Aminoacyl chain translocation catalysed by a type II thioesterase domain in an unusual non-ribosomal peptide synthetase

Non-Ribosomal Peptide Synthetases (NRPSs) assemble a diverse range of natural products with important applications in both medicine and agriculture. They consist of several multi-enzyme subunits that must interact with each other in a highly controlled manner to facilitate efficient chain transfer, thus ensuring biosynthetic fidelity. Several mechanisms for chain transfer are known for NRPSs, promoting structural diversity. Herein, we report the first biochemically characterized example of a type II thioesterase (TEII) domain capable of catalysing aminoacyl chain transfer between thiolation (T) domains on two separate NRPS subunits responsible for installation of a dehydrobutyrine moiety. Biochemical dissection of this process reveals the central role of the TEII-catalysed chain translocation event and expands the enzymatic scope of TEII domains beyond canonical (amino)acyl chain hydrolysis. The apparent co-evolution of the TEII domain with the NRPS subunits highlights a unique feature of this enzymatic cassette, which will undoubtedly find utility in biosynthetic engineering efforts.
N
onribosomal peptides (NRPs) are a structurally diverse class of secondary metabolites produced by large multimodular enzymes known as nonribosomal peptide synthetases (NRPSs). During biosynthesis, NRPSs utilize a modular thio-template mechanism, typically consisting of multidomain enzymatic subunits that must interact with each other in a tightly controlled manner during chain translocation events to ensure biosynthetic fidelity. The reactions catalysed by catalytic domains within each module fulfil a cycle of peptidyl chain elongation using at least three domains: adenylation (A), condensation (C), and thiolation (T). The A domains selectively activate and load amino acid extender units onto a 4′-phosphopantetheine (Ppant) moiety, which is post-translationally appended to each T domain in the corresponding module, and the C domains typically catalyse chain extension reactions between the growing peptidyl chain and the downstream amino acid. During the biosynthetic process, the growing peptidyl intermediate remains covalently tethered to the T domains via a thioester linkage. Chain release of the nascent peptide chain is achieved by either hydrolysis or intramolecular cyclization to generate either a linear or cyclic peptide, respectively.

While aminoacyl/peptidyl chain transfer events are commonly mediated by C domains, whereby the upstream T-bound intermediate unit is transferred to the downstream T domain through the formation of a peptide bond (Fig. 1a), Nature has evolved several other mechanisms of chain translocation to downstream pathways. For example, ketosynthase (KS)-mediated chain transfer of T-tethered intermediates are often found in NRPS-PKS/fatty acid hybrid pathways. This is exemplified in the biosynthesis of pyrrolomycin C1, where the first KS domain of Pyr25, a multidomain PKS complex, is proposed to catalyse aminoacyltransfer between a Pyr28-terminated chlorinated pyrrole intermediate and the downstream acyl carrier protein (ACP)-bound acyl unit (Fig. 1b). An alternative mechanism is to use a TEII enzyme to hydrolyse the T-tethered substrates which creates a pool of a given specialised amino acid dedicated for an A domain in the downstream NRPS. An example of this phenomenon can be found in the biosynthesis of barbamide. Here, the didomain NRPS subunit, BarA, activates L-Leu, and a T domain-tethered Leu residue is then subjected to three chlorination events at the unreactive methyl group, followed by transamination to generate BarA-tethered α-ketochloroisocaproic acid. Subsequent hydrolysis catalysed by the TEII, BarC, yields the corresponding chlorinated α-ketoacid which will be loaded onto the main NRPS assembly line, BarE (Fig. 1c).

A further example is acyltransferase (AT) domain-mediated transfer, where AT domains catalyse the condensation between a T domain-tethered aminoacyl unit and a free-standing substrate. This particular approach can be found in the biosynthesis of coumarmycin A2 where CouN7 transfers a pyrrole acid unit to ultimately assemble the pyrrole-based product. A rather unusual case of AT domain-catalysed chain translocation was found in the biosynthesis of coronomic acid 3 where an AT-like domain, CmaE, shuttles the substrate between T domains of different NRPS enzymes (Fig. 1d). The didomain NRPS, CmaA, activates isoleucine, followed by aminoacyltransfer from CmaA to a standalone T domain, CmaD, catalysed by CmaE. CmaD delivers the substrate to the subsequent enzyme for the conversion to 3. Such CmaE-mediated aminoacyltransfer is essential for the biosynthesis, as the downstream enzymes cannot accept CmaA-tethered substrates. CmaE-like AT homologues are also found in the biosynthetic pathways of syringomycin E8 and zorbamycin9. Another example involves spontaneous transthiolation between T domains which was observed in pathway reconstitution of the antibiotic natural product, pacidamycin S4 (Fig. 1e). PacP is an A-T-TE3 tridomain NRPS, of which the TE4 domain was believed to be inactive due to lack of the typical catalytic triad of TE domains. Incubation of PacP-tethered substrate with PacH, a standalone T domain, resulted in high transfer efficiency of an aminoacyl group from the T domain of PacP to PacH, indicating that the transthiolation occurs spontaneously.

Legonmycins 5 (Fig. 1f) belong to a group of bacterial pyrrolizidine alkaloids (PAs). Compared to other bacterial PAs, such as pyrrolizineamides, azabicyclene, and bohemamine derivatives (Supplementary Fig. 1a), legonmycins contain an extra methyl group at the C7 position (Fig. 1f). Unlike the biosynthetic pathways of other bacterial PAs, which have canonical NRPS complexes, the pathway for the assembly of legonmycins is rather unusual. It contains two NRPS proteins, LgnB and LgnD, with domain arrangement of A1-T1-C1-T2-C2-A2, T2-TE, respectively. The A1 domain in LgnB and A2 in LgnD were predicted to activate L-Thr and L-Pro, respectively (Supplementary Fig. 1b). It was proposed that the biosynthesis of legonmycins 5 is initiated by the activity of LgnB and LgnD, utilising L-Thr, L-Pro, and acyl-CoAs to generate the final NRPS-bound acylated-dehydrotripteryn (Ddb)-Pro intermediate, which is likely followed by a TE-catalysed ring closure, leading to the key intermediates, legonindolizidines 6 (Supplementary Fig. 1b). Biochemical analyses showed that the monooxygenase, LgnC, acts on 6 and catalyses a multistep reaction to yield 5 (Supplementary Fig. 1b). In contrast to the incorporation of Ser in other PA systems, the usage of a L-Thr residue in the biosynthesis of 5 was believed to contribute the extra methyl group at C7 of 5. Another unique feature is that the minimal biosynthetic gene cluster (BGC) (accession number: KM514925) of 5 contains an essential gene, lgnA, encoding a type II thiosterase (TEII) orthologue, which was proposed to play an important editing role of removing as-yet-known aberrant intermediates. Interestingly, two amide bonds are formed in the proposed biosynthesis of 6, consistent with two C domains in LgnD (Supplementary Fig. 1b), while three T domains were found in LgnB and LgnD19. This raises an interesting question of whether a chain translocation event between T domains is required.

Herein we report an extensive set of biochemical experiments that establish the role played by LgnA, LgnB-T0 and LgnD-T1 domains in the unusual chain translocation employed at the early stage of legonindolizidine NRPS (Fig. 1f). These experiments demonstrate that the TEII orthologue, LgnA, transfers a L-Thr unit from the LgnB-T0 to the LgnD-T1 domain, allowing condensation with isovaleryl (IV)-CoA to yield a T1-bound IV-Thr unit. This is followed by a LgnD-C2 domain-catalysed dehydra- tion to provide the key intermediate, T1-bound IV-Dhb for the downstream NRPS. Importantly, LgnA also efficiently hydrolyses T0-terminated IV-Thr units that form via aberrant activity of the LgnD-C1 domain, thus maintaining biosynthetic fidelity. Taken together, these results demonstrate that the recruitment of a specialised class of “adapter” TEII proteins to biosynthetic clusters has likely enabled evolutionary distinct NRPS subunits to communicate via an aminoacyltransfer reaction, thereby increasing the diversity of natural product structure while mitigating the risk of biosynthetic derailment.

Results

LgNA TEII is essential for production of legonindolizidin A 6. To investigate the role of the LgnB and LgnD NRPS subunits in the production of 6, we first sought to overexpress both lgnB and lgnD genes in E. coli BL21 CodonPlus (DE3)-RP. LgnB was expressed as an N terminal pHis6-recombinant protein while LgnD was a C-terminal pHis6-recombinant protein, and both were produced in a soluble form that was purified to near homogeneity by Ni-NTA chromatography and gave estimated
molecular weights of 65 kDa for LgnB and 207 kDa for LgnD as observed in SDS page analysis (Supplementary Fig. 2).

In the first instance, apo-LgnB and LgnD were converted into their holo forms, using the Ppant transferase, Sfp, in the presence of coenzyme A and MgCl2. This was followed by the addition of substrates, L-Thr, L-Pro, IV-CoA, and ATP to initiate the biosynthetic process. The assay was monitored by HPLC and UHPLC-high resolution-electrospray ionisation-mass spectrometry (HR-MS). However, no desired product was detected (Fig. 2a (i) and Supplementary Fig. 3).

Given the essential role of the lgnA gene during the biosynthesis of 6 from our previous observations19, we then
overexpressed LgnA in *E. coli* as a N-terminal His₆-recombinant protein, followed by purification to near homogeneity with the expected molecular weight (29 kDa) (Supplementary Fig. 4). Subsequently, an assay comprised of LgnA (0.5 µM), LgnB (10 µM) and LgnD (10 µM), together with L-Thr, L-Pro, IV-CoA, and ATP was performed. HPLC analysis clearly demonstrated that 6 was produced, compared to an authentic standard (Fig. 2a (ii) and (iii)), indicating that the presence of LgnA facilitates the production of 6.

In order to investigate T-bound intermediates, we applied a chemical derivatization strategy to offload the biosynthetic intermediates attached to the NRPS subunits. This approach uses an excess of cysteamine to cleave the (amino)acyl-thioesters attached to the Ppant arm of each T domain. This thioesterification reaction, followed by intramolecular rearrangement and disulphide formation with another free cysteamine unit, yields a free amine that is readily detectable by positive mode HR-MS. Based on previous studies, the chemically captured derivatives of possible biosynthetic intermediates for 6 were predicted to be the cysteamine adducts of Thr 7, IV-Thr 10, and IV-Dhb 9, IV-Thr-Pro 10, and IV-Dhb-Pro 11 as shown in Fig. 2b.

To this end, holo LgnB was first incubated with L-Thr and ATP in the presence of cysteamine. HR-MS analysis of the enzymatic mixture produced an ion corresponding to 7, indicating that the LgnB A₄ domain indeed activates L-Thr which is subsequently loaded onto the LgnB-T₀ domain (Supplementary Fig. 5). An assay of holo-LgnB and holo-LgnD together with substrates in the presence of cysteamine was carried out, which afforded the detection of new ions corresponding to 9 and 10, however, derivatives, 9 and 11, were not observed (Supplementary Fig. 6). To confirm the identity of the cysteamine adducts, we synthesized IV-Thr-Pro-SNAC 10a (Supplementary Fig. 7–9) as a synthetic mimic of protein-tethered IV-Thr-Pro. Incubation of 10a in the presence of cysteamine allowed the generation of 10, confirming the occurrence of chemical cleavage and rearrangement of the thioesters (Supplementary Fig. 10). When LgnA (0.5 µM) was introduced to the above assay, HR-MS analysis confirmed the presence of ions corresponding to derivatives 7–11, indicating that LgnA facilitates the Dhb formation (Fig. 2c and Supplementary Fig. 11). Removal of L-Pro from this assay led to the accumulation of only 7, 8, and 9 (Supplementary Fig. 12). The presence of 8 and 9 in this assay suggested that the dehydration occurs on the NRPS-tethered IV-Thr intermediate to yield the IV-Dhb intermediate, prior to the condensation with L-Pro to generate LgnD-tethered IV-Dhb-Pro.

The captured IV-Thr-Pro intermediate 10 is likely a dead-end species, arising from the condensation between an IV-Thr unit and NRPS-tethered IV-Thr (presumably T₀-tethered IV-Thr, *vide infra*) that escapes the dehydration. Interestingly, 10 was found to accumulate in both the presence and absence of LgnA. Our time-course experiments showed that a significantly larger amount of 10 was accumulated in the assay when LgnA was omitted over a period of 3 h, compared to when LgnA was present (Supplementary Fig. 13). A possible explanation for this observation is that, even when LgnA is present, the nonselective condensation domain (presumably LgnD-C₂ domain) is still able to condense the NRPS-tethered IV-Thr unit with the downstream L-Pro unit to produce NRPS-tethered IV-Thr-Pro.

LgnA facilitates the Dhb formation catalysed by LgnD-C₃ domain on LgnD-T₁ domain. To understand how LgnA facilitates Dhb formation, it was necessary to investigate the exact location of the dehydration in the NRPS assembly. Our phylogenetic analysis suggested that the LgnD-C₃ domain belongs to a group of unusual C* domains (Supplementary Fig. 14), analogous to the recently characterized AlbB-C₂ and -C₃ domains in the biosynthetic pathway of albopeptide, which catalyse the dehydration on L-Ser and L-Thr to dehydroalanine (Dha) and (E)-Dhb residues, respectively. Similar to AlbB-C₂ and -C₃, LgnD-C₃ possesses a HHxxDG catalytic motif, characteristic of conventional C domains, where the second H is proposed to catalyse the amide formation. This is distinct from the NocB-C₃ domain, in nocardicin biosynthesis, which contains a rather unique H₇₉₀-HHxxDDG motif in its active site where H₇₉₀ was shown to be responsible for the generation of a transient Dha species, while H₇₉₂ catalyses amide bond formation during the course of the β-lactam formation. However, structural information is needed to determine the residue(s) responsible for the dehydration in these C* domains including LgnD-C₂.

We chose to overexpress two truncated proteins; LgnD-C₁-T₁ didomain and LgnD-C₁-T₁-C₂ tridomain in *E. coli*, followed by purification to near homogeneity with the expected molecular weights of 66 kDa and 114 kDa, respectively, as observed in SDS page analysis (Supplementary Fig. 15). Compared to the earlier assay containing LgnA, LgnB and the full-length LgnD where both 8 and 9 were present (Fig. 2d (i)), incubation of LgnB and LgnD-C₁-T₁ in the presence or absence of LgnA confirmed the presence of one ion congruent with the captured intermediate 8 as indicated by our HR-MS analysis (Supplementary Fig. 16), suggesting that LgnA is not required for the formation of the IV-Thr unit. However, the absence of 9 (Fig. 2d (ii)) indicated that the dehydration reaction on the IV-Thr unit had not occurred in this assay. We also mutated the key Ser residue (S492) of LgnD-T₁ in the truncated LgnD-C₁-T₁ didomain to Ala to yield the LgnD-C₁-T₁(S492A) variant, preventing the apo-LgnD-T₁ domain from being converted to the holo form. HR-MS analysis of the assay of LgnA, LgnB and LgnD-C₁-T₁(S492A) showed the occurrence of only 8 (Supplementary Fig. 16), indicating that IV-Thr must be bound to the LgnB-T₀ domain in the case with the LgnD-C₁-T₁(S492A) variant.

Incubation of LgnD-C₂-T₁-C₂ with LgnA and LgnB in the presence of cysteamine allowed the accumulation of both 8 (Supplementary Fig. 17) and 9 as observed in our HR-MS analysis (Fig. 2d (iii)), confirming that the presence of the C₂ domain is a prerequisite for the dehydration reaction. To ascertain which T domain the dehydration reaction is occurring on, we generated
LgnD-C1-T1(S492A)-C2 variant. In an assay using this variant with LgnA, only intermediate 8 (Supplementary Fig. 17) was observed, but not 9 (Fig. 2d (iv)). This indicated that the presence of a holo-LgnD-T1 domain is key for the dehydration reaction on the IV-Thr unit to provide IV-Dhb species. Taken together, our results strongly support that the LgnD-C2 domain catalyses the dehydration reaction of LgnD-T1-tethered IV-Thr to yield an IV-Dhb unit.

Interestingly, when omitting LgnA in an assay of LgnB-A1-T0 and LgnD-C1-T1-C2, only intermediate 8 was observed (Supplementary Fig. 17) but not 9 (Fig. 2d (v)), demonstrating that IV-Thr unit is likely tethered to LgnB-T0 and not the LgnD-T1 domain, as a LgnD-T1-bound IV-Thr unit would undergo a dehydration reaction in the presence of LgnD-C2 domain. This observation raised the question of whether LgnA transfers Thr or

Fig. 2 HPLC and chemical capturing-based LC-MS analyses of the production of 6 and potential biosynthetic intermediates, 7-11, respectively, in the assays of the pathway reconstitution and truncated LgnD with LgnB in the presence or absence of LgnA. a HPLC analysis of the production of 6 in the in vitro reconstitution experiments, (i) the assay of LgnB and LgnD only, (ii) the authentic 6; (iii) the assay of LgnA, LgnB and LgnD. b The structures of the expected cystamine-adducted intermediates, 7-11. c LC-MS traces of chemically captured intermediates during the NRPS assembly towards 6 in the presence of LgnA. d, LC-HR-MS analysis of chemical captured intermediate, 9, in the assays of truncated LgnD constructs, (i), the extracted ion chromatograms (EICs) of 9 in the assay of LgnB with truncated LgnD-C1-T1-C2; (ii), the EICs of 9 in the assay of LgnB and truncated LgnD-C1-T1-C2 in the presence of LgnA; (iii), the EICs of 9 in the assay of LgnB and truncated LgnD-C1-T1 in the presence of LgnA. (iv), the EICs of 9 in the control by incubation of LgnB, LgnD and LgnA. (v), the EIC of 0 in the assay of LgnB, LgnD and the truncated LgnD-C1-T1(S492A)-C2 tridomain. The molar ratio of LgnA to LgnB and LgnD or truncated LgnD was 1:20:20. The individual intermediate derivative and the corresponding LC-MS traces were colour coded. All LC-MS traces were represented in the same scale of ion intensity (10^4).
IV-Thr or indeed both units, from LgnB-T₀ to LgnD-T₁ domain for the dehydration reaction.

LgnA catalyses the chain translocation of T₀-tethered Thr unit to LgnD-T₁ domain. Intact protein mass spectrometry (MS) is an invaluable tool for the interrogation of covalently tethered intermediates of polyketide synthases (PKS) and NRPSs. Taking the holo form of LgnB A₁-T₀ (Fig. 3b (i)) as the starting point, following addition of L-Thr and ATP we observed a new species from the deconvoluted mass spectrum (Fig. 3b (ii)). This new species corresponded to the addition of one L-Thr unit tethered to LgnB-A₁-T₀, as indicated by the mass shift of +101 Da relative to the mass of holo-LgnB A₁-T₀.

Next, we performed an assay using LgnB-A₁-T₀ (100 µM) and LgnD-C₁-T₁ (100 µM) and L-Thr, in the absence and presence of LgnA. Compared to the mass of holo-LgnB-C₁-T₁ (Fig. 3b (iii)), only LgnB-A₁-T₀-tethered L-Thr was observed in the absence of LgnA (Fig. 3b (iv)), with no species corresponding to LgnD-C₁-T₁ tethered L-Thr observed. In contrast, when LgnA (5 µM) was included, a +101 Da mass shift relative to holo-LgnD-C₁-T₁ was clearly observed for a new species corresponding to one unit of L-Thr tethered to LgnD-C₁-T₁ (Fig. 3b (v)), strongly indicating that LgnA catalyses an aminoacyltransfer of L-Thr units from the LgnB-T₀ domain to LgnD-T₁ domain. It should be noted that there is a population of LgnB-A₁-T₀-tethered L-Thr present due to the excess L-Thr added in the enzyme mixture (Fig. 3b (vi)). An attempt to capture the possible acylated LgnA intermediate during this aminoacyltransfer was carried out. To this end, we incubated holo-LgnB-A₁-T₀ (100 µM) and LgnA (100 µM) in the presence of L-Thr and ATP and monitored LgnA by intact MS. Compared to the control of LgnA alone (Fig. 3c (i)), incubation with holo-LgnB-A₁-T₀, L-Thr and ATP allowed almost all of LgnA to become aminoacylated with a single L-Thr unit (+101 Da) (Fig. 3c (ii)). Observation of a stable aminoacyl-enzyme intermediate suggested that LgnA does not hydrolyse L-Thr from the LgnB-T₀ domain, but acts as an aminoacyltransferase to shuttle L-Thr between NRPS subunits. LgnA possesses a canonical catalytic triad (Ser₇⁷-Asp₁₈₄-His₂₁₂) exhibited by classical TEII domains with the active site Ser residue situated in a conserved G₇₅-H₇ₙ-S₇₇-X₇₈-G₇₉ motif. These features have been observed in well-characterised hydrolytic TEII domains, such as RfiR in the polyketide pathway of rifamycin and BorB in the PKS-NRPS pathway of borrelidin (Supplementary Fig. 18). It is worth noting that TEII domains belong to the α,β-hydrolase superfamily, which also includes AT domains from polyketide synthases, where the TEII domain in a Ser-His dyad is acylated by an acyl-CoA species before being transferred to the Ppant thiol of the carrier protein domain. Therefore, it is likely that the equivalent Ser residue (Ser₇⁷) in LgnA is the active site residue that becomes acylated with L-Thr. The Ser₇⁷ residue of LgnA was then mutated to Ala in order to yield the inactive variant, LgnA(S77A). Addition of LgnA(S77A) in the pathway reconstitution assays abolished the production of 6 (Supplementary Fig. 19a) when compared to the authentic standard of 6 and the control reaction, as observed in HPLC and LC-MS analysis (Supplementary Fig. 19a and b). Chemical capturing in these reconstitution experiments also indicated the absence of key intermediate cystamine adducts, 9 and 11, in the reaction with LgnA(S77A), compared to the control (Supplementary Fig. 19c and d). These snapshots of our intact MS analysis together with site directed mutagenesis (SDM) and chemical capturing experiments strongly demonstrated that LgnA transfers an L-Thr residue from LgnB-T₀ domain to LgnD-T₁ domain via an L-Thr aminoacyl LgnA intermediate.

LgnA hydrolyses the aberrant T₀-tethered IV-Thr units. While we had managed to elucidate the aminoacyltransferase role of LgnA, the canonical hydrolytic activity of TEII domains had not been probed. To investigate this further, a pre-incubation of holo-LgnB A₁-T₀ (100 µM), holo-LgnD-C₁-T₁ (100 µM), L-Thr, ATP and LgnA (5 µM) for 30 mins allowed the T₀ and T₁ domains to be fully acylated with a L-Thr unit (Supplementary Fig. 20ii). Addition of IV-CoA into the reaction mixture resulted in observation of an IV-Thr unit attached to LgnD-C₁-T₁ as the stalled intermediate due to the absence of the downstream NRPS machinery (Supplementary Fig. 20iv), consistent with our chemical capturing results that show IV-Thr is correctly formed on LgnD-T₁ domain. Interestingly, in the same assay, IV-Thr is also found tethered to LgnB-A₁-T₀. This is probably due to full occupancy of the LgnD-T₁ domain with IV-Thr units, leaving the nonselective C₁ domain to catalyse the formation of IV-Thr-LgnB-T₀. Again, this is in agreement with our previous chemical capturing results in the assay between LgnB and LgnD-C₁-T₁ didomain (Supplementary Fig. 16). The appearance of holo-LgnB was interesting as LgnB should be fully occupied when excessive substrates were added, suggesting aminoacyl chain hydrolysis (Supplementary Fig. 20iv). This was confirmed upon observation that the hydrolytic product, IV-Thr-COOH, was accumulated in the pathway reconstitution assay with LgnA as indicated in HR-MS and MS² fragmentation analyses, while no 12 was observed in the assay without LgnA (Supplementary Fig. 21). This clearly indicates hydrolytic activity of LgnA towards the IV-Thr unit.

To probe this in more detail, a chemoenzymatic synthesis of IV-Thr-COOH was attempted, which could be used for artificial loading onto T-domains and assessing hydrolytic activity of LgnA. While the construction of IV-O-Bu-Thr-COOH 12a and pantetheine arm (pant) 13 were straightforward (Supplementary Methods and Supplementary Fig. 22). Coupling 12a with 13 proved to be challenging, resulting in unsuccessful isolation of the desired products by either flash or HPLC chromatography. Instead, IV-O-Bu-Thr-SNAC 14 was synthesized (Supplementary Fig. 23), which could be deprotected under acidic conditions to yield IV-Thr-SNAC 14a. Interestingly, two conformational isomers of 14a appeared in our LC-MS analysis (Supplementary Fig. 24), suggesting the presence of a possible equilibrium between cis- and trans-configurations of the amide bond. Incubation of 14a with LgnA or LgnA(S77A), however, showed that no hydrolytic reaction occurred (Supplementary Fig. 24).

Next, we pursued the in situ production of the IV-Thr unit on the LgnB-T₀ domain to check the hydrolytic activity of LgnA. Here, LgnB (100 µM) and LgnD-C₁-T₁ (100 µM) were co-incubated with L-Thr, IV-CoA and ATP to generate an IV-Thr unit tethered to LgnB-T₀ domain (Fig. 4d (iii)) (no IV-Thr formation on LgnD-C₁-T₁ as no LgnA was present to catalyse transfer). In order to prevent further loading/condensation events, the excess L-Thr, IV-CoA and ATP were then removed from the assay by ultrafiltration. The addition of LgnA (5 µM) into the mixture resulted in formation of holo-LgnB-A₁-T₀ (Fig. 4d (iv)), unambiguously confirming the hydrolytic activity of LgnA. The T₀-tethered IV-Thr remained intact when the LgnA(S77A) variant was used in the equivalent assay (Fig. 4d (v)), confirming that the S₇⁷ residue in LgnA is crucial for the hydrolytic reaction. Taken together, the intact MS analyses and SDM strongly supports that LgnA hydrolyses LgnB-T₀-tethered IV-Thr to 12 and also that LgnA is incapable of transferring the IV-Thr unit from LgnB-T₀ to LgnD-T₁ domain.
Interestingly, the concentration of LgnA in the enzyme mixture is crucial for the production of 6. The optimal molar ratio of LgnA to LgnB and LgnD is approximately 1:20, where the production of 6 is the most efficient (Supplementary Fig. 25a) and the least amount of the hydrolytic product, IV-Thr-COOH, 12, was observed (Supplementary Fig. 25b). Increasing the LgnA ratio in the enzyme mixture resulted in decreased production of 6 but the increased accumulation of 12 (Supplementary Fig. 25).

When the ratio of LgnA to LgnB and LgnD-C_{1-T_{1}} was 1:4, no IV-Thr was observed attached to either LgnB A_{1-T_{0}} or LgnD-C_{1-T_{1}} in our intact MS analysis, with only a Thr unit observed attached to LgnD-C_{1-T_{1}} (Supplementary Fig. 26), suggesting that
the stalled IV-Thr on LgnD-C1-T1 and the aberrant IV-Thr on LgnB-A1-T0 were hydrolysed by the excessive amount of LgnA. This is consistent with the reconstitution results that the use of a high ratio of LgnA led to decreased production of 6 and increased hydrolytic product, 12, probably due to a loss of peptide extension efficiency. A similar biochemical precedent of the stoichiometry of TEII domains and their NRPS partners was also observed in the glycopeptide biosynthesis where the inclusion of a low ratio (2.5 mol%) of TEII domain and its NRPS partner was the most efficient for the editing roles.31

Examination of the lgB GBC indicated that there are two putative promotor regions, one for LgnA and the other for lgnB, lgnC, and lgnD (Supplementary Fig. 27), suggesting that the expression level of LgnA in the producing strain may be different to the ones of LgnB-D. Further in vitro experiments will be required to warrant whether in vitro results are correlated to what the stoichiometry of LgnA with LgnD and LgnB is required for the efficient metabolite production.

LgnA TEII interacts directly with the LgnD subunit of the NRPS. Our observations that LgnA shuttles L-Thr residues from LgnB to LgnD is intriguing, and likely requires specific protein-protein interactions (PPIs) at each interface to ensure biosynthetic fidelity. To investigate possible PPIs between LgnA, LgnB and LgnD, native PAGE and ITC analyses were conducted. Native-PAGE analysis of a 1:1 solution of LgnA and apo-LgnD resulted in a dominant new band, which migrated further than both LgnA and LgnD alone in the gel (Supplementary Fig. 28a), suggesting that the interaction of LgnA with LgnD caused the change of the overall shape of the complex (and possibly charge exposure). MALDI-TOF MS peptide sequencing analysis confirmed that this band was indeed the protein complex of LgnA and LgnD (Supplementary Fig. 28b).

Titration of LgnA into apo-LgnD during isothermal titration calorimetry (ITC) experiments shed light on the thermodynamic traits of their binding (Table 1 and Supplementary Figure 29), with the curve of the heat trace fitting to the one-site binding model. The dissociation constant $K_d$ was calculated to be $2.550 \pm 0.096 \mu M$, suggesting that LgnA has strong affinity with apo-LgnD in the binary complex. To further investigate the exact location of the PPI, a series of titration of LgnA with truncated LgnD was performed. First titration of LgnA with apo-LgnD-C1-TT showed that LgnA displays an strong affinity with both truncated proteins with the dissociation constant ($K_d$) values of $1.01 \pm 0.227 \mu M$ for apo-LgnD-C1-T1 (Table 1 and Supplementary Figure 29) and $1.700 \pm 0.223 \mu M$ for apo-LgnD-C1-T1-C2 (Table 1 and Supplementary Figure 29). Next overexpression of truncated LgnD-C1-T1 and LgnD-T1 domains in E. coli was carried out (Supplementary Figure 30). Titration of LgnD-T1 domains with LgnA showed strong binding activity with a $K_d$ value of $7.21 \pm 1.61 \mu M$ (Table 1 and Supplementary Figure 31) while the LgnD-C1 domain displays no binding affinity (Table 1 and Supplementary Figure 31), indicating that LgnA must interact with LgnD-T1 domain to shuttle the L-Thr residue.

Titration of LgnA to LgnB in our ITC analysis showed no binding affinity (Table 1 and Supplementary Fig. 32). However, when the same titration was conducted with microscale thermophoresis (MST) analysis, which uses changes in fluorescence to measure the movement of molecules along local temperature gradients instead of measuring heat release32, LgnA was found to indeed have a strong affinity toward apo-LgnB-A1-T0 didomain with a $K_d$ value of $1.699 \pm 0.813 \mu M$ (Table 1 and Supplementary Figure 32), indicating that LgnA also interacts with LgnB for its aminoacyltransfer and hydrolytic activities. Our results suggest that the interaction between LgnA and LgnB is thermodynamically distinct from the other interactions measured by ITC, which might be indicative of higher order conformational changes. No binding affinity was observed when LgnB was titrated with LgnD and truncated LgnD-C1-T1 (Table 1 and Supplementary Fig. 33).

Discussion

The mechanism of aminoacyl chain translocation elucidated here for the legonindolizidine NRPS has several unique features. First, the activated L-Thr must be transferred by the standalone TEII orthologue, LgnA, from the LgnB-T0 domain to the LgnD-T1 domain for the downstream dehydrogenation reaction. Our observations of a highly specific chain translocation event in NRPSs expands the catalytic scope of the TEII enzyme family beyond canonical (amino)acyl chain hydrolysis. Second, the biosynthetic pathway involves an unusual dehydrogenation reaction on L-Thr that must occur at the T1 domain to provide the Dhb residue. Third, LgnB is an A1-T0 didomain, separated from the core mega-synthetase, LgnD. The consequence of this is that the TEII orthologue, LgnA, has been recruited to serve as an adapter to bridge LgnB-T0 and LgnD-T1 domains in order to maintain biosynthetic fidelity.

Standalone TEII domains are a common feature in many of PKS, NRPS and PKS-NRPS hybrid pathways.33 They play an important role in maintaining the efficiency of the assembly line and enhancing product titres.33 It has been well-established that TEII domains either remove nonreactive (amino)acyl residues or aberrant intermediates to restore the performance of NRPSs and PKSs, or control the starter units and participate substrate selection.33 In a few cases, TEII domains release key intermediates or the final products.35 Our data show that the standalone TEII domain, LgnA, catalyses an unprecedented transaminoacylation reaction that shuttles LgnB-T0-tethered L-Thr to the LgnD-T1 domain of the downstream megasynthetase (Fig. 5a). This is necessary, because the nonselective condensation reaction between IV-CoA species and a L-Thr unit can occur at the LgnB-T0 domain. LgnB hydrolyses the aberrant intermediate, LgnB-T0 tethered IV-Thr to liberate the “blocked” T0 domain (Fig. 5a). Despite the aminoacyltransfer and editing roles played by LgnA, it is possible that the mis-programmed NRPS-tethered IV-Thr (presumably T2-tethered IV-Thr-Pro), stemming from the nonselective condensation domains, may still accumulate (Fig. 5a), thus blocking the assembly line, as observed in our time-course experiments (Supplementary Fig. 13).
Fig. 4 LgnA hydrolys the aberrant intermediate, LgnB-T₀-tethered IV-l-Thr. a A schematic diagram of the hydrolysis of aberrant intermediate catalysed by LgnA. b Deconvoluted intact protein mass spectra of (i) holo-LgnB A₁-T₀, (ii) holo-LgnD-C₁-T₁, (iii) an assay of LgnB-A₁-holo-T₀ + holo-LgnD-C₁-T₁ + ATP + L-Thr + Mg²⁺ + IV-CoA, (iv) an assay of LgnB-A₁-holo-T₀ + holo-LgnD-C₁-T₁ + ATP + L-Thr + Mg²⁺ + IV-CoA followed by ultracentrifugation filtration to remove unreacted small molecules and subsequent addition of LgnA-TEII, (v) an assay of LgnB-A₁-holo-T₀ + holo-LgnD-C₁-T₁ + ATP + L-Thr + Mg²⁺ + IV-CoA + LgnAS77A-TEII variant. An Peaks labelled with grey dots and asterisks in the intact MS spectra indicate N-terminal acetylation or gluconoylation, respectively, both of which are known post-translational modifications of recombinant heterologous proteins in E. coli²⁴. The exact measured and observed masses for each species are detailed in Supplementary Table 3.
The TEII-mediated transaminoacylation may not be uncommon. For example, WS9326A, a specialised cyclodepsipeptide isolated from various Streptomyces, exhibits several interesting chemical features, such as an (E)-2,3-dehydrotyrosine and a (Z)-pentenylcinnamoyl moiety. Analysis of the corresponding BGC indicated the presence of two standalone upstream modules (WS22 and WS23) consisting of an A-T didomain, two TEII domains (WS5 and WS20, respectively), and the adjacent downstream module 7 with an arrangement of C-T-C-A-T, similar to LgnD. The two A domains in WS22 and WS23 were predicted to activate L-allo-Thr and L-Asn, respectively, key chemical moieties of WS9326A. A series of gene mutation experiments have been performed to investigate the biosynthetic logic of this NRPS assembly. Interpretation of the structures of the accumulated products in these variants provided conclusions that two TEII domains, WS5 and WS20, are highly likely to act as shuttling enzymes to translocate the activated L-allo-Thr and L-Asn from WS22 and WS23, respectively, to the downstream iterative module 7, which adds these two amino acids to the growing peptide chain. MST analysis revealed considerably high binding affinities between WS5-apo-WS23, WS20-apo-WS23 and WS20-apo-WS23 while relatively low binding affinities between WS5-apo-WS23. Crystallographic studies of WS5 and structural modelling suggested possible PPIs between WS5 and the T domains of WS22/23. Our ITC and MST analysis indicated that LgnA displays a strong binding affinity toward the T1 domain of LgnD and LgnB, respectively. More recently, another TEII orthologue, RthD, was proposed to transfer two Thr units into the growing peptidyl chain during the biosynthesis of rotihbins, specialised metabolites with potent herbicidal activity, identified from the phytopathogen Streptomyces scabies that is associated with common scab disease. Interestingly, although LgnA displays a very low sequence homologue (<30% AA identity) with WS5, WS20 and RthD, our phylogenetic analysis suggested that WS20 and RthD belong to the same clade of LgnA, while WS5 is grouped into TEII domains that have functions of intermediate releasing or editing. However, further biochemical characterization is needed to confirm the exact roles played by these TEII systems in the corresponding NRPS systems.

LgnA was also found to be grouped into a separate clade with other putative LgnA-like TEII domains from the classical standalone TEII enzymes, type I TE domains, aminoacyltransferases, suggesting that LgnA is likely to be diverged from typical TEII domains. With this knowledge in hand, we used phylogeny-based, mechanism-guided conserved genomic analysis to identify BGCs containing LgnA-like open reading frames (ORFs), and therefore biosynthetic pathways likely involving this TEII-mediated chain transfer. This led to the identification of multiple BGCs encoding multidomain NRPSs and standalone LgnB-like NRPSs (Supplementary Fig. 34). Representatives are distributed in Gram-positive and Gram-negative bacteria, including biologically important plant pathogens: Pseudomonas fluorescens, Xanthomonas arboricola, nitrogen fixing bacteria such as Cabal tolerantia choica, Hahella sp. KA22 and Rheinheimera sp.; the well-studied model strains, S. albus and S. avermitilis. Common features in these BGCs are that there are extra C-T didomains located in the middle of one of multidomain NRPSs. In a few cases, the C domains in these C-T didomains belong to the group of C* domains which are predicted to catalyse dehydration reactions (Supplementary Fig. 14). It is likely that LgnA-like aminoacyltransferases transfer amino acid residues activated by Lnb-like NRPSs into the extra T domain of the downstream pathway to incorporate modified amino acid residues (i.e. Dhb units) in the growing peptide chains (Supplementary Fig. 34). Taken together, the recruitment of biosynthetic TEII-like transaminoacylation enzymes to biosynthetic clusters may enable the evolution of NRPS pathways by inserting specialised amino acid residues into the growing peptide chains, thereby increasing the diversity of natural product structure.

In the broader context, there have been considerable efforts to use bioengineering approaches to directly modify NRPS machineries in vivo to generate optimized natural products. However, most attempts to achieve this have yielded impaired or non-functional biosynthetic machineries. More recently, a new NRPS domain swapping biotechnology has emerged that uses defined exchange units fused at specific positions that connect two domains and respect the original specificity of the downstream module to enable the production of new peptides. In

| Domain interaction | $K_d$ value | Domain interaction | $K_d$ value |
|--------------------|------------|--------------------|------------|
| LgnA TEII $\Rightarrow$ LgnD | 2.55 ± 0.964 μM | LgnA TEII $\Rightarrow$ LgnD-C$_1$ | 0 |
| LgnA $\Rightarrow$ LgnD-C$_1$,C$_2$ | 1.01 ± 0.227 μM | LgnA $\Rightarrow$ LgnB | 1.699 ± 0.813 μM# |
| LgnA TEII $\Rightarrow$ LgnD-C$_1$,T$_1$ | 1.70 ± 0.223 μM | LgnB $\Rightarrow$ A$_1$,C$_1$,C$_2$,A$_3$ | 0 |
| LgnA TEII $\Rightarrow$ LgnD-T$_1$ | 7.21 ± 1.61 μM | LgnD $\Rightarrow$ C$_1$,C$_2$,A$_3$,TE | 0 |
Fig. 5 Overview of the essential role of the bifunctional TEii, LgnA, in the legonindolizidine A6 NRPS assembly line. a LgnA catalyses L-Thr unit transaminoacylation reaction between LgnB-T0 and LgnD-T1 domain. The correct intermediate, LgnD-T1-bound L-Thr, undergoes the first condensation with IV-CoA to yield LgnD-T1-bound IV-Thr, followed by the C2-catalysed dehydration on L-Thr residue to provide LgnD-T1-tethered IV-Dhb for the downstream pathway. Due to the nonselective condensation reaction, LgnB-T0-bound L-Thr reacts with IV-CoA to generate the aberrant intermediate, LgnB-T0-bound IV-Thr. LgnA plays the editing role of removing LgnD-T0-bound IV-Thr to regenerate the holo-LgnB. However, the nonselective condensation domain could catalyse the reaction between LgnB-T0-tethered IV-Thr and the downstream L-Pro unit to generate the “dead-end” intermediate protein-tethered IV-Thr-Pro. b Phylogenetic analysis indicates that LgnA and LgnA-like TEii CDSs form a distinct clade from typical TEii enzymes, TEi domains, and known aminoacyltransferases in NRPSs. A maximum likelihood tree containing bootstrap values was generated using MEGA 7.0.26. The sequence information for thioesterases and aminoacyltransferases used in this analysis are listed in Supplementary Table 4.
this regard, our finding of LgnA together with its NRPS partner domains may offer a Dhb-forming cassette to generate new Dhb-containing molecules as functionalized peptides, where the Dhb moiety could also provide a bioorthogonal handle for further diversification for peptide-based compound library generation. Further molecular dissection will be required to shed insight on the logical role to contribute to natural product structural diversity.

In conclusion, our results provide important insights into a new mechanism of chain translocation from the legonmycin NRPS. We demonstrate that the TEII orthologue, LgnA, is recruited to catalyse a unique transaminase reaction among known NRPS domains. The occurrence of LgnA-like open reading frames may offer a Dhb-forming cassette to generate new Dhb-containing molecules. We have demonstrated that the TE II orthologue, LgnA, is involved in the formation of the legonmycin scai moiety and can be recruited to generate novel Dhb-forming cascades. This provides a new strategy for the development of natural product scaffolds with potential therapeutic applications.

### Methods

#### General chemicals, reagents, and analytical methods

All starting materials and reagents were bought from commercial sources and used as received. All biochemical reactions apart from where noted were carried out in triplicate. For every set of measurements, triplicate control reactions were performed to ensure that the assay were functioning correctly. All chemical reactions apart from where noted were conducted in 95% DCM. This gave the desired compound as a clear oil in a 4% yield.

#### Synthetic Chemistry

**Synthesis of S-(2-acetamidoethyl) (S)-1-((t-buty1-N-(3-methylbutanoyl)-L-threonyl) pyrroline-2-carbothioate 10b**

To an SPPS tube was added H-Pro-2-CITr resin (0.428 g, 0.30 mmol, 0.7 mole/g loading) and DMP (5 mL) this was agitated for 1 h to swell the resin. The resin was drained and a solution of Fmoc-Thr(Bu)-OH (0.298 g, 0.75 mmol), PyBOP (0.390 g, 0.75 mmol) and DIPEA (0.194 g, 1.50 mmol) in DMF (5 mL) was added and agitated for 2 h. The resin was drained and washed with DMP (3 × 5 mL). The coupling reaction was then repeated for another 2 h before the resin was once again drained and washed with DMP (3 × 5 mL). The resin was added to a solution of 20% piperidine in DMP (5 mL), and the resin agitated for 30 min. The resin was then drained and washed with DMP (3 × 5 mL) before TFA (0.303 g, 3.00 mmol) in DCM (5 mL) was added and buffered with t-butyl alcohol (0.50 mL) and trifluoroacetic anhydride (0.50 mL). The resin was then washed with a solution of 20% triethylsilane in DCM (2 mL) for 1 h. The solution was then concentrated under reduced pressure to give a yellow oil. The resin was drained and washed with DCM (5 mL), and the reaction mixture stirred for 24 h at rt. After this time the reaction mixture was concentrated under reduced pressure and the recovered residue was purified by flash column chromatography (100% DCM to 5% MeOH 95% DCM). This gave the desired compound as a clear oil (0.033 g) in a 4% yield.

**1H NMR (400 MHz, CDCl3) δ 6.61 (d, J = 8.3 Hz, 1H), 6.33 (d, J = 5.4 Hz, 1H), 4.77 (d, J = 8.4, 3.4 Hz, 1H), 4.72–4.65 (m, 1H), 4.03–3.73 (m, 4H), 3.67–3.54 (m, 4H), 3.36 (d, J = 13.9, 8.3, 3.3 Hz, 1H), 3.26–3.14 (m, 1H), 3.07 (t, J = 6.5 Hz, 1H), 2.93–2.94 (m, 1H), 2.87 (dd, J = 13.9, 6.2, 3.1 Hz, 1H), 2.00 (s, 3H), 1.24 (s, 9H), 1.11 (d, J = 6.5 Hz, 3H), 1.00–0.95 (m, 1H).**

### General microbiology and cloning methods

The primers used in this study are listed in Supplementary Table 1. The genes, lgnA, lgnB, and lgnD, as well as truncated [gnd] fragments were amplified from cosmids 7G14 using KOD hot start DNA polymerase (Novagen). The resulting PCR products were inserted into the EcoRI/HindIII site of pET-28a (+) in the cases of LgnA, LgnB, and truncated proteins, LgnD-C1-T1 and LgnD-C1-T1-C2 or into the NdeI/HindIII site of pET-28a (+) in the cases of LgnD-C1 and LgnD-T1 or into the NdeI/XhoI site of pEHISEVe in case of LgnD using In-Fusion® HD CLoning Kit (Takara Bio, Inc) or DNA Ligase (Thermo Fisher Scientific). DNA ligations were carried out using a QIAprep Spin Miniprep Kit (Qiagen). Restriction enzymes (NEB) were used according to the instructions provided by the manufacturers. Site directed mutagenesis of LgnA, LgnD-C1-T1 and LgnD-C1-T1-C2 was achieved using In-Fusion® HD Multiple-Insert Cloning. For LgsN(S77A) variant, two DNA fragments containing the desired mutations were amplified using pET-28a-Igna using a primer pair Lgna_28afor and Lgna_S77toArev, and a primer pair Lgna_S77toAfor and Lgna_28arev, respectively (see Supplementary Table 1 for primer sequences), and cloned into the EcoRI/HindIII site of pET-28a (+) vector using In-Fusion® HD CLoning Kit. For LgncD-C1-T1(S492A) or LgncD-C1- T1(S492A) variant, two PCR reactions were performed to create LgncD for and LgncD_T1_50Aorev, and a primer pair of LgncD_T1_50Afor and LgncD_T1_50Arev were used. The desired mutation in each variant was verified by DNA sequencing.

The genes were amplified using the PCR primers detailed in Supplementary Table 1.

### General methods of protein expression and purification

The above corresponding constructs were individually transformed into E. coli BL21-CodonPlus (DE3)-RP. Single colonies from each transformation were grown overnight in LB media (5 mL) containing kanamycin (50 µg/mL) and chloramphenicol (25 µg/mL). The overnight cultures were transferred to fresh LB medium (500 µL) supplemented with kanamycin (50 µg/mL) and chloramphenicol (25 µg/mL) and cultivated at 37 °C until the cell density reached an OD600 of 0.6. IPTG was added to a final concentration of 0.1 mM to induce protein expression. Cells were grown for 16–20 h at 16 °C and then harvested by centrifugation at 4 °C. The cell pellets were re-suspended in ice-cold lytic buffer (20 mM Tris-HCl 300 mM NaCl, 10 mM imidazole, pH 8.0) and further disrupted by Ultrasonic Homogenizer (Y92-II). The supernatant of cell debris was loaded onto Ni-NTA-affinity column. Bound proteins were eluted with the same Tris-HCl buffer containing different concentrations of imidazole. Bound proteins were collected using a QIAprep Spin Miniprep Kit (Qiagen). For isotopic titration calorimetry (TCC) measurements of these protein, large scale fermentation (5 L) was conducted and protein solutions from Ni affinity chromatography were further purified by gel filtration chromatography on a Superdex 75 column at 12 °C equilibrated with 20 mM Tris-HCl (pH 8.0) containing 100 mM NaCl. The proteins were stored at −80 °C in storage buffer (100 mM Tris-HCl, 0.15 M NaCl, 10% (v/v) glycerol, 1 mM DTT). The protein concentrations were determined on NanoDrop 2000 by using the corresponding extinction coefficient.

### In vitro reconstitution of legonindolizidine

A sample of LgbN (10 µM) and LgdN (10 µM) were incubated with Sfp (2 µM), coenzyme A (100 µM) and MgCl2 (10 mM) in HEPES-Na buffer (50 mM, pH 7.5) at 28 °C for 1 h. To the mixture were added LgbN (10 µM) or 0.5 µM i-isovaleryl-CoA, L-Thr-L-Pro (1 mM each) and ATP (3.5 mM) to a final volume of 50 µL. The reaction was incubated at 28 °C for 3 h and then quenched by addition of 100 µL of acetonitrile. The mixture was centrifuged at 13,000 rpm for 10 min to remove protein precipitates. The supernatant was then analyzed by Bruker MAXII D II QTOF in tandem with an Agilent 1290 Infinity IIUPHC. Samples were separated on a Phenomenex Kinetex XB-C18 (2.6 µm, 100 × 2.1 mm) column with a mobile phase of 5% ACN + 0.1% formic acid to 100% ACN + 0.1% formic acid in 10 min.

### Biochemical determination of NRPS-bound intermediates

The one-pot in vitro legonindolizidine reconstitution reactions using LgbN (10 µM) and LgdN (truncated LgdN proteins) (10 µM) with and without the addition of LgbN (0.5 µM) were conducted to capture the bioorthogonal intermediates bound to LgbN and LgdN via chemical probe cysteamine. Freshly prepared cysteine hydrochloride (2.5 µL, 1 M) was added to each reaction (50 µL) as the final component after adding all
Biochemical assay of LgnA hydrolytic activity. The one-pot in vitro lego-nindolizidine reconstitution reactions using LgNB and LgND with different concentrations of LgnA used in the one-pot transformation assay. The reaction mixtures were incubated at 28 °C for 3 h and then quenched with 100 µL of acetonitrile. Protein precipitates were removed by centrifugation. The supernatants were subjected to LC-MS analysis under the condition as described above.

Microscale Thermophoresis (MST) measurement. To test the binding affinity between LgnA and LgNB, LgnA was utilized as the ligand for the target protein LgNB. The purified LgNB was labelled by RED-tris-NTA and diluted to a final concentration of 25 nM with MST optimized buffer (HBS-ET: 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% Tween-20, PH 8.0). LgnA (20 µM) was diluted 1:1 in HBS-ET buffer (10 µL) to generate a 16-ligand serial dilution from 20 to 0.00061 µM. Labelled LgNB (10 µL) was then added to each above LgnA solution (10 µL). After 10 min incubation at room temperature, 16 capillaries were filled with each of samples of which ligand concentrations range from 10 to 0.000305 µM, and were subjected to the MST measurement to yield the respective data points. MST experiments were performed on a Monolith NT.115 instrument (NanoTemper Technologies, Germany) which was set to 80% LED power and 40% MST power at 25 °C. Data analyses were performed using MO. Affinity Analysis V2.3.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The datasets generated and analysed during the current study are available from the corresponding author upon request. Amino acid sequence GenBank accession numbers used in this study are reported in Supplementary Table 4 and Supplementary Fig. 14.

Received: 30 April 2021; Accepted: 18 November 2021;
Published online: 10 January 2022

References

1. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis—Principles and Prospects. Angew. Chem. - Int. Ed. 56, 3770–3821 (2017).
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. Reisinger, T. A. et al. Vibriobactin biosynthesis in Vibrio cholerae: VbbH is an amide synthase homologous to nonribosomal peptide synthetase condensation domains. Biochemistry 19, 15513–15521 (2000).
4. Flatt, P. M. et al. Characterization of the initial enzymatic steps of barbamide biosynthesis. J. Nat. Prod. 69, 938–944 (2006).
5. Galoncic, D. P. et al. Halogenation of unactivated carbon centers in natural product biosynthesis: trichlorination of leucine during barbamide biosynthesis. J. Am. Chem. Soc. 128, 3900–3901 (2006).
6. Garneau-Tsodikova, S. et al. Installation of the pyrrolid-2-carboxyl pharmacophore by CouN1 and CouN7 in the late biosynthetic steps of the aminocoumarin antibiotics chlorobiozinc and coumermycin A1. Biochemistry 45, 8568–8578 (2006).
7. Vaillancourt, F. H. et al. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. Nature 436, 1191–1194 (2005).
8. Singh, G. M. et al. Characterization of SyrC, an aminocyclotransferase shuffling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line. Chem. Biol. 14, 31–40 (2007).
9. Galoncic, D. P. et al. Biosynthetic gene cluster of zorbarcin, a member of the bleomycin family of antitumor antibiotics, from Streptomyces flavoviridis ATCC 21892. Mol. Biogest. 5, 77–90 (2009).
10. Zhang, W. et al. Nine enzymes are required for assembly of the pacidyacin group of peptidyl nucleoside antibiotics. J. Am. Chem. Soc. 133, 5240–5243 (2011).
11. Robertson, J. & Stevens, K. Pyrrolizidine alkaloids: Occurrence, biology, and chemical synthesis. Nat. Prod. Rep. 34, 62–89 (2017).
12. Schimming, O. et al. Structure, biosynthesis, and occurrence of bacterial pyrrolizidine alkaloids. Angew. Chem. - Int. Ed. 54, 12702–12705 (2015).
13. Hong, Z. et al. Azetidine-Containing Alkaloids Produced by a Quorum-Sensing Regulated Nonribosomal Peptide Synthetase Pathway in Pseudomonas aeruginosa. Angew. Chem. - Int. Ed. 131, 3210–3214 (2019).
14. Patteson, J. B. et al. Discovery and Biosynthesis of Azabicyclene, a Conserved Nonribosomal Peptide in Pseudomonas aeruginosa. Org. Lett. 21, 4955–4959 (2019).
15. Liu, L. et al. Activation and Characterization of Bohemamine Biosynthetic Gene Cluster from Streptomyces sp. CB02009. *Org. Lett.* 22, 4614–4619 (2020).
16. Fu, P. & Macmillan, J. B. Spithioneines A and B, Two New Bohemamine Derivatives Possessing Ergothioneine Moiety from a Marine-Derived *Streptomyces spinoverrucosus*. *Org. Lett.* 17, 3046–3049 (2015).
17. Fu, P. et al. 1,3-Oxazin-6-one Derivatives and Bohemamine-Type Pyrrolizidine Alkaloids from a Marine-Derived *Streptomyces spinoverrucosus*. *J. Nat. Prod.* 80, 2177–2184 (2017).
18. Kim, M. C. et al. Salainazinones A and B: Pyrrolidinyl-Oxazinones from Solar Saltern-Derived *Streptomyces sp.* KM-F004. *Org. Lett.* 17, 5024–5027 (2015).
19. Huang, S. et al. Discovery of a single monoxygenase that catalyzes carbamate formation and ring contraction in the biosynthesis of the legonmycins. *Angew. Chem. - Int. Ed.*** 54, 12697–12701 (2015).
20. Belecchi, K. & Townsend, C. A. Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the endyne pyrrole synthase CallE. *J. Am. Chem. Soc.* 135, 14339–14348 (2013).
21. Wang, S. et al. Discovery and biosynthetic investigation of a new antibacterial dehydrated non-ribosomal tripeptide. *Angew. Chem. - Int. Ed.*** 60, 3229–3237 (2021).
22. Gaudelli, N. M. et al. β-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature*** 520, 383–387 (2015).
23. Long, D. H. & Townsend, C. A. Mechanism of Integrated β-Lactam Formation by a Nonribosomal Peptide Synthetase during Antibiotic Synthesis. *Biochemistry*** 57, 3353–3358 (2018).
24. Arndt, M., Vögle, M. & Kienast, M. Unusual *Borrelia gladioli* double chain-initiating nonribosomal peptide synthetase assembles ‘fungal’ icosaside antibiotics. *Chem. Sci.* 10, 5489–5494 (2019).
25. Hai, Y. et al. Fungal siderophore biosynthesis catalysed by an iterative nonribosomal peptide synthetase. *Chem. Sci.* 11, 11525–11530 (2020).
26. Jenner, M. et al. Mechanism of interunivert ketosynthase-dehydratase interaction in polyketide synthases. *Nat. Chem. Biol.* 14, 455–462 (2018).
27. Claxton, H. B. et al. Structure and functional analysis of RIRβ, the type II thioesterase from the rifamycin biosynthetic pathway. *J. Biol. Chem.* 284, 5021–5029 (2009).
28. Curran, S. C. et al. Structure and Function of BorR, the Type II Thioesterase from the Borrellidin Biosynthetic Gene Cluster. *Biochemistry*** 59, 1630–1639 (2020).
29. Kinate, K. & Townend, C. A. The structures of type I polyketide synthases. *Nat. Prod. Rep.* 29, 1050–1073 (2012).
30. Agarwal, V. et al. Chemoenzymatic Synthesis of Acyl Coenzyme A Substrates Enables in Situ Labeling of Small Molecules and Proteins. *Org. Lett.* 17, 4452–4455 (2015).
31. Kanusiate, M. et al. A proof-reading mechanism for non-proteinogenic amino acid incorporation into glycopeptide antibiotics. *Chem. Sci.* 10, 9466–9482 (2019).
32. Jerabek-Willemsen, M. et al. MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* 1077, 101–113 (2014).
33. Kotowska, M. & Pawlik, K. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* 98, 7735–7746 (2014).
34. Liu, T. et al. Mechanism of Thioesterase-Catalyzed Chain Release in the Biosynthesis of the Polyether Antibiotic Nanchangmycin. *Chem. Biol.* 15, 449–458 (2008).
35. Hayashi, K. et al. WS9326A, a novel tachykinin antagonist isolated from *Streptomyces violaceoniger* no. 9326. 1. Taxonomy, fermentation, isolation, physico-chemical properties and biological activities. *J. Antibiot.* 45, 1055–1063 (1992).
36. Yu, Z. et al. New WS9326A congeners from Streptomyces sp. 9078 inhibiting Brugia malayi asparagine-tRNA synthetase. *Org. Lett.* 14, 4946–4949 (2012).
37. Hang, S. et al. New WS9326A Derivatives and One New Annimycin Derivative with Antimalarial Activity are Produced by *Streptomyces asterosporus* DSM 41452 and Its Mutant. *Chembiochem* 19, 272–279 (2018).
38. Bernhardt, M. et al. Role of Two Exceptional trans Adenylation Domains and MBH-like Proteins in the Biosynthesis of the Nonribosomal Peptide WS9324A from *Streptomyces calvus* ATCC 13382. *Chembiochem* 21, 2659–2666 (2020).
39. Kim, M. S. et al. Unprecedented Noncanonical Features of the Nonlinear Nonribosomal Peptide Synthetase Assembly Line for WS9326A Biosynthesis. *Angew. Chem. Int Ed. Engl.* 60, 19766–19773 (2021).
40. Planckaert S. et al. Identification of Novel Rothibin Analogues in *Streptomyces salubris*, Including Discovery of Its Biosynthetic Gene Cluster. *Microbiol. Spectr.* e005712 (2021).

41. Winn, M. et al. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* 33, 317–347 (2016).
42. Brown, A. S. et al. Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* 35, 1210–1228 (2018).
43. Bozhuyık, K. A. J. et al. De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* 10, 275–281 (2018).
44. Bozhuyık, 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).
45. Bogart, J. W. & Bowers, A. A. Dehydroamino acids: Chemical multi-tools for late-stage diversiﬁcation. *Org. Biomol. Chem.* 17, 3653–3669 (2019).

**Acknowledgements**

This work was supported by the Biotechnology and Biological Sciences Research Council UK (S.W. and H.D., BB/P003808X/1 and BB/R00479X/1). S.W. and H.D. are grateful for SFC Covid-19 Extension allocation and bridging Fund and UKRI Covid-19 Allocation Fund. Q.Z. and Y.Y. are grateful for financial support from the National Key Research and Development Program of China (2018YFA0900400), the National Natural Science Foundation of China (31570033, 3181153029, and 31870035 to Y.Y.). H.D. and Y.Y. are recipients of a Royal Society-NSFC Newton Mobility Grant Award (IEC/NSFC/170617). M.J. is the recipient of a BBSRC Discovery Fellowship (BB/R012121/1). The Bruker MaXis II instrument used in this study was funded by the BBSRC (BB/M017982/1). K.K. and H.D. thank Leverhulme Trust Royal Society Africa award (AA900088) and the jointly funded UK Medical Research Council-UK Department for International Development (MRC/DFID) Cordaid agreement African Research Leaders Award (MR/S00520X/1).

**Author contributions**

S.W., W.B., S.L.C., M.J., K.K., Y.Y. and H.D. designed the experiments. S.W., Z.L., M.H.T., W.B., Q.Z., K. W. and M.J. performed chemical synthesis, biochemical experiments, small molecule/intact protein mass spectrometry, ITC and bioinformatics analyses. H.D. wrote the paper with S.L.C., K.K., Y.Y., M.J. S.W., W.B. and Q.Z.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-27512-0.

Correspondence and requests for materials should be addressed to Matthew Jenner, Yi Yu, Steven L. Cobb or Hai Deng.

**Peer Review Information** *Nature Communications* thanks Max Creyle and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at [http://www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2022