In Vivo Characterization of Tandem C-Terminal Thioesterase Domains in Arthrofactin Synthetase

Niran Roongsawang, Kenji Washio, and Masaaki Morikawa

Macrocyclization of a peptide or a lipopeptide occurs at the last step of synthesis and is usually catalyzed by a single C-terminal thioesterase (Te) domain. Arthrofactin synthetase (Arf) from Pseudomonas sp. MIS38 represents a novel type of nonribosomal peptide synthetase that contains unique tandem C-terminal Te domains, ArfC_Te1 and ArfC_Te2. In order to analyze their function in vivo, site-directed mutagenesis was introduced at the putative active-site residues in ArfC_Te1 and ArfC_Te2. It was found that both Te domains were functional. Peaks corresponding to arthrofactin and its derivatives were absent in ArfC_Te1:S89A, ArfC_Te1:S89T, and ArfC_Te1:E26G/F27A mutants, and the production of arthrofactin by ArfC_Te2:S92A, ArfC_Te2:S92A/D118A, and ArfC_Te2 was reduced by 95% without an alteration of the cyclic lipoundecapeptide structure. These results suggest that Ser89 in ArfC_Te1 is essential for the completion of macrocyclization and the release of product. Glu26 and Phe27 residues are also part of the active site of ArfC_Te1. ArfC_Te2 might have been added during the evolution of Arf in order to improve macrocyclization efficiency.

Introduction

Many bioactive macrocyclic compounds, such as tyrocidine, surfactin, arthrofactin, erythromycin and epothilone are produced by microorganisms by nonribosomal peptide synthetases (NRPS), polyketide synthases (PKS) and hybrid PKS/NRPS. Having a macrocyclic structure decreases the conformational flexibility of a molecule compared to its linear analogue, and this can constrain it to a biologically active conformation.[1] NRPS are modular multifunctional enzymes that recognize, activate, modify and link amino acid intermediates to the final product.[2] Each module of NRPS can be further subdivided into domains, each of which exhibits a single enzymatic activity. The adenylation (A) domain is responsible for amino acid recognition and adenylation at the expense of ATP. The thiolation (T) domain is the attachment site of Ppant and serves as a carrier of thioesterified amino acid intermediates. The condensation (C) domain catalyzes peptide bond formation between sequential amino acids. The modifying epimerization (E) domain catalyzes the conversion of L-amino acids to D-isomers and is typically associated with the α-amino-acid-incorporating module. Lastly, the C-terminal thioesterase (Te) domain generally catalyzes the macrocyclization and release of linear intermediate peptides.

Arthrofactin (Figure 1) is a cyclic potent lipoundecapeptide biosurfactant that is produced by the Gram-negative bacterium Pseudomonas sp. MIS38.[3,4] The molecule is cyclized through the formation of an ester bond between the carboxyl group of the C-terminal Asp and the hydroxyl group of β-alloThr (Ikegami et al., unpublished data). The biosynthesis of arthrofactin is catalyzed by arthrofactin synthetase (Arf), which consists of three NRPS protein subunits: ArfA (234 kDa), ArfB (474 kDa), and ArfC (648 kDa). Arf represents a novel type of NRPS that contains a dual C/E domain and tandem C-terminal Te domains.[4,5] It is assumed that leucine is activated and coupled to the T domain of the first module of ArfA. The β-hydroxydecanoyl thioester is then coupled to the activated leucine by the action of the first C-domain and provides β-hydroxydecanoyl-L-leucine as the initial intermediate.[5,6] This intermediate is sequentially elongated into lipoundecapeptide through the concerted action of the Arf complex. During the aminoacyl/peptidyl-thioester stage, L-amino acids are epimerized to the D-configuration by dual C/E domains.[5] The full-length lipoundecapeptide is expected to be cyclized and released from Arf by the function of unique tandem Te domains.

Two types of Te domains, internal and external are generally associated with NRPS and PKS. Most NRPS and PKS have only one internal Te domain at the C terminus of the last module. This internal Te domain (type I, Tel) carries a typical GXSG (X = any amino acid residue) sequence motif with highly conserved Asp and His residues.[7] The initial function of the Tel domain involves the acceptance of the linear peptide from the last T domain to form a peptide–O–Te intermediate. Concomitant deacylation of the intermediate results in either hydrolysis, or intramolecular cyclization of a linear product.[8] The other type of Te domain is the external stand-alone Te (type II, Tell). This
protein also contains a GXSG sequence motif and highly conserved Asp and His residues, and is involved in the regeneration of misprimed T domains by removing short acyl chains from the 4'-Ppant. Moreover, a recent study has suggested that the TeII domain also hydrolyzes incorrectly loaded amino acids, which are not processed by the nonribosomal machinery.

Cyclization and release of the cyclic peptides are usually catalyzed by a single internal TeI domain of 25–35 kDa (~250 aa). However, ArfC has a larger C-terminal region of approximately 62 kDa (580 aa) and shows significant similarity with TeI. This region bears putative tandem Te domains ArfC_Te1 and ArfC_Te2, both with a set of possible catalytic triads: Ser89/Asp116/His264 and Ser92/Asp118/His259, respectively. TeI of NRPS possesses either hydrolase (e.g., ACV synthetase) or cyclase activity (e.g., surfactin synthetase), which results in the release of free carboxylate products or cyclic lactones, respectively. We wonder if ArfC_Te1 and ArfC_Te2 share coordinated hydrolase and cyclase activities, or whether either one has the cyclase activity that is responsible for the completion of the arthrofactin biosynthesis. Here, we tested the function of ArfC_Te domains in vivo by introducing a site-directed mutation at the putative active site residues.

Results and Discussion
Molecular diversity of Te domains

Both NRPS and PKS commonly have a modular architecture of repetitive catalytic units and function like an assembly-line. After the synthesis of linear intermediates, the cyclization or hydrolysis of the product from enzymes is carried out by an internal Te domain. Additionally, an external TeII domain is associated with these biosynthesis systems. In order to analyze the evolutionary relationship among Te proteins, a phylogenetic tree was constructed with various Te proteins of PKS and NRPS. A total of 120 Te proteins from bacteria and fungi were clustered according to the type of reactions that they catalyze, and by organism group (Figure 2). Te proteins are grouped into three major classes, these are Tel of NRPS, Tel of PKS, and Tel of NRPS and PKS. Tel of NRPS is the most diverse group and can be further classified into five subclasses, they are cyclase (subclass I), hydrolase (subclass II), cyclase and hydrolase of actinomycetes (subclass III), putative cyclase (subclass IV) and cyclase and hydrolase of hybrid PKS/NRPS (subclass V).

Subclass I is composed of cyclase-type Te domains from Gram-positive Bacillus and Gram-negative cyanobacteria. This cyclase produces both cyclic macrolactones, such as surfactin, lichenysin and fengycin (1JMK/LicC/LchAC/FenB_Te), and...
Figure 2. A phylogenetic tree analysis of 120 Te proteins of PKS and NRPS, mainly from bacteria. ArfC_Te1 and ArfC_Te2 are indicated by arrows. The scale bar represents 10 substitutions per 100 amino acids. Bootstrap values higher than 500 are indicated. The Te proteins used in this analysis are shown in Table S1 in the Supporting Information.
Figure 3. Amino acid sequence alignment of tandem C-terminal Te domains with FenB_Te and SrfC_Te (1JMK). The sequences analyzed here include ArfC from Pseudomonas sp. M358 (BAC67536), PfO from P. fluorescens PfO-1 (ZP_00265375), Pf5 from P. fluorescens Pf-5 (AAY91421), SyPC from P. syringae pv. syringae B301D (AAO72425), DC from P. syringae pv. tomato strain DC3000 (NP_792634), B278a from P. syringae pv. syringae B728a (ZP_00205846), GMI from Ralstonia solanacearum GMI1000 (NP_522203), SCRI from Erwiniacarotovora SCRI1043 (YP_049592), BurM from Burkholderia mallei ATCC2334 (LY_106216), BurP from B. pseudomallei K96243 (YP_111640), FenB from Bacillus subtilis F29-3 (AAB00093), and SrfC from B. subtilis 168 (Q08787). The GXG motif is underlined and the positions of the catalytic triad residues of SrfC_Te (S80/D107/H207) are indicated by asterisks. The predicted secondary structure of ArfC_Te1 / ArfC_Te2 and secondary structure of SrfC_Te are shown as arrows (β-strand) and cylinders (α-helix) on the top of sequences. Glu26 and Phe27 in ArfC_Te1, and Gln7 in ArfC_Te2 are indicated by arrow heads. The lid region is indicated by the dotted line.
cyclic macrolactam products such as tyrosidine, bacillomycin, microcystin and bacitracin (Tyc/BamC/Mcy/BacC_Te). Subclass II is composed of hydrolase-type Te domains from Gram-positive/negative bacteria and fungi, and catalyze the hydrolysis of peptide intermediates in β-lactam antibiotics synthetase (ACV/Pcb_Te) from fungi, actinomycetes, and Gram-negative bacteria. Additionally, this hydrolase-type Te is also found in pyoverdine synthetase (Pv_Te) from Gram-negative Pseudomonas species. This subclass also contains the multimodular fatty acid synthase for mycolic acids (Pks13_Te), which are high-molecular-weight α-alkyl-β-hydroxy acids that are unique to the mycobacteria. The Te of subclass III hydrolyzes linear

Figure 3 (continued).
peptide precursors of vancomycin-type antibiotics (Bpsc/CepC/TelID/StaD/ComD_Te)\textsuperscript{[7]} or cyclizes calcium-dependent antibiotics (CAD3/DptD_Te).\textsuperscript{[10]} Interestingly, the iterative cyclases of *E. coli* or *Salmonella* sp. *enterobactin* (EntF) and *Bacillus subtilis* *bacilliadin* (DhbF), an aryl cap siderophores are closely related to subclass III.\textsuperscript{[19]} This suggests that a close evolutionary relationship among these Te groups exists. Gene transfer from the filamentous bacteria to unicellular bacteria or vice versa might have happened during the process of gene evolution.

There are several putative NRPS that contain tandem internal Te domains similar to those found in arthrofactin and syringopeptin synthetases.\textsuperscript{[14,16]} These tandem Te domains, namely Te1 and Te2, (each – 280 aa) are clustered in subclass IV and V, respectively. They might have evolved from different ancestral genes, instead of gene duplication in the cell. We propose that subclass IV is a novel cyclase-type Te1 because several peptide products of this group form macrolactone structures between the C-terminal amino acid and the hydroxyl group of Thr or Ser (ArfC/SypC/P5/SyrE_Te).\textsuperscript{[14,20,21]} Notably, SyrE_Te in syringomycin synthetase contains only one internal Te, but it belongs to this group. The biochemical characterization of SyrE_Te showed that it is indeed a cyclase.\textsuperscript{[21]} The function of subclass V Te2 is as yet unknown, and we propose that this subclass is a novel type cyclase/hydrolase Te2, because it is closely related to the cyclase and hydrolase of the hybrid PKS/NRPS.\textsuperscript{[22,23]} This phylogenetic analysis also suggests that the cyclase/hydrolase Te2 is not a lineage of TeII that had been fused to internal Te1 because TeII of NRPSs and PKS forms a distinctly separate branch. Tel of PKS forms a cluster that is different from Tel of NRPS. This result would explain the different substrate specificity of these two Te classes: one is specific for polyketides and the other for peptide intermediates.

**Construction of ArfC_Te1 and ArfC_Te2 mutants**

The NRPS architecture, which is characterized by tandem Te domains, is found in several species of Gram-negative bacteria, notably *Pseudomonas* sp., *Ralstonia* sp., *Burkholderia* sp., and *Erwinia* sp. The amino acid sequences of ArfC_Te1 and ArfC_Te2 were compared with those of orthologous tandem Te domains, and also with SrfC_Te and FenB_Te, which have known crystal structures. It was found that Ser80, Asp107 and His207, which form a catalytic triad in SrfC_Te, are completely conserved among them. The only exceptions were BurM_Te1 and BurP_Te1, where Ser80 was replaced with Cys80. These results suggest that both ArfC_Te1 and ArfC_Te2 are functional (Figure 3). The secondary structure of ArfC_Te1 and ArfC_Te2 was predicted by PSIPRED.\textsuperscript{[24]} Like SrfC_Te and FenB_Te, ArfC_Te1 and ArfC_Te2 consist of a seven-stranded β-sheet.\textsuperscript{[27,25]} Further, SrfC_Te was found to form two distinct conformations at the lid region. This region (from Lys111 to Ser164) covered most of the active site of the enzyme.\textsuperscript{[27]} There are insertions of peptide at the N-terminal of the putative lid region in ArfC_Te1 and ArfC_Te2 (Figure 3). This would make the structure of both ArfC_Te domains more complex than SrfC_Te and FenB_Te.

In order to determine the function of two Te domains in Arf, site-directed mutagenesis at the putative catalytic GXSXG motif was conducted on ArfC_Te1 (Ser89) and ArfC_Te2 (Ser92). These serine residues were replaced by alanine or threonine to give ArfC_Te1:S89A, ArfC_Te1:S89T, ArfC_Te2:S92A, ArfC_Te2:S92T, and ArfC_Te2:S92A/D118A, a double mutant. Moreover, the ArfC_Te2 deletion mutant (ArfCΔTe2) was also constructed by inserting a stop codon in the boundary region between ArfC_Te1 and ArfC_Te2. This boundary region was deduced from the secondary structure prediction of ArfC_Te (Figure 3). Then, a CAA codon (Gln7), which was located at the N-terminal of ArfC_Te2 was replaced by a TGA stop codon. Integration of the plasmid into the chromosome by first crossing-over at either side of the mutation point (case 1 or 2, Figure 4A) was confirmed by PCR, and yielded a 3.4-kb fragment (figure not shown). This result suggests that the recombinant suicide plasmid was integrated at the expected position. A second crossing-over was initiated by growing the cells to the late logarithmic in a non-selective L-broth. Serial dilutions were inoculated onto L plates containing 6% sucrose without NaCl. Although two outcomes after the second crossing-over were possible, only the successful mutagenesis (case 4; Figure 4B) was obtained; a sequencing experiment confirmed that the PCR was error-free.

**Arthrofactin production by the mutants**

Production of arthrofactin by a wild-type MIS38, mutant NC1 (see the Experimental Section), ArfC_Te1:S89A, ArfC_Te1:S89T, ArfC_Te2:S92A, ArfC_Te2:S92A/D118A, and ArfCΔTe2 were compared by HPLC-UV and LC-MS (Figures 5 and 6). Peaks corresponding to arthrofactin (C10, m/z = 1354.9) and its derivatives (C9 and C12) were found in the sample from MIS38 (total amount 220 ± 3.6 mg L\textsuperscript{-1}), while they were absent in that from mutant NC1, ArfC_Te1:S89A, and ArfC_Te1:S89T. This result was reconfirmed by LC-MS (Figure not shown). It indicates that the Ser89 residue in ArfC_Te1 is essential for the completion of arthrofactin synthesis, and that the exact location of the hydroxy group in the serine side chain is important for catalytic function; serine cannot be replaced by threonine. Similarly, the production of arthrofactin in ArfC_Te2:S92A (12.5 ± 4 mg L\textsuperscript{-1}), ArfC_Te2:S92A/D118A (12.5 ± 1 mg L\textsuperscript{-1}), and ArfCΔTe2 (13.4 ± 4 mg L\textsuperscript{-1}) was reduced by 95% without alteration of the cyclic lipodecapeptide structure. These results allowed us to conclude that ArfC_Te1 and ArfC_Te2 function cooperatively to cyclize and release the peptide product. Interestingly, the proteins that resulted from the deletion of the entire Tel domain in surfactin synthetase, and the serine-to-alanine site-directed mutagenesis in fungal ACV synthetase also retained a slight but significant activity.\textsuperscript{[26,27]} This suggests that autonomous cyclization could occur without the Te domain in these synthetases. Our observation suggests that ArfC_Te2 functions similarly to Tel, and that ArfC_Te1 functions as the last acceptor of linear peptide intermediates, like the last T domain located before Tel. Meanwhile, less reduction of surfactin (84%) was observed in the external Tel mutant.\textsuperscript{[26]} Disruption of the external Tel in a modular PKS also resulted in a moderate drop (20–85%) in polyketide production.\textsuperscript{[26]} A drastic reduction of arthro-
factin production in ArfC_Te2 mutants supports the idea that ArfC_Te2 is functionally different from the external TeII.[26,28]

In order to understand the catalytic mechanism of ArfC_Te1 more deeply, we constructed two more mutants. Based on the crystal structure and amino acid sequence alignment of cyclase Te domains (SrfC_Te) and lipases (hydrolases), we focused on the amino acid at position 26, where proline (Pro26) is conserved among cyclases and glycine (Gly26) among hydrolases.[8] The 26th amino acid, which is located near the oxyanion hole in the active site (Val27 and Ala81), might determine the reaction type, that is either cyclization or hydrolysis. Tseng et al., reported that the SrfC_Te P26G mutant mainly hydrolyzes and releases its linear peptide in vitro. [8] They proposed that a change from a rigid proline to a flexible glycine could increase the conformational freedom in the region of the active site, and could result in easier access of a water molecule to the active site. The corresponding residue in ArfC_Te1 and ArfC_Te2 were identified as Glu26 and Gly26, respectively (Figure 3). Therefore, E26/F27 in ArfC_Te1 was replaced by P26/V27 (similar to SrfC_Te) and G26/A27 (similar to ArfC_Te2). Production of arthrofactin in the mutants was compared by HPLC–UV and LC–MS (Figures 5 and 6). It was found that ArfC_Te1:E26P/F27V produced approximately 1% of the amount of arthrofactin produced by MIS38 (2.2 ± 1 mg L⁻¹), and ArfC_Te1:E26G/F27A produced no arthrofactin at all (figure not shown). We could not detect linear arthrofactin intermediates in either the intracellular or extracellular fraction of the mutants. This result suggested that Glu26 and Phe27 in ArfC_Te1 also constitute the active site, and that a common cyclization mechanism is shared by SrfC_Te and ArfC_Te1. This study demonstrates that ArfC_Te1 is critical for arthrofactin synthesis because a single mutation at the Ser89 residue completely abolished arthrofactin production. ArfC_Te2 seems to be not essential however, it still supports an efficient synthesis of arthrofactin because the deletion of this domain, or mutation at Ser92 retained only slight (5%) arthrofactin production activity.

According to the SrfC_Te model, a catalytic triad in the Te domain is formed by Ser80, which acts as the nucleophile, His207, which acts as the acid–base catalyst, and Asp107 which optimally orients the histidine and serine residues.[7,8] These active-site residues effectively macrocyclize and release the product surfactin. The cyclization and release of the arthrofactin lipoundecapeptide chain from the enzyme is likely mediated by two Te domains in a series mechanism shown in Figure 7. First, the lipoundecapeptidyl chain bound to the adjacent T11 domain is directed to an invariant serine residue (Ser89) of ArfC_Te1, which has been activated by Asp116 and His264 to form a peptide–O–Te1 intermediate (Figure 7A).
Experimental Section

Bacterial strains and plasmids: Arthrofactin-producing Pseudomonas sp. Ml38 was previously isolated from oil spills in Shizuoka prefecture, Japan.

Arthrofactin-deficient Pseudomonas sp. NC1 was used as a negative control and was previously constructed by inserting a kanamycin-resistant gene cassette (Km) in the arfB gene. E. coli DH5α was used as a host strain for the construction of recombinant plasmids. E. coli SM10[pir] was used for transforming Ml38 with the suicide vector pCVD442-Km. Cloning vectors pUC18 and pGEM-T Easy were used in E. coli DH5α. pSMC32 is a derivative of pSU36 (X53938). pCVD442 is a suicide vector that contains a pir-dependent R6K replicon and sacB gene from Bacillus subtilis which allows positive selection with sucrose for loss of the vector.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.

General DNA manipulations: Genomic DNA of Ml38 was prepared by using the Sarkosyl method and was purified by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation.

DNA fragments were recovered from an agarose gel by ethidium bromide equilibrium density gradient ultracentrifugation prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.
used for PCR as shown in Table 1. The PCR products were first cloned into pGEM-T Easy vector, then the SacI fragment was excised from the plasmid and introduced into the suicide vector pCVD442. The resulting suicide vector, designated pCVD442-Km was transferred into E. coli SM10::pir by an electroporation method as follows and was subsequently used for different kinds of allelic exchanges.

**Electrotransformation of E. coli SM10::pir:** Cells were grown in L-broth until the mid-log phase (OD<sub>600</sub> = 0.4). After collection by centrifugation (5000 g for 15 min at 4 °C), the cells were washed twice with ice-cold pure H<sub>2</sub>O. Then, the cells were washed twice with glycerol (10%), and resuspended in glycerol (10%) at 3 x 10<sup>10</sup> cells per mL. A portion of this cell suspension (40 μL) was mixed with purified recombinant DNA (50 ng) and was kept on ice for 5 min. The DNA/cell mixture was transferred into a cuvette (0.1 cm electrode distance) and subjected to a high electric field pulse (14 kV cm<sup>-1</sup> with 35 μF and 5 ms) by using the Electro Gene Transfer Equipment (Shimadzu GTE-10) equipped with a time constant optimizer (Shimadzu TCO-1). Treated cells were immediately suspended in 1 mL of L-broth and grown for 1 h at 30 °C before plating onto L/Amp-agar plates (Amp = 50 μg/mL).

**Cloning of native arfC<sub>Te1</sub> and arfC<sub>Te2</sub> gene and its flanking region:** It is important that both sides of the target gene have a sufficient length (ca. 1 kb) of flanking DNA for the homologous recombination in the next step. Therefore, the native 2kb arfC<sub>Te1</sub> and arfC<sub>Te2</sub> gene fragment, which have a flanking regions of around 1 kb was amplified by the PCR method by using MIS38 chromosomal DNA as a template. The following oligonucleotide primers, Te1-XbaI/Fw and Te1-XbaI/Rv for the arfC<sub>Te1</sub> gene, and Te2-XbaI/Fw and Te2-XbaI/Rv for the arfC<sub>Te2</sub> gene, which contained the XbaI site (underlined) were used (Table 1). The PCR products were cloned into pGEM-T Easy vector to yield pGEM-Te1 and pGEM-Te2. Sequencing confirmed that the PCR experiment was error-free.

**Site-directed mutagenesis of catalytic residues in ArfC<sub>Te1</sub> and ArfC<sub>Te2</sub>:** The arfC<sub>Te1</sub> and arfC<sub>Te2</sub> genes were mutagenised by the overlap extension method. Constructs were obtained by PCR amplification of the pGEM-Te1 or pGEM-Te2 template. In the first PCR reaction, the 5'-fragment of the mutant gene was amplified by using the primers Te1-XbaI/Fw or Te2-XbaI/Fw and mutation-Rv primers, and the 3' -fragment was amplified by using the mutation-Fw and Te1-XbaI/Rv or Te2-XbaI/Rv primers (Table 1). After agarose gel purification, the two fragments were mixed together and the full-length gene was further amplified by using Te1-XbaI/Fw or Te2-XbaI/Fw primers and Te1-XbaI/Rv or Te2-XbaI/Rv primers. The blunt-ended PCR product was first cloned into pUC18 at the SmaI site and then the XbaI fragment was excised and ligated into the XbaI gap of the pCVD442-Km vector. The resulting plasmids, designated pCVD442-Km:S89A, pCVD442-Km:S89T, pCVD442-Km:S92A, pCVD442-Km:S92A/D118A, pCVD442-Km:E26P/F27V, pCVD442-Km:E26G/F27A and pCVD442-Km: ΔTe2 were transferred into E. coli SM10::pir and then introduced into an arthrofactin-producing Pseudomonas sp. MIS38, by mating with selection for kanamycin and chloramphenicol resistance. The wild-type MIS38 is resistant to high concentrations of chloramphenicol and sensitive to kanamycin.

**Isolation of mutant strains:** Donor and recipient strains were grown in L-broth until the OD<sub>600</sub> values reached to 0.5. Cells were then mixed at an equal ratio and spotted onto a L plate without antibiotics. After 18 h conjugation at 30 °C, the cells were scraped and resuspended in L-broth and spread onto an L-agar plate.
that contained chloramphenicol (34 µg mL⁻¹) and kanamycin (35 µg mL⁻¹). After an overnight incubation at 30°C, individual colonies were analyzed. Transconjugants that had the plasmid integrated into the chromosome via homologous recombination were selected by their Cmr and Kmr phenotype. One of the transconjugants was allowed to grow at 30°C for 18 h in L-broth without antibiotics. Serial dilutions were inoculated onto L agar plates containing sucrose (6%) without NaCl, and were incubated for 24 h at 37°C. The omission of NaCl from this medium was shown previously to improve the sucrose counterselection. The presence of the sacB gene in pCVD442 inhibits growth on sucrose plate. Therefore, growth on sucrose is a positive selection for the loss of the suicide vector sequences from the chromosome by second crossover. Sucrose-resistant colonies were picked and tested for Km sensitivity, which indicated the loss of the pCVD442-Km part. Such colonies were tested for the successful introduction of the mutation in arfC_Te1 or arfC_Te2 by cloning and sequencing the target gene locus. Primers for amplifying the gene from arfC_Te1 mutants are

---

**Figure 7.** Proposed mechanism of ArfC_Te1 and ArfC_Te2. A) MIS38, B) ArfC_Te2:S92A, C) ArfC_Te1:S89A. The side chain of the potential nucleophiles of ArfC_Te1:S89 and ArfC_Te2:S92 are represented by -CH₂-OH whereas -CH₃ represents the side chain of alanine. Peptidyl chain transfer and the subsequent cyclase release are abrogated in the ArfC_Te1:S89A. Each domain is similarly symbolized as in Figure 1. Only the structural formula of Thr3 and Asp11 in the peptide chain is shown. R indicates an alkyl chain.
Analysis of arthrofactin production: Wild-type MIS38 and mutants were grown in L-broth (100 mL) at 30 °C for 72 h. Arthrofactin and its derivatives were purified as described previously. Briefly, the supernatant was acidified by adding concentrated HCl to a final pH of 2.0, and then was allowed to form aggregates at 4 °C for 3 h. The aggregates were collected by centrifugation and were used for the analysis by reverse-phase HPLC as described below.

Reversed-phase HPLC was carried out on an octadecyl silica gel column (Cosmosil SCXAR 4.6 x 150 mm, Nacalai, Kyoto, Japan) attached to a system HP1100 (Hewlett-Packard, Palo Alto, California) at a flow rate 0.5 mL min⁻¹ of solvent mixture A (10% acetonitrile/0.1% TFA) and B (100% acetonitrile/0.1% TFA). The elution program was performed by changing the ratio of solvent A and B, and was optimized as follows: %B = 0 (0–5 min), %B = 0–100 (5–35 min), %B = 100 (35–40 min), and %B = 0 (40–45 min). Peaks eluting from the column were monitored by their absorbance at 210 nm. The molecular weight of each component was determined by using a mass spectrometer LCQ (Thermo Finnigan) equipped with an electrospray ion source. The yields of arthrofactin were calculated from the peak areas and by weighing the methanol extracts of the acid precipitates.

Acknowledgements

N.R. acknowledges his post-doctoral fellowship from Japan Society for the Promotion of Science (P04468). We would like to thank Dr. Deborah Hogan (Dartmouth Medical School) and Dr. Roberto Koter (Harvard Medical School) for providing plasmids pSMC32, pCDV442, and E. coli SM10 Δpir. This work was supported by the Grants-in-Aid for Scientific Research for Exploratory Research of the MEXT (No. 17510171) and Takeda Science Foundation.

Keywords: cyclic lipopeptides · cyclization · nonribosomal peptide synthetases · peptides · thioesterase domain

Table 1. Primers used in this study[a]

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSMC-Socl/Fw</td>
<td>5'-CATGAGCTCGTTTATGAGAACGAGGTCGTAAGGAA</td>
</tr>
<tr>
<td>pSMC-Socl/Rv</td>
<td>5'-CATGAGCTCGTTTATGAGAACGAGGTCGTAAGGAA</td>
</tr>
<tr>
<td>Te1-Xocl/Fw</td>
<td>5'-CATGCTAGATGCAGGATCGTGAACGAGGACGAGG</td>
</tr>
<tr>
<td>Te1-Xocl/Rv</td>
<td>5'-CATGCTAGATGCAGGATCGTGAACGAGGACGAGG</td>
</tr>
<tr>
<td>Te2-Xocl/Fw</td>
<td>5'-CATGCTAGATGCAGGATCGTGAACGAGGACGAGG</td>
</tr>
<tr>
<td>Te2-Xocl/Rv</td>
<td>5'-CATGCTAGATGCAGGATCGTGAACGAGGACGAGG</td>
</tr>
<tr>
<td>S89A-Fw</td>
<td>5'-TGGCCGGCTGCATGGATCTGCCCAGG</td>
</tr>
<tr>
<td>S89A-Rv</td>
<td>5'-ACCCCGCCGAAATCCGCAAGGAGGAGG</td>
</tr>
<tr>
<td>SacI/Fw</td>
<td>5'-CACCAGCCTGACGGGCCGAGGAGGG</td>
</tr>
<tr>
<td>SacI/Rv</td>
<td>5'-CGTCGTCAGCAGCGGT</td>
</tr>
<tr>
<td>XbaI/Fw</td>
<td>5'-TCGTCAGCAGCGGT</td>
</tr>
<tr>
<td>XbaI/Rv</td>
<td>5'-AGCCGCCGCGCTCAGTG</td>
</tr>
</tbody>
</table>

[a] Introduced mutations are bold and italicized.

Received: October 30, 2006
Published online on February 27, 2007