



# Gene editing enables rapid engineering of complex antibiotic assembly lines

Wei Li Thong<sup>1,2</sup>, Yingxin Zhang<sup>1,2</sup>, Ying Zhuo<sup>1,2</sup>, Katherine J. Robins <sup>1</sup>, Joanna K. Fyans<sup>1</sup>, Abigail J. Herbert<sup>1</sup>, Brian J. C. Law<sup>1</sup> & Jason Micklefield <sup>1</sup>✉

Re-engineering biosynthetic assembly lines, including nonribosomal peptide synthetases (NRPS) and related megasynthase enzymes, is a powerful route to new antibiotics and other bioactive natural products that are too complex for chemical synthesis. However, engineering megasynthases is very challenging using current methods. Here, we describe how CRISPR-Cas9 gene editing can be exploited to rapidly engineer one of the most complex megasynthase assembly lines in nature, the 2.0 MDa NRPS enzymes that deliver the lipopeptide antibiotic enduracidin. Gene editing was used to exchange subdomains within the NRPS, altering substrate selectivity, leading to ten new lipopeptide variants in good yields. In contrast, attempts to engineer the same NRPS using a conventional homologous recombination-mediated gene knockout and complementation approach resulted in only traces of new enduracidin variants. In addition to exchanging subdomains within the enduracidin NRPS, subdomains from a range of NRPS enzymes of diverse bacterial origins were also successfully utilized.

<sup>1</sup>Department of Chemistry and Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK. <sup>2</sup>These authors contributed equally: Wei Li Thong, Yingxin Zhang, Ying Zhuo. ✉email: [jason.micklefield@manchester.ac.uk](mailto:jason.micklefield@manchester.ac.uk)



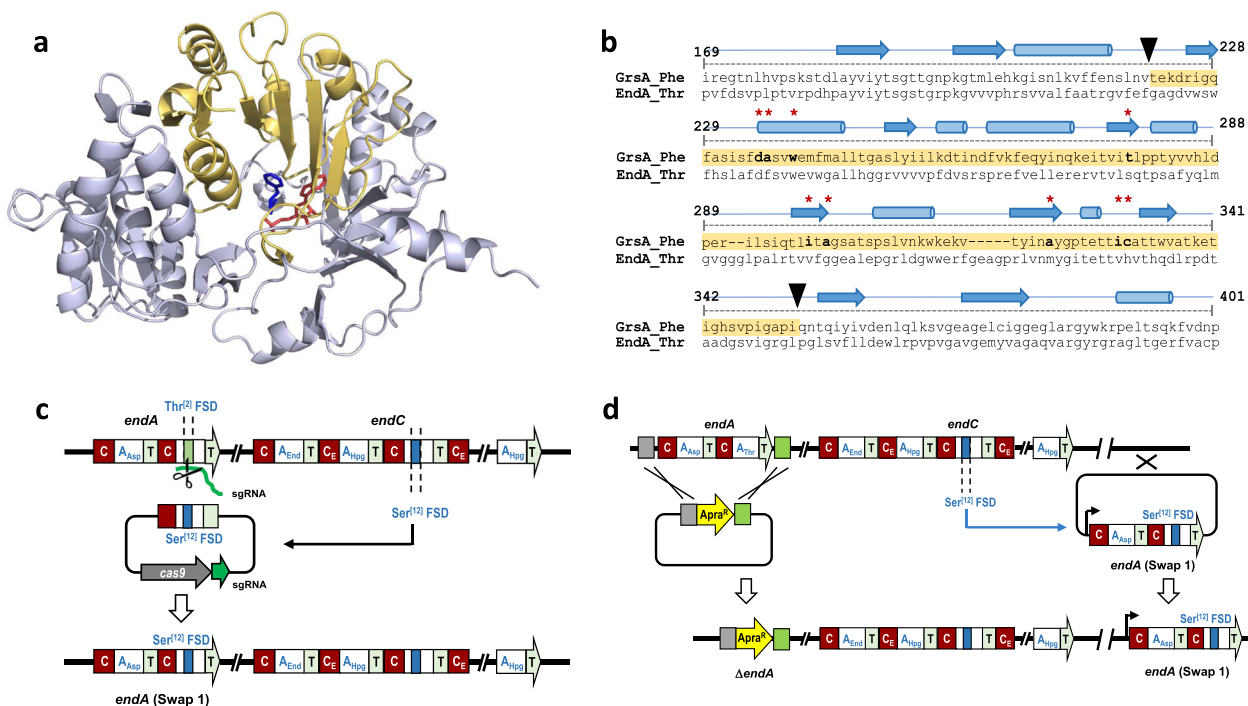
efficiently replacing FSDs in vivo at their native loci, without impacting production titres.

## Results

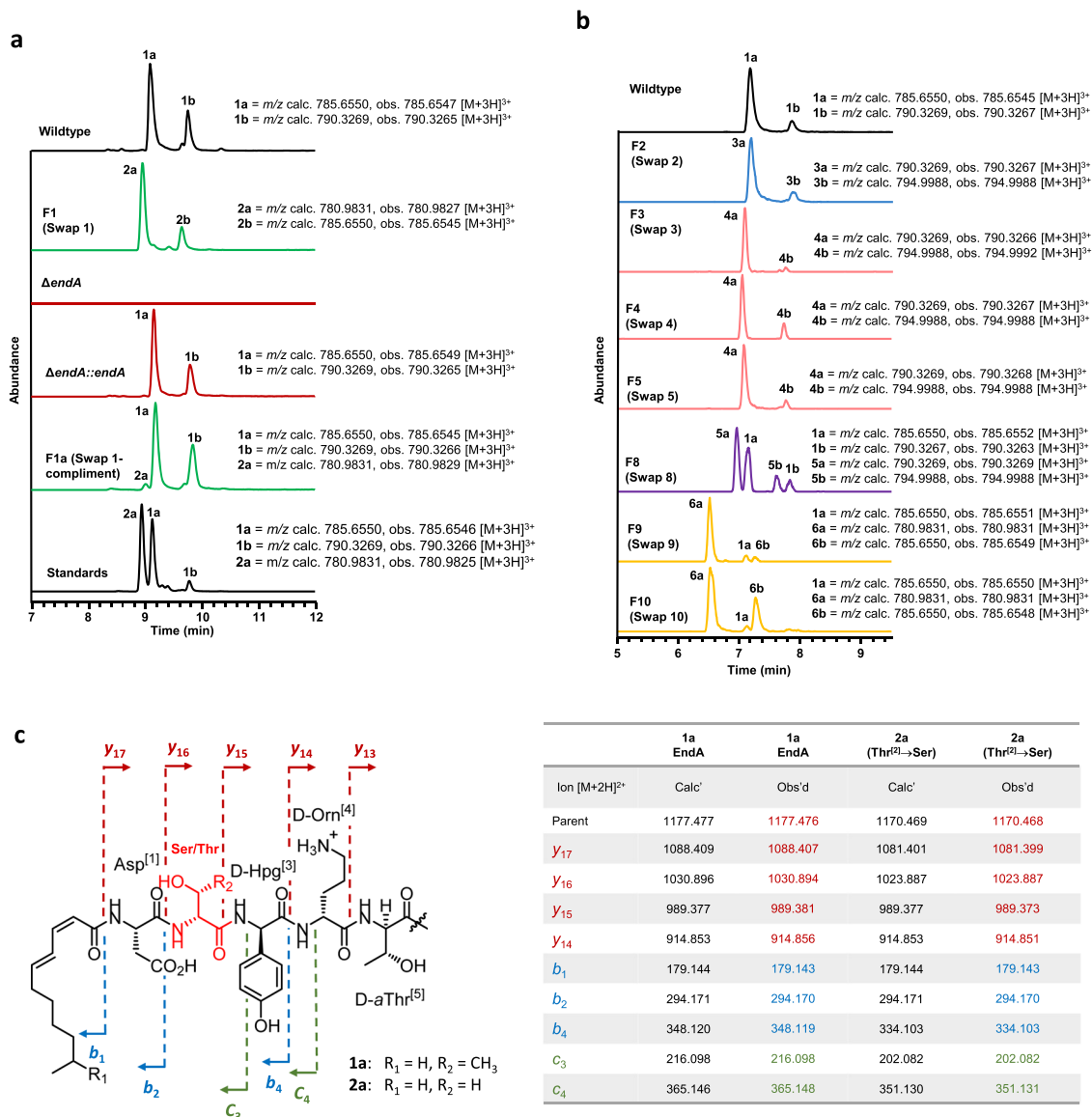
**Adenylation subdomain exchange via CRISPR-Cas9 to generate new enduracidin analogues.** Enduracidin (Fig. 1a) is produced by *Streptomyces fungicidicus* ATCC 21013 and is closely related to ramoplanin (Fig. 1b), from *Actinoplanes* sp. ATCC 33076, which entered phase III clinical trials for the treatment of vancomycin-resistant *Enterococcus*<sup>30,31</sup>. A total synthesis of ramoplanin has been developed, and used to produce variants with improved attributes, but this required over 40 chemical steps<sup>31</sup>, which is not viable for drug development. Biosynthetic engineering approaches for diversification of these lipopeptide antibiotics would be a far more attractive alternative to chemical synthesis. However, genetic manipulation of the huge and repetitive 17-module NRPS systems that deliver these lipopeptide antibiotics would be extremely challenging using the conventional double-crossover homologous recombination method<sup>16</sup>. Enduracidin biosynthesis requires four NRPSs (2.0 MDa) encoded by *endA-D* genes, encompassing 57 kb in total (Fig. 1a). The first of these, *EndA*, is relatively small and contains just two modules (Fig. 1a). The relatively low complexity of *EndA* made it a good candidate for initial testing of FSD swaps using CRISPR-Cas9, whilst also allowing comparison under optimal conditions with conventional recombinering methods (Fig. 2). An FSD comprising 139 amino acid residues within the second A domain of

*EndA* that recognizes L-Thr was identified<sup>27,32</sup>, and selected for replacement with the FSD from the L-Ser selective A domain of *EndC* (Supplementary Fig. 1a). A successful exchange of these two FSDs, which share 88% identity and 89% similarity (Figs. 2a and b, Supplementary Table 2), should result in enduracidin with L-Ser in place of L-Thr at position 2. By making this change at the native chromosomal locus, using the CRISPR-Cas9 system, any disruption to the integrity of the NRPS machinery should be limited and the creation of the mutant strains could be achieved more rapidly than with the laborious multi-step gene knockout/complementation approach.

A pCRISPOmyces-2-based plasmid<sup>28</sup> was constructed, containing a sgRNA that directed a double-stranded DNA cut within the targeted L-Thr FSD and a repair cassette comprised of *endA* with the native FSD exchanged for the alternative L-Ser FSD from *endC* (Supplementary Table 1 and Supplementary Data 1). This repair cassette should repair the resulting DNA break, via homologous recombination, with the copy of *endA* containing the desired FSD change (Swap 1, Fig. 2c, and Supplementary Table 2). The plasmid was introduced into the wild type enduracidin producer, *S. fungicidicus*, resulting in a clean exchange of the original FSD with the new L-Ser selective FSD, at the native locus, as determined by DNA sequencing (Supplementary Fig. 1b). RP-HPLC and LC-HRMS/MS analysis of the resulting mutant strain F1 (Thr to Ser) revealed production of new L-Ser containing variants of enduracidins a and b (**2a** & **2b**) with observed *m/z* 780.9827 ( $C_{106}H_{139}C_{12}N_{26}O_{31}$ ) and 785.6545 ( $C_{107}H_{141}C_{12}N_{26}O_{31}$ ) corresponding to the predicted  $[M + 3H]^{3+}$  ions (Fig. 3a) at a high titre



**Fig. 2 Subdomain swap strategies.** **a** The structure of a ligand-bound GrsA Phe-A domain (PDB: 1AMU)<sup>32</sup>. The flavodoxin-like subdomain (FSD) is highlighted in yellow and ligands phenylalanine and adenosine monophosphate (AMP) are shown in blue and red, respectively. Residues HGGISNLKVFENSLNV which form an alpha helix have been removed for greater visibility of the ligands. **b** Alignment of the *grsA* Phe-A domain with the *endA* Thr<sup>[2]</sup>-A domain, with the secondary structural features indicated and nine of the ten amino acid substrate binding residues (except Lys517) marked with red asterisks<sup>20</sup>. The FSD sequence is highlighted in yellow and the black triangles show cut sites for the subdomain swap. **c** Single-step CRISPR-Cas9 strategy for exchanging subdomains at the native locus. sgRNA-guided Cas9 cleaves within the Thr<sup>[2]</sup> A domain of *endA*. DNA repair utilizes a plasmid-borne sequence containing *endA* possessing a Thr<sup>[2]</sup> to Ser<sup>[12]</sup> subdomain swap, resulting in seamless exchange of the Thr<sup>[2]</sup> FSD with a Ser<sup>[12]</sup> FSD at its native locus within the BGC. **d** Schematic for conventional multi step gene knockout and complementation strategy. The wild type *endA* is replaced with an apramycin resistance cassette. The resultant  $\Delta endA$  strain is complemented with an integrative plasmid containing *endA* (Swap 1) where the subdomain sequence of the Thr<sup>[2]</sup> A domain is replaced with the subdomain from the Ser<sup>[12]</sup> A domain. Insertion of *endA* (Swap 1) occurs at the  $\Phi C31$  site outside of the enduracidin biosynthetic gene cluster (BGC).



**Fig. 3** Combined extracted ion chromatographs (EIC) of enduracidin a and b analogues from LC-HRMS analysis of extracts from primary *S. fungicidicus* strains (F1-10) used in this study (normalized to 100%), and tandem mass characterization of 2a. **a** LC-HRMS comparison between CRISPR-Cas9 editing and conventional gene complementation methods for introducing subdomain swaps. The F1a (Swap 1-complement) complementation strain produces mostly 1a & 1b, whereas the CRISPR-Cas9 edited strain *S. fungicidicus* F1 (Swap 1) produces Ser-containing enduracidins 2a & 2b. **b** LC-HRMS analysis of all productive mutants that generate new enduracidin analogues. **c** LC-HRMS/MS analysis of 1a and 2a confirms that Thr has been replaced with Ser in compound 2a. Fragment ions containing residue 2 are 14 mass units higher in 1a compared with 2a (methyl group of Thr). Loss of residue 2 results in fragment ions with identical m/z values between 1a and 2a.

of 65 mg/L, with only trace levels of wild type enduracidins (1a & 1b) evident (<3%). The high preference of this strain (>97%) for producing the L-Ser containing enduracidin (2a & 2b), rather than the wild type, confirms that introducing the subdomain change at the native chromosomal locus is very effective. The traces of 1a & 1b detected in the mutant strain (F1) could be due to the structural similarity between the amino acids (Thr & Ser). To confirm the structures of the new products, 2a was purified from strain F1 and compared with 1a isolated from the wild type *S. fungicidicus* ATCC 21013. Tandem MS analysis of 2a revealed fragment ions corresponding to a loss of CH<sub>2</sub> when compared to the equivalent ions of 1a (Fig. 3c and Supplementary Fig. 2). Fragment ions not predicted to contain the altered amino acid were identical for all four compounds, indicating the observed loss of CH<sub>2</sub> stems from a Thr to Ser switch at the desired position of enduracidin. The <sup>1</sup>H NMR spectra of 1a includes a doublet at

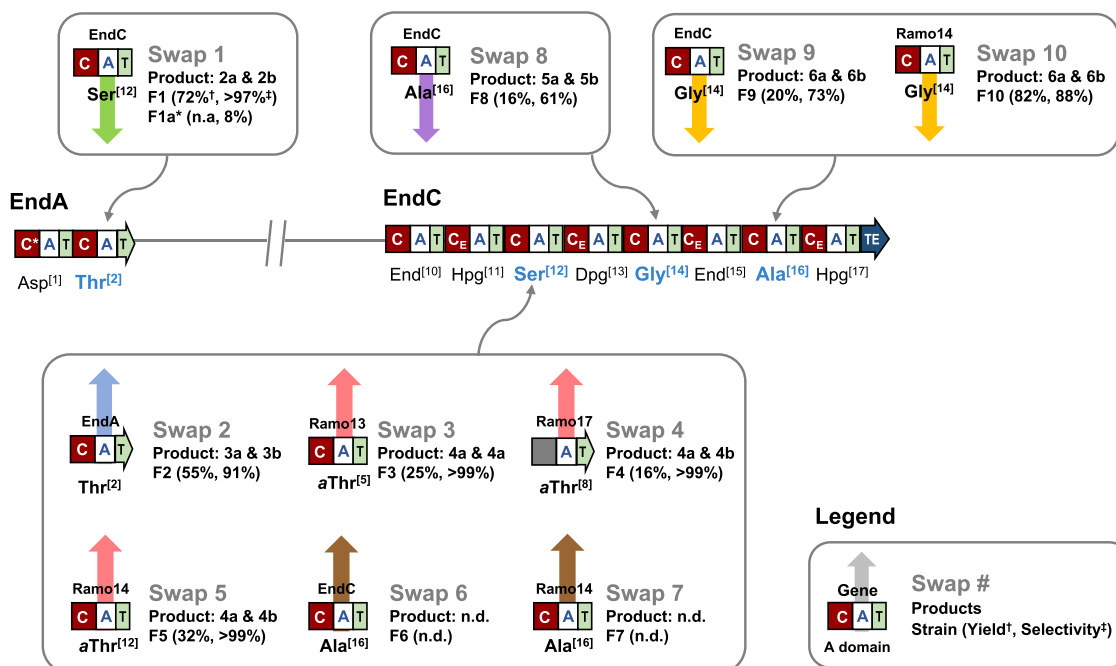
1.39 ppm, corresponding to the methyl moiety of the Thr residue, which was absent in the spectra of 2a as expected. In addition, a new NOESY correlation signal from the NH moiety of residue 2 at 8.52 ppm to a methylene (CH<sub>2</sub>) signal at 4.06 ppm in 2a, provides further support that the Thr to Ser exchange (Swap 1) has been successful (Supplementary Figs. 3–6 and Supplementary Table 3).

**A-domain FSD exchange via conventional gene deletion and complementation.** Given that *endA* is small, we chose to repeat Swap 1 using a more conventional gene knockout and complementation strategy for comparison with the CRISPR-Cas9 mediated procedure (Fig. 2d). First, we deleted *endA* from the *S. fungicidicus* chromosome by introducing an apramycin resistance cassette *apra<sup>R</sup>* in place of *endA* (Supplementary Fig. 7). The resultant  $\Delta endA$  strain was selected on apramycin-containing

media and was unable to produce wild type enduracidins **1a** & **1b** (Fig. 3a). A wild type copy of *endA* was introduced back into this strain (forming strain  $\Delta endA::endA$ ) under control of the constitutive *ermE\** promoter, using a vector integrating at the  $\Phi C31$  site, distal to the *end* biosynthetic gene cluster (BGC) locus. This restored production of **1a** & **1b**, albeit at lower levels (14 mg/L) when compared with the wild type strain (108 mg/L). Next, a copy of *endA* engineered with the L-Ser FSD from *endC* (Swap 1), was introduced into the  $\Delta endA$  strain via integration at the  $\Phi C31$  site. Surprisingly, the resulting strain (F1a) still produced wild type enduracidins (**1a** & **1b**) as the major products (9 mg/L), despite possessing a mutant EndA that is identical in sequence to that produced in the CRISPR engineered strain (F1). LC-HRMS analysis revealed additional products with *m/z* 780.9831 and 785.6550 corresponding to the predicted  $[M + 3H]^{3+}$  ions of the desired L-Ser containing enduracidin variants (**2a** & **2b**) (Fig. 3a). However, these variants were only observed at very low levels (*ca.* 8% that of **1a** & **1b**, as determined by LC-HRMS). Whole genome sequencing of strain F1a confirmed that the engineered *endA* gene was indeed inserted at a different locus on the chromosome. Given that *endA* would be expressed independently to the native *end* BGC, it is possible that the assembly and interactions of the EndA variant with the other NRPS and gene products of the *end* BGC could be perturbed. This, compounded by the selectivity of the downstream C-domain of EndB and any proofreading machinery, may result in a preference for introduction of L-Thr over L-Ser, especially if peptide assembly is slowed due to poor timing of gene expression and assembly of the NRPS machinery. The fact that the single-step CRISPR-Cas9 strategy was not only more successful in generating a successful amino acid substitution, but did so without a penalty in production titre, indicates that this is a much more efficient way of introducing A domain specificity changes than conventional means. Moreover, the multi-step conventional

approaches often leave scars, in this case, an apramycin resistance gene, which can severely limit subsequent genetic manipulation. The CRISPR-Cas9 approach, on the other hand, is scarless, facilitating further strain manipulation, which could potentially enable combinatorial FSD exchanges to be implemented.

**Exploring further FSD changes.** We next sought to explore the broader applicability of CRISPR-Cas9 subdomain swapping. Genes *endB* and *endC*, encoding six and eight NRPS modules, respectively, are significantly larger than *endA* and contain multiple repeating sequences. Knocking out *endB* or *endC* and complementation with engineered variants, possessing FSD swaps, would present a significant challenge and may not be achievable with standard methodologies. To address this, three modules within EndC, the largest enduracidin NRPS, were targeted for FSD changes by CRISPR-Cas9 (Ser<sup>[12]</sup>, Gly<sup>[14]</sup> and Ala<sup>[16]</sup>). Bioinformatics analysis of the A domain sequences of both the enduracidin and ramoplanin NRPS<sup>30,33,34</sup> (Fig. 1) guided the design of nine additional constructs for FSD replacement (Fig. 4 and Supplementary Table 2). First, the complementary exchange to Swap 1, where the L-Ser FSD of EndC was swapped with the L-Thr FSD from EndA, was performed in an analogous fashion. This exchange (Swap 2) resulted in two new enduracidin products **3a** and **3b** with L-Thr in place of the original L-Ser residue at position 12. The change was confirmed by LC-HRMS and LC-HRMS/MS (Fig. 3b and Supplementary Fig. 8). Only trace quantities of wild type enduracidin **1a** and **1b** were evident, with the desired new compounds (**3a** & **3b**) being produced with a selectivity of 91% versus the wild type. To explore if subdomains from a different gene cluster and organism could be introduced, three FSDs from the ramoplanin NRPS that are selective for L-*allo*-Thr were exchanged with the L-Ser FSD of EndC (Swaps 3–5). The selected replacement FSDs were from modules that incorporate *allo*-Thr into positions 5, 8 and 12 of ramoplanin,



**Fig. 4 Overview of the primary (Swap 1–10) subdomain swap mutants generated in this study.** Most of the subdomain swaps were designed for *endC*, the largest NRPS, containing multiple repeating sequences, which presents a significant challenge for engineering conventional homologous recombination methodologies. \*Swap via gene complementation, <sup>†</sup> Relative titre of engineered variants (**2–6**) compared titres of **1** produced by wild type, <sup>‡</sup> selectivity for new products (**2–6**) over parent enduracidin (**1**) in engineered NRPS. Coloured arrows are used to indicate different amino acid swaps: green (Thr<sup>[2]</sup> to Ser), blue (Ser<sup>[12]</sup> to Thr), pink (Ser<sup>[12]</sup> to *a*Thr), brown (Ser<sup>[12]</sup> to Ala), purple (Gly<sup>[14]</sup> to Ala) and yellow (Ala<sup>[16]</sup> to Gly). Source data for yield and selectivity are provided with this paper.

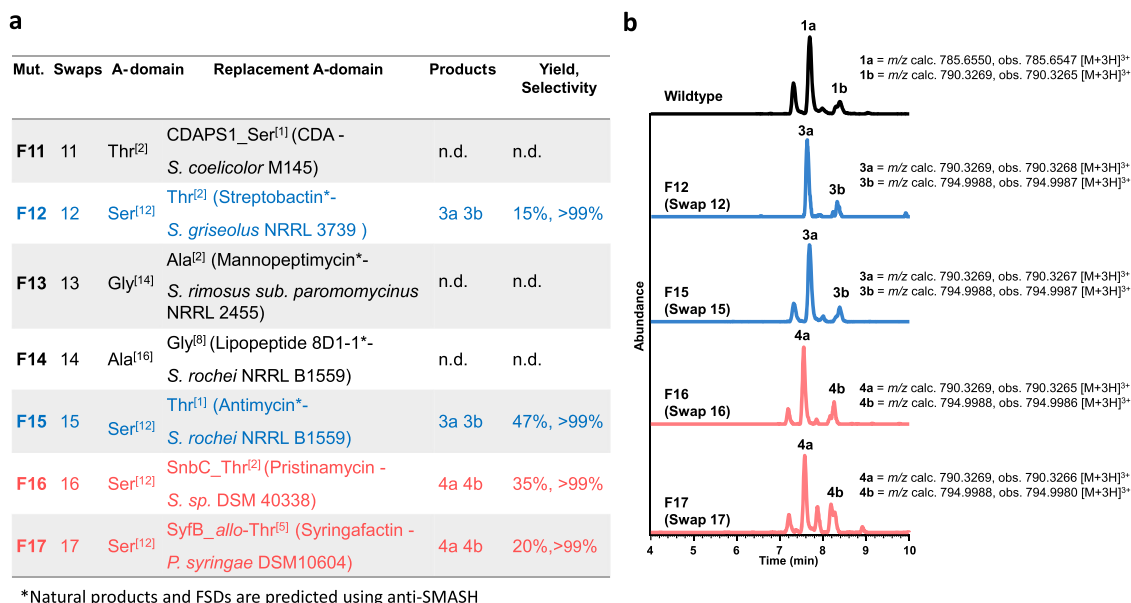
respectively (Fig. 1)<sup>31,33,34</sup>. All three resulting swaps generated two new products (**4a** & **4b**) consistent with a Ser to *allo*-Thr exchange with high selectivity as little or no **1a** and **1b** could be detected. Although the new products (**4a** & **4b**) have identical masses and MS/MS fragmentation patterns as the Thr containing **3a** and **3b**, LC-HRMS showed shifted retention times as expected for diastereoisomers (Fig. 3b).

Two further exchanges (Swap 6 & 7) were then attempted where the L-Ser FSD of EndC was replaced with Ala-activating FSDs from either EndC or the ramoplanin NRPS Ramo14. However, LC-HRMS analysis showed that these swaps failed to produce the Ala containing variants and production of enduracidin was abolished in both mutants (Supplementary Fig. 9). The FSD swaps in these two strains may have disrupted the activity of the A domain, preventing activation of Ala. Alternatively, strict specificity of the downstream C-domain could have disrupted the processing of the Ala containing variant<sup>14,15,35</sup>. The same Ala FSD from EndC was, however, successfully used to replace the Gly FSD of EndC (Swap 8), resulting in new enduracidin a and b analogues with an Ala in place of Gly (**5a** & **5b**) (Fig. 3b). The opposite replacement of the Ala FSD of EndC with the Gly-activating FSDs from both EndC and Ramo14 (Swap 9 & 10) was also successful and resulted in production of new enduracidin a and b variants consistent with a switch of Ala to Gly (**6a** & **6b**) (Fig. 3b). Production of analogues **5a**, **5b**, **6a** and **6b** was, however, accompanied with 12–39% of the wild type enduracidins **1a** & **1b**. This indicates an incomplete switch in selectivity, which may be due to the close structural similarity between alanine and glycine. To confirm the structures of the new engineered products, the most abundant enduracidin a variants (**3a**, **4a**, **5a** & **6a**) were isolated and subjected to tandem MS analysis. Analysis of fragmentation patterns enabled diagnostic fragmentation ions to be assigned, confirming the positions of the swapped amino acid residues (Supplementary Figs. 8 and 10–15).

To demonstrate the broader applicability of this method, seven additional FSD swap mutants were constructed using diverse NRPS from different bacterial strains (Fig. 5 and Supplementary

Table 1). One swap mutant for each of the positions discussed above (Thr<sup>[2]</sup>, Ser<sup>[12]</sup>, Gly<sup>[14]</sup> and Ala<sup>[16]</sup>) was constructed: Swap 11 (Thr<sup>[2]</sup> to Ser) used a Ser-selective FSD from the calcium-dependent antibiotic (CDA)<sup>36</sup> NRPS of *Streptomyces coelicolor* M145; Swap 12 (Ser<sup>[12]</sup> to Thr) used a predicted Thr activating FSD from a streptobactin-like<sup>37</sup> BGC in *Streptomyces griseolus* NRRL 3739; Swap 13 (Gly<sup>[14]</sup> to Ala) employed an FSD predicted to activate Ala from tyrobetaine-like<sup>38</sup> BGC in *Streptomyces rimosus sub. paromomycinus* NRRL 2455; and Swap 14 (Ala<sup>[16]</sup> to Gly) was constructed with a predicted Gly-selective FSD from a lipopeptide 8D1-1<sup>39</sup> BGC found in *Streptomyces rochei* NRRL B1559. Anti-SMASH<sup>40</sup> was used to identify candidate FSD for swaps 12, 13 and 14 from NRPS that exhibit high similarity to previously well characterized NRPS pathways<sup>37–39</sup>. Among these four mutants, only Swap 12 generated the desired enduracidin variants (**3a** & **3b**) with a selectivity of >99%, with only traces of the wild type enduracidins detected (Fig. 5). Swaps 13 and 14 showed abolished production, while Swap 11 showed traces (<1%) of wild type enduracidin **1a**. The fact that only Swap 12 produces the expected variants is most likely due to the high sequence identity (64%) and similarity (72%) between the FSD that were exchanged. In comparison, the FSD exchanged in Swaps 11, 13, and 14 have lower identity (29–44%) and similarity (40–59%) ranges (Supplementary Table 2).

Based on these results, further FSD exchanges were implemented: Swap 15 (Ser<sup>[12]</sup> to Thr) using an FSD from an NRPS in *Streptomyces rochei* NRRL B1559, which is predicted by anti-SMASH to produce antimycin<sup>41</sup>; Swap 16 (Ser<sup>[12]</sup> to Thr) introducing an FSD derived from the known pristinamycin pathway in *Streptomyces sp.* DSM 40338<sup>42</sup>; and finally Swap 17 (Ser<sup>[12]</sup> to *allo*-Thr) using an FSD from a BGC of *Pseudomonas syringae* DSM 10604 that is known to produce syringafactin<sup>43</sup>. All three mutants generated new enduracidin analogues, which was expected based on the high sequence similarity between the FSDs that are exchanged (Supplementary Table 2). Swap 15 produced variants **3a** & **3b** and Swaps 16 and 17 gave the compounds **4a** & **4b** at >99% selectivity (Fig. 5). Of the 17 FSD exchanges that were



**Fig. 5 Overview and extracted ion chromatographs (EIC) of enduracidins a and b analogues from LC-HRMS analysis of the additional subdomain swap mutants (F11–17) generated in this study.** **a** List of additional swap mutants constructed using FSDs from various NRPS to demonstrate the broad applicability of this method. **b** Combined EIC of enduracidins a and b analogues from LC-HRMS analysis of extracts from additional *S. fungicidicus* strains (F12 and F15–17) used in this study (normalized to 100%). The product for Swap 16 had a retention time consistent with the incorporation of *allo*-Thr. Source data for yield and selectivity are provided with this paper. \*Natural products and FSDs are predicted using anti-SMASH.

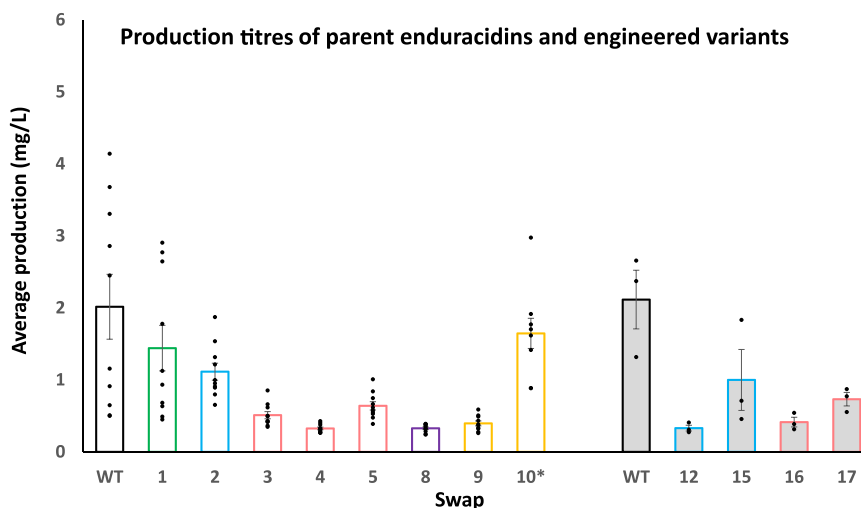
performed, 12 were successful in producing enduracidin variants. This provides a guideline for future engineering, with successful swaps occurring between FSD that have an identity of >55% and a similarity of >65% (Supplementary Table 2).

**Effect of FSD changes on production titre.** In order to quantify the effects of the subdomain changes on overall enduracidin titre, we cultivated all the subdomain swap mutants and the wild type *S. fungicidicus* under identical conditions. *Streptomyces* strains are known to be highly susceptible to variations in culture conditions; therefore ten replicates per strain were prepared and were cultured simultaneously to minimize disparities in culture time or temperature fluctuations. To accommodate parallel cultivation of a large number of strains and replicates, the fermentation conditions were optimized for higher throughput, which involved changing media components and reducing the volume of cultures. When grown under these new conditions, the wild type *S. fungicidicus* produced significantly lower enduracidin titres of  $2.02 \pm 0.45$  mg/L (**1a** & **1b**) when compared to the previous conditions that were optimized for production of **1a** and **1b** (108 mg/L) (Fig. 6, Supplementary Fig. 16 and Supplementary Table 4). However, these new conditions did allow a reliable comparison between strains to be made. Notably, several mutant strains produced enduracidin variants in yields approaching that observed by the wild type; strain F1 (Thr<sup>[2]</sup> to Ser, Swap 1), F2 (Ser<sup>[12]</sup> to Thr, Swap 2), and F10 (Ala<sup>[16]</sup> to Gly, Swap 10) produced variants (**2a** & **2b**), (**3a** & **3b**) and (**6a** & **6b**) in titres of 72%, 55% and 82% relative to the wild type. The three strains (F3–F5) containing Ser<sup>[12]</sup> to *allo*-Thr exchanges (Swaps 3–5) produced variant **4a** and **4b** in lower relative yields (16–32%) compared with the wild type. Strains F8 (Gly<sup>[14]</sup> to Ala, Swap 8) and F9 (Ala<sup>[16]</sup> to Gly, Swap 9) also gave reduced yields of 16–20% (Fig. 4, Fig. 6, and Supplementary Table 4). The strains F12 and F15–F17 which were generated from FSD derived from NRPS produced by more diverse *Streptomyces* and *Pseudomonas* species, gave similar yields of engineered enduracidins (**3a** & **3b** or **4a** & **4b**), in the range 16% to 47%, relative to the wild type strain (Fig. 6 and Supplementary Table 4).

**Whole genome sequencing to evaluate off-target effects of CRISPR-Cas9 editing.** In order to establish if the CRISPR-Cas9 gene editing caused any off-target effects that may have impacted on the mutant strains ability to produce enduracidin, six strains were subjected to PacBio whole genome sequencing. The wild type strain was also sequenced, and the mutant strains were aligned with this to identify any off-target mutations. Mutant strains with a range of enduracidin production levels were sequenced, from relatively high levels of production (F1, F2, and F10) to no production (F7). All of the mutant strains had the correct swapped subdomains, as previously confirmed by DNA sequencing. In addition, all the strains sequenced did not possess any mutations anywhere in the enduracidin gene cluster, which indicates that CRISPR-Cas9 gene editing of NRPS does not lead to any of the rearrangements that have been seen with genetic manipulations of other genes encoding related megasynthases<sup>17–19</sup>. Most mutant strains did have a number of single-nucleotide variants (SNVs) or small insertions or deletions (indels) (Supplementary Table 5). The three strains that had the highest titres of enduracidin production (F1, F2, and F10) did exhibit the lowest number of mutations across their genomes, with F2 having no mutations at all other than the FSD swap. The strain that produced no enduracidin (F7) also had a relatively low number of mutations (2 indels and 1 SNV), ruling out off-target effects of the CRISPR-Cas9 editing in explaining the lack of production in this strain. The other two strains (F8 and F9) sequenced had relatively low enduracidin production levels and a higher number of mutations within the genome (14 mutations for F8, and seven for F9), although none of these can explain the lower levels of enduracidin produced and so it is most likely that the swapped subdomain has affected the efficiency of these NRPSs.

## Discussion

In this study, we show how CRISPR-Cas9 gene editing can be used to engineer highly complex NRPS assembly lines, which to the best of our knowledge, has not been described previously. Several of the engineered NRPSs were highly selective for non-



**Fig. 6** Titres of enduracidin analogues produced by FSD swap mutant. Production levels (mg/L) for engineered enduracidin from FSD swaps compared with the wild type enduracidin (**1a** & **1b**). Non-filled bars represent a batch of ten replicates ( $n = 10$ ) for Swaps 1–5 and Swaps 8–10. Shaded bars represent a different batch of cultures carried out in triplicate ( $n = 3$ ) for Swaps 12, 15, 16 and 17. Data are presented as mean values  $\pm$  standard error. Each data point is overlaid on the graph as a dot plot. The wild type strain is included in both batches as a positive control. The production for Swap 10 was calculated from an average of eight samples, omitting two statistical outliers from flasks 1 and 5. The ratio of enduracidin a & b variants produced by wild type and engineered strains is variable and so the combined yields (a & b) are reported in each case to enable a clearer comparison of the productivity of each strain. Each bar is colour-coded for the analogues produced: green (**2a** & **2b**), blue (**3a** & **3b**), pink (**4a** & **4b**), purple (**5a** & **5b**) and yellow (**6a** & **6b**). Source data are provided with this paper.

cognate amino acid substrates, producing new lipopeptides in titres approaching the wild type strain. We observed that FSDs with identical selectivity, but from different origins, can have an effect on the production titre of the system as Swap 9 resulted in only 25% of the production of Swap 10 despite both FSDs selecting an identical amino acid (Fig. 4). Whilst the reasons for such variations are not clear, subtle differences in sequence and structure between subdomains may make some FSDs more compatible than others. Nevertheless, the high speed, efficiency and fidelity of CRISPR-Cas9 mediated replacements, compared to conventional methods, makes the screening of multiple FSD exchanges much more feasible, significantly increasing the scope to create optimized engineered assembly lines. Whole genome sequencing of six of the strains generated here indicate very accurate FSD swaps, with no other mutations introduced into the BGC or other pathways that are essential for antibiotic production. Moreover, the improvements seen with CRISPR-Cas9 compared to an identical switch using the conventional gene knockout/complementation approach also shows the benefits of this new approach.

Previously, CRISPR-Cas9 gene editing has been used to delete genes from BGC, to knock-in strong promoters for the activation of silent BGC and to enable *in vitro* engineering of PKS encoding genes for heterologous expression<sup>44–49</sup>. Here we demonstrate how CRISPR-Cas9 gene editing can be used to efficiently swap subdomains within large NRPS, in the native host, leading to engineered variants of complex lipopeptide antibiotics that would be extremely demanding to prepare synthetically. Our approach overcomes some of the limitations inherent in other NRPS engineering methods that rely on utilizing a model heterologous host, such as *E. coli*, where functional expression of large NRPS products can be problematic<sup>47</sup>. Our methodology is rapid, does not leave scars, and has the potential to be implemented in a combinatorial fashion. We therefore envisage that this method could be used to alter the sequences of a wide range of complex nonribosomal peptides in native hosts, “fine-tuning” their bioactivity and other properties. Although the scope of adenylation domain engineering may be limited by the specificity of downstream C-domains, the combination of our gene editing method with other NRPS engineering approaches, including new methods of modulating C-domain selectivity<sup>15</sup>, may allow for more significant changes in peptide structure to be realized. Additionally, the direct editing of gene clusters on the native chromosome means larger NRPS genes, such as *endC*, can now be targeted for engineering, which may not have been possible with conventional methods. Finally, we envisage that the gene editing approach described in this study could also be employed to re-programme other native pathways, including polyketide synthase (PKS) and hybrid NRPS-PKS assembly lines, which are also a prolific source of antibiotics and other therapeutic agents.

## Methods

### General fermentation of *S. fungicidicus* ATCC 21013 wild type and mutant strains, extraction and analysis of enduracidins. *Streptomyces fungicidicus*

spores were used to inoculate tryptone soy broth (TSB) media and the seed cultures cultivated at 30 °C for 2–3 days until a high cell density was reached. For fermentation, the seed culture (1.5 mL) was inoculated into a 250 mL Erlenmeyer flask containing 30 mL of enduracidin production media (3% (v/v) corn steep liquor (Sigma-Aldrich), 2% soybean flour (Holland & Barrett), 0.5% NaCl, 1% CaCO<sub>3</sub>, 2% glucose, 5% soluble starch (Sigma-Aldrich), adjusted to pH 7.0 prior to autoclaving). The fermentation cultures were incubated with 180 rpm agitation at 30 °C for 12–14 days.

For enduracidin crude extraction and analysis, the mycelia were pelleted by centrifugation and extracted with three volumes of acidic methanol (~pH 2.0). The methanol extracts were centrifuged and the supernatant analysed by RP-HPLC on a Shimadzu UFLC XR HPLC system with a Phenomenex Aeris XB-C18 peptide column 150 × 4.6 mm, 3.6 μm particle size, UV detection wavelength 267 nm, flow rate 1.5 ml min<sup>-1</sup> with the following gradient: 0–1 min 10% B, 1–1.5 min 10–30% B, 1.5–11.5 min 30–50% B, 11.6–13 min 95% B, 13.1–15 min 10% B, where mobile

phase A is H<sub>2</sub>O + 0.1% formic acid and mobile phase B is methanol + 0.1% formic acid. HPLC data analysis was performed using Shimadzu LabSolution Lite v5. LC-HRMS and MS/MS analyses of enduracidins were carried out on a Thermo Dionex Ultimate 3000 UHPLC coupled to a Thermo QExactive mass spectrometer, using a Thermo accucore C18 column, 100 × 2.1 mm, 2.6 μm particle size, flow rate 0.3 mL min<sup>-1</sup>, with the following gradient: 0–1 min 10% B, 1–1.5 min 10–25% B, 1.5–9 min 25–60% B, 9–9.2 min 60–95% B, 9.2–10.5 min 95% B, 10.5–11 min 95–10% B, 11–12.5 min 10% B, where mobile phase A is H<sub>2</sub>O + 0.1% formic acid and mobile phase B is methanol + 0.1% formic acid. For LC-MS/MS analysis an inclusion list was used to target compounds for fragmentation, with CID 25, 30, 35 eV. LC-MS data analysis was performed using Thermo Xcalibur v2.2, and ChemDraw Professional v16 was used for exact mass calculations.

**Isolation and NMR analysis of enduracidin a (1a) and analogue (2a).** Crude methanol extracts containing enduracidin were prepared as described previously. The methanol was removed under reduced pressure leaving a crude acidic aqueous extract that was extracted with two equal volumes of ethyl acetate. The aqueous layer was then adjusted to pH 7–8, and the enduracidin was extracted into butanol (two volumes). The butanol layers were pooled and the solvent removed under reduced pressure, leaving a yellow–brown solid. This was dissolved in 50% methanol and purified using semi-preparative RP-HPLC as follows: Shimadzu Prominence HPLC with a Phenomenex Gemini C18 column 250 × 10 mm, 5 μm particle size, UV detection wavelength 267 nm, flow rate 5 ml min<sup>-1</sup>, with the following gradient: 0–2 min 10% B, 2–18 min 10–60% B, 18–19 min 60–95% B, 19–21 min 95% B, 21–22 min 95–10% B, 22–25 min 10% B, where mobile phase A is H<sub>2</sub>O + 0.1% formic acid and mobile phase B is methanol + 0.1% formic acid. Fractions containing the desired enduracidins were collected, pooled and the methanol removed under reduced pressure. The residual water fraction was frozen, and the water was removed by lyophilization to leave pure enduracidin. For NMR analysis, the samples were dissolved in H<sub>2</sub>O:D<sub>2</sub>O:TFA 9:1:0.05 (v/v/v) and was performed on a B500 Bruker Avance 500 MHz NMR spectrometer. Previously reported NMR data of enduracidin was used as a reference in our NMR analysis of enduracidin 1a and analogue 2a<sup>50</sup>. NMR data analysis was performed using MestReNova v11.

**Quantification of enduracidin a and b by *S. fungicidicus* wild type and new variants by mutant strains.** To quantify the production of our new enduracidin versus wild type enduracidin, we cultivated all mutants and the wild type in replicate flasks. The mycelia were pelleted by centrifugation and extracted twice, first with three volumes of methanol followed by three volumes of acidic methanol (~pH 2.0). The methanol was then removed under reduced pressure and the remaining acidic crude aqueous extract was adjusted to pH 7–8, frozen and lyophilized to yield a dry crude extract. The extracts were then reconstituted with methanol and analysed by LC-HRMS under the same conditions as stated previously. Titres of target compounds were calculated based on the peak area of the extracted ion chromatogram of each enduracidin a and b analogue in relation to the standard curve (Supplementary Fig. 16) of a commercial standard containing enduracidin 1a and enduracidin 1b (Toku-E).

**Deletion of *end* ORFs in *S. fungicidicus* ATCC 21013.** Deletion of the *endA* gene was carried out using homologous recombination with a plasmid constructed using REDirect PCR targeting<sup>47</sup>. A 1.5 kb homologous region upstream of *endA* were PCR amplified from *S. fungicidicus* ATCC 21013 genomic DNA using primers JKF\_JM\_1F(5'-GCGCCTCGAGTCGCTATGTGTGCGCGTGCCT-3') and R(5'-GCGGAATTCTCCAGGTCCTTTCGACGCAT-3'), and inserted into a pBlue-script KS + based vector, pDWU5 (a gift from David Widdick, John Innes Centre, Supplementary Table 1), using XhoI and EcoRI restriction sites. Next, a 1.5 kb homologous region downstream of *endA* was amplified using primers JKF\_JM\_3F(5'-GCGGAATTCCCGCTGCGCCCGATGCGCTGA-3') and R(5'-GCGCTCTAGAAACCCCTCACCAGCCGCGA-3') and inserted using restriction sites EcoRI and XbaI. The resultant plasmid pDWU5-*ΔendA* containing upstream and downstream *endA* flanking arms was transformed into *E. coli* BW25113/pJ790.

An apramycin resistance cassette *apra*<sup>R</sup> was amplified from pJ773 using primers *endA\_RED\_F*(5'-CTGACAGCCGACTGAGGGCCTTCGAGGGGGAGGACGATGATTCGGGGATCCGTCGACC-3') and R(5'-GCGAACCAGCGCGGAATCATGTGGAACCCCTTTCCTTCATGTAGGCTGGAGCTGCTTC-3') to produce an oriT-*apra*<sup>R</sup>-FRT fragment that was subsequently inserted into plasmid pDWU5-*ΔendA* using PCR targeting, generating plasmid pDWU5-*ΔendA-apra*<sup>R</sup>. Introduction of the plasmid into *S. fungicidicus* ATCC 21013 was carried out by conjugal transfer from *E. coli* ET12567/pUZ8002. Ex-conjugant single crossover colonies (apramycin resistant and kanamycin resistant phenotypes) were selected on SFM medium plate containing apramycin and kanamycin. Double-crossover colonies containing the *endA* deletion were selected for apramycin resistant but kanamycin sensitive phenotypes. Confirmation of *endA* deletion was carried out by PCR (Supplementary Fig. 7).

**Complementation of the *ΔendA* strain with wild type *endA*.** The wild type *endA* gene was amplified from *S. fungicidicus* ATCC 21013 genomic DNA in two parts.



Prior to cloning the *endA* gene, the expression promoter (*ermE*\**p*) sequence was excised from pJ186 and inserted in pDWU5 using XbaI/BamHI to generate plasmid pDWU5-*endA*-1. Then, the first half of *endA* was amplified from genomic DNA using primers JFK\_JM\_5 F(5'-CGGCGGCCAGCTCGAGTTCG-3') and R(5'-GCGCGGTACCGAATCTGGGGCTCGCCG-TCGGCTTC-3') and inserted into pDWU5-*endA*-1 using XhoI and BamHI restriction sites to generate plasmid pDWU5-*endA*-2. The second half of *endA* was amplified using primers JFK\_JM\_7 F(5'-GCGCGGATCCCTTCGAGGGGGAGGACGATG-3') and R(5'-CCCCGGAAGCTCGAGTGG-3') and subcloned into pDWU5 using BamHI/XhoI to generate plasmid pDWU5-*endA*-1-2. The *ermE*\**p*-*endA* sequence from plasmid pDWU5-*endA*-2 was excised using XbaI and XhoI and inserted into pDWU5-*endA*-1-2 to generate pDWU5-*endA*, containing the full length *endA* sequence. Due to the  $\Delta$ *endA* strain containing an *apra*<sup>R</sup> cassette in place of *endA*, the *apra*<sup>R</sup> of plasmid pDWU5-*endA* was replaced with a hygromycin resistance cassette *hyg*<sup>R</sup>. The *hyg*<sup>R</sup> sequence was amplified from pIJ10700 using primers HygR F(5'-GATCGACTGATGTCATCAGCGGTGGAGTGCAATGTC-3') and R(5'-GCCCTCC AACGTCATCTCGTTCTCCGCTCATGAGCTCATCAGGCGCCGGGGCGGTGTCCG-3'). The *hyg*<sup>R</sup> fragment was inserted into pDWU5-*endA* using REDirect PCR targeting.

**Complementation of the  $\Delta$ *endA* strain with *endA*(Ser<sup>121</sup>FSD).** A section of *endA* immediately upstream of the Thr<sup>121</sup>FSD, was amplified using primers JFK\_JM123 F (5'-GCGCAAGCTTTCTAGAAAATCCCGCTTCGACGACCTCC-3') and R(5'-GCGCGATATCGAACTCGAACACCCCCCGGG-3') and inserted into pDWU5 using HindIII/EcoRV restriction sites to generate plasmid pDWU5-*endA*-1. Subsequently, the region of *endA* downstream of the Thr<sup>121</sup>FSD was amplified using primers JFK\_JM125F(5'-GCGCGATATCCCGGGTGTGCGGTGTTCTCG-3') and R(5'-GCGCGAATCTCTAGATGGGGCTCGCCGTCGGCTTC-3') and inserted into the plasmid pDWU5-*endA*-1 using EcoRV/EcoRI to generate plasmid pDWU5-*endA*-2. The middle part of *endA* was amplified from genome by using primer JFK\_JM\_149F(5'-CTGGACCGCCGGCCGAGCTC GAGTTCGCGGGCCGGCCGAC-3') and R(5'-TCGCATCAGGCGTCCCTCGC CGCTGCACC-3'). The final section of *endA* without Thr<sup>121</sup>FSD coding sequence was amplified from pDWU5-*endA*-2 by using primer JFK\_JM\_151F(5'-GGCG AGGACGCTGATGCGACGGCGGTGAC-3') and R(5'-AGCTGGGTACCG GGCCCCGAATTCTGGGGCTCGCCGTCGGCT-3'). The two PCR fragments, the middle part of *endA* and final part of *endA*, were inserted into plasmid pDWU5-*endA*-1-2 by HiFi assembly resulting in the plasmid pDWU5-*ermE*\**p*-*endA*-no-Thr<sup>121</sup>FSD. The apramycin resistance gene in pSET152 will be replaced with hygromycin resistance gene from pIJ10700 by PCR targeting (REDirect) using primers *apra2hygro*\_F(5'-GATCGACTGATGTCATCAGCGGTGGAGTGCAATGTC-3') and R(5'-GCCCTCCAACGTCATCTCGTTCTCCGCTCATGGCTCA TCAGGCGCCGGGGCGGTGTCCG-3') to create plasmid pSET152-*hygR*. The *ermE*\**p*:*endA*-no-Thr<sup>121</sup>FSD fragment was amplified using primer JFK\_JM\_213F(5'-CTCTAGAGGATCCGCGCCGCGCGCAATTCGAGCTCG GTACCAGC-3') and R(5'-ACAGCTATGACATGATTACGAATTCGAAT TCTGGGCTCGCCG-3') and inserted into the EcoRV site of pSET152-*hygR* by HiFi assembly to give pSET152-*hygR*:*ermE*\**p*:*endA*-no-Thr<sup>121</sup>FSD. The Ser<sup>121</sup>FSD-*endC* was amplified with primers JFK\_JM\_127F(5'-CCCCGGGGGTG TTCGAGTTCGGAGCAGACGACGTGTGGAGCGGCTTC-3') and R(5'-CAGG AACCCGACAACCCCGGACAGCCCGCCGATCAGCTGCCCCGTC-3') from genomic DNA of *S. fungicidicus* and inserted into the EcoRV site of the plasmid pSET152-*hygR*:*ermE*\**p*:*endA*-no-Thr<sup>121</sup>FSD to generate plasmid pSET152-*hygR*:*ermE*\**p*:*endA*-no-Thr<sup>121</sup>FSD-Ser<sup>121</sup>FSD.

**CRISPR-Cas9 editing: plasmid construction, conjugation and validation of subdomain replacement mutants.** Primers used in cloning CRISPR-Cas9 constructs are listed in Supplementary Data 1. Sequence alignments were performed using Clustal OMEGA web form and Vector NTI Suite v6. Synthetic protospacer sequences designed to cut within the subdomain sequence were synthesized by Genewiz, and inserted into pCRISPomyces-2 vector as reported in Cobb et al. 2015<sup>28</sup>. Upstream and downstream ~2–2.4 kb homology arms flanking each target A subdomains were amplified from *S. fungicidicus* genomic DNA using respective primers listed. By using HiFi assembly (NEB), these flanking arms were then inserted into the protospacer-containing pCRISPomyces-2 vector, which has been linearized at the XbaI site. Subsequently, subdomains for each corresponding exchange were amplified from genomic DNA using primers listed and inserted into the EcoRV site located between the two homology arms, generating each construct. All constructs were verified by restriction enzyme digestion and nucleotide sequencing. Introduction of each sequence verified-construct into *S. fungicidicus* was performed via intergeneric conjugation. These plasmids were first transformed into *E. coli* ET12567 /pUZ8002 and the resultant strains were used in subsequent conjugations following standard procedures<sup>16</sup>. All ex-conjugants formed (typically after 14 days of cultivation) were picked and streaked onto SFM plates supplemented with nalidixic acid (25 µg/mL) and apramycin (50 µg/mL). Colonies that grow after 7 days of incubation at 30 °C were then screened for successful subdomain exchange by colony PCR, and correct mutants were cultivated at high temperature (39 °C) to promote loss of the plasmid. Colonies that grow after the high-temperature incubation (~7 days), were screened again by replicate plating onto SFM with and without apramycin. Finally, successful subdomain swapping on

colonies that have lost apramycin resistance were confirmed by colony PCR and nucleotide sequencing.

**Whole genome sequencing of selected mutants.** Wild type *S. fungicidicus* as well as strains F1, F2, F7, F8, F9 and F10, were subject to whole genome sequencing. gDNA was isolated using NEB Monarch® Genomic DNA Purification Kit. The gDNA was then sheared using a gTube (Covaris, roughly 10 kb fragments) and prepared for sequencing following the SMRTbell Express Template Preparation Kit 2.0 protocol. Sequencing data were obtained using the Sequel system (Pacific Biosciences) with a 10 h acquisition. Assembly of the genome was performed using the HGAP4 algorithm in SMRT Link v8.0. Assembled genomes were aligned using mauve multiple genome alignment with Geneious Prime 2021.1.

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

## Data availability

All whole genome sequencing data generated in this study have been deposited in the NCBI genome database under BioProject PRJNA776125, samples SAMN22746534-SAMN22746540. The remaining data generated or analysed during this study are included in the Article, Supplementary Information, Supplementary Data file and Source Data file. Nucleotide sequences were obtained from GenBank and their accession numbers are as follows: *enduracidin* (DQ403252); *ramoplanin* (DD382878); *CDA* (CAB38518); *pristinamycin* (CAA67248); all other subdomain sequences are included in the Supplementary Methods. The original materials such as plasmids and strains that support the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Received: 9 September 2021; Accepted: 2 November 2021;

Published online: 25 November 2021

## References

- Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
- Brown, E. D. & Wright, G. D. Antibacterial drug discovery in the resistance era. *Nature* **529**, 336–343 (2016).
- Callaway, E. & Cyranoski, D. Anti-parasite drugs sweep Nobel prize in medicine 2015. *Nature* **526**, 174–175 (2015).
- van Bergeijk, D. A., Terlouw, B. R., Medema, M. H. & van Wezel, G. P. Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nat. Rev. Microbiol.* **18**, 546–558 (2020).
- Felnagle, E. A. et al. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **5**, 191–211 (2008).
- Sussmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis-principles and prospects. *Angew. Chem.* **56**, 3770–3821 (2017).
- Nam, J., Shin, D., Rew, Y. & Boger, D. L. Alanine scan of [L-Dap(2)] ramoplanin A2 aglycon: assessment of the importance of each residue. *J. Am. Chem. Soc.* **129**, 8747–8755 (2007).
- Rew, Y., Shin, D., Hwang, I. & Boger, D. L. Total synthesis and examination of three key analogues of ramoplanin: a lipoglycopeptide with potent antibiatic activity. *J. Am. Chem. Soc.* **126**, 1041–1043 (2004).
- Lohani, C. R., Taylor, R., Palmer, M. & Taylor, S. D. Solid-phase total synthesis of daptomycin and analogs. *Org. Lett.* **17**, 748–751 (2015).
- Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317–347 (2016).
- Gal, M. C., Thurston, L., Rich, P., Miao, V. & Baltz, R. H. Complementation of daptomycin *dptA* and *dptD* deletion mutations in trans and production of hybrid lipopeptide antibiotics. *Microbiology* **152**, 2993–3001 (2006).
- Miao, V. et al. Genetic engineering in *Streptomyces roseosporus* to produce hybrid lipopeptide antibiotics. *Chem. Biol.* **13**, 269–276 (2006).
- Nguyen, K. T. et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. USA* **103**, 17462–17467 (2006).
- Bozhuyuk, K. A. J. et al. De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018).
- Bozhuyuk, 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).
- Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics*. (John Innes Foundation, 2000).
- Yamanaka, K. et al. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. USA* **111**, 1957–1962 (2014).

18. Wlodek, A. et al. Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nat. Commun.* **8**, 1206 (2017).
19. Tao, W., Yang, A., Deng, Z. & Sun, Y. CRISPR/Cas9-based editing of *Streptomyces* for discovery, characterization, and production of natural products. *Front. Microbiol.* **9**, 1660 (2018).
20. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
21. Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224 (2000).
22. Uguru, G. C. et al. Active-site modifications of adenylation domains lead to hydrolysis of upstream nonribosomal peptidyl thioester intermediates. *J. Am. Chem. Soc.* **126**, 5032–5033 (2004).
23. Thirlway, J. et al. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew. Chem.* **51**, 7181–7184 (2012).
24. Kaljunen, H. et al. Structural elucidation of the bispecificity of A domains as a basis for activating non-natural amino acids. *Angew. Chem.* **54**, 8833–8836 (2015).
25. Kries, H. et al. Reprogramming nonribosomal peptide synthetases for “clickable” amino acids. *Angew. Chem.* **53**, 10105–10108 (2014).
26. Crüsemann, M., Kohlhaas, C. & Piel, J. Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chem. Sci.* **4**, 1041–1045 (2013).
27. Kries, H., Niquille, D. L. & Hilvert, D. A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol.* **22**, 640–648 (2015).
28. Cobb, R. E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**, 723–728 (2015).
29. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.* **4**, 1020–1029 (2015).
30. Yin, X. & Zabriskie, T. M. The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*. *Microbiology* **152**, 2969–2983 (2006).
31. Walker, S. et al. Chemistry and biology of ramoplanin: a lipoglycopeptide with potent antibiotic activity. *Chem. Rev.* **105**, 449–476 (2005).
32. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–4183 (1997).
33. McCafferty, D. G. et al. Chemistry and biology of the ramoplanin family of peptide antibiotics. *Biopolymers* **66**, 261–284 (2002).
34. Hoertz, A. J., Hamburger, J. B., Gooden, D. M., Bednar, M. M. & McCafferty, D. G. Studies on the biosynthesis of the lipodepsipeptide antibiotic ramoplanin A2. *Bioorg. Med. Chem.* **20**, 859–865 (2012).
35. Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta Proteins Proteom.* **1865**, 1587–1604 (2017).
36. Hojati, Z. et al. Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem. Biol.* **9**, 1175–1187 (2002).
37. Matsuo, Y. et al. Streptobactin, a tricatchol-type Siderophore from marine-derived *Streptomyces* sp. YM5-799. *J. Nat. Prod.* **74**, 2371–2376 (2011).
38. Parkinson, E. I. et al. Discovery of the tyrobetaine natural products and their biosynthetic gene cluster via metabologenomics. *ACS Chem. Bio* **13**, 1029–1037 (2018).
39. Xu, M. et al. Functional genome mining for metabolites encoded by large gene clusters through heterologous expression of a whole-genome bacterial artificial chromosome library in *Streptomyces* spp. *Appl. Environ. Microbiol.* **82**, 5795–5805 (2016).
40. Blin, K. et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* **49**, W29–W35 (2021).
41. Liu, J., Zhu, X., Kim, S. J. & Zhang, W. Antimycin-type depsipeptides: discovery, biosynthesis, chemical synthesis, and bioactivities. *Nat. Prod. Rep.* **33**, 1146–1165 (2016).
42. Mast, Y. et al. Characterization of the ‘pristinamycin supercluster’ of *Streptomyces pristinaespiralis*. *Microb. Biotechnol.* **4**, 192–206 (2011).
43. Götze, S. et al. Structure elucidation of the syringafactin lipopeptides provides insight in the evolution of nonribosomal peptide synthetases. *Chem* **10**, 10979–10990 (2019).
44. Alberti, F. & Corre, C. Editing streptomycete genomes in the CRISPR/Cas9 age. *Nat. Prod. Rep.* **36**, 1237–1248 (2019).
45. Zhang, M. M. et al. CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. *Nat. Chem. Biol.* **13**, 607–609 (2017).
46. Culp, E. J. et al. Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics. *Nat. Biotechnol.* **37**, 1149–1154 (2019).
47. Zhang, J. J. & Moore, B. S. Site-directed mutagenesis of large biosynthetic gene clusters via oligonucleotide recombineering and CRISPR/Cas9 targeting. *ACS Synth. Biol.* **9**, 1917–1922 (2020).
48. Ferrara, M., Haidukowski, M., Logrieco, A. F., Leslie, J. F. & Mule, G. A CRISPR-Cas9 system for genome editing of *Fusarium proliferatum*. *Sci. Rep.* **9**, 19836 (2019).
49. Kudo, K. et al. In vitro Cas9-assisted editing of modular polyketide synthase genes to produce desired natural product derivatives. *Nat. Commun.* **11**, 4022 (2020).
50. Castiglione, F., Marazzi, A., Meli, M. & Colombo, G. Structure elucidation and 3D solution conformation of the antibiotic enduracidin determined by NMR spectroscopy and molecular dynamics. *Magn. Reson. Chem.* **43**, 603–610 (2005).

## Acknowledgements

We thank BBSRC for support through grants BB/L013754/1 and BB/N023536/1. We also thank The Royal Society for support through the grant ICA\1\191148. The Michael Barber Centre for Collaborative Mass Spectrometry provided access to MS instrumentation. We are grateful for the Royce equipment grant (EP/P025021/1) for PacBio Sequel Next Generation Sequencing instrumentation and to Dr Andrew Currin for technical support. We thank Dr Michael Winn for helpful discussion and editing of the draft manuscript. Catherine Lowe is also acknowledged for technical experimental support.

## Author contributions

J.M. supervised the project; W.T., Y.Zha., Y.Zhu., K.J.R., J.K.F. and J.M. designed experiments; W.T., Y.Zha., Y.Zhu., K.J.R., A.J.H., J.K.F. and B.J.C.L. carried out the experiments; W.T. and J.M. wrote the manuscript; all authors analyzed data and reviewed the manuscript.

## 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/s41467-021-27139-1>.

**Correspondence** and requests for materials should be addressed to Jason Micklefield.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021