

Supporting Information

for

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

Seiji Kawai, Akito Yamada, Yohei Katsuyama and Yasuo Ohnishi

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

Additional experimental data and NMR spectra

License and Terms: This is a supporting information file under the terms of the Creative Commons Attribution License (https://creativecommons.org/ licenses/by/4.0). Please note that the reuse, redistribution and reproduction in particular requires that the author(s) and source are credited and that individual graphics may be subject to special legal provisions.

Table of contents

Table S1. ¹H and ¹³C NMR data for compound 6.

Table S2. Amino acid identity between ACPs.

Table S3. Primers used for the construction of heterologous expression plasmids.

Table S4. Primers used for the construction of recombinant protein production plasmids.

Figure S1. The BiG-SCAPE analysis of ava-like clusters (Part 1).

Figure S2. The BiG-SCAPE analysis of ava-like clusters (Part 2).

Figure S3. LC-MS analysis of the metabolites of *S. albus-cma* which produces compounds 5 and 6 and other 3,4-AHBA derivatives.

Figure S4. SDS-PAGE analysis of the recombinant proteins used in this study.

Figure S5. In vitro analysis of CmaA1, holo-CmaA3, AvaA1, and holo-AvaA3.

Figure S6. Kinetic analysis of AvaA7, fitted with the Michaelis-Menten equation.

Figure S7. Insights into the mechanism of partner ACP (CmaA3) recognition mechanism by CmaA1.

Figure S8. Phylogenetic analysis of CLFs encoded by the by *ava*-related BGCs in the database.

Figure S9. ¹H NMR spectrum of compound 6.

Figure S10. ¹³C NMR spectrum of compound 6.

Figure S11. ¹H-¹H COSY spectrum of compound 6.

Figure S12. ¹H-¹³C HMQC spectrum of compound 6.

Figure S13. ¹H-¹³C HMBC spectrum of compound 6

Supplementary tables

Table S1. ¹H and ¹³C NMR data for compound **6**. The spectra were recorded in DMSO*d*₆. The solvent peak was used as an internal standard ($\delta_{\rm C}$ 39.51, $\delta_{\rm H}$ 2.50).



	AvaA2	AvaA3	CmaA2	CmaA3
AvaA2	-	29.0%	50.5%	24.4%
AvaA3	29.0%	-	25.3	51.9%
CmaA2	50.5%	25.3%	-	22.9%
CmaA3	24.4%	51.9%	22.9%	-

Table S2. Amino acid identity between ACPs

Table S3. Primers used for the construction of heterologous expression plasmids.

name	sequence	description
<i>cmaI-D</i> F	5'-AAGGGAGCGGA <u>CATATG</u> GCTGACACCGGGAAAGCG-3'	NdeI site is underlined
<i>cmaI-D</i> R	5'-GGTCCTGCCC <u>AAGCTT</u> TCAGGTGCTGAGCCGGTACC-3'	HindIII site is underlined
<i>cmaG</i> F	5'-AAGGGAGCGGA <u>CATATG</u> TCCTCAGATCGGAGAGG-3'	NdeI site is underlined
cmaG R	5'-GCAGGTCGAC <u>TCTAGA</u> TTACCGCGCGGGTGCCAGTGC-3'	XbaI site is underlined

Table S4. Primers used for the construction of recombinant protein production plasmids.

name	sequence	description
cmaAl F	5'-TCGAAGGTAGG <u>CATATG</u> CGGCTGGTGGAGGACCT-3'	NdeI site is underlined
cmaAl R	5'-ATTCGGATCC <u>CTCGAG</u> TCAACTGATCGCCGCGCGCA-3'	XhoI site is underlined
cmaA3 F	5'-TCGAAGGTAGG <u>CATATG</u> ACCACCGAGCAGGTGCGC-3'	NdeI site is underlined
cmaA3 R	5'-ATTCGGATCC <u>CTCGAG</u> TCATCGCACCGCCCCGGAGT-3'	XhoI site is underlined
<i>cmaA6</i> F	5'-TCGAAGGTAGG <u>CATATG</u> ATCACCAAGGAAGAACGC-3'	NdeI site is underlined
cmaA6 R	5'-ATTCGGATCCCTCGAGTCACCCCTGGTCGGCGCGGG-3'	XhoI site is underlined

Supplementary figures

FAM_00108

NZ_L/12211000666.1_003707.region999: Streptomyces phaseochromogenes strain NRPIL B-1248 B-1248_contig456, whole genome shotgun sequence
JUISZ201000002 1,00008 regionmyces unbrinus strain V24F Gald55527 2, strates prome shotpun sequence
JPECKI70000011,00004 region997: Sheptomyces sp. NBC, 00278 NBC, 00
ACOTYO1000004 1 000054 region/992 Streptompces violance strain ASOP, 20 Contig, 008, whole genome shalpun sequence

FAM_00132

PEKUR1000011.1_00003.region099: Streptomyces sp. JV/78 0000011, whole penome shatpun regenerae	
OLANO 100001 r. good 1. good 1	
NZ_LICK01000081.1_000017.region399: Sheptomyces ossamyceticus strain NPR B-0422 B	822 contig_ 50, whole genome shotpun sequence
JARULD1000004.1_00012.region999: Streptomyces cosamyceticus strain DB8 Scatfold-1, whole genome stady in seque	^{***}
SPBA100073.1,00003.ngion988.Steptomyce ipomoae strain 01221 contg. 14. what genome tog	n a populariza 19 da 19
NZ_JAEMUX0100000111_000021region999: Streptomyces advisidensis strain C-MP281 & Annex	a parame shalpan sequence
	më shqipun sequence m::\
FOUV01000011.00027 region899: Streptomyces sp. c124 genome assembly, config. Gul9701184, 100, ander	annan an an anna an anna an anna an anna an an
	<mark>;an (-), (</mark>
	72, national processor and space and and a state of the space of the s
	₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩
	<mark>(,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</mark>
	₩₩₩₩₩₩ ੑ <mark>ੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑੑ</mark>
	$= \frac{1}{2} \left(\frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} $
	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
	8. and 20 process and process
	
	<mark>(),(</mark>

FAM_00127



Figure S1. The BiG-SCAPE analysis of *ava*-like clusters (Part 1). The clusters belong to FAM_00127, FAM_00108, and FAM_00132 are shown. *avaI*, *H*, *A1*, *A2*, *A3*, *A4*, *A5*, *A6*, *B*, and *A7* homologs are highlighted in yellow. *avaE* and *avaD* homologs are highlighted in purple. *avaF* and *avaA8* homologs are highlighted in red. The colors of Pfam domains are automatically generated by BiG-SCAPE.

								_	-	-	-	 —\\ī	(m))m han) 🖗 🗇 🗰 dan man da ma da man da da man da da man da man da man da da man da ma
	Г			-								 NZ_BBC	NO TODORER I, COLOR AND AND STATEMENT STATEMENT AND
	ŀг							-	-			 	Add PERFORMANCE 1, 0001001xepion/b95: Streptomyces sp. RKCA744 000000FContig. whole genome shotgun sequence
	Ц	Г	-		-				-		-		Rationances 1, 200008, regionality: Streatonyces ap. SRP1 control this in the second of the second
	l	L	-		-	-	-	-		-	-	 внекот (ІНП)-[0000051_000078.mpion9999: Briptionyces hypotospicus N1141 DAL nospacet, wiele p romet shappa nogance. }}#III.}#µ\
		-				-						 лен. 	A00 0000000000000000000000000000000000
г										-			0100000002.1_0000004.ng/un098 Strateburyes domaini stain DSM 41000 Gub/41097_02. • ••• generate d rugun requesse (JBE/JEE/BE/J/)
Ľ					-			-			-	 NZ_BBC	UM 1000011. good 2. specific Streptonyce in Systems and an INSTE 1000, while an one altogan accession.

FAM_00111





Figure S2. The BiG-SCAPE analysis of *ava*-like clusters (Part 2). The clusters belong to FAM_00091, FAM_00111, FAM_00125, and FAM_00133 are shown. Most of these clusters do not have *avaA8* and *avaC* homologs but have a *cmaG*-like gene encoding an FMN-dependent oxidoreductase and a *cmaR*-like gene encoding a LysR family transcriptional regulator. *avaI*, *H*, *A1*, *A2*, *A3*, *A4*, *A5*, *A6*, *B*, and *A7* homologs are highlighted in yellow. *avaE* and *avaD* homologs are highlighted in purple. *cmaG* and *cmaR* homologs are highlighted in red. The colors of Pfam domains are automatically generated by BiG-SCAPE.



Figure S3. LC-MS analysis of the metabolites of *S. albus-cma*, which produces compounds 5 and 6 and other 3,4-AHBA derivatives. (A) Comparison of compound 5 from the metabolites of *S. albus-cma* and authentic *p*-coumaric acid. UV chromatograms at 310 nm are shown. (B) Mass spectra of compound 6 from the metabolites of *S. albus-cma*. (C) UV spectrum of the compound 6. (D) Comparison of the UV spectra of compound 5 from the metabolites of *S. albus-cma* and authentic *p*-coumaric acid. (E) Mass spectra of compound 5 from the metabolites of *S. albus-cma* and authentic *p*-coumaric acid. (E) Mass spectra of compound 9 from the metabolites of *S. albus-cma*. [M + H]⁺ ion at m/z = 196 corresponds to that of *N*-acetyl-3,4-AHBA. (F) UV spectrum of compound 9.



Figure S4. SDS-PAGE analysis of the recombinant proteins used in this study. All recombinant proteins were produced by *E. coli* BL21(DE3). The theoretical molecular mass values (kDa) of recombinant proteins are CmaA1, 50.6; CmaA3, 10.5; and CmaA6, 62.0.



Figure S5. *In vitro* analysis of CmaA1, *holo*-CmaA3, AvaA1, and *holo*-AvaA3. (**A**) 3,4-AHBA was loaded onto *holo*-CmaA3 by AvaA1. Extracted ion chromatograms of m/z = 1087.9, which corresponds to $[M + 10H]^{10+}$ of 3,4-AHBA-CmaA3 under positive ion mode, are shown. (**B**) 3,4-AHBA was loaded onto *holo*-AvaA3 by CmaA1. Extracted ion chromatograms of m/z = 1287.0, which corresponds to $[M + 10H]^{10+}$ of 3,4-AHBA-AvaA3 under positive ion mode, are shown. (**C-F**) The mass spectra shift of *holo*-CmaA3 (**C**), 3,4-AHBA-CmaA3 synthesized by CmaA1 (**D**), 3,4-AHBA-CmaA3 synthesized by AvaA1 (**E**), and 3,4-AHBA-AvaA3 synthesized by CmaA1 (**F**).



Figure S6. Kinetic analysis of AvaA7, fitted with the Michaelis-Menten equation. The error bars represent the standard error (n = 3). (A) Kinetic analysis for NADPH. (B) Kinetic analysis for NADH, which could not be fitted by the Michaelis-Menten equation probably because of the high K_m value. (C) Kinetic analysis for 3-DAA (8) when NADPH was used as a cofactor. (D) The schematic representation of the reaction focused on this experiment.



Figure S7. Insights into the mechanism of partner ACP (CmaA3) recognition mechanism by CmaA1. (A) Amino acid alignment of AvaA2, AvaA3, CmaA2, and CmaA3. The Ser residues to which the phosphopantetheinyl arm binds are labeled with a red box. Residues labeled with blue boxes are predicted to be important for the partner ACP recognition by AvaA1 and CmaA1. (B) The structure model of CmaA1-CmaA3 complex predicted by AlphaFold2 [1]. CmaA1, green; CmaA3, cyan. (C) Enlarged view of the interface between CmaA1 and CmaA3. Met34, Trp38, and His41 in CmaA3 seem to be important for the interaction. Trp38 and His41 are conserved between AvaA3 and CmaA3.



Figure S8. Phylogenetic analysis of CLFs encoded by the by *ava*-related BGCs in the database. These enzymes are divided into three large clades. AvaA5 and CmaA5 are marked with arrows. The distance between each enzyme was determined by the global alignment using BLOSUM62.







Figure S10. ¹³C NMR spectrum of compound 6.



Figure S11. ¹H, ¹H COSY spectrum of compound 6.



Figure S12. ¹H, ¹³C HMQC spectrum of compound 6.



Figure S13. ¹H, ¹³C HMBC spectrum of compound 6.

References

J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* 2021, *596*, 583–589.