Chemistry & Biology, Vol. 10, 869–880, September, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2003.09.004

Cloning and Characterization of the Gene Cluster Encoding Arthrofactin Synthetase from *Pseudomonas* sp. MIS38

Niran Roongsawang,[1] Ken-ichi Hase,[1] Mitsuru Haruki,[1,3] Tadayuki Imanaka,[2] Masaaki Morikawa,[1,*] and Shigenori Kanaya[1]
[1]Department of Material and Life Science
Graduate School of Engineering
Osaka University
Osaka 565-0871
[2]Department of Synthetic Chemistry and Biological Chemistry
Graduate School of Engineering
Kyoto University
Kyoto 606-8501
Japan

Summary

Arthrofactin is a potent cyclic lipopeptide-type biosurfactant produced by *Pseudomonas* sp. MIS38. In this work, an arthrofactin synthetase gene cluster (*arf*) spanning 38.7 kb was cloned and characterized. Three genes termed *arfA*, *arfB*, and *arfC* encode ArfA, ArfB, and ArfC, containing two, four, and five functional modules, respectively. Each module bears condensation, adenylation, and thiolation domains, like other nonribosomal peptide synthetases. However, unlike most of them, none of the 11 modules possess the epimerization domain responsible for the conversion of amino acid residues from L to D form. Possible L- and D-Leu adenylation domains specifically recognized only L-Leu. Moreover, two thioesterase domains are tandemly located at the C-terminal end of ArfC. These results suggest that ArfA, ArfB, and ArfC assemble to form a unique structure. Gene disruption of *arfB* impaired arthrofactin production, reduced swarming activity, and enhanced biofilm formation.

Introduction

Arthrofactin is a cyclic lipoundecapeptide (Figure 1A) produced by *Pseudomonas* sp. MIS38, which was initially identified as *Arthrobacter* sp. [1, 2]. This lipoundecapeptide is one of the most effective cyclic lipopeptide biosurfactants ever reported and reduces the surface tension of water from 72 to 24 mN/m with a critical micelle concentration of 1×10^{-5} M [1]. An oil displacement assay demonstrated that arthrofactin was twice as effective as surfactin, which is produced by *Bacillus subtilis* [3] and *B. pumilus* [4]. Recently, several arthrofactin-like cyclic lipoundecapeptides with remarkable biosurfactant and antifungal properties (amphisin, lokisin, tensin [5, 6]) or an enzyme inhibitor (pholipeptin [7]) have been isolated from *Pseudomonas* spp. All of them contain 11 amino acid residues in the cyclic peptide structure and a β-hydroxydecanoyl moiety. Lokisin shares the most similar structure with arthrofactin, except for the ninth amino acid residue, which is replaced by L-Leu in lokisin.

Bacillus spp. also produces a variety of cyclic oligopeptides that are modified by a fatty acid like arthrofactin. Even a single *B. subtilis* strain has the ability to produce three kinds of lipopeptides, surfactin (a biosurfactant), bacillomycin (an antibiotic), and plipastatin (an enzyme inhibitor) [8]. Surfactin and plipastatin contain a β-hydroxy fatty acid group like arthrofactin, whereas members of the iturin family, such as bacillomycin, mycosubtilin, and iturin, carry a β-amino acid modification. The genes encoding multimodular nonribosomal peptide synthetases (NRPSs) for surfactin [9], lichenysin [10, 11], fengycin [12], mycosubtilin [13], and iturin A [14] have been cloned and characterized in *Bacillus*. NRPSs recognize, activate, modify, and link the amino acid intermediates to the product [15]. They are capable of synthesizing peptides that contain unusual amino acid residues. These include D amino acids, β amino acids, and hydroxy- or *N*-methylated residues. The modules are usually colinear to the product peptide sequences. Each module of the NRPSs 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) or peptidyl carrier protein (PCP) domain is the attachment site of 4′-phosphopantetheine cofactor and serves as a carrier of thioesterified amino acid intermediates. The condensation (C) domain catalyzes peptide bond formation of two consecutively bound amino acids. Modifying domains such as the epimerization (E) domain catalyze the conversion of L amino acids to D isomers. At last, cyclization and a release of the peptide product are catalyzed by the C-terminal thioesterase (Te) domain.

The Te domains of NRPSs and polyketide synthases (PKSs) could be classified into two groups, internal and external ones. Most NRPSs and PKSs have only one internal Te domain at the C terminus of the last module. This internal Te domain (type I) carries a typical GXSXG (X: any amino acid residue) sequence motif with highly conserved Asp and His residues at approximately 27 and 130 amino acids downstream of the C-terminal end, respectively. But some of them lack this conserved His residue [16]. The initial function of type I Te domains involves the acceptance of the linear peptide from the terminal T domain to form a peptide-O-Te intermediate. Concomitant deacylation of the intermediate by this type I Te domain results in either hydrolysis or intramolecular cyclization of a linear product [17, 18]. Surfactin and gramicidin, a cyclic peptide, synthetases have an external Te domain (type II) in addition to the internal one (type I). This type II Te domain shows sequence homology to the type II fatty acid Te domain of vertebrate origin. The protein also contains a GXSXG sequence motif and highly conserved Asp and His residues [16]. Recently, this type II Te domain has been shown to play a physiological role in deblocking acylated 4′-phosphopantetheine cofactors and thereby regenerating the functionality of NRPSs [19].

*Correspondence: morikawa@mls.eng.osaka-u.ac.jp
[3]Present address: Department of Materials Chemistry and Engineering, College of Engineering, Nihon University, Fukushima 963-8642, Japan.

Figure 1. Structure of *arf* and Its Products

(A) The primary structure of arthrofactin. This lipoundecapeptide is cyclic in nature and has a lactone structure between the carboxyl group of the C-terminal L-Asp and either the β-hydroxyl group of the fatty acid portion or the hydroxyl group of D-Thr [1, 6].

(B) The λ clones isolated in this work and the position of the walking probes for screening the λ-EMBL3 genomic library are marked. Restriction enzyme sites for subcloning are also indicated.

(C) Modular organization of *arf* and additionally identified *ORFs*. Three genes, *arfA*, *arfB*, and *arfC*, encode NRPSs ArfA, ArfB, and ArfC, respectively. The domain organizations of the NRPSs are illustrated within the individual *arf* genes. The position of the *kan* gene insert in the mutant strain NC1 is also shown. Mod1–Mod11 represent the modules responsible for incorporation of individual amino acid residues.

(D) The organization of the 11 modules in Arf. The C domain contains seven highly conserved core motifs (C1–C7). The A domain is composed

The NRPS gene cluster of syringomycin (*syr*) [20], mycosubtilin (*myc*) [13], and iturin A (*itu*) [14] represent those with unique features. Syringomycin is a lipononapeptide phytotoxin and a key virulence determinant of *Pseudomonas syringae* strain B301D. The *syrB* and *syrE* genes do not respect the colinearity rule and lack E domains. Eight modules for the first eight amino acids in SyrE are arranged in a line, but the ninth module (SyrB), which is necessary for the incorporation of the last amino acid, is located in the upstream region [20]. Meanwhile, the *mycA* and *ituA* genes encode multifunctional hybrid enzymes of a peptide synthetase, an amino transferase, and a fatty acid synthase [13, 14]. These variations reveal that the NRPSs' assembly line is more complicated than previously thought. This knowledge prompted us to isolate and analyze the genes encoding arthrofactin synthetase. Here, we report that the arthrofactin synthetase gene cluster (*arf*) contains three genes encoding NRPSs, ArfA, ArfB, and ArfC. These proteins do not contain an E domain but do contain two internal Te domains. The *arf* would represent the gene structure of the lipoundecapeptide class in *Pseudomonas* spp. Two A domains for L-Leu and D-Leu were overproduced in *E. coli*, purified, and examined for substrate specificity. A gene disruption mutant was also prepared and characterized.

Results and Discussion

Structural Organization of *arf*

The nucleotide sequence of 40,430 bases covering the entire *arf* was determined by using λ5, λ4, λ-P3, and λ-S10 phage clones (Figure 1B). The gene cluster consists of three large genes, which are designated as *arfA*, *arfB*, and *arfC* (Figure 1C). The putative promoter regions −35 (5′-TCGATG-3′) and −10 (5′-CATTTT-3′) are found in the nucleotide sequences from 1,441 to 1,446 and from 1,464 to 1,469, respectively. The gene spans 38.7 kb and is followed by a putative transcriptional terminator located 19 bp downstream of the last *arfC* gene. The three gene products show strong amino acid sequence similarities to the members of the peptide synthetase family and show the highest similarities to syringomycin synthetase (Syr) of *P. syringae* B301D (DDBJ/EMBL/GenBank: AAC80285). The *arfA* gene is composed of 6,414 bp. The translation of this gene possibly starts from the ATG codon at position 1,642, which is located 6 bp downstream of a putative ribosome binding site (RBS), AGAAGG. The *arfA* gene would encode a protein (ArfA) with a molecular mass of 234 kDa, which consists of 2,137 amino acid (aa) residues. The translation of the *arfB* gene probably starts from the ATG codon, which is located 182 bp downstream of the TAG stop codon of the *arfA* gene and is preceded by a putative RBS, AGCAGG. The *arfB* gene is composed of 13,017 bp and encodes a protein (ArfB) with a molecular mass of 474 kDa (4,338 aa). The translation of the *arfC* gene may start from the TTG codon, which is located 66 bp downstream of the TGA stop codon of the *arfB* gene and is preceded by a putative RBS, GGACG, and ends at the TGA stop codon at position 39,093. The *arfC* gene is composed of 17,775 bp and encodes a protein (ArfC) with a molecular mass of 648 kDa (5,924 aa).

Characteristics of the Arthrofactin Synthetase

Based on analogy to other NRPSs, ArfA, ArfB, and ArfC are composed of two (Mod1, Mod2), four (Mod3–Mod6), and five (Mod7–Mod11) modules, respectively (Figure 1C). Because a modular structure is responsible for the incorporation of individual amino acid residues [15], and arthrofactin consists of 11 amino acid residues, ArfA, ArfB, and ArfC would form arthrofactin synthetase (Arf). Each module commonly has three major domains known as C, A, and T domains that are characteristic for such multienzymes. Most of the consensus amino acid residues identified within other NRPS modules are highly conserved in the NRPS modules of Arf (Figure 1D). The characteristic features of the individual domains of Arf are summarized below.

C Domains

The Arf bears ten typical C domains that probably catalyze amide bond formation of consecutively bound thioesterified intermediates of arthrofactin and one additional domain. Recently, site-directed mutagenesis work in the C domain of tyrocidine synthetase B revealed that Arg in the C2 motif and His/Asp in the C3 motif are involved in amide bond formation [21]. These functional amino acid residues are conserved in C2: RHX(I/T)LR(T/S)X(F/Y) and C3: HHXXXD(H/G) motifs from the 2nd to 11th C domains in Arf (Figure 1D). An additional C domain was identified in the first module of ArfA. This arrangement has also been reported for several NRPSs [9, 10, 12, 20], and it suggests that the first amino acid could be initially acylated with a fatty acid in this domain.

E Domains

The modifying E domain that catalyzes L to D form transformation of the thioester bound intermediates is typically associated with the modules incorporating D amino acids in Gram-positive bacteria. In contrast, although 7 amino acid residues out of 11 in arthrofactin are D form, the E domain is not associated with the modules incorporating respective D amino acids. This suggests that an external racemase is involved in L to D form conversion [20, 22]. A similar structure has been reported for the gene clusters encoding Syr and the synthetase of syringopeptin, a necrosis-inducing lipodocasapeptide phytotoxin, Syp from *P. syringae* B301D [20, 23]. This information may allow us to generalize that no E domain is present in NRPSs from *Pseudomonas* spp.

A Domains

Stachelhaus et al. postulated the amino acid specificity-conferring code for various A domains by comparing their primary structures to that of a phenylalanine A domain (PheA) of gramicidin synthetase whose crystal structure had been determined [24]. The corresponding

of ten core motifs (A1–A10), whereas the T domain comprises one core motif. Two unique internal Te domains are found at the C-terminal end of the 11th module.

Abbreviations of restriction enzymes are as follows: B, BamHI; M, MluI; S, SalI; Sc, SacI.

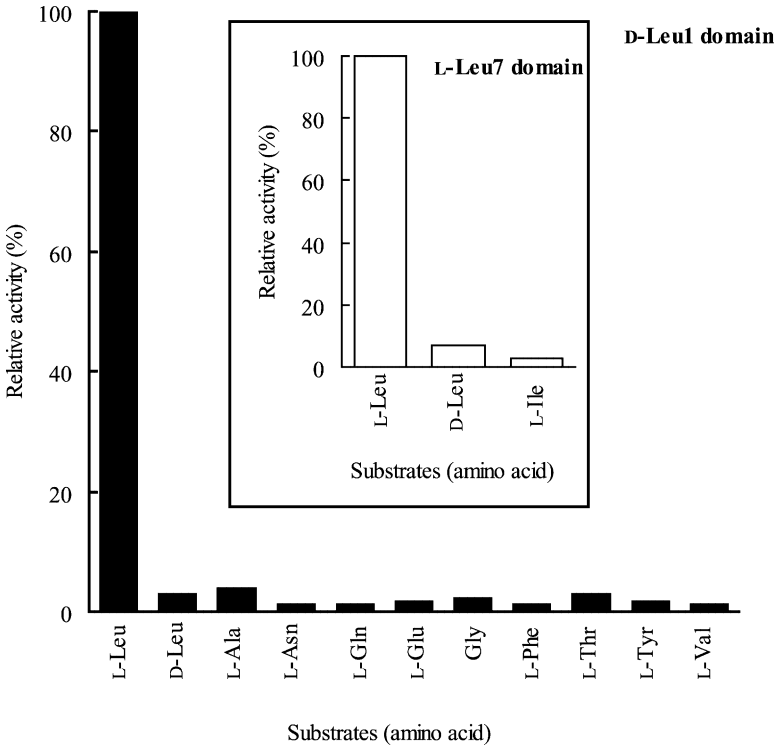

Figure 2. Relative Activities of the Purified D-Leu1 Domain for Different Amino Acid Substrates

A total of 20 typical L amino acids were tested for D-Leu1 activity; the substrate amino acids showing relative activity less than 1% of L-Leu are omitted from the graph. As for L-Leu7 activity, L-Leu, D-Leu, and L-Ile were tested, and the results are shown in a box.

amino acid sequences were also identified in Arf. The amino acid sequences of core motifs from A1 to A10 in 11 modules were aligned to the corresponding sequence of PheA. The signature sequences thus derived are as follows: (Mod1, Mod4, Mod5, Mod7) DAWFLGNVVK, (Mod2, Mod11) DSWFLGVVDK, (Mod3) DFWTIGMVHK, (Mod6, Mod8) DVWFMSLVDK, and (Mod9, Mod10) DAMFLGCTYK. All specificity-conferring codes of Mod1, Mod4, Mod5, and Mod7 are 100% identical to that of the Leu(1) conferring code and share 70% identity to Ile(1) of bacitracin, lichenysin, and surfactin synthetases. The sequences in Mod2 and Mod11 match 60% with Asp of SyrE, and, interestingly, both also share 70% identity to Ile(1). The signature of Mod3 is 90% identical to the Thr conferring code of SyrB. The Mod6 and Mod8 conferring codes are 80% identical to Ser of SyrE. Meanwhile, the signature sequences of Mod9 and Mod10 match 60% with Ile(2) of fengycin and plipastatin synthetases. These features strongly suggest that the *arf* structure is ordered by the colinearity rule.

In order to verify this assumption and investigate the mechanism of L and D amino acid incorporation into the peptide, we overproduced, purified, and determined amino acid specificities of the L- and D-Leu activating domains. The D-Leu1 and L-Leu7 A domains, representing D- and L amino acid-incorporating A domains, respectively, were tested for amino acid specificities by utilizing the ATP-PPi exchange assay with various amino acids. The results clearly showed that both D-Leu1 and L-Leu7 proteins recognized only L-Leu (Figure 2); these results are consistent with what was found for Syr [20]. These data strongly support the colinearity of the modular arrangement with the primary structure of the product lipopeptide. The first amino acid of arthrofactin is D-Leu, but the E domain was absent in the D-Leu1 module. The observation that the D-Leu1 A domain adopted only L-Leu as a substrate supports a hypothesis that an external racemase is responsible for L to D form conversion in NRPSs of *Pseudomonas* spp. It is proposed that epimerization might be concomitantly occurring at the peptidyl or aminoacyl stage [20].

T Domains

T domains in NRPS modules are generally classified into two groups according to their locations. T domains that are associated with an E domain have a highly conserved motif [F(F/Y)XXGGDSIKA(I/L)Q] (motif 1), where the active site serine residue is underlined, whereas T domains that are not associated with an E domain have another highly conserved motif [FFXXGGHSLKAXX] (motif 2). Linne et al. experimentally demonstrated that the aspartate residue in front of the invariant serine residue in motif 1 is important for proper interaction between the T and E domains. This interaction is crucial for substrate epimerization [25]. Although Arf contains no E domains, T domains in Arf could also be classified into two groups according to their locations. They are T(L) domains, responsible for transferring L amino acid residues to the intermediate peptide, and T(D) domains, responsible for transferring D amino acids. When the amino acid sequences of T(L) and T(D) domains in Arf are compared, all contain a [FFELGGHSLLA(V/M)] sequence motif (Figure 3), which is rather similar to motif 2. This supports the assumption that a different mechanism is adopted for D amino acid incorporation in Arf synthesis. Sequence alignments also reveal that significant differences between the T(L) and T(D) domains are located downstream of the conserved core motif. These differences might reflect the different topologies between A and T domains or otherwise the recognition by a putative external racemase that is necessary for D

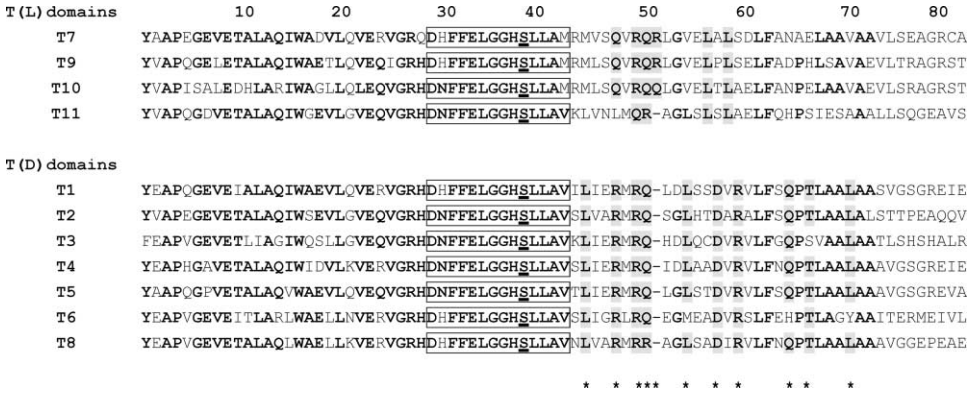

Figure 3. The Deduced Amino Acid Sequence Alignment of 11 T Domains in Arf

The T(L) domains are T7, T9, T10, and T11, whereas the remaining T domains belong to T(D) domains. Conserved motif sequences are boxed. The invariant serine residues are underlined. The asterisks show the differences found downstream of the conserved core motif.

amino acid incorporation. During posttranslational modification of NRPSs and PKSs, the phosphopantetheinyl transferases (PPTase) converts the inactive *apo* form of T domains to their active *holo* forms by transferring the phosphopantetheinyl arm from coenzyme A to the side chain hydroxyl group of an invariant serine residue [26]. A PPTase-encoding gene was not found in the vicinity of *arf*. In Gram-positive bacteria, this gene is usually located near the synthetase gene cluster, but not in Gram-negative bacteria.

Te Domains

C-terminal regions of NRPSs usually contain a Te domain with a molecular mass of 25–35 kDa (~250 aa) [17]. ArfC has a large C-terminal region, which is approximately 62 kDa (580 aa) in size. This region shows a significant similarity with the Te-like proteins and bears two putative Te domains, ArfCTe1 and ArfCTe2, which have conserved GXSXG sequence motifs (Figure 1D). It shows high amino acid sequence similarity (65% indentical) to the C-terminal region of the *P. fluorescens* PfO-1 hypothetical protein (ZP_00086026). ArfCTe2 contains two GXSXG motifs and shares amino acid sequence homology with α/β hydrolases. Phylogenetic analysis based on the differences in the amino acid sequences reveals that both domains belong to the internal Te group (data not shown) and suggests that ArfCTe2 is not an external Te that had been accidentally fused to ArfCTe1. Alignment of the amino acid sequences of the ArfCTe1 and ArfCTe2 domains (~250 aa each) with those of various internal Te domains reveals that Ser^{80}, Asp^{107}, and His^{207}, which form a catalytic triad of Te in SrfA-C [27], are conserved in both ArfCTe1 and ArfCTe2 (Figure 4). These results suggest that both ArfCTe domains are functional. Recently, the modular structure of Syp in *P. syringae* B301D has been reported [23]. Like ArfC, the C-terminal end of SypC also contains two putative Te domains. The gene cluster structure containing two Te domains with no E domain may be widely shared in *Pseudomonas* sp.

The C-terminal Te domain of NRPSs possesses either hydrolase or cyclase activity, resulting in the release of free carboxylate products or cyclic lactones/lactams, respectively [17]. ACV synthetase for penicillin biosynthesis represents hydrolase activity and releases a peptide carboxylate [28]. There are many NRPSs such as surfactin and tyrocidine synthetase that do not release free carboxylate product, but rather release cyclic lactones or lactams [18]. Both ArfCTe1 and ArfCTe2 show the highest amino acid sequence similarities to the Te domain of SyrE (~24%) from *P. syringae* B301D. They show higher sequence similarities (~20%) to eukaryotic ACV synthetase than to SrfA-C (~16%) from *B. subtilis*. ArfCTe1 and ArfCTe2 may share coordinated hydrolase and cyclase activities; however, either one may only have the cyclase activity responsible for completion of arthrofactin biosynthesis. A conserved domain of the α/β hydrolase family is contained in the ArfCTe2 domain, and the presence of this domain suggests that ArfCTe2 has higher hydrolase activity than ArfCTe1. ArfCTe2 may play a role in regenerating the misacylated NRPSs, as shown for type II Te of SrfA [19]. Trauger et al. have reported on the hydrolytic activity of a recombinant tyrocidine synthetase Te domain (TycCTe) [18]. The cyclization/hydrolysis ratio is 6:1, and this ratio is presumably higher in the intact peptide synthetase. It is of great interest to analyze the function of each Te domain in Arf.

Reprogramming of NRPSs has been successfully done to produce novel lipopeptide in vivo. This strategy includes recombination and alteration of functional domains or modules at the genetic level. The expected surfactin analog was obtained by the engineering of single module peptide synthetase [29, 30]. Amino acid sequence information of two unique internal Te domains together with 11 amino acid conferring codes in Arf would contribute to expand the limitation of novel NRPSs' design in vivo and in vitro.

Analyses of the *arf* Flanking Regions

The gene encoding a protein with 264 aa (ORF1) was found to be located in the upstream region of the *arfA* gene in an opposite direction. The amino acid sequence of ORF1 is 49% identical to SalC, a DNA binding protein from *P. syringae* B728a (AAG22806). A common helix-turn-helix motif of the LuxR regulator and Sal protein family was also conserved in the C-terminal region of ORF1. Another Sal protein family member, SalA (23.9% identical to SalC), has been reported as a regulatory

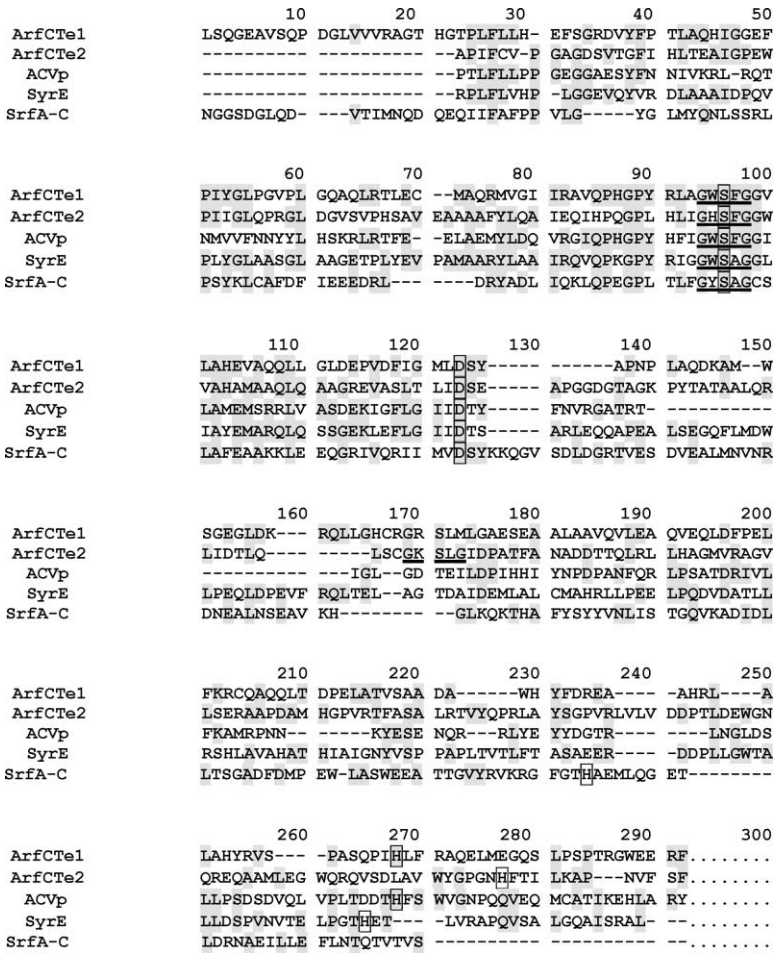

Figure 4. Alignment of the Deduced Amino Acid Sequences of Two Te Domains in ArfC, ArfCTe1 and ArfCTe2, and Various Internal Te Domains

The GXSXG motif is underlined. The positions of the catalytic triad residues of SrfA-C (S^{80}, D^{107}, H^{207}) are boxed. Abbreviations of the protein names are as follows: ACVp, delta-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase of *Penicillium chrysogenum*; SyrE, syringomycin synthetase; SrfA-C, surfactin synthetase.

protein that positively regulates the expression of a syringomycin biosynthetic gene, *syrB1* [31]. ORF1 may be a regulator protein that controls the expression level of *arf*. Recently, the GacA/GacS two-component system has been shown to control the production of lipoundecapeptide amphisin in *Pseudomonas* sp. DSS73 [32]. Sequencing analysis revealed that two additional genes, which encode an outer membrane efflux protein homolog (ORF5) and a partial ABC transporter protein (ORF6), are present in the downstream region of the *arfC* gene. The direction of the transcriptions of the *ORF5* gene, which is located just 78 bp downstream of the *arfC* gene, and the *ORF6* gene are the same as that of *arf*. These protein groups conform to an ATP-dependent transporter channel that plays a coordinating role in mutidrug efflux systems [33]. These systems have emerged as key mechanisms in mutidrug-resistant strains of *P. aeruginosa* and other Gram-negative bacteria. *Pseudomonas* sp. MIS38 is resistant to high concentrations (>20 mg/l) of antibiotics such as ampicillin and chloramphenicol. ORF5 might be involved in antibiotic resistance of the strain. On the other hand, ABC-type transporter systems flanking to the peptide synthetases are proposed to be involved in the product secretion [11, 34, 35]. Disruption of the gene encoding either the transporter or the synthetase in *Proteus mirabilis* gave the same phenotypic alteration [34]. This experimental result allows us to imagine that ORF5 and ORF6 could be members of the secretion channel of arthrofactin.

Disruption of the *arfB* Gene

In order to confirm that cloned *arf* is functional in vivo, the *arfB* gene was genetically disrupted and its effect on arthrofactin production was examined. A plasmid, pUC18ΔSacI-B2::*kan*, for the *arfB* gene disruption was prepared by inserting a kanamycin-resistant gene cassette (*kan*) in the *arfB* gene at the D-Leu4-encoding locus. The plasmid was transferred into MIS38, so that kanamycin-resistant (Km^r) and biosurfactant-nonproducing (BS^-) colonies (assayed by oil displacement activity, [1]) were obtained. One candidate, designated as strain NC1 (Km^r/BS^-), was selected for further investigation. The integration of the *kan* gene in the *arfB* gene of the NC1 genome was verified by the PCR method using primers (P1, P2 in Figure 1C) that could anneal outside of the D-Leu4 gene locus. Amplification of the single 4,160 bp PCR product suggested that the *kan* gene was integrated at the expected position by a double-crossover recombination event (data not shown). Insertion of the *kan* gene at a single site in the genome was further confirmed by Southern blot analysis (data not shown). The mutant strain NC1 grew normally and produced secretion enzymes lipase and protease at a similar level to that of the wild-type strain MIS38 [2].

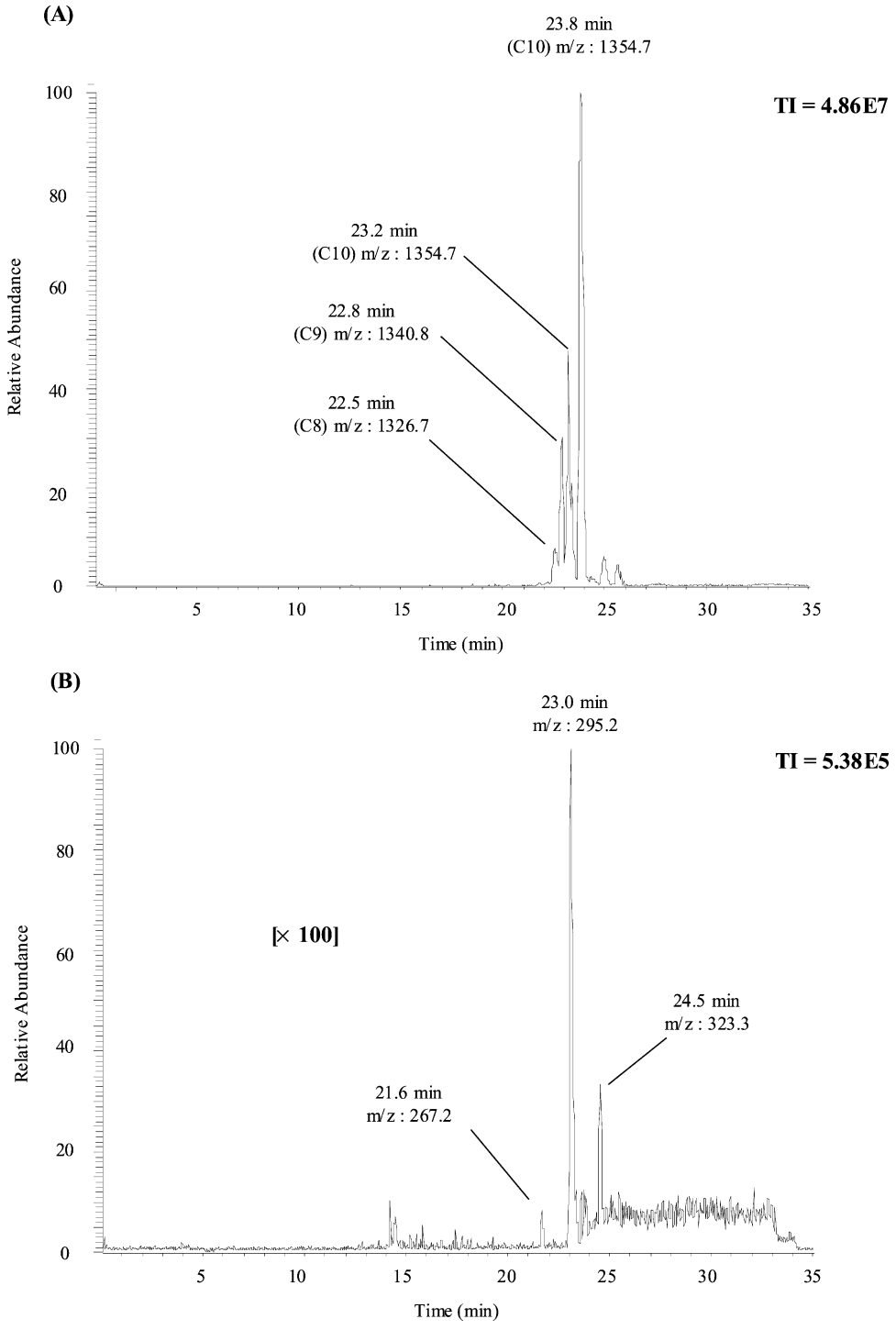

Figure 5. HPLC/MS Analysis of Methanol Extracts from the Spent Media

(A and B) (A) wild-type MIS38 and (B) mutant NC1 (magnified 100×). The mass number (m/z) is shown for the major peaks with retention time. TI; total ions.

The production of arthrofactin in MIS38 and NC1 was compared by analyzing methanol extracts of their culture broth with HPLC/MS. Peaks corresponding to arthrofactin (C_{10}, m/z 1354.7) and its cognates (C_8 and C_9) were found in the sample from MIS38 (Figure 5A), while they were completely missing in that from NC1 (Figure 5B). When the sample from NC1 was analyzed, several compounds were eluted from HPLC with similar

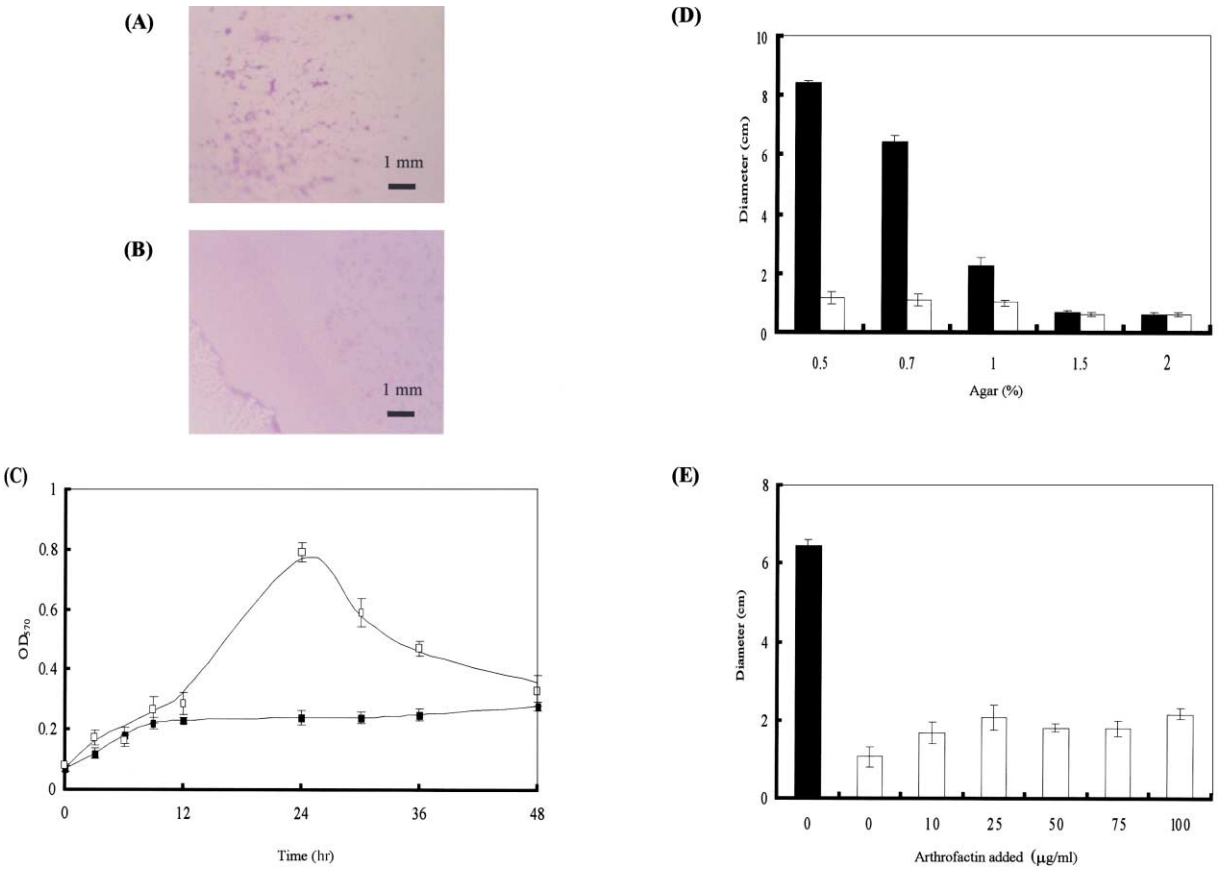

Figure 6. Biofilm Formation and Swarming Motility of the Wild-Type Strain MIS38 and the Mutant Strain NC1

(A and B) Microscopic observation of crystal violet staining of the biofilm at 24 hr is shown in (A) MIS38 and (B) NC1.

(C) The time course of biofilm formation is shown.

(D) For the swarming motility assay, colony diameter was measured after cells' growth in various concentrations of agar.

(E) Swarming motility of NC1 at different concentrations of externally added arthrofactin. The plates were prepared with 0.7% agar in this case. The data are averages of triplicate samples. MIS38 is represented by solid squares; NC1 is represented by open squares.

retention times to those of arthrofactin and its cognates. However, their molecular weights were much smaller than those of arthrofactin and its cognates. These results indicate that NC1 does not produce arthrofactin at all. Therefore, we conclude that Arf is functional in vivo.

Characteristics of Arthrofactin-Deficient Strain NC1

Biosurfactants display a wide range of biological properties, such as a role as an enzyme inhibitor and antibiotic, hemolytic, and surface activities. Biosurfactants may play crucial roles in the strategies bacteria use to dominate in their environments. In this study, a mutant strain, NC1, deficient in arthrofactin production, was characterized from these viewpoints.

Biofilm Formation

Biofilms are a bacterial community attached to a solid surface and are embedded in polymer matrices such as polysaccharides. Biofilm formation could be a strategy that cells employ to survive in a scrum under harsh natural conditions [36]. Biosurfactants are an essential factor in keeping water channels in a mature biofilm. *P. aeruginosa* is a pathogen that infects cystic fibrosis patients and has been best studied for its molecular mechanisms of biofilm formation. It has been reported that the *rhlA* mutant of *P. aeruginosa* PAO1, which is defective in the production of a glycolipid-type biosurfactant, rhamnolipid, produces a biofilm with a thick, uniform mat of bacterial cells and is unable to maintain channels [37]. Because MIS38 does not produce rhamnolipid [1], arthrofactin is postulated to play an alternative role in developing the mature biofilm. Then, the effect of the disruption of the *arfB* gene on biofilm formation was examined. Time courses of biofilm development by MIS38 and NC1 are shown in Figure 6C. The results clearly show that NC1 forms unstable, but more, biofilms than MIS38. The microscopic appearances of the biofilms formed by the wild-type and mutant strains were different from each other. Biofilms are separated by channels in MIS38, but they are uniformly flat and fused in NC1 (Figures 6A and 6B). These results suggest that a glycolipid biosurfactant and a lipopeptide biosurfactant similarly function in normal biofilm formation in *Pseudomonas* strains.

Swarming Motility

Swimming motility is attributed to flagellation and pillation of individual cells and is independent of communi-

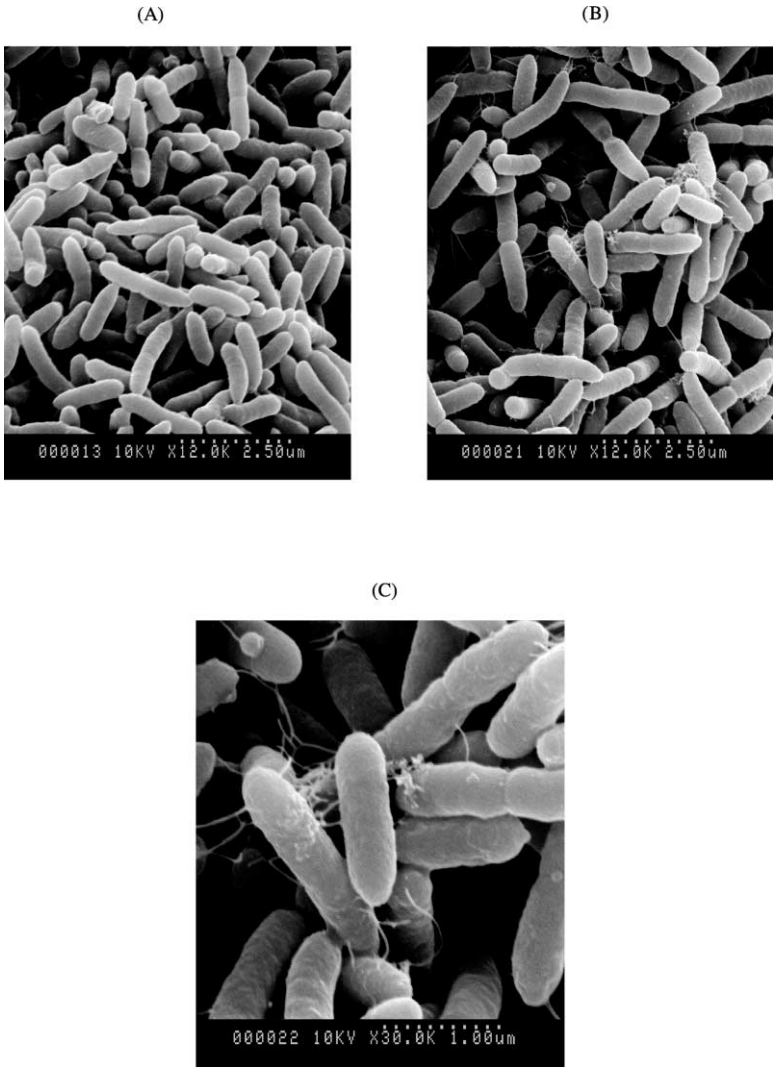

Figure 7. Scanning Electron Microscopic Observations

(A) MIS38.
(B) NC1.
(C) Closer view of NC1.

Colonies on a 0.7% agar plate were gently washed with distilled water, fixed with glutaraldehyde and OsO_4, dehydrated by critical CO_2, and sputter coated with platinum. Observation was performed by using Hitachi S800. Although there is no dramatic change in the shape of the cells, an extracellular fiber network is found only for NC1.

cation with other cells. Meanwhile, swarming motility depends on cell-cell contact together with hyperflagellation to specifically respond to cell density [38]. Swarming can be defined as an expanding colony on semisolid media [39]. In *Serratia liquefaciens*, the *flhDC* flagella regulatory master operon, temperature, nutrient status, and quorum sensing all contribute to the regulation of swarming motility [40]. Surface tension is also a critical factor that affects the swarming phenomenon [41].

The arthrofactin-producing strain MIS38 exhibited high swarming motility (Figure 6D), while the activity was lost in the mutant strain NC1, which did not produce arthrofactin. The colony size of NC1 on solid media containing 0.5% agar was almost identical to that on solid media containing 1% agar. This suggests that arthrofactin production is required for swarming motility of MIS38. To address this assumption, complementation experiments were performed (Figure 6E). It was found that the external addition of arthrofactin at the concentrations even above the critical micelle concentration, 13.5 μg/ml, did not restore the swarming activity of NC1 to the level of MIS38. It only slightly enhanced the swarming activity of the mutant NC1. This observation is inconsistent with those for serrawettin W2 [41] and amphisin [42]. Transmission electron microscopic observation of NC1 after negative staining by uranyl acetate demonstrated that the flagellation of MIS38 and NC1 were at an equivalent level (picture not shown). When we observed the cells by use of a scanning electron microscope, MIS38 and NC1 showed different cell surface architectures (Figure 7). A fiber network is formed only in NC1 (Figures 7B and 7C). The cells seem to be connected to one another by this network, so that swarming motility might be interfered with. Although it remains to be elucidated whether *arf* regulates the production of these fibers, this phenomenon recalls the gene structure of *srfA*. A competence regulatory peptide, ComS, is encoded in a part of the surfactin biosynthesis gene cluster, *srfA-B* of *B. subtilis* [43].

Significance

Cloning and sequencing of the arthrofactin synthetase gene cluster revealed that the overall modular architecture of synthetase follows the colinearity rule. However, it lacks epimerization domains and maintains two internal thioesterase domains. This arthrofactin synthetase gene cluster may represent those for the

cyclic lipoundecapeptide class of *Pseudomonas* spp. Cyclic lipoundecapeptides are a unique class of lipopeptides that have high surface activity (arthrofactin), antifungal activity (amphisin, lokisin, tensin), and inhibitory activity for phospholipase C (pholipeptin). Because these valuable properties of cyclic lipoundecapeptides accompany a high potential for industrial application, engineering of the arthrofactin synthetase gene cluster should facilitate the production of lipopeptides with improved properties. The *arfB* gene disruption mutant did not produce arthrofactin and exhibited low swarming activity but enhanced biofilm formation and extracellular fiber production. These results suggest that the arthrofactin synthetase gene has multiple functions. Characterization of the adenylation domains for L-Leu and D-Leu overproduced in *E. coli* revealed that both domains recognized only L-Leu. This result suggests that an external racemase is generally involved in nonribosomal peptide synthetases from *Pseudomonas* spp. Control of lactone formation between the carboxyl group of the C-terminal amino acid residue and either the β-hydroxyl group of the fatty acid portion or the hydroxyl group of the threonine residue is the future challenge.

Experimental Procedures

Bacterial Strains and Plasmids

Arthrofactin-producing *Pseudomonas* sp. MIS38 was previously isolated from oil spills in Shizuoka prefecture, Japan [1]. *Escherichia coli* JM109 and *E. coli* LE392 were used as host strains for the construction of recombinant plasmids and a λ-EMBL3 genomic library, respectively. *E. coli* BL21(DE3) was used for overproduction of proteins with an expression vector, pET-28a(+) (Novagen). Cloning vectors pUC18, pUC19, and pT7Blue were used in *E. coli* JM109. In the construction of the *arfB* disruption gene, plasmid pSMC32 was used as a template for polymerase chain reactions (PCR) to prepare the *kan* gene. pSMC32 is a derivative of pSU36 (X53938) [44].

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 [4]. DNA fragments were recovered from an agarose gel by using the GeneClean Kit (Bio101). 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. In order to construct a λ-EMBL3 genomic library, genomic DNA of MIS38 was partially digested with Sau3AI. The fractions containing DNA fragments from 9 to 23 kb were separated by agarose gel electrophoresis, recovered, and ligated into λ-EMBL3 BamHI arms (Stratagene). Lamda lytic suspension was used as a genomic library for gene screening.

PCR was performed in 30 cycles by using a thermal cycler, the GeneAmp PCR System 2400 (Perkin-Elmer), and KOD DNA polymerase (Toyobo). Oligodeoxyribonucleotides for PCR primers were synthesized at GENSET KK. The nucleotide sequences of the gene fragments were determined by using the dideoxy-chain termination method with the ABI Prism BigDye terminator v3.0 cycle sequencing kit and the autosequencer ABI Prism 310 (Applied Biosystems). The DNA sequence was analyzed by using DNASIS software (Hitachi Software) and BLAST programs (www.ncbi.nlm.nih.gov/BLAST/, NCBI, [45]). Amino acid sequences were compared and analyzed by the Clustal W program (http://clustalw.genome.ad.jp/, [46]).

Cloning of *arf*

A part of *arf* was amplified by the PCR method with a combination of primers, which were constructed according to the nucleotide sequences in highly conserved regions among various NRPSs. The nucleotide sequences of the primers were as follows: forward, 5′-TGTACMRNACVGGYVAYYTGG-3′ and reverse, 5′-CCSCCSAGVT CYAAYAAYTVYTC-3′, where M, R, N, V, Y, and S represent A/C, A/G, A/C/G/T, A/C/G, C/T, and C/G, respectively. The genomic DNA of MIS38 was used as a template. The resultant DNA fragment (520 bp), whose deduced amino acid sequence showed high homology to the peptide synthetase family, was used as a probe for screening of the λ-EMBL3 genomic library constructed by Max Plax Packaging Extract (AIRBROWN). Plaque hydridization was performed by using the AlkPhos Direct Labeling and Detection System (Amersham Biosciences). Two positive phage clones with a 14.5 and 17.5 kb insert, designated as λ4 and λ5, respectively, were obtained (Figure 1B). DNA sequence analysis revealed that two more clones, λ-P3 and λ-S10, were necessary to cover the entire *arf*. The DNA inserts of each positive λ phage clone were physically mapped and subcloned in pUC18 for sequence determination, mostly by using BamHI and SalI. In order to assemble the sequence data, the nucleotide sequences around the cloning site of the subclones were determined by using λ-DNAs as sequencing templates.

Overproduction and Purification of Histidine-Tagged D-Leu1 and L-Leu7 A Domains

Gene fragments encoding the first (D-Leu1) and seventh (L-Leu7) A domains were amplified by PCR. The boundary of domains was predicted as described previously [10]. The oligonucleotide primers were as follows: D-Leu1-NdeI, 5′-CTGCATATGCACTTGCTGAG CAATCTGGC-3′; D-Leu1-EcoRI, 5′-CTGGAATTCTTACAGGGCGAT TTCGACCTC-3′; L-Leu7-NdeI, 5′-CTGCATATGACGATGAGTGTCG ATGACCT-3′; L-Leu7-EcoRI, 5′-CTGGAATTCTTACGCTGCGTACT CGCGGGT-3′. Restriction enzyme sites are underlined. The blunt ended PCR products were first cloned into pUC19 at the SmaI site, and then the NdeI-EcoRI fragment was excised from the plasmid for introduction into the expression vector pET-28a(+). Addition of the TAA stop codon (set in bold in the above sequences) leads to production of the recombinant proteins containing the histidine tag only at the N-terminal region. The resulting expression plasmids designated pET-28a:D-Leu1 and pET-28a:L-Leu7 were transferred into *E. coli* BL21(DE3). *E. coli* cells harboring pET-28a:D-Leu1 and pET-28a:L-Leu7 were grown at 37°C until OD_{660} reached 0.5, added to 1.0 and 2.0 mM IPTG (isopropyl-β-D-thiogalactopyranoside), respectively, and cultivated at 25°C for 4 hr. Under this condition, target proteins were overproduced mostly in a soluble form. Purification of His_6-tagged D-Leu1 and L-Leu7 proteins was performed by disrupting the cells by sonication, followed by Ni^{2+}-charged HiTrap Chelating HP and HitrapQ HP column chromatography (Amersham Biosciences). The purity of the proteins was determined by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis with Coomassie Brilliant Blue staining. Protein concentration was determined by using the Bio-Rad Protein Assay Kit (Bio-Rad).

ATP-PPi Exchange Assay

In order to determine the amino acid specificities of purified D-Leu1 and L-Leu7 proteins, the ATP-PPi exchange assay was adopted [20]. The assay mixture contained 50 mM HEPES (pH 7.0), 100 mM NaCl, 10 mM $MgCl_2$, 2 mM DTE, 1 mM EDTA, 2 mM amino acid, and 250 nM protein. The reaction was initiated by the addition of starting mixture (2 mM ATP, 0.2 mM tetrasodium pyrophosphate, and 0.15 μCi tetrasodium [^{32}P] pyrophosphate) in a total volume of 100 μl and was incubated at 37°C for 10 min. The reaction was terminated by the addition of 0.5 ml terminating mixture (1.2% [w/v] activated charcoal, 0.1 M tetrasodium pyrophosphate, and 0.3 M $HClO_4$). The charcoal was washed by a brief centrifugation and was resuspended with 0.5 ml distilled water. Finally, 3.5 ml scintillation fluid was added to the samples, and the bound radioactivity was determined by the use of the liquid scintillation counter LPS 6500 (Beckman).

Construction of the *arfB* Gene Disruption Mutant

The internal 3,372 bp BamHI (B2) fragment of the *arfB* gene was cloned in pUC18ΔSacI, in which the SacI site of pUC18 is filled by using the DNA Blunting Kit (Takara shuzo). The resulting plasmid pUC18ΔSacI-B2 was digested with MluI and SacI, and the internal 460 bp MluI/SacI fragment was replaced by the 1,210 bp MluI/SacI fragment containing the *kan* gene and its promoter. This 1,210 bp MluI/SacI fragment had been amplified by using the PCR method using vector pSMC32 as a template. Oligonucleotide primers 5′-CATGACGCGTGTTTTATGGACAGCAAGCGA-3′ and 5′-CATGGAG

CTCCCGTCA GTAGCTGAACAGGA-3′ supplemented with MluI and SacI restriction sites (underlined) were used for PCR. The resultant gene disruption plasmid pUC18ΔSacI-B2::*kan* was then transferred into *Pseudomonas* sp. MIS38 by an electrotransformation method as follows.

Electrotransformation of Strain MIS38

An overnight culture was inoculated into 100 ml L-broth (1% inoculum size). L-broth consists, per liter, of 10 g tryptone, 5 g yeast extract, and 5 g NaCl (pH 7.2). After 5.5 hr of shaking at 30°C, the cells ($OD_{600} \approx 1.0$) were collected by centrifugation (10,000 × g for 10 min at 4°C) and washed once with cold 1 mM HEPES buffer (pH 7.0). Then, the cells were washed twice with 10% glycerol and resuspended in 0.5 ml 10% glycerol. A portion of this cell suspension (50 μl) was mixed with plasmid DNA (1 μg) 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 (10 kV/cm with 25 μF and 200 Ω) by using the Bio-Rad Gene Pulser (Bio-Rad). Treated cells were immediately suspended with 1 ml L-broth. After incubation at 30°C for 3 hr, 100 μl each of culture was spread on L-agar plates containing 35 μg/ml kanamycin and was overlaid with 30 μl crude oil. The plates were incubated at 30°C for 24–36 hr [47].

Analysis of Arthrofactin Production

Bacterial strains MIS38 and NC1 were grown in L-broth at 30°C for 40 hr. Arthrofactin and its derivatives were purified and analyzed by the following methods. Bacterial cells were removed from 250 ml culture medium by centrifugation (20,000 × g for 15 min) at 4°C. The supernatant was acidified by adding concentrated HCl to a final pH of 2.0 and was allowed to form aggregates at 4°C for 3 hr. The aggregates were collected by centrifugation and were washed three times with diluted HCl (pH 2.0). Biosurfactant containing lipophilic substances was extracted from the precipitates three times with methanol and was used for the analysis by reverse-phase HPLC as described below. When necessary, the solvent was removed by using a rotary evaporator under vacuum conditions.

Reverse-phase HPLC was carried out on an octadecyl silica gel column (Cosmosil $5C_{18}$AR 4.6 × 150 mm, Nacalai) attached to a system HP1100 (Hewlett Packard). The mobile phase was changed linearly from eluent A (10% acetonitrile and 0.1% trifluoroacetic acid) to eluent B (acetonitrile and 0.05% trifluoroacetic acid) in 30 min at a flow rate of 0.5 ml/min. Peaks eluting from the column were monitored by their absorbance at 220 nm. The molecular weight of each component was determined by using a mass spectrometer LCQ (Thermo Finnigan) equipped with an electrospray ion source, HPLC/MS.

Swarming Motility and Biofilm Formation Assay

Swarming motility of bacterial strains was examined by measuring a diameter of the colony grown on L-agar plates containing various concentrations of agar. Bacterial strains were cultivated to an OD_{600} of ~0.5 at 30°C, and then 1 μl of each culture was spotted on each L-agar plate (8.5 cm diameter plate). After incubation at 30°C for 12 hr, the diameter of the colony was measured.

As to the biofilm formation assay, cells grown overnight on L-agar plates were subsequently diluted in liquid L-broth to an OD_{600} of ~0.3. The cell suspension was further diluted by one hundred times in 1.5 ml polypropylene microcentrifuge tubes containing 300 μl L-medium. The tubes were incubated at 30°C without shaking, typically for 24 hr. Then, nonadherent cells were removed by rinsing several times with distilled water. Finally, cells attached to the inner surface of the tube were stained with 1% crystal violet solution for 25 min at room temperature, followed by rinsing several times with distilled water. The cell-associated dye was solubilized with 400 μl dimethylsulfoxide (DMSO) and was quantified by measuring the OD_{570} [48].

Plate Assay for Lipase and Protease

Lipase activity was detected as a clearing zone, halo, around the colony on a plate of LTB-agar medium; L-broth supplemented with 1% tributyrin, 0.1% Tween 80, and 1.5% agar. Protease activity was similarly detected on LCA-agar or LG agar medium; L-broth supplemented with 0.5% casein or gelatin and 1.5% agar, respectively.

Acknowledgments

We would like to thank Dr. Roberto Kolter (Harvard Medical School) for providing plasmid pSMC32 and Ms. Atsuko Takatsuka for technical assistance. Electron microscopic observations were a collaboration with Dr. Masao Kuroda and Mr. Eiji Oiki (Osaka University Graduate School of Medicine). This work was supported by a grant from the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN). N.R. acknowledges his Ph.D. fellowship from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government.

Received: June 16, 2003
Revised: July 18, 2003
Accepted: July 24, 2003
Published: September 19, 2003

References

1. Morikawa, M., Daido, H., Takao, T., Murata, S., Shimonishi, Y., and Imanaka, T. (1993). A new lipopeptide biosurfactant produced by *Arthrobacter* sp. strain MIS38. J. Bacteriol. *175*, 6459–6466.
2. Amada, K., Haruki, M., Imanaka, T., Morikawa, M., and Kanaya, S. (2000). Overproduction in *Escherichia coli*, purification and characterization of a family I.3 lipase from *Pseudomonas* sp. MIS38. Biochim. Biophys. Acta *1487*, 201–210.
3. Arima, K., Kakinuma, K., and Tamura, G. (1968). Surfactin, a crystalline peptide lipid surfactant produced by *Bacillus subtilis*: isolation, characterization and its inhibition of fibrin clot formation. Biochem. Biophys. Res. Commun. *31*, 488–494.
4. Morikawa, M., Ito, M., and Imanaka, T. (1992). Isolation of a new surfactin producer *Bacillus pumilus* A-1, and cloning and nucleotide sequence of the regulator gene, *psf*-1. J. Ferment. Bioeng. *74*, 255–261.
5. Nielsen, T.H., Sorensen, D., Tobiasen, C., Andersen, J.B., Christophersen, C., Givskov, M., and Sorensen, J. (2002). Antibiotic and biosurfactant properties of cyclic lipopeptides produced by fluorescent *Pseudomonas* spp. from the sugar beet rhizosphere. Appl. Environ. Microbiol. *68*, 3416–3423.
6. Sorensen, D., Nielsen, T.H., Sorensen, J., and Christophersen, C. (2002). Cyclic lipoundecapeptide lokisin from *Pseudomonas* sp. strain DSS41. Tetrahedron Lett. *43*, 4421–4423.
7. Ui, H., Miyake, T., Iinuma, H., Imoto, M., Naganawa, H., Hattori, S., Hamada, M., Takeuchi, T., Umezawa, S., and Umezawa, K. (1997). Pholipeptin, a novel cyclic lipoundecapeptide from *Pseudomonas fluorescens*. J. Org. Chem. *62*, 103–108.
8. Roongsawang, N., Thaniyavarn, J., Thaniyavarn, S., Kameyama, T., Haruki, M., Imanaka, T., Morikawa, M., and Kanaya, S. (2002). Isolation and characterization of a halotolerant *Bacillus subtilis* BBK-1 which produces three kinds of lipopeptides; bacillomycin L, plipastatin, and surfactin. Extremophiles *6*, 499–506.
9. Cosmina, P., Rodriguez, F., de Ferra, F., Grandi, G., Perego, M., Venema, G., and van Sinderen, D. (1993). Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. Mol. Microbiol. *8*, 821–831.
10. Konz, D., Doekel, S., and Marahiel, M.A. (1999). Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. J. Bacteriol. *181*, 133–140.
11. Yakimov, M.M., Kroger, A., Slepak, T.N., Giuliano, L., Timmis, K.N., and Golyshin, P.N. (1998). A putative lichenysin A synthetase operon in *Bacillus licheniformis*: initial characterization. Biochim. Biophys. Acta *1399*, 141–153.
12. Lin, T.P., Chen, C.L., Chang, L.K., Tschen, J.S., and Liu, S.T. (1999). Functional and transcriptional analyses of a fengycin synthetase gene, *fenC*, from *Bacillus subtilis*. J. Bacteriol. *181*, 5060–5067.
13. Duitman, E.H., Hamoen, L.W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., et al. (1999). The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide

synthetase, an amino transferase, and a fatty acid synthase. Proc. Natl. Acad. Sci. USA *96*, 13294–13299.
14. Tsuge, K., Akiyama, T., and Shoda, M. (2001). Cloning, sequencing, and characterization of the iturin A operon. J. Bacteriol. *183*, 6265–6273.
15. Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. *97*, 2651–2673.
16. Schneider, A., and Marahiel, M.A. (1998). Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. Arch. Microbiol. *169*, 404–410.
17. Keating, T.A., Ehmann, D.E., Kohli, R.M., Marshall, C.G., Trauger, J.W., and Walsh, C.T. (2001). Chain termination steps in nonribosomal peptide synthetase assembly lines: directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis. Chembiochem *2*, 99–107.
18. Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A., and Walsh, C.T. (2000). Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. Nature *407*, 215–218.
19. Schwarzer, D., Mootz, H.D., Linne, U., and Marahiel, M.A. (2002). Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. Proc. Natl. Acad. Sci. USA *99*, 14083–14088.
20. Guenzi, E., Galli, G., Grgurina, I., Gross, D.C., and Grandi, G. (1998). Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. J. Biol. Chem. *273*, 32857–32863.
21. Bergendahl, V., Linne, U., and Marahiel, M.A. (2002). Mutational analysis of the C-domain in nonribosomal peptide synthesis. Eur. J. Biochem. *269*, 620–629.
22. Nishizawa, T., Asayama, M., and Shirai, M. (2001). Cyclic heptapeptide microcystin biosynthesis requires the glutamate racemase gene. Microbiol. *147*, 1235–1241.
23. Scholz-Schroeder, B.K., Soule, J.D., and Gross, D.C. (2003). The *sypA*, *sypB*, and *sypC* synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. *syringae* B301D. Mol. Plant Microbe Interact. *16*, 271–280.
24. Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. *6*, 493–505.
25. Linne, U., Doekel, S., and Marahiel, M.A. (2001). Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases. Biochemistry *40*, 15824–15834.
26. Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily — the phosphopantetheinyl transferases. Chem. Biol. *3*, 923–936.
27. Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T., and Stubbs, M.T. (2002). Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. Structure (Camb). *10*, 301–310.
28. Kallow, W., Kennedy, J., Arezi, B., Turner, G., and von Dohren, H. (2000). Thioesterase domain of delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase: alteration of stereospecificity by site-directed mutagenesis. J. Mol. Biol. *297*, 395–408.
29. Eppelmann, K., Stachelhaus, T., and Marahiel, M.A. (2002). Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. Biochemistry *41*, 9718–9726.
30. Stachelhaus, T., Schneider, A., and Marahiel, M.A. (1995). Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. Science *269*, 69–72.
31. Kitten, T., Kinscherf, T.G., McEvoy, J.L., and Willis, D.K. (1998). A newly identified regulator is required for virulence and toxin production in *Pseudomonas syringae*. Mol. Microbiol. *28*, 917–929.
32. Koch, B., Nielsen, T.H., Sorensen, D., Andersen, J.B., Christophersen, C., Molin, S., Givskov, M., Sorensen, J., and Nybroe, O. (2002). Lipopeptide production in *Pseudomonas* sp. strain DSS73 is regulated by components of sugar beet seed exudate via the Gac two-component regulatory system. Appl. Environ. Microbiol. *68*, 4509–4516.
33. Johnson, J.M., and Church, G.M. (1999). Alignment and structure prediction of divergent protein families: periplasmic and outer membrane proteins of bacterial efflux pumps. J. Mol. Biol. *287*, 695–715.
34. Gaisser, S., and Hughes, C. (1997). A locus coding for putative non-ribosomal peptide/polyketide synthase functions is mutated in a swarming-defective *Proteus mirabilis* strain. Mol. Gen. Genet. *253*, 415–427.
35. Quigley, N.B., Mo, Y.Y., and Gross, D.C. (1993). SyrD is required for syringomycin production by *Pseudomonas syringae* pathovar *syringae* and is related to a family of ATP-binding secretion proteins. Mol. Microbiol. *9*, 787–801.
36. Watnick, P., and Kolter, R. (2000). Biofilm, city of microbes. J. Bacteriol. *182*, 2675–2679.
37. Davey, M.E., Caiazza, N.C., and O'Toole, G.A. (2003). Rhamnolipid surfactant production affects biofilm architecture in *Pseudomonas aeruginosa* PAO1. J. Bacteriol. *185*, 1027–1036.
38. Eberl, L., Molin, S., and Givskov, M. (1999). Surface motility of *Serratia liquefaciens* MG1. J. Bacteriol. *181*, 1703–1712.
39. Harshey, R.M. (1994). Bees aren't the only ones: swarming in Gram-negative bacteria. Mol. Microbiol. *13*, 389–394.
40. Givskov, M., Ostling, J., Eberl, L., Lindum, P.W., Christensen, A.B., Christiansen, G., Molin, S., and Kjelleberg, S. (1998). Two separate regulatory systems participate in control of swarming motility of *Serratia liquefaciens* MG1. J. Bacteriol. *180*, 742–745.
41. Lindum, P.W., Anthoni, U., Christophersen, C., Eberl, L., Molin, S., and Givskov, M. (1998). N-Acyl-L-homoserine lactone autoinducers control production of an extracellular lipopeptide biosurfactant required for swarming motility of *Serratia liquefaciens* MG1. J. Bacteriol. *180*, 6384–6388.
42. Andersen, J.B., Koch, B., Nielsen, T.H., Sorensen, D., Hansen, M., Nybroe, O., Christophersen, C., Sorensen, J., Molin, S., and Givskov, M. (2003). Surface motility in *Pseudomonas* sp. DSS73 is required for efficient biological containment of the root-pathogenic microfungi *Rhizoctonia solani* and *Pythium ultimum*. Microbiol. *149*, 37–46.
43. D' Souza, C., Nakano, M.M., and Zuber, P. (1994). Identification of *comS*, a gene of the *srfA* operon that regulates the establishment of genetic competence in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA *91*, 9397–9401.
44. Bartolome, B., Jubete, Y., Martinez, E., and de la Cruz, F. (1991). Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. Gene *102*, 75–78.
45. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
46. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.
47. Morikawa, M., Daido, H., Pongpobpibool, S., and Imanaka, T. (1994). Construction of a new host-vector system in *Arthrobacter* sp., and cloning of the lipase gene. Appl. Microbiol. Biotechnol. *42*, 300–303.
48. O'Toole, G.A., and Kolter, R. (1998). Initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis. Mol. Microbiol. *28*, 449–461.

Accession Numbers

All of the nucleotide sequences reported in this paper have been deposited in the DDBJ/EMBL/GenBank databases under accession number AB107223.