Corrected 22 March 2024. See full text.

RESEARCH ARTICLE SUMMARY

BIOSYNTHESIS

Evolution-inspired engineering of nonribosomal peptide synthetases

Kenan A. J. Bozhüyük*†, Leonard Präve†, Carsten Kegler†, Leonie Schenk†, Sebastian Kaiser, Christian Schelhas, Yan-Ni Shi, Wolfgang Kuttenlochner, Max Schreiber, Joshua Kandler, Mohammad Alanjary, T. M. Mohiuddin, Michael Groll, Georg K. A. Hochberg, Helge B. Bode*

INTRODUCTION: Natural products (NPs) have played a pivotal role in drug discovery, contributing to 48% of new medicines developed between 1981 and 2019. Despite their significance, there are obstacles in translating NPs into clinical drugs owing to their structural complexity and limitations to derivatize or synthesize them. Genetic engineering or synthetic biology present promising avenues for the efficient and cost-effective discovery of tailored biological drugs. Bacterial NPs, especially those derived from nonribosomal peptide synthetases (NRPSs), have emerged as ideal targets for synthetic biology and have the potential to enhance the pharmacological properties of nonribosomal peptides (NRPs) in clinical development. However, rational engineering of NRPSs remains complex despite technological advancements.

RATIONALE: Recent advancements in bioinformatic and genetic engineering technologies as well as structural data have propelled synthetic biology strategies for manipulating megasynthetases, offering innovative solutions for the production of NRP analogs. Building on the growing understanding of NRPS evolution, this study emphasizes the importance of intragenomic recombination, speciation, horizontal gene transfer, and recombination as driving forces behind the diversification and functionalization within NRP families. We hypothesized that recognizing the central evolutionary mechanisms is essential for the redesign of assembly lines, with the aim of achieving greater structural diversity and increased production yields. By using phylogenetic hidden Markov models and principal component analysis–based machine learning approaches, our study tries to expand the understanding of intragenomic recombination for NRPS engineering, identifying regions with inconsistent evolutionary histories as potential synthetic breakpoints.

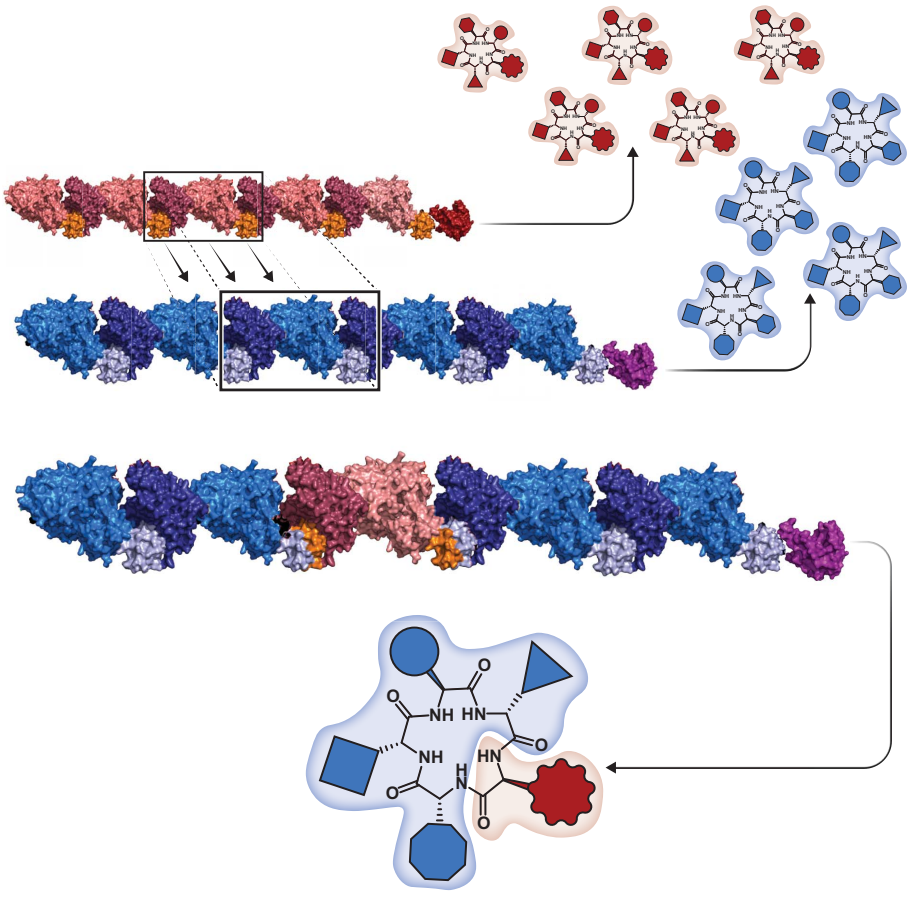


Illustration of the XUT engineering concept. Two different NRPSs are shown in red and blue producing different schematic cyclic peptides. A modified blue NRPS is illustrated with one module exchanged against a red NRPS module through the XUT approach, resulting in a rationally engineered peptide product. Structural data for illustrations were derived from PDB 6MFZ and PDB 2CB9.

RESULTS: The analysis of NRPS evolution, combined with systematic experimental analysis and in silico methods, unveils a previously undocumented recombination site within NRPSs. Using fusion point screening, we identified evolution-inspired synthetic engineering sites and designed more than 50 artificial peptides by combining building blocks from unrelated natural NRPS systems. The developed engineering framework, named the evolution-inspired eXchange Unit between T domains (XUT), aligns with structural revelations, substantially converging with the recently proposed unified model for the evolution of NRPSs. Important to this work was the identification of an additional yet undescribed recombination site within NRPS's thiolation domains, which allows the combination of NRPS building blocks that differ in taxonomy, biochemistry, and GC content.

CONCLUSION: This study applies insights into the evolution of NRPS as a foundation to improve engineering of these megasynthetases. The XUT approach broadens the synthetic biology toolkit and facilitates the creation of tailor-made bioactive peptides. This approach is versatile and complementary to previous engineering strategies and holds great potential for advancing synthetic biology and NP engineering for clinical drug discovery, development, and optimization. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: kenan.bozhueyuek@myria.bio (K.A.J.B.); helge.bode@mpi-marburg.mpg.de (H.B.B.)
†These authors contributed equally to this work.
Cite this article as K. A. J. Bozhüyük *et al.*, *Science* **383**, eadg4320 (2024). DOI: 10.1126/science.adg4320

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adg4320

RESEARCH ARTICLE

BIOSYNTHESIS

Evolution-inspired engineering of nonribosomal peptide synthetases

Kenan A. J. Bozhüyük[1,2,3]*†‡, Leonard Präve[1,2]†, Carsten Kegler[1,2]†, Leonie Schenk[1,2]†, Sebastian Kaiser[1,4], Christian Schelhas[1], Yan-Ni Shi[2], Wolfgang Kuttenlochner[5], Max Schreiber[1,2], Joshua Kandler[2], Mohammad Alanjary[6], T. M. Mohiuddin[2], Michael Groll[5], Georg K. A. Hochberg[4,7,8], Helge B. Bode[1,2,7,8,9]*

Many clinically used drugs are derived from or inspired by bacterial natural products that often are produced through nonribosomal peptide synthetases (NRPSs), megasynthetases that activate and join individual amino acids in an assembly line fashion. In this work, we describe a detailed phylogenetic analysis of several bacterial NRPSs that led to the identification of yet undescribed recombination sites within the thiolation (T) domain that can be used for NRPS engineering. We then developed an evolution-inspired "eXchange Unit between T domains" (XUT) approach, which allows the assembly of NRPS fragments over a broad range of GC contents, protein similarities, and extender unit specificities, as demonstrated for the specific production of a proteasome inhibitor designed and assembled from five different NRPS fragments.

Natural products (NPs) play a crucial role in drug discovery, accounting for 48% of new medicines developed between 1981 and 2019 (*1*). However, the translation of NPs into clinical drugs is challenging owing to their complex structures, which limit chemical derivatization or full synthesis, especially at the large scale (*2*). This complexity hinders the exploration of structure-activity relationships, impeding the development of NP leads. To address these challenges and surpass synthetic chemistry limitations, genetic engineering and synthetic biology of NPs offer promising avenues for faster and cost-effective discovery of tailor-made biological drugs (*3*). Bacterial NPs, which frequently originate from nonribosomal peptide synthetases (NRPSs) (*4*), stand out as ideal targets for synthetic biology, aiming to enhance the pharmacological properties of NP leads in clinical development.

NRPSs are genetically encoded molecular assembly lines that biosynthesize a broad range of valuable nonribosomal peptides (NRPs) or even clinical drugs, such as penicillins (*5*–*7*), bleomycin (*8*), and cyclosporin (*9*). These assembly lines consist of repeating modules of enzymatic domains, each catalyzing the incorporation and chemical modification of a specific extender unit into the growing chain before passing it on to the next module (*4*). Hundreds of different extender units, typically derived from amino acids, have been described (*10*, *11*). An adenylation (A) domain selects and activates an extender unit, covalently linking it to the prosthetic thiol (phosphopantetheine) group added posttranslationally to a thiolation (T) domain. Condensation (C) domains then linearly connect these covalently bound substrates to the growing NRP chain. Tailoring domains, including heterocyclization (Cy), epimerization (E), *N*-methylation (MT), oxidation (Ox), or reduction (R) domains, may be present to modify the NRP chain. Lastly, the full-length NRP is released by hydrolysis or macrocyclization; its release is usually catalyzed by a thioesterase (TE) domain.

The logic of this assembly line mechanism has inspired efforts to engineer megasynthetases for natural product analog production. Recent advances in technology and structural data (*12*) have accelerated the development of innovative synthetic biology strategies for megasynthetase engineering. Additionally, the continuous growth of publicly available genomic data, along with community efforts to develop processing tools for biosynthetic gene cluster (BGC) and NP identification (*13*–*16*), has led to a new trend in assembly line engineering by using evolution-driven strategies. Insightful studies suggest that understanding the evolutionary mechanisms of these often huge multifunctional enzyme machines can enhance our ability to redesign assembly line proteins, achieving greater structural diversity while maintaining high production titers, therefore having the potential to expand our therapeutic arsenal (*17*–*23*).

In the investigation of NRPS evolution, a recent in silico study emphasized the significance of intragenomic recombination, speciation, horizontal gene transfer, and recombination in NRPS evolution but also introduced a unified model for the evolution of present-day NRPSs (*17*). It highlighted the occurrence of single recombination events at multiple breakpoints within NRPS A domains, a phenomenon termed subdomain swapping. This finding not only emerged as a widespread mechanism, as shown previously (*18*, *20*–*22*, *24*, *25*), but also surfaced as a pivotal driver of diversification and functionalization within NRP families. It is important to acknowledge that this propensity for recombination within NRPS A domains results in divergent phylogenetic histories for C and A domains, extending beyond the confines of a single NRPS module.

Driven by this understanding, our focus shifted to the implications and potential applications of this phenomenon for NRPS engineering. We hypothesized that regions within a NRPS module sharing a common phylogenetic history might not be suitable for synthetic recombination owing to their relative lack of recombination events. Conversely, regions with inconsistent histories could serve as promising synthetic breakpoints given their higher likelihood of undergoing frequent recombination to produce inconsistent histories.

Therefore, we sought to decipher the evolutionary history of NRPSs to facilitate efficient BGC engineering. By using a systematic analysis of various NRPS sequences using in silico methods, we discovered a previously undocumented recombination site. Through fusion-point screening, we identified and tested evolution-inspired engineering sites through NRPS reprogramming, which ultimately led to the design of a biologically active peptide.

Results

Deciphering the evolutionary history of NRPSs

Considering the hypothesis that interfaces between diverse evolutionary histories offer favorable recombination sites, we used a maximum likelihood methodology to determine these boundaries. We used a phylogenetic hidden Markov model (HMM) to identify a specified number of independent evolutionary histories that most effectively account for

[1]Max Planck Institute for Terrestrial Microbiology, Department of Natural Products in Organismic Interactions, 35043 Marburg, Germany. [2]Molecular Biotechnology, Department of Biosciences, Goethe-University Frankfurt, 60438 Frankfurt, Germany. [3]Myria Biosciences AG, Tech Park Basel, Hochbergstrasse 60C, 4057 Basel, Switzerland. [4]Evolutionary Biochemistry Group, Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany. [5]Chair of Biochemistry, Center for Protein Assemblies, Technical University of Munich, Ernst-Otto-Fischer-Straße 8, 85748 Garching, Germany. [6]Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands. [7]Center for Synthetic Microbiology (SYNMIKRO), Phillips University Marburg, 35043 Marburg, Germany. [8]Department of Chemistry, Phillips University Marburg, 35043 Marburg, Germany. [9]LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG) & Senckenberg Gesellschaft für Naturforschung, 60325 Frankfurt, Germany.
***Corresponding author. Email: kenan.bozhueyuek@myria.bio (K.A.J.B.); helge.bode@mpi-marburg.mpg.de (H.B.B.)**
†These authors contributed equally to this work.
‡Present address: Synthetic Biology of Microbial Natural Products (SIMS), Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research, Saarland University Campus, 66123 Saarbrücken, Germany.

Corrected 22 March 2024. See full text.

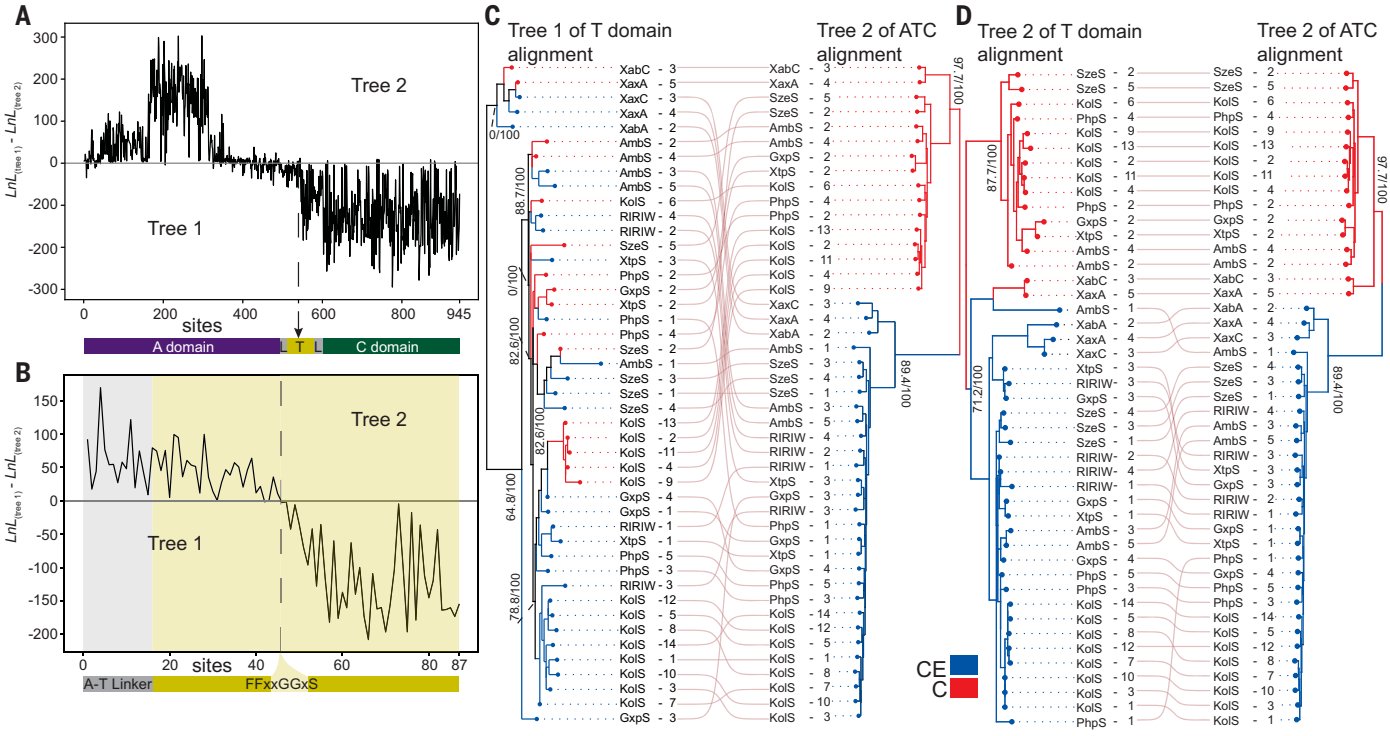


Fig. 1. Evolutionary analysis of ATC tridomains and T domains of representative NRPSs. (**A**) Site-wise log-likelihood difference plot of two phylogenetic trees of ATC tridomains that together best describe the alignment by using a phylogenetic hidden Markov model (HMM). Positive numbers indicate that sites are better described by tree 1, and negative numbers indicate sites that are better described by tree 2. (**B**) Site-wise log-likelihood difference plot as in (A), but for an alignment of T domains plus A-T linkers. The partitions into two trees detected by the HMM are separated with a dashed line. The partition lies around two conserved glycine residues, which are highlighted light yellow. (**C** and **D**) Comparison of tree 1 from the T domain alignment with tree 2 from the ATC tridomain alignment (left), and tree 2 from the T domain alignment with tree 2 from the ATC tridomain alignment (right). Names indicate the abbreviation of the NRPS, and numbers, the ATC tridomains within that NRPS. Lines connect the same NRPS and ATC tridomains between the two trees. Red branches label tridomains that contain $^{L}C_{L}$ domains, and blue branches, tridomains with CE (dual C) domains. Some clades are annotated with aLRT statistic and bootstrapping values (see materials and methods). Trees are shown as unrooted.

the alignment (*26*). Unlike a sliding window technique that infers phylogenies for short stretches along the sequence, this method simultaneously uses information from all positions in the alignment. We focused our analysis on identifying two distinct histories, expecting one to broadly align with established A domain trees and the other with recognized C domain phylogenies (*27*–*30*). To visualize the most likely history for each position, we computed the difference between site-wise logarithmic likelihoods under the two inferred trees. Positive values indicated a stronger fit with the first evolutionary history, and negative values, the second history. We verified the robustness of these assignments using an alternative method implemented in IQ-TREE, which is conceptually related to a phylogenetic HMM but instead computes the likelihood as a mixture model across prespecified trees (*31*).

The phylogenetic HMM approach was applied to two datasets. The first dataset consisted of 225 aligned amino acid sequences of NRPS ATC tridomains from various bacterial species, such as those of the genera *Photorhabdus* and *Xenorhabdus*, as well as representatives from firmicutes, actinomycetes, cyanobacteria, and other proteobacteria (data S1). The second dataset primarily focused on actinomycetes and comprised 72 aligned amino acid sequences of Leerzeichen NRPS ATC tridomains (data S2). The analysis of both datasets unveiled three key findings: (i) The second evolutionary history primarily encompassed the C domain along with the latter half of the T domain, with most sites in this region strongly aligning with this history (Fig. 1A and figs. S1 to S3). (ii) The first evolutionary history corresponded with a central stretch of the A domain, whereas (iii) sites at the C and N termini of this region, including the initial portion of the T domain, exhibited less pronounced alignment with either history. This pattern suggests infrequent recombination within the C domain and the central segment of the A domain. In line with prior analyses that describe A subdomain swaps as one driver for NRP diversification (*17*, *18*, *20*–*22*, *31*), our analysis suggested that recombination events are more frequently observed in the N and C termini of A domains. On the basis of our phylogenetic HMM analysis (Fig. 1A and figs. S1 to S3), this region extends toward approximately the latter half of the T domain, suggesting the existence of previously unidentified recombination sites within the catalytically inactive T domains. The extensive replacement of A subdomains for the purpose of rational modification in two existing recombinant (type S NRPS) assembly lines (*32*) resulted in the successful introduction of only conservative changes in amino acid specificity, albeit with good efficiency (see below). Therefore, we shifted our focus toward in-depth investigation of the putative evolutionary recombination site. For additional information on this experimental series, please refer to data S3.

To obtain a more profound understanding of the potential recombination sites identified in our initial analysis, we carried out a second phylogenetic HMM and multitree model approach (*26*, *31*), specifically targeting the T domain along with the A-T linker region (Fig. 1B and figs. S1 to S3). This procedure was chosen because the first half of the T domain exhibits weak alignment with either history in the initial analysis (Fig. 1A), likely because it does not perfectly share the evolutionary

Corrected 22 March 2024. See full text.

history of the A or C domains (Fig. 1C and figs. S4 and S5). This refined analysis revealed a distinct boundary between the two histories within the conserved FFxxGGxS motif of the T domain (Figs. 1B and figs. S1 to S7). We next used the output of the HMM to segment our alignment into these two histories by using both the Viterbi and Forward-Backward algorithms as implemented in PhyML_Multi (*26*). This detected a breakpoint before the two conserved glycines (fig. S1B and S3D). Notably, the second evolutionary history exhibited a topology akin to the C domain tree (Fig. 1D), with a clear division segregating the T domains based on the condensation reaction catalyzed by downstream C domains (*29*, *33*–*35*) (Fig. 1D and fig. S8). By contrast, the first evolutionary history did not mirror the C or A domain trees (Fig. 1C and fig. S5). Collectively, these observations (Fig. 1 and figs. S1 to S8) suggest that the A domain indeed serves as a recombination hotspot, as reported previously (*17*, *18*, *20*, *21*, *24*), but also that the T domain could potentially serve as a yet undescribed recombination site with an important breakpoint within the conserved FFxxGGxS motif (Figs. 1B and fig. S4).

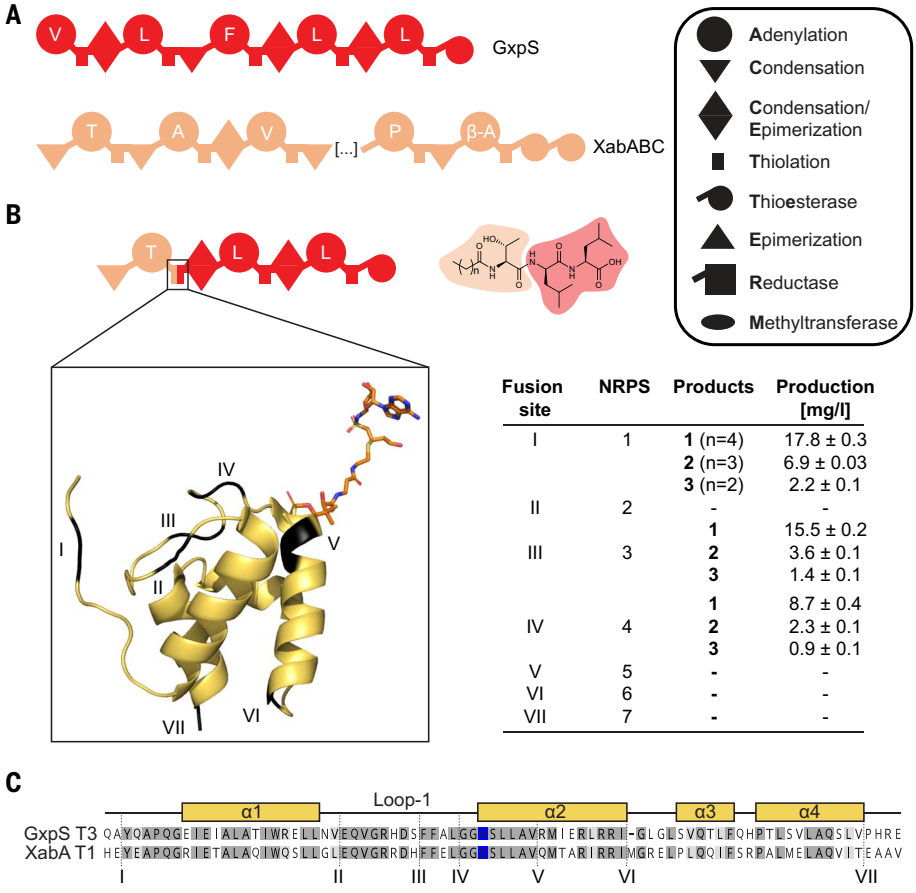
To reinforce our findings, we pursued an additional machine learning investigation involving 130,870 bacterial T domain sequences extracted from the antiSMASH database (https://antismash-db.secondarymetabolites.org) (data S4). Details concerning the workflow and results of this analysis can be found in the supplementary materials (fig. S9). Upon data curation, 77,424 sequences remained viable for analysis. These sequences were numerically transformed by using e-descriptor, a matrix that captures amino acid physicochemical properties through five values (*36*). The conducted principal component analysis (PCA) affirmed that, unlike the T domain's first half, the second half of the T domain exhibited distinct clustering based on subsequent domains [e.g., C domains accepting L-configured amino acids ($^{L}C_{L}$), condensation epimerization (CE), and E]. This finding not only corroborates the outcomes of our phylogenetic HMM analysis but also expands on previous results in a comprehensive manner (*37*, *38*).

To validate these in silico predictions and circumvent potential computational artifacts, we also subjected homologous NRPS involved in producing PAX (*39*, *40*), endopyrrole A (*41*), rhizomide A (*42*), and syringopeptin SP-25a (*43*) peptides to detailed analysis (figs. S10 to S13) (*26*). These results supported the hypothesis that recombination events within the T domain's conserved core motif (FFxxGGxS) do indeed occur (Fig. 1, A and B, and figs. S1B and S3D), leading to changes in stereochemistry (T-C versus T-CE), the interchange of T-TE domains, or modifications in the size of the NRPS system and the corresponding natural product scaffold. NRPS recombination events are thus not exclusive to the A domain but also involve the T domain, resulting in the creation of new NRPSs. Although we currently cannot determine the relative frequencies of recombination events in the T and A domains, this finding underscores the engineering potential of T domain recombinations.

Fusion point screening

The conserved core motif (FFxxGGxS) of each ~100–amino acid–spanning T domain region is located at the N terminus (loop1) of the second helix (α2), holding the invariant serine residue that becomes posttranslationally modified by a phosphopantetheinyl (Ppant) transferase (*44*–*46*). Although the T domain is the only NRPS domain without an autonomous catalytic activity, the attachment of a Ppant moiety is a functional prerequisite, not only to covalently bind activated extender units and the growing peptide chain but also to reach the active sites of A and C domains. Although our computational recombination analysis (figs. S1 and S3) identified a particular recombination site located just before the conserved glycines (FFxx||GGxS), we acknowledge that computational analysis may not pinpoint a single recombination site but rather indicates a sequence region likely to facilitate homologous recombination. Therefore, we conducted a fusion-point screening (Fig. 2) to validate fusion sites that yield optimal peptide production.

Unless stated otherwise, all NRPS and NRPS–polyketide synthase (PKS) hybrids described in ths study (data S5) were heterologously



Fusion site	NRPS	Products	Production [mg/l]
I	1	**1** (n=4)	17.8 ± 0.3
		2 (n=3)	6.9 ± 0.03
		3 (n=2)	2.2 ± 0.1
II	2	-	-
		1	15.5 ± 0.2
III	3	**2**	3.6 ± 0.1
		3	1.4 ± 0.1
		1	8.7 ± 0.4
IV	4	**2**	2.3 ± 0.1
		3	0.9 ± 0.1
V	5	-	-
VI	6	-	-
VII	7	-	-

Fig. 2. Fusion-point screening of a NRPS hybrid assembled within the T domain. (**A**) Schematic representation of precursor NRPSs GxpS (red) and XabABC (beige), which produce GameXPeptides and xenoamicins, respectively. Domain legend is depicted in the box. The A domain specificity is indicated by the single-letter amino acid code written within the schematic domain. (**B**) Schematic representation of the XabA-GxpS hybrid NRPS and produced lipopeptide and compounds titers. The color code of the peptide structure follows the NRPS color code from (A). The different fusion sites within the T domain are highlighted in black at their respective positions in the crystal structure of the T domain EntF (PDB 4ZXJ) (*83*). (**C**) Sequence alignment of GxpS T3 and XabA T1 with secondary structures of the T domain and fusion sites indicated. Grayscale indicates the identity and/or similarity of the sequences. The follow domain–correlated FFxxGG**x**S amino acid is highlighted in blue. Single-letter abbreviations for the amino acid residues referenced are as follows: R, Arg; K, Lys; I, Ile; P, Pro; F, Phe; Q, Gln; G, Gly; E, Glu; H, His; L, Leu; N, Asn; A, Ala; T, Thr; D, Asp; C, Cys; S, Ser; M, Met; Y, Tyr; W, Trp.

produced in *Escherichia coli* DH10B::*mtaA*, which encode the broad spectrum Ppant transferase MtaA (*47*). The resulting peptides (data S5 and table S5) and yields were confirmed by high-performance liquid chromatography–tandem mass spectrometry (HPLC-MS/MS) and comparison of retention times with synthetic standards (supplementary materials).

As a starting point, we chose GameXPeptide- (*48*) and xenoamicin-producing (*49*) synthetases (GxpS and XabABC, respectively) from *Photorhabdus luminescens* TTO1 and *Xenorhabdus stockiae* (Fig. 2A), respectively, to produce seven recombinant NRPSs (NRPS-1 to -7, Fig. 2B), each with a different fusion site (I to VII, Fig. 2C), with the in silico predicted breakpoint represented by fusion site IV. Briefly, this screening led to the identification of three functional fusion sites (I, III, and IV) in NRPS-1, -3, and -4, which all produced the expected lipopeptides **1** to **3**, differing only in the acyl starter originating from the fatty acid pool of *E. coli* and with titers between 1 and 18 mg L^{-1} (Fig. 2B, figs. S14 to S22, and table S6).

To uncover why fusion sites V to VII led to nonfunctional recombinant assembly lines, we conducted a comprehensive in silico analysis, again using a dataset of 130,870 bacterial T domain sequences. We aimed to determine whether the introduced changes from applying fusion sites V to VII rendered the engineered assembly lines NRPS-5 to -7 nonfunctional. The analysis focused on the amino acid position preceding the invariant serine residue (FFxxGG**x**S) in the T domain, which is known to be linked to the downstream C domain's catalyzed reaction (*37*, *38*). In our case (Fig. 2), we combined two T domains (XabA_T1 and GxpS_T3) followed by an $^{L}C_{L}$ domain and a CE domain, respectively. Our analysis revealed variations in this position for T domains, followed by $^{L}C_{L}$, TE, or R domains, with amino acids such as histidine, aspartate, or asparagine. Notably, a high conservation was observed, with aspartate common in T domains, followed by E domains and C domains accepting D-configured amino acids ($^{D}C_{L}$), and histidine present when the T domain was followed by CE domains, suggesting a role in interactions with the downstream domain (fig. S23). Fusion sites I, III, and IV were identified as preferable for a broadly applicable engineering approach, mitigating the risk of losing the specific interaction site after assembly. This finding might explain the observed lack of production for fusion sites V to VII in our experimental setup (Fig. 2C).

Our in silico observations (Fig. 1 and figs. S1 to S13) along with the results from the in vivo–conducted fusion site screening (Fig. 2) led us to the hypothesis that T-C-A units (fusion point I), $T_{1/2}$-C-A-$T_{1/2}$ units (fusion points III and IV), and combinations thereof may serve as ideal starting points to do evolution-inspired megasynthetase engineering, designated as the "eXchange Unit between T domains" (XUT) approach. However, after reviewing crystal structure data of A-T and T-C didomains, we decided to proceed with fusion sites I and IV because fusion sites III and IV are located directly adjacent (III) and within (IV) the conserved T domain motif, and the two variable positions in between the conserved motif (FF**xx**GGxS) are potentially contributing to a functional A-T interface (*50*).

Leveraging evolution-inspired eXchange Units for NRPS engineering

To verify the in silico–identified (IV) and in vivo–verified (I and IV) fusion sites on a broad scale, we targeted the NRPSs FitAB (Fig. 3) and FtrAB (fig. S24 and S51-S56 and Supplementary Table S11), which produce the NRPs fitayylide [**4**/**5**; a close derivative of xeinamide (*51*)] and faTTTVIR (named after their amino acid sequences) from *Xenorhabdus innexi* and *Xenorhabdus mauleonii*, respectively; GxpS (Fig. 4); the entolysin A–producing synthetase (EtlABC) from *Pseudomonas entomophila* (*52*); as well as the gramicidin S (GrsAB)– (*53*), tyrocidine (TycABC)– (*54*), bacitracin (BacABC)– (*55*), and surfactin (SrfABC)–producing synthetases (*56*) from Leerzeichen *Aneurinibacillus migulanus* ATCC 9999, *Brevibacillus brevis* ATCC 8185, *Bacillus licheniformis* ATCC 10716, and *Bacillus subtilis* MR168, respectively. We created 16 recombinant FitAB derivatives (NRPS-9 to -18) (Fig. 3, figs. S25 to S50, and tables S7 and S8), 10 GxpS derivatives (NRPS-19 to -28) (Fig. 4 and figs. S57 to S82), 1 FtrAB derivative (NRPS-34) (fig. S24), and 2 EtlABC derivatives (NRPS-35 and -36) (figs. S83 and S84) applying fusion sites I, IV, or both. The latter derivatives were produced in both *E. coli* DH10B::*mtaA* and *Pseudomonas putida* KT2440. Furthermore, we generated two chimeric NRPS enzymes combining GrsAB with TycABC (NRPS-37) (figs. S85 to S87) and BacABC with SrfABC (NRPS-38) (figs. S85 and S88), respectively.

The building blocks to engineer these assembly lines were selected to cover a broad range of bacterial (*Xenorhabdus*, *Photorhabdus*, *Serratia*, *Chondromyces*, *Myxococcus*, *Pseudomonas*, *Bacillus*, and *Streptomyces*) and fungal (*Mortierella*) genera with GC contents between 41 and 72% to test if the identified fusion sites have the potential to mimic horizontal gene transfer along with recombination on a rational scale suitable to reengineer NRPS.

We created 43 recombinant NRPSs. Notably, except for NRPS-16b (Fig. 3) and -28b (Fig. 4), both using fusion point IV, all recombinant NRPSs [including a NRPS-PKS assembly line in which the PKS is responsible for the polyunsaturated starter acyl moiety (NRPS-17a)] showed catalytic activity producing a broad range of cyclic and linear peptides (**4** to **67**) at titers ranging from ~0.1 (**40**, NRPS-24a) to ~168 mg L^{-1} (**34**, NRPS-22a) and from ~0.1 (**45**, NRPS-25b) to ~76 mg L^{-1} (**34**, NRPS-22b) for fusion sites I and IV, respectively (Figs. 3 and 4 and tables S12 and S13).

Comparing XUT with XUC and XU

The potential advantages of evolution-inspired engineering sites over A subdomain swap strategies are increased versatility and higher success rates. We also aimed to assess these engineering sites in comparison to the structure-based rational engineering strategies we had previously devised (*57*, *58*). These strategies included the "eXchange Unit" (XU) (*57*) and "eXchange Unit condensation domain" (XUC) concepts (*58*), which use ATC tridomain and C_{Asub}-A-T-C_{Dsub} units, respectively. In the latter case, C_{Asub} represents the acceptor site (approximately the second half), and C_{Dsub} represents the donor site (approximately the first half) of the pseudodimeric C domains.

To benchmark the XUT approach, we reconstructed NRPS-9 to -18 (Fig. 3, fig. S26, and table S12), NRPS-20 to -25 and NRPS-28 (Fig. 4 and figs. S57, S60 to S75, and S81 and S82), and NRPS-37 and NRPS-38 (figs. S85 to S88) also using the XU and XUC concepts and compared the relative peptide production yields as summarized in Fig. 5. This comparative approach highlights that both fusion sites of the evolution-inspired XUT as well as the XU (*57*) enabled us to create chimeric NRPSs from unrelated BGCs with respect to taxonomy (Gram-positive and Gram-negative bacteria) and GC content (~41 to ~71% GC). Other methods, such as A subdomain swaps (*24*, *25*) (data S3) and the XUC concept (*58*), enabled efficient reprogramming of NRPSs only within a narrow taxonomic range. Nevertheless, a correlation between the GC content of the introduced NRPS building blocks and peptide production can be observed (Figs. 3 and 4). Whereas building blocks of genera with a similar or slightly higher (~50 to ~65%) GC content (i.e., NRPS-13 and NRPS-14) (Fig. 3) were generally well tolerated, building blocks originating from the high-GC branch (≥70%; i.e., NRPS-17, -23, and -24) (Figs. 3 and 4) resulted in impaired assembly lines when recombined with NRPSs originating from the low-GC branch (~50%; i.e., NRPS from *Photorhabdus* and *Xenorhabdus*). The initial reduction of catalytic activity when building blocks of different GC content were recombined with each other might also occur naturally during homologous recombination after a horizontal gene transfer event.

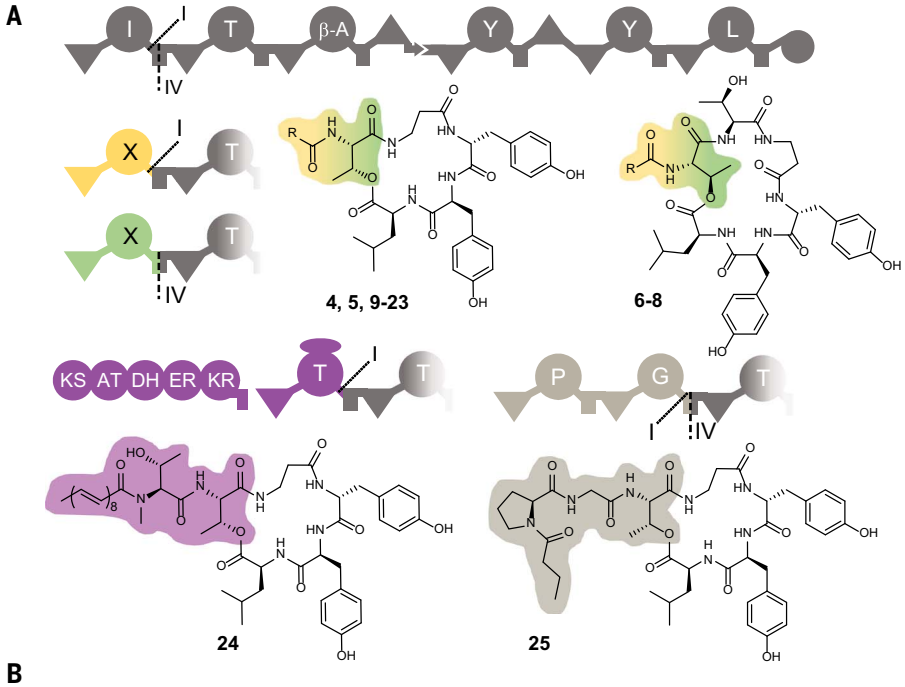
Evolution-inspired eXchange Units allow targeted peptide production

To validate XUT as an engineering approach, we designed de novo an artificial biosynthetic assembly line producing a biologically active peptide against a well-characterized target. We chose the eukaryotic proteasome as the target because it is involved in protein homeostasis

Corrected 22 March 2024. See full text.

Fig. 3. Evolution-inspired eXchange Units replacing NRPS starting modules.
(**A**) Schematic representation of the FitAB NRPS that produces fitayylide A (**4**) and B (**5**) and selected alternative starting modules from other NRPSs with indicated fusion sites I and IV. Amino acid specificities are assigned for all A domains. KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoylreductase. Selected structures of the produced peptides are shown so that, in conjunction with the table in (B), all peptide structures can be deduced. (**B**) Production data relative to the wild-type NRPS-8 and the absolute peptide yields are based on triplicate production cultures. The origin of the alternative starting module, their cognate gene cluster, and the fusion point for each starter module are indicated. Production was observed for all NRPS derivatives apart from NRPS-16b, but absolute production titers were not determined for all of them. For the latter cases, indication of production yield was indicated by + (modest) to +++ (very good). β-A, β-alanine; aa, amino acid; aa identity alignment was determined with ClustalW dual alignment of FitA first module to swapped module.

A



B

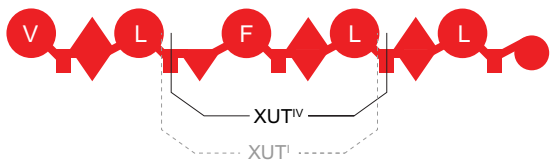
NRPS	Peptide	FA-peptide start (R)	Organism	Donor BGC	Fusion site	GC content (%)	aa identity alignment (%)	Production in % to NRPS-8	Production [mg/l]
8	**4** **5**	C2-Ile C2-Leu	*X. innexi*	*fitAB*	WT	44.9	100	100	122.7 ±1.5
9a	**6-8** **9-11**	C4/6/8 C4/6/8-Thr	*X. bovienii*	*txlAB*	XUT^{I}	43.5	51.7	46	56.2 ±3.1
9b	**6-8** **9-11**	C4/6/8 C4/6/8-Thr			XUT^{IV}	43.9	52.7	52	63.9 ±2.3
10a	**12-14**	C2/4/6-Lys	*X. doucetiae*	*prtAB*	XUT^{I}	48.2	51.2	+++	+++
10b	**12-14**	C2/4/6-Lys			XUT^{IV}	48.3	51.8	+++	+++.
11a	**15**	C2-Val	XKK7.4	*XEKKV2_1206*	XUT^{I}	45.2	67.0	36	44.4 ±2.5
11b	**15**	C2-Val			XUT^{IV}	44.8	67.0	54	66.5 ±3.3
12a	**16**	3OH C10-Ser	*X. indica*	*XINDV2_09420*	XUT^{I}	41.3	31.4	6	6.9 ±1.5
12b	**16**	3OH C10-Ser			XUT^{IV}	41.2	32.0	6	7.0 ±1.9
13a	**17/18**	3OH C8/10-Leu	*Serratia* SCBI	*swrA*	XUT^{I}	62.4	23.4	24	29.7 ±3.5
13b	**17/18**	3OH C8/10-Leu			XUT^{IV}	62.3	24.1	10	12.7 ±1.1
14a	**18** **19**	3OH C10-Leu Leu	*P. lurida*	*viscAB*	XUT^{I}	65.4	32.7	26	32.0 ±1.3
14b	**18**	3OH C10-Leu			XUT^{IV}	65.2	32.9	27	32.5 ±2.8
15a	**18** **19**	3OH C10-Leu Leu	*X. bovienii*	*xfpS*	XUT^{I}	47.6	25.9	24	29.8 ±3.4
15b	**18**	3OH C10-Leu			XUT^{IV}	47.7	26.7	22	26.7 ±3.1
16a	**20** **21** **22** **23**	3OH C16:1-Glu 3OH C16-Glu C16:1-Glu C16-Glu	*B. subtilis*	*ppsA-E*	XUT^{I}	47.4	30.7	++ ++ ++ ++	++ ++ ++ ++
16b	**20-23**		*B. subtilis*	*ppsA-E*	XUT^{IV}	47.7	31.5	-	-
17a	**24**	C18:8-(Me)Thr	*M. xanthus*	*mchABC*	XUT^{I}	70.1	21.7	+	+
18a	**25**	C4-Pro-Gly	*X. szentirmaii*	*xabABC*	XUT^{I}	51.8	33.2	34	42.5±1.4
18b	**25**	C4-Pro-Gly			XUT^{IV}	51.8	33.1	25	30.6±0.9

in the cell cycle, signal transduction, and general cell physiology. Proteasomes are a family of N-terminal nucleophilic hydrolases consisting of two sets of seven copies of α and β subunits that assemble into a barrel-shaped complex (Fig. 6C) (*59*). Peptides inhibiting the proteasome, such as the clinically used bortezomib (*60*), can lead to apoptosis, making the human proteasome a target for anticancer chemotherapy.

Inspired by nature, we decided to use the lipopeptide aldehyde fellutamide B (*61*) as the lead structure. This compound is not only active against the eukaryotic proteasome in humans and yeast but also the most potent inhibitor of the *Mycobacterium tuberculosis* proteasome tested to date. Fellutamide B consists of a C8-3OH acyl chain, L-Asn, L-Gln, and a L-Leu–aldehyde. The aldehyde moiety is responsible for the reversible binding to the active site threonine (Thr1) of the proteasome. As an alternative to TE domains, nature applies thioester reductase (R) domains (*62*–*64*), not only to release the synthesized peptide but also to introduce the aldehyde function by catalyzing an NAD(P)H-dependent two-electron reduction of the thioester.

Corrected 22 March 2024. See full text.

A



B

NRPS	Origin (NRPS)	GC content (%)	aa identity alignment (%)	A domain specificity	Fusion site	Products	Peptides	Production [mg/l]
19a	*X. innexi* (XdnAB)	50.4	21.8	Tyr	XUT^{I}	**26**	vLyL	1.65 ± 0.17
						27	lLyL	1.25 ± 0.09
20a	*X. mauleonii* (XabA)	53.9	50.1	Leu	XUT^{I}	**28**	vLlL	8.32 ± 1.4
						29	lLlL	5.23 ± 0.73
20b					XUT^{IV}	**28**	vLlL	1.12 ± 0.18
						29	lLlL	0.63 ± 0.18
21a	*X. mauleonii* (XabA)	54.1	48.2	Leu-Ile	XUT^{I}	**30**	vLllL	1.23 ± 0.11
						31	lLllL	0.17 ± 0.03
						32	cyclo(vLllL)	5.35 ± 0.49
						33	cyclo(lLllL)	2.30 ± 0.15
21b					XUT^{IV}	**30**	vLllL	1.23 ± 0.11
						31	lLllL	0.16 ± 0.06
						32	cyclo(vLllL)	5.87 ± 0.69
						33	cyclo(lLllL)	2.92 ± 0.33
22a	*X. indica* (XldS)	54.2	47.5	Ser	XUT^{I}	**34**	vLsL	168.03 ± 8.63
						35	lLsL	65.07 ± 4.46
22b					XUT^{IV}	**34**	vLsL	75.88 ± 2.49
						35	lLsL	29.74 ± 0.9
23a	*M. alpina* (MpbA)	54.3	52.5	Phe	XUT^{I}	**36**	vLfL	17.19 ± 1.02
						37	lLfL	8.33 ± 0.32
						38	vLwL	17.33 ± 1.15
						39	lLwL	7.86 ± 0.53
23b					XUT^{IV}	**36**	vLfL	13.4 ± 0.66
						37	lLfL	5.59 ± 0.32
						38	vLwL	24.06 ± 1.11
						39	lLwL	10.5 ± 0.48
24a	*M. xanthus* (MchB)	68.8	38.6	Ala	XUT^{I}	**40**	vLaL	0.1 ± 0.02
						41	lLaL	0.06 ± 0.03
24b					XUT^{IV}	**40**	vLaL	0.33 ± 0.06
						41	lLaL	0.21 ± 0.09
25b	*M. xanthus* (MchB)	69.0	33.8	Pro-Ala	XUT^{IV}	**42**	vLPaL	0.48 ± 0.03
						43	lLPaL	0.32 ± 0.04
						44	cyclo(vLPaL)	0.23 ± 0.02
						45	cyclo(lLPaL)	0.05 ± 0.01
26b	*P. entomophila (EtlC)*	69.3	46.3	Ser	XUT^{IV}	**34**	vLsL	8.08 ± 0.31
						35	lLsL	3.50 ± 0.02
27a	*Streptomyces sp.* (GriA)	69.7	38.4	Leu	XUT^{I}	**28**	vLlL	1.89 ± 0.06
						29	lLlL	0.53 ± 0.03
27b					XUT^{IV}	**28**	vLlL	0.43 ± 0.01
						29	lLlL	0.23 ± 0.03
28a	*C. crocatus* (CpnD)	71.1	42.5	Leu	XUT^{I}	**28**	vLlL	1.63 ± 0.17
						29	lLlL	0.73 ± 0.16
28b					XUT^{IV}	**-**	-	-

Fig. 4. Evolution inspired eXchange Units replacing internal modules.
(**A**) Schematic representation of the precursor NRPS GxpS producing Game-XPeptides. A T2 to T4 fragment was exchanged against XUTs from different NRPSs from various organisms. (**B**) NRPSs, the origin of the inserted eXchange Unit (XdnAB, XabA, XldS, MpbA, MchB, EtlC, GriA, and CpnD), corresponding fusion sites (XUT^{I} and XUT^{IV}), variability in GC content spanning from 50.4% (*xdnAB*) to 71.1% (*cpnD*), the respective A domain specificity, peptides, and production titers of the respective peptides. Peptide sequences are indicated by the single-letter codes, capital letters being L-configured amino acids, and small letters, D-configured amino acids. aa identity alignment was determined with ClustalW dual alignment of exchanged GxpS modules 2 and 3 to swapped modules when replaced by single-modular blocks only; module 2 from GxpS was used for the alignment. -, no production.

In a proof-of-concept experiment, we used a combination of XU and XUT fusion sites to design an artificial three-module assembly line in silico, incorporating NRPS building blocks originating from five distinct origins (Fig. 6a) with the NRPS donor fragments XabA and KolS being derived from a previously published recombinant NRPS (*57*): a C_{start} domain to introduce the acyl chain and A domains with specificities (N- to C-terminus) for L-Gln (A1), L-Ala (A2), and L-Leu (A3). To achieve the reduction of leucine into an aldehyde, the R domain of the tilivalline-producing NRPS (XtvB) (*65*) from *Xenorhabdus indica* was used as the termination domain using the fusion sites I, III, IV, and VII (Fig. 6 and figs. S89 to S96). Owing to comparative reasons, we also tested a homologous XtvB R domain from *Xenorhabdus eapokensis* DL20, by using a previously used (*65*) fusion site two amino acid positions upstream (VII_{-2}) of fusion site VII (NRPS-39, fig. S96). The

Corrected 22 March 2024. See full text.

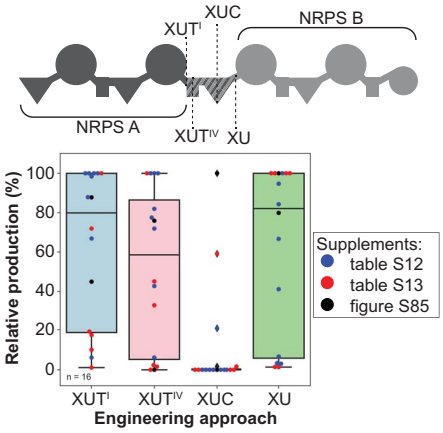


Fig. 5. NRPS engineering approach comparison. Boxplot of relative productions of artificial NRPSs assembled by the XUT^{I}, XUT^{IV}, XUC, or XU approach. Each of the 16 artificial NRPSs was generated with all four approaches, which were quantitatively compared among the respective NRPS with the highest production set to 100% relative production. Dots within the plot indicate different series of experiments, which are described in detail in the supplementary materials and datasets.

resulting assembly lines NRPS-29 to -32 and NRPS-39 all showed catalytic activity producing the desired lipopeptide aldehydes **46** to **48**, differing only in the acyl group used as a starter, with titers between ~1 and ~22 mg L^{-1} (Fig. 6B, figs. S89 to S96, and table S9). Compared with the NRPSs generated with fusion sites I (NRPS-29), VII (NRPS-32), and VII_{-2} (NRPS-39), the NRPSs generated with fusion sites III (NRPS-26) and IV (NRPS-27) produced about a fivefold-higher peptide titer. Whereas low titers of NRPS-32 and -39 were in good agreement with our initial fusion-point screening (Fig. 2), the amount of produced peptide by NRPS-29 was unexpectedly low. The impaired formation of a functional A-T domain-domain interface (*50*) in the case of NRPS-29 and a functional T-R (*62*) domain-domain interface in the case of NRPS-32 and -39 could serve as an explanation for this result, as also observed previously (*65*, *66*). Furthermore, these results highlight the advantages of the evolution-inspired fusion sites III and IV compared with fusion sites I and VII, which are located within the A-T and T-C linker regions, respectively.

To test whether the lipopeptide aldehyde **47** is indeed able to inhibit the yeast 20S proteasome core particle (yCP) by binding the active site Thr1, the half-maximum inhibitory concentration (IC_{50}) and cocrystallization of yCP together with **47** (yCP:**47** complex) was performed (Fig. 6, C to E; fig. S97; and table S10). Both experiments confirmed the expected activity of **47** against the yCP β5 subunit at 3.6 ± 0.8 μM and a binding mode to Thr1 equivalent to that of fellutamide A and B (Fig. 6, C to E). This proof-of-concept experiment not only demonstrated the efficient introduction of a reactive aldehyde moiety through the XUT approach but also highlighted the capability to biosynthetically create tailor-made bioactive peptides de novo with rational drug design strategies.

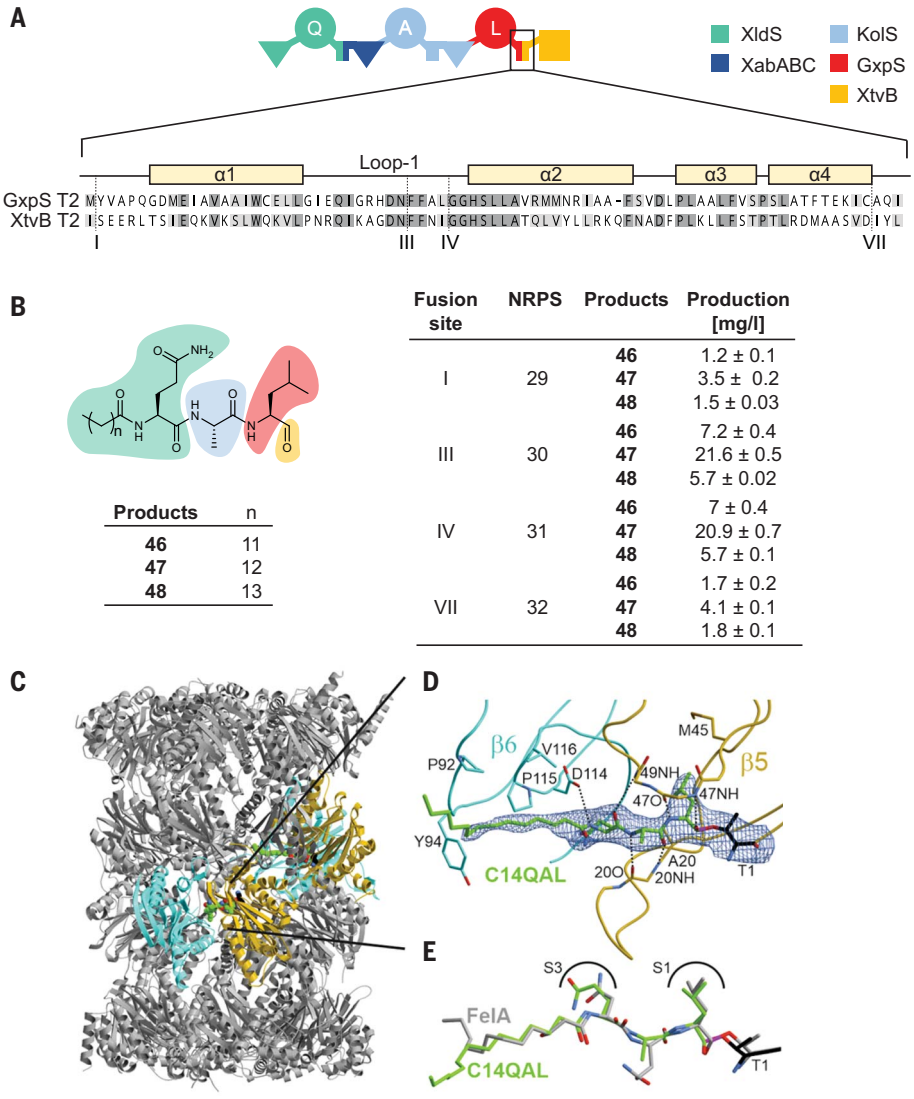


Fig. 6. XUT approach for the design of a proteasome inhibitor. (**A**) Schematic representation of reassembled NRPS-29 to -32 composed of NRPS fragments from XldS, XabA, KolS (kolossin synthetase), GxpS, and XtvB. The terminal T domain is shown as a sequence alignment of GxpS T2 and XtvB T2, indicating secondary structures, fusion sites, and similarity and/or identity in gray. (**B**) Production titers corresponding to the fusion sites within the terminal T domain. The color code in the peptides follows that of the NRPS fragments used and is shown in (A). (**C**) Crystal structure of the yeast 20S proteasome in complex with **47** (spherical model, green carbon atoms), bound to the chymotrypsin-like active sites (β5 subunits, gold; PDB ID 8BW1). (**D**) Illustration of the $2F_O–F_C$ electron density map (blue mesh, contoured to 1σ) of **47** (depicted as C14QAL), covalently linked through a hemiacetal bond (magenta) to Thr1O$^{\gamma}$. Protein residues interacting with **47** are highlighted in black. Dots illustrate hydrogen bonds between **47** and protein residues. (**E**) Superposition of **47** (depicted as C14QAL) (green) and fellutamide A (gray; PDB ID s3D29) (*84*) complex structures, highlighting similar conformations at the chymotrypsin-like active site.

Conclusions

The engineering of assembly line enzymes, particularly NRPSs, has posed formidable challenges (*67*) despite technical advancements and a deep understanding of their biochemical and structural properties (*4*). Nature, on the other hand, has masterfully navigated the intricate terrain of biosynthetic pathway engineering through BGC evolution and serves as inspiration for our XUT approach (*17*, *18*, *21*, *22*, *68*, *69*).

This method, which mimics horizontal gene transfer and recombination events, broadens structural diversity through rational engineering, surpassing the limitations of natural diversity. Our work, exemplified by the design of an artificial proteasome inhibitor, underscores the impact of evolution-driven strategies on NRPS engineering and the broader landscape of synthetic biology. In particular, the potential of the XUT approach to efficiently change the assembly lines after T domains to other than C domains, as demonstrated by reductase domain insertion (Fig. 6), complements previous XU and XUC approaches that introduce repetitive NRPS modules (*57*, *58*).

In contrast to prior efforts that focused on deciphering precise mechanisms behind contemporary NRPS families (*17*, *22*, *28*, *68*, *69*), our study found new fusion sites for NRPS engineering inspired by evolution. By unveiling a previously unexplored recombination site within the conserved core motif of T domains, represented by fusion sites III and IV, the XUT approach introduces the $T_{1/2}$-C-A-$T_{1/2}$ architecture. Aligned with recent structural revelations (*50*) and the recently proposed unified model for NRP evolution (*17*), our findings illuminate an additional recombination site beyond A subdomain exchanges (*18*, *21*, *24*, *25*). The relative obscurity of this recombination site in previous studies (*68*) is probably attributable to variations in dataset composition and methodological divergence characterized by the integration of phylogenetic HMMs and a PCA-based machine learning approach in our research (*26*). Upon reanalyzing datasets describing NRPS evolution (*43*, *68*), we unequivocally identify the T domain as an additional recombination site (figs. S13 and S98), offering a fresh perspective on the dynamics of NRPS evolution.

Materials and methods

Cultivation of strains

All *E. coli* DH10B::*mtaA* cells were cultured either on liquid or solid low salt LB medium (pH 7.5, 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl). Either kanamycin (50 μg/ml), chloramphenicol (34 μg/ml), gentamicin (20μg/ml) or spectinomycin (50 μg/ml) were added as selection markers. Solid media contained 1% (w/v) agar. Cells were cultivated at 37°C and at 22°C for peptide production cultures.

Cloning of biosynthetic gene clusters and NRPS modules

For use as template, genomic DNA (gDNA) was extracted from bacteria indicated in Table S1 by use of the Gentra Puregene Yeast/Bact. Kit (Qiagen) and the Monarch® Genomic DNA Purification Kit (NEB) which in turn was taken as template for the PCR amplification. The proof-reading PCR polymerase Q5® High-Fidelity DNA Polymerase (NEB) and *Phusion* DNA Polymerase (NEB/Thermo Fisher Scientific) in their standard and hot start variations were employed. Oligonucleotides for the PCR and the correct product size are documented in Table S4. In specified cases (Table S4) already cloned NRPS parts were used as template for the PCR. PCR products were agarose gel purified taking the Monarch® DNA Gel Extraction Kit (NEB) to be used as substrate for the Gibson cloning procedure using the Gibson Assembly® Master Mix or the NEBuilder® HiFi DNA Assembly Cloning Kit (NEB). In cases indicated in Table S4 restriction enzyme digests with enzymes indicated were used as one part of the substrate for the Gibson cloning step.

The vector pCK_0407 was cloned in a classic fashion. To this end the plasmid pCK_0407 was linearised using the restriction enzymes AvrII/XbaI and the 1.750 bp fragment ligated to the 1.933 bp fragment of the AvrII/XbaI digest of pCDFDuet (Merck-Novagen).

Heterologous expression of NRPS constructs and HPLC-MS analysis

After plasmid transformation into *E. coli* DH10B::*mtaA* or in *P. putida* KT2440, cells were grown overnight in LB medium containing all necessary antibiotics (50 μg/ml kanamycin, 34 μg/ml chloramphenicol, 100 μg/ml spectinomycin, 20 μg/ml gentamicin). 10 ml LB medium containing antibiotics, 0.002 mg/ml L-arabinose or 0.002 mg/ml vanillic acid and 2% (v/v) XAD-16 beads were inoculated with 1% overnight grown culture. After incubation for 72 hours at 22°C, XAD-16 beads were harvested and one culture volume methanol was added. Methanol extraction was conducted for 60 min at 22°C. The organic phase was filtrated and diluted 1:10 in methanol. Cleared HPLC-UV-MS analysis was conducted on an UltiMate 3000 system (Thermo Fisher) coupled to an AmaZonX mass spectrometer (Bruker) with an ACQUITY UPLC BEH C18 column (130 Å, 2.1 mm × 100 mm, 1.7-μm particle size, Waters) at a flow rate of 0.4 ml min−1 (5–95% acetonitrile/water with 0.1% formic acid, vol/vol, 16 min, UV detection wavelength 190–800 nm). HPLC-UV-HRMS analysis was conducted on an UltiMate 3000 system (Thermo Fisher) coupled to an Impact II qTof mass spectrometer (Bruker) with an ACQUITY UPLC BEH C18 column (130 Å, 2.1 mm × 100 mm, 1.7-μm particle size, Waters) at a flow rate of 0.4 ml min−1 16 min, UV detection wavelength 190–800 nm). Evaluation was performed using DataAnalysis 5.3 software (Bruker).

For peptide quantification of NRPS-8 to -18 and -66 and -67, the production medium was, deviating from above, XPP medium (*70*) without phenylalanine but with 1 mM β-alanine added. For peptide production and quantification of NRPS-19 to -28 and -35 to -36, the production medium was XPP medium.

Peptide purification

Compounds **4**, **5**, **7**, **10**, **47**, and **50** were produced in *E. coli* DH10B::*mtaA* expressing the respective NRPS variants. 4L XPP medium containing 34 μg/ml chloramphenicol, 0.02% L-arabinose and 2% XAD-16 beads was inoculated with 1% overnight grown culture as described in section S1.3. The culture was incubated at 180 rpm for 72 hours at 22°C. Subsequently, the XAD-16 beads were extracted 3 times with 500 ml methanol for 30 min, stirring. The solvent was fully removed at reduced pressure. The crude extract of compound **47** was completely solved in DMSO in order to purify it by preparative HPLC–MS (LC-MS-System 1260 Infinity II Preparative LC/MSD from Agilent). A C3 column (Agilent ZORBAX 300XB-C3) utilizing a gradient of 40-55% ACN/H_2O (+0.1% formic acid) was used. The compound was freeze-dried and the purity of the compound determined by NMR and HPLC-HR-MS.

The crude extract containing compound **4** and **5** were purified using the Agilent Prep HPLC-MS using a C18 column (Agilent C-18 22.10x250 mm) applying a 30-50% ACN/H_2O (+0.1% formic acid, 20 ml/min) gradient over 32 min. The retention time for **5** was 11.7 min and for **4** 12.8 min. The crude extract containing compound **7** and **10** were purified using the Agilent Prep HPLC-MS using a C18 column (Agilent C-18 22.10x250 mm, 20 ml/min) applying the solvent mixture 42% ACN/H_2O (+0.1% formic acid) over 30 min. The retention time for **7** was 9.0 min and the retention time for **10** was 9.6 min. The crude extract containing compound **26** was purified using the Agilent Prep HPLC-MS using a C18 column (Agilent C-18 22.10x250 mm, 20 ml/min) applying a 20-30% ACN/H_2O (+0.1% formic acid) gradient over 26 min. The retention time of 26 was 19.5-20.5 min.

Peptide quantification

The absolute production titers were calculated as previously described (*71*). Therefore, calibration curves based on pure **1a** (for quantification of **1**, **2** and **3**), **4** (for **4**, **5**, **15**, **17** and **18**), **10** (**6**, **7**, **8**, **9**, **10**, **11** and **16**), **67** (**67**, **68**, **69** and **70**), **26a** (for **26**, **27**, **38**, **39**), **28b** (for **28**, **29**, **36**, **37**, **40** and **41**), **34** (for **34** and **35**), **42a** (for **30**, **31**, **42**, **43**), **44** (for **32**, **33**, **44**, **45**) and **47** (for **46**, **47** and **48**), were prepared. The pure compounds were prepared at different concentrations: **1a** utilizing a standard curve with concentrations of 5000, 500, 50, 5 and 0.5 μg L^{-1}; **4** utilizing a standard curve with concentrations of 10, 4, 1, 0.4, 0.1, 0.04 and 0.01 mg L^{-1}, **10** utilizing a standard curve with concentrations of 10, 4, 1, 0.4, 0.1 and 0.04 mg L^{-1}, **67** utilizing a standard curve with concentrations of 40, 4, 0.4, 0.04 and 0.004 mg L^{-1}, **50**, **51**, **52** and **53** utilizing a standard curve with concentrations of 100, 20, 4, 0.8 and 0.16 mg L^{-1}, **47** utilizing a

standard curve with concentrations 10, 5, 2.5, 1.25, 0.625, 0.3125 and 0.1562 mg L^{-1} and measured by LC-MS using HPLC/MS measurements as described above. To ensure sample signals being within the range of the standard curve they were diluted when necessary. The peak area for each compound at different concentrations was calculated using Compass Data Analysis and used for the calculation of a standard curve passing through the zero point. Triplicates of all in vivo experiments were measured. The pure peptide standards **1a**, **26, 28, 30, 34, 42,** and **44** were synthesized in-house, **4, 10**, **47** and **50** were purified from production cultures.

Chemical synthesis

The linear peptide **1a** (C6/D-Thr/D-Leu/L-Leu) was synthesized on preloaded resin (0.25 mmol H-Leu-2ClTrt PS resin, Sigma Aldrich, Germany) by solid phase peptide synthesis using standard Fmoc/*t*-Bu chemistry. Fmoc protected amino acids or fatty acids were activated by mixture of 5 eq. Fmoc-AA-OH (or fatty acid), 12 eq. N,N-diisopropylethylamine (DIPEA, Iris Biotech, c = 2.4 M), 5 eq. O-(7-azabenzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HATU, Carbolution Chemicals) in 15 ml dimethylformamide (DMF, Carl Roth, Germany). The resin was incubated with the activated amino acid/fatty acid mixture for 2 hours at room temperature. After each coupling, the resin was washed with NMP (5 ×), DMF (5 ×) and DCM (5 ×). Finally, the peptide was cleaved by addition of 20 ml of a mixture of hexafluoroisopropanol (HFIP) and DCM (1:4 v/v). Subsequently, the peptide was deprotected upon addition of 2 ml Trifluoroacetic acid (TFA) incubating for 2 hours at room temperature. The linear peptide was dissolved in MeOH in order to purify it by semi-preparative HPLC–MS (Agilent LC-MS-System 1260 Infinity II Analytical-Scale LC/MSD) utilizing a C18 column (Eclipse XDB-C18 (9.4 × 250 mm, 5 μm). The purity was determined by NMR and HPLC-HR-MS analysis.

Chemical synthesis of peptides **1**, **26a, 28a, 28b, 30b, 34, 36, 40a, 40b, 42a, 44, 60** and **77** was performed using the CEM Liberty Prime 2.0 automated microwave peptide synthesizer with a preloaded Wang Resin from Sigma Aldrich. Synthesis was performed in two stages, deprotection with 25% pyrrolidine in DMF for 40 s at 110°C. Coupling reactions were performed twice with a 5 fold excess of Fmoc-AA-OH with 1:1:1.5 AA/DIC/Oxyma for 1 min at 110°C. After addition of the final amino acid and deprotection step, the resin was washed with DMF (5 ×) and DCM (5 ×). Cleavage was performed using 95:2.5:2.5 TFA/H_2O/TIS for 30 min at 38°C. Following cleavage, the peptide was dried at the V10 (Biotage) and washed two times with MeOH. The purity was determined by HPLC-HR-MS and NMR.

Expression and purification of yeast 20 S proteasome

The yeast 20S proteasome was prepared as previously described (*72*).

IC_{50} value determination with purified yCP

Concentration of purified yeast 20 S proteasome (yCP) was determined spectrophotometrically at 280 nm. yCP (final concentration: 0.05 mg/ml in 100 mM Tris-HCl, pH 7.5) was mixed with DMSO as a control or serial dilutions of fellutamide derivatives in DMSO, thereby not surpassing a final concentration of 10% (v/v) DMSO. After an incubation time of 45 min at RT, fluorogenic substrates Boc-Leu-Arg-Arg-AMC, Z-Leu-Leu-Glu-AMC and Suc-Leu-Leu-Val-Tyr-AMC (final concentration of 200 μM) were added to measure the residual activity of caspase-like (C-L, β1 subunit), trypsin-like (T-L, β2 subunit) and chymotrypsin-like (ChT-L, β5 subunit), respectively. The assay mixture was incubated for another 60 min at RT and afterwards diluted 1:10 in 20 mM Tris-HCl, pH 7.5. The AMC-molecules released by hydrolysis were measured in triplicate with a Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies) at λ_{exc} = 360 nm and λ_{em}=460 nm. Relative fluorescence units were normalized to the DMSO treated control. The calculated residual activities were plotted against the logarithm of the applied inhibitor concentration and fitted with GraphPad Prism 5. Half maximum inhibitory concentration (IC_{50}) values were deduced from the fitted data. They depend on enzyme concentration and are comparable within the same experimental settings.

Crystallization and structure determination of the yeast 20S proteasome core particle (yCP) in complex with 47

Crystals of yCP were grown in hanging drops at 20°C as previously described (*73*). The protein concentration used for crystallization was 40 mg/mL in Tris / HCl (20 mM, pH 7.5) and EDTA (1 mM). The drops contained 1 μl of protein and 1 μl of the reservoir solution [30 mM magnesium acetate, 100 mM 2-(N-morpholino) ethanesulfonic acid (pH 6.8) and 10% (wt/vol) 2-methyl-2,4-pentanediol]. Crystals appeared after two days and were incubated with a fellutamide derivative at final concentrations of 10 mM for at least 24 hours. Droplets were then complemented with a cryoprotecting buffer [30% (wt/vol) 2-methyl-2,4-pentanediol, 15 mM magnesium acetate, 100 mM 2-(N-morpholino)ethanesulfonic acid, pH 6.9] and vitrified in liquid nitrogen. The dataset from the yCP: **47** complex was collected using synchrotron radiation (λ =1.0 Å) at the X06SA-beamline (Swiss Light Source, Villingen, Switzerland). X-ray intensities and data reduction were evaluated using the XDS program package (Table S10) (*74*). Conventional crystallographic rigid body, positional, and temperature factor refinements were carried out with REFMAC5 (*75*) using coordinates of the yCP structure as starting model (PDB ID 5CZ4) (*76*). For model building, the programs SYBYL and COOT (*77*) were used. The final coordinates yielded excellent R factors, as well as geometric bond and angle values. Coordinates were confirmed to fulfill the Ramachandran plot and have been deposited in the RCSB (PDB ID 8BW1).

Evolutionary analysis of ATC tridomains from NRPS using PhyML_Multi

The amino acid sequence of NRPS were collected from our *Photorhabdus* and *Xenorhabdus* genome collection. 152 ATC tridomain sequences stem from *Photorhabdus* and *Xenorhabdus*. We also included in our analysis three ATC tridomain sequences from actinomycetes, three from cyanobacteria and 20 came from proteobacteria other than *Photorhabdus* and *Xenorhabdus*. ATC tridomains from NRPS protein sequences were extracted from our NRPS dataset using local BLAST with the second ATC tridomain from GxpS of *Photorhabdus laumondii* TTO1 as query. Sequences were aligned using MUSCLE v3.8.31 (*78*) and trimmed with trimAl v1.2 (*79*). This alignment was used for the evolutionary analysis using the software PhyML_Multi. We specified that PhyML_Multi search for two trees under a HMM that together best fit the alignment. Since PhyML_Multi does not have a model finder, the model finder of IQ-tree (*79*) with the selection of '-msub nuclear' was used. IQ-tree chose JTT (*80*) as the best fit model which was also used for the analysis with PhyML_Multi with a four-category gamma distribution of among site rate-variation. Afterwards, the log likelihood of tree 1 was deducted from the log likelihood of tree 2 and plotted.

The same dataset used for the PhyML_Multi analysis was also used for the MAST analysis. First, we ran the analysis with two trees generated from the PhyML_Multi analysis. We used as the best fit model for both trees JTT with a 4-category gamma distribution of among site rate-variation. Again, the log likelihood value of tree 1 was deducted from the log likelihood of tree 2 and plotted. Also, we operated the same analysis but now feeding it with an A and C domain tree that was individually inferred using iqtree. The corresponding log likelihood values from both trees were deducted from each other and plotted.

Evolutionary analysis of the T domain from NRPS using PhyML_Multi

The T domain dataset covered the amino acid sequence of the A-T linker and the T domain. This area was extracted from our NRPS dataset using local BLAST with the third T domain from GxpS of *Photorhabdus laumondii* TTO1 as query. The T domains were aligned using

MUSCLE and carefully trimmed manually to reduce gaps. Afterwards, the software PhyML_Multi was used to detect recombination breakpoints and phylogenetic histories within the T domain (*26*).

Here, the MAST analysis was used as well. First, we ran the analysis with two trees generated by the PhyML_Multi run. The best fit model to run the analysis for both trees was JTT with a 4-category gamma distribution of among site rate-variation. The log likelihood value of tree 1 was deducted from the log likelihood of tree 2 and plotted. Afterwards, we inferred trees of the first half of the T domain (that is until breakpoint IV) and a tree of the second half of the T domain. These two trees were used as an input for the second MAST analysis. The corresponding log likelihood values from both trees were deducted from each other and plotted.

Topological comparison of different phylogenetic trees

The four different trees generated by PhyML_Multi were pruned using the software mesquite (*81*) to reduce the number branches on the trees for visual clarity. Trees were compared using the R package phytools (*82*). To test the robustness of each topology, we first divided the alignment into partitions separated at breakpoint 4. This was necessary, because neither bootstrapping nor likelihood-based confidence measures are implemented in PhyML_Multi or Mast, which calculate joint likelihoods across topologies. We then performed nonparametric bootstraps and calculated SH-aLRT statistics for both topologies using iqtree. The number of nonparametric bootstraps was 500 for the A domain, 500 for the C domain and 100 for both parts of the T domain. For the SH-alrt statistics, 1000 replicates were used.

REFERENCES AND NOTES

1. D. J. Newman, G. M. Cragg, Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020). doi: 10.1021/acs.jnatprod.9b01285; pmid: 32162523
2. S. Huang, F. Ba, W. Q. Liu, J. Li, Stapled NRPS enhances the production of valinomycin in *Escherichia coli*. *Biotechnol. Bioeng.* **120**, 793–802 (2023). doi: 10.1002/bit.28303; pmid: 36510694
3. F. David *et al*., A Perspective on Synthetic Biology in Drug Discovery and Development-Current Impact and Future Opportunities. *SLAS Discov.* **26**, 581–603 (2021). doi: 10.1177/24725552211000669; pmid: 33834873
4. R. D. Süssmuth, A. Mainz, Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed.* **56**, 3770–3821 (2017). doi: 10.1002/anie.201609079; pmid: 28323366
5. M. F. Byford, J. E. Baldwin, C.-Y. Shiau, C. J. Schofield, The Mechanism of ACV Synthetase. *Chem. Rev.* **97**, 2631–2650 (1997). doi: 10.1021/cr960018l; pmid: 11851475
6. K. Tahlan, M. A. Moore, S. E. Jensen, δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase (ACVS): Discovery and perspectives. *J. Ind. Microbiol. Biotechnol.* **44**, 517–524 (2017). doi: 10.1007/s10295-016-1850-7; pmid: 27766439
7. J. Zhang, A. L. Demain, ACV synthetase. *Crit. Rev. Biotechnol.* **12**, 245–260 (1992). doi: 10.3109/07388559209069194; pmid: 1633622
8. H. Umezawa, Structure and action of bleomycin. *Prog. Biochem. Pharmacol.* **11**, 18–27 (1976). pmid: 63962
9. A. Fahr, Cyclosporin clinical pharmacokinetics. *Clin. Pharmacokinet.* **24**, 472–495 (1993). doi: 10.2165/00003088-199324060-00004; pmid: 8513650
10. C. T. Walsh, R. V. O'Brien, C. Khosla, Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed.* **52**, 7098–7124 (2013). doi: 10.1002/anie.201208344; pmid: 23729217
11. S. Caboche, V. Leclère, M. Pupin, G. Kucherov, P. Jacques, Diversity of monomers in nonribosomal peptides: Towards the prediction of origin and biological activity. *J. Bacteriol.* **192**, 5143–5150 (2010). doi: 10.1128/JB.00315-10; pmid: 20693331
12. K. A. Bozhüyük, J. Micklefield, B. Wilkinson, Engineering enzymatic assembly lines to produce new antibiotics. *Curr. Opin. Microbiol.* **51**, 88–96 (2019). doi: 10.1016/j.mib.2019.10.007; pmid: 31743841
13. K. Blin, S. Shaw, S. A. Kautsar, M. H. Medema, T. Weber, The antiSMASH database version 3: Increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res.* **49** (D1), D639–D643 (2021). doi: 10.1093/nar/gkaa978; pmid: 33152079
14. K. Blin *et al*., antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49** (W1), W29–W35 (2021). doi: 10.1093/nar/gkab335; pmid: 33978755
15. C. L. M. Gilchrist *et al*., cblaster: A remote search tool for rapid identification and visualization of homologous gene clusters. *Bioinform. Adv.* **1**, vbab016 (2021). doi: 10.1093/bioadv/vbab016; pmid: 36700093
16. P. Hirsch *et al*., ABC-HuMi: The Atlas of Biosynthetic Gene Clusters in the Human Microbiome. *Nucleic Acids Res.* **52** (D1), D579–D585 (2024). doi: 10.1093/nar/gkad1086; pmid: 37994699
17. M. Baunach, S. Chowdhury, P. Stallforth, E. Dittmann, The Landscape of Recombination Events That Create Nonribosomal Peptide Diversity. *Mol. Biol. Evol.* **38**, 2116–2130 (2021). doi: 10.1093/molbev/msab015; pmid: 33480992
18. T. J. Booth *et al*., Bifurcation drives the evolution of assembly-line biosynthesis. *Nat. Commun.* **13**, 3498 (2022). doi: 10.1038/s41467-022-30950-z; pmid: 35715397
19. A. Nivina, K. P. Yuet, J. Hsu, C. Khosla, Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **119**, 12524–12547 (2019). doi: 10.1021/acs.chemrev.9b00525; pmid: 31838842
20. M. J. Calcott, J. G. Owen, D. F. Ackerley, Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *Nat. Commun.* **11**, 4554 (2020). doi: 10.1038/s41467-020-18365-0; pmid: 32917865
21. M. Crüsemann, C. Kohlhaas, J. Piel, Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chem. Sci.* **4**, 1041–1045 (2013). doi: 10.1039/C2SC21722H
22. S. Meyer *et al*., Biochemical Dissection of the Natural Diversification of Microcystin Provides Lessons for Synthetic Biology of NRPS. *Cell Chem. Biol.* **23**, 462–471 (2016). doi: 10.1016/j.chembiol.2016.03.011; pmid: 27105282
23. A. Nivina, S. Herrera Paredes, H. B. Fraser, C. Khosla, GRINS: Genetic elements that recode assembly-line polyketide synthases and accelerate their diversification. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2100751118 (2021). doi: 10.1073/pnas.2100751118; pmid: 34162709
24. H. Kries, D. L. Niquille, D. Hilvert, A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol.* **22**, 640–648 (2015). doi: 10.1016/j.chembiol.2015.04.015; pmid: 26000750
25. W. L. Thong *et al*., Gene editing enables rapid engineering of complex antibiotic assembly lines. *Nat. Commun.* **12**, 6872 (2021). doi: 10.1038/s41467-021-27139-1; pmid: 34824225
26. B. Boussau, L. Guéguen, M. Gouy, A mixture model and a hidden markov model to simultaneously detect recombination breakpoints and reconstruct phylogenies. *Evol. Bioinform. Online* **5**, 67–79 (2009). doi: 10.4137/EBO.S2242; pmid: 19812727
27. R. He *et al*., Knowledge-guided data mining on the standardized architecture of NRPS: Subtypes, novel motifs, and sequence entanglements. *PLOS Comput. Biol.* **19**, e1011100 (2023). doi: 10.1371/journal.pcbi.1011100; pmid: 37186644
28. S. Noriler, J. C. Navarro-Muñoz, C. Glienke, J. Collemare, Evolutionary relationships of adenylation domains in fungi. *Genomics* **114**, 110525 (2022). doi: 10.1016/j.ygeno.2022.110525; pmid: 36423773
29. C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson, Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78 (2007). doi: 10.1186/1471-2148-7-78; pmid: 17506888
30. T. Stachelhaus, H. D. Mootz, M. A. Marahiel, The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999). doi: 10.1016/S1074-5521(99)80082-9; pmid: 10421756
31. T. K. F. Wong *et al*., MAST: Phylogenetic Inference with Mixtures Across Sites and Trees. bioRxiv 2022.10.06.511210 [Preprint] (2022); https://doi.org/10.1101/2022.10.06.511210.
32. K. A. J. Bozhueyuek, J. Watzel, N. Abbood, H. B. Bode, Synthetic Zippers as an Enabling Tool for Engineering of Non-Ribosomal Peptide Synthetases. *Angew. Chem. Int. Ed.* **60**, 17531–17538 (2021). doi: 10.1002/anie.202102859; pmid: 34015175
33. M. J. Wheadon, C. A. Townsend, Evolutionary and functional analysis of an NRPS condensation domain integrates β-lactam, D-amino acid, and dehydroamino acid synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026017118 (2021). doi: 10.1073/pnas.2026017118; pmid: 33893237
34. N. Ziemert *et al*., The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLOS ONE* **7**, e34064 (2012). doi: 10.1371/journal.pone.0034064; pmid: 22479523
35. S. R. Messenger *et al*., Metagenomic domain substitution for the high-throughput modification of nonribosomal peptides. *Nat. Chem. Biol.* **20**, 251–260 (2024). doi: 10.1038/s41589-023-01485-1; pmid: 37996631
36. W. Braun, M. S. Venkatarajan, New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. *J. Mol. Model.* **7**, 445–453 (2001). doi: 10.1007/s00894-001-0058-5
37. U. Linne, S. Doekel, M. A. Marahiel, Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases. *Biochemistry* **40**, 15824–15834 (2001). doi: 10.1021/bi011595t; pmid: 11747460
38. J. G. Owen, M. J. Calcott, K. J. Robins, D. F. Ackerley, Generating Functional Recombinant NRPS Enzymes in the Laboratory Setting via Peptidyl Carrier Protein Engineering. *Cell Chem. Biol.* **23**, 1395–1406 (2016). doi: 10.1016/j.chembiol.2016.09.014; pmid: 27984027
39. S. W. Fuchs, A. Proschak, T. W. Jaskolla, M. Karas, H. B. Bode, Structure elucidation and biosynthesis of lysine-rich cyclic peptides in *Xenorhabdus nematophila*. *Org. Biomol. Chem.* **9**, 3130–3132 (2011). doi: 10.1039/c1ob05097d; pmid: 21423922
40. T. D. Vo, C. Spahn, M. Heilemann, H. B. Bode, Microbial Cationic Peptides as a Natural Defense Mechanism against Insect Antimicrobial Peptides. *ACS Chem. Biol.* **16**, 447–451 (2021). doi: 10.1021/acschembio.0c00794; pmid: 33596038
41. S. P. Niehs *et al*., Genome Mining Reveals Endopyrroles from a Nonribosomal Peptide Assembly Line Triggered in Fungal-Bacterial Symbiosis. *ACS Chem. Biol.* **14**, 1811–1818 (2019). doi: 10.1021/acschembio.9b00406; pmid: 31283172
42. X. Wang *et al*., Discovery of recombinases enables genome mining of cryptic biosynthetic gene clusters in Burkholderiales species. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4255–E4263 (2018). doi: 10.1073/pnas.1720941115; pmid: 29666226
43. A. Isogai *et al*., Structural analysis of new syringopeptins by tandem mass spectrometry. *Biosci. Biotechnol. Biochem.* **59**, 1374–1376 (1995). doi: 10.1271/bbb.59.1374; pmid: 7670202
44. J. R. Lai, A. Koglin, C. T. Walsh, Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* **45**, 14869–14879 (2006). doi: 10.1021/bi061979p; pmid: 17154525
45. S. A. Sieber, M. A. Marahiel, Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chem. Rev.* **105**, 715–738 (2005). doi: 10.1021/cr0301191; pmid: 15700962
46. Z. Zhou, J. R. Lai, C. T. Walsh, Directed evolution of aryl carrier proteins in the enterobactin synthetase. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11621–11626 (2007). doi: 10.1073/pnas.0705122104; pmid: 17606920
47. O. Schimming, F. Fleischhacker, F. I. Nollmann, H. B. Bode, Yeast homologous recombination cloning leading to the novel peptides ambactin and xenolindicin. *ChemBioChem* **15**, 1290–1294 (2014). doi: 10.1002/cbic.201402065; pmid: 24816640
48. F. I. Nollmann *et al*., Insect-specific production of new GameXPeptides in *photorhabdus luminescens* TT01, widespread natural products in entomopathogenic bacteria. *ChemBioChem* **16**, 205–208 (2015). doi: 10.1002/cbic.201402603; pmid: 25425189

49. B. Behsaz *et al*., Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery. *Nat. Commun.* **12**, 3225 (2021). doi: 10.1038/s41467-021-23502-4; pmid: 34050176
50. J. M. Reimer *et al*., Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, eaaw4388 (2019). doi: 10.1126/science.aaw4388; pmid: 31699907
51. A. Yoshimura *et al*., *Membrane-Vesicle-Mediated Interbacterial Communication Activates Silent Secondary Metabolite Production* (Angewandte Chemie, International Edition, 2023).
52. H. B. Bode *et al*., Determination of the absolute configuration of peptide natural products by using stable isotope labeling and mass spectrometry. *Chemistry* **18**, 2342–2348 (2012). doi: 10.1002/chem.201103479; pmid: 22266804
53. J. Krätzschmar, M. Krause, M. A. Marahiel, Gramicidin S biosynthesis operon containing the structural genes grsA and grsB has an open reading frame encoding a protein homologous to fatty acid thioesterases. *J. Bacteriol.* **171**, 5422–5429 (1989). doi: 10.1128/jb.171.10.5422-5429.1989; pmid: 2477357
54. H. D. Mootz, M. A. Marahiel, The tyrocidine biosynthesis operon of Bacillus brevis: Complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843–6850 (1997). doi: 10.1128/jb.179.21.6843-6850.1997; pmid: 9352938
55. K. Eppelmann, S. Doekel, M. A. Marahiel, Engineered biosynthesis of the peptide antibiotic bacitracin in the surrogate host Bacillus subtilis. *J. Biol. Chem.* **276**, 34824–34831 (2001). doi: 10.1074/jbc.M104456200; pmid: 11448966
56. O. V. Kisil, V. S. Trefilov, V. S. Sadykova, M. E. Zvereva, E. A. Kubareva, Surfactin: Its Biological Activity and Possibility of Application in Agriculture. *Appl. Biochem. Microbiol.* **59**, 1–13 (2023). doi: 10.1134/S0003683823010027
57. K. A. J. Bozhüyük *et al*., De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018). doi: 10.1038/nchem.2890; pmid: 29461518
58. K. A. J. Bozhüyük *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). doi: 10.1038/s41557-019-0276-z; pmid: 31182822
59. W. Baumeister, J. Walz, F. Zühl, E. Seemüller, The proteasome: Paradigm of a self-compartmentalizing protease. *Cell* **92**, 367–380 (1998). doi: 10.1016/S0092-8674(00)80929-0; pmid: 9476896
60. K. Scott, P. J. Hayden, A. Will, K. Wheatley, I. Coyne, Bortezomib for the treatment of multiple myeloma. *Cochrane Database Syst. Rev.* **4**, CD010816 (2016). doi: 10.1002/14651858.CD010816.pub2; pmid: 27096326
61. G. Lin, D. Li, T. Chidawanyika, C. Nathan, H. Li, Fellutamide B is a potent inhibitor of the *Mycobacterium tuberculosis* proteasome. *Arch. Biochem. Biophys.* **501**, 214–220 (2010). doi: 10.1016/j.abb.2010.06.009; pmid: 20558127
62. S. Deshpande, E. Altermann, V. Sarojini, J. S. Lott, T. V. Lee, Structural characterization of a PCP-R didomain from an archaeal nonribosomal peptide synthetase reveals novel interdomain interactions. *J. Biol. Chem.* **296**, 100432 (2021). doi: 10.1016/j.jbc.2021.100432; pmid: 33610550
63. K. L. Kavanagh, H. Jörnvall, B. Persson, U. Oppermann, Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **65**, 3895–3906 (2008). doi: 10.1007/s00018-008-8588-y; pmid: 19011750
64. B. Persson, Y. Kallberg, Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs). *Chem. Biol. Interact.* **202**, 111–115 (2013). doi: 10.1016/j.cbi.2012.11.009; pmid: 23200746
65. A. Tietze, Y. N. Shi, M. Kronenwerth, H. B. Bode, Nonribosomal Peptides Produced by Minimal and Engineered Synthetases with Terminal Reductase Domains. *ChemBioChem* **21**, 2750–2754 (2020). doi: 10.1002/cbic.202000176; pmid: 32378773
66. M. J. Calcott, D. F. Ackerley, Portability of the thiolation domain in recombinant pyoverdine non-ribosomal peptide synthetases. *BMC Microbiol.* **15**, 162 (2015). doi: 10.1186/s12866-015-0496-3; pmid: 26268580
67. A. S. Brown, M. J. Calcott, J. G. Owen, D. F. Ackerley, Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* **35**, 1210–1228 (2018). doi: 10.1039/C8NP00036K; pmid: 30069573
68. S. Götze *et al*., Structure elucidation of the syringafactin lipopeptides provides insight in the evolution of nonribosomal peptide synthetases. *Chem. Sci.* **10**, 10979–10990 (2019). doi: 10.1039/C9SC03633D; pmid: 32953002
69. M. H. Medema, P. Cimermancic, A. Sali, E. Takano, M. A. Fischbach, A systematic computational analysis of biosynthetic gene cluster evolution: Lessons for engineering biosynthesis. *PLOS Comput. Biol.* **10**, e1004016 (2014). doi: 10.1371/journal.pcbi.1004016; pmid: 25474254
70. E. Bode *et al*., Promoter Activation in Δ*hfq* Mutants as an Efficient Tool for Specialized Metabolite Production Enabling Direct Bioactivity Testing. *Angew. Chem.* **131**, 19133–19139 (2019). doi: 10.1002/ange.201910563
71. K. A. J. Bozhüyük, J. Watzel, N. Abbood, H. B. Bode, Synthetic Zippers as an Enabling Tool for Engineering of Non-Ribosomal Peptide Synthetases. *Angew. Chem.* **133**, 17672–17679 (2021). doi: 10.1002/ange.202102859
72. N. Gallastegui, M. Groll, Analysing properties of proteasome inhibitors using kinetic and X-ray crystallographic studies. *Methods Mol. Biol.* **832**, 373–390 (2012). doi: 10.1007/978-1-61779-474-2_26; pmid: 22350899
73. M. Groll, R. Huber, Purification, crystallization, and X-ray analysis of the yeast 20S proteasome. *Methods Enzymol.* **398**, 329–336 (2005). doi: 10.1016/S0076-6879(05)98027-0; pmid: 16275340
74. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010). doi: 10.1107/S0907444909047337; pmid: 20124692
75. T. R. Schneider, G. M. Sheldrick, Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779 (2002). doi: 10.1107/S0907444902011678; pmid: 12351820
76. E. M. Huber *et al*., A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nat. Commun.* **7**, 10900 (2016). doi: 10.1038/ncomms10900; pmid: 26964885
77. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi: 10.1107/S0907444910007493; pmid: 20383002
78. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). doi: 10.1093/nar/gkh340; pmid: 15034147
79. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009). doi: 10.1093/bioinformatics/btp348; pmid: 19505945
80. D. T. Jones, W. R. Taylor, J. M. Thornton, The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992). doi: 10.1093/bioinformatics/8.3.275; pmid: 1633570
81. W. Maddison, D. R. Maddison, Mesquite: A modular system for evolutionary analysis. *Evolution* **62**, 1103–1108 (2008).
82. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012). doi: 10.1111/j.2041-210X.2011.00169.x
83. E. J. Drake *et al*., Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016). doi: 10.1038/nature16163; pmid: 26762461
84. J. Hines, M. Groll, M. Fahnestock, C. M. Crews, Proteasome inhibition by fellutamide B induces nerve growth factor synthesis. *Chem. Biol.* **15**, 501–512 (2008). doi: 10.1016/j.chembiol.2008.03.020; pmid: 18482702

ACKNOWLEDGMENTS

We thank the staff of beamline X06SA at the Paul Scherrer Institute, SLS, Villigen (Switzerland), for their assistance during data collection and are grateful to all Bode lab members for continuous discussions on further developments of NRPS engineering methods. **Funding:** This work was supported by an ERC Advanced grant (835108); the LOEWE Research Center TBG, funded by the State of Hesse; and the Max-Planck Society (to H.B.B.); as well as the German Research Foundation (SFB1035, project no. 201302640, A02 to W.K. and M.G.). **Author contributions:** K.A.J.B., L.P., C.K., L.S., M.S., and T.M.M. planned and performed all NRPS engineering experiments and peptide productions. J.K. performed A subdomain swaps and S.K., C.K., and G.K.A.H. performed all phylogenetic analyses. C.S. synthesized and purified all peptide standards used for quantification, and Y.-N.S. and L.P. isolated and elucidated the structures of selected natural products. W.K. and M.G. performed proteasome assays and crystallization. M.A. extracted the T domain dataset from the antiSMASH Database, which was analyzed by L.P. With input from all authors, K.A.J.B. and H.B.B. conceived all experiments and wrote the paper. **Competing interests:** A patent describing the XUT approach was filed by the Goethe University Frankfurt (application no. EP22190641), of which H.B.B., K.A.J.B., and C.K. are listed as inventors. K.A.J.B. and H.B.B. are cofounder and shareholder of Myria Biosciences AG, respectively, of which K.A.J.B. is also CSO. **Data and materials availability:** Coordinates of structural data have been deposited in the RCSB (PDB ID 8BW1). Material (strains and plasmids) described in this manuscript is available from H.B.B. upon request.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg4320
Tables S1 to S13
Figs. S1 to S98
References (*85*–*102*)
MDAR Reproducibility Checklist
Data S1 to S5

Submitted 23 December 2022; resubmitted 11 November 2023
Accepted 9 February 2024
10.1126/science.adg4320