Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

A transferable condensation domain linker contributes to iterative chain extension in an NRPS

Xiaolin Zou[+[1,2]], Jia-He Qiu[+[3,4]], Shuaiqiang Zhao[+[1]], Pengcheng Ji[[1,5]], Chang Zhou[[2]], Anfu Wei[[1]], Ke-Wei Chen[[3,4]], Jiakuan Zhao[[1]], Qinghong Jin[[1]], Yifei Yu[[1]], Haiqiang Wu[[1]], Mingwei Cai[[1]], Zhen Hui[[1]], Yong-Xin Li[[2]], Yun-Dong Wu*[[1,3,4,6]] & Xiaoyu Tang*[[1,7]]

[1] Institute of Chemical Biology, Shenzhen Bay Laboratory, 518132, Shenzhen, China

[2] Department of Chemistry and The Swire Institute of Marine Science, The University of Hong Kong, Hong Kong 999077, China

[3] Lab of Computational Chemistry and Drug Design, State Key Laboratory of Chemical Oncogenomics, Peking University Shenzhen Graduate School, Shenzhen 518055, China

[4] AI for Science (AI4S)-Preferred Program, Peking University Shenzhen Graduate School, Shenzhen 518055, China

[5] School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China

[6] College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

[7] Shenzhen Medical Academy of Research and Translation, Shenzhen 518107, China

[+]These authors contributed equally: Xiaolin Zou, Jia-He Qiu, Shuaiqiang Zhao.

*Email: wuyd@pkusz.edu.cn and xtang@microbechembio.org

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Abstract

Iterative nonribosomal peptide synthetases can assemble peptides longer than their module count predicts, yet the molecular features that support repeated chain extension remain poorly understood. Here we show that DidA, a dimodular synthetase, incorporates up to four glutamine residues into the *N*-acyl side chains of didemnins. Biochemical reconstitution identified the second condensation domain (C_2) and both thiolation domains as essential for repeated chain extension, while native-host mutagenesis confirmed their requirement for didemnin biosynthesis. Both condensation domains catalyzed initial acylation, but only C_2 accepted the longer *N*-acyl donors tested. Structural modelling and molecular dynamics simulations suggested conformational differences in the linker connecting the donor and acceptor subdomains of C_1 and C_2. Replacing the C_1 linker with its C_2 counterpart restored detectable didemnin B production in a C_2-inactivated background. Further dissection revealed that inserting just five residues from the C_2 linker into C_1 was sufficient for this rescue. These findings identify the donor-acceptor linker as a potential target for engineering condensation domain function in iterative peptide assembly.

Keywords: Nonribosomal peptide synthetase (NRPS); iterative chain extension; condensation domain; donor-acceptor linker; didemnin biosynthesis

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Introduction

Nonribosomal peptide synthetases (NRPSs) assemble structurally diverse natural products through the coordinated activities of adenylation (A), thiolation (T) and condensation (C) domains. In canonical assembly lines, modules act sequentially, with each module typically incorporating one building block into the growing peptide[1-3]. This collinear organization provides a framework for predicting natural product structures and engineering their biosynthesis[4,5] . Some NRPSs, however, use catalytic domains or modules more than once, allowing a compact assembly line to produce a peptide longer than its module count predicts[6-10].

Studies of nonlinear NRPSs have revealed several ways in which assembly can depart from collinearity, including intermodule substrate activation, module skipping, and repeated chain extension[6-10]. For example, functional characterization of fungal iterative NRPSs has provided insight into chain elongation and length control[7], while investigations of thalassospiramide biosynthesis have supported a model in which a growing chain returns to an upstream module for further extension[8]. For a condensation domain to catalyze successive rounds of extension, it must accept the product of one round as the donor for the next. However, the molecular features that allow some C domains to accommodate progressively longer donors remain poorly understood.

Didemnin biosynthesis offers a tractable system for addressing this question. The marine bacterial NRPS DidA is proposed to assemble the *N*-acyl side chains of didemnins X (DX, **2**) and Y (DY, **3**) (Fig. 1a, b). However, DidA contains only two C-A-T modules, whereas these side chains contain three and four glutamine residues, respectively[11]. These side chains are subsequently removed during maturation to didemnin B (DB, **1**) (Fig. 1b)[12,13]. The mismatch between module number and side-chain length suggests that DidA reuses one or more domains during chain extension. Because successive rounds of chain

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

extension require C domains to accept progressively longer donors, comparing the two C domains of DidA with the same series of substrates provides a way to assess their respective capacities to support further elongation. Furthermore, exchanging corresponding segments between them allows us to test whether local sequence differences affect didemnin biosynthesis.

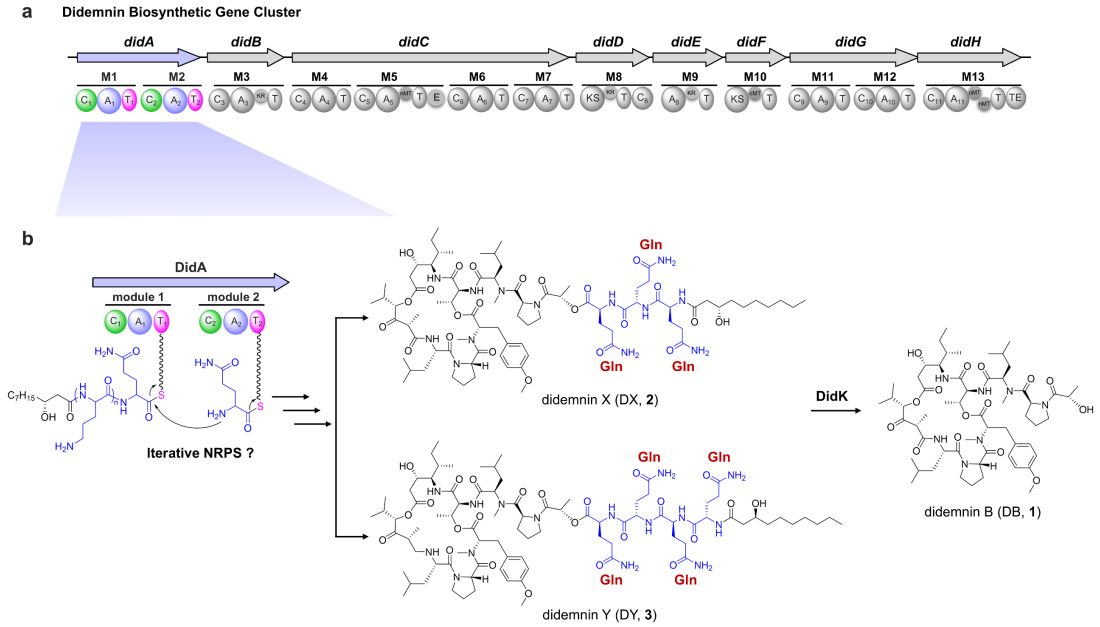

Fig. 1. The didemnin biosynthetic gene cluster and the proposed role of DidA in side-chain assembly. (**a**) Organization of the core didemnin biosynthetic gene cluster, with the two modules of DidA highlighted. Domain abbreviations: C, condensation; A, adenylation; T, thiolation; KR, ketoreductase; E, epimerization; MT, methyltransferase. KS, *β*-ketoacyl synthase; TE, thioesterase. (**b**) The two-module NRPS DidA is proposed to iteratively assemble the *N*-acyl side chains of didemnin X (DX, **2**) and didemnin Y (DY, **3**), which contain three and four glutamine residues, respectively. DidK cleaves these side chains during maturation to didemnin B (DB, **1**).

Here we combine *in vitro* reconstitution, native-host genetics, substrate profiling, and molecular modelling to investigate iterative chain extension by DidA. We show that repeated chain extension depends on C_2 and both T domains, and that although both C domains catalyze initial acylation, only C_2 accepts the longer *N*-acyl donors tested. Comparison of C_1 and C_2 models identifies a

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

donor–acceptor linker (DAL) that differs in predicted conformation between the two domains. Replacing the C_1 linker with its C_2 counterpart restores detectable DB production in a C_2-inactivated background, with insertion of just five residues from the C2 linker into C1 being sufficient for this rescue. Furthermore, molecular dynamics simulations suggest that these modifications alter the geometry of the substrate channel and the dynamics of the carrier-domain interface. Together, these findings link the distinct roles of C_1 and C_2 in iterative peptide assembly to a short linker sequence whose transfer allows C_1 to partially compensate for C_2 inactivation.

Results

DidA catalyzes iterative assembly of the *N*-acyl polyglutamine side chain

Didemnin X (DX, **2**) and didemnin Y (DY, **3**) contain *N*-acyl side chains comprising three and four glutamine-derived units, respectively, whereas DidA contains only two NRPS modules (Fig. 1 and Supplementary Fig. 1). Consistent with this organization, the AlphaFold3-predicted full-length structure of DidA revealed two tandem C-A-T modules (C_1-A_1-T_1 and C_2-A_2-T_2), with each adenylation domain further comprising core and subdomains (Supplementary Fig. 1). Structural superposition showed similar overall folds between A_1 and A_2 (RMSD = 0.563) and between T_1 and T_2 (RMSD = 0.645), supporting the proposed domain assignments (Supplementary Fig. 2). To determine whether this dimodular enzyme can incorporate more than two glutamine residues, we purified full-length DidA (Supplementary Fig. 5) and reconstituted the reaction *in vitro* using 3-hydroxydecanoic acid, L-glutamine, ATP, CoA, and an acyl-CoA synthetase from *Pseudomonas putida* (Fig. 2a)[9]. Following alkaline release of enzyme-bound intermediates, LC-MS analysis detected four products, F1G-F4G (**4**-**7**) (Fig. 2b), corresponding to *N*-acylated chains containing one to four glutamine residues (Fig. 2c-i and Fig. 2d-i). Their retention times and molecular masses matched those of chemically synthesized standards, and HRMS/MS

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

analysis further supported their assignments (Supplementary Fig. 6). The formation of F3G and F4G establishes that DidA can use its two modules iteratively to assemble chains of up to four glutamine residues.

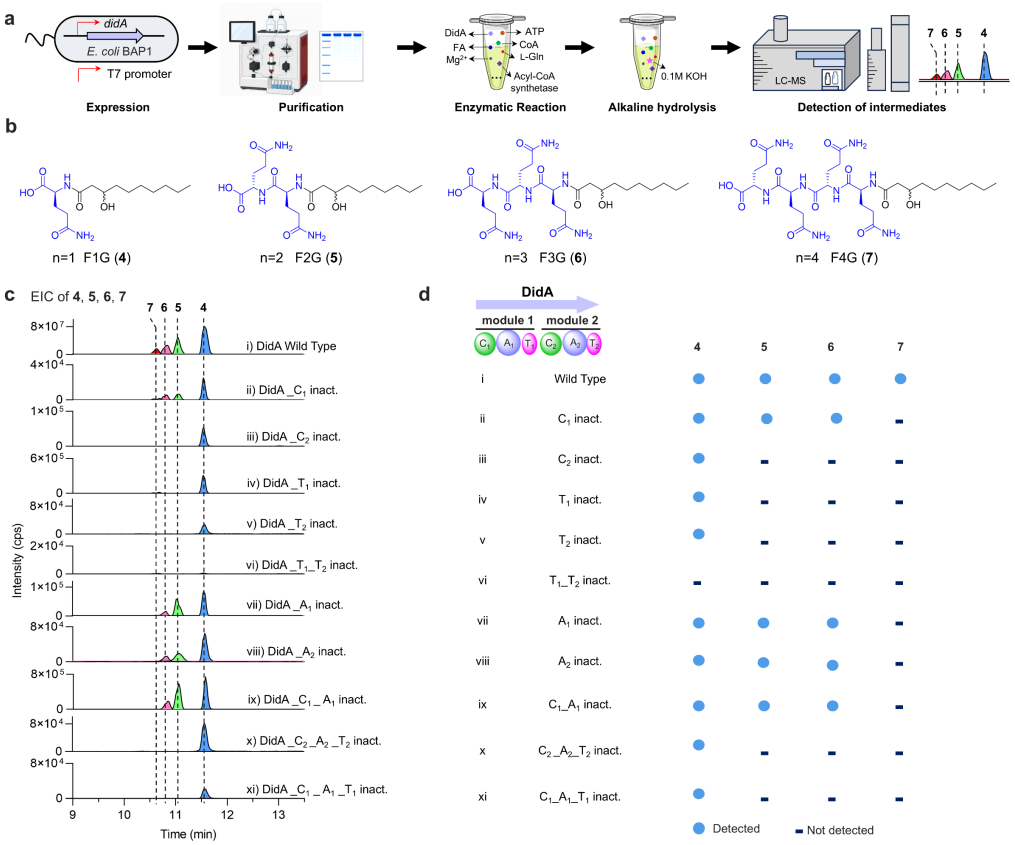

Fig. 2. DidA iteratively assembles *N*-acyl chains containing up to four glutamine residues *in vitro*. (**a**) Schematic workflow for the heterologous expression, purification, and enzymatic reconstitution of DidA. The *didA* gene is expressed in *E. coli* BAP1 under the control of a T_7 promoter, and the recombinant protein is purified by chromatography. The *in vitro* activity assay is performed in a reaction mixture containing purified DidA, 3-hydroxydecanoic acid (FA), ATP, Mg^{2+}, CoA, L-Gln, and an auxiliary acyl-CoA synthetase. Following alkaline hydrolysis (0.1 M KOH) to release covalent protein-bound intermediates, the reaction products and intermediates (**4**–**7**) are detected and analyzed by LC-MS. (**b**) Chemical structures of the products F1G-F4G (**4**-**7**), which contain one to four glutamine residues, respectively. (**c**) Extracted ion chromatograms (EICs) of F1G (**4**), F2G (**5**), F3G (**6**) and F4G (**7**) following alkaline release of enzyme-bound intermediates from reactions with wild-type DidA or the indicated domain-inactivated variants. Intensity scales differ among traces. (**d**) Product profiles of wild-type DidA

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

and the variants shown in **c**. Blue circles indicate detected products, and dashes indicate products not detected under the assay conditions. Inact., inactivated.

Iterative elongation requires C_2 and both thiolation domains

To identify the domains required for repeated chain extension, we generated full-length DidA variants carrying substitutions at conserved catalytic or substrate-loading residues in one or more C, A or T domains. We expressed the variants in *E. coli* BAP1[14], purified them and assayed them under the same conditions as wild-type DidA (Supplementary Fig. 5).

After C_1 inactivation, we still detected F1G, F2G and F3G (Fig. 2c-ii and Fig. 2d-ii), indicating that DidA can extend the chain to three glutamine residues without an active C_1 domain. By contrast, F1G was the only product detected after C_2 inactivation (Fig. 2c-iii and Fig. 2d-iii). Inactivation of either T domain produced the same profile (Fig. 2c-iv, v and Fig. 2d-iv, v), whereas simultaneous inactivation of T_1 and T_2 abolished all detectable *N*-acyl glutamine products (Fig. 2c-vi and Fig. 2d-vi). Thus, C_2 and both T domains are required for extension beyond the first glutamine unit.

We next examined whether both A domains were required for repeated glutamine incorporation. Inactivation of either A_1 or A_2 yielded F1G, F2G and F3G (Fig. 2c-vii, viii and Fig. 2d-vii, viii), indicating that either A domain can support extension to F3G. F3G remained detectable even after simultaneous inactivation of C_1 and A_1 (Fig. 2c-ix and Fig. 2d-ix). However, among all constructs tested, only wild-type DidA yielded detectable F4G (Fig. 2c-i and Fig. 2d-i). HRMS/MS data for these variants are shown in Supplementary Fig. 7–12.

Finally, we asked whether either module could support the initial acylation of glutamine when all three domains of the other module were inactivated. Both variants formed F1G as the only detectable product (Fig. 2c-x, xi and Fig. 2d-x, xi). These results indicate that either module can support initial acylation within

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

full-length DidA, whereas subsequent chain extension requires C_2 and the T domains of both modules.

Native-host genetics supports the domain requirements identified *in vitro*

To determine whether the domain requirements observed *in vitro* also applied in the native host, we used *Tistrella mobilis* L17B[15], an engineered derivative of the native producer L17 that carries an attachment site (*attB*) for site-specific integration[16]. We first deleted *didA* from this strain, which abolished detectable DB production (Fig. 3b-i and 3c-i, Supplementary Fig. 16 and Supplementary Fig. 40). We then integrated wild-type or mutant *didA* alleles at the *attB* site in the deletion strain (Fig. 3a). Complementation with wild-type *didA* restored DB production (Fig. 3b-ii and 3c-ii).

We first examined the two T domains. Replacing either Ser1001 in T_1 or Ser2046 in T_2 with alanine abolished detectable DB production, as did replacing both phosphopantetheinylation sites simultaneously (Fig. 3b-iii, iv, v and 3c-iii, iv, v). The two T domains are therefore nonredundant in the native pathway, consistent with their joint requirement for chain extension *in vitro*.

In contrast, strains carrying an inactivated A_1 or A_2 domain each produced approximately 25% of the parental-strain level (Fig. 3b-vi, vii and 3c-vi, vii, and Supplementary Fig. 39). The two C domains also differed in their contributions to DB production. The C_1-inactivated strain retained approximately 22% of the control level, whereas DB production was undetectable in the C_2-inactivated strain (Fig. 3b-viii, ix and 3c-viii, ix). Combined inactivation of C_1 with either A domain reduced DB production to approximately 0.2% of the control level (Fig. 3b-x, xi and 3c-x, xi). Together, these results show that C_2 and both T domains are required for detectable DB production in the native host, whereas the pathway retains partial activity after inactivation of C_1, A_1 or A_2 individually.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

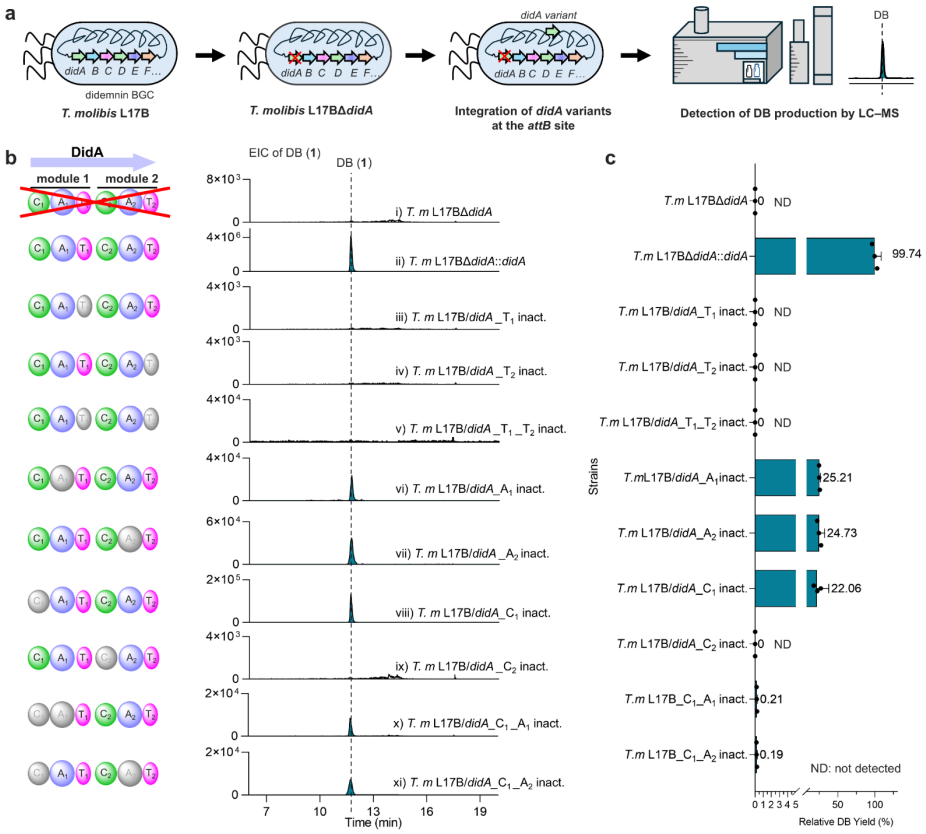

Fig. 3. DB production requires C_2 and both T domains and can be supported by either A domain. (**a**) Genetic complementation and LC-MS workflow in *T. mobilis* L17B. We deleted *didA* from the parental strain, integrated wild-type or mutant *didA* alleles at the *attB* site of the deletion strain, and measured didemnin B (DB, **1**) production by LC-MS. (**b**) DidA domain architectures and corresponding extracted ion chromatograms of DB for the indicated strains. The red cross denotes *didA* deletion, and grey shading marks individual domains inactivated by point mutation. Intensity scales differ among traces. (**c**) Relative DB yields in the strains shown in b, with DB production in the parental *T. mobilis* L17B strain set to 100%. Data points represent individual biological replicates. ND, not detected. Bars represent means ± s.d. from three independent cultures; dots indicate individual biological replicates.

C_2 accommodates longer *N*-acyl intermediates than C_1

The distinct effects of C_1 and C_2 inactivation suggested that the two domains differ in their ability to process intermediates during elongation. Sequence alignment showed that both retain the conserved C-domain HHxxxDG motif

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

(Fig. 4a and Supplementary Fig. 3)[17-19]. In structural models of C_1 and C_2, residues in this motif aligned closely with their counterparts in the crystal structure of the LgrA condensation domain (PDB: 6MFY; Fig. 4b and Supplementary Fig. 30)[20]. We therefore asked whether C_1 and C_2 differ in the lengths of the donor substrates they can accept.

We expressed and purified C_1 and C_2 separately and assayed each domain with synthetic *N*-acetylcysteamine (SNAC) thioesters as substrate mimics[21]. The substrate panel comprised glutaminyl-SNAC, fatty acyl-SNAC and a series of *N*-acyl donors containing one to three glutamine residues (Fig. 4c). We first asked whether C_2 could catalyze the initial acylation of glutamine. When supplied with fatty acyl-SNAC and glutaminyl-SNAC, wild-type C_1 and C_2 formed comparable amounts of F1G under the assay conditions (Fig. 4d). Each of the C_2 substitutions tested reduced F1G formation. In particular, the variant carrying an alanine substitution at the second histidine of the HHxxxDG motif retained approximately one-third of wild-type C_2 activity (Fig. 4d).

We next compared the ability of C_1 and C_2 to process progressively elongated *N*-acyl donors. Comparison with synthetic standards showed that C_1 produced F1G and F2G, whereas neither F3G nor F4G was detectable in the C_1 assays (Fig. 4e, traces i–v). C_2, by contrast, supported formation of all four products (Fig. 4e, traces vi–ix). Thus, both domains catalyzed initial acylation and formation of F2G, but only C_2 accepted F2G-SNAC and F3G-SNAC as donors for further extension. This difference in donor acceptance is consistent with the requirement for C_2 during iterative elongation by full-length DidA. HRMS/MS data for the C_1 and C_2 assays are shown in Supplementary Fig. 33–36.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

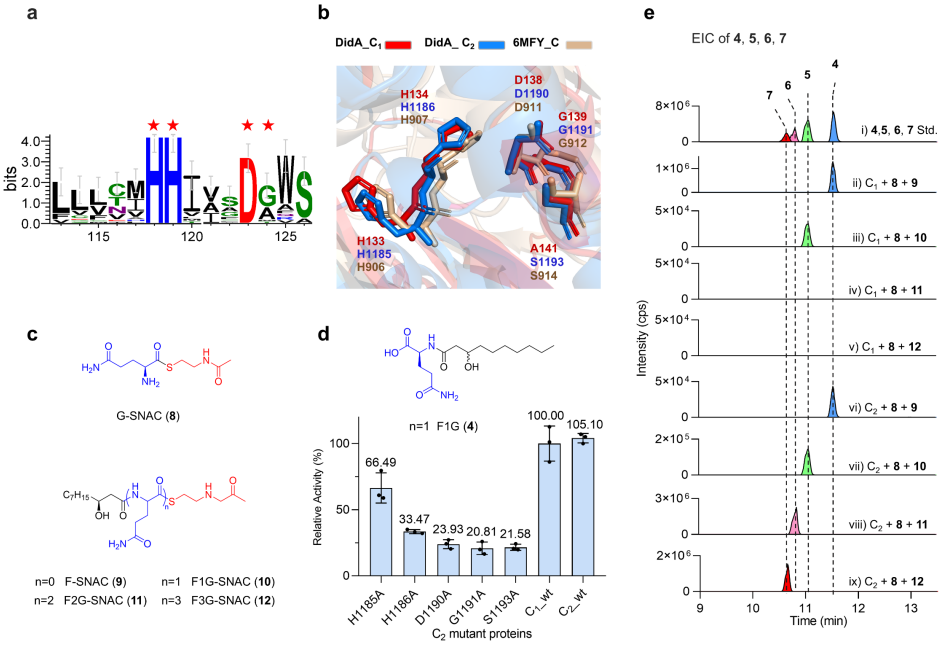

Fig. 4. C_2 accepts longer *N*-acyl donors than C_1 *in vitro*. (**a**) Sequence logo of the conserved region surrounding the HHxxxDG motif in condensation domains. Asterisks mark the two histidines, the aspartate and the glycine of the motif. (**b**) Structural superposition of selected active-site residues from the predicted DidA C_1 (red) and C_2 (blue) structures and the crystal structure of the LgrA condensation domain (*Brevibacillus parabrevis*, PDB 6MFY, tan)[20]. (**c**) Chemical structures of the SNAC thioester substrate mimics used in the assays. Glutaminyl-SNAC (**8**) served as the acceptor, and compounds **9**–**12** served as donors. (**d**) Relative F1G formation by wild-type C_1, wild-type C_2 and C_2 variants carrying individual alanine substitutions. F1G formation by wild-type C_1 was set to 100%. WT, wild type. Bars represent means ± s.d. from three independent experiments; dots indicate individual replicates. (**e**) Extracted ion chromatograms of F1G–F4G (**4**–**7**). Trace i shows the synthetic standards. Traces ii–v (C_1) and vi–ix (C_2) show the products formed with acceptor **8** and donors **9**–**12**, tested individually in the order shown. Intensity scales differ among traces.

A C_2-derived linker restores detectable DB production when transplanted into C_1

To place the functional divergence of C_1 and C_2 in an evolutionary context, we compared them with 130 condensation-related domains from 44 biosynthetic

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

gene clusters using reference sequences from NaPDoS2[22] and MIBiG 4.0[23] (Supplementary table 4). C_1 grouped with starter condensation domains, whereas C_2 clustered within an $^{L}C_L$-associated branch that included bacterial C domains involved in iterative elongation (Fig. 5a). This placement of C_2 is consistent with its ability to accept longer *N*-acyl donors *in vitro*.

We next compared AlphaFold[24] models of C_1 and C_2 to identify local differences that might contribute to their distinct activities. One such difference occurred in the region connecting the donor and acceptor subdomains, which we refer to as the donor–acceptor linker (DAL). The corresponding regions in C_1 and C_2 were designated DAL_1 and DAL_2, respectively. The models predicted a loop in DAL_1 and a short helix in DAL_2 (Fig. 5b).

To test the functional significance of this difference, we first replaced DAL_2 in C_2 with DAL_1. DB production fell to approximately 0.56% of the reference level (Fig. 5c-v and 5d-v, Supplementary Fig. 40). We then asked whether DAL_2 could support DB production when introduced into C_1. For this experiment, we used a strain carrying catalytically inactivated C_2, in which DB production was undetectable (Fig. 5c-iii and 5d-iii, and Supplementary Fig. 39). Replacing DAL_1 in C_1 with DAL_2 restored detectable DB production to approximately 0.25% of the reference level (Fig. 5c-iv and 5d-iv).

Thus, replacing DAL_2 in its native C_2 context sharply reduced DB production, yet the same linker retained a measurable biosynthetic effect when transferred to C_1. We therefore asked whether a shorter segment of DAL_2 would be sufficient for this rescue.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

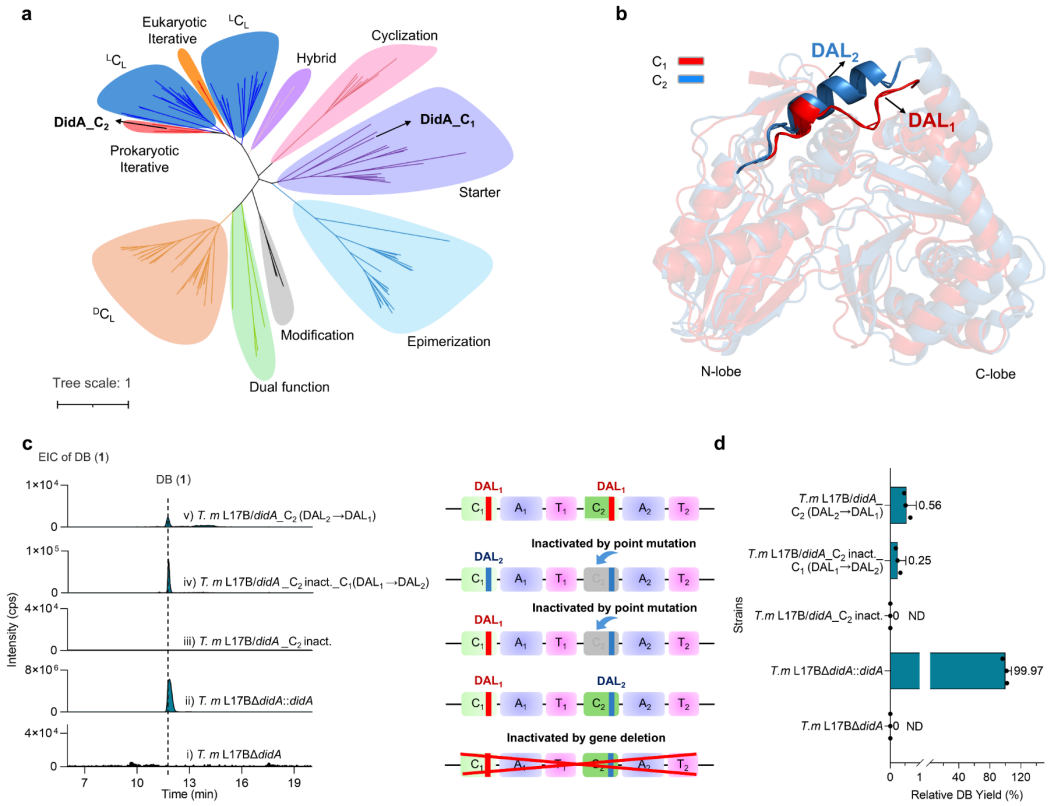

Fig. 5. Transferring DAL_2 to C_1 restores detectable DB production in a C_2-inactivated background. (**a**) Phylogenetic tree of DidA C_1 and C_2 and reference condensation-related domains. Colors indicate functional subclasses[18,25,26]. C_1 groups with starter condensation domains, whereas C_2 lies within an $^{L}C_L$-associated branch containing bacterial iterative C domains. (**b**) Superposition of the predicted structures of C_1 and C_2 (RMSD of 2.57 Å over 1,559 aligned atom pairs). The donor–acceptor linkers DAL_1 (C_1 residues 149–175) and DAL_2 (C_2 residues 1232–1258) are highlighted in red and blue, respectively. (**c**) EICs of DB (**1**) (left) and schematic representations of the corresponding DidA domain architectures (right). The red cross denotes *didA* deletion. Grey shading and blue arrows mark C_2 inactivation by point mutation. Red and blue bars indicate DAL_1 and DAL_2, respectively. Trace iv shows the strain in which DAL_2 replaces DAL_1 in C_1 while C_2 is inactive. Trace v shows the strain in which DAL_1 replaces DAL_2 in C_2. Intensity scales differ among traces. (**d**) Relative DB yields in the indicated strains quantified by LC-MS. DB production in the parental *T. mobilis* L17B strain was used as the 100% reference. Bars represent means ± s.d. from three independent cultures; dots indicate

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

individual biological replicates. ND, not detected. The broken x-axis allows the low DB yields of the linker variants to be shown alongside the complementation control.

A five-residue insertion into C_1 restores detectable DB production

We first asked whether a shorter segment of DAL_2 could retain the activity of the full linker after transfer to C_1 (Fig. 6a). To examine how residues within DAL_2 move and interact, we analyzed three independent 1-μs molecular dynamics simulations per system, focusing on the 27-residue DAL_2 region of C_2 (residues 1232–1258)[27,28]. We combined dynamic cross-correlation and residue-contact occupancy analyses to characterize interactions within this region and compared the resulting boundaries with those of the segments tested experimentally (Fig. 6b,c)[29].

We constructed six variants (EX1–EX6) to test C_2-derived segments that differed in length and position (Fig. 6c,d and Supplementary Table 5). In EX1, the full 27-residue DAL_2 segment replaced C_1 residues 151–172. In EX6, we inserted only five residues from DAL_2 (QRRRL, residues 1250–1254) between C_1 residues 168 and 169, leaving the native C_1 sequence intact. All six C_1 variants were introduced into full-length DidA carrying the C_2-inactivating mutations and tested by complementation of the *didA*-deletion strain.

LC-MS analysis detected DB production in all six engineered strains, whereas DB was undetectable in the control strain complemented with C_2-inactivated DidA carrying unmodified C_1 (Fig. 5c,d and Fig. 6h). EX6 gave the highest mean relative DB yield among the six variants (0.36%), showing that insertion of

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

QRRRL alone was sufficient for detectable rescue in this background.

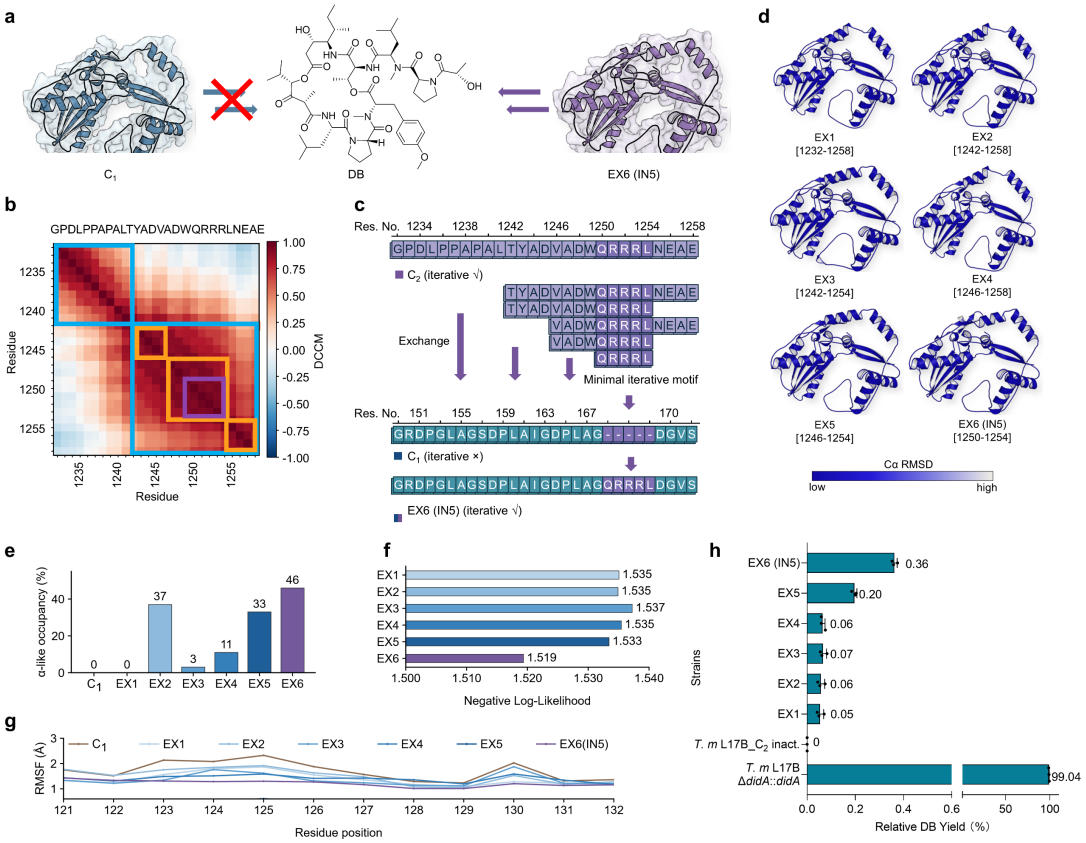

Fig. 6. Gradient partitioning and functional analysis identify a five-residue C_2 segment that restores detectable DB production when inserted into C_1. (**a**) Schematic illustrating the objective of progressively reducing the C_2-derived DAL segment introduced into C_1 while retaining detectable DB production in a C_2-inactivated background. The image was rendered using PyMOL[30]. (**b**) Residue-level dynamic cross-correlation matrix (DCCM) for DAL_2 (C_2 residues 1232–1258), 4-μs molecular dynamics trajectories[29]. Positive and negative values indicate correlated and anticorrelated residue motions, respectively. Colored boxes mark the residue ranges compared in the partitioning analysis. (**c**) Determination of the smallest functional segment tested in C_2 and its functional transplantation into C_1. Top, amino acid sequence alignment (Res. No. 1232–1258). The purple region indicates the C_2 domain (with iterative activity, √), which was progressively shortened to C_2-derived segments to identify the smallest functional segment tested (QRRRL). The "Exchange" arrows denote the stepwise transplantation of these truncated C_2 fragments into the corresponding gap (----) of the C_1 domain (blue, lacking iterative activity, ×). Bottom, the detectable DB production. (**d**) AF3

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

structures of EX1-EX6 aligned to the parental C_1 structure. Labels indicate the inclusive C2-derived residue range introduced into each construct. Coloring represents the Cα RMSD relative to the parental C_1 for positions corresponding to C_1. (**e**-**g**) Possible structural consequences of the EX series mutants. (**e**) α-Like occupancy was calculated, with frames counted when more than 60% of residues in the analyzed segment satisfied the predefined α-helical φ/ψ criterion. (f) Whole chain means NLL for AF3 structures scored with ProteinMPNN. (**g**) Cα RMSF profiles for C_1 and the corresponding positions in EX1–EX6. (**h**) Experimental evaluation of EX1-EX6 by LC-MS. DB production was detected in all six engineered strains. Bars represent means ± s.d. from three independent cultures; dots indicate individual biological replicates.

We next used structural modelling and simulations to examine how the C_2-derived segments affected C_1. Relative to parental C_1, several variants more frequently sampled helix-compatible conformations in the exchanged region (Fig. 6e). EX5 and EX6 showed the greatest native sequence-backbone compatibility (Fig. 6f). RMSF profiles further revealed construct-dependent changes at several positions surrounding the conserved HHxxxDG region (Fig. 6g), suggesting that the insertions remodel local dynamics near the catalytic pocket. These results suggest that the insertions affect regions involved in substrate accommodatio[31-34] and carrier-domain engagement[5].

The substrate assays and transplantation experiments thus connect two observations: C_2 accepts longer *N*-acyl donors than C_1, and a five-residue segment from its linker allows C_1 to support detectable DB production when C_2 is inactivated.

Discussion

Understanding iterative NRPS assembly requires identifying both how catalytic domains are reused and which molecular features allow them to support repeated chain extension. Here, we establish that the dimodular NRPS DidA assembles *N*-acyl polyglutamine chains through iterative catalysis and identify

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

a local sequence element that contributes to the functional specialization of its condensation domains. Full-length enzyme reconstitution demonstrates incorporation of up to four glutamine units, while biochemical and genetic dissection reveals that C_2 and both T domains are required for repeated extension. Moreover, replacing the C_1 linker with DAL_2 restores detectable DB production in a C_2-inactivated background, and insertion of just five residues from DAL_2 is sufficient for this rescue.

Previous studies have reported noncollinear assembly behaviors in NRPS systems and proposed models involving intermodular substrate activation, module skipping, and pass-back chain extension[6-10]. These findings suggest that catalytic domains or modules can be reused during peptide assembly. Therefore, successive rounds of extension also impose a requirement on the elongating C domain: it must accept the product of one round as the donor for the next. In DidA, C_1 and C_2 share a conserved catalytic motif, and both catalyze the initial acylation of glutamine, yet only C_2 accepts the longer *N*-acyl donors tested. This difference helps explain why DidA retains the ability to form F3G when C_1 is inactivated, whereas inactivation of C_2 restricts the detectable products to F1G (Fig. 2c,d). Both T domains are also required for extension beyond F1G (Fig. 2c,d), supporting a model in which the two carrier domains cooperate during repeated chain extension (Supplementary Fig 42). A similar requirement for two distinct T domains has been reported for SidD, in which one elongating C domain supports repeated bond formation[35]. Together, these observations suggest that repeated extension by DidA relies on C_2 to accommodate the growing chain and on both carrier domains to support successive rounds of elongation.

Functional transplantation provides a means to test which local features contribute to this specialization. Replacing DAL_2 in C_2 with DAL_1 strongly reduced DB production, indicating that the two regions are not functionally interchangeable in this context. Conversely, introducing DAL_2 into C_1 restored

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

detectable DB production in a C_2-inactivated background. Among the six variants tested, the five-residue insertion QRRRL gave the highest mean relative DB yield (0.36%; Fig. 6h). This rescue demonstrates that a five-residue insertion into C_1 can partially compensate for C_2 inactivation at the pathway level. Molecular dynamics simulations showed lower residue fluctuations in the catalytic-pocket and C-T interface regions of the engineered variants, together with a larger calculated volume for the pocket and adjoining channel. These changes suggest that the insertions may affect substrate accommodation or carrier-domain engagement, providing a possible explanation for the recovered biosynthetic activity.

Several aspects of the iterative mechanism remain to be resolved. The requirement for both T domains does not establish the order in which they engage C_2 or the route by which growing intermediates are transferred. Moreover, the engineering experiments measured DB, whose *N*-acyl polyglutamine appendage is removed during maturation. DB recovery therefore establishes that the engineered pathway is functional but leaves open how far the modified DidA extends the side chain. Characterizing the intermediates formed by the engineered variants would establish their chain lengths and help distinguish effects on elongation from effects on transfer to downstream modules. Structural and biophysical studies would complement this analysis by testing how the DAL insertions affect substrate binding and carrier-domain engagement[36,37].

The comparison between C_1 and C_2 thus links a biochemical difference between two domains in the same assembly line to a sequence element that can be exchanged between them. More broadly, the DAL provides a starting point for testing whether local sequence changes can alter the range of chain lengths accepted by other NRPS condensation domains.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

References

1 Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* **3**, 490-504 (2002).

2 Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468-3496 (2006).

3 Süssmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis-principles and prospects. *Angew Chem. Int. Ed.* **56**, 3770-3821 (2017).

4 Bozhüyük, K. A. J. *et al.* Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653-661 (2019).

5 Bozhüyük, K. A. J. *et al.* Evolution-inspired engineering of nonribosomal peptide synthetases. *Science* **383**, eadg4320 (2024).

6 Hoyer, K. M., Mahlert, C. & Marahiel, M. A. The iterative gramicidin S thioesterase catalyzes peptide ligation and cyclization. *Chem. Biol.* **14**, 13-22 (2007).

7 Yu, D., Xu, F., Zhang, S. & Zhan, J. Decoding and reprogramming fungal iterative nonribosomal peptide synthetases. *Nat. Commun.* **8**, 15349 (2017).

8 Zhang, J. J., Tang, X., Huan, T., Ross, A. C. & Moore, B. S. Pass-back chain extension expands multimodular assembly line biosynthesis. *Nat. Chem. Biol.* **16**, 42-49 (2020).

9 Cui, J. *et al.* Biosurfactant biosynthesis by *Alcanivorax borkumensis* and its role in oil biodegradation. *Nat. Chem. Biol.* **21**, 1631-1641 (2025).

10 Zhong, L. *et al.* The intricate nonribosomal assembly of a potent antifungal lipopeptide from the *Burkholderia cepacia* complex. *J. Am. Chem. Soc.* **147**, 20725-20734 (2025).

11 Xu, Y. *et al.* Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. *J. Am. Chem. Soc.* **134**, 8625 (2012).

12 Jia, K. *et al.* Abi family protein, DidK, is involved in the maturation of anticancer depsipeptide, didemnin B. *ACS Chem. Biol.* **18**, 2300-2308 (2023).

13 Zou, X. *et al.* Unveiling a CAAX protease-like protein involved in didemnin drug maturation and secretion. *Adv. Sci.* **11**, 2306044 (2024).

14 Quadri, L. E. *et al.* Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585-1595 (1998).

15 Zhang, H. *et al.* A semisynthesis platform for the efficient production and exploration of didemnin-based drugs. *Angew Chem. Int. Ed. Engl.* **63**, e202318784 (2024).

16 Zhang, J. J., Moore, B. S. & Tang, X. Engineering *Salinispora tropica* for heterologous expression of natural product biosynthetic gene clusters. *Appl. Microbiol. Biotechnol.* **102**, 8437-8446 (2018).

17 Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522-526 (2002).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

18 Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78 (2007).

19 Izoré, T. *et al.* Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity. *Nat. Commun.* **12**, 2511 (2021).

20 Reimer, J. M. *et al.* Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366** (2019).

21 Ehmann, D. E., Trauger, J. W., Stachelhaus, T. & Walsh, C. T. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Cell Chem. Biol.* **7**, 765-772 (2000).

22 Klau, L. J. *et al.* The Natural Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function. *J. Biol. Chem.* **298**, 102480 (2022).

23 Zdouc, M. M. *et al.* MIBiG 4.0: advancing biosynthetic gene cluster curation through global collaboration. *Nucleic Acids Res.* **53**, D678-D690 (2025).

24 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024).

25 Wheadon, M. J. & Townsend, C. A. Evolutionary and functional analysis of an NRPS condensation domain integrates β-lactam, D-amino acid, and dehydroamino acid synthesis. *Proc. Natl. Acad. Sci. USA* **118**, e2026017118 (2021).

26 Cao, W. *et al.* Analysis of heterocycle formation and stereochemical control by a non-ribosomal peptide synthetase condensation domain. *Nat. Synth.* **5**, 1054-1069 (2026).

27 Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195 (2008).

28 Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49-56 (2022).

29 Meng, H. *et al.* Precise DNA base editing using AlphaFold3-based contact modelling. *Nature* **656**, 463-473 (2026).

30 DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. protein crystallogr* **40**, 82-92 (2002).

31 Fushinobu, S., Nishimasu, H., Hattori, D., Song, H. J. & Wakagi, T. Structural basis for the bifunctionality of fructose-1, 6-bisphosphate aldolase/phosphatase. *Nature* **478**, 538-541 (2011).

32 Borges, P. T. *et al.* Methionine-rich loop of multicopper oxidase McoA follows open-to-close transitions with a role in enzyme catalysis. *ACS Catal.* **10**, 7162-7176 (2020).

33 Corbella, M., Pinto, G. P. & Kamerlin, S. C. Loop dynamics and the evolution of enzyme activity. *Nat. Rev. Chem.* **7**, 536-547 (2023).

34 Harding, C. J., Bischoff, M., Bergkessel, M. & Czekster, C. M. An anti-biofilm cyclic peptide targets a secreted aminopeptidase from *P. aeruginosa*. *Nat. Chem. Biol.* **19**, 1158-1166 (2023).

35 Jenner, M. *et al.* Elucidating the molecular programming of a nonlinear non-ribosomal peptide synthetase responsible for fungal siderophore biosynthesis. *Nat. Commun.* **14**, 2832 (2023).

36 Heberlig, G. W., La Clair, J. J. & Burkart, M. D. Crosslinking intermodular

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

condensation in non-ribosomal peptide biosynthesis. *Nature* **638**, 261-269 (2025).

37 Pistofidis, A. *et al.* Structures and mechanism of condensation in non-ribosomal peptide synthesis. *Nature* **638**, 270-278 (2025).

38 Maier, J. A. *et al.* ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696-3713 (2015).

39 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935 (1983).

40 Kang, W., Jiang, F. & Wu, Y.-D. Universal implementation of a residue-specific force field based on CMAP potentials and free energy decomposition. *J. Chem. Theory Comput.* **14**, 4474-4486 (2018).

41 Chen, K.-W. *et al.* Converting dimeric enzymes into efficient monomeric catalysts using artificial intelligence-assisted models. *CCS Chem.* **8**, 2348-2362 (2026).

42 Case, D. *et al.* AMBER 20. *University of California: San Francisco, CA, USA*, 1-918 (2020).

43 Roe, D. R. & Cheatham, T. E., III. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084-3095 (2013).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Methods

Bacterial strains and growth conditions

T. mobilis strains were recovered from frozen stocks stored at −80 °C. A 10-μL aliquot of each stock was spread on Luria–Bertani (LB) agar (10.0 g/L NaCl, 10.0 g/L tryptone, 5.0 g/L yeast extract and 12.0 g/L agar) containing 50 μg/mL ampicillin. After incubation at 25 °C for 2 days, three or four well-isolated large colonies were used to inoculate seed medium (1.0 g/L galactose, 5.0 g/L glycerol, 3.0 g/L yeast extract and 5.0 g/L peptone). After 2 days, the seed culture reached an OD_{600} of 0.6-0.8 and was used to inoculate 600 mL of R2 medium (4.0 g/L galactose, 20.0 g/L glycerol, 12.0 g/L yeast extract and 20.0 g/L peptone) at a 1:100 inoculum-to-medium ratio. Cultures were incubated at 25 °C with shaking at 220 rpm for 7 days. Aliquots were collected every 24 h to measure pH and OD_{600}.

E. coli strains were grown on LB agar or in LB broth at 37 °C with the appropriate antibiotic for selection (ampicillin or kanamycin, each at 50 μg/mL). For protein production, *E. coli* BAP1 cultures were grown in Terrific Broth (TB; Sangon, Shanghai, China) with the appropriate antibiotic at 37 °C for 2–3 h until they reached an OD_{600} of 0.6-0.8.

Site-directed mutagenesis and construction of *didA* expression plasmids

Genes encoding full-length DidA variants were synthesized and cloned into pET-28a(+) by Sangon Biotech (Shanghai, China). The DidA_C_1 variant contains the following substitutions in the C_1 region: H133A, H134A, D138A, and G139A. The DidA_C_2 variant contains the following substitutions in the C_2 region: H1185A, H1186A, D1190A, G1191A, and S1193A. For A_1, alanine substitutions at Thr604, Ser605, Ser607 and Thr608 generated DidA_A_1 inact., whereas alanine substitutions at Thr1646, Ser1647, Ser1649 and Thr1650 generated DidA_A_2 inact. DidA_T_1 inact. and DidA_T_2 inact. carried the single alanine substitutions S1001A and S2046A, respectively. DidA_T_1_T_2 inact.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

carried both substitutions.

E. coli BAP1 cells were transformed with the recombinant plasmids, and transformants were selected on kanamycin-containing medium. Plasmid DNA was purified using the OMEGA Spin Miniprep kit (OMEGA, USA). All constructs were verified by restriction digestion and Sanger sequencing (Sangon Biotech, Shanghai, China).

Protein overexpression and purification

The codon-optimized *didA* gene was synthesized and cloned into pET-28a(+) by Sangon Biotech (Shanghai, China). The resulting plasmid encoded DidA with an N-terminal His_6-tag under the control of the T_7 promoter. *E. coli* BAP1 was transformed with the plasmid, and transformants were selected on LB agar containing 50 μg/mL kanamycin.

A single colony was used to inoculate 10 mL of LB broth containing 50 μg/mL kanamycin, and the preculture was grown at 37 °C with shaking at 220 rpm for 4 h. A 5-mL aliquot of the preculture was then transferred to 500 mL of Terrific Broth (TB, Sangon Technology, Shanghai, China) containing 50 μg/mL kanamycin. The culture was grown at 37 °C with shaking at 220 rpm until the OD_{600} reached 0.6-0.8. Protein expression was induced with IPTG at a final concentration of 0.1 mM, and the culture was incubated at 16 °C with shaking at 160 rpm for 16 h. The cells were harvested by centrifugation at 4,500 rpm for 20 min and resuspended in 50 mM Tris-HCl (pH 8.0) containing 500 mM NaCl, 20 mM imidazole and 10% (v/v) glycerol. The cells were disrupted by sonication at 40% amplitude for 40 min using 3-s pulses separated by 9-s intervals. The lysate was centrifuged at 18,000 rpm for 30 min, and His_6-tagged DidA was purified from the clarified supernatant by using HisTrap columns (Cytiva).

The columns were first flushed with 10 column volumes of Milli-Q water and then equilibrated with 10 column volumes of lysis buffer at a flow rate of 2 mL/min using a peristaltic pump. Approximately 80 mL of clarified lysate was

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

loaded onto the column. The column was then washed with 10 column volumes of wash buffer (20 mM Tris–HCl, 500 mM NaCl, 20 mM imidazole, pH 8.0) at 1 mL/min. Bound protein was eluted with 5 column volumes of elution buffer (20 mM Tris–HCl, 500 mM NaCl, 300 mM imidazole, pH 8.0) at the same flow rate, and fractions were collected in 2-mL Eppendorf tubes. Protein-containing fractions were identified using the Bradford assay (Bio-Rad), pooled and concentrated to 2.5 mL using Amicon centrifugal filters with a molecular-weight cutoff of 100 kDa. The concentrated protein was desalted using PD-10 columns (Cytiva) pre-equilibrated with storage buffer (50 mM potassium phosphate, 10% (v/v) glycerol, pH 8.0), following the manufacturer's instructions. The final protein concentration was determined using the Bradford assay. DidA variants were expressed and purified using the same procedures as wild-type DidA.

Construction of *T. mobilis* strains carrying *didA* variants

The parental strain *T. mobilis* L17B was previously constructed by introducing a DNA fragment containing the bacterial attachment site (*attB*) for the ΦC31 integrase into the chromosome of *T. mobilis* L17[15], hereafter designated *T. m* L17B. The *didA* gene was then deleted from L17B to generate *T. mobilis* L17BΔ*didA*, following the procedure described previously[15].

Plasmids carrying *didA* variants were constructed to encode the same amino acid substitutions as those tested *in vitro*. Specifically, pXZ-*didA*_T_1 inact. encoded the S1001A substitution in T_1, pXZ-*didA*_T_2 inact. encoded the S2046A substitution in T_2, and pXZ-*didA*_T_1T_2 inact. encoded both substitutions.

Each of the three plasmids was transferred separately into *T. mobilis* L17BΔ*didA* by triparental intergeneric conjugation using *E. coli* ET12567/pUZ8002 as the helper strain. Apramycin-resistant transconjugants were isolated and designated *T. mobilis* L17B/*didA*_T_1 inact., *T. mobilis* L17B/*didA*_T_2 inact., and *T. mobilis* L17B/*didA*_T_1T_2 inact., respectively. The same procedure was used to introduce pXZ-*didA*, pXZ-*didA*_A_1 inact., pXZ-

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

*didA*_A_2 inact., pXZ-*didA*_C_1 inact. and pXZ-*didA*_C_2 inact. individually into the deletion strain. The resulting strains were designated *T. mobilis* L17B/*didA*, *T. mobilis* L17B/*didA*_A_1 inact., *T. mobilis* L17B/*didA*_A_2 inact. *T. mobilis* L17B/*didA*_C_1 inact. and *T. mobilis* L17B/*didA*_C_2 inact., respectively. The same procedure was used to generate the strains carrying combined C_1/A_1 or C_1/A_2 substitutions.

In vitro enzyme assays

DidA activity was assayed in a total volume of 500 μL in potassium phosphate buffer (50 mM, pH 7.5). The reaction mixture contained 5 μM acyl-CoA synthetase from *Pseudomonas putida*, 5 mM ATP, 5 mM DTT, 10 mM $MgCl_2$, 3 mM CoA, 3 mM 3-hydroxydecanoic acid and L-glutamine at 3, 6, 9 or 12 mM. Reactions were initiated by adding wild-type DidA or the corresponding variant to a final concentration of 100 μM and were then incubated at 30 °C for 16 h.

Reactions were quenched by adding trichloroacetic acid (TCA) to a final concentration of 10% (w/v), and precipitated proteins were collected by centrifugation. To release enzyme-bound intermediates, the protein pellet was resuspended in 500 μL of 0.1 M KOH and heated at 65 °C for 5 min. Proteins were then reprecipitated by adding 5 μL of 50% trifluoroacetic acid (TFA). After centrifugation, the supernatant containing the released products was collected for LC-MS analysis. Analyses were performed using an Agilent 6470 triple quadrupole LC-MS system and a SCIEX TripleTOF 6600+ quadrupole time-of-flight mass spectrometer coupled to a UPLC system.

HPLC and LC-MS analysis

For chromatographic separation, a 5.0 μL aliquot of each sample was injected onto either a Poroshell HPH-C18 column (2.1 × 100 mm, 2.7 μm) fitted with an Agilent HPH-C18 guard column (4.6 × 5 mm, 2.7 μm) or an Agilent Poroshell 120 EC-C18 column (100 × 4.6 mm, 2.7 μm, 100 Å). LC-MS analyses were performed using either a SCIEX TripleTOF 6600+ quadrupole time-of-flight

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

mass spectrometer or an Agilent 1290 Infinity II LC system coupled to a 6470 triple quadrupole mass spectrometer (Agilent, USA).

For high-resolution MS and MS/MS analyses on the TripleTOF 6600+, spectra were acquired in positive-ion mode over an m/z range of 100–2,000. The acquisition rate was 1.25 spectra s^{-1} for both MS and MS/MS, and the collision energy for MS/MS was fixed at 35 eV. The source temperature was set to 500 °C, the curtain gas to 35, the ion spray voltage to 5,500 V, and ion source gases 1 and 2 to 50 each. Before each run, the instrument was automatically calibrated in positive-ion mode using SCIEX tuning solution. TripleTOF data were acquired and processed using Analyst software (SCIEX, USA).

Structural modelling and molecular dynamics simulations

Structures of C_1, iterative C_2, and the engineered C_1 or C_2 variants were generated using AlphaFold 3 (AF3)[24], and were retained for ProteinMPNN analysis. Systems were processed with pdb4amber and tleap using ff14SB(ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB[38] and TIP3P water[39] solvated in 10-Å truncated-octahedral boxes and neutralized. After hydrogen-mass repartitioning with ParmEd, topologies were converted to RSFF2C with CMAP corrections[40,41]. Particle-mesh Ewald and a 9-Å real-space cutoff were used. Following 40,000-cycle minimization, 600-ps NVT heating to 300 K and 1-ns restrained NPT equilibration, unrestrained 1-μs NVT simulations were performed with GPU PMEMD in AMBER 20[42]. A 2-fs time step, SHAKE, a Langevin thermostat (2 ps-1) and 40-ps coordinate output were used.

Dynamic cross-correlation and contact-occupancy analysis

Trajectories were autoimaged, stripped of solvent and ions, and fitted to conserved-core Cα atoms while excluding common positions 159–185. For C_2, DCCM and Cα contact occupancies for residues 1232–1258 (common positions 159–185) were calculated in CPPTRAJ[43] from 4-μs trajectories

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

sampled every 0.4 ns. DCCM coefficients describe correlated or anticorrelated fluctuations, whereas contacts used an 8-Å cutoff and combined native and non-native maps. After summarization, joint edge weights were defined as the product of absolute DCCM and contact occupancy, thereby retaining pairs that were both dynamically coupled and persistently proximal. The highest 40% of non-zero edges were subjected to weighted modularity community detection (resolution 1.0); average-linkage clustering of the unthresholded matrices over 2–10 clusters provided a sensitivity analysis. Agreement with the five experimental blocks (159–168, 169–172, 173–176, 177–181 and 182–185) was evaluated using adjusted Rand index, normalized mutual information and within- versus between-block separation, with 10,000 size-preserving random assignments as the null model. α-Like occupancy was evaluated for C_1 residues 151–172 and EX1–EX6 residues 151–177: a residue was α-like when −100° ≤ φ ≤ −30° and −80° ≤ ψ ≤ −5°, and a frame was counted when more than 60% of the segment met both criteria. Cα RMSF values were calculated for residues 121–132. ProteinMPNN[28] was run in single-amino-acid scoring mode on the AF3 structures, with the native sequence supplied as context (single_aa_score = 1, use_sequence = 1; seed 111; ten batches; no redesign). For residue *i*, the conditional negative log-likelihood was calculated as $\mathrm{NLL}_i = -\ln \mathrm{P}(a_i^{\mathrm{native}} | B, a_{-i}^{\mathrm{native}})$, where *B* denotes the protein backbone, a_i is the native amino acid at position *i*, and a_{-i} denotes the native amino acids at all other sequence positions.The whole-chain mean NLL was obtained by averaging the residue-level NLL_i values over all *L* scored residues:$\overline{\mathrm{NLL}} = \frac{1}{L}\sum_{i=1}^{L} \mathrm{NLL}_i$.

Trajectory-weighted network construction and gradient partitioning

Network analyses were conducted independently for each system. Joint edge weights were defined as

$$[W_{ij} = \left|\overline{D}ij\right|\overline{C}ij,]$$

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

such that high weights required both strong dynamical coupling and persistent physical proximity. Self-edges were removed, and only the highest 40% of non-zero upper-triangle edge weights were retained for community detection.

Communities were identified without using the experimental block labels. Starting from singleton communities, pairs were greedily merged when doing so produced the largest positive increase in weighted modularity, using a resolution parameter of 1.0. Merging stopped when no further positive increase was obtained. Community detection was performed separately for the absolute-DCCM and joint networks. As a complementary sensitivity analysis, average-linkage clustering was applied to the corresponding unthresholded similarity matrices after conversion to distances:

$$[d_{ij} = \max(S) - S_{ij},]$$

where (S) denotes either the absolute-DCCM or joint-weight matrix. Partitions containing 2-10 clusters were evaluated, and the partition with the highest silhouette score was retained.

The experimental design comprised five contiguous blocks in the common numbering: 159-168, 169-172, 173-176, 177-181 and 182-185. Their post hoc agreement with the unsupervised communities was quantified using the adjusted Rand index and normalized mutual information. Block organization was further assessed as the difference between mean within-block and between-block weights. Scores calculated from the unthresholded absolute-DCCM, contact-occupancy and joint-weight matrices were compared with 10,000 random assignments preserving the five block sizes.

Catalytic pocket channel and interdomain interface analyses

The catalytic pocket and the adjoining substrate channel were defined once from the aligned C-domain models, and the same geometric definitions were applied to all constructs. Pocket stability was assessed from the distributions of

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

pocket-residue fluctuations and persistent intrapocket contacts. Pocket and channel volumes were calculated for each aligned trajectory frame with identical grid and probe settings across systems. C-T interface stability was evaluated from interdomain contact occupancies and the persistence of the interface contact network. Metrics were first summarized within each trajectory and then compared across the three independent replicas.

Acknowledgements

We thank B. S. Moore for helpful discussions and support, and the staff of the mass spectrometry and NMR platforms at the Biomedical Research Core Facilities of Shenzhen Bay Laboratory for assistance with MS and NMR experiments.

Funding

This work was supported by the Guangdong S&T Program (2024B1111160007 to X.T.), the Guangdong Basic and Applied Basic Research Foundation (2024A1515010922 to X.T.), the National Natural Science Foundation of China (82173719 to X.T.), the Shenzhen Bay Laboratory Startup Fund (21230051 to X.T.), the National Key Research and Development Program of China (2023YFA1506500 to Y.-D.W.), and the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM507 to Y.-D.W.). Additional support was provided by the Lab of Computational Chemistry & Drug Design, Shenzhen (SYSPG20241211173440006).

Author contributions

X.T. and X.Z. conceived and designed the project. X.Z. and P.J. carried out biochemical experiments, compound extraction and metabolic profiling. J.-H.Q. performed molecular docking and molecular dynamics simulations and

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.23.000398. This version posted November 23, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

developed strategies for DAL engineering under the supervision of Y.-D.W. and K.-W.C.. S.Z. performed chemical synthesis and compound isolation and assisted with NMR and MS data acquisition and analysis. C.Z. conducted bioinformatic analyses under the guidance of Y.-X.L.. A.W., J.Z., Q.J. and Y.Y. assisted with didemnin extraction and MS data analysis under the supervision of X.T.. H.W. assisted with protein overexpression and purification. M.C. contributed to phylogenetic analysis. Z.H. provided advice on genetic manipulation in *T. mobilis*. X.Z. and X.T. drafted the manuscript with input from all authors. X.T. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at XXXXX.

Correspondence and requests for materials should be addressed to Yun-Dong Wu or Xiaoyu Tang.