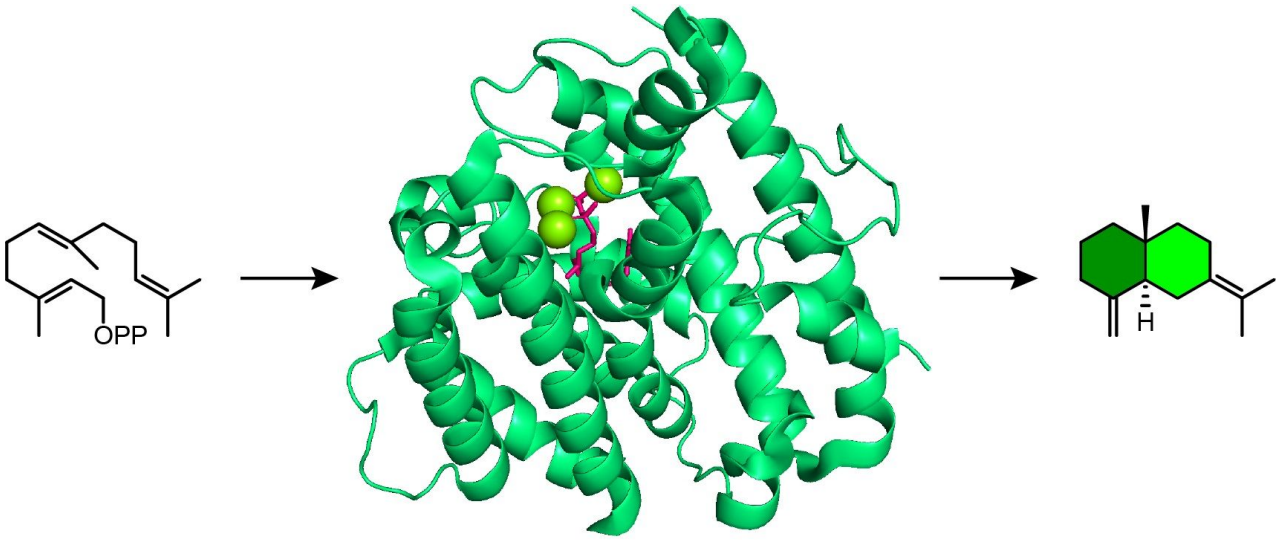
BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Enzymes in biosynthesis

Edited by Jeroen S. Dickschat



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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Enzymes in biosynthesis

Jeroen S. Dickschat

Editorial

Open Access

Address:
Kekulé-Institut für Organische Chemie und Biochemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Gerhard-Domagk-Straße 1, 53121 Bonn, Germany

Email:
Jeroen S. Dickschat - dickschat@uni-bonn.de

Keywords:
biosynthesis; enzymes in biosynthesis

Beilstein J. Org. Chem. **2022,** *18,* 1131–1132.
https://doi.org/10.3762/bjoc.18.116

Received: 05 August 2022
Accepted: 24 August 2022
Published: 30 August 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Enzymes are fascinating biocatalysts that can accelerate remarkable transformations in nature. Some of the most interesting transformations catalyzed by enzymes are known from the biosynthetic pathways towards natural products. For instance, class I terpene synthases can convert highly complex transformations of an acyclic precursor, such as farnesyl or geranylgeranyl diphosphate, into sesqui- or diterpenes, respectively. As has been described recently, even farnesylfarnesyl diphosphate can be converted into triterpenes, a substance class that was previously believed to originate exclusively from squalene by class II terpene synthases [1]. These conversions proceed through multistep cationic cascade reactions and usually produce a polycyclic terpene hydrocarbon or alcohol with multiple stereogenic centers. While these transformations require only a single enzyme, polyketide and nonribosomal peptide biosyntheses are catalyzed by megasynthases that follow an assembly line logic, with individual domains for each single step [2]. Furthermore, the domains are organized into modules, each of which is responsible for the incorporation of one extender unit into the growing polyketide or peptide chain. With our knowledge today, the function of these large enzyme factories is easier to read than for terpene synthases, the functions of which are difficult to predict, but their size makes the megasynthases much more difficult to handle in the laboratory. Besides these core enzymes of the biosynthetic machineries to some of the most important classes of natural products, nature has evolved a large number of enzyme classes for more specific transformations, including cytochromes P450 or α-ketoglutarate-dependent dioxygenases for late-stage oxidations and transferases for the attachment of sugar units, acyl, or methyl groups. Moreover, some enzymes can catalyze reactions that were first known from synthetic chemistry, e.g., pericyclases can promote pericyclic reactions such as [4 + 2]-cycloaddition, also known as Diels–Alder reaction [3]. In fact, most named reactions in organic chemistry originally discovered by synthetic chemists have an analogy in nature, requiring a sophisticated enzymology [4]. Recent developments show us that there is still much more to discover, e.g., altemicidin was shown to be enzymatically constructed from NAD^+ and SAM that usually serve as enzyme cosubstrates in redox transformations and methylations but are rarely used to construct the molecular scaffolds of natural products [5].

During the past two decades, large amounts of genome information from thousands of organisms have become available. This allows scientists today to gain direct access to the encoded catalysts through expression in easy-to-handle heterologous hosts, such as *Escherichia coli* or *Saccharomyces cerevisiae*. Besides in vitro studies with purified enzymes, heterologous expres-

sions of whole pathways for the production of compounds is possible [6]. Enzyme mechanisms can be addressed through structure-based site-directed mutagenesis, which may also lead to novel products [7]. An alternative approach is offered by computational chemistry, which is ideally performed in combination with experimental verification of the computational results, e.g., through the enzymatic conversion of isotopically labelled compounds [8].

This thematic issue will cover all different aspects of studying the roles of enzymes in the biosynthesis of natural products. Also contributions showing the application of enzymes in synthetic organic chemistry will be welcome. I am grateful to all colleagues who have contributed to this issue and to the Editorial Team of the Beilstein-Institut for their professional support. I wish the readers of this issue some stimulating new insights into enzyme research in natural product biosynthesis.

Jeroen S. Dickschat

Bonn, August 2022

ORCID® iDs

Jeroen S. Dickschat - https://orcid.org/0000-0002-0102-0631

References

1. Tao, H.; Lauterbach, L.; Bian, G.; Chen, R.; Hou, A.; Mori, T.; Cheng, S.; Hu, B.; Lu, L.; Mu, X.; Li, M.; Adachi, N.; Kawasaki, M.; Moriya, T.; Senda, T.; Wang, X.; Deng, Z.; Abe, I.; Dickschat, J. S.; Liu, T. *Nature* **2022,** *606,* 414–419. doi:10.1038/s41586-022-04773-3
2. Weissman, K. J. *Beilstein J. Org. Chem.* **2017,** *13,* 348–371. doi:10.3762/bjoc.13.39
3. Jamieson, C. S.; Ohashi, M.; Liu, F.; Tang, Y.; Houk, K. N. *Nat. Prod. Rep.* **2019,** *36,* 698–713. doi:10.1039/c8np00075a
4. Lin, C.-I.; McCarty, R. M.; Liu, H.-w. *Angew. Chem., Int. Ed.* **2017,** *56,* 3446–3489. doi:10.1002/anie.201603291
5. Barra, L.; Awakawa, T.; Shirai, K.; Hu, Z.; Bashiri, G.; Abe, I. *Nature* **2021,** *600,* 754–758. doi:10.1038/s41586-021-04214-7
6. Liu, C.; Minami, A.; Ozaki, T.; Wu, J.; Kawagishi, H.; Maruyama, J.-i.; Oikawa, H. *J. Am. Chem. Soc.* **2019,** *141,* 15519–15523. doi:10.1021/jacs.9b08935
7. Srivastava, P. L.; Escorcia, A. M.; Huynh, F.; Miller, D. J.; Allemann, R. K.; van der Kamp, M. W. *ACS Catal.* **2021,** *11,* 1033–1041. doi:10.1021/acscatal.0c04647
8. Wang, Y.-H.; Xu, H.; Zou, J.; Chen, X.-B.; Zhuang, Y.-Q.; Liu, W.-L.; Celik, E.; Chen, G.-D.; Hu, D.; Gao, H.; Wu, R.; Sun, P.-H.; Dickschat, J. S. *Nat. Catal.* **2022,** *5,* 128–135. doi:10.1038/s41929-022-00735-0

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Structural basis for endoperoxide-forming oxygenases

Takahiro Mori*[1,2,3] and Ikuro Abe*[1,2]

Review

Open Access

Address:
[1]Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, [2]Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan and [3]PRESTO, Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan

Email:
Takahiro Mori* - tmori@mol.f.u-tokyo.ac.jp; Ikuro Abe* - abei@mol.f.u-tokyo.ac.jp

* Corresponding author

Keywords:
biosynthesis; endoperoxide; enzyme; natural products; X-ray crystallography

Beilstein J. Org. Chem. **2022,** *18,* 707–721.
https://doi.org/10.3762/bjoc.18.71

Received: 03 May 2022
Accepted: 10 June 2022
Published: 21 June 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

Endoperoxide natural products are widely distributed in nature and exhibit various biological activities. Due to their chemical features, endoperoxide and endoperoxide-derived secondary metabolites have attracted keen attention in the field of natural products and organic synthesis. In this review, we summarize the structural analyses, mechanistic investigations, and proposed reaction mechanisms of endoperoxide-forming oxygenases, including cyclooxygenase, fumitremorgin B endoperoxidase (FtmOx1), and the asnovolin A endoperoxygenase NvfI.

Introduction

Endoperoxide-containing compounds form a large group of natural products with cyclic peroxide structures [1-5]. These compounds are widely distributed in nature, and many endoperoxide containing alkaloids, terpenoids, and polyketides have been isolated from plants, animals, bacteria, fungi, and other organisms (Figure 1) [6,7]. Because of the high reactivity of the cyclic peroxide O–O bond, these compounds exhibit various biological activities [1-5]. For example, natural and (semi)synthetic endoperoxides with wide structural diversity show antimalarial activity against *Plasmodium falciparum* malaria. In this case, the reductive activation of the endoperoxide ring with the homolytic cleavage of the O–O bond leads to the generation of carbon-centered free radicals that play important roles in antimalarial activity by damaging membranes, inhibiting nucleic acid and protein syntheses, and so on [8,9].

The best studied endoperoxide-containing compound is probably prostaglandin H2 (PGH2), the common precursor of biologically active prostanoids [10-12]. Artemisinin, the antimalarial agent isolated from the plant *Artemisia annua* [8,13,14], and ergosterol peroxides with anticancer and antiviral activities, identified in many fungi, algae, lichens, and plants, also belong

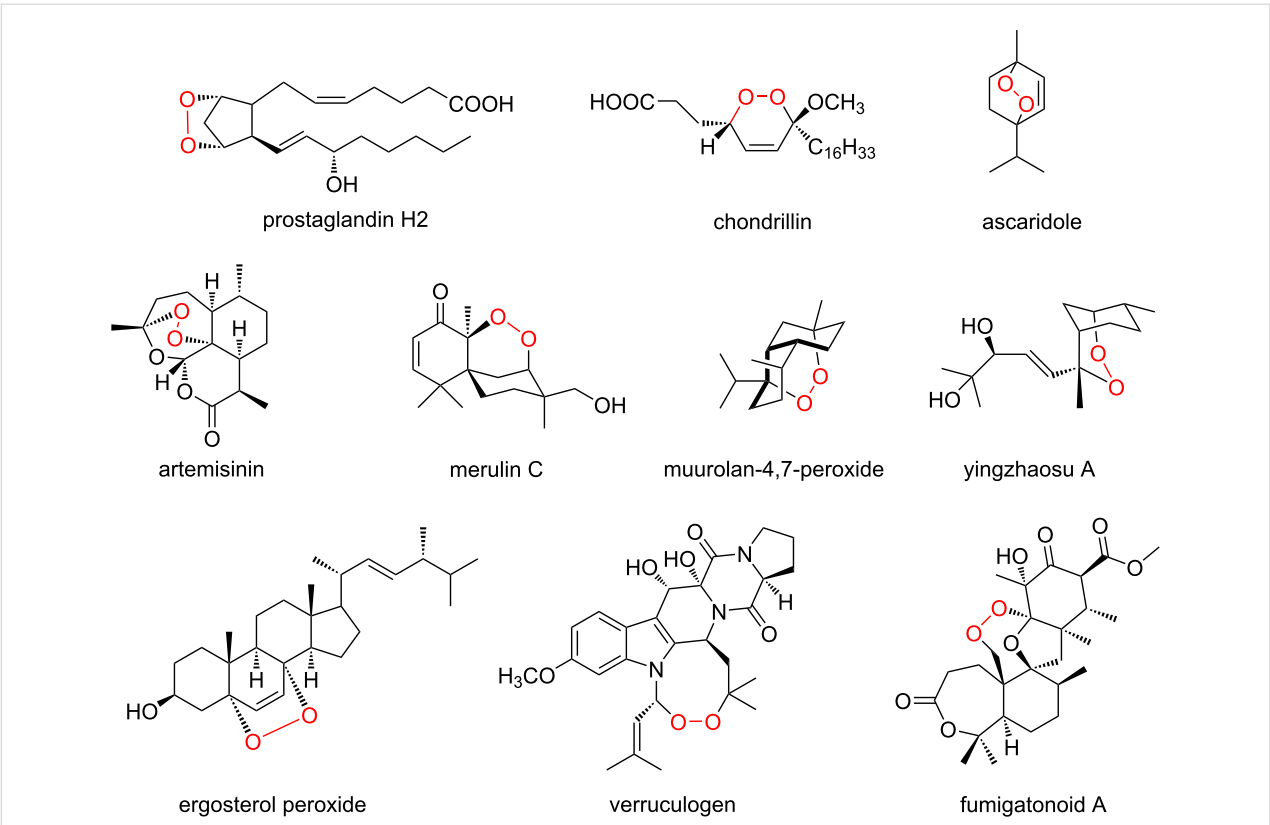


Figure 1: Examples of endoperoxide-containing natural products.

to this group [15-17]. Due to the significant biological activities of the endoperoxide-containing natural products, numerous synthetic analyses and biosynthesis of endoperoxide compounds have been reported [18-21]. In some cases; e.g., in the biosynthesis of artemisinin and ergosterol peroxides, a reactive oxygen species (ROS) such as singlet oxygen, which is generated by photosensitizers or visible light, non-enzymatically reacts with the biosynthetic intermediates to produce endoperoxide structures [16,22,23]. However, over the past three decades, only a few endoperoxide-forming enzymes have been identified, including the cyclooxygenases in the biosynthesis of prostaglandins [24], iodide peroxidase in the biosynthesis of ascaridole [25], fumitremorgin B endoperoxidase (FtmOx1) in the biosynthesis of verruculogen [26], and asnovolin A endoperoxygenase NvfI in the biosynthesis of novofumigatonin [27,28]. Among them, although a soluble iodide peroxidase has been isolated from *Chenopodium ambrosioides*, the sequence, structural, and mechanistic analyses have not been vigorously pursued [24].

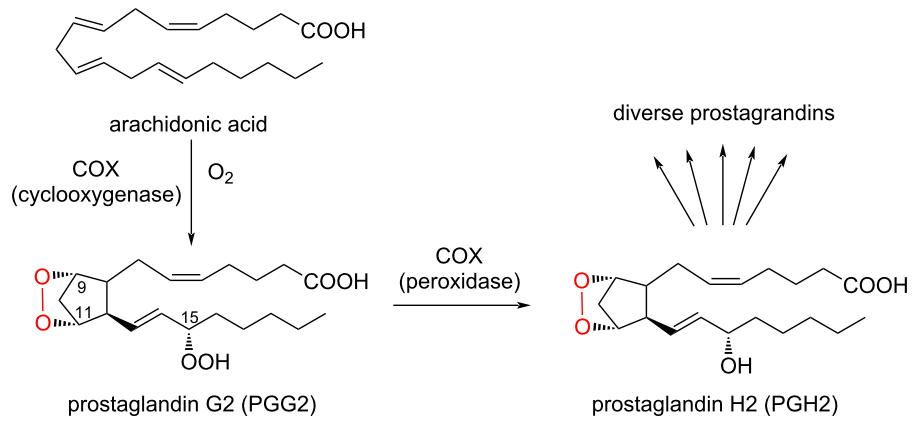
While many review articles on the structure determination, biological analysis, and synthesis of endoperoxide-containing natural products have been reported [1-7,11,16], the details of the complex biosynthetic enzymes producing these compounds have remained enigmatic. Therefore, this review will focus on the enzymatic synthesis of endoperoxide natural products, by summarizing the recent structural and mechanistic analyses of endoperoxide formation reactions by cyclooxygenases, FtmOx1, and NvfI.

Review

COX: Heme-dependent cyclooxygenases in the biosynthesis of prostaglandins

Enzyme reaction of COXs

The cyclooxygenases are the best studied and understood oxygenases among the mammalian oxygenases [29,30]. Mammals have two cyclooxygenase isoforms, COX-1 and COX-2 [31,32], which share ≈60% amino acid identity [33]. Both isoforms catalyze the incorporation of two oxygen atoms into arachidonic acid (AA) to form an endoperoxide between C9 and C11 and a peroxide at C15 to generate prostaglandin G2 (PGG2) (Scheme 1) [24,34]. Subsequently, the 15-hydroperoxide in PGG2 is reduced to produce PGH2. PGH2 is metabolized by downstream enzymes to yield a series of prostaglandins, which play important roles in inflammatory responses [35-37]. Although the active site architectures of COX-1 and COX-2 are not completely identical, the reaction



Scheme 1: Reactions of COXs.

mechanisms and catalytic residues are well conserved. While COX-1 is a constitutive enzyme present in most tissues, COX-2 is an isoenzyme induced in response to tumor promoters, growth factors, and cytokines [38-40]. Therefore, many COX-2 selective inhibitors are clinically used for treatments of inflammation, cancers, and pain [41-43].

Crystal structures of COXs

The structural basis for the di-peroxide formation reaction by COXs has been substantially elucidated by electron paramagnetic resonance (EPR), kinetic analysis, X-ray crystallography, and mutagenesis experiments [24,44,45]. The structural analyses of mammalian COXs revealed that COX-1 and COX-2 are both glycosylated by post-translational modifications [46-51]. These enzymes form homodimers, and the overall structures of COX-1 and COX-2 superimpose well, with root mean square values of ≈0.9 Å. Each COX monomer contains three domains, including the epidermal growth factor (EGF) domain, the membrane binding domain, and the catalytic domain (Figure 2A) [46-51]. The catalytic domain possesses two active sites, the cyclooxygenase- and heme-dependent peroxidase-sites, which are physically separated. The peroxidase-site activates the catalytic tyrosine residue, while the cyclooxygenase-site catalyzes the formation of di-peroxides. The active site of the peroxidase-site contains a heme cofactor in the solvent-exposed cleft on the opposite side of the membrane binding domain. Although the heme cofactor is located in the peroxidase-site and the active site of peroxidase-site and cyclooxygenase-site are separated, the heme cofactor plays a critical role in both of peroxidase reaction and cyclooxygenase reaction. The active site of the cyclooxygenase-site consists of a deep, L-shaped hydrophobic cavity, referred to as the cyclooxygenase channel, and the entrance of its active site is located on the opposite side from that of the peroxidase-site (Figure 2A).

In the complex structures of COXs with AA (PDB ID: 1CVU, 1DIY, and 3HS5), the carboxylate of the AA is located near the entrance of the cyclooxygenase channel, and the tail of the fatty acyl chain is bound deeply into the narrow hydrophobic channel (Figure 2B and 2C) [48,52,53]. Structural and biochemical studies of ovine COX-1 indicated that the ionic interaction between the carboxylate of AA and Arg120 is required for the reaction (Figure 2B) [54-56]. In contrast, this interaction is not essential in COX-2, suggesting that the hydrophobic interactions between the acyl chain and the active site residues in the cyclooxygenase channel are important for AA binding to COX-2 (Figure 2C) [57,58]. This is one of the major structural differences between COX-1 and COX-2.

Catalytic residue in the cyclooxygenase reaction

The formation of a tyrosyl radical during the catalytic cycle was proved by EPR and kinetic analyses [59-61]. Moreover, chemical and molecular biology analyses and a mutagenesis experiment identified the position of the tyrosyl radical. Treatment of the enzymes with tetranitromethane, a reagent for tyrosyl residue nitration, eliminated the cyclooxygenase activity, while the activity was not abolished in the presence of the cyclooxygenase inhibitor indomethacin, indicating that the tyrosine residue(s) is the catalytic center for endoperoxide formation. The sequence analysis of the tetranitromethane-treated enzyme indicated that three tyrosine residues, Tyr355, Tyr385, and Tyr417, were only nitrated in the absence of indomethacin (Figure 2A) [62]. A subsequent mutagenesis study of these residues indicated that the Tyr385 residue plays the catalytic role in the cyclooxygenase reaction. In the crystal structure, the catalytic Tyr385 residue is located at the interface between the active regions of the peroxidase-site and cyclooxygenase-site (Figure 2) [46-51]. Furthermore, the C13 of AA is located near the catalytic residue Tyr385 in the crystal structure, indicating

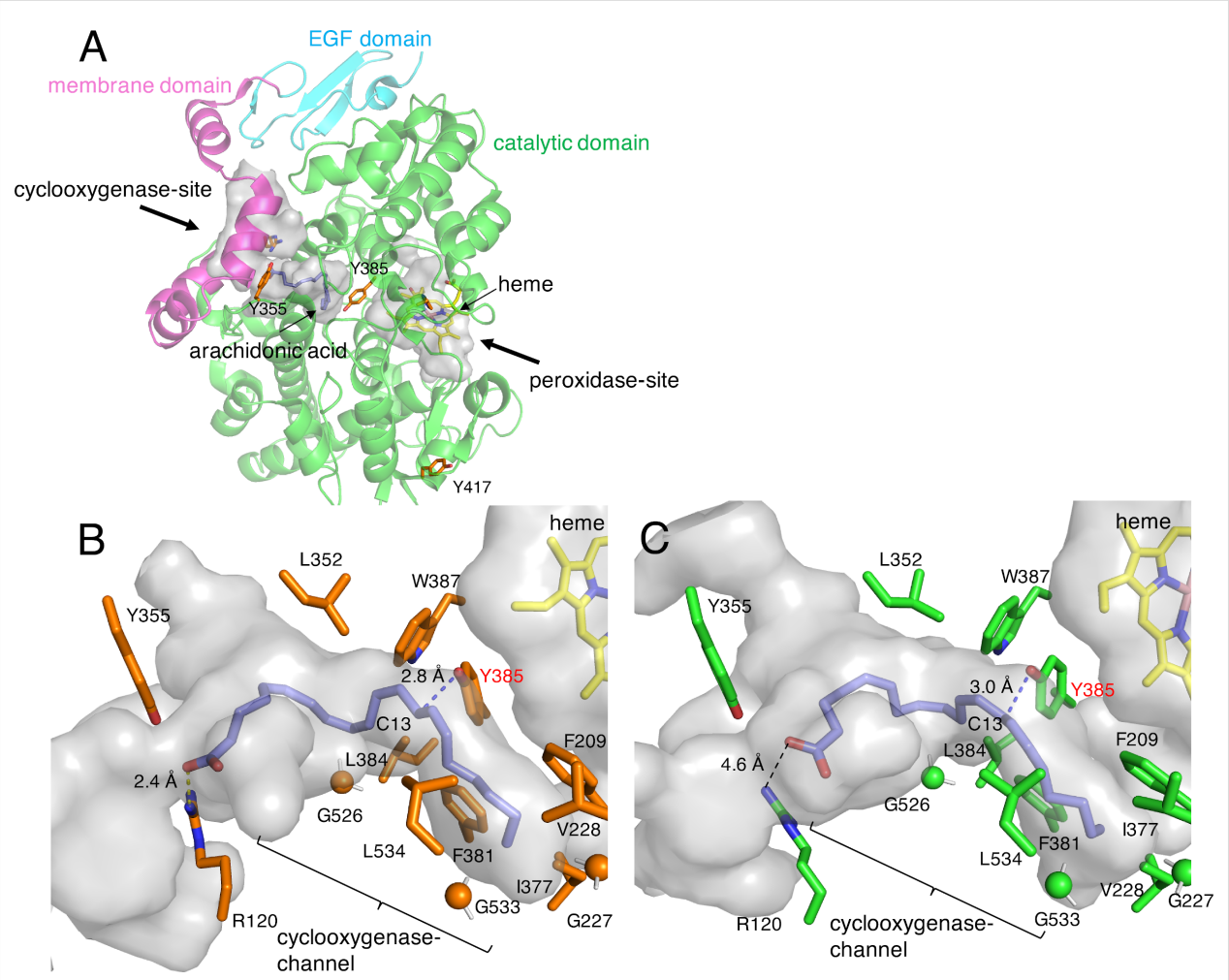


Figure 2: Structures of COXs [52,53]. (A) The overall structure of ovine COX-1. (B and C) Comparison of the cyclooxygenase sites of (B) COX-1 and (C) COX-2 in complex with AA. Yellow dashed lines show hydrogen bond interactions. The blue dashed line shows the distance between Tyr385 and C13 of AA.

that a tyrosyl radical abstracts the pro-*S* hydrogen atom from C13 of AA (Figure 2B and 2C) [48,52,53].

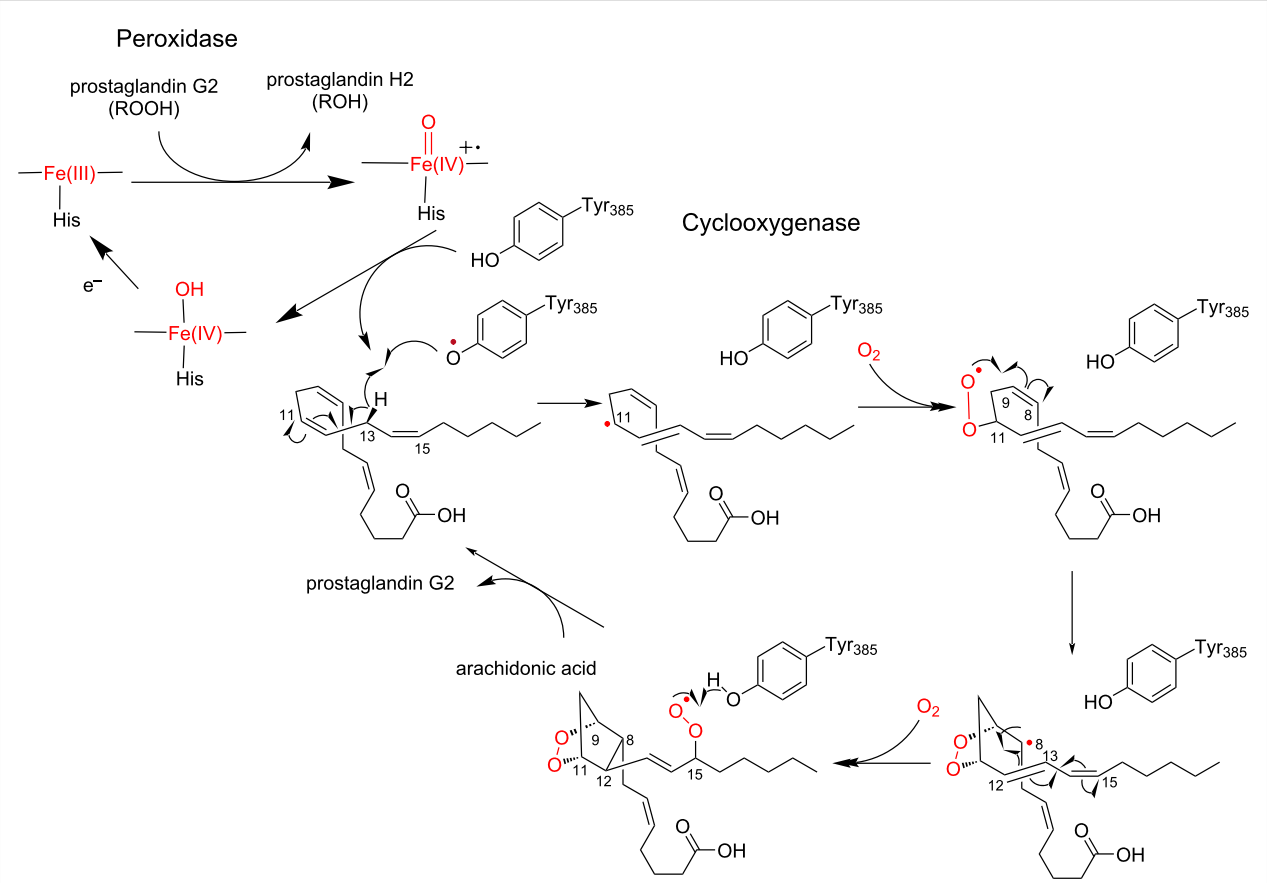
Mechanism of the cyclooxygenase reaction

The enzyme reaction is initiated upon Tyr385 activation by the oxyferryl heme cation radical, which is generated through the two-electron reduction of PGG2, to form a tyrosyl radical in the active site of the cyclooxygenase-site (Scheme 2) [24,34,63]. The tyrosyl radical then abstracts a C13-pro-*S* hydrogen atom from AA to produce the arachidonoyl radical, which is delocalized over C11 to C15. An oxygen molecule reacts with the C11 radical intermediate to produce a C11-peroxyl radical. The subsequent 5-exo cyclization with a double bond at C8–C9 forms the C8 radical intermediate with the endoperoxide bridge between C9 and C11. The cyclization between C8 and C12 forms a bicyclic radical intermediate, in which the spin density is distributed over C13–C15. The allyl radical at C15 reacts with a second molecular oxygen to afford the C15 peroxyl radical. Finally, the transfer of a hydrogen atom from the catalytic Tyr385 residue quenches the C15 peroxyl radical to yield PGG2 and a tyrosyl radical for the next round of the enzyme reaction. The released PGG2 is accepted by the peroxidase active site, and the 15-hydroperoxyl radical of PGG2 is reduced to generate PGH2.

FtmOx1: Nonheme iron-dependent endoperoxygenase in the biosynthesis of verruculogen

Enzyme reaction of FtmOx1

Fumitremorgin B endoperoxidase (FtmOx1) from *Aspergillus fumigatus* is the first identified nonheme iron and 2-oxoglutarate (Fe/2OG)-dependent endoperoxidase that catalyzes the formation of an endoperoxide in the biosynthesis of verruculogen [26].



Scheme 2: Proposed reaction mechanisms of COXs [24].

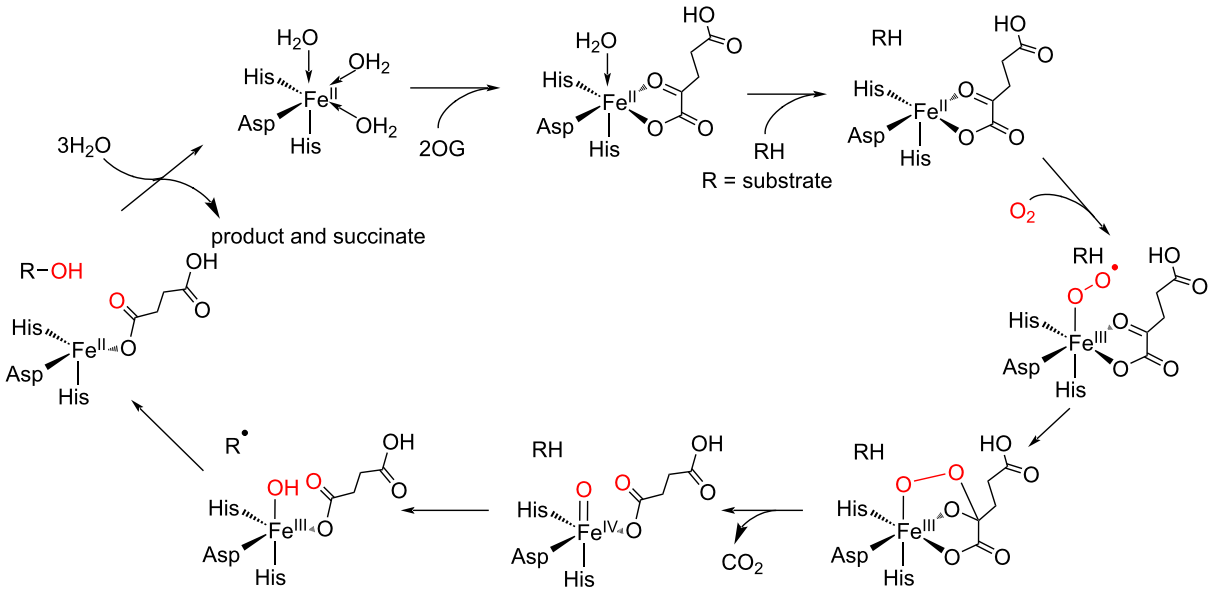
Fe/2OG oxygenases utilize Fe(II) as a cofactor and 2OG and O_2 as co-substrates (Scheme 3) [64-67]. The Fe(II) is coordinated by the conserved 2-His-1-Asp residues and a 2OG in the active site. The binding of a substrate in the active site, where it is close to the Fe(II) center, provides a coordination site for O_2. Subsequently, the oxidative decarboxylation of 2OG generates a highly reactive Fe(IV)=O species and a succinate byproduct. This Fe(IV)=O abstracts a hydrogen atom from an aliphatic C–H bond of the substrate to generate a radical intermediate. When the enzyme catalyzes the hydroxylation reaction, the radical reacts with the Fe(III)-OH species to form a hydroxylated product.

The functional analysis of FtmOx1 indicated that the enzyme accepts fumitremorgin B as a substrate and catalyzes the formation of verruculogen and a C13-oxo product, through the installation of an endoperoxide bridge between C21 and C27 of fumitremorgin B and oxidation of the C13-hydroxy group of verruculogen, respectively (Scheme 4) [68-71]. The single-turnover enzyme reaction of FtmOx1 in the absence of reductants (e.g., ascorbate) indicated that FtmOx1 consumes two oxygen molecules to generate these products in the catalytic cycle. Thus, an O_2 molecule directly reacts with the radical intermediate of fumitremorgin B, and is incorporated without cleavage of the O–O bond in the enzyme reaction catalyzed by FtmOx1.

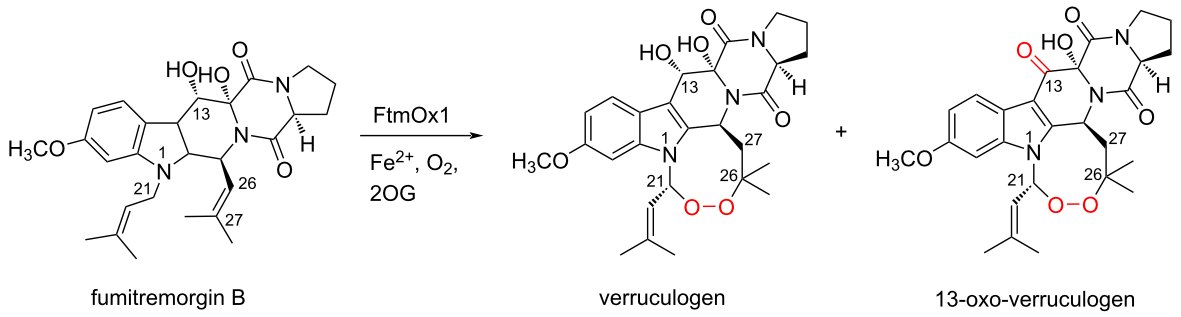
Biochemical and biophysical analyses of FtmOx1 have been reported by many different groups, and the structures of FtmOx1 wild type and variants have been solved in the apo and complex forms with 2OG or with 2OG and fumitremorgin B (PDB IDs: 4Y5T, 4Y5S, 6OXH, 6OXJ, 7DE0, 7ETL, and 7ETK) [68-71]. Based on these structural and functional analyses, the reaction mechanism of FtmOx1, and especially the role of the active site tyrosine residues in the catalysis, are heatedly debated.

COX-like mechanism of FtmOx1

The first crystal structures of FtmOx1 in the apo form and in complex with 2OG were reported in 2015 [68] (this article got retracted [69] in 2021). The overall structure of FtmOx1 exists as a functional homodimer and possesses the double-stranded β-helix (DSBH) fold observed in typical Fe/2OG-dependent oxygenases (Figure 3A). The Fe(II) is coordinated by His129, Asp131, His205, and 2OG in the binary complex structure of



Scheme 3: General reaction mechanism of Fe/2OG oxygenases.



Scheme 4: Reaction of FtmOx1 [68-71].

FtmOx1 with 2OG. In this structure, Tyr224 is close to the putative oxygen binding site (Figure 3B). Furthermore, the variants in which Tyr224 is substituted with Ala or Phe generates an N-1 dealkylated product, which is non-enzymatically created from C21 hydroxylated products, as a major product. Based on these structural analyses, as well as an EPR analysis of the enzyme reaction of FtmOx1 and a transient ultraviolet–visible (UV–vis) absorption analysis, the group proposed that Tyr224 is involved in the catalytic mechanism of the FtmOx1-catalyzed endoperoxide formation reaction as an intermediary of hydrogen atom transfer (HAT), similar to Tyr385 in the COX reaction. In this COX-like reaction mechanism (Scheme 5), the Fe(IV)=O species oxidizes Tyr224 to form a tyrosyl radical, which abstracts a hydrogen atom from C21 of fumitremorgin B to generate a radical intermediate. The insertion of molecular oxygen at C21 produces a C21 peroxyl radical intermediate, which then reacts with the C26–C27 double bond on another prenyl group to generate the endoperoxide with a C26 radical intermediate. Finally, the radical is quenched by hydrogen atom donation from Tyr224, to form verruculogen and a tyrosyl radical for the next round of the reaction. The tyrosyl radical at Tyr224 is also involved in the oxidation of the C13-hydroxy group of verruculogen to a C13-keto product, in the absence of reductants. Thus, the catalytic Tyr224 in this mechanism plays a similar role to Tyr385 in the COX enzyme reaction. In the paper, the authors also reported the complex structure of FtmOx1 with fumitremorgin B (PDB ID: 4ZON, removed recently). However, re-examination of the electron density map indicated that the density is not fit for fumitremorgin B. Since only biochemical data do not conclusively support the mecha-

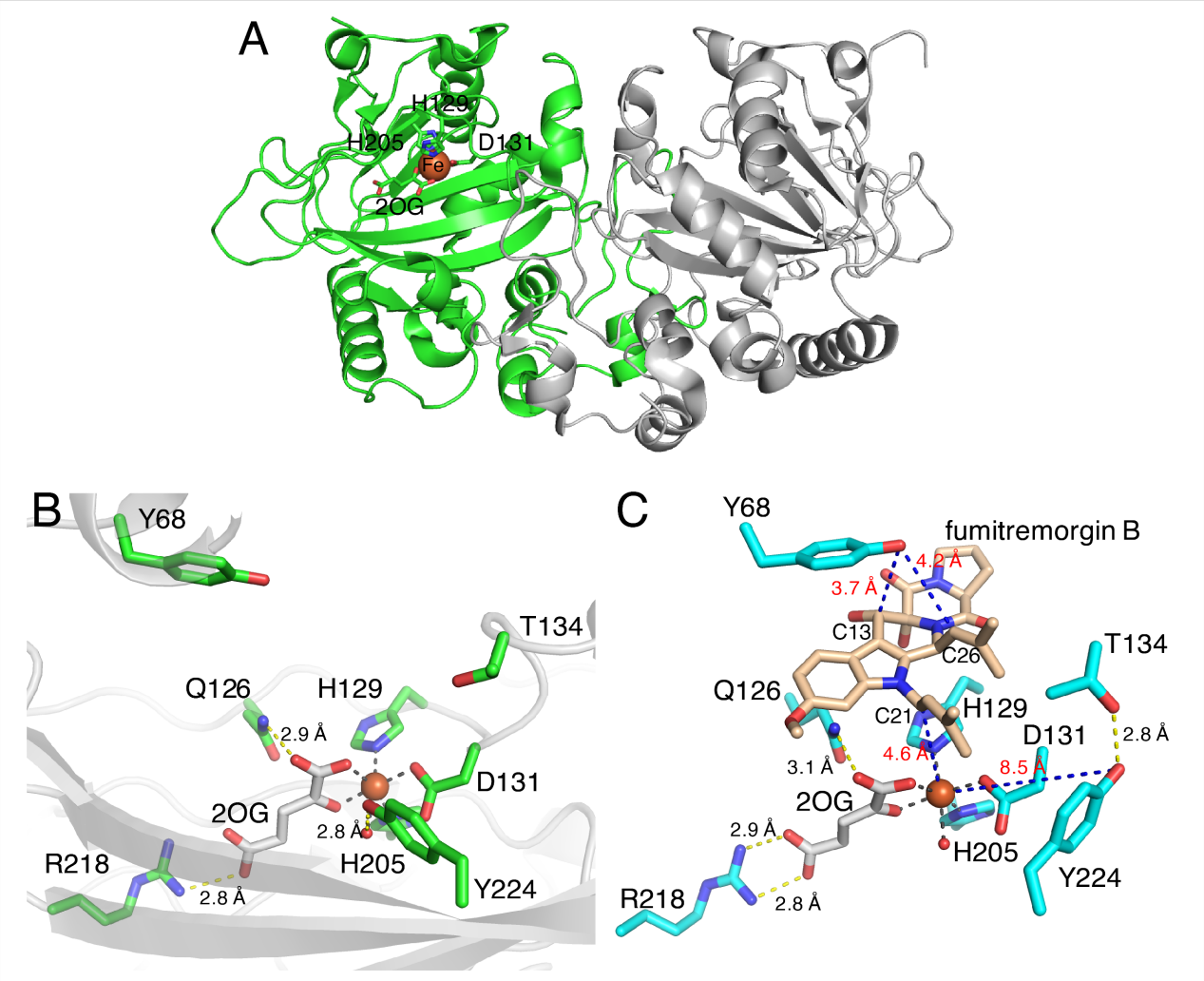


Figure 3: Structure of FtmOx1 [71]. (A) The FtmOx1 binary structure in complex with 2OG. (B and C) Comparison of the active site architectures between (B) the binary structure and (C) the ternary structure. Blue dashed lines show the distances between atoms and yellow dashed lines represent hydrogen bond interactions. Grey dashed lines show the coordination of the iron atom. Iron atoms are depicted by orange spheres.

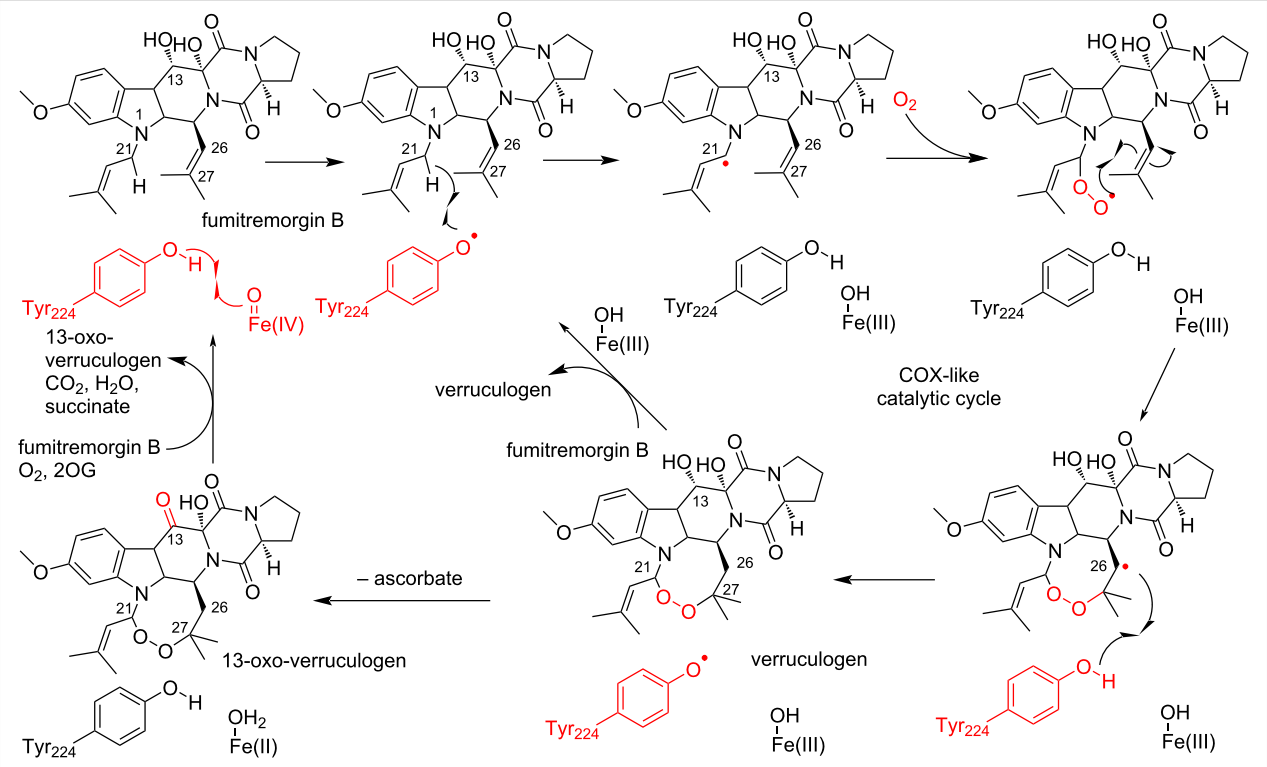
nistic role of Tyr224 in catalysis, one of the authors has agreed with the retraction, whereas the other authors stand by their data.

Recent docking and molecular dynamics analyses, as well as DFT calculations of the FtmOx1 reaction, suggested a modified mechanism [72-74]. In this mechanism, the tyrosyl radical at Tyr224 is generated by the Fe(IV)=O species, which abstracts a hydrogen atom at C21 to form an endoperoxide ring through the reaction with an O_2 molecule, as in the case of the COX-like mechanism. However, the radical quenching at C26 is achieved by reductants such as ascorbate, but not by Tyr224, in the final step of the reaction.

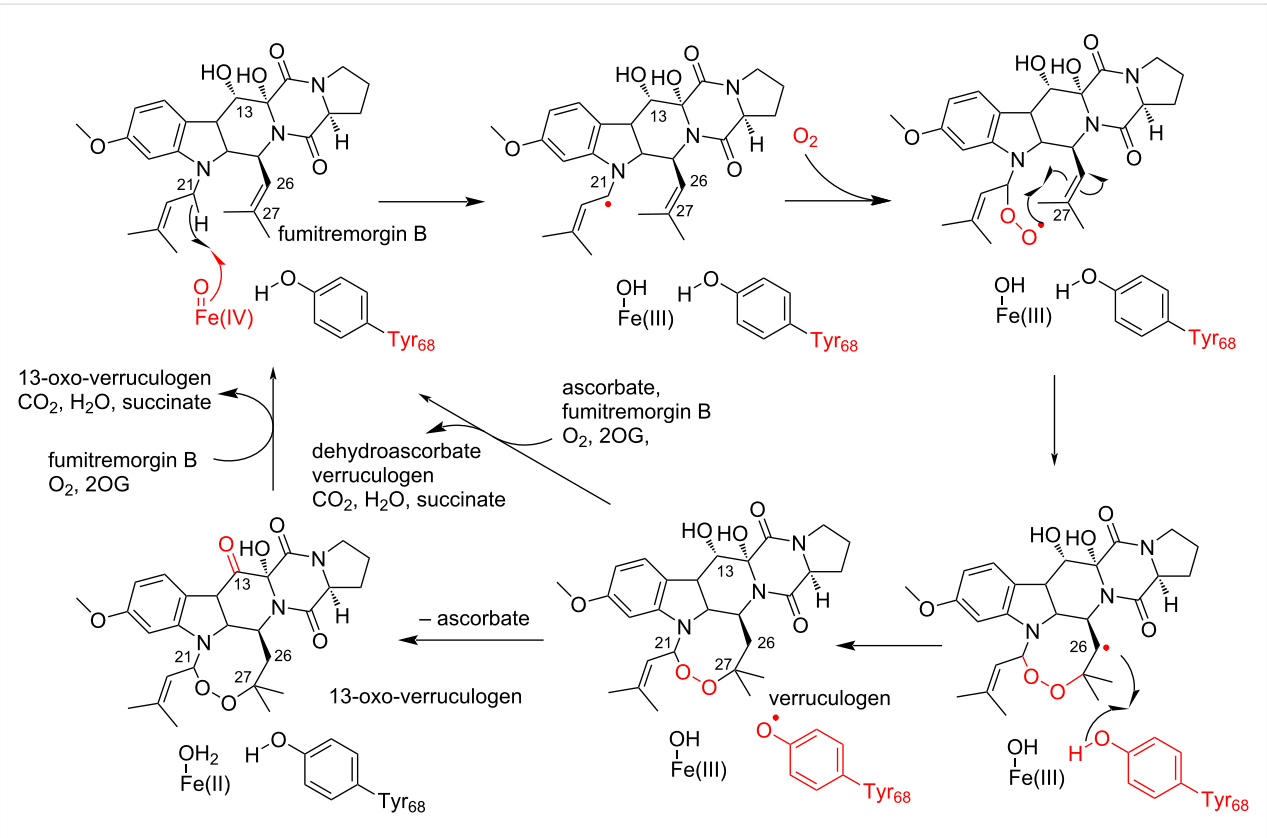
CarC-like mechanism of FtmOx1

In 2019, Bollinger and co-workers reported a detailed mechanistic study, including kinetic, UV–vis absorption, and EPR analyses of FtmOx1 wild type and its Y68F, Y74F, Y140F, and Y224F variants [70]. Their results revealed that the Y68F variant did not accumulate the initial tyrosyl radical and the formation of verruculogen was significantly decreased, while the Y224F variant showed similar reaction kinetics and verruculogen productivity to those of wild type FtmOx1. Interestingly, the Y68F variant generated an unidentified product, which was recently determined to be 26-hydroxyverruculogen [75].

Based on these observations, they proposed an alternative reaction mechanism. In this proposal, a tyrosyl radical is generated on Tyr68, instead of Tyr224, and this tyrosyl radical donates a hydrogen to the C26 radical intermediate, which is reminiscent of the carbapenem synthase (CarC) reaction mechanism. In this CarC-like mechanism (Scheme 6), the Fe(IV)=O species, but not the tyrosyl radical, first abstracts a hydrogen atom from C21



Scheme 5: Proposed COX-like mechanism of FtmOx1 [68].



Scheme 6: Proposed CarC-like mechanism of FtmOx1 [70].

to form a substrate radical intermediate. The following reaction with molecular oxygen and the formation of an endoperoxide bridge generate the C26 radical intermediate. Finally, HAT from Tyr68 produces verruculogen and a tyrosyl radical at Tyr68, which is quenched by reductants.

The ternary complex structure of FtmOx1 with 2OG and fumitremorgin B was recently reported by Zhou and co-workers [71]. Fumitremorgin B binds in the active site with a planar conformation, through hydrophobic and hydrophilic interactions (Figure 3C). While Tyr68 is located on the protein surface and solvent-exposed, the distance between C21 of fumitremorgin B and the iron center is 4.6 Å and the hydroxy group of Tyr68 is near C26 of fumitremorgin B (Figure 3C). Moreover, Tyr68 is located close to C13 of fumitremorgin B, at a distance of 3.7 Å. The comparison between the binary and ternary complex structures indicated that the Tyr224 residue rotates by 115° toward the opposite side of the iron center and forms a hydrogen bond with T134. The resulting distance between the Tyr224 hydroxy group and the iron center is 8.5 Å. These observations support the CarC-like mechanism, rather than the COX-like mechanism.

Recently, Bollinger and co-workers also reported the incorporation of non-canonical Tyr analogs, including 3-fluorotyrosine, 2,3-difluorotyrosine, 3,5-difluorotyrosine, 3-chlorotyrosine, and 4-aminophenylalanine, at Tyr68 or Tyr224 to analyze the functions of these tyrosine residues [75]. The transient-kinetic, UV–vis absorption, and EPR analyses of FtmOx1 variants containing non-canonical Tyr analogs also indicated that Tyr68, rather than Tyr224, acts as the catalytic residue in the FtmOx1 reaction, supporting the CarC-like mechanism.

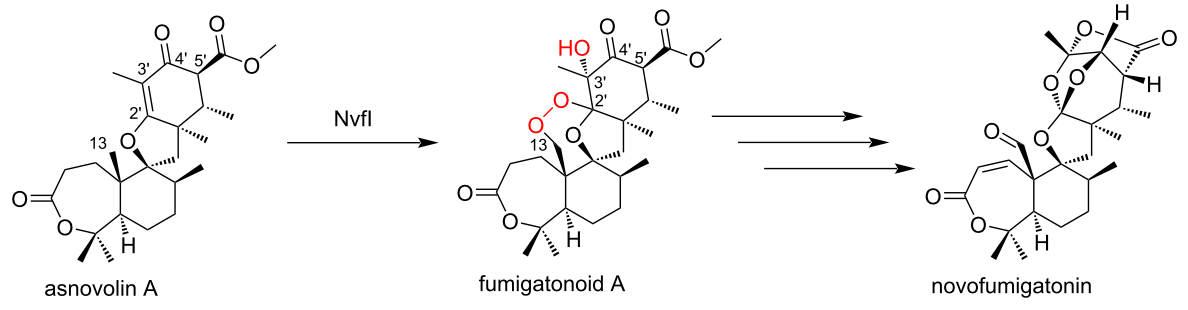
NvfI: Nonheme iron-dependent endoperoxygenase in the biosynthesis of novofumigatonin

Enzyme reaction of NvfI

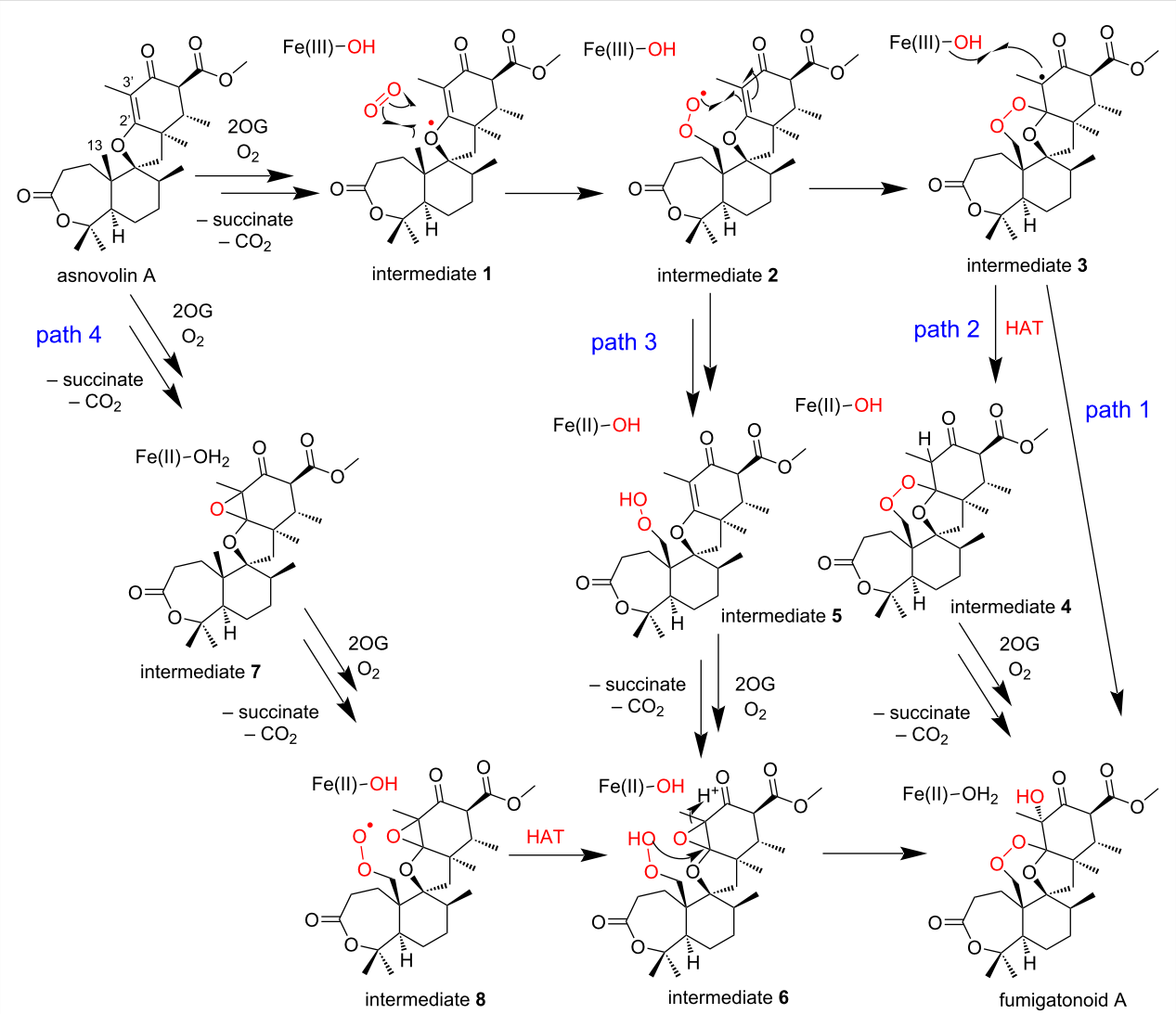
Asnovolin A endoperoxygenase NvfI is a second example of a 2OG-dependent endoperoxygenase, which is involved in the biosynthesis of novofumigatonin from *Aspergillus novofumigatus* IBT 1680611 [27,28]. The enzyme converts asnovolin A into fumigatonoid A, a biosynthetic intermediate of novofumigatonin, by introducing three oxygen atoms, including a hydroxy group at C3' and an endoperoxide bridge between C13 and C2' (Scheme 7). Although the in vitro assay of NvfI indicated that the enzyme is a 2OG-dependent endoperoxidase, NvfI shares a low amino acid sequence similarity with FtmOx1 (only 17%) and a phylogenetic analysis indicated that it is located in a different clade from FtmOx1.

Prior to the structure–function analysis of NvfI, four pathways were proposed for the formation of fumigatonoid A from asnovolin A (Scheme 8, path 1). The Fe(IV)=O species abstracts a hydrogen atom from C13 of asnovolin A to form radical intermediate **1**. Subsequently, it reacts with molecular oxygen to form a peroxyl radical intermediate **2**. The peroxyl radical attacks C2' to generate intermediate **3**, which contains an endoperoxide bridge and a C3' radical. Finally, the hydroxylation at C3' by the Fe(III)-OH species yields fumigatonoid A (path 2). At the stage of intermediate **3** in path 1, HAT from an active site residue or reductant to the C3' radical in intermediate **3** generates intermediate **4**. Then, the hydroxylation at C3' forms fumigatonoid A (path 3). The C13 peroxide intermediate **5** is generated from intermediate **2**, which is subsequently epoxidated at C2'–C3' to form intermediate **6**. The following cyclization reaction from the peroxide generates fumigatonoid A (path 4). The epoxide formation reaction occurs first at C2'–C3' of asnovolin A (intermediate **7**), and then the peroxide formation (intermediate **8**) and cyclization (intermediate **6**) reactions produce fumigatonoid A.

The stoichiometric analysis of 2OG and O_2 indicated that one equivalent of 2OG and two equivalents of O_2 are consumed to generate fumigatonoid A, suggesting that the installation of three oxygen atoms onto asnovolin A occurs in a single turnover of the enzyme reaction. Thus, pathways 2–4, which require two equivalents of 2OG for the formation of fumigatonoid A, are unlikely. Further experiments with $^{18}O_2$ and



Scheme 7: Reaction of NvfI [28].



Scheme 8: Possible reaction pathways leading to fumigatonoid A [28].

$H_2{}^{18}O$ suggested that three oxygen atoms are enzymatically incorporated into fumigatonoid A, in which the oxygen atoms of the endoperoxide are derived from the O_2 molecule and the C3' hydroxy group most likely originates from the solvent water. Although the oxygen atom in the hydroxylation reaction is usually from molecular oxygen, the oxygen atom in the Fe(III)-OH species can be exchanged with the solvent water [76]. Therefore, the fact that almost all of the oxygen atoms in the Fe(III)-OH species exchanged with the solvent in the enzyme reaction of NvfI suggested the presence of a long-lived C3' radical and the hydroxy ligand during the reaction.

Structure of NvfI

In the structural analysis of NvfI, three different active site conformations (open, partially closed, and closed) were observed (PDB IDs: 7DE2, 7EMZ, and 7ENB) (Figure 4A–C) [28]. The conformations of the loop regions between Ser122-Gly128 and Trp199-Pro209 were altered by soaking with the substrate. The substrate asnovolin A binds in the closed conformation through a hydrogen bond network with active site residues (Figure 4D). In this binding mode, C7' of the substrate is 4.2 Å away from the iron center, which is shorter than the distance between the iron center and C13 (6.5 Å). This spatial arrangement is not reasonable for the formation of fumiganoid A, because the initial step should be the abstraction of a C13 hydrogen atom by the Fe(IV)=O species. Considering the conformational changes of the active site and the docking simulation of the substrate in the active site of the partially closed conformation, the substrate binding mode in the crystal structure apparently shows a different stage of the reaction. The conformational changes of the loops, especially the flipping of Glu208, would contribute to al-

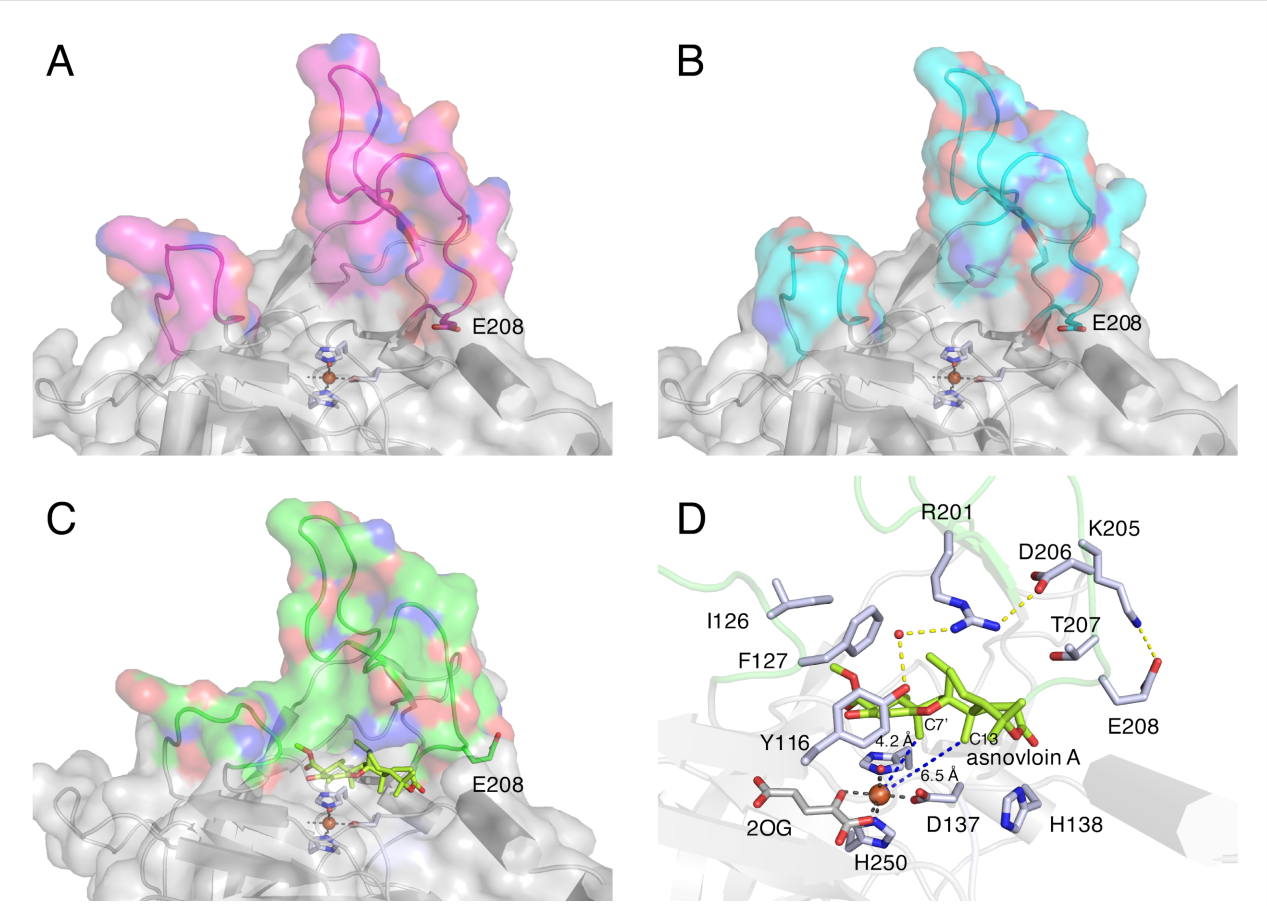
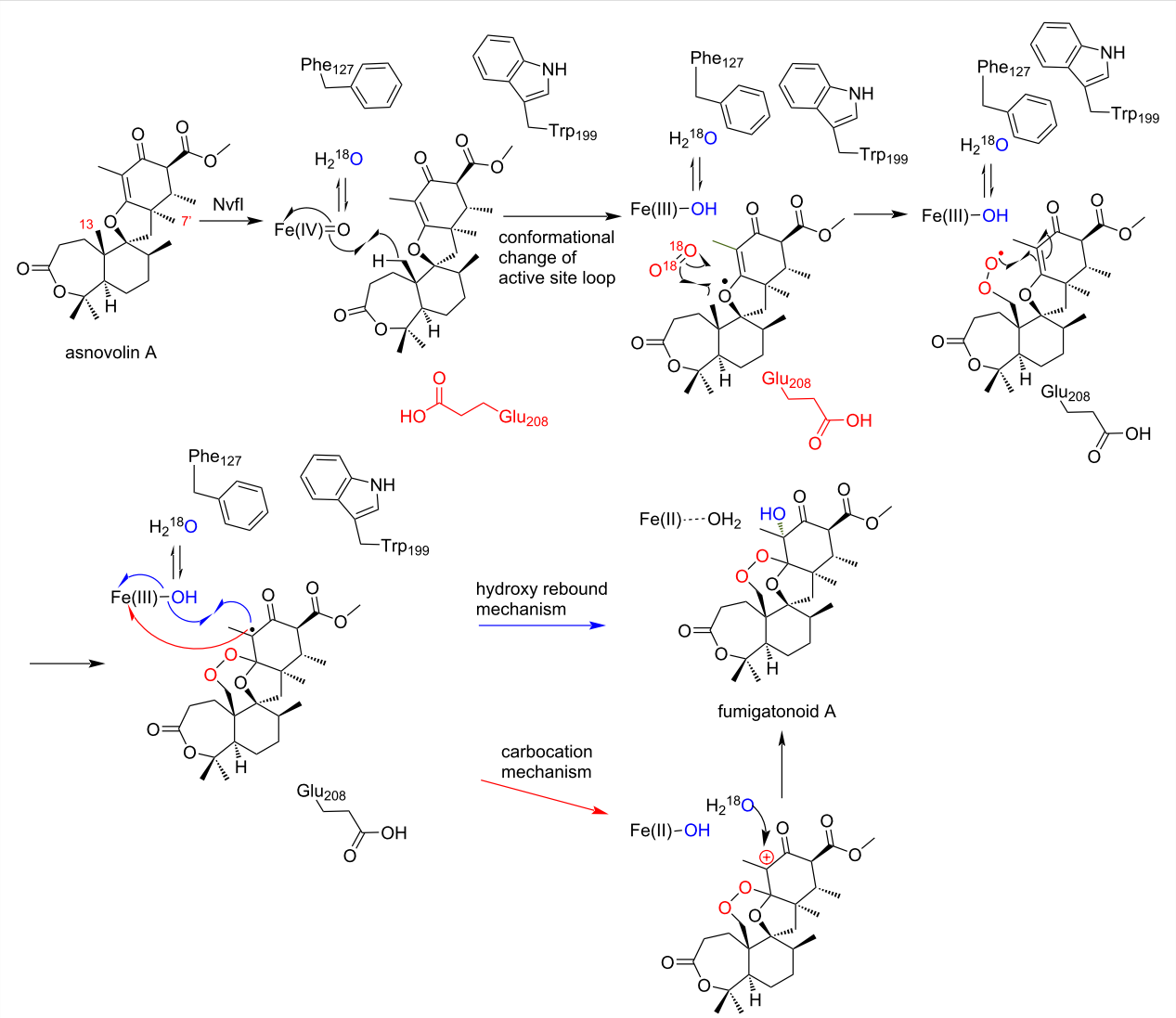


Figure 4: Structure of NvfI [28]. (A–C) Conformational changes of loop regions: (A) open conformation, (B) partially closed conformation, and (C) closed conformation of NvfI. (D) The active site of NvfI in complex with asnovolin A. Blue dashed lines show the distances between atoms and yellow dashed lines represent hydrogen bond interactions. Iron atoms are depicted by orange spheres.

terations of the binding mode of the substrate and the long-lived radical on the intermediate and Fe(III)-OH species (Figure 4A–C). Interestingly, a mutagenesis experiment of the active site residues of NvfI, including Tyr116, suggested that the enzyme does not employ any active site residues for the HAT step.

Reaction mechanism of NvfI

Based on these observations, we propose the mechanism of the NvfI-catalyzed endoperoxide formation reaction (Scheme 9). First, the substrate asnovolin A binds in the active site of the open conformation. The binding of asnovolin A would be a driving force for the conformational change of the loop to form a partially closed conformation. Then, the abstraction of a hydrogen atom from C13 is catalyzed by the Fe(IV)=O species. Here, the active site residue Glu208 would contribute to determining the position of the hydrogen atom abstraction, by forming steric hindrance with the A-ring of the substrate. Subsequently, further conformational changes of the active site residues on flexible loops would relocate the C13 radical intermediate, to prevent the hydroxy-rebound from the Fe(III)-OH species. Alternatively, the C13-radical reacts with molecular oxygen to form a peroxyl radical intermediate, which undergoes radical addition to C2' to generate an intermediate containing an endoperoxide bridge and a C3' radical. Finally, hydroxylation at C3' by the Fe(III)-OH species produces fumigatonoid A. Considering the catalysis by 2OG-dependent oxygenases, the radical mechanism of the hydroxylation is plausible. However, it is also possible that the water addition occurs on the C3' carbocation, which is generated by one electron transfer to the ferric iron from the C3' radical intermediate (carbocation mechanism in Scheme 9). Notably, in this last step, the stereochemistry of the hydroxylation reaction is regulated by the enzyme to be the *R*-configuration. These biochemical and biophysical analyses of NvfI suggested that NvfI catalyzes the endoperoxide formation reaction through a different mechanism from those of COX and FtmOx1. Further computational studies and mechanistical studies are now in progress in our laboratories to clarify the molecular dynamics of the active site during the enzyme reaction.



Scheme 9: Another possible reaction pathway for the formation of fumigatonoid A [28].

Conclusion

Endoperoxide compounds have recently attracted keen attention as a source of new drug leads, due to their unique structures and remarkable biological activities. To date, the total synthesis of endoperoxide-containing natural products and the synthesis of chemical reagents with an endoperoxide bridge have been reported [77,78]. The classical method for endoperoxide synthesis is through cycloadditions of dienes and alkenes, using singlet oxygen. Furthermore, cyclizations of hydroperoxides with pendant alkenes or alkynes have also been reported, by using a metal catalyst or Brønsted-acid catalysis [79-81]. However, the efficient regio- and stereoselective installation of the endoperoxide structure is still challenging, because of the increased reactivity of activated oxygen/peroxides and the high sensitivity of peroxide bridges to reductants. In this aspect, the chemoenzymatic synthesis would be a useful approach to synthesize endoperoxide-containing compounds with high regio- and stereoselectivities. The structural and mechanistic analyses of the enzymes for endoperoxide formation reviewed here would provide useful information about the engineering and design of enzymes for this application.

Although more than 200 endoperoxide-containing natural products have been isolated, only three enzymes responsible for the formation of endoperoxides have been characterized. The structural and mechanistic analyses of endoperoxide-forming enzymes, COX, FtmOx1, and NvfI, indicated that these enzymes employ distinct reaction mechanisms, suggesting that the enzymatic endoperoxide formation reactions individually evolve in a substrate-dependent manner. This fact makes it difficult to identify the endoperoxide-forming enzymes by a sequence-based genome mining approach. While it is possible that many endo-

peroxide compounds are produced through non-enzymatic oxidation by ROS, future biosynthetic analyses of endoperoxide natural products will lead to the discovery of novel endoperoxide-forming pathways and generate more comprehensive molecular insights into their remarkable chemistries. Finally, the detailed structural and mechanistic investigations of these enzymes will provide an excellent basis for the development of biocatalysts to generate novel compounds with endoperoxides for future drug discovery.

Funding

The 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 JP16H06443, JP19K15703, JP20H00490, JP20KK0173, and JP20K22700), the New Energy and Industrial Technology Development Organization (NEDO, Grant Number JPNP20011), AMED (Grant Number JP21ak0101164), the PRESTO program from Japan Science and Technology Agency, The Uehara Memorial Foundation, Astellas Foundation for Research on Metabolic Disorders, Takeda Science Foundation, and Noda Institute for Scientific Research.

ORCID® iDs

Ikuro Abe - https://orcid.org/0000-0002-3640-888X

References

1. Ann Casteel, D. *Nat. Prod. Rep.* **1999,** *16,* 55–73. doi:10.1039/a705725c
2. Dembitsky, V. M. *Eur. J. Med. Chem.* **2008,** *43,* 223–251. doi:10.1016/j.ejmech.2007.04.019
3. Liu, D.-Z.; Liu, J.-K. *Nat. Prod. Bioprospect.* **2013,** *3,* 161–206. doi:10.1007/s13659-013-0042-7
4. Dembitsky, V. M. *Med. Mycol.* **2015,** *53,* 5. doi:10.21767/2471-8521.100005
5. Tamez-Fernández, J. F.; Melchor-Martínez, E. M.; Ibarra-Rivera, T. R.; Rivas-Galindo, V. M. *Phytochem. Rev.* **2020,** *19,* 827–864. doi:10.1007/s11101-020-09687-4
6. Bu, M.; Yang, B. B.; Hu, L. *Curr. Med. Chem.* **2016,** *23,* 383–405. doi:10.2174/0929867323666151127200949
7. Norris, M. D.; Perkins, M. V. *Nat. Prod. Rep.* **2016,** *33,* 861–880. doi:10.1039/c5np00142k
8. Meshnick, S. R. *Int. J. Parasitol.* **2002,** *32,* 1655–1660. doi:10.1016/s0020-7519(02)00194-7
9. Rudrapal, M.; Chetia, D. *Drug Des., Dev. Ther.* **2016,** *10,* 3575–3590. doi:10.2147/dddt.s118116
10. Hamberg, M.; Svensson, J.; Wakabayashi, T.; Samuelsson, B. *Proc. Natl. Acad. Sci. U. S. A.* **1974,** *71,* 345–349. doi:10.1073/pnas.71.2.345
11. Crofford, L. J. *Gastroenterol. Clin. North Am.* **2001,** *30,* 863–876. doi:10.1016/s0889-8553(05)70217-x
12. Korbecki, J.; Baranowska-Bosiacka, I.; Gutowska, I.; Chlubek, D. *Acta Biochim. Pol.* **2014,** *61,* 639–649. doi:10.18388/abp.2014_1825
13. Haynes, R. K. *Curr. Opin. Infect. Dis.* **2001,** *14,* 719–726. doi:10.1097/00001432-200112000-00010
14. Borstnik, K.; Paik, I.-h.; Posner, G. H. *Mini-Rev. Med. Chem.* **2002,** *2,* 573–583. doi:10.2174/1389557023405620
15. Bauslaugh, G.; Just, G.; Blank, F. *Nature* **1964,** *202,* 1218. doi:10.1038/2021218a0
16. Adam, H. K.; Campbell, I. M.; McCorkindale, N. J. *Nature* **1967,** *216,* 397. doi:10.1038/216397a0
17. Lin, Y.; Wen, B.; Zhang, Z.; Cai, J. *Nat. Prod. Res.* **2021,** 1–6. doi:10.1080/14786419.2021.1949593
18. Martinez, G. R.; Ravanat, J.-L.; Medeiros, M. H. G.; Cadet, J.; Di Mascio, P. *J. Am. Chem. Soc.* **2000,** *122,* 10212–10213. doi:10.1021/ja0016452
19. Kumura, N.; Furukawa, H.; Kobayashi, M.; Onyango, A. N.; Izumi, M.; Nakajima, S.; Kim, H.-S.; Wataya, Y.; Baba, N. *Biosci., Biotechnol., Biochem.* **2009,** *73,* 217–220. doi:10.1271/bbb.80571
20. Novkovic, L.; Trmcic, M.; Rodic, M.; Bihelovic, F.; Zlatar, M.; Matovic, R.; Saicic, R. N. *RSC Adv.* **2015,** *5,* 99577–99584. doi:10.1039/c5ra13476e
21. Domeyer, S.; Bjerregaard, M.; Johansson, H.; Sejer Pedersen, D. *Beilstein J. Org. Chem.* **2017,** *13,* 644–647. doi:10.3762/bjoc.13.63
22. Li, C.; Li, J.; Wang, G.; Li, X. *J. Appl. Microbiol.* **2016,** *120,* 1466–1478. doi:10.1111/jam.13044
23. Khorobrykh, S.; Havurinne, V.; Mattila, H.; Tyystjärvi, E. *Plants* **2020,** *9,* 91. doi:10.3390/plants9010091
24. Rouzer, C. A.; Marnett, L. J. *J. Lipid Res.* **2009,** *50,* S29–S34. doi:10.1194/jlr.r800042-jlr200
25. Johnson, M. A.; Croteau, R. *Arch. Biochem. Biophys.* **1984,** *235,* 254–266. doi:10.1016/0003-9861(84)90274-1
26. Steffan, N.; Grundmann, A.; Afiyatullov, S.; Ruan, H.; Li, S.-M. *Org. Biomol. Chem.* **2009,** *7,* 4082–4087. doi:10.1039/b908392h
27. 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
28. Mori, T.; Zhai, R.; Ushimaru, R.; Matsuda, Y.; Abe, I. *Nat. Commun.* **2021,** *12,* 4417. doi:10.1038/s41467-021-24685-6
29. DuBois, R. N.; Abramson, S. B.; Crofford, L.; Gupta, R. A.; Simon, L. S.; van de Putte, L. B. A.; Lipsky, P. E. *FASEB J.* **1998,** *12,* 1063–1073. doi:10.1096/fasebj.12.12.1063
30. Fitzpatrick, F. A. *Curr. Pharm. Des.* **2004,** *10,* 577–588. doi:10.2174/1381612043453144
31. Williams, C. S.; DuBois, R. N. *Am. J. Physiol.* **1996,** *270,* G393–G400. doi:10.1152/ajpgi.1996.270.3.g393
32. Vane, J. R.; Bakhle, Y. S.; Botting, R. M. *Annu. Rev. Pharmacol. Toxicol.* **1998,** *38,* 97–120. doi:10.1146/annurev.pharmtox.38.1.97
33. Chandrasekharan, N. V.; Simmons, D. L. *Genome Biol.* **2004,** *5,* 241. doi:10.1186/gb-2004-5-9-241
34. van der Donk, W. A.; Tsai, A.-L.; Kulmacz, R. J. *Biochemistry* **2002,** *41,* 15451–15458. doi:10.1021/bi026938h
35. Williams, C. S.; Mann, M.; DuBois, R. N. *Oncogene* **1999,** *18,* 7908–7916. doi:10.1038/sj.onc.1203286
36. Smyth, E. M.; Grosser, T.; Wang, M.; Yu, Y.; FitzGerald, G. A. *J. Lipid Res.* **2009,** *50,* S423–S428. doi:10.1194/jlr.r800094-jlr200
37. Ricciotti, E.; FitzGerald, G. A. *Arterioscler., Thromb., Vasc. Biol.* **2011,** *31,* 986–1000. doi:10.1161/atvbaha.110.207449
38. Yamagata, K.; Andreasson, K. I.; Kaufmann, W. E.; Barnes, C. A.; Worley, P. F. *Neuron* **1993,** *11,* 371–386. doi:10.1016/0896-6273(93)90192-t

39. Ferreri, N. R.; An, S.-J.; McGiff, J. C. *Am. J. Physiol.* **1999,** *277,* F360–F368. doi:10.1152/ajprenal.1999.277.3.f360
40. Konturek, P. C.; Kania, J.; Burnat, G.; Hahn, E. G.; Konturek, S. J. *J. Physiol. Pharmacol.* **2005,** *56,* 57–73.
41. Brune, K.; Hinz, B. *Scand. J. Rheumatol.* **2004,** *33,* 1–6. doi:10.1080/03009740310004766
42. Blobaum, A. L.; Marnett, L. J. *J. Med. Chem.* **2007,** *50,* 1425–1441. doi:10.1021/jm0613166
43. Rayar, A.-M.; Lagarde, N.; Ferroud, C.; Zagury, J.-F.; Montes, M.; Sylla-Iyarreta Veitia, M. *Curr. Top. Med. Chem.* **2017,** *17,* 2935–2956. doi:10.2174/1568026617666170821124947
44. Smith, W. L.; DeWitt, D. L.; Garavito, R. M. *Annu. Rev. Biochem.* **2000,** *69,* 145–182. doi:10.1146/annurev.biochem.69.1.145
45. Rouzer, C. A.; Marnett, L. J. *Chem. Rev.* **2003,** *103,* 2239–2304. doi:10.1021/cr000068x
46. Picot, D.; Loll, P. J.; Garavito, R. M. *Nature* **1994,** *367,* 243–249. doi:10.1038/367243a0
47. Kurumbail, R. G.; Stevens, A. M.; Gierse, J. K.; McDonald, J. J.; Stegeman, R. A.; Pak, J. Y.; Gildehaus, D.; iyashiro, J. M.; Penning, T. D.; Seibert, K.; Isakson, P. C.; Stallings, W. C. *Nature* **1996,** *384,* 644–648. doi:10.1038/384644a0
48. Kiefer, J. R.; Pawlitz, J. L.; Moreland, K. T.; Stegeman, R. A.; Hood, W. F.; Gierse, J. K.; Stevens, A. M.; Goodwin, D. C.; Rowlinson, S. W.; Marnett, L. J.; Stallings, W. C.; Kurumbail, R. G. *Nature* **2000,** *405,* 97–101. doi:10.1038/35011103
49. Xu, S.; Hermanson, D. J.; Banerjee, S.; Ghebreselasie, K.; Clayton, G. M.; Garavito, R. M.; Marnett, L. J. *J. Biol. Chem.* **2014,** *289,* 6799–6808. doi:10.1074/jbc.m113.517987
50. Miciaccia, M.; Belviso, B. D.; Iaselli, M.; Cingolani, G.; Ferorelli, S.; Cappellari, M.; Loguercio Polosa, P.; Perrone, M. G.; Caliandro, R.; Scilimati, A. *Sci. Rep.* **2021,** *11,* 4312. doi:10.1038/s41598-021-83438-z
51. Orlando, B. J.; Malkowski, M. G. *J. Biol. Chem.* **2016,** *291,* 15069–15081. doi:10.1074/jbc.m116.725713
52. Malkowski, M. G.; Ginell, S. L.; Smith, W. L.; Garavito, R. M. *Science* **2000,** *289,* 1933–1937. doi:10.1126/science.289.5486.1933
53. Vecchio, A. J.; Simmons, D. M.; Malkowski, M. G. *J. Biol. Chem.* **2010,** *285,* 22152–22163. doi:10.1074/jbc.m110.119867
54. Mancini, J. A.; Riendeau, D.; Falgueyret, J.-P.; Vickers, P. J.; O'Neill, G. P. *J. Biol. Chem.* **1995,** *270,* 29372–29377. doi:10.1074/jbc.270.49.29372
55. Smith, W. L.; Garavito, R. M.; DeWitt, D. L. *J. Biol. Chem.* **1996,** *271,* 33157–33160. doi:10.1074/jbc.271.52.33157
56. Bhattacharyya, D. K.; Lecomte, M.; Rieke, C. J.; Garavito, R. M.; Smith, W. L. *J. Biol. Chem.* **1996,** *271,* 2179–2184. doi:10.1074/jbc.271.4.2179
57. Greig, G. M.; Francis, D. A.; Falgueyret, J.-P.; Ouellet, M.; Percival, M. D.; Roy, P.; Bayly, C.; Mancini, J. A.; O'Neill, G. P. *Mol. Pharmacol.* **1997,** *52,* 829–838. doi:10.1124/mol.52.5.829
58. Rieke, C. J.; Mulichak, A. M.; Garavito, R. M.; Smith, W. L. *J. Biol. Chem.* **1999,** *274,* 17109–17114. doi:10.1074/jbc.274.24.17109
59. Tsai, A.; Hsi, L. C.; Kulmacz, R. J.; Palmer, G.; Smith, W. L. *J. Biol. Chem.* **1994,** *269,* 5085–5091. doi:10.1016/s0021-9258(17)37658-5
60. Wilson, J. C.; Wu, G.; Tsai, A.-l.; Gerfen, G. J. *J. Am. Chem. Soc.* **2005,** *127,* 1618–1619. doi:10.1021/ja043853q
61. Dorlet, P.; Seibold, S. A.; Babcock, G. T.; Gerfen, G. J.; Smith, W. L.; Tsai, A.-l.; Un, S. *Biochemistry* **2002,** *41,* 6107–6114. doi:10.1021/bi015871f
62. Shimokawa, T.; Kulmacz, R. J.; DeWitt, D. L.; Smith, W. L. *J. Biol. Chem.* **1990,** *265,* 20073–20076. doi:10.1016/s0021-9258(17)30468-4
63. Liu, Y.; Roth, J. P. *J. Biol. Chem.* **2016,** *291,* 948–958. doi:10.1074/jbc.m115.668038
64. Aik, W. S.; Chowdhury, R.; Clifton, I. J.; Hopkinson, R. J.; Leissing, T.; McDonough, M. A.; Nowak, R.; Schofield, C. J.; Walport, L. J. Introduction to structural studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *2-Oxoglutarate-Dependent Oxygenases;* The Royal Society of Chemistry: Cambridge, UK, 2015; pp 59–94. doi:10.1039/9781782621959-00059
65. Islam, M. S.; Leissing, T. M.; Chowdhury, R.; Hopkinson, R. J.; Schofield, C. J. *Annu. Rev. Biochem.* **2018,** *87,* 585–620. doi:10.1146/annurev-biochem-061516-044724
66. Loenarz, C.; Schofield, C. J. *Trends Biochem. Sci.* **2011,** *36,* 7–18. doi:10.1016/j.tibs.2010.07.002
67. Schofield, C. J.; Zhang, Z. *Curr. Opin. Struct. Biol.* **1999,** *9,* 722–731. doi:10.1016/s0959-440x(99)00036-6
68. Yan, W.; Song, H.; Song, F.; Guo, Y.; Wu, C.-H.; Sae Her, A.; Pu, Y.; Wang, S.; Naowarojna, N.; Weitz, A.; Hendrich, M. P.; Costello, C. E.; Zhang, L.; Liu, P.; Jessie Zhang, Y. *Nature* **2015,** *527,* 539–543. doi:10.1038/nature15519
69. Yan, W.; Song, H.; Song, F.; Guo, Y.; Wu, C.-H.; Her, A. S.; Pu, Y.; Wang, S.; Naowarojna, N.; Weitz, A.; Hendrich, M. P.; Costello, C. E.; Zhang, L.; Liu, P.; Zhang, Y. J. *Nature* **2021,** *593,* 612. doi:10.1038/s41586-021-03527-x
70. Dunham, N. P.; Del Río Pantoja, J. M.; Zhang, B.; Rajakovich, L. J.; Allen, B. D.; Krebs, C.; Boal, A. K.; Bollinger, J. M., Jr. *J. Am. Chem. Soc.* **2019,** *141,* 9964–9979. doi:10.1021/jacs.9b03567
71. Wu, L.; Wang, Z.; Cen, Y.; Wang, B.; Zhou, J. *Angew. Chem., Int. Ed.* **2022,** *61,* e202112063. doi:10.1002/anie.202112063
72. Miłaczewska, A.; Borowski, T. *Dalton Trans.* **2019,** *48,* 16211–16221. doi:10.1039/c9dt02581b
73. Wang, X.; Su, H.; Liu, Y. *Phys. Chem. Chem. Phys.* **2017,** *19,* 7668–7677. doi:10.1039/c7cp00313g
74. Ji, J.-N.; Chen, S.-L. *Phys. Chem. Chem. Phys.* **2018,** *20,* 26500–26505. doi:10.1039/c8cp05637d
75. Lin, C.-Y.; Muñoz Hernández, A. L.; Laremore, T. N.; Silakov, A.; Krebs, C.; Boal, A. K.; Bollinger, J. M., Jr. *ACS Catal.* **2022,** *12,* 6968–6979. doi:10.1021/acscatal.2c01037
76. Pan, J.; Wenger, E. S.; Matthews, M. L.; Pollock, C. J.; Bhardwaj, M.; Kim, A. J.; Allen, B. D.; Grossman, R. B.; Krebs, C.; Bollinger, J. M., Jr. *J. Am. Chem. Soc.* **2019,** *141,* 15153–15165. doi:10.1021/jacs.9b06689
77. Adam, W.; Griesbeck, A. G. In *CRC Handbook of Organic Photochemistry and Photobiology;* Horspool, W. M.; Lenci, F., Eds.; CRC Press: Boca Raton, FL, USA, 1992.
78. Ando, W., Ed. *Organic Peroxides;* John Wiley & Sons: New York, NY, USA, 1992.
79. Harris, J. R.; Waetzig, S. R.; Woerpel, K. A. *Org. Lett.* **2009,** *11,* 3290–3293. doi:10.1021/ol901046z
80. Krabbe, S. W.; Do, D. T.; Johnson, J. S. *Org. Lett.* **2012,** *14,* 5932–5935. doi:10.1021/ol302848m
81. Rubush, D. M.; Morges, M. A.; Rose, B. J.; Thamm, D. H.; Rovis, T. *J. Am. Chem. Soc.* **2012,** *134,* 13554–13557. doi:10.1021/ja3052427

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Identification of the new prenyltransferase Ubi-297 from marine bacteria and elucidation of its substrate specificity

Jamshid Amiri Moghaddam[1], Huijuan Guo[1], Karsten Willing[2], Thomas Wichard[3] and Christine Beemelmanns[*1,4]

Full Research Paper

Open Access

Address:
[1]Chemical Biology Leibniz Institute for Natural Product Research and Infection Biology e.V., Hans-Knöll-Institute, Beutenbergstraße 11a, 07745 Jena, Germany, [2]Bio Pilot Plant, Leibniz Institute for Natural Product Research and Infection Biology e.V., Hans-Knöll-Institute, Beutenbergstraße 11a, 07745 Jena, Germany, [3]Institute for Inorganic and Analytical Chemistry, Friedrich Schiller University Jena, Lessingstr 8, 07743 Jena, Germany and [4]Biochemistry of Microbial Metabolism, Institute of Biochemistry, Leipzig University, Johannisallee 21–23, 04103 Leipzig, Germany

Email:
Christine Beemelmanns* - christine.beemelmanns@leibniz-hki.de

* Corresponding author

Keywords:
Flavobacteria; prenylation; Saccharomonospora; UbiA-like prenyltransferase

Beilstein J. Org. Chem. **2022,** *18,* 722–731.
https://doi.org/10.3762/bjoc.18.72

Received: 10 April 2022
Accepted: 07 June 2022
Published: 22 June 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

Aromatic prenylated metabolites have important biological roles and activities in all living organisms. Compared to their importance in all domains of life, we know relatively little about their substrate scopes and metabolic functions. Here, we describe a new UbiA-like prenyltransferase (Ptase) Ubi-297 encoded in a conserved operon of several bacterial taxa, including marine Flavobacteria and the genus *Sacchromonospora*. In silico analysis of Ubi-297 homologs indicated that members of this Ptase group are composed of several transmembrane α-helices and carry a conserved and distinct aspartic-rich Mg^{2+}-binding domain. We heterologously produced UbiA-like Ptases from the bacterial genera *Maribacter*, *Zobellia*, and *Algoriphagus* in *Escherichia coli*. Investigation of their substrate scope uncovered the preferential farnesylation of quinoline derivatives, such as 8-hydroxyquinoline-2-carboxylic acid (8-HQA) and quinaldic acid. The results of this study provide new insights into the abundance and diversity of Ptases in marine Flavobacteria and beyond.

Introduction

Marine bacteria harbor an enormous potential to produce structurally diverse natural products, including prenylated aromatic metabolites [1,2]. Prenylation of metabolites most often confers increased biological activities due to enhanced lipophilicity, solubility, and improved binding abilities to target proteins [3]. The prenylation reaction, most often a C–C-bond-forming step

between an aromatic acceptor moiety and a prenyl chain, is catalyzed by dedicated dominantly membrane-bound prenyltransferases (Ptases) [4-7]. Ptases belonging to the UbiA-superfamily are responsible for the modification of many important signaling molecules that are involved in a wide variety of crucial biological processes, such as cellular respiration, detoxification, and photosynthesis, within almost all living organisms [8].

In general, Ptases can be distinguished by their substrate preferences. While the microbial UbiA Ptase catalyzes the C–C-bond formation between an isoprenyl chain and the *meta*-position of *p*-hydroxybenzoate (PHB) in the ubiquinone-Coenzyme Q10 biosynthesis (Figure 1), Ptases of type MenA perform the key step in the menaquinone biosynthesis by prenylating 1,4-dihydroxy-2-naphthoic acid (DHNA) via an intermediate decarboxylative coupling step yielding demethylmenaquinone (DMK) [9]. In contrast, COX10 and chlorophyll synthases are known to fuse prenyl or phytyl tails to porphyrins as acceptors, while homogentisate prenyltransferases catalyze the condensation of homogentisate and geranylgeranyl diphosphate [10,11]. Another intriguing Ptase, called AuaA, has been reported to catalyze the farnesylation of 2-methyl-4-hydroxyquinoline using farnesyl diphosphate (FPP), which results in the metabolite aurachin D [12,13].

Following up on our recent exploration of the biosynthetic repertoire of marine bacteria [14,15], the diversity of the encoded and yet often unexplored bacterial Ptases of Flavobacteria and *Saccharomonospora* strains sparked our interest. In this study we investigated three yet poorly described homologous Ptases within the UbiA superfamily and evaluated their substrate scope by heterologous production and enzymatic bioassays. Results of our study showcase that marine bacteria harbor still a broad unexplored enzymatic repertoire.

Results and Discussion

In silico analysis of Ptases in marine Flavobacteria and the genus *Saccharomonospora*

In a first step, we aimed to gain insights into abundance and diversity of Ptases encoded in both, Flavobacteria and members of the Phylum Actinobacteria, as these were suggested to be involved in the production of meroterpenoids [2]. Thus, the genomes of marine Flavobacteria, including members of the genera *Maribacter*, *Zobellia*, *Algoriphagus*, *Polaribacter*, *Algibacter*, *Arenibacter*, *Echinicola*, *Flavobacterium*, and members of the genus *Saccharomonospora* were subjected to homology searches via local BLAST search. The detected Ptase-related sequences were then subjected to an all-against-all pairwise similarity network with Ptases deposited in the Uniprot database (Figure 2 and Table S1 in Supporting Information File 1). A subsequent network analysis uncovered that the genomes of analyzed bacterial genera encoded only four (G1–G4) out of eight previously reported Ptase groups.

The first group of Ptases (G1) contained close homologs of the bacterial MenA family (EC 2.5.1.74) (50–100% pairwise identity) catalyzing the key step in the menaquinone biosynthesis [2]. As marine bacteria such as *Zobellia barbeyronii* [16] *Marinithermus hydrothermalis* [17], *Marinobacter litoralis* [18], and *Marinobacter flavus* [19] have been reported to produce menaquinone 6 (MK-6), it can be speculated that G1-Ptases of this group are likely involved in its biosynthesis. The second group of Ptases (G2) included yet poorly described

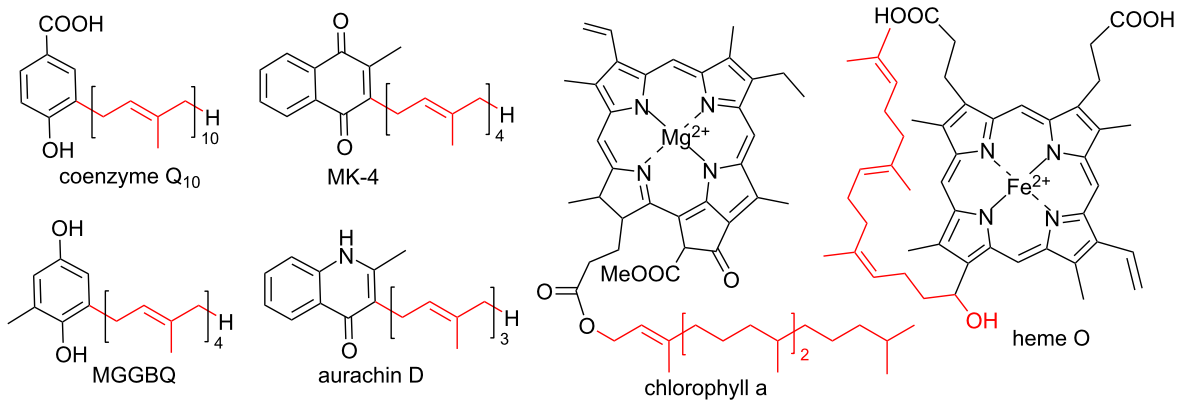


Figure 1: Prenylated aromatic metabolites are involved in cellular processes like cell respiration (coenzyme Q_{10}), cell growth and survival (menaquinone MK-4), photosynthesis (chlorophyll a), oxygen reduction (heme O), and biosynthesis of secondary metabolites (MGGBQ: 2-methyl-6-geranylgeranyl-1,4-benzoquinol and aurachin D).

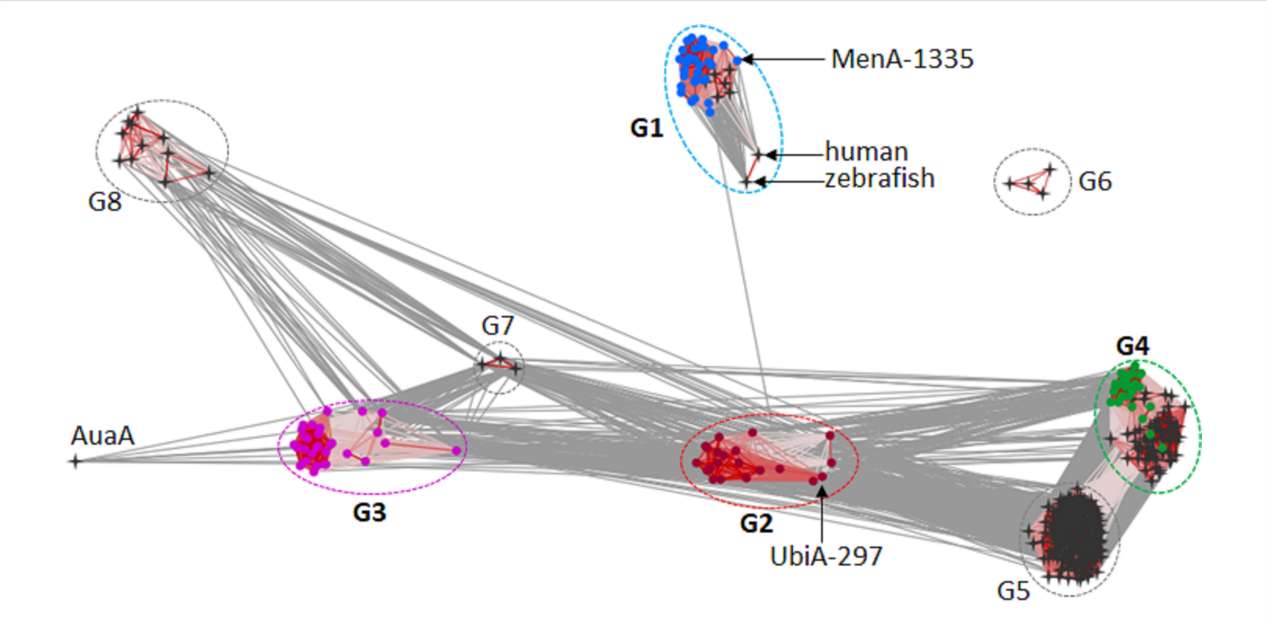


Figure 2: Homology clustering of Ptases encoded in marine *Flavobacteria* and *Saccharomonospora* species (G1–G4, colored nodes) and described Ptases in the Uniprot database (black 4-pointed stars). Visualization of the relationship between proteins is based on their all-against-all pairwise sequence similarities using CLANS. The E-values lower than 1.0E−6 were used to connect each sequence pair by edges. Edges color indicates pairwise identities lower than 10% in grey, 10–45% in light red, and above 45% in dark red. G1: 1,4-dihydroxy-2-naphthoate polyprenyltransferase (blue nodes), G2: UbiA-like Ptases (red nodes), G3: (*S*)-2,3-di-*O*-geranylgeranylglyceryl phosphate synthase (purple nodes), G4: protoheme IX farnesyltransferase (green nodes), G5: 4-hydroxybenzoate octaprenyltransferase (bacterial and mitochondrial), G6: decaprenyl-phosphate phosphoribosyltransferase, G7: chlorophyll synthase, and G8: homogentisate Ptases. AuaA is used as outgroup Ptase from *Stigmatella aurantiaca*.

UbiA-like Ptases (EC 2.5.1.39) with 43–99% pairwise identity. The third cluster of Ptases (G3) (20–100% pairwise identity) displayed similarities to geranylgeranylglyceryl phosphate synthases (EC 2.5.1.42), which are involved in the formation of polar membrane lipids of archaea and other bacteria [20], while group G4 contained close homologs of the protoheme IX farnesyltransferase (EC 2.5.1.141) (23–100% pairwise identity) responsible for heme O production in bacteria like *E. coli* [21,22]. However, no representatives of the remaining four Ptase groups G6–G8, which encode for ubiquinone biosynthesis (G5), decaprenyl-phosphate phosphoribosyltransferases (EC 2.4.2.45, G6) as reported in Mycobacteriaceae [23], chlorophyll synthases (EC 2.5.1.62, G7) [24] or homogentisate Ptases (EC 2.5.1.117, G8) [25], were found within the investigated genomes of Flavobacteria and *Saccharomonospora* representatives.

A Ptase (UbiA-297) which was putatively assigned as an aromatic Ptase of the G2 UbiA-like family (299 amino acids, calculated mass of 31.9 kDa) (Figure 2) caught our attention due to its low sequence identity on the amino acid sequence level (coverage below 50%) with other characterized Ptases. To further explore the putative function of UbiA-297, we generated a phylogenetic tree using 444 Ptase sequences with already biochemically characterized Ptases; however, none of the members grouped with UbiA-297 the G2-Ptase group (Figure S1, Supporting Information File 1). The closest characterized relatives were identified as a putatively assigned digeranylgeranylglyceryl phosphate synthase encoded in *Sulfurisphaera tokodaii* str. 7 (33% identity) and a 4-hydroxybenzoate octaprenyltransferase (ubiquinone-8 (UQ-8), 23% identity) from *Shewanella woodyi* ATCC 51908. Additionally, UbiA-297 showed sequence identity to a membrane-bound Ptase from the plant *Avena sativa* (31%) [26].

We then compared the genetic environment of the coding gene *ubiA*-297 (900 bp) within Flavobacteria and *Saccharomonospora* genomes and found a set of conserved genes within the proximity of *ubiA*-297, which were previously identified as *ebo* cluster in bacterial genomes of strains belonging to various different phyla, most notably bacteroidetes and cyanobacteria (Figure 3) [27]. The gene sequence *eboA-E* encodes five yet uncharacterized enzymes, including EboA, a putative metallo-dependent hydrolase TatD (EboB), a putative UbiA prenyltransferase (EboC), a putative 3-dehydroquinate synthase (EboD) likely catalyzing the second step in the shikimate pathway, and a TIM barrel protein (EboE) [28]. While prior studies suggested that the enzymatic reactions carried out by EboA-E include the prenylation of an undetermined substrate by eboC (UbiA-297 homologue) and modifications of a polyhydroxylated aromatic metabolite, the enzymatic reactions carried out by EboA-E have not been investigated yet in detail.

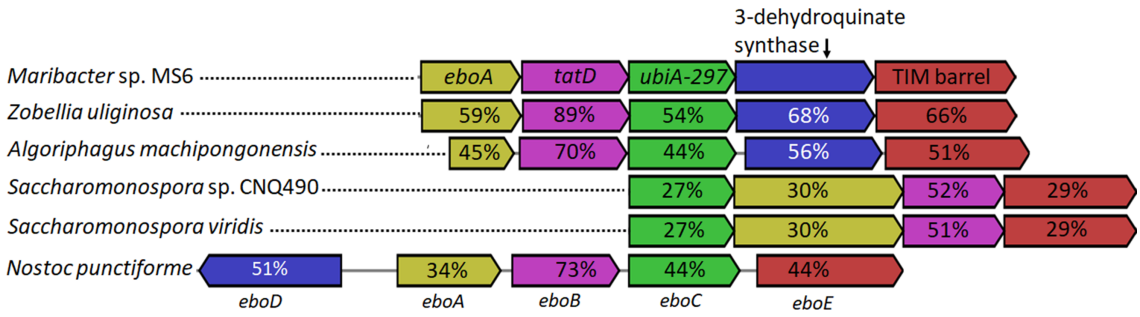


Figure 3: Regional alignment of *ubiA-297* of *Maribacter* sp. MS6 (EU359911.1) and homologous genes in *Z. uliginosa* DSM 2061 (NZ_FTOB01000006), *A. machipongonensis* PR1 (NZ_CM001023), *Saccharomonospora* sp. CNQ490 (NZ_AZUM01000003), *S. viridis* DSM 43017 (NC_013159), and *Nostoc punctiforme* PCC 73102 (NC_010628). Percentage identities of homologous genes are given in the same colored arrow boxes (*Maribacter* sp. MS6 was set as reference).

The wide spread occurrence of the *ebo* gene cluster region within genomes of bacterial symbionts associated, e.g., with marine algae, such as *Maribacter* sp. MS6 [29-31], sparked our interest. To provide more insights into the structural basis of UbiA-297, we performed sequence alignments and structure homology modelling using Swiss-Model (Figure 4) [32], which revealed the transmembrane domain consisting of ten α-helices and loops connecting the transmembrane helices (Figure S2 in Supporting Information File 1), similar to archaeal UbiA prenyltransferase enzymes [33].

The Asp-rich motif of G2-Ptases, known to coordinate Mg^{2+} ions and pyrophosphate, was detectable in all homologous sequences retrieved from marine Flavobacteria and *Saccharomonospora* genomes and was similar to the DXXDXXXD motif in *E. coli* UbiA, but distinct from the motif of other aromatic Ptases such as MenA (DXXDXXXXXD).

While most UbiA Ptases, such as of *E. coli* UbiA (C5A133) and human COQ2 (Q96H96), carry a basic arginine residue within the central cavity and which is proposed to bind the aromatic substrate (e.g., 4-hydroxybenzoate), UbiA-297 and other G2-Ptases harbor instead a tyrosine (Tyr-56) residue, which likely fulfills a similar coordinative function. Furthermore, a conserved arginine residue (UbiA-297 R145) was detectable, which was located in the neighboring α-helix (R145) and in proximity to the binding motif, and was hypothesized to be involved in the coordination process of the aromatic substrate (Figure 4B and 4C).

Heterologous production of UbiA-297

To enable investigations into the substrate scope of UbiA-297, the coding gene sequence *ubiA-297* was amplified from the genomic DNA of *Maribacter* sp. MS6, while homologous sequences encoded in *Z. uliginosa* DSM 2061 and *A. machipongonensis* sp. PR1 were synthesized. To gain more insights into the binding properties of UbiA-297 and the functional role of the conserved arginine moiety (R145), an additional point-mutated *ubiA-297* (R145A) version was synthesized codon-optimized for expression in *E. coli*. Synthesized and amplified sequences were then cloned into an expression pET28 plasmid containing an N-terminal 6-histidine tag sequence. Heterologous production of enzymes was achieved in *E. coli* BL21 and western blot analysis indicated the accumulation of His-tagged UbiA-297 (35 kD) within concentrated cell membrane fractions (Figure S2, Supporting Information File 1). However, purification of active Ptases failed despite testing different detergent-based purification protocols. Thus, we performed assays with crude protein fractions (1st membrane faction) and protein-enriched membrane fractions, which were obtained after washing and ultracentrifugation.

Substrate specificity of UbiA-297

Based on our in silico analysis and previous mass-spectrometry-guided metabolomic analysis of marine Flavobacteria and members of the genus *Saccharomonospora* [29,30], we anticipated hydroxylated aromatic or even quinoline-like acceptor compounds and farnesyl pyrophosphate (FPP) as most likely substrates for Ubi-297 (*Maribacter* sp. MS6). Thus, the 1st membrane faction containing membrane-bound UbiA-297 was subjected to an enzyme assay with farnesyl pyrophosphate (FPP) and different aromatic acceptor substrates in the presence of Mg^{2+} as co-factor. Product formation was monitored after 2 h using high-resolution tandem mass spectrometry (HRMS/MS). Membrane proteins obtained from *E. coli* BL21 cultures harboring the empty expression pET28 plasmid were used as negative control (Figure S3 in Supporting Information File 1).

As depicted in Figure 5, farnesylated products were detectable for six out of 14 tested aromatic acceptor substrates by HRMS/MS (Figures S4–S8 in Supporting Information File 1). In partic-

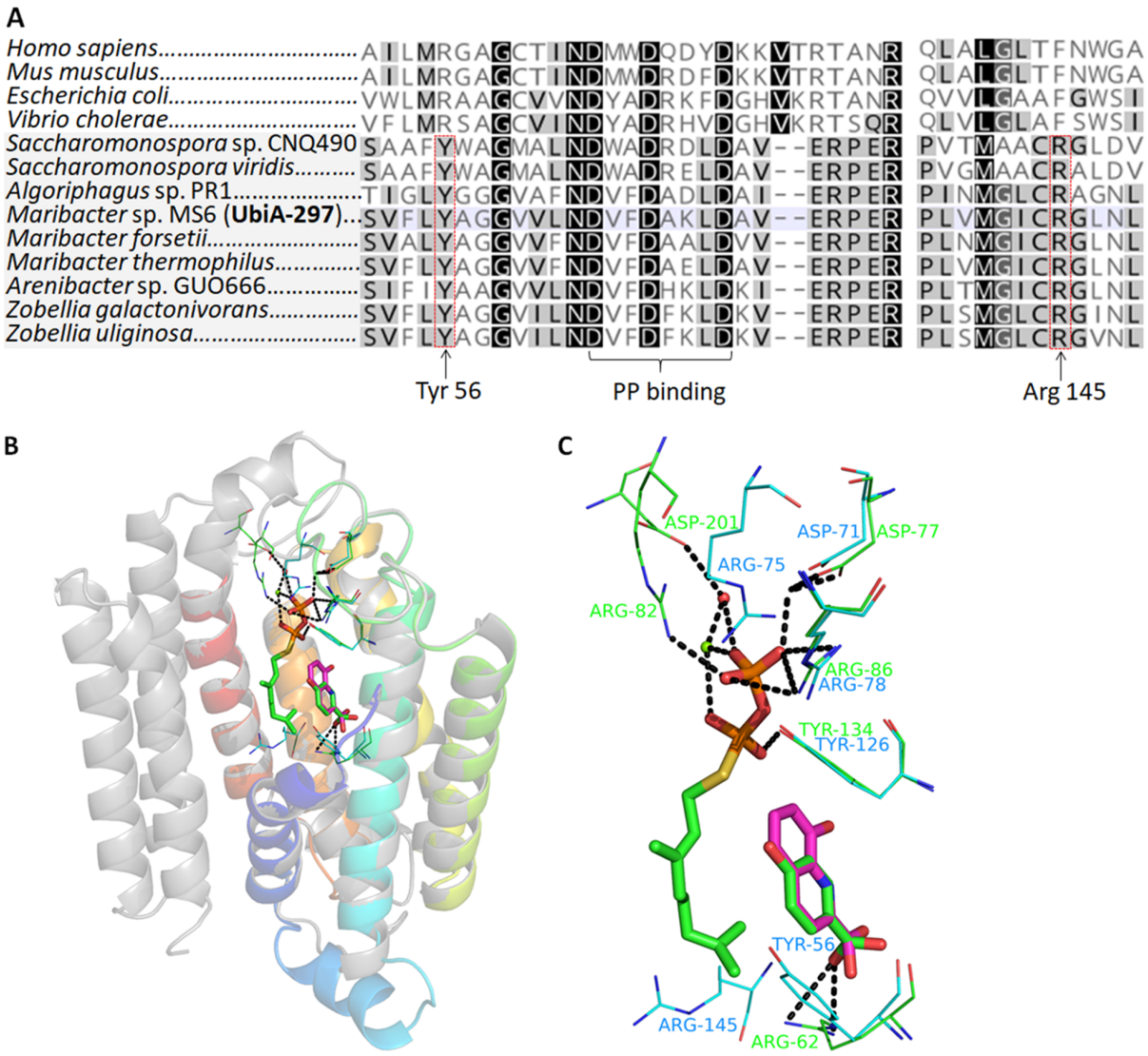


Figure 4: A) Amino acid alignment and binding residues of UbiA-297. G2-Ptases are illustrated in the grey box. B) Homology model of UbiA-297 (in color). Model is based on the substrate-bound structure of a UbiA homolog from *Aeropyrum pernix* sp. K1 (grey). C) The pyrophosphate and aromatic substrate binding sites of UbiA-297 model (shown in green) and template structure (shown in blue). 8-HQA (pink) is paired to 4-HBA (green) after energy minimization.

ular, quinoline-type substrates, such as 8-HQA and quinaldic acid, were transformed, while only moderate conversion of 8-hydroxyquinoline and 1,3-dihydroxynaphthalene were observed. Xanthurenic acid and 4-methylumbelliferone were only farnesylated in negligible amounts, and no product formation was observed for phenols or catechols (Figure 5).

Overall, 8-hydroxyquinoline-2-carboxylic acid (8-HQA) appeared to be the most favored substrate amongst the tested panel. Thus, we shortly investigated different reaction parameters using 8-HQA and FPP as substrates. First, we compared the enzyme activity of crude protein fractions directly obtained from cell lysate and enriched UbiA-297 fractions (Figure 6). As expected, higher product signals (RT 18.03 min, *m/z* $[M + H]^+$ 394.2368, calcd for $C_{25}H_{32}NO_3^+$, *m/z* $[M + H]^+$ 394.2376; *m/z* $[M - H]^-$ 392.2233, calcd for $C_{25}H_{30}NO_3^-$, $[M - H]^-$ 392.2231) were detectable for enriched protein fractions (Figure 6A–C), which suggested that the active protein requires indeed a membrane-like environment. Along these lines, no product formation was detectable when UbiA-297 was denaturated by heating prior to the assay (control), nor when a MenA-1335 (*Maribacter* sp. MS6) homolog was tested.

Furthermore, different pH values and extended incubation times were investigated, however, no significant changes in production levels were noted. Next, enriched protein UbiA-

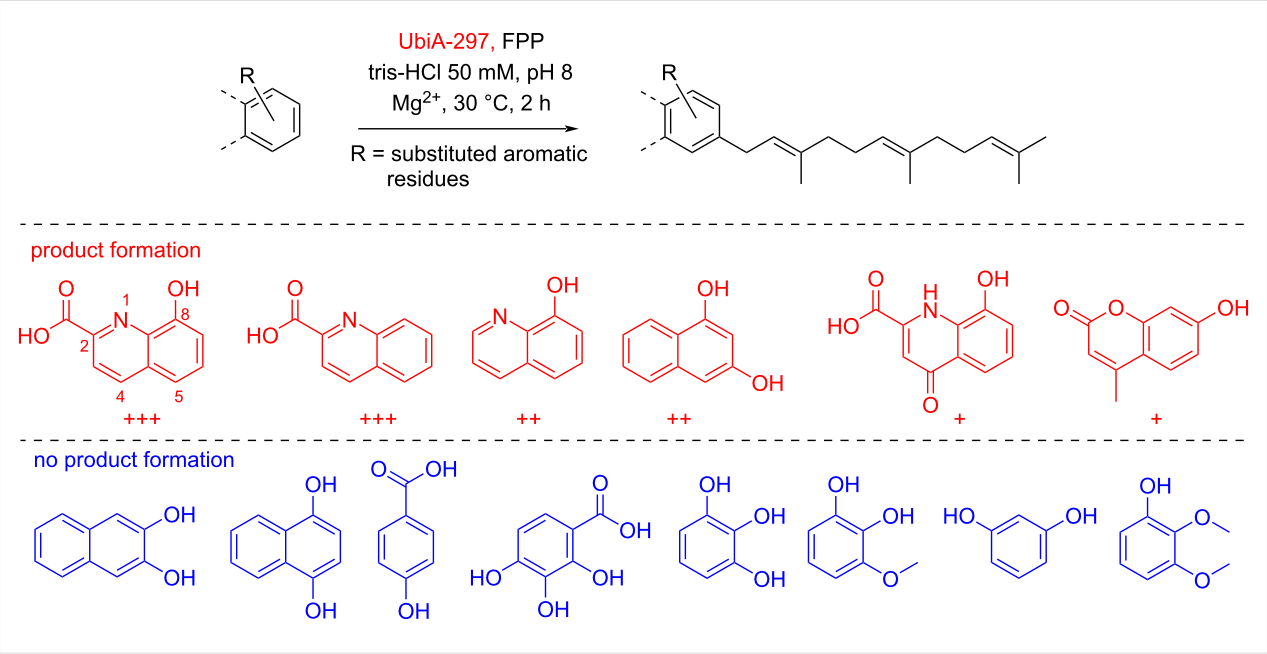


Figure 5: Evaluation of substrate specificity of UbiA-297. Accepted substrates are shown in red, while no product formation was observed for compounds colored in blue. HRMS signal intensities and the MS^2 pattern of the farnesylated product served as measure to determine preferred acceptor substrates (ion signal intensity: $>10^7$ = +++, $>10^6$ = ++, and $>10^5$ = +, n = 3) (for comparision assuming equal ionisation potential).

297(R145A), in which the conserved Arg145 (Figure 4A) was replaced by Ala145, was tested (Figure 6D). As similar production levels as UbiA-297 were observed, we concluded that Arg145 might not be crucial for the enzymatic activity.

We then tested, if heterologously produced UbiA-297 homologs from *A. machipongonensis* PR1 and *Z. uliginosa* DSM 2061 (Figure 6D) perform similarly, and indeed observed farnesylation of 8-HQA for both enzymes, which indicated again towards a conserved function of UbiA-297 enzymes within different bacterial taxa.

Furthermore, prenylation activity was also detectable when 8-HQA and farnesol were directly added to an *E. coli* BL21 culture heterologously producing UbiA-297 (Figure S3 in Supporting Information File 1). With the aim to confirm the structure of the prenylated 8-HQA product, a 30 L fermentation of *E. coli*/pET28-297 was performed. After induction of protein expression, the precursor 8-HQA and farnesol were added to the culture and incubated overnight. MS-guided purification of cell lysate and ^{1}H NMR analysis of the prenylated 8-HQA product confirmed the prenylation on the quinoline ring (Figure S9 in Supporting Information File 1).

Conclusion

In silico analysis and homology clustering of Ptases encoded in Flavobacteria and the genus *Sacchromonospora* hinted towards a yet unexplored group of membrane-bound UbiA-like Ptase named UbiA-297, which is part of the conserved *ebo* gene sequences and widespread across various bacterial lineages, including many symbiotic taxa. Heterologously produced UbiA-297 catalyzed the farneslyation of quinoline-like aromatic substrates, with a strong preference for 8-HQA. The herein obtained results build the foundation for future in-depth studies on the substrate scope of Ubi-297-like enzymes and will allow exploring the functions of prenylated 8-HQA for the bacterial producers and their symbionts.

Experimental

Chemicals: Aromatic substrates and the ammonium salt of FPP were obtained from Sigma–Aldrich, Alfa Aesar (Ward Hill, MA), and Acros Organics (Geel, Belgium).

Bioinformatic analysis: Amino acid sequences of all described prenyltransferases were retrieved from the UniProtKB/Swiss-Prot database. Other Ptase amino acid sequences of marine Flavobacteria and *Saccharomonopora* strains were obtained using Blast searches against defined genome groups within the PATRIC database (3.6.12) [34]. The sequence clustering was generated by the CLANS (CLuster ANalysis of Sequences) program [35]. In brief, the relationships between sequences were assessed based on an all-against-all BLAST search and the E-values better (lower) than 1.0E−6 were used to connect each sequence pair by edges, which was then colored based on pairwise identities. Different groups were considered based on their distance in space and a combination of E-values

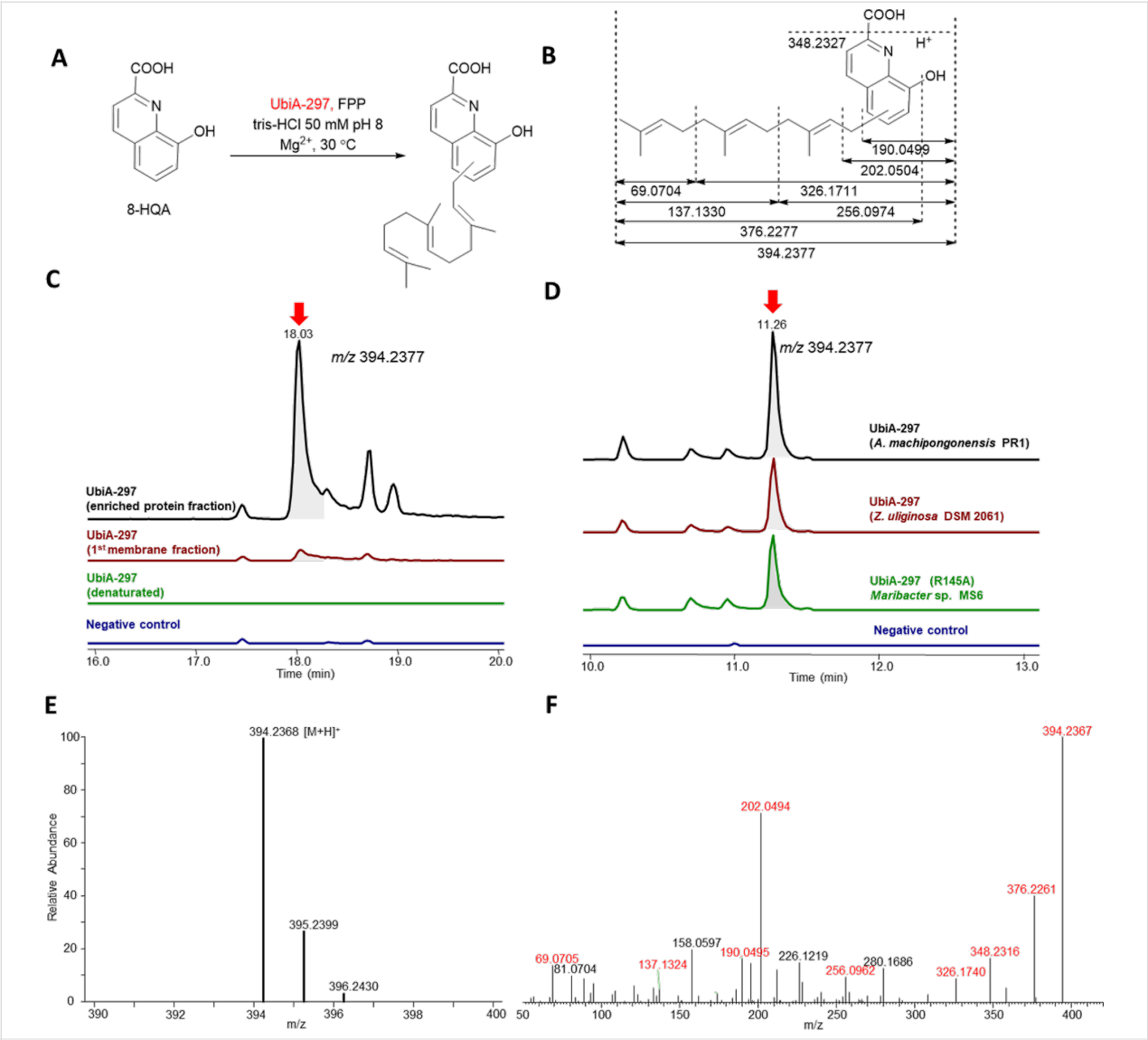


Figure 6: A) Reaction scheme of UbiA-297 catalyzing the assumed *para*-directed farnesylation of 8-HQA; B) calculated MS^2 fragmentation pattern of farnesylated 8-HQA; C) extracted ion chromatogram (EIC) for *m/z* 394.2376 (t_R 18.03 min, red arrow) of enriched UbiA-297 protein fractions, cell-membrane fraction containing UbiA-297 (*Maribacter* sp. MS6), denaturated enriched UbiA-297 (negative control), and enriched *E. coli* BL21 derived membrane fractions (negative control not shown); fixed ion intensity scale (5.00E7) was applied to all chromatograms; D) EIC for *m/z* 394.2376 (t_R 11.26 min) of assay containing membrane-bound UbiA-297 encoded in *A. machipongonensis* sp. PR1, *Z. uliginosa* DSM 2061, and point mutated UbiA-297 (R145A) from *Maribacter* sp. MS6; E and F) HRMS and MS^2 spectrum showing fragmentation of prenylated 8-HQA (*m/z* 394.2367, t_R 18.03 min).

and pairwise identities. A phylogenetic tree was created using the neighbor joining method in Geneious Prime (2020.2.3) after multiple sequence alignments using MAFFT (7.450) [36]. Structure modelling of UbiA-297 was done using the SWISS-MODEL server [32] and visualized using Pymol (2.3.3). Regional alignment of homologous Ptase genes was done using MultiGeneBlast [37].

Nucleotide sequence accession numbers: The *ubiA-297* and *menA-1335* gene sequences are deposited at GenBank with the accession numbers ON075815 and ON075816, respectively. Accession number or ID of all other sequences used to infer the pairwise similarity network and phylogenetic tree are provided in Supporting Information File 1, Table S1.

Bacterial strains, plasmids, and culture conditions: *Maribacter* sp. MS6 was initially isolated from the green macroalga *Ulva mutabilis* [29,30]. The strain was cultivated in 25 mL marine broth (Carl Roth) at 150 rpm and 28 °C for up to four days. For cloning experiments, *E. coli* DH5 alpha and the pJET 1.2 (Thermofisher) and pET28a(+) vectors were used, and *E. coli* BL21 was used as production host. Cultures were

supplemented with ampicillin (100 μg mL^{-1}) and kanamycin (60 μg mL^{-1}) for the selection of plasmids.

DNA isolation, PCR amplification, and cloning: gDNA of *Maribacter* sp. MS6 was extracted using DNeasy Blood & Tissue Kit (Qiagen) and PCR amplification of *ubiA-297* (922 bp) and *menA-1335* (952 bp) genes was carried out using S7 Fusion High-Fidelity DNA Polymerase (Biozym) and designated primer sequences (*ubiA*-297 forward 5'-CAGGATCCAGGATGTCCAATAAACTAATG-3' reverse 5'-CGAAGCTTGTCTTAGGTAATGGCAAAAAG-3'; *menA* forward 5'-CAGGATCCCCCTTACTAGTGAC-3' and reverse 5'-GTAAGCTTGATGGTTGACATTC-3'). The obtained PCR fragments were ligated into a pJET 1.2 vector yielding plasmid pJET-297 and pJET-1335, which were sequenced to confirm integrity. To create the pET28-297 and pET28-1335 expression vector, pJET-297 and pJET-1335 were digested with BamHI and HindIII, and then ligated to digested pET28a(+) with the same restriction enzymes and transformed into *E. coli* BL21 for heterologous expression of UbiA-297 and MenA-1335.

Codon-optimized *ubiA-297* homologous genes of *Z. galactanivorans* (DSM 2061) and *A. machipongonensis* (DSM 24695) were synthesized in pET28 vector (BioCAT GmbH) and transformed into *E. coli* BL21 for heterologous expression. Additionally, a modified *ubiA-297*(R145A) gene was synthesized and cloned into a pET28 vector (BioCAT GmbH) yielding pET28-297(R145A) for heterologous expression.

Preparation of enzyme extracts and protein quantification: Plasmids containing the respective gene were transformed into *E. coli* BL21 and strains harboring the plasmid were cultivated in TB medium supplemented with kanamycin (60 μg mL^{-1}). Cultures were grown at 37 °C to an OD_{600} of 0.8, then brought to 16 °C and isopropyl β-D-thiogalactoside (IPTG) was added to a final concentration of 0.1 mM. Cells were cultivated overnight at 16 °C before harvesting. Cultures were centrifuged, the cell pellet resuspended in Tris-HCl buffer (50 mM, pH 7.8) supplemented with dithiothreitol (DTT, 10 mM), and sonicated on ice (12 min, 100% C, 40% A, 2 seconds on and 3 seconds off intervals using Hielscher UP200St ultrasonic processor). The first cell membrane fraction was obtained by centrifugation (12000*g*, 20 min, 4 °C), while the enriched protein fraction, likely imbedded in lipid rafts, required ultracentrifugation of the crude protein lysates (240000*g*, 90 min). The obtained protein pellet was resuspended in the same buffer as used for the bioassays. Protein concentration was measured according to Bradford [38].

Enzyme assays: All the enzyme assays were performed using a standard reaction mixture (100 μL) by adding 50 μL Tris-HCl (50 mM, pH 7.8) containing $MgCl_2$, FPP, aromatic substrate (each 1 mM), and 50 μL of membrane-bound protein aliquots with a total protein content of 0.4 mg. The reaction mixtures were incubated at 30 °C for 2 h (extended period of time) and extracted subsequently three times with ethyl acetate (450 μL). The solvent was removed in vacuo, the residue was dissolved in methanol (100 μL), and analyzed by HRMS/MS. Denaturation of proteins was performed at 95 °C for 10 min (negative control). Additionally, *E. coli* BL21 derived enriched membrane fractions served as negative control.

In vivo assays: In vivo assays (100 mL) were performed using *E. coli* BL21 harboring one of the following vectors: pET28-297, pET28-1335, and pET28-no insert. Protein expression was performed under the same conditions explained above, and after 12 h 8-HQA, farnesol, and $MgCl_2$ were added to each culture to a final concentration of 1 mM each and cultivation was continued at 30 °C (3 d). Subsequently, cultures were extracted twice with ethyl acetate (400 mL), the solvent was removed in vacuo, the residue redissolved in MeOH, and subjected to HRMS/MS analysis.

Fermentation and purification: Fermentation of *E. coli* BL21 cells harboring the pET28-297 vector was performed in a 75 L X-Cube Bioreactor (Braun Biotech International) using 30 L of Terrific Broth medium (Carl Roth GmbH) enriched with 4 mL/L glycerol and 60 mg/L kanamycin. Similar to the in vivo culture assays, the culture was cooled to 16 °C at an OD_{600} of 1 and heterologous production induced with IPTG (final concentration of 0.1 mM). After 12 h, 8-HQA, farnesol, and $MgCl_2$ were added to the fermentation with a final concentration of 1 mM. Fermentation was continued for 3 d at 30 °C, after which the culture was centrifuged and the cell biomass was lyophilized. Dry cell mass (103.6 g) was extracted with methanol (1.0 L), dried under vacuum, and the resultant MeOH extract was then extracted with hexane (120 mL) and dried again under vacuum. The hexane extract was purified by flash chromatography (Biotage Isolera Prime) over a silica gel column (eluent: cyclohexane/EtOAc 100:0 to 80:20 to 0:100). The appropriate fraction was collected, evaporated, and purified by preparative HPLC (Shimadzu) over a phenyl-hexyl column (Luna, 5 μm, 250 × 21.2 mm, 100 Å) (eluent: H_2O/MeCN + 0.1% formic acid 80:20 to 50:50). The appropriate fraction was collected and evaporated to afford farnesylated 8-HQA.

HRMS/MS analysis of the enzymatic products: UHPLC-HESI-HRMS measurement was performed on a Dionex Ultimate3000 system combined with a Q-Exactive Plus mass spectrometer (Thermo Scientific) with a heated electrospray ion source (HESI). Metabolite separation was carried out by reversed-phase liquid chromatography at 40 °C using a Luna

Omega C18 column (100 × 2.1 mm, 1.6 µm, 100 Å, Phenomenex) preceded by a SecurityGuardTM ULTRA guard cartridge (2 × 2.1 mm, Phenomenex). Mobile phases were acidified with 0.1% formic acid and consisted of H_2O (A), and acetonitrile (B) with a flow rate of 0.3 mL/min and the injection volume was 5 µL. The prenylated products were separated under either long or short gradient runs. The long gradient run was 30 min as follows: 0–0.5 min, 5% B; 0.5–18 min, 5–97% B; 18–25 min, 97–5% B; 25–30 min, 5% B. The short gradient run was 15 min as follows: 0–0.8 min, 40% B; 0.8–10 min, 40–97% B; 10–12 min, 97% B; 12–13 min, 40% B; 13–15 min, 40% B. The retention times of the farnesylated 8HQA was 18.03 min using a long gradient run and 11.26 min using a short gradient run.

Spectroscopic analysis of the enzyme products: ^{1}H NMR spectra was carried out using a Bruker AVANCE III 600 MHz spectrometer equipped with a Bruker Cryoplatform with chemical shifts given in ppm (δ).

Supporting Information

Supporting Information File 1

Sequence analysis and copies of MS/MS and NMR spectra.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-72-S1.pdf]

Acknowledgements

We would like to thank Mrs. Heike Heinecke (HKI) for recording NMR spectra.

Funding

We are grateful for financial support from the Leibniz Association. CB greatly acknowledges funding from the European Union's Horizon 2020 research and innovation program (ERC Grant number: 802736, MORPHEUS). We are deeply thankful for funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Project-ID 239748522 – CRC 1127 (projects A01 and A06).

ORCID® iDs

Jamshid Amiri Moghaddam - https://orcid.org/0000-0002-9385-0711
Huijuan Guo - https://orcid.org/0000-0002-1308-4069
Thomas Wichard - https://orcid.org/0000-0003-0061-4160
Christine Beemelmanns - https://orcid.org/0000-0002-9747-3423

References

1. Mori, T. *J. Nat. Med.* **2020,** *74,* 501–512. doi:10.1007/s11418-020-01393-x
2. Rudolf, J. D.; Alsup, T. A.; Xu, B.; Li, Z. *Nat. Prod. Rep.* **2021,** *38,* 905–980. doi:10.1039/d0np00066c
3. Winkelblech, J.; Fan, A.; Li, S.-M. *Appl. Microbiol. Biotechnol.* **2015,** *99,* 7379–7397. doi:10.1007/s00253-015-6811-y
4. Liang, P.-H.; Ko, T.-P.; Wang, A. H.-J. *Eur. J. Biochem.* **2002,** *269,* 3339–3354. doi:10.1046/j.1432-1033.2002.03014.x
5. Bonitz, T.; Alva, V.; Saleh, O.; Lupas, A. N.; Heide, L. *PLoS One* **2011,** *6,* e27336. doi:10.1371/journal.pone.0027336
6. Xu, B.; Li, Z.; Alsup, T. A.; Ehrenberger, M. A.; Rudolf, J. D. *ACS Catal.* **2021,** *11,* 5906–5915. doi:10.1021/acscatal.1c01113
7. Zeyhle, P.; Bauer, J. S.; Kalinowski, J.; Shin-ya, K.; Gross, H.; Heide, L. *PLoS One* **2014,** *9,* e99122. doi:10.1371/journal.pone.0099122
8. Li, W. *Trends Biochem. Sci.* **2016,** *41,* 356–370. doi:10.1016/j.tibs.2016.01.007
9. Dhiman, R. K.; Pujari, V.; Kincaid, J. M.; Ikeh, M. A.; Parish, T.; Crick, D. C. *PLoS One* **2019,** *14,* e0214958. doi:10.1371/journal.pone.0214958
10. Sadre, R.; Gruber, J.; Frentzen, M. *FEBS Lett.* **2006,** *580,* 5357–5362. doi:10.1016/j.febslet.2006.09.002
11. Mène-Saffrané, L. *Antioxidants* **2017,** *7,* 2. doi:10.3390/antiox7010002
12. Pistorius, D.; Li, Y.; Sandmann, A.; Müller, R. *Mol. BioSyst.* **2011,** *7,* 3308–3315. doi:10.1039/c1mb05328k
13. Kitagawa, W.; Ozaki, T.; Nishioka, T.; Yasutake, Y.; Hata, M.; Nishiyama, M.; Kuzuyama, T.; Tamura, T. *ChemBioChem* **2013,** *14,* 1085–1093. doi:10.1002/cbic.201300167
14. Amiri Moghaddam, J.; Jautzus, T.; Alanjary, M.; Beemelmanns, C. *Org. Biomol. Chem.* **2021,** *19,* 123–140. doi:10.1039/d0ob01677b
15. Zhou, Q.; Hotta, K.; Deng, Y.; Yuan, R.; Quan, S.; Chen, X. *Microorganisms* **2021,** *9,* 2551. doi:10.3390/microorganisms9122551
16. Nedashkovskaya, O.; Otstavnykh, N.; Zhukova, N.; Guzev, K.; Chausova, V.; Tekutyeva, L.; Mikhailov, V.; Isaeva, M. *Diversity* **2021,** *13,* 520. doi:10.3390/d13110520
17. Lin, H.; Lin, D.; Zhang, M.; Ye, J.; Sun, J.; Tang, K. *Curr. Microbiol.* **2021,** *78,* 2815–2820. doi:10.1007/s00284-021-02519-4
18. Lee, D. W.; Lee, H.; Kwon, B.-O.; Khim, J. S.; Yim, U. H.; Park, H.; Park, B.; Choi, I.-G.; Kim, B. S.; Kim, J.-J. *Int. J. Syst. Evol. Microbiol.* **2018,** *68,* 3471–3478. doi:10.1099/ijsem.0.003011
19. Tang, M.; Wang, G.; Xiang, W.; Chen, C.; Wu, J.; Dai, S.; Wu, H.; Li, T.; Wu, H. *Int. J. Syst. Evol. Microbiol.* **2015,** *65,* 3997–4002. doi:10.1099/ijsem.0.000526
20. Hemmi, H.; Shibuya, K.; Takahashi, Y.; Nakayama, T.; Nishino, T. *J. Biol. Chem.* **2004,** *279,* 50197–50203. doi:10.1074/jbc.m409207200
21. Saiki, K.; Mogi, T.; Anraku, Y. *Biochem. Biophys. Res. Commun.* **1992,** *189,* 1491–1497. doi:10.1016/0006-291x(92)90243-e
22. Antonicka, H.; Leary, S. C.; Guercin, G.-H.; Agar, J. N.; Horvath, R.; Kennaway, N. G.; Harding, C. O.; Jaksch, M.; Shoubridge, E. A. *Hum. Mol. Genet.* **2003,** *12,* 2693–2702. doi:10.1093/hmg/ddg284
23. He, L.; Wang, X.; Cui, P.; Jin, J.; Chen, J.; Zhang, W.; Zhang, Y. *Tuberculosis* **2015,** *95,* 149–154. doi:10.1016/j.tube.2014.12.002
24. Proctor, M. S.; Chidgey, J. W.; Shukla, M. K.; Jackson, P. J.; Sobotka, R.; Hunter, C. N.; Hitchcock, A. *FEBS Lett.* **2018,** *592,* 3062–3073. doi:10.1002/1873-3468.13222
25. Shino, M.; Hamada, T.; Shigematsu, Y.; Banba, S. *Pest Manage. Sci.* **2020,** *76,* 3389–3394. doi:10.1002/ps.5698
26. Schmid, H. C.; Rassadina, V.; Oster, U.; Schoch, S.; Rüdiger, W. *Biol. Chem.* **2002,** *383,* 1769–1776. doi:10.1515/bc.2002.198
27. Klicki, K.; Ferreira, D.; Hamill, D.; Dirks, B.; Mitchell, N.; Garcia-Pichel, F. *mBio* **2018,** *9,* e02266-18. doi:10.1128/mbio.02266-18

28. Yurchenko, T.; Ševčíková, T.; Přibyl, P.; El Karkouri, K.; Klimeš, V.; Amaral, R.; Zbránková, V.; Kim, E.; Raoult, D.; Santos, L. M. A.; Eliáš, M. *ISME J.* **2018,** *12,* 2163–2175. doi:10.1038/s41396-018-0177-y
29. Weiss, A.; Costa, R.; Wichard, T. *Bot. Mar.* **2017,** *60,* 197–206. doi:10.1515/bot-2016-0083
30. Alsufyani, T.; Califano, G.; Deicke, M.; Grueneberg, J.; Weiss, A.; Engelen, A. H.; Kwantes, M.; Mohr, J. F.; Ulrich, J. F.; Wichard, T. *J. Exp. Bot.* **2020,** *71,* 3340–3349. doi:10.1093/jxb/eraa066
31. Wichard, T.; Beemelmanns, C. *J. Chem. Ecol.* **2018,** *44,* 1008–1021. doi:10.1007/s10886-018-1004-7
32. 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
33. Cheng, W.; Li, W. *Science* **2014,** *343,* 878–881. doi:10.1126/science.1246774
34. Davis, J. J.; Wattam, A. R.; Aziz, R. K.; Brettin, T.; Butler, R.; Butler, R. M.; Chlenski, P.; Conrad, N.; Dickerman, A.; Dietrich, E. M.; Gabbard, J. L.; Gerdes, S.; Guard, A.; Kenyon, R. W.; Machi, D.; Mao, C.; Murphy-Olson, D.; Nguyen, M.; Nordberg, E. K.; Olsen, G. J.; Olson, R. D.; Overbeek, J. C.; Overbeek, R.; Parrello, B.; Pusch, G. D.; Shukla, M.; Thomas, C.; VanOeffelen, M.; Vonstein, V.; Warren, A. S.; Xia, F.; Xie, D.; Yoo, H.; Stevens, R. *Nucleic Acids Res.* **2020,** *48,* D606–D612. doi:10.1093/nar/gkz943
35. Frickey, T.; Lupas, A. *Bioinformatics* **2004,** *20,* 3702–3704. doi:10.1093/bioinformatics/bth444
36. Kopcsayová, D.; Vranová, E. *Molecules* **2019,** *24,* 4556. doi:10.3390/molecules24244556
37. Medema, M. H.; Takano, E.; Breitling, R. *Mol. Biol. Evol.* **2013,** *30,* 1218–1223. doi:10.1093/molbev/mst025
38. Bradford, M. M. *Anal. Biochem.* **1976,** *72,* 248–254. doi:10.1016/0003-2697(76)90527-3

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

The stereochemical course of 2-methylisoborneol biosynthesis

Binbin Gu, Anwei Hou and Jeroen S. Dickschat*

Letter

Open Access

Address:
Kekulé-Institute of Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk-Straße 1, 53121 Bonn, Germany

Email:
Jeroen S. Dickschat* - dickschat@uni-bonn.de

* Corresponding author

Keywords:
biosynthesis; enantioselective synthesis; enzyme mechanisms; gas chromatography; terpenoids

Beilstein J. Org. Chem. **2022,** *18,* 818–824.
https://doi.org/10.3762/bjoc.18.82

Received: 23 May 2022
Accepted: 05 July 2022
Published: 08 July 2022

Associate Editor: C. Stephenson

Abstract

Both enantiomers of 2-methyllinalyl diphosphate (2-Me-LPP) were synthesized enantioselectively using Sharpless epoxidation as a key step and purification of enantiomerically enriched intermediates through HPLC separation on a chiral stationary phase. Their enzymatic conversion with 2-methylisoborneol synthase (2MIBS) demonstrates that (*R*)-2-Me-LPP is the on-pathway intermediate, while a minor formation of 2-methylisoborneol from (*S*)-2-Me-LPP may be explained by isomerization to 2-Me-GPP and then to (*R*)-2-Me-LPP.

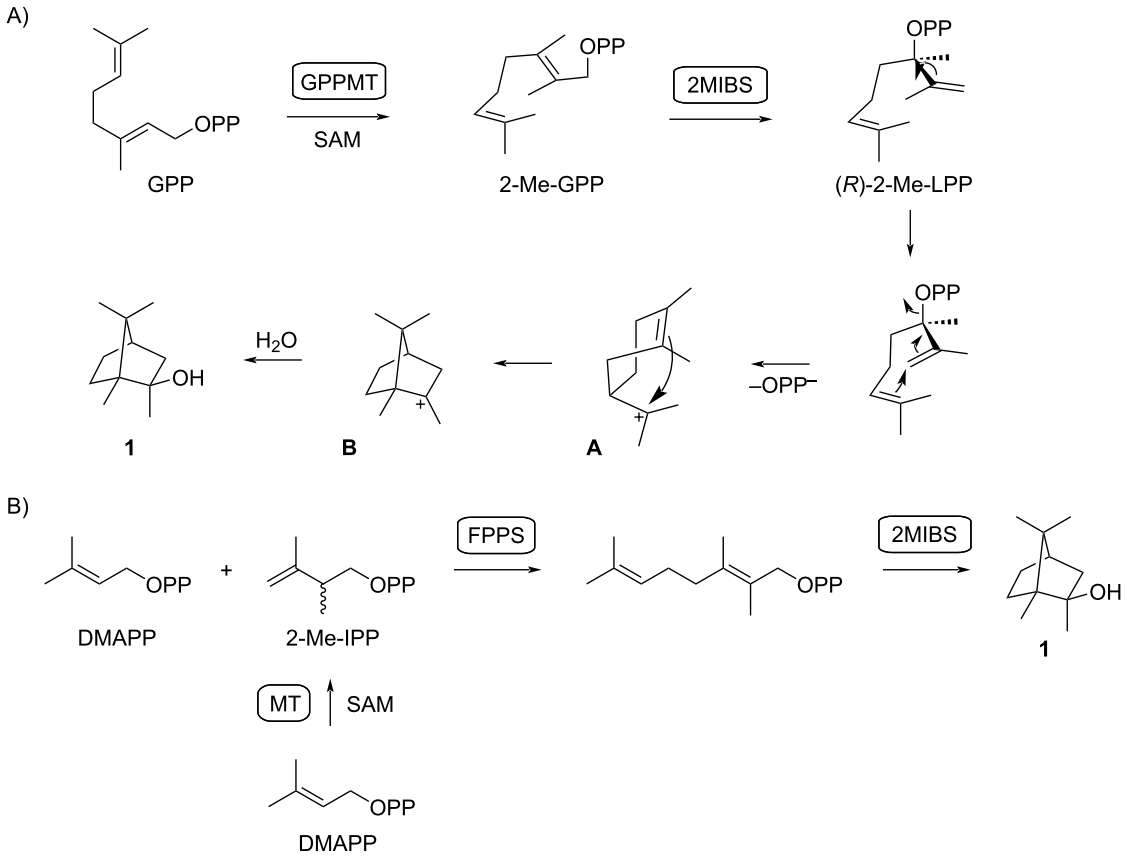
Introduction

After its first discovery from *Streptomyces* [1,2], it has been recognized that many soil bacteria including various genera from the actinobacteria [3-7] and myxobacteria [8] produce the volatile musty odour compound 2-methylisoborneol (**1**). The compound is also found in marine *Streptomyces* strains [9] and aquatic cyanobacteria that can cause drinking water contaminations in water supply systems [10,11]. In addition, the liverwort *Lophocolea heterophylla* [12] and various strains of *Penicillium* [13] have been reported as a source of compound **1**. As a consequence of cheese fermentation with *Penicillium*, compound **1** can add to the flavor of Camembert and Brie [14], but in other foodstuff such as fish and coffee contaminations with **1** are perceived as unpleasant flavor constituents [15-18]. Despite its occurrence in fungi, **1** also has moderate antifungal activity as observed for its inhibition of mycelial growth and sporulation in *Fusarium moniliforme* [19]. Recent research on its chemical ecology demonstrated that arthropodes are attracted by compound **1** which helps in the dispersion of *Streptomyces* spores [20].

The absolute configuration of (–)-**1** has been established through a synthesis from (+)-camphor [21]. The biosynthesis of compound **1** was initially suggested to proceed through degradation of a sesquiterpene [2], but first feeding experiments with ^{14}C-labeled acetate and methionine pointed to a methylated monoterpene [22]. Further investigations by feeding of

^{13}C-labeled methionine and deuterated mevalonolactone isotopomers to *Nannocystis exedens* resulted in a biosynthetic model that includes the methylation of geranyl diphosphate (GPP) to 2-methyl-GPP (2-Me-GPP), followed by cyclization to compound **1** (Scheme 1A) [8]. This process involves the isomerization of 2-Me-GPP by allylic transposition of diphosphate to 2-methyllinalyl diphosphate (2-Me-LPP), followed by a conformational change through rotation around the C2–C3 bond and cyclization to the 2-methyl-α-terpinyl cation (**A**). A second cyclization to **B** and attack of water results in 2-methylisoborneol (**1**) [8]. The stereochemical details of this cyclization cascade were first suggested by Cane, with processing through (*R*)-2-Me-LPP [23]. The GPP methyltransferase (GPPMT) and the 2-methylisoborneol synthase (2MIBS) and their coding genes were discovered and functionally characterized, giving further evidence for the biosynthetic pathway to compound **1** [23-25]. As we have recently demonstrated, the biosynthesis of **1** can also be reconstituted in vitro through coupling of dimethylallyl diphosphate (DMAPP) with 2-methyl-IPP (2-Me-IPP; IPP = isopentenyl diphosphate) to 2-Me-GPP using farnesyl diphosphate synthase (FPPS), followed by cyclization through 2MIBS to **1** [26]. A recently described methyltransferase from *Micromonospora humi* can convert DMAPP into (*R*)-2-Me-IPP with a methyltransferase [27], naturally providing the C_6 building block for this hypothetical alternative pathway towards **1** (Scheme 1B).

Today the genomes of many bacteria from the genus *Streptomyces* have been made available, showing that the genes for the biosynthesis of **1** are present in about half of the species [28], which is reflected by the frequent detection of **1** among the volatiles emitted by a large number of streptomycetes and closely related bacteria [3-7]. A series of side products of the 2MIBS has been identified by GC/MS analysis and synthesis of reference compounds [29], several of which also occur in *Escherichia coli* or yeast strains that were engineered for the biosynthesis of methylated monoterpenes derived from 2-Me-GPP [30,31]. About one decade ago, the crystal structures of GPPMT and 2MIBS have been solved [32,33]. Notably, the structure of 2MIBS has been obtained in complex with the non-reactive substrate analog 2-fluoro-GPP (2FGPP), showing the substrate surrogate in a stretched conformation in the active site



Scheme 1: Biosynthesis of 2-MIB (**1**). A) Naturally observed pathway through methylation of GPP to 2-Me-GPP by GPPMT and cyclization to **1** by 2MIBS. B) Reconstituted pathway through methylation of DMAPP to 2-Me-IPP with a hypothetical MT, followed by coupling of DMAPP and 2-Me-IPP to 2-Me-GPP by FPPS and cyclization to **1** by 2MIBS.

of 2MIBS (Figure 1A). The observed conformation of 2FGPP, if this is also relevant for the native substrate 2-Me-GPP, seems to imply that the isomerization through suprafacial allylic transposition of diphosphate should result in the intermediate (*S*)-2-Me-LPP (Figure 1B), which would be the opposite enantiomer as suggested by Cane [23]. This prompted us to investigate whether (*R*)- or (*S*)-2-Me-LPP is the true pathway intermediate towards compound **1**. For this purpose, both enantiomers of 2-Me-LPP were synthesized and enzymatically converted by 2MIBS. Here we report on the enantioselective synthesis of (*R*)- and (*S*)-2-Me-LPP and the results from the incubation experiments with 2MIBS.

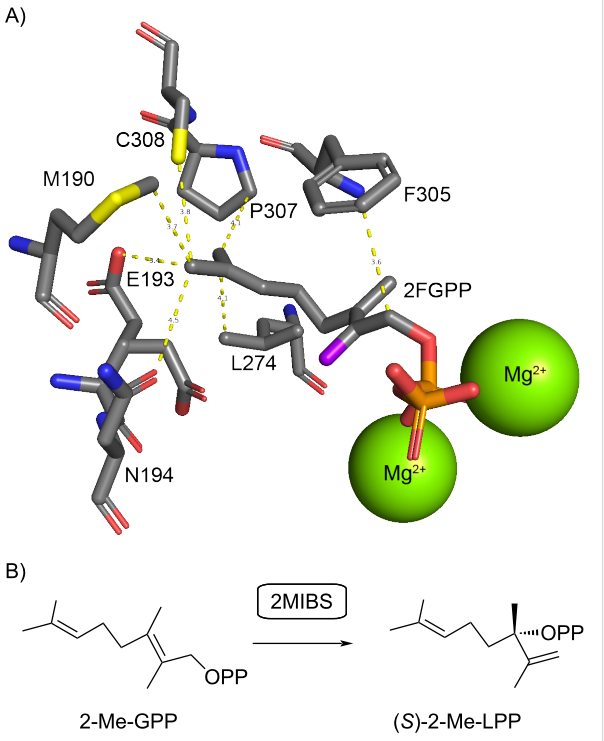
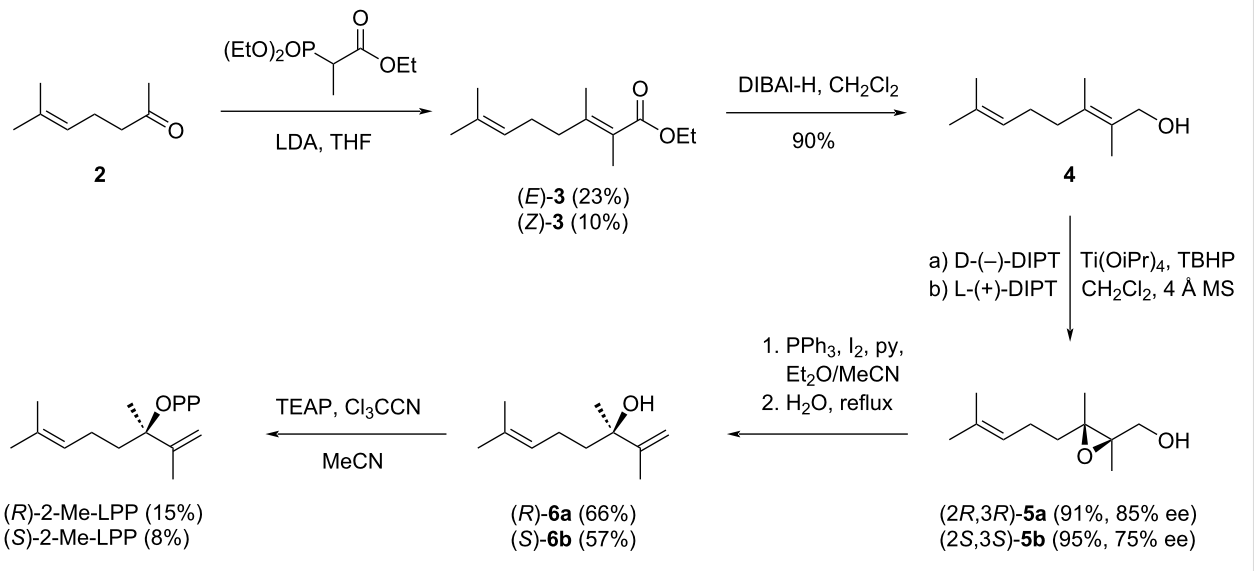


Figure 1: A) Active site of 2MIBS with the bound substrate surrogate 2FGPP (generated with Pymol from the crystal structure, PDB code: 3V1X). B) Isomerization of 2-Me-GPP to (*S*)-2-Me-LPP, the hypothetical enantiomer expected for 2-Me-GPP in the same conformation as observed for 2FGPP.

Results and Discussion

Enantioselective synthesis of 2-methyllinalyl diphosphate

The synthesis of (*R*)- and (*S*)-2-Me-LPP started with the Horner–Wadsworth–Emmons reaction [34,35] of sulcatone (**2**) with triethyl 2-phosphonopropionate to obtain ethyl 2-methylgeranate (**3**) as a mixture of the *E* and *Z* stereoisomers (5:2) that were separated by column chromatography (Scheme 2). Reduction of (*E*)-**3** with DIBAl-H gave 2-methylgeraniol (**4**) that was converted under Sharpless conditions [36] into the epoxides (2*R*,3*R*)-**5a** using D-(−)-diisopropyl tartrate (DIPT) and (2*S*,3*S*)-**5b** with L-(−)-DIPT. The enantiomeric purity of both compounds was determined by small scale conversions with (*S*)-α-methoxy-α-trifluoromethylphenylacetyl chloride (Mosher's acid chloride) [37] and ^{1}H NMR analysis of the products (Figure S1 in Supporting Information File 1), showing enantiomeric purities of 85% ee for **5a** and 75% ee for **5b**. Further conversion by treatment with PPh_3, iodine, pyridine, and water [38] gave access to (*R*)- and (*S*)-2-methyllinalool (**6a** and **6b**). The materials were subsequently converted into the



Scheme 2: Synthesis of (*R*)- and (*S*)-2-Me-LPP.

diphosphates using triethylammonium phosphate and trichloroacetonitrile [39] (Scheme 2).

Conversion of enantiomerically enriched 2-Me-LPP with 2-MIBS

The enantiomerically enriched substrates (*R*)- and (*S*)-2-Me-LPP were incubated with purified 2MIBS (Figure S2 in Supporting Information File 1), followed by extraction of the enzyme reactions with hexane and GC/MS analysis of the obtained products (Figure S3, Table S1 in Supporting Information File 1). All compounds were identified from their EI mass spectra and retention indices in comparison to synthetic standards [29]. The substrate (*R*)-2-Me-LPP gave high yields of compound **1** (62% of total enzyme products in GC), besides 2-methylenebornane (**10**, 21%) and small amounts of 2-methylmyrcene (**7**, 4%), 2-methyllimonene (**8**, 1%), 2-methyl-α-terpineol (**9**, 9%), 2-methyl-2-bornene (**11**, 1%), and 2-methylenefenchane (**12**, 2%). In contrast, (*S*)-2-Me-LPP yielded compound **9** as the main product (51%) and minor amounts of **1** (31%), besides **7** (8%), **8**, (1%), **10** (6%), **11** (0.3%), and **12** (0.6%) (structures are shown in Figure 2). Reproducibility of these results was demonstrated in triplicates. While these data showed that enantiomerically enriched (*R*)-2-Me-LPP is more efficiently converted into **1** than the enriched *S* enantiomer, the enantiomeric purity of the substrates was not sufficiently high to decide, if only one enantiomer of LPP serves as the precursor to **1**.

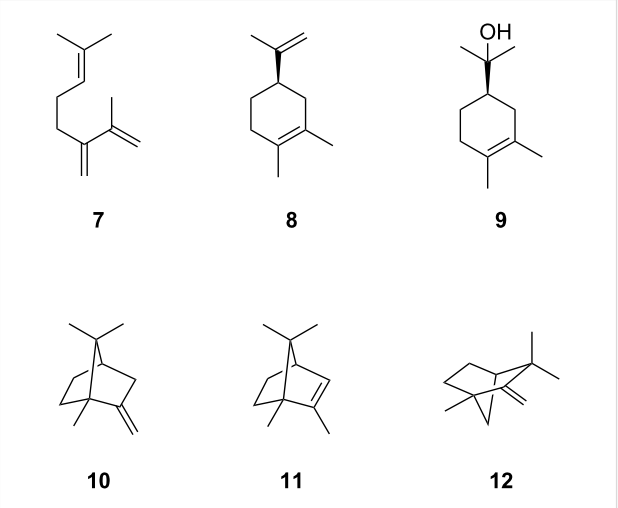


Figure 2: Structures of 2MIBS side products and spontaneous degradation products of 2-Me-LPP. The enantiomers shown are the expected on-pathway intermediates towards **1**.

Purification of the enantiomers of 2-Me-LPP

In order to obtain the enantiomers of 2-Me-LPP with high purity, the synthetically obtained enantiomerically enriched compounds **6a** and **6b** were purified by HPLC using a chiral stationary phase. Ultimately, enantiomeric purities of >99% ee were reached for both **6a** and **6b** (Figure S4, Supporting Information File 1). The compounds were subsequently converted into the diphosphates. To exclude partial racemization during this conversion, small samples of each enantiomer of 2-Me-LPP were dephosphorylated with calf intestinal phosphatase (CIP). The thus obtained compounds **6a** and **6b** were analyzed by gas chromatography on a chiral stationary phase, revealing that the materials were unchanged and still of very high enantiomeric purity (>99% ee, Figure S5 in Supporting Information File 1).

Conversion of enantiomerically pure 2-Me-LPP with 2-MIBS

Both pure enantiomers of 2-Me-LPP were incubated with 2-MIBS. GC/MS analysis of the products (Figure 3A and 3B, and Table S1 in Supporting Information File 1) showed with the substrate (*R*)-2-Me-LPP an efficient conversion into 2-methylisoborneol (**1**, 75%). Minor compounds included 2-methylenebornane (**10**, 13%), 2-methyllinalool (**6**, 8%), 2-methylenefenchane (**12**, 2%), 2-methyl-2-bornene (**11**, 1%), 2-methyl-α-terpineol (**9**, 1%), and 2-methylmyrcene (**7**, 1%). In contrast, (*S*)-2-Me-LPP yielded mainly **9** (55%), but only small amounts of **1** coeluting with **6** (sum: 15%). Further minor products included **10** (13%), **8** (13%), **12** (2%), and **7** (1%). Control experiments by incubation of (*R*)- and (*S*)-2-Me-LPP in buffer containing no enzyme revealed a major non-enzymatic formation of **9** (70%), besides smaller amounts of **6** (22%) and **7** (5%) next to traces of **8** (1%) and another unknown compound (2%, Figure 3C). Reproducibility of these data was again shown in triplicates.

These results demonstrate that (*R*)-2-Me-LPP is the on-pathway intermediate to compound **1**. However, the small amounts of **1** formed from (*S*)-2-Me-LPP (>99% ee) are too large to be explained from the minor enantiomer (*R*)-2-Me-LPP in this sample (<1%). A possible explanation is that (*S*)-2-Me-LPP can bind to the active site of 2MIBS in a non-productive conformation. Its enzyme assisted isomerization to 2-Me-GPP followed by a conformational change may allow for another isomerization to (*R*)-2-Me-LPP and thus lead to the observed minor formation of **1** (Scheme 3). In contrast to the product distribution from (*R*)-2-Me-LPP with **1** as the main and **10** as a side product of 2MIBS, the hydrocarbon **10** is formed from (*S*)-2-Me-LPP in slightly larger amounts than **1**, which could be explained by an incorrect placing or incomplete binding of the active site water involved in the formation of **1**, if (*S*)-2-Me-LPP occupies the active site of 2MIBS.

The formation of **6** and **9** in the incubations of (*R*)- and (*S*)-2-Me-LPP with 2MIBS seems to be non-enzymatic in all cases, because the enantiomeric composition of these products is nearly the same for enzymatic and non-enzymatic reactions,

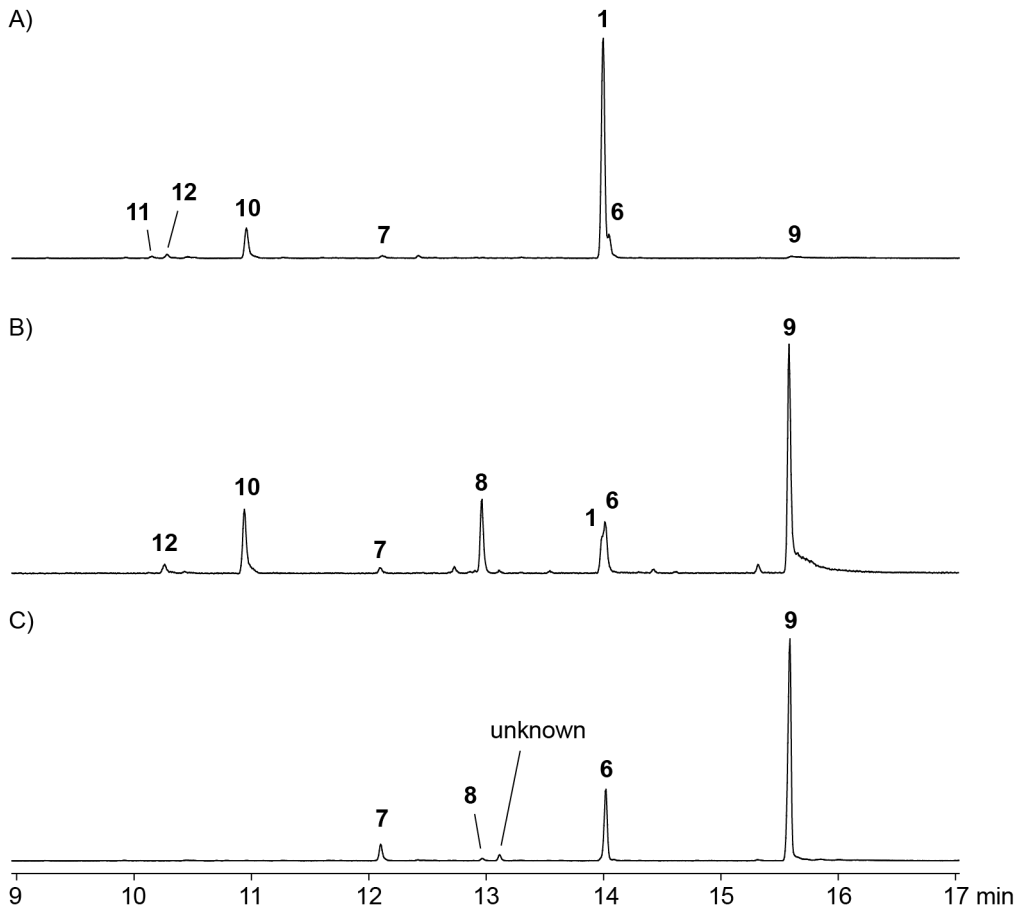
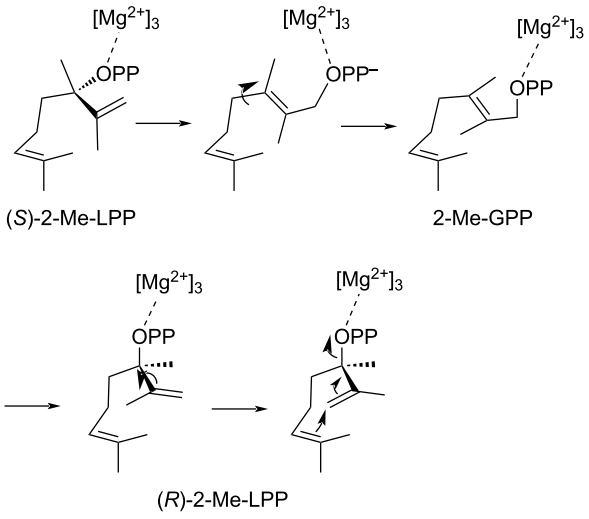


Figure 3: Total ion chromatograms of extracts from an incubation of A) enantiomerically pure (*R*)-2-Me-LPP with 2-MIBS, B) enantiomerically pure (*S*)-2-Me-LPP with 2-MIBS, and C) enantiomerically pure (*R*)-2-Me-LPP without enzyme in incubation buffer. The result for (*S*)-2-Me-LPP without enzyme in incubation buffer was the same as in C) and is not shown.



Scheme 3: Hypothetical mechanism for the isomerization of (*S*)-2-Me-LPP through 2-Me-GPP to (*R*)-2-Me-LPP.

as shown by GC on a chiral stationary phase (Supporting Information File 1, Figures S6 and S7), minor participation of the enzyme in the formation of these products cannot be excluded). In contrast, the formation of **8** from (*S*)-2-Me-LPP must involve the participation of 2MIBS, because its production is clearly enhanced in comparison to the non-enzymatic sample. Investigation of the enantiomeric composition through GC on a chiral stationary phase reveals a 2:3 ratio of enantiomers in favor of (*R*)-**8** (Figure S8, Supporting Information File 1), suggesting that a partial isomerization of (*S*)-2-Me-LPP to (*R*)-2-Me-LPP according to Scheme 3 is also relevant for the formation of **8**. The pseudoracemic mixture of the synthetic compounds **6a** and **6b**, as well as enantiomerically enriched synthetic reference compounds for **8** and **9** (Scheme S1 in Supporting Information File 1) were used for comparison in these analyses.

Conclusion

Both enantiomers of 2-Me-LPP can be selectively prepared using a Sharpless epoxidation strategy in high enantiomeric

purity of >75% ee. HPLC purification of the synthetic precursor 2-methyllinalool using a chiral stationary phase can make the pure enantiomers available, and their conversion into the enantiomers of 2-Me-LPP proceeds without noticeable racemization, leading to materials of >99% ee. Incubation of both pure enantiomers revealed that (*R*)-2-Me-LPP is the on-pathway intermediate towards **1**, while its formation from (*S*)-2-Me-LPP may be explained through isomerization to 2-Me-GPP and then to (*R*)-2-Me-LPP. Conclusively, these findings confirm Cane's mechanistic proposal [23], while the observed conformation of 2FGPP in the crystal structure of 2MIBS may not represent the required conformation for 2-Me-GPP for the production of **1**.

Supporting Information

Supporting Information File 1

Experimental.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-82-S1.pdf]

Acknowledgements

We thank Andreas Schneider for HPLC separation of the enantiomers of 2-methyllinalool.

Funding

This work was supported by the German Research Foundation (DI1536/11-1) and the International Postdoctoral Exchange Fellowship Program (No. 2020038).

ORCID® iDs

Jeroen S. Dickschat - https://orcid.org/0000-0002-0102-0631

References

1. Medsker, L. L.; Jenkins, D.; Thomas, J. F.; Koch, C. *Environ. Sci. Technol.* **1969,** *3,* 476–477. doi:10.1021/es60028a008
2. Gerber, N. N. *J. Antibiot.* **1969,** *22,* 508–509. doi:10.7164/antibiotics.22.508
3. Schöller, C. E. G.; Gürtler, H.; Pedersen, R.; Molin, S.; Wilkins, K. *J. Agric. Food Chem.* **2002,** *50,* 2615–2621. doi:10.1021/jf0116754
4. Wilkins, K.; Schöller, C. *Actinomycetologica* **2009,** *23,* 27–33. doi:10.3209/saj.saj230202
5. Citron, C. A.; Gleitzmann, J.; Laurenzano, G.; Pukall, R.; Dickschat, J. S. *ChemBioChem* **2012,** *13,* 202–214. doi:10.1002/cbic.201100641
6. Rabe, P.; Citron, C. A.; Dickschat, J. S. *ChemBioChem* **2013,** *14,* 2345–2354. doi:10.1002/cbic.201300329
7. Citron, C. A.; Barra, L.; Wink, J.; Dickschat, J. S. *Org. Biomol. Chem.* **2015,** *13,* 2673–2683. doi:10.1039/c4ob02609h
8. Dickschat, J. S.; Nawrath, T.; Thiel, V.; Kunze, B.; Müller, R.; Schulz, S. *Angew. Chem., Int. Ed.* **2007,** *46,* 8287–8290. doi:10.1002/anie.200702496
9. Dickschat, J. S.; Martens, T.; Brinkhoff, T.; Simon, M.; Schulz, S. *Chem. Biodiversity* **2005,** *2,* 837–865. doi:10.1002/cbdv.200590062
10. Izaguirre, G.; Hwang, C. J.; Krasner, S. W.; McGuire, M. J. *Appl. Environ. Microbiol.* **1982,** *43,* 708–714. doi:10.1128/aem.43.3.708-714.1982
11. Yagi, M.; Kajino, M.; Matsuo, U.; Ashitani, K.; Kita, T.; Nakamura, T. *Water Sci. Technol.* **1983,** *15,* 311–321. doi:10.2166/wst.1983.0155
12. Toyota, M.; Asakawa, Y.; Frahm, J.-T. *Phytochemistry* **1990,** *29,* 2334–2337. doi:10.1016/0031-9422(90)83066-a
13. Larsen, T. O.; Frisvad, J. C. *Mycol. Res.* **1995,** *99,* 1153–1166. doi:10.1016/s0953-7562(09)80271-2
14. Karahadian, C.; Josephson, D. B.; Lindsay, R. C. *J. Agric. Food Chem.* **1985,** *33,* 339–343. doi:10.1021/jf00063a005
15. Martin, J. F.; Bennett, L. W.; Graham, W. H. *Water Sci. Technol.* **1988,** *20,* 99–105. doi:10.2166/wst.1988.0230
16. Selli, S.; Prost, C.; Serot, T. *Food Chem.* **2009,** *114,* 317–322. doi:10.1016/j.foodchem.2008.09.038
17. Bai, Z.; Pilote, A.; Sarker, P. K.; Vandenberg, G.; Pawliszyn, J. *Anal. Chem. (Washington, DC, U. S.)* **2013,** *85,* 2328–2332. doi:10.1021/ac3033245
18. Blank, I.; Grosch, W. *J. Agric. Food Chem.* **2002,** *50,* 4653–4656. doi:10.1021/jf0255174
19. Wang, Z.; Wang, C.; Li, F.; Li, Z.; Chen, M.; Wang, Y.; Qiao, X.; Zhang, H. J. *J. Microbiol. (Seoul, Repub. Korea)* **2013,** *51,* 477–483. doi:10.1007/s12275-013-2586-y
20. Becher, P. G.; Verschut, V.; Bibb, M. J.; Bush, M. J.; Molnár, B. P.; Barane, E.; Al-Bassam, M. M.; Chandra, G.; Song, L.; Challis, G. L.; Buttner, M. J.; Flärdh, K. *Nat. Microbiol.* **2020,** *5,* 821–829. doi:10.1038/s41564-020-0697-x
21. Wood, N. F.; Snoeyink, V. L. *J. Chromatogr.* **1977,** *132,* 405–420. doi:10.1016/s0021-9673(00)82905-2
22. Bentley, R.; Meganathan, R. *FEBS Lett.* **1981,** *125,* 220–222. doi:10.1016/0014-5793(81)80723-5
23. Wang, C.-M.; Cane, D. E. *J. Am. Chem. Soc.* **2008,** *130,* 8908–8909. doi:10.1021/ja803639g
24. Komatsu, M.; Tsuda, M.; Ōmura, S.; Oikawa, H.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 7422–7427. doi:10.1073/pnas.0802312105
25. Giglio, S.; Chou, W. K. W.; Ikeda, H.; Cane, D. E.; Monis, P. T. *Environ. Sci. Technol.* **2011,** *45,* 992–998. doi:10.1021/es102992p
26. Gu, B.; Dickschat, J. S. *Chem. Commun.* **2022,** *58,* 4316–4319. doi:10.1039/d2cc00636g
27. Drummond, L.; Haque, P. J.; Gu, B.; Jung, J. S.; Schewe, H.; Dickschat, J. S.; Buchhaupt, M. *ChemBioChem* **2022,** *23,* e202200091.
28. Martín-Sánchez, L.; Singh, K. S.; Avalos, M.; van Wezel, G. P.; Dickschat, J. S.; Garbeva, P. *Beilstein J. Org. Chem.* **2019,** *15,* 1181–1193. doi:10.3762/bjoc.15.115
29. Brock, N. L.; Ravella, S. R.; Schulz, S.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2013,** *52,* 2100–2104. doi:10.1002/anie.201209173
30. Kschowak, M. J.; Wortmann, H.; Dickschat, J. S.; Schrader, J.; Buchhaupt, M. *PLoS One* **2018,** *13,* e0196082. doi:10.1371/journal.pone.0196082
31. 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
32. Köksal, M.; Chou, W. K. W.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2012,** *51,* 3003–3010. doi:10.1021/bi300109c
33. Köksal, M.; Chou, W. K. W.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2012,** *51,* 3011–3020. doi:10.1021/bi201827a

34. Horner, L.; Hoffmann, H.; Wippel, H. G.; Klahre, G. *Chem. Ber.* **1959,** *92,* 2499–2505. doi:10.1002/cber.19590921017
35. Wadsworth, W. S.; Emmons, W. D. *J. Am. Chem. Soc.* **1961,** *83,* 1733–1738. doi:10.1021/ja01468a042
36. Katsuki, T.; Sharpless, K. B. *J. Am. Chem. Soc.* **1980,** *102,* 5974–5976. doi:10.1021/ja00538a077
37. Dale, J. A.; Dull, D. L.; Mosher, H. S. *J. Org. Chem.* **1969,** *34,* 2543–2549. doi:10.1021/jo01261a013
38. Dorta, R. L.; Rodríguez, M. S.; Salazar, J. A.; Suárez, E. *Tetrahedron Lett.* **1997,** *38,* 4675–4678. doi:10.1016/s0040-4039(97)00964-7
39. Banthorpe, D. V.; Christou, P. N.; Pink, C. R.; Watson, D. G. *Phytochemistry* **1983,** *22,* 2465–2468. doi:10.1016/0031-9422(83)80141-1

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Efficient production of clerodane and *ent*-kaurane diterpenes through truncated artificial pathways in *Escherichia coli*

Fang-Ru Li, Xiaoxu Lin, Qian Yang, Ning-Hua Tan and Liao-Bin Dong*

Full Research Paper

Open Access

Address:
State Key Laboratory of Natural Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing 211198, Jiangsu, China

Email:
Liao-Bin Dong* - ldong@cpu.edu.cn

* Corresponding author

Keywords:
artificial pathway; *ent*-kaurene; *Escherichia coli*; overproduction; terpentetriene

Beilstein J. Org. Chem. **2022,** *18,* 881–888.
https://doi.org/10.3762/bjoc.18.89

Received: 30 May 2022
Accepted: 15 July 2022
Published: 21 July 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

The clerodane and *ent*-kaurane diterpenoids are two typical categories of diterpenoid natural products with complicated polycyclic carbon skeletons and significant pharmacological activities. Despite exciting advances in organic chemistry, access to these skeletons is still highly challenging. Using synthetic biology to engineer microbes provides an innovative alternative to bypass synthetic challenges. In this study, we constructed two truncated artificial pathways to efficiently produce terpentetriene and *ent*-kaurene, two representative clerodane and *ent*-kaurane diterpenes, in *Escherichia coli*. Both pathways depend on the exogenous addition of isoprenoid alcohol to reinforce the supply of IPP and DMAPP via two sequential phosphorylation reactions. Optimization of these constructs provided terpentetriene and *ent*-kaurene titers of 66 ± 4 mg/L and 113 ± 7 mg/L, respectively, in shake-flask fermentation. The truncated pathways to overproduce clerodane and *ent*-kaurane skeletons outlined here may provide an attractive route to prepare other privileged diterpene scaffolds.

Introduction

Diterpenoids, of which there are over 34,000 members (http://terokit.qmclab.com), have attracted great attention from chemists and biologists due to their intriguing chemical structures and broad pharmacological functions [1-4]. The vast structural diversity of diterpenoids arise biosynthetically from the following two stages: i) diterpene synthase (DTS, also called diterpene cyclase) act on geranylgeranyl diphosphate (GGDP) to perform regio- and stereoselective cyclizations or skeleton rearrangement reactions via carbocation chemistry to form diverse and versatile carbon skeletons; and ii) multiple post-modification enzymes, most often cytochrome P450s, decorate the carbon skeletons resulting in a large array of oxidative diversity [5-7]. Nature's ability to efficiently biosynthesize diterpenoids has attracted chemists to mimic it for the synthesis of complex diterpenoids using either pure chemical tools, exemplified by the 'two-phase strategy' pioneered by the Baran group or a combination of enzymatic and chemical tools (chemoenzymatic synthesis) [8-11]. Despite great efforts spanning several decades, de novo organic synthetic methods access to the core diterpene skeletons are still highly challenging

owing to their numerous chiral centers and polycyclic complexity [12]. Additionally, chemical transformations from commercial natural products are also tedious and currently limited to a few diterpene skeletons [8].

Engineering microbes via synthetic biology provides new opportunities to produce terpenoid carbon skeletons. All terpenoids are derived from the minimum C_5 isoprenoid building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are produced in the cell via one of two pathways: i) the mevalonate (MVA) pathway includes seven steps from acetyl-CoA (A-CoA); and ii) the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway includes eight steps starting from the condensation of pyruvate and D-glyceraldehyde 3-phosphate (G3P) [13,14] (Figure 1a). Due to these lengthy biosynthetic steps as well as complex metabolic regulations and extensive cofactor requirements, several groups have engineered elegant bypass pathways to mitigate pressures on

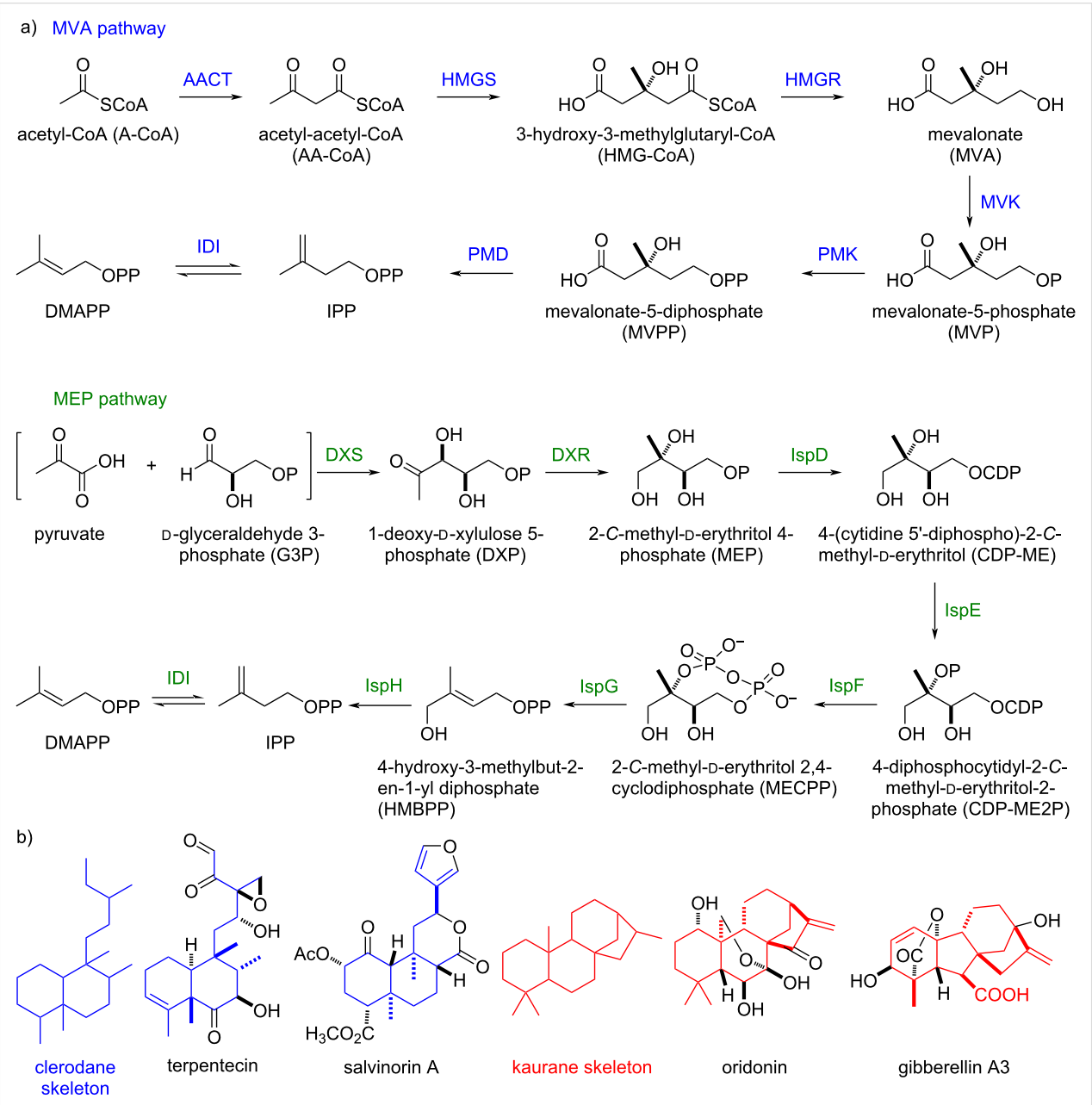


Figure 1: (a) The natural pathways (MVA: blue, MEP: green) for producing IPP and DMAPP; (b) the carbon skeletons of clerodane and kaurane diterpenes and representative bioactive natural products. acetoacetyl-CoA thiolase (AACT); HMG-CoA synthase (HMGS); HMG-CoA reductase (HMGR); mevalonate kinase (MVK); phosphomevalonate kinase (PMK); diphosphomevalonate decarboxylase (PMD); 1-deoxy-D-xylulose 5-phosphate synthase (DXS); 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR).

gatekeeper enzymes [15-18]. However, these systems are still dependent on the entry points within the MVA or MEP pathways [17,18]. Recently, the Stephanopoulos and Williams groups reported two-step artificial pathways to efficiently produce isoprenoid precursors IPP and DMAPP from isopentenol (ISO) and dimethylallyl alcohol (DMAA) [15,19]. In this strategy, two independent kinases were used. ISO and DMAA were phosphorylated to form isopentenyl monophosphate (IP) and dimethylallyl monophosphate (DMAP), respectively, which were then phosphorylated by another kinase to produce IPP and DMAPP [19-21]. Notably, this pathway successfully bypassed the limitations of native isoprenoid biosynthetic pathways, resulting in the overproduction of multiple (mero)terpenoids such as lycopene, *cis*-abienol, and prenylated tryptophan [15,19,22,23].

The clerodane and *ent*-kaurane diterpenoids are two categories of diterpenoids that are widely distributed in terrestrial plants, fungi, and a few bacteria and possess broad pharmacological bioactivities [1-3]. Representative natural products containing these skeletons are terpentecin (cytotoxic and antibiotic), salvinorin A (kappa opioid receptor), oridonin (cytotoxic), and gibberellin (phytohormone) (Figure 1b) [24-27]. How to efficiently construct the core carbon skeletons is a critical question in utilizing the advanced 'two-phase strategy' or chemoenzymatic synthesis to readily synthesize clerodane and *ent*-kaurane diterpenoids. In this paper, we report the reconstruction of truncated artificial pathways to overproduce two representative clerodane and *ent*-kaurane diterpenes, terpentetriene and *ent*-kaurene, in *E. coli*. The titers of terpentetriene and *ent*-kaurene were optimized to 66 ± 4 mg/L and 113 ± 7 mg/L, respectively, in shake-flask fermentation.

Results and Discussion

Constructing a two-step artificial pathway to overproduce IPP and DMAPP precursors

Following the Williams design, *phoN* and *ipk* from *Shigella flexneri* and *Thermoplasma acidophilum*, respectively, were codon-optimized and synthesized for overexpression in *E. coli* [19]. Isopentenyl diphosphate isomerase (IDI) from *E. coli,* which balances IPP and DMAPP in vivo, was also included in our construct. To initially test the efficiency of this two-step artificial pathway, we constructed strains DL10001 (*phoN*, *ipk* and *idi* plus the lycopene-producing genes *crtE*, *crtB*, and *crtI*) and DL10002 (only *crtE*, *crtB*, *crtI*, and *idi*) [28]. Compared to strain DL10002, DL10001 produced significantly larger amounts of lycopene after feeding 6 mM ISO and DMAA (3:1) in a 3-day fermentation (Figure S1, Supporting Information File 1). This result supported that our reconstructed two-step artificial pathway efficiently produced IPP and DMAPP and thus can be used to overproduce the clerodane and *ent*-kaurane diterpenes in *E. coli*.

Collecting the essential genes in the biosynthesis of terpentetriene and *ent*-kaurene

Terpentetriene and *ent*-kaurene are labdane-related diterpenes and biosynthetically constructed by two sequential DTSs from the common C_{20} linear allylic diphosphate GGDP [29]. Terpentetriene was the proposed biosynthetic intermediate of terpentecin, an anticancer and antibiotic natural product isolated from *Kitasatospora griseolosporeus* MF730-N6 in 1985 [24,30]. In the biosynthesis of terpentetriene, GGDP was first cyclized by a class II DTS (Cyc1) that contains a conserved DxDD motif to form terpentedienyl diphosphate (TDP) via a *syn*-labda-13-en-8-yl$^+$ diphosphate intermediate (Figure 2), which, prior to deprotonation, can be followed by rearrangement to form the clerodane skeleton. TDP was then ionized by a class I DTS (Cyc2) that contains a conserved DDxxD motif and through a deprotonation to install a terminal double bond at the side chain [31,32]. We were unable to access the original terpentetriene producing strain of *K. griseolosporeus* MF730-N6, as well as two possible producers (*Streptomyces* sp. San01 and *Kitasatospora* sp. CB02891) after a survey of the existing genome sequence databases, however, we discovered a strain, *Kitasatosporia griseola* DSM 43859, without a genome sequence disclosed, from the strain library of CGMCC. Using the primers designed from the sequences of *cyc1* and *cyc2*, we fortunately obtained two genes of expected lengths, which we named *tdps* and *ttes*, by polymerase chain reaction (PCR). The sequencing results of *tdps* and *ttes* showed extremely high identities with those of *cyc1* and *cyc2* (97% and 99%, respectively) (Table S4, Supporting Information File 1). Additionally, a GGDP synthase named GGDPS was also successfully cloned from the same strain. These results suggested that *K. griseola* DSM 43859 is a different strain with *K. griseolosporeus* MF730-N6 and likely includes a similar biosynthetic gene cluster in the production of terpentetriene. The medium screening for terpentetriene production as well as the elucidation of terpentetriene biosynthetic pathway in *K. griseola* DSM 43859 are underway in our lab. Thus, all essential genes (*phoN*, *ipk*, *idi*, *ggdps*, *tdps*, and *ttes*) for a truncated artificial pathway to produce the clerodane diterpene, terpentetriene, were fully collected.

The biosynthetic pathway towards *ent*-kaurene resembles that of terpentetriene. First, a class II DTS catalyzes the cyclization of GGDP into a diphosphate intermediate, *ent*-copalyl diphosphate (*ent*-CPP). Next, a class I DTS further cyclizes *ent*-CPP into the target tetracyclic skeleton, *ent*-kaurene (Figure 2). In this study, the *ent*-CPP synthase (*e*CDPS) gene was cloned from

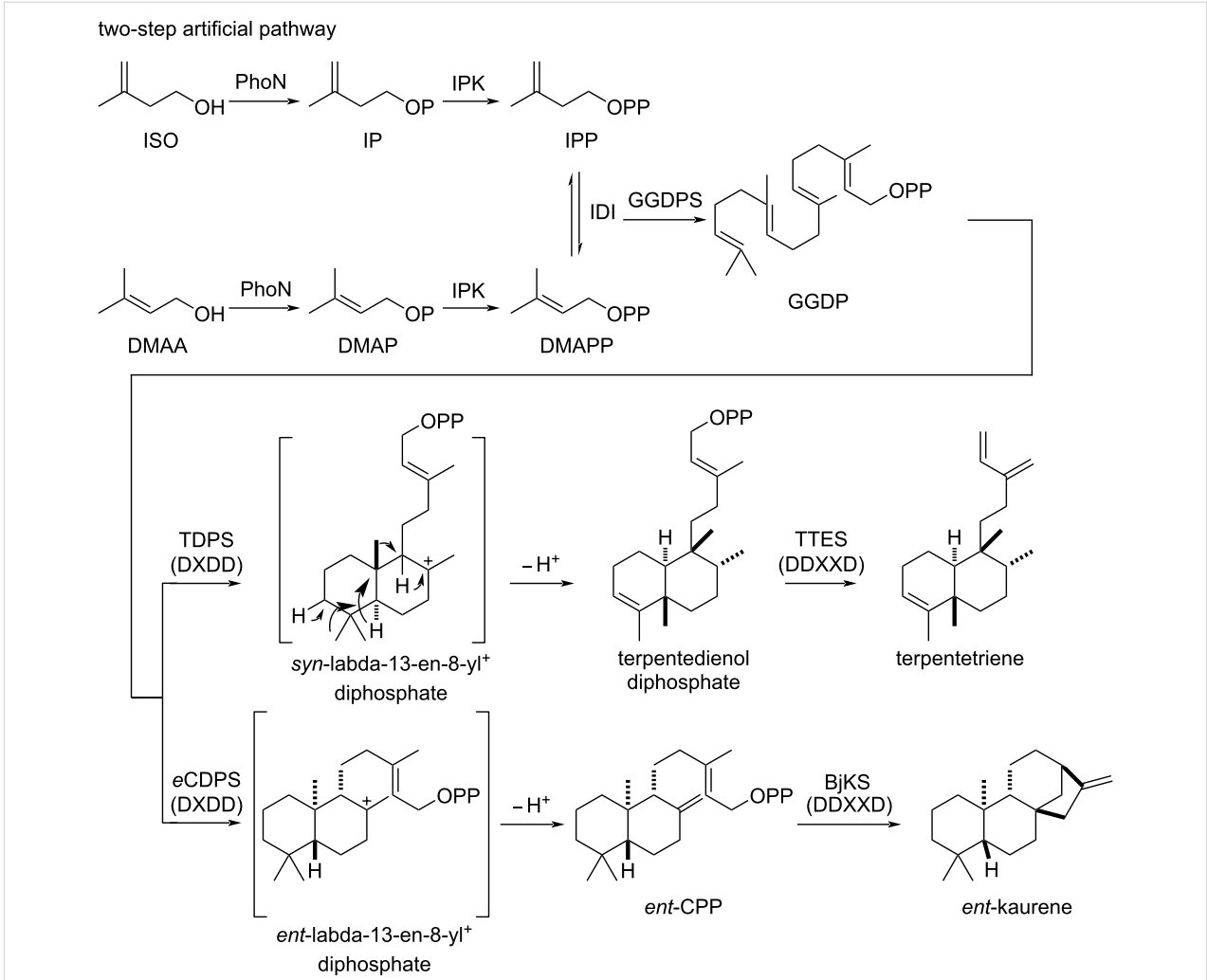


Figure 2: Truncated artificial pathways (six steps) to produce terpentetriene and *ent*-kaurene.

Streptomyces sp. NRRL S-1813, which was an alternative *ent*-kaurenol-derived antibiotic platensimycin producer [33-35], while the *ent*-kaurene synthase (BjKS) gene was from *Bradyrhizobium japonicum*, a bacterial symbiont of soybean that is known to produce the *ent*-kaurene-derived phytohormone gibberellin [36,37]. The *bjks* gene was codon-optimized and synthesized for overexpression in *E. coli*. We used the same *ggdps* from *K. griseola* DSM 43859 in this construct. Thus, the full pathway to *ent*-kaurene, possessing six genes (*phoN*, *ipk*, *idi*, *ggdps*, *ecdps*, and *bjks*), was completed.

Constructing truncated artificial pathways to produce terpentetriene and *ent*-kaurene

After collecting all the essential genes, we initially designed two different expression systems for producing terpentetriene and *ent*-kaurene from ISO and DMAA (Figure 3). To decrease the burden on the host cell [38], the six genes (*phoN*, *ipk*, *idi*, *ggdps*, *tdps*, and *ttes*) leading to the production of terpentetriene, each with a strong and inducible T7 promoter, were cloned into pETDuet-1 plasmid to form pLD10010, which was transformed into *E. coli* BL21 (DE3) to create strain DL10003. In another expression system, the upstream two genes, *phoN* and *ipk*, were cloned into pETDuet-1 plasmid, while the other four genes (*idi*, *ggdps*, *tdps*, and *ttes*) were cloned into pRSF-Duet-1 plasmid. Both plasmids were co-transformed into *E. coli* BL21 (DE3) to create strain DL10004. Using the same protocol, another two strains, DL10005 with a single plasmid expression system and DL10006 with a two-plasmid expression system, that included the whole truncated artificial pathway (*phoN*, *ipk*, *idi*, *ggdps*, *ecdps*, and *bjks*) for *ent*-kaurene production were created.

Next, DL10003–10006 were fermented in Lysogeny broth (LB) medium supplied with 1% glycerol, 6 mM ISO/DMAA 3:1, and 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) inducer. In comparison with the negative control of wild-type *E. coli*

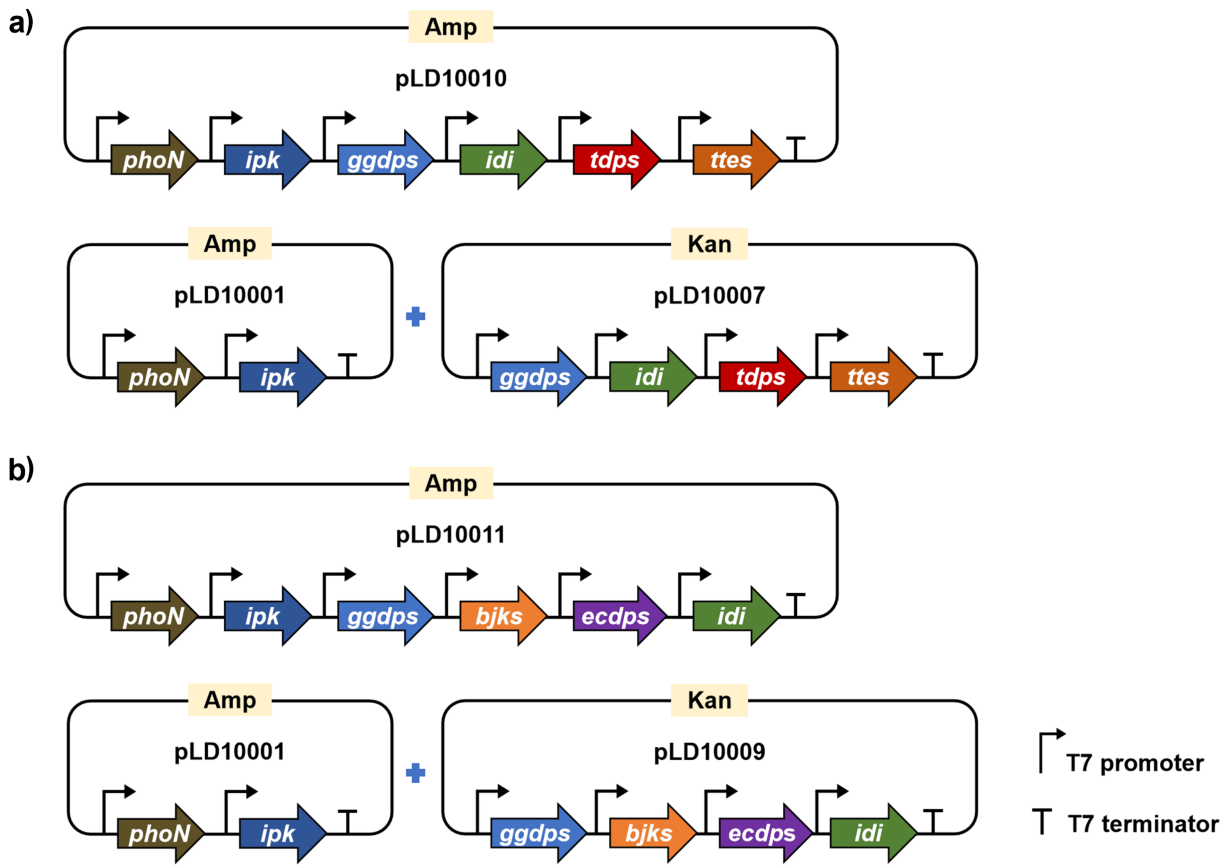


Figure 3: Construction maps of single plasmid expression system and two-plasmid expression system for overproducing terpentetriene (a) and *ent*-kaurene (b) in *E. coli*.

BL21 (DE3), all strains successfully produced new peaks in the HPLC profiles after a 3-day fermentation. Larger scale (3 L) fermentations of DL10004 and DL10006 led to the isolation of 45 mg and 90 mg of terpentetriene and *ent*-kaurene, respectively, whose ^{1}H and ^{13}C NMR spectra supported their chemical structures (Figures S4–S7 in Supporting Information File 1) [31]. These results demonstrated that the two-step artificial pathway coupled with downstream genes for the terpentetriene and *ent*-kaurene biosynthesis was successful and efficient. Though a single plasmid expression system might lower the host cell burden, our results showed that the two-plasmid expression system produced significantly more amounts (14-fold) of terpentetriene and (3-fold) of *ent*-kaurene than single plasmid expression system (22 ± 4 mg/L vs 1.6 ± 0.2 mg/L for terpentetriene, and 27 ± 3 mg/L vs 8.1 ± 0.2 mg/L for *ent*-kaurene). DL10004 and DL10006 were therefore selected for subsequent fermentation optimization.

Optimizing the ISO/DMAA concentrations

To determine the ideal concentrations of exogenous supplementary ISO/DMAA for producing terpentetriene and *ent*-kaurene, a series of feeding experiments was carried out with various concentrations of ISO/DMAA. When strains DL10004 and DL10006 that were cultured in LB medium reached an OD_{600} of 0.6, 0.1 mM IPTG and different amounts of ISO/DMAA were added. After 3-day fermentations in the absence of ISO/DMAA, both strains only produced small amounts of terpentetriene and *ent*-kaurene (1.3 ± 0.1 mg/L and 1.6 ± 0.2 mg/L, respectively), suggesting a low expression level of the endogenous MEP pathway in *E. coli* (Figure 4). When ISO, DMAA, or a mixture of ISO/DMAA 3:1 were exogenously added, the yields of terpentetriene and *ent*-kaurene increased dramatically. The highest yields of terpentetriene (55 ± 3 mg/L) and *ent*-kaurene (53 ± 2 mg/L) were observed in the presence of 25 and 10 mM of DMAA, respectively. These results demonstrated that i) the introduced two-step artificial pathway could efficiently convert exogenous supplemented ISO and DMAA into a pool of hemiterpenes; and ii) IDI effectively balances the proportion of IPP and DMAPP. Given that the commercial DMAA is cheaper than ISO, our result in merely using DMAA should be helpful in decreasing the overall cost for the production of other terpenes using this two-step artificial pathway [22].

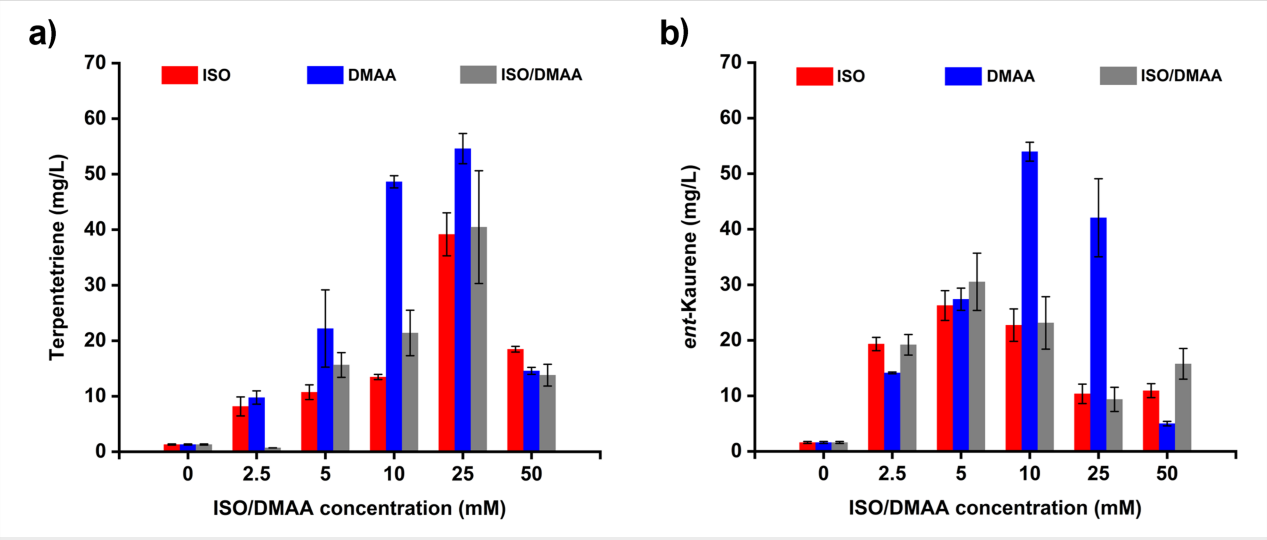


Figure 4: Optimizing the ratios of ISO/DMAA for overproducing terpentetriene (a) and *ent*-kaurene (b). Red: ISO; blue: DMAA; gray: ISO/DMAA 3:1 mixture. All product yields are reported as means ± SD of three replicates.

Optimizing the concentrations of glycerol and IPTG, and fermentation time course

Three orthogonal experiments were run to examine the effects of varying glycerol and IPTG concentrations and fermentation time, all in an effort to optimize terpentetriene and *ent*-kaurene production. Given that the carbon source is vital for overproducing natural products in *E. coli*, and glycerol is one of the most frequently used carbon sources, we first tested for optimal glycerol concentrations for terpentetriene and *ent*-kaurene production. Strains DL10004 and DL10006 were cultured in 50 mL LB medium and supplied with 0%, 1%, 2%, 5%, and 10% (v/v) glycerol, respectively. As showed in Figure 5a, the supplementary 1% (v/v) glycerol led to produce 48 ± 3 mg/L terpentetriene in DL10004, while adding 2% (v/v) glycerol resulted in the yield of 50 ± 4 mg/L *ent*-kaurene in DL10006. The production of terpentetriene and *ent*-kaurene, however, decreased significantly when more glycerol (5% or 10% (v/v)) was added, suggesting that higher glycerol concentrations might be harmful for the host cell, which was also supported by the less cell pellets harvested.

We next explored optimal IPTG concentrations (0, 0.05, 0.1, 0.25, 0.5, and 1 mM) on the production of terpentetriene and *ent*-kaurene. As shown in Figure 5b, 0.1 mM IPTG inducer led to the highest yield of terpentetriene reaching 50 ± 4 mg/L in DL10004. With an increase of IPTG concentration more than 0.1 mM, the yield of terpentetriene decreased dramatically. Additionally, although DL10006 strain could overproduce 74 ± 3 mg/L *ent*-kaurene when 0.5 mM IPTG was added, no significant decrease of the *ent*-kaurene yields under the other two IPTG concentrations (72 ± 4 mg/L for 0.1 mM and 71 ± 2 mg/L for 0.25 mM). From an economic point of view, 0.1 mM IPTG was therefore selected as the optimal concentration for both strains.

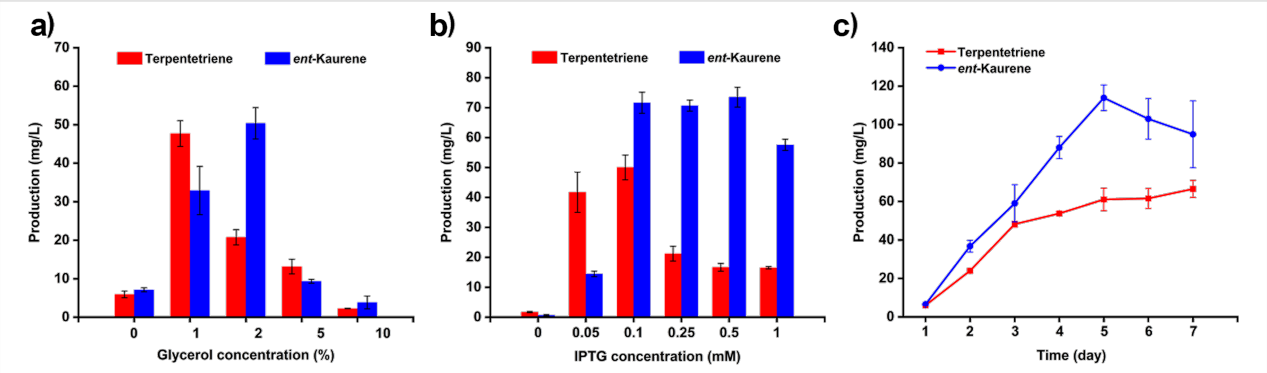


Figure 5: (a) Terpentetriene (red) and *ent*-kaurene (blue) yields supplied with various concentrations of glycerol. (b) Terpentetriene (red) and *ent*-kaurene (blue) yields induced with different concentrations of IPTG. (c) Time course analysis (1–7 days) of terpentetriene (red) and *ent*-kaurene (blue). All product yields are reported as means ± SD of three replicates.

Finally, we tested the fermentation time course of DL10004 and DL10006. Based on the above established fermentation conditions, including the ratios of ISO/DMAA and the concentrations of glycerol and IPTG, we performed a sequence of parallel assays to determine the best fermentation length. Considering the inherent growth feature of *E. coli*, we set up a maximum 7-day experimental timeline for determining the titers of terpentetriene and *ent*-kaurene. As shown in Figure 5c, the yield of terpentetriene was gradually increased with the fermentation time from day 1–7, resulting in a final titer of 66 ± 4 mg/L, while the yield of *ent*-kaurene was gradually increased until day 5 before decreasing to day 7. At day 5, the titer reached 113 ± 7 mg/L for *ent*-kaurene. Taken together, the optimal fermentation conditions of both terpentetriene and *ent*-kaurene were fully established.

Conclusion

In this study, using the truncated artificial pathways, we overproduced two clerodane and *ent*-kaurane diterpenes, terpentetriene and *ent*-kaurene, in *E. coli* by exogenously feeding ISO/DMAA to reinforce the supply of IPP and DMAPP. We optimized the ratio of ISO/DMAA, concentrations of glycerol and IPTG, and fermentation time to enhance the production of terpentetriene and *ent*-kaurene. As a result, strain DL10004, an engineered terpentetriene producer with a two-plasmid expression system, reached the titer of 66 ± 4 mg/L under the optimal conditions of supplementary 25 mM DMAA, 1% glycerol, and 0.1 mM IPTG for a 7-day shake-flask fermentation in LB medium. Strain DL0006, an engineered *ent*-kaurene producer with a two-plasmid expression system, reached the titer of 113 ± 7 mg/L under the optimal conditions of supplementary 10 mM DMAA, 2% glycerol, and 0.1 mM IPTG for a 5-day shake-flask fermentation in LB medium. Compared with the reported optimal combination of efflux pumps for *ent*-kaurene production [39], the titer was enhanced by 3.5-fold in this study. The strategy outlined here not only provides an efficient pathway to overproduce clerodane and *ent*-kaurane carbon skeletons but also offers a blueprint for coupling with emerging chemoenzymatic strategies and biocatalysis in preparation of high value diterpenoid natural products.

Supporting Information

Supporting Information File 1

Experimental part and supplementary figures and tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-89-S1.pdf]

Acknowledgements

The authors thank Dr. Jeffrey D. Rudolf from University of Florida for critical reading and helpful discussion, and Dr. Hui-Min Xu from The Public Laboratory Platform at China Pharmaceutical University for assistance with NMR techniques.

Funding

This work is financially supported, in part, by the National Science Foundation of China (82073746), the Thousand Youth Talents Program of China, the Jiangsu Specially Appointed Professor Program, and the Jiangsu Double Innovation Plan.

ORCID® iDs

Liao-Bin Dong - https://orcid.org/0000-0002-2943-1299

References

1. Rudolf, J. D.; Alsup, T. A.; Xu, B.; Li, Z. *Nat. Prod. Rep.* **2021,** *38,* 905–980. doi:10.1039/d0np00066c
2. Li, R.; Morris-Natschke, S. L.; Lee, K.-H. *Nat. Prod. Rep.* **2016,** *33,* 1166–1226. doi:10.1039/c5np00137d
3. Liu, M.; Wang, W.-G.; Sun, H.-D.; Pu, J.-X. *Nat. Prod. Rep.* **2017,** *34,* 1090–1140. doi:10.1039/c7np00027h
4. Zeng, T.; Chen, Y.; Jian, Y.; Zhang, F.; Wu, R. *New Phytol.* **2022,** *235,* 662–673. doi:10.1111/nph.18133
5. Rudolf, J. D.; Chang, C.-Y. *Nat. Prod. Rep.* **2020,** *37,* 425–463. doi:10.1039/c9np00051h
6. Dickschat, J. S. *Angew. Chem., Int. Ed.* **2019,** *58,* 15964–15976. doi:10.1002/anie.201905312
7. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
8. 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
9. Horn, E. J.; Rosen, B. R.; Chen, Y.; Tang, J.; Chen, K.; Eastgate, M. D.; Baran, P. S. *Nature* **2016,** *533,* 77–81. doi:10.1038/nature17431
10. Jørgensen, L.; McKerrall, S. J.; Kuttruff, C. A.; Ungeheuer, F.; Felding, J.; Baran, P. S. *Science* **2013,** *341,* 878–882. doi:10.1126/science.1241606
11. Kawamura, S.; Chu, H.; Felding, J.; Baran, P. S. *Nature* **2016,** *532,* 90–93. doi:10.1038/nature17153
12. Mendoza, A.; Ishihara, Y.; Baran, P. S. *Nat. Chem.* **2012,** *4,* 21–25. doi:10.1038/nchem.1196
13. Dairi, T.; Kuzuyama, T.; Nishiyama, M.; Fujii, I. *Nat. Prod. Rep.* **2011,** *28,* 1054–1086. doi:10.1039/c0np00047g
14. Frank, A.; Groll, M. *Chem. Rev.* **2017,** *117,* 5675–5703. doi:10.1021/acs.chemrev.6b00537
15. 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
16. Rinaldi, M. A.; Ferraz, C. A.; Scrutton, N. S. *Nat. Prod. Rep.* **2022,** *39,* 90–118. doi:10.1039/d1np00025j
17. Kang, A.; George, K. W.; Wang, G.; Baidoo, E.; Keasling, J. D.; Lee, T. S. *Metab. Eng.* **2016,** *34,* 25–35. doi:10.1016/j.ymben.2015.12.002
18. King, J. R.; Woolston, B. M.; Stephanopoulos, G. *ACS Synth. Biol.* **2017,** *6,* 1416–1426. doi:10.1021/acssynbio.7b00072

19. Lund, S.; Hall, R.; Williams, G. J. *ACS Synth. Biol.* **2019,** *8,* 232–238. doi:10.1021/acssynbio.8b00383
20. Tanaka, N.; Hasan, Z.; Hartog, A. F.; van Herk, T.; Wever, R. *Org. Biomol. Chem.* **2003,** *1,* 2833–2839. doi:10.1039/b304012g
21. Dellas, N.; Thomas, S. T.; Manning, G.; Noel, J. P. *eLife* **2013,** *2,* e00672. doi:10.7554/elife.00672.001
22. Valliere, M. A.; Korman, T. P.; Arbing, M. A.; Bowie, J. U. *Nat. Chem. Biol.* **2020,** *16,* 1427–1433. doi:10.1038/s41589-020-0631-9
23. Zhang, X.; Zhu, K.; Shi, H.; Wang, X.; Zhang, Y.; Wang, F.; Li, X. *J. Cleaner Prod.* **2022,** *351,* 131310. doi:10.1016/j.jclepro.2022.131310
24. Tamamura, T.; Sawa, T.; Isshiki, K.; Masuda, T.; Homma, Y.; Iinuma, H.; Naganawa, H.; Hamada, M.; Takeuchi, T.; Umezawa, H. *J. Antibiot.* **1985,** *38,* 1664–1669. doi:10.7164/antibiotics.38.1664
25. Sheffler, D. J.; Roth, B. L. *Trends Pharmacol. Sci.* **2003,** *24,* 107–109. doi:10.1016/s0165-6147(03)00027-0
26. Sun, H.-D.; Huang, S.-X.; Han, Q.-B. *Nat. Prod. Rep.* **2006,** *23,* 673–698. doi:10.1039/b604174d
27. Mander, L. N. *Nat. Prod. Rep.* **2003,** *20,* 49–69. doi:10.1039/b007744p
28. Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashima, K. *J. Bacteriol.* **1990,** *172,* 6704–6712. doi:10.1128/jb.172.12.6704-6712.1990
29. Peters, R. J. *Nat. Prod. Rep.* **2010,** *27,* 1521–1530. doi:10.1039/c0np00019a
30. Isshiki, K.; Tamamura, T.; Takahashia, Y.; Sawa, T.; Naganawa, H.; Takeuchi, T. *J. Antibiot.* **1985,** *38,* 1819–1821. doi:10.7164/antibiotics.38.1819
31. Dairi, T.; Hamano, Y.; Kuzuyama, T.; Itoh, N.; Furihata, K.; Seto, H. *J. Bacteriol.* **2001,** *183,* 6085–6094. doi:10.1128/jb.183.20.6085-6094.2001
32. Hamano, Y.; Kuzuyama, T.; Itoh, N.; Furihata, K.; Seto, H.; Dairi, T. *J. Biol. Chem.* **2002,** *277,* 37098–37104. doi:10.1074/jbc.m206382200
33. Hsu, S.-Y.; Perusse, D.; Hougard, T.; Smanski, M. J. *ACS Synth. Biol.* **2019,** *8,* 2397–2403. doi:10.1021/acssynbio.9b00261
34. Rudolf, J. D.; Dong, L.-B.; Cao, H.; Hatzos-Skintges, C.; Osipiuk, J.; Endres, M.; Chang, C.-Y.; Ma, M.; Babnigg, G.; Joachimiak, A.; Phillips, G. N., Jr.; Shen, B. *J. Am. Chem. Soc.* **2016,** *138,* 10905–10915. doi:10.1021/jacs.6b04317
35. Rudolf, J. D.; Dong, L.-B.; Shen, B. *Biochem. Pharmacol.* **2017,** *133,* 139–151. doi:10.1016/j.bcp.2016.11.013
36. Morrone, D.; Chambers, J.; Lowry, L.; Kim, G.; Anterola, A.; Bender, K.; Peters, R. J. *FEBS Lett.* **2009,** *583,* 475–480. doi:10.1016/j.febslet.2008.12.052
37. Liu, W.; Feng, X.; Zheng, Y.; Huang, C.-H.; Nakano, C.; Hoshino, T.; Bogue, S.; Ko, T.-P.; Chen, C.-C.; Cui, Y.; Li, J.; Wang, I.; Hsu, S.-T. D.; Oldfield, E.; Guo, R.-T. *Sci. Rep.* **2014,** *4,* 6214. doi:10.1038/srep06214
38. Ma, X.; Liang, H.; Pan, Q.; Prather, K. L. J.; Sinskey, A. J.; Stephanopoulos, G.; Zhou, K. *J. Agric. Food Chem.* **2022,** *70,* 3512–3520. doi:10.1021/acs.jafc.2c00014
39. Wang, J.-F.; Xiong, Z.-Q.; Li, S.-Y.; Wang, Y. *Appl. Microbiol. Biotechnol.* **2013,** *97,* 8057–8067. doi:10.1007/s00253-013-5062-z

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Anti-inflammatory aromadendrane- and cadinane-type sesquiterpenoids from the South China Sea sponge *Acanthella cavernosa*

Shou-Mao Shen[1,2], Qing Yang[2,3], Yi Zang[2], Jia Li[2,4,5], Xueting Liu[*6] and Yue-Wei Guo[*1,2,4]

Full Research Paper

Open Access

Address:
[1]School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing 210023, China, [2]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China, [3]School of Pharmaceutical Science, Nanchang University, Nanchang 330006, China, [4]Shandong Laboratory of Yantai Drug Discovery, Bohai Rim Advanced Research Institute for Drug Discovery, Yantai, Shandong 264117, China, [5]Open Studio for Druggability Research of Marine Natural Products, Pilot National Laboratory for Marine Science and Technology, Qingdao, China and [6]State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China

Email:
Xueting Liu[*] - liuxueting@ecust.edu.cn; Yue-Wei Guo[*] - ywguo@simm.ac.cn

* Corresponding author

Keywords:
Acanthella cavernosa; anti-inflammatory; biosynthetic pathway; chiral separation; marine sponge; sesquiterpenoid

Beilstein J. Org. Chem. **2022,** *18,* 916–925.
https://doi.org/10.3762/bjoc.18.91

Received: 29 May 2022
Accepted: 14 July 2022
Published: 25 July 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

One new aromadendrane-type sesquiterpenoid, namely ximaocavernosin P [(+)-**1**], and three new cadinane-type sesquiterpenoids, namely (+)-maninsigin D [(+)-**4**], (+)- and (−)-ximaocavernosin Q [(+)- and (−)-**5**], together with five related known ones [**2**, **3**, (−)-**4**, **6**, and **7**], were isolated from the Hainan sponge *Acanthella cavernosa*. Compounds **4** and **5** were isolated as racemic forms, which were further separated to the corresponding enantiomers [(+)-**4**/(−)-**4** and (+)-**5**/(−)-**5**], respectively, by using chiral-phase HPLC. The structures of new compounds were elucidated by extensive spectroscopic analysis and comparison with the reported data. In addition, the absolute configuration of optically pure (+)-**1** and **2** were determined by time-dependent density functional theory/electronic circular dichroism (TDDFT-ECD) calculations or X-ray diffraction analysis. A plausible biosynthetic pathway of these sesquiterpenoids and their internal correlation were proposed and discussed. In an in vitro bioassay, (+)-aristolone (**3**) exhibited promising anti-inflammatory activity by the inhibition of LPS-induced TNF-α and CCL2 release in RAW 264.7 macrophages.

Introduction

Marine sponges of the genus *Acanthella* (class Demospongiae, order Halichondrida, family Axinellidae) are one of the most common marine invertebrates natively distributed throughout tropical and subtropical regions of the Indo-Pacific Ocean, in particular, the South China Sea [1]. They are well-known producers of various nitrogenous sesquiterpenes and diterpenes with characteristic isocyano, isothiocyano, and formamido functionalities [2-5]. Many of these secondary metabolites merit further investigation due to their intriguing structural diversity and wide spectra of biological activities ranging from antifeedant, antifouling, and cytotoxic to antibiotic effects [3,5,6]. *Acanthella* sponges have thus attracted much attention from marine natural products chemists and pharmacologists. The title animal is the most chemically studied species among the *Acanthella* sponges. Till now, more than 100 secondary metabolites belonging to sesquiterpenoids and diterpenes [7,8], alkaloids [9], and steroids [10], have been isolated and characterized.

In connection with our continuing studies of Chinese marine invertebrates to search for novel and bioactive secondary metabolites, we have recently studied the sponge *A. cavernosa* collected from Ximao Island of Hainan Province, China, resulting in the isolation of fifteen new nitrogenous sesquiterpenoids, exemplified by (+)-ximaocavernosin A (**8**) (Figure 1) [11]. To accumulate more amounts of these sesquiterpenoids to perform more in-depth pharmacological screening, we carried out the chemical investigation of the same sample on a large scale manner (310 g, dry weight). As a result, besides the already reported nitrogenous sesquiterpenoids, nine non-nitrogenous sesquiterpenoids, including a new aromadendrane-type sesquiterpenoid [(+)-**1**] and three new cadinane-type sesquiterpenoids [(+)-**4**, (+)-**5**, and (−)-**5**], together with five related known ones [**2**, **3**, (−)-**4**, **6**, and **7**] (Figure 1), were obtained. Herein, we report the isolation, chiral separation of racemic mixtures of **4** and **5**, structural elucidation, plausible biosynthetic pathway, and biological evaluation of these isolated compounds.

Results and Discussion

By the similar workup [11] of the Et_2O-soluble portion of the dichloromethane/methanol 1:1 extract of the title animal, five optically pure sesquiterpenoids [(+)-**1** (1.2 mg), **2** (4.1 mg), **3** (3.6 mg), **6** (1.3 mg), and **7** (0.5 mg)], as well as two non-optically pure compounds **4** (2.7 mg) and **5** (4.6 mg) were isolated.

Among them, known compounds **2**, **3**, **6**, and **7** were readily identified as *ent*-4β,10α-dihydroxyaromadendrane (**2**) [12-14], (+)-aristolone (**3**) [15,16], cadalene (**6**) [17], and *trans*-4,5-dihydroxycorocalane (**7**) [18], respectively, by comparing their spectroscopic data and optical rotation values with those reported in the literature. In addition, the full structure of **2**, which was previously isolated from the soft coral *Sinularia mayi* [12], was unambiguously confirmed by X-ray diffraction analysis using Cu Kα radiation (λ = 1.54178 Å) [Flack parameter: 0.00 (11)] (Figure 2), since it was crystallized from MeCN at 4 °C in the present study.

Compound (+)-**1** was obtained as an optically active colorless oil. Its molecular formula was deduced to be $C_{15}H_{22}O_2$ on the

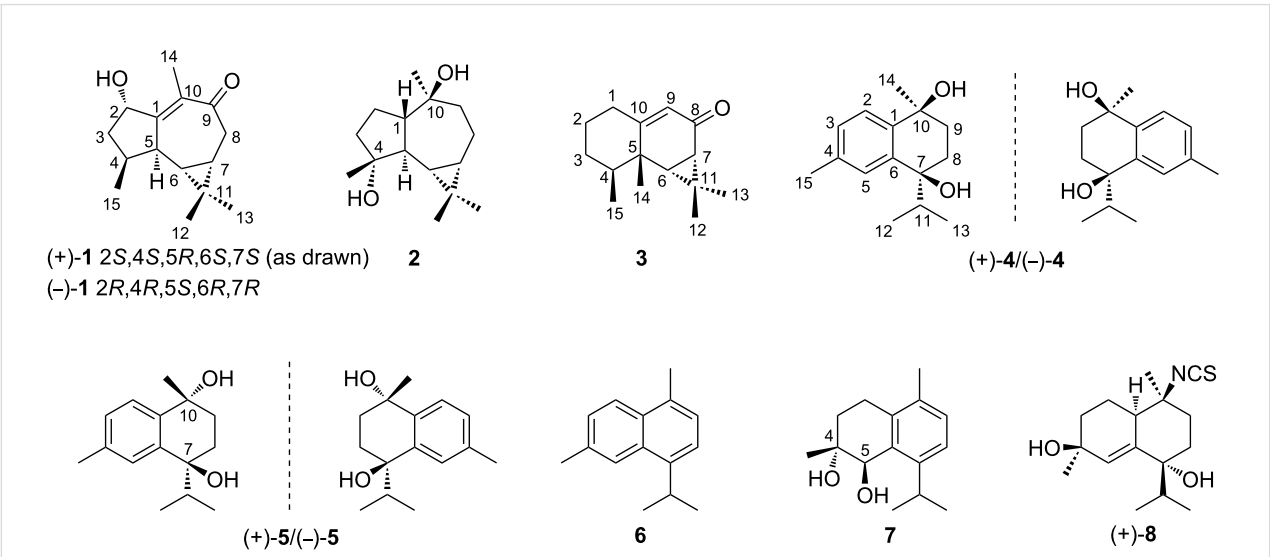


Figure 1: Chemical structures of compounds **1**–**8**.

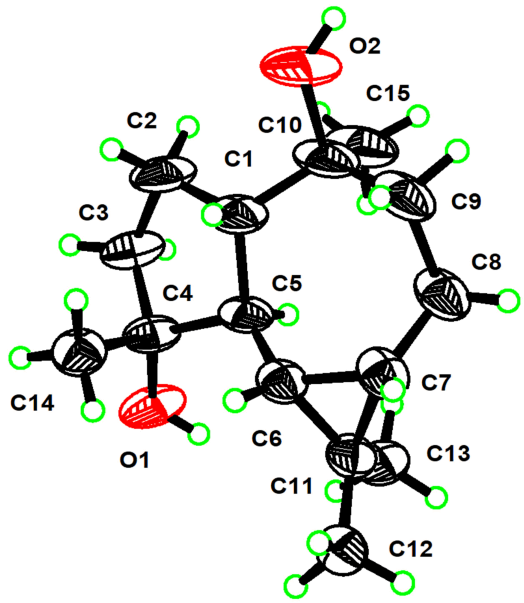


Figure 2: ORTEP drawing of **2** (displacement ellipsoids are drawn at the 50% probability level).

basis of HRESIMS [quasi-molecular ion peak at *m/z* 235.1700 ([M + H]$^+$, calcd for 235.1693)]. The IR spectrum displayed an absorption band of an α,β-unsaturated ketone group (1644 cm^{-1}), as additionally supported by the UV absorption at λ_{max} 245 nm (log ε 3.8). Careful analysis of the NMR spectra of (+)-**1** (Table 1 and Figures S8–S13 in Supporting Information File 1) showed a close similarity with those of co-occurring **2** [12-14], indicating compound (+)-**1** also being an aromadendrane-type sesquiterpenoid containing a *gem*-dimethylcyclopropyl unit [δ_H 0.80 (dd, J = 10.6, 9.6 Hz, H-6), 0.68 (m, H-7), 1.07 (s, 12-Me), 1.19 (s, 13-Me)]. Further literature survey revealed that the overall 1D and 2D NMR data of compound (+)-**1** (Table 1), except the optical rotation signal ($[\alpha]_D^{20}$ +169.6 (*c* 0.12, $CHCl_3$) for (+)-**1**, $[\alpha]_D^{20}$ −186.0 (*c* 6.30, $CHCl_3$) for (−)-**1**), were almost identical to those of 2β-hydroxyaromadendr-1(10)-en-9-one [(−)-**1**] [19], namely (1a*R*,5*R*,7*R*,7a*S*,7b*R*)-1,1a,2,5,6,7,7a,7b-octahydro-5-hydroxy-1,1,4,7-tetramethyl-3*H*-cycloprop[*e*]azulen-3-one, a biotransformation product of squamulosone [aromadendr-1(10)-en-9-one] by the fungus *Curvularia lunata* ATCC 12017 [19], indicating (+)-**1** and (−)-**1** were enantiomeric each other.

Table 1: ^{1}H (*J* in Hz) and ^{13}C NMR data of compounds (+)-**1**, **4**, and **5**.

	(+)-**1**		**4**			**5**		
no.	δ_H[a,b]	δ_C[a,c]	δ_H[b,d]	δ_C[c,d]	δ_C[a,e]	δ_H[b,d]	δ_C[c,d]	δ_C[a,e]
1	–	163.7 s	–	142.8 s	141.7	–	140.4 s	139.0
2	4.92 d (4.8)	74.5 d	7.47 d (8.4)	126.6 d	125.6	7.44 d (7.8)	127.2 d	125.8
3a	1.92 overlap	41.7 t	7.08 dd (8.4, 1.8)	129.2 d	129.0	7.08 dd (7.8, 1.8)	129.0 d	128.9
3b	1.56 dd (13.2, 4.8)	–	–	–	–	–	–	–
4	2.68 m	33.8 d	–	137.6 s	137.5	–	137.8 s	137.9
5	2.73 dd (9.6, 9.6)	43.7 d	7.28 brs	127.4 d	126.0	7.33 brs	127.9 d	126.7
6	0.80 dd (10.6, 9.6)	32.1 d	–	140.9 s	139.9	–	142.1 s	140.6
7	0.68 m	22.6 d	–	75.0 s	74.3	–	75.2 s	74.5
8a	2.38 dd (14.4, 11.4)	–	1.92 ddd (15.6, 10.8, 3.0)	28.4 t	27.9	2.26 ddd (13.8, 10.2, 3.0)	28.3 t	26.9
8b	2.84 dd (14.4, 4.8)	42.1 t	1.84 m	–	–	1.70 ddd (13.8, 8.0, 3.0)	–	–
9a	–	–	2.09 ddd (15.6, 10.8, 3.0)	35.6 t	35.6	2.00 ddd (13.8, 10.2, 3.0)	36.2 t	35.0
9b	–	201.9 s	1.81 m	–	–	1.92 ddd (13.8, 8.0, 1.2)	–	–
10	–	134.1 s	–	71.5 s	71.0	–	70.3 s	69.5
11	–	25.9 s	2.41 m	38.4 d	37.2	2.35 m	38.5 d	37.6
12	1.07 s	28.3 q	0.62 d (6.6)	16.5 q	16.2	0.71 d (6.6)	17.0 q	16.5
13	1.19 s	16.1 q	1.08 d (6.6)	18.9 q	18.7	1.08 d (6.6)	19.1 q	18.9
14	1.94 d (1.8)	14.8 q	1.40 s	30.2 q	29.7	1.54 s	31.2 q	30.6
15	1.05 d (6.6)	15.3 q	2.31 s	21.4 q	21.4	2.32 s	21.3 q	21.5

[a]Recorded in $CDCl_3$, chemical shifts refer to $CHCl_3$ (δ_H 7.26, δ_C 77.2); [b]recorded at 600 MHz; [c]recorded at 125 MHz; [d]recorded in CD_3OD, chemical shifts refer to CD_3OD (δ_H 3.31, δ_C 49.0); [e]recorded at 150 MHz.

To secure the absolute configuration of optically pure (+)-**1**, a TDDFT-ECD calculation, which has proven to be a reliable tool for the absolute configuration determination of natural products with stereogenic centers near the chromophore groups [20], was applied, since there is an α,β-unsaturated ketone chromophore nearby C-5 and C-2 in compound (+)-**1**. Thus, the theoretical ECD spectrum of (+)-**1** was calculated by the DFT calculation method at the b3lyp/6-311G** level of theory (see Supporting Information File 1 for details). A detailed comparison of the Boltzmann-averaged ECD spectrum with that of the experimental one (Figure 3) confirmed the absolute configuration of (+)-**1** as 2*S*, 4*S*, 5*R*, 6*S*, 7*S*. Consequently, the structure of (+)-**1** was assigned as depicted in Figure 1, and named (+)-ximaocavernosin P.

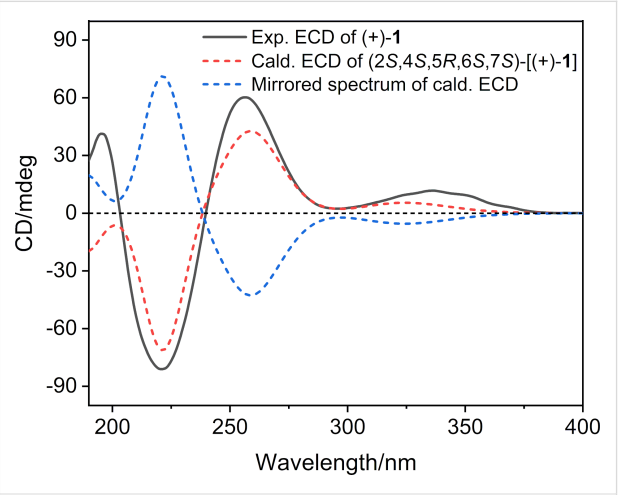


Figure 3: Experimental and calculated ECD curves of (+)-**1**.

Compounds **4** and **5** showed NMR data diagnostic of tetrahydronaphthalene-bearing cadinane-type sesquiterpenoids, in line with the co-occurring compounds **6** and **7**. In addition, **4** and **5** as optically inactive white powder implied the possibility that both of them were present as racemic mixtures.

Compound **4** possessed a molecular formula of $C_{15}H_{22}O_2$. Its structural elucidation was straightforward. The planar structure and relative configuration of **4** were immediately identified to be the same as those of (−)-maninsigin D [(−)-**4**] (Figure 1), a cadinane-type sesquiterpenoid of plant origin (leaves and stems of *Manglietia insignis*) [21], based on their identical ^{1}H and ^{13}C NMR data (Table 1 and Figures S16–S20 in Supporting Information File 1). Since the relative configuration of **4** (7*S**, 10*R**) was assigned, its absolute configuration was worth to be determined. However, the optical value at zero of **4** differed from that of (−)-maninsigin D [(−)-**4**: lit. $[\alpha]_D^{20}$ −9.4 (*c* 0.70, MeOH)] [21], suggesting that **4** should be present as a racemic mixture.

Ximaocavernosin Q (**5**) had the same molecular formula ($C_{15}H_{22}O_2$) as **4**, implying compound **5** to be isomeric with **4**. The IR absorption band at 3386 cm^{-1} suggested the presence of hydroxy groups. A careful analysis of its 1D and 2D NMR spectra revealed that the NMR spectroscopic features of compound **5** (Table 1 and Figures S24–S31 in Supporting Information File 1) extremely resembled those of **4**. The main difference between them was obvious at C-10 and its neighboring carbons (e.g., C-1 and C-14). Considering the insignificant distinction between these chemical shift values ($\Delta\delta \leq 2.4$), compounds **4** and **5** were further analyzed using the same RP-HPLC conditions to examine whether they were the same compound (see Supporting Information File 1 for details). As a result, their distinct retention time [MeCN/H_2O (35:65), 3.0 mL/min, 210 nm, **5**: t_R = 10.6 min, **4**: t_R = 17.6 min] (Figure S1, Supporting Information File 1) confirmed their different identities. Detailed analysis of their NMR spectroscopic data suggested an epimeric relationship between **5** and **4**. Epimerization at C-10 caused upfield shifts for C-10 and C-1 (δ_C 71.5 and 142.8 in **4**; δ_C 70.3 and 140.4 in **5**), whereas the resonance of C-14 was shifted downfield from δ_C 30.2 in **4** to δ_C 31.2 in **5**). Further, the diagnostic NOESY correlations between 14-Me (δ_H 1.54) and H-8b (δ_H 1.70)/H-9a (δ_H 2.00), between H-5 (δ_H 7.33) and H-11 (δ_H 2.35), as well as between H-8a (δ_H 2.26) and 12-Me (δ_H 0.71)/13-Me (δ_H 1.08), as shown in a computer-generated 3D drawing (Figure 4), determined the 7,10-*trans* relationship.

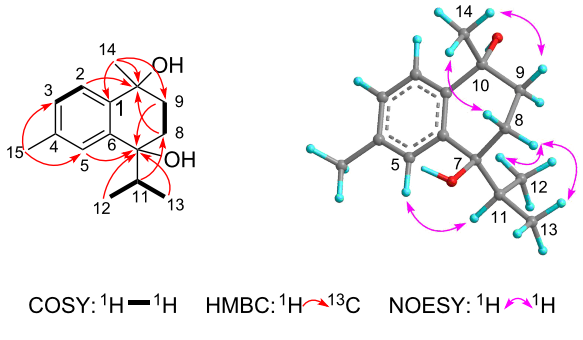


Figure 4: ^{1}H-^{1}H COSY, key HMBC correlations, and NOESY correlations of compound **5**.

To complete the full structure, the next step was to determine the stereochemistry at C-7 and C-10 of compounds **4** and **5**. Since both **4** and **5**, as mentioned above, are respective racemic mixtures, chiral-phase HPLC was then applied to separate each of them (see Figures S2 and S5 in Supporting Information File 1). As expected, new (+)-maninsigin D [(+)-**4**, 0.8 mg] and known (−)-maninsigin D [(−)-**4**, 0.8 mg] were afforded from **4** with opposite optical values [(+)-**4**: $[\alpha]_D^{20}$ +26.0 (*c* 0.08, MeOH); (−)-**4**: $[\alpha]_D^{20}$ −24.2 (*c* 0.08, MeOH)] (Figures S3 and S4 in Supporting Information File 1) and the mirror image-like

ECD curves (Figure 5). (+)-Ximaocavernosin Q [(+)-**5**, 0.8 mg] and (−)-ximaocavernosin Q [(−)-**5**, 0.7 mg] were obtained in the same way from **5** with $[\alpha]_D^{20}$ values of +8.8 (*c* 0.08, MeOH) and −8.3 (*c* 0.07, MeOH) (Figures S6 and S7 in Supporting Information File 1), and mirrored ECD curves (Figure 5). Having obtained two pairs of optically pure enantiomers [(+)-**4**/(−)-**4** and (+)-**5**/(−)-**5**], it is worth determining their absolute configuration. Unfortunately, our efforts to obtain suitable crystals for X-ray diffraction analysis were unsuccessful. The lack of secondary hydroxy groups in these molecules prevented the use of chemical approaches. In addition, their weak Cotton effects (Figure 5) also restricted the application of computational methods. Therefore, the absolute configuration for (+)-**4**/(−)-**4** and (+)-**5**/(−)-**5** remain undefined.

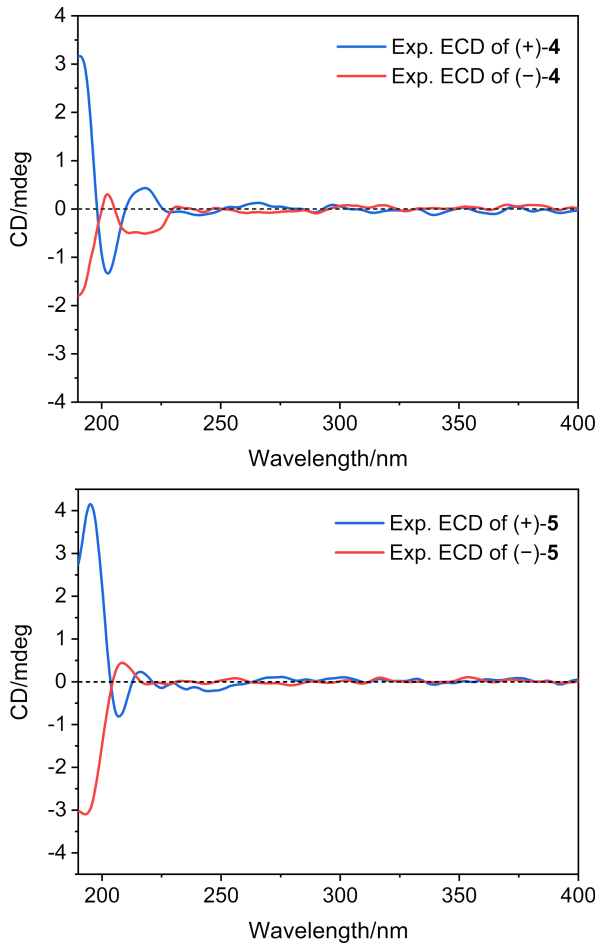
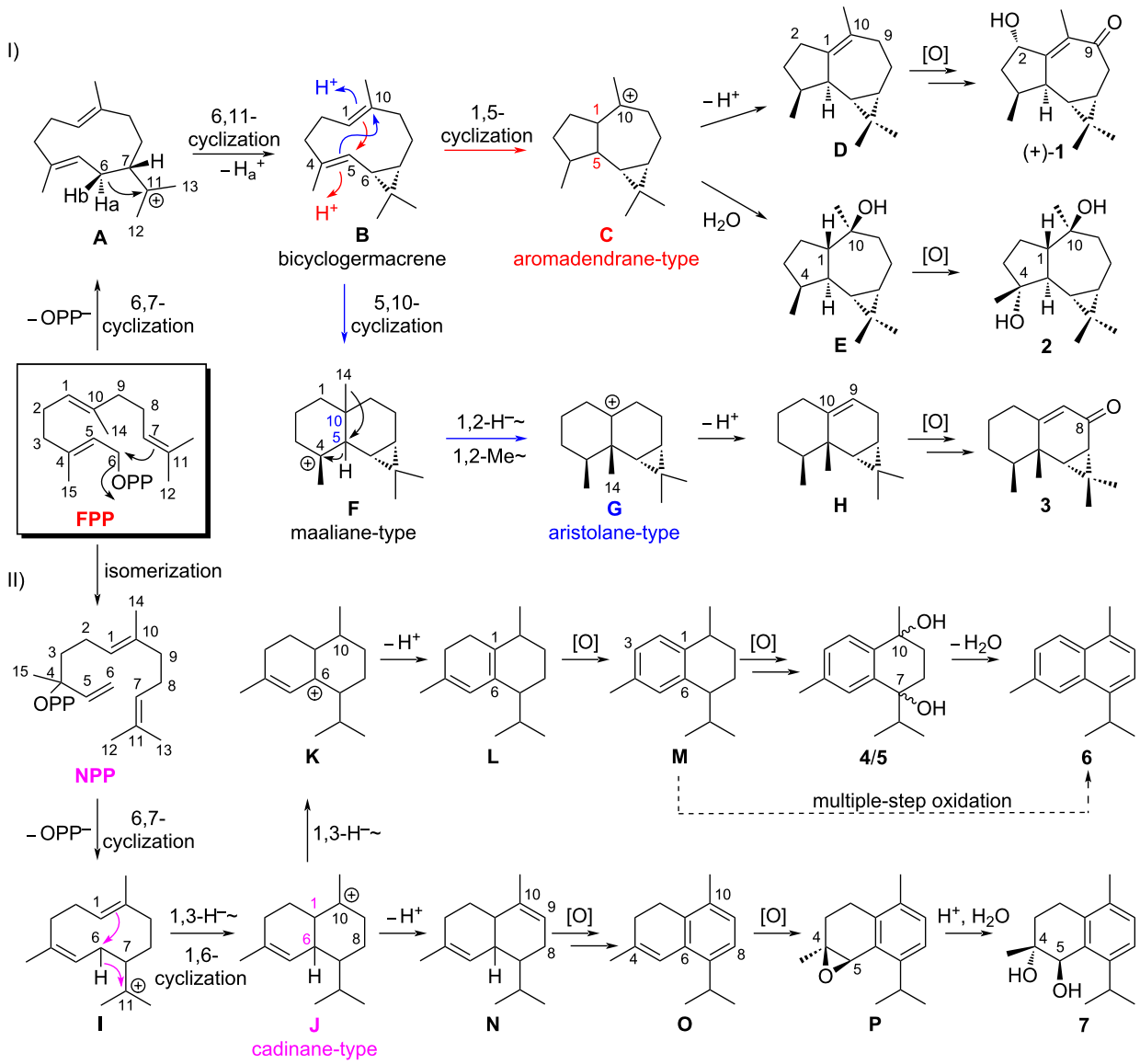


Figure 5: Experimental ECD curves of compounds (+)-**4**, (−)-**4** (top), (+)-**5**, and (−)-**5** (bottom).

As mentioned in a review by William's group [22], the enantiomeric natural products can arise from a single species or different genera and/or species. Several examples of enantiomers formation catalyzed by different terpene synthases were also reported. Jiang et al. characterized two new fungal bifunctional terpene synthases, FoFS and AtAS (identity 27.8%), that catalyzed the formation of a pair of enantiomeric sesterterpenes [23]. Two groups independently reported that the quiannulatene synthases EvQS [24] and AtTPS25 [25] (identity < 10%) from fungus and plant produced (+)- and (−)-quiannulatene, respectively. Interestingly, the cadinane-type sesquiterpenoids obtained from *A. cavernosa* generally exist in enantiomeric forms, such as our recently reported ximaocavernosins A–C, E, and G [11], as well as compounds **4** and **5** obtained in this present study. However, aromadendrane-type sesquiterpenoids from *A. cavernosa* are generally produced in optically pure forms, occasionally enantiomeric with others from a different origin, exemplified by (+)-ximaocavernosin P [(+)-**1**]. Further, these isothiocyano-containing enantiomers were usually isolated as scalable mixture (where one enantiomer predominates) with an enantiomeric excess of ca. 80% [11]. In contrast, the two neutral cadinane-type enantiomers (**4** and **5**) were produced as racemic mixture (1:1 ratio). Different enantiomeric ratios could explain the properties of the active sites in the corresponding terpene synthases, which remain unclear for further investigations [22].

The diversified structures of terpenes were constructed by terpene synthase [26] along with the post-modification enzymes, such as P450 enzymes and other oxygenases [27]. The plausible biosynthetic pathways of the isolated sesquiterpenoids **1**–**7** were proposed, as shown in Scheme 1. Aromadendrane- [(+)-**1** and **2**], aristolane- (**3**) and candinane-type sesquiterpenes [(+)-**4**/(−)-**4**, (+)-**5**/(−)-**5**, **6** and **7**)] were all originated from *E*,*E*-farnesyl diphosphate (*E*,*E*-FPP) as their linear precursor (Scheme 1).

The 6,7-bond formation was triggered by eliminating the pyrophosphate group of *E*,*E*-FPP yielding a monocyclic carbocation intermediate **A**, followed by the 6,11-closure via deprotonation to afford bicyclogermacrene (**B**) containing a *gem*-dimethylcyclopropyl unit [28]. Bicyclogermacrene (**B**) served as the branch point of aromadendrane cyclization and aristolane cyclization routes (Scheme 1, I). On one hand, the protonation on the double bond $\Delta^{1,10}$ of **B** initialized the aromadendrane cyclization, followed by 1,5-bond formation to yield carbocation intermediate **C**. Firstly, deprotonation occurred to form the double bond $\Delta^{1,10}$ and led to the formation of (−)-ledene (**D**) [29], on which the multiple-step oxidation happened at C-2 and C-9 to generate (+)-**1**. Secondly, the reaction is quenched by an H_2O attach at carbocation C-10 to form (+)-globulol (**E**) [30], which was further oxidized to afford **2**. On the other hand, the aristolane cyclization route is started from protonation on $\Delta^{4,5}$ of **B** and followed by 5,10-cyclization to give maaliane-type carbocation intermediate **F**. Then, a



Scheme 1: Proposed cyclization pathway of terpene intermediates and plausible post-modifications of compounds **1**–**7**.

successive transformation involving the concerted 1,2-hydride and 1,2-methyl shifts led to aristolane-type carbocation intermediate **G**, which was further deprotonated to afford 9-aristolene (**H**) [29]. Multiple-step oxidation on **H** furnished the structure of **3**.

The terpene cyclase catalyzed the cyclization of cadinene-type sesquiterpenes using FPP as the substrate, which is first isomerized to nerolidyl diphosphate (NPP), followed by the 6,7-bond formation to generate carbocation intermediate **I** (Scheme 1, II) [31]. Sequential 1,3-hydride shift and 1,6-cyclization occurred to afford cadinyl cation (**J**). Further 1,3-hydride shift and deprotonation on **J** resulted in cadina-1(6),4-diene (**L**) [32], the terpene precursor of compounds **4**–**6**. Meanwhile, direct deprotonation of cadinyl cation (**J**) generates the double bond $\Delta^{9,10}$ of 1,5-cadinadiene (**N**) [32], the precursor of compound **7**. Compounds **4**–**7** belong to the phenolic sesquiterpenes family, and the biosynthesis of the phenolic group has not yet been discovered. Phenolic sesquiterpenoids, for instance, illudacetalic acid and illudinine from *Omphalotus olearius*, were also discovered. Based on bioinformatics analysis, their aromatic rings were proposed to be constructed by putative P450 enzymes or oxidoreductase [33]. The Huang group characterized the function of a P450 enzyme CYP76AH1 which was responsible for the formation of the aromatic ring of ferruginol in the biosynthesis pathway of tanshinones [34]. Hence, we proposed that the oxidation

occurred on **L** to furnish the aromatic ring of calamenene (**M**) [29], followed by the hydroxylation at C-7 and C-10 to give a pair of dihydroxy epimers **4** and **5**. Compound **6** could be generated by the dehydration of **4**/**5** or by the multiple-step oxidation of **M**. Similarly, the plausible biosynthetic pathway of compound **7** could be proposed below. Multiple-step oxidation of **N** gave a phenolic sesquiterpene α-corocalene (**O**) [35]. An epoxidation at C-4/C-5 of **O** resulted in the formation of α-corocalene epoxide (**P**) [35], which was further hydrolyzed to generate the final product **7** [36].

In the anti-inflammatory assay, the transcriptional expression level of the representative inflammatory genes such as tumor necrosis factor-α (TNF-α) and C–C motif chemokine ligand 2 (CCL2) were investigated in lipopolysaccharide (LPS)-stimulated RAW 264.7 macrophages, using NF-κB inhibitor BAY-11-7082 as the positive control. Compound **3** displayed promising dose-dependent anti-inflammatory activity with the inhibition ratios of 74.1% in TNF-α and 64.1% in CCL2 at a concentration of 1 μM (Figure 6).

Conclusion

The continuous chemical investigation of the sponge *A. cavernosa* has resulted in the isolation of nine non-nitrogenous sesquiterpenoids, of which four optically pure compounds [(+)-**1**, (+)-**4,** (+)-**5**, and (−)-**5**] are new. Compounds (+)-**1**, **2**, and **3** share a common *gem*-dimethylcyclopropyl unit, belonging to aromadendrane- and aristolane-type sesquiterpenoids, respectively. Whereas compounds **4**–**7** are a small group of cadinane-type sesquiterpenoids bearing a tetrahydronaphthalene system, which are produced in nature from isoprene units rather than by the much more common routes to aromatics involving acetate or shikimate [37]. In addition, the absolute stereochemistry of (+)-**1** was unambiguously elucidated by TDDFT-ECD calculations, and the complete structure of **2** was confirmed for the first time by X-ray diffraction analysis using Cu Kα radiation. Two pairs of racemic cadinane-type sesquiterpenoids (**4** and **5**) were successfully separated, by chiral-phase HPLC, to their corresponding enantiomers [(+)-**4**/(−)-**4** and (+)-**5**/(−)-**5**]. However, the determination of the absolute configuration of (+)-**4**/(−)-**4** and (+)-**5**/(−)-**5** requires further studies.

Notably, to our best knowledge, although a series of nitrogenous sesquiterpenoids from the genus *Acanthella* have been well documented in literature, less than six non-nitrogenous sesquiterpenoids have been reported from the title sponge [38-40]. The discovery of these sesquiterpenoids not only extended the members of the sesquiterpenoid family but also enriched the chemical diversity of the *Acanthella* sponges. The plausible biosynthetic pathway of isolated compounds **1**–**7** was also proposed. Up to date, no sesquiterpene cyclase has been characterized for aromadendrane and aristolane cyclization. The ubiquitous co-existence of enantiomeric cadinane-type sesquiterpenoids in the sponge revealed the unique catalytic properties of sponge-derived terpene synthases, which attracts our attention for further investigation of their structure–function relationship. Further elucidation of the cyclization mechanism and characterization of post-modification enzymes for the biosynthetic pathway of these sesquiterpenoids will provide insights into expanding the chemical space from marine sponges.

Experimental

General experimental procedure. The melting point was recorded using an SGW X-4 micro-melting point apparatus

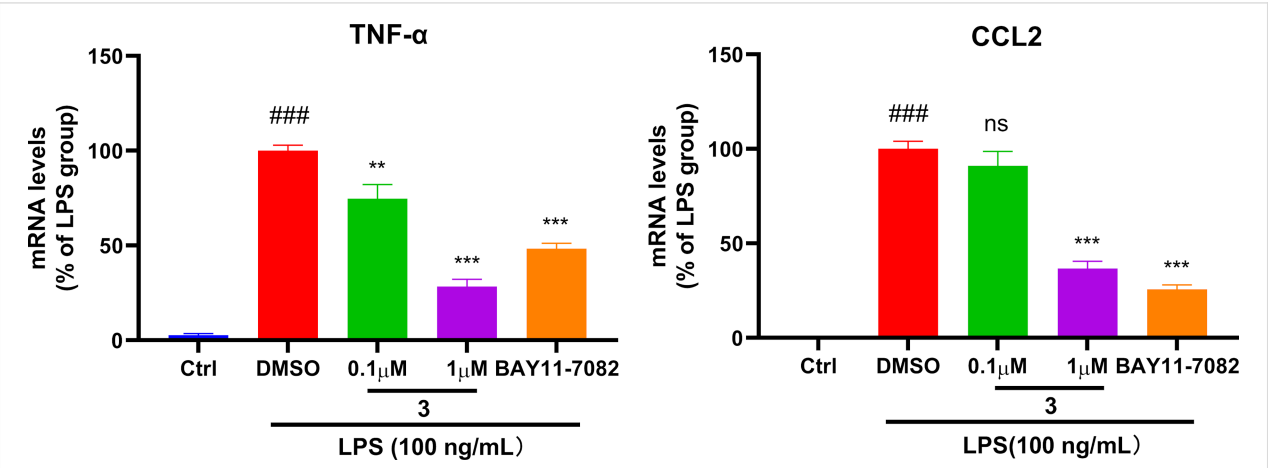


Figure 6: Compound **3** reduced the mRNA levels of TNF-α (left) and CCL2 (right) in LPS-stimulated RAW264.7 macrophages. Data were normalized by DMSO group and are presented as means ± standard errors of the mean (n = 3). $^{###}p < 0.001$ vs the control group; $^{***}p < 0.001$ vs the DMSO group.

(Shanghai Precision Scientific Apparatuses Co., Ltd, Shanghai, China). Optical rotations were measured on a Perkin-Elmer 241MC polarimeter (PerkinElmer, Fremont, CA, USA). IR spectra were recorded on a Nicolet 6700 spectrometer (Thermo Scientific, Waltham, MA, USA). ECD and UV spectra were measured with a JASCO J-815 instrument (JASCO, Japan). NMR spectra were measured on Bruker Avance III 500 and Bruker Avance III 600 instruments (Bruker Biospin AG, Rheinstetten, Germany) using TMS as an internal standard. Chemical shifts (δ) were reported with reference to the solvent signals, and coupling constants (*J*) were in Hz. HRESIMS spectra were recorded on an Agilent G6520 Q-TOF mass spectrometer, while HREIMS spectra were recorded on a Finnigan-MAT-95 mass spectrometer (Thermo Fisher Scientific, Waltham, USA). Commercial silica gel (Qingdao Haiyang Chemical Group Co., Ltd., Qingdao, China, 200–300 and 300–400 mesh), Sephadex LH-20 gel (Amersham Biosciences), and MCI gel (CHP 20P, 75–150 μm, Mitsubishi Chemical Co., Ltd., Tokyo, Japan) were used for column chromatography (CC). Reversed-phase (RP) HPLC was performed on an Agilent 1260 series liquid chromatography equipped with a DAD G1315D detector and an Agilent Eclipse XDB-C_{18} column (5 μm, 9.4 × 250 mm). In contrast, chiral HPLC separation was operated on the chromatography equipped with CHIRALPAK IB N-3 column (5 μm, 4.6 × 250 mm). All solvents used for CC and HPLC were analytical grade (Shanghai Chemical Reagents Co., Ltd., Shanghai, China) and chromatographic grade (DiKMA Technologies Inc., CA, USA), respectively.

Animal materials. See ref [11].

Extraction and isolation. For the extraction and preliminary fractionation of the extract, see ref [11]. The extract was separated by MCI column chromatography (CC) eluted with a MeOH/H_2O gradient solvent system (10% to 100%) to yield six fractions (A–F). The fraction B (570 mg) was then fractionated into subfractions B1–B6 by Sephadex LH-20 CC eluted with CH_2Cl_2. The subfraction B4 (96.2 mg) was further separated by Sephadex LH-20 CC (petroleum ether (PE)/CH_2Cl_2/MeOH 2:1:1) to give subfractions B4A and B4B. The subfraction B4B (81.0 mg) was chromatographed by RP-HPLC (35% to 50% MeCN in H_2O, 3.0 mL/min), yielding compounds **5** (4.6 mg, t_R = 9.9 min), **7** (0.5 mg, t_R = 13.8 min), (+)-**1** (1.2 mg, t_R = 18.9 min) and **2** (4.1 mg, t_R = 10.3 min). The subfraction B5 (80.1 mg) was purified by RP-HPLC (65% MeOH in H_2O, 3.0 mL/min) to give **4** (2.7 mg, t_R = 16.2 min). The subfraction C2 (66.5 mg) was successively separated by silica gel CC (PE/Et_2O 5:1 to 1:1) and RP-HPLC (75% MeCN in H_2O, 3.0 mL/min) to give **3** (3.6 mg, t_R = 9.8 min). The subfraction E (34.9 mg) was successively separated by Sephadex LH-20 CC (PE/CH_2Cl_2/MeOH, 2:1:1) and RP-HPLC (55% MeCN in H_2O, 3.0 mL/min) to afford compound **6** (1.3 mg, t_R = 12.4 min).

The separation of racemic mixtures (**4** and **5**) was performed on chiral HPLC equipped with a CHIRALPAK IB N-3 column eluted with MeCN/H_2O [**4**: 25:75, **5**: 15:85, respectively, 1.0 mL/min] at room temperature, yielding (+)-**4** (0.8 mg, t_R = 18.4 min) and (−)-**4** (0.8 mg, t_R = 16.7 min), (+)-**5** (0.8 mg, t_R = 30.4 min), (−)-**5** (0.7 mg, t_R = 31.9 min), respectively.

(+)-Ximaocavernosin P [(+)-**1**]: colorless oil; $[\alpha]_D^{20}$ +169.6 (*c* 0.12, $CHCl_3$); UV (MeCN) λ_{max}, nm (log ε): 245 (3.8) nm; ECD (MeCN) λ_{max}, nm (Δε): 222 (−9.7), 257 (+7.2), 340 (+1.3); IR (KBr) ν_{max}: 3435, 2956, 2927, 2872, 1644, 1383, 1261, 1109, 1037 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRMS–ESI (*m/z*): $[M + H]^+$ calcd for $C_{15}H_{23}O_2$, 235.1693; found, 235.1700.

ent-4β,10α-Dihydroxyaromadendrane (**2**): colorless crystal; mp 117.0–118.0 °C; $[\alpha]_D^{20}$ +23.8 (*c* 0.40, MeOH).

(+)-Maninsigin D [(+)-**4**]: white powder; $[\alpha]_D^{20}$ +26.0 (*c* 0.08, MeOH); ECD (MeCN) λ_{max}, nm (Δε): 203 (−0.5), 218 (+0.2); IR (KBr) ν_{max}: 2925, 1699, 1632, 1435, 1384, 1259, 1243, 1116, 1069 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRMS–ESI (*m/z*): $[M + H]^+$ calcd for $C_{15}H_{23}O_2$, 235.1693; found, 235.1689.

(+)-Ximaocavernosin Q [(+)-**5**]: white powder; $[\alpha]_D^{20}$ +8.8 (*c* 0.08, MeOH); ECD (MeCN) λ_{max}, nm (Δε): 207 (−0.3), 216 (+0.1); IR (KBr) ν_{max}: 3386, 2961, 2925, 2874, 1463, 1383, 1130, 993, 939, 820 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRMS–EI (*m/z*): M^+ calcd for $C_{15}H_{22}O_2$, 234.1614; found, 234.1618.

(−)-Ximaocavernosin Q [(−)-**5**]: white powder; $[\alpha]_D^{20}$ −8.3 (*c* 0.07, MeOH); ECD (MeCN) λ_{max}, nm (Δε) 208 (+0.2); IR (KBr) ν_{max}: 3386, 2961, 2925, 2874, 1463, 1383, 1130, 993, 939, 820 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRMS–EI (*m/z*): M^+ calcd for $C_{15}H_{22}O_2$, 234.1614; found, 234.1618.

X-ray crystallographic analysis. Compound **2** was crystallized from MeCN at 4 °C. The crystallographic data for compound **2** was collected on a Bruker D8 Venture diffractometer using Cu Kα radiation (λ = 1.54178 Å). The collected data integration and reduction were processed with SAINT V8.37A software, and multi-scan absorption correction was performed using the SADABS program. The structure was solved using ShelXTL and refined on F^2 by the full-matrix least-squares technique using the SHELXL-2015 program package. The

crystallographic data has been uploaded to the Cambridge Crystallographic Data Centre with CCDC number 2173439 (**2**). The data could be obtained free of charge via http://www.ccdc.cam.ac.uk/data_request/cif. Details of the crystallographic data were shown in Supporting Information File 1.

Computational section. TDDFT-ECD calculations. Conformational searches were conducted using the torsional sampling (MCMM) method and the OPLS_2005 force field. Conformers above 1% population were reoptimized at the B3LYP/6-311G** level of theory with IEFPCM (Polarizable Continuum Model using the Integral Equation Formalism variant) solvent model for acetonitrile. For the resulting geometries, ECD spectra were obtained by TDDFT calculations performed with the B3LYP/6-311G** level of theory with IEFPCM solvent model for acetonitrile. Finally, the Boltzmann-averaged ECD spectra were obtained with SpecDis1.62 [41].

In vitro anti-inflammatory assay. RAW264.7 cells, a murine macrophage cell line, was purchased from American Type Culture Collection (ATCC, Manassas, VA, USA). RAW264.7 cells were cultured in DMEM (Dulbecco's modified Eagle medium) supplemented with 10% fetal bovine serum, antibiotics (100 U/mL penicillin and 100 U/mL streptomycin), and maintained at 37 °C in a humidified incubator of 5% CO_2. RAW264.7 cells (1.6×10^5/well) were seeded in a 24-well plate overnight for cell adherence. Cells were treated with LPS alone (100 ng/mL, dissolved in DMSO) or with compounds at the indicated concentrations for 24 h. Total RNA was harvested from cells using TRIzol™ reagent (Invitrogen, USA) after being washed thrice with cold PBS. The total RNA (1 μg) was transcribed into cDNA in 20 μL reactions using the PrimeScript RT reagent kit (ABclonal, China) and then diluted to 200 μL. Real-time quantitative PCR (qPCR) was performed using SYBR GREEN QPCR KIT (Bimake, B21202) on Agilent Mx3000P following the manufacturer's instructions. β-ACTIN was used as the normalization control. All reactions were performed in triplicate. The NF-κB inhibitor Bay 11-7082 (5 μM) was used as a positive control.

Supporting Information

Supporting Information File 1

HPLC chromatograms of **4** and **5**, chiral separation of **4** and **5**, X-ray crystallographic data for **2**, spectra of compounds (+)-**1**, **4** and **5**, TDDFT-ECD calculation of compound (+)-**1**.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-91-S1.pdf]

Acknowledgements

We thank Dr. Lin Gong from Institute of Oceanology, Chinese Academy of Sciences for the taxonomic identification of the sponge material.

Funding

This work was financially supported by the National Natural Science Foundation of China (No. 81991521) and the SKLDR/SIMM Project (No. SIMM2013ZZ-06).

ORCID® iDs

Shou-Mao Shen - https://orcid.org/0000-0003-3953-2429
Xueting Liu - https://orcid.org/0000-0002-1322-8253
Yue-Wei Guo - https://orcid.org/0000-0003-0413-2070

References

1. Huang, Z.-G. *Species and distribution of Marine life in China;* China Ocean Press Publishing: Beijing, 2008.
2. Edenborough, M. S.; Herbert, R. B. *Nat. Prod. Rep.* **1988,** *5,* 229–245. doi:10.1039/np9880500229
3. Garson, M. J.; Simpson, J. S. *Nat. Prod. Rep.* **2004,** *21,* 164–179. doi:10.1039/b302359c
4. Emsermann, J.; Kauhl, U.; Opatz, T. *Mar. Drugs* **2016,** *14,* 16. doi:10.3390/md14010016
5. Massarotti, A.; Brunelli, F.; Aprile, S.; Giustiniano, M.; Tron, G. C. *Chem. Rev.* **2021,** *121,* 10742–10788. doi:10.1021/acs.chemrev.1c00143
6. Patra, A.; Chang, C. W. J.; Scheuer, P. J.; Van Duyne, G. D.; Matsumoto, G. K.; Clardy, J. *J. Am. Chem. Soc.* **1984,** *106,* 7981–7983. doi:10.1021/ja00337a061
7. Omar, S.; Albert, C.; Fanni, T.; Crews, P. *J. Org. Chem.* **1988,** *53,* 5971–5972. doi:10.1021/jo00260a034
8. Wu, Q.; Chen, W.-T.; Li, S.-W.; Ye, J.-Y.; Huan, X.-J.; Gavagnin, M.; Yao, L.-G.; Wang, H.; Miao, Z.-H.; Li, X.-W.; Guo, Y.-W. *Mar. Drugs* **2019,** *17,* 56. doi:10.3390/md17010056
9. Grkovic, T.; Blees, J. S.; Bayer, M. M.; Colburn, N. H.; Thomas, C. L.; Henrich, C. J.; Peach, M. L.; McMahon, J. B.; Schmid, T.; Gustafson, K. R. *Mar. Drugs* **2014,** *12,* 4593–4601. doi:10.3390/md12084593
10. Qiu, Y.; Deng, Z. W.; Xu, M.; Li, Q.; Lin, W. H. *Steroids* **2008,** *73,* 1500–1504. doi:10.1016/j.steroids.2008.08.006
11. Shen, S.-M.; Zhang, Z.-Y.; Yao, L.-G.; Wang, J.-R.; Guo, Y.-W.; Li, X.-W. *Chin. J. Chem.* **2022,** *40,* 235–246. doi:10.1002/cjoc.202100676
12. Beechan, C. M.; Djerassi, C.; Eggert, H. *Tetrahedron* **1978,** *34,* 2503–2508. doi:10.1016/0040-4020(78)88378-1
13. Nagashima, F.; Tanaka, H.; Toyota, M.; Hashimoto, T.; Kan, Y.; Takaoka, S.; Tori, M.; Asakawa, Y. *Phytochemistry* **1994,** *36,* 1425–1430. doi:10.1016/s0031-9422(00)89735-6
14. Sun, Z.-H.; Zhang, S.; Shi, J.; Hu, C.-Q. *Chin. J. Org. Chem.* **2004,** *24,* 806–810.
15. Piers, E.; Britton, R. W.; De Waal, W. *Can. J. Chem.* **1969,** *47,* 831–840. doi:10.1139/v69-132
16. Harrigan, G. G.; Ahmad, A.; Baj, N.; Glass, T. E.; Gunatilaka, A. A. L.; Kingston, D. G. I. *J. Nat. Prod.* **1993,** *56,* 921–925. doi:10.1021/np50096a016

17. Kitagawa, I.; Cui, Z.; Son, B. W.; Kobayashi, M.; Kyogoku, Y. *Chem. Pharm. Bull.* **1987,** *35,* 124–135. doi:10.1248/cpb.35.124
18. Kuo, Y.-H.; Yu, M.-T. *Chem. Pharm. Bull.* **1999,** *47,* 1017–1019. doi:10.1248/cpb.47.1017
19. Collins, D. O.; Buchanan, G. O.; Reynolds, W. F.; Reese, P. B. *Phytochemistry* **2001,** *57,* 377–383. doi:10.1016/s0031-9422(01)00060-7
20. Mándi, A.; Kurtán, T. *Nat. Prod. Rep.* **2019,** *36,* 889–918. doi:10.1039/c9np00002j
21. Shang, S.-Z.; Kong, L.-M.; Yang, L.-P.; Jiang, J.; Huang, J.; Zhang, H.-B.; Shi, Y.-M.; Zhao, W.; Li, H.-L.; Luo, H.-R.; Li, Y.; Xiao, W.-L.; Sun, H.-D. *Fitoterapia* **2013,** *84,* 58–63. doi:10.1016/j.fitote.2012.10.010
22. Finefield, J. M.; Sherman, D. H.; Kreitman, M.; Williams, R. M. *Angew. Chem., Int. Ed.* **2012,** *51,* 4802–4836. doi:10.1002/anie.201107204
23. Jiang, L.; Zhang, X.; Sato, Y.; Zhu, G.; Minami, A.; Zhang, W.; Ozaki, T.; Zhu, B.; Wang, Z.; Wang, X.; Lv, K.; Zhang, J.; Wang, Y.; Gao, S.; Liu, C.; Hsiang, T.; Zhang, L.; Oikawa, H.; Liu, X. *Org. Lett.* **2021,** *23,* 4645–4650. doi:10.1021/acs.orglett.1c01361
24. Okada, M.; Matsuda, Y.; Mitsuhashi, T.; Hoshino, S.; Mori, T.; Nakagawa, K.; Quan, Z.; Qin, B.; Zhang, H.; Hayashi, F.; Kawaide, H.; Abe, I. *J. Am. Chem. Soc.* **2016,** *138,* 10011–10018. doi:10.1021/jacs.6b05799
25. Huang, A. C.; Kautsar, S. A.; Hong, Y. J.; Medema, M. H.; Bond, A. D.; Tantillo, D. J.; Osbourn, A. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114,* E6005–E6014. doi:10.1073/pnas.1705567114
26. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
27. Narita, K.; Chiba, R.; Minami, A.; Kodama, M.; Fujii, I.; Gomi, K.; Oikawa, H. *Org. Lett.* **2016,** *18,* 1980–1983. doi:10.1021/acs.orglett.6b00552
28. Durán-Peña, M. J.; Botubol Ares, J. M.; Hanson, J. R.; Collado, I. G.; Hernández-Galán, R. *Nat. Prod. Rep.* **2015,** *32,* 1236–1248. doi:10.1039/c5np00024f
29. Warmers, U.; Wihstutz, K.; Bülow, N.; Fricke, C.; König, W. A. *Phytochemistry* **1998,** *49,* 1723–1731. doi:10.1016/s0031-9422(98)00283-0
30. Stodůlková, E.; Šulc, M.; Císařová, I.; Novák, P.; Kolařík, M.; Flieger, M. *Folia Microbiol.* **2008,** *53,* 15–22. doi:10.1007/s12223-008-0002-5
31. Rinkel, J.; Rabe, P.; Garbeva, P.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 13593–13596. doi:10.1002/anie.201608042
32. Ngo, K.-S.; Wong, W.-T.; Brown, G. D. *J. Nat. Prod.* **1999,** *62,* 549–553. doi:10.1021/np980289m
33. Wawrzyn, G. T.; Quin, M. B.; Choudhary, S.; López-Gallego, F.; Schmidt-Dannert, C. *Chem. Biol.* **2012,** *19,* 772–783. doi:10.1016/j.chembiol.2012.05.012
34. Guo, J.; Zhou, Y. J.; Hillwig, M. L.; Shen, Y.; Yang, L.; Wang, Y.; Zhang, X.; Liu, W.; Peters, R. J.; Chen, X.; Zhao, Z. K.; Huang, L. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110,* 12108–12113. doi:10.1073/pnas.1218061110
35. Weyerstahl, P.; Marschall, H.; Thefeld, K.; Subba, G. C. *Flavour Fragrance J.* **1998,** *13,* 377–388. doi:10.1002/(sici)1099-1026(199811/12)13:6<377::aid-ffj755>3.0.co;2-f
36. Wu, Q.; Li, S.-W.; Xu, H.; Wang, H.; Hu, P.; Zhang, H.; Luo, C.; Chen, K.-X.; Nay, B.; Guo, Y.-W.; Li, X.-W. *Angew. Chem., Int. Ed.* **2020,** *59,* 12105–12112. doi:10.1002/anie.202003643
37. Dewick, P. M. *Medicinal natural products: a biosynthetic approach;* John Wiley & Sons: Chichester, UK, 2002.
38. Okino, T.; Yoshimura, E.; Hirota, H.; Fusetani, N. *J. Nat. Prod.* **1996,** *59,* 1081–1083. doi:10.1021/np960496r
39. Hirota, H.; Tomono, Y.; Fusetani, N. *Tetrahedron* **1996,** *52,* 2359–2368. doi:10.1016/0040-4020(95)01079-3
40. Nogata, Y.; Yoshimura, E.; Shinshima, K.; Kitano, Y.; Sakaguchi, I. *Biofouling* **2003,** *19* (Suppl. 1), 193–196. doi:10.1080/0892701031000065963
41. Wu, Q.; Ye, F.; Li, X.-L.; Liang, L.-F.; Sun, J.; Sun, H.; Guo, Y.-W.; Wang, H. *J. Org. Chem.* **2019,** *84,* 3083–3092. doi:10.1021/acs.joc.8b02912

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Understanding the competing pathways leading to hydropyrene and isoelisabethatriene

Shani Zev[1], Marion Ringel[2], Ronja Driller[3,4], Bernhard Loll[3], Thomas Brück[2] and Dan T. Major[*1]

Full Research Paper

Open Access

Address:
[1]Department of Chemistry and Institute for Nanotechnology & Advanced Materials, Bar-Ilan University, Ramat-Gan 52900, Israel, [2]Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany, [3]Institute for Chemistry and Biochemistry, Structural Biochemistry Laboratory, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany, and [4]Department of Molecular Biology and Genetics, Aarhus University, Danish Research Institute of Translational Neuroscience – DANDRITE, Universitetsbyen 81, 8000 Aarhus C, Denmark

Email:
Dan T. Major[*] - majort@biu.ac.il

* Corresponding author

Keywords:
diterpenes; enzyme mechanism; quantum mechanics; terpene synthases; thermodynamic and kinetic control

Beilstein J. Org. Chem. **2022,** *18,* 972–978.
https://doi.org/10.3762/bjoc.18.97

Received: 26 May 2022
Accepted: 18 July 2022
Published: 04 August 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

Terpene synthases are responsible for the biosynthesis of terpenes, the largest family of natural products. Hydropyrene synthase generates hydropyrene and hydropyrenol as its main products along with two byproducts, isoelisabethatrienes A and B. Fascinatingly, a single active site mutation (M75L) diverts the product distribution towards isoelisabethatrienes A and B. In the current work, we study the competing pathways leading to these products using quantum chemical calculations in the gas phase. We show that there is a great thermodynamic preference for hydropyrene and hydropyrenol formation, and hence most likely in the synthesis of the isoelisabethatriene products kinetic control is at play.

Introduction

Terpenes constitute a ubiquitous class of natural molecules that are synthesized by terpene synthases (TPS). TPS generate a plethora of terpenes employing rich carbocation chemistry from a very limited number of substrates, known as geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP), and geranylgeranyl pyrophosphate (GGPP), to produce mono-, sesqui-, and diterpenes, respectively. The formation of terpenes relies on an assortment of carbocation steps like cyclization, methyl migrations, rearrangements, proton or hydride transfers, hydroxylations, and epoxidations. TPS and downstream functionalizing enzymes, like P450s, together produce more than 80,000 known terpenes and terpenoids [1-3].

Hydropyrene synthase (HpS) from *Streptomyces clavuligerus* generates a mixture of diterpenes named hydropyrene (HP) (52%) and diterpenoid named hydropyrenol (HPol) (26%) as its main GGPP cyclization products, along with two minor compounds, namely the isoelisabethatrienes (IEs) A (13%) and B (9%), respectively. Interestingly, the elisabethatriene diterpene macrocycle and its isoforms can act as biosynthetic precursors of the bioactive compounds erogorgiaene and pseudopterosin, having antibiotic and anti-inflammatory activities, respectively [4,5]. Unexpectedly, a single active site mutation, M75L, significantly shifts the product distribution and IE A becomes the dominant product (44%) in this enzyme variant [6].

As suggested by Rinkel et al., both routes (HP and IE routes) proceed from the same substrate (GGPP) with an initial C1–C10 cyclization. In other TPS enzymes, the initial fold of GGPP in the active site can result in different initial cyclization, for example C1–C6, C1–C7, C1–C10, C1–C11, C1–C14, and C1–C15. The main difference between the two pathways to HP and IE, is that in the HP pathway C1–C10 cyclization occurs immediately, while in the IE pathway a substrate transoid to cisoid conformational change occurs, shifting the covalent attachment point (i.e., C1 to C3) between the substrate hydrocarbon and the pyrophosphate group [7] (Figure 1). Presumably, this isomerization is responsible for a slightly different substrate fold inside the active site, hence shifting the product distribution in favor of the IE products in certain enzyme variants rather than the HP products.

Oxidation of IEs A and B by lipases results in the formation of the advanced pseudopterosin (P) precursor erogorgiaene and (1*R*)-epoxyelisabetha-5,14-diene (EED), respectively [6,7]. Ps, marine amphilectane-type diterpenoids from the gorgonian coral *Antillogorgia elisabethae*, feature superior anti-inflammatory properties which render them innovative target compounds for drug development [8,9]. Hence, increasing the IE products at the expense of the HP products is an important biotechnology mission for sustainable supply of the latter. In order to modulate the IE and HP enzyme pathways accordingly, it is important to understand the factors determining both synthetic routes.

In the current work, we focus on the mechanistic details of the HP and IE pathways using computational methods in the gas phase. Gas-phase studies have been crucial in understanding terpene chemistry [10-22]. This work sheds light on the thermodynamic and kinetic parameters of the inherent chemistry in these reactions and also points to some understanding of the possible thermodynamic and kinetic control in the enzyme.

Results

Reaction mechanism

To better understand the HP and IE reaction pathways, we performed quantum mechanics (QM) calculations using density functional theory (DFT). We studied the inherent chemistry of the reaction leading to HP and IE using gas-phase calculations. This provided the free energy of distinct carbocation intermediates and transition states along the proposed reaction path leading to products in the gas phase. The gas phase is a natural choice as a reference environment for terpene synthases [10-12,15,16,21-25]. The proposed reaction mechanisms yielding HP and IE and are presented in Scheme 1, while the reaction

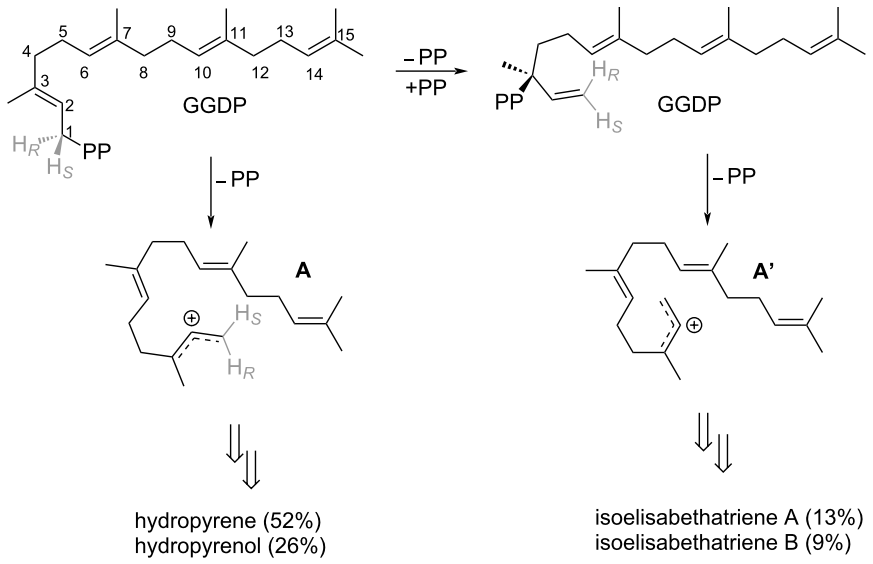
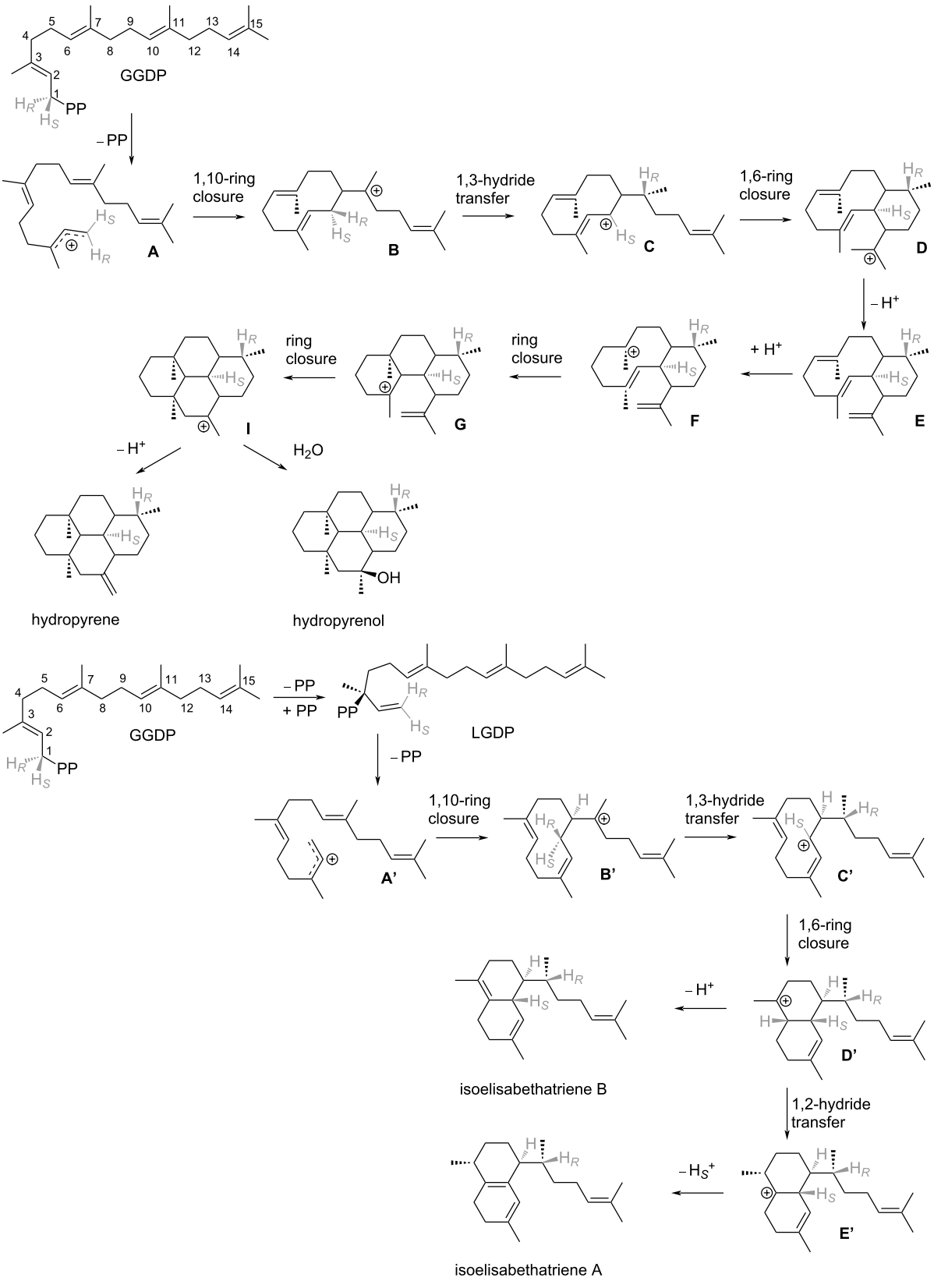


Figure 1: Summary of yields of HP and IE products in hydropyrene synthase.



Scheme 1: Proposed mechanism for HP and IE routes.

free energy profile is presented in Figure 2. Here, we modeled the transformations **A**→**I** (HP) and **A'**→**E'** (IE). The gas-phase calculations commenced with geranylgeranyl cation (**A** and **A'**) in a fully extended form.

HP pathway

The HP gas-phase pathway commences with a C1–C10 cyclization, which yields cation **B**, which is more stable than **A** by −11.4 kcal/mol. A subsequent 1,3-hydride transfer results in an allyl cation (**C**), which is −18.5 kcal/mol more stable than **A**. The barrier for the 1,3-hydride transfer is 16.2 kcal/mol for **B**→**C**. Subsequently, the double bond on C14–C15 reacts with the cationic charge on C1 to form intermediate **D**, which is slightly less stable than **C** (−17.3 kcal/mol) In the enzyme environment intermediate **D** deprotonates to form intermediate **E**, while intermediate **E** is re-protonated to form intermediate **F**,

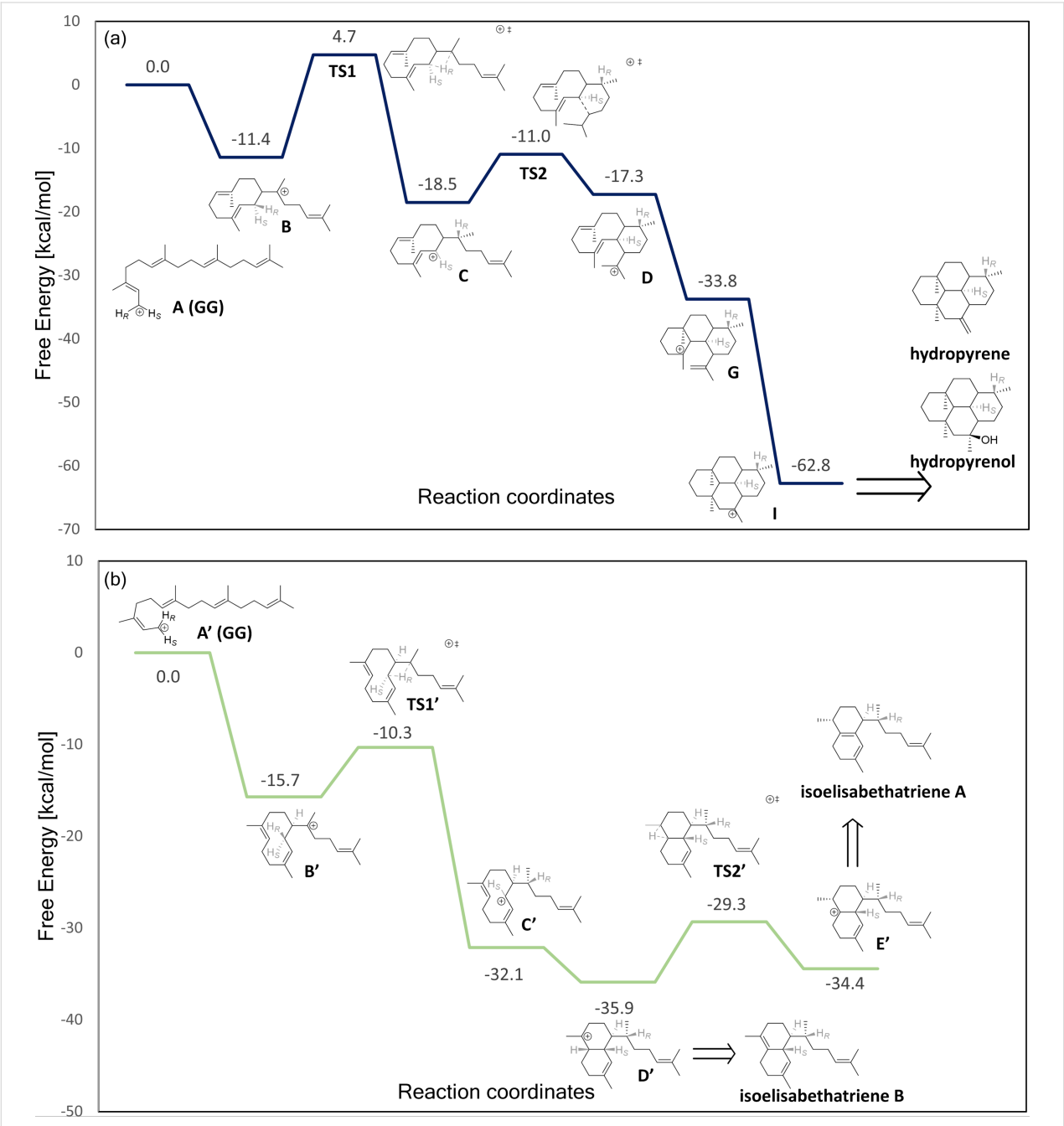


Figure 2: Free energy profile of hydropyrene cation (a), and IE cation (b) formation in the gas phase. The free energy of cation **A** and **A'** is set to zero. Bonds breaking/forming in the transition states are marked by dotted lines. All calculations were performed in the gas phase at the M06-2X/6-31G+(d,p) level of theory.

which immediately transforms to intermediate **G** (i.e., **F** is not stable). **G** is significantly more stable than **D** (by −16.5 kcal/mol). **G** then transforms to the very stable 4-ring intermediate **I** without any free energy barrier. The deprotonation and re-protonation steps are not included in our calculations. The overall exergonicity of this process which transforms four π-bonds to σ-bonds, with accompanying gains in intramolecular dispersion interactions, is −62.8 kcal/mol.

IE pathway

As described above both pathways commence with a C1–C10 cyclization. However, in the IE pathway a preliminary isomerization step occurs via rotation around the C2–C3 bond, transforming from the *trans* to the *cis* form. In the enzyme this process occurs with the help of a pyrophosphate group. The C1–C10 cyclization yields cation **B'**, which is more stable than **A'** by −15.7 kcal/mol. A subsequent 1,3-hydride transfer results in an allyl cation (**C'**, −32.1 kcal/mol relative to **A'**), with a barrier of 5.4 kcal/mol. Cation **C'** collapses into **D'** via a barrierless 1,6-ring closure (ΔG_r of −35.9 kcal/mol relative to **A'**). **D'** can be deprotonated to yield IE B or conversely may undergo a 1,2-hydride transfer, forming carbocation **E'** (ΔG_r of −34.4 kcal/mol relative to **A'**). This transformation has a $\Delta G^{\ddagger}$ of 6.6 kcal/mol. Cation **E'** may then be deprotonated to form IE A. Overall, the exergonicity for the formation of IEs A/B from carbocation **A'** is significant, due to the exchange of two π-bonds for σ-bonds, as well as gain in dispersion interactions on folding of the extended geranylgeranyl cation.

Discussion

Although the current calculations were performed in the gas phase without inclusion of the enzyme environment, we may still generate some hypotheses regarding the enzymatic process. First, considering the similar free energy of IEs A and B and the small kinetic barrier separating them, these isomers may exist in equilibrium in the enzyme active site. The relative amount of IEs A and B may then be determined by their proximity to an active site base, as has been observed in other cases, like taxadiene synthase [26]. Second, the chemical changes occurring up to intermediates **C** and **C'** in both routes are similar and diverge in the transformation of **C**→**D** and **C'**→**D'**. Here, we suggest that the ring closure in the IE pathway (**C'**→**D'**) occurs immediately due to the *cis* conformation and the short distance between the cation and the double bond (≈2 Å), which could be due to the active site substrate fold. In intermediate **C**, the *trans* conformation and possibly slightly different substrate fold, results in an alternative reaction route being followed (Figure 3). Lastly, considering the huge thermodynamic preference for the HP pathway, it is unlikely that a thermodynamic equilibrium exists between this pathway and the IE pathway. Rather, an equilibrium may exist between GGDP and LGDP, but once cyclization commences, the reactions will proceed until completion along their respective pathways. Hence, the difference in the product profile in WT and enzyme variants may be largely due to different folding of the initial substrate. Future in-enzyme studies can shed light on the preferred folding of GGDP inside the WT and variant enzymes and potentially the roles specific active site residues play during catalysis [23,24,27].

Conclusion

In the current work we performed gas-phase quantum chemistry calculations for the competing reaction pathways leading to IEs A and B as well as hydropyrene and hydropyrenol. In the former two reactions there is an exchange of two π-bonds for σ-bonds, resulting in a total exergonicity of −34.4 and −35.9 kcal/mol, respectively. In the latter two reactions which replace four π-bonds for σ-bonds and share a common final carbocation, the exergonicity is −62.8 kcal/mol. These values reflect the energetics of exchanging π-bonds for σ-bonds. In

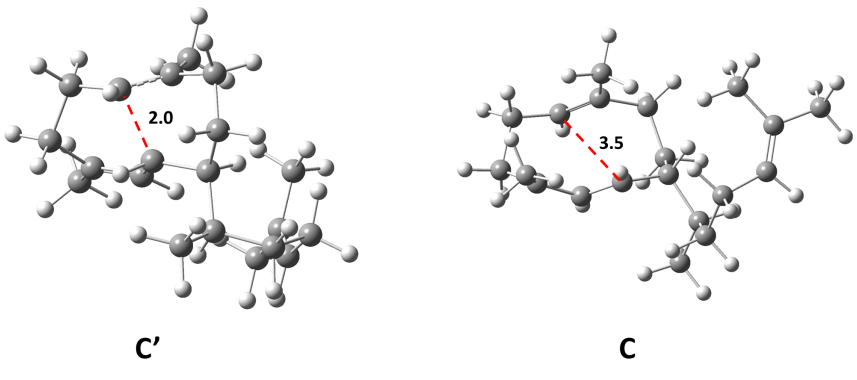


Figure 3: Structures of intermediates **C'** and **C**. The distance between the double bond and the cation in intermediate **C'** is 2.0 Å while in **C** the distance is 3.5 Å.

spite of the current calculations being performed in the gas phase, we may still generate some propositions regarding the enzymatic process. First, considering the similar free energy of IEs A and B and the low barrier between them, IEs A and B may exist in equilibrium in the enzyme active site. The proximity to an active site base may then determine the relative amount of IEs A and B. Second, it is unlikely that a thermodynamic equilibrium exists between the HP and IE pathways, due to the significant free energy barriers required for reverse barriers in the enzyme. Rather, an equilibrium may exist between GGDP and LGDP.

Experimental

Dynamics and Monte Carlo simulations

We generated conformers using simulated annealing (SA) molecular dynamics followed by SA Monte Carlo simulation using CHARMM [28]. Force field parameters were generated using CGenFF [29] and an in-house code which modifies parameters for cations from existing parameters for neutral molecules. For each intermediate in the reaction mechanism (path HP and IE) we created 100 conformers which were subsequently clustered (a cluster width of 1.0 Å was used). For each unique conformer we performed QM calculations and then chose the lowest energy conformer as representative of each carbocation intermediate in the mechanism. Hence, each intermediate state is represented by a single conformer, which is the lowest energy conformer found.

Quantum chemistry calculations

Optimizations and subsequent frequency calculations were performed using the Gaussian 16 program [30] at the M06-2X/6-31G+(d,p) level of theory [31,32]. This combination has been employed previously to TPS reactions [17,18,20,24,25,33-36]. The Gibbs free energies were calculated within the standard harmonic approximation and at a pressure of 1 bar and temperature of 298 K. We note that for some reaction steps, a transition state could not be located.

Supporting Information

Supporting Information File 1

All coordinate files for intermediates and TS structures.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-97-S1.rar]

Funding

MR and TB gratefully acknowledge funding by the Werner Siemens foundation for establishing the field of Synthetic Biotechnology at TUM. This work was supported by a grant from the Israel Science Foundation (Grant 1683/18) (DTM) and by a grant from the German-Israeli Foundation for Scientific Research and Development (Grant I-85-302.14-2018) (DTM, TB, BL).

ORCID® iDs

Shani Zev - https://orcid.org/0000-0002-3731-5968
Marion Ringel - https://orcid.org/0000-0003-0162-0765
Ronja Driller - https://orcid.org/0000-0001-8834-9087
Bernhard Loll - https://orcid.org/0000-0001-7928-4488
Thomas Brück - https://orcid.org/0000-0002-2113-6957
Dan T. Major - https://orcid.org/0000-0002-9231-0676

Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2022.37.v1

References

1. Croteau, R. *Chem. Rev.* **1987,** *87,* 929–954. doi:10.1021/cr00081a004
2. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
3. Cane, D. E. *Chem. Rev.* **1990,** *90,* 1089–1103. doi:10.1021/cr00105a002
4. Movahhed, S.; Westphal, J.; Kempa, A.; Schumacher, C. E.; Sperlich, J.; Neudörfl, J.-M.; Teusch, N.; Hochgürtel, M.; Schmalz, H.-G. *Chem. – Eur. J.* **2021,** *27,* 11574–11579. doi:10.1002/chem.202101863
5. Kohl, A. C.; Kerr, R. G. *Mar. Drugs* **2003,** *1,* 54–65. doi:10.3390/md101054
6. Ringel, M.; Reinbold, M.; Hirte, M.; Haack, M.; Huber, C.; Eisenreich, W.; Masri, M. A.; Schenk, G.; Guddat, L. W.; Loll, B.; Kerr, R.; Garbe, D.; Brück, T. *Green Chem.* **2020,** *22,* 6033–6046. doi:10.1039/d0gc01697g
7. Rinkel, J.; Rabe, P.; Chen, X.; Köllner, T. G.; Chen, F.; Dickschat, J. S. *Chem. – Eur. J.* **2017,** *23,* 10501–10505. doi:10.1002/chem.201702704
8. Berrué, F.; McCulloch, M. W. B.; Kerr, R. G. *Bioorg. Med. Chem.* **2011,** *19,* 6702–6719. doi:10.1016/j.bmc.2011.06.083
9. Mayer, A. M. S.; Jacobson, P. B.; Fenical, W.; Jacobs, R. S.; Glaser, K. B. *Life Sci.* **1998,** *62,* PL401–PL407. doi:10.1016/s0024-3205(98)00229-x
10. Hong, Y. J.; Tantillo, D. J. *Nat. Chem.* **2009,** *1,* 384–389. doi:10.1038/nchem.287
11. Tantillo, D. J. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020,** *10,* e1453. doi:10.1002/wcms.1453
12. Tantillo, D. J. *Angew. Chem., Int. Ed.* **2017,** *56,* 10040–10045. doi:10.1002/anie.201702363
13. Smentek, L.; Hess, B. A., Jr. *J. Am. Chem. Soc.* **2010,** *132,* 17111–17117. doi:10.1021/ja1039133
14. Hess, B. A., Jr.; Smentek, L. *Angew. Chem., Int. Ed.* **2013,** *52,* 11029–11033. doi:10.1002/anie.201302886
15. Tantillo, D. J. *Chem. Soc. Rev.* **2018,** *47,* 7845–7850. doi:10.1039/c8cs00298c
16. Tantillo, D. J. *Nat. Prod. Rep.* **2011,** *28,* 1035–1053. doi:10.1039/c1np00006c
17. Dixit, M.; Weitman, M.; Gao, J.; Major, D. T. *ACS Catal.* **2017,** *7,* 812–818. doi:10.1021/acscatal.6b02584

18. Major, D. T. *ACS Catal.* **2017,** *7,* 5461–5465. doi:10.1021/acscatal.7b01328
19. Weitman, M.; Major, D. T. *J. Am. Chem. Soc.* **2010,** *132,* 6349–6360. doi:10.1021/ja910134x
20. Ansbacher, T.; Freud, Y.; Major, D. T. *Biochemistry* **2018,** *57,* 3773–3779. doi:10.1021/acs.biochem.8b00452
21. Hong, Y. J.; Tantillo, D. J. *J. Am. Chem. Soc.* **2009,** *131,* 7999–8015. doi:10.1021/ja9005332
22. Hong, Y. J.; Tantillo, D. J. *J. Am. Chem. Soc.* **2014,** *136,* 2450–2463. doi:10.1021/ja4106489
23. 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
24. Major, D. T.; Freud, Y.; Weitman, M. *Curr. Opin. Chem. Biol.* **2014,** *21,* 25–33. doi:10.1016/j.cbpa.2014.03.010
25. Raz, K.; Driller, R.; Dimos, N.; Ringel, M.; Brück, T.; Loll, B.; Major, D. T. *J. Am. Chem. Soc.* **2020,** *142,* 21562–21574. doi:10.1021/jacs.0c11348
26. Schrepfer, P.; Buettner, A.; Goerner, C.; Hertel, M.; van Rijn, J.; Wallrapp, F.; Eisenreich, W.; Sieber, V.; Kourist, R.; Brück, T. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* E958–E967. doi:10.1073/pnas.1519680113
27. 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
28. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983,** *4,* 187–217. doi:10.1002/jcc.540040211
29. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. *J. Comput. Chem.* **2010,** *31,* 671–690. doi:10.1002/jcc.21367
30. *Gaussian 16,* Rev. A.03; Gaussian, Inc.: Wallingford, CT, 2016.
31. Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008,** *120,* 215–241. doi:10.1007/s00214-007-0310-x
32. Nagy, B.; Jensen, F. *Rev. Comput. Chem.* **2017,** *30,* 93–149. doi:10.1002/9781119356059.ch3
33. Das, S.; Dixit, M.; Major, D. T. *Bioorg. Med. Chem.* **2016,** *24,* 4867–4870. doi:10.1016/j.bmc.2016.07.002
34. Pahima, E.; Zhang, Q.; Tiefenbacher, K.; Major, D. T. *J. Am. Chem. Soc.* **2019,** *141,* 6234–6246. doi:10.1021/jacs.8b13411
35. Levi, S.; Zhang, Q.; Major, D. T. *ACS Catal.* **2020,** *10,* 6843–6853. doi:10.1021/acscatal.0c00278
36. Zev, S.; Gupta, P. K.; Pahima, E.; Major, D. T. *J. Chem. Theory Comput.* **2022,** *18,* 167–178. doi:10.1021/acs.jctc.1c00746

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Isolation and biosynthesis of daturamycins from *Streptomyces* sp. KIB-H1544

Yin Chen[‡1,2], Jinqiu Ren[‡1], Ruimin Yang[1], Jie Li[1,2], Sheng-Xiong Huang[1] and Yijun Yan[*1]

Full Research Paper

Open Access

Address:
[1]State Key Laboratory of Phytochemistry and Plant Resources in West China, and CAS Center for Excellence in Molecular Plant Sciences, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China and [2]University of Chinese Academy of Sciences, Beijing 100049, China

Email:
Yijun Yan[*] - yanyijun@mail.kib.ac.cn

* Corresponding author ‡ Equal contributors

Keywords:
biosynthesis; diarylcyclopentenone; polyporic acid synthetase; *p*-terphenyl; *Streptomyces*

Beilstein J. Org. Chem. **2022,** *18,* 1009–1016.
https://doi.org/10.3762/bjoc.18.101

Received: 25 May 2022
Accepted: 25 July 2022
Published: 09 August 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

Two novel diarylcyclopentenones daturamycin A and B (**1** and **2**), and one new *p*-terphenyl daturamycin C (**3**), along with three known congeners (**4**–**6**), were isolated from a rhizosphere soil-derived *Streptomyces* sp. KIB-H1544. The structures of new compounds were elucidated via a joint use of spectroscopic analyses and single-crystal X-ray diffractions. Compounds **1** and **2** belong to a rare class of tricyclic 6/5/6 diarylcyclopentenones, and compounds **3**–**6** possess a C-18 tricyclic aromatic skeleton. The biosynthetic gene cluster of daturamycins was identified through gene knockout and biochemical characterization experiments and the biosynthetic pathway of daturamycins was proposed.

Introduction

Natural products containing a terphenyl skeleton exhibit a large number of structural diversity due to the differences of the center ring and the connection among rings. Structurally, most natural terphenyls are *p*-terphenyl derivatives consisting of a C-18 tricyclic or polycyclic C-18 aromatic skeleton. Diarylcyclopentenones, which possess a rare class of a tricyclic 6/5/6 system, could also belong to *p*-terphenyl derivatives from a biosynthetic perspective. To date, more than 230 natural products containing *p*-terphenyl have been unearthed in nature, among them many *p*-terphenyls were isolated from fungi, only a few were discovered from *Streptomyces* species [1-4]. Meanwhile, these types of compounds exhibit a broad spectrum of bioactivities, including antitumor [5,6], antibacterial [7,8], antioxidant [9-12], immunosuppressive [13], and antithrombotic activities [14].

The biosynthesis of the *p*-terphenyls has been studied since the 1960s [15,16]. Stable-isotope labeling experiments confirmed

that L-tyrosine or L-phenylalanine are involved in the biosynthesis of *p*-terphenyl as metabolic origin. The precursors undergoing deamination are converted to the corresponding α-keto acid, then a quinone intermediate arises by condensation between two molecules of α-keto acid. Structurally diverse *p*-terphenyls are formed from these key intermediates by several tailoring reactions such as cyclization, tautomerization, methylation, and glycosylation. A previous study has shown that the formation of 2,5-diarylcyclopentenone proceeds via the terphenylquinone atromentin (Figure 1), followed by oxidative ring contraction [17]. However, the details of cyclopentenone formation involving ring contraction have remained unclear.

In this study, two novel diarylcyclopentenones daturamycins A amd B (**1** and **2**), one new *p*-terphenyl daturamycin C (**3**), and three known *p*-terphenyl derivatives (**4**–**6**) (Figure 1) were isolated from the fermentation extract of *Streptomyces* sp. KIB-H1544. Structurally, daturamycins A and B are uncommon diarylcyclopentenone compounds that consist of a tricyclic 6/5/6 system instead of a C-18 tricyclic skeleton in other compounds. To explore the biosynthesis of daturamycins in *S*. sp. KIB-H1544, especially the formation mechanism of the tricyclic 6/5/6 scaffold, the biosynthetic gene cluster of daturamycins was found and confirmed by gene knockout experiments. We also characterized the deduced peptide synthetase DatA, which catalyzes the Claisen–Dieckmann condensation of phenyl-pyruvic acid (**7**) to generate the key intermediate polyporic acid (**8**). Finally, we proposed a biosynthetic pathway for daturamycins.

Results and Discussion

Daturamycin A (**1**), a yellow powder, possessed a molecular formula of $C_{19}H_{16}O_5$ with 12 degrees of unsaturation based on the HRMS–ESI data (*m/z* 347.0893 $[M + Na]^+$, calcd for $C_{19}H_{16}O_5Na^+$, 347.0890) (Figure S1, Supporting Information File 1). Comprehensive analysis of the ^1H and ^{13}C NMR data (Table 1, Figures S3 and S4, Supporting Information File 1) and HSQC data (Figure S5, Supporting Information File 1) demonstrated the presence of one methyl, eleven methines (including ten aromatic ones), and seven quaternary carbons (including two carbonyl carbons). The structural information of **1** was further defined according to HMBC data (Figure S6, Supporting Information File 1). The key HMBC signals (H-2' to C-1', C-4', C-6'; H-3' to C-1', C-5'; H-4' to C-2', C-6' ; H-5' to C-1', C-3'; H-2'' to C-1'', C-4'', C-6''; H-3'' to C-1'', C-5''; H-4'' to C-2'', C-6'', and H-5'' to C-1'', C-3'') of compound **1** indicated the presence of two mono-substituted aromatic rings. The cross-peaks from H-4 to C-1, C-2, C-3, C-5, and C-6, from 6-OCH_3 to C-6 established a central five-membered carbon ring. Meanwhile, the correlations from H-4 to C-1', C-1'', from H-2' and H-6' to C-5, and from H-2'' and H-6'' to C-3 showed that two benzene rings were conjugated to the central ring at the C-3 and C-5 positions (Figure 2). The NMR data indicated that compound **1** shares a similar skeleton to the known compound (±)-tylopilusin B [18]. The additional signals suggested that three methines (δ_C/δ_H 59.2/4.49, CH-4; 129.0/7.27, CH-4'; 130.5/7.38, CH-4'') in compound **1** were hydroxylated in (±)-tylopilusin B. Furthermore, the absolute configuration of C-4 and C-5 in compound **1** was also confirmed as 4*R* and 5*S* by X-ray crystallography (Figure 2).

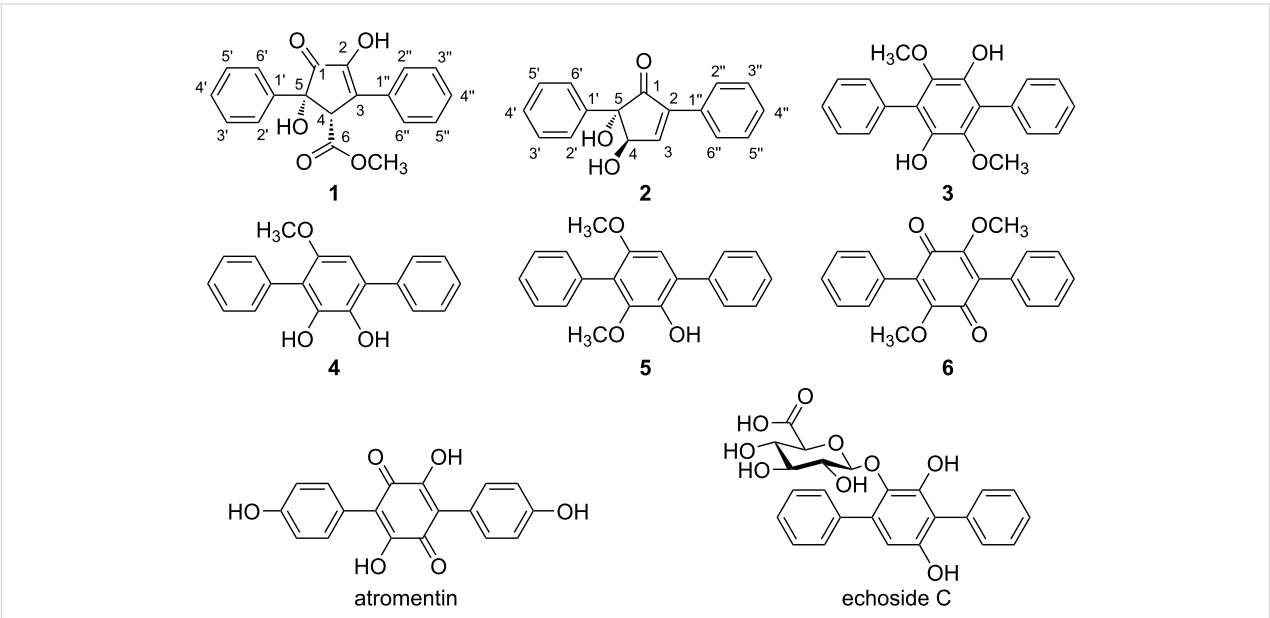


Figure 1: Structures of compounds **1**–**6**, atromentin, and echoside C.

Table 1: ^{1}H (600 MHz) and ^{13}C (150 MHz) NMR data of compounds **1** and **2**.

	1 (in CD_3OD)		**2** (in DMSO-d_6)	
position	δ_C	δ_H, mult. (*J* in Hz)	δ_C	δ_H, mult. (*J* in Hz)
1	201.2, C		204.6, C	
2	151.4, C		140.8, C	
3	135.1, C		157.7, CH	7.98, d (1.8)
4	59.2, CH	4.49, s	77.5, CH	4.84, d (4.3)
5	79.4, C		85.5, C	
6	172.8, C			
1’	143.5, C		139.6, C	
2’/6’	126.2, CH	7.44, d (7.7)	126.9, CH	7.28, m
3’/5’	129.5, CH	7.33, t (7.5)	127.3, CH	7.28, m
4’	129.0, CH	7.27, t (7.3)	126.7, CH	7.21, t (6.9)
1’’	134.7, C		130.7, C	
2’’/6’’	128.7, CH	7.95, d (7.7)	127.1, CH	7.82, d (7.3)
3’’/5’’	129.8, CH	7.44, t (7.5)	128.6, CH	7.45, t (7.3)
4’’	130.5, CH	7.38, t (7.3)	129.0, CH	7.41, t (7.3)
6-OCH_3	52.9, CH_3	3.71, s		
4-OH				5.60, d (6.3)
5-OH				6.29, s

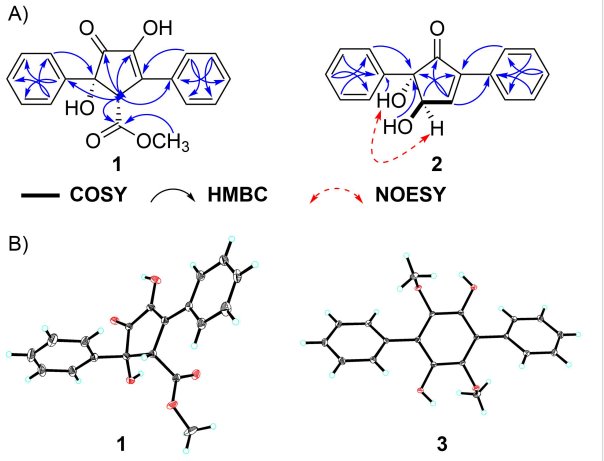


Figure 2: (A) Key 2D NMR correlations of compounds **1** and **2**. (B) X-ray crystal structure of compounds **1** and **3**.

(±)-Daturamycin B (**2**) was isolated as a white powder, and its molecular formula was determined as $C_{17}H_{14}O_3$ by HRMS–ESI data (*m/z* 289.0833 $[M + Na]^+$, calcd for $C_{19}H_{16}O_5Na^+$, 289.0835) (Figure S7, Supporting Information File 1), and the unsaturation was 11 degree. It could be deduced that compound **2** also possesses a similar core to compound **1**, based on the molecular formula and NMR data (Table 1, Figures S8–S13, Supporting Information File 1). This could be confirmed by the key COSY correlation of H-3/H-4 and the HMBC correlations (H-3 to C-1, C-4, C-1’’; H-4 to C-1, C-2, C-1’; 4-OH to C-3, C-5; 5-OH to C-1, C-4, C-1’; H-2’ to C-5; H-3’ to C-1’; H-4’ to C-2’, C-6’; H-2’’ to C-2, C-3, C-4’’; H-3’’ to C-1’’; H-4’’ to C-2’’). The NOESY correlation (Figure 2) between H-4 and 5-OH suggested that the relative configurations of C-4 and C-5 were *trans*. Therefore, the structure of compound **2** has been determined, as shown in Figure 1. However, the absolute configuration of compound **2** remained unsolved.

Daturamycin C (**3**) was obtained as a brown powder, and the ^{1}H and ^{13}C NMR spectra (Figures S14 and S15, Supporting Information File 1) indicated that compound **3** showed strong similarity with 1,4-diphenyl-2,3,5,6-tetramethoxybenzene [19], with two methyl signals were absent in the NMR data of **3**. Furthermore, the structure of **3** could be confirmed, as shown in Figure 1, based on X-ray crystallography (Figure 2).

The known congeners were determined as terferol (**4**) [20], BTH-II0204-207: D (**5**) [21], and betulinan A (**6**) [22] (Figure 1) by comparing their ^{1}H and ^{13}C NMR spectroscopic data (Figures S16–S21, Supporting Information File 1) and specific rotation values with those in the literatures.

Diarylcyclopentenones, characteristic constituents of mushrooms [23], were rarely discovered in *Streptomyces* species. These components exhibit redox activity and are involved in reducing ferric (Fe^{3+}) in the Fenton-based biological decompo-

sition of lignocellulose [24,25]. The biosynthetic pathway of *p*-terphenyl was previously identified in *Paxillus involutus*, and atromentin was supposed to be a metabolic precursor of diarylcyclopentenone [17]. However, the mechanism of cyclopentenone moiety formation has remained unclear. Two pathways may be involved in the biosynthesis of diarylcyclopentenone (Figure 3B): (path a) two molecules of phenylpyruvic acid (**7**) undergo direct condensation to form a five-membered ring intermediate or (path b) polyporic acid (**8**) undergoes oxidative ring contraction or conversion to generate the cyclopentenone skeleton.

To verify which pathway was responsible for the biosynthesis of daturamycins in *S.* sp. KIB-H1544, we conducted the following experiments. Firstly, the genome of *S.* sp. KIB-H1544 was sequenced, and bioinformatic analysis yields the three-genes cluster (NCBI accession number: ON973849) encoded by *datA* (peptide synthetase), *datB* (NAD-dependent dehydratase), and *datC* (limonene-1,2-epoxide hydrolase), which share high sequence similarity with the echoside (Figure 1) biosynthetic gene cluster [26] (EchA, 69.2% identity to DatA; EchB, 78.9% identity to DatB; EchC, 66.9% identity to DatC) (Figure 3A). Additionally, one bilirubin oxidase (DatD) and NADPH-de-

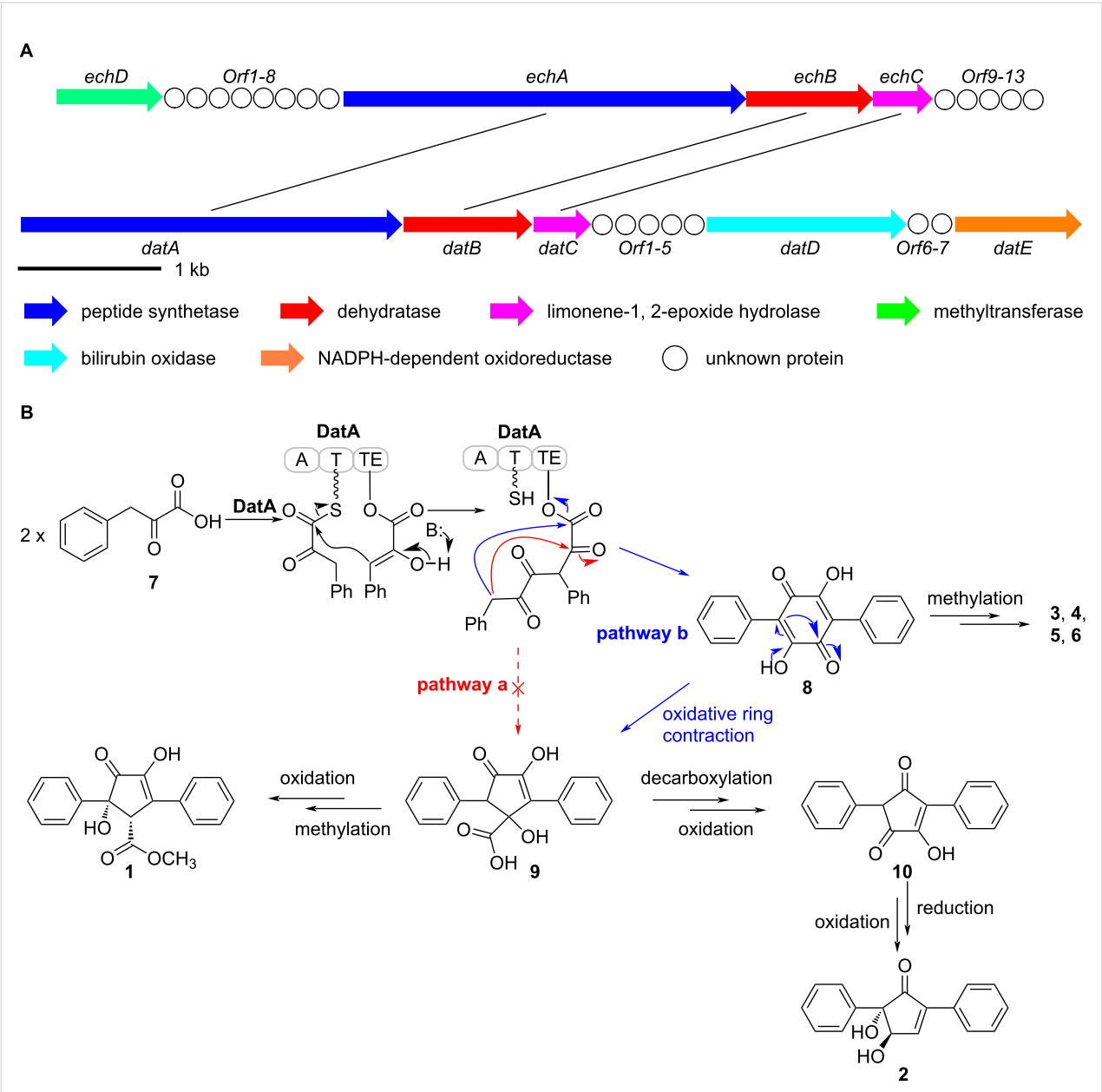


Figure 3: (A) The biosynthetic gene cluster of daturamycins. (B) Proposed biosynthetic pathway of daturamycins.

pendent oxidoreductase (DatE) are found downstream of the three-genes cluster (Table S1, Supporting Information File 1). Subsequently, we set out to knockout *datA* gene by λ-RED-mediated recombination in *S.* sp. KIB-H1544. Culture and fermentation results of mutant *S*. sp. KIB-H1544-Δ*datA* indicated that the primary production of compounds **1**–**6** in the mutant was eliminated (Figure 4A). This result suggested that the *datA* gene is responsible for the biosynthesis of daturamycins. Lastly, in order to verify the cyclopentenone rings of daturamycins A and B are generated via direct condensation of phenylpyruvic acid (**7**) catalyzed by DatA, the DatA protein was expressed and purified (Figure 4B). Incubating DatA with substrate **7** and ATP resulted in a new product **8**, which was confirmed by HRMS–ESI data (*m/z* 291.0663 [M − H]$^-$, calcd for ([$C_{18}H_{12}O_4$] − H)$^-$, 291.0663) (Figure S22, Supporting Information File 1). The product was not present in the control reaction with boiled DatA protein (Figure 4B). Furthermore, we could not detect any other intermediates in the reaction mixture through MS analysis (Figure S23, Supporting Information File 1). These results suggested that DatA could not catalyze the formation of the cyclopentenone ring. It has the same function as EchA, which only catalyzes the Claisen–Dieckmann condensation of phenylpyruvic acid (**7**) to generate polyporic acid (**8**) [26].

Based on the gene knockout and biochemical characterization experiments, we proposed a possible biosynthetic pathway for daturamycins in *S.* sp. KIB-H1544 (Figure 3B). First, two molecules of phenylpyruvic acid (**7**), which is deaminated from L-phenylalanine, are condensed into the six-membered polyporic acid (**8**) catalyzed by DatA. Then, **8** converts to compounds **3**, **4**, **5**, and **6** through methylation and oxidation. On the other hand, polyporic acid (**8**) probably undergoes rearrangement, decarboxylation, methylation reaction, and further modification to generate compounds **1** and **2**.

Conclusion

S. sp. KIB-H1544 was isolated from the rhizosphere soil of *Datura stramonium* L. Two novel diarylcyclopentenones daturamycins A and B and one new *p*-terphenyl daturamycin C, were isolated and identified. The new diarylcyclopentenones possess a rare class of a tricyclic 6/5/6 system. Based on the bioinformatics analysis, the daturamycins biosynthetic gene cluster has been identified and confirmed by the inactivation of the structural gene *datA*. The following biochemical characterization indicated that the DatA is a polyporic acid synthetase, which helped to propose a possible biosynthetic pathway for daturamycins in S. sp. KIB-H1544 (Figure 3B). Finally, the discovery of daturamycins, identification, and characterization of *dat* biosynthetic gene cluster allowed us to explore their biosynthesis, especially the mechanism of cyclopentenone formation in the diarylcyclopentenones.

Experimental

General experimental procedures

IR spectra were measured on a Bruker Tensor 27 FTIR spectrometer using KBr disks. NMR spectra were performed with a Bruker AV600 MHz spectrometer with TMS as an internal standard. HRMS–ESI data were obtained through an Agilent 1290 UPLC/6540 Q-TOF mass instrument. Column chromatography (CC) was performed using silica gel (30–400 mesh, Qingdao Marine Chemical Inc., China), Sephadex LH-20 (25–100 μm, Pharmacia Biotech Ltd., Sweden), and MCI gel (75–150 μm, Mitsubishi Chemical Corporation, Tokyo, Japan). Semipreparative HPLC was conducted on a HITACHI Chromaster system equipped with a DAD detector, a YMC-Triart

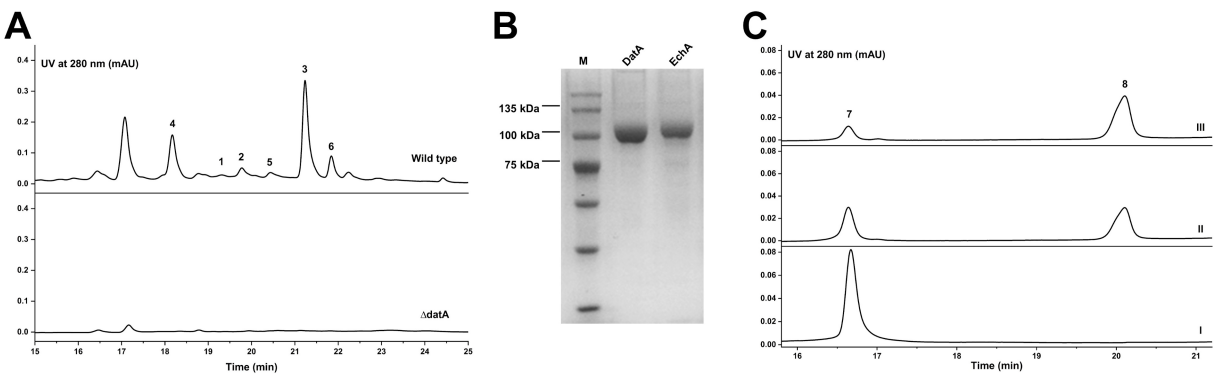


Figure 4: (A) HPLC analysis of the fermentation extracts of mutant *S*. sp. KIB-H1544-Δ*datA*. (B) SDS-PAGE analysis of purified DatA (calculated molecular mass 101.8 kDa) and EchA (calculated molecular mass 105.4 kDa) protein. (C) HPLC analysis of *in vitro* enzyme reactions. Panel I, phenylpyruvic acid with boiled DatA; panel II, phenylpyruvic acid with EchA; panel III, phenylpyruvic acid with DatA.

C18 column (250 × 10 mm, i.d. 5 μm), and a flow rate of 3.0 mL/min. LC–MS was performed using an Agilent 1200 series HPLC system coupled to an Agilent q TOF 6540 mass spectrometer with a YMC-Triart C18 column (250 × 4.6 mm, i.d. 5 μm) and a flow rate of 1.0 mL/min.

Extraction and isolation

S. sp. KIB-H1544, isolated from the rhizosphere soil of *Datura stramonium* L., was 99.9% similar to the 16S rRNA of *Streptomyces roseofulvus* NBRC15816. The strain was cultured in 20 L M15 medium: 30.0 g/L glucose, 1.0 g/L peptone, 5.0 g/L beef paste, 2.5 g/L $CaCO_3$, 5.0 g/L NaCl, 1.0 mL/L trace elements solution (1.0 g/L $FeSO_4{\cdot}7H_2O$, 0.45 g/L $CuSO_4{\cdot}5H_2O$, 1.0 g/L $ZnSO_4{\cdot}7H_2O$, 0.1 g/L $MnSO_4{\cdot}4H_2O$, 0.1 g/L K_2MoO_4). After extraction and concentration, the extract was purified by silica gel chromatography and semipreparative HPLC to yield daturamycin A (**1**) (4.5 mg), daturamycin B (**2**) (1.8 mg), daturamycin C (**3**) (8.7 mg), terferol (**4**) (2.4 mg), BTH-II0204-207: D (**5**) (2.1 mg), betulinan A (**6**) (6.2 mg).

Identification of new compounds

Daturamycin A (**1**): pale yellow solid; IR (KBr) ν_{max}: 3468, 3287, 1731, 1702, 1631, 1398 and 1202 cm^{-1}; UV (MeOH) λ_{max} (log ε): 314 (0.48), 217 (0.28), 202 (0.48); ^1H and ^{13}C NMR, see Table 1; HRMS–ESI (*m*/*z*): $[M + Na]^+$ calcd for $C_{19}H_{16}O_5Na^+$, 347.0890; found, 347.0893.

Daturamycin B (**2**): white powder; ^1H and ^{13}C NMR, see Table 1; HRMS–ESI (*m*/*z*): $[M + Na]^+$ calcd for $C_{19}H_{16}O_5Na^+$, 289.0835; found, 289.0833.

Daturamycin C (**3**): brown powder; ^1H NMR (600 MHz, $CDCl_3$) δ 7.57 (d, *J* = 7.2 Hz, 4H), 7.48 (t, *J* = 7.6 Hz, 4H), 7.39 (t, *J* = 7.4 Hz, 2H), 5.52 (s, 2H), 3.38 (s, 6H); ^{13}C NMR (150 MHz, $CDCl_3$) δ 140.9, 139.6, 132.8, 130.6, 130.3, 128.3, 128.0, 127.7, 120.7.

X-ray crystal data of compound 1: $C_{19}H_{16}O_5$, *M* = 324.32, *a* = 9.8765(14) Å, *b* = 8.9786(13) Å, *c* = 18.315(3) Å, α = 90°, β = 92.119(3)°, γ = 90°, *V* = 1623.0(4) $Å^3$, *T* = 100(2) K, space group *P*21/*n*, *Z* = 4, μ(Mo Kα) = 0.096 mm^{-1}, 16091 reflections measured, 4037 independent reflections (Rint = 0.0728). The final R1 values were 0.0526 (I > 2σ(I)). The final wR(F2) values were 0.1003 (I > 2σ(I)). The final R1 values were 0.1061 (all data). The final wR(F2) values were 0.1183 (all data). The goodness of fit on F2 was 1.003.

X-ray crystal data of compound 3: $C_{20}H_{18}O_4$, *M* = 322.34, *a* = 5.9209(8) Å, *b* = 20.882(3) Å, *c* = 7.0856(10) Å, α = 90°, β = 113.939(2)°, γ = 90°, *V* = 800.70(19) $Å^3$, *T* = 100(2) K, space group *P*21/*c*, *Z* = 2, μ(Mo Kα) = 0.093 mm^{-1}, 8836 reflections measured, 2398 independent reflections (Rint = 0.0250). The final R1 values were 0.0423 (I > 2σ(I)). The final wR(F2) values were 0.1059 (I > 2σ(I)). The final R1 values were 0.0495 (all data). The final wR(F2) values were 0.1101 (all data). The goodness of fit on F2 was 1.051.

Genomic library construction and screening

The genome of *S*. sp. KIB-H1544 was sequenced through the Illumina Genome Analyzer (Illumina, San Diego, CA) by BGI (BGI-Shenzhen, China). According to standard procedures, the genomic DNA was digested with *Mbo*I, and the 30–42 kb DNA fragments were isolated and ligated to cosmid pSuperCos I. MaxPlax Lambda packaging extracts were used for packaging. About 2,000 *E. coli* clones were picked and stored in 20 96-well microplates at −80 °C. The positive clone, named pSC-21A4, was verified by PCR. And primers used for screening were listed in Table S3 (Supporting Information File 1).

Gene inactivation

Primers designed for inactivation of *datA* gene are listed in Table S3 (Supporting Information File 1). The deletion mutant was constructed by λ-RED-mediated PCR targeting mutagenesis method [27]. The positive cosmid pSC-21A4 was transformed into *E. coli* BW25113/pIJ790 for gene inactivation. Then, the PCR fragment was introduced via electro-transformation into *E. coli* BW25113/pIJ790-pSC-21A4, in which the target gene *datA* would be replaced with the cassette. This recombinant cosmid was transformed into *E. coli* ET12567/pUZ8002, and suffered from intergeneric conjugation with *S*. sp. KIB-H1544 wild strain. *E. coli-Streptomyces* conjugation was performed on MS solid medium freshly supplemented with 10 mM $MgCl_2$. Double crossover mutants were selected based on the Kan-Apr and then confirmed by PCR using primers. Finally, the mutant strain *S*. sp. KIB-H1544-Δ*datA* was generated (Table S2, Supporting Information File 1).

Expression and purification of recombinant DatA and EchA protein

The DNA fragment containing *datA* was amplified by the primer pair Duet-DatA-F/Duet-DatA-R (Table S3, Supporting Information File 1), then cleaved by *BamH* I and *Sac* I, and inserted into the corresponding site of pETDuet-*sfp* to produce plasmid pETDuet-*sfp-datA*. As a positive control, the DNA fragment containing *echA* gene (GenBank: KJ156360.1) was synthesized by GENEWIZ company, plasmid pETDuet-*sfp-echA* was constructed using the same method as DatA expression vector construction. The recombinant plasmids were then transformed into *E. coli* BL21 (DE3), respectively. pETDuet-*sfp-datA* and pETDuet-*sfp-echA* were grown in Luria Bertani (LB) supplemented with 100 μg/mL ampicillin at 37 °C until

OD_{600} reached 0.6. IPTG (0.2 mM) was added to the final concentration of 0.4 mM and incubated at 16 °C overnight. The cells were centrifuged at 4000 rpm for 25 min and resuscitated with lysis buffer (buffer A: 50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 8.0). After ultrasonic cell crushing, the cells were centrifuged at 24,000 rpm for 60 min to remove cell fragments. The supernatant was filtered through 0.22 μm of the filter membrane and then loaded into the nickel column of the rebalanced lysate (HisTraqTM FF, GE Healthcare). The eluent was removed with buffer B (50 mM Tris-HCl, 300 mM NaCl, 250 mM imidazole, pH 8.0) at a flow rate of 2 mL/min, starting from the fourth tube of the collection tube and ending at the sixth tube. The outgoing protein was collected and poured into a precooled filter column (Ultracel series 10 KDa; GE Healthcare). Then, the collected DatA (or EchA) protein was concentrated and buffer-exchanged into storage buffer (50 mM NaH_2PO_4, 100 mM NaCl, pH 8.0, containing 10% glycerin). The proteins were tested by SDS-PAGE, frozen using liquid nitrogen, and kept at −80 °C.

In vitro reactions

DatA enzymatic activity was tested in a 100 μL reaction mixture containing 100 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 300 mM NaCl, 5 mM ATP, 0.5 mM phenylpyruvic acid and 5 μM DatA (or EchA), at 30 °C for 2 h. The reaction mixture was quenched by adding 200 μL methanal and then centrifuged at 12,000 rpm for 10 min. The supernatant was analyzed by HPLC.

Supporting Information

Supporting Information File 1

Spectroscopic data for compounds **1**–**3**, HRMS–ESI data for compound **8**, annotation of genes in the *dat* biosynthetic gene cluster, list of biological material, vectors, and primers used in this study.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-101-S1.pdf]

Funding

This research was financially supported by the National Natural Science Foundation of China (82073738, 31972852, and U1702285), the Yunnan Provincial Science and Technology Department (2019FJ007 and 2019FA034), Biological Resources Program (KFJ-BRP-009), and Research Program of Frontier Sciences (QYZDB-SSW-SMC051), Yunnan Thousand Talents Plan Young Talent and Youth Innovation Promotion Association (2018424), Chinese Academy of Sciences.

ORCID® iDs

Yijun Yan - https://orcid.org/0000-0001-7727-8084

References

1. Calì, V.; Spatafora, C.; Tringali, C. *Stud. Nat. Prod. Chem.* **2003,** *29,* 263–307. doi:10.1016/s1572-5995(03)80009-1
2. Liu, J.-K. *Chem. Rev.* **2006,** *106,* 2209–2223. doi:10.1021/cr050248c
3. Li, W.; Li, X.-B.; Lou, H.-X. *J. Asian Nat. Prod. Res.* **2018,** *20,* 1–13. doi:10.1080/10286020.2017.1381089
4. Zhou, G.; Zhu, T.; Che, Q.; Zhang, G.; Li, D. *Mar. Life Sci. Technol.* **2022,** *4,* 62–73. doi:10.1007/s42995-021-00117-8
5. Affleck, K.; Sidebottom, P. J.; Taylor, N. L.; Drake, C. S.; Todd, M.; Jowett, A.; Webb, G. *J. Antibiot.* **1999,** *52,* 89–95. doi:10.7164/antibiotics.52.89
6. Wang, D.; Wang, Y.; Ouyang, Y.; Fu, P.; Zhu, W. *J. Nat. Prod.* **2019,** *82,* 3504–3508. doi:10.1021/acs.jnatprod.9b00963
7. Belofsky, G. N.; Gloer, K. B.; Gloer, J. B.; Wicklow, D. T.; Dowd, P. F. *J. Nat. Prod.* **1998,** *61,* 1115–1119. doi:10.1021/np980188o
8. Tian, S.-Z.; Pu, X.; Luo, G.; Zhao, L.-X.; Xu, L.-H.; Li, W.-J.; Luo, Y. *J. Agric. Food Chem.* **2013,** *61,* 3006–3012. doi:10.1021/jf400718w
9. Quang, D. N.; Hashimoto, T.; Nukada, M.; Yamamoto, I.; Tanaka, M.; Asakawa, Y. *Planta Med.* **2003,** *69,* 1063–1066. doi:10.1055/s-2003-45159
10. Liu, J.-K.; Hu, L.; Dong, Z.-J.; Hu, Q. *Chem. Biodiversity* **2004,** *1,* 601–605. doi:10.1002/cbdv.200490050
11. Kuhnert, E.; Surup, F.; Herrmann, J.; Huch, V.; Müller, R.; Stadler, M. *Phytochemistry* **2015,** *118,* 68–73. doi:10.1016/j.phytochem.2015.08.004
12. Lee, I.-K.; Jung, J.-Y.; Kim, Y.-S.; Rhee, M. H.; Yun, B.-S. *Bioorg. Med. Chem.* **2009,** *17,* 4674–4680. doi:10.1016/j.bmc.2009.04.064
13. Kamigauchi, T.; Sakazaki, R.; Nagashima, K.; Kawamura, Y.; Yasuda, Y.; Matsushima, K.; Tani, H.; Takahashi, Y.; Ishii, K.; Suzuki, R.; Koizumi, K.; Nakai, H.; Ikenishi, Y.; Terui, Y. *J. Antibiot.* **1998,** *51,* 445–450. doi:10.7164/antibiotics.51.445
14. Khanna, J. M.; Malone, M. H.; Euler, K. L.; Brady, L. R. *J. Pharm. Sci.* **1965,** *54,* 1016–1020. doi:10.1002/jps.2600540714
15. Read, G.; Vining, L. C.; Haskins, R. H. *Can. J. Chem.* **1962,** *40,* 2357–2361. doi:10.1139/v62-359
16. von Massow, F. *Phytochemistry* **1977,** *16,* 1695–1698. doi:10.1016/0031-9422(71)85072-0
17. Braesel, J.; Götze, S.; Shah, F.; Heine, D.; Tauber, J.; Hertweck, C.; Tunlid, A.; Stallforth, P.; Hoffmeister, D. *Chem. Biol.* **2015,** *22,* 1325–1334. doi:10.1016/j.chembiol.2015.08.016
18. Fukuda, T.; Nagai, K.; Tomoda, H. *J. Nat. Prod.* **2012,** *75,* 2228–2231. doi:10.1021/np300428r
19. Morisaki, Y.; Tsuji, Y.; Chujo, Y. *Tetrahedron Lett.* **2014,** *55,* 1631–1634. doi:10.1016/j.tetlet.2014.01.093
20. Nakagawa, F.; Takahashi, S.; Naito, A.; Sato, S.; Iwabuchi, S.; Tamura, C. *J. Antibiot.* **1984,** *37,* 10–12. doi:10.7164/antibiotics.37.10
21. Biggins, J. B.; Liu, X.; Feng, Z.; Brady, S. F. *J. Am. Chem. Soc.* **2011,** *133,* 1638–1641. doi:10.1021/ja1087369
22. Lee, I.-K.; Yun, B.-S.; Cho, S.-M.; Kim, W.-G.; Kim, J.-P.; Ryoo, I.-J.; Koshino, H.; Yoo, I.-D. *J. Nat. Prod.* **1996,** *59,* 1090–1092. doi:10.1021/np960253z
23. Feling, R.; Polborn, K.; Steglich, W.; Mühlbacher, J.; Bringmann, G. *Tetrahedron* **2001,** *57,* 7857–7863. doi:10.1016/s0040-4020(01)00761-x

24. Eastwood, D. C.; Floudas, D.; Binder, M.; Majcherczyk, A.; Schneider, P.; Aerts, A.; Asiegbu, F. O.; Baker, S. E.; Barry, K.; Bendiksby, M.; Blumentritt, M.; Coutinho, P. M.; Cullen, D.; de Vries, R. P.; Gathman, A.; Goodell, B.; Henrissat, B.; Ihrmark, K.; Kauserud, H.; Kohler, A.; LaButti, K.; Lapidus, A.; Lavin, J. L.; Lee, Y.-H.; Lindquist, E.; Lilly, W.; Lucas, S.; Morin, E.; Murat, C.; Oguiza, J. A.; Park, J.; Pisabarro, A. G.; Riley, R.; Rosling, A.; Salamov, A.; Schmidt, O.; Schmutz, J.; Skrede, I.; Stenlid, J.; Wiebenga, A.; Xie, X.; Kües, U.; Hibbett, D. S.; Hoffmeister, D.; Högberg, N.; Martin, F.; Grigoriev, I. V.; Watkinson, S. C. *Science* **2011,** *333,* 762–765. doi:10.1126/science.1205411
25. Shah, F.; Schwenk, D.; Nicolás, C.; Persson, P.; Hoffmeister, D.; Tunlid, A. *Appl. Environ. Microbiol.* **2015,** *81,* 8427–8433. doi:10.1128/aem.02312-15
26. Zhu, J.; Chen, W.; Li, Y.-Y.; Deng, J.-J.; Zhu, D.-Y.; Duan, J.; Liu, Y.; Shi, G.-Y.; Xie, C.; Wang, H.-X.; Shen, Y.-M. *Gene* **2014,** *546,* 352–358. doi:10.1016/j.gene.2014.05.053
27. Sawitzke, J. A.; Thomason, L. C.; Bubunenko, M.; Li, X.; Costantino, N.; Court, D. L. *Methods Enzymol.* **2013,** *533,* 79–102. doi:10.1016/b978-0-12-420067-8.00007-6

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

New azodyrecins identified by a genome mining-directed reactivity-based screening

Atina Rizkiya Choirunnisa[‡1], Kuga Arima[‡1], Yo Abe[1], Noritaka Kagaya[2], Kei Kudo[3], Hikaru Suenaga[3], Junko Hashimoto[4], Manabu Fujie[5], Noriyuki Satoh[5], Kazuo Shin-ya[3], Kenichi Matsuda[*1,6] and Toshiyuki Wakimoto[*1,6]

Full Research Paper

Open Access

Address:
[1]Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060-0812, Japan, [2]Technology Research Association for Next Generation Natural Products Chemistry, Tokyo 135-0064, Japan, [3]National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan, [4]Japan Biological Informatics Consortium (JBIC), Tokyo 135-0064, Japan, [5]Okinawa Institute of Science and Technology Graduate University, Okinawa, 904-0495, Japan and [6]Global Station for Biosurfaces and Drug Discovery, Global Institution for Collaborative Research and Education, Hokkaido University, Sapporo 060-0812, Japan

Email:
Kenichi Matsuda[*] - kematsuda@pharm.hokudai.ac.jp;
Toshiyuki Wakimoto[*] - wakimoto@pharm.hokudai.ac.jp

* Corresponding author ‡ Equal contributors

Keywords:
biosynthesis; methyltransferase; natural azoxides; reactivity-based screening; *Streptomyces*

Beilstein J. Org. Chem. **2022,** *18,* 1017–1025.
https://doi.org/10.3762/bjoc.18.102

Received: 31 May 2022
Accepted: 29 July 2022
Published: 10 August 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

Only a few azoxy natural products have been identified despite their intriguing biological activities. Azodyrecins D–G, four new analogs of aliphatic azoxides, were identified from two *Streptomyces* species by a reactivity-based screening that targets azoxy bonds. A biological activity evaluation demonstrated that the double bond in the alkyl side chain is important for the cytotoxicity of azodyrecins. An in vitro assay elucidated the tailoring step of azodyrecin biosynthesis, which is mediated by the *S*-adenosylmethionine (SAM)-dependent methyltransferase Ady1. This study paves the way for the targeted isolation of aliphatic azoxy natural products through a genome-mining approach and further investigations of their biosynthetic mechanisms.

Introduction

Azoxy natural products are a rare yet intriguing class of natural products with various beneficial biological properties, such as antibacterial, antifungal, nematicidal, and cytotoxic activities (Figure 1) [1-3]. Since the discovery of the natural azoxy compound macrozamin in 1951 as the first example of a nitrogen–nitrogen bond-containing natural product [4], azoxy

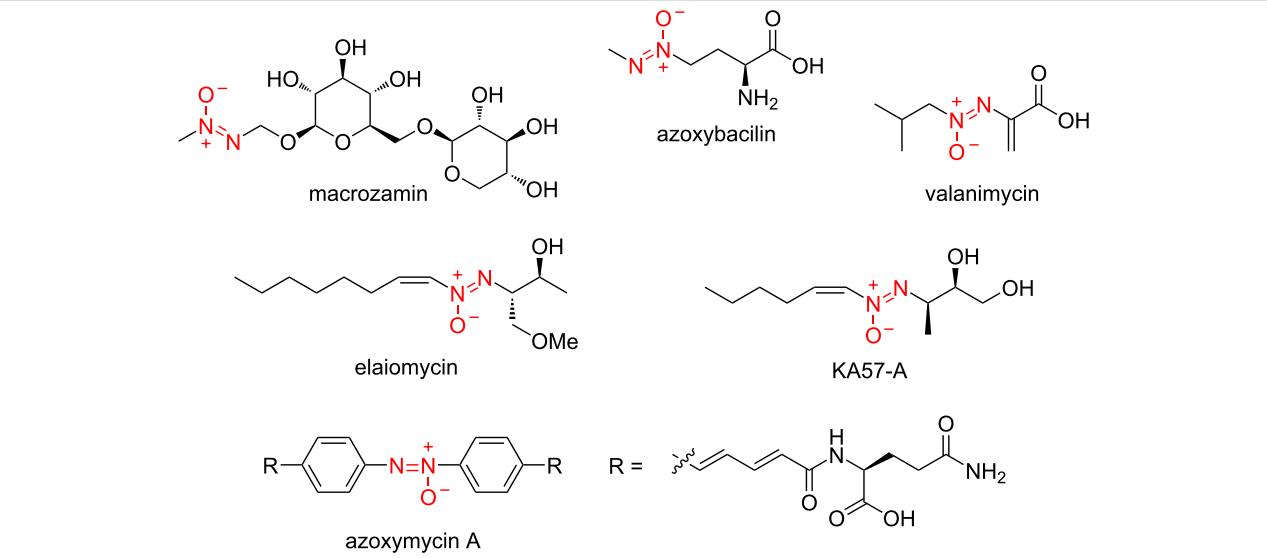


Figure 1: Representative natural azoxides.

compounds have been isolated from various natural sources including bacteria, fungi, plants, and marine sponges [1-3]. Azoxy natural products have occasionally been discovered by conventional isolation schemes guided by biological activities or physicochemical properties, which are not selective for the azoxy functionality. Consequently, there are only a few examples of azoxy natural products, despite their notable biological activities.

Reactivity-based screening is an emerging strategy in natural products discovery, in which chemical probes are used for the specific detection of the unique functionality of interest in crude metabolites [5,6]. The reactions usually facilitate the subsequent isolation process of the target molecules. This strategy has been successfully applied for detecting a range of peculiar functional groups, such as ureido [7], isocyanide [8], and alkyne [9,10]. A combination of reactivity-based screening and genome-based prioritization would allow the prediction of the producer of natural products with the functional groups of interest, leading to a higher rate of hits. We recently exploited genome mining targeting hydrazine synthetase and the generation of inorganic hydrazine N_2H_4 in acid hydrolysate as an indicator of N–N bond-containing functional groups, which led to the discovery of actinopyridazinones with a unique dihydropyridazinone scaffold [11]. N_2H_4 generates similarly in the acid hydrolysate of azoxy compounds, and this feature has been exploited as proof for the presence of N–N bonds in the plant-derived methyl azoxy compound macrozamin [4]. Although the N_2H_4 generation is an advantageous reactivity of an azoxy group, to the best of our knowledge, it has not been applied for the detection of azoxy bonds in the context of natural products discovery.

The azoxy bond is biosynthesized by two distinct mechanisms. The first is the nitrogen radical coupling mechanism in the biosynthesis of azoxymycins [12,13], which are aromatic azoxy natural products. A similar mechanism has been envisioned for the autoxidation and spontaneous dimerization of aliphatic hydroxylamines via the azoxy linkage in malleobactin D biosynthesis [14]. A distinct mechanism is employed in the biosynthesis of valanimycin, an aliphatic azoxy natural product. This involves the *N*-hydroxylation of isobutylamine, mediated by the flavin-dependent monooxygenase VlmH [15-17], and the following formation of *O*-(L-seryl)-isobutylhydroxylamine by the tRNA-utilizing enzyme VlmA [18]. This intermediate is hypothesized to be transformed into the azoxy bond-containing intermediate via an intramolecular rearrangement accompanied by a concomitant oxidation [18]. Although the exact mechanisms of azoxy bond formation remain unclear, VlmH and VlmA cooperate to biosynthesize the *N*-acyl intermediate for azoxy bonds, suggesting that a genetic region containing both genes could be a potential biosynthetic gene cluster of aliphatic azoxy natural products.

Results and Discussion

Reactivity-based isolation of azodyrecins from two *Streptomyces* strains

During our efforts toward the discovery of N–N bond-containing natural products from our in-house actinobacterial culture collection, we found two closely related potential biosynthetic gene clusters of aliphatic azoxy natural products in *Streptomyces* sp. RM72 and *Streptomyces* sp. A1C6 (Figure 2 and Tables S1 and S2 in Supporting Information File 1). They encode five genes that are also present in the biosynthetic gene

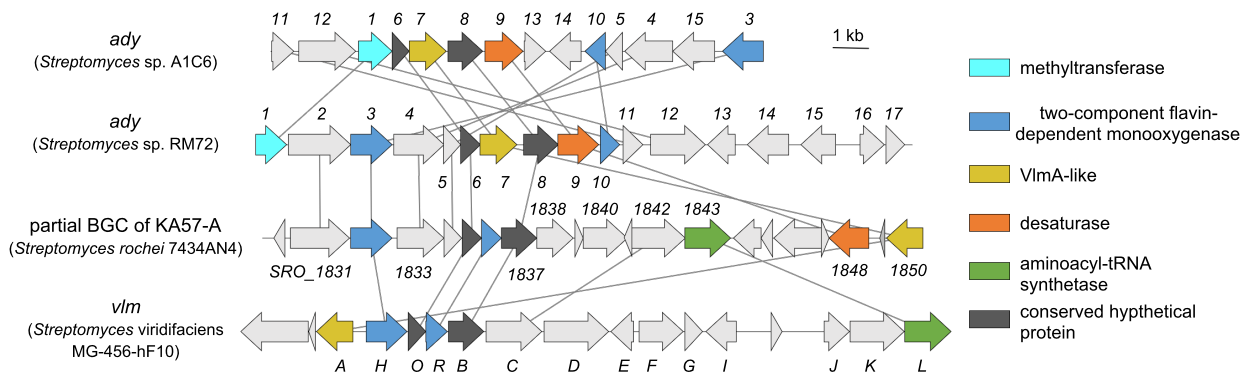
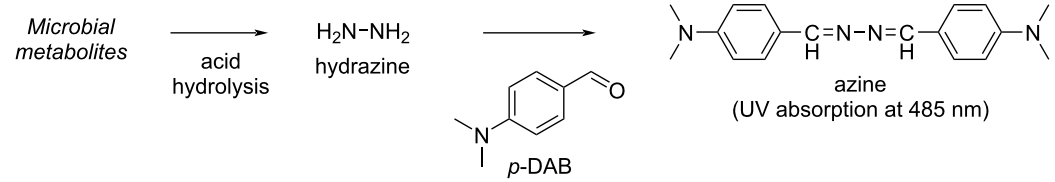


Figure 2: Biosynthetic gene clusters of aliphatic azoxy natural products. Conserved proteins are colored according to their functional annotations. Genes encoding homologous proteins with more than 30% amino acid sequence identities are linked.

clusters of valanimycin [19] and KA57-A [20]: the putative two-component flavin-dependent monooxygenase similar to VlmH/VlmR, the homologous protein of VlmA, and two additional hypothetical proteins similar to VlmB and VlmO. Based on this observation, we hypothesized that these two *Streptomyces* strains potentially produce aliphatic azoxy natural products similar to valanimycins and KA57-A, and aimed to identify their biosynthetic products. To this end, we conducted a reactivity-based screening to detect N_2H_4, which could be generated upon the acid hydrolysis of azoxy natural products. In the assay, N_2H_4 is captured by two equivalents of *p*-dimethylaminobenzaldehyde (DAB) to generate *p*-dimethylaminobenzaldazine, which can be sensitively detected by HPLC by monitoring the UV absorption at 485 nm (Scheme 1). As a result, we observed the generation of N_2H_4 upon the hydrolysis of solid-culture extracts of both *Streptomyces* strains (Supporting Information File 1, Figure S1). Therefore, we attempted to isolate the azoxy natural products in an N_2H_4-detecting assay guided manner.

The extracts of *Streptomyces* sp. RM72 were first partitioned by water and ethyl acetate, and then the organic layer was further fractionated by silica gel column chromatography. Fractionation by reversed-phase HPLC yielded ten compounds (**1**–**10**) that generate N_2H_4 upon acid hydrolysis. The combination of ^1H and ^{13}C NMR with a series of 2D NMR analyses and optical rotation revealed that six of the isolated compounds are azodyrecin A (**1**), azodyrecin B (**2**), azodyrecin C (**3**), and their geometric isomers 1'-*trans*-azodyrecin A (**4**), 1'-*trans*-azodyrecin B (**5**), and 1'-*trans*-azodyrecin C (**6**), respectively, which were previously discovered from *Streptomyces* sp. strain P8-A2 (Figure 3a) [21]. Of note, the *trans*-azodyrecins **4**–**6** are reportedly generated by the spontaneous isomerization of the *cis*-congeners, after long-term exposure to $CHCl_3$ [21]. While *trans*-azodyrecins could also be generated nonenzymatically from their *cis*-congeners during the isolation process, *trans*-azodyrecins were detected in the fresh extracts of the solid cultures from both strains, suggesting that they are also generated in vivo (Figure S2 in Supporting Information File 1).

The NMR spectra of compound **7** are similar but distinct from those of the known azodyrecins. Close inspection of **7** revealed the absence of olefinic proton signals (7.09 ppm and 6.99 ppm in **4**) and additional methylene proton signals at 4.21 ppm and 1.93 ppm in **7** (Table 1, Figure S3, Supporting Information File 1). The chemical formula predicted with HRMS ($C_{18}H_{36}O_3N_2$) indicated the presence of two additional hydrogens as compared to compounds **1** and **4**. Collectively, compound **7** was identified as a new analog with a fully saturated alkyl side chain, named azodyrecin D (**7**) (Figure 3b). Likewise, compounds **8**, **9**, and **10** also lack olefinic proton signals (7.09 ppm and 6.99 ppm in **4**) but possess methylene protons (4.21 ppm and 1.93 ppm in **8**, 4.22 ppm and 1.92 ppm in **9**, and 4.21 ppm and 1.81 ppm in **10**). HRMS showed that compound **8**



Scheme 1: N_2H_4-detecting colorimetric assay.

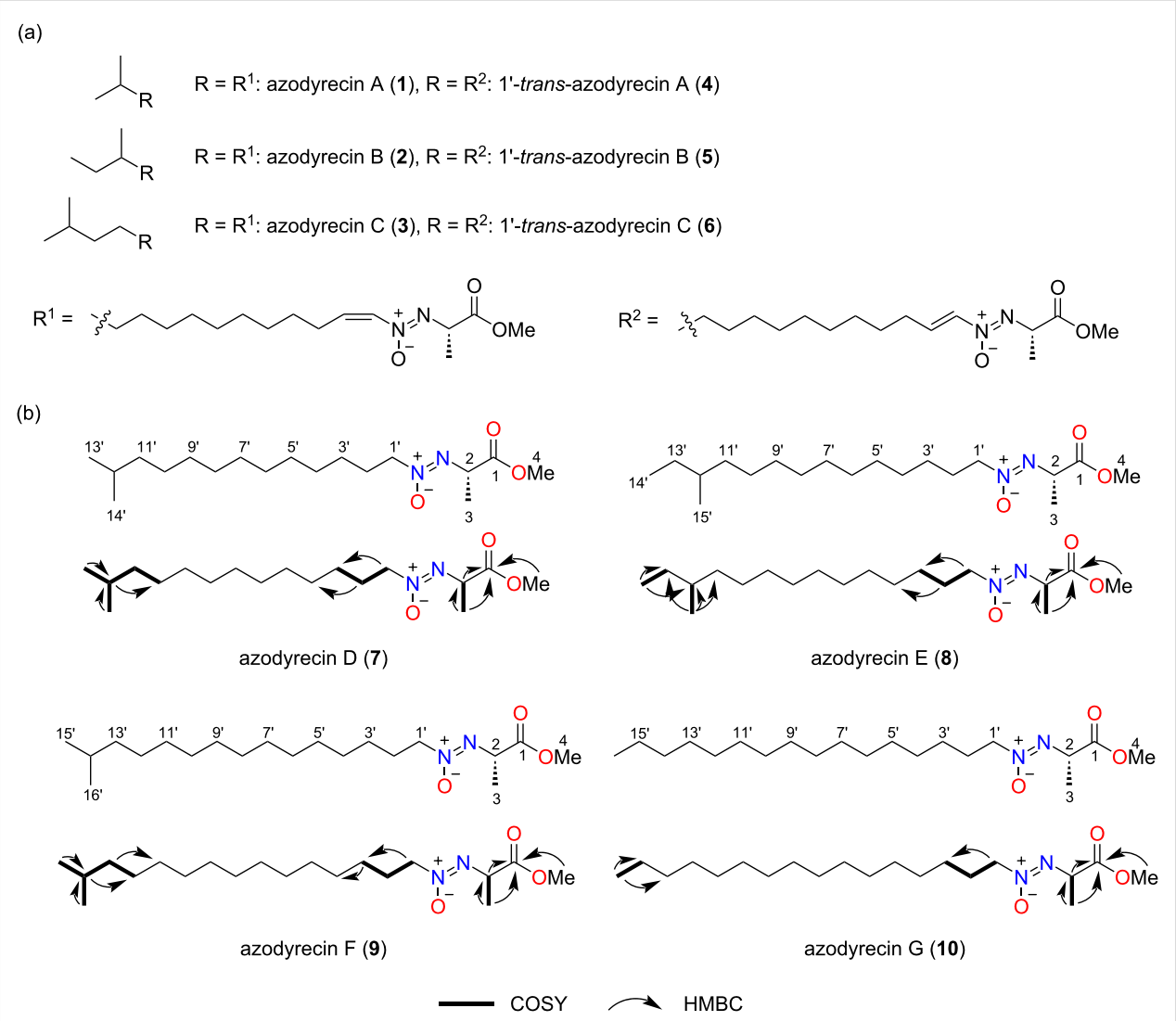


Figure 3: Structures of azodyrecins (a) and new azodyrecin derivatives, azodyrecins D–G (**7**–**10**) (b). Key correlations in 2D NMR are shown.

possesses two additional hydrogens compared to **2** and **5**, while compounds **9** and **10** possess two additional hydrogens compared to compounds **3** and **6**. These data, together with 2D NMR spectra showed that compounds **8**, **9**, and **10** were azodyrecin derivatives with saturated alkyl side chains, which were named azodyrecin E (**8**), azodyrecin F (**9**), and azodyrecin G (**10**), respectively (Figure S3, Supporting Information File 1). Although azodyrecins characteristically possess branched alkyl side chains, compound **10** is the only azodyrecin analog with a straight alkyl chain. The configurations of the azoxy groups in **7**–**10** were determined to be *Z*, as in the case of other azodyrecins, according to the characteristic UV absorption at 221 nm (Figure S4 in Supporting Information File 1) [22]. Based on optical rotation values and biosynthetic correlation to known azodyrecins, the configuration of compounds **7**–**10** was defined to be 2*S*.

Fractionation guided by the N_2H_4-detecting assay revealed that compounds **1**–**10** were also produced by *Streptomyces* sp. A1C6 (Figure S2 in Supporting Information File 1). This result identified a similar yet distinct type of azodyrecin biosynthetic gene cluster that contains several insertions and inversions, which generally make it challenging to precisely predict its biosynthetic products based solely on genome information (Figure 2). Taken together, the N_2H_4-detecting reactivity-based screening led to the identification of four new analogs and two types of biosynthetic gene clusters of azodyrecins, demonstrating its utility in natural product discovery and deorphanization of biosynthetic gene clusters.

Evaluation of cytotoxicities of azodyrecins

A previous report demonstrated the cytotoxic activity of azodyrecins [21]; however, the relationships between the struc-

Table 1: ^{1}H (500 MHz) and ^{13}C (125 MHz) NMR data for azodyrecin D (**7**), azodyrecin E (**8**), azodyrecin F (**9**), and azodyrecin G (**10**).[a]

	compound **7**			compound **8**			compound **9**			compound **10**		
pos.	δ_C	δ_H	m, *J* (Hz)	δ_C	δ_H	m, *J* (Hz)	δ_C	δ_H	m, *J* (Hz)	δ_C	δ_H	m, *J* (Hz)
1	172.7			172.7			172.9			170.6		
2	59.9	4.38	q, 7.2	59.9	4.38	q, 7.1	60.0	4.38	q, 7.2	58.1	4.33	q, 7.5
3	16.1	1.45	d, 7.2	16.1	1.45	d, 7.2	15.9	1.45	d, 7.0	15.4	1.38	d, 7.1
4	52.6	3.66	s	52.6	3.67	s	52.5	3.66	s	51.8	3.59	s
1’	70.6	4.21	m	70.6	4.21	m	70.6	4.22	m	68.9	4.21	m
2’	28.9	1.93	m	28.9	1.93	m	28.8	1.92	m	27.2	1.81	m
3’	27.3	1.36[b]	m	27.3	1.37[b]	m	27.2	1.36[b]	m	25.4	1.20–1.30	m
4’	30.2	1.25–1.35	m	30.4	1.25–1.36	m	30.0	1.26–1.36	m	28.3–29.0	1.20–1.30	m
5’	29.2–31.0	1.25–1.35	m	28.4–31.3	1.25–1.36	m	30.5–30.8	1.26–1.36	m	28.3–29.0	1.20–1.30	m
6’	29.2–31.0	1.25–1.35	m	28.4–31.3	1.25–1.36	m	30.5–30.8	1.26–1.36	m	28.3–29.0	1.20–1.30	m
7’	29.2–31.0	1.25–1.35	m	28.4–31.3	1.25–1.36	m	30.5–30.8	1.26–1.36	m	28.3–29.0	1.20–1.30	m
8’	29.2–31.0	1.25–1.35	m	28.4–31.3	1.25–1.36	m	30.5–30.8	1.26–1.36	m	28.3–29.0	1.20–1.30	m
9’	29.2–31.0	1.25–1.35	m	28.4–31.3	1.25–1.36	m	30.5–30.8	1.26–1.36	m	28.3–29.0	1.20–1.30	m
10’	28.7	1.29[b]	m	28.4–31.3	1.25–1.36	m	30.5–30.8	1.26–1.36	m	28.3–29.0	1.20–1.30	m
11’	40.4	1.18	m m	37.9	1.31[b] 1.10[b]	m m	31.0	1.26–1.36	m	28.3–29.0	1.20–1.30	m m
12’	29.2	1.53	m	35.8	1.33[b]	m	28.5	1.28[b]	m	28.3–29.0	1.20–1.30	m
13’	23.2	0.88	d, 6.7	30.7	1.15 1.25–1.36	m m	40.2	1.17	m	28.3–29.0	1.20–1.30	
14’	23.3	0.88	d, 6.7	12.0	0.87	m	29.1	1.52	m	31.3	1.20–1.30	m
15’				19.9	0.86	m	23.0	0.87	d, 6.7	22.1	1.20–1.30	m
16’							23.0	0.87	d, 6.7	13.9	0.85	d, 6.6

[a]Spectra for compounds **7**–**9** were measured in methanol-d_4 and spectrum for compound **10** was measured in DMSO-d_6; [b]determined by HSQC.

ture and cytotoxicity were not determined. With new azodyrecin analogs with saturated alkyl chains in hand, we attempted to gain insights into the effects of the double bond on cytotoxicity. To this end, we assessed the cytotoxic activities of azodyrecin B (**2**), 1’-*trans*-azodyrecin B (**5**), together with the most abundant new analogs azodyrecin D (**7**), and azodyrecin E (**8**), against human ovarian adenocarcinoma SKOV-3 cells, malignant pleural mesothelioma MESO-1 cells, immortalized T lymphocyte Jurkat cells, and P388 murine leukemia cells. The results revealed that the derivatives with unsaturated side chains, **2** and **5**, exhibited moderate cytotoxicity against all tested cell lines, with the highest potency against Jurkat cells (IC_{50} at 3.36 μM for **5**) (Table 2 and Figure S5 in Supporting Information File 1). In contrast, the saturated analogs **7** and **8** showed no or negligible cytotoxicity, clearly supporting the importance of the double bond adjacent to the azoxy bond for the cytotoxicity. This result is in contrast to the similar structure–activity relationship (SAR) profiles of elaiomycins, structurally related aliphatic azoxy natural products, in which the antimicrobial and cytotoxic activities do not depend on the

Table 2: Cytotoxicities of compound **2**, **5**, **7**, and **8**.

	IC_{50} (μM)			
	SKOV-3	MESO-1	Jurkat	P388
azodyrecin B (**2**)	7.37	9.70	8.72	11.6
1’-*trans*-azodyrecin B (**5**)	8.24	6.70	3.36	4.72
azodyrecin D (**7**)	>50	43.2	>50	>50
azodyrecin E (**8**)	>50	>50	>50	>50

double bond adjacent to the azoxy bond [23]. The difference in the SAR profiles between azodyrecins and elaiomycins suggests their distinct modes of actions.

Biosynthetic origin of the unique methyl ester in azodyrecin

The structural diversity of aliphatic azoxy natural products can be attributed to variations in the alkyl side chains and the amino acid-derived counterparts. The variation in the amino acid-derived units is considerably large, as it includes primary and secondary alcohols [24], methoxides [23,25,26], carboxylic acids [27], amides [28], ketones [29,30], an exo-olefin [31], and lactones [32]. Elucidating the mechanisms of structural diversification is essential when considering the synthesis of unnatural azoxides by a synthetic biology-based approach. However, their enzymatic basis has remained elusive except for the exo-olefin formation in valanimycin biosynthesis, which is mediated by the phosphorylation of a serine moiety by VlmJ and the subsequent dehydration by VlmK [33]. To obtain insights into the late-stage diversification mechanisms, we focused on the biosynthesis of the methyl ester, which is unique to azodyrecins. To this end, we characterized the putative SAM-dependent methyltransferase Ady1 in vitro to assess its activity against the carboxylic acid **11**, which was prepared by the hydrolysis of compound **8** under basic conditions. When acid **11** was incubated with recombinant Ady1 in the presence of SAM, it was converted to **8**, showing that Ady1 can install the methyl ester of azodyrecins (Figure 4).

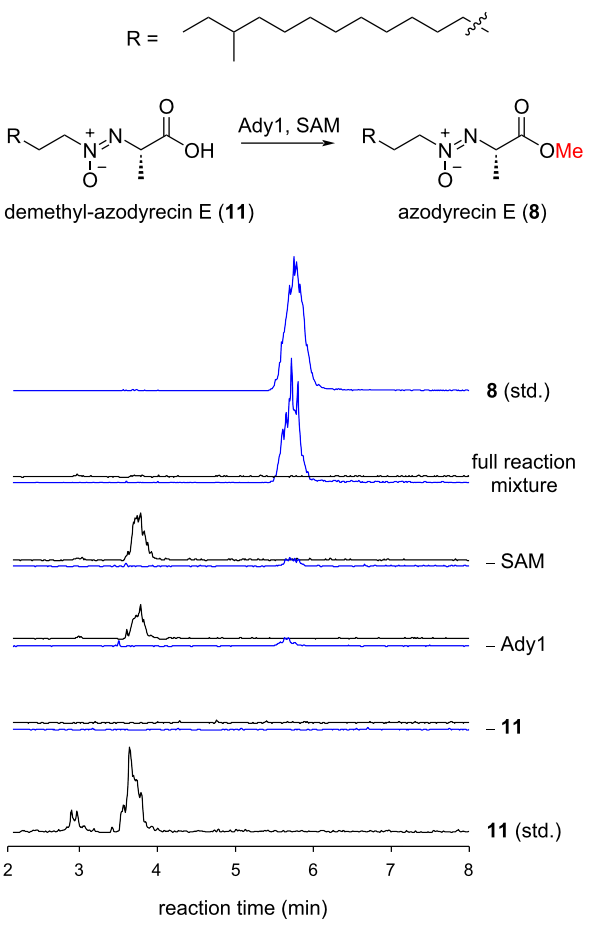
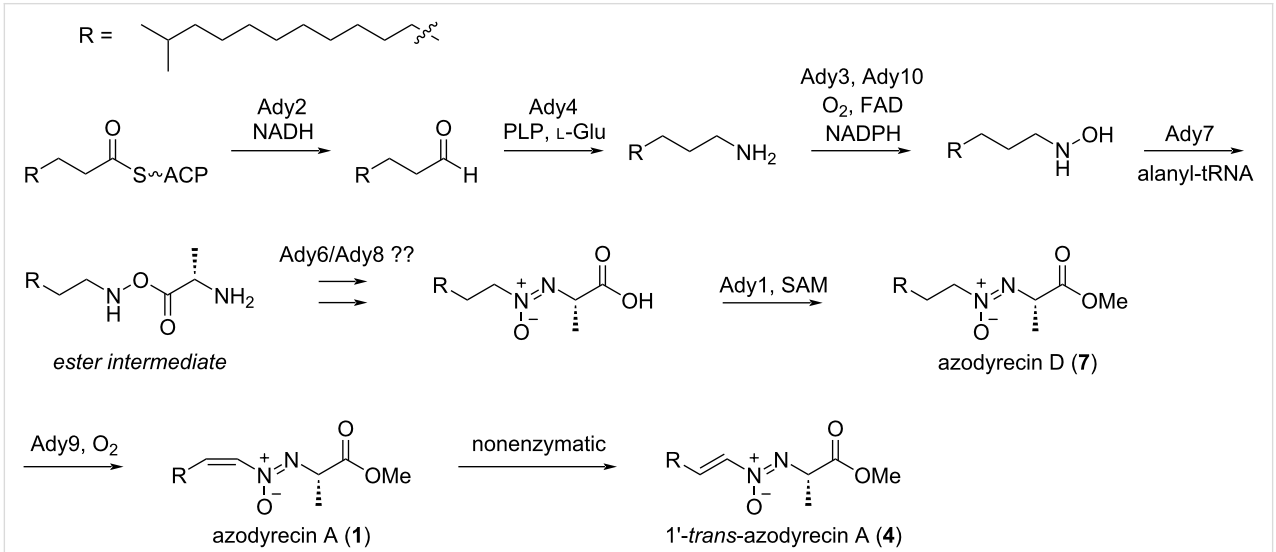


Figure 4: In vitro characterization of Ady1. Extracted ion chromatograms at *m/z* 329.3 (black) and *m/z* 343.3 (blue), corresponding to compound **11** and **8**, respectively, are shown.

The in vitro characterization of Ady1 and the functional annotation of *ady* clusters allowed the prediction of the entire biosynthetic pathway of azodyrecins (Scheme 2). The pathway is



Scheme 2: Proposed biosynthetic pathway of azodyrecin.

likely initiated by Ady2, a putative dehydrogenase that recruits fatty acids from primary metabolism to generate an aldehyde, which would be converted to an aliphatic amine by the pyridoxal phosphate (PLP)-dependent transaminase Ady4. The amine would be *N*-hydroxylated by the two-component flavin-dependent monooxygenase Ady3/Ady10, as in the valanimycin biosynthesis mediated by VlmH/VlmR [15-17]. The hydroxylamine would be conjugated to alanyl-tRNA to form an ester intermediate by the function of the tRNA-utilizing enzyme Ady7, which is homologous to VlmA. In valanimycin biosynthesis, the substrate seryl-tRNA is provided by VlmL, an additional seryl-tRNA synthetase (SerRS) encoded within the *vlm* cluster [34]. However, no aminoacyl-tRNA synthetase gene is present in the *ady* cluster, suggesting that the alanyl-tRNA is directly provided from the cellular tRNA pool in the case of azodyrecin biosynthesis. The mechanism for the subsequent rearrangement of the ester intermediate for azoxy bond formation remains unclear; however, the conservation of the two hypothetical proteins Ady6/Ady8 among the biosynthetic gene clusters of valanimycin, KA57-A, and azodyrecins may suggest their participation in this step. Ady6 shows weak homology to DUF4260 (PF14079.9), a family of integral membrane proteins with unknown functions, while Ady8 is similar to the ferritin-like superfamily protein (IPR009078). Ady6/Ady8 are homologous to VlmO/VlmB and SRO_1835/SRO_1837 in the biosynthetic gene clusters of valanimycin and KA57-A, respectively. The in vitro characterization of Ady1 suggested the late-stage biosynthetic pathway of azodyrecin: the azoxy bond formation is followed by the Ady1-mediated methyl esterification to form saturated azodyrecins, and then the subsequent installation of a *cis*-olefin on the 1,2-positions of the alkyl side chain would afford azodyrecins A–C (**1**–**3**), in a reaction likely to be mediated by Ady9, a putative fatty acid desaturase. Nevertheless, the possibility that the desaturation occurs prior to the methyl esterification could not be excluded. The elucidation of the exact order of the late-stage modifications requires further investigations, such as gene knockout experiments and substrate scope analyses of Ady1, which will be accomplished in future work.

Distribution of potential biosynthetic gene clusters of aliphatic azoxy natural products

To obtain insights into the distribution and diversity of the biosynthetic gene clusters of aliphatic azoxy natural products, we searched for a pair of VlmA and VlmH, two key enzymes in the azoxy bond formation, encoded in close genetic loci in the reference genome of 3,146 actinobacteria in the NCBI database. A stepwise HMM-based search using models for "VlmA" (PF09924: LPG_synthase_C) and the amino acid sequence of VlmH identified 179 pairs of VlmA/VlmH, indicating that approximately 5.7% of the actinobacteria present in the NCBI Refseq database are potential producers of aliphatic azoxy natural products. The sequence similarity network (SSN) generated by all-vs-all blastp with *E*-value at 1×10^{-70} classified these "VlmAs" into more than fourteen groups (Figure 5). In this network, the two Ady7s from *Streptomyces* sp. RM72 and *Streptomyces* sp. A1C6 belong to group 3, while VlmA and SRO_1850 belong to groups 12 and 1, respectively. An analysis of the genome neighborhoods of "VlmA" revealed that the three genes encoding "VlmB/O/R" are highly conserved (Figure S7, Supporting Information File 1), suggesting the functional relevance of these genes with VlmA and VlmH. Additionally, a comparison of representative gene clusters from each group indicated that several protein families are frequently observed in the genome neighborhoods of specific "VlmA" groups, such as Ady1-like methyltransferases (PF04072), homologous pairs of VlmJ/VlmK-like exo-olefin-forming enzymes (PF19279/PF03972), seryl-tRNA synthetases (PF02403), putative Trp halogenase-like enzymes (PF04820), and putative 3-oxoacyl-[acyl-carrier-protein (ACP)] synthase III-like enzymes (PF08541) (Figures S7 and S8 in Supporting Information File 1). The various protein families encoded in the proximity of the "VlmA" gene suggest the manifold biosynthetic pathways and structural diversity of aliphatic azoxy natural products. Considering that most groups in the SSN lack links to their biosynthetic products, a substantial fraction of the chemical diversity in aliphatic natural azoxides likely remains untapped.

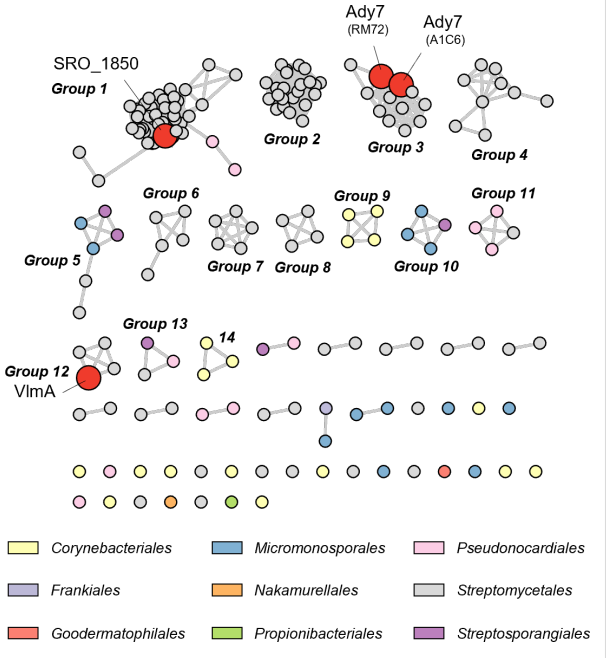


Figure 5: Sequence similarity network of VlmA-like enzymes in the actinobacterial genomes in the Refseq database. Nodes are colored according to the host organism's order. Enzymes with known biosynthetic products are colored red.

Conclusion

By using the generation of N_2H_4 as the indicator of an azoxy bond, we conducted a reactivity-based screening for aliphatic azoxy natural products. This led to the identification of two new producers of azodyrecins, as well as the new analogs **7**–**10**, demonstrating the utility of this reactivity-based strategy for natural products discovery. Bioinformatic surveys shed light on the unexplored biosynthetic potential of actinobacteria for aliphatic azoxides, setting the stage for the targeted isolation of this scarce yet valuable class of natural products with remarkable biological activities.

Supporting Information

Supporting Information File 1

Experimental procedures, characterization data (^{1}H, ^{13}C NMR, and HRMS) and biochemical characterization of recombinant Ady1.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-102-S1.pdf]

Acknowledgements

We thank to Dr. Eri Fukushi (Hokkaido University) for technical assistance in measurement of NMR.

Funding

This work was partly 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 Asahi Glass Foundation, the Naito Foundation, the Uehara Memorial Foundation, the Sumitomo Foundation−Grant for Basic Science Research Projects, Daiichi Sankyo Foundation of Life Science, the Japan Agency for Medical Research and Development JP19ae0101045, Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Science and Technology Agency (JST Grant Numbers ACT-X JPMJAX201F and A-STEP JPMJTR20US), Japan JP16H06448, JP18H02581, JP19K16390, JP21H02635, and JP22K15302.

ORCID® iDs

Atina Rizkiya Choirunnisa - https://orcid.org/0000-0001-6594-1125
Kuga Arima - https://orcid.org/0000-0002-1093-6342
Kei Kudo - https://orcid.org/0000-0001-7500-9462
Kenichi Matsuda - https://orcid.org/0000-0002-9269-688X
Toshiyuki Wakimoto - https://orcid.org/0000-0003-2917-1797

References

1. Wibowo, M.; Ding, L. *J. Nat. Prod.* **2020,** *83,* 3482–3491. doi:10.1021/acs.jnatprod.0c00725
2. Blair, L. M.; Sperry, J. *J. Nat. Prod.* **2013,** *76,* 794–812. doi:10.1021/np400124n
3. Le Goff, G.; Ouazzani, J. *Bioorg. Med. Chem.* **2014,** *22,* 6529–6544. doi:10.1016/j.bmc.2014.10.011
4. Langley, B. W.; Lythgoe, B.; Riggs, N. V. *J. Chem. Soc.* **1951,** 2309–2316. doi:10.1039/jr9510002309
5. Hughes, C. C. *Nat. Prod. Rep.* **2021,** *38,* 1684–1705. doi:10.1039/d0np00034e
6. Harris, L. A.; Mitchell, D. A. *Methods Enzymol.* **2022,** *665,* 177–208. doi:10.1016/bs.mie.2021.11.018
7. 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
8. Huang, Y.-B.; Cai, W.; Del Rio Flores, A.; Twigg, F. F.; Zhang, W. *Anal. Chem. (Washington, DC, U. S.)* **2020,** *92,* 599–602. doi:10.1021/acs.analchem.9b05147
9. Moss, N. A.; Seiler, G.; Leão, T. F.; Castro-Falcón, G.; Gerwick, L.; Hughes, C. C.; Gerwick, W. H. *Angew. Chem., Int. Ed.* **2019,** *58,* 9027–9031. doi:10.1002/anie.201902571
10. Back, D.; Shaffer, B. T.; Loper, J. E.; Philmus, B. *J. Nat. Prod.* **2022,** *85,* 105–114. doi:10.1021/acs.jnatprod.1c00798
11. Matsuda, K.; Arima, K.; Akiyama, S.; Yamada, Y.; Abe, Y.; Suenaga, H.; Hashimoto, J.; Shin-ya, K.; Nishiyama, M.; Wakimoto, T. *J. Am. Chem. Soc.* **2022,** *144,* 12954–12960. doi:10.1021/jacs.2c05269
12. Guo, Y.-Y.; Li, H.; Zhou, Z.-X.; Mao, X.-M.; Tang, Y.; Chen, X.; Jiang, X.-H.; Liu, Y.; Jiang, H.; Li, Y.-Q. *Org. Lett.* **2015,** *17,* 6114–6117. doi:10.1021/acs.orglett.5b03137
13. Guo, Y.-Y.; Li, Z.-H.; Xia, T.-Y.; Du, Y.-L.; Mao, X.-M.; Li, Y.-Q. *Nat. Commun.* **2019,** *10,* 4420. doi:10.1038/s41467-019-12250-1
14. Franke, J.; Ishida, K.; Ishida-Ito, M.; Hertweck, C. *Angew. Chem., Int. Ed.* **2013,** *52,* 8271–8275. doi:10.1002/anie.201303196
15. Parry, R. J.; Li, W. *Arch. Biochem. Biophys.* **1997,** *339,* 47–54. doi:10.1006/abbi.1996.9857
16. Parry, R. J.; Li, W. *J. Biol. Chem.* **1997,** *272,* 23303–23311. doi:10.1074/jbc.272.37.23303
17. Li, H.; Forson, B.; Eckshtain-Levi, M.; Valentino, H.; Martín del Campo, J. S.; Tanner, J. J.; Sobrado, P. *Biochemistry* **2021,** *60,* 31–40. doi:10.1021/acs.biochem.0c00679
18. Garg, R. P.; Qian, X. L.; Alemany, L. B.; Moran, S.; Parry, R. J. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 6543–6547. doi:10.1073/pnas.0708957105
19. Garg, R. P.; Ma, Y.; Hoyt, J. C.; Parry, R. J. *Mol. Microbiol.* **2002,** *46,* 505–517. doi:10.1046/j.1365-2958.2002.03169.x
20. Nindita, Y.; Cao, Z.; Fauzi, A. A.; Teshima, A.; Misaki, Y.; Muslimin, R.; Yang, Y.; Shiwa, Y.; Yoshikawa, H.; Tagami, M.; Lezhava, A.; Ishikawa, J.; Kuroda, M.; Sekizuka, T.; Inada, K.; Kinashi, H.; Arakawa, K. *Sci. Rep.* **2019,** *9,* 10973. doi:10.1038/s41598-019-47406-y
21. Wibowo, M.; Gotfredsen, C. H.; Sassetti, E.; Melchiorsen, J.; Clausen, M. H.; Gram, L.; Ding, L. *J. Nat. Prod.* **2020,** *83,* 3519–3525. doi:10.1021/acs.jnatprod.0c00339
22. Taylor, K. G.; Riehl, T. *J. Am. Chem. Soc.* **1972,** *94,* 250–255. doi:10.1021/ja00756a044

23. Ding, L.; Ndejouong, B. L. S. T.; Maier, A.; Fiebig, H.-H.; Hertweck, C. *J. Nat. Prod.* **2012,** *75,* 1729–1734. doi:10.1021/np300329m
24. Kunitake, H.; Hiramatsu, T.; Kinashi, H.; Arakawa, K. *ChemBioChem* **2015,** *16,* 2237–2243. doi:10.1002/cbic.201500393
25. Stevens, C. L.; Gillis, B. T.; French, J. C.; Haskell, T. H. *J. Am. Chem. Soc.* **1956,** *78,* 3229–3230. doi:10.1021/ja01594a078
26. Haskell, T. H.; Ryder, A.; Bartz, Q. R. *Antibiot. Chemother.* **1954,** *4,* 141–144.
27. Manderscheid, N.; Helaly, S. E.; Kulik, A.; Wiese, J.; Imhoff, J. F.; Fiedler, H.-P.; Süssmuth, R. D. *J. Antibiot.* **2013,** *66,* 85–88. doi:10.1038/ja.2012.99
28. Fukumoto, A.; Murakami, C.; Anzai, Y.; Kato, F. *J. Antibiot.* **2016,** *69,* 395–399. doi:10.1038/ja.2015.126
29. Takahashi, Y.; Nakayama, M.; Watanabe, I.; Deushi, T.; Ishiwata, H.; Shiratsuchi, M.; Otani, G. *J. Antibiot.* **1989,** *42,* 1541–1546. doi:10.7164/antibiotics.42.1541
30. McGahren, W. J.; Kunstmann, M. P. *J. Am. Chem. Soc.* **1969,** *91,* 2808–2810. doi:10.1021/ja01038a081
31. Yamato, M.; Iinuma, H.; Naganawa, H.; Yamagishi, Y.; Hamada, M.; Masuda, T.; Umezawa, H.; Abe, V.; Hori, M. *J. Antibiot.* **1986,** *39,* 184–191. doi:10.7164/antibiotics.39.184
32. Nakano, H.; Hara, M.; Katsuyama, T.; Uozaki, Y.; Gomi, K. Compound DC0110. WO Pat. Appl. WO1993009110A1, May 13, 1993.
33. Garg, R. P.; Alemany, L. B.; Moran, S.; Parry, R. J. *J. Am. Chem. Soc.* **2009,** *131,* 9608–9609. doi:10.1021/ja901243p
34. Garg, R. P.; Gonzalez, J. M.; Parry, R. J. *J. Biol. Chem.* **2006,** *281,* 26785–26791. doi:10.1074/jbc.m603675200

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

A *Streptomyces* P450 enzyme dimerizes isoflavones from plants

Run-Zhou Liu[1,2,3], Shanchong Chen[2,3] and Lihan Zhang[*2,3]

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, Westlake University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China and [3]Institute of Natural Sciences, Westlake Institute for Advanced Study, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China

Email:
Lihan Zhang[*] - zhanglihan@westlake.edu.cn

* Corresponding author

Keywords:
biaryl coupling; cytochrome P450; dimerization; isoflavone; natural product

Beilstein J. Org. Chem. **2022,** *18,* 1107–1115.
https://doi.org/10.3762/bjoc.18.113

Received: 31 May 2022
Accepted: 11 August 2022
Published: 26 August 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

Dimerization is a widespread natural strategy that enables rapid structural diversification of natural products. However, our understanding of the dimerization enzymes involved in this biotransformation is still limited compared to the numerous reported dimeric natural products. Here, we report the characterization of three new isoflavone dimers from *Streptomyces cattleya* cultured on an isoflavone-containing agar plate. We further identified a cytochrome P450 monooxygenase, CYP158C1, which is able to catalyze the dimerization of isoflavones. CYP158C1 can also dimerize plant-derived polyketides, such as flavonoids and stilbenes. Our work represents a unique bacterial P450 that can dimerize plant polyphenols, which extends the insights into P450-mediated biaryl coupling reactions in biosynthesis.

Introduction

Dimerization is a ubiquitous biotransformation in nature. Almost all forms of life, including bacteria, fungi, and plants, have the ability to produce homo- or heterodimeric natural products, which enables rapid structural diversification from simple monomers [1]. Often, the dimerized products exhibit significant biological activities due to the increased functional group density or complex stereochemistry, as seen in vinblastine [2], julichrome [3], himastatin [4], and biflavone [5]. However, the structural complexity of the dimeric natural products has hampered synthetic chemistry approaches towards these molecules, since chemo-, regio-, and atroposelective formation of the biaryl linkage remains highly challenging and often requires prefunctionalization of the substrate monomers, costly metal ligands, or tedious protection–deprotection steps [6-9].

With the advance of biosynthetic studies on natural products, a number of enzyme classes that are responsible for the dimer formation have been identified [1,10-14]. In plants and fungi, laccases and cytochrome P450 monooxygenases play pivotal roles in the biaryl bond formation of various polyketide dimers [10,15,16]. In contrast, in bacteria, P450 enzymes are the dominant catalysts, but no laccases have been reported for dimerization reactions (Figure S1, Supporting Information File 1). Due to the high reaction selectivity that the enzyme active site offers, these enzymes provide biocatalytic means for the biaryl linkage formation, and recent enzyme engineering efforts also demonstrated selective and efficient production of unnatural dimers or cross-coupling products, starting from simple monomers [17-19]. Nevertheless, our knowledge of enzyme-mediated dimerization is still limited in contrast to the numerous reported dimeric natural products.

Phenol coupling in plant polyphenol biosynthesis is one of the earliest documented biocatalytic dimerization reactions [20]. Contrary to flavone dimers and oligomers being abundant in nature, only limited dimeric compounds have been reported for isoflavones. Isoflavones bear a characteristic 3-phenylchroman skeleton, which is formed by the B-ring migration from the flavonoid scaffold catalyzed by isoflavone synthases [21]. The sporadic distribution of this isoflavone synthase limits the discovery of isoflavones in the plant kingdom [22], and the enzymes catalyzing isoflavone dimerization, to our knowledge, remain uncharacterized.

In this study, we report the discovery of a *Streptomyces* cytochrome P450 enzyme that catalyzes dimerization of plant isoflavones. By untargeted metabolomics, we isolated three new isoflavone dimers, namely cattleyaisoflavones A–C (**1**–**3**), from *Streptomyces cattleya* cultured in soy flour-containing media. We then identified a P450 enzyme, CYP158C1, which is able to dimerize the isoflavones as well as other plant-derived phenolic compounds. This work extends the insight on the substrate scope of P450-mediated dimerization and provides a potential biocatalytic tool for the synthesis of isoflavone dimers in future.

Results and Discussion

Characterization of three isoflavone dimers derived from daidzein (**4**)

During our effort of untargeted metabolomic screening of *Streptomyces* species, we identified several unknown isoflavone-like metabolites from *S. cattleya* NRRL8057 cultured on mannitol soya flour (MS) agar plates (Figure S2, Supporting Information File 1). Purification of the three major unknown compounds resulted in the isolation of the isoflavone dimers cattleyaisoflavones A (**1**), B (**2**), and C (**3**) (Figure 1). HRMS analyses suggested the molecular formula of both **2** and **3** to be $C_{30}H_{18}O_8$ and that of **1** to be $C_{31}H_{20}O_9$ (Figure S3, Supporting Information File 1), which matches the molecular formula of isoflavone dimers or dimeric derivatives.

To confirm whether **1**–**3** were biosynthesized de novo by the strain or derived from the media that contain isoflavones, we cultured *S. cattleya* on International *Streptomyces* Project-2 (ISP-2) liquid medium that does not contain isoflavones. These cultures did not produce **1**–**3**, but daidzein (**4**) supplementation in ISP-2 liquid medium restored the production of **1**–**3**,

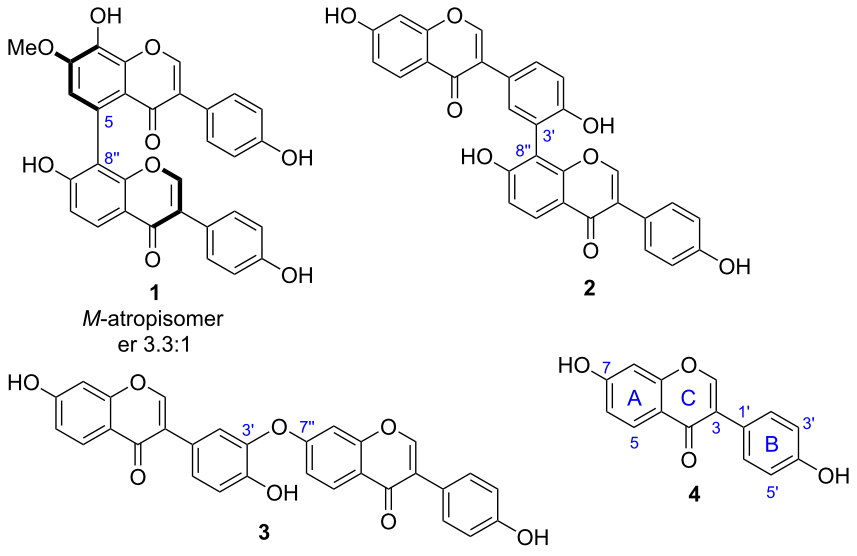


Figure 1: Structures of cattleyaisoflavones A (**1**), B (**2**), C (**3**), and daidzein (**4**).

suggesting that *S. cattleya* utilizes exogenous isoflavone to form the dimer products **1**–**3** (Figure 2a).

A comparison of the ^{1}H NMR spectra of **1** and **4** (Tables S3 and S4, Supporting Information File 1) suggested that **1** has a biaryl linkage between the two A-rings. The ^{1}H,^{1}H-COSY correlation of H-5'' (δ_H 8.06, d) and H-6'' (δ_H 7.03, d) revealed C-8'' as one of the coupling sites. The methoxy group (δ_H 3.98, s) was assigned to the C-7 site because of the ^{1}H,^{1}H-NOESY correlation with H-6 (δ_H 6.98, s). The HMBC correlations of H-6 to C-8 (δ_C 136.3), C-10 (δ_C 118.9), C-7 (δ_C 151.7), C-5 (δ_C 123.6), and C-8'' (δ_C 119.1) suggested that C-5 is the coupling site on this unit, and C-8 was substituted by a hydroxy group. Finally, the structure of **1** was determined as a dimer with a rare 5,8'' biaryl coupling skeleton. Likewise, C-8'' was assigned as one of the coupling sites of **2**. Three B-ring protons, H-2' (δ_H 7.38, d), H-5' (δ_H 7.02, d), and H-6' (δ_H 7.48, dd), suggested that C-3' was another coupling site. Thus, the structure of **2** was established as a 3',8''-coupled dimer (Table S5, Supporting Information File 1). For the dimer **3**, three hydroxy and fifteen sp^2 protons indicated a C–O bond linkage. An analysis of ^{1}H,^{1}H-COSY and HMBC spectra suggested it to be a 3'–O–7''-coupled dimer (Table S6, Supporting Information File 1).

The absolute configuration of **1** and **2** with respect to atropisomerism was assessed by chiral HPLC analysis and the comparison of experimental and calculated ECD spectra. While **1** showed a ≈3:1 atropisomer ratio, with the *M*-atropisomer being the major form, **2** did not exhibit a Cotton effect in the ECD measurement, suggesting that it is in a freely rotating state (Supporting Information File 1).

Discovery of the P450 enzyme responsible for dimerization

We next investigated which enzyme in *S. cattleya* is responsible for dimerization of **4**. The genome of *S. cattleya*

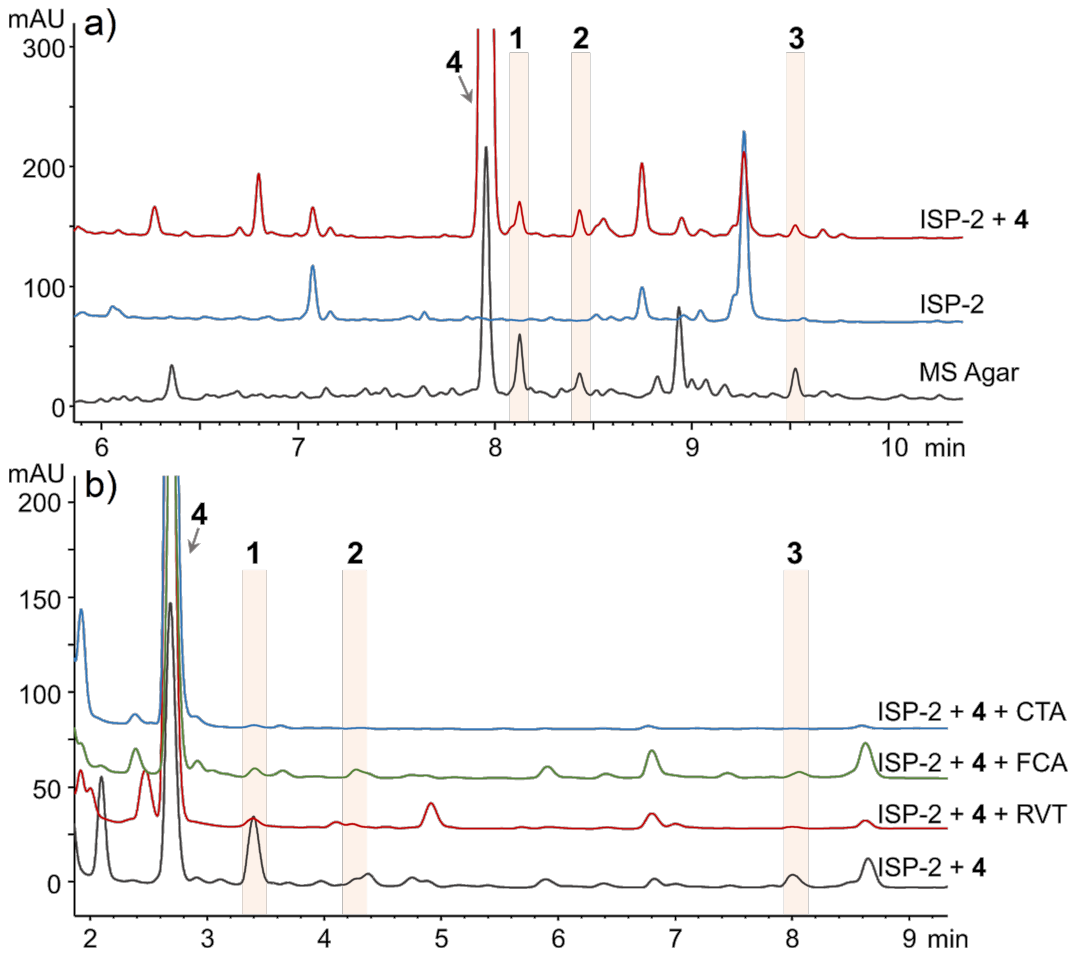


Figure 2: a) The culture in ISP-2 liquid did not produce **1**–**3**, while feeding with **4** restored the production (analytical method A according to Experimental section, 254 nm). b) Chemical genetics experiments indicated three P450 inhibitors could reduce the production of dimers derived from **4** (analytical method B to better analyze the isoflavone region, 254 nm). CTA: clotrimazole (2.5 μM), FCA: fluconazole (1000 μM), RVT: resveratrol (100 μM).

(GCF_000237305.1) did not encode any laccase homologs but encoded 41 P450 enzymes in total. To investigate whether P450 enzymes mediate this dimerization reaction, we first employed a chemical genetics screening using three P450 inhibitors, CTA (strong inhibitor), FCA (moderate inhibitor), and RVT (polyphenol inhibitor) [23,24]. All three agents reduced the production of cattleyaisoflavones in a concentration-dependent manner (Figure 2b and Figure S4, Supporting Information File 1), suggesting that P450 enzymes play a key role in isoflavone dimerization.

The known bacterial P450 enzymes involved in biaryl coupling can be classified into four types, including CYP158 enzymes for dimerization of naphthols [25-27], JulI/SetI/SptI clade for anthraquinones [3], HmtS/ClpS/LtzS clade for cyclopeptides [4,18,19], and Bmp7 clade for polybrominated substrates (Table S7 and Figure S1, Supporting Information File 1) [13,17]. These P450 enzymes tend to have broad substrate specificity and form a variety of biaryl products. Thus, we constructed a phylogenetic tree using the 41 P450 enzymes in *S. cattleya*, together with these reported dimerization P450 enzymes (Figure S5, Supporting Information File 1). Notably, a P450 (WP_014145731.1, here termed CYP158C1) was clustered into the CYP158 clade that reportedly catalyzes dimerization of type III polyketide synthase (T3PKS) products, such as naphthols. Considering the similar biosynthetic pathway of isoflavones shared, this enzyme was expressed in *E. coli* and purified for in vitro biochemical assay together with four other P450 enzymes (CYP1–4) that were successfully expressed out of 10 P450 enzymes from each representative clade (Figures S5 and S6, Supporting Information File 1). The results showed that only CYP158C1 was able to dimerize **4** (Figure 3 and Figure S7, Supporting Information File 1), indicating that CYP158C1 is the most likely enzyme for isoflavone dimerization. Interestingly, different from the in vivo extract, in vitro experiments gave a major dimer **5**, a 3',3' coupled dimer assigned by MS–MS (Figures S8 and S14, Supporting Information File 1). Although dimer **3** was not visible in the HPLC analysis, a trace amount of **3** was detected by HRMS analysis (Figure S8, Supporting Information File 1). Redox partners might be responsible for the different regioselectivity between in vivo and in vitro reactions, as it has been reported that different reductases would

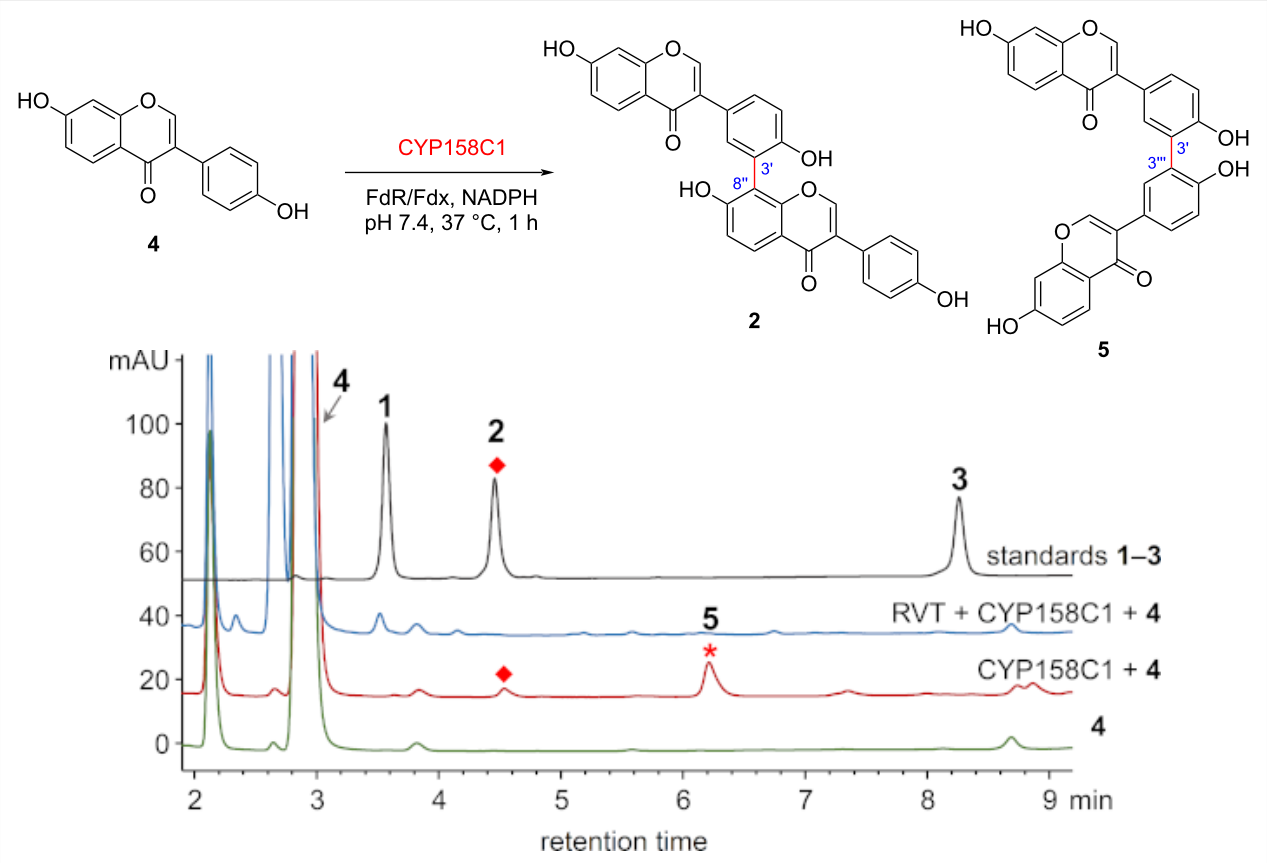


Figure 3: CYP158C1 dimerizes **4** to form dimers **2** and **5** in vitro (analytical method B, 254 nm). Control conditions (green): **4** (0.4 mM), NADPH (1 mM), and ferredoxin–NADP$^+$ reductase *Sel*FdR0978/ferredoxin *Sel*Fdx1499 (FdR/Fdx, 5 μM) in buffer (50 mM PBS, 0.2 mM DTT, 1 mM EDTA, pH 7.4) were incubated at 37 °C for 1 h. Red: the conditions above with the addition of CYP158C1 (3 μM). Blue: the conditions above with the addition of CYP158C1 (3 μM) and RVT (0.5 mM).

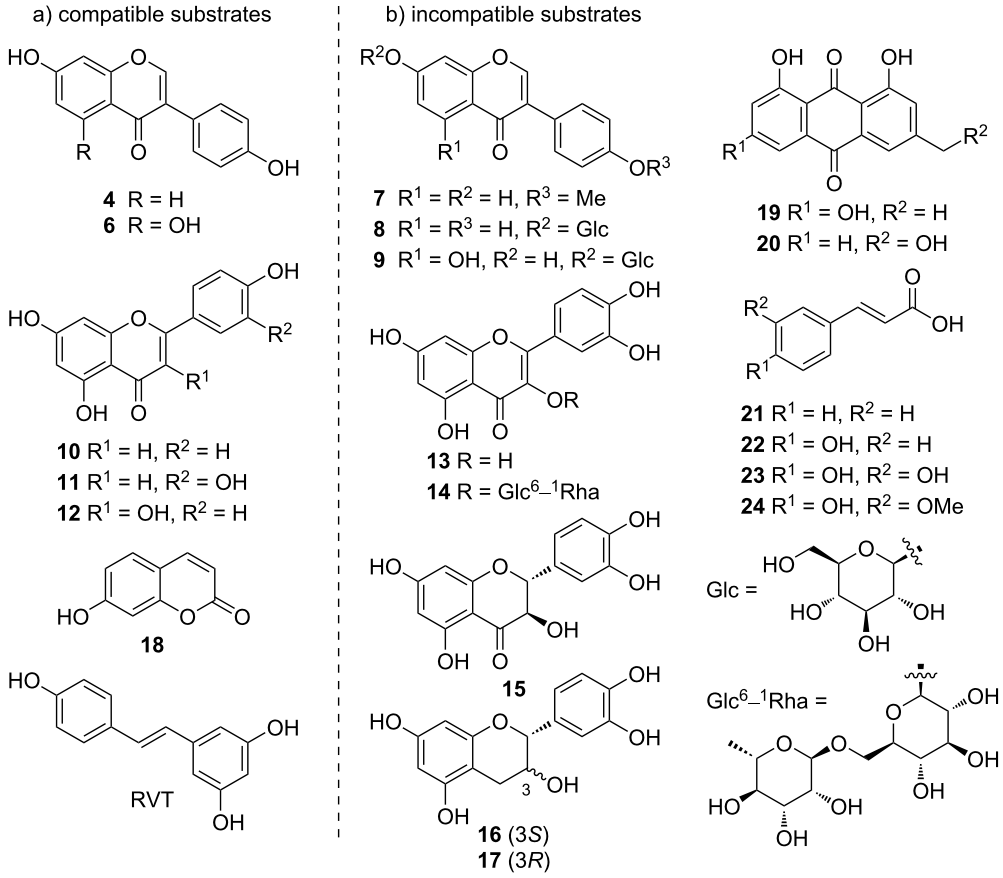
lead to different diastereomeric dimers of coclaurine [28]. The absence of the product **1** in vitro is reasonable because the biosynthesis of **1** requires an additional methyltransferase from *S. cattleya* for the methoxy group formation.

Substrate scope and reaction mechanism

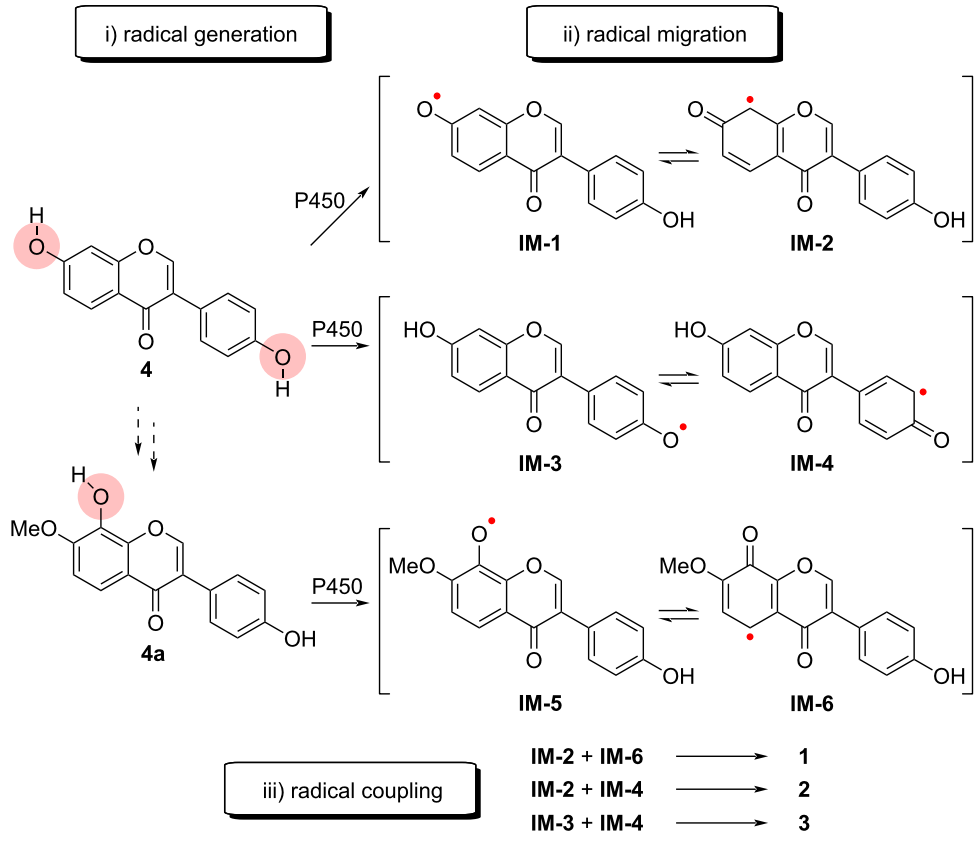
After identification of CYP158C1 as the isoflavone dimerization enzyme, we next explored the substrate scope of the catalyst using a variety of plant-derived polyketides (Scheme 1). The results revealed that the 5-OH substitution on isoflavone (see compound **6**) was tolerable for dimerization (Figure S15, Supporting Information File 1), whereas the additional 4'-methoxyl group at the B-ring (see **7**) prevented the dimerization. Two isoflavone glucosides, **8** and **9**, failed to be transformed, presumably due to the excessive molecular size. When using the other flavonoids **10**–**17** as substrates, we observed the corresponding dimeric products for the substrates **10**–**12** (Figures S16–S18, Supporting Information File 1) but not for **13**–**17**, suggesting that CYP158C1 is also compatible with compounds with flavone scaffold. Apart from flavonoids, this enzyme also dimerized umbelliferone (**18**), a substrate of the fungal P450 KtnC [14,18], and RVT with relatively low efficiency (Figures S19 and S20, Supporting Information File 1). Other plant polyketides, such as anthraquinones **19** and **20** and phenylpropanoids **21**–**24**, failed to be dimerized.

The reaction mechanism of P450-mediated phenol dimerization is believed to involve oxidative radical–radical coupling, though other mechanisms, such as radical addition, radical cation addition, and electrophilic aromatic addition, have also been proposed [1,10,29]. A proposed mechanism is depicted in Scheme 2: First, the hydroxy group on the A- or B-ring is converted into a radical by a P450-induced single-electron transformation. The resulting radical then migrates to the π-system and is stabilized in the *ortho*- and *para*-positions, generating diverse carbon radical intermediates. As a result, various dimers are formed via promiscuous coupling of these radical intermediates.

Although we cannot exclude the possibility that other P450 enzymes in *S. cattleya* may also catalyze isoflavone dimerization, the above results indicated that CYP158C1 dimerizes



Scheme 1: a) Compatible and b) incompatible substrates of CYP158C1. Products were identified using analytical method D.



Scheme 2: Proposed mechanism of CYP158C1-mediated dimerization of isoflavones.

isoflavones and several plant-derived aromatic polyketides, albeit with low catalytic efficiency. Future optimization of the reductase partner or reaction conditions may be required to improve the enzyme activity. Interestingly, we found that CYP158C1 clusters together with a T3PKS gene in the *S. cattleya* genome, which is similar to the biflaviolin [30] and naringenin [31] biosynthetic gene clusters. The native function of these type-III polyketide synthase products is believed to be involved in the protection of the host against oxidative radicals generated by UV irradiation [25,32]. To verify the antioxidative effect of the isoflavone dimers, 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS)-based antioxidant activity was performed [33]. The results showed that while **2** and **3** had an activity roughly equal to the monomers **4** and **6**, **1** displayed a roughly twofold radical scavenging capacity (Figure S9, Supporting Information File 1). This is in agreement with the previous report that the configuration of the hydroxy group of the B-ring plays a key role in the antioxidant activity [34]. Thus, CYP158C1 might provide an alternative pathway for antioxidant biosynthesis by using exogenous metabolites as substrates. The native substrate for CYP158C1, presumably produced by the clustered T3PKS, has yet to be identified.

The broad substrate specificity of CYP158C1 towards plant polyphenols suggests a potential host–microbial interaction in the rhizosphere. Flavones and isoflavones are common molecules in plant roots and can be secreted in soil, playing a role in allelopathic interference [35,36]. To our knowledge, this is the first bacterial P450 that can dimerize plant flavonoids. Since dimerization of polyphenols in plants is mostly catalyzed by laccases, the dimerization of (iso)flavones by a P450 in *Streptomyces* may represent a convergent evolution in enzyme chemistry between plants and bacteria.

Conclusion

In conclusion, we have characterized three novel isoflavone dimers from *S. cattleya* and identified the P450 enzyme CYP158C1 to be able to catalyze biaryl coupling reactions using plant-derived daidzein (**4**) as a substrate. The enzyme can dimerize various flavonoids and provides a potential synthetic tool for the dimerization of aromatic polyketides.

Experimental

General experimental procedures

NMR spectra were acquired on Bruker 500 and 600 MHz AVANCE NEO spectrometers with TMS as an internal stan-

dard. HPLC–UV analysis was performed on an Agilent 1260 Infinity II HPLC system. UHPLC–HRMS analysis was performed on a Waters Synapt-G2-Si UHPLC-ESI-QToF-MS system. UV and CD spectra were acquired under N_2 gas by a Chirascan V100 spectrometer (Applied Photophysics). Strain cultivation was performed at 28 °C in an incubator for agar plates or in a rotary shaker at 200 rpm for liquid media.

Strain preparation and fermentation

The strain *S. cattleya* NRRL 8057 was purchased from the China General Microbiological Culture Collection Center (CGMCC). Seed culture of this strain was done in ISP-2 medium (yeast extract 0.4 g, malt extract 1 g, glucose 0.4 g, distilled water up to 100 mL, pH 7.2) for 2 days. Isoflavone dimer production was performed on MS agar medium (mannitol 20 g, soy flour 20 g, agar 20 g, distilled water up to 1 L) for 15 days from the above seed culture.

Analytical methods

All HPLC analyses were performed using (A) H_2O with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid at 40 °C column temperature unless otherwise noted. The HPLC–UV analysis of feeding experiments was conducted with a YMC-Triart C18 ExRS column (150 mm × 3.0 mm, 3 µm, 12 nm, YMC) with a gradient of 5% B (0 min) to 95% B (12 min) at a flow rate of 0.5 mL/min (analytical method A). The HPLC–UV analysis of isoflavones was conducted with a Poroshell 120 SB-C18 column (100 mm × 3.0 mm, 2.7 µm, Agilent) with a gradient of 30% B (0 min) to 45% B (10 min) at a flow rate of 0.5 mL/min (analytical method B). The UHPLC–HRMS analysis of isoflavones was conducted with a HSS T3 column (100 mm × 2.1 mm, 1.8 µm, Waters) with a gradient of 30% B (0 min) to 45% B (4.5 min) at a flow rate of 0.4 mL/min (analytical method C). The UHPLC–HRMS analysis of the substrate scope was conducted with a HSS T3 column (100 mm × 2.1 mm, 1.8 µm, Waters) with a gradient of 0% B (0 min) to 100% B (5.0 min) at a flow rate of 0.4 mL/min (analytical method D). All HRMS analyses were performed in positive model with a scan range of *m*/*z* 50–1600. For MS–MS analysis, the collision-induced dissociation (CID) energy was set to 15–40 eV or 30–50 eV depending on the compounds. The HPLC–UV analysis of the chiral separation was conducted with a CHIRALCEL OX-3R column (150 mm × 4.6 mm, 3 µm, DAICEL) with isocratic 35% B at a flow rate of 1.0 mL/min at 30 °C column temperature, and an Agilent G6215B MSD detector was used for peak assignment (analytical method E).

Isolation of compounds

To isolate compounds **1**–**3**, the fermentation culture of *S. cattleya* on MS agar medium (125 plates, 80 mL/plate each) was collected and extracted using organic solvent ($CHCl_3$/MeOH 1:1, v/v, 20 L in total). After removing the solvent in vacuo, the crude extract was redissolved in methanol and separated by Sephdex LH-20 (GE healthcare) column chromatography using methanol as eluent. The yellow band at the elution tail on the LH-20 column (Figure S2, Supporting Information File 1) was collected and further purified by a Shimadzu LC-20AD semipreparative HPLC by using (A) H_2O with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid. Compounds **1** (5.0 mg), **2** (2.2 mg), and **3** (2.0 mg) were isolated using a YMC-Triart C18 column (250 mm × 10 mm, 5 µm, YMC) at 4 mL/min with a gradient of 40% B (0 min) to 60% B (12 min). A second round of purification was then performed for **1**–**3** with isocratic 35%, 38%, and 45% B, respectively, with a YMC-Triart Phenyl column (250 mm × 10 mm, 5 µm, YMC) at 4 mL/min.

Cattleyaisoflavone A (1): yellowish powder; UV (MeOH) λ_{max} (log ε) 257, 304 nm; for NMR spectral data, see Table S4, Supporting Information File 1; HRMS–ESI (*m*/*z*): $[M + H]^+$ calcd for $C_{31}H_{20}O_9$, 537.1180; found, 537.1206.

Cattleyaisoflavone B (2): yellowish powder; UV (MeOH) λ_{max} 255, 310 nm; for NMR spectral data, see Table S5, Supporting Information File 1; HRMS–ESI (*m*/*z*): $[M + H]^+$ calcd for $C_{30}H_{18}O_8$, 507.1074; found, 507.1091.

Cattleyaisoflavone C (3): white powder; UV (MeOH) λ_{max} 254, 300 nm; for NMR spectral data, see Table S6, Supporting Information File 1; HRMS–ESI (*m*/*z*): $[M + NH^+$ calcd for $C_{30}H_{18}O_8$, 507.1074; found, 507.1091.

Feeding experiments

Feeding experiments were performed in 500 mL flasks each containing 100 mL of ISP-2 liquid medium and 10 mg daidzein (**4**, as 1 mL DMSO solution added after autoclave) by 1% inoculation from the above seed culture. On the 5th day, 2 mL of medium from each sample was taken and freeze-dried. Then, 1 mL MeOH was added and the mixture was treated with ultrasound for 15 min. After centrifugation, the supernatant was directly analyzed by HPLC following analytical method A.

Chemical genetics experiments

All samples involved in inhibition experiments were set up using 100 mL flasks containing 20 mL ISP-2 medium, 0.02 mg/mL daidzein (**4**), and the P450 inhibitor, by 1% inoculation from the above seed culture. On the 9th day, 0.5 mL of medium from each sample was taken and freeze-dried. Then, 0.2 mL MeOH was added, and the mixture was treated with ultrasound for 15 min. After centrifugation, the supernatant was detected following analytical method B.

Bioinformatic analysis

Multiple sequence alignment was constructed by MAFFT 7.450 using the E-INS-I method [37], and the phylogenetic tree was constructed by FastTree 2.1.11 [38] with default parameters by using Geneious Prime 2022.1.1 (see https://www.geneious.com).

Protein purification and biochemical assays

The *Streptomyces* genome was extracted by TIANamp Bacteria DNA Kit (TIANGEN BIOTECH), and the P450 genes were amplified by PCR (Table S1, Supporting Information File 1) and cloned into a pET-28a(+) vector by the T5 exonuclease DNA assembly method [39]. The resulting expression vectors were individually introduced into *E. coli* BL21(DE3) for protein expression. Cells harboring expression plasmid were inoculated into LB medium with 50 μg/mL kanamycin and incubated at 37 °C until the culture reached an OD_{600} of 0.4–0.6. The cultures were then cooled to 16 °C and induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 h. Cells were harvested by centrifugation (10 min, 8,000 rpm, 4 °C).

For protein purification, cell pellets were resuspended in lysis buffer (50 mM NaH_2PO_4, 400 mM NaCl, 1 mM TCEP, pH 7.5) and lysed by sonication. After removing cell debris by centrifugation, the lysate supernatant was purified by Ni-NTA resin with washing buffer (50 mM NaH_2PO_4, 150 mM NaCl, 100 mM imidazole, pH 7.5) and elution buffer (50 mM NaH_2PO_4, 150 mM NaCl, 300 mM imidazole, pH 7.5) and then dialyzed into the storage buffer (50 mM NaH_2PO_4, 150 mM NaCl, 1 mM TCEP, 10% glycerol, pH 7.5) and stored at −80 °C for further use. The P450 partner proteins, FdR and Fdx, were constructed according to the reported method [40].

For in vitro biochemical assays, the reaction mixture (50 μL) contained 0.4 mM substrate, 3 μM P450 enzyme, 5 μM FdR, 5 μM Fdx, 1 mM NADPH, 0.2 mM DTT, and 1 mM EDTA in 50 mM PBS buffer (pH 7.4). The reaction mixture was incubated at 37 °C for 1 h and quenched by adding 50 μL MeOH. After centrifugation (10 min, 13,000 rpm), the mixture was subjected to HPLC–UV or UHPLC–HRMS analysis.

Antioxidant activity

The antioxidant activity was assayed using a Total Antioxidant Capacity Assay Kit with ABTS method (Beyotime Biotechnology). Briefly, the fresh ABTS working solution was prepared by mixing ABTS stock solution with oxidant solution for an overnight reaction, and then this mixture was diluted 50 times by 80% ethanol. To initiate the assay reaction, 10 μL of each sample with a concentration of 1.0 mM in 80% ethanol was mixed with 200 μL of fresh ABTS working solution. After being incubated at 30 °C for 5 min, the 734 nm absorbance was measured on a Thermo Varioskan LUX Microplate reader. Trolox was used as the reference compound for calculation of ABTS-reducing activity (in %).

Computational ECD calculation of compound 1

Merck molecular force field (MMFF) and DFT as well as TDDFT calculations were carried out with the Spartan 14 software (Wavefunction Inc.) and the Gaussian 16 program [41], respectively. Conformers within the 2 kJ/mol energy window were generated and optimized using DFT calculations at the B3LYP/6-31+G (d,p) level. Frequency calculations were performed at the same level to confirm that each optimized conformer was true minimum and to estimate the relative thermal free energy (ΔG) at 298.15 K. The two conformers were chosen for ECD calculations in methanol at the B3LYP/6-311+G (d,p) level. Solvent effects were taken into consideration using the self-consistent reaction field (SCRF) method with the polarizable continuum model (PCM). The ECD spectrum was generated by the SpecDis program [42] using a Gaussian band shape with 0.23 eV exponential half-width from dipole length rotational strengths.

Supporting Information

Supporting Information File 1

Detailed descriptions of the experimental procedures and comprehensive analytical data.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-113-S1.pdf]

Acknowledgements

We thank Dr. Xingyu Lu (Westlake University) for his help in measuring NMR spectra and the Instrumentation and Service Center for Molecular Sciences at Westlake University for the assistance in HRMS and optical analyses.

Funding

This work was supported by the Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (2020R01004) to L. Z.

ORCID® iDs

Lihan Zhang - https://orcid.org/0000-0003-4364-6242

References

1. Liu, J.; Liu, A.; Hu, Y. *Nat. Prod. Rep.* **2021,** *38,* 1469–1505. doi:10.1039/d0np00063a

2. Gigant, B.; Wang, C.; Ravelli, R. B. G.; Roussi, F.; Steinmetz, M. O.; Curmi, P. A.; Sobel, A.; Knossow, M. *Nature* **2005,** *435,* 519–522. doi:10.1038/nature03566
3. Präg, A.; Grüning, B. A.; Häckh, M.; Lüdeke, S.; Wilde, M.; Luzhetskyy, A.; Richter, M.; Luzhetska, M.; Günther, S.; Müller, M. *J. Am. Chem. Soc.* **2014,** *136,* 6195–6198. doi:10.1021/ja501630w
4. Ma, J.; Wang, Z.; Huang, H.; Luo, M.; Zuo, D.; Wang, B.; Sun, A.; Cheng, Y.-Q.; Zhang, C.; Ju, J. *Angew. Chem., Int. Ed.* **2011,** *50,* 7797–7802. doi:10.1002/anie.201102305
5. Liu, X.-G.; Lu, X.; Gao, W.; Li, P.; Yang, H. *Nat. Prod. Rep.* **2022,** *39,* 474–511. doi:10.1039/d1np00026h
6. Yin, L.; Liebscher, J. *Chem. Rev.* **2007,** *107,* 133–173. doi:10.1021/cr0505674
7. Rahman, M.; Riaz, M.; Desai, U. R. *Chem. Biodiversity* **2007,** *4,* 2495–2527. doi:10.1002/cbdv.200790205
8. Selepe, M. A.; Van Heerden, F. R. *Molecules* **2013,** *18,* 4739–4765. doi:10.3390/molecules18044739
9. Shalit, H.; Dyadyuk, A.; Pappo, D. *J. Org. Chem.* **2019,** *84,* 1677–1686. doi:10.1021/acs.joc.8b03084
10. Hüttel, W.; Müller, M. *Nat. Prod. Rep.* **2021,** *38,* 1011–1043. doi:10.1039/d0np00010h
11. Hu, J.; Li, H.; Chooi, Y.-H. *J. Am. Chem. Soc.* **2019,** *141,* 8068–8072. doi:10.1021/jacs.9b03354
12. Obermaier, S.; Thiele, W.; Fürtges, L.; Müller, M. *Angew. Chem., Int. Ed.* **2019,** *58,* 9125–9128. doi:10.1002/anie.201903759
13. 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
14. Mazzaferro, L. S.; Hüttel, W.; Fries, A.; Müller, M. *J. Am. Chem. Soc.* **2015,** *137,* 12289–12295. doi:10.1021/jacs.5b06776
15. Kudanga, T.; Nemadziva, B.; Le Roes-Hill, M. *Appl. Microbiol. Biotechnol.* **2017,** *101,* 13–33. doi:10.1007/s00253-016-7987-5
16. Fürtges, L.; Obermaier, S.; Thiele, W.; Foegen, S.; Müller, M. *ChemBioChem* **2019,** *20,* 1928–1932. doi:10.1002/cbic.201900041
17. Agarwal, V.; Moore, B. S. *ACS Chem. Biol.* **2014,** *9,* 1980–1984. doi:10.1021/cb5004338
18. 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
19. Guo, Z.; Li, P.; Chen, G.; Li, C.; Cao, Z.; Zhang, Y.; Ren, J.; Xiang, H.; Lin, S.; Ju, J.; Chen, Y. *J. Am. Chem. Soc.* **2018,** *140,* 18009–18015. doi:10.1021/jacs.8b10136
20. Ferreira, D.; Bekker, R. *Nat. Prod. Rep.* **1996,** *13,* 411–433. doi:10.1039/np9961300411
21. Sohn, S. I.; Pandian, S.; Oh, Y. J.; Kang, H. J.; Cho, W. S.; Cho, Y. S. *Front. Plant Sci.* **2021,** *12,* 670103. doi:10.3389/fpls.2021.670103
22. Al-Maharik, N. *Nat. Prod. Rep.* **2019,** *36,* 1156–1195. doi:10.1039/c8np00069g
23. Lewis, D. F. V.; Wiseman, A. *Enzyme Microb. Technol.* **2005,** *36,* 377–384. doi:10.1016/j.enzmictec.2004.07.018
24. Chun, Y. J.; Kim, M. Y.; Guengerich, F. P. *Biochem. Biophys. Res. Commun.* **1999,** *262,* 20–24. doi:10.1006/bbrc.1999.1152
25. Funa, N.; Funabashi, M.; Ohnishi, Y.; Horinouchi, S. *J. Bacteriol.* **2005,** *187,* 8149–8155. doi:10.1128/jb.187.23.8149-8155.2005
26. Zhao, B.; Guengerich, F. P.; Voehler, M.; Waterman, M. R. *J. Biol. Chem.* **2005,** *280,* 42188–42197. doi:10.1074/jbc.m509220200
27. Zhao, B.; Lamb, D. C.; Lei, L.; Kelly, S. L.; Yuan, H.; Hachey, D. L.; Waterman, M. R. *Biochemistry* **2007,** *46,* 8725–8733. doi:10.1021/bi7006959
28. Kraus, P. F.; Kutchan, T. M. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92,* 2071–2075. doi:10.1073/pnas.92.6.2071
29. Tang, M.-C.; Zou, Y.; Watanabe, K.; Walsh, C. T.; Tang, Y. *Chem. Rev.* **2017,** *117,* 5226–5333. doi:10.1021/acs.chemrev.6b00478
30. Funa, N.; Ohnishi, Y.; Fujii, I.; Shibuya, M.; Ebizuka, Y.; Horinouchi, S. *Nature* **1999,** *400,* 897–899. doi:10.1038/23748
31. Álvarez-Álvarez, R.; Botas, A.; Albillos, S. M.; Rumbero, A.; Martín, J. F.; Liras, P. *Microb. Cell Fact.* **2015,** *14,* 178. doi:10.1186/s12934-015-0373-7
32. Lim, Y.-R.; Han, S.; Kim, J.-H.; Park, H.-G.; Lee, G.-Y.; Le, T.-K.; Yun, C.-H.; Kim, D. *Biomol. Ther.* **2017,** *25,* 171–176. doi:10.4062/biomolther.2016.182
33. Ungar, Y.; Osundahunsi, O. F.; Shimoni, E. *J. Agric. Food Chem.* **2003,** *51,* 4394–4399. doi:10.1021/jf034021z
34. Heim, K. E.; Tagliaferro, A. R.; Bobilya, D. J. *J. Nutr. Biochem.* **2002,** *13,* 572–584. doi:10.1016/s0955-2863(02)00208-5
35. Weston, L. A.; Mathesius, U. *J. Chem. Ecol.* **2013,** *39,* 283–297. doi:10.1007/s10886-013-0248-5
36. Sugiyama, A. *J. Adv. Res.* **2019,** *19,* 67–73. doi:10.1016/j.jare.2019.03.005
37. Katoh, K.; Standley, D. M. *Mol. Biol. Evol.* **2013,** *30,* 772–780. doi:10.1093/molbev/mst010
38. Price, M. N.; Dehal, P. S.; Arkin, A. P. *PLoS One* **2010,** *5,* e9490. doi:10.1371/journal.pone.0009490
39. Xia, Y.; Li, K.; Li, J.; Wang, T.; Gu, L.; Xun, L. *Nucleic Acids Res.* **2019,** *47,* e15. doi:10.1093/nar/gky1169
40. Zhang, W.; Du, L.; Li, F.; Zhang, X.; Qu, Z.; Han, L.; Li, Z.; Sun, J.; Qi, F.; Yao, Q.; Sun, Y.; Geng, C.; Li, S. *ACS Catal.* **2018,** *8,* 9992–10003. doi:10.1021/acscatal.8b02913
41. *Gaussian 16,* Rev. A.03; Gaussian, Inc.: Wallingford, CT, 2016.
42. Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Bringmann, G. *Chirality* **2013,** *25,* 243–249. doi:10.1002/chir.22138

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Make or break: the thermodynamic equilibrium of polyphosphate kinase-catalysed reactions

Michael Keppler[1], Sandra Moser[2], Henning J. Jessen[2], Christoph Held[3] and Jennifer N. Andexer[*1]

Full Research Paper

Open Access

Address:
[1]Institute of Pharmaceutical Sciences, University of Freiburg, Albertstr. 25, 79104 Freiburg, Germany, [2]Institute of Organic Chemistry, University of Freiburg, Albertstr. 21, 79104 Freiburg, Germany and [3]Department of Biochemical and Chemical Engineering, TU Dortmund University, Emil-Figge-Str. 70, 44227 Dortmund, Germany

Email:
Jennifer N. Andexer[*] - jennifer.andexer@pharmazie.uni-freiburg.de

* Corresponding author

Keywords:
ATP regeneration; biocatalyst; ePC-SAFT; polyp; PPK

Beilstein J. Org. Chem. **2022,** *18,* 1278–1288.
https://doi.org/10.3762/bjoc.18.134

Received: 19 May 2022
Accepted: 30 August 2022
Published: 20 September 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

Polyphosphate kinases (PPKs) have become popular biocatalysts for nucleotide 5'-triphosphate (NTP) synthesis and regeneration. Two unrelated families are described: PPK1 and PPK2. They are structurally unrelated and use different catalytic mechanisms. PPK1 enzymes prefer the usage of adenosine 5'-triphosphate (ATP) for polyphosphate (polyP) synthesis while PPK2 enzymes favour the reverse reaction. With the emerging use of PPK enzymes in biosynthesis, a deeper understanding of the enzymes and their thermodynamic reaction course is of need, especially in comparison to other kinases. Here, we tested four PPKs from different organisms under the same conditions without any coupling reactions. In comparison to other kinases using phosphate donors with comparably higher phosphate transfer potentials that are characterised by reaction yields close to full conversion, the PPK-catalysed reaction reaches an equilibrium in which about 30% ADP is left. These results were obtained for PPK1 and PPK2 enzymes, and are supported by theoretical data on the basic reaction. At high concentrations of substrate, the different kinetic preferences of PPK1 and PPK2 can be observed. The implications of these results for the application of PPKs in chemical synthesis and as enzymes for ATP regeneration systems are discussed.

Introduction

Polyphosphate (polyP, Figure 1) is a linear polymer of up to thousands of phosphate residues connected by phosphate anhydride bonds. It serves as a phosphate storage molecule and plays a crucial role in biofilm formation and stress responses of cells [1]. So far polyP has been detected in every living organism investigated [1-3]. In 1956, Kornberg described the first polyP kinase (PPK) in *Escherichia coli* catalysing adenosine 5'-triphosphate (ATP)-dependent synthesis of polyP (Figure 2a)

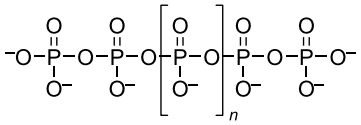


Figure 1: Polyphosphate, a ubiquitous phosphate storage molecule. Reported chain lengths range from three to several thousands.

[4]. The enzyme was reclassified as family-1 PPK (PPK1) when a structurally different PPK (family-2, PPK2) was found in *Pseudomonas aeruginosa* in 2002 [5]. PPK2 were later subdivided into three classes: PPK2-I, PPK2-II, and PPK2-III phosphorylating nucleotide diphosphates (NDPs), nucleotide monophosphates (NMPs), and both, respectively [6]. Nevertheless, these substrate profiles rather seem to be preferences, as most enzymes catalyse all phosphorylation steps during extended reaction times; also higher phosphorylated species have been detected in the reactions [7,8]. The enzymes characterised from *E. coli* (*Ec*PPK1) and *Sinorhizobium meliloti* (renamed *Ensifer meliloti, Sm*PPK2) are often regarded as model enzymes for PPK1 and PPK2 [9,10]. From a structure perspective, PPK1 enzymes form tetramers in solution with a mass of approximately 80 kDa for the monomer (Figure 2b). Although not being an integral membrane protein, the enzyme is described to be membrane-associated [11-13]. The phosphate transfer likely proceeds via formation of a phospho-enzyme intermediate

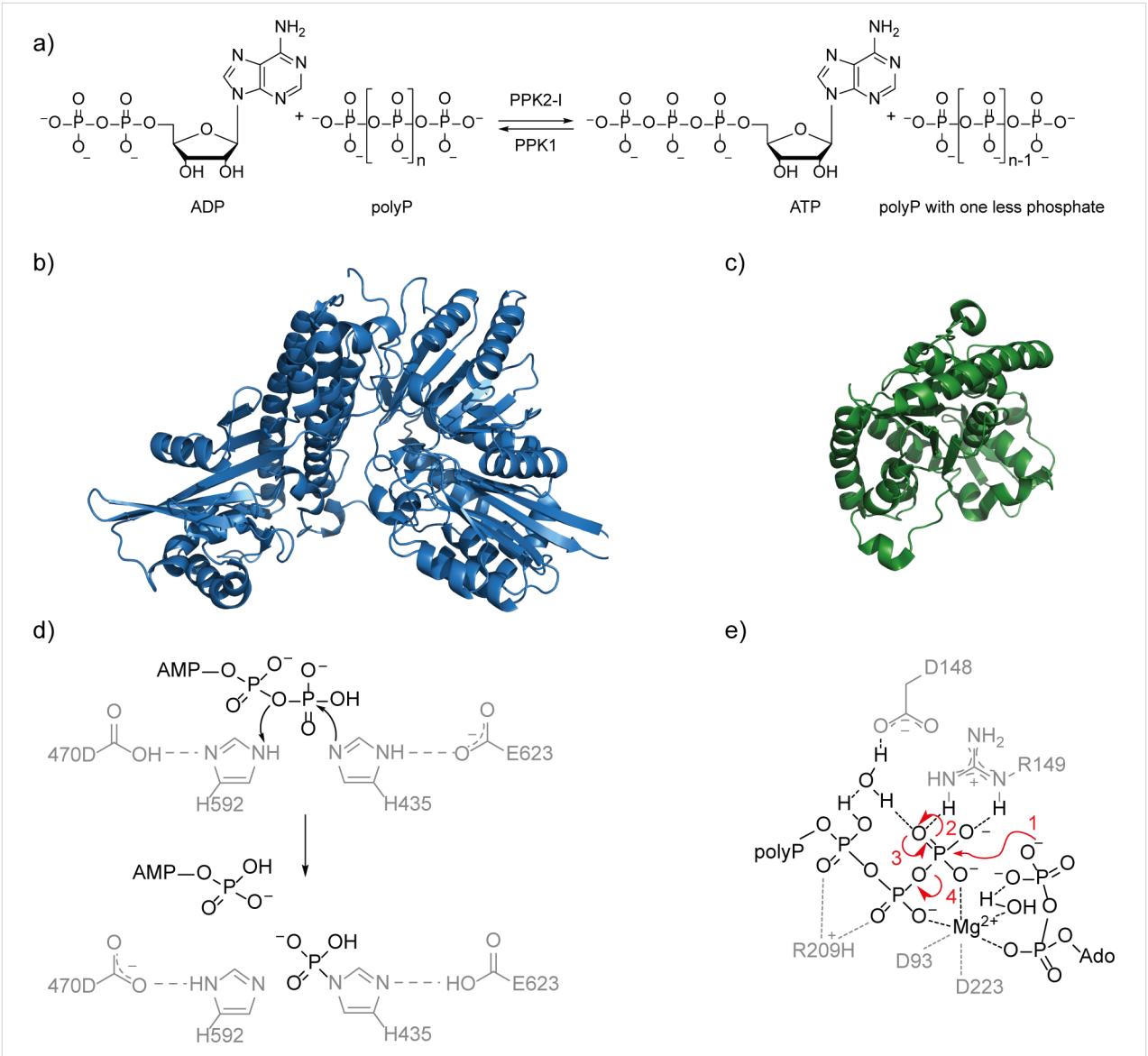


Figure 2: Comparison of PPK1 and PPK2 enzymes. a) Reaction scheme; b) structure of the *Ec*PPK1 monomer (PDB 1XDO) [13]; c) structure of the *Sm*PPK2 monomer (PDB 3CZQ) [10]; d) proposed mechanism of PPK1 with phosphate transfer via a phosphor enzyme intermediate; e) PPK2 mechanism exemplarily shown for the associative reaction pathway.

(Figure 2d). Two essential histidine residues for autophosphorylation were identified by mutagenesis experiments [9,13,14]. Variants carrying mutations at these histidine residues lost the ability to synthesise polyP or ATP in vitro, clearly demonstrating the necessity of the residues for catalysis [14]. PPK2-I enzymes are of lower molecular weight than their PPK1 counterparts, with an approximate molecular mass of 40 kDa for a monomer (Figure 2c) [5]. They form dimers or tetramers in solution and are not purified from membrane fractions [5,10,15,16]. Based on the crystal structures of three PPK2-III, the coordination of polyP and ADP by positively charged amino acids (lysine and arginine) has been suggested [16,17]. Two magnesium ions are held in place by two conserved aspartate residues that further coordinate the polyP and ADP for an in-line reaction of these two substrates. Out of this arrangement two reaction pathways have been discussed, an associative and a dissociative one. The associative one is an S_N2-like attack of ADP on the terminal phosphate of the polyP chain, while the dissociative one is an S_N1-like reaction where the terminal phosphate dissociates from the polyP chain before being attacked by the nucleotide [18]. Both mechanisms could proceed without a phosphate group transfer onto an amino acid side chain of the enzyme as in PPK1: here, the enzyme structure generates proximity and polarisation of substrate bonds (Figure 2e) [17]. Apart from the structures, the kinetic preference of either polyP synthesis (NTP usage) or NTP synthesis (polyP degradation) has been described to be a characteristic feature of PPK1 and PPK2(-I), respectively (Figure 2a). This is supported by analysis of the kinetic parameters K_M and v_{max} of selected enzymes (Table S6, Supporting Information File 1) [5,10,11,15,19-21]. A sequence-based classification of PPKs is in most cases straightforward and unambiguous. Nevertheless, there seem to be exceptions: regarding the amino acid sequence, the PPK1 from *Vibrio cholerae* is very similar to the one from *E. coli* with 82% similarity (64% identity) on the amino acid level; however, it was described to show kinetic preferences of a PPK2 [20]. While the PPK2 from *P. aeruginosa* catalyses both synthesis and usage of ATP with kinetic preference for ATP synthesis, the "model PPK2" from *S. meliloti* (*Sm*PPK2) only tested positively for ATP synthesis [5,10]. The PPK2 from *Corynebacterium glutamicum* (*Cg*PPK2) has kinetic preferences of a PPK1 although being a PPK2 regarding the amino acid sequence [15].

The biocatalytic activity of PPKs can be used as a tool for the regeneration of ATP (and other NTPs, Figure 3a) as well as for the biocatalytic production of modified NTPs (Figure 3b) [22,23]. Compared to other ATP regeneration systems using phosphate donors such as phosphoenolpyruvate, carbamoyl phosphate and acetyl phosphate, PPK catalysed reactions benefit from their stable and inexpensive phosphate donor polyP [24]. Besides the difference among PPK families, further process parameters determine the kinetic preference towards ATP synthesis or utilisation. Especially for synthetic reactions with the aim to produce and isolate phosphorylated product, the initial substrate/product ratio is an important parameter for kinetics and for the reaction equilibrium, as it defines how much conversion will be achieved [25]. For acetate kinase a conversion of at least 90% using stoichiometric amounts of ADP and acetylphosphate was reported [26]. The reaction of pyruvate kinase (phosphoenolpyruvate as phosphate donor) is strongly favouring ATP synthesis both in vivo and in vitro, this reaction was originally considered to be irreversible in cells and a point of flux control. Newer findings showed the reaction to be actually an equilibrium, although positioned far on the product side [27-29]. Also for carbamate kinase (using carbamoyl phosphate), the equilibrium lies far on the ATP side with a calculated equilibrium concentration of 3.9×10^{-4} M ADP out of 0.1 M ADP [30]. As most kinases use a phosphate donor with a high phosphate transfer potential (see Figure 3c) and a product which has to be lower in its phosphate transfer potential, the overall outcome of the reactions is expected [28,29,31,32]. The phosphate transfer potential is a measure of tendency of a molecule to transfer a phosphate group onto an acceptor molecule. A high phosphate transfer potential refers to a high energy release when the phosphate group is hydrolysed. With PPKs, this seems slightly different: The hydrolysis energy involving the terminal phosphate group of polyP ($\Delta^R G^0 \approx -30$ kJ/mol) and internal phosphate groups ($\Delta^R G^0 \approx -30$ to -40 kJ/mol) is comparable to the standard hydrolysis energy of ATP [33,34]. Under physiological conditions as well as in an in vitro system the actual ΔG may be different by coordination of cations and the ionic strength, temperature and pH of the reaction solution [29,35,36]. Compared to other ATP synthesis reactions very little is known about the thermodynamics for the PPK1 and PPK2-catalysed reactions with polyP as a phosphate donor since studies mainly focus on the kinetic characterisation of these enzymes.

Considering the growing interest in the application of PPK enzymes, knowledge about the thermodynamic course of the reaction would be useful for the optimisation of biocatalytic syntheses of nucleotides as well as nucleotide regeneration systems (Figure 3a and 3b). In a regeneration system (exemplarily shown for the hexokinase-catalysed phosphorylation of glucose (Figure 3a) the formed ADP has to be converted back to ATP to maintain a sufficient pool of ATP for the hexokinase reaction [37]. Each of the phosphate donors discussed can be used in combination with the corresponding kinase to regenerate the ATP rendering the available regeneration systems flexible and broadly applicable. While the depicted regeneration system is quite simple, the reaction can be embedded in complex biosyn-

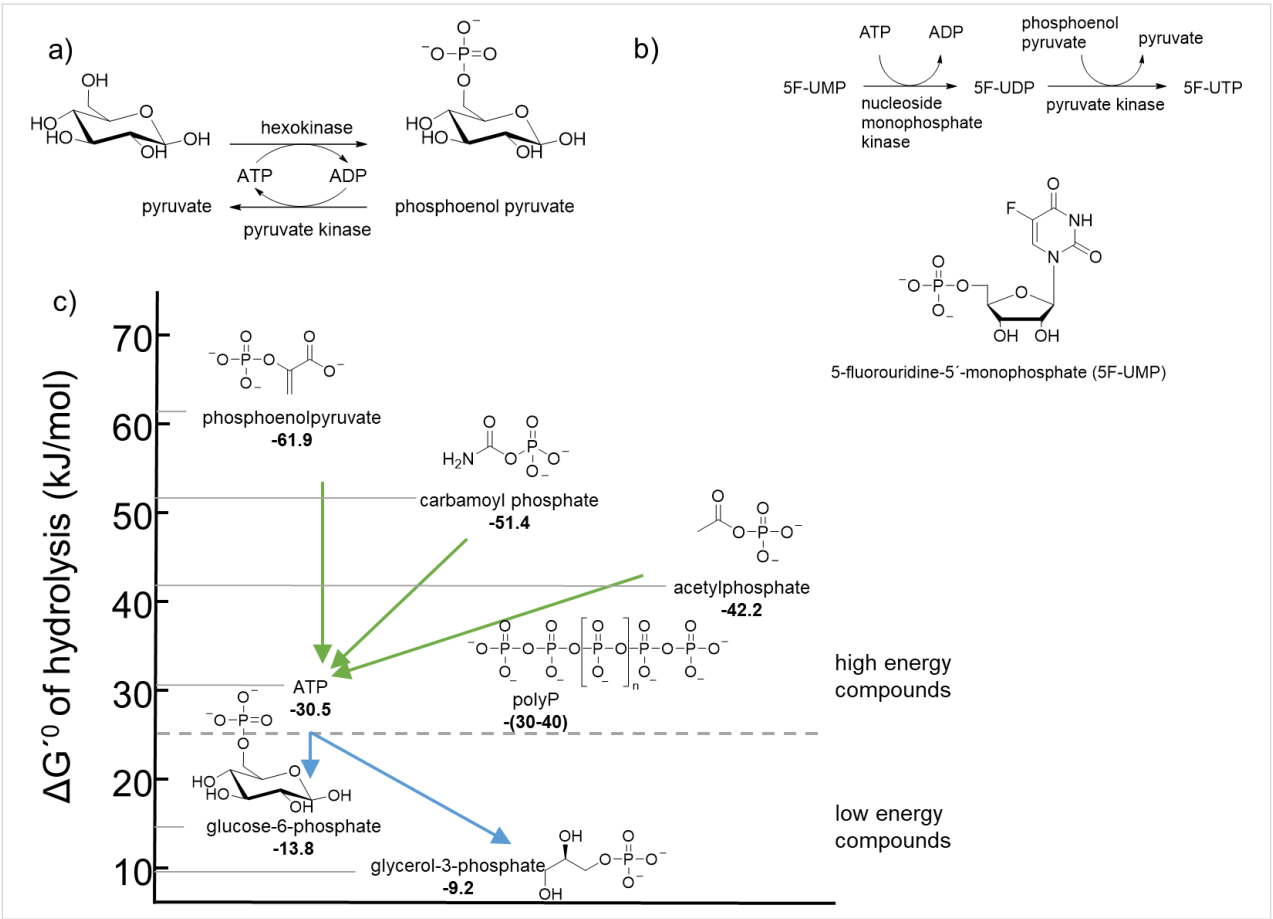


Figure 3: a) Kinases as ATP regeneration enzymes, exemplified by the hexokinase-catalysed ATP-dependent phosphorylation of glucose. The ADP formed is converted back to ATP by pyruvate kinase using phosphoenol pyruvate as the phosphate donor [37]. b) Utilisation of kinases for biocatalytic NTP production. Usually, a sequence of different kinases with a suitable phosphate donor is used, depending on the starting material different enzymes have to be employed. The example shows the final two reaction steps in the biosynthesis of 5-fluorouridine-5`-monophosphate (5F-UMP). The phosphorylation of the NMP is catalysed by the ATP dependent nucleoside monophosphate kinase yielding a 5F-UDP and ADP. 5F-UDP is then phosphorylated by pyruvate kinase under consumption of phosphoenol pyruvate [42]. c) Ranking of different phosphate donors that can be used for ATP regeneration or production because their phosphate group transfer potential is higher compared to ATP. From ATP the phosphate group can be transferred onto various substrates but the phosphate group cannot be easily transferred back to ADP from a low energy compound [28-30].

thetic networks such as in vitro *S*-adenosylmethionine (SAM)- or carbon dioxide fixation cycles, and de novo nucleobase synthesis [38-41]. For the biocatalytic synthesis of ATP or derivatives, up to three consecutive phosphorylation reactions are coupled in a linear cascade to produce the desired NTP. Figure 3b shows the reaction sequence from 5-fluorouridine-5'-monophosphate to the triphosphate in an enzymatic synthesis of an unnatural uridine nucleotide [42]. In these type of setup, the yield of the overall reaction is strongly determined by the position of the thermodynamic equilibrium of the last reaction step [22,23,42-45]. To efficiently implement such a reaction sequence, detailed kinetic and thermodynamic information has to be available to identify bottlenecks and improve the turnover of such cascade systems [25].

In the present study, we analysed a set of well-known PPK enzymes regarding the thermodynamic equilibrium of ATP synthesis and compared the experimental results obtained with theoretical calculations. The theoretical calculations addressed the equilibrium position of the considered reactions as function of the substrate concentration. In addition to the general question of the equilibria of the PPK-catalysed reactions in comparison to other kinases, we sought to evaluate the contribution of the reported characteristic "preferences" of PPK1 and PPK2 enzymes regarding polyP synthesis and polyP degradation on the reaction rate and equilibrium formation.

Results and Discussion

The enzymes investigated include all four kinds of identified PPK1/PPK2 enzymes. PPK1 from *E. coli* (*Ec*PPK1) and *V. cholerae* (*Vc*PPK1) [20] are two well investigated PPK1 with different kinetic preferences [20,21]. *Sm*PPK2 [10] is the model PPK2 enzyme and often described as the counterpart of *Ec*PPK1. *Cg*PPK2 is also designated to have different kinetic

preferences compared to the model enzyme *Sm*PPK2, favouring the usage of polyP over the synthesis of ATP [15]. All enzymes used in this work were produced in *E. coli*. While the PPK2s were produced as soluble enzymes and could be easily purified via Ni-NTA affinity chromatography using N-terminal His-tags, the PPK1 enzymes required an N-terminal maltose binding protein (MBP-tag) to improve solubility [46,47]. Trials to cleave the MBP-tag were unsuccessful and resulted in inactive protein aggregates. It has been shown that large tags such as the MBP-tag, as well as the positioning (N- or C-terminal) of the tag can influence the activity of an enzyme [48]; nevertheless, its thermodynamic characteristics should not be affected. Consequently, we decided to use the enzymes with the tags, as this likely is the most pragmatic and straightforward preparation of the enzymes for their use in chemical synthesis. The assay setup used for all PPKs is similar to established ATP regeneration systems with PPK2 enzymes, with a 10:1 excess of polyP (calculated as single phosphates [49]) over the nucleotide (default concentration in this work 2 mM) [39,50-52]. This is also a realistic scenario for biosynthetic reactions using PPKs for the production of NTPs [22,52]. The excess of polyP should prevent depletion of sufficiently long polyP chains since acceptance of very short chains ($n < 10$) might differ between different PPKs [10,15].

First experiments were conducted with the "model PPKs" *Ec*PPK1 and *Sm*PPK2. At 37 °C and pH 8, each enzyme was incubated with either ADP or ATP and polyP as co-substrate; the resulting nucleotide distributions were analysed by HPLC as a function of time (Figure 4). In all experiments, the equilibrium was reached after 90 minutes with no further changes in product concentrations upon extending incubation time. Regardless of starting the reaction with ADP or ATP, the equilibrium tends towards a ratio of 70% ATP and 30% ADP (molar ratio, ADP/ATP = 0.43). About 5% AMP was observed during the reaction, which is derived from the starting material and is not further

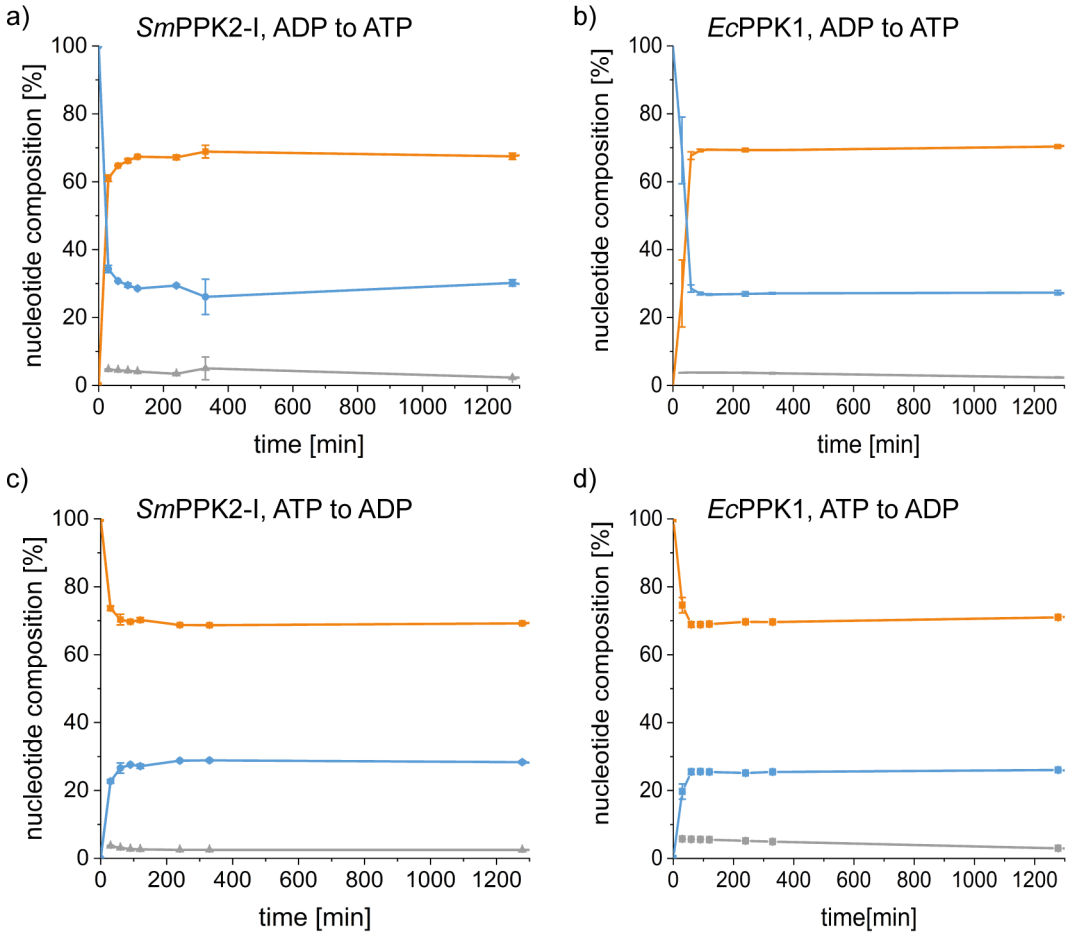


Figure 4: Time courses of reactions started with ADP catalysed by a) *Sm*PPK2 and b) *Ec*PPK1. Time courses of reactions started with ATP catalysed by c) *Sm*PPK2 and d) *Ec*PPK1. The nucleotide concentration was 2 mM, the reaction was carried out at pH 8.0 and 37 °C. Grey = AMP, blue = ADP, orange = ATP; the nucleotide composition is given in mol %.

accumulating over the course of the reaction. Higher phosphorylated compounds such as adenosine 5'-tetraphosphate, which are additional reaction products of PPK2 catalysed reactions, were not detected under the conditions applied, this usually requires higher enzyme concentrations [7,8]. A similar equilibrium concentration was observed for *Vc*PPK1 (ADP/ATP 30%:70%, Figure S1, Supporting Information File 1). In the *Cg*PPK2-catalysed reaction, the ADP/ATP ratio (ADP:ATP 35–40%:65–60%, Figure S1, Supporting Information File 1), as well as the amount of AMP formed was slightly higher compared to the other PPKs, especially when starting from ADP as substrate. This suggests that *Cg*PPK2 possesses a pronounced myokinase (2 ADP ↔ ATP + AMP) activity that cannot be suppressed at the applied conditions nor separated from the main reaction; this has been already described as a side reaction for other PPK2s [7,17]. In summary, these findings demonstrate that both, PPK1 and PPK2 enzymes catalyse the formation of the same equilibrium despite their different reaction pathways. After 30 minutes the reaction was close to equilibrium formation; thus, no kinetic effect of family 1 or 2 could be observed. The results obtained also show that polyP, despite its rather low phosphate transfer potential, is a sufficient phosphate donor for the phosphorylation of ADP, as the thermodynamic equilibrium is clearly positioned on the ATP side.

Next, we analysed the influence of the amount of the starting nucleotide for the two model enzymes *Ec*PPK1 and *Sm*PPK2: either one quarter (0.5 mM) or twice (4 mM) the default amount (2 mM) of nucleotide was used. With 0.5 mM, the equilibrium composition contained slightly lower amounts of ATP than with 2 mM after 30 min (60% ATP, Figure 4b, Figure S2, Supporting Information File 1). With 4 mM of starting nucleotide, the equilibrium is composed similar to the one of the experiments with 2 mM nucleotide (70% ATP/30% ADP, Figure S3, Supporting Information File 1). This can be explained by a concentration effect of the nucleotide on the reaction equilibrium. For this, we applied the thermodynamic activity-based framework that uses the equilibrium constant K_a, which is independent of concentration. It is expressed via the law of mass action, and exemplarily for the reaction from ADP to ATP it reads as

$$K_a = const. = \vec{m}(\vec{\gamma}) = \frac{m_{ATP}}{m_{ADP}} \frac{m_{PolyP_{n-1}}}{m_{PolyP}} \frac{\gamma_{ATP}}{\gamma_{ADP}} \frac{\gamma_{PolyP_{n-1}}}{\gamma_{PolyP}}.$$

This equation shows that any change in the reaction equilibrium (ratio of the equilibrium molalities $\vec{m}$) must be equalised by the ratio of the equilibrium activity coefficients $\vec{\gamma}$ of the reacting agents, or in other words $\vec{m} \sim (\vec{\gamma})^{-1}$. The activity coefficients describe the molecular interactions among the reaction participants in the reaction mixtures, which has been established for biochemical reactions [53,54]. The predictive electrolyte equation of state ePC-SAFT [55] was applied to predict the activity coefficients at equilibrium. ePC-SAFT is an electrolyte perturbation theory which describes physical interactions by accounting for molecular repulsion and attraction caused by van-der-Waals forces, hydrogen bonding, and Coulomb forces. The ePC-SAFT parameters of the nucleotides were fitted in previous works to experimental osmotic pressures of pseudo-binary mixtures of nucleotide and water [29,33]. As modelling polyP with high chain length is currently not possible with ePC-SAFT, we assumed that only nucleotides were present in water. The consequence of this assumption is that interactions among nucleotides and polyP were considered to be equal to interactions among nucleotides and $polyP_{n-1}$, and we focused only on the ratio of the nucleotides. Upon increasing the nucleotide concentration, the equilibrium concentration ratios shift to the ATP side of the reactions: the higher the initial substrate concentration the lower the equilibrium ratio ADP/ATP that is to be expected. This fits with the experimental data shown in Figure 5a (experimental concentration ratio at equilibrium over the initial nucleotide concentration). It can be observed from Figure 5b that the ePC-SAFT predictions are in qualitative agreement with the experimental findings, since the activity-coefficient ratio behaves reciprocally to the experimentally observed concentration ratio at equilibrium (Figure 5a). Thus, the interactions between the reacting agents (covered by $\vec{\gamma}$) cause a shift in the equilibrium position $\vec{m}$ according to the results shown in Figure 5a fulfilling the above-shown thermodynamic constraint $\vec{m} \sim (\vec{\gamma})^{-1}$. As mentioned before, the theoretical prediction procedure represented in Figure 5b ignores the presence of polyP in the reaction mixture. However, in the reaction also $polyP_n$ and $polyP_{n-1}$ take part; their concentration ratio might additionally influence the equilibrium position of the overall reaction. As it is not yet possible to characterise polyP by thermodynamic modelling due to lack of experimental data and knowledge of the precise distribution of chain lengths, in a second step the influence of orthophosphate as a representative for polyP was investigated on the qualitative behaviour of the results in Figure 5b. The results are not shown in detail here, but we found that the addition of orthophosphate did not change the qualitative course of the activity-coefficient ratios from Figure 5b.

Despite the qualitative success of the ePC-SAFT predictions, some quantitative discrepancies can be observed between Figure 5a and 5b. Using 4 mM of substrate or above, no further change in the ADP/ATP equilibrium was experimentally observed (Figure 5a), while ePC-SAFT predicts a linear behaviour with nucleotide concentration (Figure 5b). This discrepancy between model and experiment might be explained by experimental issues (measurement uncertainty, occurrence of side reactions not considered in the modelling) or by theory issues,

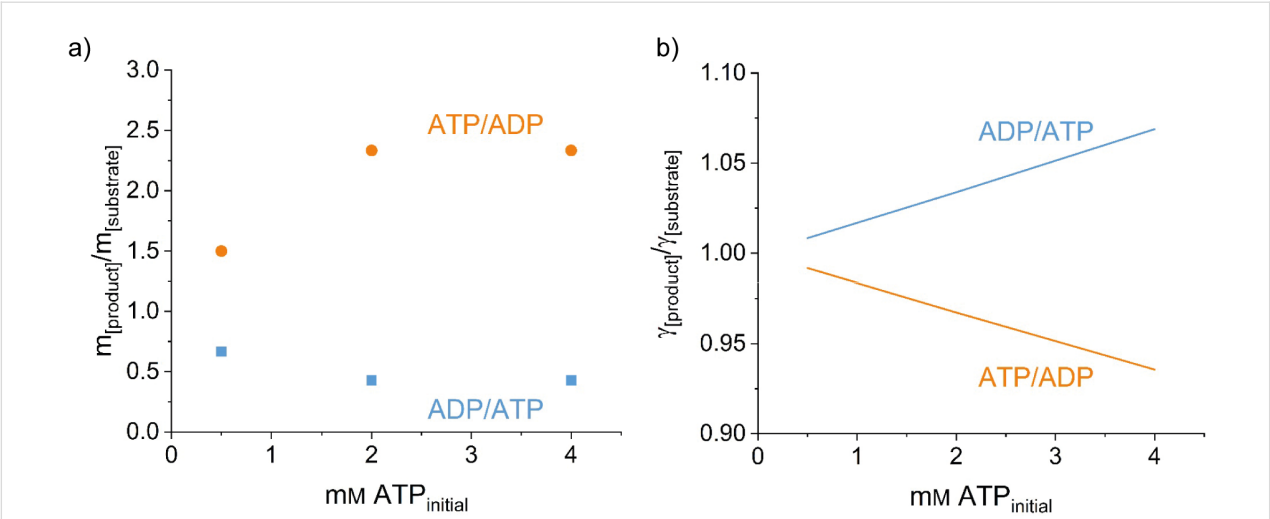


Figure 5: a) Ratio of product to substrate for the three nucleotide concentrations (0.5, 2, and 4 mM) used in this work, both synthesis directions are depicted (orange: ATP as substrate, blue: ADP as substrate). a) The ratio (moles per kg) is shown against the concentration of initial ATP. b) ePC-SAFT predicted ratio of activity coefficients for the two reactions against the concentration of initial ATP.

since, as described before, the influence of polyP was neglected in modelling with ePC-SAFT. Further, it should be noted again ePC-SAFT was used in a predictive mode, which was not fitted at all to any reaction experiment.

In contrast to the reaction equilibria, substantial kinetic differences between PPK1 and PPK2 were observed at 4 mM substrate concentration regarding the ATP synthesis reaction. Starting from ADP, the *Sm*PPK2-catalysed reaction reached the equilibrium in 30 minutes, the same reaction catalysed by *Ec*PPK1 only after 240 minutes (Figure 6). Based on literature data (Table S6, Supporting Information File 1), an opposite trend was expected for the reaction started with ATP; however, in this direction, no clear kinetic effect could be observed, and the equilibrium is generally reached very quickly. This effect is experimentally observed – it is shown in the literature that kinetics should be expressed based on the thermodynamic activity of the enzyme [56]. The enzyme activity coefficients were not taken into account in the present work, which does not allow drawing conclusions from the concentration-based kinetics presented in Figure 6 and the activity-based consideration of the equilibrium constants yielding the results in Figure 5. Nevertheless, the experimentally observed difference in velocity agrees with the described kinetic "preference" of PPK2

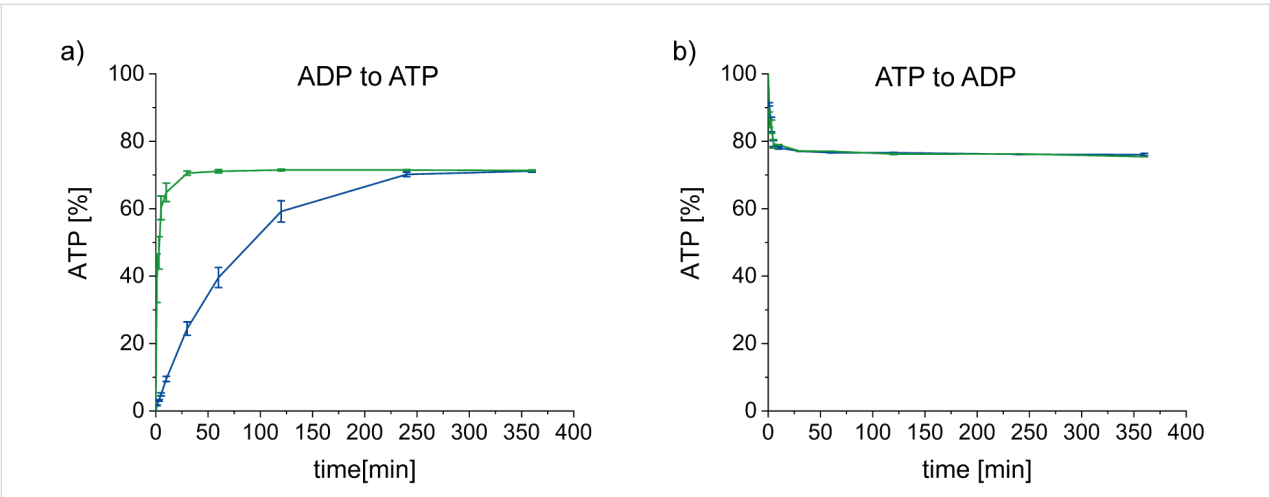


Figure 6: Time courses comparing *Ec*PPK1 (blue) and *Sm*PPK2 (green) for ATP synthesis (a) and ATP degradation (b) when 4 mM of the nucleotide was provided. While there is very little difference in the degradation reaction, *Sm*PPK2 reaches the equilibrium earlier than *Ec*PPK1 in the synthesis reaction.

for nucleotide synthesis and the tendency to use these enzymes for ATP (re)generation systems.

Another variable discussed in the context of PPK reactions is the polyP chain length. Lowered or increased activity with different chain lengths of polyP is described for various PPKs [5,15,19]. So far we used $polyP_n$ with an average chain length of n = 20 (as estimated by ^{31}P NMR, Figure S5, Supporting Information File 1) for all reactions, therefore the question remained whether or not the polyP chain length has a substantial influence on the thermodynamic equilibrium. For this reason, we conducted a reaction with a commercial $polyP_{100}$ (calculated as single phosphates, average length confirmed by ^{31}P NMR, Figure S6, Supporting Information File 1). Compared to the data obtained for the identical reaction setup with the $polyP_{20}$ (Figure 2), no differences were observed (Figure S4, Supporting Information File 1).

As discussed before, the conversion of a nucleotide diphosphate to the corresponding triphosphate is normally the last step in the biomimetic synthesis of NTPs, and might have a substantial influence on the yield of the whole cascade. A biomimetic cascade published by Whitesides and co-workers for the synthesis of ATP from adenosine uses an acetate kinase for the final conversion of ADP to ATP (Table 1) [45]. Acetylphosphate has long been preferred as phosphate donor over phosphoenolpyruvate in large-scale reactions, due to economic factors such as ease of production and atom economy [57]. On an analytical scale, this reaction reached conversions of up to 94%, which agrees with the thermodynamic equilibrium for acetate kinase of at least 90%:10% ATP:ADP [26].

A comparable one-pot synthesis was recently published by the group of Li (Table 1) using two PPKs (PPK2-II and PPK2-I) with polyP as a phosphate donor for AMP and ADP phosphorylation; this reached yields up to 75% ATP thus supporting the thermodynamic equilibrium for PPKs reported in this study [44]. Interestingly, we obtained similar results when investigating the class-III PPK2 from *Meiothermus ruber* for its polyP-dependent reaction with AMP (via ADP) to ATP (2.5% AMP/27.5% ADP/70% ATP) [16].

In theory, the addition of another, ideally irreversible reaction can be used to pull the equilibrium further to the product side. Another recent application of PPKs in the cascade synthesis of cladribine triphosphate (2-chloro-2'-deoxyadenosine 5'-triphosphate) could be discussed as an example for this (Table 1). Here, PPK2 enzymes are used to produce 2-Cl-ATP which is then reduced to the desired product. The final reduction could be considered as a pull effect on the PPK2 reaction, since it removes 2-Cl-ATP from the PPK2 reaction equilibrium. In fact, the authors observed an (compared to the adenosine-to-ATP cascades) increased yield of 80% cladribine triphosphate, with 2-Cl-dADP being the major byproduct [22]. The incomplete conversion (despite the irreversible reduction step) has been explained by the previously mentioned myokinase side reactivity of PPK2 enzymes [22]. An alternative explanation might be the PPK equilibrium: most PPKs accept a broad range of nucleoside substrates, therefore some 2-Cl-dATP produced could be a substrate for a PPK2 catalysed 2-Cl-dATP/2-Cl-dADP equilibrium.

Conclusion

PPK2, and also PPK1 enzymes are frequently used in ATP regeneration systems as well as in cascade reactions for NTP biosynthesis [22,23,39,44,50,58-60]. Therefore, the results presented here can serve to further tune and improve these multi-enzyme reactions and the yield of biomimetic NTP syn-

Table 1: Comparison of selected enzymatic syntheses of ATP (derivatives). The triphosphate-forming step is in bold, stoichiometrically added substrates are underlined.

	Reaction sequence	Yield of ATP	Reference
(1)	(a) adenosine + ATP → AMP + ADP (b) AMP + ATP → 2 ADP **(c) ADP + acetylphosphate → ATP + acetate**	94%	R. L. Baughn et al. [45]
(2)	(a) adenosine + ATP → AMP + ADP (b) AMP + $polyP_n$ → ADP + $polyP_{n-1}$ **(c) ADP + $polyP_n$ → ATP + $polyP_{n-1}$**	70–75%	C. Sun et al. [44]
(3)	(a) 2-Cl-adenine + PRPP → 2-Cl-AMP +PP_i (b) 2-Cl-AMP+ $polyP_n$ → 2-Cl-ADP + $polyP_{n-1}$ **(c) 2-Cl-ADP+ $polyP_n$ → 2-Cl-ATP + $polyP_{n-1}$** (d) 2-Cl-ATP → 2-Cl-dATP	80%	J. Frisch et al. [22]

thesis systems. Usually, reactions for polyP-dependent ATP regeneration as well as NTP biosynthesis are carried out under conditions comparable to the ones used for this study [22,39,50,52]. Overall, we can clearly see a preference of the thermodynamic system towards ATP. This behaviour was not impacted by the choice of PPK, regardless of the different reaction mechanisms. The thermodynamic equilibrium of the PPK reaction is not as close to the product side as in other ATP regenerating systems, which corresponds to the phosphate transfer potential of the different phosphate donors. This raises the question if PPK/polyP based systems are actually the best choice for biocatalytic syntheses of NTPs and other phosphorylated compounds, or if a phosphate donor with a higher phosphate transfer potential would be more useful. In our opinion, this has to be tailored to each individual system and depends on the characteristics of the starting material, and factors such as the necessity to purify the end product. Also, especially PPK2s are described to be very flexible regarding the nucleobase, which is not the case for all other kinases [10,17,61]; depending on the system this might compensate for the not ideal conversion rates. Nevertheless, as it becomes evident from the cladribine system, exactly this broad substrate range might be disadvantageous, thus highlighting the necessity to tailor the choice of enzymes to the substrates and reaction sequence in question. In ATP regeneration systems, the equilibrium issue might be less relevant, as the ATP produced will be directly used by the main reaction.

In future, it will be interesting to investigate the detailed reaction mechanism including the effects of the polyP chain length and counter ions as well as to study the thermodynamic activity of the enzymes. Especially in reaction setups where the synthesised ATP is not directly removed by follow up reactions, options to tune the reaction conditions in order to increase the conversion and yields of the final cascade product will be an important aim.

Supporting Information

Supporting Information File 1

Details of materials and methods and additional figures and tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-134-S1.pdf]

Acknowledgements

We would like to thank Katharina Strack for skilful technical assistance, and Dr. Manfred Keller (University of Freiburg) for help with ^{31}P NMR analysis of polyphosphates. Dr. Torsten Sehl (Research Centre Juelich) and Dr. Désirée Popadić (University of Freiburg) are acknowledged for helpful discussions. We appreciate the donation of pETM41-*Ec*PPK from Dr. Florian Freimoser (University of Zurich).

Funding

Work in the Jessen group (HJJ, SM) was supported by the Deutsche Forschungsgemeinschaft (DFG JE 572/8-1), JNA acknowledges support from the Heisenberg Program (DFG AN 761/3-1).

ORCID® iDs

Michael Keppler - https://orcid.org/0000-0002-6579-6885
Sandra Moser - https://orcid.org/0000-0003-3718-8521
Henning J. Jessen - https://orcid.org/0000-0002-1025-9484
Christoph Held - https://orcid.org/0000-0003-1074-177X
Jennifer N. Andexer - https://orcid.org/0000-0001-8238-5749

References

1. Kornberg, A.; Rao, N. N.; Ault-Riché, D. *Annu. Rev. Biochem.* **1999,** *68,* 89–125. doi:10.1146/annurev.biochem.68.1.89
2. Mailer, R. K. W.; Hänel, L.; Allende, M.; Renné, T. *Front. Med.* **2019,** *6,* 76. doi:10.3389/fmed.2019.00076
3. Rao, N. N.; Gómez-García, M. R.; Kornberg, A. *Annu. Rev. Biochem.* **2009,** *78,* 605–647. doi:10.1146/annurev.biochem.77.083007.093039
4. Kornberg, A.; Kornberg, S. R.; Simms, E. S. *Biochim. Biophys. Acta* **1956,** *20,* 215–227. doi:10.1016/0006-3002(56)90280-3
5. Ishige, K.; Zhang, H.; Kornberg, A. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 16684–16688. doi:10.1073/pnas.262655299
6. Motomura, K.; Hirota, R.; Okada, M.; Ikeda, T.; Ishida, T.; Kuroda, A. *Appl. Environ. Microbiol.* **2014,** *80,* 2602–2608. doi:10.1128/aem.03971-13
7. Mordhorst, S.; Singh, J.; Mohr, M. K. F.; Hinkelmann, R.; Keppler, M.; Jessen, H. J.; Andexer, J. N. *ChemBioChem* **2019,** *20,* 1019–1022. doi:10.1002/cbic.201800704
8. Frank, C.; Teleki, A.; Jendrossek, D. *Appl. Microbiol. Biotechnol.* **2020,** *104,* 9683–9692. doi:10.1007/s00253-020-10891-7
9. Sureka, K.; Dey, S.; Datta, P.; Singh, A. K.; Dasgupta, A.; Rodrigue, S.; Basu, J.; Kundu, M. *Mol. Microbiol.* **2007,** *65,* 261–276. doi:10.1111/j.1365-2958.2007.05814.x
10. Nocek, B.; Kochinyan, S.; Proudfoot, M.; Brown, G.; Evdokimova, E.; Osipiuk, J.; Edwards, A. M.; Savchenko, A.; Joachimiak, A.; Yakunin, A. F. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 17730–17735. doi:10.1073/pnas.0807563105
11. Ahn, K.; Kornberg, A. *J. Biol. Chem.* **1990,** *265,* 11734–11739. doi:10.1016/s0021-9258(19)38459-5
12. Akiyama, M.; Crooke, E.; Kornberg, A. *J. Biol. Chem.* **1992,** *267,* 22556–22561. doi:10.1016/s0021-9258(18)41708-5
13. Zhu, Y.; Huang, W.; Lee, S. S. K.; Xu, W. *EMBO Rep.* **2005,** *6,* 681–687. doi:10.1038/sj.embor.7400448
14. Kumble, K. D.; Ahn, K.; Kornberg, A. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93,* 14391–14395. doi:10.1073/pnas.93.25.14391
15. Lindner, S. N.; Vidaurre, D.; Willbold, S.; Schoberth, S. M.; Wendisch, V. F. *Appl. Environ. Microbiol.* **2007,** *73,* 5026–5033. doi:10.1128/aem.00600-07

16. Parnell, A. E.; Mordhorst, S.; Kemper, F.; Giurrandino, M.; Prince, J. P.; Schwarzer, N. J.; Hofer, A.; Wohlwend, D.; Jessen, H. J.; Gerhardt, S.; Einsle, O.; Oyston, P. C. F.; Andexer, J. N.; Roach, P. L. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115,* 3350–3355. doi:10.1073/pnas.1710741115
17. Nocek, B. P.; Khusnutdinova, A. N.; Ruszkowski, M.; Flick, R.; Burda, M.; Batyrova, K.; Brown, G.; Mucha, A.; Joachimiak, A.; Berlicki, Ł.; Yakunin, A. F. *ACS Catal.* **2018,** *8,* 10746–10760. doi:10.1021/acscatal.8b03151
18. Kamerlin, S. C. L.; Sharma, P. K.; Prasad, R. B.; Warshel, A. *Q. Rev. Biophys.* **2013,** *46,* 1–132. doi:10.1017/s0033583512000157
19. Zhang, H.; Ishige, K.; Kornberg, A. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 16678–16683. doi:10.1073/pnas.262655199
20. Ogawa, N.; Tzeng, C.-M.; Fraley, C. D.; Kornberg, A. *J. Bacteriol.* **2000,** *182,* 6687–6693. doi:10.1128/jb.182.23.6687-6693.2000
21. Kuroda, A.; Kornberg, A. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94,* 439–442. doi:10.1073/pnas.94.2.439
22. Frisch, J.; Maršić, T.; Loderer, C. *Biomolecules* **2021,** *11,* 346. doi:10.3390/biom11030346
23. Fehlau, M.; Kaspar, F.; Hellendahl, K. F.; Schollmeyer, J.; Neubauer, P.; Wagner, A. *Front. Bioeng. Biotechnol.* **2020,** *8,* 854. doi:10.3389/fbioe.2020.00854
24. Andexer, J. N.; Richter, M. *ChemBioChem* **2015,** *16,* 380–386. doi:10.1002/cbic.201402550
25. Abu, R.; Woodley, J. M. *ChemCatChem* **2015,** *7,* 3094–3105. doi:10.1002/cctc.201500603
26. Langer, R. S.; Gardner, C. R.; Hamilton, B. K.; Colton, C. K. *AIChE J.* **1977,** *23,* 1–10. doi:10.1002/aic.690230102
27. Dobson, G. P.; Hitchins, S.; Teague, W. E., Jr. *J. Biol. Chem.* **2002,** *277,* 27176–27182. doi:10.1074/jbc.m111422200
28. Nelson, D. L.; Cox, M. M.; Lehninger, A. L. *Lehninger Principles of Biochemistry,* 7th ed.; W.H. Freeman: New York, NY, USA, 2017.
29. Greinert, T.; Vogel, K.; Maskow, T.; Held, C. *Sci. Rep.* **2021,** *11,* 6125. doi:10.1038/s41598-021-85594-8
30. Alberty, R. A. *Arch. Biochem. Biophys.* **2006,** *451,* 17–22. doi:10.1016/j.abb.2006.03.025
31. Lundblad, R. L.; Macdonald, F. *Free Energies of Hydrolysis and Decarboxylation; Handbook of Biochemistry and Molecular Biology*; CRC Press: Boca Raton, FL, USA, 2018.
32. Rodwell, V. W.; Bender, D. A.; Botham, K. M.; Kennelly, P. J.; Weil, P. A. *Harper's Illustrated Biochemistry;* McGraw Hill: New York, NY, USA, 2018.
33. Meurer, F.; Do, H. T.; Sadowski, G.; Held, C. *Biophys. Chem.* **2017,** *223,* 30–38. doi:10.1016/j.bpc.2017.02.005
34. Müller, W. E. G.; Schröder, H. C.; Wang, X. *Chem. Rev.* **2019,** *119,* 12337–12374. doi:10.1021/acs.chemrev.9b00460
35. Vogel, K.; Greinert, T.; Harms, H.; Sadowski, G.; Held, C.; Maskow, T. *Biochim. Biophys. Acta, Gen. Subj.* **2020,** *1864,* 129675. doi:10.1016/j.bbagen.2020.129675
36. Vogel, K.; Greinert, T.; Reichard, M.; Held, C.; Harms, H.; Maskow, T. *Int. J. Mol. Sci.* **2020,** *21,* 7921. doi:10.3390/ijms21217921
37. Wettermark, G.; Borglund, E.; Brolin, S. E. *Anal. Biochem.* **1968,** *22,* 211–218. doi:10.1016/0003-2697(68)90308-4
38. Mordhorst, S.; Siegrist, J.; Müller, M.; Richter, M.; Andexer, J. N. *Angew. Chem., Int. Ed.* **2017,** *56,* 4037–4041. doi:10.1002/anie.201611038
39. 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
40. Schwander, T.; Schada von Borzyskowski, L.; Burgener, S.; Cortina, N. S.; Erb, T. J. *Science* **2016,** *354,* 900–904. doi:10.1126/science.aah5237
41. Schultheisz, H. L.; Szymczyna, B. R.; Scott, L. G.; Williamson, J. R. *ACS Chem. Biol.* **2008,** *3,* 499–511. doi:10.1021/cb800066p
42. Hennig, M.; Scott, L. G.; Sperling, E.; Bermel, W.; Williamson, J. R. *J. Am. Chem. Soc.* **2007,** *129,* 14911–14921. doi:10.1021/ja073825i
43. Da Costa, C. P.; Fedor, M. J.; Scott, L. G. *J. Am. Chem. Soc.* **2007,** *129,* 3426–3432. doi:10.1021/ja067699e
44. Sun, C.; Li, Z.; Ning, X.; Xu, W.; Li, Z. *Bioresour. Bioprocess.* **2021,** *8,* 117. doi:10.1186/s40643-021-00469-0
45. Baughn, R. L.; Adalsteinsson, O.; Whitesides, G. M. *J. Am. Chem. Soc.* **1978,** *100,* 304–306. doi:10.1021/ja00469a063
46. Nilsson, J.; Ståhl, S.; Lundeberg, J.; Uhlén, M.; Nygren, P.-Å. *Protein Expression Purif.* **1997,** *11,* 1–16. doi:10.1006/prep.1997.0767
47. Terpe, K. *Appl. Microbiol. Biotechnol.* **2003,** *60,* 523–533. doi:10.1007/s00253-002-1158-6
48. Zhu, Y.; Lee, S. S. K.; Xu, W. *Biochem. Biophys. Res. Commun.* **2003,** *305,* 997–1001. doi:10.1016/s0006-291x(03)00886-6
49. PolyP NMR. The average chain length of polyP can be determined by phosphorus NMR but the exact size is very challenging to determine.
50. Mordhorst, S.; Maurer, A.; Popadić, D.; Brech, J.; Andexer, J. N. *ChemCatChem* **2017,** *9,* 4164–4168. doi:10.1002/cctc.201700848
51. Liu, S.; Li, Y.; Zhu, J. *Process Biochem. (Oxford, U. K.)* **2016,** *51,* 1458–1463. doi:10.1016/j.procbio.2016.06.006
52. Ngivprom, U.; Lasin, P.; Khunnonkwao, P.; Worakaensai, S.; Jantama, K.; Kamkaew, A.; Lai, R.-Y. *ChemBioChem* **2022,** *23,* e202200071. doi:10.1002/cbic.202200071
53. Held, C.; Sadowski, G. *Annu. Rev. Chem. Biomol. Eng.* **2016,** *7,* 395–414. doi:10.1146/annurev-chembioeng-080615-034704
54. Hoffmann, P.; Voges, M.; Held, C.; Sadowski, G. *Biophys. Chem.* **2013,** *173–174,* 21–30. doi:10.1016/j.bpc.2012.12.006
55. Held, C.; Reschke, T.; Mohammad, S.; Luza, A.; Sadowski, G. *Chem. Eng. Res. Des.* **2014,** *92,* 2884–2897. doi:10.1016/j.cherd.2014.05.017
56. Knierbein, M.; Wangler, A.; Luong, T. Q.; Winter, R.; Held, C.; Sadowski, G. *Phys. Chem. Chem. Phys.* **2019,** *21,* 22224–22229. doi:10.1039/c9cp03868j
57. Crans, D. C.; Kazlauskas, R. J.; Hirschbein, B. L.; Wong, C.-H.; Abril, O.; Whitesides, G. M. *Methods Enzymol.* **1987,** *136,* 263–280. doi:10.1016/s0076-6879(87)36027-6
58. Cao, H.; Li, C.; Zhao, J.; Wang, F.; Tan, T.; Liu, L. *Appl. Biochem. Biotechnol.* **2018,** *185,* 385–395. doi:10.1007/s12010-017-2664-4
59. Resnick, S. M.; Zehnder, A. J. B. *Appl. Environ. Microbiol.* **2000,** *66,* 2045–2051. doi:10.1128/aem.66.5.2045-2051.2000
60. Zhang, X.; Wu, H.; Huang, B.; Li, Z.; Ye, Q. *J. Biotechnol.* **2017,** *241,* 163–169. doi:10.1016/j.jbiotec.2016.11.034
61. Nahálka, J.; Pätoprstý, V. *Org. Biomol. Chem.* **2009,** *7,* 1778–1780. doi:10.1039/b822549b

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Cytochrome P450 monooxygenase-mediated tailoring of triterpenoids and steroids in plants

Karan Malhotra[1] and Jakob Franke[*1,2]

Review

Open Access

Address:
[1]Institute of Botany, Leibniz University Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany and [2]Centre of Biomolecular Drug Research, Leibniz University Hannover, Schneiderberg 38, 30167 Hannover, Germany

Email:
Jakob Franke[*] - jakob.franke@botanik.uni-hannover.de

* Corresponding author

Keywords:
biosynthesis; CYPs; cytochrome P450 monooxygenases; plants; steroid; sterol; triterpene; triterpenoid

Beilstein J. Org. Chem. **2022,** *18,* 1289–1310.
https://doi.org/10.3762/bjoc.18.135

Received: 29 June 2022
Accepted: 02 September 2022
Published: 21 September 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

The cytochrome P450 monooxygenase (CYP) superfamily comprises hemethiolate enzymes that perform remarkable regio- and stereospecific oxidative chemistry. As such, CYPs are key agents for the structural and functional tailoring of triterpenoids, one of the largest classes of plant natural products with widespread applications in pharmaceuticals, food, cosmetics, and agricultural industries. In this review, we provide a full overview of 149 functionally characterised CYPs involved in the biosynthesis of triterpenoids and steroids in primary as well as in specialised metabolism. We describe the phylogenetic distribution of triterpenoid- and steroid-modifying CYPs across the plant CYPome, present a structure-based summary of their reactions, and highlight recent examples of particular interest to the field. Our review therefore provides a comprehensive up-to-date picture of CYPs involved in the biosynthesis of triterpenoids and steroids in plants as a starting point for future research.

Introduction

Triterpenoids are a large class of natural products derived from precursors containing 30 carbon atoms and composed of six isoprene units (C_5). The structural variety of triterpenoids found in plants is particularly astonishing, and so are their biological activities. To date, more than 20,000 different plant triterpenoids have been identified, and many of these have found applications in agronomic [1], food [2], cosmetics [3] and pharmaceutical industries [4]. Plant triterpenoids include primary metabolites such as phytosterols or steroidal hormones such as brassinosteroids, but also specialised metabolites that convey diverse biological functions [5]. A key factor for the structural richness of triterpenoids and steroids from plants lies in their extensive oxidative tailoring, which in many cases is carried out by cytochrome P450 monooxygenases (CYPs). CYPs represent one of the largest superfamilies of enzymes in plants; in many species, around 1% of all genes encode CYPs [6]. CYPs are well-known for their capacity to catalyse highly regio- and stereospecific reactions on complex substrates. Besides simple hydroxylations, they can also introduce oxo, carboxy, or epoxy moieties or double bonds. Such decorations often also enable

additional layers of diversification by glycosyltransferases or acyltransferases [7]. Hence, there is considerable interest in CYPs involved in triterpenoid and steroid metabolism in plants not only for improving our understanding of plant specialised metabolism, but also for synthetic biology and chemoenzymatic synthesis. In this review, we will provide an extensive overview of CYPs involved in tailoring of triterpenoids and steroids in plants. We will first introduce the nomenclature and mechanistic properties of these enzymes, before we describe the phylogenetic distribution of triterpenoid-modifying CYPs and summarise their reaction space. Lastly, we will highlight selected recent examples of multifunctional CYPs that catalyse particularly remarkable modifications of triterpenoids. We therefore hope to provide an up-to-date overview over these key enzymes in plant triterpenoid and steroid metabolism since the last comparable endeavour from Ghosh in 2017 [7]. In addition, readers might also be interested in other excellent reviews or resources providing a more general overview over plant CYPs or CYPs from other plant pathways [6,8-13].

Review

Nomenclature

Considering the enormous numbers of genes encoding cytochrome P450 monooxygenases in plants, a universal naming system is crucial to group related CYPs and to facilitate functional predictions. Hence, CYPs from all kingdoms are systematically named according to their amino acid identity by the cytochrome P450 nomenclature committee (David Nelson: dnelson@uthsc.edu). CYPs are grouped into clans, families, and subfamilies. The broadest hierarchy level is represented by clans, which comprise one or multiple families. An example CYP name is CYP51G1; here, "CYP51" designates the family, "G" refers to the subfamily within the CYP51 family, and "1" represents the isoform of CYPs in this subfamily [14]. Typically, all CYPs in the same subfamily share more than 55% amino acid sequence identity, and all CYPs in the same family more than 40%, although exceptions exist [15,16]. These thresholds also underline the remarkable sequence variety of CYPs, as even enzymes with only 60–70% amino acid identity can display almost identical biochemical activity.

Enzymatic mechanism

As monooxygenases, CYPs catalyse the transfer of a single oxygen atom from molecular oxygen to their substrates (Figure 1A). Decades of research on CYPs led to detailed insights into their mechanistic properties based on a variety of biochemical, biophysical and computational methods [17-21]. Key for the oxidative chemistry performed by CYPs is a heme prosthetic group that activates molecular oxygen using electrons from an electron donor such as NADPH. A central Fe(III) ion is coordinated by the heme porphyrine system as well as a cysteine thiolate ligand from the protein backbone (Figure 1B). The generally accepted catalytic cycle for hydroxylations is shown in Figure 1C.

In the resting state **A**, the central ferric ion is coordinated by six ligands, four from the porphyrin ring system, one cysteine thiolate group and an aqua ligand (water), resulting in an octahedral complex [17]. The oxidative reaction is initiated by displacement of the axial water molecule by the substrate (step 1), which pushes the Fe(III) ion out of the porphyrin plane in intermediate **B** [17]. This geometrical distortion promotes electron transfer from a reductase partner protein (step 2). The most common electron donors in plants are cytochrome P450 reductases which employ NADPH for the electron transfer, but several other electron transfer systems are known [22]. The reduced ferrous intermediate **C**, bearing an overall negative charge, can then efficiently bind molecular oxygen (step 3), leading to dioxygen adduct **D**. Transfer of an additional electron from a reducing partner such as cytochrome P450 reductase (step 4) generates peroxo intermediate **E**, which upon protonation (step 5) gives hydroperoxo intermediate **F**, also called compound 0. This nucleophilic and basic intermediate is prone to dehydration (step 6), leading to the strongly electrophilic and oxidising key intermediate **G**, which is commonly known as compound I (cpd I). Although there has been a lot of debate regarding the exact structure and electronic properties of compound I (intermediate **G**), it is now generally accepted as a ferryl (Fe(IV)) oxo species with a radical cation in the porphyrin system [18,23]. In the case of hydroxylations, the oxygen from compound I (intermediate **G**) can then be transferred by an oxygen rebound mechanism (steps 7 and 8) via the ferryl hydroxy intermediate **H**, also known as compound II. This leads to the hydroxylated product coordinated to ferric ion (intermediate **I**). Lastly, displacement of the product by water regenerates the resting state **A** (step 9). In addition to simple hydroxylations, slight deviations from this general mechanistic cycle can also lead to different reaction outcomes, e.g., rearrangements, desaturations or epoxidations. Multiple oxidation rounds, leading to aldehydes/ketones or carboxylic acids, are also commonly observed. Together, this versatile oxidative chemistry makes CYPs key enzymes in specialised metabolism in general [8,11], and crucial agents for the structural diversity of triterpenes.

Phylogenetic distribution of triterpenoid- and steroid-modifying plant cytochrome P450 monooxygenases

To assess the phylogenetic distribution of triterpenoid-modifying CYPs in comparison to other CYPs from plants, we collected a total of 149 CYPs from plants reported in the literature which are involved in triterpenoid or steroid metabolism

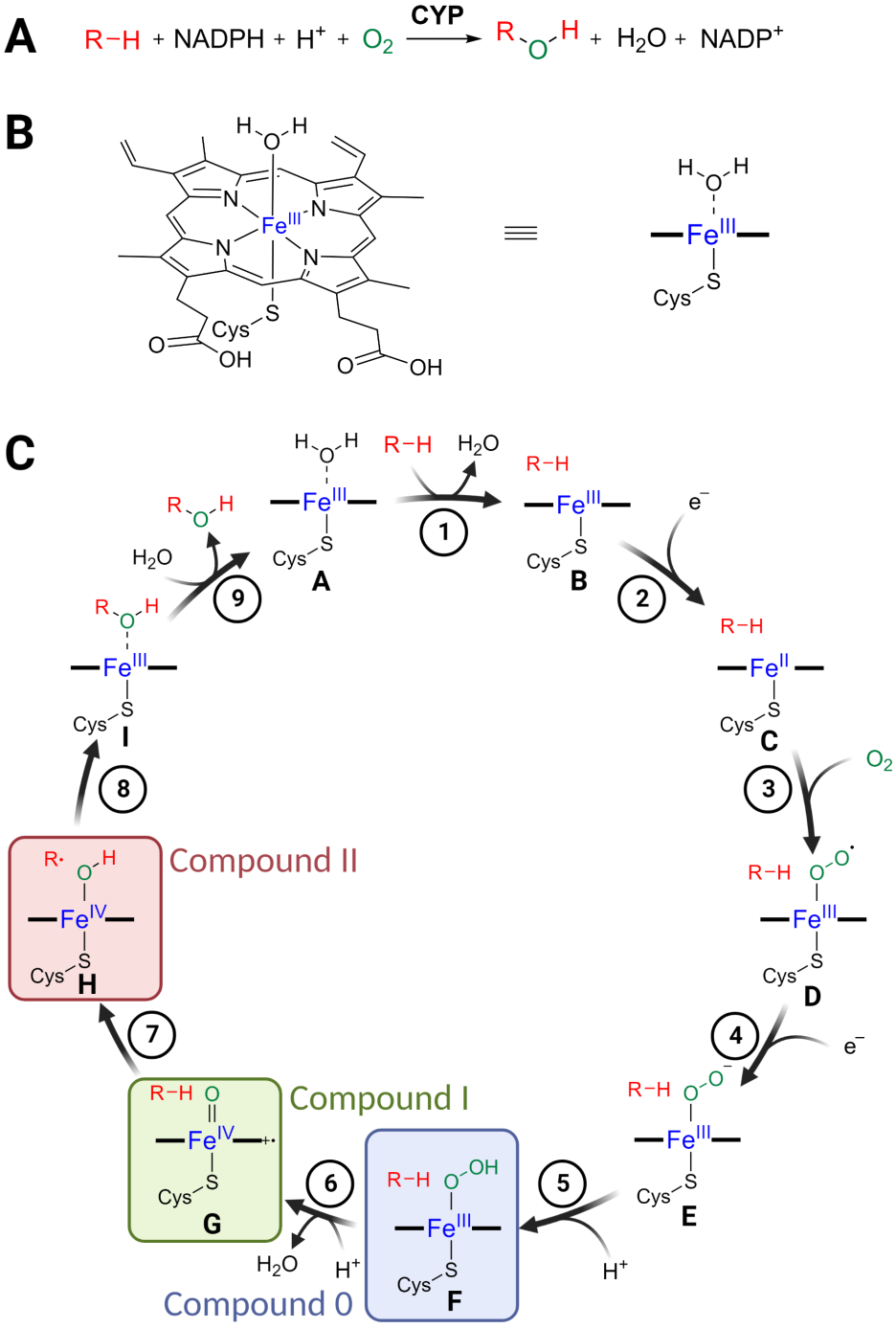


Figure 1: Enzyme function of cytochrome P450 monooxygenases (CYPs). A) Typical net reaction of CYPs, resulting in hydroxylation of a substrate. As monooxygenases, CYPs catalyse transfer of only one oxygen atom from molecular oxygen to their substrate. B) Heme prosthetic group containing the reactive Fe ion. Also shown is the abbreviated form of the cofactor used in the catalytic cycle. C) Catalytic cycle of CYPs. The key intermediates **F**, **G**, and **H**, called compound 0, compound I, and compound II, respectively, are highlighted. For details see main text. Figure 1 was created with BioRender.com. This content is not subject to CC BY 4.0.

(Table 1), and generated a neighbour-joining tree together with 266 non-triterpenoid CYPs with a different substrate scope (Figure 2). Notably, our analysis highlights that triterpenoid CYPs do not seem to occur randomly in various CYP clans and families; instead, certain clans and families represent "hotspots" of triterpenoid/steroid-modifying CYPs.

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids.

Name	Clan	Family	Species	Accession number	Scaffold	Substrate	Reaction	Product	Ref.
CYP51G1	51	51	*Sorghum bicolor*	XM_021461021 2.1	steroid	obtusifoliol	C14α demethylation	4α-methyl-5α-ergosta-8,14,24(28)-trien-3β-ol	[24]
CYP51G1	51	51	*Arabidopsis thaliana*	AB014459	steroid	obtusifoliol	C14α demethylation	4α-methyl-5α-ergosta-8,14,24(28)-trien-3β-ol	[25]
CYP51H10	51	51	*Avena strigosa*	DQ680852	pentacyclic oleanane	β-amyrin	C12–C13β epoxidation / C16 β hydroxlation	12,13-β-epoxy-16-β-hydroxy-amyrin	[1]
CYP51H14	51	51	*Brachypodium distachyon*	ON108677	pentacyclic triterpene	19-hydroxy-isoarborinol	C7 and C28 hydroxylation	7,19,28-trihydroxyisoarborinol	[26]
CYP51H15	51	51	*Brachypodium distachyon*	ON108678	pentacyclic triterpene	isoarborinol	C19 hydroxylation	19-hydroxy-isoarborinol	[26]
CYP51H16	51	51	*Brachypodium distachyon*	ON108679	pentacyclic triterpene	7,19,28-trihydroxyisoarborinol	C1 hydroxylation	1,7,19,28-tetrahydroxy-isoarborinol	[26]
CYP51H35	51	51	*Triticum aestivum*	ON108669	pentacyclic triterpene	isoarborinol	C19 hydroxylation	19-hydroxy-isoarborinol	[26]
CYP51H37	51	51	*Triticum aestivum*	ON108670	pentacyclic triterpene	19-hydroxy-isoarborinol	C25 hydroxylation and C2 oxidation	ellarinacin	[26]
CYP71A16	71	71	*Arabidopsis thaliana*	NM_123623.5	monocyclic triterpene aldehyde	marneral / marnerol	C23 hydroxylation	23-hydroxy-marneral / 23-hydroxy-marnerol	[27,28]
CYP71BQ5	71	71	*Melia azedarach*	MK803264.1	tirucallane triterpenoid	dihydroniloticin	C21 hydroxylation	melianol	[29]
CYP71CD2	71	71	*Melia azedarach*	MK803271	tirucallane triterpenoid	tirucalla-7,24-dien-3β-ol	C23 hydroxylation and C24–C25 epoxidation	dihydro-niloticin	[29]
CYP71D353	71	71	*Lotus japonicus*	KF460438	pentacyclic lupane	dihydrolupeol / 20-hydroxy-lupeol	C20 hydroxylation / C28 oxidation	20-hydroxy-lupeol / 20-hydroxy-betulinic acid	[30]
CYP71D443	71	71	*Ajuga reptans*	LC066937	steroid	3β-hydroxy-5β-cholestan-6-one	C22 hydroxylation	3β,22*R*-dihydroxy-5β-cholestan-6-one	[31]
CYP81AQ19	71	81	*Momordica charantia*	LC456843	tetracyclic triterpenoid	cucurbitadienol	C23α hydroxylation	cucurbita-5,24-dien-3,23α-diol	[32]
CYP81Q58	71	81	*Cucumis sativus*	KM655856	tetracyclic triterpenoid	19-hydroxy-cucurbitadienol	C25 hydroxylation / double bond shift	19,25-dihydroxy-cucurbita-dienol	[33]
CYP81Q59	71	81	*Cucumis melo*	Melo3C022375	tetracyclic triterpenoid	11-carbonyl-20β-hydroxy-cucurbitadienol	C2β hydroxylation	11-carbonyl-2β,20β-dihydroxy-cucurbita dienol	[34]
CYP82J17	71	82	*Trigonella foenum-graecum*	MK636709	steroid	16*S*-hydroxy-22-oxo-cholesterol	C27 hydroxy-lation / spiro-ketalisation	diosgenin	[35]

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids. (continued)

CYP93A220 / IaAO5	71	93	*Ilex asprella*	MZ508433	pentacyclic oleanane	β-amyrin	C24 oxidation	α-boswellic acid	[36]
CYP93E1	71	93	*Glycine max*	AB231332	pentacyclic oleanane	β-amyrin / sophoradiol	C24 hydroxylation	24-hydroxy-β-amyrin / soyasapogenol B	[37]
CYP93E2	71	93	*Medicago truncatula*	DQ335790	pentacyclic oleanane	β-amyrin	C24 hydroxylation	24-hydroxy-β-amyrin	[38]
CYP93E3	71	93	*Glycyrrhiza uralensis*	AB437320	pentacyclic oleanane	β-amyrin	C24 hydroxylation	24-hydroxy-β-amyrin	[39]
CYP93E4	71	93	*Arachis hypogaea*	KF906535	pentacyclic oleanane	β-amyrin	C24 hydroxylation	24-hydroxy-β-amyrin	[40]
CYP93E5	71	93	*Cicer arietinum*	KF906536	pentacyclic oleanane	β-amyrin	C24 hydroxylation	24-hydroxy-β-amyrin	[40]
CYP93E6	71	93	*Glycyrrhiza glabra*	KF906537	pentacyclic oleanane	β-amyrin	C24 hydroxylation	24-hydroxy-β-amyrin	[40]
CYP93E7	71	93	*Lens culinaris*	KF906538	pentacyclic oleanane	β-amyrin	C24 hydroxylation	24-hydroxy-β-amyrin	[40]
CYP93E8	71	93	*Pisum sativum*	KF906539	pentacyclic oleanane	β-amyrin	C24 hydroxylation	24-hydroxy-β-amyrin	[40]
CYP93E9	71	93	*Phaseolus vulgaris*	KF906540	pentacyclic oleanane	β-amyrin	C24 hydroxylation	24-hydroxy-β-amyrin	[40]
CYP705A1	71	705	*Arabidopsis thaliana*	NM_001341032.1	tricyclic triterpenoid	arabidiol	C15–C16 cleavage	14-apo-arabidiol	[28]
CYP705A5	71	705	*Arabidopsis thaliana*	NM_124173.3	tricyclic triterpenoid	7β-hydroxy-thalianol	C15–C16 desaturation	desaturated (C15–C16) 7β-hydroxy-thalianol	[41]
CYP712K1	71	712	*Tripterygium wilfordii*	MN621243	pentacyclic triterpenoid	friedelin	C29 oxidation	polpunonic acid and 29-hydroxy-friedelin	[42]
CYP712K2	71	712	*Tripterygium wilfordii*	MN621244	pentacyclic triterpenoid	friedelin	C29 oxidation	polpunonic acid and 29-hydroxy-friedelin	[42]
CYP712K3	71	712	*Tripterygium wilfordii*	MN621245	pentacyclic triterpenoid	friedelin	C29 oxidation	polpunonic acid and 29-hydroxy-friedelin	[42]
CYP712K4	71	712	*Maytenus ilicifolia*	MK829814	pentacyclic triterpenoid	friedelin	C29 oxidation	polpunonic acid or maytenoic acid	[43]
CYP72A61	72	72	*Medicago truncatula*	DQ335793	pentacyclic oleanane	24-hydroxy-β-amyrin	C22 hydroxylation	soyasapogenol B	[44]
CYP72A61v2	72	72	*Medicago truncatula*	XM_003605422	pentacyclic oleanane	24-hydroxy-β-amyrin	C22 hydroxylation	soyasapogenol B	[44]
CYP72A62v2	72	72	*Medicago truncatula*	AB558147	pentacyclic oleanane	β-amyrin	C29 oxidation	29-hydroxy-β-amyrin / epi-katonic acid	[45]
CYP72A63	72	72	*Medicago truncatula*	AB558146	pentacyclic oleanane	β-amyrin	C30 oxidation	11-deoxy-glycyrrhentinic acid	[46]
CYP72A64v2	72	72	*Medicago truncatula*	MK534548	pentacyclic oleanane	β-amyrin	C29 oxidation	29-hydroxy-β-amyrin / epi-katonic acid	[45]
CYP72A65v2	72	72	*Medicago truncatula*	XM_003628012.4	pentacyclic oleanane	β-amyrin	C21 hydroxylation	21-hydroxy-β-amyrin	[45]

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids. (continued)

CYP72A67	72	72	*Medicago truncatula*	DQ335780	pentacyclic oleanane	oleanolic acid / hederagenin / gypsogenic acid / gypsogenin	C2β hydroxylation	2β-hydroxy-oleanolic acid / bayogenin / medicagenic acid / 2β,3β-dihydroxy-olean-12-en-23-oxo-28-oic acid	[47,48]
CYP72A68	72	72	*Medicago truncatula*	DQ335782	pentacyclic oleanane	oleanolic acid / hederagenin / gypsogenin	C23 oxidation	hederagenin / gypsogenin / gypsogenic acid	[48]
CYP72A68v2	72	72	*Medicago truncatula*	XM_013608494.3	pentacyclic oleanane	oleanolic acid / hederagenin / gypsogenin	C23 oxidation	hederagenin / gypsogenin / gypsogenic acid	[44]
CYP72A69	72	72	*Glycine max*	LC143440	pentacyclic oleanane	soyasapogenol B	C21 hydroxylation	soyasapogenol A	[49]
CYP72A141	72	72	*Glycine max*	MK534532	pentacyclic oleanane	β-amyrin	C29 hydroxylation	29-hydroxy-β-amyrin	[45]
CYP72A154	72	72	*Glycyrrhiza uralensis*	AB558153	pentacyclic oleanane	β-amyrin / 11-oxo-β-amyrin	C30 oxidation	30-hydroxy-β-amyrin / glycyrrhetinic acid	[46]
CYP72A302	72	72	*Phaseolus vulgaris*	MK534537	pentacyclic oleanane	β-amyrin	C29 hydroxylation	29-hydroxy-β-amyrin	[45]
CYP72A397	72	72	*Kalopanax septemlobus*	KT150517	pentacyclic oleanane	oleanolic acid	C23 oxidation	hederagenin	[50]
CYP72A552	72	72	*Barbarea vulgaris*	MH252571	pentacyclic oleanane	oleanolic acid	C23 oxidation	hederagenin	[51]
CYP72A557	72	72	*Medicago truncatula*	MK534544	pentacyclic oleanane	β-amyrin	C21 hydroxylation	21-hydroxy-β-amyrin	[45]
CYP72A558	72	72	*Medicago truncatula*	MK534545	pentacyclic oleanane	β-amyrin	C21 hydroxylation	21-hydroxy-β-amyrin	[45]
CYP72A559	72	72	*Medicago truncatula*	MK534546	pentacyclic oleanane	β-amyrin	C21 hydroxylation	21-hydroxy-β-amyrin	[45]
CYP72A613	72	72	*Trigonella foenum-graecum*	MK636708	steroid	16*S*-hydroxy-22-oxo-cholesterol	C27 hydroxylation / spiro-ketalisation	diosgenin	[35]
CYP72A616	72	72	*Paris polyphylla*	MK636705	steroid	16*S*-hydroxy-22-oxo-cholesterol	C27 hydroxy-lation / spiro-ketalisation	diosgenin	[35]
CYP72A694	72	72	*Vigna angularis*	MK534538	pentacyclic oleanane	β-amyrin / 29-hydroxy-β-amyrin	C29 oxidation	29-hydroxy-β-amyrin / epi-katonic acid	[45]
CYP72A697	72	72	*Lotus japonicus*	MK534539	pentacyclic oleanane	β-amyrin	C29 hydroxylation	29-hydroxy-β-amyrin	[45]
CYP72A699	72	72	*Trifolium pratense*	MK534549	pentacyclic oleanane	β-amyrin / 29-hydroxy-β-amyrin	C29 oxidation	29-hydroxy-ß-amyrin / epi-katonic acid	[45]
CYP714E19	72	714	*Centella asiatica*	KT004520	pentacyclic oleanane / ursane	oleanolic acid / ursolic acid	C23 oxidation	hederagenin / 23-hydroxy-ursolic acid	[52]
CYP714E88 / IaAO4	72	714	*Ilex asprella*	MZ508437	pentacyclic oleanane / ursane	ursolic acid / oleanolic acid	C23 oxidation	23-carboxyl-ursolic acid / gypsogenic acid	[36]

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids. (continued)

CYP734A7	72	734	*Solanum lycopersicum*	AB223041	steroid	castasterone / 28-nor-castasterone / brassinolide	C26 hydroxylation	26-hydroxy-castasterone / 26-hydroxy-norcastasterone / 26-hydroxy-brassinolide	[53]
CYP749A63	72	749	*Crataegus pinnatifida*	MF596155	pentacyclic oleanane	oleanolic acid	C24 hydroxylation	4-epi-hederagenin	[54]
CYP85A1	85	85	*Arabidopsis thaliana*	AB035868	steroid	6-deoxoteasterone / 3-dehydro-6-deoxoteasterone / 6-deoxotyphasterol / 6-deoxocastasterone	C6 oxidation	teasterone / 3-dehydroteasterone / typhasterol / castasterone	[55]
CYP85A1	85	85	*Solanum lycopersicum*	U54770	steroid	6-deoxoteasterone / 3-dehydro-6-deoxoteasterone / 6-deoxotyphasterol / 6-deoxocastasterone	C6 oxidation	teasterone / 3-dehydroteasterone / typhasterol / castasterone	[56]
CYP85A2	85	85	*Arabidopsis thaliana*	AB087801	steroid	castasterone / 6-deoxocastasterone / 6-deoxotyphasterol / 3-dehydro-6-deoxoteasterone	Baeyer-Villiger oxidation / C6 oxidation	brassinolide / castasterone / typhasterol / 3-dehydroteasterone	[57,58]
CYP85A3	85	85	*Solanum lycopersicum*	AB190445	steroid	6-deoxocastasterone / castasterone	Baeyer-Villiger oxidation / C6 oxidation	castasterone / brassinolide	[58]
CYP87D16	85	87	*Maesa lanceolata*	KF318735	pentacyclic oleanane	β-amyrin	C16α hydroxylation	16α-hydroxy-β-amyrin	[59]
CYP87D18	85	87	*Siraitia grosvenorii*	HQ128570	tetracyclic triterpenoid	cucurbitadienol / 11α-hydroxy-cucurbitadienol/ 24,25-di-hydroxy-cucurbitadienol	C11 oxidation	11α-hydroxy-cucurbitadienol / 11-oxo-cucurbitadienol / mogrol	[34]
CYP87D20	85	87	*Cucumis sativus*	Csa1G044890	tetracyclic triterpenoid	cucurbitadienol / 11-oxo-cucurbitadienol	C11 oxidation / C20β hydroxylation	11-oxocucurbitadienol / 11-carbonyl-20β-hydroxy-cucurbitadienol	[34]
CYP88D6	85	88	*Glycyrrhiza uralensis*	AB433179	pentacyclic oleanane	β-amyrin	C11 oxidation	11-oxo-β-amyrin	[39]
CYP88L2	85	88	*Cucumis sativus*	Csa3G903540	tetracyclic triterpenoid	cucurbitadienol / 11-oxo-cucurbitadienol	C19 hydroxylation	19-hydroxy-cucurbitadienol	[34]
CYP88L7	85	88	*Momordica charantia*	LC456844	tetracyclic triterpenoid	cucurbitadienol	C19 hydroxylation, C5 and C19 ether bridge	cucurbita-5,24-dien-3β,19-diol and 5β,19-epoxy-cucurbita-6,24-dien-3β-ol	[32]

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids. (continued)

CYP88L8	85	88	*Momordica charantia*	LC456845	tetracyclic triterpenoid	cucurbitadienol	C7β hydroxylation	cucurbita-5,24-dien-3β,7β-diol	[32]
CYP90A1	85	90	*Arabidopsis thaliana*	X87367	steroid	6-deoxocathasterone / 6-deoxoteasterone / 22*S*-hydroxycampesterol / 22*R*,23*R*-dihydroxycampesterol	C3 oxidation	22*S*-hydroxy-5α-campestan-3-one / 3-dehydro-6-deoxoteasterone/ 22*S*-hydroxycampest-4-en-3-one / 22*R*,23*R*-dihydroxycampest-4-en-3-one	[60]
CYP90B1	85	90	*Arabidopsis thaliana*	NM_114926.4	steroid	campesterol / 24*R*-ergost-4-en-3-one / 24*R*-5α-ergostan-3-one / campestanol / 6-oxocampestanol	C22 hydroxylation	22*S*-hydroxycampesterol / 22*S*-hydroxy-24*R*-ergost-4-en-3-one / 22*S*-hydroxy-24*R*- 5α-ergostan-3-one / 6-deoxocathasterone / cathasterone	[61]
CYP90B2	85	90	*Oryza sativa*	AB206579	steroid	campesterol / campestanol	C22 hydroxylation	22*S*-hydroxycampesterol / 6-deoxocathasterone	[62]
CYP90B3	85	90	*Solanum lycopersicum*	NM_001279330.2	steroid	campesterol / 24*R*-ergost-4-en-3-one / 24*R*-5α-ergostan-3-one / campestanol	C22 hydroxylation	22-hydroxycampesterol / 22*S*-hydroxy-24*R*-ergost-4-en-3-one / 22*S*-hydroxy-24R-5α-ergostan-3-one / 6-deoxocathasterone	[63]
CYP90B27	85	90	*Veratrum californicum*	KJ869252	steroid	cholesterol / 26-hydroxycholesterol / 7ß-hydroxycholesterol	C22 hydroxylation	22*R*-hydroxycholesterol / 22,26-dihydroxycholesterol / 7ß,22-dihydroxycholesterol	[64]
CYP90B27	85	90	*Paris polyphylla*	KX904822	steroid	cholesterol	C22 hydroxylation	22*R*-hydroxycholesterol	[65]
CYP90B50	85	90	*Trigonella foenum-graecum*	MK636707	steroid	cholesterol	C22*R*, C16 dihydroxylation	16*S*,22*R*-dihydroxycholesterol	[35]
CYP90B51	85	90	*Trigonella foenum-graecum*	MK636706	steroid	cholesterol	C22*S* hydroxylation / C22*R* hydroxylation	22*S*-hydroxycholesterol / 22*R*-hydroxycholesterol	[66]
CYP90B52	85	90	*Paris polyphylla*	MK636701	steroid	cholesterol	C22*S* hydroxylation	22*S*-hydroxycholesterol	[35]
CYP90B71	85	90	*Dioscorea zingiberensis*	MN829441	steroid	cholesterol	C22*R* hydroxylation	22*R*-hydroxycholesterol	[66]

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids. (continued)

CYP90C1	85	90	*Arabidopsis thaliana*	NM_001342408.1	steroid	22*S*-hydroxy-24*R*-5α-ergostan-3-one / 3-epi-6-deoxocathasterone	C23 hydroxylation	3-dehydro-6-deoxoteasterone / 6-deoxotyphasterol	[67]
CYP90D1	85	90	*Arabidopsis thaliana*	NM_112223	steroid	22*S*-hydroxy-24*R*-5α-ergostan-3-one / 3-epi-6-deoxocathasterone	C23 hydroxylation	3-dehydro-6-deoxoteasterone / 6-deoxotyphasterol	[67]
CYP90D2	85	90	*Oryza sativa*	NM_001409071	steroid	22*S*-hydroxy-24*R*-5α-ergostan-3-one / 3-epi-6-deoxocathasterone	C23 hydroxylation	3-dehydro-6-deoxoteasterone / 6-deoxotyphasterol	[68]
CYP90D3	85	90	*Oryza sativa*	AAT44310	steroid	22*S*-hydroxy-24*R*-5α-ergostan-3-one / 3-epi-6-deoxocathasterone	C23 hydroxylation	3-dehydro-6-deoxoteasterone / 6-deoxotyphasterol	[68]
CYP90G1v1	85	90	*Veratrum californicum*	KJ869258	steroid	22*R*-hydroxycholesterol / 22,26-dihydroxycholesterol / 22-hydroxy-26-aminocholesterol	C22 hydroxylation	22-ketocholesterol / 22-keto-26-hydroxycholesterol / verazine	[64]
CYP90G1v2	85	90	*Veratrum californicum*	KJ869261	steroid	22*R*-hydroxycholesterol / 22,26-dihydroxycholesterol / 22-hydroxy-26-aminocholesterol	C22 hydroxylation	22-ketocholesterol / 22-keto-26-hydroxycholesterol / verazine	[64]
CYP90G1v3	85	90	*Veratrum californicum*	KJ869260	steroid	22*R*-hydroxycholesterol / 22,26-dihydroxycholesterol / 22-hydroxy-26-aminocholesterol	C22 hydroxylation	22-ketocholesterol / 22-keto-26-hydroxycholesterol / verazine	[64]
CYP90G4	85	90	*Paris polyphylla*	MK636702	steroid	22*R*-hydroxycholesterol	C16 oxidation	16*S*,22*R*-dihydroxycholesterol	[66]
CYP90G6	85	90	*Dioscorea zingiberensis*	MN829442	steroid	22*R*-hydroxycholesterol	C16 oxidation	16*S*,22*R*-dihydroxycholesterol	[66]
CYP708A15	85	708	*Iberis amara*	MW514550	tetracyclic triterpenoid	16β-hydroxycucurbitadienol	C22 hydroxylation	16β,22-dihydroxycucurbitadienol	[69]
CYP708A15v2	85	708	*Iberis amara*	MW514551	tetracyclic triterpenoid	16β-hydroxycucurbitadienol	C22 hydroxylation	16β,22-dihydroxycucurbitadienol	[69]
CYP708A16	85	708	*Iberis amara*	MW514556	tetracyclic triterpenoid	cucurbitadienol	C16 hydroxylation	16β-hydroxycucurbitadienol	[69]

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids. (continued)

CYP708A2	85	708	*Arabidopsis thaliana*	NM_001344756.1	tricyclic triterpenoid	thalianol	C7β hydroxylation	7β-hydroxy-thalianol	[41]
CYP716A1	85	716	*Arabidopsis thaliana*	NM_123002.2	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / lupeol	C28 oxidation	ursolic acid / oleanolic acid / betulin	[70]
CYP716A2	85	716	*Arabidopsis thaliana*	LC106013.1	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / lupeol	C16/C22α/ C28 hydroxylation	uvaol / C22α-hydroxy-β-amyrin / erythrodiol / betulin	[70]
CYP716A12	85	716	*Medicago truncatula*	FN995113	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / betulin	C28 oxidation	ursolic acid / oleanolic acid / betulinic acid	[71]
CYP716A14v2	85	716	*Artemisia annua*	KF309251	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C3 oxidation	α-amyrone / β-amyrone	[72]
CYP716A15	85	716	*Vitis vinifera*	AB619802	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / betulin	C28 oxidation	ursolic acid / oleanolic acid / betulinic acid	[73]
CYP716A17	85	716	*Vitis vinifera*	AB619803	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[73]
CYP716A44	85	716	*Solanum lycopersicum*	XM_004239248.4	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C28 oxidation	ursolic acid / oleanolic acid	[74]
CYP716A46	85	716	*Solanum lycopersicum*	XM_004243858	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C28 oxidation	ursolic acid / oleanolic acid	[74]
CYP716A51	85	716	*Lotus japonicus*	AB706297	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / lupeol	C28 oxidation	ursolic acid / oleanolic acid / betulinic acid	[75]
CYP716A52v2	85	716	*Panax ginseng*	JX036032	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[76]
CYP716A75	85	716	*Maesa lanceolata*	KF318733	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[59]
CYP716A78	85	716	*Chenopodium quinoa*	KX343075	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[77]
CYP716A79	85	716	*Chenopodium quinoa*	KX343076	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[77]
CYP716A80	85	716	*Barbarea vulgaris*	KP795926	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / lupeol	C28 oxidation	ursolic acid / oleanolic acid / betulinic acid	[78]
CYP716A81	85	716	*Barbarea vulgaris*	KP795925	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / lupeol	C28 oxidation	ursolic acid / oleanolic acid / betulinic acid	[78]
CYP716A83	85	716	*Centella asiatica*	KU878849	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C28 oxidation	ursolic acid / oleanolic acid	[79]
CYP716A86	85	716	*Centella asiatica*	KU878848	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[79]
CYP716A94	85	716	*Kalopanax septemlobus*	KT150521	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[50]
CYP716A110	85	716	*Aquilegia coerulea*	KU878864	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[79]

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids. (continued)

CYP716A111	85	716	*Aquilegia coerulea*	KY047600	pentacyclic oleanane	β-amyrin	C16β hydroxylation	16β-hydroxy-β-amyrin	[79]
CYP716A113	85	716	*Aquilegia coerulea*	KU878866	tetracyclic triterpenoid	cycloartenol	unknown regio-selectivity	hydroxy-cyclo-artenol, performs non-specific reaction of endogenous yeast compounds	[79]
CYP716A140	85	716	*Platycodon grandiflorus*	KU878853	pentacyclic oleanane / ursane	β-amyrin / 16β-hydroxy-β-amyrin / 12,13α-epoxy-β-amyrin	C28 oxidation	oleanolic acid / 16β-hydroxy-oleanolic acid	[79]
CYP716A140v2	85	716	*Platycodon grandiflorus*	LC209199	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[80]
CYP716A141	85	716	*Platycodon grandiflorus*	KU878855	pentacyclic oleanane	β-amyrin / oleanolic acid	C28 oxidation / C16β hydroxylation	oleanolic acid / 16β-hydroxy-oleanolic acid	[79,80]
CYP716A154	85	716	*Catharanthus roseus*	JN565975	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / betulin	C28 oxidation	ursolic acid / oleanolic acid / betulinic acid	[81]
CYP716A155	85	716	*Rosmarinus officinalis*	MK592859	pentacyclic lupane	lupeol	C28 oxidation	betulinic acid	[82]
CYP716A175	85	716	*Malus domestica*	XM_008392874	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / lupeol / germanicol	C28 oxidation	ursolic acid / oleanolic acid / betulinic acid / morolic acid	[83]
CYP716A179	85	716	*Glycyrrhiza uralensis*	LC157867	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / betulin	C28 oxidation / C22α hydroxylation	ursolic acid / C22α-hydroxy-amyrin / oleanolic aicd / betulinic acid	[84]
CYP716A180	85	716	*Betula platyphylla*	KJ452328	pentacyclic lupane	lupeol	C28 oxidation	betulinic acid	[85]
CYP716A210 / IaAO1	85	716	*Ilex asprella*	MK994507	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C28 oxidation	ursolic acid / oleanolic acid	[86]
CYP716A244	85	716	*Eleutherococcus senticosus*	KX354739	pentacyclic oleanane	β-amyrin	C28 oxidation	oleanolic acid	[87]
CYP716A252	85	716	*Ocimum basilicum*	JQ958967	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C28 oxidation	ursolic acid / oleanolic acid	[88]
CYP716A253	85	716	*Ocimum basilicum*	JQ958968	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C28 oxidation	ursolic acid / oleanolic acid	[88]
CYP716A265	85	716	*Lagerstroemia speciosa*	MG708187	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / lupeol	C28 oxidation	ursolic acid / oleanolic acid / betulinic acid	[89]
CYP716A266	85	716	*Lagerstroemia speciosa*	MG708188	pentacyclic ursane / oleanane / lupane	α-amyrin / β-amyrin / lupeol	C28 oxidation	ursolic acid / oleanolic acid / betulinic acid	[89]

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids. (continued)

CYP716C11	85	716	*Centella asiatica*	KU878852	pentacyclic oleanane / ursane	oleanolic acid / ursolic acid / 6β-hydroxy-oleanolic acid	C2α hydroxylation	maslinic acid / 2α-hydroxy-ursolic acid / 6β-hydroxy-maslinic acid	[79]
CYP716C49	85	716	*Crataegus pinnatifida*	MF120282	pentacyclic oleanane / ursane / lupane	oleanolic acid / ursolic acid / betulinic acid	C2α hydroxylation	maslinic acid / corosolic acid / alphitolic acid	[54]
CYP716C55	85	716	*Lagerstroemia speciosa*	MG708191	pentacyclic ursane / oleanane	ursolic acid / oleanolic acid	C2α hydroxylation	corosolic acid / maslinic acid	[89]
CYP716E26	85	716	*Solanum lycopersicum*	XM_004241773	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C6β hydroxylation	6β-hydroxy-α-amyrin / daturadiol	[74]
CYP716E41	85	716	*Centella asiatica*	KU878851	pentacyclic oleanane / ursane	oleanolic acid / ursolic acid / maslinic acid	C6β hydroxylation	6β-hydroxy-oleanolic acid / 6β-hydroxy-ursolic acid / 6β-hydroxy-maslinic acid	[79]
CYP716S1	85	716	*Panax ginseng*	JX036031	tetracyclic triterpene	protopanaxadiol	C6 hydroxylation	protopanaxatriol	[76]
CYP716S5	85	716	*Platycodon grandiflorus*	KU878856	pentacyclic oleanane	β-amyrin / oleanolic acid	C12-C13α epoxidation	C12-C13α-epoxy-β-amyrin / C12-C13α-epoxy-oleanolic acid	[79]
CYP716U1	85	716	*Panax ginseng*	JN604536	tetracyclic triterpene	dammarenediol-II	C12 hydroxylation	protopanaxadiol	[76]
CYP716Y1	85	716	*Bupleurum falcatum*	KC963423	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C16α hydroxylation	16α-hydroxy-α-amyrin / 16α-hydroxy-β-amyrin	[38]
IaAO2	85	716	*Ilex asprella*	OL604227	pentacyclic ursane / oleanane	α-amyrin / β-amyrin	C28 oxidation	ursolic acid / oleanolic acid	[36]
CYP724A1	85	724	*Arabidopsis thaliana*	NM_001343334.1	steroid	possibly brassinosteroids	C22 hydroxylation		[90]
CYP724B1	85	724	*Oryza sativa*	AB158759	steroid	campesterol / campestanol	C22 hydroxylation	22*S*-hydroxy-campesterol / 6-deoxo-cathasterone	[62]
CYP724B2	85	724	*Solanum lycopersicum*	XM_004243170	steroid	campesterol / 24*R*-ergost-4-en-3-one / 24*R*-5α-ergostan-3-one / campestanol	C22 hydroxylation	22-hydroxy-campesterol / 22*S*-hydroxy-24*R*-ergost-4-en-3-one / 22*S*-hydroxy-24*R*-5α-ergostan-3-one / 6-deoxo-cathasterone	[63]
CYP94D108	86	94	*Paris polyphylla*	MK636703	steroid	16*S*-hydroxy-22-oxo-cholesterol	C27 hydroxylation / spiroketalisation	diosgenin	[35]
CYP94D109	86	94	*Paris polyphylla*	MK636704	steroid	16*S*-hydroxy-22-oxo-cholesterol	C27 hydroxylation / spiroketalisation	diosgenin	[35]

Table 1: List of characterised plant cytochrome P450 monooxygenases (CYPs) modifying triterpenoids or steroids. (continued)

CYP94N1	86	94	*Veratrum californicum*	KJ869255	steroid	22*R*-hydroxy-cholesterol	C26 hydroxylation	22,26-di-hydroxy-cholesterol and 22-hydroxy-cholesterol-26-al	[64]
CYP710A1	710	710	*Arabidopsis thaliana*	AB219423	steroid	β-sitosterol	C22 desaturation	stigmasterol	[91]
CYP710A2	710	710	*Arabidopsis thaliana*	AB233425	steroid	β-sitosterol / 24-epi-campesterol	C22 desaturation	stigmasterol/ brassicaster-ol	[91]
CYP710A4	710	710	*Arabidopsis thaliana*	NM_128444.2	steroid	β-sitosterol	C22 desaturation	stigmasterol	[91]
CYP710A11	710	710	*Solanum lycoper-sicum*	NM_001247585.2	steroid	β-sitosterol	C22 desaturation	stigmasterol	[91]

Probably the most well-known example of a triterpenoid-biased CYP family are the CYP716s (part of the CYP85 clan) [79], but also other families of the CYP85 clan such as CYP87, CYP85, or CYP90 contain mostly triterpene-modifying CYPs to date. The small clans CYP51 and CYP710 are other important examples of groups with a high preference for triterpenoid substrates. The highly diverse CYP71 clan, in contrast, only contains a few triterpene-modifying CYPs, particularly in the families CYP93, CYP712 and CYP705. The CYP72 family (CYP72 clan) also contains several known representatives. In other clans, however, not a single triterpene-modifying CYP has been identified so far, for example CYP97, CYP74, or CYP711.

The discovery of biosynthetic genes in plants often involves the screening of large pools of gene candidates derived from sequencing studies [93-96]. Hence, efficient approaches are needed to select the most promising gene candidates, particularly for large gene families such as CYPs. Our summarised phylogenetic distribution of known triterpenoid-modifying CYPs therefore might facilitate the discovery of new CYPs in triterpenoid and steroid pathways in plants by highlighting CYP families with a known propensity to participate in these pathways.

Major reaction types of triterpenoid- and steroid-modifying CYPs

The basic polycyclic skeletons of triterpenoids and steroids are created by oxidosqualene cyclases (OSCs) from the universal substrate 2,3-oxidosqualene [5]. As different folding modes (chair–boat–chair vs chair–chair–chair) and different ring sizes can occur during this cyclisation cascade, resulting triterpene and sterol scaffolds have drastically different three-dimensional shapes. For this reason, CYPs are typically specific to a certain group of triterpenoid scaffolds. Hence, we summarised our list of 149 triterpenoid/steroid CYPs (Table 1) according to their target scaffold.

Figure 3 covers plant CYPs acting on steroid, cucurbitacin, or simple tetracyclic triterpenoid scaffolds. Important scaffolds here are campesterol (**1**), β-sitosterol (**2**), cholesterol (**3**), cucurbitadienol (**4**), and dammarenediol-II (**5**). Not surprisingly, CYPs involved in the biosynthesis of essential sterols in plants are highly conserved and play a crucial role in their growth and development. For example, members of the CYP710A subfamily were characterised as C22 desaturases in *Arabidopsis* and tomato [91,97]. Three CYPs, CYP710A1, CYP710A2 and CYP710A4 were identified in *Arabidopsis* and CYP710A11 was identified in tomato. All four CYPs could produce stigmasterol from β-sitosterol (**2**) in enzyme assays performed in vitro. However, *Arabidopsis* CYP710A2 showed substrate flexibility towards campesterol (**1**) epimers and could also produce brassicasterol from 24-epicampesterol in vitro. Enzymes of the CYP51G subfamily (CYP51 clan) function as sterol 14α-demethylases in green plants [24,25,98]. These enzymes catalyse oxidation of the C14α methyl group to trigger elimination of formic acid [24,25]. The sister subfamily CYP51H, on the other hand, is only found in monocots. AsCYP51H10 from *Avena sativa* (oat) is a multifunctional CYP that performs hydroxylation and epoxidation reactions of the β-amyrin (**6**) scaffold to produce 12,13β-epoxy-16β-hydroxy-β-amyrin [1,99]. Thus, CYP51H10 is an example of a neofunctionalised CYP recruited from primary sterol metabolism.

Two members of the CYP87D subfamily decorate the tetracyclic scaffold in plants from the Cucurbitaceae family (Figure 3B). CYP87D18 (CYP85 clan) was identified as a multifunctional C11 oxidase involved in the biosynthetic pathway of mogrosides. Mogrosides, isolated from ripe fruits of

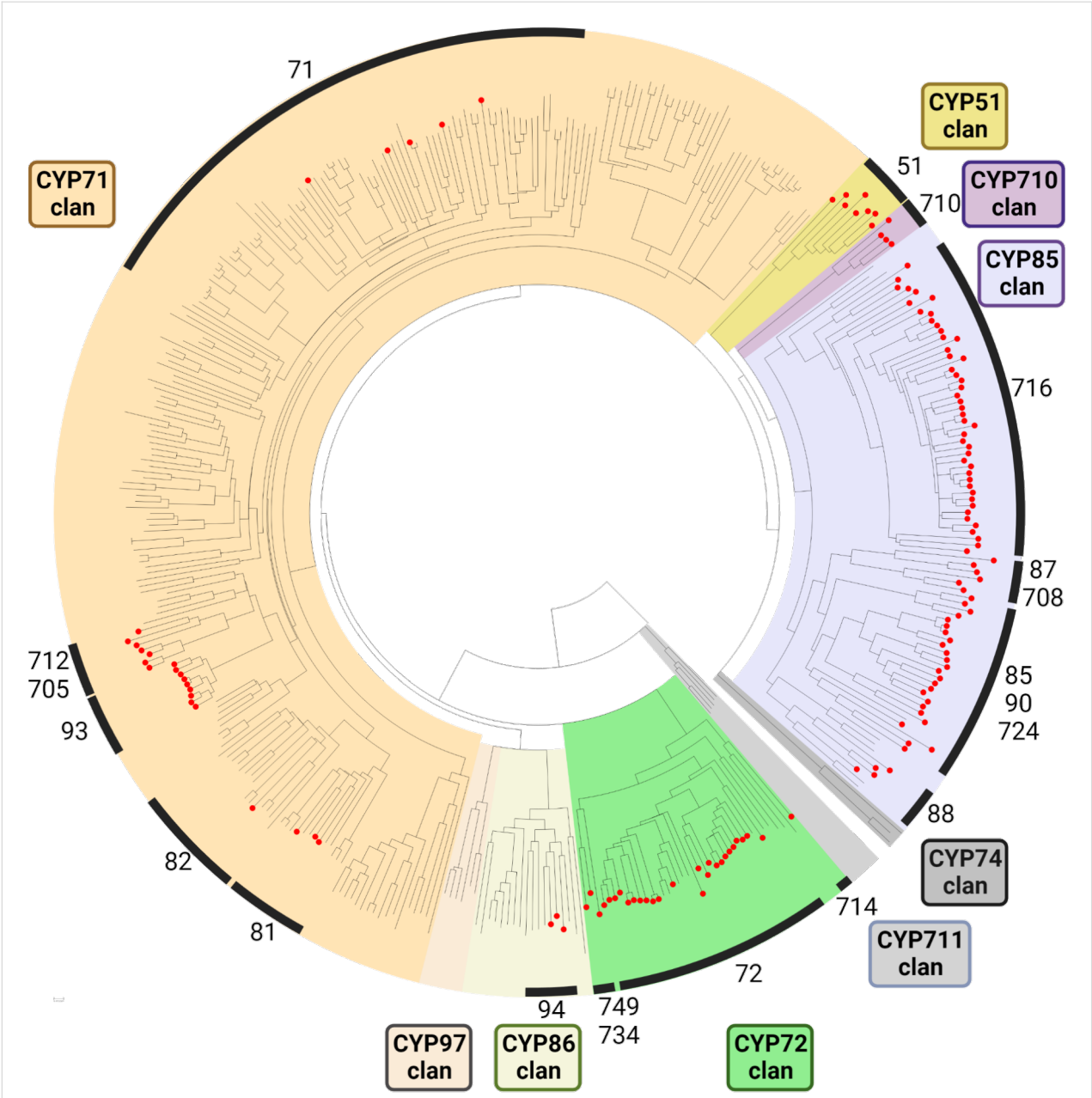


Figure 2: Phylogenetic distribution of CYPs acting on triterpenoid and steroid scaffolds (red nodes) compared to other CYPs from higher plants [15]. Numbers and black bars mark CYP families containing CYPs known to act on triterpenoids and steroids. Amino acid sequences (149 triterpenoid CYPs, 266 non-triterpenoid CYPs) were aligned using MUSCLE [92], and a neighbour-joining consensus tree of 1,000 bootstrap replicates was generated using the Jukes–Cantor model. The final tree was visualised in Python using the ete3 package. Annotations in Figure 2 were created with BioRender.com. This content is not subject to CC BY 4.0. A high resolution version with tip labels is available as Supporting Information File 1.

Siraitia grosvenorii (Cucurbitaceae) are glycosylated triterpenoid saponins with rare C24 and C25 hydroxylation [100]. Based on feeding assays in yeast it was found that CYP87D18 catalyses a two-step sequential C11 oxidation of cucurbitadienol (**4**) to 11-hydroxycucurbitadienol and 11-oxo-cucurbitadienol [101]. CYP87D18 also catalysed C11 hydroxylation of *trans*-24,25-dihydroxycucurbitadienol to form trihydroxylated mogrol in yeast [102].

CYPs acting on pentacyclic 6-6-6-6-6 triterpenes, which include the extremely important and widespread scaffolds β-amyrin (**6**), α-amyrin (**7**), and friedelin (**8**), are summarised in Figure 4. Of particular relevance in this area is the CYP716 family, which plays a central role in the diversification of triterpenoids in eudicots [79]. Members of the CYP716A subfamily were mostly identified as C28 oxidases that catalyse three-step oxidation of α-amyrin (**7**), β-amyrin (**6**) and lupeol (**10**) to ursolic

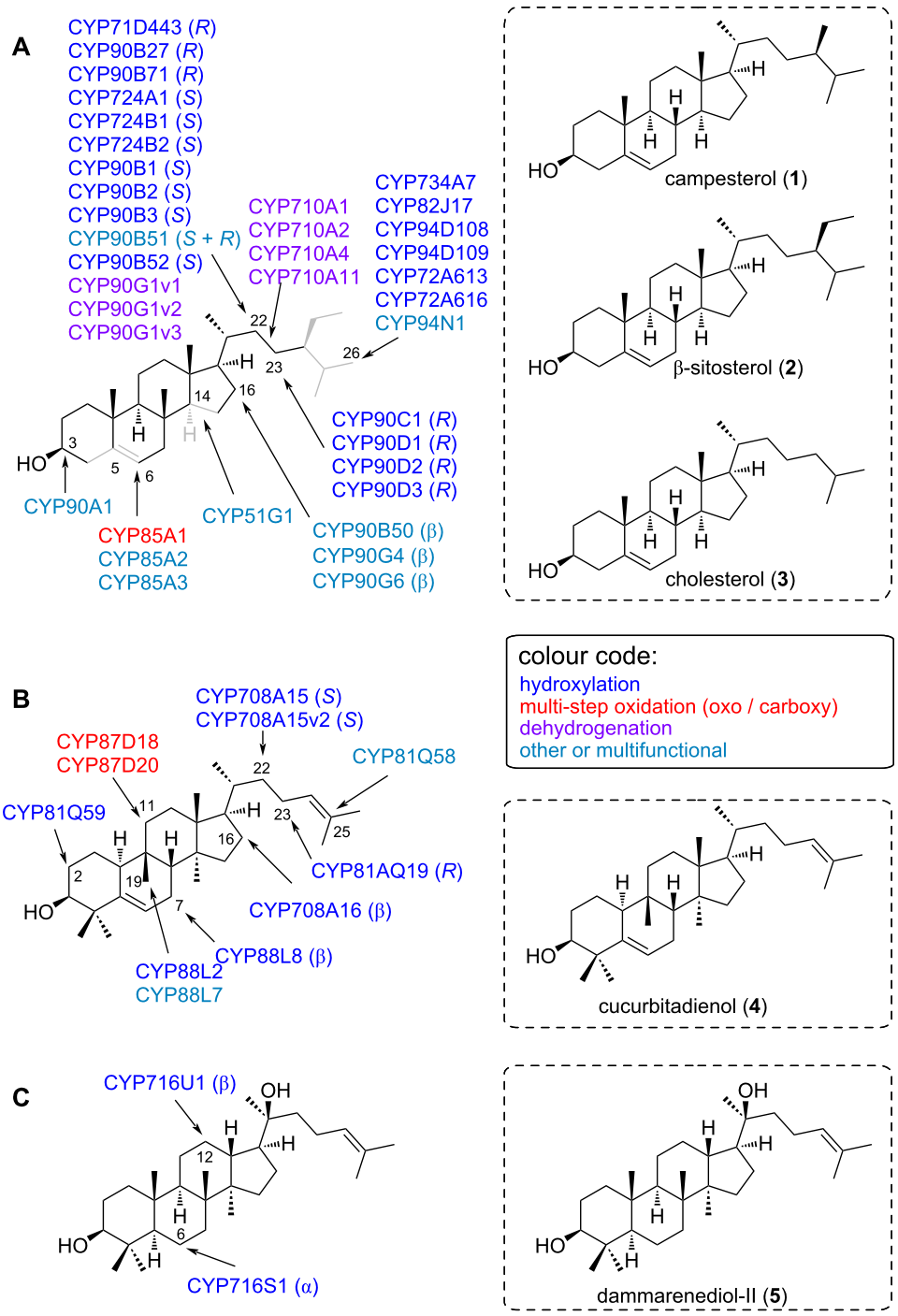


Figure 3: CYPs modifying steroid (A), cucurbitacin steroid (B) and tetracyclic triterpene (C) backbones. Substructures in grey indicate regions where major structural differences occur between different substrates of the same class. Representative substrate skeletons (not exact substrates) are shown in the dotted boxes. For exact substrate specificity see Table 1.

acid, oleanolic acid, and betulinic acid, respectively [71,73]. Nonetheless, other CYP716 enzymes have evolved to perform a wider range of modifications of triterpenoids; several CYP716 enzymes were found to catalyse C3 oxidation of α-amyrin (**7**) and β-amyrin (**6**), C16α oxidation of β-amyrin (**6**), or C22α oxidation of α-amyrin (**7**) [40,72,79]. Some members even act on triterpenoid scaffolds other than the 6-6-6-6-6 pentacyclic triterpenes. For example, two CYP716 enzymes from *Panax ginseng*

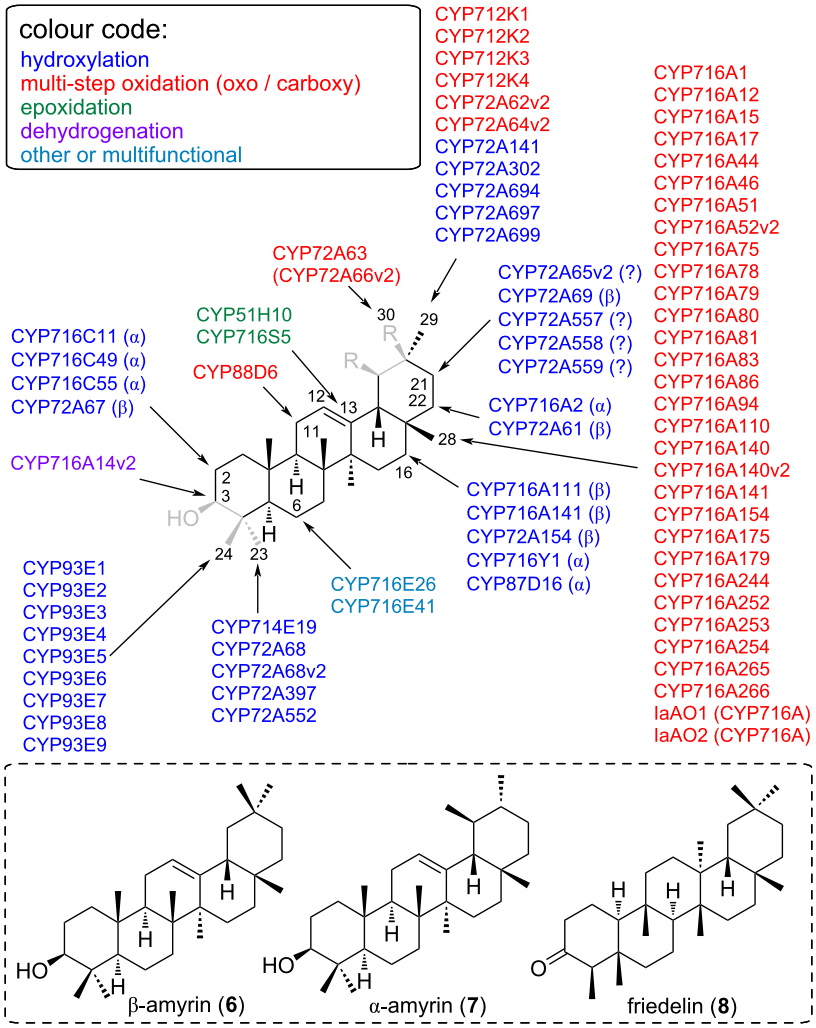


Figure 4: CYPs modifying pentacyclic 6-6-6-6-6 triterpenes. Substructures in grey indicate regions where major structural differences occur between different substrates of the same class. Representative substrate skeletons (not exact substrates) are shown in the dotted boxes. For exact substrate specificity see Table 1.

act on tetracyclic scaffolds; CYP716U1 hydroxylates dammarenediol-II (**5**) to protopanaxadiol, and CYP716S1 hydroxylates the C6 of protopanaxadiol to form protopanaxatriol [76,103]. CYP716A113v1 from *Aquilegia coerulea* hydroxylates cycloartenol with unknown regiospecificity when expressed in a yeast strain harbouring a tomato cycloartenol synthase gene [79].

CYP712 family members (clan 71) were first identified in the biosynthetic pathway of nor-triterpenoid celastrol, a potent anti-obesity metabolite [42,43]. In two independent studies, transcriptome mining and functional studies in *Nicotiana benthamiana* were used to identify the CYPs CYP712K1, CYP712K2, CYP712K3, and CYP712K4 capable of oxidising friedelin (**8**) into polpunonic acid via an aldehyde intermediate [42,43].

Members of the CYP93E subfamily are restricted to legumes and are involved in the biosynthesis of triterpenoid saponins. So far, nine CYP93E members were identified from different legume species [37,40]. All of these perform C24 hydroxylation of β-amyrin (**6**) to form 24-hydroxy-β-amyrin. CYP93E1 also catalyses the conversion of sophoradiol to soyasapogenol B [37,40,46]. Members of other CYP93 subfamilies (CYP93A, B, C and G) are ubiquitous in flowering plants and are mostly involved in flavonoid biosynthesis [25,26].

Lastly, CYPs acting on either pentacyclic 6-6-6-6-5 scaffolds, such as isoarborinol (**9**) or lupeol (**10**), or on unusual triterpene scaffolds such as arabidiol (**11**) or thalianol (**12**) are grouped in Figure 5. Enzymes from the CYP705 and CYP708 family catalyse Brassicaceae-specific reactions. The corresponding genes were found in operon-like gene clusters and catalyse the modifi-

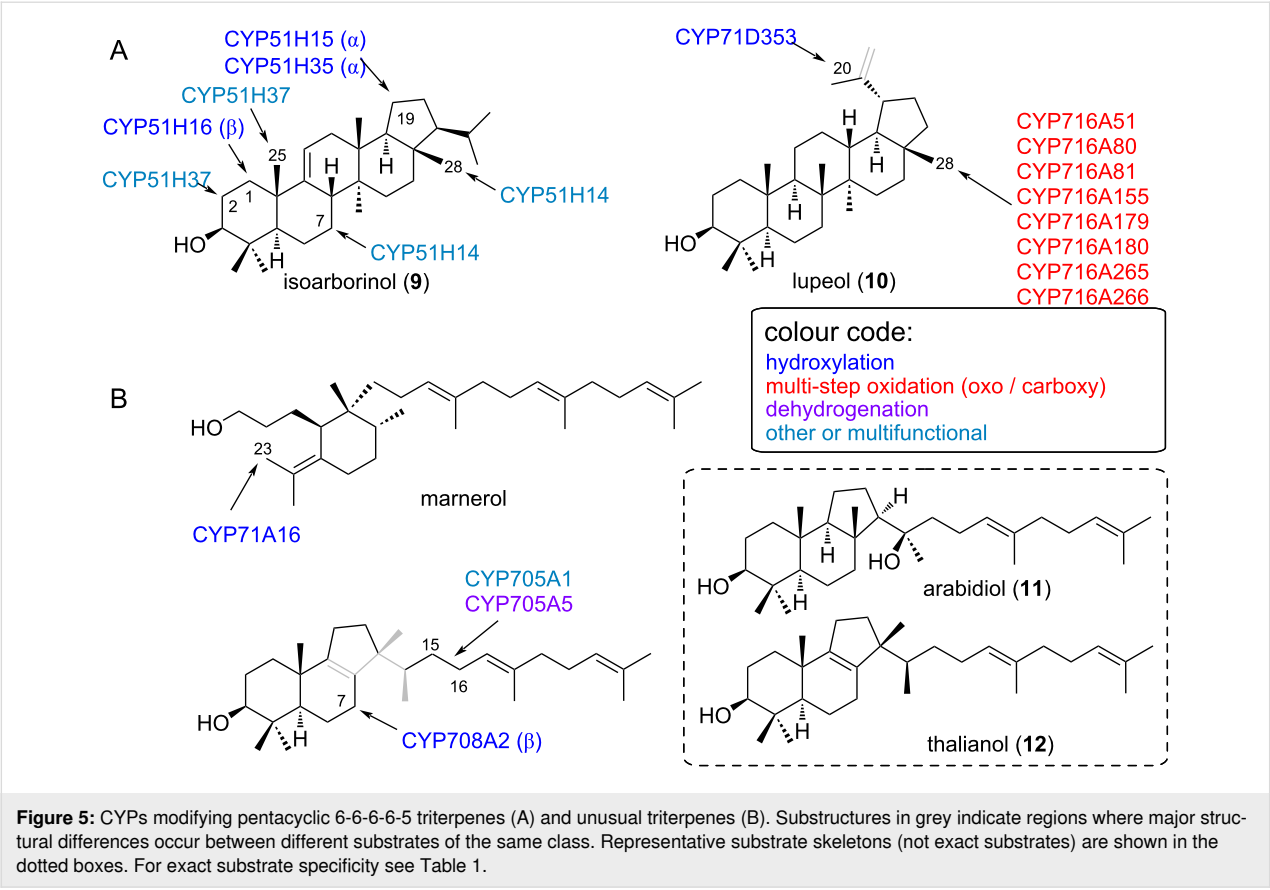


Figure 5: CYPs modifying pentacyclic 6-6-6-6-5 triterpenes (A) and unusual triterpenes (B). Substructures in grey indicate regions where major structural differences occur between different substrates of the same class. Representative substrate skeletons (not exact substrates) are shown in the dotted boxes. For exact substrate specificity see Table 1.

cation of monocyclic marnerol and tricyclic thalianol (**12**) in *Arabidopsis* [27,41]. Marneral synthase (MRN1) produces two oxidation products, one is marneral (aldehyde) and the other marnerol (alcohol). *Arabidopsis* CYP71A16 hydroxylates the allylic methyl side-chain of monocyclic marneral/marnerol to 23-hydroxymarneral/23-hydroxymarnerol. Modification of thalianol (**12**) involves CYPs from two clans. Genes encoding CYP708A2 (clan 85) and CYP705A5 (clan 71) are physically clustered with the thalianol synthase (THAS) gene, encoding the corresponding oxidosqualene cyclase. CYP708A2 oxidises the tricyclic thalianol (**12**) scaffold to 7β-hydroxythalianol, while CYP705A5 is a desaturase and introduces a double bond at C15 [41]. The related *Arabidopsis* CYP705A1 (also from clan 71) accepts a slightly different scaffold, arabidiol (**11**), triggering cleavage of the side chain at the same C15 instead of dehydrogenation. This shows that even closely related CYPs from the same subfamily can exhibit distinct differences in their substrate and reaction profiles.

Recent examples of triterpenoid and steroid cytochrome P450 monooxygenases

In this last section, we will illustrate selected examples that showcase the enzymatic versatility of CYPs in plant triterpenoid and steroid metabolism (Figure 6).

Diosgenin (**13**) is a specialised plant natural product with a unique 5,6-spiroketal moiety that serves as an inexpensive raw material for the industrial synthesis of steroidal drugs. Diosgenin (**13**) biosynthesis from cholesterol (**3**) was explored in *Paris polyphylla* (Pp; monocot), *Trigonella foenum-graecum* (Tf; dicot) and *Dioscorea zingiberensis* (Dz; monocot) (Figure 6A) [35,66]. Multifunctional CYPs PpCYP90G4/TfCYP90B50 were independently recruited from the ancient CYP90B subfamily involved in brassinosteroid biosynthesis to catalyse the initial C22,16 dihydroxylation of cholesterol (**3**) [35]; in contrast, the related CYP DzCYP90B71 was found to catalyse only the first hydroxylation at C22 [66]. This step is followed by a rate-limiting cyclisation step through unstable furostanol intermediate **14** that involves CYP-catalysed oxidative ring closure, leading to a hemiketal bridge between C16 and C22. Following these initial hydroxylations, CYPs from multiple families catalyse end-of-chain hydroxylation at C27 which is followed by spontaneous spiroketalisation to form diosgenin (**13**). The CYP pairs PpCYP90G4-PpCYP94D108 in *P. polyphylla* and TfCYP90B50-TfCYP82J17 in *T. foenum-graecum* resulted in the highest diosgenin (**13**) production. Diosgenin (**13**) biosynthesis in distantly related plants is an example of catalytic plasticity embedded within the ancient CYP90Bs. Especially CYPs from large families often show

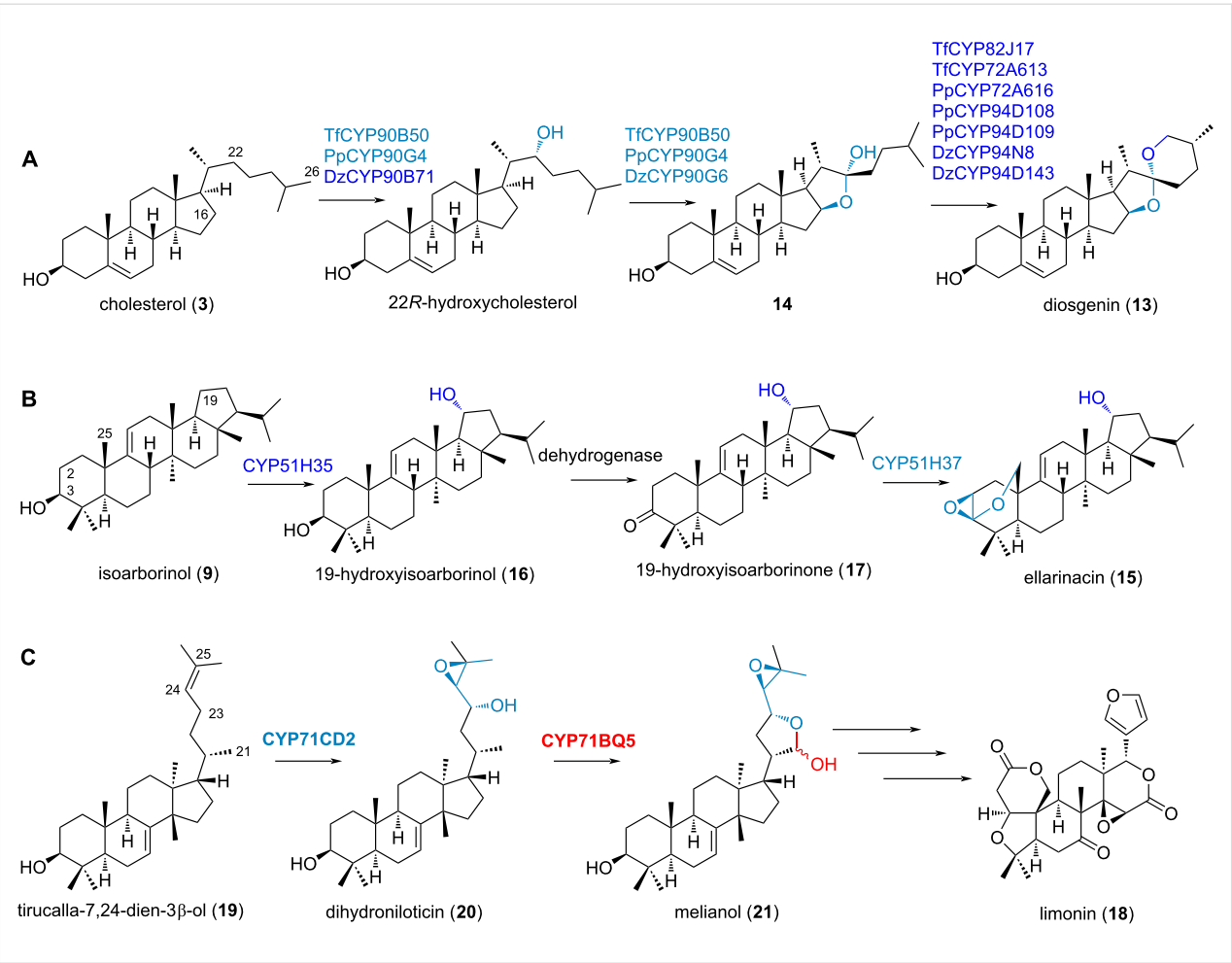


Figure 6: Recent examples of multifunctional CYPs in triterpenoid and steroid metabolism in plants that install complex oxidative modifications; in addition, their discovery also showcases modern approaches to elucidate plant specialised metabolism. A) CYPs from different plants producing diosgenin (**13**) (Tf: *Trigonella foenum-graecum*; Pp: *Paris polyphylla*; Dz: *Dioscorea zingiberensis*) [35,66]. B) Formation of the defence compound ellarinacin (**15**) in bread wheat [26]. Stereochemistry of ellarinacin (**15**) is shown as published. C) Biosynthesis of the key intermediate melianol (**21**) in the pathway to the limonoid limonin (**18**) [29]. The stereochemistry is shown as published.

high substrate promiscuity which facilitates duplication events resulting in neofunctionalisation [15].

Ellarinacin (**15**) is a defence-related arborinane-type triterpenoid that was recently discovered in bread wheat (*Triticum aestivum*) by genome mining (Figure 6B) [26]. The ellarinacin gene cluster encodes the three CYP enzymes TaCYP51H35, TaCYP51H37 and TaCYP51H13P, with the latter carrying a premature stop codon. TaCYP51H35 catalyses the C19-hydroxylation of isoarborinol (**9**) to form 19-hydroxyisoarborinol (**16**), which is oxidised to ketone **17** by a dehydrogenase (TaHID). TaCYP51H37 then carries out a remarkable double oxidation at the methyl group C25 as well as C2, leading to the highly unusual acetal-epoxide proposed for ellarinacin (**15**). This work therefore not only represents an important example how a CYP51H evolved by gene duplication and neofunctionalisation from a sterol biosynthetic gene, but also demonstrates the capacity of CYPs to catalyse unique enzymatic cascades.

Limonoids are highly oxidised, modified and truncated triterpenoids; one of the most well-known limonoids is the eponymous compound limonin (**18**), which contributes to the bitter taste of citrus products (Figure 6C) [104]. The first steps of limonoid biosynthesis were recently explored by functional characterisation in heterologous hosts [29,105-107]. There, CYP enzymes MaCYP71CD2 and MaCYP71BQ5 from *Melia azedarach* initiate the ring formation on the side chain of the triterpene precursor tirucalla-7,24-dien-3β-ol (**19**) in a sequential manner. MaCYP71CD2 is a bifunctional CYP that hydroxylates C23 and additionally introduces a C24–C25 epoxide on the side chain of tirucalla-7,24-dien-3β-ol (**19**), yielding dihydroniloticin (**20**). MaCYP71BQ5 then oxidises the methyl group C21 to a formyl group, leading to spontaneous hemi-

acetal ring formation in the product melianol (**21**). It is believed that these transformations are the starting point for formation of the characteristic furan ring of limonoids [29].

Taken together, these case studies not only represent impressive examples how CYPs create chemical complexity in plant triterpenoid and steroid metabolism, but also illustrate state-of-the-art approaches to discover and characterise new CYPs by genome mining, co-expression analyses, and efficient heterologous expression systems.

Conclusion

In this review, we provided a comprehensive overview over the phylogenetic distribution and diverse metabolic reactions catalysed by CYPs involved in the tailoring of triterpenoids and steroids from plants, covering 149 CYPs that have been functionally characterised to date (Table 1). Considering that up to 1% of all plant genes encode CYPs and that triterpenoids are one of the largest natural product classes in plants, we expect that this number will rise quickly in years to come. Several of our examples highlight the substrate promiscuity embedded within ancient CYP families, which enables rapid functional extension to acquire unique catalytic functions during duplication events [15,26,79]. The increasing availability of high-quality transcriptome and genome data even of non-model plants together with reliable and efficient expression systems in yeast and in *Nicotiana benthamiana* will facilitate future approaches to fully harness the diversity of triterpenoids and steroids found in plants. In combination with ground-breaking machine learning approaches for protein structure prediction such as AlphaFold2 [108], we anticipate that the catalytic repertoire of CYPs will be exploited much more for the biotechnological production of tailor-made triterpenoids and steroids in the near future. We hope that our review provides a good starting point for such further studies.

Supporting Information

Supporting Information File 1

High-quality version of the phylogenetic tree shown in Figure 2 with tip labels.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-135-S1.pdf]

Acknowledgements

We thank the entire Franke group for helpful discussions. The graphical abstract was created with BioRender.com.

Funding

We gratefully acknowledge financial support by the Emmy Noether programme of the Deutsche Forschungsgemeinschaft (FR 3720/3-1) and the SMART BIOTECS alliance between the Technische Universität Braunschweig and the Leibniz Universität Hannover, supported by the Ministry of Science and Culture (MWK) of Lower Saxony.

ORCID® iDs

Karan Malhotra - https://orcid.org/0000-0002-9748-7067
Jakob Franke - https://orcid.org/0000-0002-7603-6232

References

1. Qi, X.; Bakht, S.; Qin, B.; Leggett, M.; Hemmings, A.; Mellon, F.; Eagles, J.; Werck-Reichhart, D.; Schaller, H.; Lesot, A.; Melton, R.; Osbourn, A. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103,* 18848–18853. doi:10.1073/pnas.0607849103
2. Kitagawa, I. *Pure Appl. Chem.* **2002,** *74,* 1189–1198. doi:10.1351/pac200274071189
3. Zhang, M.; Sun, X.; Ren, R.; Su, L.; Xu, M.; Zheng, L.; Li, H. *Phytochem. Lett.* **2022,** *49,* 145–151. doi:10.1016/j.phytol.2022.03.020
4. Xiong, J.; Kashiwada, Y.; Chen, C.-H.; Qian, K.; Morris-Natschke, S. L.; Lee, K.-H.; Takaishi, Y. *Bioorg. Med. Chem.* **2010,** *18,* 6451–6469. doi:10.1016/j.bmc.2010.06.092
5. Thimmappa, R.; Geisler, K.; Louveau, T.; O'Maille, P.; Osbourn, A. *Annu. Rev. Plant Biol.* **2014,** *65,* 225–257. doi:10.1146/annurev-arplant-050312-120229
6. Nelson, D.; Werck-Reichhart, D. *Plant J.* **2011,** *66,* 194–211. doi:10.1111/j.1365-313x.2011.04529.x
7. Ghosh, S. *Front. Plant Sci.* **2017,** *8,* 1886. doi:10.3389/fpls.2017.01886
8. Nguyen, T.-D.; Dang, T.-T. T. *Front. Plant Sci.* **2021,** *12,* 682181. doi:10.3389/fpls.2021.682181
9. Bathe, U.; Tissier, A. *Phytochemistry* **2019,** *161,* 149–162. doi:10.1016/j.phytochem.2018.12.003
10. Zhang, Y.; Ma, L.; Su, P.; Huang, L.; Gao, W. *Crit. Rev. Biotechnol.* **2021,** 1–21. doi:10.1080/07388551.2021.2003292
11. Mizutani, M.; Sato, F. *Arch. Biochem. Biophys.* **2011,** *507,* 194–203. doi:10.1016/j.abb.2010.09.026
12. Miettinen, K.; Iñigo, S.; Kreft, L.; Pollier, J.; De Bo, C.; Botzki, A.; Coppens, F.; Bak, S.; Goossens, A. *Nucleic Acids Res.* **2018,** *46,* D586–D594. doi:10.1093/nar/gkx925
13. Shang, Y.; Huang, S. *Plant J.* **2019,** *97,* 101–111. doi:10.1111/tpj.14132
14. Bak, S.; Beisson, F.; Bishop, G.; Hamberger, B.; Höfer, R.; Paquette, S.; Werck-Reichhart, D. *Arabidopsis Book* **2011,** *9,* e0144. doi:10.1199/tab.0144
15. Hansen, C. C.; Nelson, D. R.; Møller, B. L.; Werck-Reichhart, D. *Mol. Plant* **2021,** *14,* 1244–1265. doi:10.1016/j.molp.2021.06.028
16. Nelson, D. R. Cytochrome P450 Nomenclature, 2004. In *Cytochrome P450 Protocols;* Phillips, I. R.; Shephard, E. A., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2006; pp 1–10. doi:10.1385/1-59259-998-2:1
17. Meunier, B.; de Visser, S. P.; Shaik, S. *Chem. Rev.* **2004,** *104,* 3947–3980. doi:10.1021/cr020443g

18. Poulos, T. L. *Chem. Rev.* **2014,** *114,* 3919–3962. doi:10.1021/cr400415k
19. Guengerich, F. P. *ACS Catal.* **2018,** *8,* 10964–10976. doi:10.1021/acscatal.8b03401
20. Rittle, J.; Green, M. T. *Science* **2010,** *330,* 933–937. doi:10.1126/science.1193478
21. Dubey, K. D.; Shaik, S. *Acc. Chem. Res.* **2019,** *52,* 389–399. doi:10.1021/acs.accounts.8b00467
22. Chen, C.-C.; Min, J.; Zhang, L.; Yang, Y.; Yu, X.; Guo, R.-T. *ChemBioChem* **2021,** *22,* 1317–1328. doi:10.1002/cbic.202000705
23. Ortiz de Montellano, P. R. Substrate Oxidation by Cytochrome P450 Enzyme. In *Cytochrome P450: Structure, Mechanism, and Biochemistry;* Ortiz de Montellano, P. R., Ed.; Springer International Publishing: Cham, Switzerland, 2015; pp 111–176. doi:10.1007/978-3-319-12108-6_4
24. Bak, S.; Kahn, R. A.; Olsen, C. E.; Halkier, B. A. *Plant J.* **1997,** *11,* 191–201. doi:10.1046/j.1365-313x.1997.11020191.x
25. Kushiro, M.; Nakano, T.; Sato, K.; Yamagishi, K.; Asami, T.; Nakano, A.; Takatsuto, S.; Fujioka, S.; Ebizuka, Y.; Yoshida, S. *Biochem. Biophys. Res. Commun.* **2001,** *285,* 98–104. doi:10.1006/bbrc.2001.5122
26. Polturak, G.; Dippe, M.; Stephenson, M. J.; Chandra Misra, R.; Owen, C.; Ramirez-Gonzalez, R. H.; Haidoulis, J. F.; Schoonbeek, H.-J.; Chartrain, L.; Borrill, P.; Nelson, D. R.; Brown, J. K. M.; Nicholson, P.; Uauy, C.; Osbourn, A. *Proc. Natl. Acad. Sci. U. S. A.* **2022,** *119,* e2123299119. doi:10.1073/pnas.2123299119
27. Field, B.; Fiston-Lavier, A.-S.; Kemen, A.; Geisler, K.; Quesneville, H.; Osbourn, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108,* 16116–16121. doi:10.1073/pnas.1109273108
28. Castillo, D. A.; Kolesnikova, M. D.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2013,** *135,* 5885–5894. doi:10.1021/ja401535g
29. Hodgson, H.; De La Peña, R.; Stephenson, M. J.; Thimmappa, R.; Vincent, J. L.; Sattely, E. S.; Osbourn, A. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116,* 17096–17104. doi:10.1073/pnas.1906083116
30. Krokida, A.; Delis, C.; Geisler, K.; Garagounis, C.; Tsikou, D.; Peña-Rodríguez, L. M.; Katsarou, D.; Field, B.; Osbourn, A. E.; Papadopoulou, K. K. *New Phytol.* **2013,** *200,* 675–690. doi:10.1111/nph.12414
31. Tsukagoshi, Y.; Ohyama, K.; Seki, H.; Akashi, T.; Muranaka, T.; Suzuki, H.; Fujimoto, Y. *Phytochemistry* **2016,** *127,* 23–28. doi:10.1016/j.phytochem.2016.03.010
32. Takase, S.; Kera, K.; Nagashima, Y.; Mannen, K.; Hosouchi, T.; Shinpo, S.; Kawashima, M.; Kotake, Y.; Yamada, H.; Saga, Y.; Otaka, J.; Araya, H.; Kotera, M.; Suzuki, H.; Kushiro, T. *J. Biol. Chem.* **2019,** *294,* 18662–18673. doi:10.1074/jbc.ra119.009944
33. Shang, Y.; Ma, Y.; Zhou, Y.; Zhang, H.; Duan, L.; Chen, H.; Zeng, J.; Zhou, Q.; Wang, S.; Gu, W.; Liu, M.; Ren, J.; Gu, X.; Zhang, S.; Wang, Y.; Yasukawa, K.; Bouwmeester, H. J.; Qi, X.; Zhang, Z.; Lucas, W. J.; Huang, S. *Science* **2014,** *346,* 1084–1088. doi:10.1126/science.1259215
34. Zhou, Y.; Ma, Y.; Zeng, J.; Duan, L.; Xue, X.; Wang, H.; Lin, T.; Liu, Z.; Zeng, K.; Zhong, Y.; Zhang, S.; Hu, Q.; Liu, M.; Zhang, H.; Reed, J.; Moses, T.; Liu, X.; Huang, P.; Qing, Z.; Liu, X.; Tu, P.; Kuang, H.; Zhang, Z.; Osbourn, A.; Ro, D.-K.; Shang, Y.; Huang, S. *Nat. Plants (London, U. K.)* **2016,** *2,* 16183. doi:10.1038/nplants.2016.183
35. Christ, B.; Xu, C.; Xu, M.; Li, F.-S.; Wada, N.; Mitchell, A. J.; Han, X.-L.; Wen, M.-L.; Fujita, M.; Weng, J.-K. *Nat. Commun.* **2019,** *10,* 3206. doi:10.1038/s41467-019-11286-7
36. Li, L.; Lin, S.; Chen, Y.; Wang, Y.; Xiao, L.; Ye, X.; Sun, L.; Zhan, R.; Xu, H. *Front. Plant Sci.* **2022,** *13,* 831401. doi:10.3389/fpls.2022.831401
37. Shibuya, M.; Hoshino, M.; Katsube, Y.; Hayashi, H.; Kushiro, T.; Ebizuka, Y. *FEBS J.* **2006,** *273,* 948–959. doi:10.1111/j.1742-4658.2006.05120.x
38. Moses, T.; Pollier, J.; Almagro, L.; Buyst, D.; Van Montagu, M.; Pedreño, M. A.; Martins, J. C.; Thevelein, J. M.; Goossens, A. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111,* 1634–1639. doi:10.1073/pnas.1323369111
39. Seki, H.; Ohyama, K.; Sawai, S.; Mizutani, M.; Ohnishi, T.; Sudo, H.; Akashi, T.; Aoki, T.; Saito, K.; Muranaka, T. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 14204–14209. doi:10.1073/pnas.0803876105
40. Moses, T.; Thevelein, J. M.; Goossens, A.; Pollier, J. *Phytochemistry* **2014,** *108,* 47–56. doi:10.1016/j.phytochem.2014.10.002
41. Field, B.; Osbourn, A. E. *Science* **2008,** *320,* 543–547. doi:10.1126/science.1154990
42. Hansen, N. L.; Miettinen, K.; Zhao, Y.; Ignea, C.; Andreadelli, A.; Raadam, M. H.; Makris, A. M.; Møller, B. L.; Stærk, D.; Bak, S.; Kampranis, S. C. *Microb. Cell Fact.* **2020,** *19,* 15. doi:10.1186/s12934-020-1284-9
43. Bicalho, K. U.; Santoni, M. M.; Arendt, P.; Zanelli, C. F.; Furlan, M.; Goossens, A.; Pollier, J. *Plant Cell Physiol.* **2019,** *60,* 2510–2522. doi:10.1093/pcp/pcz144
44. Fukushima, E. O.; Seki, H.; Sawai, S.; Suzuki, M.; Ohyama, K.; Saito, K.; Muranaka, T. *Plant Cell Physiol.* **2013,** *54,* 740–749. doi:10.1093/pcp/pct015
45. Fanani, M. Z.; Fukushima, E. O.; Sawai, S.; Tang, J.; Ishimori, M.; Sudo, H.; Ohyama, K.; Seki, H.; Saito, K.; Muranaka, T. *Front. Plant Sci.* **2019,** *10,* 1520. doi:10.3389/fpls.2019.01520
46. Seki, H.; Sawai, S.; Ohyama, K.; Mizutani, M.; Ohnishi, T.; Sudo, H.; Fukushima, E. O.; Akashi, T.; Aoki, T.; Saito, K.; Muranaka, T. *Plant Cell* **2011,** *23,* 4112–4123. doi:10.1105/tpc.110.082685
47. Biazzi, E.; Carelli, M.; Tava, A.; Abbruscato, P.; Losini, I.; Avato, P.; Scotti, C.; Calderini, O. *Mol. Plant* **2015,** *8,* 1493–1506. doi:10.1016/j.molp.2015.06.003
48. Tzin, V.; Snyder, J. H.; Yang, D. S.; Huhman, D. V.; Watson, B. S.; Allen, S. N.; Tang, Y.; Miettinen, K.; Arendt, P.; Pollier, J.; Goossens, A.; Sumner, L. W. *Metabolomics* **2019,** *15,* 85. doi:10.1007/s11306-019-1542-1
49. Yano, R.; Takagi, K.; Takada, Y.; Mukaiyama, K.; Tsukamoto, C.; Sayama, T.; Kaga, A.; Anai, T.; Sawai, S.; Ohyama, K.; Saito, K.; Ishimoto, M. *Plant J.* **2017,** *89,* 527–539. doi:10.1111/tpj.13403
50. Han, J. Y.; Chun, J.-H.; Oh, S. A.; Park, S.-B.; Hwang, H.-S.; Lee, H.; Choi, Y. E. *Plant Cell Physiol.* **2018,** *59,* 319–330. doi:10.1093/pcp/pcx188
51. Liu, Q.; Khakimov, B.; Cárdenas, P. D.; Cozzi, F.; Olsen, C. E.; Jensen, K. R.; Hauser, T. P.; Bak, S. *New Phytol.* **2019,** *222,* 1599–1609. doi:10.1111/nph.15689
52. Kim, O. T.; Um, Y.; Jin, M. L.; Kim, J. U.; Hegebarth, D.; Busta, L.; Racovita, R. C.; Jetter, R. *Plant Cell Physiol.* **2018,** *59,* 1200–1213. doi:10.1093/pcp/pcy055
53. Ohnishi, T.; Nomura, T.; Watanabe, B.; Ohta, D.; Yokota, T.; Miyagawa, H.; Sakata, K.; Mizutani, M. *Phytochemistry* **2006,** *67,* 1895–1906. doi:10.1016/j.phytochem.2006.05.042

54. Dai, Z.; Liu, Y.; Sun, Z.; Wang, D.; Qu, G.; Ma, X.; Fan, F.; Zhang, L.; Li, S.; Zhang, X. *Metab. Eng.* **2019,** *51,* 70–78. doi:10.1016/j.ymben.2018.10.001
55. Shimada, Y.; Fujioka, S.; Miyauchi, N.; Kushiro, M.; Takatsuto, S.; Nomura, T.; Yokota, T.; Kamiya, Y.; Bishop, G. J.; Yoshida, S. *Plant Physiol.* **2001,** *126,* 770–779. doi:10.1104/pp.126.2.770
56. Bishop, G. J.; Nomura, T.; Yokota, T.; Harrison, K.; Noguchi, T.; Fujioka, S.; Takatsuto, S.; Jones, J. D. G.; Kamiya, Y. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96,* 1761–1766. doi:10.1073/pnas.96.4.1761
57. Shimada, Y.; Goda, H.; Nakamura, A.; Takatsuto, S.; Fujioka, S.; Yoshida, S. *Plant Physiol.* **2003,** *131,* 287–297. doi:10.1104/pp.013029
58. Nomura, T.; Kushiro, T.; Yokota, T.; Kamiya, Y.; Bishop, G. J.; Yamaguchi, S. *J. Biol. Chem.* **2005,** *280,* 17873–17879. doi:10.1074/jbc.m414592200
59. Moses, T.; Pollier, J.; Faizal, A.; Apers, S.; Pieters, L.; Thevelein, J. M.; Geelen, D.; Goossens, A. *Mol. Plant* **2015,** *8,* 122–135. doi:10.1016/j.molp.2014.11.004
60. Ohnishi, T.; Godza, B.; Watanabe, B.; Fujioka, S.; Hategan, L.; Ide, K.; Shibata, K.; Yokota, T.; Szekeres, M.; Mizutani, M. *J. Biol. Chem.* **2012,** *287,* 31551–31560. doi:10.1074/jbc.m112.392720
61. Fujita, S.; Ohnishi, T.; Watanabe, B.; Yokota, T.; Takatsuto, S.; Fujioka, S.; Yoshida, S.; Sakata, K.; Mizutani, M. *Plant J.* **2006,** *45,* 765–774. doi:10.1111/j.1365-313x.2005.02639.x
62. Sakamoto, T.; Morinaka, Y.; Ohnishi, T.; Sunohara, H.; Fujioka, S.; Ueguchi-Tanaka, M.; Mizutani, M.; Sakata, K.; Takatsuto, S.; Yoshida, S.; Tanaka, H.; Kitano, H.; Matsuoka, M. *Nat. Biotechnol.* **2006,** *24,* 105–109. doi:10.1038/nbt1173
63. Ohnishi, T.; Watanabe, B.; Sakata, K.; Mizutani, M. *Biosci., Biotechnol., Biochem.* **2006,** *70,* 2071–2080. doi:10.1271/bbb.60034
64. Augustin, M. M.; Ruzicka, D. R.; Shukla, A. K.; Augustin, J. M.; Starks, C. M.; O'Neil-Johnson, M.; McKain, M. R.; Evans, B. S.; Barrett, M. D.; Smithson, A.; Wong, G. K.-S.; Deyholos, M. K.; Edger, P. P.; Pires, J. C.; Leebens-Mack, J. H.; Mann, D. A.; Kutchan, T. M. *Plant J.* **2015,** *82,* 991–1003. doi:10.1111/tpj.12871
65. Yin, Y.; Gao, L.; Zhang, X.; Gao, W. *Phytochemistry* **2018,** *156,* 116–123. doi:10.1016/j.phytochem.2018.09.005
66. Zhou, C.; Yang, Y.; Tian, J.; Wu, Y.; An, F.; Li, C.; Zhang, Y. *Plant J.* **2022,** *109,* 940–951. doi:10.1111/tpj.15604
67. Ohnishi, T.; Szatmari, A.-M.; Watanabe, B.; Fujita, S.; Bancos, S.; Koncz, C.; Lafos, M.; Shibata, K.; Yokota, T.; Sakata, K.; Szekeres, M.; Mizutani, M. *Plant Cell* **2006,** *18,* 3275–3288. doi:10.1105/tpc.106.045443
68. Sakamoto, T.; Ohnishi, T.; Fujioka, S.; Watanabe, B.; Mizutani, M. *Plant Physiol. Biochem.* **2012,** *58,* 220–226. doi:10.1016/j.plaphy.2012.07.011
69. Dong, L.; Almeida, A.; Pollier, J.; Khakimov, B.; Bassard, J.-E.; Miettinen, K.; Stærk, D.; Mehran, R.; Olsen, C. E.; Motawia, M. S.; Goossens, A.; Bak, S. *Mol. Biol. Evol.* **2021,** *38,* 4659–4673. doi:10.1093/molbev/msab213
70. Yasumoto, S.; Fukushima, E. O.; Seki, H.; Muranaka, T. *FEBS Lett.* **2016,** *590,* 533–540. doi:10.1002/1873-3468.12074
71. Carelli, M.; Biazzi, E.; Panara, F.; Tava, A.; Scaramelli, L.; Porceddu, A.; Graham, N.; Odoardi, M.; Piano, E.; Arcioni, S.; May, S.; Scotti, C.; Calderini, O. *Plant Cell* **2011,** *23,* 3070–3081. doi:10.1105/tpc.111.087312
72. Moses, T.; Pollier, J.; Shen, Q.; Soetaert, S.; Reed, J.; Erffelinck, M.-L.; Van Nieuwerburgh, F. C. W.; Vanden Bossche, R.; Osbourn, A.; Thevelein, J. M.; Deforce, D.; Tang, K.; Goossens, A. *Plant Cell* **2015,** *27,* 286–301. doi:10.1105/tpc.114.134486
73. Fukushima, E. O.; Seki, H.; Ohyama, K.; Ono, E.; Umemoto, N.; Mizutani, M.; Saito, K.; Muranaka, T. *Plant Cell Physiol.* **2011,** *52,* 2050–2061. doi:10.1093/pcp/pcr146
74. Yasumoto, S.; Seki, H.; Shimizu, Y.; Fukushima, E. O.; Muranaka, T. *Front. Plant Sci.* **2017,** *8,* 21. doi:10.3389/fpls.2017.00021
75. Suzuki, H.; Fukushima, E. O.; Shimizu, Y.; Seki, H.; Fujisawa, Y.; Ishimoto, M.; Osakabe, K.; Osakabe, Y.; Muranaka, T. *Plant Cell Physiol.* **2019,** *60,* 2496–2509. doi:10.1093/pcp/pcz145
76. Han, J.-Y.; Hwang, H.-S.; Choi, S.-W.; Kim, H.-J.; Choi, Y.-E. *Plant Cell Physiol.* **2012,** *53,* 1535–1545. doi:10.1093/pcp/pcs106
77. Fiallos-Jurado, J.; Pollier, J.; Moses, T.; Arendt, P.; Barriga-Medina, N.; Morillo, E.; Arahana, V.; de Lourdes Torres, M.; Goossens, A.; Leon-Reyes, A. *Plant Sci.* **2016,** *250,* 188–197. doi:10.1016/j.plantsci.2016.05.015
78. Khakimov, B.; Kuzina, V.; Erthmann, P. Ø.; Fukushima, E. O.; Augustin, J. M.; Olsen, C. E.; Scholtalbers, J.; Volpin, H.; Andersen, S. B.; Hauser, T. P.; Muranaka, T.; Bak, S. *Plant J.* **2015,** *84,* 478–490. doi:10.1111/tpj.13012
79. Miettinen, K.; Pollier, J.; Buyst, D.; Arendt, P.; Csuk, R.; Sommerwerk, S.; Moses, T.; Mertens, J.; Sonawane, P. D.; Pauwels, L.; Aharoni, A.; Martins, J.; Nelson, D. R.; Goossens, A. *Nat. Commun.* **2017,** *8,* 14153. doi:10.1038/ncomms14153
80. Tamura, K.; Teranishi, Y.; Ueda, S.; Suzuki, H.; Kawano, N.; Yoshimatsu, K.; Saito, K.; Kawahara, N.; Muranaka, T.; Seki, H. *Plant Cell Physiol.* **2017,** *58,* 874–884. doi:10.1093/pcp/pcx043
81. Huang, L.; Li, J.; Ye, H.; Li, C.; Wang, H.; Liu, B.; Zhang, Y. *Planta* **2012,** *236,* 1571–1581. doi:10.1007/s00425-012-1712-0
82. Huang, J.; Zha, W.; An, T.; Dong, H.; Huang, Y.; Wang, D.; Yu, R.; Duan, L.; Zhang, X.; Peters, R. J.; Dai, Z.; Zi, J. *Appl. Microbiol. Biotechnol.* **2019,** *103,* 7029–7039. doi:10.1007/s00253-019-10004-z
83. Andre, C. M.; Legay, S.; Deleruelle, A.; Nieuwenhuizen, N.; Punter, M.; Brendolise, C.; Cooney, J. M.; Lateur, M.; Hausman, J.-F.; Larondelle, Y.; Laing, W. A. *New Phytol.* **2016,** *211,* 1279–1294. doi:10.1111/nph.13996
84. Tamura, K.; Seki, H.; Suzuki, H.; Kojoma, M.; Saito, K.; Muranaka, T. *Plant Cell Rep.* **2017,** *36,* 437–445. doi:10.1007/s00299-016-2092-x
85. Zhou, C.; Li, J.; Li, C.; Zhang, Y. *BMC Biotechnol.* **2016,** *16,* 59. doi:10.1186/s12896-016-0290-9
86. Ji, X.; Lin, S.; Chen, Y.; Liu, J.; Yun, X.; Wang, T.; Qin, J.; Luo, C.; Wang, K.; Zhao, Z.; Zhan, R.; Xu, H. *Front. Plant Sci.* **2020,** *11,* 612. doi:10.3389/fpls.2020.00612
87. Jo, H.-J.; Han, J. Y.; Hwang, H.-S.; Choi, Y. E. *Phytochemistry* **2017,** *135,* 53–63. doi:10.1016/j.phytochem.2016.12.011
88. Misra, R. C.; Sharma, S.; Sandeep; Garg, A.; Chanotiya, C. S.; Ghosh, S. *New Phytol.* **2017,** *214,* 706–720. doi:10.1111/nph.14412
89. Sandeep; Misra, R. C.; Chanotiya, C. S.; Mukhopadhyay, P.; Ghosh, S. *New Phytol.* **2019,** *222,* 408–424. doi:10.1111/nph.15606
90. Zhang, R.; Xia, X.; Lindsey, K.; da Rocha, P. S. C. F. *J. Plant Physiol.* **2012,** *169,* 421–428. doi:10.1016/j.jplph.2011.10.013
91. Morikawa, T.; Mizutani, M.; Aoki, N.; Watanabe, B.; Saga, H.; Saito, S.; Oikawa, A.; Suzuki, H.; Sakurai, N.; Shibata, D.; Wadano, A.; Sakata, K.; Ohta, D. *Plant Cell* **2006,** *18,* 1008–1022. doi:10.1105/tpc.105.036012
92. Edgar, R. C. *Nucleic Acids Res.* **2004,** *32,* 1792–1797. doi:10.1093/nar/gkh340

93. Jacobowitz, J. R.; Weng, J.-K. *Annu. Rev. Plant Biol.* **2020,** *71,* 631–658. doi:10.1146/annurev-arplant-081519-035634
94. Chuang, L.; Franke, J. Rapid Combinatorial Coexpression of Biosynthetic Genes by Transient Expression in the Plant Host Nicotiana Benthamiana. In *Engineering Natural Product Biosynthesis: Methods and Protocols;* Skellam, E., Ed.; Methods in Molecular Biology; Springer US: New York, NY, USA, 2022; pp 395–420. doi:10.1007/978-1-0716-2273-5_20
95. Dang, T.-T. T.; Franke, J.; Carqueijeiro, I. S. T.; Langley, C.; Courdavault, V.; O'Connor, S. E. *Nat. Chem. Biol.* **2018,** *14,* 760–763. doi:10.1038/s41589-018-0078-4
96. Nett, R. S.; Lau, W.; Sattely, E. S. *Nature* **2020,** *584,* 148–153. doi:10.1038/s41586-020-2546-8
97. Arnqvist, L.; Persson, M.; Jonsson, L.; Dutta, P. C.; Sitbon, F. *Planta* **2008,** *227,* 309–317. doi:10.1007/s00425-007-0618-8
98. Cabello-Hurtado, F.; Taton, M.; Forthoffer, N.; Kahn, R.; Bak, S.; Rahier, A.; Werck-Reichhart, D. *Eur. J. Biochem.* **1999,** *262,* 435–446. doi:10.1046/j.1432-1327.1999.00376.x
99. Geisler, K.; Hughes, R. K.; Sainsbury, F.; Lomonossoff, G. P.; Rejzek, M.; Fairhurst, S.; Olsen, C.-E.; Motawia, M. S.; Melton, R. E.; Hemmings, A. M.; Bak, S.; Osbourn, A. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110,* E3360–E3367. doi:10.1073/pnas.1309157110
100. Kasai, R.; Nie, R.-L.; Nashi, K.; Ohtani, K.; Zhou, J.; Tao, G.-D.; Tanaka, O. *Agric. Biol. Chem.* **1989,** *53,* 3347–3349. doi:10.1080/00021369.1989.10869815
101. Zhang, J.; Dai, L.; Yang, J.; Liu, C.; Men, Y.; Zeng, Y.; Cai, Y.; Zhu, Y.; Sun, Y. *Plant Cell Physiol.* **2016,** *57,* 1000–1007. doi:10.1093/pcp/pcw038
102. Itkin, M.; Davidovich-Rikanati, R.; Cohen, S.; Portnoy, V.; Doron-Faigenboim, A.; Oren, E.; Freilich, S.; Tzuri, G.; Baranes, N.; Shen, S.; Petreikov, M.; Sertchook, R.; Ben-Dor, S.; Gottlieb, H.; Hernandez, A.; Nelson, D. R.; Paris, H. S.; Tadmor, Y.; Burger, Y.; Lewinsohn, E.; Katzir, N.; Schaffer, A. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* E7619–E7628. doi:10.1073/pnas.1604828113
103. Han, J.-Y.; Kim, H.-J.; Kwon, Y.-S.; Choi, Y.-E. *Plant Cell Physiol.* **2011,** *52,* 2062–2073. doi:10.1093/pcp/pcr150
104. Li, L.-J.; Tan, W.-S.; Li, W.-J.; Zhu, Y.-B.; Cheng, Y.-S.; Ni, H. *Int. J. Mol. Sci.* **2019,** *20,* 6194. doi:10.3390/ijms20246194
105. Pandreka, A.; Chaya, P. S.; Kumar, A.; Aarthy, T.; Mulani, F. A.; Bhagyashree, D. D.; B, S. H.; Jennifer, C.; Ponnusamy, S.; Nagegowda, D.; Thulasiram, H. V. *Phytochemistry* **2021,** *184,* 112669. doi:10.1016/j.phytochem.2021.112669
106. Lian, X.; Zhang, X.; Wang, F.; Wang, X.; Xue, Z.; Qi, X. *Physiol. Plant.* **2020,** *170,* 528–536. doi:10.1111/ppl.13189
107. Chuang, L.; Liu, S.; Biedermann, D.; Franke, J. *Front. Plant Sci.* **2022,** *13,* 958138. doi:10.3389/fpls.2022.958138
108. 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,** 1–11. doi:10.1038/s41586-021-03819-2

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Characterization of a new fusicoccane-type diterpene synthase and an associated P450 enzyme

Jia-Hua Huang‡[1], Jian-Ming Lv‡[1], Liang-Yan Xiao[1], Qian Xu[1], Fu-Long Lin[1], Gao-Qian Wang[1], Guo-Dong Chen[1], Sheng-Ying Qin*[2], Dan Hu*[1,3] and Hao Gao[1]

Full Research Paper

Open Access

Address:
[1]Institute of Traditional Chinese Medicine & Natural Products, College of Pharmacy / Guangdong Province Key Laboratory of Pharmacodynamic Constituents of TCM and New Drugs Research / International Cooperative Laboratory of Traditional Chinese Medicine Modernization and Innovative Drug Development of Ministry of Education (MOE) of China, Jinan University, Guangzhou 510632, China, [2]Clinical Experimental Center, First Affiliated Hospital of Jinan University, Guangzhou 510630, China, and [3]Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

Email:
Sheng-Ying Qin* - qinshengying78@163.com; Dan Hu* - thudan@jnu.edu.cn

* Corresponding author ‡ Equal contributors

Keywords:
cytochrome P450 enzyme; diterpene synthase; gene cluster; genome mining; site-directed mutagenesis

Beilstein J. Org. Chem. **2022,** *18,* 1396–1402.
https://doi.org/10.3762/bjoc.18.144

Received: 08 June 2022
Accepted: 19 September 2022
Published: 05 October 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: K. N. Allen

Abstract

Fusicoccane-type terpenoids are a subgroup of diterpenoids featured with a unique 5-8-5 ring system. They are widely distributed in nature and possess a variety of biological activities. Up to date, only five fusicoccane-type diterpene synthases have been identified. Here, we identify a two-gene biosynthetic gene cluster containing a new fusicoccane-type diterpene synthase gene *tadA* and an associated cytochrome P450 gene *tadB* from *Talaromyces wortmannii* ATCC 26942. Heterologous expression reveals that TadA catalyzes the formation of a new fusicoccane-type diterpene talaro-7,13-diene. D_2O isotope labeling combined with site-directed mutagenesis indicates that TadA might employ a different C2,6 cyclization strategy from the known fusicoccane-type diterpene synthases, in which a neutral intermediate is firstly formed and then protonated by an environmental proton. In addition, we demonstrate that the associated cytochrome P450 enzyme TadB is able to catalyze multiple oxidation of talaro-7,13-diene to yield talaro-6,13-dien-5,8-dione.

Introduction

Terpenoids are a large class of natural products that attract extensive attention, due to not only their potential applications in pharmaceuticals, agrochemicals, etc. but also due to their abundant structural architectures [1]. Fusicoccane (FC)-type terpenoids are a subgroup of diterpenoids possessing a unique 5-8-5 tricyclic skeleton, which can be produced by plants, fungi

and bacteria [2]. This type of diterpenoids, represented by fusicoccin A and cotylenin A, can serve as efficient modulators of 14-3-3 protein–protein interactions (PPIs) [3,4]. 14-3-3 PPIs, which refer to the binding interactions of 14-3-3 proteins with hundreds of "client" proteins, are associated with many diseases, such as cancer and neurodegenerative diseases [3,4]. As a result, lots of attempts, including traditional isolation from nature [5-8] and chemical synthesis [9], have been made to expand the structural diversity of FC-type diterpenoids for drug development.

Along with the development of low-cost sequencing technologies and tractable heterologous expression systems, genome mining has become a promising strategy for targeted discovery of natural products [10-12], which can also provide enzymatic tools toward combinatorial biosynthesis [13,14]. As terpene synthases play a fundamental role in constructing molecular skeletons, great efforts have been devoted to mining novel synthases in pursuit of new terpenoids [15-17]. It is generally accepted that the FC-type diterpene skeleton is formed from geranylgeranyl diphosphate (GGPP) via a concerted C1,11–C10,14-bicyclization, followed by a C2,6-cyclization [18-21]. Theoretically, according to the configuration of stereogenic centers at C2, C6, C10, and C14 introduced during the two cyclization steps, FC-type diterpene synthases (DTSs) could be divided into 16 subtypes [20]. Furthermore, considering the modes of potential carbocation rearrangement and final carbocation quenching, there might be more FC-type DTSs in nature. To date, only five distinct FC-type DTSs, including fungi-derived PaFS/SdnA/MgMS and bacteria-derived CotB2/CpCS, have been reported (Figure 1) [20-24], implying that there still exists a large enzymatic space remaining to be explored.

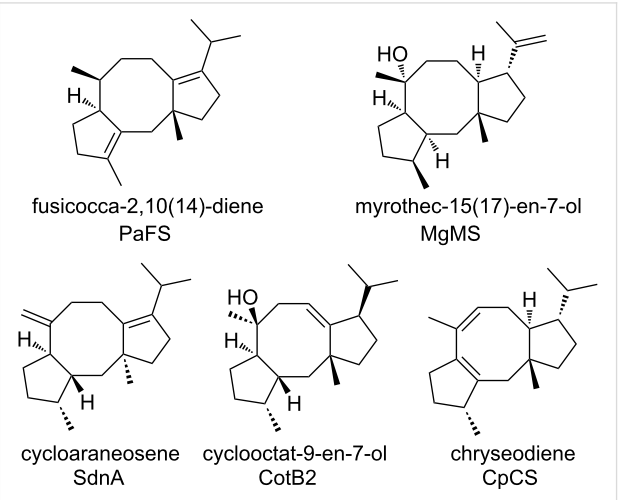


Figure 1: The five distinct FC-type DTSs and the corresponding products.

Herein, we characterize a two-gene cluster from *Talaromyces wortmannii* ATCC 26942, in which TadA is identified to be a new FC-type DTS responsible for the formation of talaro-7,13-diene, and the associated P450 enzyme TadB is characterized to be a multifunctional enzyme, converting talaro-7,13-diene to highly oxygenated talaro-6,13-dien-5,8-dione.

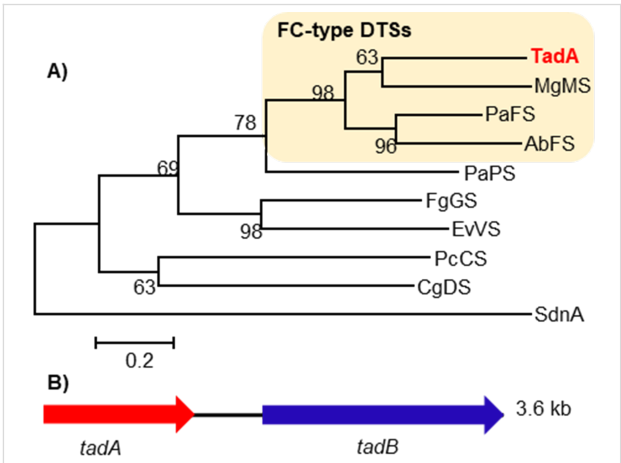
Results and Discussion

Identification of a new fusicoccane-type DTS and an associated P450 enzyme

We used MgMS, a previously identified FC-type DTS in our group [20], as a query to perform local BLAST search against our in-house fungal genome database, and then found a candidate enzyme TadA from *T. wortmannii* ATCC 26942. Further phylogenetic analysis of TadA and representative fungal DTSs showed that TadA falls within the clade of FC-type DTSs (Figure 2A). In light of showing low amino acid sequence identity (<50%) to reported fungal FC-type DTSs [20,22,24,25], TadA was annotated as a putative new FC-type DTS. We then scanned the flanking region of *tadA*, and found a cytochrome P450 gene *tadB*. Accordingly, the two-gene cluster was termed *tad* cluster (Figure 2B), and the sequence data was deposited in GenBank under the accession number ON624151.

Functional analysis of the FC-type DTS TadA

In order to analyze the function of TadA, we introduced *tadA* into the quadruple auxotrophic host *Aspergillus oryzae* NSAR1



Gene	Amino acids	Protein homolog	Similarity/ identity (%)	Putative function
tadA	319	MgMS	59/45	Terpene synthase
tadB	508	ESZ98217.1	73/56	P450

Figure 2: Bioinformatics analysis of the *tad* cluster. A) Phylogenetic tree of TadA and representative fungal DTSs. B) Putative functions of the *tad* cluster.

(*niaD*$^-$, *sC*$^-$, Δ*argB*, *adeA*$^-$) (Supporting Information 1, Tables S1 and S2) [26], which has been widely used for biosynthesis of fungi-derived natural products due to its genetic tractability [27]. The resulted transformant was cultivated in the modified Czapek–Dox medium for four days, and then mycelia were harvested and extracted for analysis. Upon gas chromatography-mass spectrometry (GC–MS) analysis, we observed an additional peak at m/z = 272 $[M]^+$ from the *tadA* harboring transformant, implying that TadA can catalyze the formation of a diterpene hydrocarbon (Supporting Information 1, Figure S1). The crude extract was also subjected to high-performance liquid chromatography–mass spectrometry (HPLC–MS) analysis. It showed that introduction of *tadA* led to the generation of two products **1** and **2**, but at low yields (Figure 3A, lines i and ii). To facilitate the isolation of these compounds, we co-expressed *tadA* with the GGPP synthase (GGPPS) gene (Supporting Information File 1, Note S1) derived from *Nodulisporium* sp. (No. 65-12-7-1) [28] in *A. oryzae* NSAR1. Although GGPP was not detected, production of **1** was increased by five-fold (Figure 3A, line iii). Finally, we isolated **1** and **2** through large-scale fermentation.

On the basis of the quasi-molecular ion at m/z 273.2599 $[M + H]^+$ (calcd. for $C_{20}H_{33}$, 273.2582) (Supporting Information File 1, Figure S2) detected by high-resolution electrospray ionization mass spectrometry (HRESIMS), the molecular formula of **1** was deduced as $C_{20}H_{32}$, indicating that **1** has five degrees of unsaturation. The ^{13}C NMR spectrum showed that there are four olefinic carbons (δ_C 153.2, 137.4, 125.1, 118.6) in **1**. We thus reasoned that **1** features a tricyclic system. Subsequently, the extensive NMR analysis established the planar structure of **1**, and its relative configuration was partially assigned as 2*S**,3*S**,6*R**,10*R** by the NOESY spectrum (Supporting Information File 1, Table S3 and Figures S3–S8). As to the stereochemistry of C11, quantum chemical calculations of ^{13}C NMR chemical shifts were performed, which enabled us to determine that C18 and H10 are located on opposite sides of the five-membered ring (Supporting Information File 1, Figures S9 and S10). The conclusion was in agreement with NOE correlations between H10 and Ha12, and between $H_3$18 and Hb12 (Supporting Information File 1, Table S3). The absolute configuration of **1** was later determined based on its oxidized product **4** generated by TadB. Based on these results, TadA was experimentally determined as a new FC-type DTS, which catalyzes the formation of talaro-7,13-diene (**1**, Scheme 1A).

According to the HRESIMS spectrum and NMR analysis (Supporting Information File 1, Table S4 and Figures S11–S17), **2** was established as the C19-hydroxylated form of **1**. As during heterologous expression in *A. oryzae*, shunt products could sometimes be generated by endogenous enzymes [29,30], we performed feeding experiments to test whether **2** was generated by *A. oryzae* NSAR1. The result showed that the heterologous host indeed converts **1** to talaro-7,13-dien-19-ol (**2**, Scheme 1A; Supporting Information File 1, Figure S18).

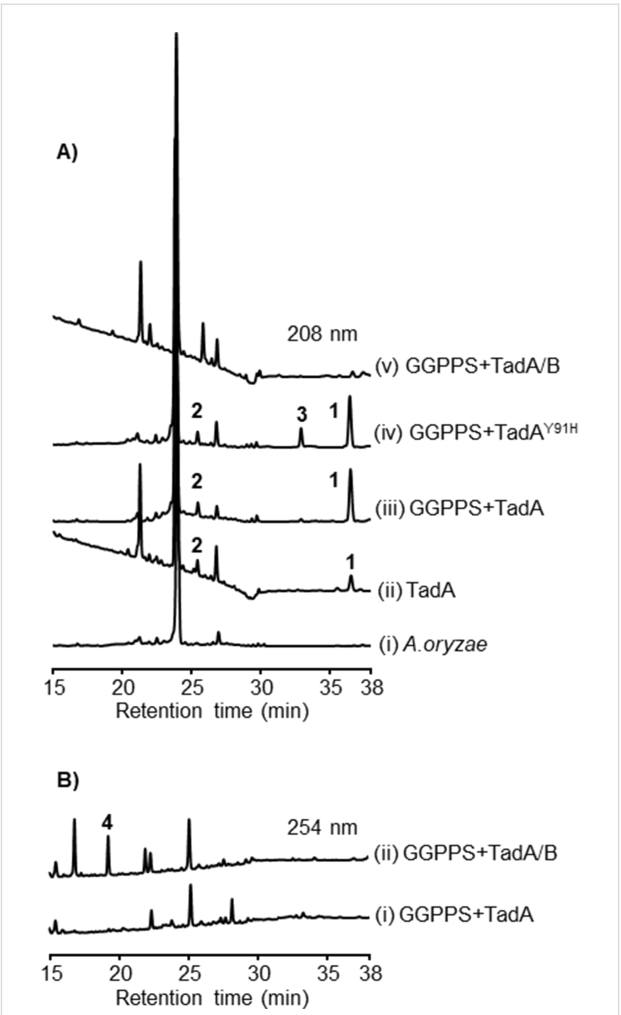
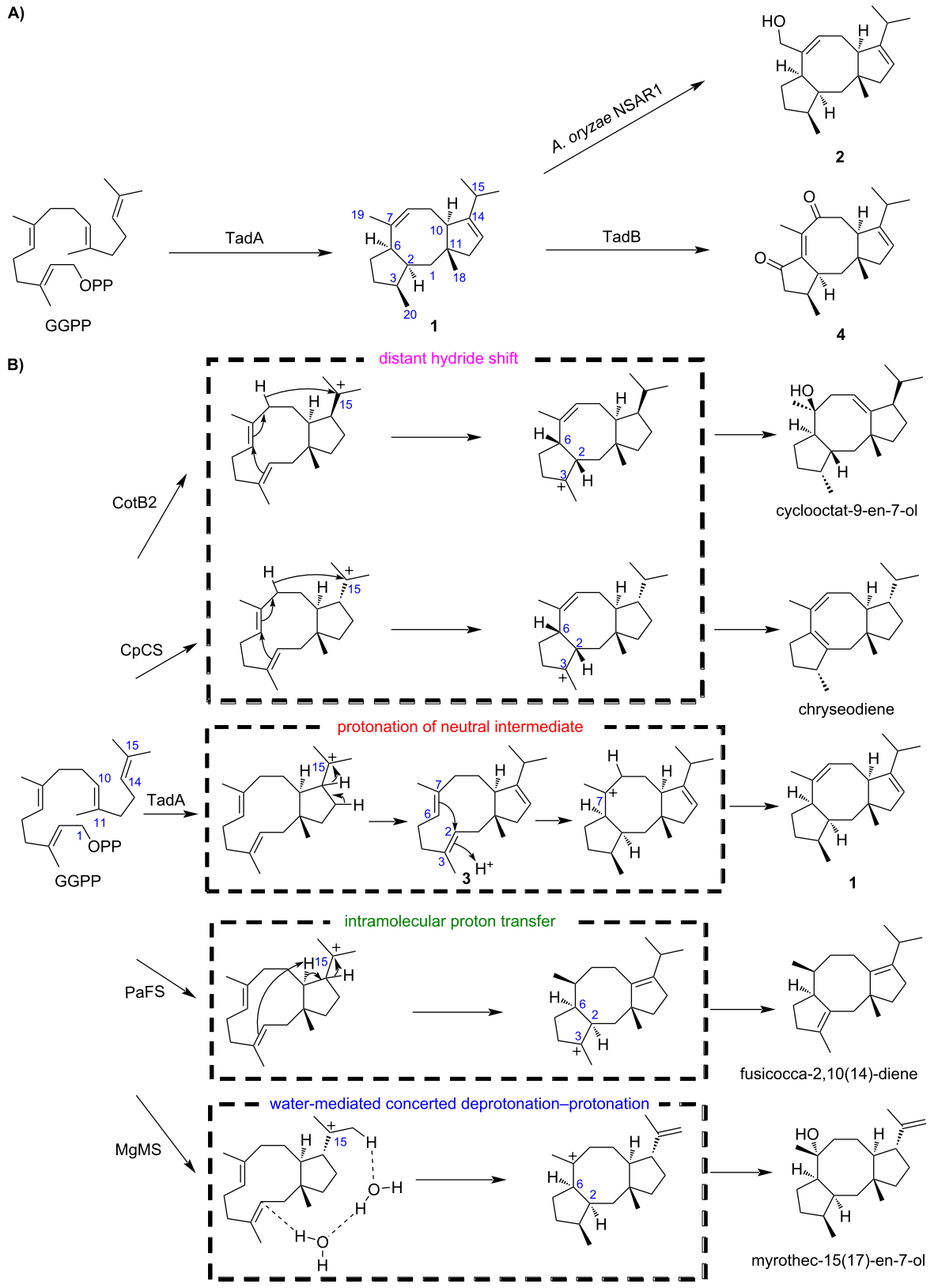


Figure 3: HPLC–MS analysis of mycelial extracts from *A. oryzae* NSAR1 transformants. A) The HPLC profiles monitored at 208 nm. B) The HPLC profiles monitored at 254 nm.

Mechanistic characterization of TadA

So far, cyclization mechanisms of FC-type diterpenes afforded by PaFS [18], MgMS [20], CotB2 [19], and CpCS [21] have been deciphered. All these enzymes undergo a common C1,11–C10,14-bicyclization to form a C15 carbocation, but differ a lot at the following C2,6 cyclization (Scheme 1B). CotB2 and CpCS trigger the C2,6 cyclization via a distant hydride shift, whereas PaFS employs an intramolecular proton transfer. We recently showed that a water-mediated concerted deprotonation–protonation is required for the MgMS-mediated cyclization [20]. In order to probe the mechanism underlying the cyclization of **1**, we used His_6-tagged TadA to carry out in



Scheme 1: Biosynthesis of FC-type diterpenoids. A) The biosynthetic pathway of **1**, **2** and **4**. B) Cyclization mechanisms of **1** and reported FC-type diterpenes.

vitro enzymatic reactions with or without addition of deuterated water (D_2O). GC–MS analysis showed that when the reaction mixture was supplemented with D_2O, the ion peak at *m*/*z* 273 was observed, indicating that exogenous deuterium was incorporated into **1** (Supporting Information File 1, Figure S19). This suggests that TadA might adopt a similar strategy as MgMS to initiate C2,6 cyclization though protonation at C2 by bulky water.

To obtain further insights into the C2,6-cyclization process of TadA, its three-dimensional (3D) protein structure was constructed with SWISS-MODEL using PaFS (PDB entry 5er8) as the template, and the proposed bicyclic neutral intermediate was docked into the active pocket of TadA (Supporting Information File 1, Figure S20). We searched for the amino acid residues surrounding C2 or C3, which might be involved in the C2 protonation, and found the candidate residue Tyr91. In the corresponding site, PaFS possesses a histidine residue (Supporting Information File 1, Figure S20). To test its role, Tyr91 in TadA was mutated to His, and then analyzed by an in vitro enzymatic assay. The result showed that the variant could give an additional product **3** (Supporting Information File 1, Figure S21). For isolation of **3**, the mutated *tadA*, along with GGPPS gene, was introduced into *A. oryzae* NSAR1, and the resulted transformant produced **3** at a titer of 0.6 mg/L (Figure 3A, line iv). By comparison of NMR data and specific optical rotation values, together with quantum chemical calculations of ^{13}C NMR chemical shifts, **3** was determined to be (3a*S*,5*E*,9*E*,12a*R*)-3,3a,4,7,8,11,12,12a-octahydro-3a,6,10-trimethyl-1-(1-methylethyl)cyclopentacycloundecene [31] (Supporting Information File 1, Figures S22–S24), indicating that Tyr91 is an essential amino acid residue involved in C2,6-cyclization. Intriguingly, through careful examination of the GC–MS profile of the transformant expressing *tadA*, we also observed the appearance of **3**, the content of which is rather lower than that in the $tadA^{Y91H}$-containing transformant (Supporting Information File 1, Figure S25). Based on these, we propose that though both MgMS and TadA use protonation-induced C2,6 cyclization, TadA likely adopts a more asynchronous process to give a neutral intermediate **3** first followed by protonation to form **1**, which is different from the highly concerted deprotonation–protonation process employed by MgMS (Scheme 1B). Further isotope labeling experiments and density functional theory (DFT) calculations are needed so as to gain deeper insight into the cyclization mechanism of **1**.

Functional analysis of the cytochrome P450 enzyme TadB

Due to the significance of tailoring enzymes in terms of structural diversification and bioactivity improvement [11], we then turned to the associated cytochrome P450 gene *tadB*. We found that introduction of *tadB* into the transformant possessing *tadA* and the GGPPS gene could result in disappearance of **1**, but no additional products were observed at 208 nm (Figure 3A, line v). When the detective wavelength was switched from 208 nm to 254 nm, additional product **4** was detected (Figure 3B, lines i and ii), which was elucidated to be the highly oxygenated product of **1** through the exhaustive NMR analysis (Supporting Information File 1, Table S5 and Figures S26–S32). Based on NMR data and the fact that **4** is derived from **1**, the relative configuration of **4** was assigned as 2*S**,3*S**,10*R**,11*R**. And as the experimental electronic circular dichroism (ECD) spectrum of **4** resembled the calculated ECD spectrum of (2*S*,3*S*,10*R*,11*R*)-**4**, the absolute configuration of **4** was determined to be 2*S*,3*S*,10*R*,11*R* (Supporting Information File 1, Figures S33 and S34). The highly oxidized properties of talaro-6,13-dien-5,8-dione (**4**) indicate that TadB is a multifunctional P450 enzyme (Scheme 1A). Since no other intermediates were obtained, we could not determine the exact order of TadB-mediated oxidation, or exclude the possibility that endogenous enzymes from *A. oryzae* NSAR1 was involved in formation of **4**.

In addition, elucidating the stereochemistry of **4** also allowed us to assign the absolute configurations of **1** and **2** as 2*S*,3*S*,6*R*,10*R*,11*R*, raising a possibility that **1** might be the biosynthetic precursor of roussoellol C, a cytotoxic FC-type diterpenoid isolated from *T. purpurogenus* PP-414 [7]. It will be important to elucidate the biosynthetic pathway of roussoellol C, providing enzymatic tools for expanding the chemical diversity of talaro-7,13-diene related FC-type diterpenoids via combinational biosynthesis [14].

Conclusion

We have identified a new fungal FC-type DTS, which is responsible for the biosynthesis of talaro-7,13-diene (**1**). Further mechanistic studies revealed that 2,6-cyclization in the formation of **1** is likely to be triggered by protonation of the neutral intermediate **3**, and Tyr91 in TadA plays a significant role in this process. The associated P450 enzyme TadB can catalyze multiple oxidation of **1** to highly oxygenated product talaro-6,13-dien-5,8-dione (**4**). This study has expanded the enzyme inventory for structural diversification of FC-type diterpenoids.

Supporting Information

Supporting Information File 1

Experimental methods, nucleotide sequence, tables, and figures.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-144-S1.pdf]

Acknowledgements

We thank Prof. K. Gomi (Tohoku University) and Prof. K. Kitamoto (The University of Tokyo) for providing the *A. oryzae* NSAR1 heterologous expression system.

Funding

This work was financially supported by grants from National Key Research and Development Program of China (2018YFA0903200 / 2018YFA0903201), the National Natural Science Foundation of China (82073046, 81925037, 32170060, 31870032, 22177037), the 111 Project of Ministry of Education of the People's Republic of China (B13038), National Foreign Experts Project (G2021199001L, China), National High-level Personnel of Special Support Program (2017RA2259, China), the Guangdong Natural Science Funds for Distinguished Young Scholar (2019B151502014, 2022B1515020028, China), Guangdong International Science and Technology Cooperation Base (2021A0505020015, China), Guangdong Science and Technology Planning Project (2020A0505100041, China), Local Innovative and Research Teams Project of Guangdong Pearl River Talents Program (2017BT01Y036, China), Innovative and Research Teams Project of Guangdong Higher Education Institution (2021KCXTD001, China), Guangzhou Science and Technology Project (202102010114, China), and Shenzhen Institute of Synthetic Biology Scientific Research Program (DWKF20210002, China).

ORCID® iDs

Dan Hu - https://orcid.org/0000-0002-1104-8646

Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2022.47.v1

References

1. Gershenzon, J.; Dudareva, N. *Nat. Chem. Biol.* **2007,** *3,* 408–414. doi:10.1038/nchembio.2007.5
2. de Boer, A. H.; de Vries-van Leeuwen, I. J. *Trends Plant Sci.* **2012,** *17,* 360–368. doi:10.1016/j.tplants.2012.02.007
3. Valeur, E.; Narjes, F.; Ottmann, C.; Plowright, A. T. *Med. Chem. Commun.* **2019,** *10,* 1550–1568. doi:10.1039/c9md00263d
4. Kaplan, A.; Andrei, S. A.; van Regteren Altena, A.; Simas, T.; Banerjee, S. L.; Kato, N.; Bisson, N.; Higuchi, Y.; Ottmann, C.; Fournier, A. E. *Cell Chem. Biol.* **2020,** *27,* 657–667.e6. doi:10.1016/j.chembiol.2020.02.010
5. Kim, S.; Shin, D.-S.; Lee, T.; Oh, K.-B. *J. Nat. Prod.* **2004,** *67,* 448–450. doi:10.1021/np030384h
6. Li, F.; Sun, W.; Guan, J.; Lu, Y.; Lin, S.; Zhang, S.; Gao, W.; Liu, J.; Du, G.; Wang, J.; Zhu, H.; Qi, C.; Hu, Z.; Zhang, Y. *Org. Biomol. Chem.* **2018,** *16,* 8751–8760. doi:10.1039/c8ob02353k
7. Wang, W.; Wan, X.; Liu, J.; Wang, J.; Zhu, H.; Chen, C.; Zhang, Y. *Mar. Drugs* **2018,** *16,* 150. doi:10.3390/md16050150
8. Bie, Q.; Chen, C.; Yu, M.; Guo, J.; Wang, J.; Liu, J.; Zhou, Y.; Zhu, H.; Zhang, Y. *J. Nat. Prod.* **2019,** *82,* 80–86. doi:10.1021/acs.jnatprod.8b00694
9. Andrei, S. A.; de Vink, P.; Sijbesma, E.; Han, L.; Brunsveld, L.; Kato, N.; Ottmann, C.; Higuchi, Y. *Angew. Chem., Int. Ed.* **2018,** *57,* 13470–13474. doi:10.1002/anie.201806584
10. Bauman, K. D.; Butler, K. S.; Moore, B. S.; Chekan, J. R. *Nat. Prod. Rep.* **2021,** *38,* 2100–2129. doi:10.1039/d1np00032b
11. Huang, J.-H.; Lv, J.-M.; Wang, Q.-Z.; Zou, J.; Lu, Y.-J.; Wang, Q.-L.; Chen, D.-N.; Yao, X.-S.; Gao, H.; Hu, D. *Org. Biomol. Chem.* **2019,** *17,* 248–251. doi:10.1039/c8ob02832j
12. Zhang, W.-Y.; Zhong, Y.; Yu, Y.; Shi, D.-F.; Huang, H.-Y.; Tang, X.-L.; Wang, Y.-H.; Chen, G.-D.; Zhang, H.-P.; Liu, C.-L.; Hu, D.; Gao, H.; Yao, X.-S. *J. Nat. Prod.* **2020,** *83,* 3338–3346. doi:10.1021/acs.jnatprod.0c00675
13. Wilkinson, B.; Micklefield, J. *Nat. Chem. Biol.* **2007,** *3,* 379–386. doi:10.1038/nchembio.2007.7
14. Song, X.; Lv, J.; Cao, Z.; Huang, H.; Chen, G.; Awakawa, T.; Hu, D.; Gao, H.; Abe, I.; Yao, X. *Acta Pharm. Sin. B* **2021,** *11,* 1676–1685. doi:10.1016/j.apsb.2020.12.007
15. Matsuda, Y.; Mitsuhashi, T.; Lee, S.; Hoshino, M.; Mori, T.; Okada, M.; Zhang, H.; Hayashi, F.; Fujita, M.; Abe, I. *Angew. Chem., Int. Ed.* **2016,** *55,* 5785–5788. doi:10.1002/anie.201601448
16. Ye, Y.; Minami, A.; Mandi, A.; Liu, C.; Taniguchi, T.; Kuzuyama, T.; Monde, K.; Gomi, K.; Oikawa, H. *J. Am. Chem. Soc.* **2015,** *137,* 11846–11853. doi:10.1021/jacs.5b08319
17. Luo, P.; Lv, J.-M.; Xie, Y.-F.; Xiao, L.-Y.; Qin, S.-Y.; Chen, G.-D.; Luo, X.-Z.; Hu, D.; Gao, H. *Org. Chem. Front.* **2022,** *9,* 3057–3060. doi:10.1039/d2qo00408a
18. Toyomasu, T.; Tsukahara, M.; Kenmoku, H.; Anada, M.; Nitta, H.; Ohkanda, J.; Mitsuhashi, W.; Sassa, T.; Kato, N. *Org. Lett.* **2009,** *11,* 3044–3047. doi:10.1021/ol901063s
19. Meguro, A.; Motoyoshi, Y.; Teramoto, K.; Ueda, S.; Totsuka, Y.; Ando, Y.; Tomita, T.; Kim, S.-Y.; Kimura, T.; Igarashi, M.; Sawa, R.; Shinada, T.; Nishiyama, M.; Kuzuyama, T. *Angew. Chem., Int. Ed.* **2015,** *54,* 4353–4356. doi:10.1002/anie.201411923
20. Lin, F.-L.; Lauterbach, L.; Zou, J.; Wang, Y.-H.; Lv, J.-M.; Chen, G.-D.; Hu, D.; Gao, H.; Yao, X.-S.; Dickschat, J. S. *ACS Catal.* **2020,** *10,* 4306–4312. doi:10.1021/acscatal.0c00377
21. Lauterbach, L.; Goldfuss, B.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2020,** *59,* 11943–11947. doi:10.1002/anie.202004691
22. Toyomasu, T.; Tsukahara, M.; Kaneko, A.; Niida, R.; Mitsuhashi, W.; Dairi, T.; Kato, N.; Sassa, T. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104,* 3084–3088. doi:10.1073/pnas.0608426104
23. Kim, S.-Y.; Zhao, P.; Igarashi, M.; Sawa, R.; Tomita, T.; Nishiyama, M.; Kuzuyama, T. *Chem. Biol.* **2009,** *16,* 736–743. doi:10.1016/j.chembiol.2009.06.007
24. Kudo, F.; Matsuura, Y.; Hayashi, T.; Fukushima, M.; Eguchi, T. *J. Antibiot.* **2016,** *69,* 541–548. doi:10.1038/ja.2016.40
25. Minami, A.; Tajima, N.; Higuchi, Y.; Toyomasu, T.; Sassa, T.; Kato, N.; Dairi, T. *Bioorg. Med. Chem. Lett.* **2009,** *19,* 870–874. doi:10.1016/j.bmcl.2008.11.108
26. Jin, F. J.; Maruyama, J.-i.; Juvvadi, P. R.; Arioka, M.; Kitamoto, K. *FEMS Microbiol. Lett.* **2004,** *239,* 79–85. doi:10.1016/j.femsle.2004.08.025
27. He, Y.; Wang, B.; Chen, W.; Cox, R. J.; He, J.; Chen, F. *Biotechnol. Adv.* **2018,** *36,* 739–783. doi:10.1016/j.biotechadv.2018.02.001

28. Wang, G.-Q.; Chen, G.-D.; Qin, S.-Y.; Hu, D.; Awakawa, T.; Li, S.-Y.; Lv, J.-M.; Wang, C.-X.; Yao, X.-S.; Abe, I.; Gao, H. *Nat. Commun.* **2018,** *9,* 1838. doi:10.1038/s41467-018-04298-2
29. He, Y.; Cox, R. J. *Chem. Sci.* **2016,** *7,* 2119–2127. doi:10.1039/c5sc04027b
30. Lv, J.-M.; Gao, Y.-H.; Zhao, H.; Awakawa, T.; Liu, L.; Chen, G.-D.; Yao, X.-S.; Hu, D.; Abe, I.; Gao, H. *Angew. Chem., Int. Ed.* **2020,** *59,* 13531–13536. doi:10.1002/anie.202004364
31. Ioannou, E.; Quesada, A.; Rahman, M. M.; Gibbons, S.; Vagias, C.; Roussis, V. *Eur. J. Org. Chem.* **2012,** 5177–5186. doi:10.1002/ejoc.201200533

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Navigating and expanding the roadmap of natural product genome mining tools

Friederike Biermann[‡1,2,§], Sebastian L. Wenski[‡1,2,§] and Eric J. N. Helfrich[*1,2]

Perspective

Open Access

Address:
[1]Institute for Molecular Bio Science, Goethe University Frankfurt, Max-von-Laue Str. 9, 60438 Frankfurt am Main, Germany and [2]LOEWE Center for Translational Biodiversity Genomics (TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

Email:
Eric J. N. Helfrich[*] - eric.helfrich@bio.uni-frankfurt.de

* Corresponding author ‡ Equal contributors
§ Authors are listed in alphabetical order.

Keywords:
genome mining; natural product biosynthesis; non-canonical pathways; PKS; NRPS; RiPP

Beilstein J. Org. Chem. **2022,** *18,* 1656–1671.
https://doi.org/10.3762/bjoc.18.178

Received: 16 August 2022
Accepted: 02 November 2022
Published: 06 December 2022

This article is part of the thematic issue "Enzymes in biosynthesis".

Associate Editor: J. S. Dickschat

Abstract

Natural products are structurally highly diverse and exhibit a wide array of biological activities. As a result, they serve as an important source of new drug leads. Traditionally, natural products have been discovered by bioactivity-guided fractionation. The advent of genome sequencing technology has resulted in the introduction of an alternative approach towards novel natural product scaffolds: Genome mining. Genome mining is an in-silico natural product discovery strategy in which sequenced genomes are analyzed for the potential of the associated organism to produce natural products. Seemingly universal biosynthetic principles have been deciphered for most natural product classes that are used to detect natural product biosynthetic gene clusters using pathway-encoded conserved key enzymes, domains, or motifs as bait. Several generations of highly sophisticated tools have been developed for the biosynthetic rule-based identification of natural product gene clusters. Apart from these hard-coded algorithms, multiple tools that use machine learning-based approaches have been designed to complement the existing genome mining tool set and focus on natural product gene clusters that lack genes with conserved signature sequences. In this perspective, we take a closer look at state-of-the-art genome mining tools that are based on either hard-coded rules or machine learning algorithms, with an emphasis on the confidence of their predictions and potential to identify non-canonical natural product biosynthetic gene clusters. We highlight the genome mining pipelines' current strengths and limitations by contrasting their advantages and disadvantages. Moreover, we introduce two indirect biosynthetic gene cluster identification strategies that complement current workflows. The combination of all genome mining approaches will pave the way towards a more comprehensive understanding of the full biosynthetic repertoire encoded in microbial genome sequences.

Introduction

In 2002, the genome sequences of the model actinomycete *Streptomyces coelicolor* A3(2) [1] and the producer of the antiparasitic drug avermectin, *Streptomyces avermitilis* [2], were published. These index cases marked the transition from the pre- to the post-genomic era in microbial natural product (NP) research [3]. The introduction of next-generation sequencing technologies [4] has led to a constant decrease in sequencing costs [5]. As a result, the number of publicly available genome sequences has rapidly increased and paved the way for a completely new avenue: genome mining. Genome mining describes the targeted bioinformatic analysis of (meta-)genomes to identify gene clusters involved in the biosynthesis of NPs [3]. NPs have been shown to act as signaling metabolites (e.g., acylhomoserine lactones (**1**) [6]), siderophores (e.g., pyoverdines (**2**) [7]), virulence factors (e.g., malleicyprol (**3**) [8-10]), toxins (e.g., bongkrekic acid (**4**) [11]), antibacterial (e.g., vancomycin (**5**) [12]) or antifungal compounds (e.g., amphotericin B (**6**) [13]) (Figure 1). The identification of almost all clinically relevant antibiotics using bioactivity-guided fractionation approaches long before the beginning of the post-genomic era initiated the field of microbial NP research. In the "golden age" of antibiotic discovery from the 1940s to 1970s, microbes and especially bacteria have been identified as an almost untapped treasure trove for the discovery of bioactive NPs. For the longest time, researchers focused on a few talented NP producers, that have mainly been isolated from soil samples [14]. Since the low hanging fruits have been picked using traditional bioactivity-based workflows, this approach frequently results in the rediscovery of known metabolites. The introduction of genome mining revolutionized NP research and helped overcome the rediscovery problem frequently encountered using traditional approaches. Contrary to earlier estimations that

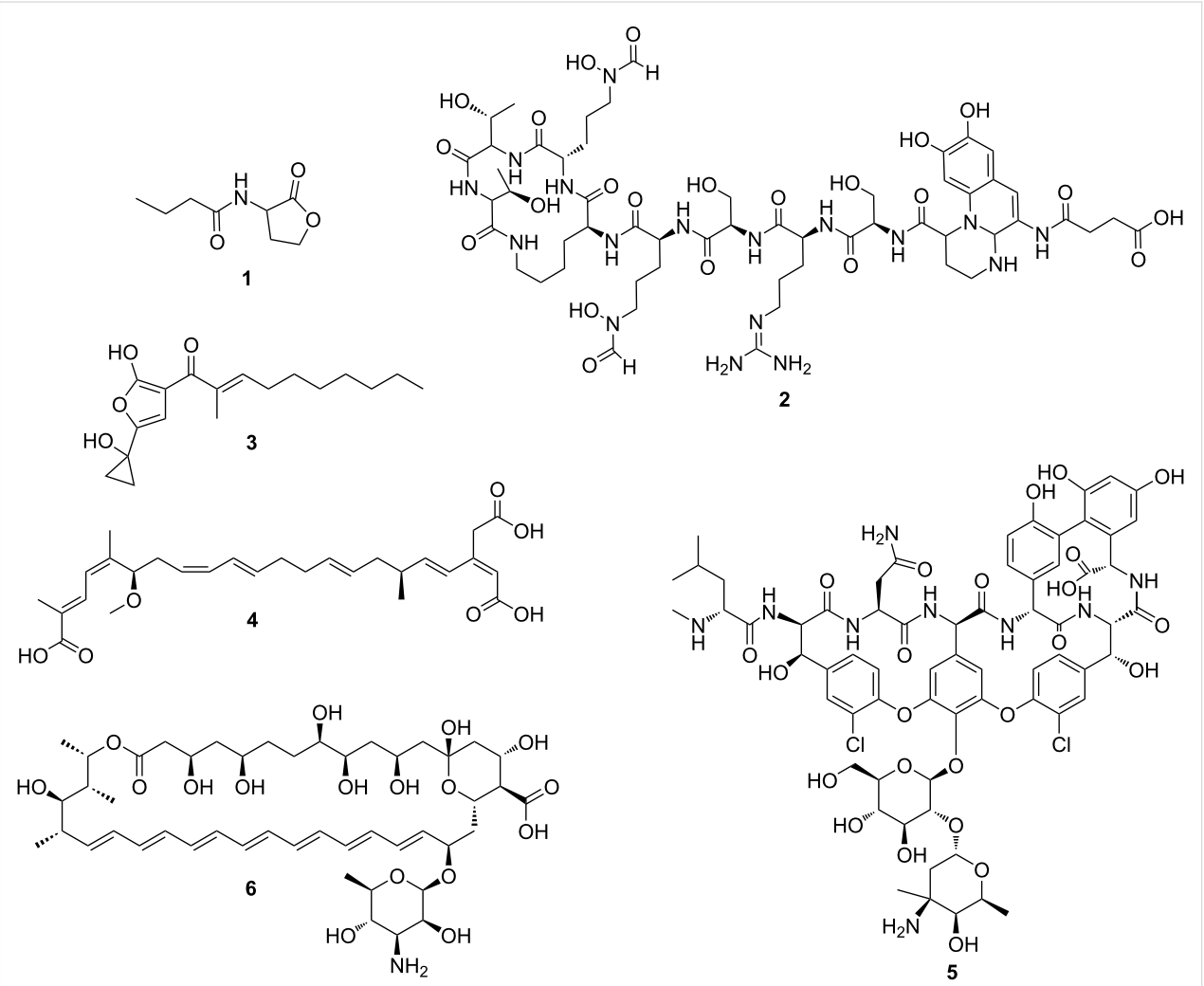


Figure 1: Examples of prominent natural products: *N*-butyrylhomoserine lactone (**1**), pyoverdin (**2**), malleicyprol (**3**), bongkrekic acid (**4**), vancomycin (**5**), and amphotericin B (**6**).

were based on bioactivity-guided discovery strategies, mining microbial genomes revealed a much higher biosynthetic potential than initially anticipated [14]. *Streptomyces hygroscopicus* sp. XM201, for instance, harbors more than 50 putative biosynthetic gene clusters (BGCs), many of which are cryptic, i.e., BGCs for which the corresponding NPs have yet to be identified [15]. A problem when it comes to the characterization of the full biosynthetic potential of an organism is the fact that many BGCs are silent. Silent BGC are not expressed under standard laboratory cultivation conditions as they might lack a specific ecological clue for their expression. As a result, two types of approaches have been developed to unleash this hidden biosynthetic potential. Several pleiotropic (non-targeted, e.g., modifying culturing conditions) and pathway-specific (e.g., heterologous expression or in situ pathway activation) approaches have been developed to awaken silent biosynthetic pathways [16]. Most importantly, however, genome mining can prevent the time-consuming re-discovery of already known metabolites [14]. In-silico dereplication can be performed on two levels: First, BGCs identified by genome mining can be compared to characterized BGCs [17]. Second, in many cases NP core structures can be predicted from genome sequence information and the predicted structures can then be used to search in NP databases for identical or related compounds [18,19]. While the BGC-centric approach might be more accurate, it is limited by the number of characterized BGCs in publicly available databases. Since significantly more NPs than NP BGCs are characterized, the search space of known metabolites is significantly larger than that of experimentally verified BGCs [20]. The accuracy of the predicted core structures on the other hand might restrict the approach.

In this perspective, we will take a closer look at the most commonly used state-of-the-art genome mining tools, ranging from algorithms based on hard-coded rules to machine learning (ML)-based approaches with regard to the natural product biosynthetic principles they are most suited for. We focus on how the different genome mining tools identify BGCs and highlight their advantages and limitations. Moreover, we will showcase two potential strategies for the targeted identification of non-canonical pathways to chart the full biosynthetic potential encoded in bacterial genomes.

Perspective

Natural product biosynthetic principles

NPs are structurally highly diverse and can be divided into several classes depending on their biosynthetic concepts. NP biosynthesis follows two fundamentally different principles: NPs can either be produced in an assembly line-like fashion (Figure 2A) or by discrete, multi-enzymatic assemblies (Figure 2B). Discrete, multi-enzymatic assemblies utilize monofunctional enzymes for the consecutive build-up and decoration of a NP scaffold. In comparison to biosynthetic assembly lines, intermediates are not permanently covalently bound to carrier proteins in discrete, multi-enzymatic assemblies. In both biosynthetic principles, the NP backbone is first assembled by core enzymes and then further modified by tailoring enzymes that decorate the NP scaffold.

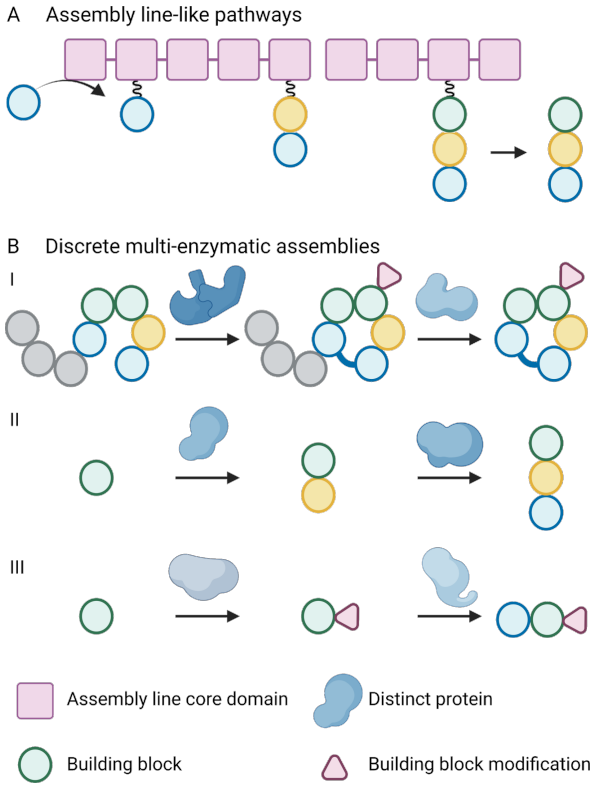


Figure 2: Biosynthetic principles of (A) assembly line-like pathways and (B) discrete multi-enzymatic assemblies. Assembly line-like pathways use large mega enzymes and generate NPs via the successive addition and/or modification of building blocks (e.g., non-ribosomal peptide biosynthesis) using conserved core domains. In discrete multi-enzymatic assemblies, distinct and mostly monofunctional enzymes catalyze the built-up of the NP scaffold and its decoration (e.g., in (I) ribosomally synthesized and post-translationally modified peptide, (II) terpene, or (III) alkaloid biosynthesis).

Assembly line-like pathways are characterized by mega enzymes, which can be subdivided into modules. Each module is responsible for the incorporation (and/or processing) of one building block into the nascent product. A “textbook” extension module minimally harbors three core domains, responsible for the activation and loading, tethering, and condensation of building blocks and intermediates. The biosynthesis is directional and starts at the N-terminal module with the activation and loading of the first building block onto the assembly line (Figure 2A) [21]. The specificity of the activating domain deter-

mines the type of building block incorporated. The growing intermediate stays permanently bound to the assembly line until the final product is released at the C-terminal module. However, modules can also be skipped, used for the modification of the nascent NP rather than its chain extension, or utilized iteratively [22,23]. In textbook assembly line-like pathways, the architecture of the mega enzyme complex correlates with the product structure, a principle that is referred to as the colinearity rule [24]. Examples of these assembly line-like pathways are canonical type I *cis*-acyltransferase polyketide synthases (PKSs) and type A non-ribosomal peptide synthetases (NRPSs) (Figure 2A) [25,26]. The substrate specificity of the specificity conferring domains in each module can be predicted from the sequences of adenylation (A) (for NRPS [26]), acyltransferase (AT) (for *cis*-AT PKS [15]), or ketosynthase (KS) domains (in *trans*-acyltransferase PKS systems [19,27]). Moreover, in the large majority of cases, the gene order within a BGC reflects the order of the corresponding enzymes during the biosynthesis of the associated NP [19]. *trans*-AT PKSs are much more complex than *cis*-AT PKS systems as they harbor non-elongating modules, cryptic domains and seemingly superfluous domains. Moreover, they frequently employ a number of *trans*-acting modifying enzymes, are characterized by modules that are split between proteins and they often harbor non-canonical module architectures and cryptic domains [19,22]. As a result, the colinearity rule cannot be applied to predict *trans*-AT PKS-derived polyketide core structures [19]. Instead, it has been observed that the amino acid sequences of the ketosynthase domains in *trans*-AT PKSs correlate with their substrate specificity [27]. This correlation can be used for the prediction of *trans*-AT PKS-derived polyketide core structures and is referred to as the correlation rule [19]. All commonly occurring domains in assembly line-like NP biosynthetic pathways as well as their non-modular homologs (e.g., type II and III PKSs) show a high degree of sequence homology. For that reason, their sequence can be used by genome mining tools as universal signature sequences to identify the genes encoding the respective domains and the remaining genes of the BGC (Figure 3A and B (e.g., bialaphos (**7**) [11])) [28].

In contrast, discrete multi-enzymatic assemblies utilize distinct, monofunctional enzymes. Examples are terpene (e.g., cyclooctatin (**8**) [29]), ribosomally synthesized and post-translationally modified peptide (RiPP), or NRPS-independent alkaloid pathways. In the case of terpene biosynthesis, terpene cyclases generate the oftentimes multicyclic, hydrocarbon scaffold via a carbocation-mediated cascade reaction [30]. Terpene cyclases are obligatory components of canonical terpene pathways and are used to identify terpene BGCs (Figure 3B) [30,31]. RiPPs, on the other hand, lack genes that are conserved across all 40 plus RiPP families [32]. However, each RiPP BGC family features genes encoding characteristic tailoring enzymes, or precursor peptides, that show a high degree of sequence conservation within the family. These conserved genes can be utilized for the targeted, family-specific identification of RiPP BGCs (Figure 3C (e.g., tryptorubin (**9**) [33])) [21]. In addition, multiple RiPP tailoring enzymes harbor a precursor peptide-binding domain, the so-called RiPP recognition element (RRE) (Figure 3D (e.g., pyrroloquinoline quinone (PQQ, **10**) [34])). RRE-derived signature motifs (i.e., short sequences that are conserved across different types of enzymes and that have a specific function) are used to identify RiPP BGCs beyond family borders as they are present in the BGCs of approximately 50% of all RiPP families [35]. BGCs without conserved signature sequences are almost impossible to identify using current bioinformatic approaches (Figure 3E (e.g., kojic acid (**11**) [36])). Therefore, the prediction of these BGCs is mainly based on the co-localization of adjacent genes encoding tailoring or additional core enzymes.

The current BGC prediction approach has its limitations, as genes involved in the biosynthesis of a NP might be dispersed (i.e., not clustered) throughout the genome and hence cannot be recognized by genome mining algorithms due to the missing proximity of the biosynthetic gene sets (BGSs) (Figure 3F (e.g., pyonitrin (**12**) [37])). NPs whose biosynthesis significantly deviates from the well-established biosynthetic principles (e.g., through the lack of signature sequences) (Figure 3E) [38] are frequently overlooked by state-of-the-art genome mining pipelines. Most genome mining algorithms rely on the identification of signature sequences (Figure 3A–D). As a result, BGCs of the most commonly studied NP classes (e.g., PKS and NRPS BGCs) can be identified with high confidence based on the sequence homology of the commonly occurring biosynthetic domains. Since chemical novelty in assembly line-like pathways is typically obtained through novel arrangements of a limited set of module architectures, a limited diversity of sequential module arrangements, and varying substrate specificities, the probability of identifying truly novel biosynthetic principles and biochemical transformations in these systems is restricted when using hard-coded biosynthetic principles that are based on the detection of the frequently encountered biosynthetic domains [21]. As a result, a lot of effort is currently being put into the development of complementing workflows to chart the “biosynthetic dark matter” (i.e., overlooked biosynthetic pathways) that we currently cannot access bioinformatically [39]. State-of-the-art genome mining tools are ideally suited for the detection of assembly line-like pathways. The focus on these pathways led to a strong bias in training sets: In the MIBiG database of characterized BGCs nearly 80% of all deposited NP BGCs are PKS, NRPS, or terpene BGCs (April 2022) [20]. As the largest database of characterized BGCs,

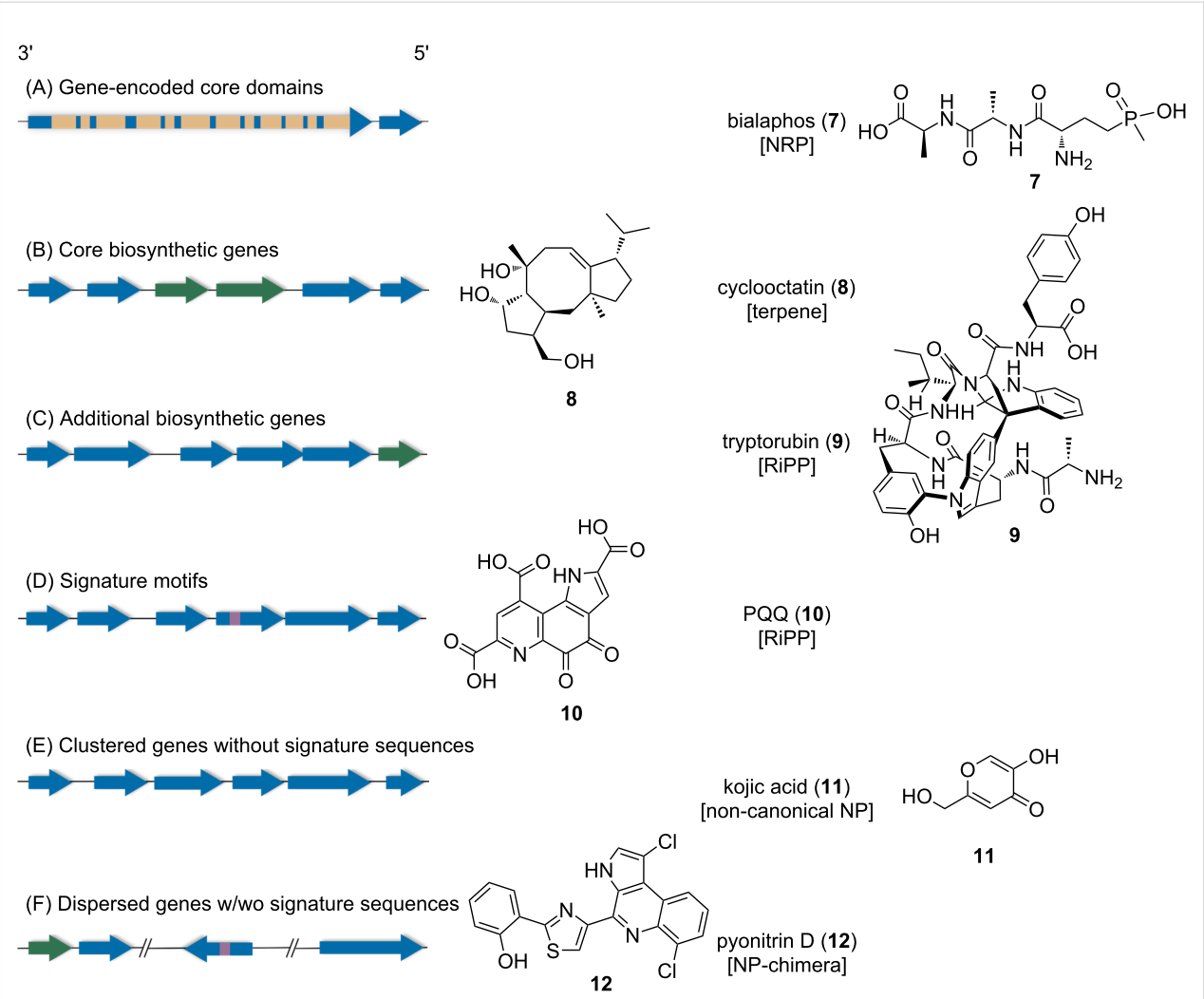


Figure 3: Universal (A, B, F) NP class and NP family-specific (C, D, F) signature sequences in NP BGCs and selected examples of each scenario. Genome mining tools utilize signature sequences to identify NP BGCs derived from (A) gene-encoded conserved core domains of assembly line-like pathways (e.g., modular NRPSs/PKSs), (B) distinct core enzymes (e.g., terpene or type II PKS biosynthesis), (C) tailoring enzymes (e.g., characterized families of RiPP biosynthetic pathways), (D) as well as signature motifs (e.g., RiPP biosynthetic pathways that utilize tailoring enzymes containing RREs). (E) BGCs without signature sequences (e.g., NRPS-independent alkaloid biosynthesis) or (F) genomically dispersed (i.e., not clustered) genes (here also referred to as biosynthetic gene sets) are difficult to identify. The enzymes encoded in core biosynthetic genes are responsible for assembling the NP backbone; additional biosynthetic genes encode tailoring enzymes and other components of a pathway (transporters, regulators, immunity enzymes). Conserved domains are depicted in yellow, genes in green and motifs in violet.

MIBiG is frequently used as a training data set for the development of genome mining algorithms. The imbalanced representation of NP BGCs in the database, however, might introduce a bias when it comes to the training of novel algorithms. Another obstacle to overcome is the efficient mining of the vast quantity of genomic data generated via next-generation sequencing, as a lot of genome mining algorithms are not capable of handling big data [40].

Genome mining principles and tools

Many genome mining tools are based on gene homology and rely on alignments of annotated open reading frames (ORFs). Yet, their purpose, functions, and additional features such as comparative analyses of BGCs, dereplication concepts, or NP structure prediction differ significantly. In addition, some tools implement alternative BGC identification methods like phylogenetic analyses or ML approaches. In many cases, these ML approaches are based on well-established strategies adopted from other disciplines (e.g., natural language processing or comparative genomics) that were adapted by the NP community [41].

In the following section, we will look at representative genome mining tools and discuss their underlying BGC detection princi-

ples, along with advantages and limitations of the BGC identification process.

Genome mining algorithms based on hard-coded biosynthetic principles

An early approach to identify NP BGCs in (meta-)genomic data sets were sequence alignments with known genes and domains using algorithms like BLAST (Figure 4) [42]. BLAST detects similar sequences to a given query sequence [42]. The first version of the tool BAGEL utilized BLAST analysis, among others, to identify putative BGCs of bacteriocins (= antimicrobial peptides and proteins) [43-46]. The advantage of such reference alignment methods that are based on sequence homology is their high confidence. The performance of these tools can be rapidly improved via the addition of new reference databases, which was contributing to their success at the beginning of the genome mining era. However, using BLAST-based approaches, the identification of real structural or biosynthetic novelty remains relatively sparse, as the BLAST algorithm is most suitable to detect close homologs of the query sequence. Up to this day, tools like BAGEL are predestined for the rapid and computationally cost-effective characterization of genomic data [43].

Hidden Markov Models (HMMs) are statistical models that are used by the NP community as a more flexible approach to identify BGCs (Figure 4). These models consist of a sequence of "states" (e.g., the occurrences of specific amino acids or nucleotides at a certain position of a protein or DNA sequence, respectively) with pre-determined transition probabilities from one state to the next (e.g., the transition probability in a sequence between one base at a given position to another base at the next position). A sequence of probabilities is calculated from given sequence alignments, for instance, of members of a given gene or protein family. By adding up all possibilities, the likelihood of the complete sequence being a member of the gene family can be calculated [47]. Derivatives of HMMs, so-called profile Hidden Markov Models (pHMMs), are additionally taking gaps and incomplete sequences into consideration. In addition to whole genes or proteins, sequences of conserved key domains of assembly line-like pathways like PKSs (e.g., acyl-carrier-proteins, AT or KS domains) [25] or NRPSs (e.g., peptidyl-carrier-proteins, A domains, condensation (C) domains) [26] are utilized for the generation of pHMMs. The resulting pHMMs recognize signature sequences of such conserved domains in genomic query sequences. pHMMs cannot only be employed to detect and annotate BGCs but also to predict substrate specificities that are essential for NP structure predictions [19,39,48]. After the identification of the core biosynthetic genes, co-localized genes are analyzed and the locus and borders of the BGC are predicted via hard-coded rules based on textbook biosynthetic knowledge, e.g., the minimum amount of domains in a typical NRPS. Due to their seemingly universal biosynthetic principles and modular composition, canonical PKS and NRPS BGCs are predestined for the high confidence detection of their encoded biosynthetic core domains using pHMMs. Structural novelty in these systems that predominantly comprise the same set of conserved domains arises from the novel arrangement of the limited set of different module architectures (e.g., around a dozen in *cis*-AT PKSs vs >150 in *trans*-AT PKSs [19]) along with varying substrate selectivities of specificity-conferring domains (e.g., A domains in NRPSs, AT domains in *cis*-AT PKSs, and KS domains in *trans*-AT PKSs). Moreover, since these assembly line-like pathways follow the same biosynthetic principle, they often form hybrids with other biosynthetic assembly line-like pathways [21].

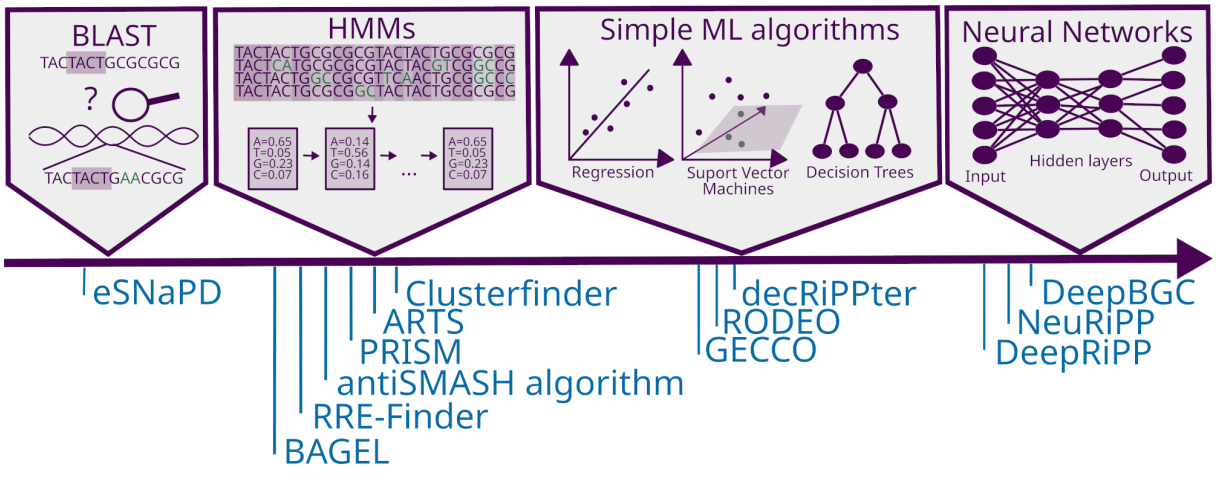


Figure 4: Concepts of algorithms in order of complexity and examples of genome mining tools that employ the respective concept.

Prominent examples of the usage of pHMMs are the original algorithm of the antibiotics & Secondary Metabolite Analysis Shell (antiSMASH) [17,29,49-52] as well as PRediction Informatics for Secondary Metabolomes (PRISM) [18,53-55]. In addition to PKSs and NRPSs, both tools identify a high number of NP classes and families using pHMMs (antiSMASH 6: 876 pHMMs, PRISM 4: 1772 pHMMS). Apart from BGC detection by pHMMs, several stand-alone tools have been implemented into antiSMASH to improve BGC identification, annotation, and substrate predictions (Table 1) (described in detail below). Therefore, we distinguish the original antiSMASH algorithm from the antiSMASH platform (Table 1). Although the BGC

Table 1: Purpose, principles, advantages, and disadvantages of selected genome mining tools. The upper part of the table contains hard-coded tools and the lower part ML-based tools. Novelty refers to the ability of a genome mining tool to chart non-canonical BGCs. Confidence refers to the ability of a genome mining tool to correctly identify a NP BGC.

Tool [first/latest version]	Purpose	BGC identification principles	(Dis-)advantages	Novelty	Confidence
antiSMASH algorithm [17] 2011/2021	identification of a broad range of NP BGC classes and families	pHMMs, hard-coded rules	comprehensive NP class detection	low	high
antiSMASH platform [17] 2011/2021 https://antismash.secondarymetabolites.org	identification of a broad range of NP BGC classes and families, functional and comparative analyses, structure prediction	ClusterFinder[a]: pHMM RODEO: BLAST, pHMM, SVMs RRE-Finder: pHMMs/HHpred database	comprehensive analysis covering many NP classes, dereplication via comparative analysis, usage of NP BGC databases	medium	high
ARTS [59] 2017/2020 http://arts.ziemertlab.com/index	target directed genome mining for antibiotics in bacteria via resistance genes	pHMMs for BGC prediction (antiSMASH), TIGRFAM for detection of housekeeping genes, phylogenetic analysis for identification of horizontal gene transfer	targeted approach for bioactivity	low	high
BAGEL [49] 2006/2018 http://bagel4.molgenrug.nl/	identification of bacterial bacteriocins and RiPPs in (meta-) genomic sequences	BLAST analysis, HMMs, hard-coded rules	restricted to RiPP and bacteriocin BGCs	low	high
CASSIS and SMIPS [60] 2016 https://sbi.hki-jena.de/cassis/	BGC detection in fungi	CASSIS: Density of transcription factor binding sites, SMIPS: Signature sequences	precise cluster borders	low	high
ClusterFinder [61] 2014 Implemented in antiSMASH	BGC detection without functional assignment of NP class	HMM for whole cluster	comprehensive NP class detection	high	low
eSNaPD [62] 2014 http://esnapd2.rockefeller.edu/	BGC detection in non-assembled bacterial metagenomic sequences	BLAST analysis against BGC database	comprehensive NP class detection of smaller BGCs that are similar to known BGCs	low	high
EvoMining [63] 2016/2019 https://github.com/nselem/evomining	identification of BGCs integrating evolutionary principles	phylogenomic analysis in combination with antiSMASH analysis	independent of commonly used signature sequences	medium	medium

Table 1: Purpose, principles, advantages, and disadvantages of selected genome mining tools. The upper part of the table contains hard-coded tools and the lower part ML-based tools. Novelty refers to the ability of a genome mining tool to chart non-canonical BGCs. Confidence refers to the ability of a genome mining tool to correctly identify a NP BGC. (continued)

PRISM [55] 2015/2020 https://prism.adapsyn.com/	identification of a broad range of NP BGCs, structure prediction	HMMs for BGC detection, BLAST analysis, protein motifs and HMMs for domain specificity prediction, support vector machines for activity prediction	comprehensive analysis covering many NP classes, several structure suggestions, dereplication via structural comparisons	low	high
SMURF [64] 2010 http://smurf.jcvi.org/run_smurf.php	identification of fungal BGCs	HMMs, hard-coded rules	comprehensive NP class detection	low	high
transATor [19] 2019	annotation of *trans*-AT PKSs and accurate structure predictions of *trans*-AT PKS-derived polyketides	pHMMs, hard-coded rules	restricted to *trans*-AT PKSs	low	high
decRiPPter [65] 2020 https://github.com/Alexamk/decRiPPter	identification of RiPP BGCs	SVMs, pan-genomic analyses	restricted to RiPP BGCs	medium	medium
DeepBGC [41] 2019 https://github.com/Merck/deepbgc	identification of bacterial and fungal BGCs	neural network with vector- represented Pfam domains (ML)	comprehensive NP class detection	high	medium
DeepRiPP [66] 2019 http://deepripp.magarveylab.ca	identification of RiPP BGCs, structure prediction	natural language processing (deep learning)	restricted to RiPP BGCs	medium	medium
GECCO [67] 2021 https://github.com/zellerlab/GECCO	identification of bacterial and fungal BGCs	conditional random fields	comprehensive NP class detection	high	medium
NeuRiPP [68] 2019 https://github.com/emzodls/neuripp	identification of RiPP precursors	neural networks	restricted to RiPP precursors	medium	medium
RODEO [69] 2017 https://rodeo.scs.illinois.edu/	identification of RiPP BGCs	BLAST analysis of tailoring enzymes, pHMMs, SVMs for precursor detection	restricted to RiPP BGCs	low	medium

[a]ClusterFinder is not available on the antiSMASH web server any longer but is incorporated into the standalone antiSMASH command line tool.

identification approach of antiSMASH and PRISM is quite similar, both tools differ in the downstream processing of the identified BGCs. While antiSMASH focuses on functional and comparative analyses of the biosynthetic genes and BGCs, the focus of PRISM lies on a comprehensive chemical structure prediction of the associated NP [56-58]. In silico dereplication to

eliminate BGCs associated with known NPs is one of the major functions of genome mining to avoid the time-consuming and costly re-isolation of known NPs. For instance, the antiSMASH platform compares putative BGCs with reference databases to detect BGCs that are similar to previously characterized BGCs [15,20,58]. However, as many NPs were isolated during the pre-genomic era, they have not been linked to their corresponding BGC. As a result, BGC databases are incomplete which is a drawback when it comes to the dereplication on a gene level. PRISM aims at overcoming this obstacle via retro-biosynthetic building block predictions of known NPs from multiple databases in combination with several BGC-derived NP structure suggestions [58].

To identify RiPP BGCs, the antiSMASH algorithm and PRISM utilize pHMMs based on RiPP-family-specific signature sequences derived from tailoring enzymes or precursor peptides (Figure 3C and D). These family-specific pHMMs are likewise used in tools like BAGEL or RODEO and enable the identification of novel members of known RiPP families [46,69]. RRE-Finder, which is integrated into the antiSMASH platform and RODEO, utilizes the presence of RREs, predicted via pHMMs, to detect RiPP BGCs (Figure 3D). Since the RRE motif is only present in approximately 50% of all RiPP families, it restricts the predictable biosynthetic space. Yet, RRE-Finder is one of the few RiPP genome mining tools which is capable of identifying RiPP BGCs in a family-independent manner [70].

Since the potential of identifying truly novel BGCs via signature sequences is limited, the tool ClusterFinder was developed and implemented into the command line version of antiSMASH [61]. ClusterFinder annotates BGCs via pHMMs from a string of contiguous Pfam domains (protein domains annotated in the protein family database) instead of individual genes. pHMMs are calculated using training sets of known BGCs and non-BGC sequences. Here, two states "BGC" and "non-BGC" are distinguished depending on the Pfam domain frequency in the training data set and the identities of adjacent domains. Consequently, the ClusterFinder algorithm is designed to detect BGCs that are overlooked by other biosynthetic pipelines. As in many other algorithms for the detection of true biosynthetic novelty, high false positive rates have to be taken into consideration, which makes the output of low-confidence/high novelty algorithms more difficult to interpret [61].

An alternative to the above mentioned classical genome mining approaches is the utilization of evolutionary information for the detection of NP BGCs. The EvoMining concept is based on the assumption that secondary metabolite biosynthetic enzymes are distant paralogs of enzymes involved in primary metabolism [63,71]. These NP biosynthetic enzymes are hypothesized to have undergone significant sequence and selectivity changes while still operating based on the same reaction mechanism (e.g., fatty acid biosynthesis → polyketide biosynthesis). As such, NP biosynthetic pathways utilize members of existing enzyme families that have evolved to perform new metabolic functions. Consequently, NP BGCs "borrow" genes encoding paralogs of enzymes that have their origin in primary metabolism and that have diverged into catalyzing alternative metabolic functions. That way, the EvoMining approach identifies members of biosynthetic enzyme families that have likely been repurposed and thus, their corresponding genes are prime targets for a closer inspection of the genomic context to identify new types of BGCs. Although EvoMining is a signature sequence independent concept and instead uses phylogenetic analysis of primary metabolite biosynthetic enzymes, it remains a "hard-coded" sequence similarity-based approach that uses phylogenetic analysis instead of pHMMs for BGC detection [63,71].

Machine learning-based genome mining tools

Some NP BGCs contain solely family-specific features, and lack universal class-specific signature sequences. In these cases, only members of the same subfamily can be identified via pHMMs. An example of the latter are RiPPs that are the most rapidly expanding NP subclass. Eighteen new RiPP families have been characterized over the span of just 8 years, suggesting that many more RiPP families have yet to be discovered [32]. To exploit these currently overlooked biosynthetic treasures, multiple recently developed genome mining tools make use of ML algorithms that have been adapted from other research fields like image recognition [65-68]. Most ML-based tools utilize "supervised learning," a strategy that employs a dataset with known classifications to train the algorithm [72]. Traditional ML algorithms include regression, decision tree-based classifiers, and support vector machines (SVM), which construct a hyperplane that splits the *n*-dimensional data-space (i.e., different features/categories serve as dimensions of this space) into different areas that correspond to the different classes (Figure 4) [72]. These algorithms usually lead to robust and interpretable predictions but are limited when it comes to solving complex problems [72].

An example of an advanced combination of different approaches and methods for the identification of RiPPs is the Data-driven Exploratory Class-independent RiPP TrackER (decRiPPter) [65]. decRiPPter uses a support vector machine algorithm trained on a set of known precursor genes to detect RiPP precursor genes semi-independently of their subclass. Subsequently, a pan-genome analysis is performed to identify the corresponding BGCs with the putative RiPP precursor genes as seeds. Putative NP BGCs are identified that are organized in

operon-like structures and prioritized based on the taxonomic distribution of the cluster. decRiPPter was successfully used for the identification of a new lanthipeptide subfamily, providing experimental validation of the algorithm [65].

A more advanced form of supervised learning is deep learning (Figure 4). An example of a deep learning architecture is the artificial neural network inspired by the human brain architecture. It consists of artificial neurons processing information organized in different layers and connected by synapses [73]. These advanced algorithms often provide higher accuracy in their prediction but are no longer interpretable as a result of their high level of abstraction [73]. NeuRiPP, for instance, utilizes a parallel convolutional neural network to predict novel RiPP precursor genes independent of their RiPP family. The neural network is trained on a RiPP precursor training set that is based on experimentally verified precursors and precursors predicted by other tools [68]. Both RiPP-specific tools, NeuRiPP and decRiPPter, allow a more flexible BGC identification than hard-coded algorithms but are biased in that the precursor identification depends on training sets consisting of precursors from known RiPP families.

In contrast to NeuRiPP and decRiPPter, DeepBGC is not restricted to a single NP class. Comparable to some hard-coded algorithms, DeepBGC is based on HMM-generated Pfam annotations. However, instead of utilizing Pfam-domains as features, it converts the arrays of Pfam annotations into numeric vectors using a shallow two-layer neural network, an approach adopted from natural language processing [41,72]. These high-dimensional vectors are then used as input for a second two-layer neural network trained on a set of BGC and non-BGC sequences to predict NP BGCs. In the last step, the NP class is predicted using a random forest classifier (Figure 4) [41,74]. DeepBGC outperforms ClusterFinder in its accuracy and false-positive rates due to its ML approach. Like with many other tools, a major disadvantage of DeepBGC is that BGCs lacking canonical biosynthetic domains and small BGCs are filtered out in a pruning stage. Consequently, small BGCs (e.g., biarylitides **15** [75] and tryptorubins **9** [33]) or those that feature solely atypical biosynthetic genes are not recognized, which reduces the likelihood of identifying true biosynthetic novelty [41].

A similar approach is utilized by GECCO, that uses conditional random fields on arrays of Pfam annotations [67]. Conditional random fields belong to the statistical methods and can be classified between HMMs and simpler machine learning algorithms. An advantage of conditional random fields is their interpretability [67]. GECCO outperforms rule-based models in terms of novelty and DeepBGC in terms of accuracy while being less computationally expensive than both [67]. Like DeepBGC, GECCO currently lacks functional proof for the identification of a novel natural product guided by the tool [67].

Challenges and potential solutions to identify currently overlooked BGCs

Genome mining was pivotal for the expansion of NP chemical space in the past two decades. Despite the development of more and more sophisticated genome mining platforms, in many cases where truly novel NP scaffolds were described, the NP was isolated first and only then linked to its corresponding BGC [38]. Notable examples include the (thio-)peptides polytheonamide A (**13**) [76], closthioamide (**15**) [77] (Figure 5), and tryptorubin A (**9**) [33]. It was not until the structure of each of these peptides was determined, that manual retrobiosynthetic analysis resulted in the proposal of biosynthetic models that were subsequently experimentally verified. Once the biosynthesis of a NP is determined using this approach, the NP family can be expanded by developing genome mining algorithms to identify BGCs that follow similar biosynthetic principles [56].

One crucial challenge in the development of novel genome mining tools is balancing novelty and confidence, as one tends to fall short as the other is optimized [39]. On the one hand, genome mining tools that are focused on detecting non-canonical BGCs (high-novelty) are usually characterized by the identification of many putative BGCs that might not be involved in NP biosynthesis (high false-positive rate). These false-positive BGCs are automatically pruned and the resulting putative BGCs need to undergo a second round of manual verification and prioritization prior to functional characterization [39]. On the other hand, hard-coded algorithms detect BGC with high confidence but are restricted when it comes to the identification of BGCs that deviate significantly from what the algorithm's pHMMs have been trained to identify (true biosynthetic novelty) [39]. As most algorithms are at least to some extent signature sequence or sequence homology based, they heavily rely on the sequence space of known BGCs. The bias of hard-coded algorithms is embedded in the biosynthetic rules used for BGC detection and the dataset used to create pHMMs. The bias of ML-based algorithms results from their training sets that usually consist of characterized, canonical BGCs that are then used for the targeted identification of non-canonical BGCs [39]. Utilizing fewer gene family-based features, like the occurrence of Pfam domains or the sequence itself, for predictions can help to avoid overfitting, i.e., the problem of getting the algorithm to perform very well on the training data but underperform on unseen data [39].

Most genome mining algorithms rely on functionally annotated ORFs for the prediction of BGCs. State-of-the-art genome annotation algorithms are not yet able to recognize all ORFs

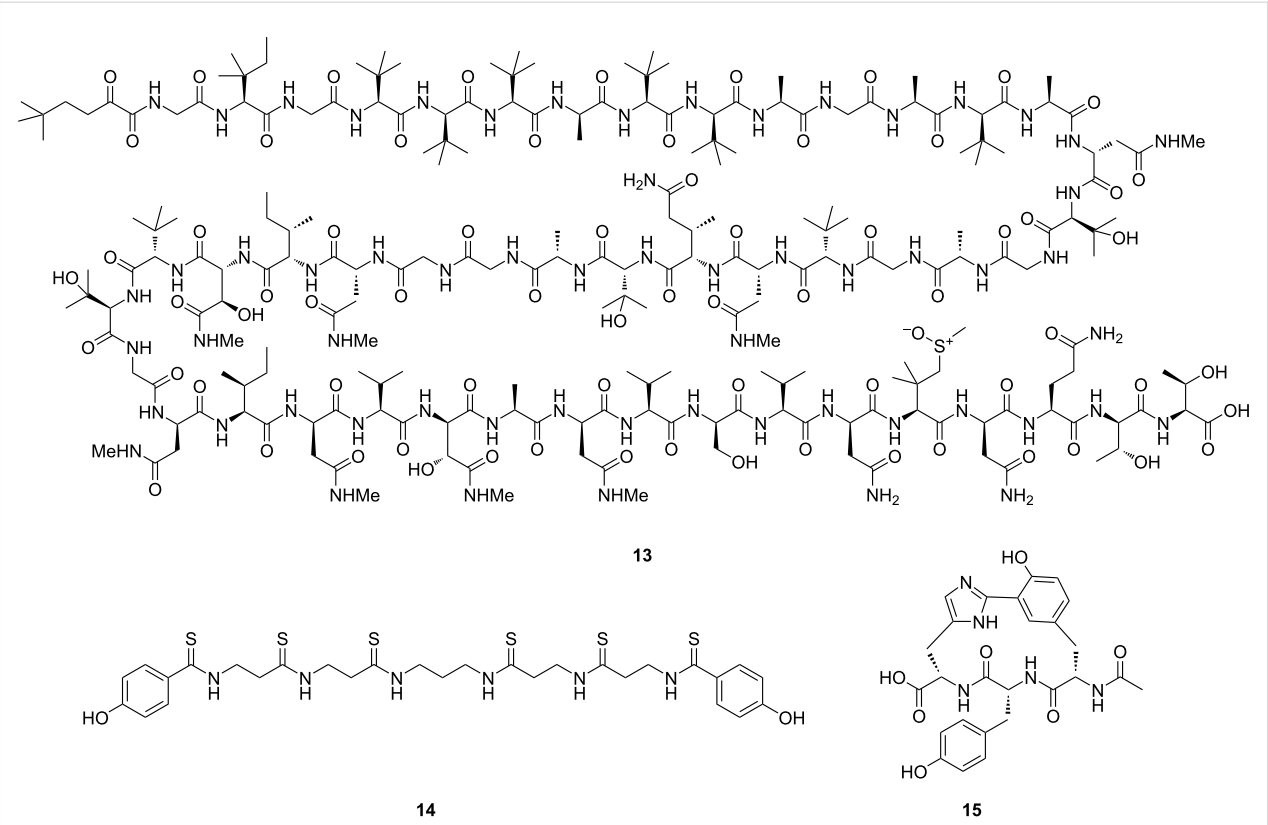


Figure 5: Examples of peptide NPs, the corresponding BGCs of which were determined through retrobiosynthetic analysis and then experimentally verified: polytheonamide A (**13**) [76], closthioamide (**14**) [77], and biarylitide YYH (**15**) [33].

correctly, especially very short ORFs like RiPP precursor genes [78]. Combined with many false ORF annotations, missing annotations impair BGC predictions downstream of the annotation process. Moreover, the BGCs of certain NP families are inherently easier to identify than others. For example, domains of canonical NRPSs and PKSs can be identified by signature sequence-based pHMMs with high confidence (Figure 3). Furthermore, pHMMs of conserved domains can be subdivided into dozens of individual pHMMs used to determine the substrate specificity of a conserved domain [17,19]. However, BGCs lacking known signature sequences are inherently more difficult to identify. In addition, the size of the BGC of interest impacts the predictive power of the algorithms: Extremely small BGCs, harboring only a few genes, are frequently overlooked as they usually do not pass hard-coded thresholds. For instance, the 1.2 kb gene cluster linked to tryptorubin (**9**) biosynthesis only encodes a 26 amino acid precursor peptide and a single cytochrome P450 monooxygenase [33,79], and hence it was overlooked by genome mining algorithms. On the other hand, large PKS or NRPS BGCs can be split across multiple contigs. This mosaic-like distribution of a single BGC makes the identification of the entire BGC a challenging endeavor especially if multiple assembly line-like BGCs are present in a genome. Moreover, the quality of assembled genomes obtained from short reads decreases with highly repetitive sequences present in many large PKS or NRPS genes [39].

Although the traditional hard-coded rule- and ML-based approaches differ fundamentally when it comes to the implementation of the respective NP BGC identification, they are both based on the same principle: The direct identification of NP BGCs. Both approaches heavily rely on training sets to generate pHMMs or to train the respective ML algorithm. As a consequence, they are both hypothesis-driven approaches resulting in an inherent bias “to identify what the algorithm was trained to identify” rather than to chart the entire biosynthetic space. This bias is largely based on the fact that both approaches use the characterized NP biosynthetic space as a training set for its expansion. Even though there might be no truly unbiased approach towards the expansion of NP biosynthetic space, indirect NP BGC detection methods might be capable of complementing the current strategies. These indirect approaches are exclusively based on the assumption that NP biosynthetic genes are clustered in microbial genomes (even though this might not be true for all NP biosynthetic pathways) and do not require prior knowledge about characterized biosyn-

thetic pathways as training data sets. Below, we are showcasing two putative solutions to complement existing approaches to expand NP biosynthetic space and to chart biosynthetic dark matter.

Genome-wide characterization of all clustered genes as an approach to identify non-canonical pathways

One concept that is based on the above outlined indirect approach is the genome-wide characterization of all clustered genes (gcBGC). In comparison to state-of-the-art genome mining tools, gcBGC inverts the current BGC identification process. Instead of identifying NP BGCs, all clustered genes involved in primary and secondary metabolite biosynthesis are identified. To specifically target non-canonical BGCs, BGCs that can be unambiguously assigned to primary metabolism and those BGCs that are detected by state-of-the-art genome mining pipelines are filtered out. Based on the initial hypothesis underlying the gcBGC approach, the remaining BGCs are likely involved in non-canonical NP biosynthesis ("biosynthetic dark matter" in Figure 6).

The gcBGC concept is based on the assumption that secondary metabolite BGCs evolve from primary metabolite biosynthetic pathways, and that the transition between both is fluid [71]. First, gcBGC identifies all clustered genes in a signature sequence-independent manner via analysis of operon-like structures (e.g., promoters or transcription start sites) as shown in fungi by the tool CASSIS/SMIPS [60]. This concept contrasts the commonly used principles that rely on the direct detection of genes via (p)HMMs- or ML-based approaches, both of which typically require a training data set.

As this approach leads to the identification of a large number of primary and secondary metabolite BGCs that are likewise detected by state-of-the-art genome mining pipelines, a filtering step is required to prioritize the putative non-canonical BGCs that are currently overlooked by existing genome mining tools [17,41,55,80] (Figure 6). Moreover, additional information on taxonomic relationships, pan-genome analyses, or whole-genome comparisons of all members of the pan-genome can be used for further prioritization (Figure 6) [81]. gcBGC is restricted to well-studied organisms where primary metabolite gene cassettes can be confidentially identified. However, the inverted BGC identification concept combined with the focus on as-of-yet unidentified BGCs suggests gcBGC-like approaches to be promising alternatives for the detection of non-canonical pathways.

A comparative genomics approach to identify non-canonical BGCs

Another concept for the expansion of NP biosynthetic space is based on a Comparative Genomics Approach (CGA). This approach relies on the fact that many BGCs are introduced into microbial genomes via horizontal gene transfer (HGT). A genome can be subdivided into groups of genes called syntenic blocks [82]. Among related strains, the order of these syntenic blocks, as well as their gene composition, is highly similar

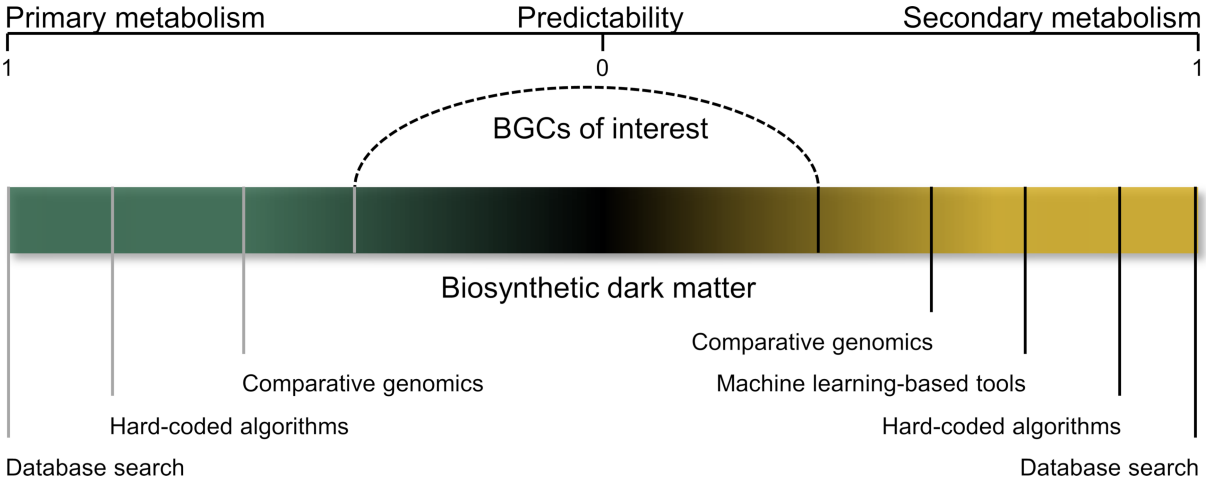


Figure 6: gcBGC prioritization process. The bar represents all identified gene clusters in one organism and is subdivided into primary (green) on the left and secondary metabolism (yellow) on the right. Gene clusters associated with primary or secondary metabolism (both ends of the spectrum) are identified using the strategies listed below the bar and subsequently excluded from further analysis. gcBGC aims at identifying as-of-yet overlooked BGCs that cannot be detected by state-of-the-art bioinformatic platforms, here referred to as "biosynthetic dark matter."

(Figure 7). Evolutionary young HGT events in single/few strains can disrupt this order, leading to the insertion of non-syntenic blocks (Figure 7) [82]. These insertions are detectable by comparing multiple closely related strains utilizing whole genome alignments, a technique adopted from the field of comparative genomics [83]. In a recent study, 10 *Aspergillus* genomes were compared to identify BGCs in non-syntenic blocks, leading to the confirmation of all previously known BGCs using the CGA concept [84]. As a proof of concept, the previously characterized kojic acid (**11**) BGC, which escaped detection by state-of-the-art genome mining algorithms, was identified [84]. The kojic acid (**11**) BGC lacks the classical biosynthetic signature sequences typically used for BGC identification, thus showing the potential of the approach (Figure 3) [84].

CGA aims at scaling this approach and comparing all sequenced strains of one genus (e.g., *Streptomyces*) to find non-syntenic blocks that might code for NP BGCs. Comparable to the genome-wide characterization of all clustered genes concept, CGA focuses on BGC detection independently of signature sequences and known NP families to expand the known NP chemical space via the identification of non-canonical pathways.

The first step of CGA consists of the homogenous functional annotation of all genes of the selected genomes to reduce false positive rates of non-syntenic blocks due to different annotations of genes using different annotation algorithms. Subsequently, all annotated genes are clustered based on sequence similarity to improve functional annotations [85]. The obtained sequential arrangements of gene annotations representing the different genomes are aligned to compare the genomes not on a sequence level, but instead on the gene-function level [86].

Whole genome alignments are performed to detect single diverging gene loci that are subsequently expanded by their genomic neighborhood to detect genomic islands. Therefore, the genomic neighborhoods of the identified genes are analyzed for differences in their synteny to detect HGT regions composed of multiple genes. In addition, these regions are analyzed for genetic characteristics like promoters or transposase genes to identify operon-like structures. Single gene duplication events are filtered out and all known BGCs are excluded in a similar fashion as in the gcBGC approach. The presence of prototypical tailoring enzymes ubiquitously distributed in secondary metabolism might serve as an additional line of evidence for a functional NP BGC. This approach is computationally expensive yet feasible with the availability of high-performance computer clusters but requires excellent quality of the analyzed genomes. This method has the advantage over simpler approaches to detect HGT events, like for example comparing GC contents of different regions, that it can be used to detect HGT events from closely related strains.

Conclusion

The development of next-generation sequencing technologies [4] and the resulting availability of a seemingly exponentially increasing number of genome sequences enabled or revolutionized several biological fields including comparative genomics [81], functional genomics [87], and NP-genome mining [88]. From simple BLAST analyses through pHMM-based algorithms to ML-based approaches, genome mining is a continuously evolving field that has benefited from other disciplines, such as mathematics, image processing, or linguistics. State-of-the-art sequence homology- and ML-based genome mining tools identify BGCs that share even low levels of similarity with known BGCs with high confidence. Traditional pHMMs-based approaches are ideally suited to chart the biosynthetic space of

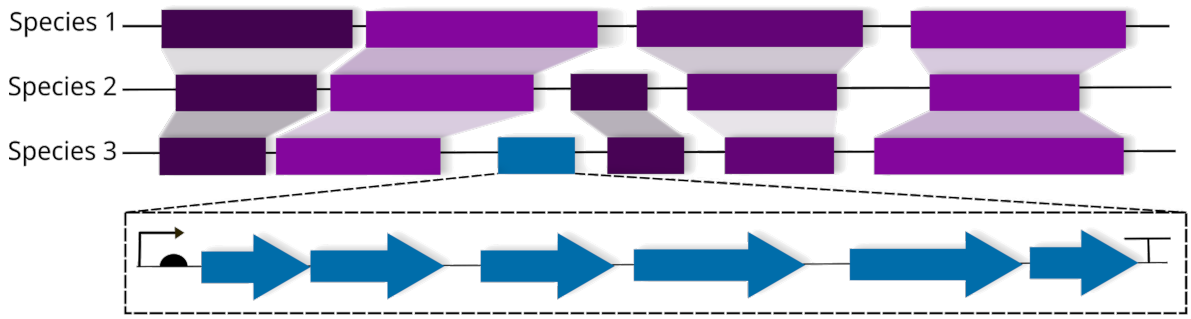


Figure 7: Genome alignments of related organisms revealing the presence of syntenic (purple) as well as non-syntenic blocks (blue). The order of syntenic blocks can be disrupted by putative HGT events, leading to the integration of non-syntenic blocks that are subsequently screened for operon-like structures containing multiple continuous gene arrangements (blue) without large gaps.

assembly line-like pathways that are typically composed of novel arrangements of recurring module architectures with varying specifications of the substrate specificity-conferring domains. ML-based approaches on the other hand are more frequently employed to target non-homogeneous NP classes such as RiPPs whose BGCs do not share sequence homologies across all 40 plus RiPP-families and to identify NP BGCs that are currently overlooked by state-of-the-art sequence homology-based tools. Even though the scope and implementation of both approaches differs significantly, the underlying concept is the same: The direct, hypothesis-driven identification of clustered NP biosynthetic genes based on a training data set that requires a database of characterized BGCs. This training data set might comprise individual domains from characterized pathways to generate pHMMs to complex features that are extracted from characterized BGCs. The bias introduced through the dependence on these reference datasets is likely to result in an inherent limitation when it comes to the identification of truly non-canonical pathways that share low to no similarity to characterized pathways. To address these limitations, indirect approaches that do not rely on training data sets of characterized BGCs might be capable of complementing the current suite of highly sophisticated genome mining tools as they might be ideally suited to identify non-canonical pathways that are overlooked by direct identification approaches. We showcased two such hypothetical indirect approaches that we named “genome-wide characterization of all clustered genes” and “comparative genomics-based identification of non-canonical BGCs”. These indirect BGC detection concepts are solely based on the assumption that biosynthetic genes are clustered in bacterial genomes. Both approaches are based on the sequence similarity-independent identification of non-canonical BGCs via recognition of operon-like structures or usage of comparative genomics to detect horizontally transferred gene clusters. In a subsequent prioritization step, clustered genes that are involved in primary metabolite biosynthesis or that can be likewise detected by state-of-the-art genome mining pipelines can be excluded to target uncharted biosynthetic space also referred to as biosynthetic dark matter. These indirect concepts might serve as an inspiration for further innovative tools for the targeted discovery of hidden biosynthetic treasures.

Acknowledgements

Figure 1 and the graphical abstract were created in BioRender.

Funding

EJNH gratefully acknowledges funding by the LOEWE Center for Translational Biodiversity Genomics (LOEWE TBG) and the German Chemical Industry Association. FB acknowledges funding from a Kekulé Fellowship of the German Chemical Industry Association.

ORCID® iDs

Friederike Biermann - https://orcid.org/0000-0002-9152-2562
Eric J. N. Helfrich - https://orcid.org/0000-0001-8751-3279

References

1. 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
2. Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Ōmura, S. *Nat. Biotechnol.* **2003,** *21,* 526–531. doi:10.1038/nbt820
3. Bachmann, B. O.; Van Lanen, S. G.; Baltz, R. H. *J. Ind. Microbiol. Biotechnol.* **2014,** *41,* 175–184. doi:10.1007/s10295-013-1389-9
4. Goodwin, S.; McPherson, J. D.; McCombie, W. R. *Nat. Rev. Genet.* **2016,** *17,* 333–351. doi:10.1038/nrg.2016.49
5. van Dijk, E. L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. *Trends Genet.* **2018,** *34,* 666–681. doi:10.1016/j.tig.2018.05.008
6. Waters, C. M.; Bassler, B. L. *Annu. Rev. Cell Dev. Biol.* **2005,** *21,* 319–346. doi:10.1146/annurev.cellbio.21.012704.131001
7. Wendenbaum, S.; Demange, P.; Dell, A.; Meyer, J. M.; Abdallah, M. A. *Tetrahedron Lett.* **1983,** *24,* 4877–4880. doi:10.1016/s0040-4039(00)94031-0
8. Biggins, J. B.; Ternei, M. A.; Brady, S. F. *J. Am. Chem. Soc.* **2012,** *134,* 13192–13195. doi:10.1021/ja3052156
9. Franke, J.; Ishida, K.; Hertweck, C. *Angew. Chem., Int. Ed.* **2012,** *51,* 11611–11615. doi:10.1002/anie.201205566
10. Trottmann, F.; Franke, J.; Richter, I.; Ishida, K.; Cyrulies, M.; Dahse, H.-M.; Regestein, L.; Hertweck, C. *Angew. Chem., Int. Ed.* **2019,** *58,* 14129–14133. doi:10.1002/anie.201907324
11. Moebius, N.; Ross, C.; Scherlach, K.; Rohm, B.; Roth, M.; Hertweck, C. *Chem. Biol.* **2012,** *19,* 1164–1174. doi:10.1016/j.chembiol.2012.07.022
12. Felnagle, E. A.; Jackson, E. E.; Chan, Y. A.; Podevels, A. M.; Berti, A. D.; McMahon, M. D.; Thomas, M. G. *Mol. Pharmaceutics* **2008,** *5,* 191–211. doi:10.1021/mp700137g
13. Caffrey, P.; Lynch, S.; Flood, E.; Finnan, S.; Oliynyk, M. *Chem. Biol.* **2001,** *8,* 713–723. doi:10.1016/s1074-5521(01)00046-1
14. Baltz, R. H. *J. Ind. Microbiol. Biotechnol.* **2019,** *46,* 281–299. doi:10.1007/s10295-018-2115-4
15. Kautsar, S. A.; Blin, K.; Shaw, S.; Weber, T.; Medema, M. H. *Nucleic Acids Res.* **2021,** *49,* D490–D497. doi:10.1093/nar/gkaa812
16. Liu, Z.; Zhao, Y.; Huang, C.; Luo, Y. *Front. Bioeng. Biotechnol.* **2021,** *9,* 632230. doi:10.3389/fbioe.2021.632230
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. Skinnider, M. A.; Dejong, C. A.; Rees, P. N.; Johnston, C. W.; Li, H.; Webster, A. L. H.; Wyatt, M. A.; Magarvey, N. A. *Nucleic Acids Res.* **2015,** *43,* 9645–9662. doi:10.1093/nar/gkv1012

19. Helfrich, E. J. N.; Ueoka, R.; Dolev, A.; Rust, M.; Meoded, R. A.; Bhushan, A.; Califano, G.; Costa, R.; Gugger, M.; Steinbeck, C.; Moreno, P.; Piel, J. *Nat. Chem. Biol.* **2019,** *15,* 813–821. doi:10.1038/s41589-019-0313-7
20. 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
21. Wenski, S. L.; Thiengmag, S.; Helfrich, E. J. N. *Synth. Syst. Biotechnol.* **2022,** *7,* 631–647. doi:10.1016/j.synbio.2022.01.007
22. Helfrich, E. J. N.; Piel, J. *Nat. Prod. Rep.* **2016,** *33,* 231–316. doi:10.1039/c5np00125k
23. He, J.; Hertweck, C. *ChemBioChem* **2005,** *6,* 908–912. doi:10.1002/cbic.200400333
24. Traitcheva, N.; Jenke-Kodama, H.; He, J.; Dittmann, E.; Hertweck, C. *ChemBioChem* **2007,** *8,* 1841–1849. doi:10.1002/cbic.200700309
25. Hertweck, C. *Angew. Chem., Int. Ed.* **2009,** *48,* 4688–4716. doi:10.1002/anie.200806121
26. Süssmuth, R. D.; Mainz, A. *Angew. Chem., Int. Ed.* **2017,** *56,* 3770–3821. doi:10.1002/anie.201609079
27. Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. *Nat. Biotechnol.* **2008,** *26,* 225–233. doi:10.1038/nbt1379
28. Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. *Nucleic Acids Res.* **2011,** *39* (Suppl. 2), W339–W346. doi:10.1093/nar/gkr466
29. Aoyama, T.; Naganawa, H.; Muraoka, Y.; Aoyagi, T.; Takeuchi, T. *J. Antibiot.* **1992,** *45,* 1703–1704. doi:10.7164/antibiotics.45.1703
30. Helfrich, E. J. N.; Lin, G.-M.; Voigt, C. A.; Clardy, J. *Beilstein J. Org. Chem.* **2019,** *15,* 2889–2906. doi:10.3762/bjoc.15.283
31. Dickschat, J. S. *Nat. Prod. Rep.* **2016,** *33,* 87–110. doi:10.1039/c5np00102a
32. 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
33. Reisberg, S. H.; Gao, Y.; Walker, A. S.; Helfrich, E. J. N.; Clardy, J.; Baran, P. S. *Science* **2020,** *367,* 458–463. doi:10.1126/science.aay9981
34. Evans, R. L., III; Latham, J. A.; Xia, Y.; Klinman, J. P.; Wilmot, C. M. *Biochemistry* **2017,** *56,* 2735–2746. doi:10.1021/acs.biochem.7b00247
35. Burkhart, B. J.; Hudson, G. A.; Dunbar, K. L.; Mitchell, D. A. *Nat. Chem. Biol.* **2015,** *11,* 564–570. doi:10.1038/nchembio.1856
36. Terabayashi, Y.; Sano, M.; Yamane, N.; Marui, J.; Tamano, K.; Sagara, J.; Dohmoto, M.; Oda, K.; Ohshima, E.; Tachibana, K.; Higa, Y.; Ohashi, S.; Koike, H.; Machida, M. *Fungal Genet. Biol.* **2010,** *47,* 953–961. doi:10.1016/j.fgb.2010.08.014
37. Mevers, E.; Saurí, J.; Helfrich, E. J. N.; Henke, M.; Barns, K. J.; Bugni, T. S.; Andes, D.; Currie, C. R.; Clardy, J. *J. Am. Chem. Soc.* **2019,** *141,* 17098–17101. doi:10.1021/jacs.9b09739
38. Biermann, F.; Helfrich, E. J. N. *mSystems* **2021,** *6,* e00846-21. doi:10.1128/msystems.00846-21
39. Blin, K.; Kim, H. U.; Medema, M. H.; Weber, T. *Briefings Bioinf.* **2019,** *20,* 1103–1113. doi:10.1093/bib/bbx146
40. Baltz, R. H. *J. Ind. Microbiol. Biotechnol.* **2021,** *48,* kuab044. doi:10.1093/jimb/kuab044
41. Hannigan, G. D.; Prihoda, D.; Palicka, A.; Soukup, J.; Klempir, O.; Rampula, L.; Durcak, J.; Wurst, M.; Kotowski, J.; Chang, D.; Wang, R.; Piizzi, G.; Temesi, G.; Hazuda, D. J.; Woelk, C. H.; Bitton, D. A. *Nucleic Acids Res.* **2019,** *47,* e110. doi:10.1093/nar/gkz654
42. 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
43. 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
44. de Jong, A.; van Heel, A. J.; Kok, J.; Kuipers, O. P. *Nucleic Acids Res.* **2010,** *38* (Suppl. 2), W647–W651. doi:10.1093/nar/gkq365
45. van Heel, A. J.; de Jong, A.; Montalbán-López, M.; Kok, J.; Kuipers, O. P. *Nucleic Acids Res.* **2013,** *41,* W448–W453. doi:10.1093/nar/gkt391
46. de Jong, A.; van Hijum, S. A. F. T.; Bijlsma, J. J. E.; Kok, J.; Kuipers, O. P. *Nucleic Acids Res.* **2006,** *34,* W273–W279. doi:10.1093/nar/gkl237
47. Eddy, S. R. *Bioinformatics* **1998,** *14,* 755–763. doi:10.1093/bioinformatics/14.9.755
48. Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* **2011,** *39* (Suppl. 2), W362–W367. doi:10.1093/nar/gkr323
49. Blin, K.; Medema, M. H.; Kazempour, D.; Fischbach, M. A.; Breitling, R.; Takano, E.; Weber, T. *Nucleic Acids Res.* **2013,** *41,* W204–W212. doi:10.1093/nar/gkt449
50. Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015,** *43,* W237–W243. doi:10.1093/nar/gkv437
51. Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H. *Nucleic Acids Res.* **2017,** *45,* W36–W41. doi:10.1093/nar/gkx319
52. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019,** *47,* W81–W87. doi:10.1093/nar/gkz310
53. Skinnider, M. A.; Merwin, N. J.; Johnston, C. W.; Magarvey, N. A. *Nucleic Acids Res.* **2017,** *45,* W49–W54. doi:10.1093/nar/gkx320
54. Skinnider, M. A.; Johnston, C. W.; Edgar, R. E.; Dejong, C. A.; Merwin, N. J.; Rees, P. N.; Magarvey, N. A. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* E6343–E6351. doi:10.1073/pnas.1609014113
55. Skinnider, M. A.; Johnston, C. W.; Gunabalasingam, M.; Merwin, N. J.; Kieliszek, A. M.; MacLellan, R. J.; Li, H.; Ranieri, M. R. M.; Webster, A. L. H.; Cao, M. P. T.; Pfeifle, A.; Spencer, N.; To, Q. H.; Wallace, D. P.; Dejong, C. A.; Magarvey, N. A. *Nat. Commun.* **2020,** *11,* 6058. doi:10.1038/s41467-020-19986-1
56. Medema, M. H.; de Rond, T.; Moore, B. S. *Nat. Rev. Genet.* **2021,** *22,* 553–571. doi:10.1038/s41576-021-00363-7
57. 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

58. Dejong, C. A.; Chen, G. M.; Li, H.; Johnston, C. W.; Edwards, M. R.; Rees, P. N.; Skinnider, M. A.; Webster, A. L. H.; Magarvey, N. A. *Nat. Chem. Biol.* **2016,** *12,* 1007–1014. doi:10.1038/nchembio.2188
59. Mungan, M. D.; Alanjary, M.; Blin, K.; Weber, T.; Medema, M. H.; Ziemert, N. *Nucleic Acids Res.* **2020,** *48,* W546–W552. doi:10.1093/nar/gkaa374
60. Wolf, T.; Shelest, V.; Nath, N.; Shelest, E. *Bioinformatics* **2016,** *32,* 1138–1143. doi:10.1093/bioinformatics/btv713
61. Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Wieland Brown, L. C.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J.; Birren, B. W.; Takano, E.; Sali, A.; Linington, R. G.; Fischbach, M. A. *Cell* **2014,** *158,* 412–421. doi:10.1016/j.cell.2014.06.034
62. Reddy, B. V. B.; Milshteyn, A.; Charlop-Powers, Z.; Brady, S. F. *Chem. Biol.* **2014,** *21,* 1023–1033. doi:10.1016/j.chembiol.2014.06.007
63. Sélem-Mojica, N.; Aguilar, C.; Gutiérrez-García, K.; Martínez-Guerrero, C. E.; Barona-Gómez, F. *Microb. Genomics* **2019,** *5,* e000260. doi:10.1099/mgen.0.000260
64. Khaldi, N.; Seifuddin, F. T.; Turner, G.; Haft, D.; Nierman, W. C.; Wolfe, K. H.; Fedorova, N. D. *Fungal Genet. Biol.* **2010,** *47,* 736–741. doi:10.1016/j.fgb.2010.06.003
65. Kloosterman, A. M.; Cimermancic, P.; Elsayed, S. S.; Du, C.; Hadjithomas, M.; Donia, M. S.; Fischbach, M. A.; van Wezel, G. P.; Medema, M. H. *PLoS Biol.* **2020,** *18,* e3001026. doi:10.1371/journal.pbio.3001026
66. Merwin, N. J.; Mousa, W. K.; Dejong, C. A.; Skinnider, M. A.; Cannon, M. J.; Li, H.; Dial, K.; Gunabalasingam, M.; Johnston, C.; Magarvey, N. A. *Proc. Natl. Acad. Sci. U. S. A.* **2020,** *117,* 371–380. doi:10.1073/pnas.1901493116
67. Carroll, L. M.; Larralde, M.; Fleck, J. S.; Ponnudurai, R.; Milanese, A.; Cappio, E.; Zeller, G. *bioRxiv* **2021**. doi:10.1101/2021.05.03.442509
68. de los Santos, E. L. C. *Sci. Rep.* **2019,** *9,* 13406. doi:10.1038/s41598-019-49764-z
69. Tietz, J. I.; Schwalen, C. J.; Patel, P. S.; Maxson, T.; Blair, P. M.; Tai, H.-C.; Zakai, U. I.; Mitchell, D. A. *Nat. Chem. Biol.* **2017,** *13,* 470–478. doi:10.1038/nchembio.2319
70. Kloosterman, A. M.; Shelton, K. E.; van Wezel, G. P.; Medema, M. H.; Mitchell, D. A. *mSystems* **2020,** *5,* e00267–20. doi:10.1128/msystems.00267-20
71. de los Santos, E. L. C. *Sci. Rep.* **2019,** *9,* 13406. doi:10.1038/s41598-019-49764-z
72. Nasteski, V. *Horizons* **2017,** *4,* 51–62.
73. Alzubaidi, L.; Zhang, J.; Humaidi, A. J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M. A.; Al-Amidie, M.; Farhan, L. *J. Big Data* **2021,** *8,* 53. doi:10.1186/s40537-021-00444-8
74. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. *arXiv* **2013,** 1301.3781. doi:10.48550/arxiv.1301.3781
75. Zdouc, M. M.; Alanjary, M. M.; Zarazúa, G. S.; Maffioli, S. I.; Crüsemann, M.; Medema, M. H.; Donadio, S.; Sosio, M. *Cell Chem. Biol.* **2021,** *28,* 733–739.e4. doi:10.1016/j.chembiol.2020.11.009
76. Oiki, S.; Muramatsu, I.; Matsunaga, S.; Fusetani, N. *Folia Pharmacol. Jpn.* **1997,** *110,* 195–198. doi:10.1254/fpj.110.supplement_195
77. Lincke, T.; Behnken, S.; Ishida, K.; Roth, M.; Hertweck, C. *Angew. Chem., Int. Ed.* **2010,** *49,* 2011–2013. doi:10.1002/anie.200906114
78. Chang, Y.-C.; Hu, Z.; Rachlin, J.; Anton, B. P.; Kasif, S.; Roberts, R. J.; Steffen, M. *Nucleic Acids Res.* **2016,** *44,* D330–D335. doi:10.1093/nar/gkv1324
79. Nanudorn, P.; Thiengmag, S.; Biermann, F.; Erkoc, P.; Dirnberger, S. D.; Phan, T. N.; Fürst, R.; Ueoka, R.; Helfrich, E. J. N. *Angew. Chem., Int. Ed.* **2022,** *61,* e202208361. doi:10.1002/anie.202208361
80. Pascal Andreu, V.; Roel-Touris, J.; Dodd, D.; Fischbach, M. A.; Medema, M. H. *Nucleic Acids Res.* **2021,** *49,* W263–W270. doi:10.1093/nar/gkab353
81. Costa, S. S.; Guimarães, L. C.; Silva, A.; Soares, S. C.; Baraúna, R. A. *Bioinf. Biol. Insights* **2020,** *14,* 1177932220938064. doi:10.1177/1177932220938064
82. Medema, M. H.; Fischbach, M. A. *Nat. Chem. Biol.* **2015,** *11,* 639–648. doi:10.1038/nchembio.1884
83. Beskrovnaya, P.; Melnyk, R. A.; Liu, Z.; Liu, Y.; Higgins, M. A.; Song, Y.; Ryan, K. S.; Haney, C. H. *mBio* **2020,** *11,* e01906-20. doi:10.1128/mbio.01906-20
84. Takeda, I.; Umemura, M.; Koike, H.; Asai, K.; Machida, M. *DNA Res.* **2014,** *21,* 447–457. doi:10.1093/dnares/dsu010
85. Emms, D. M.; Kelly, S. *Genome Biol.* **2015,** *16,* 157. doi:10.1186/s13059-015-0721-2
86. Armstrong, J.; Hickey, G.; Diekhans, M.; Fiddes, I. T.; Novak, A. M.; Deran, A.; Fang, Q.; Xie, D.; Feng, S.; Stiller, J.; Genereux, D.; Johnson, J.; Marinescu, V. D.; Alföldi, J.; Harris, R. S.; Lindblad-Toh, K.; Haussler, D.; Karlsson, E.; Jarvis, E. D.; Zhang, G.; Paten, B. *Nature* **2020,** *587,* 246–251. doi:10.1038/s41586-020-2871-y
87. Morozova, O.; Marra, M. A. *Genomics* **2008,** *92,* 255–264. doi:10.1016/j.ygeno.2008.07.001
88. Ziemert, N.; Alanjary, M.; Weber, T. *Nat. Prod. Rep.* **2016,** *33,* 988–1005. doi:10.1039/c6np00025h

License and Terms