Article

High-throughput reprogramming of an NRPS condensation domain

Ines B. Folger [1], Natália F. Frota[2], Angelos Pistofidis[2], David L. Niquille[1], Douglas A. Hansen[1], T. Martin Schmeing [2] & Donald Hilvert [1]✉

Engineered biosynthetic assembly lines could revolutionize the sustainable production of bioactive natural product analogs. Although yeast display is a proven, powerful tool for altering the substrate specificity of gatekeeper adenylation domains in nonribosomal peptide synthetases (NRPSs), comparable strategies for other components of these megaenzymes have not been described. Here we report a high-throughput approach for engineering condensation (C) domains responsible for peptide elongation. We show that a 120-kDa NRPS module, displayed in functional form on yeast, can productively interact with an upstream module, provided in solution, to produce amide products tethered to the yeast surface. Using this system to screen a large C-domain library, we reprogrammed a surfactin synthetase module to accept a fatty acid donor, increasing catalytic efficiency for this noncanonical substrate >40-fold. Because C domains can function as selectivity filters in NRPSs, this methodology should facilitate the precision engineering of these molecular assembly lines.

Nonribosomal peptides (NRPs) represent a valuable source of clinical therapeutics[1,2]. These natural products are biosynthesized by multifunctional enzymatic assembly lines, called nonribosomal peptide synthetases (NRPSs), that use dedicated modules to incorporate each building block selectively and sequentially into the peptide scaffold (Fig. 1). NRPS modules typically possess three core domains, which perform substrate selection and activation (adenylation or A domains), shuttling of activated substrates (thiolation or T domains, also called peptide carrier proteins, PCP or CP domains) and peptide bond formation (condensation or C domains), respectively. Once product assembly is complete, the final peptide is typically released by a terminal thioesterase (TE) domain. The full NRPS assembly line may be a single megaenzyme containing all modules or may be divided into subunits that associate noncovalently through small 'docking' domains[3–5].

The modularity of NRPSs is appealing from an engineering perspective[6–10]. Individual domains, modules and even entire proteins can be added, deleted or exchanged with counterparts from different pathways to create hybrid synthetases for the production of new antibiotics and other bioactive materials[11–19]. The specificities of gatekeeper A domains can also be reprogrammed to incorporate unnatural building blocks with unique properties[20–28]. Nevertheless, the success of such strategies is often limited. Exchanging subunits or modules can disrupt interdomain contacts, lowering product yields or abolishing activity[9,10]. Efficiency may be further reduced if other enzymes in the modified assembly line, particularly C domains[10,29,30], fail to process noncognate substrates.

C domains are large enzymes (~450 amino acids), typically embedded in an NRPS protein or subunit[29,30]. They catalyze amide bond formation between an amino acid tethered to the downstream T domain and an amino acid or peptide tethered to the upstream T domain, leading to elongation of the growing peptide chain. Tolerance of C domains to noncanonical substrates can vary substantially. They are often selective with respect to the acceptor substrate[31–38] and sometimes disfavor large changes in the donor substrate as well[25,38–41]. Indeed, distinct classes of C domains have evolved based on the nature of their donor substrate, including those specific for L-amino acids, D-amino acids and fatty acids[42], although these proteins do not have discernibly different substrate binding pockets. Other C domains catalyze additional chemical reactions, including the dehydration of amino acids, to further boost the structural diversity of NRPS peptides[30,43–46].

[1]Laboratory of Organic Chemistry, ETH Zurich, Zurich, Switzerland. [2]Department of Biochemistry and Centre de Recherche en Biologie Structurale, McGill University, Montréal, Quebec, Canada. ✉e-mail: hilvert@ethz.ch
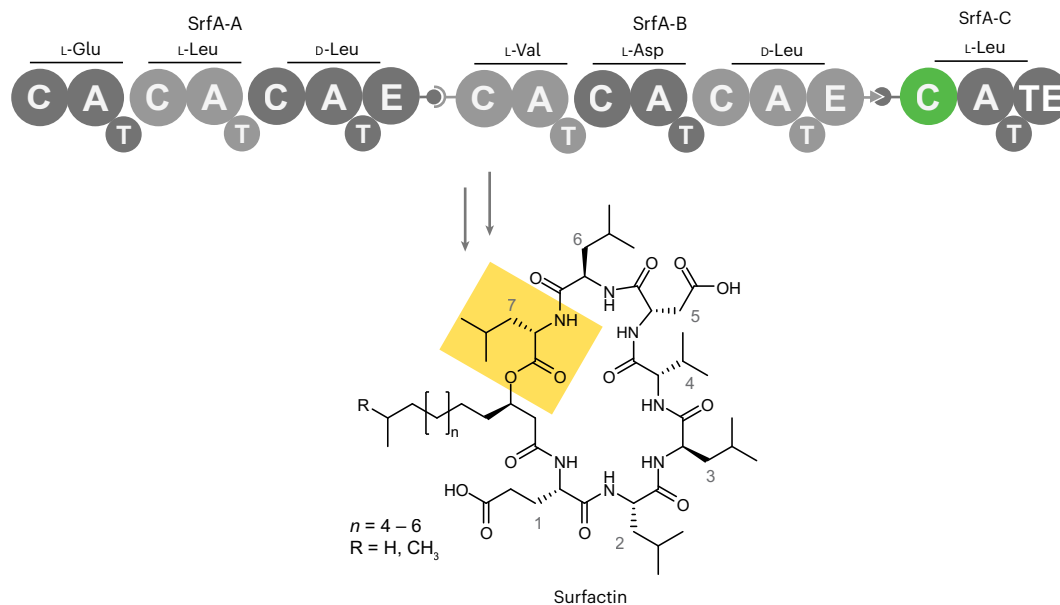
Fig. 1 | A typical NRPS. The enzymatic assembly line that produces the microbial lipopeptide antibiotic surfactin, surfactin synthetase, consists of the following three proteins: trimodular SrfA-A, trimodular SrfA-B and monomodular SrfA-C. The modules are composed of several functional domains, including C, A and T domains. In surfactin synthetase, biosynthesis is initiated by condensing a CoA-activated 3-hydroxy fatty acid to glutamate bound to the T domain of module 1. Each subsequent module performs peptide chain extension, and the final cyclic depsipeptide product is released by module 7's TE domain. Epimerase (E) domains in modules 3 and 6 invert the chirality of the respective leucine residues. Two cognate pairs of docking domains (gray ball/socket and gray arrow/pacman representation) enable selective interaction between partner protein subunits. In the current study, we re-engineered the donor substrate specificity of the SrfA-C C domain (green) to enable direct acylation of the acceptor Leu (residue 7, yellow highlight) with a fatty acid rather than an amino acid.

The absence of general rules for deducing substrate specificity[29,30,47] makes C-domain engineering more challenging than A-domain engineering. Compounding the problem, published C-domain assays[31–33,48,49] have relatively low throughput, making the discovery of rare combinations of specificity-altering mutations a formidable, labor-intensive undertaking.

Here we describe a yeast display[50] strategy that allows analysis and engineering of C-domain substrate selectivity in high throughput based on the functional display of an intact NRPS module on yeast. We used this system to reprogram the last C domain in surfactin synthetase to accept fatty acid substrates rather than amino acids. Strategic incorporation of such building blocks into NRP scaffolds could be used to target peptidyl drug candidates to cell membranes and thus enhance antimicrobial activity[51,52]. The ability to tune the properties of NRPS C domains has exciting potential to overcome many of the bottlenecks that have plagued previous efforts to adapt biosynthetic assembly lines for the sustainable production of non-natural products.

Results

High-throughput yeast display assay

NRPS-catalyzed peptide condensation requires components of two NRPS modules because the natural substrates of C domains are donor and acceptor acyl-T domains. These T domains must be loaded with their cognate building blocks, either naturally by cognate A domains or using chemoenzymatic approaches[53,54]. To monitor C-domain activity, we envisaged displaying an entire NRPS module, comprising docking, C, A and T domains, in functional form on the surface of yeast, and adding an upstream module, together with the required substrates, in solution (Fig. 2a). Productive interaction of the two modules, made possible by matching docking domains[3–5], would allow the C domain to catalyze amide bond formation between the substrate on the upstream module and the substrate on the displayed module. The resulting dipeptide product would be covalently linked to the T domain of the displayed module as a thioester and thus attached to the yeast surface. If the donor substrate contains an alkyne side chain, cells producing active

C domains can be visualized by a bioorthogonal 'click' reaction and isolated by fluorescence-activated cell sorting (FACS).

We chose the SrfA-C module from surfactin synthetase for surface display. SrfA-C, which is well characterized functionally and structurally[55], has the archetypal C, A and T domains, as well as a terminal TE domain that we omitted to prevent product off-loading after peptide bond formation. The core C–A–T domains of SrfA-C were fused to the yeast mating factor Aga2p to anchor the module to the yeast cell wall (Fig. 2a). A Myc tag was also included in the linker to allow determination of display efficiency by immunofluorescence labeling. Control experiments confirmed the successful expression and display of the NRPS module on EBY100 yeast cells (Fig. 2b). The T domain was subsequently modified post-translationally with phosphopantetheine (ppant) by addition of coenzyme A (CoA) and the 4′-ppant transferase Sfp[56], enabling activation and transfer of the acceptor substrate, L-Leu, to the ppant cofactor by the displayed A domain.

SrfA-C is known to interact productively with the first module of tyrocidine synthetase A (TycA) via compatible docking domains and T:C interactions[3]. As the upstream module, we chose a previously engineered variant of TycA, W227S TycA[22], which activates O-propargyl-L-Tyr. It was incubated with its substrate, ATP, Mg²⁺ and yeast cells displaying L-Leu-loaded SrfA-C for 15 min at room temperature. After washing the cells with phosphate buffer to remove the donor module, excess substrate and other reagents, the O-propargyl-D-Tyr-L-Leu product, tethered covalently to the ppant cofactor of the T domain on the yeast surface, was derivatized with biotin by copper(I)-catalyzed click chemistry[57,58]. Streptavidin modified with an R-phycoerythrin (R-PE) dye was added to visualize the labeled cells, which were analyzed by flow cytometry (Fig. 2b).

Although well-displayed, SrfA-C initially produced only modest yields of tethered dipeptide (Fig. 2b). Possible reasons for poor reactivity include loss of function of the displayed protein due to misfolding, product instability on yeast or unsuccessful docking of the upstream module with the displayed SrfA-C. Yeast, in contrast to many natural NRP producers, is also known to N-mannosylate proteins
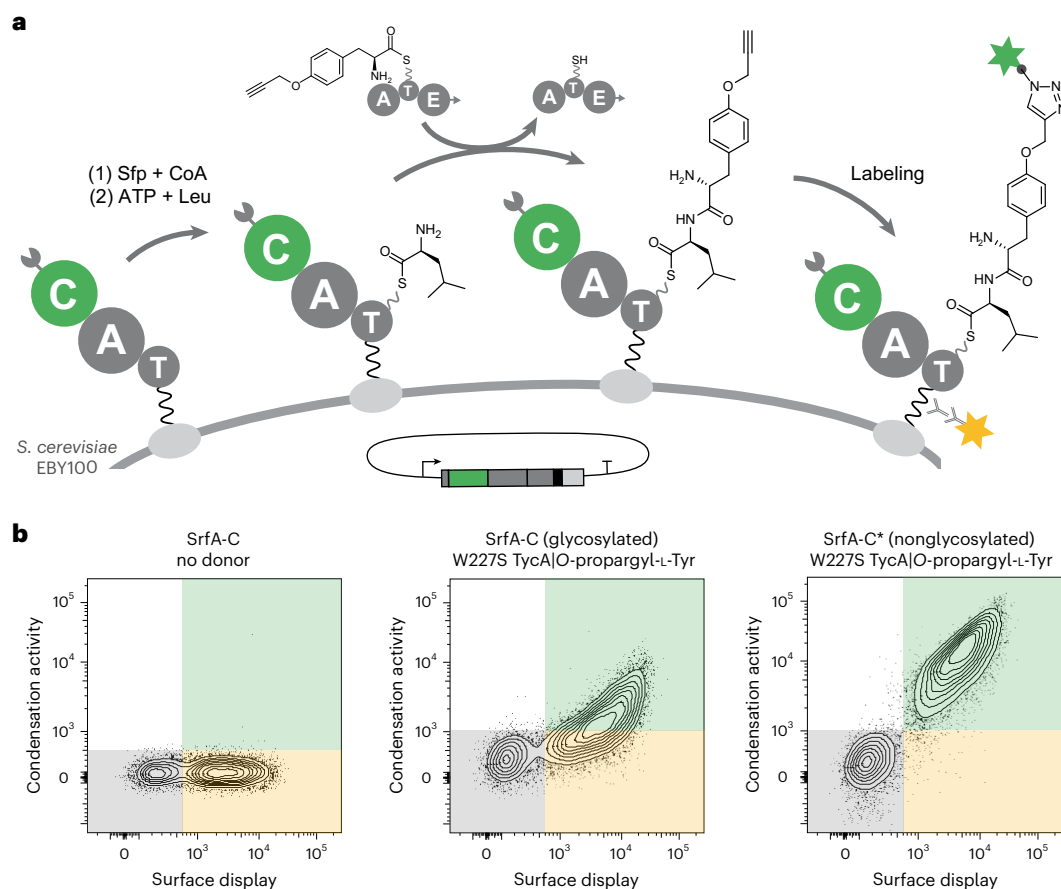
Fig. 2 | Assaying C-domain activity by yeast display. a, Schematic representation of enzymatic peptide bond formation on the surface of yeast. SrfA-C is displayed on the cell surface by fusion to the N-terminus of Aga2p via a short linker (black wavy line). The construct is primed with its ppant cofactor by the addition of CoA and the 4′-ppant transferase Sfp and loaded with the cognate amino acid, L-Leu, by action of the displayed A domain. Condensation is initiated by the addition of a donor module, W227S TycA, charged with O-propargyl-L-Tyr, in solution. Successful peptide bond formation is visualized by derivatizing the alkyne moiety in the dipeptidyl product with a biotin moiety using bioorthogonal click chemistry and adding fluorescently labeled streptavidin (R-PE, green star); a c-Myc epitope tag included in the linker was used to quantify display by immunofluorescence (FITC, yellow star). Active variants can then be isolated by FACS. The complementary docking domains are indicated by arrow and pacman appendages on the donor and acceptor domains, respectively. **b**, Flow cytometric analyses of SrfA-C modules on yeast (shown as contour plots with outliers (dots)). SrfA-C display is plotted (x axis, FITC) versus condensation activity (y axis, R-PE). Left, display of SrfA-C. Middle, yeast displaying glycosylated SrfA-C that were treated with 1.5 µM W227S TycA and 75 µM O-propargyl-L-Tyr for 15 min. Right, yeast displaying nonglycosylated SrfA-C* that were treated with 1.5 µM W227S TycA and 75 µM O-propargyl-L-Tyr for 15 min. The 'SrfA-C no donor' sample (left) was labeled with FITC and all other samples were labeled with FITC and R-PE immediately before analysis. The gray gate indicates cells that do not display any fluorescent label, the yellow gate indicates cells that show an FITC label and the green gate indicates cells that display both dyes.

at Asn-Xaa-Ser/Thr sequences (where Xaa is any amino acid except Pro)[59], which could adversely affect activity. To test whether SrfA-C is glycosylated, we reductively cleaved the displayed module from the yeast surface and analyzed it by mass spectrometry (MS). Tryptic digestion and peptide N-glycosidase F treatment indicated substantial N-glycosylation of Asn625 and Asn909 in the A domain of SrfA-C (Supplementary Fig. 1). To prevent this modification, we prepared a variant called SrfA-C* that has all potential N-glycosylation motifs disabled. Based on phylogenetic and structural analysis of similar A domains, residues Asn625, Ser787 and Asn909 were, respectively, mutated to Thr, Gln and Gln. SrfA-C*, in contrast to wild-type SrfA-C, efficiently formed the dipeptide on yeast when incubated with W227S TycA and substrates (Fig. 2b and Supplementary Fig. 2c). The signal was as high as for a positive control generated by directly ligating the authentic product precursor O-propargyl-D-Tyr-L-Leu-ppant to the displayed SrfA-C* module (Supplementary Fig. 2h), confirming that glycosylation had impaired A domain function in wild-type SrfA-C.

SrfA-C* docks with a variety of other NRPS donor modules as well, including TycA^βpY (TycA engineered to recognize O-propargyl-(S)-β-Tyr)[24], W227S/H743A TycA (an E domain knockout)[60] and W2742S

TycB3^COMᴰTycA (an excised TycB3 variant equipped with the TycA docking domain)[3,23], to generate yeast-displayed dipeptide products (Supplementary Fig. 2i–k). Negative controls with constructs lacking the ppant cofactor or inactivated by knockout mutations in the C or T domains (H147A/D151N SrfA-C* and S1003A SrfA-C*, respectively) could be displayed but did not form the dipeptide product (Supplementary Fig. 2d,e,g). Taken together, these results demonstrate that the C domain in the modified SrfA-C module effectively catalyzes dipeptide bond formation on yeast, covalently tethering the product to the displayed T domain.

Engineering the SrfA-C C domain for fatty acid recognition

Although the terminal C domain in surfactin synthetase is known to tolerate some changes in its native peptide substrate[3,10,11], we sought to expand its substrate scope to nonpeptidyl acyl donors. Incorporating lipids into NRP products is desirable because such moieties can be key to antimicrobial activity[51]. For example, the fatty acid component of the antibiotic daptomycin targets the NRP to bacterial cell membranes, resulting in microbial cell death[52]. Hence, C-domain-catalyzed incorporation of hydrophobic lipids into peptide scaffolds could be
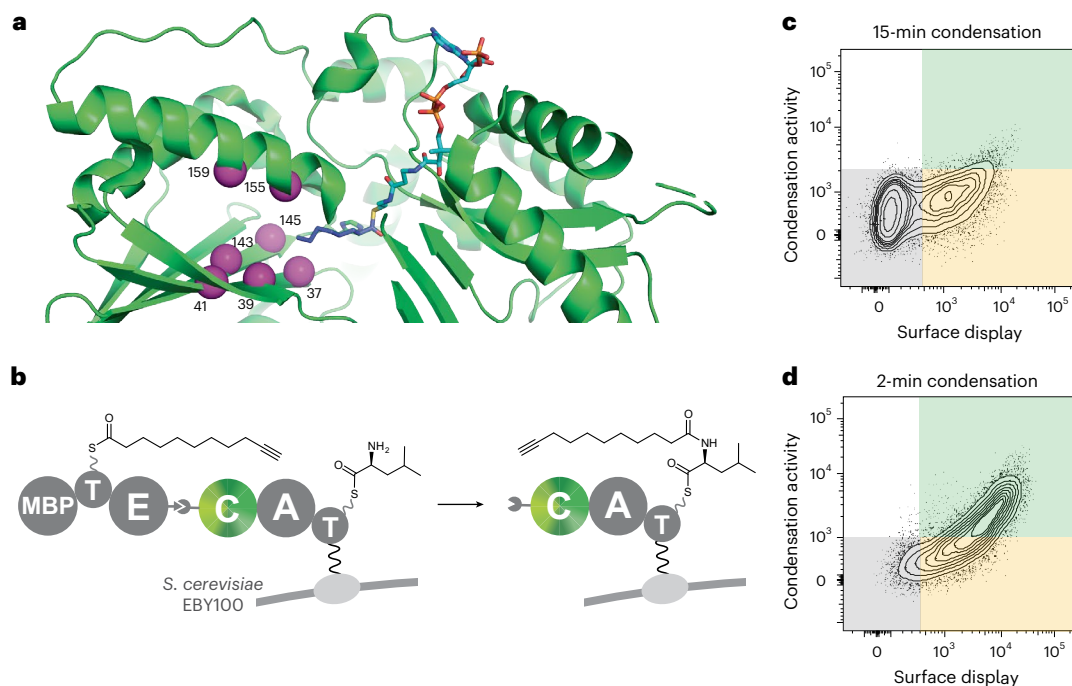
Fig. 3 | C-domain library design and sorting. a, Crystal structure of the SrfA-C C domain (PDB: 2VSQ, green cartoon)[55] overlaid with the C8-CoA fatty acyl substrate of the RzmA crystal structure (PDB: 7C1S, cyan and blue sticks)[39]. The active site residue His147 (green sticks) is positioned in close proximity to the thioester of the fatty acyl substrate. Randomized residues Glu37, Ile39, Met41, Trp143, Tyr145, Phe155 and Val159 are shown as magenta spheres. **b**, Screening of a large C-domain library displayed on the surface of yeast for fatty acylation of L-Leu. The library of SrfA-C* variants was displayed on the cell surface and reacted with 10-undecynoyl-ppant-TycA*. Active C-domain variants were labeled as described in Fig. 2 and analyzed by FACS. **c**, Contour plot of the SrfA-C*

C-domain library displayed on yeast and reacted for 15 min with 10-undecynoyl-ppant-TycA*. Cells displaying both fluorophores represent active variants that were sorted and enriched over five rounds of FACS with increasingly stringent conditions. **d**, Contour plot of the C-domain variant pool obtained after five rounds of sorting and enrichment, displayed on yeast and reacted for 2 min with 10-undecynoyl-ppant-TycA*. Surface display is shown on the x axis (FITC) and peptide bond formation on the y axis (R-PE). In gray, cells without PE and FITC labels; in yellow, cells that are labeled with FITC; in green, cells with both PE and FITC labels.

a general strategy to improve the pharmaceutical properties of many NRPs. Fatty acids are usually incorporated into NRPs by a specialized class of C domains called $C_{starter}$ domains[42], but we wondered whether yeast display could convert SrfA-C into a lipid-specific module.

To first test whether SrfA-C has any latent activity with inherently less reactive fatty acids, we examined its ability to process 10-undecynoic acid, which contains a terminal alkyne for bioorthogonal labeling. In the absence of an initiation module that can both activate this fatty acid and effectively interact with SrfA-C, we produced a modified TycA construct, TycA*, in which the A domain is replaced with maltose-binding protein (MBP) to improve solubility, and loaded the fatty acid onto its T domain using chemically synthesized 10-undecynoyl-pant in a one-pot enzymatic cascade (Supplementary Fig. 3)[53]. In vitro reaction of 10-undecynoyl-pant with the CoA biosynthetic enzymes pantothenate kinase (PanK), ppant adenylyltransferase (PPAT) and dephospho-coenzyme A kinase (DPCK) yielded 10-undecynoyl-CoA (liquid chromatography–high-resolution mass spectrometry (LC–HRMS $(M−H)^−$): calculated, 930.2275; found, 930.2280). Incubating this mixture with TycA* in the presence of Sfp afforded 10-undecynoyl-ppant-TycA* (HRMS: calculated, 107,517.2 Da; found, 107,515.0 Da). Unmodified protein was not detected, indicating near quantitative modification of the TycA* construct with the lipid substrate. As expected, however, the thioester linking the fatty acid to the T domain is not particularly stable and was hydrolyzed completely after 1-h incubation in buffer at room temperature (Supplementary Note– High-resolution electrospray mass spectrometry (ESI–MS), TycA* primed with 10-undecynoyl-ppant).

When 10-undecynoyl-ppant-TycA* was incubated with yeast displaying SrfA-C*, trace 10-undecynoyl-L-Leu (**1**) was formed (Supplementary Fig. 2l), suggesting that the fatty acid is a poor substrate for the

C domain. We quantified this activity in vitro using a bimodular assay, combining SrfA-C with 10-undecynoyl-ppant-TycA*, L-Leu, ATP and Mg^{2+} and monitoring the formation of **1** by LC–MS. Although activity was detected, the fatty acid was processed >15 times more slowly than O-propargyl-Tyr (Supplementary Fig. 4).

To engineer the C domain of SrfA-C to accept fatty acyl donors efficiently, we designed a library based on structures of RzmA-Cs, a $C_{starter}$ domain that initiates rhizomide A biosynthesis by acylation of a leucine with a short-chain fatty acid[39]. Although the C domains of RzmA and SrfA-C are only distantly related (20% sequence identity), they both have the classic C-domain fold[29] (which superimpose with a root mean square deviation of 4.2 Å over 336 Cα atoms) and the conserved HHxxxDG active site motif. The lipid substrate of RzmA-Cs sits deep within a pocket in the N-terminal lobe of the C domain, and comparison of the structures suggested that expansion of the SrfA-C substrate pocket between the core β-sheet in the N-terminal lobe and the fourth α-helix, close to the active site, might improve 10-undecynoic acid recognition. We therefore prepared a large combinatorial library in which hydrophobic (Ile39, Met41, Trp143, Tyr145, Phe155 and Val159) and negatively charged (Glu37) residues lining this site were simultaneously randomized using degenerate codons that code for small and hydrophobic amino acids (Fig. 3a and Supplementary Table 1).

The resulting library, which contained $1.5 × 10^6$ different SrfA-C variants, was displayed on yeast and screened for activity with 10-undecynoyl-ppant-TycA* by FACS. To identify the most active catalysts in the population, the cells were sorted and enriched over five consecutive cycles, increasing stringency by reducing the reaction time (Fig. 3 and Supplementary Fig. 5). Sequencing 15 variants

from the final cell pool yielded eleven distinct sequences possessing 3–7 mutations, which cluster into four closely related groups (Supplementary Fig. 6).

Biochemical characterization of the reprogrammed C domain

Based on flow cytometric analysis of selected clones from sort 5 (Supplementary Fig. 7), we produced three SrfA-C variants (lacking the glycosylation mutations and reequipped with the TE domain) in *Escherichia coli* HM0079 for study in vitro. The purified proteins were kinetically characterized by mixing with 10-undecynoyl-ppant-TycA*, L-Leu, ATP and Mg^{2+}, and monitoring the formation of 10-undecynoyl-L-Leu (**1**) by LC–MS (Fig. 4). Variant 10, harboring mutations W143T, Y145V and F155I, was found to be the best catalyst for rapid fatty acylation of L-Leu (Fig. 4 and Supplementary Figs. 4 and 8), giving an apparent rate constant of 38 ± 7 min^{-1}. We call this enzyme SrfA-C$_{FA}$. For comparison, wild-type SrfA-C is >40-fold less active under the same conditions, with an observed rate constant (k_{obs}) of 0.91 ± 0.09 min^{-1}.

Improved processing of the fatty acid by SrfA-C$_{FA}$ also manifests in higher amounts of total product formed during the reaction (210 ± 40 versus 19 ± 6 total turnovers for SrfA-C$_{FA}$ and SrfA-C, respectively; Supplementary Fig. 9). However, the reaction does not run to completion, presumably due to the limited stability of the 10-undecynoyl-ppant-TycA* thioester in the reaction buffer (Supplementary Fig. 9 and Supplementary Note). Consistent with this hypothesis, longer pre-incubation of the chemoenzymatic reaction mixture used to prepare 10-undecynoyl-ppant-TycA* decreased the amount of product formed, whereas three additions of freshly prepared 10-undecynoyl-ppant-TycA* to the in vitro reaction mixture increased product formation by 75%.

Interestingly, SrfA-C$_{FA}$ shows twofold higher activity than wild-type SrfA-C with *O*-propargyl-L-Tyr (Supplementary Fig. 4). The side chains of *O*-propargyl-L-Tyr and 10-undecynoic acid are of similar length and enlarging the binding pocket of SrfA-C$_{FA}$ presumably allows it to accommodate *O*-propargyl-L-Tyr better than the wild-type C domain. In the reactions with *O*-propargyl-L-Tyr, the donor substrate is catalytically regenerated, so considerably higher turnover numbers (1,900 ± 200) are also achieved. This suggests that an engineered TycA variant capable of catalytically reloading the lipid substrate would improve lipidation efficiency substantially.

Transplantation of the engineered C domain

Swapping individual catalytic domains is a useful means of diversifying biosynthetic assembly lines[6–10]. We therefore asked whether the engineered C domain could replace the first C domain in tyrocidine synthetase[61], which is located in the first module (TycB1) of the trimodular TycB protein. TycB1 naturally interacts with TycA and couples the upstream and downstream substrates D-Phe and L-Pro, so we expected that a modified TycB1 module containing the fatty acid-specific C domain would similarly engage productively with 10-undecynoyl-ppant-TycA* to afford lipidated products. For the swap, we chose domain boundaries based on a recently identified permissive recombination site between adjacent C and A domains[18] and inserted the SrfA-C$_{FA}$ C domain (residues Gln10–Gln430) between residues Val9 and Asn434 of TycB1 in a standalone C–A–T elongation module (TycB1$_{FA}$) and in the full-length TycB protein (TycB$_{FA}$).

As anticipated, incubation of the standalone TycB1$_{FA}$ module with 10-undecynoyl-ppant-TycA* in the presence of L-Pro, ATP and Mg^{2+} yielded 10-undecynoyl-L-Pro (**3**; Fig. 5a,b and Supplementary Fig. 10a). LC–MS analysis indicated that the TycB1$_{FA}$-catalyzed 10-undecynoylation of L-Pro is only approximately fourfold less efficient than the SrfA-C$_{FA}$-catalyzed reaction with L-Leu. Suboptimal domain boundaries between the swapped C and A domains, low tolerance for L-Pro at the acceptor site of the engineered C domain or slow off-loading of the product by the TE domain could easily account for this small decrease.
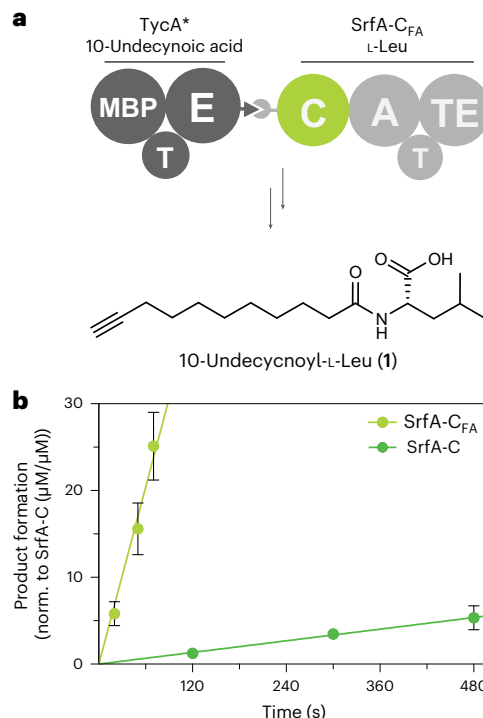

Fig. 4 | Kinetic characterization of SrfA-C$_{FA}$. a, Enzymatic fatty acylation of L-Leu catalyzed by SrfA-C and SrfA-C$_{FA}$ was performed in vitro by combining 10-undecynoyl-ppant-TycA* with SrfA-C$_{FA}$, L-Leu, ATP and Mg^{2+} to produce 10-undecynoyl-L-Leu (**1**). **b**, Reaction kinetics for the biosynthesis of 10-undecynoyl-L-Leu (**1**). Product formation over time is shown for the bimodular synthetase TycA*/SrfA-C$_{FA}$ (lime) and TycA*/SrfA-C wild type (green), normalized to 1 µM SrfA-C and SrfA-C$_{FA}$, respectively. Each data point and corresponding error bar represent the average ± s.d. of triplicate measurements with a single biological sample ($n = 1$). Four biological replicates ($n = 4$) were used to determine apparent rate constants.

Similarly, when we combined 10-undecynoyl-ppant-TycA* with trimodular TycB$_{FA}$ and the hexamodular TycC protein in the presence of all the required substrates (Fig. 5c), 10-undecynoyl-L-Pro-L-Phe-D-Phe-L-Asn-L-Gln-L-Tyr-L-Val-L-Orn-L-Leu (**4**) was formed, as detected by LC–MS and confirmed by LC–HRMS/MS (Fig. 5d,e and Supplementary Fig. 10b). Although further engineering will be required to optimize the efficiency of this artificial NRPS, the biosynthetic production of a new lipopeptide is notable considering that the system was not optimized for coupling a fatty acid to Pro or for the substrate preferences of other downstream domains. For the in vivo production of such peptides, we would also need to replace the preloaded TycA* donor module with a catalytic module evolved to activate fatty acids directly, for example, using the previously reported yeast display system to alter the substrate specificity of TycA[24]. Alternatively, the C domain could be evolved to react directly with CoA-activated fatty acid derivatives that can be produced in cells by fatty acyl-CoA ligases[62].

Structural characterization

To gain insight into how the engineered SrfA-C$_{FA}$ variant recognizes and processes a lipid substrate, we attempted to crystallize its C domain in complex with substrate analogs. Two unrelated, highly diffracting crystal forms of the SrfA-C$_{FA}$ C domain were obtained and used to solve apo structures to high resolution (Fig. 6 and Supplementary Table 2). We performed cocrystallization and soaking with 10-undecynoyl-pantetheinamide (10-undecynoyl-NH-pant) and 10-undecynoyl-NH-CoA, nonhydrolyzable small molecule analogs of the donor substrates, but as reported for other C domains[36], density for the ligands was not observed. We also undertook crystallization of the C domain covalently fused to the upstream donor
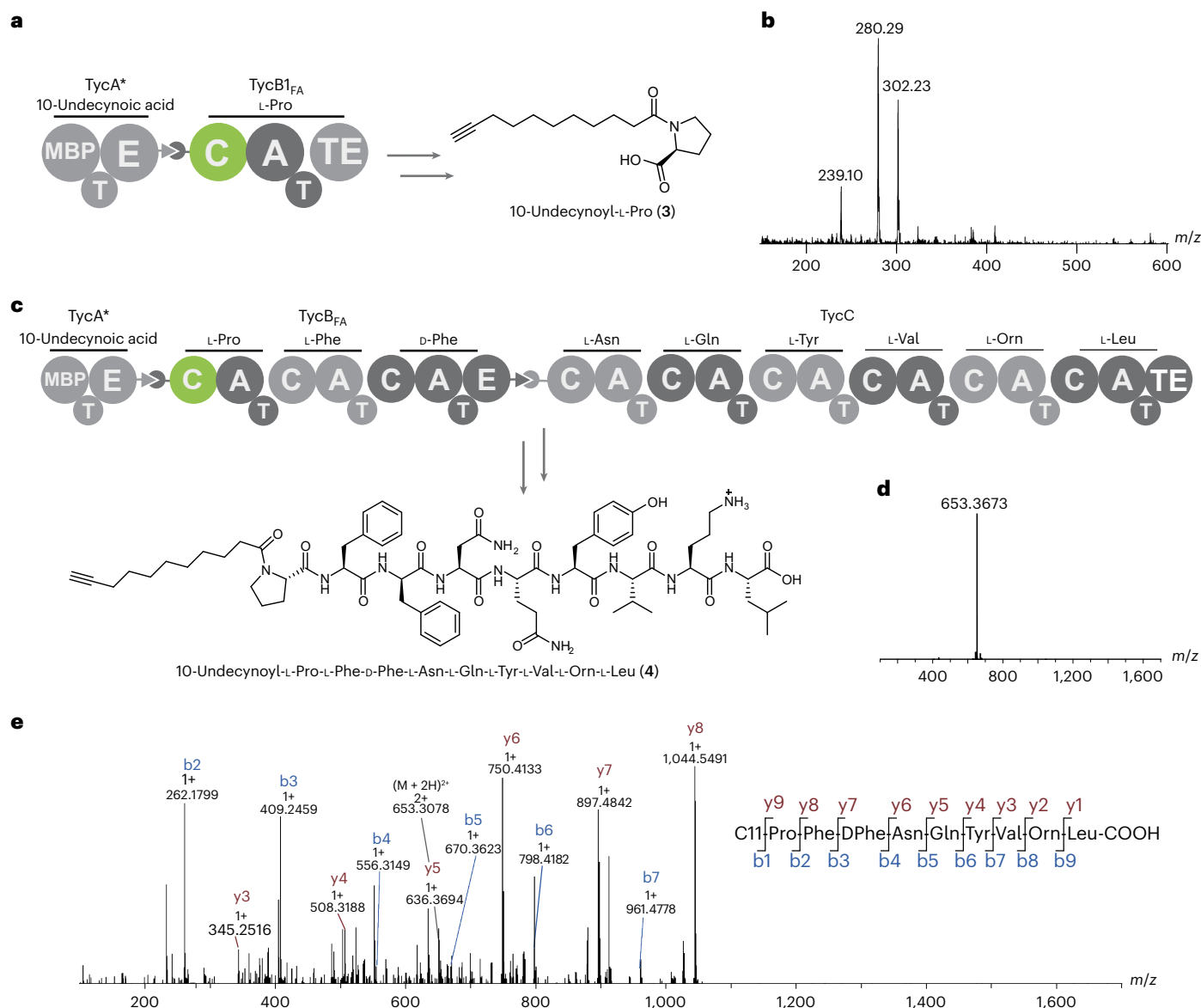
Fig. 5 | Biosynthesis of a nonribosomal lipopeptide by an engineered tyrocidine synthetase. The first C domain in tyrocidine synthetase was exchanged with the reprogrammed C domain of SrfA-C$_{FA}$. **a**, The bimodular synthetase TycA*/TycB1$_{FA}$ produces 10-undecynoyl-L-Pro (**3**) when supplied with 10-undecynoyl-ppant-TycA*, L-Pro, ATP and Mg^{2+}. **b**, Product formation was confirmed by LC–MS. ((M + H)$^+$: calculated, 280.19; found, 280.29; (M + Na)$^+$: calculated, 302.17; found, 302.23; the species 239.10 corresponds to (M + H)$^+$ of HEPES.) **c**, An engineered tyrocidine synthetase was generated by mixing

10-undecynoyl-ppant-TycA* and TycB$_{FA}$ with TycC, producing a new lipopeptide 10-undecynoyl-L-Pro-L-Phe-D-Phe-L-Asn-L-Gln-L-Tyr-L-Val-L-Orn-L-Leu (**4**). **d**, LC–HRMS/MS analysis: mass spectrum for the doubly charged lipopeptide **4** ((M + 2H)$^{2+}$: calculated, 653.3657; found, 653.3673). **e**, MS/MS spectrum of product **4**. Y ions are shown in brown and b ions in blue; respective charge and mass are indicated. All data are from one biological sample ($n = 1$). Chromatograms of LC–MS and LC–HRMS/MS analyses can be found in Supplementary Fig. 10.

complex, the 10-undecynoyl-NH-ppant-T domain, but crystals of the didomain did not grow. However, we also solved the high-resolution structures of wild-type SrfA-C, and comparison with SrfA-C$_{FA}$ structures provides insight into the origins of the specificity shift (Fig. 6a,b and Supplementary Table 2).

The structures of the SrfA-C C domain show the classic pseudodimeric V-shape and are very similar to each other and to the C domain in the previously characterized full-length SrfA-C module (Protein Data Bank (PDB): 2VSQ (ref. 55); Supplementary Fig. 11a). Small differences in side chain conformations of residues lining the putative fatty acyl binding site are observed between our wild-type structures and 2VSQ (Supplementary Fig. 11b), and the adjacent segment of the latch loop, which is poorly ordered in 2VSQ, is also disordered. In the SrfA-C$_{FA}$ structures, all mutations are clearly visible in electron density

maps (Fig. 6a,b) and introduce more space between the middle of helix 4 and the β sheet (Supplementary Fig. 11c,d). Introduction of the W143T, Y145V and F155I mutations causes the following three connected changes (Fig. 6c,d): (1) the N-terminus of helix 4 shifts inward, toward the pocket mutations, by up to 2 Å; (2) the helix's capping residue, Trp153, swings over to interact with the base of the latch; and (3) the latch shifts to pack Leu360 into the mutated pocket. However, in this conformation, the active site is not connected to the expanded mutated pocket, mainly because of the position of Glu37 and Leu360. Nevertheless, a reasonable 10-undecynoyl-ppant-TycA* model can be made that positions the distal end of the 10-undecynoyl moiety into the expanded pocket (Fig. 6e). Although this model is consistent with the mutations improving activity, cocrystal structures will ultimately be required to corroborate the proposed binding mode.
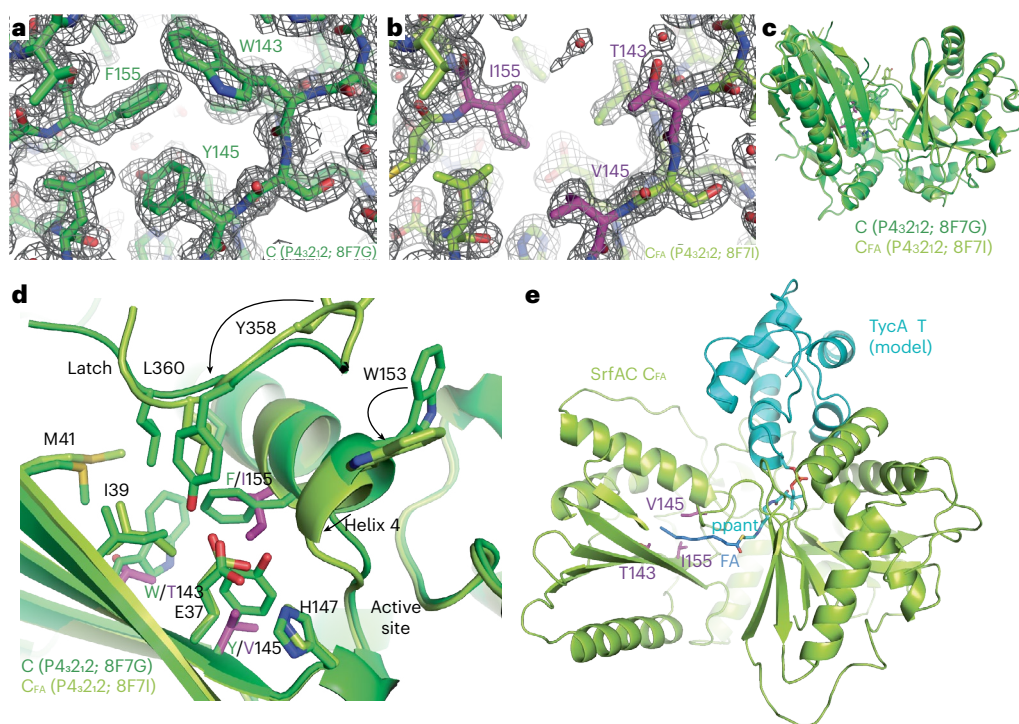
Fig. 6 | Structural analysis of the wild-type SrfA-C and SrfA-C_FA C domains. **a**, A $2F_O - F_C$ electron density map, contoured at 1σ, of the wild-type SrfA-C C domain (green). **b**, The analogous $2F_O - F_C$ electron density map, contoured at 1σ, of the fatty acid-specific C domain of SrfA-C_FA (light green with mutated residues in magenta). **c,d**, Overlay of the C domains of wild-type SrfA-C (green) and SrfA-C_FA (light green with mutated residues in magenta). **e**, An energy-minimized co-complex model of a 10-undecynoyl-ppant-T_TycA, homology model bound to the C domain of SrfA-C_FA.

Discussion

Customization of NRPS assembly lines for the biocatalytic production of bioactive peptide derivatives represents an attractive green alternative to chemical synthesis[6–10]. Nevertheless, despite some notable successes in reprogramming A domain specificities[20–28] and swapping individual domains and modules[11–19,41,63], disruption of critical domain interactions often introduces new bottlenecks in the resulting constructs that limit overall efficiency[9,10]. The substrate specificities of C domains[10,29] further complicate these engineering efforts.

New strategies for bioengineering NRPSs that feature nontraditional exchange points[16,17] and/or 'synthetic zippers' as artificial docking domains[64,65] have been developed and shown in some systems to overcome the severe drop in productivity often seen in classical NRPS engineering approaches. Harnessing combinatorial mutagenesis and high-throughput screening as described in the current study represents both a powerful alternative for addressing such problems and a complementary tool that can be applied to engineered systems to increase their productivity. By displaying full-length NRPS modules on the surface of yeast, the condensation activity of the displayed module can be directly linked to a simple fluorescent readout that enables flow cytometric screening of libraries containing up to 10^8 different variants[66,67]. Successful reprogramming of the SrfA-C C domain for coupling a fatty acid substrate with L-Leu highlights the efficacy of this approach. The >40-fold enhancement in fatty acylation obtained after a single round of mutagenesis and screening shows that C domain tolerance can be effectively expanded to building blocks substantially different from the native substrate. Moreover, the twofold improvement in O-propargyl-D-Tyr processing indicates that catalytic rates can be enhanced for multiple substrates simultaneously, broadening the substrate scope of C domains without compromising catalytic efficiency.

The small number of closely related variants found to catalyze the new lipidation reaction underscores the importance of screening large combinatorial libraries over rational design. Structure-based considerations would certainly have included mutation of Glu37 to a smaller residue, as the side chain of Glu37 blocks the fatty acid pocket in all crystal structures. However, such a substitution was not found in the best C-domain variant. Indeed, connectivity to the expanded binding pocket is not obvious in the structures of SrfA-C_FA, so although we can rationalize why this variant is more advantageous than the wild type, we would not have predicted the successful mutations a priori. Interrogating each of the target residues individually using conventional low-throughput C-domain assays, and combining beneficial mutations to find the best variant, would also have been a comparatively arduous undertaking.

In principle, any NRPS module that can be displayed in functional form on yeast will be engineerable by an analogous mutagenesis and screening strategy. Although sequence optimization may be required to prevent post-translational glycosylation or other problems that adversely affect catalytic activity, as seen for SrfA-C, a wide range of bacterial and fungal proteins should be amenable to this approach. This capability opens the door to systematic, high-throughput exploration of the poorly understood molecular recognition determinants that underlie substrate selection by C domains. Although our study focused on repurposing the binding pocket for the donor substrate, chemoenzymatic loading of the displayed A domain with a range of amino acids would similarly allow the acceptor binding site to be probed and remodeled. Beyond engineering C-domain specificity, adapting the yeast display strategy described here to gain deeper insight into the catalytic mechanism of these enzymes and to optimize critical T–C, C–A and docking domain interfaces should be relatively straightforward.

In sum, yeast surface display of functional elongation modules provides a robust method for studying and rapidly engineering C domains, the catalytic engines of NRPS assembly lines. Together with an analogous high-throughput assay for altering A domain specificity[24,25], high-throughput C-domain engineering lays the foundation for

creating increasingly complex NRPS assembly lines via scalable, sustainable and minimally invasive methods. This enabling methodology paves the way for the effective biosynthesis of life-saving therapeutic agents and other applications in synthetic biology.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-023-01532-x.

References

1. Sieber, S. A. & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **105**, 715–738 (2005).
2. Süssmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis— principles and prospects. *Angew. Chem. Int. Ed. Engl.* **56**, 3770–3821 (2017).
3. Hahn, M. & Stachelhaus, T. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc. Natl Acad. Sci. USA* **101**, 15585–15590 (2004).
4. Richter, C. D., Nietlispach, D., Broadhurst, R. W. & Weissman, K. J. Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **4**, 75–81 (2008).
5. Fage, C. D. et al. Communication breakdown: dissecting the COM interfaces between the subunits of nonribosomal peptide synthetases. *ACS Catal.* **11**, 10802–10813 (2021).
6. Cane, D. E., Walsh, C. T. & Khosla, C. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**, 63–68 (1998).
7. Schwarzer, D., Finking, R. & Marahiel, M. A. Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* **20**, 275–287 (2003).
8. Kim, E., Moore, B. S. & Yoon, Y. J. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* **11**, 649–659 (2015).
9. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317–347 (2016).
10. Brown, A. S., Calcott, M. J., Owen, J. G. & Ackerley, D. F. Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* **35**, 1210–1228 (2018).
11. Stachelhaus, T., Schneider, A. & Marahiel, M. A. Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**, 69–72 (1995).
12. Nguyen, K. T. et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. USA* **103**, 17462–17467 (2006).
13. Mootz, H. D. et al. Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J. Am. Chem. Soc.* **124**, 10980–10981 (2002).
14. Baltz, R. H. Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. *ACS Synth. Biol.* **3**, 748–758 (2014).
15. Steiniger, C., Hoffmann, S. & Süssmuth, R. D. Probing exchange units for combining iterative and linear fungal nonribosomal peptide synthetases. *Cell Chem. Biol.* **26**, 1526–1534 (2019).
16. Bozhüyük, K. A. J. et al. De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018).
17. 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).
18. Calcott, M. J., Owen, J. G. & Ackerley, D. F. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *Nat. Commun.* **11**, 4554 (2020).
19. Thong, W. L. et al. Gene editing enables rapid engineering of complex antibiotic assembly lines. *Nat. Commun.* **12**, 6872 (2021).
20. Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**, 9718–9726 (2002).
21. Thirlway, J. et al. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew. Chem. Int. Ed. Engl.* **51**, 7181–7184 (2012).
22. Kries, H. et al. Reprogramming nonribosomal peptide synthetases for 'clickable' amino acids. *Angew. Chem. Int. Ed.* **53**, 10105–10108 (2014).
23. Niquille, D. L., Folger, I. B., Basler, S. & Hilvert, D. Biosynthetic functionalization of nonribosomal peptides. *J. Am. Chem. Soc.* **143**, 2736–2740 (2021).
24. Niquille, D. L. et al. Nonribosomal biosynthesis of backbone-modified peptides. *Nat. Chem.* **10**, 282–287 (2018).
25. Camus, A., Truong, G., Mittl, P. R. E., Markert, G. & Hilvert, D. Reprogramming nonribosomal peptide synthetases for site-specific insertion of α-hydroxy acids. *J. Am. Chem. Soc.* **144**, 17567–17575 (2022).
26. Evans, B. S., Chen, Y., Metcalf, W. W., Zhao, H. & Kelleher, N. L. Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo. *Chem. Biol.* **18**, 601–607 (2011).
27. Villiers, B. & Hollfelder, F. Directed evolution of a gatekeeper domain in nonribosomal peptide synthesis. *Chem. Biol.* **18**, 1290–1299 (2011).
28. Zhang, K. et al. Engineering the substrate specificity of the DhbE adenylation domain by yeast cell surface display. *Chem. Biol.* **20**, 92–101 (2013).
29. Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta* **1865**, 1587–1604 (2017).
30. Dekimpe, S. & Masschelein, J. Beyond peptide bond formation: the versatile role of condensation domains in natural product biosynthesis. *Nat. Prod. Rep.* **38**, 1910–1937 (2021).
31. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489 (1999).
32. Ehmann, D. E., Trauger, J. W., Stachelhaus, T. & Walsh, C. T. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem. Biol.* **7**, 765–772 (2000).
33. Kaniusaite, M. et al. A proof-reading mechanism for non-proteinogenic amino acid incorporation into glycopeptide antibiotics. *Chem. Sci.* **10**, 9466–9482 (2019).
34. Doekel, S. & Marahiel, M. A. Dipeptide formation on engineered hybrid peptide synthetases. *Chem. Biol.* **7**, 373–384 (2000).
35. Ackerley, D. F. & Lamont, I. L. Characterization and genetic manipulation of peptide synthetases in *Pseudomonas aeruginosa* PAO1 in order to generate novel pyoverdines. *Chem. Biol.* **11**, 971–980 (2004).
36. Bloudoff, K., Alonzo, D. A. & Schmeing, T. M. Chemical probes allow structural insight into the condensation reaction of nonribosomal peptide synthetases. *Cell Chem. Biol.* **23**, 331–339 (2016).

37. Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. A. Peptide bond formation in nonribosomal peptide biosynthesis: catalytic role of the condensation domain. *J. Biol. Chem*. **273**, 22773–22781 (1998).

38. Luo, L. et al. Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* **41**, 9184–9196 (2002).

39. Zhong, L. et al. Engineering and elucidation of the lipoinitiation process in nonribosomal peptide biosynthesis. *Nat. Commun*. **12**, 296 (2021).

40. Kraas, F. I., Giessen, T. W. & Marahiel, M. A. Exploring the mechanism of lipid transfer during biosynthesis of the acidic lipopeptide antibiotic CDA. *FEBS Lett*. **586**, 283–288 (2012).

41. Kries, H., Niquille, D. L. & Hilvert, D. A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol*. **22**, 640–648 (2015).

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

43. Ziemert, N. et al. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **7**, e34064 (2012).

44. Patteson, J. B. et al. Structure and function of a dehydrating condensation domain in nonribosomal peptide biosynthesis. *J. Am. Chem. Soc*. **144**, 14057–14070 (2022).

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

46. Long, D. H. & Townsend, C. A. Acyl donor stringency and dehydroaminoacyl intermediates in β-lactam formation by a non-ribosomal peptide synthetase. *ACS Chem. Biol*. **16**, 806–812 (2021).

47. Izoré, T. & Cryle, M. J. The many faces and important roles of protein–protein interactions during non-ribosomal peptide synthesis. *Nat. Prod. Rep*. **35**, 1120–1139 (2018).

48. Schoppet, M. et al. The biosynthetic implications of late-stage condensation domain selectivity during glycopeptide antibiotic biosynthesis. *Chem. Sci*. **10**, 118–133 (2019).

49. Reimer, J. M. et al. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, eaaw4388 (2019).

50. Boder, E. T. & Wittrup, K. D. Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol*. **328**, 430–444 (2000).

51. Straus, S. K. & Hancock, R. E. W. Mode of action of the new antibiotic for Gram-positive pathogens daptomycin: comparison with cationic antimicrobial peptides and lipopeptides. *Biochim. Biophys. Acta* **1758**, 1215–1223 (2006).

52. Pogliano, J., Pogliano, N. & Silverman, J. A. Daptomycin-mediated reorganization of membrane architecture causes mislocalization of essential cell division proteins. *J. Bacteriol*. **194**, 4494–4504 (2012).

53. Nazi, I., Koteva, K. P. & Wright, G. D. One-pot chemoenzymatic preparation of coenzyme A analogues. *Anal. Biochem*. **324**, 100–105 (2004).

54. Worthington, A. S. & Burkart, M. D. One-pot chemo-enzymatic synthesis of reporter-modified proteins. *Org. Biomol. Chem*. **4**, 44–46 (2006).

55. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–663 (2008).

56. Lambalot, R. H. et al. A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem. Biol*. **3**, 923–936 (1996).

57. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem*. **67**, 3057–3064 (2002).

58. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl*. **41**, 2596–2599 (2002).

59. Wildt, S. & Gerngross, T. U. The humanization of N-glycosylation pathways in yeast. *Nat. Rev. Microbiol*. **3**, 119–128 (2005).

60. Stachelhaus, T. & Walsh, C. T. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry* **39**, 5775–5787 (2000).

61. Mootz, H. D. & Marahiel, M. A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol*. **179**, 6843–6850 (1997).

62. Arora, P., Vats, A., Saxena, P., Mohanty, D. & Gokhale, R. S. Promiscuous fatty acyl CoA ligases produce acyl-CoA and acyl-SNAC precursors for polyketide biosynthesis. *J. Am. Chem. Soc*. **127**, 9388–9389 (2005).

63. Crüsemann, M., Kohlhaas, C. & Piel, J. Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chem. Sci*. **4**, 1041–1045 (2013).

64. Bozhueyuek, K. A. J., Watzel, J., Abbood, N. & Bode, H. B. Synthetic zippers as an enabling tool for engineering of non-ribosomal peptide synthetases. *Angew. Chem. Int. Ed*. **60**, 17531–17538 (2021).

65. Abbood, N., Duy, Vo, T., Watzel, J., Bozhueyuek, K. A. J. & Bode, H. B. Type S non-ribosomal peptide synthetases for the rapid generation of tailormade peptide libraries. *Chem. Eur. J*. **28**, e202103963 (2022).

66. Picot, J., Guerin, C. L., Le Van Kim, C. & Boulanger, C. M. Flow cytometry: retrospective, fundamentals and recent instrumentation. *Cytotechnology* **64**, 109–130 (2012).

67. Liu, B. (ed). *Yeast Surface Display. Methods in Molecular Biology* Vol. 1319 (Humana Press, 2015).

Methods

Materials

Tyrocidine synthetase proteins and variants were produced in *E. coli* and purified as previously described[23]. Sfp and the CoA biosynthetic enzymes, PanK, PPAT and DPCK, were also produced in *E. coli* and purified using similar procedures. See Supplementary Fig. 12 for SDS–PAGE analysis of these proteins. For details on the synthesis of 10-undecynoyl-pantetheine, amino acid substrates and lipidated products, see Supplementary Note–Materials and methods.

Cloning of SrfA-C variants for yeast surface display

The plasmid pCT_tycA_AT[24] was modified for the display of SrfA-C variants. Five fragments were amplified by PCR. Fragments 1 and 2, encoding the signal peptide and linker region, were produced from pCT_tycA_AT and primer pairs SP_f/SP_r and linker_f/linker_r, respectively. The third fragment, encoding SrfA-C without the TE domain, was amplified from genomic DNA with the primer pair srfC_f/srfC_r. Fragments 4 and 5, encoding Aga2p and a portion of the vector backbone, were generated from pCT_tycA_AT using the primer pairs aga2p_f/aga2p_r and vector_f/vector_r, respectively. Fragments 1–4 and 3–5 were assembled via overlap PCR with the primer pairs SP_f/aga2p_r and linker_f/vector_r, respectively. Fragments 1–4 were digested with EcoRI and XhoI, fragments 3–5 with XhoI and XbaI and pCT_tycA_AT with EcoRI and XbaI, and the three components were subsequently ligated to yield pCTDN_srfA-C. For cloning purposes, a SalI restriction site was subsequently introduced behind the C-domain coding region. Two fragments were amplified from the plasmid using primers P1/P5 and P6/H33, reassembled by overlap PCR using P1/H33, digested with NheI and XhoI and ligated into the backbone of pCTDN_srfA-C that had been digested with the same restriction enzymes to give the display plasmid pCT_srfA-C. To prevent post-translational glycosylation of SrfA-C in yeast, three point mutations (N625T, S787Q and N909Q) were introduced into the sequence encoding the A domain using overlap PCR and Gibson assembly with the primer pairs R7/P19, P18/P23, P22/P26 and P25/H33 to give pCT_srfA-C*. For detailed cloning procedures, see Supplementary Note–Materials and methods. Primer sequences are listed in Supplementary Table 3.

Yeast display of SrfA-C and variants

Yeast surface display was adapted from the published protocol discussed in ref. 50 using optimized cell growth conditions, incubation times and substrate concentrations. pCT_srfA-C and pCT_srfA-C* were used to transform *Saccharomyces cerevisiae* EBY100 cells using the Frozen-EZ Yeast Transformation II Kit (Zymo Research) according to the manufacturer's protocol. Transformed cells were plated on synthetic dextrose–casein amino acids (SD–CAA) plates (250 ml; 20 g l^{-1} D-(+)-glucose, 1.7 g l^{-1} Difco yeast nitrogen base without amino acids and ammonium sulfate (BD Biosciences), 0.5 g l^{-1} ammonium sulfate, 5 g l^{-1} Difco casamino acids (BD Biosciences), 40 mM Na$_2$HPO$_4$, 70 mM NaH$_2$PO$_4$, 182 g l^{-1} sorbitol, 16 g l^{-1} agar, pH 6) and incubated at 30 °C for 2–3 d. A preculture of a single clone was prepared in 2 ml SD–CAA medium (20 g l^{-1} D-(+)-glucose, 1.7 g l^{-1} Difco yeast nitrogen base without amino acids and ammonium sulfate (BD Biosciences), 0.5 g l^{-1} ammonium sulfate, 5 g l^{-1} Difco casamino acids (BD Biosciences), 40 mM Na$_2$HPO$_4$, 70 mM NaH$_2$PO$_4$, pH 6) and grown overnight (18–20 h) at 30 °C and 260 r.p.m. The preculture was subsequently diluted to optical density (OD)$_{600}$ = 0.2 in 3 ml SD–CAA and regrown at 30 °C and 260 r.p.m. for 7–9 h until an OD$_{600}$ of 2–4 was reached. The fresh culture was then diluted again to OD$_{600}$ = 0.2 and stored for up to 2 d at 4 °C. For protein induction and surface display, the cell culture was regrown for 6 h at 30 °C and 260 r.p.m. and centrifuged at 500g at 4 °C for 3 min. The cell pellet was resuspended in 3 ml SG–CAA medium (20 g l^{-1} D-(+)-galactose, 1.7 g l^{-1} Difco yeast nitrogen base without amino acids or ammonium sulfate (BD Biosciences), 0.5 g l^{-1} ammonium sulfate, 5 g l^{-1} Difco casamino acids (BD Biosciences),

40 mM Na$_2$HPO$_4$, 70 mM NaH$_2$PO$_4$, pH 6) and shaken for 21–23 h at 20 °C and 230 r.p.m.

The efficiency of SrfA-C display on EBY100 was assessed by immunofluorescence labeling with the primary mouse anti-c-Myc antibody 9E10 (Roche, ROAMYC; final concentration 300 ng μl^{-1}) and the secondary anti-mouse IgG-fluorescein isothiocyanate (FITC) antibody F2012 (Sigma-Aldrich; final concentration: 65 ng μl^{-1}) in 10 μl permeabilization buffer (PMB; 7.2 mM NaH$_2$PO$_4$, 40 mM Na$_2$HPO$_4$, 137 mM NaCl, 2.7 mM KCl, 1 mM MgCl$_2$, pH 7.4) per 100 μl cells[25]. Flow cytometric analysis was performed on an LSRFortessa Cell Analyzer (BD Biosciences) at the Flow Cytometry Core Facility of ETH Zurich. Display of SrfA-C constructs on yeast usually varied between 50% and 80% and was dependent on induction time as well as cell viability.

LC–MS/MS analysis of yeast-displayed SrfA-C

Display of SrfA-C on EBY100 was performed as described above using 35 ml cell culture. After 23 h induction, the cells were centrifuged for 3 min at 5,000g at 4 °C and washed a total of four times with 50 ml PMB. The cell pellet was resuspended in 1 ml PMB containing 1 mM tris(2-carboxyethyl)phosphine (TCEP) and incubated at 4 °C for 25 min. The supernatant was transferred to a fresh tube and protein was concentrated in a Vivaspin 500 centrifugal concentrator. The sample was analyzed by SDS–PAGE using a Phast system and 7.5% gels (GE Health-Care) following the manufacturer's protocols (Supplementary Fig. 1).

For LC–MS/MS analysis by the Functional Genomics Center Zurich, the sample was digested with trypsin. Specifically, 10 μl sample was diluted to 90 μl with 50 mM triethylammonium bicarbonate buffer (TEAB), and 100 μl of 20% trichloroacetic acid was added. The sample was washed twice with acetone, dissolved in 45 μl (50 mM) TEAB (pH 8.5) mixed with 0.9 μl TCEP (2 mM) and 1.4 μl 15 mM chloroacetylacetone, and incubated for 30 min at 60 °C. Subsequently, 5 μl of a trypsin stock solution (100 ng μl^{-1}) was added, and the proteins were digested overnight. The sample was dried and subsequently dissolved in 20 μl (0.1%) formic acid.

For de-glycosylation of the digested sample, half of the solution was dried, redissolved in 20 μl (25 mM) ammonium bicarbonate buffer (pH 8.5), mixed with 1 μl PNGase F (1 U μl^{-1}) and incubated overnight at 37 °C. The sample was subsequently dried and redissolved in 20 μl (0.1%) formic acid. The trypsin-digested and glycosylated and de-glycosylated peptide samples were analyzed by LC–MS/MS, injecting 2 μl on a nanoAcquity ultra-performance liquid chromatography system coupled to a Q Exactive mass spectrometer and analyzed using the software PEAKS[68].

Assaying peptide bond formation on yeast

To equip the displayed T domains with the ppant cofactor, 100 μl induced cells displaying SrfA-C or SrfA-C* were collected by centrifugation at 700g for 30 s at room temperature, washed 2–3 times with 90 μl PMB, resuspended in a total volume of 25 μl PMB supplemented with 4 μM Sfp, 500 μM CoA and 1 g l^{-1} BSA and incubated at room temperature for 20–25 min. The acceptor substrate was loaded onto the ppant cofactor by the displayed A domain in the same or in a separate step using 2 mM Leu and 100 μM ATP. To initiate peptide formation, the cells were washed with 90 μl PMB, collected and resuspended in 25 μl assay buffer (100 mM HEPES, 100 mM NaCl, 50 mM ATP, 10 mM MgCl$_2$, pH 7.25) supplemented with 1.5 μM W227S TycA and 75 μM *O*-propargyl-L-Tyr and incubated at room temperature for 15 min. The cells were collected and washed with 90 μl PMB, and the cell pellet was resuspended in 25 μl PMB, supplemented with freshly prepared 4 mM ascorbate, 200 μM BPSA, 100 μM CuSO$_4$ and 20 μM N$_3$-PEG$_3$-biotin (Sigma-Aldrich), and incubated for 15 min at room temperature to biotinylate any propargylated dipeptide product tethered to the yeast surface. The cells were collected and washed with 90 μl PMB. The cells were labeled with 300 ng μl^{-1} anti-c-Myc antibody 9E10 in 10 μl PMB, 65 ng μl^{-1} anti-mouse IgG-FITC antibody F2012 and 65 ng μl^{-1}

streptavidin-R-PE (Thermo Fisher Scientific, S866) in 10 µl PMB[25]. The cells were collected, washed twice, resuspended in 400 µl PMB, transferred to FACS tubes and analyzed by flow cytometry as before.

As a positive control for maximal signal, the authentic dipeptide product was loaded directly onto the displayed SrfA-C* construct by treating the induced cells with 500 µM O-propargyl-ᴅ-Tyr-ʟ-Leu-pant, the enzymes PPAT (final concentration: 45 µM), DPCK (45 µM) and PanK (45 µM), and Sfp (11 µM) in assay buffer for 25 min at room temperature. Biotinylation, fluorescent labeling and flow cytometric analysis were performed as described above.

Experiments with the alternative donor modules W2742S TycB3$^{COM^D-TycA}$, which also activates O-propargyl-ᴅ-Tyr[23]; W227S/H743A TycA, which is specific for O-propargyl-ʟ-Tyr[60]; and TycA^{βpY}, which activates O-propargyl-(S)-β-Tyr[24], were performed analogously.

Preparation of 10-undecynoyl-ppant-TycA*

TycA*, the surrogate donor module for delivery of 10-undeycnoic acid, was produced from plasmid pSU18_tycA*, in which the sequence for the A domain of TycA was replaced with that of MBP (see Supplementary Note—Materials and methods for cloning details), in E. coli BL21 (DE3) cells and purified similar to wild-type TycA[23]. For post-translational modification, 10-undecynoyl-pant, PanK, PPAT, DPCK and Sfp were pre-incubated in assay buffer for 15 min at 37 °C, directly followed by addition of TycA* and incubation for an additional 5 min at 37 °C (final concentrations: 500 µM 10-undecynoyl-pant, 8 µM PanK, 8 µM PPAT, 8 µM DPCK, 2 µM Sfp and 195 µM TycA*). The reaction mixture containing the product 10-undeycnoyl-ppant-TycA* complex was used immediately in subsequent assays without further purification.

Library construction and display

The C-domain library was constructed by homologous recombination in yeast. The vector backbone was generated by digesting pCT_srfA-C* with NheI and SalI, purified by 1% agarose gel electrophoresis and extracted using the Zymoclean Gel DNA Recovery Kit (Zymo Research). The library insert, which was designed to overlap with the backbone by 100 bp on each end, contained six degenerate codons (Supplementary Table 1). It was produced by amplifying three pCT_srfA-C* fragments using the primer pairs P1/R29, R30:R31(8:1)/R32:R33(8:1) and R34/P2, followed by assembly PCR in a second step using P1/P2 (all primer sequences are listed in Supplementary Table 3). Purified vector backbone and insert were used to transform electrocompetent EBY100 cells[50], which were prepared as described in ref. 69 using 3 µg each of the insert and backbone for 800 µl competent cells. The cells were rescued in 16 ml (1 M) sorbitol:YPD (1:1) for 1 h at 30 °C and 225 r.p.m. Transformed cells were collected by centrifugation at 3,000g for 3 min at 4 °C, resuspended in minimal medium SD–CAA and plated out on SD–CAA in serial dilutions to determine the transformation efficiency. The library was regrown at 30 °C and 230 r.p.m. until an OD_{600} of 2–4 was reached and subsequently diluted to an OD_{600} of 0.2 (in 55 ml SD–CAA). The next day, the library was regrown at 30 °C and 230 r.p.m. until an OD_{600} of 2–4 was reached and subsequently diluted to an OD_{600} of 0.2. This step was repeated twice and cells were diluted to an OD_{600} of 0.2 and stored at 4 °C until further use. The integrity of the library was confirmed by Sanger sequencing.

High-throughput screening of the SrfA-C library

The SrfA-C C-domain library was displayed on yeast by regrowing 55 ml of the transformed cell culture (OD_{600} = 0.2) for 6 h at 30 °C and 260 r.p.m. The cells were collected by centrifugation at 5,000g at 4 °C for 3 min, resuspended in 55 ml SG–CAA medium, shaken for 21–23 h at 20 °C and 230 r.p.m. and modified with ppant and ʟ-Leu as before. For reactions with 10-undecynoic acid as the donor substrate, 3 ml cells were washed with 2.7 ml PMB, resuspended in a 750 µl reaction mixture containing 195 µM freshly prepared 10-undecynoyl-ppant-TycA* and incubated for 2–15 min at room temperature. After biotinylation

and fluorescent labeling, the library was sorted using the FACSAria III cell sorter (BD Biosciences; at ~7,000–25,000 events per second) at the Flow Cytometry Core Facility of ETH Zurich. In the first sort, 7.4×10^7 cells of the C-domain library were analyzed (approximately fivefold oversampling). After sorting, enriched cells were regrown at 30 °C and 250 r.p.m. in SD–CAA supplemented medium containing 100 µg ml^{-1} chloramphenicol to prevent bacterial contamination until an OD_{600} of 2–6 was reached. The cells were diluted to an OD_{600} = 0.2 and regrown until an OD_{600} of 2–4 was reached. The cells were again diluted to OD_{600} = 0.2 and stored until the next sort. Sorting and enrichment were repeated a total of five times, increasing stringency by reducing the incubation time from 15 to 2 min (Supplementary Fig. 5). After the final sort, genes of representative variants were isolated and analyzed by Sanger sequencing (Microsynth AG).

Characterization of isolated SrfA-C variants on yeast

Variants 2, 4, 7, 10 (SrfA-C$_{FA}$), 11, 13, 14 and 15 were individually displayed on yeast as described above. The samples were incubated with 50 µM 10-undecynoyl-ppant-TycA* for 100 µl cell samples at room temperature for 5 min and, following fluorescent labeling, analyzed by flow cytometry as described above. The sort 5 pool was treated identically for comparison.

Kinetic characterization of selected SrfA-C variants

The genes for the three best C-domain variants from the individual yeast display assays were amplified from the pCT vector by PCR and assembled into the full-length C–A–T–TE SrfA-C module in the pTrc99a plasmid[70]. The C-terminally hexahistidine-tagged holo SrfA-C variants were produced in E. coli HM0079, purified at pH 7.4 and stored at pH 7.25 (see Supplementary Note—Materials and methods for details)[23]. The purified variants (1–2 µM) and 2 mM Leu were added to a solution of 10-undecynoyl-ppant-TycA*, freshly prepared as described above (final concentrations: 800 µM 10-undecynoyl-pant, 8 µM PanK, 8 µM PPAT, 8 µM DPCK, 2 µM Sfp and 426 µM TycA*) and incubated at 37 °C. The formation of 10-undecynoyl-ʟ-Leu was monitored as a function of time by periodically analyzing aliquots by LC–MS. A synthetic 10-undecynoyl-ʟ-Leu standard was used for product quantification via the isolated ion count for the product ($(M + H)^+$) in LC–MS chromatograms. The experiments were performed in triplicate with four biological replicates. For details, see Supplementary Note—Materials and methods.

For the biosynthesis of O-propargyl-ᴅ-Tyr-ʟ-Leu, the SrfA-C variants (1.5 µM) were incubated with 5 mM O-propargyl-ʟ-Tyr, 5 mM Leu and 15 µM W227S TycA instead of 10-undecynoyl-pant-TycA* and 2 mM Leu, under otherwise identical conditions. The reactions were performed as time course experiments in duplicate or triplicate with three biological replicates and analyzed by HPLC. A synthetic O-propargyl-ᴅ-Tyr-ʟ-Leu standard was used for product quantification at 220 nm. A background control reaction was prepared containing all reagents and enzymes except SrfA-C, and rates were corrected for the nonenzymatic formation of O-propargyl-ʟ-Tyr-ʟ-Leu.

Replacement of the C domain of TycB1 with that of SrfA-C$_{FA}$

The gene encoding the C domain of TycB1, a truncated Pro-specific elongation module from tyrocidine A synthetase with an appended TE domain[24], and full-length TycB were replaced with the gene encoding for the C domain of SrfA-C$_{FA}$ to yield pTrc99a_tycB1_FA and pTrc99a_tycB_FA. TycB1$_{FA}$ and TycB$_{FA}$ were produced in E. coli HM0079 cells and purified as reported for the corresponding wild-type proteins[23]. For detailed protocols, see Supplementary Note—Materials and methods.

Biosynthesis of 10-undecynoyl-ʟ-Pro (3)

Enzymatic fatty acylation of ʟ-Pro was performed in vitro in HEPES buffer (100 mM HEPES, 100 mM NaCl, 50 mM ATP, 10 mM MgCl₂, 0.1 units ml^{-1} inorganic pyrophosphatase from baker's yeast, pH 7.25).

TycB1$_{FA}$ (1.7 μM) was mixed with 2 mM L-Pro, added to a solution of 10-undecynoyl-ppant-TycA*, which was freshly prepared as described above (final concentrations: 800 μM 10-undecynoyl-pant, 8 μM PanK, 8 μM PPAT, 8 μM DPCK, 2 μM Sfp and 426 μM TycA*), and incubated at 37 °C. The amount of product formed was estimated using an authentic 10-undecynoyl-L-Leu standard. Reactions were performed in duplicate, and product formation was analyzed by LC–MS after 4 h. For details, see Supplementary Note–Materials and methods.

Biosynthesis of lipopeptide (4)

Biosynthesis of 10-undecynoyl-L-Pro-L-Phe-D-Phe-L-Asn-L-Gln-L-Tyr-L-Val-L-Orn-L-Leu was performed in vitro in HEPES buffer (100 mM HEPES, 100 mM NaCl, 50 mM ATP, 10 mM MgCl$_2$, 0.1 units ml^{-1} inorganic pyrophosphatase from baker's yeast, pH 7.25). TycB$_{FA}$ (1 μM) and TycC (1 μM) were combined with all required amino acids (2 mM L-Phe, 1 mM L-Pro, 1 mM L-Asn, 1 mM L-Gln, 1 mM L-Tyr, 1 mM L-Val, 1 mM L-Orn and 1 mM L-Leu), added to a solution of 10-undecynoyl-ppant-TycA*, which was freshly prepared as described above (final concentrations: 553 μM 10-undecynoyl-pant, 8 μM PPAT, 8 μM DPCK, 8 μM PanK, 2 μM Sfp and 355 μM TycA*), and incubated at 37 °C. Reactions were performed in triplicate, and product formation was analyzed by LC–HRMS/MS after 2 h. For detailed protocols, see Supplementary Note–Materials and methods.

Production and purification of the SrfA-C$_{FA}$ C domain

The gene for the C domain of the SrfA-C$_{FA}$ module was synthesized by Bio Basic and subcloned into a pET21 vector. The vector contains an N-terminal tobacco etch virus (TEV) cleavable octa-histidine tag. For details, see Supplementary Note–Materials and methods. The protein was expressed in *E. coli* BL21 (DE3) cells grown at 37 °C in lysogeny broth medium supplemented with 50 μg ml^{-1} kanamycin. Protein expression was induced when the culture reached an OD$_{600}$ of 0.6–0.8, by adding 0.5 mM isopropyl ß-D-1-thiogalactopyranoside and shaking at 16 °C for 19 h. Cells were collected by centrifugation before protein purification. For purification of the C domain of SrfA-C$_{FA}$ for later crystallography experiments, cell pellets were resuspended in buffer A (2 mM imidazole, 150 mM NaCl, 2 mM β-mercaptoethanol (β-ME), 25 mM HEPES, pH 7.5) and lysed by sonication. The lysate was clarified by centrifugation at 20,000g for 30 min at 4 °C. The supernatant was loaded onto a 5 ml HiTrap IMAC FF column (Cytiva) charged with Ni^{2+} equilibrated in buffer A. The column was washed with buffer A, and the protein was eluted with a 10–50% gradient of buffer B (250 mM imidazole, 150 mM NaCl, 2 mM β-ME, 25 mM HEPES, pH 7.5). Fractions containing purified protein were pooled and incubated with TEV protease during dialysis into buffer A overnight at room temperature. Protein was reloaded onto the HiTrap IMAC FF column to remove uncleaved protein, and the flow-through was collected. Protein was then applied to a MonoQ HR 16/60 column (Cytiva) equilibrated in buffer Q1 (0.5 mM TCEP, 25 mM HEPES, pH 7.5). Bound protein was washed with buffer Q1 plus 100 mM NaCl and eluted using a gradient of 100–600 mM NaCl over 80 ml. Protein was pooled, concentrated, and applied to a HiLoad 16/60 Superdex S75 column (GE Healthcare) equilibrated in size exclusion buffer (150 mM NaCl, 0.5 mM TCEP, 25 mM HEPES, pH 7.5). Purity was accessed by SDS–PAGE, and pure fractions were pooled, concentrated to 38.9 mg ml^{-1}, flash-frozen in liquid nitrogen and stored at −80 °C.

Protein crystallization and structure determination

The isolated C domain of SrfA-C$_{FA}$ was used in sparse matrix crystallization trials to obtain initial crystallization conditions, with 3,000 crystallization conditions in 96-well plates at 4 °C and 22 °C assayed. The most promising crystals were optimized by fine screening in 24-well, sitting drop plates using 2 μl protein solution and 2 μl reservoir solution in the drop and a 500 μl reservoir volume. Two conditions produced diffraction-quality crystals. The first condition included 25 mg ml^{-1} of protein and a crystallization solution of 0.04 M KH$_2$PO$_4$, 16% (wt/vol) polyethylene glycol (PEG) 8 K and 20% (vol/vol) glycerol. The second condition included 25 mg ml^{-1} of protein and a crystallization solution of 0.16 M calcium acetate, 0.08 M sodium cacodylate (pH 6.5), 14.4% (wt/vol) PEG 8 K and 20% (vol/vol) glycerol. Wild-type SrfA-C C domain was also crystallized in these two conditions. Crystals were looped and flash-cooled in liquid nitrogen. Diffraction data sets were collected using beamline CMCF-ID of the Canadian Light Source, NE-CAT 24-ID-E beamline at the Advanced Photon Source (APS) and beamline 23ID-B at APS during the CCP4/APS school in macromolecular crystallography. Data were indexed and scaled with HKL2000 (ref. 71), HKL3000 (ref. 72) or DIALS[73]. Structure determination was accomplished by molecular replacement in Phaser[74] using the C domain of the SrfA-C termination module structure (PDB: 2VSQ)[55] as a search model, followed by iteratively building in Coot[75] and refinement in Phenix[76]. The data collection and refinement statistics are summarized in Supplementary Table 2.

Donor complex modeling

As co-complex structures were not forthcoming, we produced a model of 10-undecynoyl-ppant-T$_{TycA}$ bound to the SrfA-C$_{FA}$ C-domain variant. The T$_{TycA}$ model was produced using RoseTTAFold[77] and ligated to a model of 10-undecynoyl-ppant. The 10-undecynoyl-ppant-T$_{TycA}$ was placed based on aminoacyl-ppant-T[49] and fatty acyl-CoA structures[39], and energy was minimized using Phenix, with 10-undecynoyl-ppant restraints generated in eLBOW[78].

Reporting summary

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

Data availability

The X-ray crystal structures and diffraction data from this study were deposited in the Research Collaboratory for Structural Bioinformatics PDB under accession codes 8F7F, 8F7G, 8F7H and 8F7I. Source data for all figures in the main text and Supplementary Information have been supplied with the paper or as Supplementary Information files. Source data are provided with this paper.

References

68. Ma, B. et al. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **17**, 2337–2342 (2003).

69. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C. M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–159 (2010).

70. Gruenewald, S., Mootz, H. D., Stehmeier, P. & Stachelhaus, T. In vivo production of artificial nonribosomal peptide products in the heterologous host *Escherichia coli*. *Appl. Environ. Microbiol.* **70**, 3282–3291 (2004).

71. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

72. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 859–866 (2006).

73. Winter, G. et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D Biol. Crystallogr.* **74**, 85–97 (2018).

74. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

75. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

76. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

77. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).

78. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009).

Author contributions

All authors were involved in the project design. I.B.F. was responsible for assay development, directed evolution, chemical synthesis and biochemical characterization. N.F.F., A.P. and T.M.S. performed co-complex modeling and X-ray crystallography. D.L.N. established assay proof-of-concept. D.A.H. designed TycA* and carried out chemical synthesis. The paper was written by I.B.F. and D.H. and revised by all authors.

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/s41589-023-01532-x.

Correspondence and requests for materials should be addressed to Donald Hilvert.

Peer review information *Nature Chemical Biology* thanks Max Cryle and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

nature portfolio

Corresponding author(s): Donald Hilvert

Last updated by author(s): Nov 1, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	BD FACSDiva™ Software was used to collect data on LSRFortessa Cell Analyzer (BD) as well as on FACSAria III cell sorter (BD). MassLynx was used to collect and analyse data collected on LC-MS (Waters H-class UPLC/SQD-2). Data on the Thermo nanoAcquity UPLC coupled to a Q Exactive mass spectrometer was collected using Thermo Exactive software. Data run on U-HPLC (Ultimate 3000, Dionex) were collected using Chromeleon (Version 7.2 SR4). Diffraction data sets were collected using beamline CMCF-ID of the Canadian Light Source, NE-CAT 24-ID-E beamline at the Advanced Photon Source (APS) and beamline 23ID-B at APS during the CCP4/APS school in macromolecular crystallography. Data of LC-HRMS/MS and LC-HRMS measurements were collected on a Bruker maxis-ESI-Qq-TOF-MS using Bruker Compass DataAnalysis (Version 5.3).
Data analysis	Flow cytometry and flourescence activated cell sorting data was analysed using FlowJo (Versoin 10.5-10.8). MassLynx was used to analyse data collected on LC-MS (Waters H-class UPLC/SQD-2). Calculations of kinetic parameters were performed in Microsoft Excel and plots were generated using GraphPad Prism (Version 8.0.0 - 10.0.3). DNA sequences were analysed using CLC Genomics Workbench (Version 10-11). Data on the Thermo nanoAcquity UPLC coupled to a Q Exactive mass spectrometer was analysed utilizing PEAKS. Data run on U-HPLC (Ultimate 3000, Dionex) were analysed using Chromeleon (Version 7.2 SR4). Calculations of kinetic parameters were performed in Microsoft Excel and plots were generated using GraphPad Prism (Version 8.0.0 - 10.0.3). NMR data was analysed using MestReNova (MNova, Version 14). Diffraction data were indexed and scaled with softwares HKL2000 (Version 2.3.8), HKL3000 (Version 2.3.15) or DIALS (Version 3.8.0). Structure determination was accomplished using Phaser (Version 2.1), Coot and Phenix (Version 1.20.1 and Version 0.9.8.1). The data collection and refinement statistics are summarized in Supplementary Table 2.

Data of LC-HRMS/MS and LC-HRMS measurements were analysed using Bruker Compass DataAnalysis (Version 5.3).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The X-ray crystal structures and diffraction data from this study were deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) under accession codes 8F7F, 8F7G, 8F7H and 8F7I. Source Data for all figures in the main text and supplementary information have been supplied with the manuscript or as supplementary information files.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was not predetermined using statistical methods. To obtain kinetic data at least three biological replicates were performed on different days with different protein batches to account for differences in protein stability. Each biological replicate included duplicates or triplicates to account for pipetting errors. Biological replicates were reproducible and support the findings of this study. Similarly, for comparison of engineered C domain variants, two biological replicates were performed on different days. Biological replicates were reproducible and support the finding of this work. For TycA* stability assessment, one biological experiment was performed to support hypothesized TycA* instability. Although the experiment has not been replicated, similar experiments using slightly different conditions (data not shown in manuscript) supported this hypothesis.
Data exclusions	no data were excluded
Replication	Generated data was generally reproducible. Kinetic measurements were repeated 3-4 times (see above). Kinetic comparison of C domain variants was performed twice and TycA* stability test was performed once.
Randomization	Randomization was not relevant to this study as reagent preparation and culture conditions were kept constant. Experiments were performed with the exact same procedure, controls and data analysis methodology. Samples were analyzed individually and all parameters for protein expression were kept constant.
Blinding	Blinding was not relevant to study as data was collected by unbiased methods (i.e. machines: LC-MS, HPLC, flow cytometry, ...) and individual control samples have been included in every experiment performed and analysed individually.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-c-myc 9E10 from mouse IgG1κ (Millipore Sigma, ROAMYC, 11667203001); Anti-Mouse IgG (whole molecule)–FITC antibody produced in goat (Millipore Sigma, F2012)
Validation	validated by manufacturer (https://www.sigmaaldrich.com/US/en/product/roche/roamyc)

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	EBY100 originally obtained from Plückthun research group at University of Zurich who ordered the strain from ATCC.
Authentication	not authenticated
Mycoplasma contamination	not tested for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	n/a

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Proteins were displayed on cells and assayed as described in text. Cells were fluorescently labeled via streptavidin-R-PE and anti-myc-FITC and directly subjected to flow cytometry. Cells were gated for single cells via standard procedures (plotting SSC-A vs FSC-A and SSC-H vs SSC-A for flow cytometry analysis and plotting SSC-A vs FSC-A, FSC-H vs FSC-A, and FSC-H vs FSC-W for FACS) and single cells were analyzed as described/shown in manuscript.
Instrument	LSRFortessa Cell Analyzer (BD) and FACSAria III cell sorter (BD)
Software	Software of BD Instruments for data collection and data analysis using FlowJo
Cell population abundance	Single cells represented usually 65-80% of total cell population

Gating strategy | Single cells were gated (plotting SSC-A vs FSC-A and SSC-H vs SSC-A for flow cytometry analysis and plotting SSC-A vs FSC-A, FSC-H vs FSC-A, and FSC-H vs FSC-W for FACS) and single cell population was analyzed and gated using appropriate negative cell poppulation controls that were not fluorescently labeled, displayed inactive enzyme variants or were unable to produce products (see Suppl. Fig. 2 for details)

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.