Young investigators in natural products chemistry, biosynthesis, and enzymology

Edited by Jeffrey D. Rudolf, Lena Barra and Takayoshi Awakawa
Unraveling the role of prenyl side-chain interactions in stabilizing the secondary carbocation in the biosynthesis of variexenol B

Moe Nakano¹, Rintaro Gemma¹ and Hajime Sato*¹,²

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

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

The terpene cyclization cascade generally involves a multistep domino-type reaction. Therefore, it is challenging to reveal the detailed reaction mechanism solely by an experimental method.
To address this issue, computational chemistry including DFT [5-9], QM/MM [10-16] and QM/MM MD [14-17] calculations have been used for the biosynthetic studies of terpene/terpenoids [18].

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

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

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

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

In this study, we investigated the biosynthetic pathways using DFT calculations to validate the above-mentioned aspects.

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

Our research started with the application of DFT calculations to the putative biosynthetic pathway of variexenol B (Figure 1). It was revealed that the variexenol B biosynthetic pathway undergoes a two-step reaction process. Contrary to the putative biosynthetic pathway, the formation of the C1–C11 and C2–C10 bonds was found to be concerted, due to the formation of a secondary carboxylation at the C10 position. Then, the tertiary carboxylation formed at the C3 position undergoes virtu-
ally barrier-free cyclization from the exomethylene group to yield IM3.

We next investigated the effect of the prenyl side chain in the biosynthesis of variexenol B. Although several terpene compounds with prenyl side chains have been reported, it remains unclear whether these prenyl side chains are located inside or outside the active site during the cyclization process. Therefore, we searched for conformations in which the side chain is closer to the carbocation center and performed calculations.
It was found that the structure with the prenyl side chain containing the C14=C15 double bond positioned inwards was more advantageous than the pathway shown in Figure 1. Calculations based on the specified structure are shown in Figure 2. In this pathway, the C14=C15 double bond interacts with the secondary carbocation at C10, reducing the activation energy of the

\[ \text{PPO iso-GGPP I} \rightarrow IM1a/IM1b \rightarrow TS_{1a-2a}/TS_{1b-2b} \rightarrow path a \rightarrow IM2a \rightarrow TS_{2a-3a}/TS_{2b-3b} \rightarrow IM2b \rightarrow IM4a/IM4b \rightarrow TS_{3a-4a}/TS_{3b-4b} \rightarrow IM3a/IM3b \]

\[ + H_2O \]

\[ \text{variexenol B} \rightarrow IM1a/IM1b \rightarrow TS_{1a-2a}/TS_{1b-2b} \rightarrow TS_{2a-3a}/TS_{2b-3b} \rightarrow TS_{3a-4a}/TS_{3b-4b} \rightarrow IM4a \]

\[ \Delta G \text{ (kcal/mol)} \]

\[ \text{IM1a (0.0)} \rightarrow \text{IM1b (-1.6)} \rightarrow \text{IM2a (0.6)} \rightarrow \text{IM2b (-6.2)} \rightarrow \text{IM3a (0.9)} \rightarrow \text{IM3b (2.3)} \rightarrow \text{IM4a (-14.7)} \]

\[ \text{1.2-alkyl shift} \rightarrow 1\text{st annulation} \rightarrow 2\text{nd annulation} \rightarrow 3\text{rd annulation} \]

\[ \text{double ring formation (5/11)} \rightarrow \text{C–C bond cleavage} \rightarrow \text{double ring formation (4/9)} \rightarrow \text{ring skeleton completion (4/7/6)} \]

**Figure 2:** (A) Results of the DFT evaluation of the whole pathway of variexenol B including cation–π interaction from the prenyl side chain. Path a has an α-hydrogen at the C14 position in IM2, while path b has the opposite orientation. (B) Energy diagram of variexenol B with consideration of cation–π interaction. Potential energies (kcal/mol, Gibbs free energies calculated at the mPW1PW91/6-31+G(d,p)//M06-2X/6-31+G(d,p) level) relative to IM1 are shown in parentheses.
first step by approximately 4.7 kcal/mol. Moreover, due to the stabilization of the secondary carbocation-like intermediate IM2, the reaction proceeds stepwise rather than concertededly [7]. It was found that the final cyclization reaction from the exomethylene group proceeds without a barrier, similar to the previous pathway.

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

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

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

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

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

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

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

The distance between C10 and C14 is 1.71 Å, which is hardly to recognize as a single C–C bond, since the distance is greatly elongated. Moreover, the bond length of C14=C15 is 1.43 Å, which is close to the double bond length. Judging from the bond length alone, it is not impossible to conclude that the C10–C14 bond is formed, but considering the rational mechanism of organic reactions, bond cleavage does not occur immediately after the bond is formed.

On the other hand, the C10–C14 bond length of IM2b is 1.64 Å, which is the bond length when hyperconjugated and is commonly observed in terpene-forming reactions. This relatively short bond length appears to contribute to the stability of IM2b. The energy difference between IM2a and IM2b appears to be due to small conformational differences caused by the stereochemistry of H14.
In TS\textsubscript{2a–3a}, the C10 secondary carbocation is stabilized and sandwiched between the two π orbitals of C2=C3 and C14=C15. The status of this orbital interaction is depicted in Figure 3B. This interaction forms the C2–C10 bond and the reaction proceeds to IM3a.

Regarding the 4-membered ring formation, the C2–C10 bond in IM3a is 1.73 Å long, which is hard to recognize as a single bond. However, it is well consistent with the previously reported hyperconjugation in 4-membered ring formation [7]. Then, the C2–C10 bond became 1.56 Å and the 4-membered ring bond is completed (Figure 3B) when the hyperconjugation effect is eliminated by the removal of the C3 carbocation through annulation from the exomethylene group. Based on the key bond analysis, we have successfully elucidated the details of variegenol B biosynthesis. Note that there is an interaction between the empty p orbital of C10 and the π orbital of C14=C15 in Figure 3B, although the bond is not shown. The presence or absence of a bond in the GaussView depends only on the distance between the atoms and may differ from the actual bonding. If a bond is stretched due to hyperconjugation etc., it often happens that the bond is not displayed correctly. Therefore, we performed bond length analysis and NBO analysis to understand the state of the carbocation and bonding.
Conclusion
In conclusion, we have investigated the detailed reaction mechanism of the biosynthesis of variexenol. We have revealed three new insights: (i) the possibility of stabilization of the secondary carbocation by the prenyl side chain of the intermediate, (ii) the four-membered ring formation is completed by the bridging reaction, and (iii) the annulation from the exomethylene group is a barrier-free process.

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

Experimental
All calculations were carried out using the Gaussian 16 package [42]. Structure optimizations were done with the M06-2X [43] density functional theory method and the 6-31+G(d,p) basis set without any symmetry restrictions. M06-2X was selected because of its accuracy in calculating terpene-forming reactions without any symmetry restrictions. M06-2X was selected because of its accuracy in calculating terpene-forming reactions [40-41]. Intrinsic reaction coordinate (IRC) calculations [45-47] for all TSs were performed with GRRM11 [52] based on Gaussian 16. Single-point energies were calculated at the mPW1PW91/6-31+G(d,p) level based on the optimized structure by using the M06-2X method. The utility of relative Gibbs free energies (Grel) based on single-point energy at the mPW1PW91 level has been previously validated for a wide variety of terpene-forming reactions [22,41,53].

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

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

ORCID® iDs
Moe Nakano - https://orcid.org/0009-0002-0245-0075
Rintaro Gemma - https://orcid.org/0009-0006-9682-6607
Hajime Sato - https://orcid.org/0000-0001-5185-096X

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

References

Supporting Information
Supporting Information File 1
IRC plot, 3D representations of all computed structures, cartesian coordinates, energies, and imaginary frequencies. [https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-19-107-S1.pdf]
Identification of the \( p \)-coumaric acid biosynthetic gene cluster in \textit{Kutzneria albida}: insights into the diazotization-dependent deamination pathway

Seiji Kawai\(^1\), Akito Yamada\(^1\), Yohei Katsuyama\(^*1,2\) and Yasuo Ohnishi\(^1,2\)

Full Research Paper

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

Email:
Yohei Katsuyama\(^*\) - aykatsuhko@g.ecc.u-tokyo.ac.jp

\(^*\) Corresponding author

Keywords:
actinomycetes; avenalumic acid; biosynthesis; \( p \)-coumaric acid; polyketides

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

Introduction
The genomes of microorganisms possess diverse biosynthetic gene clusters (BGCs) to produce natural products [1]. In particular, actinomycetes, a major source of bioactive compounds, have numerous BGCs in their genomes, but the majority of them are thought to be silent under laboratory conditions [2].
Our research group previously identified the secondary metabolite-specific nitrous acid biosynthetic pathway, named ANS (aspartate-nitrosuccinate), from the study on cremeomycin biosynthesis [4,5]. The ANS pathway is composed of two enzymes, CreE (FAD-dependent monooxygenase) and CreD (lyase), to synthesize nitrous acid from l-aspartate and the nitrous acid is used to synthesize the diazo group of cremeomycin [4]. After the discovery of the ANS pathway, it has been shown that the ANS pathway is involved in the nitrogen–nitrogen (N–N) bond formation in the biosynthesis of several natural products [6-8]. Enzymes that catalyze N–N bond formation by using nitrous acid from the ANS pathway have also been characterized in several recent studies [9-14]. Most of them belong to the adenylation-forming enzyme superfamily (ANL superfamily) and utilize ATP to activate nitrous acid by AMPylation, with the only exception being AzpL in alazopeptin biosynthesis, which is a membrane protein that catalyzes diazotization, presumably in an ATP-independent manner [9-14]. Moreover, the genome database analysis indicated that there are many orphan BGCs containing genes encoding the ANS pathway, which implies that the biosynthesis of many unknown natural products requires nitrous acid derived from the ANS pathway [4].

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

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

Results
Bioinformatic analysis of ava and cma clusters
In our previous study, we found more than 100 BGCs that have a similar gene component to that of the ava cluster in the genome database [13]. In this study, to obtain and analyze ava cluster-related BGCs, we again searched the genome database for gene clusters that contain (i) genes responsible for 3,4-AHBA synthesis (AvaH and AvaI homologs), (ii) genes encoding a KS-CLF complex (AvaA4 and AvaA5 homologs), and (iii) a gene encoding a diazotase (AvaA6 homolog). As a result, we discovered 134 BGCs and analyzed them using anti-SMASH [17] and BiG-SCAPE (Figures S1 and S2 in Supporting Information File 1) [18]. Interestingly, approximately half of them (72 BGCs) do not have avaE and avaD homologs for the ANS pathway, suggesting three possibilities: (i) avaE and avaD homologs are encoded apart from the cluster, (ii) the strains have an alternative route to synthesize nitrous acid (as reported for alanosine biosynthesis [19]), and (iii) the strains use exogenous nitrous acid [13]. The BiG-SCAPE analysis divided the BGCs into seven groups (Figures S1 and S2 in Supporting Information File 1) and the ava cluster belongs to FAM_00127 (Figure S1). Interestingly, most of the BGCs in FAM_00091, FAM_00111, FAM_00133, and FAM_00125 (approximately 40 BGCs) do not have avaA8 (FabG-like ketoreductase) and avaC (major facilitator superfamily transporter) homologs, but have a cmaG-like gene encoding an FNM-dependent oxidoreductase and a cmaR-like gene encoding a LysR family transcriptional regulator (Figure 1, Figures S1 and S2 in Supporting Information File 1, and Table 1). We anticipated that they would be a subtype of ava-like BGCs and might produce compounds other than avenalumic acid. Among these clusters, we focused on a cluster present in the genome of a rare actinomycete, Kutzneria albida, and named it as cma cluster because the cma cluster was later shown to be responsible for the biosynthesis of p-coumaric acid (Figure 1 and Table 1, NCBI accession number CP007155.1).
Figure 1: Comparison of ava and cma clusters and the biosynthetic pathway of p-coumaric acid. A) Schematic representation of ava and cma clusters from Streptomyces sp. RI-77 and Kutzneria albida, respectively. They harbor similar gene components except for the genes encoding oxidoreductases (avaA8 and cmaG), transporter (avaC), and regulator (cmaR). B) The proposed biosynthetic pathway of p-coumaric acid by the cma cluster.

Table 1: Comparison of cma and ava gene clusters.$^a$

<table>
<thead>
<tr>
<th>Genes in cma cluster</th>
<th>Genes in ava cluster</th>
<th>Amino acid sequence identity (%)</th>
<th>Predicted function of encoded protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmaR</td>
<td>–</td>
<td></td>
<td>LysR family transcriptional regulator</td>
</tr>
<tr>
<td>cmaF</td>
<td>avaF</td>
<td>45.9</td>
<td>major facilitator superfamily transporter</td>
</tr>
<tr>
<td>cmaG</td>
<td>–</td>
<td></td>
<td>FMN-dependent oxidoreductase</td>
</tr>
<tr>
<td>–</td>
<td>avaC</td>
<td></td>
<td>major facilitator superfamily transporter</td>
</tr>
<tr>
<td>cmaI</td>
<td>avaI</td>
<td>65.4</td>
<td>3,4-AHBA synthase</td>
</tr>
<tr>
<td>cmaH</td>
<td>avaH</td>
<td>66.2</td>
<td>3,4-AHBA synthase</td>
</tr>
<tr>
<td>cmaA1</td>
<td>avaA1</td>
<td>61.4</td>
<td>acyl-ACP ligase</td>
</tr>
<tr>
<td>cmaA2</td>
<td>avaA2</td>
<td>50.5</td>
<td>acyl carrier protein</td>
</tr>
<tr>
<td>cmaA3</td>
<td>avaA3</td>
<td>51.9</td>
<td>acyl carrier protein</td>
</tr>
</tbody>
</table>
Table 1: Comparison of cma and ava gene clusters.

<table>
<thead>
<tr>
<th>cmaG</th>
<th>avaG</th>
<th>Retention Time (min)</th>
<th>Mass and UV spectra</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmaA4</td>
<td>avaA4</td>
<td>2.0</td>
<td>N-acetyl-3-aminoavenalumic acid</td>
</tr>
<tr>
<td>cmaA5</td>
<td>avaA5</td>
<td>3.2</td>
<td>3,4-AHBA derivatives</td>
</tr>
<tr>
<td>cmaA6</td>
<td>avaA6</td>
<td>4.0</td>
<td>N-acetyl-3,4-AHBA (9)</td>
</tr>
<tr>
<td>cmaB</td>
<td>avaB</td>
<td>4.5</td>
<td>3,4-AHBA derivatives</td>
</tr>
<tr>
<td>cmaA7</td>
<td>avaA7</td>
<td>5.0</td>
<td>3,4-AHBA derivatives</td>
</tr>
<tr>
<td>cmaE</td>
<td>avaE</td>
<td>5.5</td>
<td>3,4-AHBA derivatives</td>
</tr>
<tr>
<td>cmaD</td>
<td>avaD</td>
<td>6.0</td>
<td>3,4-AHBA derivatives</td>
</tr>
<tr>
<td>–</td>
<td>avaA8</td>
<td>7.0</td>
<td>3,4-AHBA derivatives</td>
</tr>
</tbody>
</table>

a:“–” denotes no counterpart.

Heterologous expression of cma cluster

To examine whether the cma cluster is responsible for the biosynthesis of avenaluminic acid or another natural product, we performed heterologous expression of the cma cluster. Plasmids named pHKO4-cma-D and pTYM3a-cmaG (Figure 2A) were introduced into S. albus by conjugation, resulting in S. albus-cma [20]. S. albus-cma possesses all the cma genes, except for the genes encoding a transporter (cmaF) and a regulator (cmaR), under the control of tipA promoter (Figure 2A).

When we cultured S. albus-cma and analyzed its metabolites by liquid chromatography–mass spectrometry (LC–MS), formation of avenaluminic acid was not observed. Instead, production of p-coumaric acid (5) was detected (Figure 2B and Figure S3A,D in Supporting Information File 1). In addition, S. albus-cma produced compound 6, which showed [M + H]^+ ion at m/z = 222, as well as other putative 3,4-AHBA derivatives, among which compound 9 was indicated to be N-acetyl-3,4-AHBA (9) by its mass and UV spectra (Figure 2B and Figure S3B,C,E,F in Supporting Information File 1). The control strain of S. albus, which harbors two empty vectors, did not produce p-coumaric acid and compound 6, suggesting that they are the biosynthetic products of the cma cluster (Figure 2B). The yield of compound 6 decreased when cmaG was removed from S. albus-cma, indicating that CmaG is involved in the production of 6, although it is not essential (Figure 2B). In addition, the ΔcmaG strain produced a higher amount of compound 9 than S. albus-cma. Because compound 9 seems to be a shunt product derived from 3,4-AHBA (1), which is the starter substrate of Cma PKS, this result indicated that CmaG is a component of Cma PKS. From NMR analysis and high-resolution (HR)MS analysis ([M + H]^+ ion at m/z = 222.0766, which corresponds to C_11H_{12}NO_4^+, calcd. 222.0761) of purified 6, the structure of 6 was determined as N-acetyl-3-aminoacenocumaric acid (Figures S9–S13 and Table S1 in Supporting Information File 1). The production of 6 is consistent with our previous work in which N-acetyl-3-aminoacenocumaric acid was produced by S. albus-ava (a recombinant S. albus strain expressing the ava cluster for 3-aminoacenocumaric acid produc-
was tracked to 435 during the AvaA7-catalyzed reaction (Figure S6D in Supporting Information File 1). The absorbance at 435 nm (A8) was measured by CmaA6. Next, we analyzed the function of CmaA6, which was predicted to catalyze the diazotization of 3-aminocoumaric acid (3-ACA, 5) in p-coumaric acid biosynthesis. As expected, 3-diazocoumaric acid (3-DCA, 4) was synthesized by CmaA6 in the presence of ATP and sodium nitrite (Figure 3B). Because we supposed that CmaA6 has a promiscuous substrate specificity, we tested whether CmaA6 could also catalyze the diazotization of 3-aminoenalamic acid (3-AAA, 6). As expected, CmaA6 also accepted 7 as a substrate to synthesize 3-DAA (8) (Figure 3C). Note that the efficiency of diazotization catalyzed by CmaA6 was much higher than that of AvaA6; almost 100% of 3-ACA and 3-AAA were converted to corresponding aromatic diazo compounds 4 and 8, respectively (Figure 3C).

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

<table>
<thead>
<tr>
<th></th>
<th>vs NADPH</th>
<th>vs 3-DAA (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K_m [µM]</td>
<td>138 ± 10</td>
<td>152 ± 31</td>
</tr>
<tr>
<td>k_cat [min⁻¹]</td>
<td>266 ± 2.0</td>
<td>615 ± 55</td>
</tr>
<tr>
<td>k_cat/K_m [min⁻¹ µM⁻¹]</td>
<td>1.9</td>
<td>4.0</td>
</tr>
</tbody>
</table>

**Discussion**

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

Although we have not yet investigated the activity of CmaG (a putative FMN-dependent oxidoreductase), it is expected to catalyze the reduction of the keto group of an β-ketocycl intermediates, similar to the ketoreductase of fatty acid synthase (FAS) and highly reducing type II PKS. This hypothesis is supported by the fact that cmaG exists in the cma cluster instead of avaA8, which encodes a putative FabG-like ketoreductase, and by the observation that the yield of compound 6 decreased considerably when cmaG was removed from the heterologous expression system (Figure 1A and Figure 2B, and Figures S1 and S2).
Figure 3: In vitro analysis of Cma proteins. A) In vitro analysis of CmaA1 and CmaA3. Extracted ion chromatograms at m/z 1087.9, which corresponds to [M + 10H]+ of 3,4-AHBA-CmaA3 (2) under positive ion mode, are shown. B, C) In vitro analysis of CmaA6. UV chromatograms at 310 nm and 310 nm are shown. CmaA6 catalyzed diazotization toward 3-ACA (3) (B) and 3-AAA (7) (C) to synthesize 3-DCA (4) and 3-DAA (8), respectively, in a high conversion ratio.
in Supporting Information File 1). It should be emphasized that the production of 5 and 6 was not completely abolished in S. albus-cma ∆cmaG. This is because FAS’s ketoreductase probably interacts with the Cma system in the absence of CmaG.

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

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

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

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

Finally, this study is important because of the success of genome mining for BGCs in a rare actinomycete, K. albida. Although 47 BGCs were indicated in the K. albida genome by the antiSMASH analysis, only aculeximycin was reported as a natural product produced by K. albida [30]. Although rare
actinomycetes have been expected to be a source of novel natural products, reports of natural product isolation from rare actinomycetes are limited because of the difficulty in cultivation and genetic manipulation. This study demonstrates that heterologous expression of BGCs combined with promoter swapping is a powerful tool for the discovery of new natural products from rare actinomycetes.

Conclusion

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

Experimental

Strains, chemicals, and enzymes

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

Construction of heterologous expression vectors

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

Heterologous expression of the cma cluster in S. albus

pHKO4-cmal-D and pTYM3a-cmaG were introduced into S. albus to construct S. albus-cma in the same way to construct S. albus-ava described in our previous study [13]. S. albus-cma was inoculated into 10 mL of Tryptone Soya Broth (TSB) medium with kanamycin (50 mg/L) and incubated with shaking (300 rpm) at 30 °C for 2 days. One milliliter of this preculture medium was inoculated into 100 mL of Waksman medium (5 g/L Bacto™ peptone, 5 g/L Bacto™ yeast extract, 20 g/L glucose, 5 g/L meat extract, 5 g/L NaCl, and 3 g/L CaCO₃) with kanamycin (50 mg/L), and incubated with shaking (120 rpm) at 30 °C for 2 days. Thioestrepton (15 mg/L) was then added to the culture, and the culture was continued for another 1 day.

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

Isolation and structural determination of compound 6

S. albus-cma was inoculated into 10 mL of TSB medium with kanamycin (50 mg/L) and incubated with shaking (300 rpm) at 30 °C for 2 days. The whole culture was transferred into 1 L of Waksman medium with kanamycin (50 mg/L) and incubated with shaking (120 rpm) at 30 °C for 2 days. Then, thioestrepton (15 mg/L) was added to the culture, and the culture was continued for another 1 day. The pH was adjusted to 4 by adding 6 M HCl to the culture, and the metabolites were extracted by ethyl acetate. After the ethyl acetate was
evaporated, the metabolites were absorbed for 5.0 g silica gel 60 (0.040–0.063 mm, Merck Millipore, Burlington, MA, USA). The silica gel holding the metabolites was applied to a normal-phase medium-pressure liquid chromatography system (MPLC, Shoko Scientific, Kanagawa, Japan) equipped with a silica column (Purif-Pack, Shoko Scientific), and the metabolites were eluted with a linear gradient of chloroform/methanol. Fractions containing compound 6 were concentrated by evaporation. The residual materials were desorbed in 1 mL DMSO and applied to a reversed-phase high-performance liquid chromatography (HPLC, Shimadzu Corp.) equipped with a COSMOCORE Packed column SC18-AR-II (10 mm ID × 250 mm, Nacalai Tesque), and the metabolites were eluted with a linear gradient of water/acetonitrile containing 0.1% formic acid. Fractions containing compound 6 were concentrated by evaporation. Compound 6 was then desorbed in DMSO-d₆, and the structure was determined by the JNM-A600 NMR system (JEOL, Tokyo, Japan) (Table S1 and Figures S9–S13 in Supporting Information File 1).

Production and purification of recombinant proteins
pColdI-cmaA1, pColdI-cmaA3, and pColdI-cmaA6 were constructed by cloning each gene (which was amplified by PCR using pHKO4-cma-D as a template) into the NdeI and XhoI sites of pColdI using In-Fusion (TaKaRa Bio Inc.). The primers used for plasmid construction are listed in Table S4 of Supporting Information File 1. Each protein expression plasmid was introduced into E. coli BL21(DE3). To obtain holo-CmaA3, pACYC-sfp [21] was co-introduced with pColdI-cmaA3 into E. coli BL21(DE3). The obtained strain was cultured in 1 L of Terrific broth (24 g/L tryptone, 12 g/L yeast extract, 12 g/L glucose, 0.8% glycerol, 0.94% K₂HPO₄, and 0.22% KH₂PO₄) with ampicillin and chloramphenicol (if necessary for sfp expression) and incubated with shaking (150 rpm) at 37 °C until the OD₆₀₀ reached 0.8. After the culture was cooled on ice for 30 min, isopropyl β-D-thiogalactopyranoside (IPTG) was added to the culture (for CmaA1 and CmaA3 expression, final concentration was 100 µM; for CmaA6 expression, IPTG was not added) and incubated with shaking (150 rpm) at 15 °C for 20 h. Cells were harvested by centrifugation and suspended in lysis buffer (20 mM HEPES-NaOH, 10% glycerol, and 200 mM NaCl; pH 8.0). The cells were disrupted by sonication on ice, and cell debris was removed by centrifugation. The recombinant protein was purified using His60 Ni Superflow Resin (TaKaRa Bio Inc.). The protein was eluted using a stepwise gradient of imidazole in lysis buffer (20–500 mM imidazole). The buffer was replaced with lysis buffer using an Amicon Ultra centrifugal filter with a suitable molecular mass cutoff (Merck Millipore).

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

In vitro analysis of CmaA1, CmaA3, AvaA1, and AvaA3
A reaction mixture (50 µL) containing 4 µM CmaA1 or AvaA1, 40 µM CmaA3 or AvaA3, 2.5 mM 3,4-AHBA, 1 mM ATP, 5 mM MgCl₂, 20 mM HEPES-NaOH (pH 8.0), 10% glycerol, and 200 mM NaCl was prepared and incubated at 30 °C for 1 h. After centrifugation, the supernatant was analyzed using LC–MS equipped with a BioResolve RP mAb polyphenyl column (2.1 mm ID × 100 mm, Waters, Milford, MA, USA). The compounds were eluted using a linear gradient of water/acetonitrile containing 0.1% formic acid.

In vitro analysis of CmaA6
A reaction mixture (50 µL) containing 5 µM CmaA6, 0.5 mM 3-ACA (3), 1 mM ATP, 5 mM NaNO₂, 5 mM MgCl₂, 20 mM HEPES-NaOH (pH 8.0), 10% glycerol, and 200 mM NaCl was prepared and incubated at 30 °C for 1 h. The reaction was quenched by adding 50 µL of methanol. After centrifugation, the supernatant was analyzed by LC–MS equipped with a COSMOCORE 2.6Pbr Packed column (2.0 mm ID × 100 mm, Nacalai Tesque). The compounds were eluted with a linear gradient of water/acetonitrile containing 0.1% formic acid.

Kinetic analysis of AvaA7
To measure the kinetic parameters of NADPH or NADH, the reaction mixture (90 µL) containing 10 µM CmaA6, 1.0 mM 3-AAA (7), 2 mM ATP, 5 mM NaNO₂, 2.5 mM MgCl₂, 20 mM HEPES-NaOH (pH 8.0), 10% glycerol, and 200 mM NaCl was prepared and incubated at 30 °C for 1 h. Then, 5.0 µL of 4.0 µM AvaA7 (final concentration: 0.2 µM) and 5.0 µM of different concentrations of NADPH or NADH (final concentration: 25–800 µM) were added to the reaction mixture (final reaction volume: 100 µL). The initial velocity of the reaction catalyzed by AvaA7 was estimated by monitoring the decrease in absorbance at 435 nm for 3-DAA using a SpectraMax M2 microplate reader (Molecular Devices, San Jose, CA, USA).
The kinetic parameters ($V_{\text{max}}$, $k_{\text{cat}}$, and $K_m$) were calculated by fitting the substrate concentration [S]–initial velocity (v) plot using the equation $v = V_{\text{max}} \times [S] / (K_m + [S])$.

To measure the kinetic parameters of 3-DAA, the reaction mixture (90 µL) containing 10 µM CmaA6, 1.0 mM 3-AAA (7), 2 mM ATP, 5 mM NaNO$_2$, 2.5 mM MgCl$_2$, 20 mM HEPES-NaOH (pH 8.0), 10% glycerol, and 200 mM NaCl was prepared and incubated at 30 °C for 1 h. Then, the concentration of 3-DAA (8) in the reaction mixture was calculated by measuring the absorbance at 435 nm for 3-DAA (8) using a SpectraMax M2 microplate reader. The initial velocity of the reaction catalyzed by AvaA7 was calculated following the same method as mentioned above, after diluting the reaction mixture to a final concentration of 25 to 400 µM and adding AvaA7 (final concentration: 0.2 µM) and NADPH (final concentration: 800 µM).

Supporting Information

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

Funding
This work was supported by Grants-in-Aid for Scientific Research on Innovative Areas (19H04645 to Y.K., 19H05685 to Y.O.), Grant-in-Aid for Transformative Research Areas (22H05130 to Y.K.), Grant-in-Aid for JSPS Fellows (22K1046 to S.K.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the World-leading Innovative Graduate Study Program in Proactive Environmental Studies, The University of Tokyo (to S.K.), and JSPS A3 Fore- sight Program grant (to Y.O.).

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

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

References

License and Terms

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

The definitive version of this article is the electronic one which can be found at: https://doi.org/10.3762/bjoc.20.1
Discovery of unguisin J, a new cyclic peptide from Aspergillus heteromorphus CBS 117.55, and phylogeny-based bioinformatic analysis of UngA NRPS domains

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

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

Introduction
Unguisins are a small family of fungal cyclic heptapeptides isolated predominantly from Aspergillus sp. [1-8]. Distinctive features of these cyclic peptides include the non-proteinogenic amino acid γ-aminobutyric acid (GABA) and the incorporation of up to five d-amino acids (Figure 1) [1-8]. The amino acids at positions 1 (d-Ala) and 7 (GABA) are conserved but there is considerable variability at positions 2–6, including the incorporation of additional non-proteinogenic amino acids β-methylphenylalanine (βMePhe) and kynurenine (Kyn) [3,4]. So far, no significant biological activities have been reported for these small peptides [3,4,9], however, unguisin A has been shown to bind a series of anions [10].
Recently the biosynthesis of unguisins A and B from *Aspergillus violaceofuscus* CBS 115571 was reported [5]. A seven module non-ribosomal peptide synthetase (NRPS; UngA) was heterologously expressed in *Aspergillus oryzae* NSAR1 yielding both unguisins A and B, which differ by the incorporation of d-Phe and d-Leu at position 3, respectively. The highly conserved d-Ala at position 1 was shown to be synthesized from l-alanine via the PLP-dependent alanine racemase UngC [5]. A hydrolase/peptidase (UngD) was also discovered that linearized the cyclic unguisins to linear peptides during in vitro investigations, although the linear peptides were not detected from the fungal cultures.

NRPS enzymes are large multifunctional enzymes that often synthesize very important bioactive molecules [11,12]. These enzymes consist of several catalytic domains organized into modules. Typically, a module possesses an adenylation (A) domain for selecting and activating amino- or keto acids, a thio-lation (T) domain for shuttling intermediates between catalytic domains, and a condensation (C) domain that catalyzes amide synthesis.
or ester bond formation. Additional common domains include epimerization (E) domains for converting naturally occurring l-amino acids to d-amino acids, methyltransferase (MT) domains that typically methylate specific N atoms, and terminal condensation (CT) domains which cyclize the growing peptide chain and facilitate release from the NRPS. Of the fungal NRPS studied to date, many appear to have some tolerance for the range of amino acids incorporated by the A domains and the C domain has been highlighted as a gatekeeper [13].

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

Results and Discussion

The cultivation of A. heteromorphus CBS 117.55 on rice solid medium yielded an organic-soluble extract, which was subjected to fractionation using preparative HPLC-PDA-ELSD and purification by semipreparative HPLC-PDA; this led to the isolation of a new cyclic peptide 1, along with unguisin B (2, Figure 2). The structure of the new compound 1 was elucidated by 1D and 2D NMR and HRESIMS/MS. Unguisin B was identified by the 1H and 13C NMR data with the reported data [1,5]. Compound 1 was obtained as a white amorphous solid optically active, with [α]D22 +23.4 (c 0.1, MeOH). Its molecular formula was established as C41H56N8O7 by HRMS ([M + H]⁺ at m/z 773.4338, calculated for C41H57N8O7⁺, m/z 773.4345, Δ 0.9 ppm; [M + Na]⁺ at m/z 795.4162, calculated for C41H56N8O7Na⁺, m/z 795.4164, Δ 0.3 ppm) and NMR data analysis, corresponding to eighteen indices of hydrogen deficiency. Its UV spectrum exhibited absorption maxima at λmax 219 and 279 nm.

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

Analysis of the NMR data of 1 allowed identifying characteristic 1H and 13C signals very similar to those of unguisin B (2) [1,5], the difference being the replacement of the Phe-4 in 1 by Val-4 in 2. This assignment was confirmed by observation of HMBC correlations from δH 7.98 (NH) and δH 8.44 (NH) to C=O (δC 173.1) and from δH 2.93 and 3.01 (H2-β) to C=O (δC 171.1) (Figure 3), together with key NOESY interactions between the NH signals at δH 7.98 ↔ 8.44 ↔ 8.14.

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

Figure 2: Chemical structures of unguisin J (1) and unguisin B (2).
Table 1: 1D and 2D NMR data for 1 (\textsuperscript{1}H: 500 MHz, \textsuperscript{13}C: 125 MHz; DMSO-\textit{d}_6).

<table>
<thead>
<tr>
<th>residue/position</th>
<th>δ_C (ppm)</th>
<th>δ_H (mult., J in Hz)</th>
<th>NOESY</th>
<th>HMBC</th>
<th>COSY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>alanine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>–</td>
<td>8.35 (d, 4.2)</td>
<td>2.14, 1.98, 7.42</td>
<td>172.1, 50.2</td>
<td>3.83</td>
</tr>
<tr>
<td>Co</td>
<td>50.2</td>
<td>3.83 (m)</td>
<td>17.0</td>
<td></td>
<td>8.35, 1.13</td>
</tr>
<tr>
<td>Cβ</td>
<td>17.0</td>
<td>1.13 (d, 6.9)</td>
<td>172.6, 50.2</td>
<td>3.83</td>
<td></td>
</tr>
<tr>
<td>C=O</td>
<td>172.1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>valine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>–</td>
<td>7.43 (d 8.9)</td>
<td>8.35, 8.14</td>
<td>172.1</td>
<td>3.98</td>
</tr>
<tr>
<td>Co</td>
<td>58.4</td>
<td>3.98 (t 8.7)</td>
<td>30.0, 18.2, 172.1</td>
<td>7.43, 1.98</td>
<td></td>
</tr>
<tr>
<td>Cβ</td>
<td>30.0</td>
<td>1.98 (m)</td>
<td>19.0, 58.4, 172.6</td>
<td>3.98, 0.68</td>
<td></td>
</tr>
<tr>
<td>Cy</td>
<td>18.2</td>
<td>0.68 (d, 6.7)</td>
<td>58.4, 30.0, 19.0</td>
<td>1.98, 0.78</td>
<td></td>
</tr>
<tr>
<td>Cy</td>
<td>19.0</td>
<td>0.78 (d, 6.7)</td>
<td>58.4, 30.0, 18.2</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>C=O</td>
<td>172.1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>leucine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>–</td>
<td>8.14 (d, 7.6)</td>
<td>7.43</td>
<td>171.1, 51.8</td>
<td>3.88</td>
</tr>
<tr>
<td>Co</td>
<td>51.8</td>
<td>3.88 (m)</td>
<td>39.4, 172.1</td>
<td>8.14, 1.36</td>
<td></td>
</tr>
<tr>
<td>Cβ</td>
<td>39.4</td>
<td>1.36 (m)</td>
<td>23.6, 23.3, 20.8</td>
<td>3.88, 0.98</td>
<td></td>
</tr>
<tr>
<td>Cy</td>
<td>23.6</td>
<td>0.98 (m)</td>
<td>–</td>
<td>1.36, 0.56</td>
<td></td>
</tr>
<tr>
<td>Cβ</td>
<td>20.8</td>
<td>0.56 (d, 6.6)</td>
<td>39.4, 23.6, 23.3</td>
<td>0.98, 0.69</td>
<td></td>
</tr>
<tr>
<td>Cβ</td>
<td>23.3</td>
<td>0.69 (d, 6.6)</td>
<td>39.4, 23.6, 20.8</td>
<td>0.98, 0.56</td>
<td></td>
</tr>
<tr>
<td>C=O</td>
<td>171.1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>phenylalanine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>–</td>
<td>8.44 (d, 4.7)</td>
<td>8.14, 7.98</td>
<td>173.1</td>
<td>4.15</td>
</tr>
<tr>
<td>Co</td>
<td>55.8</td>
<td>4.15 (m)</td>
<td>171.1, 17.5</td>
<td>8.44, 2.93, 3.01</td>
<td></td>
</tr>
<tr>
<td>Cβ</td>
<td>36.1</td>
<td>2.93 (m), 3.01 (m)</td>
<td>173.1, 129.2, 137.2</td>
<td>4.15</td>
<td></td>
</tr>
<tr>
<td>C-1</td>
<td>137.2</td>
<td>–</td>
<td>55.8, 129.2, 137.2</td>
<td>4.15</td>
<td></td>
</tr>
<tr>
<td>C-2, C-6</td>
<td>129.2</td>
<td>7.16 (dd, 7.5, 1.3)</td>
<td>126.4</td>
<td>7.24</td>
<td></td>
</tr>
<tr>
<td>C-3, C-5</td>
<td>128.3</td>
<td>7.24 (d, 7.5)</td>
<td>137.2, 128.3</td>
<td>7.16, 7.19</td>
<td></td>
</tr>
<tr>
<td>C-4</td>
<td>126.4</td>
<td>7.19 (dd, 7.5, 1.3)</td>
<td>129.2</td>
<td>7.24</td>
<td></td>
</tr>
<tr>
<td>C=O</td>
<td>171.1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>alanine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>–</td>
<td>7.98 (d, 5.1)</td>
<td>4.06, 1.12</td>
<td>172.6</td>
<td>4.06</td>
</tr>
<tr>
<td>Co</td>
<td>48.7</td>
<td>4.06 (m)</td>
<td>173.1, 17.5</td>
<td>7.98, 4.06</td>
<td></td>
</tr>
<tr>
<td>Cβ</td>
<td>17.5</td>
<td>1.12 (d, 6.9)</td>
<td>173.1, 48.7</td>
<td>4.06</td>
<td></td>
</tr>
<tr>
<td>C=O</td>
<td>172.6</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>tryptophan</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>–</td>
<td>7.86 (d, 6.9)</td>
<td>7.61</td>
<td>172.6</td>
<td>4.01</td>
</tr>
<tr>
<td>Co</td>
<td>55.3</td>
<td>4.01 (m)</td>
<td>172.6, 171.1, 110.7, 25.0</td>
<td>7.86, 3.22</td>
<td></td>
</tr>
<tr>
<td>Cβ</td>
<td>25.0</td>
<td>3.22 (m)</td>
<td>127.2, 123.7, 110.7, 55.3</td>
<td>4.01</td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>–</td>
<td>10.82 (d, 1.7)</td>
<td>136.4, 127.2, 136.4</td>
<td>7.06</td>
<td></td>
</tr>
<tr>
<td>C-2</td>
<td>123.7</td>
<td>7.06 (br s)</td>
<td>110.7</td>
<td>10.82</td>
<td></td>
</tr>
<tr>
<td>C-3</td>
<td>110.7</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C3a</td>
<td>127.2</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C-4</td>
<td>118.3</td>
<td>7.50 (br d, 8.0)</td>
<td>136.4, 121.1, 127.2w, 110.7w</td>
<td>6.97</td>
<td></td>
</tr>
<tr>
<td>C-5</td>
<td>118.4</td>
<td>6.97 (dt, 8.0, 0.9)</td>
<td>111.5, 127.2</td>
<td>7.50, 6.97</td>
<td></td>
</tr>
<tr>
<td>C-6</td>
<td>121.1</td>
<td>7.07 (dt, 8.0, 1.0)</td>
<td>136.4, 118.4, 118.3</td>
<td>6.97, 7.33</td>
<td></td>
</tr>
<tr>
<td>C-7</td>
<td>111.5</td>
<td>7.33 (dt, 8.0, 0.9)</td>
<td>127.2, 118.4, 118.3</td>
<td>7.07</td>
<td></td>
</tr>
<tr>
<td>C-7a</td>
<td>136.4</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C=O</td>
<td>171.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>GABA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>–</td>
<td>7.61 (dd, 5.4, 4.3)</td>
<td>7.86, 4.01</td>
<td>171.0</td>
<td>3.07, 2.99</td>
</tr>
<tr>
<td>Co</td>
<td>38.8</td>
<td>3.07 (m), 2.99 (m)</td>
<td>25.6</td>
<td>7.61, 1.69, 1.60</td>
<td></td>
</tr>
</tbody>
</table>
Table 1: 1D and 2D NMR data for 1 (1H: 500 MHz, 13C: 125 MHz; DMSO-d6). (continued)

<table>
<thead>
<tr>
<th></th>
<th>CB</th>
<th>Cy</th>
<th>C=O</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ (ppm)</td>
<td>25.6</td>
<td>32.9</td>
<td>172.6</td>
</tr>
<tr>
<td>J (Hz)</td>
<td>1.69 (m), 1.60 (m)</td>
<td>2.14 (m), 1.98 (m)</td>
<td>–</td>
</tr>
<tr>
<td>J (Hz)</td>
<td>38.8w</td>
<td>25.6, 38.8</td>
<td>1.69, 1.60</td>
</tr>
<tr>
<td>w.</td>
<td>weak.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Key gHMBC and gCOSY correlations, and NOESY interactions of 1.

A second peptide was isolated from the same culture of A. heteromorphus CBS 117.55. Compound 2 was obtained as an amorphous white powder, [α]D20 +37 (c 0.1, EtOH) [lit +40 (c 1.0, EtOH)] [5]; for the 1H and 13C NMR spectroscopic data, see Table S2 in Supporting Information File 1. By comparison with literature data this compound was identified as unguisin B (2) [1,5], further corroborating the identification of the new unguisin J (1).

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

The co-isolation of unguisins B and J indicates that module 4 of the NRPS is able to accept two different amino substrates and so may possess subtle differences to UngA from A. violaceofuscus CBS 115571 which has relaxed substrate specificity in module 3. We performed genome mining of the publicly available A. heteromorphus CBS 117.55 (accession number MSFL0000000.1) [15] using fungiSMASH and identified a four gene BGC encoding a seven module NRPS, an alanine racemase, a hydrolase, and a transporter. We named this BGC ung″ to distinguish it from the ung BGC present in A. violaceofuscus and the ung′ BGC in A. campestris IBT 28561 which encodes unguisins H and I [5]. Clinker analysis with the ung BGCs from A. violaceofuscus CBS 115571 and A. campestris IBT 28561 indicated a high level of homology (Figure 4). The biosynthesis of unguisins B and J therefore is proposed to arise from this single BGC, similar to the biosynthesis

Figure 4: Clinker analysis of identified unguisin-encoding BGCs. UngE′ is a methyltransferase that methylates phenylalanine and appears only in the A. campestris BGC.
of unguisins A and B in *A. violaceofuscus* CBS 115571 (Scheme 1).

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

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

The clades formed by the C domains showed higher divergence than the A domains with the C₆ domains forming their own branch and C domain from modules 1 and 3 clearly distinct to those from modules 2, 4, 5, and 6 (Figure 6). This separation of the non-terminal C domains could be due to modules 1 and 3 lacking an E domain. Again, the domains from UngA and UngA’’ were more closely related than those from UngA’ regardless of which two amino acids were condensed.

**Conclusion**

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

![Scheme 1: Proposed biosynthesis of unguisins B and J in A. heteromorphus CBS 117.55.](image-url)
Experimental

General experimental procedures

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

1D and 2D NMR experiments were recorded on a Varian INOVA 500 instrument ($^1$H: 500 MHz; $^{13}$C: 125 MHz). The chemical shifts ($\delta$) were expressed in ppm and recorded with reference to solvent signals ($^1$H NMR: DMSO-$d_6$ 2.50 ppm; $^{13}$C NMR: DMSO-$d_6$ 39.5 ppm). Optical rotation was measured on a Jasco P-2000 polarimeter with a path length of 100 mm. Analytical HPLC-PDA-MS system was a Shimadzu instrument (LC2030C 3D Plus Prominence) coupled to a Shimadzu LCMS-2020 mass spectrometer. Analyses were performed using a Phenomenex Kinetex RP$_{18}$ column (100 mm $\times$ 4.6 mm i.d., 2.6 $\mu$m) along with the Security Guard RP$_{18}$ protective guard column (4.6 mm i.d.) and eluting with $H_2O + 0.1\%$ formic acid and MeCN $+ 0.1\%$ formic acid using a gradient from 90:10 to 10:90 of $H_2O/MeCN$ over 15 min, maintaining in 10:90 $H_2O/MeCN$ for 3 min, from 10:90 to 90:10 in 1 min, and maintaining at 90:10 for 1 min, using a flow rate of 1.0 mL/min. The PDA detector scanned between $\lambda = 190$ and 700 nm. The MS was optimized using the following conditions: interface voltage 4.5 kV; interface temperature 350 °C; DL temperature 250 °C; heat block 200 °C; ESI mode, acquisition range 100 to 1000 Da; nebulizing gas 1.5 L min$^{-1}$; drying gas flow 15 L min$^{-1}$. The fractionation of the sample was performed on a Shimadzu LC-20AP preparative liquid chromatograph (SCL-40 System Controller Deliver and LH-40 Liquid Handler) coupled to a Shimadzu SPD-M40 Photo Diode Array Detector (PDA) system using a RP-18 column (Phenomenex, Kinetex 250 $\times$ 30 mm i.d., 5 $\mu$m, flow rate of 18.0 mL min$^{-1}$). The purification of compounds was performed on a Shimadzu LC-20AD liquid chromatography (CBM-20A Communication...
Bus Module, CTO-20A column oven, DGU-20A Degassing Unit and SIL-20A AutoSampler) coupled to a Shimadzu SPD-20A UV–vis Detector system using a RP-18 column (Shimadzu, Premier 250 × 10 mm i.d., 5 µm, flow rate of 3.0 mL min⁻¹). High-resolution mass spectra were recorded on an ABSciex TripleTOF 6600+ mass spectrometer. Direct infusion of compounds 1 and 2 through the high-resolution mass spectrometry (HRMS) was performed using a flow rate of 10 μL min⁻¹ which the samples were diluted at 10 ppm with a solution of MeCN/H₂O (50:50; v/v) containing 0.1% formic acid. The parameters such as declustering and entrance potentials remained constant for MS and MS/MS were set up at 150 V and 10 V, respectively. Collision energy for MS and MS² scan surveys was 10 V and 45 V, respectively, with a collision energy spread of 12 V for MS² scan survey. Precursor ion was impacted with three different collision energies (33, 45, 57 V), and the resulting MS² spectra were combined into one final MS² spectrum. The mass spectra were acquired using Turbo Spray Ionization set to 5.5 kV in positive ion mode with an accumulation time of 100 ms. The mass ranges for MS and MS² scan surveys were 500–800 amu and 30–800 amu, respectively. The curtain gas (nitrogen), nebulizing and heating gas were fixed at 25 psi, 20 psi and 15 psi, respectively. The temperature of the source was 25 °C. MS spectra were acquired and processed using Analyst TF 1.8.1 software.

Fungal growth and extraction
*Anareiomorphus* CBS 117.55 was cultivated in 2 Erlenmeyer flasks (500 mL), each containing 90 g of rice and 150 mL of H₂O [16]. The medium was autoclaved at 121 °C for 20 min. After sterilization, the medium was inoculated with the spore solution of *A. heteromorphus* (1 mL) and incubated in static mode at 25 °C for 21 days. The following day, the cultured mass in the flasks was ground and extracted with ethyl acetate (EtOAc, 3 × 100 mL). The EtOAc fraction was dried using a rotary evaporator and then dissolved in CH₃CN for defatting with hexane by partitioning. The CH₃CN fraction was evaporated, yielding 0.601 g of soluble-organic extract.

Fractionation and isolation of unguisins J and B
The soluble-organic extract was fractionated by preparative HPLC-PDA using Kinetex RP18 column (250 mm × 30 mm i.d., 5 µm) and UV detector at λ_max = 254 nm. The mobile
phase consisted of H₂O + 0.05% formic acid (eluent A) and MeCN + 0.05% formic acid (eluent B), which was eluted of 20–100% of B with flow rate of 18 mL min⁻¹, yielding 20 fractions.

Fractions Fr13 and Fr15 were subjected to a purification by semipreparative HPLC-UV using a Premier RP18 column (250 mm × 10 mm i.d., 5 μm) and UV detector at λmax = 210 nm. The gradient elution consisted from 65:35 to 35:65 of H₂O/MeCN over 20 min, using a flow rate of 3.0 mL min⁻¹. Fractions Fr13 and Fr15 resulted in the isolation of 1 (11.5 mg) and 2 (14.1 mg), respectively.

**Unguisin J (1).** Amorphous white powder. [α]D²² +23.4 (c 0.1, MeOH); UV (photodiode array, MeCN/H₂O) λmax = 219 and 279 nm; HRESIMS m/z: [M + H]⁺ calcd for C₄₁H₅₁N₇O₁₁+, 773.4345; found, 773.4338. [M + Na]⁺ calcd for C₄₁H₅₁N₇O₁₁Na⁺, 795.4164; found, 795.4162. For the ¹H and ¹³C NMR spectroscopic data, see Table 1.

**Bioinformatics**

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

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

Samples 1 and 2 (0.3 mg each) were hydrolyzed in 0.5 mL of 6 N HCl at 115 °C for 20 h. After cooling, the reaction mixture was evaporated under nitrogen gas flow. The residue was dissolved in 0.5 mL of H₂O, and dried using speedvac to remove the residual HCl. The hydrolysates of 1 and 2 as well as authentic amino acid samples were treated with 200 μL of 1 N sodium bicarbonate solution and 100 μL of 1% Marfey’s reagent (N⁶-(5-fluoro-2,4-dinitrophenyl)-o-leucinamide) in acetone [14]. The samples were incubated for 1 h at 60 °C followed by neutralization with 100 μL of 2 N HCl. The HCl was removed by using N₂ gas. Then, the samples were diluted by 2 mL of MeCN/H₂O (1:1, v/v) solution, filtered with 0.22 μm filter and analyzed by LC–MS for comparison of the retention times.

**Supporting Information**

**Supporting Information File 1**

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

**Acknowledgements**

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

**Funding**

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

**Author Contributions**

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

**ORCID® iDs**

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

**Data Availability Statement**

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

**Preprint**

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

**References**


**License and Terms**

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

The definitive version of this article is the electronic one which can be found at: https://doi.org/10.3762/bjoc.20.32
Abstract
The adenylation (A) domain is essential for non-ribosomal peptide synthetases (NRPSs), which synthesize various peptide-based natural products, including virulence factors, such as siderophores and genotoxins. Hence, the inhibition of A-domains could attenuate the virulence of pathogens. 5′-O-N-(Aminoacyl or arylacyl)sulfamoyladenosine (AA-AMS) is a bisubstrate small-molecule inhibitor of the A-domains of NRPSs. However, the bacterial cell permeability of AA-AMS is typically a problem owing to its high hydrophilicity. In this study, we investigated the influence of a modification of 2′-OH in the AMS scaffold with different functional groups on binding to target enzymes and bacterial cell penetration. The inhibitor 7 with a cyanomethyl group at 2′-OH showed desirable inhibitory activity against both recombinant and intracellular gramicidin S synthetase A (GrsA) in the gramicidin S-producer Aneurinibacillus migulanus ATCC 9999, providing an alternative scaffold to develop novel A-domain inhibitors.

Introduction
Nonribosomal peptides (NRPs) exhibit various biological activities and have been used as therapeutic agents, such as antibiotics, anticancer agents, and immunosuppressants [1]. Additionally, NRPs function as virulence factors, such as siderophores and genotoxins [2]. Therefore, inhibiting their biosynthesis by using small molecules can help to elucidate their natural functions and their potential as therapeutic targets. NRPs are synthesized by large, versatile, and multifunctional proteins called nonribosomal peptide synthetases (NRPSs), which are composed of multiple modules and subdivided domains (Figure 1) [3]. The adenylation (A) domain in NRPSs is responsible for the selection and activation of amino acids, hydroxy acids, and aryl acids upon ATP consumption (Figure 2a) [4]. The activated aminoacyladenosine monophosphate (AMP) is transferred to the thiol...
group of a phosphopantetheine prosthetic arm in an adjacent peptidyl carrier protein (PCP). The amino acid loaded on the PCP then undergoes coupling with the amino acid loaded on the downstream PCP in the condensation (C) domain. Finally, the linear peptide on the PCP in the last module is either hydrolyzed or cyclized by a thioesterase (TE) domain, consequently resulting in the formation of the final products.

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

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

Aryl acid A-domains are involved in the synthesis of several bacterial siderophores such as vibriobactin from Vibrio cholera, yersiniabactin from Yersinia pestis, and mycobactin from Mycobacterium tuberculosis [6]. 5′-O-Sulfamoyladenosine
(AMS), a bioisosteric analog of an AMP intermediate, has been used as a non-hydrolysable scaffold for developing A-domain inhibitors. Moreover, 5'-O-[N-(salicyl)sulfamoyladenosine (Sal-AMS) and its derivatives show potent inhibitory activities against the A-domain of MbtA, a component of mycobactin synthetase and antimicrobial activities against \textit{M. tuberculosis} [7]. In addition, aminoacyl (AA)-AMS has been designed to inhibit the amino acid-activating A-domains in NRPSs and has been found to be a tight-binding inhibitor (Figure 2b) [8]. Moreover, the intracellular concentrations of a series of AMS derivatives in \textit{Escherichia coli}, \textit{Bacillus subtilis}, and \textit{Mycobacterium smegmatis} have been investigated by Tan et al., demonstrating non-obvious correlations between the chemical structure and permeability among various bacteria, owing to the differences in the composition of cell membranes and presence of efflux pumps [9].

Specific protein labeling using a chemical probe can help to identify, characterize, and visualize target proteins [10,11]. The first chemical probe used for A-domains in NRPSs was reported by Marahiel et al. [8]. They introduced a pegylated biotin linker at the 2'-OH group of l-Phe-AMS and confirmed that the probe retains the binding activities toward the A-domain of GrsA, a gramicidin S synthetase. Aldrich et al. developed a Sal-AMS-based activity-based probe (ABP) to profile MbtA in \textit{M. tuberculosis} [12]. In contrast, we previously described an activity-based protein profiling (ABPP) strategy for NRPSs using ABPs that target A-domains (Figure 2b) [13-15]. The probes comprise an aminoacyl-AMS ligand and a photoaffinity group with clickable alkyne functionality appended to the 2'-OH group of adenosine. A complex structure of the GrsA A-domain with l-Phe and AMP revealed that the 2'-OH of the adenosine skeleton is oriented toward the outside of the active site of the GrsA A-domain, suggesting that chemical modification at the 2'-OH group of the adenosine skeleton would be tolerated [16] (Figure 2b). Moreover, these probes (AA-AMS-BPyne) can selectively label the A-domains corresponding to the amino acid of the ligand in both recombinant enzymes and proteomes. We
recently reported that these probes can be used to label the A-domains of endogenous NRPSs in live bacterial cells [17-19]. The intracellular labeling of the enterobactin synthetase EntF with Sal-AMS-BPyne requires carbonyl cyanide m-chlorophenylhydrazone, which collapses the proton motive force used in most efflux pumps [17]. Under the same conditions, the competitive inhibition of labeling using excess Sal-AMS is not observed, suggesting that the modification at the 2′-OH group of the adenosine in AA-AMS might affect the cell permeability of the compounds. Given the high hydrophilicity of the bisubstrate AMS scaffold, it is reasonable to conclude that the BPyne component increases hydrophobicity and facilitates accumulation in live E. coli. In the present study, we investigated the influence of the introduction of several functional groups at the 2′-OH group of the AMS scaffold on both the binding affinities for A-domains and cell permeability (Figure 2c and Figure 3). We selected the L-Phe-selective A-domain of the gramicidin S synthetase GrsA, which was previously demonstrated to be selectively labeled with L-Phe-AMS-BPyne (3) in Aneurinibacillus migulanus ATCC 9999 [19]. Intracellular competitive ABPP of GrsA using L-Phe-AMS-BPyne was also performed to reveal the cell permeability of 2′-OH-modified L-Phe-AMS derivatives.

Results and Discussion
To investigate the influence of the introduction of different alkyl groups at the 2′-OH, we prepared L-Phe-AMS derivatives 4-9. As the compounds 6, 8, and 9 were synthesized previously [20], we designed and synthesized three new L-Phe-AMS derivatives containing methyl (4), benzyl (5), and cyanoethyl (7) groups at the 2′-OH. The synthetic routes to compounds 4-9 are shown in Scheme 1. The 2′-OH of adenosine was alkylated with several alkyl halides in the presence of sodium hydride (NaH). Both the 3′-OH and 5′-OH groups of compounds 10a-e were protected by a TBS group, followed by the selective deprotection of the 5′-OH group using 25% trifluoroacetic acid in tetrahydrofuran. Subsequently, the 5′-OH group of compounds 12a-e were reacted with sulfamoyl chloride in the presence of NaH. Compounds 13a-e were coupled to pre-activated Boc-L-Phe-OSu in the presence of Cs2CO3. Removal of the Boc and TBS groups of compounds 14a-e yielded the desired L-Phe-AMS derivatives 4, 5, and 7.

We first determined the binding affinities of L-Phe-AMS derivatives 4-9 for the A-domain of GrsA using a previously developed competitive enzyme-linked immunosorbent assay technique for A-domains in NRPSs [14], which allowed us to measure the dissociation constant (Kd) of the test compounds (Figure S1a, Supporting Information File 1). The L-Phe-AMS-biotin probe was immobilized on a streptavidin-coated 96-well plate and incubated with recombinant His6-tagged GrsA in the presence or absence of inhibitors (Figure S1b, Supporting Information File 1). After washing, the wells were incubated with an anti-His6 tag antibody and subsequently with a horseradish peroxidase-conjugated secondary antibody. The amount of GrsA bound to L-Phe-AMS-biotin was determined by measuring the absorbance of o-phenylenediamine dihydrochloride at 492 nm. The Kd values of the L-Phe-AMS derivatives are listed in Table 1. Compared with the binding affinity of L-Phe-AMS 1 (Kd value, 11.4 ± 3.4 nM) for the A-domain of GrsA, all tested compounds showed slightly decreased binding affinities. Among them, inhibitors 4 (23.9 ± 0.7 nM), 7 (16.6 ± 0.6 nM), and 8 (30.2 ± 4.2 nM) displayed Kd values comparable to that of L-Phe-AMS 1, suggesting that the substitution of small functional groups such as methyl and cyanomethyl groups has a minor effect on the Kd values (Table 1 and Figures S2 and S3, Supporting Information File 1). In contrast, inhibitors 6 and 9 exhibited lower affinity than inhibitors 1, 2, 4, 5, and 7. This result suggests that the incorporation of the pegylated functionality at the 2′-OH group of the adenosine skeleton particularly perturbs binding via steric hindrance.

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

![Image](71x141 to 524x234)

**Figure 3:** Illustration of the inhibition of A-domains by aminoacyl-AMS derivatives in live bacterial cells. Cell-permeable L-Phe-AMS derivatives can penetrate cells and interact with A-domains in live bacterial cells, resulting in the competitive inhibition of the labeling by L-Phe-AMS-BPyne. The substituent group (R; gray) of L-Phe-AMS derivatives could facilitate their cell penetration. After UV irradiation (365 nm), the labeled proteins are treated with a TAMRA-N3 under copper(I)-catalyzed azide–alkyne cycloaddition conditions, followed by SDS-PAGE coupled with in-gel fluorescence scanning. AMS, 5′-O-sulfamoyladenosine.
the ABP L-Phe-AMS-BPyne (3). GrsA (1 µM) was incubated with probe 3 (1 µM) in the absence or presence of either inhibitor 1, 2, or 4–9 (100 µM). The mixtures were exposed to ultraviolet light at 365 nm to form a covalent bond between GrsA and the benzophenone moiety in probe 3. The samples were then reacted with TAMRA-N₃ (structure shown in Figure S4, Supporting Information File 1) under copper(I)-catalyzed azide–alkyne cycloaddition conditions [21] and subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis coupled with in-gel fluorescence scanning (excitation wavelength, 532 nm; emission wavelength, 580 nm). The labeling of GrsA by probe 3 was completely suppressed by the addition of inhibitors 1, 2, 4, 5, and 7 (Figure 4a). In contrast, inhibitors 6, 8, and 9 moderately inhibited labeling, which was consistent with the decreased $K_d$ values of these compounds. Subsequently, we conducted competitive labeling experiments for endogenous GrsA in the proteome of the gramicidin S-producer A. miglanus ATCC 9999. The cellular lysates of strain ATCC 9999 were treated with probe 3 (1 µM) in the absence or presence of inhibitor 1, 2, or 4–9 (100 µM), irradiated at 365 nm, and subjected to the click reaction with TAMRA-N₃. In-gel fluorescence scanning revealed that inhibitors 1, 2, 4, and 7 completely suppressed GrsA labeling by probe 3 in the proteomic environment (Figure 4b). Unlike the results obtained for the labeling of purified recombinant GrsA, inhibitor 5 moderately inhibited labeling at 100 µM.
Finally, we investigated whether the inhibitors could penetrate cells and inhibit intracellular GrsA labeling in *A. migulanus* ATCC 9999. In this study, *A. migulanus* ATCC 9999 cells were grown at 37 °C in YPG medium for 24 h, harvested, and washed with phosphate-buffered saline. The bacterial suspension was then treated with probe 3 (10 µM) in either the absence or presence of inhibitors 1, 2, or 4–9 (10 or 100 µM). Inhibitors 1, 2, 4, 5, 7, and 8 completely inhibited the labeling of endogenous GrsA at high concentrations (100 µM), suggesting that these inhibitors can penetrate cells (Figure 4c). Notably, inhibitors 1 and 7 efficiently inhibited the labeling of GrsA at 10 µM. The incorporation of the nitrile group at the 2′-OH group of the adenosine skeleton is expected to provide chemical properties that would allow the compound to retain its binding affinity and cell permeability. Overall, these results indicate that a 2′-OH modification with a cyanomethyl group represents a useful AMS scaffold for intracellular NRPS inhibition.

### Conclusion

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

Supporting Information

Supporting Information File 1
Additional Figures, experimental part and NMR spectra. [https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-39-S1.pdf]

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

ORCID® iDs
Fumihiro Ishikawa - https://orcid.org/0000-0002-8681-9396
Sho Konno - https://orcid.org/0000-0002-5851-4130
Genzoh Tanabe - https://orcid.org/0000-0002-7954-8874

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

References

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

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