A Journal of the Gesellschaft Deutscher Chemiker

Angewandte International Edition GDCh Chemie

www.angewandte.org

Accepted Article

Title: Unprecedented Noncanonical Features of the Nonlinear Nonribosomal Peptide Synthetase Assembly Line for WS9326A Biosynthesis

Authors: Myoun-Su Kim, Munhyung Bae, Ye-Eun Jung, Jung Min Kim, Sunghoon Hwang, Myoung Chong Song, Yeon Hee Ban, Eun Seo Bae, Suckchang Hong, Sang Kook Lee, Sun-Shin Cha, Dong-Chan Oh, and Yeo Joon Yoon

To be cited as: *Angew. Chem. Int. Ed.* 10.1002/anie.202103872

Link to VoR: https://doi.org/10.1002/anie.202103872

WILEY-VCH

WILEY-VCH

RESEARCH ARTICLE

Unprecedented Noncanonical Features of the Nonlinear Nonribosomal Peptide Synthetase Assembly Line for WS9326A Biosynthesis

Myoun-Su Kim,[+a] *Munhyung Bae*,[+a] *Ye-Eun Jung*,[+b] *Jung Min Kim*,[a] *Sunghoon Hwang*,[a] *Myoung Chong Song*,[a] *Yeon Hee Ban*,[a] *Eun Seo Bae*,[a] *Suckchang Hong*,[c] *Sang Kook Lee*,[a] *Sun-Shin Cha*,*[b] *Dong-Chan Oh*,*[a] *and Yeo Joon Yoon**[a]

[a] Dr. M.-S. Kim, Dr. M. Bae, J. M. Kim, Dr. S. Hwang, Dr. M. C. Song, Dr. Y. H. Ban, E. S. Bae, Prof. S. K. Lee, Prof. D.-C. Oh, Prof. Y. J. Yoon
Natural Products Research Institute, College of Pharmacy
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826 (Republic of Korea)
E-mail: dongchanoh@snu.ac.kr, yeojoonyoon@snu.ac.kr

[b] Y.-E. Jung, Prof. S.-S. Cha
Department of Chemistry and Nanoscience
Ewha Womans University
52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760 (Republic of Korea)
E-mail: chajung@ewha.ac.kr

[c] Prof. S. Hong
Research Institute of Pharmaceutical Sciences, College of Pharmacy
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826 (Republic of Korea)

[+] These authors contributed equally to this work.

Supporting information for this article is given via a link at the end of the document.

Abstract: Systematic inactivation of nonribosomal peptide synthetase (NRPS) domains and translocation of the thioesterase (TE) domain revealed several unprecedented nonlinear NRPS assembly processes during the biosynthesis of the cyclodepsipeptide WS9326A in *Streptomyces* sp. SNM55. First, two sets of type II TE (TEII)-like enzymes mediate the shuttling of activated amino acids between two sets of stand-alone adenylation (A)-thiolation (T) didomain modules and an 'A-less' condensation (C)-T module with distinctive specificities and flexibilities. This was confirmed by the elucidation of the affinities of the A-T didomains for the TEIIs and its structure. Second, the C-T didomain module operates iteratively and independently from other modules in the same protein to catalyze two chain elongation cycles. Third, this biosynthetic pathway includes the first example of module skipping, where the interpolated C and T domains are required for chain transfer. Understanding such untapped nonlinear biosynthetic strategies will provide potentially valuable tools for engineered biosynthesis of these complex natural products.

Introduction

Nonribosomal peptides (NRPs) are biosynthesized by large multifunctional enzyme complexes referred to as nonribosomal peptide synthetases (NRPSs). A typical minimal NRPS module is composed of an adenylation (A) domain, which recognizes and activates the cognate amino acid building block as an aminoacyl adenylate, a thiolation (T) domain, which tethers the amino acid through a 4′-phosphopantetheine (Ppant) linker, and a condensation (C) domain, which catalyzes the peptide bond formation in the usual C-A-T domain organization. In most cases, the thioesterase (TE) domain of the last module releases the product via hydrolysis or cyclization.[1,2] Numerous NRPS assembly lines follow the "co-linearity rule, " in which the number and order of NRPS modules correspond to the number and sequence of amino acids in the NRP. Currently, a growing number of NRPS assembly lines violating this rule and deviating from the canonical modular architecture have been discovered.[3,4,5] However, a detailed understanding of the diverse biosynthetic strategies of these "nonlinear" NRPSs is needed to further exploit their potential for combinatorial biosynthesis of modified NRPs with improved pharmaceutical properties.[6]

The cyclodepsipeptide WS9326A (**1**) (Figure 1) was first isolated from *Streptomyces violaceusniger* no. 9326.[7] Several *Streptomyces* strains (*S.* sp. 9078,[8] *S.* sp. SNM55,[9,10] *S. calvus* ATCC 13382,[11] and *S. asterosporus* DSM 41452[12]) have also been identified to produce **1** and its congeners. WS9326A has remarkable structural features, including the (*E*)-2,3-didehydrotyrosine (ΔTyr) residue, which has not been found in other NRPs. A recent gene deletion study in *S. asterosporus* DSM 41452 indicated that the *sas16* gene, encoding a cytochrome P450 monooxygenase, is involved in the formation of the double bond of ΔTyr in WS9326A.[12] Another interesting structural feature of WS9326A is the *N*-terminal attached (*Z*)-alkenylcinnamoyl moiety, which has also been observed in pepticinnamins,[13] skyllamycin,[14] and the WS9326A/B dimer-like mohangamide A/B.[10] Recently, similar cinnamoyl lipids in youssoufenes were shown to be biosynthesized by unusual type II polyketide synthase.[15] In addition, the NRPS assembly line of **1** in the WS9326A cluster of *S. calvus* ATCC 13382[11] and *S. asterosporus* DSM 41452[12] has several unique features from a biosynthetic viewpoint: the existence of two stand-alone modules consisting of an A-T didomain as well as two type II TEs (TEIIs), and one module lacking an A domain. However, this has not been investigated in detail. Here, we detail several hitherto undescribed features of the "nonlinear" NRPS assembly line in *S.* sp. SNM55 by systematic domain inactivation and the translocation of the TE domain to internal modules in combination with elucidation of the

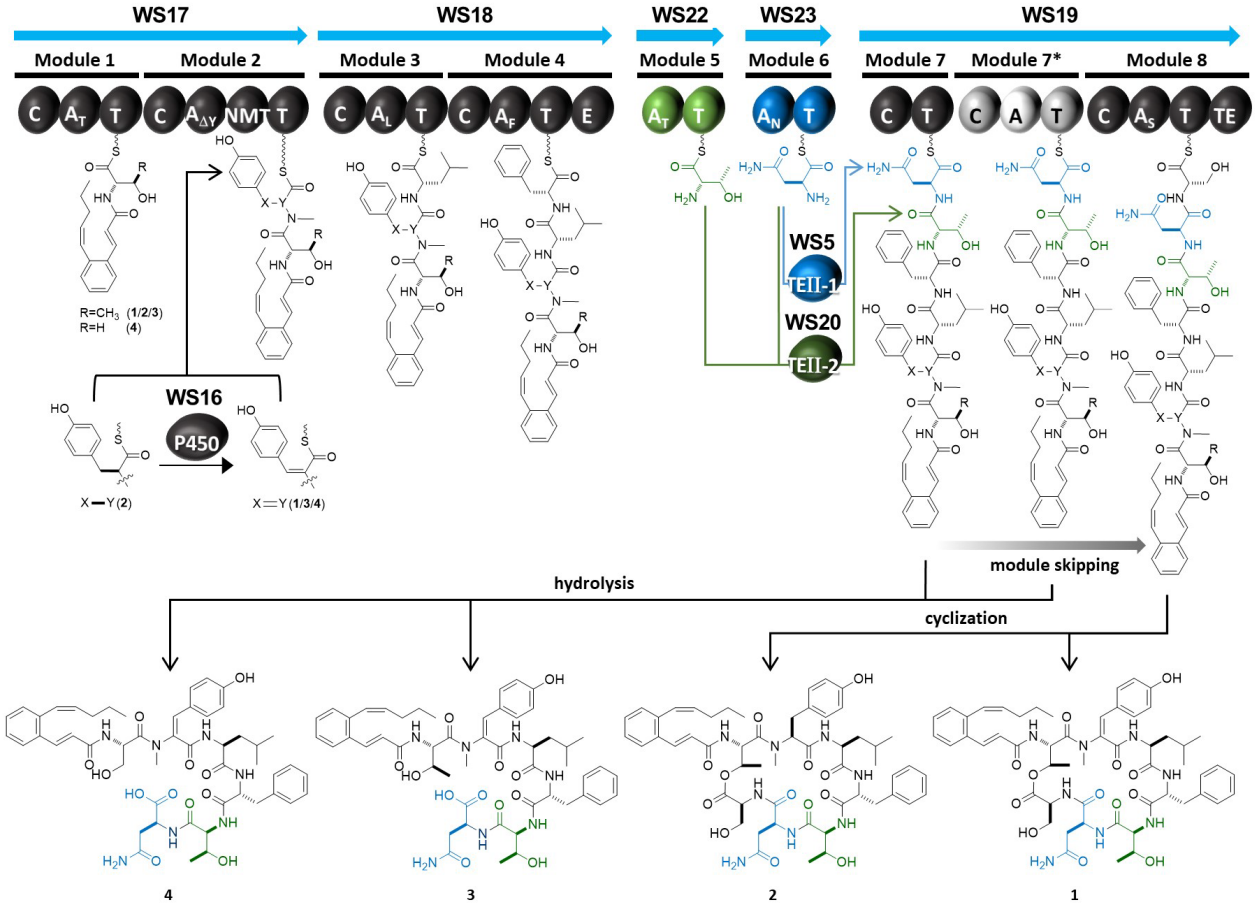


Figure 1. Structures of WS9326A and its congeners, and the schematic representation of the WS9326A NRPS assembly line with its biosynthesis in *S.* sp. SNM55. Module 7* indicated in gray is absent in *S. calvus* ATCC 13382 and *S. asterosporus* DSM 41452. Two A-T didomain modules as well as the corresponding activated amino acids are depicted in green and blue, respectively. The shuttling enzymes are depicted in dark green and blue. Domains within each module are represented by circles. A white circle indicates a domain that is not predicted to be active due to deletions of key amino acids in the conserved regions.

structure of TEII-like shuttling enzyme and the interactions between TEIIs and A-T didomains.

Results and Discussion

Genome sequencing of *S.* sp. SNM55[9] identified a NRPS gene cluster, which was highly similar to the WS9326A gene clusters in *S. calvus*[11] and *S. asterosporus*.[12] Sequence analysis of this NRPS cluster revealed that 38 of the 40 genes in the WS9326A cluster of *S. calvus* and *S. asterosporus* are well maintained and identically organized in *S.* sp. SNM55 (Figure S1 and Table S1 in the Supporting Information). Importantly, five NRPS genes (*ws17–ws19* and *ws22–ws23*) in *S.* sp. SNM55 are almost identical to those in *S. calvus* and *S. asterosporus,* except for the insertion of one extra C-A-T module between the last two modules (Figure 1). Based on a detailed analysis of the specificity-conferring sequences in A domains[16-18] (Table S2) and the structure of NRPs produced, we suggest that modules 1–4 of WS17 and WS18 are responsible for the incorporation of L-Thr or L-Ser (producing **1**/WS9326B (**2**)/WS9326F (**3**), which contains an L-Thr residue as the first amino acid and WS9326G (**4**), which contain an L-Ser residue, respectively, as shown in Figure 1), *N*-methyl-(*E*)–2,3–didehydro-Tyr (MeΔTyr), L-Leu, and D-Phe, respectively. The A domains of the stand-alone incomplete A-T didomain modules WS22 (module 5) and WS23 (module 6) seem to be specific for L-*allo*-Thr and L-Asn, respectively, which has been experimentally proven by in vitro ATP-pyrophosphate exchange assay.[19] In-frame deletion of the *ws17* gene encoding module 1 led to the loss of production of **1,** confirming the identity of this gene cluster (Figure S2). We propose that the activated L-*allo*-Thr and L-Asn are shuttled from these stand-alone modules to the T domain of the 'A-less' C-T didomain module (T7 domain) encoded by *ws19* through the action of putative type II TE(s) encoded by either *ws5* or *ws20*. There are two types of T-to-T shuttling enzymes known to date. Aminoacyltransferases containing a GxCxG motif at the active site, such as SyrC for syringomycin,[20] CmaE for coronamic acid,[21] and ZbmVIId for zorbamycin biosynthesis,[22] transfer the amino acid moiety between Ppant arms of the T domains utilizing a cysteine active site. Alternatively, type II TEs, such as BarC for barbamide[23] and CouN7 for coumermycin A_1 biosynthesis,[24] with a GxSxG motif, are predicted to either directly transfer the amino acid onto the T domain or to release the amino acid via a TE-like mechanism. Since WS5 and WS20 contain a Ser residue in the active site (GxSxG) (Figure S3), they might mediate the transfer of L-*allo*-Thr and L-Asn from WS22 and WS23 to the T7 domain of WS19 via a TE-like mechanism.[23,24] Next, we envision that module 7, consisting of a C-T didomain, might catalyze two cycles of elongation reactions with L-*allo*-Thr and L-Asn provided by WS22 and WS23. Module 8 was predicted to be responsible for the incorporation of L-Ser and the cyclized release of **1**/**2** (Figure 1).

The extra module found in *S.* sp. SNM55 (module 7*) seems to be unessential for **1**/**2** biosynthesis because this extra module is absent in both *S. calvus* and *S. asterosporus* strains and the A domain of this module 7* in *S.* sp. SNM55 lacks several key amino acids in the conserved regions, although its C and T domains appear to be active (Figure S4). Taken together, this unusual architecture of WS9326A NRPSs raises several questions: 1) is this extra module found only in *S.* sp. SNM55 superfluous or essential for WS9326A biosynthesis?; 2) what is the mechanism for module skipping, which is rare in NRPS-mediated biosynthesis?; 3) does module 7 operate iteratively?; and 4) how do the two sets of *trans*-acting A-T didomain modules cooperate with the type II TEs to load the T7 domain with L-*allo*-Thr and L-Asn?

We first examined the involvement of module 7* in the biosynthesis of **1**/**2**. Inactivation of the C domain of module 7* (C7*) through the mutation of the second His residue of the conserved active motif HHxxxDG[25] (H697V) led to complete loss of production of **1** and **2** in the resulting mutant strain (ΔC7* strain) (Figure 2a). On the other hand, a mutant strain in which the A domain of module 7* was inactivated (ΔA7* strain) through the replacement of the universally conserved Asp residue with Phe (D1411F) in the ATP-binding pocket, (S/T)GD[26], did not affect the production of **1** and **2** (Figure 2b). Mutation of the highly conserved Ser residue, which represents the binding site for the cofactor Ppant, in the core T motif GxxS (S1569F)[27] also abolished the production of both **1** and **2** in the ΔT7* strain (Figure 2c). We investigated whether T7* can be post-translationally activated by 4′-phosphopantetheinylation (Ppantylation), since the T7* domain should be activated to participate in chain transfer. Following qTOF-HR-MS analysis, we discovered that incubation of the recombinant T7* domain expressed from *Escherichia coli* (Figure S5) with the broad specificity phosphopantetheinyl transferase (PPTase) Sfp[28] produced a mass (21143 Da) consistent with *holo*-T7*. A small peak corresponding to *holo*-T7* was detected without incubation with Sfp. A previous study showed that heterologous T domains are often partially Ppantylated by endogenous PPTase(s) of *E. coli*.[29] As a negative control, the mutated T7* domain, in which the active site Ser residue was substituted with Phe in the T7* domain, was expressed in the same *E. coli* strain and purified. No mass shift corresponding to Ppantylation (+340 Da) was observed in the mutated T7* (Figure S5) upon incubation with Sfp; thus, showing that the T7* domain can be present as an active *holo*-form in *S.* sp. SNM55. These results indicate that the extra module 7* is indispensable for **1**/**2** assembly in *S.* sp. SNM55, although this module does not carry out peptide elongation. Notably this module skipping process requires C and T domains, which has not been previously described. Module skipping has been reported for an engineered polyketide synthase (PKS) system, in which the growing polyketide chain passes through the skipped module by direct transfer between the T domains, thus, requiring the interpolated T domain for the skipping.[30] Similarly, module skipping was reported in an engineered plipastatin NRPS, where the TE domain was moved to the end of the internal T domains, however, the involvement of the T domain in module skipping has not been studied.[31] There has been only one case where module skipping has been observed during the natural NRPS assembly process. In the biosynthesis of the pentapeptide myxochromide S, the growing peptide chain is directly transferred from module 3 to module 5 (skipping module 4 completely), where the T domain of

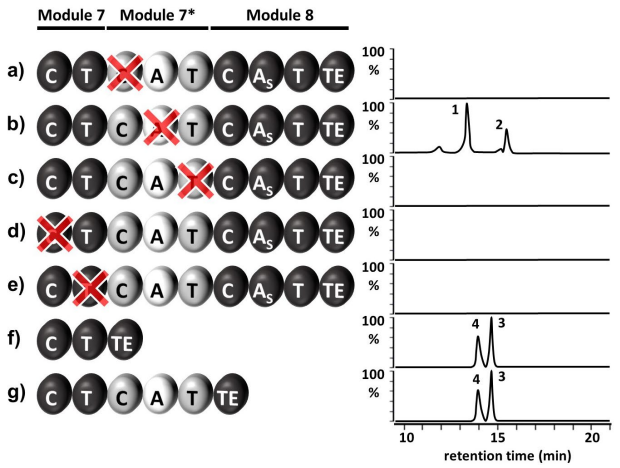


Figure 2. Analysis of strains containing mutations in module 7 and 7*, as well as TE reposition. HPLC-MS chromatogram obtained from the culture of a) ΔC7*, b) ΔA7*, c) ΔT7*, d) ΔC7, e) ΔT7, f) Mod7:TE, and g) Mod7*:TE strains, selected for **1** (for *m/z* = 1037), **2** (*m/z* = 1039), **3** (*m/z* = 968), and **4** (*m/z* = 954). A cross indicates the mutagenesis of a single amino acid essential for the domain activity.

the skipped module was not post-translationally activated by PPTase in contrast to T7*.[32] It is plausible that the active site of the C domain recognizes the Ppant arms of both the preceding and following T domains and brings these prosthetic groups into proximity to allow efficient chain transfer. This hypothesis is supported by the currently available C domain structures.[1,33-35] Additionally, repositioning module 8 directly after module 7 by deleting module 7* (ΔMod7* strain) terminated the production of **1** and **2** (Figure S6). This advocates that the non-elongating module 7* is essential for production of **1**/**2** in *S.* sp. SNM55; although, the possible negative effects of module 7* deletion on the functional integrity of WS19 and **1**/**2** production cannot be ruled out.

We hypothesized that module 7 of WS19 in *the S.* sp. SNM55 and the corresponding module in *S. calvus*/*S. asterosporus* catalyze two elongation steps to assemble the hepta-peptide core structures of **1**/**2**. To examine this possibility, we inactivated each domain of module 7 by site-directed mutagenesis in *S.* sp. SNM55. Substitution of the second His residue of the conserved C domain motif with Val (H146V) and the Ser residue from the core T motif with Ala (S512A) in module 7 of *S.* sp. SNM55 resulted in the loss of **1**/**2** production in the ΔC7 and ΔT7 strains, respectively (Figures 2d and 2e). This indicates that module 7, consisting of a C-T didomain, is indispensable for **1**/**2** assembly and is probably used iteratively. The validity of this hypothesis was further strengthened by the translocation of the TE domain to the end of the T7 domain. A mutant strain in which the TE domain containing the inter-domain linker region between the T and TE domains of module 8 was fused to the end of the T7 domain (Mod7:TE strain) was generated (Figure S7). This strain produced small amounts of **3** (acyl-Thr-MeΔTyr-Leu-Phe-Thr-Asn) and **4** (acyl-Ser-MeΔTyr-Leu-Phe-Thr-Asn), which are the linear truncated hexa-peptide analogues of **1** released from module 7 (Figures 1 and 2f; see Figures S8–S19 for their spectroscopic data). This confirms that module 7 must be used iteratively and adds two amino acids (L-*allo*-Thr and L-Asn) to the growing peptide chain. Furthermore, **3** and **4** are probably generated due to the substrate promiscuity of the A domain of module 1 and have been isolated from the large-scale culture of the wild-type *S. asterosporus* strain,[12] while not being detected

in the wild-type *S.* sp. SNM55 culture under the conditions used in this study. In addition, repositioning the TE domain at the end of the T domain of module 7* in the *S.* sp. SNM55 (Mod7*:TE strain) also produced **3** and **4** (Figure 2g and Figure S20), supporting the iterative use of module 7 as well as the skipping of module 7*. The production of **3**/**4** was also confirmed in the mutant strain of *S. calvus,* where a TE domain was fused to the end of module 7 (ScMod7:TE) (Figure S21). Taken together, these results proved that the 'A-less' module 7 of the WS9326A NRPS in both *S.* sp. SNM55 and *S. calvus* (or *S. asterosporus*) operate iteratively, catalyzing two cycles of chain elongation using two different amino acid building blocks. Several examples of the repeated use of modules in PKS-catalyzed assembly lines have been previously characterized.[4,36,37] In contrast, only the iterative use of the entire assembly line or individual domains is reported in the NRPS system.[3,4] Classic examples of complete NRPS assembly lines that operate iteratively to generate repeated peptide sequences include enterobactin[38] and gramicidin.[39] Further examples of the repeated use of an individual NRPS domain can be found in the assembly of yersiniabactin, in which the A domain of module 2 loads three different T domains with cysteine.[40] The iterative use of module 7 represents the first proven example in which an individual module within a multimodular NRPS enzyme operates iteratively and separately from other modules in the same protein.

Based on our results, it is expected that the activated L-*allo*-Thr and L-Asn would be loaded onto the T7 domain through the combined action of two sets of A-T didomain modules and TEII-like shuttling enzymes (Figure 1). To examine the function of each enzyme in this shuttling process, single amino acid mutagenesis was first introduced to each T domain of A-T didomain modules and TEII in *S.* sp. SNM55. Mutation of the Ppant-binding Ser residue in the T domain of module 5 (WS22) (S1046A; ΔT5 strain) produced a new analogue of **1**, which contained L-Asn, instead of L-*allo*-Thr as the fifth amino acid [acyl-Thr-MeΔTyr-Leu-Phe-Asn-Asn-Ser, designated as WS9326I (**5**)] (Figures 3a and 3b; see Figures S22–S28 and Table S3 for its structural elucidation). Similarly, the mutant strain where the T domain of module 6 (WS23) was inactivated (S541A; ΔT6 strain) produced a **1** derivative where the sixth L-Asn is replaced with L-*allo*-Thr [acyl-Thr-MeΔTyr-Leu-Phe-Thr-Thr-Ser, designated as WS9326J (**6**)] (Figures 3a and 3c; Figures S29–S35 and Table S4). These results confirmed that the A-T didomain modules 5 and 6 are responsible for the activation of L-*allo*-Thr and L-Asn, respectively, and that the T domains of these A-T didomain modules are required for the activation reaction. Although the production levels of new compounds **5** and **6** decreased significantly compared to the production of **1** from wild-type *S.* sp. SNM55, the production of these new analogues showed that the downstream modules possess a degree of promiscuity to recognize and process the altered substrate.

To investigate the specificity of the two shuttling enzymes, two mutant strains were constructed by replacing the nucleophilic Ser residue of the conserved catalytic triads (GxSxG) with Ala in WS5 (S97A; ΔTEII-1 strain) and WS20 (S79A; ΔTEII-2 strain). While the ΔTEII-1 strain produced **1**/**2**, the ΔTEII-2 strain produced only **5** (Figures 3d and 3e), indicating that WS5 can mediate only the shuttling of L-Asn from module 6 to the T7 domain repeatedly. Alternatively, WS20 is able to release L-*allo*-Thr and L-Asn from modules 5 and 6, respectively, and load the T7 domain with these amino acids. This model was further verified using multiple double mutants and chemical analyses. Both ΔT5ΔTEII-1 and ΔT5ΔTEII-2 strains produced **5** (Figures 3f and 3g), showing that both WS5 and WS20 can shuttle L-Asn twice in the absence of activated L-*allo*-Thr. The ΔT6ΔTEII-1 strain produced **6** (Figure 3h), indicating that L-*allo*-Thr activated by module 5 can be shuttled twice by WS20. The ΔT6ΔTEII-2 strain did not produce any WS9326 compounds (Figure 3i), further supporting that WS5 is unable to shuttle L-*allo*-Thr. Additionally, no WS9326 compound was detected in ΔT5ΔT6 and ΔTEII-1ΔTEII-2 strains (Figure S36). These results confirmed that the two *trans*-shuttling enzymes WS5 and WS20 have different specificities: WS5 is able to shuttle only L-Asn, while WS20 can shuttle both L-*allo*-Thr and L-Asn. Notably, these shuttling enzymes worked iteratively in the absence of other shuttling enzymes. In particular, the discovery of WS20, which showed substrate promiscuity associated with its acyl and T substrates, will provide a valuable tool for the engineered biosynthesis of novel NRP scaffolds with altered or improved biological activities.

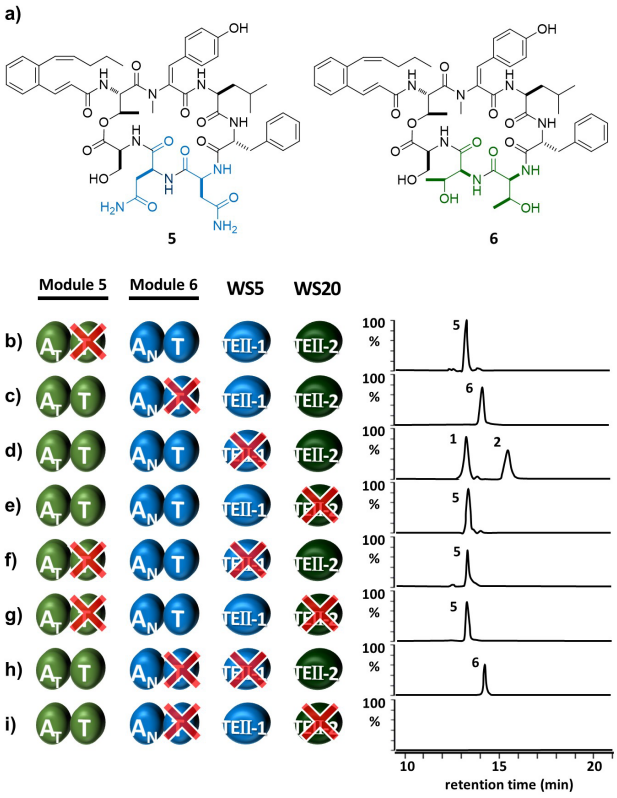


Figure 3. Analysis of strains containing mutations in modules 5 and 6, as well as shuttling thioesterases. a) Chemical structures of **5** and **6**. HPLC-MS chromatogram obtained from the culture of b) ΔT5, c) ΔT6, d) ΔTEII-1, e) ΔTEII-2, f) ΔT5ΔTEII-1, g) ΔT5ΔTEII-2, h) ΔT6ΔTEII-1, and i) ΔT6ΔTEII-2 strains, selected for **1**, **2**, **5** (*m/z* = 1050), and **6** (*m/z* = 1024).

This unique cross-talk between the A-T didomains and TEIIs was further verified by estimating their interactions using microscale thermophoresis (MST) and elucidation of the crystal structure of WS5. Since the MbtH-like protein Cal4 in *S. calvus* (corresponding to WS4 in *S.* sp. SNM55) is known to increase the soluble expression of A domains in Cal22 and Cal23[19] (corresponding to WS22 and WS23), we coexpressed WS22 and WS23 in *E. coli* with WS4. WS5 and WS20 were also expressed in *E. coli*. These recombinant proteins were purified by affinity and

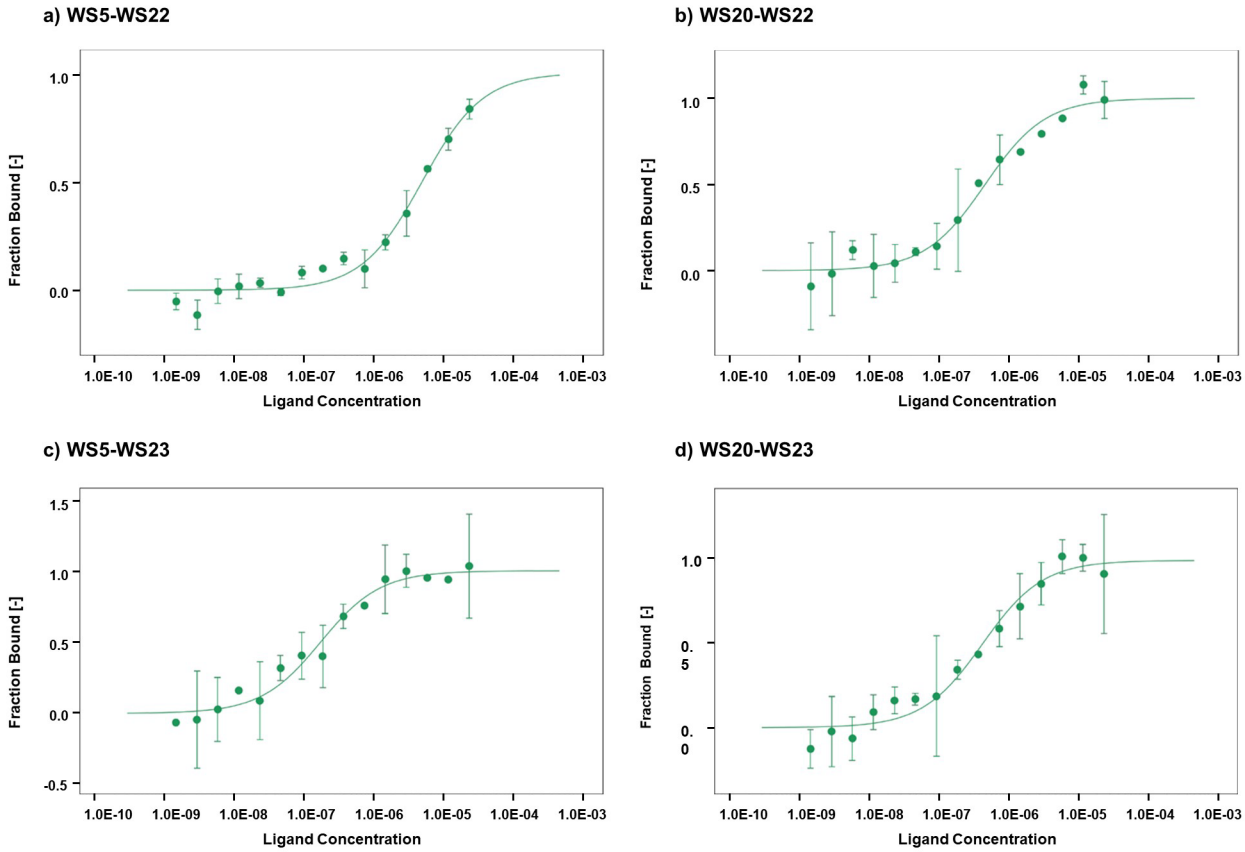


Figure 4. MST analysis for affinity determination between proteins a) WS5–WS22, b) WS20–WS22, c) WS5–WS23, and d) WS20–WS23. Binding curves for measurement of specific interactions were derived from the change in the thermophoretic signal upon WS5/WS20 titration (25 - 0.00153 μM) to a constant concentration (10 nM) of fluorescently labelled WS22/WS23. The curves were fitted to K_d model and yielded K_d = 5214.7 ± 1324.2 nM for WS5–WS22, K_d = 477.9 ± 101.5 nM for WS20–WS22, K_d = 269.6 ± 87.7 nM for WS5–WS23, and K_d = 442.7 ± 106.2 nM for WS20–WS23, respectively. The error bars mean standard deviation calculated from three independently pipetted MST measurements.

fast protein liquid chromatography to homogeneity (Figure S37). The MST assay revealed that the binding affinities between WS5–WS23, WS20–WS22, and WS20–WS23 were reasonably high (K_d = 269.6 ± 87.7 nM, 477.9 ± 101.5 nM, and 442.7 ± 106.2 nM, respectively) while the binding affinity of WS5–WS22 was relatively low (K_d = 5214.7 ± 1324.2 nM) (Figure 4 and Figure S38). These large differences in K_d values are consistent with the observations obtained from in vivo gene inactivation experiments in that WS5 was unable to mediate shutting of L-*allo*-Thr from WS22 to the T7 domain of WS19 in contrast to WS20.

To obtain a structural basis for the partner preference of WS5, we determined the crystal structure of WS5. Our group has discovered that even in the absence of an intrinsic zinc-binding site, proteins can be combined with zinc ions in zinc-containing solutions, and demonstrated successful experimental phasing through zinc derivatization.[41] Likewise, the structure of WS5 was determined by zinc single-wavelength anomalous dispersion phasing after soaking WS5 crystals in a zinc-containing solution (see Supporting Information). The native WS5 structure was then solved by molecular replacement using the zinc-bound WS5 structure as a search model. The structure of WS5 exhibits an α/β hydrolase fold in which the central sheet consisting of six parallel β-strands is sandwiched by five α-helices and this core domain is covered with a lid domain consisting exclusively of α-helices (Figure 5a). Furthermore, an active site pocket harboring the catalytic triad is situated at the domain interface. Nucleophilic Ser97 takes a position at the turn between strand β4 and helix α3, the so-called strand-turn-helix nucleophilic elbow (Figure 5a). It has been previously shown that the nucleophile of catalytic triads in α/β hydrolases resides at the turn of the nucleophilic elbow.[42] Asp208 and His236, which form the catalytic triad of WS5 with Ser97, are positioned within hydrogen bonding distances of each other (Figure 5a). To elucidate the affinity disparities revealed by MST, we modeled the complex structures between WS5 and WS22/23. For this, the T domain structures of WS22 and WS23 were generated by the SWISS-MODEL web server with PDB codes 5T3D and 5U89 as templates, respectively. Subsequently, in silico model structures of the WS5–WS22 and WS5–WS23 complexes were constructed on the basis of the crystal structure of the T-TE complex (PDB code: 3TEJ; Figure S39).[43] The first insight provided by complex modeling is associated with the WS5 structure which represents the closed state. Furthermore, there is no open portal in the binding interface (Figure 5a) for the entry of Ppant covalently linked to Ser1046/Ser541 of WS22/23 T domains, which indicates that the portal will be open on T domain binding. The second insight refers to the relatively low binding affinity between WS5 and WS22, which might be explained through electrostatic repulsion. As shown in Figure 5b, the binding interface of WS5 has a positively charged sector (p-sector). Notably, the surface areas of WS22 and WS23 that contact the p-sector on complex formation have different electrostatic potentials despite the high similarity in the charge distribution of their binding interfaces (Figure 5b). The p-sector-contacting area of WS22 is positively

charged, while the corresponding area of WS23 assumes both positive and negative charges (Figure 5b). Consequently, the p-sector of WS5 repels WS22 binding, while forming favorable interactions with WS23. To the best of our knowledge, the involvement of two shuttling enzymes with distinct specificities has never been observed in an NRPS assembly line, although a combination of stand-alone A-T didomains and shuttling enzymes (aminoacyltransferase or TEII) has been found in other nonlinear NRPS assembly lines.[20,23,24]

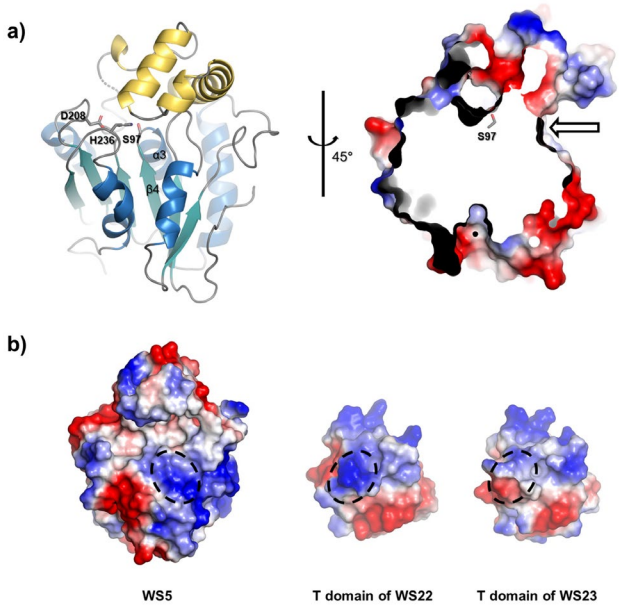


Figure 5. Structure of WS5 and binding interfaces of WS5 and T domains of WS22 and WS23. a) *left*, A cartoon diagram of WS5 shown with labeled catalytic triad, strand β4, and helix α3. Helices, strands, and loops in the core domain are colored in sky-blue, deep teal, and grey, respectively. The lid domain is in yellow. *right*, A clipped view of the surface representation of WS5. The nucleophilic Ser97 is shown in stick and an empty arrow points out the closed portal. Nitrogen and oxygen atoms in sticks are colored in blue and red, respectively. b) Surface representations with electrostatic potentials of WS5 and T domains. A dotted circle in WS5 indicates p-sector and dotted circles in T domains do p-sector-contacting areas on complex formation. In all surface representations, positive and negative charged regions are colored in blue and red, respectively.

Although the MST and structural studies explain the specificity of TEIIs (WS5 and WS20) on interaction with the two A-T didomain modules (WS22 and WS23), it remains to be investigated how the correct order in the incorporation of L-*allo*-Thr and L-Asn, activated by WS22 and WS23 respectively, at the 'A-less' module 7 is achieved. All natural WS9326 congeners contain L-*allo*-Thr and L-Asn as the fifth and sixth amino acids, respectively (Figure 1). Only in the mutants, where either L-*allo*-Thr or L-Asn was not activated, the same amino acid was incorporated twice (Figure 3). Sequence analysis identified the presence of typical communication-mediating (COM) domains, which mediate specific protein-protein interactions between the consecutive NRPS subunits,[44] at the C-terminus of the epimerization (E) domain of WS18 and N-terminus of the C domain of WS19 (Figure S40). However, there are no COM domain-like sequences at both the termini of WS22 and WS23. Moreover, comparison of both the N- and C-terminal sequences of several A-T didomains did not reveal any significantly conserved motifs. It is plausible that WS18 and WS19 form the main NRPS assembly line through the COM domain-mediated interactions; and two sets of the A-T didomain module and TEII cooperate in trans to supply activated L-*allo*-Thr and L-Asn to the T7 domain of WS19 in sequential order through a currently unknown mechanism. Even in the canonical linear NRPSs, COM domain, which is always located at the C-terminus of E (or oxidation) domain and N-terminus of C domain, is just one of the protein-protein interaction systems employed by NRPS that we understand so far. The protein-protein interaction systems in nonlinear NRPSs with all variant domains/modules, most of which have not been fully understood on a mechanistic level, need to be further investigated.

NRPSs utilizing two stand-alone TEIIs like the WS9326 biosynthetic pathway might be more prevalent in nature than previously thought. Bioinformatic analysis of the polyoxypeptin A biosynthetic gene cluster suggested that polyoxypeptin A could also be biosynthesized through a non-linear NRPS system that might use two discrete stand-alone TEIIs for shuttling the activated amino acids from two distinct free-standing T domains to the main NRPS in a similar manner to that of the WS9326 system. In the polyoxypeptin A biosynthetic gene cluster, there are five essential stand-alone NRPS domains: an A domain (PlyC), two T domains (PlyD and PlyQ), and two TEII domains (PlyI and PlyS). It was proposed that the single A domain PlyC may activate multiple amino acids and tether them to the corresponding T domain (PlyD or PlyQ). After some modification, such as *N*/β-hydroxylation, the modified amino acid building blocks may be released by discrete TEIIs (PlyI or PlyS) and loaded to the main NRPS assembly line, although further experiments are required to prove this hypothesis.[45] Interestingly, we also found several putative non-linear NRPS systems from our own actinobacterial genome collection, which contain more than two stand-alone TEIIs together with two (or more) A-T didomain modules or free-standing A and T domains (Figure S41). This indicates that the NRPSs adopting in trans shuttling processes involving discrete TEIIs, as found in the WS9326 assembly line, have not yet been fully explored and may be widely distributed in the actinobacterial genome. Further investigation of these untapped shuttling systems will provide insights on how the complicated interplay between the activation, shuttling, and elongation enzymes, such as WS22/WS23, WS5/WS20 and module 7 in the WS19 presented herein works.

It has been shown that WS9326 class peptides inhibit tumor angiogenesis.[10] The antiangiogenic effects of cyclic **1**/**2**/**5**/**6** and acyclic **3**/**4** were measured based on capillary tube network formation in vascular endothelial growth factor-induced human umbilical vein endothelial cells (Figures S42 and S43). All cyclic WS9326 peptides displayed significant inhibition of endothelial tube formation at a concentration of 20 μM, except for **6,** without noticeable cytotoxicity, indicating that the sixth amino acid, Asn, in the cyclic WS9326s might play an essential role in its activity.

Conclusion

During the past decade, numerous reports on nonlinear NRPS assembly lines have emerged, which makes it difficult to predict both the biosynthetic products and logic of the NRPS assembly lines based on the primary sequence. Understanding the unrevealed biosynthetic logic of NRPSs using structural, biochemical, and genetic analyses is a prerequisite to fully exploit their potential for the engineered biosynthesis of diverse NRPs.

The WS9326 NRPS assembly line represents one of the most complicated interplays within NRPS assembly lines, that involves two TEIIs mediating the transfer of activated amino acids from two stand-alone A-T didomain modules with different specificities to the C-T didomain, which operates iteratively by catalyzing two chain elongation steps. The unprecedented features of the nonlinear WS9326 NRPS add potential new tools to explore nature's ingenuity for producing this important class of natural products.

Acknowledgements

This work was supported by the Collaborative Genome Program of the Korea Institute of Marine Science and Technology Promotion (KIMST) (No. 20180430) and the project titled "Development of potential antibiotic compounds using polar organism resources (15250103, KOPRI Grant PM21030)" funded by the Ministry of Oceans and Fisheries (MOF), the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (MSIT) (2019R1A2B5B03069338 and 2021R1A4A2001251), the Bio & Medical Technology Development Program of the NRF funded by the MSIT (2018M3A9F3079662), Republic of Korea. We thank the staff of beamlines at the Pohang Light Source (Republic of Korea) for the help with data collection.

Keywords: biosynthesis • nonribosomal peptide synthetase • module iteration • module skipping • shuttling thioesterase

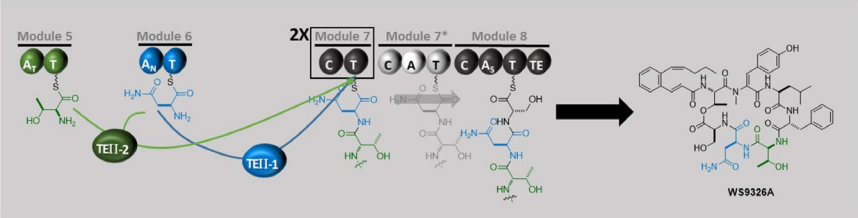
[1] R. D. Süssmuth, A. Mainz, *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 3770–3821; *Angew. Chem.* **2017**, *129*, 3824-3878.
[2] M. A. Fischbach, C. T. Walsh, *Chem. Rev*. **2006,** *106*, 3468–3496.
[3] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *ChemBioChem* **2002**, *3*, 490–504.
[4] S. C. Wenzel, R. Müller, *Curr. Opin. Chem. Biol*. **2005**, *9*, 447–458.
[5] M. J. Jaremko, T. D. Davis, J. C. Corpuz, M. D. Burkart, *Nat. Prod. Rep*. **2020**, *37*, 355–379.
[6] E. Kim, B. S. Moore, Y. J. Yoon, *Nat. Chem. Biol*. **2015**, *11*, 649–659.
[7] K. Hayashi, M. Hashimoto, N. Shigematsu, M. Nishikawa, M. Ezaki, M. Yamashita, S. Kiyoto, M. Okuhara, M. Kohsaka, H. Imanaka, *J. Antibiot.* **1992**, *45*, 1055–1063.
[8] Z. Yu, S. Vodanovic-Jankovic, M. Kron, B. Shen, *Org. Lett*. **2012**, *14*, 4946–4949.
[9] M. Bae, H. Kim, K. Moon, S.-J. Nam, J. Shin, K.-B. Oh, D.-C. Oh, *Org. Lett*. **2015**, *17*, 712–715.
[10] M. Bae, J. Oh, E. S. Bae, J. Oh, J. Hur, Y.-G. Suh, S. K. Lee, J. Shin, D.-C. Oh, *Org. Lett*. **2018**, *20*, 1999–2002.
[11] C. W. Johnston, M. A. Skinnider, M. A. Wyatt, X. Li, M. R. Ranieri, L. Yang, D. L. Zechel, B. Ma, N. A. Magarvey, *Nat. Commun*. **2015**, *6*, 8421.
[12] S. Zhang, J. Zhu, D. L. Zechel, C. Jessen-Trefzer, R. T. Eastman, T. Paululat, A. Bechthold, *ChemBioChem* **2018**, *19*, 272–279.
[13] S. Omura, D. Van der Pyl, J. Inokoshi, Y. Takahashi, H. Takeshima, *J. Antibiot (Tokyo)*.**1993**, *46*, 222–228.
[14] S. Pohle, C. Appelt, M. Roux, H.-P. Fiedler, R. D. Süssmuth, *J. Am. Chem. Soc.* **2011**, *133*, 6194–6205.
[15] Z. Deng, J. Liu, T. Li, H. Li, Z. Liu, Y. Dong, W. Li, *Angew. Chem. Int. Ed Engl.* **2021**, *60*, 153–158.
[16] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493–505.
[17] G. L. Challis, J. Ravel, C. A. Townsend, *Chem. Biol*. **2000**, *7*, 211–224.
[18] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben, R. Breitling, E. Takano, M. H. Medema, *Nucleic Acids Res*. **2015**, *43*, W237–W243.
[19] M. Bernhardt, S. Berman, D. Zechel, A. Bechthold, *ChemBiochem* **2020**, *21*, 2659–2666.
[20] G. M. Singh, F. H. Vaillancourt, J. Yin, C. T. Walsh, *Chem. Biol.* **2007**, *14*, 31–40.
[21] E. R. Strieter, F. H. Vaillancourt, C. T. Walsh, *Biochemistry* **2007**, *46*, 7549-7557.
[22] U. Galm, E. Wendt-Pienkowski, L. Wang, N. P. George, T.-J. Oh, F. Yi, M. Tao, J. M. Coughlin, B. Shen, *Mol. Biosyst.* **2009**, *5*, 77–90.
[23] Z. Chang, P. Flatt, W. H. Gerwick, V. A. Nguyen, C. L. Willis, D. H. Sherman, *Gene* **2002**, *296*, 235–247.
[24] S. Garneau-Tsodikova, A. Stapon, D. Kahne, C. T. Walsh, *Biochemistry* **2006**, *45*, 8568–8578.
[25] V. Bergendahl, U. Linne, M. A. Marahiel, *Eur. J. Biochem*. **2002**, *269*, 620–629.
[26] A. M. Gulick, *ACS Chem. Biol*. **2009**, *4*, 811–827.
[27] T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure* **2000**, *8*, 407–418.
[28] L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, C. T. Walsh, *Biochemistry* **1998**, 37, 1585–1595.
[29] Y. Huang, E. Wendt-Pienkowski, B. Shen, *J. Biol. Chem.* **2006**, *281*, 29660–29668.
[30] I. Thomas, C. J. Martin, C. J. Wilkinson, J. Staunton, P. F. Leadlay, *Chem. Biol*. **2002**, *9*, 781–787.
[31] L. Gao, H. Liu, Z. Ma, J. Han, Z. Lu, C. Dai, F. Lv, X. Bie, *Sci. Rep*. **2016**, *6*, 38467.
[32] S. C. Wenzel, P. Meiser, T. M. Binz, T. Mahmud, R. Müller, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 2296–2301; *Angew. Chem*. **2006**, *118*, 2354-2360.
[33] A. Tanovic, S. A. Samel, L. O. Essen, M. A. Marahiel, *Science* **2008**, *321*, 659–663.
[34] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. L. Allen, G. Skiniotis, C. C. Aldrich, A. M. Gulick, *Nature* **2016**, *529*, 235–238.
[35] W.-H. Chen, K. Li, N. S. Guntaka, S. D. Bruner, *ACS Chem. Biol*. **2016**, *11*, 2293–2303.
[36] H. Chen, L. Du, *Appl. Microbiol. Biotechnol*. **2016**, *100*, 541–557.
[37] W. Xu, G. Zhai, Y. Liu, Y. Li, Y. Shi, K. Hong, H. Hong, P. F. Leadlay, Z. Deng, Y. Sun, *Angew. Chem. Int. Ed. Engl*. **2017**, *56*, 5503–5506; *Angew. Chem.* **2017**, *129*, 5595-5598.
[38] C. A. Shaw-Reid, N. L. Kelleher, H. C. Losey, A. M. Gehring, C. Berg, C. T. Walsh, *Chem. Biol*. **1999**, *6*, 385–400.
[39] K. M. Hoyer, C. Mahlert, M. A. Marahiel, *Chem. Biol*. **2007**, *14*, 13–22.
[40] D. A. Miller, C. T. Walsh, *Biochemistry* **2001**, *40*, 5313–5321.
[41] S.-S. Cha, Y. J. An, C.-S. Jeong, M.-K. Kim, S.-G. Lee, K.-H. Lee, B.-H. Oh, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2012**, *68*, 1253–1258.
[42] S.-J. Lee, S. J. Kim, I.-K. Kim, J. Ko, C.-S. Jeong, G.-H. Kim, C. Park, S.-O. Kang, P.-G. Suh, H.-S. Lee, S.-S. Cha, *J. Biol. Chem*. **2003**, *278*, 44552–44559.
[43] Y. Liu, T. Zheng, S. D. Bruner, *Chem. Biol*. **2011**, *18*, 1482–1488.
[44] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. U. S. A*. **2006**, *103*, 275–280.
[45] Y. Du, Y. Wang, T. Huang, M. Tao, Z. Deng, S. Lin, *BMC Microbiol*. **2014**, *14*, 30.

10.1002/anie.202103872

WILEY-VCH

RESEARCH ARTICLE

Entry for the Table of Contents



Unprecedented NRPS assembly line: Systematic manipulation of the nonribosomal peptide synthetase (NRPS) domains in combination with elucidation of the interactions between the unique domains and structure revealed several hitherto unprecedented nonlinear NRPS assembly processes during the biosynthesis of cyclodepsipeptide WS9326A.

Myoun-Su Kim,[+] Munhyung Bae,[+] Ye-Eun Jung,[+] Jung Min Kim, Sunghoon Hwang, Myoung Chong Song, Yeon Hee Ban, Eun Seo Bae, Suckchang Hong, Sang Kook Lee, Sun-Shin Cha, Dong-Chan Oh,* and Yeo Joon Yoon**

Page No. – Page No.
Unprecedented noncanonical features of the nonlinear nonribosomal peptide synthetase assembly line for WS9326A biosynthesis