial calculation, in the SNN Finishing section, an alignment score threshold of 40% was chosen, with default values for sequence length constraint, and neighborhood connectivity disabled. Network similarity was edited with Cytoscape 3.10.1 [58].

#### AlphaFold models of the biosynthetic enzymes

A phylogenetic analysis was conducted using the closest homologs identified after a BLAST search, employing previously described parameters for CloM1, CloM2, CloPt1, and CloPt2. Additionally, tertiary structure modeling was performed using AlphaFold 2.0 [52] in its CoLab online service [59], with default parameters. Each model underwent energy minimization in Chimera [60]. Subsequently, these models were analyzed in SWISS-MODEL [61] to obtain Ramachandran analysis. Finally, structural alignment was performed using PyMOL for CloM1 and CloM2 with the CylM enzyme (PDB code: 5DZT). For the structural alignment of CloPt1 and CloPt2, the characterized protein PCAT1 (PDB code: 3QF4) was analyzed.

#### Gene expression in *E. coli* and purification of CloA1 and CloA2 precursor peptides and C39 peptidase domain

*E. coli* NiCO21 (DE3) cells were transformed with the following vectors: pRSF-Duet_C39 peptidase domain, pRSF-Duet_CloA1_CloM1, and pRSF-Duet_CloA2_CloM2. The transformed bacteria were cultivated in 1 L LB medium containing 30 mg/L kanamycin at 23 °C. Induction with 0.5 mM IPTG was performed when the $OD_{600nm}$ was between 0.6 and 0.8. Cultures were incubated for 14 h, at 200 rpm, at 18 °C. Subsequently, cells were centrifuged at 10,000 rpm for 30 min at 4 °C. The resulting cell pellets were resuspended in 20 mL of LanA starting buffer (20 mM Tris pH 7.5, 500 mM KCl, 10%

glycerol) and lysed by sonication (4.0 seconds on, 10 seconds off, for a total of 10 minutes at 4.0 potency). Finally, the sample was centrifuged for 30 min at 4 °C at 10,000 rpm. Supernatants were stored at −80 °C until purification.

The C39 peptidase domain, and the clostrisin and cellulosin lanthipeptides were linked with a 6xHis N-terminal tag to perform purifications using affinity chromatography with a 1 mL His Trap nickel affinity column. Sonication lysis was performed and centrifuged during 30 min, at 8,000 rpm, 4 °C. The supernatants were loaded into the column. The column was washed with 5 column volumes of LanA wash buffer 1 (20 mM Tris pH 7.5, 500 mM KCl, 10% glycerol, 0.5 mM imidazole), followed by 10 column volumes of LanA wash buffer 2 (20 mM Tris pH 7.5, 500 mM KCl, 10% glycerol, 30 mM imidazole). Finally, the elution was performed with 3 column volumes of LanA elution buffer (20 mM Tris pH 7.5, 500 mM KCl, 10% glycerol, 750 mM imidazole). Purifications were monitored through 20% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) [51]. For HPLC-LC analysis no native purification was performed, for each pellet of 500 mL liquid culture induction of clostrisin and cellulosin lanthipeptides. The cell pellet was resuspended with 10 mL urea buffer (8 M urea, 100 mM $NaH_2PO_4$, 10 mM Tris-base), pH 8.0, frozen at −80 °C, and then thawed at room temperature. The cell lysate was clarified by centrifugation and incubated with 1 mL Ni-NTA resin (Qiagen). The precursor protein was purified using the standard protocol from Qiagen for purifying protein from Ni-NTA resin under denaturing conditions with urea.

Expression was confirmed through Western blot analysis using the 6xHis epitope tag on a lysate sample (Figure S3A, Supporting Information File 1) for the clostrisin and cellulosin lanthipeptides, and the C39 peptidase domain. Affinity chromatography purification (Figure S7B in Supporting Information File 1) was performed for the C39 peptidase domain, resulting in a purity of approximately 96%, based on densitometric analysis of the Coomassie-stained gel. Protein yield was estimated using spectrophotometric quantification. Analysis using the NanoDrop device at 280 nm indicated a 1.25 mg/L yield. For clostrisin, the same analysis revealed a purity of around 41% and a 1 mg/L yield. In the case of cellulosin, the analysis showed a purity of 60% and a 0.8 mg/L yield.

## Protease activity of the peptidase C39 domain of CloPt1

To optimize the heterologous expression process of protease CloPt1, we refrained from attempting to purify membrane proteins due to technical complexities. Instead, we exclusively focused on cloning the peptidase C39 domain into the pRSF-Duet_C39 peptidase domain vector. The sample containing the C39 peptidase domain, was mixed with samples of CloA1 and CloA2A at a 1:5 molar ratio within a reaction buffer composed of 50 mM HEPES at pH 7.0, 150 mM NaCl, and 5 mM DTT. Cofactors were added as indicated, with the following concentrations: 0.5 mM ATP and 1 mM MgCl2. The reactions were incubated at 4 °C for 36 h. Subsequently, the samples were subjected to analysis using 20% SDS-PAGE.

## HPLC–MS–MS/MS Experiments

Mass spectra were generated using the Agilent 1260 Infinity II LC coupled with the Agilent 6530 QTOF, with the CID voltage determined by the formula $V = 0.036 \times (m/z) - 4.8$. Specifically, the MS spectrum of clostrisin and cellulosin was acquired from the non-native affinity chromatography sample. Additionally, the $MS^2$ spectra of the N27-C87 fragment with five and six dehydrations were obtained from the trypsin digestion of the cellulosin precursor peptide. Finally, the $MS^2$ spectrum of the -K26 to C42 fragment with one alkylation was acquired following an iodoacetamide derivatization reaction and subsequent trypsin digestion. These experimental approaches provide insights into the potential structure of cellulosin.

## Antimicrobial activity of clostrisin and cellulosin

Antimicrobial activity assays were conducted using the reference strains, *A. baumannii* ATCC BAA 747, *P. aeruginosa* PA14 ATCC 15442 and MIQPA25 (clinical strain), *E. coli* ATCC 10799, *S. aureus* ATCC 43300 (MRSA) and *S. epidermidis* MIQ43. Tests were performed in 96-well polystyrene microtiter plates with bacterial inoculation at an initial OD 600 nm of 0.1. The minimum inhibitory concentration (MIC) for each purified lanthipeptide was determined by testing three different concentrations in three separate wells and repeating the process in three independent experiments. For the highest tested concentration of lanthipeptide (5.6 µg/mL for clostrisin and 4.8 µg/mL for cellulosin) the well contained 178 µL of LB broth, 12 µL of purified lanthipeptide and 10 µL of bacterial culture. The plates were incubated at 37°C for 16 h, measuring the optical density (OD) at 600 nm, under shaking.

The results for MIC and percentage inhibition were averaged. The percentage of inhibition was derived by analyzing the 16 h growth curve data using the standard protocol established in our laboratory, comparing the growth curves obtained across the conducted experiments. All data processing and final figures were generated using Prism GraphPad (version 8.0.2).

## Atomic force microscopy

AFM characterization was conducted using a MultiMode 8-HR (Bruker). Samples containing *S. epidermis* at an $OD_{600} > 1$ were incubated with the lanthipeptides at room temperature for

1 and 5 hours. Subsequently, 9 µL of the samples were combined with 1 µL of a 0.1% w/v poly-L-lysin solution in water (Sigma-Aldrich) and immediately deposited onto freshly cleaved mica to be incubated for 10 min at room temperature to allow adsorption. The surface was then rinsed using 600 µL of ultrapure 0.2 µm filtered water and slowly dried using compressed air. Imaging was performed using a Digital Instruments NanoScope V, acquiring 1024 samples per line with silicon nitride cantilevers possessing a nominal spring constant of 0.32 $Nm^{-1}$ and a 0.8–1.0 Hz scan rate. Imaging was conducted at room temperature using the ScanAsyst™ air mode. Images were processed using NanoScope Analysis V.1.80.

## Supporting Information

### Supporting Information File 1

Additional Figures and Tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-159-S1.pdf]

## Acknowledgements

We thank LESP Aguascalientes, Dr. Rodolfo Garcia Contreras from the Faculty of Medicine, UNAM, Mexico, and Dr. Victor Coria, National Pediatric Institute, Mexico, for sharing the multi-resistant strains. We thank Dr. Daniel Finkelstein for the grant awarded to KPVS and KGMG for the PAPIIT project (IT20242). We acknowledge using Grammarly AI to improve the manuscript's English writing.

## Funding

This research was financially supported by DGAPA, PAPIIT (project numbers IA201721, IA205122, IN203923, IT202424) and CONACHYT (CF-2022/319816, CF-2022/319596). MAAH and EGSR acknowledge financial support from CONAHCYT, Mexico, for postgraduate studies.

## Author Contributions

Moisés Alejandro Alejo Hernandez: conceptualization; formal analysis; investigation; methodology; visualization; writing – original draft; writing – review & editing. Katia Pamela Villavicencio Sánchez: investigation; methodology; validation; visualization; writing – original draft; writing – review & editing. Rosendo Sánchez Morales: formal analysis; investigation; validation; visualization; writing – original draft; writing – review & editing. Karla Georgina Hernández-Magro Gil: formal analysis; investigation; methodology; supervision; validation; visualization; writing – original draft; writing – review & editing. David Silverio Moreno-Gutiérrez: conceptualization; formal analysis; investigation; methodology; supervision; validation; writing – original draft; writing – review & editing. Eddie Guillermo Sanchez-Rueda: formal analysis; methodology; validation; visualization; writing – original draft; writing – review & editing. Yanet Teresa-Cruz: methodology; validation; visualization; writing – original draft; writing – review & editing. Brian Choi: formal analysis; investigation; resources; supervision; visualization; writing – review & editing. Armando Hernández Garcia: formal analysis; methodology; project administration; resources; supervision; validation; writing – original draft; writing – review & editing. Alba Romero-Rodríguez: formal analysis; methodology; project administration; supervision; validation; writing – original draft; writing – review & editing. Oscar Juárez: conceptualization; data curation; formal analysis; funding acquisition; project administration; supervision; validation; visualization; writing – original draft; writing – review & editing. Siseth Martínez-Caballero: data curation; formal analysis; funding acquisition; investigation; methodology; project administration; supervision; validation; visualization; writing – original draft; writing – review & editing. Mario Figueroa: data curation; funding acquisition; investigation; project administration; supervision; validation; visualization; writing – original draft; writing – review & editing. Corina-Diana Ceapă: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; writing – original draft; writing – review & editing.

## ORCID® iDs

Moisés Alejandro Alejo Hernandez - https://orcid.org/0009-0005-4911-6404
Katia Pamela Villavicencio Sánchez - https://orcid.org/0009-0004-8006-2339
Rosendo Sánchez Morales - https://orcid.org/0009-0002-0687-6591
Karla Georgina Hernández-Magro Gil - https://orcid.org/0009-0001-6375-8017
David Silverio Moreno-Gutiérrez - https://orcid.org/0000-0002-7363-8561
Eddie Guillermo Sanchez-Rueda - https://orcid.org/0000-0001-7079-9282
Yanet Teresa-Cruz - https://orcid.org/0009-0001-4563-4846
Brian Choi - https://orcid.org/0000-0003-2976-8577
Armando Hernández Garcia - https://orcid.org/0000-0002-2401-8139
Siseth Martínez-Caballero - https://orcid.org/0000-0003-4842-8447
Mario Figueroa - https://orcid.org/0000-0001-7004-0591
Corina-Diana Ceapă - https://orcid.org/0000-0001-8661-4211

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2023.56.v1

# References


1. Murray, C. J. L.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Robles Aguilar, G.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; Johnson, S. C.; Browne, A. J.; Chipeta, M. G.; Fell, F.; Hackett, S.; Haines-Woodhouse, G.; Kashef Hamadani, B. H.; Kumaran, E. A. P.; McManigal, B.; Achalapong, S.; Agarwal, R.; Akech, S.; Albertson, S.; Amuasi, J.; Andrews, J.; Aravkin, A.; Ashley, E.; Babin, F.-X.; Bailey, F.; Baker, S.; Basnyat, B.; Bekker, A.; Bender, R.; Berkley, J. A.; Bethou, A.; Bielicki, J.; Boonkasidecha, S.; Bukosia, J.; Carvalheiro, C.; Castañeda-Orjuela, C.; Chansamouth, V.; Chaurasia, S.; Chiurchiù, S.; Chowdhury, F.; Clotaire Donatien, R.; Cook, A. J.; Cooper, B.; Cressey, T. R.; Criollo-Mora, E.; Cunningham, M.; Darboe, S.; Day, N. P. J.; De Luca, M.; Dokova, K.; Dramowski, A.; Dunachie, S. J.; Duong Bich, T.; Eckmanns, T.; Eibach, D.; Emami, A.; Feasey, N.; Fisher-Pearson, N.; Forrest, K.; Garcia, C.; Garrett, D.; Gastmeier, P.; Giref, A. Z.; Greer, R. C.; Gupta, V.; Haller, S.; Haselbeck, A.; Hay, S. I.; Holm, M.; Hopkins, S.; Hsia, Y.; Iregbu, K. C.; Jacobs, J.; Jarovsky, D.; Javanmardi, F.; Jenney, A. W. J.; Khorana, M.; Khusuwan, S.; Kissoon, N.; Kobeissi, E.; Kostyanev, T.; Krapp, F.; Krumkamp, R.; Kumar, A.; Kyu, H. H.; Lim, C.; Lim, K.; Limmathurotsakul, D.; Loftus, M. J.; Lunn, M.; Ma, J.; Manoharan, A.; Marks, F.; May, J.; Mayxay, M.; Mturi, N.; Munera-Huertas, T.; Musicha, P.; Musila, L. A.; Mussi-Pinhata, M. M.; Naidu, R. N.; Nakamura, T.; Nanavati, R.; Nangia, S.; Newton, P.; Ngoun, C.; Novotney, A.; Nwakanma, D.; Obiero, C. W.; Ochoa, T. J.; Olivas-Martinez, A.; Olliaro, P.; Ooko, E.; Ortiz-Brizuela, E.; Ounchanum, P.; Pak, G. D.; Paredes, J. L.; Peleg, A. Y.; Perrone, C.; Phe, T.; Phommasone, K.; Plakkal, N.; Ponce-de-Leon, A.; Raad, M.; Ramdin, T.; Rattanavong, S.; Riddell, A.; Roberts, T.; Robotham, J. V.; Roca, A.; Rosenthal, V. D.; Rudd, K. E.; Russell, N.; Sader, H. S.; Saengchan, W.; Schnall, J.; Scott, J. A. G.; Seekaew, S.; Sharland, M.; Shivamallappa, M.; Sifuentes-Osornio, J.; Simpson, A. J.; Steenkeste, N.; Stewardson, A. J.; Stoeva, T.; Tasak, N.; Thaiprakong, A.; Thwaites, G.; Tigoi, C.; Turner, C.; Turner, P.; van Doorn, H. R.; Velaphi, S.; Vongpradith, A.; Vongsouvath, M.; Vu, H.; Walsh, T.; Walson, J. L.; Waner, S.; Wangrangsimakul, T.; Wannapinij, P.; Wozniak, T.; Young Sharma, T. E. M. W.; Yu, K. C.; Zheng, P.; Sartorius, B.; Lopez, A. D.; Stergachis, A.; Moore, C.; Dolecek, C.; Naghavi, M. *Lancet* **2022,** *399,* 629–655. doi:10.1016/s0140-6736(21)02724-0
2. *Global Action Plan on Antimicrobial Resistance;* World Health Organization: Geneva, Switzerland, 2016.
3. Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2021,** *49,* W29–W35. doi:10.1093/nar/gkab335
4. van Heel, A. J.; de Jong, A.; Song, C.; Viel, J. H.; Kok, J.; Kuipers, O. P. *Nucleic Acids Res.* **2018,** *46,* W278–W281. doi:10.1093/nar/gky383
5. Agrawal, P.; Khater, S.; Gupta, M.; Sain, N.; Mohanty, D. *Nucleic Acids Res.* **2017,** *45,* W80–W88. doi:10.1093/nar/gkx408
6. Foulston, L. *Curr. Opin. Microbiol.* **2019,** *51,* 1–8. doi:10.1016/j.mib.2019.01.001
7. Ren, H.; Dommaraju, S. R.; Huang, C.; Cui, H.; Pan, Y.; Nesic, M.; Zhu, L.; Sarlah, D.; Mitchell, D. A.; Zhao, H. *Nat. Commun.* **2023,** *14,* 1624. doi:10.1038/s41467-023-37287-1
8. 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.; Jones, M.; Lazarides, L.; Steadman, V. A.; Cohen, D. R.; Felix, C. R.; Fetterman, K. A.; Millett, W. P.; Nitti, A. G.; Zullo, A. M.; Chen, C.; Lewis, K. *Nature* **2015,** *517,* 455–459. doi:10.1038/nature14098
9. Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackermann, Z.; Tran, V. M.; Chiappino-Pepe, A.; Badran, A. H.; Andrews, I. W.; Chory, E. J.; Church, G. M.; Brown, E. D.; Jaakkola, T. S.; Barzilay, R.; Collins, J. J. *Cell* **2020,** *180,* 688–702.e13. doi:10.1016/j.cell.2020.01.021
10. Zhong, Z.; He, B.; Li, J.; Li, Y.-X. *Synth. Syst. Biotechnol.* **2020,** *5,* 155–172. doi:10.1016/j.synbio.2020.06.002
11. van Staden, A. D. P.; van Zyl, W. F.; Trindade, M.; Dicks, L. M. T.; Smith, C. *Appl. Environ. Microbiol.* **2021,** *87,* e00186-21. doi:10.1128/aem.00186-21
12. van Kraaij, C.; de Vos, W. M.; Siezen, R. J.; Kuipers, O. P. *Nat. Prod. Rep.* **1999,** *16,* 575–587. doi:10.1039/a804531c
13. 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.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013,** *30,* 108–160. doi:10.1039/c2np20085f
14. Yang, X.; van der Donk, W. A. *Chem. – Eur. J.* **2013,** *19,* 7662–7677. doi:10.1002/chem.201300401
15. Guder, A.; Wiedemann, I.; Sahl, H.-G. *Biopolymers* **2000,** *55,* 62–73. doi:10.1002/1097-0282(2000)55:1<62::aid-bip60>3.0.co;2-y
16. Major, D.; Flanzbaum, L.; Lussier, L.; Davies, C.; Caldo, K. M. P.; Acedo, J. Z. *Molecules* **2021,** *26,* 7218. doi:10.3390/molecules26237218
17. Lin, Y.; Teng, K.; Huan, L.; Zhong, J. *Microbiol. Res.* **2011,** *166,* 146–154. doi:10.1016/j.micres.2010.05.001
18. Wang, J.; Ge, X.; Zhang, L.; Teng, K.; Zhong, J. *Sci. Rep.* **2016,** *6,* 38630. doi:10.1038/srep38630
19. Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* **2007,** *61,* 477–501. doi:10.1146/annurev.micro.61.080706.093501
20. Deisinger, J. P.; Arts, M.; Kotsogianni, I.; Puls, J.-S.; Grein, F.; Ortiz-López, F. J.; Martin, N. I.; Müller, A.; Genilloud, O.; Schneider, T. *iScience* **2023,** *26,* 106394. doi:10.1016/j.isci.2023.106394
21. Repka, L. M.; Chekan, J. R.; Nair, S. K.; van der Donk, W. A. *Chem. Rev.* **2017,** *117,* 5457–5520. doi:10.1021/acs.chemrev.6b00591
22. Lin, D. Y.-w.; Huang, S.; Chen, J. *Nature* **2015,** *523,* 425–430. doi:10.1038/nature14623
23. Baltz, R. H. *J. Ind. Microbiol. Biotechnol.* **2019,** *46,* 281–299. doi:10.1007/s10295-018-2115-4
24. Xue, D.; Older, E. A.; Zhong, Z.; Shang, Z.; Chen, N.; Dittenhauser, N.; Hou, L.; Cai, P.; Walla, M. D.; Dong, S.-H.; Tang, X.; Chen, H.; Nagarkatti, P.; Nagarkatti, M.; Li, Y.-X.; Li, J. *Nat. Commun.* **2022,** *13,* 1647. doi:10.1038/s41467-022-29325-1

25. Dickman, R.; Mitchell, S. A.; Figueiredo, A. M.; Hansen, D. F.; Tabor, A. B. *J. Org. Chem.* **2019,** *84,* 11493–11512. doi:10.1021/acs.joc.9b01253
26. Karczewski, J.; Krasucki, S. P.; Asare-Okai, P. N.; Diehl, C.; Friedman, A.; Brown, C. M.; Maezato, Y.; Streatfield, S. J. *Front. Microbiol.* **2020,** *11,* 598789. doi:10.3389/fmicb.2020.598789
27. Yates, K. R.; Welsh, J.; Udegbunam, N. O.; Greenman, J.; Maraveyas, A.; Madden, L. A. *Blood Coagulation Fibrinolysis* **2012,** *23,* 396–401. doi:10.1097/mbc.0b013e3283538875
28. Zhou, H.; Fang, J.; Tian, Y.; Lu, X. Y. *Ann. Microbiol. (Heidelberg, Ger.)* **2014,** *64,* 413–420. doi:10.1007/s13213-013-0679-9
29. Field, D.; Blake, T.; Mathur, H.; O' Connor, P. M.; Cotter, P. D.; Paul Ross, R.; Hill, C. *Mol. Microbiol.* **2019,** *111,* 717–731. doi:10.1111/mmi.14183
30. Kuthning, A.; Mösker, E.; Süssmuth, R. D. *Appl. Microbiol. Biotechnol.* **2015,** *99,* 6351–6361. doi:10.1007/s00253-015-6557-6
31. Ortiz-López, F. J.; Carretero-Molina, D.; Sánchez-Hidalgo, M.; Martín, J.; González, I.; Román-Hurtado, F.; de la Cruz, M.; García-Fernández, S.; Reyes, F.; Deisinger, J. P.; Müller, A.; Schneider, T.; Genilloud, O. *Angew. Chem., Int. Ed.* **2020,** *59,* 12654–12658. doi:10.1002/anie.202005187
32. Chapeton-Montes, D.; Plourde, L.; Bouchier, C.; Ma, L.; Diancourt, L.; Criscuolo, A.; Popoff, M. R.; Brüggemann, H. *Sci. Rep.* **2019,** *9,* 11220. doi:10.1038/s41598-019-47551-4
33. Mesa, V.; Monot, M.; Ferraris, L.; Popoff, M.; Mazuet, C.; Barbut, F.; Delannoy, J.; Dupuy, B.; Butel, M.-J.; Aires, J. *Microb. Genomics* **2022,** *8,* 000813. doi:10.1099/mgen.0.000813
34. Pei, Z.; Liu, Y.; Yi, Z.; Liao, J.; Wang, H.; Zhang, H.; Chen, W.; Lu, W. *J. Appl. Microbiol.* **2023,** *134,* lxad127. doi:10.1093/jambio/lxad127
35. Pahalagedara, A. S. N. W.; Flint, S.; Palmer, J.; Brightwell, G.; Gupta, T. B. *Int. J. Antimicrob. Agents* **2020,** *55,* 105910. doi:10.1016/j.ijantimicag.2020.105910
36. Schieferdecker, S.; Shabuer, G.; Knuepfer, U.; Hertweck, C. *Org. Biomol. Chem.* **2019,** *17,* 6119–6121. doi:10.1039/c9ob00968j
37. Chiriac, A. I.; Kloss, F.; Krämer, J.; Vuong, C.; Hertweck, C.; Sahl, H.-G. *J. Antimicrob. Chemother.* **2015,** *70,* 2576–2588. doi:10.1093/jac/dkv161
38. Kautsar, S. A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J. C.; Terlouw, B. R.; van der Hooft, J. J. J.; van Santen, J. A.; Tracanna, V.; Suarez Duran, H. G.; Pascal Andreu, V.; Selem-Mojica, N.; Alanjary, M.; Robinson, S. L.; Lund, G.; Epstein, S. C.; Sisto, A. C.; Charkoudian, L. K.; Collemare, J.; Linington, R. G.; Weber, T.; Medema, M. H. *Nucleic Acids Res.* **2020,** *48,* D454–D458. doi:10.1093/nar/gkz882
39. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990,** *215,* 403–410. doi:10.1016/s0022-2836(05)80360-2
40. Olson, R. D.; Assaf, R.; Brettin, T.; Conrad, N.; Cucinell, C.; Davis, J. J.; Dempsey, D. M.; Dickerman, A.; Dietrich, E. M.; Kenyon, R. W.; Kuscuoglu, M.; Lefkowitz, E. J.; Lu, J.; Machi, D.; Macken, C.; Mao, C.; Niewiadomska, A.; Nguyen, M.; Olsen, G. J.; Overbeek, J. C.; Parrello, B.; Parrello, V.; Porter, J. S.; Pusch, G. D.; Shukla, M.; Singh, I.; Stewart, L.; Tan, G.; Thomas, C.; VanOeffelen, M.; Vonstein, V.; Wallace, Z. S.; Warren, A. S.; Wattam, A. R.; Xia, F.; Yoo, H.; Zhang, Y.; Zmasek, C. M.; Scheuermann, R. H.; Stevens, R. L. *Nucleic Acids Res.* **2023,** *51,* D678–D689. doi:10.1093/nar/gkac1003
41. Ongey, E. L.; Giessmann, R. T.; Fons, M.; Rappsilber, J.; Adrian, L.; Neubauer, P. *Front. Microbiol.* **2018,** *9,* 1688. doi:10.3389/fmicb.2018.01688
42. Zhao, X.; van der Donk, W. A. *Cell Chem. Biol.* **2016,** *23,* 246–256. doi:10.1016/j.chembiol.2015.11.014
43. Saitou, N.; Nei, M. *Mol. Biol. Evol.* **1987,** *4,* 406–425. doi:10.1093/oxfordjournals.molbev.a040454
44. Ren, H.; Shi, C.; Bothwell, I. R.; van der Donk, W. A.; Zhao, H. *ACS Chem. Biol.* **2020,** *15,* 1642–1649. doi:10.1021/acschembio.0c00267
45. Sawa, N.; Wilaipun, P.; Kinoshita, S.; Zendo, T.; Leelawatcharamas, V.; Nakayama, J.; Sonomoto, K. *Appl. Environ. Microbiol.* **2012,** *78,* 900–903. doi:10.1128/aem.06497-11
46. Sand, S. L.; Nissen-Meyer, J.; Sand, O.; Haug, T. M. *Biochim. Biophys. Acta, Biomembr.* **2013,** *1828,* 249–259. doi:10.1016/j.bbamem.2012.11.001
47. Xin, B.; Zheng, J.; Liu, H.; Li, J.; Ruan, L.; Peng, D.; Sajid, M.; Sun, M. *Front. Microbiol.* **2016,** *7,* 1115. doi:10.3389/fmicb.2016.01115
48. Begley, M.; Cotter, P. D.; Hill, C.; Ross, R. P. *Appl. Environ. Microbiol.* **2009,** *75,* 5451–5460. doi:10.1128/aem.00730-09
49. Hyink, O.; Balakrishnan, M.; Tagg, J. R. *FEMS Microbiol. Lett.* **2005,** *252,* 235–241. doi:10.1016/j.femsle.2005.09.003
50. Yonezawa, H.; Kuramitsu, H. K. *Antimicrob. Agents Chemother.* **2005,** *49,* 541–548. doi:10.1128/aac.49.2.541-548.2005
51. Zallot, R.; Oberg, N.; Gerlt, J. A. *Biochemistry* **2019,** *58,* 4169–4182. doi:10.1021/acs.biochem.9b00735
52. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature* **2021,** *596,* 583–589. doi:10.1038/s41586-021-03819-2
53. Dong, S.-H.; Tang, W.; Lukk, T.; Yu, Y.; Nair, S. K.; van der Donk, W. A. *eLife* **2015,** *4,* e07607. doi:10.7554/elife.07607
54. Gutiérrez-Santana, J. C.; Gerónimo-Gallegos, A.; Martínez-Corona, M. B.; López-López, M.; Toscano-Garibay, J. D.; Cuevas-Schacht, F.; Coria-Jiménez, V. R. *Curr. Microbiol.* **2022,** *79,* 353. doi:10.1007/s00284-022-03048-4
55. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. *Mol. Biol. Evol.* **2018,** *35,* 1547–1549. doi:10.1093/molbev/msy096
56. Edgar, R. C. *Nucleic Acids Res.* **2004,** *32,* 1792–1797. doi:10.1093/nar/gkh340
57. The UniProt Consortium. *Nucleic Acids Res.* **2023,** *51,* D523–D531. doi:10.1093/nar/gkac1052
58. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003,** *13,* 2498–2504. doi:10.1101/gr.1239303
59. Wierzchowski, A.; Wink, D. J.; Zhang, H.; Kambanis, K.; Rojas Robles, J. O.; Rosenhouse-Dantsker, A. *J. Chem. Educ.* **2022,** *99,* 4085–4093. doi:10.1021/acs.jchemed.1c01290
60. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004,** *25,* 1605–1612. doi:10.1002/jcc.20084
61. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. *Nucleic Acids Res.* **2018,** *46,* W296–W303. doi:10.1093/nar/gky427

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.159

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Cell-free protein synthesis with technical additives – expanding the parameter space of in vitro gene expression

Tabea Bartsch[1], Stephan Lütz[1] and Katrin Rosenthal[*2]

Full Research Paper

Open Access

Address:
[1]Department of Biochemical and Chemical Engineering, TU Dortmund University, Emil-Figge-Straße 66, 44227 Dortmund, Germany and [2]School of Science, Constructor University, Campus Ring 6, 28759 Bremen, Germany

Email:
Katrin Rosenthal[*] - krosenthal@constructor.university

* Corresponding author

Keywords:
cell-free protein synthesis; cGAS; *Escherichia coli* cell-free extract; sfGFP; TX-TL

*Beilstein J. Org. Chem.* **2024,** *20,* 2242–2253.
https://doi.org/10.3762/bjoc.20.192

Received: 13 June 2024
Accepted: 22 August 2024
Published: 04 September 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra



© 2024 Bartsch et al.; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Biocatalysis has established itself as a successful tool in organic synthesis. A particularly fast technique for screening enzymes is the in vitro expression or cell-free protein synthesis (CFPS). The system is based on the transcription and translation machinery of an extract-donating organism to which substrates such as nucleotides and amino acids, as well as energy molecules, salts, buffer, etc., are added. After successful protein synthesis, further substrates can be added for an enzyme activity assay. Although mimicking of cell-like conditions is an approach for optimization, the physical and chemical properties of CFPS are not well described yet. To date, standard conditions have mainly been used for CFPS, with little systematic testing of whether conditions closer to intracellular conditions in terms of viscosity, macromolecules, inorganic ions, osmolarity, or water content are advantageous. Also, very few non-physiological conditions have been tested to date that would expand the parameter space in which CFPS can be performed. In this study, the properties of an *Escherichia coli* extract-based CFPS system are evaluated, and the parameter space is extended to high viscosities, concentrations of inorganic ion and osmolarity using ten different technical additives including organic solvents, polymers, and salts. It is shown that the synthesis of two model proteins, namely superfolder GFP (sfGFP) and the enzyme truncated human cyclic GMP-AMP synthase fused to sfGFP (*ths*cGAS-sfGFP), is very robust against most of the tested additives.

## Introduction

In addition to other applications such as biomanufacturing or biosensing, cell-free protein synthesis (CFPS) of enzymes has established itself as a tool for rapid screening of biocatalysts [1,2]. The open environment allows easy manipulation of the protein synthesis [3] and coupling to subsequent enzyme activity assays, e.g., for substrate screening [4,5]. The CFPS system

is advantageous for proteins that are difficult to express in a viable host cell, e.g., due to toxic effects on the metabolism [6]. Furthermore, protein synthesis takes only a few hours [7], making the process very fast compared to heterologous expression. CFPS relies on the transcription and translation (TX-TL) system of the donor organism [8]. In addition, the reaction solution contains the DNA-template encoding for the target protein, amino acids and nucleoside triphosphates as substrates, an energy regeneration system and other additives such as polyethylene glycol (PEG) [9].

Although CFPS has been used and improved since the 1960s, there are challenges in its application such as low production volumes, batch-to-batch reproducibility, and reliable kinetic modelling of the system [10,11]. Furthermore, the transferability of CFPS screening results to the cells is limited but important, as in vivo production is often required for preparative scale applications [11,12]. To date, the description of CFPS systems has mainly focused on individual components: the energy regeneration system, the cell extract itself, or individual buffer components [13,14]. For example, the importance of optimizing the concentrations of these reaction components has been demonstrated for four different CFPS systems, in which each system was affected differently by the individual components [15]. This shows that the different CFPS systems have different requirements for their composition, which must be highly balanced. At most, pH and ionic strength are general variables that are considered [13]. Since the main influences of the intracellular environment on the function and cellular behavior of proteins are composition, viscosity, and macromolecular crowding [16], these parameters could have a strong impact on the protein synthesis performance using CFPS. For example, the addition of chemical chaperones, such as alcohols, polyols, polyions or polymers, has a positive effect on protein stability and the soluble fraction expressed with CFPS [17]. The variable composition of CFPS systems with a high number of ingredients and possible reaction conditions [13] thus opens a large parameter space. In addition, non-physiological conditions further extend the parameter space in which CFPS can be performed. This extension would be highly desirable for a coupled CFPS and enzyme assay, where, e.g., an organic solvent is used to solubilize poorly water soluble substrates for the enzyme.

In this study, we therefore aim to fill some of the gaps in the consideration of the general physical properties and potential effects on the performance of *Escherichia coli*-based CFPS. We use technical additives, such as water-soluble macromolecular polymers and salts, which are commonly used as deep eutectic solvents (DES) and extend the properties beyond physiological ranges. We also tested several organic solvents that are miscible and non-miscible with water. For the experiments, two model proteins, the superfolder GFP (sfGFP) and the enzyme human cyclic GMP-AMP synthase-sfGFP, were used, which differ in their size (sfGFP: 27 kDa, *th*scGAS-sfGFP: 84 kDa [2,18]) and fractional yield obtained for in vitro expression (sfGFP: 58%; *th*scGAS-sfGFP: 9% [12]).

# Results and Discussion

## Effects of additives on fluid properties in CFPS

The fluid properties of the cytoplasm of *E. coli*, the CFPS system, and additives were determined to evaluate their influence on the synthesis performance of CFPS. Polymers, DES, and organic solvents were considered to modify the fluid properties.

### Polymers and deep eutectic solvents (DES) as additives in CFPS

Polymers (PEG, methylcellulose (MC), and carboxymethylcellulose (CMC)) and DES (choline chloride/urea, betaine/ethylene glycol (EG), and choline chloride/glycerol) were chosen as additives to vary the viscosity, ion concentration, amount of macromolecules, and osmolarity in CFPS. The calculated values for the properties of the CFPS system with polymers and DES added at different concentrations are shown in Table 1 in comparison with the *E. coli* cytoplasm and water.

Comparison of the fluid properties of the natural cytoplasm in *E. coli* [16,19,20] with those calculated for our standard CFPS system shows that although the values are not the same, they are not orders of magnitude different. The CPFS system has a lower viscosity than that of the *E. coli* cytoplasm [25]. The amount of macromolecules in the CFPS system is 167–265 g/L, only slightly less than in a living cell. The calculation is based on the estimated concentrations of macromolecules in the cell extract, tRNA, plasmid, and PEG. Although the molecular weight of PEG-8000 (8000 g/mol) is below the typical definition of a macromolecule (10,000 g/mol [26]), it is conventionally considered an artificial crowding agent [25] and was included in the calculation of macromolecules. The concentration of inorganic ions in the CFPS system, calculated from the magnesium and potassium glutamate concentration, is 140 mM, which is less than half the concentration reported for the cellular environment. The cytoplasmic osmolarity of about 600 mOsm [16] is 50% higher than that calculated for CFPS. Taking into account all defined components, the water content in the CFPS system is 22% higher than in the cytoplasm, as expected for a diluted system.

**Table 1:** Properties of cytoplasm, water, CFPS solution without additives, CFPS with polymers, and CFPS with DES (25 °C, 1 bar). Some values were taken from the literature as indicated. All other values were calculated (Supporting Information File 1). The line marked with the bold text is the standard composition of the in-house CFPS system and serves as reference. PEG, MC, CMC: % ≙ % w/v. Other: % ≙ % v/v.

| additive to CFPS | concentration [%] | viscosity [mPa·s] | macromolecules [g/L] | inorganic ions [mM] | osmolarity [mOsm] | water content [% v/v] |
|---|---|---|---|---|---|---|
| *E. coli* cytoplasm | n.a. | 3–9.7 [16,19] | 300–500 [16] | 300 [16] | 600[a] [16] | 70 [20] |
| water | n.a. | 0.9 [21] | – | – | – | 100 |
| PEG-8000 | **2[b]** | **1.4** [22] | **167–265** | **140** | **405** | **92** |
| | 5 | 2.2 | 197–295 | 140 | 405 | 89 |
| | 10 | 8.9 [23] | 247–345 | 140 | 405 | 84 |
| methylcellulose | 0.5 | 3.5 [24] | 172–270 | 140 | 405 | 92 |
| | 0.75 | 4.7 | 174–272 | 140 | 405 | 91 |
| | 1 | 6.0 [24] | 177–275 | 140 | 405 | 91 |
| | 2 | 12–18[c,d] | 187–285 | 140 | 405 | 90 |
| carboxymethylcellulose | 0.5 | 17.8 | 172–270 | 140 | 405 | 92 |
| | 0.75 | 71.3 | 174–272 | 140 | 405 | 91 |
| | 1 | 142.5 | 177–275 | 140 | 405 | 91 |
| | 2 | 1000–1500[c] | 187–285 | 140 | 405 | 90 |
| choline chloride/urea 1:2 | 2 | 1.5 | 167–265 | 232 | 590 | 90 |
| | 5 | 1.5 | 167–265 | 371 | 867 | 87 |
| | 10 | 1.7 | 167–265 | 602 | 1329 | 82 |
| choline chloride/glycerol 1:2 | 2 | 1.4 | 167–265 | 214 | 553 | 90 |
| | 5 | 1.5 | 167–265 | 325 | 775 | 87 |
| | 10 | 1.6 | 167–265 | 511 | 1146 | 82 |
| betaine/ethylene glycol 1:3 | 2 | 1.4 | 167–265 | 140 | 405 | 90 |
| | 5 | 1.5 | 167–265 | 140 | 405 | 87 |
| | 10 | 1.6 | 167–265 | 140 | 405 | 82 |

[a]Sum of inorganic ions and combined metabolites.
[b]Standard composition of in-house CFPS system and reference.
[c]Manufacturer specification.
[d]At 20 °C; n.a.: not applicable.

The fluid properties of the CFPS can be modified by adding various additives. The viscosity of the CFPS system can be increased by adding polymers. Different concentrations of PEG-8000, methylcellulose (MC), and carboxymethylcellulose (CMC) cover a wide range of viscosities including that of the cytoplasm up to a very viscous solution. The concentrations of methylcellulose and carboxymethylcellulose are limited to 2% as the viscosity would have become too high. Simultaneously, polymers contribute to the concentration of macromolecules. PEG, a water-soluble macromolecular polymer, is a commonly used crowding agent to mimic the cellular environment in vitro [25]. Interestingly, it has been observed that different proteins have different PEG-8000 concentration optima [27]. The standard composition of the in-house CFPS system contains 2% PEG-8000, resulting in a lower viscosity of the liquid system. By adding up to 10% PEG-8000, both the viscosity and the concentration of macromolecules in the CFPS system reach the physiological range.

To increase the concentration of inorganic ions, DESs were added to the CFPS solution. Although the viscosities of pure DES are relatively high (choline chloride/urea 1:2: 1200 mPa·s [28], choline chloride/glycerol 1:2: 300 mPa·s [28], betaine/ethylene glycol 1:3: 65 mPa·s [29]), the effect on viscosity when adding 2–10% to the CFPS system is almost negligible. The concentration of inorganic ions in CFPS reactions to which choline chloride was added ranged from 232 to 602 mM, exceeding the cytoplasmic inorganic ion concentration of 300 mM. Osmolarity is also increased by increasing salt concentrations as well. With the addition of 10% choline chloride/urea, the osmolarity of the CFPS system increases to 1329 mOsm. Next to choline chloride/urea and choline chloride/glycerol, betaine/ethylene glycol (EG) was tested as it is considered to be an environmentally friendly natural deep eutectic solvent (NADES) [30] and is widely used with proteins [31]. As it does not consist of any ions there is only a slight increase in the viscosity, but no changes are expected for

the other parameters. Except for the water content which, as with all additives, decreases with the percentage of substance added. The lowest value is a water content of 82%, which is still more than 10% above the cytosolic water content of 70% [20], but 10% below the standard conditions of our CFPS system.

### Solvents as additives in CFPS

For some applications, the usage of solvents in CFPS might be beneficial. Organic solvents as additives do not contribute to more cell-like conditions in CFPS systems, but could allow the use of poorly soluble substrates in combined enzyme assays if they are tolerated.

The effects of organic solvents on the properties of the CFPS system are different from those of polymers and DES. Some fluid properties of the pure solvents and the calculated viscosity of the CFPS system with different concentrations of water-soluble solvents are shown in Table 2. MTBE and *n*-hexane have low solubility in water and formed a second phase on top of the aqueous CFPS solution. For experiments with *n*-hexane and MTBE, the size of the vessel was reduced and the reaction volume was increased to 100 µL to avoid evaporation of the solvent in the headspace of the reaction vessel due to the high vapor pressures. In contrast to the standard volumetric ratio of 20 µL in a 1.5 mL microreaction tube, a visible gradient in the concentration of sfGFP occurred under these conditions. Therefore, shaking at 700 rpm was set to ensure adequate mixing. DMSO and methanol are highly soluble in water, which facilitates handling. Their influence on the viscosity of the CFPS system is negligible, as can be seen in Table 2. The polarities of the different solvents cover a wide range to show the effect on CFPS and give various options for soluble substances.

## Effects of technical additives on the CFPS performance

### In vitro sfGFP production with additives

The in vitro expression of sfGFP, or GFP variants in general, is well established and is often used as a model system for optimization (e.g., with active learning workflows [45]) and performance evaluation. This is convenient because product formation and even concentration can be easily quantified measuring the fluorescence intensity. To take advantage of this, fusion proteins with sfGFP can be constructed for CFPS performance evaluation [46].

Therefore, sfGFP was used to establish a reference CFPS synthesis under standard conditions containing 2% PEG-8000. A concentration of 1.77 mg/mL sfGFP was obtained after 4 hours. The calculated fractional yield of 114% based on the added amino acid concentration is higher than expected, which can be explained either by deviations in the measurements or by the undefined addition of amino acids via the cell-free extract. Irrespective of this, the high concentration of sfGFP obtained shows the high level of optimization of the synthesis. Therefore, it is not expected that the addition of technical additives will further increase the synthesis yield, but the general influence of all additives will be investigated. Figure 1 shows the results of sfGFP synthesis with different technical additives. All values were normalized in relation to the fluorescence intensity of the reference with 2% PEG-8000.

Macromolecular crowding is a known mechanism that positively influences CFPS reactions [25]. Interestingly, increasing the concentration of PEG-8000, and thus increasing the viscosity and macromolecular concentration of the CFPS solution towards the properties of the cytoplasm, has a negative effect on

**Table 2:** Properties of water and added solvents (25 °C, 1 bar). Given values are for pure substances. Viscosities of water-soluble solvents (DMSO and methanol) are additionally calculated for solutions with CFPS at displayed concentrations.

| | water | DMSO | methanol | MTBE | *n*-hexane |
|---|---|---|---|---|---|
| molecular weight [g/mol] | 18.02 [21] | 78.14 [32] | 32.04 [21] | 88.15 [33] | 86.18 [21] |
| density [g/L] | 997 [21] | 1100[a] [34] | 786 [21] | 741 [33] | 655 [21] |
| solubility in water [g/L] | – | 1000[b] [32] | 1000 [35] | 26[a] [36] | 0.009 [37] |
| vapor pressure [mmHg] | 23.8 [38] | 0.6 [32] | 127 [35] | 245 [33] | 153 [37] |
| viscosity [µPa·s] | 890 [21] | 2140[a] [34]<br>2%: 1409<br>5%: 1414<br>10%: 1423 | 544 [21]<br>2%: 1411<br>5%: 1420<br>10%: 1434 | 370[c] [39] | 298 [21] |
| polarity [D] | 2.9 [40] | 3.96 [41] | 2.61 [42] | 1.25 [43] | 1.08[b] [44] |
| log P [–] | – | –1.35[b] [32] | –0.77[b] [35] | 0.94[b] [33] | 3.9[b] [37] |

[a]At 20 °C.
[b]Temperature unknown.
[c]At 15 °C.

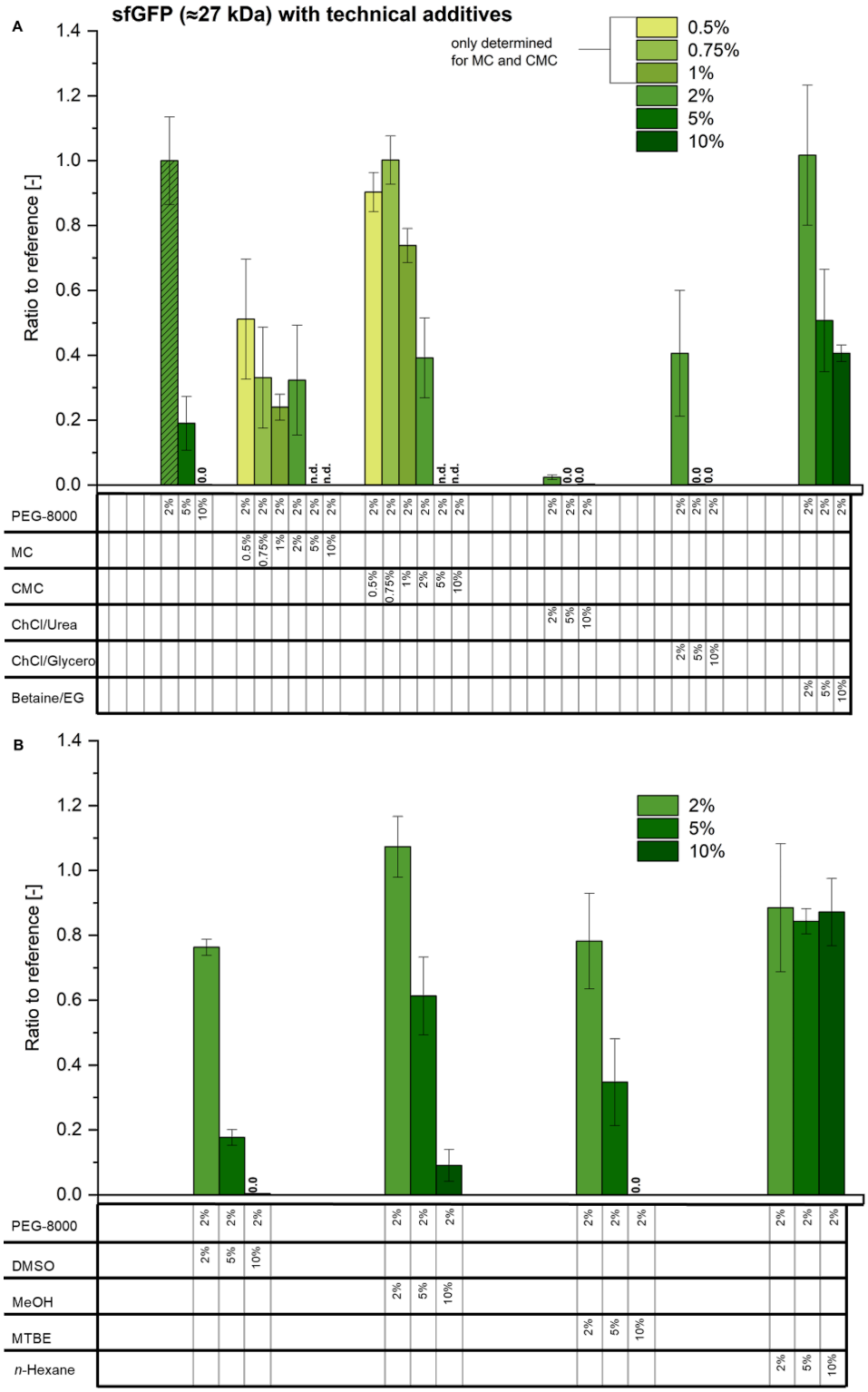



**Figure 1:** CFPS of sfGFP with different technical additives at various concentrations. Experiments with 2% PEG-8000 serves as reference, equal to 1.77 ± 0.24 mg/mL sfGFP. A) represents the additives polyethylene glycol with a molecular weight of 8000 g/mol (PEG-8000), methylcellulose (MC), carboxymethylcellulose (CMC), choline chloride/urea (ChCl/urea), choline chloride/glycerol (ChCl/glycerol), and betaine/ethylene glycol (betaine/EG); the reference experiment is labeled as shaded bars. B) represents dimethyl sulfoxide (DMSO), methanol (MeOH), methyl *tert*-butyl ether (MTBE), and *n*-hexane. Measurements in triplicates. 0.5–1% only for MC and CMC. PEG, MC, CMC: % ≙ % w/v. Other: % ≙ % v/v. Note: 2% PEG-8000 is present as a standard component in all reactions, unless otherwise stated. n.d.: not determined; 0.0: no detectable amount.

the sfGFP synthesis. This is in agreement with results on the co-optimization of PEG with phosphoenolpyruvate, which indicate an optimal concentration of 4% PEG (7.5 kDa) and a decrease in the activity of the CFPS system at concentrations above 5% [47]. PEG-8000 has been shown to have opposite effects on transcription and translation. While transcription is stable up to concentrations of 10% PEG-8000, translation is already inhibited at a concentration of 1% [48]. When the viscosities are increased with methylcellulose or carboxymethylcellulose, a less negative influence on the sfGFP synthesis is observed. This indicates that the impact of the polymer itself is greater than that of the increased viscosity. A positive effect on the stability and activity of the model enzyme β-D-glucuronidase by the addition of carboxymethylcellulose has already been reported [49]. However, the effects and states of molecular crowding in the cell are much more complex than what can be mimicked by the sole addition of a polymer. The diffusion of macromolecules depends on the perceived viscosity in the cell, which is inhomogeneous and depends on the location in the cell and its growth phase [50,51].

Comparing the salt concentrations, the concentration in the CFPS system is lower than that in the cells. It was therefore increased by the addition of DES, which have already been used for biological applications [52] and are considered as promising environmentally friendly alternative solvents [53]. The addition of choline chloride appears to have a strong negative effect on the in vitro synthesis of sfGFP. Although the concentration of inorganic ions and osmolarity are close to physiological conditions at a concentration of 2% choline chloride, a reduction of sfGFP production to only 2% was measured for choline chloride/urea and 41% for choline chloride/glycerol compared to the standard composition. This is less than for the addition of any other additive. All other parameters are constant, suggesting that increased salt concentrations have a negative effect on CFPS. Usually, higher salt concentrations can cause an increased precipitation of proteins [54], which would decrease the amount of detectable CFPS product. Other publications describe that the solubility of proteins can be improved with increasing salt concentration by the addition of NaCl [55]. Testing of other salts at high concentrations is necessary to clarify whether the salt concentration or the salt itself is responsible for the low in vitro protein production with choline chloride as an additive.

Interestingly, the addition of organic solvents has little influence on the synthesis performance at concentrations up to 2%. Even methanol concentrations of 5% are well tolerated by the system demonstrating a high robustness against these additives. With *n*-hexane, the amount of sfGFP is stable at around 85% of the reference for all concentrations tested. The reason for this might be that the influence on the CFPS system does not increase with a higher amount of *n*-hexane due to the low solubility of *n*-hexane in water, a limited interfacial area, and evaporation in the headspace. There is no clear trend for the influence of the polarity or log P of the added solvents. Among the water-miscible additives, methanol, whose dipole moment is relatively close to that of water, is better accepted than DMSO.

## In vitro *ths*cGAS-sfGFP production with additives

The CPFS system used, or CFPS in general, is not further optimized for the production of *ths*cGAS-sfGFP or other specific enzymes. By default, GFP is generally used as a model protein to optimize the composition of CFPS, as it provides a simple and robust readout, the fluorescence signal. As a result, the composition of the CFPS solution is optimized for the synthesis of GFP. In addition, larger and more complex enzymes are usually more difficult to synthesize with CFPS [56], although there are exceptions, such as the production of non-ribosomal peptide synthetases with a molecular weight higher than 100 kDa [57]. However, these enzymes are often of great interest for specific applications in biomanufacturing. Cyclic GMP-AMP synthase (cGAS) is one of these enzymes. cGAS and its biocatalytic product 2'3'-cyclic GMP-AMP (cGAMP) are part of the innate immune response in higher eukaryotes [58]. cGAMP is therefore a promising candidate for pharmaceutical applications [59]. The successful synthesis of the fusion protein of truncated human cGAS and superfolder GFP (*ths*cGAS-sfGFP) using the in-house *E. coli*-based CFPS system has already been demonstrated under standard conditions [18]. We have now repeated this experiment and tested the synthesis of *ths*cGAS-sfGFP with the addition of additives.

The average production of *ths*cGAS-sfGFP under reference conditions using the in-house CFPS system was 0.13 mg/mL, which is comparable to published data [12]. The fractional yield for *ths*cGAS-sfGFP is approximately 10.5% and therefore has potential for optimization. Figure 2 summarizes results for the production of *ths*cGAS-sfGFP. In general, the trend of the protein concentrations obtained is consistent with that observed for sfGFP. The production levels are at or below the reference value of 2% PEG-8000. Among the additives, 0.75% of carboxymethylcellulose and 2% of betaine/EG, methanol and *n*-hexane are the best with 76–93% compared to the reference. The addition of carboxymethylcellulose results in 75% for sfGFP and *ths*cGAS-sfGFP.

## Parameter scope and robustness of CFPS

The composition of reaction media for enzymatic applications, especially in the chemical industry, has been expanded to technical additives [60]. It was therefore tested whether CFPS is also compatible with the addition of technical additives. CFPS show a high robustness against various additives including cyto-

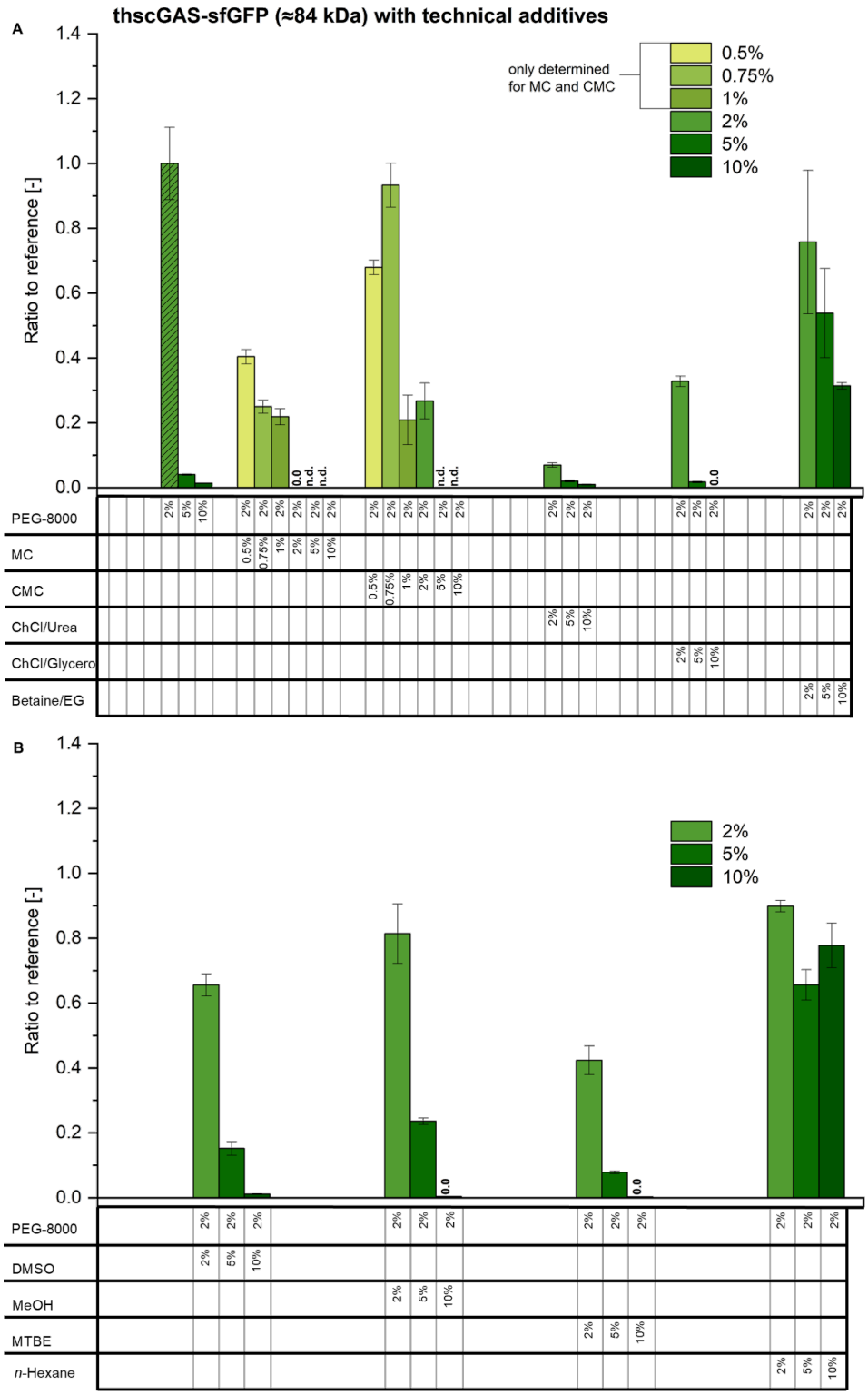



**Figure 2:** CFPS of *thsc*GAS-sfGFP with different technical additives at various concentrations. Experiments with 2% PEG-8000 serve as reference, equal to 0.13 ± 0.02 mg/mL *thsc*GAS-sfGFP. A) represents the additives polyethylene glycol with a molecular weight of 8000 g/mol (PEG-8000), methylcellulose (MC), carboxymethylcellulose (CMC), choline chloride/urea (ChCl/urea), choline chloride/glycerol (ChCl/glycerol), and betaine/ethylene glycol (betaine/EG). The reference experiment is labeled as shaded bars. B) represents dimethyl sulfoxide (DMSO), methanol (MeOH), methyl *tert*-butyl ether (MTBE), and *n*-hexane. Measurements in triplicates. 0.5–1% only for MC and CMC. PEG, MC, CMC: % ≙ % w/v. Other: % ≙ % v/v. Note: 2% PEG-8000 is present as a standard component in all reactions, unless otherwise stated. n.d.: not determined; 0.0: no detectable amount.

toxic solvents, which was demonstrated by the synthesis of sfGFP and *ths*cGAS-sfGFP. The combination of CFPS with activity assays would allow further parallelization and also miniaturization, for example into femtoliter-sized droplets [61,62].

In general, protein synthesis with CFPS is best at low additive concentrations. The general trend for the production of *ths*cGAS-sfGFP were similar compared to sfGFP, but the protein concentrations obtained were overall lower. The protein production still works for most additives up to a concentration of 10%, except when 10% of MTBE or choline chloride/glycerol was added, which completely inhibited the CFPS system. The effect of the addition of *n*-hexane in the experiments is low, which could be due to its low solubility. For the other additives tested, 5–10% seems to be the limit, with the exception of betaine/EG, which is accepted in higher concentrations.

The robustness of the transcription-translation machinery is astonishing and extends the parameter range for CFPS. Successful protein synthesis was observed at very high viscosities, increased concentrations of macromolecules, organic ions, and osmolarity. The tolerated viscosities (determined for standard conditions: 25 °C, 1 bar) range from 1.4 to about 1000 mPa·s. At concentrations of inorganic ions of up to 602 mM and an osmolarity of 1329 mOsm, small concentrations of *ths*cGAS-sfGFP were still detected. The concentration of macromolecules and the water content of the standard CFPS system are in the range of *E. coli* cytoplasm [16,20]. We were able to extend these conditions for the CFPS system to values between 167 and 265 g/L macromolecules corresponding to 82 to 92% of water content.

## Conclusion

This study shows that CFPS is robust against various technical additives. The general trend is a decrease in protein concentration with increasing concentrations of additives, but still detectable amounts of product were reported with 10% of PEG, choline chloride/urea, betaine/ethylene glycol, DMSO, methanol, and *n*-hexane. The results are most promising for betaine/ethylene, methanol, and *n*-hexane and open new potential for applications such as on-site synthesis of enzymes for subsequent biotransformation. In addition to evaluating the physical properties of a standard CFPS system, the range of parameters for CFPS was successfully extended to high values of viscosity, concentrations of inorganic ions, and osmolarity. Carboxymethylcellulose was identified as an interesting alternative crowding agent. This provides a starting point for a multifactorial approach to optimize the synthesis of non-model enzymes.

## Experimental

### Additives and preparation of deep eutectic solvents

The selected additives are supposed to shift the properties of the reaction solution in more extreme directions or are interesting for other reasons. PEG-8000 (Sigma, Darmstadt, Germany) is a molecular crowder that is used in the CFPS system by default. Carboxymethylcellulose sodium salt (Roth, Karlsruhe, Germany) and methylcellulose (VWR, Darmstadt, Germany) are polymers that expand the viscosity range. Choline chloride (VWR, Darmstadt, Germany) + urea (Roth, Karlsruhe, Germany) (molar ratio 1:2) and choline chloride + glycerol (Roth, Karlsruhe, Germany) (molar ratio 1:2) were chosen to increase the amount of salts in the solution. Betaine (Sigma, Darmstadt, Germany) + ethylene glycol (Roth, Karlsruhe, Germany) (molar ratio 1:3) was selected as additional common DES. The DES were prepared by weighting out the substances and stirring at up to 100 °C until liquid state was reached [29]. For the organic solvents, DMSO (Roth, Karlsruhe, Germany) and methanol ≥99% (Roth, Karlsruhe, Germany) were used as water-soluble, and MTBE (Arcos organics, Schwerte, Germany) and *n*-hexane (Lach:ner, Neratovice, Czech Republic) as non-water-soluble solvents. Values for properties were taken from the databases Chemistry WebBook by the National Institute of Standards and Technology (NIST) (https://webbook.nist.gov/) at 1 bar and 25 °C, PubChem by the National Center for Biotechnology and Information (https://pubchem.ncbi.nlm.nih.gov/), and GESTIS-Stoffdatenbank by Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung (https://gestis.dguv.de/), datasheets provided by manufacturers and other sources as referenced at the corresponding point. Viscosity of the standard CFPS system was estimated as that of 2% PEG in water. For the other concentrations and polymers, the value was assumed as that for the component with pure water as well, as the contribution of the other components to the viscosity is considered neglectable in comparison to the high viscosity of the polymer–water mixtures. Some of these were inter- or extrapolated from published values. For carboxmethylcellulose values were derived from the manufacturer's specification with the rule of thumb that doubling the concentration increases the viscosity by a factor of about 8 [63]. As a simple approach according to Arrhenius, the viscosity for mixtures with DES and the water-soluble solvents was calculated with Equation 1 [64], as only minor influences are assumed at the concentrations used in this work:

$$\log \eta_s = N_1 \log \eta_1 + N_2 \log \eta_2 \quad (1)$$

The concentration of macromolecules was calculated based on the cytosolic composition and the average $OD_{600}$ at the harvest of the culture for the cell-free extract. With the derived number of cells, the cellular volume of 4.4 μm³ per *E. coli* cell [65] and the intracellular concentration of macromolecules, the range for the total amount of macromolecules extracted from the culture was determined. The volume of buffer, dilution at CFPS assembly and PEG-8000, tRNA and plasmid as further macromolecules were included for the calculation of the macromolecular concentration of the CFPS mix.

Magnesium and potassium glutamate were considered as the contributing inorganic ions for the reference composition, for the DES corresponding salts were added to the value of 140 mM.

For the osmolarity the concentrations of all defined components were multiplied with their number of dissociated particles, which was assumed as 1 for most components and 2 for magnesium and potassium glutamate and combined with the calculated concentration of macromolecules.

For the water content, the amount of all known components was subtracted from the 100% of pure water, additives decreased that value by the percentage of their contribution. Calculations and further details can be found in Supporting Information File 1.

## Cell-free protein synthesis

The CFPS system was prepared and the reactions performed according to Rolf et al. [18] with the described strains *E. coli* BL21(DE3) pAR1219 for extract preparation and *E. coli* DH5α pETH6sfGFP and *E. coli* DH5α pETSUMO*ths*cGASGFP for plasmid production. Minor variations are stated in the following. The preculture for extract preparation was grown for 20 h at 200 rpm and 37 °C, centrifugation for cell harvesting and washing was performed for 20 min at 3220*g* and storage was at −70 °C. The extract contained 46–67 mg/mL protein. It was premixed with the buffer consisting of magnesium and potassium glutamate, 20 amino acids, HEPES, ATP, GTP, cytidine triphosphate (CTP), uridine triphosphate (UTP), tRNA, coenzyme A (CoA), nicotinamide adenine dinucleotide (NAD), cyclic adenosine monophosphate (cAMP), folinic acid, spermidine, 3-PGA, and PEG-8000 to obtain a master mix. The master mix was assembled with the plasmid encoding for sfGFP respectively *ths*cGAS-sfGFP and nuclease-free water, which added up the free volume to the final CFPS volume of 20 μL in a 1.5 mL microreaction tube. The final composition of the reaction was 11–16 mg/mL protein from extract, 10 mM magnesium glutamate, 130 mM potassium glutamate, 1.5 mM of each of 20 amino acids except for leucine, which is 1.25 mM, 50 mM HEPES, 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/mL tRNA, 0.26 mM CoA, 0.33 mM NAD, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30 mM 3-PGA, 2% PEG-8000, and 1 nM plasmid DNA. Reactions were incubated for 4 h at 37 °C with no shaking. Resulting fluorescence intensities were measured from 2 μL reaction solution in 98 μL 0.5 M HEPES buffer (pH 8.0) in 384-well microplates with a FLUOstar® Omega multi-mode microplate reader (BMG LABTECH, Ortenberg, Germany). The endpoint measurement was set to a gain of 1390, $\lambda_{ex}$ 485 nm and $\lambda_{em}$ 520 nm.

For the experiments with organic solvents, the corresponding volume of 2, 5, and 10% v/v was added right before the incubation. To keep the final volume at 20 μL the volume of added water was decreased by the same volume. For the non-water-soluble solvents, the scale was linearly increased to 100 μL in a 200 μL microreaction tube, incubation was with shaking at 700 rpm. DES and PEG were pre-diluted with nuclease-free water for better pipettablity and added to a final concentration of 2, 5 and 10% v/v, respectively 2, 5 and 10 % w/v for PEG. Methylcellulose and carboxymethylcellulose were added as solid powders to the master mix in the appropriate amount to set the final concentration in the reaction to 0.5, 0.75, 1 and 2% w/v.

All reactions were prepared in triplicates with an additional negative control without the addition of DNA. For all reactions with additives a triplicate of the standard composition was run at the same time and with the same cell-free extract as a reference.

## Correlation of fluorescence intensities and protein concentrations

Plasmids for CFPS were expressed in *E. coli* BL21 (DE3) and purified as described by Rolf et al. [18]. The quantifications of purified proteins and set dilution series were performed with the Bradford assay [66]. Purity of the in vivo produced proteins was checked with sodium dodecyl sulfate polyacrylamide gel-electrophoresis (SDS-PAGE) [67]. Impurities were quantified with ImageJ [68] and measured protein concentrations corrected by the results to gain concentrations of pure sfGFP and *ths*cGAS-sfGFP. Fluorescence was measured with FLUOstar® Omega multi-mode microplate reader (BMG LABTECH, Ortenberg, Germany) under the same conditions as for the in vitro-produced proteins to determine the correlation between fluorescence intensity and protein concentration for each protein.

## Fractional yield

The fractional yield is the ratio between the theoretically achievable protein concentration based on the amount of provi-

ded amino acids in a CFPS system and the sequence of the target protein and the experimentally achieved result [12]. Fractional yields in this work have been calculated using the excel sheet provided by Rolf et al. [12] and can be found in Supporting Information File 3 and Supporting Information File 4.

## Supporting Information

### Supporting Information File 1

Calculations of viscosity, macromolecules, inorganic ions, osmolarity and water content for CFPS in Table 1.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-192-S1.xlsx]

### Supporting Information File 2

Calculations of viscosity for CFPS with water-soluble solvents in Table 2.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-192-S2.xlsx]

### Supporting Information File 3

Fractional Yield of average sfGFP production with CFPS.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-192-S3.xlsx]

### Supporting Information File 4

Fractional yield of average t*hs*cGAS-sfGFP production with CFPS.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-192-S4.xlsx]

## Acknowledgements

We would like to thank Jascha Rolf for the initial preliminary experiments and Christoph Held for the discussion on the selection of additives.

## ORCID® iDs

Tabea Bartsch - https://orcid.org/0009-0006-8290-4952
Stephan Lütz - https://orcid.org/0000-0001-8534-0554
Katrin Rosenthal - https://orcid.org/0000-0002-6176-6224

## Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

## Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2024.41.v1

## References

1. Karim, A. S.; Jewett, M. C., Eds. *Cell-Free Gene Expression;* Methods in Molecular Biology; Springer US: New York, NY, USA, 2022. doi:10.1007/978-1-0716-1998-8
2. Rolf, J.; Rosenthal, K.; Lütz, S. *Catalysts* **2019,** *9,* 190. doi:10.3390/catal9020190
3. Jewett, M. C.; Calhoun, K. A.; Voloshin, A.; Wuu, J. J.; Swartz, J. R. *Mol. Syst. Biol.* **2008,** *4,* 220. doi:10.1038/msb.2008.57
4. Quertinmont, L. T.; Orru, R.; Lutz, S. *Chem. Commun.* **2015,** *51,* 122–124. doi:10.1039/c4cc08240k
5. Rolf, J.; Nerke, P.; Britner, A.; Krick, S.; Lütz, S.; Rosenthal, K. *Catalysts* **2021,** *11,* 1038. doi:10.3390/catal11091038
6. Orth, J. H. C.; Schorch, B.; Boundy, S.; Ffrench-Constant, R.; Kubick, S.; Aktories, K. *Toxicon* **2011,** *57,* 199–207. doi:10.1016/j.toxicon.2010.11.011
7. Caschera, F.; Noireaux, V. *Biochimie* **2014,** *99,* 162–168. doi:10.1016/j.biochi.2013.11.025
8. Swartz, J. R.; Jewett, M. C.; Woodrow, K. A. Cell-Free Protein Synthesis With Prokaryotic Combined Transcription-Translation. *Recombinant Gene Expression;* Methods in Molecular Biology, Vol. 267; Humana Press, 2004; pp 169–182. doi:10.1385/1-59259-774-2:169
9. Sun, Z. Z.; Hayes, C. A.; Shin, J.; Caschera, F.; Murray, R. M.; Noireaux, V. *J. Visualized Exp.* **2013,** No. 79, e50762. doi:10.3791/50762
10. Nirenberg, M. W.; Matthaei, H. *Proc. Natl. Acad. Sci. U. S. A.* **1961,** *47,* 1588–1602. doi:10.1073/pnas.47.10.1588
11. Silverman, A. D.; Karim, A. S.; Jewett, M. C. *Nat. Rev. Genet.* **2020,** *21,* 151–170. doi:10.1038/s41576-019-0186-3
12. Rolf, J.; Handke, J.; Burzinski, F.; Lütz, S.; Rosenthal, K. *Biotechnol. Prog.* **2023,** *39,* e3373. doi:10.1002/btpr.3373
13. Dopp, B. J. L.; Tamiev, D. D.; Reuel, N. F. *Biotechnol. Adv.* **2019,** *37,* 246–258. doi:10.1016/j.biotechadv.2018.12.006
14. Huang, M.; Wang, W.; Guo, T.; Long, X.; Niu, F. *Processes* **2022,** *10,* 1122. doi:10.3390/pr10061122
15. Zhang, L.; Lin, X.; Wang, T.; Guo, W.; Lu, Y. *Bioresour. Bioprocess.* **2021,** *8,* 58. doi:10.1186/s40643-021-00413-2
16. Theillet, F.-X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J.; Elcock, A. H.; Gershenson, A.; Selenko, P. *Chem. Rev.* **2014,** *114,* 6661–6714. doi:10.1021/cr400695p
17. Kai, L.; Dötsch, V.; Kaldenhoff, R.; Bernhard, F. *PLoS One* **2013,** *8,* e56637. doi:10.1371/journal.pone.0056637
18. Rolf, J.; Siedentop, R.; Lütz, S.; Rosenthal, K. *Int. J. Mol. Sci.* **2020,** *21,* 105. doi:10.3390/ijms21010105
19. Mullineaux, C. W.; Nenninger, A.; Ray, N.; Robinson, C. *J. Bacteriol.* **2006,** *188,* 3442–3448. doi:10.1128/jb.188.10.3442-3448.2006
20. Luby-Phelps, K. *Int. Rev. Cytol.* **2000,** *192,* 189–221. doi:10.1016/s0074-7696(08)60527-6
21. Lemmon, E., Huber, M.; McLinden, M. NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 9.1, Natl Std. Ref. Data Series (NIST NSRDS), National Institute of Standards and Technology, Gaithersburg, MD, 2013 [online]. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=912382 (accessed June 12, 2024).
22. Syal, V. K.; Chauhan, A.; Chauhan, S. *J. Pure Appl. Ultrason.* **2005,** *27,* 61–69.

23. González-Tello, P.; Camacho, F.; Blázquez, G. *J. Chem. Eng. Data* **1994,** *39,* 611–614. doi:10.1021/je00015a050
24. Pijper, A. *J. Bacteriol.* **1947,** *53,* 257–269. doi:10.1128/jb.53.3.257-269.1947
25. Verma, P. K.; Kundu, A.; Ha, J.-H.; Cho, M. *J. Am. Chem. Soc.* **2016,** *138,* 16081–16088. doi:10.1021/jacs.6b10164
26. Elias, H. G. *Makromoleküle;* Georg Thieme Verlag: Stuttgart, Germany, 2009. doi:10.1002/9783527626472
27. Garamella, J.; Marshall, R.; Rustad, M.; Noireaux, V. *ACS Synth. Biol.* **2016,** *5,* 344–355. doi:10.1021/acssynbio.5b00296
28. Stefanovic, R.; Ludwig, M.; Webber, G. B.; Atkin, R.; Page, A. J. *Phys. Chem. Chem. Phys.* **2017,** *19,* 3297–3306. doi:10.1039/c6cp07932f
29. Kučan, K. Z.; Perković, M.; Cmrk, K.; Načinović, D.; Rogošić, M. *ChemistrySelect* **2018,** *3,* 12582–12590. doi:10.1002/slct.201803251
30. Mulia, K.; Adam, D.; Zahrina, I.; Krisanti, E. *Int. J. Technol.* **2018,** *2,* 335–344. doi:10.14716/ijtech.v9i2.1008
31. Rodrigues, L. A.; Cardeira, M.; Leonardo, I. C.; Gaspar, F. B.; Radojčić Redovniković, I.; Duarte, A. R. C.; Paiva, A.; Matias, A. A. *J. Mol. Liq.* **2021,** *335,* 116201. doi:10.1016/j.molliq.2021.116201
32. National Center for Biotechnology Information. PubChem Compound Summary for CID 679, dimethyl sulfoxide.
33. National Center for Biotechnology Information. PubChem Compound Summary for CID 15413, methyl *tert*-butyl ether.
34. Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung. Eintrag zu Dimethylsulfoxid in der GESTIS-Stoffdatenbank.
35. National Center for Biotechnology Information. PubChem Compound Summary for CID 887, methanol.
36. Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung. Eintrag zu methyl *tert*-butylether in der GESTIS-Stoffdatenbank.
37. National Center for Biotechnology Information. PubChem Compound Summary for CID 8058, *n*-hexane.
38. Lide, D. R. *CRC Handbook of Chemistry and Physics;* CRC Press: Boca Raton, FL, USA, 2004.
39. Landaverde-Cortes, D. C.; Estrada-Baltazar, A.; Iglesias-Silva, G. A.; Hall, K. R. *J. Chem. Eng. Data* **2007,** *52,* 1226–1232. doi:10.1021/je600554h
40. Badyal, Y. S.; Saboungi, M.-L.; Price, D. L.; Shastri, S. D.; Haeffner, D. R.; Soper, A. K. *J. Chem. Phys.* **2000,** *112,* 9206–9208. doi:10.1063/1.481541
41. Nelson, R. D., Jr.; Lide, D. R., Jr.; Maryott, A. A. *Selected Values of Electric Dipole Moments for Molecules in the Gas Phase;* NSRDS–NBS 10 National Standard References Data Series – National Bureau of Standards 10, Category 3 – Atomic and Molecular Properties; U.S. Department of Commerce National Bureau of Standards: Washington, DC, USA, 1967. doi:10.6028/nbs.nsrds.10
42. Jorge, M.; Gomes, J. R. B.; Barrera, M. C. *J. Mol. Liq.* **2022,** *356,* 119033. doi:10.1016/j.molliq.2022.119033
43. Suenram, R. D.; Lovas, F. J.; Pereyra, W.; Fraser, G. T.; Walker, A. R. H. *J. Mol. Spectrosc.* **1997,** *181,* 67–77. doi:10.1006/jmsp.1996.7155
44. Cayman Chemicals. CommonSolventsNonpolar. https://cdn2.caymanchem.com/cdn/cms/caymanchem/cmsImages/news/502/CommonSolventsNonpolar.pdf (accessed Jan 17, 2024).
45. Pandi, A.; Diehl, C.; Yazdizadeh Kharrazi, A.; Scholz, S. A.; Bobkova, E.; Faure, L.; Nattermann, M.; Adam, D.; Chapin, N.; Foroughijabbari, Y.; Moritz, C.; Paczia, N.; Cortina, N. S.; Faulon, J.-L.; Erb, T. J. *Nat. Commun.* **2022,** *13,* 3876. doi:10.1038/s41467-022-31245-z
46. Kahn, T. W.; Beachy, R. N.; Falk, M. M. *Curr. Biol.* **1997,** *7,* R207–R208. doi:10.1016/s0960-9822(06)00100-x
47. Kim, D.-M.; Kigawa, T.; Choi, C.-Y.; Yokoyama, S. *Eur. J. Biochem.* **1996,** *239,* 881–886. doi:10.1111/j.1432-1033.1996.0881u.x
48. Ge, X.; Luo, D.; Xu, J. *PLoS One* **2011,** *6,* e28707. doi:10.1371/journal.pone.0028707
49. Li, J.; Jiang, Z.; Wu, H.; Liang, Y.; Zhang, Y.; Liu, J. *Carbohydr. Polym.* **2010,** *82,* 160–166. doi:10.1016/j.carbpol.2010.04.045
50. Chai, Q.; Singh, B.; Peisker, K.; Metzendorf, N.; Ge, X.; Dasgupta, S.; Sanyal, S. *J. Biol. Chem.* **2014,** *289,* 11342–11352. doi:10.1074/jbc.m114.557348
51. Śmigiel, W. M.; Mantovanelli, L.; Linnik, D. S.; Punter, M.; Silberberg, J.; Xiang, L.; Xu, K.; Poolman, B. *Sci. Adv.* **2022,** *8,* eabo5387. doi:10.1126/sciadv.abo5387
52. Ribeiro, B. D.; de Carvalho Iff, L.; Coelho, M. A. Z.; Marrucho, I. M. *Curr. Biochem. Eng.* **2019,** *5,* 57–68. doi:10.2174/2212711906666190710181629
53. Yadav, A.; Pandey, S. *J. Chem. Eng. Data* **2014,** *59,* 2221–2229. doi:10.1021/je5001796
54. Shih, Y.-C.; Prausnitz, J. M.; Blanch, H. W. *Biotechnol. Bioeng.* **1992,** *40,* 1155–1164. doi:10.1002/bit.260401004
55. Kress, C.; Sadowski, G.; Brandenbusch, C. *Eur. J. Pharm. Biopharm.* **2017,** *112,* 38–44. doi:10.1016/j.ejpb.2016.11.016
56. Gregorio, N. E.; Levine, M. Z.; Oza, J. P. *Methods Protoc.* **2019,** *2,* 24. doi:10.3390/mps2010024
57. Goering, A. W.; Li, J.; McClure, R. A.; Thomson, R. J.; Jewett, M. C.; Kelleher, N. L. *ACS Synth. Biol.* **2017,** *6,* 39–44. doi:10.1021/acssynbio.6b00160
58. Ablasser, A.; Goldeck, M.; Cavlar, T.; Deimling, T.; Witte, G.; Röhl, I.; Hopfner, K.-P.; Ludwig, J.; Hornung, V. *Nature* **2013,** *498,* 380–384. doi:10.1038/nature12306
59. Bartsch, T.; Becker, M.; Rolf, J.; Rosenthal, K.; Lütz, S. *Biotechnol. Bioeng.* **2022,** *119,* 677–684. doi:10.1002/bit.28027
60. Pätzold, M.; Siebenhaller, S.; Kara, S.; Liese, A.; Syldatk, C.; Holtmann, D. *Trends Biotechnol.* **2019,** *37,* 943–959. doi:10.1016/j.tibtech.2019.03.007
61. Zhang, Y.; Minagawa, Y.; Kizoe, H.; Miyazaki, K.; Iino, R.; Ueno, H.; Tabata, K. V.; Shimane, Y.; Noji, H. *Sci. Adv.* **2019,** *5,* eaav8185. doi:10.1126/sciadv.aav8185
62. Ribeiro, A. L. J. L.; Pérez-Arnaiz, P.; Sánchez-Costa, M.; Pérez, L.; Almendros, M.; van Vliet, L.; Gielen, F.; Lim, J.; Charnock, S.; Hollfelder, F.; González-Pastor, J. E.; Berenguer, J.; Hidalgo, A. *Microb. Cell Fact.* **2024,** *23,* 169. doi:10.1186/s12934-024-02440-y
63. Adden, R.; Hübner-Keese, B.; Förtsch, S.; Knarr, M. Cellulosics. In *Handbook of Hydrocolloids;* Phillips, G. O.; Williams, P. A., Eds.; Woodhead Publishing: Duxford, UK, 2021; pp 481–508. doi:10.1016/b978-0-12-820104-6.00011-5
64. Grunberg, L.; Nissan, A. H. *Nature* **1949,** *164,* 799–800. doi:10.1038/164799b0
65. Volkmer, B.; Heinemann, M. *PLoS One* **2011,** *6,* e23126. doi:10.1371/journal.pone.0023126
66. Bradford, M. M. *Anal. Biochem.* **1976,** *72,* 248–254. doi:10.1016/0003-2697(76)90527-3
67. Laemmli, U. K. *Nature* **1970,** *227,* 680–685. doi:10.1038/227680a0

68. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. *Nat. Methods* **2012,** *9,* 671–675. doi:10.1038/nmeth.2089

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.192

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

# Improved deconvolution of natural products' protein targets using diagnostic ions from chemical proteomics linkers

Andreas Wiest and Pavel Kielkowski*

Review

Open Access

Address:
LMU Munich, Department of Chemistry, Butenandtstr. 5-13, 81377 Munich, Germany

Email:
Pavel Kielkowski* - pavel.kielkowski@cup.lmu.de

* Corresponding author

Keywords:
chemical proteomics; diagnostic ions; mass spectrometry; target identification

*Beilstein J. Org. Chem.* **2024,** *20,* 2323–2341.
https://doi.org/10.3762/bjoc.20.199

Received: 05 April 2024
Accepted: 27 August 2024
Published: 12 September 2024

This article is part of the thematic issue "Young investigators in natural products chemistry, biosynthesis, and enzymology".

Guest Editor: L. Barra

© 2024 Wiest and Kielkowski; licensee Beilstein-Institut.
License and terms: see end of document.

## Abstract

Identification of interactions between proteins and natural products or similar active small molecules is crucial for understanding of their mechanism of action on a molecular level. To search elusive, often labile, and low-abundant conjugates between proteins and active compounds, chemical proteomics introduces a feasible strategy that allows to enrich and detect these conjugates. Recent advances in mass spectrometry techniques and search algorithms provide unprecedented depth of proteome coverage and the possibility to detect desired modified peptides with high sensitivity. The chemical 'linker' connecting an active compound–protein conjugate with a detection tag is the critical component of all chemical proteomic workflows. In this review, we discuss the properties and applications of different chemical proteomics linkers with special focus on their fragmentation releasing diagnostic ions and how these may improve the confidence in identified active compound–peptide conjugates. The application of advanced search options improves the identification rates and may help to identify otherwise difficult to find interactions between active compounds and proteins, which may result from unperturbed conditions, and thus are of high physiological relevance.

## Introduction

Natural products (NPs) have been pivotal for the development of modern medicine accompanying humans from the prehistorical age [1]. In the last century, NPs became the main source of the active scaffolds for the pharmaceutical industry with focus on principle of one compound, one target, and one disease [2,3]. In parallel, mass spectrometry (MS) has been crucial in many areas centered around the characterization of NPs [4]. First to annotate their often complex structures using diverse fragmentation techniques [4]. Nowadays, MS is applied for the identification of NPs' cellular protein targets via MS-based chemical

proteomics [5,6]. In proteomics and its sub-field of chemical proteomics, there are two major areas of development, mass spectrometry instrumentation and bioinformatics. The MS instruments acquire the mass spectra with significantly increased throughput and sensitivity. Acquisition techniques benefit from possible multistage ion separation and isolation, and diverse fragmentation techniques. All of that is nowadays combined in one instrument to provide greater flexibility of the spectra acquisition setup [7-16]. Given the sheer amount of data acquired by LC–MS/MS runs and the high density of information content in the MS spectra, another area that experienced progressive development includes search algorithms, which aim to maximize the information obtained from the acquired mass spectra and hence to improve rate and confidence of hit identifications [17-22].

All parts of the chemical proteomics experimental procedure, the sample preparation together with MS measurement and evaluation can be performed in a streamlined or automated way using sophisticated pipelines [13,17,21,25-27]. These are highly optimized to provide the desired protein coverage and robustness. Still, the critical part of the process starts with the careful design of a NP analogue and selection of the proper biological system, which may include animal models, 3D tissue-organoids, or 2D cell culture. Depending on the project aim, which typically might be to obtain the pathologically relevant protein targets or off-targets, this may further lead to optimization of the NP scaffold to improve the desired activity. In this review, we focus on the most frequently used chemical proteomic strategies that necessitate a covalent bond between a target/off-target protein and a tested NP or a similar active small compound (Figure 1). The covalent bond serving this purpose can be formed in two distinct ways: either the NP or small compound already contains a reactive, often electrophilic group directly or the reactive group, for example a photo-crosslinker, is attached synthetically to the basic scaffold (Figure 2A) [28-32]. However, this requires a structure–activity relationship (SAR) screening to ensure that the selected NP retains biological activity after

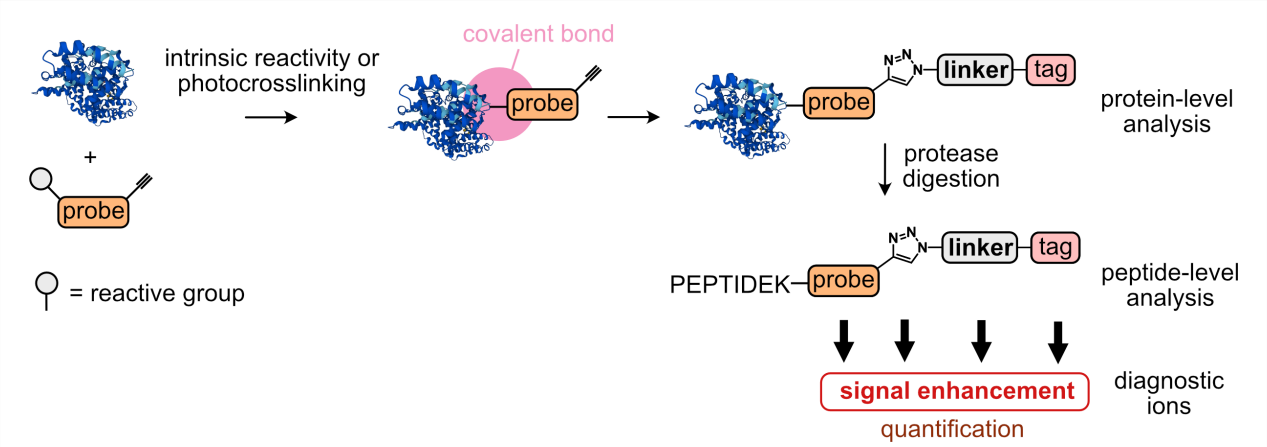



**Figure 1:** Overall chemical proteomics strategy to identify protein targets of natural products (NPs) and similar active small compounds. The example protein (blue) is an AlphaFold v2.0-generated prediction of bovine serum albumin (BSA) [23,24].

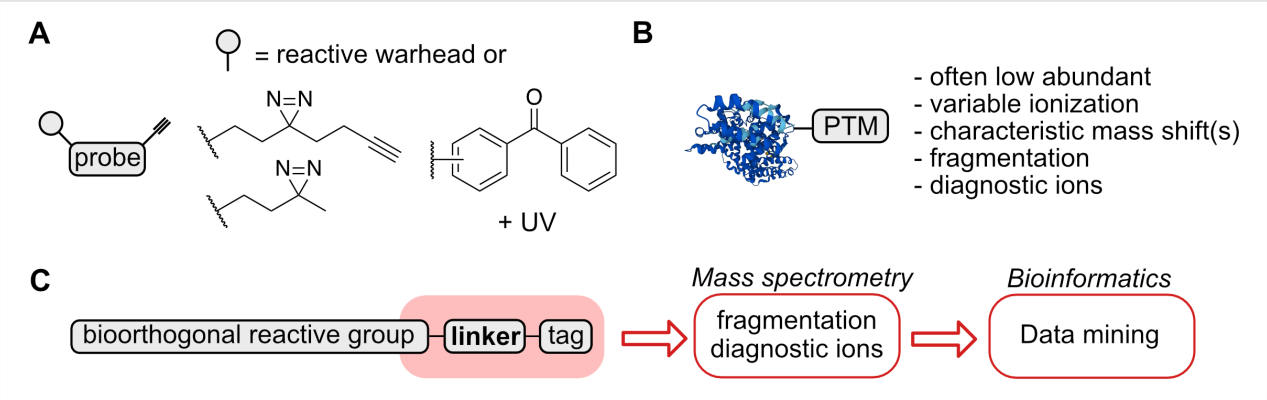



**Figure 2:** A) Design of mostly used photo-crosslinking groups. B) Mass spectrometry properties of proteins PTMs, which relate to probe–protein conjugates. C) General concept utilizing linkers to improve searches for protein–probe interactions.

modification. The term probe is further used to unite such NPs and active compounds. Furthermore, from this perspective, the formed probe–protein covalent conjugate shares many features with naturally occurring protein post-translational modifications (PTMs) offering productive exchange of the techniques and experimental setups used in both fields (Figure 2B) [33,34]. The second prerequisite required for a probe used in a chemical proteomic study is an embedded bioorthogonal handle, for example a terminal alkyne or azide, which is able to react chemoselectively with a tag facilitating unambiguous identification by a selected analytical technique, for example LC–MS/MS (Figure 1) [35-37]. A covalent bond between a probe and a protein is needed because of denaturing conditions used during the analysis. In contrast, relying on a noncovalent interaction would result in loss of the probe–protein interaction. The bioorthogonal handle is therefore essential for an enrichment or distinction of the probe–protein complex from the background. There are two levels on which the probe conjugates can be identified: either on the protein level as intact probe–protein conjugates or on the peptide level after protease digestion as a probe–peptide conjugate (Figure 1). Although the increased speed of acquisition of mass spectra leads to ever better coverage of proteomes submitted for analysis, the throughput has not yet reached the point in which all peptides and their modified forms can be identified and characterized in a single experiment. Therefore, an enrichment of the probe-modified peptides or enhancement of the signal-to-noise ratio is necessary for successful analysis (Figure 1). The future challenge to complete the chemical proteomic analysis will include the absolute quantification of those identified NP–peptide conjugates to estimate protein occupancy (Figure 1). Although mass spectrometry-based chemical proteomics represents a robust method for the identification of NP protein targets, in some cases, the synthetic effort associated with the above-described SAR study to find a functional NP probe might outweigh the advantages of this approach. Considering a complementary strategy that does not necessitate for NP modification might be desirable. This may include hydrogen–deuterium exchange mass spectrometry (HDX-MS), limited proteolysis-coupled mass spectrometry (LiP–MS), thermal proteome profiling (TPP), cellular thermal shift assay (CETSA), affinity selection-mass spectrometry (AS-MS), co-fractionation mass spectrometry (CF-MS), or mass spectrometry integrated with equilibrium dialysis for the discovery of allostery systematically (MIDAS) [38-44]. The chemical proteomics approach and related activity-based protein profiling has been reviewed comprehensively in several reviews [45-48]. Applications of cleavable linkers, which might be considered as pioneers of MS fragmentation-based platforms to search peptide-linker remainders, were reviewed thoroughly as well [49-51]. Recent advances in mass spectrometers technology and analysis pipelines enable to obtain more information from already established workflows, because they are becoming practically feasible only now. These will be discussed further in this review. We primarily focus on the linkers utilized for enrichment and identification of the probe–protein conjugates resembling probe–protein interaction in native conditions in living cells. Taken together, the linkers used in chemical proteomics workflows are applied broadly in many research areas including screening of fragment libraries, covalent modifiers, protein PTMs, protein–protein interactions, protein–nucleic acid interactions, and indeed NP–protein interactions, but they share many common features, which are of interest of this review (Figure 2C) [5,52]. These chemical linkers bridge an imaginative gap between the crude, highly complex, and structurally diverse cell or tissue lysates containing unknown probe–protein conjugate(s) and the desired analytical output with the information about the protein and conjugate identity. Thus, the linkers have not only the function to simply connect the probe scaffold with an affinity or reporter tag, but they carry previously underappreciated features enhancing the capabilities of nowadays mass spectrometric hardware and bioinformatic software tools.

# Review

## Principles of chemical proteomics workflows

The linker connects the affinity or reporter tag to the probe–protein conjugate to offer several principally different analytical pathways (Figure 3) [53,54]. Although chemical proteomics protocols may have many steps and operations in common, there is a great variability in how these individual steps are arranged into a workflow (Figure 3). The important aspect of any chemical proteomics study is the efficient coupling between the probe–protein conjugate and the selected linker bearing an affinity or reporter tag. To carry out this bioorthogonal reaction well-known chemistries were developed including traceless Staudinger ligation, Cu-catalyzed azide–alkyne cycloaddition (CuAAC), strain-promoted azide–alkyne cycloaddition (SPAAC), inverse electron-demand Diels–Alder reaction (IEDDA), and recently, azomethine imine (AMIs)–isonitrile ligation [37,55-60]. The kinetics, chemoselectivity, stability, and steric demand of the bioorthogonal tag attached on the probe are decisive factors during the selection procedure [61,62]. The most commonly used strategy is CuAAC due to its rapid reaction kinetics, robustness, and relatively small steric hindrance of the terminal alkyne, which is usually attached to the probe core scaffold to form an alkyne probe [5,63]. Once the covalent bond between the probe and protein is formed, the cells are lysed, and the probe–protein conjugates are reacted via CuAAC with an azide tag. While the CuAAC has been employed in various studies and has had large impact on many biological discoveries, an unspecific reactivity was often reported [54,64]. Recently, we described this background-forming reac-

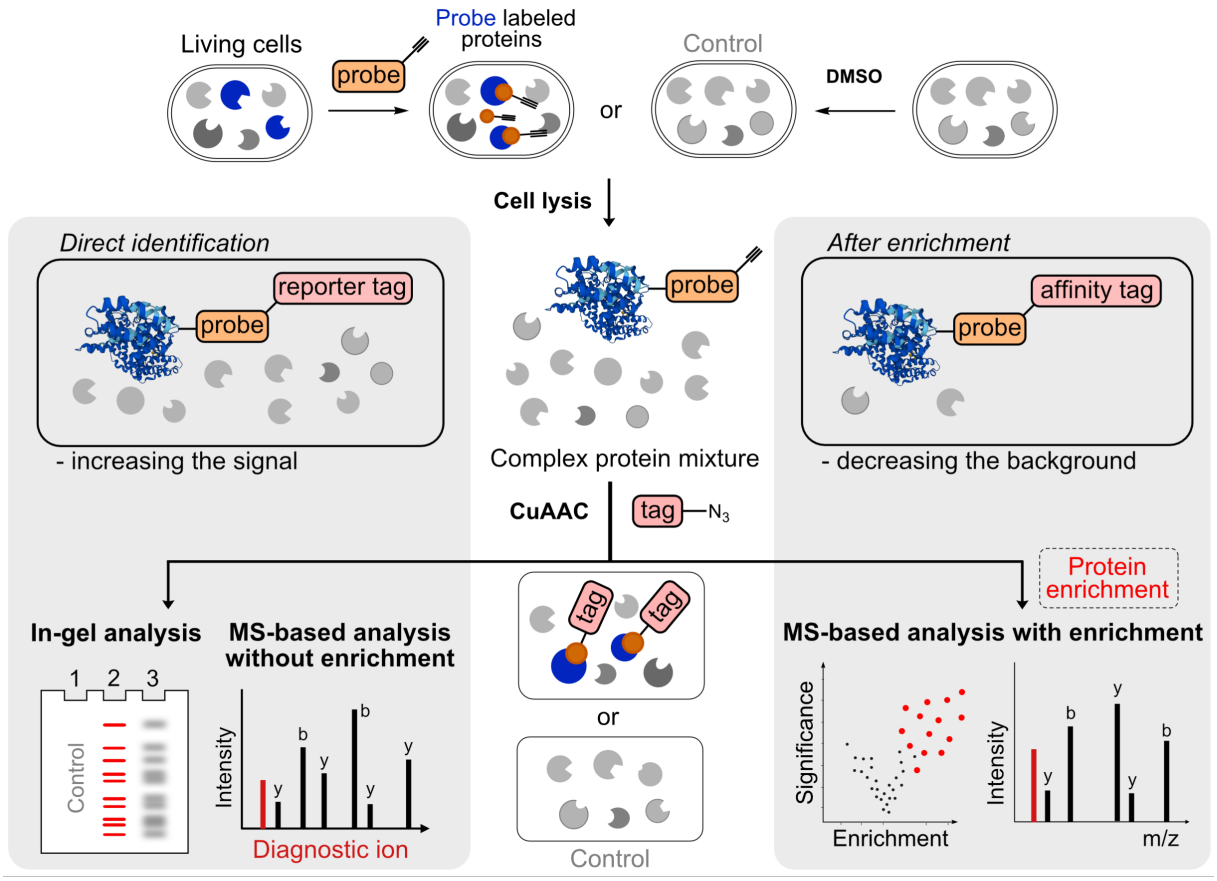



**Figure 3:** Direct and indirect approach to identify protein targets and representative chemical proteomics workflow visualizing the main steps; cells incubation with the probe, bioorthogonal reaction between probe–protein conjugates and a tag followed by analysis of those labeled proteins.

tion, which is based on an interrupted CuAAC mechanism [65]. The thiotriazole product of this reaction, which is indistinguishable in the protein-level downstream analysis, is formed by coupling between protein free thiol groups and the triazole–copper adduct (Figure 4). However, its formation can be avoided by eliminating the free alkyne probe from the CuAAC reaction mixture or by increasing the concentration of the reducing agent such as tris(2-carboxyethyl)phosphine (TCEP). The reversed chemical proteomics approach in which the azide probe and alkyne tag are used suffers from similar unspecific reactivity of the terminal alkyne with free thiols, including the protein cysteines reacting to thioalkyne conjugates [66,67]. This is caused by the relatively high concentration of the alkyne tag which is used to accelerate the reaction.

The chemical proteomics workflows can be differentiated based on their read out, which can be done on the protein or peptide level (Figure 1). The protein-level identification leads to determination of protein(s) interacting with the probe. This might be the more straightforward approach as it requires smaller

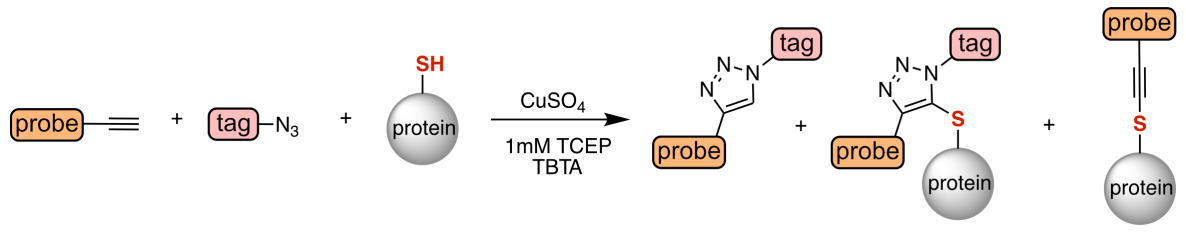



**Figure 4:** Products of the CuAAC side reactions.

amounts of protein to begin with and typically the protein identification relies on standard and well-established peptide fragmentation used in the standard MS-based proteomics. Furthermore, the advantage of significantly improved data completeness using the data-independent acquisition (DIA), which in comparison to the previously applied data-dependent acquisition (DDA), achieves a higher number of protein identifications in relatively shorter time (Figure 5A) [68,69]. The mass spectra can be searched by standard proteomics search engines incorporated into freely available tools such as FragPipe platform, MaxQuant, and DIA-NN [17,18,20,70]. However, the probe-modified peptide and amino acid remain unknown. In contrast, the peptide-level identification is focused on the characterization of the actual probe-modified peptides. It might be more challenging as a larger protein amount to start with is necessary and the fragments of the probe-modified peptides may not be trivial to elucidate. The site identification is then reconstructed based on multiple rounds of searches utilizing different setups which may include labile searches with diagnostic ion mining, open and closed searches (Figure 5B–D) [19,71-74]. Exemplarily, the FragPipe workflows describing how to set up labile, open and closed searches and how we recalculated the published MS data for different linkers in this review can be downloaded from the publicly available database (PRIDE data set identifier PXD043402) [65]. Furthermore, the data processing can be adjusted and paralleled accordingly using orthogonal fragmentation techniques [75,76]. Because of the often low number of identified peptides bearing a desired modification, it is difficult to control the false-discovery rate (FDR) of the hits, and thus, the introduction of proper negative controls is of the upmost importance.

Thus far, the linkers were proposed and designed to mainly facilitate the enrichment of the probe–protein or probe–peptide conjugates and their selective release after enrichment. In this review we would like to highlight the fragmentation properties of the different linkers and resulting ions as the valuable source of information steering the site identifications and increasing the confidence in those identifications in MS-based chemical proteomics.

## In-gel chemical proteomics

The detection of the probe–protein conjugate(s) is based on the enhanced signal stemming from the probe–protein conjugate above the background from total protein present in the lysate. The detection can be done in several different ways. To obtain an overview of how many proteins are labeled in total and what is the intensity of the labeling, the probe–protein–tag conju-

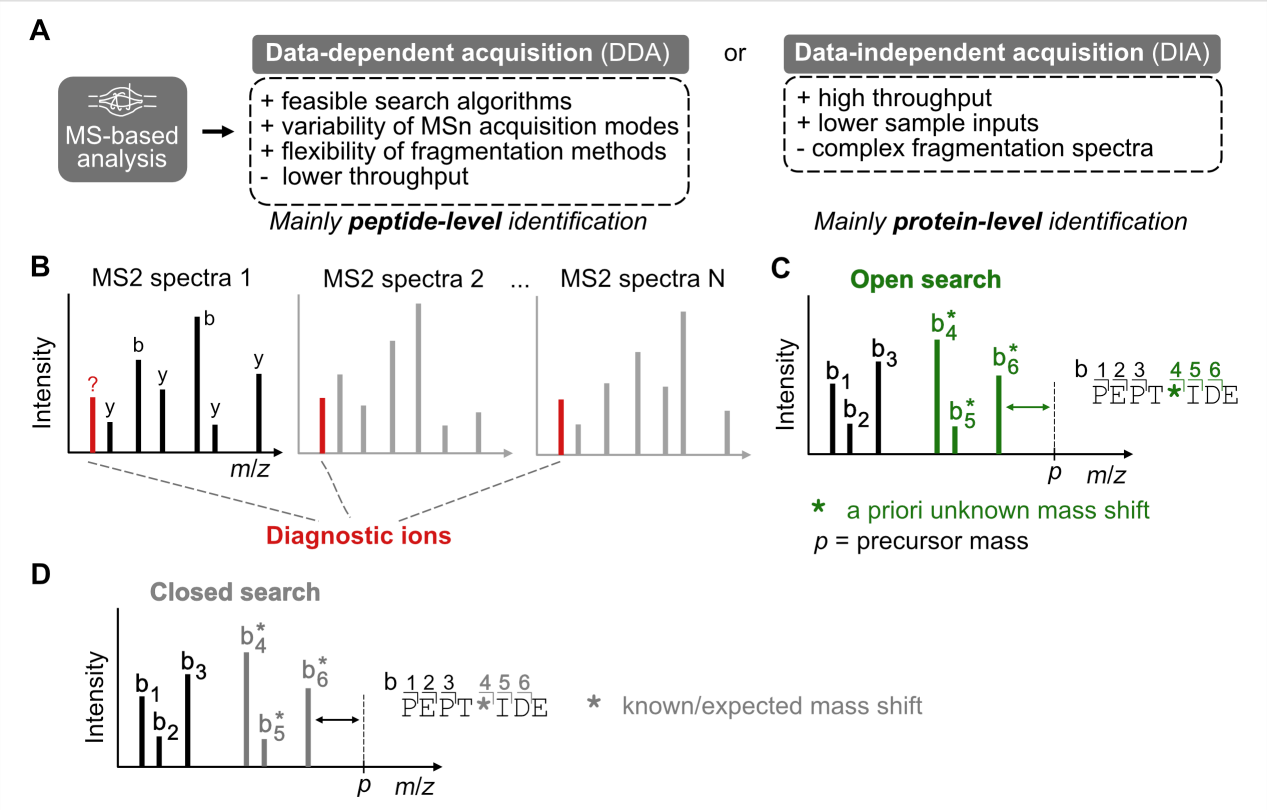



**Figure 5:** Search possibilities on peptide-level characterization. A) Comparison of DDA and DIA techniques. B) Diagnostic ions mining. C) Open search. D) Closed search.

gates are separated on sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) and visualized by in-gel fluorescence scanning (Figure 6A). In this approach the tag is a suitable fluorophore. Previously, during an introduction period of a chemical proteomics workflow of activity-based protein profiling (ABPP), the resulting fluorescent bands might be simply cut out from the gel and subjected to MS analysis to identify the labeled proteins [77,78]. The in-gel analysis is often used for target validation with corresponding protein mutants or knockout cell lines, with loss of the identified probe–protein interaction leading to disappearance of the fluorescence band from the SDS-PAGE. This approach also accounts for the protein-level identification; the exact modification site remains unknown. Here we show an example of the most used and commercially available rhodamine-derived dye (TAMRA-azide, Figure 6A). The bioorthogonal reaction between probe-labeled proteins and the TAMRA-azide yields a mixture of labeled and unlabeled proteins. The proteins are then separated on SDS-PAGE by their molecular weight and fluorescence of the labeled proteins is recorded (Figure 6A). Due to the high sensitivity of fluorescence read-out, this allows to identify protein labeling occurring already at sub-stoichiometric ratios. However, by in-gel analysis, it is not possible to quantify the extent of labeled proteins. To ensure the fidelity of the bioorthogonal reaction a control experiment in which the probe treatment is replaced by a plain solvent used to prepare the stock solution of the probe is necessary to identify any unspecific bands stemming from side reactivity of the azide–protein conjugate. In case a photo-crosslinker is used, another control experiment needs to be included, in which the cells/lysate are treated with the probe but not UV-irradiated. The non-covalent probe–protein interactions are not detected because of the denaturing conditions of the SDS-PAGE. Alternatively, the fidelity of the probe labeling can be tested by a competition experiment, in which the parent compound lacking a bioorthogonal tag is used at increasing concentration to outcompete the probe–protein interaction leading to disappearance of the fluorescent bands.

A gel shift assay is a complementary strategy to validate the labeling of a specific protein (Figure 6B) [65,79]. The assay is based on an increased retardation of the probe-labeled protein on SDS-PAGE due to the addition of a large tag, which significantly changes the molecular weight of the given protein. The shift on the gel can be then visualized by protein-specific antibody or coomassie staining. The linkers applied to perform the gel shift assay contain a linear PEG polymer with an approximate size ranging from one to dozens of kilodaltons. The main advantage of the gel shift assay is to obtain quantitative information of the protein's probe modification extent, considering high efficiency of the click reaction. Given the bulkiness of the PEG–azide linker, the bioorthogonal reaction needs some optimization, including increased equivalents of the linker, elevated temperature, and/or longer reaction time. The limiting factor of this approach might be the size of the target proteins, because for larger proteins it might be more challenging to obtain the desired significant difference between mobility of probe-modified and unmodified protein.

## MS-based chemical proteomics without enrichment

The aim of this approach is to isolate and characterize the probe–peptide conjugate and hence to provide direct evidence of the probe–protein interaction [80]. In principle, the direct characterization of the probe–peptide conjugate can be attempted, in which the coupling of the probe–peptide conjugate with a suitable reporter tag may help to improve the detectability in an excess of probe-unmodified peptides [65,81,82]. The speed of acquisition of mass spectra of the cur-

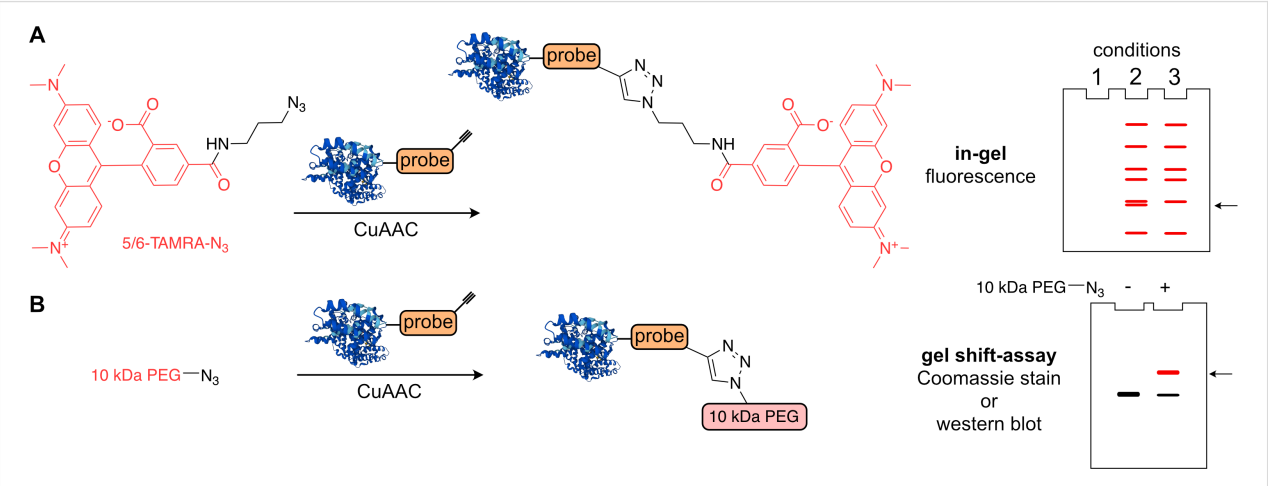



**Figure 6:** In-gel analysis using a tag with fluorophore (A) or via shift-assay (B).

rent LC–MS/MS methods does not ultimately cover all unmodified and modified peptides resulting from the proteolytic digest of a whole proteome. In other words, the complexity of the whole proteome digest is beyond the coverage and sensitivity of today's mass spectrometers. On the other hand, the enrichment of the probe–peptide conjugates necessitates larger amounts of starting protein and application of often time-demanding and tedious workflows. Thus, the chemical tools increasing the signal intensity of the modified peptide species over the majority of background signals are of utmost interest and may bridge the gap between an ultimate characterization of complete proteome and enrichment-based methods. There are several chemical proteomics linkers reported that improve ionization of the probe–peptide conjugates or release a reporter ion during fragmentation or both (Figure 7). The reporter ion is important as it may serve two purposes. First, it facilitates the

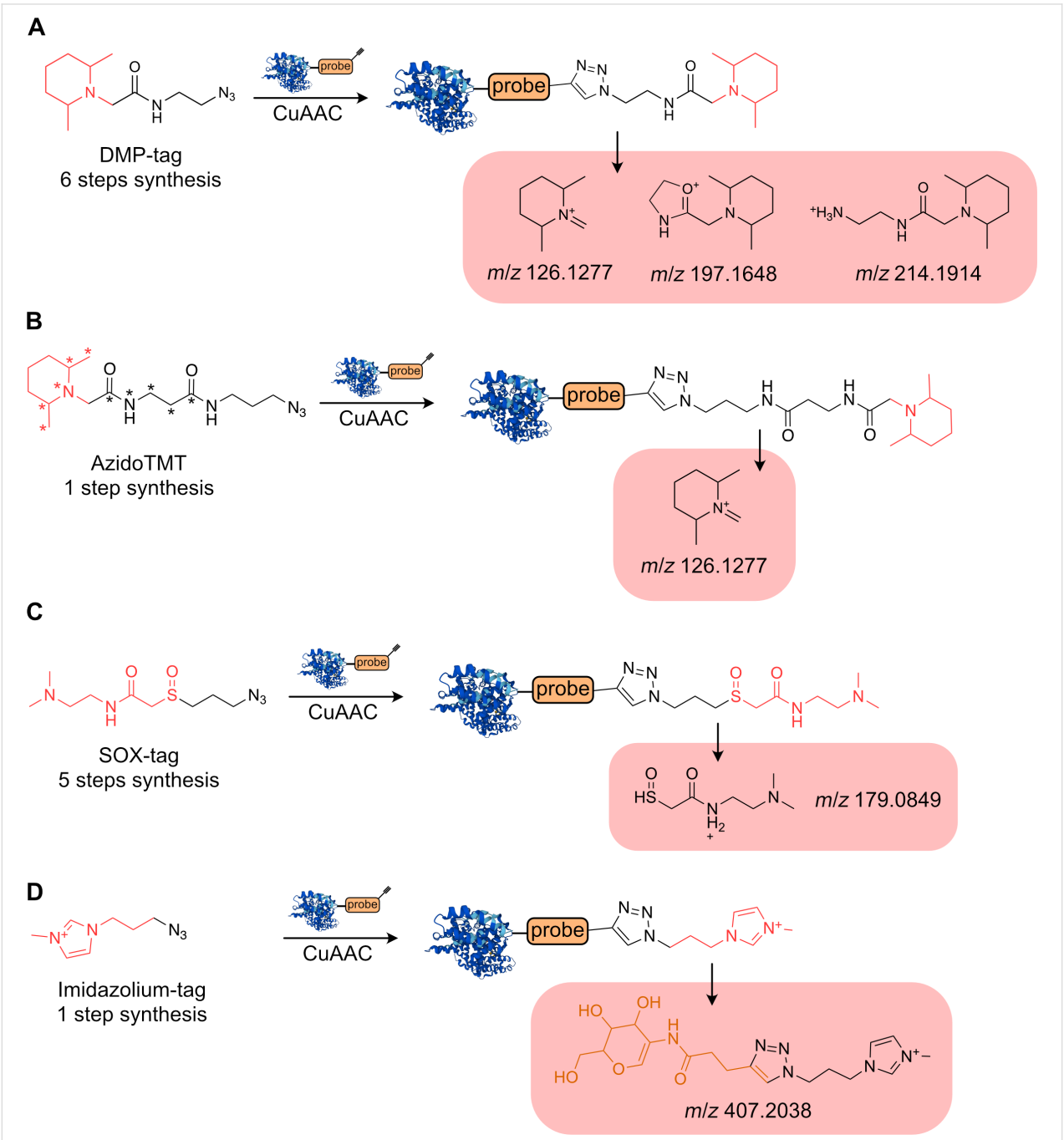



**Figure 7:** Reporter linkers. A) DMP-tag. B) AzidoTMT tag. C) SOX-tag. D) Imidazolium tag. *A star indicates the possible introduction of stable isotopes for MS quantification.

better identification of the modified peptides by search algorithms and second, if labeled with stable isotopes, it allows for quantification of the modified probe–peptide conjugates. Inspired by tandem mass spectrometry tags (TMT-tags) Makarov et al. have developed a 'clickable' 2,6-dimethylpiperidine analogue, the DMP-tag, which in contrast to standardly used isotopic tags bearing the *N*-hydroxysuccinimide ester (NHS-ester) reacting with all primary amines of lysine side-chains, allows to selectively label the probe–protein conjugates (Figure 7A) [81,83]. The DMP-tag can be prepared in a six-steps synthesis. The higher-energy collisional dissociation (HCD) releases the characteristic reporter ion at *m*/*z* 126.1277. In case the electron-transfer dissociation (ETD) is used the DMP-tag yields the reporter ion at *m*/*z* 114.1275. Of note, although ETD is more selective towards fragmentation of the peptide backbone leaving the side-chain modifications largely intact, the generation of the fluoranthene anion used for fragmentation slows down the fragmentation process and hence decreases the total amount of fragmented peptides and throughput of the LC–MS/MS measurement. To balance the throughput and fragmentation performance, the combination of HCD and ETD can be used in a single MS measurement. In such methods ETD is only triggered by probe-specific diagnostic ions released by the faster HCD fragmentation with the aim to fragment the same parent ion again with ETD to obtain a better fragmentation spectrum [84,85]. Similar to the DMP-tag, the AzidoTMT introduced by Ma et al. contains an azide to selectively label probe–peptide conjugates, but the synthesis takes advantage of the commercially available isotopically labeled 11plex-TMT (Figure 7B) [82]. The 11plexTMT reagent is reacted in one step with 3-azido-1-propanamine to give the desired AzidoTMT linker. Furthermore, Makarov et al. developed a linker containing a sulfoxide (SOX) group, which is well-known for specific fragmentation releasing the reporter ion at *m*/*z* 179.0849 (Figure 7C) [81,86,87]. The preparation of the tag is based on a five-step synthesis. However, in comparison to the DMP-tag or AzidoTMT the release of the reporter ion from the SOX-tag has somewhat lower intensity.

Calle et al. designed, synthesized, and validated a series of 'clickable' linkers for characterization of protein *O*-glycosylation containing positively charged pyridinium or imidazolium residues (Figure 7D) [88]. The presence of a stable positively charged moiety in the linker has, in case of imidazolium, increased by 70–90% the relative abundance of modified peptides. In combination with release of the glycan-related oxonium ions during fragmentation, such approach may improve the coverage of modified peptides. Taken together, the ever faster and more sensitive MS instruments in recent years drastically improved the proteome coverage of the LC–MS/MS methods to provide great insight into proteomes and enabled direct characterization of abundant modified proteins. The linkers releasing the reporter ion may serve as an interesting tool during transition from enrichment-based methods to direct 'linker-free' identification MS methods of desired probe–peptide conjugates.

## Enrichment-based chemical proteomics

The pursuit to decrease the complexity of proteomic samples led to development of numerous enrichment strategies using a variety of linkers as outlined below. The enrichment of the probe–protein or probe–peptide conjugates reduces the number of total peptides to be resolved by LC–MS/MS. The most common strategy to identify proteins interacting with a probe either directly in living cells or in cell lysates utilizes a linker with biotin (Figure 8A). This approach led for example to the identification of protein targets of the synthetic pyrethrin analogues cerulenin, hypothemicin, fimbrolide, and eupalmerin acetate [89-92]. The biotin forms a tight non-covalent complex with a glycoprotein called avidin or its non-glycosylated form streptavidin, from which the latter shows less of an unspecific protein–protein interaction due to missing glycan groups [93]. Streptavidin-coated paramagnetic agarose or acrylamide beads are used for pull-down of the probe-modified proteins [94]. The biotin-containing linker is introduced into the probe–protein conjugate via a selective bioorthogonal reaction. The main advantage of the biotin-based linkers is the possibility to apply stringent wash steps, due to the almost irreversible non-covalent complex formation between biotin and streptavidin ($K_d \approx 1 \times 10^{-14}$ M) in aqueous solution. Another advantage is that after the enrichment and proteolytic digestion of each of the enriched proteins likely releases several suitable peptides for LC–MS/MS analysis resulting in a reliable identification of protein targets and relative quantification [95-97]. On the protein level, the biotin–azide linker was used to determine many probe–protein targets [92,98-100]. However, the modified probe–peptide conjugates linked with biotin mainly remain in complex with the streptavidin beads, which restricts the site identification [101]. The difference in protein intensities from the probe and control experiments can be visualized using a volcano plot or a waterfall plot and is also referred as protein-level experiment. Next, the elution of a desired biotin–probe–peptide conjugate can be carried out by high percentage of organic solvent, for example 80% acetonitrile. Yan et al. showed that such biotin–probe–peptides afforded characteristic fragmentation of the biotin linker, thus improving the identification of probe–peptide conjugates in cysteine-profiling experiments [102]. The biotin linker yields after HCD fragmentation a set of diagnostic ions including the prominent oxonium-biotin and dehydrobiotin ions (Figure 8B). Importantly, the oxonium-biotin fragment was found in more than 99% of the assigned spectra, so-called peptide spectrum matches (PSMs).

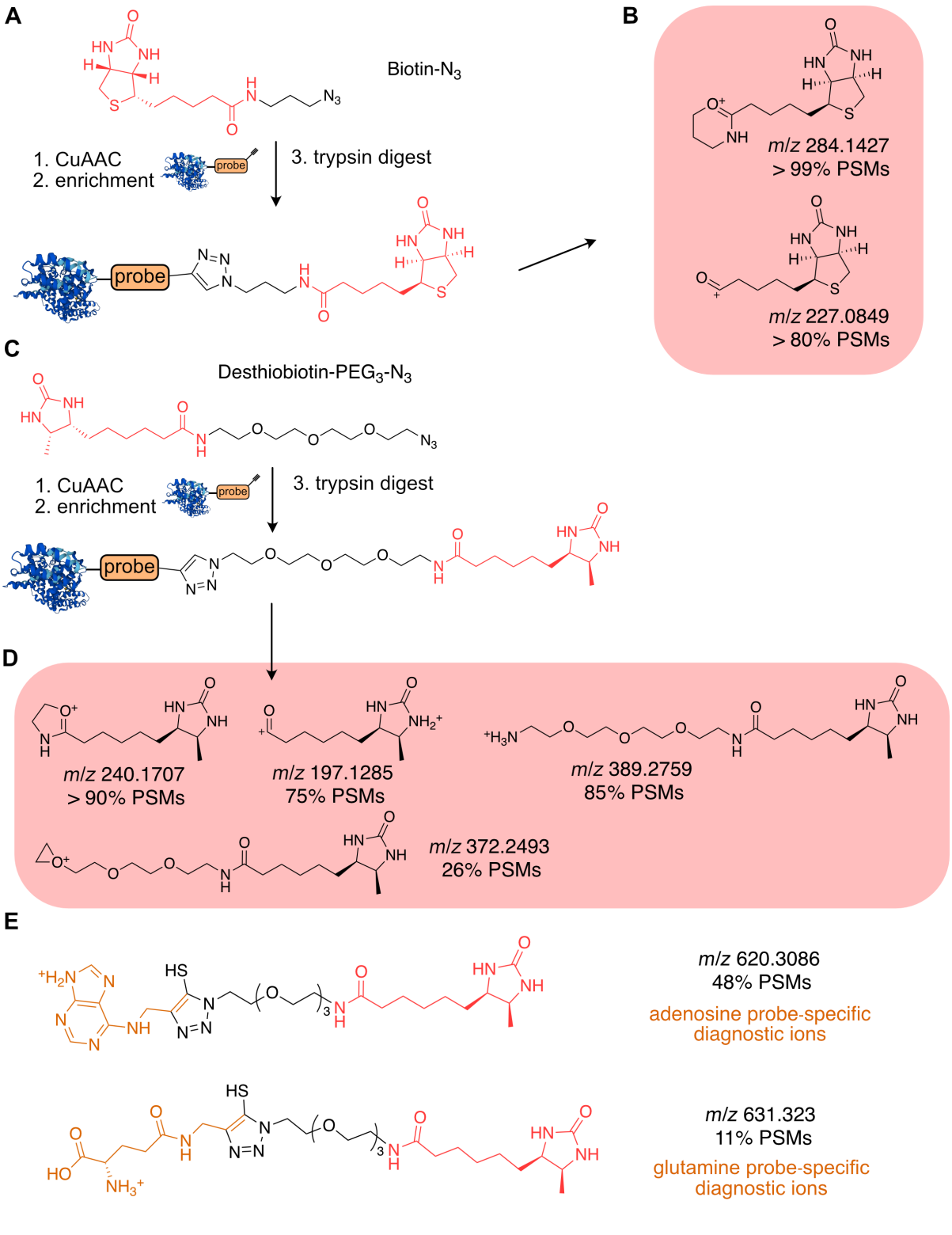



**Figure 8:** Biotin and desthiobition-based sample linkers and their associated diagnostic peaks. A) Structure of the biotin–azide linker and the CuAAC product. B) Diagnostic peaks derived from the biotin–azide linker. C) Structure of the desthiobiotin–PEG–azide linker and the corresponding CuAAC product. D) Diagnostic peaks resulting from fragmentation of the desthiobiotin–PEG–azide linker. E) Probe-specific diagnostic peaks after fragmentation of probe–peptide conjugates enriched using the desthiobiotin–PEG–azide linker.

The second most frequent diagnostic ion was the dehydrobiotin ion with occurrence in >80% of the spectra. The detailed list of all identified biotin and cysteine reactive probe related ions can be found in reference [102]. A plethora of biotin–azide linkers is commercially available, with the main difference lying in the length or presence of the PEG moiety, which has likely only minimal influence on the diagnostic peaks' generation. Taken together, consideration of the oxonium-biotin and dehydrobiotin diagnostic ions increases the confidence in the identified probe–peptide conjugates.

Next, selective elution of the probe–peptide conjugates may be carried out using desthiobiotin (DTB) as affinity tag instead of biotin [103,104]. The DTB-tag allows to elute probe–peptide–DTB conjugates from streptavidin beads using a mixture of water, organic solvent – typically acetonitrile (ACN) – and an acid such as formic acid. The desthiobiotin–azide (DTB–azide) linker still possesses strong binding to streptavidin ($K_d \approx 1 \times 10^{-11}$ M), but it can be released by mild washing conditions with organic solvent and low pH (Figure 8C) [105,106]. Similarly, desthiobiotin releases upon HCD fragmentation a set of characteristic diagnostic ions including oxonium-desthiobiotin, dehydrodesthiobiotin, and several other ions containing the PEG moiety (Figure 8D). The occurrence of these diagnostic ions ranges from >90% for oxonium-desthiobiotin and dehydrobiotin to ≈60% for the other presented diagnostic ions. During the studies of CuAAC side reactivity Wiest et al. proceeded with detailed characterization of desthiobition-linker-derived diagnostic ions and probe-specific fragments [65]. The latter are relevant for the discovery of probe–peptide conjugates as the fragments contain both desthiobiotin and probe-derived moieties (Figure 8E) [65]. In the study several small molecule probes were used including propargylamine, a glutamine derivative, $N^6$-propargyladenosine and a tyrosine derivative to document the thiotriazole formation during CuAAC. The two selected examples of adenosine and glutamine-derived diagnostic ions are shown in Figure 8E. Since, in this case, the cysteine thiols were derivatized and the cysteine C–S bond was previously described to be susceptible to fragmentation, the fragments containing the thiol provide the information on the structure of the modified amino acid side chain [107,108]. Although the study was focused on the characterization of the side-products of CuAAC, it shares multiple features applicable for determination of NP–peptide conjugates in complex proteomic samples.

Recently, an isoDTB linker was introduced, which is synthetically easily accessible and now as well commercially available. IsoDTB combines the desthiobiotin moiety and isotopically labeled residues for direct quantification (Figure 9A) [103]. Thus far, the linker has been mainly utilized for profiling of reactive cysteines. For this review the labile and open search of the data acquired by Zanon et al. using the FragPipe suite led to identification of several anticipated diagnostic ions, including the b-type ions from fragmented peptide bonds of the linker (Figure 9A). The study builds on the reactivity of free cysteine thiols, which are reacted on the proteome-wide level with iodoacetamide-alkyne via $S_N$ reaction. Similar as for the above-described DTB-linker, several probe-derived ions were observed (Figure 9B). The isoDTB-linker's availability in 'light' and 'heavy' form together with highly optimized protocols and direct presets in MSFragger for streamlined analysis make this linker a feasible choice for site identifications of probe–peptide conjugates.

An alternative approach to selectively release the probe–peptide conjugates after the affinity enrichment is the utilization of a cleavable linker [52]. From the plethora of cleavable linkers reviewed by Beard et al., we focus here on selected few, which were used in recent studies with sufficient resolution and depth [52]. In general, cleavable linkers can be divided into two main groups based on the cleavage mechanism, which can be either enzymatic or chemical. The advantage of cleavable linkers is the possibility to combine stringent enrichment, ensured by the biotin–streptavidin complex, with the selective release of modified peptide–probe conjugates afterwards. The frequently used enzymatically cleaved TEV-linker contains a peptide sequence, which is recognized by Tobacco Etch Virus nuclear inclusion A peptidase, which is orthogonal to trypsin and chymotrypsin cleavage – two most commonly used proteases [109]. Multiple modifications to the peptide sequence and linker structure were consecutively introduced by the Cravatt lab including the 'light' and 'heavy' derivatives containing stable isotopes enabling the quantification and direct comparison between the conditions [80]. The disadvantage of the TEV-linker might be the multistep synthesis. The labile and open search of the publicly available data revealed a set of diagnostic ions, both derived from the linker and the probe (Figure 10A and B). The structural motives of the diagnostic ions corroborate the above-described findings. The fragmentation of the triazole ring leaving the primary amine and b-ion resulting from the fragmentation of the TEV-recognition peptide sequence. The chemical cleavage of the linker to release probe–peptide conjugates is achieved mainly by the change of the pH or via reducing conditions to release acid, base, or redox-labile parts of the linker [52]. Most recently, Wozniak et al. mapped the binding sites of a large small molecules library, which were equipped with a photocrosslinker [110]. The study used the acid-cleavable dialkoxydiphenylsilane (DADPS) linker to release the probe–peptide conjugates (Figure 11A). The crucial advantage of the study was the introduction of a search algorithm considering generation of the chimeric spectra. The chimeric spectra may include

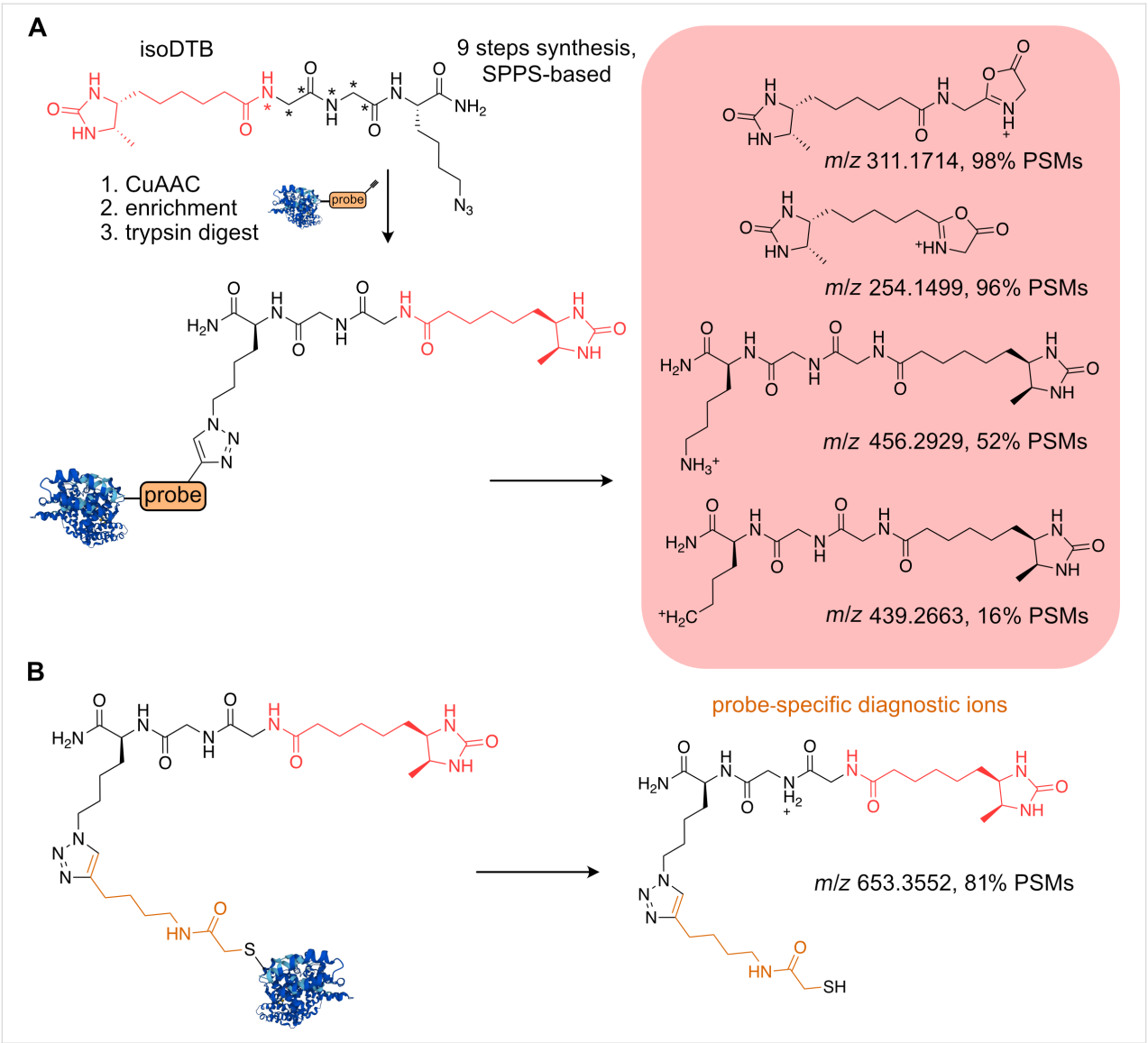



**Figure 9:** A) isoDTB linker and probe-specific diagnostic ions (B). *A star indicates the possible introduction of stable isotopes for MS quantification.

fragmentation of multiple probe–peptide conjugates which differ only in the position of the probe–peptide covalent bond. Thus, the physicochemical properties are highly similar as well as the mass of the conjugate leading to coelution from LC and cofragmentation during MS/MS. The case study of the adamantane-based probe revealed at least five probe-derived diagnostic ions from labile search, which are summarized in Figure 11B. Next, we used the diagnostic ions in a closed search with the MaxQuant search engine. For this purpose, we set two calculation runs searching for peptides carrying a mass offset of 485.3366 on C, K, R, E, D, S, T, or H. The closed search including the diagnostic ions found ≈20% more identified peptides showing the search improvement by diagnostic ions (Figure 12). The DADPS-linker containing 'light' or 'heavy' valine can be synthetized in total nine steps [111].

After the introduction of the DADPS-based linker by Szychowski et al. for the first time several interesting analogs were developed, including the shortened version described in the previous paragraph (Figure 13) [112-115]. As an interesting feature, the DADPS-linker with dibromo substitution takes advantage of the naturally occurring stable isotopes $^{79}$Br and $^{81}$Br in almost stoichiometric ratio. The predictable isotopic pattern thus introduces a searchable moiety and therefore clearly tags the MS/MS spectra derived from probe–peptide conjugates [113]. Another DADPS-based linker improves the CuAAC kinetics by using picolyl azide [115]. Together, the application of the acid-cleavable DADPS linkers is a feasible alternative to enzymatic cleavage. In particular, when considering the linker preparation, the biotin–DADPS–azide can be synthetized in only two steps [112].

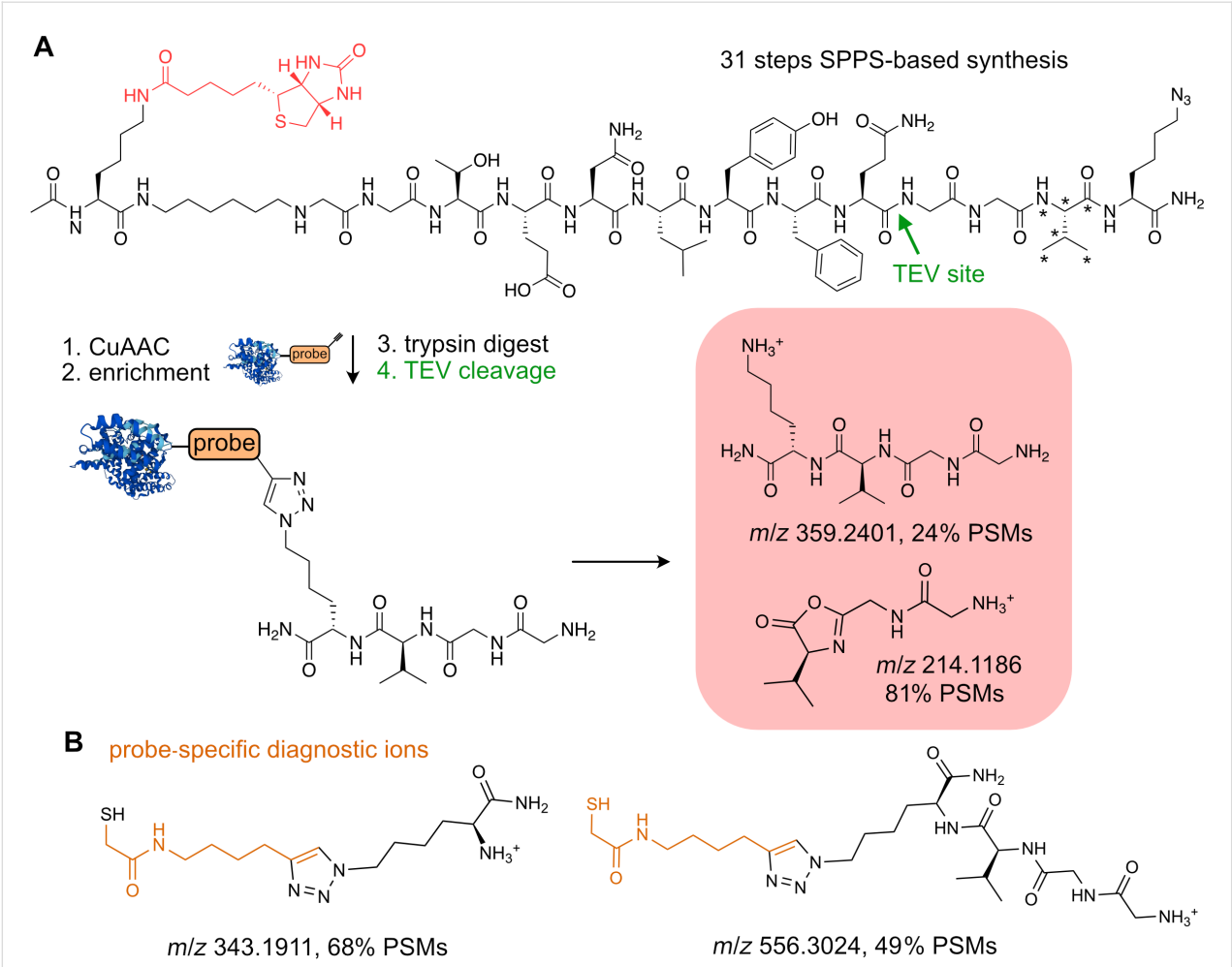



**Figure 10:** TEV-cleavable linker structure with its characteristic diagnostic ions (A) and probe-specific diagnostic ions (B). *A star indicates the possible introduction of stable isotopes for MS quantification.

Finally, the development of multifunctional linkers enables to combine the in-gel and MS-based analysis in one experiment [116-120]. The trifunctional linker AzKTB (Figure 14) was introduced by Wright et al. to study protein *N*-myristoylation and was later used for the identification of the natural product callyspongynic acid and zerumbone protein targets [117,121,122]. The AzKTB linker integrates biotin for affinity enrichment with a TAMRA fluorophore for in-gel analysis, a terminal azide for CuAAC to label probe–protein conjugates, and a trypsin-cleavage site, which allows to release the modified probe–peptide conjugates together with other unmodified tryptic peptides from the streptavidin beads (Figure 14). In principle, the release of modified peptides can be used for side identification given the relatively high abundance and ionization efficiency due to remaining lysine. The use of the AzTB linker is practical for swift optimization of enrichment workflows, because after the probe–protein conjugates enrichment on the streptavidin beads, the beads can be split into two portions from which one is analyzed by in-gel analysis first to determine the overall labeling and the other portion after on-beads trypsin digest is subjected to LC–MS/MS analysis. The AzKTB reagent can be synthetized by standard solid-phase peptide synthesis (SPPS) with overall 17% yield.

In this review we would like to also highlight the importance of deposition of the acquired MS data to publicly available depositaries with a clear description of the sample preparation protocol. This may allow the community to research the data and apply the ever-improving search algorithms to find a new features and probe–protein interactions.

## Automatization of chemical proteomics workflows

The increased throughput and less hands-on time during the chemical proteomics workflows led recently to a number of protocols, which either multiplicate the sample preparation or

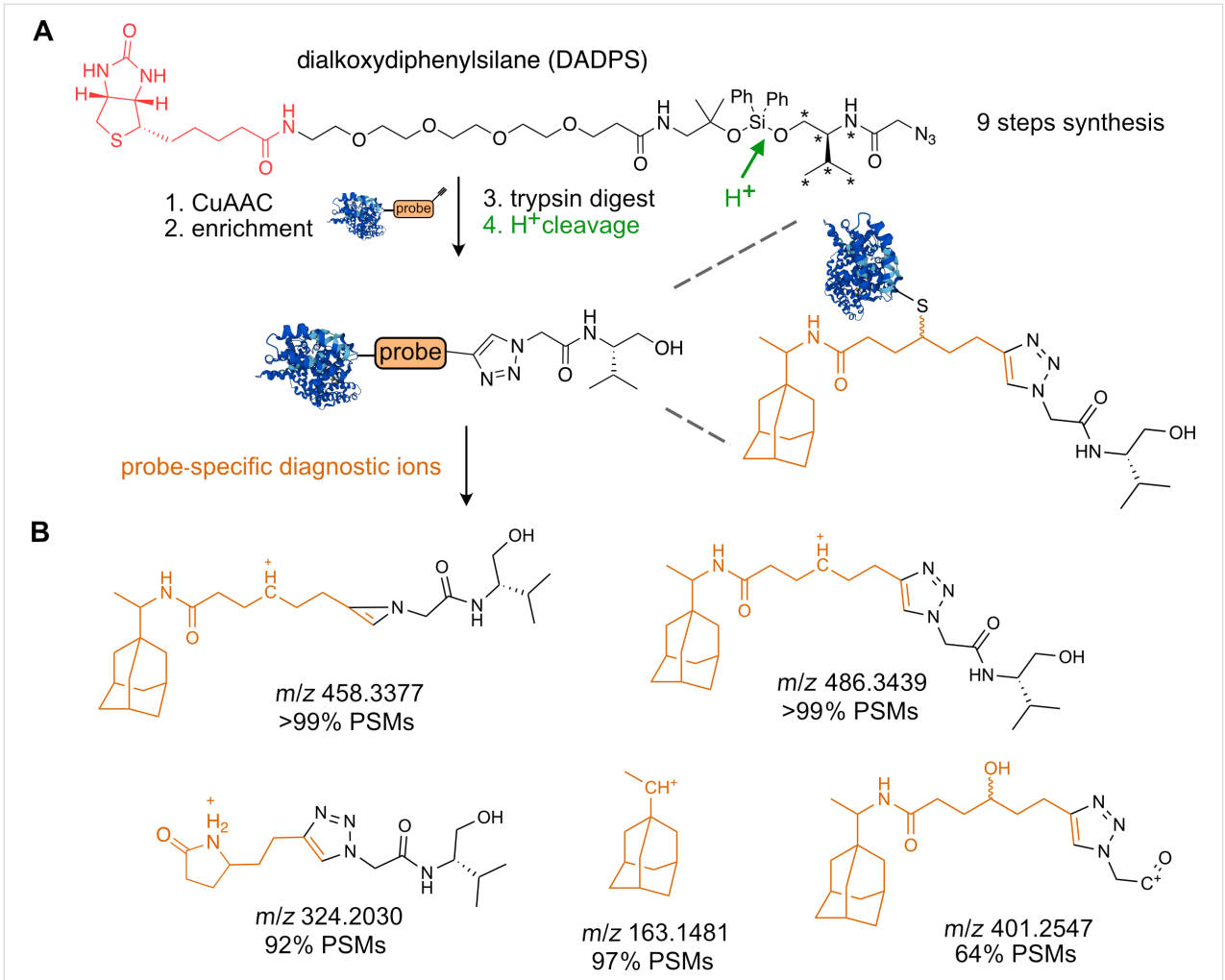



**Figure 11:** A) Structure of the full length DADPS linker and remaining part after cleavage. B) Diagnostic ions. *A star indicates the possible introduction of stable isotopes for MS quantification.

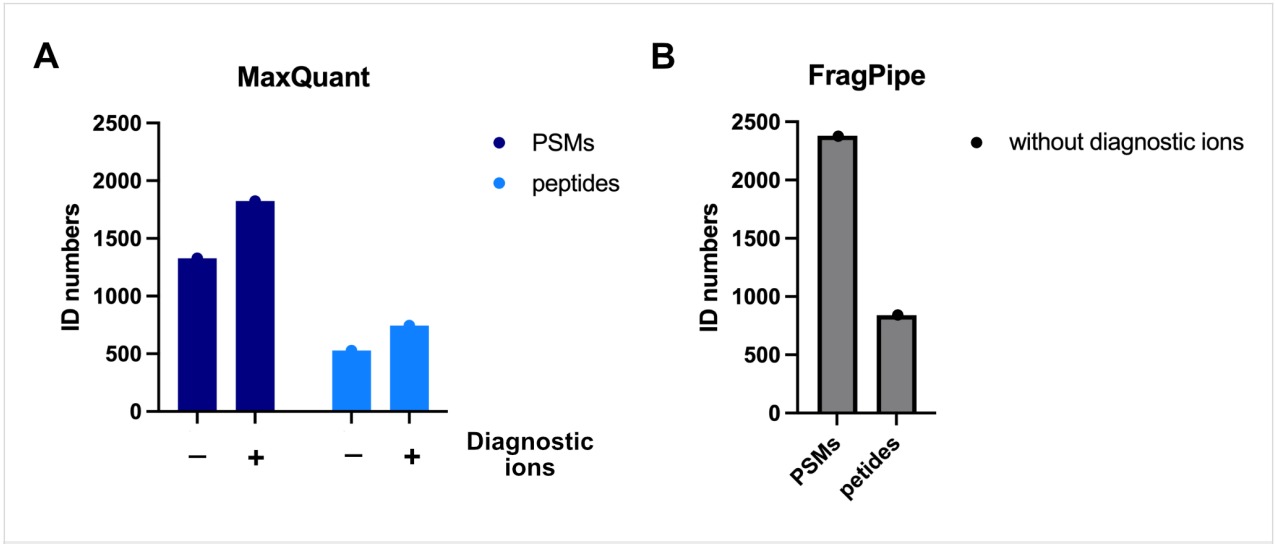



**Figure 12:** Diagnostic peaks included in the search identify higher numbers of modified PSMs and peptides using MaxQuant (A) and comparison with FragPipe (B) in case of the adamantane-based probe from the study by Wozniak et al. [110].

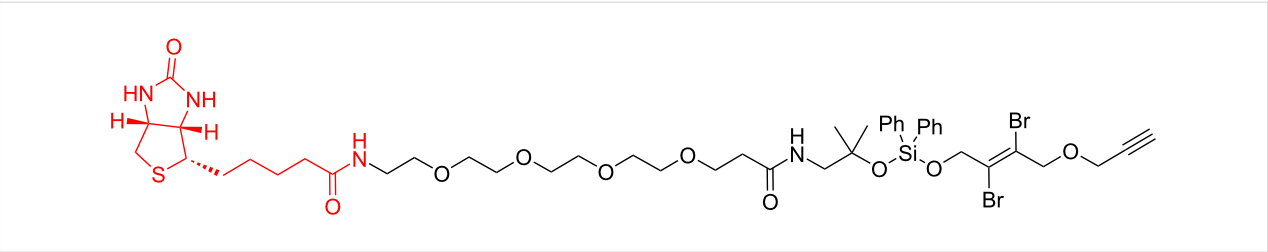


**Figure 13:** An alternative DADPS linker.

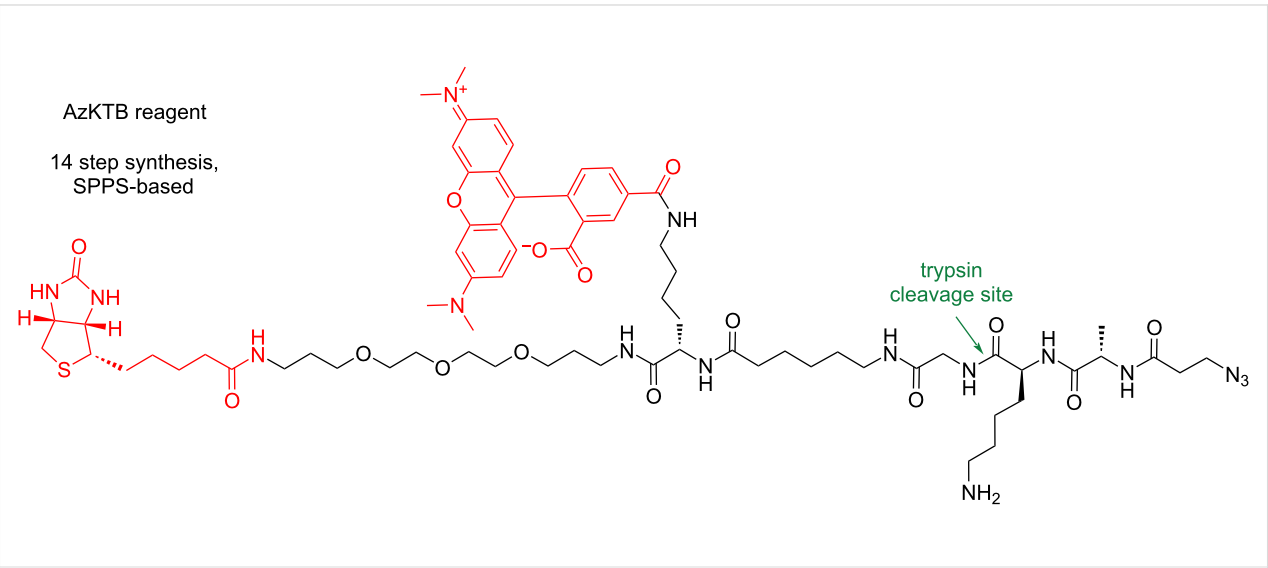



**Figure 14:** Chemical structure of the trifunctional trypsin cleavable AzKTB linker.

scale down and partially automatize the protocol or both. The standard mass spectrometry-based chemical proteomics workflow requires up to 1 mg of total protein, acetone protein precipitation after the 'click' reaction to remove the excess of the biotin-containing reagent, desalting of the peptides after on-beads digest, and concentration of peptides either by freeze-drying or on SpeedVac. This standard protocol is typically carried out in individual 1.5 mL tubes which need to be handled separately. Indeed, the standard protocol would span over several days with a limited number of samples possibly processed in one batch. With the advancement of LC–MS/MS sensitivity and speed of spectra acquisition, it is possible now to scale-down the chemical proteomics protocol. There are several approaches with significantly increased throughput available for either probe–peptide conjugates [123,124] or probe–proteins enrichment [94]. The underlying and unifying advantage of these protocols is the use of a mixture of hydrophilic and lipophilic carboxylate-coated magnetic beads to replace the protein acetone precipitation after CuAAC. After the 'click' between the probe–protein conjugates and affinity tag in cell lysate, the proteins are loaded onto the carboxylate magnetic beads and aggregated with the beads by addition of an organic solvent, for example ethanol or acetonitrile. The subsequent washes of the carboxylate magnetic beads with proteins by an organic solvent–water mixture or organic solvent alone cleans the probe–protein conjugates from excess of 'click' reagents and non-protein components present in lysates with high efficiency. The fast sequence of the wash steps is performed on a magnet separating the magnetic beads with aggregated proteins from the wash solution. Next, the proteins can be proteolytically cleaved on-beads or directly eluted for further enrichment on an affinity resin. Due to the easy handling of the magnetic beads and separation for the liquid phase using the magnet, the protocol utilizing the carboxylate magnetic beads was first developed for whole proteome proteomics, so called single-pot solid-phase enhanced sample preparation (SP3) [125,126]. The workflow can be scaled down readily to low protein amounts and small volumes compatible with the 96 and 384-well plate format fitting to liquid handling robots. An automatization of the protocol indeed lowers the necessary hands-on time and improves robustness and reproducibility of the protocol. However, even the scale-down of the protocol into microplates with manual pipetting can significantly parallelize and thus speed up the process.

## Conclusion

In the future, the aim will be to unlock MS-based chemical proteomic workflows, which will obscure the necessity of an enrichment step and thus will lead to significantly simplified and accelerated characterization of important and desired probe–protein interactions. Until then, the transfer of expertise and techniques from different areas of the chemical proteomic field have the potential to facilitate this advancement.

## Funding

This work was funded by a Ph.D. fellowship from the Studienstiftung des deutschen Volkes to A.W. and a Liebig fellowship from VCI to P.K. We are grateful for the generous support from SFB 1309 – 325871075 (Deutsche Forschungsgemeinschaft) and Boehringer Ingelheim Foundation for Plus 3 Program granted to P.K.

## Author Contributions

Andreas Wiest: conceptualization; writing – original draft; writing – review & editing. Pavel Kielkowski: conceptualization; funding acquisition; visualization; writing – original draft; writing – review & editing.

## ORCID® iDs

Andreas Wiest - https://orcid.org/0000-0003-2525-6304
Pavel Kielkowski - https://orcid.org/0000-0003-4910-6263

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

1. Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Orhan, I. E.; Banach, M.; Rollinger, J. M.; Barreca, D.; Weckwerth, W.; Bauer, R.; Bayer, E. A.; Majeed, M.; Bishayee, A.; Bochkov, V.; Bonn, G. K.; Braidy, N.; Bucar, F.; Cifuentes, A.; D'Onofrio, G.; Bodkin, M.; Diederich, M.; Dinkova-Kostova, A. T.; Efferth, T.; Bairi, K. E.; Arkells, N.; Fan, T.-P.; Fiebich, B. L.; Freissmuth, M.; Georgiev, M. I.; Gibbons, S.; Godfrey, K. M.; Gruber, C. W.; Heer, J.; Huber, L. A.; Ibanez, E.; Kijjoa, A.; Kiss, A. K.; Lu, A.; Macias, F. A.; Miller, M. J. S.; Mocan, A.; Müller, R.; Nicoletti, F.; Perry, G.; Pittalà, V.; Rastrelli, L.; Ristow, M.; Russo, G. L.; Silva, A. S.; Schuster, D.; Sheridan, H.; Skalicka-Woźniak, K.; Skaltsounis, L.; Sobarzo-Sánchez, E.; Bredt, D. S.; Stuppner, H.; Sureda, A.; Tzvetkov, N. T.; Vacca, R. A.; Aggarwal, B. B.; Battino, M.; Giampieri, F.; Wink, M.; Wolfender, J.-L.; Xiao, J.; Yeung, A. W. K.; Lizard, G.; Popp, M. A.; Heinrich, M.; Berindan-Neagoe, I.; Stadler, M.; Daglia, M.; Verpoorte, R.; Supuran, C. T. *Nat. Rev. Drug Discovery* **2021,** *20,* 200–216. doi:10.1038/s41573-020-00114-z
2. Yıldırım, M. A.; Goh, K.-I.; Cusick, M. E.; Barabási, A.-L.; Vidal, M. *Nat. Biotechnol.* **2007,** *25,* 1119–1126. doi:10.1038/nbt1338
3. Meissner, F.; Geddes-McAlister, J.; Mann, M.; Bantscheff, M. *Nat. Rev. Drug Discovery* **2022,** *21,* 637–654. doi:10.1038/s41573-022-00409-3
4. Demarque, D. P.; Crotti, A. E. M.; Vessecchi, R.; Lopes, J. L. C.; Lopes, N. P. *Nat. Prod. Rep.* **2016,** *33,* 432–455. doi:10.1039/c5np00073d
5. Wright, M. H.; Sieber, S. A. *Nat. Prod. Rep.* **2016,** *33,* 681–708. doi:10.1039/c6np00001k
6. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989,** *246,* 64–71. doi:10.1126/science.2675315
7. Sinha, A.; Mann, M. *Biochemist* **2020,** *42,* 64–69. doi:10.1042/bio20200057
8. Jiang, L.; Wang, M.; Lin, S.; Jian, R.; Li, X.; Chan, J.; Dong, G.; Fang, H.; Robinson, A. E.; Aguet, F.; Anand, S.; Ardlie, K. G.; Gabriel, S.; Getz, G.; Graubert, A.; Hadley, K.; Handsaker, R. E.; Huang, K. H.; Kashin, S.; MacArthur, D. G.; Meier, S. R.; Nedzel, J. L.; Nguyen, D. Y.; Segrè, A. V.; Todres, E.; Balliu, B.; Barbeira, A. N.; Battle, A.; Bonazzola, R.; Brown, A.; Brown, C. D.; Castel, S. E.; Conrad, D.; Cotter, D. J.; Cox, N.; Das, S.; de Goede, O. M.; Dermitzakis, E. T.; Engelhardt, B. E.; Eskin, E.; Eulalio, T. Y.; Ferraro, N. M.; Flynn, E.; Fresard, L.; Gamazon, E. R.; Garrido-Martín, D.; Gay, N. R.; Guigó, R.; Hamel, A. R.; He, Y.; Hoffman, P. J.; Hormozdiari, F.; Hou, L.; Im, H. K.; Jo, B.; Kasela, S.; Kellis, M.; Kim-Hellmuth, S.; Kwong, A.; Lappalainen, T.; Li, X.; Liang, Y.; Mangul, S.; Mohammadi, P.; Montgomery, S. B.; Muñoz-Aguirre, M.; Nachun, D. C.; Nobel, A. B.; Oliva, M.; Park, Y.; Park, Y.; Parsana, P.; Reverter, F.; Rouhana, J. M.; Sabatti, C.; Saha, A.; Skol, A. D.; Stephens, M.; Stranger, B. E.; Strober, B. J.; Teran, N. A.; Viñuela, A.; Wang, G.; Wen, X.; Wright, F.; Wucher, V.; Zou, Y.; Ferreira, P. G.; Li, G.; Melé, M.; Yeger-Lotem, E.; Barcus, M. E.; Bradbury, D.; Krubit, T.; McLean, J. A.; Qi, L.; Robinson, K.; Roche, N. V.; Smith, A. M.; Sobin, L.; Tabor, D. E.; Undale, A.; Bridge, J.; Brigham, L. E.; Foster, B. A.; Gillard, B. M.; Hasz, R.; Hunter, M.; Johns, C.; Johnson, M.; Karasik, E.; Kopen, G.; Leinweber, W. F.; McDonald, A.; Moser, M. T.; Myer, K.; Ramsey, K. D.; Roe, B.; Shad, S.; Thomas, J. A.; Walters, G.; Washington, M.; Wheeler, J.; Jewell, S. D.; Rohrer, D. C.; Valley, D. R.; Davis, D. A.; Mash, D. C.; Branton, P. A.; Barker, L. K.; Gardiner, H. M.; Mosavel, M.; Siminoff, L. A.; Flicek, P.; Haeussler, M.; Juettemann, T.; Kent, W. J.; Lee, C. M.; Powell, C. C.; Rosenbloom, K. R.; Ruffier, M.; Sheppard, D.; Taylor, K.; Trevanion, S. J.; Zerbino, D. R.; Abell, N. S.; Akey, J.; Chen, L.; Demanelis, K.; Doherty, J. A.; Feinberg, A. P.; Hansen, K. D.; Hickey, P. F.; Jasmine, F.; Kaul, R.; Kibriya, M. G.; Li, J. B.; Li, Q.; Linder, S. E.; Pierce, B. L.; Rizzardi, L. F.; Smith, K. S.; Stamatoyannopoulos, J.; Tang, H.; Carithers, L. J.; Guan, P.; Koester, S. E.; Little, A. R.; Moore, H. M.; Nierras, C. R.; Rao, A. K.; Vaught, J. B.; Volpi, S.; Snyder, M. P. *Cell* **2020,** *183,* 269–283. doi:10.1016/j.cell.2020.08.036
9. Bekker-Jensen, D. B.; Martínez-Val, A.; Steigerwald, S.; Rüther, P.; Fort, K. L.; Arrey, T. N.; Harder, A.; Makarov, A.; Olsen, J. V. *Mol. Cell. Proteomics* **2020,** *19,* 716–729. doi:10.1074/mcp.tir119.001906
10. Bian, Y.; The, M.; Giansanti, P.; Mergner, J.; Zheng, R.; Wilhelm, M.; Boychenko, A.; Kuster, B. *Anal. Chem. (Washington, DC, U. S.)* **2021,** *93,* 8687–8692. doi:10.1021/acs.analchem.1c00738
11. Bian, Y.; Zheng, R.; Bayer, F. P.; Wong, C.; Chang, Y.-C.; Meng, C.; Zolg, D. P.; Reinecke, M.; Zecha, J.; Wiechmann, S.; Heinzlmeir, S.; Scherr, J.; Hemmer, B.; Baynham, M.; Gingras, A.-C.; Boychenko, O.; Kuster, B. *Nat. Commun.* **2020,** *11,* 157. doi:10.1038/s41467-019-13973-x

12. Wang, Z.; Mülleder, M.; Batruch, I.; Chelur, A.; Textoris-Taube, K.; Schwecke, T.; Hartl, J.; Causon, J.; Castro-Perez, J.; Demichev, V.; Tate, S.; Ralser, M. *eLife* **2022,** *11,* e83947. doi:10.7554/elife.83947
13. Messner, C. B.; Demichev, V.; Wang, Z.; Hartl, J.; Kustatscher, G.; Mülleder, M.; Ralser, M. *Proteomics* **2023,** 2200013. doi:10.1002/pmic.202200013
14. Messner, C. B.; Demichev, V.; Bloomfield, N.; Yu, J. S. L.; White, M.; Kreidl, M.; Egger, A.-S.; Freiwald, A.; Ivosev, G.; Wasim, F.; Zelezniak, A.; Jürgens, L.; Suttorp, N.; Sander, L. E.; Kurth, F.; Lilley, K. S.; Mülleder, M.; Tate, S.; Ralser, M. *Nat. Biotechnol.* **2021,** *39,* 846–854. doi:10.1038/s41587-021-00860-4
15. Hosp, F.; Mann, M. *Neuron* **2017,** *96,* 558–571. doi:10.1016/j.neuron.2017.09.025
16. Aebersold, R.; Mann, M. *Nature* **2003,** *422,* 198–207. doi:10.1038/nature01511
17. Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. *Nat. Methods* **2020,** *17,* 41–44. doi:10.1038/s41592-019-0638-x
18. Yu, F.; Teo, G. C.; Kong, A. T.; Fröhlich, K.; Li, G. X.; Demichev, V.; Nesvizhskii, A. I. *Nat. Commun.* **2023,** *14,* 4154. doi:10.1038/s41467-023-39869-5
19. Polasky, D. A.; Geiszler, D. J.; Yu, F.; Li, K.; Teo, G. C.; Nesvizhskii, A. I. *Mol. Cell. Proteomics* **2023,** *22,* 100538. doi:10.1016/j.mcpro.2023.100538
20. Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. *Nat. Methods* **2017,** *14,* 513–520. doi:10.1038/nmeth.4256
21. Lou, R.; Cao, Y.; Li, S.; Lang, X.; Li, Y.; Zhang, Y.; Shui, W. *Nat. Commun.* **2023,** *14,* 94. doi:10.1038/s41467-022-35740-1
22. Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; Prianichnikov, N.; Yılmaz, Ş.; Rudolph, J. D.; Tenzer, S.; Perez-Riverol, Y.; Nagaraj, N.; Humphrey, S. J.; Cox, J. *Nat. Biotechnol.* **2021,** *39,* 1563–1573. doi:10.1038/s41587-021-00968-7
23. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature* **2021,** *596,* 583–589. doi:10.1038/s41586-021-03819-2
24. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Žídek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. *Nucleic Acids Res.* **2022,** *50,* D439–D444. doi:10.1093/nar/gkab1061
25. Reilly, L.; Lara, E.; Ramos, D.; Li, Z.; Pantazis, C. B.; Stadler, J.; Santiana, M.; Roberts, J.; Faghri, F.; Hao, Y.; Nalls, M. A.; Narayan, P.; Liu, Y.; Singleton, A. B.; Cookson, M. R.; Ward, M. E.; Qi, Y. A. *Cells Rep. Methods* **2023,** *3,* 100593. doi:10.1016/j.crmeth.2023.100593
26. Smith, L. M.; Thomas, P. M.; Shortreed, M. R.; Schaffer, L. V.; Fellers, R. T.; LeDuc, R. D.; Tucholski, T.; Ge, Y.; Agar, J. N.; Anderson, L. C.; Chamot-Rooke, J.; Gault, J.; Loo, J. A.; Paša-Tolić, L.; Robinson, C. V.; Schlüter, H.; Tsybin, Y. O.; Vilaseca, M.; Vizcaíno, J. A.; Danis, P. O.; Kelleher, N. L. *Nat. Methods* **2019,** *16,* 939–940. doi:10.1038/s41592-019-0573-x
27. He, J.-X.; Fei, Z.-C.; Fu, L.; Tian, C.-P.; He, F.-C.; Chi, H.; Yang, J. *Nat. Chem. Biol.* **2022,** *18,* 904–912. doi:10.1038/s41589-022-01074-8
28. Pan, S.; Jang, S.-Y.; Wang, D.; Liew, S. S.; Li, Z.; Lee, J.-S.; Yao, S. Q. *Angew. Chem., Int. Ed.* **2017,** *56,* 11816–11821. doi:10.1002/anie.201706076
29. Li, Z.; Hao, P.; Li, L.; Tan, C. Y. J.; Cheng, X.; Chen, G. Y. J.; Sze, S. K.; Shen, H.-M.; Yao, S. Q. *Angew. Chem., Int. Ed.* **2013,** *52,* 8551–8556. doi:10.1002/anie.201300683
30. Boike, L.; Henning, N. J.; Nomura, D. K. *Nat. Rev. Drug Discovery* **2022,** *21,* 881–898. doi:10.1038/s41573-022-00542-z
31. Kleiner, P.; Heydenreuter, W.; Stahl, M.; Korotkov, V. S.; Sieber, S. A. *Angew. Chem., Int. Ed.* **2017,** *56,* 1396–1401. doi:10.1002/anie.201605993
32. West, A. V.; Muncipinto, G.; Wu, H.-Y.; Huang, A. C.; Labenski, M. T.; Jones, L. H.; Woo, C. M. *J. Am. Chem. Soc.* **2021,** *143,* 6691–6700. doi:10.1021/jacs.1c02509
33. Parker, C. G.; Pratt, M. R. *Cell* **2020,** *180,* 605–632. doi:10.1016/j.cell.2020.01.025
34. Bagdonaite, I.; Malaker, S. A.; Polasky, D. A.; Riley, N. M.; Schjoldager, K.; Vakhrushev, S. Y.; Halim, A.; Aoki-Kinoshita, K. F.; Nesvizhskii, A. I.; Bertozzi, C. R.; Wandall, H. H.; Parker, B. L.; Thaysen-Andersen, M.; Scott, N. E. *Nat. Rev. Methods Primers* **2022,** *2,* 48. doi:10.1038/s43586-022-00128-4
35. Agard, N. J.; Baskin, J. M.; Prescher, J. A.; Lo, A.; Bertozzi, C. R. *ACS Chem. Biol.* **2006,** *1,* 644–648. doi:10.1021/cb6003228
36. Ramil, C. P.; Lin, Q. *Chem. Commun.* **2013,** *49,* 11007–11022. doi:10.1039/c3cc44272a
37. Schauenburg, D.; Weil, T. *Adv. Sci.* **2024,** *11,* 2303396. doi:10.1002/advs.202303396
38. Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; Garlish, R.; Griffin, P. R.; Gross, M. L.; Guttman, M.; Hamuro, Y.; Heck, A. J. R.; Houde, D.; Iacob, R. E.; Jørgensen, T. J. D.; Kaltashov, I. A.; Klinman, J. P.; Konermann, L.; Man, P.; Mayne, L.; Pascal, B. D.; Reichmann, D.; Skehel, M.; Snijder, J.; Strutzenberg, T. S.; Underbakke, E. S.; Wagner, C.; Wales, T. E.; Walters, B. T.; Weis, D. D.; Wilson, D. J.; Wintrode, P. L.; Zhang, Z.; Zheng, J.; Schriemer, D. C.; Rand, K. D. *Nat. Methods* **2019,** *16,* 595–602. doi:10.1038/s41592-019-0459-y
39. Schopper, S.; Kahraman, A.; Leuenberger, P.; Feng, Y.; Piazza, I.; Müller, O.; Boersema, P. J.; Picotti, P. *Nat. Protoc.* **2017,** *12,* 2391–2410. doi:10.1038/nprot.2017.100
40. Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Molina, D. M.; Jafari, R.; Dovega, R. B.; Klaeger, S.; Kuster, B.; Nordlund, P.; Bantscheff, M.; Drewes, G. *Science* **2014,** *346,* 1255784. doi:10.1126/science.1255784
41. Molina, D. M.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P. *Science* **2013,** *341,* 84–87. doi:10.1126/science.1233606
42. Annis, D. A.; Athanasopoulos, J.; Curran, P. J.; Felsch, J. S.; Kalghatgi, K.; Lee, W. H.; Nash, H. M.; Orminati, J.-P. A.; Rosner, K. E.; Shipps, G. W., Jr.; Thaddupathy, G. R. A.; Tyler, A. N.; Vilenchik, L.; Wagner, C. R.; Wintner, E. A. *Int. J. Mass Spectrom.* **2004,** *238,* 77–83. doi:10.1016/j.ijms.2003.11.022
43. Chan, J. N. Y.; Vuckovic, D.; Sleno, L.; Olsen, J. B.; Pogoutse, O.; Havugimana, P.; Hewel, J. A.; Bajaj, N.; Wang, Y.; Musteata, M. F.; Nislow, C.; Emili, A. *Mol. Cell. Proteomics* **2012,** *11,* M111.016642. doi:10.1074/mcp.m111.016642

44. Hicks, K. G.; Cluntun, A. A.; Schubert, H. L.; Hackett, S. R.; Berg, J. A.; Leonard, P. G.; Ajalla Aleixo, M. A.; Zhou, Y.; Bott, A. J.; Salvatore, S. R.; Chang, F.; Blevins, A.; Barta, P.; Tilley, S.; Leifer, A.; Guzman, A.; Arok, A.; Fogarty, S.; Winter, J. M.; Ahn, H.-C.; Allen, K. N.; Block, S.; Cardoso, I. A.; Ding, J.; Dreveny, I.; Gasper, W. C.; Ho, Q.; Matsuura, A.; Palladino, M. J.; Prajapati, S.; Sun, P.; Tittmann, K.; Tolan, D. R.; Unterlass, J.; VanDemark, A. P.; Vander Heiden, M. G.; Webb, B. A.; Yun, C.-H.; Zhao, P.; Wang, B.; Schopfer, F. J.; Hill, C. P.; Nonato, M. C.; Muller, F. L.; Cox, J. E.; Rutter, J. *Science* **2023,** *379,* 996–1003. doi:10.1126/science.abm3452
45. Rix, U.; Superti-Furga, G. *Nat. Chem. Biol.* **2009,** *5,* 616–624. doi:10.1038/nchembio.216
46. Bantscheff, M.; Scholten, A.; Heck, A. J. R. *Drug Discovery Today* **2009,** *14,* 1021–1029. doi:10.1016/j.drudis.2009.07.001
47. Leitner, A. *Anal. Chim. Acta* **2018,** *1000,* 2–19. doi:10.1016/j.aca.2017.08.026
48. Fedorov, I. I.; Lineva, V. I.; Tarasova, I. A.; Gorshkov, M. V. *Biochemistry (Moscow)* **2022,** *87,* 983–994. doi:10.1134/s0006297922090103
49. Yang, Y.; Hahne, H.; Kuster, B.; Verhelst, S. H. L. *Mol. Cell. Proteomics* **2013,** *12,* 237–244. doi:10.1074/mcp.m112.021014
50. Yang, Y.; Fonović, M.; Verhelst, S. H. L. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2017,** *1491,* 185–203. doi:10.1007/978-1-4939-6439-0_14
51. Ianni, A. D.; Ihling, C. H.; Vranka, T.; Matoušek, V.; Sinz, A.; Iacobucci, C. *JACS Au* **2024,** *4,* 2936–2943. doi:10.1021/jacsau.4c00282
52. Beard, H. A.; Korovesis, D.; Chen, S.; Verhelst, S. H. L. *Mol. Omics* **2021,** *17,* 197–209. doi:10.1039/d0mo00181c
53. Heal, W. P.; Dang, T. H. T.; Tate, E. W. *Chem. Soc. Rev.* **2011,** *40,* 246–257. doi:10.1039/c0cs00004c
54. Speers, A. E.; Cravatt, B. F. *Curr. Protoc. Chem. Biol.* **2009,** *1,* 29–41. doi:10.1002/9780470559277.ch090138
55. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002,** *41,* 2596–2599. doi:10.1002/1521-3773(20020715)41:14<2596::aid-anie2596>3.0.co;2-4
56. Meldal, M.; Tornøe, C. W. *Chem. Rev.* **2008,** *108,* 2952–3015. doi:10.1021/cr0783479
57. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004,** *126,* 15046–15047. doi:10.1021/ja044996f
58. Saxon, E.; Bertozzi, C. R. *Science* **2000,** *287,* 2007–2010. doi:10.1126/science.287.5460.2007
59. Markos, A.; Biedermann, M.; Heimgärtner, J.; Schmitt, A.; Lang, K.; Wennemers, H. *J. Am. Chem. Soc.* **2023,** *145,* 19513–19517. doi:10.1021/jacs.3c07006
60. Oliveira, B. L.; Guo, Z.; Bernardes, G. J. L. *Chem. Soc. Rev.* **2017,** *46,* 4895–4950. doi:10.1039/c7cs00184c
61. McKay, C. S.; Finn, M. G. *Chem. Biol.* **2014,** *21,* 1075–1101. doi:10.1016/j.chembiol.2014.09.002
62. Prescher, J. A.; Bertozzi, C. R. *Nat. Chem. Biol.* **2005,** *1,* 13–21. doi:10.1038/nchembio0605-13
63. Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y. *Angew. Chem., Int. Ed.* **2012,** *51,* 5852–5856. doi:10.1002/anie.201108181
64. Cutolo, G.; Shankar, S. N.; Pratt, M. R. *Bioorg. Med. Chem. Lett.* **2021,** *48,* 128260. doi:10.1016/j.bmcl.2021.128260
65. Wiest, A.; Kielkowski, P. *J. Am. Chem. Soc.* **2024,** *146,* 2151–2159. doi:10.1021/jacs.3c11780
66. Yang, Y.; Dong, W.; Guo, Y.; Rioux, R. M. *Green Chem.* **2013,** *15,* 3170–3175. doi:10.1039/c3gc41330f
67. Yang, Y.; Yang, X.; Verhelst, S. H. L. *Molecules* **2013,** *18,* 12599–12608. doi:10.3390/molecules181012599
68. Guo, T.; Aebersold, R. *Proteomics* **2023,** *23,* e2200011. doi:10.1002/pmic.202200011
69. Messner, C. B.; Demichev, V.; Muenzner, J.; Aulakh, S. K.; Barthel, N.; Röhl, A.; Herrera-Domínguez, L.; Egger, A.-S.; Kamrad, S.; Hou, J.; Tan, G.; Lemke, O.; Calvani, E.; Szyrwiel, L.; Mülleder, M.; Lilley, K. S.; Boone, C.; Kustatscher, G.; Ralser, M. *Cell* **2023,** *186,* 2018–2034.e21. doi:10.1016/j.cell.2023.03.026
70. Cox, J.; Mann, M. *Nat. Biotechnol.* **2008,** *26,* 1367–1372. doi:10.1038/nbt.1511
71. Chang, H.-Y.; Kong, A. T.; da Veiga Leprevost, F.; Avtonomov, D. M.; Haynes, S. E.; Nesvizhskii, A. I. *J. Proteome Res.* **2020,** *19,* 2511–2515. doi:10.1021/acs.jproteome.0c00119
72. Geiszler, D. J.; Polasky, D. A.; Yu, F.; Nesvizhskii, A. I. *Nat. Commun.* **2023,** *14,* 4132. doi:10.1038/s41467-023-39828-0
73. Polasky, D. A.; Yu, F.; Teo, G. C.; Nesvizhskii, A. I. *Nat. Methods* **2020,** *17,* 1125–1132. doi:10.1038/s41592-020-0967-9
74. Yu, F.; Teo, G. C.; Kong, A. T.; Haynes, S. E.; Avtonomov, D. M.; Geiszler, D. J.; Nesvizhskii, A. I. *Nat. Commun.* **2020,** *11,* 4065. doi:10.1038/s41467-020-17921-y
75. Hendriks, I. A.; Larsen, S. C.; Nielsen, M. L. *Mol. Cell. Proteomics* **2019,** *18,* 1010–1026. doi:10.1074/mcp.tir119.001315
76. Marx, H.; Lemeer, S.; Schliep, J. E.; Matheron, L.; Mohammed, S.; Cox, J.; Mann, M.; Heck, A. J. R.; Kuster, B. *Nat. Biotechnol.* **2013,** *31,* 557–564. doi:10.1038/nbt.2585
77. Speers, A. E.; Cravatt, B. F. *Chem. Biol.* **2004,** *11,* 535–546. doi:10.1016/j.chembiol.2004.03.012
78. Böttcher, T.; Pitscheider, M.; Sieber, S. A. *Angew. Chem., Int. Ed.* **2010,** *49,* 2680–2698. doi:10.1002/anie.200905352
79. Darabedian, N.; Thompson, J. W.; Chuh, K. N.; Hsieh-Wilson, L. C.; Pratt, M. R. *Biochemistry* **2018,** *57,* 5769–5774. doi:10.1021/acs.biochem.8b00648
80. Weerapana, E.; Speers, A. E.; Cravatt, B. F. *Nat. Protoc.* **2007,** *2,* 1414–1425. doi:10.1038/nprot.2007.194
81. Makarov, D.; Telek, A.; Becker, T.; von Wrisberg, M.-K.; Schneider, S.; Kielkowski, P. *J. Mass Spectrom.* **2022,** *57,* e4812. doi:10.1002/jms.4812
82. Ma, T. P.; Izrael-Tomasevic, A.; Mroue, R.; Budayeva, H.; Malhotra, S.; Raisner, R.; Evangelista, M.; Rose, C. M.; Kirkpatrick, D. S.; Yu, K. *J. Proteome Res.* **2023,** *22,* 2218–2231. doi:10.1021/acs.jproteome.2c00703
83. Li, J.; Van Vranken, J. G.; Pontano Vaites, L.; Schweppe, D. K.; Huttlin, E. L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A. M.; Thompson, A. H.; Kuhn, K.; Pike, I.; Bomgarden, R. D.; Rogers, J. C.; Gygi, S. P.; Paulo, J. A. *Nat. Methods* **2020,** *17,* 399–404. doi:10.1038/s41592-020-0781-4
84. Broncel, M.; Serwa, R. A.; Bunney, T. D.; Katan, M.; Tate, E. W. *Mol. Cell. Proteomics* **2016,** *15,* 715–725. doi:10.1074/mcp.o115.054429
85. Kielkowski, P.; Buchsbaum, I. Y.; Kirsch, V. C.; Bach, N. C.; Drukker, M.; Cappello, S.; Sieber, S. A. *Nat. Commun.* **2020,** *11,* 517. doi:10.1038/s41467-019-14235-6
86. Stadlmeier, M.; Runtsch, L. S.; Streshnev, F.; Wühr, M.; Carell, T. *ChemBioChem* **2020,** *21,* 103–107. doi:10.1002/cbic.201900611

87. Kao, A.; Chiu, C.-l.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. *Mol. Cell. Proteomics* **2011,** *10,* M110.002170. doi:10.1074/mcp.m110.002212
88. Calle, B.; Bineva-Todd, G.; Marchesi, A.; Flynn, H.; Ghirardello, M.; Tastan, O. Y.; Roustan, C.; Choi, J.; Galan, M. C.; Schumann, B.; Malaker, S. A. *J. Am. Soc. Mass Spectrom.* **2021,** *32,* 2366–2375. doi:10.1021/jasms.1c00084
89. Ismail, H. M.; O'Neill, P. M.; Hong, D. W.; Finn, R. D.; Henderson, C. J.; Wright, A. T.; Cravatt, B. F.; Hemingway, J.; Paine, M. J. I. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110,* 19766–19771. doi:10.1073/pnas.1320185110
90. Zheng, B.; Zhu, S.; Wu, X. *ACS Chem. Biol.* **2015,** *10,* 115–121. doi:10.1021/cb500758s
91. Nishino, M.; Choy, J. W.; Gushwa, N. N.; Oses-Prieto, J. A.; Koupparis, K.; Burlingame, A. L.; Renslo, A. R.; McKerrow, J. H.; Taunton, J. *eLife* **2013,** *2,* e00712. doi:10.7554/elife.00712
92. Li, J.; Cisar, J. S.; Zhou, C.-Y.; Vera, B.; Williams, H.; Rodríguez, A. D.; Cravatt, B. F.; Romo, D. *Nat. Chem.* **2013,** *5,* 510–517. doi:10.1038/nchem.1653
93. Dundas, C. M.; Demonte, D.; Park, S. *Appl. Microbiol. Biotechnol.* **2013,** *97,* 9343–9353. doi:10.1007/s00253-013-5232-z
94. Becker, T.; Wiest, A.; Telek, A.; Bejko, D.; Hoffmann-Röder, A.; Kielkowski, P. *JACS Au* **2022,** *2,* 1712–1723. doi:10.1021/jacsau.2c00284
95. Duan, X.; Li, Y.; Rajan, N. K.; Routenberg, D. A.; Modis, Y.; Reed, M. A. *Nat. Nanotechnol.* **2012,** *7,* 401–407. doi:10.1038/nnano.2012.82
96. Piran, U.; Riordan, W. J. *J. Immunol. Methods* **1990,** *133,* 141–143. doi:10.1016/0022-1759(90)90328-s
97. GREEN, N. *Biochem. J.* **1963,** *89,* 585–591. doi:10.1042/bj0890585
98. Wang, J.; Tan, X. F.; Nguyen, V. S.; Yang, P.; Zhou, J.; Gao, M.; Li, Z.; Lim, T. K.; He, Y.; Ong, C. S.; Lay, Y.; Zhang, J.; Zhu, G.; Lai, S.-L.; Ghosh, D.; Mok, Y. K.; Shen, H.-M.; Lin, Q. *Mol. Cell. Proteomics* **2014,** *13,* 876–886. doi:10.1074/mcp.m113.029793
99. Shibata, T.; Nakashima, F.; Honda, K.; Lu, Y.-J.; Kondo, T.; Ushida, Y.; Aizawa, K.; Suganuma, H.; Oe, S.; Tanaka, H.; Takahashi, T.; Uchida, K. *J. Biol. Chem.* **2014,** *289,* 32757–32772. doi:10.1074/jbc.m114.585901
100. Yang, P.-Y.; Liu, K.; Ngai, M. H.; Lear, M. J.; Wenk, M. R.; Yao, S. Q. *J. Am. Chem. Soc.* **2010,** *132,* 656–666. doi:10.1021/ja907716f
101. Rybak, J.-N.; Scheurer, S. B.; Neri, D.; Elia, G. *Proteomics* **2004,** *4,* 2296–2299. doi:10.1002/pmic.200300780
102. Yan, T.; Palmer, A. B.; Geiszler, D. J.; Polasky, D. A.; Boatner, L. M.; Burton, N. R.; Armenta, E.; Nesvizhskii, A. I.; Backus, K. M. *Anal. Chem. (Washington, DC, U. S.)* **2022,** *94,* 3800–3810. doi:10.1021/acs.analchem.1c04402
103. Zanon, P. R. A.; Lewald, L.; Hacker, S. M. *Angew. Chem., Int. Ed.* **2020,** *59,* 2829–2836. doi:10.1002/anie.201912075
104. Patricelli, M. P.; Nomanbhoy, T. K.; Wu, J.; Brown, H.; Zhou, D.; Zhang, J.; Jagannathan, S.; Aban, A.; Okerberg, E.; Herring, C.; Nordin, B.; Weissig, H.; Yang, Q.; Lee, J.-D.; Gray, N. S.; Kozarich, J. W. *Chem. Biol.* **2011,** *18,* 699–710. doi:10.1016/j.chembiol.2011.04.011
105. Hirsch, J. D.; Eslamizar, L.; Filanoski, B. J.; Malekzadeh, N.; Haugland, R. P.; Beechem, J. M.; Haugland, R. P. *Anal. Biochem.* **2002,** *308,* 343–357. doi:10.1016/s0003-2697(02)00201-4
106. Green, N. M. *Methods Enzymol.* **1970,** *18,* 418–424. doi:10.1016/0076-6879(71)18342-5
107. Chalkley, R. J.; Brinkworth, C. S.; Burlingame, A. L. *J. Am. Soc. Mass Spectrom.* **2006,** *17,* 1271–1274. doi:10.1016/j.jasms.2006.05.017
108. Ficarro, S. B.; Browne, C. M.; Card, J. D.; Alexander, W. M.; Zhang, T.; Park, E.; McNally, R.; Dhe-Paganon, S.; Seo, H.-S.; Lamberto, I.; Eck, M. J.; Buhrlage, S. J.; Gray, N. S.; Marto, J. A. *Anal. Chem. (Washington, DC, U. S.)* **2016,** *88,* 12248–12254. doi:10.1021/acs.analchem.6b03394
109. Speers, A. E.; Cravatt, B. F. *J. Am. Chem. Soc.* **2005,** *127,* 10018–10019. doi:10.1021/ja0532842
110. Wozniak, J. M.; Li, W.; Governa, P.; Chen, L.-Y.; Jadhav, A.; Dongre, A.; Forli, S.; Parker, C. G. *Nat. Chem. Biol.* **2024,** *20,* 823–834. doi:10.1038/s41589-023-01514-z
111. Conway, L. P.; Jadhav, A. M.; Homan, R. A.; Li, W.; Rubiano, J. S.; Hawkins, R.; Lawrence, R. M.; Parker, C. G. *Chem. Sci.* **2021,** *12,* 7839–7847. doi:10.1039/d1sc01360b
112. Szychowski, J.; Mahdavi, A.; Hodas, J. J. L.; Bagert, J. D.; Ngo, J. T.; Landgraf, P.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *J. Am. Chem. Soc.* **2010,** *132,* 18351–18360. doi:10.1021/ja1083909
113. Woo, C. M.; Iavarone, A. T.; Spiciarich, D. R.; Palaniappan, K. K.; Bertozzi, C. R. *Nat. Methods* **2015,** *12,* 561–567. doi:10.1038/nmeth.3366
114. Wang, J.; Zhang, C.-J.; Zhang, J.; He, Y.; Lee, Y. M.; Chen, S.; Lim, T. K.; Ng, S.; Shen, H.-M.; Lin, Q. *Sci. Rep.* **2015,** *5,* 7896. doi:10.1038/srep07896
115. Flaxman, H. A.; Miyamoto, D. K.; Woo, C. M. *Curr. Protoc. Chem. Biol.* **2019,** *11,* e75. doi:10.1002/cpch.75
116. Broncel, M.; Serwa, R. A.; Ciepla, P.; Krause, E.; Dallman, M. J.; Magee, A. I.; Tate, E. W. *Angew. Chem., Int. Ed.* **2015,** *54,* 5948–5951. doi:10.1002/anie.201500342
117. Wright, M. H.; Clough, B.; Rackham, M. D.; Rangachari, K.; Brannigan, J. A.; Grainger, M.; Moss, D. K.; Bottrill, A. R.; Heal, W. P.; Broncel, M.; Serwa, R. A.; Brady, D.; Mann, D. J.; Leatherbarrow, R. J.; Tewari, R.; Wilkinson, A. J.; Holder, A. A.; Tate, E. W. *Nat. Chem.* **2014,** *6,* 112–121. doi:10.1038/nchem.1830
118. Kallemeijn, W. W.; Lanyon-Hogg, T.; Panyain, N.; Goya Grocin, A.; Ciepla, P.; Morales-Sanfrutos, J.; Tate, E. W. *Nat. Protoc.* **2021,** *16,* 5083–5122. doi:10.1038/s41596-021-00601-6
119. Heal, W. P.; Jovanovic, B.; Bessin, S.; Wright, M. H.; Magee, A. I.; Tate, E. W. *Chem. Commun.* **2011,** *47,* 4081–4083. doi:10.1039/c0cc04710d
120. Paioti, P. H. S.; Lounsbury, K. E.; Romiti, F.; Formica, M.; Bauer, V.; Zandonella, C.; Hackey, M. E.; del Pozo, J.; Hoveyda, A. H. *Nat. Chem.* **2024,** *16,* 426–436. doi:10.1038/s41557-023-01386-9
121. Nickel, S.; Serwa, R. A.; Kaschani, F.; Ninck, S.; Zweerink, S.; Tate, E. W.; Kaiser, M. *Chem. – Eur. J.* **2015,** *21,* 10721–10728. doi:10.1002/chem.201500934
122. Kalesh, K. A.; Clulow, J. A.; Tate, E. W. *Chem. Commun.* **2015,** *51,* 5497–5500. doi:10.1039/c4cc09527h
123. Desai, H. S.; Yan, T.; Yu, F.; Sun, A. W.; Villanueva, M.; Nesvizhskii, A. I.; Backus, K. M. *Mol. Cell. Proteomics* **2022,** *21,* 100218. doi:10.1016/j.mcpro.2022.100218
124. Yang, K.; Whitehouse, R. L.; Dawson, S. L.; Zhang, L.; Martin, J. G.; Johnson, D. S.; Paulo, J. A.; Gygi, S. P.; Yu, Q. *Cell Chem. Biol.* **2024,** *31,* 565–576.e4. doi:10.1016/j.chembiol.2023.11.015
125. Müller, T.; Kalxdorf, M.; Longuespée, R.; Kazdal, D. N.; Stenzinger, A.; Krijgsveld, J. *Mol. Syst. Biol.* **2020,** *16,* e9111. doi:10.15252/msb.20199111

126. Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. *Nat. Protoc.* **2019,** *14,* 68–85. doi:10.1038/s41596-018-0082-x

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.199