Argicyclamides A-C unveil enzymatic basis for guanidine bis-prenylation.

Chin-Soon Phan, Kenichi Matsuda, Nandani Balloo, Kei Fujita, Toshiyuki Wakimoto, Tatsufumi Okino

‡Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060-0810, Japan
†Faculty of Environmental Earth Science, Hokkaido University, Sapporo 060-0812, Japan
§Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060-0812, Japan
Graduate School of Environmental Science and Faculty of Environmental Earth Science, Hokkaido University, Sapporo 060-0812, Japan
Global Station for Biosurfaces and Drug Discovery, Hokkaido University, Kita 12, Nishi 6, Sapporo 060-0812, Japan

ABSTRACT: Guanidine prenylation is an outstanding modification in alkaloid and peptide biosynthesis, but its enzymatic basis has remained elusive. We report the isolation of argicyclamides, a new class of cyanobactins with unique mono- and bis-prenylations on guanidine moieties, from Microcystis aeruginosa NIES-88. The genetic basis of argicyclamide biosynthesis was established by the heterologous expression and in vitro characterization of biosynthetic enzymes including AgcF, a new guanidine prenyltransferase. This study provides important insight into the biosynthesis of prenylated guanidines and offers a new toolkit for peptide modification.

Cyanobactins are a class of ribosomally-synthesized and posttranslationally modified peptides (RiPPs) produced by cyanobacteria. Although their chemical structures are hyper-variable in terms of amino acid sequences, cyanobactins are biosynthesized through a common pathway. Cyanobactins are directly encoded in C-terminal region of precursor peptide (IPR031036) as core peptide. Core peptide is flanked by recognition sequences (RSs) that facilitate proper recognition from biosynthetic enzymes modifying core peptide. Universal proteolytic step that remove region preceding core peptide (leader peptide) is mediated by subtilisin-like SBA protease (IPR023830) represented by PatA. Subsequently, second subtilisin-like SBA protease represented by PatG removes region following core peptide (follower peptide) to yield head-to-tail cyclic cyanobactins or linear cyanobactins. In many cases, core peptide undergoes divergent modification steps including heterocyclization to generate oxazol(in)e/thiazol(in)e, methylesterification on C-terminal carboxylic acid, and forward or reverse prenylation and geranylation. The prenylation is catalyzed by ABBA-type prenyltransferases (PTases, IPR031037) that are selective for the isoprenyl donor (dimethylallyl diphosphate: DMAPP or geranyl diphosphate: GPP) and the residue that is prenylated. On the other hand, they exhibit extremely relaxed specificity against the remainder of the prenyl acceptor, and thus cyanobactin PTases could be versatile biochemical tools for late stage peptide modifications. The catalytic variations of cyanobactin PTases revealed that the residues prenylated by these PTases are not only Tyr, Ser, Thr, and Trp, but also N- and C-termini of linear cyanobactins. However, the prenylation sites are limited to these residues, and the prenylation of a charged side chain has not been observed.

During the course of our efforts toward the discovery of natural products, by targeting cyanobacteria from the NIES collection, we found the unreported compounds 1-3 from the well-studied strain M. aeruginosa NIES-88, a producer of micropeptins and kawaguchipeptins (4, 5). The LCMS profile of compounds 1-3 showed molecular ion peaks at m/z 1058, 990 and 922, respectively (Figure S1). This is reminiscent of a homolo-

Figure 1. Structures of cyanobactins from M. aeruginosa NIES-88. Argicyclamides A-C (1-3) are newly reported in this study.
gous series that differs by 68 atomic mass units, and this difference corresponds to a prenyl group. The HRESI(+)-MS analysis of purified 1 yielded a molecular ion [M + H]+ at m/z 1058.7128, indicative of the molecular formula C₅₇H₉₁N₁₁O₈ (∆0.3 mDa). The 1H and 13C NMR spectra of 1 in CD₃OD (Table S1) showed eight signals for the α methines of α-amino acids (δC 66.5, 63.5, 62.4, 61.1, 59.8, 58.4, 57.2, 51.7; δH 4.66, 4.55, 4.24, 4.23, 4.23, 4.09, 3.78, 2.99), and six signals for amide NH protons (Figure S37), probably due to slow exchange to deuterium through inter-residual hydrogen bonding. In addition, signals for an N-prenyl unit were also observed, and two sets of them are overlapped according to the integral intensity (δC 138.6, 119.8, 40.7, 25.8, 18.1; δH 5.22, 3.82, 1.76, 1.71). A detailed analysis of the 2D NMR data established that 1 is the octapeptide composed of Phe, Ile, Val (x 2), Leu, Pro (x 2), and Arg residues (Figure 2). The sequence of these residues was determined by interpretations of the HMBC and ROESY spectra in CD₃OD, which allowed the assignment of two partial structures: the dipeptide Arg-Val(1), and the hexapeptide Phe-Ile-Val(2)-Leu-Pro(1)-Pro(2) (Figure 2a). It was difficult to connect these two fragments, due to the overlapped signals for the α-methylene protons of Arg and Val(1) at δH 4.23 in CD₃OD (Figure 2a). These two fragments were connected based on the ROESY correlations in DMSO-d₆ (Figure 2b), which allowed us to close the linear fragments into a head-to-tail cyclic octapeptide. The two prenyl units are attached to the two Nωs of guanidine in Arg, as deduced from the DQF 1H-1H COSY (in DMSO-d₆) correlations between NωH (δH 7.42) and CH₂ (δH 3.75), and NδH (δH 7.37) and CH₂ (δH 3.11), as well as the HMBC correlation between Cω and CH₂ (δH 3.75) (Figure 2b). The symmetrical pattern of bis-prenylation on Nω agrees with the overlapped signals of two prenyl groups, thus excluding the possibility of an Nω-, Nδ-asymmetric bis-prenylated guanidine. The

Figure 2. Key 2D NMR correlations of 1 in CD₃OD (a) and DMSO-d₆ (b).

Figure 3. Genetic basis of cyanobactin biosyntheses in M. aeruginosa NIES-88. (a) Cyanobactin-related genes encoded in remote genetic loci of M. aeruginosa NIES-88 chromosome. Genes are labeled with locus numbers. The locus tag prefix (MAN88) was omitted. Homologous genes are linked and amino acid identities are shown for each linkage. (b) Primary sequences of the precursor peptides AgcEs and KgpEs, with the core peptide (bold) and proposed recognition sequences RSII and RSIII (underlined). (c) Heterologous product ion of 3 in E. coli. Extracted ion chromatograms for m/z 922.5000 are shown.

(a) agcE1 A1 B1 C E1 F1 G1

(b) RSII

RSIII

(c) Retention time (min)

20

co-injection of agcE + kgpA2/G and std. (3)

agcE + kgpA2/G

agcE + kgpG

agcE + kgpA2

agcE

protease precursor peptide renyltransferase (PTase) other

Figure 3. Genetic basis of cyanobactin biosyntheses in M. aeruginosa NIES-88. (a) Cyanobactin-related genes encoded in remote genetic loci of M. aeruginosa NIES-88 chromosome.
amid bond between two Pro residues was assigned to be cis conformation based on ROESY correlations and diagnostic carbon chemical shifts ($\Delta_{CD} \sim \delta_{C}$.)$^{23}$

Compounds 2 and 3 have the molecular formulas of C$_5$H$_7$NO$_3$ (m/z 990.6495, and C$_5$H$_7$NO$_3$ (m/z 922.5873, respectively. Detailed analysis of the 2D NMR spectra revealed that 2 and 3 possess the same macrocyclic scaffold as 1, except that 2 has only one prenyl unit at Arg, while 3 lacks prenylation at Arg (Figure S2, Table S2). The Marfey’s analysis of 3 confirmed that the macrocyclic scaffold exclusively consists of L-amino acids (Figure S3). Structure of cyclic scaffold was further confirmed by total synthesis of 3, using conventional solid phase peptide synthesis, followed by ring closure with a coupling regent (Figure S4)$^{24-26}$ While 1-3 were not cytotoxic, 1 inhibited growth of *Staphylococcus aureus* with MIC of 3.12 - 6.25 µM. Notably, as the number of prenyl group increases from one to two, the antimicrobial activity was significantly enhanced. (Table S3, Figures S5-S10).

The unique N-prenylation of 1 prompted us to investigate its biosynthetic mechanism. As the cyclic scaffolds of argicyclamides exclusively consist of L-amino acids, we hypothesized that these are biosomally synthesized peptides. Based on this assumption, we searched for a putative precursor peptide of argicyclamide, *agcE1* and *agcE2*, which are encoded in the remote genetic loci (Figure 3a, Supplementary results and discussion). While no cyanobactin-related genes were encoded in the flanking region of *agcE1*, a putative enzyme *AgcF*, which shares homology with cyanobactin PTases such as Trp (AAid, 41%)$^{14}$ Trp N-prenyltransferase *AcyF* (AAid, 41%)$^{15}$ and Nα-prenyltransferase *MysF2* (AAid, 35%)$^{19}$ was found in the flanking region of *agcE2*. However, although the cyclic scaffolds of argicyclamides should be constructed by a set of PatA/G-like proteases, no such proteases were encoded in the vicinity of the *agcE1* or *agcE2*. This led us to hypothesize that ArgEs are processed by putative PatA/G-like proteases that are encoded in remote genetic loci. To search them, complete genome of *M. aeruginosa* NIES-88 was analyzed with HMM-based annotation.$^{27}$ This validated that *KgpA* and *KgpG* proteases in kawaguchipeptin biosynthesis, are the only PatA/G-like proteases (IPR023830) in *M. aeruginosa* NIES-88 chromosome. Notably, we detected two sets of near-identical kawaguchipeptin biosynthetic gene clusters ( *kgpA1* and *kgpA2*) (Figure 3a) that reside in remote loci in *M. aeruginosa* NIES-88 chromosome (Supplementary results and discussion). Genes coding for B, E, F, G proteins are identical between *kgpA1* and *kgpA2*. On the other hand, *kgpA2* encodes full-length of PatA-like protease, N-terminal region of *kgpA1* encoding catalytic triad is truncated, thus *KgpA1* is likely to be inactive. Based on these observations and the fact that the RSII (recognition sequence of PatA-like protease) and RSIII (recognition sequence of PatG-like protease) of AgcEs are similar to those of *KgpE*, kawaguchipeptin precursor peptide (Figure 3b), we hypothesized that AgcEs could be matured by *KgpA2* and *KgpG* proteases for kawaguchipeptin biosynthesis.

To test this hypothesis, we co-expressed the precursor peptides *agcE* and *kgpA2/kgpG* in the heterologous host *E. coli* BL21 (DE3). LC-MS analyses of the transformant’s metabolites revealed the production of 3, the nonprenylated cyclic peptide (Figure 3c). Omitting *kgpA2* or *kgpG* gene resulted in the loss of 3 production. To further access the production of 3 in vitro, recombinant KgpG was incubated with synthetic AgcE-core with follower peptide (RVFIVLPPFAEDGAE) (Figure S11). This resulted in KgpG-dependent production of 3 (Figure S12). These results demonstrated substantial promiscuity of kawaguchipeptin proteases and suggest that these are employed for constructing the cyclic scaffolds of both argicyclamides and kawaguchipeptins, which are structurally distinct.

Next, we investigated the biosynthetic origin of the bis-prenylated guanidine moiety in 1. To this end, the putative prenyltransferase *AgcF* was expressed in *E. coli* and characterized in vitro, using dimethylallyl diphosphate (DMAPP) as a prenyl donor and synthetic 3 as an acceptor. A time-course analysis revealed the initial generation of 2, the mono-prenyl-
ated product, followed by the accumulation of 1, the bis-prenylated product (Figure 4a). This demonstrated that AgcF catalyzes two rounds of prenylation on the guanidine moiety. This conversion is Mg^{2+}-dependent, and is optimal at 37°C, pH 8.0 (Figure S13–S14). AgcF showed no prenylation activity on synthetic analogs of 3 with Arg substituted to Trp, Tyr, Ser, Thr, or Lys (6-10, Figure S15). No prenylation was observed for linear analogs (11 and 12, Figure S15). These results show that AgcF is highly selective for 3 as prenyl acceptor. Next, specificity of AgcF on isoprenyl donor was assessed by using geranyl diphosphate (GPP) and farnesyl diphosphate (FPP). Contrary to other cyanobactin PTases that are highly selective for the isoprenyl donor,12-16,28 AgcF was capable of catalyzing the mono-prenylation of 3 (Figures S16-S17), highlighting its unusual tolerance. FPP was not accepted as an isoprenyl donor.

To gain structural insights into AgcF catalysis, we generated a computational model of AgcF, using the crystal structure of PagF, an O-Tyr PTase in prenylagaramide biosynthesis, as the template (AAId, 42%) (Figure S18).13 The Mg^{2+} and diphosphate binding site, as well as the proposed catalytic residue (Glu49 in AgcF and Glu51 in PagF) that activates the prenyl acceptor, are well conserved between AgcF and PagF. The substitution of Glu49 to Ala in AgcF abolished its prenylating activity, showing the general importance of this residue in the catalysis of cyanobactin PTases (Figure S19). Stark differences between AgcF and PagF were observed in the residues forming the active site entrance, where the bulky residues in PagF are substituted with substantially smaller residues in AgcF; for example, F69/G67, H138/G133, W271/C267, and Y292/L289, respectively (Figure S19). The enlarged active site should facilitate the accommodation of a bulky substrate and enable the sequential bis-prenylation of guanidine.

A phylogenetic analysis revealed that cyanobactin PTases form clades according to their chemo-selectivities (Figure S20). Notably, AgcF composes a small but distinct clade together with its close homologs, and all share the aforementioned substitutions at the active site entrance (Figure S21). The putative Arg-containing precursor peptides encoded in the neighboring regions of AgcF-like PTases suggest the presence of a new class of cyanobactins, with a bis-prenylated Arg residue yet to be identified (Figure S22).

In this study, we discovered argicyclamides (1-3), a new group of cyanobactins with a unique bis-prenylated Arg residue. Based on the complete genome sequence and series of biochemical analyses, we proposed unique biosynthetic route of cyanobactins, with a bis-prenylated Arg residue yet to be identified (Figure S22). The Supporting Information is available free of charge on the ACS Publications website.

REFERENCES