Successes, surprises and pitfalls in modular polyketide synthase engineering: generation of ring-contracted stambomycins

Li Su

University of Lorraine <https://orcid.org/0000-0002-7787-684X>

Laurence Hotel

University of Lorraine

Cédric Paris

University of Lorraine

Alexander Brachmann

ETH Zurich

Jörn Piel

ETH Zurich

Christophe Jacob

Lorraine University

Bertrand Aigle

University of Lorraine <https://orcid.org/0000-0001-5266-5926>

Kira Weissman (✉ kira.weissman@univ-lorraine.fr)

University of Lorraine <https://orcid.org/0000-0002-3012-2960>

Article

Keywords: polyketide synthases (PKSs), PKS engineering strategies

Posted Date: February 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-222743/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on January 26th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-27955-z>.

Abstract

The modular organization of the type I polyketide synthases (PKSs) would seem propitious for rational engineering of desirable analogues. However, despite decades of efforts, such experiments remain largely inefficient. Here, we combined multiple, state-of-the-art approaches including modification of docking domains, use of modules of varying domain composition, alternative interdomain fusion sites, and targeted adaptation of key domain-domain interfaces, to reprogram the stambomycin PKS by deleting seven internal modules – the most substantial modification to an intact system reported to date. One such system produced the target 37-membered mini-stambomycin metabolites, a reduction in chain length of 14 carbons relative to the 51-membered parental compounds, but also substantial quantities of shunt metabolites released from the multienzyme subunit upstream of the newly-installed junction. Our data also provide evidence for an unprecedented off-loading mechanism of such stalled intermediates involving the C-terminal thioesterase domain acting on chains located four modules upstream. The yields of all metabolites were substantially reduced compared to the wild type compounds, likely reflecting the poor tolerance to the non-native substrates of the modules downstream of the introduced interfaces. Taken together, our data demonstrate that even ‘optimized’ PKS engineering strategies remain inadequate for efficient production of target polyketide derivatives, and highlight several areas for future investigation.

Introduction

For almost thirty years, efforts have been made to leverage the modular genetic architecture of the type I polyketide synthases (PKSs) to generate novel derivatives, typically by modifying individual catalytic domains. Despite enormous progress in establishing domain structure-function relationships^{1,2}, such genetic manipulation remains inefficient³. Insight into factors potentially contributing to low product yields was provided by cryo-electron microscopy analysis of a model PKS module at multiple stages of its catalytic cycle^{4,5}. This work revealed that interdomain contacts are critical for establishing the various functional states of the module, and that transitions between such states rely on evolving interfaces between the domains, as well as the intervening ‘linker’ regions. In short, PKS modules appear to be highly integrated units, thus explaining why exchange of catalytic domains for heterologous counterparts is often detrimental⁶. Collectively, these observations motivate future approaches in which modules or multi-modular subunits are employed as the basic building blocks for engineering the assembly lines^{7–13}.

Nevertheless, using modules requires a clear definition of their domain composition. Traditionally, modules encompass the three invariable domains required for monomer selection and chain extension (ketosynthase (KS), acyl transferase (AT) and acyl carrier protein (ACP)), as well as any intervening β -keto processing activities (e.g. ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER)) (Fig. 1a). However, an alternative definition was recently suggested by the finding that KS domains in certain PKSs co-evolve with the tailoring domains located upstream in the assembly lines^{14,15}. Accordingly, modules begin with the modifying domains and the associated AT, and terminate with the KS that is classically

assigned to the downstream module (Fig. 2). Even before a module redefinition was suggested, engineering efforts revealed that maintaining the key ACP_n/KS_{n+1} interface can, in certain cases, be critical for the function of a hybrid PKS^{7,10,16}. Recently, we have carried out module swapping based on both of these definitions, by covalently tethering heterologous modules to a common donor module within a bimodular mini-PKS¹⁷. Overall, our data demonstrated that both module definitions led to functional hybrid PKSs, and which boundaries worked best depended on the source module¹⁸. Indeed, regardless of which extremities are employed, module exchange results in non-native interdomain interactions (ACP_n/KS_{n+1} or KS_{n+1}/ACP_{n+1}), and in the case of classical module boundaries, potential incompatibilities in terms of KS substrate specificity (Fig. 2) – both of which have been shown to reduce activity via detailed studies *in vitro*^{19–21}.

In this work, we aimed to investigate the generality of these findings for efforts to create non-native intermodular junctions when the modules are located on distinct subunits¹⁰. In such cases, the resulting non-covalent interactions are mediated by short sequences at the extreme C- and N-termini of the subunits called docking domains (DDs)²² (Figs. 1 and 2). Matched pairs of DDs form specific complexes at intersubunit interfaces, enforcing a strict subunit ordering within the PKS system. As a test case, we aimed to genetically engineer the biosynthesis of substantially smaller derivatives of the stambomycin family of polyketides in *Streptomyces ambofaciens* ATCC23877²³. The stambomycins **1** (Fig. 1b) are glycosylated macrolides which show promising anti-cancer activity against multiple human cancer cell lines²³. The six characterized family members (A – F) differ in the alkyl functionality at the C-26 position (Fig. 1a) which directly impacts their potency,^{23,24} but have in common modification by *trans*-acting cytochrome P450 hydroxylases at C-28 and C-50. Notably, at 51-members, the macrolactone ring is among the largest of all known polyketides. Thus, the stambomycins represent an attractive model system for establishing PKS engineering as a means to access structurally-simplified analogues (minimal pharmacophores²⁵) for biological evaluation, as a complement to traditional chemical synthesis.

Here we report a comprehensive series of experiments aiming to generate 37-membered ring stambomycin analogues, based on both the classical and revised module definitions. The target mini-stambomycins were detected successfully, albeit in low yields, and the identification of the derivatives allowed us to clarify the relative timing of the two cytochrome P450-catalyzed hydroxylation reactions in the pathway. Attempts to boost titers by ACP/KS interface engineering^{19,20,26} were unsuccessful, but led in certain cases to a surprising increase in the liberation of linear shunt metabolites – only the second report, to our knowledge, of inter-subunit crosstalk resulting in thioesterase (TE)-mediated chain release. Taken together with recent work by others¹⁰, our data reinforce the idea that in order to boost efficacy, strategies based on modifying PKS intersubunit interfaces must take into account the function of the modules acting downstream from the newly-established junctions.

Results

Design of engineering experiments based on classical modular boundaries. The stambomycin PKS comprises 25 modules distributed among 9 polypeptides (Pks1 – 9)²³ (Fig. 1a) (Note: throughout the text, the stambomycin genes have been numbered in accordance with ref. ²³). To access abridged derivatives using the classical module boundaries, we reasoned that we could engineer novel intersubunit interfaces by suitable manipulation of docking domains. Encouragingly, the extreme C- and N-termini of all subunits (with the exception of the N-terminus of Pks1 and the C-terminus of Pks9) contain sequences with convincing homology to previously-identified DDs^{22,27} (the C-terminal DDs are referred to hereafter as ^CDDs and their partner N-terminal DDs as ^NDDs). By bioinformatics analysis, we were able to confidently assign the DDs acting at 6 of the 8 interfaces to the type '1a' class²², and the remaining two sets of DDs as type '1b'²⁷ (**Supplementary Fig. 1**). In both cases, docking occurs between an α -helical ^CDD and a coiled-coil formed by the ^NDD, with specificity achieved via strategically-placed charge:charge interactions at the complex interface (**Supplementary Fig. 1**)^{22,27}.

Among the type 1a junctions, there were notably two which appeared compatible in terms of the translocated substrate: PKSs 3/4 + 7/8 and Pks 4/5 + 8/9 (**Supplementary Fig. 2**). Specifically, the functional groups at the critical α - and β - positions^{14,28} of the transferred chains are identical at these junctions, and correspondingly, the downstream KSs show similarities across several sequence motifs previously correlated with substrate specificity^{14,20,29} (**Supplementary Fig. 2**). Targeting such interfaces thus allowed us, at least in principle, to overcome the functional block to the engineered systems represented by poor recognition of the incoming substrate by the directly downstream KS domain²¹. Ultimately, we targeted a new interface between Pks subunits 4 and 9 for two principle reasons. Firstly, Pks4 is at the origin of the structural variation between the stambomycin family members, and thus we anticipated that maintaining the subunit within the hybrid system would give rise to a corresponding series of truncated analogues, providing important evidence for their identity. Secondly, it was genetically more practical to modify the second set of interfaces due to splitting of the PKS subunits between two loci (Fig. 1).

To establish the novel junction, we initially modified the ^CDD of Pks4 (^CDD₄) to match that of Pks8 (the natural partner of the ^NDD of Pks9 (^NDD₉)), either by site-directed mutagenesis of residues previously identified as key mediators of interaction specificity (construct ^CDD₄ SDM; **Supplementary Fig. 3** and **Supplementary Table 1**)²², or by exchange of the complete ^CDD docking α -helix of ^CDD₄ for that of ^CDD₈ (construct ^CDD₄ helix swap; **Supplementary Fig. 3** and **Supplementary Table 1**)³⁰. Modifying the ^CDD₄ specificity 'code' to match that of ^CDD₈ required mutation of 3 residues, while for the ^CDD₄ helix swap, the terminal 16 amino acids of ^CDD₄ were exchanged for the corresponding 15 residues of ^CDD₈ (**Supplementary Fig. 3** and **Supplementary Table 1**). The genetic alterations were carried out in two distinct PKS contexts: (i) in the presence of the intervening subunits 5 – 8, which allowed for the possibility of competitive interactions between modified Pks4 and both Pks5 and Pks9; and ii) removing the intervening subunits 5 – 8, thus eliminating competition for binding of Pks4 by Pks5, and of Pks9 by Pks8 (**Supplementary Fig. 3**). We further generated a mutant in which Pks subunits 5 – 8 were deleted but

no modification was made to $^{\text{C}}\text{DD}_4$, in order to judge the intrinsic capacity of Pks4 and Pks9 to interact. Furthermore, genetic engineering was carried out in parallel by both PCR-targeting³¹ and CRISPR-Cas9³², in order to directly compare the efficacy of these two approaches, as well as evaluate the effect of the short scar sequence remaining in the chromosome following PCR-targeting.

Engineering the stambomycin PKS based on the classical module definition. The $^{\text{C}}\text{DD}_4$ SDM and $^{\text{C}}\text{DD}_4$ helix swap sequences were introduced in parallel into the *S. ambofaciens* genome using PCR targeting and CRISPR-Cas9 (full experimental details are provided in the **Supplementary Methods**). As discussed previously, the modifications were made both in the presence of the intervening subunits Pks5 – 8 and in their absence (**Supplementary Fig. 3**). As previous work has shown that production from the stambomycin biosynthetic gene cluster requires activation by constitutive overexpression of a pathway-specific LAL (Large ATP-binding regulators of the LuxR family) regulator²³, we additionally introduced the LAL overexpression plasmid (pOE484) into each of the mutants, using the empty parental plasmid (pIB139³³) as a control. In total, this strategy resulted in 20 targeted strains harboring interface mutants (where K7N refers to PCR targeting and CPN to CRISPR-Cas9 engineering): K7N1/pIB139, K7N1/OE484, K7N2/pIB139, K7N2/OE484, K7N3/pIB139, K7N3/OE484, K7N4/pIB139, K7N4/OE484, K7N5/pIB139, K7N5/OE484, K7N6/pIB139, K7N6/OE484, CPN1/pIB139, CPN1/OE484, CPN2/pIB139, CPN2/OE484, CPN4/pIB139, CPN4/OE484, CPN5/pIB139, CPN5/OE484 (**Table 1, Supplementary Tables 2 – 4**; despite extensive efforts the CPN3 mutant strain was not obtained). The principal difference between the K and CPN series of constructs is the presence of a 33 bp ‘scar’ sequence between the modified *pks4* and *pks9* genes (**Supplementary Fig. 4**). Construct K7N6 was assembled specifically to test the effect of this region, without any further modification to $^{\text{C}}\text{DD}_4$ and the intervening *pks5* – *pks8* genes.

With the exception of K7N3, CPN4 and CPN5, extracts of the engineered mutant strains harboring pOE484 were analyzed by high performance liquid chromatography heated electrospray ionization high-resolution mass spectrometry (HPLC-HESI-HRMS) on a Dionex UltiMate 3000 HPLC coupled to a Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer, and compared to extracts of the control strain containing pIB139³³ as well as the wild type *S. ambofaciens* containing pOE484, using SIEVE 2.0 screening software. K7N3, CPN4 and CPN5 were analyzed subsequently, and the data inspected manually. Yield quantification was carried out with reference to a calibration curve generated with purified stambomycins A/B **1** (the limit of detection was found to be between 10 and 1 $\mu\text{g L}^{-1}$, and so any yields < 10 $\mu\text{g L}^{-1}$ must be considered an estimate). Novel metabolites not present in the control strains, and for which we obtained reliable exact masses, are listed in **Table 1** and **Supplementary Fig. 5**.

The first notable result is that the K7N6/OE484 mutant yielded a similar metabolic profile to *S. ambofaciens* wt ($22 \pm 3 \text{ mg L}^{-1}$, 73% relative yield (**Supplementary Table 5**)), showing that the scar sequence negatively impacted stambomycin production, but not dramatically (Fig. 3). By contrast, no stambomycins were observed, as anticipated, in all constructs in which Pks5 – Pks8 had been removed (K7N1 – 3; CPN1,2) (Fig. 3). Stambomycins **1** were present, however, in strains K7N4 and CPN4 harboring $^{\text{C}}\text{DD}_4$ site-directed mutations and in the $^{\text{C}}\text{DD}_4$ helix swap strain CPN5, all of which still contained Pks5 –

Pks8, albeit at reduced amounts relative to the wild type (18%, 23% and 14% of wt, respectively) (Fig. 3 and **Supplementary Table 5**). (Surprisingly, the metabolic profile of K7N5 reproducibly differed from that of CPN5, as no stambomycin-related metabolites were detected (Fig. 3)). These data suggested that while the mutations introduced into CDD_4 reduced interaction with NDD_5 , they were not sufficient to disrupt natural chain transfer between Pks4 and Pks5, arguing that DD engineering to alter partner choice should be accompanied by removal of competing intersubunit interactions.

We did not find any evidence in the DD engineering experiments for any of the target 37-membered metabolites (**Supplementary Figs. 3 and 5**). However, all strains in which stambomycin production was abolished (**Table 1**) exhibited four new peaks in common (Fig. 3b and **Supplementary Fig. 5**) (other peaks corresponding to potentially novel compounds were observed, but none were shared between multiple strains). The determined exact masses and mass spectra (as exemplified by strain CPN2/OE484, Fig. 3b) correspond to truncated derivatives of stambomycins A/B and C/D respectively, following premature release from modules 13 and 12 of Pks4 (compounds **2 – 5**, Fig. 3d and **Supplementary Fig. 5**; ca. 8-fold greater yield of the module 13 products (**Supplementary Table 5**)). Further support for the identity of these shunt compounds was obtained by grafting the chain-terminating (type I) thioesterase (TE) domain from the C-terminal end of Pks9 to the C-terminus of Pks4 in order to force chain release at this stage. Indeed, identical compounds were produced, but at 17-fold increased yield relative to CPN2/OE484, consistent with active off-loading of the chains (Fig. 3c, **Supplementary Fig. 6** and **Supplementary Table 5**).

Based on their masses, both sets of shunt metabolites were hydroxylated on a single carbon, while none were found to bear the β -mycaminose of the mature stambomycins, consistent with the absence of the tetrahydropyran moiety to which it is normally tethered. To determine the location of the hydroxylation and therefore the hydroxylase responsible, we inactivated in mutant CPN2/OE484 the genes *samR0478* and *samR0479* encoding respectively, the stambomycin C-28 and C-50 cytochrome P450 hydroxylases³⁴. While extracts of CPN2/OE484/ Δ 478 were unchanged relative to CPN2/OE484 (i.e. the hydroxyl group was still present), the CPN2/OE484/ Δ 479 mutant exhibited four new peaks with masses corresponding to the dehydroxylated shunt products (Fig. 3, **Supplementary Fig. 7** (compounds **6 – 9**) and **Supplementary Table 5**). Taken together, these data show that the unusual on-line modification catalyzed by SamR0479³⁴ which is necessary for macrocyclization, occurs prior to chain extension by Pks5. While SamR0478 has also been speculated to act during chain assembly³⁴, hydroxylation evidently occurs downstream of Pks4, at least. The intriguing substrate structural and/or protein-protein recognition features controlling the timing of hydroxylation by these P450 enzymes remain to be elucidated.

Role of TE domains in release of the shunt metabolites. We attributed the observed shunt metabolites **2 – 5** to the lack of productive chain translocation between Pks4 and Pks9, causing intermediates to accumulate on ACPs 12 and 13. To evaluate whether these were released by spontaneous hydrolysis or enzymatically, we further investigated the role of Pks9 TEI³⁴ in chain release, as well as that of SamR0485, a proof-reading type II TE³⁵ located in the cluster. Both TEs were disabled by site-directed mutagenesis of the active site serines (Ser $\rightarrow$ Ala) (**Supplementary Fig. 6**).

Interestingly, inactivation of both the type I and type II TEs reduced the yields of shunt products **2 – 5** relative to the parental strain CPN2/OE484 (by 66% and 27%, respectively; average of duplicate experiments) (**Supplementary Fig. 6** and **Supplementary Table 5**). These data clearly show that premature release of the chains is catalyzed, at least in part, by both TEs in the cluster, although spontaneous liberation also occurs. While type II TEs typically interact with acyl-ACPs *in trans* to release blocked chains³⁵, the effect of the Pks9 TEI is less readily explained. One possibility is that the new productive docking interaction between Pks4 and Pks9 allows Pks9 to adopt an alternative conformation from which the TE can off-load intermediates bound to ACPs 12 and 13 of Pks4 (**Supplementary Fig. 6**). Although this mechanism is reminiscent of that used by the pikromycin PKS to generate both 12- and 14-membered rings³⁶, the pikromycin TEI is separated from its alternative ACP target by a single module, while Pks9 TEI is located five or four modules downstream from ACPs 12 and 13 in the engineered system, which would seem to necessitate substantial inter-subunit acrobatics.

Understanding the docking domain engineering via studies in vitro with recombinant domains. To better understand the results of the DD engineering, we studied *in vitro* the wild type DD pairs (^CDD₄/^NDD₅ and ^CDD₈/^NDD₉), as well as binding between the modified versions of ^CDD₄ and wild type ^NDD₉. Design of suitable expression constructs in *E. coli* (**Supplementary Tables 1 – 3**) was based on bioinformatics analysis of the C-terminal ends of Pks4 and Pks8, and the N-termini of Pks5 and Pks9, and secondary structure analysis using PSIPRED³⁷ (**Supplementary Fig. 8**). Overall, we expressed and purified the following proteins in recombinant form from *E. coli*: ^CDD₄ wt, ^CDD₄ SDM, ^CDD₄ helix swap, ^NDD₅, and ^CDD₈ (**Supplementary Fig. 8, Supplementary Table 4**). As ^NDD₉ proved insoluble when expressed in *E. coli*, two versions with alternative start sites were obtained as synthetic peptides (Met and Val; **Supplementary Fig. 8, Supplementary Table 4**). Analysis of the individual ^CDDs by circular dichroism (CD) confirmed their expected high α -helical content (^CDD₄ wt (100 μ M): 58%; ^CDD₈ wt (100 μ M): 49%), and showed no evident effect of the introduced mutations on secondary structure (**Supplementary Fig. 8**). All of the constructs were further confirmed to be homodimeric by size exclusion chromatography-multi-angle light scattering (SEC-MALS) (**Supplementary Fig. 8**).

The two ^NDDs also exhibited α -helical character, though less pronounced than the ^CDDs (^NDD₅ (100 μ M): 27%; ^NDD₉ Met (100 μ M): 21%; ^NDD₉ Val (100 μ M): 25%), and were monomeric by SEC-MALS (**Supplementary Fig. 8**). The latter result was surprising, as type 1a ^NDDs classically form a homodimeric coiled-coil domain (Fig. 1, **Supplementary Fig. 1**), but we recently identified functional, monomeric type 1 ^NDDs³⁸. Indeed, we detected binding between the native pairs by isothermal titration calorimetry (ITC), with affinities in the range of those determined previously for matched pairs of DDs^{27,38–40} (^CDD₄ + ^NDD₅, $K_d = 14.5 \pm 0.9 \mu$ M; ^CDD₈ + ^NDD₉ Met, $K_d = 33 \pm 2 \mu$ M; ^CDD₈ + ^NDD₉ Val, $K_d = 22 \pm 1 \mu$ M) (**Supplementary Fig. 8**). Thus, while stable homodimerization of the ^NDDs may depend on the presence of a downstream homodimeric KS domain, their monomeric character did not preclude interaction with their ^CDD partners. Based on the higher affinity of the interaction, we could identify the ^NDD₉ Val as the physiologically

relevant construct. The observed binding stoichiometry (1 homodimeric CDD_4 :2 monomeric NDD_9), is consistent with the known structure of a type 1a complex in which two monomers of each DD are present (Fig. 1, **Supplementary Fig. 1**)²². As expected, no non-specific interaction was detected between native CDD_4 and NDD_9 , explaining the lack of productive interaction between unmodified subunits Pks4 and Pks9 when the intervening multienzymes are deleted (strain K7N3) (Fig. 3a).

Analysis by ITC of binding between CDD_4 SDM or CDD_4 helix swap and NDD_5 revealed the complete absence of interaction (**Supplementary Fig. 8**), and therefore that the introduced modifications were sufficient to disrupt communication between the native pair. Thus, the continued production of stambomycins **1** by K7N4, CPN4 and K7N5 harboring Pks5 – Pks8 must be due to additional contacts between Pks4 and Pks5 beyond the docking domains, likely including the compatible ACP_{13}/KS_{14} interface. On the other hand, no interaction was detected between CDD_4 SDM and NDD_9 , showing that this limited number of mutations was inadequate to induce productive contacts. This result is fully in accord with the absence of the expected mini-stambomycin products from these strains (K7N1/CPN1, Fig. 3a). By contrast, the CDD_4 helix swap exhibited essentially the same binding to NDD_9 Val as CDD_8 ($K_d = 21.0 \pm 0.3 \mu M$), demonstrating that exchange of just this helix is sufficient to redirect docking specificity³⁰. Thus, inefficient docking is not at the origin of the failure of the CDD_4 helix swaps to yield chain-extended products *in vivo* (strains K7N2/CPN2, Fig. 3a). We could therefore conclude at this stage that the problem arose from the non-native interface generated between ACP_{13} and KS_{21} , poor acceptance by KS_{21} of the incoming substrate during chain transfer and/or chain extension, and/or low activity towards the modified chain of domains/modules acting downstream.

Attempted optimization of the stambomycin DD mutants. We aimed next to improve the novel Pks4/Pks9 intersubunit interface in strain CPN2 (CDD_4 helix swap + deletion of Pks5 – 8) by targeting helix αI of ACP_{13} , as the first 10 residues of this helix have been implicated previously in governing the interaction with the downstream KS domain at hybrid junctions²⁶. Notably, multiple sequence alignment of all ACPs in the stambomycin PKS located at intersubunit junctions, revealed a unique sequence for each ACP in the helix αI region, consistent with a recognition ‘code’ for the KS partner, and the idea that mismatching these contacts might hamper productive chain transfer (**Supplementary Fig. 9**). Indeed, as mentioned previously, even when docking is interrupted, contacts between ACP_{13} and KS_{14} are apparently sufficient to enable chain translocation between Pks4 and Pks5 (Fig. 3a). In addition, an analogous strategy of optimizing the ACP_n/KS_{n+1} chain transfer interface was shown recently to substantially improve interaction between an ACP (JamC) derived from the jamaicamide B biosynthetic pathway, and the first chain extension module of the lipomycin PKS (LipPKS1)⁴¹.

In our case, the first six residues of ACP_{13} helix αI were modified using CRISPR-Cas9 (EADQRR $\diamond$ PSERRQ), so that the full 10-residue recognition sequence matched that of ACP_{20} , the natural partner of KS_{21} (**Supplementary Fig. 9**). Analysis of extracts of the resulting strain CPN2/OE484/ACP13 SDM by HPLC-MS revealed at best minute amounts (highest yield of $0.5 \mu g mL^{-1}$) of target cyclic mini-

stambomycins A/B (**11**), lacking the hydroxyl group introduced by SamR0478 (Fig. 4 and **Supplementary Fig. 9**). Thus, while this experiment finally yielded the first evidence for successful chain transfer between Pks4 and Pks9 followed by subsequent chain extension by Pks9 and TE-catalyzed release, the overall efficiency of the system remained poor. Interestingly, however, the yields of the four shunt metabolites **2 – 5** were as much as 48-fold higher from the ACP₁₃ helix swap mutant than from CPN2/OE484, showing that improved interactions between ACP₁₃ and KS₂₀ facilitated release of the stalled intermediates from ACPs 12 and 13, presumably via remote action by the TEI domain.

Engineering mini-stambomycins by maintaining the native ACP₁₃/KS₁₄ junction (alternative module definition). Cumulatively, the results obtained with the docking domain engineering identified KS₂₁ as one potential bottleneck in the engineered PKS. Our parallel strategy based on the alternative module definition (Fig. 2) allowed us to directly test this idea. Specifically, we investigated the effects of preserving the native ^CDD₄/^NDD₅ pair and either the majority of KS₁₄, or a little more than half of the domain, resulting in a KS₁₄/KS₂₁ hybrid. For this, we used two different splice sites in KS₁₄: i) at the end of the domain in a highly-conserved region (GTNAHV) exploited recently to efficiently swap downstream AT domains⁴²; and, ii) at a site corresponding to a recombination hot spot identified during induced evolution of the rapamycin (RAPS) PKS⁴³, yielding the KS₁₄/KS₂₁ chimera (Fig. 4 and **Supplementary Fig. 10**). Both of these modifications were introduced into *S. ambofaciens* using CRISPR-Cas9, while simultaneously removing Pks5 – Pks8, yielding respectively after co-transformation with pOE484 and the control plasmid pIB139, strains ATCC/OE484/hy59_S1, ATCC/pIB139/hy59_S1, ATCC/OE484/hy59_S2, and ATCC/pIB139/hy59_S2.

Analysis of culture extracts revealed the presence in both ATCC/OE484/hy59_S1 and ATCC/OE484/hy59_S2 relative to the controls, of a novel series of 37-membered metabolites (Fig. 4). The measured masses were consistent with the desired mini-stambomycins either as their free acids or in cyclic form (metabolites **10 – 12**, Fig. 4). Signals corresponding to the A/B and C/D derivatives of all metabolites were detected, providing important evidence for their identities, as well as both the C-14 hydroxylated **12** and non-hydroxylated **11** forms of the cyclic mini-stambomycins (C-14 corresponds to C-28 in the parental compounds (Fig. 1)). It is not surprising that the corresponding E and F forms were not detected, as their yields even from the wild type are much lower than the A – D derivatives (Fig. 3a). The observation of non-hydroxylated **11** shows notably that internal hydroxylation by SamR0478 is not an absolute prerequisite for TE-catalysed macrolactonization, and argues that hydroxylation of the mini-stambomycins only takes place on the macrocyclic compound. Although compounds **11** and **12** incorporate the tetrahydropyran moiety of the parental stambomycins **1** which undergoes glycosylation, derivatives bearing β-mycaminose were not observed, presumably due to poor recognition of the overall modified macrocycle by glycosyl transferase SamR0481²³.

The yields of the target compounds were minor relative to wild type stambomycins (metabolites **10**, **11** and **12** from ATCC/OE484/hy59_S1 were obtained at highest yields of 0.3, 3.2 and 1.0 μg L⁻¹ (4.5 μg L⁻¹ total), respectively), but nonetheless approximately 4-fold higher from ATCC/OE484/hy59_S2

incorporating the hybrid KS₁₄/KS₂₁ than from the full KS₁₄ swap (3.7, 9.7 and 3.6 $\mu\text{g L}^{-1}$ (17 $\mu\text{g L}^{-1}$ total); 1500-fold lower yields than **1**) (Fig. 4 and **Supplementary Fig. 10**). Although the low titers of these compounds precluded their structure elucidation by NMR, we obtained additional confirmatory evidence for their identities by inactivation of *samR0479* (which introduces the hydroxyl used for macrocyclization), which resulted in exclusive production of linear dehydroxy mini-stambomycins **13** (**Supplementary Fig. 10**).

As observed previously, the strains also produced substantial quantities of the shunt products **2 – 5** (inactivation of *samR0479* led correspondingly to the dehydroxy versions of these compounds **6 – 9** (**Supplementary Figs. 7 and 10**)). The yields were ca. 80-fold higher than those of the corresponding mini-stambomycins, with the highest titers observed in the strain incorporating the hybrid KS₁₄/KS₂₁. The amount of shunt metabolites was also approximately 123-fold higher than from strain CPN2/OE484 (which incorporates an ACP₁₃-^CDD₄ swap/^NDD₉-KS₂₁ interface) (Figs. 3a and 4, **Supplementary Table 5**). Thus, contrary to expectation, although using the KS as a fusion site improved communication between Pks4 and Pks9, it also substantially boosted TEI-mediated off-loading of stalled upstream intermediates.

In principle, such stalling could result from a slow rate of chain extension in the now hybrid acceptor module (for example, in the full KS swap construct, KS₁₄ and ACP₂₁ are completely mismatched for chain extension). To evaluate this idea, we modified ACP₂₁ within ATCC/OE484/hy59_S1 incorporating the full-length KS₁₄, targeting a sequence region previously identified as mediating intramodular communication between the KS and ACP during chain extension (**Supplementary Fig. 11**)^{19,26}. Specifically, we exchanged loop 1 and the initial portion of helix all of ACP₂₁ for the corresponding sequence of ACP₁₄, using CRISPR-Cas9 (**Supplementary Fig. 11**). As we anticipated that creation of this substantially hybrid ACP might engender structural perturbation, we also engineered a minimal mutant of ACP₂₁ in which only one of the two most critical residues in the recognition motif was mutated to the corresponding amino acid in ACP₁₄ (G₁₄₉₉ of Pks9 $\diamond$ D; the second residue, R, of the motif is already common to the two ACPs) (**Supplementary Fig. 10**). Analysis of the loop/helix all swap by HPLC-MS showed that all mini-stambomycin production had been abolished (**Supplementary Fig. 10**), consistent with the anticipated disruption to ACP₁₄ structure. Production by the ACP site-directed mutant was not any better than by the full KS swap construct (Fig. 4 and **Supplementary Fig. 10**), as only metabolite **11** remained above the limits of detection.

In principle, the hybrid KS₁₄/KS₂₁ domain may have worked better than KS₁₄ for chain extension due to improved interaction with ACP₂₁, with stalling displaced to later modules. If this were the case, we might expect to see accumulation in the medium of shunt metabolites corresponding to the intermediate generated by module 21. Indeed, in the case of strain hy59_S2 (chimeric KS₁₄/KS₂₁) but not hy59_S1 (KS₁₄), we detected masses consistent with the A/B and C/D forms of intermediate **14** generated by module 21, at yields comparable to those of the final mini-stambomycins (Fig. 4, **Supplementary Fig. 10**) (and correspondingly, **15**, the dehydroxylated analogue of **14**, was detected in the SamR0479 mutant

(**Supplementary Fig. 10**)). The same metabolite **14** was identified from the ACP₂₁ G $\diamond$ D mutant (Fig. 4 and **Supplementary Fig. 11**), consistent with interrupted chain transfer to KS₂₂. Taken together, these data confirm module 22 as a new blockage point in the engineered systems.

Relative efficacy of PKS engineering using PCR-targeting and CRISPR-Cas9. As multiple of our core constructs were generated by both PCR-targeting and CRISPR-Cas9, we were able to directly compare the efficiency of the two techniques (Fig. 3 and **Supplementary Fig. 4**). Globally, our results confirm that both approaches can be employed to introduce large-scale modifications to PKS biosynthetic genes (i.e. deletions of single or multi-gene regions)^{32,44–46}. We have also demonstrated, for only the second time to our knowledge, that CRISPR-Cas9 can be leveraged to specifically modify modular PKS domains⁴⁷. Of the two methods, CRISPR-Cas9 was the more rapid, as the corresponding constructs were engineered in approximately half of the time. In addition, while CRISPR-Cas9 allowed for direct modification of the host genome, PCR-targeting relied on the availability of suitable cosmids housing the target genes, and resulted in a 33 bp *attB*-like ‘scar’ sequence in the genome (**Supplementary Fig. 4**)⁴⁸. In addition to hampering iterative use of the approach, the scar apparently provoked a moderate reduction in stambomycin yields in mutant K7N6 compared to the wild type, an effect also noted upon comparison of several analogous mutant strains (e.g. K7N4 vs. CPN4, Fig. 3). Nonetheless, we did encounter certain difficulties with use of CRISPR-Cas9 (i.e. failure to obtain construct CPN3, occasional reversions to wild type, etc.), observations motivating ongoing efforts in other laboratories to further enhance the suitability of CRISPR-Cas9 for editing PKS pathways^{47,49–54}.

Discussion

In this work, we have utilized an approach based on the state-of-the-art in PKS engineering to modify the stambomycin PKS (Fig. 5). Specifically, we aimed to remove the four PKS subunits between Pks4 and Pks9 in the assembly line which together house seven chain extension modules, to generate a series of 37-membered ‘mini-stambomycins’. While in principle such a change might have been possible by directly fusing Pks4 and Pks9 via a suitable intermodular linker, this approach would have resulted in a heptamodular subunit whose size is far in excess of the tetramodular multienzymes present in the system. We have also demonstrated recently the low efficacy of this strategy when the module downstream of the linker is N-terminal in its native subunit context (as with module 21 of Pks9)¹⁸.

As an initial approach (Fig. 5), we modified ^CDD₄ to render it compatible with ^NDD₉, with the aim of inducing productive communication between Pks4 and Pks9, while leaving all modular units intact. This modified PKS relied for function on both a non-native chain transfer interface (ACP₁₃/KS₂₁), and the intrinsic tolerance of the downstream KS/modules to the incoming substrate. We were optimistic this experiment might work given the structural similarities between the native substrates of KS₁₄ and KS₂₁ at least directly adjacent to the acyl terminus, as well as the fact that the stambomycin PKS generates a small family of metabolites, and therefore must exhibit some intrinsic tolerance to structural variation. Although we showed *in vitro* with recombinant DD pairs that a docking helix-swapped mutant of ^CDD₄

communicated effectively with ^NDD₉, chain transfer across the engineered interface did not occur *in vivo*, as evidenced by the accumulation of multiple shunt products. While our attempt to render the ACP₁₃/KS₂₁ junction more native by site-directed mutagenesis did result in certain target metabolites, the most significant effect was to increase the yields of the truncated chains.

Having narrowed down the biosynthetic block to events occurring downstream of the engineered junction, we next carried out interface engineering based on proposed alternative module boundaries, leveraging fusion points within the KS domain (Fig. 5). In this case, sites were selected to either maintain essentially the whole of KS₁₄, or to create a hybrid KS₁₄/KS₂₁ domain. This strategy at once preserved key elements of the ACP₁₃/KS₁₄ chain transfer junction, and in the case of the almost full-length KS₁₄, ensured that the domain had the appropriate substrate specificity for the incoming chain. Interestingly, the construct incorporating the chimeric KS functioned best, producing the desired mini-stambomycins in both linear and macrocyclic forms. These data identify this location in the middle of the KS⁴³ as a potentially general fusion site, perhaps because it preserves elements of both the chain transfer and chain extension interfaces with the two partner ACP domains (ACP₁₃ and ACP₂₁, in this case). Intriguingly, a site within the analogous condensation (C) domains of modular nonribosomal peptide synthetase (NRPS) systems, has also recently emerged as a useful fusion point for generating engineered hybrids⁵⁵.

Our results also showcase the intrinsically high tolerance of the Pks9 TEI domain towards shorter substrates. Indeed, the data also demonstrate that this TEI domain participates in off-loading the shunt metabolites from the upstream subunit, and that this activity interferes with passage of the chain to subsequent modules. Unfortunately, our attempts to boost yields of the mini-stambomycins by engineering the condensation interface between KS₁₄ and ACP₂₁ were unsuccessful, both when the full ACP₂₁ recognition loop/helix α II region was swapped for that of ACP₁₄, and when a single site-directed mutation was made at a putatively critical position (Fig. 5). This result is surprising in light of the beneficial effects reported *in vitro* of both of these modifications on chain extension carried out by mismatched KS and ACP domains sourced from the erythromycin PKS (DEBS)²⁶. Apparently, the introduced changes were not sufficient to ensure effective communication between the KS₁₄ and ACP₂₁ domains (or were in fact deleterious to function), and/or any benefit was masked by the poor tolerance of the downstream modules to the modified intermediates.

To fully judge the efficacy of this work, it is instructive to compare it to the other two examples in the literature in which full biosynthetic systems have been re-engineered to remove multiple internal modules¹⁰. In the first, recently-reported case, the neoauerothin (Nor) hexamodular PKS was 'morphed' into the evolutionarily-related auerothin (Aur) tetramodular PKS by removing the second bimodular subunit, NorA'. As in our work, the authors initially attempted to engineer a new interaction between the monomodular subunits NorA and NorB flanking NorA' using compatible docking domains, by exchanging the type 1b ^NDD of NorB for the type 1a ^NDD of NorA' (the natural partner of NorA). When the target metabolite was not obtained, they relocated the fusion site to the KS-AT linker downstream of the conserved KS region in NorB, thereby maintaining the native NorA ACP-^CDD/^NDD-KS NorA' junction.

Ultimately, several linker variants had to be evaluated before a functional sequence was identified, in part by serendipity (indeed it is 1 residue longer than the native linker). Overall, the yields of the targeted chain-shortened metabolites were reduced approximately 18-fold compared to the parental neoareothin (to ca. 2.5 mg L^{-1}), a much less significant penalty than engendered by our engineering strategy. Presumably, the superior titers obtained in this experiment reflect the much higher intrinsic amenability of the Nor PKS to conversion into an Aur PKS, as the Nor PKS likely evolved from an Aur PKS by subunit insertion¹⁰. Nevertheless, the newly-created NorA/NorB interface was also only partially functional, as product corresponding to the intermediate generated by iterative action of the upstream subunit NorA was still obtained.

The second relevant investigation concerns the accelerated evolution (AE) of the RAPS PKS, based on spontaneous induced homologous recombination between its component modules⁴³. As mentioned earlier, several of the resulting systems incorporated intermodular fusion sites essentially at the mid-point of the respective KS domains, and so can be compared to our best performing construct hy59_S2. Notably, yields from the hybrid RAPS PKSs from which either 3 or 6 modules were removed, were reduced by a maximum of 33% relative to that of the parental compound. We propose two explanations for the higher functionality of these systems relative to hy59_S2. First of all, in every case, the module downstream of the newly-formed junctions in the contracted RAPS PKSs was internal to its respective subunit (unlike module 21 of Pks9), consistent with the idea that such modules boast intrinsically broader substrate specificity¹⁸. Secondly, the KS domains of the RAPS PKS exhibit unusually high mutual sequence identity (85 – 90%)⁵⁶. This means, in effect, that the same KS domain acts on a large variety of substrates of differing length and functionality, contributing to the tolerance of all modules downstream of the fusion site. In contrast, the KSs in the stambomycin PKS have substantially lower sequence identity (62 – 80%), and thus KSs 22 – 24 in hy59_S2 likely represent a specificity barrier to efficient transfer and extension of the modified intermediates.

Taken together, this set of results shows that contracting PKS systems represents a viable approach to accessing truncated polyketide derivatives of variable length, including macrocycles. Whether such systems are generated rationally or using an AE process, the most efficient hybrids will likely result: i) from PKSs whose modules (and in particular KS domains) exhibit a substantial degree of mutual sequence identity and thus intrinsically high substrate tolerance (or which can be adapted by mutagenesis to broaden their specificity²⁹); and, ii) when novel junctions are created with downstream modules which are situated at internal positions within their subunits. The data also reinforce the idea that in cases where communication at modified interfaces occurs via non-covalent protein-protein interactions, at least a portion of the KS downstream from the docking domains should be included to boost efficiency^{10,11,16}. Finally, our work has identified an increase in TEI-mediated proof-reading provoked by such interface engineering. Elucidating the mechanism underlying this unexpected intersubunit release activity, and thus how to effectively suppress it, should be a profitable avenue for further boosting product titers.

Methods

Bioinformatics analysis. To underpin the interface engineering strategy, the extremities of all the stambomycin PKS subunits were analyzed to identify the boundaries of the most C-terminal and N-terminal function domains (ACP and KS, respectively), and thus the regions potentially containing docking domains (DDs). The resulting sequences were compared by multiple sequence alignment using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo>)⁵⁷, to bonafide and putative DD sequences from multiple DD classes, including those present at the DEBS 2/DEBS 3 interface (type 1a, PDB ID:1PZQ, 1PZR²²) and the PikAIII/PikAIV junction (type 1b, PDB ID: 3F5H²⁷), to allow for type classification. To identify suitable boundaries for DD heterologous expression in *E. coli*, the secondary structure of the putative DD regions was predicted using PSIPRED 4.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>)³⁷. Analysis for potential specificity-conferring residues in the stambomycin PKS ketosynthase (KS) domains was carried out by multiple sequence alignment against model KS domains^{14,20,29}, using Clustal Omega⁵⁷.

General methods. All reagents and chemicals were obtained from Sigma-Aldrich, except the following: BD (tryptone, yeast extract, TSB powder), Thermo Fisher Scientific (Tris), VWR (glycerol, NaCl, NaNO₃), ADM, France (NutriSoy flour), and New England Biolabs (T4 DNA ligase, restriction enzymes). Oligonucleotide primers and two additional synthetic DNA fragments for CPN4 and CPN5 constructs were synthesized by Sigma-Aldrich (**Supplementary Table 2**). The docking domains ^NDD₉ Val and ^NDD₉ Met (**Supplementary Table 1**) were obtained as synthetic peptides from GeneCust. DNA sequencing of PCR products was performed by Sigma-Aldrich and Eurofins.

PCR reactions were performed with Taq DNA polymerase (Thermo Fisher Scientific) or Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) when higher fidelity was required. Isolation of DNA fragments from agarose gel, purification of PCR products and extraction of plasmids were carried out using the NucleoSpin® Gel and PCR Cleanup or NucleoSpin® Plasmid DNA kits (Macherey Nagel, Hoerd, France).

Strains and media. *E. coli* BL21 strains were obtained from Novagen. Unless otherwise specified, all *E. coli* strains were cultured in LB medium (yeast extract 10 g, tryptone 5 g, NaCl 10 g, distilled water up to 1 L, pH 7.0)⁵⁸ or on LB agar plates (LB medium supplemented with 20 g L⁻¹ agar) at 37°C. *Streptomyces ambofaciens* ATCC23877 and the derived mutants were grown in TSB (TSB powder 30 g (tryptone 17 g, soy 3 g, NaCl 5 g, K₂HPO₄ 2.5 g, glucose 2.5 g), distilled water up to 1 L, pH 7.3) or on TSA plates (TSB medium supplemented with 20 g L⁻¹ agar), and sporulated on SFM⁵⁹ agar plates (NutriSoy flour 20 g, D-mannitol 20 g, agar 20 g, tap water up to 1 L) at 30°C. All strains were maintained in 20% (v/v) glycerol in 2 mL Eppendorf tubes and stored at -80°C.

For fermentation of *S. ambofaciens* ATCC23877 and its mutants, spores were streaked on TSA with appropriate antibiotics and after incubation 48 h at 30°C, a loop of mycelium was used to inoculate 7 ml of MP5 medium (yeast extract 7 g, NaCl 5 g, NaNO₃ 1 g, glycerol 36 mL, MOPS 20.9 g, distilled water up to 1 L, pH 7.4) supplemented with selective antibiotics and sterile glass beads, followed by incubation at

30°C and 200 rpm for 24 – 48 h. Finally, the seed culture was centrifuged and resuspended into 2 mL fresh MP5 before being inoculated into 50 mL MP5 medium in a 250 mL Erlenmeyer flask, and cultivated at 200 rpm and 30°C for 4 days.

PCR-targeting-based genetic engineering (exemplified by mutant ATCC/OE484/K7N1, **Supplementary Fig. 4**). To render the BAC BAA9ZA8 proficient for selection following conjugation, its chloramphenicol resistance gene was replaced using a PCR-targeting approach, by a kanamycin resistance gene cassette sourced from pIJ776³¹, resulting in BAC1 (**Supplementary Table 3**). The cassette *attL + aac(3)IV + oriT + attR* was amplified from the plasmid pSPM88T⁶⁰ (**Supplementary Fig. 4**) using primers 9996 and 9997 (**Supplementary Table 2**), affording PCR amplicon *PCR-K7N1*. The PCR fragment was then electro-transformed into *E. coli* BW25113/pKD20/BAC1⁶¹, giving rise to mutant BAC1_K7N1 (in which the C-terminus of *pks4* was modified and the genes *pks5 – pks8*, were deleted)³¹. The BAC1_K7N1 was then introduced into *E. coli* ET12567/pUZ8002⁶² and then transferred to *S. ambifaciens* wild type via intergeneric conjugation. The resulting exconjugants (ATCC/K7N1_ *aac(3)IV + oriT*) were selected for their apramycin resistance and kanamycin sensitivity (i.e. a phenotype consistent with successful double cross-over). The correct mutations were confirmed by PCR and sequencing. Subsequently, the disruption cassette was excised using the excisionase and integrase of pSAM2 encoded by pOSK111 as described⁶⁰, leaving a 33 bp ‘scar’ sequence (mutant ATCC/K7N1). Successful removal of the cassette was verified by PCR and DNA sequencing. Finally, the LAL regulator overexpression plasmid pOE484²³ or the parental vector pIB139³³ was introduced into the strain giving rise to mutants K7N1/OE484 and K7N1/pIB139, respectively. The same overall procedure was applied to construct mutants K7N2/OE483, K7N2/pIB139, K7N3/OE483, K7N3/pIB139, K7N4/OE483, K7N4/pIB139, K7N5/OE483, K7N5/pIB139, K7N6/OE483 and K7N6/pIB139. An analogous PCR-targeting approach (**Supplementary Fig. 4**) was also employed to inactivate *samR0478* and *samR0479* using appropriate BACS³⁴ (**Supplementary Table 3**).

CRISPR-Cas9-mediated genetic engineering. Plasmids pCRISPomyces-2 (and associated cloning and screen protocols)³² (used for construction of all mutants except ATCC/hy59_S1 and ATCC/hy59_S2) and pCRISPR-Cas9⁴⁵, respectively, were used for CRISPR-Cas9-based genome editing, as described previously. The two systems differ in the way in which Cas9 is expressed; in the case of pCRISPomyces-2, the nuclease is expressed constitutively, while in the pCRISPR-Cas9 system, its expression is under inductive control by thiostrepton (Tsr). The crRNA sequence was selected to match the DNA segment which contains NGG on its 3’ end (N is any nucleotide, and the NGG corresponds to the protospacer-adjacent motif (PAM)). The annealed crRNA fragment and two homologous arms (HAL and HAR, flanking the target region) were sequentially inserted into the delivery plasmid pCRISPomyces-2 using the respective restriction sites *BbsI* and *XbaI*, to afford the specific recombinant plasmid for each mutant (**Supplementary Fig. 4**). Correspondingly, an sgRNA cassette (tracrRNA + sgRNA) and two homologous arms were inserted into the plasmid pCRIPR-Cas9 using sites *NcoI*, *SnaBI* and *StuI*, respectively (**Supplementary Fig. 10**). In addition, the crRNA was designed to be located within the region to be deleted (**Supplementary Fig. 4**) to avoid Cas9-catalyzed cleavage occurring in the genome of the resulting mutant. In the case of site-directed mutants, additional DNA fragments containing the targeted mutations

were inserted between the two homologous arms. In addition, the DNA sequence with the fragments identical to the crRNA was modified, so as to avoid subsequent Cas9-catalyzed cleavage of the obtained mutants (**Supplementary Figs. 9 and 11**).

Overexpression and purification of docking domains. The wild-type docking domains (CDD_4 , NDD_5 , CDD_8 , NDD_9 Val and NDD_9 Met) and mutant docking domains (CDD_4 SDM, CDD_4 helix swap) were amplified from genomic DNA of *S. ambofaciens* wild type and the relevant mutants, using forward and reverse primers incorporating *Bam*HI and *Hind*III restriction sites, respectively (**Supplementary Table 2**). The PCR amplicons were digested using FD *Bam*HI and FD *Hind*III, and then ligated into the equivalent sites of vector pBG-102 (Center for Structural Biology, Vanderbilt University). In the case of all CDD s which lacked aromatic residues, a tyrosine residue (codon TAT incorporated in the forward primer, **Supplementary Table 1**) was added at the N-terminal ends (so as not to interfere with docking with the NDD partner) to allow efficient monitoring by UV-Vis during the purification, as well as reliable measurement of protein concentration necessary for binding studies by ITC.

The resulting constructs pBG102- NDD_5 , pBG102- CDD_8 and pBG102- NDD_9 were used to transform *E. coli* BL21 (DE3). For CDD_4 and its mutants, these were transformed into Rosetta™ 2(DE3), as these constructs contain 8 codons rarely used in *E. coli*. Positive transformants were selected on LB agar supplemented with kanamycin ($50 \mu\text{g mL}^{-1}$) ($25 \mu\text{g mL}^{-1}$ chloramphenicol was also added for expression in Rosetta™ 2(DE3)). A single colony was transferred to LB (10 mL) supplemented with antibiotics, and the culture grown at 37°C and 200 rpm for overnight. The 1 mL overnight culture was used to inoculate LB media (1 L) supplemented with appropriate antibiotics, and then incubated at 37°C and 200 rpm to an optical density of 0.8, at which point protein synthesis was induced by the addition of IPTG (final concentration 0.1 mM). After incubation at 18°C and 200 rpm for 18 h, cells were collected by centrifugation at 8000g for 30 min, resuspended in 40 mL protein purification buffer A (50 mM Tris-HCl, 400 mM NaCl, 10 mM imidazole, pH 8.0), and lysed by sonication. Following centrifugation at 20000g and filtration using a 0.45 μm membrane, the soluble cell lysates were loaded onto 2 · 5 mL HisTrap HP (GE Healthcare) columns (two 5 mL columns in series) equilibrated in buffer A, and purified by preparative protein purification chromatography using an ÄKTA Avant system. The following program was applied: sample loading, 1 mL min^{-1} ; washing, 2 mL min^{-1} , 10 column volumes of buffer A; elution, 2 mL min^{-1} , 5 column volumes of buffer B (50 mM Tris-HCl, 400 mM NaCl, 250 mM imidazole, pH 8.0); elution, 2 mL min^{-1} , 2 column volumes of buffer C (50 mM Tris-HCl, 400 mM NaCl, 500 mM imidazole, pH 8.0).

All His₆-SUMO-tagged proteins were collected (fractions containing the protein of interest were selected based on the UV chromatography and SDS-PAGE gel), and transferred into dialysis bag containing His₆-tagged human rhinovirus 3C protease (H3C) (1 – 2 μM). The dialysis bag was then placed into a container filled with buffer D (50 mM Tris-HCl, 400 mM NaCl, pH 8.0), and the cleavage allowed to proceed at 4°C overnight. The resulting proteins, which incorporated a non-native N-terminal GPGS sequence, were then separated from the remaining His₆-tagged SUMO and His₆-tagged human rhinovirus 3C protease by re-loading onto the 2 · 5 mL HisTrap HP columns pre-equilibrated in buffer A. Purification was then carried

out with the following program: sample loading, 1 mL min⁻¹; washing, 2 mL min⁻¹, 4 column volumes of buffer A; elution, 2 mL min⁻¹, 2 column volumes of buffer B; elution, 2 mL min⁻¹, 2 column volumes of buffer C. The untagged docking domains passed through the column during the washing step, and were collected and concentrated to 5 – 7 mL using an Amicon Ultra 3000 MWCO centrifuge filter (Millipore Corp).

Subsequently, the concentrated docking domains were loaded onto a size exclusion chromatography column (Superdex 75 26/60 column, GE Healthcare) equilibrated in buffer GF (20 mM HEPES, 100 mM NaCl, 0.5 mM TCEP, pH 7.5). Following a concentration step, the purity of the purified proteins was verified by SDS-PAGE, and their concentrations were determined by NanoDrop (Thermo Scientific) with extinction coefficients calculated using the ExPASy ProtParam tool⁶³.

Isothermal Titration Calorimetry measurements. ITC measurements were performed at 20°C in buffer GF using a MicroCal ITC200 (Malvern Instruments) (A2F Plateforme ASIA: Approches fonctionnelles et Structurales des InterActions cellulaires). A 300 µL aliquot of ^NDD₅ at 70 µM was placed in the calorimeter cell and titrated with 700 µM of the ^CDD₄s (^CDD₄ wild type, ^CDD₄ SDM and ^CDD₄ helix swap) in the syringe. In the case of the binding experiments between ^NDD₉ Met and ^CDD₈, the ^CDD₈ (700 µM) was added to the ^NDD₉ Met (80 µM in the cell), while for the binding between ^NDD₉ Val and the ^CDDs (^CDD₈, ^CDD₄ wild type, ^CDD₄ SDM and ^CDD₄ helix swap), the ^CDDs (700 µM) were added to ^NDD₉ Val in the cell (120 µM). The ITC experiments were then carried out as followed: initial waiting time 120 s, initial injection of 0.5 µL over 1 s followed by 19 serial injections of 2 µL over 4 s, separated by an interval of 120 s. For each experiment, the reference power was set to 5 µcal⁻¹, stirring speed to 750 rpm, and the high feedback mode was selected. Two independent titrations were performed for each combination of DDs. The heat of reaction per injection (µcal s⁻¹) was determined by integration of the peak areas using the Origin 7.0 (OriginLab) software, assuming a one-site binding model (consistent with the solved structures of the types of DDs^{22,27}), yielding the best-fit values for the heat of binding (ΔH), the stoichiometry of binding (N) and the dissociation constant (K_d). The heats of dilution of the DDs were determined by injecting them into the cell containing buffer only, and these were subtracted from the corresponding binding data prior to curve fitting.

In some cases, when a plateau (binding saturation) was not reached at the final titration step, and the problem could not be solved by increasing the concentration of DD in syringe, we initially placed ^CDD/^NDD complex in the ITC cell (at the concentration of the two partners reached in the previous titration), filled the syringe with additional DD, and performed a second titration experiment. This procedure was then repeated until binding saturation was reached. To fit the data, the MicroCal ITC concatenation software was used to combine two ITC data files together. Most importantly, the critical parameter dimensionless constant (C-value) was calculated as follows:

$$C = NK_a[M]_T$$

where K_a is the binding constant, $[M]_T$ is the total macromolecular concentration in the cell, and N is the stoichiometry of interaction. A reliable ITC binding isotherm is evidenced by ITC data with C -values > 1 (the optimal range is $5 < C < 500$)⁶⁴, as was the case for all of our measurements.

Circular Dichroism measurements. CD spectra were recorded on a Chirascan CD (Applied Photophysics, United Kingdom) (IBS-Lor UMS 2008 Plateforme de Biophysique et Biologie Structurale) at 0.5 nm intervals in the wavelength range of 180 – 260 nm at 20°C, using a temperature-controlled chamber. A 0.01 cm quartz cuvette containing 30 μ L of docking domain at 100 μ M, a 0.1 cm cuvette with 200 μ L of sample at 10 μ M, and a 1 cm cuvette containing 1.5 mL of sample at 1 μ M, were used for all the measurements. All measurements were performed at least in triplicate, and sample spectra were corrected for buffer background by subtracting the average spectrum of buffer alone. The CD spectra were deconvoluted using the deconvolution software CDNN2.1⁶⁵ to estimate the secondary structure present in the docking domains.

SEC-MALS analysis of docking domains. The oligomeric state of all the docking domains was determined by size-exclusion chromatography multi-angle light scattering (SEC-MALS) on the A2F Plateforme ASIA. For this, SEC was first carried out on a Superdex75 10/300 column (GE Healthcare) at 20°C using a flow rate of 0.5 mL min⁻¹ in HEPES buffer (20 mM HEPES, 100 mM NaCl, 0.5mM TCEP, pH 7.5) using a ÄKTA-Purifier FPLC (GE Healthcare). Multi-angle light scattering (MALS) was measured using a MiniDAWN TREOS II (Wyatt Technology), while refractometry was monitored using an Optilab T-rEX (Wyatt Technology). Data processing was carried out with the manufacturer supplied software (ASTRA 6.1, Wyatt Technology) to determine the protein oligomerization state.

HPLC-MS analysis of fermentation metabolites and purified docking domains. The fermentation broth of *Streptomyces* was centrifuged at 4000g for 10 min. As described previously²³, the stambomycins and their derivatives were then extracted from the mycelia, by first resuspending the cells in 40 mL distilled water, followed by centrifugation (4000g, 10 min, repeated 3x) to remove water-soluble components. After decanting the water, the cell pellets were weighed and extracted with methanol by shaking at 150 rpm for 2 h at room temperature. Thereafter, the methanol extracts were filtered to remove the cell debris, followed by rotary evaporation to dryness. The obtained extracts were then dissolved in methanol, whose volume was determined according to the initial weight of the mycelia (70 μ L methanol to 1 g of initial cell pellet). The resulting mycelial crude extracts were then passed through a 0.4 μ m syringe filter and analyzed in positive electrospray mode (ESI⁺) by HPLC-HRMS (Thermo Scientific Orbitrap LTQ XL or an Orbitrap ID-X Tribrid Mass Spectrometer) (Plateau d'Analyse Structurale et Métabolomique (PASM)-SF4242 EFABA) using an Alltima™ C18 column (2.1 · 150mm, 5 μ m particle size). Separation was carried out with Milli-Q water containing 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B) using the following elution profile: 0 – 48 min, linear gradient 5 – 95% solvent B; 48 – 54 min, constant 95% solvent B; 54 – 60 min, constant 5% solvent B. Mass spectrometry operating parameters were: spray voltage, 5 kV; source gases were set respectively for sheath gas, auxiliary gas and sweep gas at 30, 10 and 10 arbitrary units min⁻¹; capillary temperature, 275°C; capillary voltage, 4 V; tube lens, split lens and front lens voltages 155

V, -28 V and -6 V, respectively. Due to the much lower sensitivity of the Orbitrap LTQ XL relative to the Orbitrap ID-X Tribrid as evidenced by comparative analysis of identical samples on the two instruments, we introduced a 10× correction factor to the the yields determined using the Orbitrap LTQ XL (**Supplementary Table 5**).

The purified docking domains in buffer GF were diluted with Milli-Q water to a concentration of 50 µM and injected onto an Alltima™ C18 column (2.1 × 150mm, 5 µm particle size). Analysis was carried out with Milli-Q water containing 0.1% TFA (A) and acetonitrile containing 0.1% TFA (B), using the elution profile: 0 – 15 min, constant 10% solvent B; 15 – 20 min, linear gradient of 10% solvent B to 95%; 20 – 25 min, constant 10% solvent B. Mass spectrometry operating parameters were set as above.

Metabolite profiling of engineered strains. Comparative analysis of fermentation extracts (of all strains containing pOE484 except K7N3, CPN4 and CPN5, relative to control mutant containing empty plasmid pIB139) was conducted on a Dionex Ultimate 3000 HPLC system coupled to a Thermo Scientific™ Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer. MS-settings: spray voltage 3.5 kV; capillary temperature 320°C; sheath gas (52.50), auxiliary gas (13.75), sweep gas (2.75); probe heater 437.50°C; S-Lens RF (50), positive mode, resolution 70.000; AGC target 1e6, microscans 1, maximum IT 75 ms, scan range 200 – 1800 *m/z*. Chromatographic separation was obtained using a Phenomenex Kinetex 2.6 µm XB-C18 150 × 4.6 mm column with solvents (A, H₂O + 0.1% formic acid) and (B, MeCN + 0.1% formic acid) and the following gradient: flow rate 0.7 mL min⁻¹, 20% B for 2 min, 20–98% B over 18 min, 98% B for 5 min, 98–20% B in 0.5 min and 20% B for 4 min. Metabolic differences within the obtained data (**Supplementary Table 6**) were identified using SIEVE 2.0 screening software, applying the default settings for component extraction of small molecules, except that of the base peak minimum intensity, which was set to 5000000.

Quantification of metabolites. To quantify the yields of native stambomycins and the newly-generated derivatives by HPLC-MS, we generated a calibration curve with previously-purified stambomycin A/B **1** as the standard (using a concentration range between 1 µg L⁻¹ and 50 mg L⁻¹). This approach yielded a linear correlation between the quantity of **1** and the respective peak area in the extracted ion chromatogram (EIC) (the areas of all peaks corresponding to the parental ions were used) (**Supplementary Fig. 12, Supplementary Table 5**). The titers of all stambomycin derivatives were then determined using this calibration curve, based on the peak areas corresponding to the parental ions in their respective EIC chromatograms (**Supplementary Table 5**). As the limit of detection in these experiments lay between 1 and 10 µg L⁻¹, yields in this range must be viewed as estimates.

Declarations

Acknowledgements

We acknowledge financial support from the Agence Nationale de la Recherche (grant number ANR-16-CE92-0006-01, PKS STRUCTURE to K.J.W.), the Université de Lorraine, the Centre National de la Recherche

Scientifique (CNRS), and the IMPACT Biomolecules project of the Lorraine Université d'Excellence (Investissements d'avenir - ANR 15-004). J.P. acknowledges funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 742739). S. Collin and B. Chagot are thanked for help with the analysis of the DDs *in vitro*, A. Kriznik for assistance with the biophysical experiments, J. Grosjean for help with the HPLC-MS analysis and S. Rousselot for contributing to developing the genetic engineering strategy.

Author contributions

L.S. constructed plasmids, generated and fermented recombinant strains, analyzed and interpreted HPLC-MS data, expressed and purified recombinant docking domains, carried out biophysical analysis of DD interactions, and generated all of the manuscript figures. L.H. and C.J. constructed certain plasmids, mutant BACs and/or mutant strains. C.P. performed HPLC analyses. A.B. and J.P. carried out HPLC-MS analyses and metabolic profiling. B.A., C.J. and K.J.W. designed the research and supervised the project. K.J.W. performed *in silico* analyses and wrote the manuscript, with contributions from all authors.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/>

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

References

1. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 660–670 (2015).
2. Keatinge-Clay, A. T. The uncommon enzymology of *cis*-acyltransferase assembly lines. *Chem. Rev.* **117**, 5334–5366 (2017).
3. Weissman, K. J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat. Prod. Rep.* **33**, 203–230 (2016).
4. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).
5. Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564 (2014).
6. Hans, M., Hornung, A., Dziarnowski, A., Cane, D. E. & Khosla, C. Mechanistic analysis of acyl transferase domain exchange in polyketide synthase modules. *J. Am. Chem. Soc.* **125**, 5366–5374 (2003).
7. Ranganathan, A. *et al.* Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: a route to simple statin analogues. *Chem. Biol.* **6**, 731–741

- (1999).
8. McDaniel, R., Kao, C. M., Hwang, S. J. & Khosla, C. Engineered intermodular and intramodular polyketide synthase fusions. *Chem. Biol.* **4**, 667–674 (1997).
 9. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171–1176 (2005).
 10. Peng, H., Ishida, K., Sugimoto, Y., Jenke-Kodama, H. & Hertweck, C. Emulating evolutionary processes to morph aureothin-type modular polyketide synthases and associated oxygenases. *Nat. Commun.* **10**, 3918 (2019).
 11. Miyazawa, T., Hirsch, M., Zhang, Z. & Keatinge-Clay, A. T. An *in vitro* platform for engineering and harnessing modular polyketide synthases. *Nat. Commun.* **11**, 80 (2020).
 12. Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).
 13. Wu, N., Tsuji, S. Y., Cane, D. E. & Khosla, C. Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J. Am. Chem. Soc.* **123**, 6465–6474 (2001).
 14. Zhang, L. *et al.* Characterization of giant modular PKSs provides insight into genetic mechanism for structural diversification of aminopolyol polyketides. *Angew. Chem. Int. Ed Engl.* **56**, 1740–1745 (2017).
 15. Keatinge-Clay, A. T. Polyketide synthase modules redefined. *Angew. Chem. Int. Ed Engl.* **56**, 4658–4660 (2017).
 16. Chandran, S. S., Menzella, H. G., Carney, J. R. & Santi, D. V. Activating hybrid modular interfaces in synthetic polyketide synthases by cassette replacement of ketosynthase domains. *Chem. Biol.* **13**, 469–474 (2006).
 17. Cortes, J. *et al.* Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268**, 1487–1489 (1995).
 18. Massicard, J.-M., Soligot, C., Weissman, K. J. & Jacob, C. Manipulating polyketide stereochemistry by exchange of polyketide synthase modules. *Chem. Commun.* **56**, 12749–12752 (2020).
 19. Kapur, S., Chen, A. Y., Cane, D. E. & Khosla, C. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22066–22071 (2010).
 20. Klaus, M. *et al.* Protein-protein Interactions, not substrate recognition, dominate the turnover of chimeric assembly line polyketide synthases. *J. Biol. Chem.* **291**, 16404–16415 (2016).
 21. Klaus, M., Buyachuihan, L. & Grninger, M. The ketosynthase domain constrains the design of polyketide synthases. *ACS Chem. Biol.* **15**, 2422–2432 (2020).
 22. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).

23. Laureti, L. *et al.* Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6258–6263 (2011).
24. Aigle, B., Challis, G., Laureti, L., Song, L. & Leblond, P. Stambomycin and derivatives, their production and their use as drugs. Patent WO2011009938A2 (2011).
25. Wang, S., Dong, G. & Sheng, C. Structural simplification of natural products. *Chem. Rev.* **119**, 4180–4220 (2019).
26. Kapur, S. *et al.* Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4110–4115 (2012).
27. Buchholz, T. J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4**, 41–52 (2009).
28. Nguyen, T. *et al.* Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–233 (2008).
29. Murphy, A. C., Hong, H., Vance, S., Broadhurst, R. W. & Leadlay, P. F. Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *Chem. Commun.* **52**, 8373–8376 (2016).
30. Weissman, K. J. The structural basis for docking in modular polyketide biosynthesis. *ChemBioChem* **7**, 485–494 (2006).
31. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1541–1546 (2003).
32. 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).
33. Wilkinson, C. J. *et al.* Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.* **4**, 417–426 (2002).
34. Song, L. *et al.* Cytochrome P450-mediated hydroxylation is required for polyketide macrolactonization in stambomycin biosynthesis. *J. Antibiot.* **67**, 71–76 (2014).
35. Heathcote, M. L., Staunton, J. & Leadlay, P. F. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **8**, 207–220 (2001).
36. Kittendorf, J. D., Beck, B. J., Buchholz, T. J., Seufert, W. & Sherman, D. H. Interrogating the molecular basis for multiple macrolactone ring formation by the pikromycin polyketide synthase. *Chem. Biol.* **14**, 944–954 (2007).
37. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
38. Risser, F. *et al.* Towards improved understanding of intersubunit interactions in modular polyketide biosynthesis: Docking in the enacyloxin IIa polyketide synthase. *J. Struct. Biol.* **212**, 107581 (2020).

39. Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340–1351 (2013).
40. Dorival, J. *et al.* Characterization of intersubunit communication in the virginiamycin *trans*-acyl transferase polyketide synthase. *J. Am. Chem. Soc.* **138**, 4155–4167 (2016).
41. Porterfield, W. B., Poenateetai, N. & Zhang, W. Engineered biosynthesis of alkyne-tagged polyketides by type I PKSs. *iScience* **23**, 100938 (2020).
42. Yuzawa, S. *et al.* Comprehensive *in vitro* analysis of acyltransferase domain exchanges in modular polyketide synthases and its application for short-chain ketone production. *ACS Synth. Biol.* **6**, 139–147 (2017).
43. Wlodek, A. *et al.* Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nat. Commun.* **8**, 1206 (2017).
44. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin.* **47**, 231–243 (2015).
45. 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).
46. Zeng, H. *et al.* Highly efficient editing of the actinorhodin polyketide chain length factor gene in *Streptomyces coelicolor* M145 using CRISPR/Cas9-CodA(sm) combined system. *Appl. Microbiol. Biotechnol.* **99**, 10575–10585 (2015).
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. Myronovskyi, M., Rosenkränzer, B. & Luzhetskyy, A. Iterative marker excision system. *Appl. Microbiol. Biotechnol.* **98**, 4557–4570 (2014).
49. Mo, J. *et al.* Efficient editing DNA regions with high sequence identity in actinomycetal genomes by a CRISPR-Cas9 system. *Synth. Syst. Biotechnol.* **4**, 86–91 (2019).
50. Tong, Y. *et al.* Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20366–20375 (2019).
51. Zhao, Y., Li, G., Chen, Y. & Lu, Y. Challenges and advances in genome editing technologies in *Streptomyces*. *Biomolecules* **10**, (2020).
52. Alberti, F. & Corre, C. Editing streptomycete genomes in the CRISPR/Cas9 age. *Nat. Prod. Rep.* **36**, 1237–1248 (2019).
53. Wang, Q. *et al.* Dual-function chromogenic screening-based CRISPR/Cas9 genome editing system for actinomycetes. *Appl. Microbiol. Biotechnol.* **104**, 225–239 (2020).
54. Ye, S., Enghiad, B., Zhao, H. & Takano, E. Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in *Streptomyces*. *J. Ind. Microbiol. Biotechnol.* **47**, 413–423 (2020).
55. Bozhüyük, K. A. J. *et al.* Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661 (2019).

56. Aparicio, J. F. *et al.* Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169**, 9–16 (1996).
57. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
58. Sambrook, J., Fritsch, E. F. & Maniatis, T. Molecular cloning: a laboratory manual. (Cold Spring Harbor Laboratory Press, New York, 1989).
59. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics*. (John Innes Foundation, Norwich, 2000).
60. Raynal, A., Karray, F., Tuphile, K., Darbon-Rongère, E. & Pernodet, J.-L. Excisable cassettes: new tools for functional analysis of *Streptomyces* genomes. *Appl. Environ. Microbiol.* **72**, 4839–4844 (2006).
61. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).
62. MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61–68 (1992).
63. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2005).
64. Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137 (1989).
65. Böhm, G., Muhr, R. & Jaenicke, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* **5**, 191–195 (1992).
66. Keatinge-Clay, A. T. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **14**, 898–908 (2007).

Table 1

Table 1 Summary of various strains generated, as well as the novel metabolites detected.

Strain	Modifications introduced	Stambomycins 1	Novel metabolites
ATCC/OE484	Wild type	ü	n.d.
K7N6/OE484 ^a	33 bp scar ^c	ü	n.d.
K7N5/OE484	^C DD ₄ helix swap, 33 bp scar	û	2, 3, 4, 5
CPN5/OE484 ^b	^C DD ₄ helix swap	ü	n.d.
K7N4/OE484	^C DD ₄ site-directed mutagenesis (SDM), 33 bp scar	ü	n.d.
CPN4/OE484	^C DD ₄ SDM	ü	n.d.
K7N3/OE484	<i>Dpks5-8</i> , 33 bp scar	û	2, 3, 4, 5
K7N2/OE484	^C DD ₄ helix swap, <i>Dpks5-8</i> , 33 bp scar	û	2, 3, 4, 5
CPN2/OE484	^C DD ₄ helix swap, <i>Dpks5-8</i>	û	2, 3, 4, 5
K7N1/OE484	^C DD ₄ SDM, <i>Dpks5-8</i> , 33 bp scar	û	2, 3, 4, 5
CPN1/OE484	^C DD ₄ SDM, <i>Dpks5-8</i>	û	2, 3, 4, 5
ATCC/OE484/Pks4+TEI	TEI fused to Pks4	û	2, 3, 4, 5
CPN2/OE484/TEI SDM	TEI inactivation (Ser à Ala), ^C DD ₄ helix swap, <i>Dpks5-8</i>	û	2, 3, 4, 5
CPN2/OE484/TEII SDM	TEII inactivation (Ser à Ala), ^C DD ₄ helix swap, <i>Dpks5-8</i>	û	2, 3, 4, 5
CPN2/OE484/D478	<i>DsamR0478</i> , ^C DD ₄ helix swap, <i>Dpks5-8</i>	û	2, 3, 4, 5
CPN2/OE484/D479	<i>DsamR0479</i> , ^C DD ₄ helix swap, <i>Dpks5-8</i>	û	6, 7, 8, 9
CPN2/OE484/ACP ₁₃ SDM	ACP ₁₃ H1 modified ^d , ^C DD ₄ helix swap, <i>Dpks5-8</i>	û	2, 3, 4, 5, 11
ATCC/OE484/hy59_S1	^N DD ₉ +KS ₂₁ replaced by ^N DD ₅ +KS ₁₄ , <i>Dpks5-8</i>	û	2, 3, 4, 5, 10, 11, 12
ATCC/OE484/hy59_S2	^N DD ₉ +KS ₂₁ replaced by ^N DD ₅ +KS _{14/21} , <i>Dpks5-8</i>	û	2, 3, 4, 5, 10, 11, 12, 14
ATCC/OE484/hy59_S1/D479	<i>DsamR0479</i> , ^N DD ₉ +KS ₂₁ replaced by ^N DD ₅ +KS ₁₄ , <i>Dpks5-8</i>	û	6, 7, 8, 9, 13
ATCC/OE484/hy59_S2/D479	<i>DsamR0479</i> , ^N DD ₉ +KS ₂₁ replaced by ^N DD ₅ +KS _{14/21} , <i>Dpks5-8</i>	û	6, 7, 8, 9, 13, 15
ATCC/OE484/hy59_S1/ACP ₂₁	ACP ₂₁ L1+H2	û	2, 3, 4, 5

region swap	modified ^e , ^N DD ₉ +KS ₂₁ replaced by ^N DD ₅ +KS ₁₄ , <i>Dpks5-8</i>		
ATCC/OE484/hy59_S1/ACP ₂₁ GtoD	ACP ₂₁ L1 modified ^f , ^N DD ₉ +KS ₂₁ replaced by ^N DD ₅ +KS ₁₄ , <i>Dpks5-8</i>	û	2, 3, 4, 5, 11, 14

^aK7N (pronounced “cassette number”) refers to mutants generated by PCR-targeting.
^bCPN refers to mutants generated using CRISPR-Cas9.
^cUse of PCR-targeting technique introduced a ‘scar’ sequence between *pks4* and *pks9*; for details, see **Supplementary Fig. 4**.
^dH1 modified refers to mutation of six residues within the helix α1 region of ACP₁₃ (EADQRR à PSERRQ); for details, see **Supplementary Fig. 9**.
^eL1+H2 modified refers to exchange of the loop 1+helix α2 region of ACP₂₁; for details, see **Supplementary Fig. 11**.
^fL1 modified indicates that one residue within the loop 1 region of ACP₂₁ was mutated (G₁₄₉₉ à D); for details, see **Supplementary Fig. 11**.
n.d. indicates no novel metabolites were detected.

Figures

carrier protein; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; TE, thioesterase; CDD, C-terminal docking domain; NDD, N-terminal docking domain.

Figure 2

A schematic representation of classical and alternative module definitions. A classical module (black) is defined as the catalytical unit responsible for incorporation of one building block into the growing polyketide chain, and associated functional group modifications. CDD/NDD pairs (shown) mediate communication between such traditionally-defined modules. According to this definition, KS/ACP chain extension interactions (left-pointing arrows) occur within the modules, whereas ACP/KS chain transfer interactions (right-pointing arrows) occur between modules, and the incoming substrate for the KS domain is generated by the upstream module. The alternative module boundaries (shown in green) were inspired by the evolutionary co-migration in certain systems of the KSs and the upstream processing domains^{14,15}. Under this revised definition, the KS/ACP chain extension interaction (left-pointing arrows) is inter-modular, while the ACP/KS chain transfer contacts (right-pointing arrows) are intra-modular.

point, CPN2 PKS, contains a functional CDD4 helix swap/NDD9 docking interaction (swapped docking a-helix shown in dark yellow), but a mis-matched ACP13/KS21 interdomain interaction. This PKS yielded only shunt products 2-5. The CPN2/ACP13 SDM PKS, in which the ACP13 helix $\alpha 1$ has been modified to match that of ACP20 (dark yellow ball), generates mini-stambomycin derivatives (11, a cyclic form lacking the internal hydroxy, Supplementary Fig. 9). The ATCC/hy59_S1 and S2 constructs were based on the alternative module definition, as the engineering point was selected downstream of the CDD4/NDD5 interface within the KS21 domain. Of the two junctions, that in which the fusion was located essentially at the midpoint of the domain (ATCC/hy59_S2) functioned better than that which included the majority of KS14 (ATCC/hy59_S1), although both PKSs led to successful generation of three novel mini-stambomycins, both as their free acids (10) and in macrolide form (11 and 12) (Supplementary Fig. 10). In an attempt to boost yields from ATCC/hy59_S1, two further constructs were created by modification of ACP21 – either by swapping a region implicated in KS/ACP communication during extension (ATCC/hy59_S1/ACP21 region swap), or by mutating a single key residue within the motif (ATCC/hy59_S1/ACP21 GtoD). The ATCC/hy59_S1/ACP21 region swap yielded only the shunt metabolites 2-5, while solely 11 was detected from ATCC/hy59_S1/ACP21 GtoD, showing that the introduced changes did not work as intended. We observed in addition from ATCC/hy59_S2 PKS and ATCC/hy59_S1/ACP21 GtoD, shunt product 14 corresponding to the chain released from module 21 (Supplementary Figs. 9 and 10), identifying the downstream module 22 as a blockage point.

Figure 5

Summary of the engineering strategies applied in this work to the stambomycin PKS. Inset are the six distinct approaches used, and the structures of the resulting metabolites are shown. The strategies giving rise to the target mini-stambomycins 10–12 are indicated in red. The hydroxyl group shown in pink is introduced by the P450-hydroxylase SamR0478, and that in red, by SamR0479.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryinformationSuetal.docx](#)