Chemistry & Biology, Vol. 12, 1189–1200, November, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.chembiol.2005.08.010

Generation of D Amino Acid Residues in Assembly of Arthrofactin by Dual Condensation/Epimerization Domains

Carl J. Balibar, Frédéric H. Vaillancourt, and Christopher T. Walsh*
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
Boston, Massachusetts 02115

Summary

The first 6 residues of the biosurfactant lipopeptidolactone arthrofactin have the D configuration, yet none of the 11 modules of the nonribosomal peptide synthetase assembly line have epimerization domains. We show that the two-module ArfA subunit and the first module of the ArfB subunit, which act in tandem to produce the *N*-acyl-D-Leu$_1$-D-Asp$_2$-D-Thr$_3$-*S*-protein intermediate, activate the L amino acids and epimerize them as the aminoacyl-*S*-pantetheinyl T domain intermediates before the next downstream condensation. The condensation (C) domains are shown to have $^{D}C_{L}$ chirality in peptide bond formation. The upstream aminoacyl/peptidyl moiety is epimerized before condensation only when the condensation domains are simultaneously presented with the L-aminoacyl-*S*-pantetheinyl acceptor. These $^{D}C_{L}$ catalysts are dual function condensation/epimerization domains that can be predicted by bioinformatics analysis to be responsible for incorporation of all D residues in arthrofactin and of D residues in syringomycin, syringopeptin, and ramoplanin synthetases.

Introduction

Nonribosomal peptide synthetases (NRPSs) are large, multifunctional enzymes that synthesize a vast array of biologically relevant natural products [1–4]. NRPS systems are organized into an assembly line of distinct modules, the order of which is often colinear with the primary sequence of the final peptide product [5]. Each module is composed of a set of semiautonomous domains, each responsible for carrying out a specific catalytic or carrier function [6]. The core domains within a module include the adenylation (A) domain, which is responsible for substrate recognition and activation as an aminoacyl-*O*-AMP at the expense of ATP [7], and its cognate peptidyl carrier protein (PCP) or thiolation (T) domain, to which the substrate is covalently tethered as a thioester to a 4′-phosphopantetheine (Ppant) moiety appended posttranslationally to an invariant serine in a highly conserved region of the T domain [8, 9]. In addition to the A and T domains that comprise the basic units of an initiation module, elongation modules contain condensation (C) domains that catalyze peptide bond formation by facilitating attack of the nucleophilic free amine of the downstream aminoacyl-*S*-Ppant on the upstream nascent peptidyl-*S*-Ppant thioester [10, 11]. Thus, the growing peptidyl chain is transferred to downstream modules in an elongating cascade that culminates in the release of the final product through either hydrolysis or cyclization [1–4].

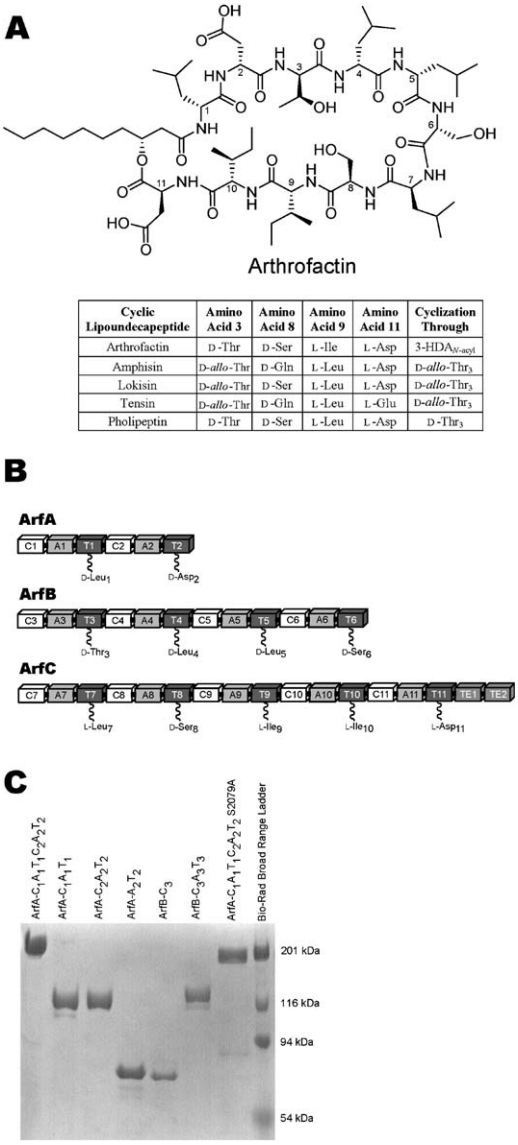
In addition to functionality for monomer recognition, activation, loading, and coupling, NRPSs often contain auxiliary domains that can modify the tethered building blocks, thus contributing significantly to the structural diversity of nonribosomal peptide products [12]. One such example is the epimerization (E) domain. A hallmark of NRPS products is that they may contain nonproteinogenic D amino acid residues. The D configurations may play a functional role by slowing the degradation of NRPS products by naturally L-specific proteases or may serve structural roles by properly orienting side chain conformers for subsequent processing steps [13–16], e.g., in vancomycin. The inclusion of D amino acids increases the variety of nonribosomal peptides produced and is important for bioactivity.

Although one rare mechanism for the incorporation of D amino acids is their direct selection and activation by the A domain [17], D amino acids most often arise in situ by the action of embedded epimerization domains after the L amino acid has been activated by the A domain and covalently tethered to its respective T domain [3, 15, 18]. Epimerization domains are usually found directly after the T domain loaded with the amino acid to be epimerized. This is the case for the NRPSs of gramicidin S, tyrocidine, bacitracin, vancomycin, cephalosporin, and penicillin [14, 19], which are produced by the gram-positive bacteria *Bacillus brevis*, *Bacillus licheniformis*, and *Amycolatopsis orientalis* and the fungi *Cephalosporium acremonium* and *Penicillium chrysogenum*, respectively. In one case, epimerization domains can be found between the A and T domains for the amino acid to be epimerized, a strategy employed by pyochelin synthetase from the gram-negative bacterium *Pseudomonas aeruginosa* [20].

The nonribosomal peptide arthrofactin produced by *Pseudomonas* sp. MIS38, an extremely potent biosurfactant [21], has become the founding member of a new class of cyclic lipoundecapeptides isolated from *Pseudomonas* spp. These molecules—including amphisin, lokisin, and tensin, which possess antifungal properties, and pholipeptin, which possesses enzyme inhibitory activity against phospholipase C—all contain a β-hydroxydecanoyl moiety and 11 amino acids that differ at four positions between the five molecules (Figure 1A). Arthrofactin is necessary for proper biofilm formation and contributes highly to swarming motility, two processes which are of critical importance in bacterial spread and survival [22].

Sequencing of the arthrofactin gene cluster revealed that it encodes a three protein, 11 module NRPS. The three enzymes, termed ArfA, ArfB, and ArfC contain two, four, and five functional modules, respectively, which adhere to the colinearity rule (Figure 1B) [22]. Although this NRPS follows the canonical C-A-T rule for module organization, there is a discrepancy between the encoded genes and the structure of the final

*Correspondence: christopher_walsh@hms.harvard.edu



Cyclic Lipoundecapeptide	Amino Acid 3	Amino Acid 8	Amino Acid 9	Amino Acid 11	Cyclization Through
Arthrofactin	D-Thr	D-Ser	L-Ile	L-Asp	3-HDA$_{N\text{-}acyl}$
Amphisin	D-*allo*-Thr	D-Gln	L-Leu	L-Asp	D-*allo*-Thr$_3$
Lokisin	D-*allo*-Thr	D-Ser	L-Leu	L-Asp	D-*allo*-Thr$_3$
Tensin	D-*allo*-Thr	D-Gln	L-Leu	L-Glu	D-*allo*-Thr$_3$
Pholipeptin	D-Thr	D-Ser	L-Leu	L-Asp	D-Thr$_3$

Figure 1. Arthrofactin Structure and NRPS Organization

(A) Structure of arthrofactin and differences between lipoundecapeptide family members. HDA signifies 3-hydroxydecanoyl acid.

(B) Organization of arthrofactin synthetase.

(C) SDS-7.5% PAGE of the Arf constructs with molecular weight standards.

arthrofactin molecule. Seven of the 11 amino acids in the lipopeptide product are D amino acids, yet there are no epimerization domains in any of the 3 components of the NRPS. In other NRPS systems that synthesize cyclic lipopeptides, such as surfactin [23] and fengycin [24] from *Bacillus subtilis*, there is a corresponding E domain for every D amino acid incorporated into the final nonribosomal peptide product. Furthermore, unlike the biosynthetic gene cluster for cyclosporine synthetase that encodes a dedicated alanine racemase to provide substrate for the D-Ala-specific A domain in the first module [25], the arthrofactin gene cluster does not encode an external racemase, and the first A domain is L-Leu specific, whereas the final product contains D-Leu [22].

Sequence alignment of arthrofactin synthetase with other known NRPSs reveals that it shares the greatest homology with two other lipopeptide synthetases from *Pseudomonas syringae*, the products of which are the phytotoxins syringomycin and syringopeptin. Interestingly, both syringomycin and syringopeptin also contain D amino acids without any epimerization domains present in their respective NRPS modules [26, 27].

In this study, we reconstitute the activity of the first three modules of arthrofactin synthetase in vitro and demonstrate that the epimerization activity is cryptically embedded within the synthetase itself. Using sequence alignment from four NRPS systems that produce peptides containing D amino acids in the absence of epimerization domains, we attribute epimerization activity to a novel subtype of C domain with dual catalytic activity for both epimerization and condensation.

Results

Cloning and Expression of Arf Constructs

Several constructs derived from the first two NRPS components of arthrofactin synthetase, namely, full-length ArfA-$C_1A_1T_1C_2A_2T_2$, ArfA-$C_1A_1T_1$, ArfA-$C_2A_2T_2$, ArfA-A_2T_2, ArfB-C_3, ArfB-$C_3A_3T_3$, and the T_2 mutant ArfA-$C_1A_1T_1C_2A_2T_2$ S2079A, were amplified from a λ5 clone of the arthrofactin gene cluster from *Pseudomonas* sp. MIS38 [22] and were cloned into either N-terminal or C-terminal His-tagged vectors. Sequencing of pArfA-$C_1A_1T_1$ and pArfA-$C_1A_1T_1C_2A_2T_2$ (three independent PCR amplifications for each plasmid) revealed a single nucleotide change of cytosine 581 to adenine, which alters Ala_{194} in the GenBank sequence (accession number AB107223) to Asp. Sequencing of pArfB-$C_3A_3T_3$ (three independent PCR amplifications) revealed two nucleotide changes from the deposited sequence: adenine 2648 to guanine, which alters His_{893} to Arg, and adenine 2994 to guanine, which is silent at Leu_{998}. Prevalence of these mutations throughout all clones implies that the GenBank sequence is likely to be incorrect at these three positions. Expression in *E. coli* at reduced temperature, either 25ºC or 15ºC, with 0.1 mM IPTG induction, yielded 4–6 mg/L protein for all constructs. Using nickel-affinity chromatography and gel filtration in tandem, all proteins were purified to homogeneity (Figure 1C).

ATP-PP$_i$ Exchange Assays to Evaluate A Domain Stereoselectivity

In order to determine whether cryptic epimerase activity exists within one or more domains of arthrofactin synthetase, it was necessary to establish the stereospecificity of the A domains activating the residues that were assigned D chirality in the final arthrofactin molecule. Previously, only the first A domains of ArfA (A_1) and ArfC (A_7), responsible for incorporating D-Leu_1 and L-Leu_7, were purified and shown to possess specificity for activating L-Leu. All other A domains were assigned predicted substrate specificity based on sequence homology to other known A domains and by using the colinearity rule for NRPSs [22]. Furthermore, no kinetic studies had been performed on any of the A domains in arthrofactin synthetase. Employing a standard ATP-PP$_i$ exchange assay, k_{obs} at 1 mM substrate and Michaelis-Menten kinetic parameters were determined for the

Table 1. Michaelis-Menten Kinetics of the First Three A Domains of Arthrofactin Synthetase

Construct	Amino Acid	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m ($mM^{-1}min^{-1}$)
ArfA-$C_1A_1T_1$	L-Leu	0.07 (0.02)	283 (10)	3880 (1200)
	D-Leu	0.30 (0.08)	122 (6)	405 (130)
	L-Ile	6.5 (0.7)	42 (2)	6.5 (1.0)
ArfA-$C_2A_2T_2$	L-Asp	1.6 (0.2)	320 (10)	200 (30)
	D-Asp	>17	CNBD[a]	0.41 (0.01)
	L-Asn	1.5 (0.3)	5.6 (0.3)	4 (1)
ArfB-$C_3A_3T_3$	L-Thr	0.76 (0.09)	133 (4)	175 (25)
	D-Thr	>17	CNBD[a]	0.17 (0.07)
	L-Ser	>17	CNBD[a]	1.8 (0.1)

Values in parentheses represent standard errors.
[a] Could not be determined due to high K_m.

first three A domains in arthrofactin synthetase, each of which activates amino acids that attain the D configuration in the final arthrofactin product.

The first A domain of ArfA (by using ArfA-$C_1A_1T_1$) was shown to activate the following amino acids (k_{obs} at 1 mM): L-Leu (156 min^{-1}), D-Leu (66 min^{-1}), L-Ile (12.8 min^{-1}), L-Val (1.6 min^{-1}), and L-Ala (0.13 min^{-1}). The second A domain of ArfA (by using ArfA-$C_2A_2T_2$) was shown to activate the following amino acids (k_{obs} at 1 mM): L-Asp (143 min^{-1}), D-Asp (2.0 min^{-1}), L-Asn (8.9 min^{-1}), and L-Glu (2.2 min^{-1}). No activity was detected with L-Ile and L-Ala. The second A domain of ArfA was also tested by using the ArfA-A_2T_2 construct, and the following k_{obs} values at 1 mM were obtained: L-Asp (41 min^{-1}) and D-Asp (0.18 min^{-1}). The substrate preference of the ArfA-$C_2A_2T_2$ and ArfA-A_2T_2 are in agreement in terms of which substrate isomer is preferentially adenylated, but the k_{obs} values for the ArfA-A_2T_2 construct are 4- to 10-fold lower than that of the ArfA-$C_2A_2T_2$ construct. This is presumed to be a consequence of truncating the module and exposing the N-terminal portion of the A domain. Therefore, because it was shown that the ArfA-A_2T_2 construct is functional, further kinetics were performed only using ArfA-$C_2A_2T_2$. Interestingly, A_2 from ArfA, which shows 70% homology to A domains that activate L-Ile and only 60% homology to A domains that activate L-Asp [22], showed no activity toward adenylation of L-Ile. The A_3 domain of ArfB (by using ArfB-$C_3A_3T_3$) was shown to activate the following amino acids (k_{obs} at 1 mM): L-Thr (73 min^{-1}), D-Thr (2.6 min^{-1}), L-Ser (8.6 min^{-1}), and L-Val (0.15 min^{-1}). No activity was detected with L-Ala.

From the initial k_{obs} obtained, it was evident that, in all three cases, the A domains showed specificity for activating the L amino acid rather than the D amino acid. This finding was confirmed by determining Michaelis-Menten parameters for the L and D isomers of the cognate amino acid as well as the noncognate amino acid showing the highest k_{obs} at 1 mM substrate (Table 1). The k_{cat} was highest for the cognate L amino acid in all three A domains tested, and, similarly, the K_m was lowest for the cognate L amino acid in all cases. The specificity (k_{cat}/K_m) was highest for the cognate L amino acid over any other substrate in all three cases. In fact, successive A domains within the synthetase appear to become more specific for activation of the L amino acid versus the D amino acid. ArfA-A_1 has a 10-fold higher specificity for L-Leu over D-Leu, ArfA-A_2 has a 485-fold higher specificity for L-Asp over D-Asp, and ArfB-A_3 has a 1030-fold higher specificity for L-Thr over D-Thr. Taken together, these data suggest that the A domains are in fact specific for activation of L amino acids and that epimerizations must occur at subsequent stages of nonribosomal peptide synthesis.

Phosphopantetheinylation and Covalent Amino Acid Loading

Before assaying for intrinsic epimerization activity, it was necessary to determine whether the constructs were competent for covalent thioesterification of the amino acid substrate to the target T domain. Reaction with CoASH and the phosphopantetheinyl transferase Sfp [28, 29] converted the purified apo proteins to post-translationally modified holo forms. Maximal covalent incorporation of ^{14}C-labeled substrate onto the phosphopantetheinyl prosthetic group on the T domains was achieved within 1 min of initiating the adenylation reaction for all constructs, as measured by TCA precipitation followed by liquid scintillation counting or autoradiography of aminoacylated protein (data not shown). With the general catalytic activity of the A domains and reconstitution of the holo forms of the T domains in all constructs verified, condensation and epimerization were then assayed.

Leu_1 Epimerization and Condensation with L-Asp_2

Utilizing the full-length ArfA construct, ArfA-$C_1A_1T_1C_2A_2T_2$, autoaminoacylated with both L-Leu (T_1) and L-Asp (T_2), epimerization of L-Leu to D-Leu and condensation of both L-Leu and D-Leu to L-Asp catalyzed by C_2 was observed (Figure 2A). The data in Figure 2A are from a 10 min incubation with 0.12 nmoles of enzyme active sites. At most, a stoichiometric equilibration of L- and D-Leu and subsequent condensation with L-Asp will occur on the aminoacyl-*S*-pantetheinyl proteins. To enable detection, L-[^{14}C]Leu or L-[^{14}C]Asp were used, and the amino acids were released by thioesterase treatment [30] prior to chiral TLC analysis. When L-Leu was radiolabeled, four spots were observed on the TLC corresponding to D-Leu, L-Leu, D-Leu-L-Asp, and L-Leu-L-Asp, in ascending order (Figure 2A). When L-Asp was radiolabeled, three spots were observed on the TLC corresponding to L-Asp, D-Leu-L-Asp, and L-Leu-L-Asp, in ascending order (Figure 2A). It should be noted that the diffuse spot attributed to D-Leu-L-Asp is thought to result from different protonation states of the L-Asp moiety on the dipeptide and was observed intermittently in both reactions and standards. The difficulty in controlling the Asp protonation state made subsequent analysis of the Asp_2 epimerization by TLC unsuccessful and prompted a change to HPLC evaluation.

In contrast, when full-length ArfA was loaded with only L-Leu, epimerization was not observed, and obviously no dipeptide was formed (Figure 2A). Incubation of the first module, ArfA-$C_1A_1T_1$, with either only L-Leu or both L-Leu and L-Asp also did not yield any epimerization (Figure 2A). This result implies that epimerization requires an aminoacylated downstream T domain that is primed for peptide bond formation.

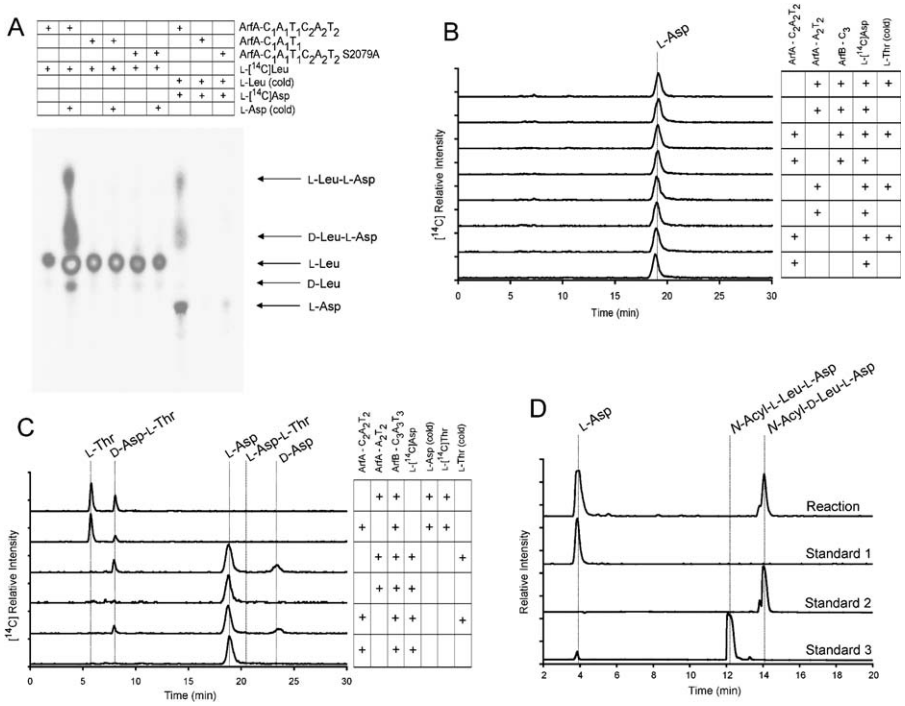


Figure 2. Epimerase Activity Embedded within Arthrofactin Synthetase

(A) Chiral radioTLC images depicting various reactions performed by using L-[^{14}C]Leu loaded onto T_1 of ArfA or L-[^{14}C]Asp loaded onto T_2 of ArfA.

(B) Chiral radioHPLC traces depicting various reactions performed by using L-[^{14}C]Asp loaded onto T_2 of ArfA.

(C) Chiral radioHPLC traces depicting various reactions performed by using L-[^{14}C]Asp loaded onto T_2 of ArfA or L-[^{14}C]Thr loaded onto T_3 of ArfB.

(D) RadioHPLC trace of a 70/30 mixture of *N*-acyl-L-Leu-CoA/CoASH loaded onto ArfA-$C_1A_1T_1C_2A_2T_2$ and reacted after covalent aminoacylation of L-[^{14}C]Asp onto T_2 (Reaction). Standard 1 was monitored by ^{14}C radioactive counts; Standard 2 and Standard 3 were monitored at 220 nm.

Detailed reaction conditions for (A), (B), (C), and (D) are described in Experimental Procedures.

In order to demonstrate that L-Asp must be covalently loaded on the downstream T domain and not simply in solution, and to test whether the cryptic epimerase activity could be triggered with free amino acid rather than a downstream aminoacyl-T domain, a full-length ArfA T_2 knockout construct was created. This construct is ArfA-$C_1A_1T_1C_2A_2T_2$ S2079A, in which the conserved phosphopantetheinylated serine in the T_2 domain was mutated to alanine, thus irreversibly confining the second module to the apo form. Upon incubation with either L-Leu alone or both L-Leu and L-Asp, no epimerization and no condensation were observed (Figure 2A). This finding supports the notion that the second module must be loaded with the downstream aminoacyl thioester for the first amino acid to be epimerized.

Asp_2 Epimerization and Condensation with L-Thr_3

Similar to the results observed for the Leu_1 epimerization reaction, Asp_2 loaded onto T_2 of ArfA was only epimerized when it interacted with ArfB-$C_3A_3T_3$ loaded with L-Thr. When ArfA-$C_2A_2T_2$ or ArfA-A_2T_2 was covalently loaded with L-Asp in the presence or absence of L-Thr, no epimerization to D-Asp was observed in the chiral radioHPLC traces (Figure 2B). Correspondingly, interaction of L-Asp-loaded ArfA-$C_2A_2T_2$ or ArfA-A_2T_2 with ArfB-C_3 in the presence or absence of L-Thr or ArfB-$C_3A_3T_3$ in the absence of L-Thr also yielded no epimerization to D-Asp (Figure 2B). However, when ArfA-$C_2A_2T_2$ or ArfA-A_2T_2 loaded with L-Asp interacted with ArfB-$C_3A_3T_3$ loaded with L-Thr, epimerization to D-Asp was observed, and only the D-Asp-L-Thr dipeptide was formed (Figure 2C).

Formation of only the D-Asp-L-Thr dipeptide was confirmed by repeating the experiments in the reciprocal manner, with L-Thr radiolabeled instead of L-Asp radiolabeled (Figure 2C). These results support a mechanism in which epimerization precedes condensation to a downstream amino acid but only occurs if the downstream amino acid is loaded onto its T domain and primed for condensation. It strongly suggests that ArfB C_3 is a $^{D}C_{L}$ catalyst.

The β-Hydroxyacyl Moiety *N*-Acylated to Leu_1 Confers Specificity for D-Leu-L-Asp Condensation

As described above, epimerization of the upstream amino acid for the first two modules of arthrofactin synthetase requires that the downstream amino acid be loaded onto its T domain and be presented simultaneously to the upstream C domain. However, a difference was observed in the condensation products produced by the two condensation domains involved in the process. Whereas the ArfB C_3 domain formed only the D-Asp-L-Thr dipeptide product, the ArfA C_2 domain formed both the L-Leu-L-Asp and D-Leu-L-Asp

dipeptides. We decided to test whether the β-hydroxy-*N*-acyl group, presumed to be condensed with Leu_1 at the initial stages of arthrofactin formation, could enhance specificity for $^{D}C_{L}$ catalysis by the ArfA C_2 domain.

Initially, the reaction was attempted in *trans* with the ArfA-$C_1A_1T_1$ and ArfA-$C_2A_2T_2$ constructs rather than the full-length ArfA-$C_1A_1T_1C_2A_2T_2$ construct because Sfp is indiscriminate as to which T domain it phosphopantetheinylates. However, the in *trans* reaction did not yield any epimerization or condensation products when tested with the positive control of Sfp priming with CoASH followed by loading of L-Leu and L-Asp. Therefore, the reaction was performed with the full-length ArfA-$C_1A_1T_1C_2A_2T_2$ construct and a 70/30 mixture of 3-hydroxybutyryl-L-Leu-CoA (*N*-acyl-L-Leu-CoA) and CoASH, respectively. Skewing of the mixture away from 50/50 was done because Sfp more readily accepts free CoASH as a substrate than peptidyl-CoAs [31]. This reaction produces three undesired species—both T domains loaded with CoASH, both T domains loaded with *N*-acyl-L-Leu-CoA, and T_1 loaded with CoASH and T_2 with *N*-acyl-L-Leu-CoA—as well as the desired product—T_1 loaded with *N*-acyl-L-Leu-CoA and T_2 loaded with CoASH. Although there are four species, this does not complicate analysis of condensation products because only L-Asp is radiolabeled. Therefore, only the two species loaded with CoASH on their second Ts will be labeled in the autoaminoacylation reaction by A_2. In turn, the Asp radiolabel will identify any *N*-acyl-Leu-Asp condensation product.

Analysis of the products released from T_2 by LiOH hydrolysis and assayed as the acylated dipeptides established that only *N*-acyl-D-Leu-L-Asp was formed, with no *N*-acyl-L-Leu-L-Asp observed (Figure 2D). This finding demonstrates that the acyl group appended to the Leu_1 amino group confers specificity for $^{D}C_{L}$ catalysis by C_2 of ArfA. Taken together with data from the previous experiments, these data suggest that epimerization occurs after condensation to the upstream peptidyl-*S*-T domain, but before condensation to the downstream aminoacyl-*S*-T domain.

Identification of C Domains with Possible Dual Activity for Peptidyl Condensation and Epimerization

Previous bioinformatic analysis of C domain primary sequences has led to the identification of two subcategories, one of which directs the condensation of two L amino acids, termed $^{L}C_{L}$ catalysts for the L/L peptide bond formed, and one that is found directly after a traditional epimerase domain that catalyzes condensation of an upstream D-aminoacyl thioester with a downstream L-aminoacyl thioester, termed $^{D}C_{L}$ catalysts for the D/L peptide bond formed [32, 33]. In this study, we find a new, to our knowledge, subclass of C domains that follow directly after T domains, as in the canonical NRPS architecture without E domains present [34]; however, the amino acid loaded onto that T domain appears as a D isomer in the final product. As noted in Figure 3, these C domains are found in four NRPS assembly lines that have been annotated and sequenced: arthrofactin [22], syringomycin [26], syringopeptin [27], and ramoplanin [35] synthetases. These C domains are proposed to have dual catalytic roles for condensation and epimerization and to be $^{D}C_{L}$ catalysts.

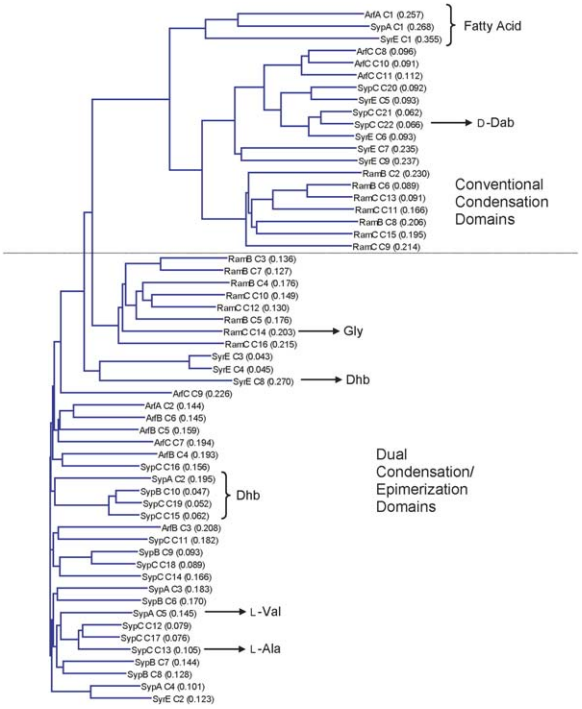


Figure 3. Phylogenetic Analysis of C Domains from Arthrofactin, Syringomycin, Syringopeptin, and Ramoplanin Synthetases

Those predicted to be conventional C domains are above the dotted line, and those predicted to be dual C/E domains are below, with exceptions noted by identification of the upstream donor substrate. Clustering was performed in the Vector NTI program by using the neighbor-joining algorithm of Saitou and Nei [45]. Numbers denote the distance from a common ancestor.

Sequence alignment of each individual C domain from these four NRPS systems generates a phylogenetic tree (Figure 3) that correctly predicts and clusters those C domains that simply carry out condensation and those that have dual roles for condensation and epimerization, with three exceptions: C_{22} from syringopeptin synthetase C, which is not predicted to have dual C/E activity but is found following a D-diaminobutyrate (Dab) residue, and C_5 from syringopeptin synthetase A and C_{13} from syringopeptin synthetase C, which are predicted to have dual C/E activity but are found following L-Val and L-Ala residues, respectively (Figure 3). Sequence alignment of the A domains and T domains from these systems shows no such pattern or clustering of D amino acids away from L amino acids when analyzing either the amino acid activated by the A domain, the amino acid one module upstream of the A domain (Figure S1; see the Supplemental Data available with this article online), or the amino acid loaded onto the T domain.

Analysis of the primary sequence of the proposed dual C/E domains reveals that they contain an N-terminal sequence that, to our knowledge, is not present in the other sets of C domains. This sequence is composed of ~50 amino acids that terminate in a slightly elongated His motif conforming to the sequence HHI/LxxxxGD (Figure 4B). This elongated His motif is present in addition to the conventional His motif (and arginine

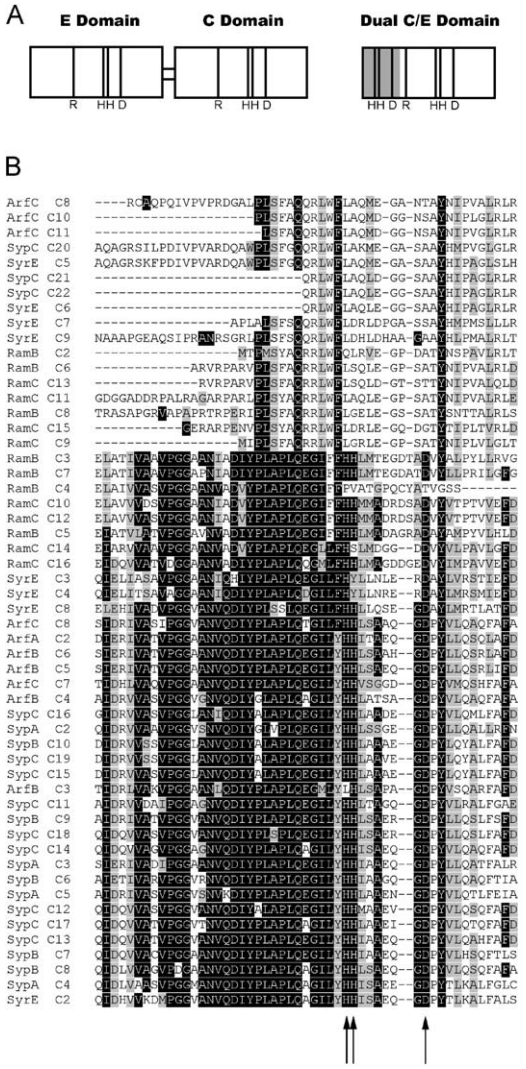


Figure 4. Unique N-Terminal Sequence Present in C Domains with Dual Condensation and Epimerization Activity

(A) Depiction of the conserved Arg, His, and Asp triad found in all E, C, and dual C/E domains. The unique N-terminal sequence in C/E domains is highlighted in gray with the secondary elongated His motif identified.

(B) Sequence alignment of the N-terminal regions of C domains from arthrofactin, syringomycin, syringopeptin, and ramoplanin synthetases indicating a unique N-terminal sequence in proposed dual C/E domains, which are grouped in the bottom half of the alignment, starting with RamB C3. All C domains between ArfC C8 and RamC C9 are conventional C domains. Alignments were performed in the Vector NTI program suite. Residues highlighted in black are highly conserved, and residues highlighted in gray are moderately conserved. Arrows identify possible catalytic residues in the elongated His motif.

residue) found in all known C domains (Figure 4A). The second His and terminal Asp in the His motif have been shown to be critical for catalysis in both traditional condensation domains and epimerase domains [15, 36].

Discussion

One feature of certain nonribosomal peptide synthetases is that they incorporate D amino acid residues into their peptide products. This is the case for the siderophore pyochelin, the antibiotics penicillin, vancomycin, and tyrocidine, the biosurfactant surfactin, and the immunosuppressant cyclosporin A [3, 37]. There have been two characterized mechanisms by which D amino acids can be incorporated into nonribosomal peptides. The free D amino acid can be directly selected for activation by a corresponding A domain and loaded onto the module responsible for its incorporation [3, 15, 18], as is the case for D-Ala_1 of cyclosporin synthetase [25]. However, the more conventional method entails epimerization of proteinogenic and nonproteinogenic L amino acids on their respective T domains during initiation and elongation by embedded adjacent 50 kDa epimerase domains, which act in *cis* on aminoacyl-*S*-Ppant and peptidyl-*S*-Ppant substrates, respectively [3, 15, 18]. For example, the nonribosomal peptide tyrocidine has two D-Phe residues at positions one and four in the heptapeptide scaffold, and there are E domains in module one and four of tyrocidine synthetase from the gram-positive *Bacillus brevis* [32].

The lipoundecapeptide arthrofactin is an analogous NRPS product in which 7 of the 11 amino acids are of the D configuration. However, sequencing of the biosynthetic gene cluster from the gram-negative *Pseudomonas* sp. MIS38 revealed that there were no epimerase domains present in the system [22]. In this study, we have demonstrated that the first three A domains in arthrofactin synthetase responsible for incorporation of D-Leu_1, D-Asp_2, and D-Thr_3 are specific for activating the L amino acid isomer. Michaelis-Menten kinetics revealed that the catalytic specificity for activation of the L amino acid substrate versus the D amino acid substrate increases from 10-fold to over 1000-fold from module 1 to module 3. This indicates that epimerization to the D isomer must occur after activation and loading of the L amino acid onto its cognate T domain.

Although it was initially postulated that external racemases must catalyze such a transformation [22], we have shown that the epimerase activity is actually cryptically embedded within the relevant modules of the synthetase. Sequence alignment of arthrofactin synthetase with three other systems that produce nonribosomal peptides containing D amino acids in the absence of epimerase domains (syringomycin, syringopeptin, and ramoplanin synthetases) reveals that in addition to the standard variety of condensation domain, these systems contain a novel, to our knowledge, condensation domain that can be discerned by analysis of primary amino acid sequence (Figure 3). These condensation domains always lie directly downstream of the T domain initially acylated with the amino acid that is epimerized in the final product in all four systems [34]. Therefore, we postulate that these C domains actually possess dual activity for condensation and epimerization.

Of critical importance to the enzymology of NRPS assembly lines is the proper timing and coordination of the several processing steps that must occur on each T-tethered intermediate as the peptide chain grows. In addition to avoiding misinitiation of noncognate amino acids by the A domain, the assembly line must avoid stalling and premature hydrolytic termination of peptidyl-*S*-T intermediates. C domains are presumed to act as key gate keepers controlling the flow of

elongating, covalently tethered intermediates in NRPSs. When epimerization domains are present, an important interplay must occur between the upstream and downstream condensation domains in order to prevent stalling due to failed recognition of the correct isomer to be elongated into the growing nonribosomal peptide product [14, 19, 32, 38].

Generally, it appears that condensation domains always incorporate a downstream L amino acid monomer into the growing peptide chain [32]. If the condensation domain comes after an epimerization domain, then it is a $^{D}C_{L}$ catalyst, and it will only accept a carboxyterminal D-aminoacyl thioester on the donor peptidyl chain for condensation, as is the case for tyrocidine, gramicidin, actinomycin, and pyochelin synthetases [14, 20, 32, 38]. Otherwise, condensation domains act as $^{L}C_{L}$ catalysts. This paradigm has the implication that unless epimerization domains are found in initiation modules, as is the case for gramicidin (GrsA) and tyrocidine (TycA) synthetases, which begin with A-T-E modules, epimerization domains act on the peptidyl-*S*-T intermediates rather than the aminoacyl-*S*-T monomers [16, 19, 32, 38]. Thus, E domains may serve as secondary gate keepers controlling the kinetics of downstream chiral condensation domains. No $^{D}C_{D}$ condensation domains have been reported.

The use of separate epimerization and chiral condensation domains to control chain elongation in NRPSs raises the interesting question of how seven modules in arthrofactin synthetase, utilizing a single domain with dual catalytic activity, are able to control the timing of epimerizations and condensations to ensure that the proper amino acid isomers are condensed and elongated. Unlike conventional condensation domains that follow epimerization domains with chiral specificity for a single enantiomer, these dual C/E domains must interact with both the L and D forms of the donor aminoacyl groups they have to incorporate into the growing chain.

Analysis of epimerization of Leu_1 and Asp_2 by arthrofactin synthetase reveals that epimerization precedes condensation to a downstream amino acid, as shown by the formation of uncondensed D amino acid, and that subsequent condensation is $^{D}C_{L}$ specific, just as it is for the conventional subcategory of C domains that follow standard epimerization domains. When using ArfA-$C_2A_2T_2$ or ArfA-A_2T_2 as an initiation module for condensation of Asp_2 to Thr_3, only the D-Asp-L-Thr dipeptide is formed. However, when using ArfA-$C_1A_1T_1$ as an initiation module for condensation of Leu_1 to Asp_2, both the D-Leu-L-Asp and L-Leu-L-Asp dipeptides are formed. ArfA does not begin with a standard A-T initiation module, but rather contains a full C-A-T organization like those normally found in elongation modules [22]. This arrangement suggests that Leu_1 is acylated with a fatty acid before peptide elongation begins. If this is the case, then perhaps the acyl group is necessary to confer specificity for $^{D}C_{L}$ catalysis by the ArfA C_2 domain. A similar case has been witnessed before in tyrocidine synthetase. Condensation of L-Phe loaded onto TycB$_3$-AT$_4$E with L-Asn loaded onto TycC$_1$-C$_5$AT$_5$ yields a 2:1 mixture of the L-L and D-L dipeptides, respectively. However, when the full natural tetrapeptidyl substrate, D-Phe-L-Pro-L-Phe-L-Phe, was loaded onto TycB$_3$-AT$_4$E, condensation with L-Asn yielded only the expected $^{D}C_{L}$ product, D-Phe-L-Pro-L-Phe-D-Phe-L-Asn [32, 38]. A similar switch to full $^{D}C_{L}$ chirality in ArfA module 1 was detected by using 3-hydroxybutyryl-L-Leu-*S*-T_1 as a donor. The chiral selectivity of a $^{D}C_{L}$ peptide condensation domain may be governed by the identity of the donor. Moreover, this result has strong implications for the timing of acylation in the synthesis of arthrofactin. If the acyl chain is necessary for chiral specificity of the downstream C domain, then Leu_1 is most likely acylated before epimerization and condensation to Asp_2. The acylation reaction could be catalyzed by the first C domain of ArfA based on homology with the quinoxaline biosynthesis system in which functional crosstalk between fatty acid synthases and NRPS was shown [39].

We have not yet dissected relative rates of epimerization of L-aminoacyl-*S*-T domains, nor have we dissected relative rates of subsequent stereoselective condensation of the D-aminoacyl-*S*-T-donor in the epimerization equilibrium, much less absolute rates. Such studies will require rapid quench, single turnover studies.

In addition to calibrating the timing of epimerization and chirality of condensation, this study shows that the epimerization reaction does not take place unless the downstream T domain is loaded with the amino acid to which the epimerized product is subsequently condensed. Epimerization of Leu_1 by ArfA-C_2 does not occur unless Asp_2 is loaded onto ArfA-T_2, and epimerization of Asp_2 by ArfB-C_3 does not occur unless Thr_3 is loaded onto ArfB-T_3. Such a gating mechanism implies that these dual C/E domains assume three conformational states that are populated dynamically depending on the availability of the upstream donor and downstream acceptor substrates (Figure 5A). First, there exists a resting C/E state that is capable of interacting with both the upstream and downstream T domains loaded with their cognate L amino acids, but is incompetent for either condensation or epimerization. Once the upstream peptidyl-*S*-Ppant and downstream aminoacyl-*S*-Ppant are present and interacting with the dual C/E domain, a C/E′ conformation arises in which the domain becomes competent for epimerization of the upstream aminoacyl thioester. Finally, in the presence of the upstream D-aminoacyl-*S*-T, a C/E″ conformation develops that is now condensation competent, resulting in elongation of the peptidyl chain with D/L chirality. Mechanistic studies remain to be conducted, but it is likely that removal of the Cα-*H* of the donor L-aminoacyl-*S*-T_{n-1} occurs only when the downstream acceptor substrate is presented, perhaps to drive a conformational change in the C domain (Figure 5B).

It is of interest to note that in the phylogenetic tree of C domains from arthrofactin, syringomycin, syringopeptin, and ramoplanin synthetases, the C domains that lie downstream of T domains that are responsible for incorporating 2,3-dehydroaminobutyric acid (Dhb) cluster with the dual C/E domains (Figure 3), even though these are not chiral residues. Evaluation of the A domains that supposedly activate these amino acids show that they cluster with those A domains responsible for activating β-branched amino acids, like L-Val, L-Thr, and L-Ile (Figure S1). Furthermore, radioactive feeding experiments done in *Pseudomonas syringae* demonstrate that these Dhb moieties are derived from

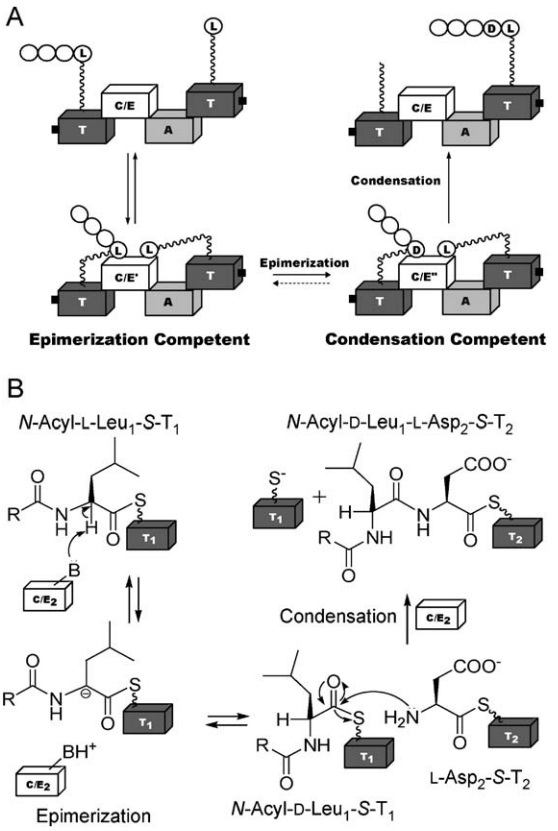


Figure 5. Proposed Conformational State and Mechanism for the Dual Condensation and Epimerization Activity of C/E Domains

(A) Scheme depicting the three conformational states of dual C/E domains. When the resting C/E state interacts with the upstream peptidyl-*S*-Ppant donor and the downstream aminoacyl-*S*-Ppant acceptor, an epimerization-competent C/E′ state arises, which subsequently leads to a condensation-competent C/E″ conformation after epimerization. It is unknown whether the epimerization reaction is reversible.

(B) Chemical schematic for epimerization of the donor aminoacyl/peptidyl-*S*-T domain to the D isomer utilized for condensation with the downstream L-aminoacyl-*S*-T domain.

L-Thr [40]. Therefore, it is possible that these dual C/E domains could also function as dual condensation/dehydration domains with or without prior epimerization. It will be of interest to see if a D-threonyl-*S*-T gets selectively dehydrated and then condensed.

In gram-positive NRPS assembly lines, it has been noted that E domains and C domains are homologous and that E domains may have evolved from C domains by domain duplication [41] followed by divergent evolution. E domains in the gram-positive assembly lines are often at the C termini of subunits and may pair up with the downstream C domains for specificity. This snapshot of the gram-negative NRPS assembly lines may reveal epimerization-competent condensation domains before C to E domain duplications arose. Evidence for a similar paradigm can be found in the VibF subunit of vibriobactin synthetase. Whereas cyclization (Cy) domains normally catalyze both heterocyclization and condensation, VibF contains tandem Cy domains, with one only capable of catalyzing condensation and the other only capable a catalyzing heterocyclization [42].

One eventual goal of NRPS research is the combinatorial swapping of modules into engineered synthetases to produce novel biologically active natural products. If one wishes to incorporate D amino acids into these new compounds utilizing standard epimerization domains, it will also be necessary to swap in a corresponding $^{D}C_{L}$ chiral condensation domain. Furthermore, it has been demonstrated that T domains upstream of epimerization domains show signature differences from standard T domains in the residues surrounding the phosphopantetheinylated conserved Ser, most notably adopting a DSI sequence in the CoreT motif rather than the standard HSL sequence [19]. Therefore, in constructing chemical libraries by using engineered nonribosomal peptide synthetases, one would have to swap in at least three unique domains (T_E-E-C_E) for each D amino acid incorporated. The novel, to our knowledge, dual C/E domains described in this study should serve as usefully economic alternatives for incorporating D amino acids because they integrate the catalytic functions of two domains into one, and they lie downstream of what appears to be the standard variety of T domain conforming to the HSL sequence.

Significance

D amino acid residues are key constituents of many bioactive nonribosomal peptides. Free D amino acid levels are low in microbial NRP producers. Typically, L amino acid monomers are activated and epimerized on the multimodular NRPS assembly lines. While distinct 50 kDa epimerase (E) domains have previously been found in NRPS modules of several strains (mostly gram-positive), there are no comparable E domains in NRPS assembly lines from multiple gram-negative *Pseudomonas* strains. We demonstrate that for the pseudomonal arthrofactin synthetase modules, which activate L amino acids and epimerize them as covalently tethered pantetheinyl thioesters, the condensation (C) domains have both peptide bond-forming condensation and upstream aminoacyl-*S*-pantetheinyl donor epimerization activities. Dual C/E condensation domains may have been precursors to separate E domains and offer promise as portable domains for the generation of D amino acid residues in combinatorial biosynthesis.

Experimental Procedures

Materials and General Methods

Standard recombinant DNA, molecular cloning, and microbiological procedures were performed as described [43]. Competent Top10 and BL21 (DE3) *E. coli* strains were purchased from Invitrogen. Oligonucleotide primers were purchased from Integrated DNA Technologies. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. Herculase DNA polymerase was purchased from Stratagene. Plasmids pET28b and pET37b were purchased from Novagen. DNA sequencing to verify PCR fidelity was performed on double-stranded DNA by the Molecular Biology Core Facilities of the Dana Farber Cancer Institute (Boston, MA). Plasmid DNA preparation was performed by using the Qiaprep kit from Qiagen, PCR cleanup was carried out by using the Qiaquick kit from Qiagen, and gel extraction of DNA fragments as well as restriction endonuclease cleanup were done by using the GFX kit from GE Healthcare. Ni-NTA Superflow resin was from Qiagen. FPLC purification of proteins was performed by using a HiLoad 26/60

Superdex 200 prep grade column run on a P-920 pump equipped with a UPC-900 detector and a Frac-950 fraction collector (GE Healthcare) with a running buffer of 20 mM Tris (pH 8.0), 50 mM NaCl, 2 mM $MgCl_2$, and 1 mM DTT. SDS-PAGE gels were from Bio-Rad. Protein samples were concentrated by using a 10KMWCO Amicon Ultra device from Millipore, and final protein concentrations were calculated by using the protein's absorbance at 280 nm and the predicted molar extinction coefficient.

HOBt, HBTU, Pybop, Boc- and Fmoc-protected amino acids, and 2-chlorotrityl resin were purchased from NovaBiochem. Preparative HPLC was performed on a Beckman Coulter System Gold instrument with a Vydac Proteins and Peptides C18 column (10 μm, 22 × 250 mm). LCMS identification was carried out on a Shimadzu LCMS-QP8000α equipped with two LC-10AD*VP* liquid chromatography pump modules, a SPD-10AV*VP* UV-vis detector, a SIL-10AD*VP* autosampler module, and a Vydac C18 Mass Spec column (5 μM, 2.1 × 250 mm).

Radiolabeled L-[^{14}C]Leu (331 mCi/mmol) was purchased from Perkin-Elmer Life Sciences; D-[^{14}C]Leu (55 mCi/mmol), [^{32}P]PP_i, L-[^{14}C]Asp (208 mCi/mmol), and D-[^{14}C]Asp (55 mCi/mmol) were purchased from American Radiolabeled Chemicals, Inc. (ARC), and L-[^{14}C]Thr (184.2 mCi/mmol) was purchased from Sigma. Purified *Bacillus subtilis* phosphopantethienyl transferase Sfp [28, 29] was provided by M. Fischbach. Purified *Bacillus brevis* TycF [30] was provided by E. Yeh. BioMax centrifugal filters were purchased from Millipore. Chiral silica TLC plates from Aldrich were exposed to a BAS-IIIs image plate for approximately 24 hr and subsequently read by a Typhoon 9400 variable mode imager (GE Healthcare) followed by analysis with ImageQuant Software. Radio-HPLC was performed by using a Beckman Coulter System Gold instrument equipped with a β-Ram module 3 radioisotope detector (IN/US Systems).

All other chemicals and HPLC solvents were purchased from Sigma-Aldrich and were used without further purification.

Cloning of Arf Genes, Overproduction, and Purification of Arf Proteins

All constructs were obtained from PCR amplification of a λ5 clone of the arthrofactin gene cluster, which was a gift from Dr. Masaaki Morikawa, Osaka, Japan [22]. This template was amplified by using the following oligonucleotide primers (italics: modified sequences, underlined: restriction site): ArfA-$C_1A_1T_1$: 5′-*GGAATTC*CATATGCC GCCTATTTCTGCCCCTGCGG-3′ and 5′-*CG*GAATTC*ATCA*ATGGGT GCAATCGGCGGCGATGCG-3′; ArfA-$C_1A_1T_1C_2A_2T_2$: 5′-*GGAATTC* CATATGCCGCCTATTTCTGCCCCTGCGG-3′ and 5′-*CCG*CTCGAG CCGGCGTTTGCGATTGAGGCTC-3′; ArfA-$C_2A_2T_2$: 5′-*GGAATTC* CATATGATCACTCCGGCAATGCTGCCGCTG-3′ and 5′-*CCG*CTCG AGCCGGCGTTTGCGATTGAGGCTC-3′; ArfA-A_2T_2: 5′-*GGAATTC* CATATGGAGGTCGATTGTCCGCTGGAC-3′ and 5′-*TGCCC* AAGCTT*CTACTA*GAGCCGGCGTTTGCGATTGAG-3′; ArfB-C_3: 5′-*GGAATTC*CATATGCATTTCAGCGAACTGATGGC-3′ and 5′-*CCG* CTCGAGGTAGCTCAGCGTCTGCTCTTCGAAC-3′; ArfB-$C_3A_3T_3$: 5′-*GGAATTC*CATATGCATTTCAGCGAACTGATGGC-3′ and 5′-*CCG* CTCGAGGTAAATATCCTGCACATTGCCGAC-3′. ArfA-$C_1A_1T_1C_2A_2T_2$, ArfA-$C_2A_2T_2$, ArfB-C_3, and ArfB-$C_3A_3T_3$ PCR products were digested with NdeI/XhoI and ligated into similarly digested pET37b to create C-terminal His8-tagged constructs, the ArfA-$C_1A_1T_1$ PCR product was digested with NdeI/EcoRI and ligated into similarly digested pET28b to create an N-terminal His6-tagged construct, and the ArfA-A_2T_2 PCR product was digested with NdeI/HindIII and ligated into similarly digested pET28b to create an N-terminal His6-tagged construct.

The described expression plasmids were transformed into *E. coli* BL21 (DE3)-competent cells. The four constructs larger than two domains were grown at 15ºC, and the other two constructs were grown at 25ºC in Luria-Bertani media supplemented with 5 mM $MgCl_2$ and 40 μg/ml kanamycin to an OD_{600} of 0.5–0.8 when the culture was induced with 100 μM IPTG and then grown for an additional 12–36 hr. The cells were harvested by centrifugation at 4000 × g for 16 min and stored as pellets at −80ºC until further use. Cell pellets from 4 L of culture were thawed and resuspended in 35 ml buffer (25 mM Tris [pH 8.0], 500 mM NaCl) and then lysed with two passes on an Emulsiflex-C5 cell disruptor (Avestin). The lysate was cleared by ultracentrifugation at 95000 × g for 35 min and then transferred to 700 μl Ni-NTA resin for incubation at 4ºC for 2 hr. The resin was then transferred to a column, and the protein was eluted with an imidazole gradient by using steps of 10 ml of 0 and 5 mM imidazole and then 5 ml of 25 and 200 mM imidazole mixed into lysis buffer. After running an SDS-PAGE gel to verify which fractions contained the protein, the 25 and 200 mM imidazole fractions were dialyzed overnight against 20 mM Tris (pH 8.0), 50 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT, and 10% glycerol. The dialyzed protein was then concentrated to 1.5 ml and subjected to gel filtration purification. After confirming which fractions contained the protein of interest by using SDS-PAGE, the resulting fractions were concentrated, supplemented with 10% glycerol, flash frozen in liquid nitrogen, and stored at −80ºC.

Mutagenesis and Purification of ArfA-$C_1A_1T_1C_2A_2T_2$ S2079A

The gene for the full-length ArfA second T domain knockout mutant was constructed via the splicing by overlap extension (SOE) method [44] by using the pET37b ArfA-$C_1A_1T_1C_2A_2T_2$ plasmid as template. In the first round of PCR amplification, the 5′ fragment of the mutant was amplified by using the primers (underlined: restriction site, bold: mutation) 5′-CCTGCCATTCGGCCTGCAGGAAGTGCAGGGC GACGG-3′ and 5′-CACCGCCAGCAA**CGC**ATGCCCACCGAG-3′, and the 3′ fragment of the mutant was amplified by using the primers 5′-CTCGGTGGGCAT**GCG**TTGCTGGCGGTG-3′ and 5′-CTCGAATCTCGAGGGTGGCGTTGAACGTGATCAG-3′. After PCR purification, the two fragments were mixed together and further amplified by using the forward first primer from the 5′ fragment and the reverse second primer from the 3′ fragment. The final PCR product was digested with SbfI/XhoI and ligated into a similarly digested pET37b ArfA-$C_1A_1T_1C_2A_2T_2$ plasmid. The mutant ArfA enzyme was purified as described for the wild-type enzyme.

ATP-PP_i Exchange Assay for ArfA Domain Substrate Specificity

Reactions (100 μl) contained 75 mM Tris (pH 7.5), 10 mM $MgCl_2$, 5 mM DTT, 5 mM ATP, 1 mM sodium [^{32}P]pyrophosphate (0.18 μCi), 100 μg/ml BSA with 0.05 or 0.1 μM enzyme or only 1 μM enzyme, and substrate amino acid concentrations ranging from 0.05 to 17 mM. Reactions were incubated at room temperature for either 5 or 10 min and were quenched by the addition of 500 μl 1.6% (w/v) activated charcoal, 200 mM tetrasodium pyrophosphate, and 3.5% perchloric acid in water. The charcoal was pelleted by centrifugation and washed twice with 500 μl 200 mM tetrasodium pyrophosphate and 3.5% perchloric acid in water. The radioactivity bound to the charcoal was then measured by liquid scintillation counting. Note that enzyme concentrations and reaction times were chosen such that ATP-PP_i exchange remained under 15% of equilibrium levels.

Analysis of Phosphopantetheinylation of Apo T Domain Constructs

Reactions (20 μl) to determine whether the T domains were functional for phosphopantetheinylation contained 50 mM Tris (pH 7.5), 5 mM $MgCl_2$, 1 mM DTT, 100 μM CoASH, 10 μM A domain-containing enzyme, and 3 μM Sfp. After incubation at room temperature for 1 hr, amino acid loading onto the T domain was initiated with 300 μM ^{14}C radiolabeled substrate and 5 mM ATP. At various time points ranging from 2 s to 10 min, reactions were quenched into 100 μl 10% TCA with 5 μl 40 mg/ml BSA. Samples were centrifuged and washed twice with 200 μl 10% TCA, and the protein pellet containing covalently bound radiolabeled amino acid was resolubilized in 200 μl formic acid and quantified by liquid scintillation counting. Alternatively, the reactions were quenched into 20 μl of 2× SDS-PAGE running buffer (without reducing agent), then loaded and run on a 5% polyacrylamide gel. The gel was then dried down, exposed to a phosphorimager plate overnight, and then scanned to see if radioactivity was bound to the protein band.

Synthesis of Dipeptide Standards: L-Leu-L-Asp, D-Leu-L-Asp, L-Asp-L-Thr, and D-Asp-L-Thr

0.33 mmol H-Asp(OtBu)-2-ClTrt resin or 0.33 mmol H-Thr(OtBu)-2-ClTrt resin was swelled in DMF then reacted with 3 equivalents of Boc-D-Leu-OH/Boc-L-Leu-OH or Boc-D-Asp(OtBu)-OH/Boc-L-Asp(OtBu)-OH, respectively, 3 equivalents of HBTU, 1 equivalent of HOBt, and 12 equivalents of DIEA dissolved in 5 ml DMF while

shaking at room temperature for 2 hr. Afterwards, the resin was washed three times with 3 ml DMF and three times with 3 ml CH_2Cl_2, followed by deprotection and cleavage from the resin overnight by using 2 ml 90/5/5 TFA/$H_2$0/TIS. The TFA was rotovapped off, and the product was redissolved in H_2O and purified by preparative HPLC with a gradient of 0%–25% acetonitrile over 25 min starting in 0.1% TFA in $H_2$0. Product peaks were identified by LCMS and lyophilized to dryness: L-Leu-L-Asp 247.13 [(M + H^+)] calculated, 246.90 observed; D-Leu-L-Asp 247.13 [(M + H^+)] calculated, 246.90 observed; L-Asp-L-Thr 235.09 [(M + H^+)] calculated, 234.80 observed; D-Asp-L-Thr 235.09 [(M + H^+)] calculated, 234.85 observed.

Synthesis of *t*-Butyl-*R*-3-Hydroxybutyrate

R-3-hydroxybutyrate methyl ester (16.93 mmol) was dissolved in 28.22 ml CH_2Cl_2 and sealed in an airtight vial to which 295 μl H_3PO_4 (made by adding 0.4 g P_2O_5 to 1.1 ml 86% H_3PO_4 to remove the water), 703 μl $BF_3 \cdot Et_2O$, and 14.2 ml isobutylene was added in respective order. The reaction was mixed at −75ºC for 2.5 hr and then overnight at room temperature. The resulting reaction mixture was dissolved in 100 ml ethyl acetate, washed twice with 50 ml saturated sodium bicarbonate and once with 50 ml NaCl brine. The product was redissolved in a minimal amount of CH_2Cl_2 after the ethyl acetate was removed by rotovap and then subjected to purification over a silica gel column. A wash of 300 ml 5:1 hexanes:CH_2Cl_2 removed what appeared to be polymer formed in the reaction. The *t*-butyl-protected product eluted from the column in a subsequent 400 ml wash of 7:1 hexanes:ethyl acetate as determined by TLC. Product was rotovapped to dryness.

Deprotection of the methyl ester was performed by dissolving 2.87 mmol *t*-butyl-*R*-3-hydroxybutyrate methyl ester with 2 equivalents of LiOH in 10 ml 1:1 H_2O:MeOH and reacting it at room temperature for 1.5 hr. The solution was then acidified to pH 1.5 with ice-cold HCl and quickly extracted three times with 7 ml ethyl acetate. The product was rotovapped to dryness and confirmed by LCMS: *t*-butyl-*R*-3-hydroxybutyrate 159.10 [(M − H)$^-$] calculated, 158.95 observed.

Synthesis of *N*-Acyl-Dipeptide Standards: *N*-Acyl-L-Leu-L-Asp and *N*-Acyl-D-Leu-L-Asp

0.2 mmol H-Asp(OtBu)-2-ClTrt resin was swelled in DMF then reacted with 3 equivalents of either Fmoc-D-Leu-OH or Fmoc-L-Leu-OH, 3 equivalents of HBTU, 1 equivalent of HOBt, and 12 equivalents of DIEA dissolved in 3 ml DMF while shaking at room temperature for 3 hr. Afterwards, the resin was washed three times with 3 ml DMF and three times with 3 ml CH_2Cl_2, followed by deprotection with 2 × 3 ml 20% piperidine for 10–15 min. After washing three times with 3 ml DMF and three times with 3 ml CH_2Cl_2, the second coupling reaction was performed by the addition of 1.5 equivalents of *t*-butyl-*R*-3-hydroxybutyrate, 1.5 equivalents of HBTU, 1 equivalent of HOBt, and 6 equivalents of DIEA dissolved in 1.5 ml DMF while shaking at room temperature for 2.5 hr. After washing three times with 3 ml DMF and three times with 3 ml CH_2Cl_2, the product was deprotected and cleaved from the resin for 3 hr by using 3 ml 90/5/5 TFA/$H_2$0/TIS. The TFA was rotovapped off, and the product was redissolved in H_2O and purified by preparative HPLC with a gradient of 0%–100% acetonitrile over 40 min starting in 0.1% TFA in $H_2$0. Product peaks were identified by LCMS and lyophilized to dryness: *N*-acyl-L-Leu-L-Asp 331.15 [(M − H)$^-$] calculated, 331.10 observed; *N*-acyl-D-Leu-L-Asp 331.15 [(M − H)$^-$] calculated, 331.05 observed.

Synthesis of *N*-Acyl-L-Leu-CoA Substrate

0.417 mmol H-Leu-2-ClTrt resin was swelled in DMF then reacted with 1.5 equivalents of *t*-butyl-*R*-3-hydroxybutyrate, 1.5 equivalents of HBTU, 1 equivalent of HOBt, and 6 equivalents of DIEA dissolved in 3.125 ml DMF while shaking at room temperature for 2.5 hr. After washing three times with 3 ml DMF and three times with 3 ml CH_2Cl_2, the product was cleaved from the resin without deprotection with 3 ml 1:1:3 AcOH:TFE:CH_2Cl_2 while shaking at room temperature for 2 hr. The CH_2Cl_2 was rotovapped off, and the product was redissolved in H_2O and purified by preparative HPLC with a gradient of 0%–100% acetonitrile over 40 min starting in 0.1% TFA in H_2O. The product peak was identified by LCMS and lyophilized to dryness: *t*-butyl-*R*-3-hydroxybutyryl-L-Leu 272.18 [(M − H)$^-$] calculated, 272.15 observed.

73.1 μmol *N*-acyl-L-Leu was reacted with 2 equivalents of CoASH, 4 equivalents of Pybop, and 8 equivalents of DIEA dissolved in 1.5 ml 1:1 THF:H_2O while stirring at room temperature for 3 hr. The reaction was then purified by direct injection onto the preparative HPLC with a 0%–100% acetonitrile gradient over 40 min starting in 0.1% TFA in H_2O. The product peak was identified and lyophilized to dryness: *t*-butyl-*R*-3-hydroxybutyryl-L-Leu-CoA 1021.30 [(M − H)$^-$] calculated, 1021.00 observed. Deprotection was achieved by reacting with 1 ml 95/5/5 TFA/H_2O/TIS for 3 hr at room temperature. The solution was then diluted to 5 ml with H_2O and injected onto the preparative HPLC with a gradient of 0%–100% acetonitrile over 40 min starting in 0.1% TFA in H_2O. The product peak was identified and then lyophilized to dryness: *N*-acyl-L-Leu-CoA 965.25 [(M − H)$^-$] calculated, 964.90 observed.

Analysis of Epimerization Activity on Aminoacyl-*S*-T Substrates

10 μM of each enzyme included in the 120 μl reaction was primed by incubation with 50 mM Tris (ArfA C2 containing reactions) or HEPES (ArfB C3 containing reactions) (pH 7.5), 5 mM $MgCl_2$, 100 μM CoASH, and 3 μM Sfp at room temperature for 1 hr. Amino acid loading, epimerization, and condensation were then assayed by initiation with 37.5 mM Tris or HEPES (pH 7.5), 5 mM ATP, 100–500 μM ^{14}C radiolabeled amino acid (depending on how concentrated the radioactive stock was) and 500 μM cold partner amino acid (if included). After 10 min, reactions were transferred to BioMax centrifugal filters and washed 4 × 500 μl with 50 mM Tris or HEPES (pH 7.5) with centrifugation between each wash. 50 μM of the external thioesterase TycF was then added to the reaction, which was at a total volume of 25 μl. Cleavage of the T domain bound products proceeded at room temperature for 1 hr, and then the products were filtered out with three 100 μl washes with H_2O. After lyophilization, the products were redissolved in 10 μl 50/50 MeOH/H_2O. The reactions containing ArfA C2 to assay Leu_1 were analyzed by using chiral silica TLC with a developing buffer of 200/50/25/25 ACN/AcOH/MeOH/H_2O followed by phosphorimaging. The reactions containing ArfB C1 to assay Asp_2 were analyzed by using chiral HPLC analysis on a Phenomenex 250 × 4.6 mm 3126 Chirex column, with 95% 2 mM copper sulfate in 5% isopropanol, isocratic, monitoring ^{14}C radioactive counts and absorbance at 220 nm.

Analysis of Epimerization Activity on *N*-Acyl-L-Leu-Peptidyl-*S*-T Substrate

10 μM ArfA $C_1A_1T_1C_2A_2T_2$ included in the 300 μl reaction was primed by incubation with 50 mM Tris (pH 7.5), 5 mM $MgCl_2$, 140 μM *N*-acyl-L-Leu-CoA, 60 μM CoASH, and 3 μM Sfp at room temperature for 1 hr. Substrate loading, epimerization, and condensation were then assayed by initiation with 37.5 μM Tris (pH 7.5), 5 mM ATP, and 100 μM L-[^{14}C]Asp. After 10 min, the reaction was quenched with 1.2 ml 10% TCA, pelleted by centrifugation, and washed three times with 100 μl with 10% TCA. Cleavage of the T domain bound products were performed by resuspending in 100 μl 0.1 M LiOH and heating at 60ºC for 15 min. The reaction was then quenched by acidifying with 20 μl 50% TCA and centrifuged to pellet the protein. The supernatant containing the released products was analyzed by HPLC on a Vydac 250 × 4.6 mm C18 small pore column, with a 0%–50% acetonitrile gradient over 25 min starting in 0.1% TFA in H_2O, monitoring ^{14}C radioactive counts and absorbance at 220 nm.

Supplemental Data

Supplemental Data including phylogenetic analysis of A domains from arthrofactin, syringomycin, syringopeptin, and ramoplanin synthetases are available at http://www.chembiol.com/cgi/content/full/12/11/1189/DC1/.

Acknowledgments

We thank Dr. Masaaki Morikawa for providing us with a λ5 clone of the arthrofactin gene cluster. We thank Michael Fischbach and Ellen Yeh for a careful reading of the manuscript and for providing Sfp and TycF, respectively. This work was supported in part by

National Institutes of Health GM 20011 (C.T.W.), the Pharmacological Sciences Training Grant from the National Institute of General Medical Sciences (C.J.B.), a Merck-sponsored Fellowship of the Helen Hay Whitney Foundation (F.H.V.), and a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship (F.H.V.).

Received: July 7, 2005
Revised: August 16, 2005
Accepted: August 16, 2005
Published: November 18, 2005

References

1. Cane, D.E., and Walsh, C.T. (1999). The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. Chem. Biol. *6*, R319–R325.
2. Cane, D.E., Walsh, C.T., and Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. Science *282*, 63–68.
3. Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. *97*, 2651–2674.
4. Schwarzer, D., and Marahiel, M.A. (2001). Multimodular biocatalysts for natural product assembly. Naturwissenschaften *88*, 93–101.
5. Keating, T.A., and Walsh, C.T. (1999). Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. Curr. Opin. Chem. Biol. *3*, 598–606.
6. Weber, T., and Marahiel, M.A. (2001). Exploring the domain structure of modular nonribosomal peptide synthetases. Structure (Camb) *9*, R3–R9.
7. Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the nonribosomal biosynthesis of gramicidin S. EMBO J. *16*, 4174–4183.
8. Stachelhaus, T., Huser, A., and Marahiel, M.A. (1996). Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. Chem. Biol. *3*, 913–921.
9. Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., and Holak, T.A. (2000). Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. Struct. Fold. Des. *8*, 407–418.
10. Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science *284*, 486–489.
11. Stachelhaus, T., Mootz, H.D., Bergendahl, V., and Marahiel, M.A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. J. Biol. Chem. *273*, 22773–22781.
12. Walsh, C.T., Chen, H., Keating, T.A., Hubbard, B.K., Losey, H.C., Luo, L., Marshall, C.G., Miller, D.A., and Patel, H.M. (2001). Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Curr. Opin. Chem. Biol. *5*, 525–534.
13. Konz, D., and Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? Chem. Biol. *6*, R39–R48.
14. Luo, L., Kohli, R.M., Onishi, M., Linne, U., Marahiel, M.A., and Walsh, C.T. (2002). Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. Biochemistry *41*, 9184–9196.
15. Stachelhaus, T., and Walsh, C.T. (2000). Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. Biochemistry *39*, 5775–5787.
16. Trauger, J.W., Kohli, R.M., and Walsh, C.T. (2001). Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. Biochemistry *40*, 7092–7098.
17. Dittmann, J., Wenger, R.M., Kleinkauf, H., and Lawen, A. (1994). Mechanism of cyclosporin A biosynthesis. Evidence for synthesis via a single linear undecapeptide precursor. J. Biol. Chem. *269*, 2841–2846.
18. Stein, T., Kluge, B., Vater, J., Franke, P., Otto, A., and Wittmann-Liebold, B. (1995). Gramicidin S synthetase 1 (phenylalanine racemase), a prototype of amino acid racemases containing the cofactor 4′-phosphopantetheine. Biochemistry *34*, 4633–4642.
19. 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.
20. Patel, H.M., Tao, J., and Walsh, C.T. (2003). Epimerization of an L-cysteinyl to a D-cysteinyl residue during thiazoline ring formation in siderophore chain elongation by pyochelin synthetase from *Pseudomonas aeruginosa*. Biochemistry *42*, 10514–10527.
21. 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.
22. Roongsawang, N., Hase, K., Haruki, M., Imanaka, T., Morikawa, M., and Kanaya, S. (2003). Cloning and characterization of the gene cluster encoding arthrofactin synthetase from *Pseudomonas* sp. MIS38. Chem. Biol. *10*, 869–880.
23. 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.
24. Tosato, V., Albertini, A.M., Zotti, M., Sonda, S., and Bruschi, C.V. (1997). Sequence completion, identification and definition of the fengycin operon in *Bacillus subtilis* 168. Microbiol. *143*, 3443–3450.
25. Hoffmann, K., Schneider-Scherzer, E., Kleinkauf, H., and Zocher, R. (1994). Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. J. Biol. Chem. *269*, 12710–12714.
26. 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.
27. Scholz-Schroeder, B.K., Soule, J.D., and Gross, D.C. (2003). The sypA, sypS, 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.
28. Lambalot, R.H., and Walsh, C.T. (1995). Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. J. Biol. Chem. *270*, 24658–24661.
29. Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., and Walsh, C.T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry *37*, 1585–1595.
30. Yeh, E., Kohli, R.M., Bruner, S.D., and Walsh, C.T. (2004). Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-*S*-enzyme that cannot be elongated. ChemBioChem *5*, 1290–1293.
31. Sieber, S.A., Walsh, C.T., and Marahiel, M.A. (2003). Loading peptidyl-coenzyme A onto peptidyl carrier proteins: a novel approach in characterizing macrocyclization by thioesterase domains. J. Am. Chem. Soc. *125*, 10862–10866.
32. Clugston, S.L., Sieber, S.A., Marahiel, M.A., and Walsh, C.T. (2003). Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C_5 domain of tyrocidine synthetase is a $^{D}C_{L}$ catalyst. Biochemistry *42*, 12095–12104.
33. von Dohren, H., Dieckmann, R., and Pavela-Vrancic, M. (1999). The nonribosomal code. Chem. Biol. *6*, R273–R279.
34. Farnet, C., and Staffa, A. February 2004. U.S. patent US 2004/0033581 A1.
35. Farnet, C.M., Zazopoulos, E., and Staffa, A. November 2002. U.S. patent US 2002/0164747-A1.
36. 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.
37. von Dohren, H., Keller, U., Vater, J., and Zocher, R. (1997). Multifunctional peptide synthetases. Chem. Rev. *97*, 2675–2706.

38. Linne, U., and Marahiel, M.A. (2000). Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. Biochemistry *39*, 10439–10447.
39. Schmoock, G., Pfennig, F., Jewiarz, J., Schlumbohm, W., Laubinger, W., Schauwecker, F., and Keller, U. (2005). Functional cross-talk between fatty acid synthesis and nonribosomal peptide synthesis in quinoxaline antibiotic-producing streptomycetes. J. Biol. Chem. *280*, 4339–4349.
40. Grgurina, I., and Mariotti, F. (1999). Biosynthetic origin of syringomycin and syringopeptin *22*, toxic secondary metabolites of the phytopathogenic bacterium *Pseudomonas syringae* pv. *syringae*. FEBS Lett. *462*, 151–154.
41. Keating, T.A., Marshall, C.G., Walsh, C.T., and Keating, A.E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. Nat. Struct. Biol. *9*, 522–526.
42. Marshall, C.G., Hillson, N.J., and Walsh, C.T. (2002). Catalytic mapping of the vibriobactin biosynthetic enzyme VibF. Biochemistry *41*, 244–250.
43. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd Edition (Plainview, NY: Cold Spring Harbor Press).
44. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene *77*, 51–59.
45. Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.