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

Niran Roongsawang, Kenji Washio, and Masaaki Morikawa*[a]

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 ArfCΔTe2 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 4′-phosphopantetheine cofactor (4′-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 D-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 D-allo-Thr (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, TeI) carries a typical GXSXG (X = any amino acid residue) sequence motif with highly conserved Asp and His residues.[7] The initial function of the TeI 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, TeII). This

[a] Dr. N. Roongsawang, Dr. K. Washio, Prof. Dr. M. Morikawa
Division of Biosphere Science
Graduate School of Environmental Science, Hokkaido University
Sapporo 060-0810 (Japan)
Fax: +(81) 11-706-2253
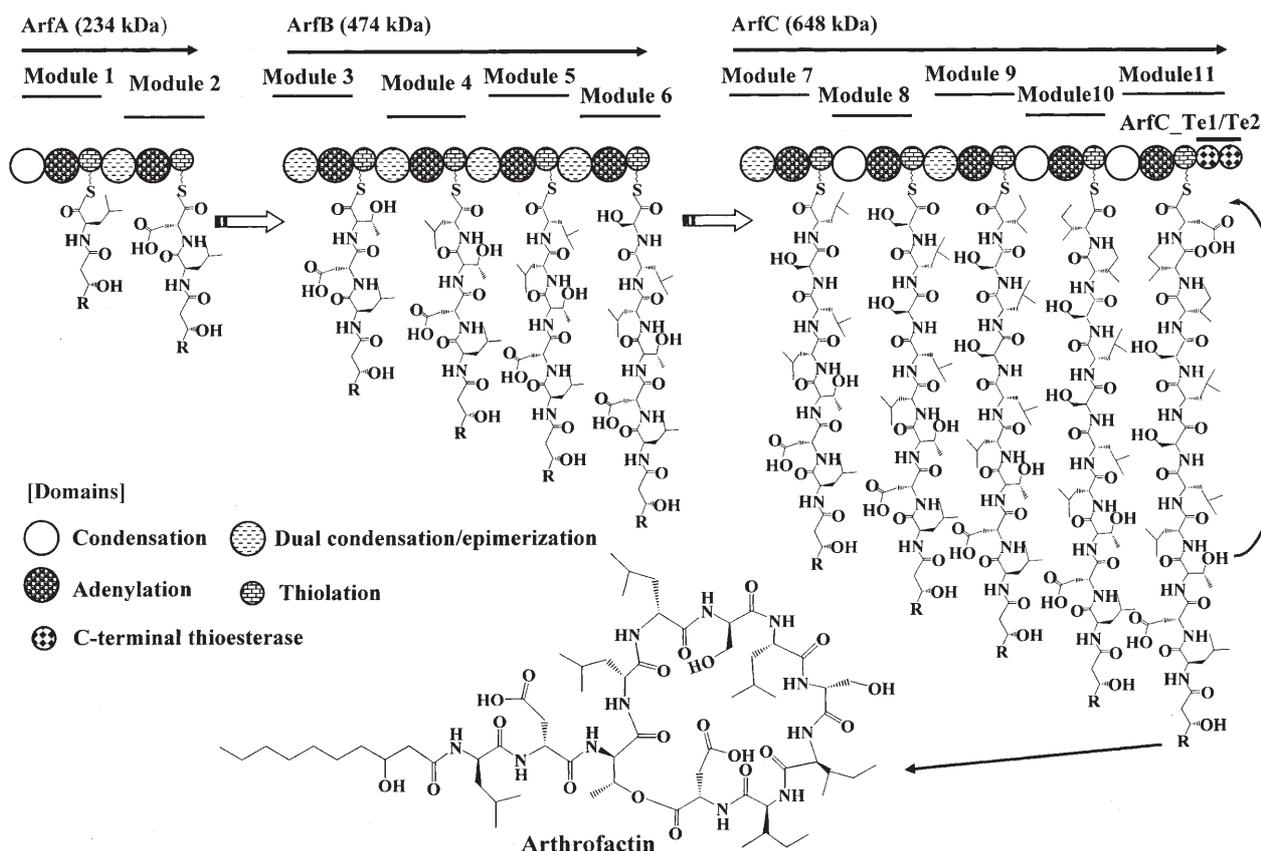E-mail: morikawa@ees.hokudai.ac.jp

Figure 1. The arthrofactin assembly line. The multienzyme complex consists of eleven modules that are specific for the incorporation of eleven amino acids. Thirty-three domains are required for peptide elongation, while the last two Te domains are unique and expected to be required for peptide release by cyclization.

protein also contains a G*XS*X*G* sequence motif and highly conserved Asp and His residues,[9] and is involved in the regeneration of misprimed T domains by removing short acyl chains from the 4'-Ppant.[10] Moreover, a recent study has suggested that the TeII domain also hydrolyzes incorrectly loaded amino acids, which are not processed by the nonribosomal machinery.[11]

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.[12] 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 TeI domain. Additionally, an external TeII domain is associated with these biosynthesis systems.[9] 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 TeI of NRPS, TeI of PKS, and TeII of NRPS and PKS. TeI 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),[8] 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.
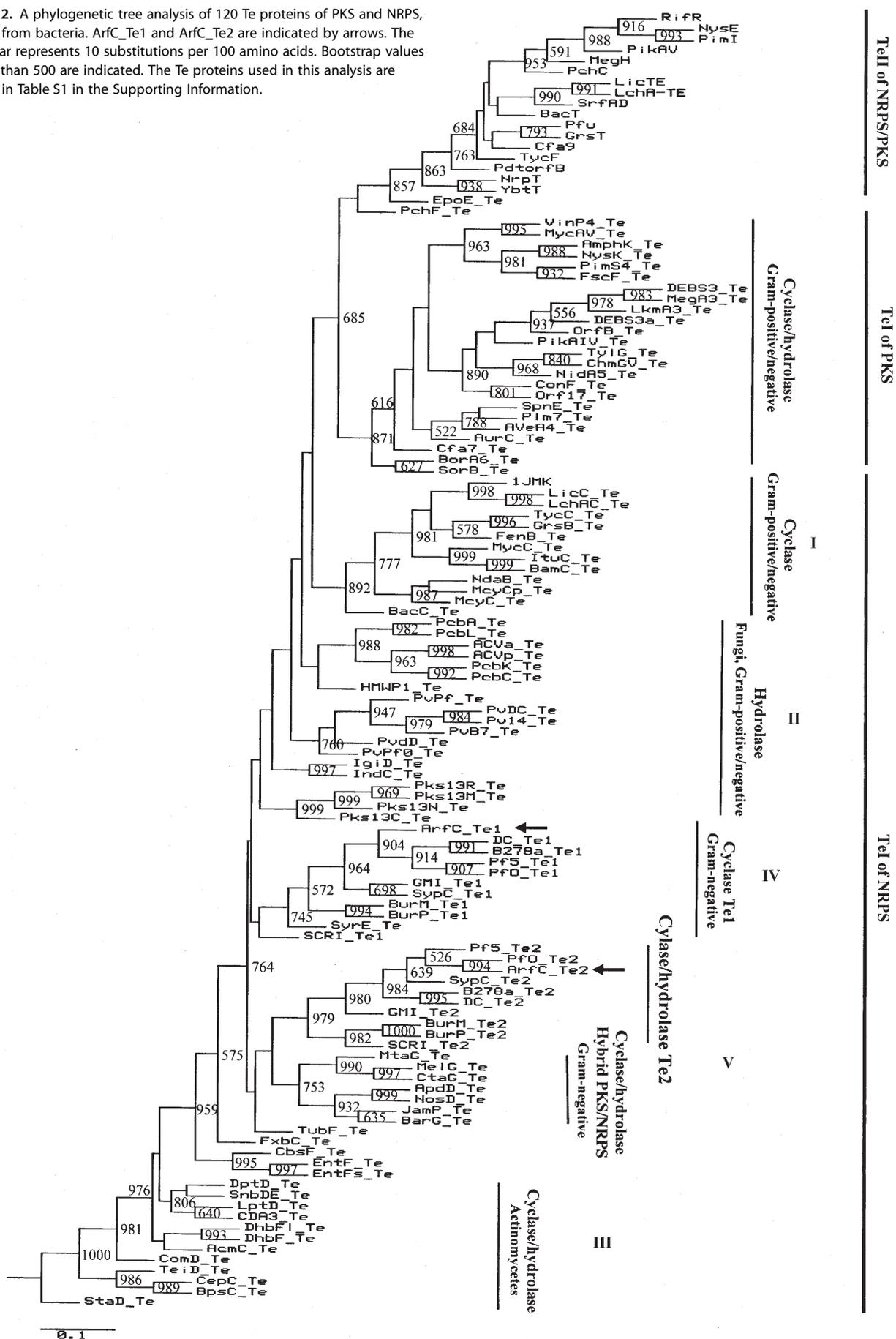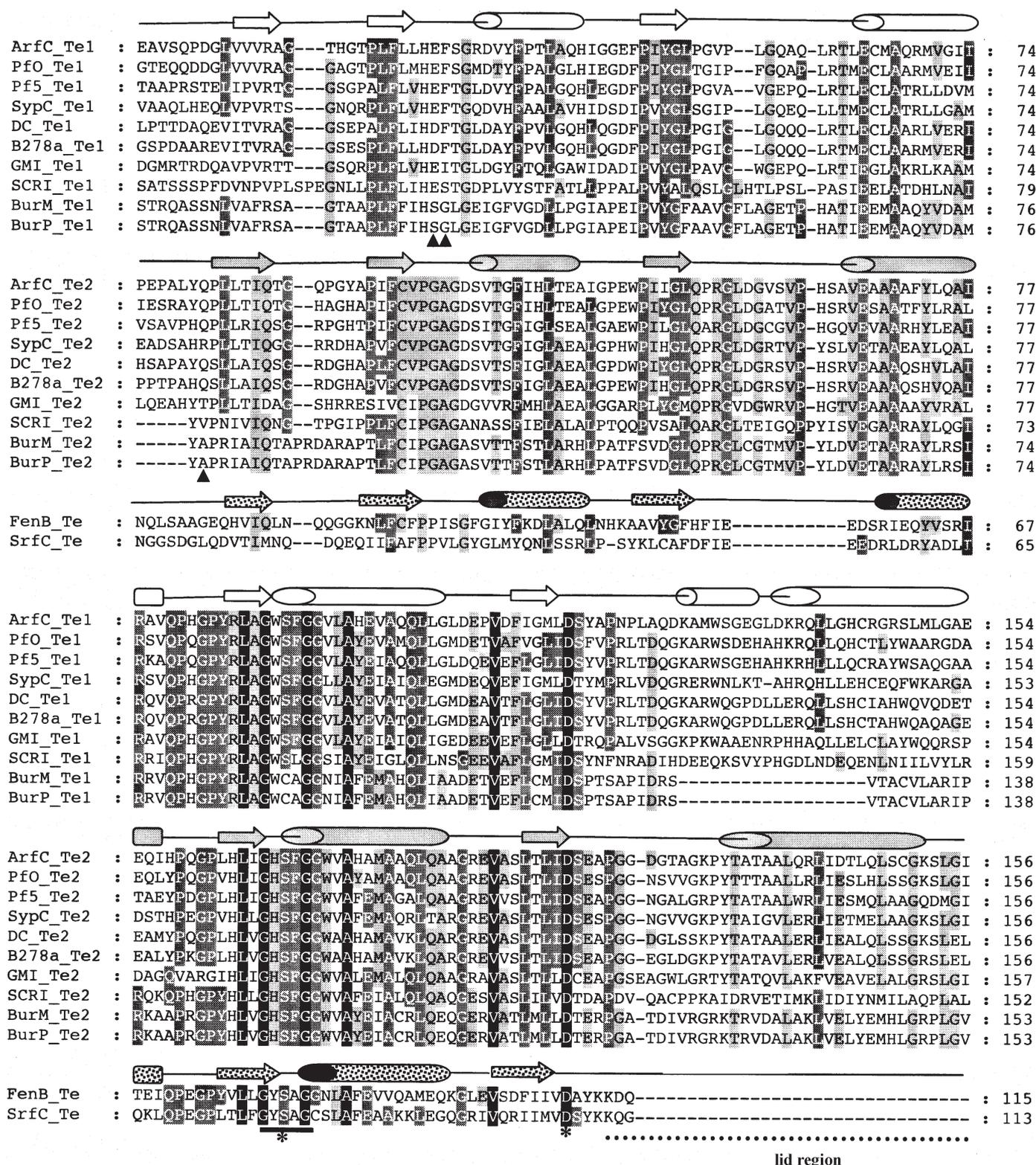
Figure 3. Amino acid sequence alignment of tandem C-terminal Te domains with FenB_Te and SrfC_Te (1 JMK). The sequences analyzed here include ArfC from *Pseudomonas* sp. MIS38 (BAC67 536), PfO from *P. fluorescens* PfO-1 (ZP_00265 375), Pf5 from *P. fluorescens* Pf-5 (AAY91 421), SypC from *P. syringae* pv. *syringae* B301D (AAO72 425), DC from *P. syringae* pv. *tomato* str. DC3000 (NP_792 634), B278a from *P. syringar* pv. *syringae* B728a (ZP_00205 846), GMI from *Ralstonia solanacearum* GMI1000 (NP_522 203), SCRI from *Erwinia carotovora* SCRI1043 (YP_049 592), BurM from *Burkholderia mallei* ATCC23 344 (YP_106 216), BurP from *B. pseudomallei* K96 243 (YP_111 640), FenB from *Bacillus subtilis* F29-3 (AAB00 093), and SrfC from *B. subtilis* 168 (Q08 787). The G*X*S*X*G 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.
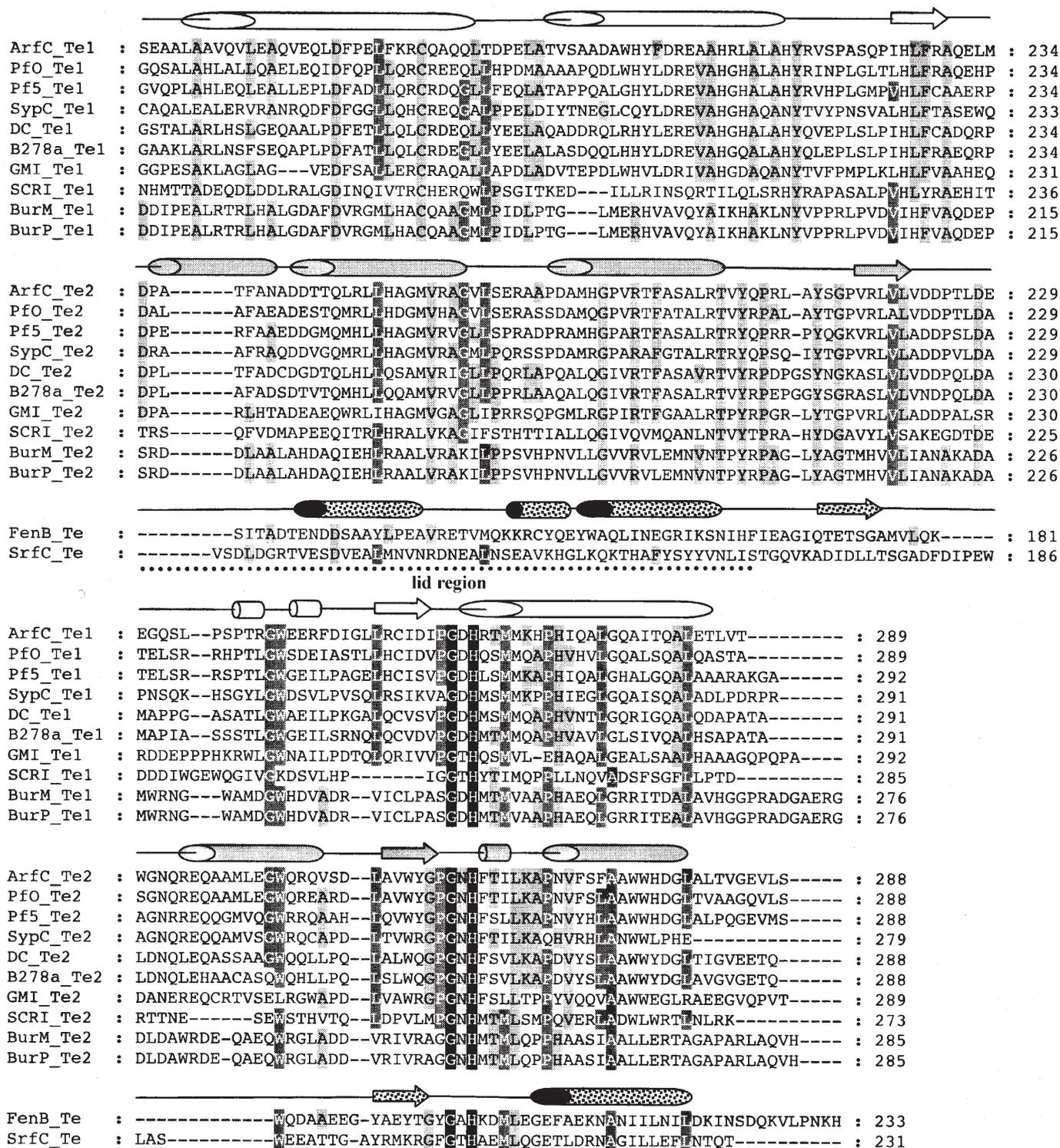
```
ArfC_Te1  : SEAALAAVQVLEAQVEQLDFPE FKRCQAQQLTDPELATVSAADAWHY DREAAHRLALAHYRVSPASQPIH RAQELM : 234
PfO_Te1   : GQSALAHLALLQAELEQIDFQP LQRCREEQL HPDMAAAAPQDLWHYLDREVAHGHALAHYRINPLGLTLH RAQEHP : 234
Pf5_Te1   : GVQPLAHLEQLEALLEPLDFAD LQRCRDQ L FEQLATAPPQALGHYLDREVAHGHALAHYRVHPLGMP HLFCAAERP : 234
SypC_Te1  : CAQALEALERVRANRQDF DFGG LQHCREQ A PPELDIYTNEGLCQYLDREVAHGHAQANYTVYPNSVALH FTASEWQ : 233
DC_Te1    : GSTALARLHSLGEQAALPDFET LQLCRDEQL YEELAQADDRQLRHYLEREVAHGHALAHYQVEPLSLPIH FCADQRP : 234
B278a_Te1 : GAAKLARLNSFSEQAPLPDFAT LQLCRDE L YEELALASDQQLHHYLDREVAHGQALAHYQLEPLSLPIH RAEQRP : 234
GMI_Te1   : GGPESAKLAGLAG---VEDFSA LERCRAQAL APDLADVTEPDLWHVLDRIVAHGDAQANYTVFPMPLKLH VAAHEQ : 231
SCRI_Te1  : NHMTTADEQDLDDLRALGDINQIVTRCHERQW PSGITKED---ILLRINSQRTILQLSRHYRAPASALP HLYRAEHIT : 236
BurM_Te1  : DDIPEALRTRLHALGDAFDVRGMLHACQAA M PIDLPTG---LMERHVAVQYAIKHAKLNYVPPRLPVD IHFVAQDEP : 215
BurP_Te1  : DDIPEALRTRLHALGDAFDVRGMLHACQAA M PIDLPTG---LMERHVAVQYAIKHAKLNYVPPRLPVD IHFVAQDEP : 215


ArfC_Te2  : DPA------TFANADDTTQLRL HAGMVRA VI SERAAPDAMHGPVRTFASALRTVYQPRL-AYSGPVRL LVDDPTLDE : 229
PfO_Te2   : DAL------AFAEADESTQMRL HDGMVHA VI SERASSDAMQGPVRTFATALRTVYRPAL-AYTGPVRLALVDDPTLDA : 229
Pf5_Te2   : DPE------RFAAEDDGMQMHL HAGMVRV LI SPRADPRAMHGPARTFASALRTRYQPRR-PYQGKVRL LADDPSLDA : 229
SypC_Te2  : DRA------AFRAQDDVGQMRL HAGMVRA M  PQRSSPDAMRGPARAFGTALRTRYQPSQ-IYTGPVRL LADDPVLDA : 229
DC_Te2    : DPL------TFADCDGDTQLHL QSAMVRI L PQRLAPQALQGIVRTFASAVRTVYRPDPGSYNGKASL LVDDPQLDA : 230
B278a_Te2 : DPL------AFADSDTVTQMHL QQAMVRV L PPRLAAQALQGIVRTFASALRTVYRPEPGGYSGRASL LVNDPQLDA : 230
GMI_Te2   : DPA------RLHTADEAEQWRLIHAGMVGA LI PRRSQPGMLRGPIRTFGAALRTPYRPGR-LYTGPVRL LADDPALSR : 230
SCRI_Te2  : TRS------QFVDMAPEEQITR HRALVKA IFSTHTTIALLQGIVQVMQANLNTVYTPRA-HYDGAVYL SAKEGDTDE : 225
BurM_Te2  : SRD------DLAALAHDAQIEH RAALVRAKI PPSVHPNVLLGVVRVLEMNVNTPYRPAG-LYAGTMHV LIANAKADA : 226
BurP_Te2  : SRD------DLAALAHDAQIEH RAALVRAKI PPSVHPNVLLGVVRVLEMNVNTPYRPAG-LYAGTMHV LIANAKADA : 226


FenB_Te   : ---------SITADTENDDSAAYLPEAVRETVMQKKRCYQEYWAQLINEGRIKSNIHFIEAGIQTETSGAMVLQK----- : 181
SrfC_Te   : -------VSDLDGRTVESDVEA MNVNRDNEA NSEAVKHGLKQKTHAFYSYYVNLISTGQVKADIDLLTSGADFDIPEW : 186
                                              lid region
```

```
ArfC_Te1  : EGQSL--PSPTRGWEERFDIGL RCIDI GDHRT MKH HIQA GQAITQA ETLVT--------- : 289
PfO_Te1   : TELSR--RHPTLGWSDEIASTL HCIDV GDHQS MQA HVHV GQALSQA QASTA--------- : 289
Pf5_Te1   : TELSR--RSPTLGWGEILPAGE HCISV GDHLS MKA HIQA GHALGQA AAARAKGA------ : 292
SypC_Te1  : PNSQK--HSGYLGWDSVLPVSQ RSIKVA GDHMS MKP HIEG GQAISQA ADLPDRPR------ : 291
DC_Te1    : MAPPG--ASATLGWAEILPKGA QCVSV GDHMS MQA HVNT GQRIGQA QDAPATA------- : 291
B278a_Te1 : MAPIA--SSSTLGWGEILSRNQ QCVDV GDHMT MQA HVAV GLSIVQA HSAPATA------- : 291
GMI_Te1   : RDDEPPPHKRWLGWNAILPDTQ QRIVV GTHQS VL-EHAQA GEALSAA HAAAGQPQPA---- : 292
SCRI_Te1  : DDDIWGEWQGIV KDSVLHP-------IGGTHYTIMQP LLNQV DSFSGF LPTD---------- : 285
BurM_Te1  : MWRNG---WAMDGWHDVADR--VICLPAS DHMT VAA HAEQ GRRITDA AVHGGPRADGAERG : 276
BurP_Te1  : MWRNG---WAMDGWHDVADR--VICLPAS DHMT VAA HAEQ GRRITEA AVHGGPRADGAERG : 276


ArfC_Te2  : WGNQREQAAMLEGTQRQVSD--LAVWYG GNHFTILKA NVFSF AWWHDG ALTVGEVLS----- : 288
PfO_Te2   : SGNQREQAAMLEGTQREARD--LAVWYG GNHFTILKA NVFS AWWHDG TVAAGQVLS----- : 288
Pf5_Te2   : AGNRREQQGMVQGWRRQAAH--LQVWYG GNHFSLLKA NVYH LAWWHDG ALPQGEVMS----- : 288
SypC_Te2  : AGNQREQQAMVS WRQCAPD--LTVWRG GNHFTILKAQHVRH NWWLPHE-------------- : 279
DC_Te2    : LDNQLEQASSAA QQLLPQ--LALWQG GNHFSVLKA DVYS AWWYDG TIGVEETQ------ : 288
B278a_Te2 : LDNQLEHAACAS QHLLPQ--LSLWQG GNHFSVLKA DVYS AWWYDG AVGVGETQ------ : 288
GMI_Te2   : DANEREQCRTVSELRGWAPD--LVAWRG GNHFSLLTP YVQQV AWWEGLRAEEGVQPVT----- : 289
SCRI_Te2  : RTTNE------SEWSTHVTQ--LDPVLM GNHMT LSM QVER ADWLRT NLRK---------- : 273
BurM_Te2  : DLDAWRDE-QAEQWRGLADD--VRIVRAGGNHMT LQP HAASI AALLERTAGAPARLAQVH---- : 285
BurP_Te2  : DLDAWRDE-QAEQWRGLADD--VRIVRAGGNHMT LQP HAASI AALLERTAGAPARLAQVH---- : 285


FenB_Te   : --------------WQDAAEEG-YAEYTGYGAHKD LEGEFAEKN NIILNI DKINSDQKVLPNKH : 233
SrfC_Te   : LAS----------WEEATTG-AYRMKRGFGTHAE LQGETLDRN GILLEF NTQT---------- : 231
                                      *                    .
```

Figure 3 (continued).

cyclic macrolactam products such as tyrosidine, bacillomycin, microcystin and bacitracin (Tyc/BamC/Mcy/BacC_Te).[13] 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.[14] Additionally, this hydrolase-type Te is also found in pyoverdine synthetase (Pv_Te) from Gram-negative *Pseudomonas* species.[15] 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.[16] The Te of subclass III hydrolyzes linear

peptide precursors of vancomycin-type antibiotics (BpsC/CepC/TeiD/StaD/ComD_Te)[17] or cyclizes calcium-dependent antibiotics (CAD3/DptD_Te).[18] Interestingly, the iterative cyclases of *E. coli* or *Samonella* sp. enterobactin (EntF) and *Bacillus bacillibactin* (DhbF), an aryl cap siderophores are closely related to subclass III.[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.[4, 20] 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 by 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/Pf5/SyrE_Te).[4, 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.[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.[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 NRPS and PKS forms a distinctly separate branch. TeI of PKS forms a cluster that is different from TeI 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.[24] Like SrfC_Te and FenB_Te, ArfC_Te1 and ArfC_Te2 consist of a seven-stranded β-sheet.[7, 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.[7] 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 G*X*S*X*G 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 respectively. A highly conserved Asp118 in ArfC_Te2 was also replaced by alanine to give 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 phase 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[4] (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 (C_{10}, $m/z = 1354.9$) and its derivatives (C_9 and C_{12}) were found in the sample from MIS38 (total amount 220 ± 3.6 mg L^{-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^{-1}), ArfC_Te2:S92A/D118A (12.5 ± 1 mg L^{-1}), and ArfCΔTe2 (13.4 ± 4 mg L^{-1}) was reduced by 95% without alteration of the cyclic lipoundecapeptide 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 TeI domain in surfactin synthetase, and the serine-to-alanine site-directed mutagenesis in fungal ACV synthetase also retained a slight but significant activity.[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 TeI, and that ArfC_Te1 functions as the last acceptor of linear peptide intermediates, like the last T domain located before TeI. Meanwhile, less reduction of surfactin (84%) was observed in the external TeII mutant.[26] Disruption of the external TeII in a modular PKS also resulted in a moderate drop (20–85%) in polyketide production.[28] A drastic reduction of arthro-
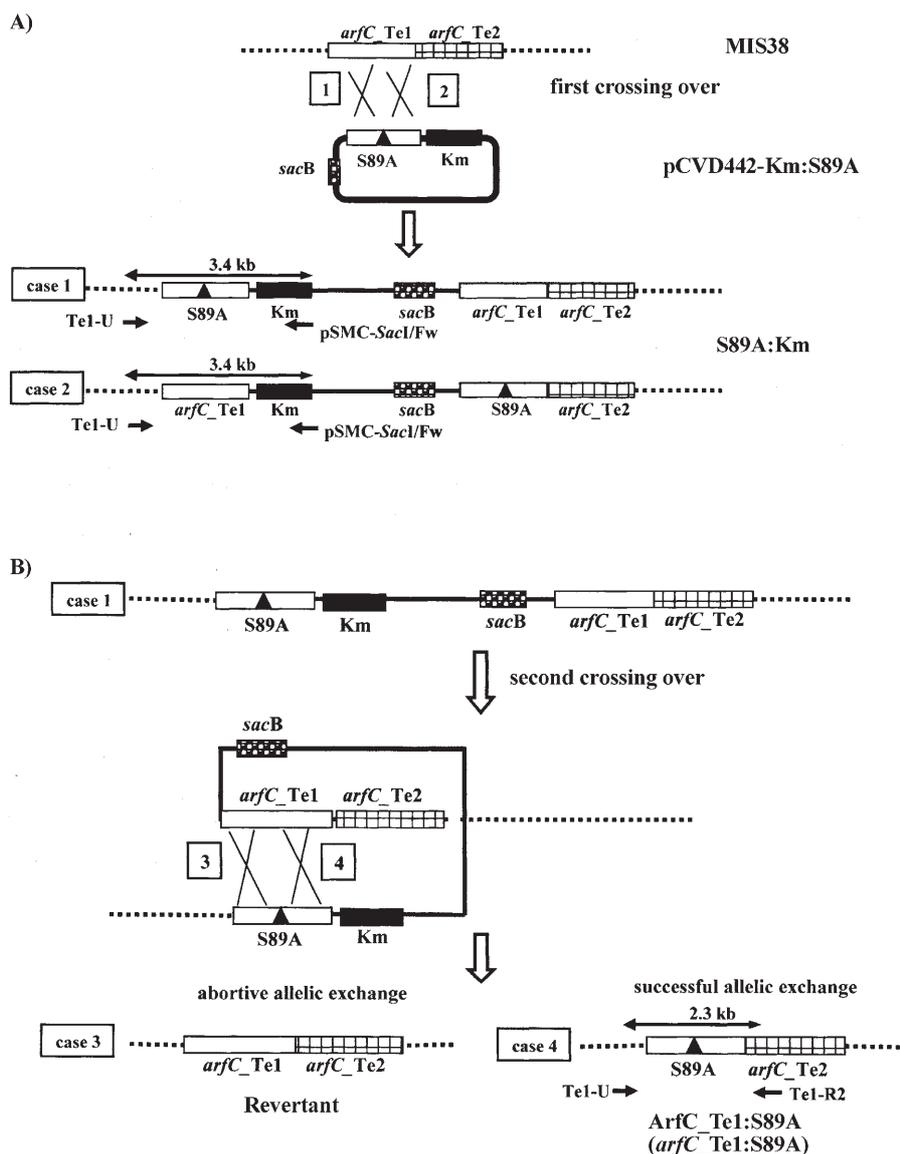
Figure 4. Strategy for the site-directed mutagenesis (S89A) in *arfC*_Te1. A) First crossing-over event. The first crossing-over can occur on either side of the mutation point (case 1 or 2). Amplification of the *arfC*_Te1 flanking region in a kanamycin resistant colony (S89A:Km) was confirmed by the PCR method by using Te1-U and pSMC-*Sac*I/Fw primers. B) Second crossing-over event. The second crossing-over is shown only for case 1. Recombination on either side of the mutation point (case 3 or 4) resulted in either an abortive or successful allelic exchange. The DNA flanking region in *arfC*_Te1:S89A was amplified by the PCR method by using Te1-U and Te1-R2 primers.

factin production in ArfC_Te2 mutants supports the idea that ArfC_Te2 is functionally different from the external Te11.[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 P26V27 (similar to SrfC_Te) and G26A27 (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 ($2.2 \pm 1\ \mathrm{mg\,L^{-1}}$), 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%) ar-

throfactin 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 T_{11} 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).
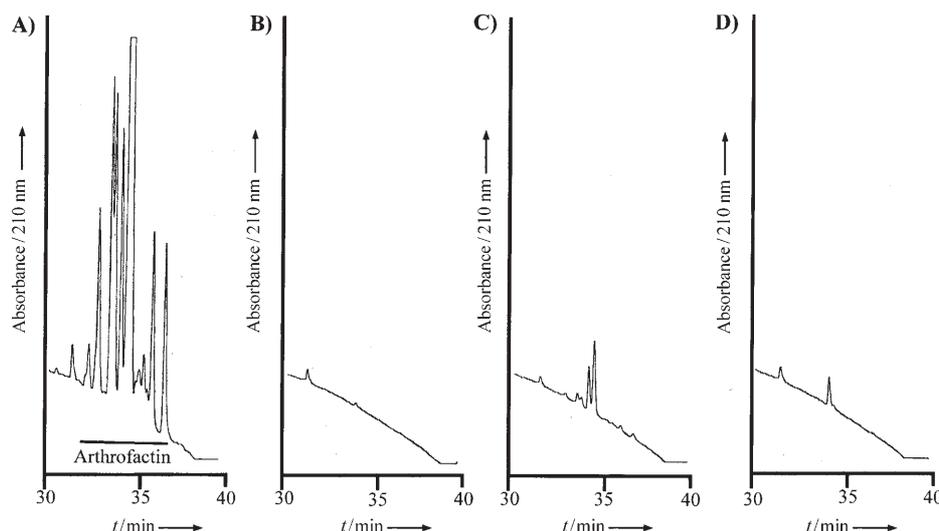
Figure 5. HPLC–UV analysis of methanol extracts from acid precipitates. A) MIS38, B) ArfC_Te1:S89A, C) ArfC_Te2:S92A, and D) ArfC_Te1:E26P/F27V. No production of arthrofactin was observed for NC1, ArfC_Te1:S89A, ArfC_Te1:S89T, and ArfC_Te1:E26G/F27A, then only HPLC–UV analysis of ArfC_Te1:S89A is shown. Similarly, the productivity of arthrofactin was reduced by 95% for ArfC_Te2:S92A, ArfC_Te2:S92A/D118A, and ArfCΔTe2, then only HPLC–UV analysis of ArfC_Te2:S92A is shown.

Second, the lipoundecapeptidyl chain is further transferred onto an invariant serine residue (Ser92) in the active site of the ArfC_Te2, which is activated by Asp118 and His259. In the deacylation step, the hydroxyl group of peptide threonine forms the lactone by an intramolecular nucleophilic attack on the acyl-enzyme ester bond. The peptidyl chain could not be transferred to ArfC_Te2 when Ser92 was changed to Ala, however, the cyclization of peptide–O–Te1 intermediate could still occur by attack of the hydroxyl group of the peptide threonine (Figure 7B). This mutation resulted in an inefficient production of arthrofactin. On the other hand, no arthrofactin was produced in ArfC_Te1:S89A (Figure 7C) at all. This should be because no peptidyl intermediate was transferred from the last T_{11} domain to ArfC_Te1. Direct transfer of peptidyl intermediates to the active site of ArfC_Te2 might not have happened due to a bulky ArfC_Te1 domain. We do not know why the autonomous cyclization did not occur in the peptide–S–T_{11} intermediate as it did in surfactin and ACV synthetase. The difference in the length of peptide chain and/or the position of lactone formation between arthrofactin and surfactin could explain this phenomenon. The exact functions of ArfC_Te1 and ArfC_Te2 still remain to be clarified. Recently, electrospray ionization Fourier-Transform mass spectrometry (ESI-FTMS) has been used to investigate the NRPS and PKS systems.[29] ESI-FTMS can be used to understand the substrate tolerance, the timing of covalent linkages, the timing of tailoring reactions and the transfer of substrates and biosynthetic intermediates from domain to domain. This technique might be able to take a snapshot of the peptidyl-transfer from the T to the ArfC_Te1 domain, and from the ArfC_Te1 to the ArfC_Te2 domain, and would help to clarify these reactions in even more detail.

Experimental Section

Bacterial strains and plasmids: Arthrofactin-producing *Pseudomonas* sp. MIS38 was previously isolated from oil spills in Shizuoka prefecture, Japan.[3] 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.[4] *E. coli* DH5α was used as a host strain for the construction of recombinant plasmids. *E. coli* SM10λpir[30] was used for transforming MIS38 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).[31] 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.[30, 32]

General DNA manipulations: Genomic DNA of MIS38 was prepared by using the Sarkosyl method and was purified by CsCl–ethidium bromide equilibrium density gradient ultracentrifugation.[33] DNA fragments were recovered from an agarose gel by using the QIAquick Gel Extraction Kit (QIAGEN). The large-scale preparation of plasmid DNA was done by using a Qiagen plasmid Maxi Kit (QIAGEN). All other DNA manipulations were performed according to standard protocols.[33] PCR was performed in 30 cycles by using a thermal cycler, the Takara Dice Standard (Takara Bio, Ohtsu, Japan), and *ExTaq* (Takara Bio) or KOD plus DNA polymerase (Toyobo, Osaka, Japan). Oligodeoxyribonucleotides for PCR primers were synthesized at Hokkaido System Science (Sapporo, Japan). The nucleotide sequences of the gene fragments were determined by using the dideoxy-chain termination method with the ABI Prism BigDye terminator v3.1 cycle sequencing kit and the autosequencer ABI Prism 3100 (Applied Biosystems, Foster City, CA).

Phylogenetic analysis of C-terminal Te domain and external TeII: The amino acid sequences of Te proteins in various PKS and NRPS were retrieved from publicly accessible databases (http://www.ncbi.nlm.nih.gov/entrez/). The sequences of Te proteins were aligned by the ClustalW program[34] provided by the DNA Data Bank of Japan, DDBJ. Phylogenetic trees were constructed by using the distance method and the character-based method from the PHYLIP package v3.6[35] as described previously.[6] Both methods gave similar tree topology, but only the tree that was constructed by the distance method is shown in this paper.

Construction of pCVD442-Km: The suicide vector pCVD442 carries the *bla* gene, which confers resistance to amplicillin (Amp) however, this selectable marker was found to be useless due to the high tolerance of *Pseudomonas* sp. MIS38 to Amp. Therefore, we introduced the Km gene from plasmid pSMC32 into *Sac*I site of pCVD442 as follows. The Km gene fragment, including its promoter was amplified by the PCR method by using vector pSMC32 as a template. The oligonucleotide primers pSMC-*Sac*I/Fw and pSMC-*Sac*I/Rv which contain the *Sac*I restriction sites (underlined) were
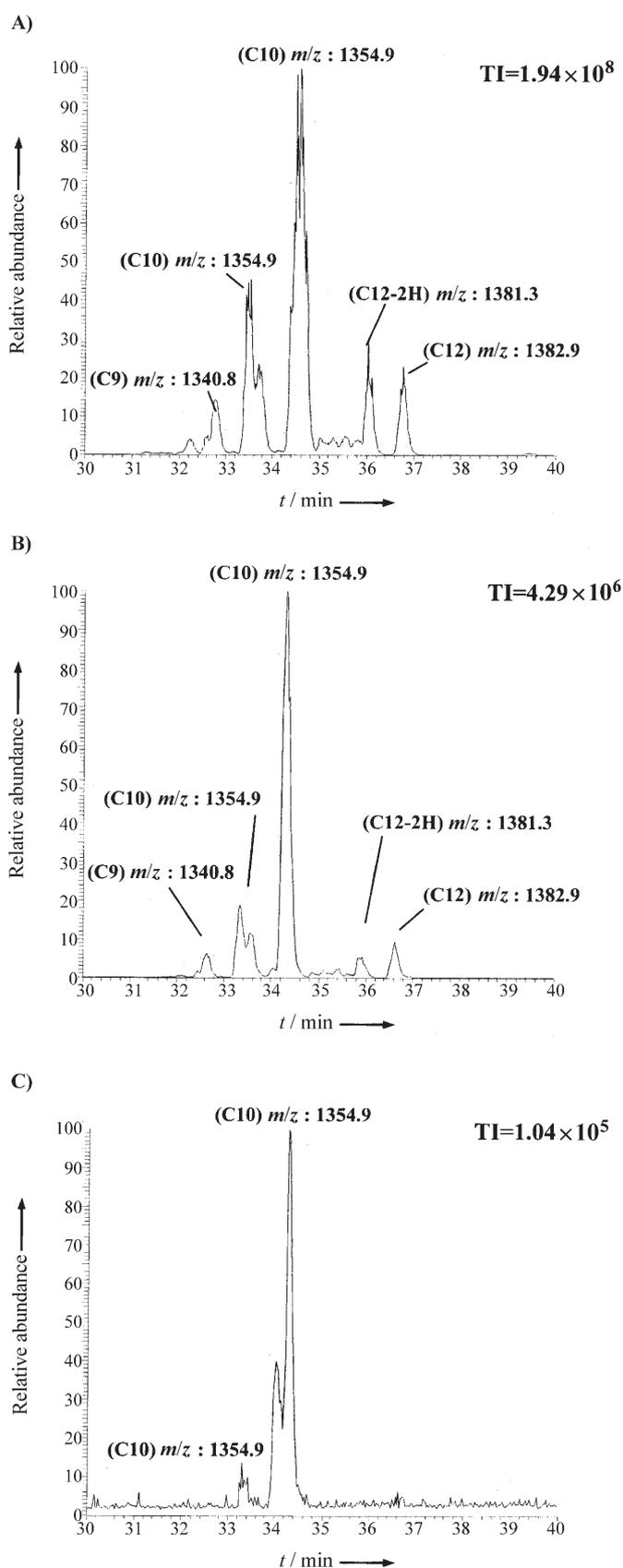
A)

B)

C)

Figure 6. LC-MS analysis of methanol extracts from acid precipitates. A) MIS38, B) ArfC_Te2:S92A, and C) ArfC_Te1:E26P/F27V. ArfC_Te2:S92A, ArfC_Te2:S92A/D118A, and ArfCΔTe2 gave similar result, then only the LC-MS analysis of ArfC_Te2:S92A is shown. TI = total ions.

used for PCR as shown in Table 1. The PCR products were first cloned into pGEM-T Easy vector, then the *Sac*I 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* SM10λpir by an electrotransformation method as follows and was subsequently used for different kinds of allelic exchanges.

Electrotransformation of *E. coli* SM10λpir: Cells were grown in L-broth until the mid-log phase (OD$_{600}$ ≈ 0.4). After collection by centrifugation (5000 g for 15 min at 4 °C), the cells were washed once with ice-cold pure H$_2$O. Then, the cells were washed twice with glycerol (10 %), and resuspended in glycerol (10 %) at 3×10^{10} 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^{-1} 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^{-1}).

Cloning of native *arfC*_Te1 and *arfC*_Te2 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.[32] Therefore, the native 2kb *arfC*_Te1 and *arfC*_Te2 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-*Xba*I/Fw and Te1-*Xba*I/Rv for the *arfC*_Te1 gene, and Te2-*Xba*I/Fw and Te2-*Xba*I/Rv for the *arfC*_Te2 gene, which contained the *Xba*I 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_Te1 and ArfC_Te2: The *arfC*_Te1 and *arfC*_Te2 genes were mutagenised by the overlap extension method.[33] 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-*Xba*I/Fw or Te2-*Xba*I/Fw and mutation-Rv primers, and the 3′-fragment was amplified by using the mutation-Fw and Te1-*Xba*I/Rv or Te2-*Xba*I/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-*Xba*I/Fw or Te2-*Xba*I/Fw primers and Te1-*Xba*I/Rv or Te2-*Xba*I/Rv primers. The blunt-ended PCR product was first cloned into pUC18 at the *Sma*I site and then the *Xba*I fragment was excised and ligated into the *Xba*I 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* SM10λpir 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$_{600}$ 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
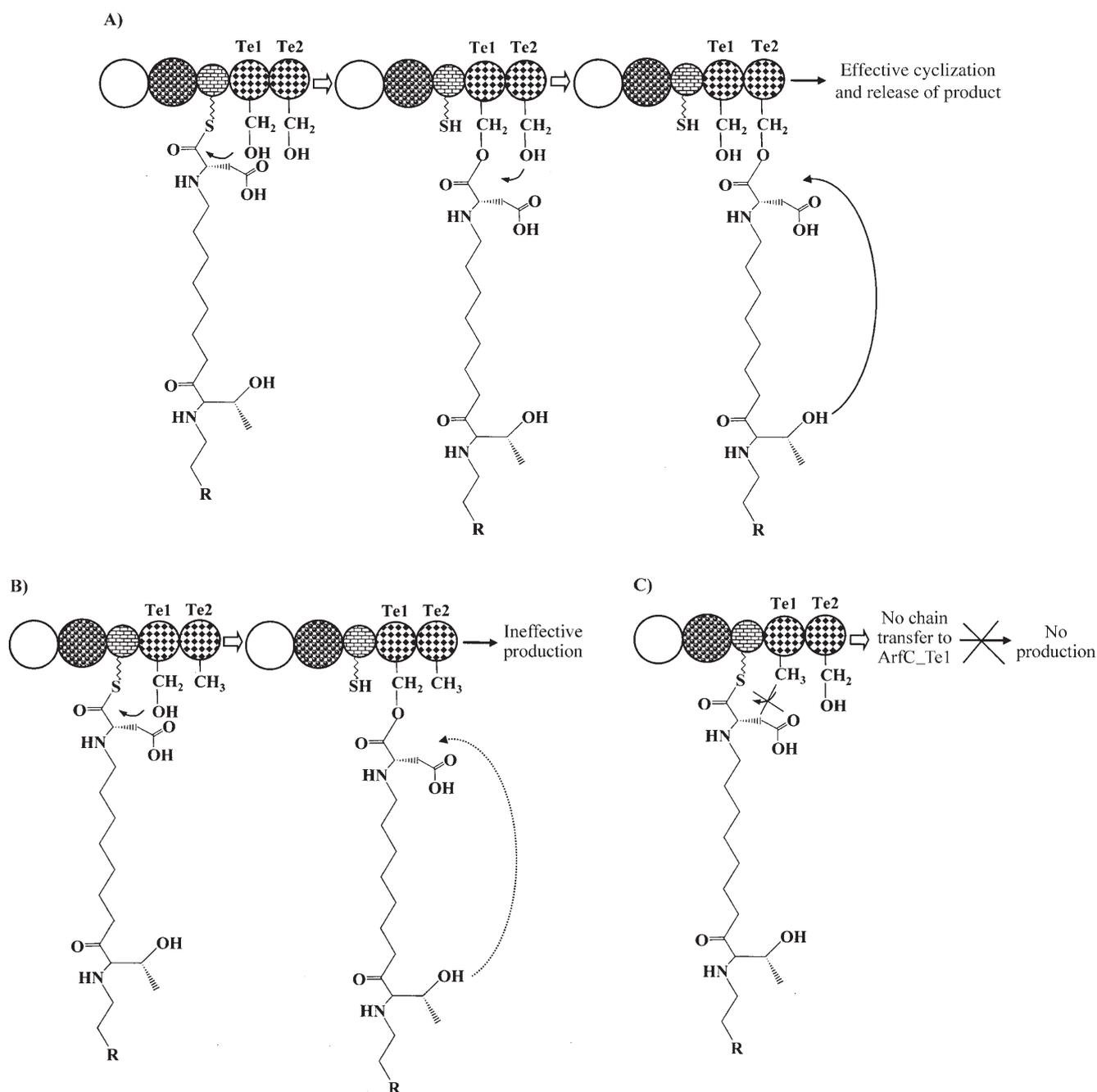
Figure 7. Proposed mechanism of ArfC_Te1 and ArfC_Te2. A) MIS38, B) ArfC_Te2:S92A, C) ArfC_Te1:S89A. The side chain of the potential nucleophiles of ArfC_Te1:S89 and ArfC_Te2:S92 are represented by -CH$_2$–OH whereas -CH$_3$ represents the side chain of alanine. Peptidyl chain transfer and the subsequent 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.

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 Cm^r and Km^r 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 previ-

ously to improve the sucrose counterselection.[36] The presence of the *sac*B 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 *arf*C_Te1 or *arf*C_Te2 by cloning and sequencing the target gene locus. Primers for amplifying the gene from *arf*C_Te1 mutants are

Table 1. Primers used in this study.[a]

Name	Sequences
pSMC-*Sac*I/Fw	5′-CATG<u>GAGCTC</u>GTTTTATGGACAGCAAGCGA
pSMC-*Sac*I/Rv	5′-CATG<u>GAGCTC</u>CCGTCAGTAGCTGAACAGGA
Te1-*Xba*I/Fw	5′-CATG<u>TCTAGA</u>TGAGCAACACTCGGCTGTAC
Te1-*Xba*I/Rv	5′-CATG<u>TCTAGA</u>TTGCCACAGGACAACTGCAG
Te2-*Xba*I/Fw	5′-CATG<u>TCTAGA</u>GTGGCGAGTTCCCGATTTAC
Te2-*Xba*I/Rv	5′-CATG<u>TCTAGA</u> ATCTCTTTGGTCTGCTTGAG
S89A-Fw	5′-TGGCGGGCTGG*GCA*TTCGGCGGGGT
S89A-Rv	5′-ACCCCGCCGAA*TGC*CCAGCCCGCCA
S89T-Fw	5′-TGGCGGGCTGG*ACG*TTCGGCGGGGT
S89T-Rv	5′-ACCCCGCCGAA*CGT*CCAGCCCGCCA
S92A-Fw	5′-CTGATCGGCCAT*GCA*TTCGGCGGCT
S92A-Rv	5′-AGCCGCCGAA*TGC*ATGGCCGATCAG
D118A-Fw	5′-CTGACCTTGATCG*CC*AGCGAGGCACCGGGC
D118A-Rv	5′-GCCCGGTGCCTCGCT*G*GCGATCAAGGTCAG
E26P/F27V-Fw	5′-TCCTGCTGCAT*CCGG*TCAGCGGCAGGGAC
E26P/F27V-Rv	5′-GTCCCTGCCGCTGA*CCGG*ATGCAGCAGGA
E26G/F27A-Fw	5′-TCCTGCTGCATG*GTGCC*AGCGGCAGGGAC
E26G/F27A-Rv	5′-GTCCCTGCCGCT*GGCAC*CATGCAGCAGGA
ArfCΔTe2-Fw	5′-ACCGGCGCTGTATT*TG*ACCGCTGCTGACGA
ArfCΔTe2-Rv	5′-TCGTCAGCAGCGGT*CA*ATACAGCGCCGGT
Te1-U	5′-CACCAGCCTGACCGATGTGCTCAAC
Te1-R2	5′-GCAGCAGTCGCAGTTGCGTGGTGTC
Te2-F	5′-TCGCTGGCCGAACTGTTCCAGCATC
Te2-R2	5′-TGATCTGCGCATCCAGCGACAGCAG

[a] Introduced mutations are bold and italicized.

Te1-U and Te1-R2; the Te2-F and Te2-R2 primers were used for *arfC*_Te2 mutants (Figure 4B, Table 1).

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.[4] 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 washed 3 times with dilute HCl (pH 2.0). Biosurfactant-containing lipophilic substances were extracted from the precipitates three times with methanol, 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 5C$_{18}$AR 4.6×150 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 total arthrofactin were calculated from the peak area and by weighing the methanol extracts of the acid precipitates.

Acknowledgements

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

[1] R. M. Kohli, J. W. Trauger, D. Schwarzer, M. A. Marahiel, C. T. Walsh, *Biochemistry* **2001**, *40*, 7099–7108.

[2] J. Grunewald, M. A. Marahiel, *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 121–146.

[3] M. Morikawa, H. Daido, T. Takao, S. Murata, Y. Shimonishi, T. Imanaka, *J. Bacteriol.* **1993**, *175*, 6459–6466.

[4] N. Roongsawang, K. Hase, M. Haruki, T. Imanaka, M. Morikawa, S. Kanaya, *Chem. Biol.* **2003**, *10*, 869–880.

[5] C. J. Balibar, F. H. Vaillancourt, C. T. Walsh, *Chem. Biol.* **2005**, *12*, 1189–1200.

[6] N. Roongsawang, S. P. Lim, K. Washio, K. Takano, S. Kanaya, M. Morikawa, *FEMS Microbiol. Lett.* **2005**, *252*, 143–151.

[7] S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh, M. T. Stubbs, *Structure* **2002**, *10*, 301–310.

[8] C. C. Tseng, S. D. Bruner, R. M. Kohli, M. A. Marahiel, C. T. Walsh, S. A. Sieber, *Biochemistry* **2002**, *41*, 13350–13359.

[9] U. Linne, D. Schwarzer, G. N. Schroeder, M. A. Marahiel, *Eur. J. Biochem.* **2004**, *271*, 1536–1545.

[10] D. Schwarzer, H. D. Mootz, U. Linne, M. A. Marahiel, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14083–14088.

[11] E. Yeh, R. M. Kohli, S. D. Bruner, C. T. Walsh, *ChemBioChem* **2004**, *5*, 1290–1293.

[12] T. A. Keating, D. E. Ehmann, R. M. Kohli, C. G. Marshall, J. W. Trauger, C. T. Walsh, *ChemBioChem* **2001**, *2*, 99–107.

[13] J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel, C. T. Walsh, *Nature* **2000**, *407*, 215–218.

[14] A. P. MacCabe, H. van Liempt, H. Palissa, S. E. Unkles, M. B. Riach, E. Pfeifer, H. von Dohren, J. R. Kinghorn, *J. Biol. Chem.* **1991**, *266*, 12646–12654.

[15] J. Ravel, P. Cornelis, *Trends Microbiol.* **2003**, *11*, 195–200.

[16] D. Portevin, C. De Sousa-D'Auria, C. Houssin, C. Grimaldi, M. Chami, M. Daffe, C. Guilhot, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 314–319.

[17] H. T. Chiu, B. K. Hubbard, A. N. Shah, J. Eide, R. A. Fredenburg, C. T. Walsh, C. Khosla, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8548–8553.

[18] J. Grunewald, S. A. Sieber, M. A. Marahiel, *Biochemistry* **2004**, *43*, 2915–2925.

[19] J. J. May, T. M. Wendrich, M. A. Marahiel, *J. Biol. Chem.* **2001**, *276*, 7209–7217.

[20] B. K. Scholz-Schroeder, J. D. Soule, D. C. Gross, *Mol. Plant-Microbe Interact.* **2003**, *16*, 271–280.

[21] S. A. Sieber, J. Tao, C. T. Walsh, M. A. Marahiel, *Angew. Chem.* **2004**, *116*, 499–504; *Angew. Chem. Int. Ed.* **2004**, *43*, 493–498.

[22] D. Hoffmann, J. M. Hevel, R. E. Moore, B. S. Moore, *Gene* **2003**, *311*, 171–180.

[23] B. Silakowski, H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blocker, G. Hofle, S. Beyer, R. Muller, *J. Biol. Chem.* **1999**, *274*, 37391–37399.

[24] K. Bryson, L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi, D. T. Jones, *Nucleic Acids Res.* **2005**, *33* (Web Server issue), W36–38.

[25] S. A. Samel, B. Wagner, M. A. Marahiel, L. O. Essen, *J. Mol. Biol.* **2006**, *359*, 876–889.

[26] A, Schneider, M. A. Marahiel, *Arch. Microbiol.* **1998**, *169*, 404–410.

[27] W. Kallow, J. Kennedy, B. Arezi, G. Turner, H. von Dohren, *J. Mol. Biol.* **2000**, *297*, 395–408.

[28] B. M. Harvey, H. Hong, M. A. Jones, Z. A. Hughes-Thomas, R. M. Goss, M. L. Heathcote, V. M. Bolanos-Garcia, W. Kroutil, J. Staunton, P. F. Leadlay, J. B. Spencer, *ChemBioChem* **2006**, *7*, 1435–1442; and references therein.

[29] P. C. Dorrestein, N. L. Kelleher, *Nat. Prod. Rep.* **2006**, *23*, 893–918.

CHEMBIOCHEM

M. Morikawa et al.

[30] M. S. Donnenberg, J. B. Kaper, *Infect. Immun.* **1991**, *59*, 4310–4317.

[31] B. Bartolome, Y. Jubete, E. Martinez, F. de La Cruz, *Gene* **1991**, *102*, 75–78.

[32] N. Philippe, J. P. Alcaraz, E. Coursange, J. Geiselmann, D. Schneider, *Plasmid* **2004**, *51*, 246–255.

[33] J. Sambrook, D. W. Russell, *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press, New York, **2001**.

[34] J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **1994**, *22*, 4673–4680.

[35] J. Felsenstein, *Methods Enzymol.* **1996**, *266*, 418–427.

[36] I. C. Blomfield, V. Vaughn, R. F. Rest, B. I. Eisenstein, *Mol. Microbiol.* **1991**, *5*, 1447–1457.